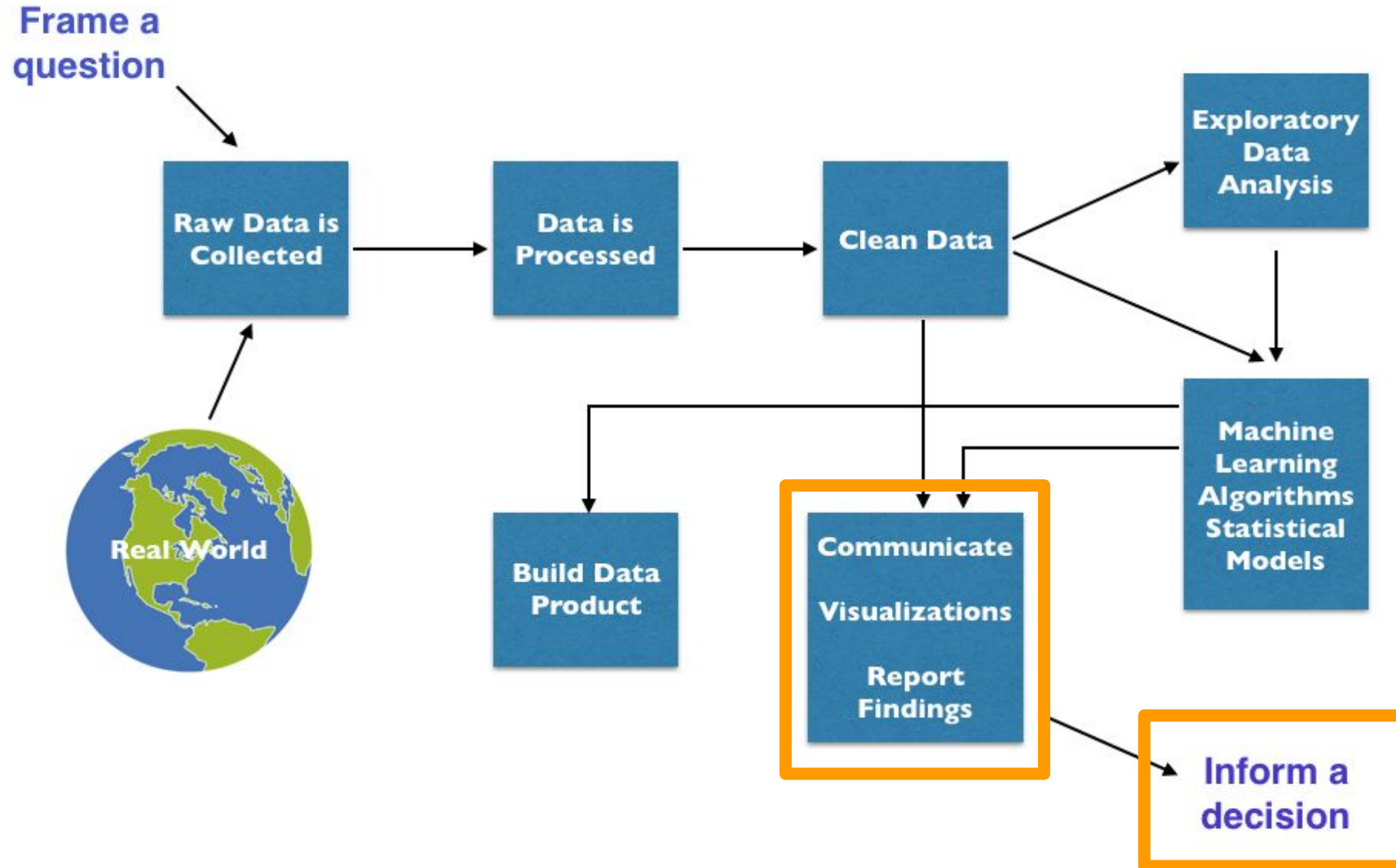


Introduction to Data Science

Data Science Essentials



Data Science Process



Data Storytelling:

The blending of data and human communication.

Data Storytelling

Think:

- What does my audience need to *know*?
- What does my audience need to *do*?

Stories can help make your message stick, moreso than just showing a bunch of plots.

Data Storytelling

Don't make your audience wait for it.

Consider leading with the recommendation or “big idea”.

This sets context for what you are going to show them and they know why they should care about what you are showing them.

<http://www.storytellingwithdata.com/blog/2014/07/lead-with-story>

Data Storytelling

Don't make your audience work for it.

Use preattentive attributes (color, size, added marks, and spatial position) to let your audience know where to direct their attention.

These can do two things:

- 1) direct the audience's eyes
- 2) establish a visual hierarchy of information

<http://www.storytellingwithdata.com/blog/2014/07/lead-with-story>

Effective **data visualization** can
make or break data storytelling.

Data Visualization Principles

Clarify – set a clear objective that people care about

Simplify – present only the visualization style that is most appropriate for the type of data being analyzed

Compare – display side-by-side comparisons for easy absorption

Attend – draw the viewer's attention to the important/relevant data

<https://www.clicdata.com/blog/the-few-the-proud-11-key-principles-of-effective-data-visualization/>

Clarify

You have probably created tons of plots throughout this project.

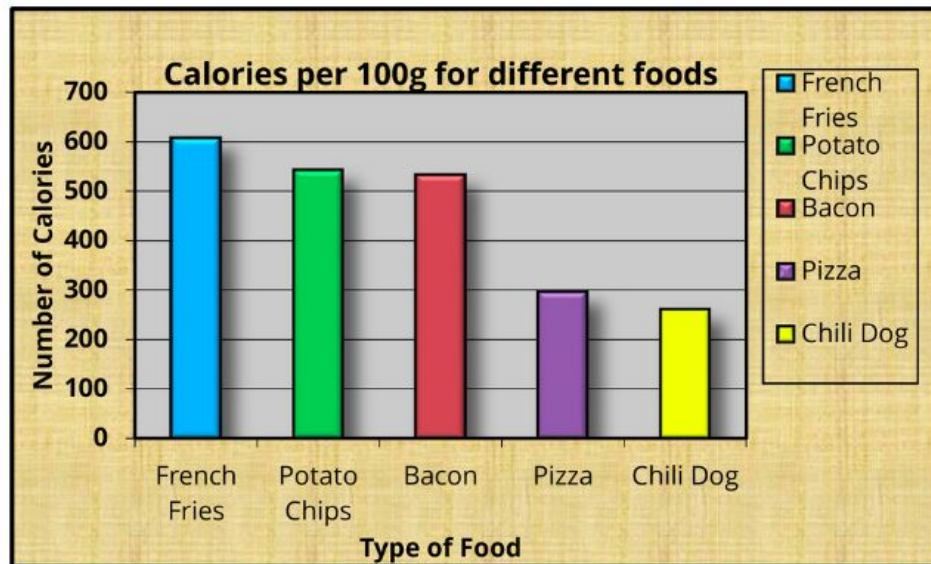
Pick the ones relevant to your story.

What is your message?

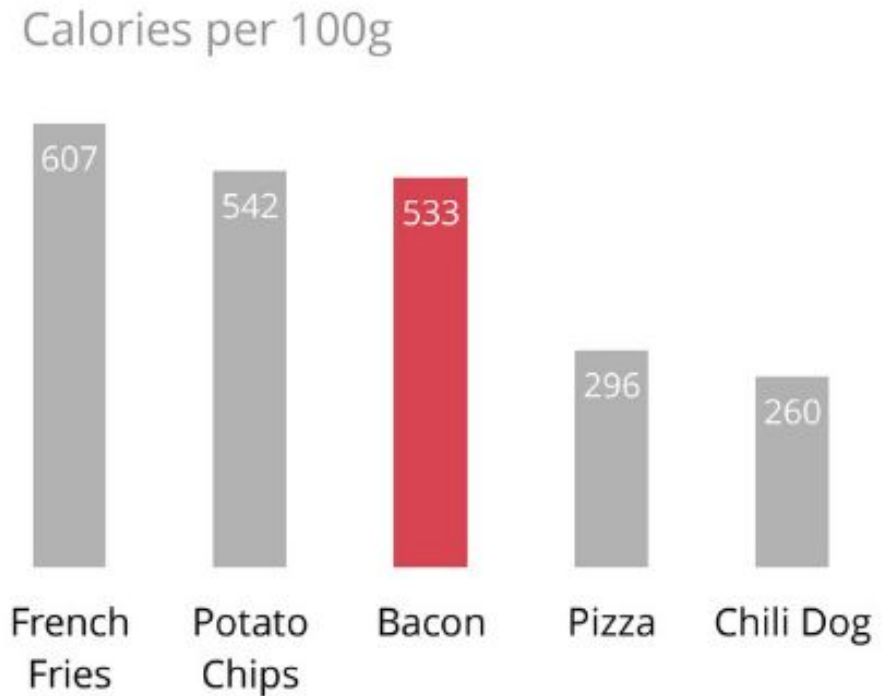
What should the key takeaways be?

Simplify

Maximize the **data-to-ink ratio**.



vs.



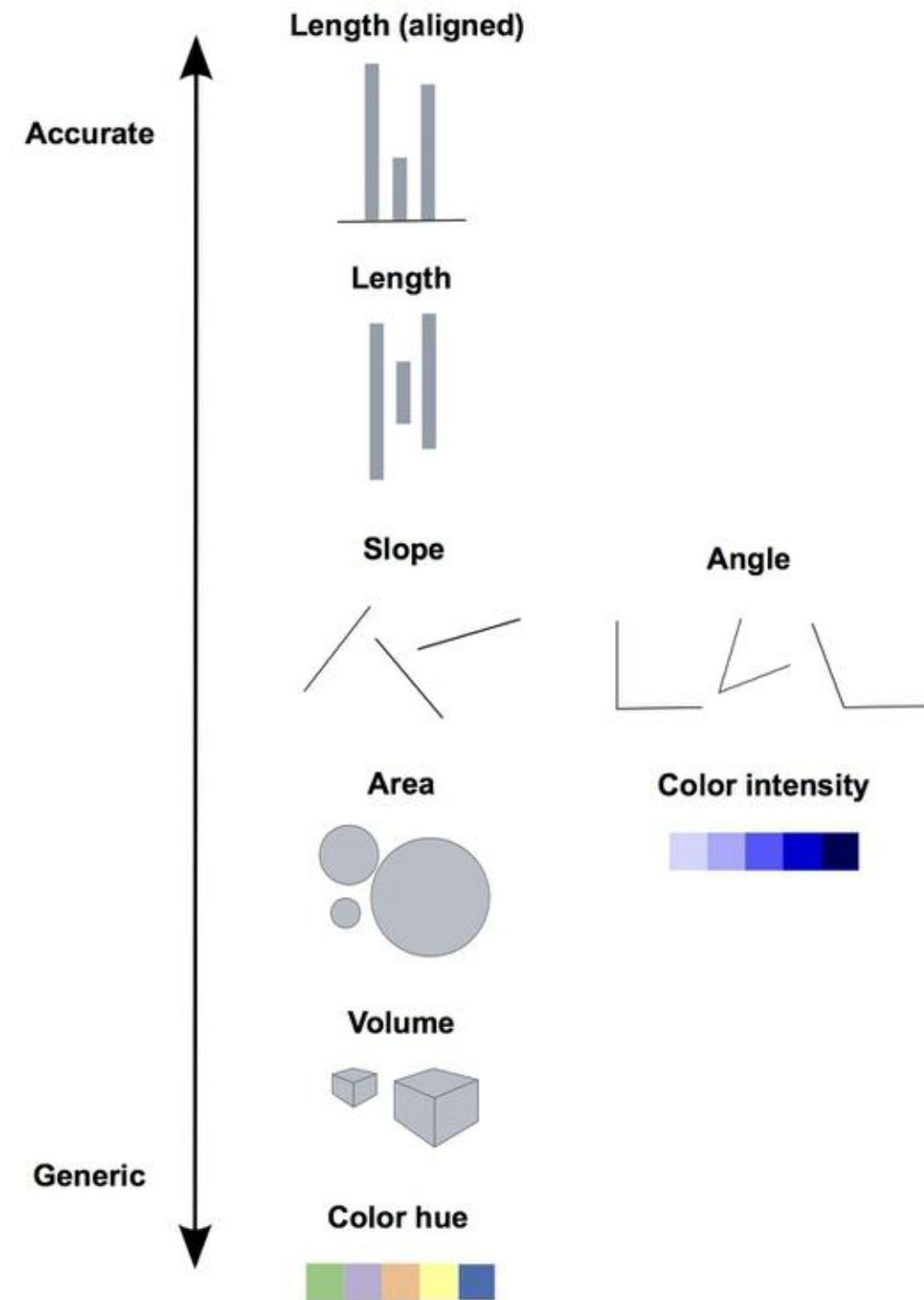
<https://www.darkhorseanalytics.com/blog/data-looks-better-naked>

Comparisons

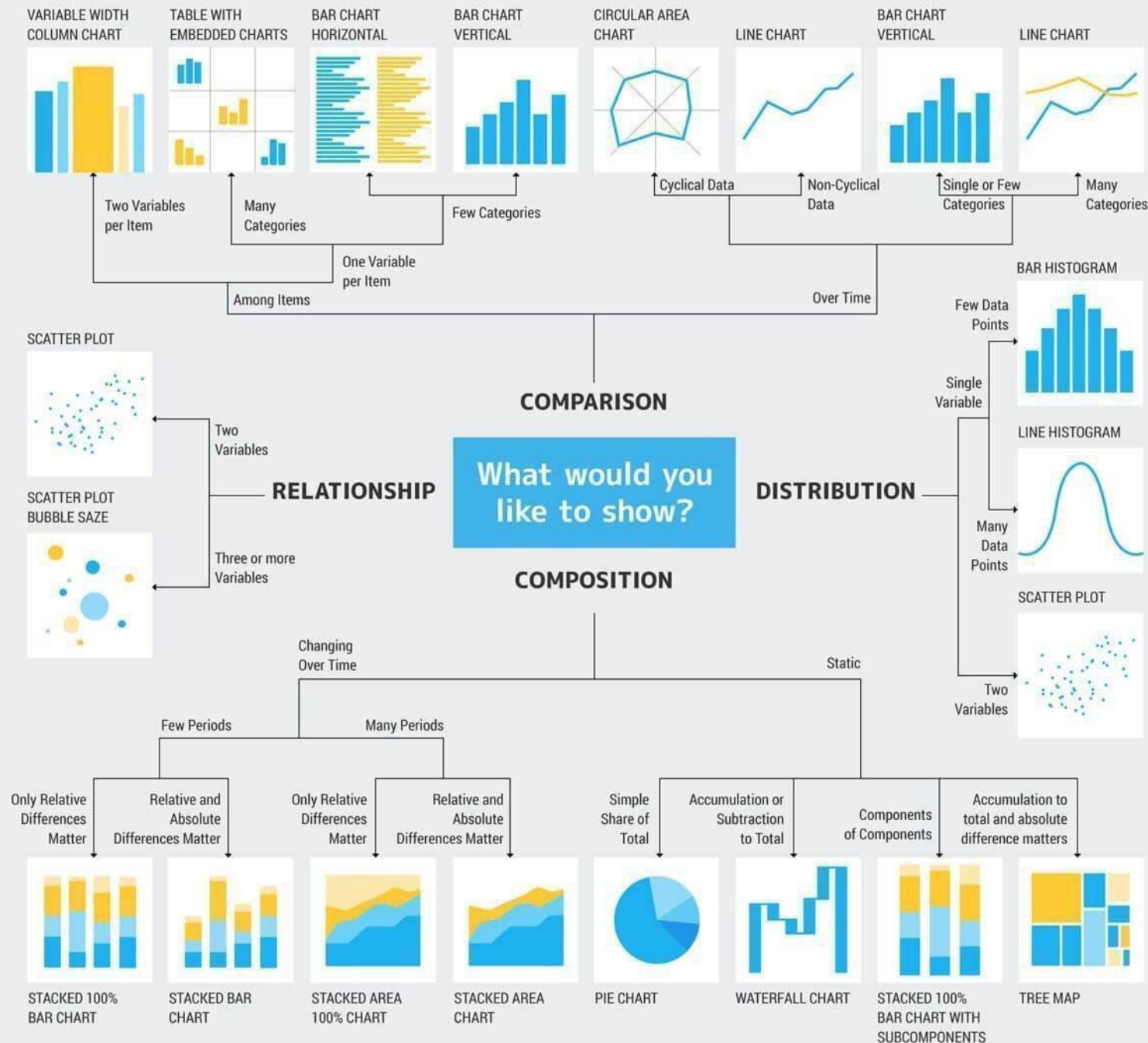
Graphical Perception and Graphical Methods for Analyzing Scientific Data. William S. Cleveland Robert McGill, *Science* 30 Aug 1985

Studied how accurately people were able to perceive quantitative information, based on how it was presented.

<https://paldhous.github.io/ucb/2016/dataviz/week2.html>



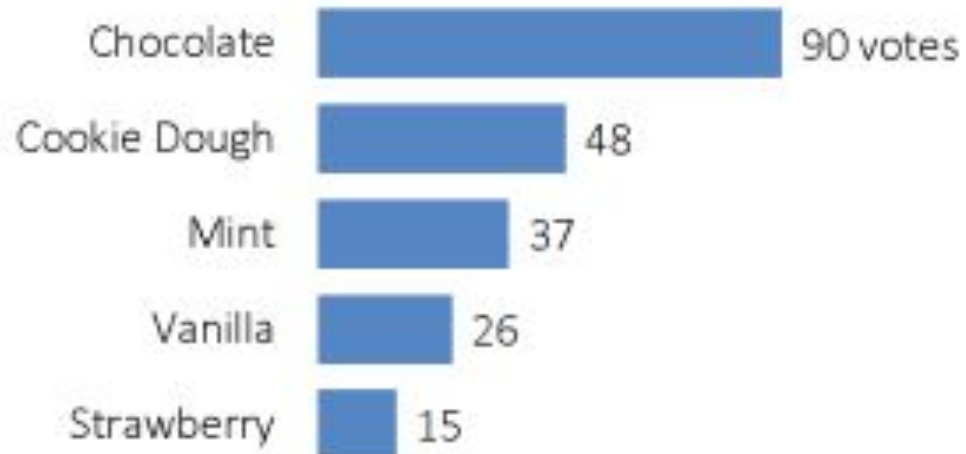
Which Type of Plot Should I Use?



<https://towardsdatascience.com/data-visualization-with-matplotlib-using-python-a7bfb4628ee3>

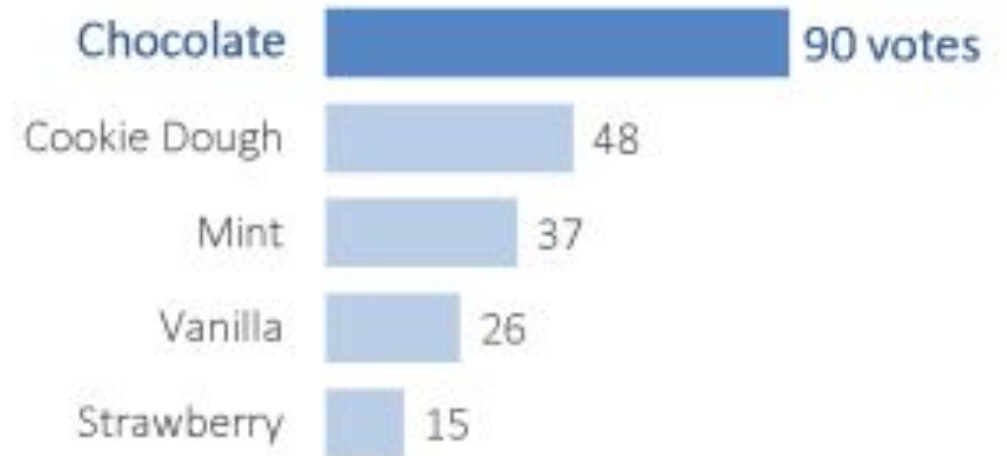
Attend: Draw the Viewer to Important Information

Ice cream flavor preferences based on 2014 survey of elementary school students (n=216)



or

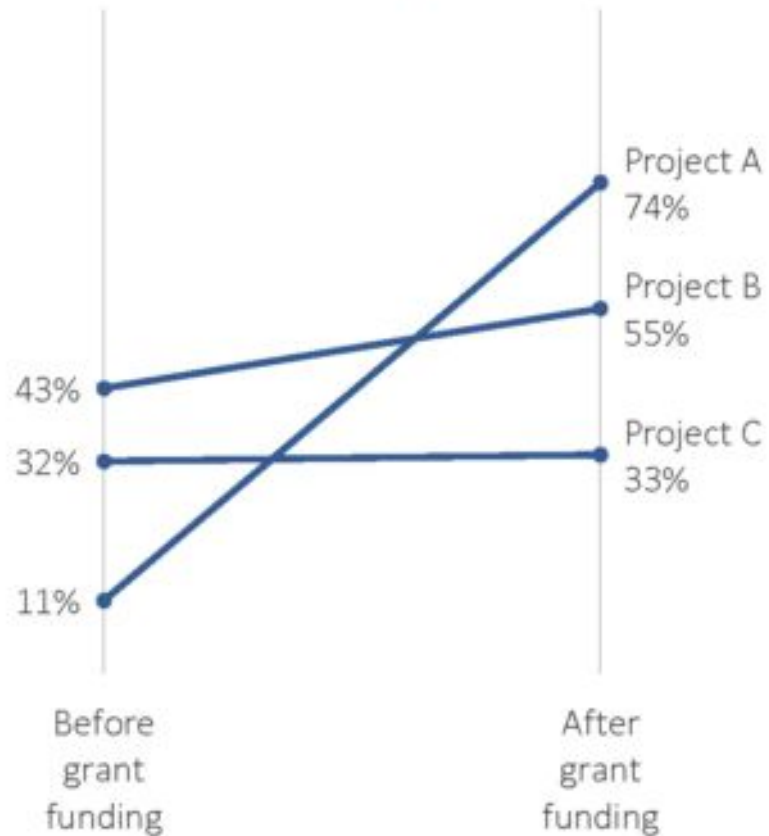
Chocolate was most popular flavor
among elementary students surveyed



Source: 2014 survey of elementary school students (n=216)

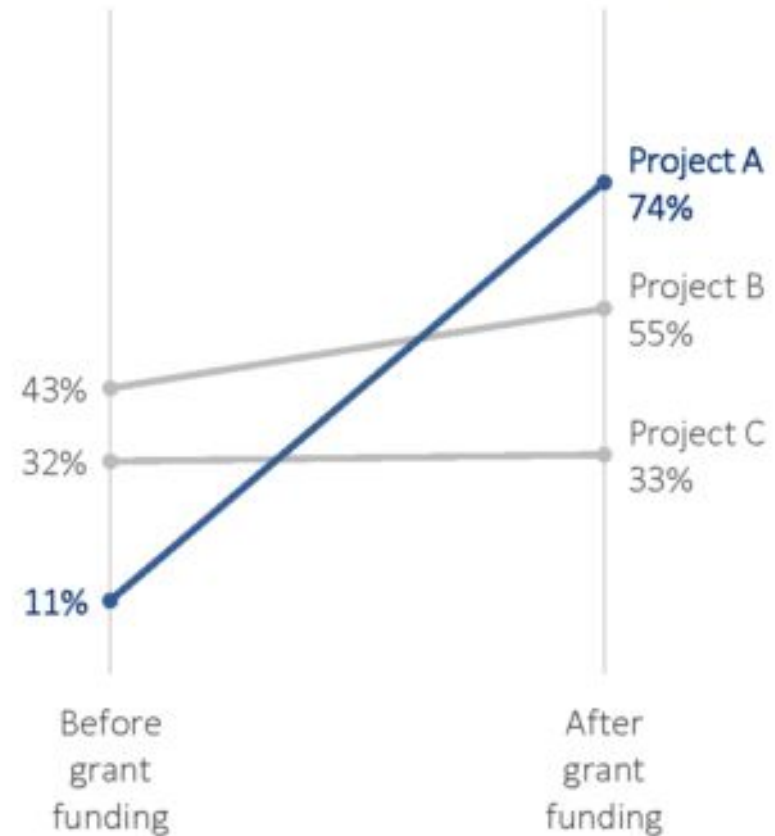
Attend: Draw the Viewer to Important Information

Project results before and after implementation of grant



or

Project A had the greatest gains after the four-year grant funding



Labels

Make sure that everything is labeled.

Make sure that labels are readable.

Make sure that you aren't over-labeling.

Data Visualization - Colors

Think carefully about the colors you are using.

How to pick more beautiful colors for your visualization:

<https://blog.datawrapper.de/beautifulcolors/>

ColorHexa: <https://www.colorhexa.com/>

Choosing Colors:

https://www.perceptualedge.com/articles/b-eye/choosing_colors.pdf

Data Visualization: Some Resources

Junk Charts: https://junkcharts.typepad.com/junk_charts/

Fundamentals of Data Visualization (Book):
<https://clauswilke.com/dataviz/>

Visualizing Data: <https://www.visualisingdata.com/>

Two videos for you to watch on your own:

Harvard's CS109

24:55 - 30:49

<https://matterhorn.dce.harvard.edu/engage/player/watch.html?id=7f968df9-404a-46a2-ae5f-e35479875f95>

Hans Rosling's TED Talk (along with some others)

<https://bigdata-madesimple.com/best-ted-videos-on-data-visualization/>



Next week:

1. **Presentations: 10 minutes each team**
2. **Panel Discussion: pivoting from PhD to Data Scientist**

