Homework 6

Due April 29, 2020 (by 11:59pm)

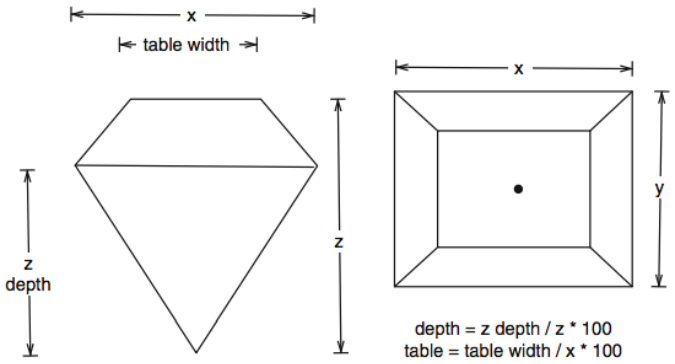
Question 1:

Consider the `diamonds` dataset built in `ggplot2` package, which consists of price and quality information for about 54,000 diamonds:

```
library(ggplot2)
head(diamonds)

## # A tibble: 6 x 10
##   carat cut       color clarity depth table price     x     y     z
##   <dbl> <ord>     <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl>
## 1 0.23  Ideal     E     SI2      61.5    55   326  3.95  3.98  2.43
## 2 0.21  Premium   E     SI1      59.8    61   326  3.89  3.84  2.31
## 3 0.23  Good      E     VS1      56.9    65   327  4.05  4.07  2.31
## 4 0.290 Premium   I     VS2      62.4    58   334  4.2   4.23  2.63
## 5 0.31  Good      J     SI2      63.3    58   335  4.34  4.35  2.75
## 6 0.24  Very Good J     VVS2     62.8    57   336  3.94  3.96  2.48
```

In this data set, there are four variables of diamond quality: carat, cut, colour and clarity; and five physical measurements: depth, table, x, y and z, as described below:



1. Create a barplot of variable `cut`.
2. Create a barplot that incoporates information from two variables `cut` and `color`.
3. Create an appropriate plot to demonstrate the distribution of `price`. What interesting patterns do you see?
4. How does the distribution of `price` vary with `clarity`?
5. Create a boxplot to demonstrate the distribution of `depth` conditioning on `cut`, and then comment on the relationship.
6. Use an appropriate plot to investigate the relationship between `carat` and `price`.
7. Use an appropriate plot to investigate the relationship between `carat` and `price`, conditioning on `cut`.