Name:_____ Date:_____ AP Statistics

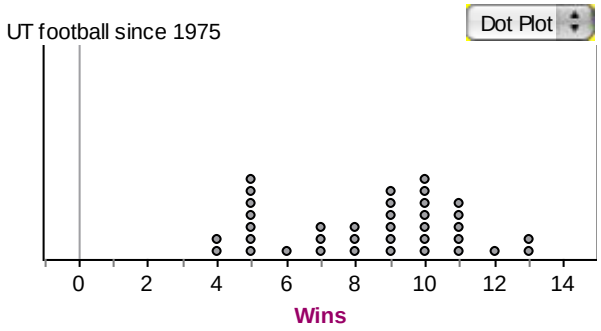
Displaying Distributions with Graphs

A dotplot is a statistical chart (graph) consisting of a group of data points plotted on a simple scale. Typically each number receives one column.

Example 1: The number of wins UT football has had since the 1975 season.

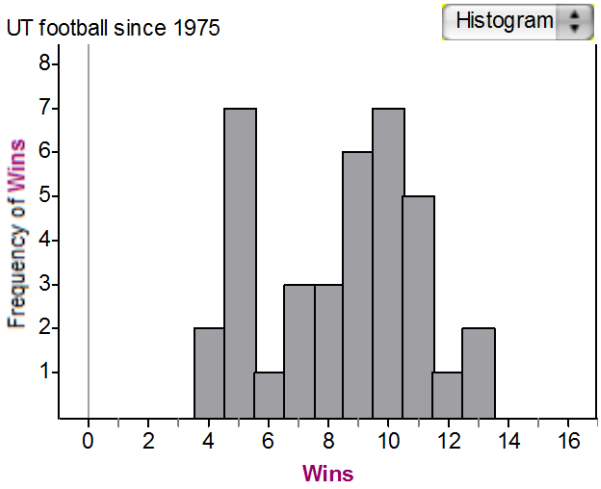
5 13 12 10 10 13 11 10 11 11 9 9 9 4 8 10 8 5 5 6 5 10 5 4 7 5 8 7 11 9 10 7 9 9 11 5 10

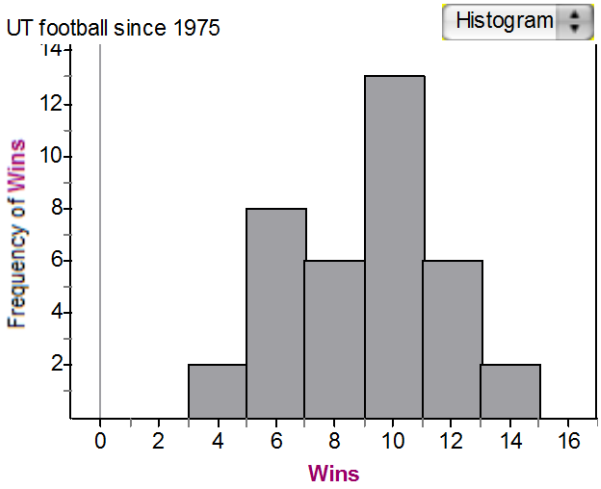
Below is the display of the data as a dotplot. Notice the dotplot has one stack of dots for each number. Also notice there is not vertical axis. Describe the important parts of the distribution.



A histogram is a graphical display consisting of a group of continuous data points displayed as frequencies grouped into fixed range "bins."

Below are two different histograms, but the same data. The first has a "bin" width of 1, so this matches our dotplot, but the second has a bin width of 2. Notice there is now a vertical axis of frequency.





We must be mindful of how the bin width can change our perception of the shape.

A **frequency distribution** is a table that displays the categories, frequencies, relative frequencies and/or cumulative relative frequencies.
The **frequency** for a particular category is the number of observed responses that fall into that category.
The corresponding **relative frequency** is the fraction or proportion of observed responses in the category.
The **cumulative relative frequency** is the fraction or proportion of observed responses in all categories so far including the current.

We will convert the second histogram into a table to explore relative frequency and cumulative relative frequency. Notice, so 5 actually falls into the next category.

Categories are written 3– <5, which mean all numbers from 3 to just less than 5.

This matches the frequency on histogram.

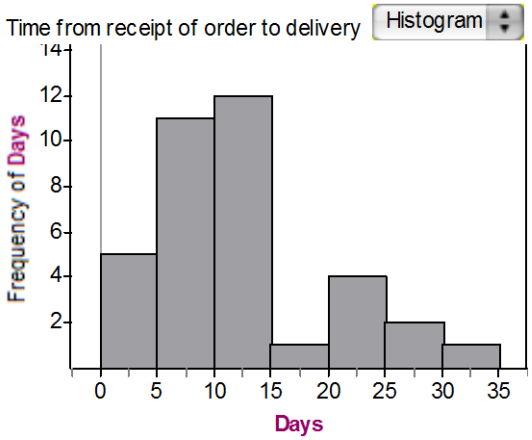
Convert the frequency to a decimal for relative frequency.

This is a running total of the relative frequency with 1 as the last number (100%).

Category	Frequency	Relative frequency	Cumulative relative frequency
3 – < 5	2	2/37 = .0541	.0541
5 – < 7	8	8/37 = .2162	(.2162 + .0541) = .2703
7 – < 9	6	6/37 = .1622	(.1622 + .2703) = .4325
9 – < 11	13	13/37 = .3514	(.3514 + .4325) = .7839
11 – < 13	6	6/37 = .1622	.9461
13 – < 15	2	2/37 = .0541	1

A **cumulative relative frequency** tells us the percent of data that fall from that category and less. In other words, we know 78.39% of the UT seasons have less than 11 wins.

Example 2: Describe the histogram below and then fill out the corresponding chart below.



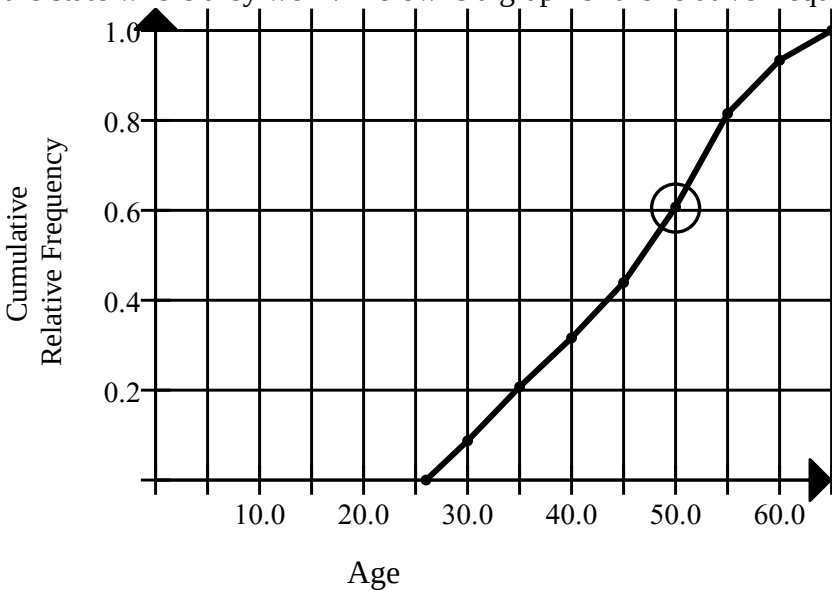
Describe the important parts of the distribution.

Category	Frequency	Relative freq.	Cumulative relative freq.
0 – < 5			

How can your cumulative relative frequency help you verify that you stated the center correctly?

The cumulative relative frequencies can be displayed as a histogram or in a graph like below.

Example 3: The 7th edition of *Who's Who Among America's Teachers*® honors approximately 134,000 outstanding teachers in four separate regional volumes. Teachers are listed alphabetically in the volume of the state where they work. Below is a graph of the relative frequencies.



a) What percent of the teachers chosen were less than 35 years old?
b) What percent of the teachers chosen were between 50 and 60?
c) What percent of the teachers chosen were older than 50?

d) Interpret the circled value in context.

STEMPLOTS

Unlike histograms, stem & leaf plots retain the original observations. It is an effective way to organize data without expending much effort. Each observation is regarded as consisting of two pieces. One or more of the leading digits make up the **stem** and the trailing digit constitutes the **leaf**.

Example 4: Construct a stemplot for the following typing speeds (net words per minute)

68	72	91	75	63	55	65	35	47	52
84	45	58	61	69	22	46	55	66	71

For example, we will take 68 and split into the stem and leaf 6 | 8 (stem: 6; leaf: 8)

To build the stemplot, we identify the lowest and highest numbers. This will guide to set the smallest and largest stems. So, since the lowest number is 22, our smallest stem is 2, and the highest number is 91, our largest stem is 9. Now, add each data point by finding the appropriate stem and writing the matching leaf.

All the 60s will be her

Start	Add 68	Add 72	Add 91	Add 75	Done with 1st row. You finish.
9 |	9 |	9 |	9 | 1	9 | 1	9 | 1
8 |	8 |	8 |	8 |	8 |	8 |
7 |	7 |	7 | 2	7 | 2	7 | 2 5	7 | 2 5
6 |	6 | 8	6 | 8	6 | 8	6 | 8	6 | 8 3 5
5 |	5 |	5 |	5 |	5 |	5 | 5 2
4 |	4 |	4 |	4 |	4 |	4 | 7
3 |	3 |	3 |	3 |	3 |	3 | 5
2 |	2 |	2 |	2 |	2 |	2 |

Example 5. The following table lists the running times (in minutes) of the videotape versions of 22 movies directed by Alfred Hitchcock. Construct a stemplot of this distribution. REMEMBER, leaf is typically only one digit while the stem can be as many digits as needed.

FILM	TIME	FILM	TIME
The Birds	119	Psycho	108
Dial M for Murder	105	Rear Window	113
Family Plot	120	Rebecca	132
Foreign Correspondent	120	Rope	81
Frenzy	116	Shadow of a Doubt	108
I Confess	108	Spellbound	111
The Man Who Knew Too Much	120	Strangers on a Train	101
Marnie	130	To Catch a Thief	103
North by Northwest	136	Topaz	126
Notorious	103	Under Capricorn	117
The Paradine Case	116	Vertigo	128

BACK TO BACK STEMPLOTS

Stemplots can often be placed back to back to facilitate a comparison.

Here are the number of home runs that Babe Ruth hit in his 15 years with the New York Yankees, 1920 to 1934:

54 59 35 41 46 25 47 60 54
46 49 46 41 34 22

For the second data set, construct a stemplot on the opposite (left) side. Babe Ruth's home run record for a single year was broken by another Yankee, Roger Maris, who hit 61 home runs in 1961. Here are Maris's home run totals for his 10 years in the American league.

13 23 26 16 33 61 28 39 14 8

Fill in the following table:

Type of graph	Advantages	Disadvantages
Dotplot		
Histogram		
Stemplot		