

WHERE SHOULD I START WHEN MAKING A VISUALIZATION OR MAP?

1 HAVE A QUESTION OR GOAL

What do you want to learn or find out?

What story or message do you want to tell?

2 DEFINE THE AUDIENCE

Who will use the information?

How will they use it?

interactive / online presentation
one pager
poster

least
↓
most information dense

Why will they use it?

- to learn
- to understand
- to make decisions
- as a platform to discuss data
- ...

3 EXPLORE & CLEAN THE DATA

Do the data make sense?

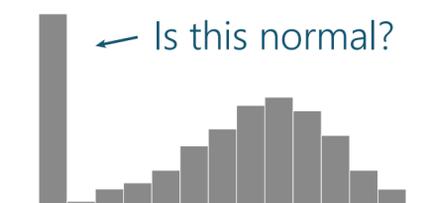
- How are the data distributed?
- Are there outliers?
- Are there missing data?
- Do the data fall within a reasonable range?

What do they mean?

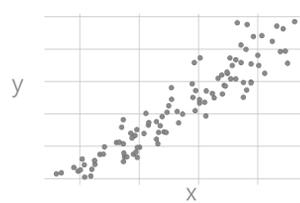
- Are the data related?

- Do new variables shed more insight?

Plot bar charts (categorical data) or histograms (numerical data)



Plot scatter or line plots between two variables



x seems to be positively correlated with y.

Transform data

- Average (point or running), calculate a percent, convert to comparable units
- Normalize, create ratios, reduce dimensions by calculating an index

4 DEFINE WHAT COMPARISONS TO MAKE

What do you want to show?

How do you want to show it?

Fill in the blanks!

I want to show the relationship between _____ and _____.

I want _____ to use this info to _____.

I want to represent this with <<plot type>>.

5 TEST IT OUT!

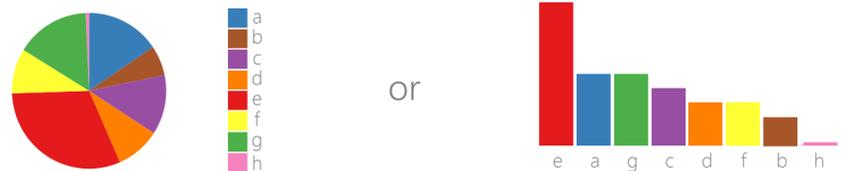
Sketch, make a mock-up, test it in your software, and refine it.

Sketch, test, refine

6 IS THIS THE BEST WAY TO REPRESENT THE INFO?

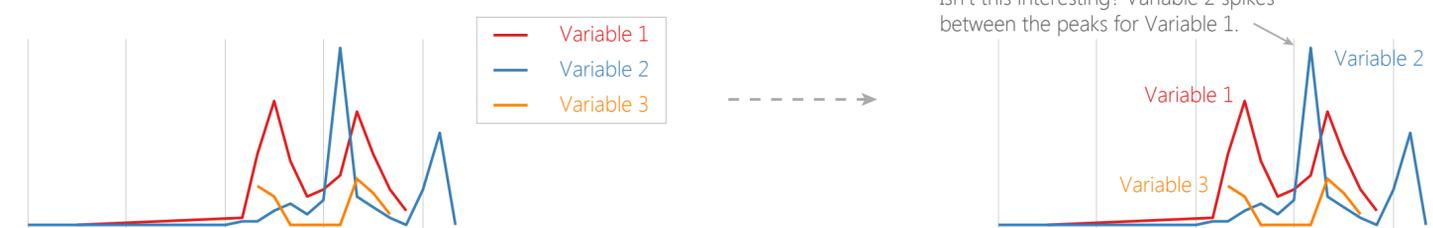
Is the plot successful? Refine it by asking yourself:

- Does the plot show the relationship I want?
- Is the plot type successful?
Is there a better way?
- Will it be useful to the audience?



- Can you understand the plot with little verbal explanation?

Annotations are your friend. Use them to explain how to read the graph, and/or what's interesting about it. Directly label things where possible. Only use legends if you have to.



- Is the plot a faithful representation of the data?

Plots can lie (or at least distort the truth). Don't do that.

- Is every dot, symbol, color, line, and variable necessary?

Keep things simple, consistent, and meaningful

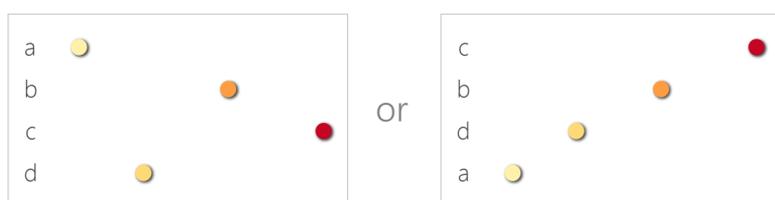


- Is the plot more effective as small multiples?



- How should things be ordered?

- alphabetically
- by ranked value
- by group or theme



- Should I group (average) the variables together?

Does the average smooth out noise, or does it wash away the signal?

