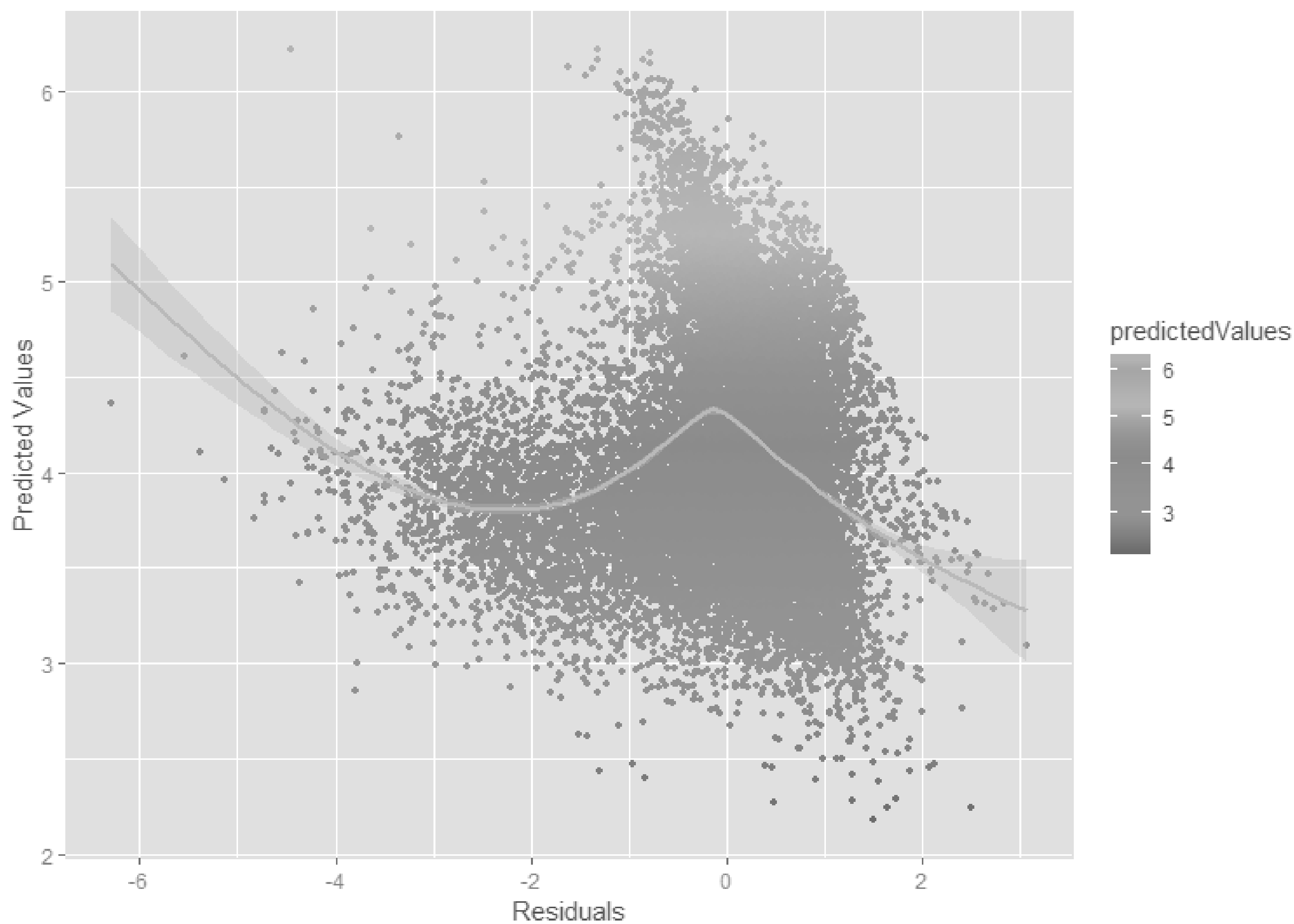
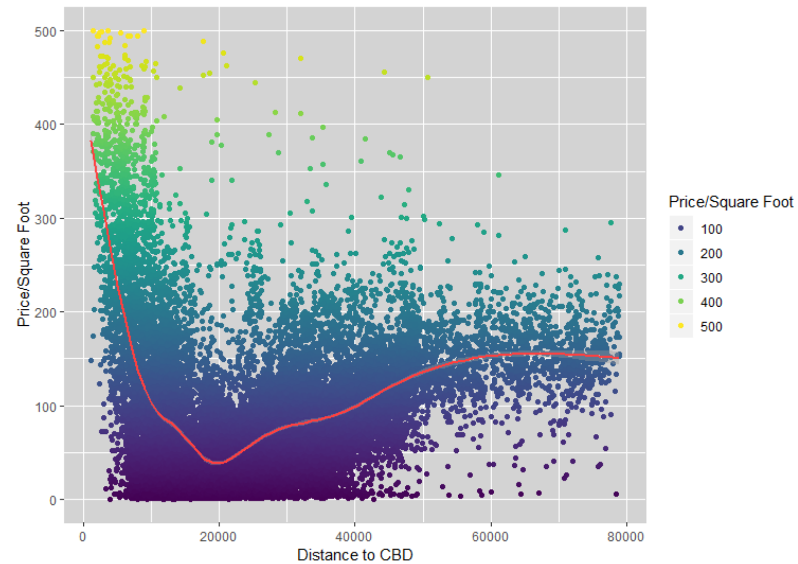


HW7 - Willingness to Pay for Transit

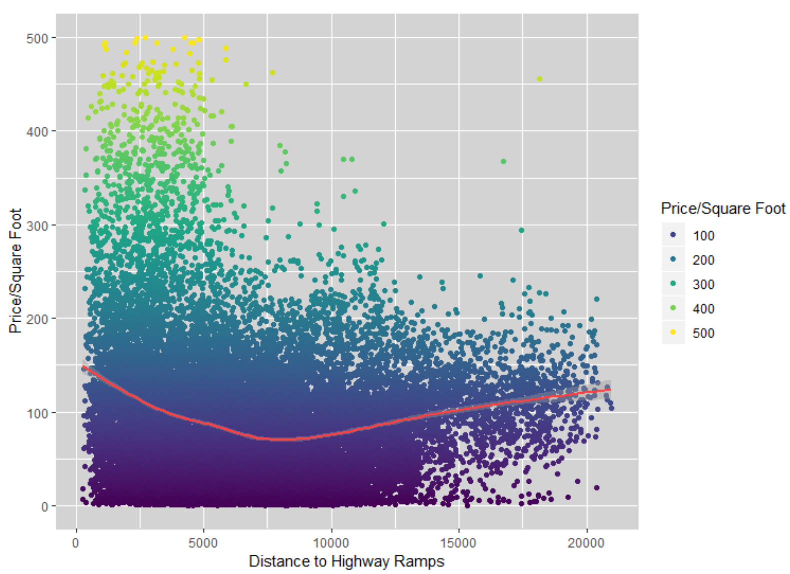
MUSA 507
Spatial Analysis for Urban & Environmental Planning
Jiaxin, WU

The relationship between Home Prices & Expainable Variables

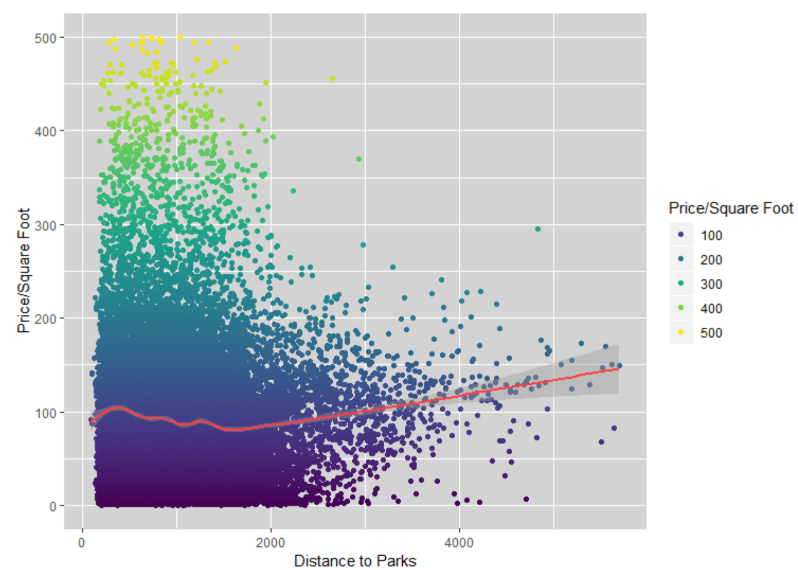
Home price goes up as the distance to crimes increases. This trend is condierably obvious among areas closest to crimes and becomes flat gradually.



The highest home prices cluster in areas close to CBD and sharply decrease as distance to CBD increases. However, home price again rises after a inflection potin, that is, 20000 units of distance.



The home price goes down and up again as the distance to highway ramps increases, most of which range from 75 to 150 dollars per square foot.

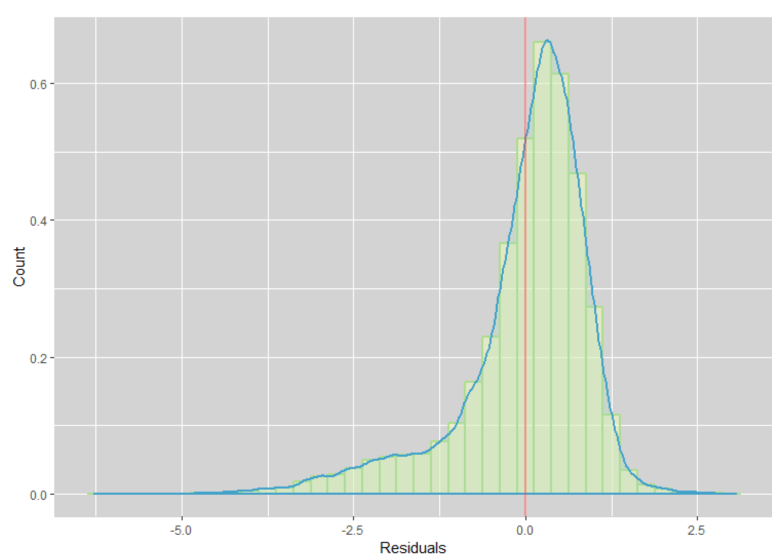


The home price slightly fluctuates around 100 dollars per square foot as distance to parks increases, most of which cluster within 2000 units of distance.

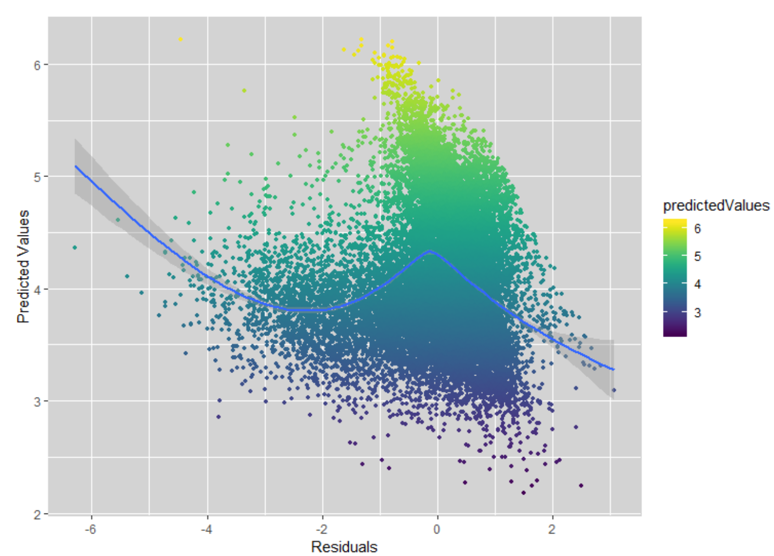
Final Kitchen Sink Regression Model

```
Call:
lm(formula = log(inf_prc_ft) ~ d_septa + log(d_parks) + log(d_crime) +
    log(d_business) + log(d_cbd) + log(d_h_ramps), data = hed)

Residuals:
    Min      1Q  Median      3Q     Max
-6.2860 -0.3174  0.2013  0.5886  3.0602

Coefficients:
                    Estimate   Std. Error t value              Pr(>|t|)
(Intercept)      3.924390575  0.164239347  23.894 < 0.0000000000000002 ***
d_septa          0.000018185  0.000001243  14.631 < 0.0000000000000002 ***
log(d_parks)    -0.043521615  0.010288564  -4.230            0.0000234 ***
log(d_crime)     0.719703860  0.010967025  65.624 < 0.0000000000000002 ***
log(d_business) -0.243460912  0.014736697 -16.521 < 0.0000000000000002 ***
log(d_cbd)      -0.251856050  0.014037916 -17.941 < 0.0000000000000002 ***
log(d_h_ramps)  -0.032266125  0.010352429  -3.117              0.00183 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9384 on 23251 degrees of freedom
Multiple R-squared:  0.2748,    Adjusted R-squared:  0.2746
F-statistic:  1468 on 6 and 23251 DF,  p-value: < 0.00000000000000022
```

In the final ***Kitchen Sink Regression Model***, 6 variables are kept, including d_septa and logs of d_parks, d_crime, d_business, d_cbd and d_h_ramps. Ech log-transformed variable are more normally distributed than before, for the purpose that the residuals of the regression model are more likely to be normal. The estimate coefficient of d_septa is 0.000018185, which means as the distance to SEPTA goes up with one unit, home price in the form of dollars per square foot increases $\mathbf{(e^{0.000018185} - 1) \times 100\% \approx 0.18\%}$.

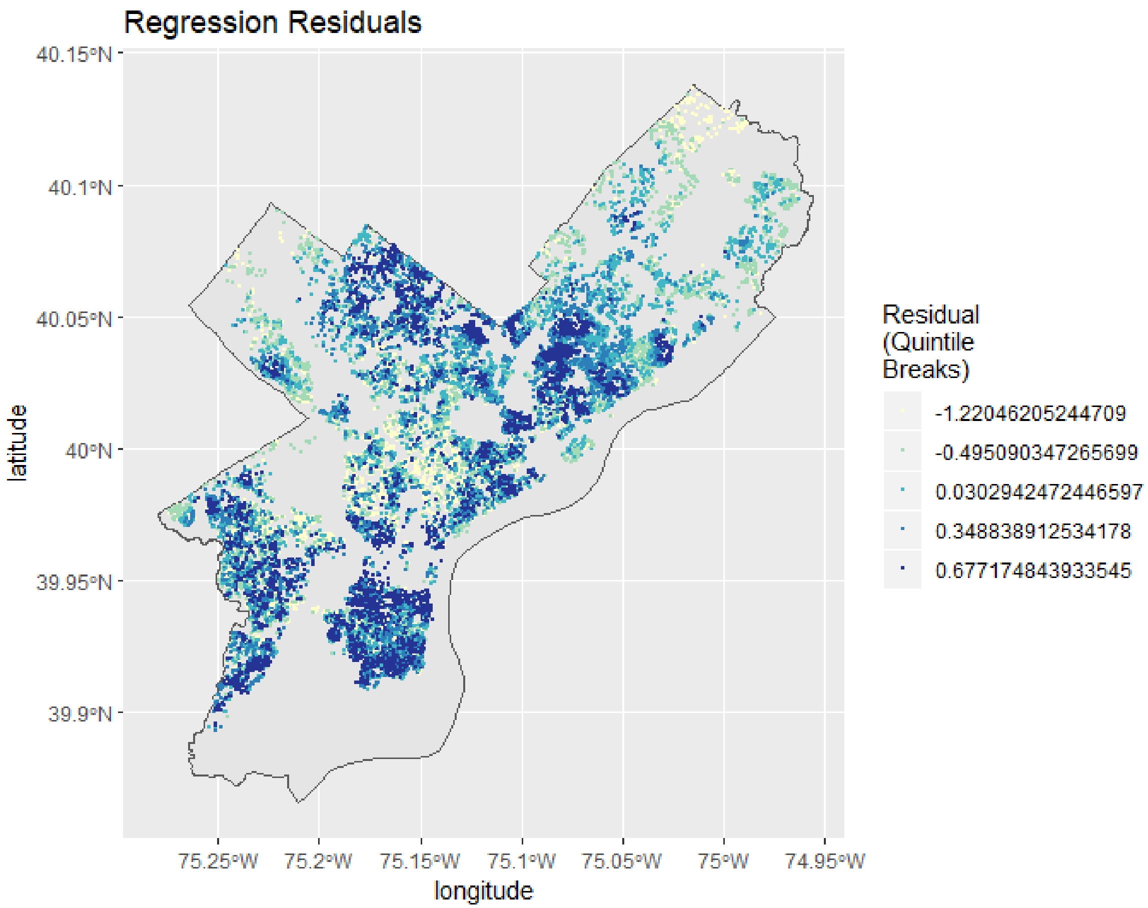


The histogram indicates that final ketchen sink regression model follows the assumption that residuals are normally distributed generally.



The residuals of the final kitchen sink regression model is not random. There is a cusp in the fitted curve when residuals equal to 0 approximately, which indicates that residuals correlate to high predicted values while they should have distribute more randomly and evenly no matter high or low

Residual Map & Moran's I Statistic




```
        Moran I test under randomisation

data:  regF$residuals
weights: nb2listw(spatialWeights, style = "W")

Moran I statistic standard deviate = 82.995, p-value < 0.00000000000000022
alternative hypothesis: greater
sample estimates:
Moran I statistic       Expectation          Variance
    0.36004209178    -0.00004299781     0.00001882392
```

Clustered Residuals

Through observation of the residuals map, I find that residuals are spatially clusted in some specific areas, which violates the assumption that residuals should be random. It may result from the problems of the model inherent. For instance, the relationship between dependent and independent variables is not linear. Or some variables that contribute to the dependent variable are missing.

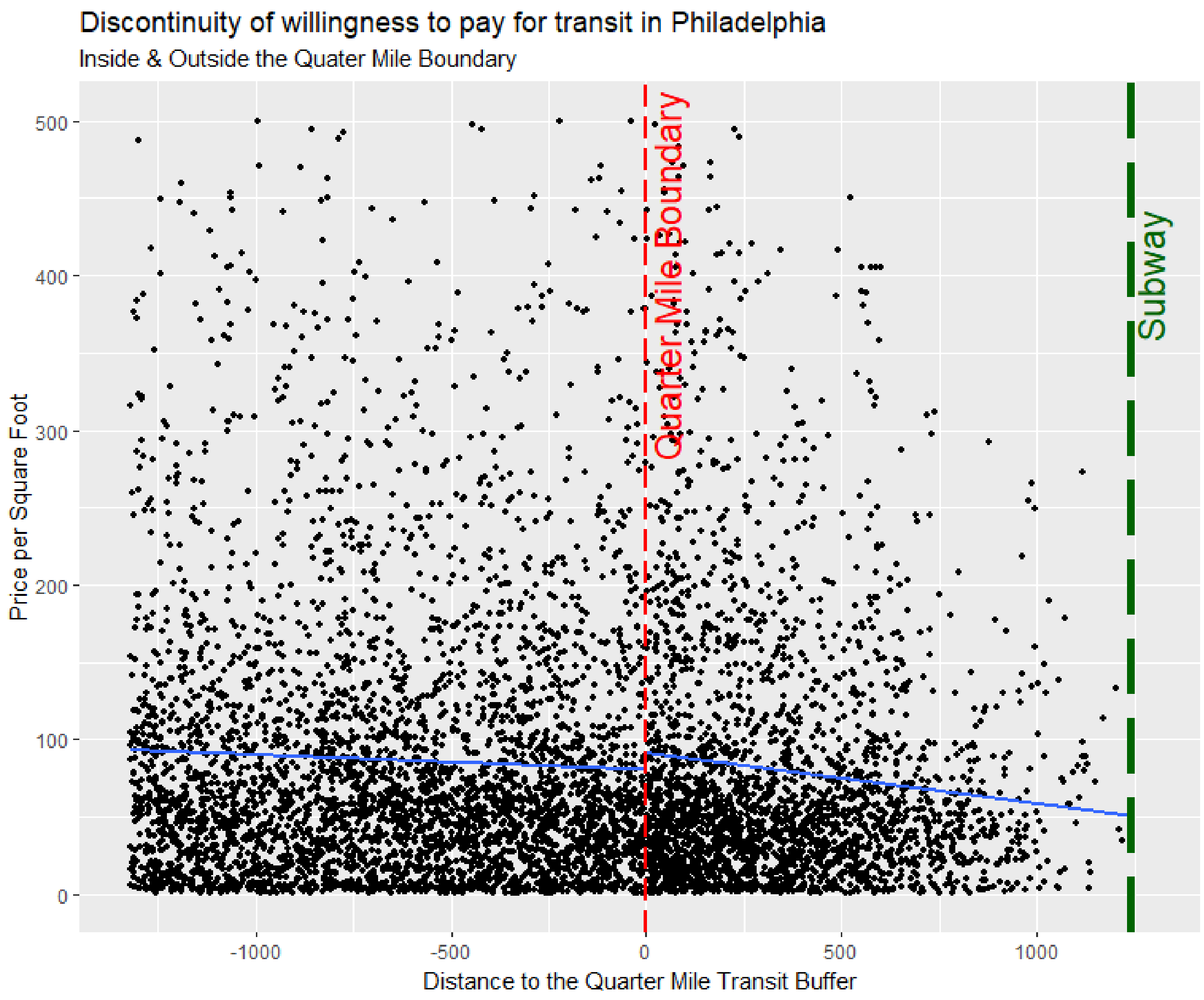
Moran's I Statistic

In the Spatial Autocorrelation test, the Moran's I Statistic is around 0.36004 with a extremely small p-value. Thus, we can reject the Null Hypothesis for the Alternative Hypothesis that a positive spatial autocorrealtion exists among home prices. It supports the conclusion that I draw from residuals map above.

Not Perfect

Therefore, it seem like that the model is not useful for us to understand the willingness to pay for transit, since distance to transit is not a primary factor when people choose where to live, at least from the map and Moran's I Statistic. In other words, some other parameters have more influences on home price rather than the accessibility to public transit.

Discontinuity Plot



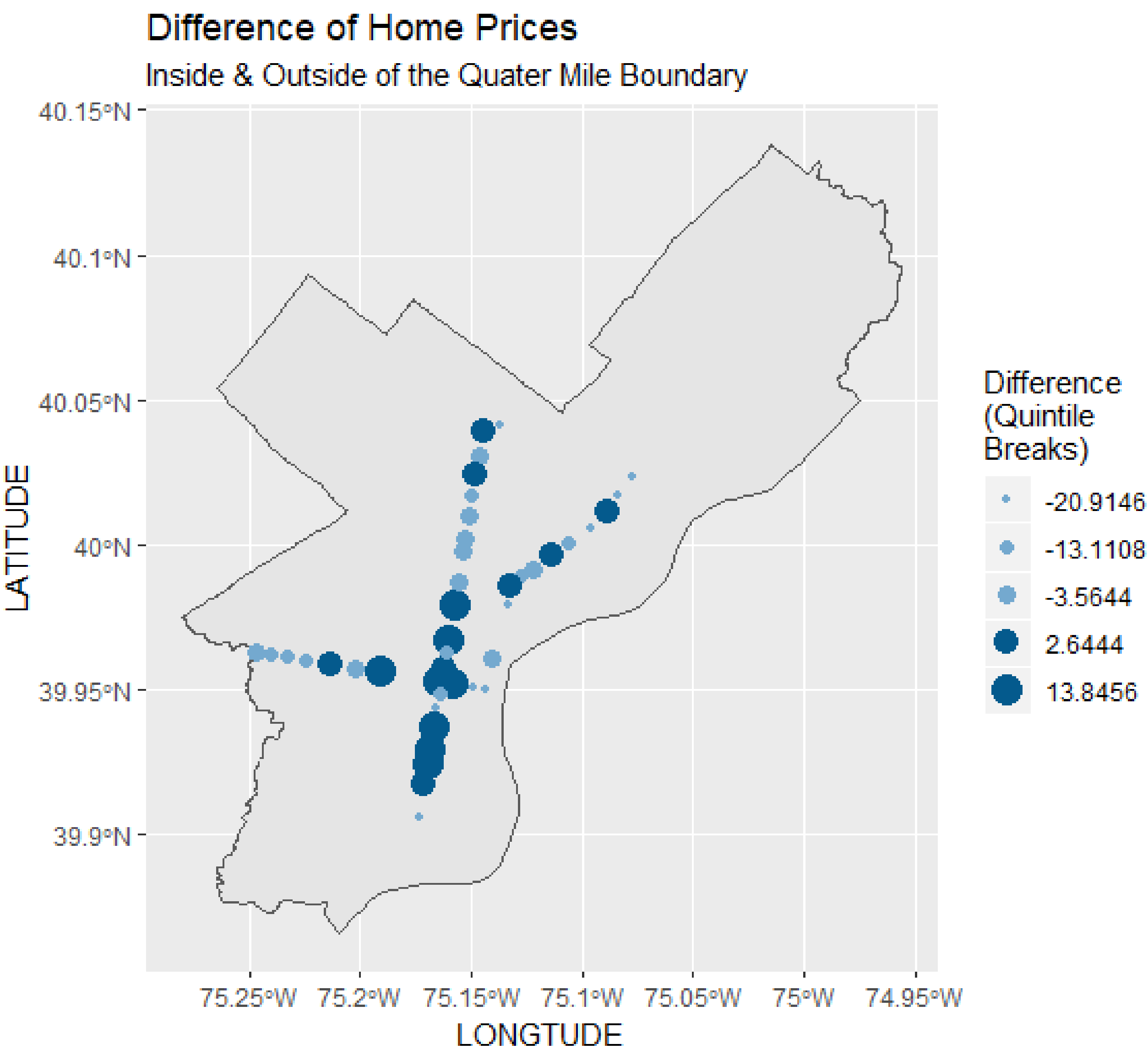
Not Obvious Discontinuity

The discontinuity of willingness to pay for transit is very ***Slight*** between home prices outside and inside the quarter mile boundary. What's more, home prices even ***Go Down*** as distance to SEPTA stations decreases, which totally vialates the original guess that people are more willing to live close to transit, for which the noise or other negative factors may account. They all indicate that the willingness to pay for transit is not obvious in Philadephia.

The Explaination for the Discontinuity

However, the slight discontinuity, to some extent, implies the existence of preference for transit, at least for some people even though it's very mild. I infer that areas around the quarter mile boundary is a suitable choice, where residents can easily secure the access to transit but avoid some potential bad impacts of public transit. That may explain the slight increase of home prices between inside of the boundary and outside.

Local Variation in the Willingness to Pay for Transit



Clustering

The local variation in the willingness to pay for transit doesn't evenly distribute. In general, the postive differences of home prices between inside and outside of the quarter mile boundary, which means high willingness to pay for transit around these areas, locate on the ***Intersection*** and ***End Points*** of these two SEPTA lines.

Intersection

The intersection of these two lines is also central city area, where the accessibility to transit is the highest in terms of both facilities and services. Then the payment for transit is spent more efficiently. In other words, the same amount of money paid for home prices can exchange back more convenience. It makes sense that people are more likely to pay more on home prices if they benefit more.

End Points

As for the end points of the lines, people living there have less choices to travel, so they have to rely more on SEPTA, especially if they don't have automobiles. It's the supply-demand relationship that results in the higher willingness to pay for transit there.

Regression Table

The R^2 of reg1 is as small as 0.0012 with a p-value less than 0.01, which means less than 1% of the variance of the home prices are successufully expained in this regression model. Similarly, in the reg3 which contains other variables expect for station fixed effcts, the R^2 is also very low. Also, the coefficient of lt_qrtMi in reg3 is low, refering to the weak correlation between home prices and whether houses locate inside or outside the quarter mile boundary, which again indicates that the willingness to pay for transit in Philadelphia is not strong.

In the reg2 which includes fixed station effects, the R^2 goes up to 0.379. It's high enough since more one third of the variance of home prices have been explained, which results from the explaination that the spatial pattern in the willingness to pay for transit has been included into the model, hence spatial autocorrelation of residuals has been reduced.

	Dependent variable:		
	log(inf_prc_ft)		
	Just the fixed effect	w/ fixed effects	w/ other variables
	-1	-2	-3
15TH STREET		0.2983	
2ND STREET		0.1147	
30TH STREET		0.1679	
34TH STREET		-0.5911	
40TH STREET		-1.5877***	
46TH STREET		-1.3493***	
52ND STREET		-2.2064***	
56TH STREET		-2.2627***	
5TH STREET		0.2402	
60TH STREET		-2.3030***	
63RD STREET		-1.9060***	
ALLEGHENY		-2.1575***	
BERKS		-1.3154***	
CECIL B MOORE		-2.0553***	
CHURCH		-1.8937***	
ELLSWORTH-FEDERAL		-1.0866***	
ERIE		-2.4807***	

	Dependent variable:		
	log(inf_prc_ft)		
	Just the fixed effect	w/ fixed effects	w/ other variables
	-1	-2	-3
ERIE-TORRESDALE		-1.4154***	
FAIRMOUNT		-1.1919***	
FERN ROCK TRANSPORTATION CENTER		-1.4779***	
FRANKFROD TRANSPORTATION CENTER		-1.3789***	
HUNTING PARK		-2.5144***	
HUNTINGDON		-2.5218***	
LOGAN		-1.9561***	
LOMBARD-SOUTH		-0.0812	
MARGARET-ORTHODOX		-1.8744***	
NORTH PHILADELPHIA		-3.0459***	
OLNEY		-1.5325***	
OREGON		-0.6467*	
PATTISON		-0.1022	
RACE-VINE		0.0988	
SNYDER		-0.9134**	
SOMERSET		-2.5338***	
SPRING GARDEN		-0.2758	

	Dependent variable:		
	log(inf_prc_ft)		
	Just the fixed effect	w/ fixed effects	w/ other variables
	-1	-2	-3
SPRING GARDEN (BROAD STREET)		-0.2805	
SUSQUEHANNA-DAUPHIN		-2.7773***	
TASKER-MORRIS		-1.2568***	
TIOGA		-1.8578***	
WALNUT-LOCUST		0.2096	
WYOMING		-2.3517***	
YORK-DAUPHIN		-2.4767***	
lt_qrtMi	-0.0879***	0.0163	-0.0248
d_abate			-0.0001***
Constant	3.8867***	5.5343***	4.3645***
	-0.0196	-0.3685	-0.0252
Observations	6612	6612	6612
R^2	0.0012	0.379	0.1072
Adjusted R^2	0.0011	0.375	0.1069
Residual Std. Error	1.2320 (df = 6610)	0.9745 (df = 6569)	1.1649 (df = 6609)
F Statistic	8.0760*** (df = 1; 6610)	95.4502*** (df = 42; 6569)	396.7686*** (df = 2; 6609)
Note:	*p **p ***p<0.01		

Residuals Map of Regression Model × 3