**Assignment 1A**

**Stat 302 Spring 2020**

**Due**

Assignment 1A is to be completed individually. The purpose of this assignment is to give you practice working with quantitative and categorical variables. You will need to use R for some simple tasks.

**Bikeshare**

Capital Bikeshare is a bike rental company in Washington D.C. The Company has provided data for number of bikes rented every day over 2011-2012. Variables related to time and weather have also been included in the data set. The Company wishes to identify important factors to use for predicting the number of bikes rented on a given day. The variables included in the data are defined below. An ID number for day and the date are also included in the data set.

|  |  |
| --- | --- |
| Season | 1=winter,  2=spring,  3=summer,  4=fall |
| yr | 0=2011,  1=2012 |
| weather | 1=clear/partly cloudy,  2=mist/cloudy/partly cloudy,  3=light rain/snow,  4=heavy rain/snow/storm |
| temp | normalized air temperature, ranging 0-1 where 0 is coldest and 1 is hottest |
| cnt | number of bike rentals |

**Do the following for each variable.**

There will be an upload spot in Crowdmark for each variable.

1. Identify the variable type (quantitative (continuous or discrete), categorical (nominal, binary, or ordinal). Explain your reasoning.
2. Using R, create an appropriate graphic showing the distribution of the data for the respective variable. Remember labels and titles.
3. Discuss the distribution for each variable based on the graphs in #2.
   * **For categorical**, compare counts or proportions between categories. Do they look as you’d expect, given the definitions of the variables?
   * **For quantitative**, discuss the approximate center, range, shape, and outliers (if any).
4. Compute appropriate summary statistics for each variable.

**(see the next page for tips on managing your work in Jupyter.)**

**Managing your work in Jupyter Notebooks:**

Give your “written” responses and interpretations in Markdown cells next to the respective output. Use multiple Markdown cells to separate text for different parts of a question. (There are other tips on Canvas for the Stat Workshop at a link in R>R in Jupyter.)

You will need to upload results to several spots in Crowdmark. You can use a separate Jupyter Notebook for each upload spot. Alternatively, you can use the code “\pagebreak” in a markdown cell where you want a page break. This will make it possible for you to upload a single PDF to the first upload spot in Crowdmark and then drag pages to the appropriate spots within Crowdmark. Make sure the pages are in the correct order before submitting. (I personally prefer to work with shorter programs, when possible. Less scrolling, and less potential for weird things to happen because of running code out of order.)