AP Statistics
Summer Packet

Name: _____

Date: _____

Teacher: _____Mr. Zukatus_____

AP Statistics Summer Assignment

Welcome to AP Statistics! The purpose of this assignment is to make you more comfortable exploring data analysis. This packet should be completed by your return to school in September. You are expected to complete each part of each problem and to construct all data displays neatly. This assignment will be graded, and it will count as a quiz grade in the first marking period of the school year.

The summer assignment is composed of three parts.

1. Reading and Vocabulary: You will use a free online Statistical tutoring site that will give you information on variable and data displays. While reviewing the information on the site you will be completing a vocabulary list (See pages 3-7). Follow the steps below:

- Go to www.stattrek.com
- On the left side, click on "AP Statistics"
- A list of general topics will appear on the left-hand side of the screen. When you click on the general topic a list of subtopics will appear. You will read or watch the video on the following subtopics to complete the vocabulary list.

General Topic: The basics
Variables
Population vs Sample
Mean and Median
Variability
Position
General Topic: Charts and graphs
Patterns in data
Dotplots
Histograms
Stemplots
Boxplots
Cumulative plots
Scatterplots
Comparing data sets
General Topic: Charts and graphs
One-way tables
Two-way tables

2. **Practice Problems.** After reading all the material above and from your book, you should be able to complete the vocabulary list and questions in the remaining pages of this packet. You may do so in the spaces provided. If you have any issues or questions you can email me at zukatusjoh@hhschools.org. I will check my email Friday mornings throughout the summer.

AP Statistics Summer Assignment

Part 1: Vocabulary List

Please define each of the following terms from the information on the stattrek website. When asked, provide a UNIQUE example or sketch of the word, i.e. one NOT given on the website and NOT the one your friends use.

1. Categorical Variables:

   Example:

2. Quantitative Variables:

   Example:

3. Discrete Variables:

4. Continuous:

5. Univariate Data:

6. Bivariate Data:

7. Population:

   Example:

8. Sample:

   Example:

9. Median:

10. Mean:

    Formula:

11. Outlier:

12. Parameter:

13. Statistics:

14. Range:

15. Standard Score (z-score):

    Formula:

16. Center:

17. Spread:

18. Variance:

    Formula:

19. Standard Deviation:

    Formula:

20. Symmetry:

    Sketch:

21. Unimodal:

    Sketch:

22. Bimodal:

    Sketch:

23. Skewness:

    Sketch Skewed left:

    Sketch Skewed right:

24. Uniform:

Sketch:

25. Gaps:

Sketch:

26. Outliers:

Sketch:

27. Dot plots:

28. Bar chart:

29. Histogram:

30. Difference between bar chart and histogram:

31. Stemplots:

32. Boxplots:

33. Quartiles:

34. Range:

35. Interquartile Range:

36. Four ways to describe data sets:

A.

B.

C.

D.

37. Types of graphs that can be used for comparing data

Part 2: Practice Problems

CATEGORICAL OR QUANTITATIVE

Determine if the variables listed below are *quantitative* or *categorical*.

1. Time it takes to get to school
2. Number of people under 18 living in a household
3. Hair color
4. Temperature of a cup of coffee
5. Teacher salaries
6. Gender
7. Smoking
8. Height
9. Amount of oil spilled
10. Age of Oscar winners
11. Type of Depression medication
12. Jellybean flavors
13. Country of origin
14. type of meat
15. number of shoes owned

STATISTIC – WHAT IS THAT?

A statistic is a number calculated from data. Quantitative data has many different statistics that can be calculated. Determine the given statistics from the data below on the number of homeruns Mark McGuire has hit in each season from 1982 – 2001.

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

Mean	
Minimum	
Maximum	
Median	
Q1	
Q3	
Range	
IQR	

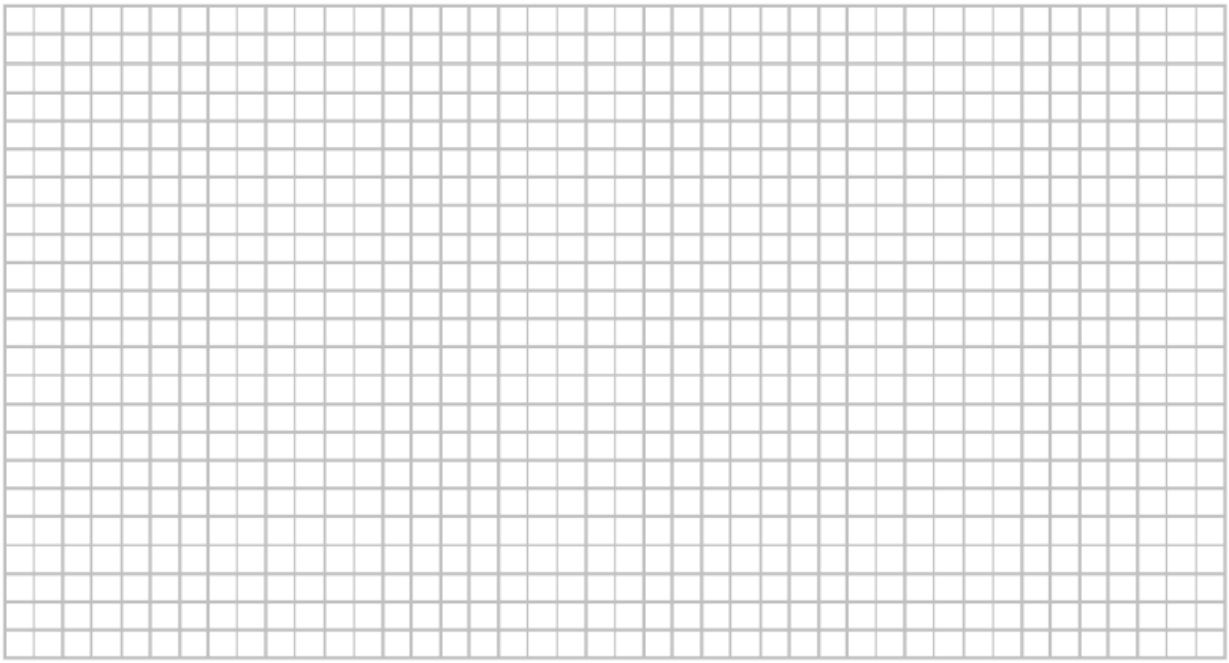
ACCIDENTAL DEATHS

In 1997 there were 92,353 deaths from accidents in the United States. Among these were 42,340 deaths from motor vehicle accidents, 11,858 from falls, 10,163 from poisoning, 4051 from drowning, and 3601 from fires. The rest were listed as "other" causes.

a. Find the percent of accidental deaths from each of these causes, rounded to the nearest percent.

b. What percent of accidental deaths were from "other" causes?

c. NEATLY create a well-labeled **bar graph** of the distribution of causes of accidental deaths. Be sure to include an "other causes" bar.



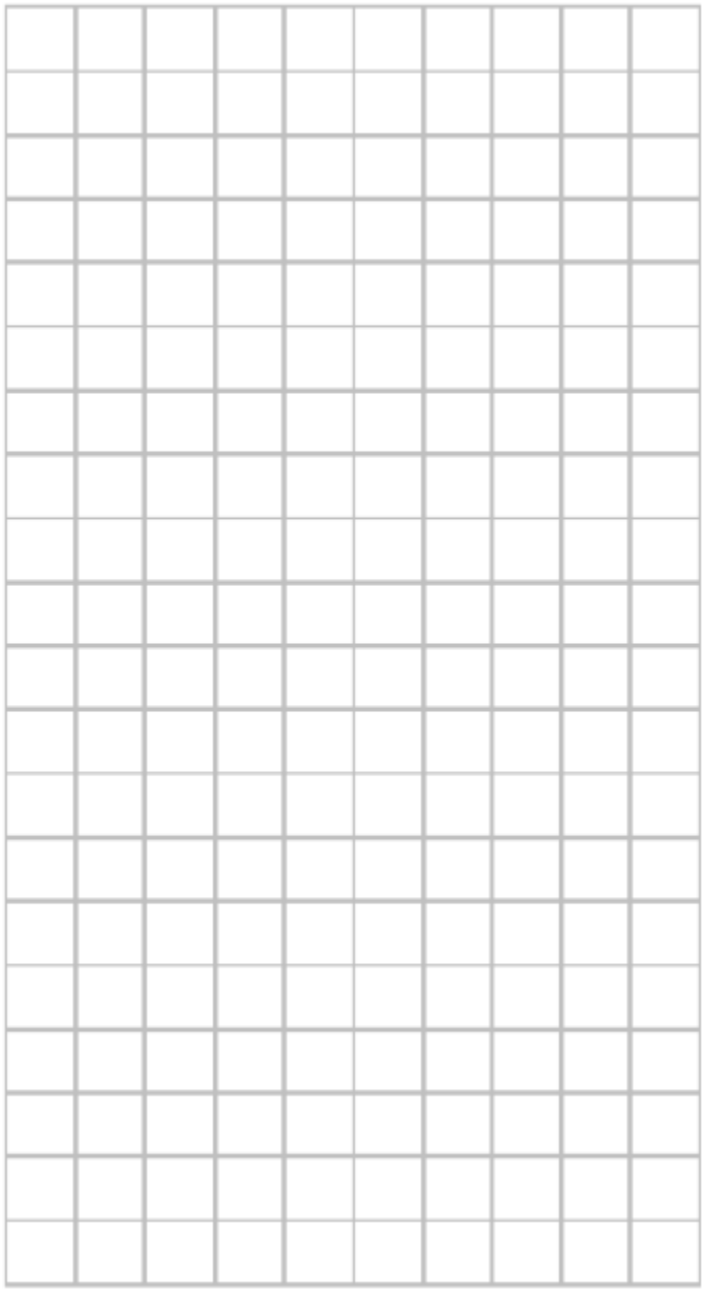
d. A pie chart is another graphical display used to show all the categories in a categorical variable relative to each other. Create a pie chart for the accidental death percentages. You may try using a software or internet source to make one and paste in the space below. (*Microsoft Excel works well*)

IT'S A TWISTA

The data below gives the number of hurricanes that happened each year from 1944 through 2000 as reported by *Science* magazine.

3	2	1	4	3	7	2	3	3	2	5	2	2	4	2	2	6	0	2	5	1	3	1	0
3	2	1	0	1	2	3	2	1	2	2	2	3	1	1	1	3	0	1	3	2	1	2	1
1	0	5	6	1	3	5	3																

a. Make a dotplot to display these data. Make sure you include appropriate labels, title, and scale. The graph paper should help ensure you space your markings (you may use x's or dots) consistently.



SHOPPING SPREE!

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

3	9	9	11	13	14	15	16	17	17
18	18	19	20	20	20	21	22	23	24
25	25	26	26	28	28	28	28	32	35
36	39	39	41	43	44	45	45	47	49
50	53	55	59	61	70	83	86	86	93

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

WHERE DO OLDER FOLKS LIVE?

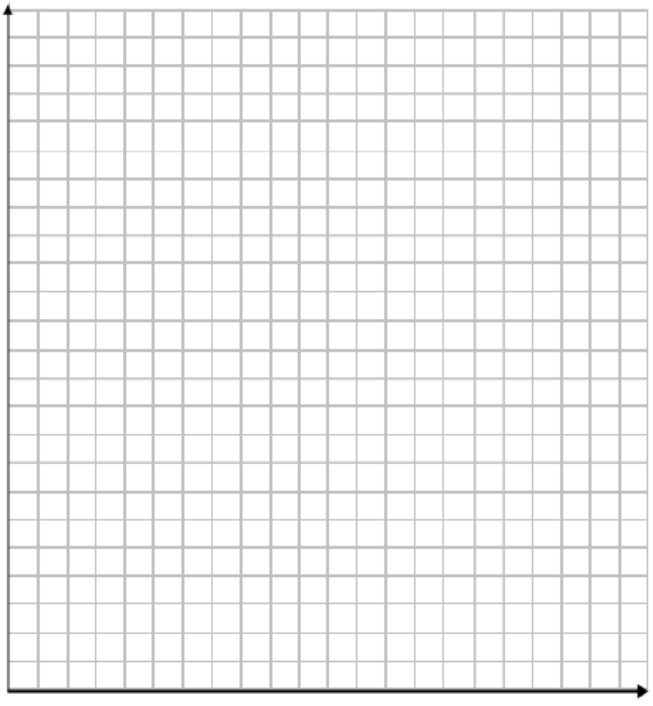
This table gives the percentage of residents aged 65 or older in each of the 50 states.

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

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

a. Finish the chart of Bin widths and then create a histogram using those bins on the grid below. Make sure you include appropriate labels, title and scale.

Bin Widths	Frequency
4 to < 6	1
6 to < 8	
8 to < 10	



SSHA SCORES

Here are the scores on the Survey of Study Habits and Attitudes (SSHA) for 18 first-year college women:

154 109 137 115 152 140 154 178 101 103 126 126 137 165 165 129 200 148

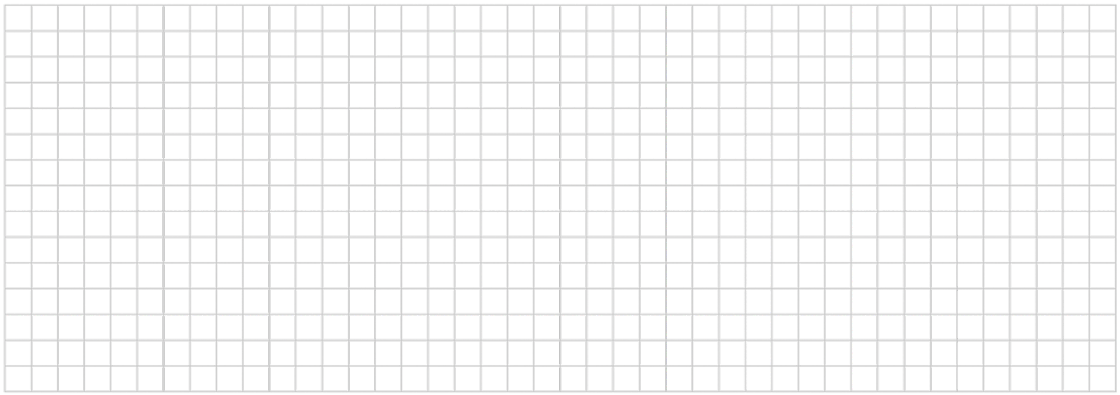
and for 20 first-year college men:

108 140 114 91 180 115 126 92 169 146 109 132 75 88 113 151 70 115 187 104

a. Put the data values in order for each gender. Compute numeral summaries for each gender.

Women			**Men**	
Mean			Mean	
Minimum			Minimum	
Q1			Q1	
Median			Median	
Q3			Q3	
Maximum			Maximum	
Range			Range	
IQR			IQR	

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



ALGEBRA PAGE!

The prerequisite for AP Statistics is Algebra II. You will not find very much equation solving in this course, but some quick review of Algebra I and Algebra II content will be helpful.

To answer the following refer to the readings on www.stattrek.com "Survey Sampling Methods".

The 7 types of sampling designs are:

A. voluntary response	B. convenience	C. simple random	D. stratified
E. cluster	F. multistage	G. systematic	

1. The Maryland division of Weight Watchers is doing research to determine how many people on the Weight Watchers diet cheat at least once a week. They decide that anonymous surveys will give them an accurate representation but do not have time to get responses from ALL the Maryland Weight Watchers people. *Read the scenarios below and determine which of the 7 sampling methods best describes it.*

_____ I. Randomly select 10 members from each of the WW centers in the Maryland division.

_____ II. Use an alphabetical listing of all Maryland division members. Randomly choose a starting person on the list. Then select every 20th person thereafter.

_____ III. Randomly select 2 or 3 branches of the Maryland division and survey every member of that center.

_____ IV. Send out the survey to every member of the Maryland division. Place drop boxes in each WW center. Anyone who returns a survey will be in the sample.

_____ V. The Maryland regional office is in Baltimore so they survey members at the WW center in Baltimore.

_____ VI. From a numbered list of all Maryland division members use a computer to randomly select 100 numbers and survey all members with those corresponding numbers.

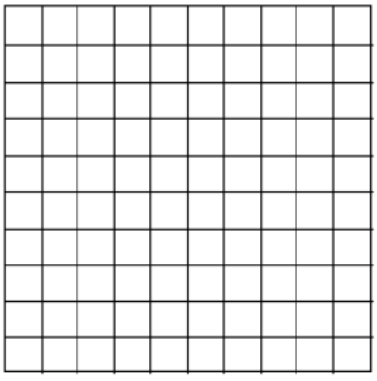
2. What is the population of interest in the WW situation?

3. Hilary wonders if people of similar heights tend to date each other. She measures herself, her dormitory roommate, and the women in the adjoining rooms; then she measures the next man each woman dates. Here are the data (heights in inches):

Women:	66	64	66	65	70	65
Men:	72	68	70	68	74	69

A. Construct a scatterplot of the data.

B. Describe the association between the heights of the women and the men they date.



Here is a formula that is used often in AP Statistics: $z = \frac{x - \bar{x}}{s}$

1. If z = 2.5, x = 102, and $\bar{x}$ =100. What is s? Show your work.

2. If z = -3.35, x = 60, and s = 4, what is $\bar{x}$? Show your work.

3. Solve $0.05 = 1.96\sqrt{\frac{0.5^2}{n}}$ for n.

4. If $-1.64 = \frac{60 - \mu}{\sigma}$ and $1.96 = \frac{95 - \mu}{\sigma}$, solve for μ and σ.

It is expected that you have a thorough understanding of linear functions.

1. The USDA reported that in 1990 each person in the United States consumed an average of 133 pounds of natural sweeteners. They also claim this amount has decreased by about 0.6 pounds each year.
    a. Write a linear equation that relates years since 1990 to the average consumption of natural sweeteners. Define your variables.

    b. What is the slope and what is the y-intercept?

    c. Predict the average consumption of sweeteners per person for the year 2005.

2. The following equation can be used to predict the average height of boys anywhere between birth and 15 years old: $y = 2.79x + 25.64$, where x is the age (in years) and y is the height (in inches).

    a. What does the slope represent in this problem? Interpret it in context.

    b. What does the y-intercept represent in this problem? Interpret it in context.

You are expected to have a basic understanding of simple probability.

1. A special lottery is to be held to select the student who will live in the only deluxe room in a dormitory. There are 100 seniors, 150 juniors, and 200 sophomores who applied. Each senior's name is placed in the lottery 3 times; each junior's name, 2 times; and each sophomore's name, 1 time. What is the probability that a senior's name will be chosen?

   A. $\frac{1}{8}$ B. $\frac{2}{9}$ C. $\frac{2}{7}$ D. $\frac{3}{8}$ E. $\frac{1}{2}$

2. Which of the following has a probability closest to 0.5?

   A. The sun will rise tomorrow.
   B. It will rain tomorrow.
   C. You will see a dog with only three legs when you leave the room.
   D. A fair die will come up with a score of 6 four times in a row.
   E. There will be a plane crash somewhere in the world within the next five minutes.

3. If a coin is tossed twice, what is the probability that on the first toss the coin lands heads and on the second toss the coin lands tails?
   A. 1/6
   B. 1/3
   C. ¼
   D. ½
   E. 1

4. If a coin is tossed twice what is the probability that it will land either heads both times or tails both times?
   A. 1/8
   B. 1/6
   C. 1/4
   D. 1/2
   E. 1

5. Calculate the following probabilities and arrange them in order from least to greatest.

   I. The probability that a fair die will produce an even number. _____
   II. A random digit from 1 to 9 (inclusive) is chosen, with all digits being equally likely. The probability that when it's squared it will end with the digit 1. _____
   III. The probability that a letter chosen from the alphabet will be a vowel. _____
   IV. A random number between 1 and 20 (inclusive) is chosen. The probability that its square root will not be an integer. _____

   ORDER: _____, _____, _____, _____