



UNIVERSIDAD DE SANTIAGO DE CHILE

FACULTAD DE CIENCIA

DEPARTAMENTO DE MATEMÁTICAS Y CIENCIAS DE LA COMPUTACIÓN

INGENIERÍA ESTADÍSTICA

SEMINARIO DE TESIS II

---

# Análisis Espacio-Temporal de la Tasa de Contagios de Covid-19 en base a la Movilidad en la Región Metropolitana, Chile

---

Darinka Emilith Carvajal Miranda

## *Comisión Evaluadora*

*Felipe Elorrieta López*

*Claudio Vargas Rona*

*Pablo Villalobos Dintrans*

*Profesor Guía, Facultad de Ciencias, USACH*

*Facultad de Ciencias, USACH*

*Facultad de Ciencias Médicas, USACH*

# Resumen

El mundo actualmente se encuentra en pandemia a causa de un virus llamado SARS-Cov2, del que cada día se van descubriendo nuevas cepas. Su efecto puede medirse a partir del número de contagios y las variaciones de este indicador, se ven influenciadas por factores como la movilidad. Este último favorece la propagación del virus pues, no se respetan las medidas sanitarias puestas por el gobierno, como el distanciamiento social. Esta tesis presenta un análisis espacio-temporal de la tasa de contagios en base a la movilidad de las personas, considerando las comunas de la Región Metropolitana (RM), en tiempos de pandemia de Covid-19. Para esto, dentro del análisis teórico realizado, se comienza con una revisión de conceptos previos relacionados con el análisis espacio-temporal, se continúa con una revisión literaria acerca de los modelos espacio-temporales y se revisan los test de asociación espacial y serial. Se prosigue con el análisis práctico, donde se parte realizando un análisis descriptivo, mostrando gráficos de líneas y mapas de la región, se continúa analizando las posibles dependencias con el espacio y tiempo, corroborando que los datos poseen dependencia espacial y serial y por último, se implementan los modelos que incorporan los efectos de dependencia espacial y serial.

Con este estudio, se logra conocer la relación entre la tasa de contagio y las 2 variables de movilidad consideradas. La dependencia entre ellas es baja, pero se acentúa si los datos se consideran por periodo. Además se obtiene que hay una fuerte dependencia espacial en las variables de movilidad y en la tasa de contagios y también existe una fuerte dependencia temporal, por lo que, el modelo que considera estas componentes en el modelo es un Modelos Espacial de Datos de Panel, Correlación Espacial y Serial, SEMRE.

# Índice general

Resumen	I
Lista de figuras	V
Lista de tablas	VII
<b>1. Introducción</b>	<b>1</b>
<b>2. Objetivos</b>	<b>4</b>
2.1. Objetivo General . . . . .	4
2.2. Objetivos Específicos . . . . .	4
<b>3. Metodología</b>	<b>5</b>
<b>4. Marco Teórico</b>	<b>7</b>
4.1. Estadística Espacial . . . . .	7
4.1.1. Datos de área (Lattice Data) . . . . .	8
4.2. Series de tiempo . . . . .	10
4.2.1. Datos temporales . . . . .	11
4.3. Datos espacio-temporales . . . . .	11
4.3.1. Datos de Panel . . . . .	11
4.4. Autocorrelación o Dependencia Espacial . . . . .	12
4.4.1. Criterio de Vecindad . . . . .	13
4.4.2. Matriz de Contigüidad o Pesos Espaciales . . . . .	14
4.5. Test de Asociación Espacial . . . . .	17
4.5.1. Test de Asociación Espacial Global . . . . .	18
4.5.2. Test de Asociación Espacial Local . . . . .	19
4.6. Autocorrelación o Dependencia Temporal o Serial . . . . .	20

4.7. Test de Asociación Temporal . . . . .	21
4.7.1. Test de Box-Ljung . . . . .	21
4.8. Modelos Espacio-Temporales . . . . .	21
4.9. Modelos Espaciales . . . . .	22
4.10. Modelos Espaciales con Datos de Panel . . . . .	26
4.10.1. Modelos de Panel Espaciales con Efectos Aleatorios, Correlación Espacial y Correlación Serial . . . . .	27
4.10.2. Modelos de Panel Espaciales con Retraso Espacial, Efectos Aleatorios, Corre- lación Espacial y Correlación Serial . . . . .	31
4.10.3. Consideraciones de la Implementación de los modelos en R . . . . .	32
4.10.4. Pruebas del Multiplicador de Lagrange (Pruebas LM) . . . . .	33
4.11. Indicadores para la evaluación del Modelo . . . . .	36
<b>5. Datos</b>	<b>38</b>
5.1. Naturaleza de los datos . . . . .	38
<b>6. Análisis Práctico</b>	<b>41</b>
6.1. Análisis Descriptivo de los Datos . . . . .	41
6.1.1. Gráficos de las variables . . . . .	42
6.1.1.1. Gráfico de Líneas Tasa de Contagios . . . . .	42
6.1.1.2. Gráfico de Líneas Variación de Movilidad Completa . . . . .	43
6.1.1.3. Gráfico de Líneas Variación de Movilidad Laboral . . . . .	44
6.2. Mapas de la Región Metropolitana . . . . .	45
6.2.1. Mapas de las Tasas de Contagios . . . . .	46
6.2.2. Mapas de la Variación de Movilidad Completa . . . . .	48
6.2.3. Mapas de la Variación de Movilidad Laboral . . . . .	50
6.3. Relación entre las Variables . . . . .	53
6.3.1. Tasa de Contagios v/s Variación de Movilidad . . . . .	53
6.3.2. Correlaciones entre las Variables . . . . .	56
6.4. Análisis de Autocorrelación Espacial . . . . .	57
6.4.1. Cálculo de la Matriz de Pesos Espaciales . . . . .	57
6.4.2. Análisis de Autocorrelación Espacial Global . . . . .	59
6.5. Análisis de Autocorrelación Temporal . . . . .	61

<i>ÍNDICE GENERAL</i>	IV
<b>7. Análisis Práctico de los Modelos Espaciales con Datos de Panel</b>	<b>66</b>
7.1. Implementación de las Pruebas LM . . . . .	67
7.2. Parámetros Estimados de cada Modelo . . . . .	68
<b>8. Conclusión</b>	<b>72</b>
<b>9. Bibliografía</b>	<b>74</b>
<b>Apéndices</b>	<b>76</b>
<b>A. Gráficos de Líneas de cada Variable</b>	<b>77</b>
A.1. Gráfico de Líneas, Tasa de Contagios . . . . .	77
A.2. Gráfico de Líneas, Variación de Movilidad Completa . . . . .	78
A.3. Gráfico de Líneas, Variación de Movilidad Laboral . . . . .	78
<b>B. Mapas de los Quintiles</b>	<b>79</b>
B.1. Mapas de los Quintiles de la Tasa de Contagios . . . . .	79
B.2. Mapas de los Quintiles de la Variación de Movilidad Completa . . . . .	88
B.3. Mapas de los Quintiles de la Variación de Movilidad Laboral . . . . .	97
<b>C. Correlaciones, período entre olas</b>	<b>106</b>
<b>D. Tablas con los p-valores del test de Ljung-Box</b>	<b>107</b>
D.1. P-valores del test de Ljung-Box considerando todas las semanas y algunas comunas	107
D.2. P-valores del test de Ljung-Box considerando el periodo 1, para algunas comunas . .	112
D.3. P-valores del test de Ljung-Box considerando el periodo 2, para algunas comunas . .	112
D.4. P-valores del test de Ljung-Box considerando el periodo 3, para algunas comunas . .	113
<b>E. Tablas de las Pruebas LM</b>	<b>114</b>
<b>F. P-valores de los parámetros estimados</b>	<b>115</b>
F.1. Considerando el Periodo Total . . . . .	115
F.2. Considerando la Primera Ola . . . . .	116
F.3. Considerando el Periodo entre Olas . . . . .	117
F.4. Considerando la Segunda Ola . . . . .	118

# Lista de Figuras

4.1. Unidades espaciales de área de la región $R$ . . . . .	9
4.2. Mapa de Galicia dividido en áreas regulares e irregulares . . . . .	9
4.3. Serie de tiempo de datos mensuales . . . . .	10
4.4. Autocorrelación Positiva . . . . .	13
4.5. Autocorrelación Negativa . . . . .	13
4.6. Tipos de Contigüidad . . . . .	16
4.7. Modelos Espaciales . . . . .	25
6.1. Gráfico de Líneas de la Tasa de Contagios para algunas comunas . . . . .	43
6.2. Gráfico de Líneas Variación de Movilidad Completa para algunas comunas . . . . .	44
6.3. Gráfico de Líneas Variación de Movilidad Laboral para algunas comunas . . . . .	45
6.4. Leyenda para los mapas de la Tasa de Contagios . . . . .	46
6.5. Quintiles de la Tasa de contagios de la Región Metropolitana, para algunas semanas	47
6.6. Leyenda para los mapas de la Variación de Movilidad Completa . . . . .	48
6.7. Quintiles de la Variación de Movilidad Completa, para algunas semanas . . . . .	49
6.8. Leyenda para los mapas de la Variación de Movilidad Laboral . . . . .	50
6.9. Quintiles de la Variación de Movilidad Laboral para algunas semanas . . . . .	52
6.10. Tasa de Contagios v/s Variación de Movilidad para algunas comunas . . . . .	54
6.11. Correlaciones entre las variables Tasa de Contagios y Variación de Movilidad . . . . .	57
6.12. Mapa de las comunas vecinas de Alhué, Región Metropolitana . . . . .	58
6.13. Valores del Índice de Moran Global significativos y no significativos . . . . .	60
6.14. P-valores del test de Ljung-Box para la comuna de Las Condes . . . . .	62
6.15. P-valores del test de Ljung-Box considerando el primer periodo . . . . .	63
6.16. P-valores del test de Ljung-Box considerando el primer periodo . . . . .	63
6.17. P-valores del test de Ljung-Box considerando el segundo periodo . . . . .	64
6.18. Gráficos de líneas de la Tasa de Contagios, para el periodo 2 . . . . .	65

6.19. P-valores del test de Ljung-Box considerando el tercer periodo . . . . .	65
7.1. Pruebas LM, considerando el periodo total . . . . .	67
7.2. Parámetros Estimados, considerando el Periodo Total . . . . .	69
7.3. Parámetros Estimados, considerando la Primera Ola . . . . .	70
7.4. Parámetros Estimados, considerando el Periodo entre Olas . . . . .	70
7.5. Parámetros Estimados, considerando la Segunda Ola . . . . .	71
D.1. P-valores de 10 lags para la serie de Santiago . . . . .	107
D.2. P-valores de 10 lags para la serie de Las Condes . . . . .	108
D.3. P-valores de 10 lags para la serie de Ñuñoa . . . . .	108
D.4. P-valores de 10 lags para la serie de Quinta Normal . . . . .	109
D.5. P-valores de 10 lags para la serie de Vitacura . . . . .	109
D.6. P-valores de 10 lags para la serie de Puente Alto . . . . .	110
D.7. P-valores de 10 lags para la serie de La Florida . . . . .	110
D.8. P-valores de 10 lags para la serie de Pudahuel . . . . .	111
D.9. P-valores de 10 lags para la serie de Cerrillos . . . . .	111
D.10.P-valores de 1 lag para cada serie seleccionada . . . . .	112
D.11.P-valores de 1 lag para cada serie seleccionada . . . . .	112
D.12.P-valores de 1 lag para cada serie seleccionada . . . . .	113

# Lista de Tablas

4.1. Ejemplo estructura de Datos de Panel . . . . .	12
4.2. Tipos de Modelos Espaciales . . . . .	24
4.3. Tipos de Modelos Espaciales de Panel con Efectos Aleatorios, Correlación Espacial y Correlación Serial . . . . .	28
4.4. Tipos de Modelos Espaciales de Panel, Retardo Espacial, Efectos Aleatorios, Correlación Espacial y Correlación Serial . . . . .	32

# Capítulo 1

## Introducción

A principios del año 2020, el mundo comenzó a vivir una crisis sanitaria, humana y económica debido a la pandemia del Covid-19. Cada día habían más casos de contagios y no existía vacuna que generara inmunidad contra el virus.

El coronavirus es una familia de virus que causa enfermedades tanto en animales como en humanos. En humanos, causa infecciones respiratorias, desde el resfriado común hasta enfermedades más graves como la neumonía, el síndrome respiratorio de Oriente Medio (MERS) y el síndrome respiratorio agudo severo (SARS). El SARS-Cov2 (coronavirus tipo 2) descubierto más recientemente, causa el Covid-19. (OMS, 2020). Esta cepa es una enfermedad infecciosa que no se conocía previamente.

El brote del Covid-19 se originó en la ciudad de Wuhan, China, en diciembre del 2019. (CuidePlus, 2020). La Organización Mundial de la Salud (OMS) y la Organización Panamericana de la Salud, reconocieron el estado de pandemia que supone el desarrollo de esta enfermedad. (OMS, 2020). Actualmente, es una pandemia que afecta a casi todo el mundo y se están descubriendo nuevas cepas.

En este último tiempo, diversos estudios han destacado múltiples causas naturales que tienen relación con el virus. Se ha investigado que la demografía de los países, el efecto de las condiciones de salud, y el impacto de condiciones socioeconómicas, son factores para explicar los resultados relacionados con el Covid-19. (Villalobos, Castillo, De la Fuente y Maddaleno, 2020). Sin embargo, en general estos estudios se han realizado desde un punto de vista espacial o temporal y hay pocos estudios que abarcan estos dos aspectos en conjunto, tanto espacio como tiempo. Esto se debe prin-

principalmente, a la escasez de softwares que trabajan con el tipo de datos espacio-temporal, y además que la recolección de los datos es difícil. Aunque existen paquetes para estimar modelos espaciales transversales en R, Stata y Matlab, son escasos los paquetes que estiman modelos espaciales de datos de panel. (Millo y Piras, 2012, citado en Carracedo y Debón, 2017)

El efecto del virus se puede medir a partir de indicadores como el número de contagios, el número de hospitalizaciones o el número de fallecidos. Según los informes epidemiológicos de Covid-19 del año 2020 publicados por el Ministerio de Salud (MINSAL), las variaciones en los contagios se vieron influenciados por diversos factores que favorecen la propagación del virus, tales como, no respetar las medidas sanitarias que se imponen y los altos índices de movilidad. Los datos de los contagios varían según espacio y tiempo, estas dos dimensiones ¿aportan significativamente en la evolución de los contagios?, ¿los contagios pueden ser explicados a través de la movilidad?. La primera interrogante, plantea la existencia de una relación significativa entre el número de contagios con el espacio y el tiempo, y la segunda plantea una asociación significativa entre las variables, que también varía de acuerdo al espacio y al tiempo. Por ende, la metodología considerada debe incorporar ambas dimensiones en el análisis.

Para el estudio, es interesante considerar las comunas de la Región Metropolitana (RM) como dimensión espacial, pues es la región más poblada de Chile. Y como dimensión temporal, se consideran las semanas epidemiológicas (toman como inicio de semana el día domingo) desde el inicio de los casos en la región, hasta los datos más actualizados que se tengan. Con estos datos, se tiene la suposición de 2 hipótesis. La primera dice que el aumento de los casos de contagios inició en el sector oriente de la RM y se fue moviendo a la zona sur-poniente y la segunda indica que durante las primeras semanas de la pandemia, la movilidad era mayor en el sector sur-poniente en comparación al sector oriente de la RM.

Según Elhorst (2014), la Econometría Espacial es un subcampo de la Econometría que se ocupa de los efectos de interacción espacial entre las unidades geográficas, ya sea, ciudades, regiones, países, etc. dependiendo de la naturaleza del estudio. La Econometría Espacial es de gran utilidad cuando se considera el uso de variables vinculadas al espacio. (citado en Castro, Foster, Ortega, 2015). Los modelos de regresión espacial son una extensión de los modelos de regresión lineal, que incorpora la dependencia espacial. Estos modelos capturan o corrigen las relaciones de dependencia impidiendo las estimaciones sesgadas. Además, la Econometría Espacial con el tiempo a centrado sus intereses en la especificación y estimación de Modelos Espaciales basados en la información que contienen los datos de panel, debido a que los datos de panel contienen más información y más

variabilidad que los datos de corte transversal o series temporales, ofreciendo a los investigadores otras posibilidades de modelado. (Elhorst, 2014, citado en Carracedo et al., 2017).

La Econometría Espacial, es un campo de investigación que ha evolucionado mucho en las últimas décadas. Hasta hoy en día se tiene una gran variedad de modelos y técnicas diferentes. Anselin (1988), consideró la correlación espacial entre unidades espaciales en Modelos Espaciales. Baltagi (2005), consideró la heterogeneidad no observada de datos de panel en Modelos de Datos de Panel. Baltagi et al. (2003), considera la heterogeneidad no observada y la correlación espacial en modelos de datos de panel espaciales. (Citados en Baltagi, Song, Jung y Koh, 2007).

Los modelos espacio-temporales basados en datos de panel son probablemente uno de los temas más prometedores en la literatura sobre econometría espacial. Recientemente ha aparecido una serie de artículos teóricos que desarrollan procedimientos de estimación para diferentes modelos. Una de las investigaciones más recientes dentro de la literatura de paneles espaciales, considera la posible correlación serial en el término del error.

La finalidad del estudio es analizar si la tasa de contagios de Covid-19 efectivamente tiene asociación espacial y temporal, es decir, si los datos dependen del tiempo y el espacio y de ser así se pretende explicar esta asociación a través de los modelos espacio-temporales, considerando la movilidad como variable explicativa.

# Capítulo 2

## Objetivos

### 2.1. Objetivo General

Realizar un análisis espacio temporal del número de contagios y el índice de movilidad durante la pandemia del Covid-19 en Chile.

### 2.2. Objetivos Específicos

- Realizar una revisión literaria sobre los modelos clásicos de análisis espacio-temporal.
- Realizar una revisión literaria de la Econometría Espacial.
- Indagar sobre los test de asociación espacial y temporal que puedan servir para el análisis.
- Confeccionar la base de datos con las variables relevantes medidas en espacio y tiempo.
- Observar gráficamente la relación entre las variables de interés con el tiempo y el espacio.
- Verificar si existe dependencia espacial y/o dependencia temporal en los datos.
- Estimar varios modelos espacio-temporales sobre la tasa de contagios de Covid-19.
- Realizar una selección de modelos de acuerdo a criterios de bondad de ajuste y validación de supuestos.

# Capítulo 3

## Metodología

Para un correcto estudio del proyecto se debe realizar una investigación sobre la Estadística Espacial. En particular, familiarizarse con la Econometría Espacial, para luego ahondar en los modelos de análisis espacio-temporal, con la finalidad de entenderlos para posteriormente aplicarlos.

Se analizará el indicador epidemiológico más relevante de esta pandemia, que es el número de contagios, y se analizarán las variables de movilidad para comprender su comportamiento y cómo están siendo calculadas. Para esto, se realizará un análisis descriptivo de los datos, y además con esto podrán notarse posibles irregularidades.

Para el análisis se considerarán las 52 comunas de la Región Metropolitana como nivel espacial, y se pretende trabajar con los datos en semanas epidemiológicas como nivel temporal. Teniendo las variables medidas en ambas dimensiones, se verificará la relación espacio-tiempo, corroborando si se cumplen las hipótesis 1 y 2 gráficamente.

Además, para poder implementar algún modelo espacio-temporal, se debe verificar la existencia de dependencia o autocorrelación espacial y temporal de las variables. Para esto se utilizará un test de asociación espacial considerando los datos para cada semana y se utilizará un test de asociación temporal considerando los datos para cada comuna. Estos test son definidos posteriormente.

Una vez, que se tengan los conocimientos necesarios sobre el modelo a utilizar y conociendo la información que se tiene sobre los datos, se procede a implementar la metodología de los modelos considerados.

Como se estableció anteriormente, el objetivo de esta tesis es realizar un análisis espacio-temporal del indicador asociado al número de contagios, por lo que, teniendo conocimientos de las características y supuestos de los modelos espacio-temporales, se realizará la selección del mejor modelo de acuerdo a los criterios de bondad de ajuste y la validación de supuestos, considerando como óptimo el que tenga un mejor ajuste y logre cumplir con los supuestos correspondientes. Este análisis se implementará en el Software R-Studio.

Para todos los análisis que lo requieran, el nivel de significancia considerado será de  $\alpha = 0,05$ .

# Capítulo 4

## Marco Teórico

En este capítulo, se comienza definiendo los conceptos básicos relacionados con el análisis espacio-temporal, como lo es la Estadística Espacial, Series de Tiempo y los Datos Espacio-Temporales. Luego se define la Correlación Espacial y Temporal que puede estar presente en los datos espacio-temporales y sus respectivos Test que comprueban la existencia de la dependencia espacial y/o serial.

Definido esto, se procede a explicar la literatura de los Modelos Espacio-Temporales provenientes de la Econometría Espacial y su generalización con datos de panel llamados Modelos Espaciales con Datos de Panel o Modelos de Panel con Estructura Espacial.

Finalmente se definen indicadores para la evaluación y la elección del modelo más óptimo.

### 4.1. Estadística Espacial

La Estadística Espacial es una rama de la estadística que analiza y trabaja con datos georreferenciados o datos espaciales, es decir, datos que tienen asociada una posición en el espacio. Es ampliamente utilizada y se establecen tres grandes áreas basadas en el tipo de dato espacial (Giraldo, 2016):

- **Datos Geoestadísticos o Geoestadística:** El investigador analiza un conjunto fijo de unidades espaciales de interés, sobre la superficie terrestre. Este conjunto es continuo, es decir, se tiene un valor para cada punto en la tierra. Su objetivo principal es estimar valores desconocidos a partir de valores ya conocidos. Se conoce el espacio, pero se desconoce el valor de la variable, por lo que, esta es aleatoria.

- **Patrones espaciales o Procesos Puntuales:** El investigador no escoge las unidades espaciales, estas son aleatorias y pueden ser continuas o discretas. Su objetivo principal es analizar los puntos o áreas en los que puede ocurrir el evento de interés. Se suele asimilar a un proceso de Poisson (cuenta la cantidad de eventos hasta un cierto lapso  $T$ ), pues cuenta la cantidad de eventos en una determinada región.
- **Datos de Lattice, Área, Celda o Grilla:** El investigador analiza el fenómeno de interés en un conjunto fijo y discreto de unidades espaciales. Usualmente los datos de área están asociados a la epidemiología.

Los datos espaciales, como se mencionó anteriormente, están asociados a una localización en el espacio y su representación geométrica puede estar dada por:

- **Puntos:** Expresados en coordenadas terrestres, latitud y longitud o en coordenadas cartesianas, eje  $x$  y eje  $y$ .
- **Líneas:** Objetos abiertos que comunican varios puntos. Pueden ser considerados como arcos debido a la forma esférica de la tierra. Por ejemplo, carreteras, vías de trenes, etc.
- **Polígonos:** Figuras planas conectadas por distintos objetos cerrados que cubren un área determinada. Por ejemplo, países, comunas, regiones, etc.

El área de estudio y la representación de los datos espaciales, dependerá del interés estadístico y del tipo de dato espacial con el que se cuente para poder realizar el análisis correspondiente.

#### 4.1.1. Datos de área (Lattice Data)

Los datos de área corresponden a un número finito de áreas provenientes de la partición espacial de una región determinada. Estas áreas, son áreas geográficas delimitadas por un polígono y pueden ser regulares o irregulares. Se espera que la variable de estudio tenga un comportamiento específico en las distintas áreas, debido a las diferentes características del lugar geográfico.

Se puede observar en la figura 4.1, la partición de la región  $R$  en 3 áreas. Esta posee tres unidades de área,  $R_1$ ,  $R_2$  y  $R_3$ . Estas áreas podrían corresponder a países, ciudades, regiones o comunas, entre otras.

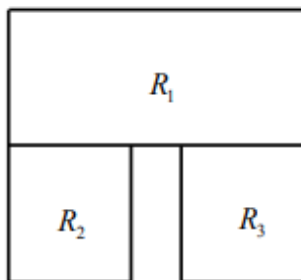


Figura 4.1: Unidades espaciales de área de la región  $R$

Como se mencionó anteriormente, las particiones de la región pueden ser irregulares o regulares espacialmente, como se puede apreciar en la figura 4.2. Las regulares son subregiones divididas idénticamente en el espacio total. Normalmente estas divisiones son rectangulares, siguiendo los cuadrantes formados por los paralelos y meridianos del planeta. Las irregulares son áreas distintas en tamaños y formas y suelen ser divisiones ya estipuladas.



Figura 4.2: Mapa de Galicia dividido por áreas regulares (derecha) y por áreas irregulares (izquierda).

Cada una de estas divisiones o áreas, se comporta como una única observación y los valores de los datos asociados a estas divisiones, representan medidas de resumen para la unidad del área total. Esto permite tener comparaciones más directas entre los valores de cada área.

En este análisis los datos espaciales corresponden a datos de área, pues los valores observados de las variables están asociados a ubicaciones fijas, específicamente los datos observados son tomados de cada comuna de la Región Metropolitana. Es por esto que se le dió mayor énfasis a esta definición.

Se consideran las comunas como dimensión espacial, porque son divisiones ya establecidas en la región. Cada una posee características propias de la comuna, que pueden servir para el análisis.

Dividir la región en áreas regulares, dificulta poder realizar conclusiones en relación a las unidades espaciales, ya que, se tendría poca o casi nada de información que caracterice a cada área. Además la densidad de la población está muy concentrada en la parte central, es decir, al realizar la división regular, habrían áreas que se llevarían mucho más peso y no se tendría como cuantificar.

## 4.2. Series de tiempo

Las series de tiempo son un conjunto de datos registrados y recolectados en intervalos de tiempo regulares (diarios, semanales, semestrales, anuales, etc.). Se denotan como  $\{Y_t : t \in T\}$ , donde el tiempo es estrictamente creciente.

En una serie, es de interés observar el comportamiento de la variable de estudio en el tiempo y al analizar la serie, los objetivos pueden ser: realizar predicciones, retropredicciones o imputaciones. Predicciones para ver el futuro comportamiento de la serie, retropredicciones para saber sobre el comportamiento pasado de la serie o imputaciones para tener información de un dato no observado en la serie.

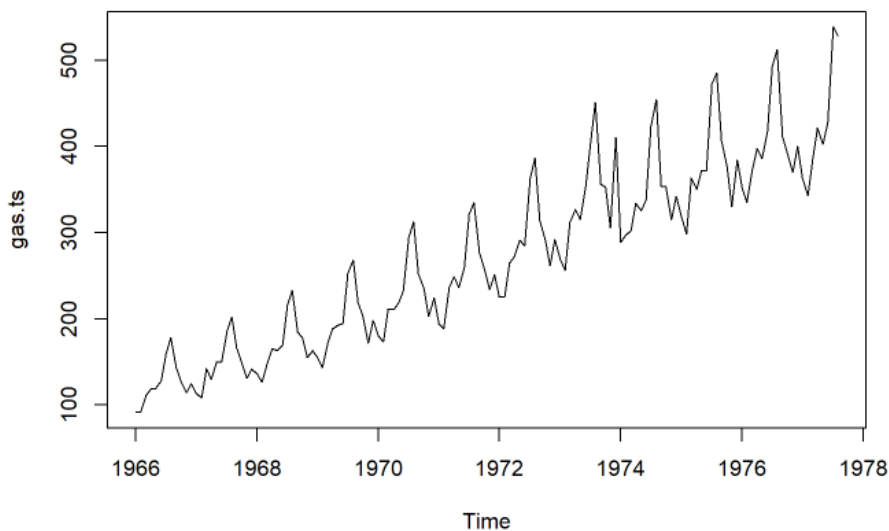


Figura 4.3: Datos mensuales del consumo de gasolina en España entre enero de 1966 y agosto de 1977

A modo de ejemplo, en la figura 4.3 se puede observar una serie de tiempo con datos mensuales recolectados entre enero de 1966 y agosto de 1977.

### 4.2.1. Datos temporales

Los datos temporales, como su nombre lo indica son datos que identifican el tiempo. Generalmente complementan los datos de otra variable, pues pueden indicar cuándo y cada cuánto fueron recolectados y/o registrados (días de la semana, fechas, horas, años, número de semana del año, etc.).

Para el análisis, este tipo de datos dará a conocer el número de la semana epidemiológica en el que fue recolectado cada dato. Por lo que, para cada comuna los datos de las variables son series de tiempo, pues los intervalos de tiempo son regulares.

## 4.3. Datos espacio-temporales

Este tipo de datos, es la combinación de datos espaciales y temporales. Como su nombre lo indica, los datos espacio-temporales son datos recolectados en distintas posiciones de un espacio determinado durante varios instantes de tiempo.

Algunos datos medidos en estas dos dimensiones, son los datos de panel espaciales.

### 4.3.1. Datos de Panel

Los datos de panel se relacionan con una sección transversal de observaciones (individuos, grupos, países, regiones) repetido durante varios períodos de tiempo. (Millo y Piras, 2012). Los datos transversales, son datos recogidos en un mismo instante de tiempo, pero no representan necesariamente localizaciones en el espacio. Es decir, espacialmente no tienen relación.

Las secciones transversales tienen relación unas con otras cuando los valores de la variable de estudio se ven influenciadas debido a sus posiciones en el espacio, ya sea que las secciones transversales estén más cercas o más lejos que otras. Estos datos son llamados datos de panel espacial, pues generan que haya relación entre cada observación espacial.

Los datos de panel, se ordenan primero según espacio y luego según el tiempo, siendo esta última la dimensión que va más “rápido”. La estructura es similar a la registrada en la tabla 4.1

Obs N°	Individuo	Año	Sexo	Edad
1	1	1890	F	20
2	1	1895	F	25
3	1	1900	F	30
4	2	1890	M	33
5	2	1895	M	38
6	2	1900	M	43
⋮	⋮	⋮	⋮	⋮

Tabla 4.1: Ejemplo estructura de Datos de Panel

Dependiendo del número de datos considerado tanto en la dimensión espacial como temporal, los datos de panel se clasifican en:

- **Paneles micro:** Paneles con más observaciones espaciales que temporales.
- **Paneles macro:** Paneles con más observaciones temporales que espaciales.

Y dependiendo de la existencia o ausencia de datos faltantes, los datos de panel pueden ser:

- **Paneles balanceados o completos:** Todas las observaciones se observan a lo largo del período de estudio, es decir, el rango temporal, es el mismo para todas las unidades espaciales.
- **Paneles no balanceados o incompletos:** El rango temporal varía entre los espacios, es decir, existen datos faltantes en alguno de los paneles.

## 4.4. Autocorrelación o Dependencia Espacial

Cuando se trabaja con datos georreferenciados, es frecuente encontrar que los valores de la variable de estudio dependan del lugar en el que se está observando el valor, posiblemente debido a características geográficas. Pues, los valores de la variable en cierto lugar suelen ser parecidos con los valores de los alrededores cercanos, considerando que se conservan o son similares las características geográficas de los lugares próximos. Como se encuentran dentro del mismo espacio o entorno, corresponde que los valores sean similares o parecidos para decir que hay presencia de autocorrelación espacial.

La primera ley de la geografía de Tobler (1970), dice que “Todos los lugares están relacionados, pero los lugares cercanos están más relacionados que los lugares lejanos.”, por otra parte Flint, Harrower y Edsall (2000) afirman que “Existe dependencia espacial cuando el valor de la variable dependiente en una unidad espacial de análisis, es parcialmente una función del valor de la misma variable en unidades vecinas.” (Citados en Vilalta, 2005). Esto más que nada avala la dependencia espacial que se encuentra en datos espaciales, con fundamentos de grandes conocedores del área.

La autocorrelación espacial puede ser positiva o negativa. Si los valores de la variable de estudio en una unidad espacial son similares a los valores de las unidades espaciales que la rodean, favoreciendo la concentración de esos valores en ese espacio, se dice que hay presencia de autocorrelación positiva. Esto se puede observar en la imagen 4.4. Por el contrario, la autocorrelación será negativa si el valor de la variable de estudio en una unidad espacial genera que en todos sus alrededores los valores de esas unidades espaciales cercanas sean muy diferentes. Esto indica una tendencia a la dispersión de las unidades espaciales. Dicho de otra forma, áreas que tengan un alto valor de la variable, estarían rodeadas de áreas con valores bajos o viceversa. Esto se puede notar en la imagen 4.5.

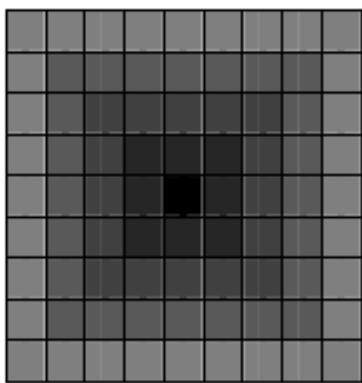


Figura 4.4: Autocorrelación Positiva

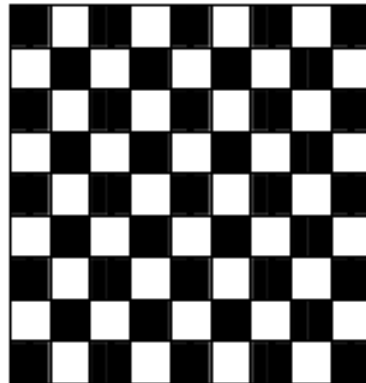


Figura 4.5: Autocorrelación Negativa

Dicho esto, se puede decir, que si la variable de estudio se distribuyera de forma aleatoria en todo el espacio que se estudia, la autocorrelación espacial no existiría.

#### 4.4.1. Criterio de Vecindad

Considerando que los datos utilizados en este análisis son del tipo de área, para notar la dependencia espacial entre las observaciones, se debe tener claro el concepto de vecindad, es decir, bajo qué criterio se considerará que las observaciones son vecinas o cercanas. Se considera relevante

definir las características que deben tener dos áreas para que puedan ser consideradas vecinas.

Cada área está representada por cada observación, pero para definir los criterios de vecindad se debe tener un punto específico que represente a dicha área. Generalmente en áreas regulares, se suele usar el centroide de cada unidad espacial y en áreas irregulares, se suele considerar algún punto importante del área. Por ejemplo, si se trabaja con países podría ser la capital, provincia o municipio.

Existen diversos criterios para definir la vecindad, estos pueden ser considerando la posición geográfica, utilizando fuentes de información socioeconómica o basándose en la similaridad o disimilaridad que existe entre las áreas. Lo que se debe tener en cuenta, es que todos estos criterios de vecindad deben poder diferir y crear dos conjuntos disjuntos que logren diferenciar las que son áreas vecinas y las que no lo son, de cada una de las áreas de estudio.

Dicho esto, los criterios de vecindad más considerados son los criterios geográficos. Estos pueden ser especificados con base en distancias (usualmente euclidiana) o en el comportamiento de bordes comunes que tenga cada área. Se tienen entonces 2 criterios:

- Vecinos por contigüidad: Este criterio considera como áreas vecinas a aquellas que están aledañas en el mapa. No hay que pasar por una tercera para ir de una a otra.
- Vecinos basados en distancias: Este criterio considera como áreas vecinas a las que cumplen con cierto criterio referente a alguna función de distancia que separa a las áreas.

Estos criterios se representan a través de la construcción de la matriz de contigüidad.

#### 4.4.2. Matriz de Contigüidad o Pesos Espaciales

La matriz de contigüidad es una medida de cercanía que cuantifica la relación de vecindad. Según Moreno y Vallá (2004), “La matriz de pesos espaciales es un instrumento básico para reflejar las interdependencias en el espacio”. (Citado en Polo, 2017). Representa la intensidad que tiene cada una de las áreas con las demás áreas del espacio que se estudia. Por tanto, la matriz está compuesta de tantas filas y columnas como tantas áreas se consideren. Cada una de las filas y columnas representa un área del espacio.

La matriz se define con la letra  $W$  y para construirla existen varias formas. Considerando  $N$  áreas, la matriz se representa de la siguiente manera:

$$W = \begin{bmatrix} 0 & w_{1,2} & \dots & w_{1,j} & \dots & w_{1,N} \\ w_{2,1} & 0 & \dots & w_{2,j} & \dots & w_{2,N} \\ \vdots & \vdots & \ddots & \vdots & \dots & \vdots \\ w_{i,1} & w_{i,2} & \dots & 0 & \dots & w_{i,N} \\ \vdots & \vdots & \dots & \vdots & \ddots & \vdots \\ w_{N,1} & w_{N,2} & \dots & w_{N,j} & \dots & 0 \end{bmatrix}$$

Donde  $w_{i,j}$  corresponden a los pesos espaciales entre el área  $i$  (fila) y el área  $j$  (columna), donde  $i, j = 1, \dots, N$ . Estos pesos indican la relación e intensidad de dependencia entre cada par de áreas. La diagonal principal, siempre está compuesta de ceros, pues indica que ninguna área puede estar relacionada consigo misma, es decir, no puede ser vecina de si misma. Esta matriz suele ser simétrica, estocástica, es decir, no definida al azar, cuadrada y se define positiva.

Los valores que toman  $w_{ij}$  dependerán del criterio de vecindad considerado.

- **Vecinos por contigüidad:** Este criterio es el más simple y el más usado, sobretodo cuando se cuenta con poca información acerca de las características de los datos. Los pesos espaciales toman valores binarios y la suma de cada fila de la matriz, corresponde al número de vecinos que tiene cada área. Ese valor representa el peso del área, la intensidad de vecindad.

$$w_{ij} = \begin{cases} 1 & \text{si las áreas } i \text{ y } j \text{ son contiguas, es decir, vecinas} \\ 0 & \text{si las áreas } i \text{ y } j \text{ no son contiguas, es decir, no son vecinas} \end{cases}$$

El criterio de vecinos por contigüidad, considera como vecinos a las polígonos que comparten límites. Esto se puede observar y considerar de tres formas:

- **Contigüidad de Torre:** Considera a las áreas al norte, sur, este y oeste como vecinas.
- **Contigüidad de Alfil:** Considera a las áreas sureste, suroeste, noreste y noroeste como vecinas.
- **Contigüidad de Reina:** Considera a todas las áreas que comparten límite como vecinas.



Figura 4.6: Tipos de Contigüidad

Observando la imagen 4.6, las áreas más oscuras alrededor del área central, son las que reciben el valor 1 en la matriz de pesos, y el resto toma el valor 0.

- **Vecinos basados en distancias:** Este criterio considera los centroides de cada área, para poder calcular las distancias entre polígonos. Algunas de estas son:
  - **Distancia Inversa ( $1/d$ ):** toma la distancia entre vecinos cercanos como 1 y el resto de las distancia entre las áreas las calcula entre cada par, considerando como pesos la expresión:
 
$$w_{ij} = \frac{1}{d_{ij}}$$
 Así, a mayor distancia entre áreas, menor es la relación.
  - **Distancia Inversa ( $1/d^2$ ):** igual que la anterior, pero considera la distancia al cuadrado. Esta suele ser la más usada.
- **Otros:** La literatura también ha combinado los criterios de distancia o contigüidad con otra información geográfica como el perímetro del polígono, la extensión de la frontera en común, y otras características geográficas. (Herrera, 2017).

Es común, luego de haber definido la matriz de pesos, trabajar con alguna transformación de esta, debido a que mejora las propiedades estadísticas de los estimadores. La transformación más utilizada suele ser la normalización por filas. (Herrera, 2017). Esto es, se divide cada elemento  $w_{ij}$  por la suma de cada fila a la que pertenece. De esta forma, la suma de los pesos asociados a cada área es igual a 1. Otra opción sería normalizar por columna.

El problema de construir la matriz de contactos ha estado siempre presente en la literatura de la Econometría Espacial, ya que, la matriz  $W$  debe captar las relaciones multidireccionales. ¿Cómo se debe construir esta matriz? “Hay muy pocas restricciones que deban considerarse cuando se construye esta matriz (ceros en la matriz diagonal, no puede ser idempotente y debe estar asintóticamente acotada, Kapoor et al., 2007). Bajo estas premisas, la decisión final corresponde al usuario, lo que introduce un alto grado de arbitrariedad en la solución.” (Citado por Paelinck, Mur y Trávez, 2015). Se pueden considerar dos enfoques:

- (i) Especificar la matriz de forma exógena.
- (ii) Especificar la matriz de forma endógena.

“ El exógeno es el enfoque más popular entre los usuarios y está basado en juicios previos, subjetivos en gran medida, relativos a la estructura geográfica de los datos. Ejemplos de este enfoque son el criterio de vecindad, los  $k$ -vecinos, los umbrales de distancia o los diferentes kernels en los que se utiliza la distancia como argumento (Harris et al., 2011). El segundo enfoque toma en consideración tanto la topología del espacio como la naturaleza de los datos para construir la matriz a partir de ellos. Bodson y Peeters (1975) pueden considerarse los pioneros en este enfoque al que después se han sumado muchos otros como Getis y Aldstadt (2004), Aldstadt y Getis (2006) con el procedimiento AMOEBA, los filtros espaciales de Tiefelsdorf y Griffith (2007), el enfoque de máxima entropía de Fernández et al. (2009) o el criterio CCC de Mur y Paelinck (2011).” (Citados por Paelinck et al., 2015).

## 4.5. Test de Asociación Espacial

Los test explicados a continuación ayudarán a identificar si los datos poseen asociación espacial, es decir, si existe dependencia o autocorrelación espacial entre ellos. Antes de ahondar en los test de asociación espacial, hay que considerar que la autocorrelación espacial se estudia desde dos perspectivas, global y local. Se definen a continuación.

- **Autocorrelación espacial Global:** Se presenta cuando todas las unidades espaciales de la muestra están correlacionadas, es decir, todas las unidades que componen la muestra no están distribuidas aleatoriamente, lo hacen conforme a un patrón determinado. Dicho de otra forma, existe una estructura espacial general en la distribución de la variable, considerando el espacio geográfico completo.

Los estadísticos de autocorrelación espacial global, analizan las observaciones de la muestra

total en forma conjunta, por lo que, no resultan sensibles ante situaciones donde exista una importante inestabilidad en la distribución espacial de la variable. Esto quiere decir, que no contemplan la posibilidad de que la dependencia no se mantenga en todas las unidades analizadas, pues al calcular el valor del indicador, puede que no todas las áreas aporten el mismo peso. Esta limitación está contemplada en las pruebas de autocorrelación espacial Local.

- **Autocorrelación espacial Local:** Se presenta cuando los valores de cada unidad espacial están correlacionadas con los valores de las unidades espaciales de su alrededor, geográficamente hablando.

Los estadísticos de autocorrelación espacial local, contemplan la posibilidad de que la dependencia no se mantenga en todas las unidades, pues es capaz de considerar situaciones de dependencia que se presenten en un área determinada, ya que, calcula un valor para cada unidad espacial.

Existen varios índices globales y locales para medir la autocorrelación espacial. Para el análisis se considerará el índice basado en covarianzas, es decir, el Índice de Moran Global con su respectivo índice Local asociado. Los valores del índice de Moran varían entre -1 y 1, estos valores son más intuitivos para una correlación. Esto sucede sólo cuando la matriz de pesos está normalizada por filas.

#### 4.5.1. Test de Asociación Espacial Global

El Índice de Moran Global, es un indicador de asociación espacial que indica la intensidad de la dependencia espacial de la variable, de todas las áreas consideradas. Este test permite contrastar la hipótesis nula de no autocorrelación espacial Global, es decir, contrastar la hipótesis nula que afirma la existencia de una distribución aleatoria de la variable a lo largo del territorio. Por lo tanto, se define:

$H_0$  : No existe autocorrelación espacial Global  $v/s$   $H_1$  : Existe autocorrelación espacial Global

o

$H_0 : I = 0$   $v/s$   $H_1 : I \neq 0$

El índice de Moran Global, está basado en las covarianzas entre áreas y se calcula de la siguiente manera:

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{i=1}^n (y_i - \bar{y})^2} \stackrel{a}{\sim} N(0, 1) \quad n \rightarrow \infty$$

donde,  $n$  es el número de áreas de la muestra,  $y_i$  el valor de la variable de estudio en el área  $i$ ,  $y_j$  el valor de la variable de estudio en el área  $j$ ,  $w_{ij}$  el valor de la matriz de pesos de la fila  $i$ , columna  $j$  y  $\bar{y}$  es el promedio de la variable  $y$ , considerando  $i \neq j$ .

Su valor varía entre -1 y 1:

$$I \begin{cases} > 0 & \text{indica autocorrelación espacial positiva} \\ = 0 & \text{no existe autocorrelación espacial} \\ < 0 & \text{indica autocorrelación espacial negativa} \end{cases}$$

Para el análisis espacio-temporal que se realizará, se espera rechazar la hipótesis nula para cada variable.

### 4.5.2. Test de Asociación Espacial Local

El índice de Moran local es un indicador local de la asociación espacial (LISA), que fue introducido por Anselin (1995). (Citado en Carracedo, 2017). Este índice ayuda a detectar si la dependencia espacial global se mantiene en todas las áreas estudiadas o en grupos de áreas detectando clusters donde la autocorrelación es significativa. Se definen las hipótesis como:

$$H_0 : \text{No existe autocorrelación espacial Local} \quad v/s \quad H_1 : \text{Existe autocorrelación espacial Local}$$

o

$$H_0 : L_i = 0 \quad v/s \quad H_1 : L_i \neq 0$$

El índice de Moran Local corresponde al mismo Índice de Moran ocupado para calcular la asociación Global, pero se calcula para cada área individualmente. La suma de los valores del índice de Moran Local de cada área corresponde al valor del índice de Moran Global.

$$L_i = \frac{(y_i - \bar{y})}{\sum_{k=1}^n \frac{(y_k - \bar{y})^2}{(n-1)}} \sum_{j \in J_i} w_{ij} (y_j - \bar{y})$$

donde,  $y_i$  es el valor de la variable para la área  $i$ ,  $y_j$  es el valor de la variable de la área vecina a  $i$ ,  $y_k$  es el valor de la variable para la área  $k$ ,  $J_i$  es el conjunto de regiones vecinas del área  $i$ ,  $\bar{y}$  es el promedio de la variable  $y$  y  $n$  es el número de áreas de la muestra.

$$L_i \begin{cases} > 0 & \text{indica autocorrelación espacial positiva} \\ = 0 & \text{no existe autocorrelación espacial} \\ < 0 & \text{indica autocorrelación espacial negativa} \end{cases}$$

Se destaca que para este test existe un problema en su aplicación, ya que, se debe realizar  $N$  veces, una vez para cada unidad espacial. Es por esto que surge un problema por la existencia de múltiples test, ya que, el valor de significancia global no será equivalente al nivel de significancia local. Para resolver esto, se deben ajustar los  $p$ -valores. Esto se puede hacer de diversas formas, ya sea, considerando el número de vecinos de cada área, considerando el número de comparaciones realizadas, entre otras. Es decir, se debe hacer alguna corrección. Algunas de ellas son: la corrección de Bonferroni, la corrección de Holm, la tasa de falsos descubrimientos, entre otras.

## 4.6. Autocorrelación o Dependencia Temporal o Serial

La dependencia temporal se presenta cuando las observaciones en un cierto instante  $t$  están estrechamente relacionadas con las observaciones en otro instante  $t$ . Dicho de una forma más correcta, un cierto valor en el tiempo  $t$ , se ve influenciado por valores en tiempos pasados,  $t-1$ ,  $t-2$ ,  $t-3$ , ..., etc.

Dentro del campo de series de tiempo, los modelos planteados tienen como uno de los supuestos, que no debe haber autocorrelación o dependencia temporal o serial en las perturbaciones (errores) del modelo, pues la autocorrelación temporal sesga las estimaciones de la varianza del error. Es decir, podría existir una subvaloración o una sobrevaloración del valor del estimador. Dicho de otra forma, los parámetros del modelo no estarían bien estimados, dando lugar a determinar inferencias erróneas. Además, si los errores no poseen autocorrelación temporal, significa que el modelo planteado logró capturar el patrón de comportamiento en las componentes del modelo.

Las principales causas de la presencia de autocorrelación temporal, son cuando existe tendencia y/o estacionalidad (algún patrón que se mantenga a lo largo de la serie) en los datos.

## 4.7. Test de Asociación Temporal

El test explicado en esta sección ayudará a identificar si los datos poseen asociación o dependencia temporal.

### 4.7.1. Test de Box-Ljung

El test de Box-Ljung es un contraste para determinar la existencia de autocorrelación temporal entre los datos. Las hipótesis se plantean como:

$H_0$  : Los datos no están autocorrelacionados  $v/s$   $H_1$  : Los datos están autocorrelacionados

El estadístico de este test es:

$$Q = T(T + 2) \sum_{h=1}^L \frac{1}{T - h} \hat{\delta}_h^2 \sim \chi_L^2$$

donde  $T$  es el número de observaciones temporales consideradas en la muestra de una unidad espacial y  $\hat{\delta}_h^2$  es la autocorrelación estimada para el  $i$ -ésimo lag  $L$ .

Este test se debe realizar para cada unidad espacial con sus respectivos  $T$  datos.

## 4.8. Modelos Espacio-Temporales

Los Modelos Espacio-Temporales pretenden modelar datos medidos en espacio y tiempo. Estos modelos tienen como finalidad capturar y considerar la posible dependencia, tanto espacial como temporal que pueda presentarse en los datos.

Como se definió anteriormente, en la sección 4.3, “Datos Espacio-Temporales”, los datos de panel pueden ser considerados datos espacio-temporales teniendo en cuenta que las secciones transversales son independientes unas con otras, es decir, consideran que no existe dependencia entre las unidades espaciales. Los modelos para estos datos, se denominan Modelos de Datos de Panel y contemplan los problemas de heterogeneidad no observada tanto en espacio como en tiempo, pero dado que consideran independencia, estos modelos no controlan ni consideran la posible dependencia espacial ni temporal.

Los modelos que consideran la dependencia espacial y serial, son los Modelos con Datos de Panel Espaciales, que son una generalización de los Modelos Espaciales junto con la estructura de

los paneles espaciales, los cuales consideran que hay dependencia entre cada unidad espacial. Estos modelos tienen en cuenta las relaciones de vecindad entre las unidades espaciales, explicando así el comportamiento de unidades que estén relacionadas entre sí.

Otros modelos espacio-temporales, provienen de una generalización de series de tiempo. En estos modelos se debe considerar gran cantidad de datos temporales para que sea posible la correcta estimación de parámetros. En este análisis no se cuenta con gran cantidad de datos temporales, por lo que, el análisis de múltiples series temporales no se abarcará como metodología.

Por tanto, el enfoque de este análisis se realizará según los Modelos con Datos de Panel y Estructura Espacial o también llamado Modelos Espaciales de Panel. Con estos modelos es posible modelar la dependencia espacial y temporal presente en los datos. Es importante recalcar la consideración de la correlación espacial y serial que puede presentarse en los datos, pues de no ser considerada tal dependencia, las estimaciones de parámetros y pruebas de significancia que se obtienen no son estables.

Actualmente, los Modelos Espaciales de Panel que consideran dependencia espacial y temporal son bastante recientes y su implementación en softwares no está tan desarrollada como los métodos de análisis de datos espaciales y de series temporales por separado, pero frente a la clara dependencia temporal y/o espacial que puede existir en los datos, no es posible obviar alguno de estos aspectos. El poco desarrollo de estos modelos en softwares limita a la hora de utilizarlos.

Aunque existen paquetes para estimar modelos espaciales transversales en R, Stata y Matlab, son escasos los paquetes que estiman modelos espaciales de datos de panel (Millo y Piras 2012). Hasta ahora, están disponibles el código que complementa a Kapoor, Kelejian y Prucha (2007) en Stata, el código de Elhorst (2011) en MATLAB y por último el paquete `spml` (Millo y Piras 2012) en R. (Citados en Carracedo et al., 2017).

Para el análisis, la implementación de estos modelos se realizará en el Software R, más específicamente se utilizará el paquete `spml`.

## 4.9. Modelos Espaciales

Los Modelos Espaciales o Modelos de Corte Transversal provienen de la Econometría Espacial como solución a problemas de dependencia espacial en datos de corte transversal, es decir, datos de unidades espaciales observados en un mismo y único instante de tiempo.

La Econometría Espacial es de gran utilidad cuando se considera el uso de variables vinculadas al espacio, ya que, este tipo de datos suelen presentar (por su naturaleza) relaciones multidireccionales, traducidas como dependencia en el espacio o autocorrelación espacial (similar a los problemas de dependencia temporal o autocorrelación serial, presente en series de datos temporales), que pueden invalidar el uso de otros modelos.

La presencia de autocorrelación en los datos, impide que los modelos usuales y convencionales, sean una buena herramienta para su modelación, pues su existencia genera varias consecuencias en el análisis estadístico. Aunque en muchos casos, los modelos que incluyen dependencia, son más realistas que los que no lo hacen.

Generalmente, en un análisis estadístico convencional o habitual, se suelen considerar algunos supuestos sobre los datos antes de realizar la modelación. Se supone que un conjunto de observaciones son independientes e idénticamente distribuidas, pero la dependencia espacial viola este supuesto, pues los valores de la variable no son independientes, por lo tanto, la variable y los residuos tendrán correlación. Además, cuando se considera una muestra de los datos espaciales, para realizar el análisis de estos, al no ser aleatoria, contiene menos información que una muestra no correlacionada. Esta pérdida de información se debe considerar en la estimación, pues si no, las estimaciones no serían representativas ni correctas. Los parámetros estimados con base en muestras que no se distribuyen aleatoriamente en el espacio son sesgados, por lo que, los supuestos que se consideran sobre los datos antes del análisis, no son válidos. La autocorrelación también, genera redundancia entre los datos, por lo que, cada observación adicional aportará menos información nueva.

¿Cómo se puede contemplar la dependencia espacial en el modelo? Existen diferentes tipos de dependencia espacial, las principales se producen debido a los errores de medida y/o debido a la propia interacción espacial entre las observaciones en unidades espaciales contiguas. Pérez (2006), señala que “la dependencia espacial en modelos de regresión aparece como consecuencia de autocorrelación en variables dependientes o independientes y por la aparición de un esquema de dependencia espacial en las perturbaciones.” (Citado en Castro, Foster y Ortega, 2015). Por lo tanto, la dependencia espacial puede estar presente en variables dependientes, independientes o en las perturbaciones del modelo. Para poder incorporar estas dependencias espaciales al modelo, se utiliza la matriz de pesos espaciales,  $W$ , definida en la subsección de Matriz de Contigüidad (4.4.2). Esta matriz permite incorporar los tipos de dependencia espacial causadas por las distintas variantes.

El modelo general más completo que planteó Cliff-Ord, que incorpora los 3 tipos de dependencia espacial es:

$$\begin{aligned} y &= \lambda Wy + x\beta + \theta Wx + u \\ u &= \rho Wu + \varepsilon \end{aligned} \tag{4.1}$$

donde,  $y$  es un vector de dimensión  $N \times 1$  de la variable dependiente,  $N$  es el número de unidades espaciales consideradas,  $x$  es una matriz  $N \times k$  de las  $k$  variables independientes,  $\beta$  es el vector de parámetros de orden  $N \times k$ ,  $W$  es la matriz de pesos espaciales de dimensiones  $N \times N$ ,  $\lambda$  es el coeficiente de rezago espacial,  $\theta$  es el coeficiente espacial de dependencia local,  $\rho$  es el coeficiente de error espacial y  $\varepsilon$  es el componente de error de dimensión  $N \times 1$ .

Este modelo espacial general (GSM), contiene los 3 tipos de efectos de interacción espaciales:

- $Wy$ : efecto de interacción espacial o retardos espaciales de la variable dependiente
- $Wx$ : efectos de interacción espacial o retardos espaciales de las variables independientes
- $Wu$ : efecto de interacción espacial o retardos espaciales en el término del error

Tipo de Modelo	Efectos de interacción espacial	Definición
OLS	ningún efecto	Modelo de Mínimos Cuadrados Ordinarios
SAR o SLM	$Wy$	Modelo de Retardo Espacial o Modelo Espacial Autorregresivo (incorpora la dependencia sustantiva)
SEM	$Wu$	Modelo de Error Espacial (incorpora la dependencia en el error)
SLX	$Wx$	Modelo de Retardo Espacial en X (incorpora la dependencia en los regresores)
SAREM, SAC o SARAR	$Wy, Wu$	Modelo Espacial Autorregresivo con Autorregresión en los Errores
SDM	$Wy, Wx$	Modelo Espacial de Durbin
SDEM	$Wx, Wu$	Modelo de Error Espacial de Durbin
GSM	$Wy, Wx, Wu$	Modelo Espacial General

Tabla 4.2: Tipos de Modelos Espaciales

Dependiendo de los efectos espaciales que se incluyan en el modelo, se definen los distintos tipos de modelos espaciales en la tabla 4.2. Y en la imagen 4.7 se muestra el orden metódico, dependiendo de la existencia de cada parámetro  $\rho$ ,  $\theta$ ,  $\lambda$ .

Para la elección del modelo óptimo de los datos, se puede comenzar analizando el modelo más general o el modelo simple. El enfoque habitual consiste en iniciar con un modelo simple, no espacial y lineal, y contrastar si hay suficiente evidencia en su contra para incorporar los efectos de interacción espacial necesarios. Este enfoque es denominado específico a general (Stge). El enfoque opuesto es el general a específico, denominado Gets, en el que se inicia con el modelo más general. Para este caso, se intenta simplificar el modelo general descartando hipótesis débiles y redundantes. (Paelinck et al., 2015)

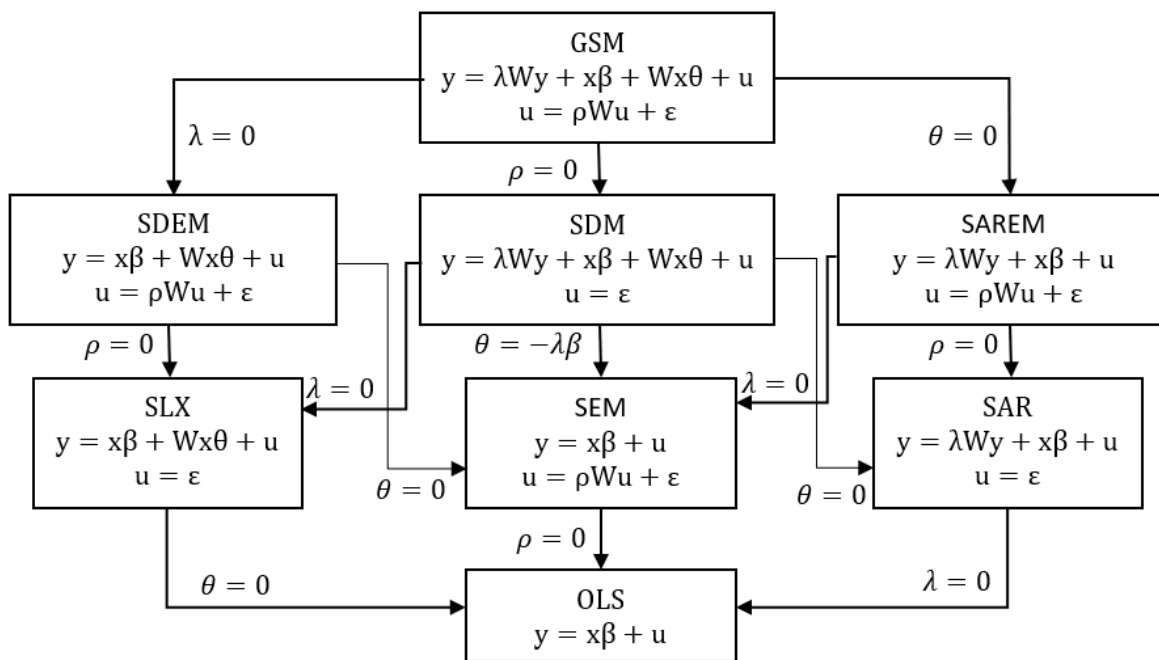


Figura 4.7: Modelos Espaciales

Para extender estos modelos espaciales de  $N$  observaciones a modelos espacio-temporales, donde para cada observación espacial se tienen  $T$  observaciones temporales, se agrega el sub-índice  $t$ , donde  $t = 1, \dots, T$ , y además se deben incorporar efectos no observables, considerados en los Modelos con Datos de Panel, ya que, de no hacerlo se pueden obtener estimaciones sesgadas e inconsistentes. Estos modelos se denominan Modelos Espaciales con Datos de Panel.

## 4.10. Modelos Espaciales con Datos de Panel

Los Modelos de Panel con Estructura Espacial, son una generalización de los modelos espaciales con estructura de datos de panel. Como se mencionó en la subsección de Datos de Panel (4.3.1), los paneles espaciales están asociados a observaciones que tienen relación con una posición en el espacio. Al suceder esto, se genera que haya una relación entre los datos de posiciones cercanas. Para representar esta relación, se utiliza la matriz de pesos espaciales,  $W$ .

La Econometría Espacial, es un campo de investigación que ha evolucionado mucho en las últimas décadas. Hasta hoy en día se tiene una gran variedad de modelos y técnicas diferentes. Una de las investigaciones más recientes dentro de la literatura de paneles espaciales, considerada la posible correlación serial en el término del error en el contexto de modelos con efectos aleatorios.

Los Modelos de Panel con Estructura Espacial consideran los efectos fijos o aleatorios para capturar la heterogeneidad no observada de los individuos y además consideran la correlación espacial. Según Millo (2016), el tema de la correlación serial en los errores es particularmente sensible en el caso de los paneles de efectos fijos. De hecho, la técnica estándar para eliminar los efectos fijos individuales, induce a una correlación serial artificial en los residuos transformados que puede combinarse con la correlación serial original, si ya está presente. La extensión de los Modelos de Panel Espaciales en dirección a la correlación serial en los errores, en la literatura está contemplada en los Modelos de Panel Espaciales con Efectos Aleatorios. (Baltagi et al., 2007). Por lo tanto, se considera el Modelo de Regresión de Datos de Panel con correlación serial en cada unidad espacial a lo largo del tiempo, así como con dependencia espacial entre las unidades espaciales en cada punto en el tiempo.

Cabe destacar que muchas de las aplicaciones empíricas de los métodos espacio-temporales se ven obstaculizadas por la falta de softwares disponibles. Como se mencionó en un principio, aunque hay bibliotecas en R, Matlab y Stata para modelos espaciales transversales, hay pocos procedimientos de softwares disponibles para estimar modelos de datos de panel espacial. (Millo et al., 2012). Los más destacados últimamente son R y MATLAB, que cada vez ofrecen más funciones para estimar Modelos de Panel Espaciales. Para el análisis de la implementación de estos modelos, se considera el Software R. El paquete **spml**, contiene varios comandos que se pueden utilizar para la implementación de los Modelos de Panel Espaciales. El comando más reciente es el **sprem1**, con el cual se puede trabajar para estimar Modelos de Pane Espacailles con Efectos Aleatorios, Correlación Espacial y Serial en los Errores. (Millo et al., 2012).

### 4.10.1. Modelos de Panel Espaciales con Efectos Aleatorios, Correlación Espacial y Correlación Serial

Estos modelos son la generalización más reciente de los Modelos de Panel Espaciales que puede implementarse en el software R. Millo et al (2012), describen la implementación de estos modelos en el documento “splm: Spatial Panel data models in R”.

Baltagi et al. en 2007, introducen los Modelos de Panel Espaciales con Efectos Aleatorios, Correlación Espacial y Serial a la Econometría Espacial. Parten considerando un modelo de regresión de datos de panel equilibrado, con  $N$  unidades espaciales observadas durante  $T$  períodos de tiempo:

$$y_{it} = x'_{it}\beta + u_{it} \quad i = 1, \dots, N \quad t = 1, \dots, T \quad (4.2)$$

donde  $y_{it}$  es la observación del  $i$ -ésimo espacio del  $t$ -ésimo tiempo de la variable dependiente,  $x_{it}$  es el vector de  $k \times 1$  de los regresores no estocásticos (de las variables explicativas fijas) y  $u_{it}$  corresponde a la perturbación del modelo. En forma vectorial, para cada unidad espacial  $i = 1, \dots, N$ ,  $u_t$  es la suma de un término de efectos aleatorios individuales  $\mu$  y un término de residuos espacialmente autocorrelacionados  $\varepsilon$ .

$$u_t = \mu + \varepsilon_t \quad (4.3)$$

$$\varepsilon_t = \rho W \varepsilon_t + v_t \quad (4.4)$$

El término de perturbación restante sigue un proceso autorregresivo de primer orden:

$$v_t = \psi v_{t-1} + e_t \quad (4.5)$$

donde,  $u'_t = (u_{1t}, \dots, u_{Nt})$ ,  $\varepsilon'_t = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})$  y  $v'_t = (v_{1t}, \dots, v_{Nt})$ .  $\mu' = (\mu_1, \dots, \mu_N)$  corresponde al vector de los efectos aleatorios individuales con  $\mu \stackrel{iid}{\sim} N(0, \sigma_\mu^2)$ .  $\rho$  es el coeficiente espacial autorregresivo con  $|\rho| < 1$  y  $\psi$  el coeficiente de correlación serial con  $|\psi| < 1$ .  $W$  es la matriz de  $N \times N$  de pesos espaciales, cuyos elementos diagonales son cero. Finalmente,  $v_{i0} \sim N(0, \frac{\sigma_e^2}{1-\psi^2})$ ,  $e_{it} \stackrel{iid}{\sim} N(0, \sigma_e^2)$ . Se asume que  $\varepsilon$  y  $\mu$  son independientes.

Este modelo es el más general en el término del error, puesto que considera la presencia de efectos aleatorios espaciales, la dependencia espacial entre unidades espaciales para cada tiempo y permite la correlación serial en cada unidad espacial a lo largo del tiempo.

La ausencia de alguno(s) o todos los parámetros,  $\sigma_\mu^2$ ,  $\rho$  y  $\psi$ , en el término del error del modelo, genera distintos modelos. Estos están presentes en la siguiente tabla:

	$\rho \neq 0$ $\psi \neq 0$	$\rho \neq 0$ $\psi = 0$	$\rho = 0$ $\psi \neq 0$	$\rho = 0$ $\psi = 0$
$\sigma_\mu^2 \neq 0$	SEMSRRE	SEMRE	SRRE	RE
$\sigma_\mu^2 = 0$	SEMSR	SEM	SR	OLS

Tabla 4.3: Tipos de Modelos Espaciales de Panel con Efectos Aleatorios, Correlación Espacial y Correlación Serial

donde:

- **SEMSRRE:** Es el modelo completo. Modelo Espacial de Panel con Correlación Espacial, Efectos Aleatorios y Correlación Serial.
- **SEMRE:** Modelo Espacial de Panel con Correlación Espacial y Efectos Aleatorios.
- **SRRE:** Modelo Espacial de Panel con Efectos Aleatorios y Correlación Serial.
- **RE:** Modelo Datos de Panel con Efectos Aleatorios.
- **SEMSR:** Modelo Espacial de Panel con Correlación Espacial y Correlación Serial.
- **SEM:** Modelo Espacial de Panel con Correlación Espacial.
- **SR:** Modelo Espacial de Panel con Correlación Serial.
- **OLS:** Modelo de Mínimos Cuadrados Ordinarios.

Los efectos aleatorios y el componente de error, pueden escribirse de manera matricial como:

$$\begin{aligned}
 y &= X\beta + u \\
 u &= (l_T \otimes I_N)\mu + \varepsilon \\
 \varepsilon &= \rho(I_T \otimes W)\varepsilon + v \\
 v_t &= \psi v_{t-1} + e_t
 \end{aligned}
 \tag{4.6}$$

donde  $y$  es de dimensión  $NT \times 1$ ,  $X$  es de  $NT \times k$ ,  $\beta$  es de  $k \times 1$ ,  $u$  es de  $NT \times 1$ ,  $\otimes$  es el producto de Kronecker que permite considerar matricialmente ambas dimensiones, tiempo y espacio,  $l_T$  es un vector de unos de dimensión  $T$ ,  $I_N$  es una matriz identidad de dimensión  $N$ ,  $\mu$  es un vector de  $N \times 1$ ,  $\varepsilon$  es de  $NT \times 1$ ,  $I_T$  es una matriz identidad de dimensión  $T$ ,  $W$  es de  $N \times N$  y  $v$  es de dimensión  $NT \times 1$

Los parámetros del modelo, se pueden estimar con el método de Máxima Verosimilitud (ML). Para los modelos con componentes de error espacialmente autocorrelacionados o para un modelo de efectos aleatorios, el método de Minimos Cuadrados Ordinales (MCO) es ineficiente.

La siguiente probabilidad es de forma general para cualquier modelo de estructura de covarianza de error  $\Omega_u$ , es decir, para los modelos que fueron definidos en la tabla 4.3. Se asume la normalidad de los errores,  $u \sim N(0, \Omega_u)$ , para realizar el método de ML. Se define el logaritmo de la verosimilitud como:

$$\log(L) = -\frac{NT}{2} \ln(2\pi\sigma_e^2) - \frac{1}{2} \ln|\Sigma| - \frac{1}{2\sigma_e^2} u' \Sigma^{-1} u \quad (4.7)$$

considerando que  $u = Y - X\beta$  y  $\Omega_u = \sigma_e^2 \Sigma$ , donde la matriz de varianza-covarianza  $\Sigma$ , está escalada por la varianza del error,  $\sigma_e^2$ , a modo de conveniencia para el procedimiento de estimación de parámetros.

A modo de ejemplo para mostrar el procedimiento, se considera la estimación mediante el método de ML, para el modelo completo, SEMSRRE. La estimación de los modelos más simples, de menos parámetros, puede ser obtenida para los casos especiales realizando el mismo procedimiento. Para mayor detalle, se puede revisar Millo (2014) ó Baltagi et al. (2007).

Para calcular  $\log(L)$ , se debe calcular la matriz de varianza-covarianza escalada del componente del error  $\Sigma$ , su inversa  $\Sigma^{-1}$  y su determinante  $|\Sigma|$ .

El término del error compuesto del modelo,  $u$ , agrupa la parte de efectos aleatorios, la correlación espacial y serial, por lo tanto, en forma matricial puede escribirse como:

$$u = (I_T \otimes I_N)\mu + (I_T \otimes B^{-1})v \quad (4.8)$$

donde  $B = I_N - \rho W$ ,  $B$  es no singular, es decir, es invertible, y las demás definiciones son las mismas que en la ecuación 4.6.

La matriz de varianza-covarianza  $\Omega$ , del término del error  $\mu$ , se puede escribir como:

$$\Omega_u = \sigma_\mu^2 (J_T \otimes I_N) + (V \otimes (B'B)^{-1}) \quad (4.9)$$

donde  $J_T$  es una matriz de unos de dimensión  $T \times T$  y  $V$  es la matriz de varianza-covarianza de

dimensión  $T \times T$  del proceso autorregresivo de primer orden,  $v_t \sim AR(1)$ . Con:

$$\begin{aligned}
 V &= \sigma_e^2 \left( \frac{1}{1 - \psi^2} \right) V_1 \\
 V_1 &= \begin{bmatrix} 1 & \psi & \psi^2 & \dots & \psi^{T-1} \\ \psi & 1 & \psi & \dots & \psi^{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \psi^{T-1} & \psi^{T-2} & \psi^{T-3} & \dots & 1 \end{bmatrix} \\
 V &= \sigma_e^2 V_\psi ; \quad V_\psi = \frac{1}{1 - \psi^2} V_1
 \end{aligned} \tag{4.10}$$

Por lo tanto,  $\Omega_u$  puede reescribirse como:

$$\Omega_u = \sigma_\mu^2 (J_T \otimes I_N) + (\sigma_e^2 V_\psi \otimes (B'B)^{-1}) \tag{4.11}$$

Y se sabe que la matriz de varianza-covarianza del error se denota como  $\Omega$  si no está escalada y se denota como  $\Sigma$  si está escalada por la varianza del error,  $\sigma_e^2$ , por tanto,  $\Omega = \sigma_e^2 \Sigma$ . La expresión de  $\Sigma_u$  es:

$$\begin{aligned}
 \Omega_u &= \sigma_\mu^2 (J_T \otimes I_N) + (\sigma_e^2 V_\psi \otimes (B'B)^{-1}) \\
 \Sigma_u &= \frac{\sigma_\mu^2}{\sigma_e^2} (J_T \otimes I_N) + (V_\psi \otimes (B'B)^{-1})
 \end{aligned} \tag{4.12}$$

Para derivar la expresión de la probabilidad, Baltagi et al. (2007c) utilizan una transformación de Prais-Winsten del modelo con efectos aleatorios y autocorrelación espacial. (Citado en Millo et al, 2012). Pero Millo et al. (2012) sólo consideran la notación simplificadora, no la transformación de Prais-Winsten. Se definen las notaciones:

$$\begin{aligned}
 \phi &= \frac{\sigma_\mu^2}{\sigma_e^2} \\
 \alpha &= \sqrt{\frac{1 + \psi}{1 - \psi}} \\
 d^2 &= \alpha^2 + (T - 1)
 \end{aligned} \tag{4.13}$$

Por lo tanto, la matriz de covarianza de error  $\Sigma$ , su inversa  $\Sigma^{-1}$  y el determinante  $|\Sigma|$  se pueden escribir respectivamente como:

$$\begin{aligned}
 \Sigma_u &= \phi (J_T \otimes I_N) + (V_\psi \otimes (B'B)^{-1}) \\
 \Sigma_u^{-1} &= V_\psi^{-1} \otimes (B'B) + \frac{1}{d^2(1 - \psi)^2} (V_\psi^{-1} J_T V_\psi^{-1}) \otimes ([d^2(1 - \psi)^2 \phi I_N + (B'B)^{-1}]^{-1} - B'B) \\
 |\Sigma_u| &= \frac{|d^2(1 - \psi)^2 \phi I_N + (B'B)^{-1}| \cdot |(B'B)^{-1}|^{T-1}}{(1 - \psi^2)^N}
 \end{aligned} \tag{4.14}$$

Finalmente, siguiendo la forma general de la probabilidad descrita en la ecuación 4.7, se tiene lo siguiente:

$$\begin{aligned} \log(L) &= -\frac{NT}{2}\ln(2\pi\sigma_e^2) - \frac{1}{2}\ln|\Sigma| - \frac{1}{2\sigma_e^2}u'\Sigma^{-1}u \\ \log(L) &= -\frac{NT}{2}\ln(2\pi\sigma_e^2) + \frac{N}{2}\ln(1-\psi^2) - \frac{1}{2}\ln|d^2(1-\psi)^2\phi I_N + (B'B)^{-1}| \\ &\quad - \frac{(T-1)}{2}\ln|B| - \frac{1}{2\sigma_e^2}u'\Sigma^{-1}u \end{aligned} \quad (4.15)$$

Millo et al. (2012) afirman que “Implementamos un procedimiento iterativo para obtener las estimaciones de máxima verosimilitud (para detalles sobre la implementación ver Millo y Piras, 2009). Partiendo de los valores iniciales para  $\rho$ ,  $\psi$  y  $\phi$ , obtenemos estimaciones para  $\beta$  y  $\sigma_e^2$  a partir de las condiciones de primer orden:”

$$\begin{aligned} \beta &= (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y \\ \sigma_e^2 &= \frac{u'\Sigma^{-1}u}{NT} = \frac{(y - X\beta)'\Sigma^{-1}(y - X\beta)}{NT} \end{aligned} \quad (4.16)$$

Dicho en otras palabras, el procedimiento iterativo parte asignando valores iniciales a  $\rho$ ,  $\psi$  y  $\phi$ , luego calcula la expresión de  $\Sigma^{-1}$ , y calcula los valores de  $\beta$  y  $\sigma_e^2$  determinados con el método de Mínimos Cuadrados Generalizados (GLS). Estos pasos se realizan hasta llegar a la convergencia y se logre maximizar  $\log(L)$ .

Los modelos con término del error compuesto por efectos aleatorios, correlación espacial y serial, pueden generalizarse aún más, considerando diversas estructuras espaciales vistos en la sección de Modelos Espaciales 4.9. Lo que restringe la implementación de estos modelos es la falta de implementaciones en softwares, ya que, sólo está la teoría en la literatura.

#### 4.10.2. Modelos de Panel Espaciales con Retraso Espacial, Efectos Aleatorios, Correlación Espacial y Correlación Serial

Millo et al. (2012), aparte de la generalización de los modelos SEMRE agregando la correlación serial al modelo (SR), también implementaron la generalización de los modelos con efectos aleatorios y retraso espacial (SARRE) agregando la correlación serial al término del error y también implementaron la generalización de los modelos con efectos aleatorios, retraso espacial y correlación espacial (SAREMRE) agregando la correlación serial al término del error. Es decir, a los modelos de la tabla 4.3 se les agrega el componente de retraso espacial (SAR).

El modelo más completo es el SEMSRRE más el retraso espacial, este modelo se designa como SAREMSRRE y se define como:

$$\begin{aligned}
 y &= \lambda W y + X\beta + u \\
 u &= (I_T \otimes I_N)\mu + \varepsilon \\
 \varepsilon &= \rho(I_T \otimes W)\varepsilon + v \\
 v_t &= \psi v_{t-1} + e_t
 \end{aligned}
 \tag{4.17}$$

Los modelos que se desprenden del modelo más general, son todos los vistos en la tabla 4.3, agregándoles el parámetro  $\lambda \neq 0$ . La tabla 4.4 muestra como quedan las siglas de estos modelos.

	$\lambda \neq 0$	$\lambda \neq 0$	$\lambda \neq 0$	$\lambda \neq 0$
	$\rho \neq 0$	$\rho \neq 0$	$\rho = 0$	$\rho = 0$
	$\psi \neq 0$	$\psi = 0$	$\psi \neq 0$	$\psi = 0$
$\sigma_\mu^2 \neq 0$	SAREMSRRE	SAREMRE	SARSRRE	SARRE
$\sigma_\mu^2 = 0$	SAREMSR	SAREM	SARSR	SAR

Tabla 4.4: Tipos de Modelos Espaciales de Panel, Retardo Espacial, Efectos Aleatorios, Correlación Espacial y Correlación Serial

Estos modelos también se estiman por el método de Máxima Verosimilitud. El logaritmo de la verosimilitud de forma general, para cualquier modelo con estructura de retardo espacial SAR combinado con cualquier estructura de covarianza de error  $\Sigma$ , se determina de la siguiente manera:

$$\log(L) = -\frac{NT}{2} \ln(2\pi\sigma_e^2) - \frac{1}{2} \ln|\Sigma| + T \ln|A| - \frac{1}{2\sigma_e^2} u' \Sigma^{-1} u
 \tag{4.18}$$

con  $u = (I_T \otimes A)y - X\beta$  y  $A = I_N - \lambda W$  donde  $A$  es una matriz no singular, es decir, es invertible y  $\lambda$  es el parámetro de retraso espacial.

Por tanto, para estimar un modelo SAREMSRRE, se toman las matrices de varianza-covarianza de la estructura de error SEMSRRE, las determinadas en la ecuación 4.14, y se realiza la optimización.

### 4.10.3. Consideraciones de la Implementación de los modelos en R

Según Millo (2012), los modelos particulares de SEMSRRE y SAREMSRRE, vistos en las tablas 4.3 y 4.4 respectivamente, que consideran la correlación serial, es decir, los modelos con las siglas SR

incluidas, son nuevos en la literatura y su implementación también. Millo señala algunos problemas con respecto al rendimiento y la estabilidad al estimar estos modelos en el software R:

- Computacionalmente es bastante pesado realizar la optimización para la estimación de los parámetros.
- Se deben considerar muestras de tamaño moderado.
- Los valores iniciales son importantes para la estabilidad y confiabilidad, no así para la velocidad.
- El escalado de parámetros es crucial para reducir los problemas de las estimaciones de varianzas negativas para parámetros no significativos.

Además, los modelos completos consideran muchos parámetros y dependiendo de la cantidad de datos, estos pueden ser difíciles de estimar numéricamente. Una estrategia viable computacionalmente y parsimoniosa para el modelo de error (considerando efectos aleatorios, correlación espacial y serial) es probar las pruebas condicionales LM descritas por Baltagi et al.(2007), para poder estimar un modelo más simple.

#### 4.10.4. Pruebas del Multiplicador de Lagrange (Pruebas LM)

Baltagi et al. (2007), generalizó los estudios anteriores de pruebas conjuntas, al derivar estadísticas de prueba para el modelo de panel espacial con correlación serial.

Las pruebas del Multiplicador de Lagrange (LM) se han utilizado ampliamente para probar efectos aleatorios, correlación espacial o serial en modelos de datos de panel. Según Millo et al. (2012), al requerir sólo la estimación de la especificación del modelo restringido, las pruebas LM son particularmente atractivas en un entorno de efectos aleatorios espaciales debido a las dificultades computacionales relacionadas con el enfoque ML.

Las pruebas que consideran modelos que permiten la correlación serial y que se pueden implementar, son las pruebas conjuntas y condicionales derivadas de Baltagi et al. (2007c). (Citado en Millo et al., 2012)

Se considera una prueba conjunta de los parámetros y tres pruebas condicionales.

1.  $H_0 : \rho = \psi = \sigma_\mu^2 = 0$  v/s  $H_1 : \text{alguno de los componentes no es cero. (J)}$

2.  $H_0 : \rho = 0$  dado que  $\psi \neq 0$  y  $\sigma_\mu^2 > 0$  v/s  $H_1 : \rho \neq 0$  dado que  $\psi \neq 0$  y  $\sigma_\mu^2 > 0$  (C.1)

3.  $H_0 : \psi = 0$  dado que  $\rho \neq 0$  y  $\sigma_\mu^2 > 0$  v/s  $H_1 : \rho \neq 0$  dado que  $\psi \neq 0$  y  $\sigma_\mu^2 > 0$  (C.2)

4.  $H_0 : \sigma_\mu^2 = 0$  dado que  $\rho \neq 0$  y  $\psi \neq 0$  v/s  $H_1 : \sigma_\mu^2 \neq 0$  dado que  $\rho \neq 0$  y  $\psi \neq 0$  (C.3)

donde si el valor de LM es significativamente cercano a cero, se rechaza  $H_0$ .

El estadístico LM de cada prueba se determina como:

$$LM = \tilde{D}_\theta' \tilde{J}_\theta^{-1} \tilde{D}_\theta \quad (4.19)$$

donde  $\theta'$  es el vector de parámetros presentes en  $H_0$ ,  $\tilde{J}_\theta$  es parte de la matriz de información de  $\theta$  evaluada en su estimación ML,  $\tilde{\theta}$ , del modelo restringido según  $H_0$  y  $\tilde{D}_\theta$  es un vector de derivadas parciales de la función de verosimilitud con respecto a cada parámetro  $\theta$  evaluado en su estimación ML del modelo restringido. Las derivadas de cada parámetro se pueden calcular según  $\tilde{D}_{\theta_r}$ , considerando  $\theta' = (\sigma_e^2, \sigma_\mu^2, \psi, \rho)$ :

$$\tilde{D}_{\theta_r} = \frac{\partial L(\tilde{\theta})}{\partial \theta_r} = -\frac{1}{2} \text{tr} \left[ \Omega^{-1} \frac{\partial \Omega}{\partial \theta_r} \right] + \frac{1}{2} u' \left( \Omega^{-1} \frac{\partial \Omega}{\partial \theta_r} \Omega^{-1} \right) u \quad , \quad r = 1, 2, 3, 4 \quad (4.20)$$

El parámetro  $\Omega$  corresponde a la matriz de varianza-covarianza del error considerando el modelo restringido según  $H_0$  y el parámetro  $u$  es un vector de residuos obtenido de la estimación ML bajo el modelo restringido de  $H_0$ .

A continuación se presentan las expresiones de los estadísticos de cada prueba. Si se desea ver el detalle del procedimiento y de las derivadas para calcular el estadístico, se puede ver Baltagi et al. (2007).

El estadístico LM de la prueba (J) está dado por:

$$LM_J = \frac{NT^2}{2(T-1)(T-2)} [A^2 - 4AF + 2TF^2] + \frac{N^2T}{b} H^2 \sim \chi_3^2 \quad (4.21)$$

donde:

$$\begin{aligned} A &= \frac{\tilde{u}'(J_T \otimes I_N)\tilde{u}}{\tilde{u}'\tilde{u}} - 1 \quad ; \quad F = \frac{\tilde{u}'(G_T \otimes I_N)\tilde{u}}{2\tilde{u}'\tilde{u}} \\ H &= \frac{\tilde{u}'(I_T \otimes (W' + W))\tilde{u}}{2\tilde{u}'\tilde{u}} \quad ; \quad b = \frac{\text{tr}(W + W')^2}{2} \end{aligned} \quad (4.22)$$

Además,  $G$  es una matriz  $T \times T$  con elementos bidiagonales iguales a uno y  $\tilde{u}$  denota los residuos de un modelo OLS, es decir, del modelo restringido según  $H_0$ .

El estadístico LM de la prueba C.1, denotado como  $LM_{\rho|\psi\mu}$ , considera el modelo restringido, según  $H_0$ , que permite la existencia de correlación de error en serie y efectos aleatorios. Este está dado por:

$$LM_{\rho|\psi\mu} = \tilde{D}(\rho)^2 \tilde{J}_{44}^{-1} \sim \chi_1^2 \quad (4.23)$$

donde  $\tilde{J}_{44}$ , corresponde al elemento de la matriz de información para el parámetro  $\rho$  y  $\tilde{D}(\rho)$  corresponde a la derivada en función de  $\rho$ . Se definen estas expresiones a continuación

$$\begin{aligned} \tilde{J}_{44} &= b(T - 2cg + c^2g^2) \\ g &= tr(V^{-1}J_T) = \frac{1 - \psi}{\sigma_e^2}(2 + (T - 2)(1 - \psi)) \\ b &= \frac{tr(W + W')^2}{2} \\ \tilde{D}(\rho) &= \left. \frac{\delta L}{\delta \rho} \right|_{H_0} = \frac{1}{2} \hat{u}' [V^{-1} - 2cV^{-1}J_TV^{-1} + c^2(V^{-1}J_T)^2V^{-1}] \otimes (W' + W) \hat{u} \end{aligned} \quad (4.24)$$

donde  $\tilde{u}$  es un vector de residuos obtenido de la estimación ML de un modelo con componentes de error de efectos aleatorios y correlación serial.

El estadístico LM de la prueba C.2, denotado como  $LM_{\psi|\rho\mu}$ , considera el modelo restringido, según  $H_0$ , que permite la existencia de correlación espacial y efectos aleatorios en el término del error. Este está dado por:

$$LM_{\psi|\rho\mu} = \tilde{D}(\psi)^2 \tilde{J}_{33}^{-1} \sim \chi_1^2 \quad (4.25)$$

donde  $\tilde{J}_{33}$ , corresponde al elemento de la matriz de información  $\tilde{J}_\theta$  para el parámetro  $\psi$  evaluado bajo  $H_0$  y  $\tilde{D}(\psi)$  corresponde a la derivada en función de  $\psi$ . Se definen estas expresiones a continuación

$$\begin{aligned} \tilde{J}_{33} &= \frac{N}{T^2}(T^3 - 3T^2 + 2T + 2) + tr[Z(B'B)^{-1}]^2 \frac{2(T-1)^2 \hat{\sigma}_e^4}{T^2} \\ Z &= [T\sigma_\mu^2 I_N + \sigma_e^2(B'B)^{-1}]^{-1} \\ \tilde{D}(\psi) &= -\frac{T-1}{T}(\hat{\sigma}_e^2 tr(Z(B'B)^{-1}) - N) \\ &\quad + \frac{\hat{\sigma}_e^2}{2} \hat{u}' \left[ \frac{1}{\hat{\sigma}_e^4} (E_T G E_T) \otimes (B'B) + \frac{1}{\hat{\sigma}_e^2} (\hat{J}_T G E_T) \otimes Z \right. \\ &\quad \left. + \frac{1}{\hat{\sigma}_e^2} (E_T G \hat{J}_T) \otimes Z + (\hat{J}_T G \hat{J}_T) \otimes Z(B'B)^{-1} Z \right] \hat{u} \end{aligned} \quad (4.26)$$

donde  $\tilde{u}$  es un vector de residuos obtenido de la estimación ML de un modelo con componentes de error de efectos aleatorios y correlación serial.

El estadístico LM de la prueba C.3, denotado como  $LM_{\mu|\psi\rho}$ , considera el modelo restringido, según  $H_0$ , que permite la existencia de correlación espacial y serial en el término del error. Este está dado por:

$$LM_{\mu|\psi\rho} = \tilde{D}(\sigma_\mu^2)^2 \tilde{J}_{22}^{-1} \sim \chi_1^2 \quad (4.27)$$

donde  $\tilde{J}_{22}$ , corresponde al elemento de la matriz de información  $\tilde{J}_\theta$  para el parámetro  $\sigma_\mu^2$  evaluado bajo  $H_0$  y  $\tilde{D}(\sigma_\mu^2)$  corresponde a la derivada en función de  $\sigma_\mu^2$ . Se definen estas expresiones a continuación

$$\begin{aligned} \tilde{J}_{22} &= \frac{g^2 \text{tr}(B'B)^2}{2} \\ \tilde{D}(\sigma_\mu^2) &= -\frac{g}{2} \text{tr}(B'B) + \frac{1}{2\sigma_e^4} \hat{u}' [V_\psi^{-1} J_T V_\psi^{-1} \otimes (B'B)^2] \hat{u} \end{aligned} \quad (4.28)$$

donde  $g$  está definida en 4.24 y  $\tilde{u}$  es un vector de residuos obtenido de la estimación ML de un modelo con componentes de correlación espacial y serial en el término del error.

Estas 4 pruebas permitir ver la posibilidad de estimar un modelo más simple y están implementadas para probar efectos aleatorios, correlación espacial y serial en el Software R en el paquete **spml** a través del comando **bsjkttest**.

Las pruebas LM no se han derivado para modelos que contengan un retraso espacial (SAR).

## 4.11. Indicadores para la evaluación del Modelo

Los indicadores de bondad de ajuste de un modelo estadístico, ayudan a establecer qué modelo es el más conveniente. Los que se consideran para este análisis son:

- Raíz del Error Cuadrático Medio (RMSE)

La raíz del error cuadrático medio, cuantifica la diferencia que existe entre los valores estimados y los observados del modelo. Este indicador da un valor general del comportamiento de los

errores y penaliza los errores de mayor magnitud. Por ende, mientras menor sea el valor del RMSE, mejor será el ajuste del modelo. Se define como:

$$RMSE = \sqrt{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (\hat{y}_{it} - y_{it})^2} \quad (4.29)$$

donde  $T$  es el número de observaciones por área,  $N$  es el número total de áreas y  $y$  es la variable de estudio.

- Coeficiente de Determinación ( $R^2$ )

El R-cuadrado indica la bondad o la aptitud del modelo, a menudo se utiliza con fines descriptivos y muestra qué tan bien las variables independientes explican la variabilidad de la variable dependiente.  $R^2$  se define como:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (4.30)$$

donde, SSR es la suma de los errores de cuadrados, SSE es la suma explicada de cuadrados y SST es la suma de cuadrados totales.

$$\begin{aligned} SST &= \sum_{i=1}^{NT} (y_i - \bar{y})^2 \\ SSR &= \sum_{i=1}^{NT} (y_i - \hat{y}_i)^2 \\ SSE &= \sum_{i=1}^{NT} (\hat{y}_i - \bar{y})^2 \\ SST &= SSR + SSE \end{aligned} \quad (4.31)$$

Para un buen modelo, se espera que  $SSR \ll SSE$ .

# Capítulo 5

## Datos

Como se ha mencionado en esta tesis, es bastante ventajoso trabajar con el tipo de dato espacio-temporal debido a que aporta mayor información a los análisis, aunque es bastante difícil encontrar variables que estén medidas en el mismo espacio y tiempo. Si bien la pandemia ha sido reciente, los datos del número de contagios (variable considerada como el indicador relevante para este análisis) se pudieron recolectar diariamente y por comuna pero, la recolección de otras variables no ha sido considerada de forma tan detallada en relación al mismo tiempo. Por lo tanto, para tener poca pérdida de información y por conveniencia para el análisis, las variables que se logran recolectar se toman a nivel semanal y a nivel comunal.

### 5.1. Naturaleza de los datos

Se trabaja con datos relacionados con el Covid-19, en la Región Metropolitana de Chile. Se considera factible tomar los datos de la región más poblada de Chile a nivel semanal, para poder analizar la evolución de las variables. Los datos están definidos en el mismo espacio y tiempo y para poder identificar estas dimensiones se considera una variable para las comunas de la Región Metropolitana y otra para las semanas epidemiológicas (consideran el día domingo como inicio de semana). Ambas variables identifican de forma única a cada observación de la data.

En este estudio, se analiza el indicador más relevante sobre el Covid-19, el número de contagios, pero es trabajado como tasa de contagios, es decir, se divide el número de contagios entre la cantidad de habitantes de cada comuna y se multiplica por 100.000 habitantes, pues esta variable permite una óptima comparación entre comunas. Estos datos son recolectados de los datos publicados por el Ministerio de Ciencias en el repositorio de Datos-COVID19. (2021, MinCiencia)

La variable que se considera para poder explicar el indicador, es una variable de movilidad. Cabe destacar que se pretendía considerar variables como nivel socioeconómico, o rango etario (edades), ya que, en Chile la edad está altamente correlacionada con el número de contagios, hospitalizaciones o muertes del COVID19, pero no se encuentran los datos medidos en el espacio y tiempo considerado y para el análisis espacio-temporal las variables deben variar en ambas dimensiones, pues sino, se generan problemas en el análisis de los modelos.

Se consideran 2 variables de movilidad recolectadas, pues se cree que quizás una explique mejor que la otra a la variable de interés, ya que, están calculadas de distinta forma. Una de ellas, corresponde al índice de movilidad proveniente del análisis realizado por el Instituto de Data Science (IDS) de la Facultad de Ingeniería de la Universidad del Desarrollo (UDD), con el apoyo de Telefónica Chile y CISCO, que consideraron el movimiento de los teléfonos móviles conectados a la red de telefónica en el territorio nacional. Este se calculó considerando el número de viajes que se realizaron dentro de una comuna específica, normalizado por el número de habitantes de la comuna. Además el Instituto consideró por separados los viajes realizados fuera de la comuna y dentro de la comuna, generando así las variables IMinterno e IMexterno y la suma de estos dos índices genera el índice de movilidad total (IM). Por otro lado, según el equipo multidisciplinario a cargo, para poder comparar el comportamiento de las distintas comunas se determinó la variación del IM de cada semana en comparación a la semana base del 9 al 15 de marzo. Para mayor información, consultar Bravo y Ferres, (2020).

La segunda variable de movilidad corresponde a una variación de movilidad proveniente del ISCI, Instituto de Sistemas Complejos de Ingeniería, que utilizó datos estadísticos e información del uso de infraestructura de telecomunicaciones, entregados de manera anonimizada y agregada por Entel, agrupadas a nivel de zona censal. Según Olivares, Goic, Weintraub, Covarrubia y Escobedo, (2020), “para cada día laboral (lunes a viernes), determinamos el flujo desde cada zona hogar a otras zonas, durante horarios de trabajo. Estos flujos pueden ser dentro de la misma comuna o a otras comunas (pero fuera de la zona hogar) y los interpretamos como asociados a actividades laborales. Luego, tomamos promedios semanales a nivel de comuna (excluyendo fines de semana) de estos flujos..<sup>El</sup>los determinaron dos variables referentes a la variación de movilidad en cada comuna, tomando como referencia las dos primeras semanas de marzo 2020. Una es variacion\_week que considera los días de lunes a viernes (excluyendo festivos) y la otra es variacion\_weekend que considera fines de semana (sábado y domingo). Si se desea saber más sobre la variable, consultar Olivares et al. (2020).

Para poder utilizar las variables de movilidad en este estudio, se realizan algunas modificaciones. La Universidad del Desarrollo determinó la variable índice de movilidad total diariamente, por lo que, se determina un promedio semanal para poder trabajar con ella. Después de eso, se determina la variación en base al promedio de la movilidad que hubo durante la semana 10 y 11, ya que, durante estas semanas había una movilidad normal en las comunas. De las variables de variación de movilidad provenientes del ISCI, sólo se considera `variacion.week`, ya que, como esta se midió en los días de lunes a viernes, estos datos se pueden asociar a las semanas epidemiológicas que toma los días de domingo a sábado, pues los días están contenidos. Esto no se puede considerar igual para la variable `variacion.weekend`, que considera los días sábado y domingo, ya que, no se sabría a que semana epidemiológica asociar esos datos.

Por otro lado, ambas variables de movilidad presentan datos faltantes en las dimensiones de espacio y tiempo. La variable movilidad proveniente de la Universidad del Desarrollo, considera datos desde la semana 10-2020 hasta 17-2021 pero, posee ausencia de datos para las 52 comunas de la Región Metropolitana en la semana 50-2020 y la variable movilidad proveniente del ISCI considera datos desde la semana 10-2020 hasta la 11-2021 pero, posee ausencia de datos para la comuna San Pedro y además para las 51 comunas restantes, posee ausencia de datos durante la semana 37-2020. Para poder comparar ambas variables de movilidad sin omitir información temporal, los valores de las semanas faltantes se imputan considerando la media entre los datos del tiempo  $t - 1$  y  $t + 1$ , para cada comuna. Por el lado espacial, la variable que no posee ningún dato para la comuna San Pedro, se puede considerar igual, ya que, la comuna esta ubicada en las afueras del área de estudio contemplada, y no presentará inconvenientes en la matriz de vecindad, por tanto esta comuna puede ser omitida.

Los valores que toman ambas variables de variación de movilidad varían alrededor del valor 0. Donde el 0 significa que no hay variación de movilidad. Si el valor es positivo indica que la variación de movilidad aumentó y si es negativo significa que la variación de la movilidad disminuyó.

# Capítulo 6

## Análisis Práctico

A modo resumen, las variables a considerar son:

- **Semana Epidemiológica:** variable que identifica la dimensión temporal. Se distingue como “yyyyxx”, siendo “yyyy” el año (2020 o 2021) y “xx” siendo la semana epidemiológica del año correspondiente. O también a lo largo del análisis, puede encontrarse sólo como la semana (usualmente en el eje  $x$  de los gráficos), considerando que la primera semana del año 2021 corresponde a la semana 54 y no a la 01 del año 2021 (ya que, la última semana del año 2020 es la 53).
- **Comuna:** variable que identifica a cada una de las 52 comunas de la Región Metropolitana.
- **Tasa de Contagios:** variable que indica la relación entre el número de contagios y la cantidad de población por comuna y semana epidemiológica, por cada 100.000 habitantes.
- **Variación de Movilidad Completa:** variable de variación de movilidad determinada con los datos del índice de movilidad proveniente de la Universidad del Desarrollo, asociados al flujo total de domingo a sábado. Periodo considerado de tiempo semana 10-2020 a
- **Variación de Movilidad Laboral:** variable de variación de movilidad proveniente del ISCI, asociados al flujo en horario laboral de lunes a viernes sin considerar festivos.

### 6.1. Análisis Descriptivo de los Datos

En esta sección se determinan algunas estadísticas descriptivas de las variables y algunos gráficos para la visualización de los datos con el fin de conocer la información con la que se cuenta.

Este análisis también será útil para comprobar las hipótesis planteadas. Estas son que el aumento de los casos de contagios inició en el sector oriente de la Región Metropolitana y se fue moviendo a la zona sur-poniente y además permitirá comprobar que la movilidad, durante las primeras semanas de la pandemia, era mayor en el sector sur-poniente en comparación al sector oriente de la Región Metropolitana.

### 6.1.1. Gráficos de las variables

Para observar el comportamiento que tienen las variables y analizar sus datos, se realizan gráficos de líneas de las variables para algunas comunas de la Región Metropolitana durante las distintas semanas epidemiológicas.

Para una mejor distinción del comportamiento de cada variable, se toman en consideración 9 comunas. Estas son: Cerrillos, La Florida, Las Condes, Ñuñoa, Pudahuel, Puente Alto, Quinta Normal, Santiago y Vitacura. Comunas con ruralidad baja o muy baja según la SEREMI, (2019).

#### 6.1.1.1. Gráfico de Líneas Tasa de Contagios

En el gráfico 6.1 se puede observar el comportamiento de la variable de 9 comunas durante las semanas epidemiológicas 10 a la 70 (202010 y 202117).

La variable tiene un comportamiento similar para cada comuna, donde durante las semanas 20 a 24 aproximadamente, se concentran los picks de la tasa de contagios, siendo los más altos los de las comunas Quinta Normal Y Pudahuel. Por la semana 24 comienzan a bajar los casos y se mantienen más controlados desde la semana 30 a la 59 y ya casi por la semana 60 nuevamente vuelven a aumentar. Dicho esto, se pueden distinguir 3 períodos. Uno donde se ve que los casos suben y disminuyen, otro donde los casos están más controlados y otro donde nuevamente vuelven a subir y bajar.

Por otra parte, al inicio de la pandemia, se puede notar que las líneas de las comunas Vitacura y las Condes están por sobre las demás, por lo que, sólo considerando las comunas seleccionadas, se puede decir que en esas comunas comenzó el aumento de casos. Durante la semana 13 y 14, estas comunas bajan sus contagios y las demás comienzan a aumentar. Esto no comprueba totalmente la hipótesis 1, pero no la contradice.

Para ver el gráfico de todas las comunas, se puede ver el Anexo A.

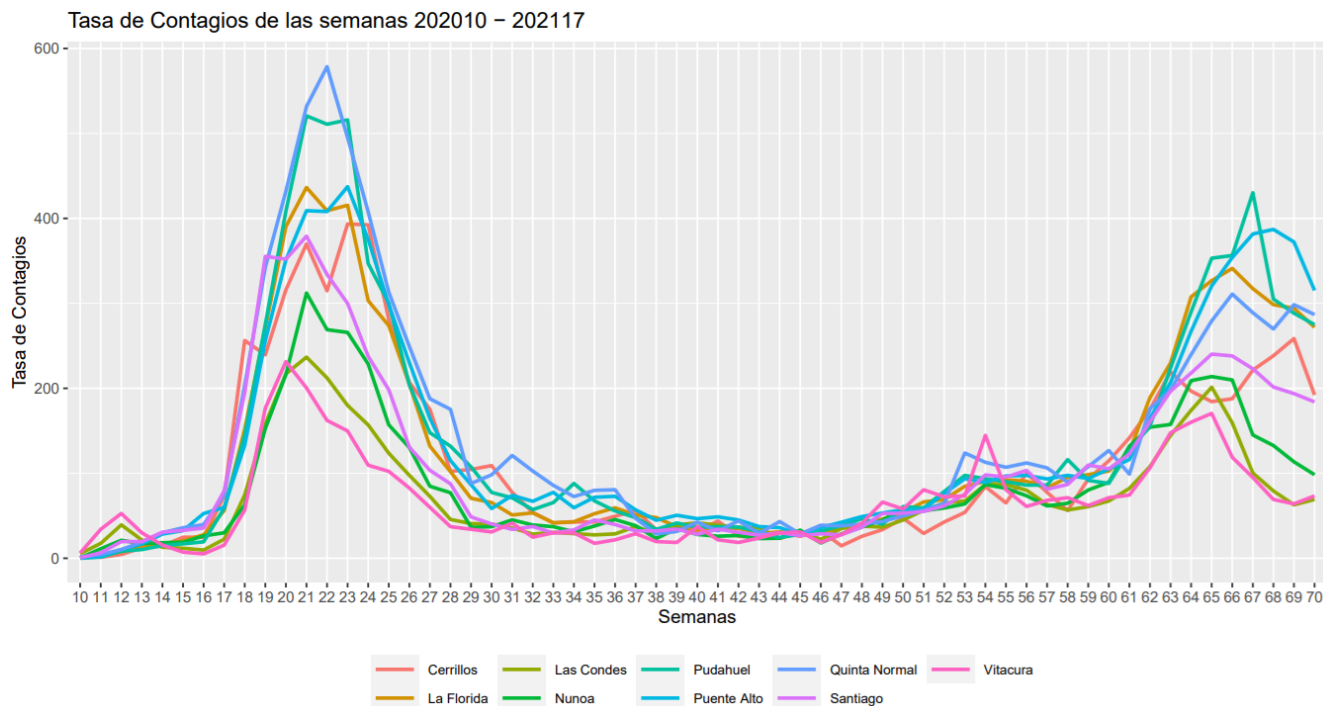


Figura 6.1: Gráfico de Líneas de la Tasa de Contagios para algunas comunas

### 6.1.1.2. Gráfico de Líneas Variación de Movilidad Completa

En el gráfico 6.2 se puede observar el comportamiento de 9 comunas seleccionadas, durante las semanas epidemiológicas 10 a la 64 (202010 y 202111).

La variable variación de movilidad tiene un comportamiento similar para todas las comunas consideradas. Además las comunas Vitacura y Las Condes, comunas que tienen un alto nivel socio-económico, durante todo el periodo de estudio son las que poseen mayor disminución en la movilidad en relación a la que tenían antes de la pandemia y en comparación con las otras comunas (a excepción de Santiago). Cabe destacar que estas comunas están ubicadas en el sector oriente de la Región, por lo que, sólo observando estas 9 comunas seleccionadas, esta inferencia apoya la hipótesis 2.

Observando los datos después de la semana 22 se puede notar que en general todas las comunas comienzan a aumentar su variación de movilidad hasta la semana 48, donde se produce una leve disminución y ya por la semana 50 hay un notorio aumento, donde ya algunas comunas sobrepasan el nivel de movilidad que tenían antes de iniciar la pandemia.

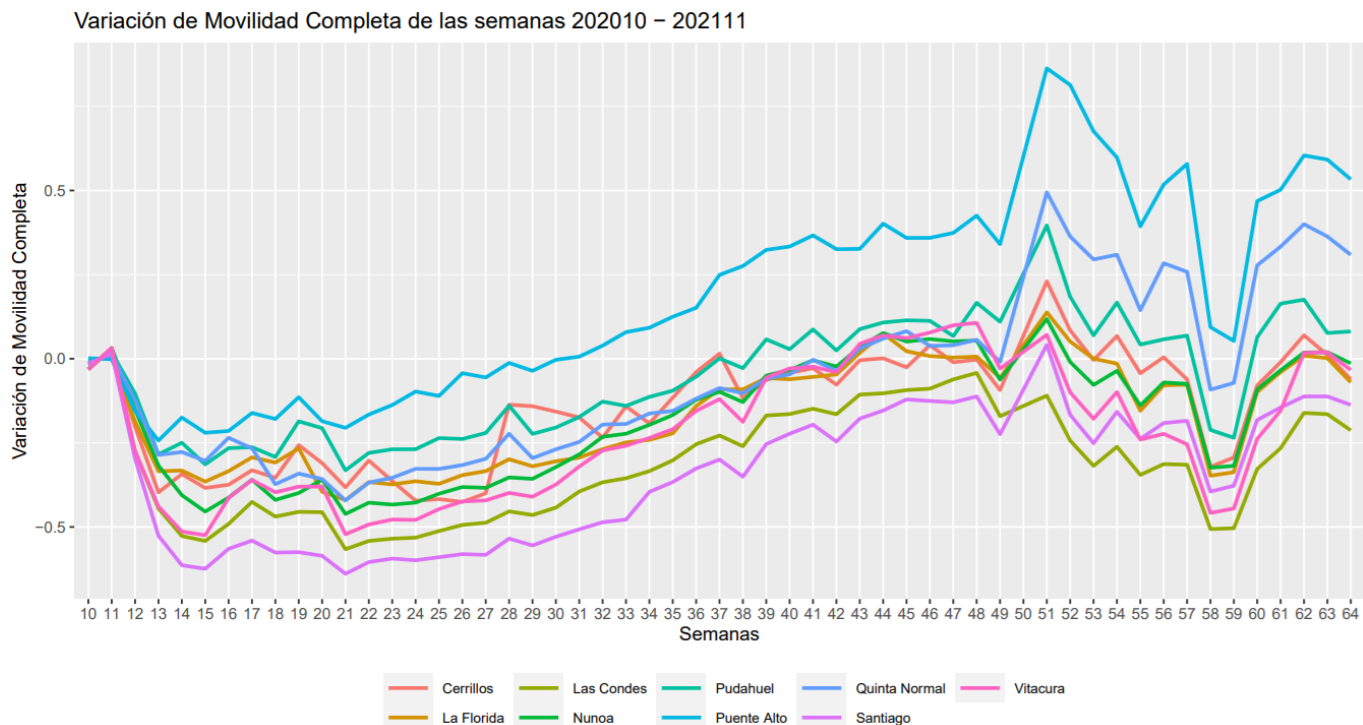


Figura 6.2: Gráfico de Líneas Variación de Movilidad Completa para algunas comunas

En el Apéndice A se puede ver el gráfico de la variación de movilidad completa de todas las comunas.

### 6.1.1.3. Gráfico de Líneas Variación de Movilidad Laboral

La segunda variable de variación de movilidad, se ve graficada en la figura 6.3, donde se consideran las 9 comunas y el mismo período de tiempo considerado en la variable tasa de contagios, semanas 10 a la 70 (202010 y 202111).

Se puede decir que de forma general la movilidad comienza a disminuir notoriamente al iniciar la pandemia. Por la semana 21 comienza un progresivo aumento en todas las comunas y ya por la semana 50 hay unas leves bajas en la movilidad. Además, al igual que la variación de movilidad completa, pero de forma un poco más marcada, las comunas de mayor nivel socioeconómico como lo son Vitacura y Las Condes, son las que más disminuyen su movilidad a lo largo del tiempo. Esto se puede deber a que estas comunas no tienen la necesidad de movilizarse para realizar actividades laborales de lunes a viernes pues, su nivel socioeconómico se los permite. Esta variable de movilidad también ayuda a respaldar la hipótesis 2.

Cabe destacar, que la variable movilidad laboral indica que ninguna comuna a superado los niveles de movilidad que habían antes de la pandemia.

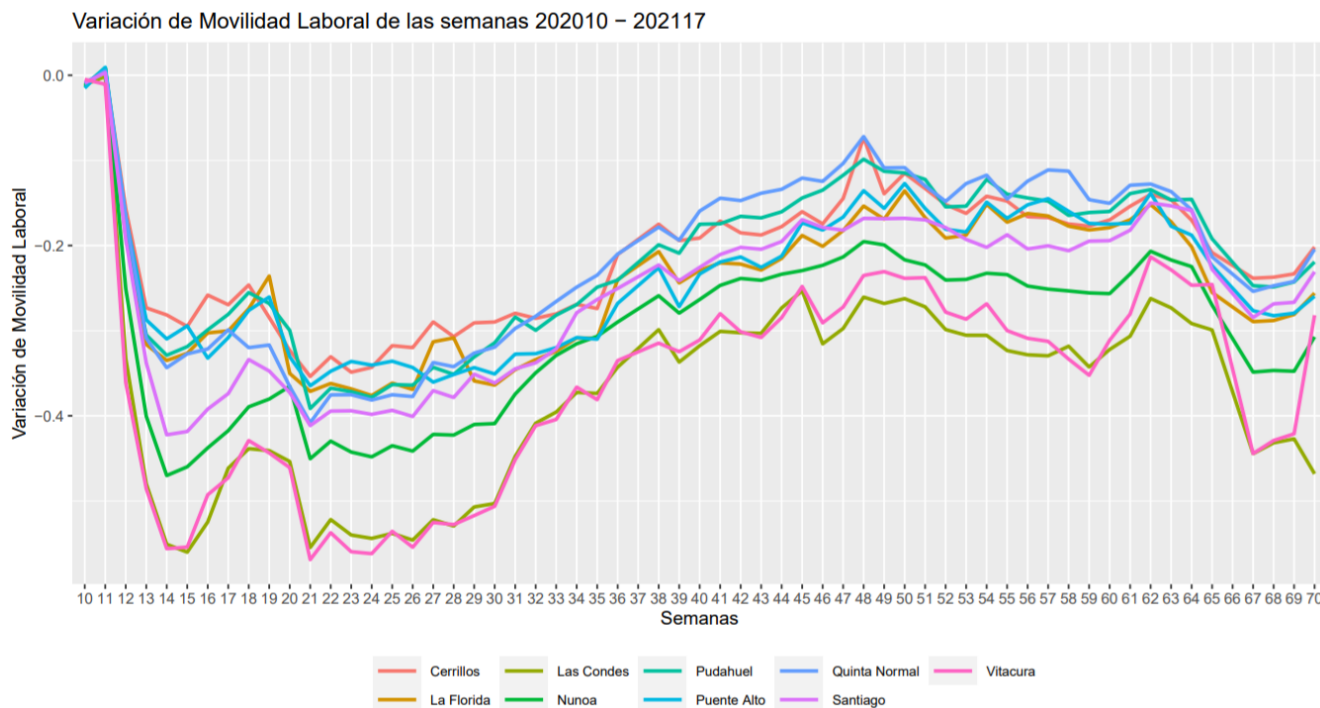


Figura 6.3: Gráfico de Líneas Variación de Movilidad Laboral para algunas comunas

En el Apéndice A se puede ver el gráfico de la variación de movilidad laboral de todas las comunas.

## 6.2. Mapas de la Región Metropolitana

Se realizan mapas de la Región Metropolitana para distintas semanas con el fin de analizar la relación espacio-temporal de las variables. Se considera la ubicación espacial de las comunas, es decir, cada una tiene asociada una localización en el espacio. Esto permite identificar patrones a nivel espacial y observar como varían durante el tiempo. Con este análisis también se puede comprobar si efectivamente las hipótesis se cumplen.

Para la realización de los mapas, es factible usar los quintiles de las variables con la finalidad de poder identificar, por la tonalidad del color, cuando aumenta o disminuye el valor de la variable.

Los quintiles se determinan considerando el periodo entre las semanas 202010 y 202111 para todas las variables. Se priorizaran sólo algunos períodos, ya que mostrar todos sería muy engorroso. Para ver el resto de los mapas, ir al Apéndice B.

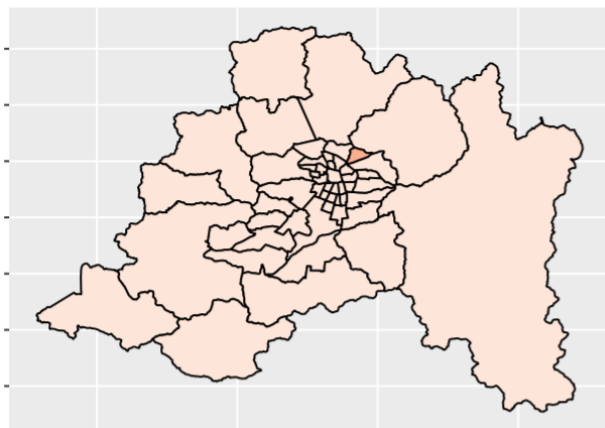
### 6.2.1. Mapas de las Tasas de Contagios

La imagen 6.4 indica la leyenda de los quintiles de la tasa de contagios que sirve para el análisis de los mapas de la figura 6.5. En esta figura, se pueden observar 8 mapas correspondientes a las semanas 202013 a 202020 de las 52 comunas de la Región Metropolitana.

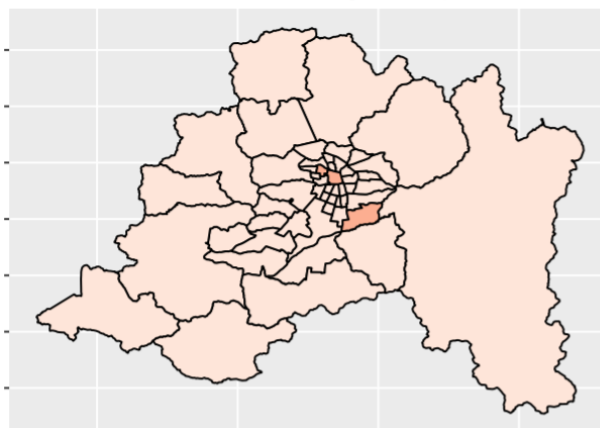


Figura 6.4: Leyenda para los mapas de la Tasa de Contagios

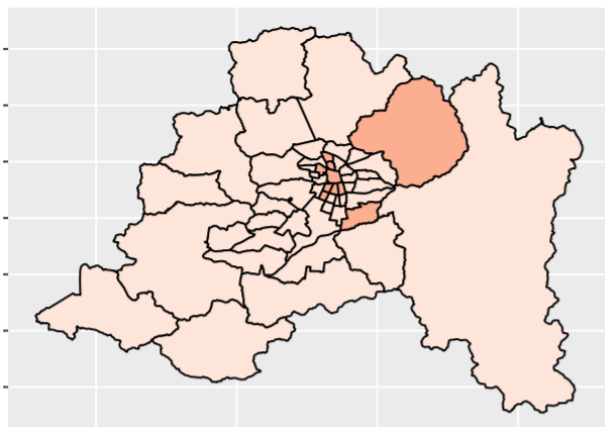
Tasa de Contagios por comuna de la semana 202013  
Considerando 202010 - 202111, Región Metropolitana



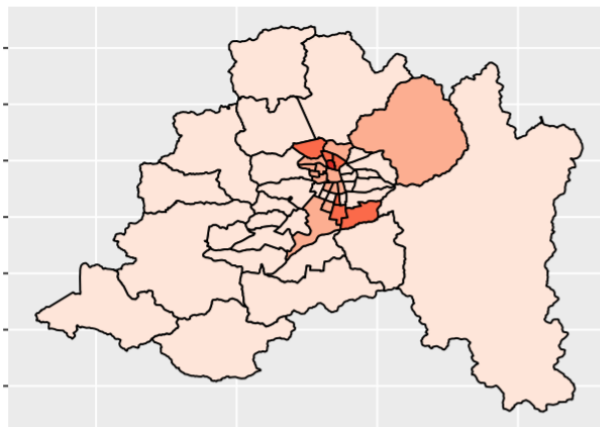
Tasa de Contagios por comuna de la semana 202014  
Considerando 202010 - 202111, Región Metropolitana



Tasa de Contagios por comuna de la semana 202015  
Considerando 202010 - 202111, Región Metropolitana



Tasa de Contagios por comuna de la semana 202016  
Considerando 202010 - 202111, Región Metropolitana



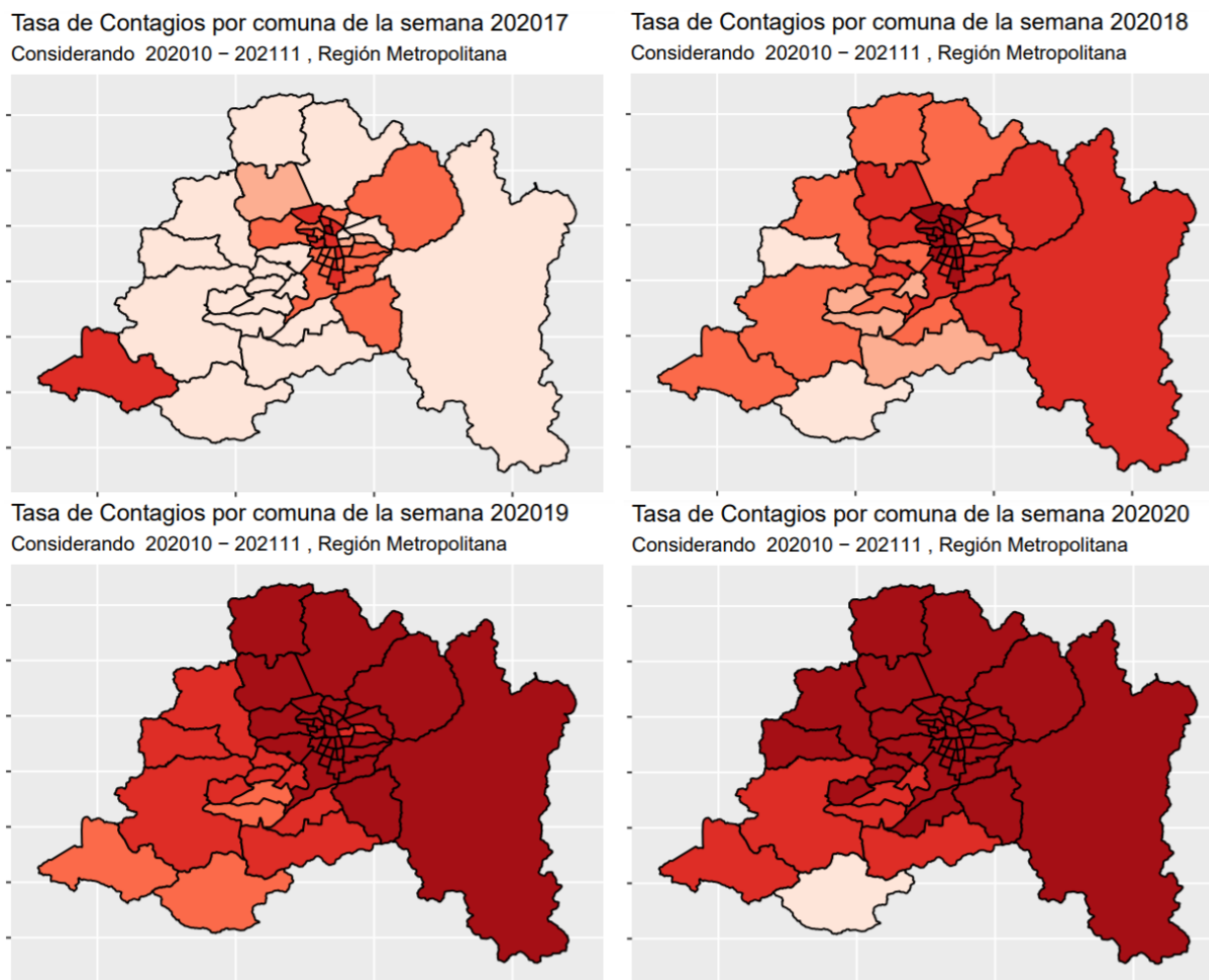


Figura 6.5: Quintiles de la Tasa de contagios de la Región Metropolitana, para algunas semanas

En los mapas se puede notar que con el paso de las semanas las tasas van aumentando en distintas comunas. En la semana 202013, se marca una comuna del sector oriente, luego en las semanas 202014, 202015 y 202016, comunas de los alrededores del mismo sector, comienzan a destacarse con una tonalidad más notoria. Es decir, los contagios comenzaron en el sector oriente de la RM.

Ya por la semana 202017, en comunas del sector poniente y otras del centro de la región, comienzan a aumentar las tasas de contagios. Y ya para las demás semanas seleccionadas, se observa que gran parte de los mapas tiene fuertes tonalidades rojas, por lo que, varias comunas poseen altas tasas de contagios. Recordar que durante esas semanas, 202018, 202019 y 202020, en el gráfico de líneas de la tasa de contagios, se muestra que hay un aumento significativo de los casos. Además en el gráfico de líneas también se puede observar que las comunas Vitacura y Las Condes

son las que comienzan con el aumento de los casos y aquí en los mapas otra vez se pudo observar eso.

Hecho este análisis, queda demostrada la hipótesis 1 que decía que el aumento de los casos de contagios de Covid-19 inició en el sector oriente de la Región Metropolitana y se fue moviendo a las comunas del sector sur-poniente de la región.

### 6.2.2. Mapas de la Variación de Movilidad Completa

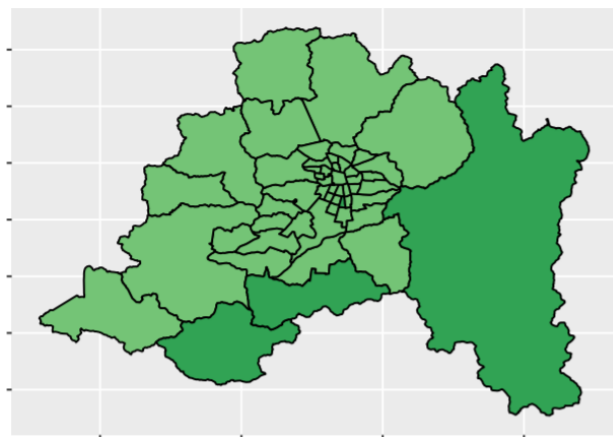
La imagen 6.6 indica la leyenda de los quintiles de la variable de variación de movilidad completa, que sirve para el análisis de los mapas de la figura 6.7.



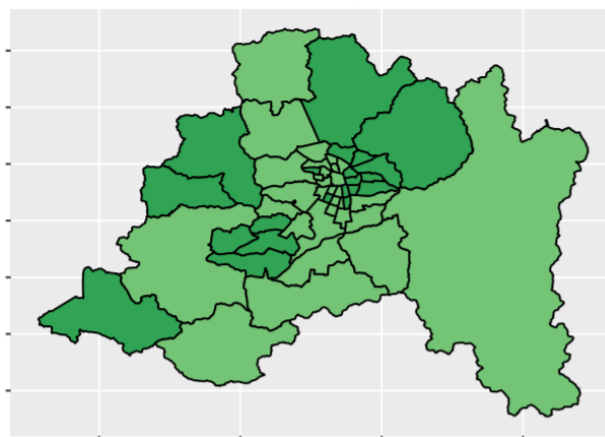
Figura 6.6: Leyenda para los mapas de la Variación de Movilidad Completa

En la leyenda, los 2 colores más claros indican variaciones negativas, siendo el quintil 1 el que indica un mayor incremento negativo en la variación de movilidad, es decir, corresponde a los valores de movilidad más bajos en comparación con los valores de movilidad normal. Los 2 colores más oscuros, indican una variación positiva, siendo el color más oscuro el quintil 5, que indica un incremento positivo en la variación de movilidad, es decir, corresponde a los valores de movilidad más altos en comparación a los valores de movilidad normal. Los valores del quintil 3, el color medio, se consideran dentro de un rango de variación de movilidad normal.

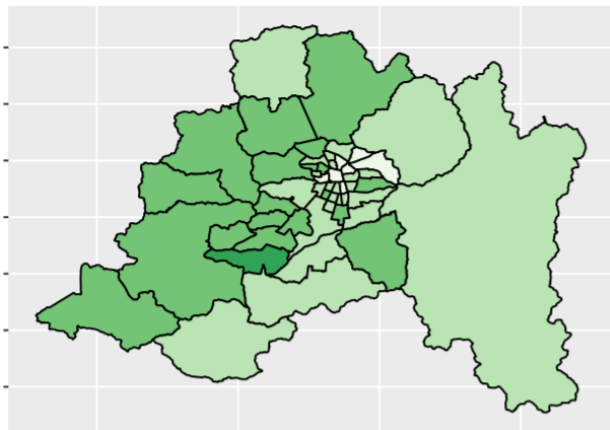
Var. Mov. Completa por comuna de la semana 202010  
Considerando 202010 – 202111, Región Metropolitana



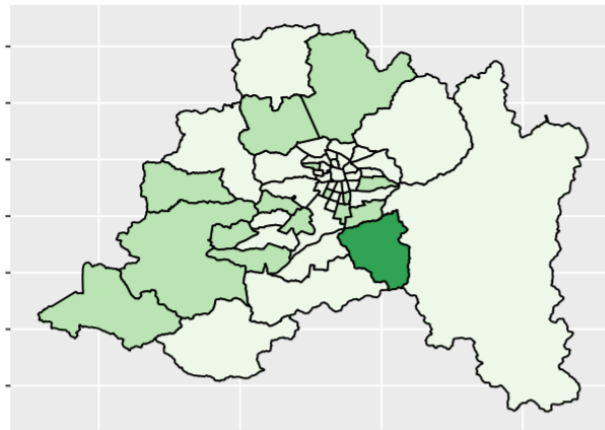
Var. Mov. Completa por comuna de la semana 202011  
Considerando 202010 – 202111, Región Metropolitana



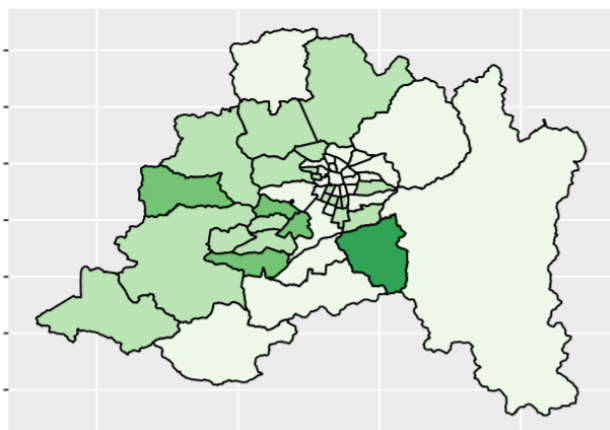
Var. Mov. Completa por comuna de la semana 20212  
Considerando 202010 - 202111, Región Metropolitana



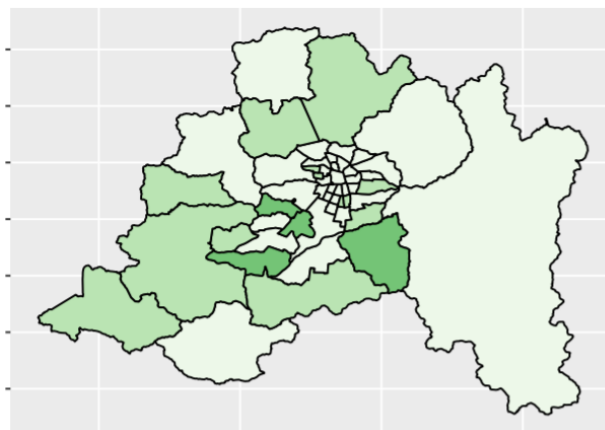
Var. Mov. Completa por comuna de la semana 20213  
Considerando 202010 - 202111, Región Metropolitana



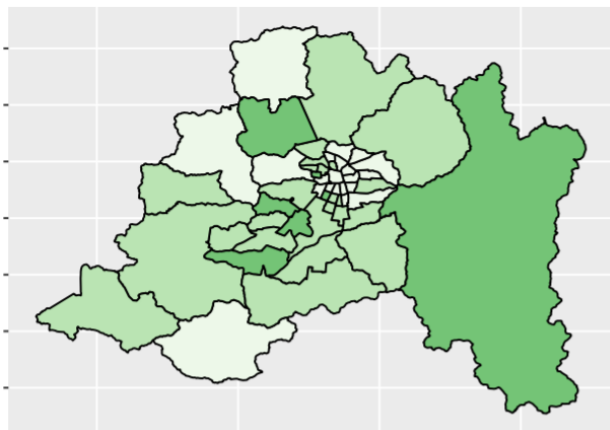
Var. Mov. Completa por comuna de la semana 20214  
Considerando 202010 - 202111, Región Metropolitana



Var. Mov. Completa por comuna de la semana 20215  
Considerando 202010 - 202111, Región Metropolitana



Var. Mov. Completa por comuna de la semana 20216  
Considerando 202010 - 202111, Región Metropolitana



Var. Mov. Completa por comuna de la semana 20217  
Considerando 202010 - 202111, Región Metropolitana

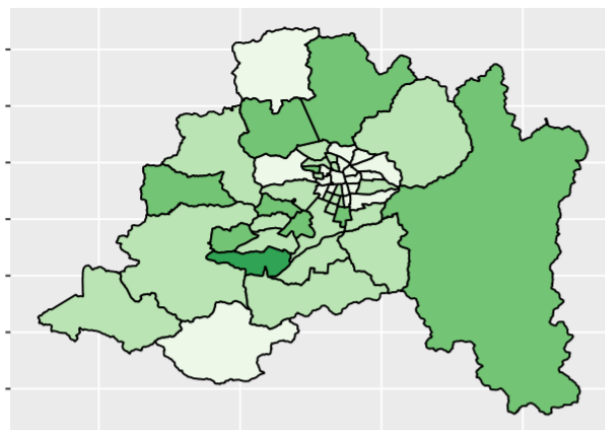


Figura 6.7: Quintiles de la Variación de Movilidad Completa, para algunas semanas

En la figura 6.7 se pueden observar 8 mapas de las semanas 202010 a 202017 de las 52 comunas de la Región Metropolitana. Analizando estos mapas se puede notar que con el paso de las semanas la variación de movilidad es bastante dispersa. En las semanas 10, 11, 12 y 13 las tonalidades del color verde van siendo cada vez más claras, eso significa que la variación de la movilidad incrementa negativamente a medida que el color es más claro.

En las semanas 202014 y 202015 se puede ver casi el mismo patrón, donde claramente la movilidad disminuyó, ya que las tonalidades de los colores son más claras pero, los colores claros se ven en las comunas del sector poniente y colores más claros aún, se ven en comunas del sector oriente, es decir, la movilidad efectivamente baja en estas semanas, pero la variación de movilidad incrementa negativamente mucho más en las comunas del sector oriente.

En los mapas de las últimas dos semanas seleccionadas, las comunas de la Provincia de Santiago del sector oriente siguen siendo las que más variación negativa posee (comunas como Vitacura, Las Condes y Providencia). Y las comunas del sector sur-poniente de la región, continúan con una variación negativa, pero la variación es menor. En los gráficos de línea de la variación de movilidad completa, también se logró ver que esto ocurría, y que considerando las 9 comunas seleccionadas, esto ocurría durante todo el período de estudio.

Con este análisis, queda demostrada la hipótesis 2, que afirmaba que durante las primeras semanas de la pandemia, la movilidad era mayor en comunas del sector sur-poniente en comparación a la movilidad de comunas del sector oriente.

### 6.2.3. Mapas de la Variación de Movilidad Laboral

La imagen 6.8 indica la leyenda de los quintiles de la variable de variación de movilidad laboral que sirve para el análisis de los mapas de la figura 6.9.

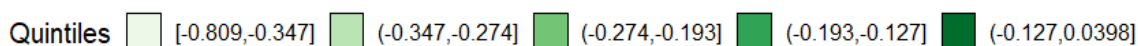
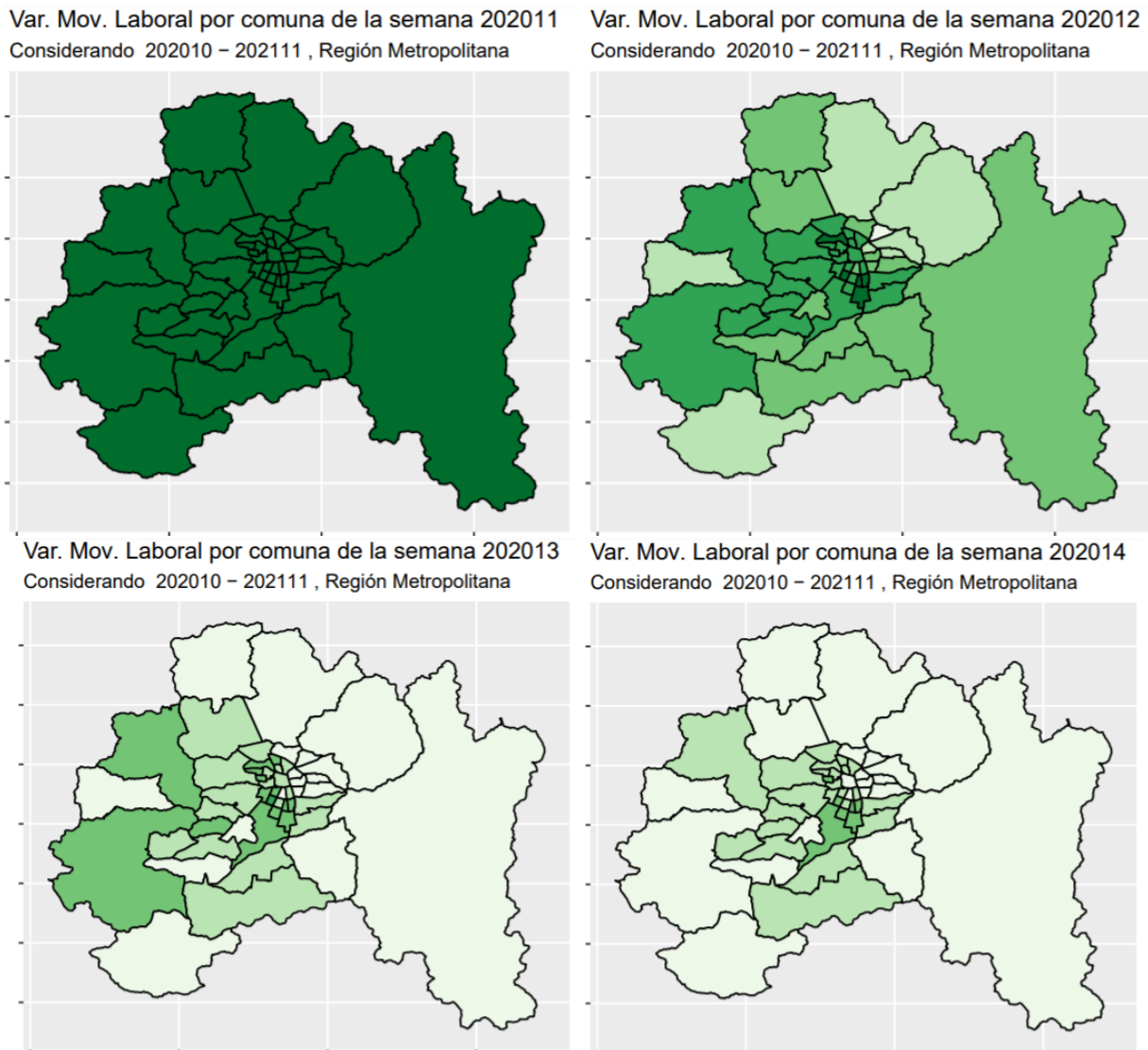


Figura 6.8: Leyenda para los mapas de la Variación de Movilidad Laboral

En la leyenda, los 4 colores más claros indican variaciones netamente negativas, siendo el quintil 1 el que indica un mayor incremento negativo en la variación de movilidad, es decir, corresponde a los valores de movilidad más bajos en comparación con los valores de movilidad normal. El color más oscuro corresponde al quintil 5, y contiene valores de baja variación positiva y negativa, por lo

que, este quintil se considera como valores de variación de movilidad normal.

En la figura 6.9 se pueden observar 8 mapas correspondientes a las semanas 202011 a 202018 de las 51 comunas de la Región Metropolitana (La comuna faltante, San José, ubicada al extremo inferior-izquierdo del mapa no está considerada en el mapa).



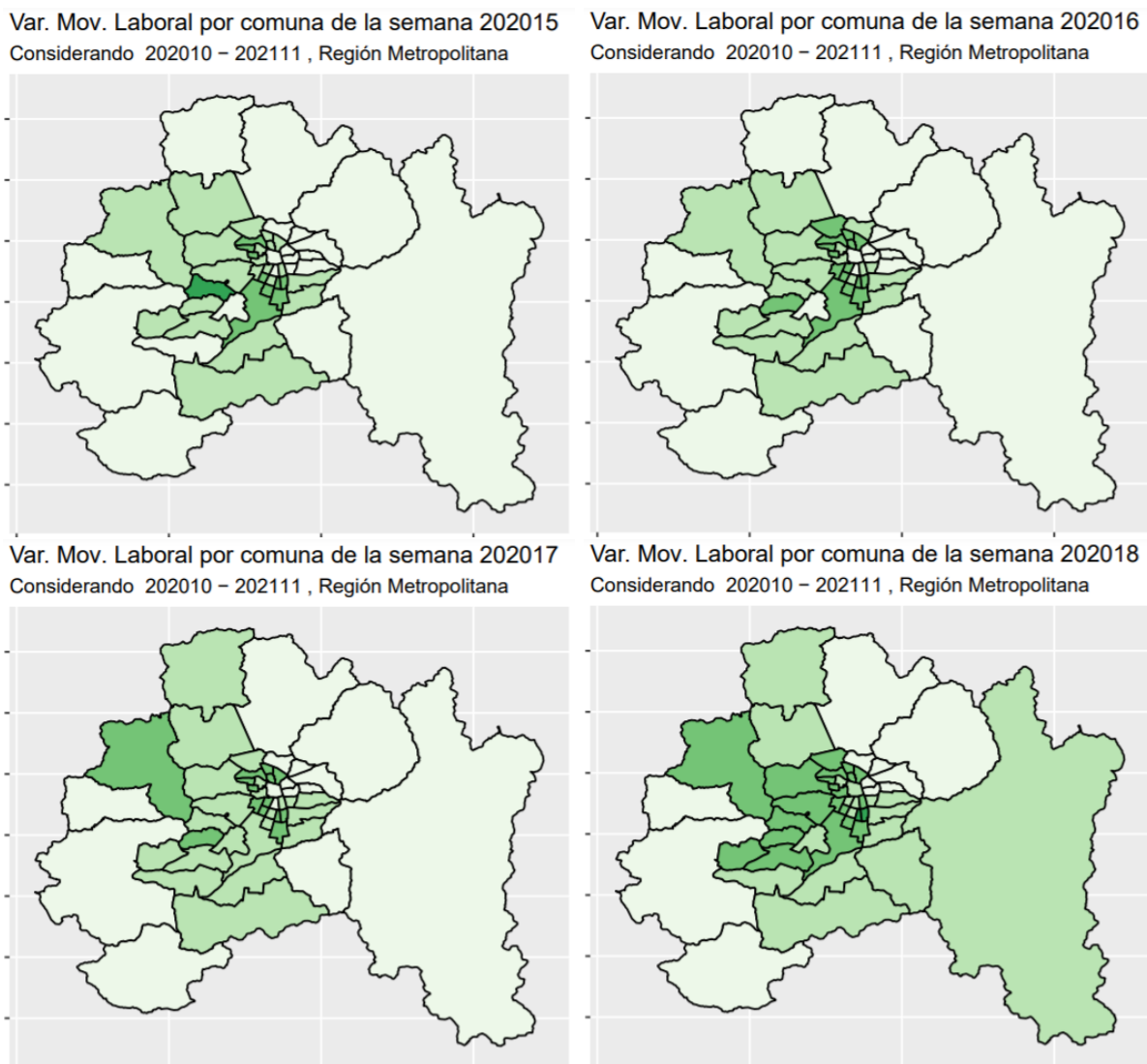


Figura 6.9: Quintiles de la Variación de Movilidad Laboral para algunas semanas

Al analizar esta variable, se espera que tenga un comportamiento similar a la variable anterior, pues ambas miden la movilidad de las personas durante la pandemia, con la diferencia de que los datos de esta variable están más enfocados a la movilidad de las personas en horarios y días de trabajo. Además, con esta variable, se espera que la hipótesis 2 se vea más marcada pues, para el planteamiento de la hipótesis 2 se intuyó que la mayor movilidad de las personas del sector poniente de la región en comparación a las comunas del sector oriente, se debe a la necesidad laboral.

Observando los mapas de las semanas 202011 y 202012, se puede notar que al iniciar la pandemia la movilidad disminuye comenzando en las comunas del sector oriente, pues los colores más

claros comienzan a notarse en ese sector. En la semana 13, todo el sector oriente tiene valores de variación de movilidad pertenecientes al primer quintil, es decir, en este sector el incremento de la variación negativa es el mayor. Se observa también que las comunas del sector poniente, disminuyen la variación pero, con un menor incremento. En el resto de las semanas seleccionadas, se observa el mismo patrón.

Al pasar las semanas, las comunas del sector oriente mantienen un mayor incremento negativo en la variación de movilidad, mientras que en el sector sur-poniente en algunas comunas sigue habiendo una variación de movilidad negativa, pero de menor incremento.

Esta variable, demuestra aún más claramente que la hipótesis 2 se cumple, y se puede inferir que la movilidad de comunas del sector poniente es mayor que las comunas del sector oriente por temas laborales, pues esta variable de variación centra el registro de sus datos en horarios y días laborales.

### 6.3. Relación entre las Variables

En esta sección se analiza la posible relación que tienen las variables, ya que, en el posterior análisis se usará la variación de movilidad para explicar la tasa de contagios. Además, se puede saber qué variable de variación de movilidad tiene más relación con la tasa de contagios. Esto se realizará observando gráficamente las variables y calculando las correlaciones entre ellas.

Para observar mejor la relación entre las variables, se grafica considerando sólo un grupo de comunas, el mismo considerado en los gráficos de líneas.

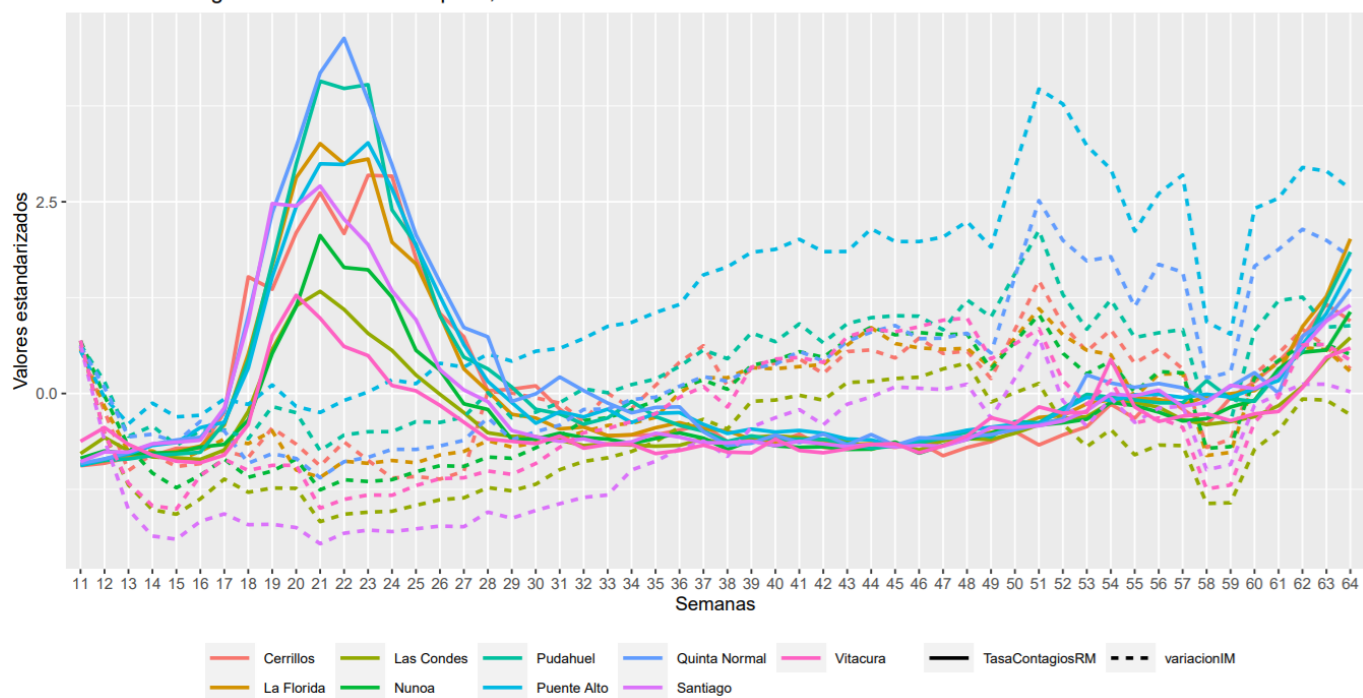
#### 6.3.1. Tasa de Contagios v/s Variación de Movilidad

Como se observó en los gráficos de líneas de ambas variables de movilidad, estas tienen un comportamiento similar, por lo que, el análisis con la variable tasa de contagios también será parecido.

Para poder analizar correctamente la comparación entre las variables, se estandarizan los valores de cada una y se grafica la tasa de contagios con cada una de las variables de movilidad.

En la figura 6.10 se pueden observar dos gráficos. El primero corresponde a la relación entre las variables tasa de contagios y variación de movilidad completa y en el segundo se grafica la relación entre las variables tasa de contagios y la variable variación de movilidad laboral.

Tasa de Contagios v/s Var. Mov. Completa, durante las semanas 202011–202111



Tasa de Contagios v/s Var. Mov. Laboral, durante las semanas 202011–202117

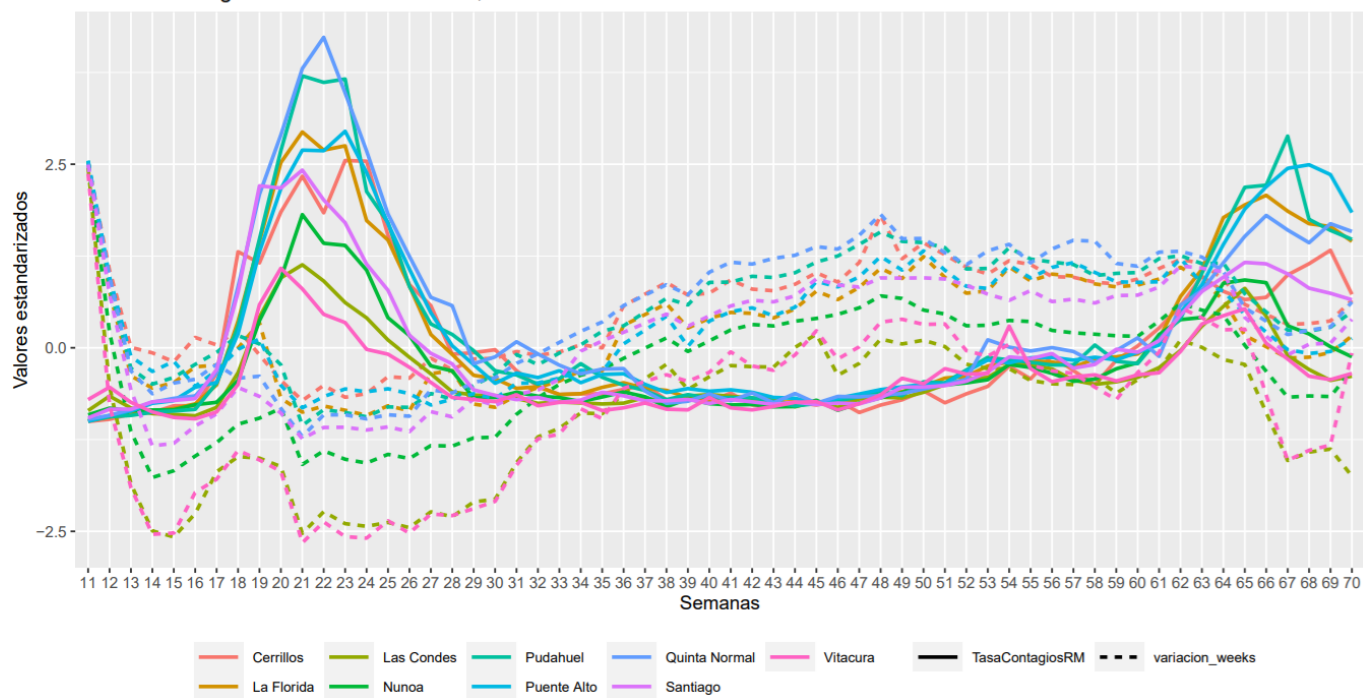


Figura 6.10: Tasa de Contagios v/s Variación de Movilidad para algunas comunas

Se destaca que el rango de tiempo considerado para el primer gráfico es menor que el del segundo gráfico, ya que, la variable variación de movilidad completa posee menos datos temporales.

Observando ambos gráficos, se ve claramente que las variables no poseen una relación lineal, esta es bastante variada. Es más factible analizar la relación por periodos. Para definirlos, se debe considerar que la variable de interés es la tasa de contagios, por ende, se definen de acuerdo a su comportamiento (que se analizó detalladamente al realizar el gráfico de líneas) y de acuerdo a la relación con la variable de movilidad.

El primer período contempla las semanas 11 a la 30, ya que, se puede ver que la variable tasa de contagios aumenta y llegando a un punto máximo según cada comuna, los casos comienzan a disminuir, mientras que en este mismo período, las variables variación de movilidad disminuyen notoriamente y a pesar de tener algunas variaciones, en general no aumentan drásticamente hasta pasar la semana 30.

El siguiente período, según el comportamiento de la variable tasa de contagios, se considera desde la semana 31 a la 59, ya que, en este rango de tiempo la variable está más acotada, es decir, los casos de contagios están más controlados, mientras que la variable variación de movilidad comienza a aumentar lentamente y pasada la semana 48, el comportamiento cambia según la variable de movilidad considerada. En el gráfico 1, después de la semana 48 el comportamiento es bastante variable aunque de forma general la variable disminuye y en el gráfico 2, pasada la semana 48 se observa que los valores de la variable de movilidad comienzan a disminuir lentamente.

El último periodo, se considera completo para la variable variación de movilidad laboral, ya que a la otra variable de movilidad le faltan más datos temporales. Por lo tanto, este período contempla las semanas 60 a la 70, donde observando el segundo gráfico se puede notar claramente, que la tasa de contagios comienza a aumentar nuevamente y llegado un punto disminuye, mientras que la variable de variación de movilidad disminuye y en algunas comunas en la última semana, aumenta nuevamente.

Por lo tanto, según el análisis, las variables poseen relación aunque marcada en algunos periodos. Como el objetivo es explicar la tasa de contagios a través de la variable de movilidad, esta última se contempla como variable independiente y la tasa de contagios como variable dependiente. Es decir, una variación en la variable de movilidad implica una variación en la variable tasa de contagios. Pero, si la tasa de contagios varía también puede implicar una variación en la variable de movili-

dad. Dicho de otra manera, si la tasa de contagios disminuye la movilidad aumenta o si la tasa de contagios aumenta la movilidad disminuye. Esto se produce porque hay un efecto endógeno en la variable, pues se trata de explicar los contagios a través de la movilidad pero, la movilidad también depende de los contagios debido a que a mayor contagios, el gobierno de Chile coloca medidas de sanitarias como cuarentenas en las comunas y eso impacta en la movilidad.

El efecto endógeno está más presente cuando la relación entre las variables también esté presente. Esto se puede observar más en el periodo 1.

### 6.3.2. Correlaciones entre las Variables

Para ver si efectivamente hay dependencia entre las variables y cuantificar la relación entre ellas, se determina el coeficiente de correlación por cada semana. El valor de este coeficiente varía entre -1 y 1, donde estos extremos indican una fuerte asociación entre las variables y el valor 0 indica una nula relación entre ellas.

En la tabla (a) de la figura 6.11 se observan las correlaciones absolutas, para las semanas 11 a la 30, entre la tasa de contagios y la variación de movilidad completa (columna 2) y las correlaciones entre la tasa de contagios y la variación de movilidad laboral (columna 3). Donde claramente se observa que en su gran mayoría las correlaciones son más altas en la columna 3, indicando que la relación entre variables es más fuerte considerando la variación de movilidad laboral. En general, estos valores están bastante concentrados entre 0.2 y 0.6.

Para el periodo entre las semanas 31 a la 59, nuevamente las correlaciones son mayores entre las tasas de contagios y la variación de movilidad laboral, pero los valores van desde 0 a 0.34, siendo muy bajas para decir que la relación es fuerte. En el Apéndice C se puede ver la tabla.

Para el último periodo, que contempla las semanas 60 hasta la 70, las correlaciones se pueden observar en la tabla (b) de la figura 6.11. Donde las correlaciones entre la tasa de contagios y la variación de movilidad laboral se concentran en su mayoría entre los valores 0.1 y 0.4. Por lo que, hay una relación débil en comparación a la del primer periodo.

Semana	Corr.VMCompleta	Corr.VMLaboral
202011	0.211	0.296
202012	0.46	0.653
202013	0.337	0.196
202014	0.271	0.175
202015	0.169	0.418
202016	0.065	0.422
202017	0.127	0.452
202018	0.234	0.481
202019	0.195	0.41
202020	0.314	0.344
202021	0.001	0.433
202022	0.14	0.492
202023	0.308	0.513
202024	0.284	0.532
202025	0.263	0.526
202026	0.388	0.218
202027	0.077	0.049
202028	0.049	0.025
202029	0.095	0.006
202030	0.069	0.08

Semana	Corr.VMCompleta	Corr.VMLaboral
202107	0.206	0.141
202108	0.098	0.146
202109	0.059	0.217
202110	0.173	0.291
202111	0.236	0.127
202112	NA	0.271
202113	NA	0.363
202114	NA	0.142
202115	NA	0.246
202116	NA	0.463
202117	NA	0.405

(a) Semanas 11 a 30 (202011 - 202030)

(b) Semanas 60 a 70 (202107 - 202117)

Figura 6.11: Correlaciones entre las variables Tasa de Contagios y Variación de Movilidad

## 6.4. Análisis de Autocorrelación Espacial

Este análisis permite saber si las variables poseen dependencia espacial significativa, es decir, si los valores de la variable dependen significativamente de la comuna que se observa. Para determinar esto, se calcula el Índice de Morán Global, el cual considera la matriz  $W$  de pesos espaciales para su cálculo. Esta matriz es la que indica la relación entre cada par de comunas.

### 6.4.1. Cálculo de la Matriz de Pesos Espaciales

Se define la matriz de pesos espaciales para analizar la asociación espacial de los datos. El criterio utilizado para definir esta matriz, es el de vecindad contigua (por ser el criterio de vecindad más abarcable en este análisis) de tipo Reina, debido a la irregularidad de las áreas (no es factible considerar el tipo Alfil o Torre, ya que, no se pueden definir). Es decir, todas las comunas que

comparten límite con una comuna específica, se consideran vecinas de esa comuna.

La matriz formada es cuadrada de dimensión  $N \times N$ , donde  $N$  representa el número de comunas y además es simétrica. Se consideran dos matrices de pesos, una de dimensión  $52 \times 52$  y otra de dimensión  $51 \times 51$ , ya que, la variable de variación de movilidad laboral no presenta los datos de la comuna San Pedro, por lo que, la base que contiene esta variable sólo considera 51 comunas. Las filas y columnas de la matriz, representan 52 o 51 comunas ordenadas de forma alfabética.

### Comunas vecinas de la comuna Alhué

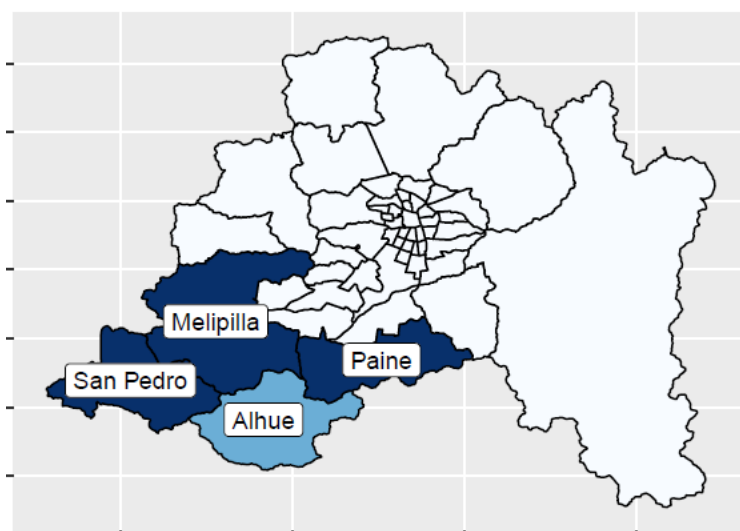


Figura 6.12: Mapa de las comunas vecinas de Alhué, Región Metropolitana

A modo de ejemplo, se toma en cuenta la comuna de Alhué (alfabéticamente esta es la primera). En la figura 6.12 se puede observar que la comuna de Alhué tiene como vecinas a las comunas Melipilla, Paine y San Pedro.

En la matriz  $W$ , a Alhué le corresponde la primera fila, esto indica que, la fila 1 con cada columna indicará la relación que tenga la comuna de Alhué con cada una del resto de las comunas. Cada columna corresponde a cada una de las 52 comunas ordenadas alfabéticamente, por lo tanto, la posición  $W_{1,1}$  corresponde a la relación entre Alhué con Alhué. Los valores de la diagonal son 0, pues no se puede tener una relación de vecindad entre la misma comuna. El resto de los valores entre las relaciones de cada comuna con Alhué, será de 1 si las comunas son vecinas de Alhué y 0 si no son vecinas.

Las posiciones de las columnas para las comunas Melipilla, Paine y San Pedro, son 28, 32 y 47 respectivamente, por tanto, los valores de  $W_{1,28}$ ,  $W_{1,32}$  y  $W_{1,47}$  serán 1 y el resto será 0. La matriz  $W$ , completada con la fila 1, quedaría entonces de la siguiente manera:

$$W_{52 \times 52} = \begin{matrix} & i,1 & i,2 & \dots & i,27 & i,28 & i,29 & \dots & i,31 & i,32 & i,33 & \dots & i,46 & i,47 & i,48 & \dots & i,52 \\ \left[ \begin{array}{cccccccccccccccccccc} 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ & 0 & & & & & & & & & & & & & & & \\ & & & & 0 & & & & & & & & & & & & \\ & & & & & & & & \dots & & & & & & & & \\ & & & & & & & & & & & & 0 & & & & \\ & & & & & & & & & & & & & & & & \\ & & & & & & & & & & & & & & & & 0 \end{array} \right] \end{matrix}$$

Luego de calcular la matriz, se normaliza por fila, ya que, según la literatura esto mejora las propiedades estadísticas de los estimadores.

### 6.4.2. Análisis de Autocorrelación Espacial Global

Se calcula el índice de Moran global, para cada una de las semanas, considerando los valores de cada variable de estudio asociados a cada comuna y considerando la relación de vecindad entre las comunas a través de la matriz  $W$ .

Cada valor del índice tiene asociado un p-valor, que se espera sea menor al nivel de significancia considerado ( $\alpha = 0,05$ ), para rechazar la hipótesis nula definida en la subsección 4.5.1 “Test de Asociación Espacial Global”, con el fin de identificar la existencia de autocorrelación espacial global en la semana analizada.

En la figura 6.13, se observan los índices de Moran Global significativos y no significativos de cada una de las variables por cada semana considerada. En general se puede ver que para las variables de variación de movilidad, existe evidencia significativa que indica la presencia de asociación espacial global para todas las semanas consideradas. Mientras que en la tasa de contagios, en algunas semanas, existe evidencia significativa que permite asegurar la existencia de la asociación espacial global pero, para otras semanas no.

En la gráfico también se puede observar que se contemplan los periodos anteriormente definidos, con el fin de analizar si existe relación entre el periodo y la autocorrelación espacial global. Esta dependerá sólo de la tasa de contagios, pues las variables de movilidad poseen autocorrelación espacial en todas las semanas.

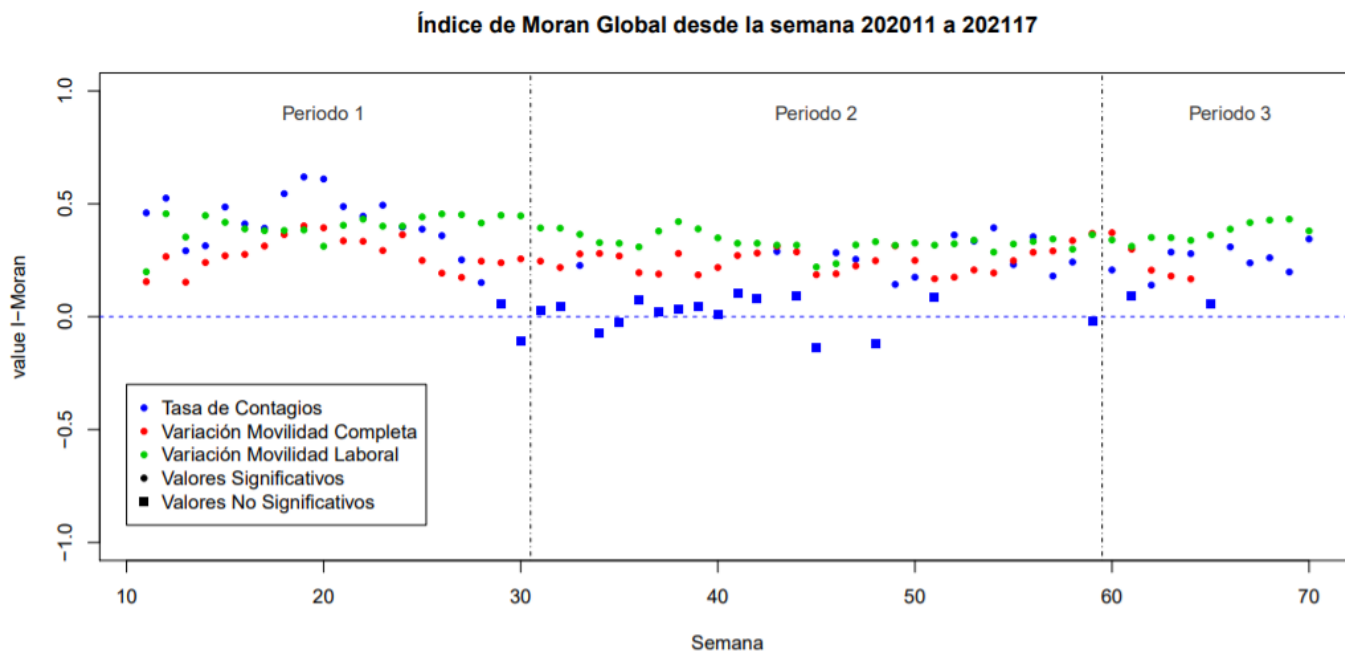


Figura 6.13: Valores del Índice de Moran Global significativos y no significativos

Se puede notar que el primer periodo está representado por una notoria presencia de asociación espacial global de la variable de interés. Pues, casi todos los puntos de la tasa de contagios, a excepción de dos valores de las últimas semanas, son significativos, es decir, la mayoría indica que durante las semanas 11 a la 30, la tasa de contagios posee autocorrelación espacial global.

El periodo 2, se ve destacado por ser el rango de tiempo donde menos existe asociación espacial global significativa para la variable tasa de contagio, pues los valores del índice de esta variable, son mayoritariamente no significativos.

Por último, el tercer periodo, a pesar de ser más pequeño que el periodo 1, también se puede notar que existe dependencia espacial global en la variable de interés. Exceptuando dos valores asociados a la tasa de contagios que no son significativos, el resto de los valores si indica una autocorrelación espacial global.

Cabe destacar, que la dependencia espacial global se dice significativa si se mantiene en todas las comunas consideradas o si está presente en algunos grupos de comunas. Para saber esto, se pueden calcular los índices de Moran Local para cada comuna durante cada semana, ya que, las comunas que posean valores significativos indicarán que existe autocorrelación espacial entre la comuna y sus comunas vecinas. Esto permite visualizar patrones.

Para esta tesis, el análisis de autocorrelación espacial local es complementario, ya que, basta con saber que en algunas semanas hay presencia de dependencia espacial global para que al momento de modelar las variables, la autocorrelación espacial existente sea considerada en el modelo, ya que, de no ser así las estimaciones de los parámetros serían sesgadas, produciendo que las pruebas de significancia no sean estables.

## 6.5. Análisis de Autocorrelación Temporal

En este análisis se estudia la correlación a nivel temporal, es decir, se analiza si los datos de cada comuna poseen una asociación significativa con el tiempo. Para determinar esto se implementa el test de Ljung-Box para cada variable según los datos de cada comuna.

Considerando todo el período temporal de estudio para las 52 comunas (dependiendo de la variable analizada), los p-valores asociados al estadístico de Ljung-Box son menores al nivel de significancia, es decir, hay evidencia significativa para no dudar de la existencia de la autorrelación espacial en cada comuna y para cada variable. Estos valores no se muestran todos, ya que, sería demasiado engorroso pero, a modo de ejemplo se considera el análisis de una de las comunas.

En la figura 6.14 se puede observar una tabla con los p-valores y sus respectivos estadísticos, asociados a 10 lags o retardos de la serie de la comuna de Las Condes, para cada variable. Las siglas TC hacen referencia a la variable Tasa de Contagios, VMC hace referencia a la Variación de Movilidad Completa y la sigla VML significa Variación de Movilidad Laboral.

Se puede observar que todos los p-valores son 0, por lo que, efectivamente se rechaza la hipótesis nula, definida en la subsección 4.7.1 “Test de Box-Ljung”, a favor de la presencia de autocorrelación temporal. Estos valores son similares para el resto de todas las comunas, los p-valores son casi 0. En el apéndice D se pueden observar las tablas para algunas comunas consideradas.

P-valores Ljung-Box, Comuna de Las Condes, considerando todas las semanas

lag	p.value.TC	statistic.TC	p.value.VMC	statistic.VMC	p.value.VML	statistic.VML
1	0	52.3	0	41.195	0	36.691
2	0	85.37	0	70.712	0	60.684
3	0	100.032	0	93.708	0	77.268
4	0	103.779	0	110.841	0	89.824
5	0	103.917	0	125.569	0	101.339
6	0	104.546	0	138.682	0	113.472
7	0	106.985	0	147.468	0	124.995
8	0	110.676	0	153.584	0	134.212
9	0	114.278	0	157.059	0	140.798
10	0	117.11	0	157.979	0	143.813

Figura 6.14: P-valores del test de Ljung-Box para la comuna de Las Condes

Al igual que en el estudio de la autocorrelación espacial, se considera relevante estudiar qué pasa con la autocorrelación temporal en los distintos periodos previamente definidos. Este análisis se mostrará para un grupo de comunas, ya que, con el resto el comportamiento es similar. Y además se considera sólo un retardo, porque la cantidad de datos de los 3 períodos es pequeña.

La tabla de la figura 6.15, contiene los p-valores y estadísticos asociados a las series del primer periodo, este considera 20 datos temporales. Se puede observar que, algunos de los p-valores superan el nivel de significancia en algunas comunas asociadas a datos de movilidad. Estas comunas son Las Condes, Ñuñoa, Vitacura, La Florida y Pudahuel. Es decir, para las variables de variación de movilidad, considerando el periodo entre las semanas 202011 a 202030, hay evidencia significativa para asegurar que no existe autocorrelación temporal para las comunas mencionadas. Los valores de las variables de movilidad no dependen del tiempo. Para la variable tasa de contagios, por el contrario, para todas las comunas hay evidencia que indica que existe autocorrelación temporal, es decir, los valores de la tasa de contagio dependen del tiempo. Para visualizar de mejor forma esto, se puede observar la figura 6.16, donde para las comunas mencionadas efectivamente se ve que los puntos superan el nivel 0.05 para las variables de movilidad.

P-valores Ljung-Box, periodo considerado entre las semanas 202011 a 202030

lag	p.value.TC	statistic.TC	p.value.VMC	statistic.VMC	p.value.VML	statistic.VML	Comuna
1	0	17.320	0.041	4.163	0.034	4.487	Santiago
1	0	17.898	0.057	3.616	0.087	2.923	Las Condes
1	0	17.490	0.018	5.612	0.075	3.174	Nunoa
1	0	18.274	0.017	5.717	0.021	5.289	Quinta Normal
1	0	16.099	0.100	2.713	0.114	2.494	Vitacura
1	0	18.255	0.005	8.013	0.031	4.663	Puente Alto
1	0	18.192	0.058	3.582	0.055	3.696	La Florida
1	0	18.075	0.112	2.532	0.027	4.905	Pudahuel
1	0	16.824	0.040	4.212	0.025	5.001	Cerrillos

Figura 6.15: P-valores del test de Ljung-Box considerando el primer periodo

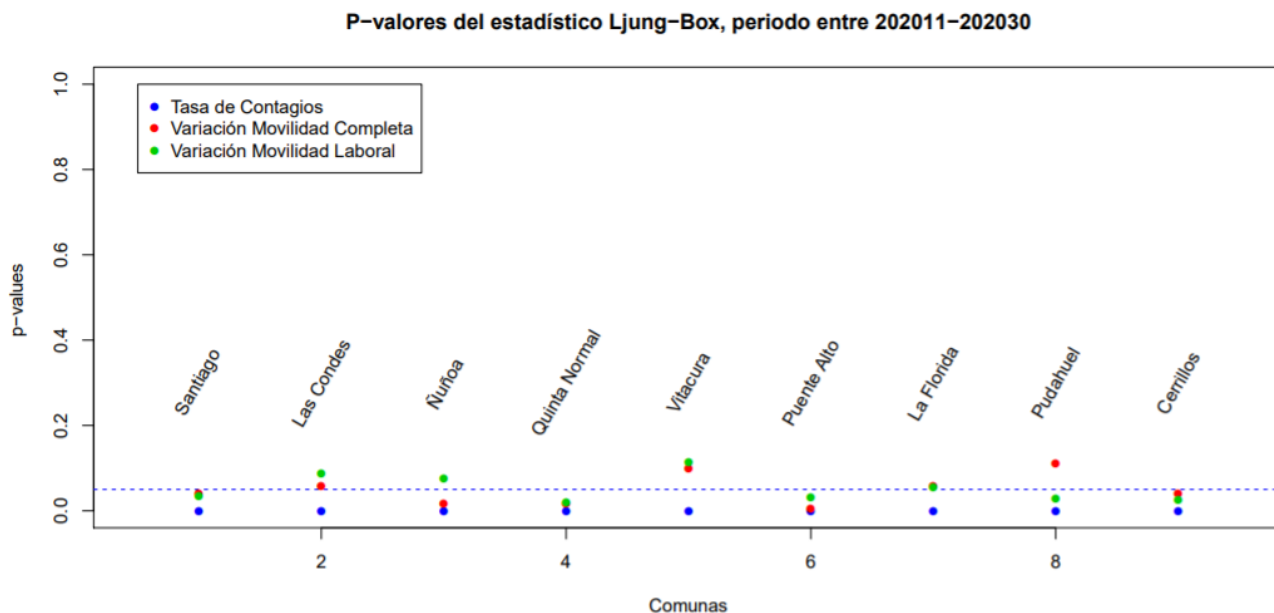


Figura 6.16: P-valores del test de Ljung-Box considerando el primer periodo

Analizando el periodo 2, que considera los datos de contagios cuando están más controlados, se espera que no exista autocorrelación temporal al menos en la variable de contagios, pues en los gráficos de líneas no se ve una notoria variación, el comportamiento es casi lineal y horizontal. Observando la tabla de la figura 6.17, se puede verificar esto. Los p-valores de todas las variables, son valores nulos o cercanos a 0, por lo que, esto indica que hay evidencia significativa para asegurar que los datos si dependen del tiempo.

P-valores Ljung-Box, periodo considerado entre las semanas 202031 a 202106

lag	p.value.TC	statistic.TC	p.value.VMC	statistic.VMC	p.value.VML	statistic.VML	Comuna
1	0.000	22.511	0.000	16.261	0	22.507	Santiago
1	0.000	25.398	0.000	18.942	0	16.920	Las Condes
1	0.000	21.976	0.000	15.311	0	21.812	Nunoa
1	0.000	20.143	0.000	19.619	0	22.781	Quinta Normal
1	0.000	16.741	0.000	19.278	0	19.011	Vitacura
1	0.000	24.754	0.000	18.116	0	23.973	Puente Alto
1	0.000	25.417	0.000	16.574	0	21.879	La Florida
1	0.000	22.676	0.000	14.475	0	25.674	Pudahuel
1	0.001	10.919	0.001	10.538	0	21.008	Cerrillos

Figura 6.17: P-valores del test de Ljung-Box considerando el segundo periodo

Como se mencionó al iniciar este análisis del periodo 2, se esperaba que la asociación espacial no existiera pero, observando más de cerca los datos, ¿cambiará esta noción? La figura 6.18, muestra que los datos efectivamente poseen variación, la tendencia que siguen parece bajar y subir levemente, por lo que, debido a este comportamiento los resultados del test de Ljung-Box muestran que efectivamente si hay asociación temporal.

Por último, analizando el periodo 3, que contempla las semanas 202107 a 202117 y en el cual sólo se consideran la tasa de contagios y la variación de movilidad laboral, se puede observar en el gráfico de la figura 6.19, que la mayoría de los p-valores se encuentran por debajo del nivel de significancia, por lo que, existe evidencia para no dudar de la dependencia espacial de las variables con el tiempo en las distintas comunas. La tabla de estos valores, se puede encontrar en el apéndice D.

Al estudiar la dependencia espacial, se expresó que para esta tesis es importante saber si los datos poseen autocorrelación, ya sea, espacial o temporal, para que esta pueda ser contemplada al modelar las variables. Por lo tanto, basta con la dependencia temporal que poseen algunas o todas las comunas para no obviar esta información y contemplarla posteriormente.

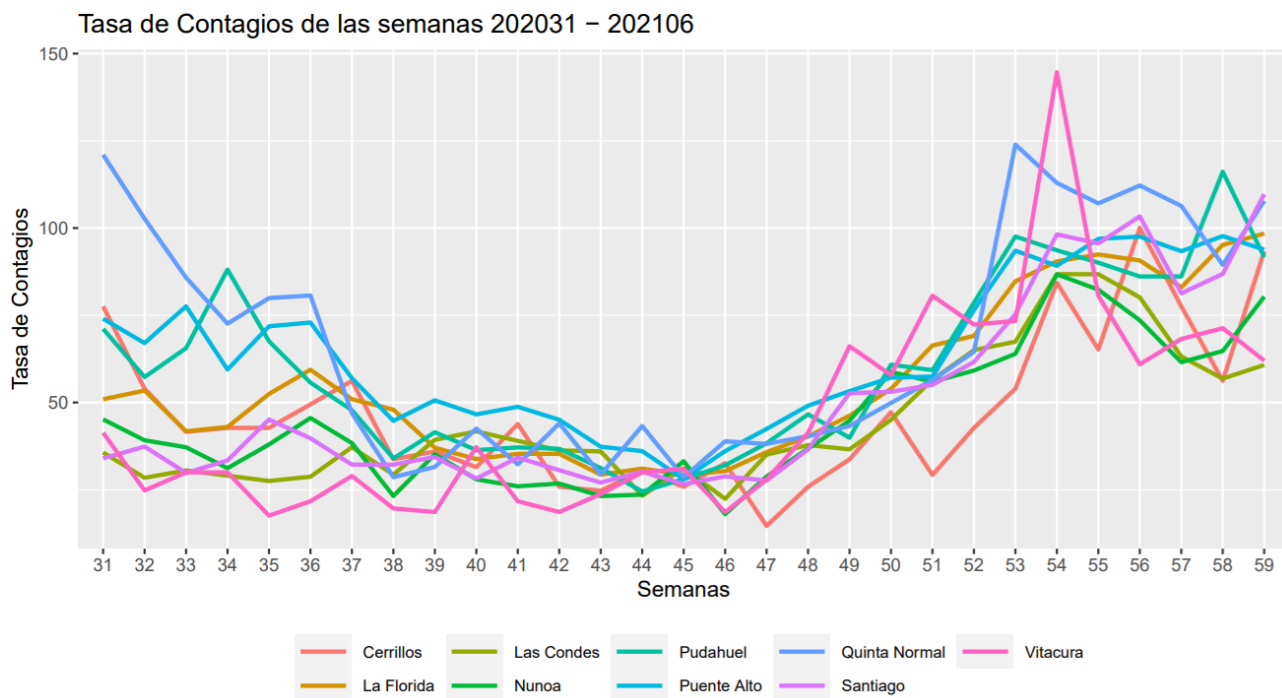


Figura 6.18: Gráficos de líneas de la Tasa de Contagios, para el periodo 2

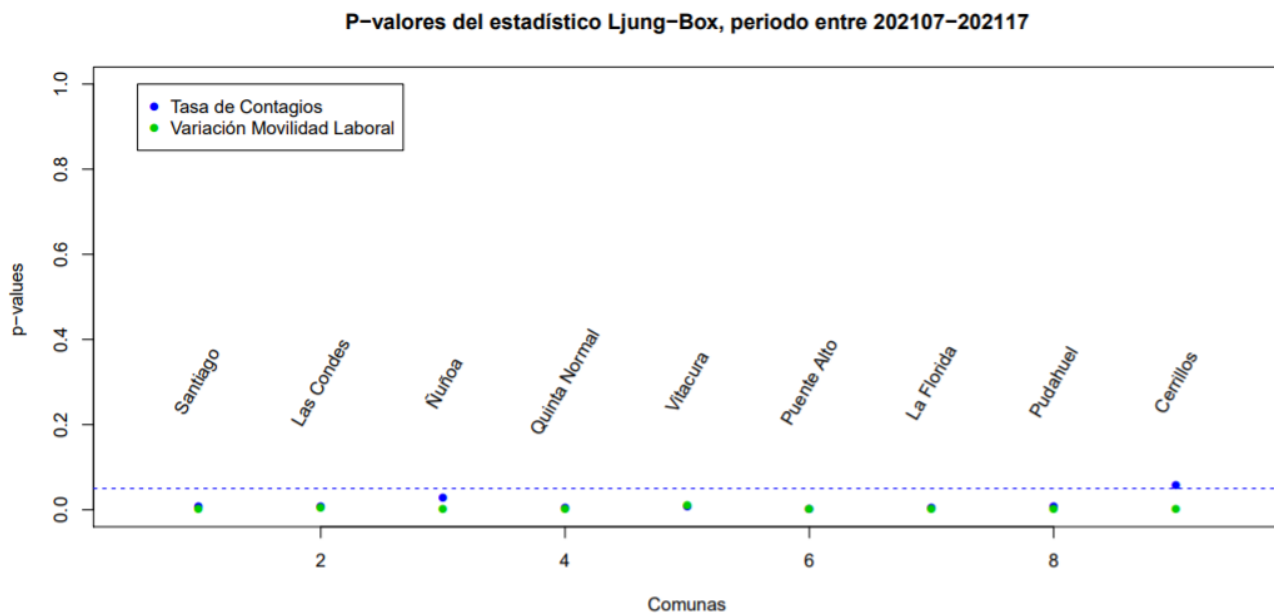


Figura 6.19: P-valores del test de Ljung-Box considerando el tercer periodo

# Capítulo 7

## Análisis Práctico de los Modelos Espaciales con Datos de Panel

Para comenzar con la implementación de los modelos, se definen las relaciones entre variables que se contemplan para ingresar al modelo. Con el estudio previo, se pudo notar que la tasa de contagios tiene más relación con la variación de movilidad laboral, pero aún así se consideran ambas variables de movilidad para estudiar ambas relaciones. Las ecuaciones de variables para implementar en los modelos son:

- Relación 1: Tasa de Contagios  $\sim$  Variación de Movilidad Completa
- Relación 2: Tasa de Contagios  $\sim$  Variación de Movilidad Laboral

Además con el estudio previo, también se pudo determinar que al dividir el periodo de estudio según el comportamiento de la variable de interés, la asociación espacial y/o temporal caracterizaban a cada uno de ellos. Estos periodos también serán considerados al implementar los modelos. Quedan definidos como:

- Periodo Total: Considera las semanas 202011 a 202117 para la relación 1 y desde la semana 202011 a 202111 para la relación 2.
- Primera Ola: Considera las semanas 202011 a 202030 para ambas relaciones.
- Periodo entre olas: Considera las semanas 202031 a 202106 para ambas relaciones.
- Segunda Ola: Considera las semanas 202107 a 202117 sólo para la relación 2.

## 7.1. Implementación de las Pruebas LM

Los modelos espaciales con datos de panel, vistos en la sección 4.10 “Modelos Espaciales con Datos de Panel”, son varios los que se pueden implementar. El análisis anterior de los datos, permitió identificar que los datos poseen asociación espacial y temporal, por lo tanto, los modelos más óptimos para implementar serían el SEMSR o el SAREMSR.

Para poder verificar esto o respaldar esto, se aplican las pruebas de Efectos Aleatorios, Correlación Espacial y Serial en el Error o Pruebas LM conjunta o condicionales de Baltagi, con el fin de analizar qué modelo más simple que el general, es más óptimo para implementar. Estas se determinan usando el comando `bsjktest` del paquete `splm` en el software R.

En la figura 7.1 se observa una tabla que contiene los estadísticos de las pruebas LM con sus respectivos p-valores, para cada relación de variables, considerando el periodo total del estudio.

**Test de Efecto Aleatorio, Correlación Espacial y  
Correlación Serial en los errores, Periodo Total**

test	variables	LM	p_value
J	TasaContagios ~ V.M.Completa	7321.227	0
C1	TasaContagios ~ V.M.Completa	55.211	0
C2	TasaContagios ~ V.M.Completa	43.341	0
C3	TasaContagios ~ V.M.Completa	-0.368	1
J	TasaContagios ~ V.M.Laboral	8348.175	0
C1	TasaContagios ~ V.M.Laboral	45.483	0
C2	TasaContagios ~ V.M.Laboral	45.784	0
C3	TasaContagios ~ V.M.Laboral	-0.626	1

Figura 7.1: Pruebas LM, considerando el periodo total

Los p-valores de la tabla son similares para cada relación de variables, por lo que, el análisis es el mismo para ambas.

En el test J, el p-valor asociado indica un rechazo de la hipótesis nula, es decir, existe evidencia significativa para asegurar que alguno de los componentes, ya sea, efectos aleatorios, correlación espacial o correlación serial, es significativamente distinto de cero. Debido a esto, es necesario proseguir con el análisis de los test condicionales, ya que, cada uno analiza la existencia de los parámetros

por separado, condicionándolos a la existencia de los otros 2.

El test C1 y C2, tienen asociados estadísticos cuyos p-valores son menores al nivel de significancia, es decir, se rechaza la hipótesis nula asociada a cada test. Esto indica para el test C1, que al considerar la existencia de los parámetros asociados a la correlación serial y a los efectos aleatorios, el parámetro de correlación espacial es significativamente distinto de cero. Es decir, aún se deben considerar los 3 parámetros en el modelo, ya que, la existencia de los parámetros condicionales no implica que se pueda omitir un parámetro asociado a la correlación espacial.

Por otro lado, en el test C2, se concluye que considerando la existencia de los parámetros de correlación espacial y efectos aleatorios, aún queda un efecto temporal significativo en los errores que debe ser considerado en el modelo, a través del parámetro de autocorrelación serial.

Por último, el test C3 tiene asociado p-valores mayores al nivel de significancia para ambas relaciones de variables, es decir, en ambos casos no se rechaza la hipótesis nula asociada. En otras palabras, existe evidencia para asumir que dado que existen parámetros asociados a la correlación serial y espacial, el parámetro asociado a los efectos aleatorios no es significativo para considerarlo en el modelo. Este test, permite considerar un modelo más simple y más parsimonioso que el general.

Para este caso, en el cual se tuvo en consideración el periodo total de los datos, el modelo más indicado sería implementar un SEMSR, que justamente contempla los dos parámetros de correlación, el espacial y el serial. En caso de que los residuos de este modelo no lograrán cumplir los supuestos del modelo, se puede considerar agregar el parámetro de retardo espacial, implementando el modelo SAREMSR.

Los resultados de las Pruebas LM para cada periodo acotado definido previamente, son muy similares a los observados en la tabla de la figura 7.1, por lo que, para cada periodo se llega a la misma conclusión con respecto a los modelos. Si se desean ver las tablas asociadas, ir al Apéndice E.

## 7.2. Parámetros Estimados de cada Modelo

En la sección anterior se determinó que lo más óptimo es implementar un modelo SEMSR, el cual considera en el término del error un parámetro para controlar la correlación espacial y otro para la correlación serial. Además, a modo de comparación, se implementan también los modelos

particulares del SEMSR, es decir, los modelos SEM y SR. El SEM sólo considera un parámetro de correlación espacial y el SR sólo considera un parámetro de correlación serial en el término del error. Y adicionalmente, se implementan los modelos SAREMSR, para comparar los resultados al agregar un parámetro de retardo espacial, y OLS para ver que efectivamente se debe controlar el efecto espacial y temporal en el modelo.

Todos estos modelos se implementan en el software R, utilizando el comando **spreml** del paquete **splm** del software R.

Se destaca que las siguientes tablas de las figuras 7.2, 7.3, 7.4, 7.5, están asociadas a los parámetros estimados de cada modelo, por período. En ellas los valores de los parámetros que estén destacados en rojo, indican que no son significativos. Además, para ver los p-valores asociados a cada estimación, se puede revisar el apéndice F.

En estas tablas, la columna de **rho**,  $\rho$ , corresponde a las estimaciones del coeficiente espacial autorregresivo, donde  $|\rho| < 1$ , la de **psi**,  $\psi$ , se asocia a estimaciones del coeficiente de correlación serial, donde  $|\psi| < 1$  y **lambda**,  $\lambda$ , se asocia al coeficiente de rezago espacial. Mientras más cercanos a 1 sean los valores de estos parámetros, más contribuyen para explicar la variable, disminuyendo el ruido en el error final del modelo. Además, **b0** indica los valores del intercepto y **b1** es el parámetro asociado a la variable de movilidad.

Parámetros estimados de cada modelo, Periodo Total

modelo	b0	b1	lambda	psi	rho	variables
ols	102.167	-112.440	NA	NA	NA	TasaContagios ~ V.M.Completa
sem	105.291	-10.569	NA	NA	0.908	TasaContagios ~ V.M.Completa
sr	109.659	-13.622	NA	0.924	NA	TasaContagios ~ V.M.Completa
semsr	106.584	-9.418	NA	0.781	0.770	TasaContagios ~ V.M.Completa
saremsr	199.543	-0.378	-0.897	0.780	0.944	TasaContagios ~ V.M.Completa
ols	69.775	-207.310	NA	NA	NA	TasaContagios ~ V.M.Laboral
sem	137.872	31.671	NA	NA	0.912	TasaContagios ~ V.M.Laboral
sr	97.783	-112.867	NA	0.925	NA	TasaContagios ~ V.M.Laboral
semsr	123.725	-4.198	NA	0.792	0.773	TasaContagios ~ V.M.Laboral
saremsr	7.412	-11.763	0.918	0.777	-0.674	TasaContagios ~ V.M.Laboral

Figura 7.2: Parámetros Estimados, considerando el Periodo Total

Parámetros estimados de cada modelo, Primera Ola

modelo	b0	b1	lambda	psi	rho	variables
ols	121.842	-172.875	NA	NA	NA	TasaContagios ~ V.M.Completa
sem	170.227	42.952	NA	NA	0.916	TasaContagios ~ V.M.Completa
sr	88.529	-103.710	NA	0.899	NA	TasaContagios ~ V.M.Completa
semsr	124.892	-37.354	NA	0.788	0.810	TasaContagios ~ V.M.Completa
saremsr	233.336	-24.027	-0.825	0.799	0.947	TasaContagios ~ V.M.Completa
ols	93.666	-213.942	NA	NA	NA	TasaContagios ~ V.M.Laboral
sem	201.537	95.453	NA	NA	0.922	TasaContagios ~ V.M.Laboral
sr	75.702	-123.499	NA	0.899	NA	TasaContagios ~ V.M.Laboral
semsr	135.518	-1.747	NA	0.791	0.830	TasaContagios ~ V.M.Laboral
saremsr	5.867	-9.002	0.937	0.798	-0.670	TasaContagios ~ V.M.Laboral

Figura 7.3: Parámetros Estimados, considerando la Primera Ola

Parámetros estimados de cada modelo, Periodo entre Olas

modelo	b0	b1	lambda	psi	rho	variables
ols	56.153	-17.746	NA	NA	NA	TasaContagios ~ V.M.Completa
sem	56.460	-16.532	NA	NA	0.535	TasaContagios ~ V.M.Completa
sr	59.339	-7.325	NA	0.749	NA	TasaContagios ~ V.M.Completa
semsr	58.674	-6.521	NA	0.703	0.280	TasaContagios ~ V.M.Completa
saremsr	8.232	-1.314	0.859	0.648	-0.985	TasaContagios ~ V.M.Completa
ols	49.331	-27.017	NA	NA	NA	TasaContagios ~ V.M.Laboral
sem	53.926	-11.441	NA	NA	0.619	TasaContagios ~ V.M.Laboral
sr	55.965	-16.381	NA	0.767	NA	TasaContagios ~ V.M.Laboral
semsr	56.301	-12.255	NA	0.709	0.338	TasaContagios ~ V.M.Laboral
saremsr	11.998	-10.766	0.756	0.678	-0.704	TasaContagios ~ V.M.Laboral

Figura 7.4: Parámetros Estimados, considerando el Periodo entre Olas

Parámetros estimados de cada modelo, Segunda Ola

modelo	b0	b1	lambda	psi	rho	variables
ols	207.306	-61.721	NA	NA	NA	TasaContagios ~ V.M.Laboral
sem	233.606	28.569	NA	NA	0.792	TasaContagios ~ V.M.Laboral
sr	154.642	-176.032	NA	0.845	NA	TasaContagios ~ V.M.Laboral
semsr	205.812	-3.012	NA	0.724	0.691	TasaContagios ~ V.M.Laboral
saremsr	25.469	-16.317	0.859	0.703	-0.525	TasaContagios ~ V.M.Laboral

Figura 7.5: Parámetros Estimados, considerando la Segunda Ola

El parámetro asociado a la variable de movilidad,  $b_0$ , en ningún modelo SEMRE es significativo. Agregando un parámetro de retardo espacial al modelo, el parámetro  $b_0$ , sigue siendo no significativo, pero si quitamos algún parámetro espacial o temporal, la variable de movilidad en algunos casos si es significativa. Por otro lado, los parámetros espaciales y temporales son significativos en todos los modelos, es decir, efectivamente hay un efecto espacial y otro serial que se deben considerar, y estos parámetros la modelan. El parámetro de retardo espacial también es significativo en todos los modelos, aún en los modelos donde ya se consideran los parámetros espacial y serial, por lo que, puede estar capturando la dependencia espacial de la variable tasa de contagios que no está logrando explicar la variable de movilidad o que no está siendo explicada por ausencia de otras variables significativas que no están siendo consideradas en el modelo. Además, el análisis por período indica que en la primera ola, hay más modelos que consideran significativa la variable de movilidad y el periodo entre olas parece ser el que menos considera a la variable explicativa como significativa. Por último, con respecto a las 2 variables de movilidad consideradas en las relaciones entre variables, no se logra distinguir cuál de las 2 explica mejor a la tasa de contagios.

Para saber qué modelo logró capturar en su mayoría la correlación espacial y serial presentes en las variables, se deben analizar los supuestos de los residuos y comparar con medidas de bondad de ajuste. Esto no se puede realizar por limitancias con el software R en el cálculo de los residuos del modelo. El inconveniente es que, el software, luego de realizar las estimaciones correspondientes de todos los parámetros, arroja los residuos estimados del modelo calculados como  $\hat{u} = y - X\hat{\beta}$ . Es decir, estos residuos siguen contemplando los valores asociados a la correlación espacial y serial, pues no se están restando los efectos espaciales y seriales que los parámetros lograron capturar. El cálculo tampoco se pudo realizar a mano, por lo que, no se continúa con el análisis.

# Capítulo 8

## Conclusión

El objetivo de esta investigación era analizar si la tasa de contagios de Covid-19 poseía relación con el tiempo y el espacio, para posteriormente poder modelar la tasa en base a la movilidad.

El estudio previamente observado y analizado, permite concluir que la tasa de contagios efectivamente tiene asociación con las semanas y las comunas consideradas, es decir, depende del tiempo y el espacio. Para observar esto de mejor forma, se dividió el periodo total de semanas considerando el comportamiento de la tasa de contagios. Este paso fue relevante, ya que, se marcaron claras diferencias entre los periodos durante parte del análisis. Al considerar las semanas de la primera ola, la dependencia tanto espacial como temporal de la tasa de contagios era muy significativa. Para las semanas entre olas, la dependencia espacial dejaba de ser significativa, es decir, en estas semanas no había o había muy poca relación entre comuna y tasa de contagios. La segunda ola, al igual que la primera, indicaba una fuerte asociación entre espacio y tiempo. Con este análisis se responde también a la primera pregunta planteada en la introducción, que hacía alusión a si el espacio y tiempo aportaban significativamente en la evolución de la tasa de contagios. Por otro lado, las variables de movilidad también poseían fuerte asociación con las dimensiones tiempo y espacio, pues la autocorrelación espacial y temporal estaba presente significativamente en las variaciones de movilidad, aún cuando se analizaban por periodo.

Considerando el análisis de los mapas de cada variable por semana, claramente se pudo ver que las hipótesis planteadas al iniciar este estudio, eran ciertas. Es decir, el aumento de los casos de contagios, sí inició en el sector oriente de la RM y los casos continuaron moviéndose a la zona sur-poniente de la región. Asimismo, durante las primeras semanas de la pandemia, la movilidad de las personas era mayor en el sector sur-poniente en comparación al sector oriente de la región.

Se verificó que existe correlación entre las variables tasa de contagios y movilidad, estudiada por semana. Esta relación no se mantiene y no es tan fuerte a lo largo de todo el periodo, pero se destaca que en las semanas del periodo de la ola 1, esta asociación es más fuerte en comparación a los demás períodos analizados. Con esto, se da respuesta a la segunda pregunta planteada en la introducción, que se relacionaba con si los contagios podían ser explicados a través de la movilidad de las personas. Cabe enfatizar que la variación de movilidad Laboral, en general parece ser la que mejor podría explicar a la tasa de contagios, ya que, presenta valores mayores en las correlaciones en comparación a la otra variable de movilidad. Uno de los puntos más destacados del análisis hecho sobre la relación de las variables, es que existe un claro efecto de endogeneidad, pues la variable de movilidad explica a la variable de contagios, pero la movilidad también puede ser explicada por la tasa de contagios. Se produce un efecto endógeno en la variable de movilidad, que puede deberse a que al aumentar los casos de contagios, el gobierno implementa medidas sanitarias en las comunas, como las cuarentenas y esta impacta en gran medida a la movilidad.

Para terminar, el último capítulo, estudia el modelado de los datos. Las pruebas LM, permitieron identificar que evidentemente se debían considerar parámetros que capturarán la correlación espacial y serial presentes en el término del error, es decir, se debía implementar un modelo SEMSR o un SAREMSR. Aún así, se estimaron varios modelos, contemplando desde el más simple, una regresión lineal simple, hasta el más general posible, que contempla los parámetros de correlación espacial, serial y retardo espacial, el SAREMSR. Esto, con la finalidad de comparar los supuestos y medidas de bondad de ajuste de cada modelo y ver si efectivamente los parámetros capturaban la correlación existente. De todos los modelos, se esperaba que el de mejor ajuste fuera un modelo SEMSR considerando los datos de la primera ola, porque captura tanto los efectos espaciales como temporales y en este periodo había mayor relación entre las variables, sobre todo con la variación de movilidad laboral. Pero, estimados los modelos, ningún SEMSR considera como significativa a la variable de movilidad. Finalmente, los residuos de los modelos no pudieron ser calculados, pero en algún posterior análisis donde se puedan calcular, se podrá analizar si efectivamente los residuos siguen o no correlacionados y concluir en base a esos resultados.

Las claras limitaciones de este estudio, apuntan a la escasez de variables medidas en espacio y tiempo para considerar en el modelo. Además la única variable considerada posee un efecto endógeno, que no está siendo considerado en el modelo, por lo que, puede producir inferencias erróneas.

# Capítulo 9

## Bibliografía

- Baltagi B., Song S., Jung B. y Koh W., 2007, “Testing for serial correlation, spatial autocorrelation and random effects using panel data panel”, Journal of Econometrics.
- Bravo L. y Ferres L., 2020, “Movilidad en Región Metropolitana de Santiago 28 de febrero al 29 de noviembre.”, Universidad del Desarrollo.
- Carracedo P., 2017, “Metodología espacio-temporal con datos de panel. Estudio de la mortalidad europea.”, Universidad Politécnica de Valencia.
- Carracedo P. y Debón A., 2017, “Selección de Modelos Espacio-Temporales con Datos de Panel en Matlab y R.”, Revista Electrónica de Comunicaciones y Trabajos de ASEPUMA, Volumen 18, páginas 93 a 118.
- Castro M., Foster W. y Ortega J., 2015, “ANÁLISIS DE CORRELACIÓN ESPACIAL EN EL PRECIO DE LA TIERRA DE USO AGRICOLA EN CHILE PERIODO 2000-2008”, AMECEDIR, México.
- Cuidate Plus, 2020, “Coronavirus”, Enfermedades, Infecciosas.
- Elhorst J., 2014, “Spatial econometrics: from cross-sectional data to spatial panels.”, Springer.
- Flint C., Harrower M. y Edsall R., 2000, “But How Does Place Matter? Using Bayesian Networks to Explore a Structural Definition of Place”, University of Colorado, Boulder.
- Giraldo R., 2016, “Introducción a la Geoestadística. Teoría y Aplicación”, Universidad Nacional de Colombia, Bogotá.

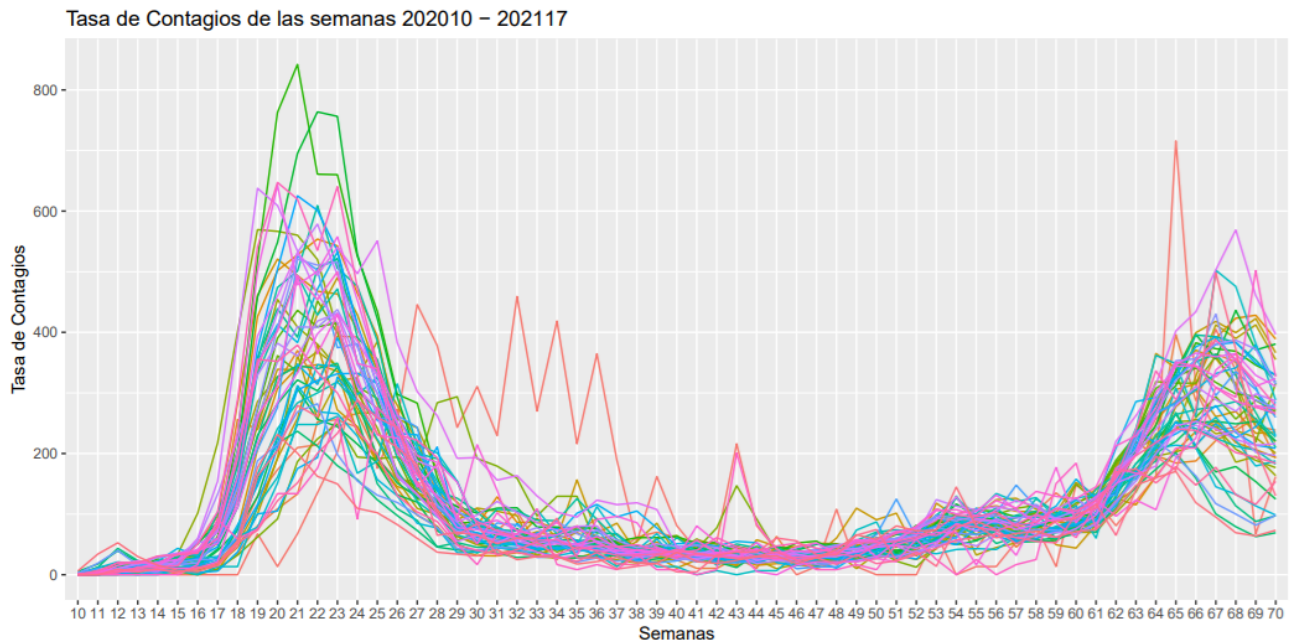
- Herrera M., 2017, “Fundamentals of Applied Spatial Econometrics”, National University of Salta.
- Huertado C., 2016, “Stationary, Homoscedastic and Cross-Sectional Spatial Models”, University of Illinois at Urbana-Champaign.
- Millo G., 2012, “Cross-sectional and spatial dependence in panels with R”, Assicurazioni Generali S.p.A, Innsbruck.
- Millo G., 2014, “Maximum likelihood estimation of spatially and serially correlated panel with random effects”, Assicurazioni Generali S.p.A., Trieste (Italy)
- Millo G., 2016, “Testing for Serial Correlation in Spatial Panels”.
- Millo G. y Piras G., 2009, “splm: econometric analysis of spatial panel data”, Universidad de Trieste, Rennes.
- Millo G. y Piras G., 2012, “splm: Spatial Panel data models in R”, Journal of Statistical Software
- Ministerio de Ciencia, Tecnología, Conocimiento e Innovación, 2021, “Base de Datos COVID-19”, <https://github.com/MinCiencia/Datos-COVID19/>
- Ministerio de Desarrollo Social y Familia, 2017, “Encuesta CASEN”, Información Social.
- Olivares M., Goic M., Weintraub G., Covarrubia J. y Escobedo C., 2020, “El dispar impacto de las cuarentenas en las comunas de la Región Metropolitana.”, ICSI.
- OMS, 2020, “Preguntas y respuestas sobre la enfermedad por coronavirus (COVID-19)”.
- Paelinck J., Mur J. y Trávez F., 2015, “Modelos para datos espaciales con estructura transversal o de pael. Una revisión.”, Asociación Internacional de Economía Aplicada, Valladolid, España.
- Polo M., 2017, “Descripción Metodológica del Modelo Espacial Autorregresivo en el Error”, Universidad Nacional Agraria La Molina, Facultad de Economía y Planificación.
- Romo J., 2019, “Modelos Espaciales y Espacio-Temporales para la modelación de datos con exceso de Ceros”, Centro de Investigación en Matemáticas.
- Vilalta C., 2005, “Cómo enseñar autocorrelación espacial”.
- Villalobos P., Castillo C., De la Fuente F. y Maddaleno M., 2020, “Incidencia y Mortalidad de COVID-19 en la Región Metropolitana de Chile: tiempo, espacio y factores estructurales”.

# Apéndices

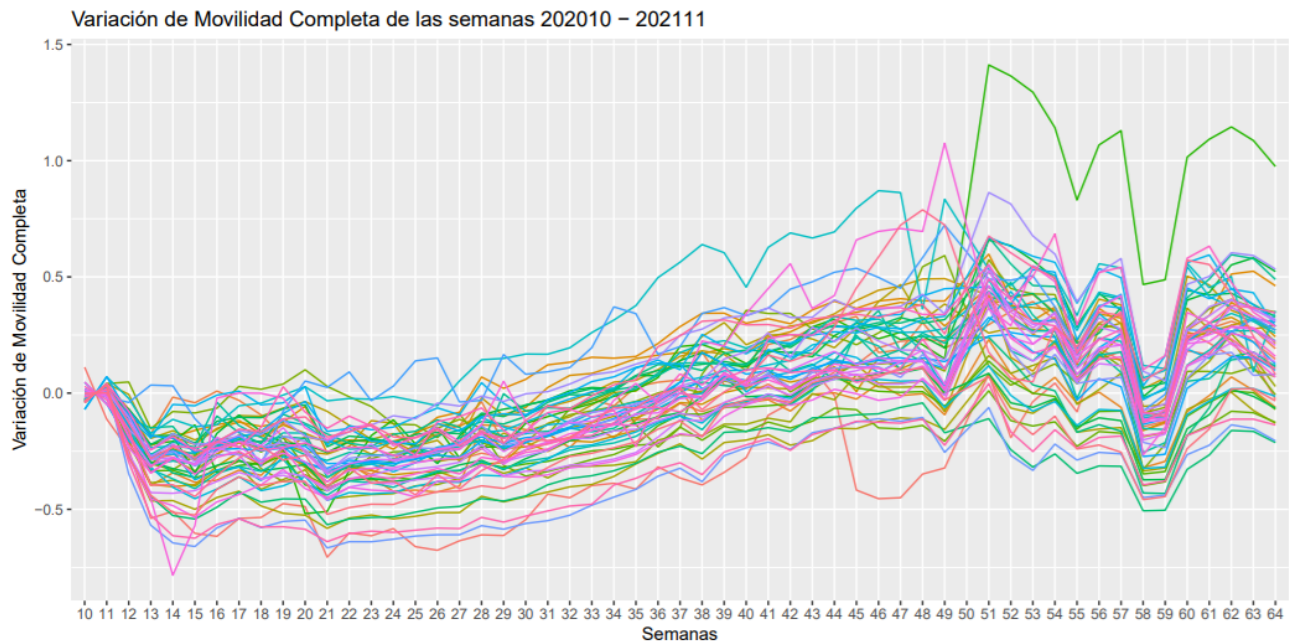
# Apéndice A

## Gráficos de Líneas de cada Variable

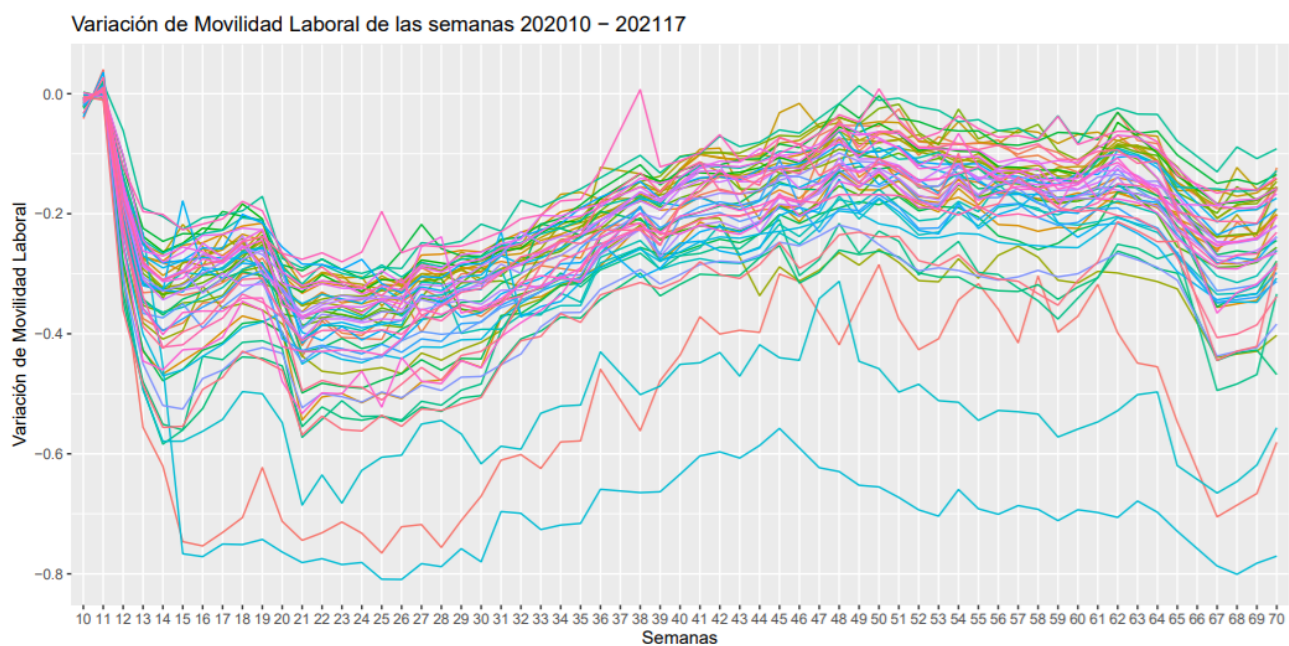
### A.1. Gráfico de Líneas, Tasa de Contagios



## A.2. Gráfico de Líneas, Variación de Movilidad Completa



## A.3. Gráfico de Líneas, Variación de Movilidad Laboral

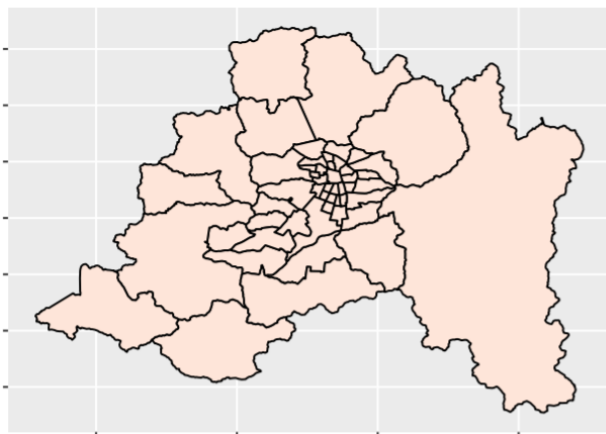


# Apéndice B

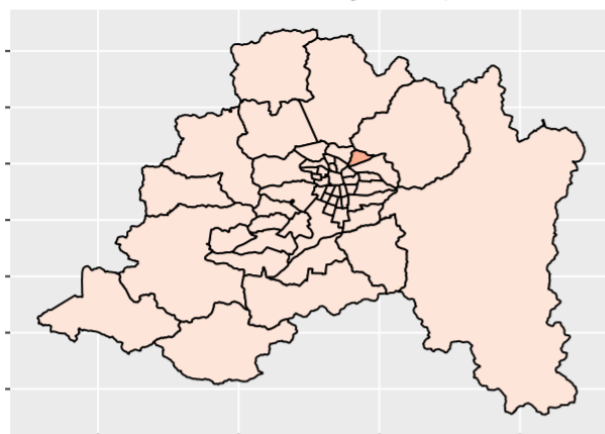
## Mapas de los Quintiles

### B.1. Mapas de los Quintiles de la Tasa de Contagios

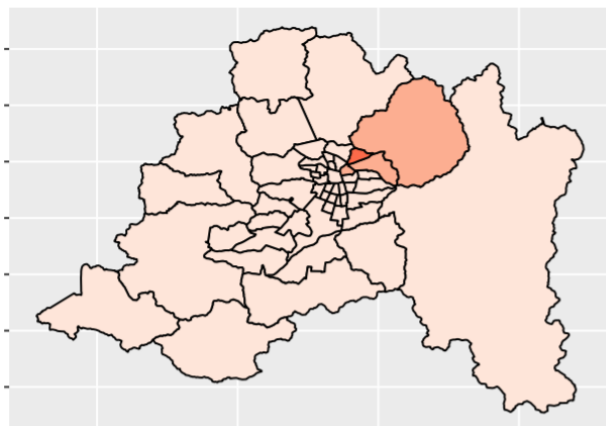
Tasa de Contagios por comuna de la semana 202010  
Considerando 202010 - 202111 , Región Metropolitana



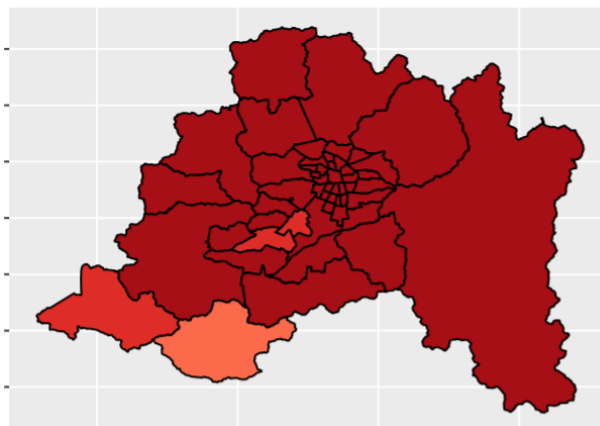
Tasa de Contagios por comuna de la semana 202011  
Considerando 202010 - 202111 , Región Metropolitana



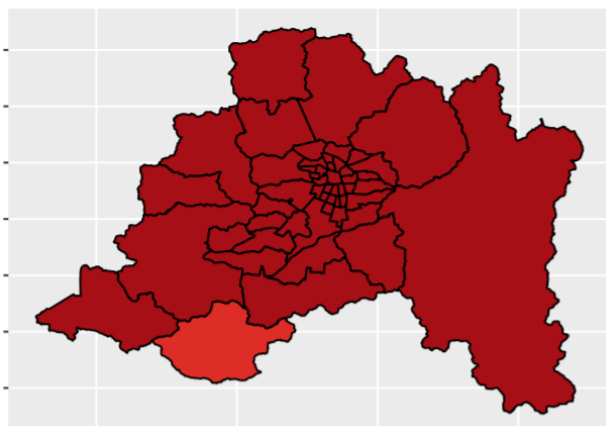
Tasa de Contagios por comuna de la semana 202012  
Considerando 202010 - 202111, Región Metropolitana



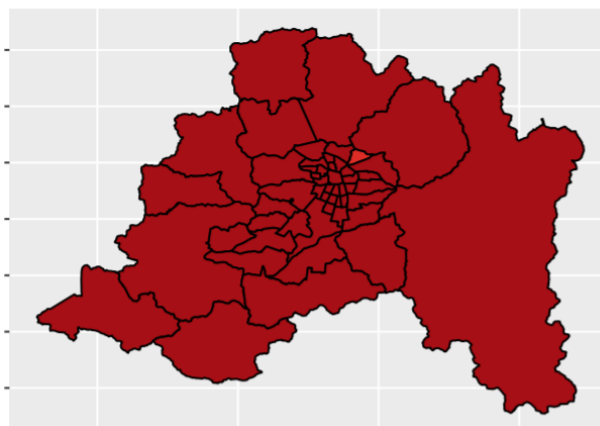
Tasa de Contagios por comuna de la semana 202021  
Considerando 202010 - 202111, Región Metropolitana



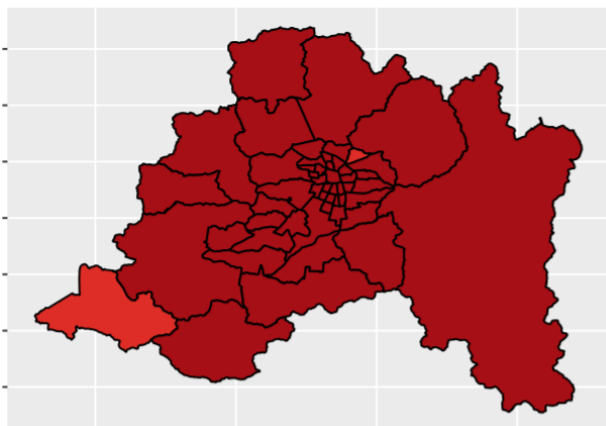
Tasa de Contagios por comuna de la semana 202022  
Considerando 202010 - 202111, Región Metropolitana



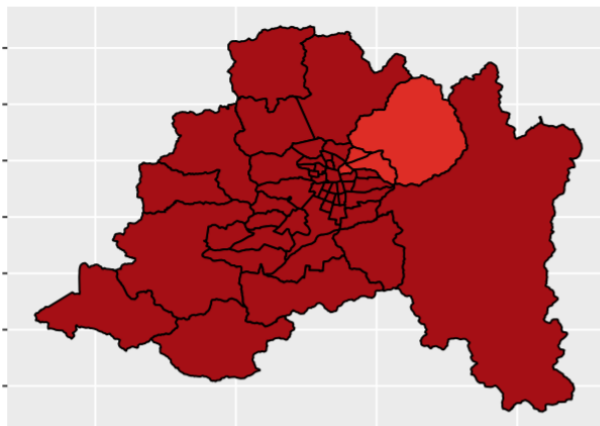
Tasa de Contagios por comuna de la semana 202023  
Considerando 202010 - 202111, Región Metropolitana



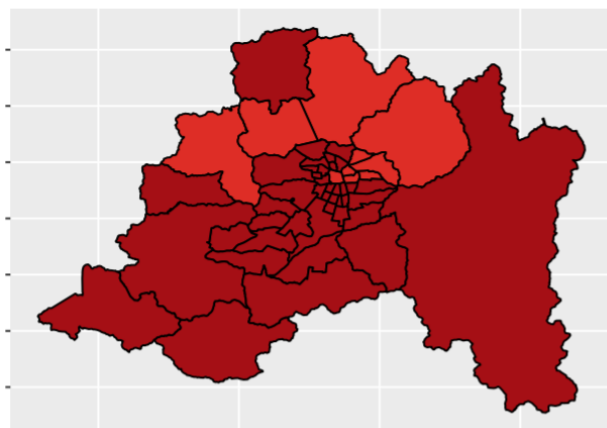
Tasa de Contagios por comuna de la semana 202024  
Considerando 202010 - 202111, Región Metropolitana



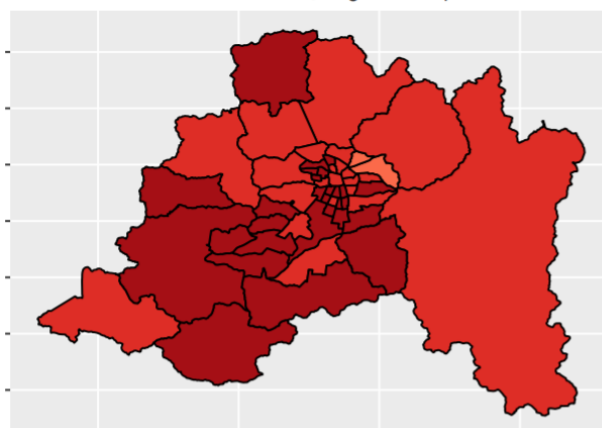
Tasa de Contagios por comuna de la semana 202025  
Considerando 202010 - 202111, Región Metropolitana



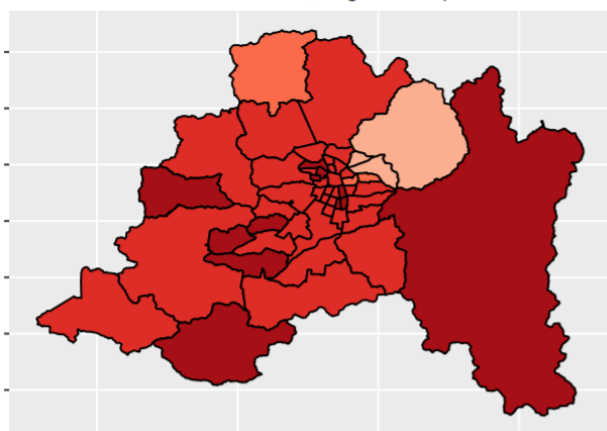
Tasa de Contagios por comuna de la semana 202026  
Considerando 202010 - 202111, Región Metropolitana



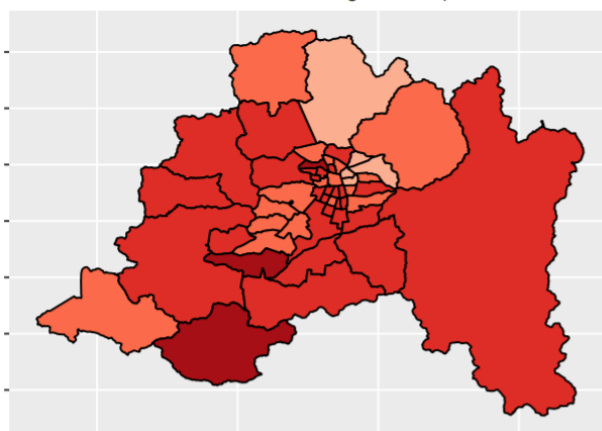
Tasa de Contagios por comuna de la semana 202027  
Considerando 202010 - 202111, Región Metropolitana



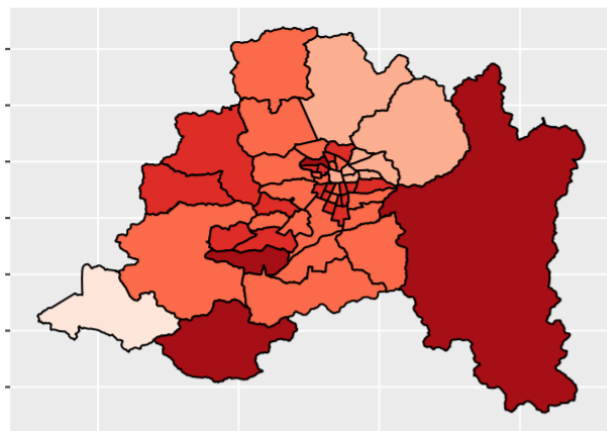
Tasa de Contagios por comuna de la semana 202028  
Considerando 202010 - 202111, Región Metropolitana



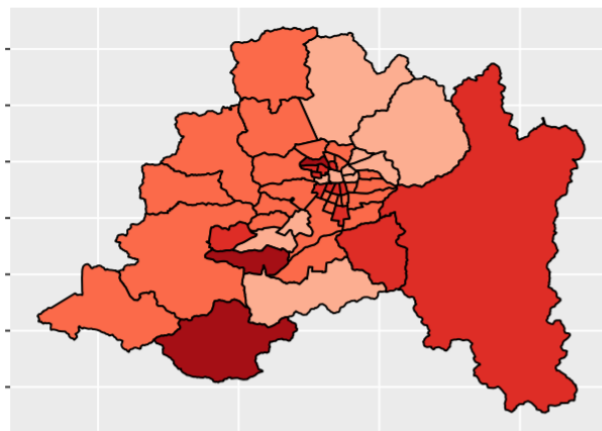
Tasa de Contagios por comuna de la semana 202029  
Considerando 202010 - 202111, Región Metropolitana



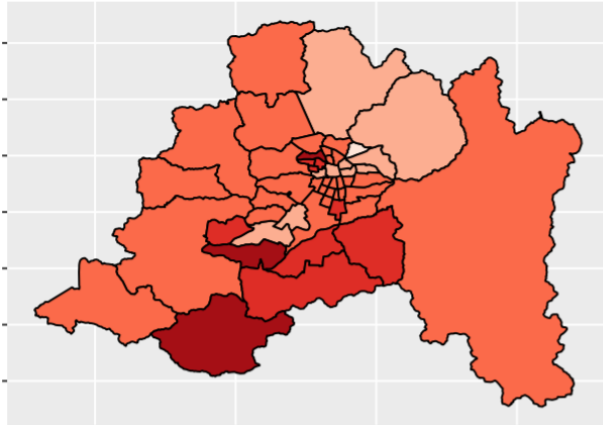
Tasa de Contagios por comuna de la semana 202030  
Considerando 202010 - 202111, Región Metropolitana



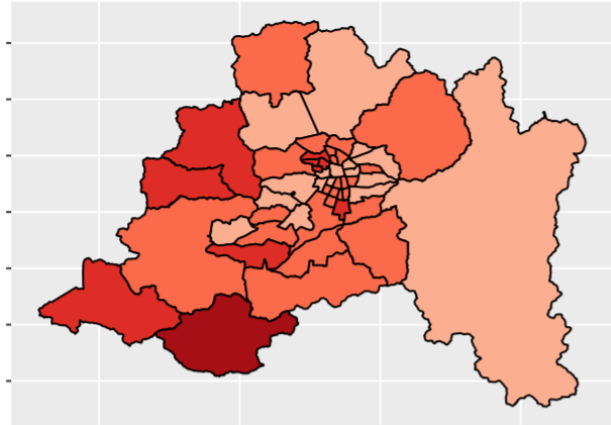
Tasa de Contagios por comuna de la semana 202031  
Considerando 202010 - 202111, Región Metropolitana



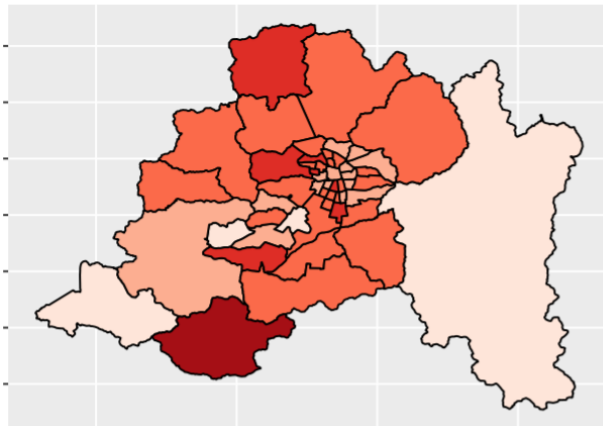
Tasa de Contagios por comuna de la semana 202032  
Considerando 202010 - 202111, Región Metropolitana



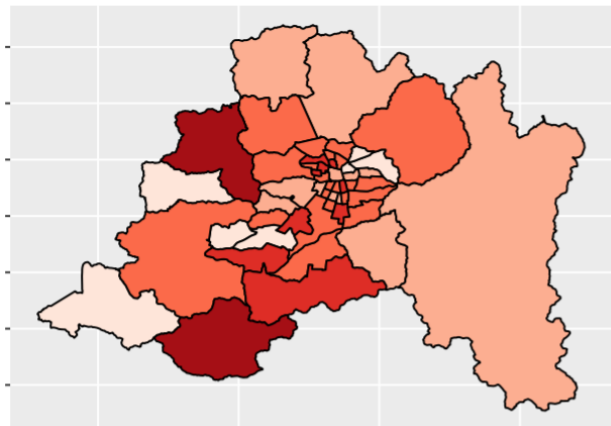
Tasa de Contagios por comuna de la semana 202033  
Considerando 202010 - 202111, Región Metropolitana



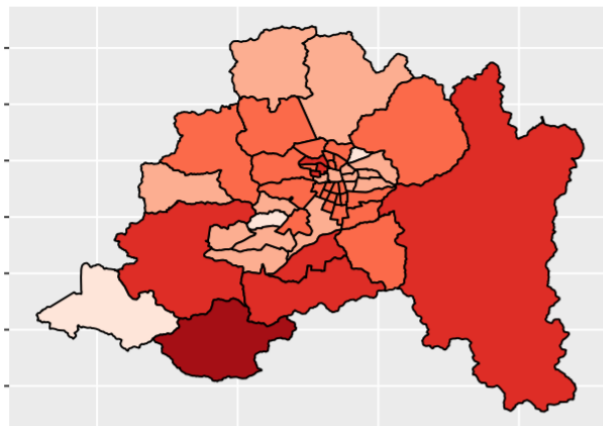
Tasa de Contagios por comuna de la semana 202034  
Considerando 202010 - 202111, Región Metropolitana



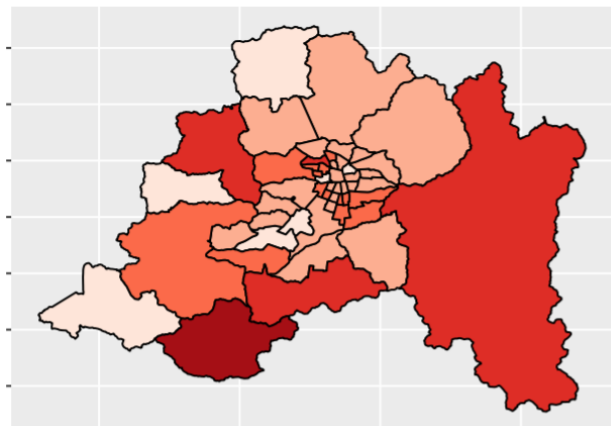
Tasa de Contagios por comuna de la semana 202035  
Considerando 202010 - 202111, Región Metropolitana



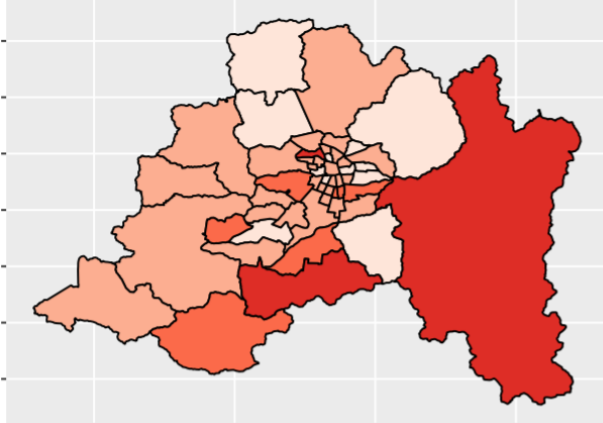
Tasa de Contagios por comuna de la semana 202036  
Considerando 202010 - 202111, Región Metropolitana



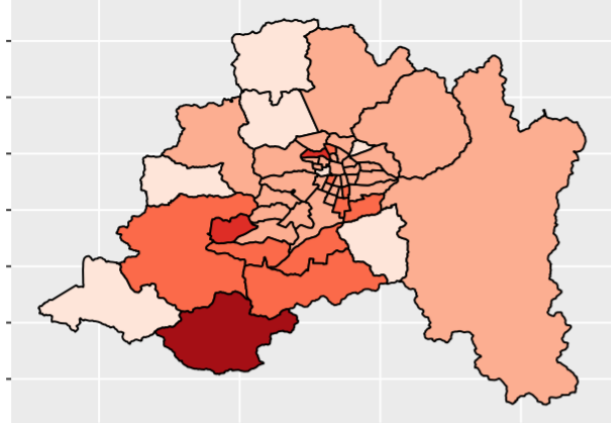
Tasa de Contagios por comuna de la semana 202037  
Considerando 202010 - 202111, Región Metropolitana



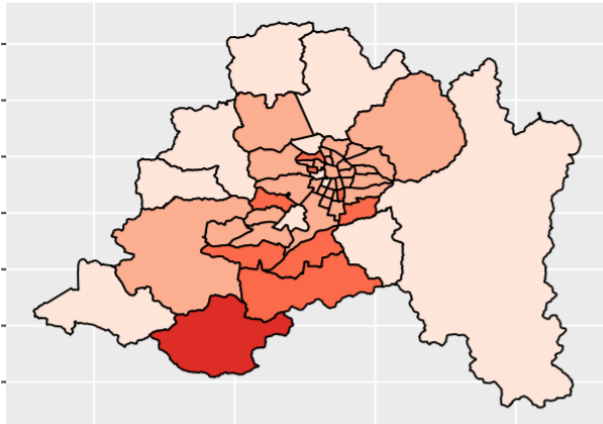
Tasa de Contagios por comuna de la semana 202038  
Considerando 202010 - 202111, Región Metropolitana



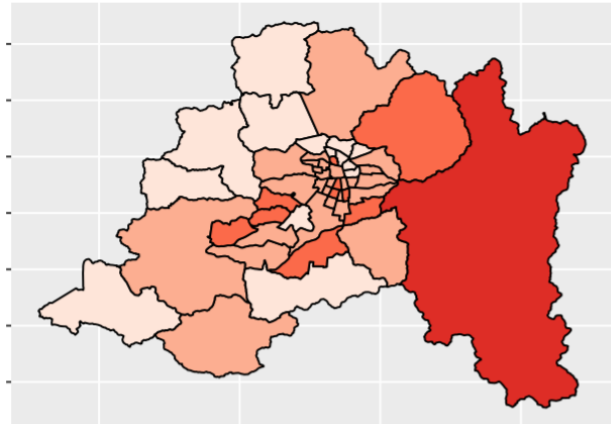
Tasa de Contagios por comuna de la semana 202039  
Considerando 202010 - 202111, Región Metropolitana



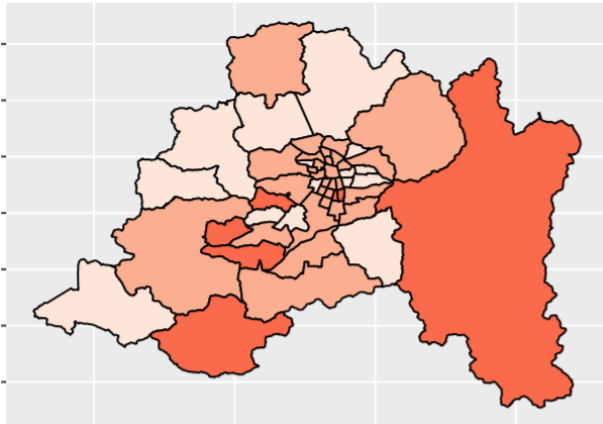
Tasa de Contagios por comuna de la semana 202040  
Considerando 202010 - 202111, Región Metropolitana



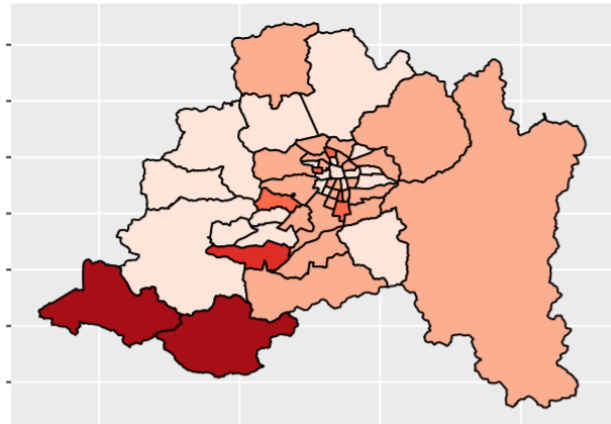
Tasa de Contagios por comuna de la semana 202041  
Considerando 202010 - 202111, Región Metropolitana



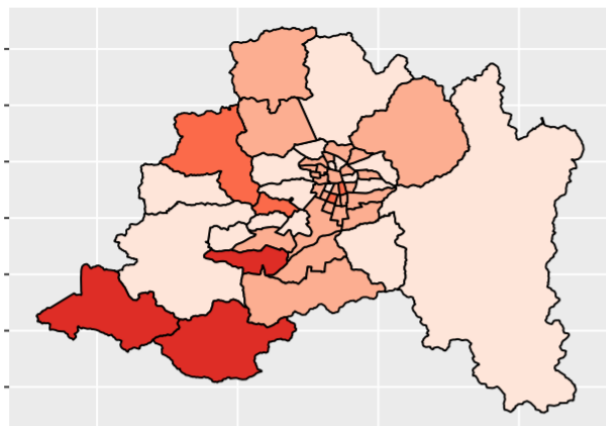
Tasa de Contagios por comuna de la semana 202042  
Considerando 202010 - 202111, Región Metropolitana



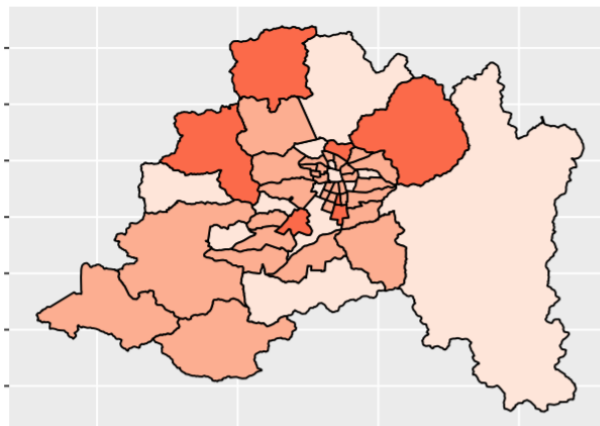
Tasa de Contagios por comuna de la semana 202043  
Considerando 202010 - 202111, Región Metropolitana



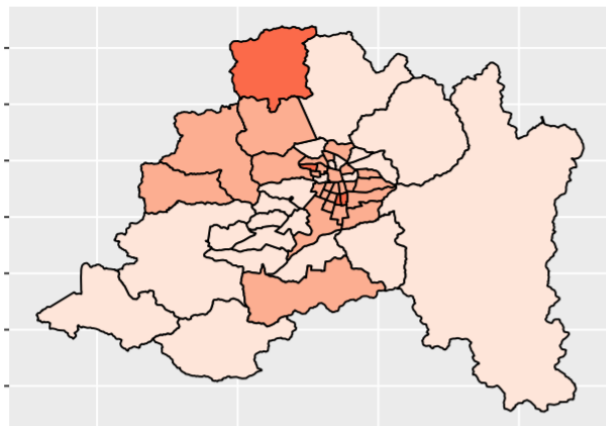
Tasa de Contagios por comuna de la semana 202044  
Considerando 202010 - 202111, Región Metropolitana



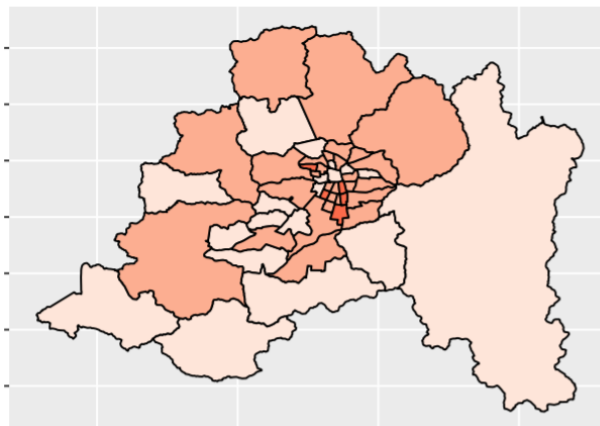
Tasa de Contagios por comuna de la semana 202045  
Considerando 202010 - 202111, Región Metropolitana



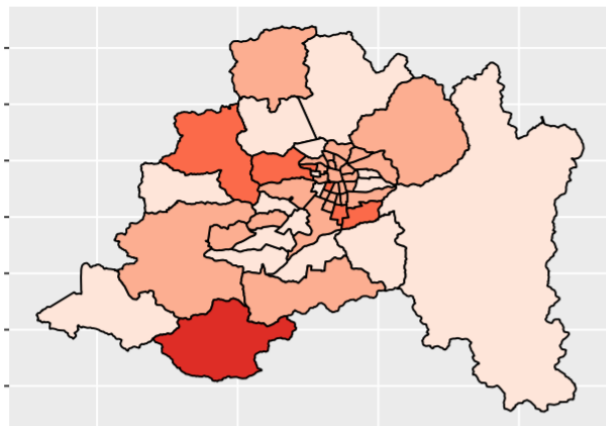
Tasa de Contagios por comuna de la semana 202046  
Considerando 202010 - 202111, Región Metropolitana



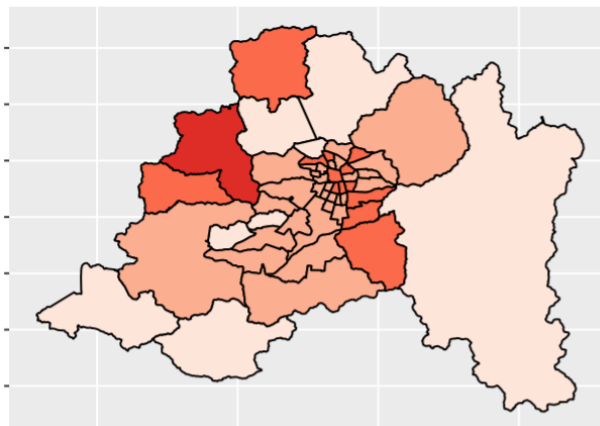
Tasa de Contagios por comuna de la semana 202047  
Considerando 202010 - 202111, Región Metropolitana



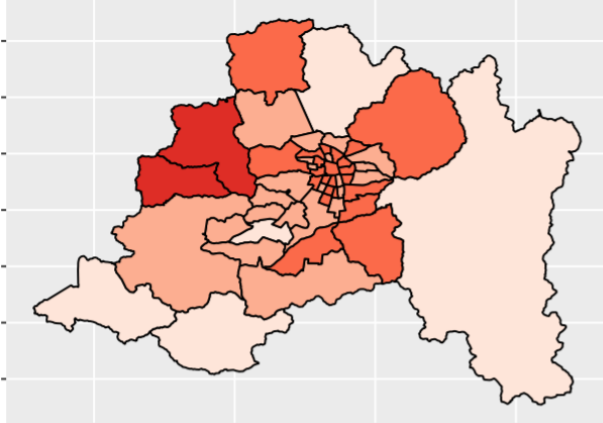
Tasa de Contagios por comuna de la semana 202048  
Considerando 202010 - 202111, Región Metropolitana



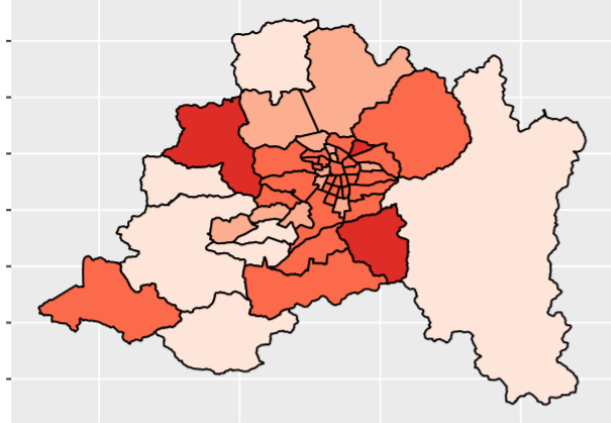
Tasa de Contagios por comuna de la semana 202049  
Considerando 202010 - 202111, Región Metropolitana



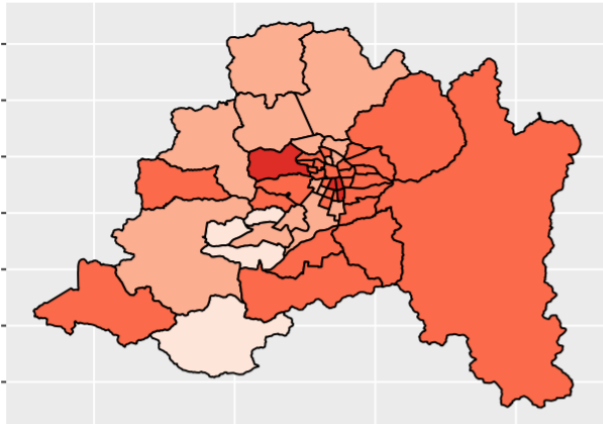
Tasa de Contagios por comuna de la semana 202050  
Considerando 202010 - 202111, Región Metropolitana



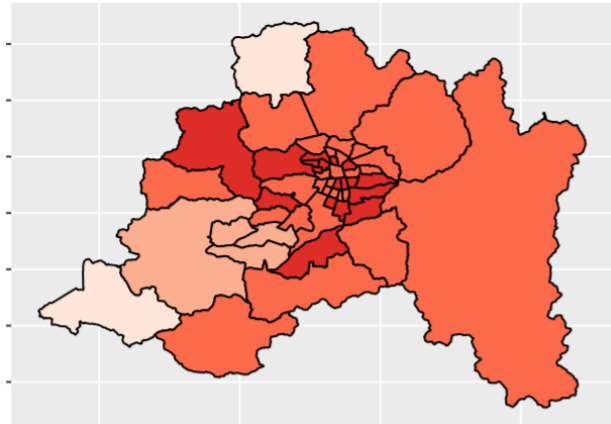
Tasa de Contagios por comuna de la semana 202051  
Considerando 202010 - 202111, Región Metropolitana



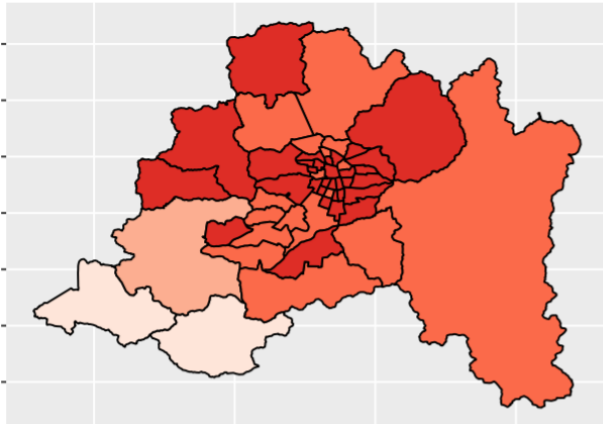
Tasa de Contagios por comuna de la semana 202052  
Considerando 202010 - 202111, Región Metropolitana



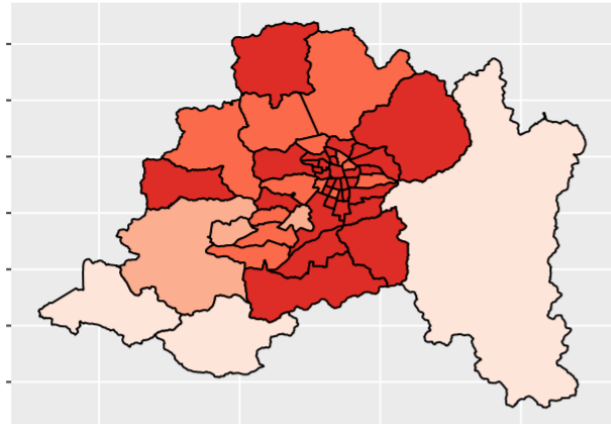
Tasa de Contagios por comuna de la semana 202053  
Considerando 202010 - 202111, Región Metropolitana



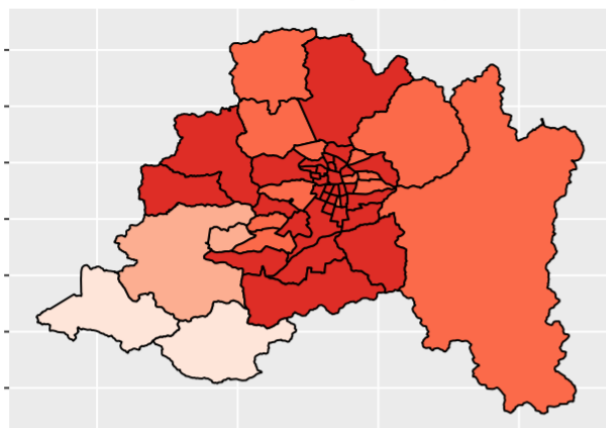
Tasa de Contagios por comuna de la semana 202101  
Considerando 202010 - 202111, Región Metropolitana



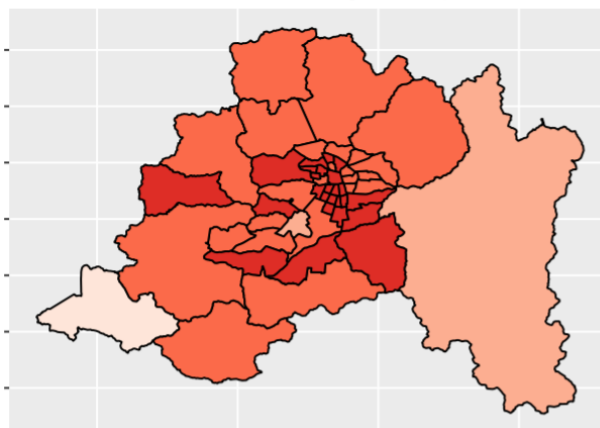
Tasa de Contagios por comuna de la semana 202102  
Considerando 202010 - 202111, Región Metropolitana



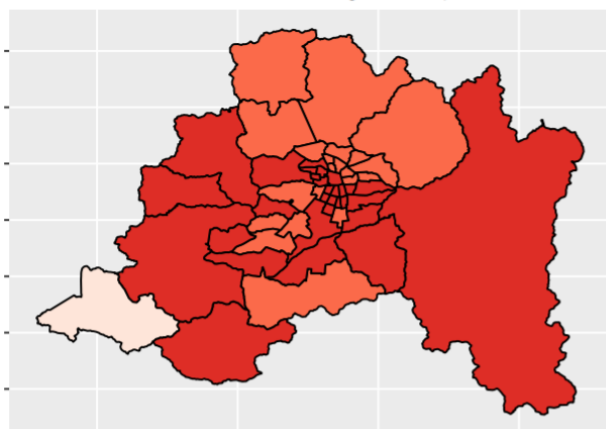
Tasa de Contagios por comuna de la semana 202103  
Considerando 202010 - 202111, Región Metropolitana



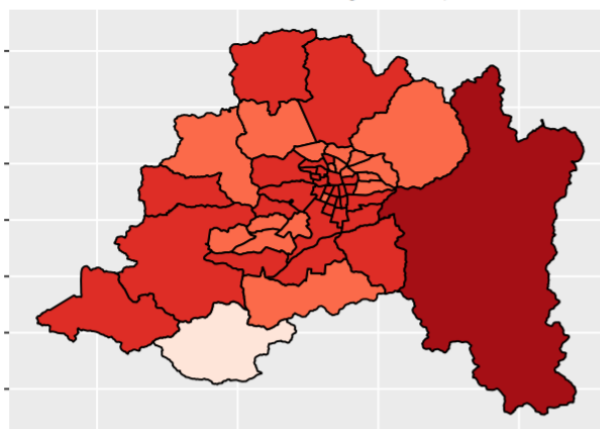
Tasa de Contagios por comuna de la semana 202104  
Considerando 202010 - 202111, Región Metropolitana



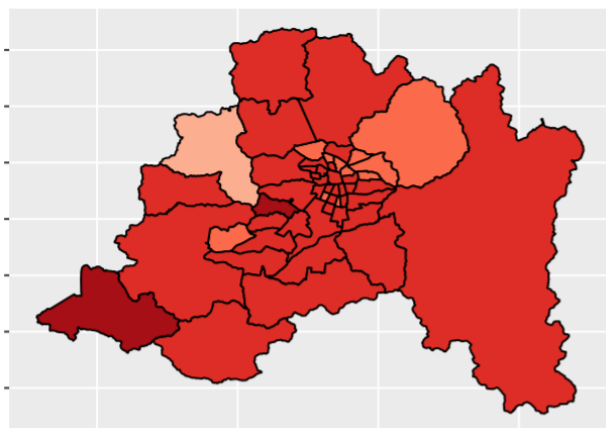
Tasa de Contagios por comuna de la semana 202105  
Considerando 202010 - 202111, Región Metropolitana



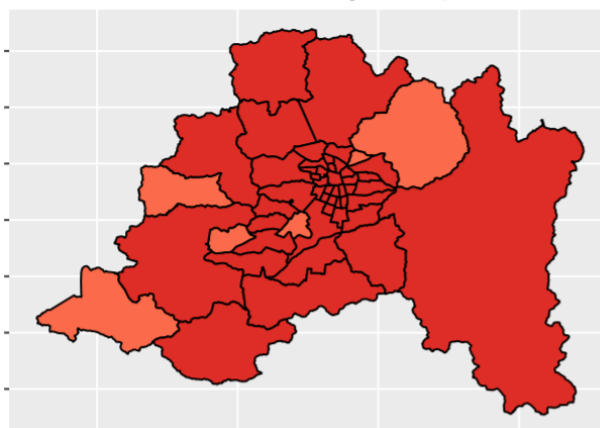
Tasa de Contagios por comuna de la semana 202106  
Considerando 202010 - 202111, Región Metropolitana



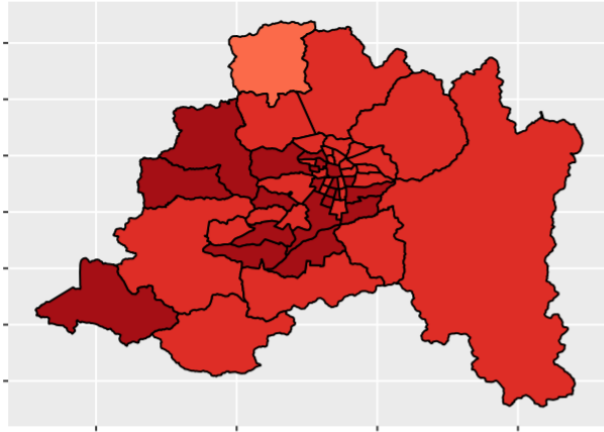
Tasa de Contagios por comuna de la semana 202107  
Considerando 202010 - 202111, Región Metropolitana



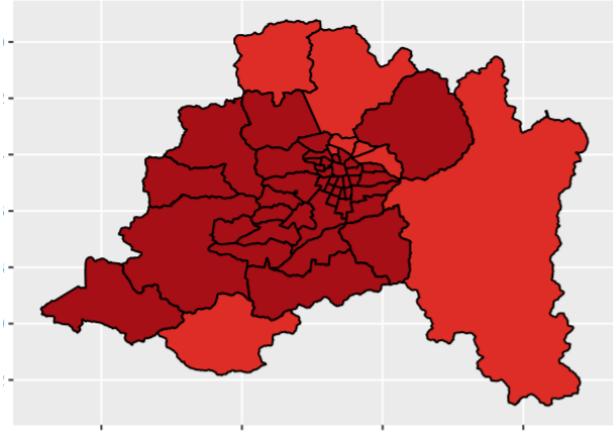
Tasa de Contagios por comuna de la semana 202108  
Considerando 202010 - 202111, Región Metropolitana



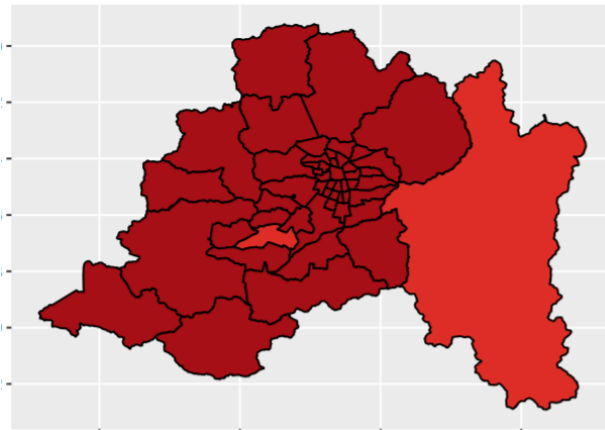
Tasa de Contagios por comuna de la semana 202109  
Considerando 202010 - 202111, Región Metropolitana



Tasa de Contagios por comuna de la semana 202110  
Considerando 202010 - 202111, Región Metropolitana

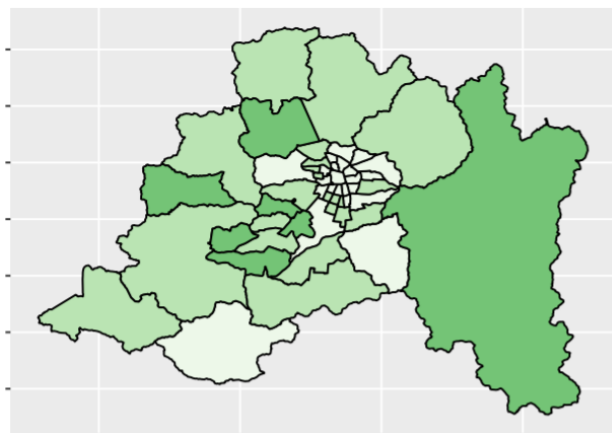


Tasa de Contagios por comuna de la semana 202111  
Considerando 202010 - 202111, Región Metropolitana

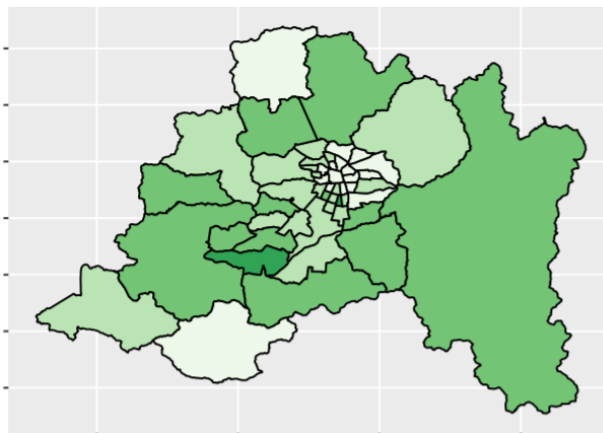


## B.2. Mapas de los Quintiles de la Variación de Movilidad Completa

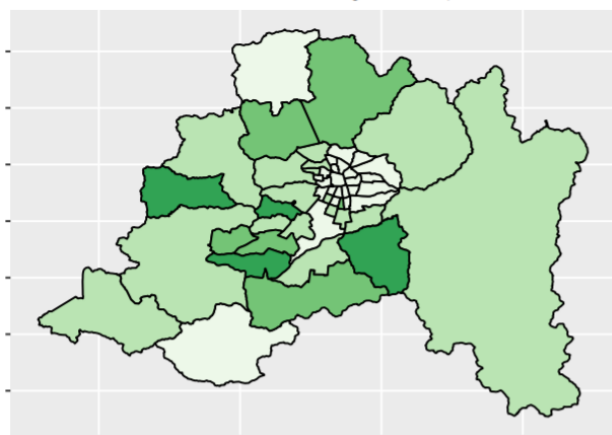
Var. Mov. Completa por comuna de la semana 202018  
Considerando 202010 - 202111, Región Metropolitana



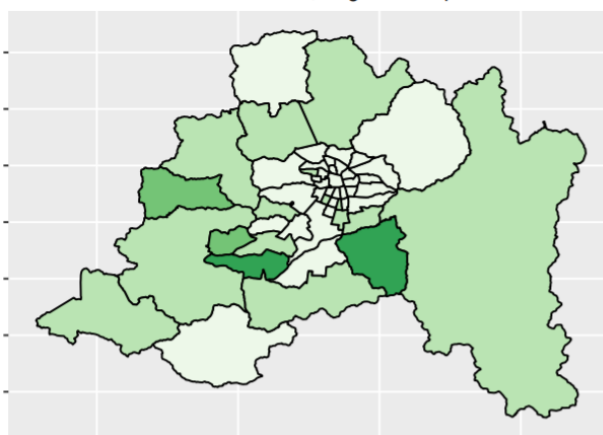
Var. Mov. Completa por comuna de la semana 202019  
Considerando 202010 - 202111, Región Metropolitana



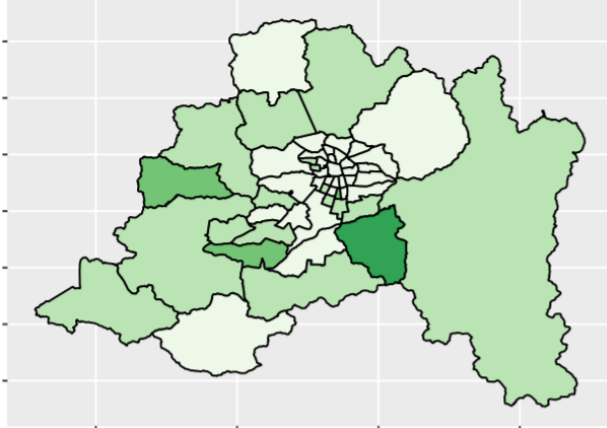
Var. Mov. Completa por comuna de la semana 202020  
Considerando 202010 - 202111, Región Metropolitana



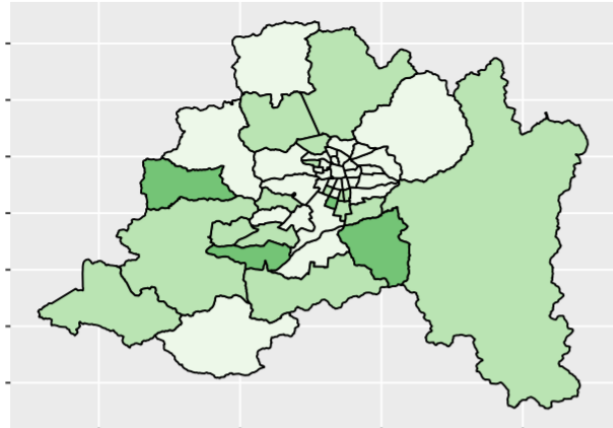
Var. Mov. Completa por comuna de la semana 202021  
Considerando 202010 - 202111, Región Metropolitana



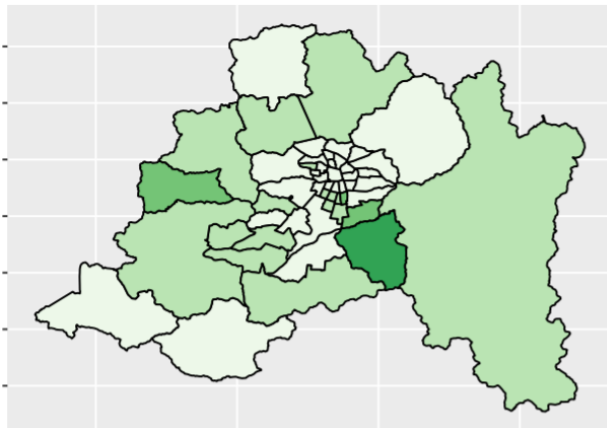
Var. Mov. Completa por comuna de la semana 202022  
Considerando 202010 - 202111, Región Metropolitana



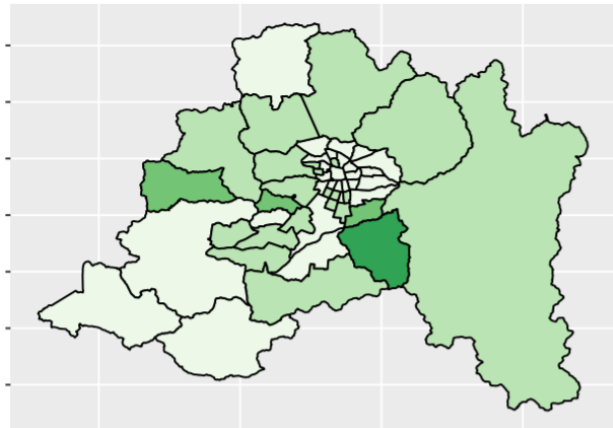
Var. Mov. Completa por comuna de la semana 202023  
Considerando 202010 - 202111, Región Metropolitana



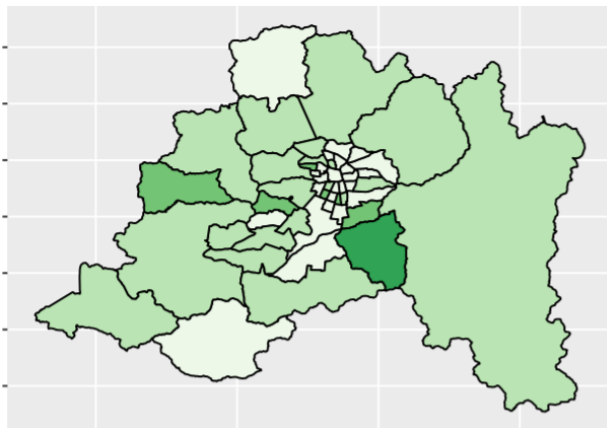
Var. Mov. Completa por comuna de la semana 202024  
Considerando 202010 - 202111, Región Metropolitana



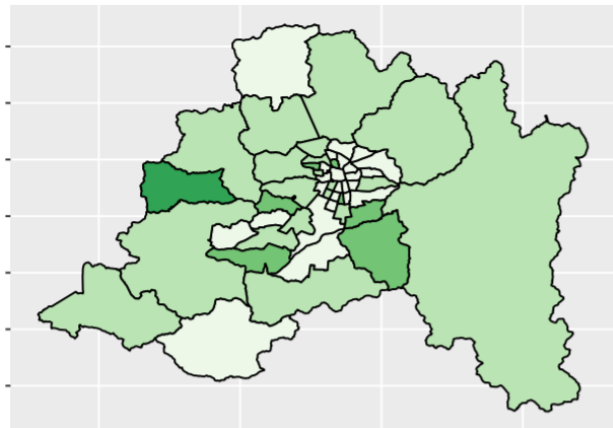
Var. Mov. Completa por comuna de la semana 202025  
Considerando 202010 - 202111, Región Metropolitana



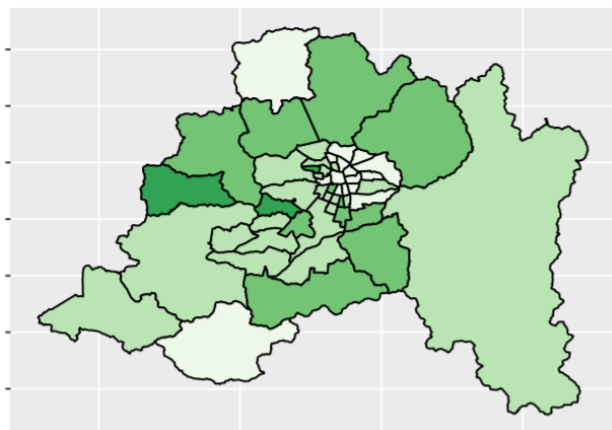
Var. Mov. Completa por comuna de la semana 202026  
Considerando 202010 - 202111, Región Metropolitana



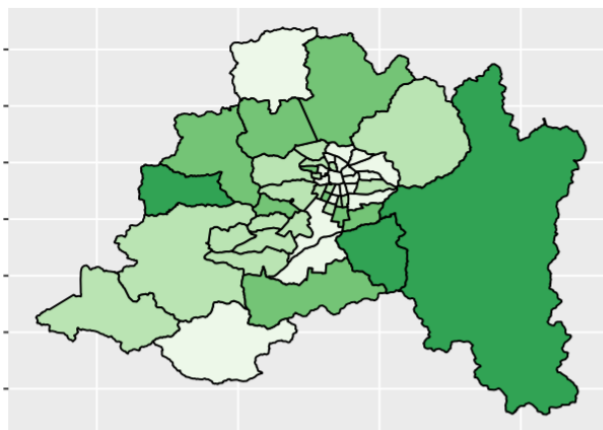
Var. Mov. Completa por comuna de la semana 202027  
Considerando 202010 - 202111, Región Metropolitana



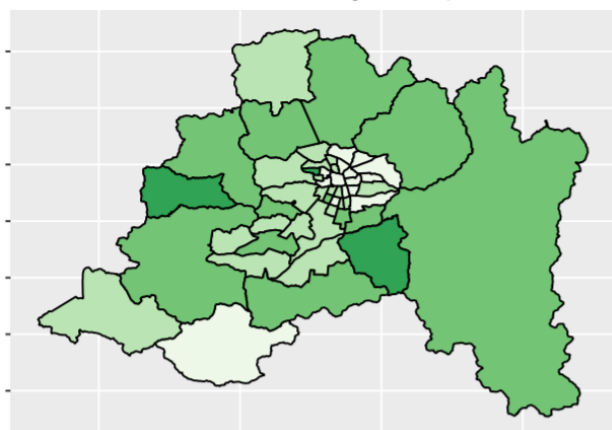
Var. Mov. Completa por comuna de la semana 202028  
Considerando 202010 – 202111 , Región Metropolitana



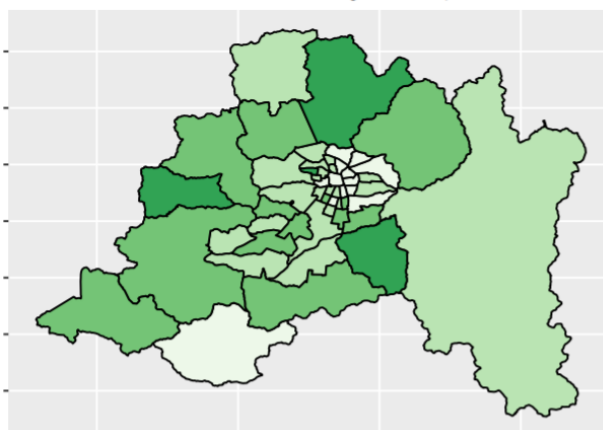
Var. Mov. Completa por comuna de la semana 202029  
Considerando 202010 – 202111 , Región Metropolitana



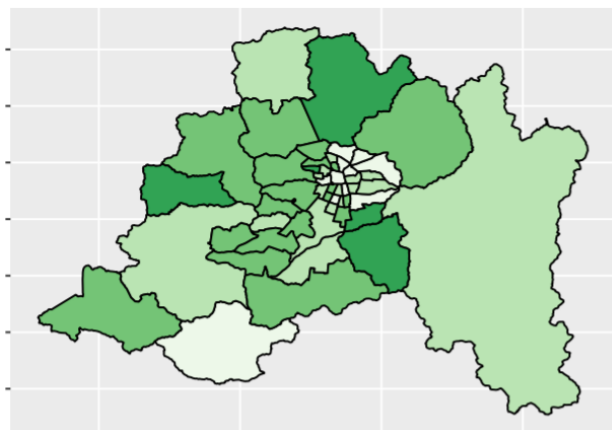
Var. Mov. Completa por comuna de la semana 202030  
Considerando 202010 – 202111 , Región Metropolitana



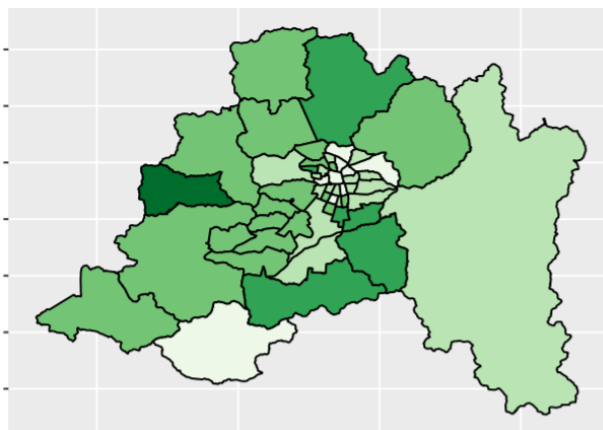
Var. Mov. Completa por comuna de la semana 202031  
Considerando 202010 – 202111 , Región Metropolitana



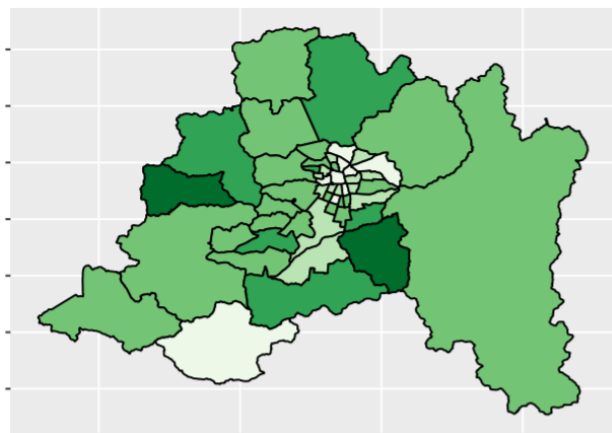
Var. Mov. Completa por comuna de la semana 202032  
Considerando 202010 – 202111 , Región Metropolitana



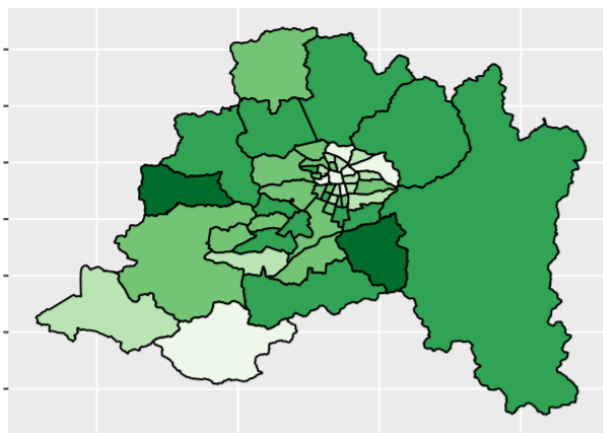
Var. Mov. Completa por comuna de la semana 202033  
Considerando 202010 – 202111 , Región Metropolitana



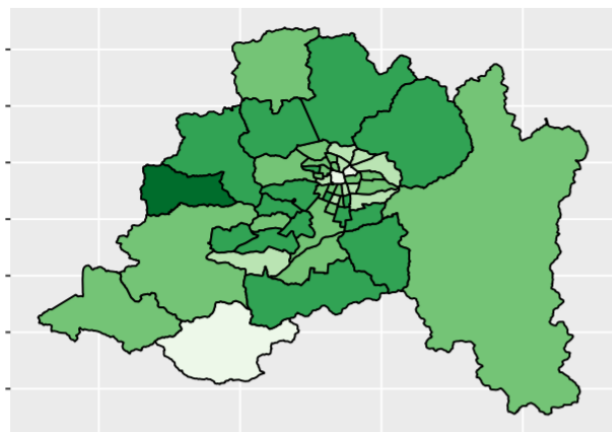
Var. Mov. Completa por comuna de la semana 202034  
Considerando 202010 - 202111, Región Metropolitana



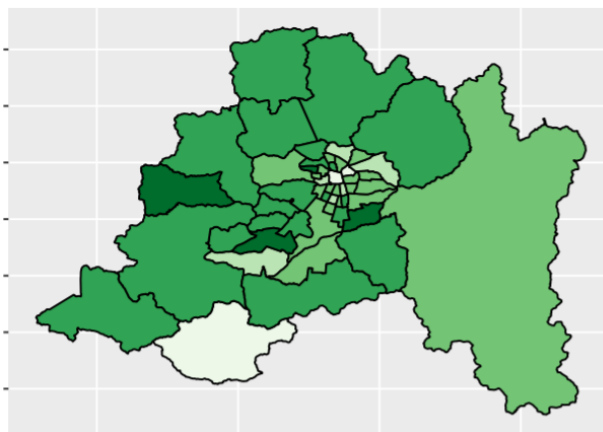
Var. Mov. Completa por comuna de la semana 202035  
Considerando 202010 - 202111, Región Metropolitana



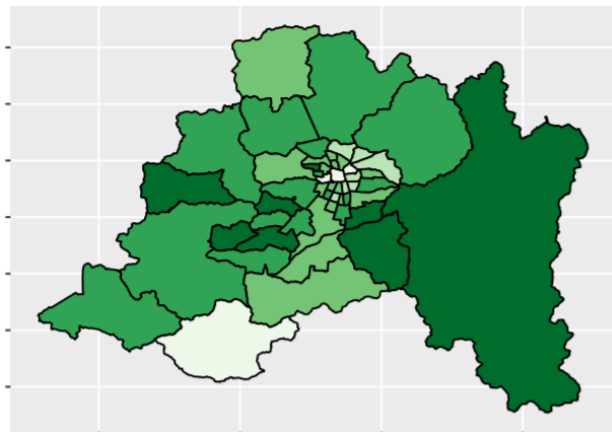
Var. Mov. Completa por comuna de la semana 202036  
Considerando 202010 - 202111, Región Metropolitana



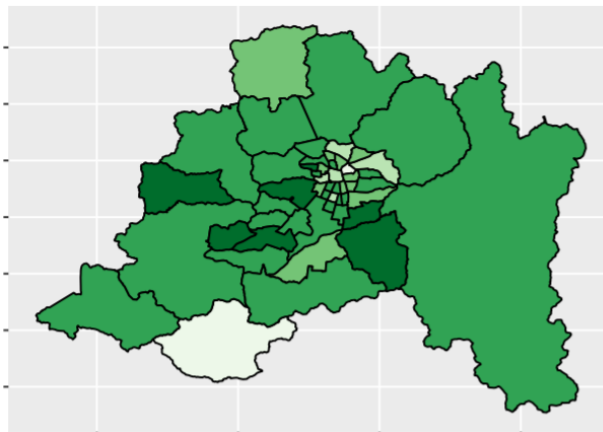
Var. Mov. Completa por comuna de la semana 202037  
Considerando 202010 - 202111, Región Metropolitana



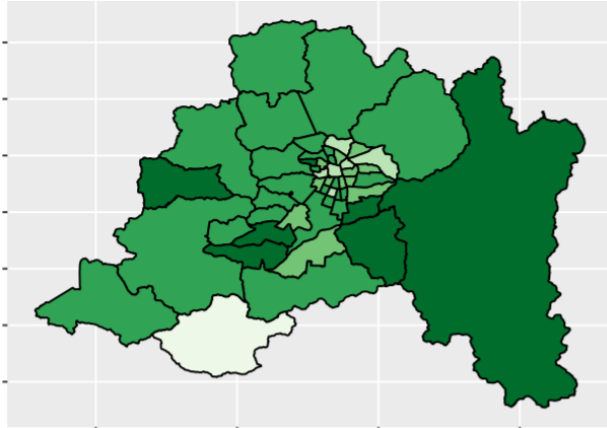
Var. Mov. Completa por comuna de la semana 202038  
Considerando 202010 - 202111, Región Metropolitana



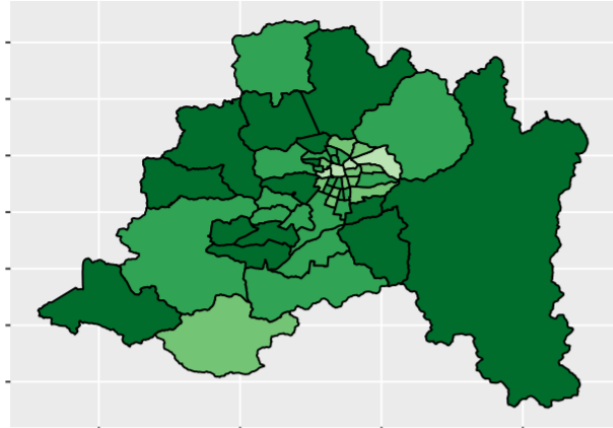
Var. Mov. Completa por comuna de la semana 202039  
Considerando 202010 - 202111, Región Metropolitana



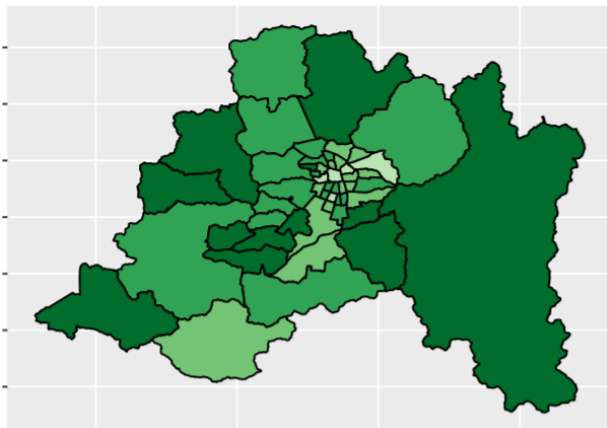
Var. Mov. Completa por comuna de la semana 202040  
Considerando 202010 – 202111 , Región Metropolitana



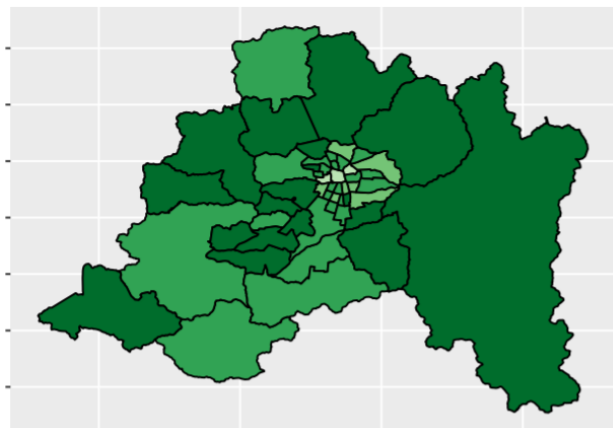
Var. Mov. Completa por comuna de la semana 202041  
Considerando 202010 – 202111 , Región Metropolitana



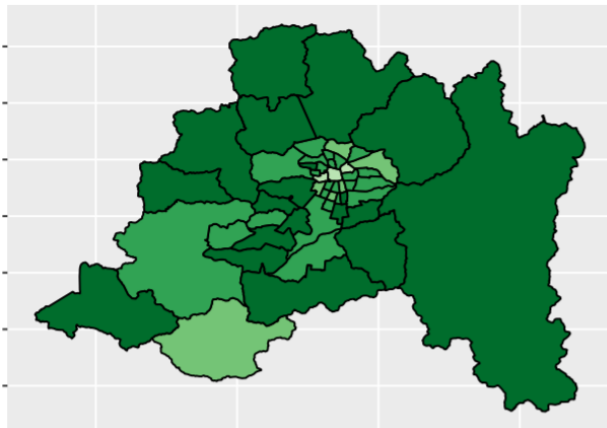
Var. Mov. Completa por comuna de la semana 202042  
Considerando 202010 – 202111 , Región Metropolitana



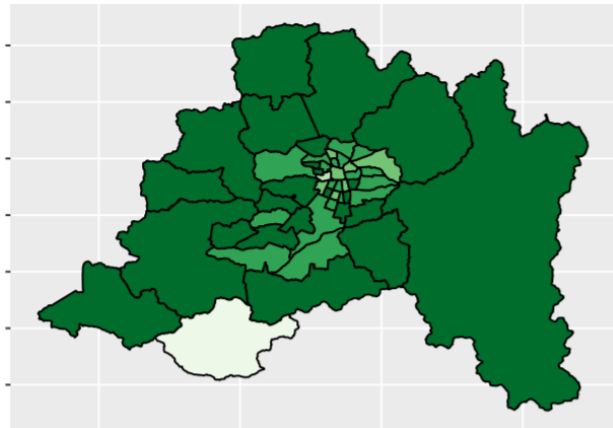
Var. Mov. Completa por comuna de la semana 202043  
Considerando 202010 – 202111 , Región Metropolitana



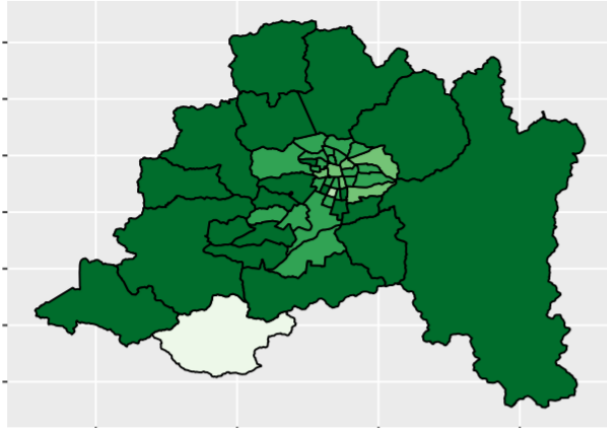
Var. Mov. Completa por comuna de la semana 202044  
Considerando 202010 – 202111 , Región Metropolitana



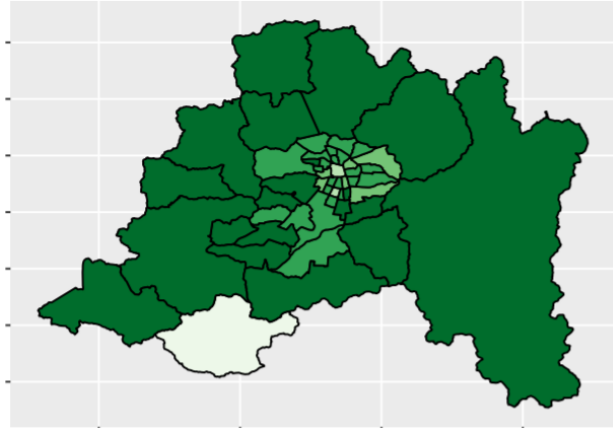
Var. Mov. Completa por comuna de la semana 202045  
Considerando 202010 – 202111 , Región Metropolitana



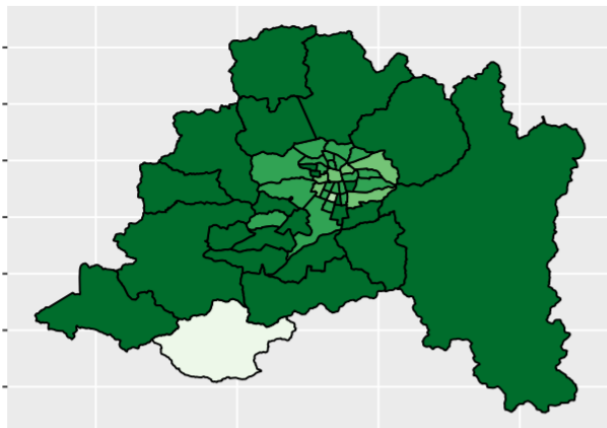
Var. Mov. Completa por comuna de la semana 202046  
Considerando 202010 – 202111 , Región Metropolitana



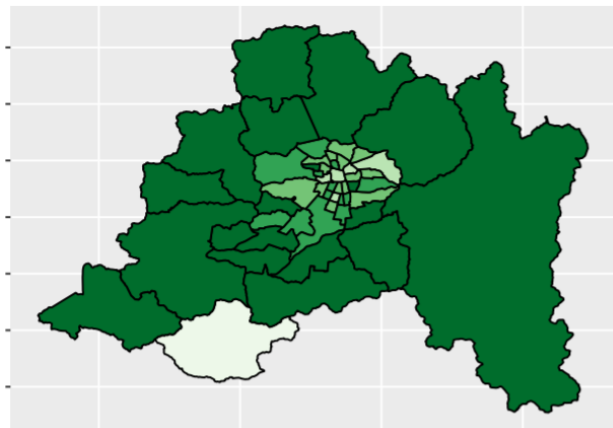
Var. Mov. Completa por comuna de la semana 202047  
Considerando 202010 – 202111 , Región Metropolitana



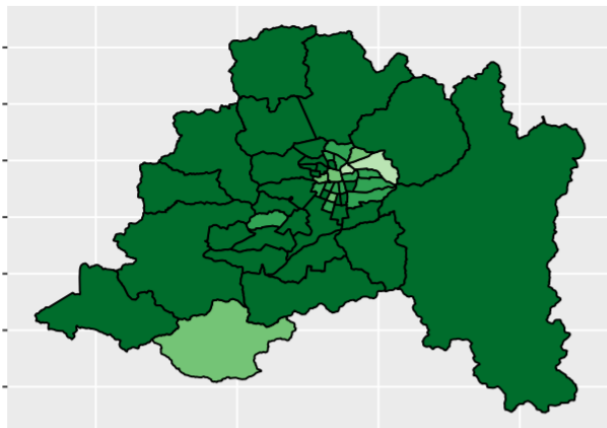
Var. Mov. Completa por comuna de la semana 202048  
Considerando 202010 – 202111 , Región Metropolitana



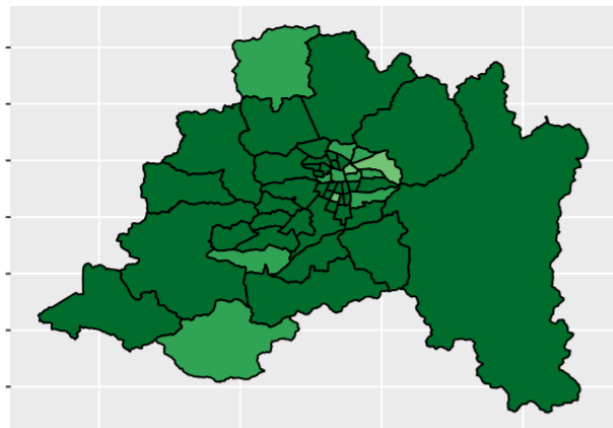
Var. Mov. Completa por comuna de la semana 202049  
Considerando 202010 – 202111 , Región Metropolitana



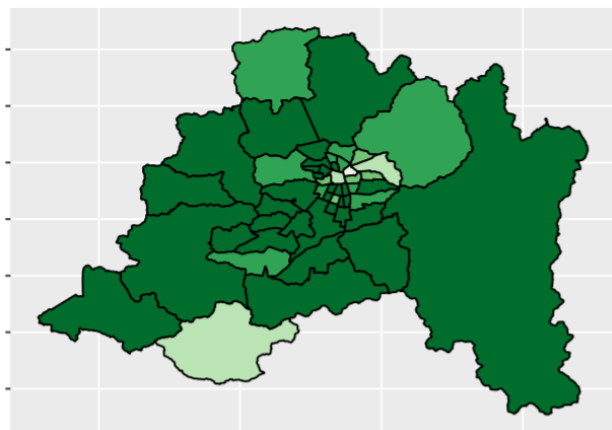
Var. Mov. Completa por comuna de la semana 202050  
Considerando 202010 – 202111 , Región Metropolitana



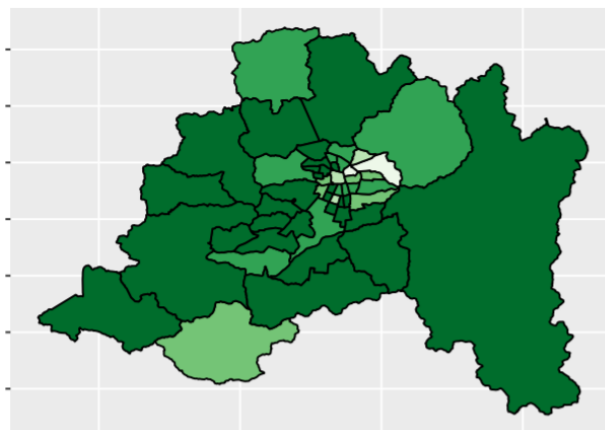
Var. Mov. Completa por comuna de la semana 202051  
Considerando 202010 – 202111 , Región Metropolitana



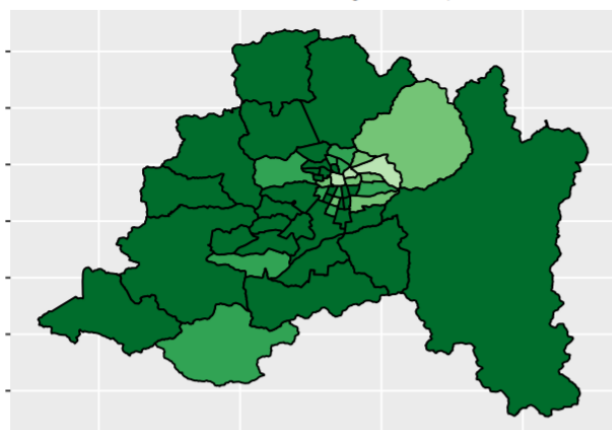
Var. Mov. Completa por comuna de la semana 202052  
Considerando 202010 – 202111 , Región Metropolitana



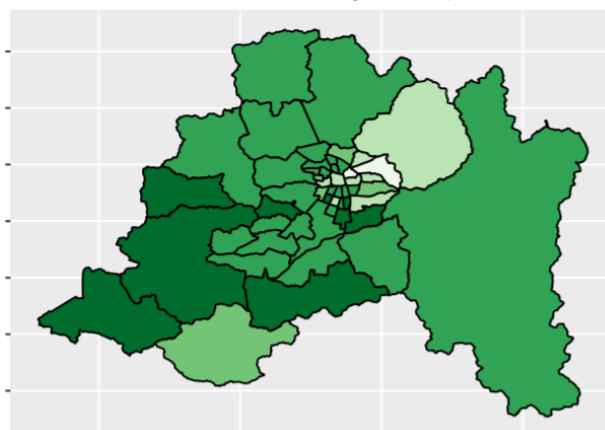
Var. Mov. Completa por comuna de la semana 202053  
Considerando 202010 – 202111 , Región Metropolitana



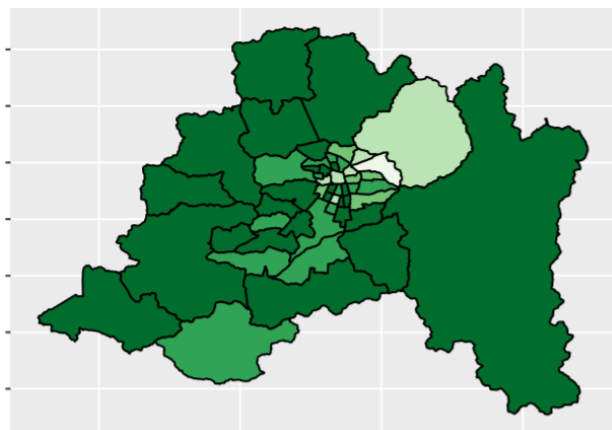
Var. Mov. Completa por comuna de la semana 202101  
Considerando 202010 – 202111 , Región Metropolitana



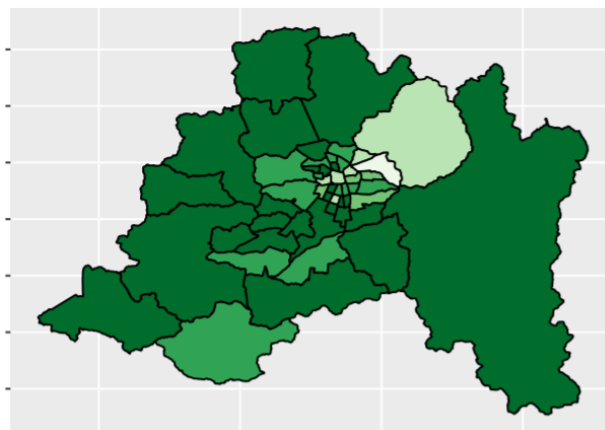
Var. Mov. Completa por comuna de la semana 202102  
Considerando 202010 – 202111 , Región Metropolitana



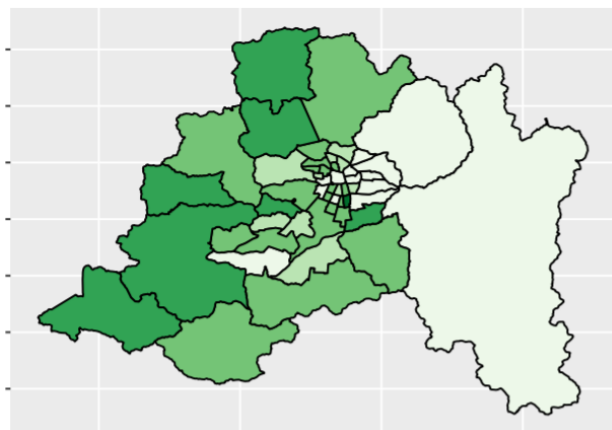
Var. Mov. Completa por comuna de la semana 202103  
Considerando 202010 – 202111 , Región Metropolitana



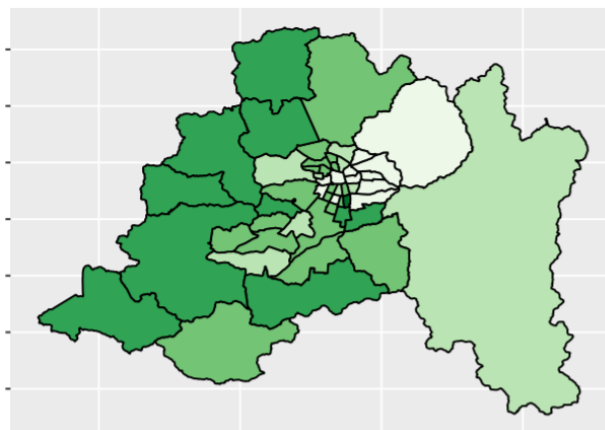
Var. Mov. Completa por comuna de la semana 202104  
Considerando 202010 – 202111 , Región Metropolitana



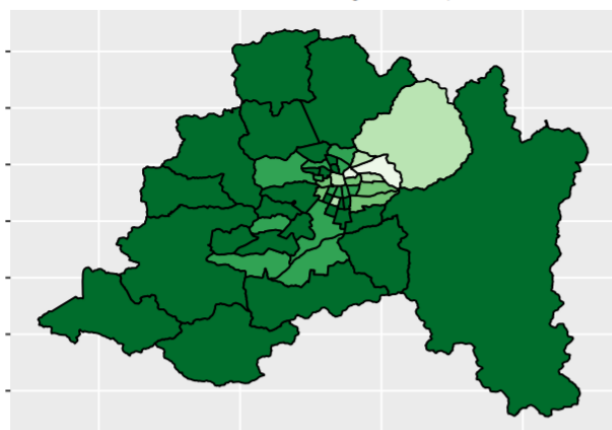
Var. Mov. Completa por comuna de la semana 202105  
Considerando 202010 – 202111 , Región Metropolitana



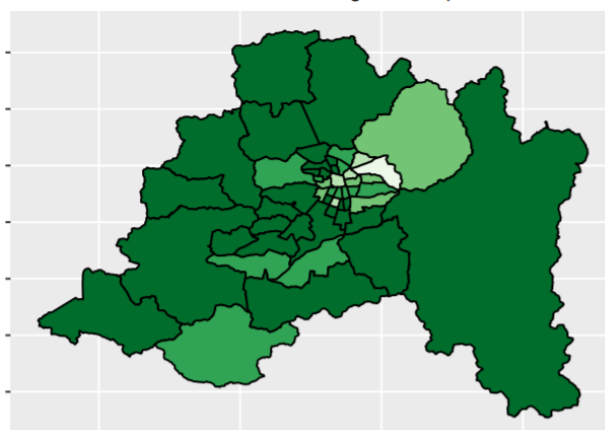
Var. Mov. Completa por comuna de la semana 202106  
Considerando 202010 – 202111 , Región Metropolitana



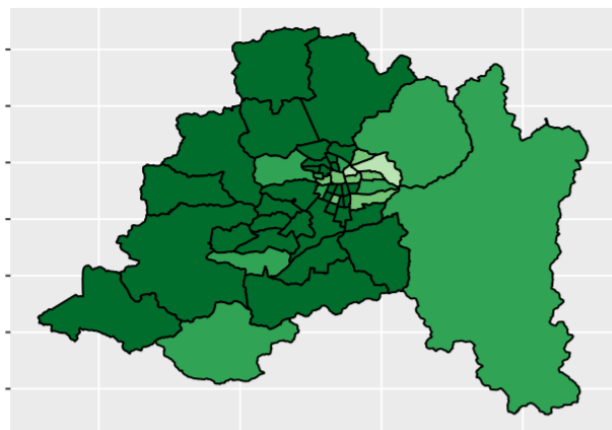
Var. Mov. Completa por comuna de la semana 202107  
Considerando 202010 – 202111 , Región Metropolitana



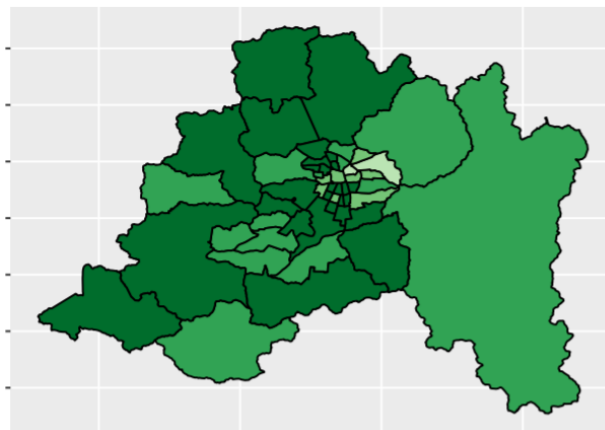
Var. Mov. Completa por comuna de la semana 202108  
Considerando 202010 – 202111 , Región Metropolitana



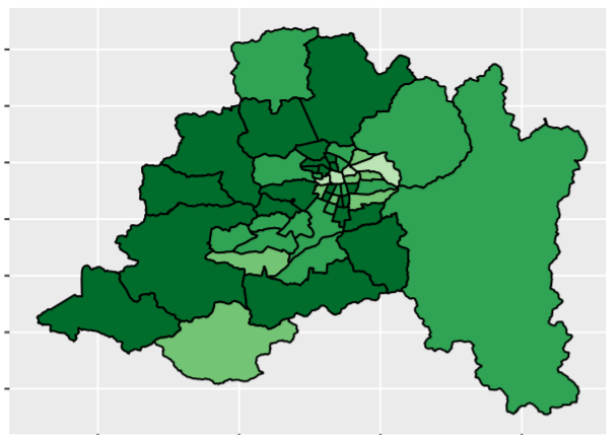
Var. Mov. Completa por comuna de la semana 202109  
Considerando 202010 – 202111 , Región Metropolitana



Var. Mov. Completa por comuna de la semana 202110  
Considerando 202010 – 202111 , Región Metropolitana

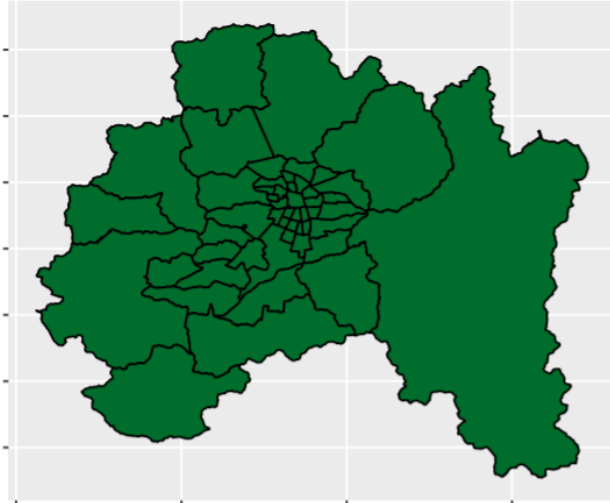


Var. Mov. Completa por comuna de la semana 202111  
Considerando 202010 – 202111 , Región Metropolitana

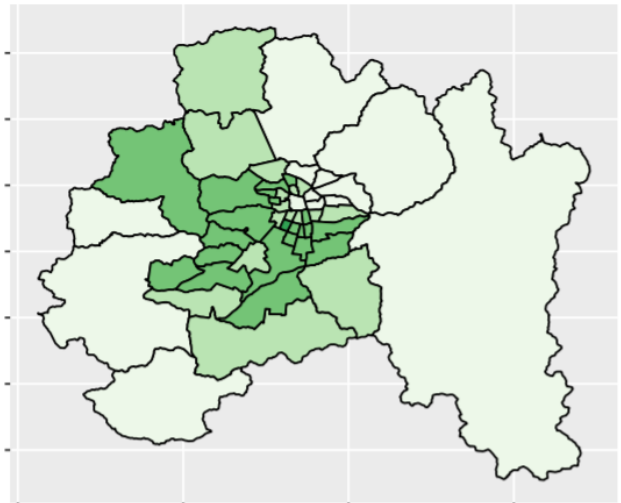


### B.3. Mapas de los Quintiles de la Variación de Movilidad Laboral

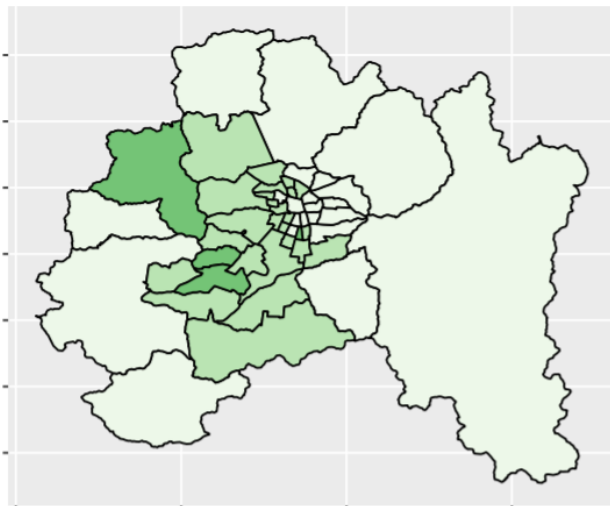
Var. Mov. Laboral por comuna de la semana 202010  
Considerando 202010 – 202111 , Región Metropolitana



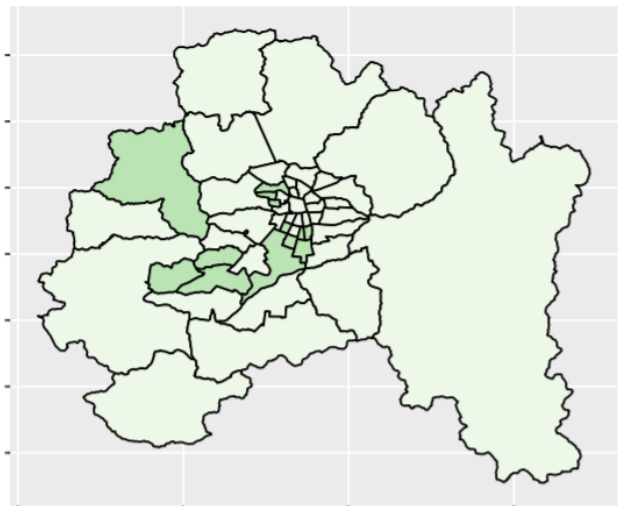
Var. Mov. Laboral por comuna de la semana 202019  
Considerando 202010 – 202111 , Región Metropolitana



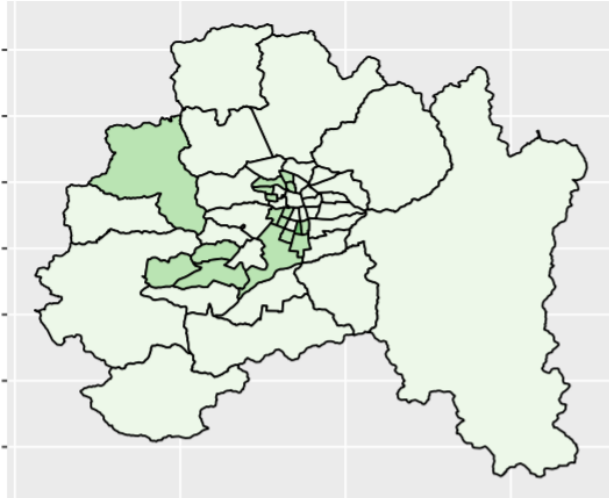
Var. Mov. Laboral por comuna de la semana 202020  
Considerando 202010 – 202111 , Región Metropolitana



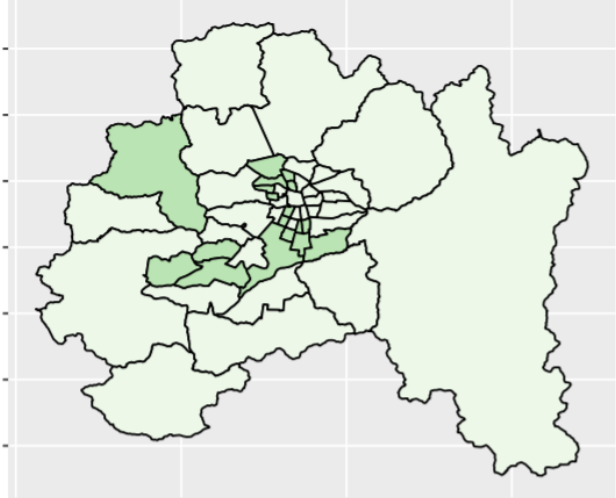
Var. Mov. Laboral por comuna de la semana 202021  
Considerando 202010 – 202111 , Región Metropolitana



Var. Mov. Laboral por comuna de la semana 202022  
Considerando 202010 - 202111, Región Metropolitana



Var. Mov. Laboral por comuna de la semana 202023  
Considerando 202010 - 202111, Región Metropolitana



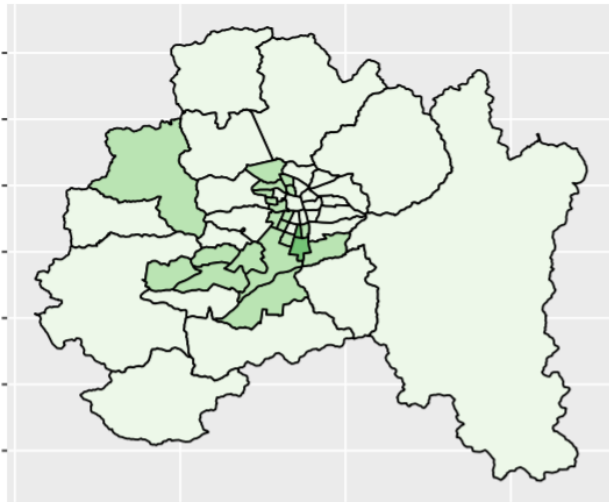
Var. Mov. Laboral por comuna de la semana 202024  
Considerando 202010 - 202111, Región Metropolitana



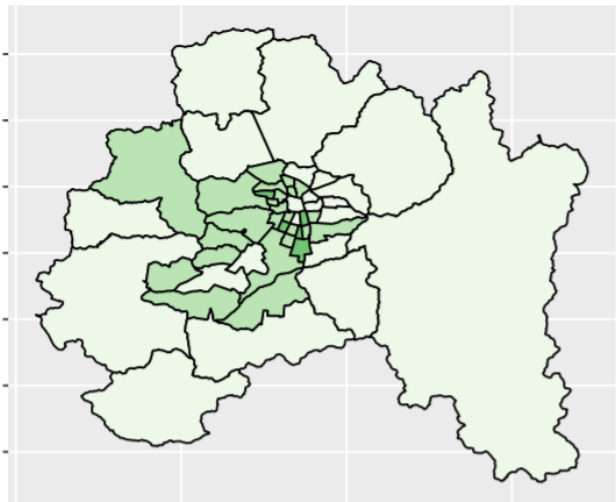
Var. Mov. Laboral por comuna de la semana 202025  
Considerando 202010 - 202111, Región Metropolitana



Var. Mov. Laboral por comuna de la semana 202026  
Considerando 202010 - 202111, Región Metropolitana



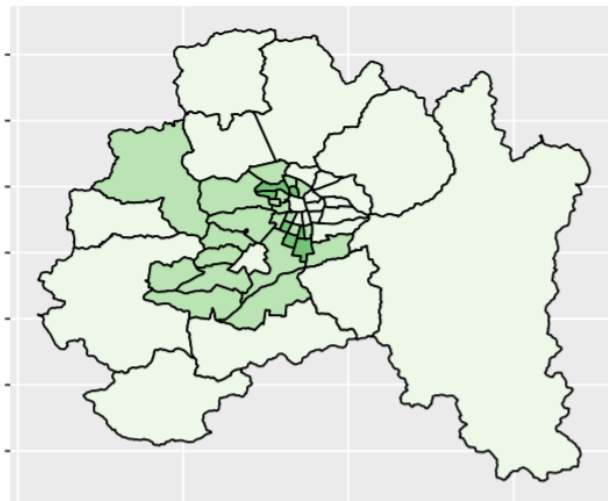
Var. Mov. Laboral por comuna de la semana 202027  
Considerando 202010 - 202111, Región Metropolitana



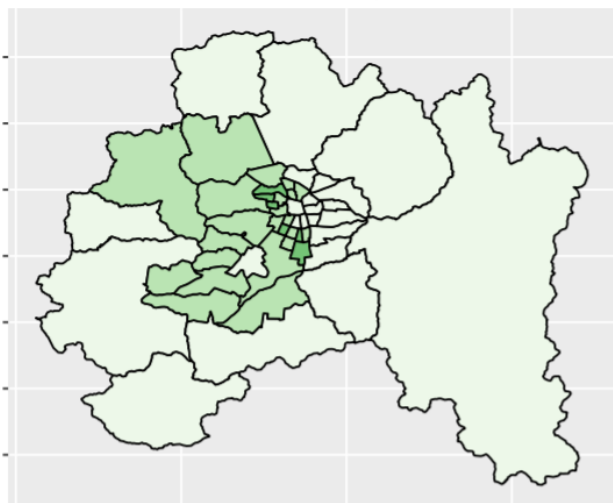
Var. Mov. Laboral por comuna de la semana 202028  
Considerando 202010 – 202111 , Región Metropolitana



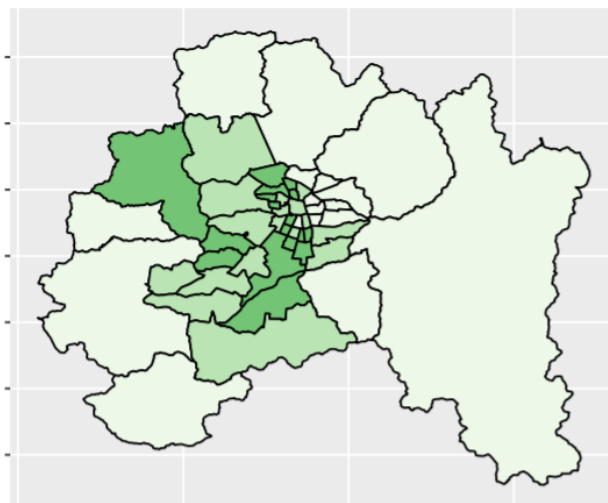
Var. Mov. Laboral por comuna de la semana 202029  
Considerando 202010 – 202111 , Región Metropolitana



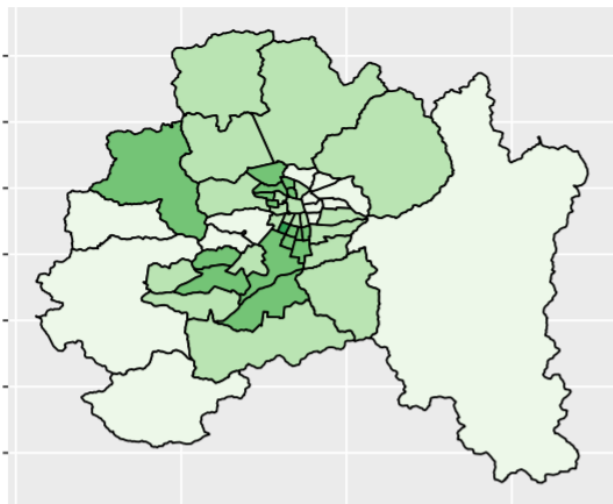
Var. Mov. Laboral por comuna de la semana 202030  
Considerando 202010 – 202111 , Región Metropolitana



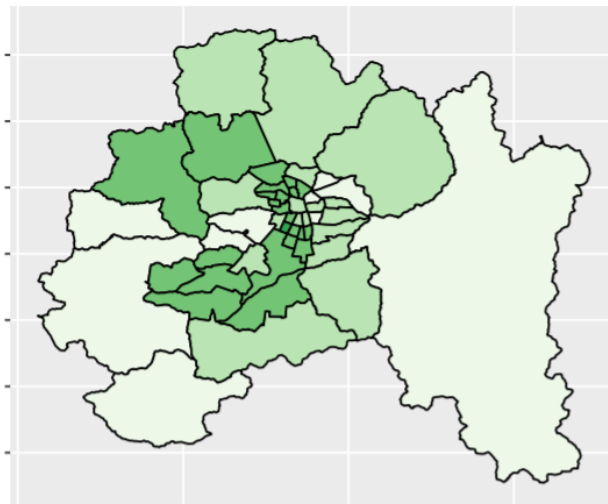
Var. Mov. Laboral por comuna de la semana 202031  
Considerando 202010 – 202111 , Región Metropolitana



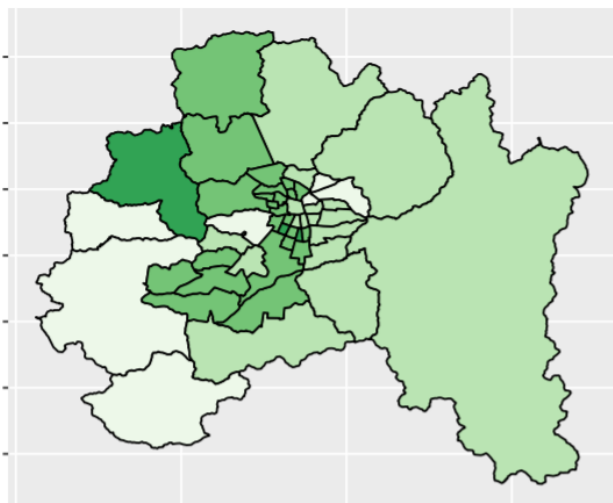
Var. Mov. Laboral por comuna de la semana 202032  
Considerando 202010 – 202111 , Región Metropolitana



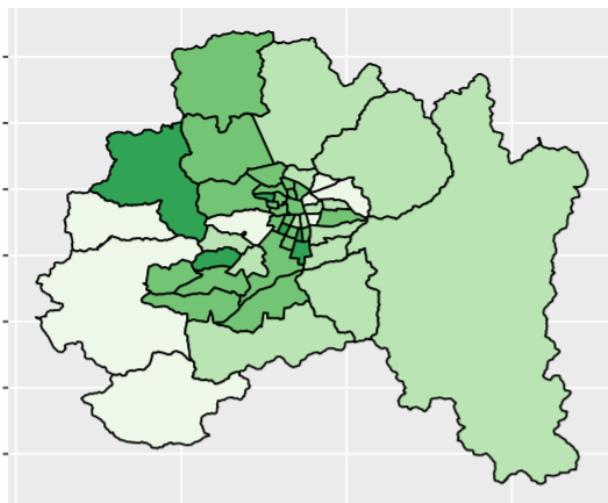
Var. Mov. Laboral por comuna de la semana 202033  
Considerando 202010 – 202111 , Región Metropolitana



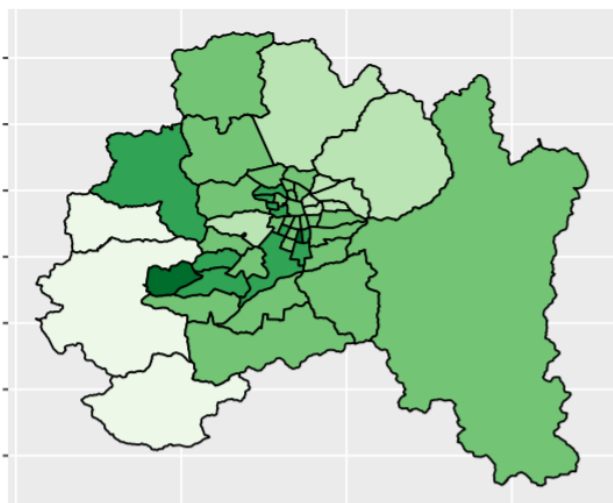
Var. Mov. Laboral por comuna de la semana 202034  
Considerando 202010 - 202111, Región Metropolitana



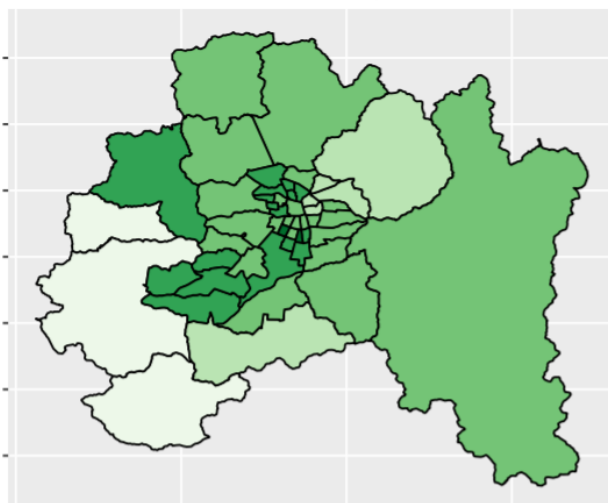
Var. Mov. Laboral por comuna de la semana 202035  
Considerando 202010 - 202111, Región Metropolitana



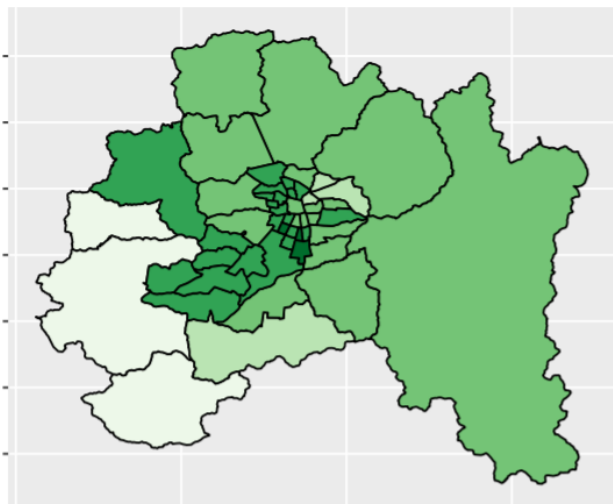
Var. Mov. Laboral por comuna de la semana 202036  
Considerando 202010 - 202111, Región Metropolitana



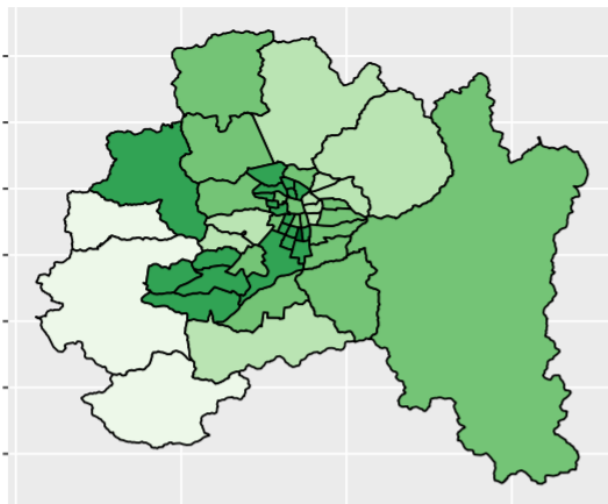
Var. Mov. Laboral por comuna de la semana 202037  
Considerando 202010 - 202111, Región Metropolitana



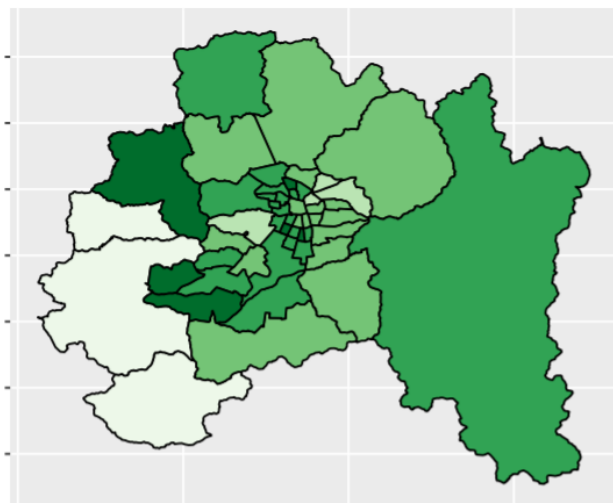
Var. Mov. Laboral por comuna de la semana 202038  
Considerando 202010 - 202111, Región Metropolitana



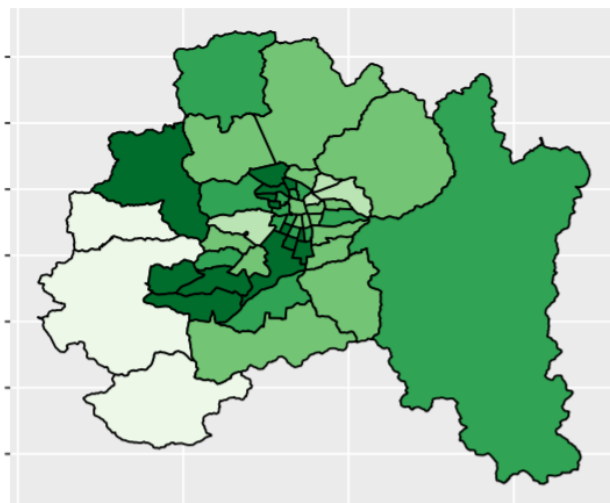
Var. Mov. Laboral por comuna de la semana 202039  
Considerando 202010 - 202111, Región Metropolitana



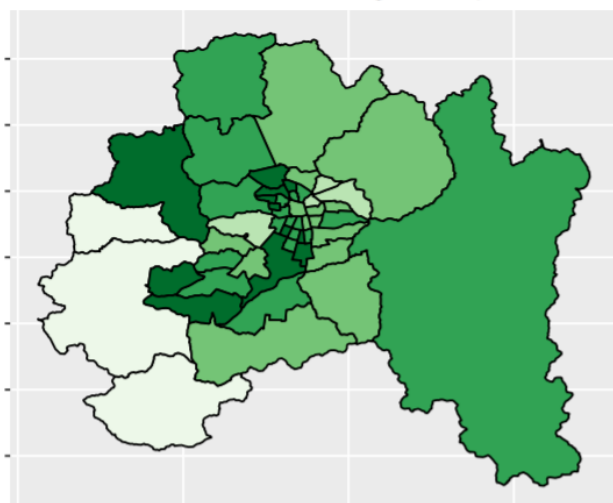
Var. Mov. Laboral por comuna de la semana 202040  
Considerando 202010 - 202111, Región Metropolitana



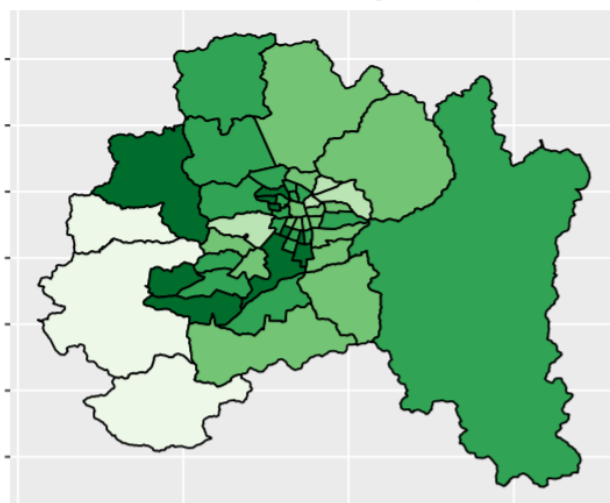
Var. Mov. Laboral por comuna de la semana 202041  
Considerando 202010 - 202111, Región Metropolitana



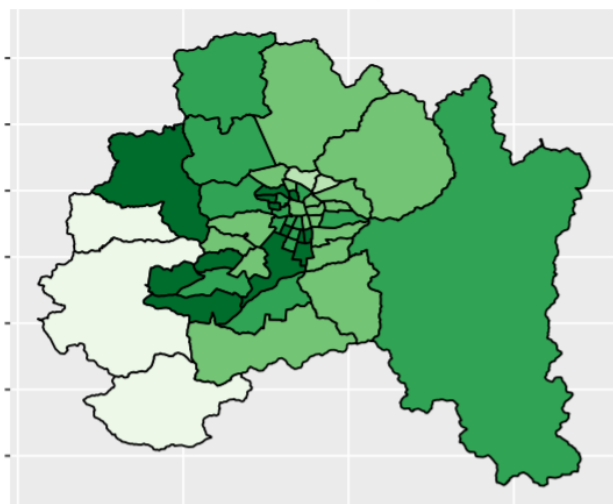
Var. Mov. Laboral por comuna de la semana 202042  
Considerando 202010 - 202111, Región Metropolitana



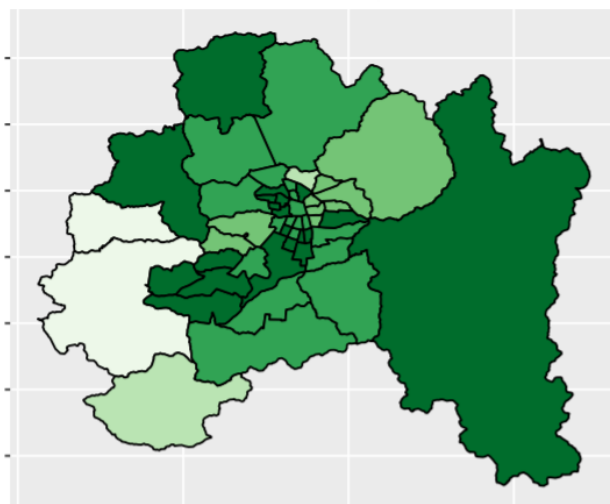
Var. Mov. Laboral por comuna de la semana 202043  
Considerando 202010 - 202111, Región Metropolitana



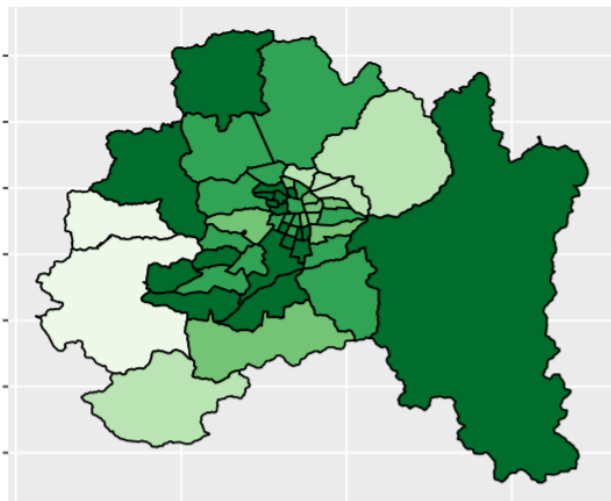
Var. Mov. Laboral por comuna de la semana 202044  
Considerando 202010 - 202111, Región Metropolitana



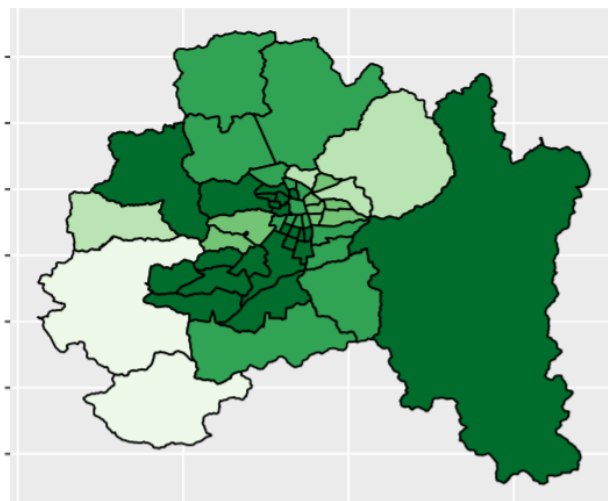
Var. Mov. Laboral por comuna de la semana 202045  
Considerando 202010 - 202111, Región Metropolitana



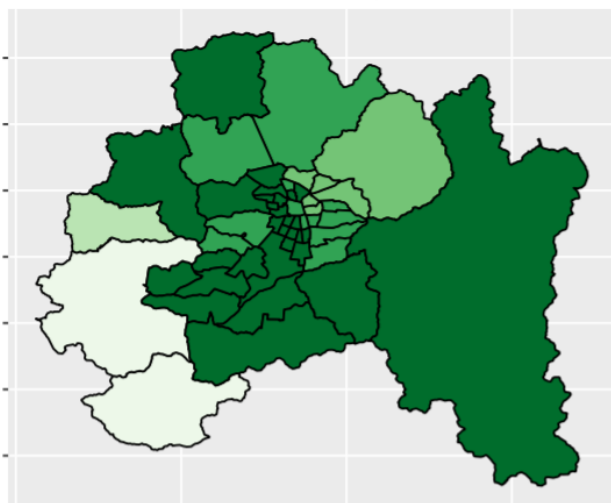
Var. Mov. Laboral por comuna de la semana 202046  
Considerando 202010 - 202111, Región Metropolitana



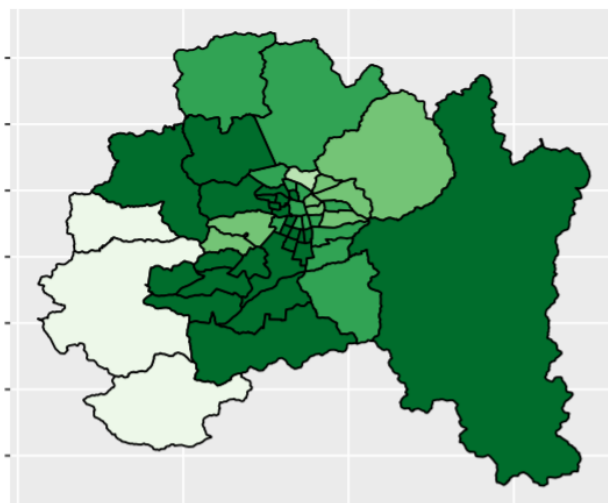
Var. Mov. Laboral por comuna de la semana 202047  
Considerando 202010 - 202111, Región Metropolitana



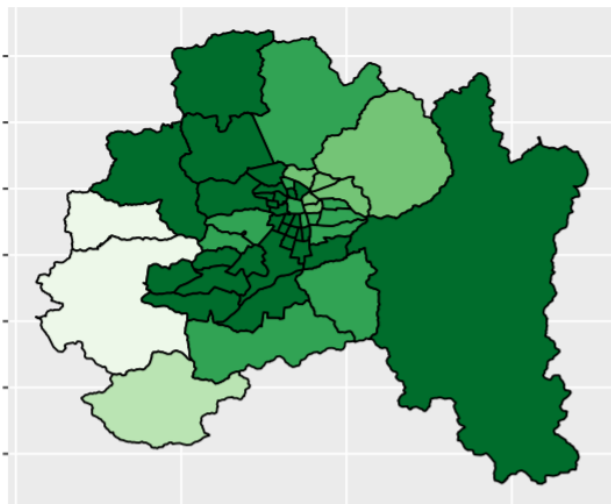
Var. Mov. Laboral por comuna de la semana 202048  
Considerando 202010 - 202111, Región Metropolitana



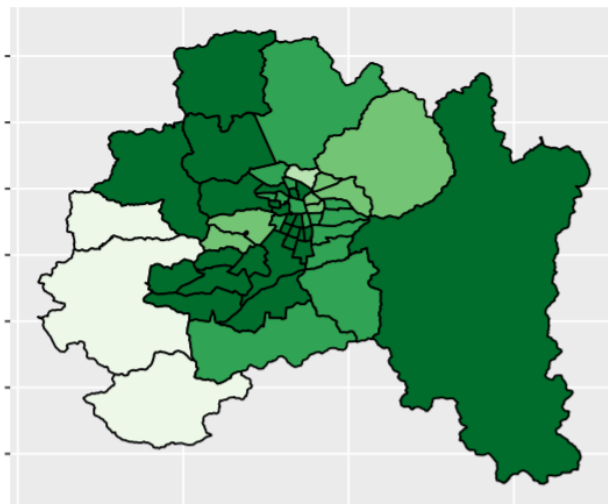
Var. Mov. Laboral por comuna de la semana 202049  
Considerando 202010 - 202111, Región Metropolitana



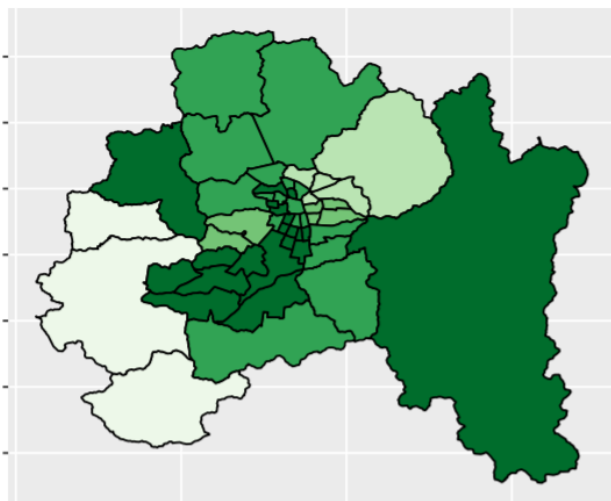
Var. Mov. Laboral por comuna de la semana 202050  
Considerando 202010 - 202111, Región Metropolitana



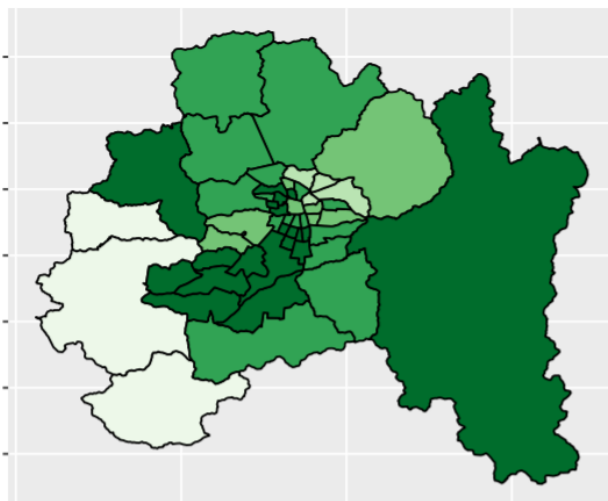
Var. Mov. Laboral por comuna de la semana 202051  
Considerando 202010 - 202111, Región Metropolitana



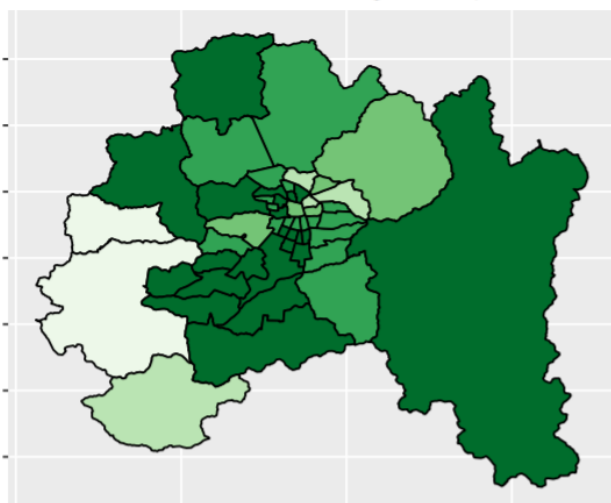
Var. Mov. Laboral por comuna de la semana 202052  
Considerando 202010 – 202111 , Región Metropolitana



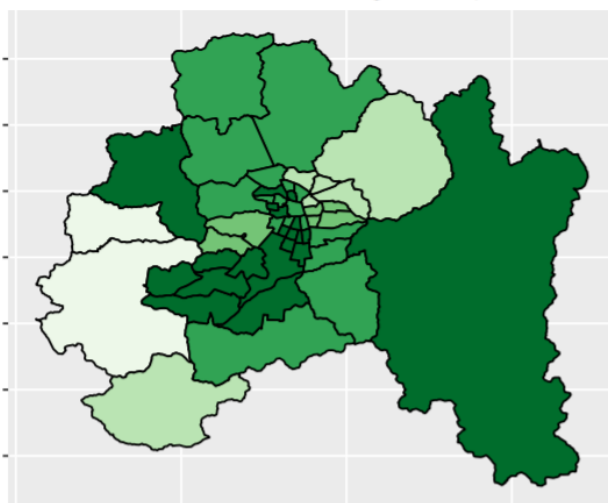
Var. Mov. Laboral por comuna de la semana 202053  
Considerando 202010 – 202111 , Región Metropolitana



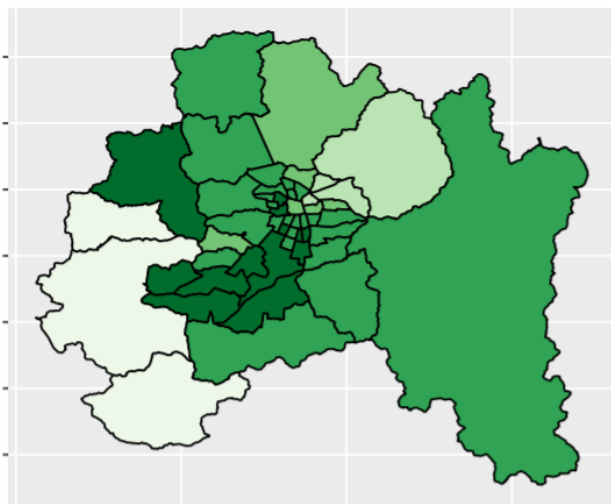
Var. Mov. Laboral por comuna de la semana 202101  
Considerando 202010 – 202111 , Región Metropolitana



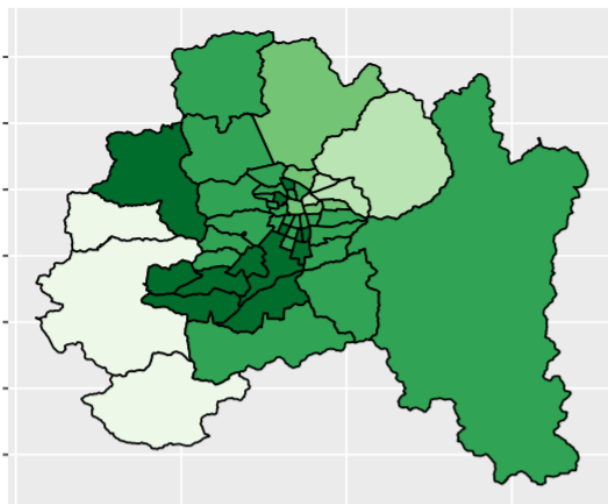
Var. Mov. Laboral por comuna de la semana 202102  
Considerando 202010 – 202111 , Región Metropolitana



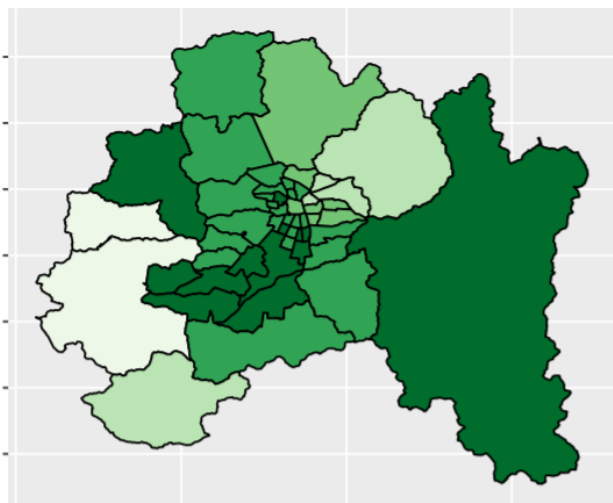
Var. Mov. Laboral por comuna de la semana 202103  
Considerando 202010 – 202111 , Región Metropolitana



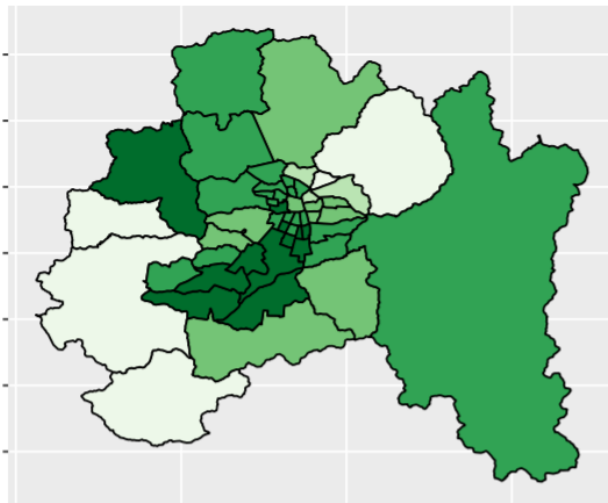
Var. Mov. Laboral por comuna de la semana 202104  
Considerando 202010 – 202111 , Región Metropolitana



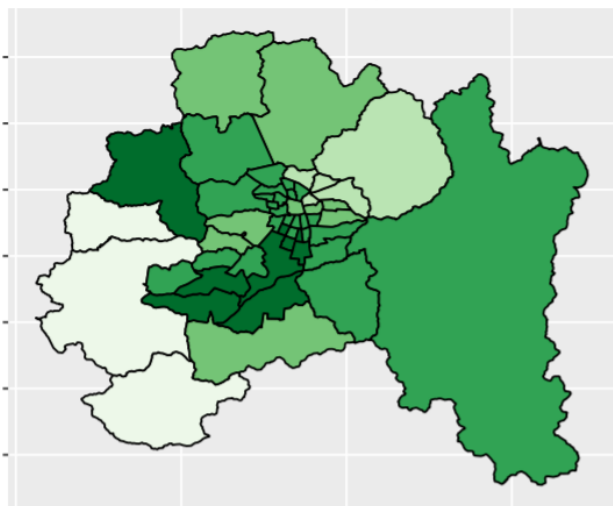
Var. Mov. Laboral por comuna de la semana 202105  
Considerando 202010 - 202111, Región Metropolitana



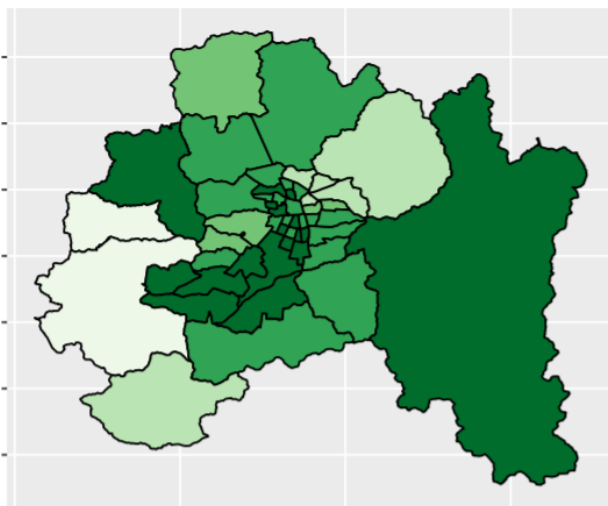
Var. Mov. Laboral por comuna de la semana 202106  
Considerando 202010 - 202111, Región Metropolitana



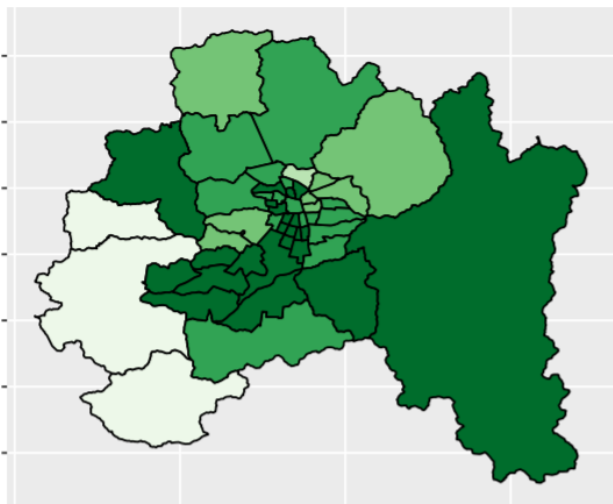
Var. Mov. Laboral por comuna de la semana 202107  
Considerando 202010 - 202111, Región Metropolitana



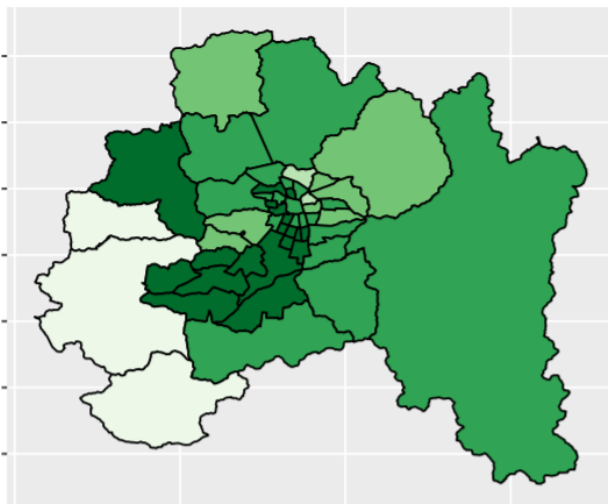
Var. Mov. Laboral por comuna de la semana 202108  
Considerando 202010 - 202111, Región Metropolitana



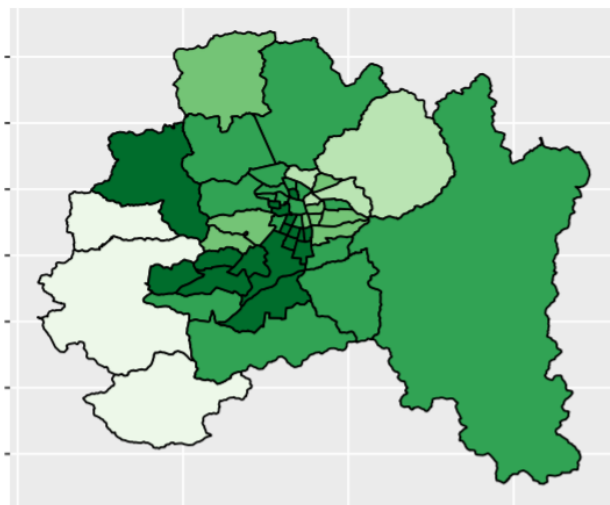
Var. Mov. Laboral por comuna de la semana 202109  
Considerando 202010 - 202111, Región Metropolitana



Var. Mov. Laboral por comuna de la semana 202110  
Considerando 202010 - 202111, Región Metropolitana



Var. Mov. Laboral por comuna de la semana 202111  
Considerando 202010 - 202111 , Región Metropolitana



# Apéndice C

## Correlaciones, período entre olas

Semana	Corr.VMCompleta	Corr.VMLaboral
202031	0.025	0.007
202032	0.123	0.244
202033	0.035	0.311
202034	0.14	0.286
202035	0.083	0.001
202036	0.151	0.338
202037	0.166	0.233
202038	0.104	0.03
202039	0.19	0.201
202040	0.134	0
202041	0.013	0.243
202042	0.037	0.303
202043	0.011	0.068
202044	0.147	0.108
202045	0.137	0.179
202046	0.132	0.296
202047	0.103	0.168
202048	0.398	0.293
202049	0.096	0.069
202050	0.044	0.083
202051	0.023	0.162
202052	0.137	0.028
202053	0.023	0.21
202101	0.178	0.17
202102	0.113	0.12
202103	0.074	0.265
202104	0.107	0.128
202105	0.077	0.049
202106	0.152	0.129

## Apéndice D

### Tablas con los p-valores del test de Ljung-Box

#### D.1. P-valores del test de Ljung-Box considerando todas las semanas y algunas comunas

P-valores Ljung-Box, Comuna de Santiago, considerando todas las semanas

lag	p.value.TC	statistic.TC	p.value.VMC	statistic.VMC	p.value.VML	statistic.VML
1	0	54.447	0	43.321	0	45.018
2	0	94.881	0	76.089	0	76.843
3	0	119.519	0	105.461	0	100.489
4	0	131.427	0	131.269	0	121.657
5	0	135.116	0	155.526	0	141.908
6	0	135.395	0	178.949	0	161.737
7	0	135.772	0	197.66	0	180.867
8	0	137.664	0	212.299	0	196.026
9	0	140.343	0	224.806	0	207.066
10	0	143.374	0	234.415	0	214.502

Figura D.1: P-valores de 10 lags para la serie de Santiago

P-valores Ljung-Box, Comuna de Las Condes, considerando todas las semanas

lag	p.value.TC	statistic.TC	p.value.VMC	statistic.VMC	p.value.VML	statistic.VML
1	0	52.3	0	41.195	0	36.691
2	0	85.37	0	70.712	0	60.684
3	0	100.032	0	93.708	0	77.268
4	0	103.779	0	110.841	0	89.824
5	0	103.917	0	125.569	0	101.339
6	0	104.546	0	138.682	0	113.472
7	0	106.985	0	147.468	0	124.995
8	0	110.676	0	153.584	0	134.212
9	0	114.278	0	157.059	0	140.798
10	0	117.11	0	157.979	0	143.813

Figura D.2: P-valores de 10 lags para la serie de Las Condes

P-valores Ljung-Box, Comuna de Ñuñoa, considerando todas las semanas

lag	p.value.TC	statistic.TC	p.value.VMC	statistic.VMC	p.value.VML	statistic.VML
1	0	53.081	0	46.673	0	39.618
2	0	91.398	0	83.217	0	65.788
3	0	112.513	0	115.261	0	84.394
4	0	120.878	0	141.869	0	100.222
5	0	122.563	0	165.338	0	114.676
6	0	122.575	0	186.993	0	128.493
7	0	123.743	0	204.164	0	142.258
8	0	126.924	0	218.078	0	154.502
9	0	131.008	0	229.563	0	164.147
10	0	135.031	0	237.325	0	169.181

Figura D.3: P-valores de 10 lags para la serie de Ñuñoa

P-valores Ljung-Box, Comuna de Quinta Normal, considerando todas las semanas

lag	p.value.TC	statistic.TC	p.value.VMC	statistic.VMC	p.value.VML	statistic.VML
1	0	54.572	0	47.457	0	50.196
2	0	95.205	0	84.511	0	90.683
3	0	120.523	0	117.104	0	126.562
4	0	132.588	0	144.493	0	159.628
5	0	136.171	0	168.982	0	189.958
6	0	136.45	0	191.924	0	217.332
7	0	136.772	0	212.31	0	240.332
8	0	138.428	0	232.466	0	258.595
9	0	141.06	0	252.619	0	270.85
10	0	144.077	0	270.302	0	278.865

Figura D.4: P-valores de 10 lags para la serie de Quinta Normal

P-valores Ljung-Box, Comuna de Vitacura, considerando todas las semanas

lag	p.value.TC	statistic.TC	p.value.VMC	statistic.VMC	p.value.VML	statistic.VML
1	0	46.12	0	45.023	0	38.462
2	0	71.934	0	79.013	0	65.52
3	0	82.812	0	105.312	0	84.436
4	0	85.413	0	123.798	0	100.095
5	0	85.547	0	138.841	0	115.543
6	0	85.782	0	151.935	0	130.383
7	0	86.842	0	161.88	0	145.666
8	0	88.476	0	169.735	0	158.944
9	0	89.662	0	174.972	0	168.862
10	0	90.288	0	177.413	0	173.979

Figura D.5: P-valores de 10 lags para la serie de Vitacura

P-valores Ljung-Box, Comuna de Puente Alto, considerando todas las semanas

lag	p.value.TC	statistic.TC	p.value.VMC	statistic.VMC	p.value.VML	statistic.VML
1	0	55.107	0	48.466	0	41.767
2	0	96.365	0	87.012	0	71.77
3	0	121.82	0	120.816	0	97.867
4	0	133.991	0	149.44	0	121.758
5	0	137.669	0	174.107	0	141.309
6	0	137.923	0	196.269	0	160.237
7	0	138.329	0	216.112	0	178.352
8	0	140.306	0	235.275	0	192.918
9	0	143.501	0	254.996	0	201.142
10	0	146.896	0	273.245	0	205.954

Figura D.6: P-valores de 10 lags para la serie de Puente Alto

P-valores Ljung-Box, Comuna de La Florida, considerando todas las semanas

lag	p.value.TC	statistic.TC	p.value.VMC	statistic.VMC	p.value.VML	statistic.VML
1	0	55.047	0	44.02	0	38.866
2	0	96.591	0	75.678	0	64.164
3	0	122.231	0	103.151	0	84.757
4	0	134.386	0	125.163	0	102.778
5	0	138.052	0	144.805	0	119.231
6	0	138.273	0	164.028	0	134.675
7	0	138.782	0	180.164	0	150.064
8	0	141.028	0	193.489	0	164.612
9	0	144.598	0	204.478	0	172.847
10	0	148.242	0	212.818	0	177.628

Figura D.7: P-valores de 10 lags para la serie de La Florida

P-valores Ljung-Box, Comuna de Pudahuel, considerando todas las semanas

lag	p.value.TC	statistic.TC	p.value.VMC	statistic.VMC	p.value.VML	statistic.VML
1	0	53.757	0	42.698	0	46.695
2	0	93.461	0	74.127	0	82.63
3	0	116.272	0	102.313	0	113.021
4	0	125.473	0	125.313	0	140.211
5	0	127.604	0	146.416	0	165.536
6	0	127.62	0	166.179	0	189.737
7	0	128.431	0	181.435	0	212.672
8	0	130.795	0	194.779	0	231.345
9	0	134.159	0	208.129	0	244.113
10	0	137.703	0	219.786	0	251.121

Figura D.8: P-valores de 10 lags para la serie de Pudahuel

P-valores Ljung-Box, Comuna de Cerrillos, considerando todas las semanas

lag	p.value.TC	statistic.TC	p.value.VMC	statistic.VMC	p.value.VML	statistic.VML
1	0	52.457	0	38.62	0	41.734
2	0	93.014	0	63.23	0	72.122
3	0	121.611	0	82.687	0	97.535
4	0	137.323	0	97.481	0	119.372
5	0	144.692	0	112.052	0	140.054
6	0	146.741	0	127.325	0	159.108
7	0	146.741	0	140.001	0	177.144
8	0	147.68	0	152.2	0	189.734
9	0	150.315	0	163.597	0	198.269
10	0	154.463	0	172.909	0	203.589

Figura D.9: P-valores de 10 lags para la serie de Cerrillos

## D.2. P-valores del test de Ljung-Box considerando el periodo 1, para algunas comunas

P-valores Ljung-Box, periodo considerado entre las semanas 202011 a 202030

lag	p.value.TC	statistic.TC	p.value.VMC	statistic.VMC	p.value.VML	statistic.VML	Comuna
1	0	17.320	0.041	4.163	0.034	4.487	Santiago
1	0	17.898	0.057	3.616	0.087	2.923	Las Condes
1	0	17.490	0.018	5.612	0.075	3.174	Nunoa
1	0	18.274	0.017	5.717	0.021	5.289	Quinta Normal
1	0	16.099	0.100	2.713	0.114	2.494	Vitacura
1	0	18.255	0.005	8.013	0.031	4.663	Puente Alto
1	0	18.192	0.058	3.582	0.055	3.696	La Florida
1	0	18.075	0.112	2.532	0.027	4.905	Pudahuel
1	0	16.824	0.040	4.212	0.025	5.001	Cerrillos

Figura D.10: P-valores de 1 lag para cada serie seleccionada

## D.3. P-valores del test de Ljung-Box considerando el periodo 2, para algunas comunas

P-valores Ljung-Box, periodo considerado entre las semanas 202031 a 202106

lag	p.value.TC	statistic.TC	p.value.VMC	statistic.VMC	p.value.VML	statistic.VML	Comuna
1	0.000	22.511	0.000	16.261	0	22.507	Santiago
1	0.000	25.398	0.000	18.942	0	16.920	Las Condes
1	0.000	21.976	0.000	15.311	0	21.812	Nunoa
1	0.000	20.143	0.000	19.619	0	22.781	Quinta Normal
1	0.000	16.741	0.000	19.278	0	19.011	Vitacura
1	0.000	24.754	0.000	18.116	0	23.973	Puente Alto
1	0.000	25.417	0.000	16.574	0	21.879	La Florida
1	0.000	22.676	0.000	14.475	0	25.674	Pudahuel
1	0.001	10.919	0.001	10.538	0	21.008	Cerrillos

Figura D.11: P-valores de 1 lag para cada serie seleccionada

#### D.4. P-valores del test de Ljung-Box considerando el periodo 3, para algunas comunas

P-valores Ljung-Box, periodo considerado entre las semanas 202107 a 202117

lag	p.value.TC	statistic.TC	p.value.VML	statistic.VML	Comuna
1	0.009	6.881	0.002	9.269	Santiago
1	0.008	7.074	0.005	8.024	Las Condes
1	0.028	4.815	0.001	10.200	Nunoa
1	0.005	7.894	0.001	10.122	Quinta Normal
1	0.008	7.123	0.011	6.514	Vitacura
1	0.003	9.137	0.002	9.634	Puente Alto
1	0.006	7.446	0.001	10.202	La Florida
1	0.009	6.878	0.001	10.262	Pudahuel
1	0.057	3.624	0.001	10.122	Cerrillos

Figura D.12: P-valores de 1 lag para cada serie seleccionada

# Apéndice E

## Tablas de las Pruebas LM

Test de Efecto Aleatorio, Correlación Espacial y Correlación Serial en los errores, Primera Ola

test	variables	LM	p_value
J	TasaContagios ~ V.M.Completa	2736.047	0
C1	TasaContagios ~ V.M.Completa	24.755	0
C2	TasaContagios ~ V.M.Completa	21.912	0
C3	TasaContagios ~ V.M.Completa	-0.078	1
J	TasaContagios ~ V.M.Laboral	2724.851	0
C1	TasaContagios ~ V.M.Laboral	25.935	0
C2	TasaContagios ~ V.M.Laboral	19.615	0
C3	TasaContagios ~ V.M.Laboral	-0.053	1

Test de Efecto Aleatorio, Correlación Espacial y Correlación Serial en los errores, Periodo entre Olas

test	variables	LM	p_value
J	TasaContagios ~ V.M.Completa	1395.703	0.000
C1	TasaContagios ~ V.M.Completa	14.080	0.000
C2	TasaContagios ~ V.M.Completa	289.663	0.000
C3	TasaContagios ~ V.M.Completa	-14.595	1.000
J	TasaContagios ~ V.M.Laboral	1474.167	0.000
C1	TasaContagios ~ V.M.Laboral	11.714	0.001
C2	TasaContagios ~ V.M.Laboral	242.470	0.000
C3	TasaContagios ~ V.M.Laboral	-13.157	1.000

Test de Efecto Aleatorio, Correlación Espacial y Correlación Serial en los errores, Segunda Ola

test	variables	LM	p_value
J	TasaContagios ~ V.M.Laboral	910.536	0
C1	TasaContagios ~ V.M.Laboral	15.841	0
C2	TasaContagios ~ V.M.Laboral	50.066	0
C3	TasaContagios ~ V.M.Laboral	-1.474	1

# Apéndice F

## P-valores de los parámetros estimados

### F.1. Considerando el Periodo Total

P-valores de los parámetros estimados de cada modelo, Periodo Total

Modelo	Parametros	Estimate	Std. Error	t-value	Pr(> t )	variables
ols	b0	102.167	2.139	47.772	0.000	TasaContagios ~ V.M.Completa
ols	b1	-112.440	7.531	-14.931	0.000	TasaContagios ~ V.M.Completa
sem	b0	105.291	9.184	11.464	0.000	TasaContagios ~ V.M.Completa
sem	b1	-10.569	5.063	-2.087	0.037	TasaContagios ~ V.M.Completa
sem	rho	0.908	0.006	143.364	0.000	TasaContagios ~ V.M.Completa
sr	b0	109.659	9.278	11.820	0.000	TasaContagios ~ V.M.Completa
sr	b1	-13.622	7.944	-1.715	0.086	TasaContagios ~ V.M.Completa
sr	psi	0.924	0.007	137.460	0.000	TasaContagios ~ V.M.Completa
semsr	b0	106.584	10.953	9.731	0.000	TasaContagios ~ V.M.Completa
semsr	b1	-9.418	8.310	-1.133	0.257	TasaContagios ~ V.M.Completa
semsr	psi	0.781	0.013	61.999	0.000	TasaContagios ~ V.M.Completa
semsr	rho	0.770	0.013	59.210	0.000	TasaContagios ~ V.M.Completa
saremsr	b0	199.543	36.268	5.502	0.000	TasaContagios ~ V.M.Completa
saremsr	b1	-0.378	6.922	-0.055	0.956	TasaContagios ~ V.M.Completa
saremsr	lambda	-0.897	0.034	-26.756	0.000	TasaContagios ~ V.M.Completa
saremsr	psi	0.780	0.012	66.019	0.000	TasaContagios ~ V.M.Completa
saremsr	rho	0.944	0.005	174.537	0.000	TasaContagios ~ V.M.Completa

P-valores de los parámetros estimados de cada modelo, Periodo Total

Modelo	Parametros	Estimate	Std. Error	t-value	Pr(> t )	variables
ols	b0	69.775	4.600	15.168	0.000	TasaContagios ~ V.M.Laboral
ols	b1	-207.310	15.817	-13.107	0.000	TasaContagios ~ V.M.Laboral
sem	b0	137.872	10.290	13.399	0.000	TasaContagios ~ V.M.Laboral
sem	b1	31.671	10.105	3.134	0.002	TasaContagios ~ V.M.Laboral
sem	rho	0.912	0.006	153.000	0.000	TasaContagios ~ V.M.Laboral
sr	b0	97.783	10.340	9.456	0.000	TasaContagios ~ V.M.Laboral
sr	b1	-112.867	18.792	-6.006	0.000	TasaContagios ~ V.M.Laboral
sr	psi	0.925	0.006	146.029	0.000	TasaContagios ~ V.M.Laboral
semsr	b0	123.725	12.927	9.571	0.000	TasaContagios ~ V.M.Laboral
semsr	b1	-4.198	20.443	-0.205	0.837	TasaContagios ~ V.M.Laboral
semsr	psi	0.792	0.012	66.661	0.000	TasaContagios ~ V.M.Laboral
semsr	rho	0.773	0.013	59.081	0.000	TasaContagios ~ V.M.Laboral
saremsr	b0	7.412	2.099	3.531	0.000	TasaContagios ~ V.M.Laboral
saremsr	b1	-11.763	6.672	-1.763	0.078	TasaContagios ~ V.M.Laboral
saremsr	lambda	0.918	0.008	119.100	0.000	TasaContagios ~ V.M.Laboral
saremsr	psi	0.777	0.012	66.089	0.000	TasaContagios ~ V.M.Laboral
saremsr	rho	-0.674	0.040	-16.768	0.000	TasaContagios ~ V.M.Laboral

## F.2. Considerando la Primera Ola

P-valores de los parámetros estimados de cada modelo, Primera Ola

Modelo	Parametros	Estimate	Std. Error	t-value	Pr(> t )	variables
ols	b0	121.842	9.380	12.989	0.000	TasaContagios ~ V.M.Completa
ols	b1	-172.875	31.932	-5.414	0.000	TasaContagios ~ V.M.Completa
sem	b0	170.227	22.910	7.430	0.000	TasaContagios ~ V.M.Completa
sem	b1	42.952	15.848	2.710	0.007	TasaContagios ~ V.M.Completa
sem	rho	0.916	0.010	91.868	0.000	TasaContagios ~ V.M.Completa
sr	b0	88.529	15.040	5.886	0.000	TasaContagios ~ V.M.Completa
sr	b1	-103.710	24.810	-4.180	0.000	TasaContagios ~ V.M.Completa
sr	psi	0.899	0.011	85.268	0.000	TasaContagios ~ V.M.Completa
semsr	b0	124.892	26.477	4.717	0.000	TasaContagios ~ V.M.Completa
semsr	b1	-37.354	21.931	-1.703	0.089	TasaContagios ~ V.M.Completa
semsr	psi	0.788	0.018	43.554	0.000	TasaContagios ~ V.M.Completa
semsr	rho	0.810	0.018	44.710	0.000	TasaContagios ~ V.M.Completa
saremsr	b0	233.336	81.186	2.874	0.004	TasaContagios ~ V.M.Completa
saremsr	b1	-24.027	18.493	-1.299	0.194	TasaContagios ~ V.M.Completa
saremsr	lambda	-0.825	0.059	-14.019	0.000	TasaContagios ~ V.M.Completa
saremsr	psi	0.799	0.017	47.280	0.000	TasaContagios ~ V.M.Completa
saremsr	rho	0.947	0.009	111.033	0.000	TasaContagios ~ V.M.Completa

P-valores de los parámetros estimados de cada modelo, Primera Ola

Modelo	Parametros	Estimate	Std. Error	t-value	Pr(> t )	variables
ols	b0	93.666	13.327	7.029	0.000	TasaContagios ~ V.M.Laboral
ols	b1	-213.942	36.353	-5.885	0.000	TasaContagios ~ V.M.Laboral
sem	b0	201.537	25.370	7.944	0.000	TasaContagios ~ V.M.Laboral
sem	b1	95.453	24.152	3.952	0.000	TasaContagios ~ V.M.Laboral
sem	rho	0.922	0.010	95.655	0.000	TasaContagios ~ V.M.Laboral
sr	b0	75.702	16.415	4.612	0.000	TasaContagios ~ V.M.Laboral
sr	b1	-123.499	29.704	-4.158	0.000	TasaContagios ~ V.M.Laboral
sr	psi	0.899	0.011	84.113	0.000	TasaContagios ~ V.M.Laboral
semsr	b0	135.518	30.628	4.425	0.000	TasaContagios ~ V.M.Laboral
semsr	b1	-1.747	36.470	-0.048	0.962	TasaContagios ~ V.M.Laboral
semsr	psi	0.791	0.018	44.115	0.000	TasaContagios ~ V.M.Laboral
semsr	rho	0.830	0.017	48.271	0.000	TasaContagios ~ V.M.Laboral
saremsr	b0	5.867	3.814	1.538	0.124	TasaContagios ~ V.M.Laboral
saremsr	b1	-9.002	9.336	-0.964	0.335	TasaContagios ~ V.M.Laboral
saremsr	lambda	0.937	0.010	92.200	0.000	TasaContagios ~ V.M.Laboral
saremsr	psi	0.798	0.017	47.075	0.000	TasaContagios ~ V.M.Laboral
saremsr	rho	-0.670	0.066	-10.119	0.000	TasaContagios ~ V.M.Laboral

### F.3. Considerando el Periodo enmtre Olas

P-valores de los parámetros estimados de cada modelo, Periodo entre Olas

Modelo	Parametros	Estimate	Std. Error	t-value	Pr(> t )	variables
ols	b0	56.153	0.921	60.998	0.000	TasaContagios ~ V.M.Completa
ols	b1	-17.746	3.418	-5.192	0.000	TasaContagios ~ V.M.Completa
sem	b0	56.460	1.665	33.917	0.000	TasaContagios ~ V.M.Completa
sem	b1	-16.532	3.780	-4.374	0.000	TasaContagios ~ V.M.Completa
sem	rho	0.535	0.026	20.955	0.000	TasaContagios ~ V.M.Completa
sr	b0	59.339	2.148	27.625	0.000	TasaContagios ~ V.M.Completa
sr	b1	-7.325	4.859	-1.508	0.132	TasaContagios ~ V.M.Completa
sr	psi	0.749	0.017	43.041	0.000	TasaContagios ~ V.M.Completa
semsr	b0	58.674	2.510	23.372	0.000	TasaContagios ~ V.M.Completa
semsr	b1	-6.521	5.411	-1.205	0.228	TasaContagios ~ V.M.Completa
semsr	psi	0.703	0.020	35.597	0.000	TasaContagios ~ V.M.Completa
semsr	rho	0.280	0.035	7.976	0.000	TasaContagios ~ V.M.Completa
saremsr	b0	8.232	0.626	13.145	0.000	TasaContagios ~ V.M.Completa
saremsr	b1	-1.314	2.068	-0.635	0.525	TasaContagios ~ V.M.Completa
saremsr	lambda	0.859	0.019	45.965	0.000	TasaContagios ~ V.M.Completa
saremsr	psi	0.648	0.021	31.522	0.000	TasaContagios ~ V.M.Completa
saremsr	rho	-0.985	0.047	-21.149	0.000	TasaContagios ~ V.M.Completa

P-valores de los parámetros estimados de cada modelo, Periodo entre Olas

Modelo	Parametros	Estimate	Std. Error	t-value	Pr(> t )	variables
ols	b0	49.331	1.772	27.834	0.000	TasaContagios ~ V.M.Laboral
ols	b1	-27.017	7.496	-3.604	0.000	TasaContagios ~ V.M.Laboral
sem	b0	53.926	2.469	21.837	0.000	TasaContagios ~ V.M.Laboral
sem	b1	-11.441	7.922	-1.444	0.149	TasaContagios ~ V.M.Laboral
sem	rho	0.619	0.025	24.550	0.000	TasaContagios ~ V.M.Laboral
sr	b0	55.965	3.898	14.358	0.000	TasaContagios ~ V.M.Laboral
sr	b1	-16.381	14.783	-1.108	0.268	TasaContagios ~ V.M.Laboral
sr	psi	0.767	0.017	45.641	0.000	TasaContagios ~ V.M.Laboral
semsr	b0	56.301	4.132	13.627	0.000	TasaContagios ~ V.M.Laboral
semsr	b1	-12.255	14.848	-0.825	0.409	TasaContagios ~ V.M.Laboral
semsr	psi	0.709	0.020	35.763	0.000	TasaContagios ~ V.M.Laboral
semsr	rho	0.338	0.038	8.833	0.000	TasaContagios ~ V.M.Laboral
saremsr	b0	11.998	1.843	6.510	0.000	TasaContagios ~ V.M.Laboral
saremsr	b1	-10.766	7.630	-1.411	0.158	TasaContagios ~ V.M.Laboral
saremsr	lambda	0.756	0.032	23.627	0.000	TasaContagios ~ V.M.Laboral
saremsr	psi	0.678	0.021	32.969	0.000	TasaContagios ~ V.M.Laboral
saremsr	rho	-0.704	0.070	-9.983	0.000	TasaContagios ~ V.M.Laboral

### F.4. Considerando la Segunda Ola

P-valores de los parámetros estimados de cada modelo, Segunda Ola

Modelo	Parametros	Estimate	Std. Error	t-value	Pr(> t )	variables
ols	b0	207.306	8.507	24.369	0.000	TasaContagios ~ V.M.Laboral
ols	b1	-61.721	32.232	-1.915	0.056	TasaContagios ~ V.M.Laboral
sem	b0	233.606	14.476	16.137	0.000	TasaContagios ~ V.M.Laboral
sem	b1	28.569	27.702	1.031	0.302	TasaContagios ~ V.M.Laboral
sem	rho	0.792	0.027	29.551	0.000	TasaContagios ~ V.M.Laboral
sr	b0	154.642	16.873	9.165	0.000	TasaContagios ~ V.M.Laboral
sr	b1	-176.032	57.874	-3.042	0.002	TasaContagios ~ V.M.Laboral
sr	psi	0.845	0.020	41.980	0.000	TasaContagios ~ V.M.Laboral
semsr	b0	205.812	21.618	9.521	0.000	TasaContagios ~ V.M.Laboral
semsr	b1	-3.012	50.926	-0.059	0.953	TasaContagios ~ V.M.Laboral
semsr	psi	0.724	0.028	26.137	0.000	TasaContagios ~ V.M.Laboral
semsr	rho	0.691	0.038	18.296	0.000	TasaContagios ~ V.M.Laboral
saremsr	b0	25.469	6.695	3.804	0.000	TasaContagios ~ V.M.Laboral
saremsr	b1	-16.317	25.688	-0.635	0.525	TasaContagios ~ V.M.Laboral
saremsr	lambda	0.859	0.032	26.826	0.000	TasaContagios ~ V.M.Laboral
saremsr	psi	0.703	0.029	24.146	0.000	TasaContagios ~ V.M.Laboral
saremsr	rho	-0.525	0.113	-4.639	0.000	TasaContagios ~ V.M.Laboral