Refer to the syllabus for instructions on how to complete this assignment. Questions marked with *** denote that a teacher solution is required.

1. Which statistical measure(s) (mean, median, mode, range, standard deviation) would you want to use to answer the following questions? More than one can apply. Also say (no need to draw) what graph(s) you would use to represent relevant data. Justify your answers.

   (a) Which of the seven Harry Potter books is the favorite?
   (b) How many M&M’s are in a regular 8oz bag?
   (c) How many hours per week does the average college student spend exercising?
   (d) What is the average starting salary for a UW graduate?
   (e) How predictable is the actual arrival time of the 8am bus?

2. (a) In 1960, the mean family income was about 11 percent above the median. By 1970, that figure was 12.5 percent. By 1980, it was 14 percent. A big jump came in the 1980’s, during which it rose to 21 percent. In 2000, the mean family income exceeded the median by nearly 29 percent.
   
   i. Discuss what it might mean that the median family income is lower than the mean.
   ii. Discuss what it might mean that this difference is growing.

   (b) A poll reports that out of 100 families surveyed, the mean number of children per family was 2.038, the median was 1.9, and the mode was 1.82. I claim that each of these numbers was incorrectly calculated. How do I know? (Hint: Don’t compare; look at each alone.)

3. Make sure to justify your reasoning in each of the following; if you construct a set or pair of sets, explain your thought process, rather than simply showing it satisfies the requirements.

   (a) For each of the following, either construct a data set satisfying the requirements (note: simple and small is fine), or explain why it is impossible.
   
   i. A set with a positive mean and a negative median
   ii. A set with a standard deviation of 0, and with a mean different from the median

   (b) For each of the following, either construct a pair of data sets satisfying the requirements, or explain why it is impossible.
   
   i. Two sets that have the same mean and the same median, but different ranges
   ii. Two sets that have different means, but the same range and the same standard deviation
4. Consider the following statistics scenarios.

(a) A technology firm discovered that 40% of all sick days were taken on a Friday or a Monday, so they immediately clamped down on sick leave to avoid their workers using sick days just to extend their weekends. Why is this somewhat silly?

(b) During the First World War, tin helmets were introduced to the front line (previously they had been cloth). After the introduction of tin hats, the number of injuries to the head increased dramatically. Should one conclude that tin helmets were worse than the older helmets? Think about what data is not mentioned here.

(c) Here is an excerpt from a 2007 survey: “By some estimates, the average [i.e. mean] American household has over $9,300 in credit card debt. Yet, despite Americans’ concern about their spending habits, few people are willing to own up to their balances: over 90 percent of survey respondents believe they had the same amount (as) or less debt (than) the average American.” The survey concludes that 90% of Americans are either liars or in denial about how much they owe on credit cards–but this conclusion is bogus. Why?

5. *** Suppose a lake has only two kinds of fish, bass and pike. The bass population has a mean weight of 6 (pounds) and a median of 2. The pike population has a mean weight of 7 and a median of 6.5.

(a) What might it mean, in context, that the mean and median for the bass population are far apart? Similarly, what might it mean that the numbers for pike are close?

(b) Suppose the DNR (Department of Natural Resources) is interested in the mean and median weight of the lake’s entire fish population (so, bass and pike combined). Explain why you cannot find these numbers given only the above information.

(c) Suppose I tell you that there are twice as many pike as bass in the lake. Which statistic can you now calculate: the mean or the median? Why, and what is it? Why are you still unable to calculate the other statistic?

(d) Find the five-number summaries and draw the box plots for this data of the weights of the bass population and pike population respectively:

\{0.25, 0.25, 0.5, 2, 3, 15, 21\}
\{0.5, 0.5, 1, 1, 2, 6, 6.5, 6.5, 7, 8, 10, 11, 15, 23\}

(e) Use your two box plots to compare these two data sets. Also describe what known information the box plots do not show.

6. Colleen is in a bike race. The following graph depicts the total distance she’s traveled as time progresses.
(a) Write a short story about the graph that explains the graph’s general shape.

(b) What does the steepness of the graph tell you about Colleen’s biking?

(c) Draw a graph below depicting Colleen’s speed as time progresses, and explain how this graph relates to the one above. You should be careful that important events in the first graph match up in time with the same events in the new graph.