

Categorical vs Quantitative
Which Graph with which type
What is wrong with this graph
 Missing parts
 Broken rules
Make a pie
 Calculate percents
Distributions
 Identify them
 Understand them

- Types of graphs
- How types of graphs and data relate
- Elements of a graph
- Handling one variable vs two variables
 - o Distributions
 - o Scatterplots
- Good graphs vs bad graphs

Two kinds of data:

Variables: Values that describe a feature of the things you are talking about.

Categorical

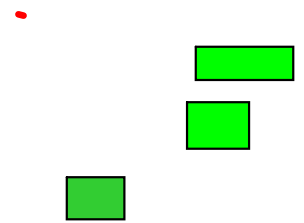
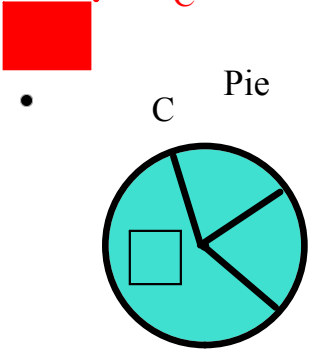
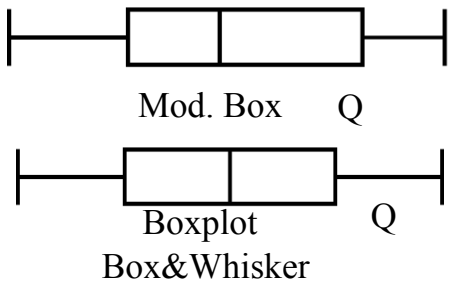
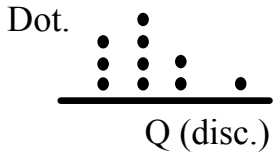
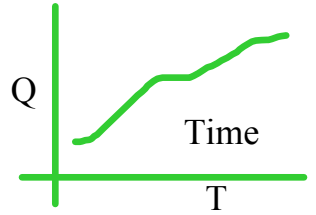
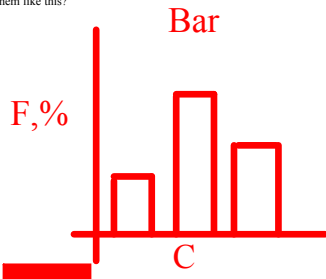
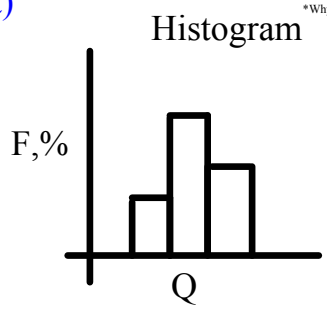
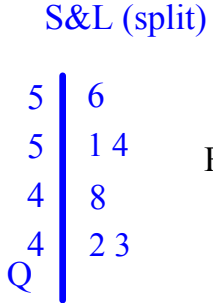
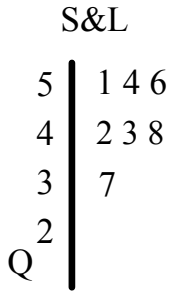
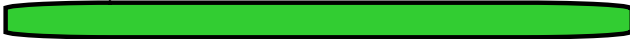
Quantitative

Categorical variable: The feature described is a group that the thing* belongs to.

Quantitative variable: The feature described is measured or counted.

* The right word for each "thing" is "individual."

Graphs: 1) Some kinds are for C and some for Q
 2) Four of these graphs are basically the same thing, which 4



Statisticians present

Tables

and/or

Graphs

- 1) That summarize the data to make a point/tell a story.
- 2) With commentary to tell people why you are showing them that table or graph.

(If it doesn't tell a story, don't make it)

Make a graph

Table: Summarizing numbers with numbers

Fund Type	Number	Percent
US Equity ETF	407	36%
Global Equity ETF	321	29%
Fixed Income ETF	141	13%
Commodity Based ETF	231	21%
Asset Allocation ETF	17	2%
	1117	100%

Table of a categorical variable

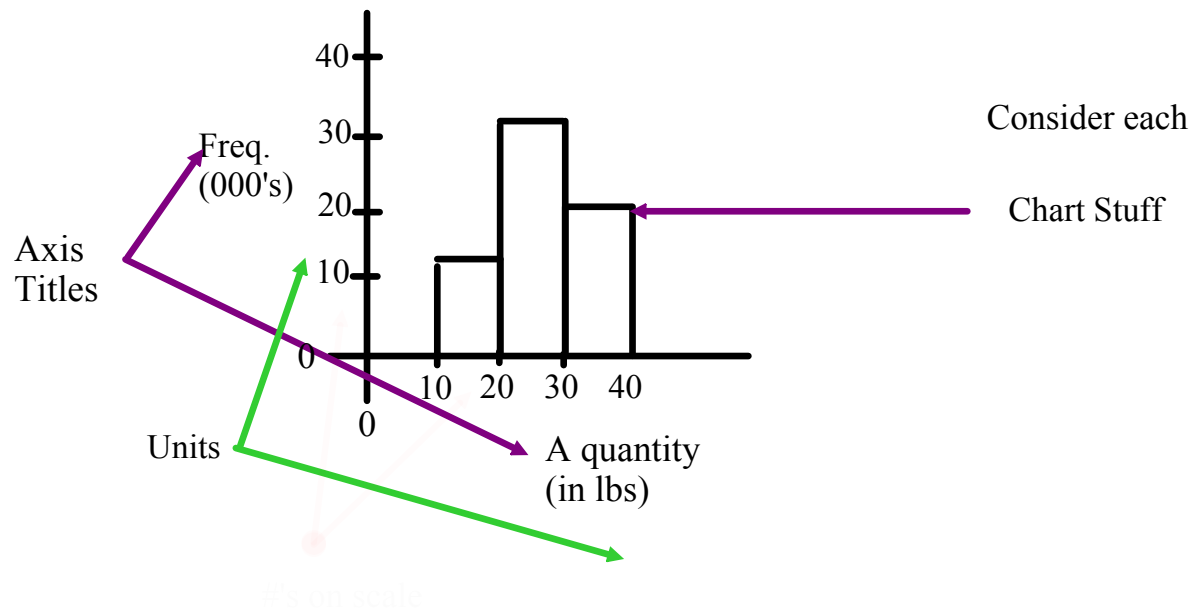
Funds with names:	Number	Percent
Longer than 60 less shorter than 80 characters	44	4%
Longer than 40 less shorter than 60 characters	432	39%
Longer than 20 less shorter than 40 characters	626	56%
Longer than 0 less shorter than 20 characters	15	1%
	1117	

Companies providing:	Number	Percent
More than 15% less than 20% of the funds	1	0%
More than 10% less than 15% of the funds	2	0%
More than 5% less than 10% of the funds	3	0%
Less than 5% of the funds	30	3%
	36	

Company	Number	Percent
AdvisorShares	10	1%
ALPS	2	0%
Claymore	4	0%
Cohen	1	0%
Columbia	5	0%
DBX	10	1%
Direxion	53	5%
EGShares	17	2%
ESG	2	0%
ETFS	7	1%
FactorShares	5	0%
Fidelity	1	0%
First	60	5%
Focus™	15	1%
Global	33	3%
GlobalX	2	0%
GreenHaven	1	0%
Guggenheim	37	3%
HOLDERS	18	2%
IndexIQ	14	1%
iShares	225	20%
Jefferies	4	0%
JETS	1	0%
Market	33	3%
PIMCO	14	1%
PowerShares	127	11%
ProShares	121	11%
REVENUESHARES	6	1%
RussellETF	17	2%
Rydex	36	3%
Schwab	14	1%
SPDR	101	9%
Teucrium	3	0%
United	9	1%
Vanguard	64	6%
WisdomTree	45	4%

Title

Histogram for demo
of the 6 parts



Good graphs

- Labels and legends:

 - Give the reader clear, complete information

 - Don't make your reader guess, help them understand

- KISS => Keep it simple, student

 - Let well organized data speak for itself

 - Don't add art

 - Don't put more info than the reader can relate.

 - Be really careful with info about more than 2 variables

 - X-axis, y-axis, and color

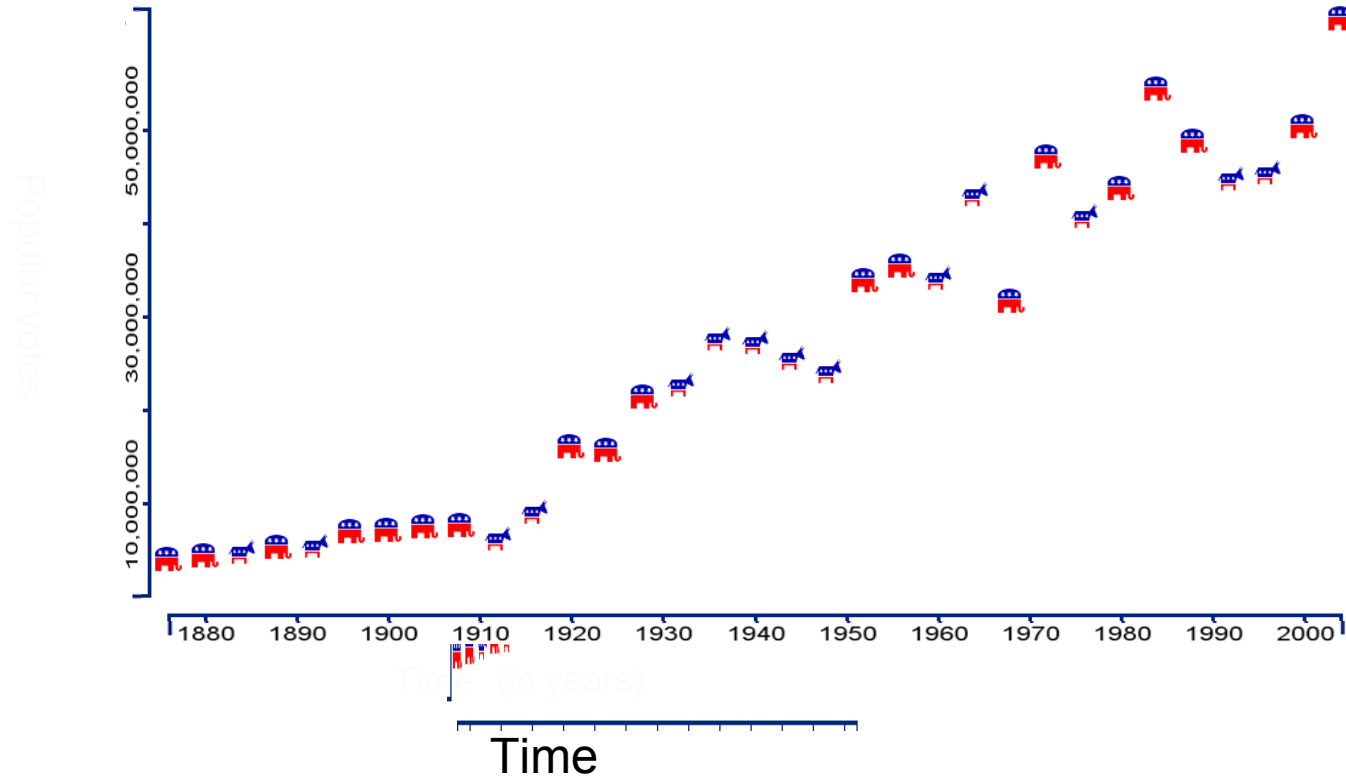
 - Labels giving new info

 - Actually 3d graphs

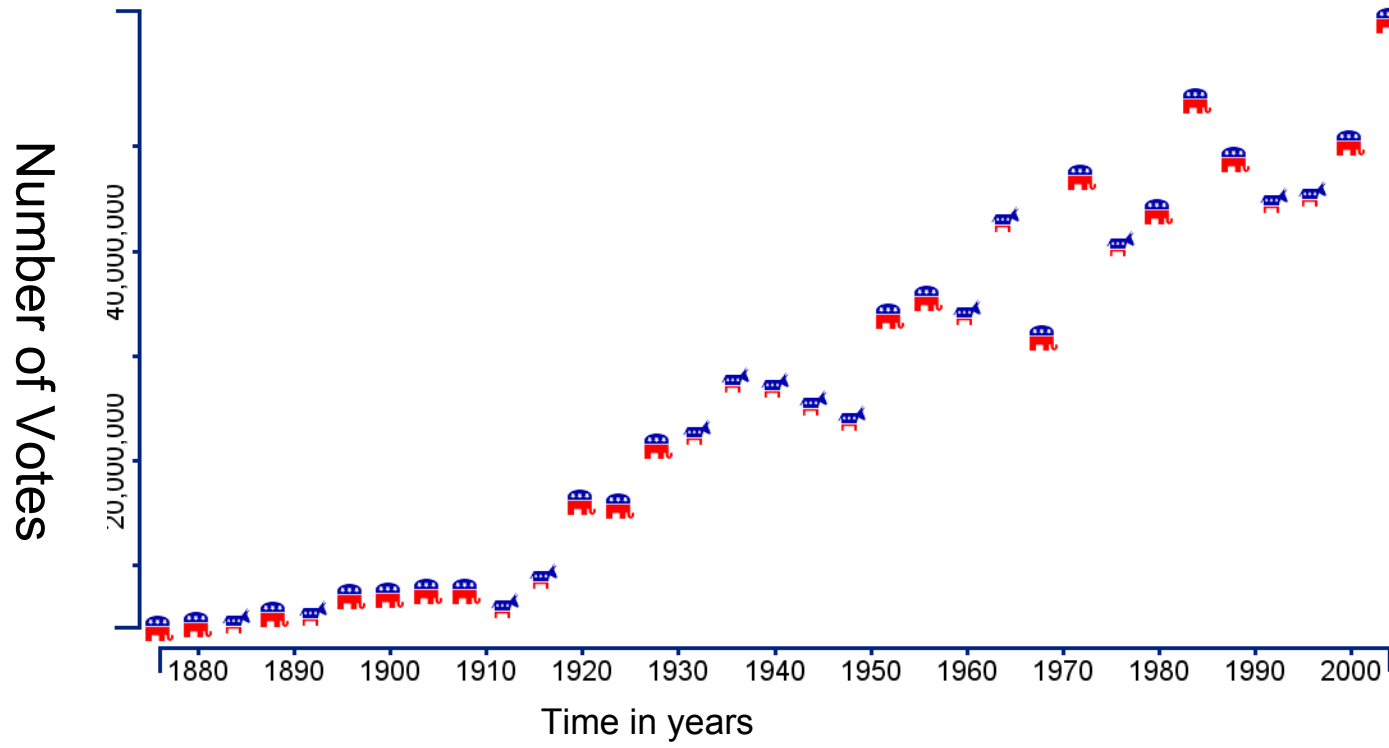
<http://www.gapminder.org/>



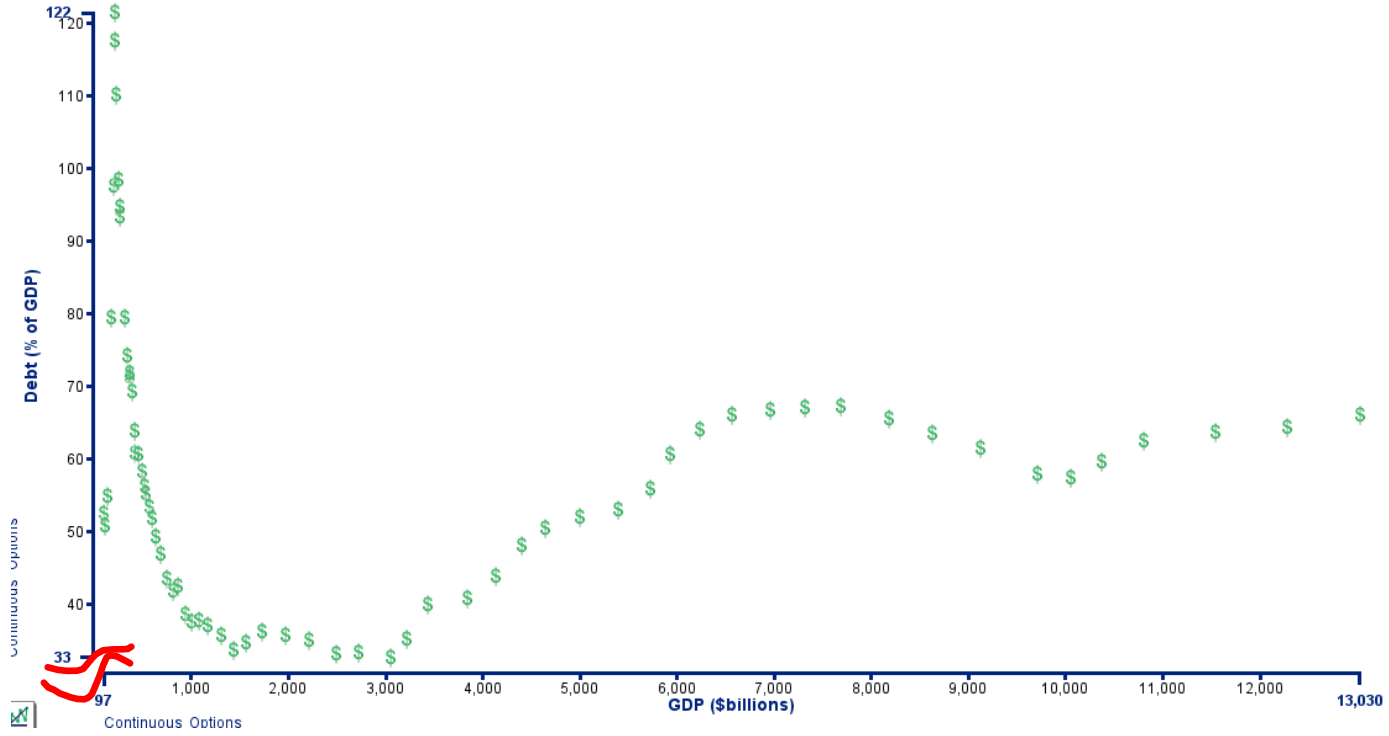
Votes Received by Winner of Presidential Election

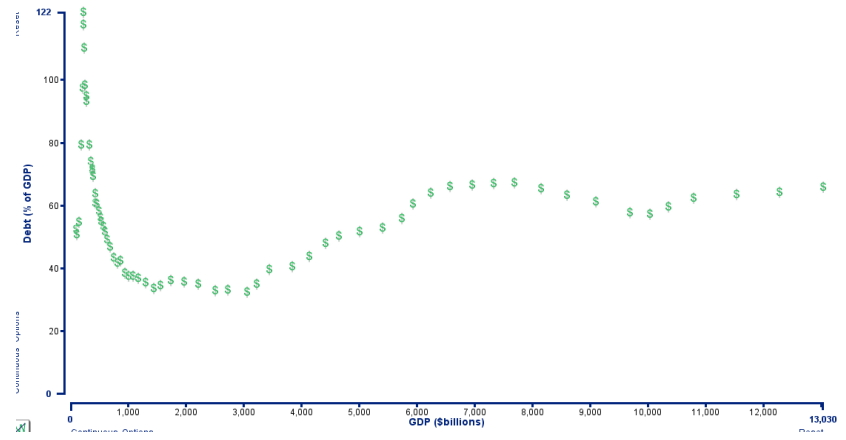
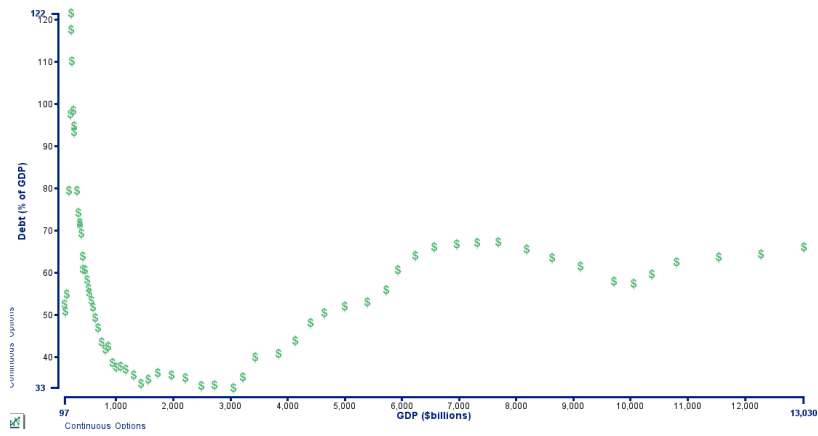


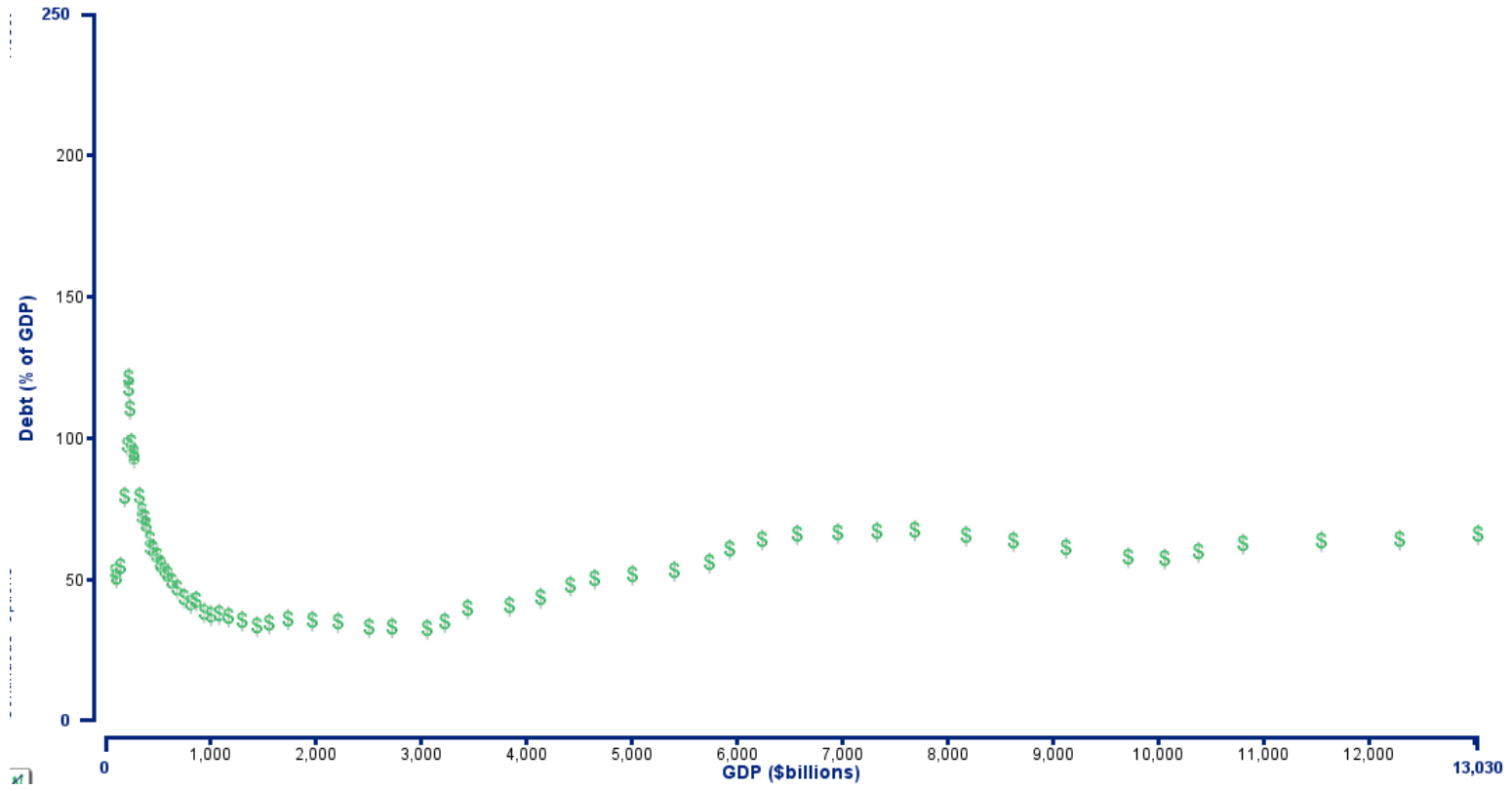
Votes Received by Winner of Presidential Election



National Debt as a Percent of GDP vs. GDP







Do not distort scales

- Look for a square
- Keep the data filling the X and Y direction
- Beware breaks
- Beware pictographs

Most basic story - ONE VARIABLE:

The distribution

A distribution does two things:

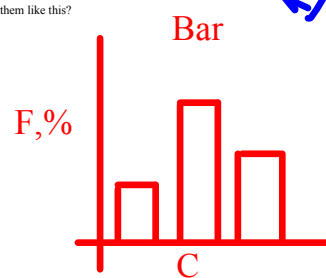
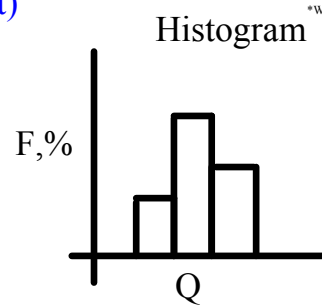
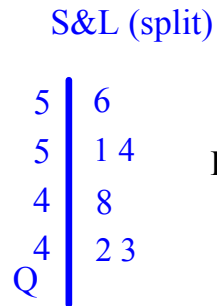
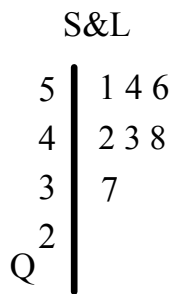
- 1) Shows* all the values a variable can have
- 2) How often it takes each value

How do the tables relate to the idea of distribution?

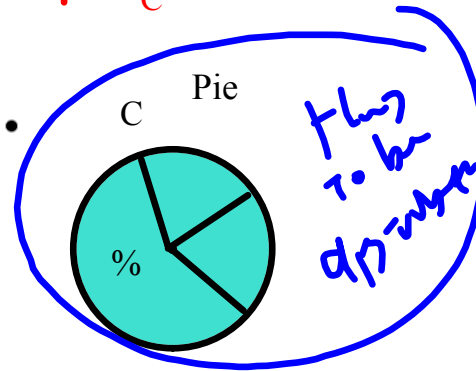
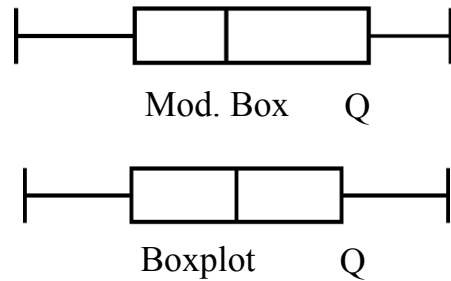
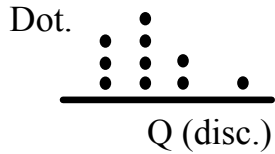
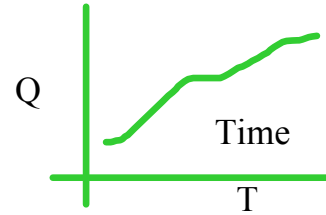
* It does not have to list every value.

What kind of graph(s) suit this data? Could you make the graph?

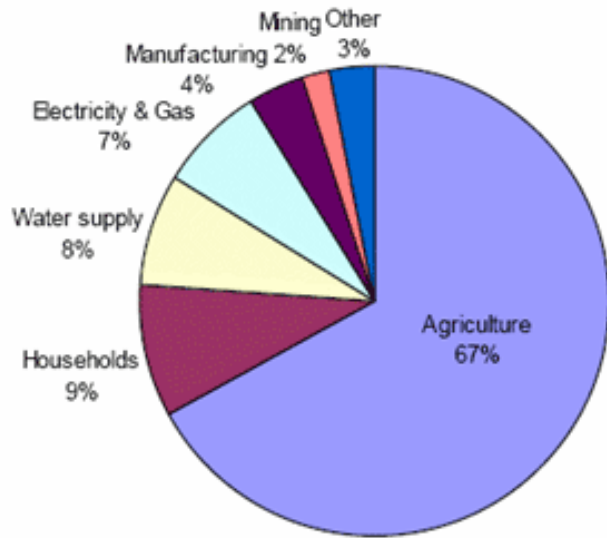
Which can be used to show distributions?
Which one must show a distribution?



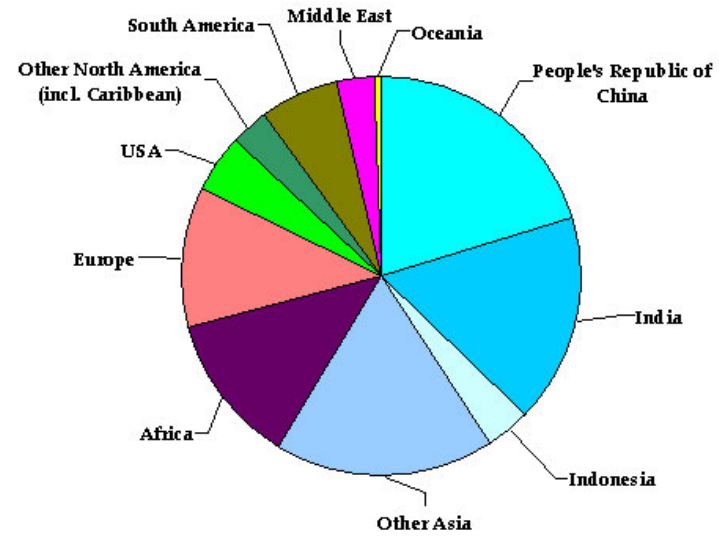
Both



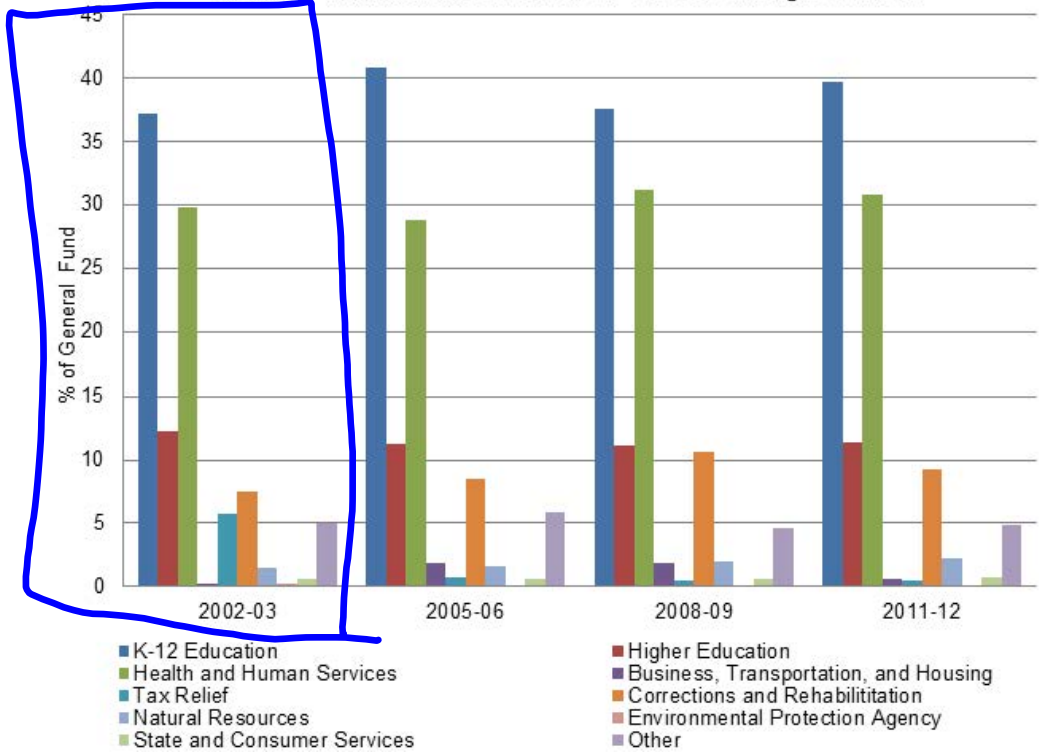
Distribution of Water Use



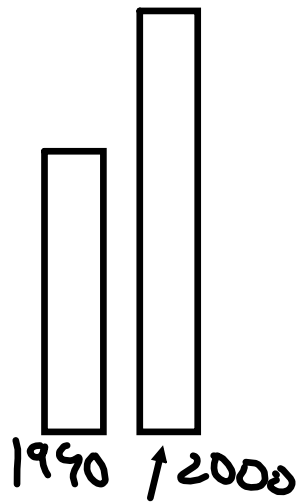
Distribution of World Population in 2005



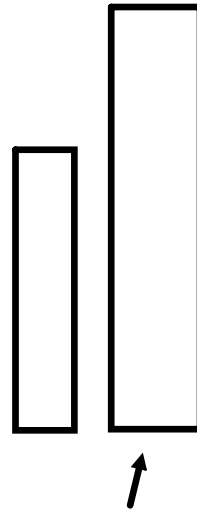
Percent of General Fund Appropriated for Major Policy Areas
 In Three-Year Intervals between FY 2002-03 through 2011-12



Warnings:
Avoid Pictograph.

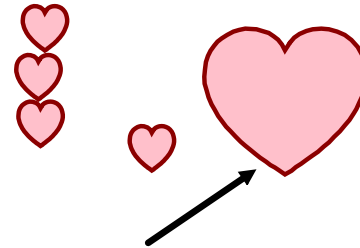


Just 1.5x
as tall,
has 1.5x
the area.



1.5x as tall
and 1.5x as wide
has 2.25 the area

Valentines day Sales
1990 vs 2000



How much greater a quantity is
this heart meant to represent?

3x as tall and wide
9x the area.



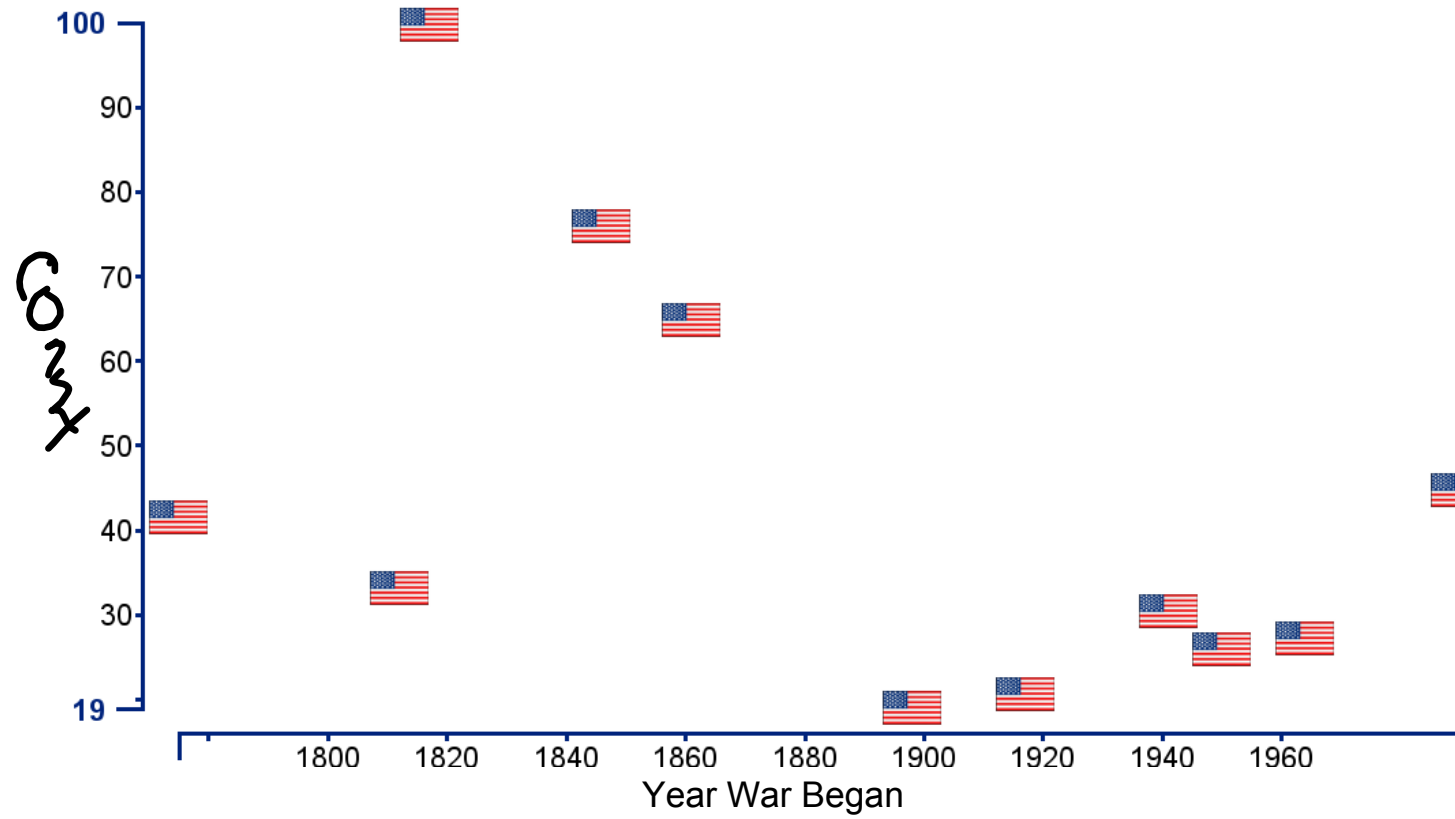
Time line is really a scatter plot
- where one variable is time
- where the points are connected

The connection suggests a trend, that time affects the variable

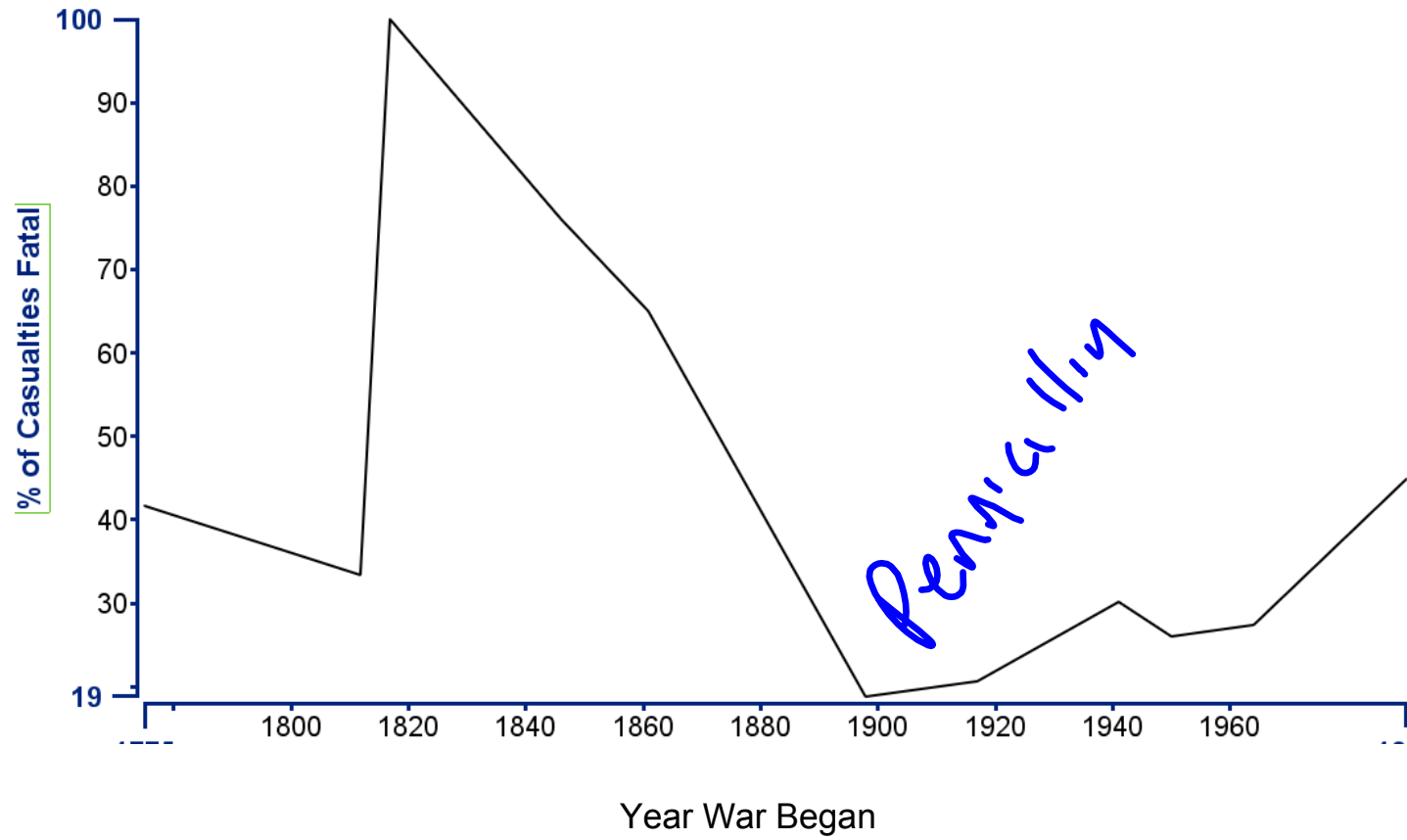
Our brains like that idea

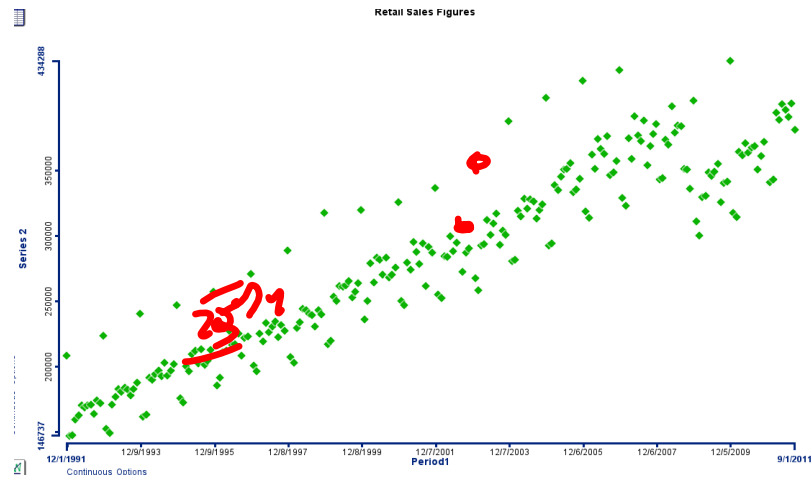
Beware

% of Casualties Resulting in Death for American Wars by Year the War Began

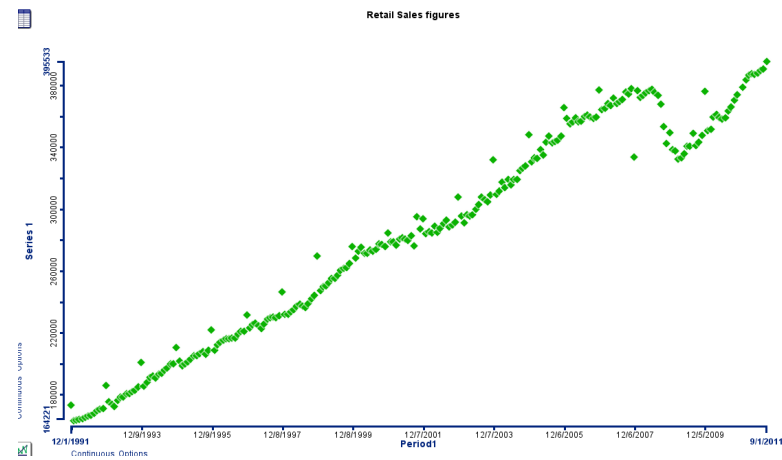


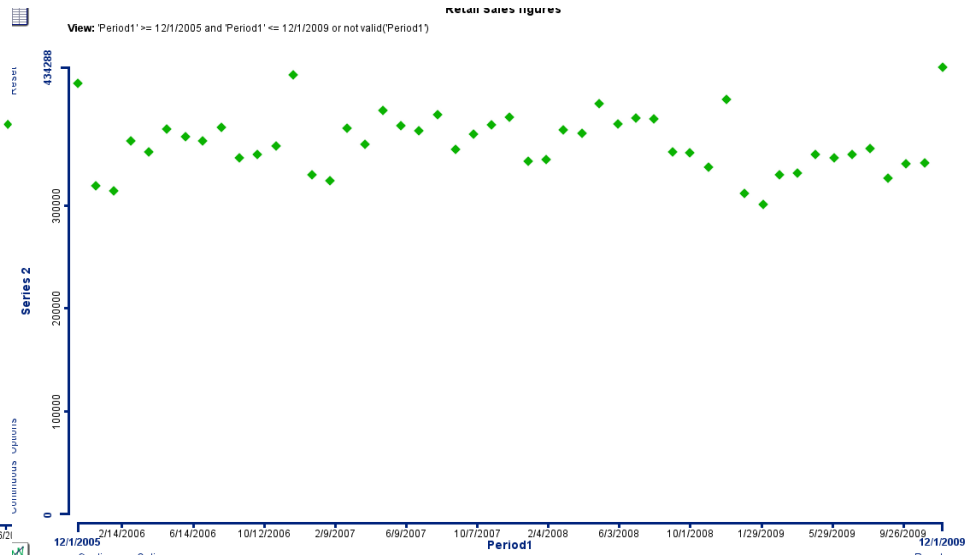
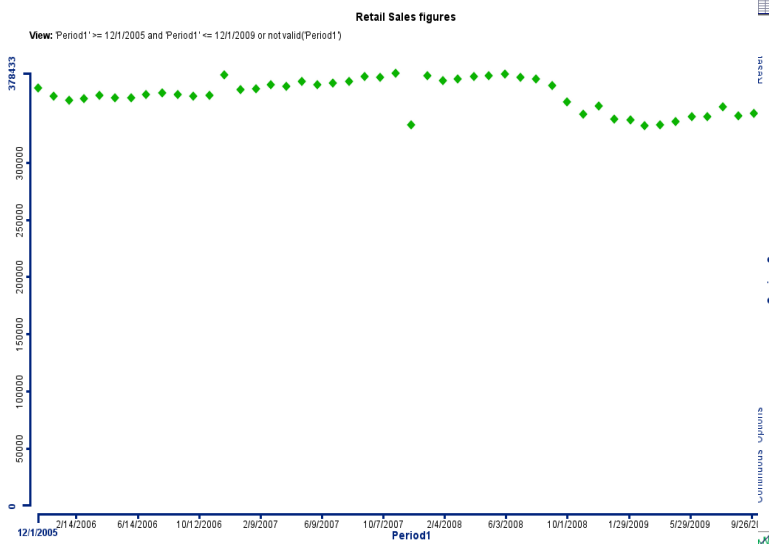
% of Casualties Resulting in Death for American Wars by Year the War Began





Which one is seasonally adjusted?





Which one is seasonally adjusted?

Time line

People look for

- Trend(s) /Patterns - This is the whole point
- Outliers/deviations - must be explained
- Seasonality - We adjust for it.

Your top story: Describe the pattern:

It goes up over time

It goes down over time

It stays the same over time.

Your second story: Exceptions, things that break the pattern.

It goes up except...

It goes down except...

It stays the same except...

The 1 Month Lending Rate Around 1990

