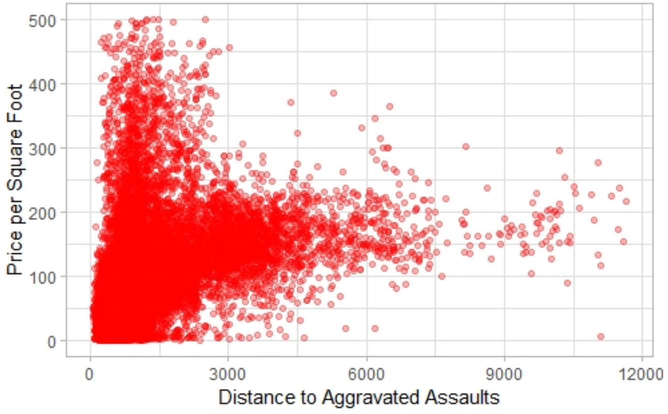
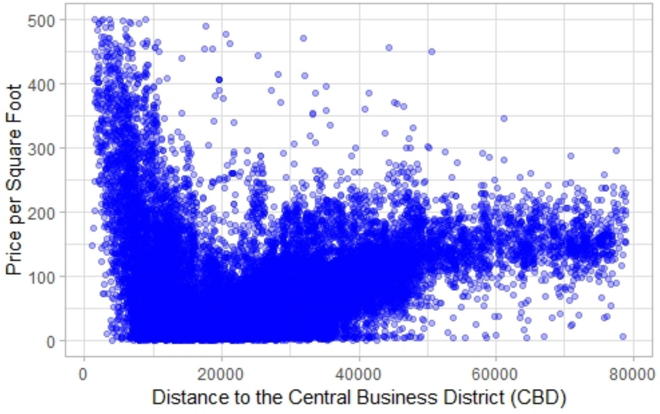
Hedonic regression & willingness to pay for transit access in Philadelphia

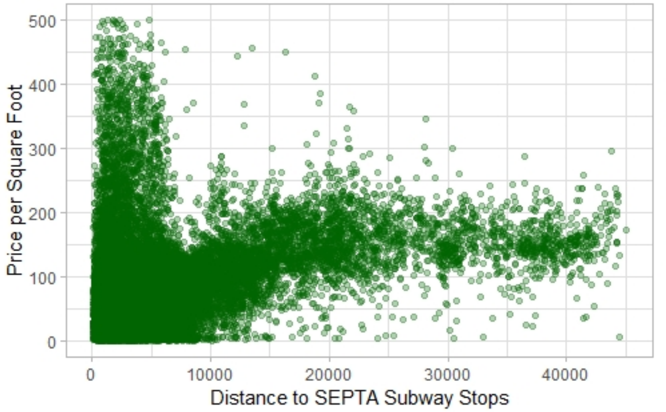
Overview: This statistical analysis studies home prices in relation to Transit Oriented Development (TOD) in Philadelphia. The first part develops a “kitchen sink” model that tries to identify the transit price premium by controlling for neighborhood factors. The second part attempts to develop a quasi-experimental design that can perhaps, more accurately identify the willingness to pay.

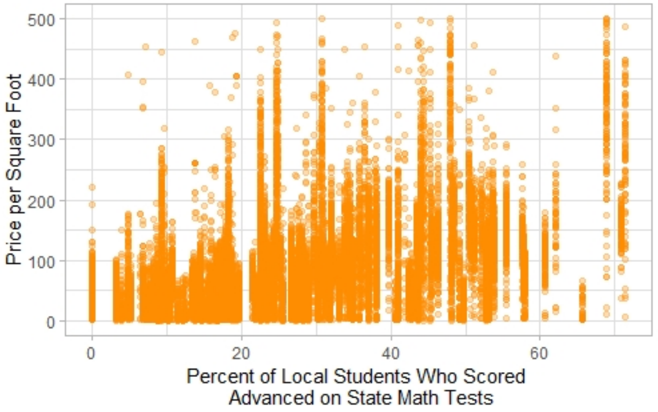
Software used: R

Predictors of Home Prices per Square Footage









When comparing the predictor “Distance to Aggravated Assaults” with the dependent variable “House prices per square foot,” it can be seen that with a closer proximity to aggravated assaults, the range of house prices vary much more than when there is a greater distance away from aggravated assaults. Additionally, there seems to be a higher density of lower house values when proximity to aggravated assaults is closer. When the distance to aggravated assaults become about 3000 ft, it is possible to note that the house prices become noticeably more expensive.

The scatterplot comparing the “Distance to the CBD” and the “House prices” show that the range of house values vary more across a greater distance to the CBD. The range of house prices per square foot is a lot greater when the house is located closer to the CBD, and the values also seem to be relatively consistent for a larger distance, meaning the prices approximate to roughly more or less about $200 per square feet whether the house is about 2 miles (~10000 ft) or 7.5 miles (~40000 ft) from the CBD.

The “Distance to SEPTA Subway Stops” x “Price per square foot” scatterplot follows a similar trend to the first graph. This predictor was chosen with the assumption that people value proximity to transit stops to perform daily activities. House prices vary more greatly when they are closer to the subway stations and vary less as they move further away from these transit stops. Similar to the previous graph, the house values don’t seem to vary much after the distance becomes 20000 ft, where the price seems to fall on average, $150 per square foot. This implies that after a certain distance, distance does not affect much of the house price. For all three graphs previously mentioned, it is clear that house values are higher when they are closer observed predictors and the values and decrease as they move further away from the predicted values.

The last predictor shows a different trend, where house prices do not seem to be very dependent on how well students perform on math tests. This predictor was chosen on the assumption that perhaps, more affluent students (possibly living in more expensive houses) would perform better since they have more resources (such as tutoring) to do well in class. Generously speaking, this could possibly be reflected at the right end of the graph, where 70% of students performing well live in higher priced houses – yet this is a very weak observation as the price range is so great. Generally speaking, the percent of local students’ performance reflected through math scores roughly range more from 0 to 60% with house price similarly ranging up to $300 per square foot, with some going higher than that (spikes on the graph).

Kitchen Sink Regression Analysis

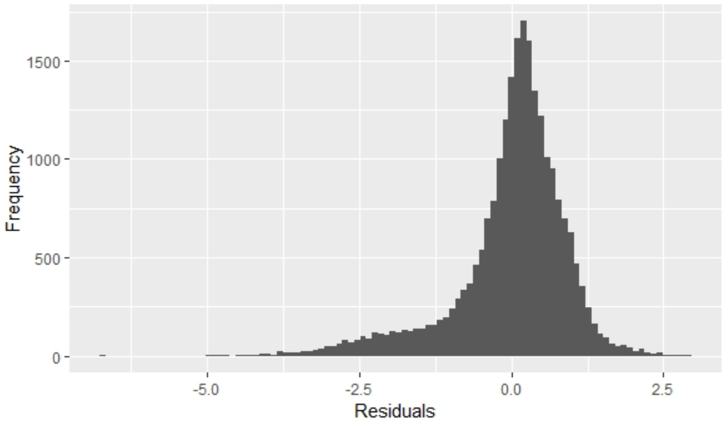
Fitted model				
Price Per Square Foot ~ Distance to SEPTA Stops + Distance to CBD + Distance to Aggravated Assaults + Distance to Highway Ramps + Percentage Non-White + Distance to Off-Site Owners				
Variable	**Coefficient**	**Std. Error**	**t-value**	**Associated p-value**
Beta-Constant	4.68 e+00	1.97 e-02	238.44	0.00
Distance to SEPTA Stops	1.82 e-05	1.65 e-06	11.01	0.00
Distance to CBD	-1.55 e-05	7.60 e-07	-20.37	
Distance to Crime	8.38 e-05	9.20 e-06	9.12	0.00
Distance to Highway Ramps	3.25 e-05	1.86 e-06	17.51	0.00
Percentage Non-White	-1.39 e+00	2.02 e-02	-68.43	
Distance to Off-Site Owners	6.64 e-04	4.51 -05	14.72	0.00
R^2 = 0.3049				
Adjusted R^2 =0.3047				
F-ratio =1700				
p-value =0.00				

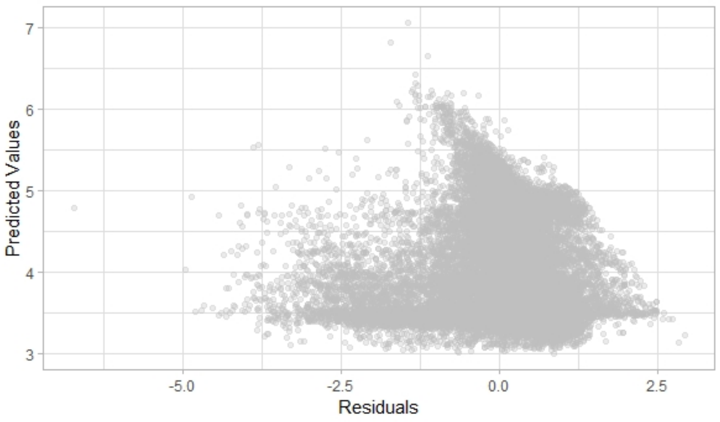
Relative to the Distance to SEPTA coefficient, it can be seen that a 1-foot change in distance to a SEPTA subway station leads to a 0.001817 price change* ($) in home value, holding all other variables constant. Because the 1-ft value is so small, it does not seem to have much effect on home values – rather when the distances become greater, the value changes more greatly. For example, a 1000 ft (~0.2 mi) change in distance to a SEPTA subway station leads to roughly a $1.82 price change in home value.

Looking into the R^2 value, it can be seen that about 30.5% of the variance is explained by the model. Given that the p-value associated with the F-ratio is less than 0.05, we can reject the null hypothesis that the beta coefficient of all variable is 0. This means that the explanatory power of the model is not due to chance.

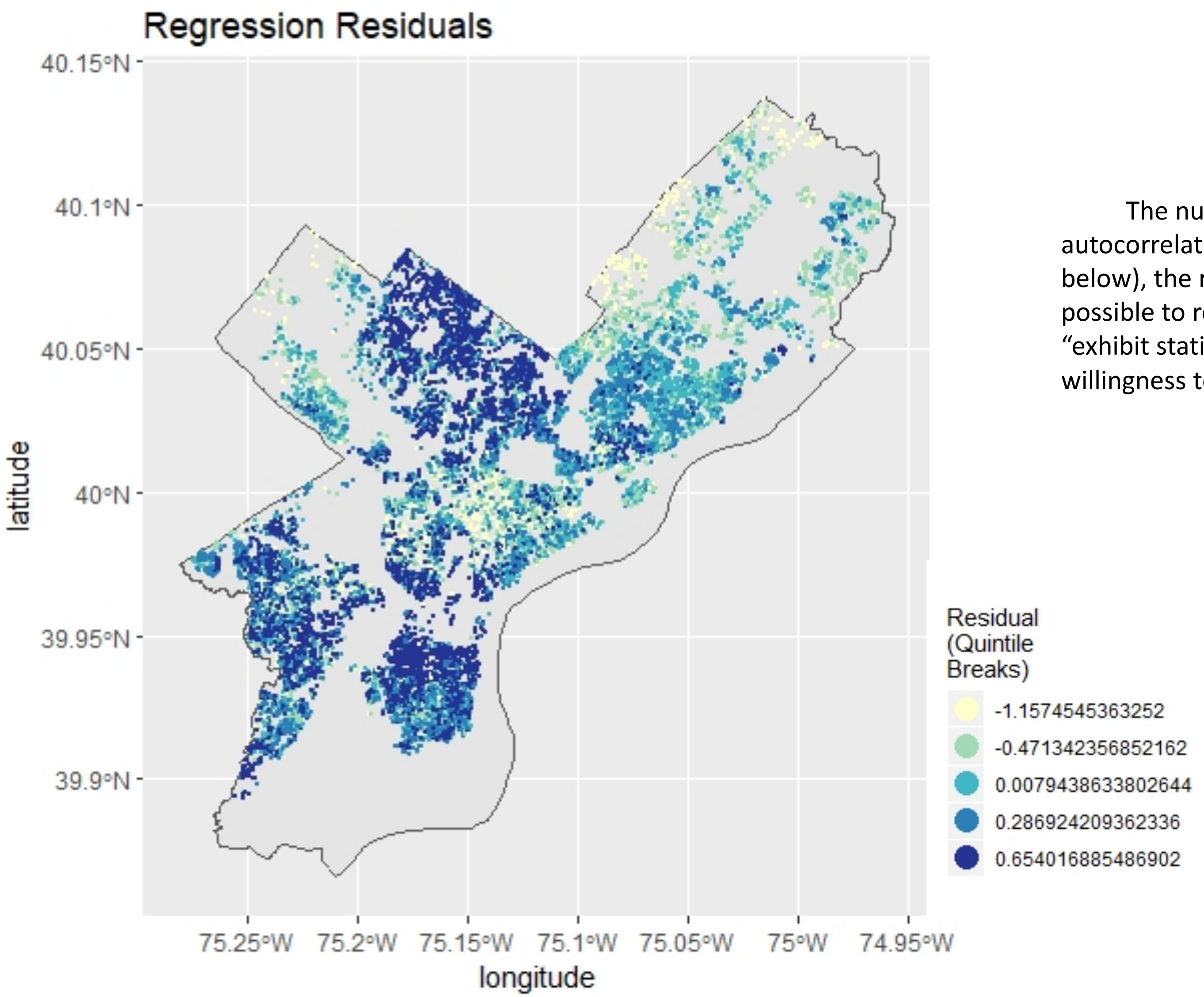
*Value obtained by the equation 100 * (exp(1.817e-05)-1)

Tests for Normality and Homoskedacity





Regression Residuals and Moran's I Statistic Interpretation



The null hypothesis for the Moran's I test claims that there is no spatial autocorrelation, or clustering in the residuals. By running the Moran's I test (shown below), the results reveal that a p-value of less than 0.05 is significant, which makes it possible to reject the null hypothesis for the alternative, which tells that the residuals "exhibit statistically significant clustering." From this, it can be concluded that people's willingness to pay for transit is influenced by the location.

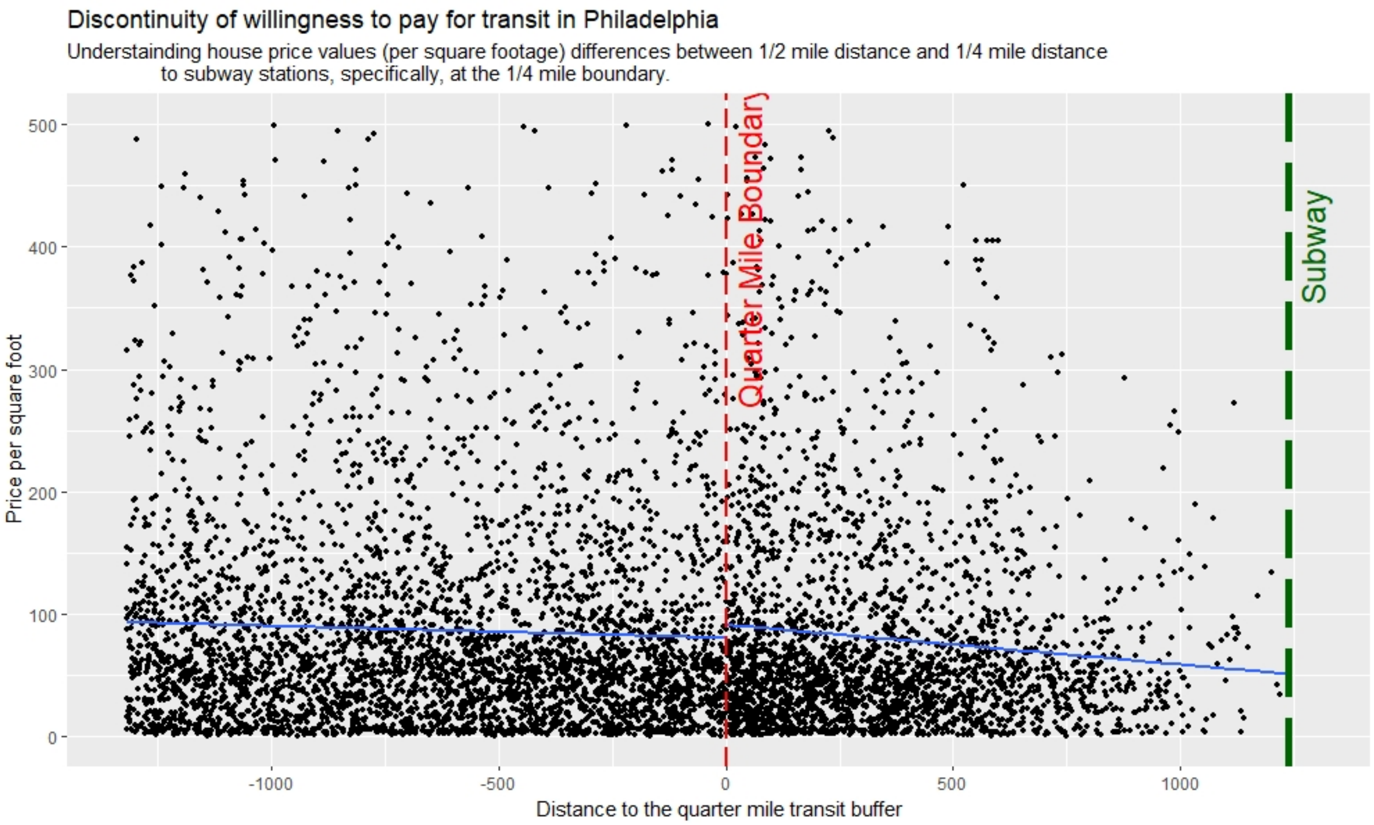
```
> moran.test(regF$residuals,nb2listw(spatialWeights, style='W'))

        Moran I test under randomisation

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

Moran I statistic standard deviate = 77.195, p-value < 2.2e-16
alternative hypothesis: greater
sample estimates:
Moran I statistic       Expectation          Variance
     3.348790e-01     -4.299781e-05      1.882403e-05
```

Discontinuity Plot – Understanding house price value differences at different distances to stations

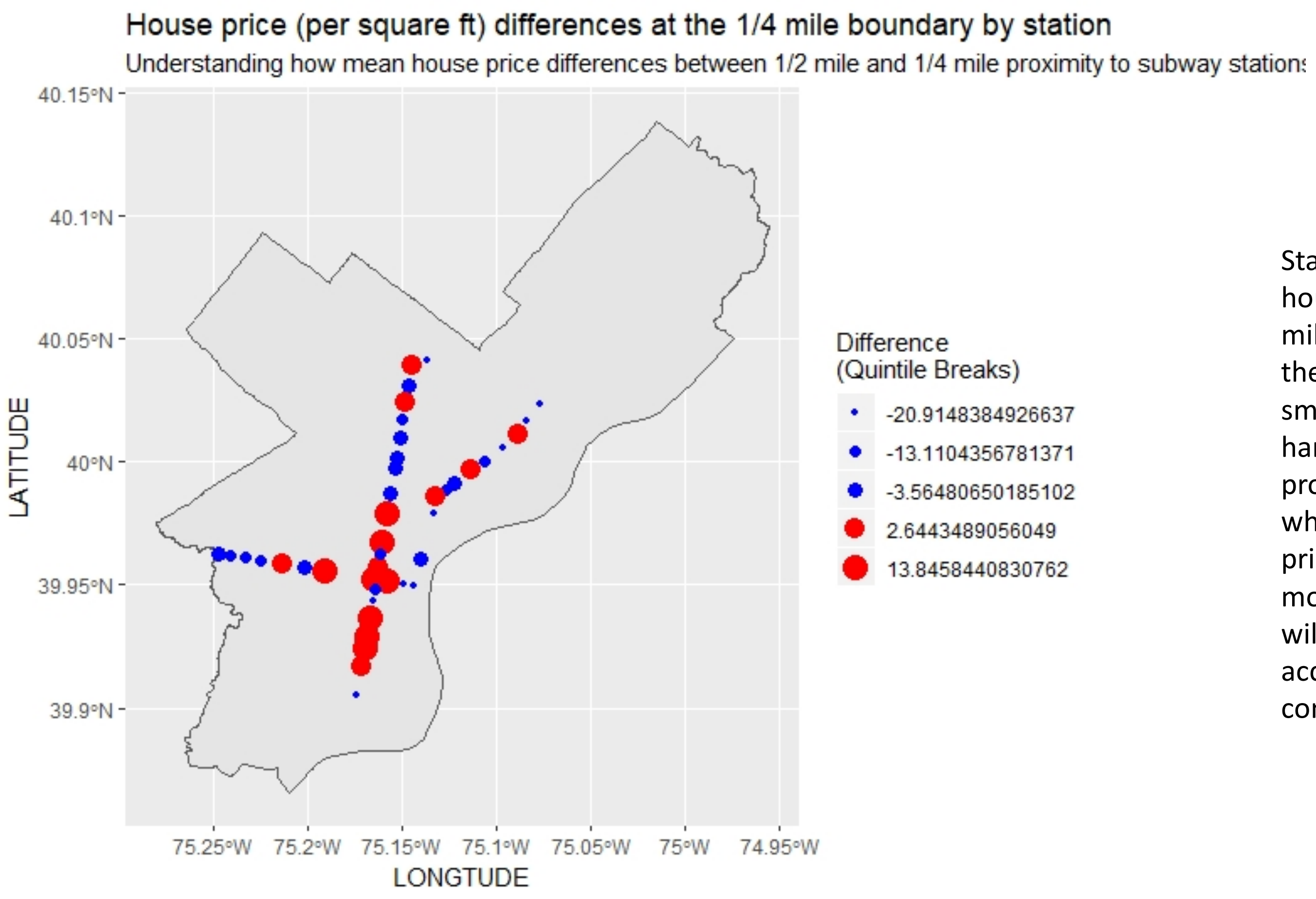


The scatterplot above helps understand how price values (per square foot) vary in areas where houses are located within 1/4 mile to 1/2 mile of subway stations (the control group on the left side of the graph) and areas within 1/4 mile distance to subway stations (treatment group on the right side of the graph).It can be noted that prices vary less on the control group side and slightly more on the treatment group side. Generally speaking, it can be noted that the average price range per square foot of the houses is about \$80 to \$100 for the control group, and about \$50 to a little less than \$100 for the treatment group. Two interesting observations can be made. First, it would be expected that prices be higher close to the subway stations, but the scatterplot shows less observations/house price records within very close proximty to stations. Instead, it is possible to see a higher concentration of houses closer to the 1/4 mile boundary.

This could be explained by the idea that there are less houses within too much proximity ('right next to') of subway stations, which can be disruptive to the people living there (such as noise and safety concerns); rather, it is more likely retail stores are more fit for these locations, while houses are a little further away, but still within a reasonable distance from these transit stops. Perhaps, these negative externalities could explain the lower square foot prices for homes much closer to the subway stops. Prices rise as houses move further away from the transit stops until they reach a 1/4 mile boundary. This could possibly imply that a 1/4 mile distance from transit stops are an ideal location for homes to be located, since they are not too far to walk and not too close to the chaos from subway stations (where there are a lot of people movement and activity).

The second interesting observation is that house prices drop beyond the quarter mile boundary and slowly rise as it moves further away. This contradicts the idea that if the ¼ mile boundary is an ideal distance from transit stops, then prices would be highest at the quarter mile boundary, then drop lower as it moves further away. Yet the graph seems to be showing an opposite trend. This could be explained by where the regression line was divided – perhaps it would show a different trend had the interval been 1/8 mile ~ 3/8 mile. Another explanation for the possible increase in home prices as it moves farther away from the quarter mile boundary is the quality of the house. As houses are built further away from the subway stops (where it is more likely to be compact), those houses are newer which could possibly drive the house prices higher. Ultimately, it could be concluded that in the big picture, prices increase with a greater distance from subways stations. Additionally, prices vary more among houses within a 1/4 mile proximity to subway stops than those beyond the quarter mile boundary.

The Transit Discontinuity – Understanding Home Price Differences Between Subway Stations



This map shows there is local variation in the willingness to pay. Stations with a "negative difference" (represented in blue) show that mean house price differences (per square foot) between homes within the quarter mile boundary and beyond the quarter mile boundary is small. The smaller the value (smaller circle), the smaller the difference. This shows that there a smaller willingness to pay for transit access in these areas. On the other hand, red circles show that the difference in home prices is more pronounced in certain stations, most notably in center city. In conclusion, when looking at Philadelphia county as a whole, the difference in home prices is greater closer to center city, whereas the difference is less as it moves away from the center. This means that people have a higher willingness to pay to live in places like center city, where there is more access to amenities and the closer proximity to transit stops makes it more convenient to get to different places.

Appendix: Regression Results Analysis

	Dependent variable:		
	log(inf_prc_ft)		
	Just the fixed effect	w/ fixed effects	w/ other variables
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***	
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	
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_grtMi	-0.0879***	0.0163	
d_parks			-0.0002***
pct_non_wh			-1.5085***
Constant	3.8867***	5.5343***	4.9407***
Observations	6,612	6,612	6,612
R^2	0.0012	0.379	0.1993
Adjusted R^2	0.0011	0.375	0.1991
Residual Std. Error	1.2320 (df = 6610)	0.9745 (df = 6569)	1.1032 (df = 6609)
F Statistic	8.0760*** (df = 1; 6610)	95.4502*** (df = 42; 6569)	822.5873*** (df = 2; 6609)
Note:			*p**p***p<0.01