CHAPTER 1: Intro to Statistical Graphs

Section 1.1 Types of Data

Given a large set of data, we would like to "make sense" of it in some way. Because data comes in many different types, we frequently have different ways of making sense of the different types of data.

Learning Goal:

There are many ways we can distinguish between different types of data. We have different methods to analyze and graphically represent different types of data. In this section, we want to understand one of the fundamental distinctions: qualitative data vs. quantitative data.

Specific Learning Objectives:

1. You will understand the difference between qualitative and quantitative data.
2. You will understand the difference between the population and a sample.
3. You will understand the difference between a statistic and a parameter.
4. You will understand the difference between discrete and continuous data.

Qualitative vs. Quantitative Data

Quantitative Data measures the amount of something and uses numbers. The prefix "quant" means "amount of" (like the word "quantity").

Qualitative Data measures a quality. It is not an amount, and serves as more of a description or name.

Example 1: Suppose I poll 60 elementary school students and ask them:

"What is your favorite ice cream flavor?"

Ice cream flavor is NOT a numerical quantity. This question does not result in quantitative data. We call data this type of data **qualitative data**. (Instead of thinking about "quantities" we are thinking about "qualities". Qualities include things like gender, major, car brands, names, places, etc.)

In practice, one could have many, many qualities that result from a single poll question. For example,

"What is your name?"

is a question that results in a lot of qualitative data. Unfortunately, analyzing this type of data is often difficult because almost every piece of data is different. What we typically do is break qualitative data down into a manageable number of categories that we can analyze. For this reason, the terms "qualitative data" and **categorical data** are frequently used interchangeably in the context of statistics.

Example 2: Suppose I poll 100 college students and ask:

"How many times did you go to office hours to seek help when you had difficulty in class?"

The type of data that results from this type of question is **quantitative data**. ("quant" is the root of the word "quantity" which measures "how much of something we have." We need numbers to measure such things.)

Example 3: Sometimes, data made up of numbers is actually qualitative data. For example,

"The 210 freeway" is qualitative data even though it has a number in it, because 210 is not measuring the amount of anything it is just serving as a name and an indicator that it is the freeway that connects highway 2 to the 10 freeway.

Example 3 is a great example of data that involves numbers but is *not* quantitative data. Zip codes and social security numbers are also not quantitative data.

A good way for you to tell whether a number is acting as quantitative data is to ask yourself – "Is this number measuring the amount of anything?" If not, then the number is probably not quantitative data.

Example 4: Sometimes we can take qualitative data and turn it into quantitative data. For example, Customer satisfaction ratings of "Poor", "Fair", "Good", and "Excellent" could be represented with numbers as 1, 2, 3, and 4 respectively. By converting the qualitative data to quantitative data, we can apply various statistical techniques.

This last example is what we do when we calculate your GPA. You are taking qualitative data (your letter grade) and turning it into quantitative data (grade points).

Discrete vs. Continuous Data

Quantitative Data can be further broken two into two types:

Discrete data is data that you can count on your fingers (assuming you had an infinite supply of fingers). For instance, we can count the number of eggs a hen lays similar to how we count on our fingers. Or we could count the number of people in a room on our fingers. But we couldn't count a distance on our fingers with any great accuracy. Nor could we count the exact amount of water a person can drink in an hour on our fingers.

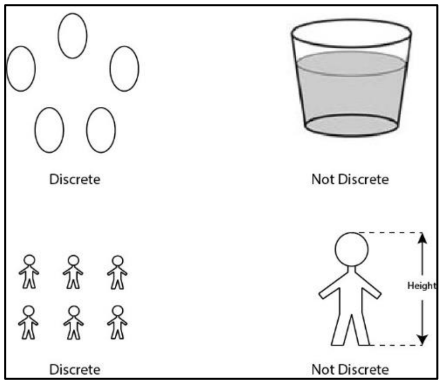
When we can count things in a way similar to how we can use our fingers to count we say these things are **discrete**.

Examples of discrete data: 12 eggs, 5 dollars, 3 classes, 28 students, number of pans in your kitchen

Continuous data is quantitative data that cannot be measured by counting on your fingers. An infinite number of values is possible.

For example, suppose I want to measure precisely how tall someone is. Our convention is to give our height in feet and inches, but this is really an **approximation**. It gives our height to the nearest inch. Two people who claim to be 5 feet 4 inches tall are probably not *exactly* the same height.

Examples of continuous data: 6ft 2in tall, 128 lbs, amount of water a person can drink in 30 seconds

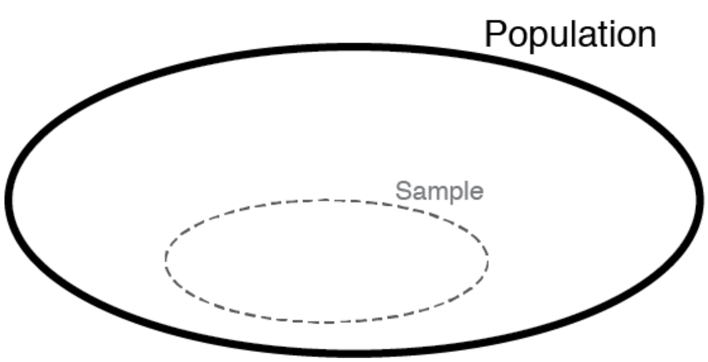


Population vs. Sample

In statistics, we frequently want to understand data that comes from a particular group. For example, suppose I want to have a better understanding of the SCC student body.

The **population** is the entire group we are looking to study. In this case, our population is all students at SCC.

Since it is generally not possible to collect data on an entire population (we don't have the time to survey all 22,735 students) we collect data from a smaller group or subset taken from the population. This smaller group is called the **sample**.



An example of a sample for our SCC student population would be all the students in our Math 112 class. In general, we would like our sample to do a good job of representing our population. Using Math 112 students as our sample probably won't do a good job of representing all of the students at SCC. Math 112 was designed for certain majors, not for all students at SCC.

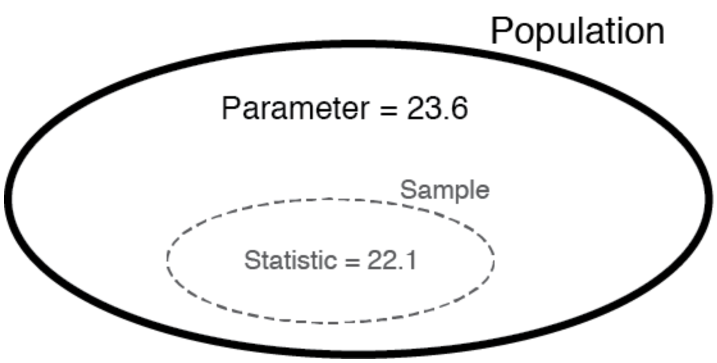
Parameter vs. Statistic

A number the represents a characteristic of the population is called a **parameter**.

A number the represents a characteristic of the sample is called a **statistic**.

For example, I could define my population to be all SCC students.

- If we calculated the average age of all SCC students and found it to be 23.6 years old, this average would be a parameter, because it was calculated using the ages of ALL the students at SCC.
- If I randomly selected 100 SCC students and calculated that their average age was 22.1 this would be a statistic, because I only used the ages of a sample of the whole population.



Now that we understand some of the basic vocabulary of Statistics, we can talk about the characteristics of data. In this class we will study three main characteristics of data:

1. **Measures of Center** – what is the middle of the data set?
2. **Measures of Spread** – is there a way to indicate how spread out the data is?
3. **Distribution** – what is the nature or shape of the spread of data?

Classwork

Classify each of the following as Quantitative or Qualitative data.

1. The president of SCC is tall. ________________________________
2. The fire fighter is 5 ft 10 inches tall. ________________________________
3. The area code for SCC is 707. ________________________________

Classify each of the following pieces of quantitative data as either Continuous or Discrete.

4. Number of math classes you have taken. _________________________________
5. The speed of the Amtrak train. ____
6. The length of your foot. ________________________________
7. Your shoe size. ________________________________

For each pair determine which is the sample and which is the population.

8. The SCC women's Basketball team and students at SCC

9. People with college degrees and the teachers teaching Math 112

10. Students at SCC and Students enrolled at a community college.

Determine which is a statistic and which is a parameter.

11. The average height of the SCC's basketball team and the average height of students at SCC.

Homework

For each of the following, write your answer on the lines provided.

1. Classify each of the following as qualitative or quantitative:
    a. Color of a picture frame is black: ____________________
    b. Picture weighs 8.5 pounds: ____________________
    c. Picture costs $300: ____________________
    d. Serving temperature of a latte is 150 degrees: ____________________
    e. Latte size is “tall”: ____________________
    f. 12 ounces of latte: ____________________
    g. He has a friendly demeanor: ____________________
    h. Class has 20 females and 8 males: ____________________
    i. Social Security Numbers: ____________________
    j. Letter grades: ____________________
    k. Zip codes: ____________________

2. Classify each piece of quantitative data as continuous or discrete:
    a. Number of children in a household: ____________________
    b. Height of a hobbit: ____________________
    c. Time to wake up in the morning: ____________________
    d. Number of languages a person speaks: ____________________
    e. Speed of a train: ____________________
    f. Distance to the nearest Starbucks: ____________________
    g. Temperature: ____________________

3. Classify each group as a population or sample:
    a. 150 randomly selected SCC students: ____________________
    b. All iPhone owners in the world: ____________________
    c. 5 students from the class chosen randomly: ____________________

4. Samples and Populations: For each of the following, create a sample or population that fits the situation by filling in the blank (note: a **good** sample is one that is **randomly selected**, so these examples are not **good** samples).
    a. Students in this class is a sample of the population___________________
    b. ____________________ is a sample of the population of all Students in this Class.
    c. Students taking Math 112 is a sample of the population __________________.
    d. ____________________ is a sample of the population of all Students taking Math 112.
    e. All SCC students is a sample of the population__________________.
    f. ____________________ is a sample of the population of all SCC Students.

5. Classify each numerical measure as either a statistic or a parameter:
    a. In the Substance Abuse and Mental Health Services Administration survey, 13.2% respondents said they have driven under the influence of alcohol: ______
    b. There are 50 state capitols of the U.S.: ____________________
    c. Class average of all Math 112 students last semester was 81.3%: ____________________
    d. In a 2011 study by the Williams Institute at the UCLA School of Law, the percentage of LGBT youth that participated in the study was 1.7%: ____________________
    e. In a Gallup poll of 1023 renters, it was found that 58.2% of them said that they had only wireless phones.

       The 1023 number is a______________. The 58.2% value is a ________________.

Section 1.2 Graphs for Qualitative Data with percent review.

One way to make sense of a large set of data is to make graphs to display the data. In this section we will learn about graphs that are used to display qualitative data.

Learning Goal:

To learn how to display qualitative data in an appropriate graph, and to learn how to use qualitative graphs to analyze situations.

Specific Learning Objectives:

1. You will be able to create a frequency distribution, bar graph, and circle graph by hand and/or using technology.
2. You will be able to use the information provided in a frequency distribution, bar graph, and circle graph to answer questions.
3. You will be able to covert between the decimal, fraction, and percent forms of a number.

1.2.1: Frequency Distributions

Definition: A frequency is a count of how many data values lie in the given category. A frequency distribution is a table which lists all of the categories, and their frequencies.

Example 1: Consider the following information: A student took a survey of the 15 students in her English class. She asked the students which was their eye color.
The results were: Blue, Green, Brown, Brown, Brown, Brown, Blue, Blue, Brown, Brown, Brown, Brown, Brown, Blue, Green.

In this example there are 3 categories of responses (Blue, Green, and Brown).
Counting the frequency (number of occurrences) for each category we see that 2 of the students responded green, 4 students responded blue, and 9 students responded brown.

The frequency distribution that displays this information would be:

Category	Green	Blue	Brown
Frequency	2	4	9

The total number of students (15) that are represented in the table can be found by adding together the frequencies: 2 + 4 + 9 = 15

Example 2: Consider the following frequency distribution for favorite ice cream flavor when answering the following questions.

Flavor	Frequency
Chocolate	17
Strawberry	23
Mint Chip	14
Vanilla	34
Other	12

How many people are represented in this frequency distribution (this is called the sample size)? Like in the previous frequency distribution we can determine the total number of responses by summing the frequencies of each category. 17 + 23 + 14 + 34 + 12 = 100

How many people identified that their favorite flavor was not Vanilla? We can answer this questions in several ways. One way would be to sum all the frequencies that are not vanilla: Chocolate (17) + Strawberry (23) + Mint (14) + Other (12) = 66
A second way that we could answer this questions would be to subtract the people who responded vanilla from the total number of people: 100 – 34 = 66. Either way we determined that 66 people of the 100 surveyed preferred a flavor other then vanilla.

What fraction of the people surveyed preferred chocolate? Since 17 of the people of 100 responded they preferred chocolate, the fraction is $\frac{17}{100}$.

What fraction of the people surveyed preferred a flavor other than chocolate?
First we need to determine how many of the 100 people preferred a flavor other than chocolate. Since 17 people preferred chocolate out the 100 surveyed the number who did not select chocolate is 100 – 17 = 83 people. Since 83 of the 100 people choose a flavor other than chocolate the fraction that represents them is $\frac{83}{100}$.

Note: Fractions should be reduced, if possible. In this case neither of the fractions we wrote could be reduced since the numerator and denominators do not share a common denominator.

1.2.1 Class Activity

Example 1: Class Survey: Let's gather some class data: Please consider your favorite color (of the categories listed below. As your class responds keep track of the responses.

Which of these colors is your favorite Color?

Blue
Green
Red
Pink
Purple
Yellow
Other

Let's Make a frequency Distribution:

Color	Blue	Green	Red	Pink	Purple	Yellow	Other
Frequency							

Using the frequency distribution, what is the sample size? ________

Using the frequency distribution, how many people preferred a color other than green? ___________

Using the frequency distribution, how many people preferred either red or pink? ___

What fraction of the people surveyed preferred green? _______

What fraction of the people surveyed preferred a color other than green? _________

1.2.1 HOMEWORK

1) Consider the following: The data was collected for the question: What grade did you earn in your last math class?

Data: A, A, A, A, A, A, A, B, B, B, B, B, B, B, B, B, B, B, B , C, C, C, C, C, C, C, C, C, C, C, D, D, D, D, D, D, D, D, F, F, F

A) Fill in the frequency distribution:

Grade	A	B	C	D	F
Frequency					

B) What is the sample size? _____________

C) How many people passed (earned a C or higher)? __________

D) How many people did not earn an A? _________

E) What fraction of the people in the survey earned an A? _________

F) What fraction of the people surveyed did not earn an A? ________

G) What fraction of the people surveyed passed the class? ________

H) What fraction of the people surveyed did not pass the class? ________

I) OF the students who passed, what fraction passed with an A? ______

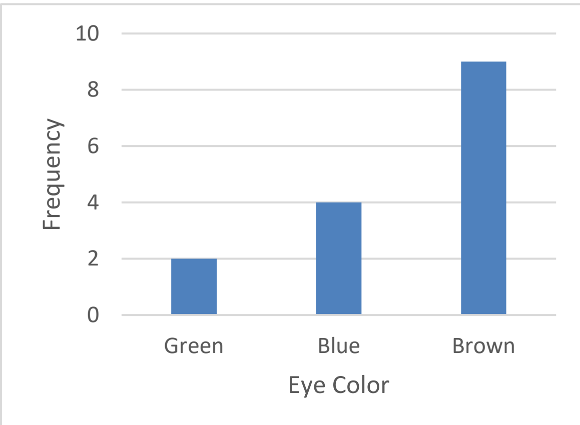
Tip: In this case the fraction is not out of all people surveyed, but only out of those who passed (earned a grade of A, B, or C).

1.2.2 Bar Graphs

A bar graph is a visual representation of the frequency distribution for qualitative data. The vertical axis records the frequency. The horizontal axis we record the categories (in an order we choose). Then for each category we draw a bar. The height of the bar is the frequency for that category. In bar graphs the bars do not touch.

Let's look at the bar graph for the frequency distribution in 1.2.1 example 1:

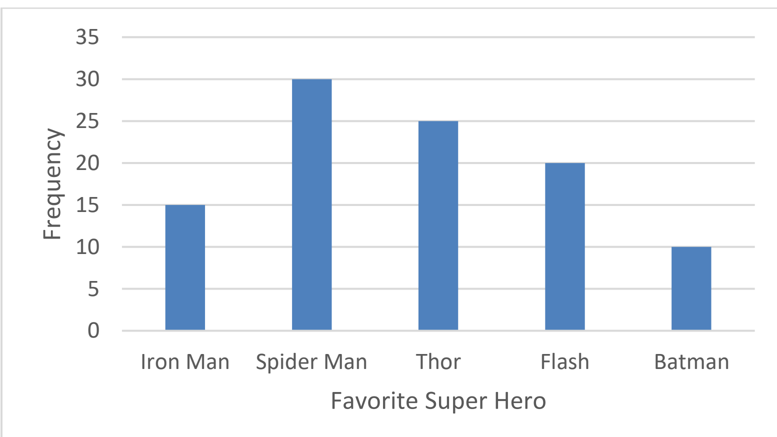
Category	Green	Blue	Brown
Frequency	2	4	9



On the graph, notice the scale is 2 (meaning that each mark on the vertical axis represents 2).

It is also important to remember that the bars can be placed in any order. We can't make any conclusions based on their order, but we can visually compare the frequencies by looking at the difference in heights of the bars.

Example 2: Consider the following bar graph:



We can tell that (according to this graph) Spider Man was the most popular superhero among those surveyed because it has the tallest bar. We can also determine that Batman was the least popular because that category has the shortest bar. While this information can also be determined from the frequency distribution, when there are many categories the visual heights can make comparisons quickly.

On this graph notice that the marks on the y-axis (the scale) is counting by 5's. This is because the numbers are much larger, so to reach the needed frequencies we need to use a larger scale then in the previous bar graph.

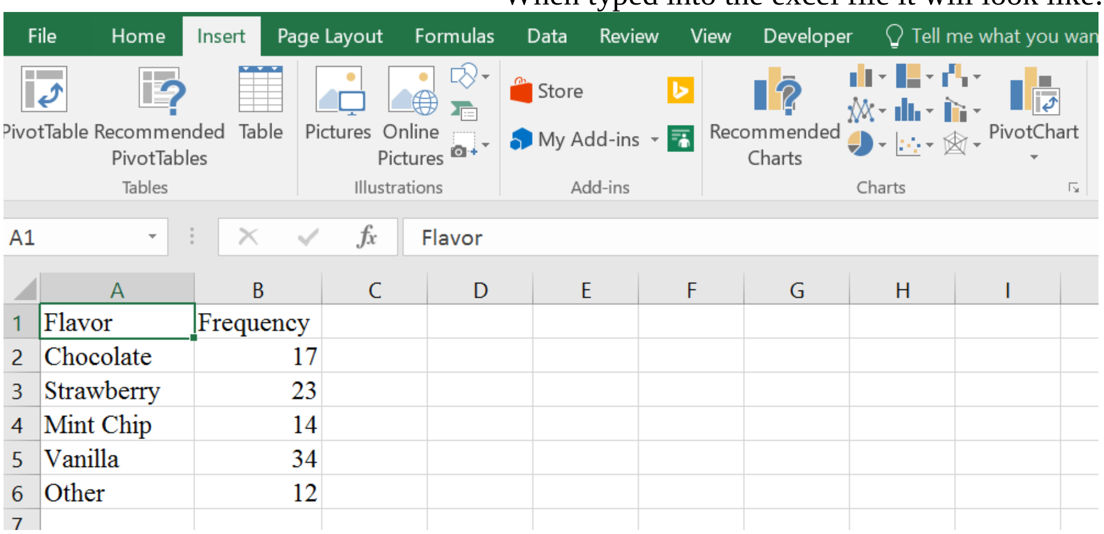
Making a bar graph using Excel/Google Sheets

Step 1: Gather your data into a frequency distribution.
Step 2: Type the categories and frequencies into the excel/ google sheet document
Step 3: Highlight the entire frequency distribution
Step 4: Select "Insert", then choose the bar graph you desire.
Step 5: Edit chart features as desired to make a nice presentation. In excel you can edit the chart features by clicking on the + you see on the right side of the chart.

Example: Consider the frequency distribution for Favorite Ice Cream flavor in 1.2.1 example 2:

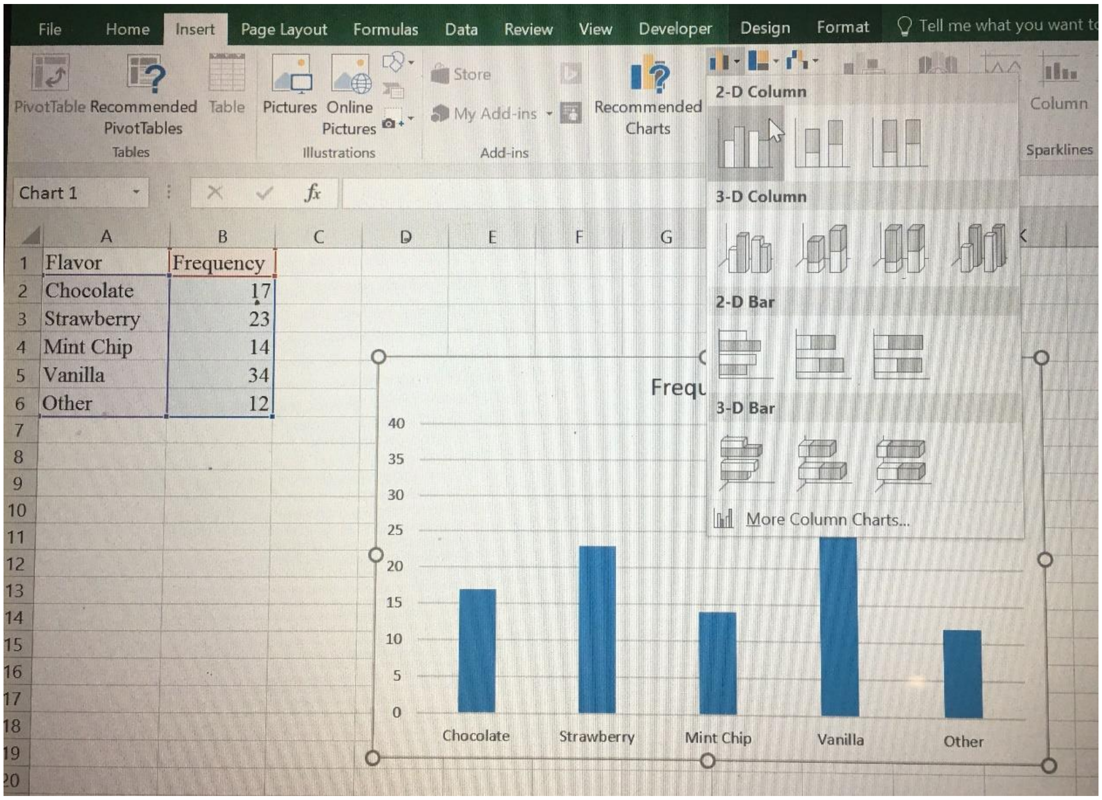
Flavor	Frequency
Chocolate	17
Strawberry	23
Mint Chip	14
Vanilla	34
Other	12

When typed into the excel file it will look like:

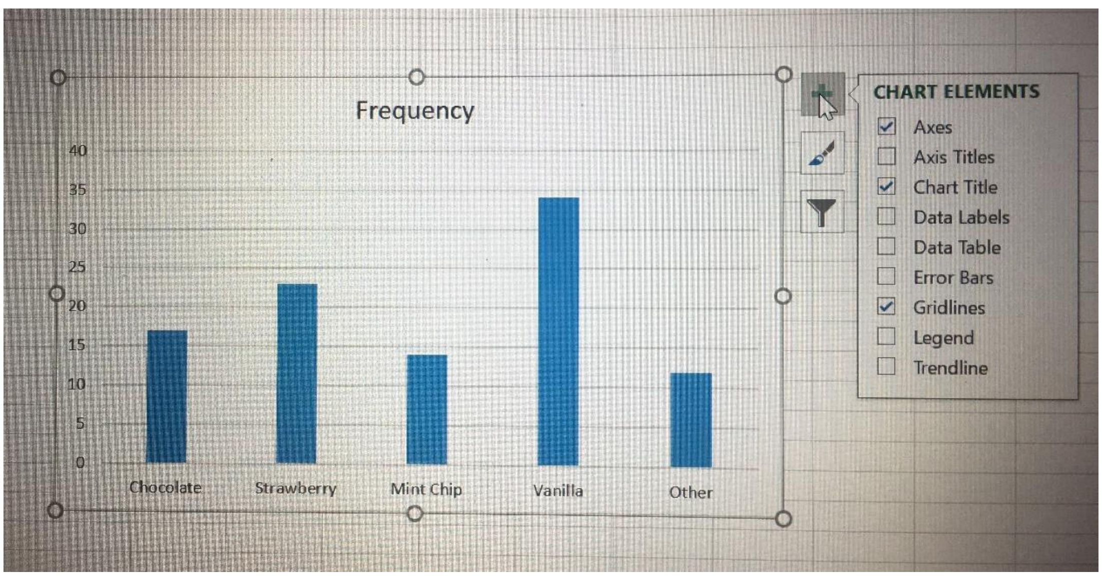


Each entry in the excel document is called a cell. We refer to each cell by the letter at the top of the column, and the number of the row. For example, the word Flavor is in cell A1 and the number 23 is in cell B3.

Next to create the bar graph we select Insert on the top. Then press the arrow next to the graph that looks like a bar graph, and select the first option. (As shown on the next page).



For the last step, click on the + on the right side of the graph and adjust any of the features that make your graph clearer. You can also edit any of axis or chart labels by clicking on the labels.



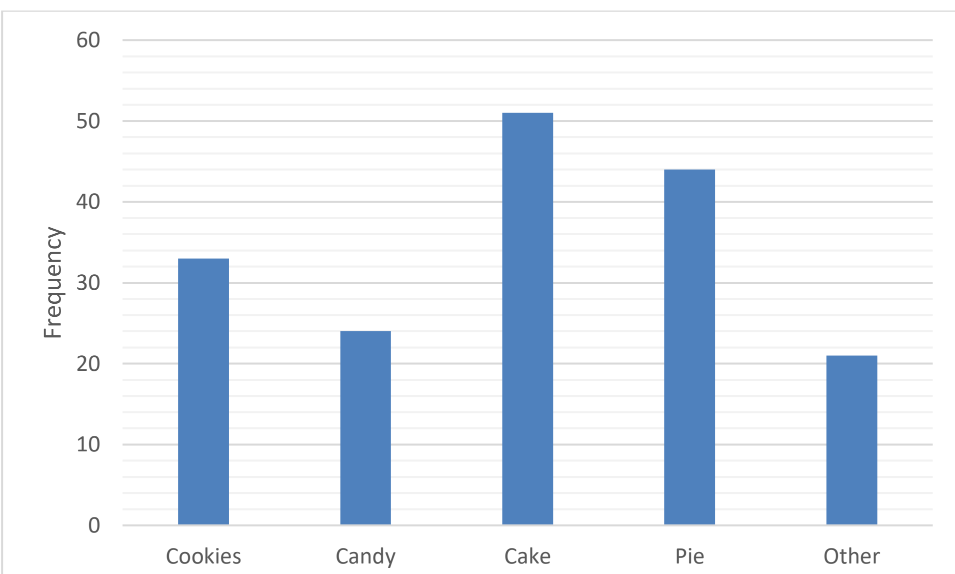
Some key elements:
Axis Titles: Allows you to title each axis
Chart Title: Allows you to title the entire chart
Data Labels: Will display the height of each bar
Gridlines: Those are the horizontal lines that help to compare the bars to the vertical axis.

1.2.2 Class Activity

1) For the class data on favorite color (located on page 11) create a bar graph that displays that data. Consider the scale for the y-axis that will help you display the data best. Make sure to label both the horizontal and vertical axis.

Color	Blue	Green	Red	Pink	Purple	Yellow	Other
Frequency							

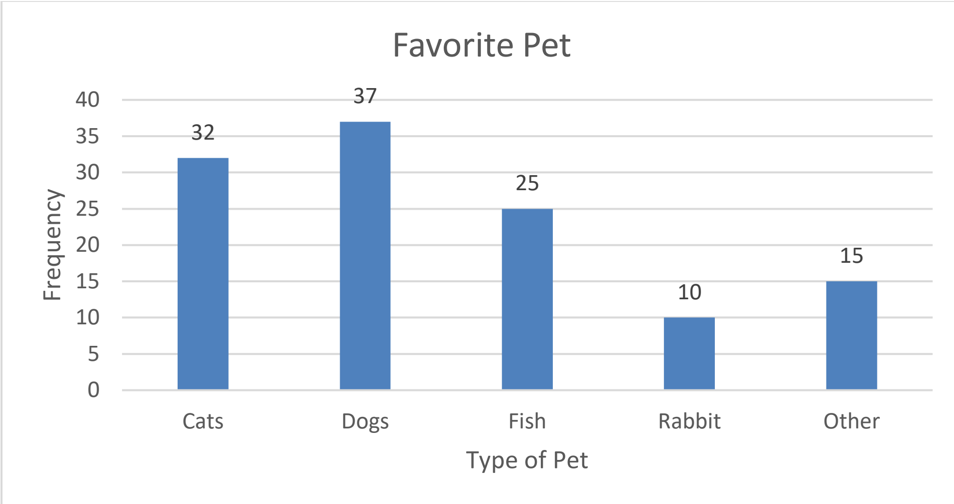
2) For the following graph, fill in the frequency distribution that it represents:



Category					
Frequency					

What is the scale for this graph? ______________

3) Use Excel to reproduce the following graph:



1.2.2 Homework

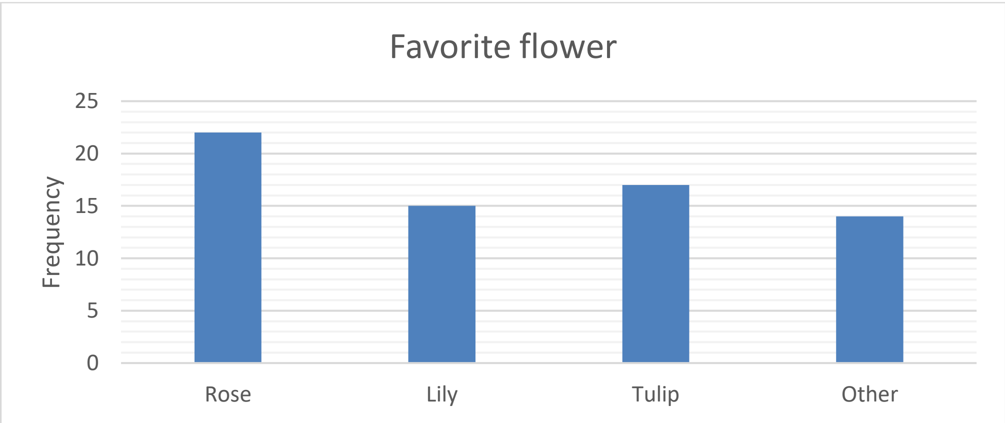
1) Consider the following data from question 1 in homework 1.2.2: What grade did you earn in your last math class

Data: A, A, A, A, A, A, A, B, B, B, B, B, B, B, B, B, B, B, B , C, C, C, C, C, C, C, C, C, C, C, D, D, D, D, D, D, D, D, F, F, F

A) Draw (by hand) a bar graph to display the data below. Make sure to label the chart, and each axis and choose a scale that well displays the data.

B) Use EXCEL of Google sheets to create the graph. Position the graph so that you can see both the frequency distribution you typed and the graph and then print the graph and attach it to your homework.

2) For the given graph, fill in the frequency distribution:



Flower				
Frequency				

1.2.3 Percent review

Percent is defined as per hundred and is denoted by the symbol %. Since percent indicates per hundred, percent are converted into fractions with denominators of 100 or into decimals with the decimal place moved 2 spaces to the left which is a quick shortcut for division by 100. Below some common percentages are written in equivalent fraction and decimal forms.

$$1\% = \text{1 per every hundred} = \frac{1}{100} = 0.01$$

$$10\% = \text{10 per every hundred} = \frac{10}{100} = \frac{1}{10} = 0.10$$

$$20\% = \text{20 per every hundred} = \frac{20}{100} = \frac{1}{5} = 0.20$$

$$25\% = \text{25 per every hundred} = \frac{25}{100} = \frac{1}{4} = 0.25$$

$$50\% = \text{50 per every hundred} = \frac{50}{100} = \frac{1}{2} = 0.50$$

$$75\% = \text{75 per every hundred} = \frac{75}{100} = \frac{3}{4} = 0.75$$

$$100\% = \text{100 per every hundred} = \frac{100}{100} = 1$$

$$200\% = \text{200 per every hundred} = \frac{200}{100} = 2$$

Example 1 Write the following as fractions: 37% 84% 3.5%

$$37\% = \frac{37}{100}$$

$$84\% = \frac{84}{100} = \frac{\cancel{4}(21)}{\cancel{4}(25)} = \frac{21}{25}$$

$$3.5\% = \frac{3.5}{100} = \frac{35}{1000} = \frac{\cancel{5}(7)}{\cancel{5}(200)} = \frac{7}{200}$$

The diagram below displays the relationship between percent, decimal, and fraction formats. The decimal format serves as the middleman between the percent and fraction formats.

Percent to Decimal
Remove the % symbol and divide by 100
Move decimal point 2 places to the left

Decimal to Fraction
Write decimal as a
fraction and reduce

$$75\% \quad = \quad 0.75 \quad = \quad \frac{3}{4}$$

Decimal to Percent
Insert the % symbol and multiply by 100
Move decimal point 2 places to the left

Fraction to Decimal
Divide denominator
into numerator

Below halves (think of a 50 cent coin), thirds, fourths (think of quarters), fifths (think of double dimes), tenths (think of dimes), and twentieths (think of nickels) are written in both fraction and percent formats. Viewing fractions with denominators 2, 4, 5, 10 and 20 respectively as half dollars, quarters, double dimes, dimes, and nickels allows for a quick mental conversion of these fractions into percent form without the use of a calculator.

Halves (half dollars)	1/2 = 50%				
Thirds	1/3 ≈ 33%	2/3 ≈ 67%			
Fourths (quarters)	1/4 = 25%	2/4 = 50%	3/4 = 75%		
Fifths (double dimes)	1/5 = 20%	2/5 = 40%	3/5 = 60%	4/5 = 80%	
Tenths	1/10 = 10%	2/10 = 20%	3/10 = 30%	4/10 = 40%	5/10 = 50%
(dimes)	6/10 = 60%	7/10 = 70%	8/10 = 80%	9/10 = 90%	
Twentieths	1/20 = 5%	3/20 = 15%	5/20 = 25%	7/20 = 35%	9/20 = 45%
(nickels)	11/20 = 55%	13/20 = 65%	15/20 = 75%	17/20 = 85%	19/20 = 95%

Example 2 Fill in the following table

Percent	Decimal	Fraction
28%		
	0.45	
		$\mathbf{\frac{3}{4}}$
120%		
		$\mathbf{\frac{8}{11}}$

Convert 28% to a decimal by removing the percent symbol and moving the decimal point two places to the left. Write the resulting decimal 0.28 as a fraction and reduce.

$$\mathbf{28\%} = 0.28 = \frac{\overset{7}{\cancel{28}}}{\underset{25}{\cancel{100}}} = \frac{7}{25}$$

Convert 0.45 to a percent by inserting the percent symbol and moving the decimal point two places to the right. Write the decimal 0.45 as the fraction 45/100 and reduce.

$$45\% = \mathbf{0.45} = \frac{\overset{9}{\cancel{45}}}{\underset{20}{\cancel{100}}} = \frac{9}{20}$$

Recognize the fraction 3/4 as the decimal 0.75 which equals 75%.

$$75\% = 0.75 = \mathbf{\frac{3}{4}}$$

Convert 120% to a decimal by removing the percent symbol and moving the decimal point two places to the left. Write the resulting decimal 1.20 as the fraction 1 20/100 and reduce.

$$\mathbf{120\%} = 1.20 = 1\frac{\overset{1}{\cancel{20}}}{\underset{5}{\cancel{100}}} = 1\frac{1}{5}$$

Write the fraction 8/11 as a decimal using calculator. The resulting repeating decimal when rounded to three significant digits is 0.727 which equals 72.7%

$$\mathbf{\frac{8}{11}} \approx 0.727 = 72.7\%$$

Percent	Decimal	Fraction
28%	0.28	$\frac{7}{25}$
45%	**0.45**	$\frac{9}{20}$
75%	0.75	$\mathbf{\frac{3}{4}}$
120%	1.20	$1\frac{1}{5}$
72.7%	0.727	$\mathbf{\frac{8}{11}}$

Example 3 The table below gives the grades and gender of students in a class. Use this table to answer the following first as a fraction then as a percent rounded to 1 decimal place.

Gender/Grade	A	B	C	D	F	W	**Total**
Female	5	5	8	3	2	4	**27**
Male	4	5	5	2	3	5	**24**
Total	**9**	**10**	**13**	**5**	**5**	**9**	**51**

Find the percentage of the females in the class that are female A students.
There are 5 female A students out of 27 total female students, approximately 18.5% of the total female students in this class received an A grade.

$$5/27 \approx 0.185 = 18.5\%$$

Find the percentage of all the students in the class that are female A students.
There are 5 female A students out of 51 total students, approximately 9.8% of the total students in this class are females that received an A grade.

$$5/51 \approx 0.098 = 9.8\%$$

Find the percentage of all the students in class that received an A grade.
Since gender is not mentioned, there are 9 A students out of 51 total students, thus approximately 17.6% of the total students in this class received an A grade.

$$9/51 \approx 0.176 = 17.6\%$$

Find the percentage of males that are male students that passed the class.
There are 14 male A, B, or C students out of 24 total male students, thus approximately 58.3% of the male students in this class passed the course.

$$14/24 \approx 0.583 = 58.3\%$$

Find the percentage of all the students that are male students that passed.
There are 14 male A, B, or C students out of 51 total students, thus approximately 27.5% of the total students in this class are males that passed the course.

$$14/51 \approx 0.275 = 27.5\%$$

Find the percentage of A students that are female.
There are 5 female A students out of 9 total A students, thus approximately 55.6% of A students in this class are females.

$$5/9 \approx 0.556 = 55.6\%$$

For fractions, the key word "of" indicates multiplication when written between a fraction and a number such as 3/4 of 12. Similarly for percentages, the key word "of" indicates multiplication when written between a percent and a number such as 75% of 12.

Fractional part of a number

$$\frac{3}{4} \textbf{ of } 12 = \frac{3}{\cancel{4}_1}\left(\cancel{12}^{3}\right) = 9$$

Percentage of a number

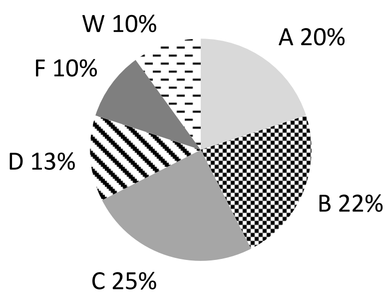
$$75\% \textbf{ of } 12 = 75\%(12) = .75(12) = 9$$

Example 4 Evaluate 3% of 170

To multiply the percent times the number, first write the percent in decimal format by removing the percent symbol % and moving the decimal point two places to the left and then multiply the resulting decimal times the number.

$$3\% \textbf{ of } 170 = 3\%(170) = 0.03(170) = 5.1$$

Example 5 The grade distribution in a class with 40 students is shown below. How many students earned an A grade? How many earned a C grade? How many students were successful in this course?



To determine the number of students that earned an A grade find 20% of the 40 total students as shown below which results with 8 students receiving an A grade.

$$20\% \textbf{ of } 40 = 20\%(40) = 0.20(40) = 8$$

To determine the number of students that earned a C grade find 25% of the 40 total students as shown below which results with 10 students receiving a C grade.

$$25\% \textbf{ of } 40 = 25\%(40) = 0.25(40) = 10$$

Successful students are defined as those that earned an A, B or C grade, to find the success rate in this course add 20%, 22% and 25%. Then, find 67% of the 40 total students as shown below which results with 27 students successful in this course.

$$67\% \textbf{ of } 40 = 67\%(40) = 0.67(40) = 26.8 \approx 27$$

COMMON CALCULATIONS: Calculations involving 1%, 10%, 50%, 100% and 200% of a number can be calculated mentally using the steps shown below.

1% of a number is equal to **one hundredth of** the number
To find 1% of a number multiply by 0.01 which moves the decimal point in the number two places to the left.

10% of a number is equal to **one tenth of** the number
To find 10% of a number multiply by 0.10 which moves the decimal point in the number one place to the left.

25% of a number is equal to **one fourth of** the number
To find 25% of a number multiply by 0.25 which results in one fourth of the number which is calculated by simply dividing the number by four

50% of a number is equal to **half of** the number
To find 50% of a number multiply by 0.50 which results in half of the number.

100% of a number is equal to **all of** the number
To find 100% of a number multiply by 1.00 which leaves the number unchanged.

200% of a number is equal to **double** the number
To find 200% of a number multiply by 2.00 which doubles the number.

Example 6 Evaluate the following mentally without showing steps.

1%(80) = 0.**80** = 0.80
To find a hundredth of 80, move the decimal point 2 places to the left

10%(80) = 8.**0** = 80
To find a tenth of 80, move the decimal point 1 places to the left

25%(80) = 20
To find 25% of 80, calculate one-fourth of 80 by dividing 80 by 4

50%(80) = 40
To find 50% of 80, calculate one half of 80 by dividing 80 by 2

100%(80) = 80
100% of 80 is 80

200%(80) = 160
To find 200% of 80, simply double 80

Example 7 Evaluate the following without the use of a calculator:

100%(400) = 400 Leave total unchanged

10%(400) = 40.**0** = 40 Move the decimal place 1 space to the left

1%(400) = 4.**00** = 4 Move the decimal place 2 spaces to the left

25%(400) = 100 Divide 400 by 4 (same as multiply by 1/4)

50%(400) = 200 Divide 400 by 2 (same as multiply by 1/2)

300%(400) = 3(100%)(400) = 3(400) = 1200
300% is 3 times 100%, first multiply by one then triple

70%(400) = 7(10%)(400) = 7(40) = 280
70% is 7 times 10%, first move decimal place one space to left then multiply by 7

2%(400) = 2(1%)(400) = 2(4) = 8
Since 2% is 2 times 1%, first move decimal place two spaces to the left then double

Example 8 Assuming that 10% of humans are left handed how many left handed students are expected in a class with 30 students

10% of 30 = 3.**0** = 3
Move the decimal place 1 space to the left
3 left handed students are expected in a typical class of 30 students

Example 9 A $90 dress is discounted 40%, find the discount and the sale price.

Discount = 40%($90) = 4(10%)($90) = 4($9) = $36

Sale Price = Regular Price – Discount = $90 – $36 = $54

The sale price can be calculated without first finding the discount by finding the difference of 100% and 400% and finding sale price by taking 60% of regular price as shown below.

Sale Price = 60%($90) = 6(10%)($90) = 6($9) = $54

Example 10 In a restaurant a customer tips 10% if the service is okay, 15% if the service is good, and 20% if the service is excellent. Mentally estimate the tips if the bill is $64.20

Start by rounding the $64.20 bill to the nearest dollar which is $64.

Okay Service

10%($64) = $6.**4** = $6.40

Move the decimal place 1 space to the left

Excellent Service

20%($64) = 2(10%)($64) = 2($6.40) = $12.80

Move the decimal place 2 spaces to the left and then double

Good Service

15%($64) = 10%($64) + 5%($64) = $6.40 + $3.20 = $9.60

15% is equal to 10% + 5% and 5% is half of 10%, to determine 15% first find 10% by moving decimal place 1 space to the left and then calculate half of 10% and finally add the two values.

1.2.3 Class Activity

1) Convert 3% to a fraction and a decimal

2) Convert $\frac{3}{10}$ to a percent and a decimal

3) Convert 0.6 to a percent and a fraction

4) A pet rescue center has the following dogs available for adoption. Fill in the total categories and use this table to answer the following first as a fraction then as a percent rounded when necessary to 1 decimal place.

Age/Size	Small	Medium	Large	**Total**
Puppies	5	2	1	
Adult	3	4	9	
Senior	6	3	8	
Total				

4A. Find the percentage of the total dogs that are small puppies.

4B. Find the percentage of the puppies that are small puppies.

4C. Find the percentage of the total dogs that are either medium or large size.

4D. Find the percentage of large dogs that are not seniors.

4E. Find the percentage of large dogs that are puppies.

Homework 1.2.3

1-4. Convert the following percent into an equivalent decimal and fraction.

1. 32% 2. 85% 3. 4.5% 4. 235%

5-8. Convert the following fraction into a percent.

5. $\frac{13}{20}$ 6. $\frac{7}{16}$ 7. $\frac{5}{12}$ 8. $\frac{123}{90}$

9-12. Convert the following decimal into a percent

9. 0.92 10. 0.234 11. 0.032 12. 5.8

13-20. Fill in the following tables.

13.

Percent	Decimal	Fraction
35%		

14.

Percent	Decimal	Fraction
	0.32	

15.

Percent	Decimal	Fraction
		$\mathbf{\frac{11}{40}}$

16.

Percent	Decimal	Fraction
140%		

17.

Percent	Decimal	Fraction
	0.05	

18.

Percent	Decimal	Fraction
		$\mathbf{\frac{4}{5}}$

19.

Percent	Decimal	Fraction
8.5%		

20.

Percent	Decimal	Fraction
	0.004	

21-32. Write the following fractions in percent format without a calculator.

21. 3/4 22. 2/3 23. 7/10 24. 3/20
25. 1/2 26. 1/4 27. 2/5 28. 7/20
29. 9/10 30. 11/20 31. 4/5 32. 3/10

33-38. Evaluate the following. When necessary round to one decimal place.

33. 35% of 120 34. 8% of 125 35. 3.5% of 280
36. 1.20% of 90 37. 125% of 48 38. 16.7% of 36

39-50. Evaluate the following without a calculator.

39.	10% of 140	40.	1% of 140	41.	50% of 140
42.	100% of 24	43.	200% of 24	44.	20% of 24
45.	2% of 35	46.	40% of 35	47.	100% of 35
48.	25% of 12	49.	30% of 12	50.	4% of 12

51-66. Solve the following application problems. Show the calculations. Solve the bolded numbered problems without using a calculator.

51. A vitamin tablet contains 25% of the daily recommended value of zinc which is 16 mg. How many milligrams of zinc are in one tablet of this vitamin?

52. Approximately 60% of the 8000 total shoppers surveyed indicated that they preferred receiving a gift card as a present. Find the number of shoppers surveyed that preferred receiving a gift card. Also find how shoppers did not prefer a gift card as a present.

53. As a class project 60 students were surveyed and asked if they preferred online to regular classes. If 35% preferred online course, how many of the students surveyed responded that they preferred online courses. Also find how many students did not prefer online classes.

54. 71% of the total earth's surface area of 510 million square kilometers is covered by water. How many million square kilometers of the earth are covered by water? How many million square kilometers of the earth are covered by land? *Round the final answers to the nearest million*

55. Assuming the current world population is 7.2 billion people. If approximately 17% of the world's population lives in India, estimate how many people live in India and how many people live in other countries in the world. *Round the final answers to one decimal place*

56. Approximately 65% of the total students passed a course. If there are 34 students in the class, find how many students passed the course. Also find how many students did not pass the course.

57. A basketball player scores approximately 22% of his team's total points. If the team averages 87 points per game during the season, find how many points this player averages per game during the season.

58. A dishwasher is on sale at 10% off the regular price. If the regular price is $420, find the discount and the sale price of the dishwasher.

59. A bed is on sale at 30% off the regular price. If the regular price is $400, find the discount and the sale price of the bed.

60. At a clearance sale, clothes are marked down 40%. Find the discount and the sale price of a dress with a $28 regular price.

61. A grocer markups up cereal by 35%. If the wholesale cost of a box of cereal is $3.00, find the markup and the price of a box of this this cereal.

62. A jeweler gets a ring wholesale for $400 and marks it up by 150%. Find the markup and list price of this ring.

63. Lilly earns a base salary of $2500 per month plus 2% commission on her total monthly sales. If her total monthly sales are $40,000 in June, find her totally monthly earning in June.

64. Kelli earns a base salary of $500 per week plus 1% commission on her total sales. If Kelli's total sales weekly sales are $18,500 find her weekly earning that week.

65. A laptop computer is advertised for $630. If the sales tax rate is 8.5%, find the sales tax amount and the total cost of the laptop including sales tax.

66. At a fast food restaurant for lunch Eduardo orders some menu items for $7.35. If the sales tax rate is 6% find the sales tax amount and the total cost for the lunch including the sales tax.

67. A pet rescue center has the following dogs available for adoption. Fill in the total categories and use this table to answer the following first as a fraction then as a percent rounded when necessary to 1 decimal place.

Age/Size	Small	Medium	Large	**Total**
Puppies	3	2	3	
Adult	1	5	8	
Senior	2	3	7	
Total				

67A. Find the percentage of the total dogs that are small puppies.
67B. Find the percentage of the puppies that are small puppies.
67C. Find the percentage of the total dogs that are medium sized dogs.
67D. Find the percentage of the total dogs that are either medium or large size.
67E. Find the percentage of large dogs that are not seniors.
67G. Find the percentage of large dogs that are puppies.

68. Below are the results of a survey of beverage preferences for teens and adults. Fill in the totals, and use the table below to answer the following first as a fraction then as a percent rounded when necessary to 1 decimal place.

Age/Drink	Soft drinks	Juice	Coffee	**Total**
Teens	7	4	3	
Adults	4	5	7	
Total				

68A. Find the percentage of the total surveyed that prefer soft drinks.
68B. Find the percentage of the total surveyed that are adults that prefer coffee
68C. Find the percentage of teens that prefer juice.
68D. Find the percentage of adults that prefer coffee.
68E. Find the percentage of coffee drinkers that are adults.
68F. Find the percentage of juice drinkers that are not adults.

69. In fiscal year 2015 the U.S. federal government's income was 2,407 billion dollars. This income was generated from the following sources: $770 billion from social security, Medicare, and unemployment taxes, $939 billion from personal income taxes, $312 billion from corporate income taxes, $168 billion from excise, customs, gift, and miscellaneous taxes, and $218 billion from borrowing to cover debt.

69A. What percent of the total 2015 federal income is the personal income tax?
69B. What percent of the total 2015 federal income is the corporate income tax?

70. In fiscal year 2015 the U.S. federal government expenditures were 2,655 billion dollars. These funds were spent in the following categories: 36% for social security, Medicare, and other retirement program, 23% for national defense, veterans, and foreign aid, 8% for interest on the national debt, 12% for physical, human, and community development, 19% for social programs, and 2% for law enforcement and general government.

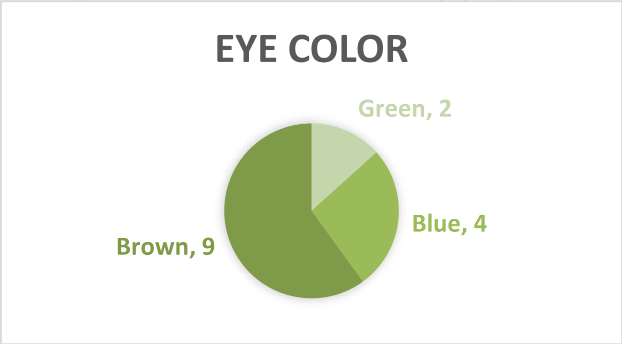
70A. What was the total federal spending on social programs in 2015?
70B. What was the total federal spending for interest on the national debt in 2015?

1.2.4 Pie Graphs

A pie graph is a graph for qualitative data in the shape of a circle, which is divided into slices. Each slice represents a part of the whole. The entire circle represents the whole survey. Each different population would require its own circle graph. The size of the slice is proportional to the percentage of the data in that slice (a larger slice has a larger frequency then a smaller slice).

Understanding Pie Graphs:

Example 1: Consider the following graph



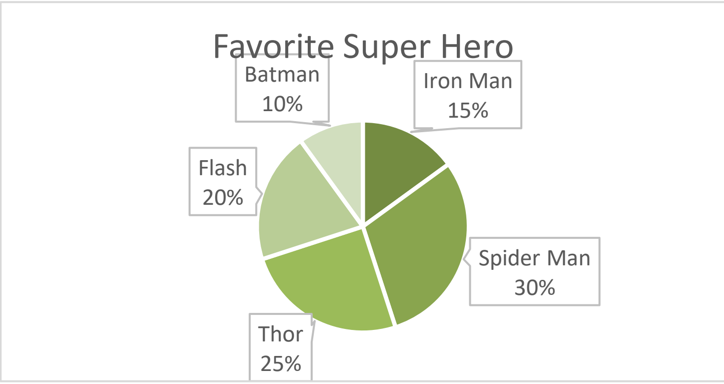
Pie graphs can display the data by telling the frequency for each slice or the percentage of the whole represented by that slice. In this case the graph is displaying the frequency.

We can see that 9 responders have Brown eyes. If we want to calculate the percent who have brown eyes we can first express it as the fraction $\frac{Number\ with\ brown\ eyes}{Total\ number\ of\ people}$

To calculate the total number of people we need to sum the individual frequencies (since the problem did not tell us). Here we see that there are 9+2+4=15 total people.

Then the fraction who have brown eyes is $\frac{9}{15}$ which simplifies to $\frac{3}{5}$. Dividing 3/5 we get the decimal equivalent of 0.6. Converting 0.6 to a percent we get that 60% of the people in the survey have brown eyes.

Example 2: Considering the following graph. This graph represents a sample of 200 students who were asked who their favorite superhero is.



This pie chart displays the data in percentages. If we wish to know the frequency for a category we can find it by taking the percent for that category of the total (which in this case we have been told is 200 people).

For example, 30% responded Spider Man. 30% of the total (200 people) would be 30% times 200 = 0.3(200) = 60 people.

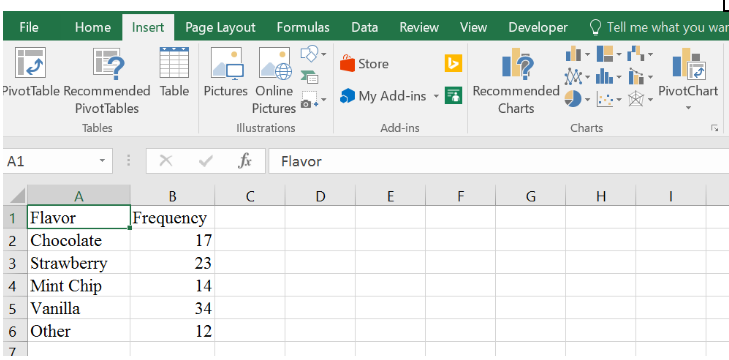
Making a Pie Chart in EXCEL

Step 1: Write the data in a frequency distribution (may use percent in the category instead of frequency)
Step 2: Type the frequency distribution in excel
Step 3: Highlight the entire frequency distribution.
Step 4: Select "insert" and then choose the graph that looks like a pie chart, select the first option.
Step 5: You can customize the graph by clicking on the + on the right side of the graph and choosing the options that best display the data (including choosing to display the information as frequency or percent in each slice).

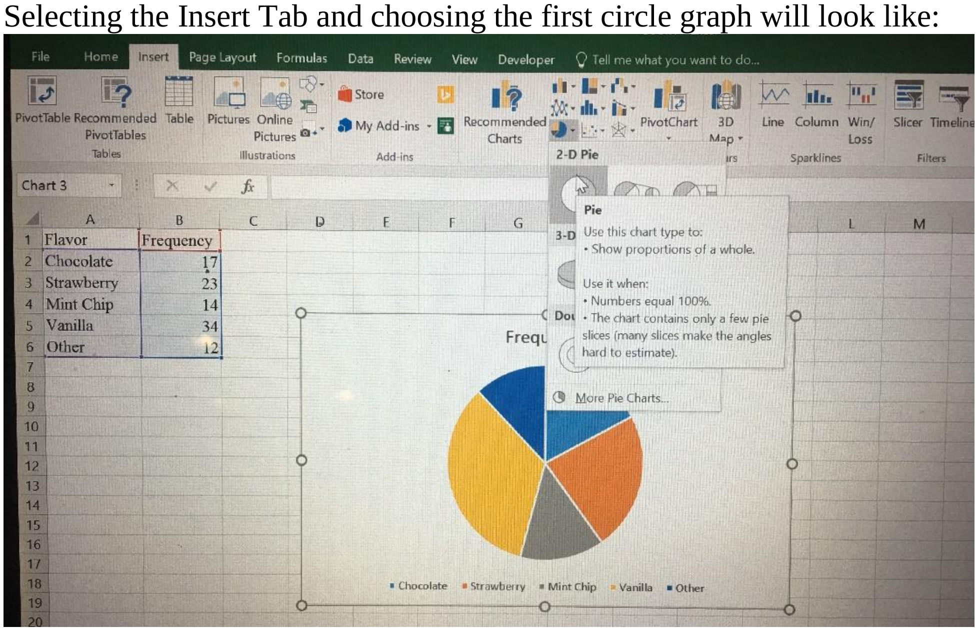
Example:
Consider the frequency distribution:

Flavor	Frequency
Chocolate	17
Strawberry	23
Mint Chip	14
Vanilla	34
Other	12

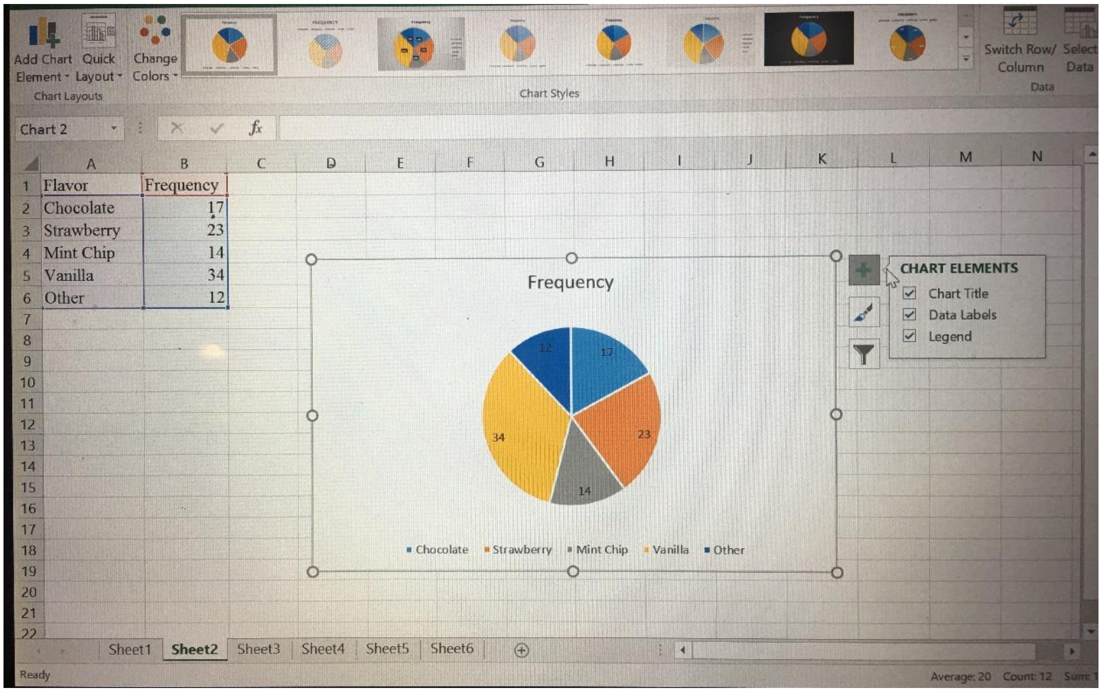
When typed into EXCEL it will look like:



Selecting the Insert Tab and choosing the first circle graph will look like:



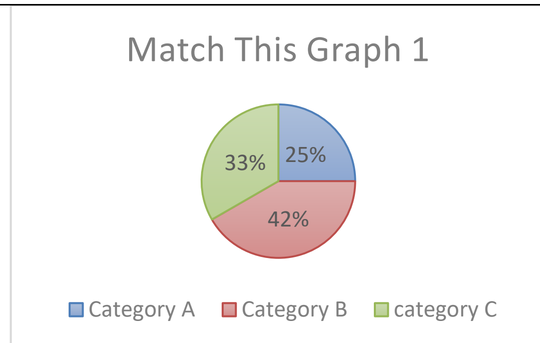
Then clicking on the + on the right side of the graph we can customize the chart elements to best display the data.



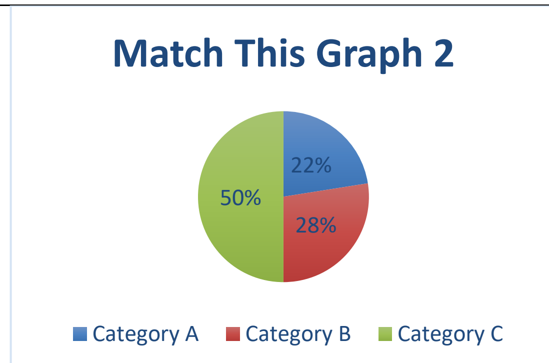
1.2.4 Class Activities

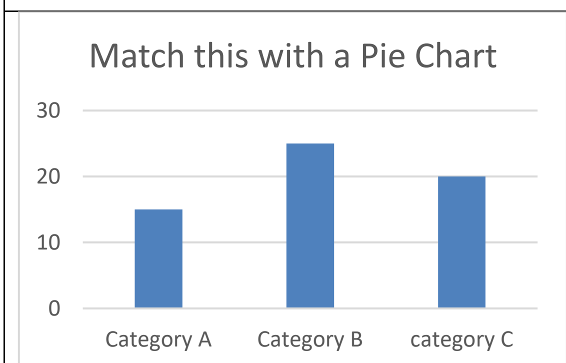
1) Matching: Match the frequency distributions and bar graphs in column 1 with the pie graphs in column 2.

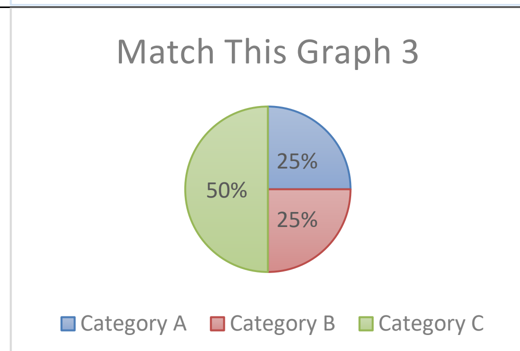
Category	Frequency
A	18
B	22
C	40



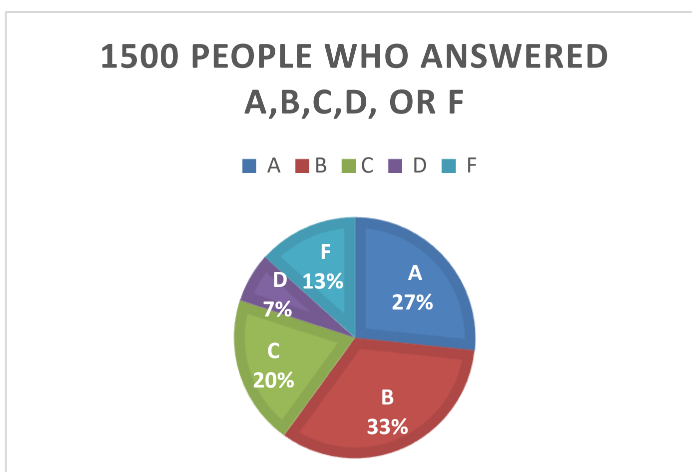
Data:
A,A,A,A,A,A,A,A,B,B,B,B,B,B,B,B,C,C,
C,C,C,C,C,C,C,C,C,C,C,C,C,C







2) If the following circle graph represents 1500 people, find how many are in each category (if needed round to the nearest person.



3) CLASS DATA: Question: What is your eye color. Record results below:

EYE COLOR	QUANTITY	Percent (from graph)
Brown		
Blue		
Green		
Hazel		
Other		

Let’s make a pie chart for this data in EXCEL/Google Sheets:

Step 1: Type the frequency distribution into the spreadsheet
Step 2: Insert the graph

Use the graph to find the percent column for each category.

4) If you were to assume the people in the college have the same proportion of eye color as we see in this class, and assume there are 12,000 students enrolled at Solano. How many students would you expect to have each color eye?

Homework 1.2.4

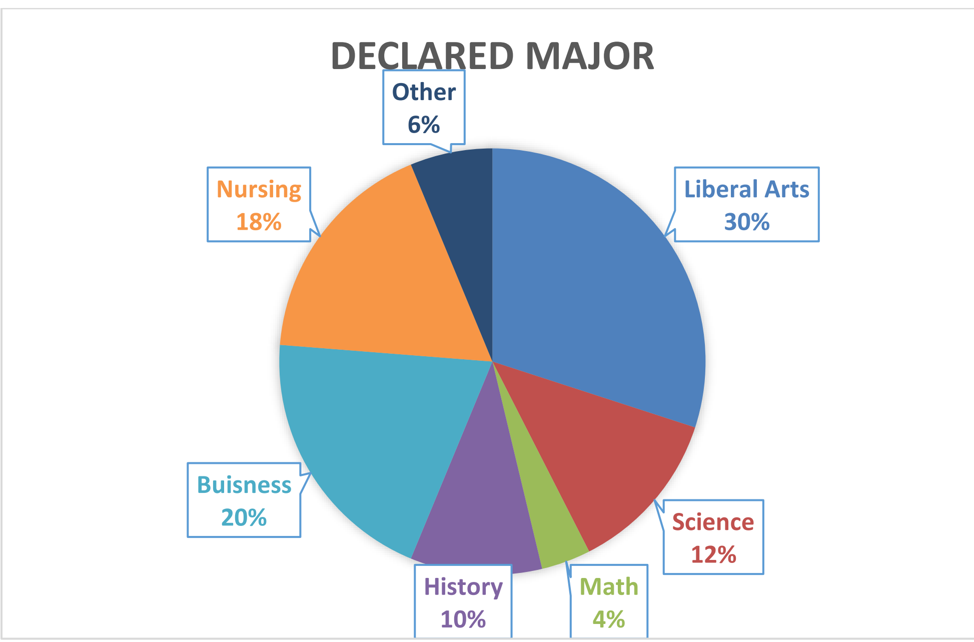
1) Consider the data:
   Data: Question: Grade on last quiz for a class of history students
   A,A,A,A,A,A,A,A,A,A,A,B,B,B,B,B,B,B,B,B,B,B,B,B,B,B,B,C,C,C,C,C,C,C,C,C,C,C,C,C,C,C,C,D,D,D,D,D,D,F,F,F,F,F,F

   A) For the given data, make a frequency distribution, bar graph, and pie chart. You may use technology to create the graphs. Print or draw them to turn them in.

   B) What fraction of the students earned an A?
   C) What percent of the students earned an A?
   D) What percent of the students passed the quiz?
   E) What percent of the students failed the quiz?
   F) How many students passed?
   G) Of the students who passed, what fraction of them passed with an A?
   H) Explain how question 1B and 1G are different questions.

2) Make bar graphs to represent the following pie-chart:

2,000 students are represented on the following pie chart (Fictitious Data).



Tip: The first step will be to figure out how many students are represented by each category.

3) For the graph in question 2, assuming the proportion in each major stays the same, how many students would need to be enrolled to have 500 math majors?

Section 1.2.5 Percent Applications

There are various techniques to solve application problems involving percentages including using formulas or setting up proportions. In this section, applied percent problems are solved using the percent formula (*percent*)(*total*) = (*amount*) abbreviated as (P)(T) = A. In solving these problems two of the three variables are given and the percent formula is then evaluated to find the value of the remaining variable. It is essential that the percent, total, and amount are identified correctly. The total is often located in a sentence after the word of. For instance, if a basketball player successfully makes 35 out of 50 attempted foul shots, the 50 attempted foul shots represent the total.

Example 1 Find 25% of 80?

P = 25%	.25(80) = A
T = 80	20 = A
A =	Check that 25% of 80 is 20

Example 2 21 is 30% of what number?

P = 30%	(.30)T = 21
T =	T = 21/0.30 = 70
A = 21	Check that 30% of 70 is 21

Example 3 13 out of 50 is what percent?

P =	P(50) = 13
T = 50	P = 13/50 = 26%
A = 13	Check that 26% of 50 is 13 (*calculator*)

Example 4 A tablet contains 500 IU (international units) of Vitamin D which is 125% of the daily recommended value. Find the daily recommended value of Vitamin D?

P = 125%	1.25(T) = 500
T =	T = 500/1.25 = 400
A = 500 IU	Check that 125% of 400 is 500 (*calculator*)

The daily recommended dosage of Vitamin D is currently 400 grams.

Use the following nutritional information to answer the following question, 1 gram of fat contains 9 Calories, each gram of carbohydrate contains 4 Calories, and each gram of protein contains 4 Calories.

Example 5 A can of lentil soup contains two servings and each serving contains 3 grams of total fat, 19 grams of total carbohydrates including 9 grams of fiber, and 8 grams of protein. Find the calories per serving and the calories in this can of lentil soup and the percentage of daily Calories based on a 2000 Calorie daily diet.

First calculate the calories per serving and the calories per can of this lentil soup.

Calories per serving 3(9) + 19(4) + 8(4) = 135 Calories

Calories per can of soup 2(135) = 270 Calories

P =
T = 2000 Cal
A = (2)(135 Cal) = 270 Cal

P(2000) = 270
P = 270/2000 = 13.5%

This can of soup contains 13.5% of daily calories based on a 2000 Calorie diet.

Example 6 18K indicates that 75% of an item is made from gold. An 18K bracelet contains 8.4 grams of gold, what is the total weight of the bracelet?

P = 75%
T =
A = 8.4 *g*

.75(T) = 8.4
T = 8.4/.75 = 11.2 *g*

The bracelet weighs 11.2 grams of which 8.4 grams are gold.

Example 7 A college softball team played 32 games of which they lost 14 games. Find the percentage of games won by this team.

P =
T = 32
A = 32 – 14 = 18

P(32) = 18
P = 18/32 ≈ 56.3%

The softball team won 18 out of 32 games which is a winning percentage of 56.3%

In some percent applications, the base or original quantity represents the total and the amount of increase or decrease represents the amount. For these applications the following formula is used.

(*percent*)(*total, original, or base*) = (*amount of increase or decrease*)

Example 8 The value of a house increased from \$300,000 to \$360,000. Find the percentage increase in the value of the house.

\$300,000 represents the original value of the house, the percent of increase in the value of the house is the unknown, and the amount of increase in the home's value is \$60,000. The value of this house has increased by 20%.

P (incr) =
T (orig) = \$300,000
A (incr) = \$360,000 – \$300,000 = \$60,000

$$P(300{,}000) = 60{,}000$$
$$P = 60{,}000/300{,}000 = 20\%\uparrow$$

Note the above could be found without a calculator by simply reducing 60,000/300,000 which equals 6/30 = 1/5 = 20% (double dime)

Example 9 The value of a house decreased from \$300,000 to \$260,000. Find the percentage decrease in the value of the house.

\$300,000 represents the original value of the house, the percent of decrease in the value of the house is the unknown, and the amount of decrease in the home's value is \$40,000. The value of this house had decreased by 13.3%.

P (decr) =
T (orig) = \$300,000
A (decr) = \$300,000 – \$260,000 = \$40,000

$$P(300{,}000) = 40{,}000$$
$$P = 40{,}000/300{,}000 \approx 13.3\%\downarrow$$

Class Activity 1.2.5____________________________

1. Nicole earns $18 per hour. If she worked 40 hours and received a paycheck for $618 which reflected her wages minus payroll taxes, find how much in payroll taxes did Nicole pay and what percentage of her total earnings are payroll taxes.

2. A can of soup contains 2.5 serving and each serving contains 24 grams of carbohydrates. If the daily recommended intake of carbohydrates is 300 grams, find the daily percentage of carbohydrates in a can of this soup?

3. A computer system valued at $1800 three years ago is now worth $1000. Find the percentage decrease in the value of the computer system?

4. A boiled egg contains approximately 80 calories. Find the percentage of calories contained in a three boiled egg breakfast for a person on a 2200 calorie daily diet.

5. A new edition of a textbook is now available. The old edition costs $95 and the new edition costs $115. Find the percentage increase in the price of the textbook.

6. A 2 ounce salmon filet contains 3.5 grams of fat, given that that the recommended daily fat intake is 65 grams of fat what percentage of daily recommended fat is in an 6 ounce portion of salmon.

Homework 1.2.5

1-27. Use the percent formula (P)(T) = A to solve the following. **Show steps.**

1.	50 is 20% of what number	2.	Find 37% of 210
3.	35 out of 45 is what percent	4.	45% of what number is 90
5.	What percent of 47 is 11	6.	Find 75% of 12
7.	18 is what percent of 12	8.	Find 120% of 800
9.	80 is 2% of what number	10.	170 out of 215 is what percent

11. On a test, Maria answered seventeen of the twenty questions correctly. What percent of the question did Maria answer correctly? What percent of the questions did Maria answer incorrectly?

12. A can of soup contains 2 serving and each serving contains 32 grams of carbohydrates. If the daily recommended intake of carbohydrates is 300 grams, find the daily percentage of carbohydrates in a can of this soup?

13. A multivitamin contains 250 milligrams of vitamin C in each tablet. If the daily recommended dosage of vitamin C is 600 milligrams, what percentage of the recommended daily dosage is contained in 3 vitamins?

14. 80% of all employees at a local company are fulltime workers. If the company has 60 fulltime workers, how many total employees does the company have?

15. A boiled egg contains approximately 80 calories. Find the percentage of calories contained in a two boiled egg breakfast for a person on a 2500 calorie daily diet.

16. Nicole earns $15 per hour. If she worked 40 hours and received a paycheck for $518 which reflected her wages minus payroll taxes, find how much in payroll taxes did Nicole pay and what percentage of her total earnings are payroll taxes.

17. A food truck starts the lunch shift with 200 premade burgers. If at the end of the shift there are still 28 burgers left, what percent of the premade burgers are sold during the lunch shift.

18. A patient's cholesterol level last year was 250 and this year it is 182. Find the percentage decrease in this patient's cholesterol level.

19. A new edition of a textbook is now available. The old edition costs $90 and the new edition costs $103. Find the percentage increase in the price of the textbook.

20. From 2014 to 2015 the average retail price of gasoline in the U.S. decreased from $3.36 to $2.44 per gallon. Find the percentage decrease in the price of a gallon of gas during this period?

21. The list price of a new compact car is $14,000 and the car dealer is offering a cash rebate of $1500. Find the percentage decrease in the price of the car?

22. A web site increased the number of annual visitors from 2.5 million to 3.2 million per year. Find the percentage increase in web site visitors.

23. A computer system valued at $1500 three years ago is now worth $600. Find the percentage decrease in the value of the computer system?

24. The number of miles driven by Americans was decreasing for the past decade until in recent years the price of gasoline fell. In 2015 Americans drove 3.06 trillion miles in 2015 which increased to 3.17 trillion driven in 2016. Find the percentage increase in miles driven between 2015 and 2016.

25. In January 2017 the price of a first class stamp increased from 47 to 49 cents. Find the percentage increase in the price of a stamp.

26. A 2 ounce salmon filet contains 3.5 grams of fat, given that that the recommended daily fat intake is 65 grams of fat what percentage of daily recommended fat is in an 8 ounce portion of salmon.

27. Brett quiz scores are 45, 33, 39, 42, 43 and his test scores are 92, 79, and 81. Find Brett's percentage average in the course if each quiz is worth 50 points and each test 100 points and the lowest quiz grade is dropped.

1.4 Graphs for Qunatitative Data

Learning Goal: For the distribution of a quantitative variable, describe the overall pattern (shape, center, and spread) and striking deviations from the pattern.

Specific Learning Objectives:

- Understand Frequency Distributions, Dot plots, and histograms as a way of analyzing Quantitative Data.
- Develop a way to describe and distinguish graphs of a quantitative variable.
- Identify reasonable explanations for what might explain the differences seen in different data sets.

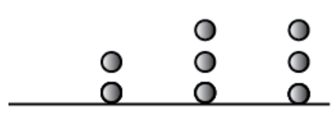
1.4.1 DOT PLOTS:

We would like to be able to create diagrams and graphs that show the **distribution** or shape of the data. An easy way we can show the distribution of the data is by creating **dot plots**. The number of dots above each value corresponds to the frequency.

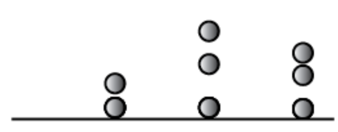
To create a dot plot:

1. Draw a number line so that the full range of data can be plotted on the number line (you do not necessarily need to start at 0). You do need to make sure you can represent each piece of data. And your values should be equally spaced.
2. For each data value, put a dot over the corresponding value on the number line.
3. When you have duplicate values, put a dot above the previous dot so that that dots are "stacked".
4. Make sure that the vertical spaces between the dots are equal. Thus, all the dots will be in nice rows and columns.

Like this:



Not like this:



You want to create a GOOD visual for the data

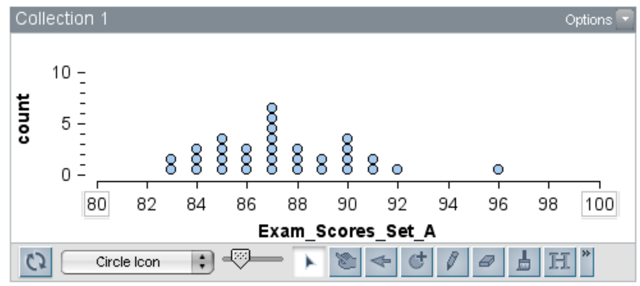
1.4.1 CLASS ACTIVITY:

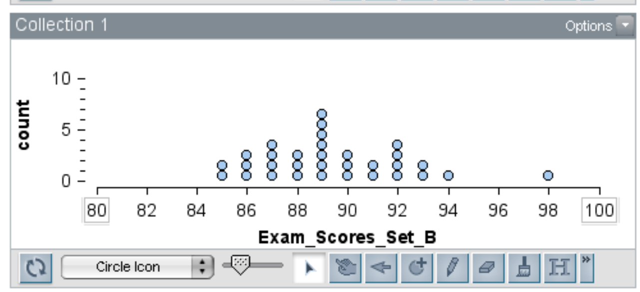
Introduction: Statisticians collect data in order to study and analyze groups. They focus on the group's data in aggregate, instead of focusing on individuals. For a statistician, creating and comparing graphs is often the first step in analyzing data.

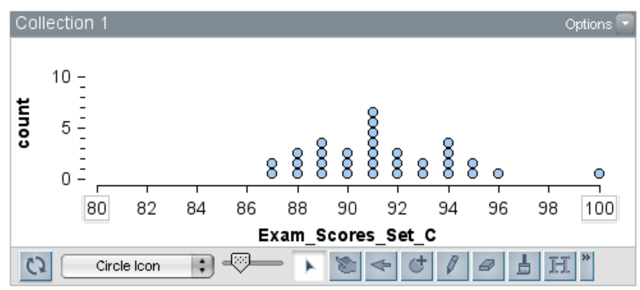
In this activity you will compare and contrast the graphs of hypothetical sets of exam scores. In other words, these graphs show made-up data. These hypothetical data sets have been constructed to help you begin to "see" like a statistician. By comparing these data sets, you will begin to develop an informal understanding of the key features of a graph that statisticians use to describe data.

During this activity do your best to describe what you see. Jot down notes to capture your thinking as you go. You can use your notes during our class discussion of this activity. This activity does not require you to remember anything or to apply previous knowledge.

1) Compare and contrast the three graphs shown at the right.







a) How are the graphs similar? How are they different? What is the most distinctive feature that distinguishes these three graphs from each other?

b) For each graph, pick a single exam score to summarize the overall performance of the students. In other words, summarize each set of data with one number.

Set A ___________
Set B ___________
Set C ___________

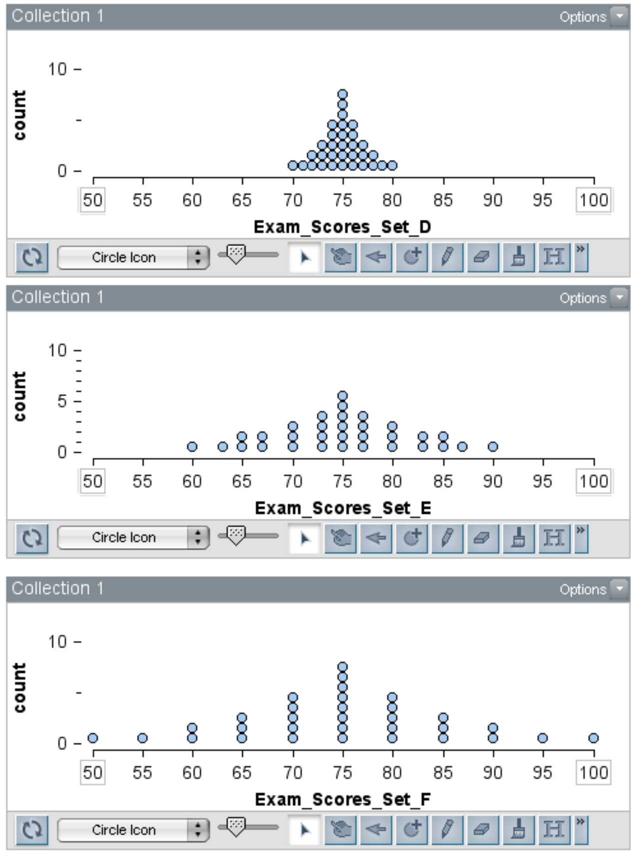
c) The average score for Set A is 87.4. What do you think the average score is for Set B and for Set C? (See if you can answer this without doing any calculations.)

d) Which, if any, of the statements below is a reasonable explanation for the differences in the graphs? Why?

- The graphs represent different classes. Different groups of students will obviously perform differently on an exam.
- The graphs represent a single class after the teacher adjusted the grades. The teacher realized that some of the exam questions were not well written, so she adjusted the grades by adding points to the original exam scores.
- There are no differences in the graphs. They look the same.

2) Compare and contrast the three graphs shown at the right.

a) How are the graphs similar? How are they different? What is the most distinctive feature that distinguishes these three graphs from each other?



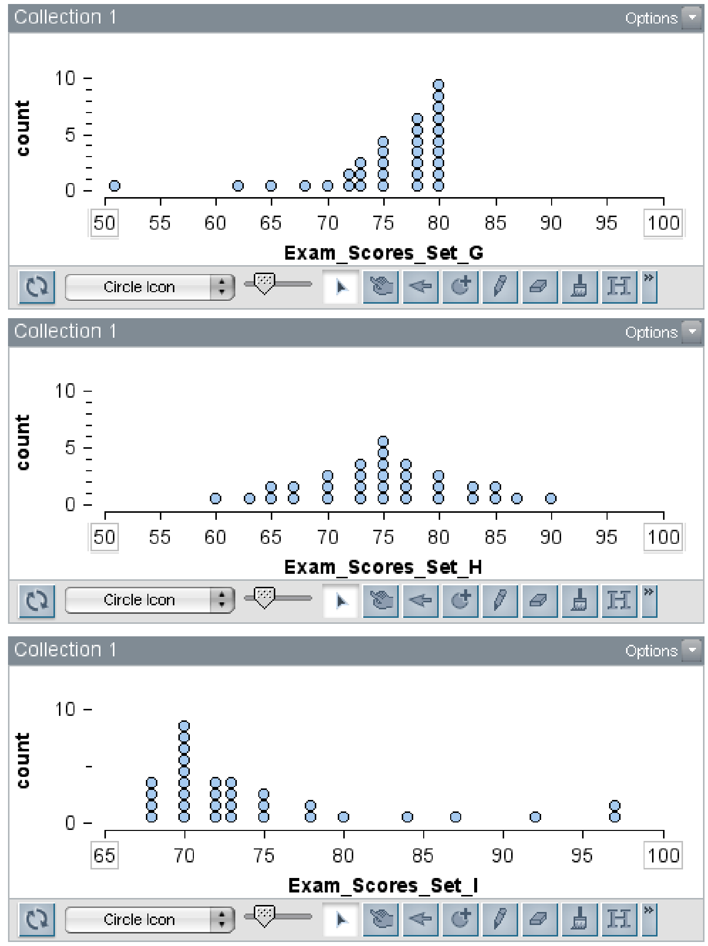
b) The average exam score is the same for each set of data (D, E and F). What do you think the average is?

c) Which of the statements below, if any, is a reasonable explanation for the differences in the graphs? Why?

- The graphs represent a single class after the teacher adjusted the grades. The teacher realized that some of the exam questions were not well written, so she adjusted the grades by adding points to the original exam scores.
- The graphs represent three different classes with teachers that have different grading standards. One is an easy grader. One is a really hard grader.
- The differences could be explained by whether the teachers allowed the students to work together on the exam and how much time the students were given to finish the exam.

3) Compare and contrast the 3 graphs shown at the right.

a) In each of these 3 graphs, the average score is the same. The average score is 75. In what other ways are the graphs similar? How are they different? What is the most distinctive feature that distinguishes these three graphs from each other?



b) What might explain the differences in these graphs?

1.4.2 Describing the shape of a distribution:

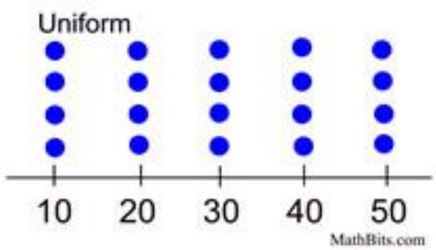
Learning Goal: For the distribution of a quantitative variable, describe the overall pattern (shape, center, and spread) and striking deviations from the pattern.

Specific Learning Objectives:

- Distinguish between categorical and quantitative variables;
- Identify graphs that represent the distribution of a quantitative variable;
- Analyze the distribution of a quantitative variable using a dot plot. Describe the shape, give a general estimate of center, and determine the overall range.

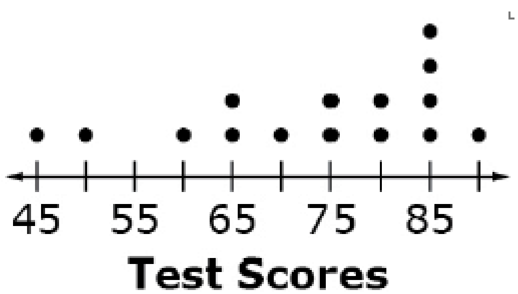
Definitions:
A uniform distribution is one where the data is evenly spread out. On a dot plot all the heights will be the same. Each value occurs with the same frequency.

Picture:



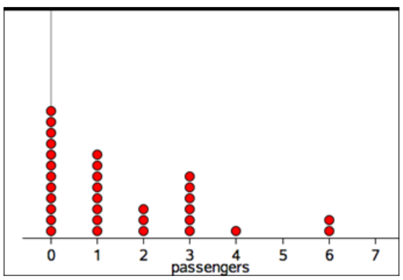
A graph is skewed left if the data looks pushed to the right side of the graph with a tail leading off to the left.

Picture:



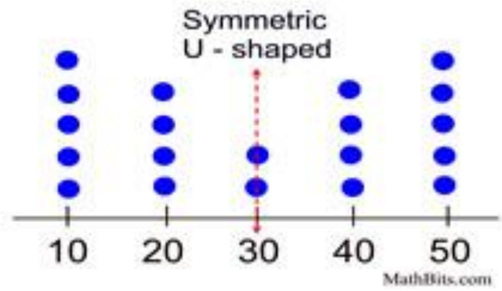
A graph is skewed right if the data looks pushed to the left side of the graph with a tail leading off to the right.

Picture:



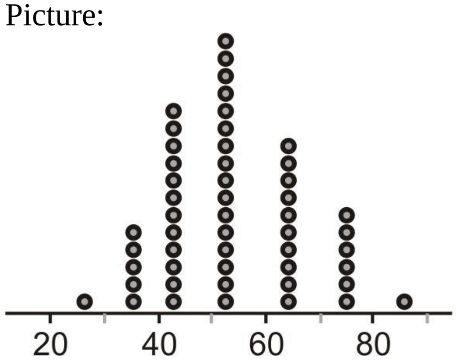
A graph is Symmetric if the right and left sides of the graph are mirror images of each other

Picture:



A graph is approximately normal if the graph appears to be symmetric, with the data in the center and tails out in both directions.

Picture:



VOCAB:

Distribution of a variable	A statistical distribution is an arrangement of the values of a variable (often a graph) showing their observed frequency of occurrence. Here is an other definition: "a representation that shows the possible values of a variable and how often the variable takes those values."
Describing the distribution of a quantitative variable	To describe a distribution, describe the *shape*, *center*, *spread* and *outliers*. Shape: describe the overall trends in the data. Typical ways to describe shape include symmetric, left or right skew, uniform. Center: give a single number that represents the data; a typical value or average. You may report the *mode*, a value or range of values that occur most often. We will discuss the *mean* and the *median* later. Spread: give a single number that measures how much the data varies. Range (maximum value minus minimum value) is one way to measure spread. Outliers: unusual data values

<u>1.4.2 CLASS ACTIVTY</u>:

In this activity you will practice analyzing the distributions of quantitative variables using descriptions of shape, center and spread. We will focus on dot plots. If you encounter vocabulary that you do not know, check out the vocabulary list on page ______.

1) Here is a partial spreadsheet of 2011-2012 data for a set of hatchback cars.

Car make and model	City miles per gallon	Drive	EPA size class	Engine
Acura ZDX	16 mpg	All wheel drive	Sport utility	6 cylinder
Audi TTS	14 mpg	All wheel drive	Midsize car	10 cylinder
Chevrolet Aveo	25 mpg	Front wheel drive	Compact car	4 cylinder

a) Who are the individuals described by this data?

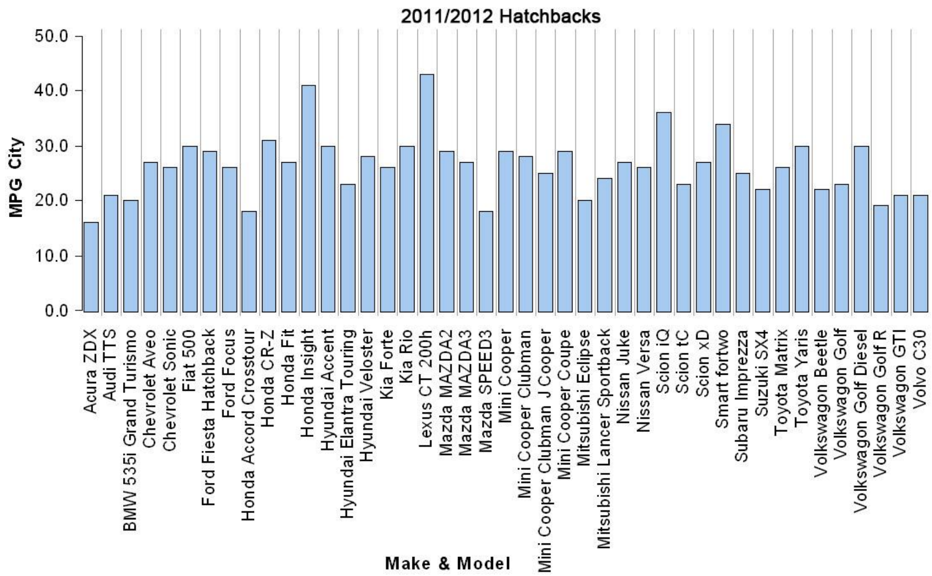
b) How many variables are shown in the spreadsheet?

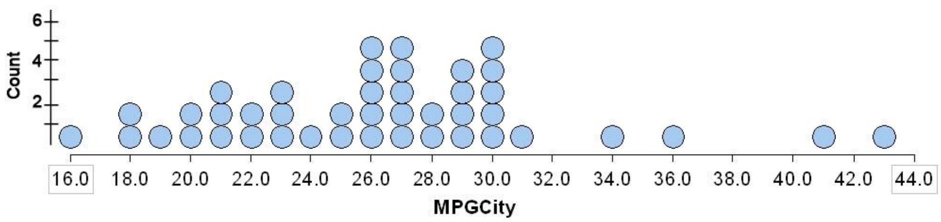
c) Which variables are categorical variables?

d) Which variables are quantitative variables?

e) Describe what the values in the spreadsheet tell us about the Chevrolet Aveo.

2) Below you will see a case-value graph and a dot plot graph of the 2011-2012 data for a set of hatchback cars. Use one or both graphs to answer the questions. Jot down notes or draw on the graphs to show how you determined your answers.





a) What is the best city mpg for this group of hatchbacks? ______________

What model of hatchback gets the highest city miles per gallon? ___________________

What is the worst city mpg for this group of hatchbacks? _____________

What model of hatchback gets the worst city mpg? ________________________________

For each of the above questions, indicate which graph (case-value or dot plot) was the easiest to use to answer the question.

b) How many hatchbacks get 25 mpg in the city? Which graph was the easiest to use to answer this question?

c) What are the mpg rates that occur most frequently for hatchbacks? Which graph was the easiest to use to answer this question?

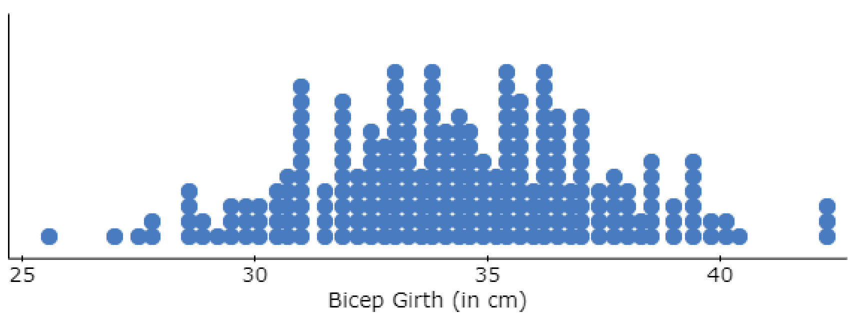
d) If you had to pick one mpg rate to represent this data, what would it be? Why did you choose that value?

3) The case-value graph and the dotplot are two different graphs of the same data. What do you see as the advantages and disadvantages of each type of graph?

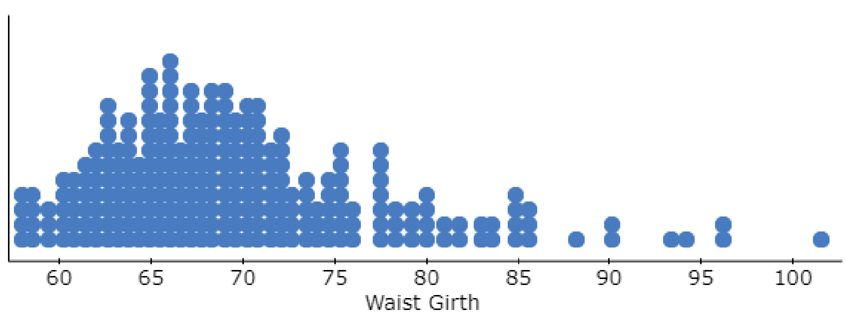
4) Statisticians make graphs to summarize data, so they prefer to use graphs that show the distribution of the data. A statistical distribution is defined as "an arrangement of the values of a variable showing their observed frequency of occurrence." Here is another definition of a statistical distribution: "a representation that shows the possible values of a variable and how often the variable takes those values." Which graph, the case-value graph or the dotplot, shows the distribution of the variable MPGCity?

5) For each of the following dotplots, draw a smooth curve outlining the distribution, and then describe the shape of the distribution using course vocabulary (See the definitions of symmetric, skewed left, skewed right.)

Bicep Girth for 247 men who were exercising several times a week

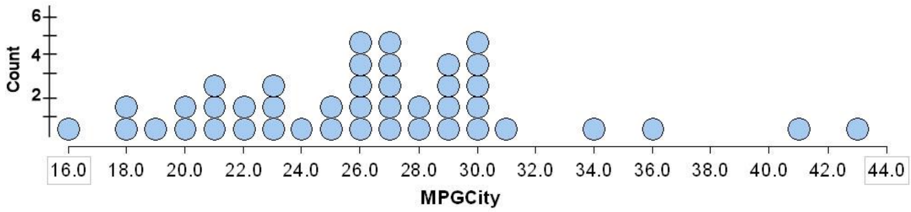


Waist girth for 260 women who were exercising several times a week



6) Suppose that almost everyone does well on the first exam with a few people (who did not study) performing poorly. What is the shape of the distribution of exam scores?

7) Thinking like a statistician: describing shape, center and spread.



a) *Thinking about shape:* Our shape descriptions (symmetric, skewed left, skewed right) don't fit this distribution. How would you describe this distribution's shape?

b) *Thinking about center:* Previously, you chose one value to represent the distribution of mpg ratings for hatchbacks. What was that value? ________

To represent the distribution, Ann decided to calculate the average mpg rating using the mean. She got 34 mpg. Do you think Ann made a mistake? How can you tell without doing any calculations?

c) *Thinking about spread:* Spread is a description of the variability we see in the data.

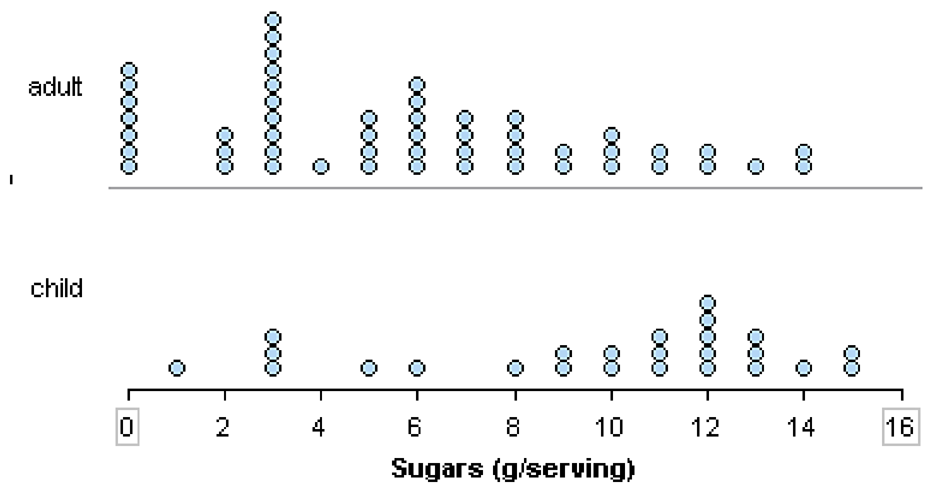
Here the city mpg for these hatchbacks varies from a low of ________ mpg to a high of ________ mpg.

What is the overall range for the mpg values (max – min)? ______________

Typical hatchbacks have a city mpg that varies from about ________ mpg to about __________mpg.

d) *Thinking about deviations from the pattern*: Are there any unusually larger or unusually small mpg rates? If so, what are the unusual values?

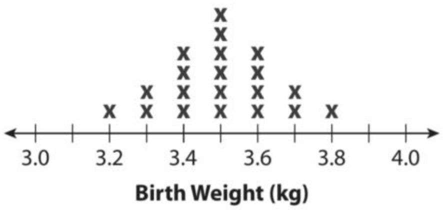
8) Here are dot plots of the sugar content (grams per serving) for some adult cereals and child cereals. Compare the two distributions by comparing shapes, estimates of center, and spreads.



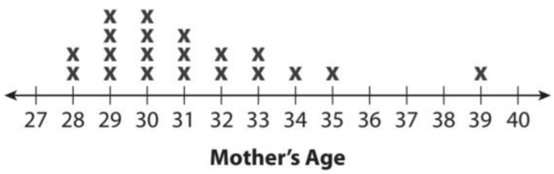
1.4.2 Homework

1. Below are three dot plots, one for the birth month for a sample of babies, one for the birthweight of the sample of babies, and one for the ages of their mothers. Answers the questions below.

a. Describe the shape of the distribution of birth months.



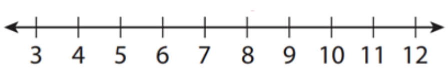
b. Describe the shape of the distribution of birth weights.



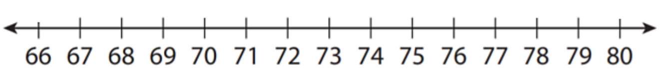
c. Describe the shape of the distribution of mothers' ages.

2. Make a dot plot of the data. Based on the shape of the distribution, describe the type of distribution it is.

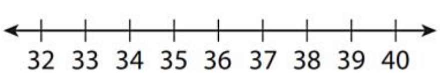
a. Ages of children: 4, 9, 12, 8, 7, 8, 7, 10, 8, 9, 6, 8



b. Scores on a test: 80, 78, 70, 77, 75, 77, 76, 66, 77, 76, 75, 77



c. Salaries (in thousands of dollars) of employees: 35, 35, 36, 40, 37, 36, 37, 35, 35, 38, 36, 34



1.4.3 Making Histograms

Learning Goal: For the distribution of a quantitative variable, describe the overall pattern (shape, center, and spread) and striking deviations from the pattern.

Specific Learning Objective:

- Distinguish between categorical and quantitative variables;
- Identify graphs that represent the distribution of a quantitative variable;
- Analyze the distribution of a quantitative variable using a histogram. Describe shape, give a general estimate of center and the overall range, and calculate relevant percentages.

A histogram is a graph for quantitative data that displays the data by dividing the range of values into equal sized sections (classes) and counting the frequency of the data values for each class. The data is then displayed by graphing bars (one per class) where the height represents the frequency for that class.
A histogram is similar to a bar graph, with a few key differences:

1) Histograms are for quantitative data, bar graphs are for qualitative data.
2) In histograms the bars touch, in bar graphs they do not.
3) In histograms the x axis is numbers (each class represents an interval on the real number line), whereas the categories are listed on the x-axis for bar graphs.
4) The order of the bars for histograms is fixed in numeric order, whereas for bar graphs the bars could be presented in any order.
5) It makes sense to discuss shape of a histogram (because the bars must be in the order they appear), but for bar graphs you cannot discuss the shape of the bar graph.

Creating the frequency distribution:
Step 1: Determine the number of bars desired for the data.
Step 2: Find the minimum and maximum values.
Step 3: Range = maximum – minimum
Step 4: Class width = Range/number of desired bars.
Step 5: Let the lower value for the first class be the minimum value. Then to find the lower value for each of the next classes add the class width found in step 4.
Step 6: Fill in the upper value foe each class by writing a number a little smaller then the next lower value. For the last class let the upper value be the maximum value.

Example of making a frequency distribution for quantitative data: Data:
1,1,1,1,2,2,3,4,5,6,6,6,6,6,6,6,6,7,7,7,7,7,7,7,8,8,8,8,8,8,8,8.5,9,9,9,9,9,9,9,9,9,10,10,10.2,10.5,11,11,11,11, 11.1, 12, 12.3, 13

Step 1: Let's choose to have 6 bars.
Step 2: Minimum = 1, Maximum = 13
Step 3: Range = 13-1 = 12
Step 4: Class width = 12/6 = 2
Step 5: Find the lower value for each class.

Class 1: Lower value = minimum = 1
Class 2: Lower value = previous + width = 1 + 2 = 3
Class 3: Lower value = 3+2 = 5
Class 4: Lower value = 5 + 2 = 7
Class 5: Lower value = 7 + 2 = 9
Class 6 (last class since 6 bars) = 9 + 2 = 11

Step 5: Since the lowest value for class 2 is 3, and data points are to the tenth we go one tenth smaller.
Class 1: Upper value = 3 – 0.1 = 2.9
Class 2: Upper value = 5 – 0.1 = 4.9 (notice you could also find this by adding the class width to the upper value from class 1).
Class 3: Upper value = 7 – 0.1 = 6.9
Class 4: Upper value =9 – 0.1 = 8.9
Class 5: Upper Value = 11-0.1= 10.9
Class 6 (Last class): Upper value = maximum = 13

So Class 1 contains all the numbers between 1 and 2.9
Class 2 contains all the numbers between 3 and 4.9
Class 3 contains all the numbers between 5 and 6.9
Class 4 contains all the numbers between 7 and 8.9
Class 5 contains all the numbers between 9 and 10.9
Class 6 contains all the numbers between 11 and 13.

Frequency distribution:
Data:
1,1,1,1,2,2,3,4,5,6,6,6,6,6,6,6,6,7,7,7,7,7,7,7,8,8,8,8,8,8,8,8.5,9,9,9,9,9,9,9,9,9,10,10,10.2,10.5,11,11,11,11, 11.1, 12, 12.3, 13

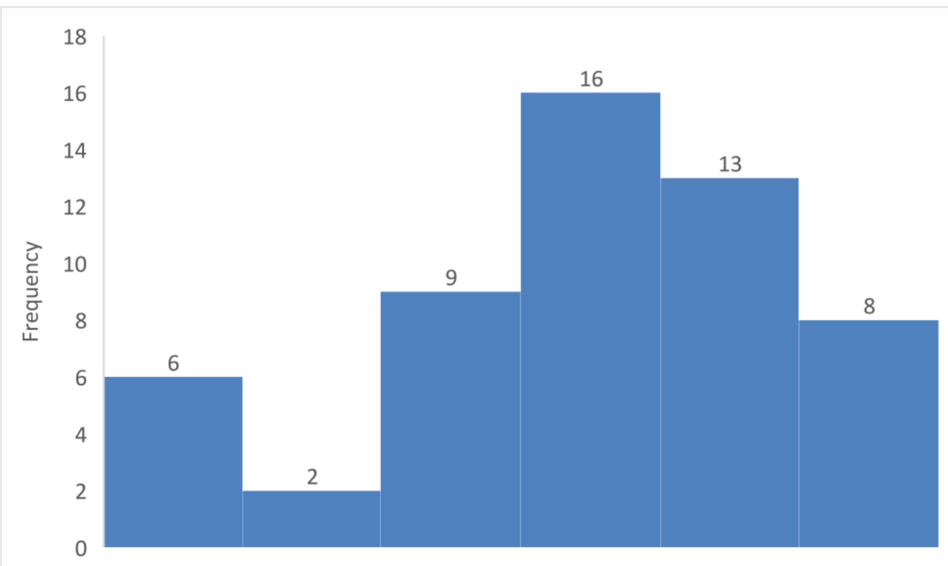
Class	Data in this class	Frequency
1-2.9	1,1,1,1,2,2	6
3-4.9	3,4	2
5-6.9	5,6,6,6,6,6,6,6,6	9
7-8.9	7,7,7,7,7,7,7,8,8,8,8,8,8,8,8.5	15
9-10.9	9,9,9,9,9,9,9,9,9,10,10,10.2,10.5	13
11-13	11,11,11,11,11.1, 12, 12.3, 13	8

Making a Histogram:

Step 1: Create a frequency distribution. In order to do this you will need to define classes (if the problem hasn't chosen them for you).
Step 2: Draw the grid. Decide on a good scale for the x and y axis.
Step 3: For each class, draw the bars at the height of the frequency. The bars should touch. If a class has a frequency of 0 then the bar at that height is 0.

Picture of the histogram for the data in this frequency distribution:

Class	Frequency
1-2.9	6
3-4.9	2
5-6.9	9
7-8.9	15
9-10.9	13
11-13	8



<u>Making a Histogram on the TI-83/84 Graphing Calculator</u>

Step 1: Enter the Data into the list.

Press: STAT, ENTER, then type in the numbers (press ENTER after each number).

Step 2: Turn STATPLOT On

Press 2^{nd}, y=, ENTER. Select the graph that looks like the histogram by scrolling to it and press ENTER. Make sure that the list you entered your data into is the one listed. To change the list, type 2^{nd}, then the list number.

Step 3: Set the WINDOW

Press WINDOW
Xmin=minimum value
Xmax = maximum value
Xscl = Class width
Ymin = 0
Ymax = (any number bigger then the largest frequency. Choose one that displays the heights well).
Yscl = 1 (make larger if needed to prevent thick line on the left).
Xres = 1

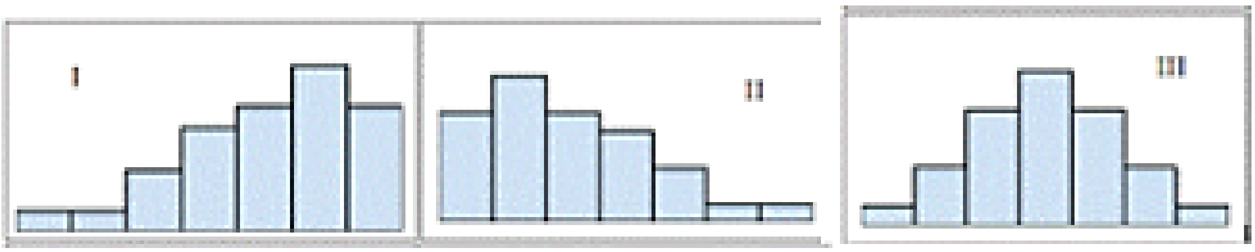
Step 4: GRAPH

Note: If the graphing calculator draws other lines/curves press y= and delete the formulas you see.

1.4.3 Classwork:

In this activity you will again practice analyzing the distributions of quantitative variables using descriptions of shape, center and spread. This is the same type of thinking you did in with dot plots, but this time the data will be summarized in a histogram.

1) Match the following descriptions to the histograms I-III.



A. Scores on an easy exam for a class of students

B. Scores on a hard exam for a class of students

C. Number of siblings for a large sample of U.S. adults

D. Exact volume of soda in a one-liter bottle for a case of 24 bottles

E. Dates on the pennies I have in my car ashtray

F. Weights for a large sample of newborn babies

2) For each research question, (1) identify the individuals of interest (the group or groups being studied), (2) identify the variable(s) (the characteristic that we would gather data about), and (3) determine whether each variable is categorical or quantitative.

   A. What is the average number of hours that community college students work each week?

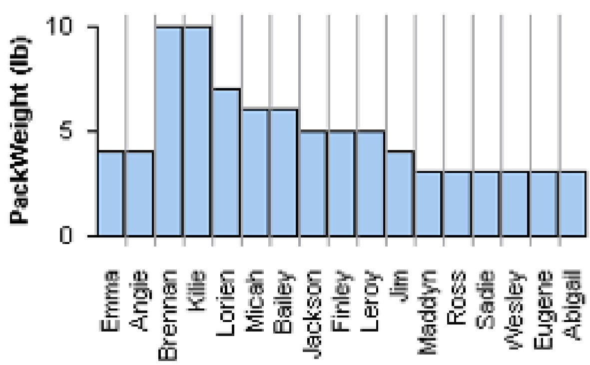
   B. What proportion of all U.S. college students are enrolled at a community college?

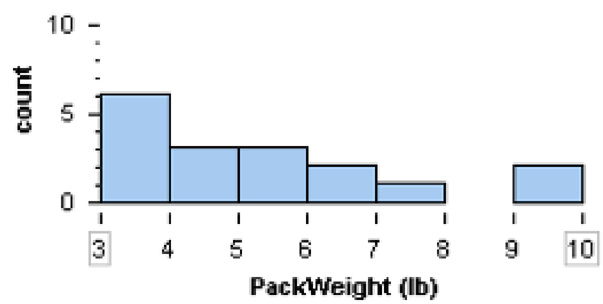
   C. In community colleges, do female students have a higher GPA than male students?

   D. Are college athletes more likely than non-athletes to receive academic advising?

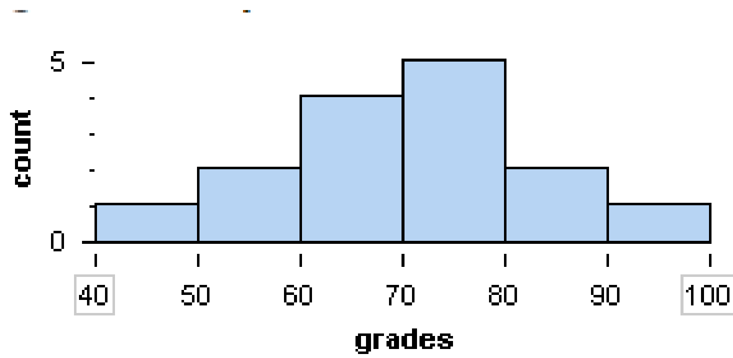
3) If we gathered data to answer the previous research questions, which data could be analyzed using a histogram? How do you know?

4) Which of the graphs below is a histogram? How do you know?





5) This histogram shows the distribution of exam scores for 15 students in a Biology class.



a) How would you describe the shape of this distribution of exam scores? (Use course vocabulary.)

b) Give an interval that describes typical grades on this exam.

c) Estimate the overall range of grades on this exam. (Range = Max – Min)

d) What percentage of the students made a D on the exam (a grade of 60-69%)?

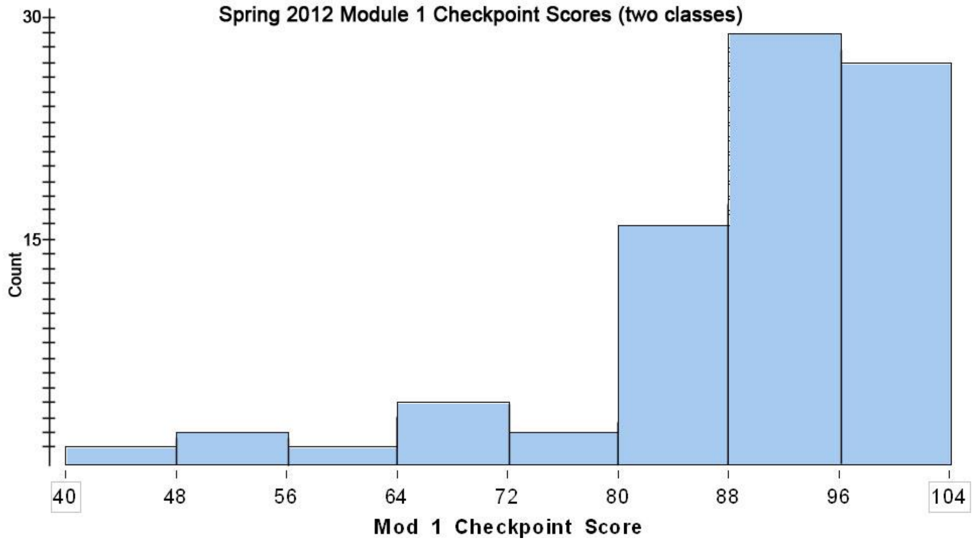
e) What percentage of the students passed the exam with a 70 or better?

f) What percentage of the students made an A or a B?

g) What percentage of the students who passed the exam made an A or a B?

h) What percentage of the students who failed the exam (grades lower than 70) made a D (a grade of 60-69%)?

6) The following is a histogram indicating the distribution of scores on an assignment for 82 students in Statistics classes.



a) How would you describe the shape of this distribution of quiz scores? (Use course vocabulary.)

b) Give an interval that describes typical performance on this quiz.

c) Estimate the overall range of grades on this quiz. (Range = Max – Min)

For each of the following questions, answer the question if the histogram provides enough information to answer it. If not, write "not enough information".

d) What percentage of students scored below 80%?

e) How many students made an A (scored a 90% or higher)?

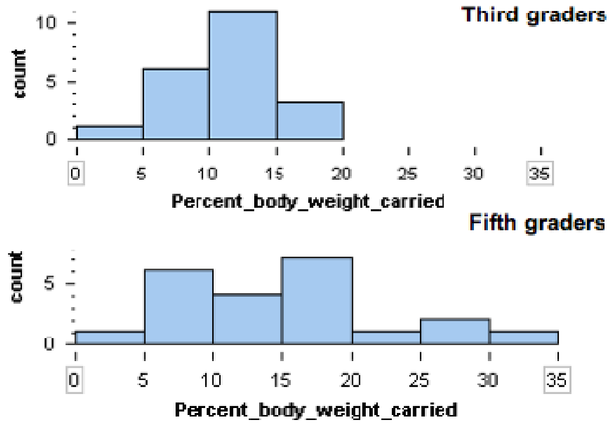
f) What is the lowest grade on the Module 1 Checkpoint?

g) What percentage of the students aced the quiz (a score of 100%)?

h) What is the average (mean) quiz score?

i) Did the majority of students pass the quiz (70% or better)?

7) The data graphed in these histograms describes 43 elementary school children. The variable is “percent of body weight carried in the school backpack.” A child who weighs 60 pounds and carries 9 pounds has a variable value of 15% since 9÷60=0.15=15%. The American Chiropractic Association (ACA) recommends that children carry no more than 10% of their body weight.



% of body weight carried in backpack	Third graders	Fifth graders
0-5%	1	1
5-10%	6	6
10-15%	11	4
15-20%	3	7
20-25%	0	1
25-30%	0	2
30-35%	0	1
Total	21	22

a) Of the 3rd graders, how many are following the ACA recommendation?

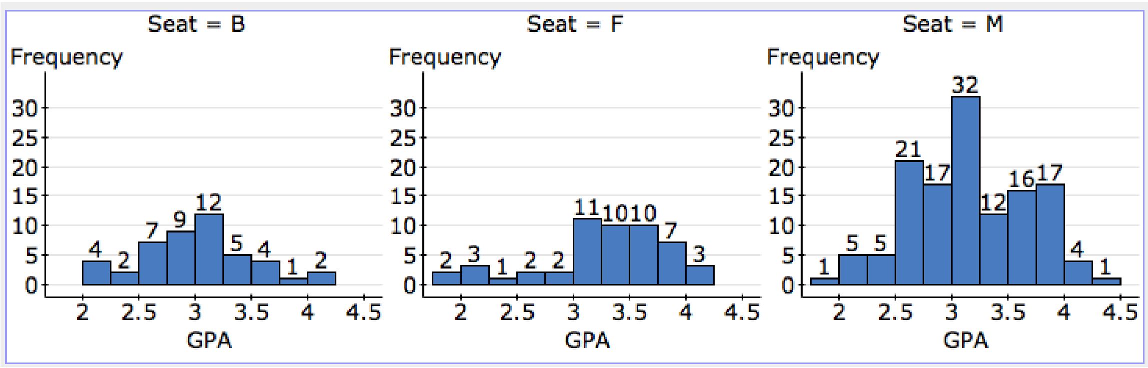
b) Of the 3rd graders what percentage is following the ACA recommendation?

c) Of the 5th graders, what percentage is following the ACA recommendation?

d) Of all the children in this study, what percentage is NOT following the ACA recommendation?

e) Of the 5th graders who are NOT following the ACA recommendation, what percentage are carrying more than 25% of their body weight?

8) This data comes from a survey of 218 students enrolled at Carnegie Mellon University in Pennsylvania. Students were asked if they preferred to sit in the back (B), front (F), or middle (M) of the classroom. 51 reported a preference for sitting at the front; 121 prefer to sit in the middle of the classroom; 46 prefer the back. They also reported their college GPA.



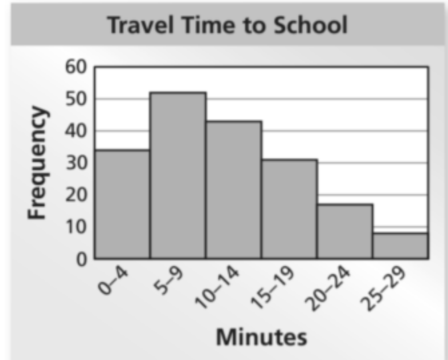
a) Do students who sit at the front of the class tend to have higher GPAs compared to students who sit at the back of class? Calculate percentages to support your answer.

b) Do the students who sit at the back of the class tend to have lower GPAs compared to students who sit at the front or in the middle?

c) To answer the questions in (a) and (b), it is better to use percentages than counts. Why is this?

1.4.3 Homework

1. The following is a histogram of students along with the time it takes them to get to school.



a. Describe the shape of the distribution.

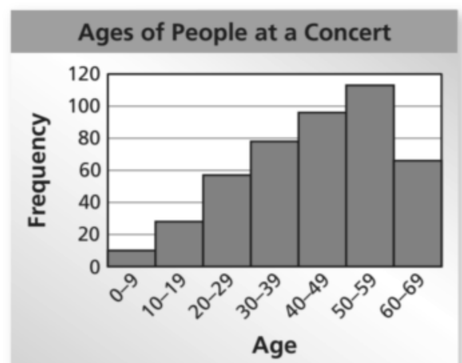
b. Give an interval that describes the typical amount of travel time to school.

c. Estimate the percentage of students who travel 15 minutes or more to school.

d. Estimate the percentage of students who travel less than 15 minutes to school.

e. What would you say is the mean (average) amount of travel time to get to school for these students?

2. The following is a histogram for a population of people who attended a concert.



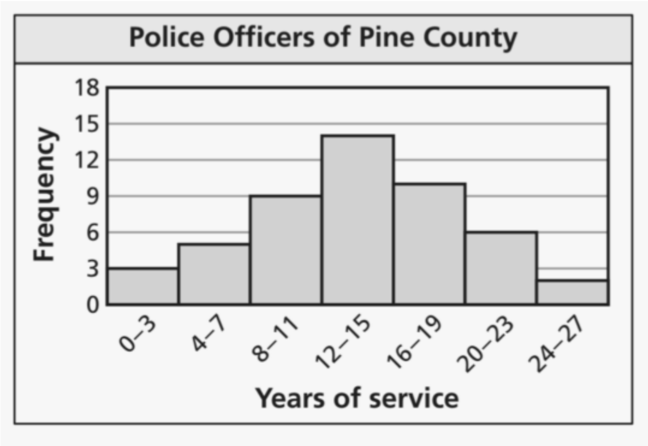
a. Describe the shape of the distribution.

b. Give an interval that describes the typical age of those who attended the concert.

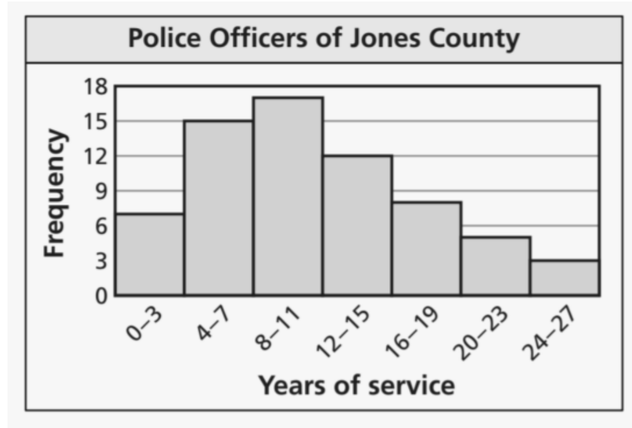
c. Estimate the percentage of people at the concert who were less than 30 years old.

d. Based on the histogram, what musical artist would likely be playing at this concert?

3. The following shows two histograms regarding police officers and their years of service in two different counties.



a. Describe the shape of each distribution.



b. For the police officers of Pine County, approximately how many of them have 16 years of service or more?

c. Give an interval that describes the typical range for years of service in each county.

d. Approximately what percentage of officers have less than 12 years of service in Jones County?

e. Do the officers of Pine County tend to have more years of service than the officers of Jones County? Calculate percentages to support your answer.

4. The following is a histogram of the frequency of amount of waiting times within a local restaurant over the course of a week.

a. Describe the shape of the distribution.

b. Can you give an interval that describes the typical waiting time for this restaurant? Why or why not?

c. What percent of customers waited 30 minutes or more to be seated?

d. What logical reason might explain why this histogram is shaped the way it is?

1.4.4 More on Histograms

Learning Goal: For the distribution of a quantitative variable, describe the overall pattern (shape, center, and spread) and striking deviations from the pattern.

Specific Learning Objectives:

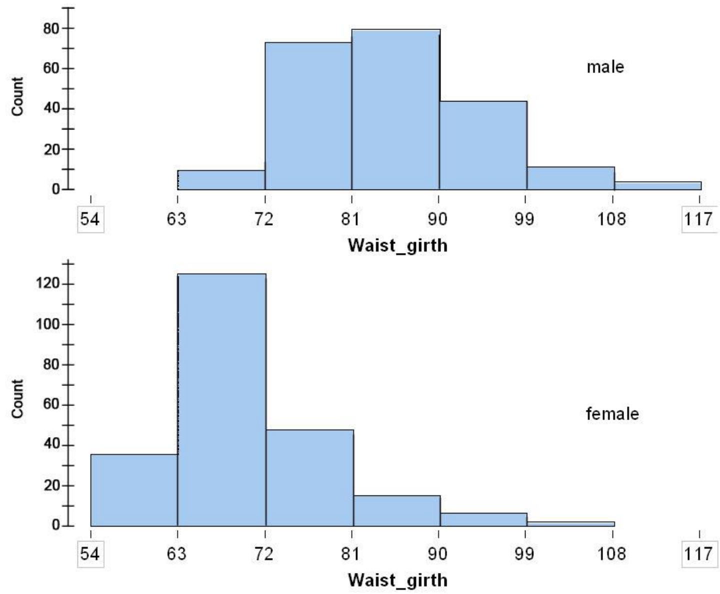
- Compare and contrast the distributions of a quantitative variable for two groups using histograms. Describe shape, give a general estimate of center and the overall range, and calculate relevant percentages.

When comparing two distributions, it can be useful to compare some of the following traits:

1) Do they have the same shape?
2) Compare their variability. Is one of the distributions more spread out then the other?
3) Compare the range of values. Does one of the distributions tend to be located at higher/lower values of x? It can be useful to write the graphs on the same scale when comparing this.
4) Compare the average values/ intervals for each distribution which describes the values you would consider common values for the distribution.
5) Compare different percentages above/below certain numbers.

1.4.4 CLASSWORK

1) Here are data from adults (247 men and 260 women) who exercise regularly. The variable is waist girth, measured in centimeters. Indicate whether you think the following statements are valid (true) or invalid (false). Jot down a few notes to explain how the histograms support or contradict each statement.

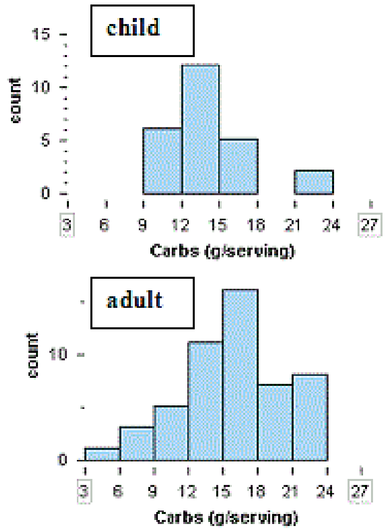


a) In this data set, typical females have a smaller waist girth than typical males.

b) There is less variability in waist girth for females.

c) Here the distributions of waist girth measurements are skewed to the right for both males and females, with only a small percentage of each group having waist girths exceeding 99 cm.

2) These histograms show the distribution of complex carbohydrates (grams per serving) for 76 cereals.



Carbs (g/serving)	Number of Child Cereals	Number of Adult Cereals
3-6	0	1
6-9	0	3
9-12	6	5
12-15	12	11
15-18	5	16
18-21	0	7
21-24	2	8
Totals	25	51

a) How many adult cereals contain 12-15 grams of complex carbs in a serving? How many child cereals contain this amount?

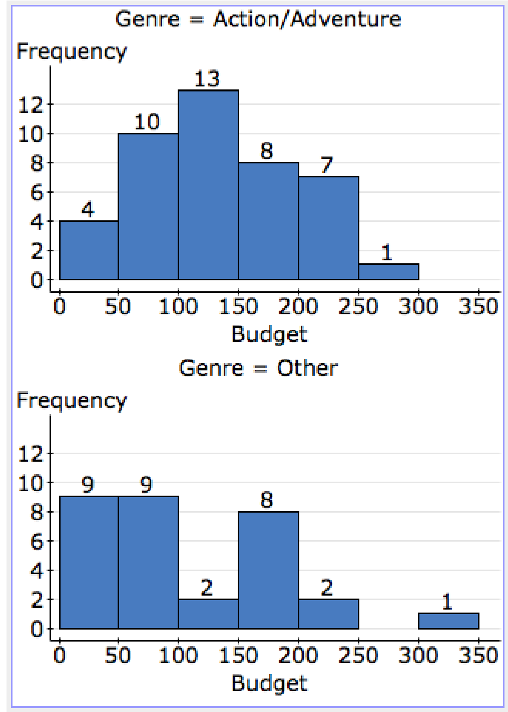
b) Of the adult cereals, what percentage contains 12-15 grams of complex carbs in a serving? Of the child cereals, what percentage contains this amount?

c) Is it better to compare numbers (like in part a) or percentages (like in part b)? Why?

d) The distributions of complex carbs for adult and child cereals overlap quite a bit. This suggests that adult and child cereals have similar amounts of complex carbs in a serving. Use the data to give one piece of precise evidence that supports this statement. Write at least one sentence to explain how your evidence supports this conclusion. (Hint: calculate the percentage within a given interval or give a description of center or spread.)

e) Despite the overlap in the distributions, there are differences. Use the data to give one piece of precise evidence that supports this statement: When compared to child cereals, adult cereals tend to have a larger amount of complex carbs in a serving. Write at least one sentence to explain how your evidence supports this conclusion.

3) These histograms show the budget in millions of dollars for a sample of 74 movies listed in the top 100 USA box office sales of all time. The movies are divided into two genres: Action/Adventure (with 43 movies) and Other (with 31 movies).



a) Describe the shape of each distribution. What does the shape tell us about where most of the data fall?

b) Which genre (Action/Adventure or Other) has the movie with the largest budget?

c) When we take all of the data into account, which genre tends to have larger budgets? (To answer this question, give an interval that represents typical budget amounts for each genre. Use these intervals to support your answer.)

d) Which genre has more variability in budget amounts? (To answer this question, estimate the overall range of budget amounts for each genre. Use your estimates to support your answer.)

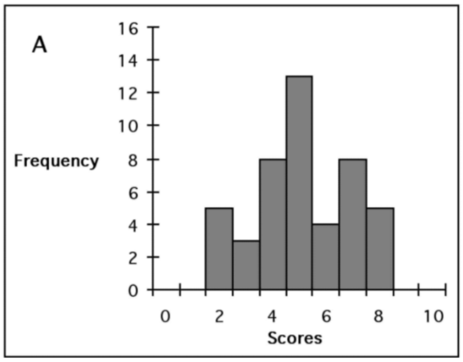
e) Pick the statement that you think is most strongly supported by the data:

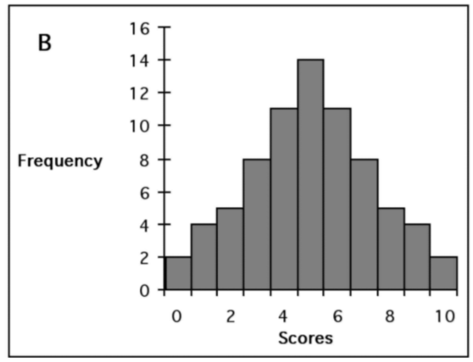
- Action/Adventure movies tend to have larger budgets than other movies.
- Budget amounts are similar for Action/Adventure movies and other movies.

For the statement you picked, support it with at least three precise observations from the histograms. Explain how your observations support the statement you chose.

1.4.4 Homework

1. The following two histograms show the scores received on two quizzes in a Math 112 class.





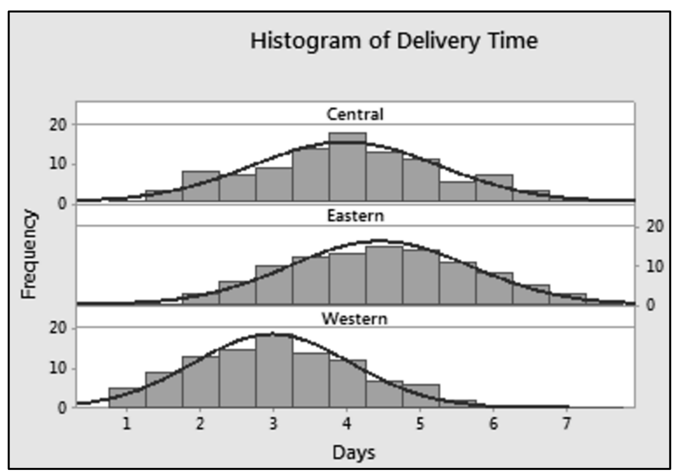
a. Which of the following distributions shows more variability?

A has more variability_________ B has more variability_________

b. Circle the statement (or statements) that led you to select your answer above.

i. Because it's bumpier
ii. Because it's more spread out
iii. Because it has a larger number of different scores
iv. Because the values differ more from the center
v. Other (please explain)

2. The following histograms show the typical delivery times (in days) for a delivery company that has locations in the Central, Eastern, and Western regions of a certain state.

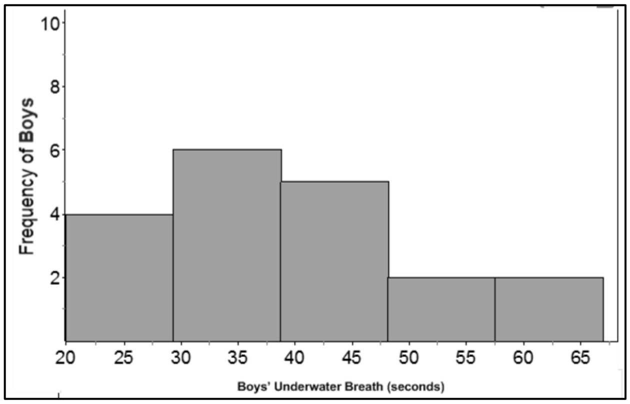


a. Which location has the least variability in their delivery time? Why?

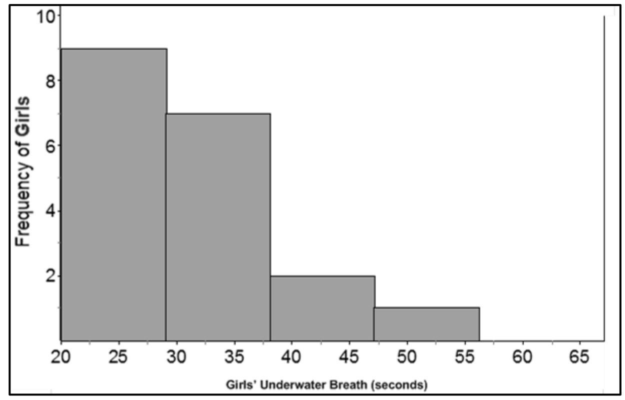
b. Which location has the most variability in their delivery time? Why?

c. Which company tends to have the longest delivery times? (To answer this question, give intervals that represent a typical range for delivery times for each location).

3. The following histograms show how long (in seconds) a group of boys and a group of girls can hold their breath underwater.



a.



About how many boys were able to hold their breath underwater for more than 29 seconds? How many girls were able to?

b. From the boys' histogram, about what percentage were able to hold their breath underwater for more than 29 seconds? What percentage of girls were able to?

c. Is it better to compare numbers (like in part a) or percentages (like in part b)? Why?

d. The distributions in number of seconds that these girls and boys can hold their breath overlap quite a bit. This suggests that boys and girls have similar amounts of time that they can hold their breath. Use the data to give one piece of precise evidence that supports this statement. Write at least one sentence to explain how your evidence supports this conclusion. (Hint: calculate the percentage within a given interval or give a description of center or spread.)

e. Despite the overlap in the distributions, there are differences. Use the data to give one piece of precise evidence that supports this statement: When compared to girls, boys tend to be able to hold their breath for a longer period of time. Write at least one sentence to explain how your evidence supports this conclusion.

1.5 Measures of center

Introduction

Sometimes we want to find a single number to represent a set of data. Usually, we want to use a number that represents the "center" of the data. We call these numbers **measures of center.**

Specific Learning Objectives:

1. You will be familiar with the notations for **population mean** and **sample mean**.
2. Given a data set, you will be able to calculate **mean**, **median**, **mode**, and **midrange**.
3. You will know which measures of center are very sensitive to **outliers**.
4. You will be able to figure out what value is needed to attain a specified mean.
5. You will be able to round your mean appropriately.
6. Given a data set, you will know which measure of center is the most appropriate.

Mean

One of the most common measures of center is something called a **mean** or **average**. The symbol we use to denote a mean that comes from a sample is $\bar{x}$.
The symbol we use to denote a population mean is μ. It is the Greek letter for lowercase "m" and is pronounced "mu".

How to Calculate the Mean

To find the mean of a set of values, we add all the values together and divide by the total number of values. The notation for **summation** (adding) is the Greek letter for uppercase "S": $\sum$ (pronounced "sigma").

We use the **variable** x to represent the data values. Thus $\sum x$ means add all the data values. If we want to write the mean using math notation, it would look like this:

$$\bar{x} = \frac{\sum x}{n}$$

where the number n is the number of values in our data set.

Example 1: Suppose a statistics student named Maria gets a 70 on the first test, an 80 on the second test, a 90 on the third test, and a 84 on the fourth test. Her test average is:

$$\bar{x} = \frac{70+80+90+84}{4} = 81$$

Example 2: Now suppose that Maria oversleeps and misses the fifth test so she earns a 0. Let's calculate her new test average:

$$\overline{x} = \frac{70+80+90+84+0}{5} = 64.8$$

So Maria went from a low B to a middle D!

Example 2 shows that the mean is VERY sensitive to extreme values. We call these extreme values **outliers** and say that the "mean is sensitive to outliers".

Rounding Your Answers

Because the calculation of the mean involves division, there will be times your mean calculation results in a number with many decimal places. In general, we will use the following round off rule:

1. Determine which digit is the last one you want to keep (i.e. round to the nearest tenths place means you want to keep the digit to the right of the decimal point).
2. Leave it the same if the next digit to the right is less than 5.

Example 3: Round 123.456 to the nearest tenth.

Since we are rounding to the tenth place, we look first at the digits in the tenths place and find that 4. Since the number to the right of 4 is 5 (or more), we need to round up.

Thus, our rounded value is 123.5.

Statistical Convention for Rounding

When presented with a data set with many numbers and no specific round off instructions, the standard practice is to round one decimal place beyond what's given in the data.

Example 4: Find the mean of the following numbers: 88.9, 98.7, 102.7

$$\bar{x} = \frac{88.9 + 98.7 + 102.7}{3} = 96.666666666...$$

According to the statistical convention, we want to round to the nearest hundredths place since the original data was given with accuracy to the nearest tenths place. Looking to the right of the hundredths place, we see a 6 which is greater than 5, which means we need to increase the value in the tenths place by 1.

Thus, our rounded value would be 96.67

Finding a Data Value to end up with a Particular Mean

Example 5: Recall that Maria's first five test scores were: 70, 80, 90, 84, and 0. Suppose Maria has one more test in the semester and wants to finish with a mean of least 70. What does she need to get on her last test?

Let x = the test score Maria needs on her 6th test. We want the mean to be 70. Thus,

$$\frac{70 + 80 + 90 + 84 + 0 + x}{6} = 70$$

$$\frac{324 + x}{6} = 70$$

$$324 + x = 420$$

$$x = 96$$

Thus, if Maria wants to get exactly 70 she needs to get a 96 on her 6th test. Of course, she wants AT LEAST a 70 so she needs a 96 or higher. Thus our answer is:

$$x \geq 96$$

When is Mean NOT the Best Measure of Center?

There are certain situations where mean is not the best **measure of center**. As we saw in the example involving Maria's missed test score, the mean is very sensitive to **outliers** (unusually large or small values). As a consequence, the mean may give a misleading value.

Example 6: Here is a sample of 5 home prices in the Bay Area (in thousands of dollars):

570 387 410 483 1060

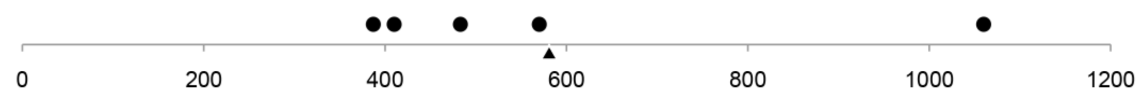
Notice that there is one extremely expensive house in our sample. If we calculate the **mean** we have:

$$\bar{x} = \frac{570+387+410+483+1060}{5} = 582$$

Does \$582,000 seem like a good measure of the center?

582 does NOT seem like a good measure of the center since four of the five values are less than 582! The presence of the outlier 1060 moves the mean so high that it is no longer a good representation of the center of the data.

Here is a picture that depicts this situation with the mean labeled as a small triangle:



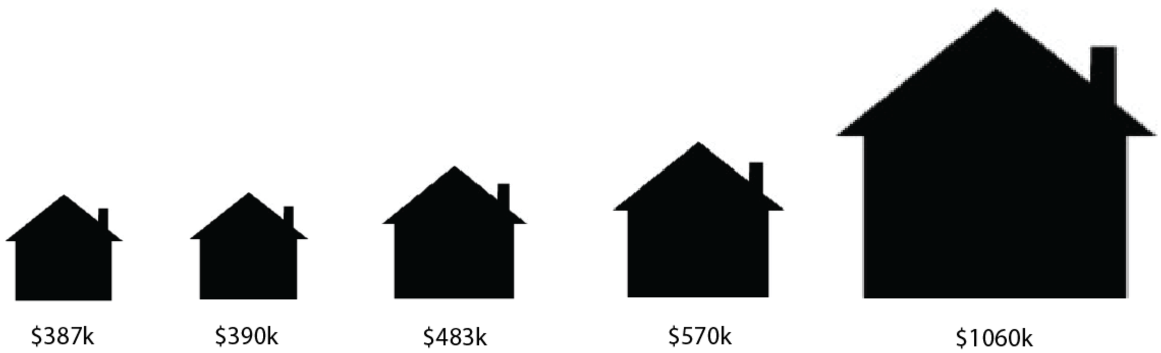
A consequence of this is that if a person is looking to buy a house in this neighborhood, they might mistakenly think that they can't afford a home in this neighborhood because the mean house price is so high (\$582k).

Perhaps there is a different measure of center that is NOT sensitive to outliers… stay tuned!

Median

Arguably, the most common **measure of center** is the **mean** (or average). One problem with the mean is that it is very sensitive to **outliers** (or unusual values). This sensitivity can lead you to make poor decisions or reach inappropriate conclusions. We saw this when Maria missed one test. The zero (an outlier compared to her other test scores) pulled her mean way down.

Example 7: Suppose you are thinking about buying a house in a particular neighborhood. If you calculate the **mean house price** you may be misled to believe you cannot afford a house in that neighborhood because the mean house price is so high. However, because the mean is so sensitive to outliers the presence of a single extremely expensive home can throw off the calculations.



Example 8: If you were to look at mean salaries of graduates by major of University of North Carolina graduates in 1984 you would've discovered that the major with the highest median salary was Geography! Yes, Geography! The reason is that there were not that many Geography majors and one of them happened to be Michael Jordan, who made lots and lots of money as (arguably) the world's best basketball player and highest paid geography major.

The previous examples are classic examples where mean is not the best measure of center because of the presence of outliers. In these two situations (home prices and income) the standard measure of center to use is **median**.

Median is not only a measure of center, it is also referred to as a **measure of relative standing**. A measure of relative standing is a number that indicates where a particular value lies in relation to the rest of the values in a set of data. (i.e. "relative" means "in relation to other values" and "standing" means "where it is")

The **median** is the middle value of an ordered list. It is the number that has half the values smaller than it and half the values bigger than it.

How to Calculate the Median

1. Arrange the numbers in either ascending (low to high) or descending (high to low) order.
2. Find the middle number.
3. If there is an **even** number of values, take the average of the two middle values.

Example 9: Calculate the median of a sample of 5 home prices in the Bay Area (in thousands):

570 387 390 483 600

If we re-arrange the values in ascending order we get:

387 390 **483** 570 600

Since there are five numbers, the middle value is the third one. Thus, the **median is 483**. If we calculate the mean we get:

$$\bar{x} = \frac{387 + 390 + 483 + 570 + 600}{5} = 486$$

We see that the mean \$486k and the median \$483k are quite close.

Example 10: Suppose we replace the 600 value with 1060. Calculate the new median:

387 390 483 570 1060

The median is still 483, because increasing the value of the largest number does not change the position of the median.

In contrast, if we calculate the mean we get:

$$\bar{x} = \frac{387 + 390 + 483 + 570 + 1060}{5} = 578$$

Notice that the one very large number changed the mean quite a bit. Now the mean is larger than all but one of the values!

These last two examples show that the mean is very sensitive to outliers (unusually large or unusually small values) while the median is not.

Example 11: Calculate the median of this sample of 6 home prices in the Bay Area (in thousands):

570, 387, 390, 483, 600, 452

If we re-arrange the values in ascending order we get:

387, 390, **452,** **483**, 570, 600

Since there are six numbers, we have an even number of data values. The median is the average of the two middle values 452 and 483. Thus, the median is

$$\text{median} = \frac{452 + 483}{2} = 467.5$$

Mode

The **mode** is the most frequently occurring data value.

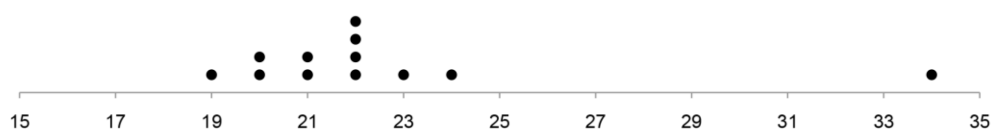
How to Calculate the Mode

To find the mode, you should first arrange the data values in order. This will make it easier to count the number of occurrences of each data value. The most frequently occurring value is the **mode**. If there appears to be two modes, we say the data is **bimodal** (the prefix "bi" means "two", as in *bi*cycle and *bi*focals). If there are more than two modes, we say the data is **multi-modal**. If no data value appears more than once, we say there is no mode.

Example 12: Calculate the mode of the following set of student age data:

17 18 19 20 20 21 21 **22** **22** **22** **22** 23 24 34

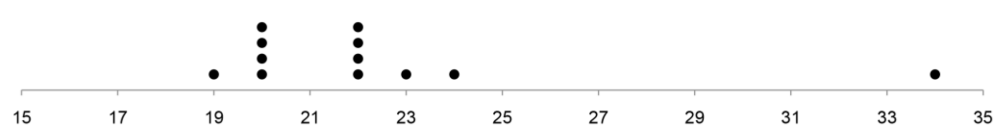
Luckily, the data is arranged in order already. We see that 22 appears four times, which is more than any other data value. Therefore, 22 is the mode. Here is a **dotplot** that describes this data (notice how the **mode** has the most dots above it):



Example 13: Calculate the mode of the following set of student age data:

17 18 19 20 20 20 20 22 22 22 22 23 24 34

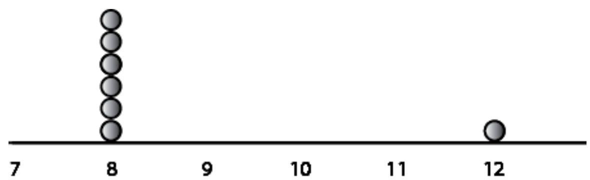
In this case, we see that both 20 and 22 appear four times. In this case, there are TWO modes: 20 and 22. When we have two modes we say the data is **bimodal**. A data set with more than two modes is called **multimodal**. Here is a dotplot that depicts this bimodal situation:



When is Mode the Best Measure of Center?

There are certain circumstances where mode is the ONLY measure of center. Unlike mean and median, we can find the mode of **qualitative** or categorical data. For example, if your data set consisted of all the surnames of citizens of South Korea, the mode would be Kim – the most frequently occurring Korean last name.

There are also other situations that depend on the distribution of data where mode may be the best measure of center. For example, if your data looked like this:



one could certainly argue that mode was the best measure of center, since every single value other than 12 is 8. The value 12 is an outlier.

Unlike the **mean**, the mode is NOT sensitive to outliers. If we added the value 29 to the data set above, our mode would remain unchanged.

Midrange

The **range** is a measure of spread that tells us how spread out the data is.

range = highest value – lowest value

The **midrange** is a measure of center that is the value exactly in the middle of the highest and lowest values. In other words, the midrange is just the mean of the highest and lowest value.

$$\text{midrange} = \frac{\text{highest value} + \text{lowest value}}{2}$$

Example 14: Suppose we have the following data set is the number of dollars the each airline company made in baggage fee revenue (in millions of dollars):

833, 625, 506, 212, 144, 112, 96

The range = \$833 – \$96 = \$737 million dollars. This means that that gap between the airline that made the most from baggage fees and the airline that made the least in this sample is 737 million dollars.

The midrange = $\frac{\$833 + \$96}{2}$ = \$464.5 million dollars

Strengths of midrange:

- Easy to calculate – only uses two data values

Weaknesses of midrange:

- Extremely sensitive to outliers
- Not a great measure of center as it only uses two data values

1.5.1 Classwork

1) Here are two sets of exam scores, one for a class that has 4 students and one for a class that has 15 students.

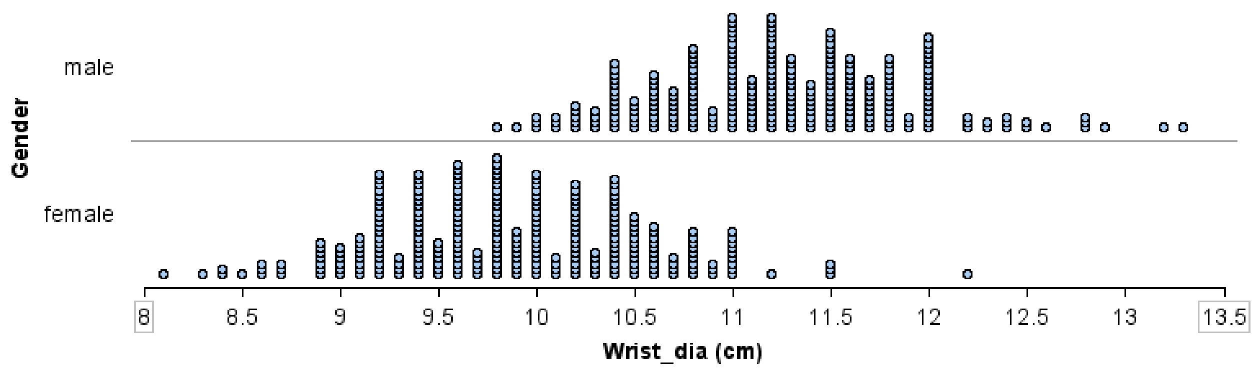
   Class A: 80, 90, 90, 100

   Class B: 60, 65, 65, 70, 70, 70, 75, 75, 80, 80, 80, 80, 80, 85, 100

   a) Without doing any calculations, which class do you think will have a larger mean? Why?

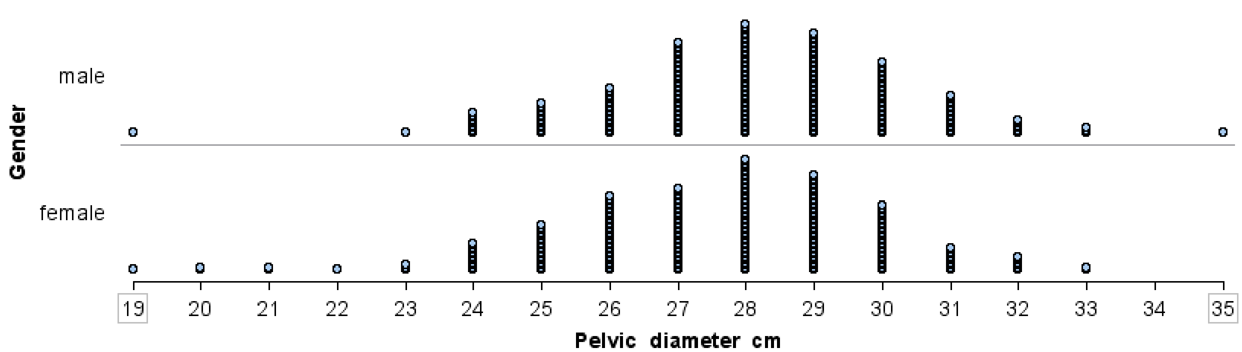
   b) Now calculate the mean for each class. Which is larger? Why does this make sense?

2) For this data, is the mean wrist measurement for men (larger than, smaller than, or about the same as) the mean wrist measurement for

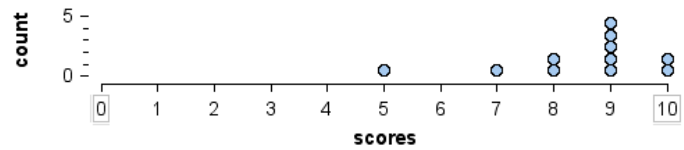


women? (Obviously, you can't calculate the means, so jot down notes about how you thought about this.)

3) For this data, is the mean pelvic diameter for men (larger than, smaller than, or about the same as) the mean pelvic diameter for women? (Obviously, you can't calculate the means, so jot down notes about how you thought about this.)



4) The dot plot gives quiz scores for a small class.



a) What is the mean? Show your work or explain how you got your answer.

b) What is the median? Show your work or explain how you got your answer.

c) Which measure (the mean or the median) is the better way to represent typical performance on this quiz? Why?

5) This table gives quiz scores for a different class.

Scores	Number of students
5	1
6	3
7	5
8	3
9	1

a) What is the mean? Show your work or explain how you got your answer.

b) What is the median? Show your work or explain how you got your answer.

c) Which measure (the mean or the median) is the better way to represent typical performance on this quiz? Why?

6) Construct a data set where neither the mean nor the median is a reasonable "typical" value.

7) For this problem, use the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.

a) List five digits that have a median of 7 and a mean of 7 (repeats allowed). Find a different set of 5 digits that work.

b) List five digits that have a median of 7 and a mean that is less than 7 (repeats allowed.) Give the mean of your 5 digits. Find a different set of 5 digits that work.

c) List five digits that have a median of 7 and a mean that is more than 7 (repeats allowed.) Give the mean of your 5 digits. Find a different set of 5 digits that work.

8) Imagine that you have a bag filled with 9 numbers. The mean and the median of the numbers in the bag are both 6.

   a) You draw a number out of the bag. It is a 4. You replace it with a 1. Does the mean of the numbers in the bag get bigger, smaller, or stay the same? What about the median? Jot down some notes to explain how you figured this out.

   b) You draw a number out of the bag. It is an 8. You replace it with 8 ones. Does the mean of the numbers in the bag get bigger, smaller, or stay the same? What about the median? Jot down some notes to explain how you figured this out.

9) Even though statisticians call the mean and the median measures of "center", it might be less confusing to think of the mean and the median as ways to summarize a distribution with a single number. In other words, the mean and the median are ways to represent a "typical" measurement in a data set. When do you think it is better to use the median and when it is better to use the mean to summarize a distribution?

1.5.1 Homework

If it is necessary to round: round money to the nearest penny. Otherwise use our statistical convention for rounding.

1. In a few sentences explain how you calculate each of the following:
   a. Mean b. Median c. Mode d. Midrange
2. Round to the indicated place
   a. 50.384523 to the nearest thousandth.
   b. 86.75309 to the nearest hundredth.
3. Calculate the mean, median, mode, and midrange of these temperatures from Anchorage Alaska: –6°, 4°, 14°, 21°, 25°, 16°, 18°, 12°, 2°
4. If Janice gets 17, 16, 13, 19, and 18 on her first few quizzes, what is her mean quiz score? What is her median quiz score?
5. In the first 6 games of the 2015 NBA playoffs, Lebron James scored 20, 30, 31, 27, 19, and 33 points. Calculate the median number of points scored over those 6 games. Also find the average number of points scored.
6. Temperatures at the State Capitols was recorded on March 15 for the following states:

State	In Fahrenheit (ºF)	In Celsius (ºC)
Connecticut	49.0	9.4
Maine	41.0	5.0
Massachusetts	47.0	8.8
New Hampshire	50.1	10.1
Rhode Island	43.8	6.6
Vermont	42.9	6.1

   a. Find the mean, median, and midrange temperature measured in degrees Fahrenheit for the states in the table.
   b. Find the mean, median, and midrange temperature measured in degrees Celsius for the states in the table.
7. Kirk has the following four quiz scores (out of 20 points): 15 18 13 10
   a. What is his mean score now?
   b. To have a B average on his quizzes, he needs a mean of 16 or 17. What does he need to get on the next quiz to bring his grade up to a B?

8. Agnes scored 87 on the first test, 73 on the second test, and 81 on the third test. Suppose Agnes has one more test in the semester and wants to finish with a mean of least 80. What does she need to get on her last test?

9. Jay is training for a marathon. On Monday he ran 7 miles. On Tuesday, he ran 10 miles. On Wednesday he ran 2 miles, and on Thursday he ran 6 miles. If Jay wants to average 7 miles for the week, how far must he run on Friday (he doesn't run on the weekends).

10. Roger, Jay's arch-rival, is also training for a marathon. On Monday he ran 5 miles. On Tuesday, he ran 11 miles. On Wednesday he ran 6 miles, and on Thursday he ran 7 miles. If Roger wants to average 7.5 miles for the week, how far must he run on Friday (he doesn't run on the weekends either).

11. Gasoline Prices:

    a. Find the mean and median of the following gasoline prices per gallon in California (regular, mid, premium, diesel):

       $4.014 $4.116 $4.214 $4.320

    b. Find the mean and median of the following gasoline prices per gallon in Missouri (regular, mid, premium, diesel):

       $3.910 $4.003 $4.104 $4.023

    c. Do gas prices appear to be higher in California than in Missouri? Why?

    d. How would we calculate the "average amount" of money a Californian paid for gas for an entire month?

    e. For part d, why do you think it makes sense for us to work with the mean rather than the median or mode?

12. Use this table to answer the following questions:

Age	**Frequency**
18	20
20	9
21	8
23	7
29	6

    a. How many people were surveyed?

    b. What is the mean Age?

    c. What is the mode Age?

    d. The median is 20 years. Why is it 20 years and not 21 years?

    e. What is the midrange?

13. On the last test the lowest score was 65 out of 100. When Mr. Nguyen computed the mean he got an answer of 58. Explain how he knew that was not the correct mean?

1.5.2 Measures of Center (Mean as a balancing point)

Learning Goal: For the distribution of a quantitative variable, describe the overall pattern (shape, center, and spread) and striking deviations from the pattern.

Specific Learning Objective: Discover why the mean is called a balancing point.

The goal of each of the problems in this exercise is to add more elements to the set but keep the mean of the set the same.

To help you understand let's look at the following set: 5,5,5,5

The mean of the set is $\frac{5+5+5+5}{4} = 5$

What would happen to the mean if we added the number 4 to the set?

New set: 5,5,5,5,4

Mean = $\frac{5+5+5+5+4}{5} = 4.8$ This is an example to show that if we add a number that is smaller than the mean of the set, the new mean will be smaller.

What would happen if instead we had added the number 6 to the set?

New set: 5,5,5,5,6

Mean = $\frac{5+5+5+5+6}{5} = 5.2$ This is an example to show that if we add a number that is larger than the mean of the set, the new mean will be larger. Also notice that 4 was one smaller than the original mean (5) and 6 was one bigger. They had similar effects on the mean (adding the 4 decreased the mean by 0.2 and the 6 increased the mean by 0.2).

What would happen if you added both the 4 and the 6?

New set = 5,5,5,5,4,6

Mean = $\frac{5+5+5+5+4+6}{6} = 5$

The 4 being 1 smaller than the mean, and the 6 being one larger than the mean balance each other out and the mean remained unchanged.

Consider the set: 5,5,5,5 (which has a mean of 5). If you added 3 to the set, what additional number should you add in order to keep the set the same?

Since 3 is 2 smaller than the mean we need to balance it with a number that is 2 larger than the mean. 5+2 = 7. We should include 7 as our additional number.

New set: 5,5,5,5,3,7

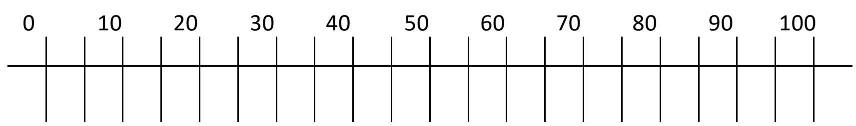
Mean: $\frac{5+5+5+5+3+7}{6} = \frac{30}{6} = 5$

1.5.2 CLASS ACTIVITY

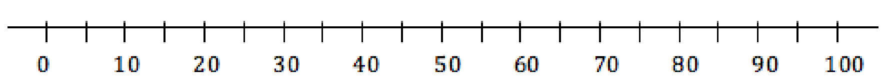
1) Emma is struggling in her math class (she failed to apply herself at first, but her effort is steadily improving). To date, her quiz scores are 0, 43, 56, 66, 68, 74, 83, and 90 (each quiz is worth 100 points).

   a) What is her mean score? $\bar{x}$ = ________

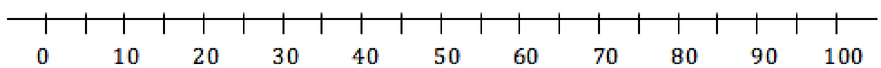
   b) Student A has the same mean score as Emma, but his quiz scores are all the same score (unlike Emma whose scores improved over time.) Make a dot plot of Student A's 8 quiz scores and mark the mean.



   c) Student B also has the same mean score as Emma. She has 6 quizzes with a score of 60 and one quiz with a score of 70. What did she score on the 8th quiz? Make a dot plot of Student B's 8 quiz scores and mark the mean. (Jot down a few notes to remind yourself how you figured this out.)



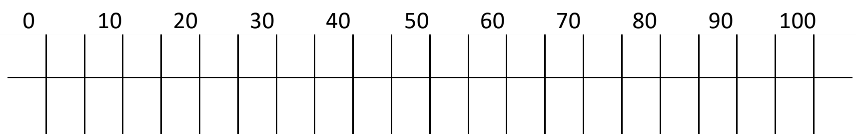
d) Student C also has the same mean score as Emma. She has 5 quizzes with a score of 60 and two quizzes with a score of 70. What did she score on the 8th quiz? Make a dot plot of Student C's 8 quiz scores and mark the mean. (Jot down a few notes to remind yourself how you figured this out.)

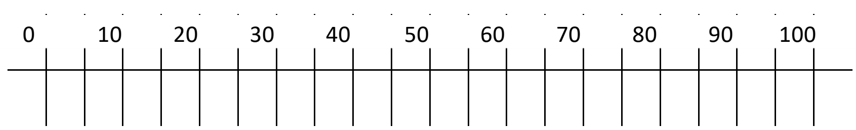


e) Student D also has the same mean score as Emma. Is it possible for Student D to have a 100 on a quiz? What is the maximum number of 100-point quiz scores possible? How do you know?

2) Draw two dotplots of 8 quiz scores such that:

- both have the same mean as Emma's (indicate the mean with a triangle);
- one dot plot has very little spread;
- one dot plot has a lot of spread.





3) Write Emma's mean quiz score below. For each of her quiz scores, record the signed distance from the mean by calculating *Score – Mean*. (Note: A score below the mean will have a negative distance. A score above the mean will have a positive distance.)

Emma's Mean Quiz Score: $\bar{x} =$ __________

Emma's Quiz Score	0	43	56	66	68	74	83	90
Signed Distance from the Mean (quiz score minus the mean)								

Now add up the signed distances from the mean. What did you get?

____

4) For one of the distributions you created in (2), record the quiz scores in the table below. For each quiz score, record its signed distance from the mean by calculating *Score – Mean*. (As before, a score below the mean will have a negative distance. A score above the mean will have a positive distance.)

Mean Quiz Score: $\bar{x} =$ __________ (Should be the same mean as Emma's.)

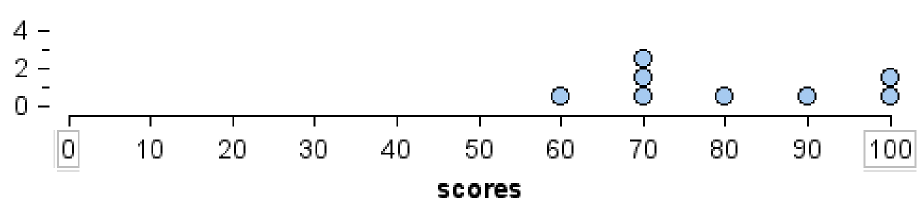
Quiz Score								
Signed Distance from the Mean (quiz score minus the mean)								

Now add up the signed distances from the mean. What did you get?

____

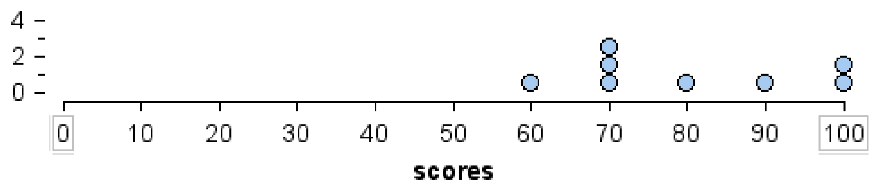
5) What do you notice when you add the signed distances from the mean? Why do you think this happens?

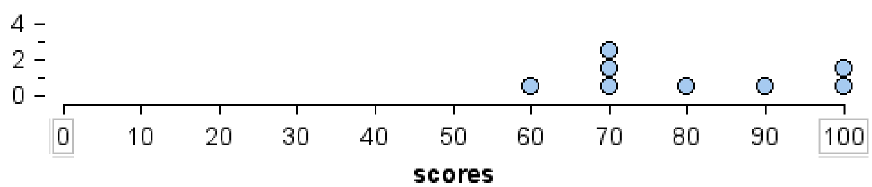
6) For the dotplot of quiz scores below, find the mean. If the student takes another quiz and it doesn't change his mean score, what did he score on this quiz? Explain why this makes sense using signed distances from the mean.

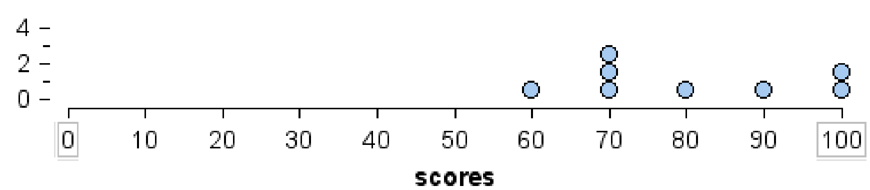


7) The student with the distribution of quiz scores below takes two more quizzes. One score is above his mean and the other is below his mean, but his mean does not change. What are some possible pairs of quiz scores for this student? Show three possibilities using the dotplots below.

How can signed distances from the mean be used to find possible pairs?



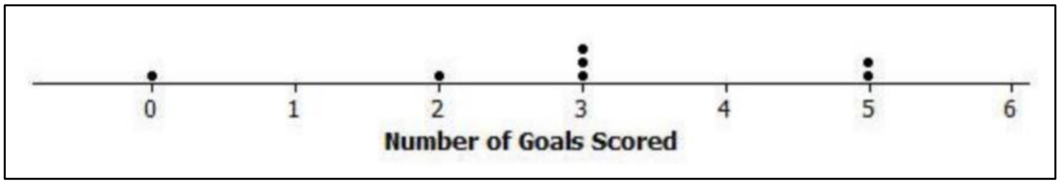




8) Why do you think that statisticians call the mean the “balancing point” of a distribution?

1.5.2 Homework

1. The dot plot below shows the number of goals scored by a college's soccer team in 7 games so far this season.



Use the balancing process to explain why the mean number of goals scored is 3.

2. The times (rounded to the nearest minute) it took each of six new gym goers to run a mile are **7**, **9**, **10**, **11**, **11**, and **12** minutes.

a. Draw a dotplot representation of the data.

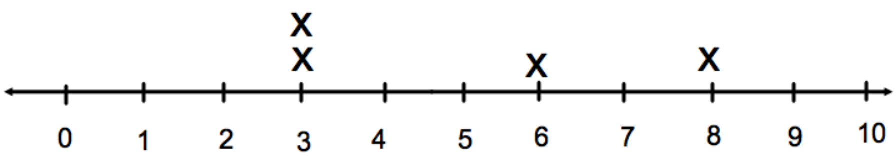
b. Suppose that a classmate thinks the mean is **11** minutes. Is he correct? Explain your answer.

3. The number of phones (landline and cell) owned by the members of each of nine families is **3**, **5**, **6**, **6**, **6**, **6**, **7**, **7**, **8**.

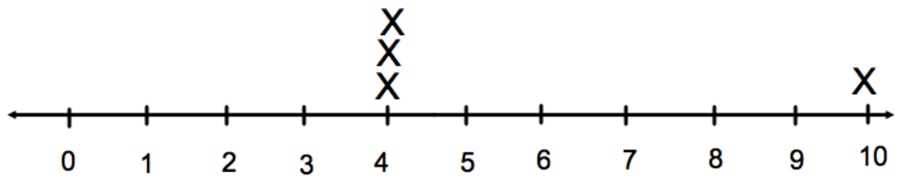
a. Use the mathematical formula for the mean to find the mean number of phones owned for these nine families.

b. Draw a dot plot of the data, and verify your answer in part (a) by using the balancing process.

4. On the following dotplot, find the balancing point (mean).



5. The following dot plot shows the amount of movies watched by students. What two missing data points would make the mean 4? Explain your reasoning.



6. Nine teachers were asked how many years they have been teaching and they answered as follows:
1, 3, 4, 4, 4, 5, 7, 8, 9.

a. Create a dot plot of the data.

b. Where is the "balance point" of your data?

c. What does this balance point represent?

d. How can you find the balance point without using the formula for mean?

e. What happens to the mean if a new number, 2, is added to the data set?

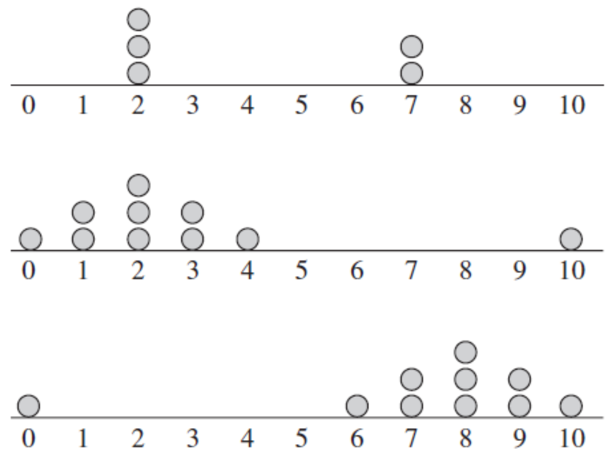
f. What happens to the mean if a new number, 8, is added to the data set?

g. What happens to the mean if a new number, 0, is added to the data set?

h. Find two numbers that can be added to the data set and not change the mean. Explain how you chose these two numbers.

i. Find three numbers that can be added to the data set and not change the mean. Explain how you chose these three numbers.

7. For each of the next dot plots, guess the approximate location of the mean by thinking about where the balance point for the data would be. Then check how close your guess was by calculating the mean.

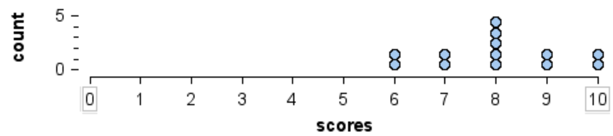


1.5.3 Shape and Measures of Center

Learning Goal: For the distribution of a quantitative variable, describe the overall pattern (shape, center, and spread) and striking deviations from the pattern.

Specific Learning Objective: Relate measures of center to the shape of the distribution. Choose the appropriate measure for different contexts.

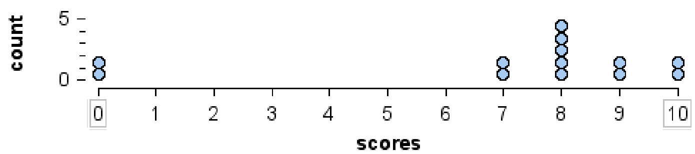
1) Here is a dot plot of Hilda's quiz scores in her math class.



a) What is the shape of this distribution?

b) Determine Hilda's median quiz score. Calculate Hilda's mean quiz score.

2) Oops, we made a mistake recording two of Hilda's scores. The two scores recorded as 6 should have been 0. Here's the revised dot plot of her quiz scores.



a) What is the shape of this distribution?

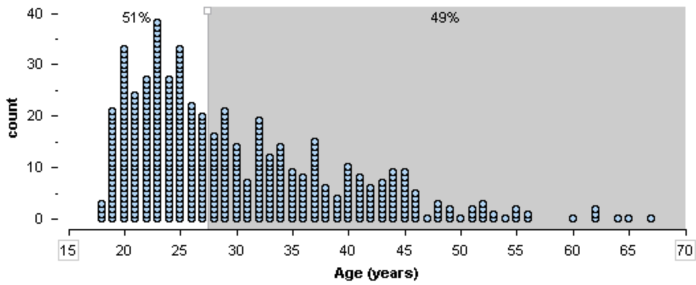
b) Find her revised median quiz score. Is it less than, more than, or about the same as her previous median quiz score? Why does this make sense?

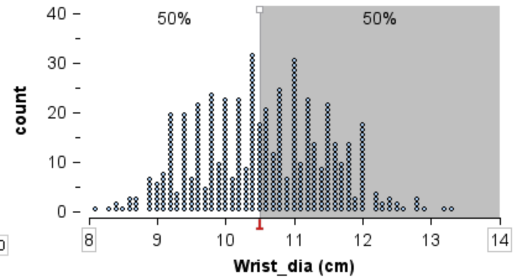
c) Find her revised mean quiz score. Is it less than, more than, or about the same as her previous mean quiz score? Why does this make sense?

d) Which measure of center (mean or median) would best represent her typical performance on the quizzes? Explain.

3) Write a few sentences that explain how the shape of the distribution might influence whether the mean is larger than, smaller than, or about the same as the median.

4) For each distribution give an estimate for the median. Then say whether the mean is probably greater than, less than, or about equal to the





median. Jot down some notes to explain how you figured this out.

5) Which of the following distributions is likely to have a mean that is smaller than the median? Jot down some notes to explain how you figured this out.

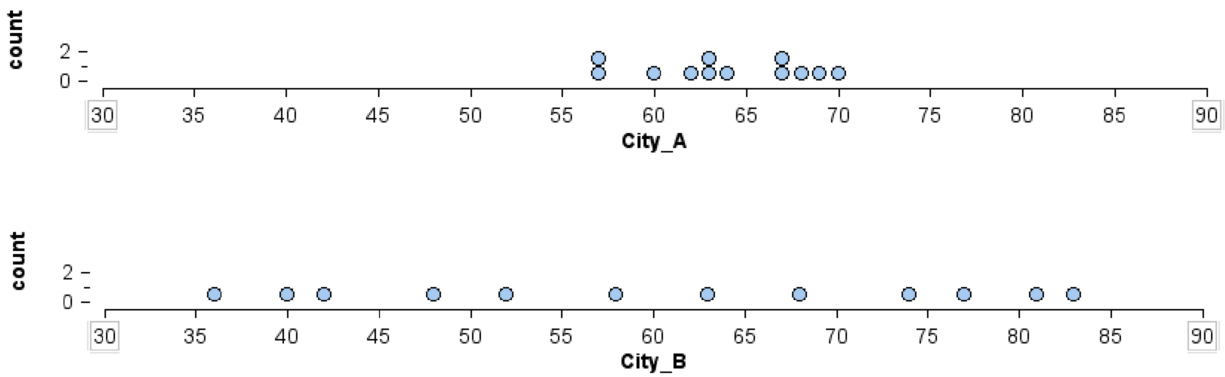
   a) scores on an easy test, where almost all of the grades were A's and B's (scores of 80 to 100), with two students earning a zero by missing class.

   b) repeated weighings of peanut M&M's in a "one-pound" bag using a digital bathroom scale.

   c) salaries of NBA basketball players (This distribution is strongly right skewed by the few superstars who make much more money that the rest of the players.

6) According to the U.S. Bureau of Census, the median family income in the United States in 2008-2012 was $53,046. Why do you think the median, rather than the mean, was reported by the U.S. Bureau of Census?

7) The dot plots below show the average monthly high temperatures for New York City and San Francisco over a period of 10 years.



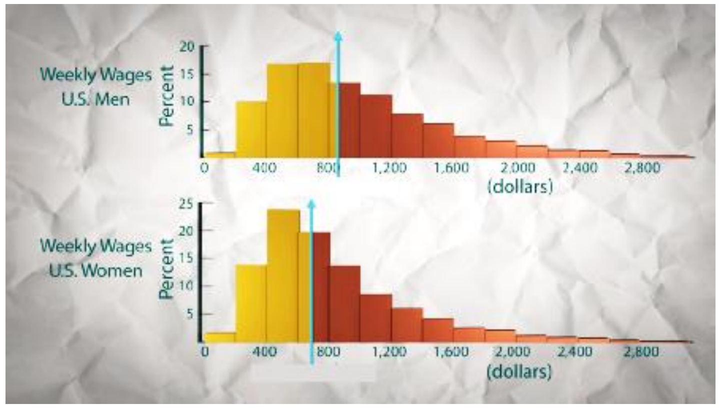
a) Is San Francisco City A or City B? How do you know?

b) One city has a median of 60.5°F; the other has a median of 63.5°F. Which is the median monthly high temperature in San Francisco? How do you know?

8) Summarize what you have learned about how to choose between the mean and median as the best measure of center based on the shape of the distribution.

1.5.3 Homework

1. The histograms below show the weekly wages for Americans in 2011, separated by gender.



For each distribution give an estimate for the median. Then say whether the mean is probably greater than, less than, or about equal to the median. Jot down some notes to explain how you figured this out.

2. Here are the starting salaries, in thousands of dollars, offered to the 20 students who earned degrees in computer science in 2011 at a university.

63 56 66 77 50 53 78 55 90 65 64 69 59 76 48 54 49 68 51 50

a. Find the median salary.
b. Find the mean salary.
c. Find the mode of the salaries.

d. Is the mean about the same as the median or not? What feature of the distribution explains the difference between the mean and the median? Is the mode a good measure of the center for these data?

3. Each month, the Commerce Department reports the "average" price of new single-family homes. For August 2012, the two "averages" reported were $256,900 and $295,300. Which of these numbers was the mean price and which was the median price? Explain your answer.

4. In 1961 New York Yankee outfielder Roger Maris held the major league record for home runs in a single season, with 61 home runs. That record held for 37 years. Here are Maris's home run totals for his 10 years in the American League.

13, 23, 26, 16, 33, 61, 28, 39, 14, 8

a. Find the mean number of home runs that Maris hit in a year, both with and without his record 61. How does removing the record number of home runs affect his mean number of runs?

b. Find the median number of home runs that Maris hit in a year, both with and without his record 61. How does removing the record number of home runs affect his median number of runs?

c. If you had to choose between the mean and median to describe Maris's home run hitting pattern, which would you use? Why?

Variability Relative to the Mean

Learning Objective: Use the concept of average deviation from the mean to estimate standard deviation from the mean.

Statisticians invented **standard deviation** to measure variability about the mean. ***The standard deviation is roughly the average distance that the data points vary from the mean.*** In this activity we will use average distance from the mean (ADM) to estimate standard deviation (SD) about the mean.

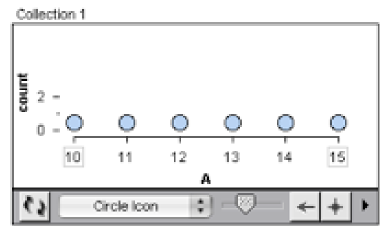
$$ADM = \frac{\sum|x - \bar{x}|}{n}$$

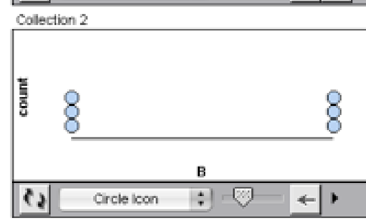
1) Each dot plot shown varies from 10 to 15 with a mean of 12.5.

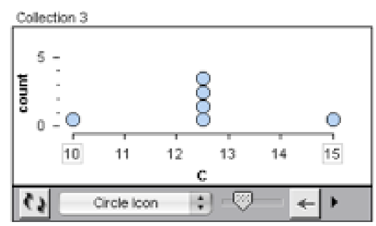
   a) Which dot plot has the least variability about the mean?

   b) Which has the most variability about the mean?

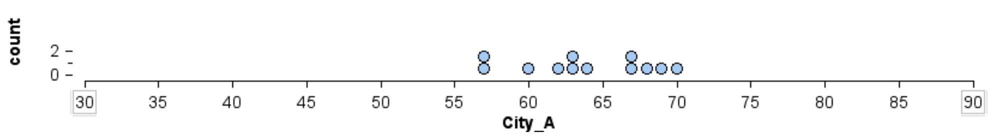
   c) Estimate the standard deviation from the mean (SD) by calculating the average deviation from the mean (ADM) for each set of data. Check to see if calculations agree with your answers above.

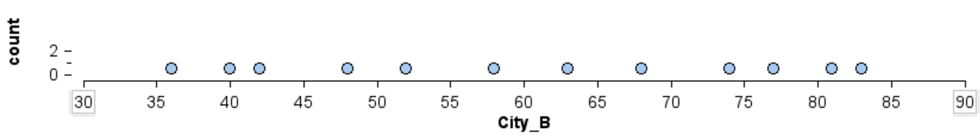






2) If we calculate the ADM for City A and City B using the average high temperatures, we get 3.6 and 14.2. Which is the ADM for San Francisco? How do you know? (See if you can answer this without doing any calculations!)

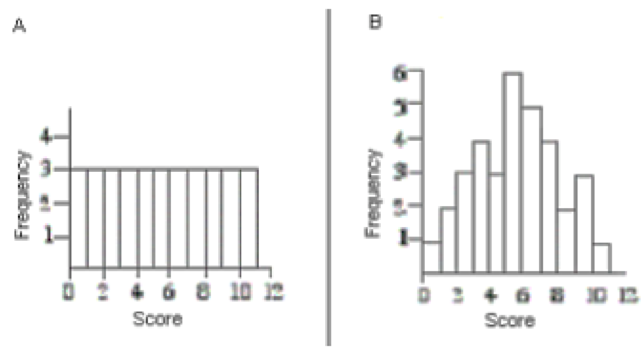




3) Which set of data will have a larger ADM? Why do you think so?

- The ages of children at a local elementary school
- The ages of people living in Pittsburg

4) Which distribution has the larger ADM? Explain how you made your decision.



5) Make up two data sets with the same mean but different ADM. Jot down a few notes about how you figured this out.

6) Make up two data sets with different means but the same ADM. Jot down a few notes about how you figured this out.

7) Make up two data sets so that one set of data has both a larger range and a smaller ADM compared to the other data set. Jot down a few notes about how you figured this out.

Measuring Variability Relative to the Mean

Learning Goal: For the distribution of a quantitative variable, describe the overall pattern (shape, center, and spread) and striking deviations from the pattern.

Specific Learning Objective: Estimate and calculate the standard deviation from the mean.

Warm-up:

1) To estimate the standard deviation from the mean, we developed the idea of average deviation from the mean. To review, find the ADM of this data set.

   {5, 6, 10, 11, 18}

2) Here is the formula for standard deviation. Compare and contrast the formula for ADM with the formula for standard deviation (SD).

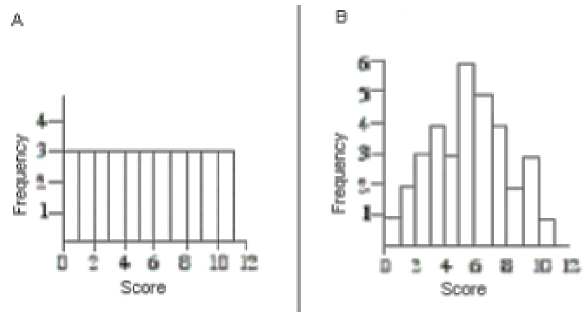
$$SD = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

3) The ADM will always be slightly smaller than the SD. Show that this is true by calculating the standard deviation for {5, 6, 10, 11, 18} and comparing the SD to the ADM you calculated in (1).

4) Which do you think will have a larger standard deviation? Why?
   a. The amount that a random sample of 30 SCC students spend per unit.

   b. The amount that a random sample of 30 college students in the U.S. spend on per unit.

5) Which distribution has the smaller standard deviation? Explain how you made your decision.

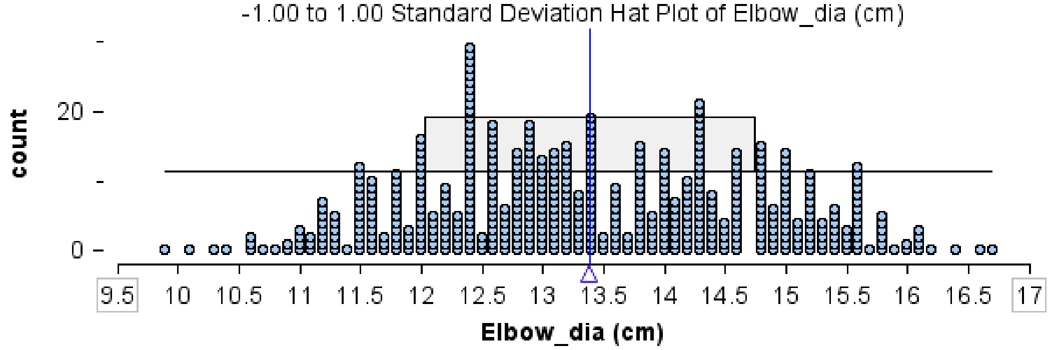


6) If the standard deviation of quiz score is zero, what do we know? Jot down a few notes to capture your thinking.

a. everyone made a 100% on the quiz
b. everyone failed the quiz
c. everyone made the same score on the quiz
d. it is impossible to tell

7) Here is a dot plot of elbow girth measurements for 507 adults with a standard deviation hat plot. The standard deviation hat plot consists of a box around the mean that captures all of the data that is within one standard deviation of the mean.

In other words, the left edge of the box is Mean – SD; the right edge of the box is Mean + SD.



The mean elbow diameter is about 13.4 cm. Which of the following choices is the most reasonable estimate for the standard deviation? Why?

a. about 2.8cm
b. about 1.4cm
c. about 7cm

8) Describe in words what the standard deviation measures. (Think about how we have been estimating it in previous activities.)

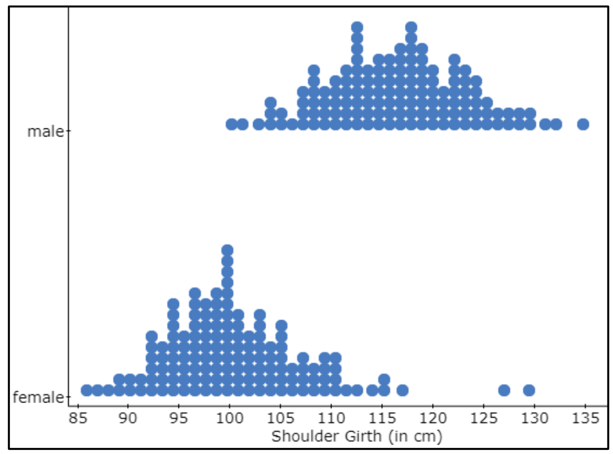
Measuring Variability Relative to the Mean

Learning Goal: For the distribution of a quantitative variable, describe the overall pattern (shape, center, and spread) and striking deviations from the pattern.

Specific Learning Objective: Use the mean and standard deviation to create intervals of typical measurements.

When we compare two distributions, we want to describe an interval of typical data values. If we use the mean as an average, we can use the interval within one SD of the mean: (mean – SD, mean + SD).

Example: Here we have data on shoulder girth measurements for 247 men and 260 women who exercise regularly.



1) For this sample of men and women, would you argue that men tend to have shoulder girths that are larger than women? Why or why not?

2) The mean shoulder girth for men is 116.5 cm vs. 100.3 cm for women. This tells us that on average men have larger shoulders. But there is variability in the data. To take variability into account, we can give an interval of typical measurements for each gender, instead of relying on a comparison of just the mean.

   Typical men have shoulder measurements within one standard deviation of the mean. The standard deviation for men is 6.5 cm.

   Mean – SD = 116.5 – 6.5 = 110

   Mean + SD = 116.5 + 6.5 = 123

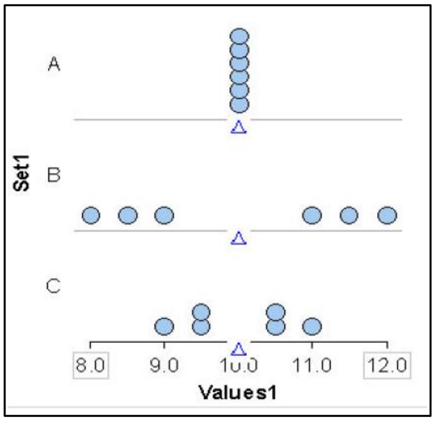
   Based on this data, typical men have a shoulder girth between 110 and 123 cm.

   The standard deviation in women's shoulder girths is also 6.5cm. Find an interval of typical shoulder measurements for women using the mean and SD:

3) Do the intervals of typical measurements overlap? How does this observation support your answer to (1)?

1.6 Homework

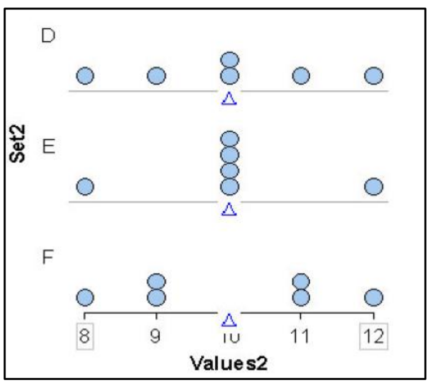
1. Below is a set of three distributions labeled A, B, and C. Assume the scales to be the same on each horizontal axis.



a. Use your own visual sense to order this set of distributions from least amount of variability to most amount of variability.

b. Explain what you observed or did to order this set of distributions from least amount of variability to most amount of variability.

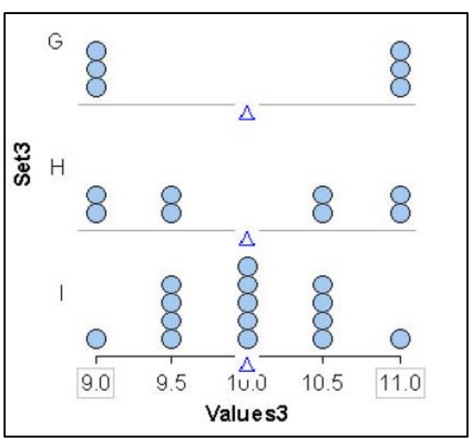
2. Below is another set of three distributions labeled D, E, and F. Assume the scales to be the same on each horizontal axis.



a. Use your own visual sense to order this set of distributions from least amount of variability to most amount of variability.

b. Explain what you observed or did to order this set of distributions from least amount of variability to most amount of variability.

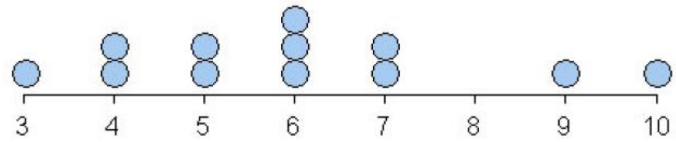
3. Below is another set of three distributions labeled G, H, and I. Assume the scales to be the same on each horizontal axis.



a. Use your own visual sense to order this set of distributions from lest amount of variability to most amount of variability.

b. Explain what you observed or did to order this set of distributions from least amount of variability to most amount of variability.

4.

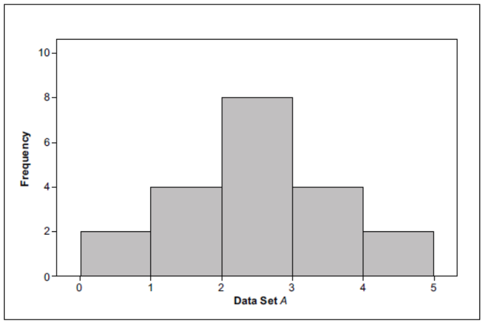


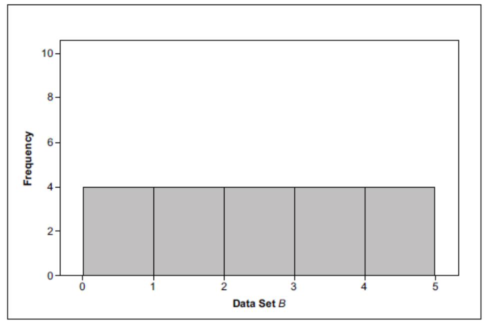
a. Find the ADM and then the standard deviation for the following set of data.

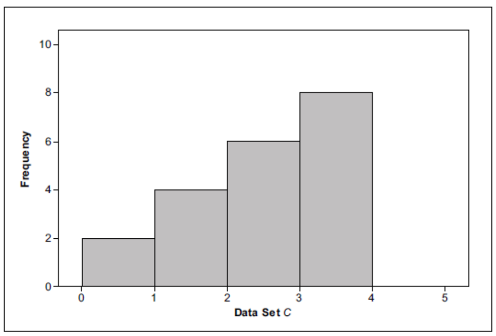
b. Add 5 to each data point, and compute the standard deviation. How does the standard deviation change? Why do you think this is?

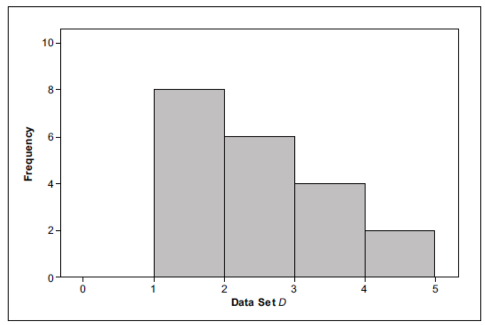
5. You want to determine how variable the high temperature is in two exotic isles. You note that in Jamaica the weather has been 87 and then 93 for two days. But for the Bahamas you have more information. The temperatures are 84, 86, 87, 88, 93, 94, and then 94 for seven days. Use the range and the standard deviation to determine variability of weather of the two islands. Decide which island has more variable weather.

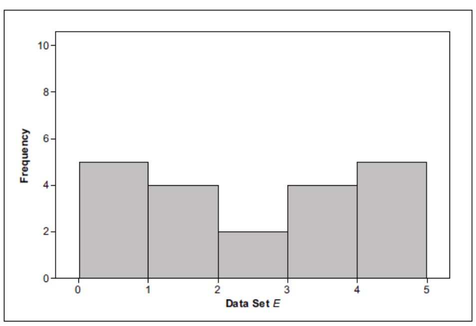
6. Data Sets *A* – *E* displayed graphically in Figures 6.5 – 6.9 all have means of 2.5. So, the mean does not provide any information that could be used to distinguish one data set from another. In parts a – c, determine from the histograms which of the two data sets has the larger standard deviation, or if the standard deviations are about the same. In each case, give a justification of your answer.











a. Data Set A and Data Set B.

b. Data Set C and Data Set D.

c. Data Set D and Data Set E.

d. The standard deviations of the Data Sets A – E are given below in random order. Match each standard deviation with its data set.

1.589 1.026 1.124 1.026 1.451