Assignment 1

CIS 335

Monica Klosin

1. Dataset downloaded from https://www.kaggle.com/shivam2503/diamonds-
2. Dataset downloaded, coding in R.
3. The response variable will be the price of the diamond. The explanatory variables and their attributes are listed in the table below.

|  |  |
| --- | --- |
| **Variable Name** | **Variable Type** |
| Carat | Quantitative |
| Cut | Categorical |
| Color | Categorical |
| Clarity | Categorical |
| Depth | Quantitative |
| Table | Quantitative |
| x | Quantitative |
| y | Quantitative |
| z | Quantitative |

1. Identify the mean, median, mode of the numeric attributes. Also identify which attributes are positively or negatively skewed.

Carat:

mean: 0.7979

median: 0.7

mode: 0.3

skew: carat and price of a diamond has a positive correlation.

Depth:

mean: 61.74

median: 61.8

mode: 62

skew:

Table:

mean: 57.45

median: 57

mode: 56

skew:

x:

mean: 5.73

median: 5.7

mode: 4.37

skew:

y:

mean: 5.73

median: 5.71

mode: 4.34

skew:

z:

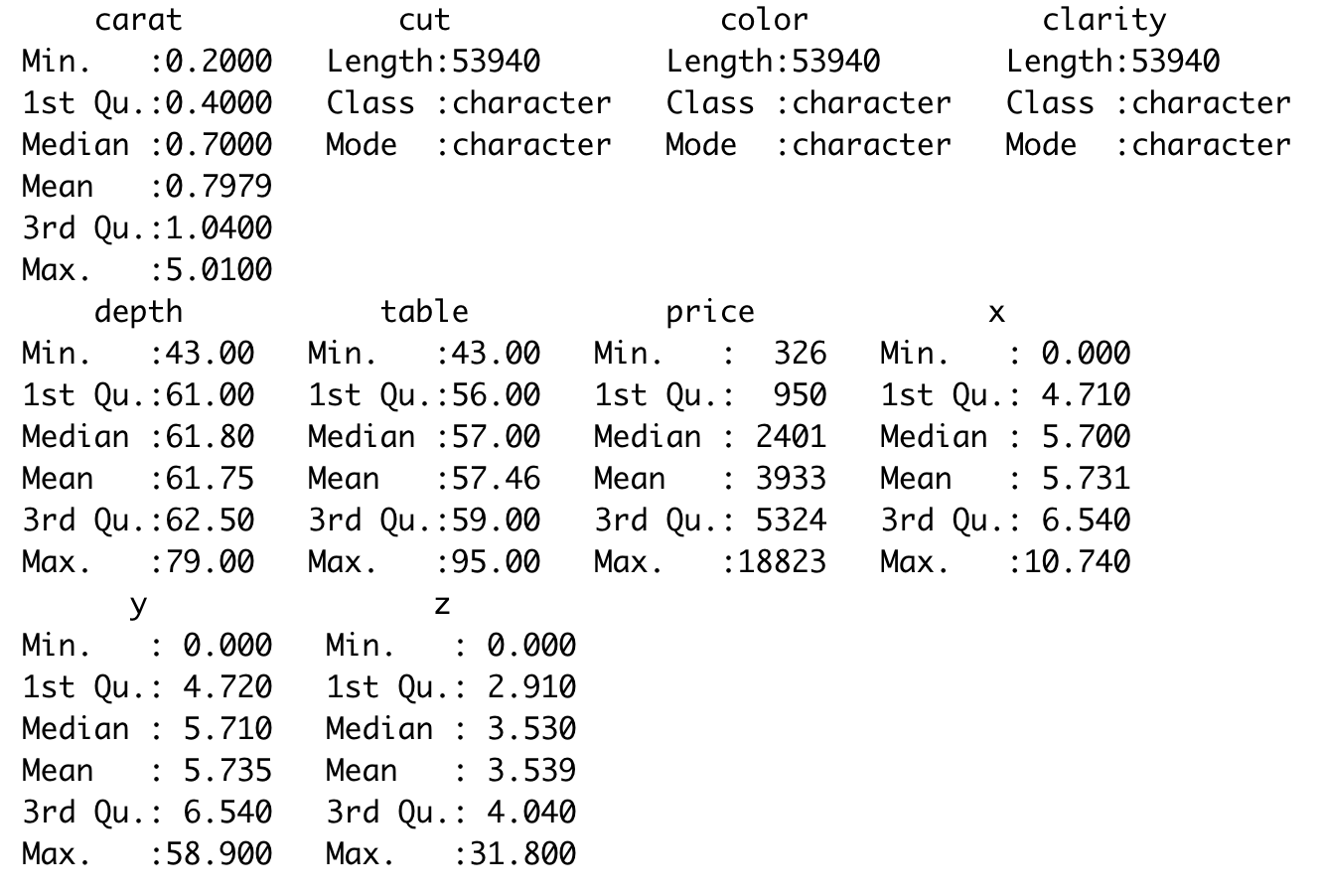
mean: 3.54

median: 3.53

mode: 2.7

skew:

1. IQR



Determine outliers, and if you will keep them and why.

1. Correlation between carat and price?

Correlation between cut and price?

1. Boxplots

