Exploring Evidence of a Relationship Between Canopy Cover and Zillow Home Value Index During Quarantine

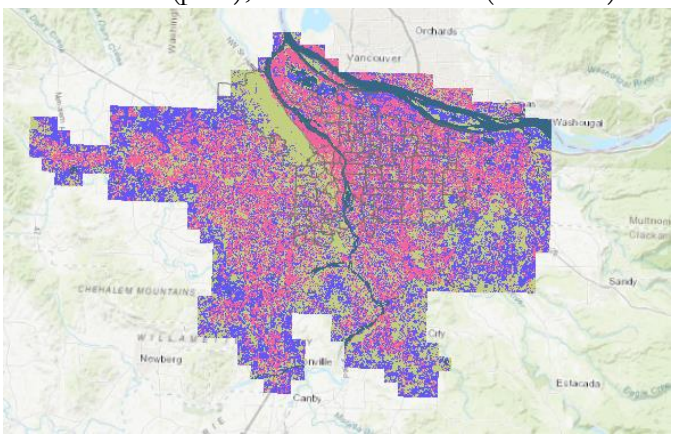
Geography 493 – Ismael Padin – Fall 2020

1. Abstract

The question investigated is 'Have quarantine requirements put a significant financial premium on green space?' As a practical matter, it is quite complex to separate out all the variables influencing home value and sale price; this exploration was examining a linear relationship between Zillow Home Value Index (ZHVI) (by neighborhood) ~ Percent Canopy Coverage (by neighborhood). The null hypothesis is that houses in neighborhoods with more percent canopy coverage do not have a significant difference in their rate of ZHVI appreciation over the months of March – April, 2020 from houses in layers with less canopy coverage. This investigation was unable to find evidence to disprove the Null Hypothesis, but served as a practical exercise in data scrubbing, analysis in GIS, and reformatting for visualization.

2. Data

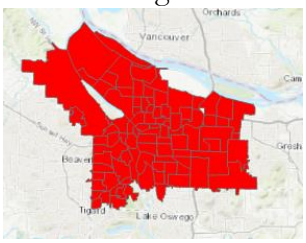
3 meter raster of ground cover in the Portland Metro area, sourced from the PSU I drive. Contains 4 types of ground cover: all canopy (trees, green), low canopy (shrubs and grassland, light blue), built environment (pink), and surface water (dark blue).



Zillow Home Value Index table, downloaded from Zillow, containing records of ZHVI by neighborhood for each month from March 2020 – October 2020.

RegionID	SizeRank	RegionNa	RegionTyp	StateNam	State	City	Metro	CountyNa	1/31/1996	2/29/1996
274772	0	Northeast	Neighborl	TX	TX	Dallas	Dallas-For	Dallas Cou	130854	130902
112345	1	Maryvale	Neighborl	AZ	AZ	Phoenix	Phoenix-M	Maricopa County		
192689	2	Paradise	Neighborl	NV	NV	Las Vegas	Las Vegas-	Clark Cour	136182	136110
270958	3	Upper We	Neighborl	NY	NY	New York	New York-	New York	237777	237402
118208	4	South Los	Neighborl	CA	CA	Los Angel	Los Angel	Los Angel	134690	134560

Portland neighborhood boundaries shapefile, sourced from https://gis-pdx.opendata.arcgis.com/.



3. Methodology

3.1 Scrubbing Data, Part 1

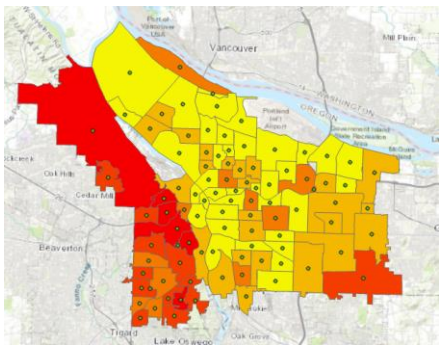
The raw zhvi data was loaded into RStudio and filtered for neighborhoods in the City of Portland, for the dates March – October 2020. It was then visually examined and neighborhood names were compared to those in the neighborhood boundary shapefile, and regular expressions were used to change the names in the zhvi data table to match the shapefile.

3.2 ArcGIS Terrain Analysis

The ground cover raster and neighborhood boundary shapefile were loaded into ArcGIS Pro. Since many boundaries overlap, individual polygons were merged in a way to help neighborhood polygon name attributes match the neighborhood names in the zhvi table. This edited neighborhood boundary shapefile was then converted to a raster, zonal statistics were calculated using the neighborhood raster for zones and the ground coverage as the input feature. This yielded a neighborhood raster whose attribute table contained counts of each type of ground cover by neighborhood – from these values, percentage of area covered by tall canopy was calculated per neighborhood. This tall canopy area percentage was taken as a proxy for neighborhood greenspace.

The zhvi table data was then joined to the neighborhood raster, ultimately making an attribute table that, for each neighborhood, has ground coverage counts, tree canopy coverage as a percent, and a series of home values in dollars for each month from March – October.

The attribute table from this final raster was exported as a csv file for re-formatting and analysis in RStudio.



Neighborhood raster classified according to percentage tree canopy from which a table containing all needed information was exported.

3.3 Reformatting Data

The exported data table contained too much information in too little space: fields containing zhvi data were identified not only by the value in the cell but by the date on which they occurred. In order to use RStudio for visualization, the table was divided into two related tables containing the same information using Neighborhood as the key field.

Data Frame 1: neighborhoods

Neighborhood	Canopy_sqft	Built_sqft	Shrub_Grass_sqft	Surface_water_sqft	Percent_Canopy
FOREST PARK - LINNTON	424008900	26624763	70357716	13800645	0.7928482
CATHEDRAL PARK	5071104	11689686	4345425	7478379	0.1774069
UNIVERSITY PARK	7294446	11752722	8885160	8884269	0.1981293
PIEDMONT	5876406	18472779	7241571	469476	0.1832927

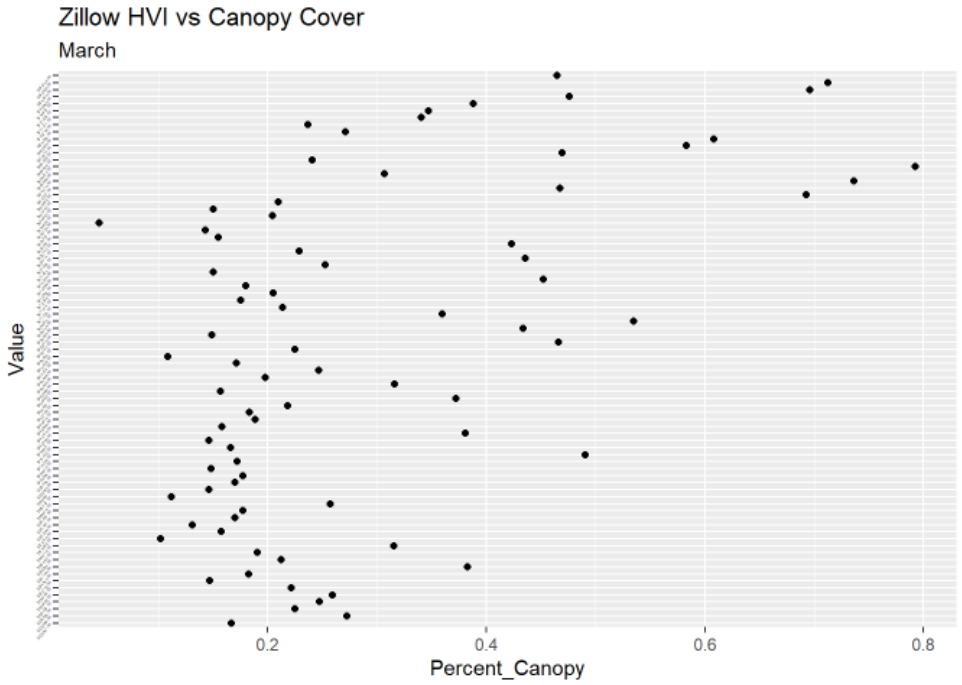
Data Frame 2: zhvi

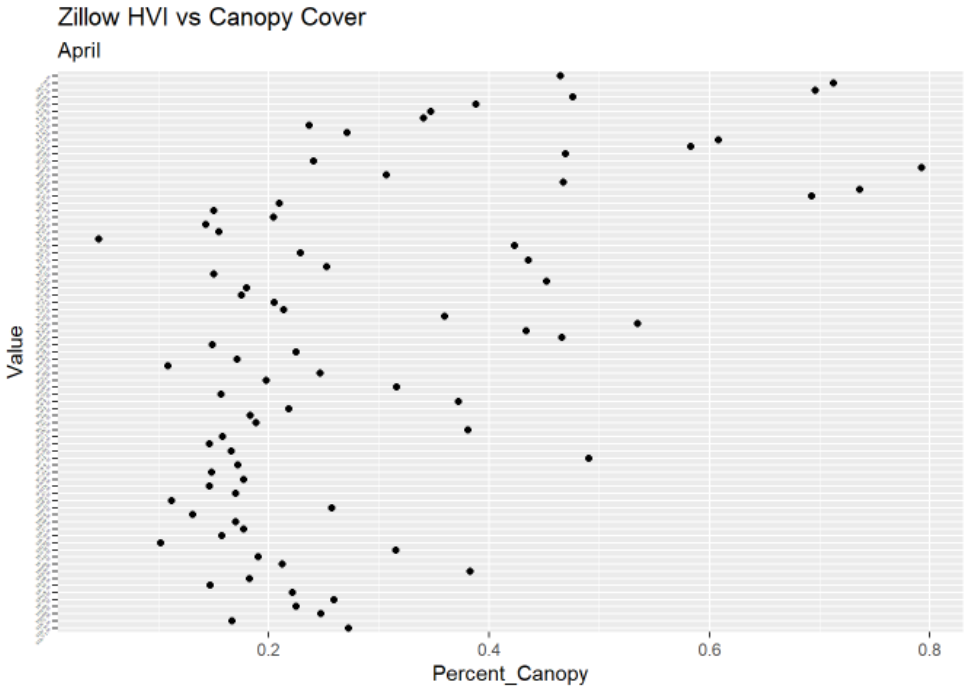
Neighborhood	Date	Value
FOREST PARK - LINNTON	3/30/2020	$590,370.00
FOREST PARK - LINNTON	4/30/2020	$593,415.00
FOREST PARK - LINNTON	5/30/2020	$595,510.00
FOREST PARK - LINNTON	6/30/2020	$598,285.00
FOREST PARK - LINNTON	7/30/2020	$603,192.00
FOREST PARK - LINNTON	8/30/2020	$608,763.00
FOREST PARK - LINNTON	9/30/2020	$616,347.00
FOREST PARK - LINNTON	10/30/2020	$621,098.00
CATHEDRAL PARK	3/30/2020	$397,485.00
CATHEDRAL PARK	4/30/2020	$400,235.00
CATHEDRAL PARK	5/30/2020	$404,026.00

3.4 Visualization

The two source data frames (neighborhoods and zhvi) were loaded into RStudio and joined using the neighborhood name field as the linking key. Percent canopy values were manually binned into equal value interval regions of 20%, and neighborhood home values were plotted against neighborhood canopy coverage percentage for each month of the pandemic.

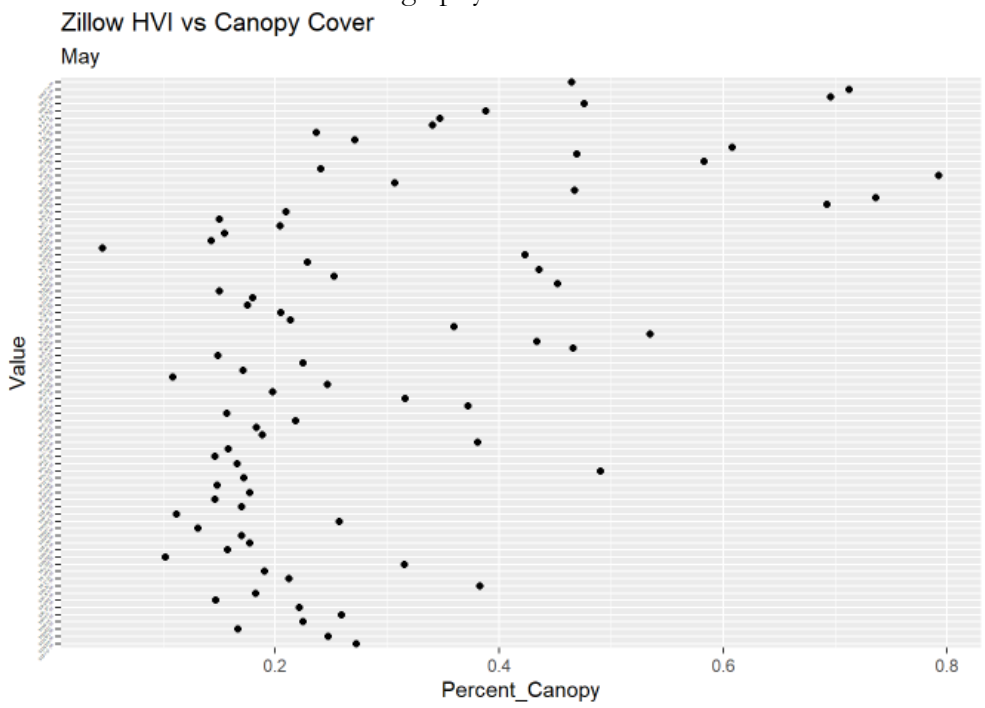
4. Results

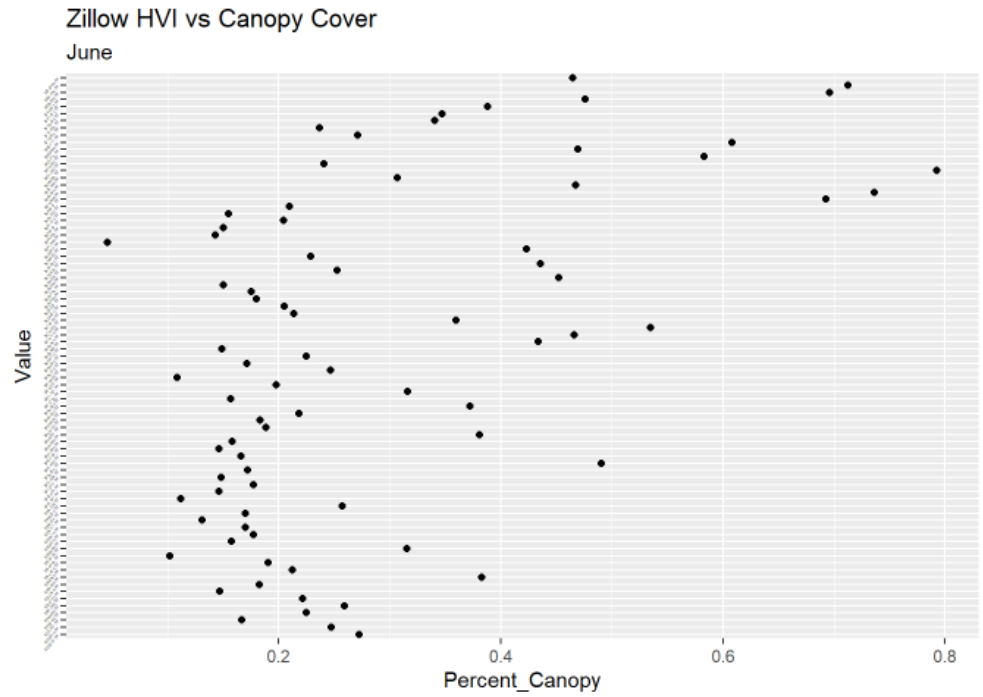

Zillow HVI vs Canopy Cover
March
Value
0.2
0.4
0.6
0.8
Percent_Canopy


Zillow HVI vs Canopy Cover
April
Value
0.2
0.4
0.6
0.8
Percent_Canopy

Exploring Evidence of a Relationship Between Canopy Cover and Zillow Home Value Index During Quarantine

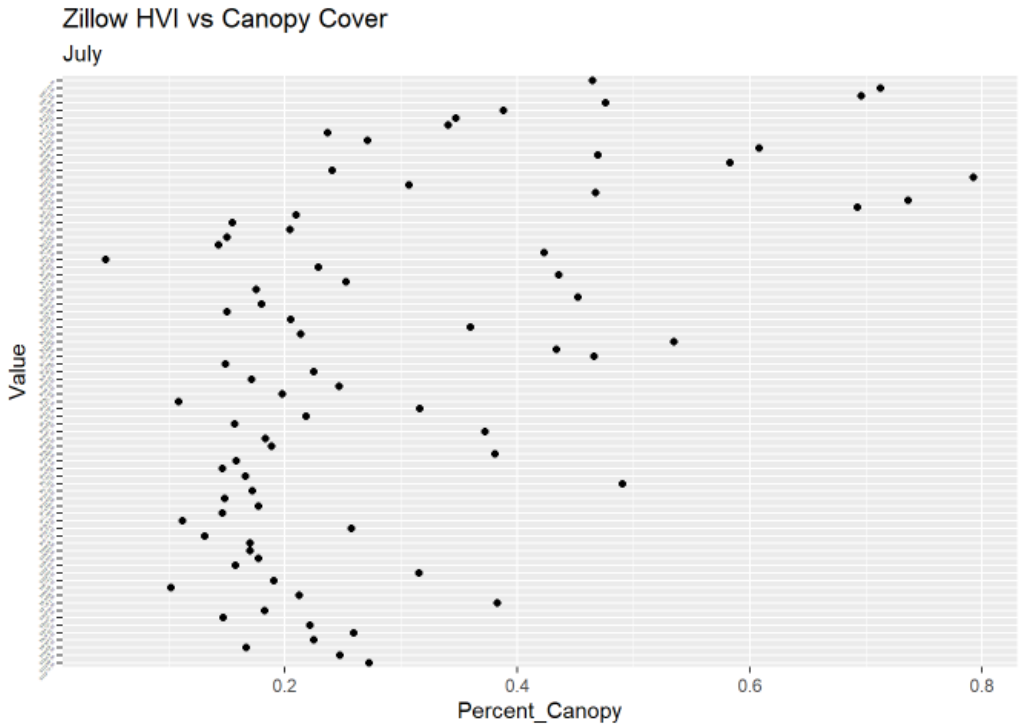
Geography 493 – Ismael Padin – Fall 2020

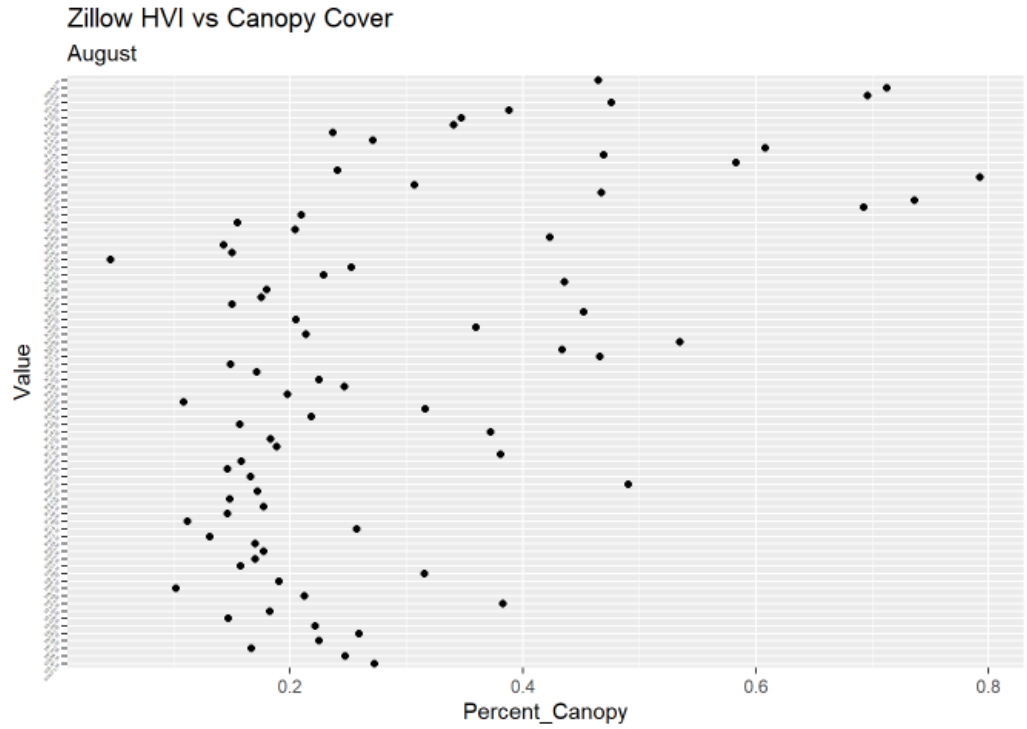




Exploring Evidence of a Relationship Between Canopy Cover and Zillow Home Value Index During Quarantine

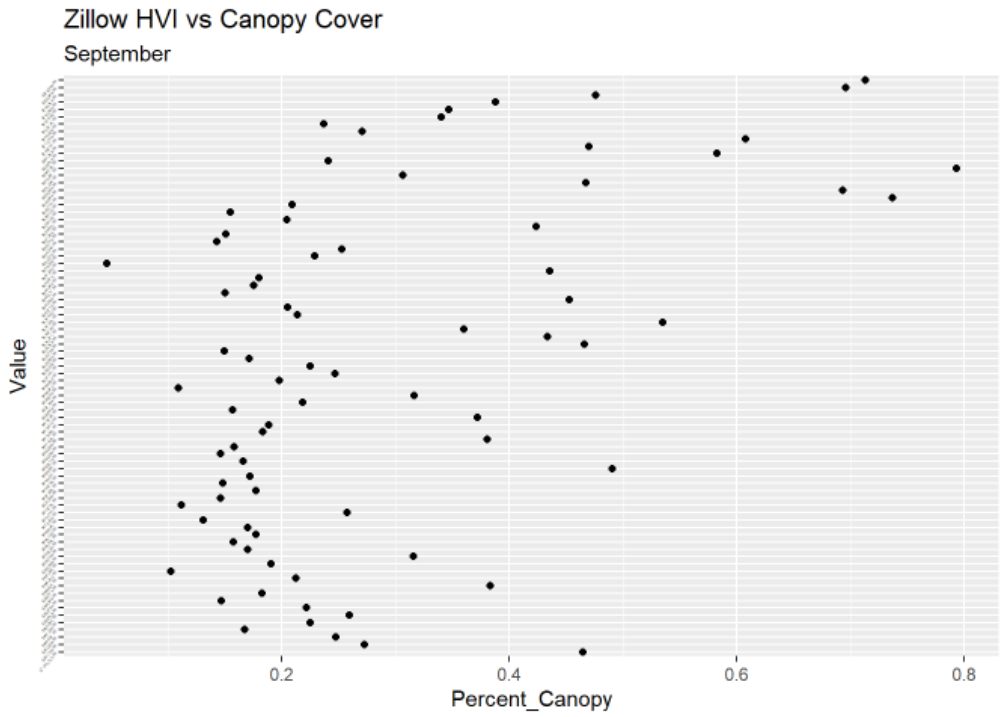
Geography 493 – Ismael Padin – Fall 2020

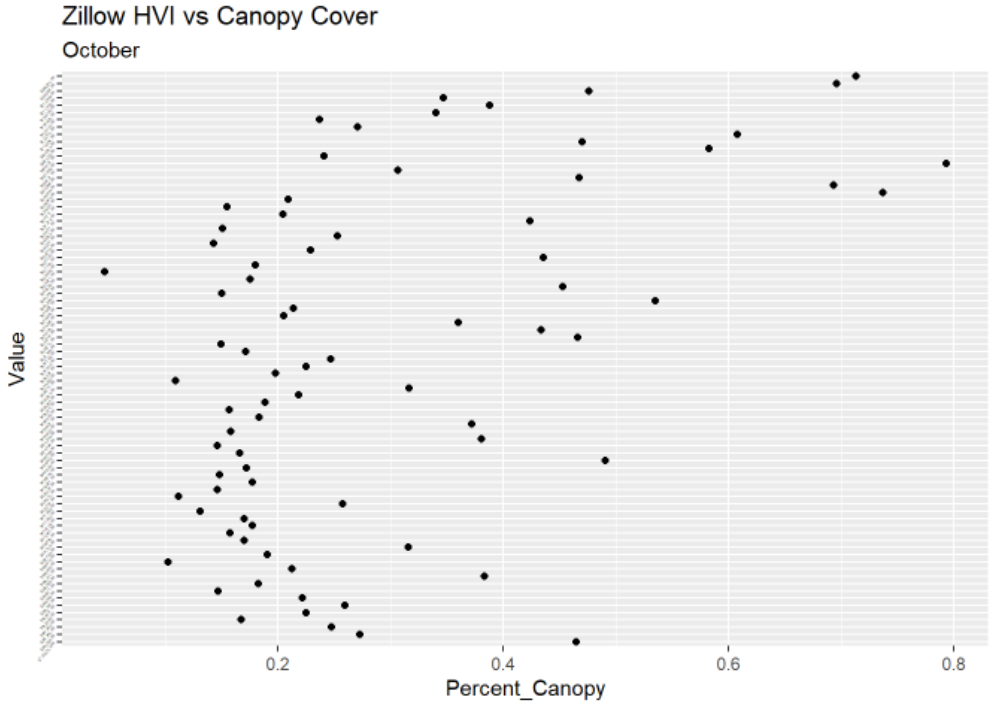




Exploring Evidence of a Relationship Between Canopy Cover and Zillow Home Value Index During Quarantine

Geography 493 – Ismael Padin – Fall 2020





5. Analysis

As can be seen from the figures, there is no significant difference in the rate at which high-canopy neighborhoods appreciated in home value verse the rate at which low-canopy neighborhoods appreciated in home value. There is a clear positive correlation between canopy percentage and magnitude of home value, but without digging deeper and developing a more complex model that takes into account more variables, it is unlikely that a discernible effect on home value from canopy or greenspace during quarantine can be identified.

6. Discussion

6.1 Limitations

There are several limitations to this exploration. Chief among these are: the Zillow Housing Value Index itself (complex, but likely uses greenspace as one of its linearly summed model components), and the implicit assumption of correlation between neighborhood value index as a whole with neighborhood greenspace. Since canopy coverage and greenspace vary throughout a neighborhood, the overall area percentage of canopy cover may not actually reflect greenspace next to individual houses; further, comparing between neighborhoods of different area is complicated by the lack of normalization.

6.2 Improvements

This exploration, while inconclusive, provides a roadmap for further analysis that may be able to parse out a correlation between greenspace and housing value. By including 'low canopy', a more representative metric for actual neighborhood greenspace could be used. Further, by comparing rate of home value increase for those same months in previous years, the coronavirus quarantine effect could be better isolated and other factors (summer price increases, noise, etc) could be controlled for better.

Normalization can be achieved by doing deeper statistical analysis of neighborhood canopy area coverage and returning a mean canopy coverage area and standard deviation, against which the relative 'weight' of a neighborhood's canopy area can be identified along a bell curve. This would make comparison between neighborhoods more of an apples-to-apples affair.