Statistics is a way of reasoning, along with a collection of tools and methods, designed to help us understand the world.

Statistics is about

Data are values along with their context. Context is provided by the "W"s

Why do we care about the data? (What are you trying to figure out?)

Who are described by the data? (Which individuals or things?)

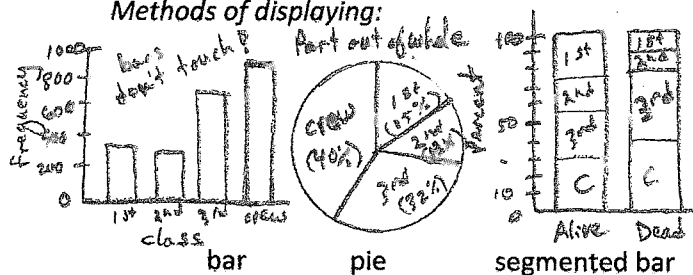**What** variables do the data contain? (What questions were asked?)

When ⎤
Where ⎬ was the data collected?
How ⎦

(Necessary)

Data is either:

For histograms:
$$\# \text{ of bars} = \frac{\# \text{ observations}}{5}$$
$$\text{bar width} = \frac{max - min}{\# \text{ of bars}}$$

or   of individuals into groups or categories.

Methods of displaying:

bar   pie   segmented bar

Numerical Summaries:

Contingency tables - categorize the individuals on all variables at once to reveal possible patterns.

Class

Survival		First	Second	Third	Crew	Total
	Alive	203	118	178	212	711
	Dead	122	167	528	673	1490
	Total	325	285	706	885	2201

		First	Second	Third	Crew	Total
Alive	Count	203	118	178	212	711
	% of Row	28.6%	16.6%		29.8%	100%
	% of Col.	62.5%	41.4%		24.0%	32.3%
	% of Table	9.2%	5.4%		9.6%	32.3%
Dead	Count	122	167	528	673	1490
	% of Row		11.2%		45.2%	
	% of Col.		58.6%		76.0%	
	% of Table		7.6%		30.6%	
Total	Count	325	285	706	885	2201
	% of Row	14.8%		32.1%		100%
	% of Col.	100%		100%		
	% of Table	14.8%		32.1%		

If rows (columns) of contingency table have equal distributions, then variables are
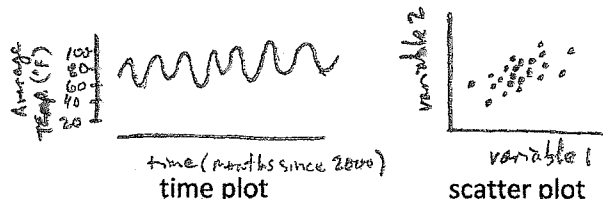o.w. they are associated. (see segmented bar)

that measure some characteristic of each individual.

Methods of displaying:

pulse rate
4 | 4 = 44 beats/min

stem & leaf   histogram   box plots
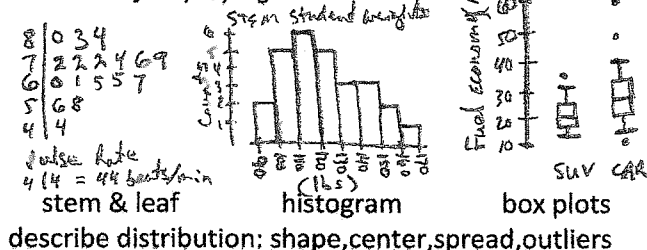
describe distribution: shape, center, spread, outliers

time (months since 2000)
time plot   scatter plot

Numerical Summaries:

5 number summary ————————→ Max

If shape is

spread =   center =

outlier if > Q3 + (1.5)(IQR)

   or < Q1 - (1.5)(IQR)   Min

If shape is

center =   spread =

enables comparison of   to

_____ is ____ " tall and _____ is ____ "

Using the   as a ruler we create z-scores that tell us how unusual a value is, in from its   Then we use models (mathematical curves) to change z-scores to

So 2 SDs above the mean places you in the ____