Homework 2

Problem 1

Let's plot! This problem uses the `diamonds` data frame in `ggplot2`.

Since this data frame is large, and let's just sample some observations from this data frame. First of all, what many observations this data frame contain?

Issue '?diamonds' in R to find out more information about this data frame. Then we do a 1000 random sample from this data frame.

You need to use qplot function for 1, 2, 3 1. How to get a hisogram and density of the variable 'carat'? 2. I think there is a relation between the carat of a diamond and the price of a diamond. How do we see this in a plot? 3. How to get different histograms of carat for different clarity levels? i.e, I want to draw different histograms of carat with and use diffrent colors to represent different clarity levels.

You need to use ggplot function to 4, 5

4. We would like to explore the relation between the price and carat in the logarithmic scale. How to do so using ggplot() function?

5. How to see the relation between the price and carat in the logarithmic scale for different clarity levels of diamonds? (check facet_grid)

Problem 2

This problem may seem involving a lot of statistics or mathematics. But do not panic. We will work on this problem together and I will provide the tools you will need for this problem.

Suppose U_1 and U_2 are uniformly generated. Someone claims that the random value

$$Z = \sqrt{-2\log(U_1)} * \cos(2\pi U_2)$$

has a standard normal distribution.

- There is a simple way to check this:

1. First generate 1,000 independent uniform random variables for U_1 and U_2($runif$ function can be useful).Type(?runif in R to check the details of this function)

- And then we do following comparisons:

1. Calculate the proportion that $Z >= 0$, name it as p_1, compare it with q1= 1-pnorm(0) in R. What's pnorm in R?

2. Calculate the proportion that $0 < Z <= 1$, name it as p_2, compare it with q2= pnorm(1)-pnorm(0) (In class we talk about how to find the number of variables greater than 0 for normal random variables, so you may want to check or look at the notes.)

3. Calculate the proportion that $1 < Z <= 2$, name it as p_3, compare it with q3=pnorm(2)-pnorm(1)

4. Calculate the proportion that $z > 2$, name it as p_4, compare it with q4=q3=1- pnorm(2)

What did you find? (This question is optional). Note that this is only an approximation and therefore if you want to be more accurate, you should generate more than 1,000 random variables.

Problem 3

We learned a lot of basic things in R, and I would like you to write down some interesting or useful functions you discovered or learned. They do not have to come from this course. I am not looking for any specific functions so you can just write down what you think are useful.