

Inequidades socioeconómicas en relación a los  
espacios verdes y cobertura de copa: un  
enfoque espacial aplicado a la ciudad de Cali,  
Colombia.

Juan Fernando Correa Caicedo

*Presentado para obtener al título de Maestría en SIG (M.Sc. UNIGIS)*

*Universidad de Salzburgo*

*2018-01-07*



# Contents

<b>Presentación</b>	<b>15</b>
<b>Agradecimientos</b>	<b>17</b>
<b>Resúmen</b>	<b>19</b>
<b>1 Introducción</b>	<b>21</b>
1.1 Problema de investigación . . . . .	22
1.2 Hipótesis . . . . .	23
1.3 Objetivos . . . . .	23
1.4 Justificación . . . . .	24
<b>2 Revisión de la literatura</b>	<b>25</b>
2.1 Las instituciones ambientales en Santiago de Cali . . . . .	25
2.2 Árboles y desarrollo urbano . . . . .	29
2.3 Servicios ecosistémicos y su valoración . . . . .	30
2.4 La perspectiva de la justicia ambiental . . . . .	36
2.5 Modelamiento y análisis espacial de variables ambientales y sociales . . . . .	39
<b>3 Metodología</b>	<b>53</b>
3.1 Aproximación general . . . . .	53
3.2 Área de estudio . . . . .	53
3.3 Datos . . . . .	57

3.4	Métodos y técnicas . . . . .	66
<b>4</b>	<b>Procesamiento y análisis de datos</b>	<b>69</b>
4.1	Capas de información geográfica . . . . .	71
4.2	Datos del censo arbóreo 2015 . . . . .	89
4.3	Datos del censo de población . . . . .	105
4.4	Análisis estadísticos . . . . .	118
4.5	Análisis geoestadísticos . . . . .	178
<b>5</b>	<b>Discusión</b>	<b>239</b>
5.1	Sobre la cobertura de copa . . . . .	239
5.2	Sobre los espacios verdes . . . . .	250
<b>6</b>	<b>Conclusiones</b>	<b>259</b>

# List of Tables

2.1	Indicadores ambientales urbanos usados en el Expediente Municipal . . . . .	28
2.2	Clasificación de servicios ecosistémicos importantes en zonas urbanas y funciones y componentes subyacentes del ecosistema . . . . .	31
2.3	Métricas para caracterizar servicios del AU . . . . .	40
2.4	Métricas para caracterizar servicios del EV . . . . .	45
2.5	Métricas para caracterizar aspectos socioeconómicos de la población y estructura de las unidades espaciales del EV . . . . .	46
3.1	Variables para caracterizar el AU . . . . .	59
3.2	Muestra de los datos del censo arbóreo . . . . .	60
3.3	Variables sobre la población . . . . .	62
3.4	Variables sobre las viviendas . . . . .	63
4.1	Resumen CA2015 por tipo de vegetación . . . . .	90
4.2	Resumen CA2015 por edad . . . . .	90
4.3	Resumen CA2015 por emplazamiento . . . . .	91
4.4	Resumen selección CA2015 por tipo de vegetación . . . . .	91
4.5	Resumen selección CA2015 por edad . . . . .	92
4.6	Resumen selección CA2015 por emplazamiento . . . . .	92
4.7	Total de población en la ciudad de Cali . . . . .	108
4.8	Resumen ajuste modelos área copa preliminares . . . . .	142
4.9	Resumen ajuste de porcentaje de cobertura de copa preliminares . . . . .	145
4.10	Resumen ajuste OLS: log(area copa) . . . . .	148

4.11	Resumen ajuste OLS: cobertura copa.ap . . . . .	150
4.12	Comparación OLS para el área de copa (AC) con variables de población y con otras dimensiones . . . . .	155
4.13	Comparación OLS para el porcentaje de cobertura de copa con variables de población y otras dimensiones . . . . .	158
4.14	Comparación OLS del modelos con todas las variables y el modelo simplificado del porcentaje de area de espacio verde . . . . .	170
4.15	Comparación OLS de los modelos con todas las variables y el modelo simplificado de índice de acceso ‘ia.areas.dist’ . . . . .	175
4.16	Metricas de ajuste para los modelos de área de copa . . . . .	201
4.17	Metricas de ajuste para los modelos de porcentaje de área de copa . . . . .	209
4.18	Metricas de ajuste para los modelos de porcentaje de área de EV . . . . .	226
4.19	Metricas de ajuste para los modelos de áreas-distancia de EV . . . . .	236
5.1	Coeficientes del modelo SD de área de copa . . . . .	241
5.2	Coeficiente de autocorrelación modelo SD de área de copa . . . . .	241
5.3	Coeficientes del modelo SAR de porcentaje de área de copa . . . . .	241
5.4	Coeficiente de autocorrelación modelo SAR de área de copa . . . . .	242
5.5	Coeficientes del modelo SEM de porcentaje de área de espacio verde . . . . .	251
5.6	Coeficiente de autocorrelación modelo SEM de porcentaje de área de espacio verde . . . . .	251
5.7	Coeficientes del modelo SD del indicador de acceso área-distancia . . . . .	252
5.8	Coeficiente de autocorrelación modelo SD del indicador de acceso área-distancia . . . . .	252

# List of Figures

3.1	Área de estudio . . . . .	54
3.2	Mapa de Santiago de Cali,tomada de Google Maps . . . . .	55
3.3	Vista satelital de Santiago de Cali,tomada de Google Maps . . . . .	56
3.4	Capas usadas para el procesamiento de los espacios verdes y las caraterísticas de las manzanas . . . . .	58
3.5	Árboles seleccionados para el análisis. El tamaño de puntos que representan los arboles es proporcional al area de la copa en unidades del mapa. . . . .	61
3.6	División en barrios y sectores urbanos de Santiago de Cali . . . . .	64
3.7	División en barrios y sectores urbanos de Santiago de Cali . . . . .	65
4.1	Sectores Urbanos del Censo del 2005. Los sectores seleccionados están parcial o totalmente contenidos en el perímetro urbano 2015 . . . . .	72
4.2	Coropleta del tamaño de manzana.Se usaron 10 grupos con aprox. el mismo número de observaciones . . . . .	73
4.3	Espacio verdes consolidados y sectores urbanos . . . . .	74
4.4	Espacio verde por categoría . . . . .	75
4.5	Small Multiple del espacio verde por categoría . . . . .	76
4.6	Small Multiple de los indices de acceso a EV basados en distancia en escala continua . . . . .	79
4.7	Small Multiple de los indices de acceso a EV basados en distancia usando deciles . . . . .	80
4.8	Espacio verdes y rango de 1 km desde centriodes de SU . . . . .	81

4.9	Small Multiple de los indices de acceso a EV basados en área usando escala continua . . . . .	82
4.10	Small Multiple de los indices de acceso a EV basados en distancia usando deciles . . . . .	83
4.11	Small Multiple de los indices de acceso a EV basados en área y distancia usando escala continua . . . . .	84
4.12	Small Multiple de los indices de acceso a EV basados en distancia usando deciles . . . . .	85
4.13	Small Multiple de los indices de acceso a EV basados en distancia usando deciles . . . . .	86
4.14	características por tipo de vegetacion . . . . .	94
4.15	Caraterísticas por tipo de vegetación y emplazamiento . . . . .	95
4.16	Variabilidad del diámetro de copa por emplazamiento . . . . .	96
4.17	Variabilidad de la altura de los arboles por emplazamiento . . . . .	97
4.18	Small multiples de los individuos arbóreos por emplazamiento . . . . .	99
4.19	Suma de cobertura por hexagonos . . . . .	100
4.20	Area de copa por sector censal . . . . .	101
4.21	Agregación de area de copa por hexagonos y SU . . . . .	101
4.22	Agregación de número de árboles por hexagonos y SU . . . . .	102
4.23	Métricas de cobertura de copa: área neta, porcentaje respecto del sector censal y porcentaje respecto del área pública . . . . .	103
4.24	Métricas de cobertura de copa: área neta, porcentaje respecto del sector censal y porcentaje respecto del área pública . . . . .	103
4.25	Histograma de las métricas de la cobertura de copa . . . . .	104
4.26	Mapas de las variables de población seleccionadas (escala contínua) . . . . .	106
4.27	Mapas de las variables de población seleccionadas (en deciles) . . . . .	107
4.28	Histogramas de las variables de población . . . . .	109
4.29	Mapas de las variables de población seleccionadas como porcentajes (escala contínua) . . . . .	110

4.30	Mapas de las variables de población seleccionadas como porcentajes (en deciles) . . . . .	111
4.31	Histogramas de las variables de población como porcentaje . . . . .	112
4.32	Mapas de las variables sobre el tipo de uso de los predios como porcentaje de la cantidad de predios (escala continua) . . . . .	114
4.33	Mapas de las variables sobre el tipo de uso de los predios como porcentaje de la cantidad de predios (en deciles) . . . . .	115
4.34	Histogramas de las variables de uso de predios como porcentaje . . . . .	115
4.35	Mapas de las variables sobre el tipo viviendas como porcentaje (escala continua) . . . . .	116
4.36	Mapas de las variables sobre el tipo viviendas como porcentaje (en deciles) . . . . .	117
4.37	Histogramas de los tipo de vivienda como porcentaje . . . . .	117
4.38	Mapa de los sectores excluidos . . . . .	120
4.39	Mapa de los sectores excluidos por criterio usado . . . . .	121
4.40	Mapa de los sectores excluidos para el análisis de EV . . . . .	122
4.41	Mapa de los sectores excluidos por criterio usado para el análisis de EV . . . . .	123
4.42	Sectores urbanos de las variables dependientes sobre cobertura de copa . . . . .	125
4.43	Comparación por pares entre predictores de población . . . . .	127
4.44	Comparación por pares entre predictores de población porcentuales . . . . .	129
4.45	Comparación por pares entre predictores de población porcentuales y predictores en unidades de personas . . . . .	130
4.46	Coeficiente Pearson entre variables de población . . . . .	131
4.47	Coeficiente Spearman entre variables de población . . . . .	132
4.48	Coeficiente Pearson entre coberturas de copa y variables de población . . . . .	133
4.49	Coeficiente Spearman entre coberturas de copa y variables de población . . . . .	133
4.50	Gráficas de dispersión entre predictores de población y variables dependientes de la cobertura de copa (regresión lineal + coeficiente de Pearson) . . . . .	134

4.51 Gráficas de dispersión entre predictores de población (usando el inverso de la variable con correlaciones negativas: $f(x) = \frac{1}{x+1}$ ) y las variables dependientes de la cobertura de copa ( regresión lineal + coeficiente de Pearson) . . . . .	135
4.52 Histogramas de las variables de cobertura de copa transformadas . . . . .	136
4.53 Gráficas de dispersión entre predictores de población y variables dependientes de la área de copa transformadas (regresión lineal + coeficiente de Pearson) . . . . .	137
4.54 Gráficas de dispersión entre predictores de población y variables dependientes de la cobertura de copa transformadas (regresión lineal + coeficiente de Pearson) . . . . .	137
4.55 Gráficas diagnósticas para el análisis de regresión lineal del área de copa .	147
4.56 Gráficas diagnósticas para el análisis de regresión lineal del porcentaje de cobertura de copa . . . . .	149
4.57 Coeficiente Pearson entre variables de uso de los predios y espacios verdes en los sectores urbanos . . . . .	151
4.58 Coeficiente Pearson entre variables de uso de los predios y espacios verdes en los sectores urbanos . . . . .	152
4.59 Coeficiente Pearson entre coberturas de copa y variables de predios y EV .	153
4.60 Coeficiente Pearson entre coberturas de copa y variables de predios y EV .	154
4.61 Gráficas diagnósticas para el análisis de regresión lineal de área de copa con los nuevos términos . . . . .	156
4.62 Gráficas diagnósticas para el análisis de regresión lineal de porcentaje de cobertura de copa con los nuevos términos . . . . .	159
4.63 Variable dependiente, valor ajustado por el modelo y residuos del OLS para log.area.copa normalizada . . . . .	160
4.64 Variable dependiente, valor ajustado por el modelo y residuos del OLS para cobertura.copa.ap normalizada . . . . .	161
4.65 Metricas de acceso a espacio verdes . . . . .	161

4.66	Metricas de acceso a espacio verdes seleccionadas . . . . .	162
4.67	Coeficiente Pearson entre acceso a EV y variables de población . . . . .	163
4.68	Coeficiente Spearman entre acceso a EV y variables de población . . . . .	164
4.69	Coeficiente Pearson entre acceso a EV y variables de uso de los predios . .	164
4.70	Coeficiente Spearman entre acceso a EV y variables de uso de los predios .	165
4.71	Coeficiente Pearson entre acceso a EV y variables sobre aspectos físicos de las manzanas y SU . . . . .	166
4.72	Coeficiente Spearman entre acceso a EV y variables sobre aspectos físicos de las manzanas y SU . . . . .	166
4.73	Distribución de variables independientes respecto de los índices de acceso a EV . . . . .	167
4.74	Gráficas diagnósticas para el análisis de regresión lineal de porcentaje de area de espacio verde . . . . .	171
4.75	Variable dependiente, valor ajustado por el modelo y residuos del OLS para 'area ep.porcentaje' normalizada . . . . .	172
4.76	Gráficas diagnósticas para el análisis de regresión lineal de 'ia.areas.dist' .	176
4.77	Variable dependiente, valor ajustado por el modelo y residuos del OLS para 'ia.areas.dist' normalizada . . . . .	177
4.78	Grafo de vecindad entre todos los SU de la ciudad de Cali . . . . .	179
4.79	Grafo de vecindad entre los SU seleccionados para el análisis . . . . .	180
4.80	Matrices de vecindad del análisis espacial . . . . .	181
4.81	Gráfico de Moran para los residuos del modelo lineal para del área de cobertura de copa usando $W_q$ . . . . .	185
4.82	Mapas LISA para la matriz $W_q$ de 'log.area copa' . . . . .	185
4.83	Mapas LISA para la matriz $W_d$ de 'log.area copa' . . . . .	186
4.84	Gráfico de Moran para los residuos del modelo lineal para el porcentajes cobertura de copa usando $W_q$ . . . . .	187
4.85	Mapas LISA para la matriz $W_q$ de 'cobertura copa.ap' . . . . .	188
4.86	Mapas LISA para la matriz $W_d$ de 'cobertura copa.ap'	189

4.87 Gráfico de Moran para los residuos del modelo lineal para del área de cobertura de copa usando $W_q$ y $W_d$ . . . . .	190
4.88 Mapas LISA para la matriz $W_q$ de los residuos del modelo lineal para del área de copa . . . . .	191
4.89 Mapas LISA para la matriz $W_d$ dende los residuos del modelo lineal para del área de copa . . . . .	191
4.90 Mapas LISA para la matriz $W_q$ de los residuos del modelo lineal para el porcentaje de cobertura copa . . . . .	194
4.91 Mapas LISA para la matriz $W_d$ dende los residuos del modelo lineal para del área de copa . . . . .	194
4.92 Gráfico de Moran para los residuos de los modelos espaciales de área de copa usando $W_q$ . . . . .	203
4.93 Diagnóstico comparativo entre modelos . . . . .	204
4.94 Gráfico de Moran para los residuos de los modelos espaciales del porcenatje de área de copa usando $W_q$ . . . . .	210
4.95 Diagnóstico comparativo entre modelos de porcentaje de copa . . . . .	211
4.96 Matrices de vecindad del análisis espacial de espacios verdes . . . . .	213
4.97 Mapas LISA para la matriz $W_q$ de ambos indicadores de acceso a EV . . . . .	217
4.98 Gráfico de Moran para los residuos del modelo lineal para el porcentaje de área de espacio verde . . . . .	218
4.99 Gráfico de Moran para los residuos del modelo lineal para el indicador ‘ia.areas.dist’ . . . . .	220
4.100 Mapas LISA para la matriz $W_q$ de los residuos de los modelos lineales para los indicadores de acceso a EV . . . . .	221
4.101 Gráfico de Moran para los residuos de los modelos espaciales de porcentaje de área de EV $W_q$ . . . . .	227
4.102 Diagnóstico comparativo entre modelos . . . . .	228
4.103 Gráfico de Moran para los residuos de los modelos espaciales de acceso área-distancia $W_q$ . . . . .	237

4.104	Diagnóstico comparativo entre modelos . . . . .	238
5.1	Mapas LISA - Porcentaje población Afro . . . . .	242
5.2	Mapas LISA - Porcentaje población sin estudios . . . . .	243
5.3	Mapas LISA - Porcentaje población con estudios superiores . . . . .	244
5.4	Mapas LISA - Porcentaje cobertura de copa . . . . .	245
5.5	Mapas LISA - Área de cobertura de copa . . . . .	246
5.6	Mapas LISA - Densidad de Población . . . . .	247
5.7	Mapas LISA - Porcentaje de viviendas tipo cuarto . . . . .	248
5.8	Mapas LISA - Porcentaje de espacios verdes . . . . .	249
5.9	Distribucion de lo indicadores de acceso local a EV . . . . .	253
5.10	Distribucion de lo indicadores de acceso a EV área-distancia . . . . .	255
5.11	Mapas LISA - Porcentaje de espacios verdes y área media de manzana . .	256
5.12	Mapas LISA - Porcentaje de unidades economicas y personas con alguna limitación . . . . .	257



# Presentación

Borrador del texto. Tesis para obtener el título de master en Geographical Information Science & Systems (UNIGIS M.Sc.).



# Agradecimientos

A Elizabeth, Augusto, Ana María y Catalina. Este trabajo fue posible gracias a su ayuda y cariño.



# Resumen

Este proyecto hace uso del censo arbóreo urbano realizado en Santiago de Cali(Colombia) entre el año 2014 y 2015, de datos del censo de población de 2005 y los datos de la estructura ecológica del municipio. El propósito es identificar si existen patrones espaciales estadísticamente significativos que muestren una relación negativa entre las coberturas de copa, el acceso a espacios verdes y zonas con población en condiciones socioeconómicas de vulnerabilidad, señalando espacialmente inequidades socioeconómicas en el acceso a servicios ambientales. Para lograrlo se hace uso de modelos de regresión espacial que capturan fenómenos de derrame, o agrupamiento y dispersión a través la inclusión de términos de autocorrelación espacial en la regresión lineal. Una matriz de vecindad determina la influencia espacial en los términos autoregresivos, de retardo espacial o en el error. Al encontrar evidencia de autocorrelación espacial en los residuos de la regresión lineal se prueban los modelos espaciales de regresión usando dos tipos de matrices distintas para observar el efecto de la topología de interacción entre las variables y los resultados del ajuste. Se apoya los análisis con gráficos estadísticos, mapas de temáticos y de LISA para la identificación de zonas con niveles en el indicador espacialmente relacionados en el modelo. Esto permite tener coeficientes más confiables para determinar la importancia de una variable dentro de un modelo predictivo o explicativo del acceso a un beneficio o servicio.

Los resultados muestran que existen inequidades explicadas con variables de estatus como el acceso a educación superior, que están negativamente correlacionadas con el porcentaje de afrocolombianos y la concentración de sectores urbanos (SU) con mejor nivel de porcentaje cobertura de copa en el área total del SU y el porcentaje de cobertura de copa

en el área pública del SU. En relación a los espacio verdes (EV), se identifican patrones espaciales de las personas con limitaciones que son coincidentes con una distribución baja de acceso a EV representada por un indicador de acceso que relaciona área y distancia, en busca de capturar el disfrute posible de EV al efectuar desplazamiento a otros SU. El indicador de porcentaje de área de EV de un SU se aproxima mejor a una idea local del disfrute, en donde los beneficios son para quien mora dentro del SU. El área media de manzana es el predictor con el coeficiente más alto en los modelos de acceso a EV, lo que invita a pensar que un nivel alto de nivel de acceso se puede deber a tener manzanas más grandes con andenes más amplios o parque más grandes, característica determinada por los urbanizadores y las autoridades locales.

# Chapter 1

## Introducción

En las agendas municipales, a nivel mundial, ha crecido la importancia de las relaciones y patrones de distribución de los beneficios de áreas verdes en comparación con la distribución espacial de variables sociales y económicas como el ingreso, acceso al trabajo, la etnicidad o el género, con miras a reducir las desigualdades entre los ciudadanos en el acceso y disfrute de los servicios ambientales. La definición y valoración de estos beneficios hace uso de medidas como la abundancia, la cobertura de las copas de los árboles, índices de vegetación, diferentes distancias a las zonas verdes dentro de un marco alineado con conceptos como la justicia ambiental, equidad y la sostenibilidad. Estos indicadores son calculados con datos de censos de población, encuesta de calidad de vida, censos arbóreos, imágenes satelitales, cartografías y bases de datos de entidades oficiales y académicas. Los beneficios tienen cargas y costos de mantenimiento para la administración y gestión de los recursos y servicios ambientales, lo que exige que se identifique las zonas, condiciones de los recursos y de la población para la ejecución de acciones eficaces y eficientes por parte de los gobiernos y autoridades ambientales.

Este proyecto hace uso del censo arbóreo urbano realizado en Santiago de Cali (Colombia) entre el año 2014 y 2015, de datos del censo de población de 2005 y los datos de la estructura ecológica del municipio. El propósito es motivar el desarrollo de políticas, proyectos

e instrumentos que potencien al árbol y el acceso a espacio verdes como estrategia para la mejora de la calidad de vida de los ciudadanos. Para ello se plantea identificar si existen patrones espaciales estadísticamente significativos que muestre una relación negativa entre las coberturas de copa, el acceso a espacios verdes y zonas con población en condiciones socioeconómicas de vulnerabilidad. El proyecto se apoya en los aportes de tipo metodológico y estadístico de la literatura especializada sobre modelos de regresión lineal y modelos de regresión espacial. Estos métodos son enriquecidos con la exploración visual de los datos.

## 1.1 Problema de investigación

La distribución espacial equitativa de los beneficios que proveen el arbolado urbano y las zonas verdes de espacio público, que constituyen un bien común, financiado y de responsabilidad de las administraciones municipales<sup>1</sup>, es un componente cuantificable por medio métodos estadístico y técnicas de análisis espacial con miras a construir evidencia sobre el disfrute y acceso a los beneficios ambientales en espacios urbanos y su relación con las condiciones de vida de la población que habita ese territorio.

Este trabajo pretende enfocarse en describir los patrones espaciales y establecer la correlación entre métricas para representar los beneficios del arbolado urbano(AU) y los espacios verdes (EV) con las variables sociales y económicas de la población. El valor explicativo de la posición en el plano geográfico sobre las métricas ambientales y socioeconómicas permite seleccionar los modelos de regresión apropiados para cuantificar el grado de correlación que existe (Fotheringham et al., 1998). El problema comprende la exploración de las variables sociales y ambientales, el cálculo de indicadores, identificación de los valores de referencia para la interpretación de las cifras, la cuantificación de la correlación, identificar las zonas con acumulación de desigualdades y la presentación visual de la información consolidada para la discusión.

---

<sup>1</sup> Así está expresado en las leyes ambientales que dan forma al Sistema Nacional Ambiental, SINA, (LEY 99 DE 1993) y reglamentan los planes de desarrollo y de ordenamiento territorial (LEY 388 DE 1997), así como la creación de organismos en la estructura municipal (Acuerdo 01 de 1996).

Algunas de las preguntas a las que se enfrenta esta investigación son: ¿a parte de la cobertura de las copas qué otras variables biofísicas, por ejemplo la edad, la especie(frutales), o la diversidad de especies arbóreas, que expresan beneficios, pueden incluirse en el análisis de correlación? ¿Cuáles son las zonas que muestran mayores correlaciones negativas entre las variables sociales y la cobertura de copas o el acceso a zonas verdes? ¿Es igual tener acceso a un parque pequeño que a uno grande? ¿ Existen aspectos de tipo urbanístico y de la estructura de los barrios que condicionan el desarrollo del arbolado? ¿Están estos factores de estructura y uso relacionados con las condiciones de la población? ¿Qué tipo de modelos son los más apropiados para capturar la dependencia espacial en los datos?

## **1.2 Hipótesis**

Las zonas urbanas con más desventajas sociales y económicas tienen menos acceso a espacios verdes y a los beneficios del arbolado urbano.

## **1.3 Objetivos**

### **1.3.1 Objetivo general**

Identificar y analizar espacialmente inequidades socioeconómicas en relación a los espacios verdes y el arbolado en la zona urbana de Cali.

### **1.3.2 Objetivos específicos**

- Generar métricas de cobertura de espacios verdes y arbolado urbano.
- Identificar y caracterizar las variables sociales y económicas para ser relacionadas con acceso a espacios verdes y cobertura.

- Evaluar las relaciones entre los diferentes indicadores ambientales, sociales y económicos.

## 1.4 Justificación

La inclusión de los componentes ambientales en el ámbito de la planificación urbana en los planes de ordenamiento territorial que exige la legislación colombiana<sup>2</sup> necesita de la creación de medidas y la elaboración de análisis sobre su relación con las condiciones de vida de la población. Las herramientas para establecer políticas públicas y el seguimiento a las acciones realizadas por las administraciones municipales deben estar asociadas a características medibles y objetivas para su implementación. Se espera que los objetivos y proyectos estén sustentados en estudios científicos que identifiquen brechas y oportunidades para la intervención y mejoramiento de los servicios ambientales de los cuales es responsable el gobierno local. En esta medida este estudio contribuye a la identificación de relaciones de inequidad en la distribución de los beneficios que provee el arbolado urbano a través del análisis espacial de la cobertura arbórea, el acceso a zonas verdes y la distribución de las variables sociales y económicas de la población. El estudio promete ser un punto de partida para la identificación de zonas de intervención del arbolado con el fin de cerrar brechas relacionadas con el desarrollo sostenible y la justicia ambiental.

Ante la oportunidad que brinda contar con los datos del censo arbóreo de Santiago de Cali, se abre la posibilidad de hacer análisis de estos beneficios ambientales para la población usando datos con alta resolución espacial que permiten hacer inferencia estadística a escalas apropiadas para la intervención y aprovechamiento de los recursos naturales de la ciudad, y que explotan el potencial que ofrece la información censal y los conjuntos de datos espaciales de los que dispone la administración municipal (Schwarz et al., 2015). Los resultados de esta investigación buscan aportar al debate académico y enriquecer el proceso de la toma de decisiones y la planificación de la ciudad.

---

<sup>2</sup>LEY 388 DE 1997

# **Chapter 2**

## **Revisión de la literatura**

### **2.1 Las instituciones ambientales en Santiago de Cali**

Los árboles son pieza clave de los ecosistemas donde la vida humana ha prosperado. Son hogar y fuente de alimento de muchas especies (Osorio and Molina, 2009); forman espacios con condiciones climáticas y funcionales que complejizan el paisaje y las posibles relaciones entre los animales (Chapman and Onderdonk, 1998). Se puede ubicar la década de los 70 el inicio de un pensamiento ambiental que empieza a ser relevante en el discurso económico mundial (Leff, 2012) y que se consolida con la publicación del Informe Brundtland en 1987. Sin embargo, las preocupaciones sobre la sostenibilidad y conservación de los ecosistemas que sustentan la vida en el planeta son relativamente nuevas en la economía mundial, si la comparamos con la simbiosis entre árboles y humanos que dan origen a la agricultura y lo convierten en un elemento simbólico de gran riqueza en el universo religioso y cultural de la humanidad (León Calle and others, 2011).

En Colombia fue el Instituto Nacional de los Recursos Naturales Renovables y del Ambiente (INDERENA) el pionero de la gestión ambiental, fundado en 1968 y adscrito al ministerio de Agricultura. En 1993 con la ley 99 se crea el Ministerio del Medio Ambiente (MinAmbiente), reemplazando al INDERENA que había cumplido su ciclo de vida, dando

paso a esta nueva entidad fortalecida económicamente y en sintonía con la dinámica ecológica mundial (Becerra, 2012). Su homólogo municipal, el Departamento Administrativo de Gestión del Medio Ambiente (DAGMA), entidad responsable de la gestión del medio ambiente y de los recursos naturales, de la política y de la acción ambiental en la ciudad de Cali, fue creada 1996 (Acuerdo 01 de 1996).

Los ambientes naturales y estructuras ambientales de las ciudades dependen fuertemente los proyecto de urbanización y de las medidas que tomen las autoridades locales para mantenerlos y promoverlos(Konijnendijk et al., 2005; Mincey et al., 2013). Investigaciones como (Dobbs et al., 2011; Ponce Donoso and Piedrahita, 2009) expresan la importancia conformar esquemas y construir modelos de valoración y gestión de este componente del paisaje urbano. A pesar del reconocimiento de los ecosistemas urbanos en el discurso de las instituciones a nivel municipal desde hace un poco más dos décadas en Colombia, los avances en la consolidación de políticas, instrumentos de seguimiento y acciones eficaces para la gestión y de los bosques urbanos y los espacios verdes tienen grandes falencias técnicas. Por lo tanto, es clave que las autoridades establezcan directrices que orienten a los diferentes entes territoriales sobre el adecuado manejo del arbolado urbano, y avances en la consolidación de sistema de información y monitoreo (Tovar-Corso, 2013).

En Cali, el progreso en este tema ha sido lento, pero existe.

Con la formulación del POT2014 se estructuró un análisis que se compone de un marco legislativo, teórico, conceptual y la definición de los objetos geográficos y zonas que componen la estructura ecológica del municipio. En el informe técnico del POT2014 se evidencia las crecientes dificultades ambientales que presenta el municipio, en particular la zona urbana, en la calidad del aire<sup>1</sup>, el estado de las cuencas hidrográficas, contaminación auditiva<sup>2</sup>, entre otros. El agravante de esta situación es “*la inexistencia de un sistema de información y monitoreo, observatorio ambiental o sistema municipal de indicadores ambientales, aso-*

<sup>1</sup> “Cali es una de las capitales más contaminadas de Colombia, por debajo de Bogotá, igualando a Medellín y superando a Barranquilla, Cartagena y Pereira, según el CONPES.” noa (b), pp. 213

<sup>2</sup> “La contaminación por ruido es un factor crítico en la ciudad. En todas las comunas de Cali, los niveles de ruido superan los 70 decibeles, que están por encima de los estándares permisibles fijados por la normatividad ambiental nacional vigente y la OMS (60dB)” noa (b), pp. 215

*ciado al Expediente Municipal (EM) ha contribuido a la deficiente gobernabilidad sobre la base ecosistémica, el deterioro del sistema ambiental, el aumento en los niveles de riesgo, y la degradación de los servicios ambientales.” (noa, b, pp. 233)*

El POT2014 construye un análisis y consolida datos que sustentan y proponen directrices sobre la estructura ecológica principal (EEP) del municipio<sup>3</sup>, contenida en su mayoría por la parte rural del municipio, pero poco hay de concreto en cifras sobre la línea de base del arbolado urbano y sobre la estructura ecológica complementaria (EEC)<sup>4</sup> que incluye las zonas verdes, cinturones ecológicos y los árboles de calle. Justamente es el POT2014 quien en concordancia con el Estatuto Arbóreo de 2013 (noa, 2013) recomienda la realización de un censo arbóreo y un diagnóstico de base para el plan de silvicultura urbana.

El censo arbóreo fue entregado en 2015 (CA2015). Gracias a ello el Expediente Municipal (EM), instrumento encargado del seguimiento al POT, publicó avances en la consolidación de datos de medición en temas ambientales con indicadores para el área urbana agregados principalmente por comuna (ver tabla 2.1), al igual que en el informe que acompaña el CA2015 (noa, 2015).

La comuna es la unidad administrativa en la cual se subdivide el área urbana de una ciudad media o principal en Colombia, agrupa barrios.

---

<sup>3</sup> “La Estructura Ecológica Principal del municipio de Santiago de Cali está compuesta por las áreas de protección y conservación ambiental en los términos del artículo 4 del Decreto 3600 del 2007 y del artículo 35 de la Ley 388 de 1997 y se constituye como suelo de protección ambiental<sup>331</sup>, y como determinante ambiental del ordenamiento territorial de conformidad con lo establecido en el artículo 15 de la misma ley. A esta categoría pertenecen los elementos del territorio que mantienen características de composición, estructura y/o función que les permite prestar servicios ecológicos y/o ambientales y cuyas características permiten mantener los procesos ecológicos esenciales del territorio convirtiéndose en el principal componente de la Estructura Ecológica Municipal.” noa (b), pp. 715

<sup>4</sup> “El segundo componente de la Estructura Ecológica Municipal es la Estructura Ecológica Complementaria (EEC), y corresponde a los elementos del territorio construido que, en su mayoría, hacen parte de los sistemas estructurantes del territorio y que a pesar de su transformación cultural y de no ser de carácter estrictamente natural, mantienen características, de composición, estructura y/o función que les permite seguir prestando servicios ecológicos y/o ambientales. Las características de estos elementos les permiten fortalecer y apoyar la definición de la Estructura Ecológica Municipal y la conservación de los recursos naturales. Las áreas que pertenecen a la Estructura Ecológica Complementaria no constituyen suelo de protección, sin embargo deben mantener y/o mejorar sus características ambientales para cumplir con su aporte a la estructura ecológica del municipio.” noa (b), pp. 715

Table 2.1: Indicadores ambientales urbanos usados en el Expediente Municipal

Indicador
Número de árboles en zona urbana por comuna.
Área de copa de árboles en zona urbana por comuna.
Área de copa de árboles en la EEC en zona urbana por comuna.
Árboles por habitante en zona urbana por comuna.
Vitalidad de los árboles en zona urbana por comuna.
Árboles sembrados por comuna en el marco del proyecto 1865 de la CVC.

La Organización Mundial de la Salud (OMS) recomienda que una ciudad tenga un árbol por cada tres habitantes, incluyendo espacios públicos y privados. Cali debería tener una población arbórea cercana a los 800.000 árboles, pero de acuerdo con los resultados del censo arbóreo de 2014–2015, se han contabilizado cerca de 290.000 árboles solo en el espacio público<sup>5</sup>. La siembra de 20.000 árboles en la zona urbana, en las comunas que muestran temperaturas por muy por encima de la media (las comunas 3, 4, 5, 8 y 13) propuesta en (CIAT, 2015) contrasta con los 100.000 árboles que tiene como meta en Plan de Desarrollo Municipal 2016-2019; ambas cifras están lejos de impactar el déficit arbóreo de 500.000 individuos; urge la reforestación, “*sin embargo, la ubicación de los sitios de siembra ha sido uno de los principales inconvenientes del proyecto, ya que no existen zonas de espacio público adecuadas o disponibles para tal fin.*” (CIAT, 2015, p. 19)

El otro instrumento consignado en el Estatuto Arbóreo de 2013 (noa, 2013) para la reforestación es el esquema de compensación por intervención del arbolado urbano, que tiene un método de cálculo para la compensación por tala, traslado e intervenciones, pero no dice como se ubicaran las zonas de siembra y compensación. La tala, poda de árboles y el corte de pastos de las zonas verdes son en gran medida las actividades de gestión

---

<sup>5</sup>Suele ser fina la diferenciación de este término entre los urbanistas.

que ha realizado históricamente el DAGMA; las labores de fumigación y mantenimiento preventivo como el control de plagas o fertilización son actividades esporádicas, escasas y contingentes.

## 2.2 Árboles y desarrollo urbano

El hombre ha materializado espacios urbanos con dimensiones que retan la imaginación y llevan al límite los sistemas de infraestructura, abastecimiento y gobernabilidad. En la empresa de consolidar antroposferas, las ciudades que construimos han desplazados muchos de los ecosistemas naturales de los territorios que fueron la razón de escoger justamente esos sitios para el asentamiento, trasladándolos más allá de los límites de la ciudad, atenuando su presencia/visibilidad en el mundo de los ciudadanos. Son reemplazados por vías, zonas verdes, áreas industriales, comerciales y residenciales(Azócar et al., 2007).

El desarrollo urbano de Cali se fue dando en forma de barrios. Cali entra al siglo XX con unos de 25000 habitantes y una 100 hectáreas (Vásquez, 1990); en 2005 tiene una población 2200000 habitantes según el Censo de Población del 2005 y una área urbana 12.101.326 hectáreas. Durante todo este siglo experimenta un rápido crecimiento demográfico, rápida expansión y crecimiento de la vivienda por las migraciones, causadas en gran medida por los fenómeno de desplazamiento por conflictos en las áreas rurales y el fortalecimiento de la industria y el empleo en la ciudad (Vásquez, 1990). Los barrios se desarrollaron no sólo como unidades de crecimiento urbano con características físicas diferenciadas por los diseños de las constructoras según el tipo de viviendas y las capacidades económicas de la población; la ciudad creció con la especulación de la tierra de las familias adineradas de la región que fueron parcelando sus latifundios para su urbanización.

## **2.3 Servicios ecosistémicos y su valoración**

Las preocupaciones sobre el crecimiento de la población mundial, su concentración en centros urbanos y las transformaciones que trae consigo el proceso de cambio climático, nos obligan a pensar en cómo maximizar los beneficios que nos brindan las zonas verdes y el bosque urbano como estrategia para mitigar los efectos negativos de estos procesos (Nesbitt and Meitner, 2016; Laredo and Mirtha, 2011). A este escenario se suma trabajos como (Nowak and Greenfield, 2012) que revelan patrones de decaimiento estadísticamente significativos del arbolado urbano en 17 de 20 ciudades norteamericanas o (Restrepo et al., 2015) que reporta la reducción de las condiciones de vitalidad del arbolado del Valle de Aburra en Colombia derivado de la interacción de causas naturales y antrópicas, las cuales afectan directa o indirectamente la fisiología y salud de los árboles en los espacios urbanos. Nowak and Greenfield (2012) se pregunta si los administradores locales conocen los cambios que presenta las coberturas arbóreas, puesto que esta es una representación simple pero confiable (y ampliamente aceptada) para tasar la extensión de los beneficios derivados de los bosque urbanos, dado que los servicio que proveen los árboles están relacionados con la salud y el funcionamiento de sus hojas.

Podemos definir un servicio ambiental o ecosistémico como los beneficios para la población humana que se derivan directa o indirectamente de los ecosistemas (Bolund and Hunhammar, 1999). Los servicios dependen entonces de los tipos de los ecosistemas con los que cuente el entorno urbano. Bolund and Hunhammar (1999) distinguen 7 tipos de ecosistemas urbanos: árboles de calle, zonas verdes, bosques urbanos, tierras cultivadas, humedales, lagos/lagunas, y ríos/arroyos. Todos ellos en conjunto benefician a la población, y muchos estudios se han encargado de cuantificar el impacto de estos beneficios, en particular se cuentan los que se relacionan con los ecosistemas de árboles de calle y las zonas verdes, como reducción de las temperaturas (Ripoll et al., 2010), reducción de la polución en el aire (Durán Rivera and Alzate Guarín, 2009), secuestro de carbono (Nowak and Crane, 2002; McPherson et al., 2013) , mitigación de los efectos de calentamiento por gases de efecto invernadero (Laredo and Mirtha, 2011) y mantenimiento del agua en los

ecosistemas o proveyendo alimento como en el caso de los árboles frutales (Konijnendijk et al., 2005; Nolazco, 2012), reducción de los niveles de ruido (Bolund and Hunhammar, 1999). Otros estudios argumentan que los beneficios ambientales de los ecosistemas urbanos pueden medirse directamente en la salud de la población (Bolund and Hunhammar, 1999; Gómez-Baggethun and Barton, 2013). Una forma de clasificar todos estos beneficios e indagar sobre las medidas usadas las resume Gómez-Baggethun and Barton (2013) en una tabla que reproduczo a continuación (ver tabla 2.2).

Table 2.2: Clasificación de servicios ecosistémicos importantes en zonas urbanas y funciones y componentes subyacentes del ecosistema

Funciones y componentes	Servicio Ecosistémico	Ejemplo	Ejemplo de indicadores/proxies
Conversión de energía en plantas comestibles a través de la fotosíntesis	Suministro de alimentos	Hortalizas producidas por lotes urbanos y áreas periurbanas	Producción de alimentos [toneladas/año]
Percolación y regulación de la escorrentía y la descarga del río	Regulación del caudal de agua y mitigación de escorrentía	El suelo y la vegetación percolan el agua durante eventos de precipitación intensa y / o prolongada	Capacidad de infiltración del suelo; [%] Sellado con respecto a la superficie permeable [ha]

Funciones y componentes	Servicio Ecosistémico	Ejemplo	Ejemplo de indicadores/proxies
Fotosíntesis, sombreado y evapotranspiración	Regulación urbana de la temperatura	Los árboles y otra vegetación urbana proporcionan sombra, crean humedad y bloquean el viento	Índice de área foliar; Disminución de la temperatura [° C] por cobertura arbórea[m <sup>2</sup> ] en parcelas cubierta de árboles
Absorción de ondas sonoras por la vegetación y el agua	Reducción de ruido	Absorción de ondas sonoras por barreras vegetales, especialmente vegetación espesa	Superficie de la hoja [m <sup>2</sup> ] y distancia a las carreteras [m]; Reducción de ruido [dBA] / unidad de vegetación [m]
Filtración y fijación de gases y partículas	Purificación de aire	Eliminación y fijación de contaminantes por la vegetación urbana en hojas, tallos y raíces	O <sub>3</sub> , SO <sub>2</sub> , NO <sub>2</sub> , CO y PM10 m removido en [toneladas/año] multiplicado por la cobertura arbórea (m <sup>2</sup> )

Funciones y componentes	Servicio Ecosistémico	Ejemplo	Ejemplo de indicadores/proxies
Barrera física y absorción en energía cinética	Moderación de los extremos ambientales	Tormentas, inundaciones y amortiguación de olas por barreras vegetales;	Cubrir la densidad de las barreras de vegetación que separan las áreas construidas del mar
Absorción de calor durante olas de calor severas			
Eliminación o descomposición de nutrientes xénicos	Tratamiento de desechos	Filtración de efluentes y fijación de nutrientes por humedales urbanos	P, K, Mg y Ca en mgkg <sup>-1</sup> en comparación con las normas de calidad del suelo y del agua
Secuestro y fijación de carbono en la fotosíntesis	Regulación climática	Secuestro y almacenamiento de carbono por la biomasa de arbustos urbanos y de tres	Secuestro de CO <sub>2</sub> por árboles (carbono multiplicado por 3.67 para convertir a CO <sub>2</sub> )

Funciones y componentes	Servicio Ecosistémico	Ejemplo	Ejemplo de indicadores/proxies
Movimiento de los gametos florales por la biota	Polinización y dispersión de semillas	El ecosistema urbano provee hábitat para aves, insectos y polinizadores	Diversidad de especies y abundancia de aves y abejorros
Ecosistemas con valores recreativos y educativos	Recreación y desarrollo cognitivo	Los parques urbanos ofrecen múltiples oportunidades para la recreación, la meditación y la pedagogía	Superficie de los espacios públicos verdes [ha / habitante (o cada 1000 habitantes)]
Disposición del hábitat para las especies animales	Avistamiento de animales	El espacio verde urbano proporciona un hábitat para las aves y otros animales a los que les gusta ver	Abundancia de aves, mariposas y otros animales valorados por sus atributos estéticos

Existen diferentes perspectivas para la evaluación de los servicios ecosistémicos, en razón de ellos se crean diferentes indicadores, métodos estadísticos y metodologías para capturar de forma directa o indirecta los beneficios ambientales. (Gómez-Baggethun and Barton, 2013) propone 2 tipos de valoración: la económica y la sociocultural. La valoración sociocultural se enfoca en la percepción y preferencias de los ciudadanos que están ligadas

a sus costumbres y sistemas de valores. Resalta lo difícil de medir y que suele ser mejor abordado por instrumentos cualitativos, construcción de escalas y el uso de narrativas. Un ejemplo de este tipo de trabajos es (Garzón et al., 2004) se hace referencia a la vegetación en las ciudades y su incidencia en la vida de las personas, sobre todo en aquellas comunidades de menores recursos. “Los árboles están estrechamente vinculados a la historia de la humanidad; están cargados de historia, son testigos de varias generaciones, que enlazan su existencia con la anterior, constituyéndose en documentos vivos de acontecimientos culturales.” (Ferro Medina, 2010). (Konijnendijk et al., 2005) propone además que se valore lo ambiental, la biodiversidad y sostenibilidad por su carácter fundamental en la existencia misma del los ecosistemas.

Las metodologías de valoración económica se basan principalmente en análisis costo beneficios(ACB) que buscan un aprovechamiento eficiente del uso del suelo (Bolund and Hunsaker, 1999). La inversión en el arbolado urbano y las zonas verdes arroja resultados positivos consistentemente en la literatura p.e (McPherson et al., 1997) reporta que los beneficios exceden los costos en ciudades como Chicago y en Adelaide (Australia) según (Killicoat et al., 2002) cada árbol da beneficios por AUD\$172.

Algunos análisis no se limitan a evaluar el costo de ahorro o el dispuesto a pagar por la población, para valorar económicamente los servicios ecosistémicos se basan en modelos biofísicos de los individuos arbóreos e incluyen variables ambientales, características de los suelos, la infraestructura de las zonas y sus habitantes (Nelson et al., 2009) p.e el modelo UFORE (Nowak and Crane, 2000) o CBAT (McPherson et al., 1997) , y que permiten tomar acciones específicas sobre el tipo de vegetación y su distribución. Sin embargo es importante tener en cuenta que este tipo de análisis pueden generar incentivos para la conversión indeseable de ecosistemas urbanos en infraestructura construida, con la consiguiente pérdida de servicios de los ecosistemas (Gómez-Baggethun and Barton, 2013).

La lógica económica de los servicios ecosistémicos puede conducir también a incentivar procesos paradójicos, como el incrementos de los precios de las casas y arriendos que

derivan en procesos de gentrificación y desplazamiento de la población que fue beneficiada por la estrategias de mejoramiento de EV y AU con el propósito de resolver problemas relacionados con la justicia ambiental. (Wolch et al., 2014)

## 2.4 La perspectiva de la justicia ambiental

La justicia ambiental es un concepto que ha evolucionado desde su aparición en la década de los 1980 a través de organizaciones dedicadas al activismo ambiental y las redes que ellas forman en conjunto con la academia, acuñaron conceptos de ecología política como justicia ambiental, deuda ecológica, epidemiología ambiental, racismo ambiental, justicia climática, soberanía alimentaria, y responsabilidad ambiental empresarial, que han sido adoptados también por académicos y por tomadores de decisiones (Martinez Alier et al., 2014; Cerdà, 2011).

La planeación urbana, las relaciones entre los ciudadanos con los espacios públicos y con las instituciones que los rigen son la base de la idea de justicia ambiental propuesta en (Low, 2013), y que tiene 3 componentes que la definen: *i*) la justicia distributiva, que en términos de espacio público se basa asegurar disponibilidad y acceso equitativo de los espacios y servicios a los ciudadanos; *ii*) la justicia procedural o procesal, que se refieren a los procesos de negociación y toma de decisiones, en concreto a la percepción de los individuos sobre qué tan justos y equilibrados son, y por tal motivo más dispuesto a aceptar los resultados aunque no les favorezca; *iii*) la justicia interaccional, que refiere al comportamiento y trato de la personas en el espacio público que configuren comportamiento violento o discriminatorios sobre grupos de la población. El autor argumenta que las condiciones ambientales provocadas por los procesos de urbanización y/o contaminación deben analizarse pensando en satisfacer las tres dimensiones, de lo contrario no es posible hablar de justicia. (Schlosberg, 2013) lleva la reflexión un poco más lejos, argumentando que las nuevas extensiones de la justicia ambiental se han movido del discurso a un nuevo dominio, donde lo natural y ambiental crean las condiciones para la justicia social.

En (Braverman, 2008) se exploran las implicaciones entre las intervenciones en el arbolado urbano y el control de fenómenos como el crimen y la gobernabilidad dada las relaciones afectivas y morales de la población con los árboles. El uso de los árboles puede verse también una forma de discriminación y de discurso político o tecnología de gobierno. El uso de zonas verdes y arboles ha sido usado también como una forma de simbolica de estatus y de poder, y esta afirmación es consistente que la tendencia en tener distribuciones inequitativas (Braverman, 2008).

La perspectiva distributiva de la justicia ambiental busca relacionar entonces métricas usadas para cuantificar los servicios ecosistémicos con métricas sobre la población y sus condiciones de vida usando unidades espaciales para caracterizar su comportamiento en el área de estudio. Típicamente se usan variables ambientales que representan aspectos biofísicos de los ecosistemas p.e superficie de la hoja, índice de área foliar, el área de cobertura de la copa o los efectos directos e indirectos de los ecosistemas sobre variables climáticas p.e temperatura o humedad, o físico-químicas para representar la composición del suelo y del aire, o mediciones de la capacidad de secuestrar carbono de los árboles y la de filtrar agua del suelo.

Los primeros trabajos enmarcados en la justicia ambiental se enfocaron en la ubicación de plantas de residuos y manejo de desechos relacionándolos con variables sociales como la etnicidad e indicadores de segregación racial o en comunidades con bajos ingresos (Heynen et al., 2006) y que desarrollan trabajos como (Chakraborty and Armstrong, 1997; Cutter et al., 1996). Posteriormente con el avance de modelos y tecnologías de la información para la caracterización de la infraestructura ecológica urbana y la cuantificación de servicios ambientales se desarrollan metodologías para establecer relaciones entre distribuciones desiguales adoptando el uso de la cobertura de copa de los árboles como variable que se consolida para este tipo de estudios, con variables socioeconómicas principalmente de raza/etnicidad de la población, que tradicionalmente ha sido una preocupación en el estudio de las desigualdades sociales (Heynen et al., 2006; Landry and Chakraborty, 2009; Phelps, 2012; Zhou and Kim, 2013; Schwarz et al., 2015; Watkins et al., 2016).

De los espacios verdes suelen caracterizarse dimensiones sobre el acceso, su dimensiones físicas, los equipamientos y servicios que prestan, y si los beneficiarios de son del ámbito local al EV o más amplio. Igualmente es importante preguntarse por la calidad del parque y zonas verdes así como los uso, que pueden variar dependiendo de las comunidades que están disfrutando del parque (Kabisch and Haase, 2014). El acceso suele ser un concepto complejo de medir, pues el análisis espacial de los datos arroja variaciones significativas usando diferentes métricas que proviene de los diferentes conceptos de acceso usados para su definición (Talen and Anselin, 1998).

En Cali se han realizado trabajos que caracterizan procesos de desigualdades entre la población analizando factores como la segregación racial, brechas salariales, empleabilidad e índices socioeconómicos de segregación espacial(Arroyo Mina et al., 2016; Mora and Arcila, 2014; Cerón and Escobar, 2014). Aunque todos ellos incluyen una dimensión espacial, en tanto que usan las comunas como unidades de agregación de las variables socioeconómicas, su análisis no hace uso de los datos espaciales de los objetos geográficos, su base teórica son los modelos econométricos de regresión de mínimos cuadrados o la creación de índices (escalas) para clasificar las unidades gráficas con la escala p.e (Cerón and Escobar, 2014) que hace uso del escalograma de Guttman, un método para normalizar las diferentes escalas usando igual número de rangos discretos y construir un índice acumulativo de todas las dimensiones. Mora and Arcila (2014) estudian la discriminación, tanto racial como por sexo a través de los modelos econométricos de Oaxaca-Blinder usado para analizar las diferencias salariales entre dos grupos usando una función que tiene un componente de discriminación. Para incluir en la ecuación la procedencia se crean los conglomerados a priori y se agregan dummies por conglomerado de comunas en la ecuación. Arroyo Mina et al. (2016) evalúan la calidad del empleo bajo el supuesto de ser buen proxy de la calidad de vida y encuentra que en Cali existe evidencia de que las poblaciones afro son discriminadas laboralmente, y que esta discriminación está explicada por el lugar de residencia. Sin embargo, el método usado es similar al de (Mora and Arcila, 2014) e inclusive usa los mismo conglomerados para las variables dummies de la procedencia.

Los pocos estudios que se han llevado a cabo sobre segregación espacial y socioeconómica muestran que la exclusión de grupos étnicos en Cali tiende a coincidir espacialmente con la segregación de los grupos socioeconómicos de estratos bajos (Cerón and Escobar, 2014). Tendencia que también se ven estudios de ciudades norteamericanas (Heynen et al., 2006; Landry and Chakraborty, 2009; Zhou and Kim, 2013; Nesbitt and Meitner, 2016). Quizá el mapa de distribución de los árboles y los espacios verdes se sume a las inequidades que acumulan los habitantes de ciertas zonas de la ciudad, y este trabajo aporte a la utilización de modelos provenientes de la estadística espacial.

## **2.5 Modelamiento y análisis espacial de variables ambientales y sociales**

Las fuentes de datos con información espacial usada para capturar variables ambientales relacionadas con la vegetación provienen principalmente de imágenes satelitales(Landry and Chakraborty, 2009; Troy et al., 2007; Vásquez Fuentes and Romero Aravena, 2008; Nesbitt and Meitner, 2016), imágenes aéreas (Azócar et al., 2007; Heynen et al., 2006; Tratalos et al., 2007), datos de Lidar (Shanahan et al., 2014; Schwarz et al., 2015) e inventarios producidos por muestras o censos arbóreos y de espacios verdes (Comber et al., 2008; Killicoat et al., 2002; Nowak and Crane, 2000, Nowak and Crane (2002); Talen and Anselin, 1998). Las imágenes satelitales son usadas en gran cantidad de los estudios dadas su creciente disponibilidad y la frecuencia de actualización —Landsat 8 revisita un mismo punto sobre la superficie de la tierra cada 16 días con un desfase de 8 días con respecto al satélite Landsat 7—, lo que permite hacer monitoreo y seguimiento a escalas entre los 15 m a 100 m por ancho de píxel. Los indicadores de cobertura calculados con base en datos de imágenes satelitales son estimados usando la escala de resolución de la imagen en combinación con el índice de Vegetación de Diferencia Normalizada (NDVI) que permite diferenciar entre densidad y tipo de vegetación leñosa o vegetación herbosa (Nesbitt and Meitner, 2016).

El uso de inventarios arbóreos permite estudios más detallados sobre las características del arbolado y son usados para evaluar la salud y estructura de los individuos arbóreos y las capacidades específicas de las especies para proveer servicios ambientales. (Killicoat et al., 2002; Nowak and Crane, 2002; Cowett, 2014).

Entre los distintos indicadores desarrollados para capturar la extensión y distribución de los servicios ambientales la cobertura de copas ha probado ser sensible y eficaz para cuantificar hasta qué punto los árboles y bosques están proporcionando servicios críticos a los residentes (Nowak et al., 2010). Se usan otros indicadores además de la cobertura en la literatura, y en muchas ocasiones se normalizan los valores de las variables ambientales por unidad de área, usando unidades geográficas definidas e introduciendo métricas sobre la densidad y cantidad de población beneficiada en el cálculo. De hecho (Cowett, 2014) propone que para analizar con precisión la distribución espacial de los árboles de las calles y los beneficios que proporcionan es importante migrar hacia métricas en las que las especies arbóreas y el tamaño del árbol sean un factor en el cálculo, pues la mayoría de los servicios de los ecosistemas arbóreos son proporcionales a la cantidad de área superficial de la hoja; en esta medida las especies de árboles de mayor estatura típicamente proporcionan muchos más beneficios que las especies de menor estatura. Trabajos como (Alanís et al., 2014) usan indicadores ecológicos de las especies como abundancia, dominancia y frecuencia para construir índices de importancia para valorar las especies nativas y estructura de los bosques urbanos, cuya conservación también hace parte de las metas de manejo del AU (Nowak et al., 2010).

La tabla 2.3 resume los indicadores para cuantificar servicios/beneficios y estado de los ecosistemas arbóreos usados en la literatura revisada.

Table 2.3: Métricas para caracterizar servicios del AU

Métrica	Referencia
Producción de alimentos [toneladas/año]	(Gómez-Baggethun and Barton, 2013)

Metrica	Referencia
Índice de área foliar	(Gómez-Baggethun and Barton, 2013)
Disminución de la temperatura [° C] por cobertura arbórea[m2] en parcelas/sitios cubiertas de árboles	(Gómez-Baggethun and Barton, 2013)
Superficie de la hoja [m2]	(Gómez-Baggethun and Barton, 2013)
O3, SO2, NO2, CO y PM10 m removido en [toneladas/año] multiplicado por la cobertura arbórea (m2)	(Gómez-Baggethun and Barton, 2013)
Secuestro de CO2 por árbol (carbono multiplicado por 3.67 para convertir a CO2)	(Gómez-Baggethun and Barton, 2013)
Abundancia (por especie) [individuos]	(Alanís et al., 2014)
Densidad [individuos/m2]	(Nowak et al., 2010)
Cobertura de copa de árbol por persona[m2/persona]	(Nowak et al., 2010)
Cobertura de copa de árbol[m2]	(Nowak et al., 2010) y muchos antes también
porcentaje de cobertura de copa de árbol en un área	(Nowak et al., 2010)
Dominancia de una especie en función a la cobertura de copa [m2/ha]	(Alanís et al., 2014)
Dominancia relativa de la especie i respecto a la dominancia total	(Alanís et al., 2014)
Frecuencia absoluta de una especie (porcentaje de presencia en los sitios)	(Alanís et al., 2014)
Abundancia relativa de la especie i respecto a la abundancia total en el área de estudio	(Alanís et al., 2014)
Árboles por habitante	POT2014

Metrica	Referencia
índice de cobertura de copa de árbol al rededor de un punto de muestreo	(Zhou and Kim, 2013)
Porcentaje de cobertura de copa de árbol en servidumbres y espacios públicos	(Landry and Chakraborty, 2009)

En cuanto a los espacios verdes se valoran típicamente los aspectos geométricos y el acceso. Los aspectos físicos se caracterizan con medidas de superficie de los espacios, el porcentaje de área o número de espacios respecto de las unidades espaciales del análisis; este tipo de medidas son llamadas por (Talen and Anselin, 1998) la aproximación contenedor (como el representado en la ecuación (2.1)), y aunque muy extendidas en la literatura para caracterizar acceso a los EV asumen que los beneficios del espacio verde tiene un impacto local: *“Sin embargo, si el investigador está seguro de que la esfera de influencia de un servicio dado se limita a un límite geográfico específico, puede seguir siendo apropiado el enfoque unidimensional tradicional de la accesibilidad por medio de conteos por unidad”* (Talen and Anselin, 1998). En esta misma línea, (Kabisch and Haase, 2014) asegura que aunque el secuestro de carbono tiene beneficios a nivel de toda la ciudad, algunos procesos biofísicos que rigen los beneficios de los EV ocurren a nivel local. Aunque no es fácil definir la escala a la que operan los procesos biofísicos, los beneficiarios de los servicios ambientales son a menudo aquellos que viven cerca de los EV.

Respecto de las unidades espaciales de agregación de las variables (Cutter et al., 1996) encontró que el coeficiente de correlación tiende a aumentar con el aumento de la escala de las unidades espaciales de agregación. Para verificar la robustez de las estimaciones de correlación entre la cantidad de instalaciones tóxicas/desperdicios e indicadores socioeconómicos Cutter et al. (1996) usó pruebas de hipótesis para probar que tanto explican los indicadores de etnia y de ingreso la presencia de las instalaciones, agregando los datos a diferentes escalas de la cartografía censal en EE.UU (condado > sectores > bloque). Los resultados sugieren que los sectores y bloques son la escala espacial más apropiada para

evaluar las desigualdades debido a la alta variabilidad de los indicadores socioeconómicos dentro del condado en comparación con los sectores y los bloques.

Estos hallazgos ponen en cuestión la selección del tamaño de las unidades espaciales del análisis en relación con el área de estudio. Para el caso de Cali se suele usar las comunas como unidad espacial, pues es la unidad sobre la que se define la inversión, pero puede no representar muy bien la escala de los procesos que dominan las transformaciones en el AU y el acceso a los espacios verdes cuando se presenta alta variabilidad de los indicadores socioeconómicos y ambientales.

En esta línea es interesante analizar variables que capturan aspectos físicos de la estructura del espacio disponible. (Landry and Chakraborty, 2009) relaciona las coberturas con las áreas de los sectores censales completos y los compara con las áreas de espacio público sin los lotes de las manzanas de cada sector censal, introduciendo un indicador espacial que apunta a espacios donde la responsabilidad es claramente de las instituciones estatales por el énfasis en el área de jurisdicción para la intervención de la estructura arbórea y de los espacios verdes (ver tabla 2.3).

Además de las medidas de superficie y sus variantes para caracterizar los beneficios en una unidad espacial definida, se usan medidas que relacionan origen destino, como la distancia mínima del centroide de la unidad espacial al borde del EV más cercano, que proporciona una estimación fiable de la distancia media desde cualquier punto dentro de una unidad de análisis. Este método trata a todos los espacios verdes por igual, independientemente del tamaño y la propiedad (Nesbitt and Meitner, 2016).

Para medir el acceso se pueden usar medidas de distancia, que operan bajo el concepto de costo de viaje. Una forma es calcular la suma de las distancias desde el centroide de una unidad espacial a **todos** los EV, como lo hace la ecuación (2.2) y su variante (2.3) que divide el costo de viaje entre el número de EV (Talen and Anselin, 1998). Una alternativa a la suma de distancias es usar la distancia del centroide al espacio verde más cercano (ecuación (2.4)). Otra métrica relevante de acceso es la distancia de red a través de la

estructura de las vías de la ciudad o distancia a pie, que produce una versión más realista de la experiencia de acceso, y puede usarse para calcular la distancia mínima promedio de puntos aleatorios o para cada manzana, por ejemplo, dentro de una unidad de espacial. Estas métricas exigen marcar los puntos de acceso a las zonas verdes, tarea que se realiza de forma no automática (Zhou and Kim, 2013).

*índice contenedor*

$$A_i^C = \sum_j s_j \quad \forall j \in I \quad (2.1)$$

*costo de viaje*

$$A_i^T = \sum_j d_{ij} \quad (2.2)$$

*costo de viaje normalizado*

$$\bar{A}_i^T = \sum_j d_{ij}/N \quad (2.3)$$

*distancia mínima*

$$A_i^M = \min |d_{ij}| \quad (2.4)$$

*índice de accesibilidad a pie*

$$A^W = \sum_{\int R_b} (r_{min}/d_j) \quad \forall d_j > 0, \quad r_{min} < R_b \quad (2.5)$$

El cálculo de estos indicadores e índices de acceso se ha complejizado con base en recomen-

dación de las agencias ambientales que definen valores de referencia como la cantidad de EV mínimo disponible ( 2 ha ) en un radio determinado (300m). Esto puede lograrse, por ejemplo, con la suma de los cocientes entre el radio recomendado y la distancia de un punto de muestreo dentro de la unidad espacial a cada uno de las EV en el radio de búsqueda  $R_b$ , que suele ser mayor que la distancia de acceso recomendada  $r_{min}$ . Se pueden seleccionar un número de muestras para cada unidad espacial y sumar o promediar las distancias obtenidas (Zhou and Kim, 2013) o usar el centriode como único punto (ecuación (2.5)). Modificaciones a este índice pueden incluir el área de la zona verde para cuantificar el acceso y las características del EV al que se accede (Comber et al., 2008).

El tipo de espacios usualmente aceptados como espacio verde urbano en los estudios incluyen parques y jardines públicos, corredores verdes (por ejemplo, adyacentes a ríos y canales), reservas naturales locales, áreas comunes, pequeñas áreas de bosques con sotobosque, sitios de importancia para la conservación de la naturaleza, áreas de drenaje (es decir, áreas regularmente inundadas cercanas a los ríos), cementerios, instalaciones desocupadas cubiertas de vegetación, lagos y humedales, campos de golf, áreas privados de asociaciones de propietarios, terrenos escolares, parcelas y hasta senderos que no son lo suficientemente grandes para calificar como parques(Nesbitt and Meitner, 2016; Comber et al., 2008; Kabisch and Haase, 2014; Zhou and Kim, 2013).

La tabla 2.4 resume los indicadores usados en la literatura revisada sobre el acceso a EV.

Table 2.4: Métricas para caracterizar servicios del EV

Métrica	Referencia
Superficie de los espacios públicos verdes [ha]	varios
Número de instalaciones o servicios contenidos en una unidad dada (por ejemplo, distrito censal o división político-administrativa)	(Talen and Anselin, 1998)
Potencial de la gravedad	(Talen and Anselin, 1998)
Distancia al EV más cercano[m]	(Talen and Anselin, 1998)

Métrica	Referencia
Coste medio de viaje EV más cercano	(Talen and Anselin, 1998)
Superficie de los espacios públicos verdes por habitante (o cada 1000 habitantes)[ha/habitantante]	
Cobertura de espacios verdes[%]	varios
Distancias entre puntos o nodos en la red de movilidad urbana	(Comber et al., 2008)
Índice de accesibilidad caminando al parque dentro de un radio	(Zhou and Kim, 2013)

Para la caracterización de las condiciones de vida de la población en relación con el disfrute a servicios ambientales los indicadores sociales usados suelen provenir de los censos de población, encuestas de trabajo o calidad de vida, registros catastrales para el avalúo de predios y uso de los suelo. La tabla 2.5 resume algunos de los indicadores socioeconómicos usados para la evaluación de inequidades ambientales en la revisión de la literatura realizada.

Table 2.5: Métricas para caracterizar aspectos socioeconómicos de la población y estructura de las unidades espaciales del EV

Métrica	Referencia
Ingreso medio del hogar	varios
Valor medio vivienda ocupada por el propietario	(Cowett, 2014)
Porcentaje de viviendas ocupadas	(Cowett, 2014)
Porcentaje de población blanca	(Cowett, 2014)
Porcentaje de personas con pregrado	(Cowett, 2014)
Porcentaje de población afro	[Cowett (2014)]

Métrica	Referencia
Porcentaje de viviendas arrendadas	(Heynen et al., 2006)
Densidad de población [residentes/ha]	(Troy et al., 2007)
Densidad de hogares[hogares/ha]	(Troy et al., 2007)
Valor medio vivienda ocupada	(Troy et al., 2007)
Porcentaje de viviendas desocupadas	(Troy et al., 2007)
Índice de delincuencia	(Troy et al., 2007)
Porcentaje de viviendas que son viviendas unifamiliares	(Troy et al., 2007)
Tamaño medio del hogar	(Landry and Chakraborty, 2009)
Tamaño medio del lote de vivienda	[Shanahan et al. (2014)]
Densidad de viviendas	
Grados escolares cursados[años]	
Porcentaje de habitantes con grado profesional	(Nesbitt and Meitner, 2016)
Porcentaje de habitantes con maestría	(Nesbitt and Meitner, 2016)
Edad del barrio	
Edad media de las construcciones	(Zhou and Kim, 2013)

### 2.5.1 Enfoque estadístico

El instrumento matemático más popular para establecer relaciones de dependencia y asociación entre dos variables aleatorias son los índices de correlación, sin que la relación cuantificada sea necesariamente de causalidad. La causalidad entre las variables suele ser una apuesta del investigador y de su conocimiento o intuición sobre los procesos que dominan las características de las variables aleatorias (Gibbons and Overman, 2012). El

coeficiente de correlación de Pearson es medida usada para caracterizar la fuerza de relaciones lineales, mientras que coeficientes de correlación como el de Spearman son usados para encontrar relaciones no necesariamente lineales, pues se calcula observando si los cambios en una de las variables está relacionado con incrementos o decrementos de la otra, lo que permite cuantificar dependencias no lineales. Estos coeficientes son análisis eficaces para seleccionar variables candidatas a ser incluidos en modelos de regresión, tanto para seleccionar las variables que tienen relación fuerte con la variable a predecir, como la independencia lineal de las variables explicativas, condiciones necesarias para que los métodos de ajuste de parámetros de los modelos lineales (Gibbons and Overman, 2012). Los coeficientes de correlación, como estadística de resumen, no pueden reemplazar el examen visual de los datos y la construcción de relaciones que tengan un fundamento teórico.

Los modelos de regresión lineal son ampliamente usados en la econometría para construir modelos explicativos con términos que describen componentes teóricos de los procesos de estudio e inferir el comportamiento o tendencias en una población con base en una muestra. De esta forma se pretende cuantificar el cambio de la variable dependiente ( $y$ ) ante aumentos o disminuciones del valor de una de las  $p$  variables dependientes  $x_j$  donde  $j = 1...p$ . Cada una de las observaciones  $i$ ,  $i = 1...n$  del vector  $X_p$  denominadas  $x_{ij}$  forman un sistema de ecuaciones lineales donde el parámetro  $\beta_j$  pesa el aporte explicativo de la variable correspondiente al estimarlo (ecuación (2.6) y (2.7)). Así, los coeficientes estimados representan la importancia de una variable independiente en los cambios de la dependiente. Los métodos para resolver son estimadores de mínimos cuadrados (OLS ordinary least-square) o de máxima verosimilitud (ML maximum likelihood). Si se cumple que las variables independientes del modelo y los términos del error tienen una distribución normal y una varianza constante (homocedasticidad), entonces el estimador es óptimo entre todos los estimadores no sesgados lineales y no lineales. Es clave indagar el cumplimiento de las condiciones de normalidad, en especial en los residuos. (Gibbons and Overman, 2012).

modelo lineal de la  $i$ -esima observación

$$y_i = \beta_0 + \beta_1 \cdot x_{i1} + \dots + \beta_j \cdot x_{ij} + \dots + \beta_p \cdot x_{ip} + e_i \quad (2.6)$$

forma matricial del sistema de ecuaciones a resolver

$$Y_{n \times 1} = X_{n \times p} \beta_{p \times 1} + e_{(n \times 1)} \quad (2.7)$$

Estudios como (Heynen et al., 2006; Vásquez Fuentes and Romero Aravena, 2008) usan coeficientes de correlación entre los pares de variables ambientales y socioeconómicas agregadas en unidades censales o administrativas para luego usar modelos de regresión de lineal con el fin de estimar la importancia explicativa de las variables a través de la estimación de los coeficientes.

Tratalos et al. (2007) indaga sobre relaciones entre indicadores ambientales y el estatus social de los residente en cinco ciudades del Reino Unido; se usa coeficientes de correlación de Spearman para seleccionar las variables usadas en un modelo de regresión lineal usando OLS para probar que los desarrollos urbanos de alta densidad generalmente se asociaban con un pobre desempeño ambiental, medido por el tamaño del parche en el espacio verde y los niveles de provisión de servicios ambientales.

### 2.5.2 Enfoque geoestadístico

Cuando se analiza el resultado de los residuos  $e_i$  debería ser ruido. Sin embargo si se encuentra que existe algún tipo de relación de los residuos con la posición de las observaciones  $i$  en plano geográfico, entonces se puede usar variaciones en el modelo lineal, que consiste en incluir un término lineal usando la variable dependiente (modelo autoregresivo SAR, ecuación (2.8)), las independientes ( spatial lag o retardo espacial en  $X$  SLX, ecuación (2.10)), en el error (modelo espacial del error SEM,ecuación (2.9)) o usando una combinación del modelo de error y autoregresivo (modelo espacial de Durbin SD,ecuación

(2.11)). Todas estas aproximaciones introducen una matriz de  $W_{n \times n}$ , donde  $n$  es el número de sitios, que captura la influencia de las variables en relación con su proximidad.

Esto significa que se propone a priori una relación espacial al dar estructura a  $W$ , por ejemplo usando solo valores de 1 (vecino) y 0 (no vecino) a las unidades espaciales contiguas o expresa una relación global usando una métrica de distancia en los valores de la matriz  $W$ , por ejemplo  $w_{ij} = 1/d_{ij}$  donde  $d_{ij}$  es la distancia euclíadiana entre los centroides de los sitios  $i$  y  $j$ . La idea es que los vecinos más cercanos influencian más la unidad  $i$ , basados en la primera ley de la geografía o Ley de Tobler<sup>6</sup> (Tobler, 1970). Este es un supuesto fuerte, y debe ser coherente con el fenómeno y los datos que lo representan. Justamente con la matriz  $W$  se calcula el coeficiente de Moran's I (Moran, 1950), que se usa para probar una asociación global entre una variable y su posición en el plano. Si los residuos  $e_i$  muestran una asociación fuerte con la estructura de  $W$  en la prueba de Moran's I, se sugiere usar alguno de los modelos espaciales (SAR, SLX, SEM o SD), pues esto significa que no se puede confiar en los coeficientes estimados por la regresión lineal. Se aduce que  $W$  puede ayudar a explicar el proceso que se está modelando. De lo contrario es mejor usar un modelo lineal u otra técnica. (Páez and Scott, 2005; Kissling and Carl, 2008; Anselin, 2002) hacen una revisión de los procesos espaciales, sus supuesto y los criterios de ajuste como errores cuadráticos y criterio de información de Akaike para la selección de un modelo y diseño bien formulado y ajustado.

### Modelo espacial autoregresivo (SAR)

$$Y = X\beta + \rho WY + e \quad (2.8)$$

### Modelo de error espacial (SEM)

$$Y = X\beta + uu = \rho Wu + e \quad (2.9)$$

---

<sup>6</sup>“Todas las cosas están relacionadas entre sí, pero las cosas más próximas en el espacio tienen una relación mayor que las distantes.”

Modelo de retardo espacial en los terminos independiente (SLX)

$$Y = X\beta + WX\lambda + e \quad (2.10)$$

Modelo espacial de Durbin SD

$$Y = X\beta + \rho WY + WX\lambda + e \quad (2.11)$$

Estudios como (Landry and Chakraborty, 2009; Schwarz et al., 2015; Zhou and Kim, 2013; Shanahan et al., 2014) hacen uso de modelos autorregresivos y de retardo espacial para explicar los errores en el modelo lineal y mejorar el ajuste comparando modelos que relacionan de las variables ambientales(cobertura de copa y acceso a EV) con indicadores socioeconómicos.

Sin embargo trabajos como [LeSage and Pace (2014); Kissling and Carl (2008); Gibbons & Overman, 2012] hacen una crítica muy relevante a esta aproximación, y muestran que la inclusión de los términos de auto regresión (SAR) no aportan mucho, pues existe un acople entre los estimadores de los coeficientes que le quita peso a la posibilidad de interpretar los coeficientes como factores del efecto aislado de uno de las variables explicativas. Además, muestra que el ajuste no suele ser muy sensible a variaciones en  $W$ , lo que derrumba un poco la idea de importancia de la vecindad definida a priori. (LeSage and Pace, 2014) propone que si la meta de usar modelos de regresión espacial es tener aproximadamente medidas escalares correctas de resumen de los efectos directos, indirectos y totales sobre la variable dependiente que surgen de cambios en las variables explicativas, entonces si parece bastante factible que se usen estos modelos (LeSage and Pace, 2014).

Además del uso del análisis de autocorrelación espacial global, se puede explorar métodos que hacen análisis local de la variabilidad y sensibilidad de los parámetros usados para las estimaciones OLS. Si no hay autocorrelación global o ninguna agrupación, todavía

podemos encontrar grupos a nivel local utilizando autocorrelación espacial local. El hecho de que de la Morán I es una suma de productos cruzados individuales es explotado por los “indicadores locales de asociación espacial” (LISA) para evaluar la agrupación de las unidades individuales mediante el cálculo del índice de Moran local para cada unidad espacial y la evaluación de la significación estadística p-value para cada sitio  $i$  (Talen and Anselin, 1998). En Fotheringham et al. (1998) se muestra que la regresión geográficamente ponderada (Geographically weighted regression GWR) produce resultados más informativos con respecto a la variación de parámetros en el espacio que la simple utilización de los agregados estadísticos. Estos métodos puede ser utilizado para examinar visualmente la variabilidad espacial de los resultados de la regresión o de las variables del estudio a través de una región y así informar sobre la presencia de no-estacionariedad espacial. La GWR es usada con éxito en (Comber et al., 2011) para analizar los posibles factores que pueden ayudar a identificar y caracterizar las dimensiones a tener en cuenta para mejorar el acceso a la salud.

### 2.5.3 Otras técnicas

(Nesbitt and Meitner, 2016) investigan la relación de usos de suelo para establecer diferencias entre las coberturas usando análisis de varianza (ANOVA). En (Kabisch and Haase, 2014) se usa el análisis cluster para identificar grupos de población con características socioeconómicas similares y similar acceso a espacios verdes, para luego analizar si existen diferencias entre la distribución de los espacios verdes usando el coeficiente de Gini para diferentes grupos de población definidos en un tipo de análisis de disimilitud.

# Chapter 3

## Metodología

### 3.1 Aproximación general

Este trabajo se concentra en indagar en particular sobre la justicia ambiental distributiva por medio de modelos estadísticos, geoestadísticos apoyados por el análisis visual de los patrones espaciales. Para ellos se hará uso de datos del censo arbóreo de 2015<sup>1</sup>, el censo de población del 2005<sup>2</sup> y aspectos estructurales delentre el espacio público y privado de las unidades espaciales de análisis<sup>3</sup>.

### 3.2 Área de estudio

El municipio de Santiago de Cali se encuentra ubicado al suroccidente colombiano. Es la capital del departamento del Valle del Cauca y es la tercera ciudad más poblada del

---

<sup>1</sup>Desafortunadamente los datos del censo arbóreo no están disponibles al público general por motivos desconocidos, pues según la legislación Colombiana no existe ningún tipo de impedimento ni restricciones de privacidad sobre estos datos, que fueron recolectados con el auspicio de dineros públicos. Trabajé con una copia a la que tuve acceso por vínculos laborales con la Alcaldía de Cali

<sup>2</sup>Disponibles en (cen; DANE, 2005; geo, a)

<sup>3</sup>Disponibles en (geo, b; noa, b)

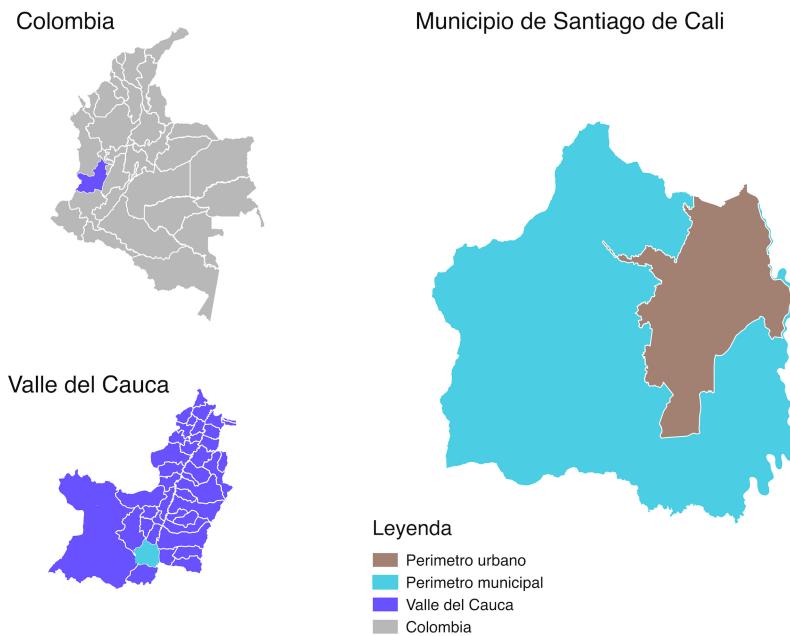


Figure 3.1: Área de estudio

país, después de Bogotá y Medellín, con 2420114 habitantes segun Cali en Cifras 2015. El municipio tiene un área 561.7 Km<sup>2</sup>, un área del perímetro urbano 120.4 Km<sup>2</sup> (21.4 %).(Ver figura 3.1).La división administrativa de la zona urbana son comunas y las comunas se componen de barrios.

Santiago de Cali presenta dos zonas topográficas: el valle del río Cauca hacia el oriente, el terreno más plano donde se ubica el casco urbano, y la zona de piedemonte hacia el occidente sobre la margen derecha de la cordillera Occidental. El area urbana limita al oeste y sur con el área rural del municipio, al este con el río Cauca y los municipios de Palmira y Candelaria, y al norte con el municipio de Yumbo.

El clima del municipio varía en relación al rango altitudinal que abarca entre 916 y 1,438 msnm. En la zona plana, se presenta un clima cálido con características semihúmedas hacia el sur y semiáridas hacia el norte mientras el piedemonte presenta condiciones de clima



Figure 3.2: Mapa de Santiago de Cali,tomada de Google Maps

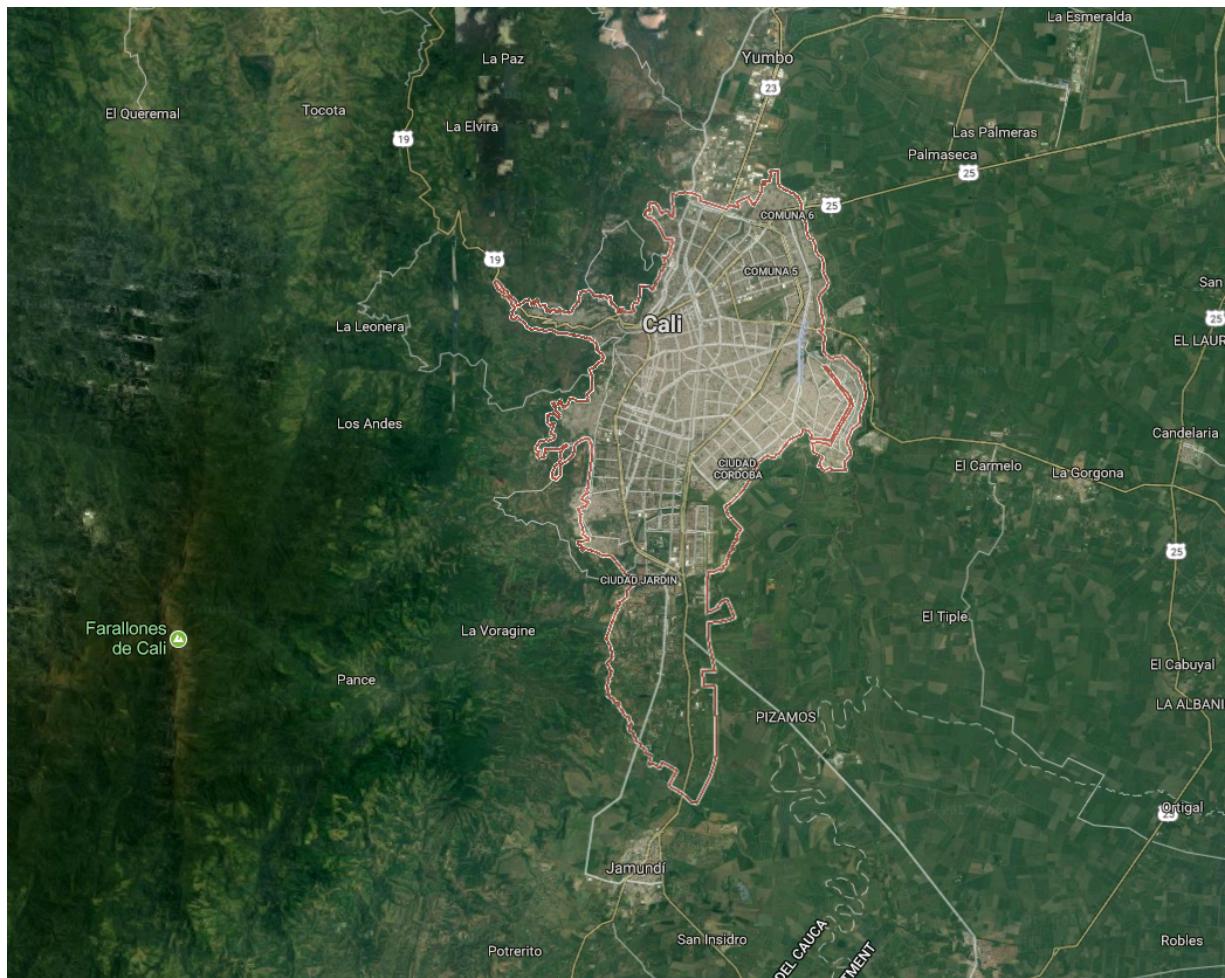


Figure 3.3: Vista satelital de Santiago de Cali,tomada de Google Maps

templado. La precipitación anual promedio es de 1.500 mm y la temperatura promedio anual es de 24 °C aproximadamente. (CIAT, 2015). La ciudad de Cali es de clima caliente, donde la sombra y la brisa son bien valoradas por sus habitantes.

## **3.3 Datos**

### **3.3.1 Datos de registros oficiales del municipio**

La cartografía disponible en la Infraestructura de Datos Espaciales de Santiago de Cali, IDESC (geo, b), incluye información sobre los objetos geográficos naturales, de infraestructura urbana, límites y divisiones político administrativas y la clasificación de predios en cuanto a espacio público disponibles en coordenadas planas del sistema (noa, a). Además está la base de datos geográfica del Plan de Ordenamiento territorial de Cali 2014, POT2014<sup>4</sup>. Del POT2014 se seleccionaron conjuntos de datos de equipamientos y espacio público contenido en la estructura ecológica complementaria (ECC) que incluye cementerios, universidades, EV de acceso no restringido aunque algunos sea predios privados contenidos en EEC. De la IDESC se seleccionó la capa de barrios, espacio público, humedales, ríos y corredores ambientales disponibles vía WFS.

En la figura 3.4 se muestra un mapa con las capas seleccionadas para el realizar el procesamiento y los análisis.

### **3.3.2 El censo arbóreo**

En el año 2015 la ciudad de Santiago de Cali (Colombia) concretó la realización de un censo arbóreo que dejó como resultado una base de datos de aproximadamente 290.000 individuos censados. Los datos dan cuenta de la identificación de especies, sus características dasométricas, de emplazamiento y estado fitosanitario. Estos datos constituyen un insumo fundamental para la caracterización de los beneficios y cargas que supone el mantenimiento y desarrollo del arbolado urbano. De hecho, su realización ocurre en el marco del

---

<sup>4</sup>Toda la información del POT2014 se encuentra en la web de la Alcaldía y puede descargarse como archivo GDB compatible con ArcGIS 10.4 o consumirse de Geoserver de IDESC como WFS o mapas en formato pdf del acuerdo

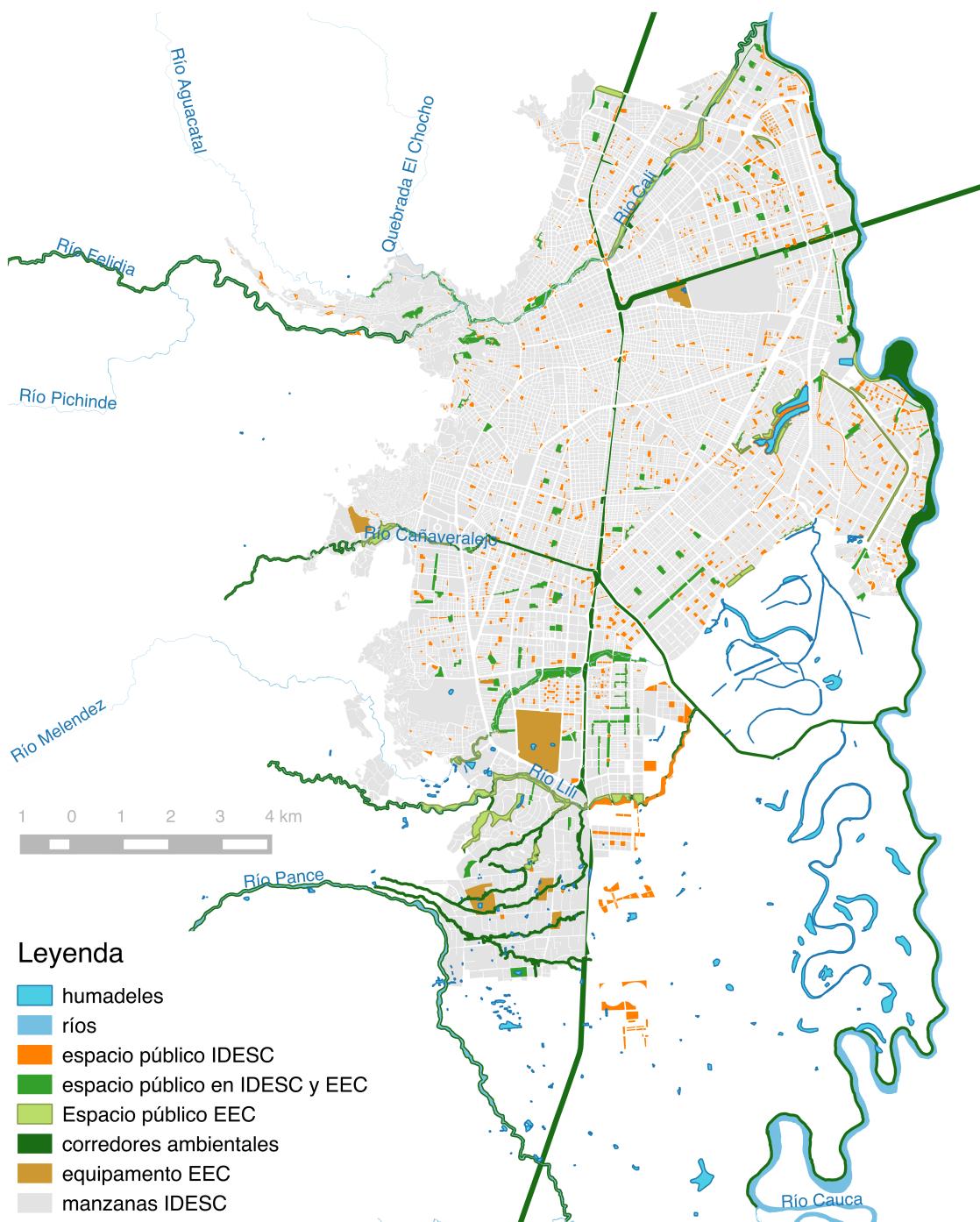


Figure 3.4: Capas usadas para el procesamiento de los espacios verdes y las características de las manzanas

proceso de formulación del plan silvicultura urbana o estatuto arbóreo<sup>5</sup> (Acuerdo 0335 de 2013). Las variables biofísicas recolectadas y la georeferenciación de los individuos permite agregar las características del arbolado a diferentes escalas de las unidades administrativas p.e divisiones censales, para identificar y caracterizar su distribución espacial y correlación con variables sociales o/y económicas. Las variables seleccionadas se resumen en la tabla 3.1 y en la tabla 3.2.

Table 3.1: Variables para caracterizar el AU

variable	{valores}[unidades]
id_arbol	número entero único
diametro	[m <sup>2</sup> ]
copa	
altura	[m]
arbol	
vitalidad	{Regular, Sano, Seco, Muerto}
edad	{Juvenil, Maduro, Longevo}
emplazamiento	{Anden, Bahias de estacionamiento, Bulevares, Corredor Ferreo, Escenario deportivo y/o Cultural, Glorieta, Parque Urbano, Paseos, Plaza, Plazoleta, Ronda de rios, Rondas de canales, Separador Vial}
vegetación	{Arbol, Arbusto, Bambu, Muerto, Palma, Planta arbustiva, Seco}
Este	[m] MAGNA - SIRGAS-CALI

---

<sup>5</sup>El proyecto de censo arbóreo se formuló en dos fases; la primera se ejecutó mediante convenio No 095 de 2013 entre la CVC y la Universidad Autónoma de Occidente, y la segunda fase mediante convenio No 049 de 2014 entre las mismas entidades. Los datos no se encuentran publicados y fueron solicitados mediante un derecho de petición.

Table 3.2: Muestra de los datos del censo arbóreo

id	vegetacion	edad	emplazamiento	diametro_copa	altura_arbol
0199G41070768	Arbol	Maduro	Ronda de rios	12	11
0199G41070769	Arbol	Maduro	Ronda de rios	7	8
0199G41070770	Arbol	Maduro	Ronda de rios	5	3
0199G41070771	Arbol	Maduro	Ronda de rios	6	7
0199G41070772	Arbol	Maduro	Ronda de rios	9	8
0199G41070773	Arbol	Maduro	Ronda de rios	10	7

variable	{valores}[unidades]
Norte	[m] MAGNA - SIRGAS-CALI

Los árboles incluidos en el censo arbóreo de 2015 están separados por 10 años de los datos de caracterización socioeconómica y de la características de ocupación de los sectores urbanos. Esta brecha puede cerrarse un poco excluyendo los árboles catalogados como jóvenes. Otro factor que ayuda a matizar la distancia entre los datos de ambos censos es que el AU se desarrolla lento y los tamaños actuales de las copas y las alturas de los árboles, aunque son mayores que en el 2005, hablan del potencial que desarrollaron esas zonas. Dado que esa es la disponibilidad de datos poblacionales hay que asumir estas diferencias. Finalmente, para este trabajo usaremos árboles, palmas y bambú de más de 1.9 m de altura (descartando arbustos y plantas arbustivas) para garantizar que son individuos que proveen sombra a los transeúntes.

En la siguiente figura 3.5 se observa todos los arboles seleccionados para el análisis.

### 3.3.3 El censo de población

El último censo de población en Colombia se realizó en el año 2005, y los datos se pueden consultar y agregar en las diferentes unidades censales (sector, sección, manzana) usando

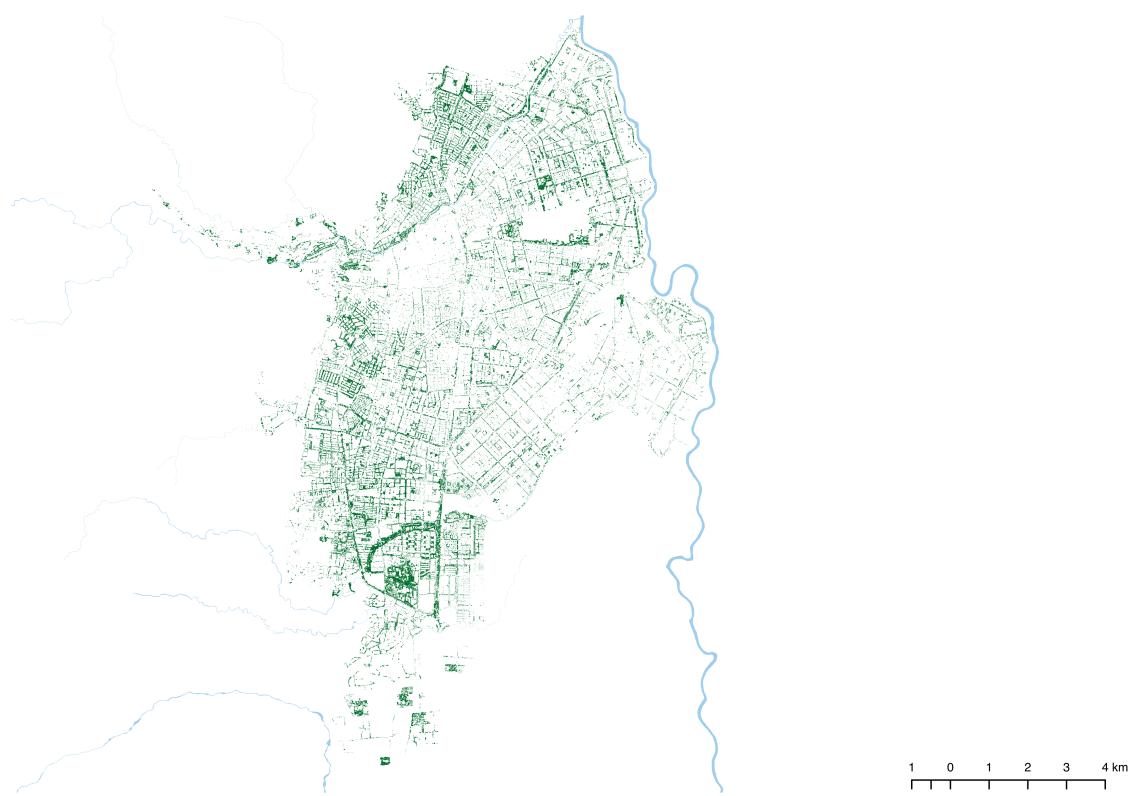


Figure 3.5: Árboles seleccionados para el análisis. El tamaño de puntos que representan los arboles es proporcional al area de la copa en unidades del mapa.

una sistema de consulta web de censos Redatam<sup>6</sup>. Estos datos sirven para caracterizar la población con base en indicadores y rasgos de las personas, aspectos sobre el uso del suelo y los tipos de vivienda. Las variables disponibles para el análisis están resumidas en las tablas 3.3 y 3.4.

El otro componente de los datos es la cartografía censal del DANE (geo, a) disponible para las diferentes unidades espaciales de agregación en el sistema de coordenadas WGS84. Para el análisis se tiene en cuenta todos las unidades censales que se interceptan con el perímetro urbano disponible en la IDESC, pues el censo arboreo se limitó al este prerímetro.(ver figura 3.6)

Table 3.3: Variables sobre la población

variable	{valores}[unidades]
Pertenencia Étnica	[personas]{indígenas, ROM, gitanos, raizales del Archipiélago de San Andrés, Providencia y Santa Catalina, palenqueros de San Basilio, afrocolombianos}
Con alguna limitación	[personas]{sí,no}
Con estudios superiores o postgrado	[personas]

---

<sup>6</sup>El sistema de consulta es el (CEPAL), que se puede acceder directamente desde (DANE, 2005) y en la página web del DANE dónde está organizada la documentación metodológica y otros servicios del censo.

variable	{valores}[unidades]
Ningún estudio	[personas]

Table 3.4: Variables sobre las viviendas

variable	{valores}[unidades]
tipo vivienda	[viviendas] {Casa,Casa indígena,Apartamento,Tipo cuarto,Otro tipo de vivienda}
uso vivienda	[predio]{Uso Vivienda.Uso Unidad Económica,Uso LEA}
cantidad predios	[predios]
cantidad viviendas	[viviendas]

Una de las apuestas del proyecto es incluir aspectos de los procesos de desarrollo urbano a través de la idea de barrio: como unidad de identidad cultural urbana y estructural, de características físicas y habitacionales en las que confluyen las transformaciones que hacen los habitantes y los diseños urbanos e intervenciones arquitectónicas de los planeadores y constructores en la ciudad. Sin embargo, la información socioeconómica disponible está en unidades de la cartografía censal del 2005, que no coincide exactamente con los límites de barrios. Un primer supuesto es que los sectores censales aproximan bien a los barrios, pues las diferencias no tan drásticas (ver figura 3.7). La unidad espacial de análisis sobre la cual se harán todas las agregaciones el sector urbano (SU) de la cartografía censal 2005.

Los demás conjuntos de datos, cuyos elementos sobrepasan los límites del centro poblado

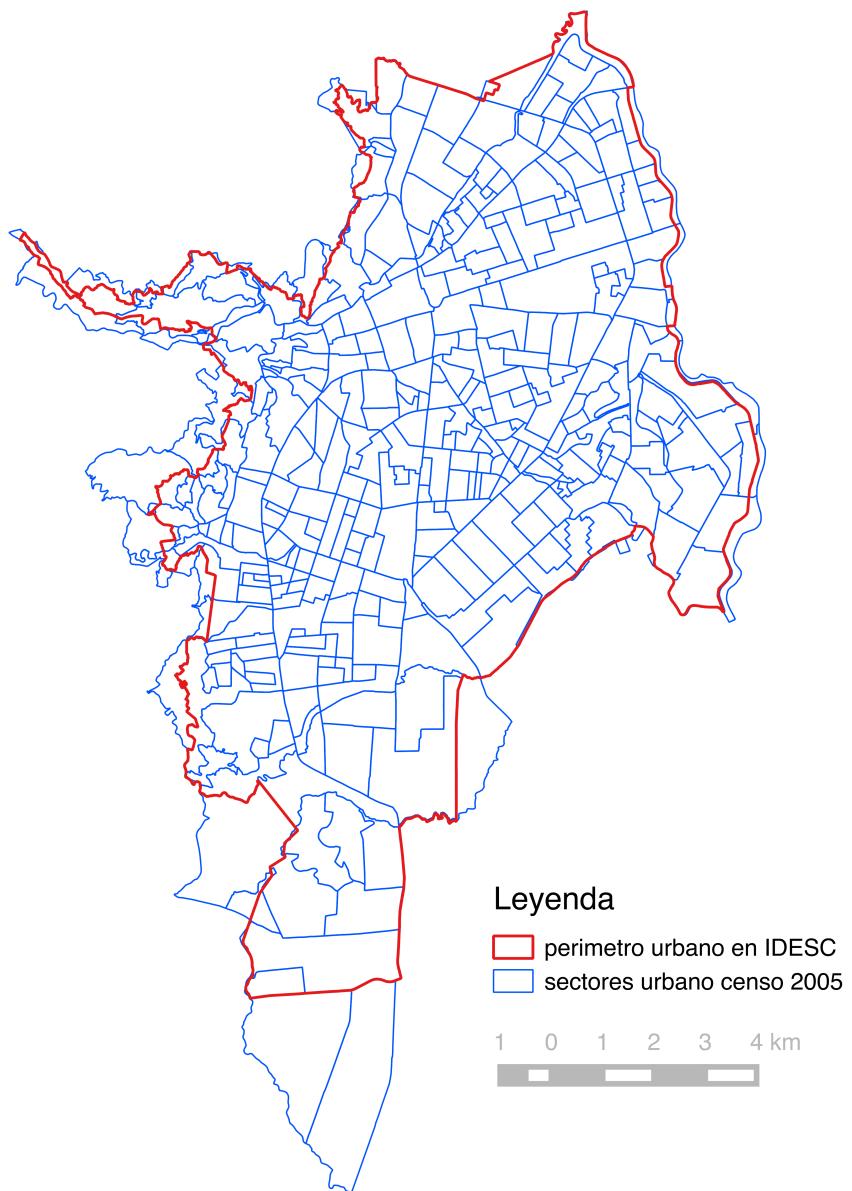


Figure 3.6: División en barrios y sectores urbanos de Santiago de Cali

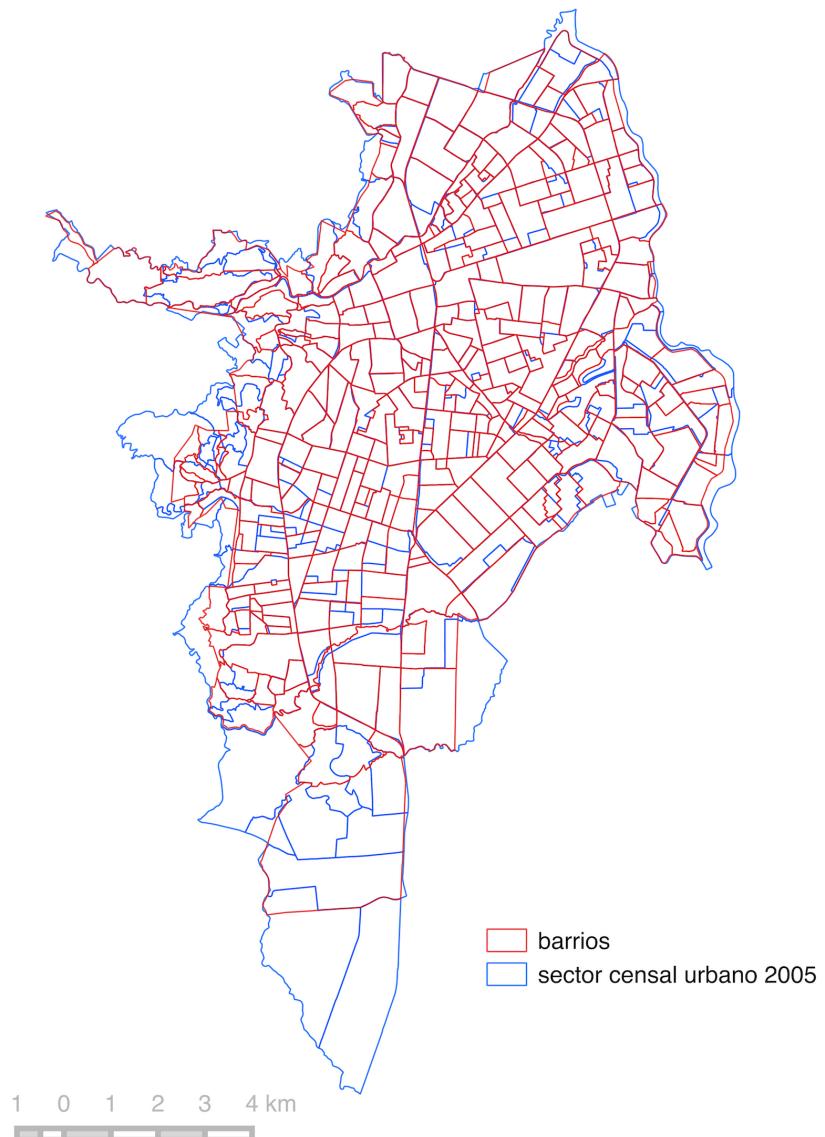


Figure 3.7: División en barrios y sectores urbanos de Santiago de Cali

conformado por los SUs seleccionados son preprocesados para excluir los datos y zonas fuera del perímetro conformado por los SUs.

## 3.4 Métodos y técnicas

El análisis propuesto se compone de las siguientes actividades:

1. Preparación de los datos: una tarea común pero crucial para el análisis de datos. La estandarización de las variables categóricas y la identificación de valores atípicos o inconsistentes es una base firme para la estimación de parámetros y obtener soluciones confiables y sensibles de interpretación. Los datos suelen estar usualmente en formatos para la lectura humana o con distinta estructura de las variables de los modelos. La preparación de los datos consume la mayor parte de los esfuerzos de las tareas de procesamiento de los datos. Los datos del censo arbóreo se encuentra en tablas bien conformadas lo que facilita su manipulación. Los datos de consulta del censo de población vienen en tablas independientes para cada unidad espacial seleccionada, con diferentes longitud de variables. A esto se suma el componente espacial, donde hay que prestar particular atención a los sistemas de coordenadas y usar coordenadas planas consistentes con las unidades de espacio.
2. Procesamiento y análisis estadístico: cálculo de indicadores de cobertura, acceso y variables socioeconómicas. Cálculo de estadísticos para probar normalidad, normalización de las variables e indicadores, cálculos de coeficientes de correlación Pearson y de Spearman entre todos los pares de variables.
3. Inspección visual de los datos: hacer uso de gráficas estadísticas, mapeos y mapas para evaluar y seleccionar los indicadores a usar en un modelo de regresión lineal.
4. Evaluar los residuos usando la prueba de correlación espacial de Moran'I usando al menos dos diseños de matriz W. Si la prueba muestra una correlación y un valor de significancia alta, se prueban modelos tipo SAR, SEM o SLX para comparar su desempeño.

5. Selección del modelo que mejor se ajusta usando métricas de error y de ajuste como R2 y el criterio de Akaike.



# Chapter 4

## Procesamiento y análisis de datos

El procesamiento de los datos se realizó principalmente en R Core Team (2017). Se usó QGIS para conectarse a los servicios WFS del IDESC y previsualizar las capas de información geográfica recolectada y la realización de algunos de los mapas detallados.

El código que implementa los análisis está dividido en archivos para facilitar su lectura, cada uno de los cuales se encargan de transformar los datos de las fuentes y construir estructuras de datos necesarias para realizar las regresiones, las gráficas y los análisis de tipo estadístico y geoestadístico. Cada script implementa una fase de la metodología y produce resultados intermedios que facilitan seguir y reproducir dichas transformaciones sobre los datos de un dominio del problema. El archivo de `funciones.R` agrupa funciones que encapsulan funcionalidades recurrentes dentro del desarrollo del análisis. El script de `geodata.R` opera sobre los fuentes de datos geográficas necesarias para consolidar los índices de acceso a espacios verdes (EV), los indicadores y variables de la estructura de física de los sectores censales y unidades geográficas del análisis. El script `arboles.R` consolida la información de cada uno de los individuos del censo arbóreo agregandolos por sector censal. El script `censopoblacion.R` consolida los datos del Censo de Población 2005. Los scripts `consolidarDatos.R` y `analisis_exploratorio.R` consolidan una única estructura con todos los datos y produce una serie de gráficas y medidas de correlación,

que son base para la identificación de supuestos y selección de las variables independientes para los análisis estadísticos y las regresiones espaciales. Finalmente los script de `analisis_estadistico.R` y `analisis_geoestadistico.R` implementan las regresiones lineales y las regresiones espaciales respectivamente, así como los test y tablas para la verificación de los supuesto matemáticos y la verificación de la calidad de los resultados. Todos estos están reunidos en un script que carga las librerías necesarias y ejecuta secuencialmente cada de los scripts descritos.

```
# Script principal para la ejecución de los .R

# librerias

library(rgdal)
library(rgeos)
library(raster)
library(sp)

library(tidyverse)
library(magrittr)
library(stringr)

library(viridis)
library(RColorBrewer)
library(gridExtra)

library(visdat)
library(GGally)
library(wesanderson)

library(ggrepel)
```

```
# correr los script en el orden correcto para realizar todos los calculos

source("funciones.R")
source("geodata.R")
source("arboles.R")
source("censopoblacion.R")
source("consolidarDatos.R")
source("analisis_exploratorio.R")
source("analisis_estadistico.R")
source("analisis_geoestadistico.R")
```

## 4.1 Capas de información geográfica

Para usar la información geográfica de la cartografía censal y la información del IDESC es necesario establecer un sistema de coordenadas común, en unidades métricas, que facilite integrar la información y produzca resultados consistentes. El sistema de coordenadas proyectadas que vamos a usar es noa (a). Para cargar y manipular los datos espaciales hacemos uso de las librerías `rgdal` (Bivand et al., 2017), `rgeos` (Bivand and Rundel, 2017) y `sp` (Pebesma and Bivand, 2017).

El siguiente mapa muestra los sectores urbanos con sus respectivos códigos de identificación descritos en la documentación que acompaña la cartografía.

La capa de manzanas es necesaria para refinar las capas de espacio verde y poder calcular el área de calle , área privada y otras métricas sobre la estructura de cada sector sector censal y que servirán como criterios para la selección de sectores urbanos a incluir en los análisis de regresión.



Figure 4.1: Sectores Urbanos del Censo del 2005. Los sectores seleccionados están parcial o totalmente contenidos en el perímetro urbano 2015

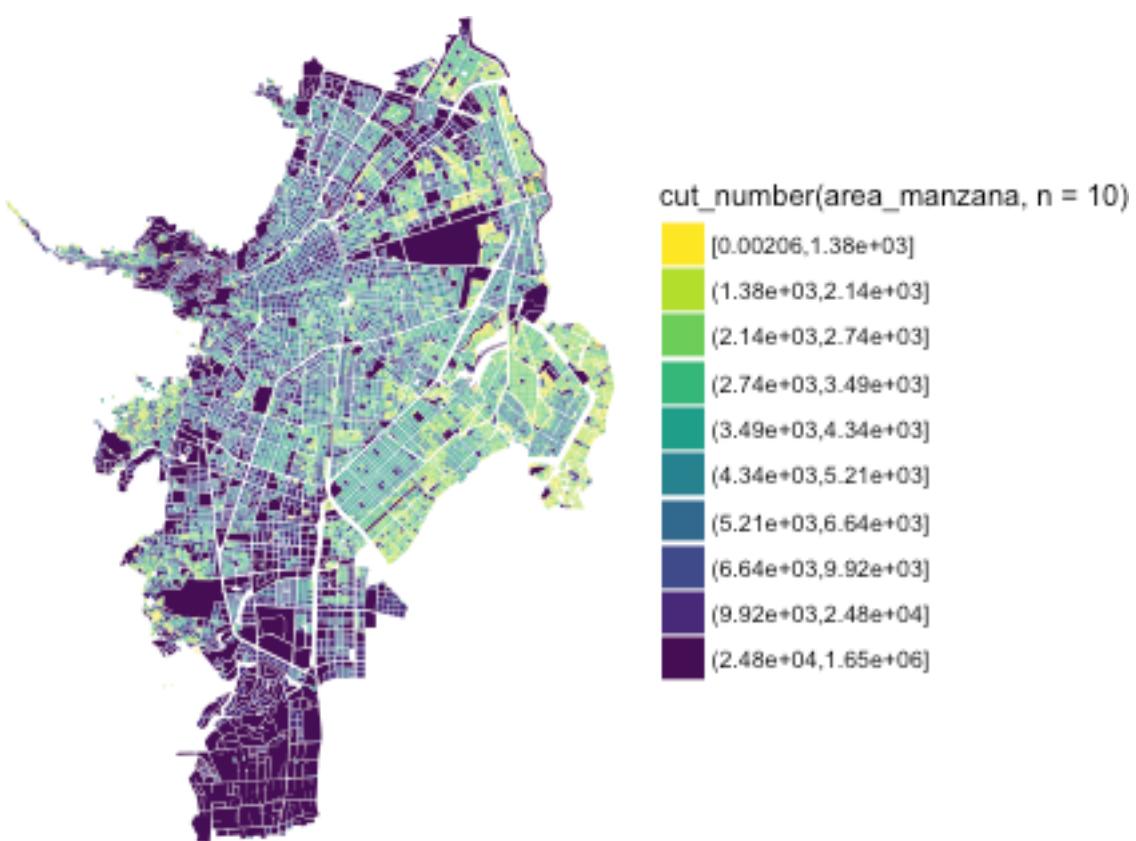


Figure 4.2: Coropleta del tamaño de manzana. Se usaron 10 grupos con aprox. el mismo número de observaciones

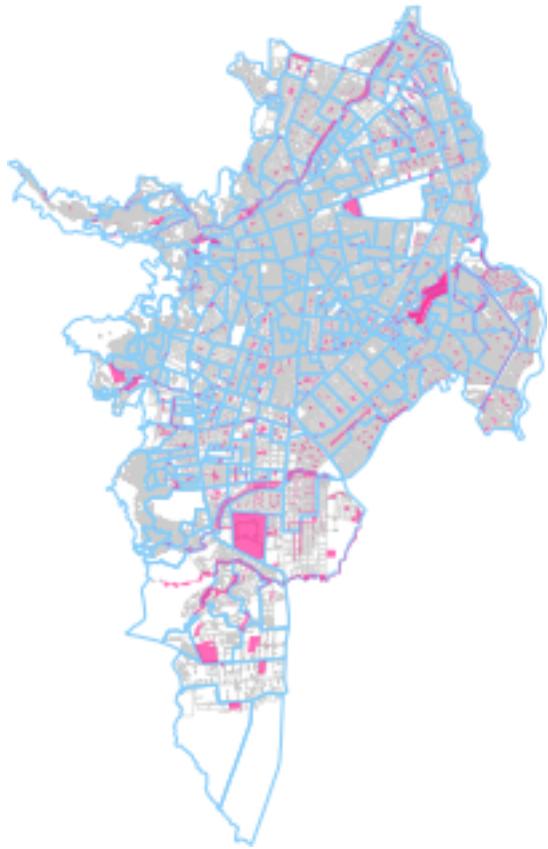


Figure 4.3: Espacio verdes consolidados y sectores urbanos

Las capas de equipamiento de la EEC y espacio público se consolidan en una sola capa conservando la mayor cantidad de información sobre la clasificación de los tipos de espacios disponibles. El resultado puede verse de forma total (ver figura 4.3) o por tipo de espacio (figuras 4.5 y 4.4).

#### 4.1.1 Índices de acceso a espacios verdes

Para mejorar la lectura de esta sección se incluyen a continuación las ecuaciones que definen los índices de acceso seleccionados y las variantes definidas en este trabajo.

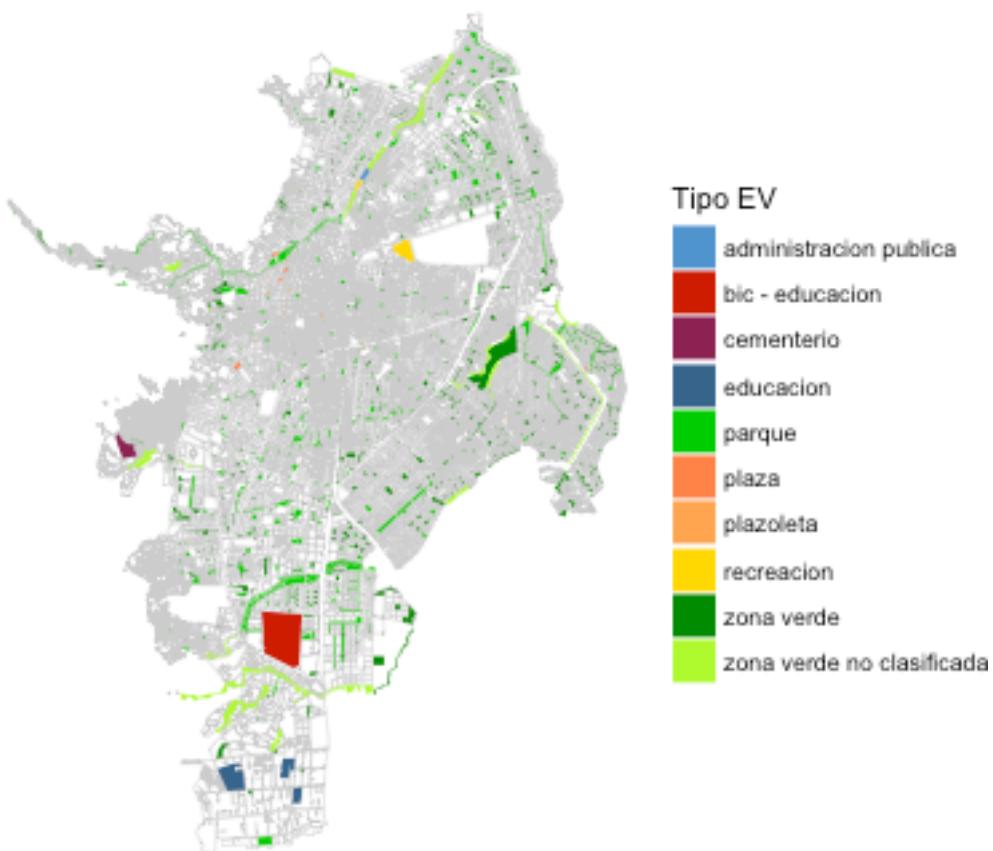


Figure 4.4: Espacio verde por categoría

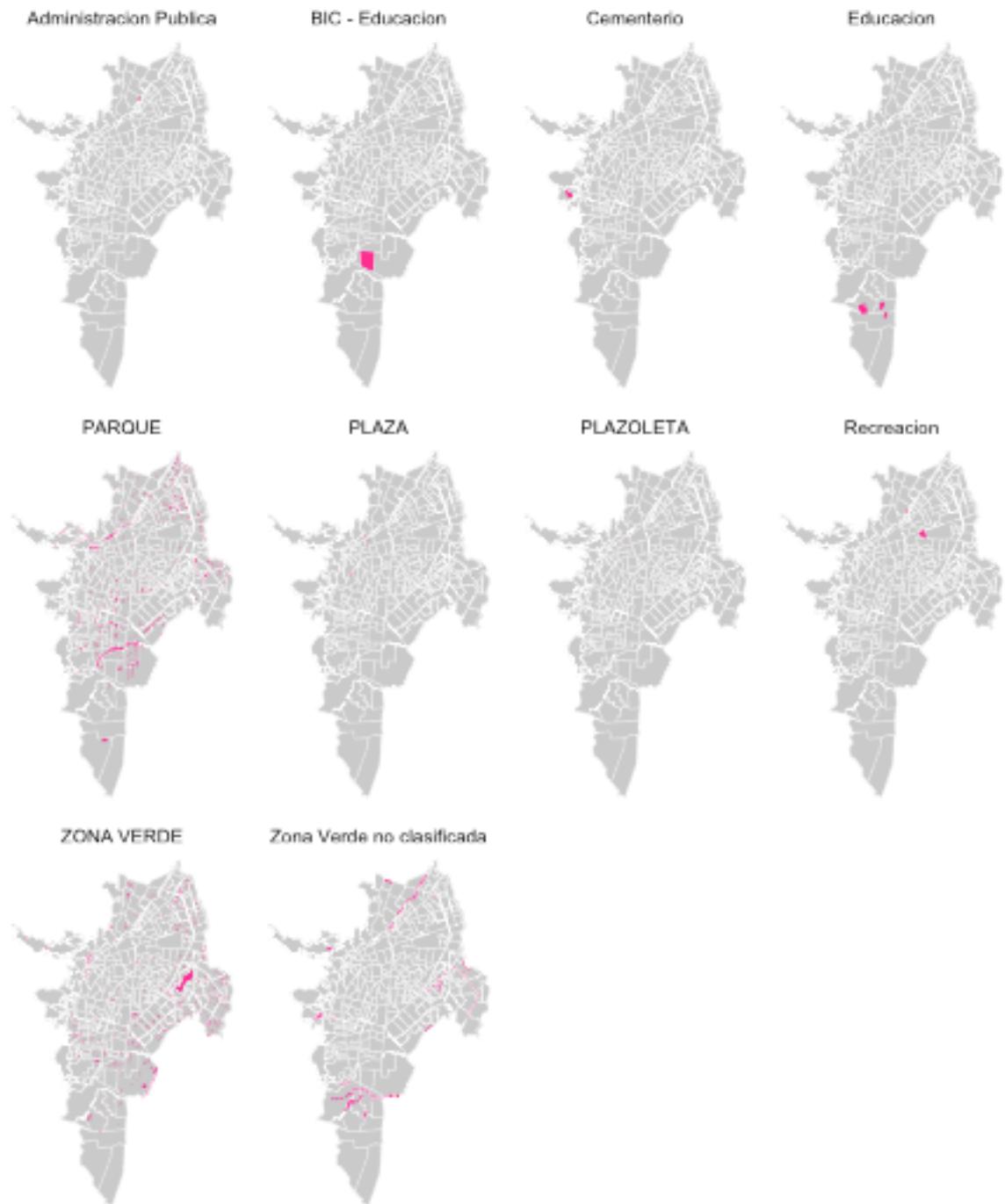


Figure 4.5: Small Multiple del espacio verde por categoría

## Índices de acceso basados en área

**índice contenedor (area\_ep)**

$$A_i^C = \sum_j s_j \quad \forall j \in I \quad (4.1)$$

donde  $s_j$  es el área de cada espacio verde  $j$  que pertenece al conjunto  $I$  de EV dentro del sector  $i$ .

**índice contenedor porcentual (area\_ep.porcentaje)**

$$A_i^{C_p} = 1/a_i \sum_j s_j \quad \forall j \in I \quad (4.2)$$

donde  $a_i$  es el área del sector  $i$ .

**índice área disponible en radio (ia.areas.1000)**

$$A_i^{AR} = \sum_{\int R_b} (s_j \quad \forall j \in I_{R_b}) \quad (4.3)$$

donde  $R_b$  es el radio de búsqueda,  $s_j$  es el área de cada espacio verde  $j$  que pertenece al conjunto  $I_{R_b}$  de EVs en el radio de búsqueda.

**índice porcentual de área disponible en radio (ia.areas.1000.porcentaje)**

$$A_i^{AR_p} = 1/a_t \sum_{\int R_b} (s_j \quad \forall j \in I_{R_b}) \quad (4.4)$$

donde  $a_t$  es el área total de espacio verde en la ciudad.

## Índices de acceso basados en distancia

**costo de viaje** (ia.costoviaje)

$$A_i^T = \sum_j d_{ij} \quad \forall j \in I_t \quad (4.5)$$

donde  $d_{ij}$  es la distancia del centriode del sector  $i$  al espacio  $j$  e  $I_t$  es el conjunto de todos los espacios verdes de la ciudad.

**costo de viaje normalizado** (ia.costo.n)

$$\bar{A}_i^{T_n} = A_i^T / N \quad (4.6)$$

donde  $N$  es el número total de espacio verdes en la ciudad.

**distancia mínima** (ia.mindist)

$$A_i^M = \min |d_{ij}| \quad \forall j \in I_t \quad (4.7)$$

## Índices de acceso mixtos

**razón área distancia** (ia.A.D)

$$A_i^{AD} = \frac{\sum s_j / d_{ij}}{\int R_b} \quad \forall j \in I_{R_b} \quad (4.8)$$

donde  $R_b$  es el radio de búsqueda,  $s_j$  es el área de cada espacio verde  $j$ ,  $d_{ij}$  es la distancia del centriode del sector  $i$  al espacio  $j$  que pertenecen al conjunto  $I_{R_b}$  de EVs en el radio de búsqueda.

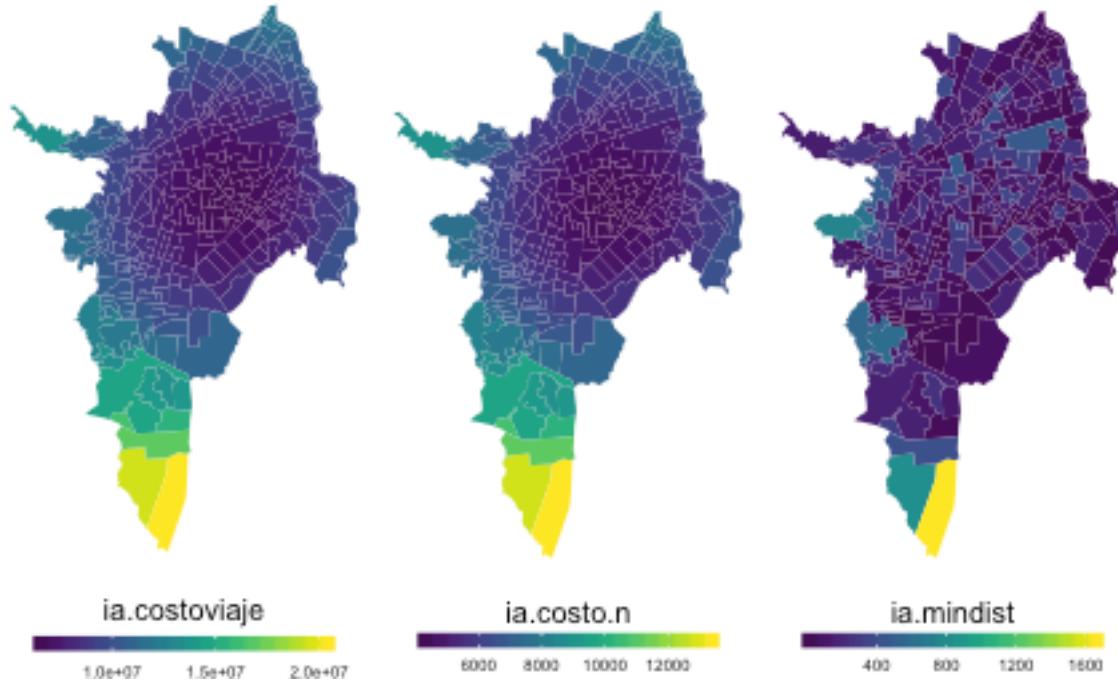


Figure 4.6: Small Multiple de los índices de acceso a EV basados en distancia en escala continua

#### razón área disponible distancia (ia.areas.dist)

$$\bar{A}_i^{AD} = \frac{\sum_{j \in I_{R_b}} s_j}{\sum_{j \in I_{R_b}} d_{ij}} \quad \forall j \in I_{R_b} \quad (4.9)$$

Con los datos consolidados se calculan las áreas de los espacios verdes al interior de cada sector censal, para calcular los índices de acceso tipo contenedor (ecuación (4.1)) interceptando los espacio verdes con los sectores urbanos (ecuación (4.2)). También se obtuvo una versión del índice contenedor como en porcentaje del área del SU. El cálculo de los índices de costo de viajes (ecuación (4.5)) y costo de viaje normalizado (ecuación (4.6)) se obtiene creado una matriz de distancia entre los centroides de los sectores censales y cada uno de los espacios verdes. De esta matriz de distancia también se obtiene el índice de distancia mínima (ecuación (4.7)).

Los mapas de los índices de acceso basados en distancia se muestran en la figura 4.6

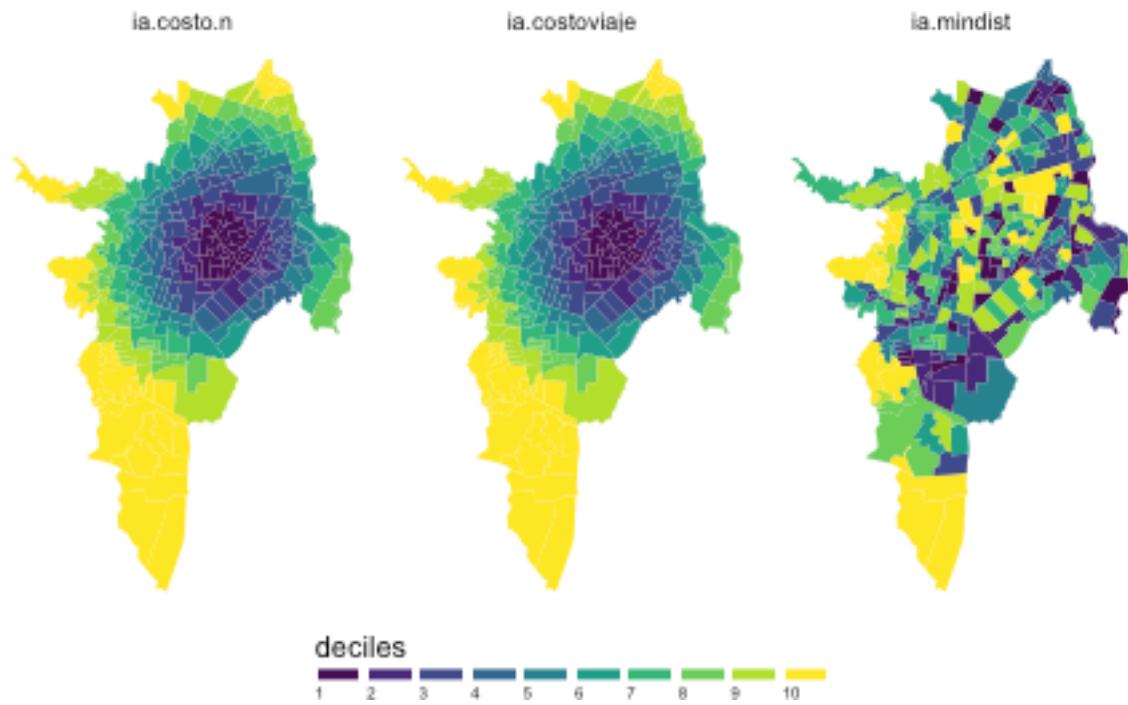


Figure 4.7: Small Multiple de los índices de acceso a EV basados en distancia usando deciles

Además de los índices basados en distancia y el tipo contenedor se calcularon índices de acceso basados en el área de espacio verde en un radio de 1000 metros. Estos índices muestran un dimensión relacionada no con solo con el acceso sino con la cantidad de espacio disponible en el radio de búsqueda definido desde el centroide del sector censal. Para hacernos una idea del radio de búsqueda seleccionado, el siguiente mapa muestra los radios búsqueda y los espacios verdes.

Los índices de acceso basados en área descritos en las ecuaciones (4.1),(4.2) y(4.3) se resumen en la siguiente gráfica.

En la búsqueda de índices de acceso más complejos que reflejen el acceso en distancia y la cantidad de área disponible desde cada sector urbano se construyeron índices similares a el índice de distancia de a pie (ecuación (2.5) que se basan en la razón entre el área a la que se accede y la distancia a la que se encuentra del centroide del sector. Dos nuevos índices se proponen en este trabajo: ia.areas.dist (ecuación (4.9)) como la suma de las áreas en el rango de 1 km desde el centroide del SU dividido la suma de las distancia a

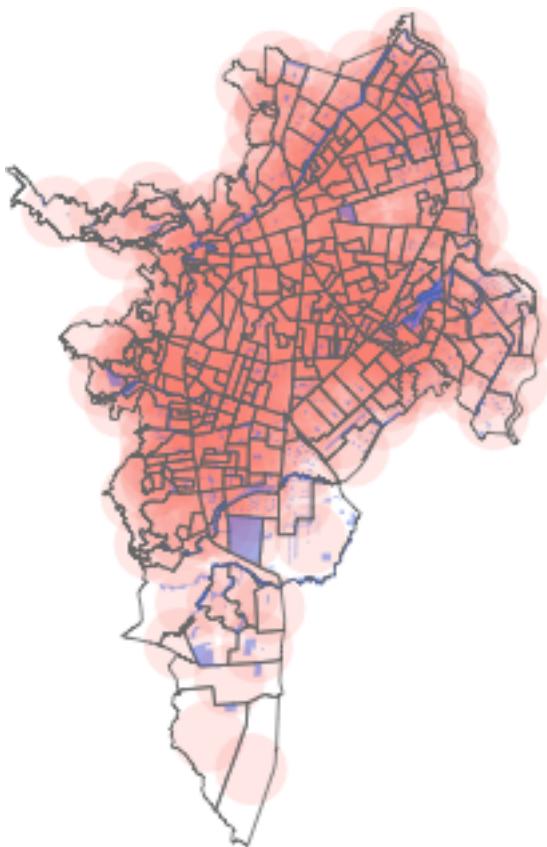


Figure 4.8: Espacio verdes y rango de 1 km desde centriodes de SU

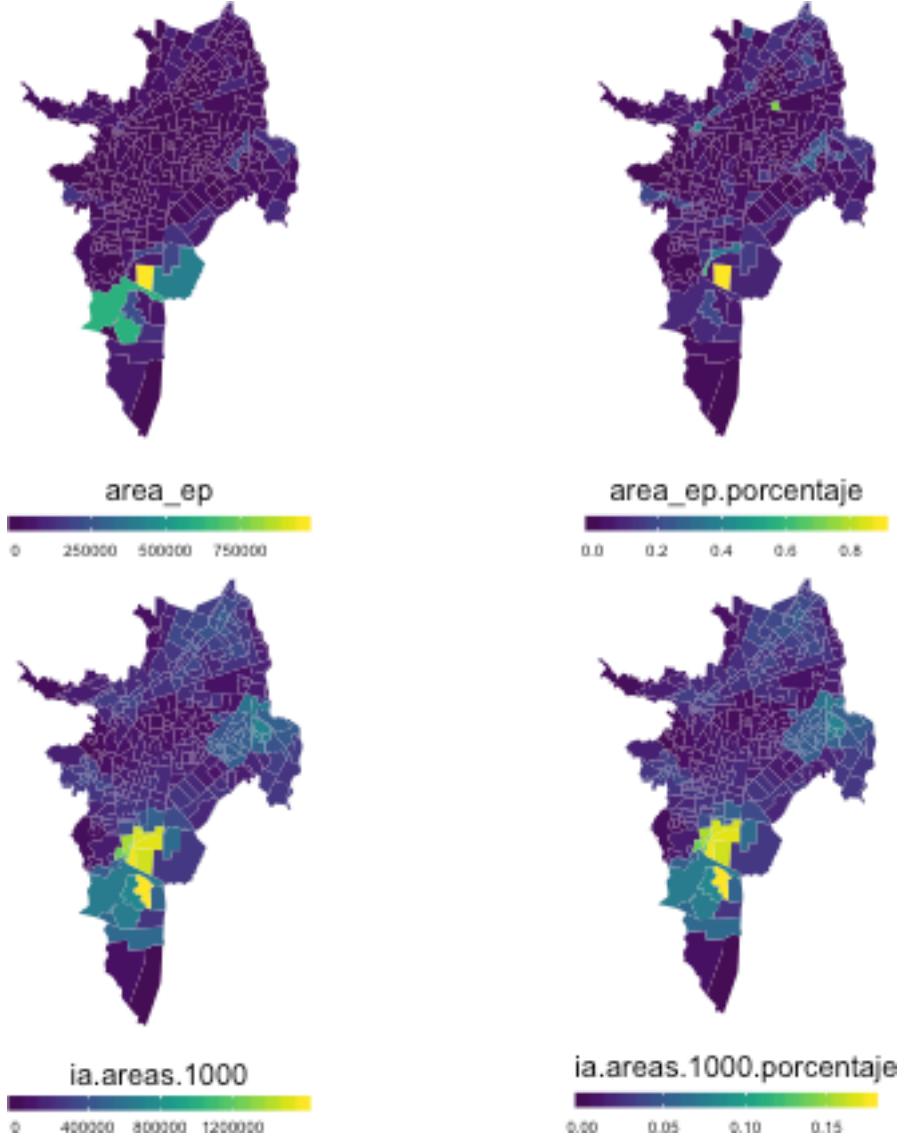


Figure 4.9: Small Multiple de los indices de acceso a EV basados en área usando escala continua

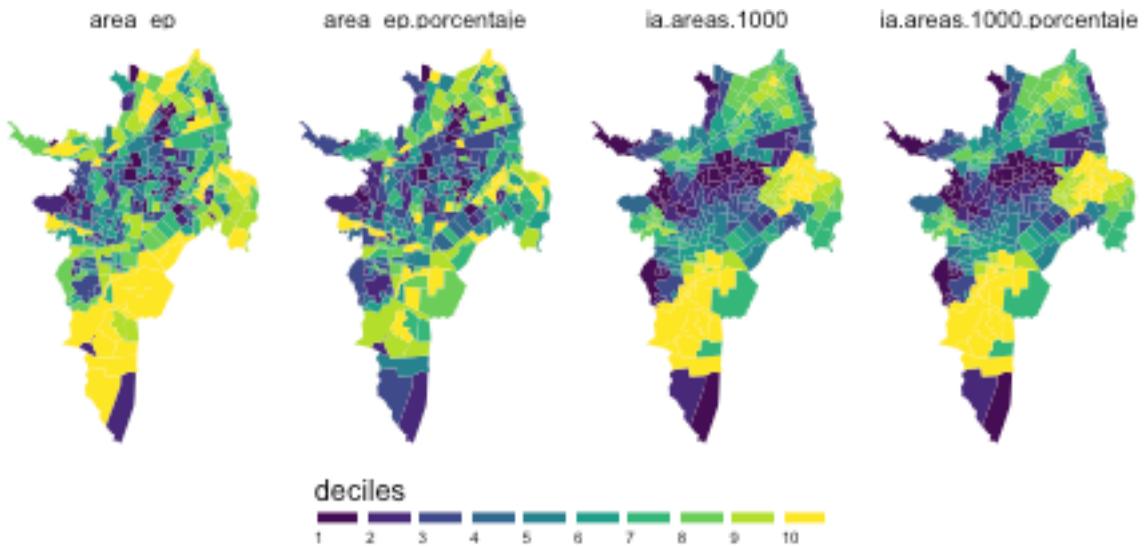


Figure 4.10: Small Multiple de los indices de acceso a EV basados en distancia usando deciles

esos EV; ia.A.D (ecuación (4.8)) , que es la suma de las razones entre el área del espacio verde  $j$  dividido entre la distancia  $d_{ij}$  desde el centroide del SU  $i$  al EV  $j$ . La siguiente gráfica muestra las métricas propuestas.

Además de los mapas es de interés observar la distribución en frecuencia de las métricas de acceso. La siguiente gráfica se observan los histogramas de las metricas calculadas.

Finalmente se muestra el código en R que calcula los índices presentados, y el resumen de los índices calculados.

```
##  ia.costoviaje      ia.costo.n      ia.mindist      area_ep
##  Min.   : 6538022   Min.   : 4293   Min.   : 10.00   Min.   :    0
##  1st Qu.: 7299897   1st Qu.: 4793   1st Qu.: 35.94   1st Qu.: 2158
##  Median : 8376735   Median : 5500   Median : 95.58   Median : 7710
##  Mean   : 8789929   Mean   : 5771   Mean   : 135.94  Mean   : 25712
##  3rd Qu.: 9836587   3rd Qu.: 6459   3rd Qu.: 178.27  3rd Qu.: 25288
##  Max.   :20407619   Max.   :13400   Max.   :1661.44  Max.   :959105
##
##  area_ep.porcentaje ia.areas.1000      ia.areas.1000.porcentaje
```

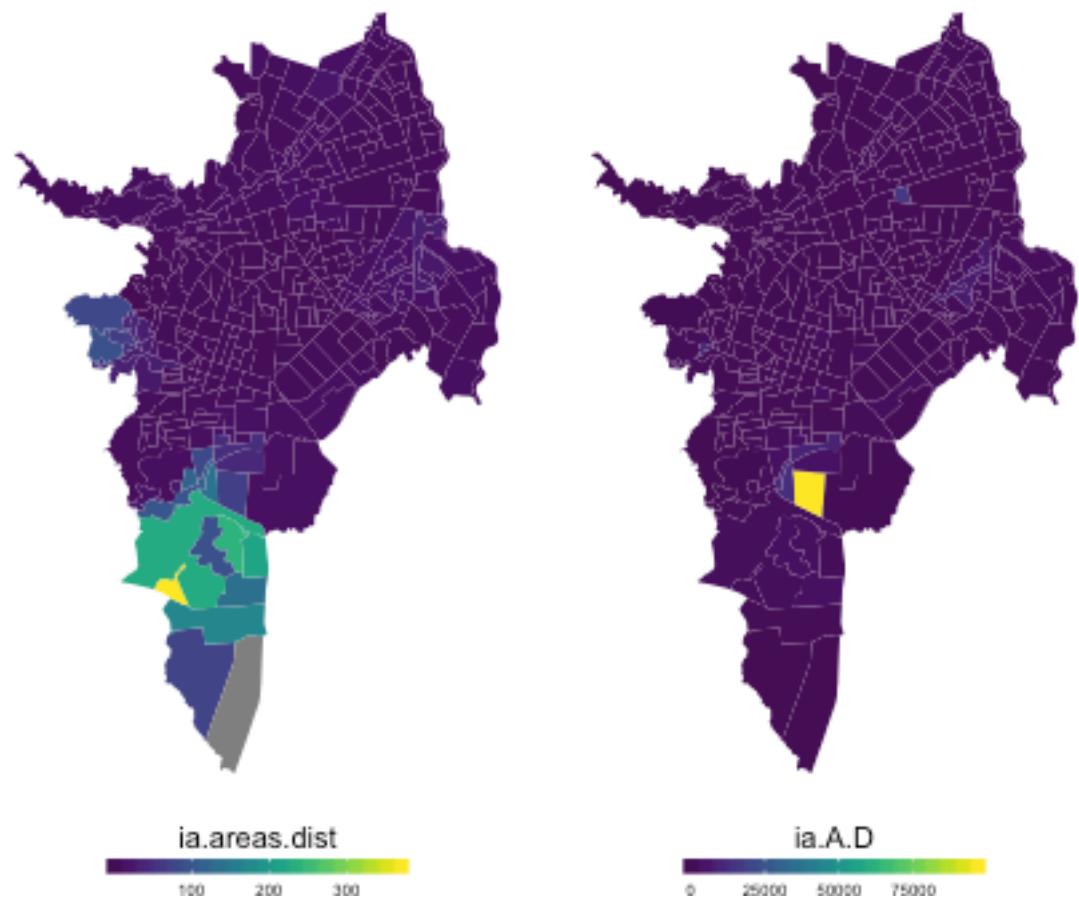


Figure 4.11: Small Multiple de los indices de acceso a EV basados en área y distancia usando escala continua

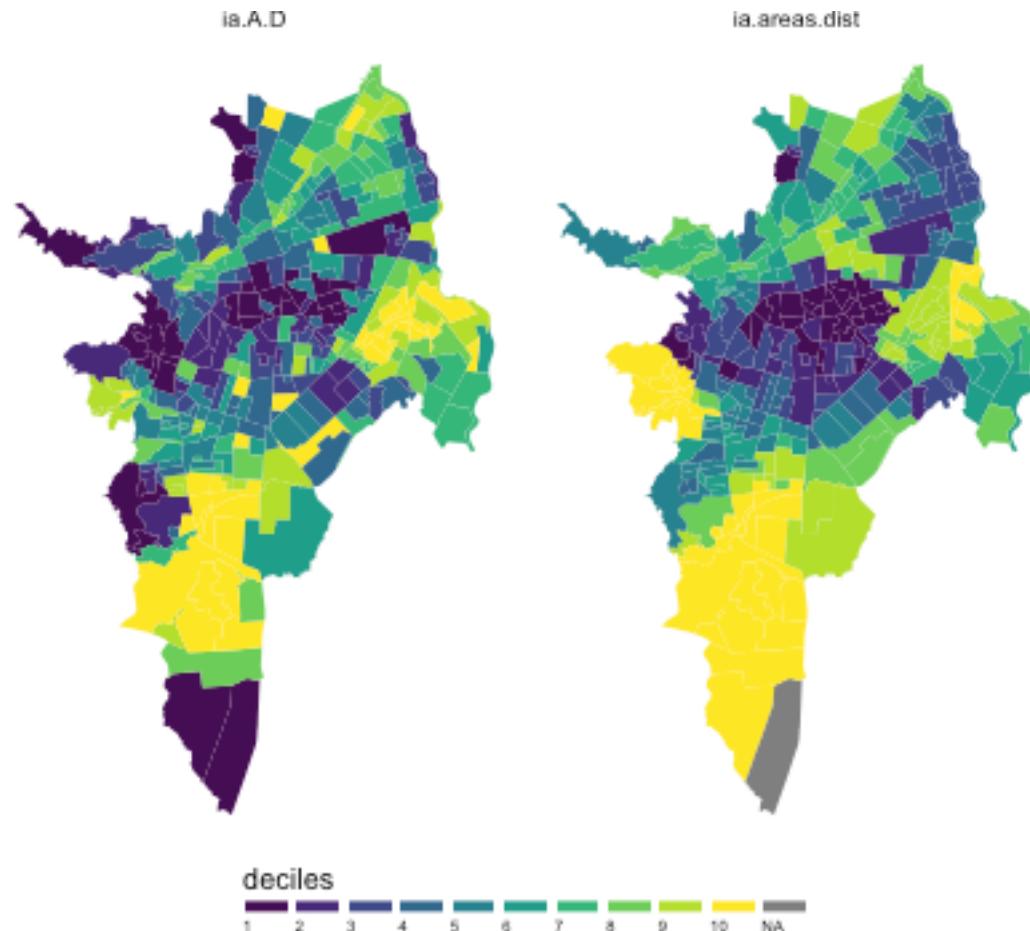


Figure 4.12: Small Multiple de los indices de acceso a EV basados en distancia usando deciles

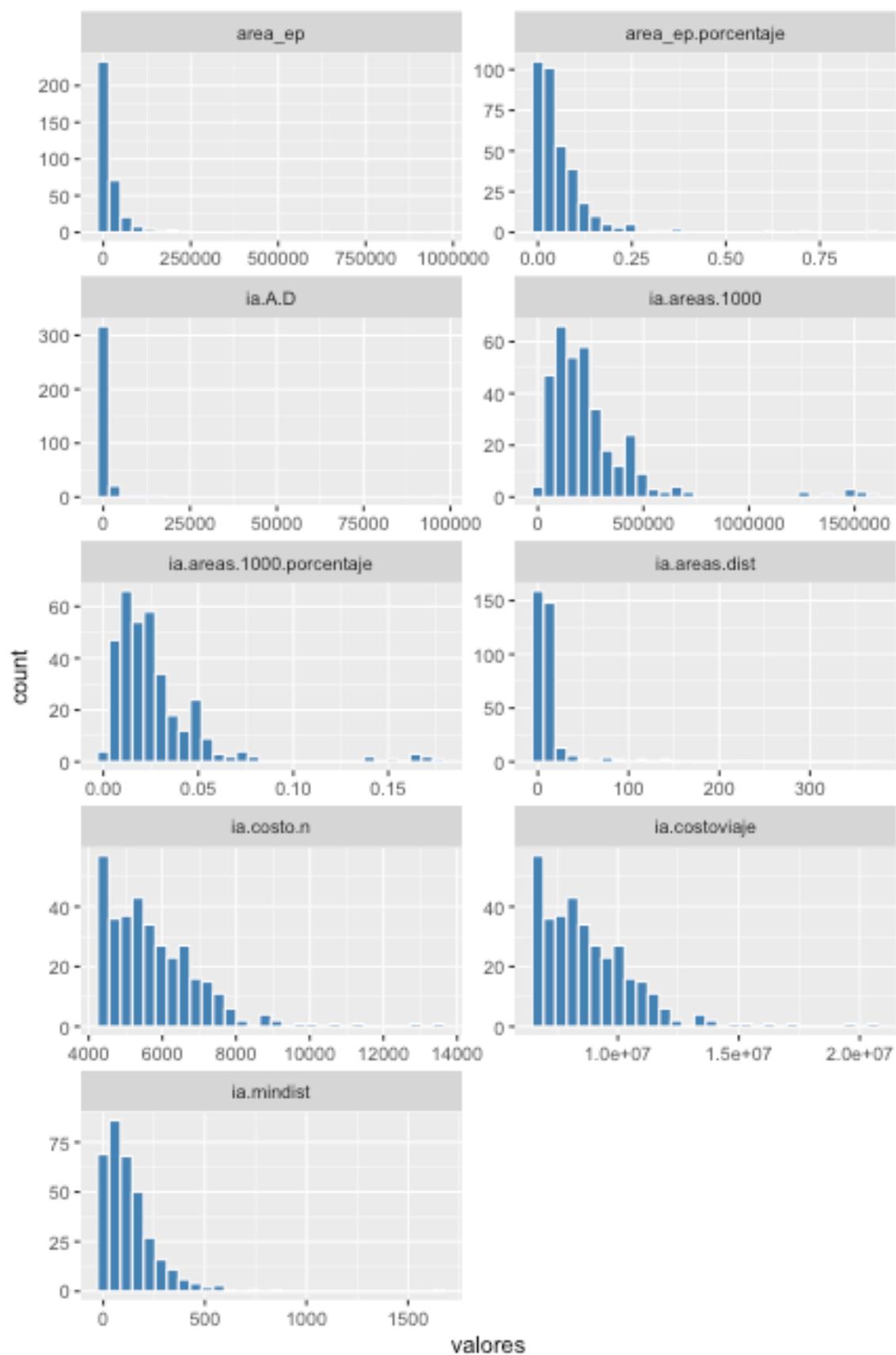


Figure 4.13: Small Multiple de los indices de acceso a EV basados en distancia usando deciles

```

## Min. :0.00000   Min. : 0   Min. :0.00000
## 1st Qu.:0.00964 1st Qu.: 117574 1st Qu.:0.01306
## Median :0.03410 Median : 193016 Median :0.02145
## Mean   :0.05866 Mean   : 249755 Mean   :0.02775
## 3rd Qu.:0.07671 3rd Qu.: 292520 3rd Qu.:0.03250
## Max.   :0.89245 Max.   :1589082 Max.   :0.17657
##
## ia.areas.dist          ia.A.D
## Min. : 0.4185  Min. : 0.0
## 1st Qu.: 3.9850 1st Qu.: 240.8
## Median : 6.7780 Median : 434.4
## Mean   : 15.1800 Mean   : 1119.4
## 3rd Qu.: 11.4793 3rd Qu.: 849.7
## Max.   :370.2725 Max.   :96664.6
## NA's   :1

```

```

# matriz distancia entre centriodes y espacios
m.dist.ctrdsu.ep <- gDistance(ep.cali, centroides.su, byid = T)
# cuando el punto esta dentro del poligono el valor que retorna es 0 por
# ese motivo le pondremos 10 a valores = 0 o entre 0 y 10 para evitar
# problemas al invertir la matriz y mejorar la consistencia del indices con
# valores inversos a la distancia
m.dist.ctrdsu.ep[m.dist.ctrdsu.ep < 10] <- 10
# matriz consicion de estar a 1000 del centriodo
is.1000.ep <- gWithinDistance(ep.cali, centroides.su, 1001, byid = T)
# cantidad de espacio verdes en radio de 1000 m de los centriodes del sector
# urbano
num.ep.1000 <- apply(is.1000.ep, 1, function(x) sum(x, na.rm = T))
# distancias de Espacio publicos a 1000 del centriode de SU
a <- m.dist.ctrdsu.ep * is.1000.ep

```

```

# distancia minima distinta de 0
ia.mindist <- apply(m.dist.ctrdsu.ep, 1, function(x) min(x[x != 0]))
index.min <- apply(m.dist.ctrdsu.ep, 1, function(x) which.min(x[x != 0]))
ia.area.mindist <- ep.calib$area_ep[index.min]/ia.mindist

# suma de las distancias a cada EP por centriode
ia.costoviaje <- apply(m.dist.ctrdsu.ep, 1, sum)

# suma de las distancias a cada EP ubicado a menos de 1000 m del centriode
ia.1000 <- apply(a, 1, function(x) sum(x))
ia.1000[ia.1000 == 0] <- NA
ia.1000.n <- ia.1000/num.ep.1000

# indice de la suma de las areas en el rango de un 1 km del sector censal
ia.Areas.1000 <- is.1000.ep %*% ep.calib$area_ep %>% as.vector()

# indice de area disponible en el radio de 1km como porcentaje del area
# total disponible

ia.Areas.1000.porcentaje <- ia.Areas.1000/sum(ep.calib$area_ep)

# matriz de distancias inversas de centriode su a espacios verdes
m.dist.ctrdsu.1000.ep.inv <- 1/a
b <- m.dist.ctrdsu.1000.ep.inv * is.finite(m.dist.ctrdsu.1000.ep.inv) # eliminar infin
# suma de inverso de las distancias a cada EP ubicado a menos de 1000 m del
# centriode
ia.1000.inv <- apply(b, 1, function(x) sum(x, na.rm = T))
ia.1000.inv[ia.1000.inv == 0] <- NA

# razon entre Area del EP y distancia al centriode
A.D <- t(b) * ep.calib$area_ep

# sumatoria de la razon entre Area del EP y distancias de ese EP al

```

```

# centriode

ia.A.D <- apply(A.D, 1, function(x) sum(x, na.rm = T))
class(ia.costoviaje)
summary(ia.costoviaje)
length(ia.costoviaje)
summary(ia.1000.inv)
length(ia.A.D)
summary(ia.A.D)

# consolidacionde indices calculados

ia.ev <- data.frame(su$SETU_CCDGO, ia.costoviaje)
ia.ev$ia.costo.n <- ia.ev$ia.costoviaje/dim(m.dist.ctrdsu.ep)[2]
ia.ev <- bind_cols(ia.ev, data.frame(ia.1000, ia.1000.inv, ia.1000.n, ia.Areas.1000,
                                      ia.Areas.1000.porcentaje))
ia.ev$ia.r300 <- 300 * ia.1000.inv
ia.ev <- ia.ev %>% dplyr::rename(SETU_CCDGO = su.SETU_CCDGO)
ia.ev$ia.mindist <- ia.mindist
ia.ev$ia.area.mindist <- ia.area.mindist
ia.ev$ia.A.D <- ia.A.D
smry.area <- summary(ep.california$area_ep)
ia.ev$ia.r300.Amedia <- 300/smry.area[4] * ia.ev$ia.A.D
ia.ev$ia.r300.Amediana <- 300/smry.area[3] * ia.ev$ia.A.D
ia.ev$ia.Areas.dist <- ia.Areas.1000/ia.1000

```

## 4.2 Datos del censo arbóreo 2015

El censo arbóreo del año 2015 consolidó un inventario de la vegetación de la ciudad compuesto por 296438 individuos. Entre las variables que categorizan los individuos censados

Table 4.1: Resumen CA2015 por tipo de vegetación

Tipo de vegetación	altura media	diámetro medio	cantidad
Arbol	7.32	7.14	200528
Arbusto	3.34	2.91	42872
Bambu	11.73	3.34	419
Muerto	2.30	1.00	1
Palma	5.53	3.93	49777
Planta arbustiva	2.77	2.83	2176
Seco	5.79	3.63	665

Table 4.2: Resumen CA2015 por edad

Edad	altura media	diámetro medio	cantidad
Juvenil	3.63	3.10	49473
Maduro	6.60	6.11	214042
Longevo	9.39	9.16	32923

están el tipo de emplazamiento, tipo de suelo que cubre la vegetación, la edad, la vitalidad, tipo de vegetación y sus caratetiristicas dasometricas p.ej la altura, el diamtro de la copa, altura y diametro del pecho, etc ..., entro otras relacionadas con el estado fitosanitario y daños físicos.

A continuación se presentan una serie de tablas que resumen las características seleccionadas en la tabla 3.1 para el análisis (por tipo de vegetación 4.1, por edad 4.2 y por emplazamiento 4.3), antes de aplicar los criterios de selección de los individuos arbóres para este estudio.

Existe una diferencia de 10 años entre censo de población de 2005 y el censo arbóreo de la ciudad de Cali. Aunque esto pueda parecer una situación que reduce la legitimidad de los resultados que se hayen en este estudio, autores como Boone et al. (2010) y Schwarz et al. (2015) reconocen que los paisajes que vemos hoy son legados de patrones de consumo pasados, y que en el caso de la vegetación urbana tratamos con organismos de larga vida que pueden tardar mucho tiempo en establecerse y crecer. En contraste, la estructura social de las ciudades puede cambiar más rápidamente.

Table 4.3: Resúmen CA2015 por emplazamiento

Emplazamiento	altura media	diámetro medio	cantidad
Anden	5.96	5.49	136948
Bahias de estacionamiento	6.61	6.23	1927
Bulevares	4.30	3.03	85
Corredor Ferreo	6.99	7.22	1534
Escenario deportivo y/o Cultural	7.14	6.47	32757
Glorieta	6.01	5.69	777
Parque Urbano	6.83	6.29	56137
Paseos	5.81	5.34	2573
Plaza	7.50	6.63	65
Plazoleta	6.92	5.80	1773
Ronda de rios	8.21	7.15	7424
Rondas de canales	6.13	5.88	11721
Separador Vial	6.51	6.32	42717

Table 4.4: Resúmen selección CA2015 por tipo de vegetación

Tipo de vegetación	altura media	diámetro medio	cantidad
Arbol	8.00	7.90	168133
Bambu	9.79	4.25	235
Palma	6.51	4.48	34357
Seco	6.61	5.67	387

Como se menciona en la metodología, la apuesta para reducir la brecha es la exclusión de los árboles jóvenes del inventario, que posiblemente no estaban ahí en 2005. Aunque no conocemos las tasa anual de tala de árboles en la ciudad, y dado es posible que una parte importante de los árboles jóvenes haya reemplazado a los que fueron talados, no parece realista mantener el inventario entero.

Aunque en general toda la vegetación aporta beneficios ambientales a los habitantes, en este estudio descartamos la vegetación arbustiva y los árboles, palmas y bambú de menos de 1.9 m de altura para circunscribirnos a los individuos más desarrollados.

Una vez aplicado este filtro contamos con 203112 individuos. Las tablas de resumen para la selección de individuos con base en estos criterios se muestran a continuación.

Table 4.5: Resúmen selección CA2015 por edad

Edad	altura media	diámetro medio	cantidad
Maduro	7.38	6.90	172776
Longevo	9.85	9.67	30336

Table 4.6: Resúmen selección CA2015 por emplazamiento

Emplazamiento	altura media	diámetro medio	cantidad
Anden	7.25	6.82	92198
Bahias de estacionamiento	8.04	7.66	1348
Bulevares	8.74	8.11	17
Corredor Ferreo	8.56	9.03	1060
Escenario deportivo y/o Cultural	8.25	7.55	24861
Glorieta	7.29	7.06	496
Parque Urbano	8.35	7.81	37756
Paseos	7.42	7.02	1623
Plaza	9.07	7.83	49
Plazoleta	8.47	7.00	1206
Ronda de rios	9.18	8.25	5759
Rondas de canales	7.41	7.23	7815
Separador Vial	7.89	7.82	28924

Para indagar sobre la distribución de estos individuos y no quedarnos con los resúmenes estadísticos se muestran a continuación los datos desagregados gráficamente cada una de las variables categóricas y las características físicas del arbolado: altura, diámetro de la copa y su ubicación en la ciudad.

En primer lugar indagamos sobre las diferencias de diámetro y altura por tipo de vegetación. La figura 4.14 muestra claramente las diferencias físicas entre los árboles (desarrollan mayor tamaño y con mayor número de individuos), el bambú y las palmas (más altos que anchos y en menor número) de la ciudad. Los árboles catalogados como secos, hace 10 años estaban vivos y los mantenemos en la selección de individuos.

Otra característica interesante para buscar condiciones que afectan el desarrollo del arbolado, representado por la altura y el diámetro son el tipo de lugares que conforman el espacio público donde se encuentran el mayor número de ellos. En la figura 4.15 se observa la desagregación de los individuos por tipo de emplazamiento en gráficas individuales de altura y diámetro, y en color el tipo de vegetación. Es notorio en la figura que los parques urbanos y escenarios deportivos son los equipamientos que mayor cantidad de individuos y más desarrollos alojan. Caso aparte son los andenes y separadores viales como se ve en la tabla 4.6 y el gráfico alojan 59.63% de los individuos. Esto puede ser un hecho que incita a incluir elementos estructurales en los modelos que explican desde aspectos estructurales de los barrios la cobertura de copa. Por supuesto las rondas de ríos y canales dada su disponibilidad de agua condicionan el desarrollo de los individuos arbóreos y su cantidad.

Otra forma de ver los datos de resumen de las tabla 4.6 y que completa la visión sobre la distribución de los datos en relación al desarrollo físico de los individuos arbóreos es la distribución de los diámetros (figura 4.16) y las alturas (figura 4.17) en relación con el emplazamiento. En ambas gráficas el punto rojo representa el valor promedio

Finalmente en la figura 4.18 graficamos los individuos en pequeños mapa por tipo de emplazamiento para observar su distribución geográfica. Es notorio que los árboles están distribuidos en toda la ciudad dada la disponibilidad de andenes aptos para alojarlos.

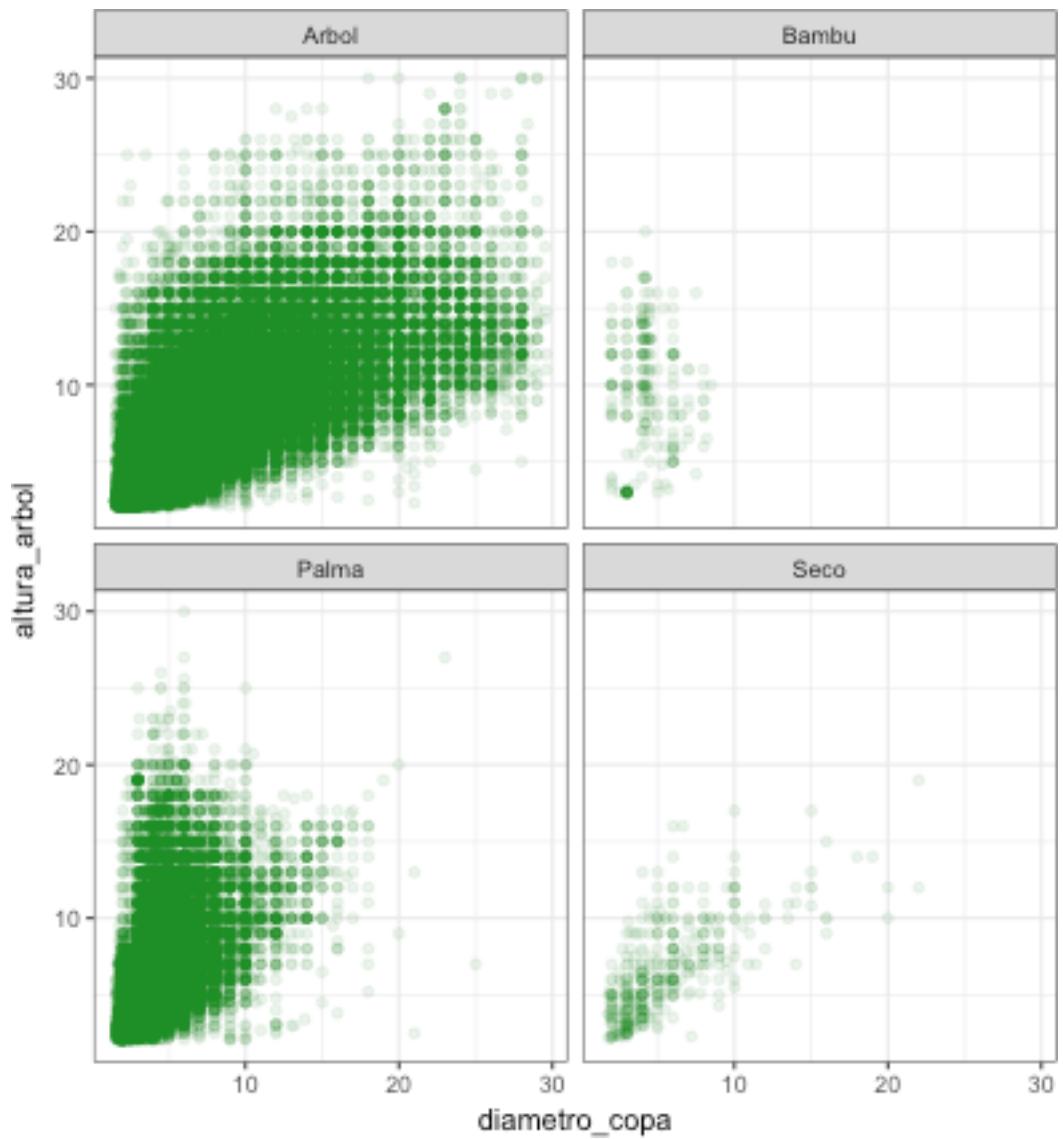


Figure 4.14: caraterísticas por tipo de vegetacion

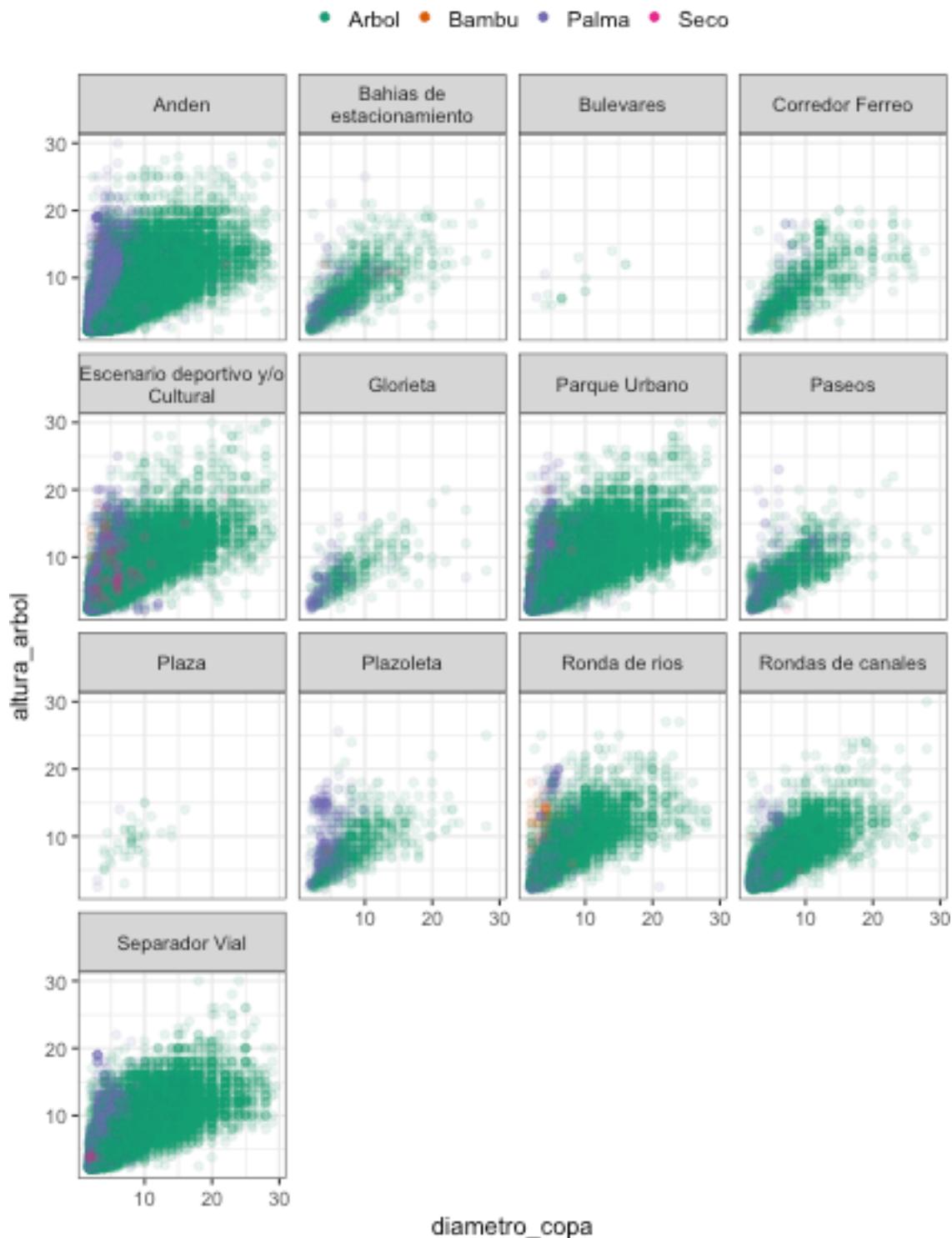


Figure 4.15: Características por tipo de vegetación y emplazamiento

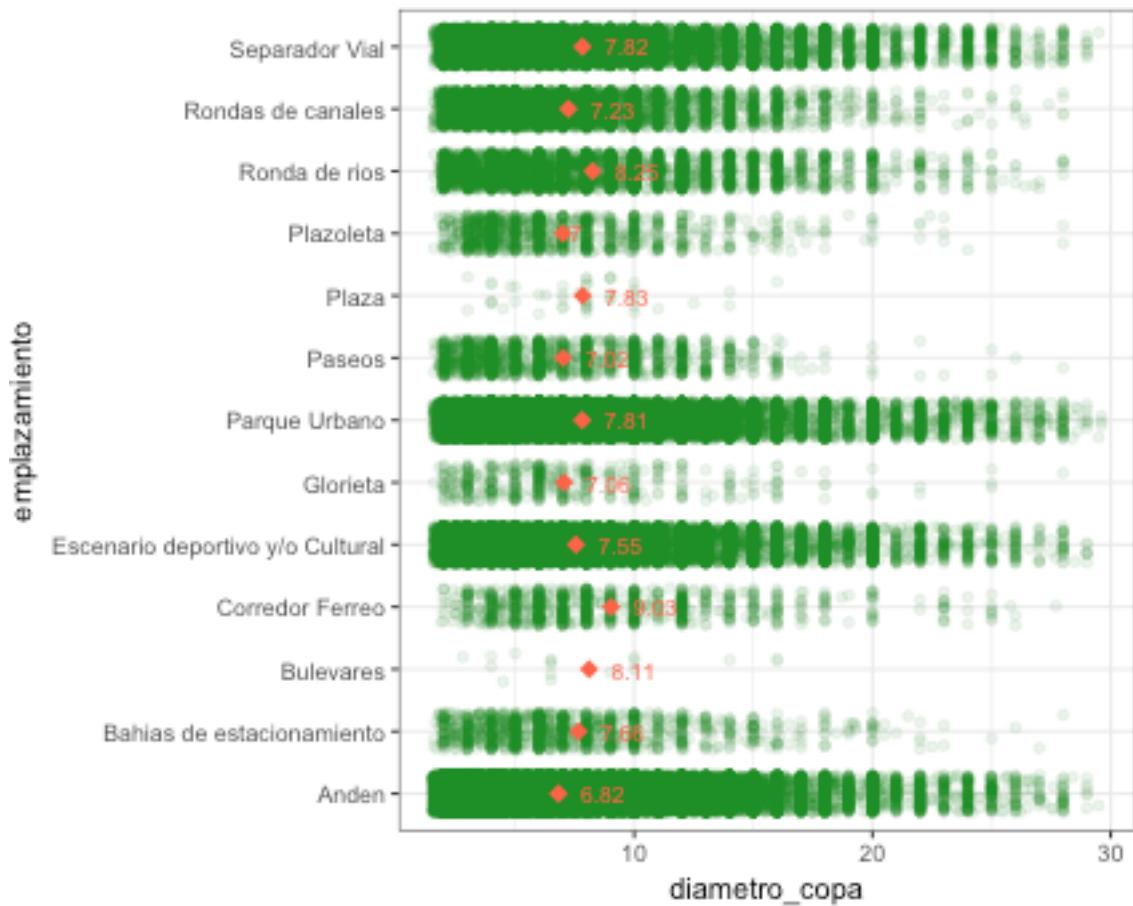


Figure 4.16: Variabilidad del diámetro de copa por emplazamiento



Figure 4.17: Variabilidad de la altura de los árboles por emplazamiento

Esto invita a explorar la relación que pueda existir entre el tamaño de los andenes de los barrios y el desarrollo físico de los individuos, ya que las vías con separadores viales y en consecuencia con más espacio para que árboles de mayor tamaño puedan desarrollarse es consistente con el valor medio de la altura y el diámetro sean mayores que el de los andenes. Sin embargo este hecho escapa del alcance de este trabajo y puede ser indagado por otras investigaciones.

Antes de agregar (enmascarar) los datos usando los sectores censales es interesante inspeccionar el efecto que tiene usar unidades regulares o de tamaños no uniformes como los sectores urbanos en las coberturas de copa. Para ello podemos usar hexágonos de 250 metros de ancho que cubren completamente el territorio. La figura 4.19 se evidencia que existen cinturones y lugares de alta concentración de individuos y en consecuencia de mayor cobertura de copa.

En la figura 4.20 usamos los SUs para agregar los valores de área de copa. Se observa como se reduce un poco la continuidad, y se intensifica el efecto de la agregación en algunos sectores y se atenúa en otros (las figuras 4.21 y 4.22 muestran el efecto de la agregación en el área de copa y en el número de individuos, respectivamente ).

Esto sugiere que usar valores porcentuales para la cobertura en relación con el área de los SU o del área no-privada de cada sector puede ser una mejor medida para caracterizar el beneficio real en cada sector censal. En razón a esta consideración y las reflexiones sobre el tema en Schwarz et al. (2015), que sugiere que la distinción entre áreas privadas y públicas puede ayudar a determinar las superficies plantables totales que están disponibles para aumentar la cobertura de copa. Calculamos entonces el porcentaje de cobertura de copa respecto del área total sector urbano y respecto del área de espacio público (área del sector censal menos el área de las manzanas privadas)<sup>1</sup>. Las medidas porcentuales respecto del área total y pública permiten hacer una comparación más justa entre las diferentes

<sup>1</sup>La capa de espacio público consolidada previamente nos permite identificar las manzanas de un SU que son espacio público, y por tanto podemos obtener el área que es vía pública, las manzanas que son privadas y las manzanas que son espacio público. Así el área pública es igual a la suma del área de calle más las manzanas de espacio público o al área de SU menos el área de manzanas privadas

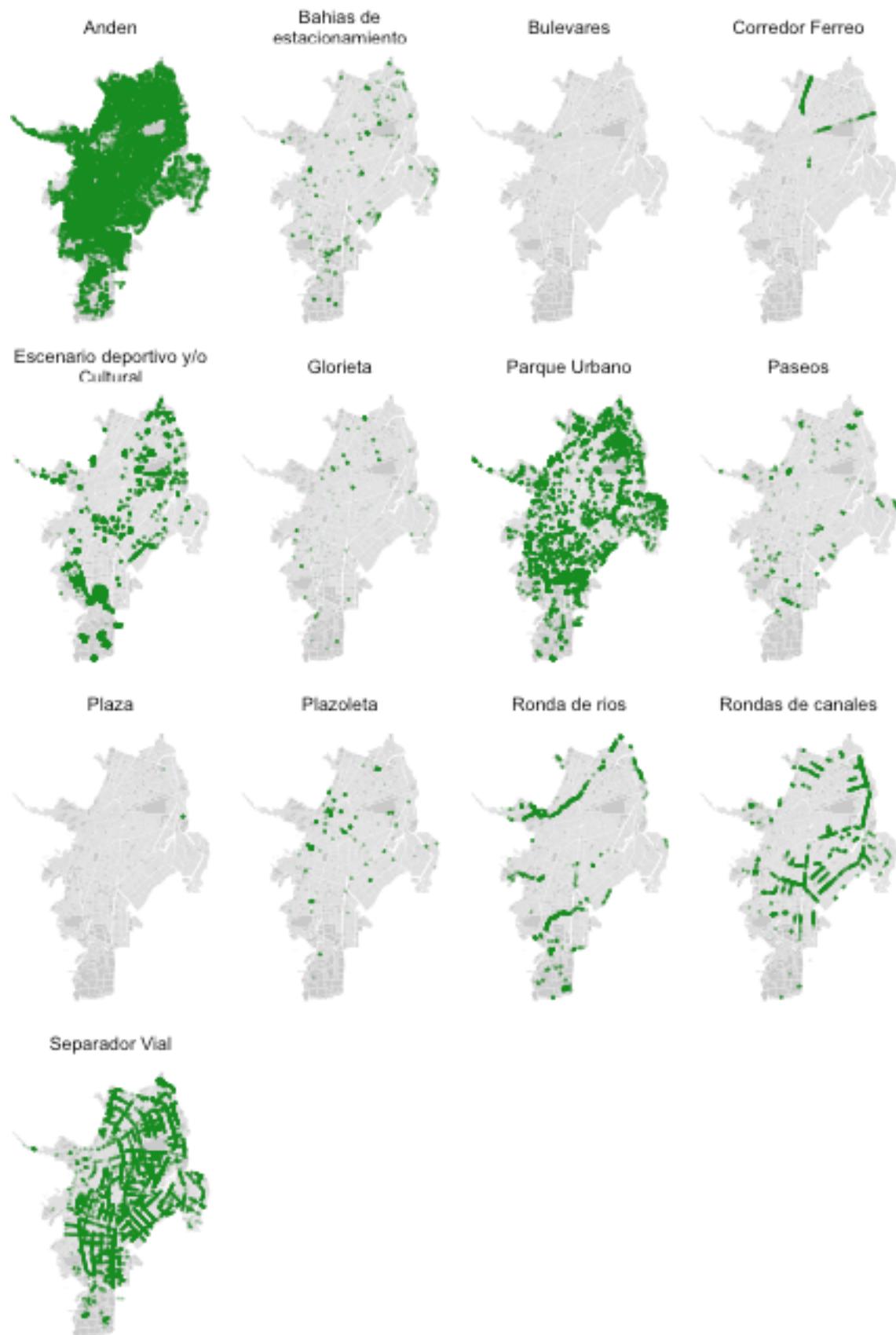


Figure 4.18: Small multiples de los individuos arbóreos por emplazamiento

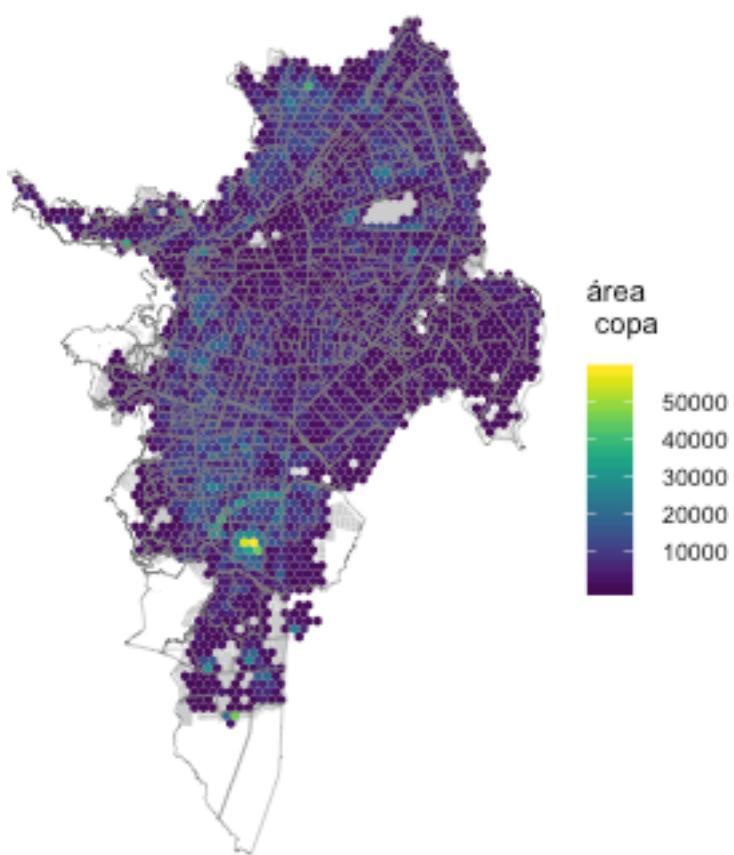


Figure 4.19: Suma de cobertura por hexagonos

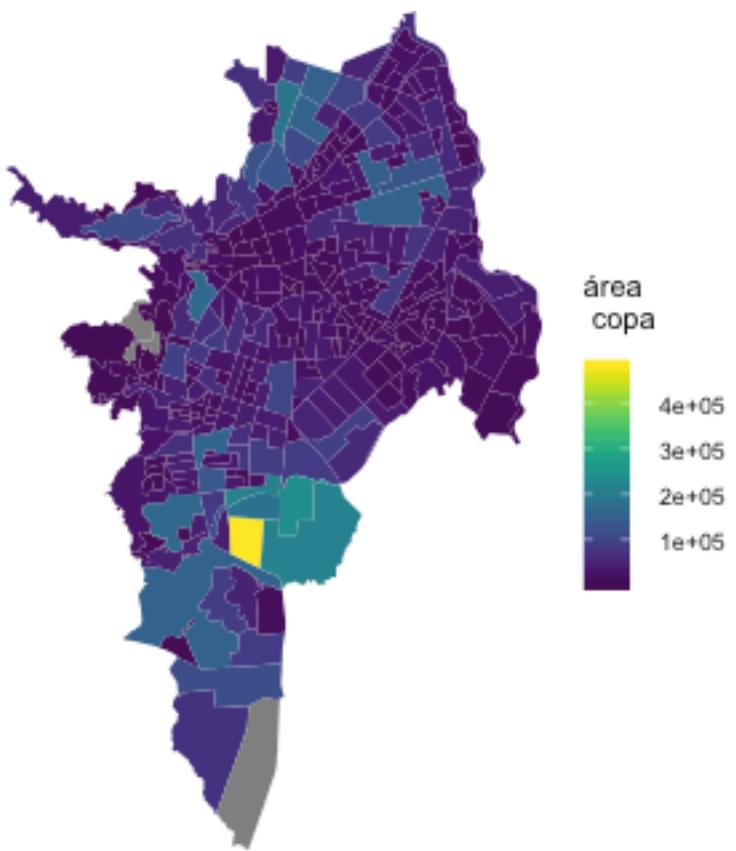


Figure 4.20: Área de copa por sector censal

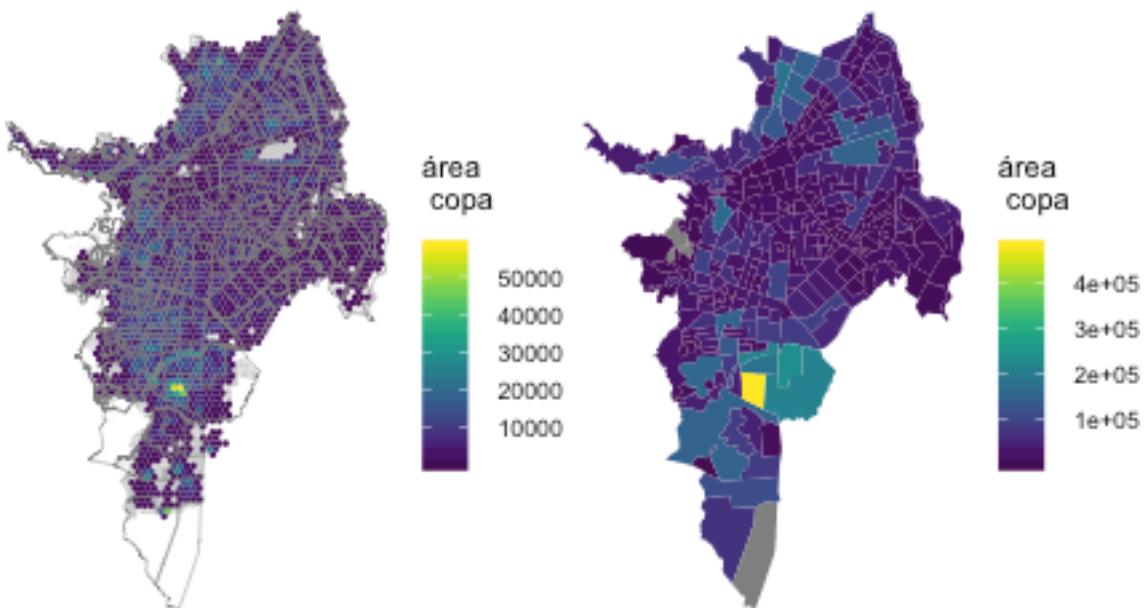


Figure 4.21: Agregación de área de copa por hexagonos y SU

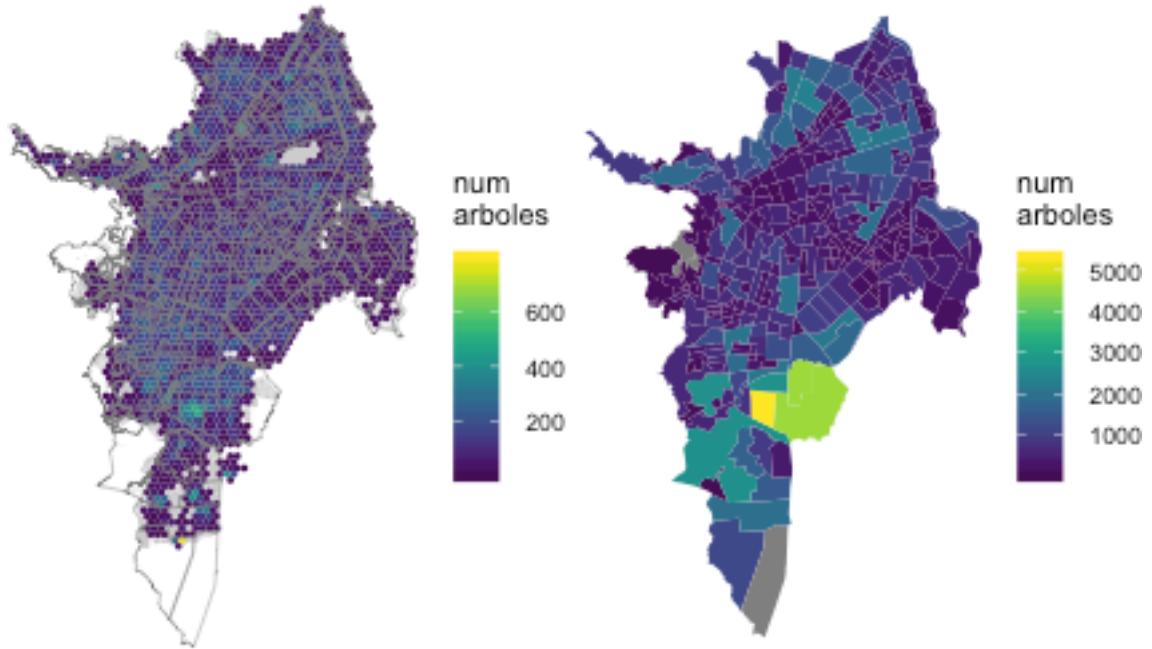


Figure 4.22: Agregación de número de árboles por hexagonos y SU

unidades pues relativiza los niveles totales de área de copa. Un hecho que apoya el uso de medidas en relación al espacio público de un sector censal es que el CA2015 solo se realizó para la vegetación en lugares públicos, sobre la calle o vía pública.

En consecuencia se calcularon las métricas de área de copa en relación al área del sector urbano (`cobertura_copa.su`) y al área pública del sector urbano (`cobertura_copa.ap`). En la figura 4.23 se ven los mapas en escala continua y en la figura 4.24 se reproducen los mismos mapas usando una escala en deciles. Visualmente, cobertura de copa parece en espacio público parecerse mejor a los patrones de distribución que se evidencia cuando usamos la división uniforme del terreno en hexágonos (figura 4.19), razón por la cual la preferiremos sobre `cobertura_copa.su` para los análisis.

Finalmente se muestra los histogramas y el resumen de las métricas de cobertura arbórea calculadas.

```
##      area_copa          cobertura_copa.su  cobertura_copa.ap
##  Min.    : 429.2   Min.    :0.000769   Min.    :0.000877
```

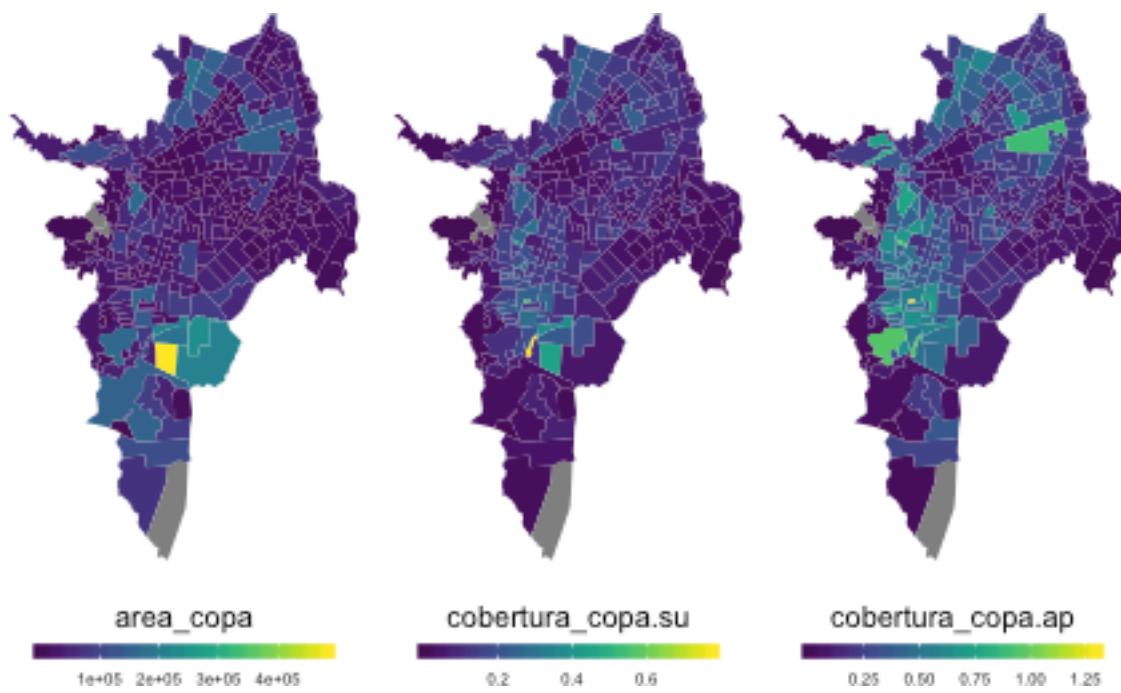


Figure 4.23: Métricas de cobertura de copa: área neta, porcentaje respecto del sector censal y porcentaje respecto del área pública

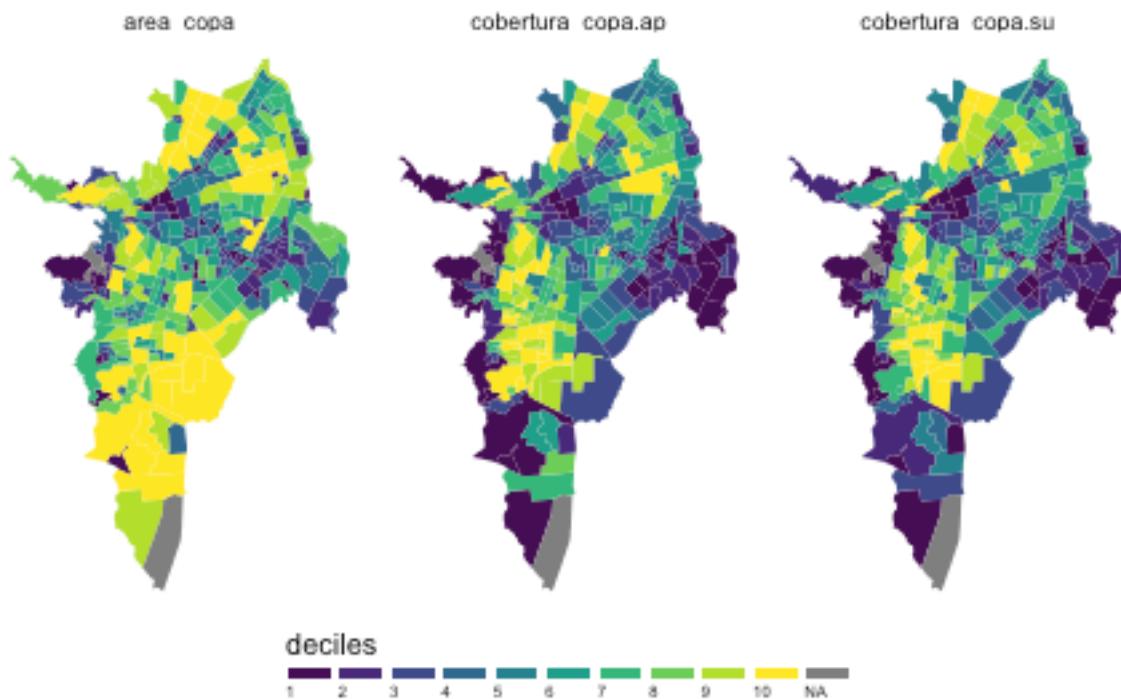


Figure 4.24: Métricas de cobertura de copa: área neta, porcentaje respecto del sector censal y porcentaje respecto del área pública

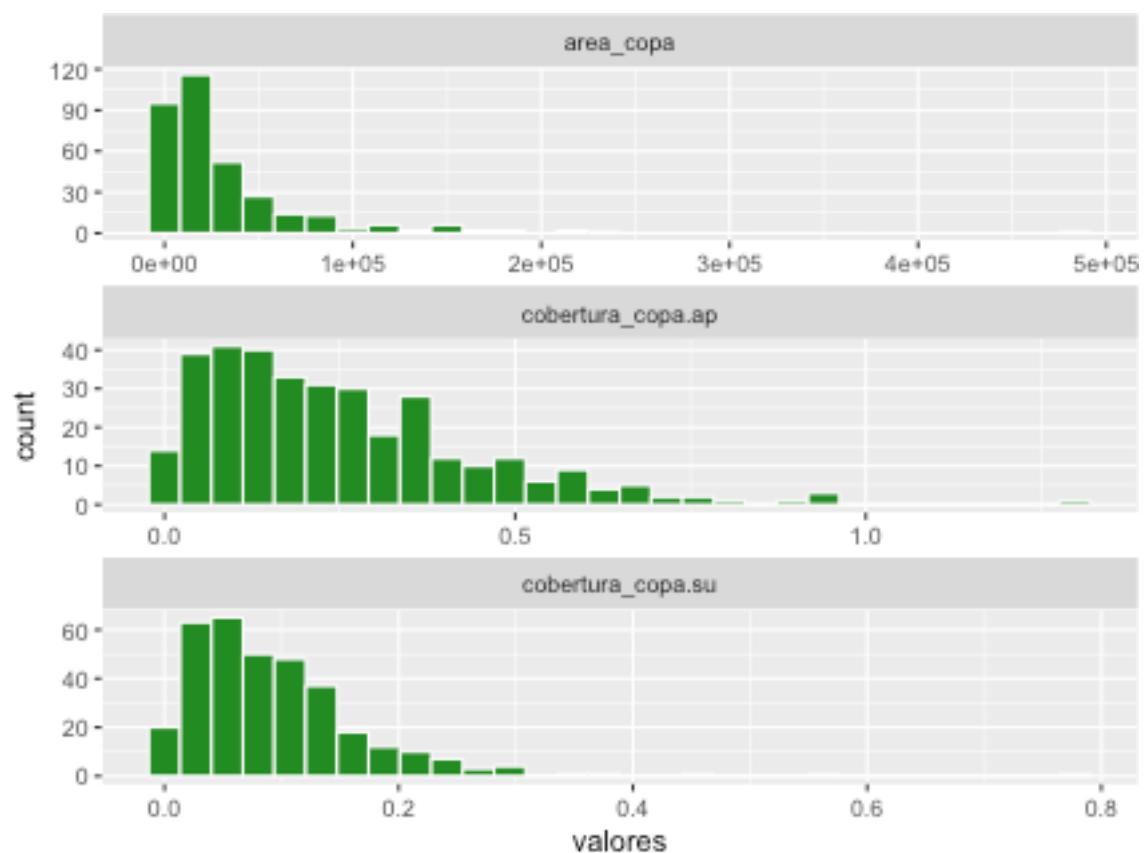


Figure 4.25: Histograma de las métricas de la cobertura de copa

```

## 1st Qu.: 7708.5    1st Qu.:0.040890    1st Qu.:0.101541
## Median : 17651.6    Median :0.076651    Median :0.208246
## Mean    : 32949.9    Mean    :0.095847    Mean    :0.251747
## 3rd Qu.: 39116.3    3rd Qu.:0.127404    3rd Qu.:0.349845
## Max.    :483719.9    Max.    :0.778169    Max.    :1.298912
## NA's    :4           NA's    :4           NA's    :4

```

## 4.3 Datos del censo de población

Los datos del censo fueron descargados del aplicativo con que cuenta el DANE para dar acceso al censo de población 2005 (cen). Los datos agregados por sectores censales los inspeccionamos a través de los resúmenes estadístico, histogramas y mapas, de forma análoga a lo realizado hasta ahora con el resto de variables.

### 4.3.1 Característica de la población

La tabla 3.3 resumen las variables consideradas inicialmente en este trabajo, sin embargo, algunas de ellas no contienen suficiente variabilidad o el número de individuos es muy bajo en comparación con el total de la población. En la tabla 4.7 se observa el bajo número de personas que pertenecen al pueblo Rom (gitanos), Palenqueros de San Basilio (departamento de Bolívar) y de Raizales del Archipiélago de San Andrés, Providencia y Santa Catalina (SAI), por lo que son descartados del análisis. La población indígena es también baja, pero no tanto como para descartarla inmediatamente.

La figura 4.26 muestra los datos de las variables de población en espacio geográfico de la ciudad usando una escala continua y la figura 4.27 lo hace usando una escala discreta (deciles).

Los histogramas de estas variables en la figura 4.28 siguen la tendencia que se ha venido

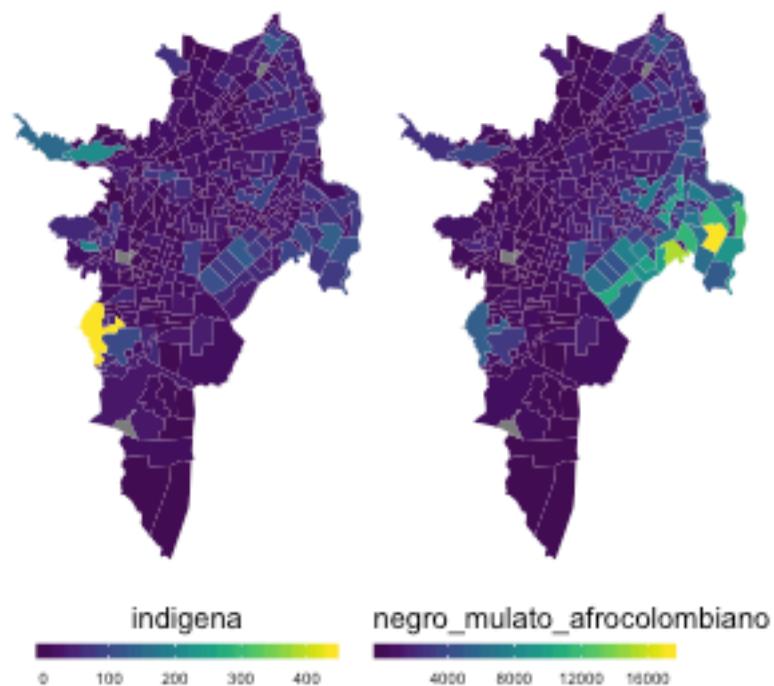
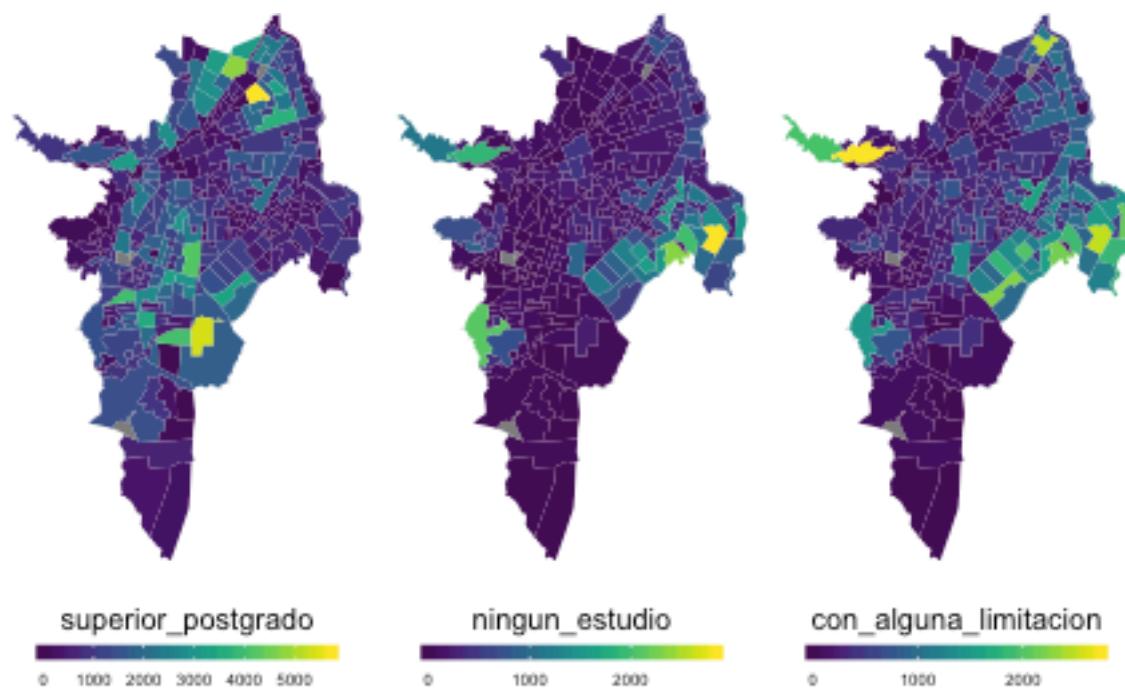


Figure 4.26: Mapas de las variables de población seleccionadas (escala continua)

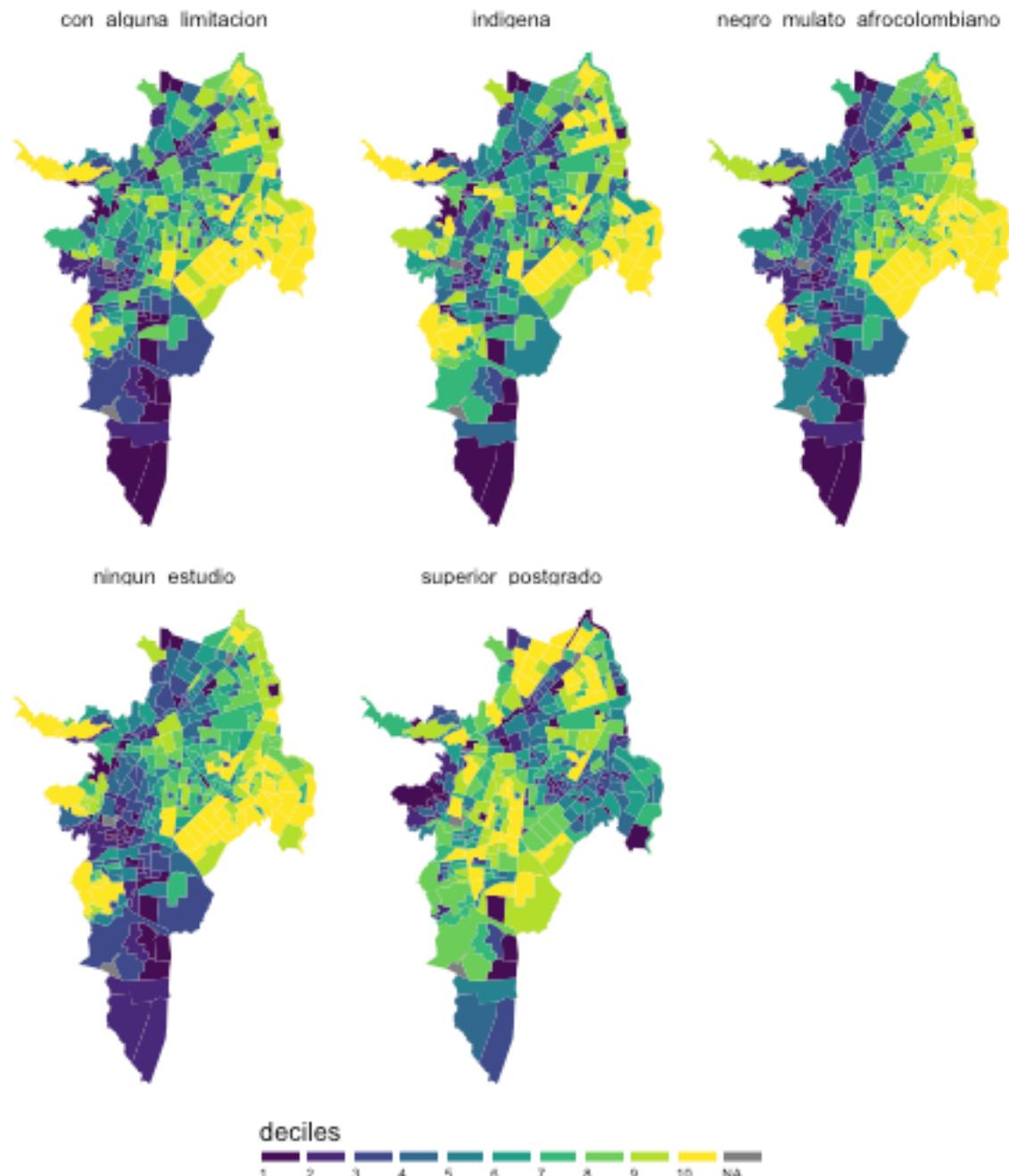


Figure 4.27: Mapas de las variables de población seleccionadas (en deciles)

Table 4.7: Totales de población en la ciudad de Cali

Tipo	Cantidad
Población Total	2027024
Población afrodescendiente, negros o mulatos	530990
Población indígena	9195
Población Rom	690
Población Palenqueros	1
Población raizales de SAI	851

observando en los histogramas de cobertura arbórea y de variables de acceso a espacios verdes : las distribuciones no son *normales*, tienen una inclinación a la derecha.

Además de las variables seleccionadas podemos calcular indicadores que se relacionan con el teóricamente como la densidad de población: dado que los árboles compiten por el espacio con los seres humanos es de esperarse que a mayor cantidad de personas haya menos lugar para los árboles. Podemos de nuevo calcular indicadores porcentualización de las condiciones de la población para facilitar la comparaciones y acentuar las diferencias entre los diferente sectores.

A continuación se muestran los mapas en escala continua (figura 4.29), discreta (figura 4.30) e histogramas (figura 4.31) de los indicadores porcentuales de las condiciones y la densidad de población.

Podemos notar que la transformación realizada al expresar las variables como porcentaje reduce la inclinación hacia la derecha de los histogramas e inclusive la corrige para el caso de personas con alguna limitación.

Para finalizar con la inspección de los datos sobre la población se proveen el resumen estadístico de las variables.

```
## superior_postgrado ningun_estudio con_alguna_limitacion
## Min.   : 0.0   Min.   : 0.0   Min.   : 0.0
## 1st Qu.: 234.8  1st Qu.: 43.0  1st Qu.: 102.0
## Median : 648.5  Median : 113.0  Median : 261.5
```

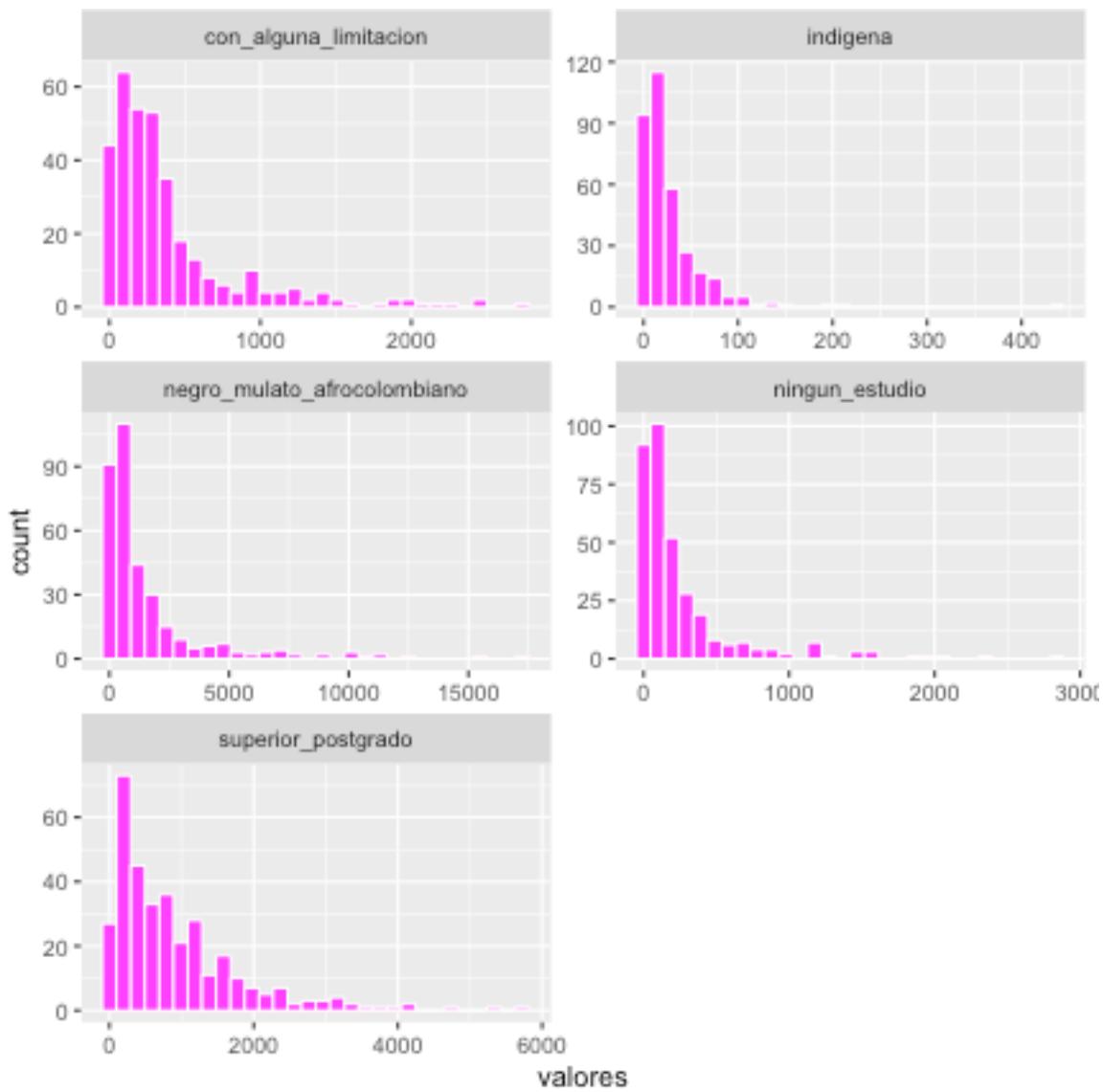


Figure 4.28: Histogramas de las variables de población

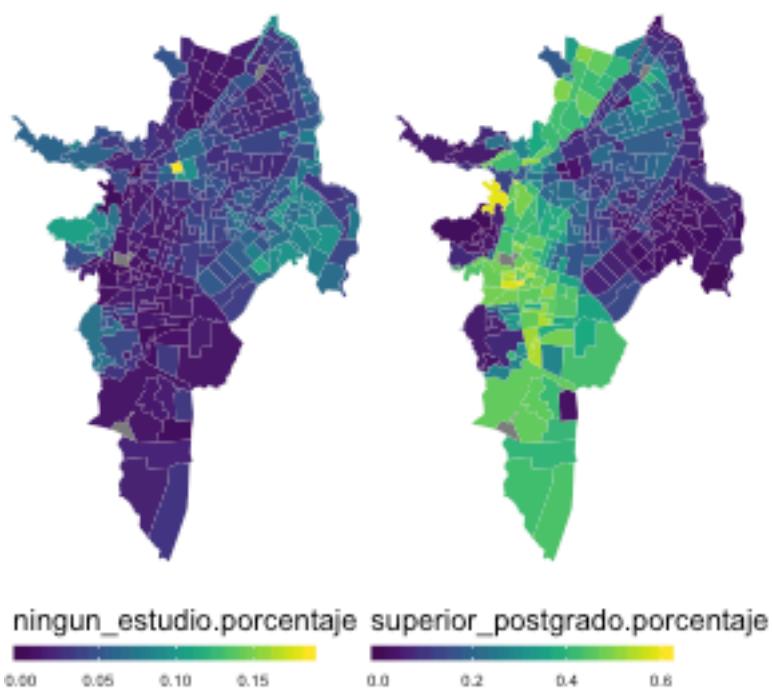
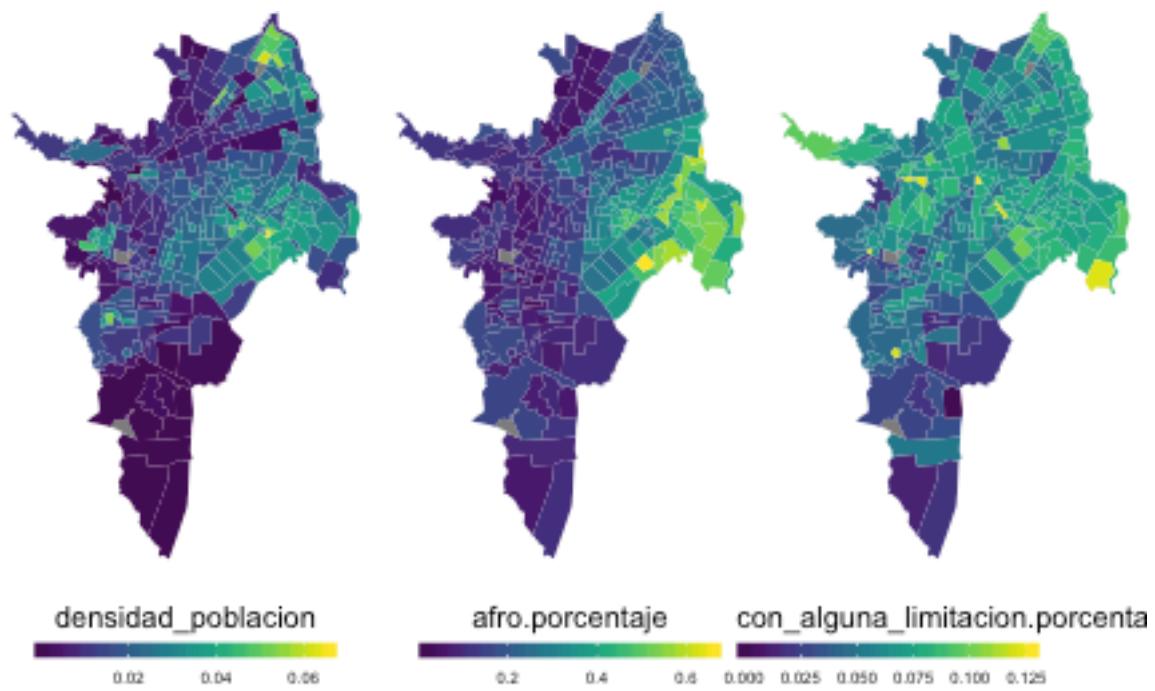


Figure 4.29: Mapas de las variables de población seleccionadas como porcentajes (escala continua)

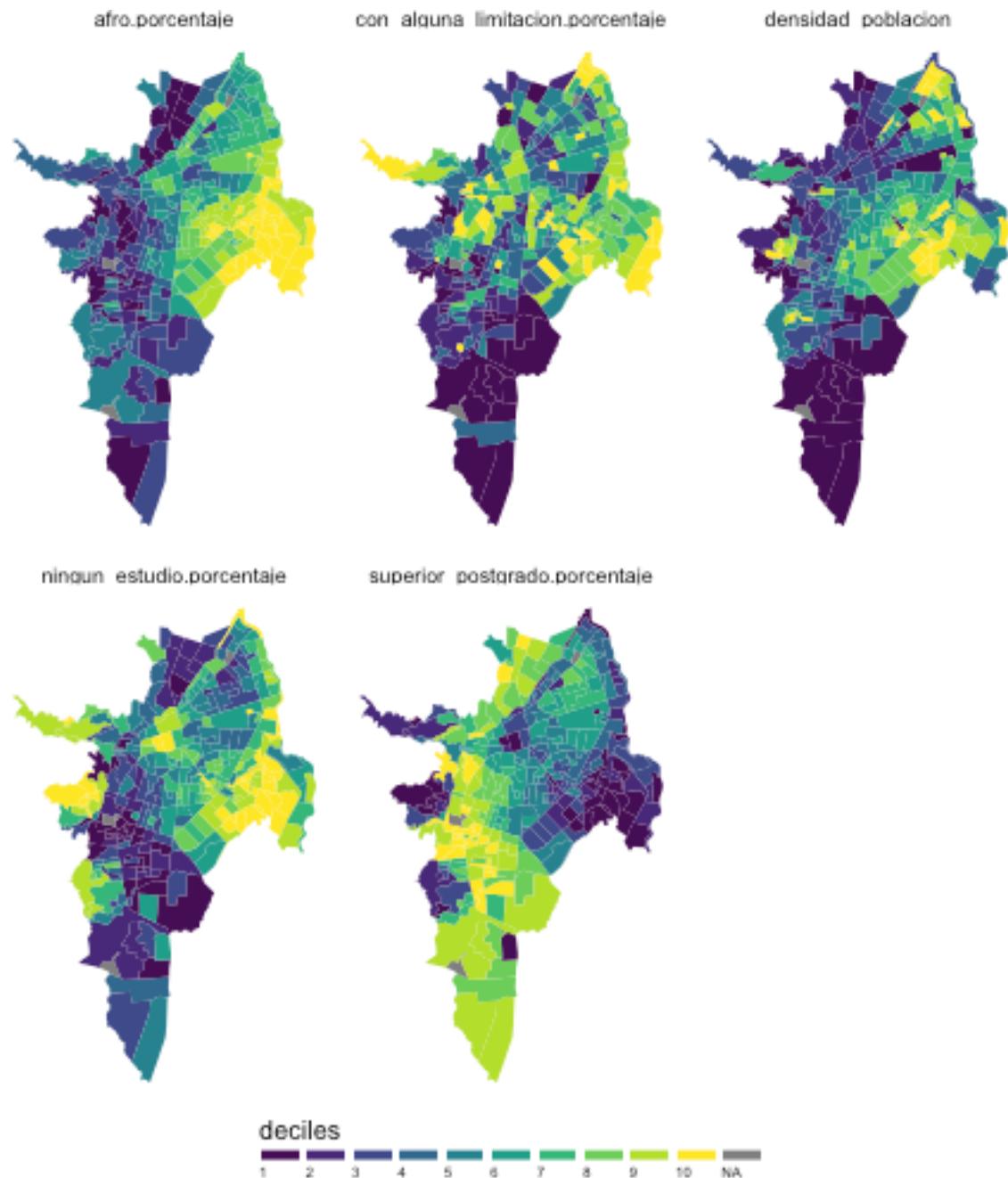


Figure 4.30: Mapas de las variables de población seleccionadas como porcentajes (en deciles)

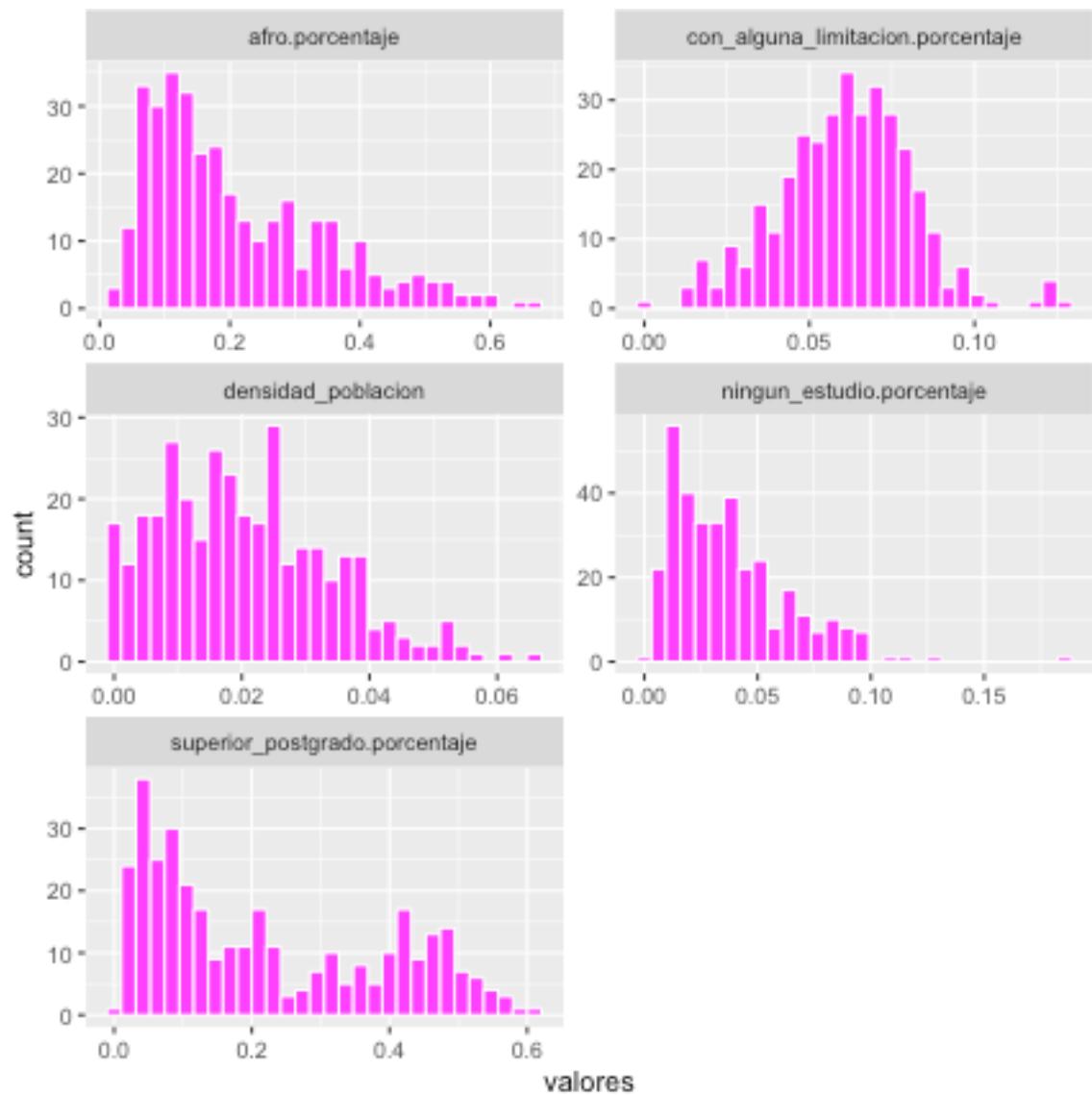


Figure 4.31: Histogramas de las variables de población como porcentaje

```

##   Mean    : 907.2      Mean    : 254.3      Mean    : 394.5
##   3rd Qu.:1229.8      3rd Qu.: 282.5      3rd Qu.: 454.8
##   Max.    :5728.0      Max.    :2838.0      Max.    :2736.0
##   NA's    :4           NA's    :4           NA's    :4
##   indigena      negro_mulato_afrocolombiano densidad_poblacion
##   Min.    : 0.00      Min.    :     1.0      Min.    :0.000033
##   1st Qu.:  6.00      1st Qu.: 273.5      1st Qu.:0.009483
##   Median  : 16.00      Median : 661.0      Median  :0.018837
##   Mean    : 26.89      Mean    : 1552.6      Mean    :0.020423
##   3rd Qu.: 34.00      3rd Qu.: 1664.5      3rd Qu.:0.028885
##   Max.    :437.00      Max.    :17264.0      Max.    :0.065882
##   NA's    :4           NA's    :4           NA's    :4
##   afro.porcentaje  con_alguna_limitacion.porcentaje
##   Min.    :0.01683     Min.    :0.00000
##   1st Qu.:0.10407     1st Qu.:0.04866
##   Median  :0.17150     Median :0.06197
##   Mean    :0.21087     Mean    :0.06109
##   3rd Qu.:0.28770     3rd Qu.:0.07422
##   Max.    :0.66255     Max.    :0.12729
##   NA's    :4           NA's    :4
##   ningun_estudio.porcentaje superior_postgrado.porcentaje
##   Min.    :0.00000     Min.    :0.00000
##   1st Qu.:0.01690     1st Qu.:0.06914
##   Median  :0.03161     Median :0.16845
##   Mean    :0.03766     Mean    :0.22020
##   3rd Qu.:0.05050     3rd Qu.:0.38509
##   Max.    :0.18548     Max.    :0.61056
##   NA's    :4           NA's    :4

```

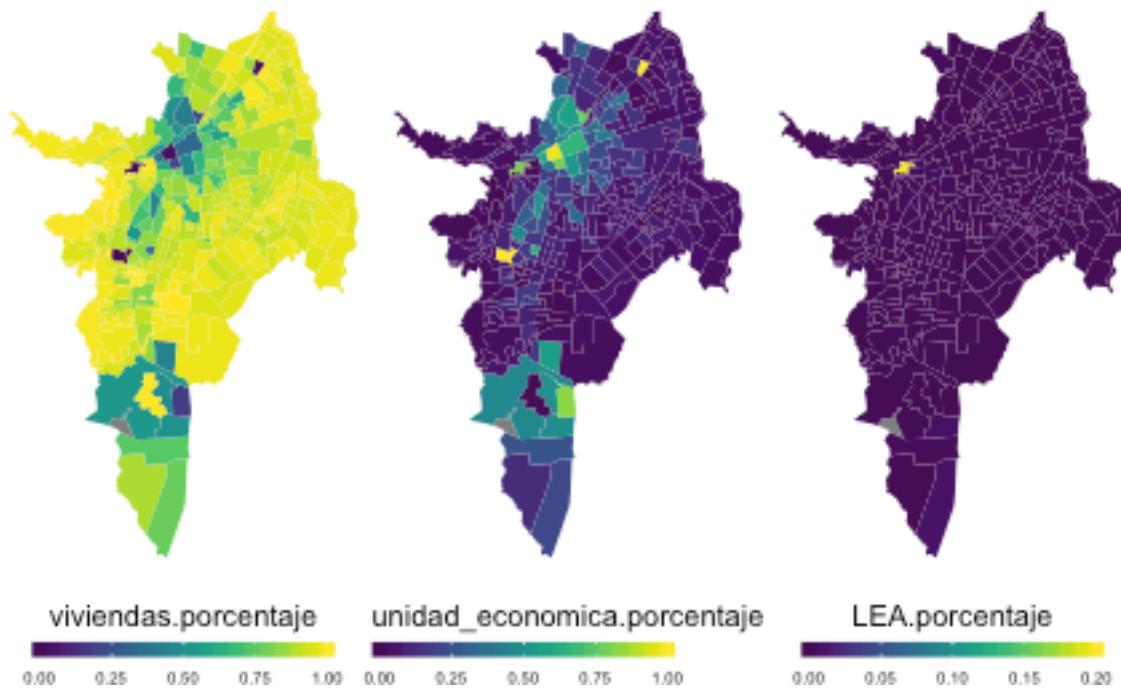


Figure 4.32: Mapas de las variables sobre el tipo de uso de los predios como porcentaje de la cantidad de predios (escala continua)

### 4.3.2 Características de las viviendas

Además de los rasgos étnicos, condiciones de estudio y limitaciones de la población el censo de 2005 tiene disponibles datos sobre el tipo de viviendas (casa, apartamento, tipo cuarto, casa indígena, otros), y el uso habitacional, comercial y la cantidad de unidades especiales de alojamiento L.E.A dado a los predios. La vocación comercial o residencial de un barrio puede ser un factor en el desarrollo del arbolado urbano, ya sea por las condiciones físicas como por la intervención de sus habitantes. Estas variables pueden también expresarse como porcentaje de la cantidad de predios de vivienda en el caso de los tipos o como porcentaje de la cantidad de predios en el caso del uso como unidad de vivienda, económica o L.E.A.

A continuación presentamos el resumen, los mapas por sector urbano (figuras 4.32 y 4.33) y los histogramas (figura 4.34) de las variables sobre el uso de los predios.

El uso de L.E.A tiene una distribución concentrada en uno pocos SU, por lo que podemos

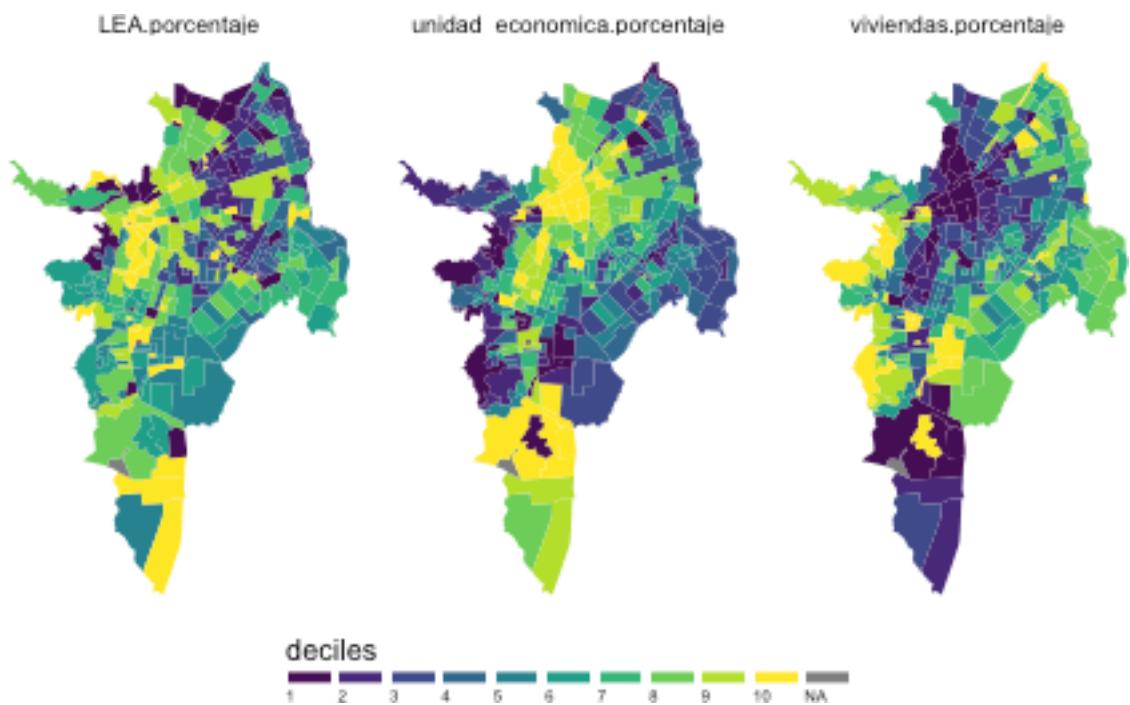


Figure 4.33: Mapas de las variables sobre el tipo de uso de los predios como porcentaje de la cantidad de predios (en deciles)

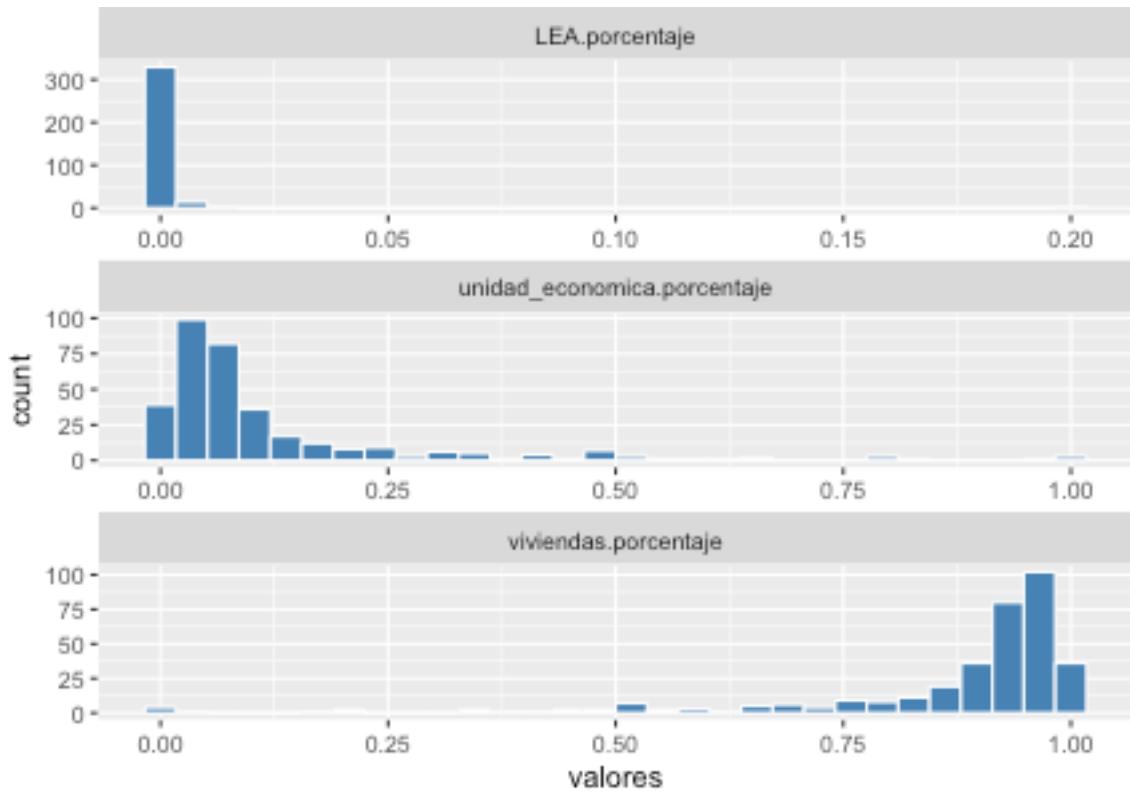


Figure 4.34: Histogramas de las variables de uso de predios como porcentaje

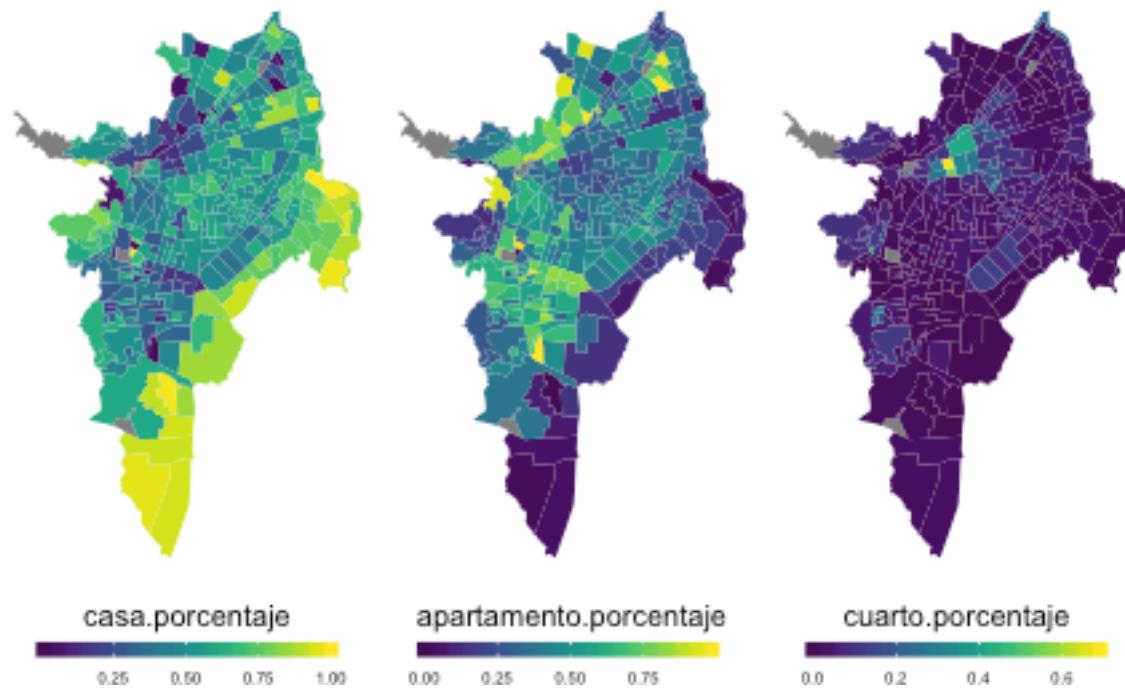


Figure 4.35: Mapas de las variables sobre el tipo viviendas como porcentaje (escala continua)

descartarla para los análisis de regresión. Existe también cierta complementariedad entre el uso de vivienda y los usos económicos de los predios, porque seguramente, si existe una correlación entre estas variables y la cobertura de copa o el acceso a espacios verdes una de las dos puede bastar para incluir esta dimensión en los modelos de regresión.

A continuación presentamos el resumen, los mapas por sector urbano (figuras 4.35 y 4.36) y los histogramas (figura 4.37) de las variables sobre los tipos de vivienda.

Esta primera inspección a los datos permitió descartar algunas variables a usar en los modelos de regresión sobre las coberturas y los espacios verdes, conocer la distribución espacial de los datos y perfilar las posibles variables a incluir. Sin embargo, la percepción suele ser engañosa, vemos patrones en todas partes, así que es necesario acompañar estas intuiciones con métricas estadísticas y gráficas sobre las relaciones entre las variables dependientes e independientes para ser más asertivos en las decisiones del proceso de modelado.

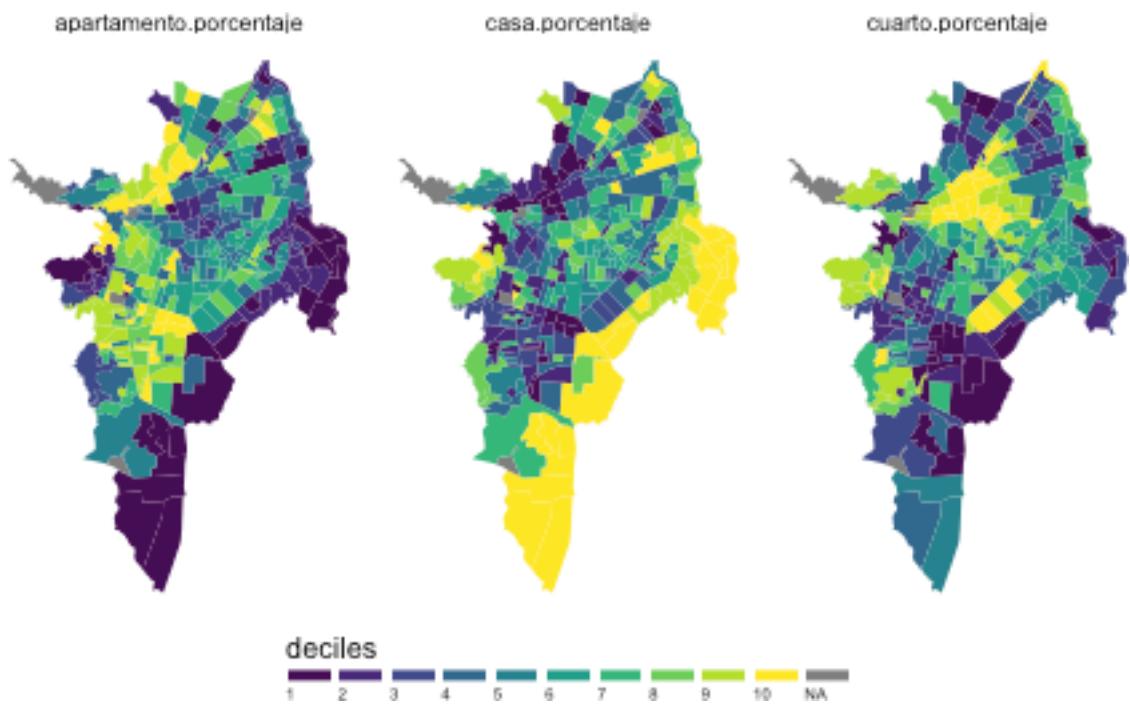


Figure 4.36: Mapas de las variables sobre el tipo viviendas como porcentaje (en deciles)

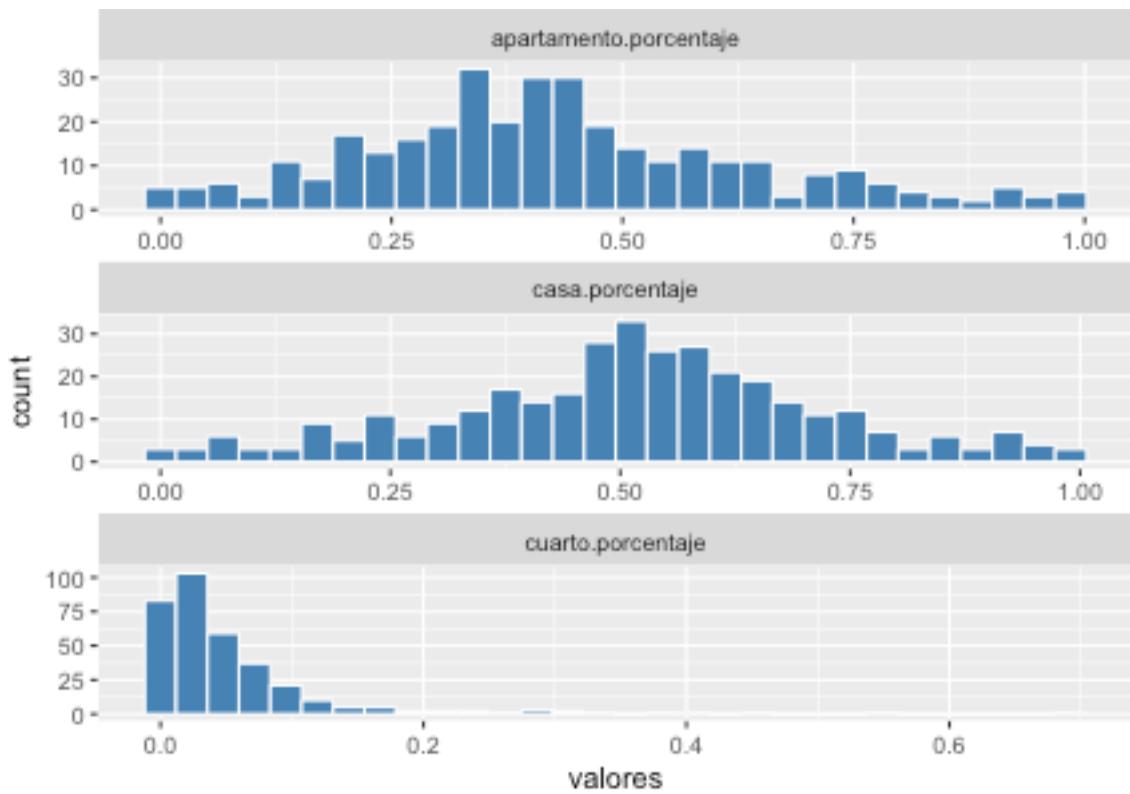


Figure 4.37: Histogramas de los tipo de vivienda como porcentaje

## 4.4 Análisis estadísticos

### 4.4.1 Criterios y selección de sectores censales

Antes de iniciar un analaisis de regresion es importante establecer ciertos criterios pra inclusion o no de ciertos datos dentro del conjunto de varores para la regresion y calculo de la correlacion. Estos criterios estan ligados creterios de excepcion de sectores urbanos a ser incluidos en el analisis de regresion:

- sectores sin personas
- sectores sin viviendas
- sectores area de espacio publico mayor que el 60 % del area del sector
- sectores area de calle mayor que el 80 % del area del sector
- sectores area privada mayor que el 90 % del area del sector

```
analisis.cali.df %>%
  filter(is.na(personas_edad)) %>%
  select(SETU_CCDGO) -> sin_personas

analisis.cali.df %>%
  filter(uso_vivienda == 0) %>%
  select(SETU_CCDGO) -> sin_viviendas

analisis.cali.df %>%
  filter(area_ep.porcentaje > 0.6 ) %>%
  select(SETU_CCDGO) -> ep_60

analisis.cali.df %>%
  filter(area_privada.porcentaje > 0.85 ) %>%
  select(SETU_CCDGO) -> privada_85
```

```

analysis.cali.df %>%
  filter(area_calle.porcentaje > 0.8 ) %>%
  select(SETU_CCDGO) -> calle_80

analysis.cali.df %>%
  filter(is.na(area_copa)) %>%
  select(SETU_CCDGO) -> sin_arboles_censado

su.exc.apriori<-c("0204","1736",# sector con alto porcentaje no urbanizado
                  "1709",# mayoria del area por fuera del perimetro urbano
                  "1317")# laguna del pandaje

```

Además de estos criterios se excluyeron los sectores donde está la Laguna el Pondaje, que cubre una porción muy importante del sector que no se ve reflejado en las otras métricas, los sectores con una porción mayor al 60% por fuera del perímetro urbano o sin urbanización visible en las imágenes satelitales. Así los sectores excluidos del análisis se muestran en los mapas 4.38 y 4.39 por criterio usado.

Para el análisis de las zonas verdes, no se tiene en cuenta el criterio de exclusión de SU sin árboles y se incluye la Laguna del Pondaje. Los mapas de las zonas se ven en la figura 4.40 y 4.41.

#### 4.4.2 Modelando la cobertura de copa

Las variables a incluir en los modelos lineales deben cumplir una serie de condiciones para ser elegidas como candidatas:

- *Mostrar una correlación fuerte* (típicamente mayor a 0.6 se considera una asociación

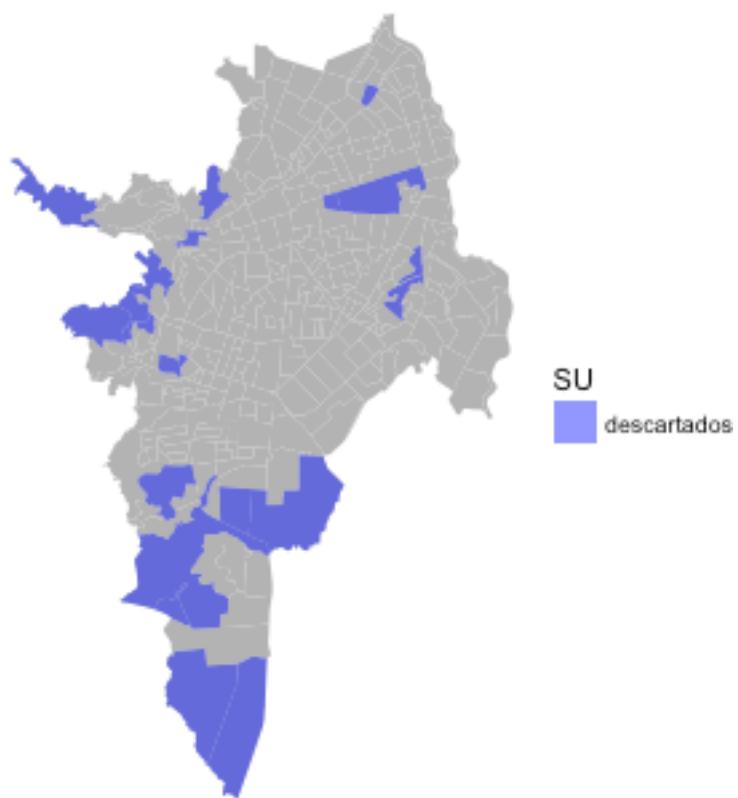


Figure 4.38: Mapa de los sectores excluidos

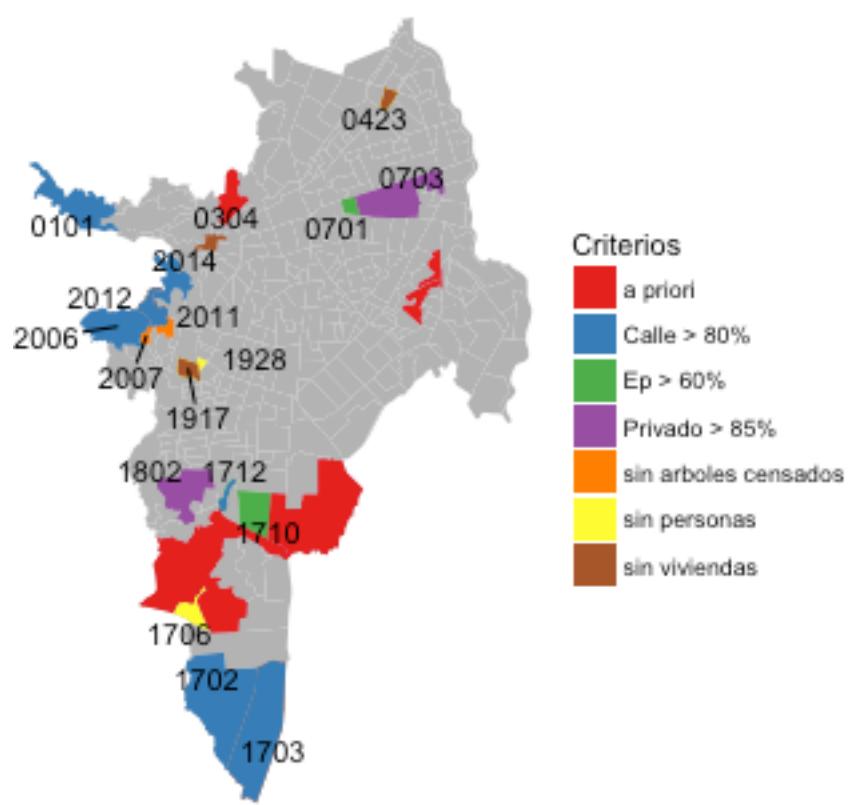


Figure 4.39: Mapa de los sectores excluidos por criterio usado

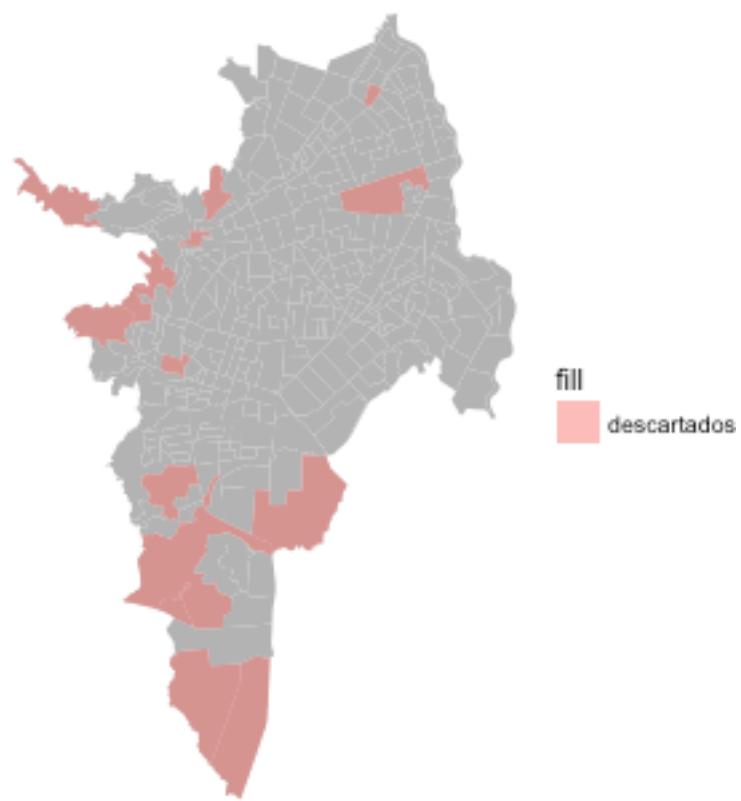


Figure 4.40: Mapa de los sectores excluidos para el análisis de EV

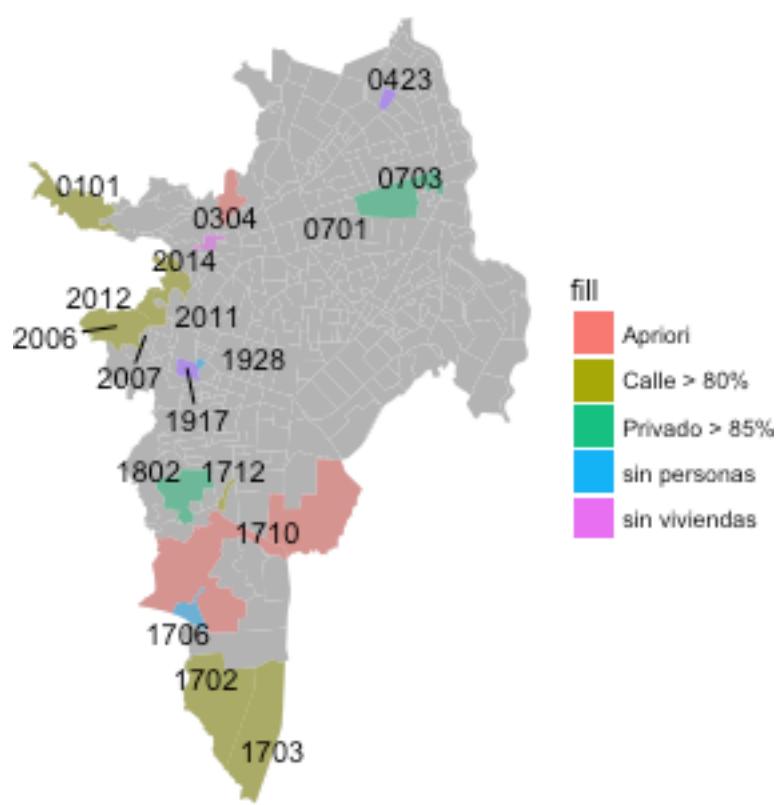


Figure 4.41: Mapa de los sectores excluidos por criterio usado para el análisis de EV

fuerte).

- *Las variables independientes o predictoras no deben estar fuertemente correlacionadas entre ellas.*
- *Las observaciones deben ser independientes.* En nuestro caso significa que no debe existir relación espacial o temporal entre los diferentes sectores. Justamente esto se pondrá a prueba con los test estadísticos y los graficos de diagnostico sobre la distribución de los residuos de la regresión: se espera que dicha dependencia esté motivada por la vecindad de los sectores.
- *Las variables dependientes e independientes deben tener una distribución normal.* Esta condición no suele ser estricta, pues lo importante es que al calcular los coeficientes de la regresión obtengamos una distribución normal de los residuos (sin ningún patrón, ruido). De no ser así, es posible que las variables no sean independientes o que exista información significativa en los residuos, por ejemplo, porque existe autocorrelación espacial en la variable dependiente y entonces la regresión lineal no obtiene resultados confiables para los coeficientes.

Para hacer más tratable y gradual el proceso de complejizar el modelo iniciaremos el análisis con las variables sobre la población, que son las de mayor interés en un estudio dado su enfoque en la justicia ambiental, para luego incorporar las variables de los dominios relacionados con el uso de los predios, los tipos de viviendas y la existencia de espacios verdes como parques, bulevares, escenarios deportivos o plazas, que como se vio en la sección 4.2 (revisar figura 4.15) alojan una cantidad considerable de los individuos arbóreos de la ciudad.

Las variables a modelar son la área de copa en metros<sup>2</sup> (**area\_copa**) y la cobertura de copa como porcentaje del área pública total (**cobertura\_copa.ap**), conformda por la vías y calles más el área de espacio públicos) (ver figura 4.42).

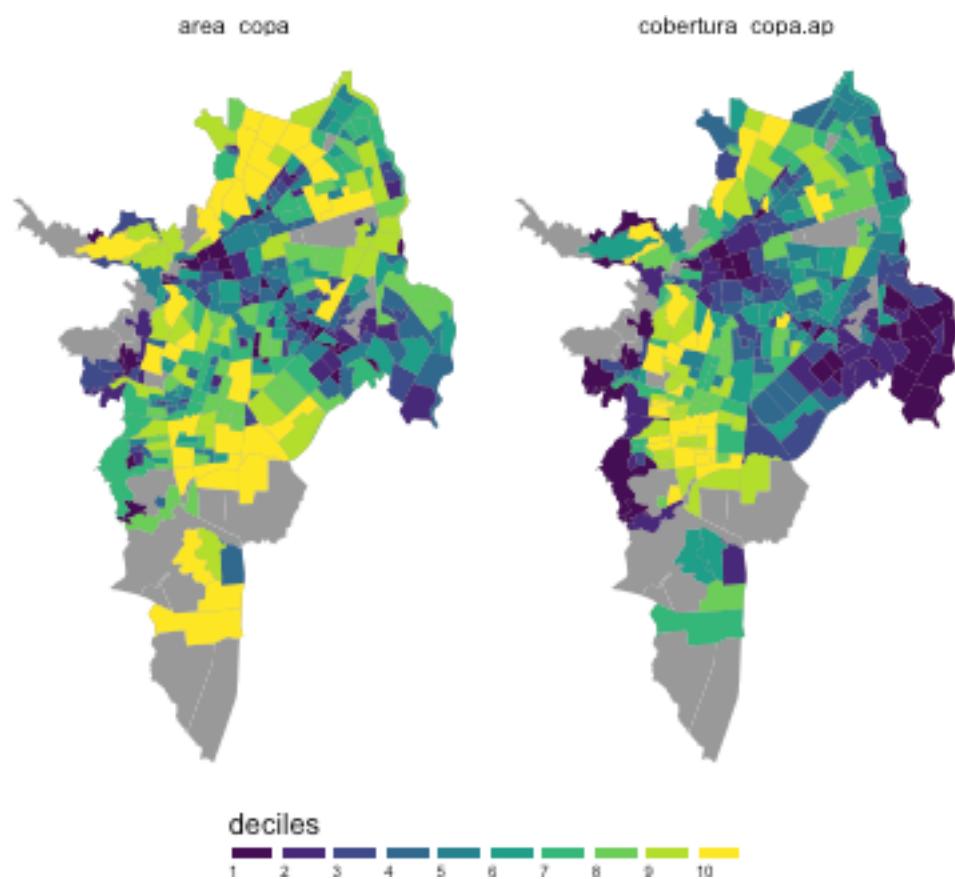


Figure 4.42: Sectores urbanos de las variables dependientes sobre cobertura de copa

#### 4.4.2.1 Correlaciones y gráficos de dispersión entre pares

Para asegurarnos de que las variables no están correlacionadas entre si, usaremos los coeficientes de correlación de Pearson, usado para detectar relaciones lineales, usualmente en variables con distribución normal, y el coeficiente de Spearman para detectar relaciones en variables con otras distribuciones o que exhiben relaciones no lineales. Para tener una idea más amplia sobre esa relación que expresan los coeficiente de correlación se incluyen gráficas de dispersión entre las variables independientes, y con las dependientes.

En la figura 4.43 se explora las relaciones entre las variables de población en las unidades originales de los datos (número de personas); la matriz triangular superior muestra los coeficientes de correlación de Pearson, la diagonal contiene el histogram de frecuencias de la variable y la matriz triangular inferior muestra un gráfico de dispersión y la línea de tendencia usando un modelo lineal entre cada par de variables. Es notoria la alta correlación entre tener ningún estudio y tener alguna limitación física (0.88); pertenecer a una comunidad afrodescendiente y carecer de estudios (0.92) o ser afrodescendiente y tener alguna limitación (0.88). Esto representa una suma de condiciones desfavorables relacionadas entre sí, que desde el punto de vista del modelo sólo podrán ser representadas por una de las variables, la que mejor se relacione con la cobertura de copa y evitar así colinealidad entre los predictores.

Cuando realizamos la misma comparación entre las variables porcentuales (más la densidad poblacional) se observan patrones similares(ver figura 4.44): existe una alta correlación negativa entre el porcentaje de población afro de un sector y la tenencia de estudios superiores (-0.71), una fuerte asociación positiva entre el porcentaje de personas afro de un sector y el porcentaje de personas que carecen de estudios (0.68). También hay una fuerte relación inversa entre el porcentaje de personas de un sector sin estudios y el porcentaje de ellos que tiene estudios superiores (-0.8). Estas variables evitaremos usarlas como predictores en una misma formulación para no sesgar la estimación con problemas de colinealidad. Existe también una asociación, no tan fuerte pero importante,

Comparación por pares entre predictores de población

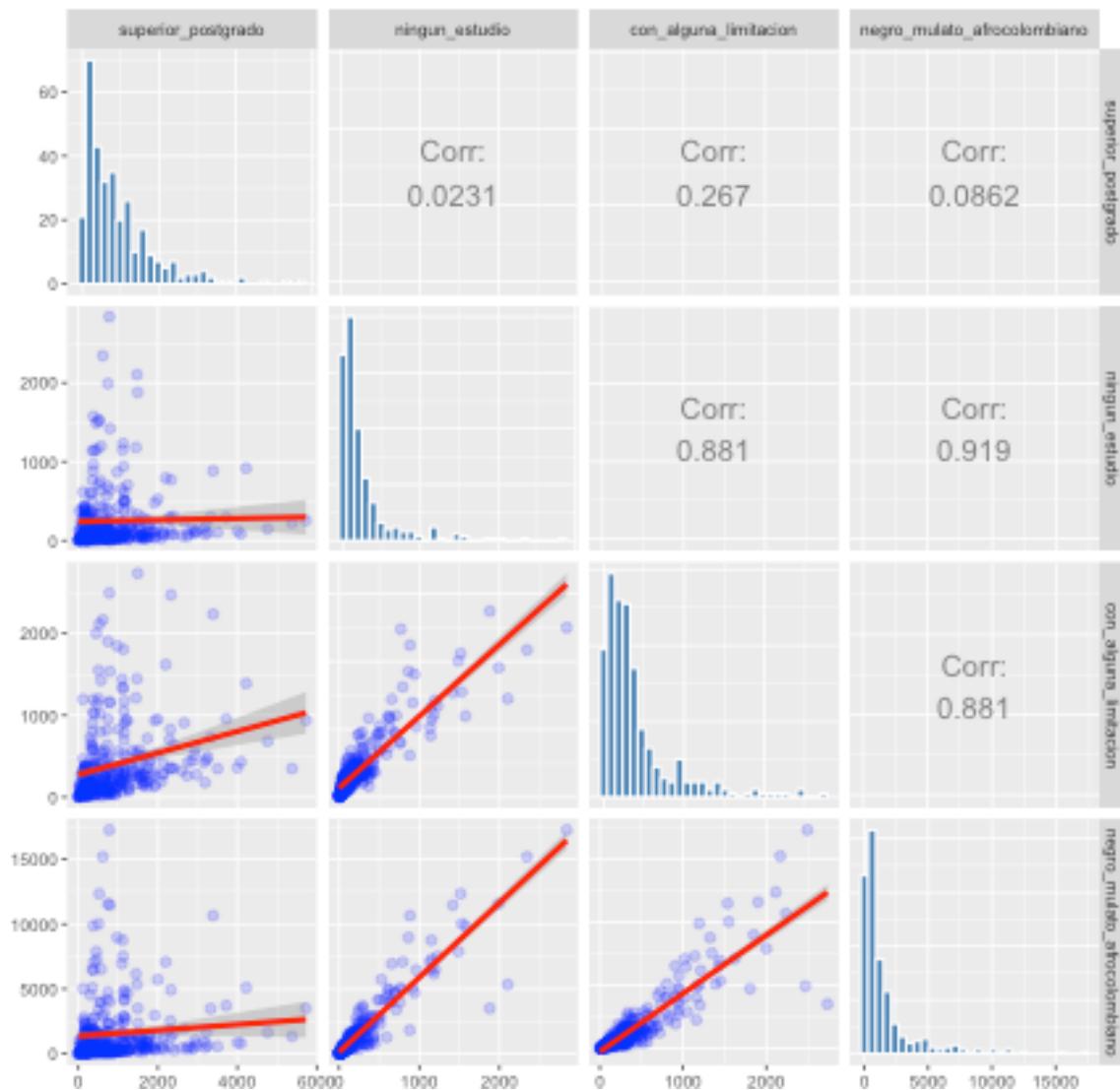


Figure 4.43: Comparación por pares entre predictores de población

entre la densidad de población y sectores con mayor porcentaje de personas afro (0.47) y una asociación negativa entre la densidad de población y el porcentaje de personas con estudios superiores (-0.51). Estos resultados hablan de una concentración de condiciones desfavorables para la población, posiblemente acompañado de una segregación racial alta.

Para completar la inspección de las relaciones entre predictores la figura 4.45 muestra la relación entre las variables provistas en por el censo de población en unidades de personas y las versiones porcentuales que calculamos. El número en el recuadro de cada subgráfica es el coeficiente de correlación de Pearson. Aunque procuraremos incluir cada una de las condiciones evitando redundar al incluir la misma variable en sus dos versiones en un mismo modelo, es interesante que la división entre el número de personas totales del sector atenúa la correlación entre la variable observada y su contraparte porcentual.

Ya que hemos explorado visualmente la dispersión entre los datos crudos, podemos usar forma resumida, usando gráficos de azulejos o de matriz para consultar la intensidad de estas relaciones, no solo las lineales que revela el coeficiente de Pearson, sino usando el coeficiente de Spearman. La figura @ref() sintetiza las relaciones lineales entre las variables dependientes, mientras que la figura @ref() lo hace para las no lineales.

Para seleccionar las variables que mejor predicen la cobertura de copa aplicamos un procedimiento análogo al realizado con las variables dependientes entre sí. Con base en los coeficientes de correlación de Pearson (figura 4.48) y Spearman (figura 4.49) entre las variables dependientes e independientes, y teniendo en cuenta las restricciones de colinealidad entre las variables dependientes, seleccionamos las variables a usar en el modelo lineal.

Así pues, para el área de copa (`area_copa`) los predictores seleccionados son `superior_postgrado`, `densidad_poblacion`, `con_alguna_limitacion.porcentaje`, `afro.porcentaje` y para la cobertura de copa en área pública (`cobertura_copa.ap`) los predictores seleccionados son `superior_postgrado.porcentaje`, `densidad_poblacion`, `con_alguna_limitacion.porcentaje`, `afro.porcentaje`

Se recomienda especificar correctamente los modelos lineales cuando existen relaciones con

Comparación por pares entre predictores de población porcentuales y la densidad de población

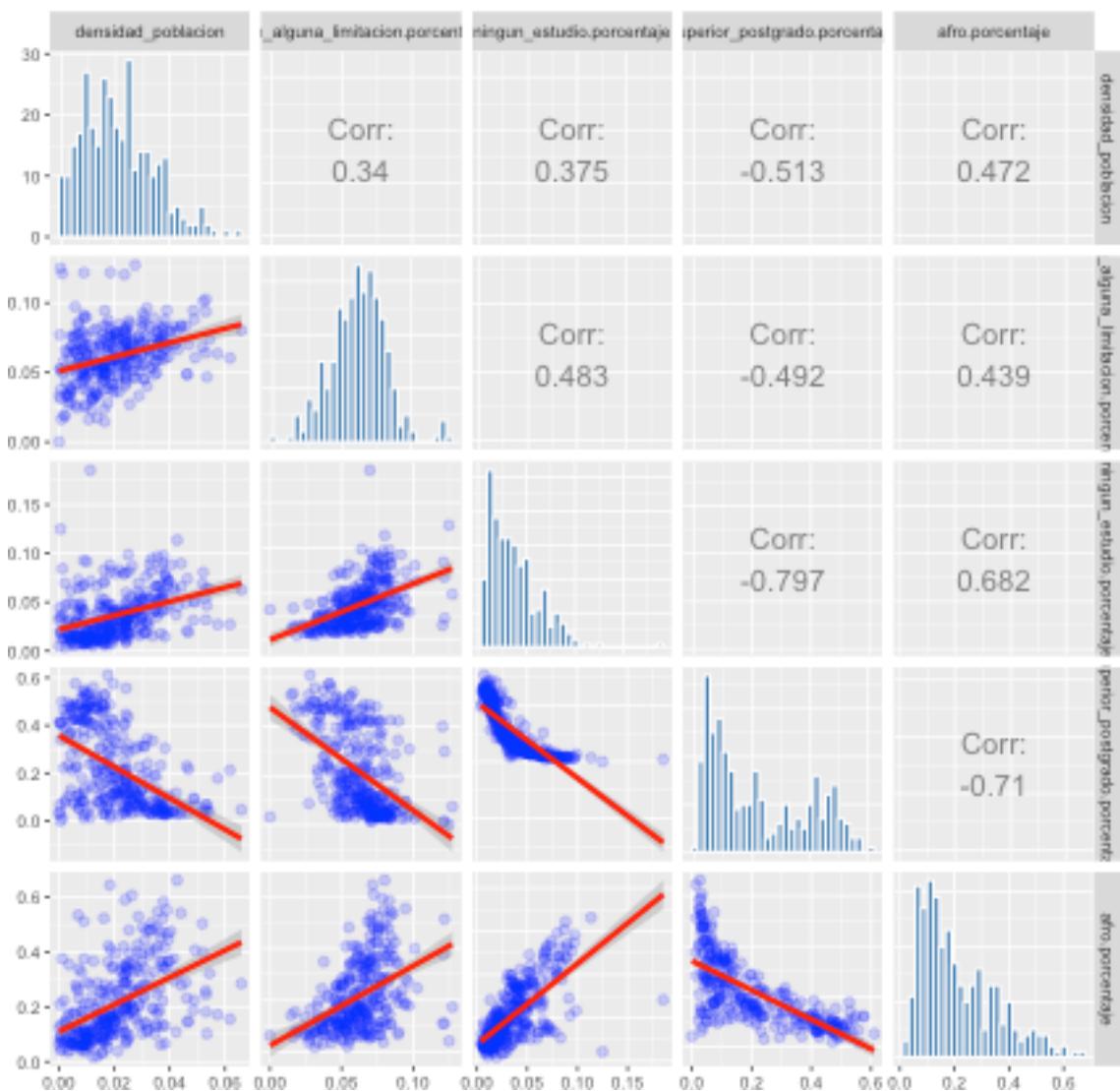


Figure 4.44: Comparación por pares entre predictores de población porcentuales

Comparación por pares entre predictores en unidades de personas y porcentuales

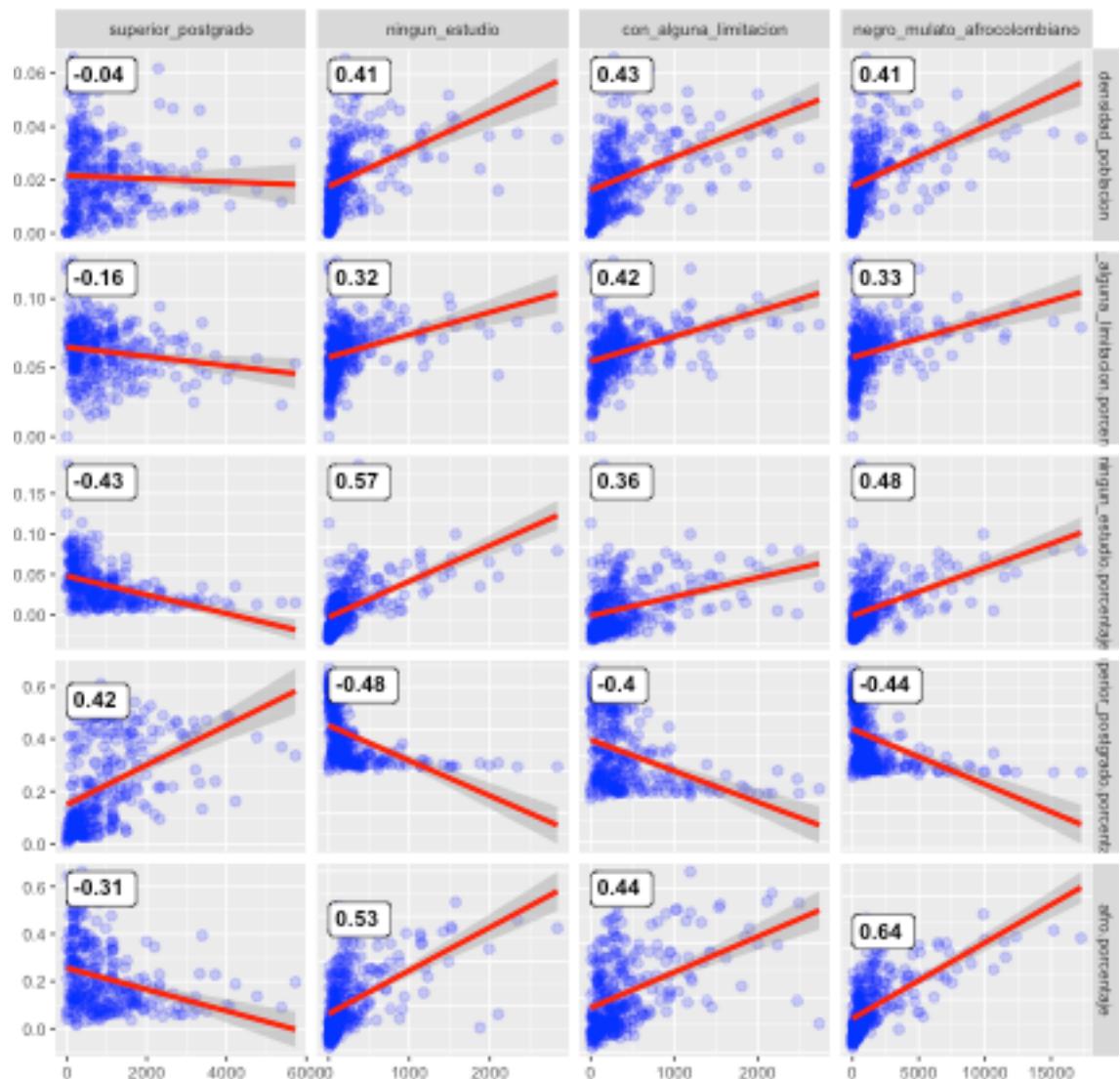


Figure 4.45: Comparación por pares entre predictores de población porcentuales y predictores en unidades de personas

### Coeficiente Pearson entre variables de población

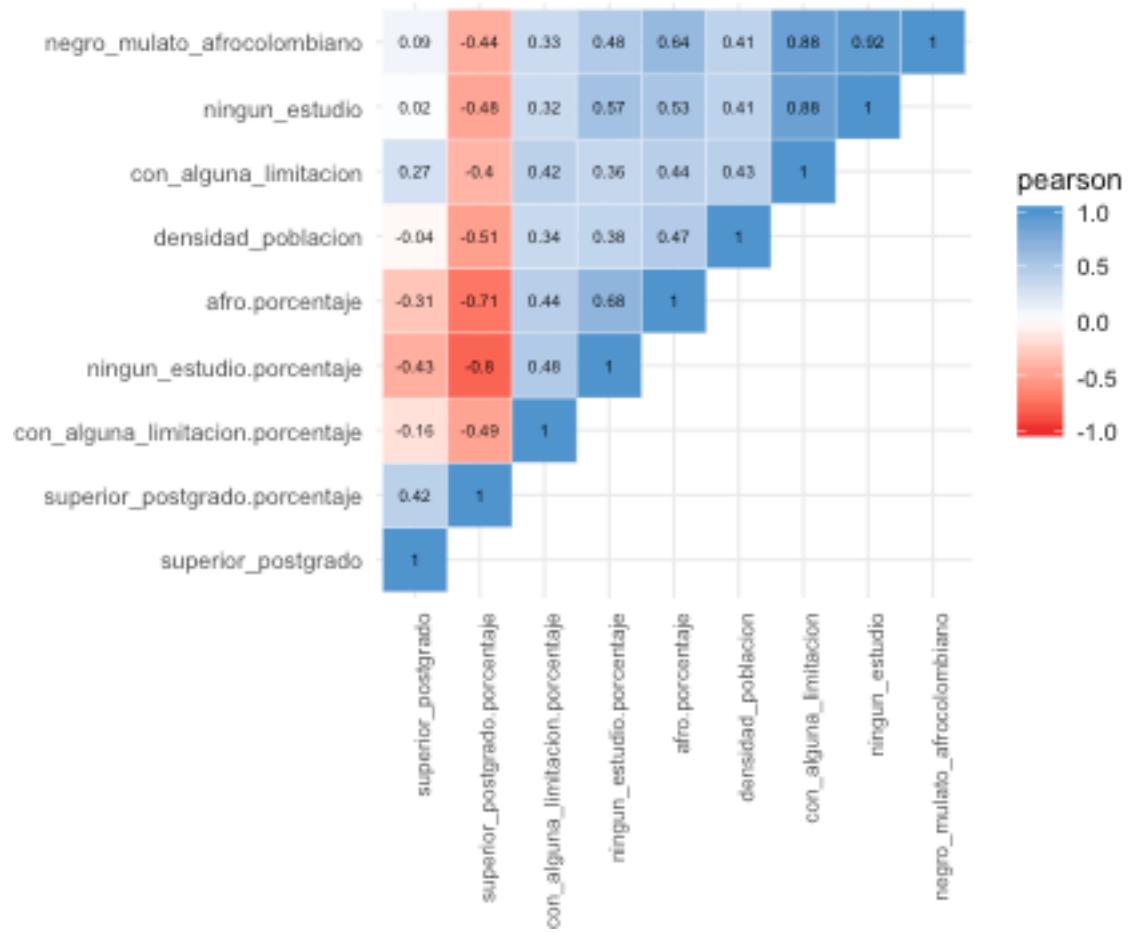


Figure 4.46: Coeficiente Pearson entre variables de población

Coeficiente Spearman entre variables de población

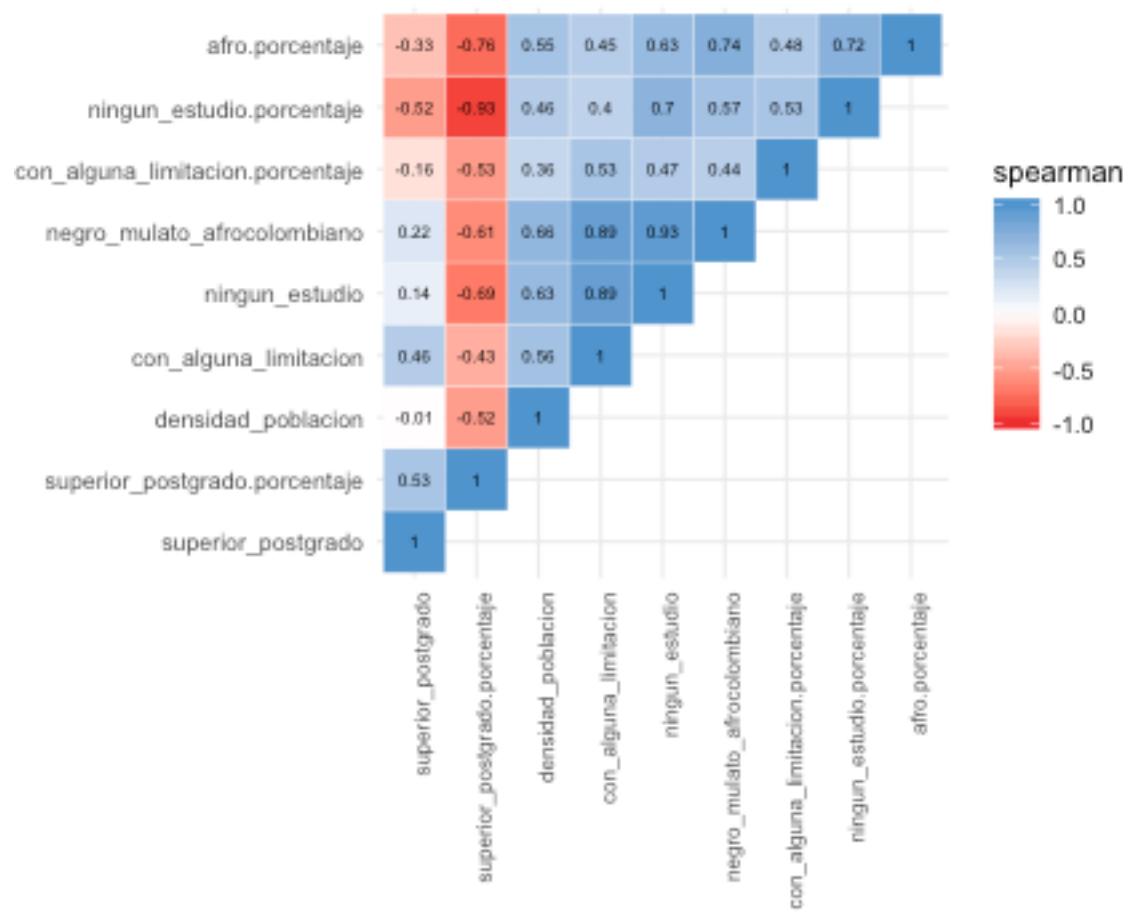


Figure 4.47: Coeficiente Spearman entre variables de población

Pearson entre cobertura de copa y variables de población

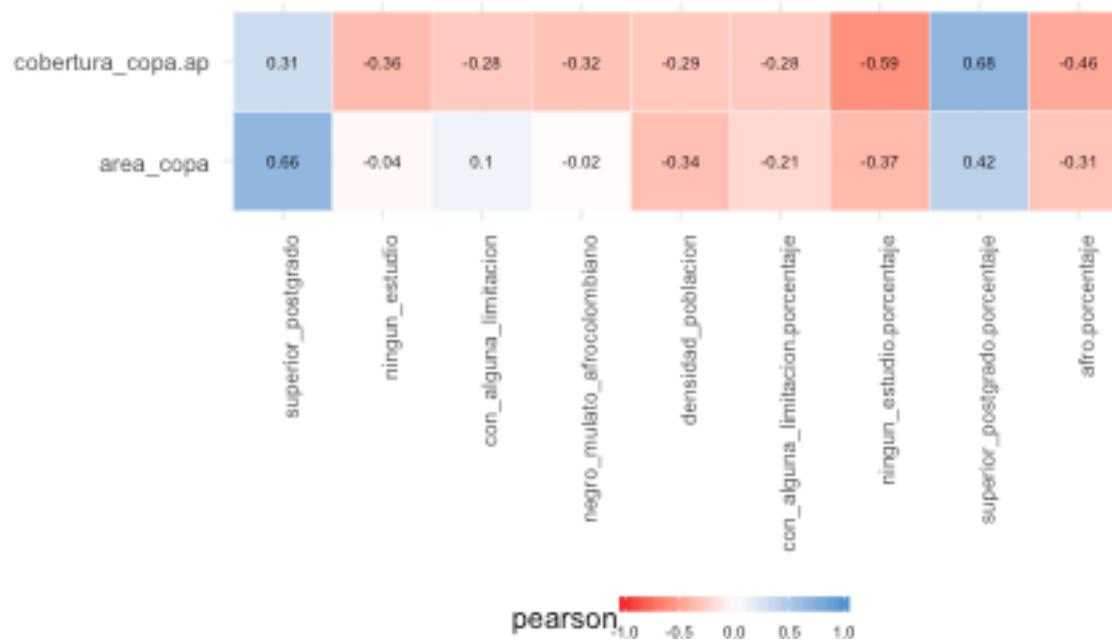


Figure 4.48: Coeficiente Pearson entre coberturas de copa y variables de población

Spearman entre cobertura de copa y variables de población

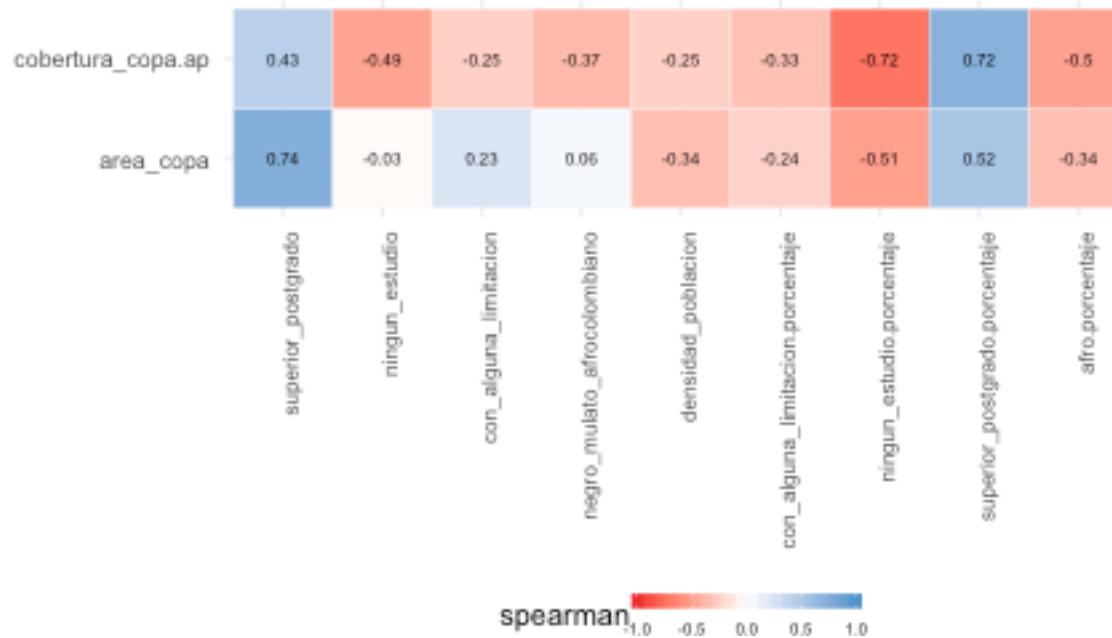


Figure 4.49: Coeficiente Spearman entre coberturas de copa y variables de población

Predictores de población vs. cobertura de copa (regresión lineal + Pearson)

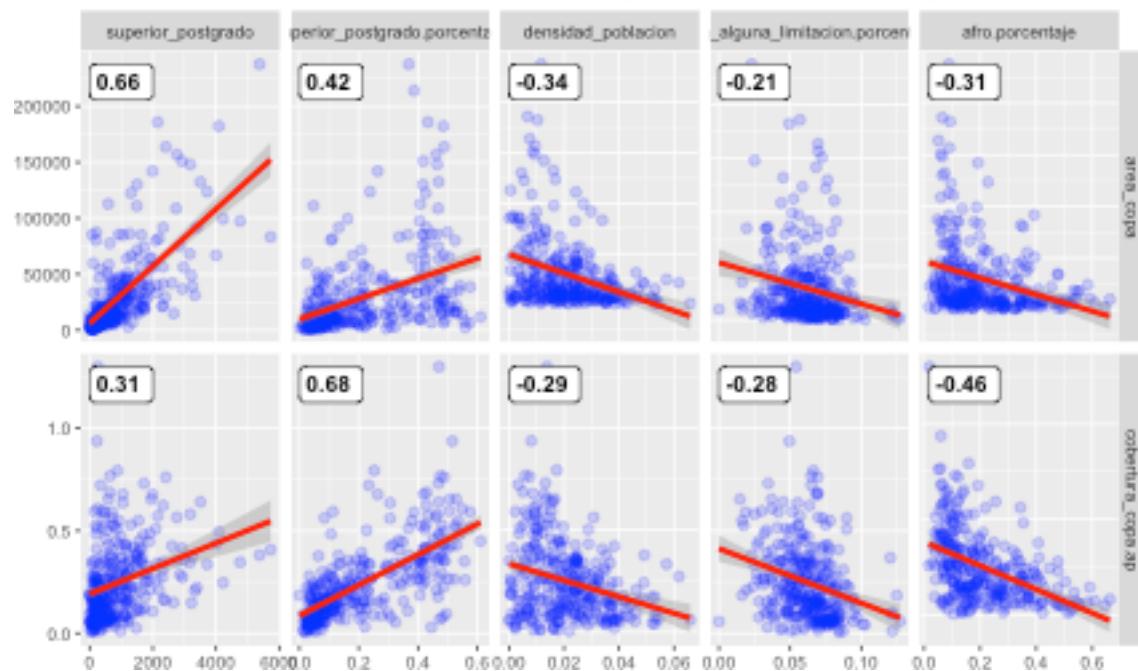


Figure 4.50: Gráficas de dispersión entre predictores de población y variables dependientes de la cobertura de copa (regresión lineal + coeficiente de Pearson)

correlaciones negativas, pues es posible que sea mejor usar el inverso de la variable. Para probar si al hacerlo dichas variables tienen incrementos en los coeficientes de correlación respecto de las variables dependientes. Como se observa en las figuras 4.50 y 4.51 las mejoras son solo en la variable de `afro.porcentaje.inv` en el segundo dígito decimal. Como son tan pequeñas las diferencias opto por no usar las variables transformadas por la función inversa pues esto reduce el intervalo en que varian, reduciendo así su poder explicativo.

#### 4.4.2.2 Modelos de regresión lineal

Antes ajustar los modelos suele ser común en los modelos de regresión ajustar la distribución de las variables dependientes (y a veces las independientes) por motivos teóricos usando transformaciones logarítmicas o de raíz cuadra para eliminar no linealidades entre las variables dependientes y las independientes, y reducir posibles fenómenos de hetero-

Predictores de población (invertido) vs. cobertura de copa (regresión lineal + Pearson)

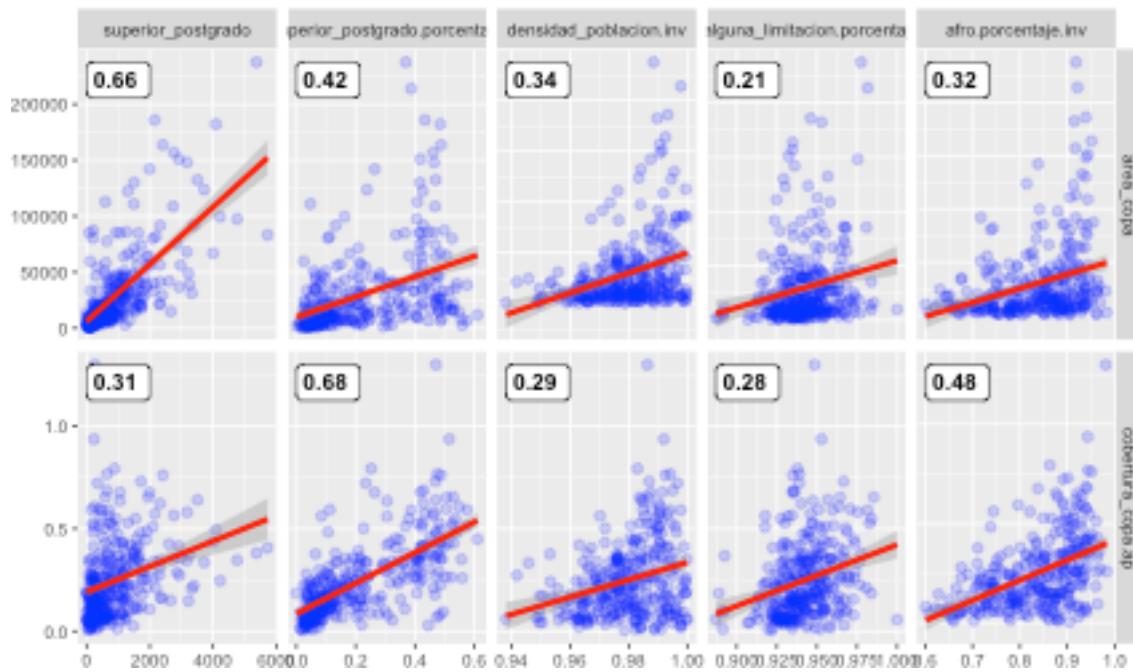


Figure 4.51: Gráficas de dispersión entre predictores de población (usando el inverso de la variable con correlaciones negativas:  $f(x) = \frac{1}{x+1}$ ) y las variables dependientes de la cobertura de copa ( regresión lineal + coeficiente de Pearson)

cedasticidad debido a estas no-linealidades. Siguiendo el ejemplo, ajustaremos 5 modelos: uno por cada variable dependiente en sus tres versiones, sin transformación, con una transformación logarítmica (excepción para la cobertura\_copa.ap pues por tener valores ya entre [0,1] produce valores negativos) y otra de la raíz cuadrada.

Los histogramas de las variables dependientes transformadas se muestran en la figura 4.52.

Las mejoras sobre el modelo de regresión que puedan hacer estas transformaciones pueden apreciarse en los gráficos de dispersión de las variables independientes transformadas vs. las dependientes o/y en los coeficientes de correlación de Pearson. Como se observa en las figuras 4.53 y 4.54 hay leves mejoras al aplicar las transformaciones.

También es sabido que dividir o multiplicar por alguna constante no tiene ningún efecto en la calidad de la estimación , pero sí sobre los coeficientes de la regresión. Esto suele ser sensible a la hora de interpretar los cambios marginales de cada una de las variables

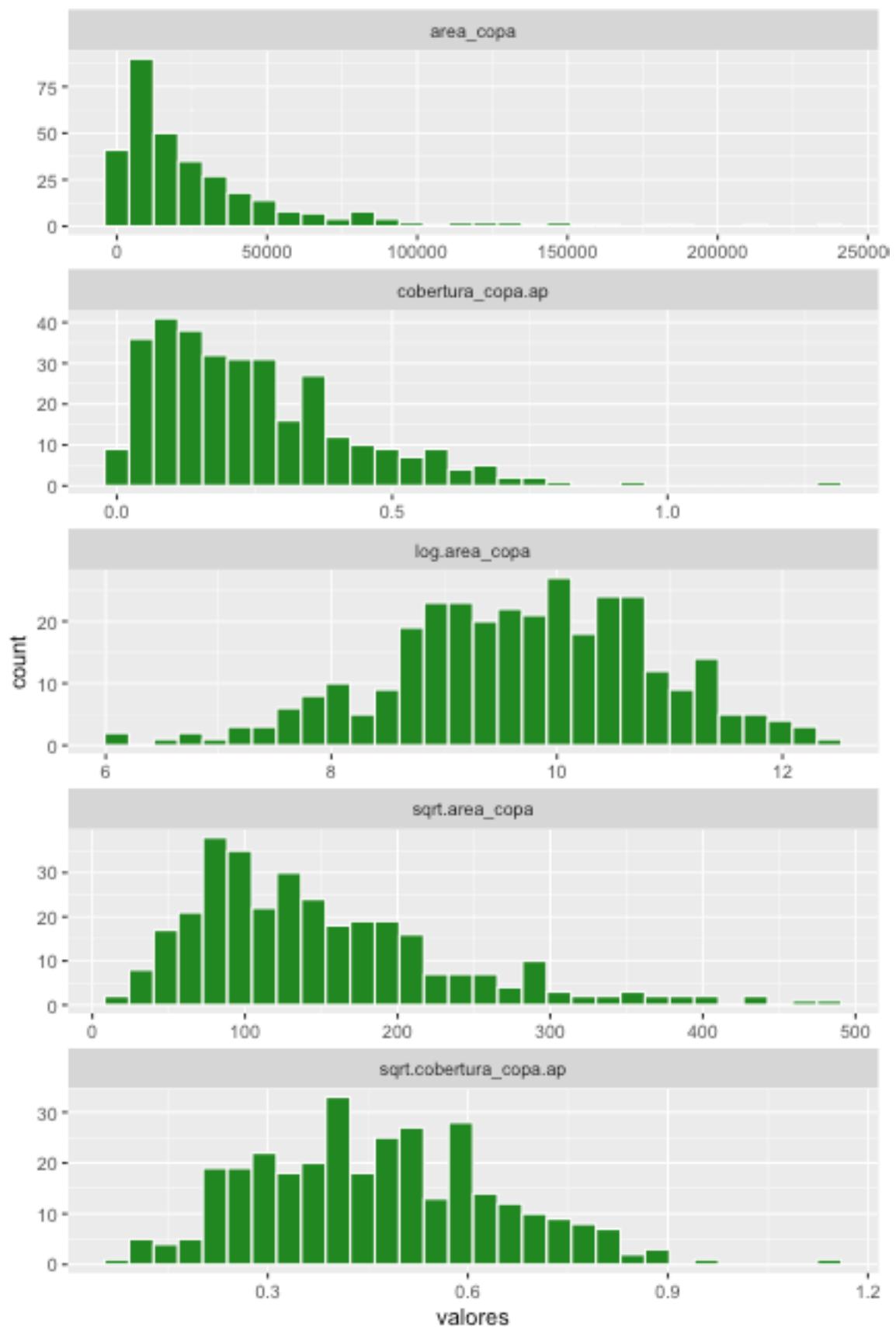


Figure 4.52: Histogramas de las variables de cobertura de copa transformadas

Predictores de población vs. área de copa transformada (regresión lineal + Pearson)

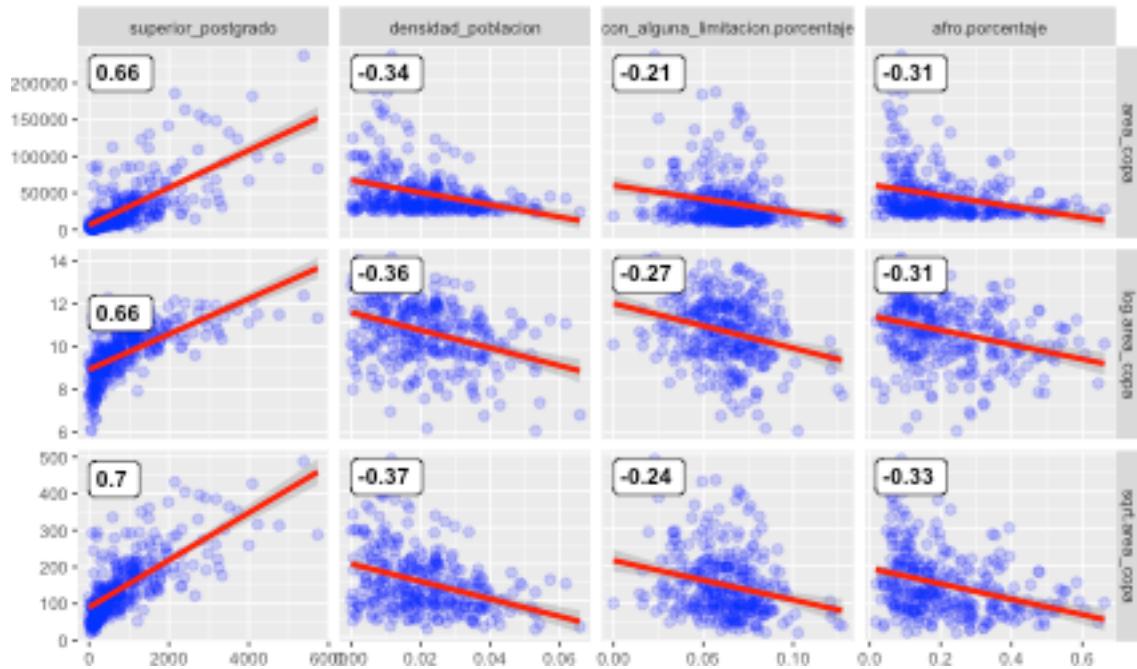


Figure 4.53: Gráficas de dispersión entre predictores de población y variables dependientes de la área de copa transformadas (regresión lineal + coeficiente de Pearson)

Predictores de población vs. cobertura de copa transformada (regresión lineal + Pearson)

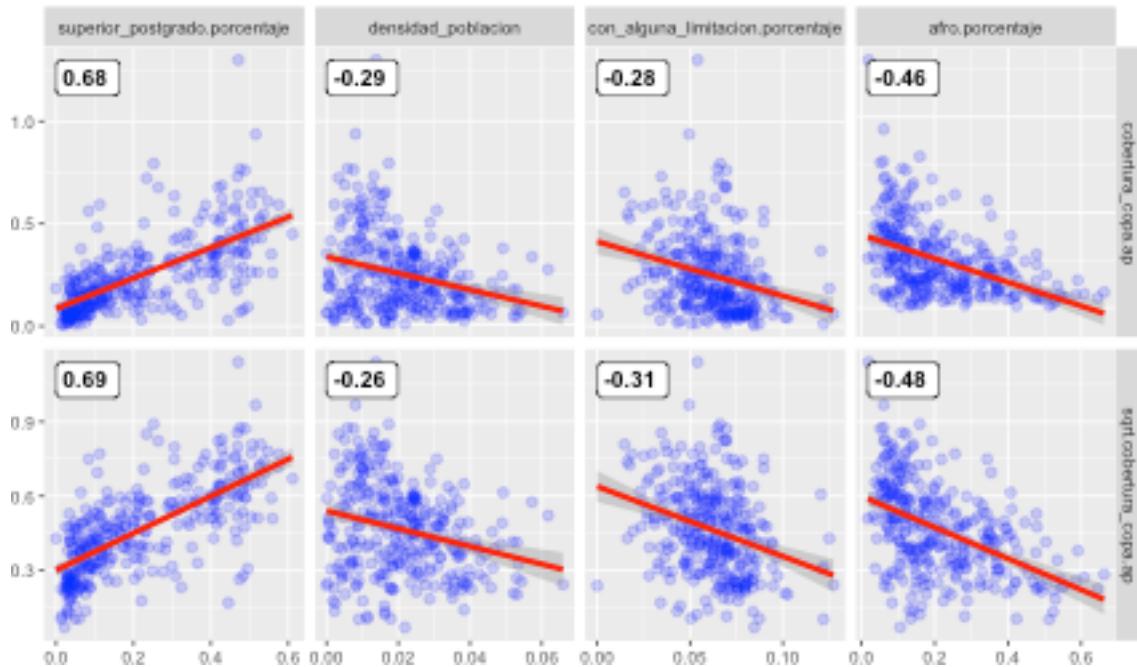


Figure 4.54: Gráficas de dispersión entre predictores de población y variables dependientes de la cobertura de copa transformadas (regresión lineal + coeficiente de Pearson)

independientes y su efecto sobre la variable dependiente. Sin embargo, lo que interesa para este estudio no es la interpretación de esos cambio sino la importancia relativa de cada variable y comparar los cambios de los coeficientes de regresión para el ajuste de cada modelo y/o las mejoras que pueda operar un modelos autorregresivo en caso de encontrarse autocorrelación en los residuos de la regresión lineal. Por esta razón, normalizar los valores puede ser una ventaja pues mantiene los coeficiente mejor acotados. La normalización se aplica posterior a las transformaciones propuestas y se realiza dividiendo por el máximo valor de los datos de cada variable para mantener valores en el intervalo [0,1], dado que los valores son todos iguales o mayores que 0.

Al aplicar test para verificar que las condiciones de un buen ajuste (no hay sesgos en el estimador o una mala especificación del modelo) de un modelo lineal se cumplen:

- La media de los residuos es 0 o muy cercana.
- La distribución de los residuos es normal.
- Los residuos muestran homocedasticidad (la varianza es constante)

Para verificar la normalidad de los residuos usaremos el test de Shapiro–Wilk (Shapiro and Wilk, 1965, ) y para la verificar si existe homocedasticidad el test de Breusch–Pagan (Breusch and Pagan, 1979). Además se acompaña de gráficas de diagnóstico de los resultados de cada modelo. El siguiente código en **R** crea las 5 regresiones lineales propuestas y se resumen en la la tablas 4.8 y 4.9

```
dependiente <- "area_copa"
independientes <- indep.poblacion.copa.sel
var_names <- c(dependiente, names(regresion.arboles[, independientes]))
regresion.arboles.mn <- max_nomalization(regresion.arboles, var_names)
lm.mn.area_copa.sel <- crear_lm_from_df(regresion.arboles.mn)
sm <- summary(lm.mn.area_copa.sel)
sm
##
```

```

## Call:
## lm(formula = form, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.24268 -0.05887 -0.01648  0.03275  0.75376
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)               0.10059   0.02129  4.725 3.45e-06
## superior_postgrado.mxn   0.61703   0.03751 16.450 < 2e-16
## densidad_poblacion.mxn  -0.25404   0.03354 -7.574 3.94e-13
## con_alguna_limitacion.porcentaje.mxn -0.01734   0.04215 -0.411   0.681
## afro.porcentaje.mxn      0.04459   0.03494  1.276   0.203
##
## (Intercept)                 ***
## superior_postgrado.mxn    ***
## densidad_poblacion.mxn    ***
## con_alguna_limitacion.porcentaje.mxn
## afro.porcentaje.mxn
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1036 on 319 degrees of freedom
## Multiple R-squared:  0.5378, Adjusted R-squared:  0.532
## F-statistic: 92.78 on 4 and 319 DF,  p-value: < 2.2e-16

dependiente <- "log.area_copa"
independientes <- indep.poblacion.copa.sel
var_names <- c(dependiente, names(regresion.arboles[, independientes]))

```

```

regresion.arboles.mn <- max_nomalization(regresion.arboles, var_names)

lm.mxn.log.area_copa.sel <- crear_lm_from_df(regresion.arboles.mn)
modelo <- crear_lm_from_df(regresion.arboles.mn)
sm <- summary(lm.mxn.log.area_copa.sel)
sm

##
## Call:
## lm(formula = form, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.209393 -0.034801  0.000263  0.043219  0.180274
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  0.78745   0.01301 60.539 < 2e-16
## superior_postgrado.mxn      0.38208   0.02292 16.671 < 2e-16
## densidad_poblacion.mxn     -0.16544   0.02049 -8.073 1.42e-14
## con_alguna_limitacion.porcentaje.mxn -0.05635   0.02575 -2.188  0.0294
## afro.porcentaje.mxn          0.04423   0.02135  2.072  0.0391
##
## (Intercept)                 ***
## superior_postgrado.mxn      ***
## densidad_poblacion.mxn      ***
## con_alguna_limitacion.porcentaje.mxn *
## afro.porcentaje.mxn          *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

## 
## Residual standard error: 0.06329 on 319 degrees of freedom
## Multiple R-squared:  0.556,  Adjusted R-squared:  0.5504
## F-statistic: 99.86 on 4 and 319 DF,  p-value: < 2.2e-16

dependiente <- "sqrt.area_copa"
independientes <- indep.poblacion.copa.sel
var_names <- c(dependiente, names(regresion.arboles[, independientes]))
regresion.arboles.mn <- max_nomalization(regresion.arboles, var_names)
lm.mxn.sqrt.area_copa.sel <- crear_lm_from_df(regresion.arboles.mn)
summary(lm.mxn.sqrt.area_copa.sel)

```

```

## 
## Call:
## lm(formula = form, data = df)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.28930 -0.07310 -0.01329  0.06219  0.60651
## 
## 
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)               0.28925   0.02274 12.720  <2e-16
## superior_postgrado.mxn   0.75574   0.04007 18.862  <2e-16
## densidad_poblacion.mxn  -0.31411   0.03583 -8.768  <2e-16
## con_alguna_limitacion.porcentaje.mxn -0.04616   0.04502 -1.025   0.306
## afro.porcentaje.mxn        0.05709   0.03732  1.530   0.127
## 
## (Intercept) *** 
## superior_postgrado.mxn ***

```

Table 4.8: Resumen ajuste modelos área copa preliminares

medidasfit	AC	Log(AC)	Sqrt(AC)
Shapiro-Wilk	0.82608	0.99140	0.94358
SW p-value	0.00000	0.05605	0.00000
Breusch-Pagan	25.76218	12.02058	23.46806
BP p-value	0.00004	0.01720	0.00010
Media Residuos	0.00000	0.00000	0.00000
MSE	0.01057	0.00394	0.01206
adj-Rsquare	0.53196	0.55042	0.60400
AIC	-542.76569	-862.01544	-500.03911
Log likelihood	277.38284	437.00772	256.01956

```
## densidad_poblacion.mxn ***  
## con_alguna_limitacion.porcentaje.mxn  
## afro.porcentaje.mxn  
## ---  
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.1107 on 319 degrees of freedom  
## Multiple R-squared: 0.6089, Adjusted R-squared: 0.604  
## F-statistic: 124.2 on 4 and 319 DF, p-value: < 2.2e-16
```

Como se muestra en la tabla 4.8 De las 3 formulaciones de usadas para modelar el área de copa la que mejor ajuste obtuvo fue el que transforma la variable dependiente usando la raíz cuadrada (`log.area_copa`). Aunque el modelo de `sqrt.area_copa` explica el 60% de la variabilidad en los datos (Adjusted R-squared: 0.604), la versión logarítmica tiene un MSE y Akaike menor, y los residuos son normales (el test Shapiro). Este modelo considera las variables `superior_postgrado` y `densidad_poblacion` como muy significativas, ambos con coeficientes obteniendo valores consistentes con el signo de las correlaciones calculadas previamente en análisis bivariado de la figura 4.53.

Los modelos del porcentaje de cobertura de copa son lo siguientes:

```

dependiente <- "cobertura_copa.ap"
independientes <- indep.poblacion.copa.ap.sel
var_names <- c(dependiente, names(regresion.arboles[, independientes]))
regresion.arboles.mn <- max_nomalization(regresion.arboles, var_names)
lm.mxn.cobertura_copa.ap <- crear_lm_from_df(regresion.arboles.mn)
sm <- summary(lm.mxn.cobertura_copa.ap)

sm

##
## Call:
## lm(formula = form, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.28517 -0.05598 -0.01502  0.04070  0.65940
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  0.007612  0.033577  0.227   0.821
## superior_postgrado.porcentaje.mxn 0.387778  0.032525 11.923 <2e-16
## densidad_poblacion.mxn        0.052744  0.034832  1.514   0.131
## con_alguna_limitacion.porcentaje.mxn 0.053220  0.044071  1.208   0.228
## afro.porcentaje.mxn           0.005960  0.040894  0.146   0.884
##
## (Intercept)
## superior_postgrado.porcentaje.mxn ***
## densidad_poblacion.mxn
## con_alguna_limitacion.porcentaje.mxn
## afro.porcentaje.mxn

```

```

## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.105 on 319 degrees of freedom
## Multiple R-squared:  0.4644, Adjusted R-squared:  0.4576
## F-statistic: 69.14 on 4 and 319 DF,  p-value: < 2.2e-16

dependiente <- "sqrt.cobertura_copa.ap"
independientes <- indep.poblacion.copa.ap.sel
var_names <- c(dependiente, names(regresion.arboles[, independientes]))
regresion.arboles.mn <- max_nomalization(regresion.arboles, var_names)
lm.mxn.sqrt.cobertura_copa.ap <- crear_lm_from_df(regresion.arboles.mn)
sm <- summary(lm.mxn.sqrt.cobertura_copa.ap)
sm

##
## Call:
## lm(formula = form, data = df)
##
## Residuals:
##    Min     1Q   Median     3Q    Max 
## -0.38864 -0.06666 -0.00395  0.06375  0.42095 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)               0.201407  0.036441  5.527 6.78e-08  
## superior_postgrado.porcentaje.mxn 0.444226  0.035300 12.584 < 2e-16  
## densidad_poblacion.mxn        0.100341  0.037804  2.654  0.00835  
## con_alguna_limitacion.porcentaje.mxn 0.033052  0.047831  0.691  0.49006  
## afro.porcentaje.mxn          -0.003885  0.044384 -0.088  0.93031  

```

Table 4.9: Resumen ajuste de porcentaje de cobertura de copa preliminares

medidasfit	%CC	Sqrt(%CC)
Shapiro-Wilk	0.89928	0.98221
SW p-value	0.00000	0.00048
Breusch-Pagan	18.48790	12.36483
BP p-value	0.00099	0.01483
Media Residuos	0.00000	0.00000
MSE	0.01085	0.01278
adj-Rsquare	0.45764	0.48584
AIC	-534.19892	-481.14452
Log likelihood	273.09946	246.57226

```
##
## (Intercept) ***
## superior_postgrado.porcentaje.mxn ***
## densidad_poblacion.mxn **
## con_alguna_limitacion.porcentaje.mxn
## afro.porcentaje.mxn
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1139 on 319 degrees of freedom
## Multiple R-squared: 0.4922, Adjusted R-squared: 0.4858
## F-statistic: 77.3 on 4 and 319 DF, p-value: < 2.2e-16
```

Como se observa en la tabla 4.9 el porcentaje de cobertura copa en el área pública, el mejor ajuste se obtuvo con la variable sin transformación `cobertura_copa.ap`. Este modelo considera que solo la variable `superior_postgrado.porcentaje` como muy significativas, y explica casi el 46% de la variabilidad (Adjusted R-squared: 0.4576384) con MSE y Akaike menores. El coeficiente de la densidad de población obtiene un valor positivo y poco significativo, contrario al signo de la correlación obtenida en el análisis bivariado de la figura 4.54, por lo que es descartado.

Reducemos ambos modelos eliminando las variables no significativas. Para el mejor modelo del área de copa el siguiente código realiza el trabajo resumir los resultados en gráficas diagnósticas y los tests de normalidad y homocedasticidad en la tabla 4.10.

```

dependiente <- "log.area_copa"
independientes <- c("superior_postgrado", "densidad_poblacion")
var_names <- c(dependiente, names(regresion.arboles[, independientes]))
regresion.arboles.mn <- max_nomalization(regresion.arboles, var_names)
lm.best.area_copa <- lm(log.area_copa.mxn ~ superior_postgrado.mxn + densidad_poblacion.
                           data = regresion.arboles.mn)
sm <- summary(lm.best.area_copa)
sm

##
## Call:
## lm(formula = log.area_copa.mxn ~ superior_postgrado.mxn + densidad_poblacion.mxn,
##      data = regresion.arboles.mn)
##
## Residuals:
##    Min     1Q Median     3Q    Max 
## -0.228080 -0.034443  0.003531  0.044043  0.191950 
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 0.773649  0.007707 100.380 <2e-16 ***
## superior_postgrado.mxn 0.373575  0.021809 17.130 <2e-16 ***
## densidad_poblacion.mxn -0.158838  0.017814 -8.917 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##

```

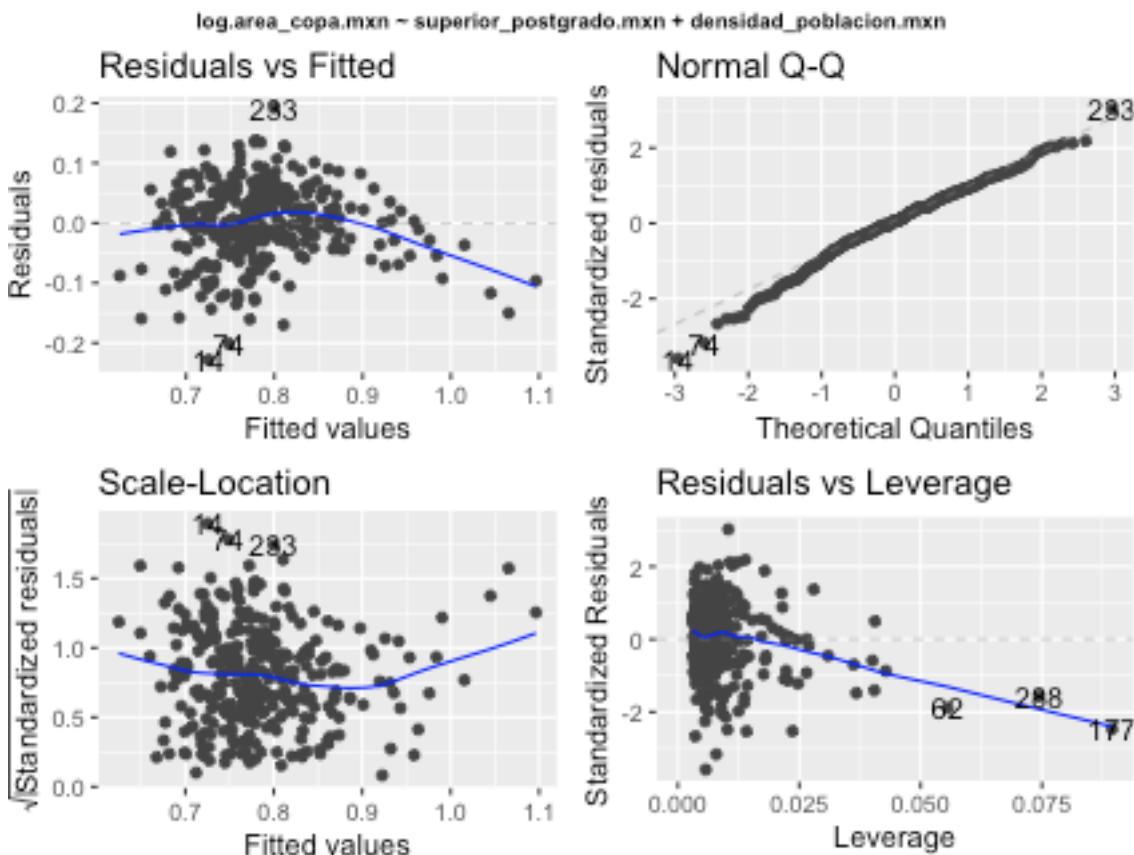


Figure 4.55: Gráficas diagnósticas para el análisis de regresión lineal del área de copa

```
## Residual standard error: 0.06378 on 321 degrees of freedom
## Multiple R-squared:  0.5463, Adjusted R-squared:  0.5435
## F-statistic: 193.3 on 2 and 321 DF,  p-value: < 2.2e-16
```

Los resultados de los tests (tabla 4.10) rechazan que los residuos sean normales, pero la varianza pasa el test Breusch-Pagan de homocedasticidad. Es posibles que los estimadores sean inefficientes y que existan efecto no considerados en el modelo dada la no normalidad de los residuos. Las gráficas diagnosticas (figura 4.55) muestran que existe no linealidades en los residuos, pues se observa como la linea de tendencia hace curvas respecto de la horizontal. Para probar si es posible tener un modelo mejor especificado se introduce en la sección análisis geoestadísticos basados en la autocorrelación espacial.

El mismo procedimiento lo aplicamos para el mejor modelo de cobertura de copa. El siguiente código construye el modelo y realiza los test descritos anteriormente.

Table 4.10: Resúmen ajuste OLS: log(area copa)

medidasfit	Area Copa
Shapiro-Wilk	0.98764
SW p-value	0.00732
Breusch-Pagan	5.24932
BP p-value	0.07246
Media Residuos	0.00000
MSE	0.00403
adj-Rsquare	0.54348
AIC	-859.02772
Log likelihood	433.51386

```

dependiente <- "cobertura_copa.ap"
independientes <- c("superior_postgrado.porcentaje")
# max normalizado
var_names <- c(dependiente, independientes)
regresion.arboles.mn <- max_nomalization(regresion.arboles, var_names)
lm.best.cobertura.ap <- lm(cobertura_copa.ap.mx ~ superior_postgrado.porcentaje.mx,
                             data = regresion.arboles.mn)
sm <- summary(lm.best.cobertura.ap)
sm

##
## Call:
## lm(formula = cobertura_copa.ap.mx ~ superior_postgrado.porcentaje.mx,
##      data = regresion.arboles.mn)
##
## Residuals:
##    Min     1Q   Median     3Q    Max 
## -0.29895 -0.05645 -0.01560  0.04094  0.66402 
## 
## Coefficients:

```

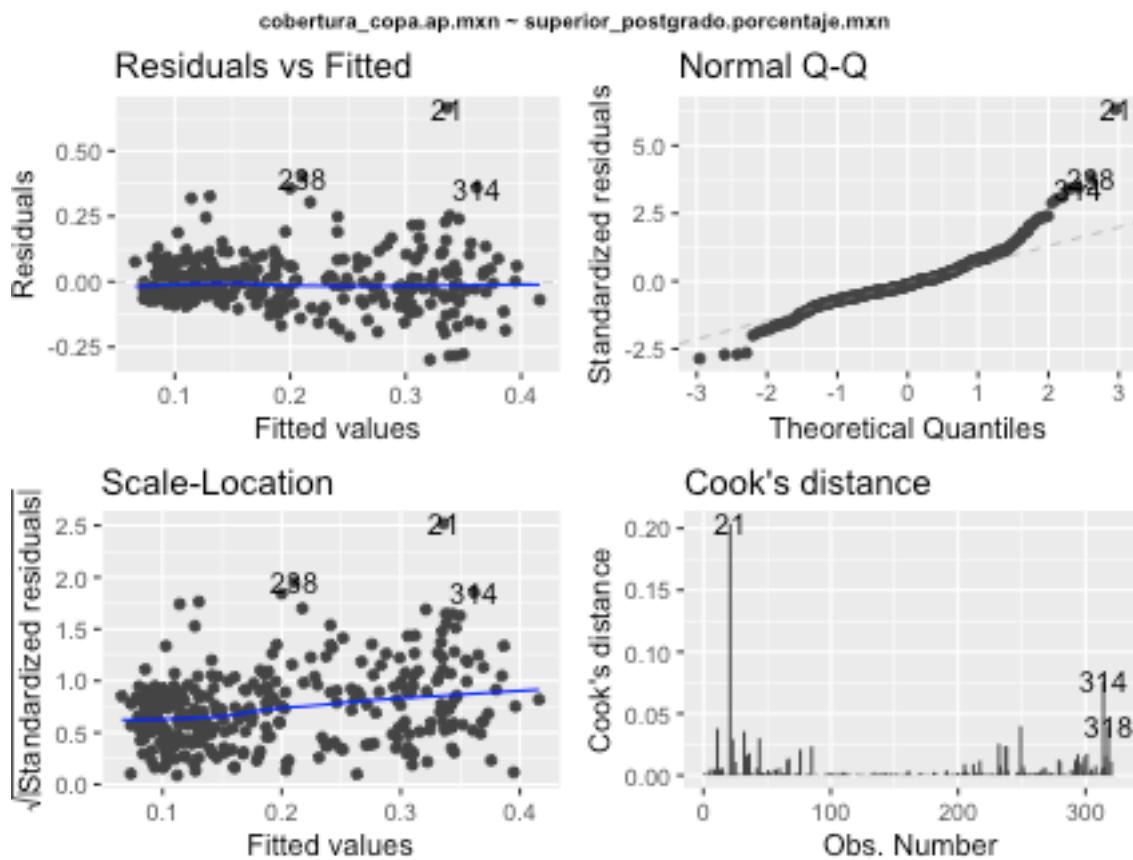


Figure 4.56: Gráficas diagnósticas para el análisis de regresión lineal del porcentaje de cobertura de copa

```

##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  0.065794   0.009627  6.835 4.13e-11 ***
## superior_postgrado.porcentaje.mxn 0.350233   0.021283 16.456 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1052 on 322 degrees of freedom
## Multiple R-squared:  0.4568, Adjusted R-squared:  0.4551
## F-statistic: 270.8 on 1 and 322 DF,  p-value: < 2.2e-16

```

Los resultados de los tests rechazan que los residuos sean normales y que la varianza sea constante (tabla 4.11), sin embargo, es notorio el aumento de la varianza con el aumento

Table 4.11: Resumen ajuste OLS: cobertura copa.ap

medidasfit	%Cobertura de Copa
Shapiro-Wilk	0.90492
SW p-value	0.00000
Breusch-Pagan	16.82734
BP p-value	0.00004
Media Residuos	0.00000
MSE	0.01100
adj-Rsquare	0.45512
AIC	-535.66435
Log likelihood	270.83218

de los valores ajustados; la curva de tendencia de los residuos vs. los valores ajustados sigue una linea de tendencia que se distancia de la horizontal confirmando la heterocedasticidad de los residuos (figura 4.56). Igualmente veremos si la introducción de elementos de la teoría geoestadística permite mejorar la especificación del modelo.

#### 4.4.2.3 Introducción de dimensiones no poblacionales al modelo lineal

Para la inclusión de otras variables agregadas por sector urbano aplicamos el procedimiento realizado para la selección de las variables de población usando como criterio la inclusión de variables con coeficientes de correlación que muestren una asociación fuerte con las variables de cobertura de copa. Las figuras 4.57 y 4.58 se ven los coeficientes de Pearson y Spearman, respectivamente, entre las variables sobre el uso los predios, tipo de viviendas y área (y porcentaje) de espacios verdes en cada sector urbano. Como se observa, existe una fuerte (perfecta) asociación negativa entre el porcentaje de casas y apartamentos, lo que obliga a solo escoger una de las dos en caso de haber una fuerte relación entre alguna de ellas con las variables de cobertura de copa. También hay una fuerte asociación positiva entre el área de espacios verdes y el porcentaje de área de espacio verdes en un sector urbano. Como mencionamos antes solo incluiremos una de las dos en caso de que ambas resulten fuertemente asociadas con las variables dependientes.

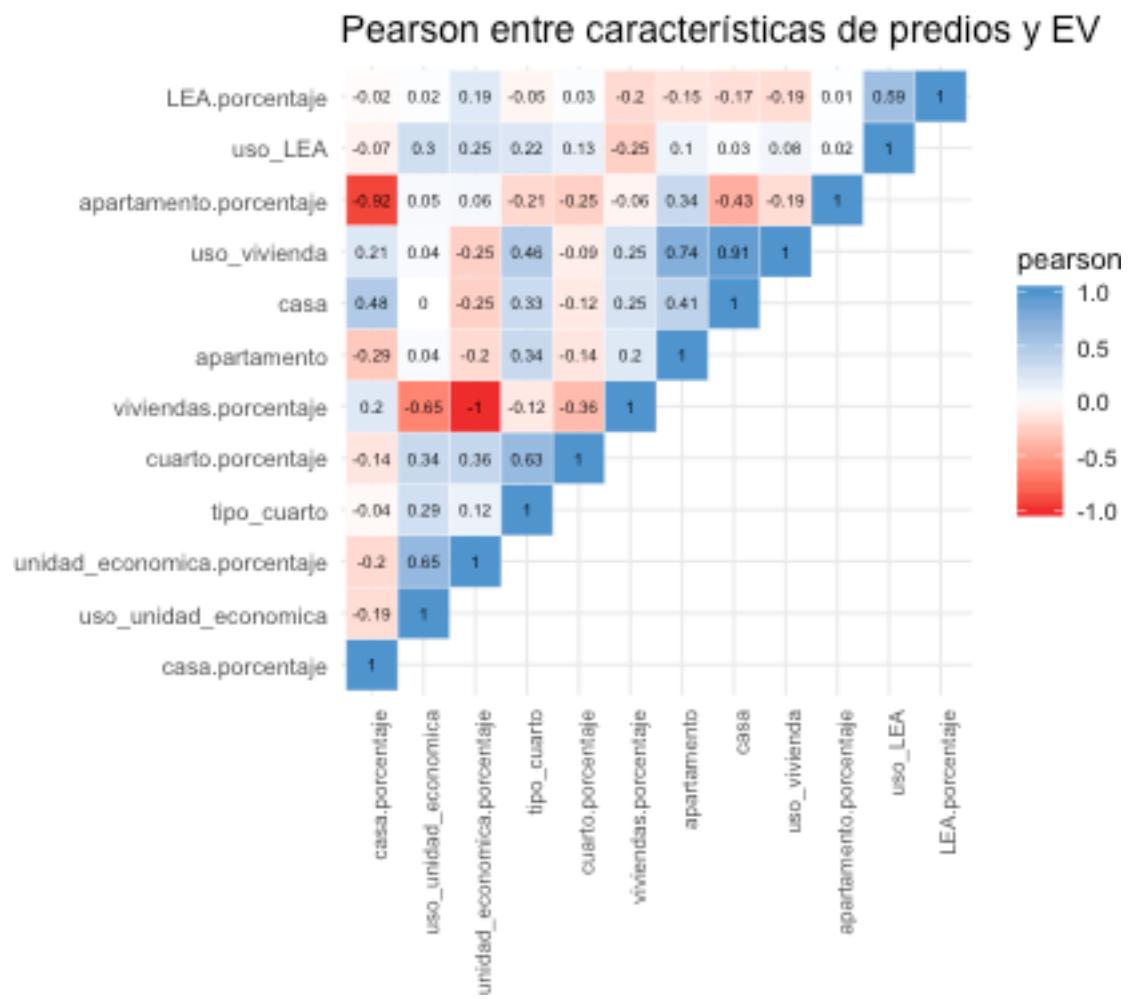


Figure 4.57: Coeficiente Pearson entre variables de uso de los predios y espacios verdes en los sectores urbanos

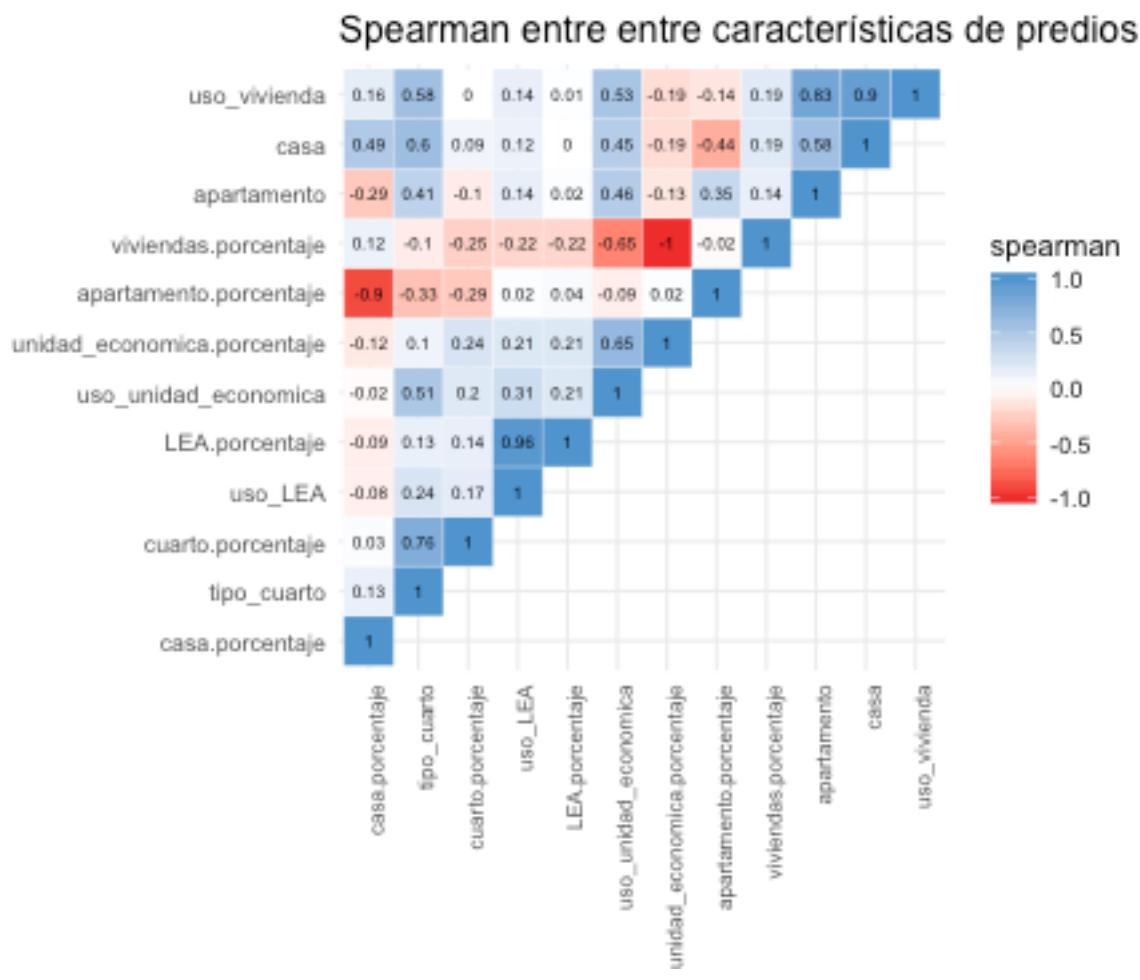


Figure 4.58: Coeficiente Pearson entre variables de uso de los predios y espacios verdes en los sectores urbanos

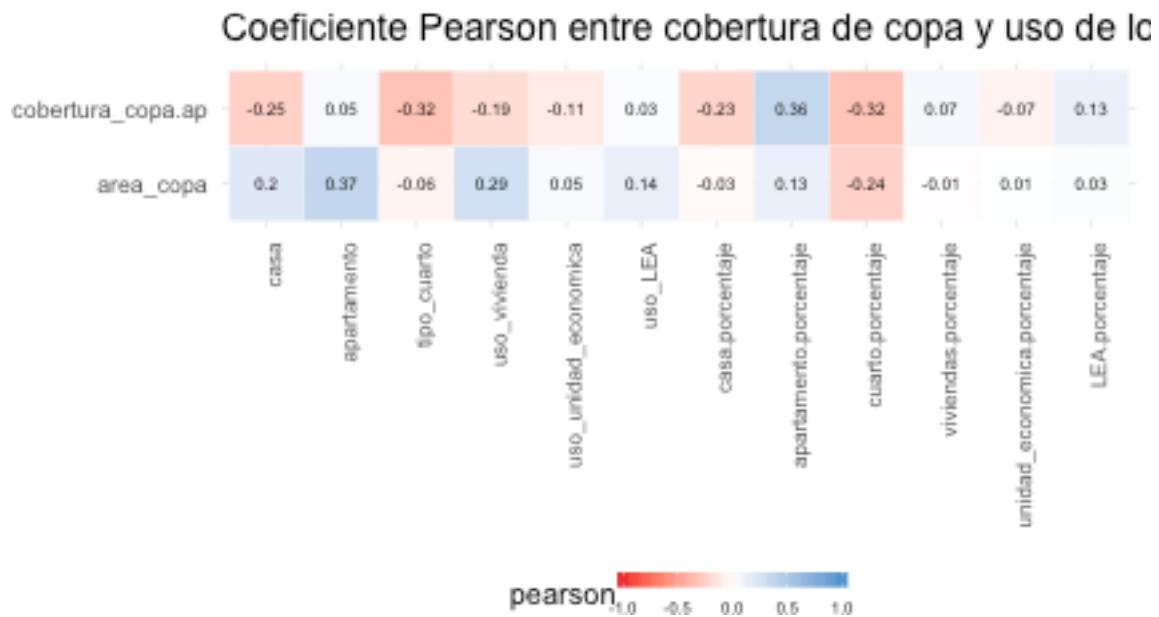


Figure 4.59: Coeficiente Pearson entre coberturas de copa y variables de predios y EV

Las figuras 4.59 y 4.60 muestran la correlación entre las potenciales nuevas variables a incluir en el modelo y las variables dependientes de cobertura y área de copa. Para el área de copa se seleccionan el área de espacios verdes (`area_ep`) y el porcentaje de viviendas tipo cuarto (`cuarto.porcentaje`). Para el modelo de porcentaje de cobertura de se seleccionan `apartamento.porcentaje`, `cuarto.porcentaje` y `area_ep.porcentaje`.

La nueva formulación del modelo de área de copa obtiene los siguientes resultados:

```
# introducimos las variables nuevas
dependiente <- "log.area_copa"
independientes <- c("superior_postgrado", "densidad_poblacion", indep.predios.copa.sel)
# max normalizado
var_names <- c(dependiente, names(regresion.arboles[, independientes]))
regresion.arboles.mn <- max_nomalization(regresion.arboles, var_names)
lm.mod.area_copa <- lm(log.area_copa.mxn ~ superior_postgrado.mxn + densidad_poblacion.mxn + cuarto.porcentaje.mxn + area_ep.mxn, data = regresion.arboles.mn)
sm <- summary(lm.mod.area_copa)
sm
```

Coeficiente Spearman entre cobertura de copa y uso de suelo

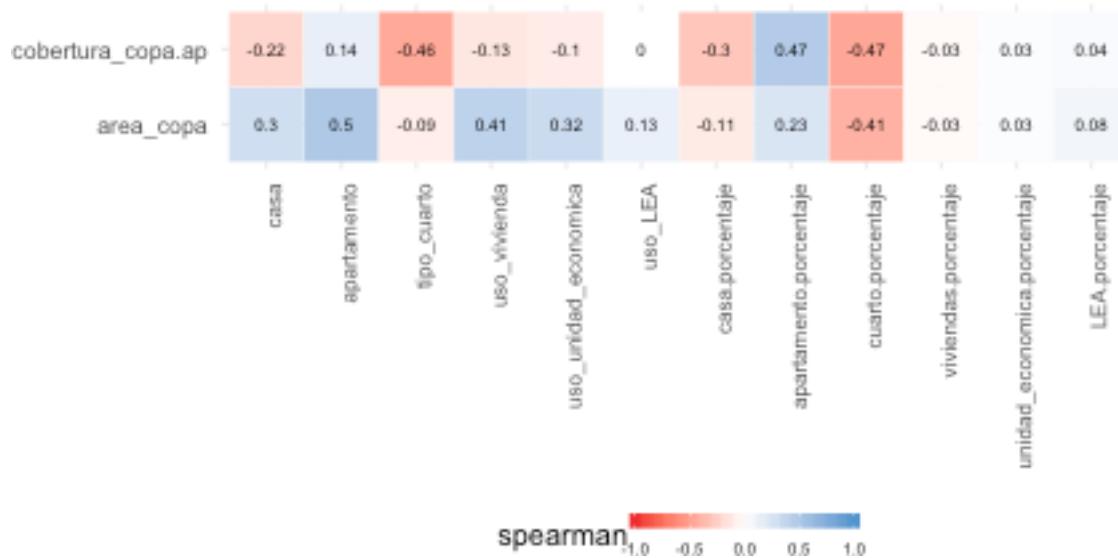


Figure 4.60: Coeficiente Pearson entre coberturas de copa y variables de predios y EV

```
##
## Call:
## lm(formula = log.area_copa.mxn ~ superior_postgrado.mxn + densidad_poblacion.mxn +
##     cuarto.porcentaje.mxn + area_ep.mxn, data = regresion.arboles.mn)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.219304 -0.030935  0.004071  0.041703  0.125492
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)               0.781751  0.008458  92.431 < 2e-16 ***
## superior_postgrado.mxn  0.309340  0.022118  13.986 < 2e-16 ***
## densidad_poblacion.mxn -0.146445  0.016984 -8.623 3.11e-16 ***
## cuarto.porcentaje.mxn -0.149299  0.029954 -4.984 1.02e-06 ***
## area_ep.mxn              0.123103  0.026560  4.635 5.21e-06 ***
## ---
##
```

Table 4.12: Comparación OLS para el área de copa (AC) con variables de población y con otras dimensiones

medidasfit	Log(AC)~pob	log(AC)~pob+otras
Shapiro-Wilk	0.98764	0.97939
SW p-value	0.00732	0.00013
Breusch-Pagan	5.24932	6.70653
BP p-value	0.07246	0.15223
Media Residuos	0.00000	0.00000
MSE	0.00403	0.00345
adj-Rsquare	0.54348	0.60639
AIC	-859.02772	-905.09560
Log likelihood	433.51386	458.54780

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05922 on 319 degrees of freedom
## Multiple R-squared:  0.6113, Adjusted R-squared:  0.6064
## F-statistic: 125.4 on 4 and 319 DF,  p-value: < 2.2e-16
```

Las variables introducidas tienen todas p-value significativos, y a pesar de la inclusión de más términos al modelo existe una mejora en el Criterio de Información de Akaike (baja de  $-859.0277185$  a  $-905.0955982$ ). Los resultados de los test evidencian que aunque hay mejoras en Adjusted R-squared ( de 0.5434765 a 0.6063903) y en MSE (de 0.0040304 a 0.0034533) sigue existiendo no normalidad en los residuos y posibles no linealidades, como se observa también en las gráficas diagnosticas de la regresión 4.61.

Para el caso del porcentaje de cobertura de copa los resultados obtenidos con la introducción de las nuevas variables son los siguientes:

```
# coberbtura AP
dependiente <- "cobertura_copa.ap"
independientes <- c("superior_postgrado.porcentaje", indep.predios.copa.ap.sel)
# max normalizado
var_names <- c(dependiente, independientes)
```

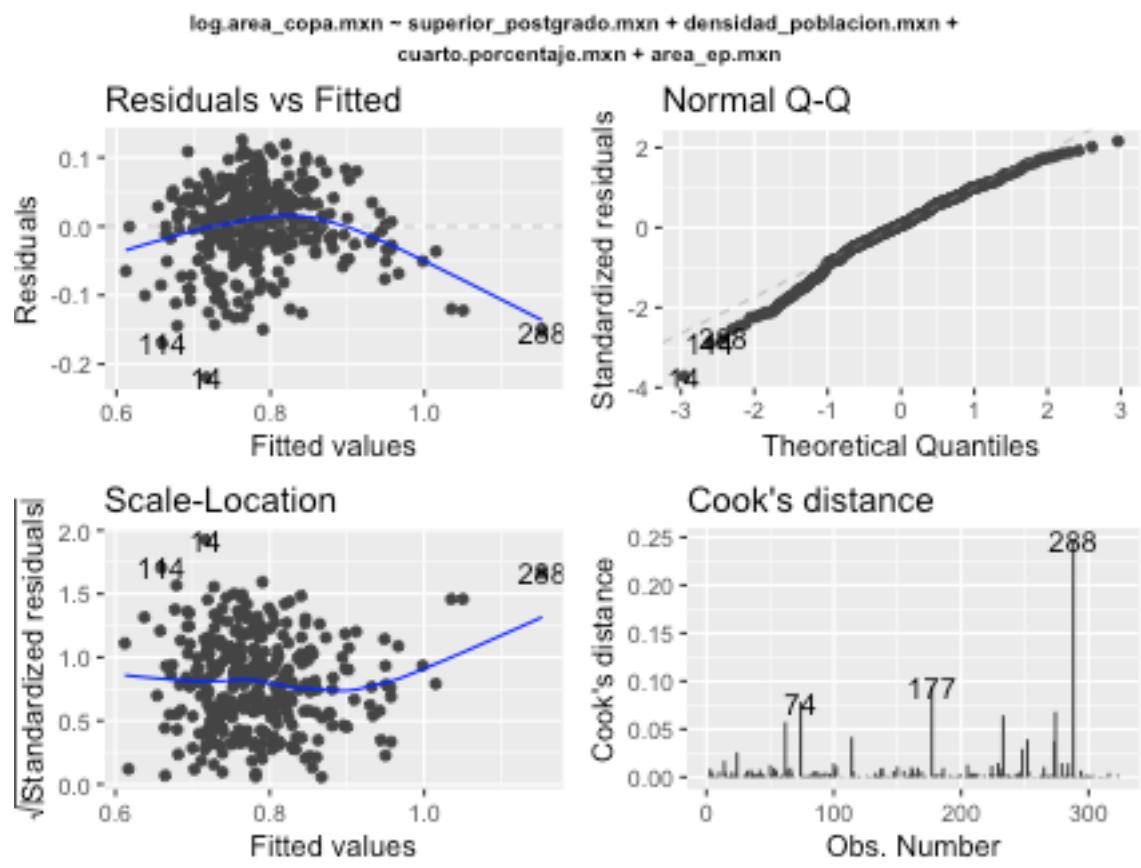


Figure 4.61: Gráficas diagnósticas para el análisis de regresión lineal de área de copa con los nuevos términos

```

regresion.arboles.mn <- max_nomalization(regresion.arboles, var_names)

lm.mod.cobertura.ap <- lm(cobertura_copa.ap.mxn ~ superior_postgrado.porcentaje.mxn +
  apartamento.porcentaje.mxn + cuarto.porcentaje.mxn + area_ep.porcentaje.mxn,
  data = regresion.arboles.mn)

sm <- summary(lm.mod.cobertura.ap)

sm

## Call:
## lm(formula = cobertura_copa.ap.mxn ~ superior_postgrado.porcentaje.mxn +
##     apartamento.porcentaje.mxn + cuarto.porcentaje.mxn + area_ep.porcentaje.mxn,
##     data = regresion.arboles.mn)

## Residuals:
##      Min       1Q   Median       3Q      Max 
## -0.29520 -0.05565 -0.00936  0.03833  0.65485 

## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                 0.08310   0.01664   4.995  9.7e-07 ***
## superior_postgrado.porcentaje.mxn 0.34914   0.02728  12.797  < 2e-16 ***
## apartamento.porcentaje.mxn      -0.03315   0.03408  -0.973   0.331    
## cuarto.porcentaje.mxn          -0.08121   0.05569  -1.458   0.146    
## area_ep.porcentaje.mxn         0.02887   0.03735   0.773   0.440    
## ---                        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Residual standard error: 0.1051 on 319 degrees of freedom
## Multiple R-squared:  0.4635, Adjusted R-squared:  0.4568

```

Table 4.13: Comparación OLS para el porcentaje de cobertura de copa con variables de población y otras dimensiones

medidasfit	%CC ~ pob	%CC ~ pob+otras
Shapiro-Wilk	0.90492	0.90720
SW p-value	0.00000	0.00000
Breusch-Pagan	16.82734	18.75791
BP p-value	0.00004	0.00088
Media Residuos	0.00000	0.00000
MSE	0.01100	0.01087
adj-Rsquare	0.45512	0.45681
AIC	-535.66435	-533.70683
Log likelihood	270.83218	272.85342

```
## F-statistic: 68.91 on 4 and 319 DF,  p-value: < 2.2e-16
```

Las variables introducidas tienen todas p-value no significativos. La inclusión de más términos al modelo empeoran el Criterio de Información de Akaike (sube de  $-535.6643522$  a  $-533.7068318$ ). Los resultados de los test no dan evidencias de mejoras importantes en Adjusted R-squared ( de 0.4551186 a 0.456814 ) o en MSE (de 0.011002 a 0.0108656). Sigue existiendo no normalidad en los residuos y heterocedasticidad, como se verifica en las gráficas diagnosticas de la regresión 4.62.

Finalmente presento los mapas de las variables dependientes de los 2 modelos seleccionados, comparandolas con el modelo ajustado y los residuos de ajuste, con el fin de observar donde se localizan los errores en la predicción. La figura 4.63 corresponde al ajuste del área de copa y la figura 4.64 al porcentaje de la cobertura de copa.

#### 4.4.3 Acceso a espacios verdes

En el caso de los EV contamos con una gran variedad de medidas sobre el acceso en relación con la distancia, el área disponible y diferentes formulaciones para aproximarse al concepto de acceso a un servicio ambiental (Figura 4.65). Para acotar el alcance de este trabajo, nos concentramos en dos métricas: el porcentaje de área de espacio verde de un sector censal

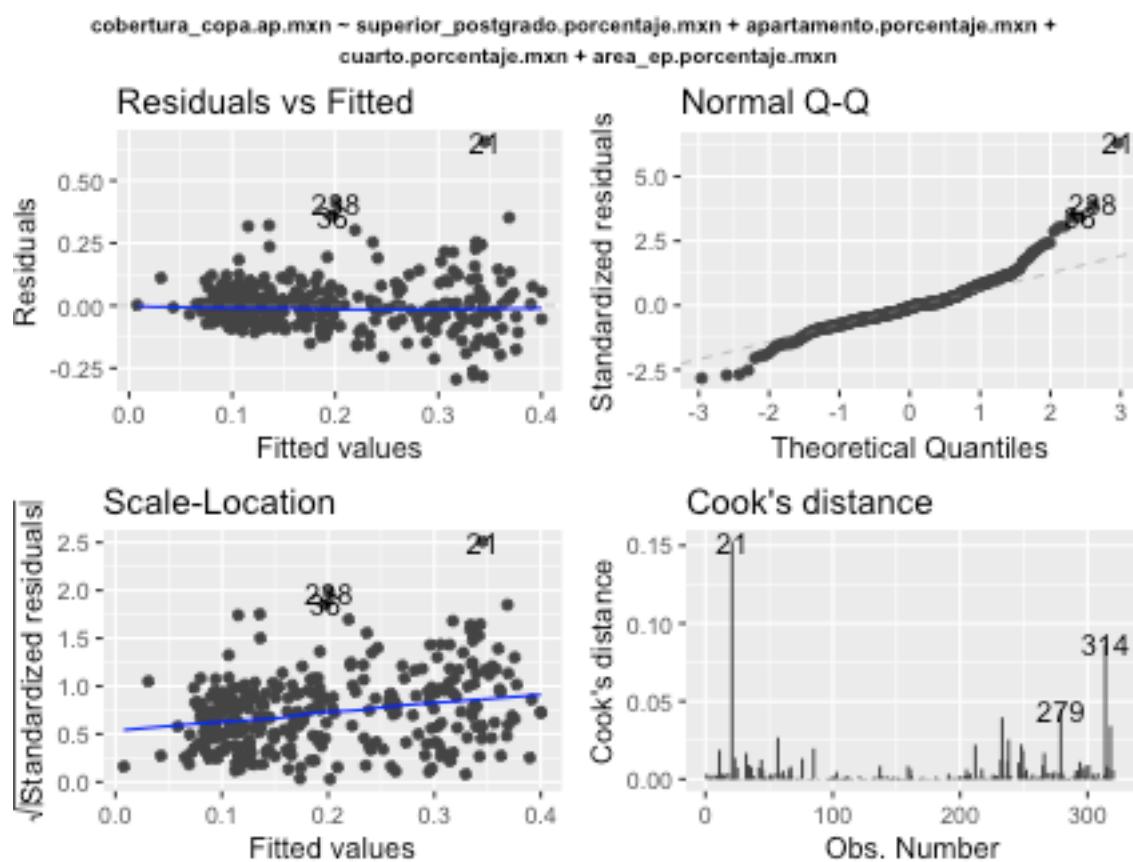


Figure 4.62: Gráficas diagnósticas para el análisis de regresión lineal de porcentaje de cobertura de copa con los nuevos términos

$\log.\text{area\_cpa.mx} \sim \text{superior\_postgrado.mx} + \text{densidad\_poblacion.mx} + \text{cuarto.porcentaje.mx} + \text{area\_ep.mx}$

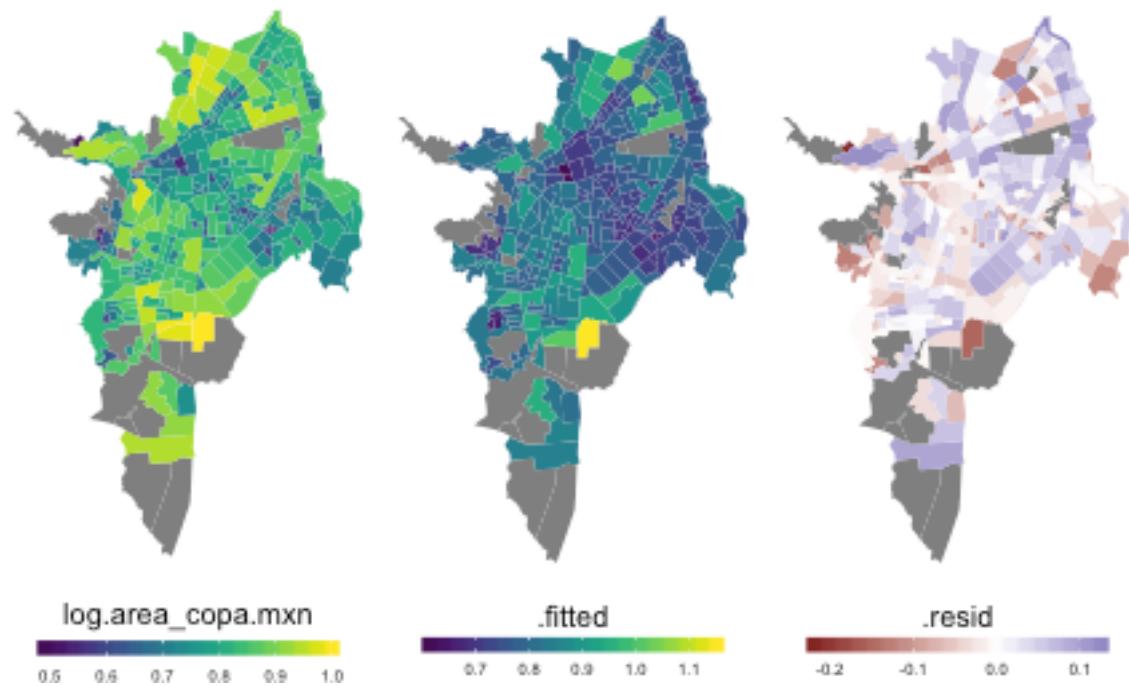


Figure 4.63: Variable dependiente, valor ajustado por el modelo y residuos del OLS para  $\log.\text{area\_cpa}$  normalizada

`cobertura_copa.ap.mx ~ superior_postgrado.porcentaje.mx`

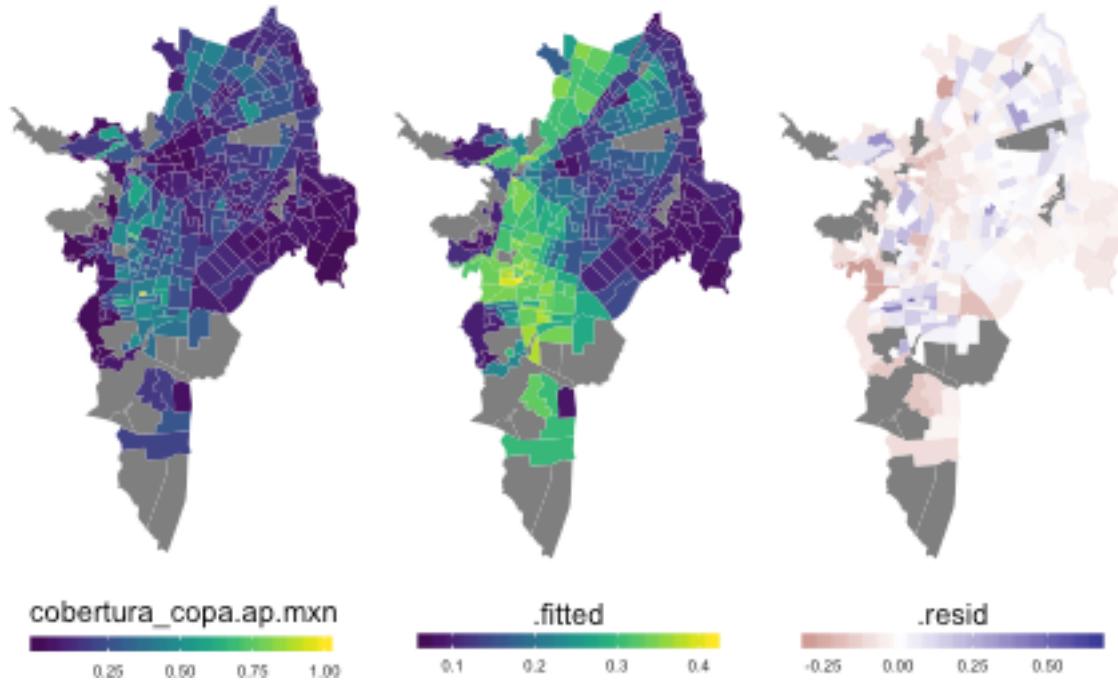


Figure 4.64: Variable dependiente, valor ajustado por el modelo y residuos del OLS para cobertura\_copa.ap normalizada

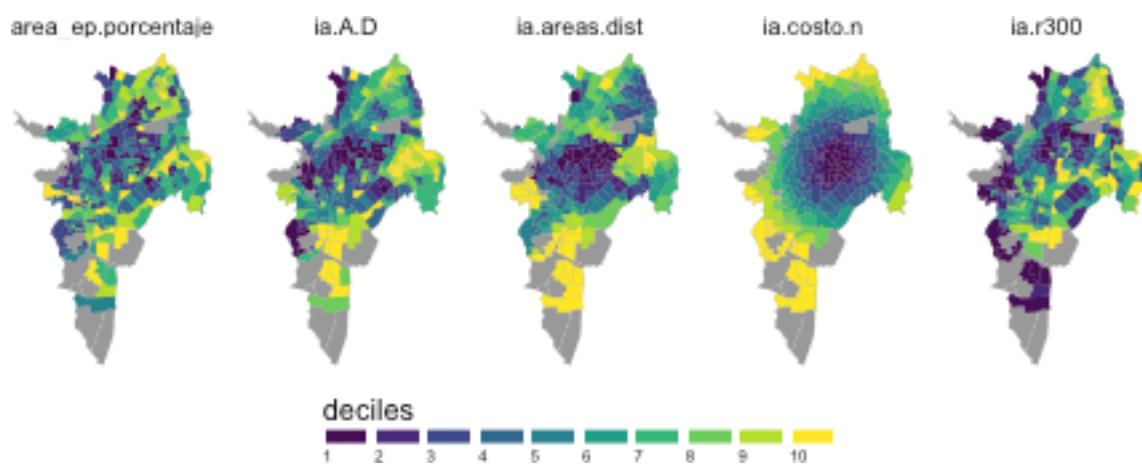


Figure 4.65: Metricas de acceso a espacio verde

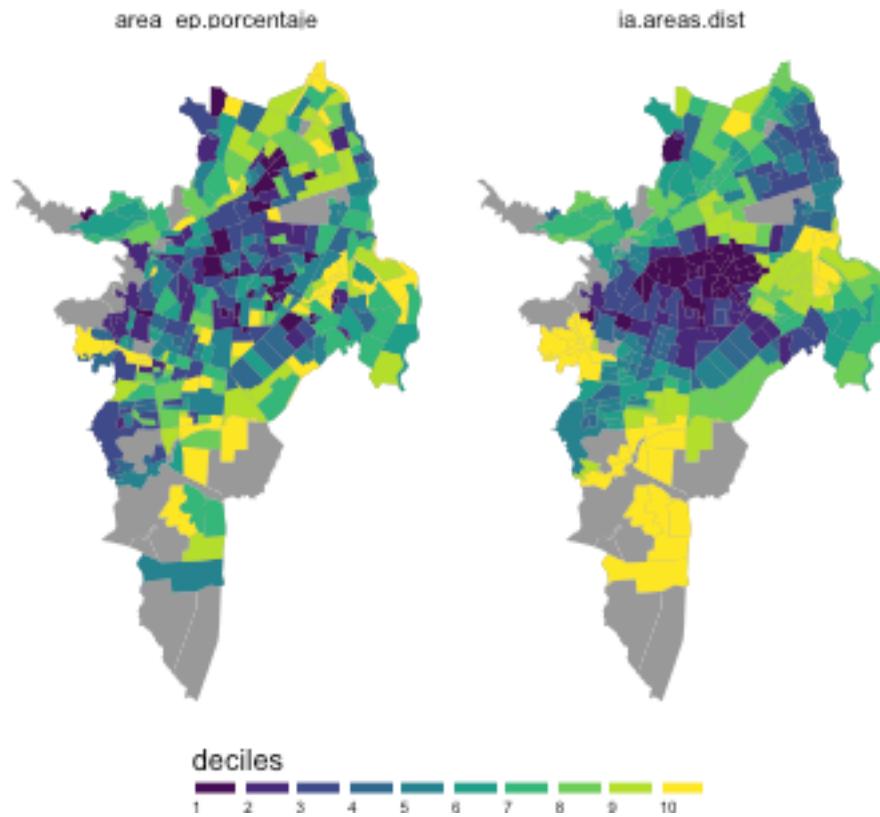


Figure 4.66: Metricas de acceso a espacio verdes seleccionadas

(`area_ep.porcentaje`), para aproximarse desde la idea de los beneficios principalmente a nivel local, y la razón área disponible distancia (`ia.areas.dist`), que tiende a formar agrupaciones de SU al rededor de las sectores donde se ubican espacios verdes, ya sea por número o por tamaño, que contemplan el fenómeno del acceso o beneficio como un proceso acotado por la distancia escogida de 1 kilómetro como radio de búsqueda; considerándola una distancia caminable para viajar en una ciudad como Cali. (Figura 4.66)

#### 4.4.3.1 Correlaciones y distribuciones bivariadas

Las variables de población y su correlación con los indicadores de acceso seleccionados nos sirven para seleccionar una vez más las variables independientes para usar en los modelos de regresión lineal. Las figuras 4.67 y 4.68 resumen los resultados del cálculo de los coeficientes de Pearson y Spearman respectivamente. Como se observa, esta

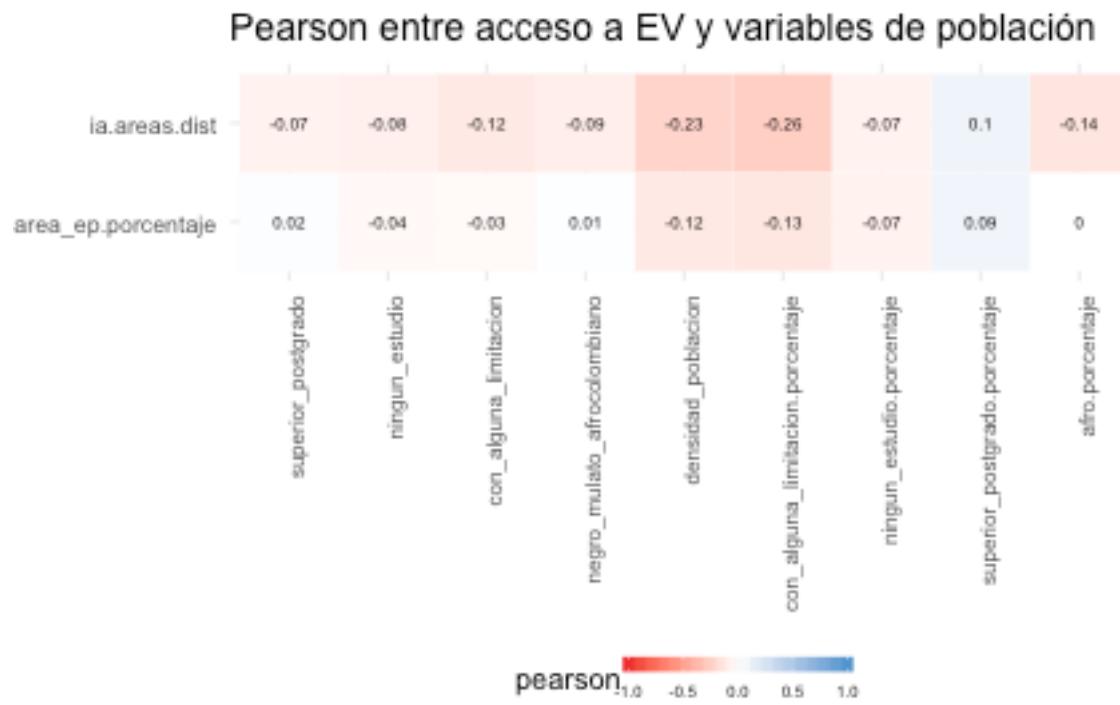


Figure 4.67: Coeficiente Pearson entre acceso a EV y variables de población

relación es muy débil, y en todos las variables (y para ambos coeficientes de correlación) es inferior a 0.3, un valor considerado muy bajo para incluir alguna de estas variables para que tenga éxito una aproximación lineal o no lineal al predecir o ajustar valores. Sin embargo, y como parte del proceso para indagar sobre el efecto en la estimación de parámetros de los modelos geoestadísticos, seleccionaremos las variables con mayor correlación: `densidad_poblacion,con_alguna_limitacion.porcentaje` mejor relacionadas con el índice de acceso `ia.areas.dist` y `ningun_estudio.porcentaje` para `area_ep.porcentaje`.

Resulta interesante ver, al igual que en el modelado de la cobertura de copa, si otras variables no poblacionales explican los resultados de los índices de acceso seleccionados. El conjunto de variables sobre el uso de los predios y sus coeficientes de correlación se muestran en las figuras 4.69 y 4.70. De nuevo las correlaciones son bajas, aparentemente poco explicativas de los índices de acceso. Las variables de uso de los predios que mejor se relacionan con los índices son: `unidad_economica.porcentaje` y el `cuarto.porcentaje`.

### Spearman entre acceso a EV y variables de población

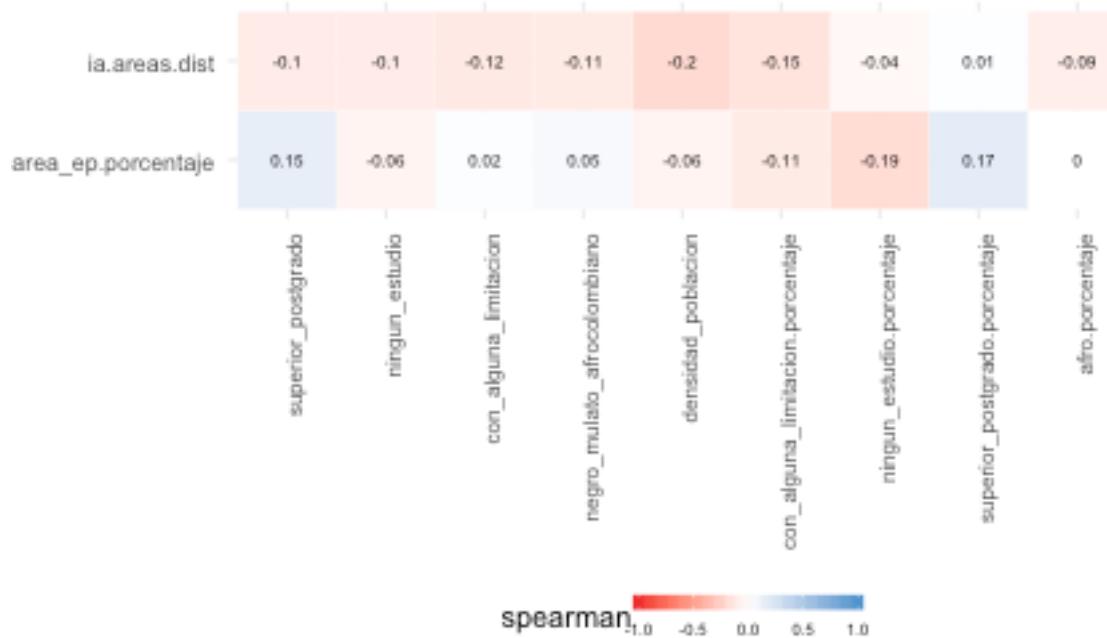


Figure 4.68: Coeficiente Spearman entre acceso a EV y variables de población

### Coeficiente Pearson entre dependiente EV e independientes uso de predios

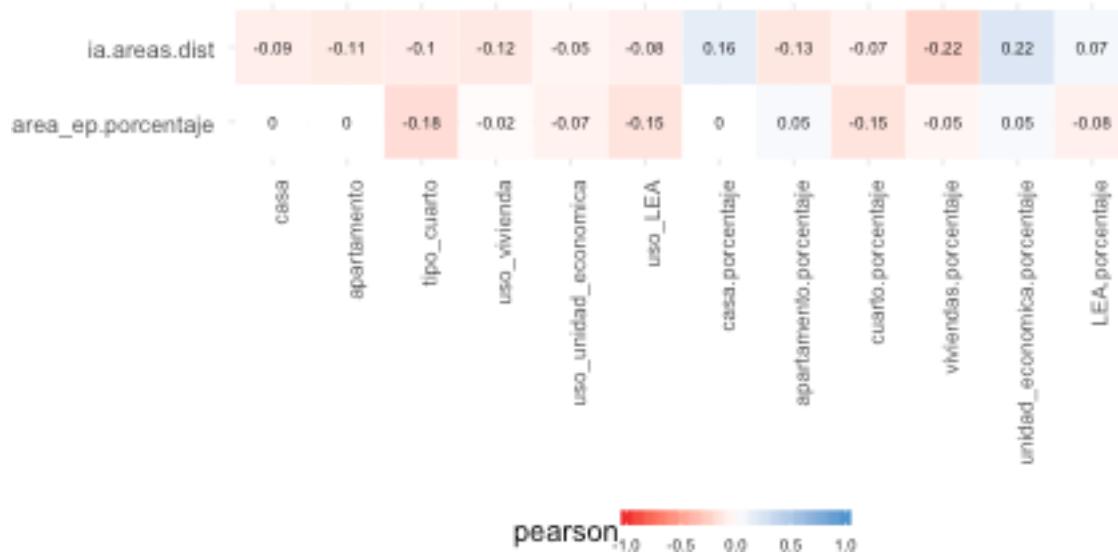


Figure 4.69: Coeficiente Pearson entre acceso a EV y variables de uso de los predios

### Coeficiente Spearman entre dependiente EV e independientes uso de predios

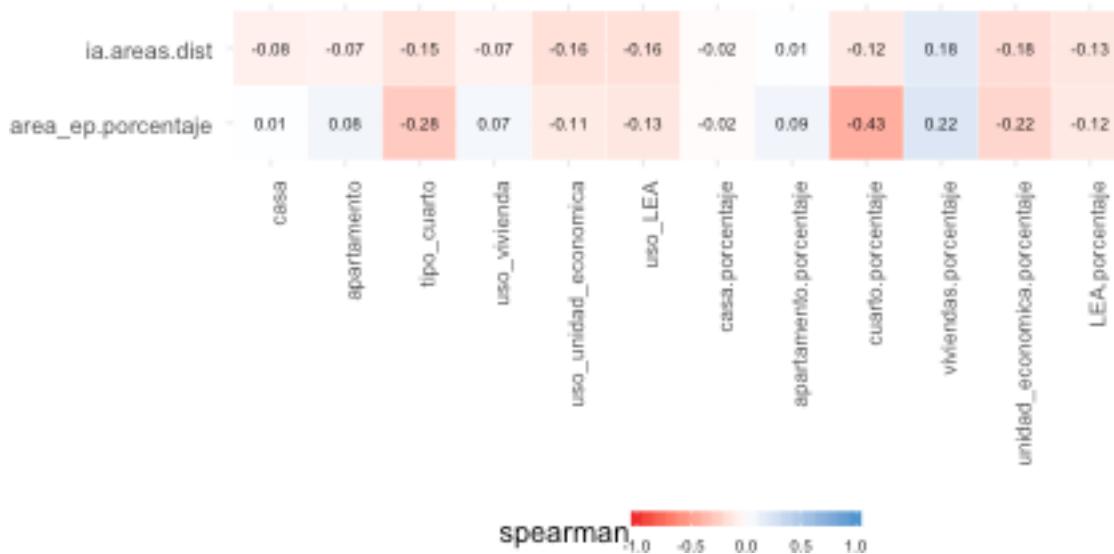


Figure 4.70: Coeficiente Spearman entre acceso a EV y variables de uso de los predios

El último bloque de variables indaga sobre las áreas y proporciones de las manzanas de cada sector censal y la vocación como pública o privada de los espacios dentro de un sector urbano. La figura 4.71 y 4.72 muestran que el área media de las manzanas (`area_media_manzana`) de los sectores urbanos se relaciona de forma positiva con ambos índices de acceso, mucho más fuertemente que las variables exploradas hasta el momento. Aunque parece haber una fuerte correlación de los indicadores de acceso con las áreas privadas, públicas y del sector urbano, estas hacen parte de los cálculos que generan estos índices, produciendo en efecto ficticio en la correlación, razón por la que no haremos uso de ellas en la modelación.

A manera de resumen e información de apoyo al proceso de modelación se incluye las distribuciones de bivariadas de términos independientes escogidos con el análisis de correlación y los índices de acceso a modelar.

### Coeficiente Pearson entre dependientes EV e independientes físicas de predios y SU

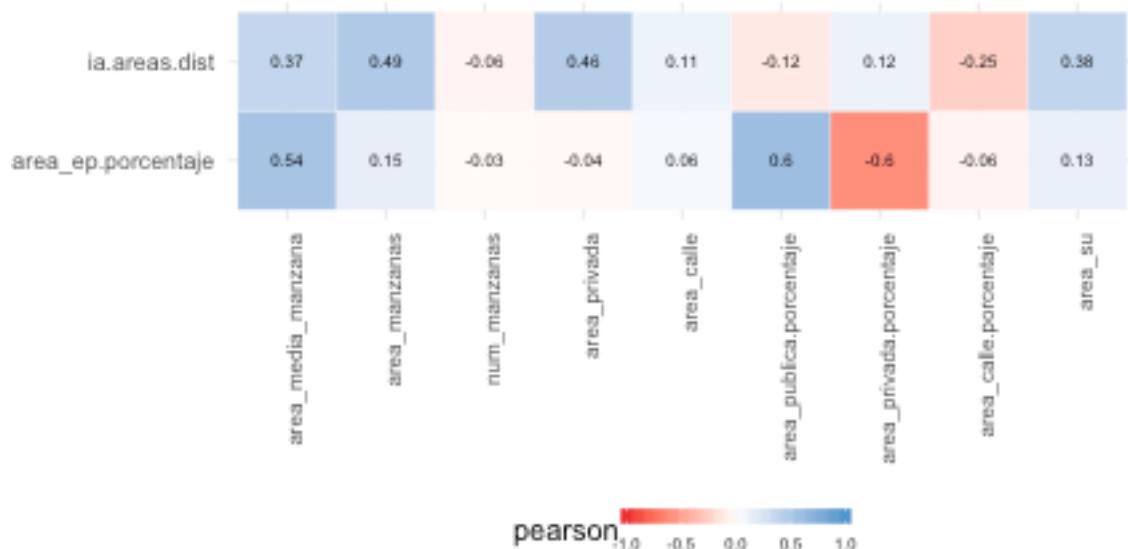


Figure 4.71: Coeficiente Pearson entre acceso a EV y variables sobre aspectos físicos de las manzanas y SU

### Coeficiente Spearman entre dependientes EV e independientes físicas de predios y SU



Figure 4.72: Coeficiente Spearman entre acceso a EV y variables sobre aspectos físicos de las manzanas y SU

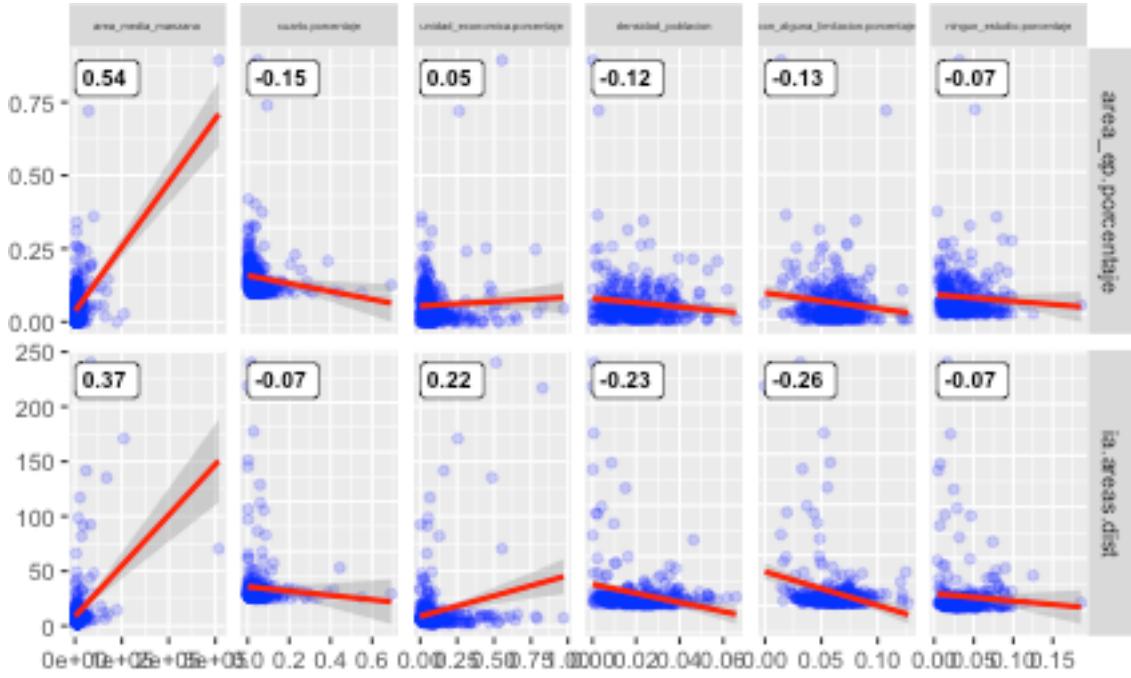


Figure 4.73: Distribución de variables independientes respecto de los índices de acceso a EV

#### 4.4.3.2 Modelos de regresión lineal

A diferencia de la cobertura de copa, donde el proceso de modelado fue incluyendo términos gradualmente para observar otros componentes no poblacionales, para los espacios verdes usaremos todos los términos seleccionados con los coeficientes de correlación para luego ver la significancia de las variables en el modelo y elegir el modelo con mejor ajuste usando criterio de información de Akaike (AIC) seleccionando los coeficientes significativos y comparando el AIC de la versión con todas las variables con el modelo simplificado. Para estos índices de acceso no usaremos variantes transformadas de la variable dependiente, solo se aplica una normalización a los datos. Para el índice de acceso de porcentaje de área de espacios verdes en un sector urbano el modelo ajusta el siguiente:

```
dependiente <- "area_ep.porcentaje"
independientes <- independientes.EV
# max normalizado
var_names <- c(dependiente, names(regresion.EV[, independientes]))
```

```

regresion.EV.mn <- max_nomalization(regresion.EV, var_names)

lm.area_ep.ptje <- crear_lm_from_df(regresion.EV.mn)
summary(lm.area_ep.ptje)

##

## Call:

## lm(formula = form, data = df)

##

## Residuals:

##      Min       1Q   Median       3Q      Max 
## -0.26917 -0.04344 -0.01571  0.02577  0.69572

## 

## Coefficients:

##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)               0.038434  0.016805  2.287  0.02284  
## area_media_manzana.mxn    0.760522  0.070424 10.799 < 2e-16  
## cuarto.porcentaje.mxn   -0.126650  0.048389 -2.617  0.00928  
## unidad_economica.porcentaje.mxn -0.000400  0.033502 -0.012  0.99048  
## densidad_poblacion.mxn    0.001654  0.026925  0.061  0.95105  
## con_alguna_limitacion.porcentaje.mxn 0.002491  0.032620  0.076  0.93918  
## ningun_estudio.porcentaje.mxn     0.055332  0.043902  1.260  0.20845  
## 

## (Intercept) * 
## area_media_manzana.mxn *** 
## cuarto.porcentaje.mxn ** 
## unidad_economica.porcentaje.mxn 
## densidad_poblacion.mxn 
## con_alguna_limitacion.porcentaje.mxn 
## ningun_estudio.porcentaje.mxn

```

```

## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07843 on 322 degrees of freedom
## Multiple R-squared: 0.3062, Adjusted R-squared: 0.2933
## F-statistic: 23.69 on 6 and 322 DF, p-value: < 2.2e-16

```

Los resultados muestran que las variables cuarto.porcentaje area\_media\_manzana son significativas. Con el modelo simplificado obtenemos los siguientes resultados:

```

dependiente <- "area_ep.porcentaje"
independientes <- independientes.EV
# max normalizado
var_names<-c(dependiente,names(regresion.EV[,independientes]))
regresion.EV.mn<-max_nomalization(regresion.EV,var_names)
lm.area_ep.ptje.sel<-lm(area_ep.porcentaje.mxn~cuarto.porcentaje.mxn+
                           # unidad_economica.porcentaje.mxn+
                           area_media_manzana.mxn#+
                           # densidad_poblacion.mxn+
                           #ningun_estudio.porcentaje.mxn
                           ,
                           data = regresion.EV.mn)

summary(lm.area_ep.ptje.sel)

##
## Call:
## lm(formula = area_ep.porcentaje.mxn ~ cuarto.porcentaje.mxn +
##     area_media_manzana.mxn, data = regresion.EV.mn)
##
## Residuals:
##       Min        1Q    Median        3Q       Max

```

Table 4.14: Comparación OLS del modelos con todas las variables y el modelo simplificado del porcentaje de area de espacio verde

medidasfit	Todas las variables	Variables seleccionadas
Shapiro-Wilk	0.77196	0.76782
SW p-value	0.00000	0.00000
Breusch-Pagan	23.18480	12.98572
BP p-value	0.00074	0.00151
Media Residuos	0.00000	0.00000
MSE	0.00602	0.00607
adj-Rsquare	0.29328	0.29656
AIC	-732.38377	-737.84913
Log likelihood	374.19189	372.92456

```
## -0.26869 -0.04398 -0.01599  0.02522  0.69949
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.049241   0.005703   8.635 2.66e-16 ***
## cuarto.porcentaje.mxn -0.093855   0.037643  -2.493  0.0132 *
## area_media_manzana.mxn  0.744371   0.065197  11.417 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07825 on 326 degrees of freedom
## Multiple R-squared:  0.3008, Adjusted R-squared:  0.2966
## F-statistic: 70.14 on 2 and 326 DF,  p-value: < 2.2e-16
```

En la tabla comparativa 4.14 que resume las medidas de ajuste de ambos modelos se evidencia los residuos presentan heterocedasticidad y no normalidad, lo que lo hace candidato a explorar el modelo ajuste geoestadístico. Existe una mejora lógica en el AIC del modelo simplificado debido a la reducción de términos del modelo. Los gráficos 4.74 muestran los resultados del ajuste en relación con los residuos y 4.75 espacialmente.

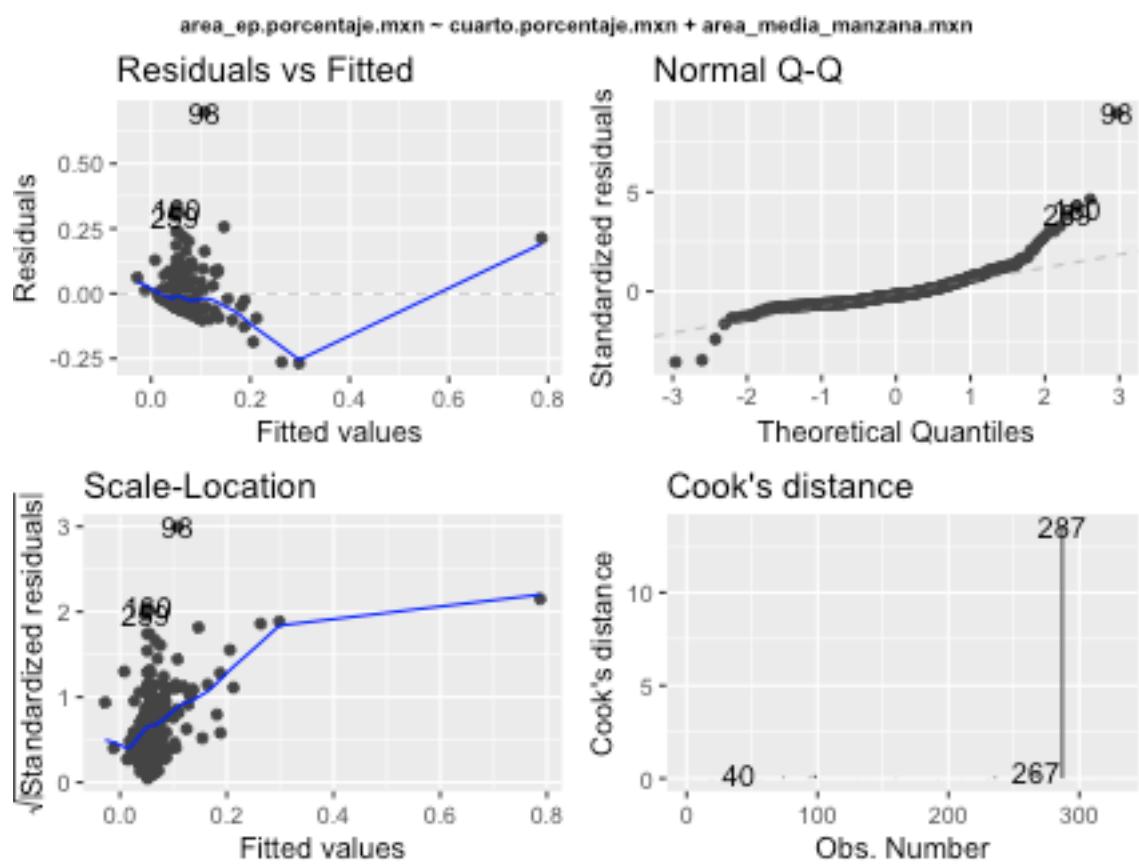


Figure 4.74: Gráficas diagnósticas para el análisis de regresión lineal de porcentaje de área de espacio verde

`area_ep.porcentaje.mxn ~ cuarto.porcentaje.mxn + area_media_manzana.mxn`

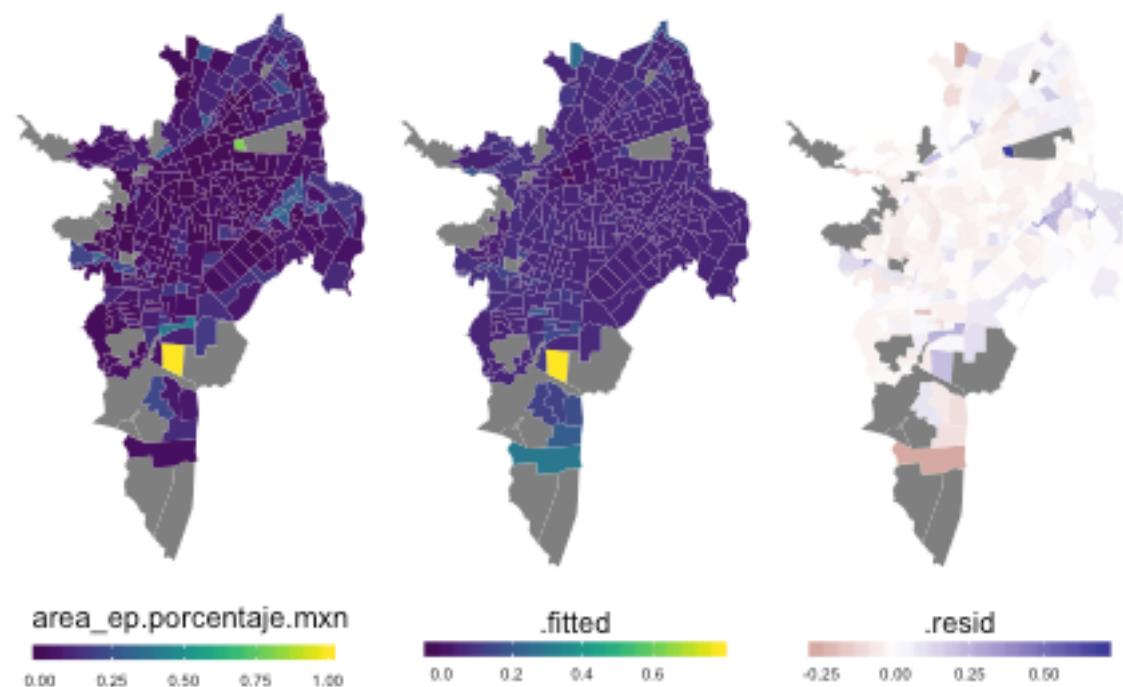


Figure 4.75: Variable dependiente, valor ajustado por el modelo y residuos del OLS para ‘area ep.porcentaje’ normalizada

Para el indice de acceso ia.areas.dist el modelo con todos los terminos es el siguiente:

```
dependiente <- "ia.areas.dist"
independientes <- independientes.EV
# max normalizado
var_names <- c(dependiente, names(regresion.EV[, independientes]))
regresion.EV.mn <- max_normalization(regresion.EV, var_names)
lm.ia.areas.dist <- crear_lm_from_df(regresion.EV.mn)
summary(lm.ia.areas.dist)

##
## Call:
## lm(formula = form, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29640 -0.03455 -0.01586  0.01009  0.82344
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  0.09193   0.02051   4.483 1.03e-05
## area_media_manzana.mxn      0.43284   0.08594   5.036 7.92e-07
## cuarto.porcentaje.mxn     -0.16017   0.05905  -2.712 0.007043
## unidad_economica.porcentaje.mxn  0.11734   0.04089   2.870 0.004376
## densidad_poblacion.mxn     -0.04921   0.03286  -1.498 0.135208
## con_alguna_limitacion.porcentaje.mxn -0.13414   0.03981  -3.369 0.000845
## ningun_estudio.porcentaje.mxn      0.14228   0.05358   2.656 0.008309
##
## (Intercept) ***
## area_media_manzana.mxn ***
```

```

## cuarto.porcentaje.mxn          **
## unidad_economica.porcentaje.mxn    **
## densidad_poblacion.mxn
## con_alguna_limitacion.porcentaje.mxn ***
## ningun_estudio.porcentaje.mxn      **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09572 on 322 degrees of freedom
## Multiple R-squared: 0.2071, Adjusted R-squared: 0.1923
## F-statistic: 14.02 on 6 and 322 DF, p-value: 3.537e-14

```

Solo la densidad de población no es significativa, sin embargo, como se muestra en la tabla 4.15 , al eliminarla, empeora levemente el AIC, el MSE y el adj-Rsquare . Consideraremos el modelo sin simplificar para los ajustes geoestadístico. Los gráficos 4.76 muestran los resultados del ajuste del modelo con todas las variables en relación con los residuos y 4.77 espacialmente.

```

##
## Call:
## lm(formula = ia.areas.dist.mxn ~ cuarto.porcentaje.mxn + unidad_economica.porcentaje.
##     area_media_manzana.mxn + ningun_estudio.porcentaje.mxn +
##     con_alguna_limitacion.porcentaje.mxn, data = regresion.EV.mn)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.31567 -0.03596 -0.01589  0.00814  0.82468
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
```

Table 4.15: Comparación OLS de los modelos con todas las variables y el modelo simplificado de índice de acceso ‘ia.Areas.dist’

medidasfit	Todas las variables	Variables seleccionadas
Shapiro-Wilk	0.57372	0.56228
SW p-value	0.00000	0.00000
Breusch-Pagan	57.30768	57.77627
BP p-value	0.00000	0.00000
Media Residuos	0.00000	0.00000
MSE	0.00897	0.00903
adj-Rsquare	0.19231	0.18921
AIC	-601.32470	-601.04097
Log likelihood	308.66235	307.52049

```

## (Intercept)          0.08100  0.01920  4.218 3.20e-05
## cuarto.porcentaje.mxn -0.14781  0.05859 -2.523 0.012121
## unidad_economica.porcentaje.mxn 0.13705  0.03879  3.533 0.000470
## area_media_manzana.mxn      0.45622  0.08468  5.388 1.38e-07
## ningun_estudio.porcentaje.mxn 0.11603  0.05073  2.287 0.022817
## con_alguna_limitacion.porcentaje.mxn -0.14140  0.03959 -3.572 0.000409
##
## (Intercept)          ***
## cuarto.porcentaje.mxn *
## unidad_economica.porcentaje.mxn ***
## area_media_manzana.mxn      ***
## ningun_estudio.porcentaje.mxn *
## con_alguna_limitacion.porcentaje.mxn ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0959 on 323 degrees of freedom
## Multiple R-squared:  0.2016, Adjusted R-squared:  0.1892
## F-statistic: 16.31 on 5 and 323 DF,  p-value: 2.386e-14

```

```

ia.areas.dist.mxn ~ cuarto.porcentaje.mxn + unidad_economica.porcentaje.mxn +
area_media_manzana.mxn + ningun_estudio.porcentaje.mxn +
con_alguna_limitacion.porcentaje.mxn

```

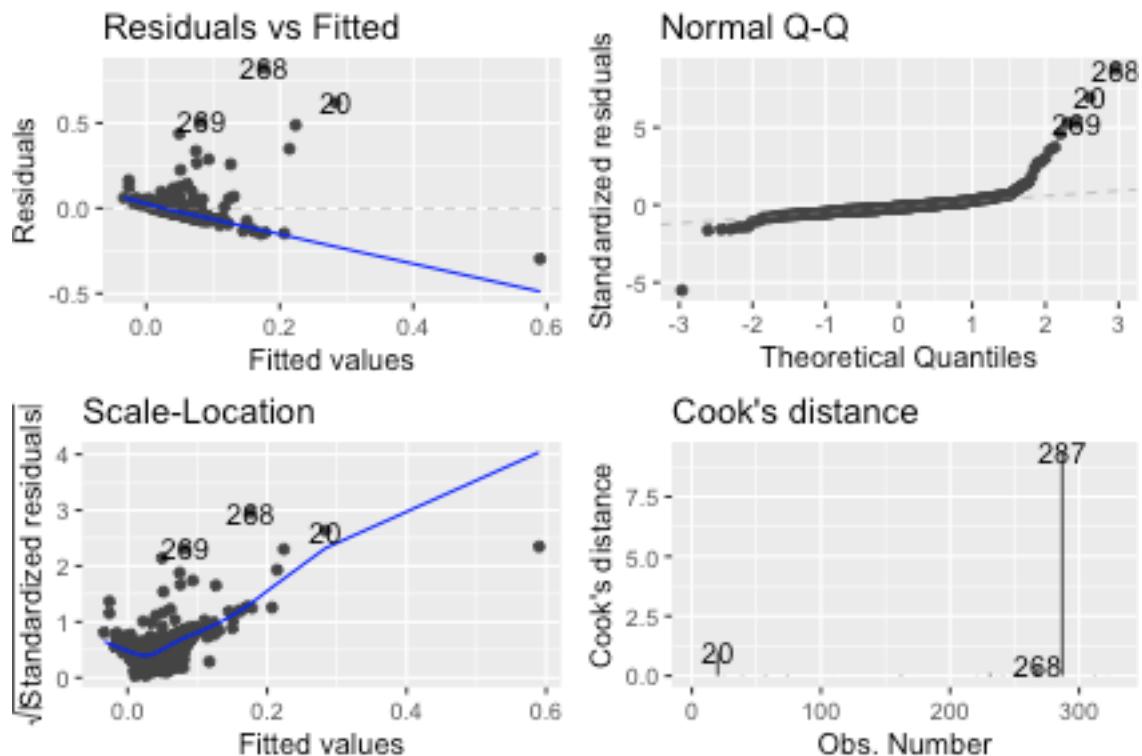


Figure 4.76: Gráficas diagnósticas para el análisis de regresión lineal de 'ia.areas.dist'

ia.Areas.dist.mx ~ area\_media\_manzana.mx + cuarto.porcentaje.mx +  
unidad\_economica.porcentaje.mx + densidad\_poblacion.mx +  
con\_alguna\_limitacion.porcentaje.mx + ningun\_estudio.porcentaje.mx

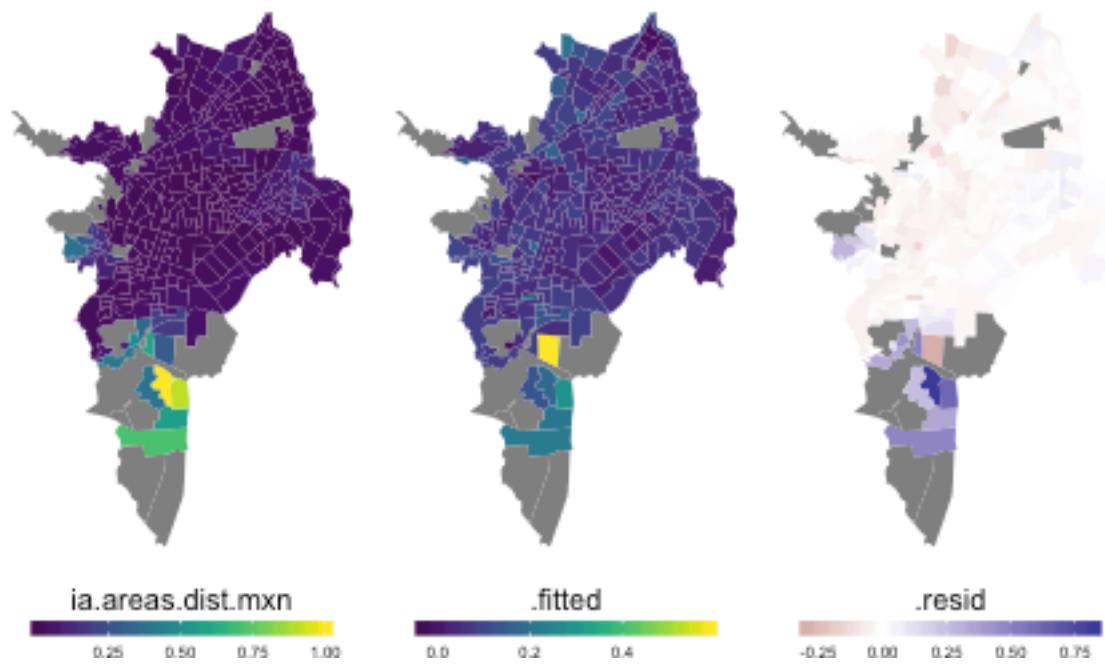


Figure 4.77: Variable dependiente, valor ajustado por el modelo y residuos del OLS para ‘ia.Areas.dist’ normalizada

## 4.5 Análisis geoestadísticos

Para los análisis geoestadísticos introducimos los modelos autoregresivos para obtener mejoras en la estimación de los coeficientes y en el ajuste de los modelo si existe algún tipo de autocorrelación espacial en los residuos. Existe una variedad de estos modelos que capturan diferentes tipos de efectos: modelo autoregresivo SAR que capturan efectos de la variable dependiente, ecuación (2.8), sobre las variables independientes (spatial lag o retardo espacial en  $X$  SLX, ecuación (2.10), en el error (modelo espacial del error SEM, ecuación (2.9) o usando una combinación del modelo de error y autoregresivo (modelo espacial de Durbin SD, ecuación (2.11)). Todas estas aproximaciones introducen una matriz de  $W_{n \times n}$ , donde  $n$  es el número de sitios, que captura la influencia de las variables en relación con su proximidad. Esta matriz  $W$  es una estructura que restringe la influencia a priori en el modelo. Para observar el efecto que tiene esta matriz sobre los resultados del modelo usaremos 2 matrices distintas, y veremos su impacto en la estimación.

Para los análisis espaciales usaremos la librería **spdep** (Bivand, 2017)

### 4.5.1 Matrices de vecindad

La matriz  $W$  representa la topología de vecindad entre los sectores censales. Existen en la literatura diferentes tipos de vecindad: rook, bishop y queen son las más referenciadas. Esta vecindad está representada en la matriz con 1 cuando existe vecindad y 0 cuando no. Otra forma de cuantificar la interacción de esa vecindad es usando una matriz de inversos de la distancia entre los centroides de los sectores censales, con el fin de atenuar la interacción entre sectores muy alejados y tener una variable continua que representa esa influencia. En la figura 4.78 se muestra la matriz  $W$  definida para vecinos que comparten un lado del polígono (vecindad rook) para todos los sectores de la ciudad de Cali.

Sin embargo, las regresiones se realizaron sobre un subconjunto de los datos, y por tanto la estructura de esta matriz debe tener esto en cuenta, o mejor, no tener en cuenta la



Figure 4.78: Grafo de vecindad entre todos los SU de la ciudad de Cali



Figure 4.79: Grafo de vecindad entre los SU seleccionados para el análisis

influencia de estos sectores excluidos. Así la matriz de vecindad para los SU usados para la estimación de los coeficientes de las regresiones lineales se ve en la figura 4.79.

Las matrices de vecindad construidas para el análisis espacial son la *queen*  $W_q$ , que considera vecino a todos los sectores que comparten un lado o una esquina con un sector censal; y una matriz de distancia inversas entre los centroides de los SU, restringiendo la vecindad a aquellos centroides que están a menos de 1 km ( $W_d$ ). El valor de un kilómetro es arbitrario, aunque razonable en la escala humana. Los grafos que representan las 2 matrices  $W$  se muestran en la figura 4.80.

La matriz  $W_d$  impone un estructura de interacciones que puede relacionar sectores en una zona que no necesariamente comparten ningún lado o esquina pero que están cercanos, mientras que  $W_q$  se restringe a condiciones de vecindad sólo entre sectores contiguos. En esa medida puede existir juego para dar interpretación teórica al fenómeno de derrame

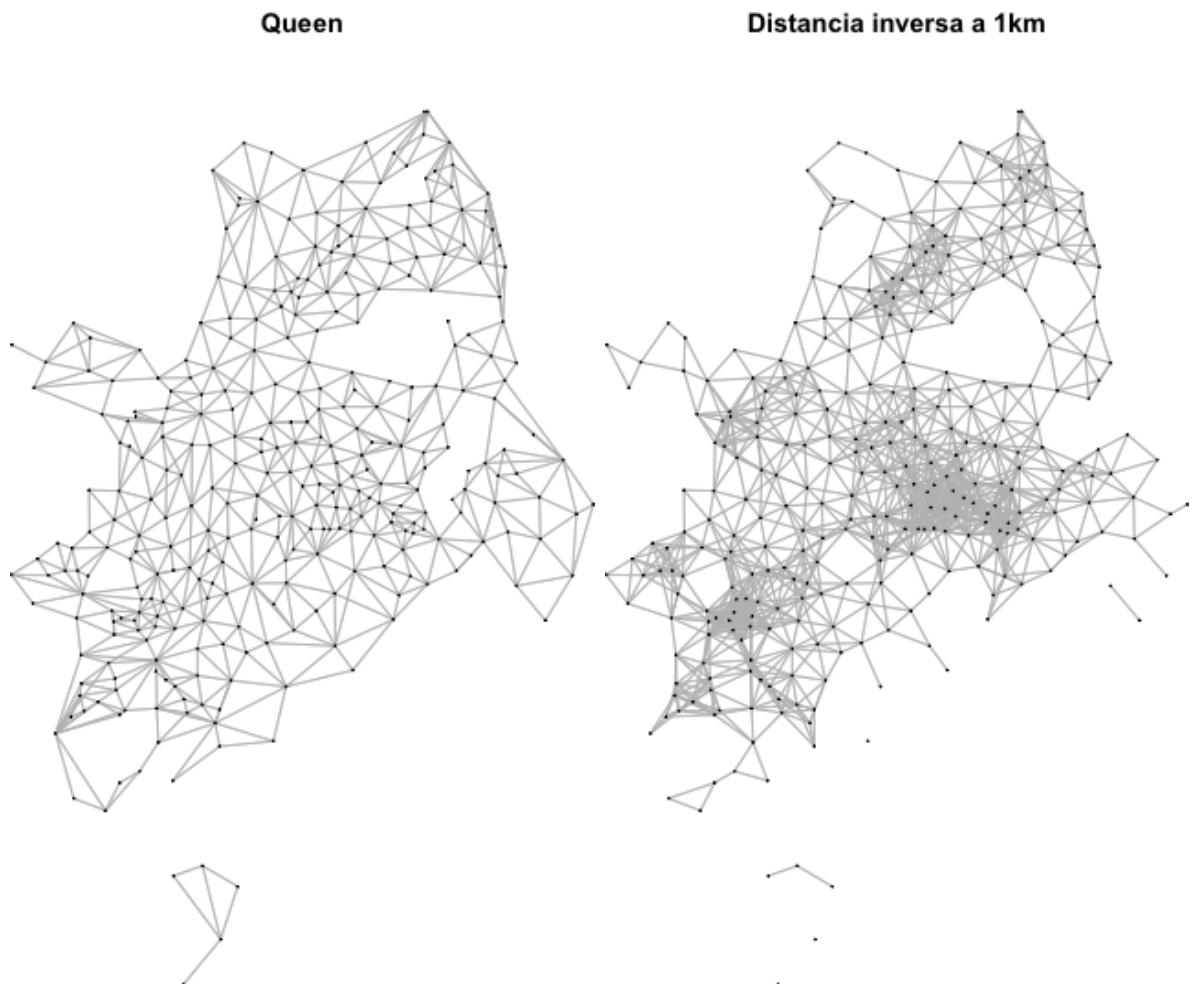


Figure 4.80: Matrices de vecindad del análisis espacial

o influencia que ejercen sobre el ajuste de los modelos. Un ejemplo puede ser que la dependencia espacial de las coberturas de árboles que expresa  $W_d$  es una característica en una zona no limitada por las divisiones del territorio con base en los desarrollos urbanísticos (barrios) sino que se ajustan más a fenómenos de dispersión continuos con base en el alcance escogido. Así, la  $W_q$  puede interpretarse como una forma de dar relevancia a la continuidad entre barrios y su importancia como unidad de desarrollo urbano en las variaciones de la variable a predecir.

## 4.5.2 Modelado espacial de la cobertura de copa

### 4.5.2.1 Autocorrelación espacial

Para indagar sobre la información o patrones espaciales de los residuos de los modelos de regresión usaremos el índice de Moran I. El índice de Moran I es el coeficiente de correlación para la relación entre una variable y sus valores circundantes. Si encontramos una correlación espacial significativa en los residuos, esto sugiere que agregando esa estructura en el modelo podremos obtener una estimación más eficiente de los coeficientes, y en consecuencia un mejor entendimiento de la relación entre esas variables. Hay que recordar que en este ejercicio no estamos queriendo entender una población por una muestra, estamos calculando estos coeficientes sobre el total de la población, y por tanto los coeficientes pueden interpretarse como la fuerza de esa relación. La confianza en esa estimación depende de que los residuos obtenidos sean tenga un valor medio de 0, y que no puedan distinguirse del ruido. La ecuación (4.10) define matemáticamente el índice:

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2} \quad (4.10)$$

donde  $N$  es el número de unidades espaciales indexados por  $i$  y  $j$ ;  $X$  es la variable de interés;  $\bar{X}$  es la media de  $X$ ; y  $w_{ij}$  es un elemento de una matriz de pesos espaciales  $W$ . Un valor de 0 de Moran'I indica un patrón espacial aleatorio. Si existe autocorrelación

los valores son positivos y el máximo es 1. Si los valores son negativos decimos que existe dispersión, siendo -1 el mínimo valor posible representando la dispersión perfecta.

El gráfico de Moran es una forma de observar el valor de la pendiente (el índice de autocorrelación) graficando los valores retardados (spatial lag: es como dijimos previamente el valor medio de los valores vecinos) de la variable en cuestión en el eje *y* y la variable en el *x*. El valor *p* del test estadístico nos dice qué tan seguros estamos que esa pendiente no es plana, por lo que se espera que sean menores que el valor límite de significancia  $\alpha = 0.05$

El hecho de que de la Morán I es una suma de productos cruzados individuales es explotado por los “indicadores locales de asociación espacial” (LISA) para evaluar la agrupación de las unidades individuales mediante el cálculo de la I de Moran local para cada unidad espacial y la evaluación de la significación estadística para cada I. (Wikipedia, 2017)

Los mapas que acompaña el resultado numérico y el grafico de Moran son el valor z-normalizado del LISA, el valor *p* y el mapa de clusters. En este ultimo mapa las regiones resaltadas en rojo tienen valores altos de la variable y tienen vecinos con valores altos también (high-high). El área azul es low-low los vecinos presentab valores bajos al igual que sus vecinos. Mientras que las regiones azul pálido son low-high y las áreas rosadas son high-low muestran correlación negativa, es decir valores muy diferentes a los de sus vecinos. Las regiones fuertemente coloreadas son aquellas que contribuyen significativamente a un resultado positivo de autocorrelación espacial global, mientras que los colores más claros contribuyen significativamente a un resultado de autocorrelación negativo.

Examinemos primero la autocorrelación de las variables dependientes.

#### 4.5.2.1.1 Variables dependientes

```
moran.test(lm.mod.area_copa$model$log.area_copa.mxn, listw = W_queen)
```

```
##
```

```

## Moran I test under randomisation
##
## data: lm.mod.area_copa$model$log.area_copa.mxn
## weights: W_queen
##
## Moran I statistic standard deviate = 10.958, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##          0.374881115     -0.003095975     0.001189781
moran.test(lm.mod.area_copa$model$log.area_copa.mxn, listw = W_dist1000.inv,
            zero.policy = TRUE)

##
## Moran I test under randomisation
##
## data: lm.mod.area_copa$model$log.area_copa.mxn
## weights: W_dist1000.inv
##
## Moran I statistic standard deviate = 9.5098, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##          0.2817402626    -0.0031250000    0.0008972897

```

La matrix  $W_q$  capatura mucho mejor de forma global la autocorrelacion del area de copa. Los mapas LISA muestran los focos de esta autocorrelación usando la matriz para  $W_q$  (figura 4.82) y para  $W_d$  (figura 4.83) .

En los estos mapas de LISA de ambas matrices se observan que los sectores en rojo (H-H)

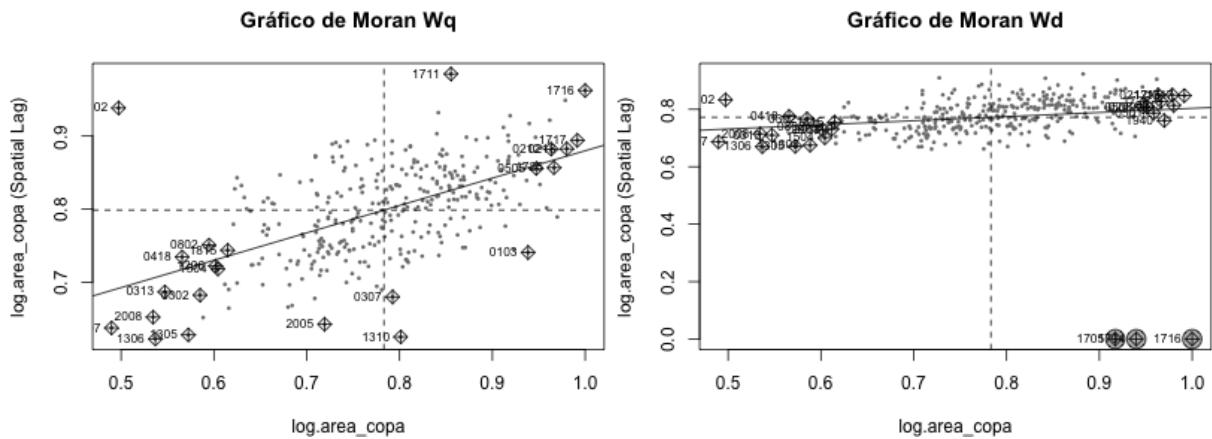


Figure 4.81: Gráfico de Moran para los residuos del modelo lineal para del área de cobertura de copa usando  $W_q$

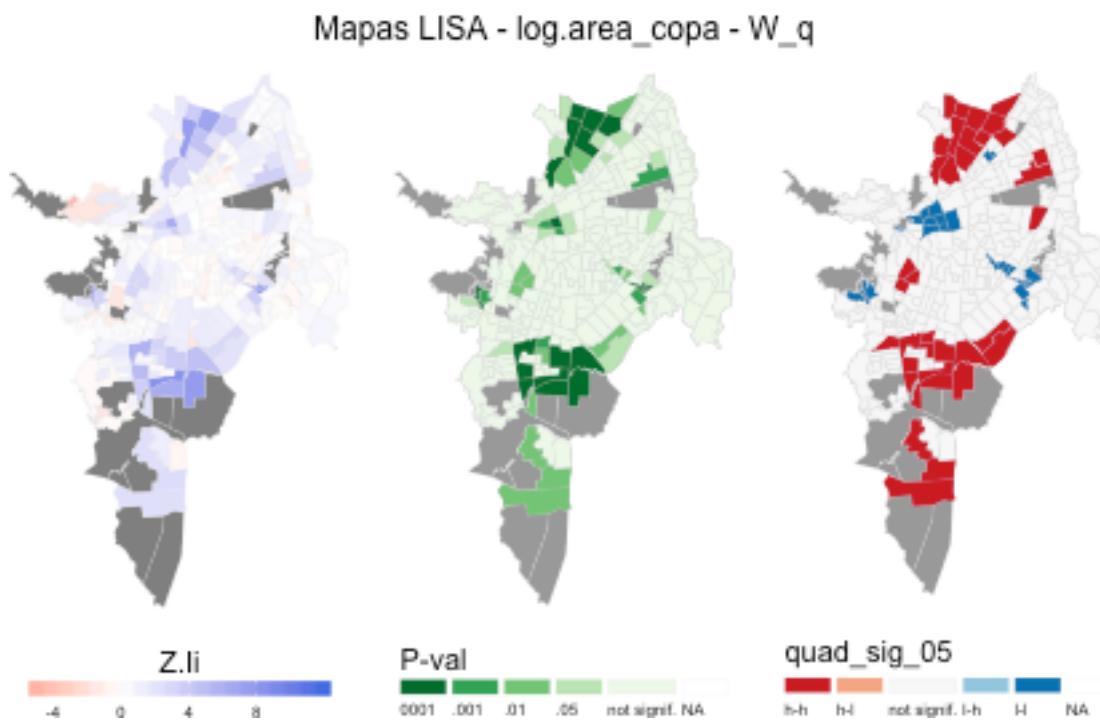


Figure 4.82: Mapas LISA para la matriz  $W_q$  de 'log.area copa'

Mapas LISA - log.area\_copa - W\_d

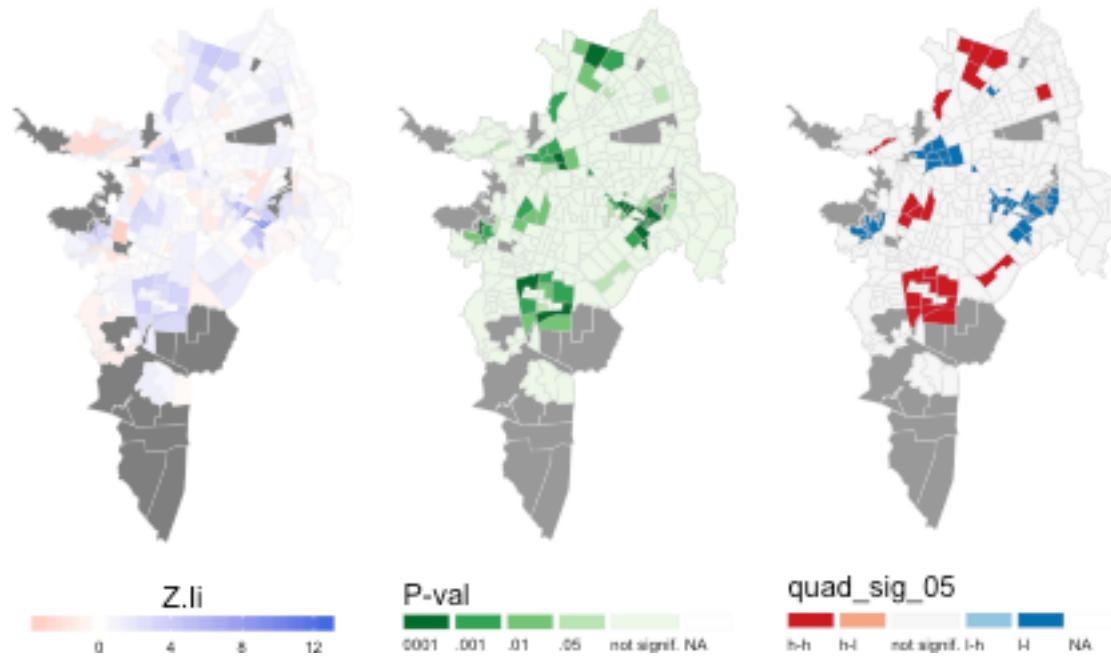


Figure 4.83: Mapas LISA para la matriz  $W_d$  de ‘log.area copa’

y azul (L-L) identifican los lugares con autocorrelación positiva, formando grupos. No se presentan valores negativos de autocorrelación.

Para la variable dependiente porcentaje de cobertura de copa los resultados del análisis de autocorrelación se presentan a continuación.

```
moran.test(lm.best.cobertura.ap$model$cobertura_copa.ap.mx, listw = W_queen)

##
## Moran I test under randomisation
##
## data: lm.best.cobertura.ap$model$cobertura_copa.ap.mx
## weights: W_queen
##
## Moran I statistic standard deviate = 14.107, p-value < 2.2e-16
## alternative hypothesis: greater
```

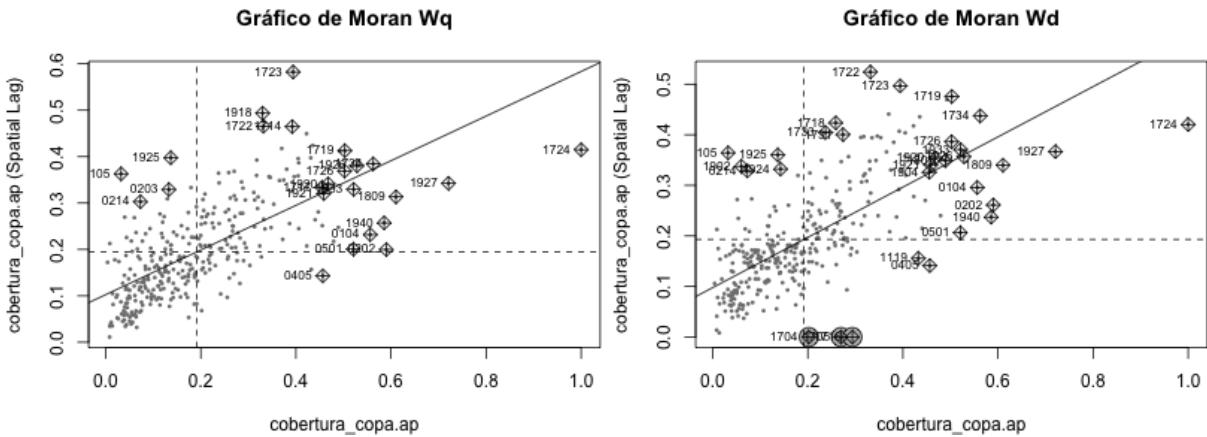


Figure 4.84: Gráfico de Moran para los residuos del modelo lineal para el porcenatajes cobertura de copa usando  $W_q$

```

## sample estimates:
## Moran I statistic      Expectation      Variance
##          0.481041232     -0.003095975    0.001177840

moran.test(lm.best.cobertura.ap$model$cobertura_copa.ap.mx, listw = W_dist1000.inv,
            zero.policy = TRUE)

##
## Moran I test under randomisation
##
## data: lm.best.cobertura.ap$model$cobertura_copa.ap.mx
## weights: W_dist1000.inv
##
## Moran I statistic standard deviate = 16.957, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##          0.5022505518     -0.0031250000    0.0008881907

```

Ambos diseños de matriz revelan precencia clara de autocorrelación espacial, pero la matrix  $W_d$  captura un legeramente mejor con un valor positivo y significativa, produciendo *clusters*

Mapas LISA - cobertura\_copa.ap - W\_q

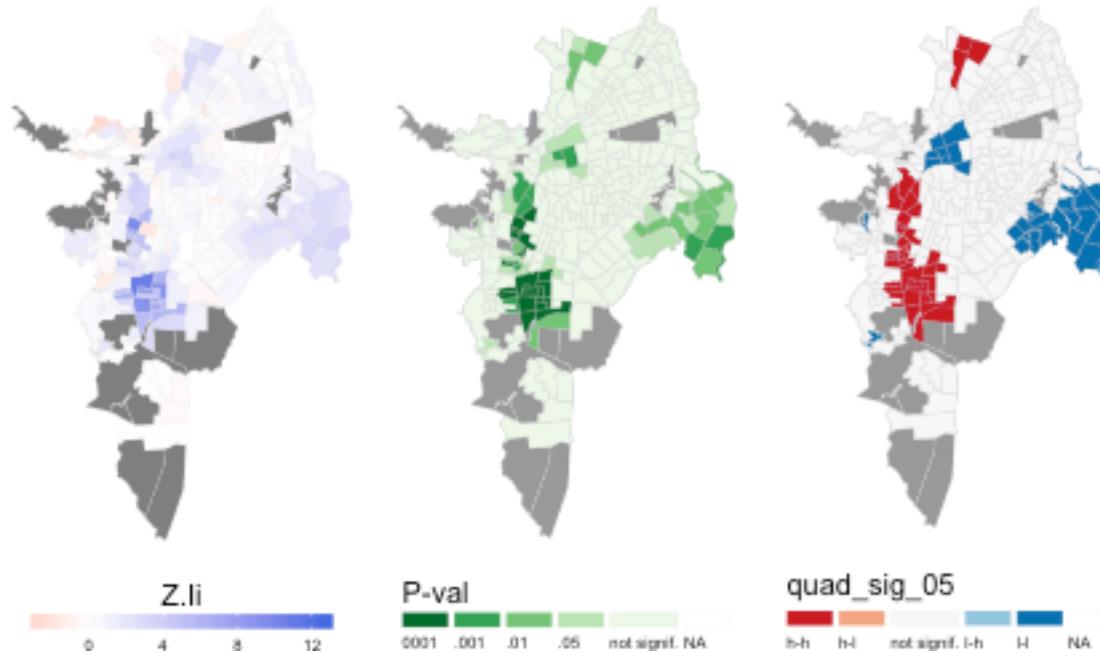


Figure 4.85: Mapas LISA para la matriz  $W_q$  de ‘cobertura copa.ap’

de sectores urbanos más con más sectores urbanos. Los mapas LISA muestran los focos de esta autocorrelación usando la matriz para  $W_q$  (figura 4.85) y para  $W_d$  (figura 4.86) . No se presentan valores negativos de autocorrelación.

#### 4.5.2.1.2 Residuos de los OLS

El siguiente código calcula los índices de autocorrelación de los residuos del mejor modelo lineal de área de copa (`sqrt.copa_area.mxn ~ superior_postgrado.mxn + densidad_poblacion.mxn + cuarto.porcentaje.mxn + area_ep.mxn`) para ambas matrices  $W$  y construye los gráficos de Moran y mapas LISA 4.87 para  $W_q$  y para  $W_d$ .

```
lm.morantest(lm.mod.area_copa, W_queen, alternative = "two.sided", zero.policy = T)

## Global Moran I for regression residuals
##
```

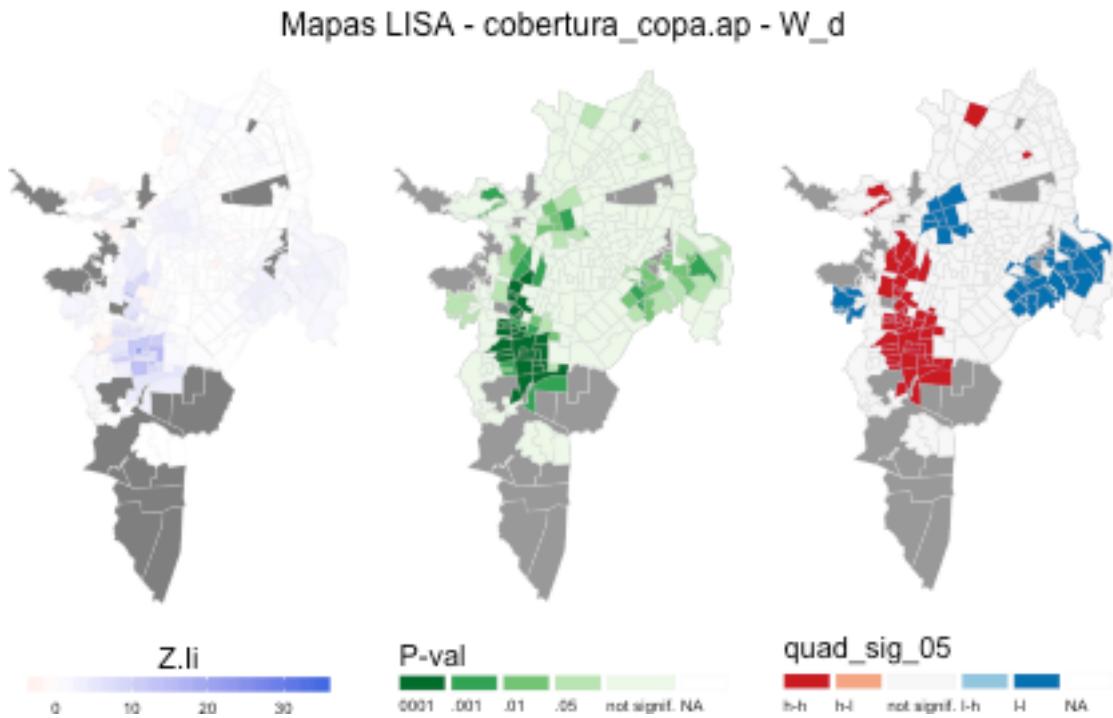


Figure 4.86: Mapas LISA para la matriz  $W_d$  de ‘cobertura copa.ap’

```

## data:
## model: lm(formula = log.area_copa.mxn ~ superior_postgrado.mxn +
## densidad_poblacion.mxn + cuarto.porcentaje.mxn + area_ep.mxn, data
## = regresion.arboles.mn)
## weights: W_queen
##
## Moran I statistic standard deviate = 3.0842, p-value = 0.002041
## alternative hypothesis: two.sided
## sample estimates:
## Observed Moran I           Expectation          Variance
##          0.096598455      -0.008456178      0.001160269
lm.morantest(lm.mod.area_copa, W_dist1000.inv, alternative = "two.sided", zero.policy
## 
## Global Moran I for regression residuals

```

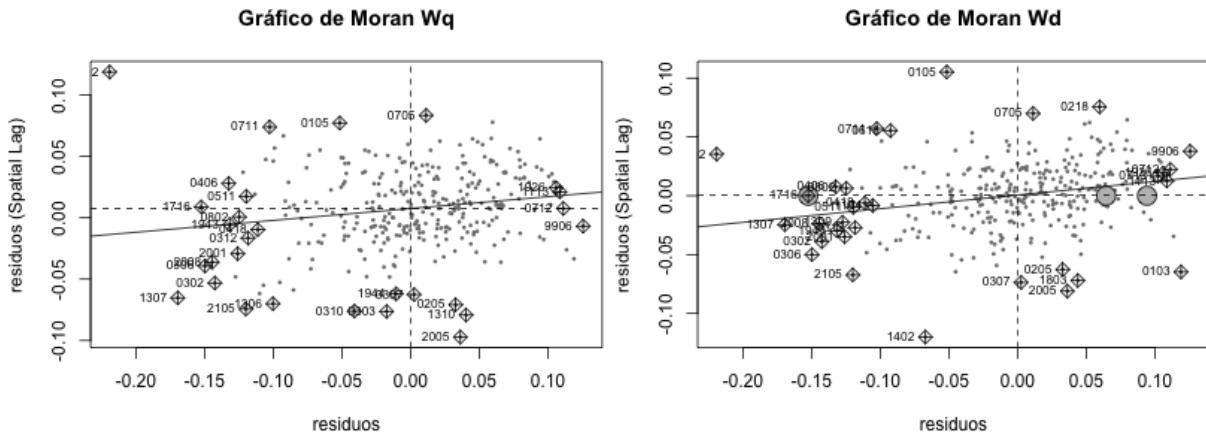


Figure 4.87: Gráfico de Moran para los residuos del modelo lineal para el área de cobertura de copa usando  $W_q$  y  $W_d$

```
##
## data:
## model: lm(formula = log.area_copa.mxn ~ superior_postgrado.mxn +
## densidad_poblacion.mxn + cuarto.porcentaje.mxn + area_ep.mxn, data
## = regresion.arboles.mn)
## weights: W_dist1000.inv
##
## Moran I statistic standard deviate = 4.1688, p-value = 3.063e-05
## alternative hypothesis: two.sided
## sample estimates:
## Observed Moran I      Expectation      Variance
## 0.1157938766 -0.0070709189  0.0008686406
```

Para ambos casos el valor de Moran Global es mayor que 0 y significativo, aunque no es muy alto, si existe una tendencia en los datos y ambas matrices capturan el efecto.  $W_d$  lo hace mejor en los residuos, mientras  $W_q$  lo hace mejor con la variable dependiente. En las figuras 4.88 y 4.89 observamos que los grupos de sectores son más pequeños que los de las variables dependientes, lo que podría deberse a que las variables independientes presentan un patrón espacial similar, y que por consiguiente, al introducir estos retardos al modelo van a ayudar a absorver esa diferencias para mejorar la estimación.

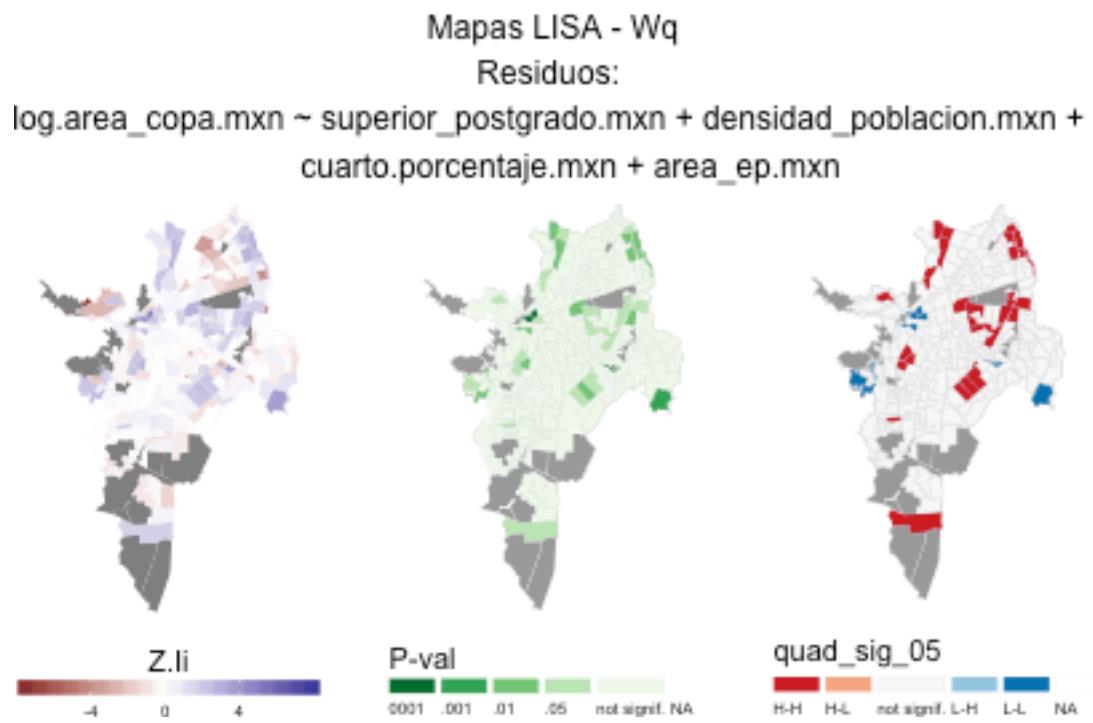


Figure 4.88: Mapas LISA para la matriz  $W_q$  de los residuos del modelo lineal para del área de copa

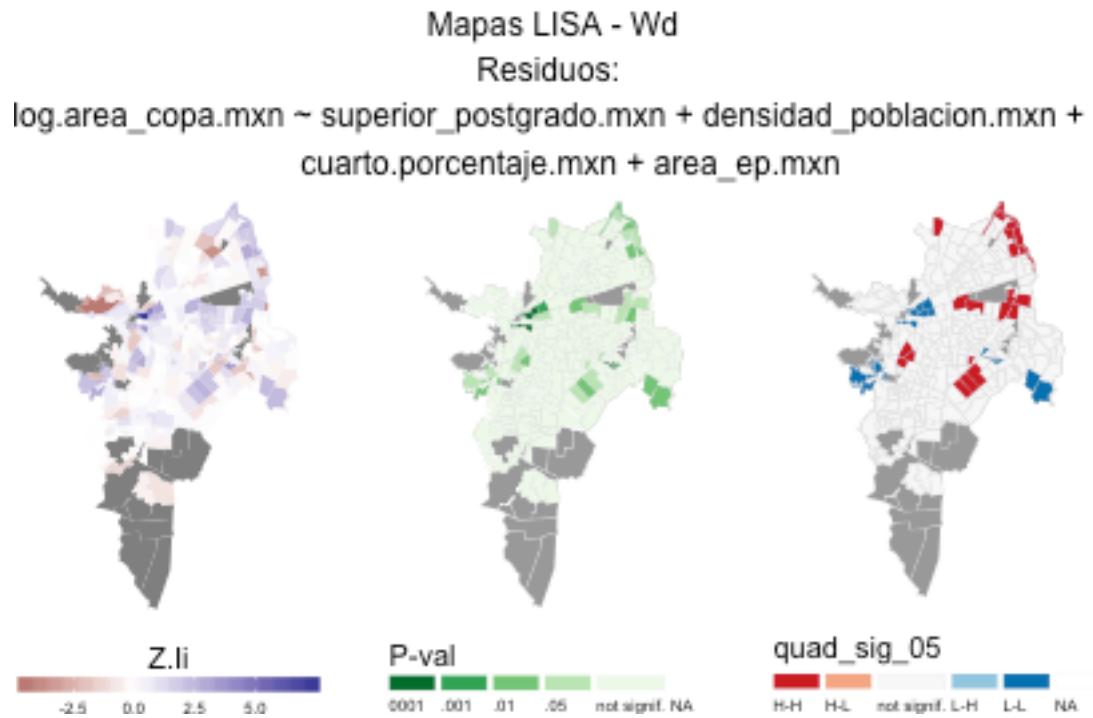


Figure 4.89: Mapas LISA para la matriz  $W_d$  dede los residuos del modelo lineal para del área de copa

El mismo análisis se aplica al modelo de porcentaje de cobertura de copa (`cobertura_copa.ap.mxn ~ superior_postgrado.porcentaje.mxn`).

```
lm.morantest(lm.best.cobertura.ap, W_queen, alternative = "two.sided", zero.policy = T)

## Global Moran I for regression residuals
## data:
## model: lm(formula = cobertura_copa.ap.mxn ~
## superior_postgrado.porcentaje.mxn, data = regresion.arboles.mn)
## weights: W_queen
##
## Moran I statistic standard deviate = 5.8191, p-value = 5.915e-09
## alternative hypothesis: two.sided
## sample estimates:
## Observed Moran I      Expectation      Variance
##          0.194361167     -0.005463538     0.001179183

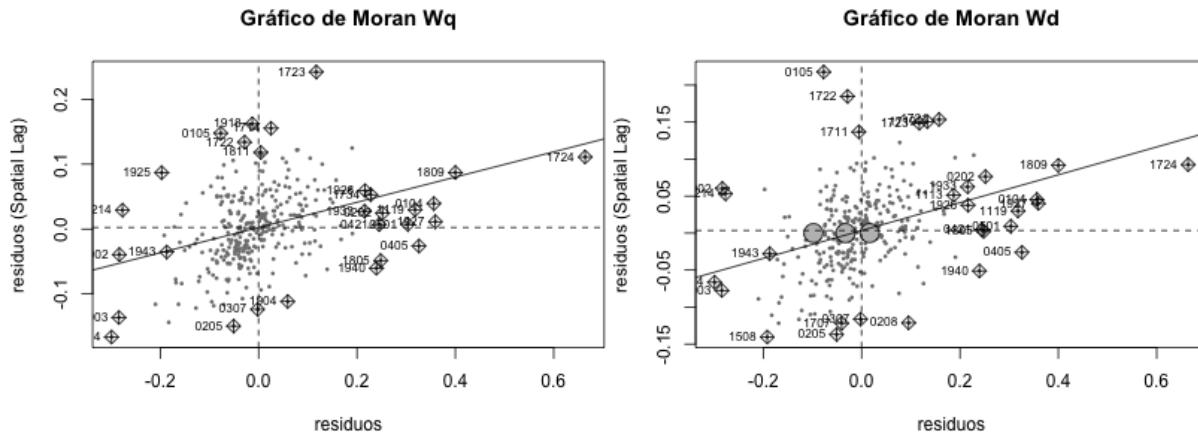
lm.morantest(lm.best.cobertura.ap, W_dist1000.inv, alternative = "two.sided",
zero.policy = T)

## Global Moran I for regression residuals
## data:
## model: lm(formula = cobertura_copa.ap.mxn ~
## superior_postgrado.porcentaje.mxn, data = regresion.arboles.mn)
## weights: W_dist1000.inv
##
## Moran I statistic standard deviate = 6.5028, p-value = 7.885e-11
```

```

## alternative hypothesis: two.sided
## sample estimates:
## Observed Moran I      Expectation      Variance
##      0.1882096881     -0.0053153442    0.0008856801

```



\begin{figure}

\caption{Gráfico de Moran para los residuos del modelo lineal para el porcentaje de cobertura de copa usando  $W_q$  y  $W_d$ } \end{figure}

Para ambos matrices el valor de Moran Global es mayor que 0 y significativo, casi el doble más alto que para el modelo de área de copa; el indicador porcentual de cobertura y el residuo del modelo para ajustarlo muestra el mismo trocamiento en la idoneidad de las matrices: el efecto es más fuerte con  $W_d$  para la variable dependiente y levemente mayor con  $W_q$  en los residuos. En las figuras 4.90 y 4.91 una vez más observamos que los grupos de sectores son más pequeños que los de las variables dependientes, lo que, como dijimos, se debe a que las variables independientes presentan un patrón espacial similar, y se espera que al introducir estos efectos espaciales al modelo mejora la confiabilidad de la estimación.

Los resultados nos invitan a proceder a realizar un ajuste en ambos modelos introduciendo algún tipo de estructura espacial. Posiblemente la matriz  $W_q$  sea la candidata para el área de copa si se piensa en un modelo autorregresivo, pues parece capturar mejor en los test de autocorrelación global la existencia de grupos formados por fenómenos de dispersión o derrame en la distribución de la copa. En el caso de usar modelos de error espacial SEM

Mapas LISA -  $W_q$   
 Residuos:  
 $\text{cobertura\_copa.ap.mxn} \sim \text{superior\_postgrado.porcentaje.mxn}$

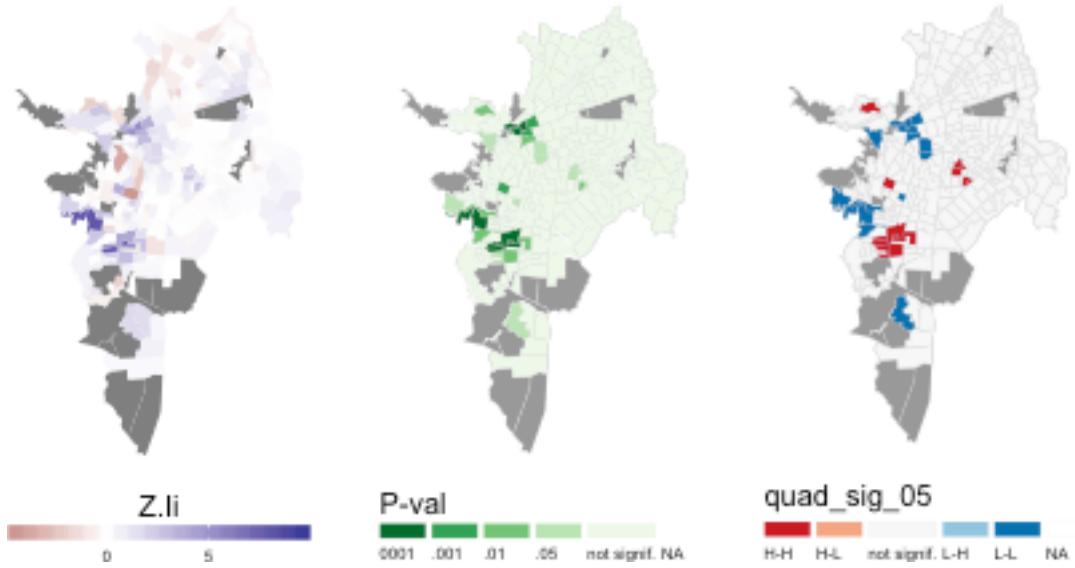


Figure 4.90: Mapas LISA para la matriz  $W_q$  de los residuos del modelo lineal para el porcentaje de cobertura copa

Mapas LISA -  $W_d$   
 Residuos:  
 $\text{cobertura\_copa.ap.mxn} \sim \text{superior\_postgrado.porcentaje.mxn}$

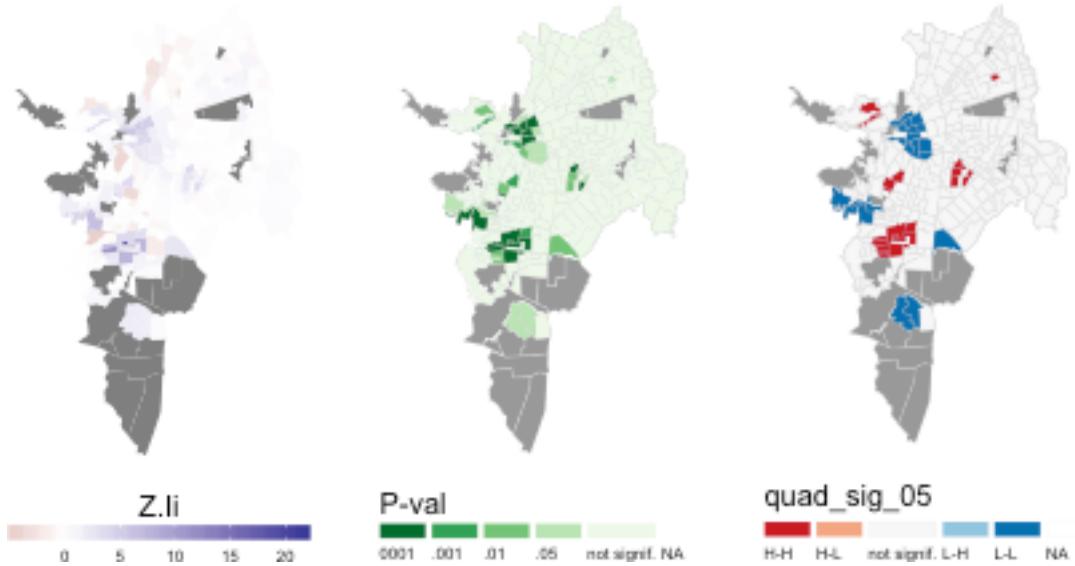


Figure 4.91: Mapas LISA para la matriz  $W_d$  dede los residuos del modelo lineal para del área de copa

posiblemente la candidata sea  $W_d$  por capturar mejor la autocorrelación en los residuos. En busca de simplificar el análisis y concentrarnos entre los modelos usaremos  $W_q$  para el área de copa; para el porcentaje de cobertura de copa la matriz que usaremos  $W_d$  usando el mismo criterio.

#### 4.5.2.2 Ajuste de modelos espaciales

Mejorar la especificación de los modelos lineales incluyendo términos de retardo espacial en la variable dependiente (SAR (2.8)) se hace para obtener una adecuada estimación de los coeficientes de las otras covariables en el modelo. Si optamos por un modelo de error espacial (SEM (2.9)) implica que no es necesario plantear efectos distintivos de la variable dependiente rezagada, y que es posible que ese efecto sea por otras variables no tenidas en cuenta: el agrupamiento espacial observado en la variable dependiente se explica simplemente por el patrón geográfico de variables independientes medidas y no medidas. El modelo SAR, en cambio, incorpora la influencia de variables independientes no medidas, pero también estipula un efecto adicional de valores de atributos vecinos, es decir, la variable dependiente rezagada. Si incluimos el retardo sólo de las variables independientes (SLX (2.10)) esperamos que los cambios en las dimensiones expresadas con las predictores produce un efecto de derrame o influencia en los sectores vecinos.

¿Qué significa decir que la cantidad de cobertura de copa está relacionada con la de los sectores vecinos? ¿Son los procesos de reproducción del arbolado urbano un fenómeno independiente de las intervenciones de sus habitantes y de los urbanizadores? ¿Los habitantes que ven árboles en las cuadras o barrios aledaños deciden sembrar árboles en su vecindario? ¿Existen similitudes en las condiciones estructurales de los barrios en ciertas zonas de la ciudad que prefieren las personas con mejores condiciones sociales (tener estudios superiores p.e.)? ¿Qué tipo de pérdidas en la cobertura de copa están motivadas por la densificación de un sector? ¿Cómo afectan los cambios en los tipos de oferta habitacional en un sector la cobertura de copa de los sectores vecinos? (las viviendas tipo cuarto suelen ofrecerse en pensiones y ser más económicas que las casas o apartamentos).

La pregunta a hacerse es cómo saber cuál de los diferentes modelos espaciales es el que mejor representa el fenómeno que estamos modelando y si los datos respaldan nuestras convicciones teóricas. Si el modelo de retraso espacial que especifique es realmente el correcto, entonces ninguna dependencia espacial debe permanecer en los residuos, y podremos elaborar sobre el tipo de procesos que pueden verse representados.

Una alternativa metodológica es probar los 4 tipos de modelos con la matriz  $W_d$  que resultó capturar mejor la asociación espacial en los datos y comparar sus resultados.

#### 4.5.2.2.1 Área de copa

```
# SAR

sar.mod.log.area_copa.wq <- lagsarlm(formula = as.formula(lm.mod.area_copa) ,
  data = lm.mod.area_copa$model, listw = W_queen, zero.policy = T, tol.solve = 1e-30)
sar.sm <- summary(sar.mod.log.area_copa.wq, Nagelkerke = T)
sar.sm

##
## Call:
## lagsarlm(formula = as.formula(lm.mod.area_copa), data = lm.mod.area_copa$model,
##   listw = W_queen, zero.policy = T, tol.solve = 1e-30)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.2529145 -0.0308535  0.0040657  0.0415388  0.1300425
##
## Type: lag
## Coefficients: (asymptotic standard errors)
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)           0.614070  0.048506 12.6598 < 2.2e-16
```

```

## superior_postgrado.mxn  0.288636  0.022492 12.8327 < 2.2e-16
## densidad_poblacion.mxn -0.119617  0.017950 -6.6640 2.665e-11
## cuarto.porcentaje.mxn  -0.116631  0.030027 -3.8842 0.0001027
## area_ep.mxn             0.118796  0.025882  4.5899 4.435e-06
##
## Rho: 0.20055, LR test value: 11.903, p-value: 0.00056042
## Asymptotic standard error: 0.058641
##      z-value: 3.4199, p-value: 0.00062646
## Wald statistic: 11.696, p-value: 0.00062646
##
## Log likelihood: 464.4993 for lag model
## ML residual variance (sigma squared): 0.0033034, (sigma: 0.057476)
## Nagelkerke pseudo-R-squared: 0.62529
## Number of observations: 324
## Number of parameters estimated: 7
## AIC: -915, (AIC for lm: -905.1)
## LM test for residual autocorrelation
## test value: 0.18965, p-value: 0.66321

# SEM

sem.mod.log.area_copa.wq <- errorsarlm(formula = as.formula(lm.mod.area_copa),
                                           data = lm.mod.area_copa$model, listw = W_queen, zero.policy = T, tol.solve = 1e-30)
sem.sm <- summary(sem.mod.log.area_copa.wq, Nagelkerke = T)
sem.sm

##
## Call:
## errorsarlm(formula = as.formula(lm.mod.area_copa), data = lm.mod.area_copa$model,
##             listw = W_queen, zero.policy = T, tol.solve = 1e-30)
##
```

```

## Residuals:
##      Min       1Q    Median     3Q    Max
## -0.244147 -0.027587  0.003208  0.040352  0.125335
##
## Type: error
## Coefficients: (asymptotic standard errors)
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)          0.775902  0.009157 84.7331 < 2.2e-16
## superior_postgrado.mxn 0.303996  0.022762 13.3556 < 2.2e-16
## densidad_poblacion.mxn -0.138182  0.018723 -7.3804 1.579e-13
## cuarto.porcentaje.mxn -0.132682  0.032233 -4.1163 3.850e-05
## area_ep.mxn           0.130410  0.028109  4.6395 3.493e-06
##
## Lambda: 0.21687, LR test value: 6.9218, p-value: 0.0085151
## Asymptotic standard error: 0.080758
## z-value: 2.6855, p-value: 0.0072429
## Wald statistic: 7.2117, p-value: 0.0072429
##
## Log likelihood: 462.0087 for error model
## ML residual variance (sigma squared): 0.0033501, (sigma: 0.05788)
## Nagelkerke pseudo-R-squared: 0.61948
## Number of observations: 324
## Number of parameters estimated: 7
## AIC: -910.02, (AIC for lm: -905.1)

# SD

sd.mod.log.area_copa.wq <- lagsarlm(formula = as.formula(lm.mod.area_copa),
  data = lm.mod.area_copa$model, listw = W_queen, zero.policy = T, tol.solve = 1e-30,
  type = "mixed")

```

```

sd.sm <- summary(sd.mod.log.area_copa.wq, Nagelkerke = T)
sd.sm

##
## Call:
## lagsarlm(formula = as.formula(lm.mod.area_copa), data = lm.mod.area_copa$model,
##   listw = W_queen, type = "mixed", zero.policy = T, tol.solve = 1e-30)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.2295259 -0.0323844  0.0047241  0.0421406  0.1330323
##
## Type: mixed
## Coefficients: (asymptotic standard errors)
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)          0.636994  0.064814  9.8281 < 2.2e-16
## superior_postgrado.mxn 0.287135  0.023882 12.0232 < 2.2e-16
## densidad_poblacion.mxn -0.095813  0.025330 -3.7826 0.0001552
## cuarto.porcentaje.mxn -0.059241  0.041204 -1.4377 0.1505058
## area_ep.mxn           0.141098  0.030091  4.6890 2.745e-06
## lag.superior_postgrado.mxn -0.015543  0.046962 -0.3310 0.7406618
## lag.densidad_poblacion.mxn -0.049618  0.036582 -1.3563 0.1749910
## lag.cuarto.porcentaje.mxn -0.145767  0.065484 -2.2260 0.0260154
## lag.area_ep.mxn         -0.074189  0.040493 -1.8321 0.0669325
##
## Rho: 0.20067, LR test value: 6.4147, p-value: 0.011318
## Asymptotic standard error: 0.080724
##      z-value: 2.4859, p-value: 0.012922
## Wald statistic: 6.1798, p-value: 0.012922

```

```

## 
## Log likelihood: 468.5857 for mixed model
## ML residual variance (sigma squared): 0.0032211, (sigma: 0.056755)
## Nagelkerke pseudo-R-squared: 0.63462
## Number of observations: 324
## Number of parameters estimated: 11
## AIC: -915.17, (AIC for lm: -910.76)
## LM test for residual autocorrelation
## test value: 0.079401, p-value: 0.77811

# SLX

slx.mod.log.area_copa.wq <- lmSLX(formula = as.formula(lm.mod.area_copa), data = lm.mod.

  listw = W_queen, zero.policy = T)

slx.sm <- summary(slx.mod.log.area_copa.wq)
slx.sm

## 
## Call:
## lm(formula = y ~ x - 1, weights = weights)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.208083 -0.032453  0.003926  0.043776  0.139282
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.79551   0.01364  58.341 < 2e-16 ***
## xsuperior_postgrado.mxn 0.28719   0.02455  11.699 < 2e-16 ***
## xdensidad_poblacion.mxn -0.09948   0.02603 -3.821  0.00016 ***

```

Table 4.16: Metricas de ajuste para los modelos de área de copa

medidasfit	OLS	SLX	SAR	SEM	SD
Globla Moran'I	0.09660	0.10143	-0.01056	-0.01056	-0.00157
GMI p-value	0.00192	0.00122	0.58578	0.58575	0.48239
Shapiro-Wilk	0.97939	0.98495	0.97959	0.97638	0.98307
SW p-value	0.00013	0.00183	0.00015	0.00004	0.00073
Breusch-Pagan	6.70653	11.43612	4.36429	5.08935	10.28094
BP p-value	0.15223	0.17819	0.35894	0.27825	0.24586
Media Residuos	0.00000	0.00000	0.00000	0.00000	0.00000
MSE	0.00345	0.00331	0.00330	0.00335	0.00322
adj-Rsquare	0.60639	0.99453	NA	NA	NA
Nagelkerke pseudo-R-squared	NA	NA	0.62529	0.61948	0.63462
AIC	-905.09560	-910.75682	-914.99865	-910.01741	-915.17149
Log likelihood	458.54780	465.37841	464.49933	462.00870	468.58574

```

## xcuarto.porcentaje.mxn      -0.06373   0.04234  -1.505  0.13326
## xarea_ep.mxn                 0.13788   0.03092   4.459  1.14e-05 ***
## xlag.superior_postgrado.mxn  0.05195   0.03998   1.299  0.19472
## xlag.densidad_poblacion.mxn -0.07698   0.03529  -2.181  0.02990 *
## xlag.cuarto.porcentaje.mxn  -0.17821   0.06576  -2.710  0.00710 **
## xlag.area_ep.mxn              -0.04655   0.04051  -1.149  0.25139
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05836 on 315 degrees of freedom
## Multiple R-squared:  0.9947, Adjusted R-squared:  0.9945
## F-statistic:  6543 on 9 and 315 DF,  p-value: < 2.2e-16

AIC(slx.mod.log.area_copa.wq)

## [1] -910.7568

```

Al comparar los resultados de las métricas de ajuste se identifica al modelo SD o mixto(variables con retardo y autocorrelación) con el mejor AIC. El modelo SD ha

logrado eliminar la autocorrelación espacial global en los residuos, al igual que SAR y SEM (ver figura 4.92). Aunque persiste la no normalidad de los residuos y la heterocedasticidad como los muestran los test y las gráficas diagnósticas, el error cometido disminuye y los coeficientes pueden considerarse más confiables. El  $\rho$  y  $\lambda$  de las estimaciones con términos de autoregresivos es significativo, lo que implica que la inclusión de los retardos funciona correctamente, aunque solo en los coeficientes de la variable retardada `cuarto.porcentaje` es significativo, aunque esta no tiene un coeficiente significativo como variable sin retardo, sugiriendo que interesa su efecto sobre el área de copa en los sectores vecinos a los focos de mayor concentración de viviendas tipo cuarto, pero no es importante en el resto de sectores. Como era de esperarse el porcentaje de área de espacios verdes es significativo en las variables sin retardo, pero no genera ningún efecto de derrame en la cobertura de copa de los sectores vecinos. A pesar de que los test de normalidad y heterocedasticidad no son existosos, las gráficas diagnósticas muestran que los problemas se presentan en los valores extremos.

#### 4.5.2.2.2 Porcentaje de cobertura de área de copa

El mismo ejercicio se aplica al porcentaje de cobertura de copa.

```
# SAR

sar.cobertura.ap.wq <- lagsarlm(formula = as.formula(lm.best.cobertura.ap),
  data = lm.best.cobertura.ap$model, listw = W_queen, zero.policy = T, tol.solve = 1e-30)
sar.sm <- summary(sar.cobertura.ap.wq, Nagelkerke = T)
sar.sm

##
## Call:
## lagsarlm(formula = as.formula(lm.best.cobertura.ap), data = lm.best.cobertura.ap$model,
##   listw = W_queen, zero.policy = T, tol.solve = 1e-30)
##
```

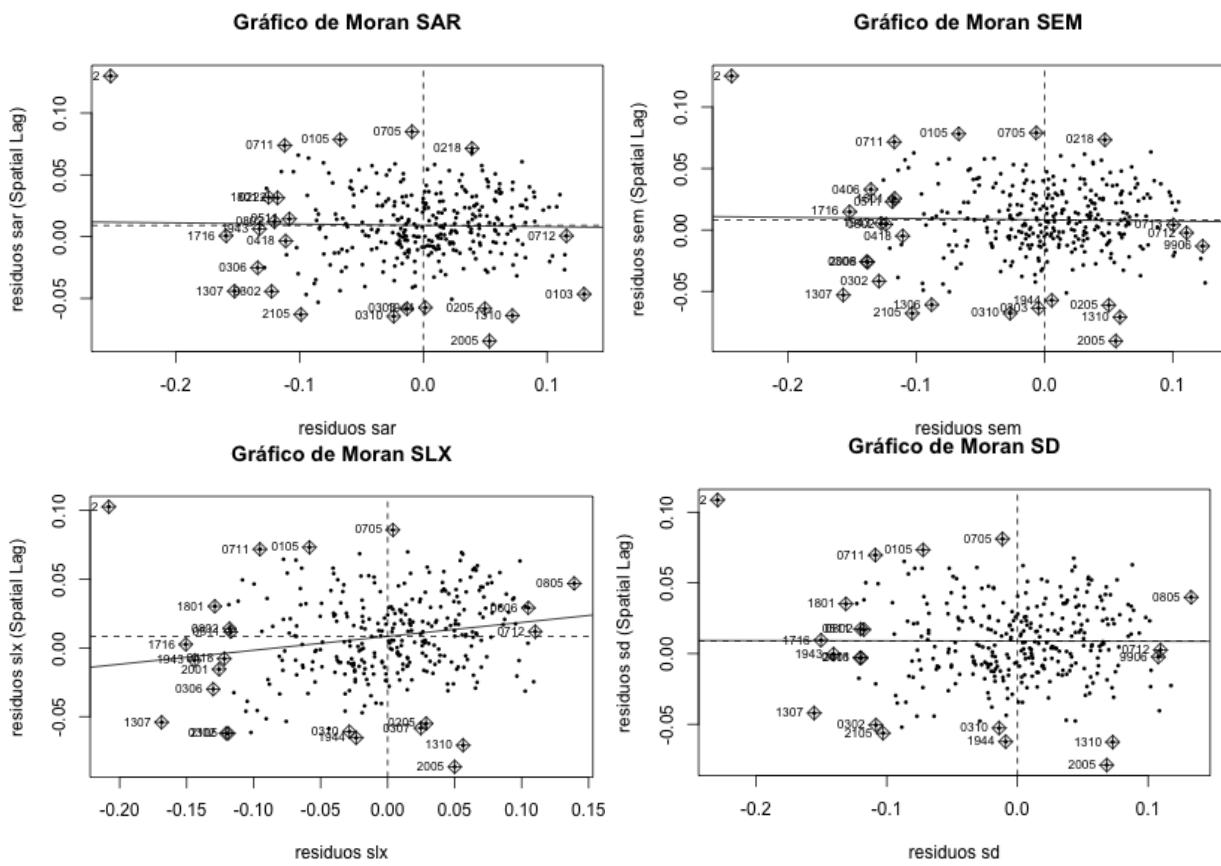


Figure 4.92: Gráfico de Moran para los residuos de los modelos espaciales de área de copa usando  $W_q$

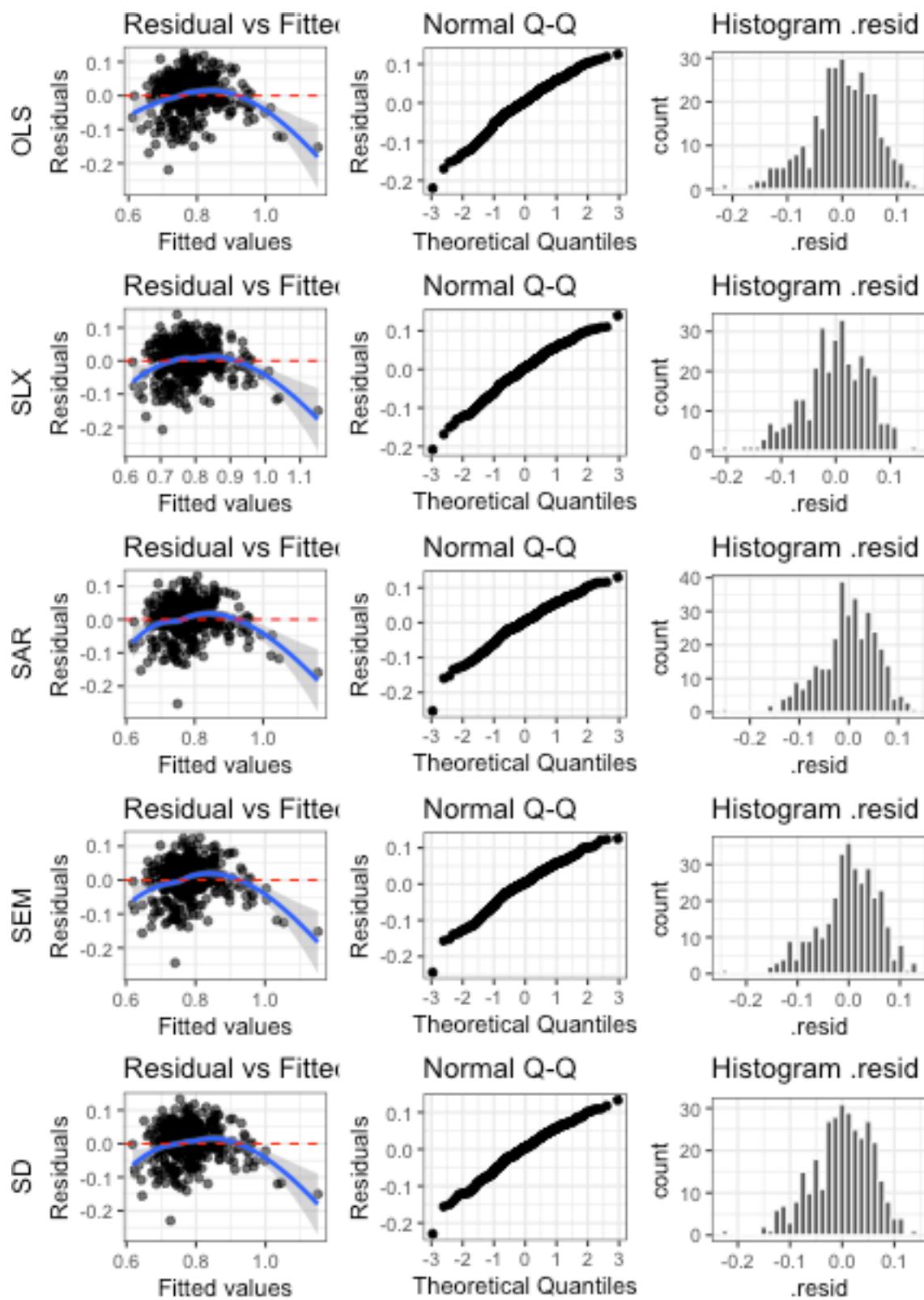


Figure 4.93: Diagnóstico comparativo entre modelos

```

## Residuals:
##      Min       1Q    Median       3Q      Max
## -0.270499 -0.049562 -0.010666  0.035328  0.624252
##
## Type: lag
## Coefficients: (asymptotic standard errors)
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 0.030066  0.010654  2.8220 0.004773
## superior_postgrado.porcentaje.mxn 0.246343  0.027300  9.0234 < 2.2e-16
##
## Rho: 0.37521, LR test value: 28.579, p-value: 8.9958e-08
## Asymptotic standard error: 0.064408
##      z-value: 5.8255, p-value: 5.6938e-09
## Wald statistic: 33.937, p-value: 5.6938e-09
##
## Log likelihood: 285.1216 for lag model
## ML residual variance (sigma squared): 0.0097906, (sigma: 0.098947)
## Nagelkerke pseudo-R-squared: 0.50267
## Number of observations: 324
## Number of parameters estimated: 4
## AIC: -562.24, (AIC for lm: -535.66)
## LM test for residual autocorrelation
## test value: 0.19089, p-value: 0.66218

# SEM

sem.cobertura.ap.wq <- errorsarlm(formula = as.formula(lm.best.cobertura.ap),
  data = lm.best.cobertura.ap$model, listw = W_queen, zero.policy = T, tol.solve = 1
sem.sm <- summary(sem.cobertura.ap.wq, Nagelkerke = T)
sem.sm

```

```

## Call:
## errorsarlm(formula = as.formula(lm.best.cobertura.ap), data = lm.best.cobertura.ap$mo
##   listw = W_queen, zero.policy = T, tol.solve = 1e-30)
##
## Residuals:
##       Min        1Q      Median        3Q       Max
## -0.281664 -0.052052 -0.012380  0.034632  0.625256
##
## Type: error
## Coefficients: (asymptotic standard errors)
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)             0.070011  0.013596  5.1495 2.612e-07
## superior_postgrado.porcentaje.mxn 0.334060  0.028075 11.8989 < 2.2e-16
##
## Lambda: 0.39903, LR test value: 26.77, p-value: 2.2916e-07
## Asymptotic standard error: 0.071019
##      z-value: 5.6187, p-value: 1.9242e-08
## Wald statistic: 31.57, p-value: 1.9242e-08
##
## Log likelihood: 284.2172 for error model
## ML residual variance (sigma squared): 0.0098055, (sigma: 0.099023)
## Nagelkerke pseudo-R-squared: 0.49988
## Number of observations: 324
## Number of parameters estimated: 4
## AIC: -560.43, (AIC for lm: -535.66)

# SD

sd.cobertura.ap.wq <- lagsarlm(formula = as.formula(lm.best.cobertura.ap), data = lm.be

```

```

listw = W_queen, zero.policy = T, tol.solve = 1e-30, type = "mixed")
sd.sm <- summary(sd.cobertura.ap.wq, Nagelkerke = T)
sd.sm

##
## Call:
## lagsarlm(formula = as.formula(lm.best.cobertura.ap), data = lm.best.cobertura.ap$mo
##   listw = W_queen, type = "mixed", zero.policy = T, tol.solve = 1e-30)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.275705 -0.051349 -0.011907  0.033539  0.621785
##
## Type: mixed
## Coefficients: (asymptotic standard errors)
##                               Estimate Std. Error z value
## (Intercept)          0.032144  0.010945  2.9370
## superior_postgrado.porcentaje.mxn  0.278081  0.044417  6.2607
## lag.superior_postgrado.porcentaje.mxn -0.051197  0.058161 -0.8803
##                               Pr(>|z|)
## (Intercept)          0.003314
## superior_postgrado.porcentaje.mxn  3.833e-10
## lag.superior_postgrado.porcentaje.mxn  0.378714
##
## Rho: 0.40097, LR test value: 27.538, p-value: 1.5403e-07
## Asymptotic standard error: 0.070765
##   z-value: 5.6663, p-value: 1.4592e-08
## Wald statistic: 32.107, p-value: 1.4592e-08
##

```

```

## Log likelihood: 285.5239 for mixed model
## ML residual variance (sigma squared): 0.0097234, (sigma: 0.098607)
## Nagelkerke pseudo-R-squared: 0.5039
## Number of observations: 324
## Number of parameters estimated: 5
## AIC: -561.05, (AIC for lm: -535.51)
## LM test for residual autocorrelation
## test value: 1.633, p-value: 0.20129

# SLX

slx.cobertura.ap.wq <- lmSLX(formula = as.formula(lm.best.cobertura.ap), data = lm.best.

  listw = W_queen, zero.policy = T)

slx.sm <- summary(slx.cobertura.ap.wq)
slx.sm

## Call:
## lm(formula = y ~ x - 1, weights = weights)
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29497 -0.05404 -0.01494  0.04472  0.66363
## Coefficients:
##                               Estimate Std. Error t value
## x(Intercept)                  0.05912   0.01080   5.472
## xsuperior_postgrado.porcentaje.mxn  0.29297   0.04733   6.190
## xlag.superior_postgrado.porcentaje.mxn  0.07542   0.05570   1.354
## Pr(>|t|)
```

Table 4.17: Metricas de ajuste para los modelos de porcentaje de área de copa

medidasfit	OLS	SLX	SAR	SEM	SD
Globla Moran'I	0.19436	0.19945	0.00734	-0.01131	-0.00863
GMI p-value	0.00000	0.00000	0.37996	0.59509	0.56441
Shapiro-Wilk	0.90492	0.90851	0.89936	0.89238	0.89685
SW p-value	0.00000	0.00000	0.00000	0.00000	0.00000
Breusch-Pagan	16.82734	16.80794	16.22458	15.21537	17.19946
BP p-value	0.00004	0.00022	0.00006	0.00010	0.00018
Media Residuos	0.00000	0.00000	0.00000	0.00000	0.00000
MSE	0.01100	0.01094	0.00979	0.00981	0.00972
adj-Rsquare	0.45512	0.80624	NA	NA	NA
Nagelkerke pseudo-R-squared	NA	NA	0.50267	0.49988	0.50390
AIC	-535.66435	-535.50972	-562.24324	-560.43439	-561.04776
Log likelihood	270.83218	271.75486	285.12162	284.21719	285.52388

```

## x(Intercept)           8.99e-08 ***
## xsuperior_postgrado.porcentaje.mxn    1.84e-09 ***
## xlag.superior_postgrado.porcentaje.mxn   0.177
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1051 on 321 degrees of freedom
## Multiple R-squared:  0.808,  Adjusted R-squared:  0.8062
## F-statistic: 450.4 on 3 and 321 DF,  p-value: < 2.2e-16

AIC(slx.cobertura.ap.wq)

## [1] -535.5097

```

Al comparar los resultados de las métricas de ajuste se identifica al modelo SAR con el mejor rendimiento en el ajuste del modelo. El coeficiente de correlación  $\rho$  es alto y muy significativo, lo que nos dice que es una mejora la inclusión de las características espaciales en los datos, y se reflejados en el índice de Akaike que tiene una mejora visible (de  $-535.6643522$  a  $-562.2432413$ ). Consistentemente la variable de **estudios superiores**

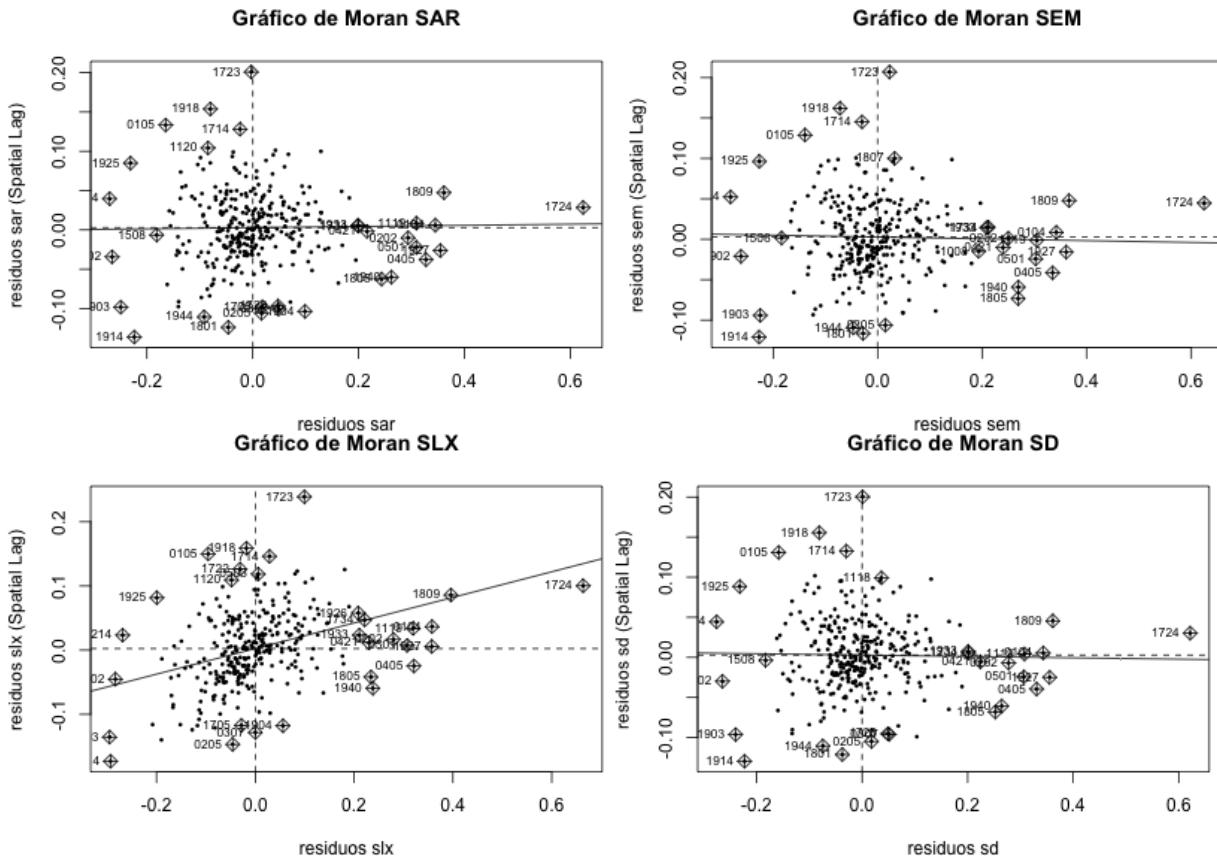


Figure 4.94: Gráfico de Moran para los residuos de los modelos espaciales del porcentaje de área de copa usando  $W_q$

en la población refleja el patrón de agrupamiento de la cobertura de copa teniendo la mayor importancia para la estimación; pero es poco significativa como variable retardada, y es por esto que los modelos de error espacial SEM y el mixto (SD) no son mejores que el autorregresivo. Todos los modelos espaciales logran reducir la autocorrelación en los residuos (ver figura 4.94) lo que hace más confiables los coeficientes estimados, pero en el modelo SAR se mantiene una importancia en el efecto de la densidad de población como determinante para la reducción de la cobertura de copa.

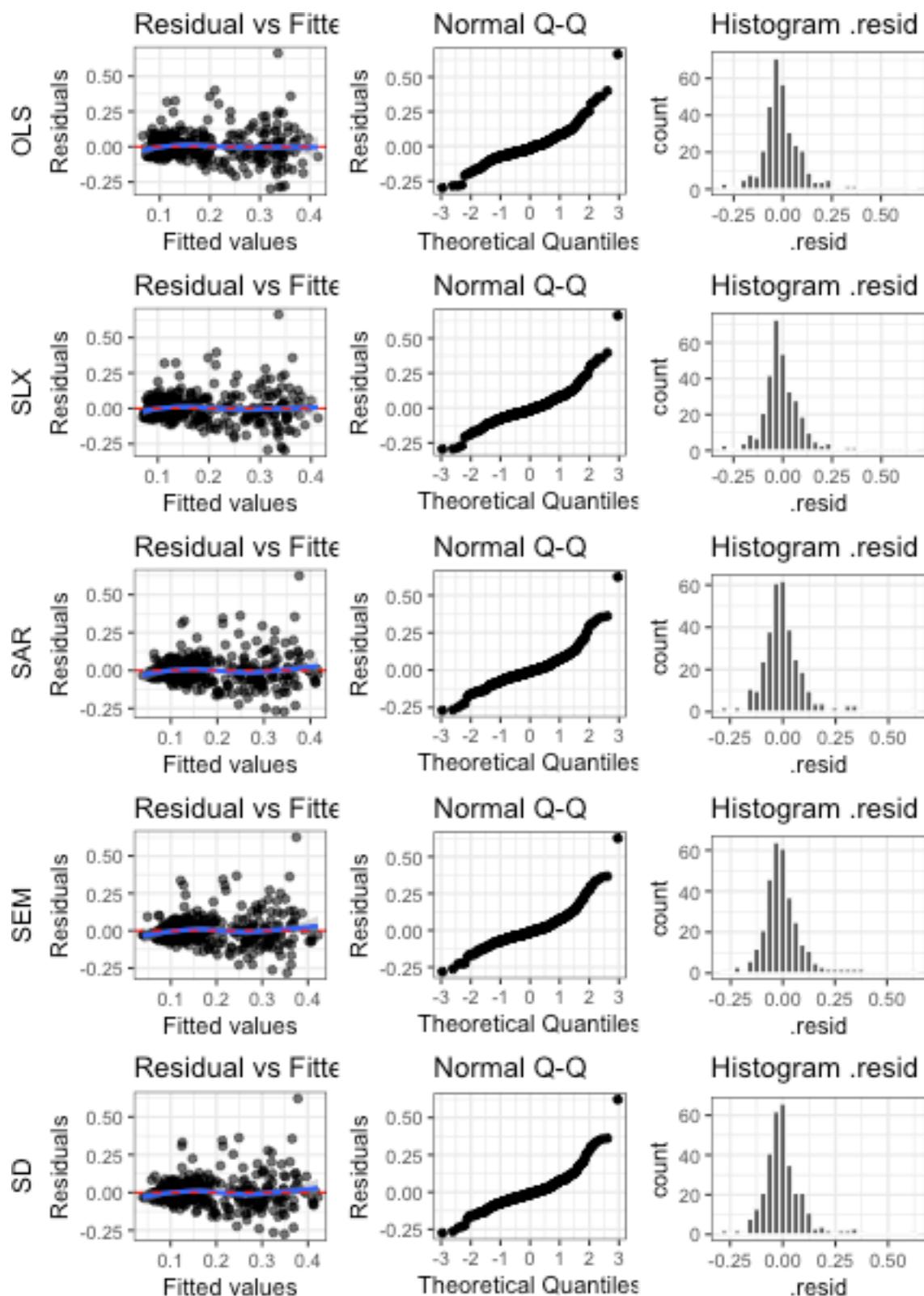


Figure 4.95: Diagnóstico comparativo entre modelos de porcentaje de copa

### 4.5.3 Modelado espacial de espacios verdes

El proceso de ajuste de los modelos geoestadísticos para el análisis de espacios verdes hace uso de los mismos elementos metodológicos usados para la cobertura de copa. Se construyen dos matrices de velocidad usando un kernel de vecindad Queen  $W_q$  y otro con base en un radio de búsqueda de 1 kilómetro  $W_d$ . El primer paso es evaluar cuál de las matrices captura mejor la autocorrelación de los residuos de los modelos lineales y de las variables dependientes. Seguidamente se compara los diferentes modelos espaciales para seleccionar el de mejor ajuste y finalmente se evalúa la significancia de las variables y el valor de los coeficientes de la regresión y la mejora en el ajuste con relación a los modelos lineales.

#### 4.5.3.1 Matrices de vecindad

Las dos matrices de vecindad construidas con los SU seleccionados para el análisis de regresión lineal. Las dos matrices resultantes se muestran en la figura ??

#### 4.5.3.2 Autocorrelación espacial

##### 4.5.3.2.1 Variables dependientes

Comparemos el efecto de cada una de las matrices sobre el indicador `area_ep.porcentaje`

```
moran.test(regresion.EV$area_ep.porcentaje, listw = W_dist1000.inv.ev, zero.policy = TRUE)

##
## Moran I test under randomisation
##
## data: regresion.EV$area_ep.porcentaje
## weights: W_dist1000.inv.ev
##
```

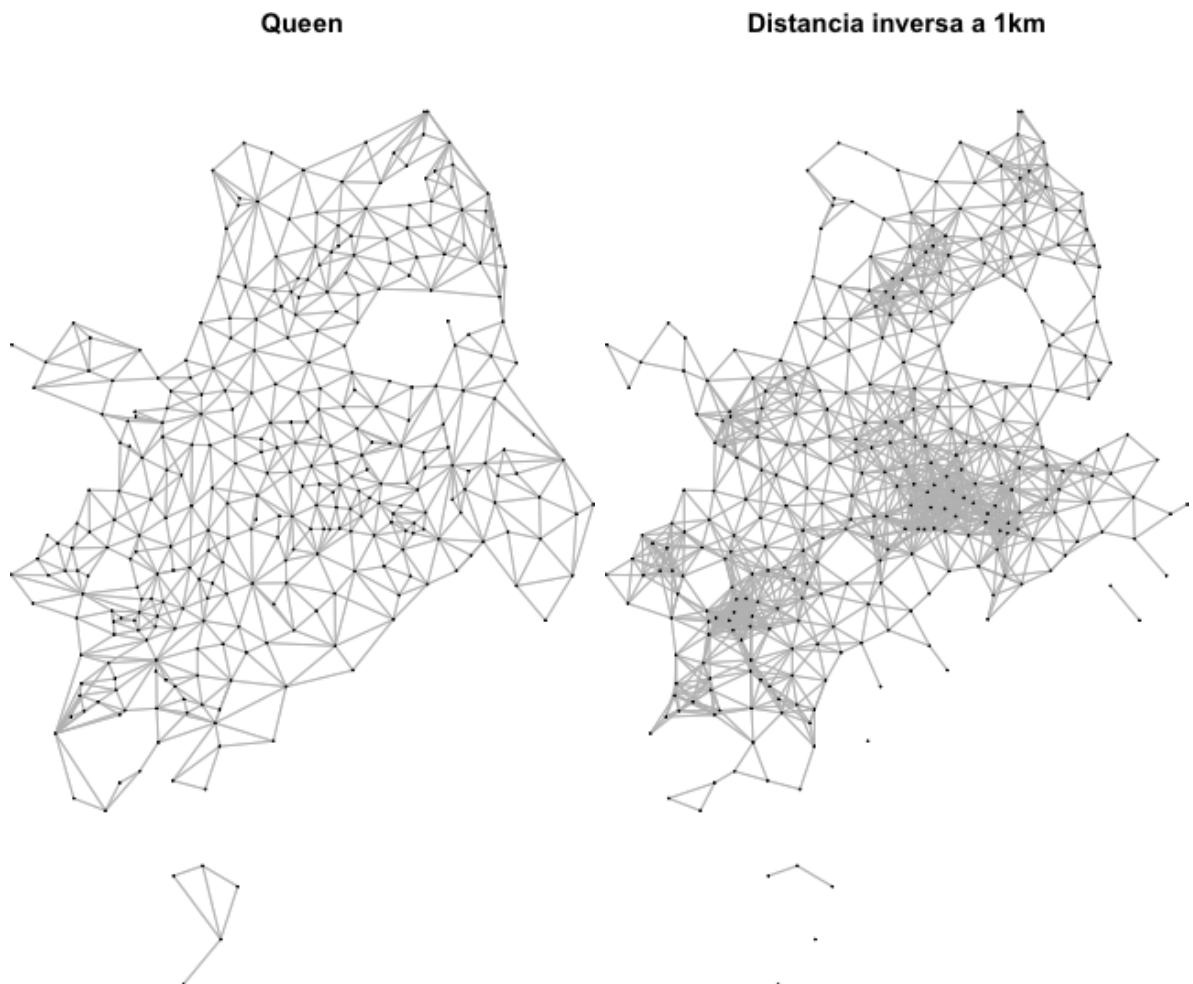


Figure 4.96: Matrices de vecindad del análisis espacial de espacios verdes

```

## Moran I statistic standard deviate = 2.0602, p-value = 0.01969
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##          0.0537662593    -0.0030769231     0.0007612826

moran.test(regresion.EV$area_ep.porcentaje, listw = W_queen.ev)

##
## Moran I test under randomisation
##
## data: regresion.EV$area_ep.porcentaje
## weights: W_queen.ev
##
## Moran I statistic standard deviate = 3.3938, p-value = 0.0003447
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##          0.104622145    -0.003048780     0.001006547

moran.test(regresion.EV$ia.areas.dist, listw = W_dist1000.inv.ev, zero.policy = TRUE)

##
## Moran I test under randomisation
##
## data: regresion.EV$ia.areas.dist
## weights: W_dist1000.inv.ev
##
## Moran I statistic standard deviate = 24.511, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:

```

```

## Moran I statistic      Expectation      Variance
##          0.677646493     -0.003076923     0.000771291

moran.test(regresion.EV$ia.areas.dist, listw = W_queen.ev)

##
## Moran I test under randomisation
##
## data: regresion.EV$ia.areas.dist
## weights: W_queen.ev
##
## Moran I statistic standard deviate = 24.615, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##          0.782948232     -0.003048780     0.001019626

```

La matriz  $W_q$  captura con mayor fuerza la autocorrelación espacial de `area_ep.porcentaje`. Al repetir el test para `ai.areas.dist` una vez más  $W_q$  captura con mayor intensidad la autocorrelación espacial del indicador, que es mucho mayor que en `area_ep.porcentaje`, posiblemente por representar una característica local del acceso. Hay que anotar aquí que el cálculo del índice `ia.areas.dist` en su construcción usa una distancia de radio de búsqueda de 1 kilómetro; en su definición el indicador está influenciado por sus vecinos por lo que se forman grupos o clusters alrededor de ciertos sectores urbanos. Resulta pues interesante no sea  $W_d$  la que capture mejor el agrupamiento.

Para indagar sobre los patrones espaciales de los dos indicadores usando  $W_q$  se muestran los mapas de LISA en la figura 4.97. Se aprecia que se forman cluster alrededor de tres zonas en el caso del porcentaje de área de espacio verde y dos para el indicador de la relación `areas-distancia`, coincidentes con el anterior. Ahí se encuentran equipamientos de ciudad como un cementerio de gran tamaño, las universidades, zonas conservadas de

riveras de ríos. El grupo que se forma al oriente de la ciudad es donde se encuentra la laguna del Pondaje.

#### 4.5.3.2.2 Residuos de los OLS

Examinemos ahora los residuos de cada modelo lineal seleccionado. Aplicamos el test de Moran primero al modelo de `area_ep.porcentaje` con ambas matrices de vecindad. La figura 4.98 muestra el mismo resultado gráficamente.

```
lm.morantest(lm.area_ep.ptje.sel, W_queen.ev, alternative = "two.sided", zero.policy = T)

## 
## Global Moran I for regression residuals
##
## data:
## model: lm(formula = area_ep.porcentaje.mxn ~ cuarto.porcentaje.mxn
## + area_media_manzana.mxn, data = regresion.EV.mn)
## weights: W_queen.ev
##
## Moran I statistic standard deviate = 3.5892, p-value = 0.0003316
## alternative hypothesis: two.sided
## sample estimates:
## Observed Moran I      Expectation      Variance
##          0.116355499     -0.005017782     0.001143508

lm.morantest(lm.area_ep.ptje.sel, W_dist1000.inv.ev, alternative = "two.sided",
zero.policy = T)

## 
## Global Moran I for regression residuals
##
## data:
```

Mapas LISA - area\_ep.porcentaje - Wq

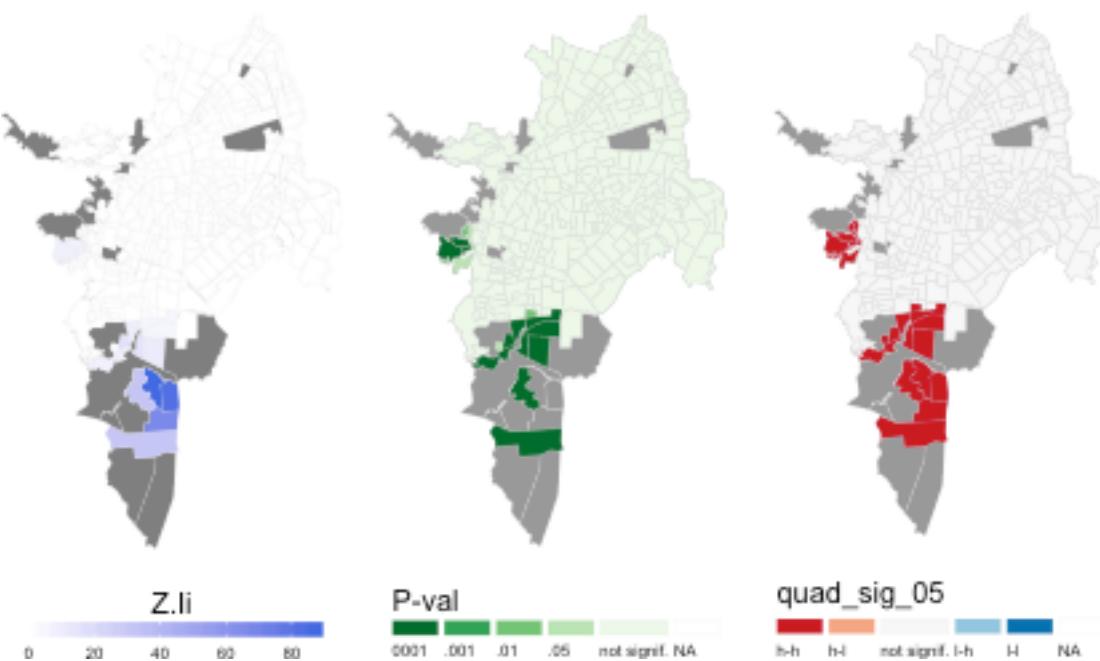
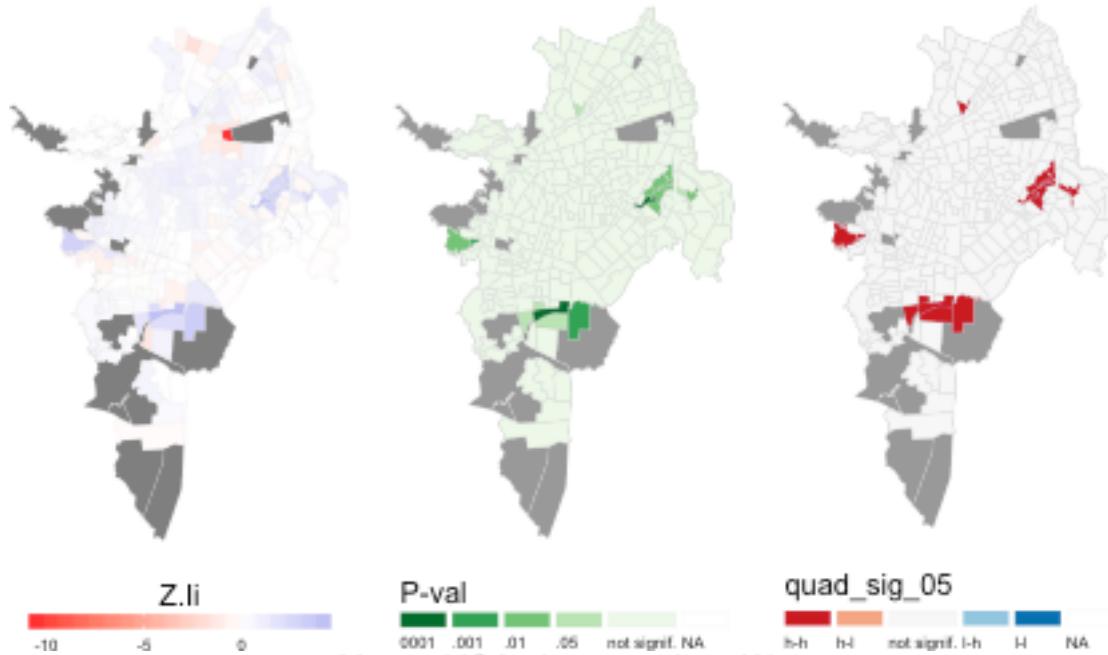


Figure 4.97: Mapas LISA para la matriz  $W_q$  de ambos indicadores de acceso a EV

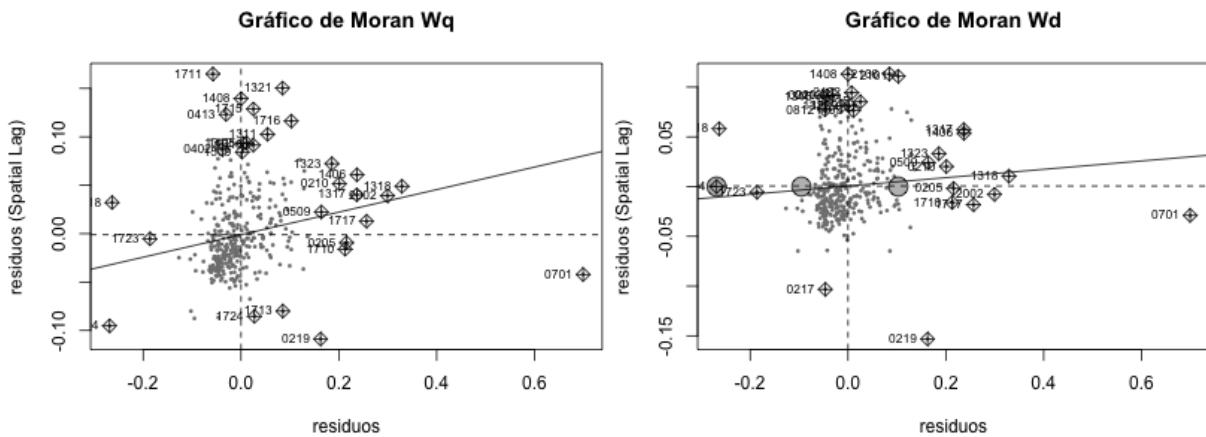


Figure 4.98: Gráfico de Moran para los residuos del modelo lineal para el porcentaje de área de espacio verde

```
## model: lm(formula = area_ep.porcentaje.mxn ~ cuarto.porcentaje.mxn
## + area_media_manzana.mxn, data = regresion.EV.mn)
## weights: W_dist1000.inv.ev
##
## Moran I statistic standard deviate = 1.579, p-value = 0.1143
## alternative hypothesis: two.sided
## sample estimates:
## Observed Moran I      Expectation      Variance
## 0.0420095674 -0.0044075679  0.0008641599
```

Al igual que con la variable dependiente,  $W_q$  captura con mayor intensidad autocorrelación en los residuos. Los resultados de aplicar el test a `ia.areas.dist` confirman que la matriz de vecindad que usaremos para el ajuste es  $W_q$ .

```
lm.morantest(lm.ia.areas.dist.sel, W_queen.ev, alternative = "two.sided", zero.policy =
##
## Global Moran I for regression residuals
##
## data:
## model: lm(formula = form, data = df)
```

```

## weights: W_queen.ev

##
## Moran I statistic standard deviate = 17.802, p-value < 2.2e-16
## alternative hypothesis: two.sided
## sample estimates:

## Observed Moran I      Expectation      Variance
##          0.585347665    -0.010137536    0.001118915

lm.morantest(lm.ia.areas.dist.sel, W_dist1000.inv.ev, alternative = "two.sided",
             zero.policy = T)

## Global Moran I for regression residuals
## data:
## model: lm(formula = form, data = df)
## weights: W_dist1000.inv.ev
## Moran I statistic standard deviate = 19.214, p-value < 2.2e-16
## alternative hypothesis: two.sided
## sample estimates:

## Observed Moran I      Expectation      Variance
##          0.5464778843   -0.0088593824   0.0008353882

```

Los mapas de LISA para los residuos de ambos modelos con la matriz  $W_q$  se muestran a continuación.

#### 4.5.3.3 Ajuste de modelos espaciales

Probamos los 4 tipos de modelos con la matriz  $W_q$  que resultó capturar mejor la asociacion espacial en los datos y comparamos sus resultados.

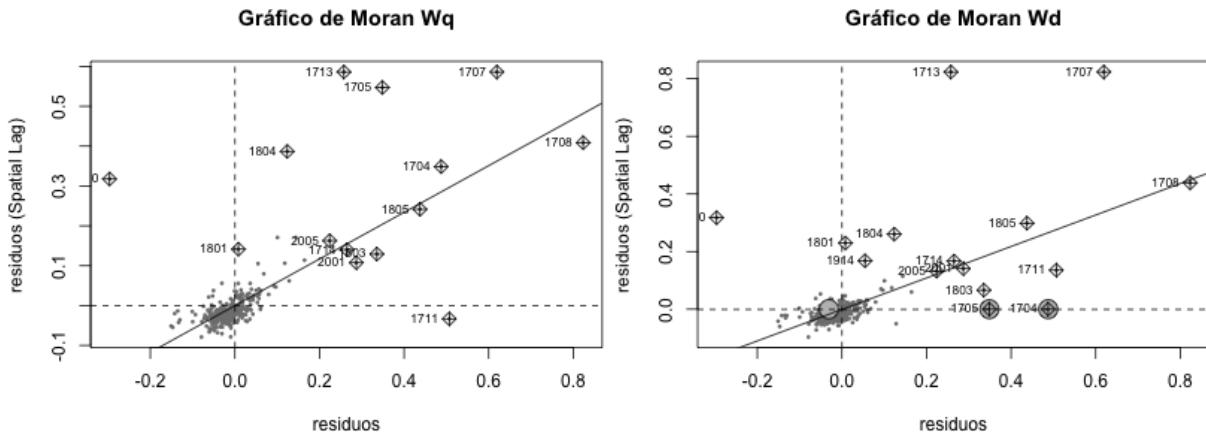


Figure 4.99: Gráfico de Moran para los residuos del modelo lineal para el indicador ‘ia.areas.dist’

#### 4.5.3.3.1 Porcentaje de espacio verde

```
# SAR
```

```
sar.areas_ep <- lagsarlm(formula = as.formula(lm.area_ep.ptje.sel), data = lm.area_ep.ptje.sel,
  listw = W_queen.ev, zero.policy = T, tol.solve = 1e-30)

sar.sm <- summary(sar.areas_ep, Nagelkerke = T)
sar.sm
```

```
##
```

```
## Call:
```

```
## lagsarlm(formula = as.formula(lm.area_ep.ptje.sel), data = lm.area_ep.ptje.sel$model,
##           listw = W_queen.ev, zero.policy = T, tol.solve = 1e-30)
```

```
##
```

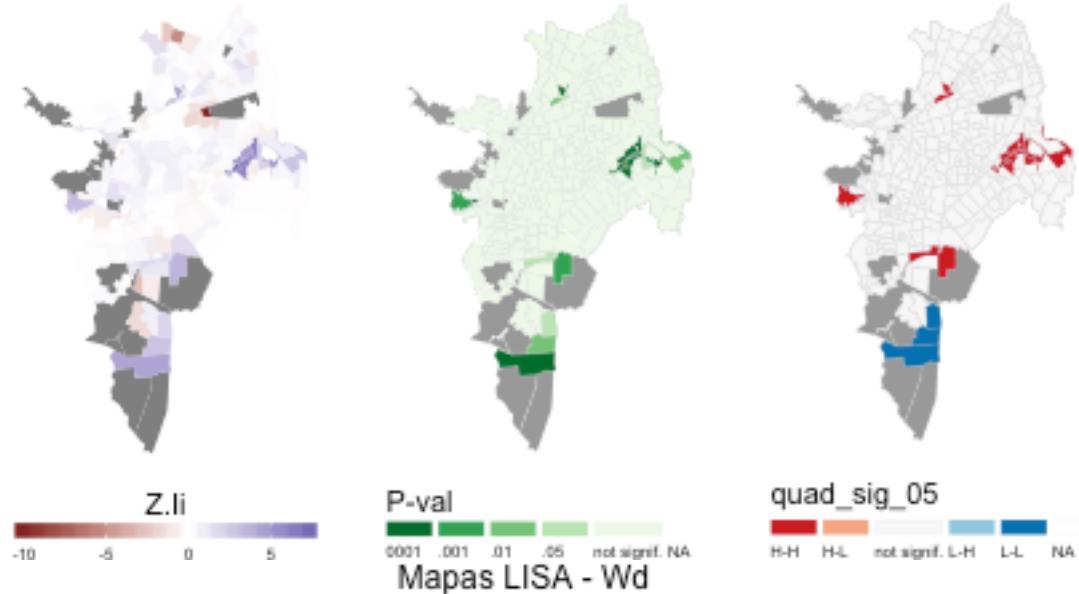
```
## Residuals:
```

```
##      Min       1Q    Median       3Q      Max
## -0.273098 -0.040970 -0.014876  0.019167  0.707957
```

```
##
```

```
## Type: lag
```

Mapas LISA - Wd  
Residuos:  
area\_ep.porcentaje.mxn ~ cuarto.porcentaje.mxn + area\_media\_manzana.mxn



ia.Areas.dist.mxn ~ area\_media\_manzana.mxn + cuarto.porcentaje.mxn + unidad\_economica.porcentaje.mxn + densidad\_poblacion.mxn + con\_alguna\_limitacion.porcentaje.mxn + ningun\_estudio.porcentaje.mxn

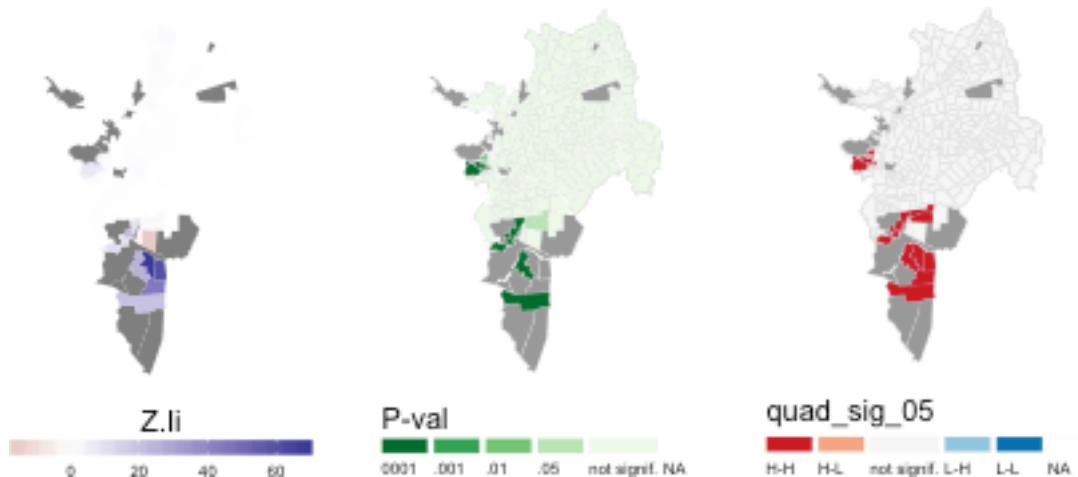


Figure 4.100: Mapas LISA para la matriz  $W_q$  de los residuos de los modelos lineales para los indicadores de acceso a EV

```

## Coefficients: (asymptotic standard errors)

##                               Estimate Std. Error z value Pr(>|z|)

## (Intercept)          0.0388741  0.0073986  5.2542 1.486e-07
## cuarto.porcentaje.mxn -0.0807630  0.0376324 -2.1461   0.03186
## area_media_manzana.mxn  0.7335412  0.0654545 11.2069 < 2.2e-16
## 

## Rho: 0.15309, LR test value: 4.5439, p-value: 0.033036

## Asymptotic standard error: 0.072979

##      z-value: 2.0977, p-value: 0.03593

## Wald statistic: 4.4004, p-value: 0.03593

## 

## Log likelihood: 375.1965 for lag model

## ML residual variance (sigma squared): 0.0059579, (sigma: 0.077187)

## Nagelkerke pseudo-R-squared: 0.31043

## Number of observations: 329

## Number of parameters estimated: 5

## AIC: -740.39, (AIC for lm: -737.85)

## LM test for residual autocorrelation

## test value: 5.7788, p-value: 0.016221

```

#### # SEM

```

sem.areas_ep <- errorsarlm(formula = as.formula(lm.area_ep.ptje.sel), data = lm.area_ep.

listw = W_queen.ev, zero.policy = T, tol.solve = 1e-30)

sem.sm <- summary(sem.areas_ep, Nagelkerke = T)

sem.sm

```

```
##
```

```
## Call:
```

```

## errorsarlm(formula = as.formula(lm.area_ep.ptje.sel), data = lm.area_ep.ptje.sel$mode

##      listw = W_queen.ev, zero.policy = T, tol.solve = 1e-30)

```

```

##  

## Residuals:  

##       Min        1Q    Median        3Q       Max  

## -0.279173 -0.039125 -0.014652  0.021290  0.708489  

##  

## Type: error  

## Coefficients: (asymptotic standard errors)  

##  

##                         Estimate Std. Error z value Pr(>|z|)  

## (Intercept)          0.0491056  0.0069003 7.1165 1.107e-12  

## cuarto.porcentaje.mxn -0.0979094  0.0415838 -2.3545  0.01855  

## area_media_manzana.mxn  0.7712382  0.0653415 11.8032 < 2.2e-16  

##  

## Lambda: 0.25902, LR test value: 10.215, p-value: 0.0013934  

## Asymptotic standard error: 0.078665  

## z-value: 3.2928, p-value: 0.00099206  

## Wald statistic: 10.842, p-value: 0.00099206  

##  

## Log likelihood: 378.0319 for error model  

## ML residual variance (sigma squared): 0.0058066, (sigma: 0.076201)  

## Nagelkerke pseudo-R-squared: 0.32222  

## Number of observations: 329  

## Number of parameters estimated: 5  

## AIC: -746.06, (AIC for lm: -737.85)

# SD

sd.areas_ep <- lagsarlm(formula = as.formula(lm.area_ep.ptje.sel), data = lm.area_ep.ptje,
listw = W_queen.ev, zero.policy = T, tol.solve = 1e-30, type = "mixed")
sd.sm <- summary(sd.areas_ep, Nagelkerke = T)
sd.sm

```

```

##
## Call:
## lagsarlm(formula = as.formula(lm.area_ep.ptje.sel), data = lm.area_ep.ptje.sel$model,
##           listw = W_queen.ev, type = "mixed", zero.policy = T, tol.solve = 1e-30)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.281818 -0.041196 -0.015641  0.025087  0.701126
##
## Type: mixed
## Coefficients: (asymptotic standard errors)
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)          0.0377326  0.0086946  4.3398 1.426e-05
## cuarto.porcentaje.mxn -0.1366936  0.0544000 -2.5128  0.011979
## area_media_manzana.mxn  0.7929675  0.0665592 11.9137 < 2.2e-16
## lag.cuarto.porcentaje.mxn  0.0985071  0.0829144  1.1881  0.234811
## lag.area_media_manzana.mxn -0.3647870  0.1326357 -2.7503  0.005954
##
## Rho: 0.26168, LR test value: 10.614, p-value: 0.0011222
## Asymptotic standard error: 0.078409
##      z-value: 3.3374, p-value: 0.0008458
## Wald statistic: 11.138, p-value: 0.0008458
##
## Log likelihood: 379.4362 for mixed model
## ML residual variance (sigma squared): 0.0057556, (sigma: 0.075866)
## Nagelkerke pseudo-R-squared: 0.32798
## Number of observations: 329
## Number of parameters estimated: 7
## AIC: -744.87, (AIC for lm: -736.26)

```

```

## LM test for residual autocorrelation
## test value: 1.9206, p-value: 0.16579

# SLX

slx.areas_ep <- lmSLX(formula = as.formula(lm.area_ep.ptje.sel), data = lm.area_ep.ptje,
listw = W_queen.ev, zero.policy = T)

slx.sm <- summary(slx.areas_ep)
slx.sm

##
## Call:
## lm(formula = y ~ x - 1, weights = weights)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.26922 -0.04573 -0.01615  0.02428  0.69203
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## x(Intercept)           0.050863  0.007944  6.402 5.37e-10 ***
## xcuarto.porcentaje.mxn -0.133837  0.056072 -2.387  0.0176 *
## xarea_media_manzana.mxn  0.776764  0.068460 11.346 < 2e-16 ***
## xlag.cuarto.porcentaje.mxn  0.066635  0.085226  0.782  0.4349
## xlag.area_media_manzana.mxn -0.158851  0.124959 -1.271  0.2046
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0782 on 324 degrees of freedom
## Multiple R-squared:  0.5316, Adjusted R-squared:  0.5243

```

Table 4.18: Metricas de ajuste para los modelos de porcentaje de área de EV

medidasfit	OLS	SLX	SAR	SEM	SD
Globla Moran'I	0.11636	0.12094	0.04015	-0.00817	-0.00602
GMI p-value	0.00014	0.00008	0.09378	0.56220	0.53618
Shapiro-Wilk	0.76782	0.77525	0.76040	0.75813	0.76159
SW p-value	0.00000	0.00000	0.00000	0.00000	0.00000
Breusch-Pagan	12.98572	14.59961	13.97239	11.07822	13.54161
BP p-value	0.00151	0.00561	0.00092	0.00393	0.00891
Media Residuos	0.00000	0.00000	0.00000	0.00000	0.00000
MSE	0.00607	0.00602	0.00596	0.00581	0.00576
adj-Rsquare	0.29656	0.52434	NA	NA	NA
Nagelkerke pseudo-R-squared	NA	NA	0.31043	0.32222	0.32798
AIC	-737.84913	-736.25817	-740.39306	-746.06371	-744.87233
Log likelihood	372.92456	374.12908	375.19653	378.03185	379.43617

```
## F-statistic: 73.53 on 5 and 324 DF,  p-value: < 2.2e-16
```

```
AIC(slx.areas_ep)
```

```
## [1] -736.2582
```

Al comparar los resultados de las métricas de ajuste se identifica al modelo SEM con el mejor AIC. El modelo SEM logra eliminar la autocorrelación espacial global en los residuos, al igual que SD (ver figura 4.101). Aunque persiste la no normalidad de los residuos y la heterocedasticidad como los muestran los test y las gráficas diagnósticas, el error cometido disminuye y los coeficientes pueden considerarse más confiables. El  $\lambda$  de las estimaciones con términos autoregresivos es significativo, esto sugiere que no es necesario plantear efectos distintivos de la variables dependiente rezagada, y que es posible que ese efecto sea por otras variables no tenidas en cuenta: el agrupamiento espacial observado en la variable dependiente se explica simplemente por el patrón geográfico de variables independientes medidas y no medidas, pero no genera ningún efecto de derrame en el acceso a espacios verdes de los sectores vecinos. A pesar de que los test de normalidad y heterocedasticidad no son existosos, las gráficas diagnósticas muestran que los problemas se presentan en los valores extremos.

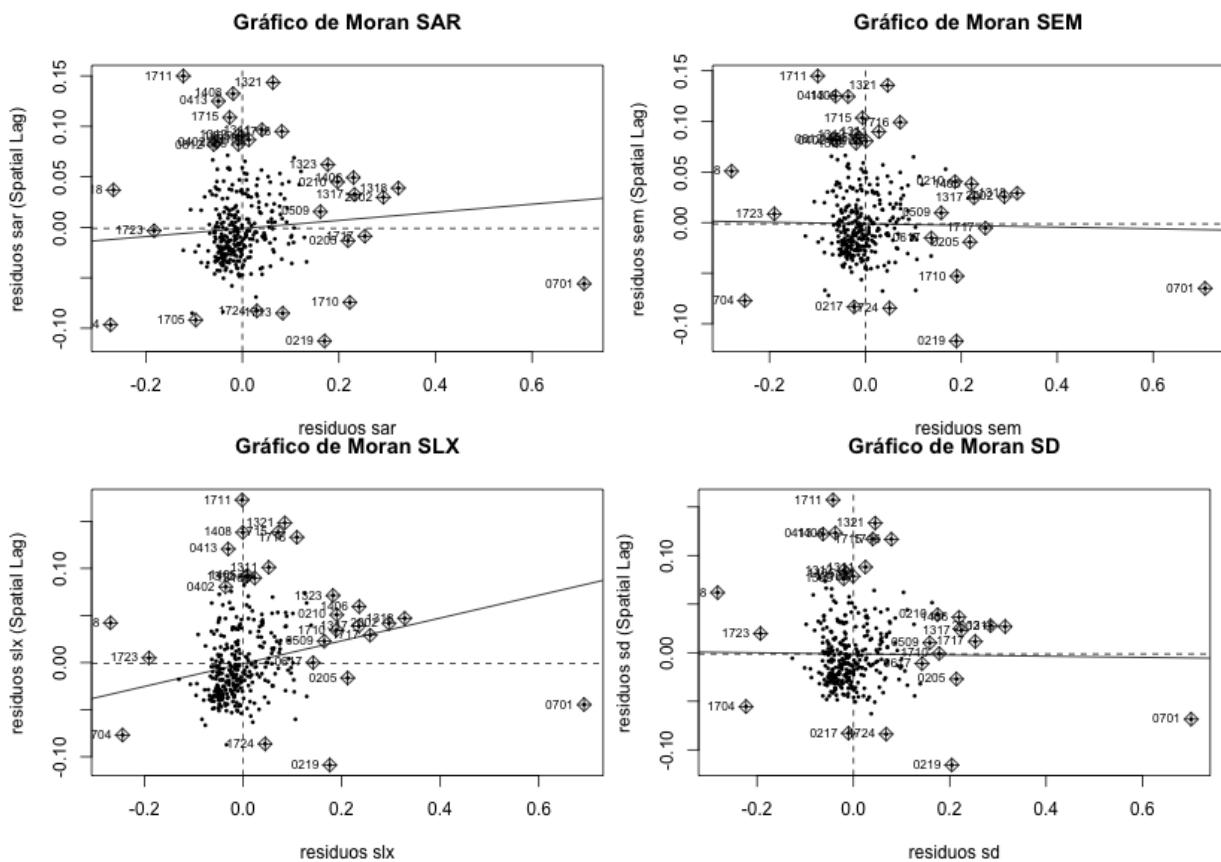


Figure 4.101: Gráfico de Moran para los residuos de los modelos espaciales de porcentaje de área de EV  $W_q$

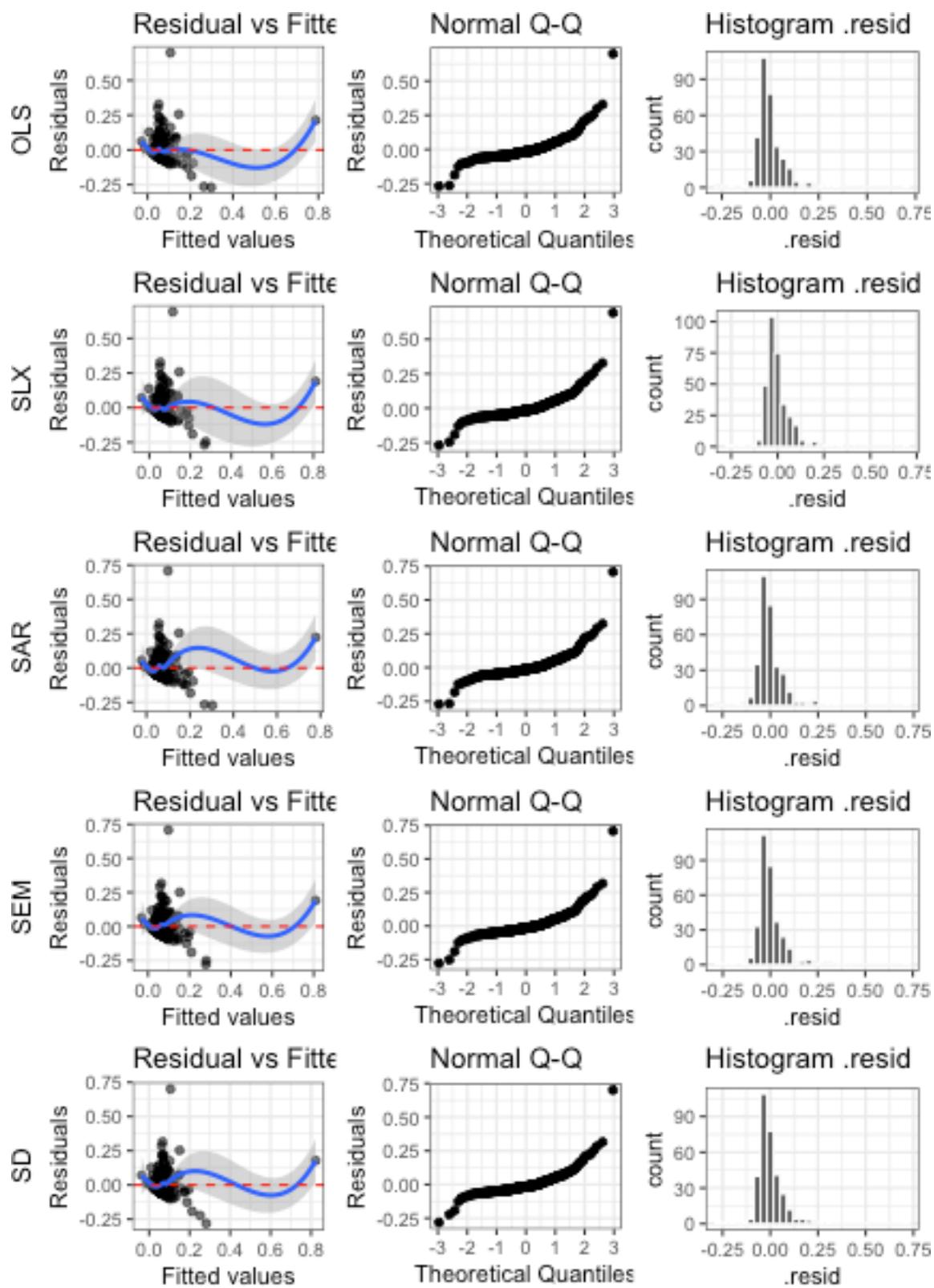


Figure 4.102: Diagnóstico comparativo entre modelos

#### 4.5.3.3.2 Índice de acceso área-distancia

```
# SAR

sar.areas.dist <- lagsarlm(formula = as.formula(lm.ia.areas.dist.sel), data = lm.ia.areas.dist,
  listw = W_queen.ev, zero.policy = T, tol.solve = 1e-30)

sar.sm <- summary(sar.areas.dist, Nagelkerke = T)
sar.sm

##
## Call:
## lagsarlm(formula = as.formula(lm.ia.areas.dist.sel), data = lm.ia.areas.dist.sel$mo,
##   listw = W_queen.ev, zero.policy = T, tol.solve = 1e-30)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.2780316 -0.0122483 -0.0042047  0.0053984  0.4480937
##
## Type: lag
## Coefficients: (asymptotic standard errors)
##                               Estimate Std. Error z value
## (Intercept)                 0.0222143  0.0116341  1.9094
## area_media_manzana.mxn     0.0405084  0.0474367  0.8539
## cuarto.porcentaje.mxn    -0.0480580  0.0325583 -1.4761
## unidad_economica.porcentaje.mxn  0.0543726  0.0227267  2.3925
## densidad_poblacion.mxn    -0.0096962  0.0181819 -0.5333
## con_alguna_limitacion.porcentaje.mxn -0.0445825  0.0220078 -2.0258
## ningun_estudio.porcentaje.mxn     0.0419333  0.0295313  1.4200
##
## Pr(>|z|)
```

```

## (Intercept)          0.05621
## area_media_manzana.mxn    0.39313
## cuarto.porcentaje.mxn     0.13993
## unidad_economica.porcentaje.mxn 0.01674
## densidad_poblacion.mxn      0.59383
## con_alguna_limitacion.porcentaje.mxn 0.04279
## ningun_estudio.porcentaje.mxn    0.15562
##
## Rho: 0.81872, LR test value: 325.38, p-value: < 2.22e-16
## Asymptotic standard error: 0.032949
##      z-value: 24.848, p-value: < 2.22e-16
## Wald statistic: 617.41, p-value: < 2.22e-16
##
## Log likelihood: 471.3541 for lag model
## ML residual variance (sigma squared): 0.0027833, (sigma: 0.052757)
## Nagelkerke pseudo-R-squared: 0.70508
## Number of observations: 329
## Number of parameters estimated: 9
## AIC: -924.71, (AIC for lm: -601.32)
## LM test for residual autocorrelation
## test value: 90.367, p-value: < 2.22e-16

# SEM

sem.areas.dist <- errorsarlm(formula = as.formula(lm.ia.areas.dist.sel), data = lm.ia.areas.dist,
                                listw = W_queen.ev, zero.policy = T, tol.solve = 1e-30)
sem.sm <- summary(sem.areas.dist, Nagelkerke = T)
sem.sm

##
## Call:

```

```

## errorsarlm(formula = as.formula(lm.ia.areas.dist.sel), data = lm.ia.areas.dist.sel$)
##      listw = W_queen.ev, zero.policy = T, tol.solve = 1e-30)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.2739137 -0.0131061 -0.0043052  0.0043362  0.4534217
##
## Type: error
## Coefficients: (asymptotic standard errors)
##
##                                     Estimate Std. Error z value
## (Intercept)                  0.0715688  0.0256916 2.7857
## area_media_manzana.mxn      -0.0984051  0.0471448 -2.0873
## cuarto.porcentaje.mxn       0.0015127  0.0445226  0.0340
## unidad_economica.porcentaje.mxn 0.0644496  0.0292204  2.2056
## densidad_poblacion.mxn     -0.0168968  0.0228046 -0.7409
## con_alguna_limitacion.porcentaje.mxn -0.0379299  0.0242676 -1.5630
## ningun_estudio.porcentaje.mxn  0.0181776  0.0426649  0.4261
##
##                                     Pr(>|z|)
## (Intercept)                  0.005342
## area_media_manzana.mxn      0.036861
## cuarto.porcentaje.mxn       0.972897
## unidad_economica.porcentaje.mxn 0.027410
## densidad_poblacion.mxn      0.458730
## con_alguna_limitacion.porcentaje.mxn 0.118056
## ningun_estudio.porcentaje.mxn  0.670068
##
## Lambda: 0.85804, LR test value: 320.82, p-value: < 2.22e-16
## Asymptotic standard error: 0.029167
##      z-value: 29.418, p-value: < 2.22e-16

```

```

## Wald statistic: 865.41, p-value: < 2.22e-16
##
## Log likelihood: 469.0722 for error model
## ML residual variance (sigma squared): 0.0027445, (sigma: 0.052388)
## Nagelkerke pseudo-R-squared: 0.70096
## Number of observations: 329
## Number of parameters estimated: 9
## AIC: -920.14, (AIC for lm: -601.32)

# SD

sd.areas.dist <- lagsarlm(formula = as.formula(lm.ia.areas.dist.sel), data = lm.ia.areas.dist,
                           listw = W_queen.ev, zero.policy = T, tol.solve = 1e-30, type = "mixed")
sd.sm <- summary(sd.areas.dist, Nagelkerke = T)
sd.sm

##
## Call:
## lagsarlm(formula = as.formula(lm.ia.areas.dist.sel), data = lm.ia.areas.dist.sel$mode,
##           listw = W_queen.ev, type = "mixed", zero.policy = T, tol.solve = 1e-30)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.2816309 -0.0138513 -0.0016856  0.0080895  0.4185725
##
## Type: mixed
## Coefficients: (asymptotic standard errors)
##                               Estimate Std. Error z value
## (Intercept)                 -0.01783647  0.01770210 -1.0076
## area_media_manzana.mxn      0.02936373  0.05041803  0.5824

```

```

## cuarto.porcentaje.mxn           -0.00306343  0.04350159 -0.0704
## unidad_economica.porcentaje.mxn 0.06683674  0.02814454  2.3748
## densidad_poblacion.mxn          -0.00520232  0.02224585 -0.2339
## con_alguna_limitacion.porcentaje.mxn -0.04263328  0.02319303 -1.8382
## ningun_estudio.porcentaje.mxn    0.03469916  0.04147606  0.8366
## lag.area_media_manzana.mxn       0.73732180  0.10534926  6.9988
## lag.cuarto.porcentaje.mxn        -0.08904581  0.06971761 -1.2772
## lag.unidad_economica.porcentaje.mxn 0.01183435  0.04319296  0.2740
## lag.densidad_poblacion.mxn       0.02328887  0.03742061  0.6224
## lag.con_alguna_limitacion.porcentaje.mxn -0.00095268  0.04277167 -0.0223
## lag.ningun_estudio.porcentaje.mxn 0.08486275  0.06097300  1.3918
##
## Pr(>|z|)
## (Intercept)                      0.31365
## area_media_manzana.mxn            0.56029
## cuarto.porcentaje.mxn             0.94386
## unidad_economica.porcentaje.mxn  0.01756
## densidad_poblacion.mxn            0.81510
## con_alguna_limitacion.porcentaje.mxn 0.06603
## ningun_estudio.porcentaje.mxn    0.40281
## lag.area_media_manzana.mxn        2.581e-12
## lag.cuarto.porcentaje.mxn         0.20152
## lag.unidad_economica.porcentaje.mxn 0.78409
## lag.densidad_poblacion.mxn        0.53371
## lag.con_alguna_limitacion.porcentaje.mxn 0.98223
## lag.ningun_estudio.porcentaje.mxn 0.16398
##
## Rho: 0.70879, LR test value: 213.29, p-value: < 2.22e-16
## Asymptotic standard error: 0.041987
##      z-value: 16.881, p-value: < 2.22e-16

```

```

## Wald statistic: 284.97, p-value: < 2.22e-16
##
## Log likelihood: 499.677 for mixed model
## ML residual variance (sigma squared): 0.0024857, (sigma: 0.049857)
## Nagelkerke pseudo-R-squared: 0.75173
## Number of observations: 329
## Number of parameters estimated: 15
## AIC: -969.35, (AIC for lm: -758.06)
## LM test for residual autocorrelation
## test value: 43.015, p-value: 5.4335e-11

# SLX

slx.areas.dist <- lmSLX(formula = as.formula(lm.ia.areas.dist.sel), data = lm.ia.areas.dist,
  listw = W_queen.ev, zero.policy = T)

slx.sm <- summary(slx.areas.dist)
slx.sm

##
## Call:
## lm(formula = y ~ x - 1, weights = weights)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.32181 -0.02997 -0.00644  0.01483  0.58277
##
## Coefficients:
##                               Estimate Std. Error t value
## x(Intercept)                0.032545  0.026528  1.227
## xarea_media_manzana.mxn    0.316823  0.070305  4.506

```

```

## xcuarto.porcentaje.mxn      -0.049277  0.065171 -0.756
## xunidad_economica.porcentaje.mxn    0.063792  0.042200  1.512
## xdensidad_poblacion.mxn      0.012811  0.033332  0.384
## xcon_alguna_limitacion.porcentaje.mxn -0.058996  0.034711 -1.700
## xningun_estudio.porcentaje.mxn     0.076113  0.062079  1.226
## xlag.area_media_manzana.mxn       1.508539  0.142559 10.582
## xlag.cuarto.porcentaje.mxn      -0.261660  0.103093 -2.538
## xlag.unidad_economica.porcentaje.mxn 0.177661  0.062842  2.827
## xlag.densidad_poblacion.mxn      0.002892  0.056115  0.052
## xlag.con_alguna_limitacion.porcentaje.mxn -0.206310  0.063527 -3.248
## xlag.ningun_estudio.porcentaje.mxn   0.351589  0.090015  3.906
##
##                                         Pr(>|t|)

## x(Intercept)                      0.220801
## xarea_media_manzana.mxn          9.3e-06 ***
## xcuarto.porcentaje.mxn          0.450140
## xunidad_economica.porcentaje.mxn 0.131617
## xdensidad_poblacion.mxn         0.700989
## xcon_alguna_limitacion.porcentaje.mxn 0.090182 .
## xningun_estudio.porcentaje.mxn   0.221088
## xlag.area_media_manzana.mxn      < 2e-16 ***
## xlag.cuarto.porcentaje.mxn      0.011626 *
## xlag.unidad_economica.porcentaje.mxn 0.004996 **
## xlag.densidad_poblacion.mxn      0.958926
## xlag.con_alguna_limitacion.porcentaje.mxn 0.001289 **
## xlag.ningun_estudio.porcentaje.mxn   0.000115 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07476 on 316 degrees of freedom

```

Table 4.19: Metricas de ajuste para los modelos de áreas-distancia de EV

medidasfit	OLS	SLX	SAR	SEM	SD
Globla Moran'I	0.58535	0.53678	-0.18235	-0.17202	-0.12515
GMI p-value	0.00000	0.00000	1.00000	1.00000	0.99991
Shapiro-Wilk	0.57372	0.66730	0.51412	0.50303	0.61029
SW p-value	0.00000	0.00000	0.00000	0.00000	0.00000
Breusch-Pagan	57.30768	63.03543	20.13534	6.43731	81.79722
BP p-value	0.00000	0.00000	0.00262	0.37602	0.00000
Media Residuos	0.00000	0.00000	0.00000	0.00000	0.00000
MSE	0.00897	0.00537	0.00278	0.00274	0.00249
adj-Rsquare	0.19231	0.60810	NA	NA	NA
Nagelkerke pseudo-R-squared	NA	NA	0.70508	0.70096	0.75173
AIC	-601.32470	-758.06361	-924.70818	-920.14450	-969.35400
Log likelihood	308.66235	393.03180	471.35409	469.07225	499.67700

```
## Multiple R-squared:  0.6236, Adjusted R-squared:  0.6081
```

```
## F-statistic: 40.27 on 13 and 316 DF,  p-value: < 2.2e-16
```

```
AIC(slx.areas.dist)
```

```
## [1] -758.0636
```

Al comparar los resultados de las métricas de ajuste se identifica al modelo SD con el mejor AIC. El modelo SD logra eliminar la autocorrelación espacial global en los residuos al exhibir un valor de significancia mucho mayor que 0.05, al igual que SEM y SAR (ver figura 4.103). Aunque persiste la no normalidad de los residuos y la heterocedasticidad como los muestran los test y las gráficas diagnósticas, el error cometido disminuye y los coeficientes pueden considerarse más confiables. El  $\rho$  de las estimaciones con términos autoregresivos es significativo, esto sugiere que una parte de los efectos de derrame de las variables significativas (`lag.area_media_manzana`, `unidad_economica.porcentaje`) en el acceso a espacios verdes de los sectores vecinos explica bien los grupos que se forman. Es posible que estar cerca de un sector con manzanas grandes y que tal vez sus parque pueden ser mas grandes explica la influencia positiva en el acceso p.e. de tener equipamientos de ciudad grandes en algun sector aledaño. En el modelo SD una parte de la mejora en el ajuste puede provenir de dimensiones no modeladas. Los test de normalidad y

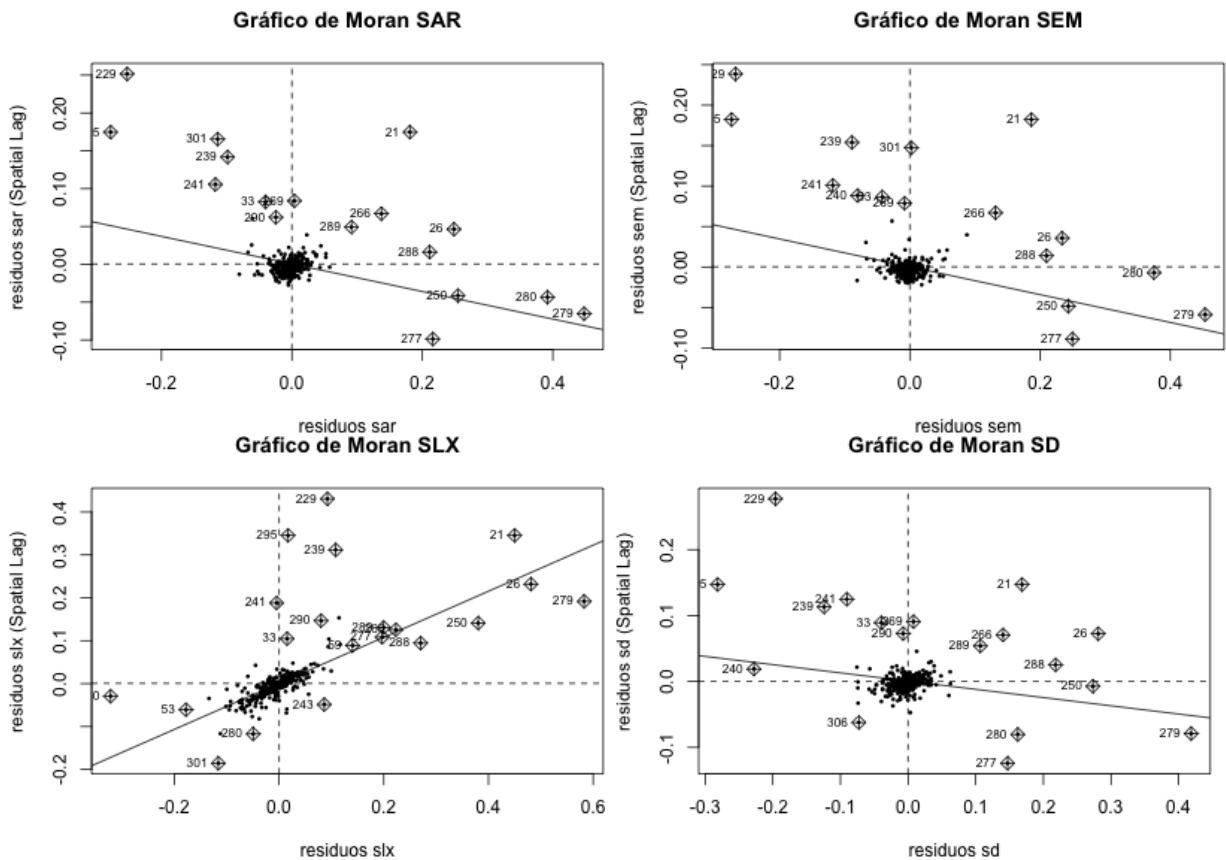


Figura 4.103: Gráfico de Moran para los residuos de los modelos espaciales de acceso área-distancia  $W_q$

heterocedasticidad no son existosos, las gráficas diagnósticas muestran que los problemas se presentan en los valores extremos.

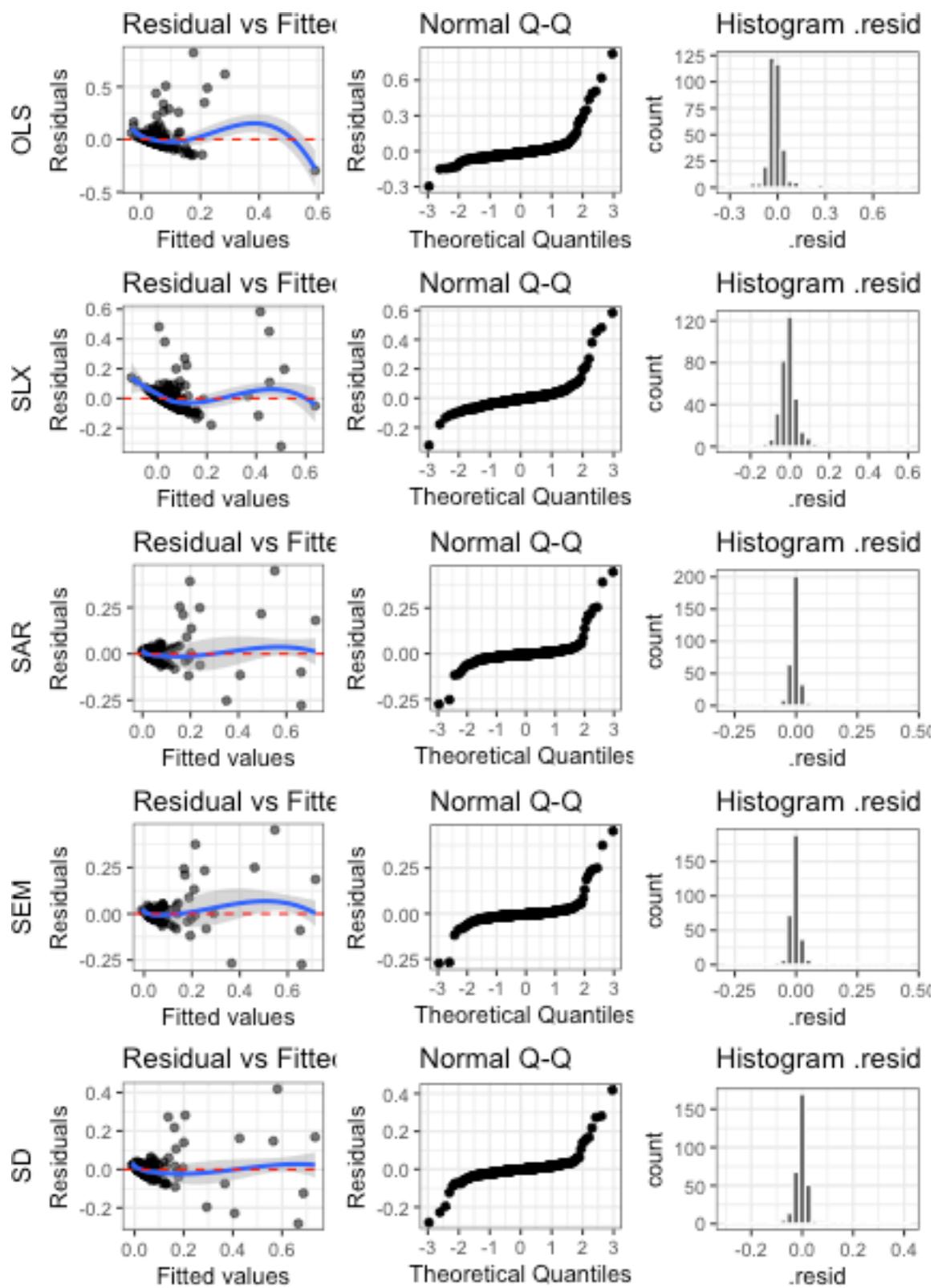


Figure 4.104: Diagnóstico comparativo entre modelos

# Chapter 5

## Discusión

Antes de iniciar con la discusión de los resultados es importante recordar que los modelos realizados son un intento por capturar evidencia empírica sobre el nivel explicativo de un conjunto de condiciones de la población y las condiciones de vida que provee la ciudad de Cali en lo que beneficios ambientales se refiere. El modelo no es la realidad ni prueba de causalidad, es un esfuerzo por cuantificar la correlación entre algunas de esas dimensiones y tener certeza de su potencial asociación. El uso de modelos espaciales indaga sobre los patrones espaciales de esas variables, lo que permite la identificación de zonas que agrupan características altas de esos servicios o de las condiciones de la población y de la estructura urbana. Los análisis no buscan hacer inferencia sobre la población, pues no son una muestra, son la población completa de árboles, espacio verde y los datos del censo. En este sentido los coeficientes y su representatividad son sobre la población, y los interpretaremos como la importancia relativa de esas variables sobre las otras.

### 5.1 Sobre la cobertura de copa

Aunque el área de copa y el porcentaje de cobertura de copa configuran dos modelos individuales y son indicadores que expresan de diferente forma el nivel de acceso a servicios

del arbolado urbano, ambos son visiones complementarias sobre el mismo fenómeno, la interpretación no aísla los modelos, los relaciona.

A pesar de la persistencia de problemas en la estimación como la no normalidad de los residuos y la heterocedasticidad —esto debidos ya no la autocorrelación de la variable dependiente ni a la de los errores sino a posibles no linealidades entre los predictores y la cobertura de copa— las estimaciones mejoran con los modelos con estructura espacial. Inclusive parece que las matrices de vecindad construidas, aunque muy similares entre sí, muestran que la inclusión a priori de una estructura espacial es un mecanismo eficaz para identificar grupos y probar aspectos teóricos de la estructura de la autocorrelación si se profundiza e incluye criterios teóricos o conocimientos del desarrollo histórico de la ciudad en su formalización. Es posible que dichos problemas de la estimación provengan de dimensiones no incluidas en el modelo dado que los modelos SEM también tienen un rendimiento similar al de los mejor rankeados SAR y SD, escogidos como mejores candidatos para la estimación del porcentaje de cobertura de copa y área de copa, respectivamente. En este sentido traemos a colación aspectos como el importante número de árboles en los andes y separadores viales (figura 4.18), factor que no puede no ser capturado en las métricas construidas, pero que puede explicar el agrupamiento de sectores censales con buena cobertura de copa al rededor de la calle 5ta que se observa en el mapa de clusters del porcentaje de cobertura de copa (ver figura 5.4) o el bajo desarrollo de los individuos en sectores de calles estrechas y con pocas vías principales.

Así pues los coeficientes de los dos modelos ajustados se detallan en las tablas 5.1 y 5.2 para el modelo SD y 5.3 y 5.4 para el modelo SAR. En consecuencia con el ajuste de autocorrelación en ambos modelos (SAR y SD) el  $\rho$  es significativo y proporcional a las diferencias en el Moran's I Global de cada modelo. Es interesante que en ambos modelos la variable más significativa y de mayor valor en el coeficiente es **estudios superiores**. Cuando se formulan los modelos se excluye el uso de variables como porcentaje de población afro o personas sin estudios por tener una alta correlación negativa con personas con **estudios superiores o postgrado** o con su versión porcentual. Esto implica que los beneficios am-

Table 5.1: Coeficientes del modelo SD de área de copa

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	0.637	0.065	9.828	0.000
superior_postgrado.mxn	0.287	0.024	12.023	0.000
densidad_poblacion.mxn	-0.096	0.025	-3.783	0.000
cuarto.porcentaje.mxn	-0.059	0.041	-1.438	0.151
area_ep.mxn	0.141	0.030	4.689	0.000
lag.superior_postgrado.mxn	-0.016	0.047	-0.331	0.741
lag.densidad_poblacion.mxn	-0.050	0.037	-1.356	0.175
lag.cuarto.porcentaje.mxn	-0.146	0.065	-2.226	0.026
lag.area_ep.mxn	-0.074	0.040	-1.832	0.067

Table 5.2: Coeficiente de autocorrelación modelo SD de área de copa

\$\rho\$	Likelihood ratio	p-value
0.201	6.415	0.011

bientales son mejores en sectores con población mejor educada (figura 5.3), posiblemente una razón para preferir o habitar espacios con buena arborización o tal vez con andes con suficiente espacio y zonas blandas para arborizar el barrio. En cualquier caso, esto muestra un grado de segregación de los clusters de sectores con mayor arborización (figura 5.3) y la población afro (figura 5.1) y sin estudios (figura 5.2), ambos altamente correlacionadas, como se aprecia en los mapas de LISA y las matriz de correlación 4.46 y 4.47

La densidad de población es una variable que produce resultados inconsistentes: aunque es significativo en ambos modelos, y tiene un coeficiente negativo para área de copa y positivo para el porcentaje de cobertura. Los patrones espaciales del área de copa y porcentaje no coinciden completamente, muestran dos patrones del mismo fenómeno que

Table 5.3: Coeficientes del modelo SAR de porcentaje de área de copa

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	0.030	0.011	2.822	0.005
superior_postgrado.porcentaje.mxn	0.246	0.027	9.023	0.000

Table 5.4: Coeficiente de autocorrelación modelo SAR de área de copa

$\$\\rho\$$	Likelihood ratio	p-value
0.375	28.579	0

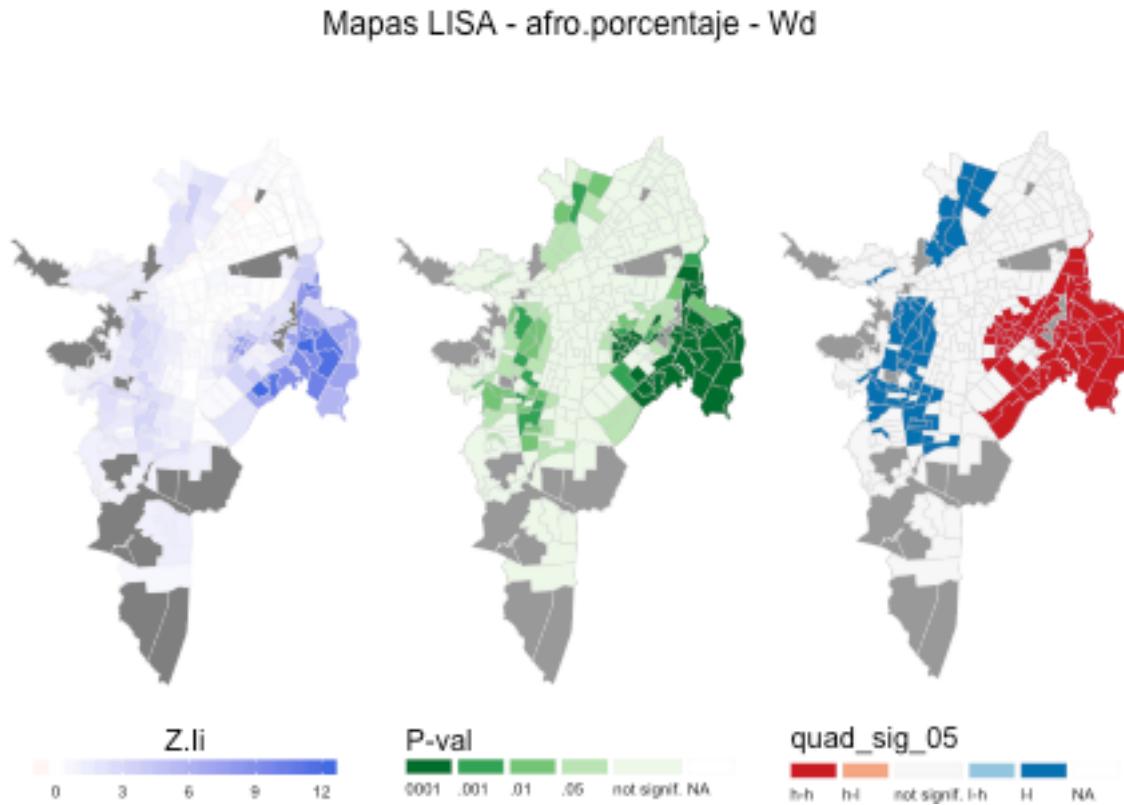


Figure 5.1: Mapas LISA - Porcentaje población Afro

Mapas LISA - ningun\_estudio.porcentaje - Wd

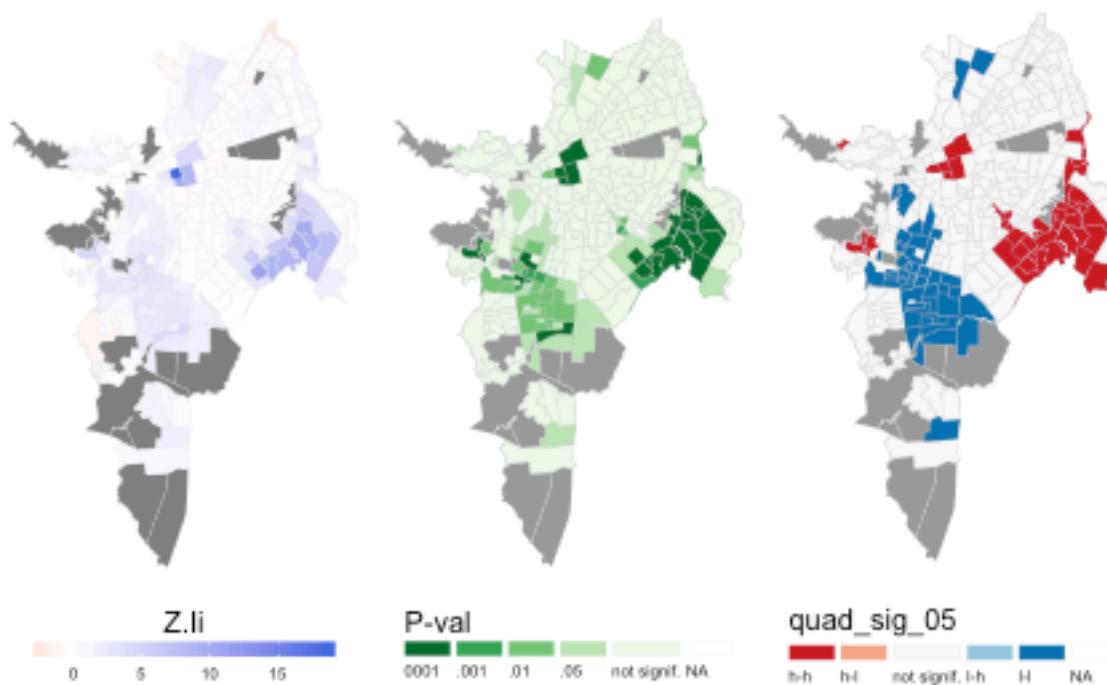


Figure 5.2: Mapas LISA - Porcentaje población sin estudios

Mapas LISA - superior\_postgrado.porcentaje - Wd

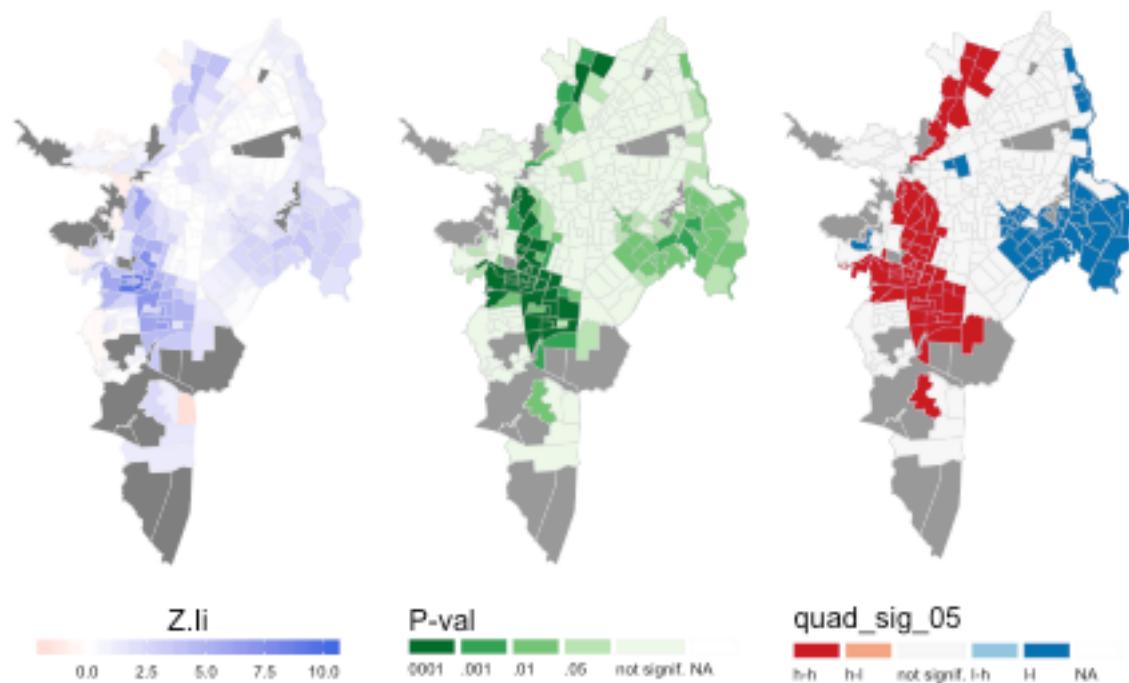


Figure 5.3: Mapas LISA - Porcentaje población con estudios superiores

Mapas LISA - sqrt.cobertura\_copa.ap - Wd

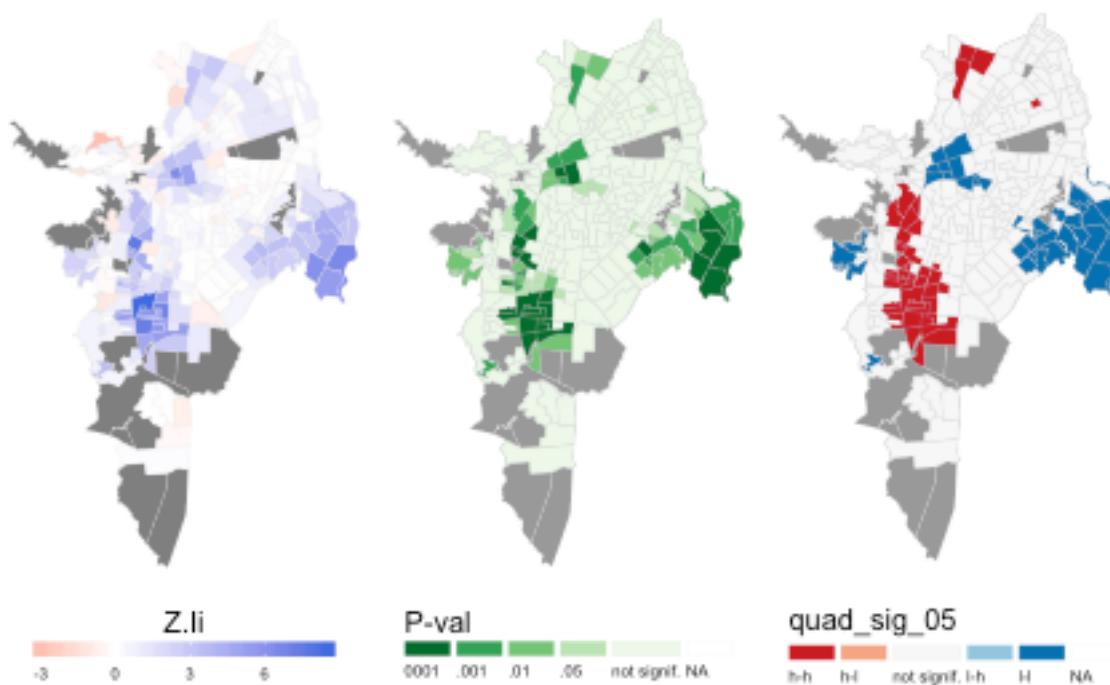


Figure 5.4: Mapas LISA - Porcentaje cobertura de copa

### Mapas LISA - sqrt.area\_copa - Wq

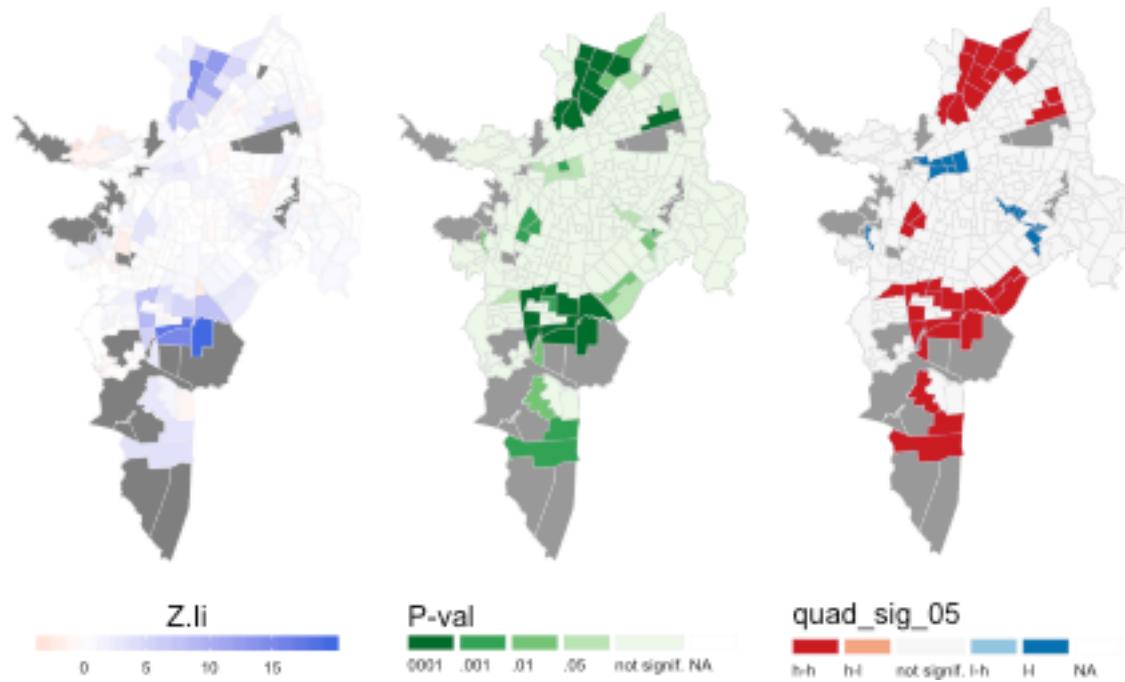


Figure 5.5: Mapas LISA - Área de cobertura de copa

se ve alterado por la heterogeneidad en los tamaños de los sectores censales de un indicador absoluto a uno porcentual y que nos muestran concentraciones en zonas con cierto grado de complementariedad y continuidad con zonas de traslape de los sectores que los componen (figuras 5.5 y 5.4). Quizá esta situación se la causa de la inconsistencia en el coeficientes de la densidad de población que toca ambos patrones de la cobertira pero no de forma simetrica.

En el modelo SD de las variables retardadas ninguna es significativa. El porcentaje de viviendas tipo cuarto no es significativo pues solo aporta como predictor para la zona del centro de la ciudad (ver figura 5.7) que ya es cubierto por la distribución espacial de la variable sobre los estudios superiores. Es significativo el porcentaje de área de espacio público, con el segundo coeficiente más elevado explicando de forma consiste el gran número de árboles que se encuentran en estos lugares, pero que explica la cobertura en puntos muy específicos de la ciudad (figura 5.8). No existe chance de que tengan un

Mapas LISA - densidad\_poblacion - Wq

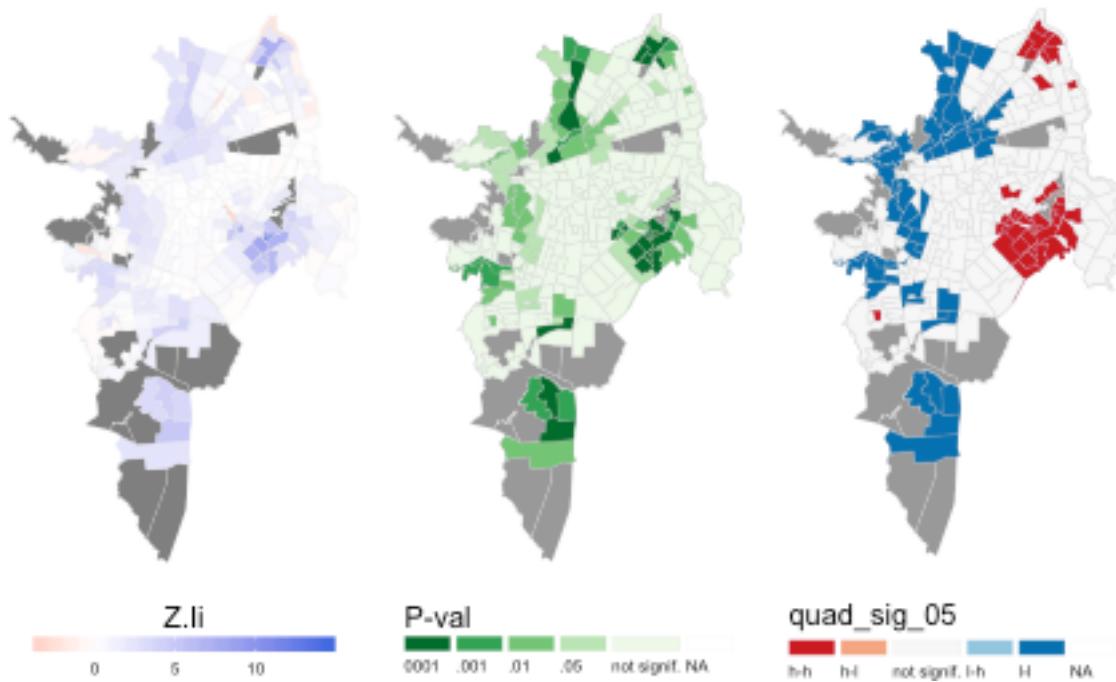


Figure 5.6: Mapas LISA - Densidad de Población

### Mapas LISA - cuarto.porcentaje - Wq

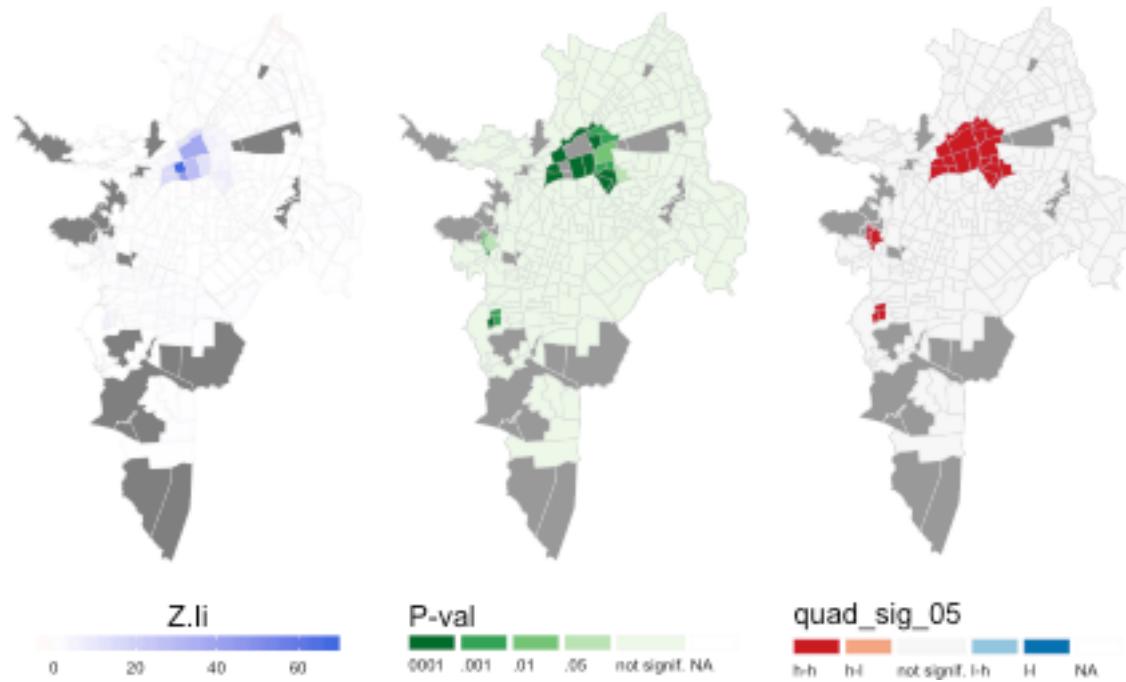


Figure 5.7: Mapas LISA - Porcentaje de viviendas tipo cuarto

efecto de difusión o derrame pues la creación de estas zonas verdes es algo que ocurre con la urbanización del sector y poco o nada cambia después de su creación.

Se puede afirmar que la hipótesis que se plantea este trabajo se cumple: “Las zonas urbanas con más desventajas sociales y económicas tienen menos acceso a beneficios del arbolado urbano.” Existe una fuerte asociación entre los patrones espaciales de la cobertura de copa y un estatus social más alto representado por al acceso a educación superior, que en Colombia no es de acceso universal y la oferta educacional está cubierta principalmente por instituciones privadas. Es además fuerte la asociación negativa entre este indicador de estatus social con zonas con prevalencia de población afro y población sin estudios, que además de la alta segregación laboral que exhiben (Arroyo Mina et al., 2016; Mora and Arcila, 2014; Cerón and Escobar, 2014) está menos beneficiada de los servicios ambientales provisto por el arbolado urbano.

Mapas LISA - area\_ep.porcentaje - Wq

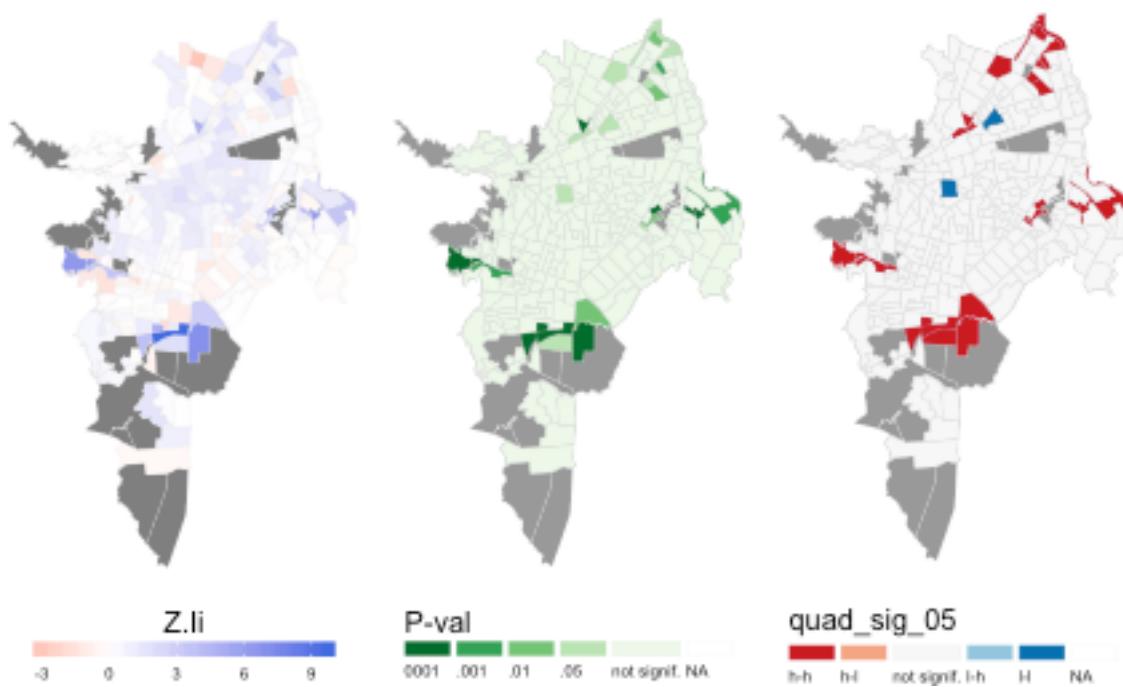


Figure 5.8: Mapas LISA - Porcentaje de espacios verdes

Este tipo de desigualdades son una responsabilidad de los urbanizadores y de los gobiernos locales. Es de su interés y responsabilidad dialogar para construir y propiciar espacios para la siembra y desarrollo de los individuos arbóreos, pues es fundamental su agencia para reducir brechas sociales, crear una ciudad con espacio e infraestructura natural que sostenga los ecosistemas de los que depende la ciudad y la calidad de vida de sus ciudadanos. Ciudadanos para quienes lo ambiental viene cobrando más importancia como valor social, y que además se convierte en una estrategia para mitigar los cambios que pueda traer consigo el calentamiento global.

Este tipo de métodos para establecer condiciones objetivas de la población y las interrelaciones de estas condiciones son relevantes para construir políticas públicas sobre métricas que permitan hacer seguimiento a los impactos de los planes y proyectos sobre el manejo del arbolado urbano y la infraestructura necesaria para transformar el paisaje y la selección de zonas que necesitan de mayor atención.

## 5.2 Sobre los espacios verdes

Análogamente a la cobertura arbórea, las métricas de acceso son aspectos complementarios sobre el mismo fenómeno de acceso aún beneficio; una representa un beneficio local, para el que vive en ese sector (`area_ep.porcentaje`) y la otra una forma (`ia.areas.dist`) que involucra un grupos de sectores beneficiarios, y que mide la relación entre el área del espacio verde y la distancia desde el centroide del sector urbano para luego sumar los valores de los espacios en un rango de distancia: intenta cuantificar la relación de costo de traslado por beneficio en área. El modelo de mejor rendimiento en la estimación medido usando el AIC fue el SEM para `area_ep.porcentaje` y SD para `ia.areas.dist`. Los coeficientes de los dos modelos ajustados se detallan en la tablas 5.5 y 5.6 para el modelo SEM y 5.7 y 5.8 para el modelo SD.

En el SEM todos los coeficientes son significativos y el valor de área media de manzana es la variable más influyente. El coeficiente de autocorrelación en el error es muy significativo,

Table 5.5: Coeficientes del modelo SEM de porcentaje de área de espacio verde

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	0.049	0.007	7.116	0.000
cuarto.porcentaje.mxn	-0.098	0.042	-2.355	0.019
area_media_manzana.mxn	0.771	0.065	11.803	0.000

Table 5.6: Coeficiente de autocorrelación modelo SEM de porcentaje de área de espacio verde

\$\lambda\$	Likelihood ratio	p-value
0.259	10.215	0.001

lo que sugiere que si existe información de patrones espaciales en el error pero en variables no modeladas. La importancia que tiene en el modelo el área media de manzana puede interpretarse como una característica estructural de barrio o sector que por tener manzanas más grandes las zonas destinadas para parque o espacio verde son en consecuencia más grandes y el beneficio mayor o/y que algunas manzanas que albergan equipamientos de ciudad, zonas verdes como riveras de ríos o lagunas determinan el alto valor del indicador al ser el promedio sensible a valores extremos. Lo cierto es que desde este punto de vista local no existe evidencia sobre una relación concluyente con ninguna variable poblacional, por lo que no puede decirse que se vulnere a ciertas comunidades o grupos diferenciados, al menos en lo que respecta a las pocas variables que proveen los sistemas de consulta del censo del 2005. Pero puede decirse la la distribución de los espacio es muy desigual solo ver en el histograma y el gráfico de dispersión (figura 5.9) que muestra que mas de 203 de 329 sectores en la ciudad tiene menos del 5% de espacio verde.

El indicador de acceso area-distancia fue mejor ajustado por el modelo SD, que tuvo como coeficientes significativos a `unidad_economica.porcentaje` con un valor bajo pero positivo al igual que `con_alguna_limitacion.porcentaje` que tuvo un valor bajo pero negativo y significativo. La zona que muestra el mapa de LISA el porcentaje de unidades económicas en el sector (figura 5.12) coinciden de nuevo con equipamientos de ciudad como el Boulevar del Río mientras que las concentraciones de personas con limitación altas

Table 5.7: Coeficientes del modelo SD del indicador de acceso área-distancia

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.018	0.018	-1.008	0.314
area_media_manzana.mxn	0.029	0.050	0.582	0.560
cuarto.porcentaje.mxn	-0.003	0.044	-0.070	0.944
unidad_economica.porcentaje.mxn	0.067	0.028	2.375	0.018
densidad_poblacion.mxn	-0.005	0.022	-0.234	0.815
con_alguna_limitacion.porcentaje.mxn	-0.043	0.023	-1.838	0.066
ningun_estudio.porcentaje.mxn	0.035	0.041	0.837	0.403
lag.area_media_manzana.mxn	0.737	0.105	6.999	0.000
lag.cuarto.porcentaje.mxn	-0.089	0.070	-1.277	0.202
lag.unidad_economica.porcentaje.mxn	0.012	0.043	0.274	0.784
lag.densidad_poblacion.mxn	0.023	0.037	0.622	0.534
lag.con_alguna_limitacion.porcentaje.mxn	-0.001	0.043	-0.022	0.982
lag.ningun_estudio.porcentaje.mxn	0.085	0.061	1.392	0.164

Table 5.8: Coeficiente de autocorrelación modelo SD del indicador de acceso área-distancia

\$\rho\$	Likelihood ratio	p-value
0.709	213.29	0

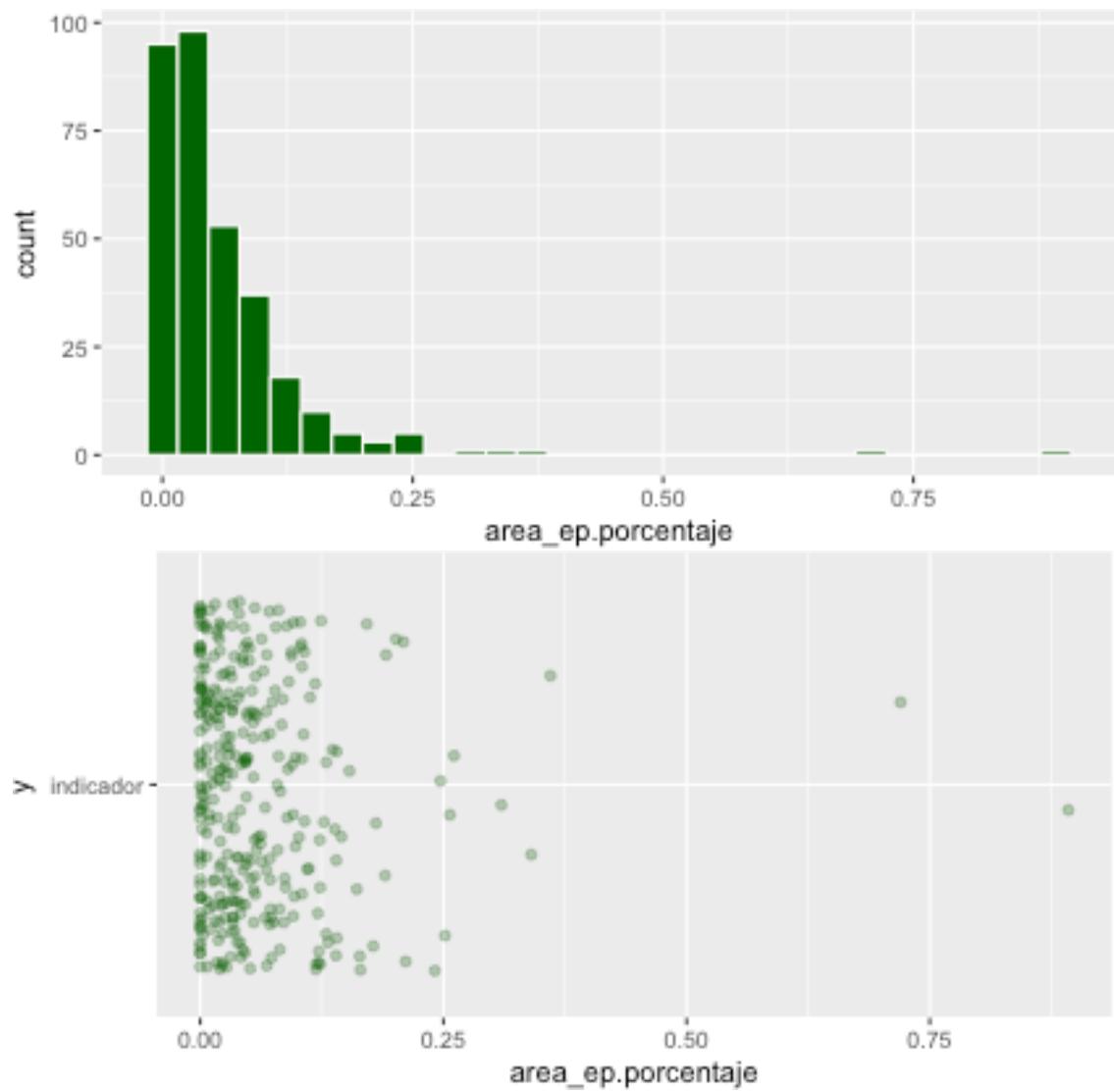


Figure 5.9: Distribucion de los indicadores de acceso local a EV

coinciden con poco espacio verde como bien indica el signo del coeficiente. La variable de área media de manzana retardada `lag.area_media_manzana` es significativa y es la que tiene el valor más alto entre los coeficientes, confirmando el efecto de derrame de los grandes espacio verdes que son parte del equipamiento de ciudad y que producen beneficios en un radio más amplio que el área de sector donde el disfrute exige que la gente se desplace, justo como lo propone el indicador de acceso. El coeficiente de autocorrelación espacial

$\rho$ . es muy alto y significativo, lo que indica que la introducción de autocorrelación espacial en la variable dependiente mejora el ajuste pues es evidente la existencia de grupos de sectores con valores similares.

A pesar de que se mantienen problemas como no normalidades en los errores, el ajuste de la variable dependiente presenta mejoras en las métricas de error y hace más confiables la estimación de los coeficientes. En cuanto al acceso a espacio verdes que involucran desplazamientos o una cobertura a grupos de sectores urbanos no puede decirse de forma irrefutable que está estrechamente relacionado con desventajas en la población como el estatus o la etnicidad, pero si marca que hay patrones espaciales en la población de limitaciones físicas o de salud que coinciden con niveles pobres de acceso. Al igual que con el indicador de acceso local existe una distribución del índice que muestra gran cantidad de sectores con muy bajos valores y algunos con valores altos cuyos vecinos se ven beneficiados(ver figura 5.10).

La estrategia de construir equipamientos con espacios verdes para mejorar el acceso puede verse optimizada de este tipo de análisis, pues resalta las zonas donde el impacto de estas obras puede llegar a más personas si se ubican correctamente. Marca también la responsabilidad de los urbanizadores y las autoridades para generar proyectos de renovación urbana que integren y den relevancia al acceso a EV de manera razonada. Estos resultados apuntan a tener en cuenta aspectos de la planeación urbana al nivel del tamaño de las manzanas y la existencia de andenes que permitan alojar parques de mayor tamaño y espacio para que se desarrolle el potencial recreativo de esos espacios, así como las especies

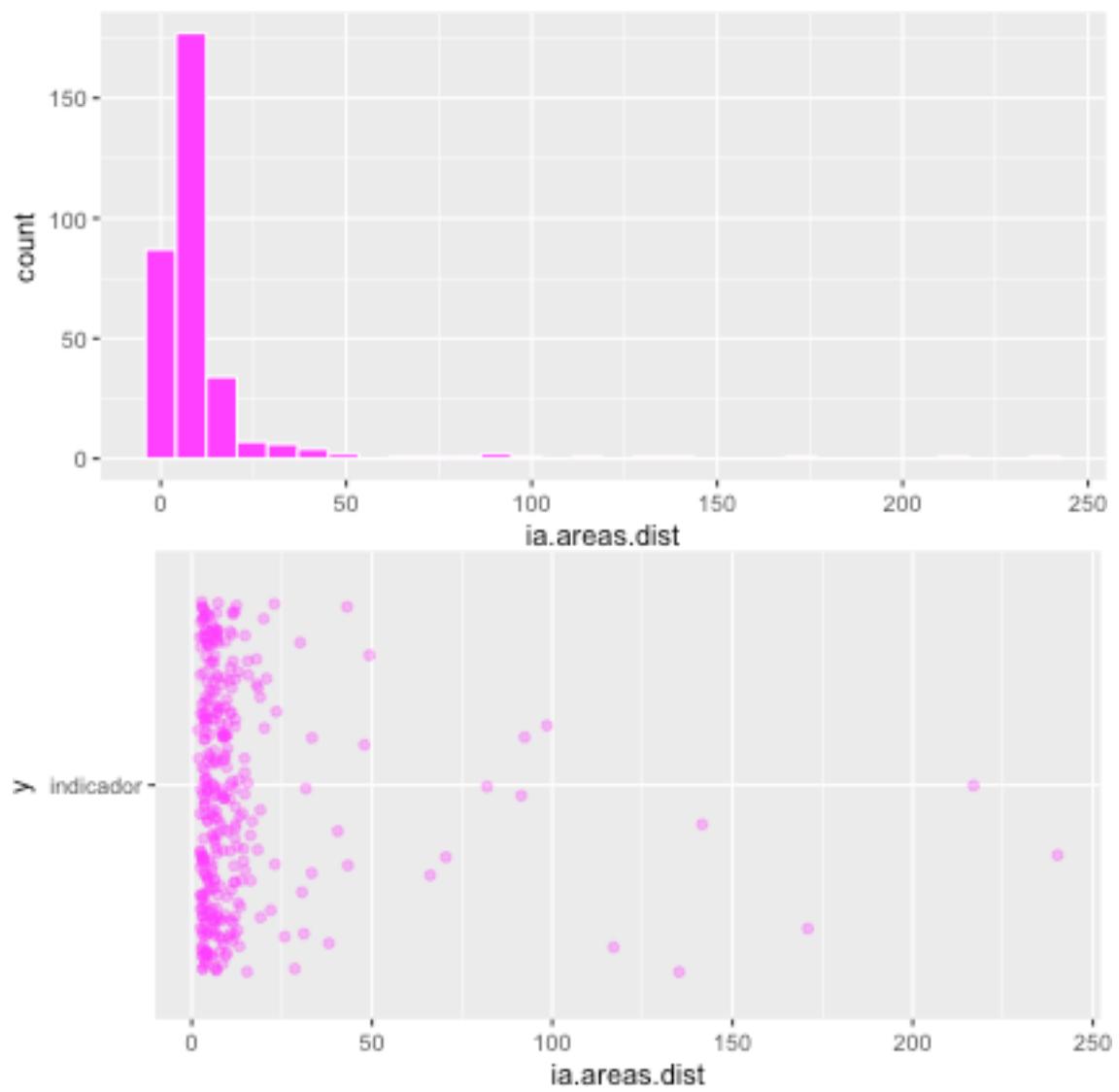


Figure 5.10: Distribucion de los indicadores de acceso a EV área-distancia

arbóreas.

Mapas LISA - area\_ep.porcentaje - Wq

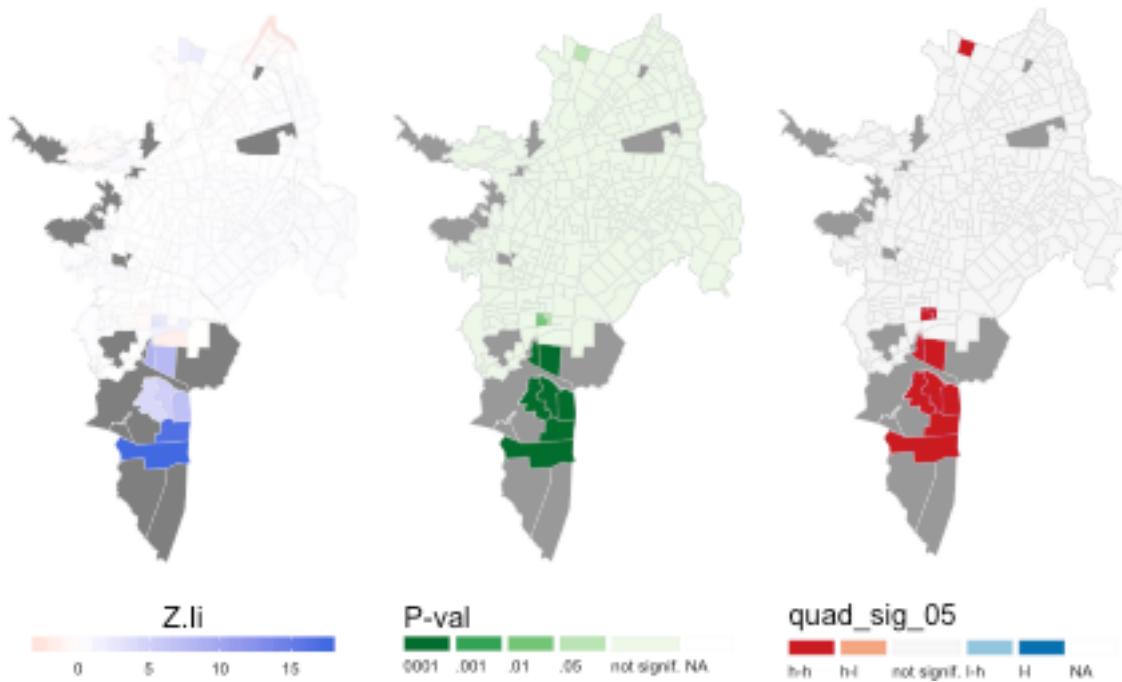
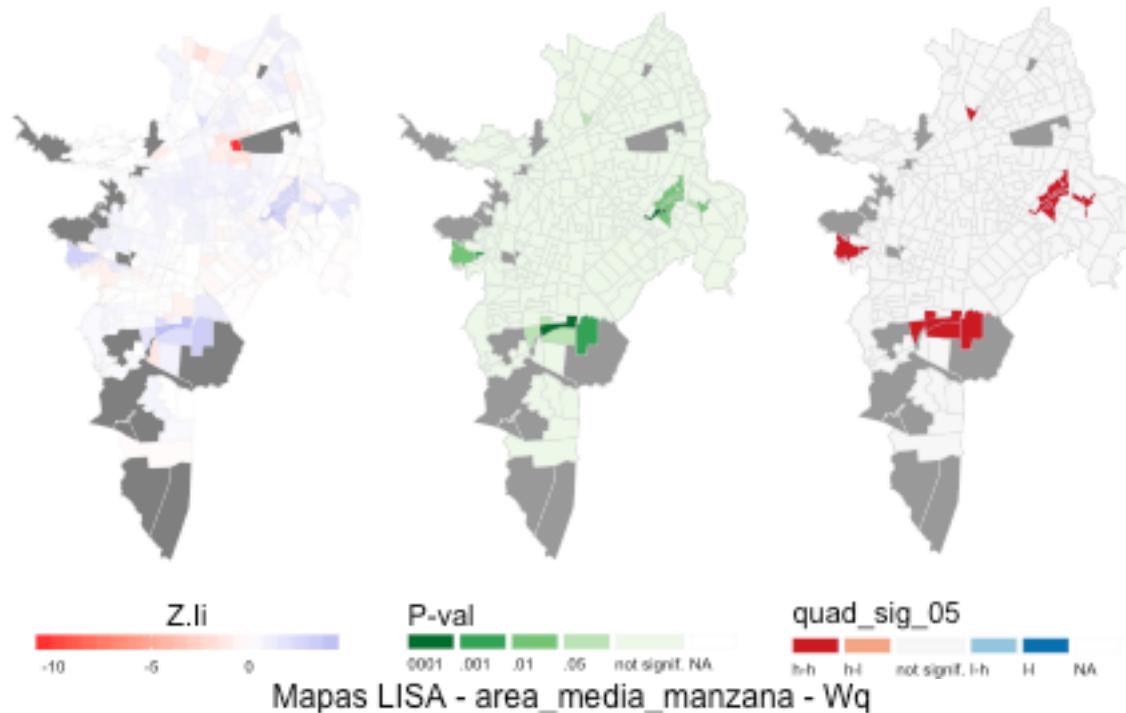
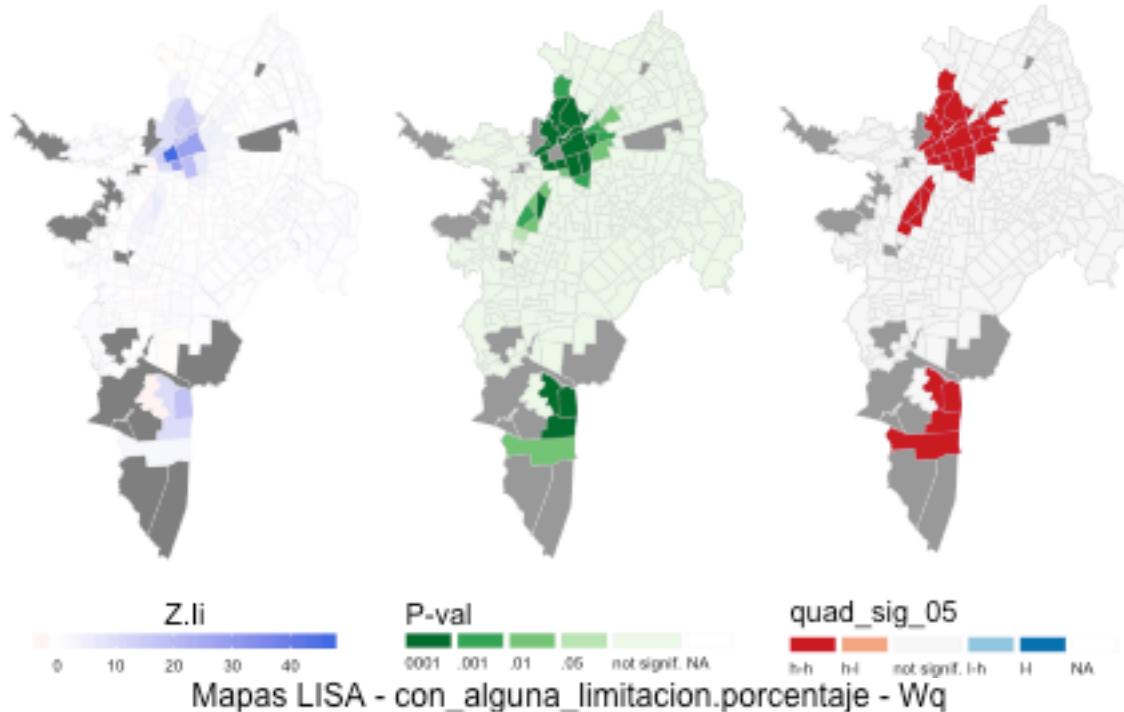


Figure 5.11: Mapas LISA - Porcentaje de espacios verdes y área media de manzana

Mapas LISA - unidad\_economica.porcentaje - Wq



Mapas LISA - con\_alguna\_limitacion.porcentaje - Wq

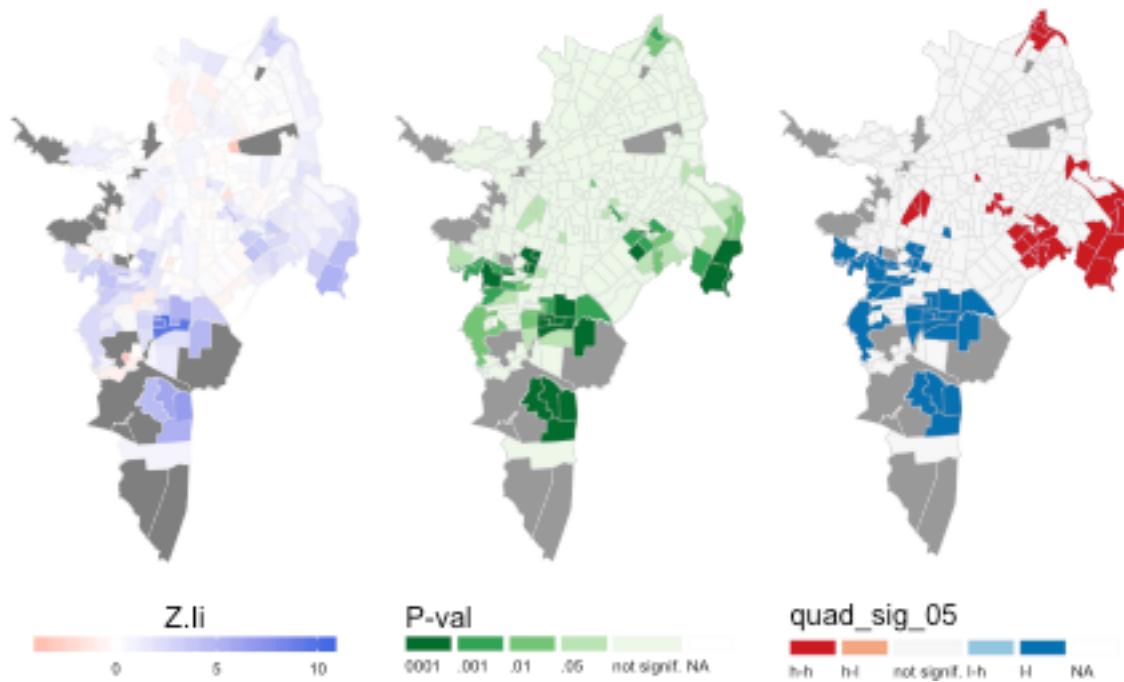


Figure 5.12: Mapas LISA - Porcentaje de unidades economicas y personas con alguna limitación



# Chapter 6

## Conclusiones

Quiero resaltar el uso de herramientas libres como *R*, Rstudio y los diferentes paquetes del ecosistema del CRAN, que favorecen la reproductibilidad y diseminación del conocimiento científico. El esfuerzo y la curva de aprendizaje de herramientas como Bookdown [R-bookdown] y knitr (Xie, 2017) para la generación de los contenidos y textos en diferentes formatos, la reproductibilidad del uso de scripts y la posibilidad de publicar un repositorio del resultado, del proceso y de los datos, no solo vale la pena como base para la construcción de flujos de trabajo eficientes en el manejo y divulgación de datos e información —geográfica—, sino que me parece un compromiso ineludible con la construcción de la ciencia de nuestro tiempo y de la ética investigativa.

En cuanto a lo metodológico en este trabajo se hizo esfuerzos por conjugar la exposición gráfica de los datos crudos y de los resultados estadísticos con la rigurosidad y capacidad explicativa de los modelos de regresión estadísticos y geoestadísticos. Entre las posibilidades que ofrece el código como herramienta cartográfica está la de producir y reproducir múltiples mapas, en lugar de uno solo con mucho detalle como respuestas a las limitaciones tecnológicas y de costos de la cartografía impresa, lo que redunda en mejor información para los análisis. Los métodos gráficos como los mapas de LISA permiten una interpretación más clara de los mecanismos de ajuste de los modelos espaciales y de la

combinación lineal de los términos. Estos mapas junto con los mapas temáticos de cada una de las variables, los gráficos de dispersión de las distribuciones multivariadas son una estrategia eficaz para lidiar con las dificultades para analizar conjuntos de datos multidimensionales y espaciales. Los métodos de ajuste espacial y la comparación entre distintos modelos permitieron seleccionar el resultado más confiable y eficiente de la estimación, dando firmeza a las conclusiones de los análisis.

En relación a lo temático este trabajo demuestra que en lo que respecta al acceso de beneficios del arbolado urbano en Santiago de Cali existe una fuerte relación de la variable de estatus social representada por el acceso a estudios superiores, que explica la variabilidad en los datos y los patrones espaciales de la cobertura de copa. Los modelos de regresión espacial confirman que los beneficios ambientales son mejores en sectores con población mejor educada, posiblemente una razón para preferir o habitar espacios con buena arborización o tal vez con condiciones suficientes de espacio y zonas blandas para arborizar el barrio y propiciar el desarrollo del potencial de los individuos arbóreos. En cualquier caso, esto muestra que los clusters de sectores urbanos con mayor arborización excluyen al grueso de la población afro y con bajos niveles de estudios, como prueban las altas correlaciones negativas entre los valores de estos grupos de población y la cobertura de copa. El mapa de los beneficios ambientales se suma pues a una serie de inequidades encontradas en la literatura que aborda problemas relacionados con la segregación racial y la baja empleabilidad de poblaciones afrocolombianas en la ciudad de Cali. Los casos de análisis espacial en relación al acceso de servicios ambientales fueron descritos ampliamente en la literatura y nos muestran una preocupación creciente sobre la integración y restauración de ecosistemas urbanos, una sofisticación metodológica y fundamentación teórica que fueron claves para aplicar exitosamente el análisis al contexto de Cali. Además coinciden en señalar problemas de segregación e inequidades ambientales en el continente americano.

En cuanto a los espacios verdes el ejercicio no es mucho más alentador. Este trabajo mostró cómo existe una importante cantidad de sectores urbanos con menos del 5% del área del

sector de espacio verde: una alta concentración del área disponible se encuentra unos pocos sectores, que sirven para suplir las carencias de espacio verde en sus vecinos. Aunque el coeficiente de la dimensión poblacional representada por porcentaje de personas con alguna limitación física o mental, es bajo, coincide espacialmente con las zonas de bajo acceso a estos espacios. Sin embargo en la medición del acceso como un beneficio local las variables poblacionales no pasaron los criterios de selección para ser candidatas en el modelo de regresión. El área media de manzana de un sector fue muy representativa y significativa de la variabilidad de los datos, y señala indirectamente la responsabilidad de las autoridades y los urbanizadores que deben garantizar la provisión de los equipamientos de ciudad, los tamaños mínimos y deseados, y en general velar por el desarrollo la estructura ecológica del municipio y del casco urbano.

Analizar varios indicadores con variaciones en el cálculo y en las variables incorporadas en su definición permite construir una visión más compleja y detallada del acceso. Este trabajo se limitó a analizar 2 indicadores por beneficio ambiental, explorando las diferencia entre indicadores porcentuales que representan las diferencias entre medir el beneficio en el área total de un sector o solo sobre su espacio público, para establecer una matiz entre lo que compete a las autoridades y el alcance de un beneficio que se irradia más allá de las fronteras de las calles y los límite administrativos. También explora dos experiencias distintas del acceso a espacio verdes, la local, al interior de un sector urbano y otra en la escala de las agrupaciones de sectores. Resultado de esta se puede pensar explorar complejizar estos indicadores para incorporar aspectos de la calidad de los espacio o incluir dimensiones poblacionales como la densidad o normalizarlos con respecto del número de habitantes.

Más interesante puede ser contar con dos conjuntos de datos separados en el tiempo tanto de las variables poblacionales y del arbolado urbano, para indagar sobre cómo la matriz de vecindad puede capturar las variaciones de la estructura de las características poblacionales de los individuos humanos y arbóreos; o analizar el impacto de la inclusión de nuevos espacio verdes a la topología de beneficios y estimar cambios en la estructura

de grupos de sectores urbanos con acumulacion de desvenajas, por ejemplo.

Esto me lleva al último punto de las conclusiones y el primer momento dentro del proceso de esta investigación: la disponibilidad de los datos. Sin la existencias de los servicios de información geográfica de la Alcaldía de Cali ni la disposición de la cartografía del censo y los datos agregados a nivel de sector urbano del DANE, o la existencia de un censo arbóreo es imposible progresar en el entendimiento de nuestro desarrollo como ciudad, región o nación. Sin embargo, aún son escasas las variables del censo de población que están disponibles a este nivel, no existe un repositorio público de los datos del censo arbóreo, ni de las acciones de mantenimiento y estado del arbolado urbano a pesar de que no existe ningún impedimento técnico ni legal para que pueda estarlo. Los resultados dependen pues no solo de la disponibilidad sino de la calidad y nivel de actualización de los datos, que pueden y deben mejorar en los servicios de información que presta el el municipio.

# Bibliography

Geoportal DANE.

Geoportal IDESC.

MAGNA-SIRGAS-CALI.

Plan de Ordenamiento Territorial - POT año 2014.

Sistema de consulta censo 2005.

(2013). Estatuto Arbóreo de Santiago de Cali-Acuerdo No 0353.

(2015). Informes del censo arbóreo.

Alanís, E., Jiménez, J., MORA, O., Canizales, P., and Rocha, L. (2014). Estructura y composición del arbolado urbano de un campus universitario del norte de México. *Revista Iberoamericana de Ciencias*, 1(7):93–101.

Anselin, L. (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural economics*, 27(3):247–267.

Arroyo Mina, J. S., Pinzón Gutiérrez, L. F., Mora, J. J., Gómez Jaramillo, D. A., and Cendales, A. (2016). Afrocolombianos, discriminación y segregación espacial de la calidad del empleo para Cali. *Cuadernos de Economía*, 35(69):753–783.

Azócar, G., Romero, H., Sanhueza, R., Vega, C., Aguayo, M., and Muñoz, M. D. (2007). Urbanization patterns and their impacts on social restructuring of urban space in

- Chilean mid-cities: The case of Los Angeles, Central Chile. *Land Use Policy*, 24(1):199–211.
- Bivand, R. (2017). *spdep: Spatial Dependence: Weighting Schemes, Statistics and Models*. R package version 0.7-4.
- Bivand, R., Keitt, T., and Rowlingson, B. (2017). *rgdal: Bindings for the 'Geospatial' Data Abstraction Library*. R package version 1.2-16.
- Bivand, R. and Rundel, C. (2017). *rgeos: Interface to Geometry Engine - Open Source ('GEOS')*. R package version 0.3-26.
- Bolund, P. and Hunhammar, S. (1999). Ecosystem services in urban areas. *Ecological economics*, 29(2):293–301.
- Boone, C. G., Cadenasso, M. L., Grove, J. M., Schwarz, K., and Buckley, G. L. (2010). Landscape, vegetation characteristics, and group identity in an urban and suburban watershed: why the 60s matter. *Urban Ecosystems*, 13(3):255–271.
- Braverman, I. (2008). Everybody loves trees: Policing American cities through street trees. *Duke Envtl. L. & Pol'y F.*, 19:81.
- Breusch, T. S. and Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, pages 1287–1294.
- CEPAL. REDATAM.
- Cerdà, M. O. (2011). Origen y evolución del movimiento de justicia ambiental. *Ecología política*, (41):17–24.
- Cerón, W. L. and Escobar, Y. C. (2014). Índice de segregación espacial y socioeconómico (ises) en las comunas de Santiago de Cali. *Cuadernos de Vivienda y Urbanismo*, 7(13).

Chakraborty, J. and Armstrong, M. P. (1997). Exploring the use of buffer analysis for the identification of impacted areas in environmental equity assessment. *Cartography and Geographic Information Systems*, 24(3):145–157.

Chapman, C. A. and Onderdonk, D. A. (1998). Forests without primates: primate/plant codependency. *American Journal of Primatology*, 45(1):127–141.

CIAT (2015). Identificación de Zonas y Formulación de Propuestas para el Tratamiento de Islas de Calor.

Comber, A., Brunsdon, C., and Green, E. (2008). Using a GIS-based network analysis to determine urban greenspace accessibility for different ethnic and religious groups. *Landscape and Urban Planning*, 86(1):103–114.

Comber, A. J., Brunsdon, C., and Radburn, R. (2011). A spatial analysis of variations in health access: linking geography, socio-economic status and access perceptions. *International Journal of Health Geographics*, 10:44.

Cowett, F. D. (2014). Methodology for Spatial Analysis of Municipal Street Tree Benefits. *Arboriculture & Urban Forestry*, 40(2).

Cutter, S. L., Holm, D., and Clark, L. (1996). The role of geographic scale in monitoring environmental justice. *Risk Analysis*, 16(4):517–526.

DANE (2005). R+::CEPAL/CELADE - R+SP WebServer.

Dobbs, C., Escobedo, F. J., and Zipperer, W. C. (2011). A framework for developing urban forest ecosystem services and goods indicators. *Landscape and Urban Planning*, 99(3–4):196–206.

Durán Rivera, B. and Alzate Guarín, F. (2009). Intercepción de partículas suspendidas totales (PST) por cinco especies de árboles urbanos en el Valle de Aburrá.

Ferro Medina, G. (2010). *Árboles ciudadanos, en la memoria y en el paisaje cultural de Bogotá*. Bogotá: Instituto Distrital de Patrimonio Cultural.

- Fotheringham, A. S., Charlton, M. E., and Brunsdon, C. (1998). Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and planning A*, 30(11):1905–1927.
- Garzón, B., Brañes, N., Abella, M. L., and Auad, A. (2004). Vegetación urbana y hábitat popular: el caso de san miguel de tucumán. *Revista invi*, 18(49).
- Gibbons, S. and Overman, H. G. (2012). Mostly pointless spatial econometrics? *Journal of Regional Science*, 52(2):172–191.
- Gómez-Baggethun, E. and Barton, D. N. (2013). Classifying and valuing ecosystem services for urban planning. *Ecological Economics*, 86:235–245.
- Heynen, N., Perkins, H. A., and Roy, P. (2006). The political ecology of uneven urban green space the impact of political economy on race and ethnicity in producing environmental inequality in Milwaukee. *Urban Affairs Review*, 42(1):3–25.
- Kabisch, N. and Haase, D. (2014). Green justice or just green? Provision of urban green spaces in Berlin, Germany. *Landscape and Urban Planning*, 122:129–139.
- Killicoat, P., Puzio, E., and Stringer, R. (2002). The economic value of trees in urban areas: estimating the benefits of Adelaide's street trees. *TREENET*, page 90.
- Kissling, W. D. and Carl, G. (2008). Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography*, 17(1):59–71.
- Konijnendijk, C., Gauthier, M., and Van Veenhuizen, R. (2005). Árboles y Ciudades Creciendo Juntos. *Revista Agricultura Urbana*, 13:1–7.
- Landry, S. M. and Chakraborty, J. (2009). Street trees and equity: evaluating the spatial distribution of an urban amenity. *Environment and Planning A*, 41(11):2651–2670.
- Laredo, D. R. and Mirtha, D. (2011). La gestión del verde urbano como un criterio de mitigación y adaptación al cambio climático. *Revista de la Asociación Argentina de Ecología de Paisajes*, 2(2):123–130.

Leff, E. (2012). Pensamiento ambiental latinoamericano: patrimonio de un saber para la sustentabilidad. *Environmental Ethics*, 34(Supplement):97–112.

LeSage, J. P. and Pace, R. K. (2014). The Biggest Myth in Spatial Econometrics. *Econometrics*, 2(4):217–249.

León Calle, S. and others (2011). *Árboles, simbolismo, cultura, memoria e identidad. Representaciones en el paisaje arbóreo de Gualaquiza*. PhD thesis, Universidad Andina Simón Bolívar, Sede Ecuador.

Low, S. (2013). Public space and diversity: Distributive, procedural and interactional justice for parks. *The Ashgate Research companion to planning and culture*. Surrey, UK: Ashgate Publishing, pages 295–310.

Martinez Alier, J., Anguelovski, I., Bond, P., Del Bene, D., Demaria, F., Gerber, J.-F., Greyl, L., Haas, W., Healy, H., Marín-Burgos, V., and others (2014). Between activism and science. *Journal of Political Ecology*, 21:19–60.

McPherson, E. G., Nowak, D., Heisler, G., Grimmond, S., Souch, C., Grant, R., and Rowntree, R. (1997). Quantifying urban forest structure, function, and value: the Chicago Urban Forest Climate Project. *Urban ecosystems*, 1(1):49–61.

McPherson, E. G., Xiao, Q., and Aguaron, E. (2013). A new approach to quantify and map carbon stored, sequestered and emissions avoided by urban forests. *Landscape and Urban Planning*, 120:70–84.

Mincey, S. K., Hutten, M., Fischer, B. C., Evans, T. P., Stewart, S. I., and Vogt, J. M. (2013). Structuring institutional analysis for urban ecosystems: A key to sustainable urban forest management. *Urban Ecosystems*, 16(3):553–571.

Mora, J. J. and Arcila, A. M. (2014). Brechas salariales por etnia y ubicación geográfica en Santiago de Cali. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 18:34–53.

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2):17–23.

Nelson, E., Mendoza, G., Regetz, J., Polasky, S., Tallis, H., Cameron, D., Chan, K., Daily, G. C., Goldstein, J., Kareiva, P. M., and others (2009). Modeling multiple ecosystem services, biodiversity conservation, commodity production, and tradeoffs at landscape scales. *Frontiers in Ecology and the Environment*, 7(1):4–11.

Nesbitt, L. and Meitner, M. J. (2016). Exploring Relationships between Socioeconomic Background and Urban Greenery in Portland, OR. *Forests*, 7(8):162.

Nolazco, S. (2012). Diversidad de aves silvestres y correlaciones con la cobertura vegetal en parques y jardines de la ciudad de Lima. *Boletín Informativo UNOP*, 7(1):4–16.

Nowak, D. J. and Crane, D. E. (2000). The Urban Forest Effects (UFORE) Model: quantifying urban forest structure and functions.

Nowak, D. J. and Crane, D. E. (2002). Carbon storage and sequestration by urban trees in the USA. *Environmental pollution*, 116(3):381–389.

Nowak, D. J. and Greenfield, E. J. (2012). Tree and impervious cover change in US cities. *Urban Forestry & Urban Greening*, 11(1):21–30.

Nowak, D. J., Stein, S. M., Randler, P. B., Greenfield, E. J., Comas, S. J., Carr, M. A., and Alig, R. J. (2010). Sustaining America's urban trees and forests: a Forests on the Edge report.

Osorio, J. and Molina, L. (2009). A vuelo de pájaro Las ciudades como refugio para las aves. *Revista nodo*, 4(7).

Pebesma, E. and Bivand, R. (2017). *sp: Classes and Methods for Spatial Data*. R package version 1.2-5.

Phelps, J. B. (2012). *The association between tree canopy cover and socio-demographics in Lubbock, Texas*. PhD thesis, Texas Tech University.

- Ponce Donoso, M. and Piedrahita, P. (2009). Valoración económica del arbolado urbano en 28 comunas de Chile. *Quebracho. Revista de Ciencias Forestales*, 17(1-2):88–100.
- Páez, A. and Scott, D. M. (2005). Spatial statistics for urban analysis: a review of techniques with examples. *GeoJournal*, 61(1):53–67.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Restrepo, O., Iván, H., Moreno, H., Hoyos, E., Helena, C., and others (2015). Incidence of progressive deterioration of urban trees in the Aburrá valley, Colombia. *Colombia Forestal*, 18(2):225–240.
- Ripoll, M. V., Kurbán, A., Papparelli, A., Cúnsulo, M., and Roca, G. (2010). Condiciones térmicas de un espacio verde urbano en clima árido. *Avances en Energías Renovables y Medio Ambiente*, 14:11–09.
- Schlosberg, D. (2013). Theorising environmental justice: the expanding sphere of a discourse. *Environmental Politics*, 22(1):37–55.
- Schwarz, K., Fragkias, M., Boone, C. G., Zhou, W., McHale, M., Grove, J. M., O’Neil-Dunne, J., McFadden, J. P., Buckley, G. L., Childers, D., Ogden, L., Pincetl, S., Pataki, D., Whitmer, A., and Cadenasso, M. L. (2015). Trees Grow on Money: Urban Tree Canopy Cover and Environmental Justice. *PLoS ONE*, 10(4).
- Shanahan, D. F., Lin, B. B., Gaston, K. J., Bush, R., and Fuller, R. A. (2014). Socio-economic inequalities in access to nature on public and private lands: a case study from Brisbane, Australia. *Landscape and Urban Planning*, 130:14–23.
- Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611.
- Talen, E. and Anselin, L. (1998). Assessing spatial equity: an evaluation of measures of accessibility to public playgrounds. *Environment and Planning a*, 30(4):595–613.

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic geography*, 46(sup1):234–240.

Tovar-Corso, G. (2013). Aproximación a la silvicultura urbana en Colombia. *Bitácora Urbano Territorial*, 1(22).

Tratalos, J., Fuller, R. A., Warren, P. H., Davies, R. G., and Gaston, K. J. (2007). Urban form, biodiversity potential and ecosystem services. *Landscape and urban planning*, 83(4):308–317.

Troy, A. R., Grove, J. M., O’Neil-Dunne, J. P., Pickett, S. T., and Cadenasso, M. L. (2007). Predicting opportunities for greening and patterns of vegetation on private urban lands. *Environmental management*, 40(3):394–412.

Vásquez, . (1990). Historia del desarrollo económico y urbano en Cali”. *Boletín socioeconómico*, 20:1–28.

Vásquez Fuentes, A. and Romero Aravena, H. (2008). Vegetación urbana y desigualdades socio-económicas en la comuna de Peñalolen, Santiago de Chile. Una perspectiva de justicia ambiental.

Watkins, S. L., Mincey, S. K., Vogt, J., and Sweeney, S. P. (2016). Is Planting Equitable? An Examination of the Spatial Distribution of Nonprofit Urban Tree-Planting Programs by Canopy Cover, Income, Race, and Ethnicity. *Environment and Behavior*, page 0013916516636423.

Wikipedia (2017). Indicadores de asociación espacial — wikipedia, la enciclopedia libre. [Internet; descargado 29-septiembre-2017 ].

Wolch, J. R., Byrne, J., and Newell, J. P. (2014). Urban green space, public health, and environmental justice: The challenge of making cities ‘just green enough’. *Landscape and Urban Planning*, 125:234–244.

Xie, Y. (2017). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.17.

Zhou, X. and Kim, J. (2013). Social disparities in tree canopy and park accessibility: A case study of six cities in Illinois using GIS and remote sensing. *Urban forestry & urban greening*, 12(1):88–97.