

```
## 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

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

#Run in console
#limpse(call_file)
#limpse(cityStats)

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

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

```
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 data frame information by different data points such as - County - City - Neighborhood - Year - Respective Means

```
#Information by Year
call_2003 <- call_file(call_file$year == 2003,)
call_2008 <- call_file(call_file$year == 2008,)
call_2013 <- call_file(call_file$year == 2013,)
call_2018 <- call_file(call_file$year == 2018,)
```

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

```
#Neighborhood Average Rent Price
meanHood <- aggregate(call_file$price, by=list(call_file$hhood), FUN=mean)
colnames(meanHood) <- c("Neighborhood", "MeanRentPrice")
meanHood$MeanRentPrice <- round(meanHood$MeanRentPrice, digits = 2)
```

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

```
#City Average Rent Price
meanCities <- aggregate(call_file$price, by=list(call_file$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(call_2003$price, by=list(call_2003$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(call_2008$price, by=list(call_2008$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(call_2013$price, by=list(call_2013$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(call_2018$price, by=list(call_2018$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(call_2003$price, by=list(call_2003$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(call_2008$price, by=list(call_2008$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(call_2013$price, by=list(call_2013$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(call_2018$price, by=list(call_2018$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
meanHood2003 <- aggregate(call_2003$price, by=list(call_2003$hhood), FUN=mean)
colnames(meanHood2003) <- c("Neighborhood", "MeanRentPrice")
meanHood2003 <- data.frame(meanHood2003, 2003)
colnames(meanHood2003)[3] <- "Year"
meanHood2003$MeanRentPrice <- round(meanHood2003$MeanRentPrice, digits = 2)

meanHood2008 <- aggregate(call_2008$price, by=list(call_2008$hhood), FUN=mean)
colnames(meanHood2008) <- c("Neighborhood", "MeanRentPrice")
meanHood2008 <- data.frame(meanHood2008, 2008)
colnames(meanHood2008)[3] <- "Year"
meanHood2008$MeanRentPrice <- round(meanHood2008$MeanRentPrice, digits = 2)

meanHood2013 <- aggregate(call_2013$price, by=list(call_2013$hhood), FUN=mean)
colnames(meanHood2013) <- c("Neighborhood", "MeanRentPrice")
meanHood2013 <- data.frame(meanHood2013, 2013)
colnames(meanHood2013)[3] <- "Year"
meanHood2013$MeanRentPrice <- round(meanHood2013$MeanRentPrice, digits = 2)

meanHood2018 <- aggregate(call_2018$price, by=list(call_2018$hhood), FUN=mean)
colnames(meanHood2018) <- c("Neighborhood", "MeanRentPrice")
meanHood2018 <- data.frame(meanHood2018, 2018)
colnames(meanHood2018)[3] <- "Year"
meanHood2018$MeanRentPrice <- round(meanHood2018$MeanRentPrice, digits = 2)

#-----#

#Combined DataFrames (Location type by year)
meanCitiesYear <- rbind(meanCities2003, meanCities2008, meanCities2013, meanCities2018)
meanCountyYear <- rbind(meanCounty2003, meanCounty2008, meanCounty2013, meanCounty2018)
meanHoodYear <- rbind(meanHood2003, meanHood2008, meanHood2013, meanHood2018)
```

```
##Lineplots by Location and Year

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

meanHoodYearLineLine <- ggplot(meanHoodYear, 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(color = "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 = "#9cf2ff") +
  ggtitle("California Average Rent Boxplot")

meanCities2003_boxplot <- ggplot(meanCities2003) + geom_boxplot(aes(x = MeanRentPrice2003), color = "black", fill = "#9cf2ff") +
  ggtitle("2003 California Average Rent Boxplot")

meanCities2008_boxplot <- ggplot(meanCities2008) + geom_boxplot(aes(x = MeanRentPrice2008), color = "black", fill = "#9cf2ff") +
  ggtitle("2008 California Average Rent Boxplot")

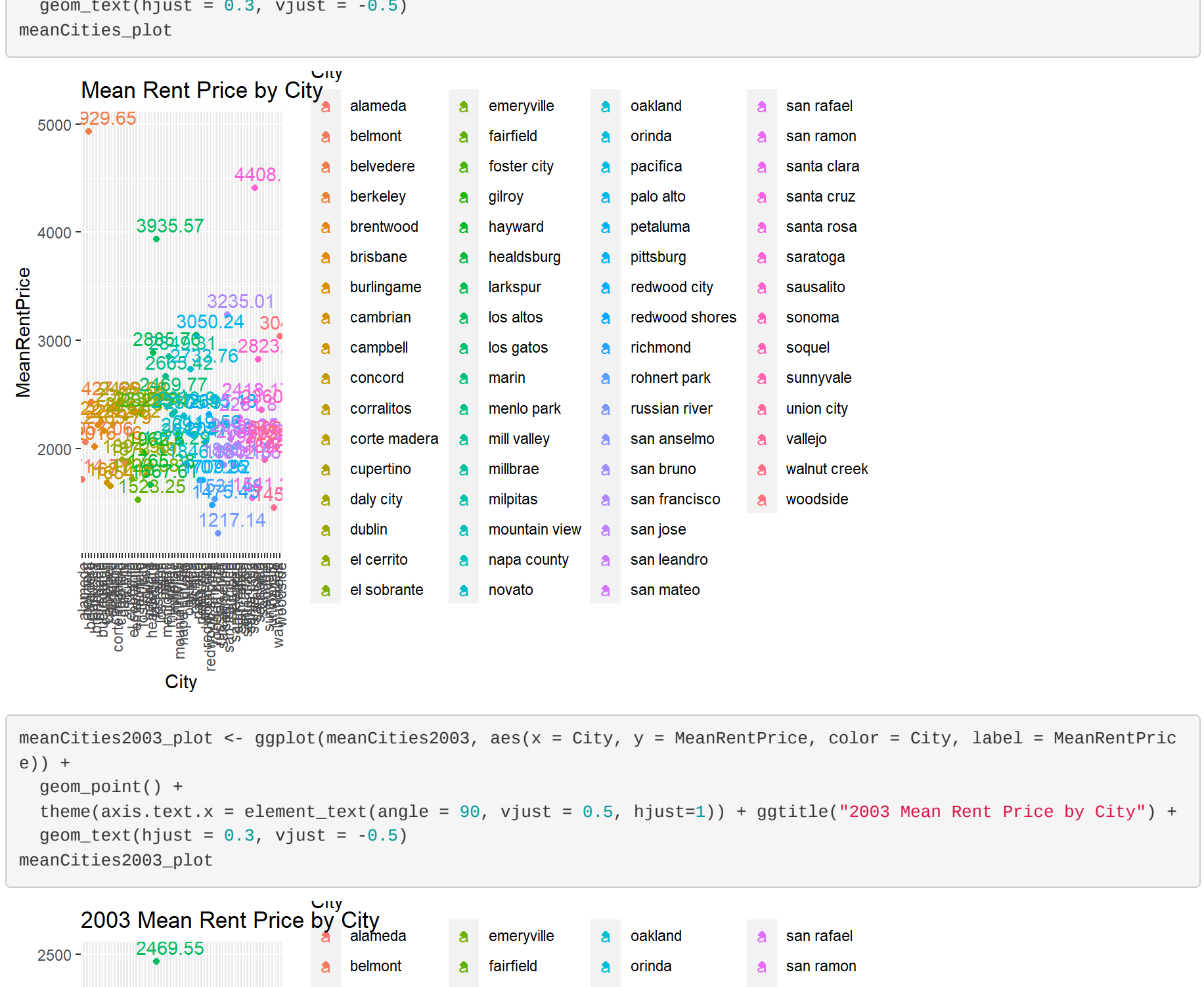
meanCities2013_boxplot <- ggplot(meanCities2013) + geom_boxplot(aes(x = MeanRentPrice2013), color = "black", fill = "#9cf2ff") +
  ggtitle("2013 California Average Rent Boxplot")

meanCities2018_boxplot <- ggplot(meanCities2018) + geom_boxplot(aes(x = MeanRentPrice2018), color = "black", fill = "#9cf2ff") +
  ggtitle("2018 California Average Rent Boxplot")

#Histogram of California Rent Prices across 2003 - 2018
call_file_hist <- ggplot(call_file) + geom_histogram(aes(x = price), fill = "white", color = "black", binwidth = 200)
call_file_hist <- call_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)) +
  geom_point() +
```



Year	City	Price
2000	Belvedere	2283.33
2000	Berkley	188.67
2000	Brentwood	194.56
2000	Brisbane	1872.5
2000	Burlingame	1762.62
2000	Cambrian	1684.71
2000	Foster City	1687.49
2000	Gilroy	1596.37
2000	Hayward	1594.45
2000	Headlandsburg	1226.17
2000	Larkspur	1226.76
2000	Los Altos	1226.76
2000	Pacific	1226.76
2000	Palo Alto	1226.76
2000	Petaluma	1226.76
2000	Pittsburg	1226.76
2000	Redwood City	1226.76
2000	Redwood Shores	1226.76
2000	Santa Clara	1226.76
2000	Santa Cruz	1226.76
2000	Santa Rosa	1226.76
2000	Saratoga	1226.76
2000	Sausalito	1226.76
2000	Sonoma	1226.76



	berkeley	gilroy	palo alto	santa cruz
rentPrice	3792.58			
	brentwood	hayward	petaluma	santa rosa
	brisbane	healdsburg	pittsburg	saratoga
	burlingame	larkspur	redwood city	sausello
	cambrian	los altos	redwood shores	sonoma
	campbell	los gatos	richmond	soquel
	daly city	san geronimo	san jose	stanford

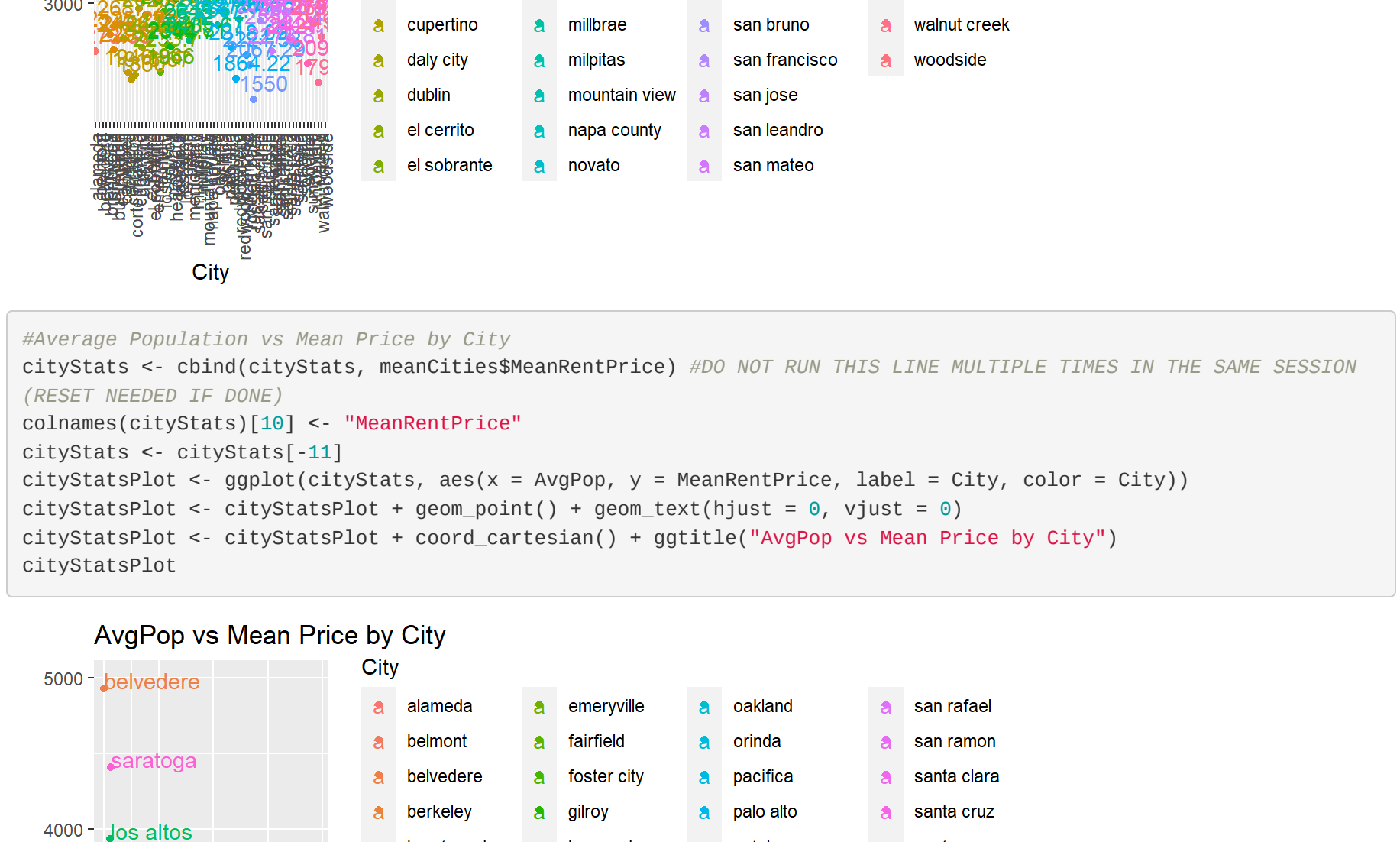


City	Year	MeanRent/PctChg
brentwood	2014	4144.05
hayward	2014	3993.15
petaluma	2014	3982.15
santa rosa	2014	3719
brisbane	2014	3475
healdsburg	2014	3142.86
pittsburg	2014	3065.42
saratoga	2014	2996
burlingame	2014	2892
redwood city	2014	2769.49
san jose	2014	2769.49
sausalito	2014	2769.49
cambrian	2014	2769.49
redwood shores	2014	2769.49
sonoma	2014	2769.49
los altos	2014	2769.49
richmond	2014	2769.49
soquel	2014	2769.49
campbell	2014	2769.49
rohnert park	2014	2769.49
concord	2014	2769.49
marin	2014	2769.49
sunnyvale	2014	2769.49



MeanRentPerSqFt vs City

City	Neighborhood	MeanRentPerSqFt (approx.)
San Francisco	burlingame	4800
San Francisco	larkspur	4800
San Francisco	redwood city	4800
San Francisco	sausalito	4800
San Jose	cambrian	4500
San Jose	los altos	4500
San Jose	redwood shores	4500
San Jose	sonoma	4500
San Jose	campbell	4200
San Jose	los gatos	4200
San Jose	richmond	4200
San Jose	soquel	4200
San Jose	concord	4000
San Jose	marin	4000
San Jose	rohnert park	4000
San Jose	sunnyvale	4000
San Jose	corralitos	3800
San Jose	merit park	3800
San Jose	russian river	3800
San Jose	union city	3800
San Jose	corte madera	3500
San Jose	mill valley	3500
San Jose	san anselmo	3500
San Jose	vallejo	3500



4000-	los altos	brentwood	hayward	petaluma
0		brisbane	healdsburg	pittsburg

	united	newberry	pinetree	sanford
burlingame				
cambrian				
larkspur				
los altos				
redwood city				
redwood shores				

Mean depth (m)

Area (km²)

###Linear Regression

Location	Area (km²)	Mean depth (m)
san mateo	1000	2800
san leandro	1000	2700
san jose	1000	2600
san francisco	1000	2500
san carlos	1000	2400
san antonio	1000	2300
san diego	1000	2200
san gabriel	1000	2100
san rafael	1000	2000
san joaquin	1000	1900
san benito	1000	1800
san luis obispo	1000	1700
san bernardino	1000	1600
san gabriel	1000	1500
san rafael	1000	1400
san joaquin	1000	1300
san benito	1000	1200
san luis obispo	1000	1100
san bernardino	1000	1000
san gabriel	1000	900
san rafael	1000	800
san joaquin	1000	700
san benito	1000	600
san luis obispo	1000	500
san bernardino	1000	400
san gabriel	1000	300
san rafael	1000	200
san joaquin	1000	100
san benito	1000	0
san luis obispo	1000	0
san bernardino	1000	0
san gabriel	1000	0
san rafael	1000	0
san joaquin	1000	0
san benito	1000	0
san luis obispo	1000	0
san bernardino	1000	0
san gabriel	1000	0
san rafael	1000	0
san joaquin	1000	0
san benito	1000	0
san luis obispo	1000	0
san bernardino	1000	0
san gabriel	1000	0
san rafael	1000	0
san joaquin	1000	0
san benito	1000	0
san luis obispo	1000	0
san bernardino	1000	0
san gabriel	1000	0
san rafael	1000	0
san joaquin	1000	0
san benito	1000	0
san luis obispo	1000	0
san bernardino	1000	0
san gabriel	1000	0
san rafael	1000	0
san joaquin	1000	0
san benito	1000	0
san luis obispo	1000	0
san bernardino	1000	0
san gabriel	1000	0
san rafael	1000	0
san joaquin	1000	0
san benito	1000	0
san luis obispo	1000	0
san bernardino	1000	0
san gabriel	1000	0
san rafael	1000	0
san joaquin	1000	0
san benito	1000	0
san luis obispo	1000	0
san bernardino	1000	0
san gabriel	1000	0
san rafael	1000	0
san joaquin	1000	0
san benito	1000	0
san luis obispo	1000	0
san bernardino	1000	0
san gabriel	1000	0
san rafael	1000	0
san joaquin	1000	0
san benito	1000	0
san luis obispo	1000	0
san bernardino	1000	0
san gabriel	1000	0
san rafael	1000	0
san joaquin	1000	0
san benito	1000	0
san luis obispo	1000	0
san bernardino	1000	0
san gabriel	1000	0
san rafael	1000	0
san joaquin	1000	0
san benito	1000	0
san luis obispo	1000	0
san bernardino	1000	0
san gabriel	1000	0
san rafael	1000	0
san joaquin	1000	0
san benito	1000	0
san luis obispo	1000	0
san bernardino	1000	0
san gabriel	1000	0
san rafael	1000	0
san joaquin	1000	0
san benito	1000	0
san luis obispo	1000	0
san bernardino	1000	0
san gabriel	1000	0
san rafael	1000	0
san joaquin	1000	0
san benito	1000	0
san luis obispo	1000	0
san bernardino	1000	0
san gabriel	1000	0
san rafael	1000	0
san joaquin	1000	0
san benito	1000	0
san luis obispo	1000	0
san bernardino	1000	0
san gabriel	1000	0
san rafael	1000	0
san joaquin	1000	0
san benito	1000	0
san luis obispo	1000	0
san bernardino	1000	0
san gabriel	1000	0
san rafael	1000	0
san joaquin	1000	0
san benito	1000	0
san luis obispo	1000	0
san bernardino	1000	0
san gabriel	1000	0
san rafael	1000	0</

```

#Linear Model/Linear Regression to see whether or not city influences price
call_1903.city.lm <- summary(lm(price ~ city, data = call_1file))
call_2003.city.lm <- summary(lm(price ~ city, data = call_2file))
call_2008.city.lm <- summary(lm(price ~ city, data = call_3file))
call_2013.city.lm <- summary(lm(price ~ city, data = call_4file))
call_2018.city.lm <- summary(lm(price ~ city, data = call_5file))

call_1file.rhoad.lm <- summary(lm(price ~ rhoad, data = call_1file))
call_2003.rhoad.lm <- summary(lm(price ~ rhoad, data = call_2file))

```

```
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))
```