Analysis of the Gentrification Process in the City of Miami

Sabrina Romero[1], Reinaldo Sanchez-Arias, Ph.D[2]

[1]Miami Dade College, InterAmerican Campus, Miami FL

[2]St. Thomas University, School of Science, Miami Gardens, FL

Introduction

Gentrification is a process involving the movement of a high-income group to working-class neighborhoods nearby amenities, and that usually results in the displacement of the original inhabitants. Changes like alteration of local services, increasing prices, and unaffordable housing are derived from the process, and the negative repercussions it might have in our community, grant great significance to its forecasting. Our goal for this project was to offer a comprehensive analysis of the main indicators of the gentrification process in Miami. By finding significant patterns of these indicators, we can help local governments acknowledge this reality, and thus, intervene in an effort to balance development and tradition. The data that was collected and analyzed, along with the different patterns found for the indicators considered in this study, will help in the development of a forecasting model for gentrification of the neighborhoods in the City of Miami.

Methods

In order to analyze how the City of Miami was being affected by gentrification, three of the main indicators were chosen:

- Local tax rates
- Records of building permits
- Housing prices

The datasets conveying such information were filtered and analyzed using R [1], an open source statistical programming language widely used in the data science community. Furthermore, the public opinion of each of the neighborhoods was obtained from social media, specifically, from Twitter. For this endeavor, sentiment analysis was implemented.

RStudio

Programming was essential in the development of this project. R is an open source language widely used in the data science community, with focus on statistical data analysis, data visualization and machine learning methods. During this project, tools from the "tidyverse" package [2] were used for data wrangling and data visualization with the help of RStudio, an open source integrated development environment (IDE) for R.
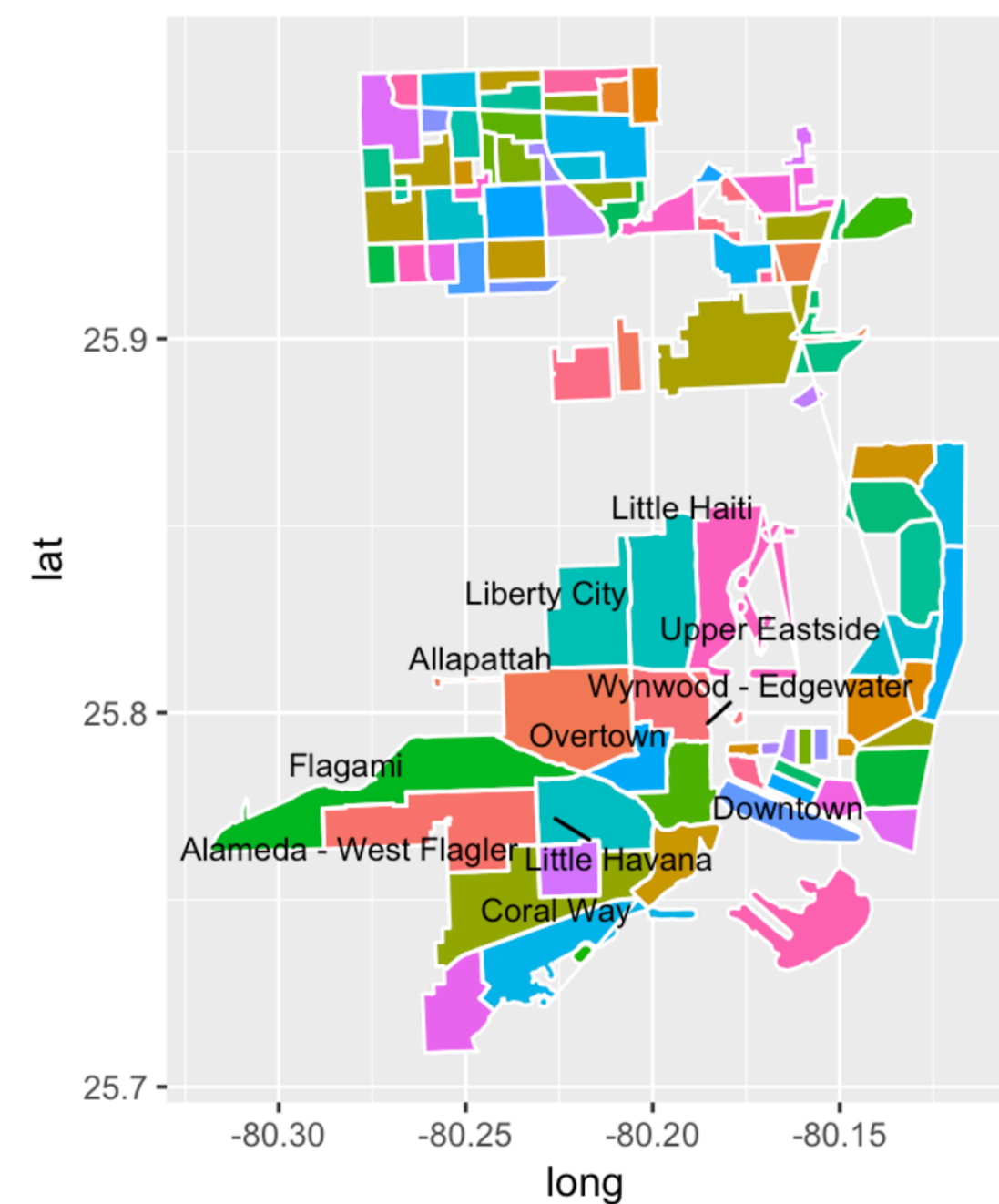
Fig 1: RStudio logo. RStudio makes R easier to use. It includes a code editor, debugging and visualization tools.

Fig 2: Neighborhoods. Map created in R showing the different neighborhoods in the City of Miami that were considered in this study

Exploratory Data Analysis and Methods

Local Business Taxes

The dataset used gave a detailed portray of each tax paid or to be paid in the year 2017 by businesses in Miami [3]. The neighborhood where the business was located was determine using the coordinates also available in the dataset.
The following results were obtained, showing Downtown, as expected, as the neighborhood with more local businesses

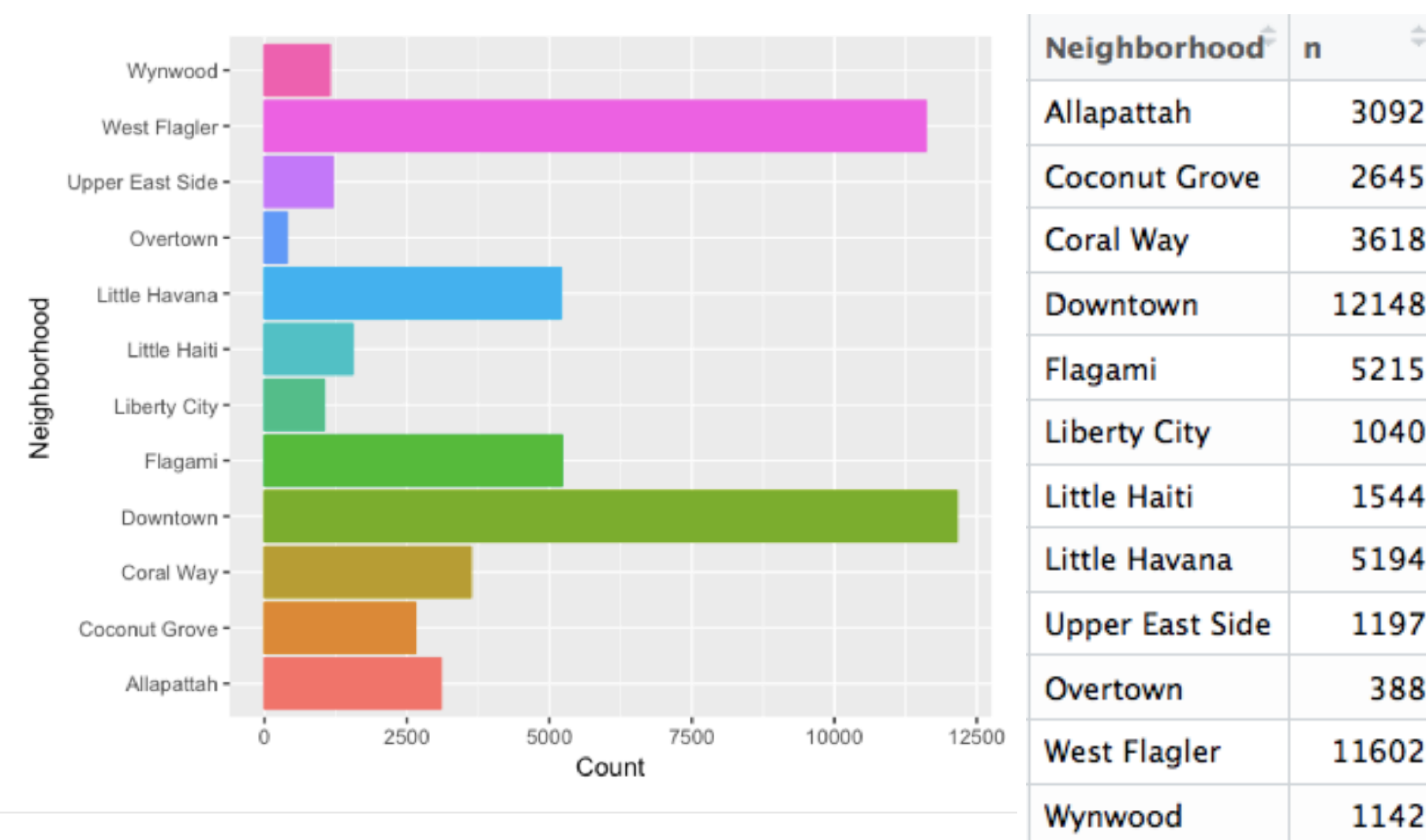| Neighborhood	n
Allapattah	3092
Coconut Grove	2645
Coral Way	3618
Downtown	12148
Flagami	5215
Liberty City	1040
Little Haiti	1544
Little Havana	5194
Upper East Side	1197
Overtown	388
West Flagler	11602
Wynwood	1142

Fig 3: Local Taxes. West Flagler and Little Havana resulted as the next neighborhoods having more businesses, which indicates these can be gentrifying.

Building Permits

The dataset used conveyed the amount of building permits per neighborhood in the years 2014, 2015, and 2016 [3]. After filtered, the following results were obtained:
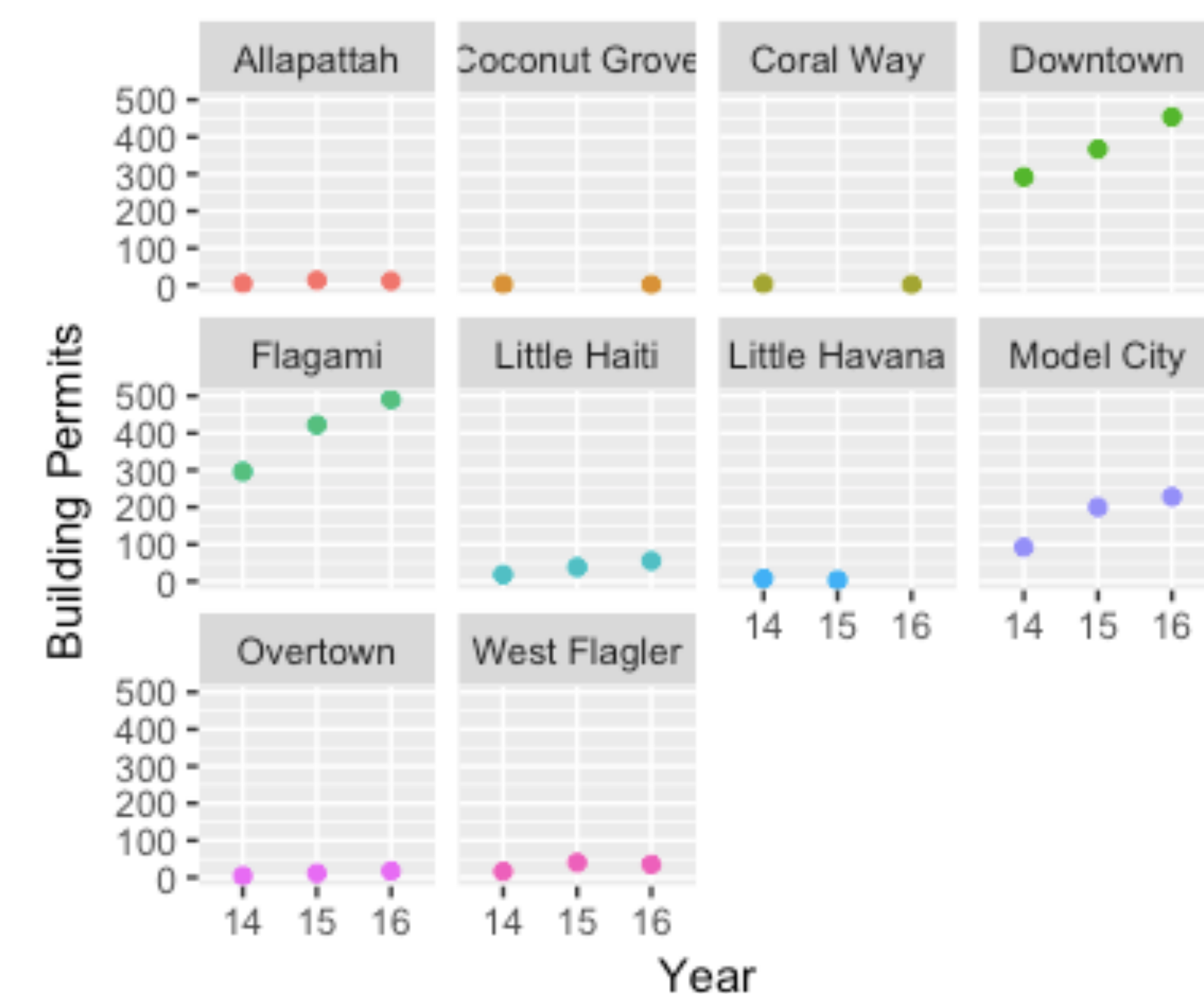
Fig 4: Building Permits. A trend of increasing building permits, can show the gentrification of an area. The neighborhoods showing an increasing trend over the years are Downtown, Flagami, Little Haiti, Liberty City, and Overtown

Housing prices

Information on housing prices was obtained from Zillow [4], an online real estate database company that has data on 110 million homes across the United States.

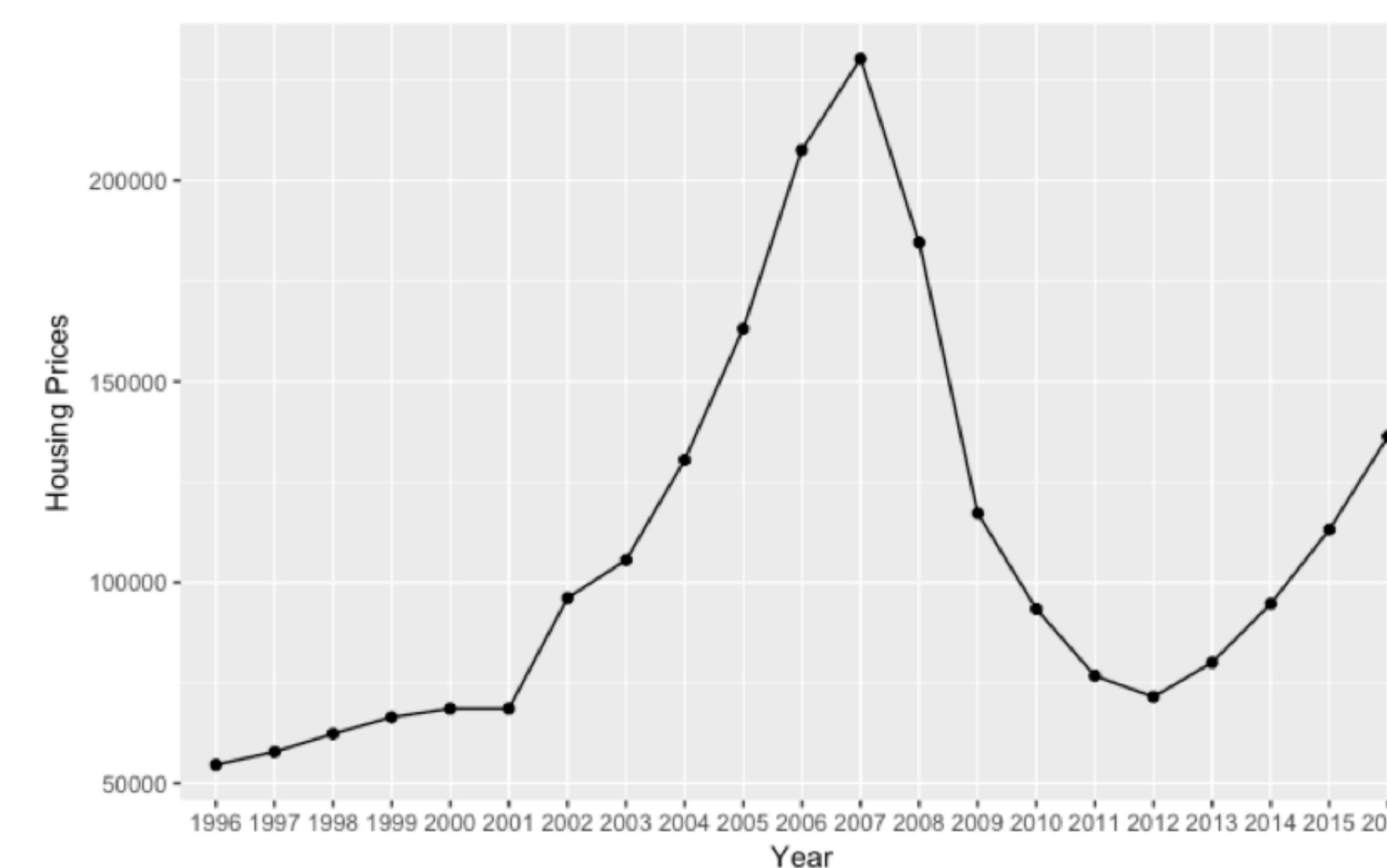The dataset contains the housing prices per neighborhood from 1996 to 2016.

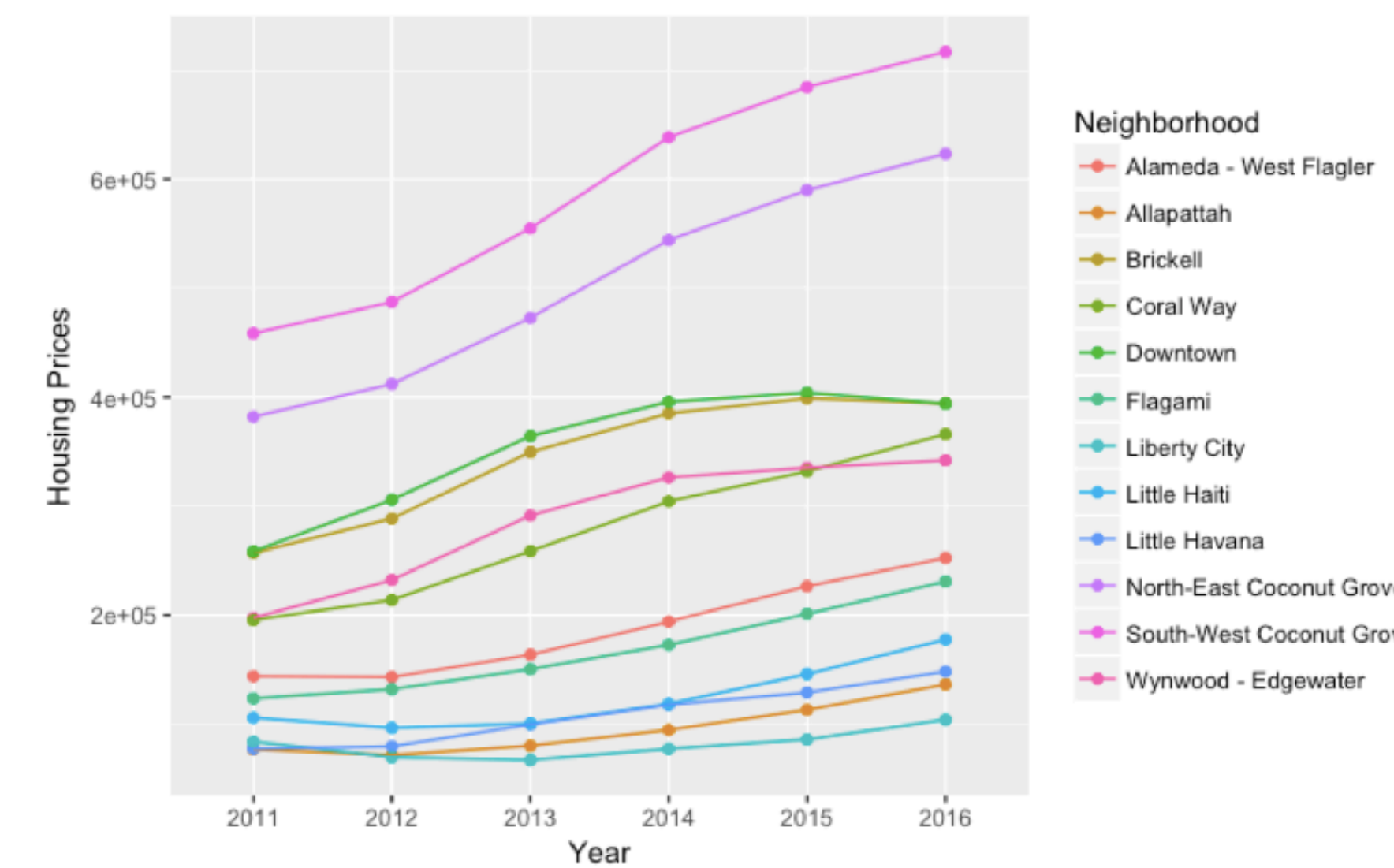Fig 5: Housing Prices in Allapattah since 1996.

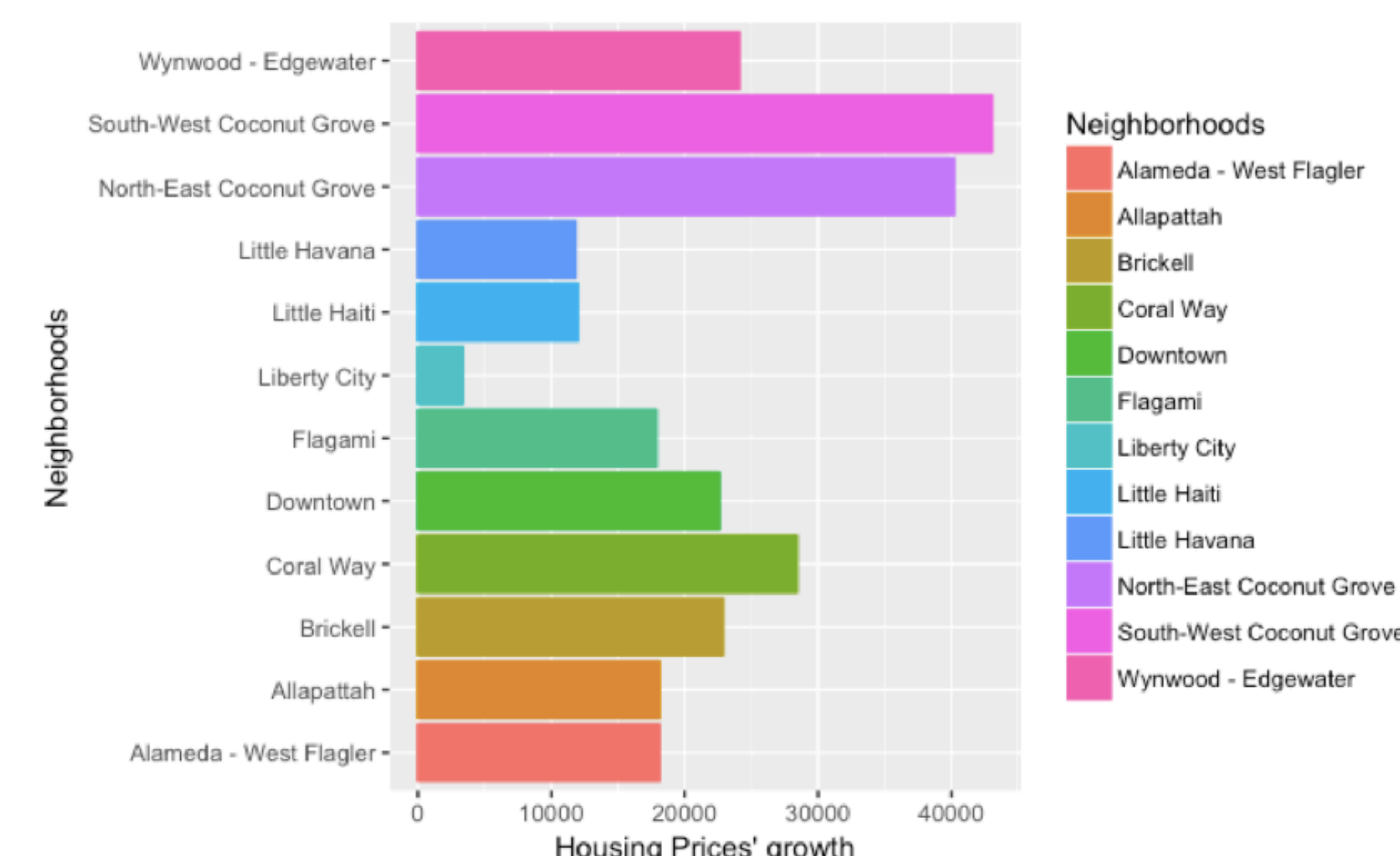Fig 6: Housing Prices in Miami's neighborhood since 2011 (after the crisis).

Fig 7: Housing Prices growth

Sentiment Analysis

Sentiment analysis can be thought of as the exercise of taking a sentence, paragraph, document, or any piece of natural language, and determining whether that text's emotional tone is positive, negative or neutral.

```
score.sentiment = function(tweets, pos.words, neg.words){
require(plyr)
require(stringr)
scores = laply(tweets, function(tweet, pos.words, neg.words) {
tweet = gsub('https://','',tweet)
tweet = gsub('http://','',tweet)
tweet=gsub('[^[:graph:]]', ' ',tweet)
tweet = gsub('[[:punct:]]', '', tweet)
tweet = gsub('[[:cntrl:]]', '', tweet)
tweet = gsub('\\d+', '', tweet) # removes numbers
tweet=str_replace_all(tweet,"[^[:graph:]]", " ")
tweet = tolower(tweet)
word.list = str_split(tweet, '\\s+')
words = unlist(word.list)
pos.matches = match(words, pos.words)
neg.matches = match(words, neg.words)
pos.matches = !is.na(pos.matches)
neg.matches = !is.na(neg.matches)
score = sum(pos.matches) - sum(neg.matches)
return(score)
}, pos.words, neg.words )
scores.df = data.frame(score=scores, text=tweets)
return(scores.df)
}
```

Fig 8. Sentiment Analysis. We used R to extract information from the social media platform Twitter, and to perform a sentiment analysis of what users tweeted about the different neighborhoods. Tweets from July, 2017 that mentioned any of the neighborhoods considered in this study.

Fig 9. Sentiment Analysis. Public opinion of each neighborhood via Twitter (data from July 2017)
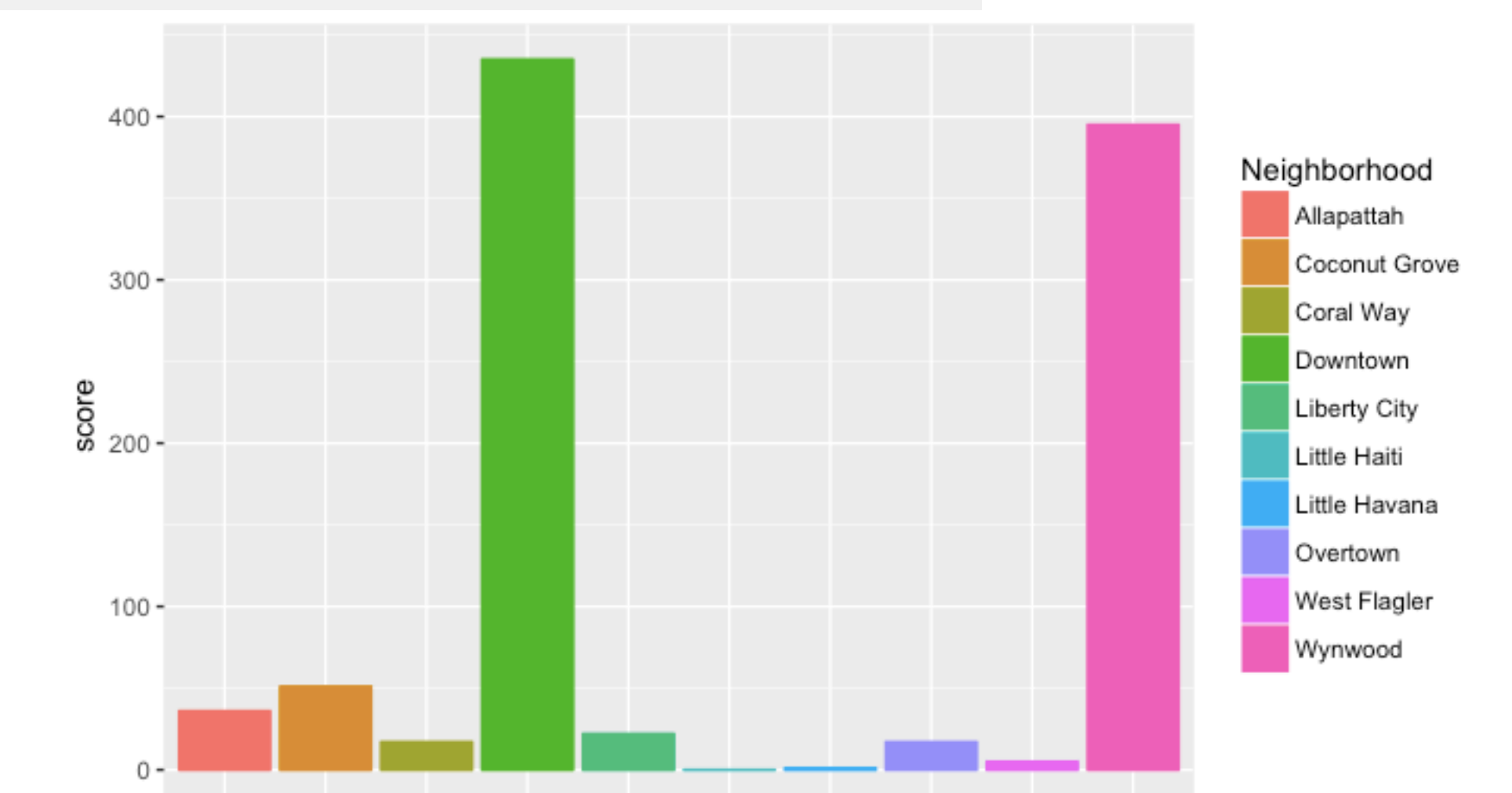
Conclusions

Using open access datasets and data mining techniques, results from our analysis confirm the high development standards of the Downtown, Wynwood, Coconut Grove, and Brickell neighborhoods. Further research including data on crime rates, demographics, tax incentives tools, public transportation and other indicators, will help to build a forecasting model that can be used to predict gentrification in the city.

References

[1] R: free software environment for statistical computing and graphics. https://www.r-project.org/
[2] tidyverse: an opinionated collection of R packages designed for data science. https://www.tidyverse.org/

[3] Miami-Dade County GIS Open Data http://gis-mdc.opendata.arcgis.com/datasets/
[4] Zillow Data. https://www.zillow.com/research/data/#median-home-value

Acknowledgements