Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**How and Why We Analyze Data**

**What Is Data Analysis?**

Data analysis is the process of interpreting the meaning of the data by organizing the data and displaying it in the form of a table, bar chart, line graph, or other representation. The process involves looking for patterns—similarities, disparities, trends, and other relationships—and thinking about what these patterns might mean.

**How We Analyze Data:**

*When analyzing data, students should ask the following questions:*

* Is my data source valid?
* What are the different categories of data?
* What problem am I trying to solve?
* Is there missing information?
* Is there extra, unnecessary data?
* Are there places where the data are concentrated or clumped?
* What trends are visible in the data?
* Are there values for which there are no data?
* Are there data points that have unusual values?
* Are these values consistent with my predictions?
* Do I have enough background information to analyze this data?
* What level of accuracy is required in my analysis?
* What is the range, mean, median, mode, and so on?
* Why might there be an increase/decrease in the data?
* How might this data change if...? (This may be a change in history, market value, audience questioned, etc.)
* What is the most appropriate way to summarize my findings?
* Based on this data, what will happen in the future? Can I change the future using this data?
* For what other data sample might I use a similar graph?
* What patterns do I see?
* What does this graph say?
* Who could use this data? How could they use it?
* Why is this data shown in a line graph, etc…?

The process of collecting, organizing, and analyzing data is not always a simple, sequential process; sometimes a preliminary analysis of a data set may prompt us to look at the data in another way, or even to go back and collect additional data to test an emerging hypothesis. For example, a student could survey their classmates on how they are transported to school (such as by car, by bus, by foot, or another way), and then display the data in a circle graph.

After analyzing the data in this graph, the student might look at the data in a different way. The student might be interested in finding out more about people who are transported to school by car. Why do they ride in a car to school? Are they on a bus route? Do they carpool with other students? Are they close enough to school to walk, but choose to ride? Is the neighborhood between home and school too dangerous to walk through? Do the people who walk sometimes ride in a car, also? They might discover that most students in the "other" category ride their bikes to school, and decide to create an additional category.

Students should be able to look at graphical displays and describe them by identifying aspects such as the greatest value, the least value, and the relationship of one data point to another. Students need to be able to summarize or characterize a data set in greater depth by determining the range and two measures of center, the mode and median. Lastly student should find the third measure of center, the mean, and also determine quartiles, identify outliers, and, for scatterplots, calculate a line or curve of best fit and describe any resulting correlation

Students should always be thinking about the meaning of the data and ***why it is important***.

**Why Is It Important?**

The ability to make inferences and predictions based on data is a critical skill for all students to develop. In studying data and statistics, students can also learn that solutions to some problems depend on assumptions and have some degree of uncertainty. Data analysis is crucial to the development of theories and new ideas. By paying close attention to patterns, the stories behind outliers, relationships between and among data sets, and the external factors that may have affected the data, we may come to have a deeper understanding of the crucial distinction between theory and evidence. All of this together ***helps us solve the problems of yesterday, today, and tomorrow!***

**Where do I begin?**

1. Look at the table of contents in the 2011 Missouri State Analysis Center report and decide what subset of the data you would like to study. Some examples of crash related problem areas you could choose to analyze are:
* Seat belts
* Alcohol
* Age of driver
* Urban vs. rural
* Holiday driving
* Day of week
* Time of day
* Texting
* Speeding
* Road conditions
* Males vs females
1. Find two or more sets of data in the report that you have observed a noticeable pattern or trend that might lead you to make a conjecture about safe driving. Remember our essential question, “How can we be safe?”.
2. Now you should summarize your data by finding the measures of center (mean, median, and mode), identifying range, outliers, quartiles, and so on.

Mean = \_\_\_\_\_\_\_\_\_ Median = \_\_\_\_\_\_\_\_\_ Mode = \_\_\_\_\_\_\_\_\_\_

Range = \_\_\_\_\_\_\_\_\_\_\_ Outlier = \_\_\_\_\_\_\_\_\_ Upper Quartile = \_\_\_\_\_\_\_\_

Lower Quartile = \_\_\_\_\_\_\_\_\_ Upper extreme = \_\_\_\_\_\_\_\_\_ Lower extreme = \_\_\_\_\_\_\_\_

1. Create a scatterplot of the data and analyze the data by answering these questions:
* What two variables are being compared?
* What is the relationship between the two variables?
* Is there a positive, negative, or no correlation?
* Why do you think there is a relationship?
* Was this a valid sample of the population?
* Would the data be the same in a state other than Missouri?
* What can you infer about the data?
* What conclusions can be drawn about the data? Why?
1. Now decide what other type of graph would show a different aspect of your data and create it.
	* + Line graph: best to show how data changes over time.
		+ Circle graph: best to show how parts of a data set compare to the whole set.
		+ Stem-and-leaf plot: best to show your data in order from least to greatest.
		+ Box-and-Whisker plot: best to show how the data is distributed.
		+ Bar Graph: best to show data in categories.
		+ Histogram: best to show data divided into intervals.
		+ Double Line and Bar graphs: best to compare multiple groups.
2. Now put it all together. Can you make a conjecture about an aspect of safe driving that can be supported and justified by the data. If so, then write your conjecture here.
3. Remember your data was from the year 2011 and only included crash data for the state of Missouri. Do you think your conjecture about safe driving could be generalized for all states? Why or why not?
4. Would the crash data collected from Illinois look different from Missouri? What factors might contribute to a difference? Explain.
5. Would the crash data collected in 2005 look any different from 2011? What factors might contribute to a difference? Explain.
6. In what ways could there be bias in the reporting of this data? Explain.
7. Would you be able to verbally defend your conjecture using the data to back you up? Write your defense here.