Module 6 - Presentation and Visualization

EDSP Challenge Mentoring

Marck Vaisman

Contents

1. The difference between designing for you vs. for designing for an audience
2. Choosing the right visualization
    - All about asking questions
    - Decomposing your chart
    - Understanding encodings
3. Visualization critique discussion
4. Final conceptual and design considerations
    - Making readable graphics
    - Deep dive into Tufte's principles
    - Bringing it all together.

Litarate programming

- https://en.wikipedia.org/wiki/Literate_programming
- http://literateprogramming.com/index.html
    - https://jupyter.org/
    - https://css-skills.uchicago.edu/posts/2021-11-16-literate-programming-with-r-markdown/

Let's begin

Data visualization is both *art* and *science*.

Art

Science

Data <----> grammar

- Data visualization is as valuable to anyone working with data as grammar is to anyone working with words
- Just as you should not write an essay without proper grammar, you should not create a graph without first mastering data visualization best practices

Visualization is an iterative process

How *thought* leadership helps drive data science impact

- Data **alone** is not **insight**.
  - Humans can tell a better story than data can by itself
  - Numbers rarely speak for themselves. Need context
- Data scientist with storytelling skills have greater business impact
  - Influencing for impact often comes down to conveying a compelling narrative around data and what it means
  - Data scientists who develop this skill typically have the edge in getting their work noticed and acted upon, and enhance their abilities to be acknowledged as an expert by peers and leaders

In your work as data scientists, in addition to doing modeling and machine learning work, you will be responsible (either individually or as part of a team) for providing the following as part of a project:

- **Findings:** what does the data say?
- **Conclusions:** what is your interpretation of the *findings*?
- **Recommendations:** what can be done as a result of the *findings* and *conclusions*?

Source: MLADS Talk

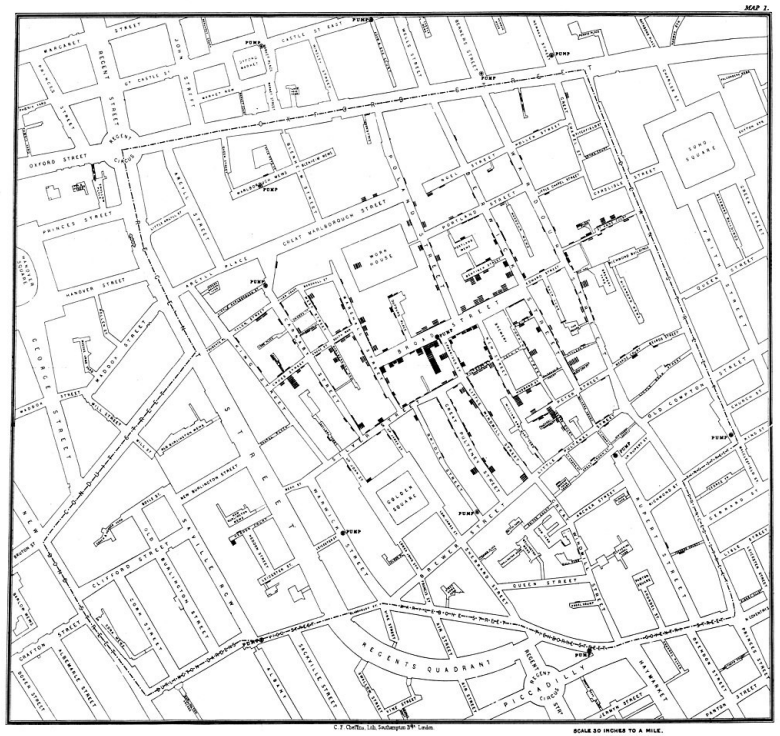
Practice, practice, practice

Why we visualize and a *very* short history of dataviz

Visualization takes advantage of our human capability to understand visual patterns quickly and often intuitively

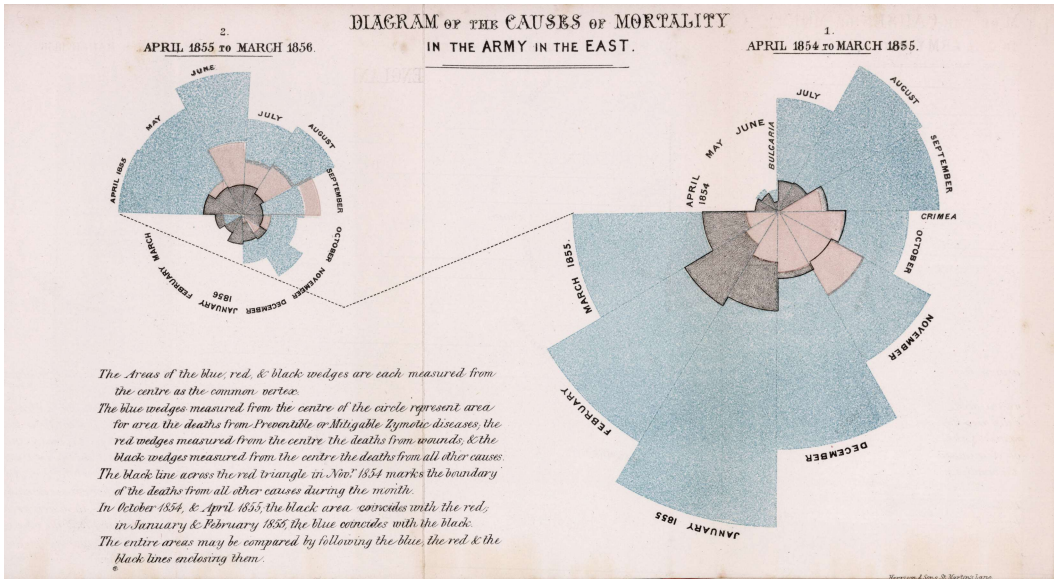
Reasons for visualizing data

Spotting trends



John Snow cholera clusters in London, 1854

Spotting trends



Florence Nightingale, Diagram of the causes of mortality in the army in the East

Analyzing and exploring

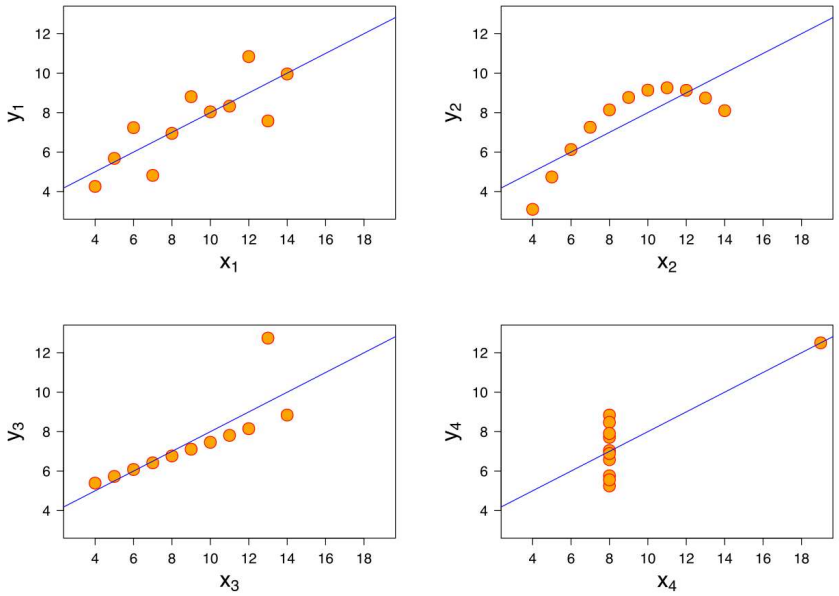
They Rule is a website that allows you to create maps of the interlocking directories of the top 100 companies in the US in 2001.

Telling a story

Classical data summaries can lie!

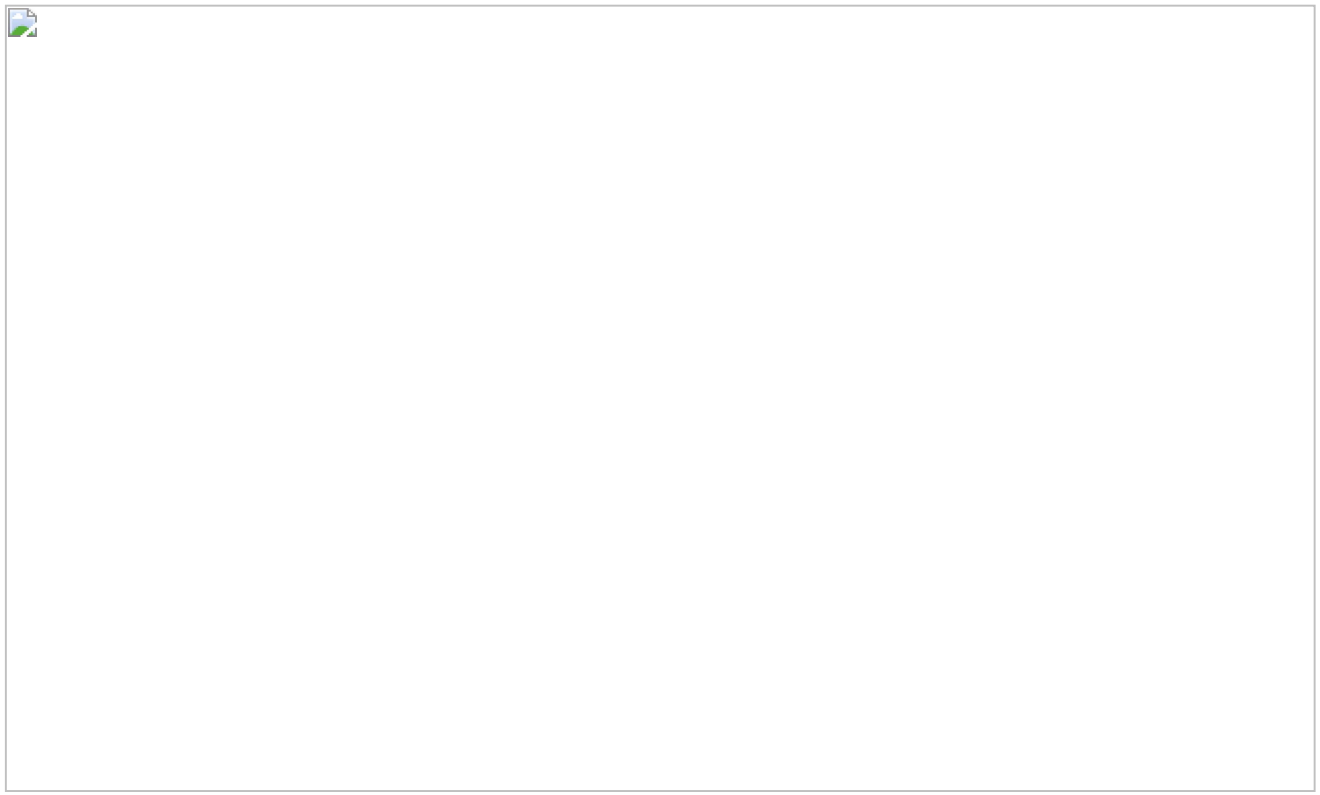
Four datasets with identical properties

Anscombe Quartet



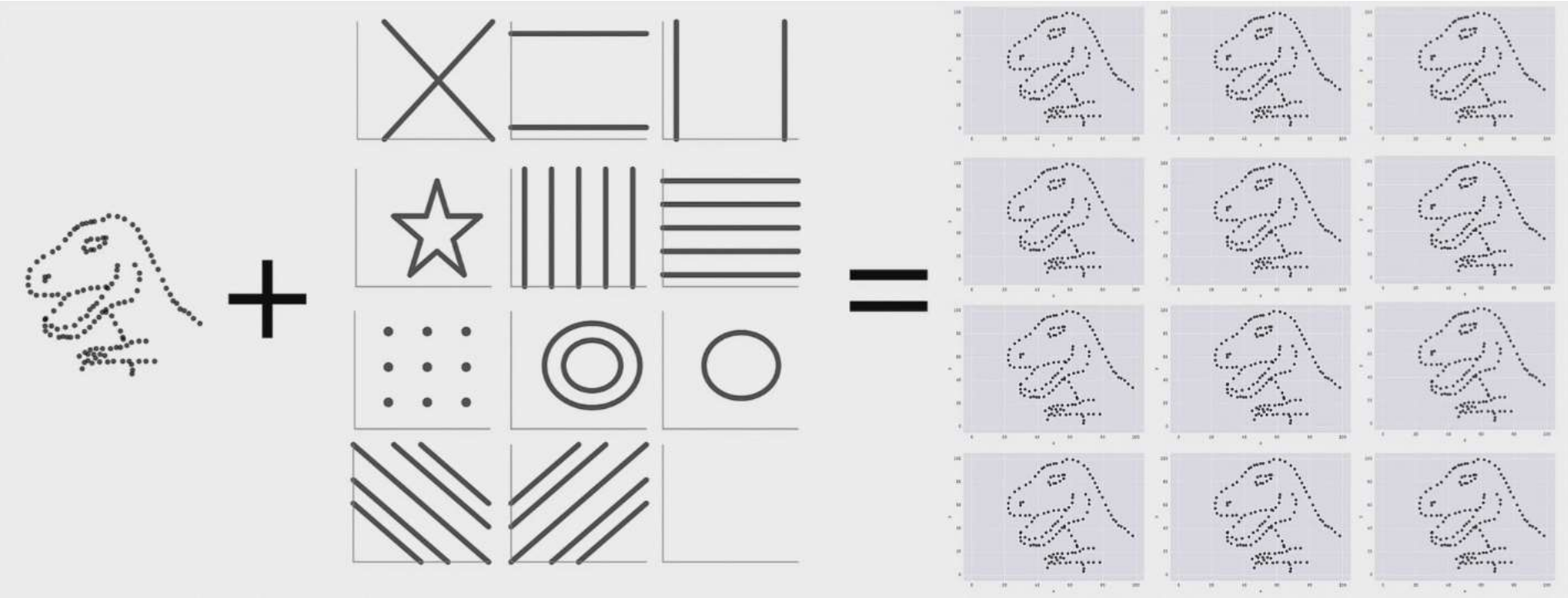
Thirteen datasets with identical properties

Datasaurus



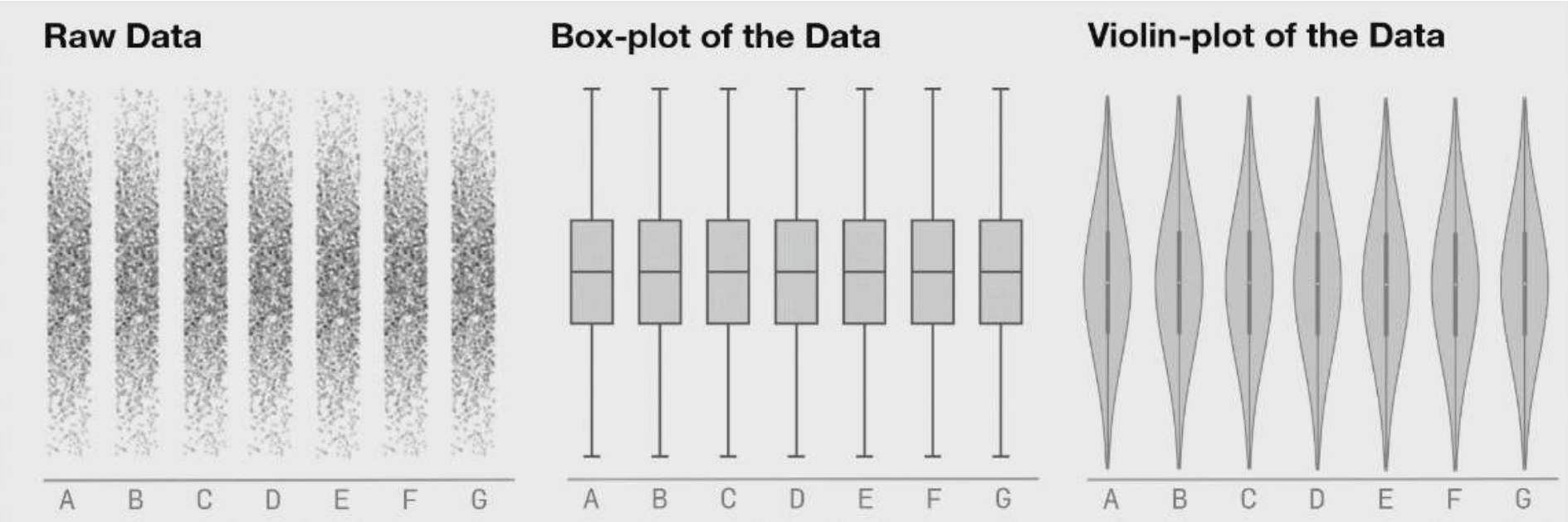
Matejka & Fitzmaurice, 2017

Making the Datasaurus



Matejka & Fitzmaurice, 2017

Summaries don't differentiate



Matejka & Fitzmaurice, 2017

A *very* short history of data visualization

- Astronomical data presentation for navigation
- 1700s: Rene Descartes
- 1800-1900: Graphs and pie charts (WilliamPlayfair)
- 1913: Iowa State University: Introduced some of the first courses in “graphing” data 1. 1977: Princeton University:
- Statistics Professor John Tukey Developed the first exploratory data analysis (EDA) using visualizations.
- 1983: Edward Tufte published “The Visual Display of Quantitative Information” which showed effective visualization methods.
- 1984, Apple Computer introduced the first popular and affordable computer that focused in graphics (GIU) as a mode of interaction and display. This was huge and persists today.
- 1999: The words, “information visualization” were so first named in the book: “Readings in Information Visualization: Using Vision to Think”, Card, Mackinlay, Shneiderman.
- Around 2000ish - : In Few’s opinion, the IBM PC detracted from the value of making graphs by hand as it offers a mouse and computer-application based option. When people made graphs by hand, they took the time to be responsible.

Evolution of computing visualization tools

- Hand drawn
- 1970s: CAD/CAM
- 1980s: Scientific visualization, business visualizations (Harvard Graphics)
- 1990s: Excel, Powerpoint, R
- 2000s: Open source, interactive, web

Back to our original programming...

Think about when you first get a dataset... what might happen?

- You don't know what to expect when you first open up a data file
- You make, remake and review summary charts to get a sense of what you are dealing with
- You throw the dataset into some kind of automatic tool for a quick overview
- You see something odd or interesting and you poke some more in that area

a-ha!

Exploratory data analysis is detective work—numerical detective work—or counting detective work—or graphical detective work.

John Tukey, 1997
Exploratory Data Analysis

You kind of know what you're looking for, but you don't know what you're going to find yet. You work with your bag of tools through the available resources.

You then present to your team...


JOB LOSS BY QUARTER
15 MIL
13.5 MIL
9 MIL
7 MIL
DEC '07
SEPT '08
MARCH '09
JUNE '10
FOX NEWS .COM
SOURCE: BLS
AMERICA'S NEWSROOM
FAIRFAX, VA... BYRD WAS ADMITTED TO THE HO
S&P 3.08

Oh no!

- Your labels are not aligned
- You have to tilt your head to see the data in the right way
- The chart type you have chosen is not ideal
- Your color choice is not color-blind friendly (and your customer is color blind)
- You used ComicSans font
- A variable is named *%$%#@!!*
- Your chart doesn't show properly on a mobile device and the CEO is looking at it from the beach


Who cares, right?

Two broad things you MUST care about:

- How your chart looks for your own use
- How your chart will work for your audience

Adjusting for differences between visualization for analysis and for an audience

A reader who lands on your chart (and the underlying data) may not have the same luxury of developing and answering questions like you did.

Some might know little about data or making sense of it. Some might know more, but they don’t want to analyze the dataset.

They want to know the results.

With this in mind, let's look over the main differences between you making charts for people to consume versus you using charts for analysis.

Visualization for analysis

- tool for understanding datasets
- you ask questions and quickly answer them
- iterate to develop insights

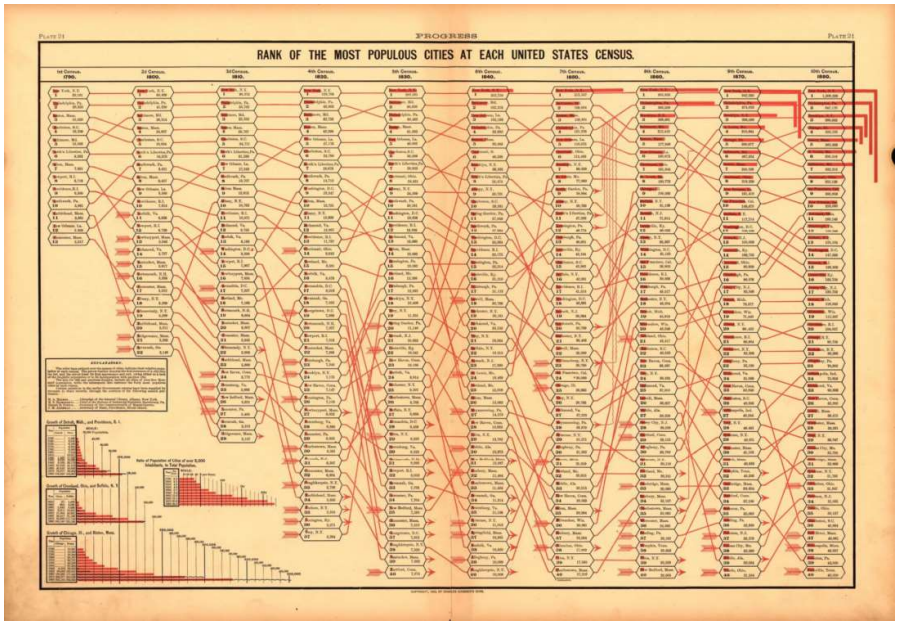
Visualization for presentation

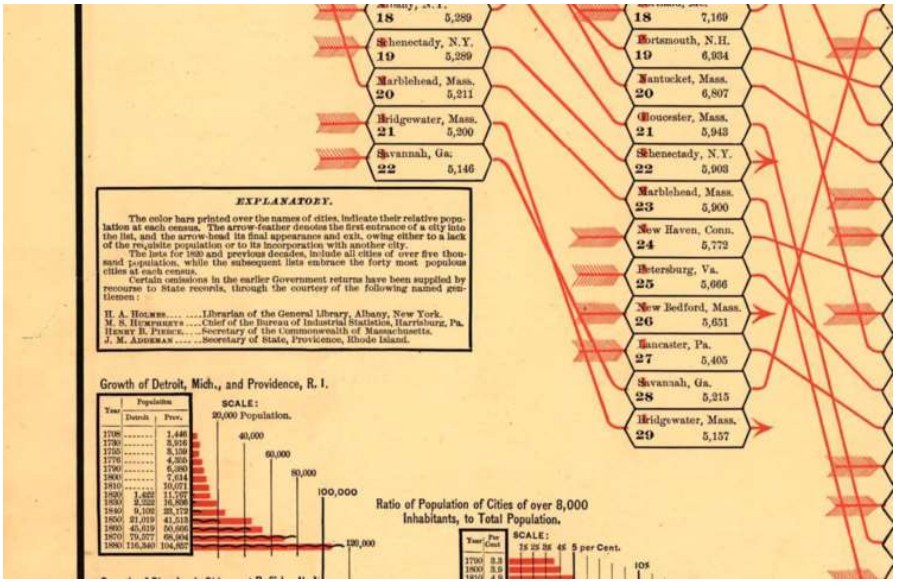
- designed to communicate something useful
- can be a form of entertainment

Four ways to adjust for these differences:

1. Explain the encodings
2. Provide context
3. Focus on readability
4. Develop aesthetics

Explain the encodings





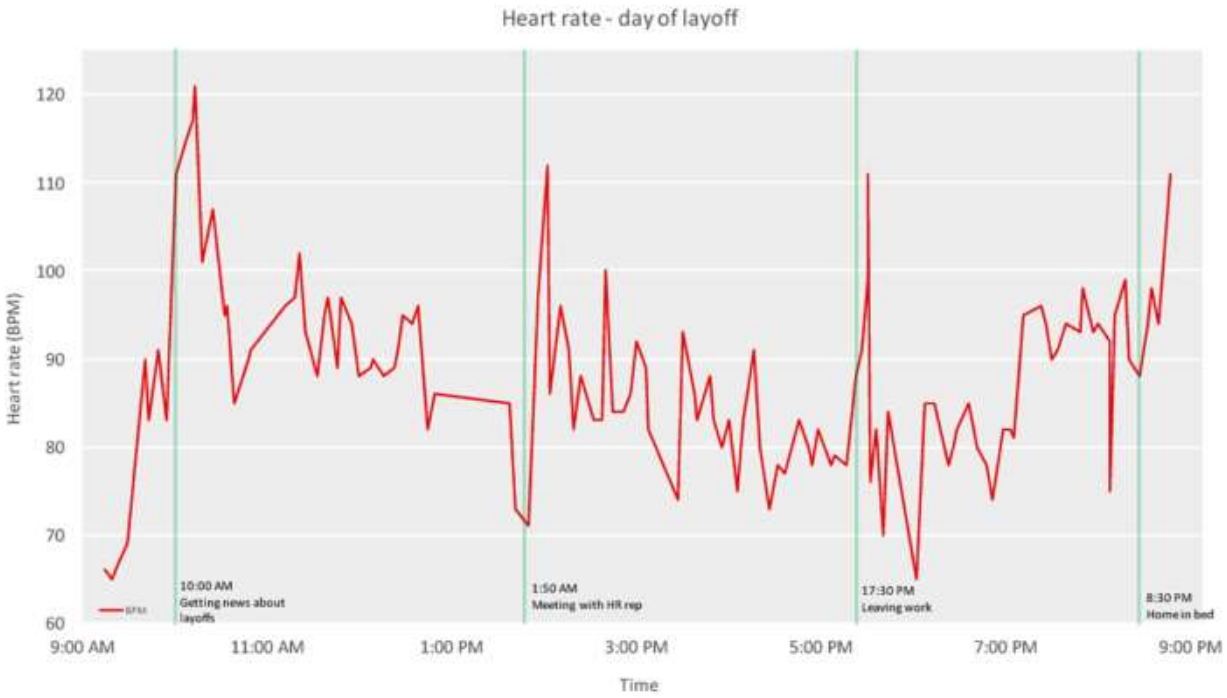
The Statistical Atlas of the United States, produced in the late 1800s

Provide context

When readers can decode the shapes, colors and geometries on your chart, you are more than half way there to producing an awesome chart.

However, **readers also need to understand the context of the data.**

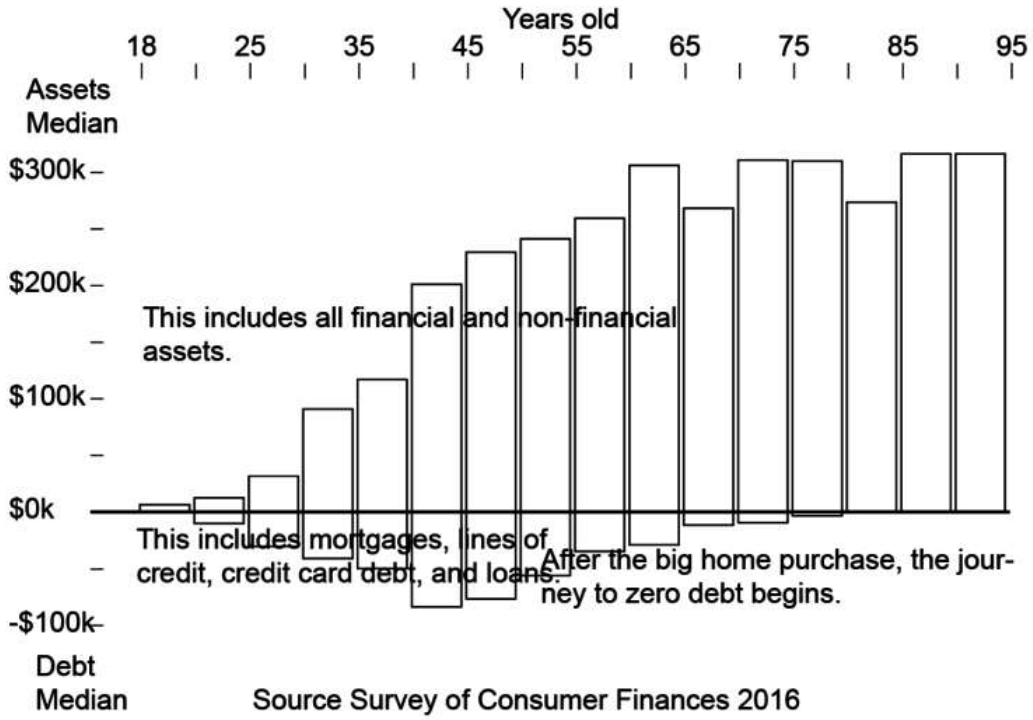
Another context example



The chart itself is not novel or unique, but the annotations make it relevant and contextual.

Improve readability

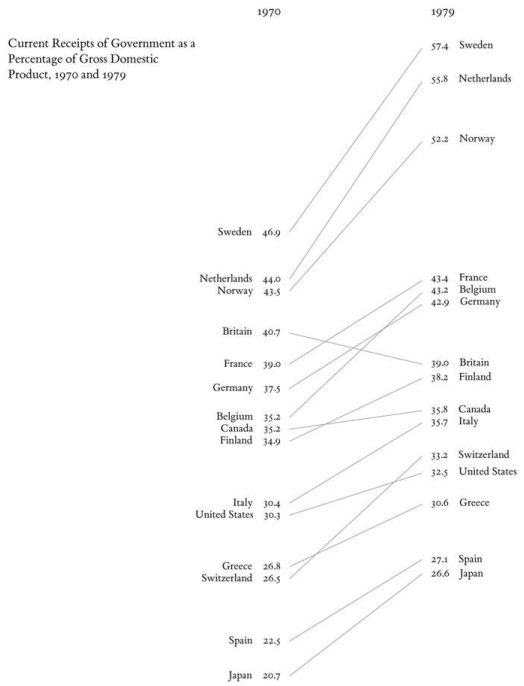
Charts should read like text. At the most basic level, it should be obvious what the chart is about and how to interpret it.



Improve readability

Develop aesthetics

- Default setting in the tools are generic and designed in sucha a way that they would work with many datasets and visualization types
- You can (and should) develop *aesthetics* (your own visual style) to make your charts less ugly

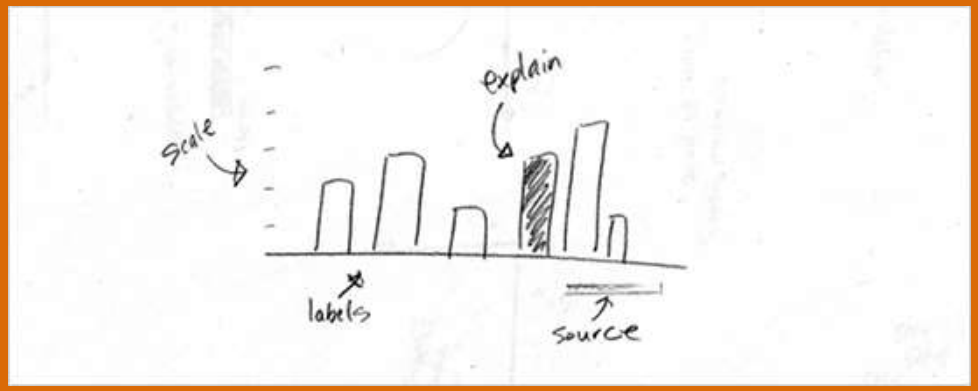


Develop aesthetics (continued)

Using these guidelines

- They're more continuous than binary. Your charts might need more or less explanations, more or less context, etc.
- Depends on your audience and the purpose behind your chart. If your audience is a small group who has the same background as you, then you might not need to provide as much context for the data you show. If your audience is already excited about a dataset, then you probably don't need to make it too flashy. If you make charts for a research paper, there are probably publisher guidelines that you need to follow, which limits what you can do (sometimes a good thing).
- Think of the above adjustments as *continuous knobs* that you can turn up or down. The more charts you make, the better you'll get at deciding how much to turn.

Basic design rules for making charts



Two leading figures

Ed Tufte

Nathan Yau

Design is choice. The theory of the visual display of quantitative information consists of principles that generate design options and that guide choices among options. The principles should not be applied rigidly or in a peevish spirit; they are not logically or mathematically certain; and it is better to violate any principle than to place graceless or inelegant marks on paper. Most principles of design should be greeted with some skepticism, for word authority can dominate our vision, and we may come to see only through the lenses of word authority rather than with our own eyes.

--- Edward Tufte, *The Visual Display of Quantitative Data*

Tufte's Principles of Graphical Integrity

1. Show data variation, not design variation
2. Do not use graphics to quote data out of context
3. Use clear, detailed, thorough labelling.
4. Representation of numbers should be directly proportional to numerical quantities
5. Don't use more dimensions than the data require

Tufte's Principles of Graphical Integrity

1. Show data variation, not design variation
    - Don't get fancy, let the data speak
2. Do not use graphics to quote data out of context
    - Maintain accuracy
3. Use clear, detailed, thorough labelling.
    - Use annotations to make your point
4. Representation of numbers should be directly proportional to numerical quantities
    - This is essential for fair representation
5. Don't use more dimensions than the data require
    - Be appropriate in use of 3D graphics, for example

Tufte's Fundamental Principles of Design

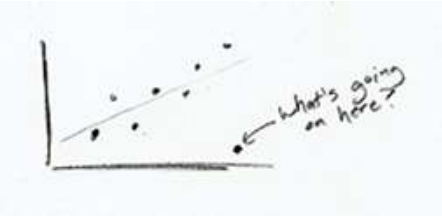
1. Show comparisons
2. Show causality
3. Use multivariate data
4. Completely integrate modes (like text, images, numbers)
5. Establish credibility
6. Focus on content

Nathan Yau's Seven Basic Rules for Making Charts and Graphs

1. Check the data
2. Explain encodings
3. Label axes
4. Include units
5. Keep your geometry in check
6. Include your sources
7. Consider your audience

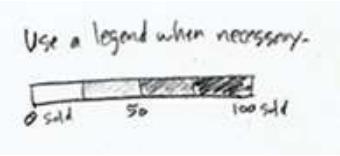
Nathan Yau, Flowing Data
https://flowingdata.com/2010/07/22/7-basic-rules-for-making-charts-and-graphs/

1) Check the data



- This should be obvious
- If your data is weak, your chart is weak
- Start with simple graphs to see if there are any outliers

2) Explain encodings



- Don't assume the reader knows what everything means
- Provide a legend
- Label shapes
- Explain color scales

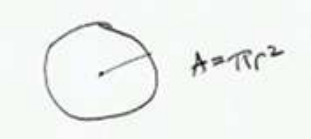
3) Label axes

- Axes without labels or explanation are just decoration
- Describe the scale (incremental, exponential, logarithmic?)
- Have axes values start at zero

4) Include units

- Numbers without units are meaningless
- Remove the guesswork

5) Keep your geometry in check



- This is something that is immediately noticeable
- Don't use area to compare two units unless they are an area. An increase in a unit squares the area.
- Tip: size circles and other 2D shapes by area, unless it's a bar chart

6) Include your sources

- This is another obvious one
- Always include the source of your data
- Makes your graphic more reputable
- Allows for others to dig deeper

7) Consider your audience

- What purpose do your charts have and who are they for?
- Avoid quirky fonts
- Make good design choices

Choosing the right chart type

Why do we visualize data?

Record information

- Blueprints, photographs, seismographs, ...

Analyze data to support reasoning

- Develop and assess hypotheses
- Find patterns and discover anomalies in data

Communicate information to others

- Share and persuade
- Collaborate and revise

What is the best way to visualize your data?

- What do you want to show?
  - What do you want to emphasize?
- Why do you want to show it?
  - What is the message you want to convey
- Who are you showing it to?
  - Understand what your audience will be receptive to
  - What is *their* context?

Is choosing the right visualization a straightforward choice?

Smaller datasets

- Look at the data
- Use multiple looks to understand the data
- Choose which patterns you want to visualize

Larger datasets

- Use random sampling to look at smaller sub-samples
- Experiment
- Methods are advancing to enable big data visualization (later this semester)

The chart selection process is *not* mechanical

Just as you can't

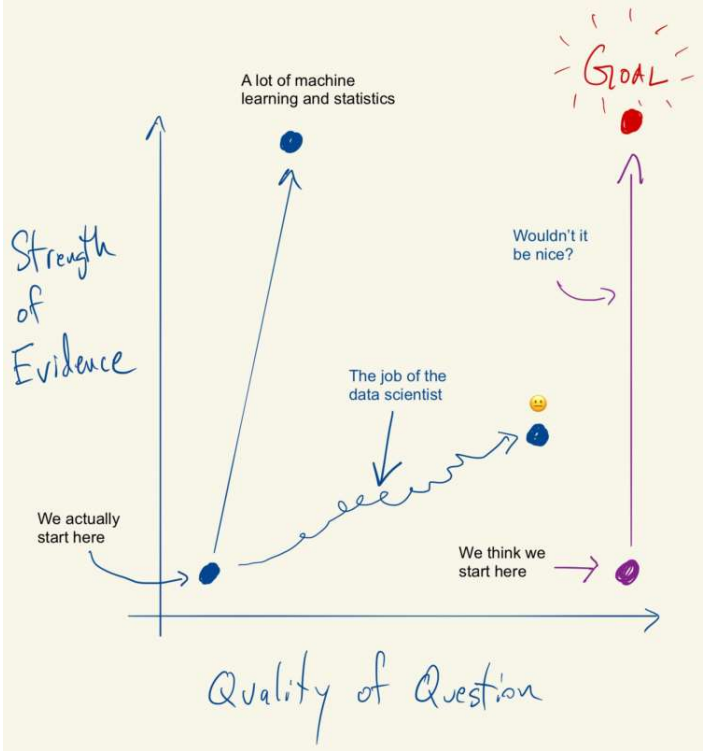
- randomly place a bunch of words together to make a book
- randomly record videos and get a finished film out of them
- randomly grab ingredients from the pantry, toss them in the pan and expect a great meal...

You cannot just put a chart together as a sequence of steps.

However, there is still a method and a mental model

Ask and answer questions

- There are many different ways to express a story from data
  - Blind men and the elephant (different perspectives)
  - Changing vantage points (different views)
  - You can change your vantage point and how you want to see the data
  - Nathan Yau shows 25 ways to see a data
- Meaningful analysis requires
  - context,
  - background, and
  - a human in the loop
- Different questions can lead to different chart types and focus



Choosing your data format

Recipes for selecting the right chart

From Data to Viz by Yan Holtz and Conor Healy

The Data Viz Project by ferdio

Multiple views on how to choose a visualization by Steven Franconeri

Slide Chooser by Andrew Abela

The Graphic Continuum by Jon Schwabish and Severino Ribecca

However, this *isn't a "if this, then that" scenario*

There can be multiple views that show different aspects of the data

All can be useful, and equally "correct"

The real question is, does the visualization convey your story in a way that is accurate and that your audience can receive, digest and understand

Creating a chart by splitting it into components

No chart is made completely in a single pass

- A chart is not a single **monolithic** element, so don't think of it as one
- Perhaps this thought (single element) may work for standard charts like bar charts, line charts and scatterplots because most software tools provide quick ways of creating them, with *reasonable* defaults
- What do you do when even a basic chart or a single element is off?

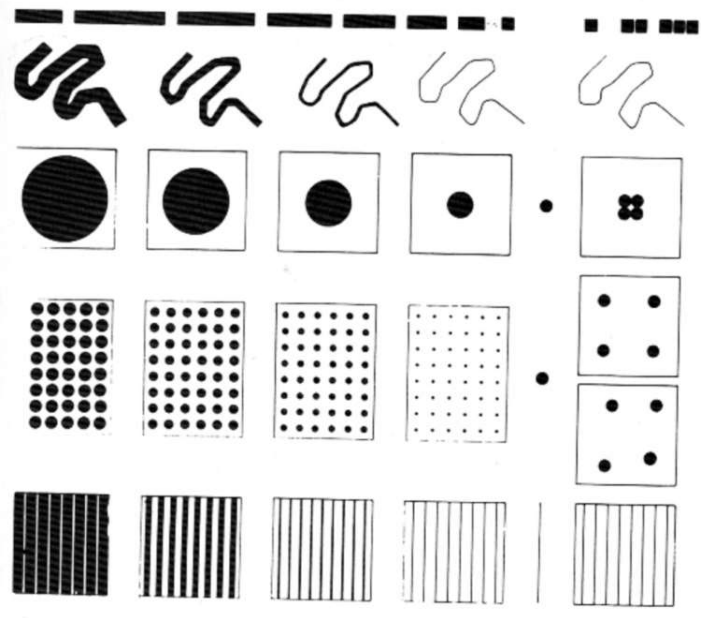
You split the chart into components

- The basic mental model is that charts are **compositional**
  - There are building blocks and ways to put them together
  - If you understand the relevant parts, you can **compose** charts by mixing and matching and layering and joining

This is a very powerful model

Plane and *retinal variables*

A **plane** is like the coordinate system that defines how geometries are placed in a space. A **retinal variable** defines how to encode data into visuals.



Jacques Bertin, *Semiology of Graphics*, 1967

The Grammar of Graphics

William S. Cleveland, in his 1994 book *The Elements of Graphing Data*, lists the "basic elements of graph construction" as **scales, captions, plotting symbols, reference lines, keys, labels, panels, and tick marks**.

In *The Grammar of Graphics*, published in 2005, Leland Wilkinson built off the work by Bertin and more formally defined the components of a graphic:

Statistical graphic specifications are expressed in six statements:

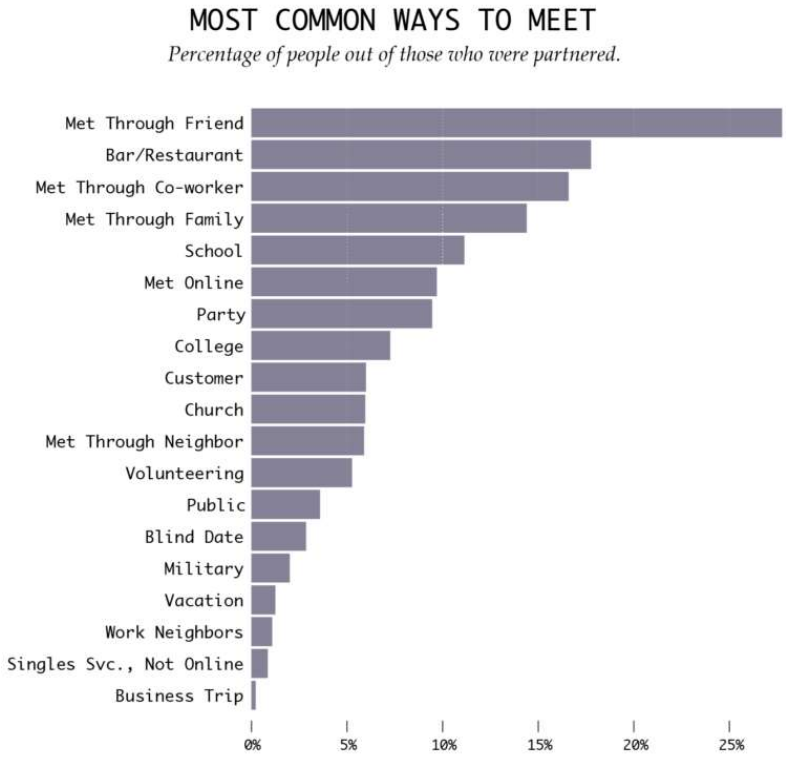
Statement	Description
DATA	a set of data operations that create variables from datasets
TRANS	variable transformation (e.g. rank)
SCALE	scale transformations (e.g. log)
COORD	a coordinate system (e.g. polar)
ELEMENT	graphs (e.g. points) and their aesthetic attributes (e.g. color)
GUIDE	one or more guides (axes, legends, etc.)

Hadley Wickham implemented Wilkinson's grammar in R with the popular `ggplot2` package.

Strategies for breaking charts into individual components

- The **data** drives all decisions
  - The purpose is to convey the **information** in the data
- The **visual encodings** dictate the geometry and/or colors of a graphic
  - This forms the **aesthetics** of the visualization
  - This most influences how the visualization is received
- The **coordinate system** (Cartesian, polar, or geographic) specifies the space in which the visual encodings reside.
  - This provides the canvas, scales and orientations upon which we visualize
- The **context** communicates what the data is about, where it is from, and why it exists.
  - This can be provided through textual annotations, legends, etc.

Example 1: breaking up a chart

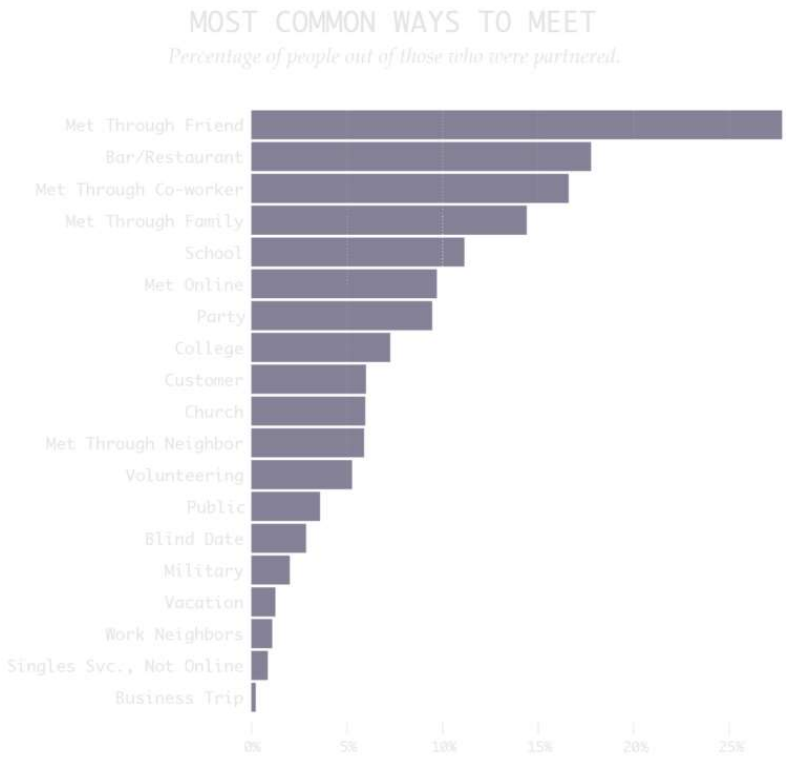


A study on how people first met: https://data.stanford.edu/hcmst2017

The *data* is aggregated percentages

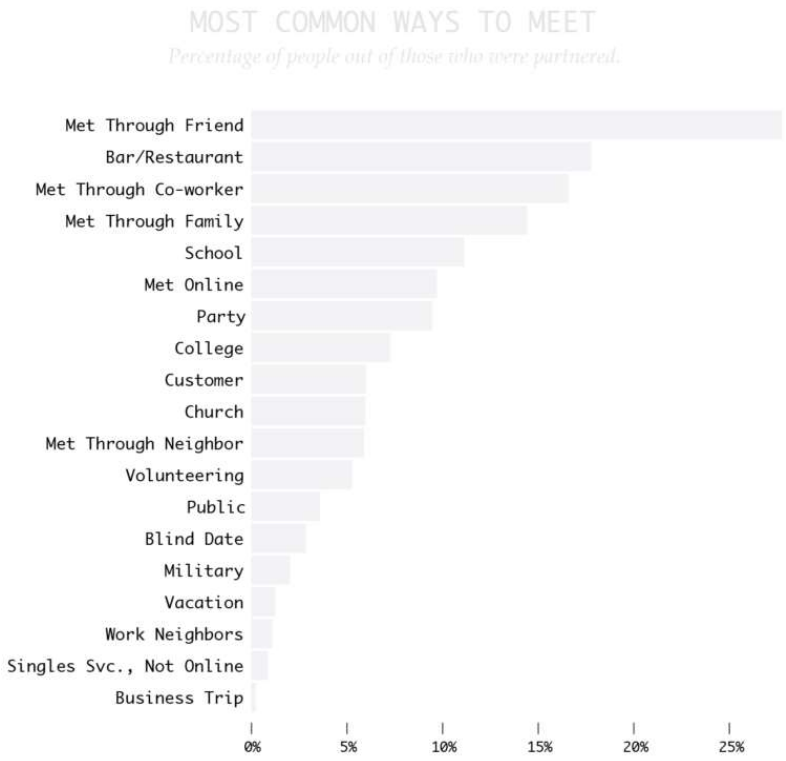
```
waymet           n           p
18         business_trip   7.260476 0.002214773
17   single_serve_nonint  27.895815 0.008509483
19        work_neighbors  35.471720 0.010820476
16              vacation  40.959558 0.012494514
8                    mil  65.504234 0.019981748
15           blind_date  93.696132 0.028581549
...
```

The bars are the *visual encoding*



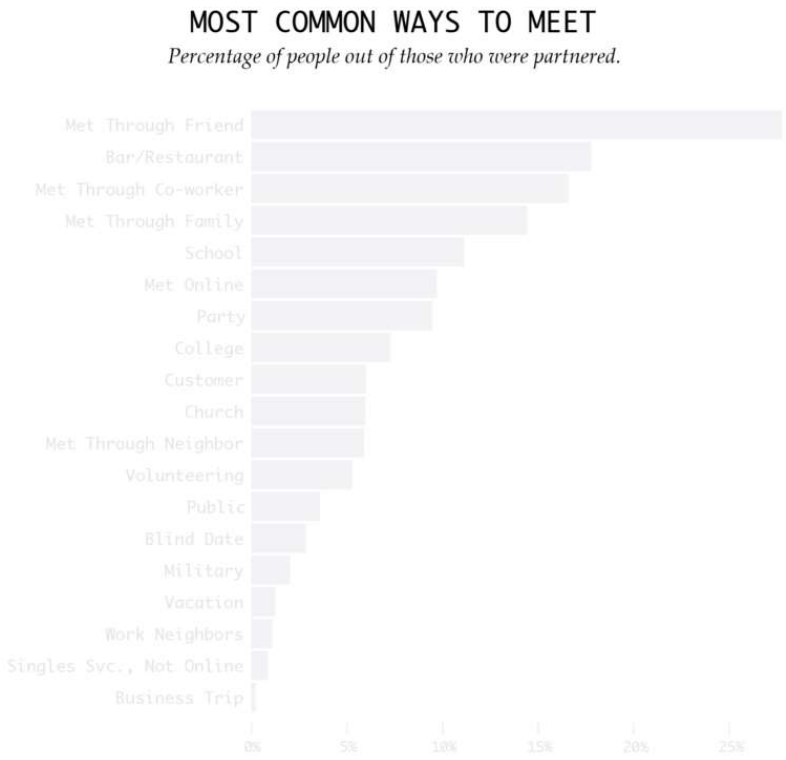
Length represents a percentage

The *coordinate system* is Cartesian



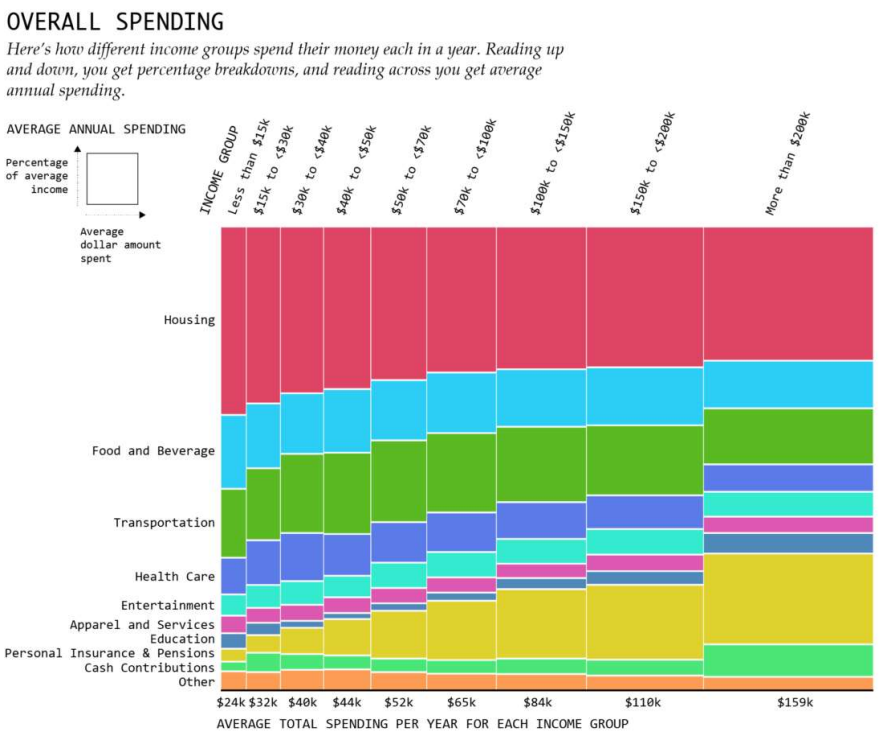
A *linear* scale on the horizontal and a *categorical* scale on the vertical axis

Additional information provides *context*



Titles, labels, markings, etc.

Example 2

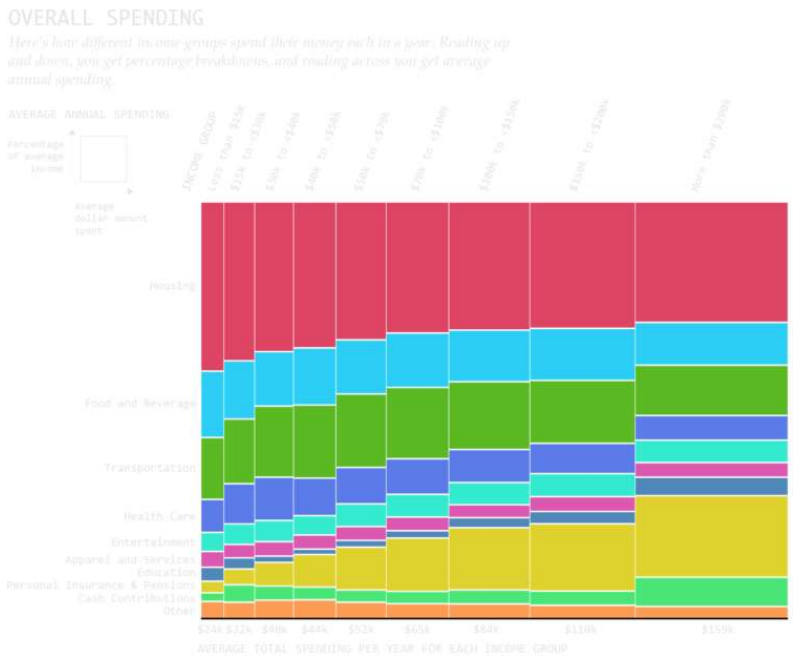


A *mosaic plot* that shows average spending by income group in the United States

The *data* is the average dollar amount for each income group

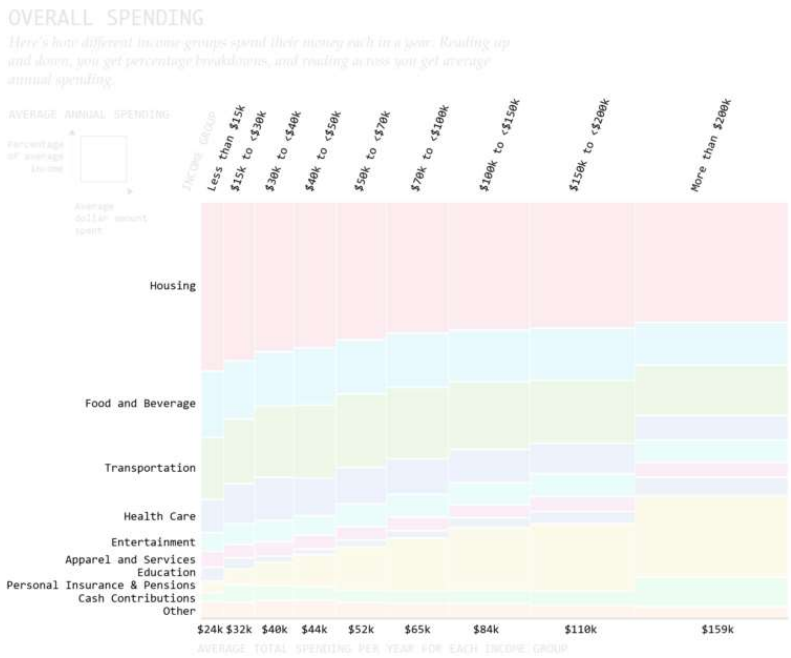
```
item    hierarchy_level all Less than $15,000   $15,000 to $29,999  $30,000 to $39,999  $40,000 to $
Food and Beverage   1   7216    3771    4453    5226    6040    6744    8453    10362   13571   161(
Food at Home    2   4049    2450    2904    3064    3656    3893    4772    5554    6718    7135
Food Away From Home 2   3154    1318    1533    2157    2371    2847    3664    4797    6832    891(
Alcoholic Beverages 2   484 133 215 280 320 420 596 734 1169    1659
Housing 1   18886   9698    12268   14533   15575   17331   20564   26003   33319   46076
...
```

What are the *visual encodings*?



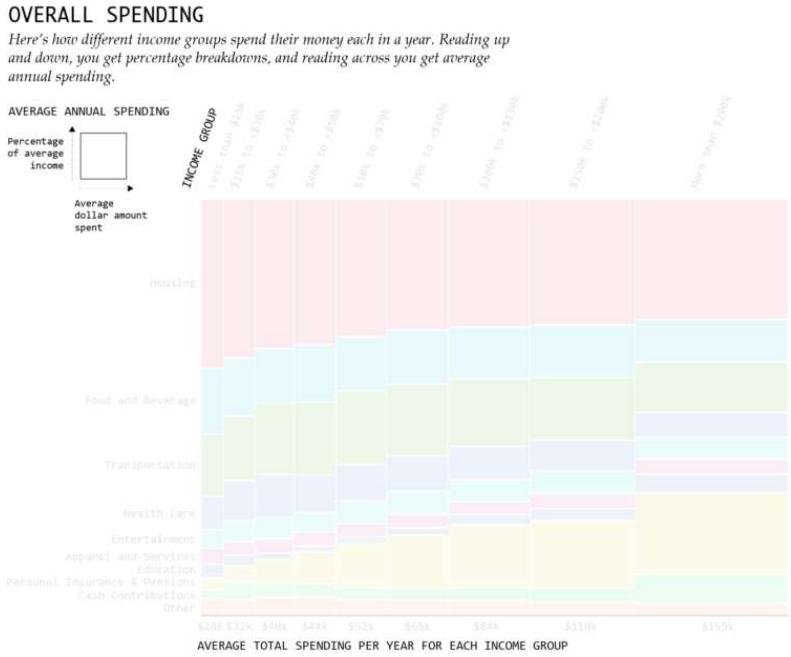
Height: percentage, width: average total spending for each income group, color: spending category

What is the coordinate system?

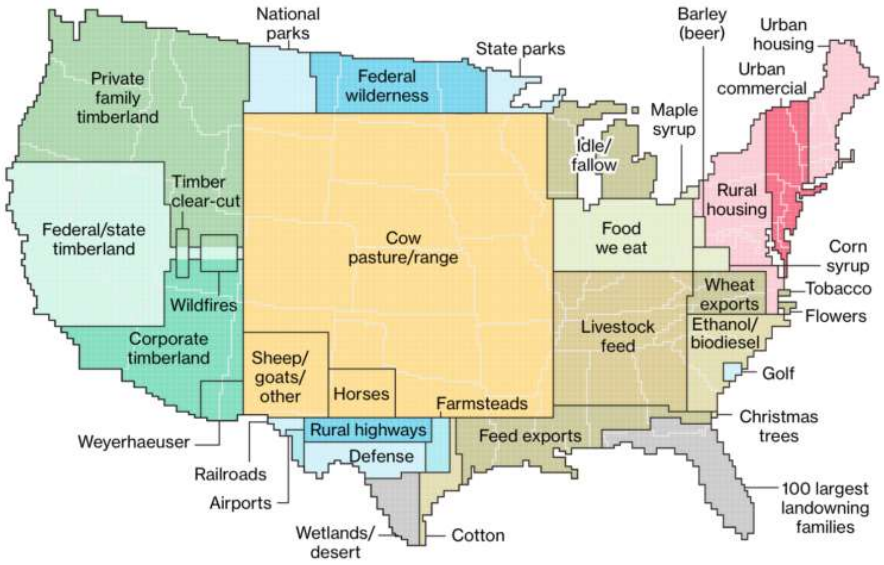


Cartesian: x-axis is total dollars spend, y-axis is 0-100% for the group

What is the *context*?



Let's decompose this one

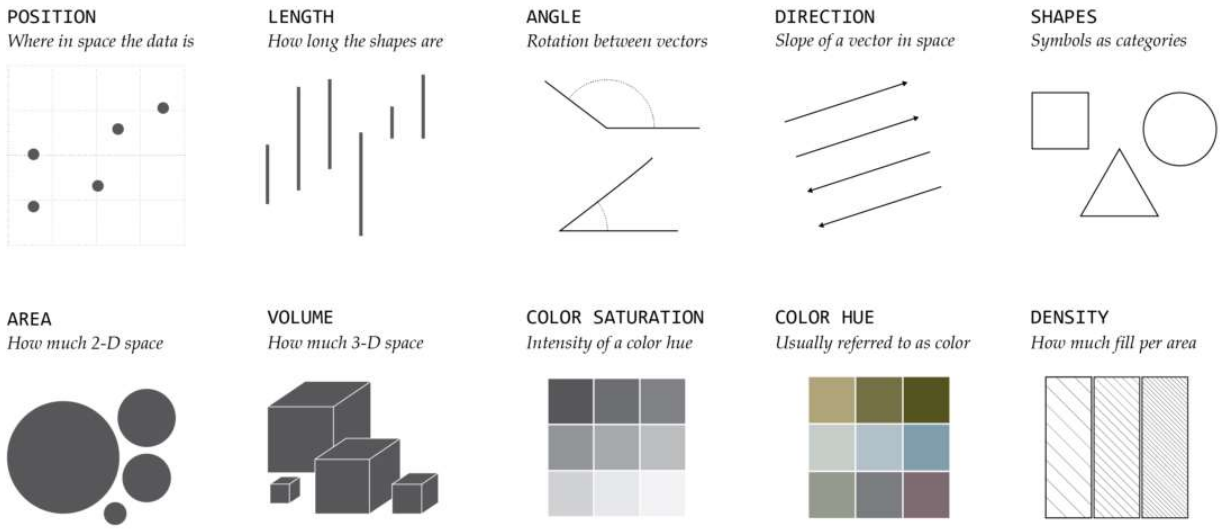


Discuss

- Data
- Visual encodings
- Coordinate system
- Context

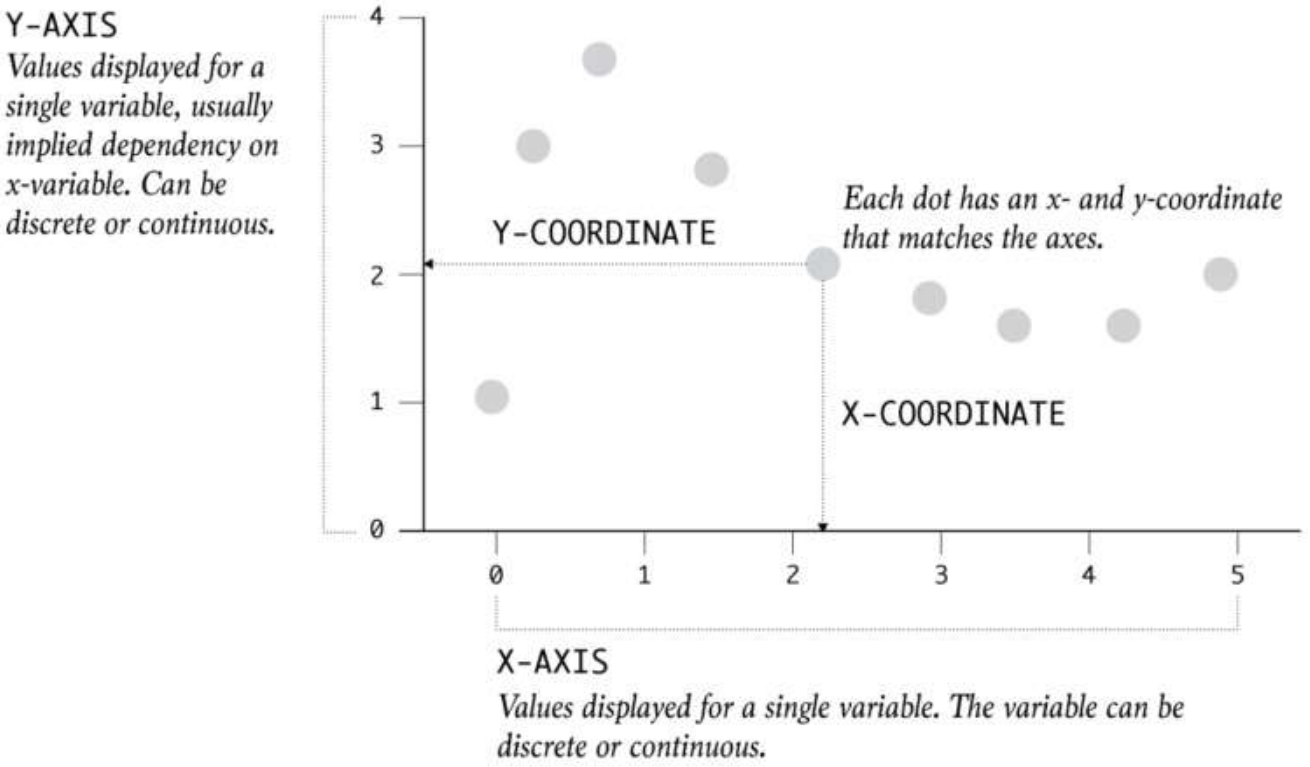
Visual Encodings

Visual encodings can be categorized into the main groups below

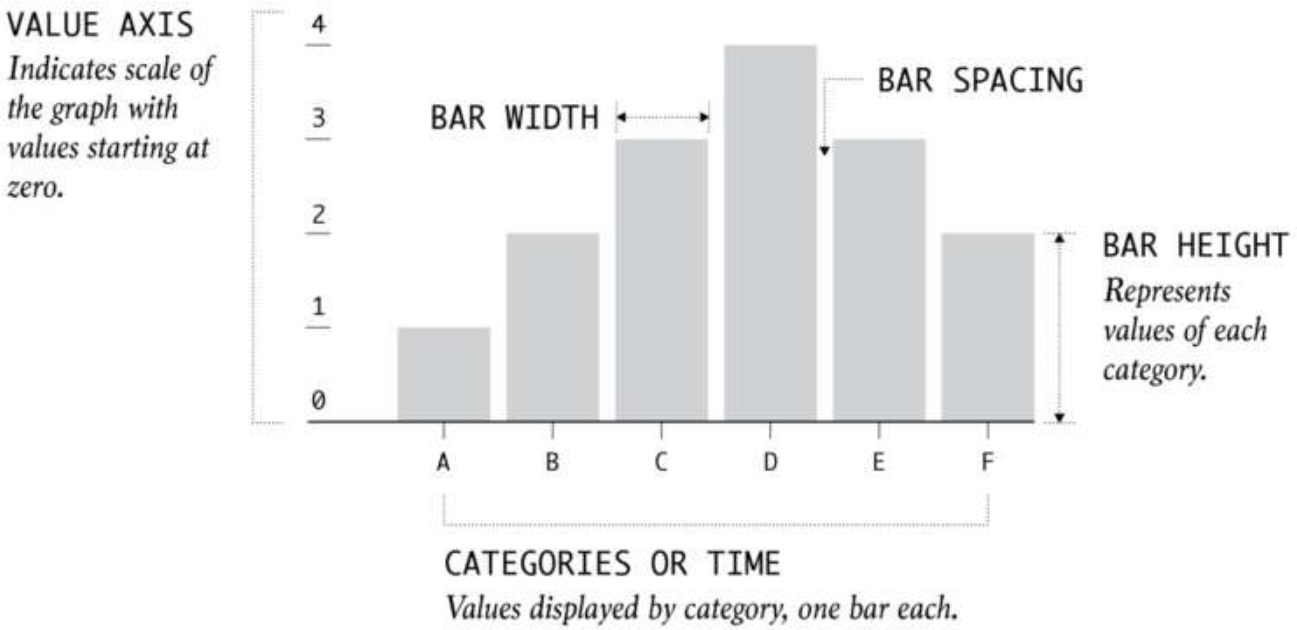


All visualizations use some combination of these

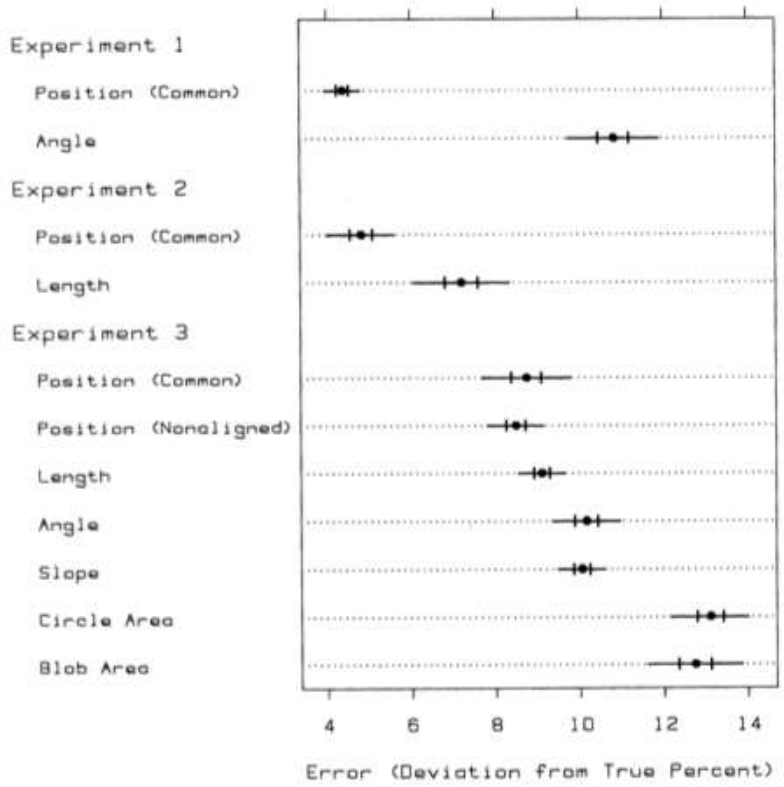
Example: a *scatterplot* uses position on an x-y scale



A *bar chart* uses length to show values



Some encodings work better than others



Decoding Error

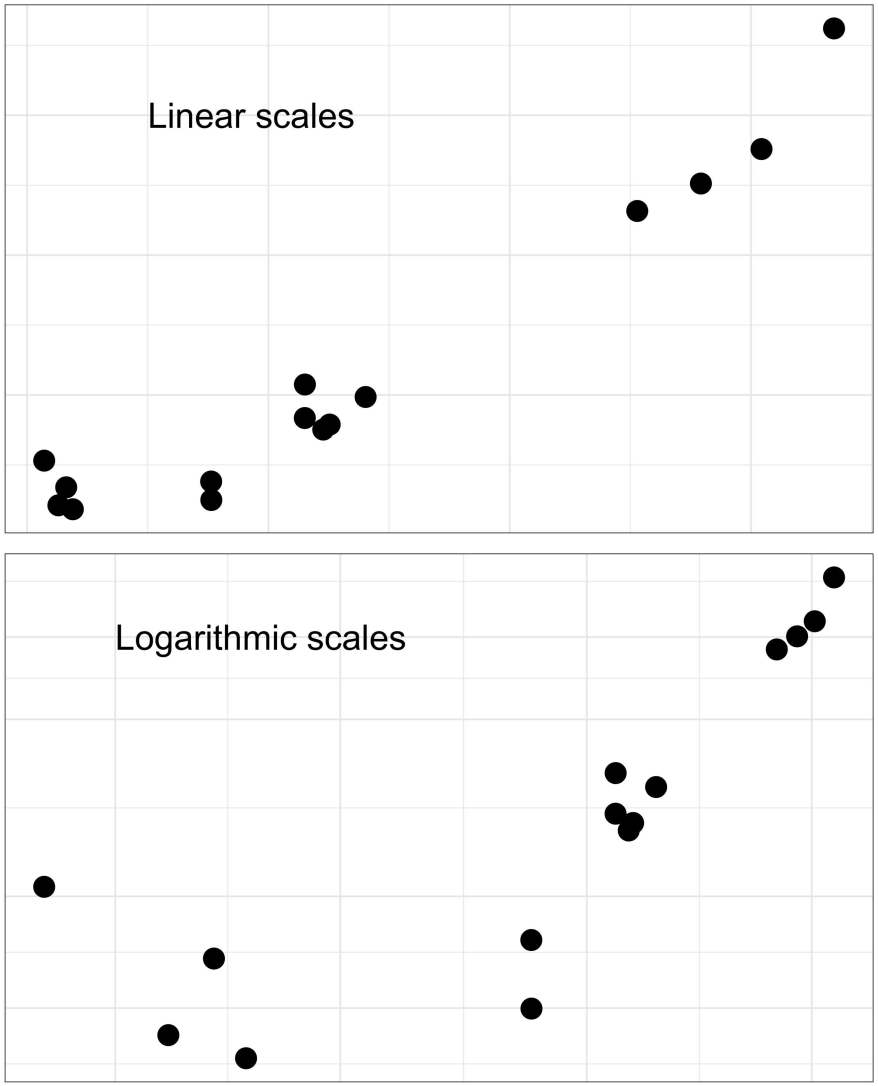
1. Position along a common scale
2. Length
3. Angle and slope
4. Area
5. Volume, density, and color saturation
6. Color hue

Position

Position allows you to compare values based on where they are placed with reference to a coordinate system.

Considerations

- Be aware of the scales you are using (linear vs logarithmic)
  - The scale changes the interpretation of distance
  - It can also change the perceived patterns



Position

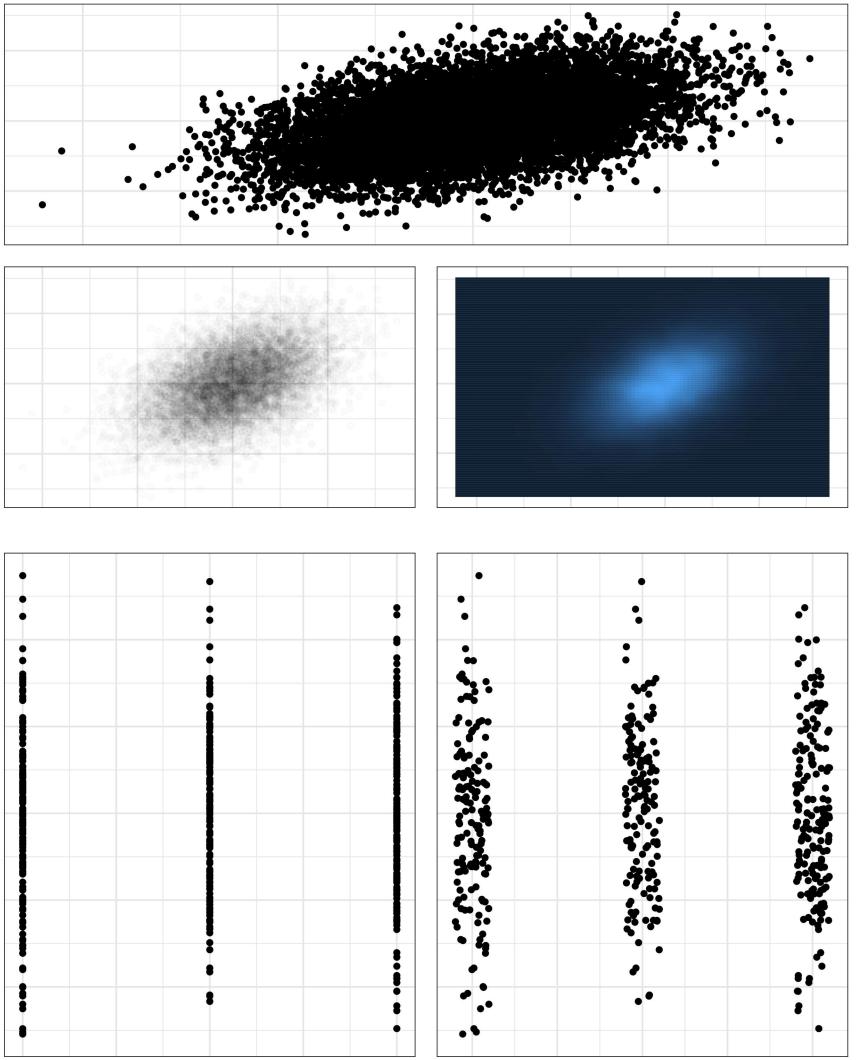
Position allows you to compare values based on where they are placed with reference to a coordinate system.

Considerations

- Avoid overplotting since many points can occupy the same space and obscure one another

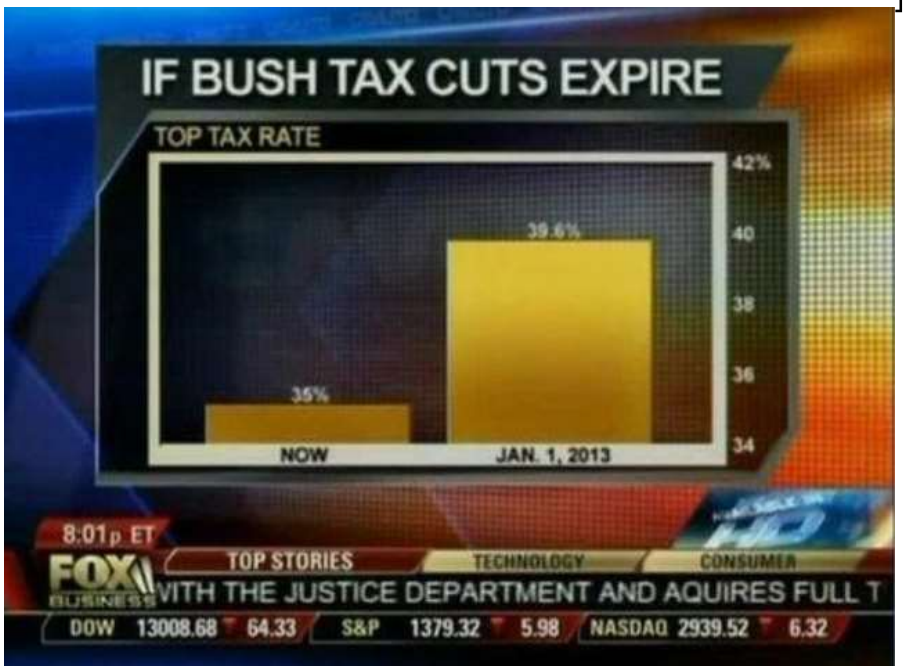
Solutions

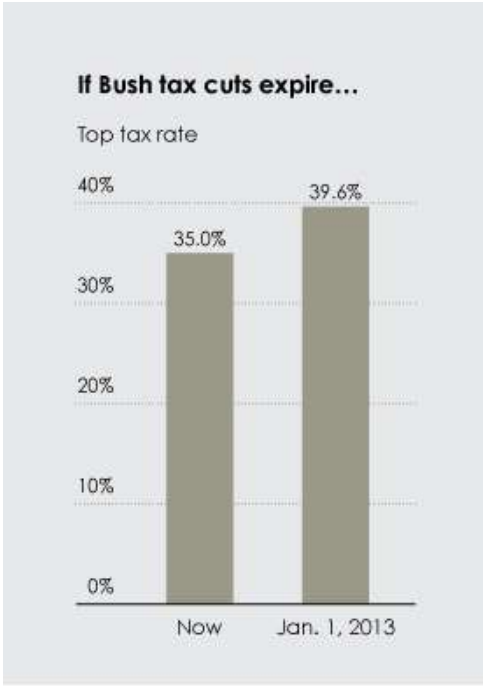
- **Use transparency** so that overlapping points make darker areas
- **jitter** (add noise so points no longer are on top of each other)
- **Use binning** to show aggregate data per pixel



Length

Length is most commonly used in the context of bar charts. The longer a bar is, the greater the value. **Don't truncate bar charts, use length in its entirety!**





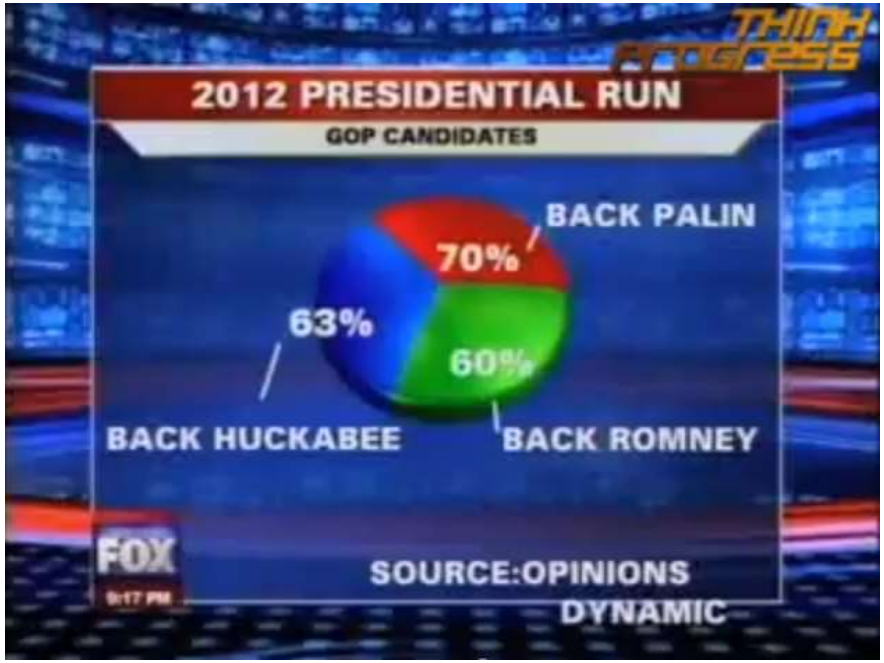
BAD FOX news, again...

Angle

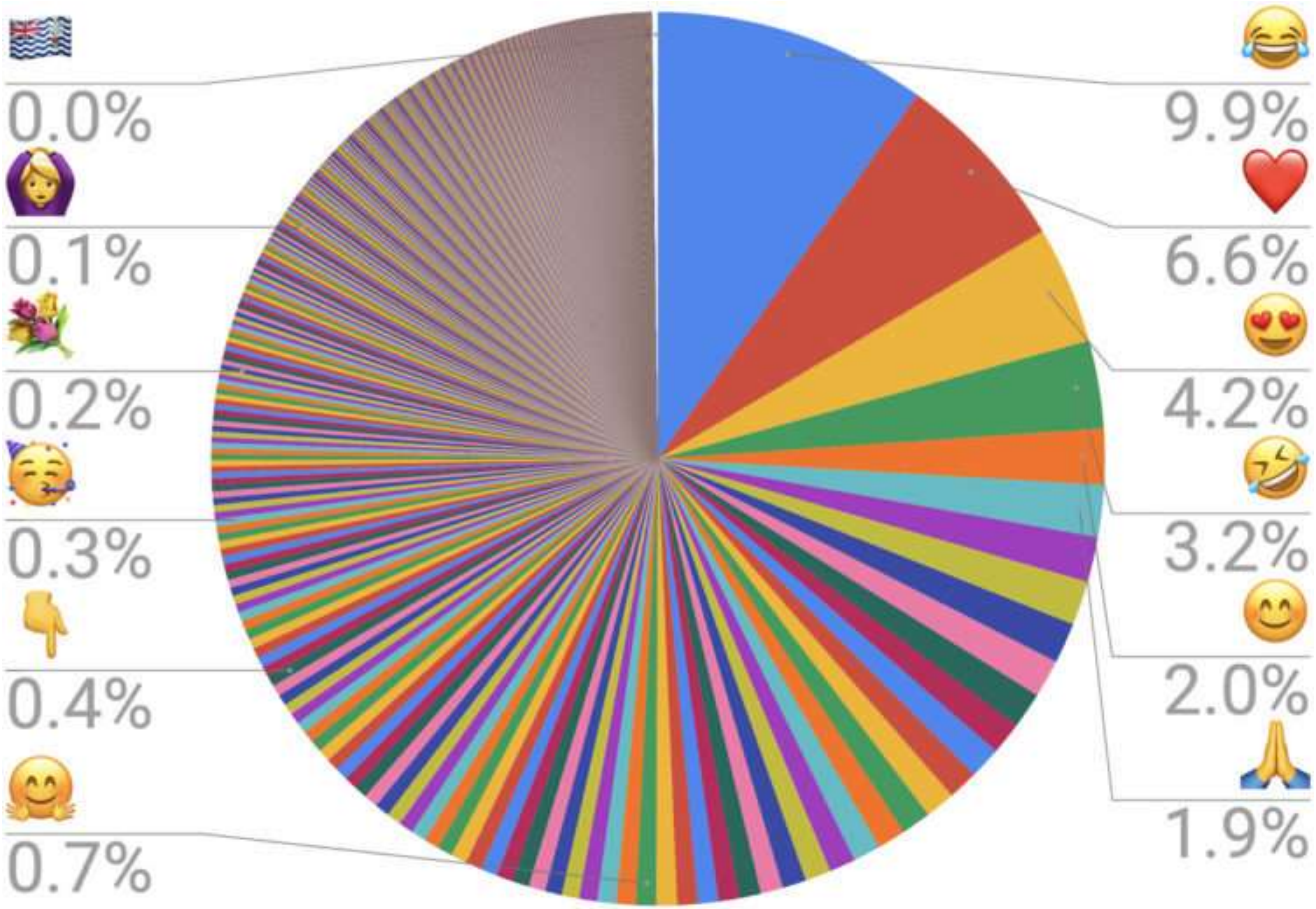
Angles range from 0 to 360 degrees in a circle.

Considerations

- Angles are most associated with *pie charts*. Pie chart is made up of parts that make up a whole.
- Don't use too many categories (bar chart is better)
- **The sum of all percentages should equal 100%!**



Don't even think about this!



Slope

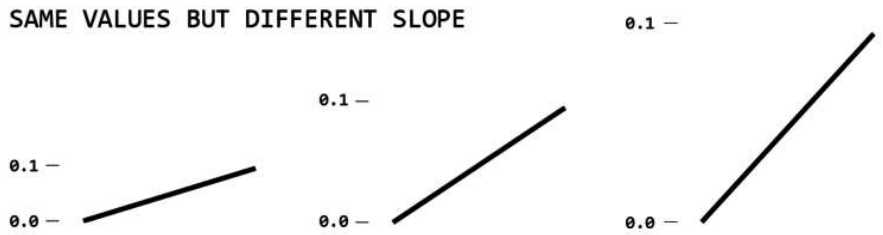
Slope is similar to *angle*. Line charts are the most common use of slope to encode data.

Considerations

- Slope magnitude: steeper = greater change, flatter = lesser change
- The aspect ratio
- Visual change should match the context of the change

Cleveland, McGIll & McGill (1988) suggested that the average slope in a line chart should be 45^{o}, in order to make neutral comparisons between lines

This is still a good rule of thumb

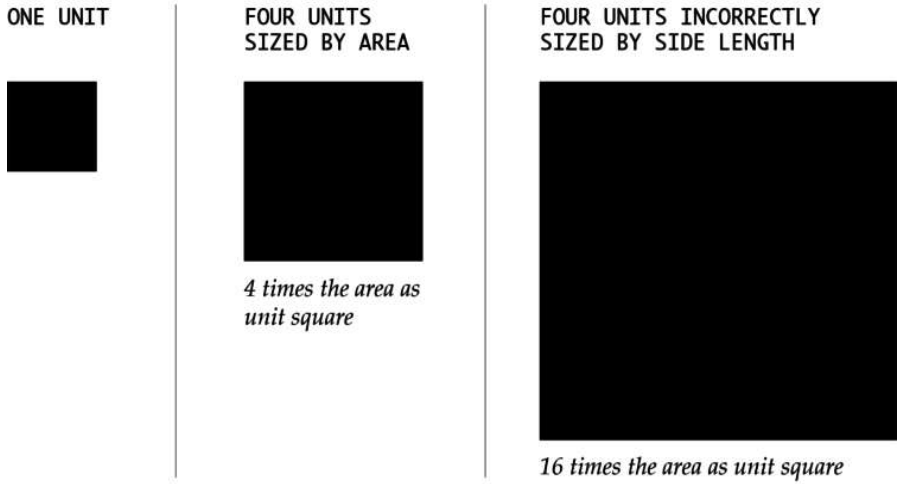


Area

Like *length*, *area* can be used to represent data with size, but with two dimensions instead of one.

Considerations

- While the encoding might not be as precise from a visual perception perspective, area can provide a more intuitive, less abstract view for some types of data
- Make sure you scale by area, not edge (remember, area gets squared per unit increase)
  - This means you should encode the length of a side as $\sqrt{x}$



A *treemap* uses rectangle areas to show hierarchical data

taken from **Flowing Data**

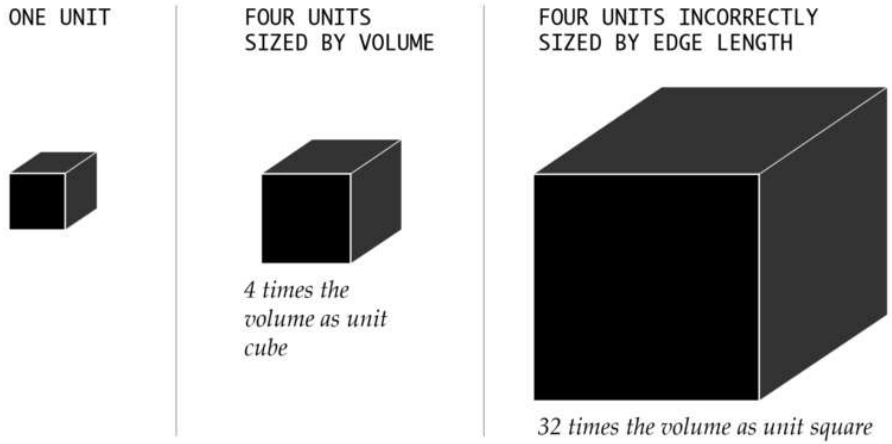
Volume

Volume can used in the same way as *area* but has one more dimension.

Considerations

- Make sure you scale by volume, not edge (remember, volume gets cubed per unit increase)
  - This means you would encode the side of a "box" as $x^{1/3}$

For 3-D encodings, you need to take the volume as proportional to the data



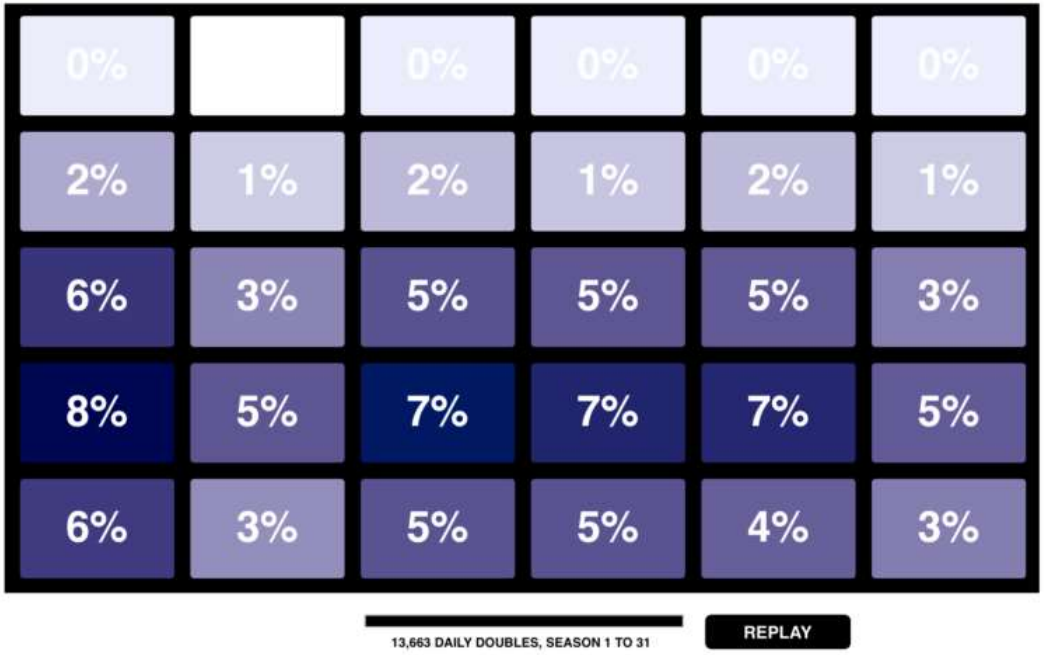
The volume, or 3D perspective representation can make tangible data more relatable

Color

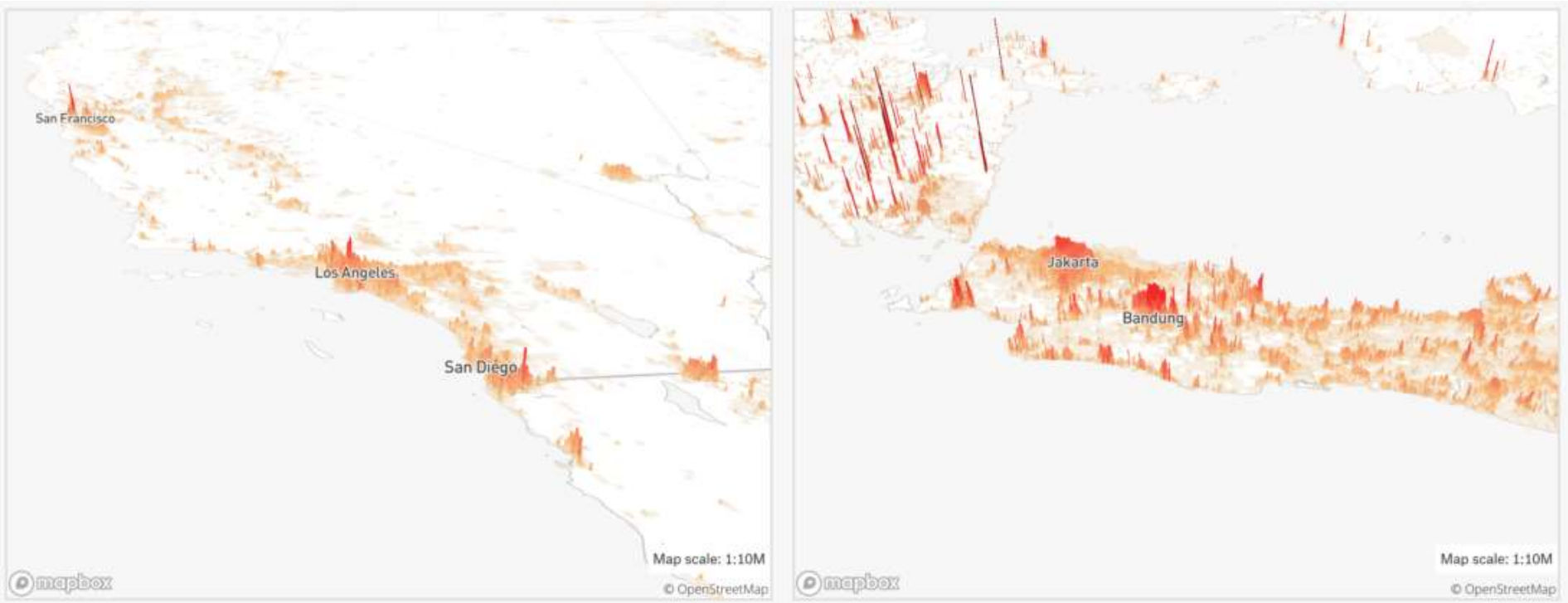
Color as a visual encoding can be split into two categories: **hue** and **saturation**. **Hue** is what most people refer to as color (red, green, blue, etc.) **Saturation** is the amount of **hue** in a color.

- Qualitative: every color represents a distinct attribute (category)
- Sequential: color represents a range (**saturation**) from low to high (or vice-versa)
- Diverging: multiple hues represent a point of inflection of the data

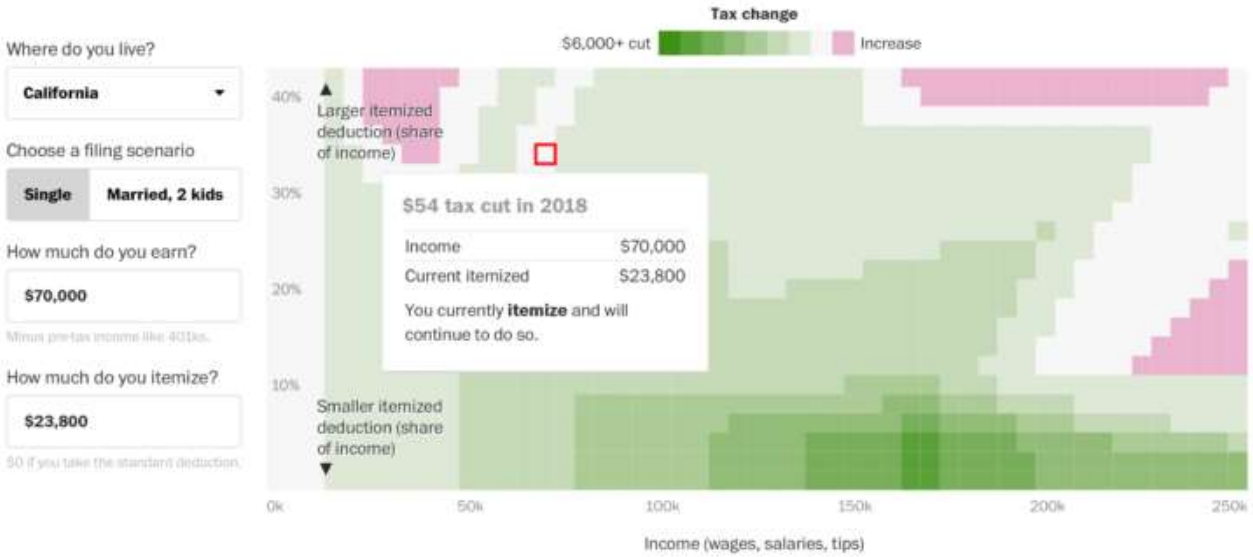
Sequential example



Another sequential example

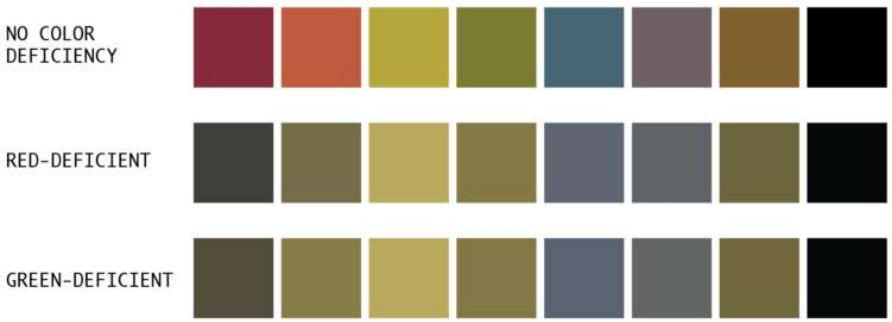


Divergent example



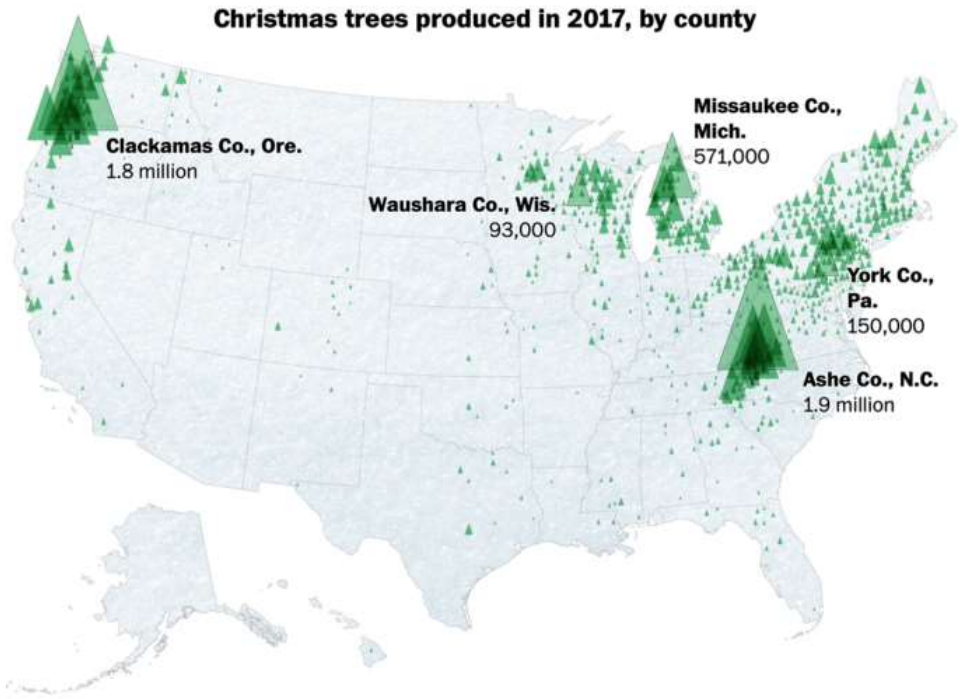
Most of these palettes are available to both **ggplot2** and **matplotlib**. For R, you may have to load packages like **RColorBrewer** or **viridis**

Consider color blindness



Consider printing

Color can provide context



Where Christmas trees come from

In summary, work with the following attributes to encode your data

- Position
- Length
- Angle
- Slope
- Area
- Volume
- Color
- Density *
- Shape *

Or any combination thereof!

* not discussed today

Visualization critiques: what is wrong with this picture?

- What is the first thing you notice about this visualization?
- What point is the visualization trying to make?
- Who is the intended audience?
- What is the visualization doing well?
- What problems do you see with the visualization design?

Bad example 1

Bad example 2

Bad example 3

Bad example 4

Bad example 5

Some cool visualization links

Visual Capitalist

Scientific American: The Pulsar Chart That Became a Pop Icon Turns 50: Joy Division’s Unknown Pleasures

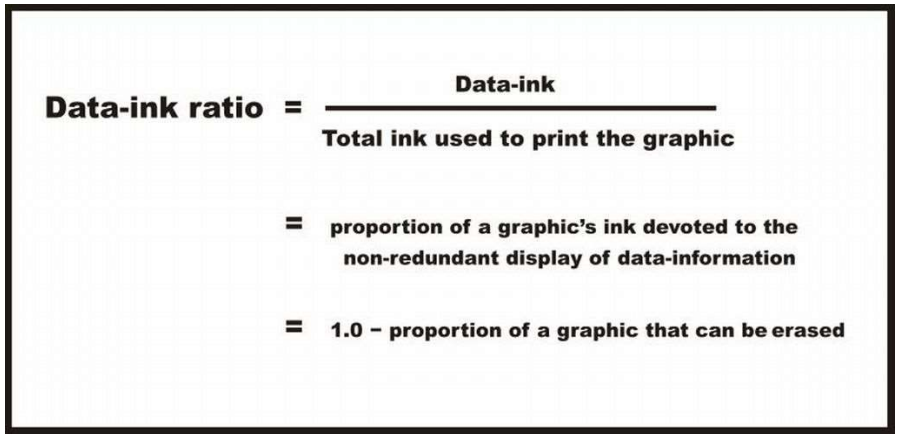
Stitch Fix: What Color is This?

Reddit (yes, Reddit) r/dataisbeautiful

What makes a readable graphic?

- It depends on who you ask
- Many go by the *data-ink ratio* as described by Tufte:

> A large share of ink on a graphic should present data-information, the ink changing as the data change. Data-ink is the non-erasable core of a graphic, the non-redundant ink arranged in response to variation in the numbers represented.



- It depends

Data is fluid and visualization represents that fluidity

- Real world is complicated
- There are visualization rules that cannot be broken related to the technical aspects of how a chart is constructed
- However, there are principles and guidelines (fuzzier aspects of chart design) that you need to adapt to the data and the context:
  - The baseline *always* needs to start at zero. *But what if the data has no zeros?*
  - Pie charts are terrible, never use them. *But people know how to read pie charts and it's fine for this specific dataset.*
  - A bar chart would have been better. *Insert some snarky remark here.*

Tradeoffs

When visualizing for *an audience* there are always factors to consider that can conflict with visual efficiency

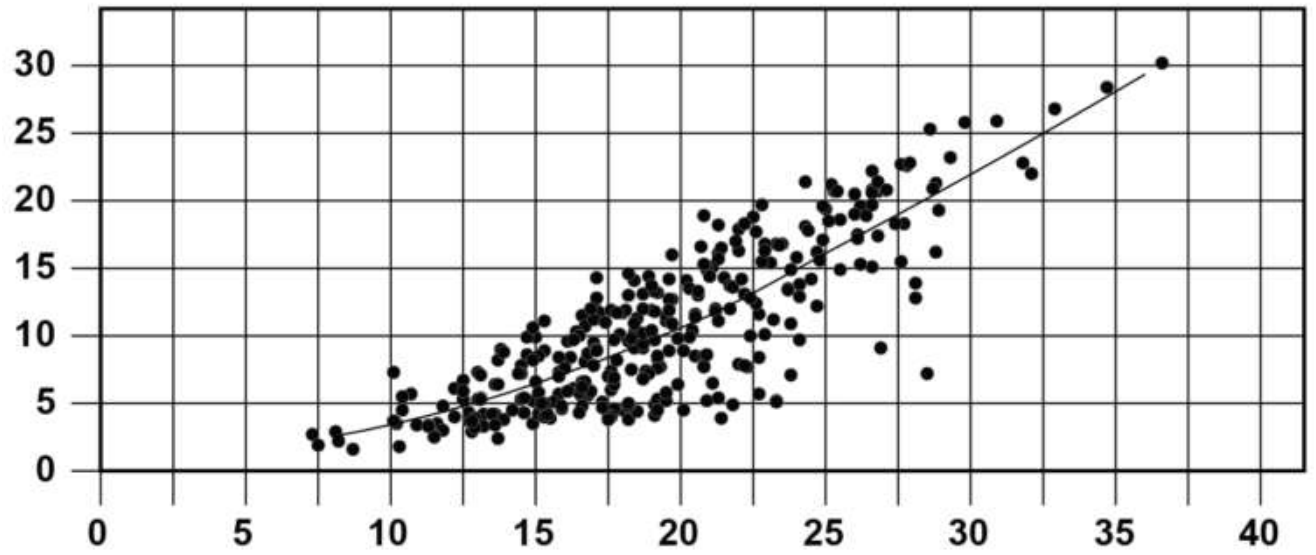
A *readable chart*:

- Provides **clarity** (removes confusion)
- Has a **clear purpose**
- Uses **visual encodings that make sense** for the **context** of the data
- Has a **clear direction for how to interpret**

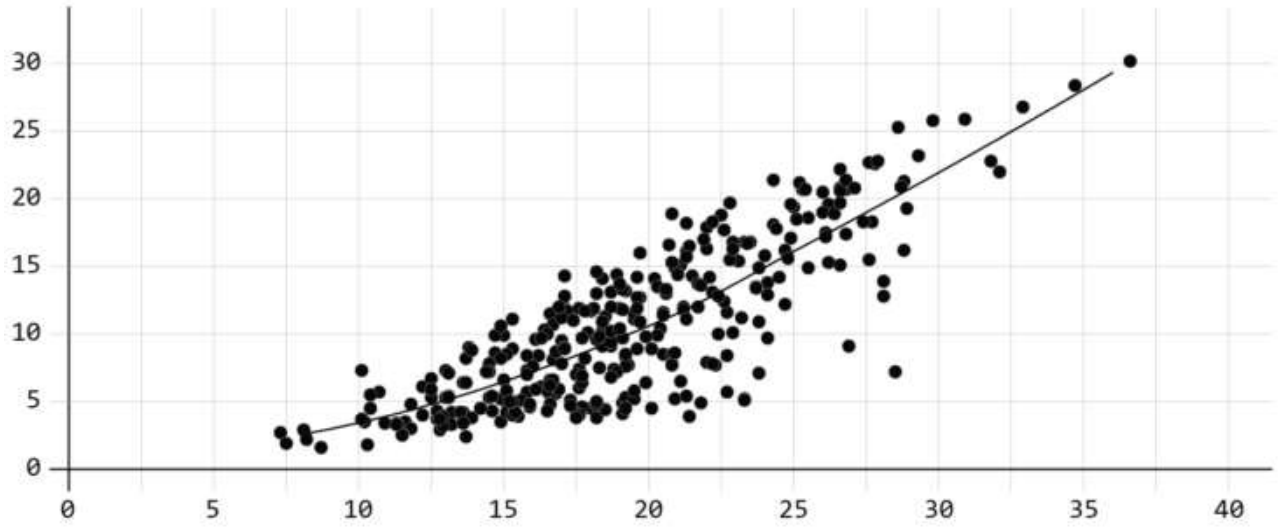
Visual Hierarchy

When you make a chart using *default settings*, you usually get a flat graphic where everything — from the tick marks, to the encoded data, to the title — gets the same amount of importance visually

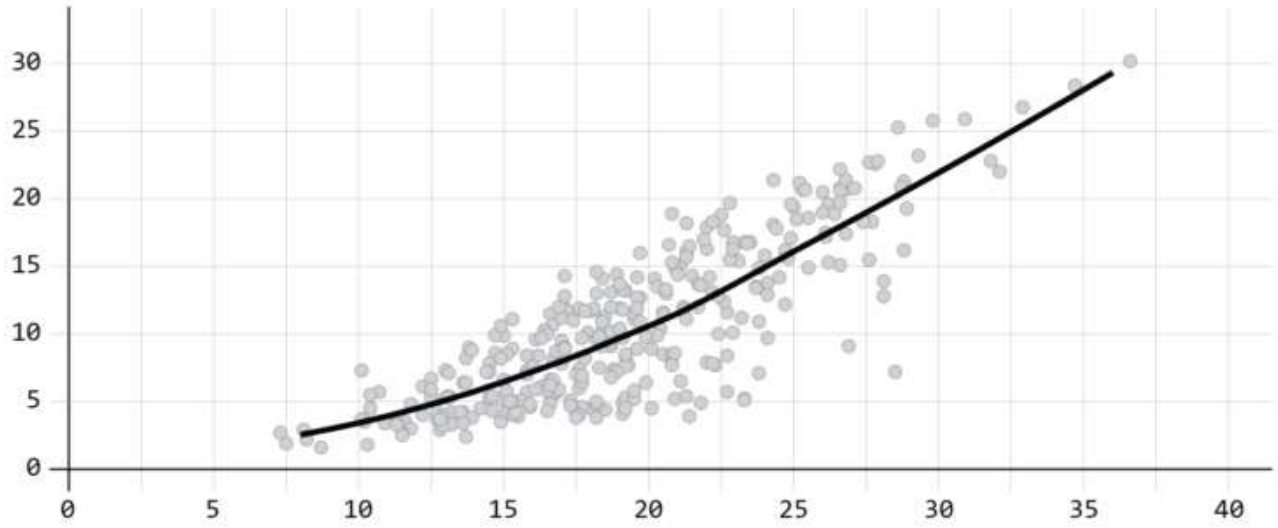
Lines, colors, border box, etc. are on the same level of importance as the data itself. Nothing stands out.



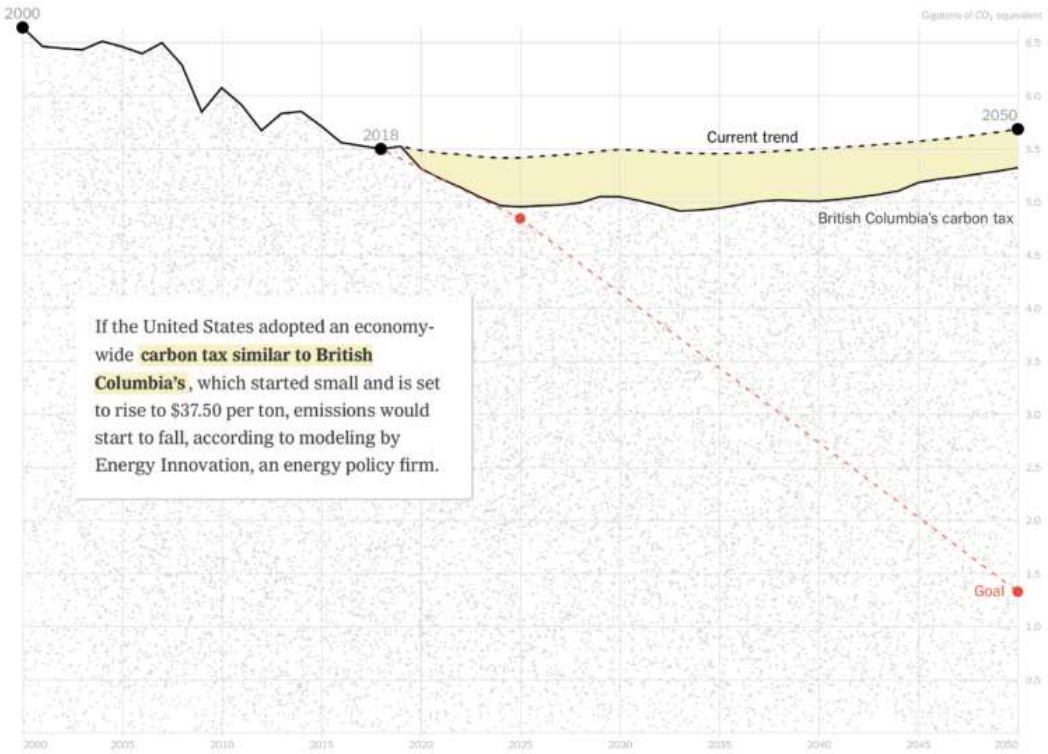
Small adjustments can help the data appear more prominently and the other parts move back to support.



It’s more obvious what part of the chart is the actual data and what part is background context



Example

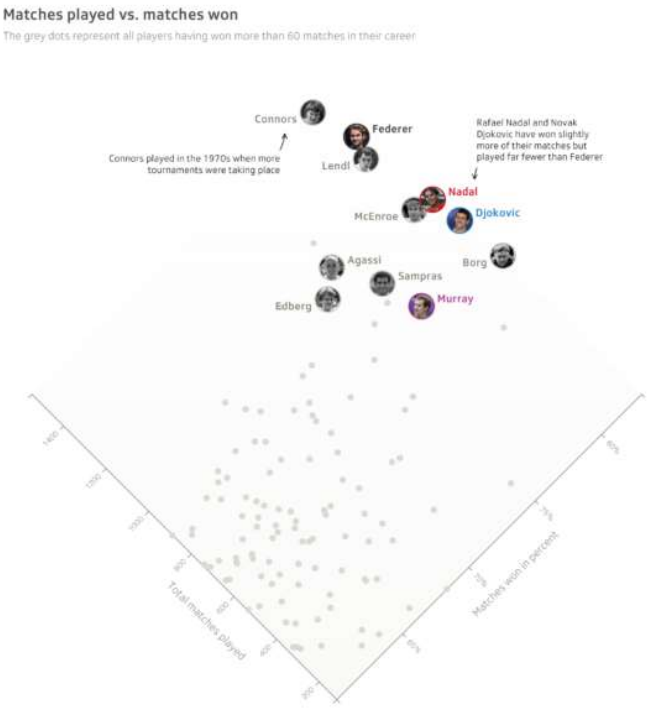


Interactive in NYT's site

Color contrast

- Color makes parts of chart stand out
- Brighter and bolder appear more prominent than greyed or faded colors
- To increase the visibility of your data, make it appear *higher* in the visual hierarchy

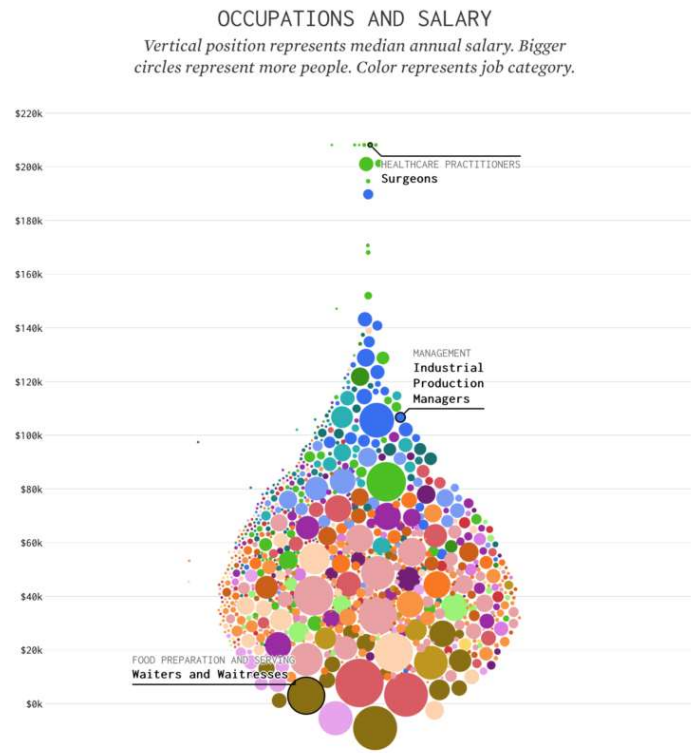
20 Years, 20 Titles - Roger Federer



Size

- Objects that use more space on the screen or paper will naturally draw more attention
- Vary the sizes in your chart to bring more attention to points of interest
- One obvious case is the size of text

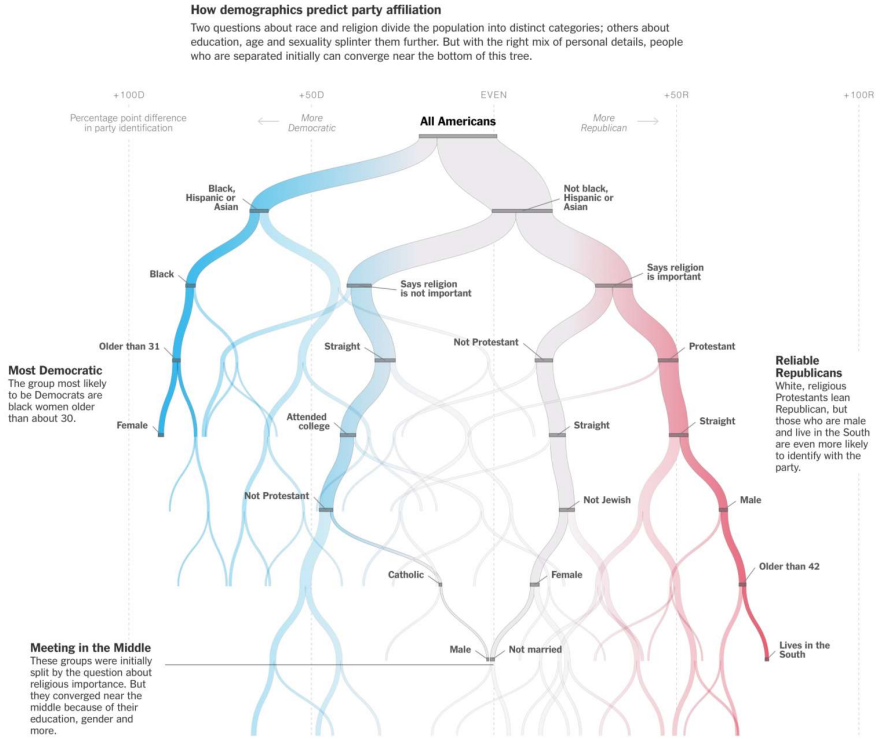
Salary and Occupation



Placement

- Where you put your data — top, bottom, left, right — also affects visual hierarchy
- Things placed at the top of a chart appear more important than things placed at the bottom.
- For example, in government and politics, *left* and *right* might be linked to certain ideologies

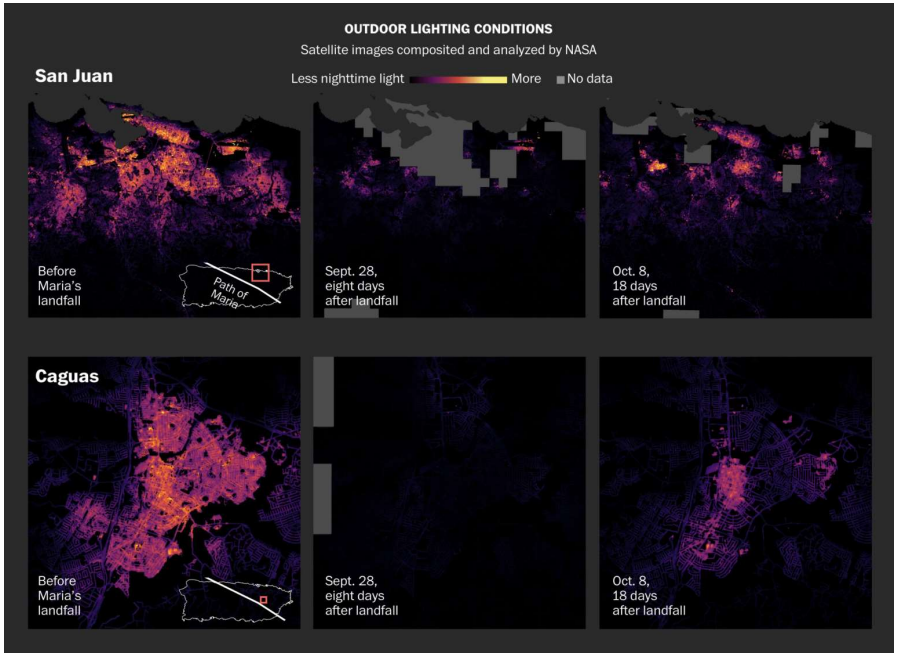
Are you a Democrat or a Republican?



Highlighting

- Use highlighting to call out specific areas of a visualization to direct readers' eyes to what is important

Hurricane Maria

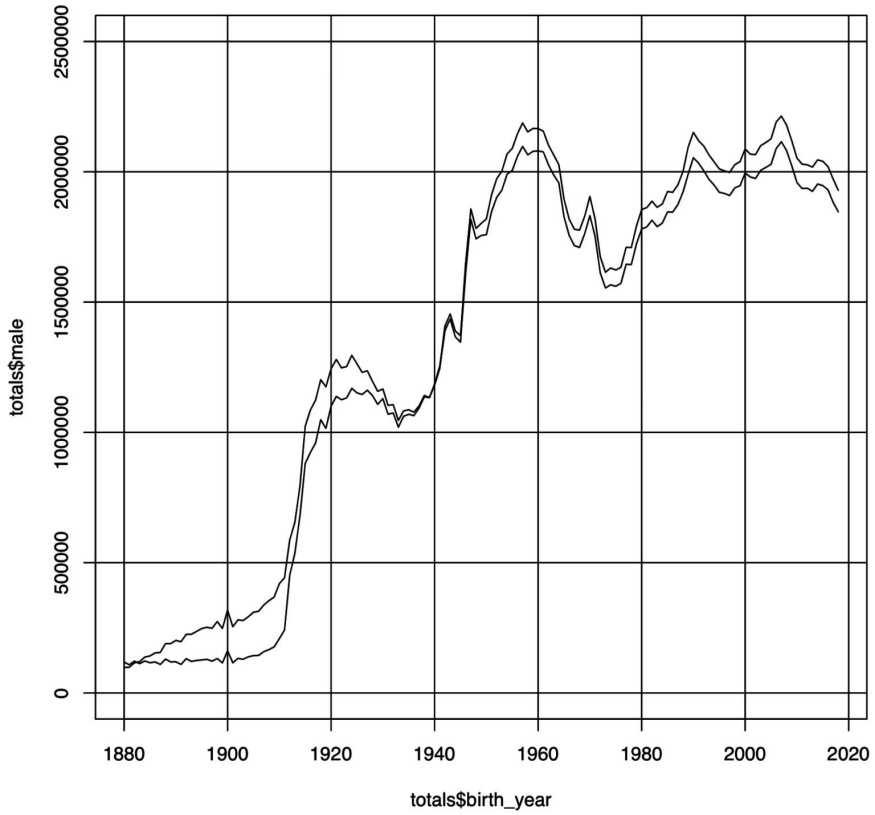


Layering

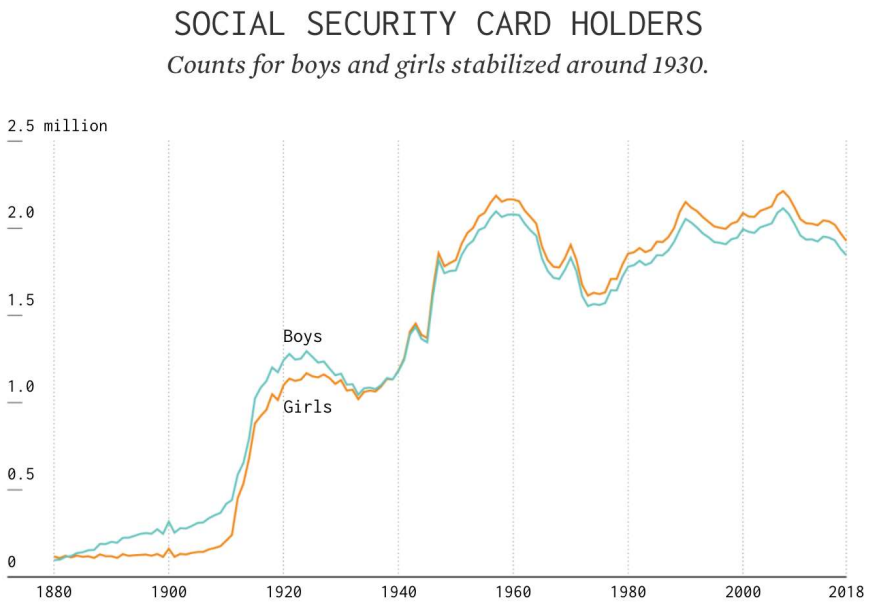
- Think of the visual hierarchy as layers
  - The most important items gets placed on the top of the stack
  - Items that are less important, or rather, more boilerplate, can fall to the back
- The layering metaphor is especially helpful when you implement or design your visualization.
- For example, Adobe Illustrator or Inkscape already uses layers, so you can stack things on top of each other based on your goals
- If you're using code, the code for a bottom layer tends to run before the top layers.

From the reader perspective, it's more obvious where to focus attention. They can spend less time trying to interpret the chart and the data and more time understanding your own interpretations of the data.

Give this flat chart some hierarchy



Much better!



> The Golden Ratio, also known as the *divine ratio* is a fundamental ratio in nature, and creates an aesthetically pleasing balance between dimensions. The ratio is $\frac{1+\sqrt{5}}{2} \approx 1.618$. Often we'll see rectangles where the longer side and shorter side are in this ratio.

Providing Context

Tips for Providing Context

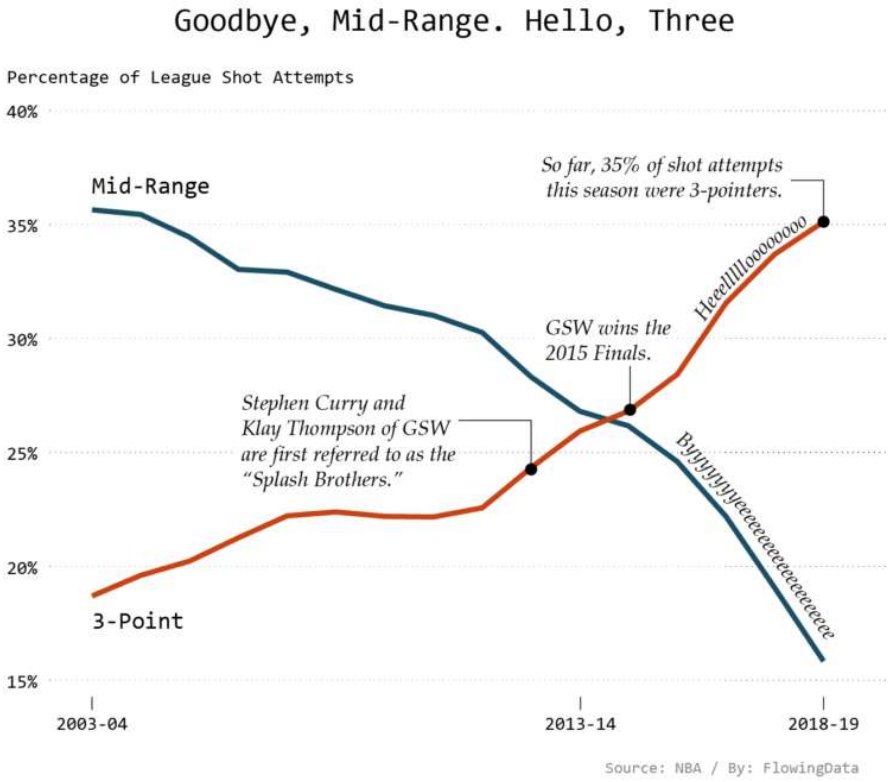
- Annotation
- Tone
- Direct Labeling
- Font Selection
- Point of Reference

Annotation

- Annotation is the quickest and most straightforward way to add context to your charts. However, under the false security of “letting the data speak”, oftentimes these words are missing from default charts.
- Add the extra layer of information, and you draw attention to specific areas and points, help explain visual encodings, and describe what a reader is seeing.
- Words can set expectations, so that readers know what they’re about to see. Here’s Hidy Kong on her group’s research on visualization titles:
- Visualization titles influence how people interpret, perceive bias in, and trust data visualizations.
- Sometimes it doesn’t even matter that a title contradicted the chart. The title could say that something increased over time when the chart showed a clear decrease, and **the reader would take away the context of the title over the chart.**

Tone

- The words you use describe your data can change the tone of your charts, which can change how people interpret them
- Using casual language could signal to readers that your chart presents a less serious topic
- Using more technical language might seem like it was meant for a technical audience
- Choose your words wisely



Goodbye, midrange shot

Direct Labeling

- Most visualization software lets you add legends to your charts to describe what each visual encoding represents
- The challenge for readers is that they have to refer to the legend and **look away** from the actual chart
- Try to directly label visual encodings

However…

- Most statistical software (R, Python, MATLAB, etc) cannot do words on graphics and typography
  - This situation is improving with newer packages
- These programs lack control of typography – in contrast to maps, where words can be placed directly on the labelled map element.
- Additional post-processing is usually done with Illustrator or some other tool

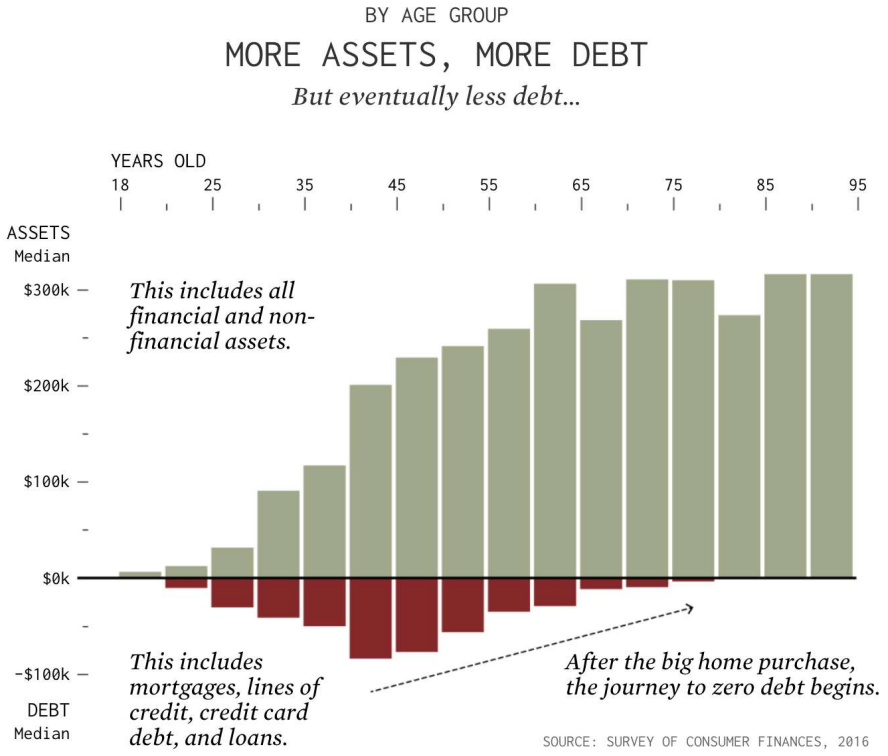
Font Selection

There are primarily two classes of annotations:

1. labels that help readers decode the visualization (axis, tick, and category labels)
2. annotations that explain the data, which is usually required to provide contextfor a specific dataset

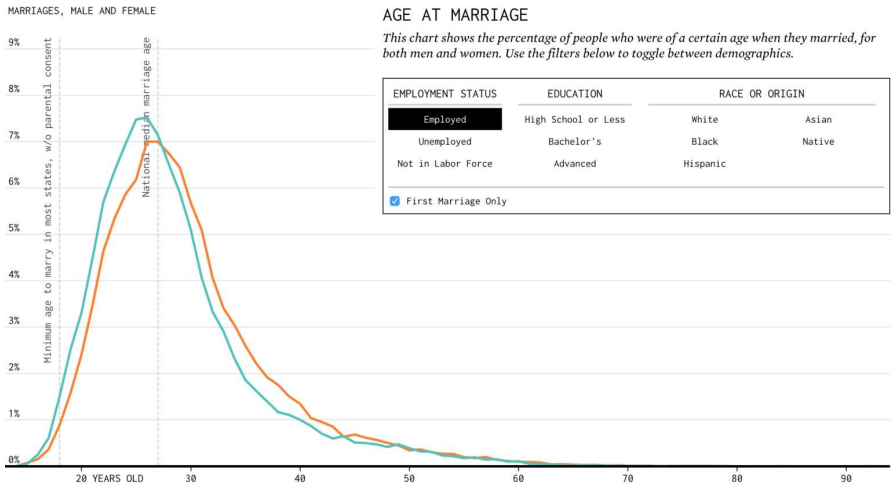
Nathan Yau uses `monospace fonts` general labels and an *italicized serif font* for contextual annotation

Reaching $100k in savings



Point of Reference

- Visualization is all about comparison
- If it is difficult to compare visual encodings, then it is difficult to interpret a chart, much less get anything useful out of it
- Providing a point of reference is a straightforward remedy
- With time series data, it can be useful to use a specific time as a point of reference



Marrying Age

Aesthetics

Aesthetics are subjective and can provide more *clarity*

Put effort into aesthetics, and it can help readers understand your charts better and also differentiate your own style

Aesthetics can provide the following:

- Beauty
- Readability
- Identity
- Expectations

Elements of aesthetics:

- Organization and arrangement
- Sizes and weights
- Color palette
- Medium

Tufte's Principles

The Visual Display of Quantitative Information

This book discusses **statistical graphics**, **charts and tables**, as well as the theory behind the **design of information graphics or data graphics**.

The book goes into a **detailed analysis of successful ways in which to display complex,** statistical information with quick, easy and **effective design techniques.**

The first edition was published in 1983.

Envisioning Information

This book tackles the problem of **conveying multiple variable information on a 2-d space.**

It teaches us ways in which we can **communicate more information per unit to make good**, clear and smart presentations.

This book won 17 awards, and was published in 1990.

Tufte's Principles of Graphical Integrity

1. Show data variation, not design variation
    - Don't get fancy, let the data speak
2. Do not use graphics to quote data out of context
    - Maintain accuracy
3. Use clear, detailed, thorough labelling.
    - Use annotations to make your point
4. Representation of numbers should be directly proportional to numerical quantities
    - This is essential for fair representation
5. Don't use more dimensions than the data require
    - Be appropriate in use of 3D graphics, for example

Tufte's Fundamental Principles of Design

1. Show comparisons
2. Show causality
3. Use multivariate data
4. Completely integrate modes (like text, images, numbers)
5. Establish credibility
6. Focus on content

Sparklines

Invention of the **sparkline**, most commonly used in stock activity.

> A sparkline is a small intense, simple, word-sized graphic with typographic resolution. Sparklines mean that graphics are no longer cartoonish special occasions with captions and boxes, but rather sparkline graphic can be everywhere a word or number can be: embedded in a sentence, table, headline, map, spreadsheet, graphic.

Tufte, May 27, 2004

Small Multiples

The **small multiples** method is one that Tufte uses often to portray multiple graphs of information.

> At the heart of quantitative reasoning is a single question: Compared to what? Small multiple designs, multivariate and data bountiful, answer directly by visually enforcing comparisons of changes, of the differences among objects, of the scope of alternatives. For a wide range of problems in data presentation, small multiples are the best design solution.

Tufte, Envision Information, page 67

Graphical Integrity

- The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented
- Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graphic itself. Label important events in the data.
- Graphics must not quote data out of context

Data-Ink

]

A large share of ink on a graphic should present data-information, the ink changing as the data change.
Data-ink is the non-erasable core of a graphic, the non-redundant ink arranged in response to variation in the numbers represented.

How to maximize the data-ink ratio, within reason:

1. Erase non-data-ink, within reason
2. Erase redundant data-ink
3. Revise and edit

Chartjunk

- Forgo chartjunk, including moiré vibration, the grid and the duck
- The interior decoration of graphics generates a lot of ink that does not tell the viewer anything new.
- The purpose of decoration varies — to make the graphic appear more scientific and precise, to enliven the display, to give the designer an opportunity to exercise artistic skills.
- **All non-data- ink or redundant data-ink is often chartjunk.**

Multifunctioning Graphical Elements

- **Mobilize every graphical element**, perhaps several times over, to show the data.
- The graphical element that actually locates or plots the data is the data measure.
- The complexity of multifunctioning elements can sometimes **turn data graphics into visual puzzles**, crypto- graphical mysteries **for the viewer to decode**.

Escaping Flatland

- Introduce **multiple dimensions on a two-space surface**
- Focus more on the point than on the presentation, good design strategies are transparent.
- Find **pattern**
- Words may not be the most appealing to everyone but symbols are universal and understood by all
- **More small images in sequence** allow more comparison with your eyes and a better understanding

Layering and Separation

Would Tufte approve of this diagram?

Layering and Separation

No! He would not. To make the visual depictions more effective, reduce them down, by using:

- Macro annotation, which can help explain micro detail
- Use light, color and space effectively
- Remove the weight, avoid vibration

Bringing it Together

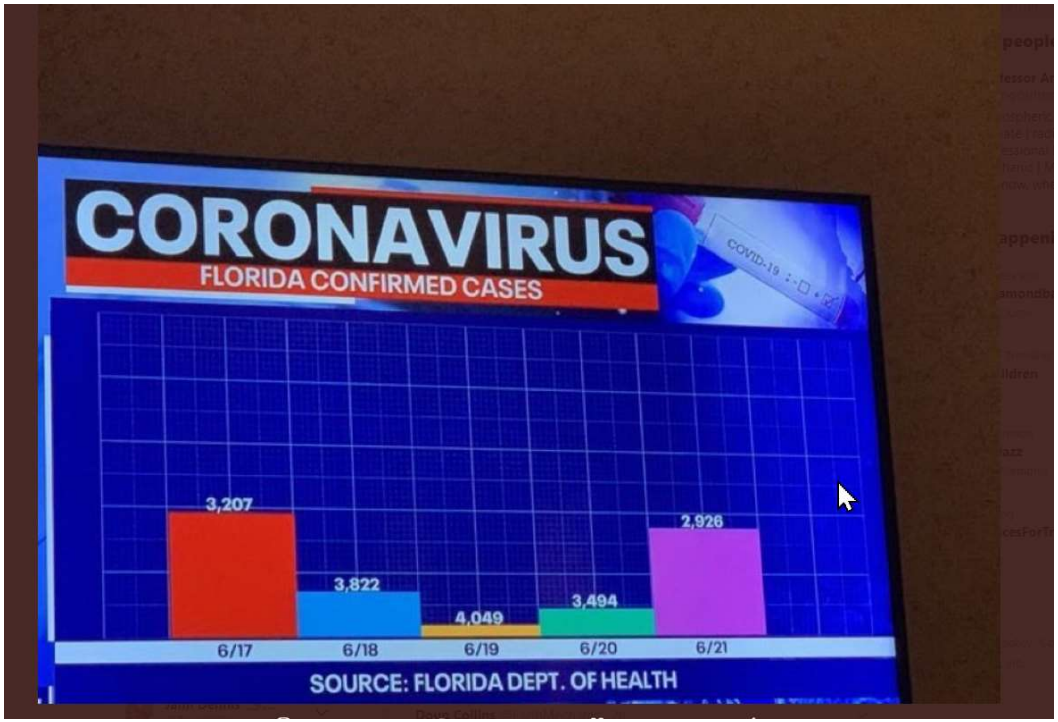
So far you've learned about

- designing for an audience
- picking the right visualization
- making readable graphics

Now what?

Practice, practice, practice

You make awesome charts



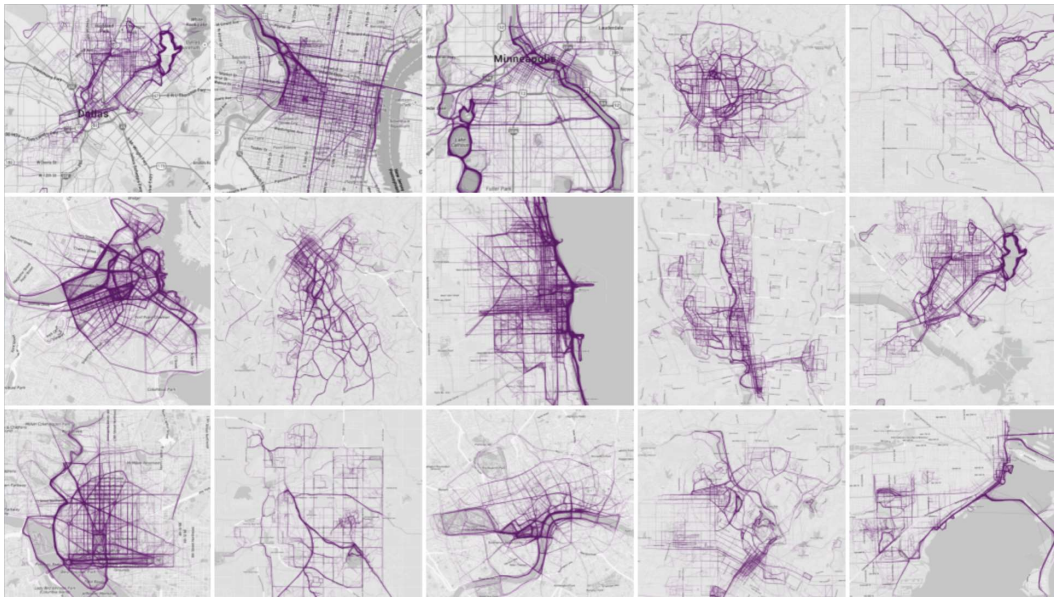
You are a Dataviz G.O.A.T. (greatest of all time)

nah!

You work with *MORE* data!

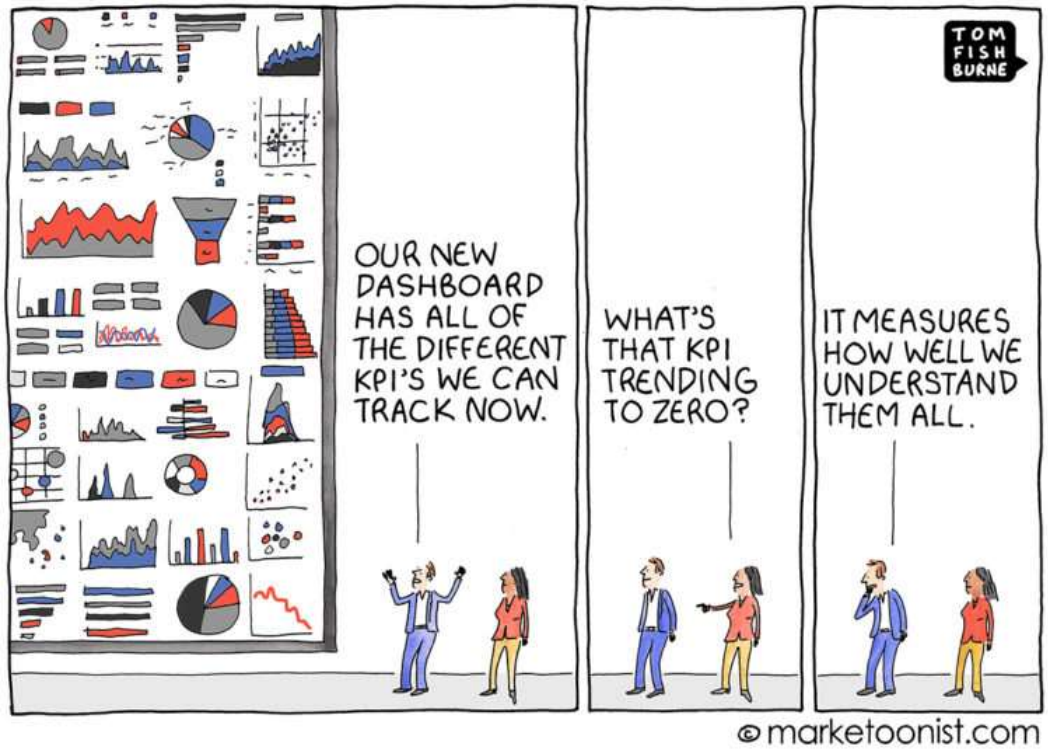
- more *complexity*
- larger files
- missing values
- incorrect encodings

Show something instead of showing everything



Where people run

Too much leads to overload



KPI Overload

Start asking questions

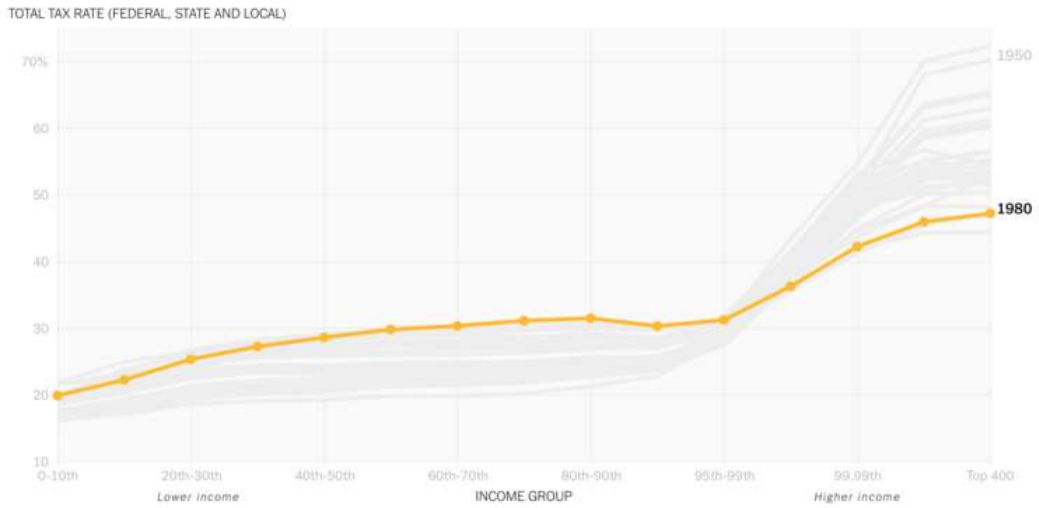
ASK -> ANSWER -> ASK NEW -> ANSWER AGAIN -> REPEAT

- What does the data look like?
- Does anything stand out?
- What is the mean and median?

Start simple and work your way up to more complex questions

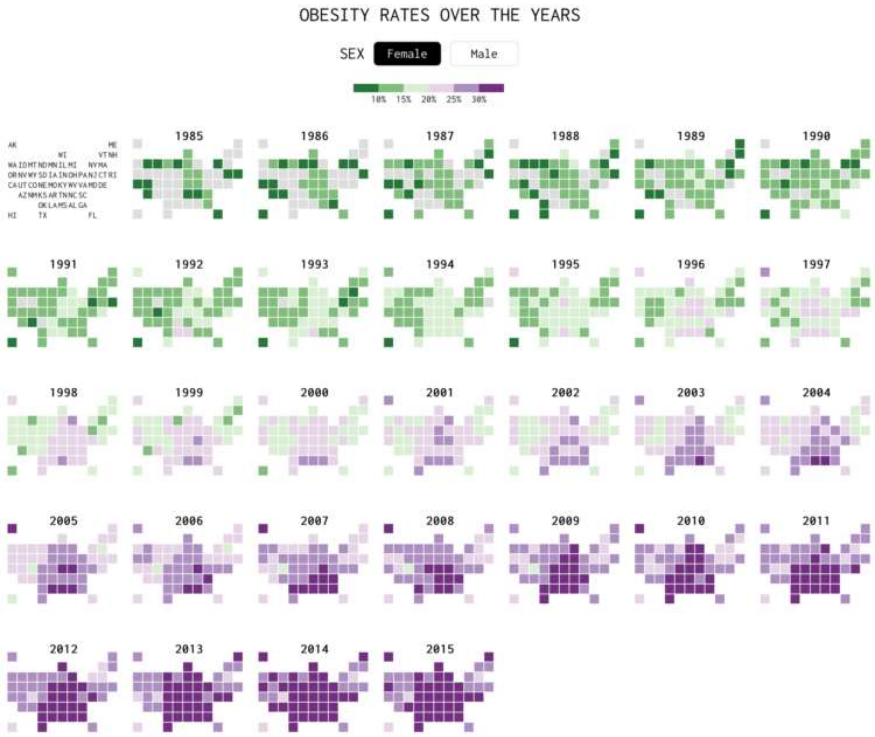
Break it down

Highlighting specific parts of a dataset



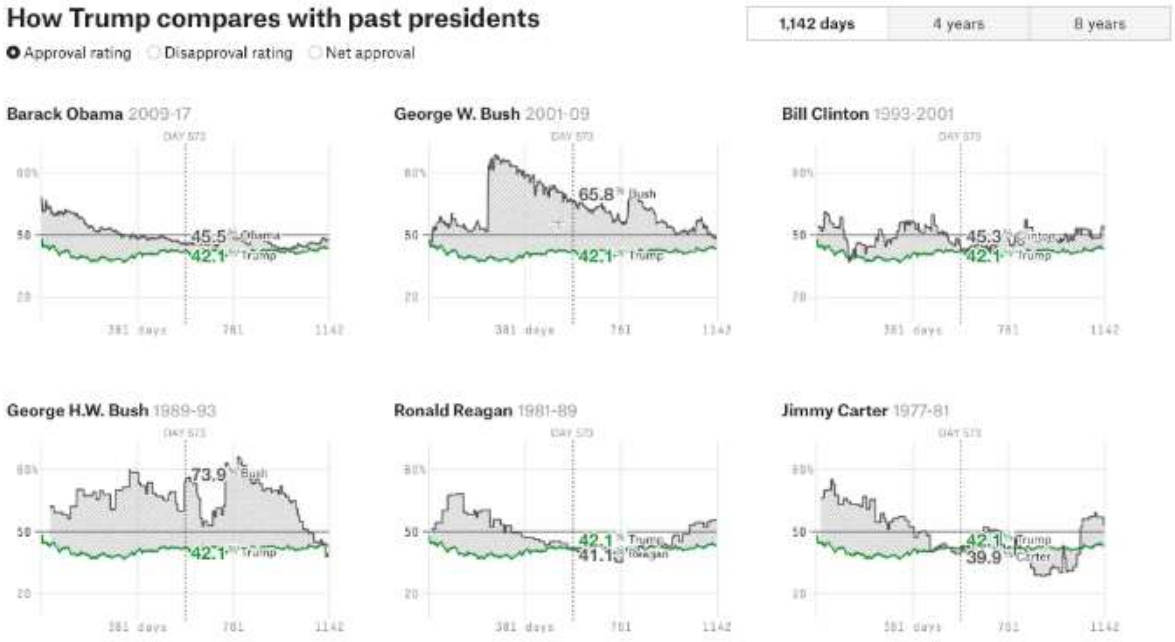
The Rich really pay lower taxes than you

Multiple charts



Mapping the spread of obesity

Linked views (dynamic)



How unpopular is Donald Trump

Build it up

Follow the data

The Follower Factory