1. If the standard deviation of a dataset is small, what does this tell you about the data points?

(A) They are widely dispersed.
(B) They are clustered closely around the mean.
(C) The data is skewed.
(D) The data is symmetrical.

2. Consider the small data set from the notes.

6 7 7 7 8 8 9 9 10 11 11

If every value in this data set is increased by 10, what happens to:

- the mean and median of the set?

(A) Both remain the same.
(B) Both increase by 10.
(C) The mean increases by 10, but the median remains the same.
(D) The mean remains the same, but the median increases by 10.

- the variance and standard deviation of the set?

(A) Both remain the same.
(B) Both increase by 10.
(C) The variance increases by 10, but the standard deviation remains the same.
(D) The variance remains the same, but the standard deviation increases by 10.

3. For which one of the following distributions will the median be a better measure of center than the mean?

(A) Salary data for players in the National Basketball Association (NBA) where most of the players earn the league minimum and a few superstars earn very high salaries in comparison.
(B) Repeated weight measurements of the same 1.6-ounce bag of Peanut M&Ms by students in a large chemistry class using a non-digital balance scale.
(C) Height data from a large random sample of men.
(D) Exam scores with a central peak around an average test score and a few students scoring lower than the average and a few scoring higher.

4. On the next page, sketch your best sense of the distribution of the following variable(s). **Just make up plausible data and values - there is no real data here.** For each, please:

i. Use a form of statistical graphic that emphasizes the important elements of the distribution.
ii. Label the axes and provide plausible values for the tick marks.
iii. (If non-categorical) Describe in words the shape of the distribution.
iv. (If non-categorical) State which measure of center and spread would be most appropriate and approximate their values.

Make a note of any assumptions you're making in interpreting these variable names.

How do Berkeley students get to campus?

Scores on a quiz where most do well but a few score poorly

The `mpg` data frame is available as a part of the `tidyverse` library. It contains information on fuel consumption for 38 models of car between 1999 and 2008. *Datasets can have help files, too!* You do not need to include code for loading in libraries or accessing help files in your answers to the below questions.

5. Write `dplyr` code to calculate the median and IQR city miles per gallon for the vehicles in the data frame and copy it below. The result of your code should be one data structure.

6. Write `dplyr` code to calculate the mean and standard deviation city miles per gallon for the vehicles in the data frame *for each class of car* and copy it below. The result of your code should be one data structure.