Welcome future statisticians! This summer you will start a wonderful journey exploring data analysis. The purpose of this assignment is to make you comfortable with the six primary data displays used during this AP Statistics class and to review skills that we expect you to have before starting the course.

***The summer assignment is composed of two parts.*** Start the summer assignment early to allow for time to receive clarification (if necessary) and to complete it by AUGUST 7th.

* + **Reading and Vocabulary:** You may use free online Statistics tutoring sites that will provide information on variables and data displays and assist you in completing the vocabulary list (see page 2 and 3). An excellent site to utilize is [www.stattrek.com](http://www.stattrek.com/).

|  |  |
| --- | --- |
| **General Topic: The Basics** | |
| Subtopics: | Variables |
|  | Populations vs. Samples |
|  | Central tendency |
| **General Topic: Charts and Graphs** | |
| Subtopics: | Patterns in data |
|  | Dotplots |
|  | Histograms |
|  | Stemplots |
|  | Boxplots |
|  | Cumulative plots |
|  | Comparing data sets |
| **General Topic: Categorical Data** | |
| Subtopics: | One-way tables |
|  | Two-way tables |

1. **Practice Problems:** After reading all the material above you should be able to complete the questions in the remaining pages of this packet. You may do so in the spaces provided.

**Materials:** You will need a **graphing calculator** (recommended for homework and for the AP Statistics exam) the textbook will explains how to use the TI-84 Plus.

**Resources/supplies:** Our textbook is **Stats Modeling the World (AP Edition)**. This textbook is aligned to the AP Statistics curriculum and the sample problems and activities will prepare you well for the AP Statistics exam.

**Caution about Academic Integrity:** DO NOT copy the answers to any assignment from sources you find on the internet or the textbook solutions. Answers you find on the internet are sometimes incorrect and will often be stated using terminology and techniques that are not specific to the current AP Statistics curriculum. In addition, this will make it easy for me to identify work that is not your own.

# Part I: Vocabulary List

Please define each of the following terms from the information on the stattrek website. When asked to provide an example of the word, provide a unique example of the word NOT given on the website.

1. Categorical Variables

Example:

1. Quantitative Variables

Example:

1. Discrete
2. Continuous
3. Univariate Data

1. Bivariate Data

1. Median

1. Mean
   1. formula

1. Population

Example:

1. Sample

Example:

1. Center

1. Variance (formerly known as *Spread*)
   1. formula
2. Standard Deviation
   1. formula

1. Symmetry

1. Unimodal and Bimodal

1. Skewness

Sketch Skewed left: Sketch Skewed right:

1. Uniform

1. Gaps

1. Outliers

1. Dot plots
2. Scatterplots

1. Stem plots

1. Boxplots

1. Quartiles
2. Range
   1. formula

1. Interquartile Range (IQR)

* 1. formula

1. 5-Number Summary

1. Difference between a frequency table and relative frequency table

1. Parameter

1. Statistic

1. Marginal Distribution

1. Conditional Distribution

1. Segmented Bar Chart

1. What are the W’s of data

16.

Part 2: Practice Problems

**WHAT ARE YOU?**

Determine if the variables listed below are quantitative data (Q), categorical data (C), a population (Pop), a sample (S), a parameter (Par), or a statistic (Stat). The word “true” means population.

|  |  |  |
| --- | --- | --- |
| Variable | Q, C, Pop, S, Par, or Stat | Reasoning |
| 1) Types of Dog Breeds |  |  |
| 2) Names of Students in a Class |  |  |
| 3) True mean height of everyone living in California |  |  |
| 4) Students in a School |  |  |
| 5) Heights of Students in a Class |  |  |
| 6) Daily Temperature in a Given  Month |  |  |
| 7) 50 Dogs in a City |  |  |
| 8) Proportion of Test Grades for 30  Students in a Class |  |  |
| 9) Favorite Breakfast Cereal |  |  |
| 10) Mean Amount of Liquid in 100  Selected Bottles |  |  |
| 11) 30 oz. of a Certain Juice |  |  |
| 12) 80 Families in a Country |  |  |
| 13) True Mean Number of Family  Members in Wisconsin |  |  |
| 14) Colors of Shirts |  |  |
| 15) True Proportion of Students  Wearing Glasses in a School |  |  |

1. Define and provide the formula for
   * Standard Deviation
   * Variance
   * Standard Score (z-score)
2. HOME RUNS

Determine the statistics from the data below on the number of homeruns Mark McGuire hit in each season from 1982 – 2001.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 70 | 52 | 22 | 49 | 3 | 32 | 58 | 39 |
| 39 | 65 | 42 | 29 | 9 | 32 | 9 | 33 |

|  |  |
| --- | --- |
| Mean |  |
| Standard Deviation |  |
| Minimum |  |
| Q1 |  |
| Median |  |
| Q3 |  |
| Maximum |  |
| Range |  |
| IQR |  |

1. Provide a sketch of the following
   * Unimodal
   * Bimodal
   * Skewed Left
   * Skewed Right
2. Draw a dot plot of the data. Then calculate the sample mean, median, and standard deviation. 3, 3, 4, 7, 7, 1, 5, 9, 8, 7, 9, 4, 1, 2, 6, 1, 9, 6, 6, 4

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

1. Calculate the sample mean, standard deviation, the five-number summary and IQR. Then draw a box plot of the data.

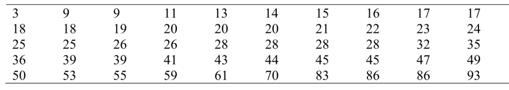
19, 22, 23, 23, 23, 26, 26, 27, 28, 29, 29, 31, 32

1. Make a pie chart for each of the following:

Diego loves to play video games. His parents added a feature to his gaming system that tracks the total time on each game and send them an email every week. In the past week he played a total of 470 minutes. The email reported that Fortnite was played 180 minutes, PUBG was played 40 minutes, 38 minutes were used in The Show, Madden 2017 was played 94 minutes, God of War was played 91 minutes, and Rock Band was played 27 minutes.

1. Shopping Spree!

A Market consultant observed 50 consecutive shoppers at a supermarket. One variable of interest was how much each shopper spent in the store. Here are the data (rounded to the nearest dollar) arranged in increasing order.



* 1. Make a stemplot using ten of dollars as the stem and dollars as the leaves. Make sure you include appropriate labels, title, and key.

23. WHERE DO OLDER FOLKS LIVE?

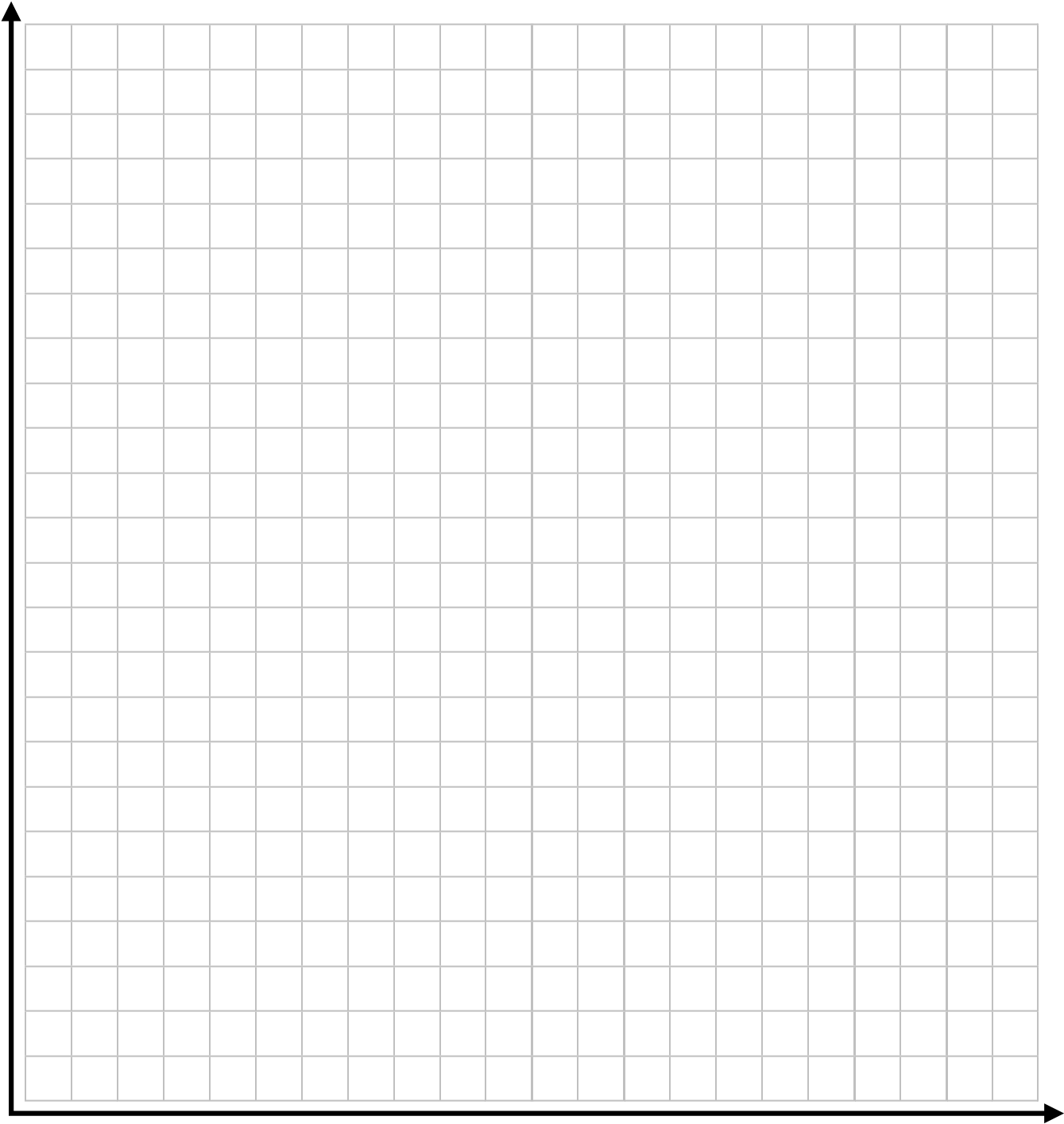
This table gives the percentage of residents aged 65 or older in each of the 50 states.

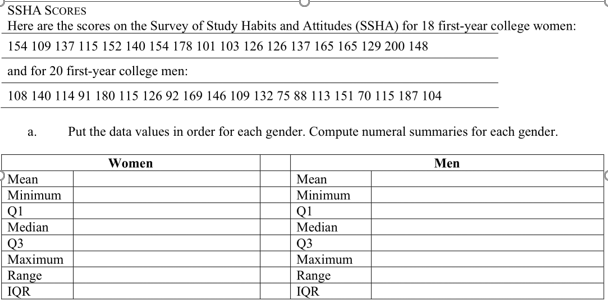
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| State | Percent | State | Percent | State | Percent |
| Alabama  Alaska  Arizona  Arkansas  California  Colorado  Connecticut  Delaware  Florida  Georgia  Hawaii  Idaho  Illinois  Indiana  Iowa  Kansas  Kentucky | 13.1  5.5  13.2  14.3  11.1  10.1  14.3  13.0  18.3  9.9  13.3  11.3  12.4  12.5  15.1  13.5  12.5 | Louisiana  Maine  Maryland  Massachusetts  Michigan  Minnesota  Mississippi  Missouri  Montana  Nebraska  Nevada  New Hampshire  New Jersey  New Mexico  New York  North Carolina  North Dakota | 11.5  14.1  11.5  14.0  12.5  12.3  12.2  13.7  13.3  13.8  11.5  12.0  13.6  11.4  13.3  12.5  14.4 | Ohio  Oklahoma  Oregon  Pennsylvania  Rhode Island  South Carolina  South Dakota  Tennessee  Texas  Utah  Vermont  Virginia  Washington  West Virginia  Wisconsin  Wyoming | 13.4  13.4  13.2  15.9  15.6  12.2  14.3  12.5  10.1  8.8  12.3  11.3  11.5  15.2  13.2  11.5 |

Histograms are a way to display groups of quantitative data into bins (the bars). These bins have the same width and scale and are touching because the number line is continuous. To make a histogram you must first decide on an appropriate bin width and count how many observations are in each bin. The bins for percentage of residents aged 65 or older have been started below for you.

|  |  |
| --- | --- |
| **Bin Widths** | **Frequency** |
| 4 to < 6 | 1 |
| 6 to < 8 |  |
| 8 to < 10 |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

(23. is continued on next page)





1. Using the Minimum, Q1, Median, Q3, and Maximum from each gender, make parallel boxplots to compare the distributions.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

1. New Grading Policy

A new grading policy has been proposed by the dean of the College of Education for all education majors. All faculty and students in the college were asked to give their opinions about the new policy. The results are given below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Favor | Neutral | Opposed | Row Total |
| Students | 353 | 75 | 191 | 619 |
| Faculty | 11 | 5 | 18 | 34 |
| Column Total | 364 | 80 | 209 | 653 |

1. State the variables and if they are categorical or quantitative.

1. What percent of responses were from students favoring the policy change? \_\_\_\_\_\_

What percent of students favored the policy change? \_\_\_\_\_\_\_\_\_\_

What percent favoring the policy change were students? \_\_\_\_\_\_\_\_

1. What is the marginal distribution of the grading policy change?

1. What is the distribution of the grading policy among just students?

1. What is the distribution of the grading policy among just faculty?

## DESCRIBING GRAPHS

Describe each distribution by its shape, center, variance, and outliers (if they exist).

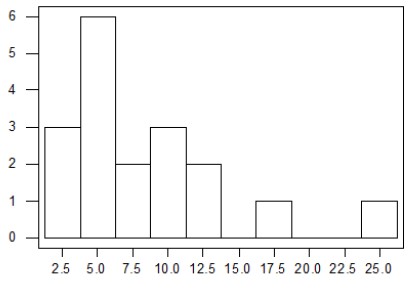
Examples of  **S**hape: symmetric, skewed right, skewed left

**C**enter: approximate where the center of the data is (be sure to indicate if you used a mean or median)

**V**ariance(Spread): Calculate the range of the data, IQR, or standard deviation (be sure to indicate which one you calculated)

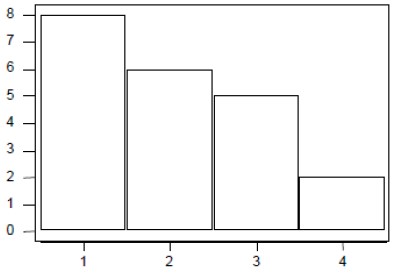
**O**utliers: approximate the location of extreme values, if they exist

(26. is continued on next page)

A)

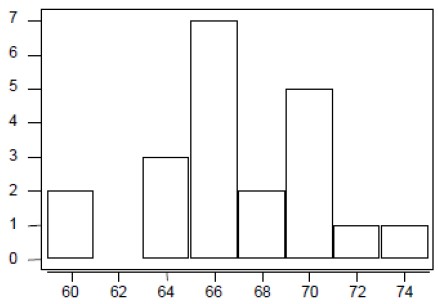
Shape \_\_\_\_\_\_\_\_\_\_ Center \_\_\_\_\_\_\_\_\_\_

Outliers \_\_\_\_\_\_\_\_\_\_ Variance \_\_\_\_\_\_\_\_\_\_

B)

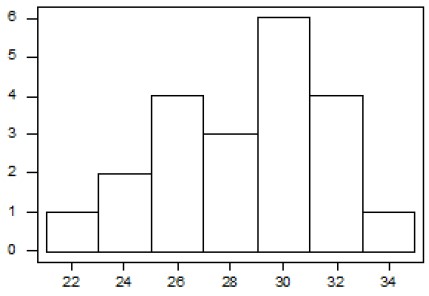
Shape \_\_\_\_\_\_\_\_\_\_ Center \_\_\_\_\_\_\_\_\_\_

Outliers \_\_\_\_\_\_\_\_\_\_ Variance \_\_\_\_\_\_\_\_\_\_

C)

Shape \_\_\_\_\_\_\_\_\_\_ Center \_\_\_\_\_\_\_\_\_\_

Outliers \_\_\_\_\_\_\_\_\_\_ Variance \_\_\_\_\_\_\_\_\_\_

D)

Shape \_\_\_\_\_\_\_\_\_\_ Center \_\_\_\_\_\_\_\_\_\_

Outliers \_\_\_\_\_\_\_\_\_\_ Variance \_\_\_\_\_\_\_\_\_\_

**ALGEBRA PAGE!**

The prerequisite for AP Statistics is Algebra II. You will find very much equation solving in this course, but some quick review of Algebra I and Algebra II content will be helpful.

Calculator function: normCdf(𝑙𝑜𝑤𝑒𝑟, 𝑢𝑝𝑝𝑒𝑟, 𝜇, 𝜎).



## Here is a formula that is used often in AP Statistics:

## 

1. If z = 2.5, x = 102, and *x* =100. What is s? Show your work.

1. If z = -3.35, x = 60, and s = 4, what is *x* ? Show your work.

1. Solve 0.05 =1.96 for n.

2

0.5

*n*

**Read the following…**

*“Teen Automobile Crash Rates Are Higher When School Starts Earlier”*

*ScienceDaily (June 10, 2010) — Earlier school start times are associated with increased teenage car crash rates, according to a research abstract presented June 9, 2010, in San Antonio, Texas, at SLEEP 2010, the 24th annual meeting of the Associated Professional Sleep Societies LLC.*

*Results indicate that in 2008 the teen crash rate was about 41 percent higher in Virginia Beach, Va., where high school classes began at 7:20 a.m., than in adjacent Chesapeake, Va., where classes started more than an hour later at 8:40 a.m. There were 65.4 automobile crashes for every 1,000 teen drivers in Virginia Beach, and 46.2 crashes for every 1,000 teen drivers in Chesapeake.*

*"We were concerned that Virginia Beach teens might be sleep restricted due to their early rise times and that this could eventuate in an increased crash rate," said lead author Robert Vorona, MD, associate professor of internal medicine at Eastern Virginia Medical School in Norfolk, Va. "The study supported our hypothesis, but it is important to note that this is an association study and does not prove cause and effect."*

*The study involved data provided by the Virginia Department of Motor Vehicles. In Virginia Beach there were 12,916 drivers between 16 and 18 years of age in 2008, and these teen drivers were involved in 850 crashes. In Chesapeake there were 8,459 teen drivers and 394 automobile accidents. The researchers report that the two adjoining cities have similar demographics, including racial composition and per-capita income.*

4. Answer the following questions regarding the above excerpt:

1. *Who* is being studied?
2. *What* about those individuals is being recorded / analyzed (i.e. what are the variables?)? Do you think the variables are categorical or quantitative in nature?

1. *When* was the data collected?

1. *Where* was the data collected (more accurately: what geographical area is associated with the data)?

1. *Why* do you think this data was collected and analyzed?

1. *How* was the data collected and analyzed? In other words, what methods were used?

1. Whydo you think the authors of the study mentioned that “it is important to note that this is an association study and does not prove cause and effect?”