Impact of regional preference for property investment on perceived financial stability of regions in Somerville

By Anant Jain

PPUA 5262 Fall 2018 --- Prof. Daniel T. O'Brien, Northeastern University

Introduction

Investors are attracted to cities where there are opportunities to make money. They will be drawn to cities which offer them the best combination of scale, risk and return. Cities are complex economies, so a huge number of factors impact this return. Investors take in account a wide range of characteristics and behaviors when assessing the attractiveness of a city; which includes quality and affordability of infrastructure and surrounding places[1]. It would be beneficial for a city to know where the money is coming from, in which areas it is going to and who is going to continue investing in these areas by keeping putting in money for their upkeep. Every city official knows that if you're not running a global capital, you're going to have to work twice as hard to attract investment. Most city officials are looking for an epiphany that will bring investment flooding in but in reality, investors simply want to go somewhere that feels relevant and vibrant.[2] To that end, this paper puts forth a perceived financial stability index which helps in indicating how well-kept and attractive certain regions in the city are and further investigates who is responsible for keeping which areas attractive, both on a parcel and a neighborhood level.

[1] https://www.centreforcities.org/reader/investors-want-guide-cities/makes-city-attractive-investors/
[2] http://www.grantthornton.am/insights/growthiq/five-ways-to-attract-business-investment-to-your-city/

Methods

Datasets:

Permit applications can indicate investment in property and determine the degree of investment put in upkeep; therefore, I utilize City of Somerville's ISD Building Permit Daily Applications dataset which is publicly available on City of Somerville's data website[3]. It comprises of about 37000 relevant permit applications. Each line contains the metadata for one permit application. The meta data is cut and comma separated into the following relevant field names: Parcel ID, Address, Location (Latitude & Longitude), Applicant Name, Applicant's Home Address and Total Permit Fees.

In addition to the building permits dataset, I needed statistical facts about the properties to create my measure and therefore, I utilize City of Somerville's FY 2018 Tax Assessor's dataset, publicly available at Data World[4]. It comprises of about 19000 relevant records where each record details the assessment of one property. The records are cut and comma separated into the following relevant field names: Property ID, Total Value, Year Built, Style, Number of Rooms, Stories and Area in Acres.

I pruned the building permits dataset first by eliminating unwanted columns; and then by eliminating uncomplete set of rows which couldn't be completed by using computational techniques (36%). Missing location information (40%) was filled by reverse geocoding the addresses of the properties for which the permit applications were filed. Same criteria were used for pruning the Tax Assessor's dataset. The Tax Assessor's dataset was further filtered to contain records for only those properties for which there were permit applications. After both pruning and fixing the datasets, I was left with about 26000 records in the building permits dataset and about 7000 records in the Tax Assessor's dataset.

The Building Permits dataset include Permit Address and the Applicant's Home Address. There were many instances where they were not the same and even the city information differed. This means that many people who are not local to Somerville are investing in a property within Somerville and many

---

[3] Available at https://data.somervillema.gov/City-Services/ISD-Building-Permit-Daily-Applications/q3yh-mp87
[4] Available at https://data.world/city-of-somerville/ubdh-uik5

people who are local to Somerville are investing in different regions of Somerville (could or could not be in a different neighborhood). This helps in estimating the regional preferences of people, local and non-local to Somerville, when it comes to investing in a property within Somerville. At this point, the analysis was divided into two parts: Local Investment and Non-Local Investment.

In order to uncover these preferences on a neighborhood level, neighborhood information for Permit Address and Applicant's Home Address was required. In case of locals, neighborhoods in which the Permit Address and Applicant's Home Address reside, were extracted using the neighborhood and parcel shape files of Somerville[5]. Geo-coding of Applicant's Home Address was necessary as there was no parcel information for the same. No need of neighborhood extraction was necessary in case of non-locals. New variables denoting extracted neighborhoods were created, namely: NBHD, to contain the neighborhoods in which Permit Addresses reside and HNBHD, to contain the neighborhoods in which Applicants' Home Addresses reside.

In order to calculate a measure which could explain the perceived financial stability of different areas in Somerville, two new measures from the building permits dataset were calculated for each part of the analysis and are enlisted in table 1.

[5] Available at https://data.somervillema.gov/browse?category=GIS+data

New Measure	Description	Computation
n_same	Extent of recurring same-neighborhood investment for a property	Aggregated count of number of permit applications per property, cut at 3rd quartile and scaled
cost_same	Degree of same neighborhood investment for a property	Aggregated summation of Total Permit Fees per property, cut at 3rd quartile and scaled
n_different	Extent of recurring different -neighborhood investment for a property	Aggregated count of number of permit applications per property, cut at 3rd quartile and scaled
cost_different	Degree of different neighborhood investment for a property	Aggregated summation of Total Permit Fees per property, cut at 3rd quartile and scaled
n_nonlocal	Extent of recurring non-local investment for a property	Aggregated count of number of permit applications per property, cut at 3rd quartile and scaled
cost_nonlocal	Degree of non-local investment for a property	Aggregated summation of Total Permit Fees per property, cut at 3rd quartile and scaled

Table 1: New Measures

Additionally, five measures were borrowed from the Tax Assessor's dataset on the basis of correlation with the newly created measures from the building permits dataset; and were transformed in order to help in the analysis. These borrowed measures are enlisted in table 2.

Borrowed Measure	Description	Transformation
TOTAL_VAL	Total value of the property	Cut at 3rd quartile and scaled
YEAR_BUILT	Year of construction of the property	Cut at 3rd quartile and scaled
STORIES	Number of floors the property has	Cut at 3rd quartile and scaled
NUM_ROOMS	Number of rooms the property has	Cut at 3rd quartile and scaled
AREA_ACRES	Area of the property in acres	Cut at 3rd quartile and scaled

Table 2: Borrowed Measures

Finally, the PFS measure indicating the degree of Perceived Financial Stability was calculated by taking row sums of the new and borrowed measures and scaling the summation.

Results

In order to find out where the investments are coming from and where they are going, permit applications were aggregated on the basis of HNBHD-NBHD pairs, counted and arranged in the form of a square matrix, where the position and count reflected the preferences of people belonging to a particular neighborhood towards all the other neighborhoods of Somerville. Contractors were separated out on a manual cut-off and research basis[6]. Figures 1-4 show neighborhood preference of local non-contractors, local contractors, non-local non-contractors and non-local contractors for property investment in Somerville, respectively.

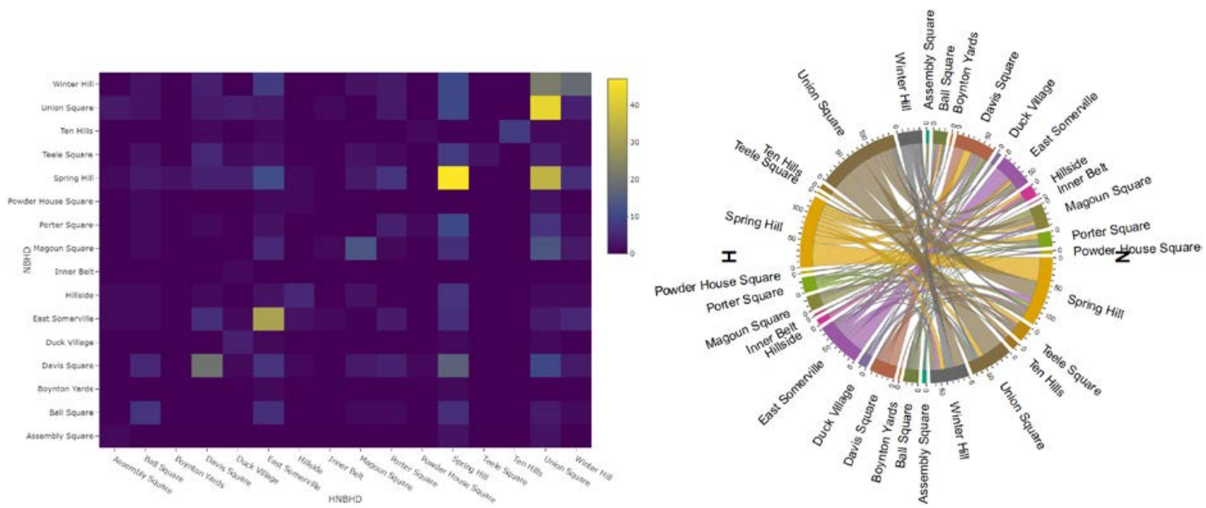

Figure 1: Neighborhood Preference of Local Non-Contractors for Property Investment in Somerville

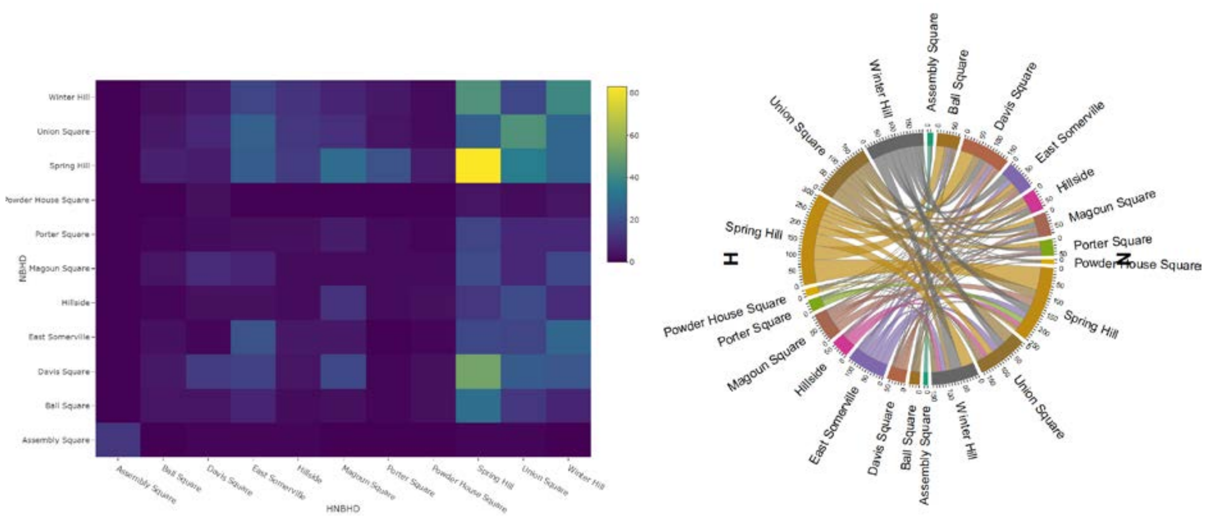

Figure 2: Neighborhood Preference of Local Contractors for Property Investment in Somerville

[6] Counting the number of properties for each applicant, determining their value and then doing a quick google search of the applicant's name to check for keywords which give the indication of the applicant being a contractor

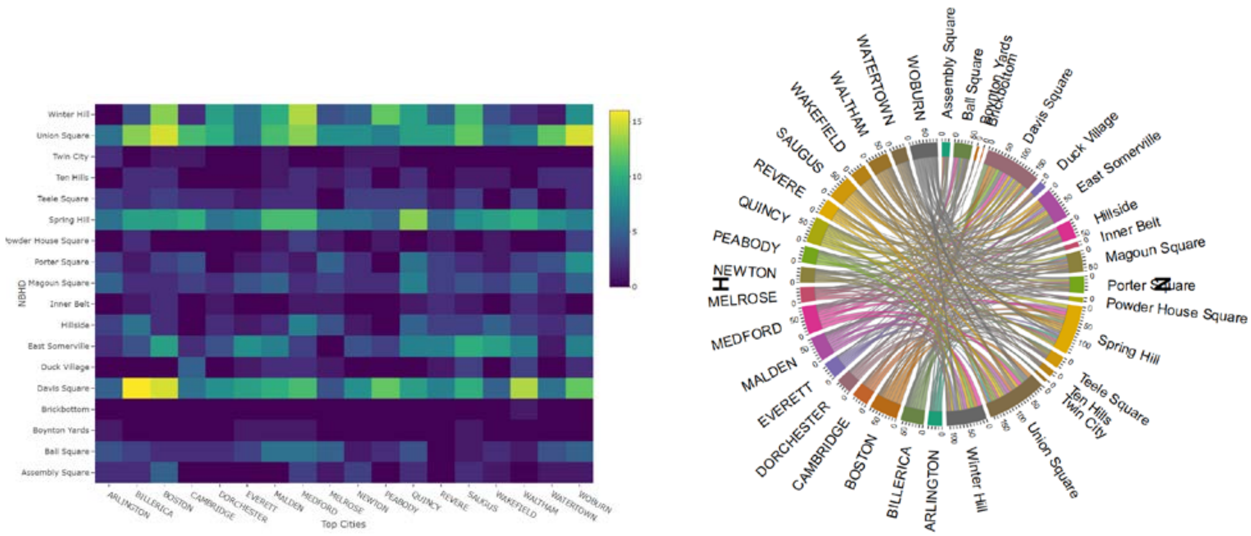

Figure 3: Neighborhood Preference of Non-Local Non-Contractors for Property Investment in Somerville

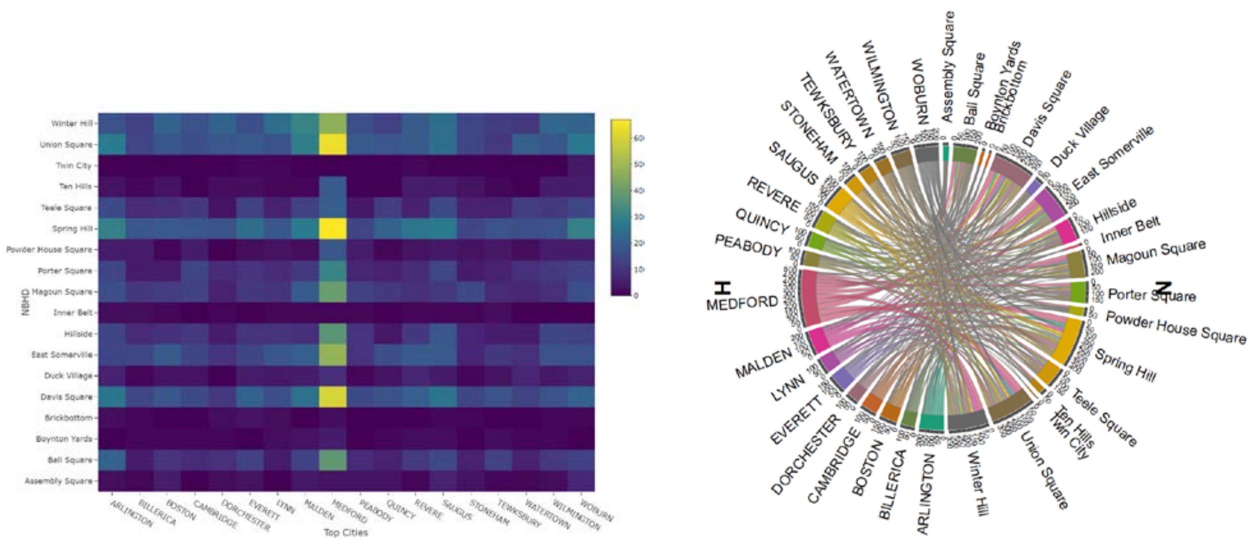

Figure 4: Neighborhood Preference of Non-Local Contractors for Property Investment in Somerville

We see neighborhoods like Spring Hill, Union Square and Winter Hill come out to be the top three neighborhoods where most locals and non-locals want to invest in property. Davis Square is especially a hit with the non-locals and Medford, being the closest city to Somerville, is especially interested in investing in property within Somerville than other cities. In addition, we also see that big chunks of investment for the top neighborhoods are originating and going to the same neighborhoods, therefore, it makes sense to divide the analysis for locals further into two parts, same-neighborhood investment and different neighborhood investment.

In order to analyze and compare how the perceived financial stability of different regions is affected by the regional preference of locals and non-locals in Somerville, I first created violin plots to compare the distribution of PFS measure of properties owned by different types of investors and then ran an ANOVA with post-hoc Tukey HSD test to actually measure this comparison, on both parcel and neighborhood level.

Parcel Level:

The distribution of PFS measure of properties owned by different types of investors on Parcel Level is visualized in figure 5. The Tukey HSD ANOVA test results are tabularized in table 4. Table 3 summarizes the PFS measure on parcel level.

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
0.1609	0.8401	0.9347	0.9745	1.0632	2.9563

Table 3: Summary of PFS measure on Parcel Level

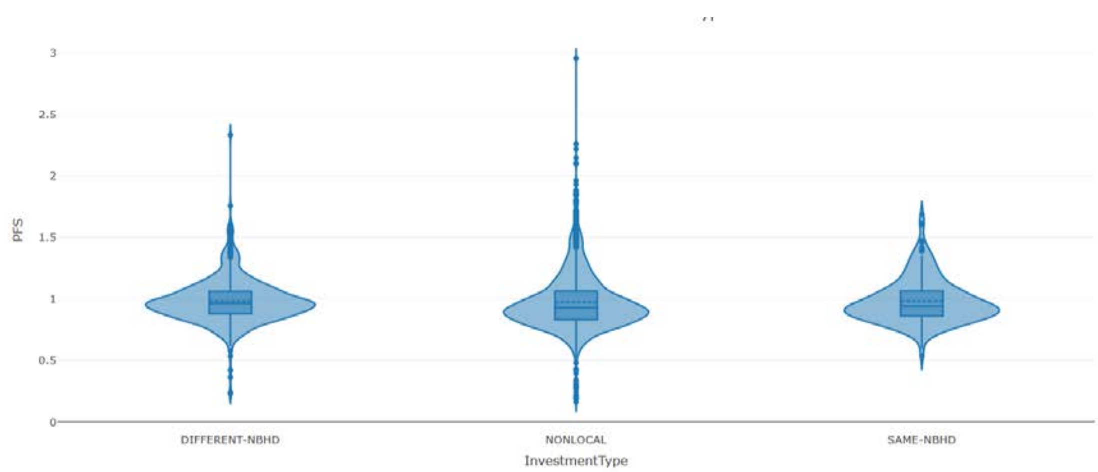

Figure 5: Distribution of PFS measure of properties owned by different types of investors on Parcel Level

Investment Type	diff	lwr	upr	p adj
NONLOCAL–DIFFERENT-NBHD	-0.0097261780	-0.03342027	0.01396791	0.6007036
SAME-NBHD–DIFFERENT-NBHD	-0.0009125602	-0.04132611	0.03950099	0.9984558
SAME-NBHD–NONLOCAL	0.0088136178	-0.02677428	0.04440151	0.8304749

Table 4: ANOVA with post-hoc Tukey HSD Test on Parcel Level

We see there's little difference in the distribution of perceived financial stability of areas where the investment was made by same-neighborhood, different-neighborhood and non-local investors, as shown in figure 5. The same is visible looking at the p adj value of the Tukey HSD ANOVA test results as displayed in table 4. This means that there are areas which locals and non-locals are attracted towards, and are responsible for their upkeep, in almost every neighborhood. These attractive areas are scattered throughout the municipality of Somerville.

Additionally, in order to actually find these areas of financial stability I decided to bin the whole municipality of Somerville and plot a layer of hex-bins depicting the aggregation of the permit sites within that bin and encode their aggregated PFS measure using color. The darker the color, the higher the stability.

Using these hex-bin maps we can easily track these areas of attractiveness throughout the city and find where they are getting concentrated; and we can see that for each type of investor i.e. same-neighborhood, different neighborhood and non-local in figures 6-8, respectively.

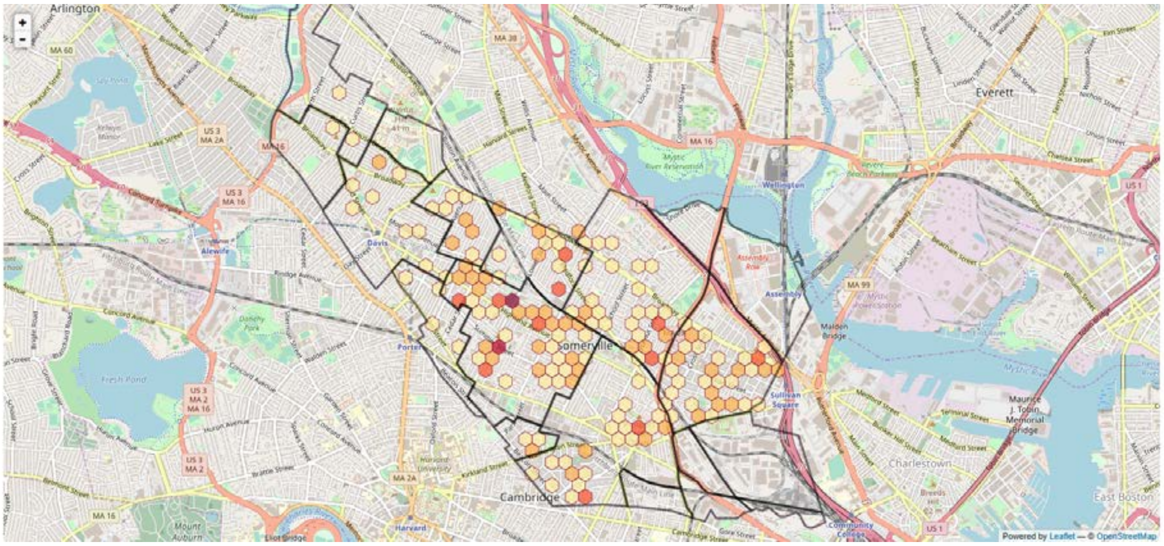

Figure 6: Hex-binned map of aggregated PFS measure for same neighborhood investment

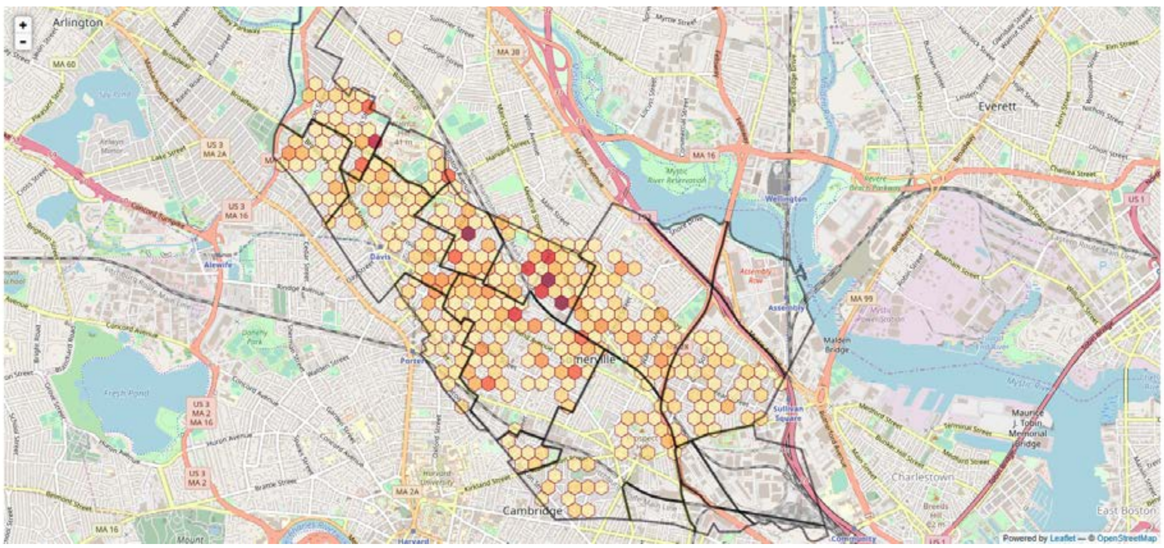

Figure 7: Hex-binned map of aggregated PFS measure for different neighborhood investment

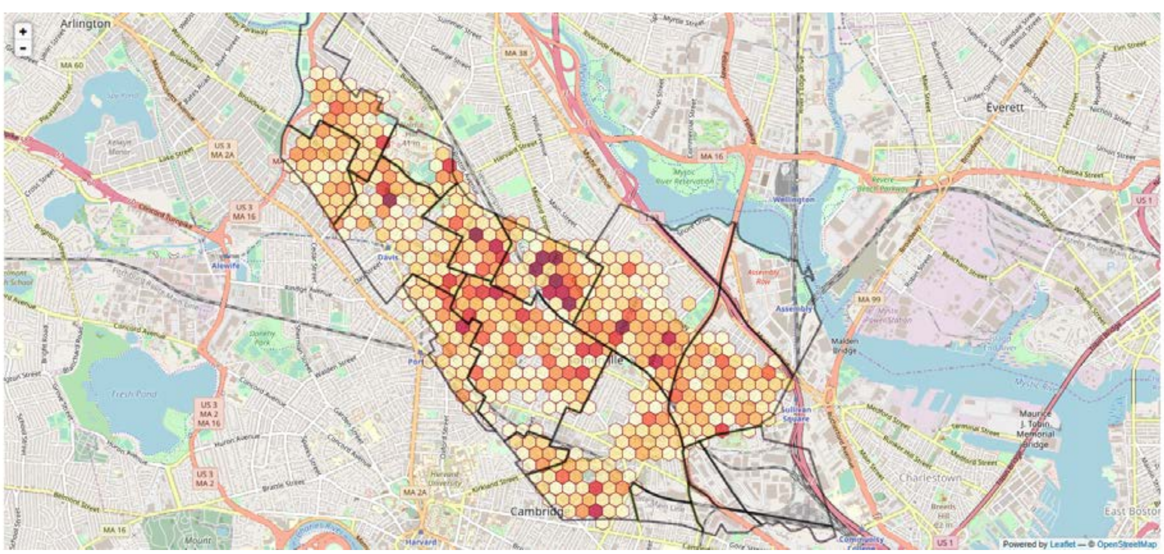

Figure 8: Hex-binned map of aggregated PFS measure for non-local investment

Neighborhood Level:

The distribution of PFS measure of properties owned by different types of investors on Neighborhood Level is visualized in figure 9. The Tukey HSD ANOVA test results are tabularized in table 6. Table 5 summarizes the PFS measure on neighborhood level.

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
0.005829	0.152257	0.287699	0.638282	1.033796	3.216514

Table 5: Summary of PFS measure on Neighborhood Level

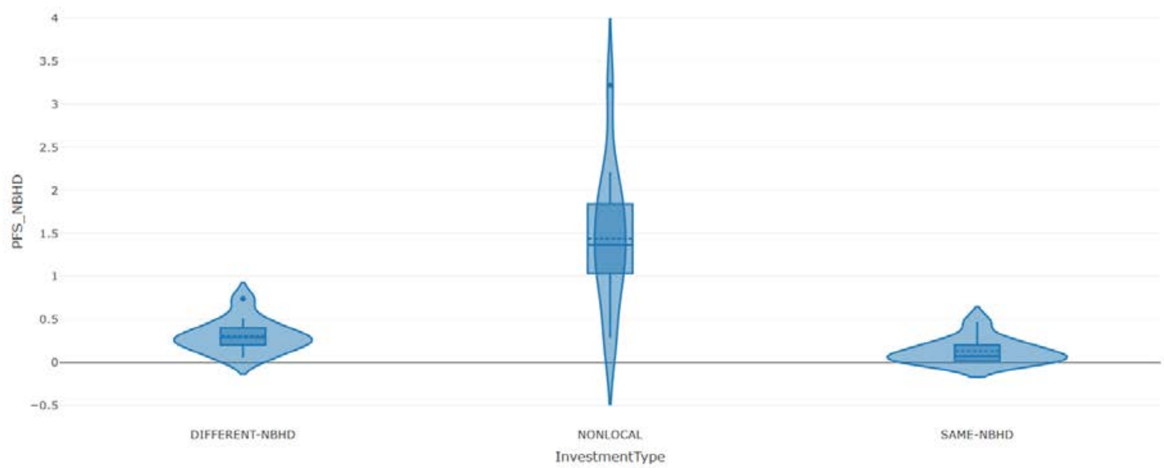

Figure 9: Distribution of PFS measure of properties owned by different types of investors on Neighborhood Level

Investment Type	diff	lwr	upr	p adj
NONLOCAL–DIFFERENT-NBHD	1.13236	0.6329547	1.6317663	0.0000110
SAME-NBHD–DIFFERENT-NBHD	-0.17266	-0.6832898	0.3379699	0.6868546
SAME-NBHD–NONLOCAL	-1.30502	-1.8156503	-0.7943906	0.0000014

Table 6: ANOVA with post-hoc Tukey HSD Test on Neighborhood Level

On neighborhood level, we see there is a big difference in the distribution of perceived financial stability for same-neighborhood, different-neighborhood and non-local investors, as shown in figure 9. The same is visible looking at the p adj value of the Tukey HSD ANOVA test results as displayed in table 6. This means that some neighborhoods which are especially preferred by same-neighborhood investors, are not very well kept in whole as compared to neighborhoods preferred by different-neighborhood investors, and neighborhoods which are especially preferred by non-local investors which are kept very well and tend to have the highest degree of perceived financial stability.

Additionally, I wanted to see how the PFS ranking of neighborhoods change for each type of investor and visualize the results. Therefore, I divided the whole municipality of Somerville by neighborhood and colored the neighborhood polygons by rank using viridis-color scale[7]. Darker the color, more attractive the neighborhood.

Using these neighborhood maps we can easily compare the attractiveness of different neighborhoods throughout the city and rank them for each type of investor i.e. same-neighborhood, different neighborhood and non-local. The ranked neighborhood maps are shown in figures 6-8. The legends show the rankings for each investor type respective to the figure.

[7] https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html

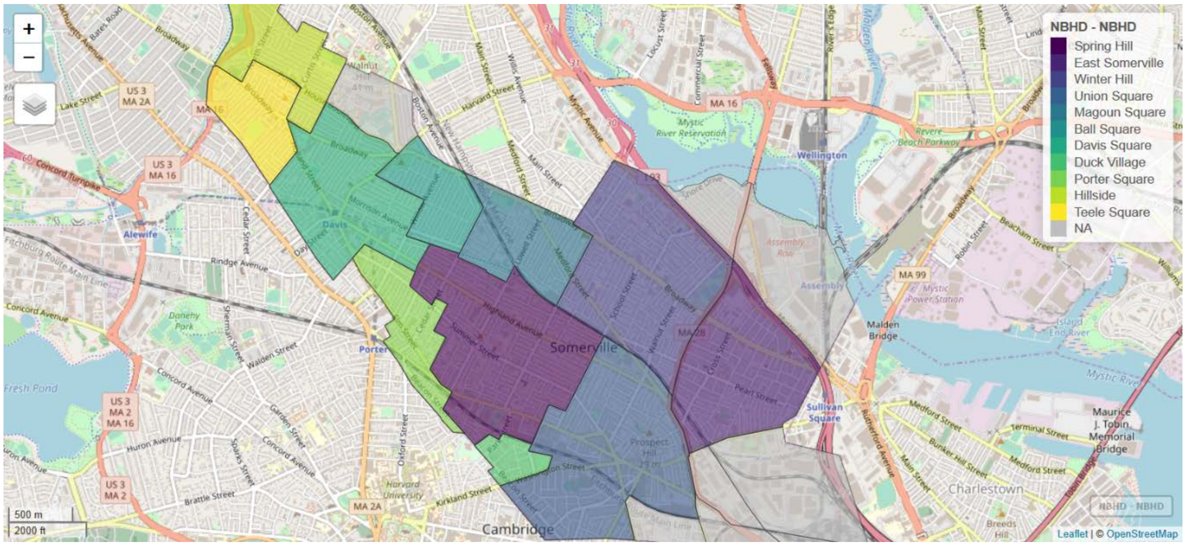

Figure 10: PFS measure map ranked by neighborhood for same neighborhood investment

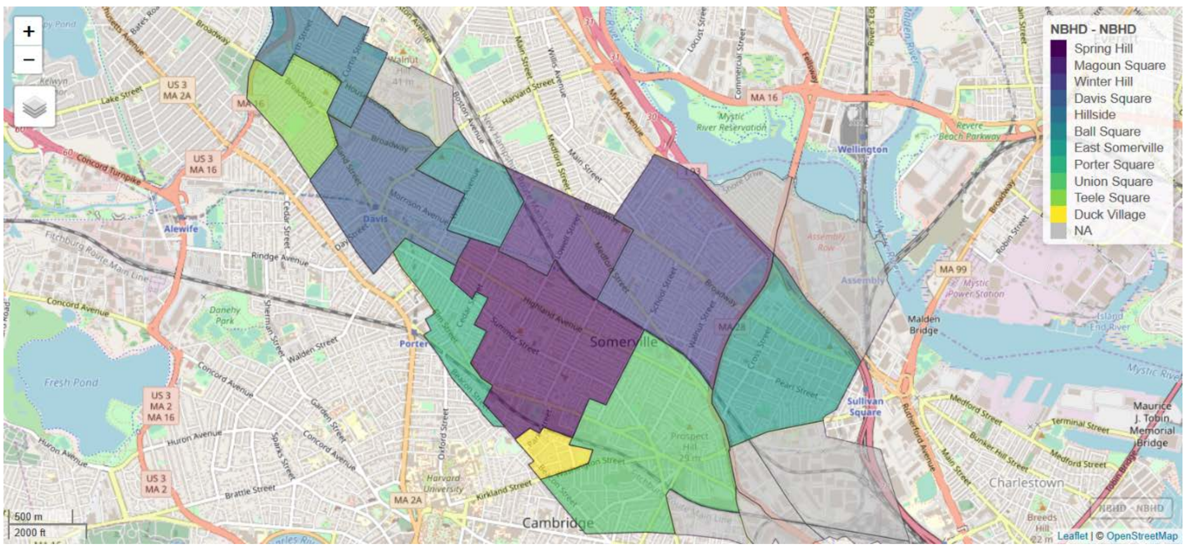

Figure 11: PFS measure map ranked by neighborhood for different neighborhood investment

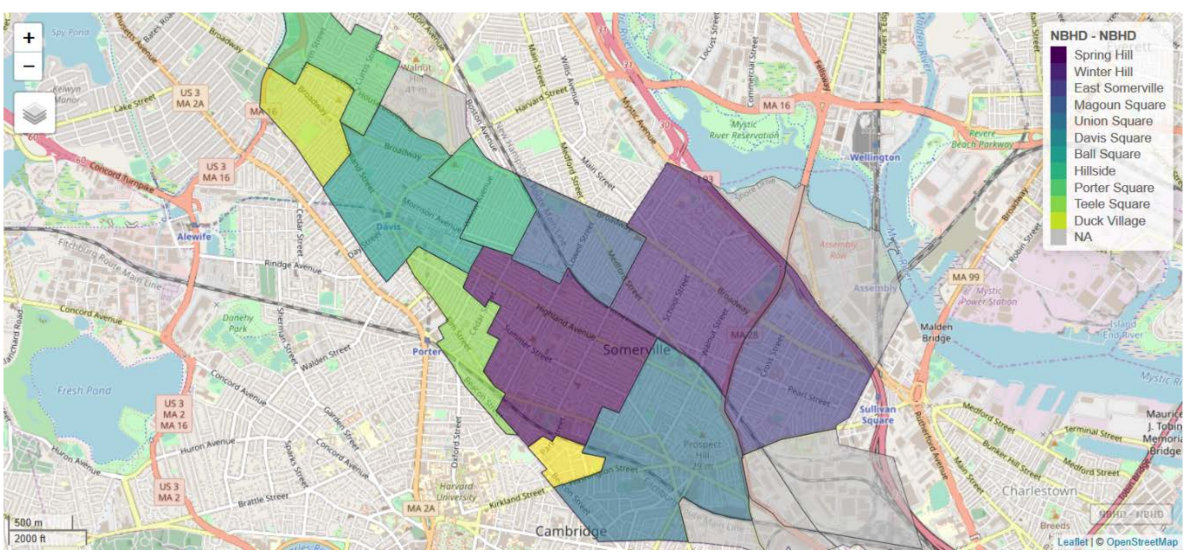

Figure 12: PFS measure map ranked by neighborhood for non-local investment

Discussion

It's beneficial for a city to know where the money is coming from, in which areas it is going to and who is going to continue investing in these areas by keeping putting in money for their upkeep. Using the techniques discussed in this paper, we can answer all of these questions for any city having such kind of permit data.

The PFS measure especially can be used in so many different ways and in so many different domains like, tracking investment in a city, tracking good neighborhoods, estimating property prices, locating the need of gentrification in certain areas or simply driving targeted business agendas. Although, it can definitely be optimized by incorporating financial information about the property owners themselves, it does well enough to visually see the difference in real life, as confirmed during the city walk assignments.

Besides cities, the measure can also be scaled upwards in ranking composite regions like the states of the US[8] and can be checked for scaling behaviors which are followed to estimate attractiveness on different geographical levels using additional datasets like UN's global migrations dataset or US migrations dataset.

[8] https://arxiv.org/abs/1606.08132

Annotated R Syntax

```
# Imports

library(tidyverse)
library(plotly)
library(geojsonio)
library(mapview)
library(chorddiag)
# ----------------------------------------------------------------------------
---
?chorddiag


# Somerville
load('Somerville_ISD_Permits_Cleaned_Corrected_AggregateReady.RData')
# ----------------------------------------------------------------------------
---

# Fix missing NBHDs

Somerville$Address <- sub("\\#.*","",Somerville$Address)
Somerville$Address <- sub("\\..*","",Somerville$Address)
Somerville$Address <- sub("\\,.*","",Somerville$Address)
Somerville$Address <- sub("\\UNIT.*","",Somerville$Address)
Somerville$Address <- sub("\\APT.*","",Somerville$Address)
Somerville$Address <- sub("\\SUITE.*","",Somerville$Address)
Somerville$Address <- sub("\\ U .*","",Somerville$Address)

write_csv(Somerville[is.na(Somerville$NBHD), c('Address', "Latitude",
"Longitude")] %>%
            group_by(Address) %>%
            summarise(Latitude=median(Latitude),
Longitude=median(Longitude)), "missing.csv")

missing <- data.frame(geojson_read('missing.geojson', what='sp'))

Somerville$NBHD[is.na(Somerville$NBHD)] <- with(missing,
NBHD[match(Somerville$Address[is.na(Somerville$NBHD)],missing$Address)])
# ----------------------------------------------------------------------------
---

# Divide the analysis in three parts

load('locals.Rdata')

summary(locals$TotalPermitFees)
summary(Somerville$TotalPermitFees)

locals <- locals %>%
  filter(TotalPermitFees<1000)

Somerville <- Somerville %>%
  filter(TotalPermitFees<1000)

# 1.
same <- locals %>%
  filter(NBHD == HNBHD) %>%
  group_by(MBL, NBHD) %>%
  summarise(n_same=n(), cost_same=sum(TotalPermitFees))

# 2.
```

```
different <- locals %>%
  filter(NBHD != HNBHD) %>%
  group_by(MBL, NBHD) %>%
  summarise(n_different=n(), cost_different=sum(TotalPermitFees))

# 3.
nonlocals <- Somerville %>%
  filter(ApplicantHomeCity!='SOMERVILLE') %>%
  group_by(MBL, NBHD) %>%
  summarise(n_nonlocal=n(), cost_nonlocal=sum(TotalPermitFees))

nonlocals <- na.omit(nonlocals)
# ----------------------------------------------------------------------------
---

# Tax Assessor's
TA <- read_csv('Somerville_TAcensus.csv')

# Extract relevant from Tax Assessor's Data
TA <- TA[c(3,8,30,34,35,36,62)]

# Fix PROP_ID
TA$PROP_ID <- gsub("_", "-", TA$PROP_ID, fixed=TRUE)
data <- str_split_fixed(TA$PROP_ID, "-", 4)[,1:3]
TA$PROP_ID <- paste(data[,1],data[,2],data[,3], sep='-')

# Merge NHBD info
ParcelsNBHD <- geojson_read('SomervilleParcelsNBHD.geojson', what="sp" )
TA <- na.omit(left_join(TA, na.omit(data.frame(ParcelsNBHD) %>% select(MBL,
NBHD)), by = c('PROP_ID'= 'MBL')))

TA <- TA[!duplicated(TA[c("PROP_ID")]),]

summary(TA)
TA <- TA %>%
  filter(TOTAL_VAL<=1000000, STORIES<=3, NUM_ROOMS<=15, AREA_ACRES<48)

save(TA, file="SomervilleTA.RData")
load("SomervilleTA.RData")
# ----------------------------------------------------------------------------
---

# Merge TA data
same <- na.omit(left_join(same,TA %>% select(- NBHD), by =
c('MBL'='PROP_ID')))

different <- na.omit(left_join(different,TA %>% select(- NBHD), by =
c('MBL'='PROP_ID')))

nonlocals <- na.omit(left_join(nonlocals,TA %>% select(- NBHD), by =
c('MBL'='PROP_ID')))
# ----------------------------------------------------------------------------
---

# Check Correlations

# Correlations
corrs <- as.data.frame(lapply(same[c(-1, -2, -7)], as.numeric))
corrs <- cor(corrs[1:2], corrs[3:7])

plot_ly(x=colnames(corrs), y=rownames(corrs), z = corrs, type = "heatmap")%>%
  layout(autosize=F,
         width=1000,
```

```
        height=800,
    margin=list( l = 150, r = 50, b = 100, t = 50),
    title = "Correlations between Investment Variables (from same NBHD) and
Census Tract Variables on Parcel Level")

# Correlations
corrs <- as.data.frame(lapply(different[c(-1, -2, -7)], as.numeric))
corrs <- cor(corrs[1:2], corrs[3:7])

plot_ly(x=colnames(corrs), y=rownames(corrs), z = corrs, type = "heatmap")%>%
  layout(autosize=F,
         width=1000,
         height=800,
         margin=list( l = 150, r = 50, b = 100, t = 50),
         title = "Correlations between Investment Variables (from different
NBHD) and Census Tract Variables on Parcel Level")

# Correlations
corrs <- as.data.frame(lapply(nonlocals[c(-1, -2, -7)], as.numeric))
corrs <- cor(corrs[1:2], corrs[3:7])

plot_ly(x=colnames(corrs), y=rownames(corrs), z = corrs, type = "heatmap")%>%
  layout(autosize=F,
         width=1000,
         height=800,
         margin=list( l = 150, r = 50, b = 100, t = 50),
         title = "Correlations between Investment Variables (from different
city) and Census Tract Variables on Parcel Level")

# ----------------------------------------------------------------------------
---

# Calculate the measure of Percieved Financial Stability

same[c(-1, -2, -7)]<-scale(same[c(-1, -2, -7)], center = F)
same$PFS <- scale(rowSums(same[c(-1, -2, -7)]), center = F)
different[c(-1, -2, -7)]<-scale(different[c(-1, -2, -7)], center = F)
different$PFS <- scale(rowSums(different[c(-1, -2, -7)]), center = F)
nonlocals[c(-1, -2, -7)]<-scale(nonlocals[c(-1, -2, -7)], center = F)
nonlocals$PFS <- scale(rowSums(nonlocals[c(-1, -2, -7)]), center = F)
# ----------------------------------------------------------------------------
---

# Combine the three parts in a singular dataframe and label rows by part
same <- same[c(1,2,7,11)]
same$InvestmentType <- 'SAME-NBHD'
different <- different[c(1,2,7,11)]
different$InvestmentType <- 'DIFFERENT-NBHD'
nonlocals <- nonlocals[c(1,2,7,11)]
nonlocals$InvestmentType <- 'NONLOCAL'

PFS <- rbind(same, different, nonlocals)
# ----------------------------------------------------------------------------
---

# ANOVA on Parcel Level

summary(PFS$PFS)

p1 <- PFS %>% plot_ly(x=~InvestmentType, y=~PFS, type = "violin",
                box = list(
                        visible = T
                      ),
```

```
                      meanline = list(
                        visible = T
                      )
                )%>%
  layout(title = "Distribution of PFS measure across different Investment
Types on Parcel Level")

TukeyHSD(aov(PFS$PFS~PFS$InvestmentType))

# ANOVA on NBHD Level

Mode <- function(x) {
  ux <- unique(x)
  ux[which.max(tabulate(match(x, ux)))]
}

PFS_NBHD <- PFS %>%
  group_by(NBHD, InvestmentType) %>%
  summarise(MOST_COMMON_STYLE=Mode(STYLE), PFS_NBHD=sum(PFS))

PFS_NBHD$PFS_NBHD <- as.numeric(scale(PFS_NBHD$PFS_NBHD,center = F))

summary(PFS_NBHD$PFS_NBHD)

p2 <- PFS_NBHD %>% plot_ly(x=~InvestmentType, y=~PFS_NBHD, type = "violin",
                box = list(
                  visible = T
                ),
                meanline = list(
                  visible = T
                )
)%>%
  layout(title = "Distribution of PFS measure across different Investment
Types on Neighborhood Level")

TukeyHSD(aov(PFS_NBHD$PFS_NBHD~PFS_NBHD$InvestmentType))

# ---------------------------------------------------------------------------
---

NBHD <- sf::st_read("Neighborhoods.shp", quiet = TRUE)
BASE <- mapview(NBHD, zcol = "NBHD")

PFS_NBHD_SAME <- PFS_NBHD %>%
  filter(InvestmentType=="SAME-NBHD")
PFS_NBHD_DIFFERENT <- PFS_NBHD %>%
  filter(InvestmentType=="DIFFERENT-NBHD")
PFS_NBHD_NONLOCAL <- PFS_NBHD %>%
  filter(InvestmentType=="NONLOCAL")

PFS_NBHD_SAME$NBHD <- factor(PFS_NBHD_SAME$NBHD, levels =
PFS_NBHD_SAME$NBHD[order(PFS_NBHD_SAME$PFS_NBHD, decreasing = TRUE)])
NBHD$NBHD <- factor(NBHD$NBHD, levels = levels(PFS_NBHD_SAME$NBHD))
PFS_NBHD_SAME <- mapview(NBHD, zcol = "NBHD")
PFS_NBHD_SAME
PFS_NBHD_DIFFERENT$NBHD <- factor(PFS_NBHD_DIFFERENT$NBHD, levels =
PFS_NBHD_DIFFERENT$NBHD[order(PFS_NBHD_DIFFERENT$PFS_NBHD, decreasing =
TRUE)])
NBHD$NBHD <- factor(NBHD$NBHD, levels = levels(PFS_NBHD_DIFFERENT$NBHD))
PFS_NBHD_DIFFERENT <- mapview(NBHD, zcol = "NBHD")
PFS_NBHD_DIFFERENT
PFS_NBHD_NONLOCAL$NBHD <- factor(PFS_NBHD_NONLOCAL$NBHD, levels =
PFS_NBHD_NONLOCAL$NBHD[order(PFS_NBHD_NONLOCAL$PFS_NBHD, decreasing = TRUE)])
```

```
NBHD$NBHD <- factor(NBHD$NBHD, levels = levels(PFS_NBHD_NONLOCAL$NBHD))
PFS_NBHD_NONLOCAL <- mapview(NBHD, zcol = "NBHD")
PFS_NBHD_NONLOCAL
save(PFS_NBHD_SAME, PFS_NBHD_DIFFERENT, PFS_NBHD_NONLOCAL, p1, p2,
file="Maps.RData")
sync(BASE, PFS_NBHD_SAME, PFS_NBHD_DIFFERENT, PFS_NBHD_NONLOCAL)
# ---------------------------------------------------------------------------
---

data <- Somerville[c("MBL", "Latitude", "Longitude")]
data <- data[!duplicated(data[c("MBL")]),]

PFS_PARCEL_SAME <- left_join(PFS %>% filter(InvestmentType=="SAME-NBHD"),
data, by = c('MBL'= 'MBL'))
geojson_write(PFS_PARCEL_SAME , lat = "Latitude", lon = "Longitude", file =
"PFS_PARCEL_SAME.geojson")

PFS_PARCEL_DIFFERENT <- left_join(PFS %>% filter(InvestmentType=="DIFFERENT-
NBHD"), data, by = c('MBL'= 'MBL'))
geojson_write(PFS_PARCEL_DIFFERENT , lat = "Latitude", lon = "Longitude",
file = "PFS_PARCEL_DIFFERENT.geojson")

PFS_PARCEL_NONLOCAL <- left_join(PFS %>% filter(InvestmentType=="NONLOCAL"),
data, by = c('MBL'= 'MBL'))
geojson_write(PFS_PARCEL_NONLOCAL , lat = "Latitude", lon = "Longitude", file
= "PFS_PARCEL_NONLOCAL.geojson")

ggplot(PFS_NBHD, aes(x=NBHD, y=PFS_NBHD, fill=InvestmentType)) +
  geom_bar(stat='identity')
```