Math 3339
Homework 2 (Chapter 2 & Sections 9.1 – 9.2)

Name:_____ PeopleSoft ID:_____

Instructions:
- Homework will NOT be accepted through email or in person. Homework must be submitted through CourseWare BEFORE the deadline.
- Print out this file use or software and complete the problems.
- Write in black ink or dark pencil or type your solutions in the space provided. You must show all work for full credit.
- Submit this assignment at http://www.casa.uh.edu under "Assignments" and choose **hw2**.
- Total possible points: **15**

1. The *reacttimes* data set from our textbook has 50 observations of human reaction times to a physical stimulus. (https://www.math.uh.edu/~wwang/MATH3339/dataset/reacttimes.txt) The reaction times are named **Times** and arranged in increasing order.

   a. Find the mean and median of **Times** without using R. You may use your calculator.
   b. Import *reacttimes* into your workspace as a data frame with a single variable **Times.** Then calculate the mean of **Times** by using R's mean( ) function and the median with R's median( ) function.
   c. Find the five number summary of **Times** and interpret each number.
   d. Find the 60th percentile of **Times** without using R.
   e. Find the 60th percentiles of **Times** using R.
   f. Take off the 1st observation from *reacttimes* data, find the 60th percentiles of **Times** by using R.

   Hint:
   Part b:
   reacttimes <- read.table("https://www.math.uh.edu/~wwang/MATH3339/dataset/reacttimes.txt",header = T)
   mean(reacttimes$Times)
   median(reacttimes$Times)
   Part e: quantile(reacttimes$Times, 0.6, type=5)
   Part f: newdata <-reacttimes$Times[-1]
         quantile(newdata,0.6, type=5)

2. The average public school teacher salaries in thousands of dollars for all 51 states and Washington D.C. are in the data set "teacher_salaries.txt".

   a. Use R to make a boxplot of the variable *Pay*.
   b. Describe the distribution of the variable *Pay*. Give, the shape, median, range, and any outliers.
   c. Make a bar plot of the variable *Region*.

3. Section 2.6.4, problem 6

4. The following are the top 25 gross movies in 2018. You can use this link to import into R: https://www.math.uh.edu/~cathy/Math3339/data/movie.csv

Rank	Movie Title	Gross per $1 million
1	Black Panther	$ 700.06
2	Avengers: Infinity War	$ 678.82
3	Incredibles 2	$ 608.58
4	Jurassic World: Fallen Kingdom	$ 416.77
5	Deadpool 2	$ 318.49
6	Dr. Seuss' The Grinch (2018)	$ 269.61
7	Aquaman	$ 259.72
8	Mission: Impossible - Fallout	$ 220.16
9	Ant-Man and the Wasp	$ 216.65
10	Solo: A Star Wars Story	$ 213.77
11	Venom (2018)	$ 213.37
12	A Star is Born (2018)	$ 202.11
13	Bohemian Rhapsody	$ 193.67
14	A Quiet Place	$ 188.02
15	Ralph Breaks the Internet	$ 187.16
16	Crazy Rich Asians	$ 174.04
17	Hotel Transylvania 3: Summer Vacation	$ 167.51
18	Halloween (2018)	$ 159.34
19	Fantastic Beasts: The Crimes of Grindelwald	$ 158.06
20	The Meg	$ 145.44
21	Ocean's 8	$ 140.22
22	Mary Poppins Returns	$ 138.73
23	Ready Player One	$ 137.69
24	Spider-Man: Into The Spider-Verse	$ 133.86
25	Mamma Mia! Here We Go Again	$ 120.63

a. Plot a histogram of the Gross for these movies. Is the distribution symmetric, skewed left, skewed right or bimodal?
b. Determine the mean, standard deviation and the median of the Gross. Justify why the mean higher than the median.
c. Determine the five number summary. Determine if there are any outliers in this data?
d. Take away the outliers determine the mean and the median. Comment on what happens to the mean and the median when these outliers are removed. Does it make sense removing the data why or why not?
e. Determine the 90th percentile. (Hint: quantile(Gross, 0.9, type=5)).

5. Below is a stem-plot of the birth weights of male babies born to the smoking group. The stems are in units of kg.

The decimal point is at the |

```
1 | 9
2 | 346778889
3 | 223467899
4 | 12234
5 | 3556
6 | 1
```

a. Find the median birth weight.
b. Find the mean birth weight.
c. Find the sample standard deviation of the birth weight.
d. Which measurement would be best to use for measuring the center? Justify your answer.

6. The following information is from a sample of 15 women. They recorded the women's weight in pounds and height in inches. You can use R to answer all of these questions, no need to do by hand.

Height	58	58	59	61	61	63	64	65	67	67	67	70	72	72	72
Weight	117	117	120	123	123	139	142	142	146	146	150	150	150	154	164

a. Draw a scatterplot of the data. Describe the relationship between height and weight.
b. Determine the covariance and correlation coefficient (r) between height and weight. Give the interpretation of the correlation.
c. Suppose we want to predict the weight of a women based on the height. Give the least-squares equation to predict weight based on height.
d. If a women is 71 inches tall, what is the predicted weight?
e. Determine the coefficient of determination (R^2). Give the interpretation of this value.
f. If a women weights 148 lbs. and is 70 inches tall, what is the residual?