Stat 3301: Homework 1

Nathan Johnson.9254

Due by date and time specified on Carmen

Setup:

```
library(alr4)
library(tidyverse)
```

Instructions

- Replace "FirstName LastName (name.n)" above with your information.
- Provide your solutions below in the spaces marked "Solution:".
- Include any R code that you use to answer the questions; if a numeric answer is required, show how you calculated it in R. Use the option `echo = TRUE` to make sure the R code is displayed.
- Knit this document to HTML and upload both the HTML file and your completed Rmd file to Carmen
- Make sure your solutions are clean and easy-to-read by
  - formatting all plots to be appropriately sized, with appropriate axis labels.
  - only including R code that is necessary to answer the questions below.
  - only including R output that is necessary to answer the questions below (avoiding lengthy output).
  - providing short written answers explaining your work, and writing in complete sentences.

Question 1 Import the data set `ames_real_estate.csv` (which is available on Carmen) and use it to do the following:

a) Find the smallest and largest observed sale prices separately for each neighborhood.
b) Make a scatterplot with the square footage of the house on the x-axis and the sale price of the house on the y-axis.
c) Write a sentence describing the relationship between `SalePrice` and `SqFt`.

```
library(readr)
ames = read_csv('ames_real_estate.csv')
```

Solution to Question 1

```
## Rows: 1598 Columns: 17
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (5): Neighborhood, Style, GarageType, CentralAir, FireplaceQu
## dbl (12): LotArea, YearBuilt, YearRemod, SqFt, FullBath, HalfBath, Bedrooms,...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

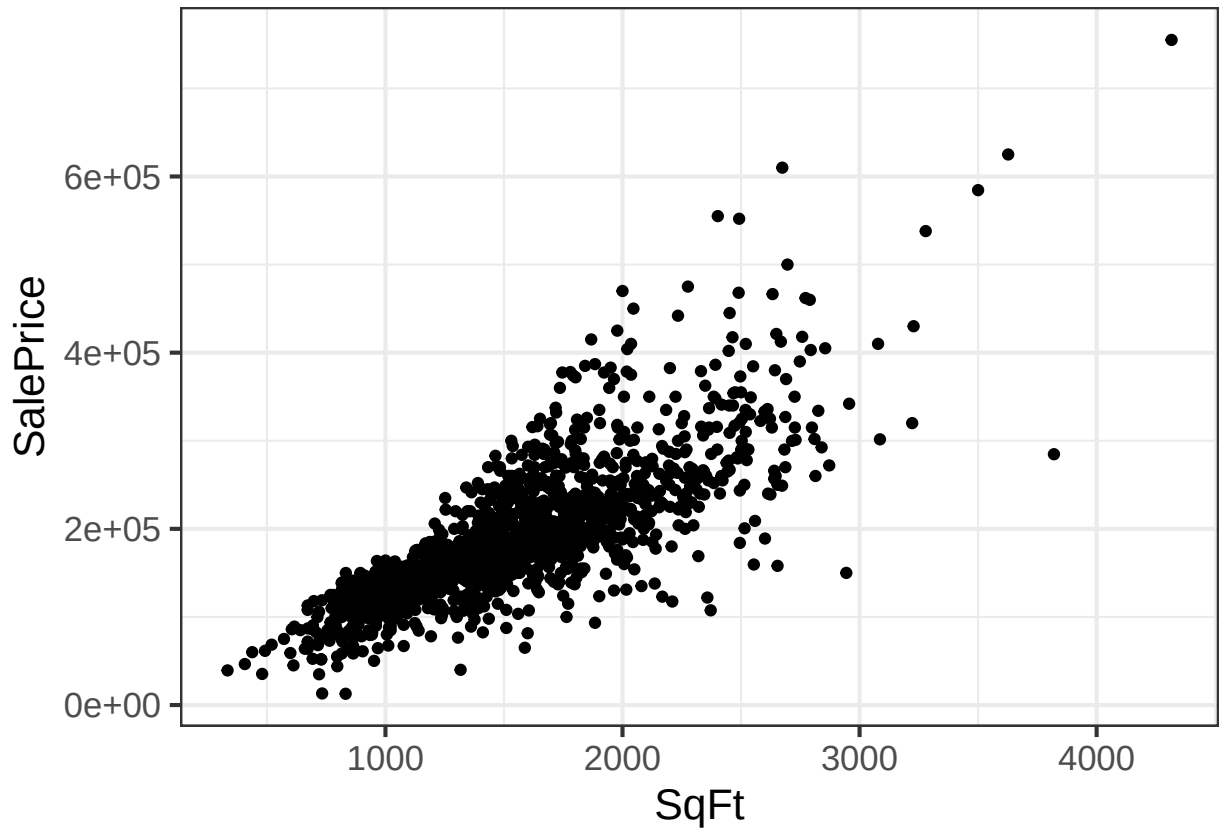
a. Here are the smallest and largest observed sale prices separately for each neighborhood:

```
ames %>% group_by(Neighborhood)%>% summarize(minSalePrice = min(SalePrice),
                          maxSalePrice = max(SalePrice)) %>% print(n = Inf)
```

```
## # A tibble: 21 x 3
##    Neighborhood minSalePrice maxSalePrice
##    <chr>               <dbl>        <dbl>
##  1 Blmngtn            159895       159895
##  2 BrkSide             39300       223500
##  3 ClearCr            143000       328000
##  4 CollgCr            110000       475000
##  5 Crawfor             90350       335000
##  6 Edwards             58500       320000
##  7 Gilbert            115000       377500
##  8 IDOTRR              13100       212300
##  9 Mitchel             81500       300000
## 10 NAmes               68000       301600
## 11 NWAmes              82500       306000
## 12 NoRidge            190000       755000
## 13 NridgHt            214000       610000
## 14 OldTown             12789       265979
## 15 SWISU               60000       189000
## 16 Sawyer              62383       219000
## 17 SawyerW             67500       320000
## 18 Somerst            176000       468000
## 19 StoneBr            240000       538000
## 20 Timber             150000       425000
## 21 Veenker            150000       385000
```

b. Here is the scatterplot of ames with square feet on the x-axis and sale price on the y-axis.

```
ames %>% ggplot(aes(x = SqFt, y = SalePrice)) + geom_point() + theme_bw(16)
```



c. As square feet (`SqFt`) in a house increases on average, the sales price (`SalePrice`) of the house also increases on average.

Question 2 Use the data set `Heights` from the library`alr4` to calculate the average height for daughters whose mothers are **at least** 60 inches tall. (The purpose of this question is to make sure you have the package installed correctly.)

```
Heights %>% filter(mheight >= 60) %>% summarize(mean(dheight))
```

Solution to Question 2

```
##   mean(dheight)
## 1      64.09704
```

The mean height of daughters whose mothers were at least 60 inches tall is 64.09704 inches.