DS311 - R Lab Assignment

Alexia Kim

2024-11-09

R Assignment 1

- In this assignment, we are going to apply some of the build in data set in R for descriptive statistics analysis.
- To earn full grade in this assignment, students need to complete the coding tasks for each question to get the result.
- After finished all the questions, knit the document into HTML format for submission.

Question 1

Using the **mtcars** data set in R, please answer the following questions.

```
# Loading the data
data(mtcars)

# Head of the data set
print(mtcars)
```

```
##                      mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4           21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag       21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710          22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive      21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout   18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Valiant             18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Duster 360          14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 240D           24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
## Merc 230            22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
## Merc 280            19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
## Merc 280C           17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
## Merc 450SE          16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
## Merc 450SL          17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
## Merc 450SLC         15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
## Cadillac Fleetwood  10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0    3    4
## Chrysler Imperial   14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## Fiat 128            32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
## Honda Civic         30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
## Toyota Corolla      33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
## Toyota Corona       21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
## Dodge Challenger    15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
## AMC Javelin         15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
```

```
## Camaro Z28          13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
## Pontiac Firebird    19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
## Fiat X1-9           27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
## Porsche 914-2       26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
## Lotus Europa        30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
## Ford Pantera L      15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
## Ferrari Dino        19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## Maserati Bora       15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
## Volvo 142E          21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
```

a. Report the number of variables and observations in the data set.

```
# Enter your code here!
ncol(mtcars)

## [1] 11

nrow(mtcars)

## [1] 32

# Answer:
print("There are total of __11___ variables and __32___ observations in this
data set.")

## [1] "There are total of __11___ variables and __32___ observations in this
data set."
```

b. Print the summary statistics of the data set and report how many discrete and continuous variables are in the data set.

```
# Enter your code here!
discrete=mtcars[c(2,4,8,9,10,11)]
ncol(discrete)*nrow(discrete)

## [1] 192

contin=mtcars[c(1,3,5,6,7)]
ncol(contin)*nrow(contin)

## [1] 160

# Answer:
print("There are __192___ discrete variables and __160___ continuous
variables in this data set.")

## [1] "There are __192___ discrete variables and __160___ continuous
variables in this data set."
```

c. Calculate the mean, variance, and standard deviation for the variable **mpg** and assign them into variable names m, v, and s. Report the results in the print statement.

```
# Enter your code here!
m=mean(mtcars$mpg)
```

```r
v=var(mtcars$mpg)

s=sd(mtcars$mpg)


print(paste("The average of Mile Per Gallon from this data set is ",m, " with
variance ", v , " and standard deviation", s, "."))

## [1] "The average of Mile Per Gallon from this data set is  20.090625  with
variance  36.3241028225806  and standard deviation 6.0269480520891 ."
```

d. Create two tables to summarize 1) average mpg for each cylinder class and 2) the standard deviation of mpg for each gear class.

```r
# Enter your code here!
#1
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

mtcars %>%
group_by(cyl) %>%
summarize(Avgmpg = mean(mpg))

## # A tibble: 3 × 2
##     cyl Avgmpg
##   <dbl>  <dbl>
## 1     4   26.7
## 2     6   19.7
## 3     8   15.1

#2
mtcars %>%
group_by(gear) %>%
summarize(Sdmpg = sd(mpg))

## # A tibble: 3 × 2
##    gear Sdmpg
##   <dbl> <dbl>
## 1     3  3.37
## 2     4  5.28
## 3     5  6.66
```

e. Create a crosstab that shows the number of observations belong to each cylinder and gear class combinations. The table should show how many observations given the car has 4 cylinders with 3 gears, 4 cylinders with 4 gears, etc. Report which combination is recorded in this data set and how many observations for this type of car.

```
# Enter your code here!
crosstab<-table(mtcars$cyl, mtcars$gear)
print(crosstab)

##
##      3  4  5
##   4  1  8  2
##   6  2  4  1
##   8 12  0  2

print("The most common car type in this data set is car with __8__ cylinders and __3__ gears. There are total of __32___ cars belong to this specification in the data set.")

## [1] "The most common car type in this data set is car with __8__ cylinders and __3__ gears. There are total of __32___ cars belong to this specification in the data set."
```

Question 2

Use different visualization tools to summarize the data sets in this question.

a. Using the **PlantGrowth** data set, visualize and compare the weight of the plant in the three separated group. Give labels to the title, x-axis, and y-axis on the graph. Write a paragraph to summarize your findings.

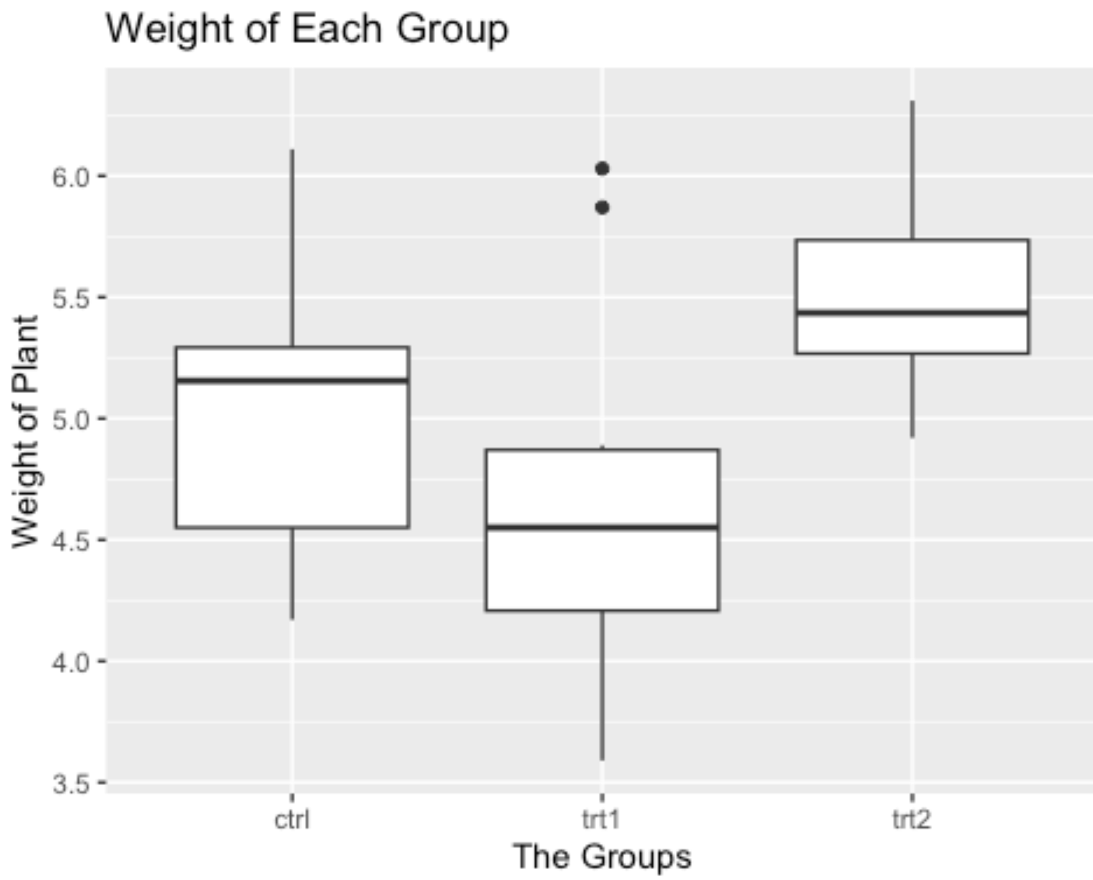
```
# Load the data set
data(PlantGrowth)
print(PlantGrowth)

##    weight group
## 1    4.17  ctrl
## 2    5.58  ctrl
## 3    5.18  ctrl
## 4    6.11  ctrl
## 5    4.50  ctrl
## 6    4.61  ctrl
## 7    5.17  ctrl
## 8    4.53  ctrl
## 9    5.33  ctrl
## 10   5.14  ctrl
## 11   4.81  trt1
## 12   4.17  trt1
## 13   4.41  trt1
```

```
## 14   3.59  trt1
## 15   5.87  trt1
## 16   3.83  trt1
## 17   6.03  trt1
## 18   4.89  trt1
## 19   4.32  trt1
## 20   4.69  trt1
## 21   6.31  trt2
## 22   5.12  trt2
## 23   5.54  trt2
## 24   5.50  trt2
## 25   5.37  trt2
## 26   5.29  trt2
## 27   4.92  trt2
## 28   6.15  trt2
## 29   5.80  trt2
## 30   5.26  trt2
```

```r
library(ggplot2)

# Head of the data set
head(PlantGrowth)
```

```
##   weight group
## 1   4.17  ctrl
## 2   5.58  ctrl
## 3   5.18  ctrl
## 4   6.11  ctrl
## 5   4.50  ctrl
## 6   4.61  ctrl
```

```r
# Enter your code here!
ggplot(PlantGrowth, aes(y=weight, x=group)) + geom_boxplot() +
ggtitle("Weight of Each Group") +labs(y="Weight of Plant", x="The Groups")
```



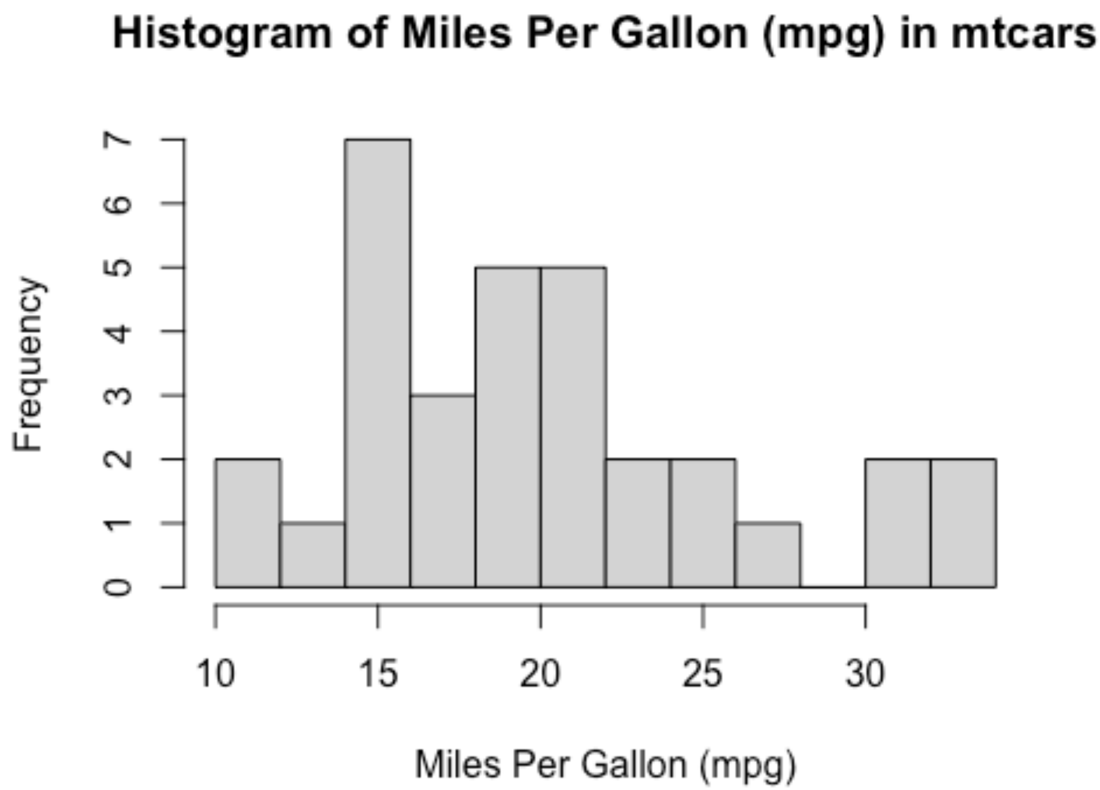
Result:

=> Report a paragraph to summarize your findings from the plot!

The line in in the boxes indicate the median. We can see from the boxplot graph that trt2 group has the highest median compared to the other groups.the ctrl group is the biggest box, meaning that the weight values for that group are more dispersed than the other groups. The ctrl group is left-skewed (negatively skewed), the trt1 group is symmetric, and trt2 is right-skewed. For trt1, there are outliers, meaning that there are some weight values that fall outside of the normal majority range.

b. Using the **mtcars** data set, plot the histogram for the column **mpg** with 10 breaks. Give labels to the title, x-axis, and y-axis on the graph. Report the most observed mpg class from the data set.

```
hist(mtcars$mpg,
     breaks = 10,
     main = "Histogram of Miles Per Gallon (mpg) in mtcars",
     xlab = "Miles Per Gallon (mpg)",)
```

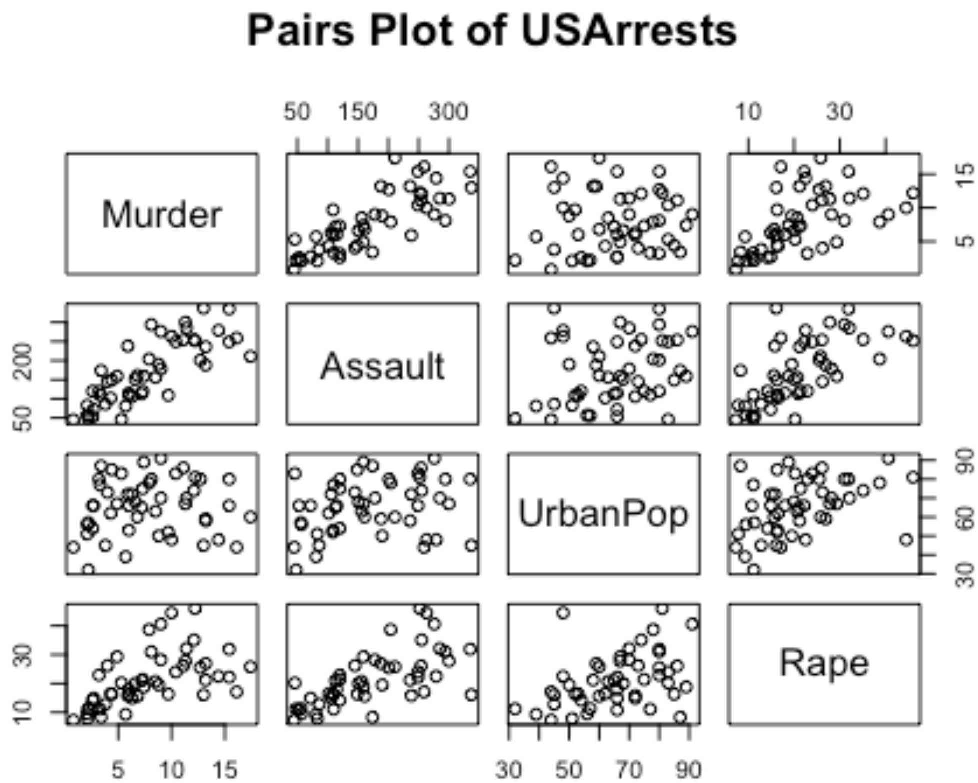



```
print("Most of the cars in this data set are in the class of __15____ mile
per gallon.")

## [1] "Most of the cars in this data set are in the class of __15____ mile
per gallon."
```

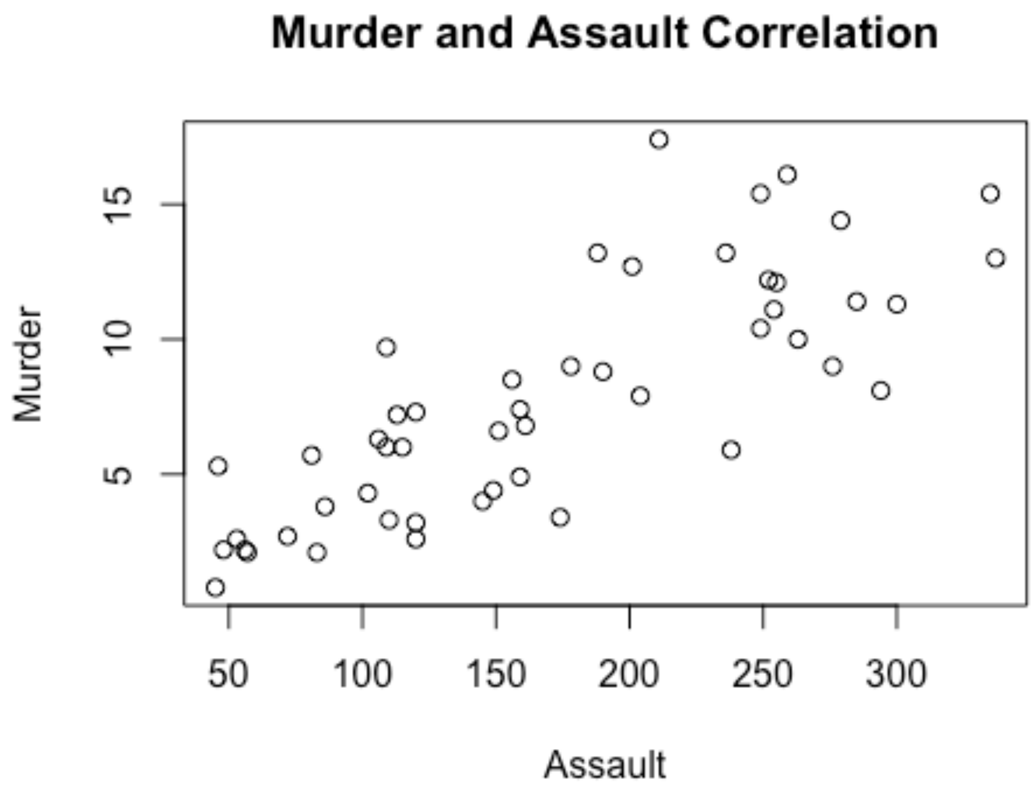
c. Using the **USArrests** data set, create a pairs plot to display the correlations between the variables in the data set. Plot the scatter plot with **Murder** and **Assault**. Give labels to the title, x-axis, and y-axis on the graph. Write a paragraph to summarize your results from both plots.

```
# Load the data set
data(USArrests)

# Head of the data set
head(USArrests)

##            Murder Assault UrbanPop Rape
## Alabama      13.2     236       58 21.2
## Alaska       10.0     263       48 44.5
## Arizona       8.1     294       80 31.0
## Arkansas      8.8     190       50 19.5
## California    9.0     276       91 40.6
## Colorado      7.9     204       78 38.7
```

```r
# Enter your code here!

# pairs plot
pairs(USArrests, main = "Pairs Plot of USArrests")
```




```r
#scatterplot
plot(Murder ~ Assault, data=USArrests, main="Murder and Assault Correlation")
```



Result:

=> Report a paragraph to summarize your findings from the plot! PAIRS PLOT: The scatter plots of Murder and Assault values show a positive linear pattern, meaning that the correlation between Murder and Assault is strong. The scatter plots of Murder and UrbanPop show a random pattern, meaning there is little to no correlation between them. The correlation between Murder and Rape values is quite strong due to the scatter plots showing a positive linear pattern. The scatter plots showing the relationship between Assault and Urban Pop show a weak correlation as the patterns are really scattered and random. The relationship between Assault and Rape values show a positive linear pattern, meaning the correlation is strong. The relationship between Urban Pop and Rape values show somewhat a positive linear pattern, meaning the correlation is quite strong as well.

Question 3

Download the housing data set from www.jaredlander.com and find out what explains the housing prices in New York City.

Note: Check your working directory to make sure that you can download the data into the data folder.

a. Create your own descriptive statistics and aggregation tables to summarize the data set and find any meaningful results between different variables in the data set.

```r
# Head of the cleaned data set
head(housingData)

##   Neighborhood Market.Value.per.SqFt      Boro Year.Built
## 1    FINANCIAL                200.00 Manhattan       1920
## 2    FINANCIAL                242.76 Manhattan       1985
## 4    FINANCIAL                271.23 Manhattan       1930
## 5      TRIBECA                247.48 Manhattan       1985
## 6      TRIBECA                191.37 Manhattan       1986
## 7      TRIBECA                211.53 Manhattan       1985

# Enter your code here!

agg_withlabels<-
aggregate(list(AvgMarketValuePerSqft=housingData$Market.Value.per.SqFt),list(
YearBuilt=housingData$Year.Built),mean)
print(agg_withlabels)

##     YearBuilt AvgMarketValuePerSqft
## 1        1825              76.36000
## 2        1836             273.77000
## 3        1853             152.79000
## 4        1860             159.64500
## 5        1874             111.17000
## 6        1875             166.05000
## 7        1879             194.52000
## 8        1881             109.70500
## 9        1883             172.10000
## 10       1890             113.28750
## 11       1891              72.83000
## 12       1892              95.21000
## 13       1893             168.85000
## 14       1894             110.62000
## 15       1895             151.77500
## 16       1896             117.26500
## 17       1897              40.83000
## 18       1898              83.25000
## 19       1899             108.16000
## 20       1900             137.55908
## 21       1901             172.36778
## 22       1902             167.62167
## 23       1903             147.97000
## 24       1904             123.09333
## 25       1905             187.76583
## 26       1906             169.03364
## 27       1907             173.80000
## 28       1908             150.35000
## 29       1909             135.23667
```

```
## 30         1910                147.36257
## 31         1911                179.76067
## 32         1912                159.51636
## 33         1913                175.93500
## 34         1914                160.29286
## 35         1915                147.08673
## 36         1916                128.20714
## 37         1917                 73.87000
## 38         1918                181.84000
## 39         1919                 63.11000
## 40         1920                145.30862
## 41         1921                122.39125
## 42         1922                118.33250
## 43         1923                115.47625
## 44         1924                165.94091
## 45         1925                147.51316
## 46         1926                148.36423
## 47         1927                131.63357
## 48         1928                153.68375
## 49         1929                106.32121
## 50         1930                142.28936
## 51         1931                129.51731
## 52         1932                 91.74333
## 53         1933                 40.97000
## 54         1934                203.80000
## 55         1935                176.23000
## 56         1936                 46.04333
## 57         1937                 51.77250
## 58         1938                 99.23857
## 59         1939                 93.65083
## 60         1940                154.89857
## 61         1941                111.83733
## 62         1942                128.38600
## 63         1947                113.13500
## 64         1948                186.25000
## 65         1949                 44.98000
## 66         1950                141.96182
## 67         1951                132.98833
## 68         1952                 97.95143
## 69         1954                 81.56500
## 70         1955                130.17538
## 71         1956                178.42786
## 72         1957                127.24091
## 73         1958                159.77000
## 74         1959                108.62692
## 75         1960                104.91200
## 76         1961                106.63000
## 77         1962                129.26294
## 78         1963                152.82937
## 79         1964                103.15000
```

```
## 80        1965              121.01313
## 81        1966               79.94375
## 82        1967               91.94000
## 83        1968              126.76000
## 84        1969              157.28000
## 85        1970              214.59000
## 86        1971               57.60000
## 87        1972              185.72000
## 88        1973              196.75500
## 89        1974              124.42500
## 90        1975              201.26667
## 91        1977              161.32250
## 92        1978              254.69000
## 93        1979              155.71333
## 94        1980              161.74500
## 95        1981              175.96800
## 96        1982              151.30364
## 97        1983              114.79917
## 98        1984              179.48333
## 99        1985              182.66868
## 100       1986              157.62328
## 101       1987              142.14055
## 102       1988              126.43686
## 103       1989              109.25390
## 104       1990               99.31500
## 105       1991              145.76105
## 106       1992               83.92333
## 107       1993               55.45000
## 108       1994               73.13500
## 109       1995               75.77375
## 110       1996              152.36750
## 111       1997              137.41364
## 112       1998              138.25125
## 113       1999              145.93217
## 114       2000              165.47296
## 115       2001              124.74295
## 116       2002              117.92442
## 117       2003              121.56193
## 118       2004              113.79702
## 119       2005              122.70817
## 120       2006              119.73598
## 121       2007              134.12665
## 122       2008              144.34935
## 123       2009               96.52619
## 124       2010               90.36667
```

```
avg_perNeighborhood<-
aggregate(list(AvgMarketValuePerSqft=housingData$Market.Value.per.SqFt),list(
Neighborhood=housingData$Neighborhood),mean)
print(avg_perNeighborhood)
```

```
##                  Neighborhood AvgMarketValuePerSqft
## 1               ALPHABET CITY             148.35500
## 2         ARROCHAR-SHORE ACRES             57.75000
## 3                      ASTORIA             91.48167
## 4                   BATH BEACH             70.34000
## 5                    BAY RIDGE             68.03500
## 6                     BAYSIDE             71.42111
## 7         BEDFORD PARK/NORWOOD             38.24500
## 8           BEDFORD STUYVESANT             83.24172
## 9                      BELMONT             56.45000
## 10                 BENSONHURST             71.70429
## 11                BERGEN BEACH             73.27000
## 12                 BOERUM HILL             96.57600
## 13                BOROUGH PARK             64.10857
## 14                   BRIARWOOD             75.36250
## 15              BRIGHTON BEACH             81.91429
## 16               BRONX-UNKNOWN             32.06500
## 17                   BRONXDALE             28.94333
## 18            BROOKLYN HEIGHTS            114.11778
## 19               BUSH TERMINAL             60.95000
## 20                    BUSHWICK             76.13500
## 21                    CANARSIE             46.58000
## 22             CARROLL GARDENS             93.40556
## 23                     CHELSEA            215.94932
## 24                   CHINATOWN            154.17952
## 25                 CITY ISLAND             40.83000
## 26                CIVIC CENTER            174.06696
## 27                     CLINTON            176.70032
## 28                CLINTON HILL             88.97385
## 29                 COBBLE HILL            120.69800
## 30            COBBLE HILL-WEST             85.71125
## 31               COLLEGE POINT             65.05000
## 32                CONEY ISLAND             55.05750
## 33                      CORONA             94.20706
## 34               CROWN HEIGHTS             64.26286
## 35       DOWNTOWN-FULTON FERRY            103.26857
## 36        DOWNTOWN-FULTON MALL            132.42500
## 37          DOWNTOWN-METROTECH            122.48000
## 38               DYKER HEIGHTS             68.36000
## 39               EAST NEW YORK             36.99167
## 40                EAST TREMONT             72.33333
## 41                EAST VILLAGE            207.46115
## 42                    ELMHURST             69.80564
## 43                FAR ROCKAWAY             74.88500
## 44                     FASHION            194.81067
## 45                   FINANCIAL            199.30917
## 46            FLATBUSH-CENTRAL             65.71167
## 47   FLATBUSH-LEFFERTS GARDEN             46.27000
## 48              FLATBUSH-NORTH             54.00000
## 49                    FLATIRON            223.30311
```

```
## 50       FLUSHING MEADOW PARK              58.59000
## 51             FLUSHING-NORTH              80.16992
## 52             FLUSHING-SOUTH              89.62750
## 53               FOREST HILLS              70.20706
## 54                FORT GREENE              81.76900
## 55                  GLENDALE              57.39667
## 56                   GOWANUS              82.45333
## 57                  GRAMERCY             188.68471
## 58                GRANT CITY              47.60000
## 59                 GRAVESEND              75.63526
## 60               GREAT KILLS              33.74000
## 61                GREENPOINT              86.18053
## 62  GREENWICH VILLAGE-CENTRAL             142.57767
## 63     GREENWICH VILLAGE-WEST             202.13667
## 64               GRYMES HILL              50.09000
## 65                   HAMMELS             139.07200
## 66            HARLEM-CENTRAL             102.79106
## 67               HARLEM-EAST             139.93972
## 68              HARLEM-UPPER              79.25667
## 69               HARLEM-WEST              95.20500
## 70  HIGHBRIDGE/MORRIS HEIGHTS              61.82000
## 71                 HILLCREST              53.95000
## 72                    HOLLIS             109.56000
## 73              HOWARD BEACH              55.06000
## 74                    INWOOD              62.05500
## 75           JACKSON HEIGHTS              47.79238
## 76                   JAMAICA             104.76600
## 77           JAMAICA ESTATES              79.69500
## 78             JAVITS CENTER             125.09000
## 79                KENSINGTON              56.87500
## 80               KEW GARDENS              69.64300
## 81   KINGSBRIDGE HTS/UNIV HTS              23.86000
## 82    KINGSBRIDGE/JEROME PARK              58.37800
## 83                 KIPS BAY             191.31769
## 84              LITTLE ITALY             142.52308
## 85               LITTLE NECK              65.85000
## 86          LONG ISLAND CITY             108.16667
## 87           LOWER EAST SIDE             173.56262
## 88                   MADISON              71.26000
## 89          MANHATTAN VALLEY             111.30043
## 90                   MASPETH              53.32750
## 91            MIDDLE VILLAGE              78.35857
## 92               MIDTOWN CBD             234.36154
## 93              MIDTOWN EAST             211.04750
## 94              MIDTOWN WEST             222.06489
## 95                   MIDWOOD              79.50273
## 96       MORNINGSIDE HEIGHTS              74.63000
## 97      MORRIS PARK/VAN NEST              26.90000
## 98       MORRISANIA/LONGWOOD              44.21250
## 99     MOTT HAVEN/PORT MORRIS              30.96000
```

```
## 100               MURRAY HILL              206.26795
## 101              NEW BRIGHTON               41.47667
## 102   NEW BRIGHTON-ST. GEORGE               41.06000
## 103           NEW SPRINGVILLE               40.47000
## 104           OAKLAND GARDENS               66.94000
## 105                OCEAN HILL               37.92900
## 106       OCEAN PARKWAY-NORTH               76.51111
## 107       OCEAN PARKWAY-SOUTH               75.08000
## 108                OZONE PARK               54.10000
## 109                PARK SLOPE               88.01774
## 110          PARK SLOPE SOUTH               95.84200
## 111               PARKCHESTER               32.67500
## 112      PELHAM PARKWAY SOUTH               30.55000
## 113          PROSPECT HEIGHTS               79.16200
## 114                 REGO PARK               62.13630
## 115                 RIDGEWOOD               64.28667
## 116                 RIVERDALE               57.10176
## 117             ROCKAWAY PARK               88.13600
## 118  SCHUYLERVILLE/PELHAM BAY               49.68000
## 119            SHEEPSHEAD BAY               79.79704
## 120               SILVER LAKE               35.80500
## 121                      SOHO              162.72473
## 122                 SOUNDVIEW               43.40333
## 123          SOUTH OZONE PARK               40.78000
## 124               SOUTHBRIDGE              159.53333
## 125                 SUNNYSIDE               61.61818
## 126               SUNSET PARK               80.58348
## 127               THROGS NECK               53.70667
## 128             TOMPKINSVILLE               35.81000
## 129                   TRIBECA              180.18473
## 130   UPPER EAST SIDE (59-79)              216.83715
## 131   UPPER EAST SIDE (79-96)              202.45179
## 132  UPPER EAST SIDE (96-110)              167.41600
## 133   UPPER WEST SIDE (59-79)              200.24391
## 134   UPPER WEST SIDE (79-96)              171.84515
## 135  UPPER WEST SIDE (96-116)              134.09353
## 136  WASHINGTON HEIGHTS LOWER               65.29600
## 137  WASHINGTON HEIGHTS UPPER               93.50833
## 138          WEST NEW BRIGHTON               39.69000
## 139                WHITESTONE               72.90000
## 140            WILLIAMSBRIDGE               42.46000
## 141      WILLIAMSBURG-CENTRAL               79.97017
## 142         WILLIAMSBURG-EAST               84.32605
## 143        WILLIAMSBURG-NORTH               84.10577
## 144        WILLIAMSBURG-SOUTH               82.27618
## 145           WINDSOR TERRACE               70.21200
## 146                 WOODHAVEN               38.61000
## 147                  WOODSIDE               80.52625
## 148           WYCKOFF HEIGHTS               84.93000
```

```
total_sum_eachNeighborhood<-
aggregate(list(TotalSumMarketValuePerSqFt=housingData$Market.Value.per.SqFt),
list(Neighborhood=housingData$Neighborhood), sum)
print(total_sum_eachNeighborhood)
```

```
##                 Neighborhood TotalSumMarketValuePerSqFt
## 1              ALPHABET CITY                    3560.52
## 2       ARROCHAR-SHORE ACRES                      57.75
## 3                    ASTORIA                    2744.45
## 4                 BATH BEACH                     422.04
## 5                  BAY RIDGE                     544.28
## 6                    BAYSIDE                     642.79
## 7       BEDFORD PARK/NORWOOD                      76.49
## 8         BEDFORD STUYVESANT                    4828.02
## 9                    BELMONT                      56.45
## 10               BENSONHURST                     501.93
## 11              BERGEN BEACH                      73.27
## 12               BOERUM HILL                    1448.64
## 13              BOROUGH PARK                    2243.80
## 14                 BRIARWOOD                     301.45
## 15            BRIGHTON BEACH                    1720.20
## 16             BRONX-UNKNOWN                      64.13
## 17                 BRONXDALE                      86.83
## 18          BROOKLYN HEIGHTS                    1027.06
## 19             BUSH TERMINAL                      60.95
## 20                  BUSHWICK                     152.27
## 21                  CANARSIE                     186.32
## 22           CARROLL GARDENS                     840.65
## 23                   CHELSEA                   19003.54
## 24                 CHINATOWN                    3237.77
## 25               CITY ISLAND                      40.83
## 26              CIVIC CENTER                    4003.54
## 27                   CLINTON                    5477.71
## 28              CLINTON HILL                    1156.66
## 29               COBBLE HILL                     603.49
## 30          COBBLE HILL-WEST                     685.69
## 31             COLLEGE POINT                     195.15
## 32              CONEY ISLAND                     220.23
## 33                    CORONA                    1601.52
## 34             CROWN HEIGHTS                     899.68
## 35     DOWNTOWN-FULTON FERRY                    1445.76
## 36      DOWNTOWN-FULTON MALL                     264.85
## 37        DOWNTOWN-METROTECH                     367.44
## 38             DYKER HEIGHTS                      68.36
## 39             EAST NEW YORK                     221.95
## 40              EAST TREMONT                     217.00
## 41              EAST VILLAGE                    5393.99
## 42                  ELMHURST                    3839.31
## 43              FAR ROCKAWAY                     299.54
## 44                   FASHION                    2922.16
```

```
## 45                   FINANCIAL                       7175.13
## 46             FLATBUSH-CENTRAL                      394.27
## 47   FLATBUSH-LEFFERTS GARDEN                         46.27
## 48               FLATBUSH-NORTH                      432.00
## 49                    FLATIRON                      10048.64
## 50         FLUSHING MEADOW PARK                       58.59
## 51               FLUSHING-NORTH                     10662.60
## 52               FLUSHING-SOUTH                      717.02
## 53                 FOREST HILLS                     1193.52
## 54                  FORT GREENE                      817.69
## 55                    GLENDALE                       172.19
## 56                     GOWANUS                       247.36
## 57                    GRAMERCY                      3207.64
## 58                  GRANT CITY                        95.20
## 59                   GRAVESEND                      1437.07
## 60                 GREAT KILLS                        33.74
## 61                  GREENPOINT                      1637.43
## 62  GREENWICH VILLAGE-CENTRAL                       8554.66
## 63     GREENWICH VILLAGE-WEST                      11521.79
## 64                 GRYMES HILL                        50.09
## 65                     HAMMELS                       695.36
## 66              HARLEM-CENTRAL                      9662.36
## 67                 HARLEM-EAST                      5037.83
## 68                HARLEM-UPPER                       951.08
## 69                 HARLEM-WEST                       190.41
## 70  HIGHBRIDGE/MORRIS HEIGHTS                         61.82
## 71                   HILLCREST                        53.95
## 72                      HOLLIS                       109.56
## 73                HOWARD BEACH                       275.30
## 74                      INWOOD                       124.11
## 75             JACKSON HEIGHTS                      1003.64
## 76                     JAMAICA                       523.83
## 77             JAMAICA ESTATES                       159.39
## 78               JAVITS CENTER                       125.09
## 79                  KENSINGTON                       113.75
## 80                 KEW GARDENS                       696.43
## 81   KINGSBRIDGE HTS/UNIV HTS                         23.86
## 82    KINGSBRIDGE/JEROME PARK                        291.89
## 83                    KIPS BAY                      2487.13
## 84                LITTLE ITALY                      1852.80
## 85                 LITTLE NECK                        65.85
## 86            LONG ISLAND CITY                      2596.00
## 87             LOWER EAST SIDE                      7289.63
## 88                     MADISON                      1353.94
## 89            MANHATTAN VALLEY                      2559.91
## 90                     MASPETH                       213.31
## 91              MIDDLE VILLAGE                       548.51
## 92                 MIDTOWN CBD                      3046.70
## 93                MIDTOWN EAST                     11818.66
## 94                MIDTOWN WEST                     10437.05
```

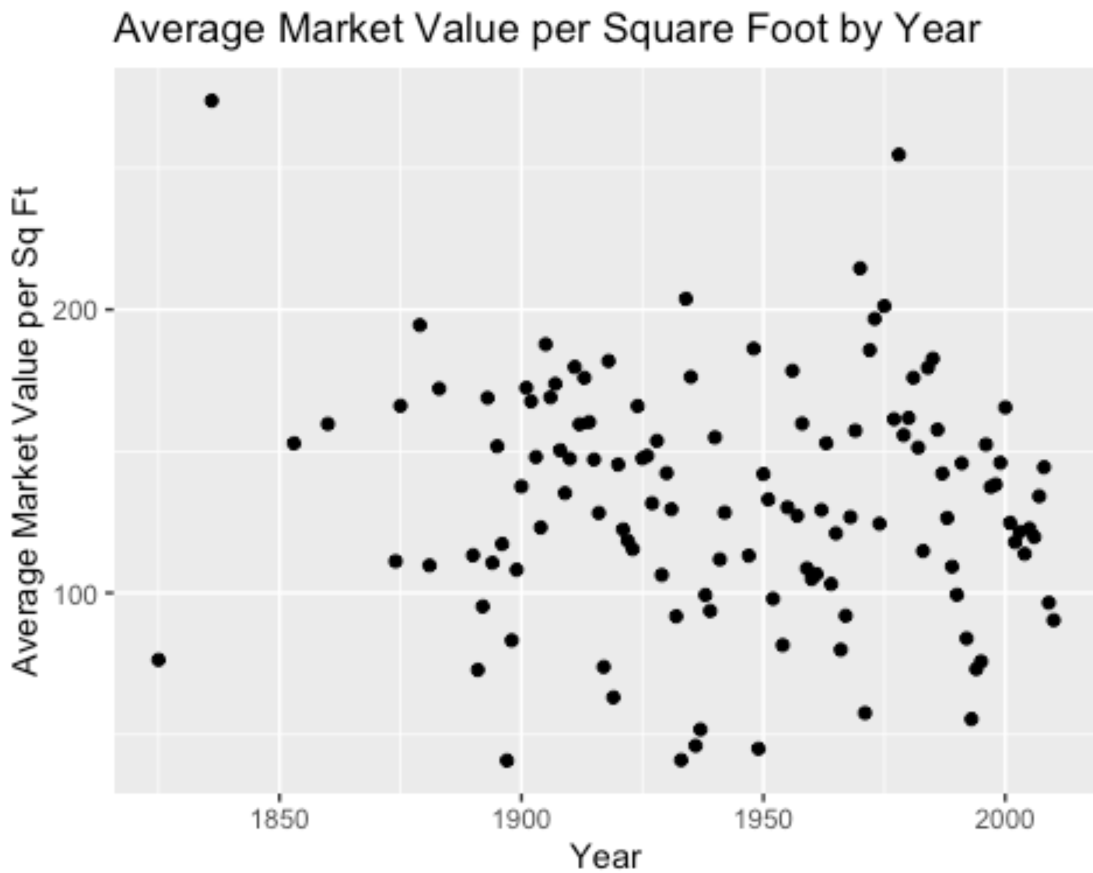
```
## 95                    MIDWOOD                         874.53
## 96        MORNINGSIDE HEIGHTS                          74.63
## 97       MORRIS PARK/VAN NEST                          26.90
## 98        MORRISANIA/LONGWOOD                         707.40
## 99     MOTT HAVEN/PORT MORRIS                          30.96
## 100               MURRAY HILL                        8044.45
## 101              NEW BRIGHTON                         124.43
## 102   NEW BRIGHTON-ST. GEORGE                          82.12
## 103           NEW SPRINGVILLE                         364.23
## 104           OAKLAND GARDENS                          66.94
## 105                OCEAN HILL                         379.29
## 106       OCEAN PARKWAY-NORTH                         688.60
## 107       OCEAN PARKWAY-SOUTH                          75.08
## 108               OZONE PARK                           54.10
## 109                PARK SLOPE                        2728.55
## 110          PARK SLOPE SOUTH                         958.42
## 111               PARKCHESTER                          65.35
## 112      PELHAM PARKWAY SOUTH                          30.55
## 113          PROSPECT HEIGHTS                         395.81
## 114                 REGO PARK                        1677.68
## 115                 RIDGEWOOD                         192.86
## 116                 RIVERDALE                         970.73
## 117             ROCKAWAY PARK                         440.68
## 118  SCHUYLERVILLE/PELHAM BAY                          49.68
## 119             SHEEPSHEAD BAY                       2154.52
## 120                SILVER LAKE                          71.61
## 121                      SOHO                        8949.86
## 122                 SOUNDVIEW                         260.42
## 123          SOUTH OZONE PARK                          40.78
## 124                SOUTHBRIDGE                       1435.80
## 125                 SUNNYSIDE                         677.80
## 126               SUNSET PARK                        1853.42
## 127               THROGS NECK                         161.12
## 128             TOMPKINSVILLE                          35.81
## 129                   TRIBECA                       13333.67
## 130   UPPER EAST SIDE (59-79)                       26670.97
## 131   UPPER EAST SIDE (79-96)                       15791.24
## 132  UPPER EAST SIDE (96-110)                         837.08
## 133   UPPER WEST SIDE (59-79)                       17421.22
## 134   UPPER WEST SIDE (79-96)                       11341.78
## 135  UPPER WEST SIDE (96-116)                        4559.18
## 136  WASHINGTON HEIGHTS LOWER                         326.48
## 137  WASHINGTON HEIGHTS UPPER                         561.05
## 138         WEST NEW BRIGHTON                          39.69
## 139                WHITESTONE                         145.80
## 140            WILLIAMSBRIDGE                          84.92
## 141      WILLIAMSBURG-CENTRAL                        4798.21
## 142         WILLIAMSBURG-EAST                        3204.39
## 143        WILLIAMSBURG-NORTH                        2186.75
## 144        WILLIAMSBURG-SOUTH                        2797.39
```

```
## 145          WINDSOR TERRACE                          351.06
## 146                WOODHAVEN                          154.44
## 147                 WOODSIDE                          644.21
## 148          WYCKOFF HEIGHTS                          254.79
```

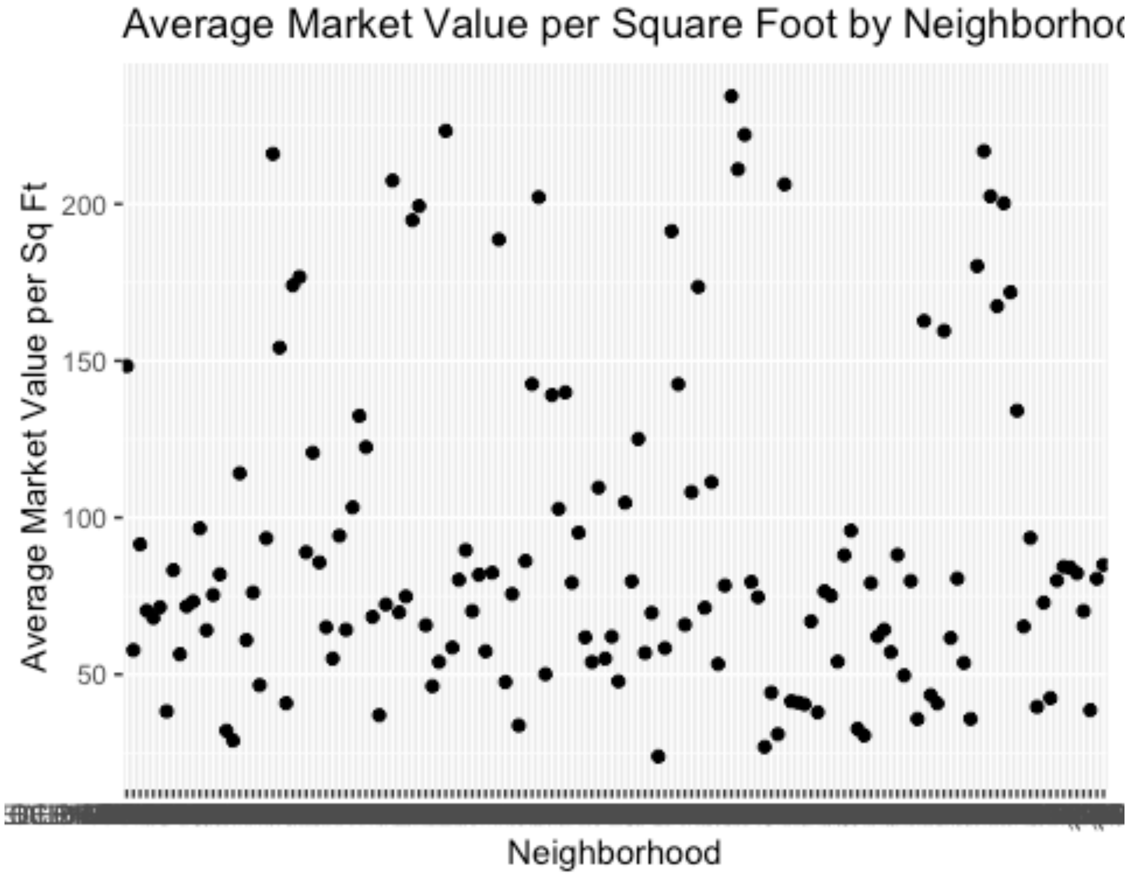
#MeaningfulResults? The years with a really high average market value per square foot probably means that those years were when the rich neighborhoods were built. As for the average market value for each Neighborhood, the cost for living of neighborhoods like financial, Chelsea, and Downtown are really high and expensive.

b. Create multiple plots to demonstrates the correlations between different variables. Remember to label all axes and give title to each graph.

```
# Enter your code here!

agg_withlabels<-
aggregate(list(AvgMarketValuePerSqft=housingData$Market.Value.per.SqFt),list(
YearBuilt=housingData$Year.Built),mean)

avg_perNeighborhood<-
aggregate(list(AvgMarketValuePerSqft=housingData$Market.Value.per.SqFt),list(
Neighborhood=housingData$Neighborhood),mean)


ggplot(agg_withlabels, aes(x = YearBuilt, y = AvgMarketValuePerSqft)) +
  geom_point() +
  labs(
    title = "Average Market Value per Square Foot by Year",
    x = "Year",
    y = "Average Market Value per Sq Ft"
  )
```




```
ggplot(avg_perNeighborhood, aes(x = Neighborhood, y = AvgMarketValuePerSqft))
+
  geom_point() +
  labs(
    title = "Average Market Value per Square Foot by Neighborhood",
    x = "Neighborhood",
    y = "Average Market Value per Sq Ft"
  )
```



c. Write a summary about your findings from this exercise.

=> Enter your answer here! Overall, this exercise helped me learn how to make tables using summarize and aggregate functions as well making graphs/plots derived from the data of the tables. I had to come up with descriptive statistics (such as mean) of values for each group. I also had to interpret what the plot graphs show and describe the correlation between certain variables, which is an essential part to data analyzing. In addition, I learned a simple way to get rid of all missing values (NA values) on R which is using the na.omit() function. That function will be highly useful to simply just get rid of all missing values that are unnecessary to analyze.