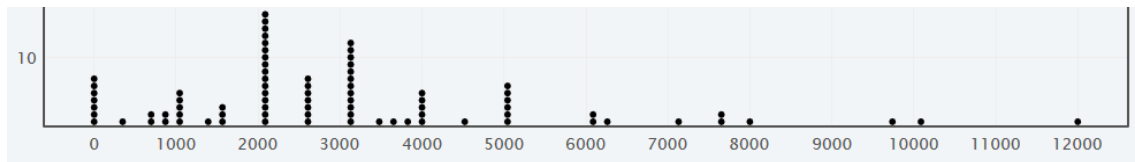

The goal of this lab is to analyze mean, median, and mode in certain data sets. Please open up MeanMedianMode.py file

1 MEAN, MEDIAN, MODE: SOME BASIC FACTS

1. Every semester, I ask my intro stats students to report their annual salary. A dot plot of their results are below. Based on this dot plot, what would you say a typical college student makes per year? Do not do any math or code to answer this. Just look at the plot. Write your answer here: 4000



2. In the MeanMedianMode.py file, there is a list called salaries that contains the salaries in the dotplot above. Use the statistical methods we learned in class to find the mean, median, mode of salaries as well as the number of elements in salaries and the sum of salaries. Write your answers below.

Mean: 3012.81 Sum: 24404838
 Median: 2500 Number of Data Points: 81
 Mode: 2000

3. Now, depending on the application, people often report either the mean/median/mode as a number that represents a “usual data point.” Fill in the blanks below with the info you found above. Do all sentences seem sensible? Do you agree with one more than another?

- College students make about 3012.81 per year. (mean)
- College students make about 2500 per year. (median)
- College students make about 2000 per year. (mode)

4. Now, let’s say Bill Gates decides to audit my stats class. Bill Gates makes 11.5 billion dollars a year (Sp \$11,500,000,000/year). First, let’s wrap our heads around how big that is. Write code in your Python file to determine the following information and write your answer below.

\$11.5 billion/year is the same as:

\$ 31506849.32/day
 \$ 1312785.39/hour
 \$ 21879.76/min
 \$ 364.66/sec

5. Since Bill Gates is auditing my stats class, I have to unclude his information in the salaries data above. Add an element to the salaries list for Bill Gates annual salary. Rerun your code and fill in the information below.

Based on the salary data we have with Bill Gates in our class:

- College students make about 140246878.51 per year. (mean)
- College students make about 2500 per year. (median)
- College students make about 2000 per year. (mode)

6. Do all sentences in #5 seem sensible? Do you agree with one more than another? Which of the mean, median, mode would you report as representing a typical college student's salary now? Why?

I agree with the median in this case cause good old bill gates has schewed that data with such an outlier

7. What you are noticing is that outliers can skew a mean so that it falls left or right of the median. Let's get some more practice. Would you expect the data sets below to have relative frequency distributions which are symmetric, skewed right, or skewed left? Be able to explain. (Hint: When I do these problems, I ask myself, "Who are the outliers?")

- The grades on an easy test. low scores
- The grades on a difficult test. high scores
- The amount of time students in your class studied last week. (hopefully) the kid who didn't study
- The amount of time all PA majors on campus studied last week. the kid who didn't study
- The ages of NFL players. Adam Vinatieri is 46 and still playing

8. Of course, there are more than 3 just distributions that are possible. For example, the histogram below shows the weights for a subset of 2018 NFL teams.

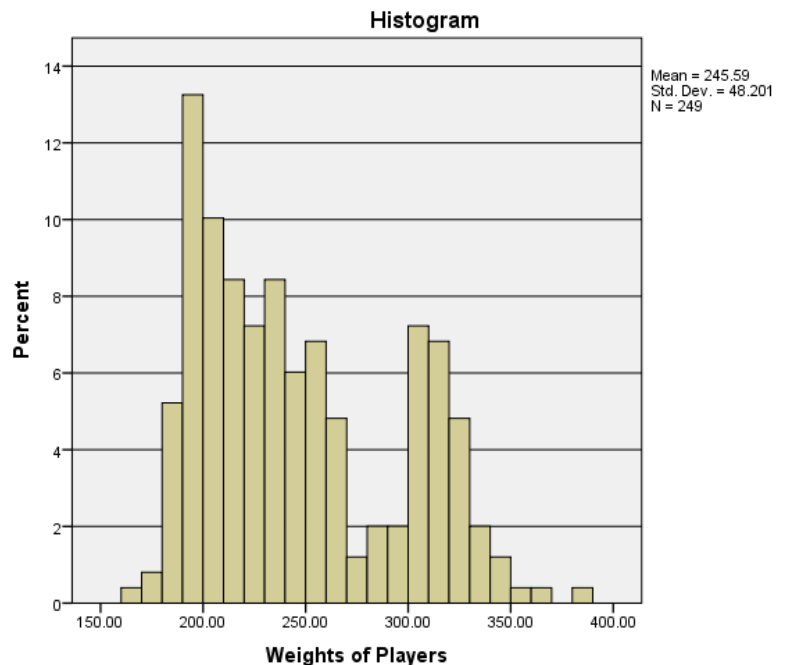
- a. We call this a bimodal ("two modes") distribution. Why does this name make sense?
cuase there are 2 possible modes

- b. Notice the average weight for these players is about 246 pounds. Would you be willing to say that a typical NFL Player weighs 246 pounds? If not, what would you report?

I would not becuase of the huge amount of ones that weigh 200 and those that weigh 310

- c. Explain what about NFL player weights would cause a distribution shape like this.

You have defensive linebackers who need to be larger and then you need recivers who are a bit faster



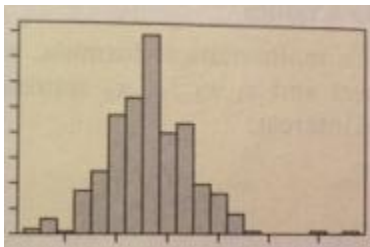
2 SUMMARIZING SOME MEAN, MEDIAN, MODE INFORMATION.

1. True or False: Every data set has a mean. True
2. True or False: Every data set has a median. True
3. True or False: Every data set has a mode. False
4. True or False: Every data set has a single, unique mean. True
5. True or False: Every data set has a single, unique median. True
6. True or False: Every data set has a single, unique mode. False
7. True or False: The mean is affected by outliers. True
8. True or False: The median is affected by outliers. false
9. True or False: The mode is affected by outliers. false
10. A fast food restaurant reports that the average employee salary is \$13.23 per hour, but an employee states that most employees make minimum wage (\$7.25 per hour). Is this possible? Explain.

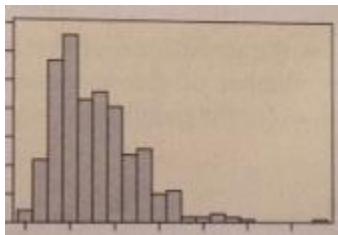
Some employees may be making a significant amount more and thus skewing the data

11. As in the NFL example, reducing an entire data set to one number (like mean, median, mode) does not always give you a full picture of the data like a visual representation does. A visual representation like a histogram can help you visual a data set greatly. Consider this example.

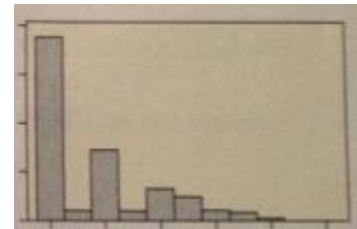
Let's say I collected data from each of my students in my classes asking how many hours of exercise per week the student gets, the Pulse rate (beats per minute) of the student, the number of body piercings (counting each hole once). Below are the resulting 3 frequency distributions for this variables. Guess which is which and be able to explain your guess.



symetric



skewed right



skewed right

3 SUMMARIZING SOME MEAN, MEDIAN, MODE INFORMATION.

1. In the MeanMedianMode.py file, there is a list called baseballSalaries that contains the salaries for all baseball players in the 2015 NBA season. Write code to find these values below.

Mean: 4.21

Sum: 3658.36

Median: 1.65

Number of Players Recorded: 868

Mode: .51

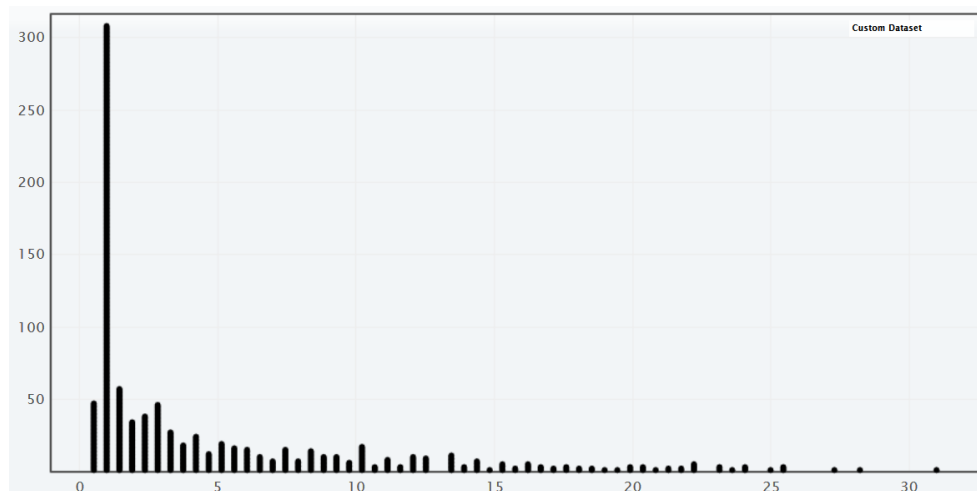
2. Based on the above numbers, is the data skewed left, skewed right, or symmetric? skewed Right

3. The salaries of **superstar professional athletes** like NBA players receive much attention in the media, and it is common that these superstar athletes have multimillion contracts. However, we often hear in the news that players' associations and team owners are negotiating for additional salary benefits for **all players on a team, not just the few superstars**. If the NBA wishes to argue for an overall higher salary for its players, should the association report the mean or median of the players' salaries? **Why? Answer BRIEFLY, in 1-2 CONCISE/COMPLETE sentences.**

The NBA should report the median because The median is more accurate for the whole team and is mostly unaffected by the larger outliers

4. As discussed above, reducing the data set to a few summary numbers does not give you a full picture. Below is the dotplot for the baseball player salary data. How does this visualization support your previous 2 answers?

the large amount of approx. 2 will skew all of the possible data and pretty



SUBMISSION INFO

TO GET CREDIT FOR THIS LAB, UPLOAD THESE DOCUMENTS TO THE SUBMISSION AREA.

- LabCh3E.pdf (should have your guess work)
 - MeanMedianMode.py (make sure your name is on top)
-