Welcome to AP Statistics! This can be a thought provoking, challenging, real-life math class that has applications in almost every career you could choose to go into. This summer packet is a series of assignments to get you ready for the learning we will do in AP Stats as well as cover a bit of the content to free up a day or two for some hands-on activities that will reinforce important concepts. As with all AP classes the timeline is very tight to teach all the material before the AP Exam in May. It is especially difficult to cover the necessary material for us since our school district starts in September and we have an extra break in February. So it is important that you do the assignments in this summer packet so you are prepared for class. Work through the assignments in order. Using your own lined paper, label each assignment with the homework number. Bring all the homework with you to the first day of Stats in the fall. Email me with questions but be patient, I'm not checking my email every day during the summer.

HW #1:

1. Without looking anything up, off the top of your head, what do you think the study of statistics is? What experiences or lessons have you had in the past with statistics? (2-6 sentences)

2. Read this definition and transcription of a lecture below.

Sta·tis·tics Source: http://www.merriam-webster.com/dictionary/statistics

Etymology: German Statistik: study of political facts and figures, from New Latin statisticus: of politics, from Latin status: state. Date: 1770

1 : a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data *[note: this is for Statistics with a uppercase S]*

2 : a collection of quantitative data *[note: this is for statistics with a lowercase s]*

What is Statistics? by Jordan Neus (http://www.fiu.edu/~neusj/whatisstatistics.html)

Statistics is becoming increasingly more important in modern society with passing time. We are constantly being bombarded with charts, graphs, and statistics of various types in an attempt to provide us with succinct information to make decisions. Sometimes this information is presented in a manner so as to sway us toward a particular view. As consumers and decision makers we must be aware of this. Which drug should we take? Which car should we buy? Where will the economy go? Who is infected with a particular deadly disease? These are all examples of questions which are usually relegated to the statistician for analysis and dissemination. This lecture will attempt to introduce some of the reasoning behind the necessity of statistical inference.

In order to realistically understand the subject of Statistics it is important to appreciate the rationale behind why and how Statistics is used by the world, at large. That is, why do we need Statistics anyway? This, perhaps, is a bit philosophical, yet I cannot over emphasize the need for thinking along these lines. Without proper perspective, Statistics becomes a mere mathematical exercise, diverging from the true nature of the subject.

In order to begin our analysis as to why Statistics is a necessary type of reasoning we must begin by addressing the nature of science and experimentation. A characteristic method used by scientists is to study a relatively small collection of objects, say 2500 people, and a characteristic, say longevity, and through experimentation or observation, draw a conclusion appropriate for the entire class of objects (i.e. people, in general). For example, suppose a study published results suggesting **"people who own pets live longer".** Would this mean that every person who owns a pet are likely to live long lives? Does owning a pet **cause** longevity? Suppose the people in the study, by chance, were on the whole, very healthy people, and therefore lived long lives: Would

this invalidate the researcher's assertion that people who own pets live longer? The obvious problem with this type of reasoning is that these issues can never be proved absolutely. This type of scientific reasoning is called *inductive reasoning* and is inherently flawed. One can never study a sample and expect conclusions to hold true for the entire population with absolute certainty. This is exactly why Statistics is needed.

In contrast to the lack of certainty associated with inductive reasoning, the type of logic used in Mathematics is absolutely certain. The mathematician begins with general principles and logically concludes more specific relationships. This type of reasoning from the general to the particular is called *deductive reasoning*. A rather simplistic (but nevertheless correct) example is based on the principle that two numbers can be added in any order, thereby giving the same sum. This is called the axiom of commutativity. An example of deductive reasoning would be to assert that since this holds for any two numbers, surely this must hold for the numbers two and three, in particular. We are, therefore, absolutely certain that 2 + 3 = 3 + 2, given the axiom of commutativity.

In its applied form, Statistics then becomes a bridge between the inductive uncertainty of science and the deductive certainty of Mathematics. In his classic book, ***The Design of Experiments***, Sir Ronald A. Fisher expresses this idea beautifully: "***We may at once admit that any inference from the particular to the general must be attended with some degree of uncertainty, but this is not the same as to admit that such inference cannot be absolutely rigorous, for the nature and degree of the uncertainty may itself be capable of rigorous expression."*** Statistics, therefore, is the mathematical method by which the uncertainty inherent in the scientific method is rigorously quantified.

3. Write a paragraph in response to this definition and lecture by Jordan Neus. Anything new you learned? What is the lecture telling you? What questions do you still have?

HW #2:

1. Read the article below.
The New York Times - August 6, 2009
For Today's Graduate, Just One Word: Statistics By STEVE LOHR

MOUNTAIN VIEW, Calif. — At Harvard, Carrie Grimes majored in anthropology and archaeology and ventured to places like Honduras, where she studied Mayan settlement patterns by mapping where artifacts were found. But she was drawn to what she calls "all the computer and math stuff" that was part of the job.

"People think of field archaeology as Indiana Jones, but much of what you really do is data analysis," she said. Now Ms. Grimes does a different kind of digging. She works at Google, where she uses statistical analysis of mounds of data to come up with ways to improve its search engine.

Ms. Grimes is an Internet-age statistician, one of many who are changing the image of the profession as a place for dronish number nerds. They are finding themselves increasingly in demand — and even cool. "I keep saying that the sexy job in the next 10 years will be statisticians," said Hal Varian, chief economist at Google. "And I'm not kidding."

The rising stature of statisticians, who can earn $125,000 at top companies in their first year after getting a doctorate, is a byproduct of the recent explosion of digital data. In field after field, computing and the Web are creating new realms of data to explore — sensor signals, surveillance tapes, social network chatter, public records and more. And the digital data surge only promises to accelerate, rising fivefold by 2012, according to a projection by IDC, a research firm.

Yet data is merely the raw material of knowledge. “We’re rapidly entering a world where everything can be monitored and measured,” said Erik Brynjolfsson, an economist and director of the Massachusetts Institute of Technology’s Center for Digital Business. “But the big problem is going to be the ability of humans to use, analyze and make sense of the data.”

The new breed of statisticians tackle that problem. They use powerful computers and sophisticated mathematical models to hunt for meaningful patterns and insights in vast troves of data. The applications are as diverse as improving Internet search and online advertising, culling gene sequencing information for cancer research and analyzing sensor and location data to optimize the handling of food shipments.

Even the recently ended Netflix contest, which offered $1 million to anyone who could significantly improve the company’s movie recommendation system, was a battle waged with the weapons of modern statistics.

Though at the fore, statisticians are only a small part of an army of experts using modern statistical techniques for data analysis. Computing and numerical skills, experts say, matter far more than degrees. So the new data sleuths come from backgrounds like economics, computer science and mathematics.

They are certainly welcomed in the White House these days. “Robust, unbiased data are the first step toward addressing our long-term economic needs and key policy priorities,” Peter R. Orszag, director of the Office of Management and Budget, declared in a speech in May. Later that day, Mr. Orszag confessed in a blog entry that his talk on the importance of statistics was a subject “near to my (admittedly wonkish) heart.”

I.B.M., seeing an opportunity in data-hunting services, created a Business Analytics and Optimization Services group in April. The unit will tap the expertise of the more than 200 mathematicians, statisticians and other data analysts in its research labs — but that number is not enough. I.B.M. plans to retrain or hire 4,000 more analysts across the company.

In another sign of the growing interest in the field, an estimated 6,400 people are attending the statistics profession’s annual conference in Washington this week, up from around 5,400 in recent years, according to the American Statistical Association. The attendees, men and women, young and graying, looked much like any other crowd of tourists in the nation’s capital. But their rapt exchanges were filled with talk of randomization, parameters, regressions and data clusters. The data surge is elevating a profession that traditionally tackled less visible and less lucrative work, like figuring out life expectancy rates for insurance companies.

Ms. Grimes, 32, got her doctorate in statistics from Stanford in 2003 and joined Google later that year. She is now one of many statisticians in a group of 250 data analysts. She uses statistical modeling to help improve the company’s search technology.

For example, Ms. Grimes worked on an algorithm to fine-tune Google’s crawler software, which roams the Web to constantly update its search index. The model increased the chances that the crawler would scan frequently updated Web pages and make fewer trips to more static ones. The goal, Ms. Grimes explained, is to make tiny gains in the efficiency of computer and network use. “Even an improvement of a percent or two can be huge, when you do things over the millions and billions of times we do things at Google,” she said.

It is the size of the data sets on the Web that opens new worlds of discovery. Traditionally, social sciences tracked people’s behavior by interviewing or surveying them. “But the Web provides this amazing resource for observing how millions of people interact,” said Jon Kleinberg, a computer scientist and social networking researcher at Cornell.

For example, in research just published, Mr. Kleinberg and two colleagues followed the flow of ideas across cyberspace. They tracked 1.6 million news sites and blogs during the 2008 presidential campaign, using algorithms that scanned for phrases associated with news topics like “lipstick on a pig.” The Cornell researchers found that, generally, the traditional media leads and the blogs follow, typically by 2.5 hours. But a handful of blogs were quickest to quotes that later gained wide attention.

The rich lode of Web data, experts warn, has its perils. Its sheer volume can easily overwhelm statistical models. Statisticians also caution that strong correlations of data do not necessarily prove a cause-and-effect link.

For example, in the late 1940s, before there was a polio vaccine, public health experts in America noted that polio cases increased in step with the consumption of ice cream and soft drinks, according to David Alan Grier, a historian and statistician at George Washington University. Eliminating such treats was even recommended as part of an anti-polio diet. It turned out that polio outbreaks were most common in the hot months of summer, when people naturally ate more ice cream, showing only an association, Mr. Grier said.

If the data explosion magnifies longstanding issues in statistics, it also opens up new frontiers. “The key is to let computers do what they are good at, which is trawling these massive data sets for something that is mathematically odd,” said Daniel Gruhl, an I.B.M. researcher whose recent work includes mining medical data to improve treatment. “And that makes it easier for humans to do what they are good at — explain those anomalies.”

2. Write a paragraph about what you learned in this article. Was any part of it surprising to you? What do you still have questions about?

HW #3:

1. Read the information below from: http://satterthwaiteapstat.blogspot.com/2012/08/5-ws.html

The 5 W's are who, what, when, where, why, and how is also included. These are important to statistics so the person looking at the information knows the circumstances of the data, as well as how to think of the data.

- Who- This tells people who the information is about. It can be one person, or more of a group.
- What- The what shows what they are looking at, as well as what the information is about.
- When- When shows how recent and up to date the information is. It is important because a survey from the 1800's is going to display results a lot different from now.
- Where- Different parts of the world, or nation, are going to think and believe different things about a topic, so the where gives people information on what the things involved are like.
- Why- This is the reason why the statistics of a topic were calculated, and show people the purpose of the information.
- How- A one-on-one survey might give different results than an anonymous one, so some topics might require information on how the information was gathered.

2. Read this article: “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.

3. Answer the following questions regarding the above excerpt:

a) **Who**: Who was being studied for this project (be specific – it wasn't just "teens in general")?

b) **What**: What about those individuals is being recorded / analyzed (i.e. what are the variables?)?

c) **When**: When was the data collected?

d) **Where**: In what geographical area was the data collected?

e) **Why**: For what purpose do you think this data was collected and analyzed?

f) Why do 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?" What else could be affecting the accidents besides early school starts?

HW #4:

1. Watch a youtube video or read an article that explains the difference between Categorical Data and Quantitative Data, and the difference between Discrete Data and Continuous Data. (Don't have to read/watch all, these are just examples)
https://www.youtube.com/watch?v=2zSYAlonQIQ
https://www.youtube.com/watch?v=7bsNWq2A5gI
https://courses.lumenlearning.com/wmopen-concepts-statistics/chapter/what-is-data/
https://support.minitab.com/en-us/minitab-express/1/help-and-how-to/modeling-statistics/regression/supporting-topics/basics/what-are-categorical-discrete-and-continuous-variables/
https://www.dummies.com/education/math/statistics/types-of-statistical-data-numerical-categorical-and-ordinal/

A good way to test if the data is Quantitative is if it makes sense to find an average. Can we find the average (mean) GPA for our class? Yes, this is Quantitative data. Can we find the average car type the class drives? No, this is Categorical data. There ARE some type of data that are numerical but also Categorical. For instance Zip Codes. Zip Codes are categorical because they tell what part of the US a person lives in, they put you in a category. It would not make sense to find the average Zip Code for our class. We would be averaging numbers like 98072, 98021, and 98012. The result would have no meaning. So don't just assume that because it is a number, it is quantitative. Other examples of numerical categorical data are 1 representing Female, 2 representing Male, or categories for income level like $0-10,000, $10,000 -$50,000, $50,000+.

2. For the following situation, properly identify the variable type.

The FAA monitors airlines for safety and customer service. For each flight, the carrier must report the type of aircraft, flight number, number of passengers, and whether or not the flights departed and arrived on schedule. What *variables* are reported for each flight, and are they quantitative or categorical?

Variables	Quantitative or Categorical	
(1) Type of aircraft	☐ Quantitative	☐ Categorical
(2) Flight Number	☐ Quantitative	☐ Categorical
(3) Number of Passengers	☐ Quantitative	☐ Categorical
(4) Arrived/Departed on Schedule	☐ Quantitative	☐ Categorical

3. Check your answers at the bottom of the page. Then complete the following:

Determine if the variables listed below are *quantitative* or *categorical*. Neatly print "Q" for quantitative and "C" for categorical.

_____ 1. Time it takes to get to school

_____ 2. Number of shoes owned

_____ 3. Hair color

_____ 4. Temperature of a cup of coffee

_____ 5. Teacher salaries

_____ 6. Gender

_____ 7. Facebook user

_____ 8. Height

_____ 9. Amount of oil spilled

_____ 10. Age of Oscar winners

_____ 11. Type of pain medication

_____ 12. Jellybean flavors

_____ 13. Country of origin

_____ 14. Type of meat

4. Go back to the variables you identified as Quantitative and circle the ones that are discrete variables.

HW #5:

Recall, that there are three ways to measure the approximate center of a set of data: the mean, the median, and the mode. Each one of these measures of center can be useful, depending on the situation.

1. The ***mean*** is often called the average. It is calculated by adding up the data pieces (finding the total) and dividing by how many data pieces you have. We will call the number of data pieces "n".

a. All vocabulary tests in your English class are out of 10 points. These are your scores:10, 1, 8, 9, 7, 9, 9, 8, 7. Find the mean of your test score.

b. A small software company has just a few employees, a secretary, two junior programmers, two senior programmers and a Chief Executive Officer (CEO). Their salaries are: $23000, $32000, $32000, $37000, $37000, $90000. Find the mean of the salaries.

Type of Aircraft: Categorical, Flight Number: Categorical (average would make no sense), # of passengers: Quantitative, On Time: Categorical

c. For a science project, you have been recording the daily high temperatures for your town each day for the last week, they are: 45˚, 53˚, 50˚, 40˚, 68˚, 55˚, 47˚. Find the mean of the temperatures.

2. The ***median*** is the middle number of an ordered set of data. To calculate the median, you put the data in order from least to greatest. If there is an odd number of data pieces, the middle piece of data is your median. If there is an even number of data pieces, you need to average the two middle data pieces.
a. Calculate the median for the vocabulary test data. First, write the data in order from least to greatest and then circle the middle number.
b. Calculate the median for the salary data.
c. Calculate the median for the temperature data.

3. The ***mode*** is the most common piece of data in a data set. It is possible to have more than one mode. It is also possible to have no mode if there are no duplicate numbers in a set of data. The mode must occur more than once. Find the mode for each of the data sets in #1.

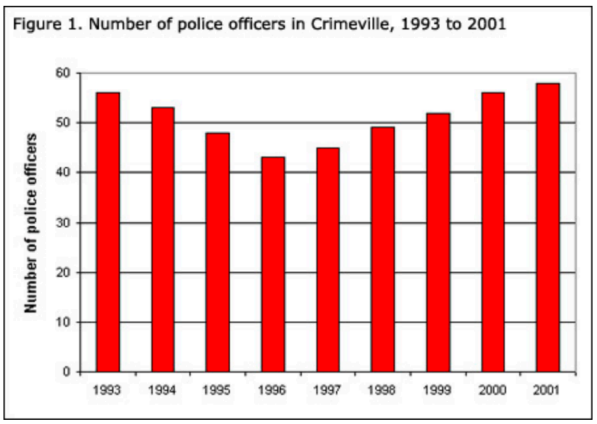
4. Look back at your answers for the mean, median, and mode for each situation to answer the following questions.
a. If your teacher allowed you to choose, which grade went on your report card: the mean, median, or mode, which would you choose?
b. If you were the CEO at the computer company and you wanted to entice someone to come work for you by showing them how great the salary is at your company. Would you tell them the mean, median, or mode salary for your company?
c. Your science hypothesis for you experiment was that the temperatures were going to be high, will you report the mean, median or mode of your data?

5. a. The mean for the salary data was much higher than both the median and the mode. Look back at the data, what could have caused it to be so high?
b. The mean for the vocabulary test data is lower than both the median and the mode. Look back at the data, what could have caused it to be so low?
c. An outlier is an extremely high or low piece of data. Do you think an outlier will effect the mean? Why or why not?
d. Do you think an outlier will effect the median or mode? Why or why not?

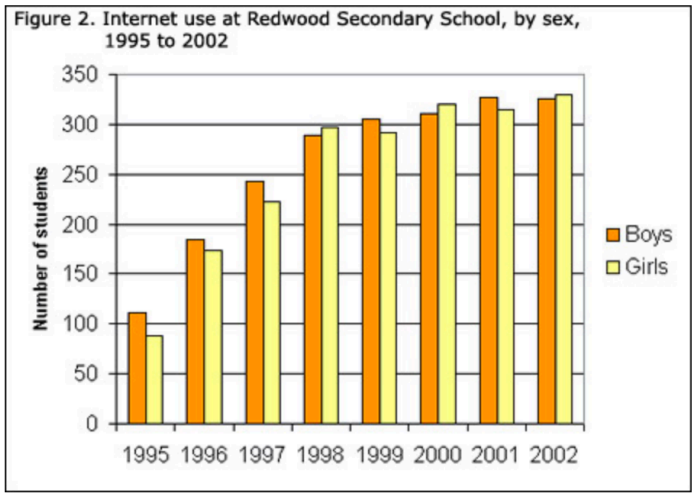
HW #6:

1. Read through the following: There are different types of graphs for different types of data. We are going to look at graphs for Categorical Data in this assignment. Graphing data shows us the distribution of the variable –what values the variable takes and how often it takes these values. Graphs give us information to help us to understand the data. For all graphs, it is important to label axes for clarity.

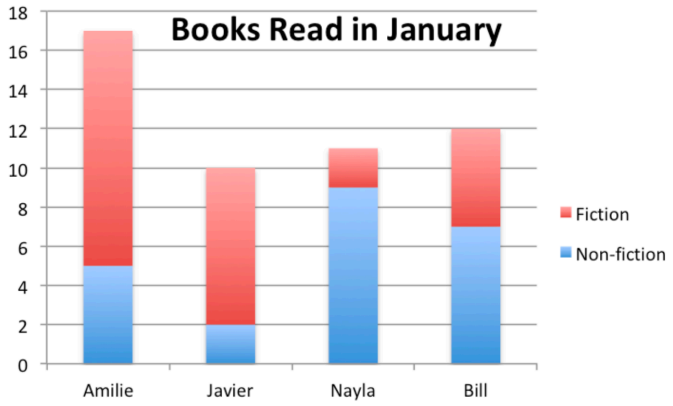
Categorical Data Graphs:
I. Bar Graph – Usually used to compare the number of data points in each category. a) Bars are separated by space. (Not touching each other.)
b) Each bar is labeled with a category.
c) Vertical axis shows frequency (count) or percentage and should have a consistent scale.
d) Allows quick comparisons of the frequencies of each category.



•There are different types of bar graphs.
Double or side-by-side bar graphs can show a comparison of two different data sets. (to the right)
•Stacked bar graphs can be used to divide larger categories into smaller categories and show comparisons between all part. (to the right, below)
You will be asked to create bar graphs on occasion in AP Stats. Be sure you know that they are for comparing categorical data.



II. Pie Chart – Usually used to compare the percent of each category to each other and to the whole.
a) Show what part of the whole each category represents. (Must have a "whole" to break up into parts.)
b) Must include all of the categories that make up the whole. (Cannot leave out a category.)
c) May include a category labeled "Other". You may combine categories under a common label.



You will not be asked to create circle graphs in AP Stats unless it is a sketch or using computer software. Creating a circle graph by hand can be a lot of work because you have to calculate angles to correspond with percentages and use a protractor to create them. You will be much more likely to be asked to interpret data from a circle graph than to create one.

2. Create a stacked bar graph for the following data. The data show the enrollment (in hundreds) of male and female students at a particular college in the years 1995, 2000, and 2005.

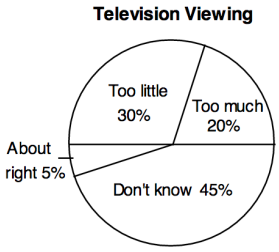
Year	1995	2000	2005
Male	30	34	32
Female	28	35	33

3. Create a side-by-side bar graph for the following data. This is consumer and expert ratings for different versions of a new phone that a company is planning to sell.

	A	B	C
1		Consumer rating	Expert rating
2	Version 1	8.5	9
3	Version 2	6.7	7.2
4	Version 3	8.8	6.4
5	Version 4	7	7.3

4. Look back at your two bar graphs. Do they each have a title? Do they each have clear, accurate labels. Are the vertical scales consistent? Did you use color to accentuate differences and make a key to show what each color represents?

5. Students were asked whether they spend too much or too little time watching television. The circle graph shows the responses of 120 students. How many students thought they watched too little television?



HW #7:

In this lesson we are going to review some tables created from Quantitative Data.
Stem and Leaf Plots: Recall that a Stem and Leaf Plot is a table where each data value in the data set is split into a "leaf" (usually the last digit) and a "stem" (the other digits). Here is an example of a Stem and Leaf Plot

for the number of employees in some different companies in one town. The original data was 23, 24, 33, 34, 35, 36, 39…etc… Notice that each number on the right (the leaves) represents the last digit of one of the data pieces. The leaves are put in order from **least to greatest** without commas. The key is very important for a stem and leaf plot because it tells you what numbers the data values in the table represent.

Company sizes (number of employees)

Stem	Leaf
2	3 4
3	3 4 5 6 9
4	1 4 7 7 7
5	2 6 9
6	4 7 8 9

Key: 6|4 means 64

1. Answer the following questions using this Stem and Leaf Plot.

a. How many companies had 47 employees?

b. How many companies do we have information on? (How many pieces of data are there?)

c. One company had more employees than all of the others, how many employees did this company have?

d. Without using your calculator find the median for this set of data.

e. If I added a piece of data for a company that had 19 employees, where would it go on the stem and leaf plot? Be specific, what would the stem be?

To compare two sets of data, you can use a 'back-to-back' Stem and Leaf Plot. For instance, here is a Back-to-Back Stem and Leaf Plot for heart rates during PE. Group 1 data was taken before a one mile jog. Group 2 data was the same students after the jog. Notice that the leaves for Group 1 are arranged from least to greatest starting at the stem and moving to the left.

Heart rates

Group 1	*Stem*	*Group 2*
7 5 5	11	
9 9 9 9 8 8 7 7 6	12	5 5 6
6 6 4 4 0	13	4 6 8 8
4 2 0	14	2 3 6 6 7 8 8
	15	3 5 5 7
	16	2 4

Key: 7|12| = 127 bpm
|13|4 = 134 bpm

2. Answer the following questions from this given plot.

a. How many students are in this PE class?

b. What was the lowest heart rate before the jog? What was the lowest heart rate after the jog?

c. What is the median of the heart rates after the jog?

d. What is the mode of the heart rates before the jog?

Frequency Tables:

The frequency of a particular data value is the number of times the data value occurs. For example, if four students have a score of 80 on the chemistry test, then the score of 80 is said to have a frequency of 4. The frequency of a data value is often represented by *f*. A Frequency Table is constructed by arranging collected data values with their frequencies. You can often add a column for tallies to make counting the frequency a bit easier. Consider the Frequency Table here. On a math quiz out of 9 points these are the scores from one class.

Score x	Tally	Frequency f
3	|	1
4	||	2
5	|||	3
6	||||	4
7	𝍸 |	6
8	𝍸	5
9	||||	4

a. What was the highest score on the quiz? How many students got that score?

b. What was the lowest score on the quiz? How many students got that score?

c. How many students took this quiz in this class?

d. What is the mode of this data?

e. When asked to find the median for this data set here is what one group did:

•Muriel lined up the scores 3, 4, 5, 6, 7, 8, 9 and circled the 6 because it was in the middle.

•Robert lined up the frequencies from the last column 1, 2, 3, 4, 6, 5, 4 and circled the 4 because it was in the middle.

•Eric said that Robert needed to put the frequencies in order first 1, 2, 3, 4, 4, 5, 6 but he still circled the 4.

•Will said "Didn't we say that there were 25 students in this class? We should be writing out 25 scores to find the median."

Who do you agree with most? Why? Then, find the median.

4. Consider the new Frequency Table to the right that displays the weights of the students in one class rounded to the nearest kilogram.

Class interval x(weight in kg)	Tally	Frequency f
40 - 44	||	2
45 - 49	||||	4
50 - 54	𝍸	5
55 - 59	𝍸 |||	8
60 - 64	𝍸	5
65 - 69	||||	4
70 - 74	||	2

a. This is the first table we've looked at where we don't know the specific data. Why can't we list the exact data pieces here?

b. For this table the mode would be expressed as a range. What is the mode?

c. How many students are in this class?

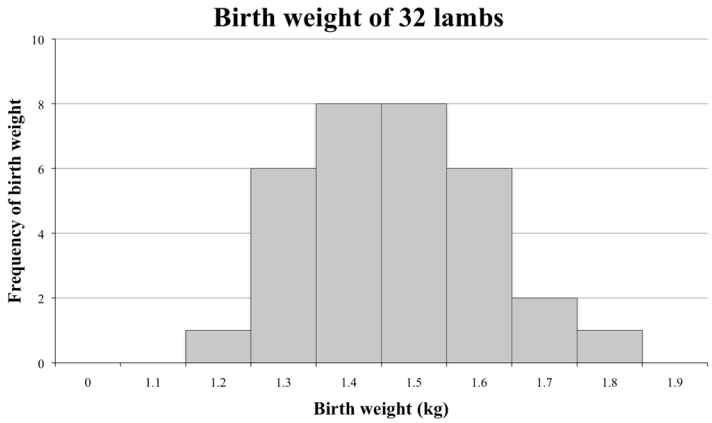
d. What is the heaviest weight in this class? How many students weigh this much?

HW #8:

The most important Quantitative data graph we will study is the histogram. Histograms show the distribution of one list of quantitative data. It is important for seeing the distribution of the data.

Histograms:

Histograms are visual displays of frequency tables. They look a lot like bar graphs but are different in several important ways. The x-axis on a histogram is always numbers, and the y-axis is always frequency (how often those numbers occur). The histogram represents the birth weights of 32 lambs on one farm last spring.

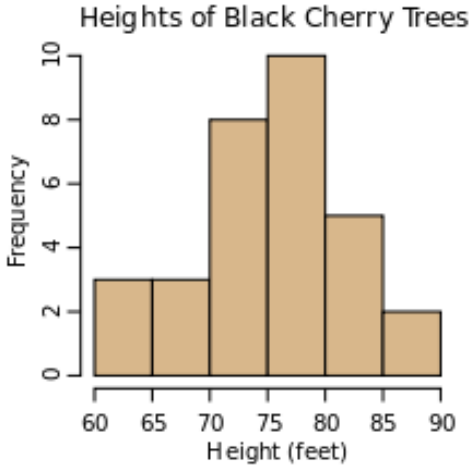


1. a. Study the histogram to the right. What does the bar above 1.4 mean? How many lambs does this bar represent? How much did those lambs weigh?

b. What was the heaviest lamb's weight? How many lambs weighed that much?

c. What was the most common weight of the lambs on this farm last spring?

d. Without using your calculator, see if you can estimate the median weight of these lambs.



2. The histogram to the right is of the Heights of Black Cherry Trees at one orchard. If a data piece falls right on the edge of a bar (for example 65 ft or 70 ft) then it goes in the bar to the right.

a. How many Black Cherry trees do they have at this orchard?

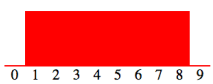
b. How tall is the tallest tree at this orchard? How many trees are this same height?

c. What is the mode of this data?

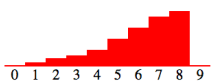
d. Approximate the median for this data.

e. The bar from 60-65 and the bar from 65-70 are the same height, what does that mean in terms of the heights of these trees?

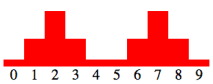
The following pictures show how we describe the shape of the distribution of the data on a histogram.



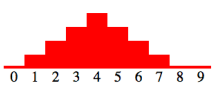
Uniform



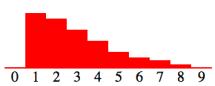
Skewed left



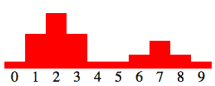
Symmetric, bimodal



Symmetric, unimodal, bell-shaped



Skewed right



Non-symmetric, bimodal

If it helps, watch a youtube video or read online about the "Shape of a Histogram".

3. Study the histogram to the right and answer the following questions.

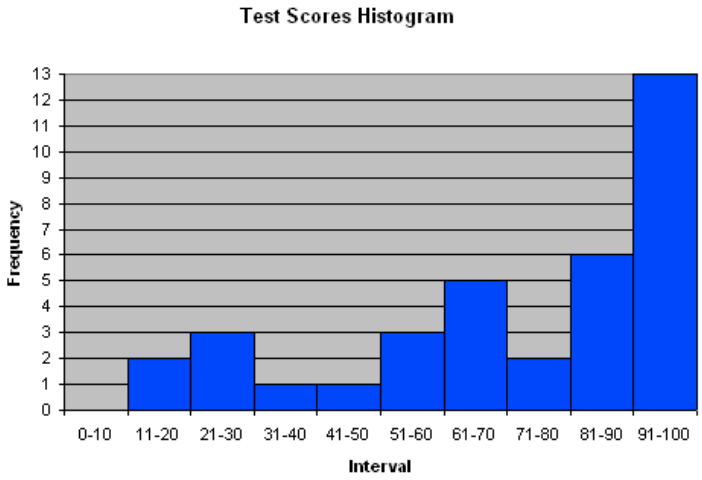
a. How many students are in this class?

b. What is the mode of this data?

c. Approximate the median of this data.

d. What does the bar above 11-20 mean?

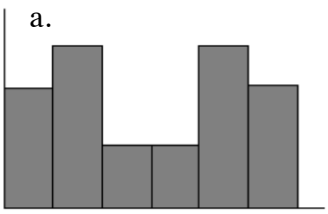
e. What is the shape of this data?



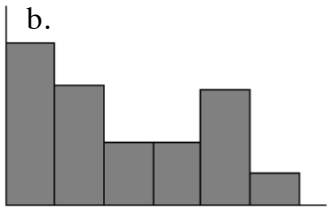
Points on Chapter 5 test

4. Name the shape of each of the following histograms.

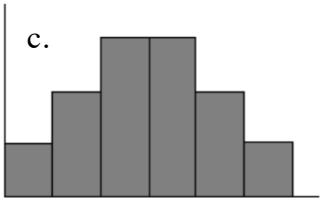
a.



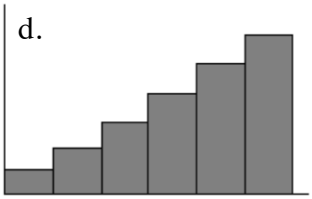
b.



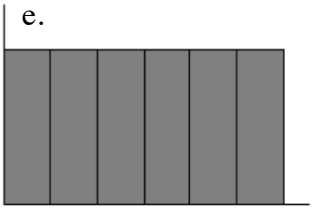
c.



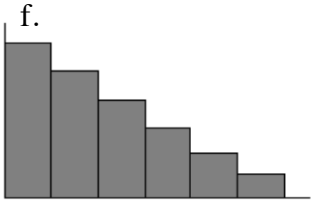
d.



e.



f.



HW #9:

Drawing a Histogram: We will create a histogram for the height of high school students in the theater production. We have their heights in inches to the right.

Individual Height, Measured in Inches				
69.9	68.9	68.2	66.0	71.0
69.0	70.0	68.5	66.5	72.5
69.6	69.5	70.0	67.5	73.0
68.5	70.4	66.8	68.3	69.0
65.0	71.1	69.0	68.2	71.3
65.9	71.0	69.3	69.1	68.2
67.2	72.5	69.1	70.2	68.5
67.5	73.1	69.4	69.5	70.0
68.0	68.8	68.5	70.5	67.0
68.6	71.3	65.5	70.8	69.2

1. Create a histogram following these steps below.

- Count the number of data points (50 in our height example).
- Determine the range of the sample - the difference between the highest and lowest values (73.1-65, or 8.1 inches for these heights)
- Determine the number of intervals (or bins) we will put the data into. Each bin will be represented by one bar. You will usually like to have 5-10 bars for most data sets. Because we have a relatively small range for our data (8.1) we might just choose to go by ones on the x-axis. So we would start with 64 or 65 and go by ones. (Draw the x-axis on your paper.)
- Develop a table or spreadsheet with frequencies for each interval. In other words figure out how many pieces of data are between 65 and 66, 66 and 67, etc… Remember if the data point is exactly on the edge of a bar (like 68 or 71) it goes in the bar on the right.
- Once you have figured out how many data pieces are in each bar, you can create the vertical axis with frequencies. Do that on your graph now.
- Draw in your bars. Be sure that your bars are touching. (No space in between like a bar graph.)
- Be sure to label your x-axis “Height in inches” as well as the numbers at the edge of each bar. Label your y-axis “frequency”.

Customer Wait Time in Seconds (n=20)	
43.1	42.2
35.6	45.5
37.6	30.3
36.5	31.4
45.3	35.6
43.5	45.2
40.3	54.1
50.2	45.6
47.3	36.5
31.2	43.1

2. Create a histogram for the wait time data to the right.

HW #10:

The **Five Number Summary** is a way to look a the spread of a set of data (how spread out the data is). The Five Number Summary is the minimum, lower quartile, median, upper quartile, and maximum. The data below is the number of speeding tickets an officer wrote each day of work for a two week period. The minimum and maximum have been underlined and the median has been circled. 1, 5, 7, 8, 10, (13), 14, 17, 21, 30, 42

1. We are going to find the five number summary for this data set. Write down the minimum.

a. Look at the 5 numbers lower than the median, find the middle number of those 5 numbers and write that down for the lower quartile. Then write down the median (circled). Do the same thing for the numbers above the medianthat you did for the lower quartile, this is the upper quartile. Then write down the maximum.

b. Find the range of this data.

c. The interquartile range (IQR) is the upper quartile minus the lower quartile. Find the IQR for this data.

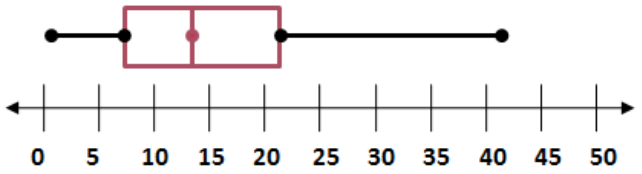
d. Will the median always be one of the data pieces? Explain.

e. Will the lower and upper quartile always be one of the data pieces? Explain.

f. Will the min and max always be one of the data pieces? Explain.

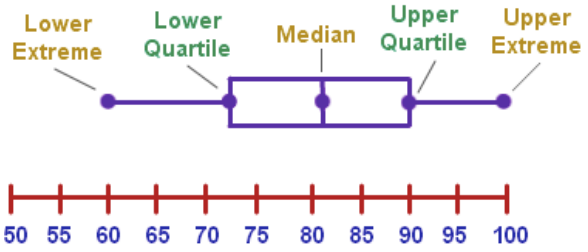
Box and Whisker Plots

A Box and Whisker Plot is a visual representation for the five number summary. The Boxplot below represents the speeding ticket data from #1. Notice how the five number summary separates the data into 4 sections. Each section has 25% of the data in it. In this way we can see how spread out the data is. The upper whisker is very long compared to the other parts, this means that the upper part of the data is more spread out than the rest. Also the box part in the middle represents the middle 50% of the data. The length of this box is the interquartile range.



2. Looking at just the boxplot (not knowing the data from #1) can you tell how many pieces of data there were? Why or why not?

3. The Box and Whisker Plot to the right represents the Chapter 8 math quiz scores from Mr. Smith's 1st period class. The test was out of 100 points.



Per 1 Chapter 8 math scores

a. Can you tell how many students were in the class?

b. Estimate the five number summary for this data.

c. Estimate the range and interquartile range for this data.

d. Complete the following sentences:

Half the test scores were above _____.

75% of the test score were below_____.

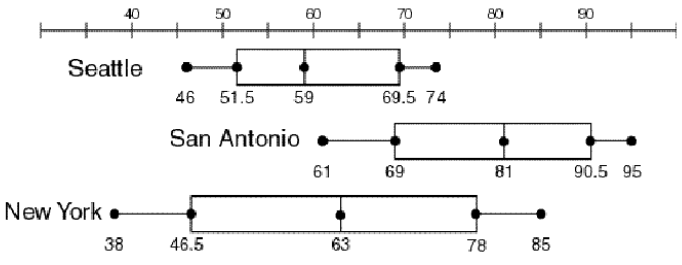
The top 25% of the class got scores between _____ and _____.

e. Would you say the data above the median or the data below the median is more spread out? Explain.

f. Is there any way to tell the mean or mode of this data?

g. Would you say this class did well on this test? Explain.

4. Boxplots are often used to compare data sets. The average temperatures were recorded each month for a year in the cities of Seattle, San Antonio, and New York.



a. Which city has more varied temperatures? How can you tell from the graphs?

b. Calculate the interquartile range for each city.

c. What do the IQRs tell us about the data?

d. Does a short whisker represent data that is more concentrated or more spread out? Explain.

e. Find the shortest section of all the graphs. What city? What section?