hw1-Sayan-Biswas

Sayan Biswas

15 January 2019

Part A

Problem 1

```
selectCols <- function(data,...){
  list_args <- list(...)
  for(i in seq_along(list_args))
  {
    if(is.numeric(list_args[[i]])){
      col <- list_args[[i]]
      list_args[[i]] <- colnames(data[col])
    }
  }
  #List with column names converted to a vector
  list_vect <- unlist(list_args)
  #Duplicate column names are removed from the vector
  list_vect <- unique(list_vect)
  data[1:10,list_vect]
}

selectCols(mpg,1,2,"year")
```

```
## # A tibble: 10 x 3
##    manufacturer model       year
##    <chr>        <chr>      <int>
##  1 audi         a4          1999
##  2 audi         a4          1999
##  3 audi         a4          2008
##  4 audi         a4          2008
##  5 audi         a4          1999
##  6 audi         a4          1999
##  7 audi         a4          2008
##  8 audi         a4 quattro  1999
##  9 audi         a4 quattro  1999
## 10 audi         a4 quattro  2008
```

Problem 2

```
plotCols <- function(data){
  for (i in seq_along(data))
  {
    name <- names(data[i])
```

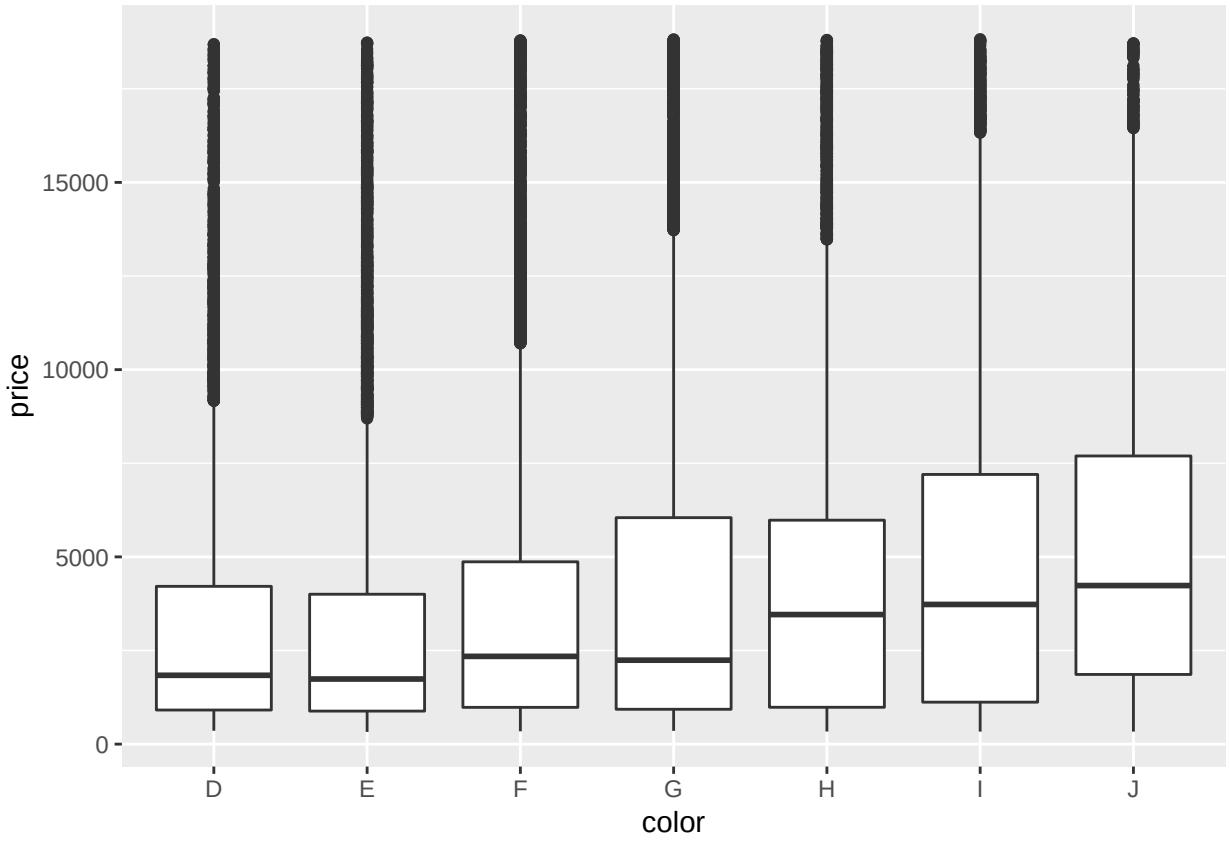
```
  if(is.numeric(data[[i]]))
  {
    g <- ggplot(data=data,mapping=aes_string(name))+
      geom_histogram()

  } else
  {
    g <- ggplot(data=data,mapping=aes_string(name))+
      geom_bar()
  }
  print(g)
  }
}
```

Part B

Problem 3

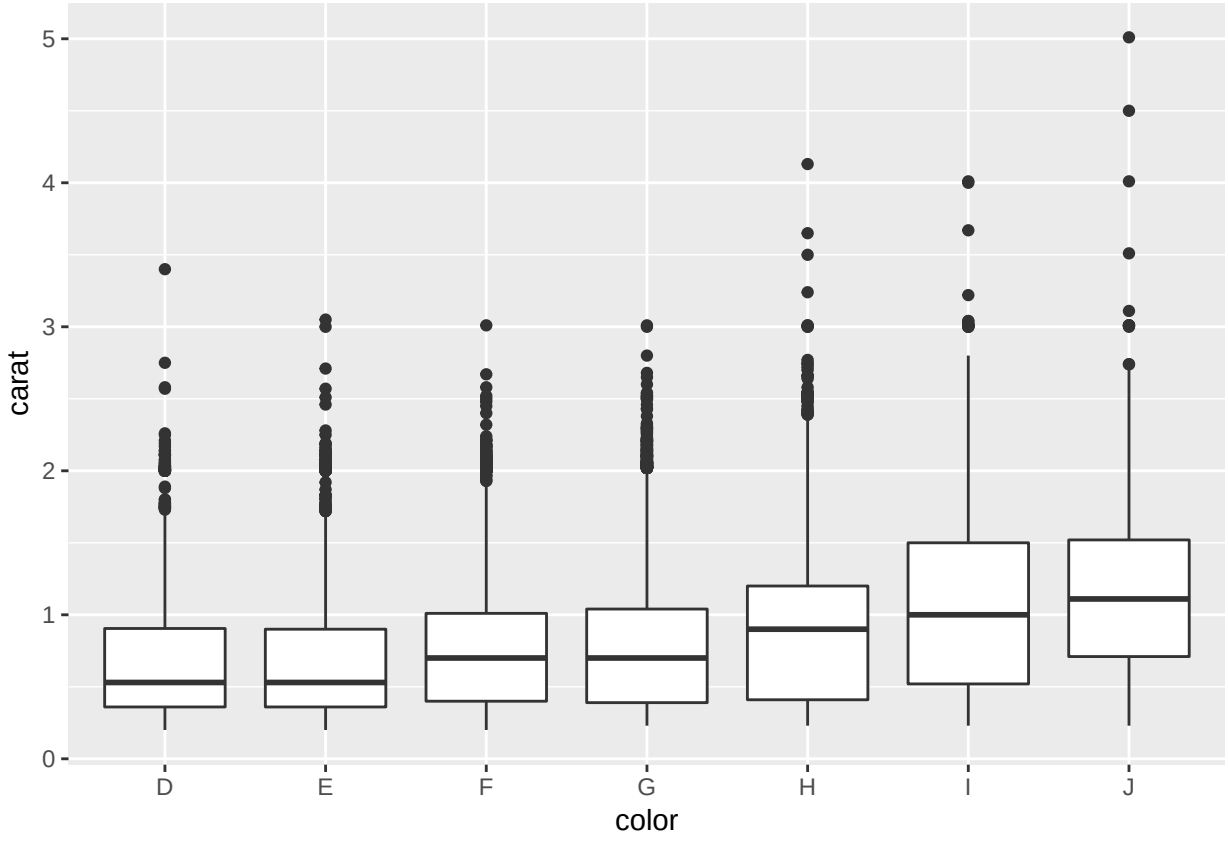
```
ggplot(data=diamonds,mapping=aes(x=color,y=price))+
  geom_boxplot()
```



The above visualization of price for each level of color for the diamonds dataset indicates that the median price of the diamond increases as the color of the diamond becomes worse. This plot is not quite clear so as to get the information as why the diamonds with the worst color has the highest median price. Also all of the boxplots are right skewed and have a large number of outliers. There is a large amount of variation of price for each color of the diamonds.

Problem 4

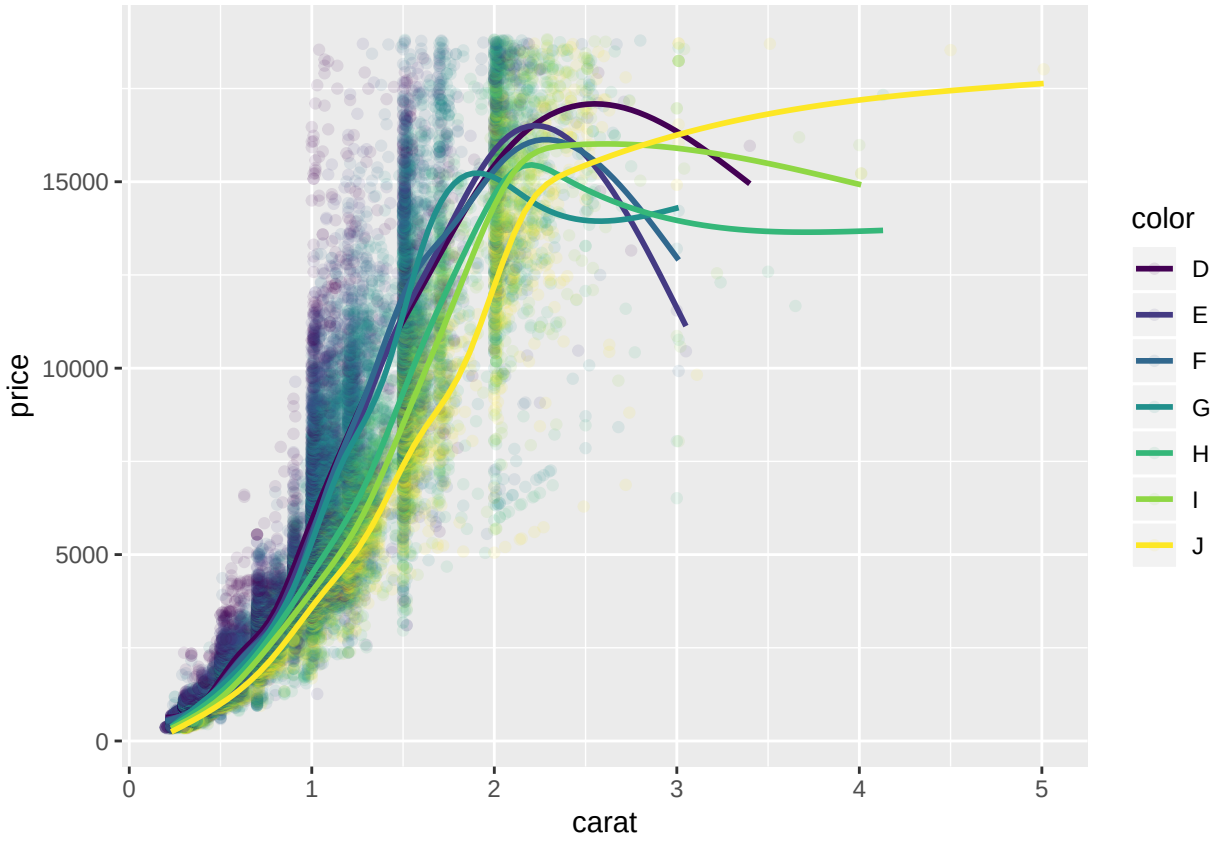
```
ggplot(data=diamonds,mapping=aes(x=color,y=carat))+
  geom_boxplot()
```



The above visualization of carat for each level of color for the diamonds dataset indicates that the median carat (i.e. the weight of the diamond) increases as the color becomes worse. This plot helps us understand the behavior of the previous plot which is, the price of the diamond is not only a factor of its color but also a factor of the carat(weight of the diamond). The price of the heavier and the worst color diamond is higher than the price of the lighter diamonds with relatively good colors.

Problem 5

```
ggplot(data=diamonds,mapping=aes(x=carat,y=price,color=color))+
  geom_point(alpha=1/10)+
  geom_smooth(se=FALSE)
```



For each of the color, there is a positive co-relation between carat and price and the price increases as the carat(or size of the diamond) increases.The price of the good colors are relatively higher than the price of the worse color till the carat size is ~2. When the carat size increases beyond ~2 carats, the prices falls steeply for the color D, E and F, whereas the price falls slowly for the colors G, H and I; for color J the price linearly increases till 5 carats.