Exam 2 Question 1

# Question 1a.   
# Reading in the eBay data  
  
library(tidyverse)

## -- Attaching packages ----------------------------------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2 v purrr 0.3.4  
## v tibble 3.0.3 v dplyr 1.0.2  
## v tidyr 1.1.2 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.5.0

## -- Conflicts -------------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

exfile <- "http://www.itk.ilstu.edu/faculty/jrwolf/ebay.csv"  
df <- read\_csv(exfile)

## Parsed with column specification:  
## cols(  
## Bids = col\_double(),  
## Color = col\_character(),  
## Price = col\_double(),  
## BuyerFB = col\_double(),  
## lnBuyerFB = col\_double(),  
## SellerFB = col\_double(),  
## lnSellerFB = col\_double(),  
## ShipCost = col\_double(),  
## GenderInd = col\_double(),  
## New = col\_double(),  
## Used = col\_double(),  
## Cond = col\_character(),  
## Gender = col\_character()  
## )

head(df)

## # A tibble: 6 x 13  
## Bids Color Price BuyerFB lnBuyerFB SellerFB lnSellerFB ShipCost GenderInd  
## <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 White 24.0 228 5.43 15 2.71 6 0  
## 2 1 White 11 911 6.82 205 5.32 4.8 0  
## 3 1 White 20 164 5.11 85 4.44 5.5 0  
## 4 1 White 25 7 2.08 175 5.16 10 0  
## 5 2 White 7.49 270 5.60 4725 8.46 7.95 0  
## 6 4 White 15.5 195 5.28 26 3.26 5 0  
## # ... with 4 more variables: New <dbl>, Used <dbl>, Cond <chr>, Gender <chr>

#Question 1b.  
#Examine the data  
  
#look at the head of the data   
head(df)

## # A tibble: 6 x 13  
## Bids Color Price BuyerFB lnBuyerFB SellerFB lnSellerFB ShipCost GenderInd  
## <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 White 24.0 228 5.43 15 2.71 6 0  
## 2 1 White 11 911 6.82 205 5.32 4.8 0  
## 3 1 White 20 164 5.11 85 4.44 5.5 0  
## 4 1 White 25 7 2.08 175 5.16 10 0  
## 5 2 White 7.49 270 5.60 4725 8.46 7.95 0  
## 6 4 White 15.5 195 5.28 26 3.26 5 0  
## # ... with 4 more variables: New <dbl>, Used <dbl>, Cond <chr>, Gender <chr>

# Look at the tail of the data  
tail(df)

## # A tibble: 6 x 13  
## Bids Color Price BuyerFB lnBuyerFB SellerFB lnSellerFB ShipCost GenderInd  
## <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 26 Blue 36.0 774 6.65 23 3.14 8.98 0  
## 2 26 Blue 35.5 9 2.30 22396 10.0 6.3 1  
## 3 26 Blue 48 53 3.99 922 6.83 0 1  
## 4 30 Blue 51 7 2.08 52 3.95 0 1  
## 5 34 Blue 40 8 2.20 737 6.60 0 1  
## 6 42 Blue 46 31 3.47 2 0.693 59.6 0  
## # ... with 4 more variables: New <dbl>, Used <dbl>, Cond <chr>, Gender <chr>

# Question 1c  
  
# Examine the structure (str) of the data  
str(df)

## tibble [1,083 x 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ Bids : num [1:1083] 1 1 1 1 2 4 4 5 7 8 ...  
## $ Color : chr [1:1083] "White" "White" "White" "White" ...  
## $ Price : num [1:1083] 23.95 11 20 25 7.49 ...  
## $ BuyerFB : num [1:1083] 228 911 164 7 270 195 131 48 90 32 ...  
## $ lnBuyerFB : num [1:1083] 5.43 6.82 5.11 2.08 5.6 ...  
## $ SellerFB : num [1:1083] 15 205 85 175 4725 ...  
## $ lnSellerFB: num [1:1083] 2.71 5.32 4.44 5.16 8.46 ...  
## $ ShipCost : num [1:1083] 6 4.8 5.5 10 7.95 5 5.95 5 4.5 12 ...  
## $ GenderInd : num [1:1083] 0 0 0 0 0 0 1 1 1 0 ...  
## $ New : num [1:1083] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Used : num [1:1083] 1 1 1 1 1 1 1 1 1 1 ...  
## $ Cond : chr [1:1083] "Used" "Used" "Used" "Used" ...  
## $ Gender : chr [1:1083] "M" "M" "M" "M" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. Bids = col\_double(),  
## .. Color = col\_character(),  
## .. Price = col\_double(),  
## .. BuyerFB = col\_double(),  
## .. lnBuyerFB = col\_double(),  
## .. SellerFB = col\_double(),  
## .. lnSellerFB = col\_double(),  
## .. ShipCost = col\_double(),  
## .. GenderInd = col\_double(),  
## .. New = col\_double(),  
## .. Used = col\_double(),  
## .. Cond = col\_character(),  
## .. Gender = col\_character()  
## .. )

# Question d1.  
  
# Find the Mean (mean) and Standard Deviation (sd) by old/new of Price  
# Filter old condition out of the data  
old\_data <- filter(df, Cond == "Used")  
old\_data$Price <- as.numeric(old\_data$Price)  
  
  
# Filter new condition out of the data  
New\_data <- filter(df, Cond == "New")  
New\_data$Price <- as.numeric(New\_data$Price)

# Find the Mean (mean) and Standard Deviation (sd) by old of price  
  
# Find the Mean (mean) by old of price  
mean(old\_data$Price)

## [1] 31.40495

# Find the Standard Deviation (sd) by old of price  
sd(old\_data$Price)

## [1] 11.18313

# Find the Mean (mean) and Standard Deviation (sd) by new of Price  
  
# Find the Mean (mean) by New = 0 of price  
mean(New\_data$Price)

## [1] 42.79198

# Find the Standard Deviation (sd) by New = 0 of price  
sd(New\_data$Price)

## [1] 8.770916

# Question 1 d2.  
# Find the Mean (mean) and Standard Deviation (sd) by male/female of Price  
  
# Filter the male data out of the data  
male\_data <- filter(df, Gender=="M")  
male\_data$Price <- as.numeric(male\_data$Price)  
  
# Filter the female data out of the data  
female\_data <- filter(df, Gender=="F")  
female\_data$Price <- as.numeric(female\_data$Price)

# Find the Mean (mean) by male of Price  
  
mean(male\_data$Price)

## [1] 37.83654

# Find theStandard Deviation (sd) by male of Price  
sd(male\_data$Price)

## [1] 12.20783

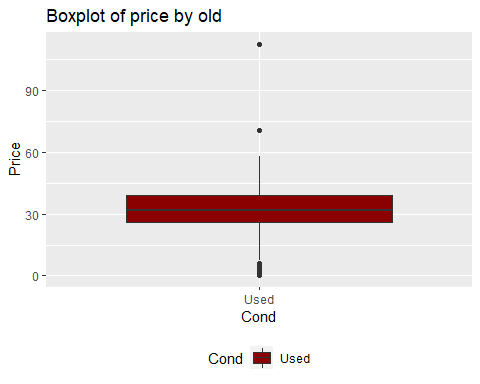
# Find the Mean (mean) by female of Price  
mean(female\_data$Price)

## [1] 38.48699

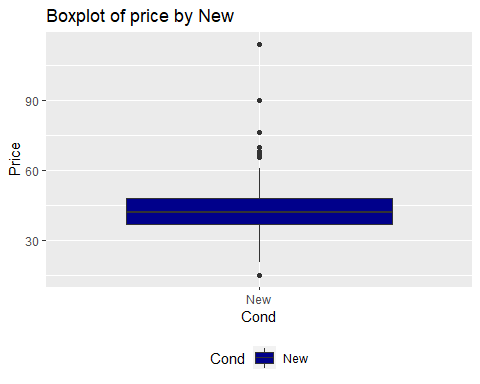
# Find the Standard Deviation (sd) by female of Price  
sd(female\_data$Price)

## [1] 10.33858

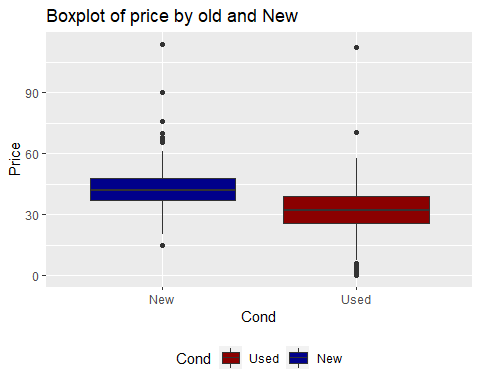
# e1. Create a botplot of Price (by old/new )  
library(ggplot2)  
library(dplyr)  
  
# botplot of Price by old  
ggplot(old\_data, aes(x=Cond, y= Price, fill = Cond)) + geom\_boxplot() + theme(legend.position = "bottom") + scale\_fill\_manual(breaks = "Used", values = 'darkred') + ggtitle("Boxplot of price by old")



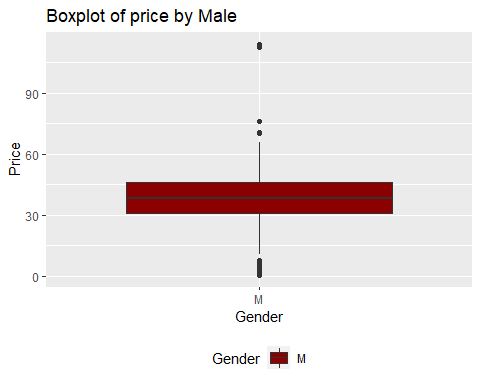
# botplot of Price by New  
ggplot(New\_data, aes(x=Cond, y= Price, fill= Cond)) + geom\_boxplot() + theme(legend.position = "bottom") + scale\_fill\_manual(breaks = "New", values = 'darkblue') + ggtitle("Boxplot of price by New")



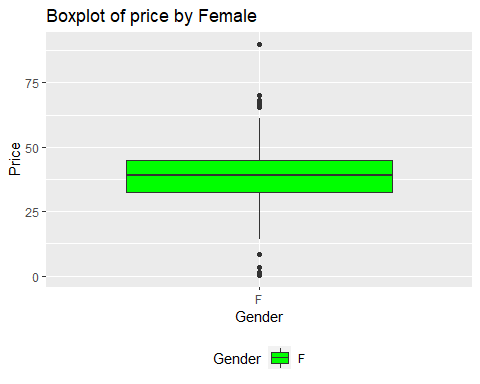
# We can show the relationship between price and old/new as  
ggplot(df, aes(x=Cond, y= Price, fill= Cond)) + geom\_boxplot() + theme(legend.position ="bottom") + scale\_fill\_manual(breaks = c("Used", "New"), values = c('darkred', 'darkblue')) + ggtitle("Boxplot of price by old and New")



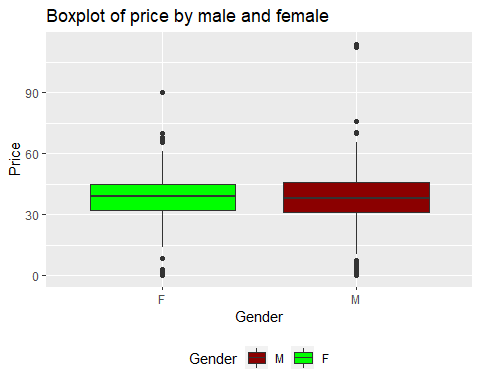
# e2. Create a botplot of Price (by male/female )  
  
# botplot of Price by male  
ggplot(male\_data, aes(x = Gender, y = Price, fill = Gender)) + geom\_boxplot() + theme(legend.position = "bottom") + scale\_fill\_manual(breaks = "M", values = 'darkred') + ggtitle("Boxplot of price by Male")



# botplot of Price by female  
ggplot(female\_data, aes(x = Gender, y = Price, fill = Gender)) + geom\_boxplot() + theme(legend.position = "bottom") + scale\_fill\_manual(breaks = "F", values = 'green') + ggtitle("Boxplot of price by Female")



# We can show the relationship between price and male/female as  
ggplot(df, aes(x=Gender, y= Price, fill= Gender)) + geom\_boxplot() + theme(legend.position ="bottom") + scale\_fill\_manual(breaks = c("M", "F"), values = c('darkred', 'green')) + ggtitle("Boxplot of price by male and female")



# f. Create a histogram of Price (by color)  
library(ggthemes)

## Warning: package 'ggthemes' was built under R version 4.0.3

## Warning: package 'ggthemes' was built under R version 4.0.3  
ggplot(df, aes(x = Price, color=Color, fill= Color)) +  
geom\_histogram(position = "identity", alpha= 0.5, bins= 30, col= "black") + scale\_fill\_manual(values = c("Blue", "Gray", "Green", "Orange", "Pink", "Purple", "Red", "#c0c0c0", "White"))+ scale\_color\_manual(values = c("blue", "gray", "green", "orange", "pink", "purple", "red", "#c0c0c0", "white")) + theme\_economist() + ggtitle("Histogram of Price by color")

