IT 497 Exam #2

- For each of the 3 questions below, create **an *.Rmd file** containing **all text and code needed to answer the questions** associated.
  - For example, Question 1 has a-f.
  - So, you would create an *.Rmd file that includes **all the code and any additional text needed to answer questions a-f**.

- **Please note that several questions (or question parts) require text -- not code answers.**
  - For example, **Question 2 Part F2** requires you to "*Write a two sentence explanation of the graph.*"
  - Include this text (and similar text answers in your *.Rmd file).

- Knit all three *.Rmd files to either a Word document or HTML file.
- Submit all 6 files via Reggienet.

Any submission without all 6 files will receive a ZERO for the Exam.

Question 1a. Read in the following eBay data:

```
library(tidyverse)
exfile <- "http://www.itk.ilstu.edu/faculty/jrwolf/ebay.csv"
df <- read_csv(exfile)
head(df)

# b. Examine the data (look at head and tail)
# c. Examine the structure (str) of the data
# d1. Find the Mean (mean) and Standard Deviation (sd) by old/new of Price
# d2. Find the Mean (mean) and Standard Deviation (sd) by male/female of Price
# e1. Create a botplot of Price (by old/new )
# e2. Create a botplot of Price (by male/female )
# f. Create a histogram of Price (by color)
```

Question 2. a. Read in the following Milk Cow data:

```
milkcow <-
"https://github.com/rfordatascience/tidytuesday/raw/master/data/2019/2
019-01-29/milkcow_facts.csv"
df <- read_csv(milkcow)
```

b. Examine the data (look at head and tail)
c. Examine the structure (str) of the data

d. Use select to obtain the columns: year, avg_milk_cow_number, milk_per_cow, avg_price_milk, milk_cow_cost_per_animal, alfalfa_hay_price and slaughter_cow_price.

e. Omit all other columns
f1. Using a line graph, show milk_per_cow and year.
f2. Write a two sentence explanation of the graph.
g1. Using a line graph, show avg_milk_cow_number and year.
g2. Write a two sentence explanation of the graph above .
h1. Find the scatter plot of avg_milk_cow_number and avg_price_milk.
h2. Does the plot reveal any relationship between the variables? Write a two sentence explanation of the graph.

Question 3. a. Read in the following Horror Movie data:

```
horror_movies <-
read_csv("https://raw.githubusercontent.com/rfordatascience/tidy
tuesday/master/data/2019/2019-10-22/horror_movies.csv")
```

b. Examine the data (look at head and tail)

c. Examine the structure (str) of the data

d. Use select to obtain the columns title, release_country, review_rating, and budget.

e. Omit other columns

f. Obtain the movies released in the USA with a movie rating of either PG-13 or R (Omit movies with other ratings)
g. Omit any rows with missing data in the columns title, release_country, review_rating, or budget
h. Omit any rows with budget in any currency other than US Dollars (this is probably the most
difficult part of the question).

i. Omit any rows with budgets less than $500,000

j. Convert movie_run_time to numeric. You can do this any way you like, but may want to use parse_number from the readr package.

k1. Using a bargraph (or histogram), show the average review rating for both PG-13 and R movies
K2. Write a two sentence explanation of the graph.

l1. Using a bargraph (or histogram), show the average budgets for both PG-13 and R movies and
l2. Write a two sentence explanation of the graph.

m1. Find the scatter plot of the review ratings and movie run times.
m2. Does the plot reveal any relationship between the variables? Write a two sentence explanation of the graph.

Knit all three *.Rmd files to either a Word document or HTML file.

Submit all 6 files via Reggienet.

Any submission without all 6 files will receive a ZERO for the Exam.