STAT 6021 Project 1

Arne Newman

10/11/2020

Start by loading in the dataset and exploring it

```
setwd("C:/Users/newma/Data_v1/stat6021/project1")
data <- read.csv("diamonds4.csv", header=TRUE)
attach(data)
names(data)
```

```
## [1] "carat"   "clarity" "color"   "cut"     "price"
```

```
str(data)
```

```
## 'data.frame':    1214 obs. of  5 variables:
##  $ carat  : num  0.24 0.4 3.35 0.61 0.4 0.5 1.76 1.02 0.7 1.01 ...
##  $ clarity: chr  "VVS2" "VS2" "IF" "VS2" ...
##  $ color  : chr  "G" "H" "F" "G" ...
##  $ cut    : chr  "Very Good" "Very Good" "Good" "Ideal" ...
##  $ price  : int  379 605 56151 1947 684 1553 24451 3958 2091 6079 ...
```

Potential analyses to consider:

- Develop an appropriate model to predict the price of diamonds.
  - Test assumptions along the way.
  - Simplify the model to the greatest extent possible for ease of use and interpretation.
  - Start with SLRs
  - Generate scatterplot matrix/heat map to assess correlations between variables
  - P rogress to MLR (with indicators)
- Determine the variable with the greatest impact on price.
- Assess multicollinearity.
- Assess interaction between variables.

##Variable Descriptions ###Carat

###Clarity In descending order, the Blue Nile site shows groupings of clarity:

- I1, I2, I3 Included Diamonds
- SI1, SI2 Slightly Included (SI) Diamonds
- VS1, VS2 Very Slightly Included (VS) Diamonds
- VVS1, VVS2 Very, Very Slightly Included (VVS) Diamonds
- Internally Flawless (IF) Diamonds
- Flawless (FL) Diamonds

###Color The GIA diamond color grades range from D (colorless) to Z (light yellow or brown). Less color, thus lower letters in the alphabet, are considered better.

###Cut Ideal: This rare cut represents roughly the top 3% of diamond cut quality. It reflects most of the light that enters the diamond.

Very Good: This cut represents roughly the top 15% of diamond cut quality. It reflects nearly as much light as the ideal cut, but for a lower price.

Good: This cut represents roughly the top 25% of diamond cut quality. It reflects most of the light that enters, but not as much as a Very Good cut grade.

Astor by Blue Nile™: These diamonds are crafted to gather and reflect the most light possible. Cut from the finest raw material (rough stones with as few impurities or inclusions as possible), they meet rigorous quality requirements and exhibit outstanding brilliance, fire, and scintillation. In addition to being graded by the GIA, all Astor by Blue Nile™ diamonds are certified by GemEx®