STEM ONLY | AP Statistics Summer Assignment

IMPORTANT: Read this page first...

INSTRUCTIONS:

I. Complete the guided notes
- A. One of the most important things to do when taking the guided notes is to describe in detail. Reading the book is a strongly recommended source. The book often uses examples to explain certain concepts. A copy of Chapter 1 has been linked below.
- B. The linked videos are optional but are a great resource.

Topics in the Summer Packet
- I. Exploring One-Variable Data
  - A. Chapter 1 - Data Analysis
    1. Analyzing Categorical Data
    2. Displaying Quantitative Data with Graphs
    3. Describing Quantitative Data with Numbers
    4. Vocabulary

Resources
- I. Links to videos in each vocabulary section
- II. Textbook -
  - A. Chapter 1 Data Analysis
  - B. Full Text: The Practice of Statistics 6th Edition by Starnes.pdf

PACING:

You should pace yourself to work on this assignment at least a few hours a week leading up to the start of school in September. If you complete the packet at the end of June or early in July, it will not be very helpful in preparation for the start of school. Also, it will not be helpful if you try to complete the entire packet a night or two before school starts. Pace yourself by setting a calendar reminder and scheduling blocks of time to focus on this assignment as you prepare to return to school in September.

GRADING:

· On the first day of school, your math teacher will check for full completion of this Summer Assignment and the supporting work for your responses (no work = no credit). This part will be weighted at 50% - this is the grade that represents your effort and following of directions.
· Your teacher will then review the assignment and provide remediation as needed.
· Upon completion of your teacher's review, you will be given an assessment (a "test") based on the topics covered in this assignment. This assessment will be weighted at 50% - this is the grade that represents your mastery of the skills.
· The two weighted scores combined will count as one Project grade for the 1st trimester.
· Acceptance of late assignments will be limited and subject to point deductions.

We are looking forward to meeting you in September. Go Bulldogs!

Introduction: *Describe/Define the following*	
Statistics	
Individual	
Variable	
Categorical Variable	
Quantitative Variable	
Discrete Variable	
Continuous Variable	
Distribution	

1. Your principal gives your class an assignment to collect data from the student body. The data you will collect are ages, gender, grade level, GPA, number of siblings, and the name of the students' guidance counselors.
   a. Who are the individuals and the variables provided in the data?

   b. Identify the variables as categorical or quantitative. Then, identify the quantitative variables as discrete or continuous.

Categorical	Quantitative	
	Discrete	Continuous

2. When it comes to quantitative data, it should make sense to calculate the _____ of the variable. However, not all numerical data is quantitative.

3. Give 5 examples of numerical data where it does not make sense to find the mean of the variable.

1.1 Analyzing Categorical Data: *Describe/Define the following*	
Resources: • Make a Relative Frequency Bar Graph • Misleading Graphs • Marginal and Joint Relative Frequency • Conditional Relative Frequency	
Frequency Table	
Relative Frequency Table	
Bar Graph	
Pie Chart	
Two-Way Table Also known as: Contingency Table	
Marginal Relative Frequency	
Joint Relative Frequency	
Conditional Relative Frequency	
Side-by-Side Graph	

Segmented Bar Graph	
Mosaic Plot	
Association	

1. An article in the *Journal of the American Medical Association* reports the results of a study designed to see if the herb St. John's wort is effective in treating moderately severe cases of depression. The study involved 338 patients who were being treated for major depression. The subjects were randomly assigned to receive one of three treatments: St. John's wort, Zoloft (a prescription drug), or placebo (an inactive treatment) for an 8-week period. The two-way table summarizes the data from the experiment.

|		Treatment		
		St. John's wort	Zoloft	Placebo
Change in Depression	Full Response	27	27	37
	Partial Response	16	26	13
	No Response	70	56	66

a. What proportion of subjects in the study were randomly assigned to take St. John's wort? Explain why this value makes sense.

b. Find the distribution of change in depression for subjects assigned to the placebo in this study using relative frequencies.

c. What percent of subjects took Zoloft and showed a full response?

2. The American Statistical Association sponsors a web-based project that collects data about primary and secondary school students using surveys. We used the site's Random Sampler to choose 40 high school students who completed the survey in a recent year. One of the questions asked:
"Which would you prefer to be? Select one"
   a. Rich
   b. Happy
   c. Famous
   d. Healthy

Here are the responses from the 40 randomly selected students:

Famous	Healthy	Healthy	Famous	Rich	Famous	Happy	Famous
Happy	Famous	Happy	Happy	Happy	Happy	Rich	Rich
Happy	Happy	Happy	Happy	Happy	Happy	Happy	Happy
Rich	Healthy	Rich	Famous	Happy	Rich	Healthy	Happy
Happy	Happy	Happy	Happy	Rich	Famous	Rich	Happy

Create a frequency distribution table	Make a relative frequency bar graph of the data	Describe what you see.
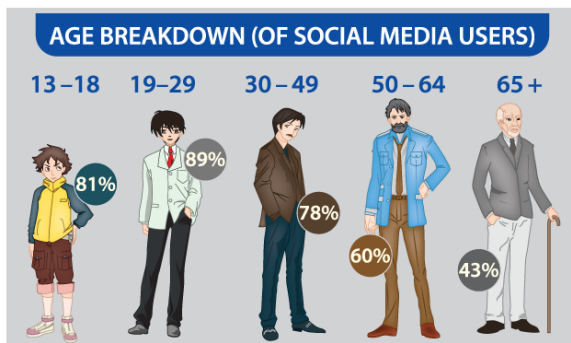
AGE BREAKDOWN (OF SOCIAL MEDIA USERS)
13–18 19–29 30–49 50–64 65+
81% 89% 78% 60% 43%

3. The Pew Research Center surveyed a random sample of U.S. teens and adults about their use of social media. The following pictograph displays some results. Explain how this graph is misleading.

1.2 Displaying Quantitative Data with Graphs: *Describe/Define the following*

Resources:
- Making a Dotplot
- Describing a Dotplot
- Making a stemplot
- Constructing a Histogram

|	
How to make a Dotplot?	
Symmetric	
Skewed to the right	
Skewed to the left	
How to describe the distribution of a quantitative variable?	
How to make a Stemplot	
Histogram	

1. Knoebels Amusement Park in Elysburg, Pennsylvania, has earned acclaim for being an affordable, family-friendly entertainment venue. Knoebels does not charge for general admission or parking, but it does charge customers for each ride they take. How much do the rides cost at Knoebels? The table shows the cost for each ride in a sample of 22 rides in a recent year.

Merry Mixer	$1.50	Whipper	$1.25	Paradrop	$1.25	Flying Turns	$3.00	Twister	$2.50
Italian Trapeze	$1.50	Skooters	$1.75	The Phoenix	$2.50	Flyer	$1.50	Cosmo	$1.75
Galleon	$1.50	Ribbit	$1.25	Gasoline Alley	$1.75	Haunted Mansion	$1.75	Paratroop	$1.50
Satellite	$1.50	Roundup	$1.50	Looper	$1.75	Stratos Fear	$2.00	Downdraft	$1.50

Make a dotplot of the data.	Describe the shape of the distribution. When describing shapes of distribution: ALWAYS include in your explanation if it is skewed left/right or roughly symmetric

2. For a statistics class project, Jonathan and Crystal hosted an ice-cream-eating contest. Each student in the contest was given a small cup of ice cream and instructed to eat it as fast as possible. Jonathan and Crystal then recorded each contestant's gender and time (in seconds), as shown in the dotplots. Compare the distributions of eating times for males and females.
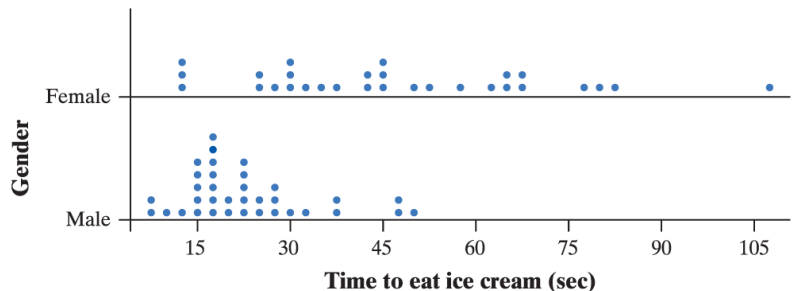
Starnes & Tabor, *The Practice of Statistics*, 6e, © 2018 Bedford, Freeman & Worth High School Publishers

3. Data is displayed for students' scores on the Chapter 1 AP Statistics Exam.
   15 43 43 48 48 50 53 53 55 55 55 60 60 63 63 63 65 68 68 68 70 75 78 85 88 88 90 93 95 95
   a. Make a stemplot of these data and describe the shape of the distribution.
   b. Are there any possible outliers?
   c. What percent of the students scored between 60 and 75?

4. The data here displays the top 18 Countries for medals won in the 2020 Summer Olympics. Make a frequency histogram to display the data and describe the shape of the distribution. Explain your reasoning for the number of bins you chose to make.

Country	Medal Count	Country	Medal Count	Country	Medal Count
France	33	Hungary	20	Ukraine	19
New Zealand	20	Canada	24	Australia	46
ROC	71	Brazil	21	Cuba	15
Great Britain	65	People's Republic of China	88	Italy	40
Republic of Korea	20	United States of America	113	Germany	37
Spain	17	Japan	58	Netherlands	36

1.3 Describing Quantitative Data with Numbers: *Describe/Define the following*

Resources:
- Variance and standard deviation
- Median and quartiles
- Boxplots and five-number summary
- Comparing boxplots

Mean	
Statistic	
Parameter	
Resistant	
Median	
Range	
Standard Deviation	
Variance	
Quartiles	
First Quartile	
Third Quartile	
Interquartile Range (IQR)	

Five-Number Summary	
Boxplot	

1. Once again, refer to the student data for scores from the Chapter 1 AP Statistics Exam.
   15 43 43 48 48 50 53 53 55 55 55 60 60 63 63 63 65 68 68 68 70 75 78 85 88 88 90 93 95 95
   a. Construct a boxplot from the data.
   b. Describe the center of the data. When describing the center of a dataset, you may want to tie in mean and median.
   c. Describe the spread of the data. When describing the spread of a dataset, you may want to tie in range and IQR.

2. The American Statistical Association (www.amstat.org) has an Instagram account (@amstatnews). Here are the number of Instagram likes for 10 posts selected at random:

16	4	8	7	8	6	15	2	9	5

   a. Calculate the mean number of Instagram likes for these 10 posts
   b. Calculate the variance and standard deviation of Instagram likes for these 10 posts

3. The American Statistical Association (ASA) has an Instagram account (@amstatnews). Bedford, Freeman & Worth (BFW), the publisher of this textbook, also has an Instagram account (@bfwhighschool). Here are the number of Instagram likes for 10 randomly selected posts from each account, along with parallel boxplots and numerical summaries of the data:

| ASA | 16 | 4 | 8 | 7 | 8 | 6 | 15 | 2 | 9 | 5 |
| BFW | 19 | 21 | 18 | 14 | 15 | 18 | 8 | 16 | 11 | 38 |

Group name	n	mean	SD	min	Q_1	med	Q_3	max
1: ASA	10	8	4.472	2	5	7.5	9	16
2: BFW	10	17.8	8.08	8	14	17	19	38
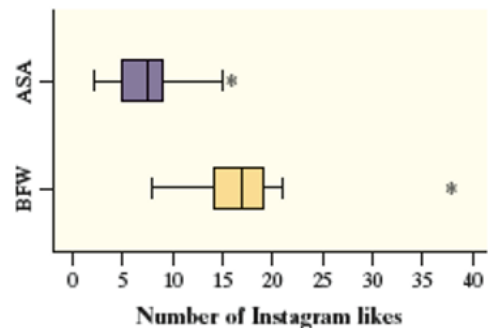
   a. Compare the distributions of the number of Instagram likes for ASA and BFW.
   b. Write the five-number summary for both the ASA dataset and the BFW dataset.