CaliFile

B

2022-12-13

CaliFile Markdown

This Markdown File is the California Rent Analysis. This includes the needed code to analyze rent drivers and other needed information.

First Section:

Loading Libraries and CSV/Excel Files Adding In Lists

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.2.2
```

```
## ── Attaching packages
## ───────────────────────────────────────
## tidyverse 1.3.2 ──
```

```
## ✔ ggplot2 3.4.0     ✔ purrr   0.3.5
## ✔ tibble  3.1.8     ✔ stringr 1.5.0
## ✔ tidyr   1.2.1     ✔ forcats 0.5.2
## ✔ readr   2.1.3
```

```
## Warning: package 'ggplot2' was built under R version 4.2.2
```

```
## Warning: package 'stringr' was built under R version 4.2.2
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
library(readxl)
library(ggplot2)
library(shinydashboard)
```

```
## Warning: package 'shinydashboard' was built under R version 4.2.2
```

```
##
## Attaching package: 'shinydashboard'
##
## The following object is masked from 'package:graphics':
##
##     box
```

```
library(imputeTS)
```

```
## Warning: package 'imputeTS' was built under R version 4.2.2
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
#Make sure to put in your own path to read the csv file
#Main Files to Use

cali_file <- read.csv("C:/Users/bliun/Desktop/Cali-File.csv")
cityStats <- read_excel("More.xlsx") #Contains Population Information by Year
colnames(cityStats)[9] <- "AvgPop"

#Run in console
#glimpse(cali_file)
#glimpse(cityStats)

#List of Unique Cities (This is Important as each city is recorded within each year timeframe)
cityList <- unique(cali_file$city)
cityList <- sort(cityList)

#Color and Number Lists for Cities
dummyCityList <- c(0:64)
colorList <- c(1:65)

#Needed Interpolation
cali_file$baths <- na_interpolation(cali_file$baths)
cali_file$baths <- round(cali_file$baths, digits = 2)
cali_file$sqft <- na_interpolation(cali_file$sqft)
cali_file$sqft <- round(cali_file$sqft, digits = 2)
```

DataFrame Adding

In this section, we are adding dataframe information by different data points such as: - County - City - Neighborhood - Year - Respecitve Means

```
#Information by Year
cali_2003 <- cali_file[cali_file$year == 2003,]
cali_2008 <- cali_file[cali_file$year == 2008,]
cali_2013 <- cali_file[cali_file$year == 2013,]
cali_2018 <- cali_file[cali_file$year == 2018,]

#Important Rent Price Averages
#Year Average Rent Price
meanYear <- aggregate(cali_file$price, by=list(cali_file$year), FUN=mean)
colnames(meanYear) <- c("Year", "MeanRentPrice")
meanYear$`MeanRentPrice` <- round(meanYear$`MeanRentPrice`, digits = 2)

#Neighborhood Average Rent Price
meanNhood <- aggregate(cali_file$price, by=list(cali_file$nhood), FUN=mean)
colnames(meanNhood) <- c("Neighborhood", "MeanRentPrice")
meanNhood$MeanRentPrice <- round(meanNhood$MeanRentPrice, digits = 2)

#County Average Rent Price
meanCounty <- aggregate(cali_file$price, by=list(cali_file$county), FUN=mean)
colnames(meanCounty) <- c("County", "MeanRentPrice")
meanCounty$MeanRentPrice <- round(meanCounty$MeanRentPrice, digits = 2)

#City Average Rent Price
meanCities <- aggregate(cali_file$price, by=list(cali_file$city), FUN=mean)
colnames(meanCities) <- c("City", "MeanRentPrice")
meanCities$MeanRentPrice <- round(meanCities$MeanRentPrice, digits = 2)

#Population Stats
cityStats <- cbind(cityStats, meanCities$MeanRentPrice)
colnames(cityStats)[10] <- "MeanRentPrice"

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

#Average Rent Price by City & Year
meanCities2003 <- aggregate(cali_2003$price, by=list(cali_2003$city), FUN=mean)
colnames(meanCities2003) <- c("City", "MeanRentPrice")
meanCities2003 <- data.frame(meanCities2003, 2003)
colnames(meanCities2003)[3] <- "Year"
meanCities2003$MeanRentPrice <- round(meanCities2003$MeanRentPrice, digits = 2)

meanCities2008 <- aggregate(cali_2008$price, by=list(cali_2008$city), FUN=mean)
colnames(meanCities2008) <- c("City", "MeanRentPrice")
meanCities2008 <- data.frame(meanCities2008, 2008)
colnames(meanCities2008)[3] <- "Year"
meanCities2008$MeanRentPrice <- round(meanCities2008$MeanRentPrice, digits = 2)

meanCities2013 <- aggregate(cali_2013$price, by=list(cali_2013$city), FUN=mean)
colnames(meanCities2013) <- c("City", "MeanRentPrice")
meanCities2013 <- data.frame(meanCities2013, 2013)
colnames(meanCities2013)[3] <- "Year"
meanCities2013$MeanRentPrice <- round(meanCities2013$MeanRentPrice, digits = 2)

meanCities2018 <- aggregate(cali_2018$price, by=list(cali_2018$city), FUN=mean)
colnames(meanCities2018) <- c("City", "MeanRentPrice")
meanCities2018 <- data.frame(meanCities2018, 2018)
colnames(meanCities2018)[3] <- "Year"
meanCities2018$MeanRentPrice <- round(meanCities2018$MeanRentPrice, digits = 2)

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

#Average Rent Price By County & Year
meanCounty2003 <- aggregate(cali_2003$price, by=list(cali_2003$county), FUN=mean)
colnames(meanCounty2003) <- c("County", "MeanRentPrice")
meanCounty2003 <- data.frame(meanCounty2003, 2003)
colnames(meanCounty2003)[3] <- "Year"
meanCounty2003$MeanRentPrice <- round(meanCounty2003$MeanRentPrice, digits = 2)


meanCounty2008 <- aggregate(cali_2008$price, by=list(cali_2008$county), FUN=mean)
colnames(meanCounty2008) <- c("County", "MeanRentPrice")
meanCounty2008 <- data.frame(meanCounty2008, 2008)
colnames(meanCounty2008)[3] <- "Year"
meanCounty2008$MeanRentPrice <- round(meanCounty2008$MeanRentPrice, digits = 2)

meanCounty2013 <- aggregate(cali_2013$price, by=list(cali_2013$county), FUN=mean)
colnames(meanCounty2013) <- c("County", "MeanRentPrice")
meanCounty2013 <- data.frame(meanCounty2013, 2013)
colnames(meanCounty2013)[3] <- "Year"
meanCounty2013$MeanRentPrice <- round(meanCounty2013$MeanRentPrice, digits = 2)

meanCounty2018 <- aggregate(cali_2018$price, by=list(cali_2018$county), FUN=mean)
colnames(meanCounty2018) <- c("County", "MeanRentPrice")
meanCounty2018 <- data.frame(meanCounty2018, 2018)
colnames(meanCounty2018)[3] <- "Year"
meanCounty2018$MeanRentPrice <- round(meanCounty2018$MeanRentPrice, digits = 2)

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

#Average Rent Price By Neighborhood & Year
meanNhood2003 <- aggregate(cali_2003$price, by=list(cali_2003$nhood), FUN=mean)
colnames(meanNhood2003) <- c("Neighborhood", "MeanRentPrice")
meanNhood2003 <- data.frame(meanNhood2003, 2003)
colnames(meanNhood2003)[3] <- "Year"
meanNhood2003$MeanRentPrice <- round(meanNhood2003$MeanRentPrice, digits = 2)

meanNhood2008 <- aggregate(cali_2008$price, by=list(cali_2008$nhood), FUN=mean)
colnames(meanNhood2008) <- c("Neighborhood", "MeanRentPrice")
meanNhood2008 <- data.frame(meanNhood2008, 2008)
colnames(meanNhood2008)[3] <- "Year"
meanNhood2008$MeanRentPrice <- round(meanNhood2008$MeanRentPrice, digits = 2)

meanNhood2013 <- aggregate(cali_2013$price, by=list(cali_2013$nhood), FUN=mean)
colnames(meanNhood2013) <- c("Neighborhood", "MeanRentPrice")
meanNhood2013 <- data.frame(meanNhood2013, 2013)
colnames(meanNhood2013)[3] <- "Year"
meanNhood2013$MeanRentPrice <- round(meanNhood2013$MeanRentPrice, digits = 2)

meanNhood2018 <- aggregate(cali_2018$price, by=list(cali_2018$nhood), FUN=mean)
colnames(meanNhood2018) <- c("Neighborhood", "MeanRentPrice")
meanNhood2018 <- data.frame(meanNhood2018, 2018)
colnames(meanNhood2018)[3] <- "Year"
meanNhood2018$MeanRentPrice <- round(meanNhood2018$MeanRentPrice, digits = 2)

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

#Combined Dataframes (Location type by year)
meanCitiesYear <- rbind(meanCities2003, meanCities2008, meanCities2013, meanCities2018)
meanCountyYear <- rbind(meanCounty2003, meanCounty2008, meanCounty2013, meanCounty2018)
meanNhoodYear <- rbind(meanNhood2003, meanNhood2008, meanNhood2013, meanNhood2018)
```

##Lineplots by Location and Year

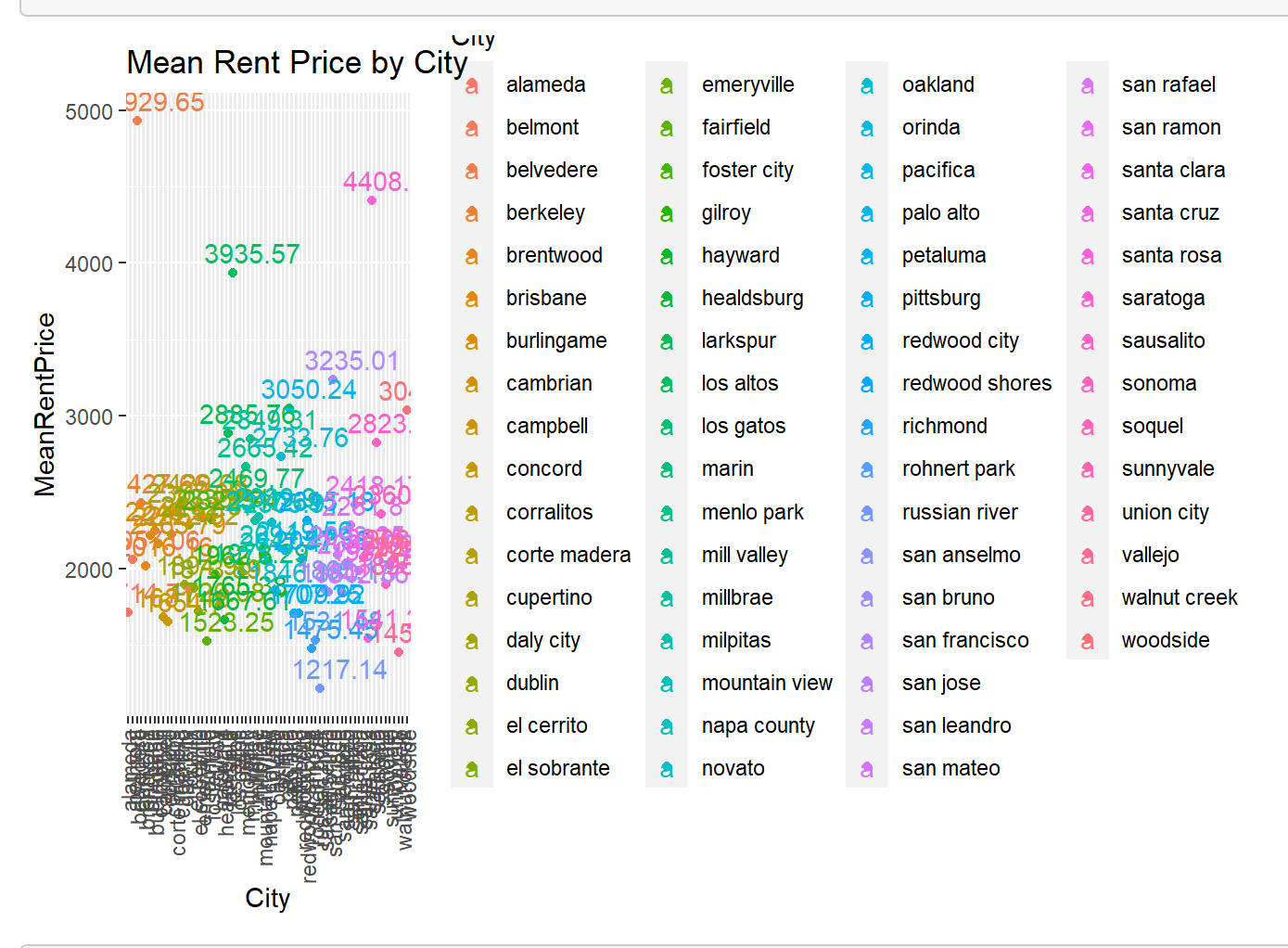
```
meanCitiesYearLine <- ggplot(meanCitiesYear, aes(x = Year, y = MeanRentPrice)) + geom_line(aes(group = City, colo
r = City)) +
  geom_point(aes(color = City)) + ggtitle("California Average Rent by City (2013-2018)")

meanCountyYearLine <- ggplot(meanCountyYear, aes(x = Year, y = MeanRentPrice)) + geom_line(aes(group = County, co
lor = County)) +
  geom_point(aes(color = County)) + ggtitle("California Average Rent by County (2013-2018)")

meanNhoodYearLine <- ggplot(meanNhoodYear, aes(x = Year, y = MeanRentPrice)) +
  geom_line(aes(group = Neighborhood, color = Neighborhood)) +
  geom_point(aes(color = Neighborhood)) + ggtitle("California Average Rent by City (2013-2018)")

#Line Chart for Average Rent Price by Recorded Year
meanYear_line <- ggplot(meanYear) + aes(x = Year, y = MeanRentPrice) + geom_line(color = "blue") + geom_point(col
or = "black")
meanYear_line <- meanYear_line + ggtitle("Average Rent Price By Recorded Year")
```

##Statistical Graphics

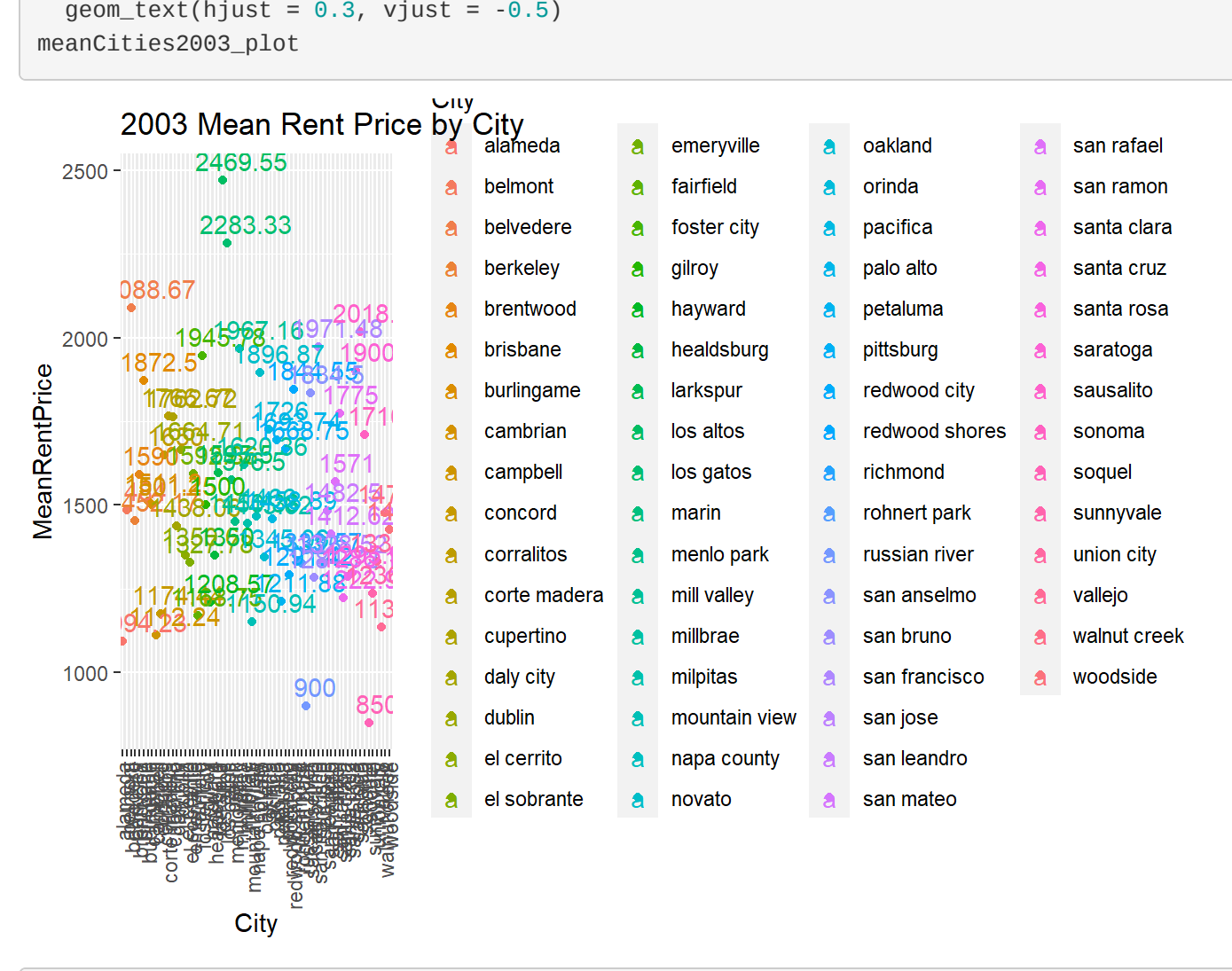
```
#Means by City & Year Boxplots
meanCities_boxplot <- ggplot(meanCities) + geom_boxplot(aes(x = MeanRentPrice), color = "black", fill = "#9cfff2"
) +
  ggtitle("California Average Rent Boxplot")

meanCities2003_boxplot <- ggplot(meanCities2003) + geom_boxplot(aes(x = MeanRentPrice2003), color = "black", fill
= "#c7ffb3") +
  ggtitle("2003 California Average Rent Boxplot")

meanCities2008_boxplot <- ggplot(meanCities2008) + geom_boxplot(aes(x = MeanRentPrice2008), color = "black", fill
= "#fff385") +
  ggtitle("2008 California Average Rent Boxplot")

meanCities2013_boxplot <- ggplot(meanCities2013) + geom_boxplot(aes(x = MeanRentPrice2013), color = "black", fill
= "#ff9a52") +
  ggtitle("2013 California Average Rent Boxplot")

meanCities2018_boxplot <- ggplot(meanCities2018) + geom_boxplot(aes(x = MeanRentPrice2018), color = "black", fill
= "#ffa6f2") +
  ggtitle("2018 California Average Rent Boxplot")

#Histogram of California Rent Prices across 2003 - 2018
cali_file_hist <- ggplot(cali_file) + geom_histogram(aes(x = price), fill = "white", color = "black", binwidth =
200)
cali_file_hist <- cali_file_hist + ggtitle("California Price Rent Histogram")
```

##Graphing Plots

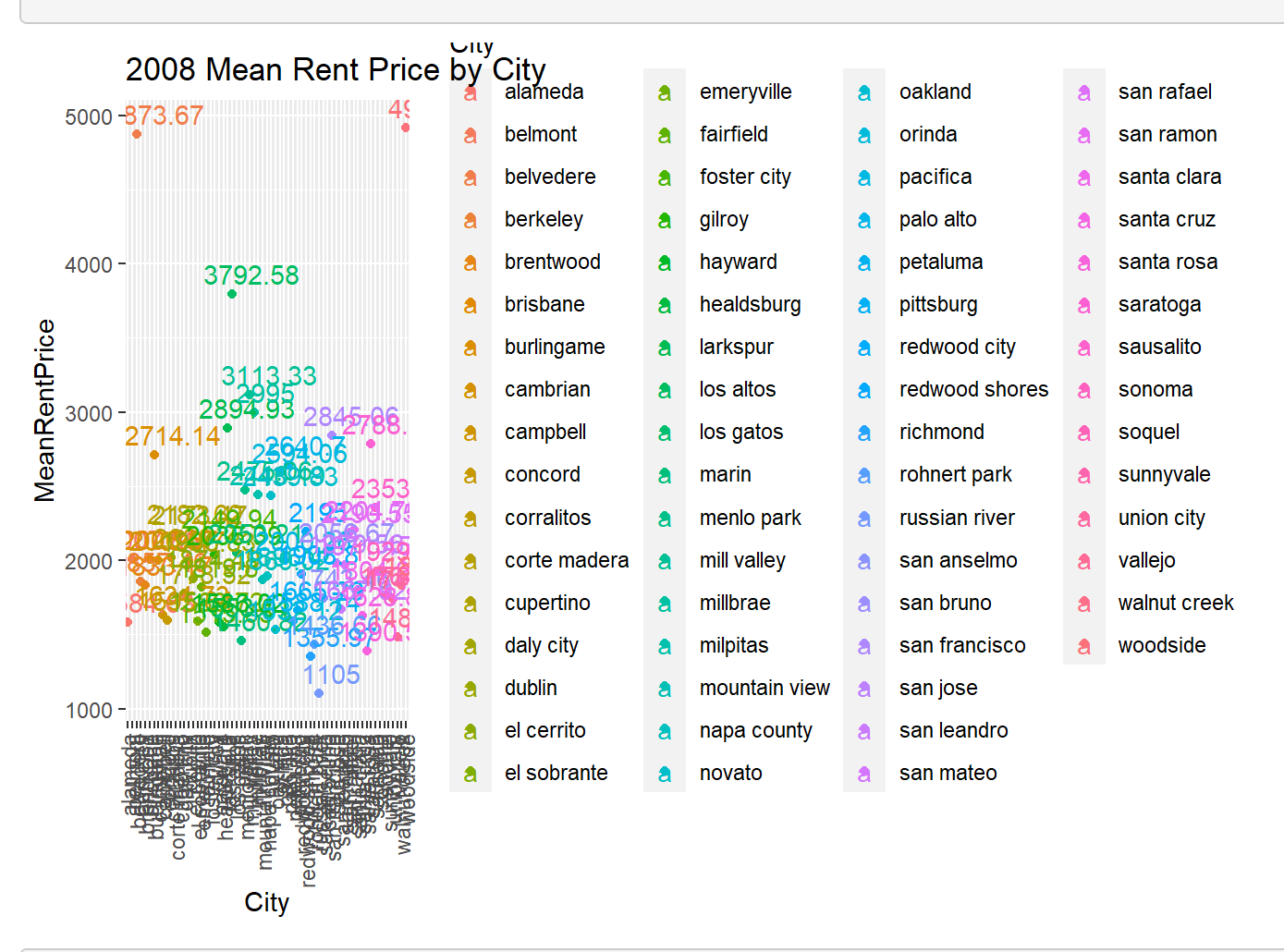
```
#Mean Rent Price by City Plot
meanCities_plot <- ggplot(meanCities, aes(x = City, y = MeanRentPrice, color = City, label = MeanRentPrice)) + ge
om_point() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) + ggtitle("Mean Rent Price by City") +
  geom_text(hjust = 0.3, vjust = -0.5)
meanCities_plot
```



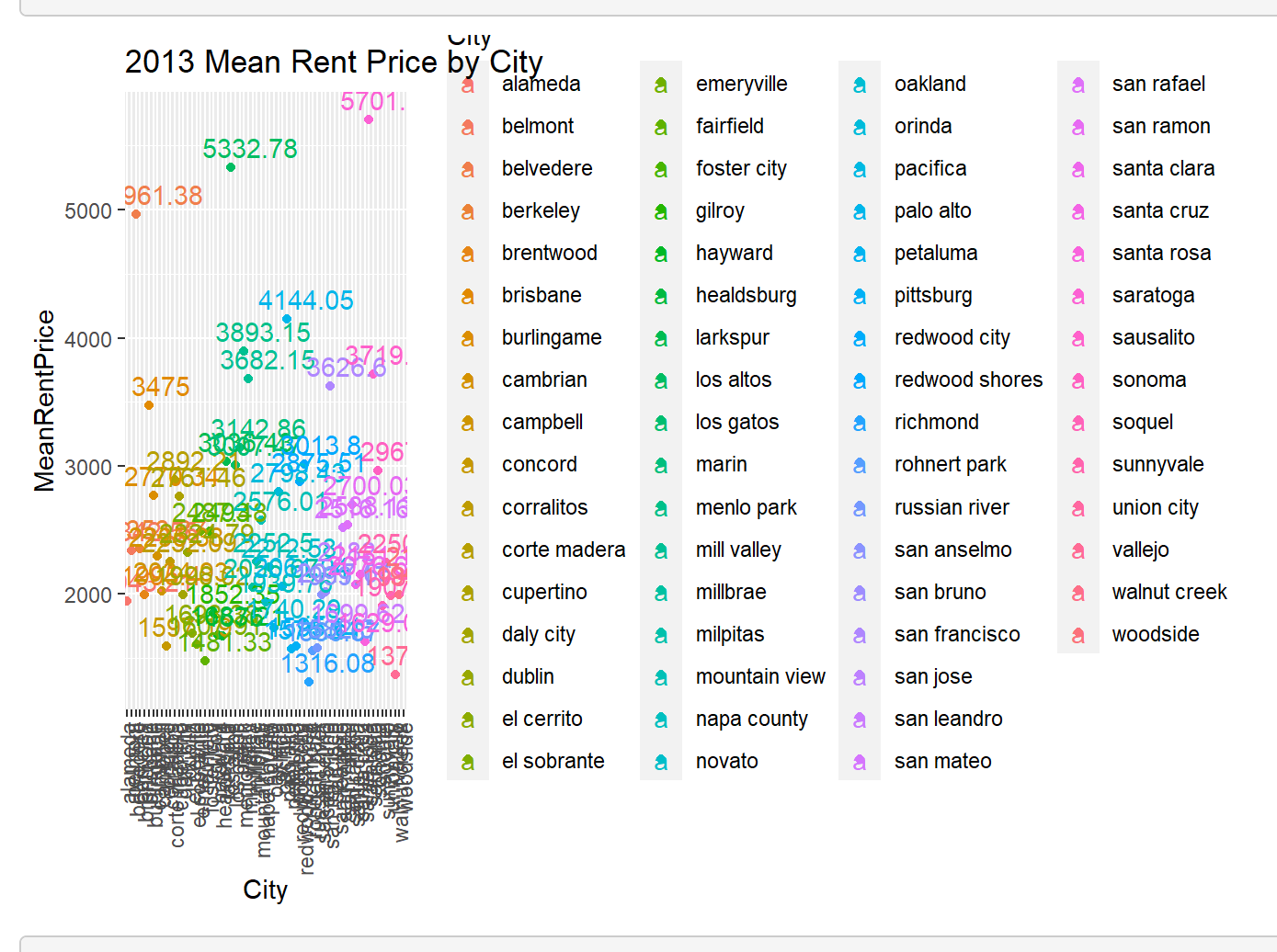

```
meanCities2003_plot <- ggplot(meanCities2003, aes(x = City, y = MeanRentPrice, color = City, label = MeanRentPric
e)) +
  geom_point() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) + ggtitle("2003 Mean Rent Price by City") +
  geom_text(hjust = 0.3, vjust = -0.5)
meanCities2003_plot
```



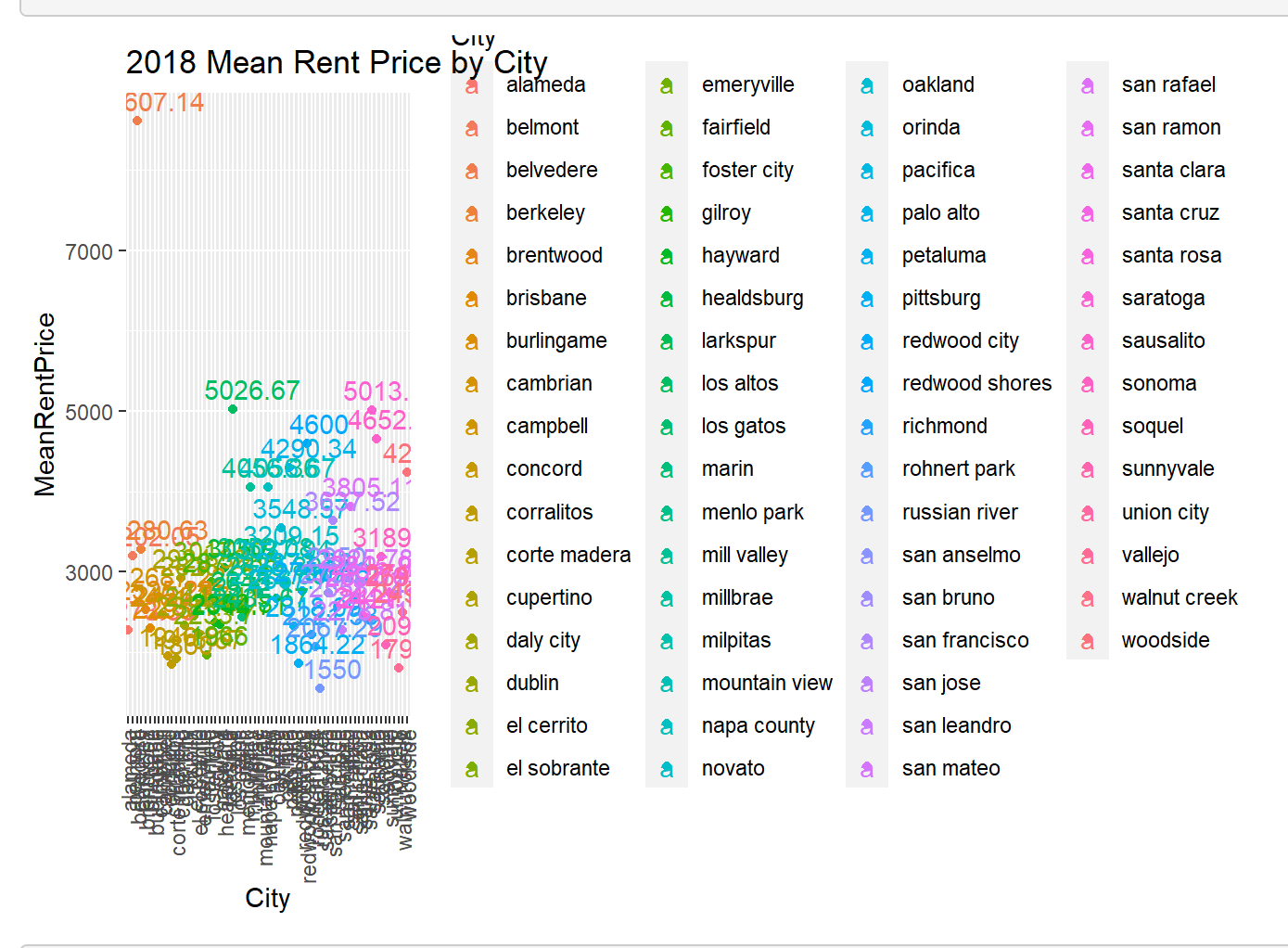

```
meanCities2008_plot <- ggplot(meanCities2008, aes(x = City, y = MeanRentPrice, color = City, label = MeanRentPric
e)) +
  geom_point() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) + ggtitle("2008 Mean Rent Price by City") +
  geom_text(hjust = 0.3, vjust = -0.5)
meanCities2008_plot
```




```
meanCities2013_plot <- ggplot(meanCities2013, aes(x = City, y = MeanRentPrice, color = City, label = MeanRentPric
e)) +
  geom_point() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) + ggtitle("2013 Mean Rent Price by City") +
  geom_text(hjust = 0.3, vjust = -0.5)
meanCities2013_plot
```




```
meanCities2018_plot <- ggplot(meanCities2018, aes(x = City, y = MeanRentPrice, color = City, label = MeanRentPric
e)) +
  geom_point() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) + ggtitle("2018 Mean Rent Price by City") +
  geom_text(hjust = 0.3, vjust = -0.5)
meanCities2018_plot
```




```
#Average Population vs Mean Price by City
cityStats <- cbind(cityStats, meanCities$MeanRentPrice) #DO NOT RUN THIS LINE MULTIPLE TIMES IN THE SAME SESSION
(RESET NEEDED IF DONE)
colnames(cityStats)[10] <- "MeanRentPrice"
cityStats <- cityStats[-11]
cityStatsPlot <- ggplot(cityStats, aes(x = AvgPop, y = MeanRentPrice, label = City, color = City))
cityStatsPlot <- cityStatsPlot + geom_point() + geom_text(hjust = 0, vjust = 0)
cityStatsPlot <- cityStatsPlot + coord_cartesian() + ggtitle("AvgPop vs Mean Price by City")
cityStatsPlot
```

##Linear Regression

In this section, we have the linear regression results: formula = **price ~ each city**

```
#Linear Model/Linear Regression to see whether or not city influences price
cali_file.city.lm <- summary(lm(price ~ city, data = cali_file))
cali_2003.city.lm <- summary(lm(price ~ city, data = cali_2003))
cali_2008.city.lm <- summary(lm(price ~ city, data = cali_2008))
cali_2013.city.lm <- summary(lm(price ~ city, data = cali_2013))
cali_2018.city.lm <- summary(lm(price ~ city, data = cali_2018))

cali_file.nhood.lm <- summary(lm(price ~ nhood, data = cali_file))
cali_2003.nhood.lm <- summary(lm(price ~ nhood, data = cali_2003))
cali_2008.nhood.lm <- summary(lm(price ~ nhood, data = cali_2008))
cali_2013.nhood.lm <- summary(lm(price ~ nhood, data = cali_2013))
cali_2018.nhood.lm <- summary(lm(price ~ nhood, data = cali_2018))

cali_file.county.lm <- summary(lm(price ~ county, data = cali_file))
cali_2003.county.lm <- summary(lm(price ~ county, data = cali_2003))
cali_2008.county.lm <- summary(lm(price ~ county, data = cali_2008))
cali_2013.county.lm <- summary(lm(price ~ county, data = cali_2013))
cali_2018.county.lm <- summary(lm(price ~ county, data = cali_2018))
```