Types of Data
Quantitative (or measurement) Data
These are data that take on numerical values that actually represent a measurement such as size, weight, how many, how long, score on a test, etc. For these data, it makes sense to find things like "average" or "range" (largest value-smallest value). For instance, it doesn't make sense to find the mean shirt color because shirt color is not an example of a quantitative variable. Some quantitative variables take on **discrete** values, such as shoe size (6, 6 ½ , 7,...) or the number of soup cans collected by a school. Other quantitative variables take on **continuous** values, such as your height (60 inches, 72.999993 inches, 64.039 inches, etc...) or how much water it takes to fill up your bathtub (73.296 gallons or 185.4 gallons or 99 gallons, etc).

Categorical (or qualitative) Data
These are data that take on values that describe some characteristics of something, such as the color of shirts. These values are "categories" of a population, such as M or F for gender of people, Don't Drive or Drive for the method of transportation used by students to get to school. These are examples of **binary** variables. These variables only have two possible values. Some categorical variables have more than two values, such as hair color, brand of jeans, and so on.

Two types of Variables

Quantitative	Categorical
Discrete Continuous	Binary More than 2 Categories
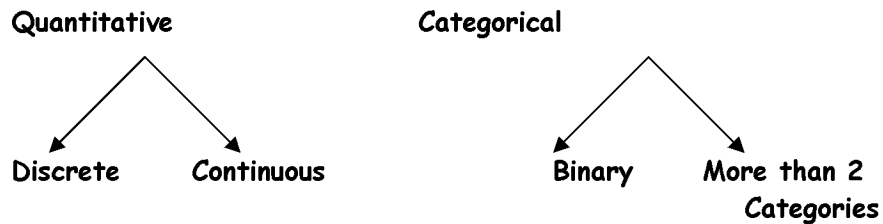
Exercises: Answer the following questions and then decide if the data is quantitative or categorical (Q or C)

	Answer	Type
1. In what grade did you take Alegbra 1?	_____	_____
2. How many CDs do you own?	_____	_____
3. How old was your father when you were born?	_____	_____
4. How old was your mother when you were born?	_____	_____
5. Choose a random integer from 1 to 20.	_____	_____
6. How many siblings do you have? (all, whether you live with them or not)	_____	_____
7. How many cousins do you have?	_____	_____

8. How tall are you (**in inches**)? _____ _____

9. How many AP classes will you be taking
THIS year? _____ _____

10. What gender are you? _____ _____

11. Where did you eat your last meal?
(1= home, 2= restaurant, 3= other) _____ _____

12. How long have you lived in this area? _____ _____

13. How far away from school do you live? _____ _____

Numerical Descriptions of Quantitative Data

Measures of Center

Mean: The sum of all the data values divided by the number (n) of data values.

> *Example:*
> Data: 4,36,10,22,9   Mean= $\overline{x} = \sum \frac{x_i}{n} = \frac{4+36+22+7}{5} = \frac{81}{5} = 16.2$

Median: The middle element of an ordered set of data

> *Examples:*
> Data: 4,36,10,22,9 = 4, 9, **10**, 22, 36 → Median = 10
>
> Data: 4, 36, 10, 22, 9, 43 = 4, 9, 10| 22, 36, 43→ Median= (10+22)/2= 16

Measures of Spread

Range: Maximum value- minimum value

> *Example:*
> Data: 4, 36, 10, 22, 9   =   4, 9, 10, 22, 36
>
> Range = Max. – Min. = 36 – 4= **32**

Interquartile Range (IQR):   The difference between the 75th percentile (Q3) and the 25th percentile (Q1). This is Q3-Q1. Q1 is the median of the lower half of the data and Q3 is the median of the upper half. In neither case is the median of the data included in these calculations.

The IQR contains 50% of the data. Each quartile contains 25% of the data.

Examples:
Data: 4, 36, 10, 22, 9 = 4,   9, <u>10,</u> 22,   36

Q1= 6.5   Q3= 29

So, the IQR = 29 – 6.5 = 22. 5

Data: 4, **9**, 10, | 22, **36**, 43

Q1   Q3

So, the IQR = 36 – 9 = 27

Exercises

Last year students collected data on the age of their moms and das when they (the students...) were born. The following are their results:

Dad:	41	27	23	31	30	33	26	32	43	25	34
	27	34	27	26	28	32	32	35	27	33	34
	34	34	35								

Mom:	39	26	23	30	28	33	23	32	38	23	35
	24	33	24	23	24	32	23	30	24	29	34
	35	26	31								

Now type the data into your calculator using the list function : **STAT → ENTER → type the Dad into L1**. If you make a mistake, you can go to the error and DELETE. If you forget an item, you can go to the line below where it is supposed to be and press 2nd DEL to insert it. Do the same for the Mom data, but put into L2.

NOTE: If the lists you are using already have numbers in them before you start, you can clear them this way: Arrow up to the line where L1 is shown. Press CLEAR, then the down arrow.

1. Find the mean and the median for the Dad data. To find the mean using your calculator, go to 2nd STAT → MATH → 5 and then type L1 by typing 2nd→ 1. This will add all the values in the list. Then divide by 26 to get the mean.

To find the median, sort the data in the lists: STAT → 2 → L1. The median is exactly in the middle between the 13th and 14th value.

Mean _____   Median_____

Are they the same? _____

If not, which is larger? _____

2. Find the mean and the median for the Mom data.

Mean _____   Median_____

Are they the same? _____

If not, which is larger? _____

3. Now compare the two means you calculated. Which is larger? _____ Is this result what you expected?_____ Why/why not?

4. Calculate the range for each set of data. Dad _____Mom _____

5. Are these ranges about the same? _____ If no, what are some reasons that might cause this difference?

6. Find Q1 and Q3 for the Dad data.   Q1_____Q3_____

7. Find Q1 and Q3 for the Mom data.   Q1_____Q3_____

8. You have now calculated the "Five-Number Summary". This can also be used as a way to determine the spread of a set of data. The five-number summary consists of:

          Miniumum   Q1   Median   Q3   Maxiumum

Write the five number summary for the Dad data: _____

Write the five number summary for the Mom data: _____

9. Now calculate the IQR for each of the two sets of data.

Dad_____

Mom_____

Graphical Displays of Univariate (one variable) Data

Quantitative Data: Dot plot
 Boxplot (Box and Whiskers)
 Stemplot (Stem and Lead)
 Histogram

Student GPA's
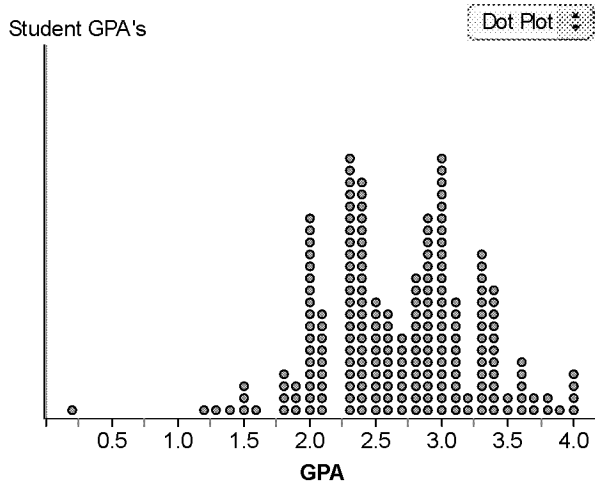
To make a Dotplot:

1. Draw and label a number line so that all the values in your dataset will fit.

2. Graph each of the data values with a dot. Be sure to line the dots up vertically as well as horizontally so that you can really see the shape of the graph.

Stemplot of Student GPAs

```
1 | 23
1 | 444
1 | 67
1 | 88888999
2 | 0000000000000000000111111111
2 | 33333333333333333333333
2 | 44444444444444444444445555555555
2 | 666666666666677777
2 | 88888888889999999999999999
3 | 00000000000000000000111111111
3 | 2233333333333333
3 | 44444444455
3 | 6666677
3 | 889
```
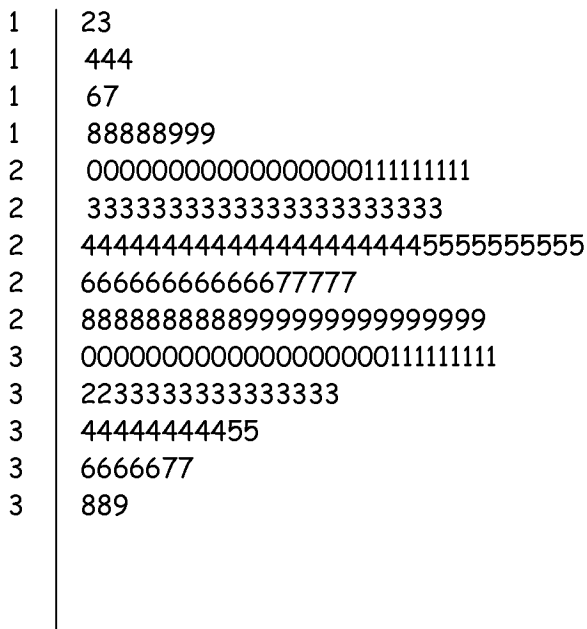
To make a Stemplot:
   1. Put the data in ascending order.

   2. Use only the last digit of the number as a leaf (see the numbers to the right of the line- each digit is the last digit of a larger number).

   3. Use one, two, or more digits as the stem. (Sometimes, you can truncate data when there are too many digits in each data value- i.e. the number 20,578 would be come 20 | 5, where the "20" is in thousands. Note that this is different from rounding.)

   4. Place the "stem" digit(s) to the left of the line and the leaf digit to the right of the line. Do this for each data value. You should then arrange the "leaves" in ascending order.

   5. Sometimes, there are many numbers with the same "stem". In this situation it might be useful to break the numbers with the same stem into either two distinct groups (each on a separate line; say, "leaves" from 0-4 on the first line and 5-9 on the second). or into five distinct groups as is shown in the graph to the right. Here, the first line for each stem contains all the 0-1 leaves, the next line contains the 2-3 leaves and so on. This technique is called "splitting the stems". It is useful in some cases in order to show the shape of the data more clearly.

Student GPA's

GPA
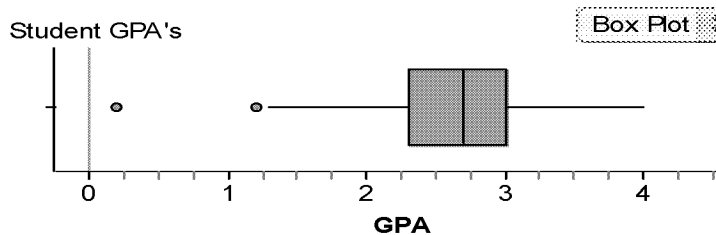
To make a box plot:
1. **Draw and label a number line** that includes the minimum and the maximum values for the set of data.

2. Calculate the five-number summary and make a dot for each of these summary numbers above the number line.

3. Draw a line between the 1st and 2nd dot, showing the "lower quartile"; then draw a line from the 4th to the 5th dot to show the "upper quartile". These are commonly called the "whiskers".

4. Draw a rectangular box from the 2nd to the 4th dot and draw a line through the box on the middle dot—the median.

NOTE: In AP Statistics, a "modified boxplo"t is used. This shows any "outliers." An outlier is a data point that does not fit the pattern of the rest of the data. When your calculator or computer software graphs a modified box plot, an algorithm is used to determine what it takes to "not fit the pattern of the rest of the data". This algorithm is 1.5 (IQR) away from the "box" part of the graph. (above and below the box). These outliers are shown with dots or stars, or any small symbol.

Student GPA's

GPA
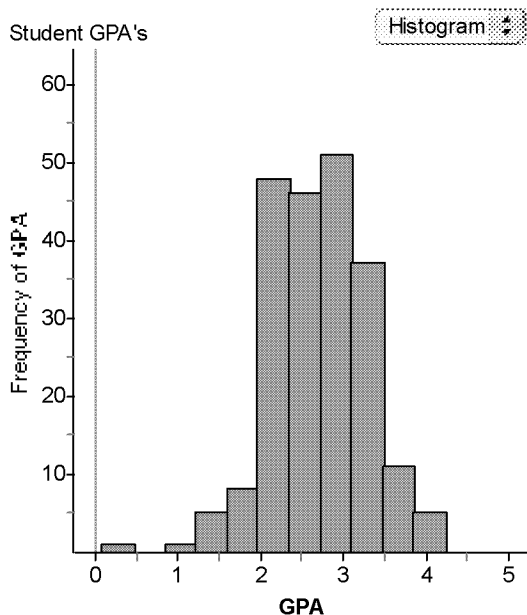
To make a histogram:

1. Put the data into ascending order.

2. Decide upon evenly spaced intervals into which to divide the set of data (such as 0,10,20,30 etc.) and then count the number of values that fall within each interval. This number is called the "frequency". If you divide each of these frequencies by the size of the data set, n, making percents then you have what you are called "relative frequencies".

3. Draw and **label** a 1st quadrant graph using scales appropriate for the data. Be sure to include a title for the x- and y- axes.

4. Graph the frequencies that you calculated in step 2.

Categorical Data: Bar Graph
                          Circle Graph

I'm assuming that you already know how to make these two types of graphs.


Assessing the *Shape* of a Graph

There are two basic shapes that we will examine: ***Symmetric*** and ***Skewed***.

Symmetric: One can tell if a graph is symmetric if a
Vertical line in the "center" divides the graph into two
Fairly congruent shapes. (A graph does *not* have to be "bell-shaped" to be considered symmetric.)
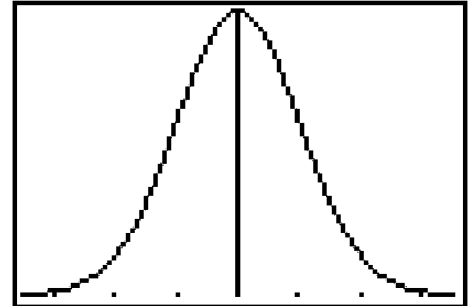
Mean ~ Median in a symmetric distribution

Skewed: One can tell that a graph is skewed if the graph has a big clump of data on either the left (skewed right) or on the right (skewed Left) with a tendency to get flatter and flatter as the values of data increase (skewed right) or decrease (skewed left). A common misconception is that the "skewness" occurs at the big clump.
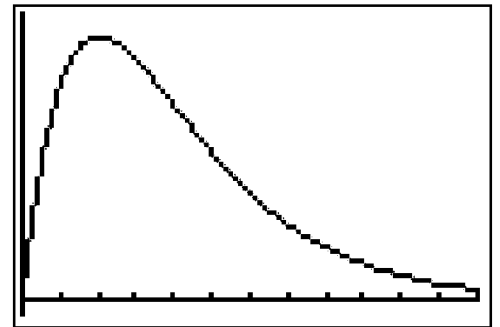
Relationship between Mean and Median in a skewed distribution:
Skewed *Left,* the mean is Less.   Skewed *Right,* the mean is Might.

Gathering Information from a Graphical Display
The first thing that should be done after gathering data is to examine it graphically and numerically to find out as much information about the various features of the data as possible. These will be important when choosing what kind of procedures will be appropriate to use to find out an answer to a question that is being investigated.
The features that are the most important are Shape, Center, Spread, Clusters and gaps: SCSCO. Most of these can only be seen in a graph. However, sometimes the shape is indistinct – difficult to discern. So, in this instance (usually because of a very small set of data), it's appropriate to label the shape "indistinct."

Exercises

1. Construct a boxplot for each of the following sets of data taken from consumer ratings of different brands of peanut butter in the September, 1990 issue of *Consumer Reportrs*. **Use the same number line for both graphs**. (You could do it this way: Draw a number line. Above this line construct the "Crunchy" boxplot. Then, above the "Crunchy" boxplot, construct the "creamy" boxplot.)

Crunchy: 62 53 75 42 47 40 34 62 52 50
                  34 42 36 75 80 47 56 62

Creamy: 56 44 62 36 39 50 53 45 65 40
                  56 68 41 30 40 50 56 30 22

a. Find the range for: Creamy_____ Crunchy_____

b. Find the median for: Creamy_____ Crunchy_____

c. Looking at your boxplots and comparing the medians, what type of peanut butter do consumers tend to prefer?

2.   The following data is taken from the *Statistical Abstract of the United States* (112th Edition). These are the ages of drivers arrested for DUI from a random sample of size 50. Make a stemplot to show the distribution of this age data.

45 16 41 26 22 33 30 22 36 34 63 24 26 18 27 24 31 38 26
55 31 47 27 43 35 22 64 40 58 20 49 37 53 25 29 32 23 49
39 40 24 56 30 51 21 45 27 34 47 35

a.   What is the shape of this graph?_____

b.   Using your stemplot, find the median of this data. _____

c.   Which data display is better- a boxplot or a stemplot?_____
      Why? (More detail?)

3. For the following graphs, find the shape, center (**just do the median**), and spread (find only the **range**). If there are any notable features evident in the graph (clusters, gaps, or outliers), then say where they are. Otherwise do not comment on clusters, gaps, or outliers. (Note:
To find the center of these graphs, use the frequencies found on the y-axis. Count how many are in each bar. Add these up and divide by 2. This tells you where the median is located. Which bar is this value in? That's the median. For graph A, n=21, so the middle value is 10.5. Starting with the first bar count 1+2+4+3+6.... So the median is in the bar that contains the 10.5 value (bigger than 10 anyway). That's 30. So, the median is 30. To find a **VERY** rough estimate of the mean, take the frequency for each bar and multiply it by the value along the x-axis for that bar. Add these up for all the bars and then divide by 21. You get the mean = 28.571.
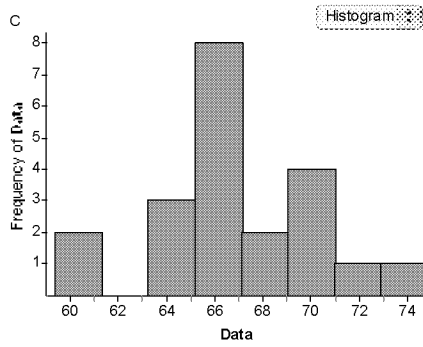
A

Histogram

A

Frequency of Data
7
6
5
4
3
2
1

20 22 24 26 28 30 32 34 36
Data

Shape_____
Center_____
Spread_____

B

Histogram

B

Frequency of data
12
10
8
6
4
2

0 20 40 60 80 100 120 140 160
data
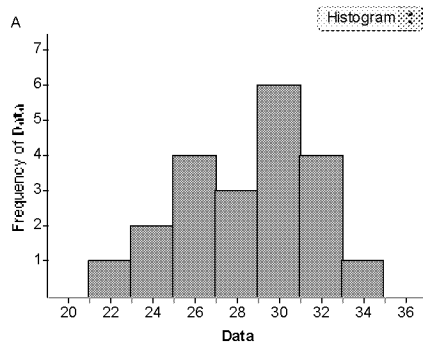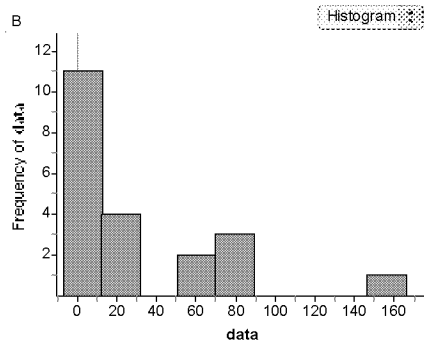
Shape_____
Center_____
Spread_____
Clusters,Gaps?____
Where?_____

Outliers?_____Where?

C
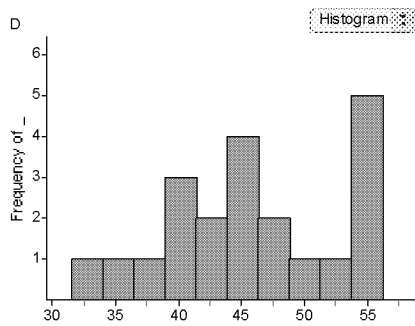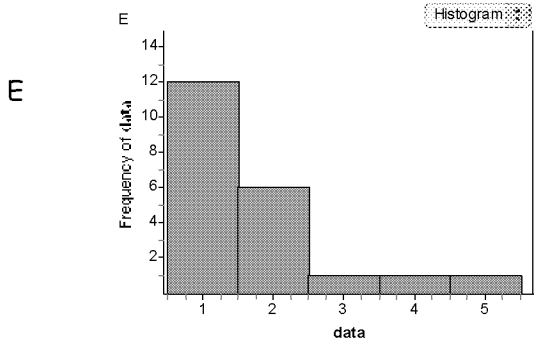
Histogram

C

Frequency of Data
8
7
6
5
4
3
2
1

60 62 64 66 68 70 72 74
Data

Shape_____
Center_____
Spread_____
Clusters, Gaps?_____   Where?

D

Histogram

D

Frequency of _
6
5
4
3
2
1

30 35 40 45 50 55
_

Shape_____
Center_____
Spread_____

E

E

14
12
10
8
6
4
2

Frequency of data

1   2   3   4   5

data

Shape_____
Center_____
Spread_____
Clusters?_____Where?

F

9
8
7
6
5
4
3
2
1

Frequency of data

0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5

data

Shape_____
Center_____
Spread_____

G

Histogram

G

12
10
8
6
4
2

Frequency of data

0   5   10 15 20 25 30

data

Shape_____
Center_____
Spread

H

Histogram

H

9
8
7
6
5
4
3
2
1

Frequency of data

0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5

data

Shape_____
Center_____
Spread_____

I
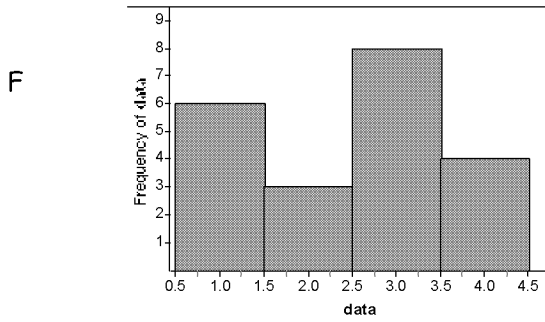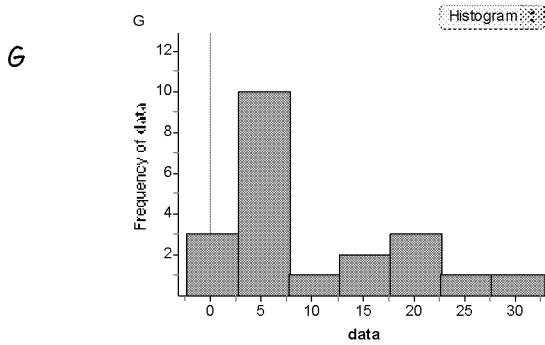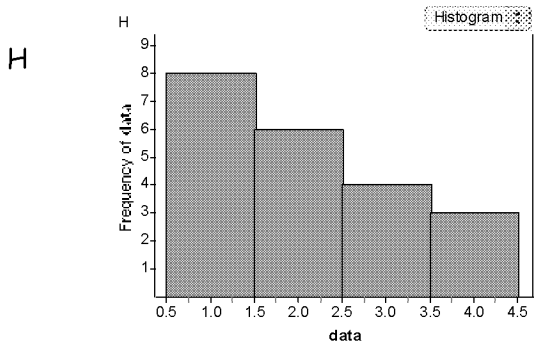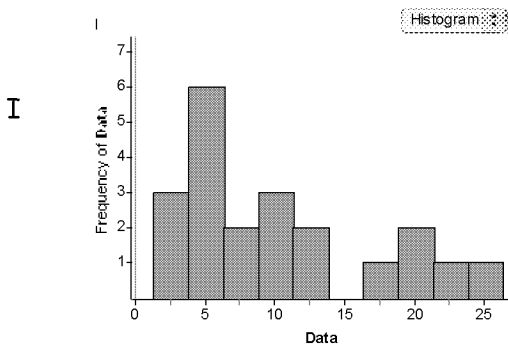
Shape_____
Center_____
Spread_____
Clusters, Gaps?_____Where?

4. Use the following list of variables to identify which of the graphs in Question 4 *could* be a graphical display of the answers for a typical class of students. (Look at things like scale, range, min, max.) Write the letter of the correct graph in the blank provided. There are more variables than there are graphs, so don't worry if you have extras.

Variables

Grade when a student takes Algebra 1_____

Average # of CD's you own_____

Age of your father when you were born _____

Age of your mother when you were born _____

Age of your stat teacher this year_____(guess, even if you don't know me!)

# of siblings you have _____

# of cousins you have _____

Your height (in inches) _____

# of AP classes you will be taking this year _____

How long you have lived in this area _____

How far away from school you live (in miles) _____

Amount of change in your pocket on the first day of school _____

6. **Reading Assignment:**   Read the following letters written to you by some of the AP Stat students in the year 2007-08.

7. **Writing Assignment:**   Write your own letter about what you are going to do to ensure success this next year in this and all of your AP classes.

On the first Monday after school starts, you will be given a "summer packet" test. Not only will you be test on all material in this packet, but the packet must also be turned in and complete (50 points). The test will count for an additional 50 points. This material that is covered in the first chapter of the textbook is very important.

Questions???? Email me here: adam.shipman@fcps.edu

Dear New Student,

Hello! We're supposed to explain to you what this class is like, so here goes. Basically, if you're thinking this is just a plug and chug number math class and you may have even heard in passing that it's an easy math AP... you're just wrong. This is more like an integrated math and English class. There's a lot of analysis, but instead of analyzing words and dead people, you're analyzing numbers and data.

The class may seem easy at first, because you're working with basic graphs and you'll think to yourself that you learned all this material in elementary school. However, as the year goes on, you'll realize that you need to pay attention because the material gets more complicated. Also, do not be worried if you don't understand the material when he teaches you at first, because for most of the units we didn't understand what was going on when he initially taught it, but we got it in the end.

Always do your homework. It helps a ridiculous amount on the test... you don't have to do it right away, but have it done before quiz and test time. It will really help you grasp the material. If he asks you if you have any questions, try to ask questions because he likes to answer. Do the review packet. I repeat. DO THE REVIEW PACKET. Often times the questions on the test are very very similar to the ones on the review packet. So do it!

Oh, sometimes if you're lucky, there are random kids in the hallways being annoying and Mr. Shipman will go out and yell at them. He can be intimidating, but if you work hard he'll like you. :)

Well, this is about it. We're not going to tell you to have fun, because there will be many occasions throughout the year that you will wish you weren't sitting here. But when it comes down to it, you'll be glad you took this class. Good Luck!!

~ Jennie and Lisa

Dear Future Student of Statistics,

Welcome to AP Statistics! The purpose of this letter is to advise you on what to expect for this coming year. This class will probably seem pretty easy at first (you begin with basic graphs and displays) but make sure you continue to pay attention throughout the entire year (well, until the AP test). One major thing to realize is that at the beginning of every unit, you will probably have close to no idea of what is going on, but by the

middle of the unit, after doing myriad worksheets, I promise you'll understand it if you do the homework and pay attention in class.

Another really big thing to realize is that there is a lot of writing. You'll be doing lots of equations, but you'll have to explain how you got these equations and what they mean. All the writing gets tedious after a while (wait until inference tests in the middle of the year), but it will save you on the AP Test and earn you many points. For the first semester of the year, you'll be doing a lot of different units which don't really seem connected, but they'll all make sense beginning second semester when you begin inference. Inference is all the same concepts just using different types of data, and everything you learned in first semester will fall into inference. So definitely retain all the information you learn throughout the year until the AP test.

A note about the AP Test…you'll feel overwhelmed but just relax! Start reviewing fairly early, even if it's just a bit, or else you'll be cramming a year's worth of equations, formulas, and concepts into your brain, which really isn't fun (I and many of my classmates know this from experience!) So yeah, just relax and don't cram!

A note about class: about once a month we do candy labs which is always fun because we get tons of skittles, M&Ms, oreos…doughnuts and bagels the day of the AP Test…all good stuff! And although the material can get a bit crass, class is usually fun and Mr. Shipman is interesting when he teaches…and if you're not paying attention, he'll make you look like an idiot…so watch out!

More helpful information—DO YOUR HOMEWORK! Mr. Shipman checks it all, and the points add up, and, it will really help you practice for tests and quizzes. Corrections can be done after tests for 1/3 of points back, and the tests will get curved if necessary. Also, ASK QUESTIONS IF YOU NEED HELP! Mr. Shipman loves when you ask questions because that means you're actually paying attention. He's always happy to help you after school, during class (if you're doing a worksheet or something), before school, etc. He wants you to understand the material and it really is imperative that you understand the material to succeed in the class.

So…enjoy the class and have a good year!
-Sabrina Caldwell

P.S. The hard work all year is worth the relaxing last month of class: a fairly easy project, and lots of card games with Mr. Shipman. It would be to your benefit to learn how to play Hearts!

How to do well in AP Stat

The best advice I can give anyone taking AP Statistics would be to do the homework diligently. The tests and quizzes come almost directly from the homework so if you do them you will do well. The catch is there is no one looking over your shoulder making sure you do the homework when it's assigned. You have to be accountable and you actually HAVE TO do the homework.

Joe Russo

Dear New Statistics Student,

AP Statistics is a very interesting course. It isn't a math course as you're used to. It requires very little mathematical comprehension, but requires a very open mind. In return it'll open your mind to many concepts and perspectives you've never experienced before. I apologize for the poetry, I was just signing yearbooks. Anyways, Mr. Shipman is a very good teacher. This will only be his third year teaching the course and every year he gets so much better at teaching it. At the beginning of each unit, you'll be very confused and lost, but towards the end, you'll realize that it's really just the same as all the other units, but with a few twists. Good luck this year and don't stress, just do the homework or pay attention in class.

Sincerely,
Warren Juba

During the second half of my junior year I was left with a decision: to take a higher math class, or to take AP statistics. Although I eventually decided to take both, I still believe that to a college freshman, knowing statistics is crucial. AP Statistics teaches us how to make logical choices base on data instead of emotions, which is going to be helpful throughout our life. AP Statistics is not an easy class. For many seniors, it may seem like another AP to put on your resume, but I assure you that you will feel sorry for yourself if you are not willing to make the effort. The workload is reasonable, tests and quizzes take up most of the spaces on your grade sheet, but believe it or not, those are the easy grades. I enjoyed the class as much as I enjoyed talking to Mr. Shipman, he's the best woohooooo

Yufei Zhou

Dear prospective AP Statistics Student:

Unlike other math classes you may have taken before, AP Stat deals largely with vocabulary and interpreting what you are being asked. Answers to questions can require little mathematical work and instead require an entire paragraph of explanation. Sometimes you'll answer a question and it will look like a short answer English essay.

So when you start class in September, keep in mind that as a different kind of math class there are different kinds of demands. You have to understand the principles behind what you are doing because you have to explain it. Memorizing a formula will not help you at all in stats.

The best advice I can give you is what you would do for any class.
- Do the homework.
- Tests are way more influential than quizzes, so if you bomb a few don't worry.
- Read the book.

- Listen to Mr. Shipman when he tells you easier ways to do things on your calculator.
- Utilize test corrections.
- Go to class.

Good luck in AP Statistics and with your senior year.

Best Wishes,

Alaura Maglio

Dear Future AP Stat Student,

You made the right choice to take this class! It is a very fun and interesting course, and I'm really glad I took it.

You should remember that this is an AP course, so there is hard work involved! There is a decent amount of homework, but not enough to kill you. Mr. Ship will give you an assignment sheet with all the homework for the chapter at the beginning of the unit so you know how much homework you have. Mr. Ship doesn't check to see if you did the homework until the end of the unit on the test day, which has good and bad implications. The good news is if you don't have time to do the homework one night, it doesn't hurt you! But if you are a procrastinator like me, and wait to do all the homework the night before it is due, it's kinda rough. So try to keep up with the homework, it will make your life so much easier!

The concepts you learn in AP Stat may be overwhelming at first. You may think to yourself, "I am never going to understand this," but not to worry! Mr. Ship is always open for questions and he makes himself available during Lunch and Learning for extra help. So don't be intimidated to ask questions if you really don't get something!

Tests are somewhat difficult in this course. If you do poorly on a test, don't fret. Mr. Ship gives you the opportunity to do test corrections to receive a third of your lost points back. This definitely helped me throughout the year. You have to do the corrections during school, so I would advise you to do them during in Lunch and Learning because again, Mr. Ship will be there to answer any questions on a problem you didn't understand and explain why you missed points.

I hope this advice will help you have a successful year!

Sincerely, Michelle Zimmerman

Dear Student who is going to take AP Statistics,

AP stat is an insightful class where you will learn many interesting things. Mr. Shipman is a good teacher, however you cannot always rely on the lectures and notes to pass. You also need to put effort into the work and study on your own. The lectures are important, however they will only help you on the free response. Read the text book to help with the multiple choice questions. Although people say this is an easy AP class, if you walk in

thinking that you might be surprised. There are actually some confusing material when you first see it, however it will soon become second nature as the year goes by. Doing the homework helps a lot, and for the online quizzes, do them before any quiz or test. They are great review.

<div align="right">

Tian-Hao Wang

June 6th, 2008

</div>

Dear Future AP Statistics Student,

As a junior or senior, you will have many academic and extracurricular obligations, including other AP courses, sports practices and games, and college applications. It is easy to let your math homework become a less important priority when it is not getting checked daily. The best advice I can offer you as you begin your year in AP Statistics is to do your homework when it is assigned and thoroughly. It doesn't take long and you will remember it a lot better than if you do it all the night before a test. (Homework for the unit is checked on the day of the unit test). Make sure your class notes are complete and COME TO CLASS. If you do your homework and pay attention, you will do well. The material in this class starts out easy and then gets progressively harder, so don't be fooled by the simplicity of the material in the first unit.

Don't worry about the AP, if you have paid attention in class and done your homework, you will do absolutely fine. It's not too difficult and there is a tremendous curve.

Good luck next year!

Sincerely,

Lara Iglesias
Class of 2008

P.S. I'm serious, do your homework.