AP STATISTICS

SUMMER ASSIGNMENT

My Dear AP Students

I am excited that you are enrolled for this class and I hope that you will find it to be challenging and interesting as the applications of statistics are seen everywhere in our world. You will have an opportunity to analyze and interpret data that you encounter every day.

The purpose of your summer assignment is to get some basic review and terminology out of the way before class begins. Because this is a college level class and the AP Exan1 is scheduled for the beginning of May we need to start into the material as quickly as we can.

This is a challenging class and this assignment will allow you to become familiar with some terminology and basic information about statistics. All material will be checked and graded on the first day of class.

You will need to purchase a TI-84 calculator. You will need this in class EVERYDAY! It will be your biggest tool in statistics. If this is a problem, let me know BEFORE class begins. Some of the summer assignment will require you to use your TI-84. If you have difficulty following directions, use the internet to help you find a tutorial.

Please email me if you have any difficulty with the assignment or have any questions.
jpmeister@henrico.k12.va.us These graphs and data will be used the first week of school to continue the concepts for chapter 1.

Enjoy your- summer but don't forget about your summer work!! I will see you in September.

(DATA SETS ARE ALL INCLUDED AT THE BACK OF THIS PACKET)

ASSIGNMENT 1: Read the Transcripts about statistics and data.

1. Write the definitions for the words listed below. (Use the transcripts or goole) (20 points)
   a. Descriptive Statistics:

   b. Inferential Statistics:

   c. Data organization and analysis:

   d. Exploratory Data Analysis:

   e. Probability-based inference:

   f. Data:

   g. Numerical Data

   h. Categorical Data:

   i. Continuous Data:

   j. Discrete Data:

2. Do: Identifying Types of Data and Statistics Worksheet

ASSIGNMENT 2: Use the data for the 2009 Tampa Bay Rays and the 2009 Boston Red Sox salaries.

1. Calculate the following for each group. Show the work on the actual data page for finding the quartiles and median. (36 points)

	Mean	Median	Q1	Q3	IQR	Range
Tampa Rays						
Red Sox						

Notes:

- Quartile 1 is the median of the first half. (Do not include the actual median).
- Quartile 3 is the median of the second half. (Do not include the actual median).
- The median is the same thing as Quartile 2.
- InterQuartile Range (IQR) is Q3-Q1 and this is the range of the middle 50% of the data.

2. Answer the following questions:
   a. Which team has the highest mean salary?

   b. Which team has the highest median salary?

   c. Explain why there is a difference between the mean and median.

   d. Which team had the larger spread? Why is this important to note?

   e. Suppose the rays traded their three highest paid players and got new players who each made $400,000 a year. What are the new values for:
      i. Mean –

      ii. Median –

      iii. IQR –

   f. Which of the above changed the most? Why?

ASSIGNMENT 3: Use the data for the 2009 Tampa Bay Rays and the 2009 Boston Red Sox salaries.

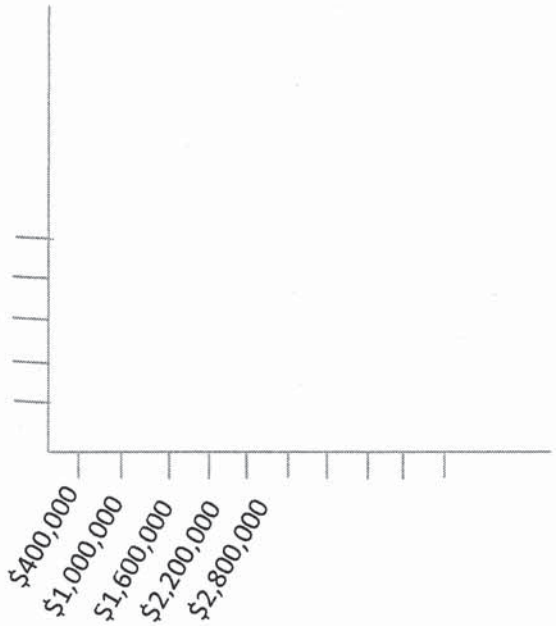
1. Complete the frequency table for each team. Use the table to create a histogram for each team using the following graph. Remember that histograms are bar graphs that show quantitative data. The bars are touching and show frequencies within a given range. (Use the same scales for the Red Sox histogram)

TAMPA BAY RAYS Salary Range	Count
$400,000 - $999,999	
$1,000,000 – $1,599,999	
$1,6000,000 – $2,199,999	
$2,200,000 -	

=29

BOSTON RED SOX Salary Range	Count
$400,000 - $999,999	
$1,000,000 – $1,599,999	
$1,6000,000 – $2,199,999	
$2,200,000 -	

=30



Salaries for Tampa Bay Rays

ASSIGNMENT 4: Complete the Independent Study – Introduction to Stem and Leaf Plots Worksheet.

Transcript: Statistics

Now we will turn to the stuff you will be tested on. Here, you do need to take notes.

Let's first look at a major split in the subject: the division between descriptive and inferential statistics. This distinction is based on what you're trying to do with your data.

With **descriptive statistics**, you describe or summarize things you definitely know.

Let's say, for example, that you have a list of all the senior SAT scores at a high school.

You want to get a better feel for the kind of students you have. To do this, you could make a graph, chart, or a table displaying their SAT scores.

For instance, this chart shows the mean (or average) SAT score for the school. This is an example of descriptive statistics: summarizing or displaying the facts known to you.

With **inferential statistics**, you're doing more than just describing. You go on to compare groups, test the hypothesis, or make predictions.

For example, suppose you hear a television commercial claiming that the gas mileage for a certain model of car is 30 mpg. You doubt the claim, and you want to test it.

The car company will make hundreds of thousands of these cars every year, and it's not practical for you to test the gas mileage on every one.

But you might go to various dealerships that sell this model and randomly pick and drive only 50 of these cars. Then you can use the information to figure out whether *all* cars of this model average approximately the gas mileage claimed by the ad.

We assume that the 50 cars we picked are representative of the hundreds of thousands of cars in the population. When we say the 50 cars are representative of the population, this means we hope that these 50 cars aren't somehow different from all the cars out there of the same make and model that we *could* have chosen. We use this small group of cars to *infer*, or conclude something about, the population of cars they came from.

This technique is part of inferential statistics. It's how we say something about a group when we haven't been able to test each group member.

The big difference between these two divisions is this: When we illustrate or summarize facts about data we already have, we're doing descriptive statistics. When we attempt to compare, test, or predict something about a group we haven't examined directly, we're doing inferential statistics.

Transcript: Phases

One way to think of a statistical study is to break it into three phases. Descriptive statistics includes the first two. Inferential statistics includes all three.

The first phase, **data gathering**, is any process that gets you data. Data gathering includes telephone surveys, written questionnaires, or simply counting something. It also includes the use of computer programs to reorganize old data.

In one sense, data gathering is the most crucial statistical phase. That's because errors made in data gathering are the hardest to correct. The usual cure when data gathering has been biased, or flawed, is to start over.

The **data organization and analysis** phase includes making graphs, charts, or tables from the data.

It also involves calculating statistics, which are numbers you get from the data. An average is an example of a statistic. Most importantly, in this phase you look for patterns in the graphs or statistics. Sometimes a pattern is found even when the statisticians weren't looking for anything in particular!

For instance, scientists looked at this graph showing oil production over many years. It raised the question whether world oil production is increasing exponentially. By "exponentially," I mean that as time passes, oil production is increasing at an increasing rate.

It turns out that it is. When a study proceeds without trying to answer any particular question, we call it **exploratory data analysis.**

More often, a study has a definite purpose. For example, in the gas mileage study we considered earlier, someone doubted the advertiser's claim.

The question is, "Does this model of car really average 30 mpg?" In the data analysis phase of the study, you might make a graph of all of your collected data and look at the average miles per gallon for the 50 cars you drove — the cars in your sample.

In the final phase of inferential statistics, we draw conclusions, or infer something, using probability. That's why we call this phase **probability-based inference**. By "probability," I mean the mathematical rules about chance that tell us how likely or unlikely something is. To understand this better, let's look more at that study about gas mileage. Suppose we'd already test-driven many cars and recorded their actual mileage.

That's the **data collection** phase.

We followed this with the **data organization and analysis** phase by creating a graph of our results and by calculating statistics. Suppose this phase revealed an average miles per gallon in our sample of 27.

Now comes the final phase, called **probability-based inference**, or simply, the inference phase.

In this phase we're guided by theories about probability, or chance. In the car mileage study, probability will tell us whether conclusion A or B is more likely.

Conclusion A says that our sample result of a 27 mpg average was a fluke. As chance would have it, the 50 cars that we sampled had unusually low gas mileage and did not represent the many other cars that we did not sample.

Conclusion B says that our sample result of a 27 mpg average was typical. As chance would have it, the 50 cars that we sampled did represent the many other cars that we did not sample. Therefore, our sample results are representative. That would mean the advertiser's claim of 30 mpg was misleading.

Probability-based inference is a fascinating part of statistics, and you'll hear more about it if you continue your study of statistics.

Transcript: Data

Since so much of statistical study depends on data, let's turn our attention to how we get it and what it looks like.

First we'll look at that term **data.**

One bit of information is a **datum**.

Many bits of information are **data**; notice that the word "data" is the plural of the word "datum." When we use the word "data," we say "these data are," not "this data is."

We can also call many bits of information a **data set**, which is a singular word meaning one **set** of data.

Data can be in the form of numbers, such as weights of anteaters trapped in an anteater preserve. This is **numeric data**.

True numeric data can be sorted and worked with mathematically. For instance, you can sort these numbers from lightest to heaviest anteater.

You can also add the numbers together, divide by the number of anteaters, and get the average weight.

Data can also be in nonnumeric forms. In an election, each ballot contains names of candidates, which are **nonnumeric data**.

Such data is often called **categorical data**.

What if the candidates were robots and for names they had numbers?

You could vote for candidate 3.14, or candidate 2.1818. While the ballots have numbers, they are still nonnumeric because the numbers are used as names. It would be meaningless to add 3.14 and 2.1818, divide by 2, and get the average robot name.

Now let's focus on numeric data.

Can your height be somewhere between 5 ft 3 in and 5 ft 4 in? Of course. But can you have somewhere between 2 and 3 brothers? Obviously not. What's the difference? Height is measured, but brothers are counted. The difference between data that are measured and data that are counted is very important to statisticians.

When data are gathered by measuring, we call them **continuous data**. Examples of continuous data are lengths, weights, volumes, and elapsed time — anything measured.

When data are gathered by counting, we call them **discrete data**. Examples of discrete data are numbers of people, outcomes with dice — anything counted.

Statistics Assignment
Identifying Types of Data and Statistics

Part 1: (14 points)
Note: Justify your answers. For example, when a question asks if a particular example of data is qualitative of quantitative, state your answer and then state *why* the data are quantitative or categorical. Show the instructor your reasoning process for arriving at your answer.

1. Identify each variable as quantitative or qualitative:

A. Amount of time it takes to assemble a simple prize

B. Number of students in a first-grade classroom

C. Rating of a newly elected politician (excellent, good, fair, poor)

D. State in which a person lives

2. Identify the following quantitative variables as discrete or continuous:

A. Population in a particular area of the United States

B. Weight of newspapers recovered for recycling on a single day

C. Time to complete a sociology exam

D. Number of consumers in a poll of 1000 who consider nutritional labeling on food products important

3. A medical researcher wants to estimate the survival time of a patient after the onset of a particular type of cancer and after a particular regimen of radiotherapy.

A. What is the variable of interest to the medical researcher?

B. Is the variable in part A qualitative, quantitative discrete, or quantitative continuous?

4. Fifty people are grouped into four categories—A, B, C, and D—and the number of people who fall into each category is shown in the table:

Category	Frequency
A	11
B	14
C	20
D	5

A. What is the variable being measured? Is it qualitative or quantitative?

5. The 1960s generation was never radical as it was portrayed. According to an opinion poll in *The American Enterprise*, when a group of 30-40-year-olds were asked to describe their political views in the 1960s and early 1970s, they gave these responses:

Conservative:	**28%**
Moderate:	**35%**
Liberal:	**31%**
Radical:	**6%**

Source: Karlyn Bowman, ed., "Opinion Pulse: '60s Kids: The Way They Were," *The American Enterprise,* May/June 1997; p. 91.

A. Is the variable qualitative or quantitative?

Part 2: (11 points)
Answer these questions.

1. This July, the U.S. House of Representatives voted to cut taxes for American citizens.

A. How would you classify the record of the vote in the 435-member House; is it a descriptive statistical study or an inferential statistical study?

B. A poll asks 435 citizens whether they want the tax cut bill passed. Is this study descriptive or inferential?

2. *National Geographic Magazine* (July, 1999) published an article called "The Shrinking World of Hornbills." (Hornbills are a genus of birds including 54 species.) Suppose you're a naturalist studying these birds. Please answer the following questions about your study of hornbills.

A. If you were to study the total number of eggs produced in one month by female Red-Knobbed Hornbills, would your variable (number of eggs) be categorical, discrete quantitative, or continuous quantitative?

B. If you wanted to know the weights of the eggs produced by Red-Knobbed Hornbills, would that variable (weight) be categorical, discrete quantitative, or continuous quantitative?

2009 / Boston Red Sox

Player	Salary	Position
Drew, J.D.	$ 14,000,000	Outfielder
Ortiz, David	$ 13,000,000	Designated Hitter
Lowell, Mike	$ 12,500,000	Third Baseman
Beckett, Josh	$ 11,166,666	Pitcher
Lugo, Julio	$ 9,250,000	Shortstop
Matsuzaka, Daisuke	$ 8,333,333	Pitcher
Bay, Jason	$ 7,800,000	Outfielder
Papelbon, Jonathan	$ 6,250,000	Pitcher
Youkilis, Kevin	$ 6,250,000	First Baseman
Smoltz, John	$ 5,500,000	Pitcher
Penny, Brad	$ 5,000,000	Pitcher
Varitek, Jason	$ 5,000,000	Catcher
Wakefield, Tim	$ 4,000,000	Pitcher
Okajima, Hideki	$ 1,750,000	Pitcher
Pedroia, Dustin	$ 1,750,000	Second Baseman
Kotsay, Mark	$ 1,500,000	Outfielder
Saito, Takashi	$ 1,500,000	Pitcher
Lopez, Javier	$ 1,350,000	Pitcher
Lester, Jonathan	$ 1,000,000	Pitcher
Green, Nick	$ 550,000	Infielder
Baldelli, Rocco	$ 500,000	Outfielder
Delcarmen, Manny	$ 476,000	Pitcher
Ellsbury, Jacoby	$ 449,500	Outfielder
Ramirez, Ramon	$ 441,000	Pitcher
Masterson, Justin	$ 415,500	Pitcher
Lowrie, Jed	$ 414,000	Shortstop
Carter, Chris	$ 400,000	Designated Hitter
Gonzalez, Miguel	$ 400,000	Pitcher
Kottaras, George	$ 400,000	Catcher
Van Every, Jon	$ 400,000	Outfielder

2009 / Tampa Bay Rays

Player	Salary	Position
Crawford, Carl	$ 8,250,000	Outfielder
Pena, Carlos	$ 8,000,000	First Baseman
Burrell, Pat	$ 7,000,000	Outfielder
Kazmir, Scott	$ 6,000,000	Pitcher
Percival, Troy	$ 4,445,000	Pitcher
Bradford, Chad	$ 3,666,666	Pitcher
Iwamura, Akinori	$ 3,250,000	Second Baseman
Wheeler, Dan	$ 3,200,000	Pitcher
Navarro, Dioner	$ 2,100,000	Catcher
Bartlett, Jason	$ 1,981,250	Shortstop
Shields, James	$ 1,500,000	Pitcher
Balfour, Grant	$ 1,400,000	Pitcher
Shouse, Brian	$ 1,350,000	Pitcher
Nelson, Joe	$ 1,300,000	Pitcher
Niemann, Jeff	$ 1,290,000	Pitcher
Gross, Gabe	$ 1,255,000	Outfielder
Kapler, Gabe	$ 1,000,018	Outfielder
Aybar, Willy	$ 975,000	Third Baseman
Isringhausen, Jason	$ 750,000	Pitcher
Cormier, Lance	$ 675,000	Pitcher
Longoria, Evan	$ 550,000	Third Baseman
Upton, B.J.	$ 435,000	Outfielder
Howell, J.P.	$ 433,700	Pitcher
Garza, Matt	$ 433,300	Pitcher
Sonnanstine, Andy	$ 430,100	Pitcher
Zobrist, Ben	$ 415,900	Shortstop
Riggans, Shawn	$ 413,900	Catcher
Joyce, Matt	$ 410,400	Outfielder
Perez, Fernando	$ 402,800	Outfielder

Statistics Independent Study
Study Sheet
Introduction to Stem-and-Leaf Plots

Most of what you need to know about stem-and-leaf plots you can learn from your textbook.

1. Construct a stem and leaf plot for the data.

Washington	67	Tyler	71	Hayes	70	Harding	57
J. Adams	90	Polk	53	Garfield	49	Coolidge	60
Jefferson	83	Taylor	65	Arthur	56	Hoover	90
Madison	85	Fillmore	74	Cleveland	71	F. D. Roosevelt	63
Monroe	73	Pierce	64	B. Harrison	67	Truman	88
J. Q. Adams	80	Buchanan	77	McKinley	58	Eisenhower	78
Jackson	78	Lincoln	56	T. Roosevelt	60	Kennedy	46
Van Buren	79	A. Johnson	66	Taft	72	L. Johnson	64
W. H. Harrison	68	Grant	63	Wilson	67	Nixon	81

Source: Robert Famighetti, ed., *The World Almanac and Book of Facts, 1997,* Mahwah, NJ, 1996.

Fill in the leaves on the stem for these data, below.

4|
5|
6|
7|
8|
9|

2. You might notice that the stem we used in Question 1 is a little crowded with leaves. Take the same data and make a stem-and-leaf plot with two lines for each stem. The first line for each stem can have any leaves numbered 0-4, and the second line can have any leaves numbered 5-9. How does this change the shape of the plot?

4|4
5|5
6|6
7|7
8|8
9|9

4. The data below are standardized birth rates, to the nearest tenth, for teenagers in 1994 by state. (*Standardized* means that the numbers are all converted to a common scale.) The states have been divided into east and west. Using the stem below, fill in the leaves for the standardized teen birth rates comparing eastern and western states. Is a stem-and-leaf plot a very efficient way to organize these data? Why not?

Western States	Birth Rate	Eastern States	Birth Rate
Alaska	55.4	Alabama	63.6
Arizona	71.1	Connecticut	45.1
Arkansas	78.0	Delaware	62.7
California	56.5	Florida	59.2
Colorado	56.9	Georgia	68.9
Hawaii	44.6	Illinois	59.3
Idaho	55.1	Indiana	63.2
Iowa	57.0	Kentucky	67.5
Kansas	63.7	Maine	42.1
Louisiana	58.9	Maryland	42.9
Minnesota	54.3	Massachusetts	47.8
Missouri	61.5	Michigan	54.8
Montana	43.9	Mississippi	62.2
Nebraska	60.2	New Hampshire	35.6
Nevada	75.1	New Jersey	37.2
New Mexico	59.4	New York	41.3
North Dakota	41.7	North Carolina	72.2
Oklahoma	69.5	Ohio	59.8
Oregon	65.0	Pennsylvania	56.2
South Dakota	45.7	Rhode Island	64.2
Texas	63.5	South Carolina	58.0
Utah	52.4	Tennessee	69.5
Washington	57.9	Vermont	32.3
Wyoming	49.1	Virginia	50.1
		West Virginia	52.3
		Wisconsin	55.2

Source: Centers for Disease Control and the National Center for Health Statistics (NCHS) in a report called the "Monthly Vital Statistics Report," vol. 45, no. 5(S), December 19, 1996.

Statistics Independent Study
Study Sheet
Introduction to Stem-and-Leaf Plots

(highest - lowest) (lowest - highest

Western States		Eastern States
	32	
	33	
	34	
	35	
	36	
	37	
	38	
	39	
	40	
	41	
	42	
	43	
	44	
	45	
	46	
	47	
	48	
	49	
	50	
	51	
	52	
	53	
	54	
	55	
	56	
	57	
	58	
	59	
	60	
	61	
	62	
	63	
	64	
	65	
	66	
	67	
	68	
	69	
	70	
	71	
	72	
	73	
	74	
	75	
	76	
	77	
	78	

Statistics Independent Study
Study Sheet
Introduction to Stem-and-Leaf Plots

5. Below is another data set divided into eastern and western states. These data are the average monthly expenditure per person to the nearest dollar for Aid to Families with Dependent Children (AFDC) in 1994. Make a back-to-back stem-and-leaf plot that compares eastern and western states. Was average monthly expenditure per person generally higher in eastern or western states? use split stems

Western States	Dollars	Eastern States	Dollars
Alaska	$247	Alabama	$ 58
Arizona	110	Connecticut	199
Arkansas	69	Delaware	120
California	192	Florida	100
Colorado	111	Georgia	91
Hawaii	219	Illinois	107
Idaho	109	Indiana	88
Iowa	128	Kentucky	79
Kansas	118	Maine	139
Louisiana	57	Maryland	118
Minnesota	169	Massachusetts	198
Missouri	91	Michigan	142
Montana	117	Mississippi	43
Nebraska	113	New Hampshire	170
Nevada	106	New Jersey	132
New Mexico	117	New York	193
North Dakota	129	North Carolina	88
Oklahoma	105	Ohio	124
Oregon	144	Pennsylvania	126
South Dakota	107	Rhode Island	180
Texas	58	South Carolina	69
Utah	129	Tennessee	60
Washington	174	Vermont	194
Wyoming	109	Virginia	108
		West Virginia	92
		Wisconsin	156

Source: AFDC Website

Acknowledgements
Question 1
This is question 1.27 (b) from pages 31-32 of *Introduction to Probability and Statistics*, Tenth Edition, by W. Mendenhall, R. Beaver, and B. Beaver. Copyright © 1999 by Brooks Cole, division of Thompson Learning Incorporated. Further reproduction is prohibited without permission of the publisher.