Data Visualization

Visual tools for understanding your data

ICME Fundamentals of Data Science – Summer Workshop Series

ICME

Yesterday's Recap and Today's Plan

Yesterday - Design for Yourself

- Intros and Course Plan
- Python for data visualization
- Exploratory Data Analysis (EDA)
- Plotting basics in Python

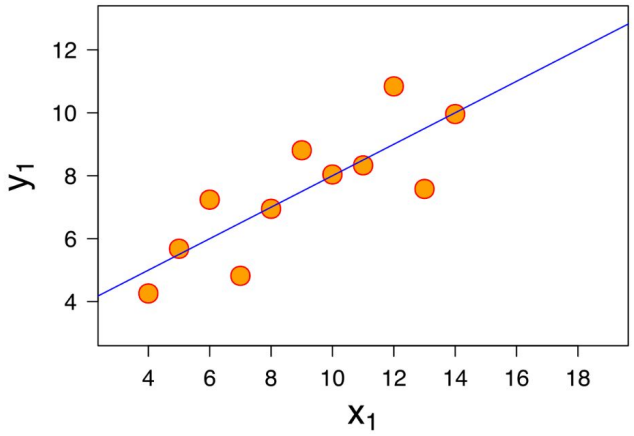
Today - Design for Others

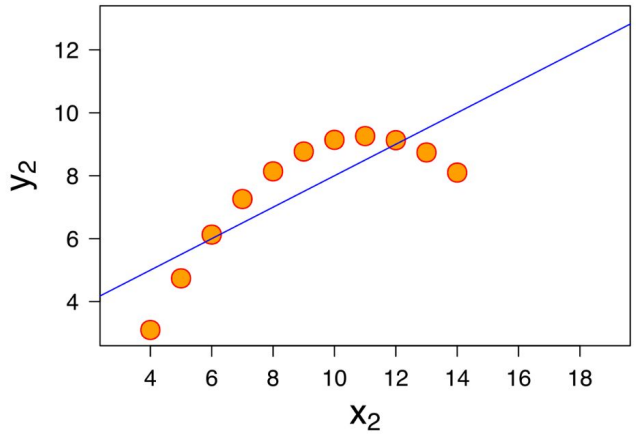
- Perception Considerations
- Polishing Plots in Python
- Data Storytelling
- Visualization Evaluation

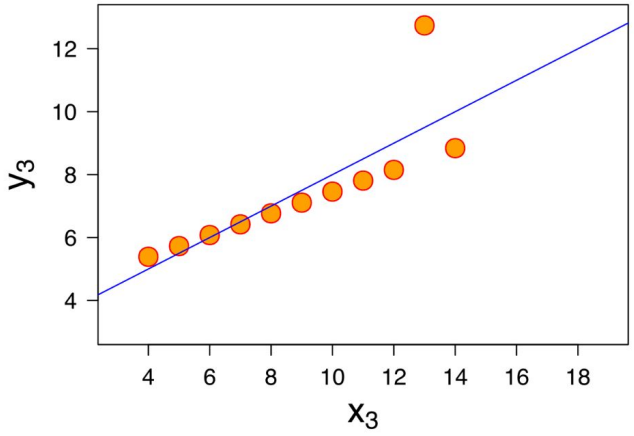
Course Website and Resources

bit.ly/icme-vis

Anscombe's Quartet







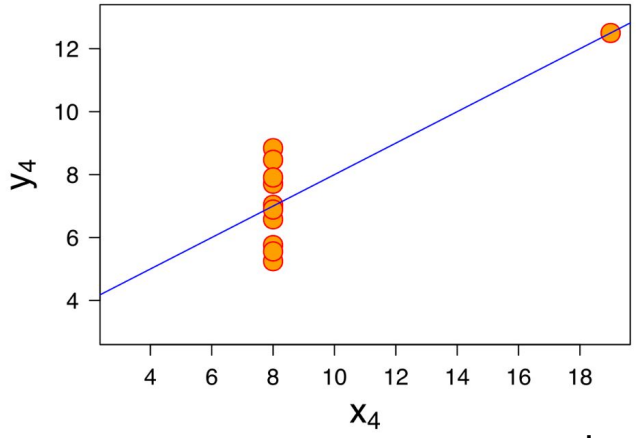
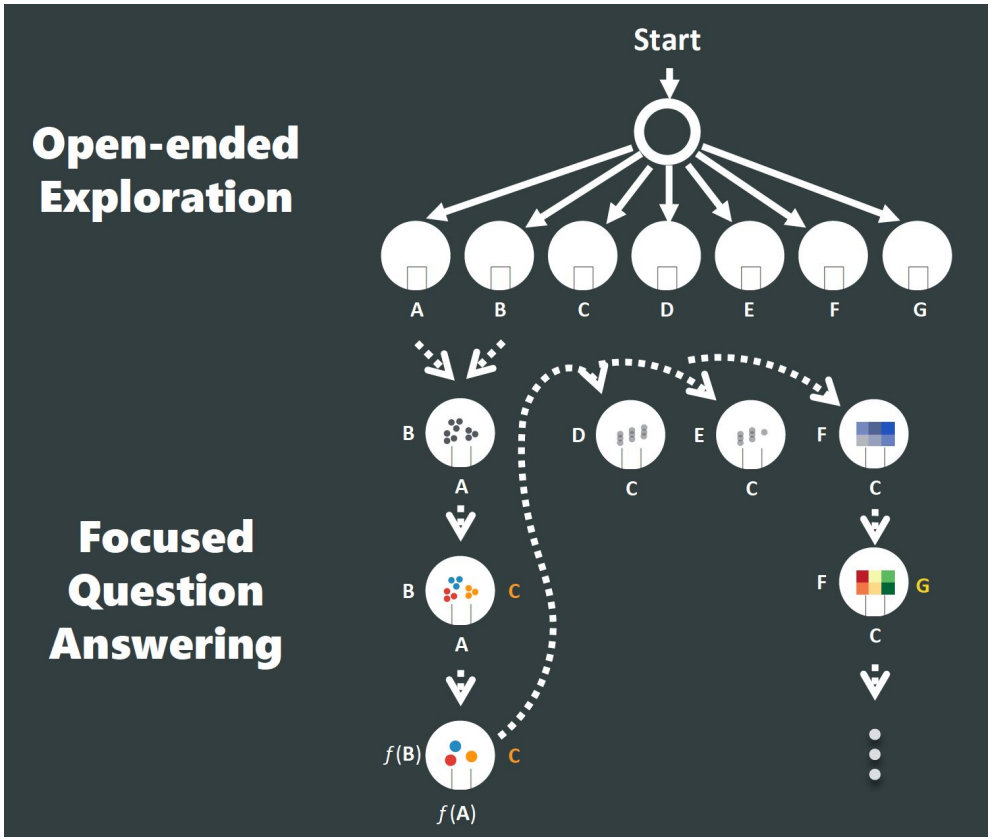


Image source: [Wikipedia]

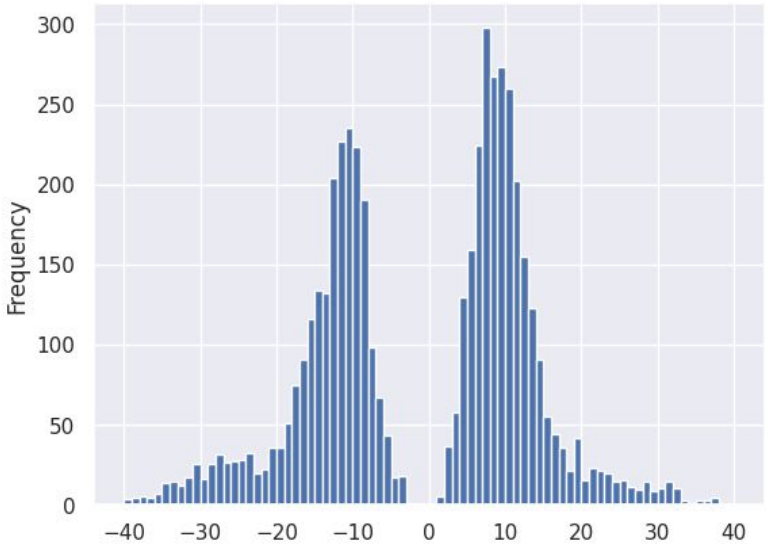
The exploratory data analysis process

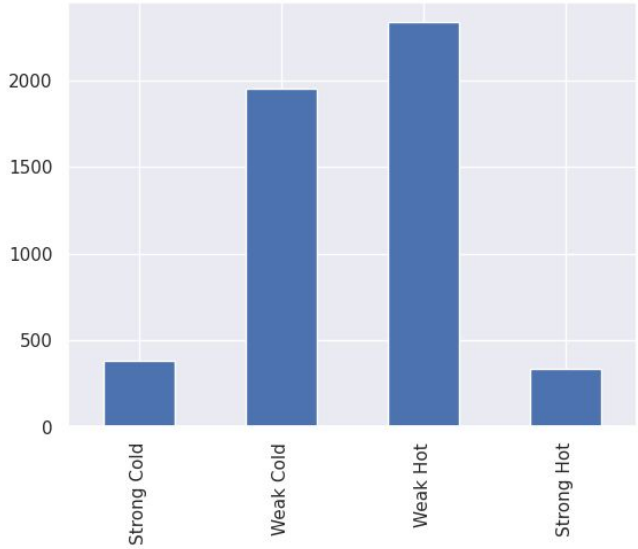


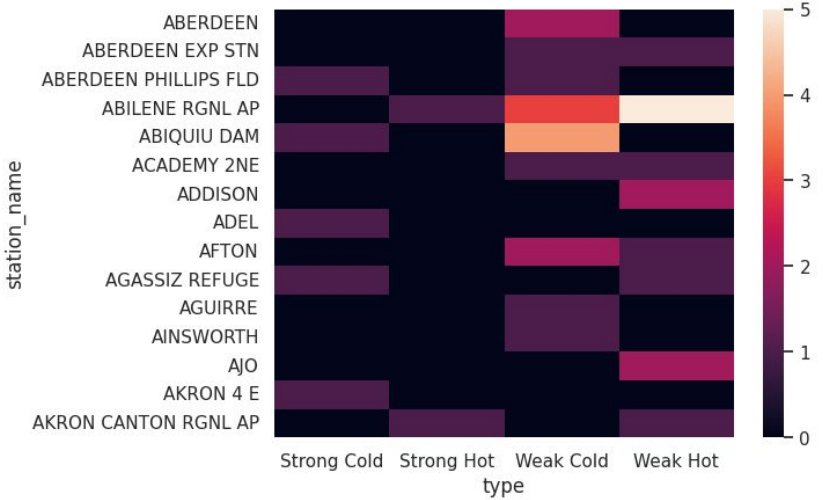
YData

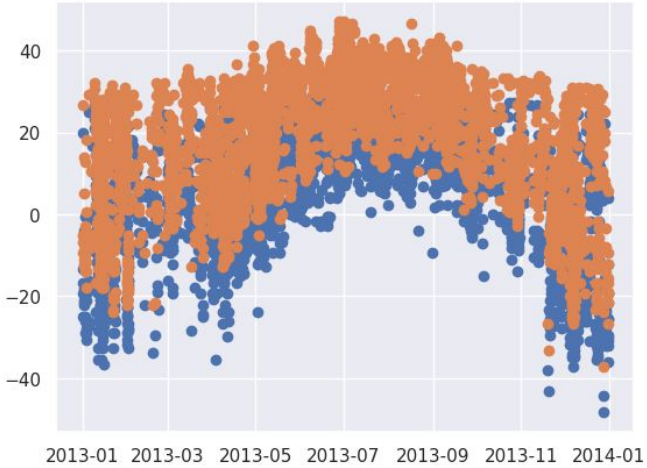
Figure from Hari Subramonyam, adapted from Hullman

Basic Plotting









Yesterday's Homework

1. **Explore the Airline Delay dataset** using the tools you learned today.
2. **Create a scatter plot of departure delay vs arrival delay.** How correlated are the two? What does this suggest about why flights are delayed?
3. **Develop 1-3 candidate ideas** for a visualization to polish tomorrow.

slido

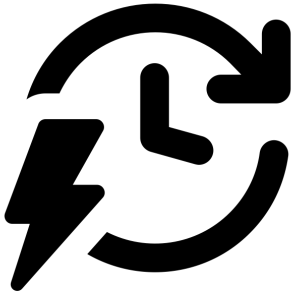
Join at slido.com
#1315187

ⓘ Click **Present with Slido** or install our Chrome extension to display joining instructions for participants while presenting.

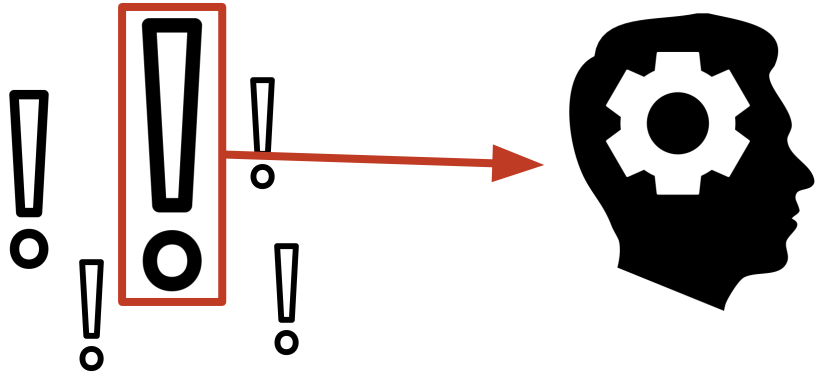
Perception Design Considerations

Pre-attentive Processing

Definition: *Subconscious* accumulation of information used for prioritizing what we pay attention to

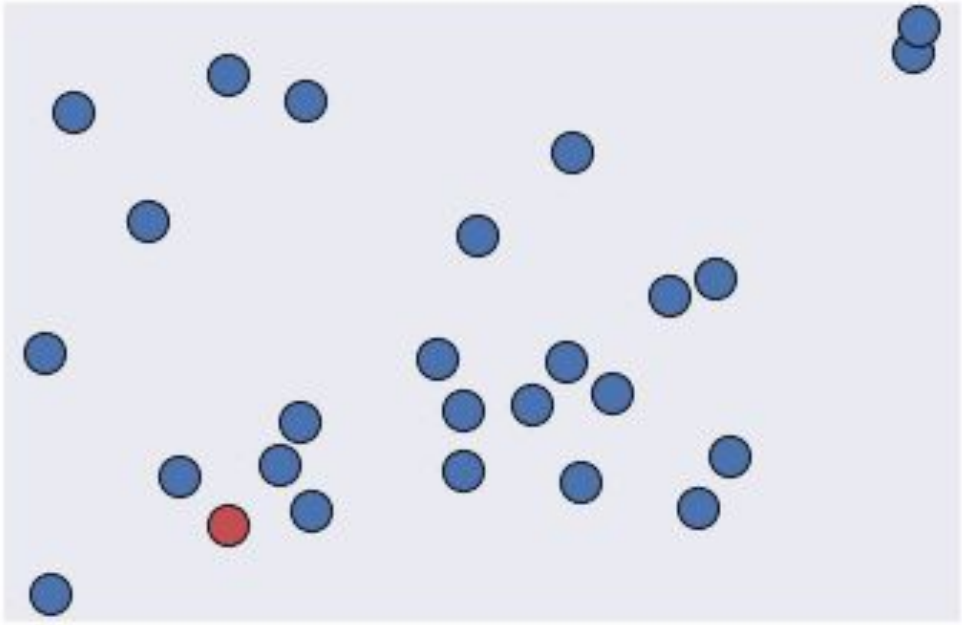


Extremely fast, automatic

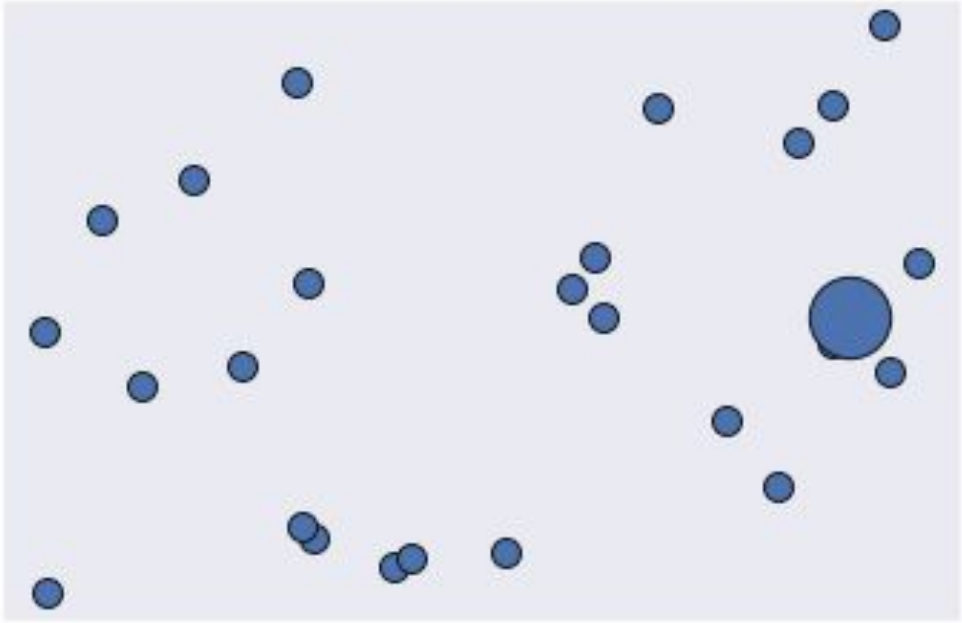


Highest salience items get passed on for conscious (attentive) processing

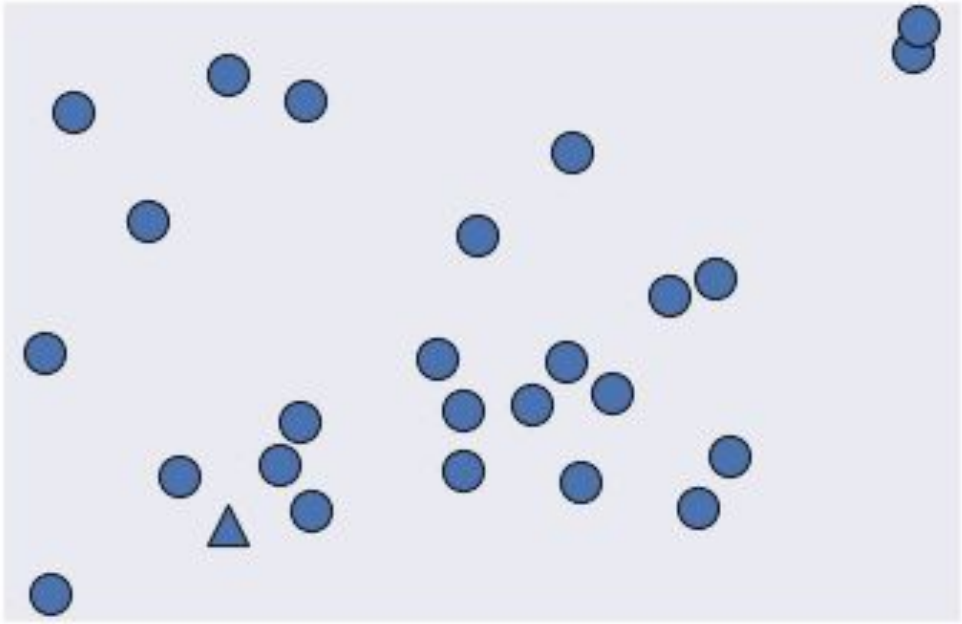
Find the red dot



Find the big dot

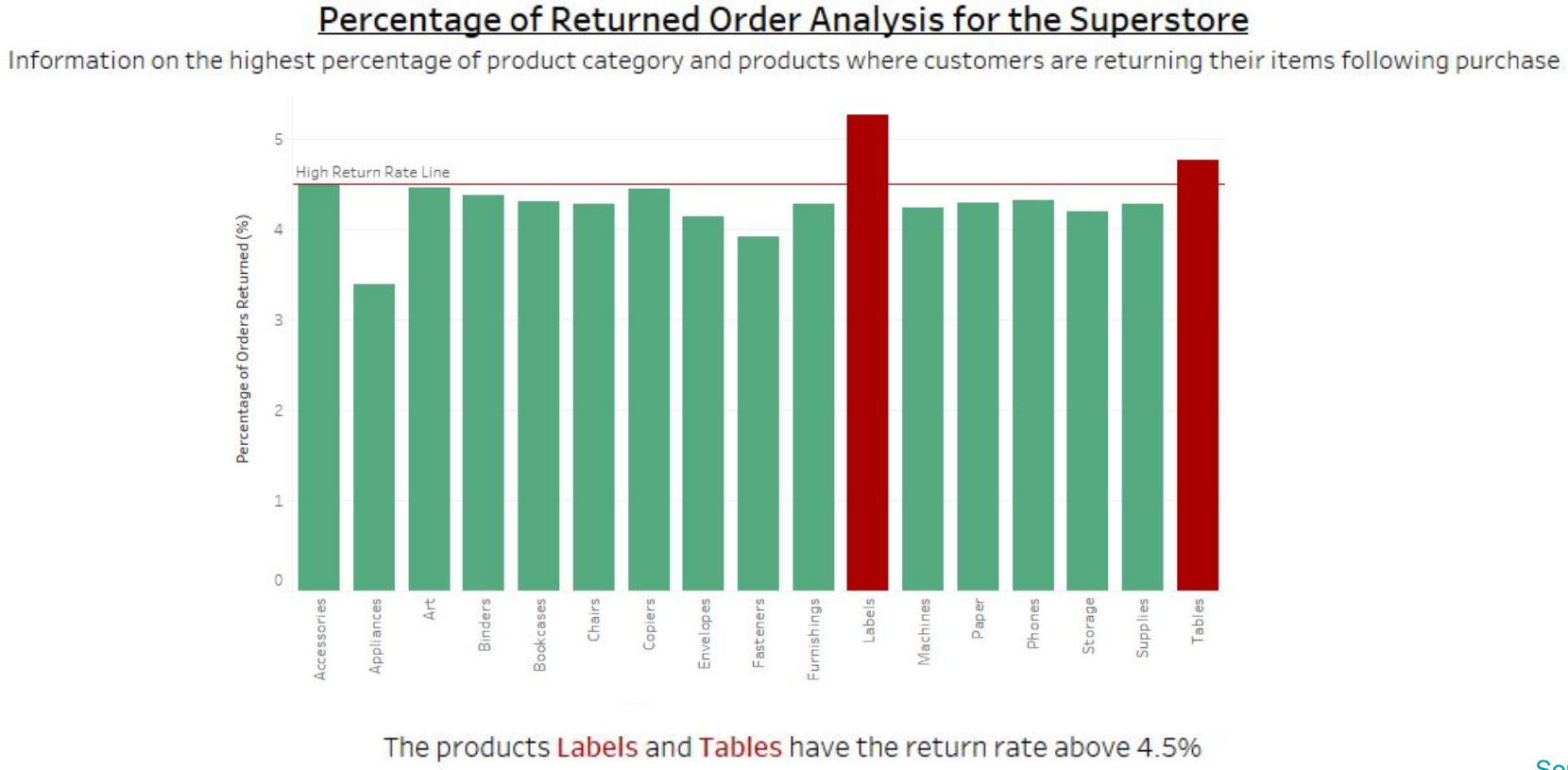


Find the triangle



Takeaway: We can leverage what the brain automatically focuses on to get our viewer to pay attention to what we want.

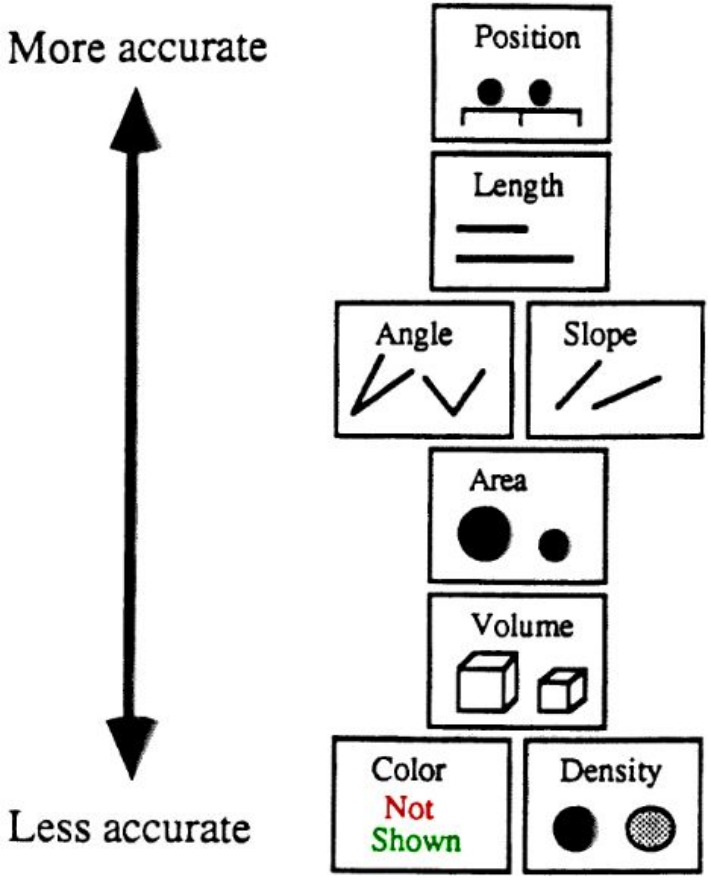
Example: Leveraging Pre-Attentive Processing



Source

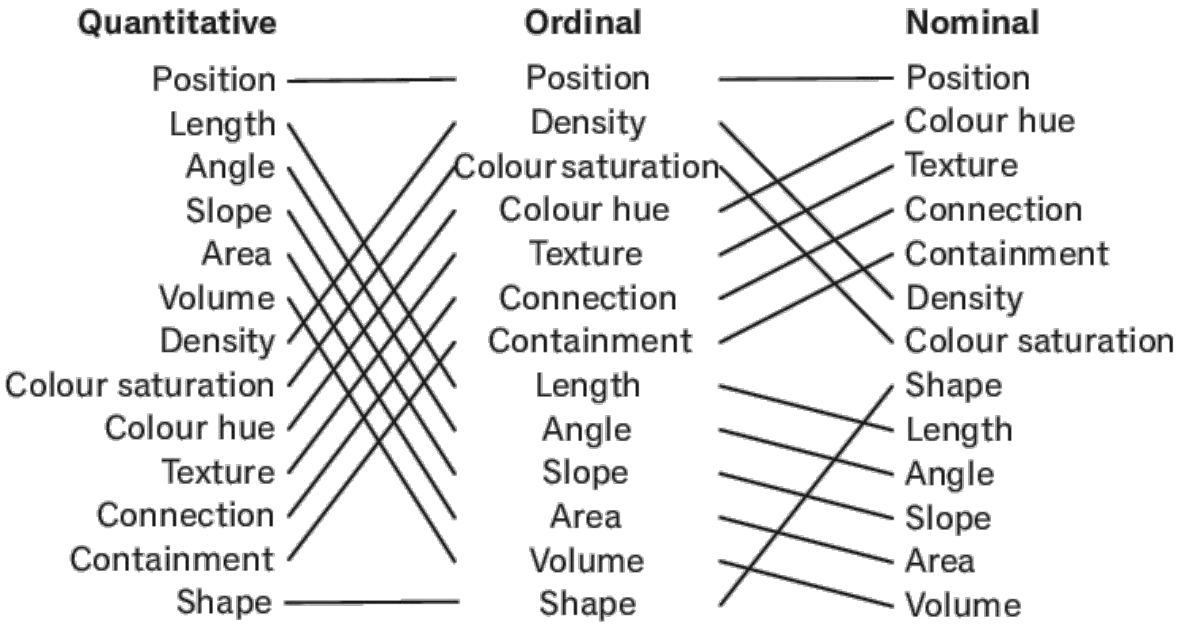
We can represent data using different kinds of “marks.”

Some encodings are easier for the human eye to differentiate than others.



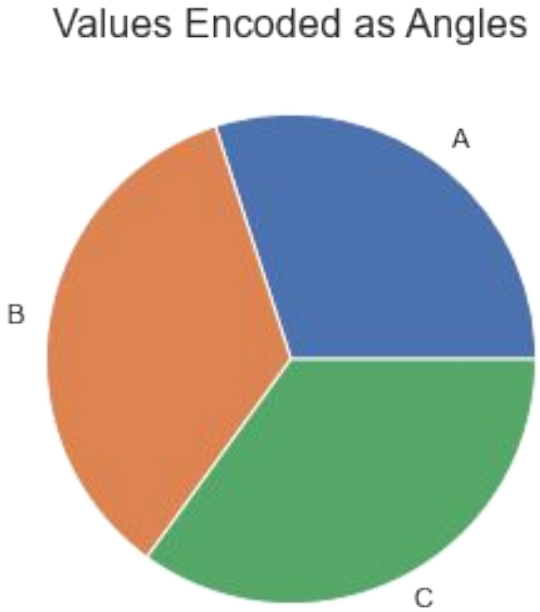
Jock Mackinlay. 1986. Automating the design of graphical presentations of relational information. ACM Trans. Graph. 5, 2 (April 1986), 110–141. https://doi.org/10.1145/22949.22950

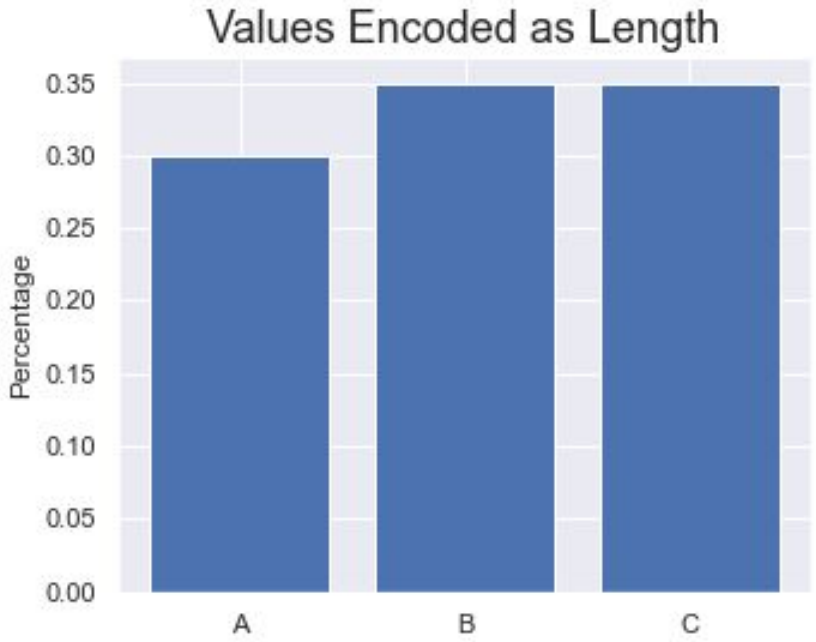
Different marks are more effective for different kinds of data.



Jock Mackinlay. 1986. Automating the design of graphical presentations of relational information. ACM Trans. Graph. 5, 2 (April 1986), 110–141. https://doi.org/10.1145/22949.22950

Example: Angle vs. Length Encoding

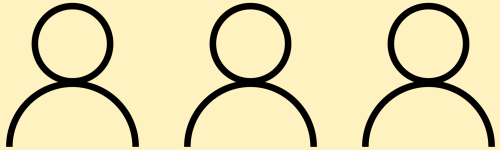




Expressiveness and Effectiveness

Expressiveness

The ability of a visualization to represent all the data accurately and without ambiguity.



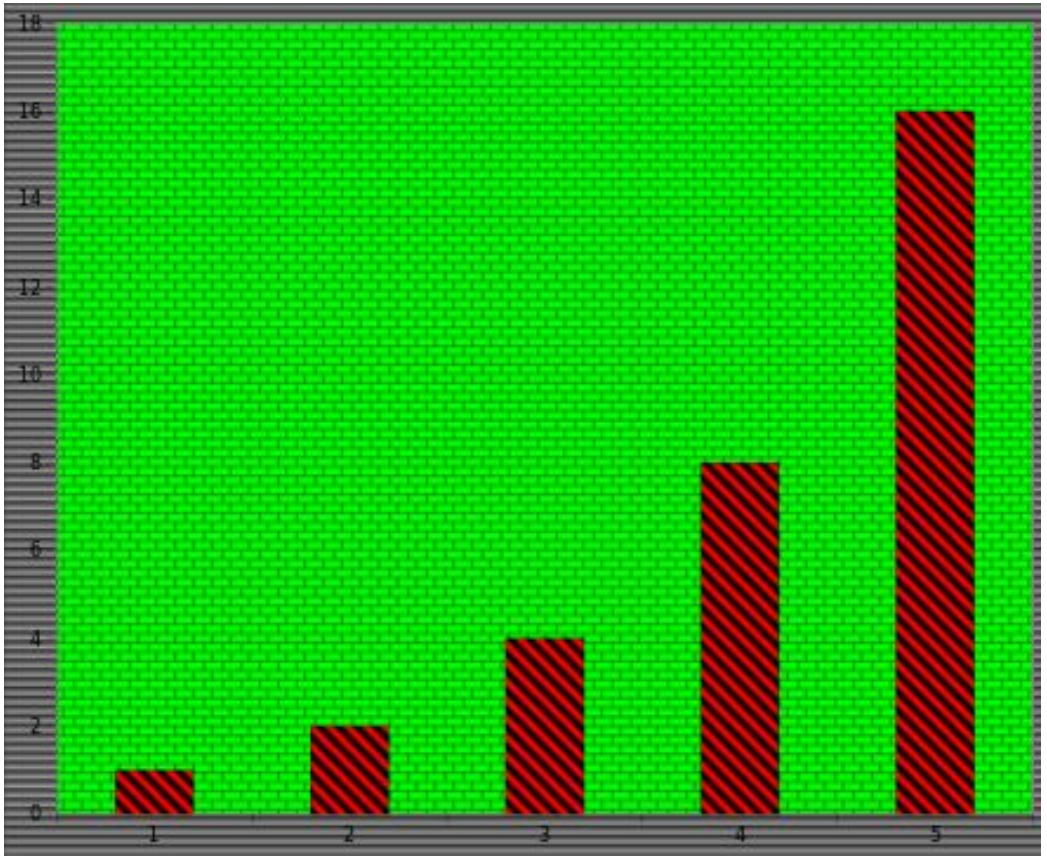
Effectiveness

How quickly, easily, and accurately the viewer can understand the information being presented.

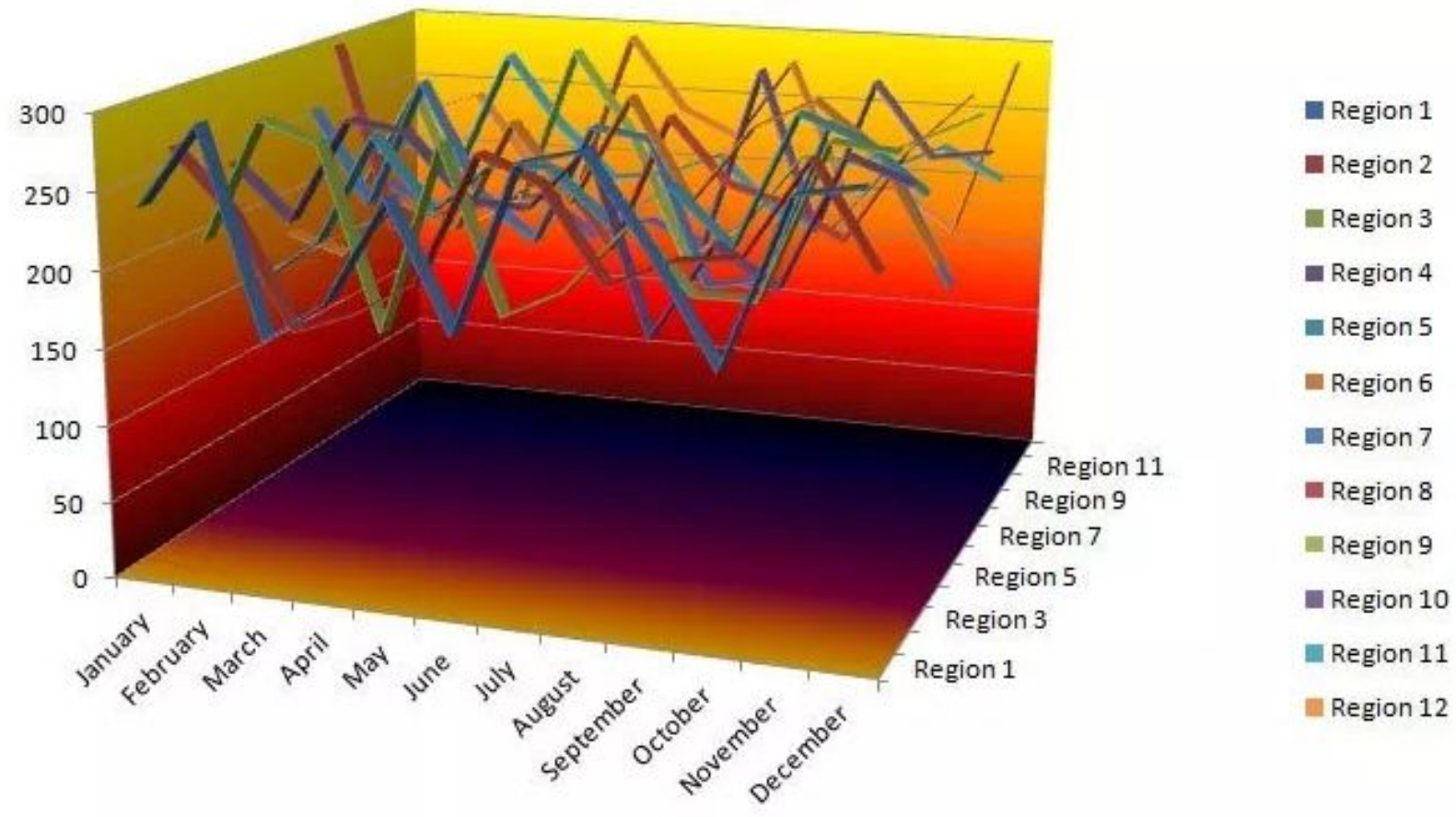
Not necessarily at odds, but sometimes tension between completeness and clarity!

Chart Junk

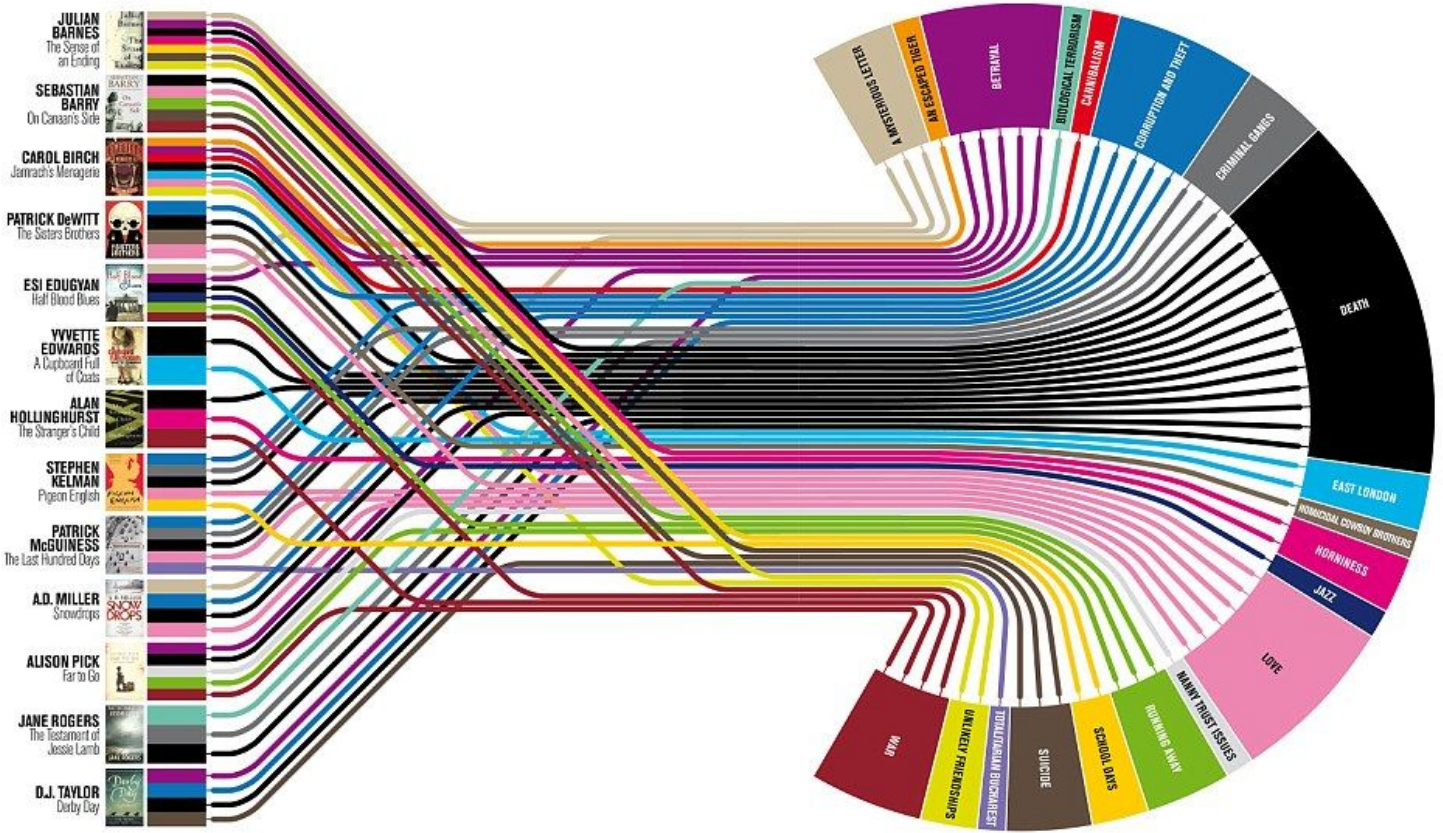
Definition:
Unnecessary or confusing elements in charts and graphs that do not contribute to the viewer's understanding of the data. Coined by Edward Tufte.



Source


300
250
200
150
100
50
0
January
February
March
April
May
June
July
August
September
October
November
December
Region 1
Region 3
Region 5
Region 7
Region 9
Region 11
Region 1
Region 2
Region 3
Region 4
Region 5
Region 6
Region 7
Region 8
Region 9
Region 10
Region 11
Region 12

Source



Plot lines

What makes a prize-winning novel? As Julian Barnes wins the Booker Prize, Delayed Gratification's Johanna Kamradt charts the themes of this year's longlisters.

Source


MONSTROUS COSTS
Total House and Senate campaign expenditures, in millions
$300
250
200
150
100
50
1972
'74
'76
'78
'80
'82 est.
VOTE

Source

How to avoid creating Chart Junk

Simplicity: Every element in your visualization should serve a purpose.

- Decorative elements
- 3D plots
- Fancy fonts
- Non-strategic extra colors

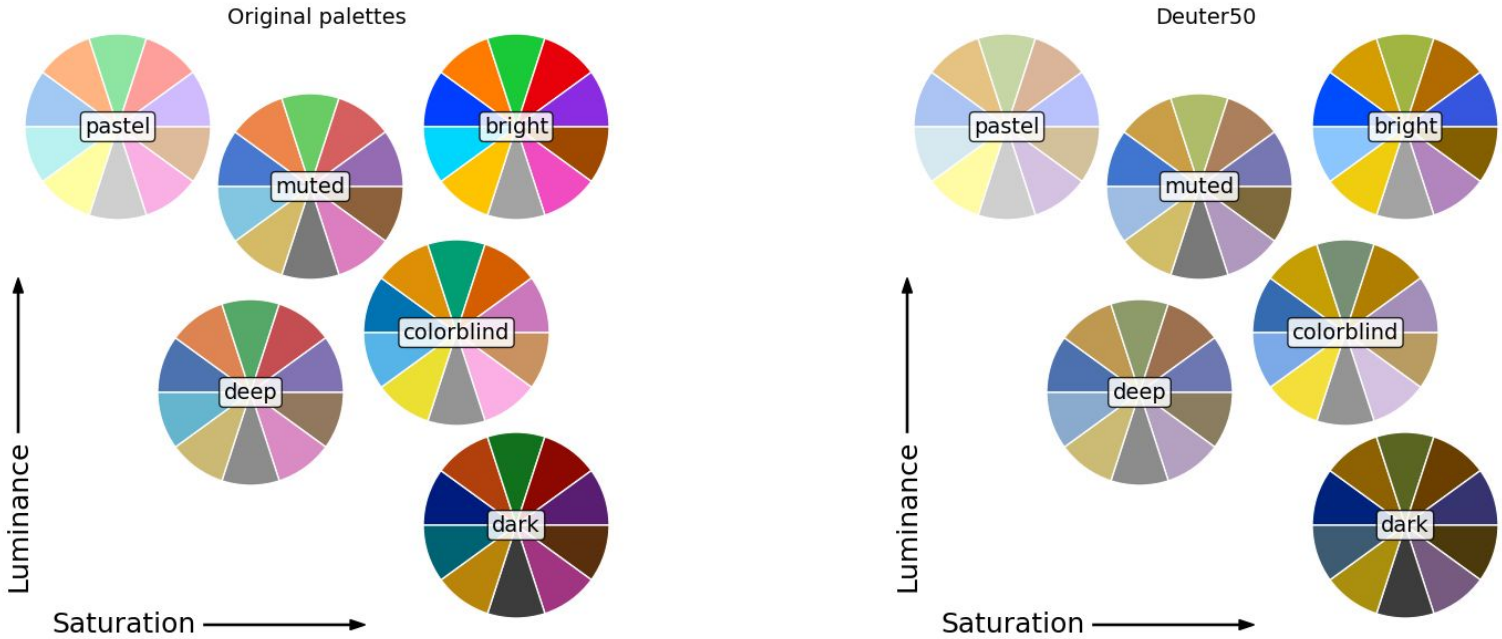
- One font per visualization
- Familiar chart types
- Colors that go together

slido

Audience Q&A Session

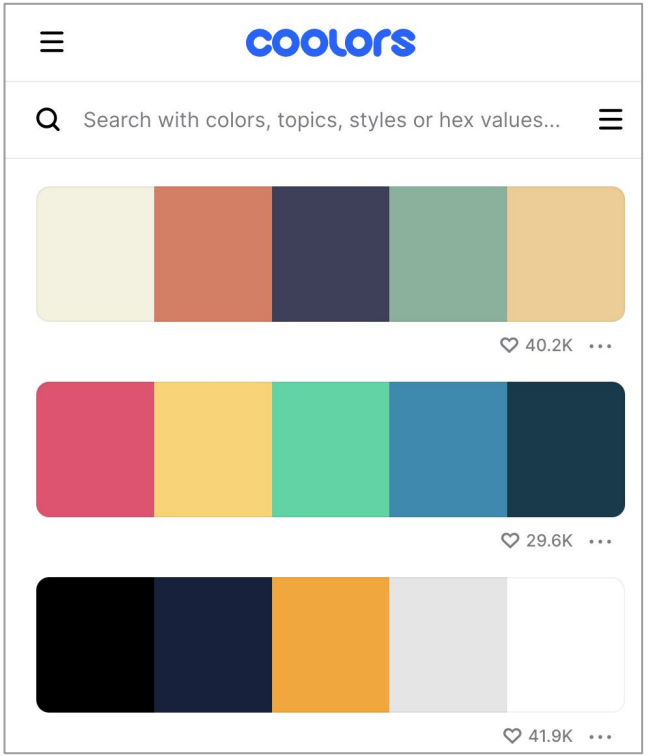
ⓘ Click **Present with Slido** or install our Chrome extension to show live Q&A while presenting.

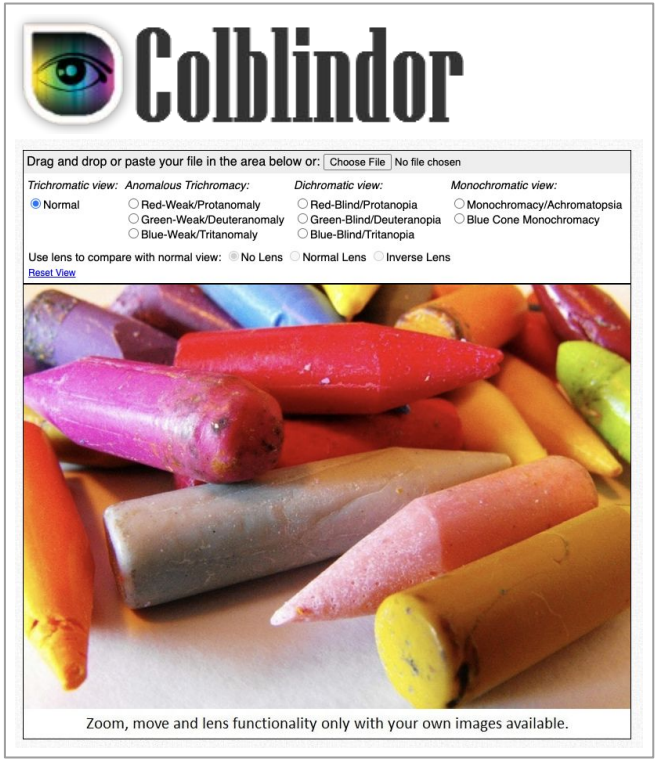
Accessibility: Designing for Color Blindness



Source

Accessibility: Color Blindness Resources





List of color resources will be posted on the workshop website.

Implementation

Beyond the Bare Bones Visualizations in Python

- Python implementation of:
  - Titles
  - Axis Labels
  - Annotations
  - Opacity
  - Legends
  - Shared Axes
  - Small Multiples

Don’t worry about the exact details here!
The important part is that you have an idea of what’s possible.

The completed version of this notebook is available on the website so you can copy from the examples.

< Code Demo >

Coffee Break

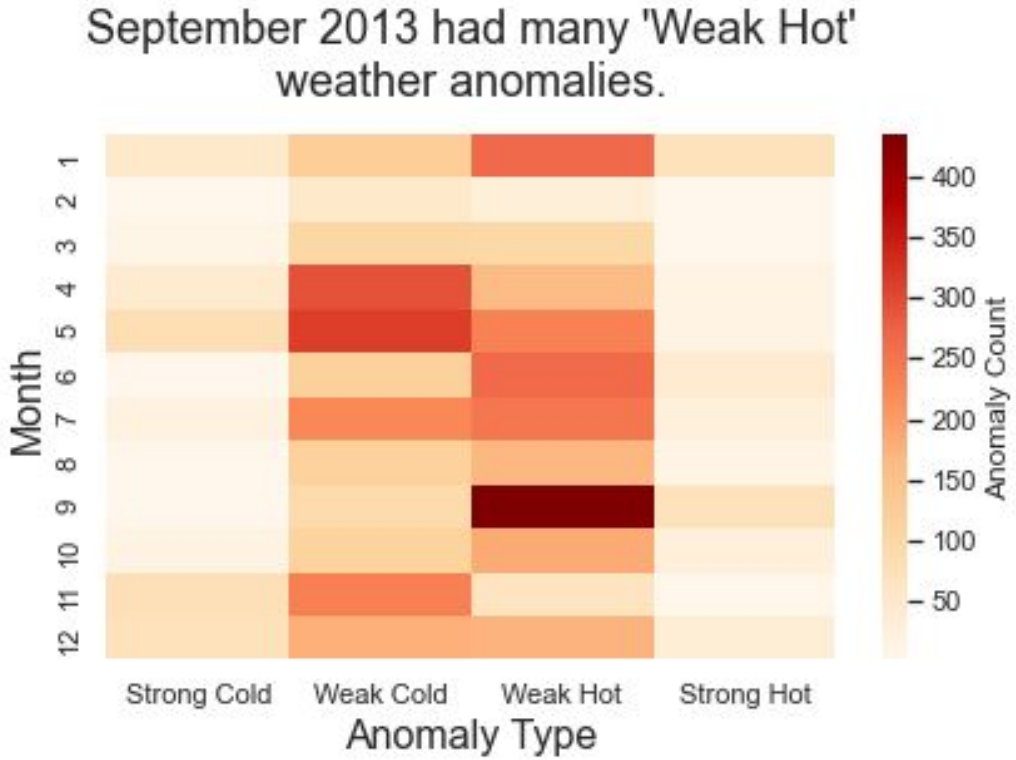
15:00

Your Task

Refine a visualization based on one of the findings in the exploratory data analysis process. Think about a descriptive title, labels, and legends!

Don’t worry if you don’t finish. You’ll have the chance to share your ideas with a peer later and get feedback.

Example with Weather Anomaly Data



15:00

colab

Your Task

Refine a visualization based on one of the findings in the exploratory data analysis process. Think about a descriptive title, labels, and legends!

Don’t worry if you don’t finish. You’ll have the chance to share your ideas with a peer later and get feedback.

slido

Audience Q&A Session

ⓘ Click **Present with Slido** or install our Chrome extension to show live Q&A while presenting.

Data Storytelling

You probably didn’t take this course just to make beautiful visualizations.

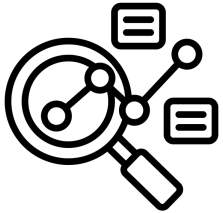
You probably want to be compelling and persuasive data communicator.

Designing a story with your data.

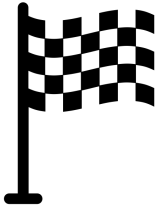
Data storytelling can include more than just the figure (e.g., written text, slides)

Typically want a narrative structure:

Beginning
Setting up the Problem



Middle
Sharing the Analysis



End
Takeaways or Call to Action

Designing for your Audience

Who is your audience?

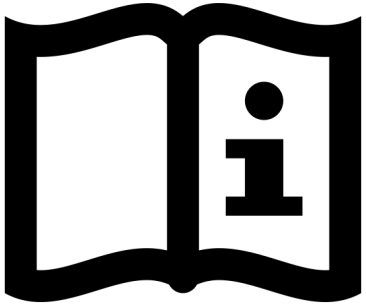
Motivation: What do ***they*** need to get out of this presentation?

Technical Depth: What level of technical complexity is appropriate and feasible? Will any jargon be understood?

Context: How familiar is the audience with the problem you are addressing? How is their context different from yours?

#1 problem for technical communication – insufficient context

Resources and Examples of Data Storytelling



Design guides for data storytelling

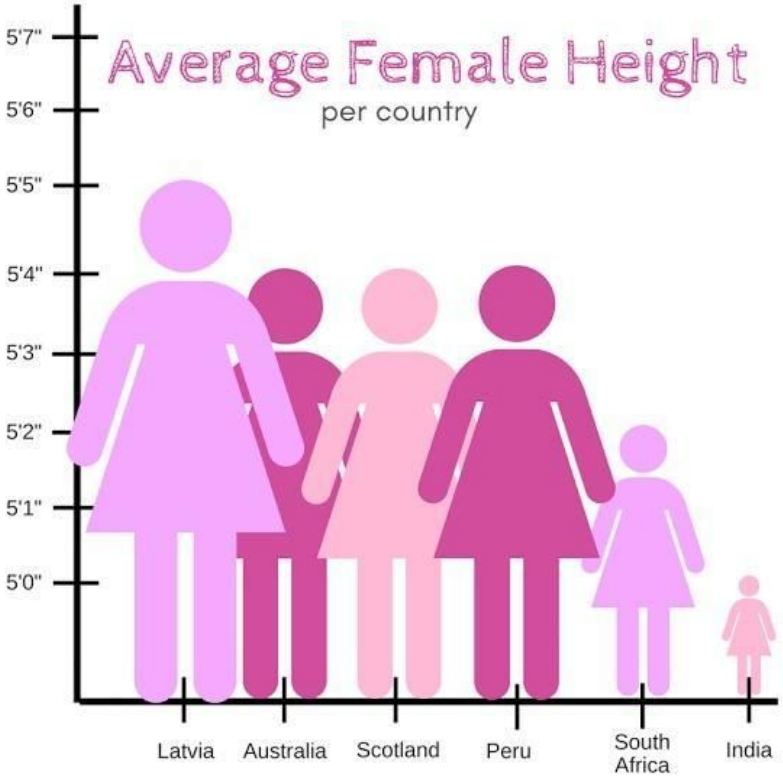
Examples of data storytelling in action

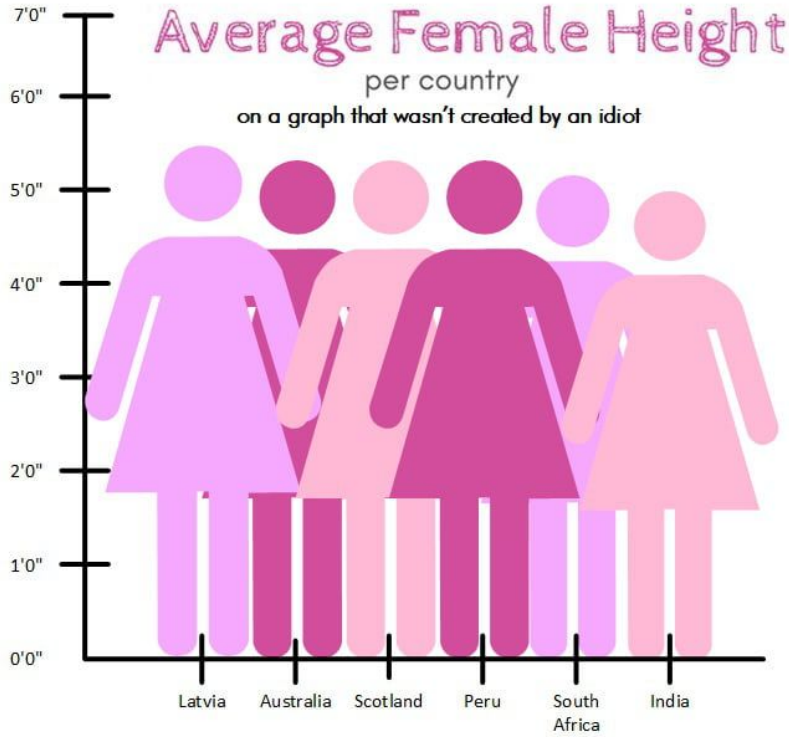
Links to resources are on workshop website

Data Visualization Critique

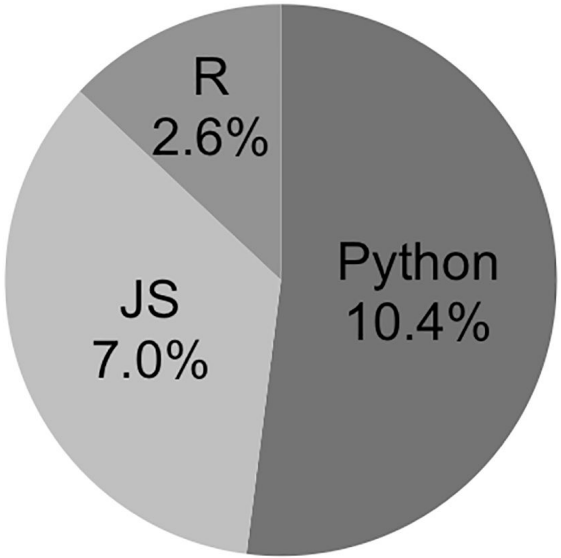
Data Visualization Critique

- Somewhat subjective – people react to visualizations differently
- Critique questions:
  - Is it expressive? Does it convey the data accurately?
  - Is it effective? Do you easily come away with the intended takeaway?
  - Are there any points of ambiguity?
  - Does it look nice?
  - Is there anything you could do to help your reader?

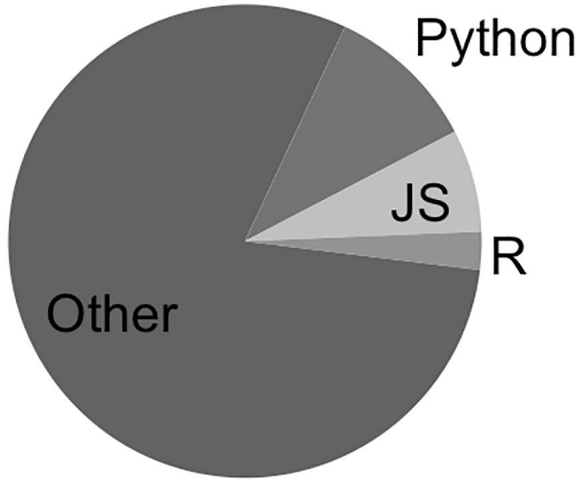

Average Female Height
per country
5'7"
5'6"
5'5"
5'4"
5'3"
5'2"
5'1"
5'0"
Latvia
Australia
Scotland
Peru
South Africa
India


Average Female Height
per country
on a graph that wasn't created by an idiot
7'0"
6'0"
5'0"
4'0"
3'0"
2'0"
1'0"
0'0"
Latvia
Australia
Scotland
Peru
South Africa
India

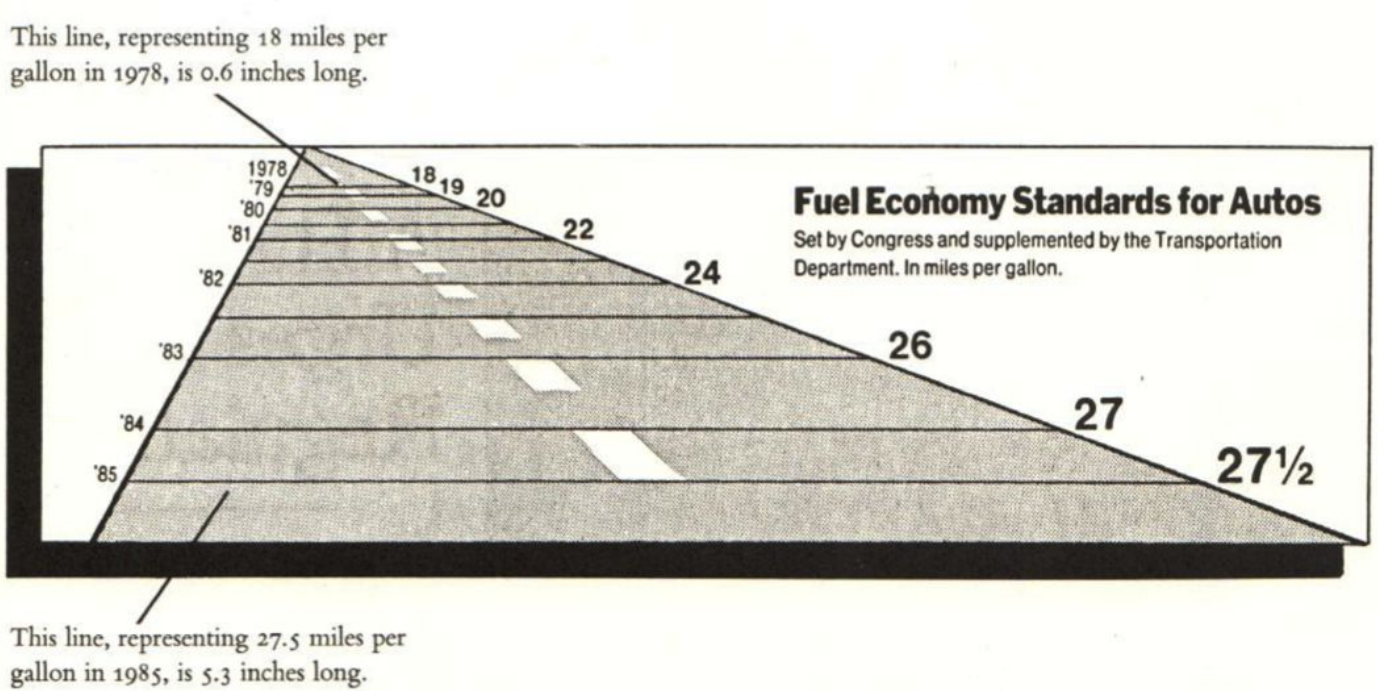
Google tutorial searches



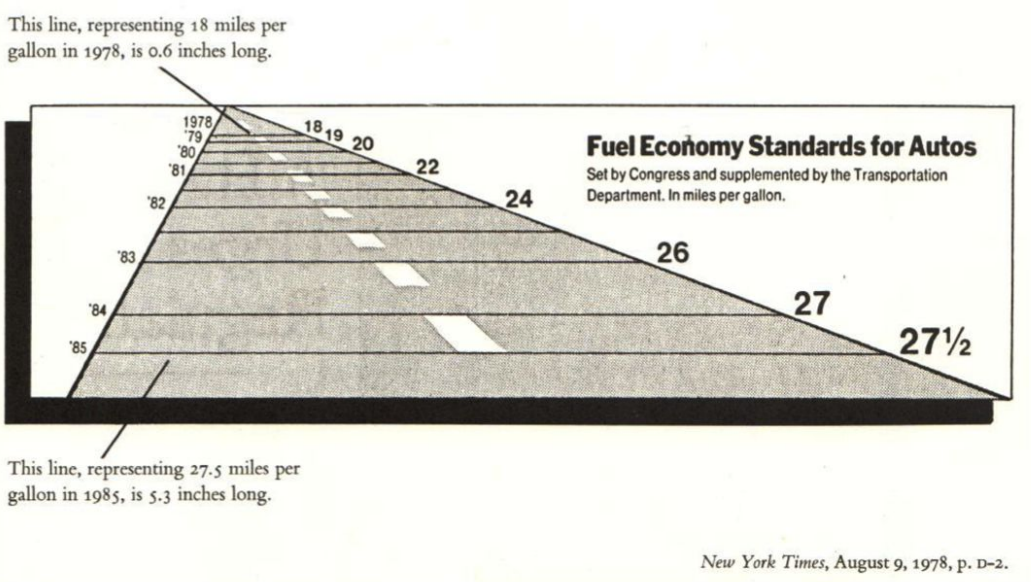
Google tutorial searches



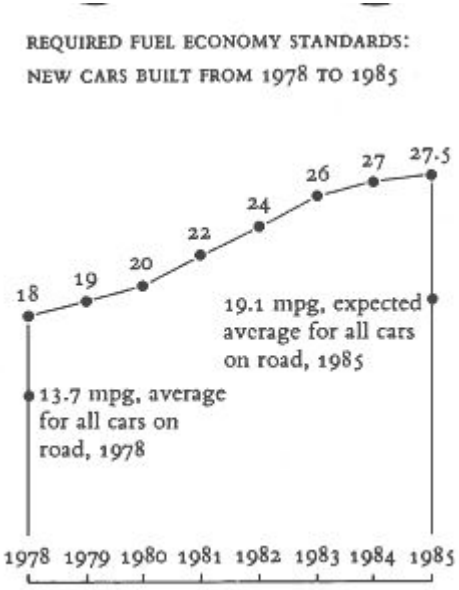
https://www.oreilly.com/content/in-defense-of-the-pie-chart/



Tufte, “Visual Display of Quantitative Information”

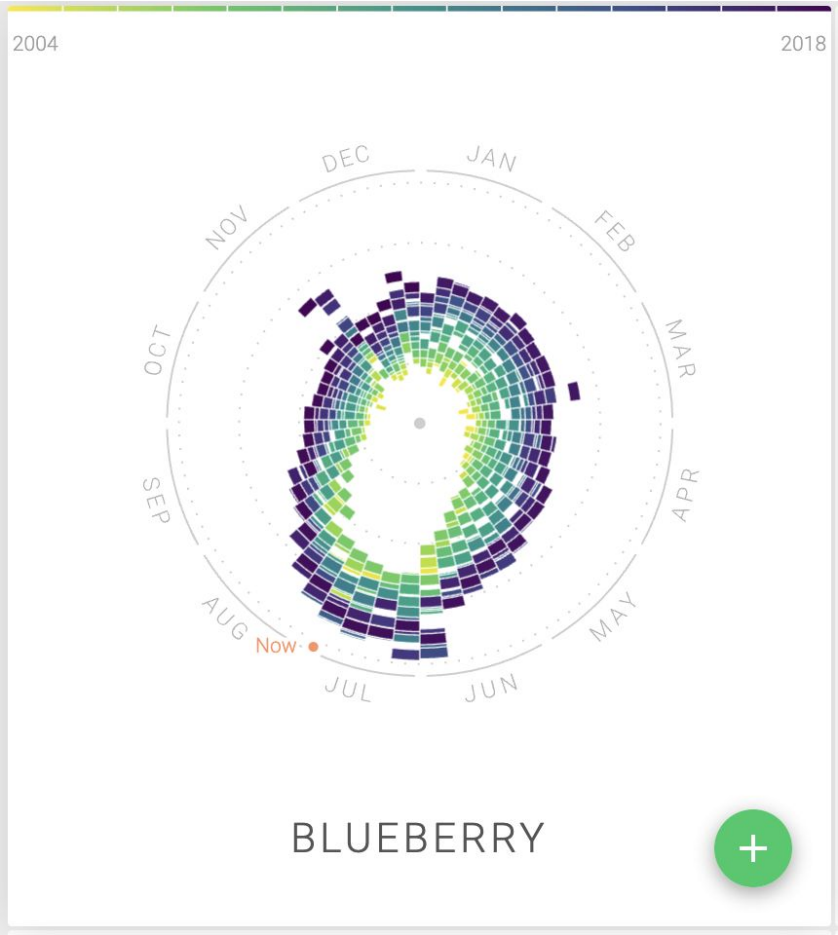


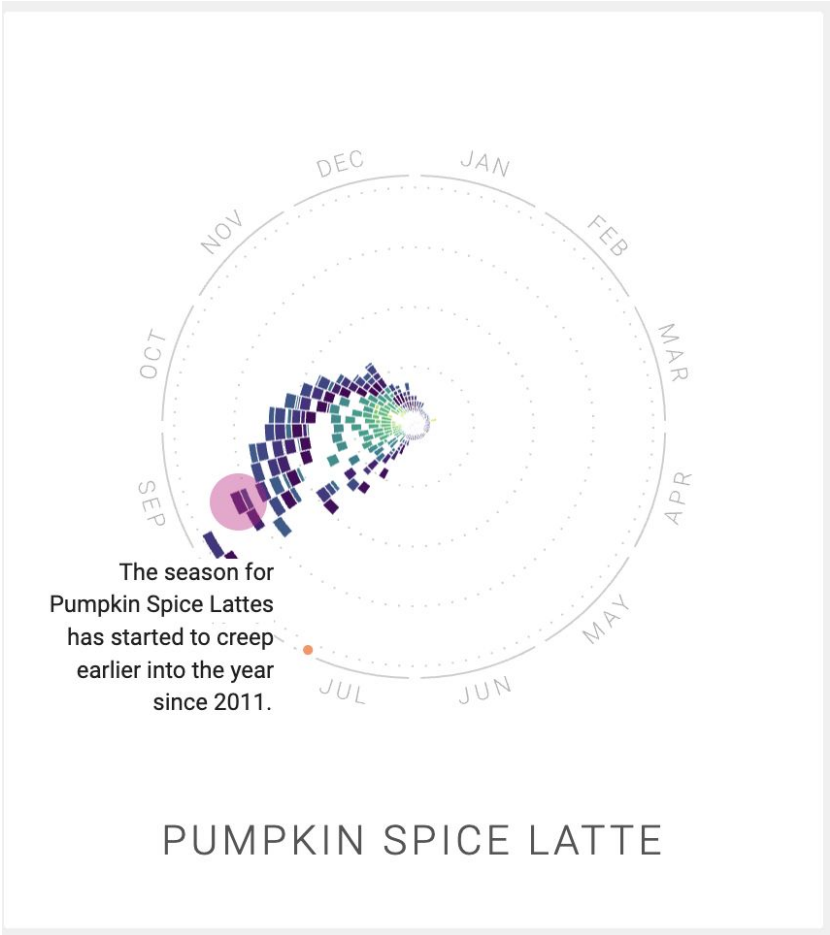
New York Times, August 9, 1978, p. D-2.



Tufte, “Visual Display of Quantitative Information”

https://rhythm-of-food.net/

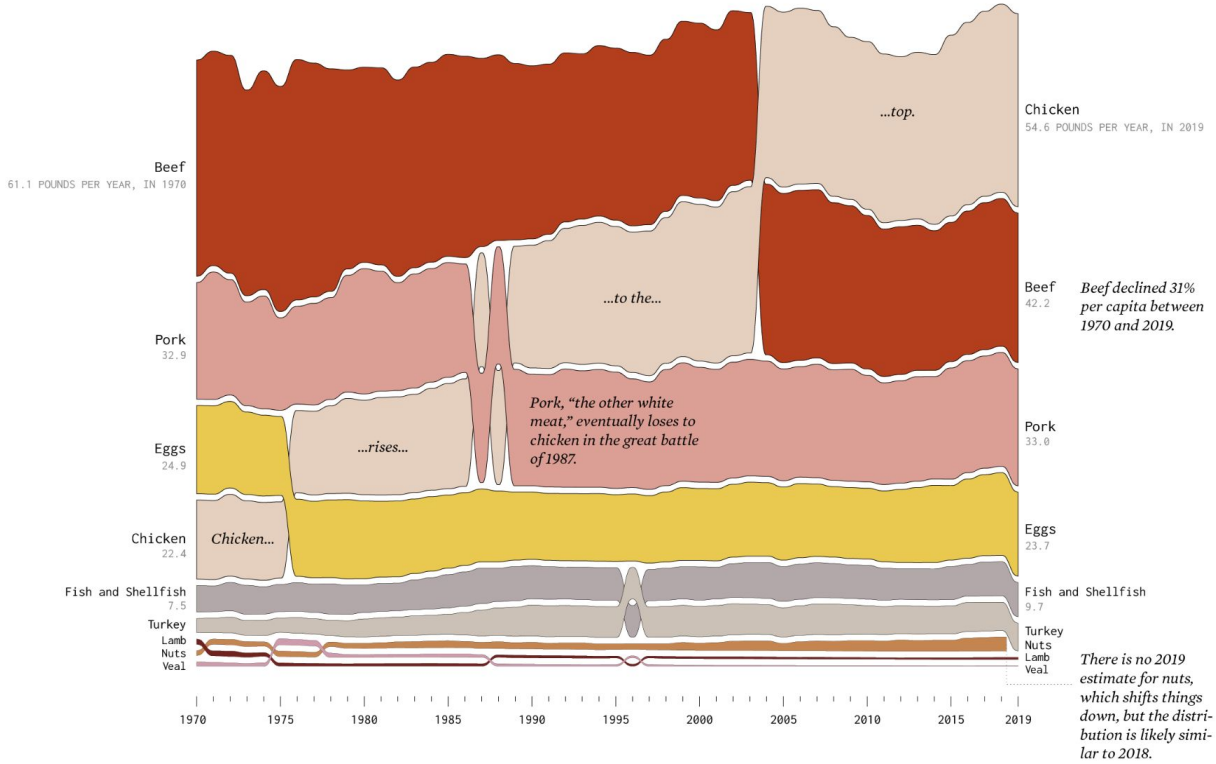




Proteins

pounds per year per capita

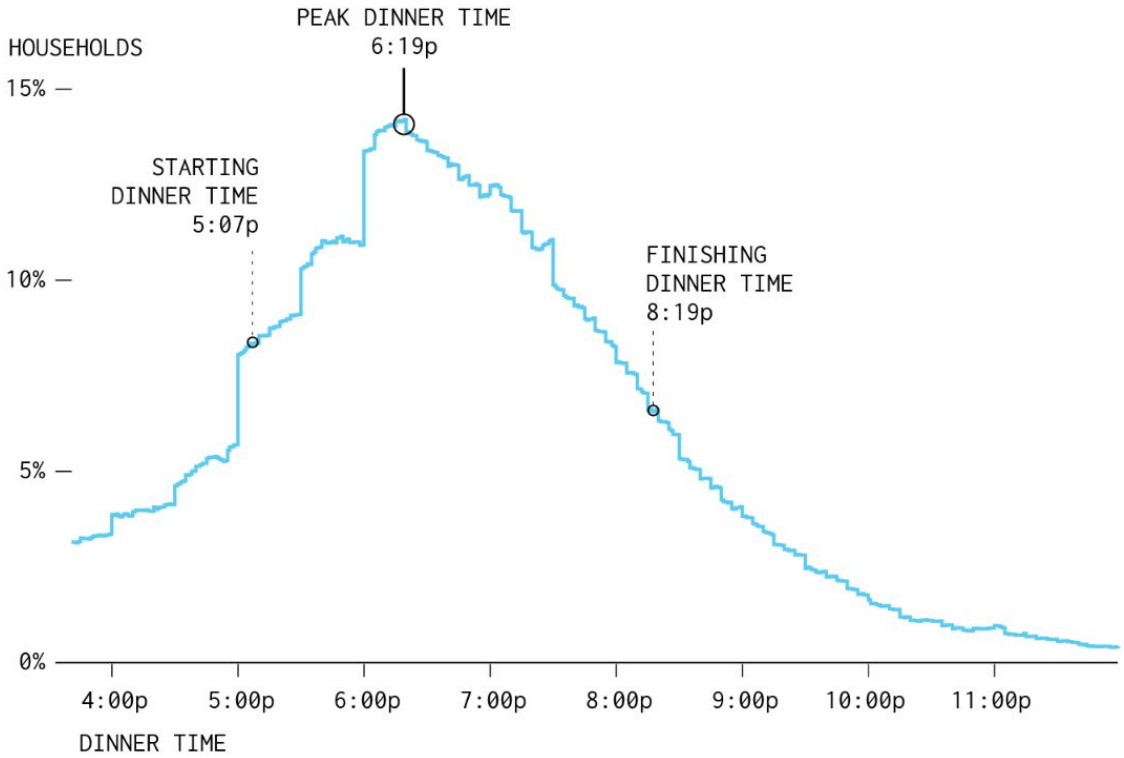
Beef still makes up a large portion of protein consumed, but pounds per capita continues to decrease, whereas chicken continues to increase. Chicken took took away the top spot from beef in 2004 and has reigned supreme ever since.



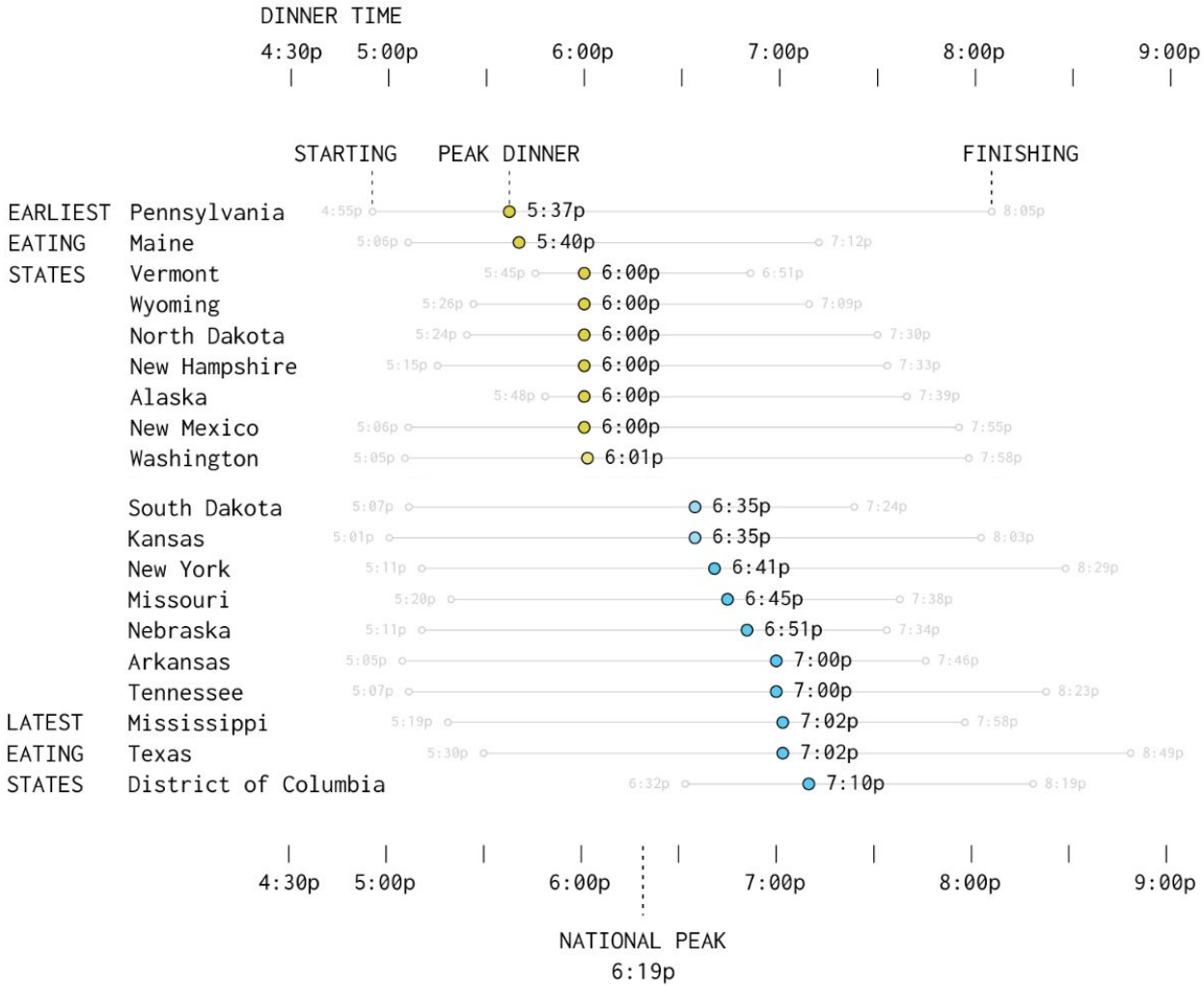
Source

WHEN AMERICANS EAT DINNER

Most households eat dinner between 5:07p and 8:19p, with peak dinner time at 6:19p.



Source


DINNER TIME
4:30p
5:00p
6:00p
7:00p
8:00p
9:00p
STARTING
PEAK DINNER
FINISHING
EARLIEST EATING STATES
Pennsylvania 4:55p 5:37p 8:05p
Maine 5:06p 5:40p 7:12p
Vermont 5:45p 6:00p 6:51p
Wyoming 5:26p 6:00p 7:09p
North Dakota 5:24p 6:00p 7:30p
New Hampshire 5:15p 6:00p 7:33p
Alaska 5:48p 6:00p 7:39p
New Mexico 5:06p 6:00p 7:55p
Washington 5:05p 6:01p 7:58p
South Dakota 5:07p 6:35p 7:24p
Kansas 5:01p 6:35p 8:03p
New York 5:11p 6:41p 8:29p
Missouri 5:20p 6:45p 7:38p
Nebraska 5:11p 6:51p 7:34p
Arkansas 5:05p 7:00p 7:46p
Tennessee 5:07p 7:00p 8:23p
LATEST EATING STATES
Mississippi 5:19p 7:02p 7:58p
Texas 5:30p 7:02p 8:49p
District of Columbia 6:32p 7:10p 8:19p
4:30p
5:00p
6:00p
7:00p
8:00p
9:00p
NATIONAL PEAK
6:19p

[Source]

slido

Audience Q&A Session

ⓘ Click **Present with Slido** or install our Chrome extension to show live Q&A while presenting.

15:00

ICME

colab

Breakout Rooms

1. Introduce yourself to your partner
2. The person whose first name comes first alphabetically shows their visualization first, and the non-visualization designer will explain back what they think the point of the visualization is.
3. Discuss what went well and what the designer can improve to communicate more clearly
4. Swap roles and repeat

slido

Audience Q&A Session

ⓘ Click **Present with Slido** or install our Chrome extension to show live Q&A while presenting.

Wrap Up

What We've Covered

Yesterday - Design for Yourself

- Intros and Course Plan
- Python for data visualization
- Exploratory Data Analysis (EDA)
- Plotting basics in Python

Today - Design for Others

- Perception Considerations
- Polishing Plots in Python
- Data Storytelling
- Visualization Evaluation

Some Takeaways

- Data visualization is very important to understand your data.
- Exploratory Data Analysis is an iterative process of asking and answering questions with data.
- Good visualizations leverage how the human brain processes visual information.
- When in doubt, choose simple and clean visualizations.
- Tell a story with your data – and be sure you give enough context for your audience to understand.
- Feedback from others is critical to designing the most effective visualizations.

thanks!

Resources will remain available on the workshop website.

Contact: kmentzer@stanford.edu