Homework 6

Andrew Estes

9/26/2021

Install dataset and libraries

```
library(tinytex)
library(AmesHousing)
```

```
## Warning: package 'AmesHousing' was built under R version 4.1.1
```

```
library(tidyverse)
```

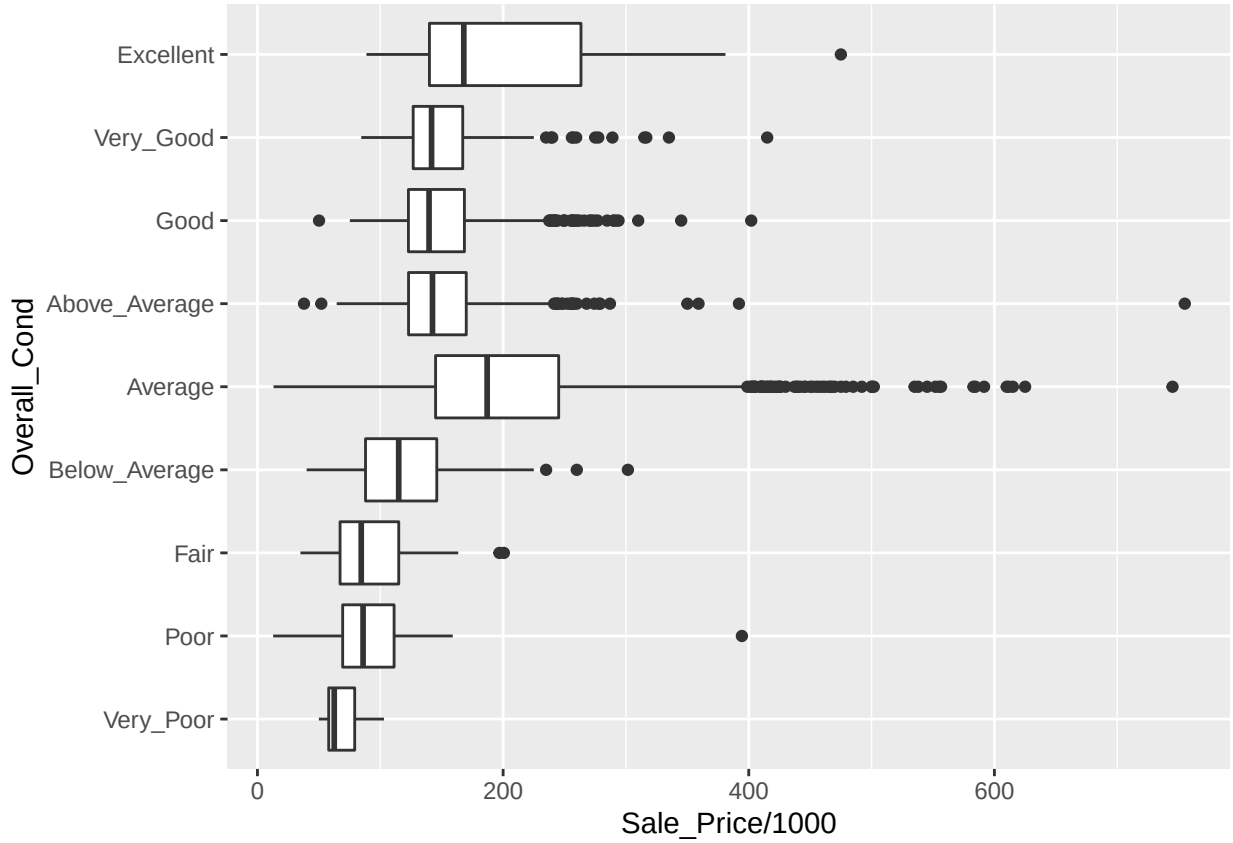
```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.3     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.0     v forcats 0.5.1

## Warning: package 'tidyr' was built under R version 4.1.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
Ames <- make_ames()
```

1

Make a boxplot between Sale Price and Overall Condition

```
p <- ggplot(Ames, aes(Sale_Price/1000, Overall_Cond))
p + geom_boxplot()
```

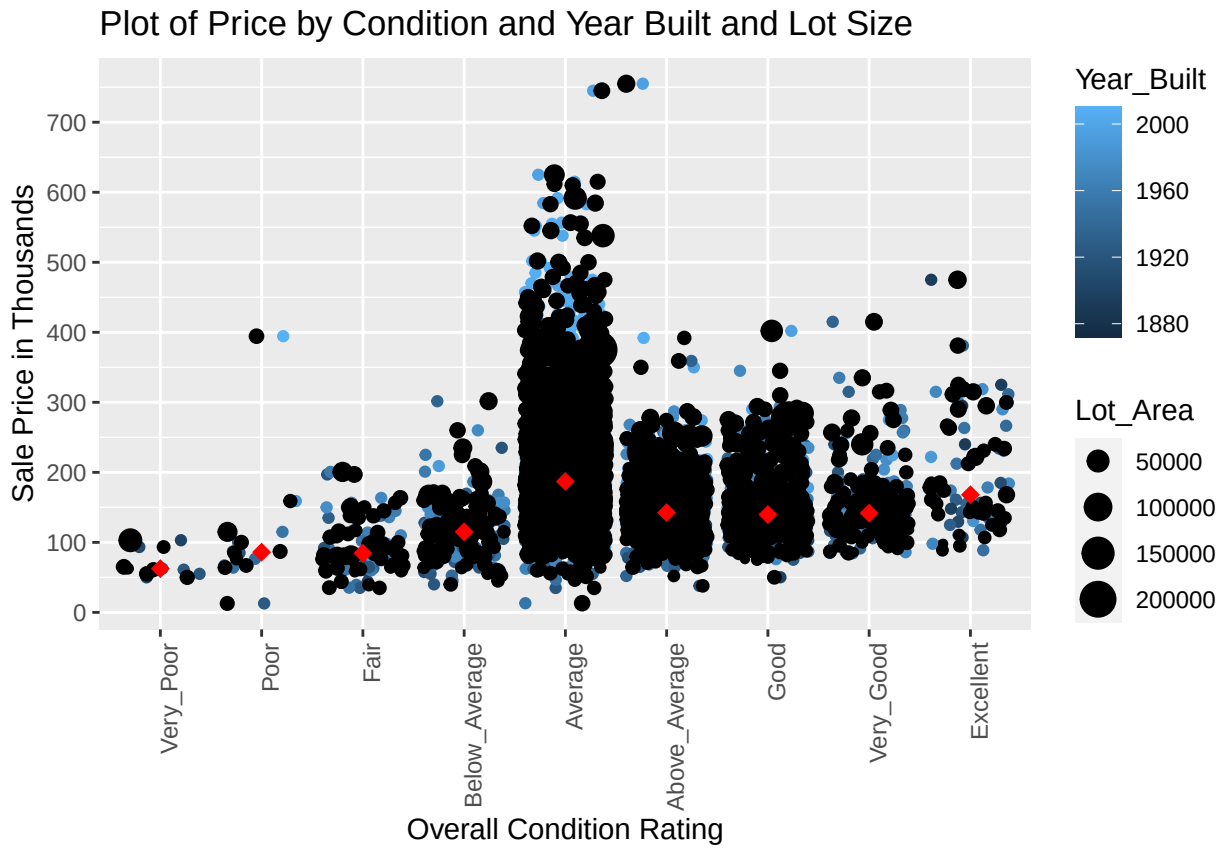


Typically overall condition correlates to price The Average boxplot is different due to several outliers on the higher end This indicates that there is another factor influencing price within the average class

2

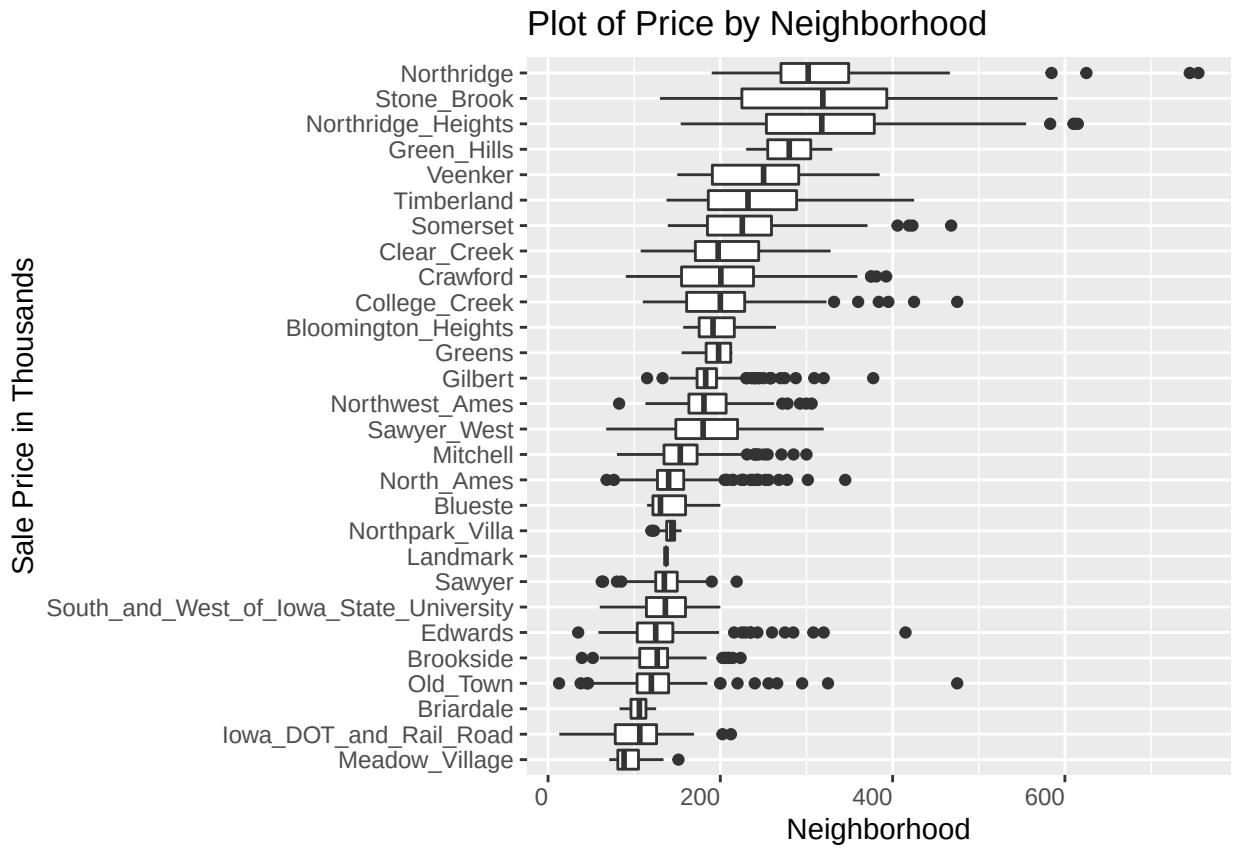
a) Make a jitter plot
b) Does Year Built or Lot Area impact Sales Price more?
c) Make your own plot analyzing a variable's impact on Sale Price
d) Write about your findings

```
#Part A
d <- p + geom_jitter(aes(colour = Year_Built)) +
          labs(title="Plot of Price by Condition and Year Built and Lot Size",x="Sale Price in Thousands
          theme(axis.text.x = element_text(angle=90, hjust=1)) +
          scale_x_continuous(breaks=seq(0, 900, by=100))
z <- d + coord_flip()
t <- z + geom_jitter(aes(size = Lot_Area))
t <- t + stat_summary(fun = median, geom="point", shape=18, size=3, color="red")
t
```



Year Built impacts the Sales Price for the Average Condition more than the Lot Area does

```
#Part C
a <- ggplot(data = Ames) +
  geom_boxplot(mapping = aes(x=reorder(Neighborhood, Sale_Price/1000, na.rm = TRUE), y = Sale_Price/1000
  theme(axis.text.x = element_text(angle=0, hjust = 1)) +
  labs(title="Plot of Price by Neighborhood",x="Sale Price in Thousands", y = "Neighborhood")
a <- a + coord_flip()
a
```



Location, location, location! Where you live is often the greatest indicator of sales price. The average rent for a 740 sqft apartment in San Francisco is over 3000 dollars/month according to rentcafe.com. Similarly, a 760 sqft apartment in Des Moines costs less than 1000 dollars per month. Similary, in Ames, the average house price in in North Ridge is over 250k meanwhile in down in Meadow Village that figure drops to less than 100k. A 2.5x difference.