Seattle Housing Prices: Assesing the effects of acessibility to public goods

Code ▾

Ahmed Almotaileq, Abigail Edelmen, Christopher Lacrampe

March 08, 2018

Research goals and question:

The aim of this project is to ascertain the effects of access to public goods on housing prices for the Seattle area. A more specific question is selected, asessing the impact of the distance to light rail stations, as public policy has recently pushed for further development of such stations for the areas East of the Seattle area, and homeowners may be interested in seeing if their property values will be effected by the newly constructed stations.

Given the constraints of the available data (only recently listed porperties), we can only extrapolate the effects in the long run (we are unable to asess the imediate impact of light rail construction on East side homes). Furthermore, this extrapolation is also contingent upon the assumption that the characteristics of property-seekers on the East side are analagous to those of property-seekers in Seattle.

Importing the Libraries

Code

The data for this project was collected from:

1. Housing data: Redfin (https://www.redfin.com/city/16163/WA/Seattle) on February 20th, 2018
2. Public goods' locations filtered from Seattle.gov's (https://data.seattle.gov/Community/My-Neighborhood-Map/82su-5fxf/data) "My Neighborhood Map" collection.
    - Light rail data: Seattle.gov (https://data.seattle.gov/Community/Light-Rail-Map/5f4s-t4jf/data) on February 20th, 2018
    - Public parks data: Seattle.gov (https://data.seattle.gov/Community/Parks-Map/rbbt-rarz) on February 23, 2018.
    - Public schools data: Seattle.gov (https://data.seattle.gov/Community/Schools-Map/2tje-83f6) on February 23, 2018.
    - Public hospital data: Seattle.gov (https://data.seattle.gov/Community/Hospitals/khp7-mz6q) on February 23, 2018.

Importing and Cleaning the Data

Code

Special thanks to stack overflow user eclark for providing the framework for the functions used to calculate minimum distances. See his response to a simmilair query from this stackoverflow thread. (https://stackoverflow.com/questions/31732281/finding-minimum-distance-between-two-sets-of-points-in-two-sets-of-r)

Calculateing the distance from the property location to the nearest light rail station, public park, public school, and public hospital using the following code chunks and minimizing the results from solving Vincenty's Inverse Problem (https://en.wikipedia.org/wiki/Vincenty%27s_formulae):

More specificaly, minimizing the results for s where:

$$s = bA(\sigma - \Delta\sigma)$$

and s is measured in meters, and accurate to 0.06mm.

Hide

```
# Preparing lat/lon dataframes for function
redfinCoords <- data.frame(location_id=redfinDat$ADDRESS,LATITUDE=redfinDat$LATITUDE,LONGITUDE=redfinDat$LONGITUDE)
lightRailCoords <- data.frame(location_id=lightRailDat$Common.Name,LATITUDE=lightRailDat$Latitude,LONGITUDE=lightRailDat$Longitude)

# Setting up DistLinkFun to find distance from property to closest lightrail station

DistLinkFun <- function(ID){
 TMP <- redfinCoords[redfinCoords$location_id==ID,]
 TMP1 <- distGeo(TMP[,3:2],lightRailCoords[,3:2]) # uses distGeo() function from geosphere package to calculate dist from lat and lon
 TMP2 <- data.frame(redfinCoordsID=ID,lightRailCoordsID=lightRailCoords[which.min(TMP1),1],distanceToLink=min(TMP1))
 print(ID)
 return(TMP2)
}

# Distance output of DistLinkFun parameters as redfinCoords$location_id, output is in meters
DistanceLinkMatrix <- rbind_all(lapply(redfinCoords$location_id, DistLinkFun))

# Taking distance variable and adding to original redfinDat dataframe
redfinDat$distanceToLink <- DistanceLinkMatrix$distanceToLink # values are in meters

# Adding the closest lightRail station variable to redfinDat
redfinDat$ClosestStation <- DistanceLinkMatrix$lightRailCoordsID
```

Code

Code

Code

Model preperation and outputs:

Code

Statistic	N	Mean	St. Dev.	Min	Max
ZIP	350	98,119.470	21.268	98,101	98,199
PRICE	350	1,053,413.000	889,135.800	115,000	7,500,000
BEDS	350	2.891	1.854	0	16
BATHS	337	2.146	1.121	0.750	8.000
SQUARE.FEET	339	1,984.263	1,594.193	33	13,390
LOT.SIZE	283	10,476.950	25,210.550	1	361,500
YEAR.BUILT	340	1,973.468	39.373	1,900	2,019
DAYS.ON.MARKET	350	11.589	12.062	1	134
X..SQUARE.FEET	339	595.372	328.926	178	3,485
HOA.MONTH	134	642.903	380.451	11	1,954
LATITUDE	350	47.623	0.057	47.488	47.735
LONGITUDE	350	-122.335	0.037	-122.416	-122.220
distanceToLink	350	3,544.922	2,486.051	274.195	9,576.834
distanceToPark	350	410.904	335.378	23.355	2,373.706
distanceToSchool	350	870.586	473.391	101.417	3,464.190
distanceToHospital	350	2,069.736	1,386.455	103.511	9,476.451
inter1	350	0.049	0.215	0	1
inter2	350	0.191	0.394	0	1
inter3	350	0.280	0.450	0	1

Hide

```
redfinDat$inter1 <- as.factor(redfinDat$inter1)
mod4 <- lm(log(PRICE) ~ BEDS + BATHS + PROPERTY.TYPE + SQUARE.FEET + LOT.SIZE + YEAR.BUIL
T + distanceToPark + distanceToSchool + distanceToHospital + distanceToLink + distanceToL
ink*inter1, data = redfinDat)


redfinDat$inter2 <- as.factor(redfinDat$inter2)
mod5 <- lm(log(PRICE) ~ BEDS + BATHS + PROPERTY.TYPE + SQUARE.FEET + LOT.SIZE + YEAR.BUIL
T + distanceToPark + distanceToSchool + distanceToHospital + distanceToLink + distanceToL
ink*inter2, data = redfinDat)


redfinDat$inter3 <- as.factor(redfinDat$inter3)
mod6 <- lm(log(PRICE) ~ BEDS + BATHS + PROPERTY.TYPE + SQUARE.FEET + LOT.SIZE + YEAR.BUIL
T + distanceToPark + distanceToSchool + distanceToHospital + distanceToLink + distanceToL
ink*inter3, data = redfinDat)



stargazer(mod1,mod2,mod3,mod4,mod5,mod6, type = "html",
          covariate.labels = c("No. of bedrooms", "No. of bathrooms", "Mobile/Manufacture
d Home", "Multi-Family(2-4 unit)", "Multi-Family(5+ unit)", "Other Property Type", "Singl
e Family, Residential", "Townhouse", "Square Feet of property", "Lot Size (sqr feet)", "Y
ear built", "Distance to nearest public green space (meters)", "Distance to nearest schoo
l (meters)", "Distance to nearest hospital (meters", "distance to nearest station (meter
s)", "D500", "Distance to nearest station interacted with D500", "D1000", "Distance to ne
arest station interacted with D1000", "D1500", "Distance to nearest station interacted wi
th D1500"),
          column.labels = c("Model 1", "Model 2", "Model 3", "Model 4", "Model 5", "Model
 6"),
          dep.var.labels.include = FALSE,
          dep.var.caption = "log(Price)")
```

	log(Price)					
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	(1)	(2)	(3)	(4)	(5)	(6)
No. of bedrooms	-0.006	0.006	0.005	0.001	0.002	0.003
	(0.020)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
No. of bathrooms	0.240***	0.216***	0.216***	0.225***	0.223***	0.216***
	(0.037)	(0.035)	(0.034)	(0.034)	(0.034)	(0.034)
Mobile/Manufactured Home	-0.555*	-0.436	-0.432	-0.434	-0.457	-0.483
	(0.325)	(0.302)	(0.300)	(0.299)	(0.296)	(0.296)
Multi-Family(2-4 unit)	-0.589***	-0.489***	-0.500***	-0.441**	-0.438**	-0.419**
	(0.196)	(0.184)	(0.183)	(0.184)	(0.183)	(0.182)
Multi-Family(5+ unit)	-0.733***	-0.717***	-0.686***	-0.662***	-0.682***	-0.665***
	(0.209)	(0.193)	(0.192)	(0.192)	(0.190)	(0.190)
Other Property Type	-0.601**	-0.553**	-0.544**	-0.541**	-0.580***	-0.614***
	(0.234)	(0.217)	(0.216)	(0.215)	(0.214)	(0.214)
Single Family, Residential	0.0005	0.118*	0.155**	0.178**	0.172**	0.163**
	(0.070)	(0.068)	(0.070)	(0.070)	(0.069)	(0.069)
Townhouse	0.110	0.145**	0.164**	0.168**	0.168**	0.146**
	(0.074)	(0.070)	(0.070)	(0.070)	(0.070)	(0.069)
Square Feet of property	0.0002***	0.0002***	0.0002***	0.0002***	0.0002***	0.0002***
	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)

Lot Size (sqr feet)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Year built	-0.001	-0.0003	-0.0002	-0.0001	-0.0001	0.00003
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Distance to nearest public green space (meters)		-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Distance to nearest school (meters)		0.0001	0.0001*	0.0001*	0.0001**	0.0001**
		(0.00005)	(0.00005)	(0.00005)	(0.00005)	(0.00005)
Distance to nearest hospital (meters		-0.00001	-0.00001	-0.00001	-0.00001	-0.00001
		(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
distance to nearest station (meters)			-0.00002**	-0.00002**	-0.00003***	-0.00003***
			(0.00001)	(0.00001)	(0.00001)	(0.00001)
D500				0.534		
				(0.506)		
Distance to nearest station interacted with D500				-0.002		
				(0.001)		
D1000					-0.218	
					(0.170)	
Distance to nearest station interacted with D1000					0.00002	
					(0.0003)	
D1500						-0.288**
						(0.125)
Distance to nearest station interacted with D1500						0.0001
						(0.0001)
Constant	14.382***	13.408***	13.282***	13.083***	13.041***	12.896***
	(1.278)	(1.192)	(1.186)	(1.184)	(1.176)	(1.176)
Observations	271	271	271	271	271	271
R^2	0.729	0.771	0.775	0.779	0.782	0.783
Adjusted R^2	0.717	0.758	0.762	0.764	0.767	0.768
Residual Std. Error	0.320 (df = 259)	0.296 (df = 256)	0.294 (df = 255)	0.292 (df = 253)	0.290 (df = 253)	0.290 (df = 253)
F Statistic	63.271*** (df = 11; 259)	61.518*** (df = 14; 256)	58.484*** (df = 15; 255)	52.452*** (df = 17; 253)	53.427*** (df = 17; 253)	53.641*** (df = 17; 253)
Note:						*p<0.1;* ***p<0.05;*** p<0.01

Variables and coefficient results:

- dependent variable is **log(price) in dollars** (not thousands)
- **No. of bedrooms** represents the number of bedrooms for that property in integer form. The coefficient for most of the models is fairly small and insignificant at most thresholds.
- **No. of bathrooms** represents the number of bedrooms for that propery in integer form. The coefficients for this variable are significant at most thresholds, and positive across the models, indicating that as the number of bathrooms increases, the price of the property increases by 24% holding all else constant.
- **Mobile/Manufactured home** represents a dummy variable indicating the style of the home as mobile/manufactured. The coefficient for this model indicates a negative relationship to the price of the property relative to a Condo/co-op style home and holding all other variables constant. However, the coefficient is insignificant for most of the models at most thresholds.

- **Multi-Family home** represents the style of property that includes 2-4 units for multiple family living space. The coefficients for this variable are significant at most thresholds and are negative for all of the models, indicating that a home with a multi-family style with 2-4 units reduces the price by about 40-50% holding all other variables constant and relative to the condo/co-op style.
- **Multi-Family home** represents the style of property that includes 5+ units for multiple family living space. The coefficients for this variable are significant at most thresholds and are negative for all of the models, indicating that a home with a multi-family style with 5+ units reduces the price by about 60-70% holding all other variables constant and relative to the condo/co-op style.
- Houses listed in the **"other" style** of property are represented by the "other" factor level. The coefficients for this variable are significant at most thresholds and are negative for all of the models, indicating that a home listed in the other category reduces the price by about 50-60% holding all other variables constant and relative to the condo/co-op style.
- **Single Family, Residential** represents the style of property that includes a factor level for properties listed in the Single Family, Residential category. The coefficients for this variable are significant at most thresholds (outside of the first two models) and are positive for all of the models, indicating that a home with a Single Family, Residential style increases the price by about 0-20% holding all other variables constant and relative to the condo/co-op style.
- **Townhouse** represents the style of property that includes a factor level for properties listed in the Townhouse category. The coefficients for this variable are significant at most thresholds (outside of the first model) and are positive for all of the models, indicating that a home with a Townhouse style increases the price by about 15% holding all other variables constant and relative to the condo/co-op style.
- The **Square Feet of property** variable captures the effects of increaseing the square feet by one on the price. The coefficient is significant at the 1% level and remains fairly consistant at .02% increase in price per an increase of one square foot holding all other variables constant.
- The **lot size (sqr Feet)** variable captures the effects of increasing lot size (different from internal property size) by one square foot. The coefficient for this variable is insignificant and extremely small for each of the models, perhaps its highly correlated with the square Feet of property and should not be included.
- The **Year built** variable is an integer style variable captures the effects of increasing the age of the house on the price. The coefficient is fairly insignificant, but consistantly negative. *One important change to make to this variable would be to track the age of the house by taking the current year and subtracting year build (this may have more effects than year built variable)*.
- **Distance to nearest public green space** captures the effects of increasing the distance to the nearest public green space by 1 meter. The coefficients are significant at most thresholds and are negative for all of the models that include the variable. Increasing the distance by 1 meter is associated with a .04% decrease in the price of the propetry holding all other variables constant.
- **Distance to nearest public school** captures the effects of increasing the distance to the nearest public school by 1 meter. The coefficients are insignificant at most thresholds and are positive for all of the models that include the variable. Increasing the distance by 1 meter is associated with a .001% increase in the price of the propetry holding all other variables constant.
- **Distance to nearest public hospital** captures the effects of increasing the distance to the nearest public hospital by 1 meter. The coefficients are insignificant at most thresholds and are positive for all of the models that include the variable. Increasing the distance by 1 meter is

associated with a .005% increase in the price of the propetry holding all other variables constant.

- **Distance to nearest station** captures the effects of increasing the distance to the nearest light rail station by 1 meter. The coefficients are significant at most thresholds and are negative for all of the models that include the variable. Increasing the distance by 1 meter is associated with a .0002% decrease in the price of the propetry holding all other variables constant.
- **D500/1000/1500** and their **interactions**: Each of these represents an attempt to identify a threshold distance for where the distance to the nearest light rail likely no-longer has an impact on housing price. By selecting a threshold close to 0, we can artificially create a category for property prices that are not effected and are effected by the distance to the light rail. Each of these dummy variables have coefficients that are different from 0 by a large margin, however from extrapolating the switches between threshold levels, the desired threshold appears to be **between 500 and 1000 meters** *(the coefficient for 500 is positive and 1000 is negative)*.