Stat 1201: Problem Set 1

Read Chapter 1, then complete the exercises below. Circle your answers clearly where appropriate. Make sure your homework is legible. Handwritten assignments should be scanned to a single pdf file. You may discuss strategies with classmates, but all work submitted must be your own. Work done in R may be submitted as a separate html file or included in the pdf. You may use technology (or not) as you see fit.

This assignment is due **Tues, Sept 15, at 11:59pm.**

(If you are admitted into the class *after* Friday, Sept 11, you can take until Mon, Sep 21 to submit this assignment.)

Lateness policy for this assignment:

Up to 24 hours late: 4 point deduction
Up to 48 hours late: 8 point deduction
After 48 hours: no credit

You are allowed one "free" up to 24 hours late assignment with no penalty. In other words, we will not start deducting points until the second late assignment, as long as the first late assignment was turned in no later than 24 hours after the due date/time.

Problems:

1. **Inkjet Printers**

The following data come from a *PC Magazine* review of 20 printers.

available in the R package **Lock5withR**, dataset: `InkjetPrinters`

More info: https://rdrr.io/cran/Lock5withR/man/InkjetPrinters.html

The data is also available here: inkjet.csv as well as in the Files section of CourseWorks.

(To download the file, click "Raw" and then File, Save in your browser.)

Price of printers (in dollars):
60, 70, 79, 79, 87, 99, 99, 100, 129, 149, 150, 150, 189, 199, 199, 199, 248, 299, 300, 348

Time to print a 4x6 color photo (in seconds):
41, 42, 42, 43, 44, 44, 45, 58, 63, 66, 66, 67, 77, 94, 96, 97, 134, 143, 170, 170

a) Draw two histograms of the printer prices, one with a binwidth of 50 and the other with a binwidth of 25, using *right closed* bins, i.e. (50, 100]. Compare the shape of each, using vocabulary for histogram shapes in Section 1.2 as well as your own narrative description.

b) Draw a stem-and-leaf plot of time to print a 4x6 color photo. What is the shape of the distribution?

2. **Create your own data**

a) Create two datasets – 10 values each – one of which is skewed right and the other skewed left.

b) Calculate the mean and median of each dataset.

c) What pattern did you find?

3. **Standard deviation**

Use the following data for this question:

35, 39, 52, 67, 89

a) Calculate $\bar{x}$ and the deviations from the mean for each value.

b) Use the deviations calculated in part (a) to determine the sample variance (s^2) and the sample standard deviation (s).

c) Subtract 10 from each observation, then calcuate the sample variance of the transformed values. How does it compare to s^2 for the original data?

4. **College Fight Songs**

Data Source: *Our Guide to the Exuberant Nonsense of College Fight Songs* (FiveThirtyEight)

available in the R package **fivethirtyeight**, dataset: `fight_songs`

More info: https://rdrr.io/cran/fivethirtyeight/man/fight_songs.html

bpm for fight songs written by student writers (32 values):

65, 75, 75, 76, 77, 79, 79, 80, 81, 83, 112, 117, 133, 135, 135, 136, 137, 137, 138, 140, 142, 145, 148, 150, 152, 153, 153, 155, 159, 160, 162, 162

bpm for fight songs written by nonstudent writers (33 values):

71, 72, 72, 75, 77, 81, 90, 132, 134, 137, 139, 139, 139, 144, 145, 145, 146, 146, 147, 149, 149, 150, 150, 151, 151, 152, 153, 153, 158, 168, 168, 178, 180

The data is also available here: fight_songs.csv as well as in the Files section of CourseWorks.

(To download the file, click "Raw" and then File, Save in your browser.)

a) Draw a density histogram for bpm for fight songs written by students. Use a binwidth of 20.

b) Draw a cumulative frequency histogram for bpm for songs written by students (same data as in part b.)

c) Draw boxplots comparing beats per minute (bpm) for songs written by students versus those not written by students. How do they compare?