

Below is a smaller version of the data from a future lab called `flights_mini`. It contains all flights out of Oakland (OAK) from December 2020. This data frame is used to create the plot that follows. `distance` refers to the distance a given plane travels on its flight, measured in miles. `carrier` refers to the carrier code for a specific airline.



- Which of the following interpretations of the plot above are true? (select all that apply)
  - The carrier with the most heavily skewed distance distribution is HA.
  - The median distance of the flights operated by DL, G4, and OO are roughly equivalent.
  - The minimum distance traveled in this data set is roughly 200.
  - There is no clear association between the carrier and the distance of their flights.
  - The carrier with the greatest variability in distance, as measured by the IQR, is AS.

Consider the small data set from the notes.

6 7 7 7 8 8 9 9 10 11 11

- The data set above was measured in meters, but what would have happened if it had been measured in decimeters (10 decimeters to a meter)? Provide reasoning for would happen to the measures of center - mean, median, mode - if it had instead been measured in decimeters. Repeat the exercise for three measures of spread: range, standard deviation, and IQR. Which measures remain the same after a multiplicative change in units?

3. Sketch your best sense of the distribution of the following variable(s). For each, please:
- Use a form of statistical graphic that emphasizes the important elements of the distribution.
  - Label the axes and provide plausible values for the tick marks.
  - Describe in words the shape of the distribution.
  - 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.

*Number of body piercings* among Stat 20 students

*Scores on an easy quiz* among Stat 20 students

The mpg dataset 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.

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