

COMP 7990

*Principles and Practices of
Data Analytics*

Data Visualization

Acknowledgements:

Lan Liang, School of Communication, HKBU

Katy Börner, School of Informatics and Computing, Indiana University, USA

Robert Putnam, Research Computing, IS&T, Boston University, USA

Slide Prepared by Dr. Yucheng Jin

Warm up...

- You are given the following data, think about how to present them:

China's global navigation satellite system

China recently joined an elite group by launching its own global navigation system, BeiDou, named after the Big Dipper asterism. But how does it stack up against the other major systems?

How do satellites orbit Earth?

There are essentially three types of Earth orbits: low Earth orbit (LEO), medium Earth orbit (MEO) and high Earth orbit (HEO), which includes the geosynchronous equatorial orbit (GEO).

Orbit height of China's BeiDou Navigation Satellite System (BDS) 21528km, America's Global Positioning System (GPS): 20200km, Russia's Global Navigation Satellite System (Glonass): 19100km, and the European Union's Galileo 23222km.

Three-step development strategy

The China National Space Administration reports BDS was developed in three stages.

BDS-1

BeiDou launched **three** geosynchronous equatorial orbit (GEO) satellites from **2000** to **2003** to provide China – and surrounding regions – with its own navigation system. China became the third country to build its own satellite navigation system, after the United States and Russia.

BDS-2

Over the next eight years, between **2004** and **2012**, China created a mixed constellation of **14** satellites which was able to cover the Asia-Pacific region. Five were GEO satellites, five inclined geosynchronous orbit (IGSO) satellites and four MEO satellites.

BDS-3

On June 23, **2020**, a Long March-3B rocket blasted off from the Xichang satellite launch centre in southwest China's Sichuan Province carrying the satellite which would complete the BDS-3 system, China's own global navigation system. It consists of a total of **30** satellites, including **24** MEO satellites, **three** IGSO satellites and **three** GEO satellites.

Location accuracy depends on factors such as atmospheric conditions, signal blockage and receiver quality. BeiDou upgraded its performance after the completion of its third phase, reaching accuracy at 1 m for public use and 1 cm for encrypted military use. Similarly, the much-anticipated Galileo, which is expected to be settled this year, will be as accurate as BeiDou. GLONASS now achieves an accuracy of 2.8 m. To improve GLONASS' s ground segment, positioning ground stations are under construction in Russia, Antarctic, Brazil and Indonesia. Those infrastructures are expected to lower GLONASS' s accuracy down to 0.6m or better by 2020. GPS-enabled mobile phones are normally accurate at 4.9 m under the open sky. Using dual-frequency and/or augmentation, DGPS coupled with CPGPS eliminates error source, realizing absolute accuracy at 20–30 cm.

<https://multimedia.scmp.com/infographics/news/china/article/3098087/beidou-satellite/index.html>



China's global navigation satellite system

China recently joined an elite group by launching its own global navigation system, BeiDou, named after the Big Dipper asterism. But how does it stack up against the other major systems?

Outline

- **Concepts of Data Visualization**
 - Value, goals
 - History
- **Design of Visualization**
 - Workflow of data visualization
 - Color, Size, Text, Titles, Labels
 - Choosing the right chart
- **Case study: COVID-19**
 - Mask wearing trends
 - People concerned about the impact of COVID-19

Outline

- **Concepts of Data Visualization**
 - Value, goals
 - History
- **Design of Visualization**
 - Workflow of data visualization
 - Color, Size, Text, Titles, Labels
 - Choosing the right chart
- **Case study: COVID-19**
 - Mask wearing trends
 - People concerned about the impact of COVID-19

Value of Information Visualization

- Exploring information collections becomes increasingly difficult as the volume grows
- With minimal effort, the human visual system can process a large amount of information in a parallel manner
- The occurrence of advanced graphical software and hardware enables the large-scale visualization and the direct manipulation of interfaces

Goals for Information Visualization

- Provide insight
 - Explain data to solve specific problems
 - Support the analytical task, showing the comparison or causality
 - Explore large data sets for better understanding
- Relieve the cognitive overload
 - Conveying abstract information in intuitive ways

In today's world, successful decision-making has everything to do with **turning data insights into action**. And because the goal of data visualization is **impact**, not numbers.

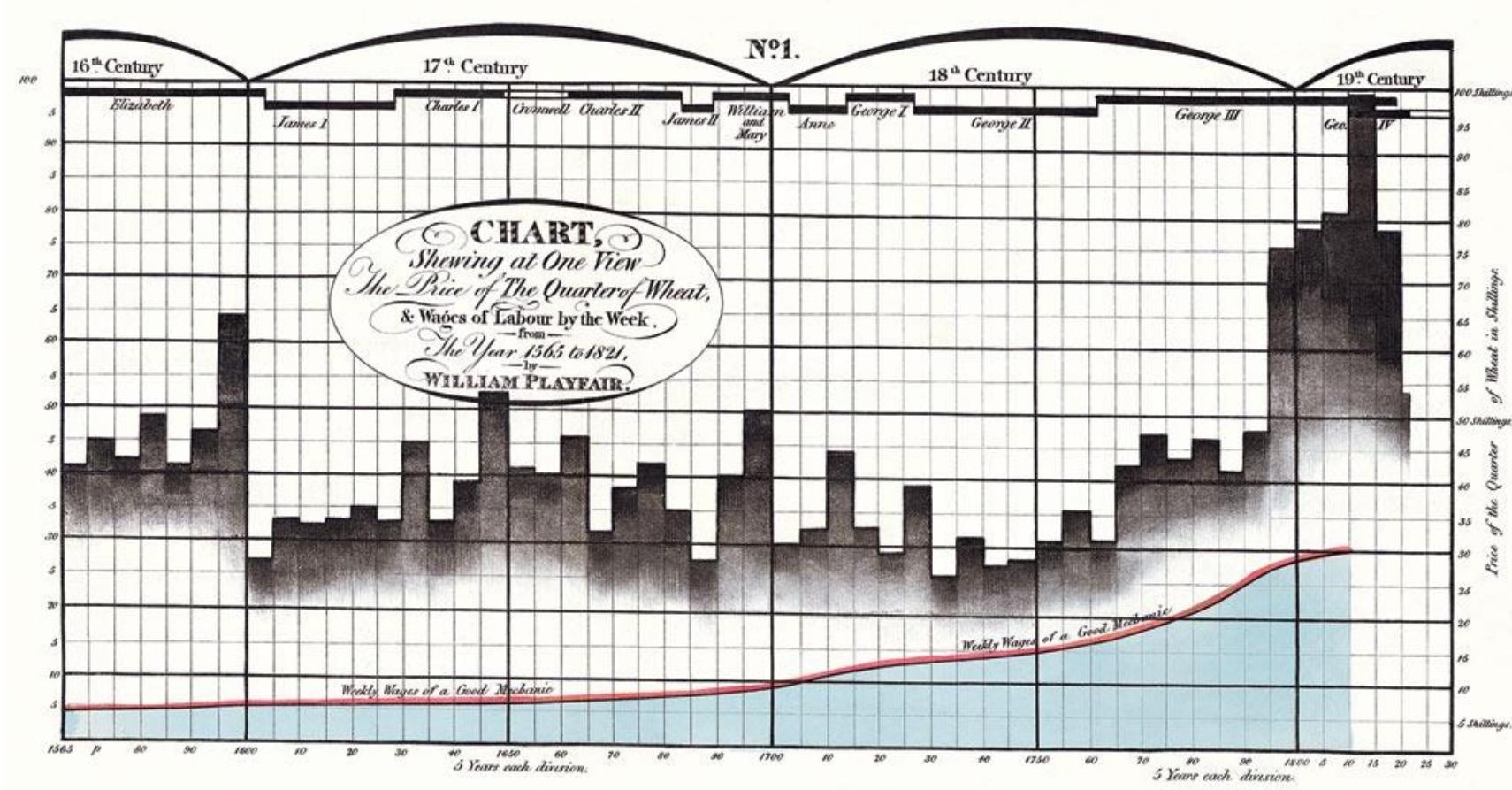
What is Visualization?

- Visualization is any technique for creating images, diagrams, or animations to communicate a message.
- Visualization through visual imagery has been an effective way to communicate both abstract and concrete ideas since the dawn of man.
 - Cave paintings
 - Hieroglyphs
 - Maps
 - ...

Brief History

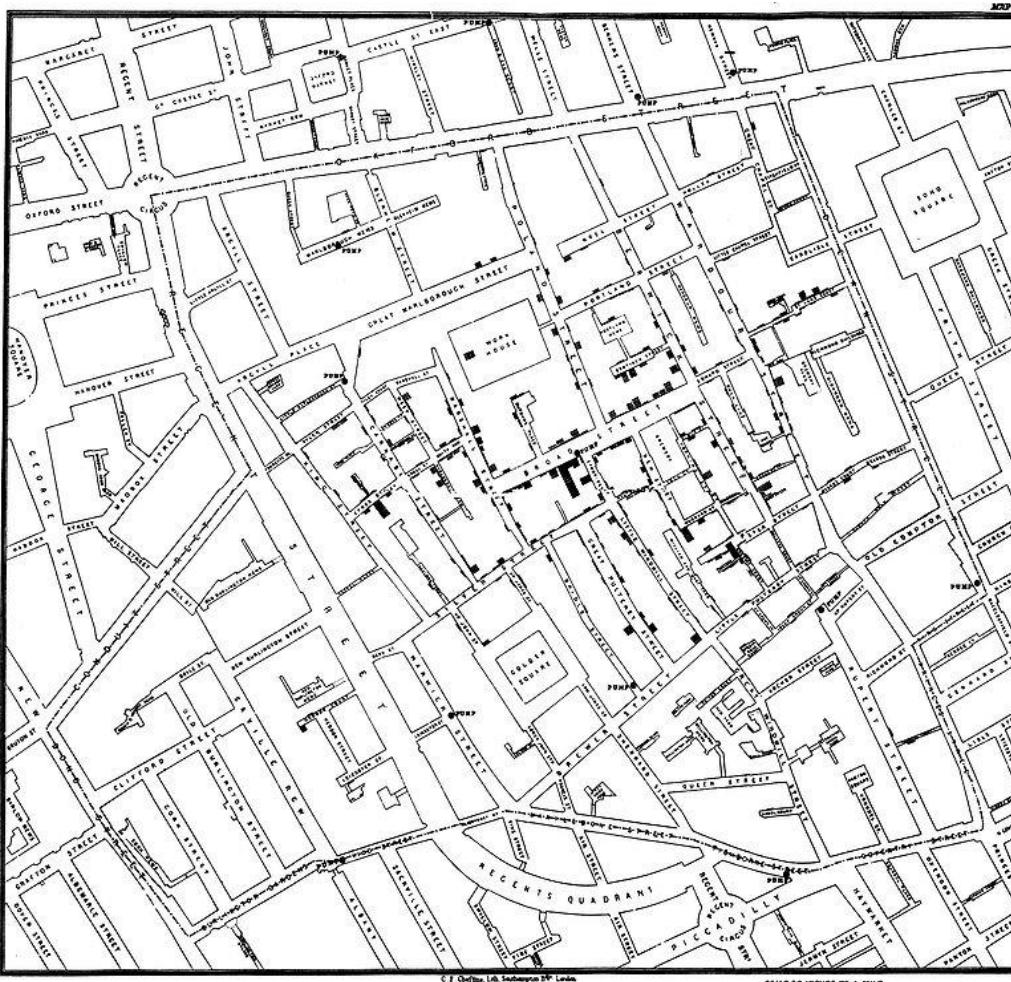
- First known map: 12th century (Tegarden, 1999)
- First presentation graphics: William Playfair (1786)
- Multidimensional representations appeared in 19th century (Tufte, 1983)
- Examples:
 - William Playfair (1821) – Chart showing at one view the price of the quarter of wheat, & wages of labor by the week, from the year 1565 to 1821
 - John Snow (1854) – Cholera Map in London
 - Charles Joseph Minard (1861) – Napoleon and The Russian Campaign of 1812

Brief History



- A graph by Playfair (1821) shows the price of wheat, weekly wages, and reigning monarch over a two hundred fifty year span from 1565 to 1820. Integration of bar charts and line graph

Brief History

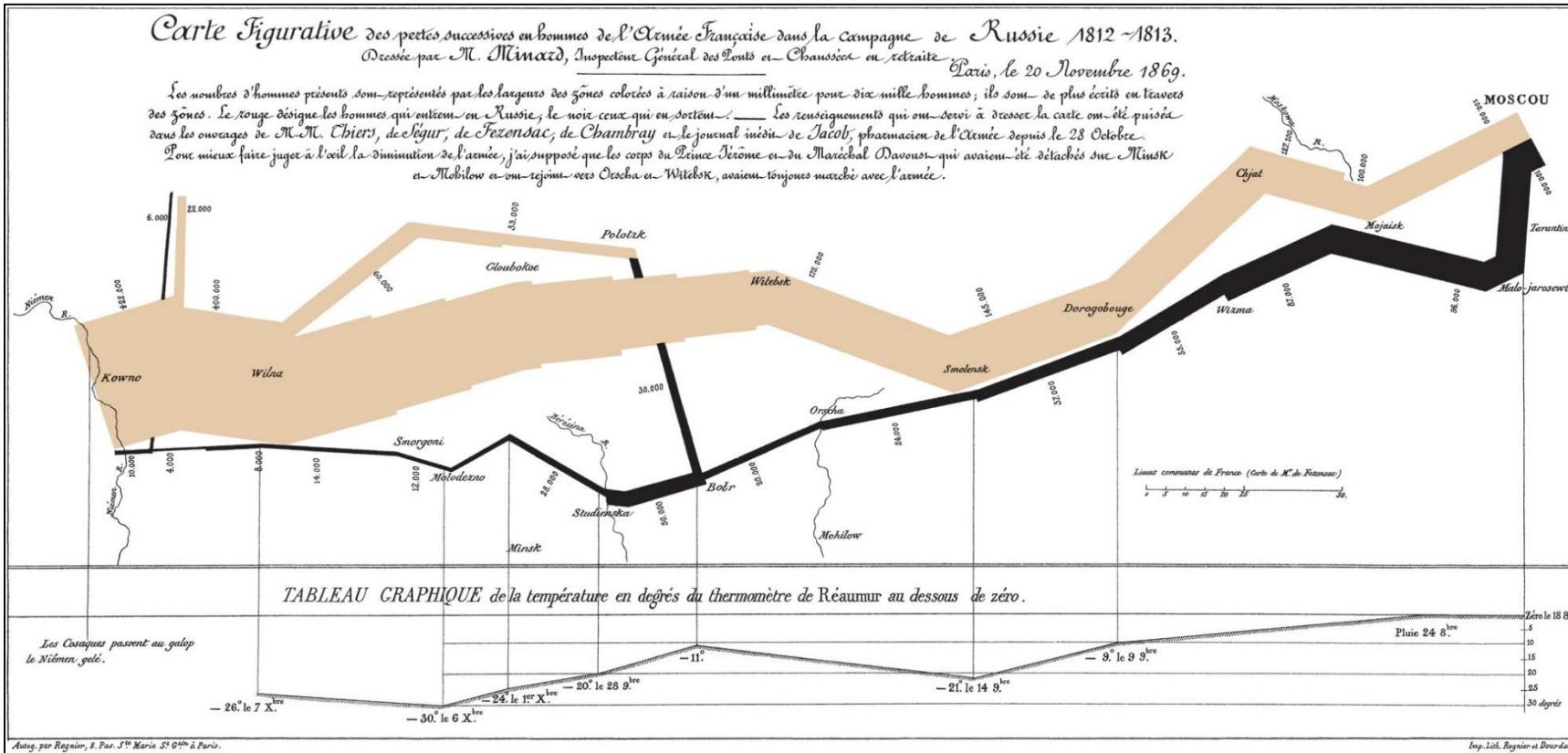


John Snow mapped the cholera cases on the map. The cases were clustered around the pump in Broad street.

- Each death case is a bar
- “Spot map” – Geo-spatial based mapping

John Snow (1854)

Brief History



In this visualization, Minard visualized:

- Napoleon's marching and retreat routes
- Army Size
- Temperature during retreat

Charles Joseph Minard (1861)

Outline

- **Concepts of Data Visualization**
 - Value, goals
 - History
- **Design of Visualization**
 - Workflow of data visualization
 - Color, Size, Text, Titles, Labels
 - Choosing the right chart
- **Case study: COVID-19**
 - Mask wearing trends
 - People concerned about the impact of COVID-19

A Visualization Workflow

A needs-driven workflow

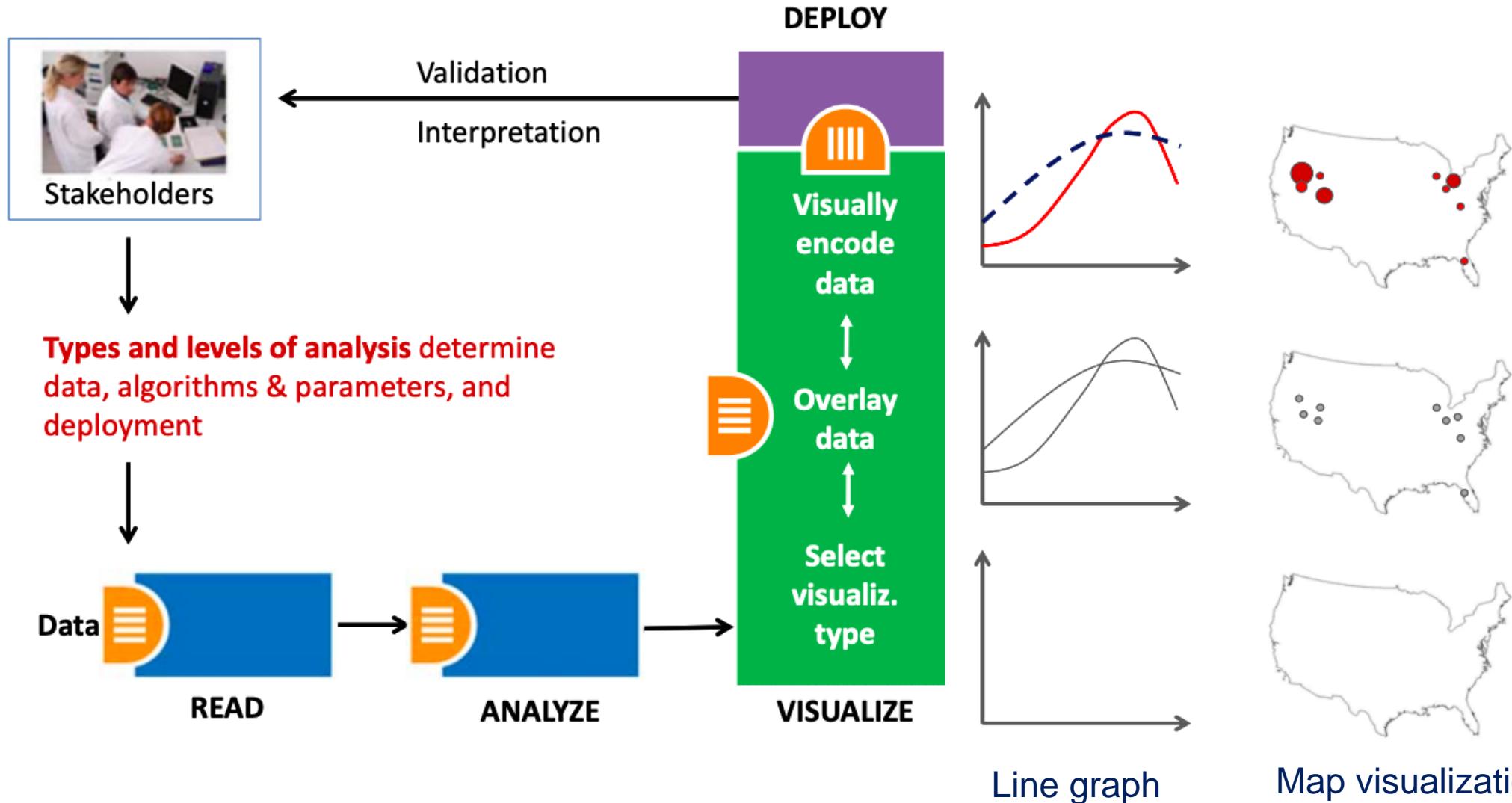
- How to select visualization types?
- How to visually encode data?
- How to visualize dynamically? – Animation and Interaction

Types and Levels of Analysis

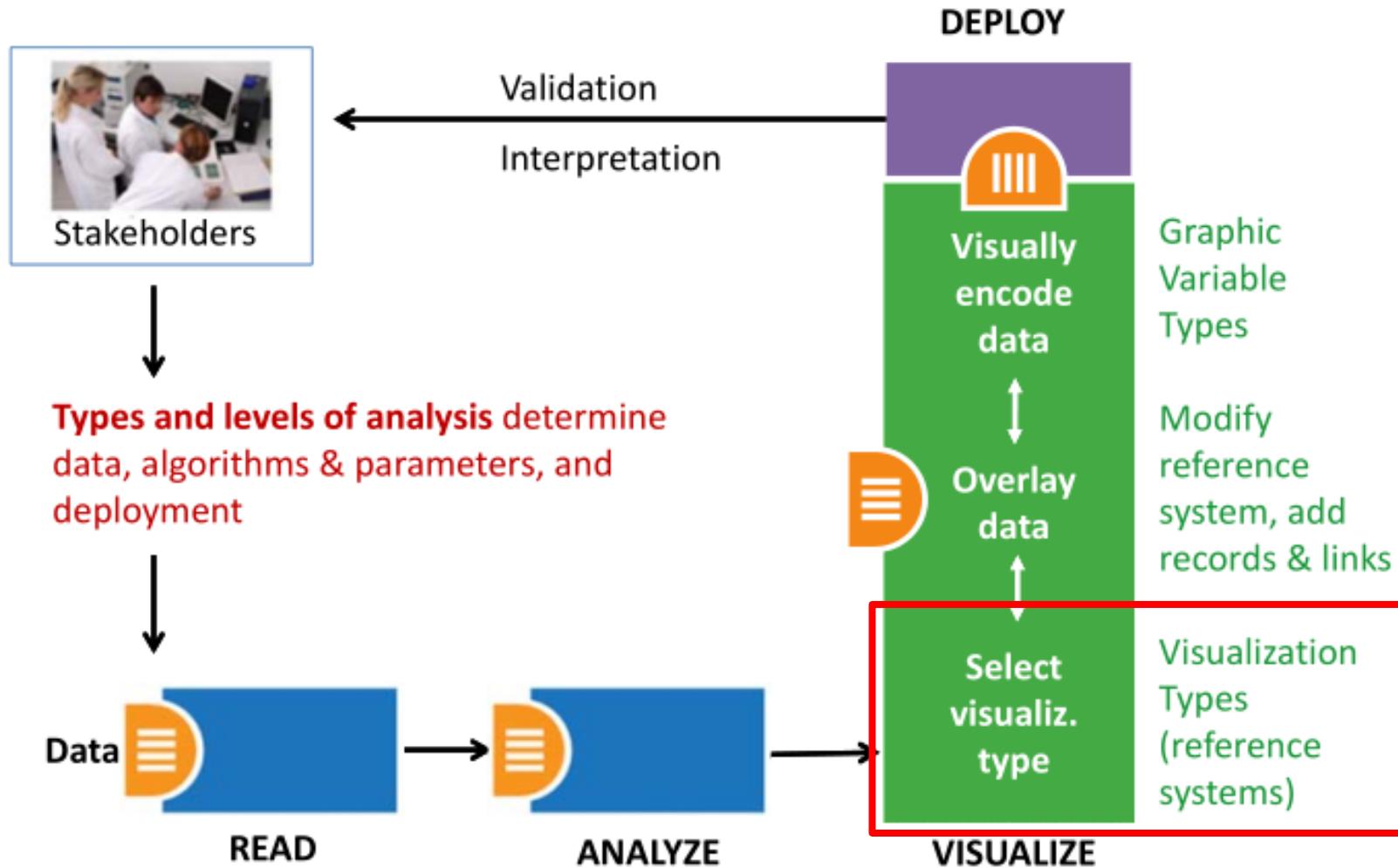
Types	Levels	<i>Micro/Individual (1-100 records)</i>	<i>Meso/Local (101–10,000 records)</i>	<i>Macro/Global (10,000 < records)</i>
Statistical Analysis/Profiling		Individual person and their expertise profiles	Larger labs, centers, universities, research domains or states	All of NSF grants, all of science
Temporal Analysis (When)		Funding portfolio of one individual	Topic bursts of PNAS	113 years of physical research
Geospatial Analysis (Where)		Career trajectory of one individual	Mapping a scientist's intellectual landscape	PNAS publications
Topical Analysis (What)		Chemistry research	Knowledge flows	World map of scientific activity
Network Analysis (With Whom?)		NSF grants	Collaboration network	NIH's global reach

Needs-Driven Workflow

Börner (2014) – *Visual Insights*



Needs-Driven Workflow – Select Visualization Type



Visualization Types

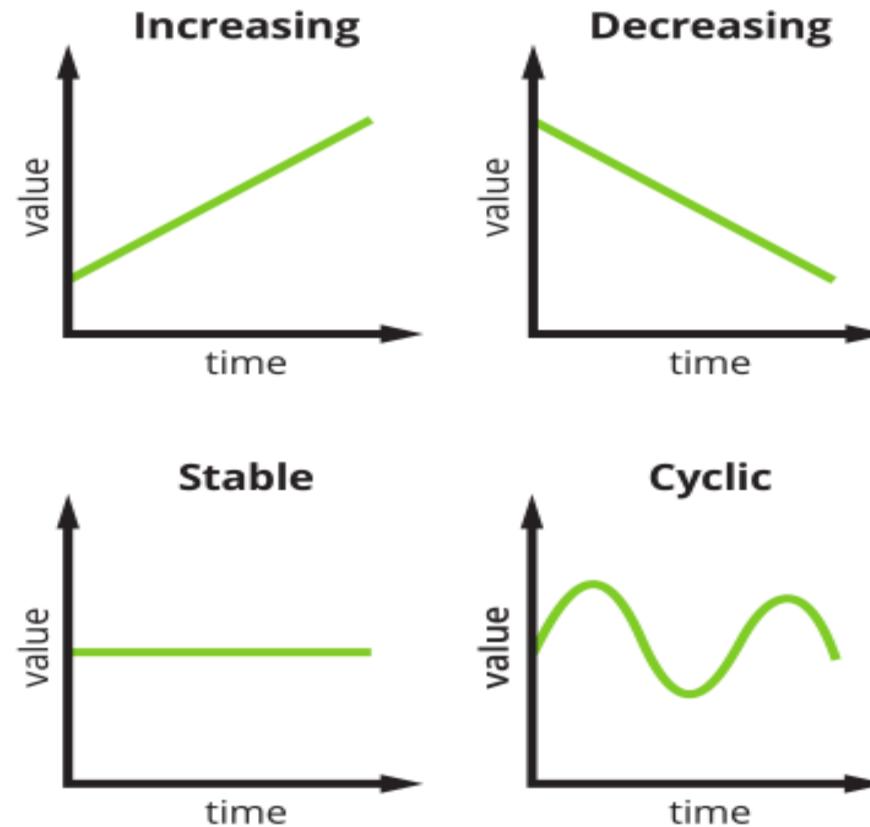
- Börner (2014) emphasized five major types:
 - Charts
 - No reference system—e.g., Wordle.com, pie charts
 - Tables
 - Categorical axes that can be selected, reordered; cells can be color coded and might contain proportional symbols. Special kind of graph.
 - Graphs
 - Quantitative or qualitative (categorical) axes. Timelines, bar graphs, scatter plots.
 - Geospatial maps
 - Use latitude and longitude reference system. World or city maps.
 - Network graphs
 - Node position might depends on node attributes or node similarity.
 - Tree graphs: hierarchies, taxonomies, genealogies.
 - Networks: social networks, migration flows.

Visualization Type Selection: Temporal

- Temporal data analysis and visualization answer “WHEN” question and can help to
 - Understand the temporal distribution of datasets
 - Identify growth rates, latency to peak times, or decay rates
 - See patterns in time-series data, such as trends, seasonality, or bursts.

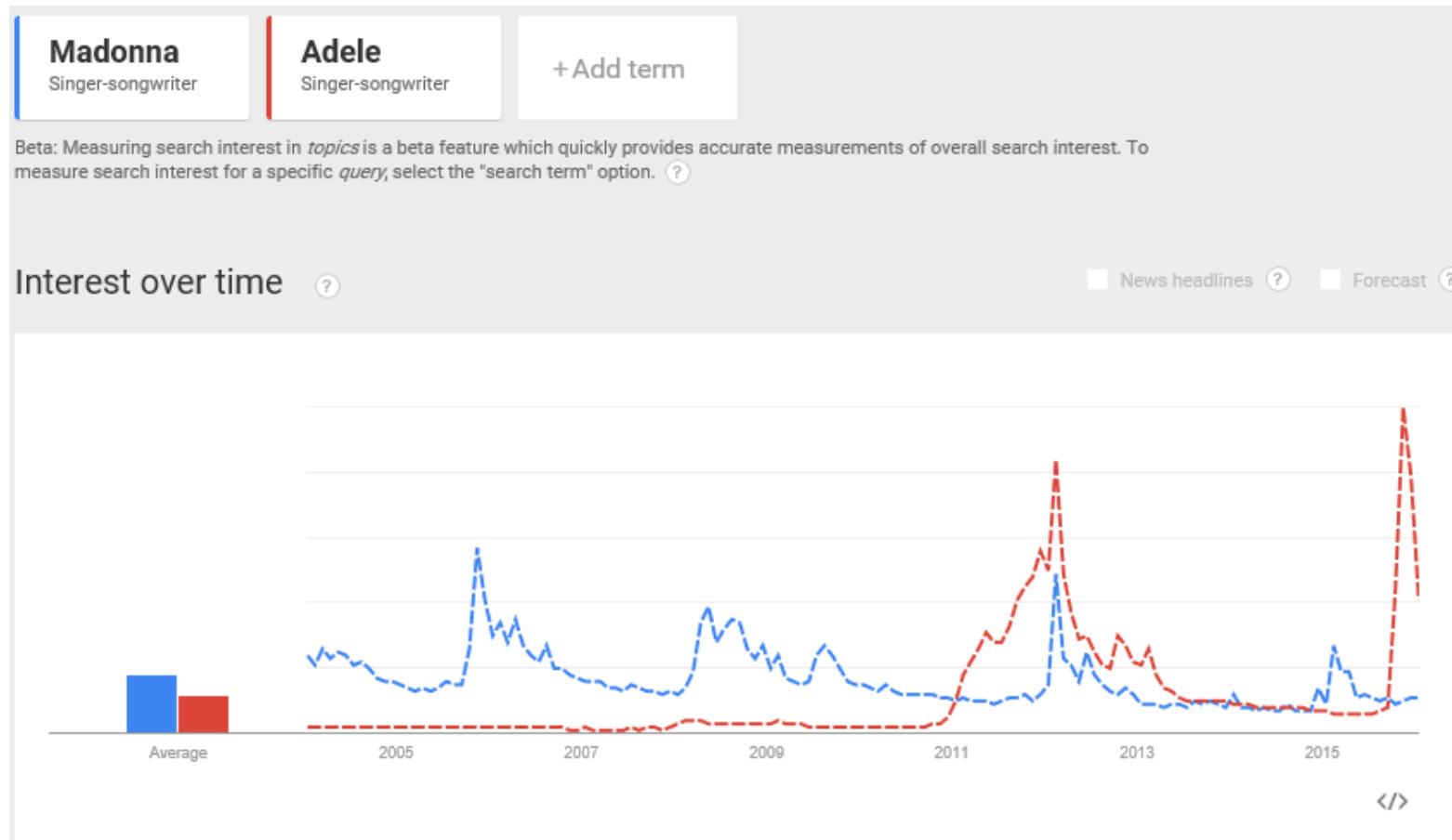
Visualization Type Selection: Temporal

- Temporal trends
 - Increasing
 - Decreasing
 - Stable
 - Cyclic



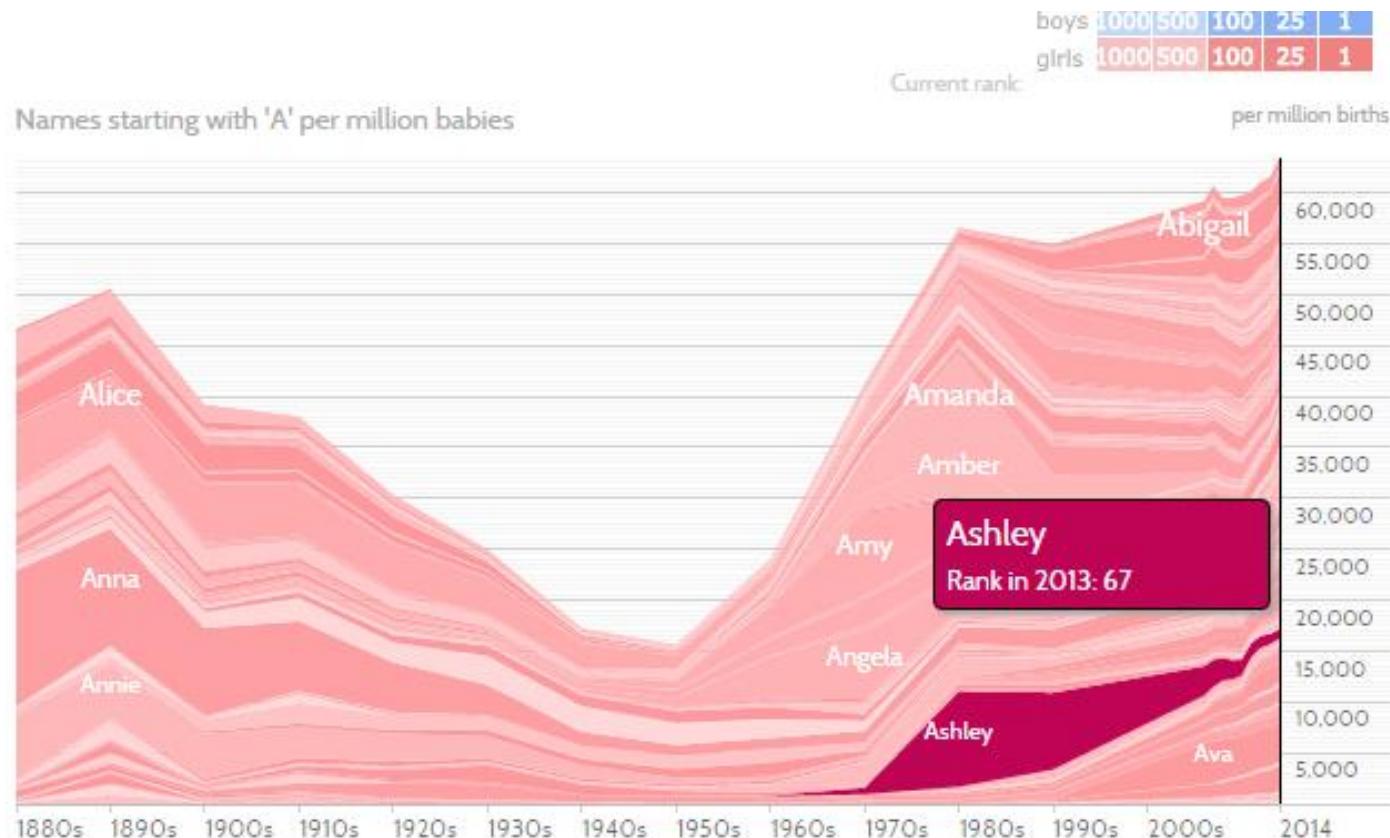
Visualization Type Selection: Temporal

- Line graph shows the trends over time.



Visualization Type Selection: Temporal

- Stacked graph illustrates individual and total trends.

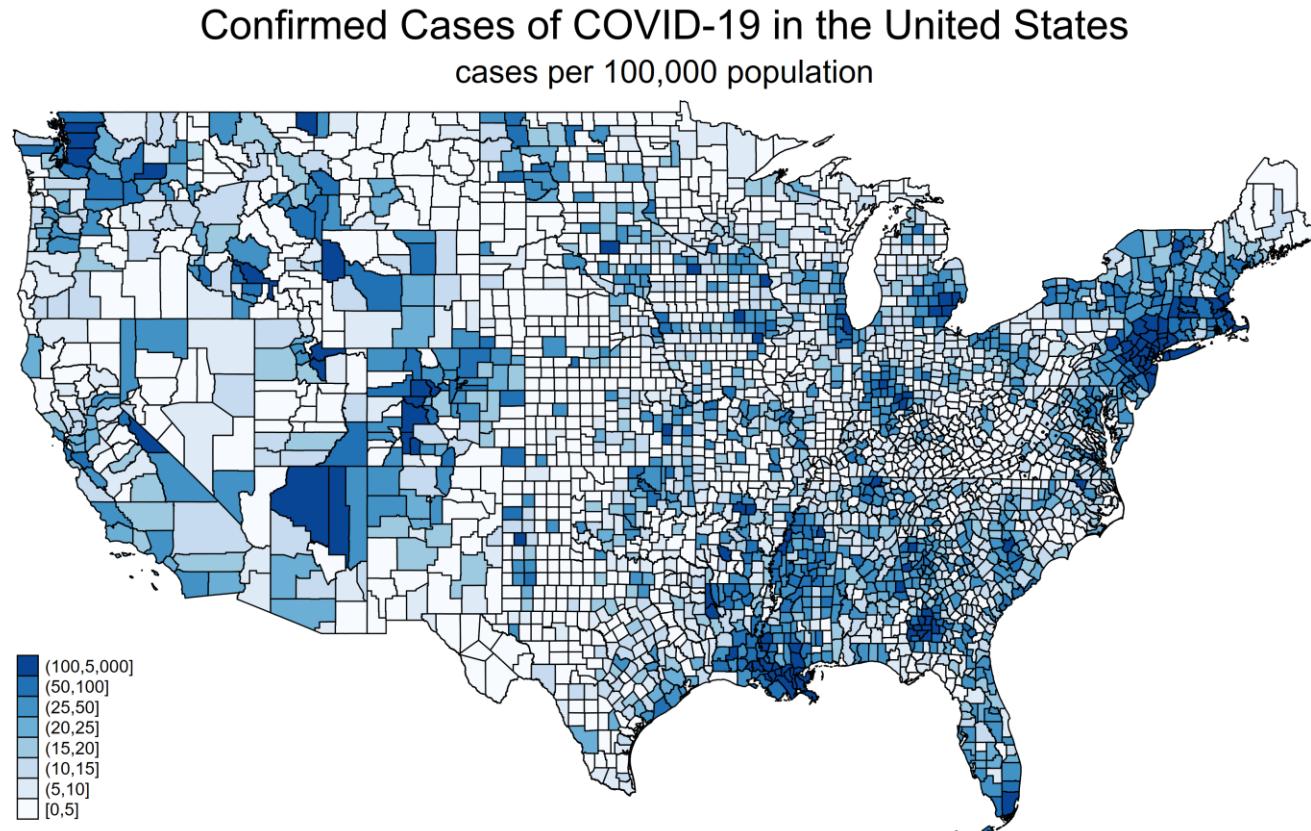


Visualization Type Selection: Geospatial

- Geospatial data analysis and visualization originated in geography and cartography, but are increasingly common in statistics, information visualization, and many other areas of science.
- The analyses aim to answer “WHERE” questions that use location information to identify their position or movement over geographic space.

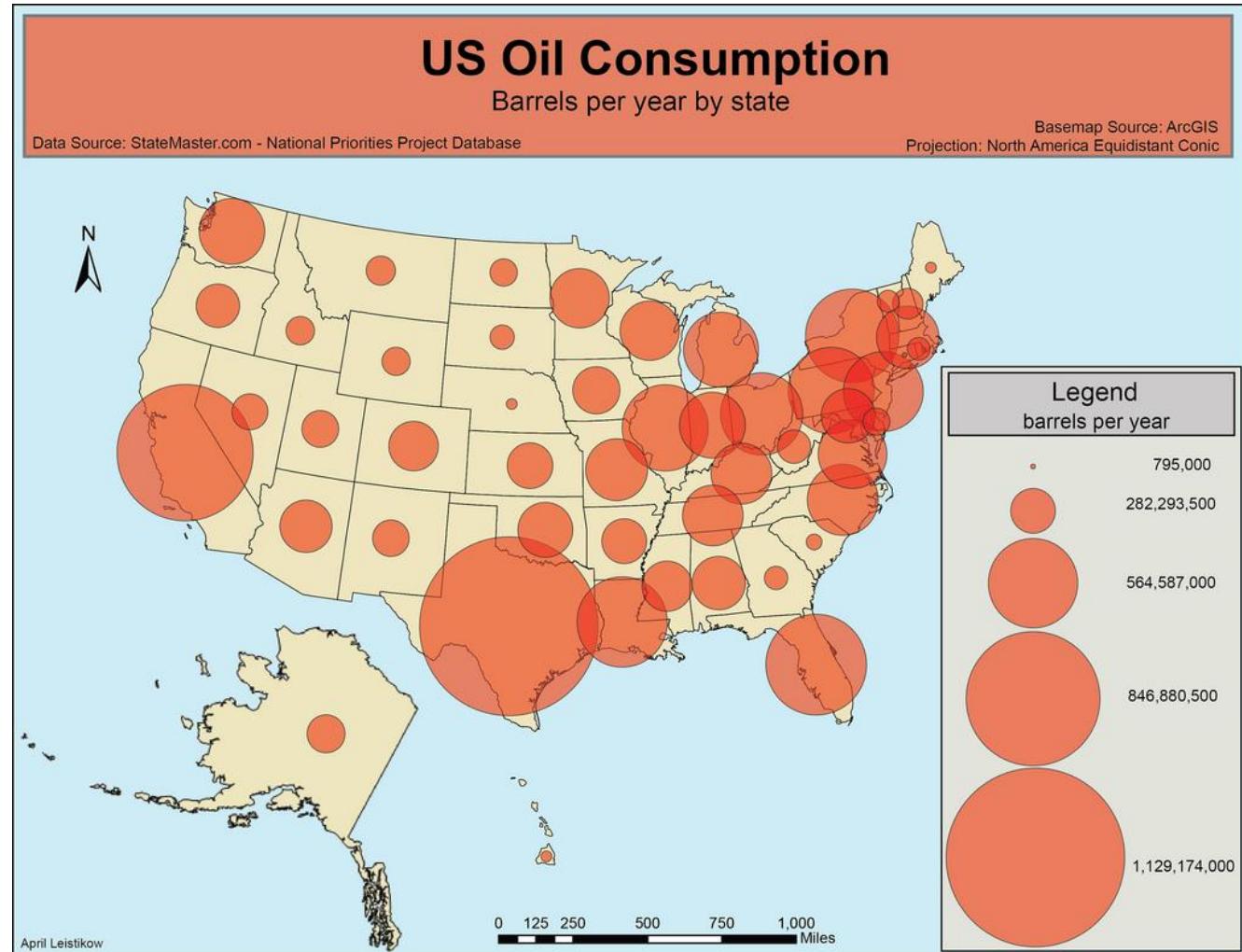
Visualization Type Selection: Geospatial

- **Choropleth map** represents data variables such as densities, ratios, or rates by proportionally colored or patterned areas.



Visualization Type Selection: Geospatial

- **Proportional symbol map** represents data variables by symbols that are sized, colored, etc. according to their amount. Data is (or can be) aggregated at points within areas.



Visualization Type Selection: Topical

- To answer “WHAT” question, we will be using texts to identify major topics, their interrelations, and their evolution over time at different levels of analysis – micro to macro.
- To generate visualization from text, text processing or natural language processing is needed to generate qualitative or quantitative features of the text.

Visualization Type Selection: Topical

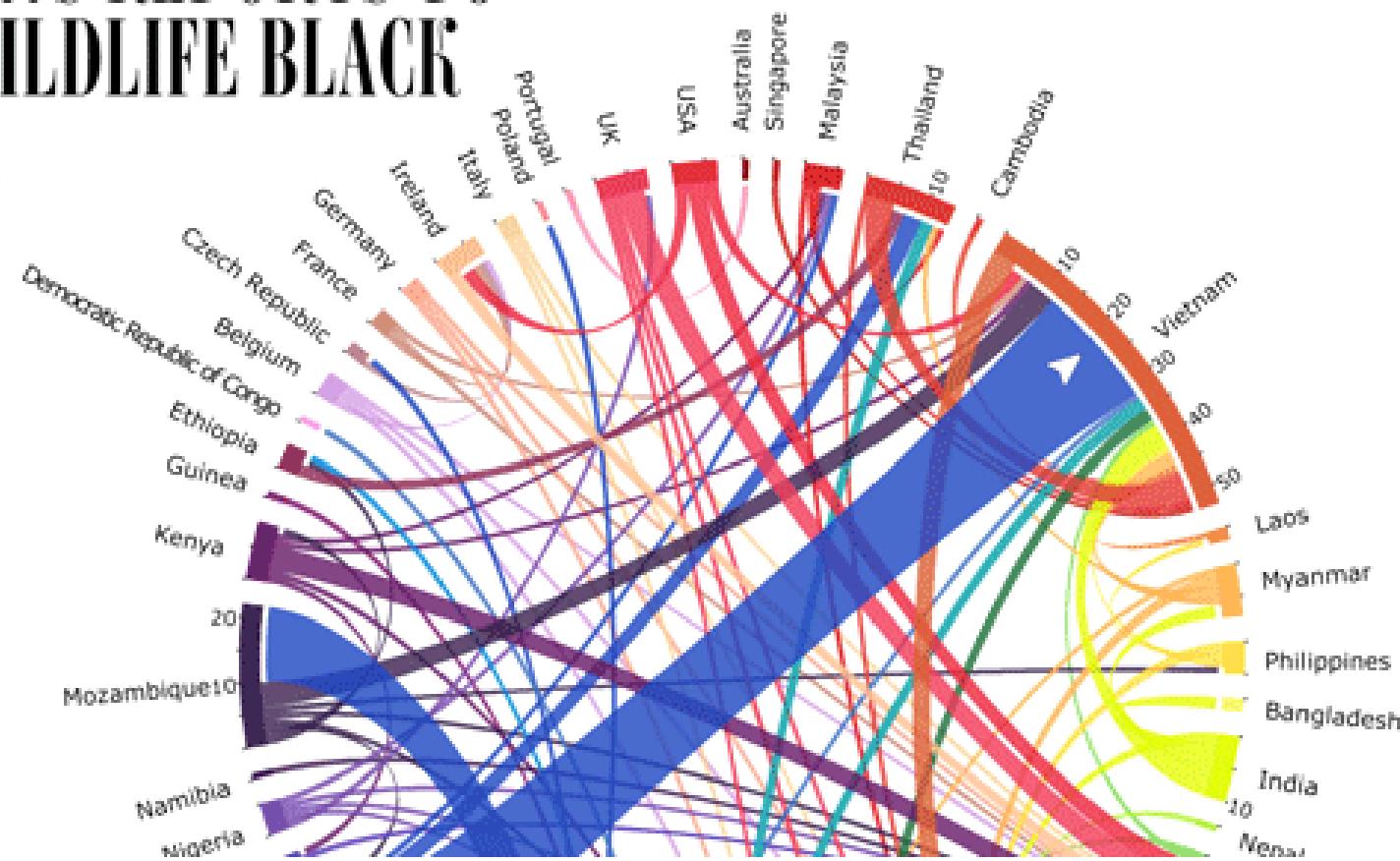
- **Word cloud** visualize the frequency of the words.
 - IMDB movie titles word cloud created with [Wordle](#).



Visualization Type Selection: Topical

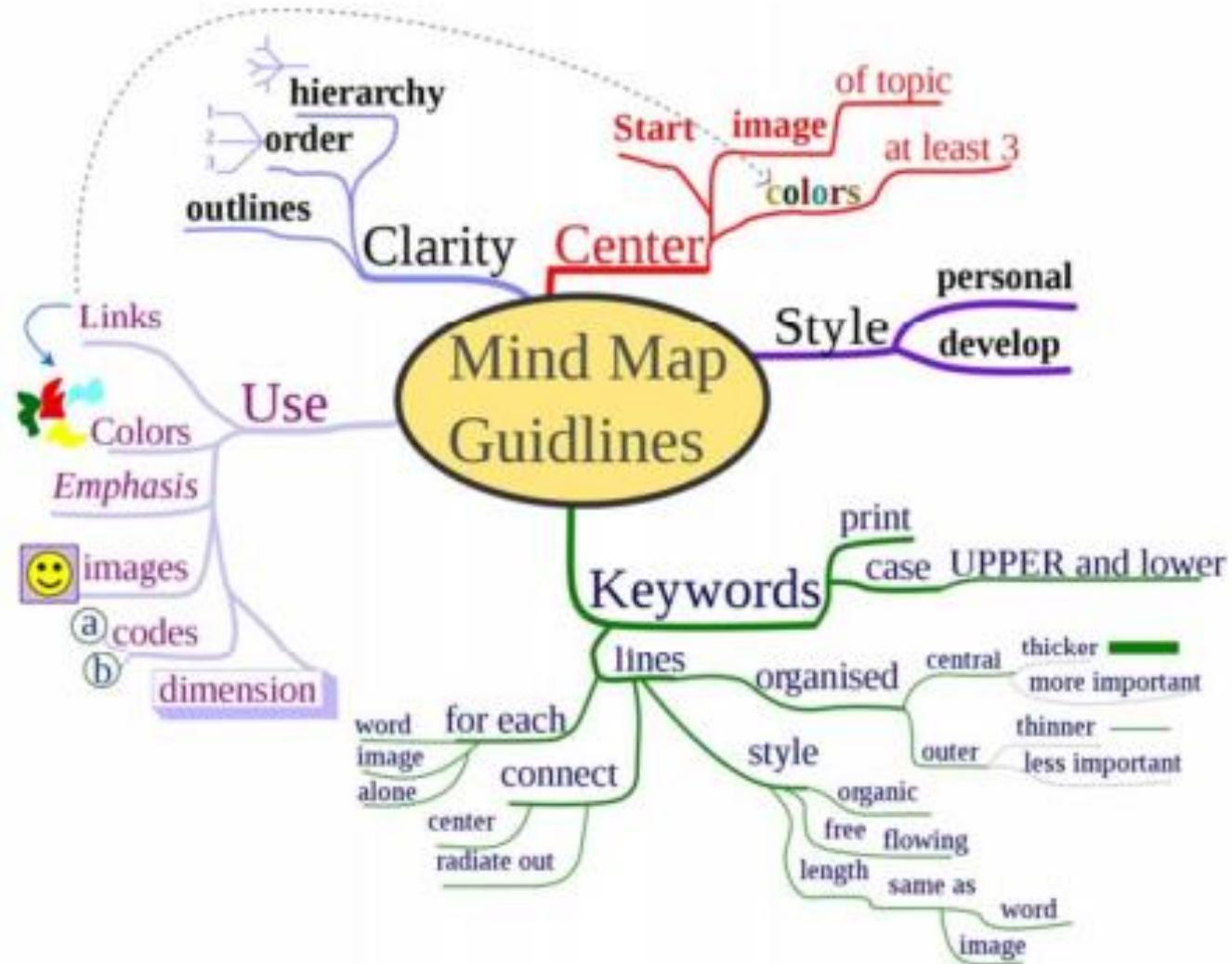
- Circular Visualization shows the relationships between entities.

USING NEWS REPORTS TO TRACK WILDLIFE BLACK MARKETS



Visualization Type Selection: Topical

- Concept maps are network graphs that show the relationships among concepts.

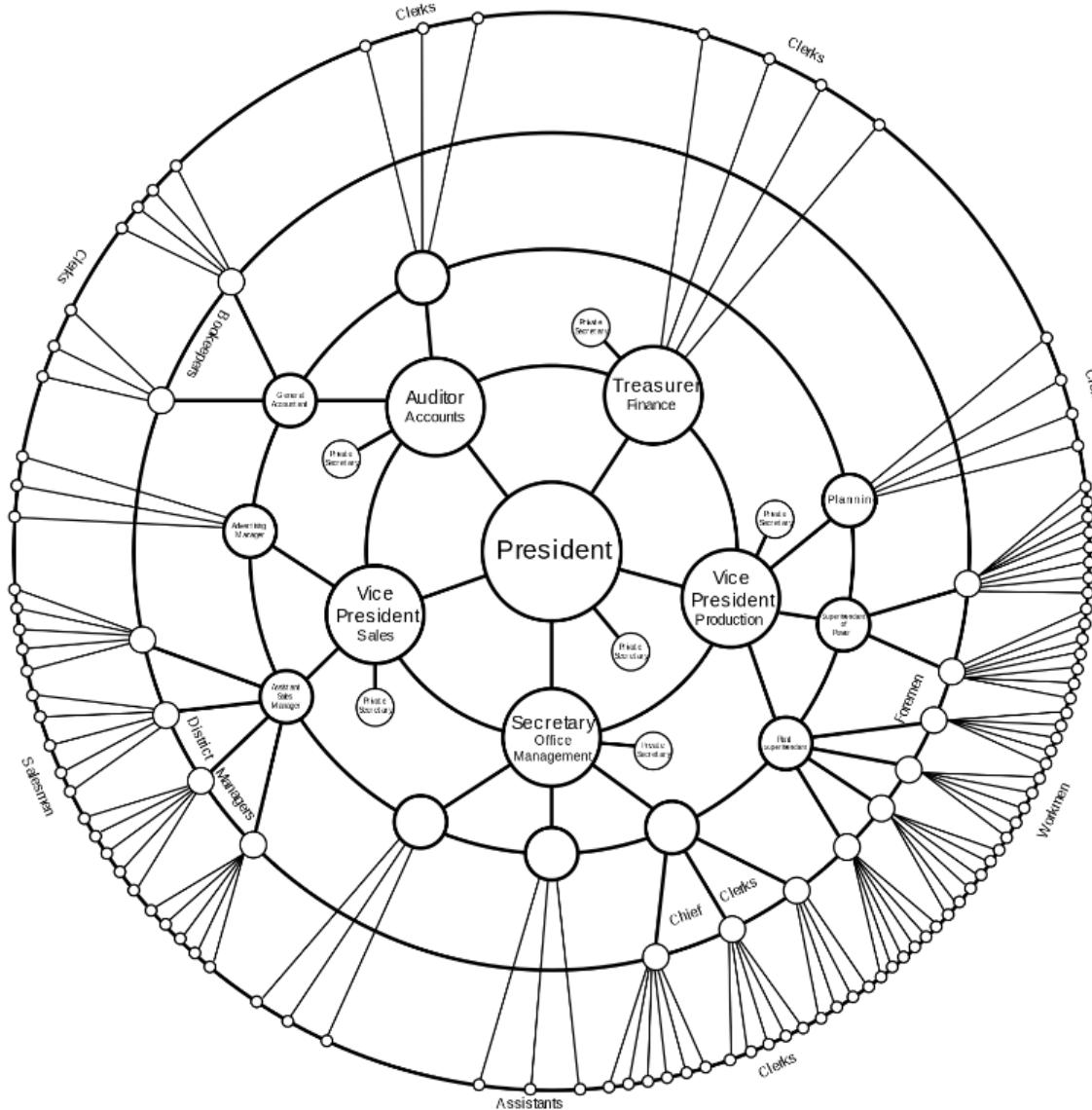


Visualization Type Selection: Network

- The “WITH WHOM” question can be answered by tree and network visualization.
 - Tree visualization utilizes structural, hierarchical data.
 - Network visualization can deal with more complex relationships between nodes.

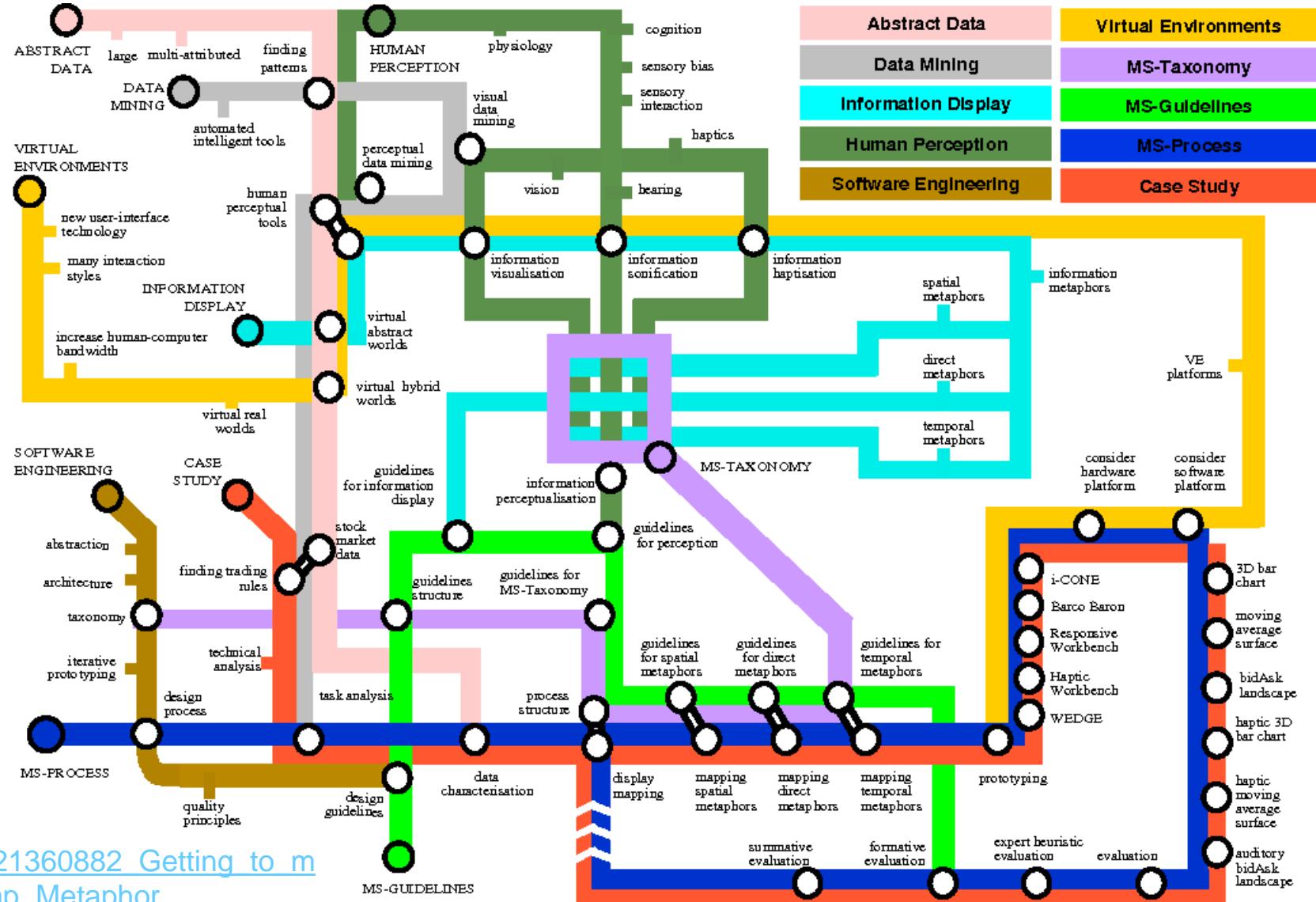
Visualization Type Selection: Network

- **Radial Tree Map** is a method of displaying a tree structure in a way that expands outwards, radially.



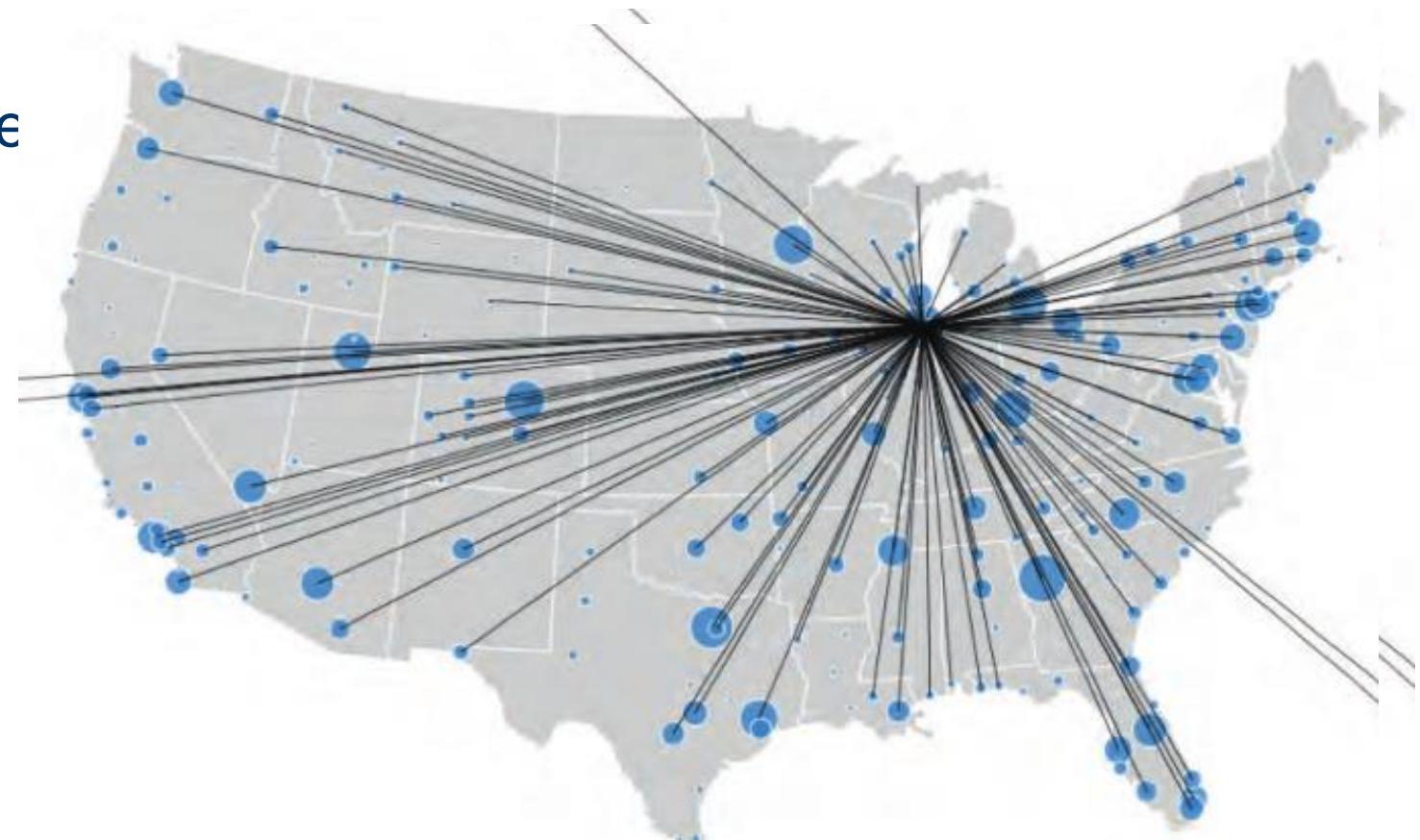
Visualization Type Selection: Network

- Subway map layout
 - Develop a diagram for explaining the major disciplines covered in the thesis and how the contributions of the thesis are derived from these disciplines.
 - Metro Maps – thesis map

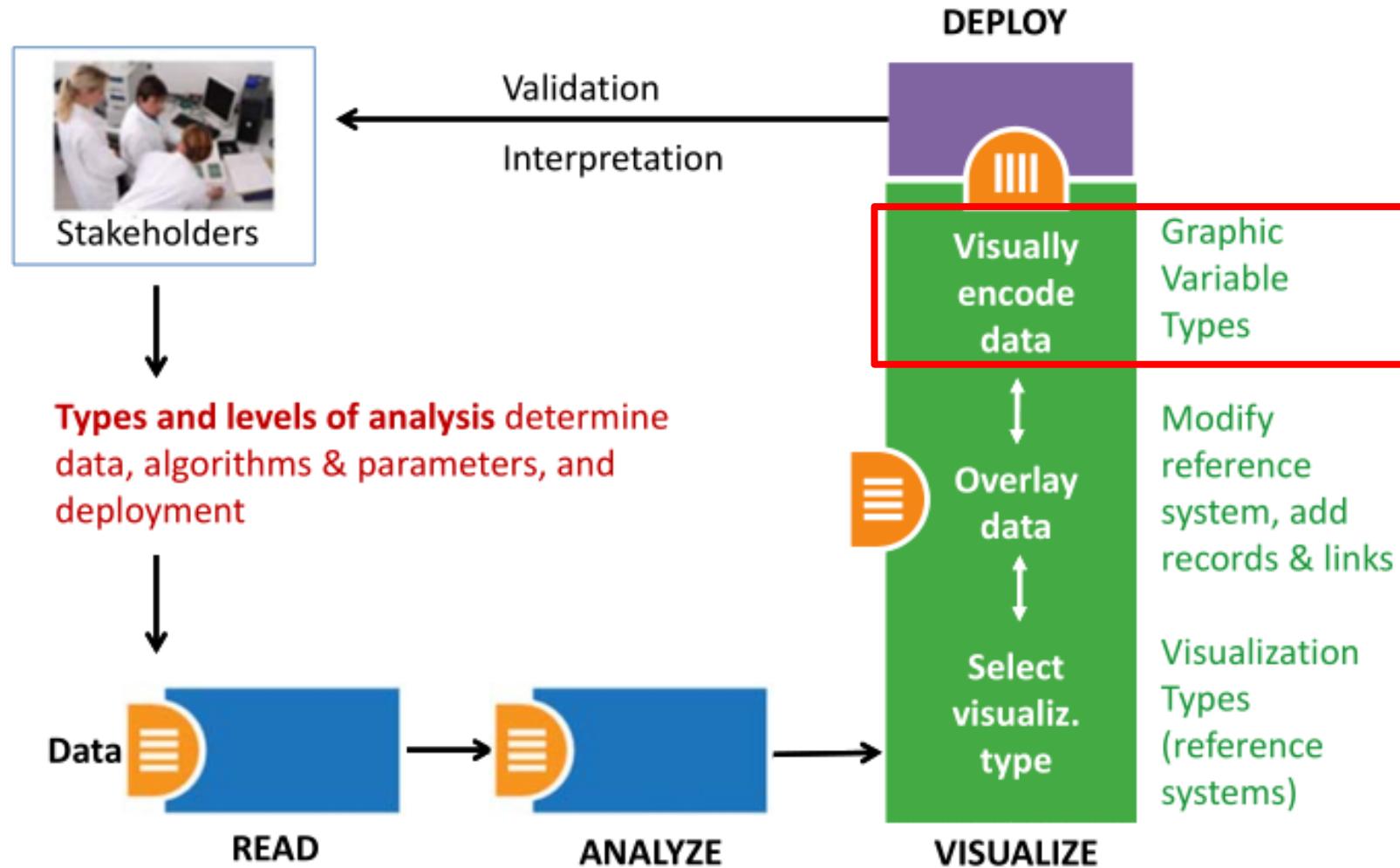


Visualization Type Selection: Network

- **Network overlays on geospatial maps**
 - Use a geospatial reference system to place nodes
 - E.g., airport traffic



Needs-Driven Workflow – Visually Encode data



Data Scale Types

- **Categorical** (nominal): A categorical scale, also called nominal or category scale, is **qualitative**. Categories are assumed to be non-overlapping.
- **Ordinal**: An ordinal scale, also called sequence or ordered, is **qualitative**. It rank-orders values representing categories based on some intrinsic ranking but not at measurable intervals.
- **Interval**: An interval scale, also called value scale, is a **quantitative** numerical scale of measurement where the distance between any two adjacent values (or intervals) is equal but the zero point is arbitrary.
- **Ratio**: A ratio scale, also called proportional scale, is a **quantitative** numerical scale. It represents values organized as an ordered sequence, with meaningful uniform spacing, and has a true zero point.

Data Scale Types – Example

- **Categorical** (nominal): Words or numbers constituting the names and descriptions of people, places, things, or events.
- **Ordinal**: Days of the week, degree of satisfaction and preference rating scores (e.g., Likert scale), or rankings such as low, medium, high.
- **Interval**: Temperature in degrees or time in hours. Spatial variables such as latitude and longitude are interval.
- **Ratio**: Physical measures such as weight, height, (reaction) time, or intensity of light; number of published papers, co-authors, citations.

Graphic Variable Types

- Quantitative
 - Position
 - x, y; possibly z
 - Form
 - Size
 - Color
 - Value (Lightness, Brightness)
 - Saturation (Intensity)
 - Texture
 - Pattern
 - Rotation
 - Coarseness
 - Size
 - Density gradient
- Qualitative
 - Form
 - Shape
 - Orientation (Rotation)
 - Color
 - Hue (tint)
 - Optics
 - Crispness
 - Transparency
 - Shading

Dynamic Visualization

- Previous sections demonstrates the workflow to create a static visualization.
- In order to present more information and make visualization dynamic, we can show multiple static images side by side or as an animation.
- Another way to make dynamic visualization is to introduce interaction into the design process.

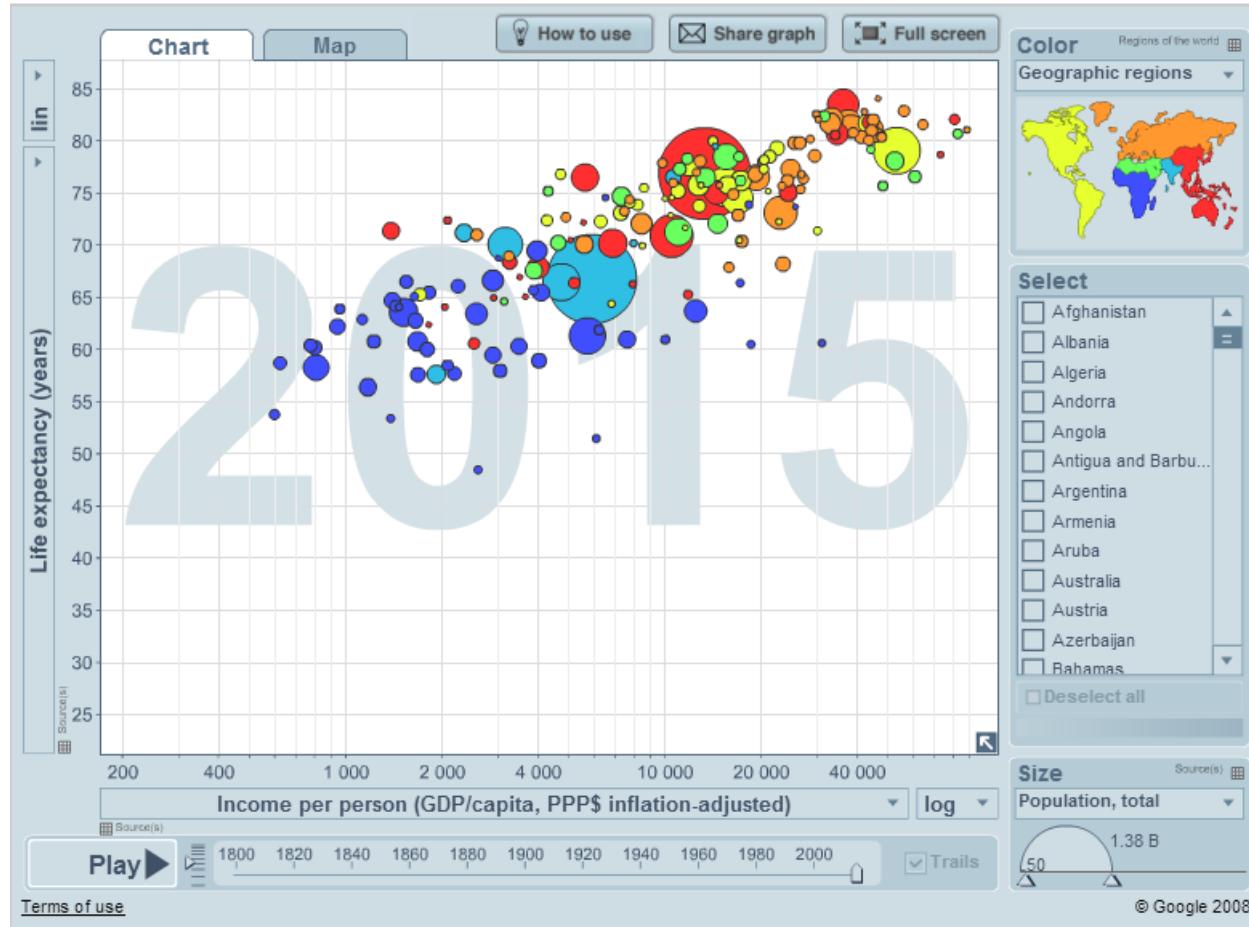
Shneiderman's Mantra

- User-Interface Interaction
 - Immediate interaction not only allows direct manipulation of the visual objects displayed but also allows users to select what to be displayed (Card et al., 1999)
 - Shneiderman (1996) summarizes six types of interface functionality
 - Overview
 - Zoom
 - Filtering
 - Details on demand
 - Relate
 - History
 - **“Overview first, zoom and filter, then details-on-demand.”**

Two Interaction Approaches

- User-Interface Interaction
 - Overview + detail
 - First overview provides overall patterns to users; then details about the part of interest to the user can be displayed. (Card et al., 1999)
 - Spatial zooming & semantic zooming are usually used
 - Focus + context
 - Details (focus) and overview (context) dynamically on the same view. Users could change the region of focus dynamically.
 - Information Landscape(Andrews, 1995)
 - Cone Tree (Robertson et al., 1991)
 - Fish-eye (Furnas, 1986)

Interactive Vis. Example: Gapminder World



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

<https://www.youtube.com/watch?v=BPt8EITQMIg> (Hans Rosling telling story)

40 years

Life expectancy at birth (years)

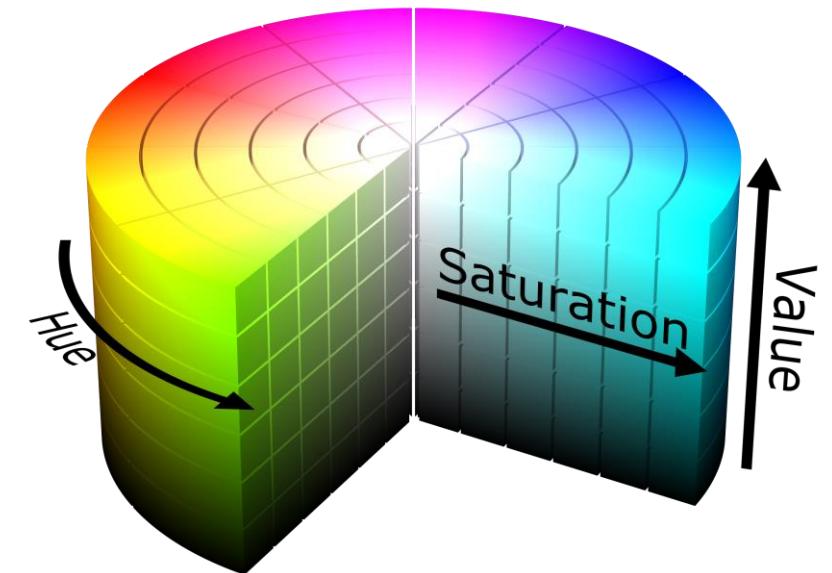


Outline

- **Concepts of Data Visualization**
 - Value, goals
 - History
- **Design of Visualization**
 - Workflow of data visualization
 - **Color, Size, Text, Titles, Labels**
 - Choosing the right chart
- **Case study: COVID-19**
 - Mask wearing trends
 - People concerned about the impact of COVID-19

Color

Color is one of the **most powerful aesthetic features** because it's an attention-grabber. It's the first thing we notice, and it can immediately highlight specific insights or identify outliers. The data, not **personal favorites or brand colors**, should drive the use of color to make a point.



Measurable

- Does the color scale match my data?
- Does the color move from light to dark, or is it stepped to best represent what you're measuring?



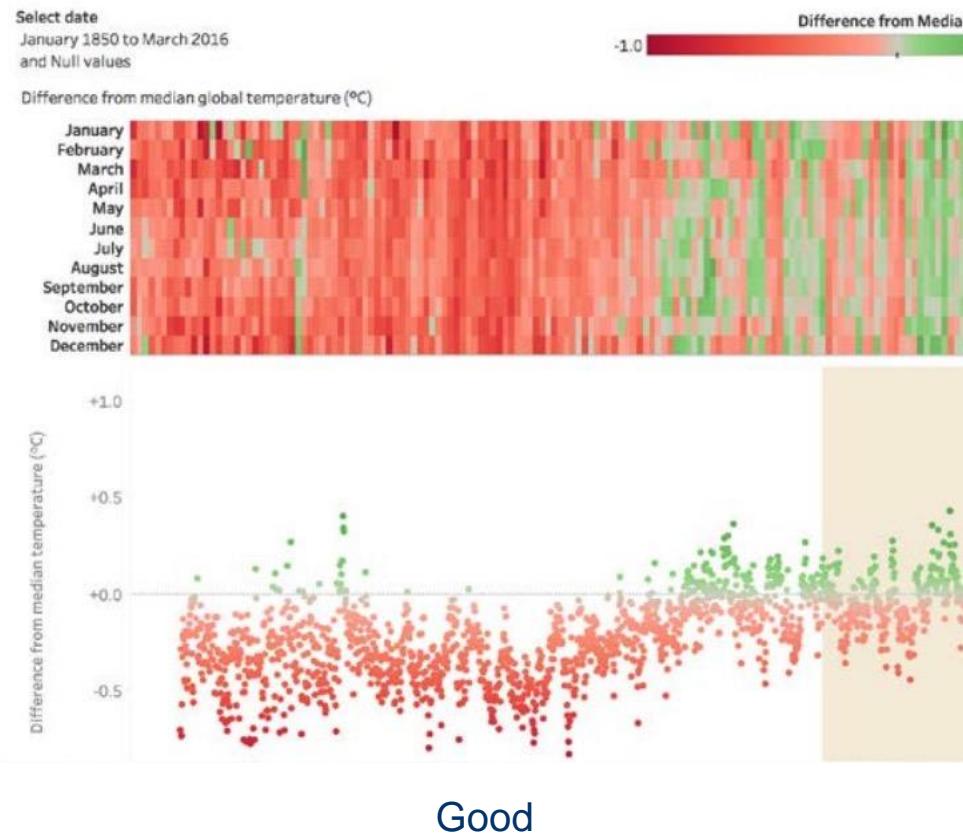
Relatable

- Semantically-resonant colors help people process information faster.
- So use yellow to depict bananas, red to represent heat.

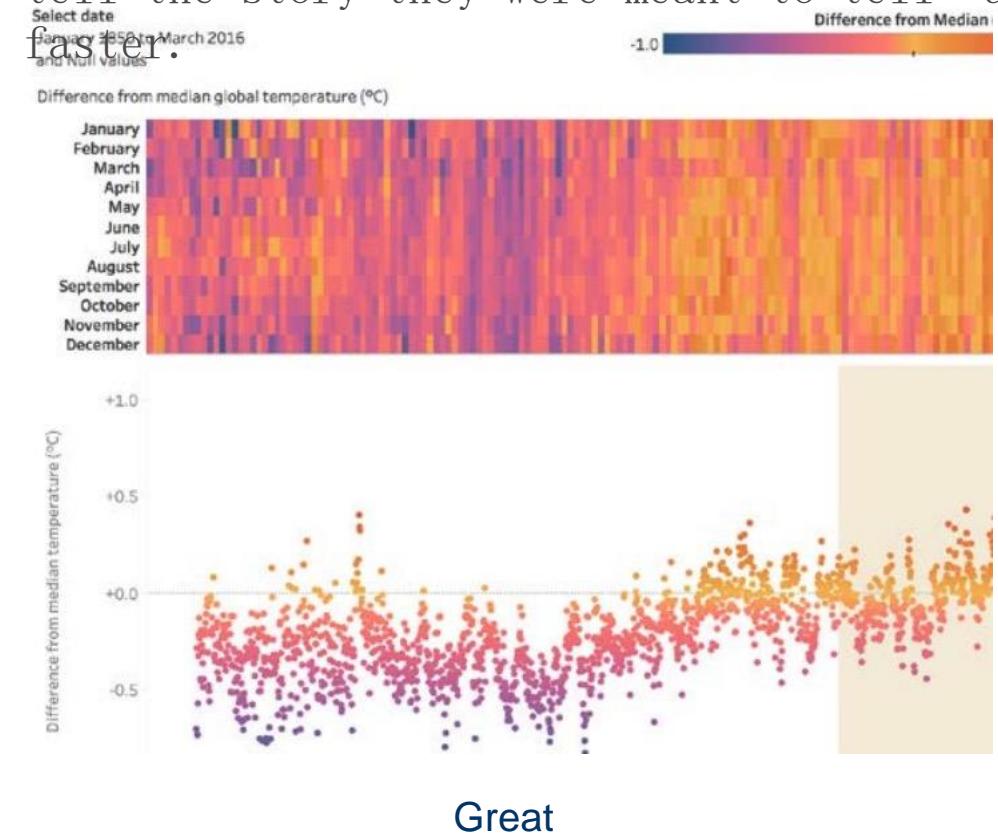


Data from 100 observations of global surface temperatures (°C) around the world from 1961- 1990

While these visualizations are accurate, the color **red** represents **cooler** temperatures, and doesn't resonate with the information the data is trying to portray.



Swapping in semantically- resonant colors to the same dashboard really **heats things up**. With a little extra thought into color choices and pallets, the data points now tell the story they were meant to tell—and faster.



Color Blind Friendly - Accessibility



Left: Normal vision, Right: Deutanopia

1. Avoiding **problematic** color combinations, e.g., red & green / green & brown / green & blue...
2. Selecting color blind friendly **palettes**
3. Using different **textures** and patterns to highlight important information—not just color
4. Using **symbols and icons** to supplement color-coded messages, warnings and alerts
5. Using highly **contrasted** color combinations
6. Adopting **minimalistic** design to help avoid unnecessary confusion.

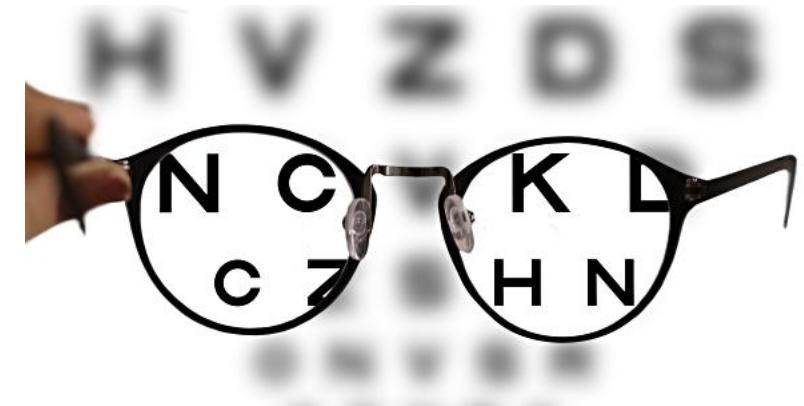
Size

The bigger the object, the bolder it looks.
Bold shapes and colors might work well with
bar charts and area charts, but they may
also look **gaudy and garish** when used in a
different chart, like a treemap. Use size to
draw emphasis to your key message, not
obscure it.

The image shows two large, bold, black capital letters 'A' positioned side-by-side. They are rendered in a clean, sans-serif font. The letters are identical in size and style, emphasizing the concept of uniformity or emphasis through size.

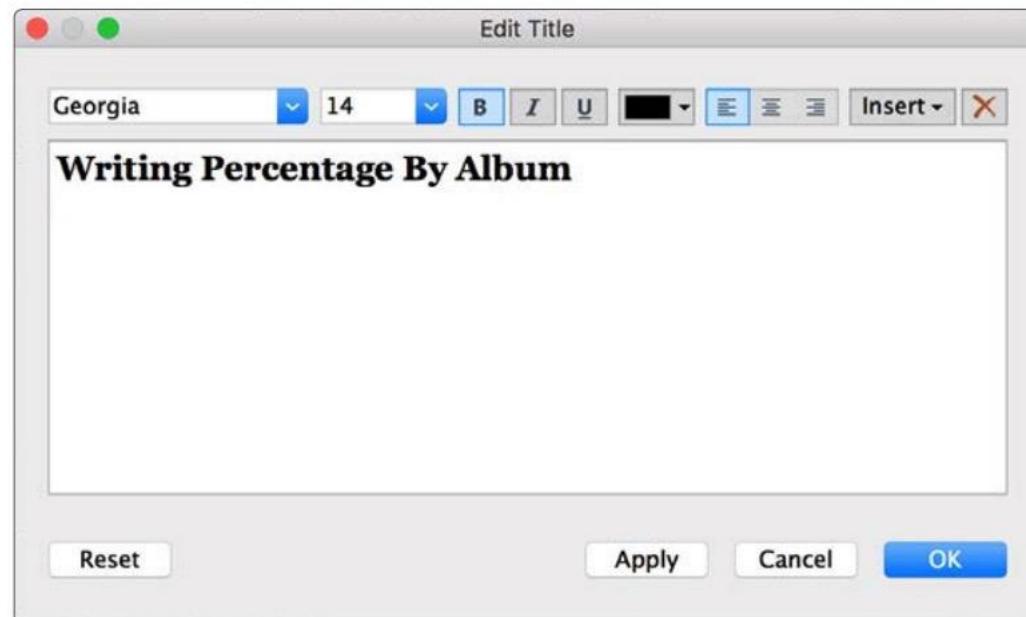
Text

- Readability is essential.
- Make the most important information stand out.



Titles

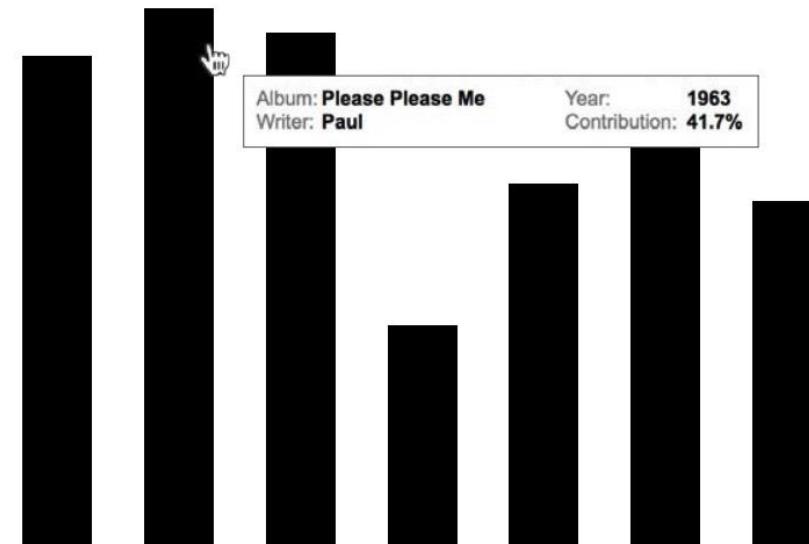
- Keep them short, but powerful.
- Convey the point, message or story in the fewest words possible.



Labels

Find the sweet spot. Too many mark labels can be very distracting.

Try labeling the **most recent mark, or min/max**. Save additional and more detailed information for **tooltips**.



Outline

- **Concepts of Data Visualization**
 - Value, goals
 - History
- **Design of Visualization**
 - Workflow of data visualization
 - Color, Size, Text, Titles, Labels
 - Choosing the right chart
- **Case study: COVID-19**
 - Mask wearing trends
 - People concerned about the impact of COVID-19

Choosing the Right Chart

- There are many different charts, how to choose the effective chart for your own purpose?

REMEMBER THIS IMPORTANT NUMBER

23%

COMPARE 2 OR MORE THINGS

Side by Side **Icon Array** **Pie/Donut** **Bar/Column**

Benchmark Line **Slopegraph** **Back-to-Back** **Dot Plot**

COMPARE TO A TARGET

Combo **Bullet Chart** **Indicator Dots**

Metric A **Metric B** **Metric C**

WATCH FOR OVERLAPPING POINTS

SHOW SURVEY RESPONSES

Stacked Bar **Small Multiples** **Diverging Bar** **Bar/Column** **Number & Icon** **Nested**

THESE ARE THE PARTS OF THIS WHOLE

Pie/Donut **Stacked Bar** **Histogram** **Tree Map** **Map**

FOR BRANCHING QUESTIONS!

VISUALIZE OPEN-ENDED COMMENTS

Quote & Pic **Word cloud** **Stock photo Rep** **Heat Map** **Prezi**

HEY, THINGS CHANGED OVER TIME

Line **Stacked Column** **Deviation Bar** **Slopegraph** **Dot Plot** **Sankey**

THIS THING CHANGES WHEN THAT THING DOES

Scatterplot **Draw It**

FOR MORE SEE
STEPHANIEEVERGREEN.COM/TAG/STEP-BY-STEP
STEPHANIEEVERGREEN.COM/BLOG
[PRESENTING DATA EFFECTIVELY](#)

CHART CHOOSER 3.0
BY STEPHANIE EVERGREEN

Choosing the Right Chart

- When a single number is important
- How two or more numbers are alike or different
- How we are better or worse than a benchmark
- When there are parts of a whole
- How things change over time

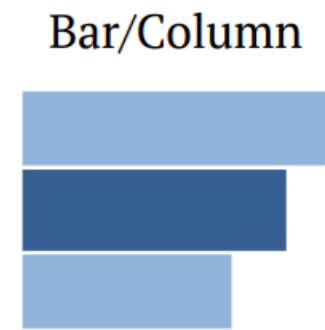
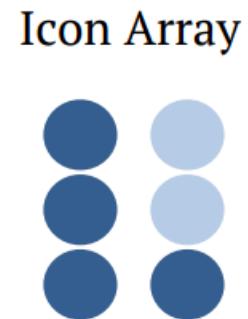
When a single number is important

- Single number
 - Showing count, frequency, mean, median etc
- Stories can be told with a single number
 - On average, our mathematics student achievement is 64%
 - We served 123,000 clients last year
 - 9 out of 10 consumers prefer our pizza over our competitors
 - Only 27% of children in this key neighborhood had a dental visit in the past year
 - Chances of dying from a snake bite are just 1 in 50 million
 - ...

Displaying a Single Number

- Simple text with a single large number
- Icon Array
- Donut or Pie Graph
- Bar

Big Number
23%



Simple text with a single large number

- Simple text can be a great way to communicate if you have just a number or two to share.
- Solely using the number and a few support words to clearly make your point.

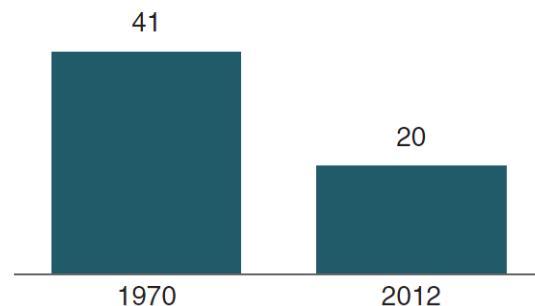
1 out of 50 million will die of a snake bite.

SOURCE: Institute of Food and Agricultural Sciences

Simply Text

Children with a "Traditional" Stay-at- Home Mother

*% of children with a married
stay-at-home mother with a
working husband*



Original Graph

20%

of children had a
traditional stay-at-home mom
in 2012, compared to 41% in 1970

Simple Text

- You have some numbers does not mean that you need a graph!
- A simple sentence would suffice in this particular example.
- **Concern:** Overload. Use simple text if you have a single large number you want to convey.

Icon Array

- Icon arrays
 - One icon (e.g. square, circle) is repeatedly 10, 100 or 1,000 times and then some of the icons are colored to represent a percentage or proportion.
 - Icon arrays have been shown to be especially effective for people with low numeracy skills.

9 out of 10 consumers prefer our pizza over those other guys.



Combining icon array with a single large number

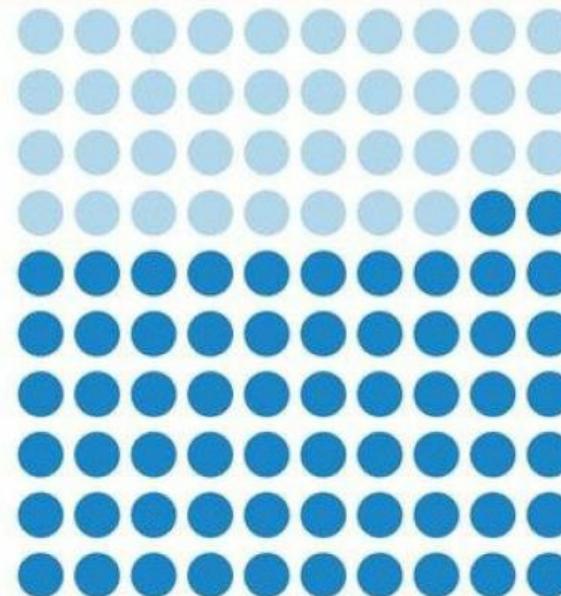
Before

Not so good news

- They have low capacity (in 62% of organizations, evaluation is done by internal nonprofessional evaluator)
- 15% of organizations report trouble knowing where or how to find an external evaluator
- In three quarters of organizations, less than 5% of budget is spent on evaluations
- They use rudimentary methods (65% before and after evaluations)

After

WHO Does the Evaluation?

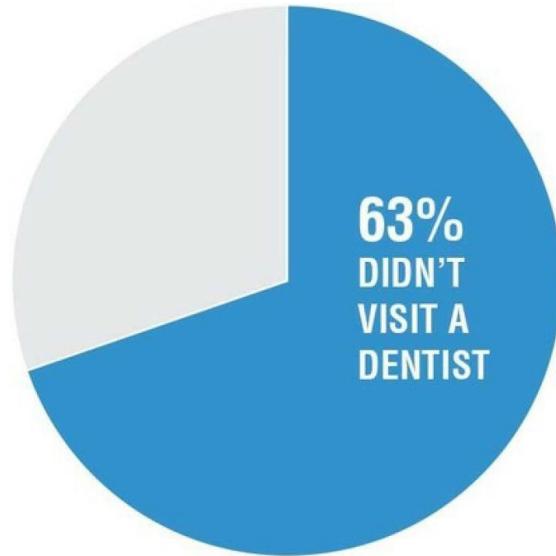


62%

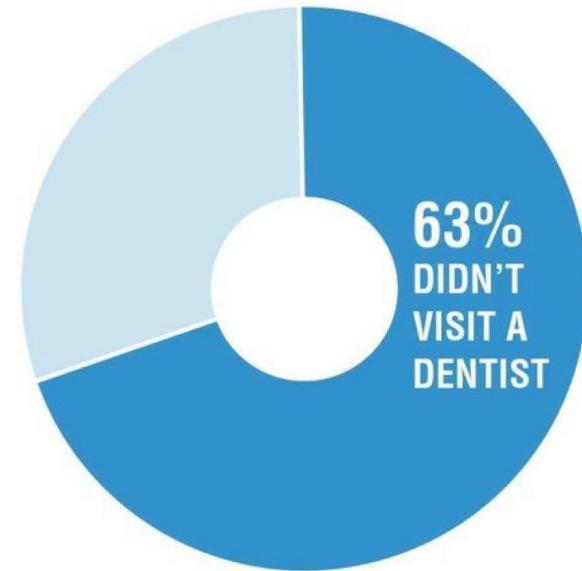
nonprofit
evaluations done
by nonprofessional

Donut or Pie Chart

- Donut chart is very similar to pie chart.
- One major difference: the middle of pie is gone for donut chart.



Pie Chart



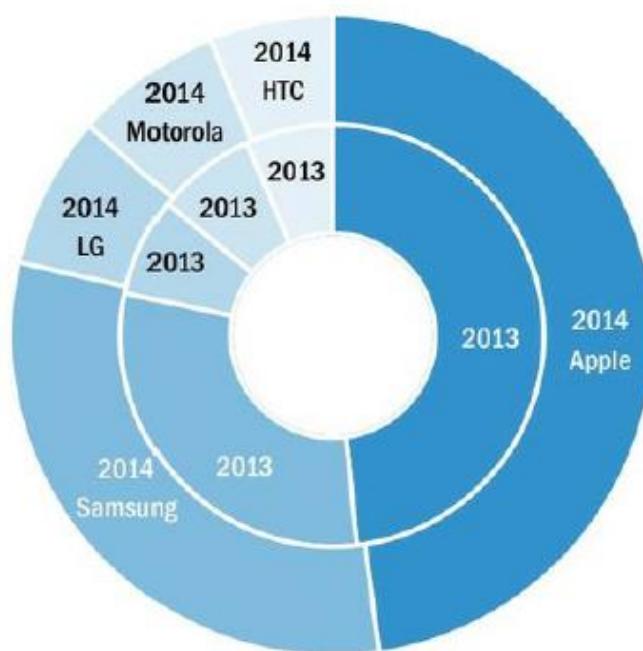
Donut Chart

The weakness of donut or pie chart

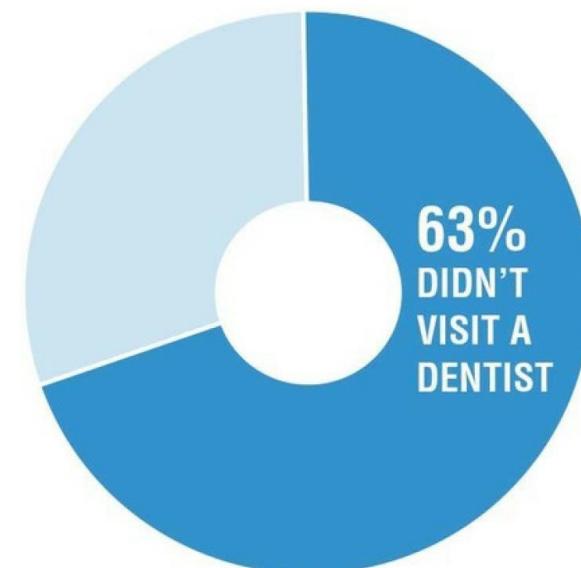
- People are not good at interpreting angles and curvature.

Smartphone market share continues to be dominated by Apple, followed closely by Samsung.

Source: comScore, Inc.

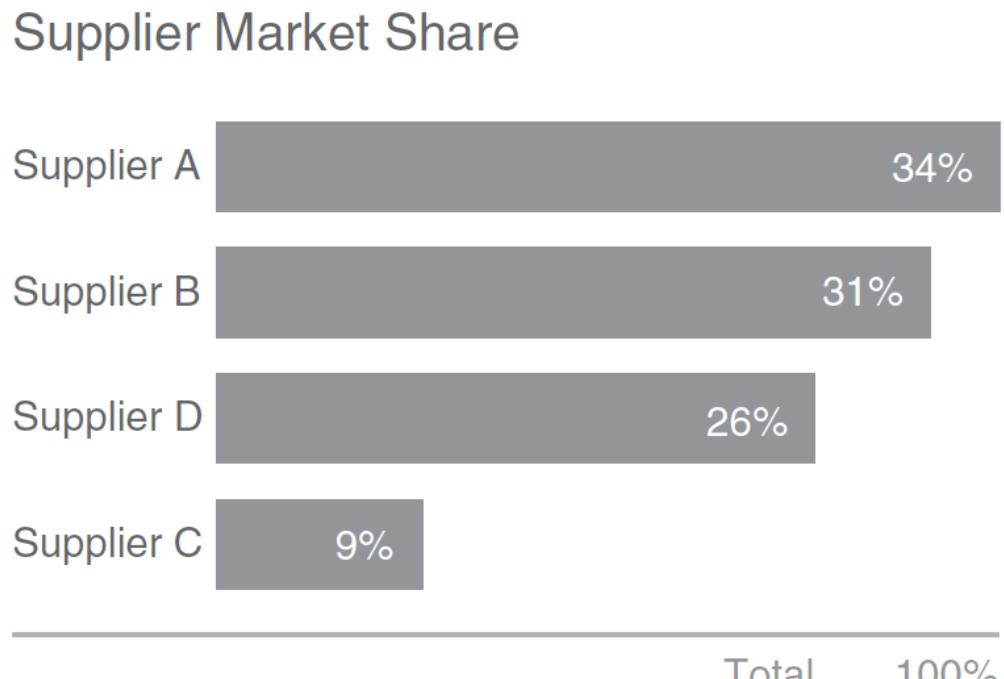


The best way to use a pie/donut chart is to **highlight just one chunk** to support a single number.

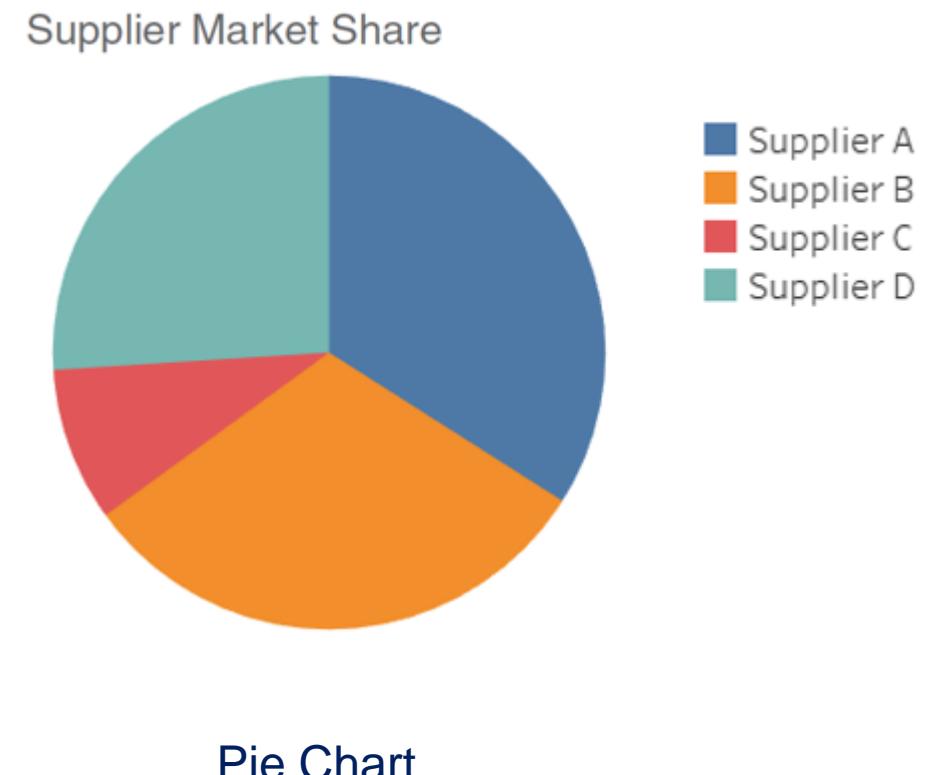


Bar Chart

- Bar chart can be more effective than pie chart
- In the bar chart, the baseline are aligned, we can easily compare the end points.

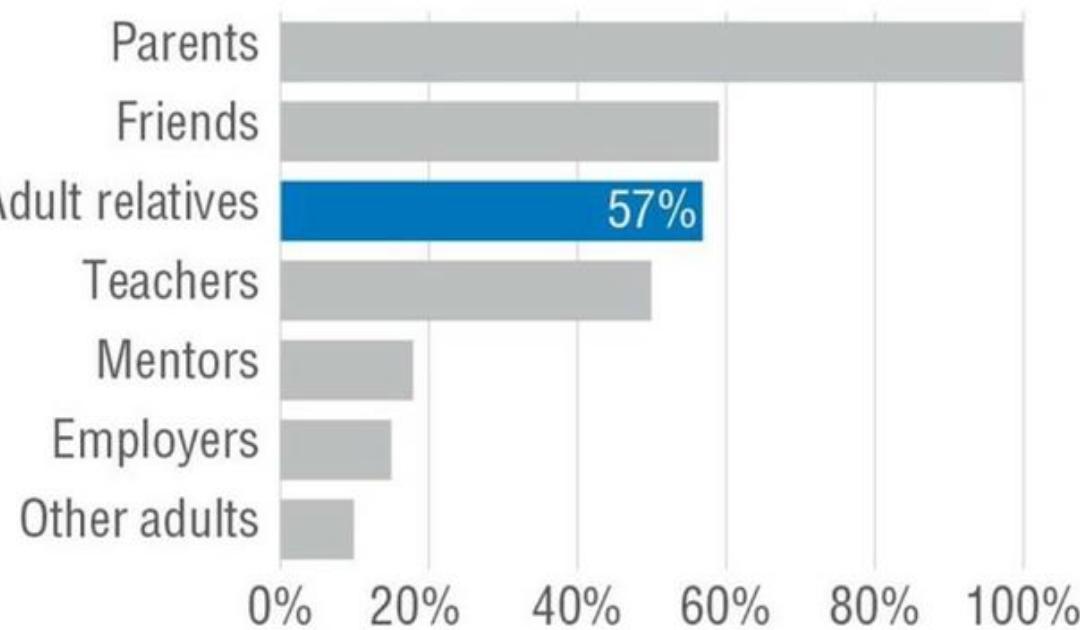


Bar Chart



Bar Chart

Nonparental **adult relatives** are important to Hispanic/Latino students. 57% reported talking with them about college aspirations.



- A simple bar can be focused to help audience to remember a single number.

Choosing the Right Chart

- When a single number is important
- How two or more numbers are alike or different
- How we are better or worse than a benchmark
- When there are parts of a whole
- How things change over time

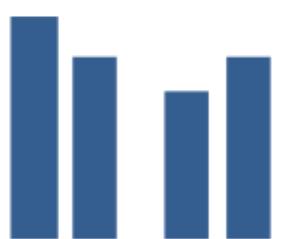
How two or more numbers are alike or different: visualizing comparisons

- Stories can be told about comparison of two numbers
 - Option A is better than Option B
 - Group F and Group G disagreed on five of the eight measures
 - There were no differences between the treatment and control groups
 - Youth rates are significantly lower than adult rates
 - We have equal participation from boys and girls with the exception of girls aged 10 to 13
 - All sites increased performance, but Site L did so at a faster rate than other sites
 - ...

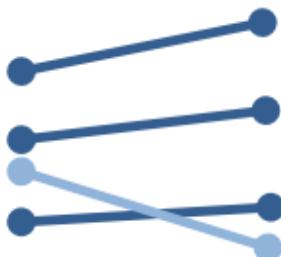
Visualizing Comparisons

- Side by Side Column or Bar Charts
- Slopegraph
- Back-To-Back Bars
- Dot Plot
- Dumbbell Dot
- Small Multiples

Side by Side



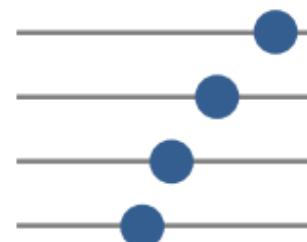
Slopegraph



Back-to-Back



Dot Plot



Dumbbell Dot

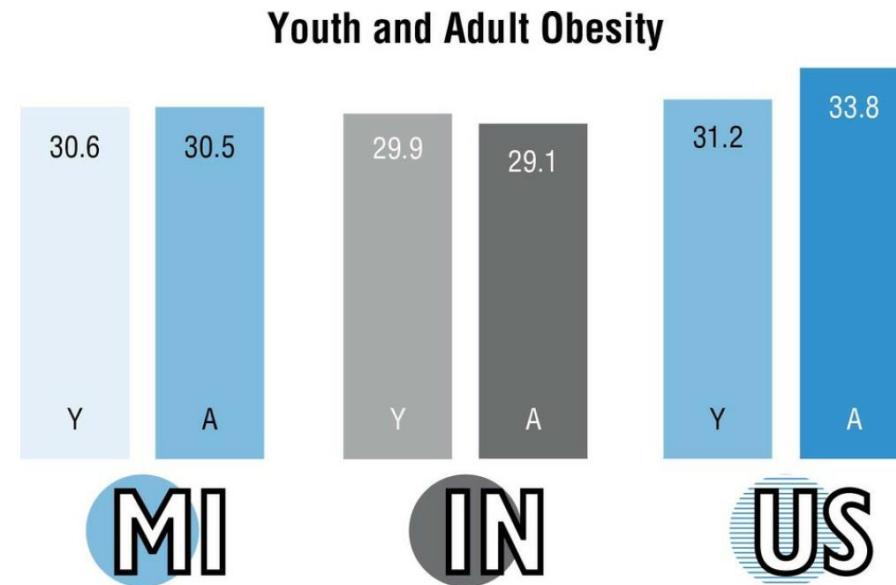


Small Multiples



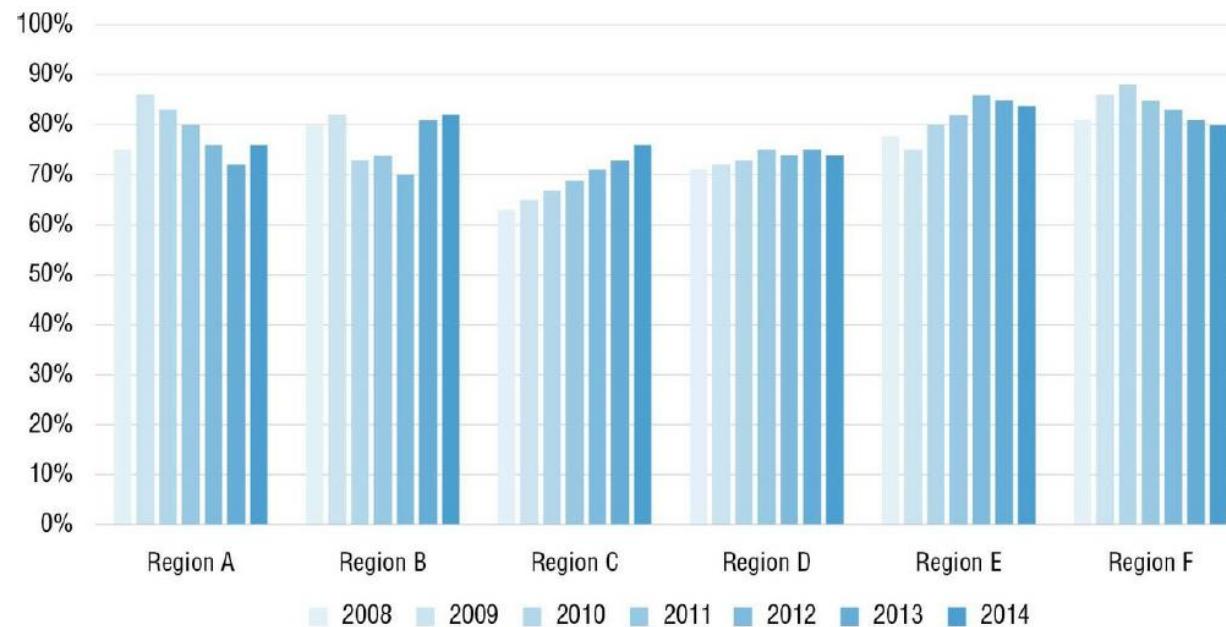
Side by Side Column Charts (or Bar Charts)

- Column charts and bar charts are similar. Their difference lies in orientation.
 - A bar chart is horizontally orientated whereas the column chart is vertically orientated.
- Side by side column charts (or bar charts) are possibly the most ubiquitous way to showing how two or more numbers compare.
 - Side by side column charts (or bar charts) are also referred to as clustered columns or bars



Limitation of side by side column charts (or bar charts)

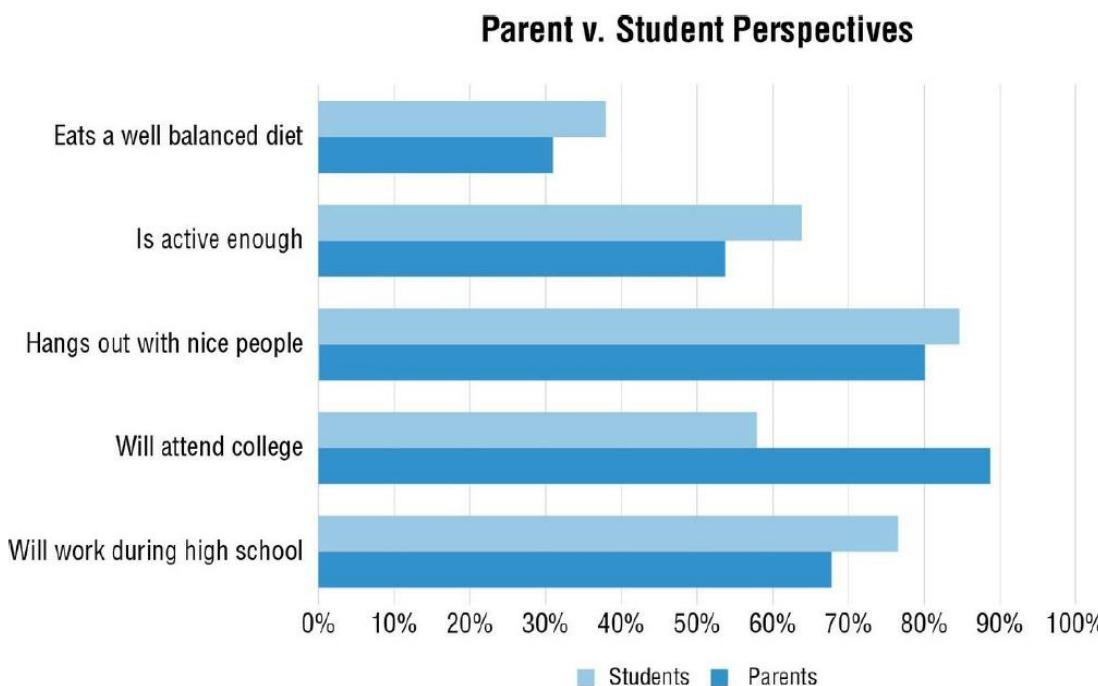
- Effectiveness really stops with two comparisons
 - It is difficult for people to compare between nonadjacent bars.
- Once we start slicing in a third, fourth, or fifth column within each group, we are asking brains to do too much



This side by side column has too many columns

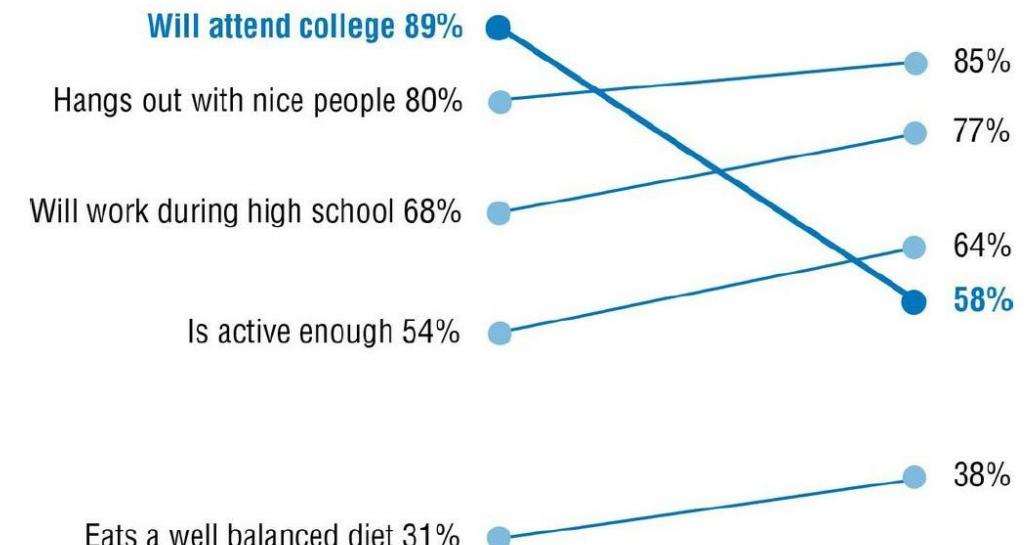
Slopegraphs

- Slopegraphs are perfect for highlighting the story of how just one category decreased when other categories increased
- Or to show that one change at a rate much faster than others.



Side-by-Side Bar Chart

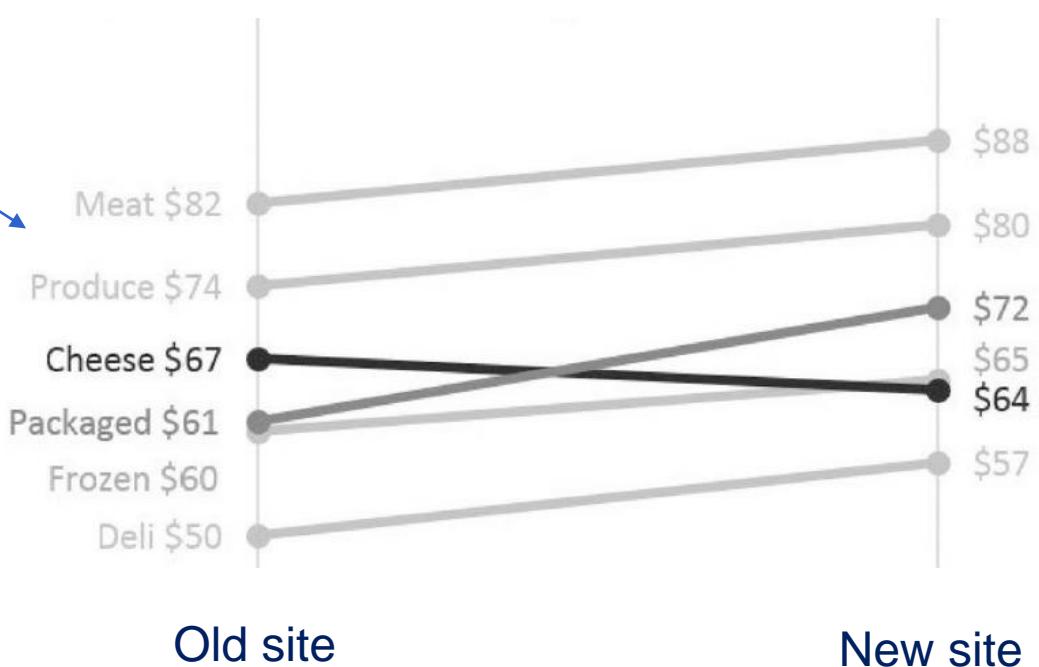
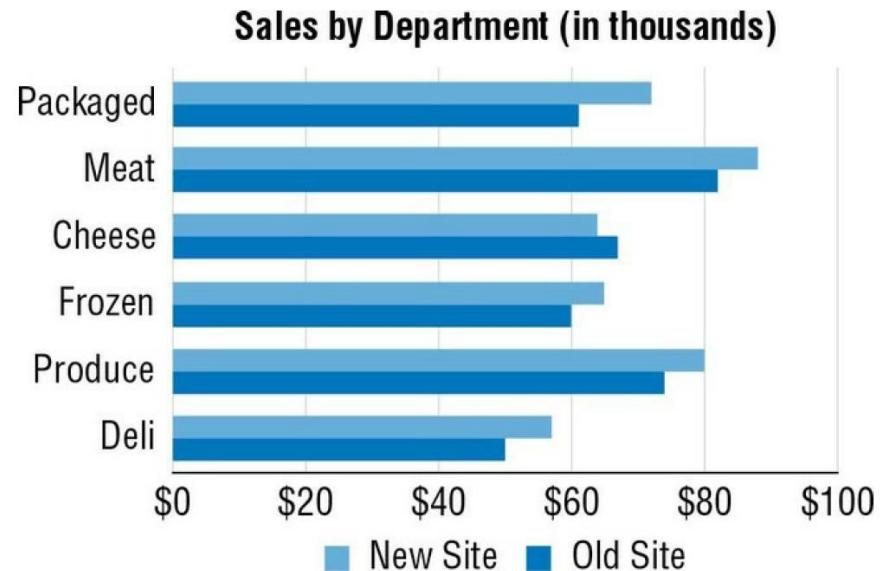
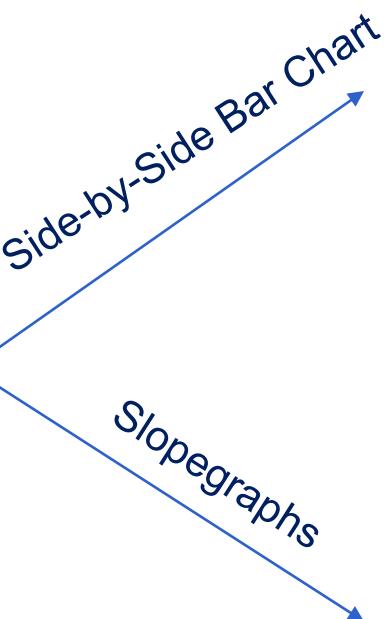
Surprisingly, students have lower expectations to go to college than their parents.



Slopegraph

Slopegraphs

Sales by Department (in thousands)		
	Old Site	New Site
Deli	\$50	\$57
Produce	\$74	\$80
Frozen	\$60	\$65
Cheese	\$67	\$64
Meat	\$82	\$88
Packaged	\$61	\$72

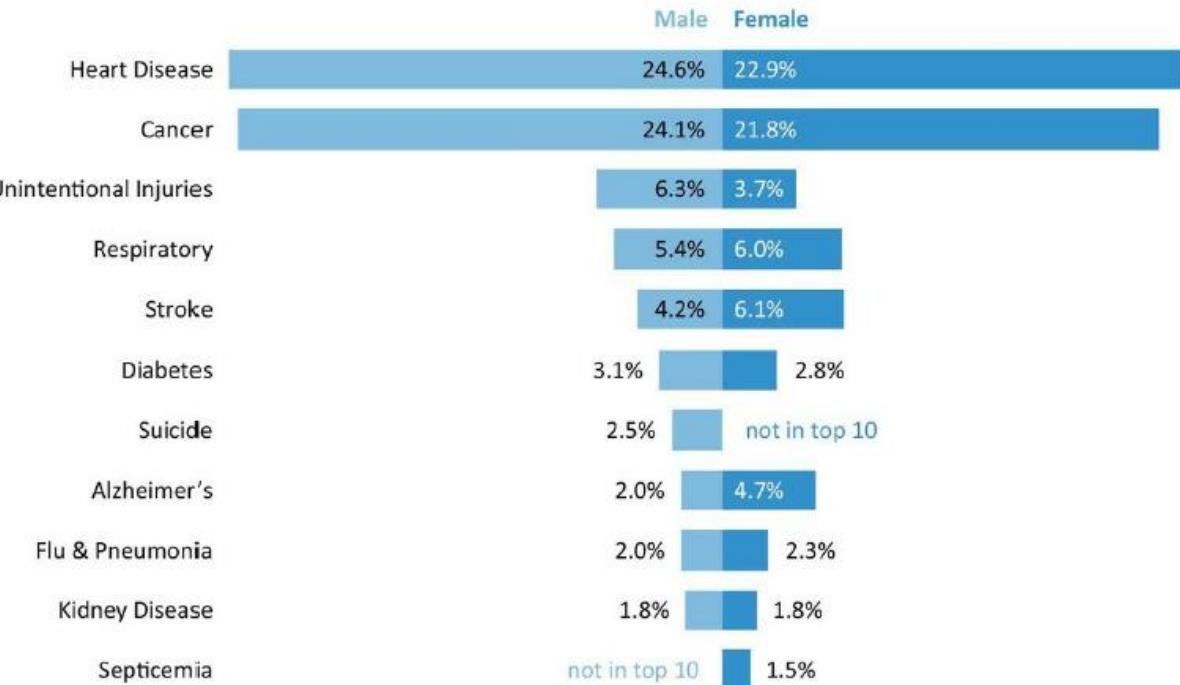


- It is a bit hard to see that sales of cheese are actually down at the new site from side-by-side bar charts (only one new site bar is shorter than its corresponding old site bar).
- Slopegraph will make the story clearer.

Back-To-Back Bars

- Two sets of bar graphs that share a spine
- The objective here is **NOT** to ask people to compare specific values from each bar graph. Such comparisons are difficult to make when the bars are back-to-back.
- These visualizations are best for showing the shape of the overall dataset. When **you are trying to show natural distributions or bell curves in the two datasets**, this is the graph type for you.

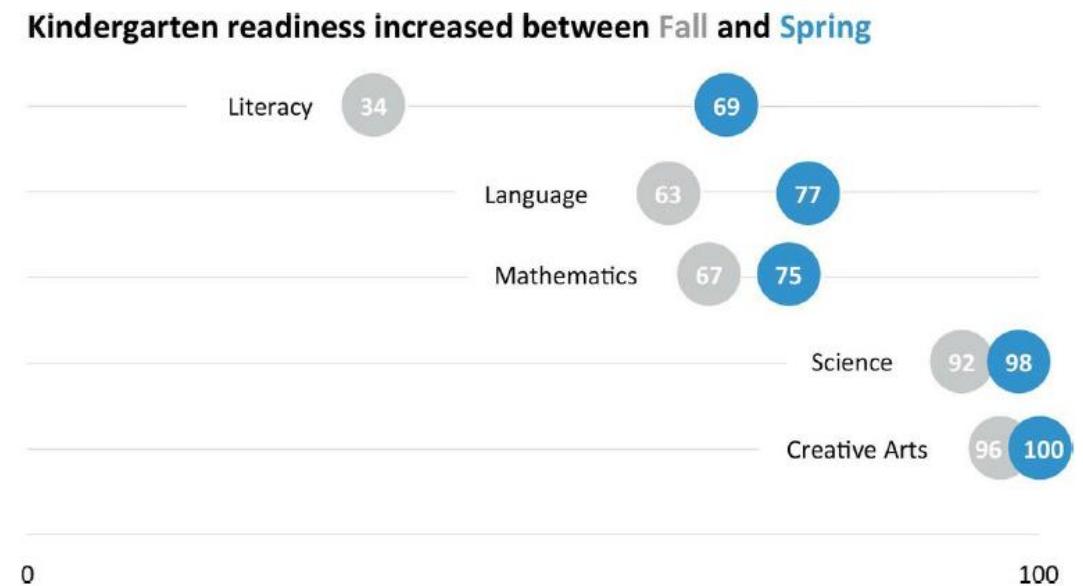
Leading causes of death for males and females are similar, with a few notable exceptions.



Dot plot

- Dot plots are flexible and easy to read.
 - Humans are accurately interpret location on line, when those lines share a common axis

	Fall	Spring
Literacy	34	69
Language	63	77
Mathematics	67	75
Science	92	98
Creative Arts	96	100

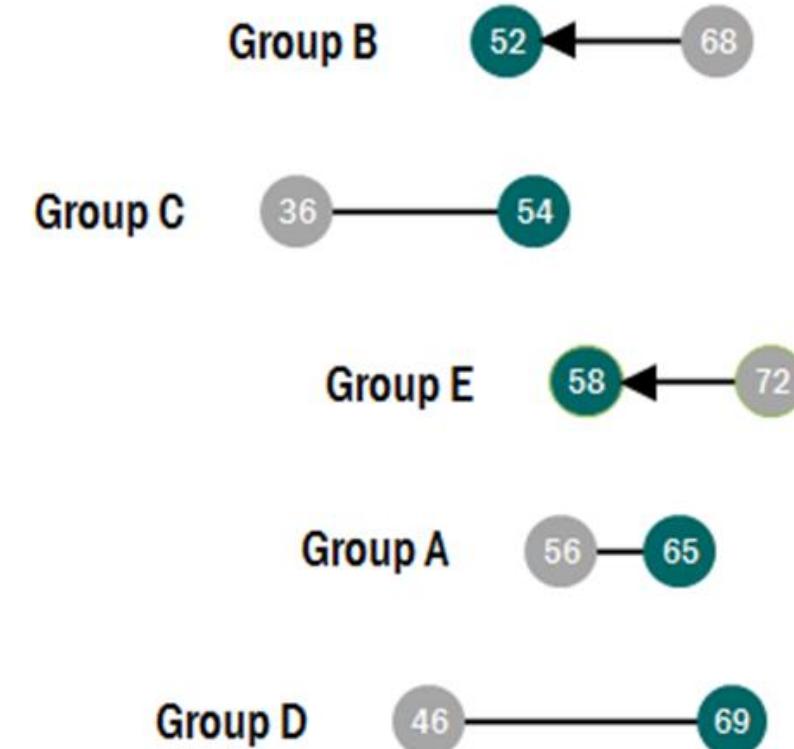


Dumbbell dot plot

- Compared with dot plot: the line that connects each pair of dots emphasizes their distance.
- Good at telling a story about the gap or growth between two things.

Groups C, A, & D all made significant increases between pretest & posttest.

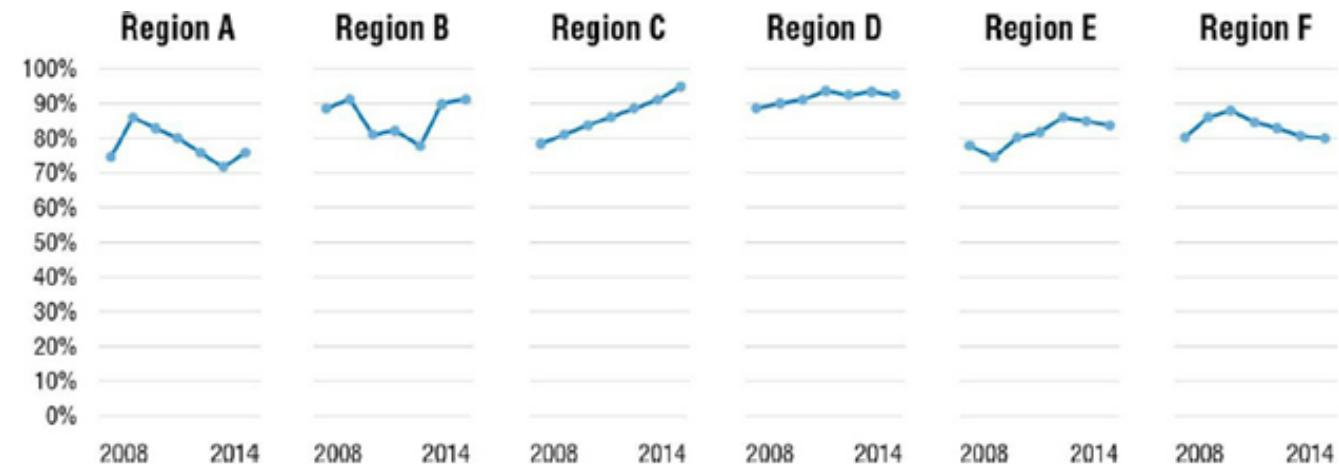
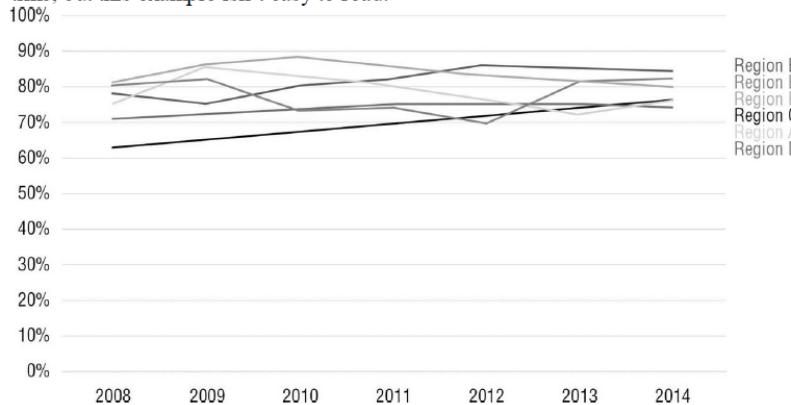
Groups A & D ended with the highest posttest scores.



Small Multiples

- Trying to compare too many numbers at once can quickly become overwhelming and meaningless.
- When we are trying to compare multiple categories, but the graph gets complicated, one good solution is to break it apart into small multiples.
- Small multiples are several graphs, all on the same scale, arranged together.

Figure 3.45 The line graph may be the most traditional way to compare trends over time, but this example isn't easy to read.



Choosing the Right Chart

- When a single number is important
- How two or more numbers are alike or different
- **How we are better or worse than a benchmark**
- When there are parts of a whole
- How things change over time

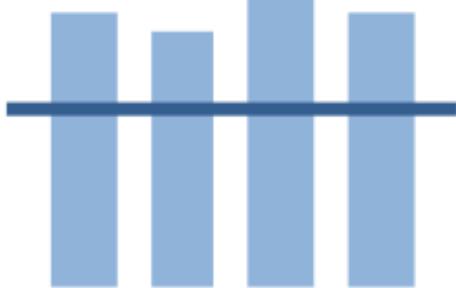
How we are better or worse than a benchmark: displaying relative performance

- Help a reader to determine whether performance was good, bad or close to the benchmark.
- Stories
 - Our key indicators met the pre-established targets in three out of seven areas
 - Regions A and B did not meet quarterly benchmarks
 - Here is how our groups compared to the national norm
 - Students in the Chemistry Department are above average on final exams this year
 - We did not meet our fundraising goal, but we got very close
 - Our click rate was 25.6% while the industry standard is 4.3%
 - ...

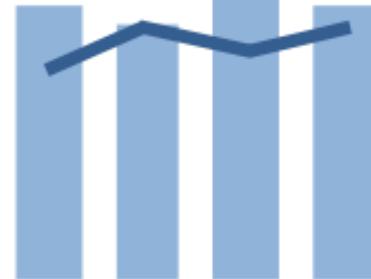
Displaying Relative Performance

- Benchmark Line
- Combo chart
- Bullet Chart
- Indicator Dots

Benchmark Line



Combo



Bullet Chart



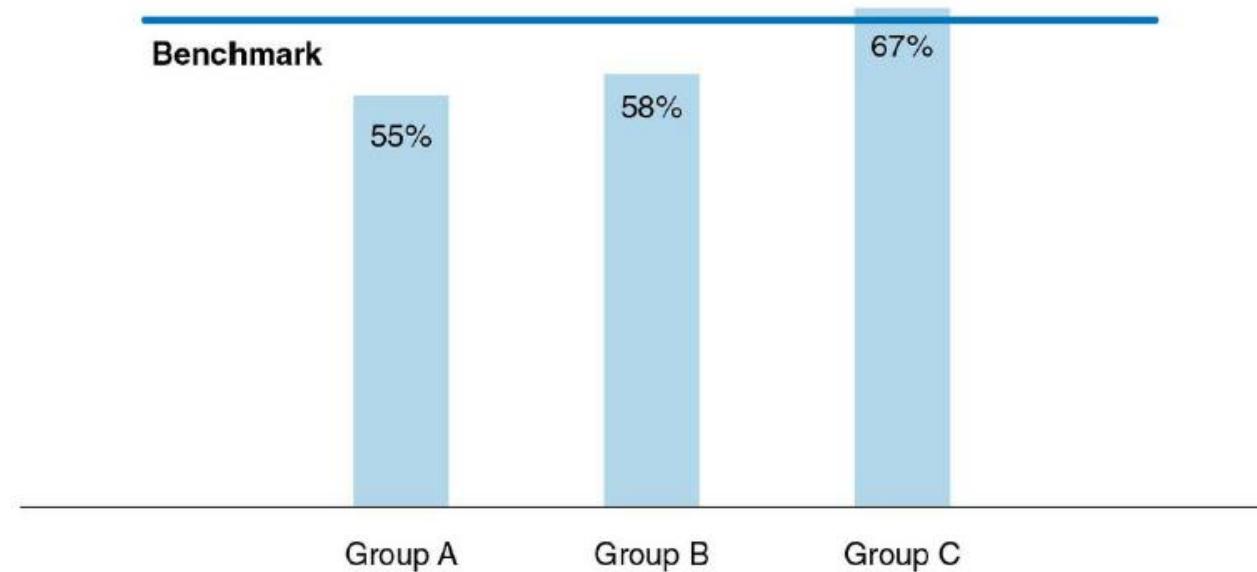
Indicator Dots



Benchmark Line

- Adding a benchmark line to a graph gives loads of context for the viewer.
- This simple line packs so much power.

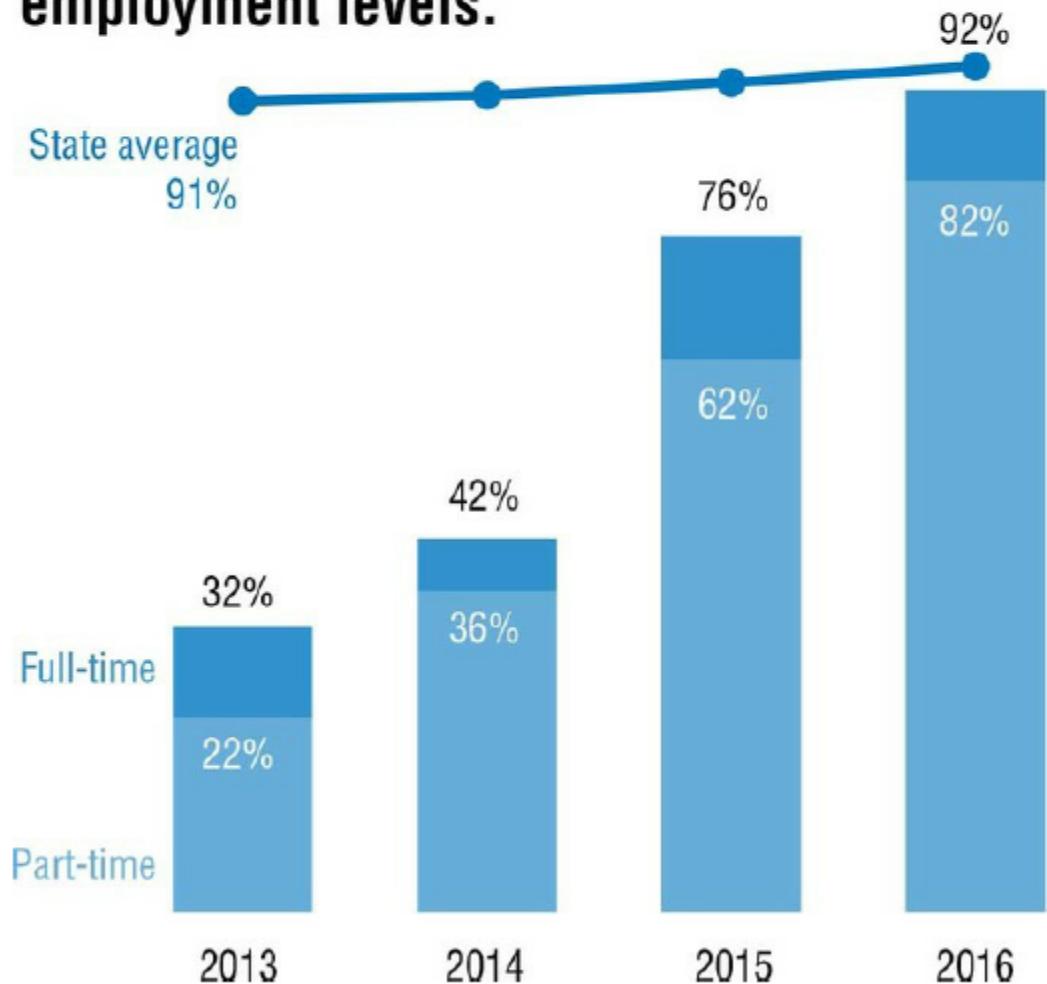
Two of the three grantee groups did not meet performance benchmarks this year.



Combo Chart

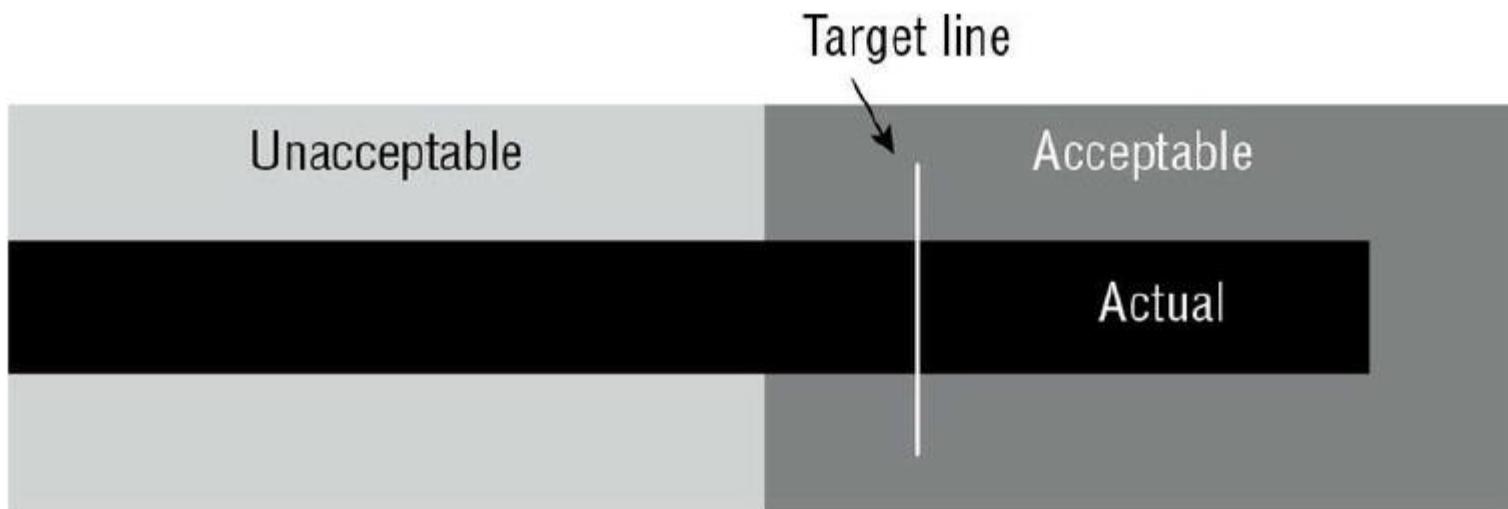
- Benchmark line works fine when the benchmark is the same for every group. But, what about when relative performance is different for each group?
- Combo Chart

Program participants gained employment during their work with us and are nearly meeting average state employment levels.

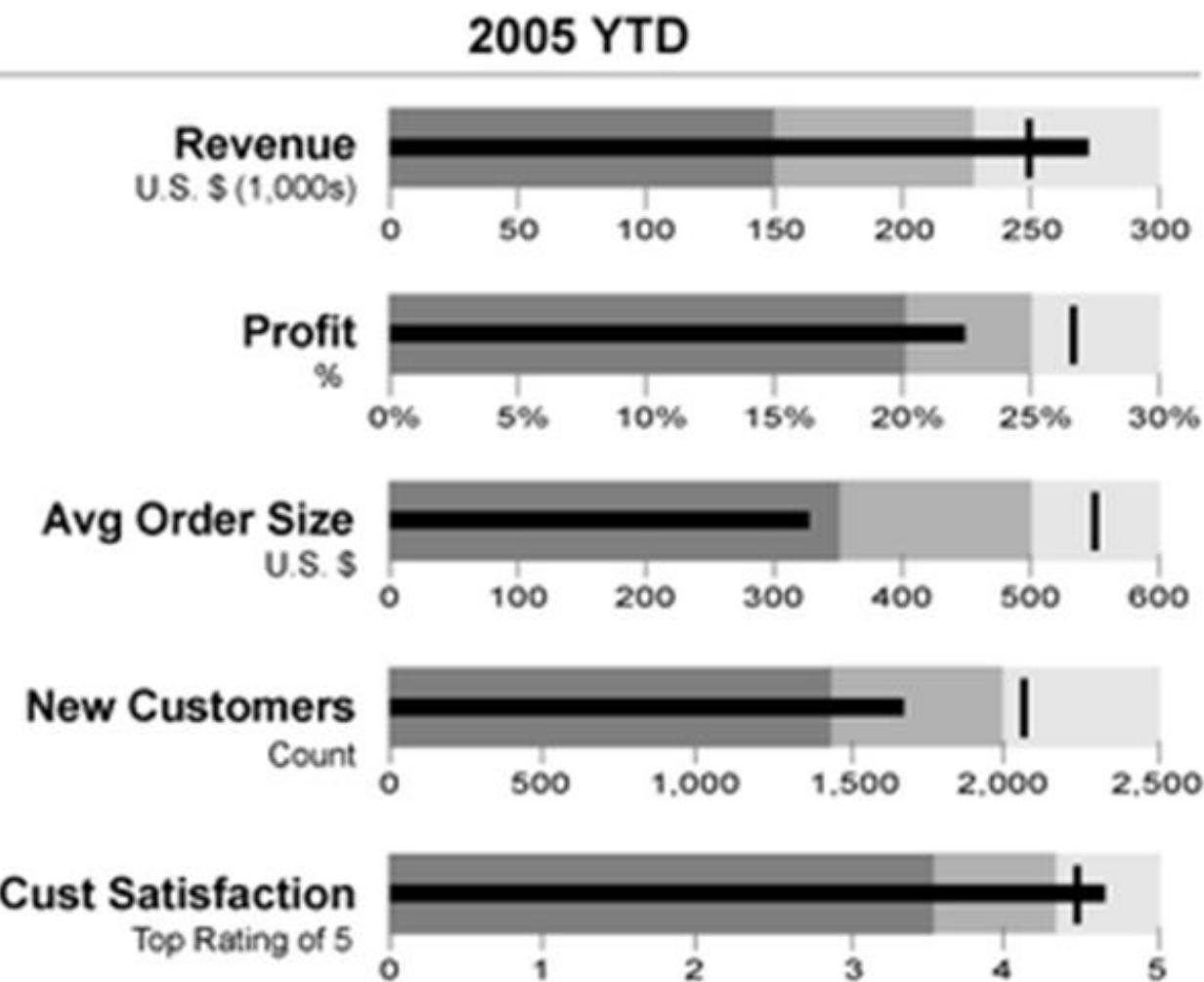


Bullet Graph

- A bullet graph is a variation of a bar graph.
- It has several key design elements
 - The dark center line represents actual value
 - The white vertical line represents target value
 - The colored bands represents ranges (e.g., unacceptable and acceptable)



Bullet Graph



- The dark center line represents the actual value.
- The dark vertical line represents a target value.
- The colored bands represent ranges, such as poor, average, and good.

Indicator Dots

- Indicator dots are little colored markers that show up next to a target value listed in a table.
- They indicate whether or not the target has been met.

A	B	C	D	E	F	G	H	I	J
1	Student ID	Homework1	Homework2	First exam	Homework3	Group Project	Homework4	Average to date	
2	1	67	75	78	85	48	62	69.2	
3	2	86	75	57	63	48	87	69.3	
4	3	46	36	52	46	48	47	45.8	
5	4	53	62	35	47	48	75	53.3	
6	5	57	66	73	75	74	78	70.5	
7	6	88	87	86	80	74	85	83.3	
8	7	76	74	67	66	74	60	69.5	
9	8	97	96	99	96	74	94	92.7	
10	9	85	85	69	75	78	68	76.7	
11	10	32	37	33	36	78	44	43.3	
12	11	78	74	72	77	78	73	75.3	
13	12	45	47	64	66	78	75	62.5	
14	13	34	36	32	38	37	48	37.5	
15	14	64	62	74	61	37	59	59.5	

Show students who are failing (score under 50)

Show students who are on the borderline (score between 50 and 60)

Choosing the Right Chart

- When a single number is important
- How two or more numbers are alike or different
- How we are better or worse than a benchmark
- When there are parts of a whole
- How things change over time

When There Are Parts of a Whole: Visualizing Beyond the Pie Chart

- Stories:
 - Most students persisted to graduation, but some dropped out and some had incomplete attendance
 - Our employees are mainly female
 - Our newest phone already has 5% of the market share
 - The student body composition is 75% Free and Reduced Lunch and 25% paid lunch
 - This map shows the percentage of our users who come from each state in the United States
 - The majority of our visitors are ages 10 and younger and 56 and older
 - Market data show we have three main profiles of customers
 - ...

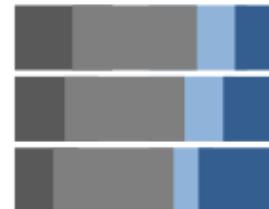
Visualize The Parts of a Whole

- Pie/Donut
- Stacked Bar
- Histogram
- Tree Map
- Map

Pie/Donut



Stacked Bar



Histogram



Tree Map

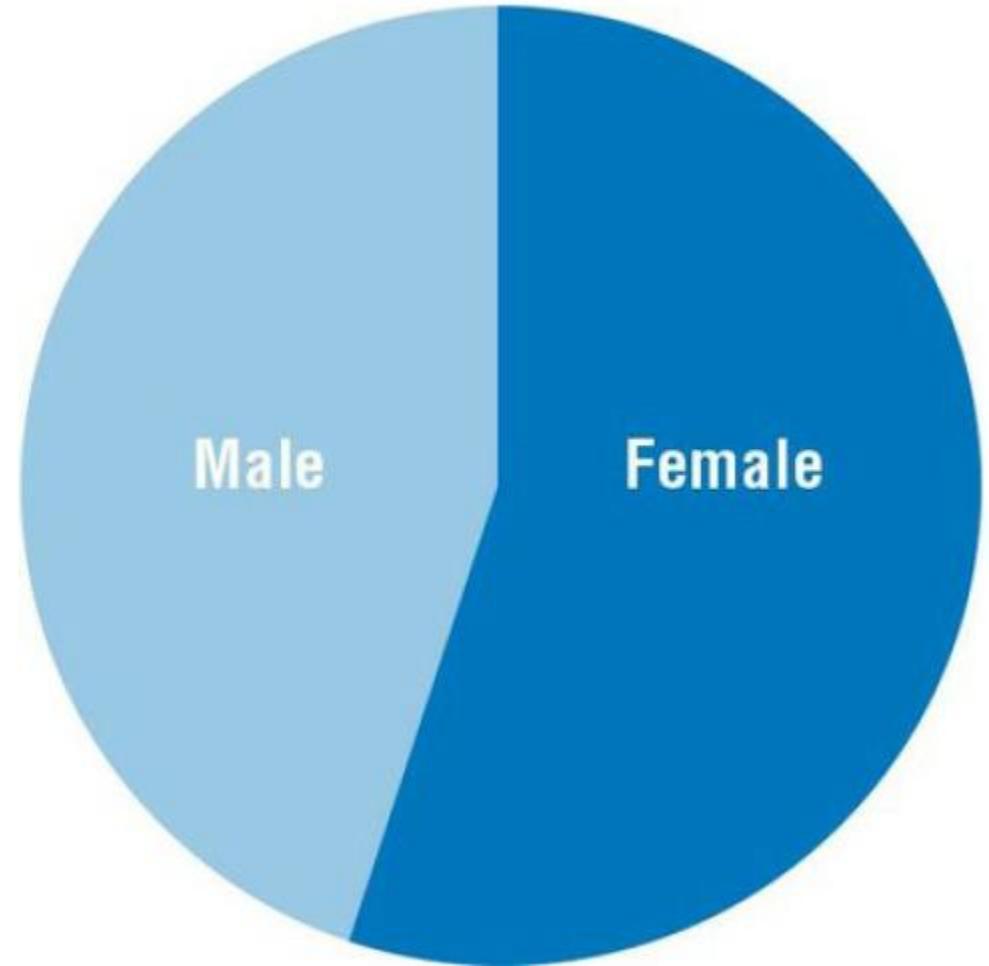


Map

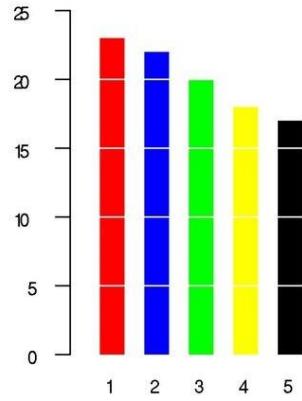
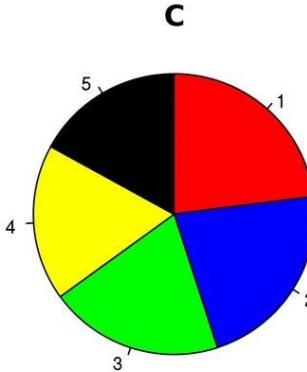
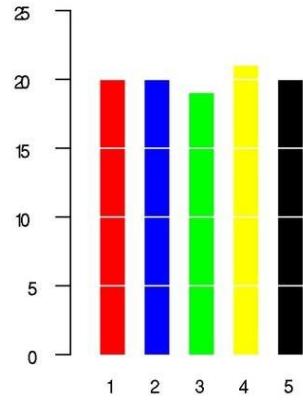
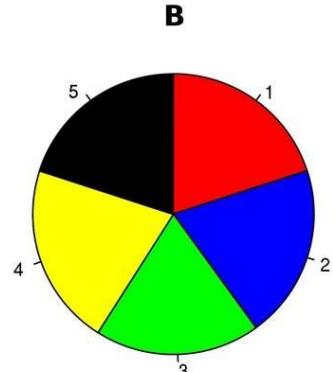
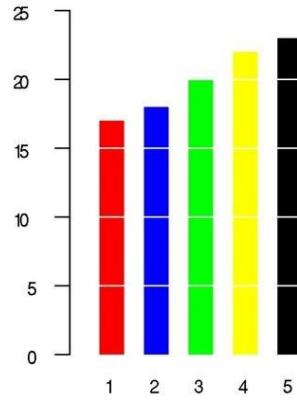
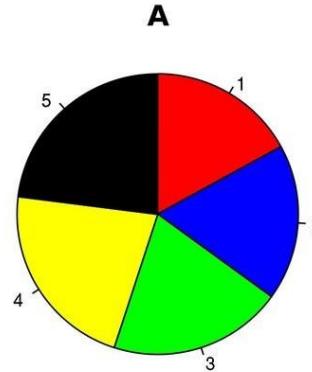


Pie Chart

- Typically, we show parts of a whole as pie chart.
- Pie charts work best when
 - Roughly four or fewer wedges
 - Wedges are really different (or really similar) to another



Where is my Pie?



1. Don't use pie charts if your data involves time.
2. Don't use pie charts if you have too many slices.
3. Don't use too many labels.
4. Keep the color-scheme flat.
5. Two-dimensional is better.
6. Use six categories or fewer.
7. Data distribution matters.

<https://www.idashboards.com/blog/2017/03/14/have-your-pie-eat-it-too-when-to-use-pie-charts/>

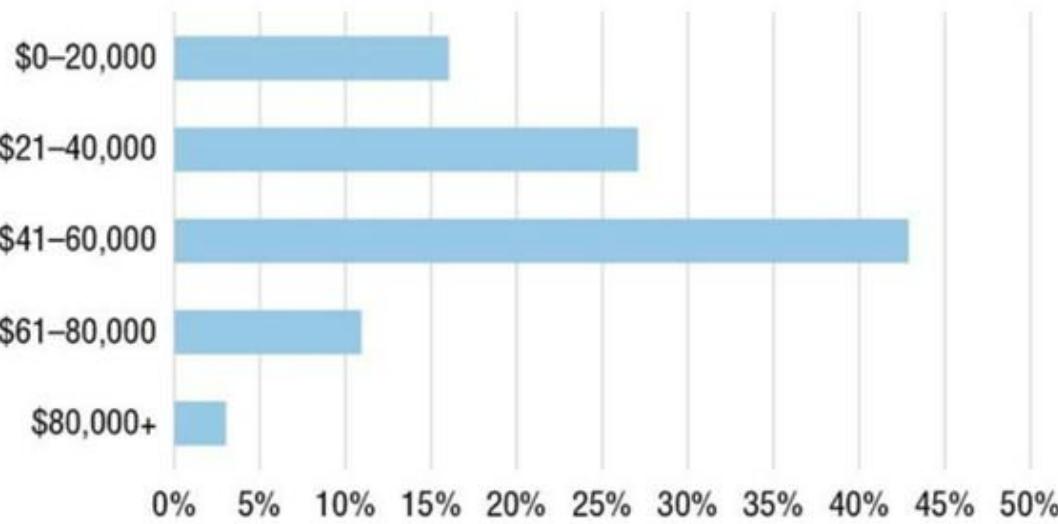
<https://www.linkedin.com/pulse/eat-pies-dont-share-them-avinash-kaushik>

<https://www.perceptualedge.com/articles/visual-business-intelligence/save-the-pies-for-dessert.pdf>

100% Stacked Bar

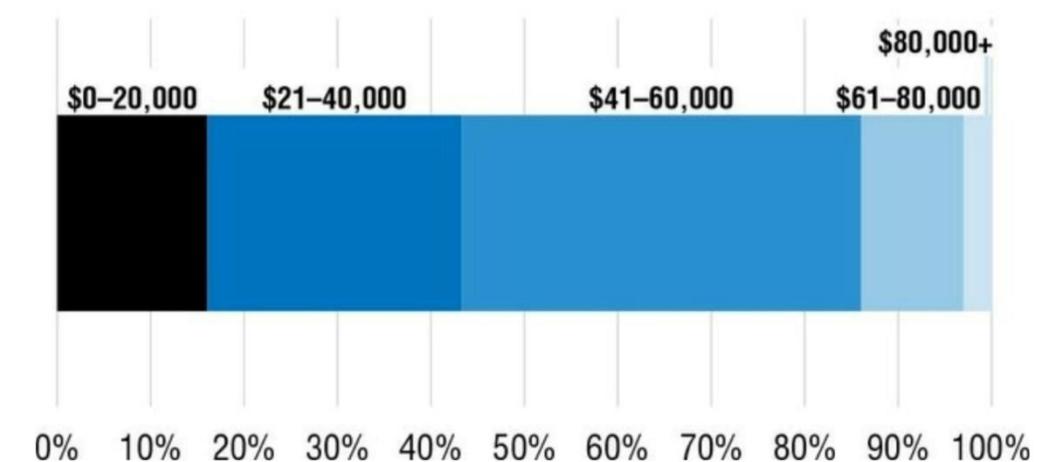
- Swap the pie for bar chart if there are more than four categories in your dataset

Few campus salaries exceeded \$61,000 per year.



Bar Chart

Few campus salaries exceeded \$61,000 per year.

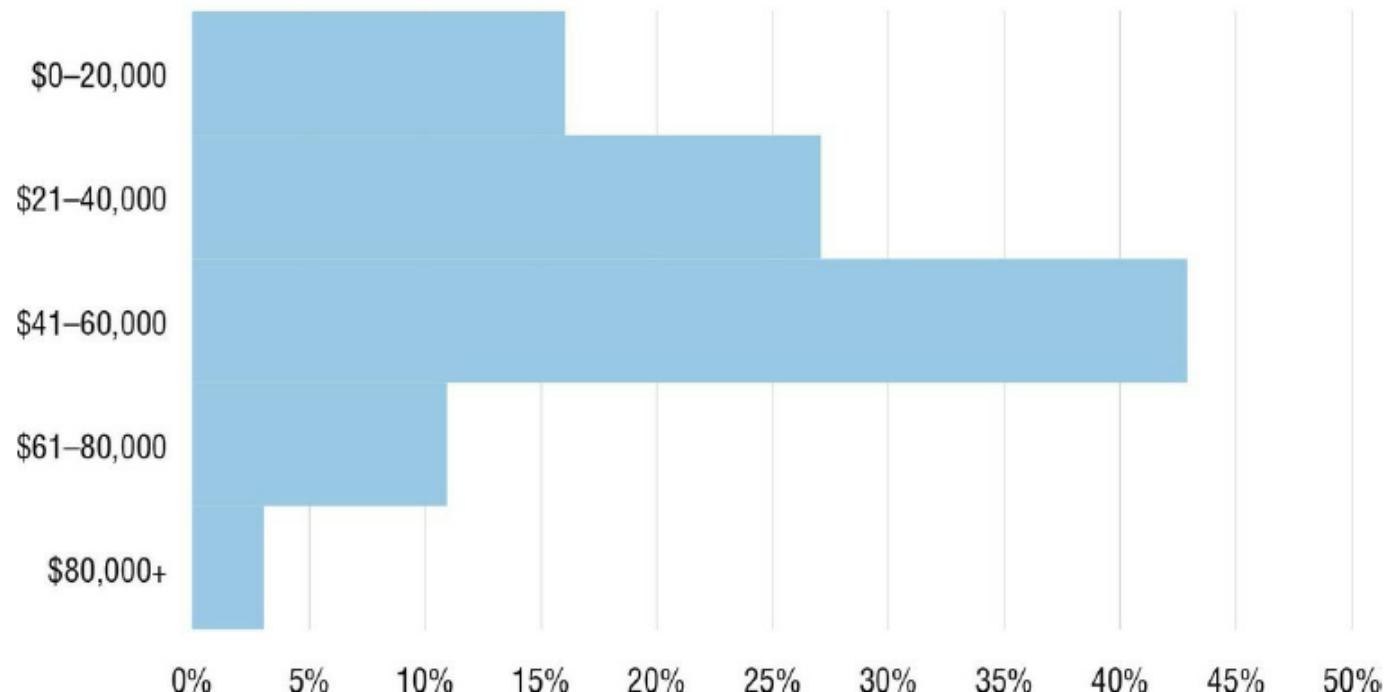


100% Stacked Chart

Histogram

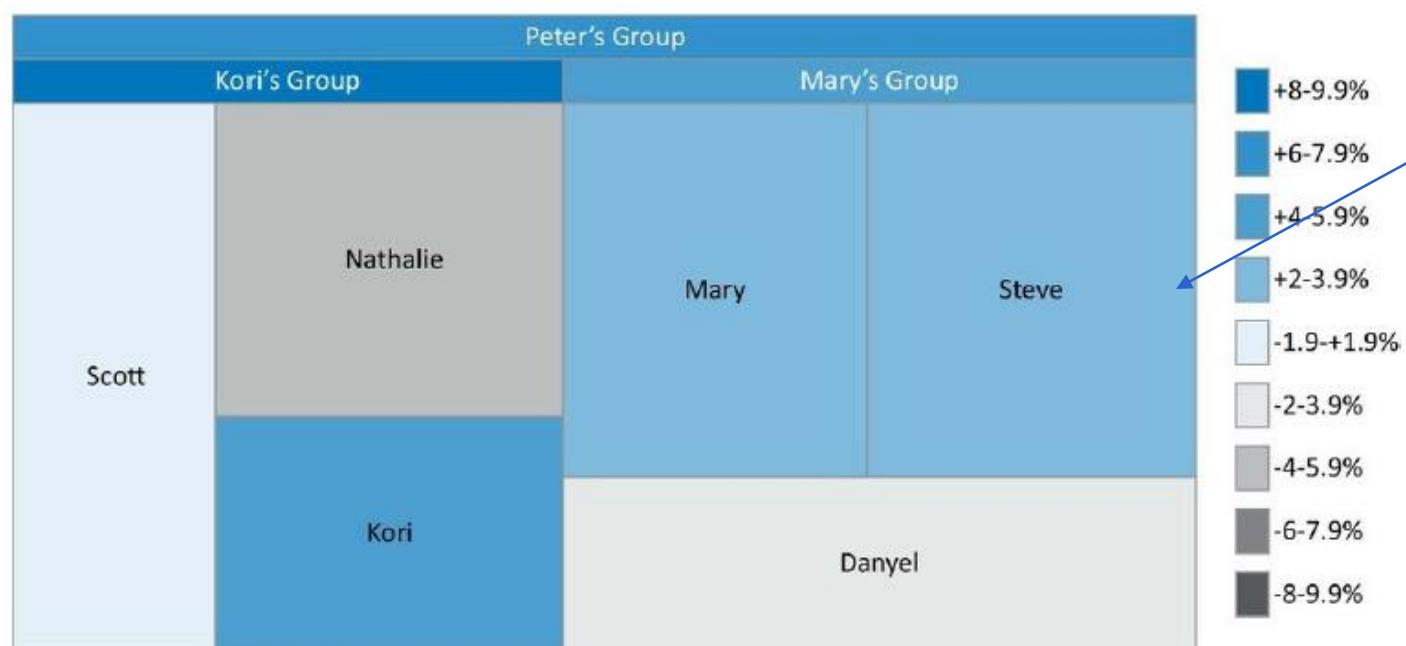
- A histogram is nothing more than a standard bar or column chart, with absolutely no space between the bars or columns.
- It gives a good picture of overall shape of the dataset.

Few campus salaries exceeded \$61,000 per year.



Treemap

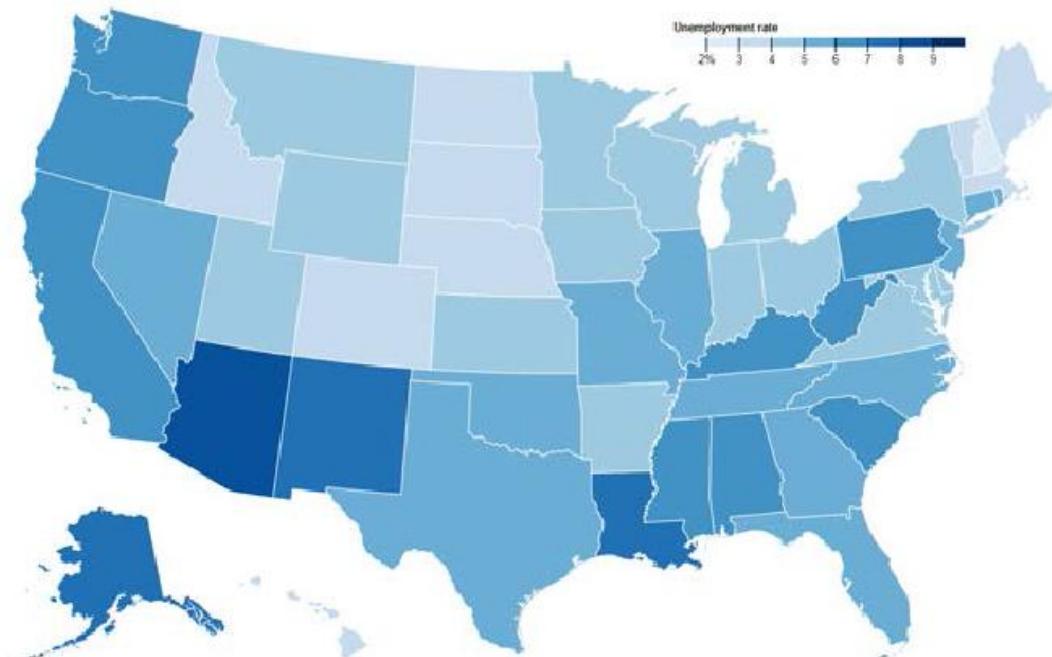
- Treemaps are square or rectangle shapes that represent parts, all positioned inside a larger square or rectangle that represents a whole.
- Treemaps can visualize hierarchy



- Peter oversees Kori and Mary, and that Kori and Mary both oversee two other employees.
- The sizes of the boxes represent sales. So that Mary's group has higher sales than Kori's group
- Color represents the growth in sales over the previous year

Maps

- Increasingly, our social sciences understand the importance of geolocation when studying outcomes. How respondents or constituents are spread over location can reveal patterns and insights that can't be seen any other way.



Choosing the Right Chart

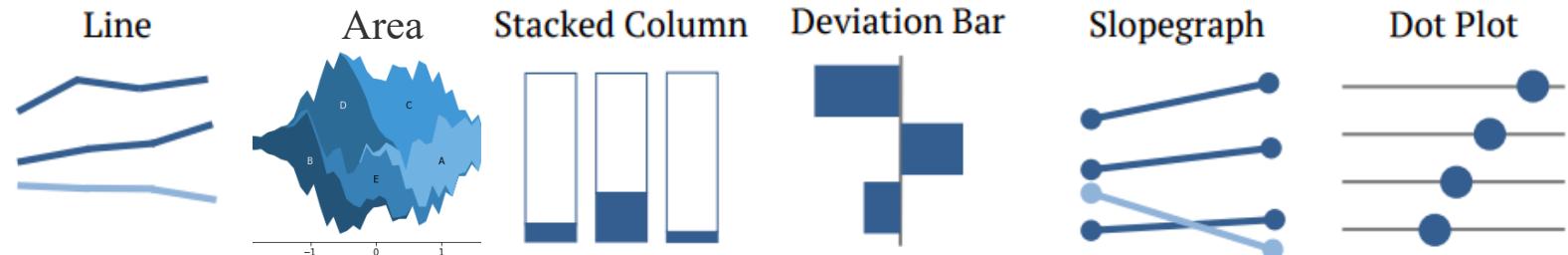
- When a single number is important
- How two or more numbers are alike or different
- How we are better or worse than a benchmark
- When there are parts of a whole
- How things change over time

How things changed over time: depicting trends

- Stories:
 - Things changed/didn't change
 - After is so much better/worse/the same as before
 - We started this intervention and the outcome improved as a result
 - When the legislation took effect, we saw peaks in services for several years
 - Sales increased 0.5% over the last quarter
 - Maternal services have steadily increased their proportion of hospital use
 - Here's how much change occurred on this measure in the last decade

Visualize How Things Changed Over Time

- Line Graph
- Area Graph
- Stacked Column
- Deviation Bar
- Slopegraph
- Dot Plot

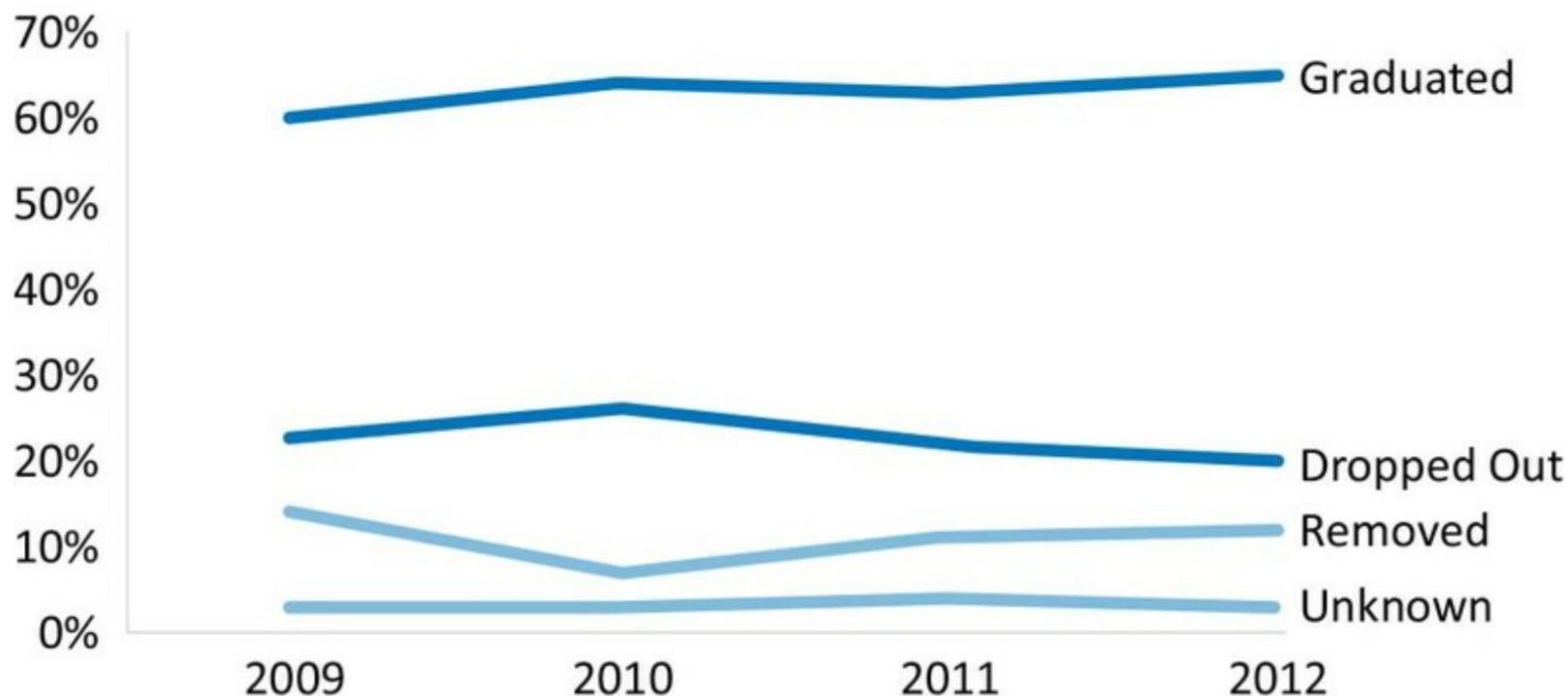


Line Graph

- Line Graphs are the most common way to depict trends over time

Graduation rates are slowly increasing.

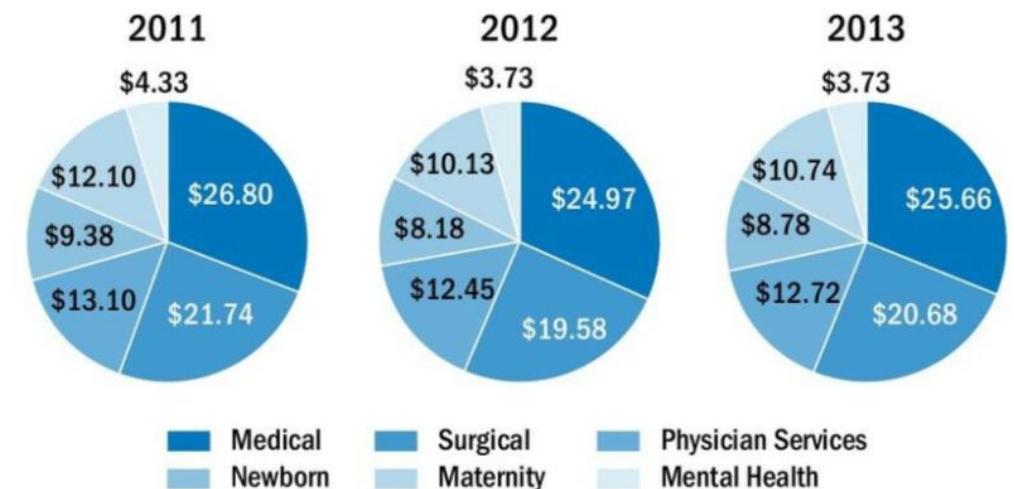
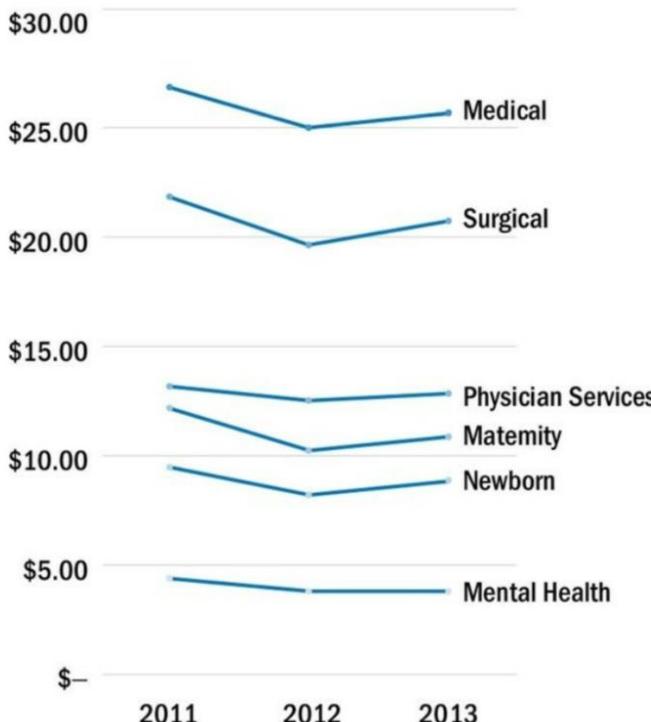
We are up 2% since last year.



Area Graph

- Objective: to show trends that make up 100%

Overall, inpatient costs have decreased since 2011.



Line graph can show the trends but can not represent parts of a whole.

Pie chart can show parts of a whole but can not show the trends over time.

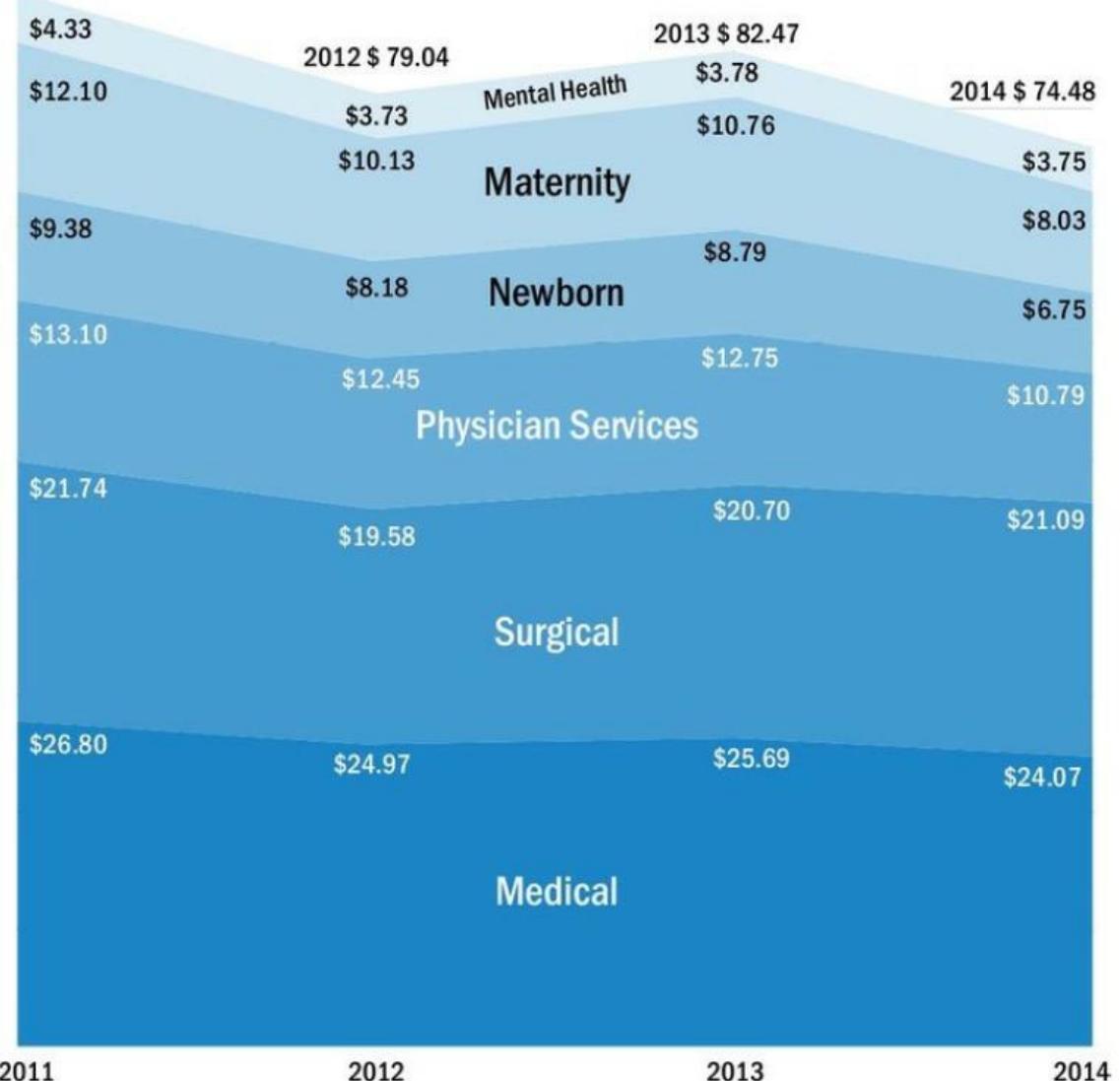
Area Graph

- Objective: to show trends that make up 100%
- Area Graph
 - A line graph with each segment stacked on top of one another.
 - Can show both parts of whole and trends.

Overall, inpatient costs have decreased since 2011.

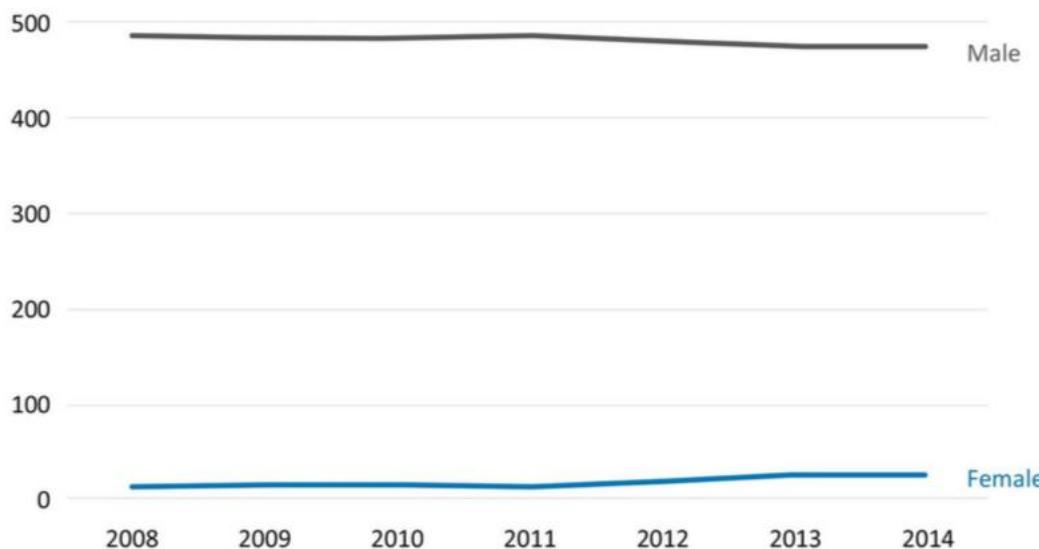
Figures are in U.S. dollars per member, per month

2011 total: \$ 87.45

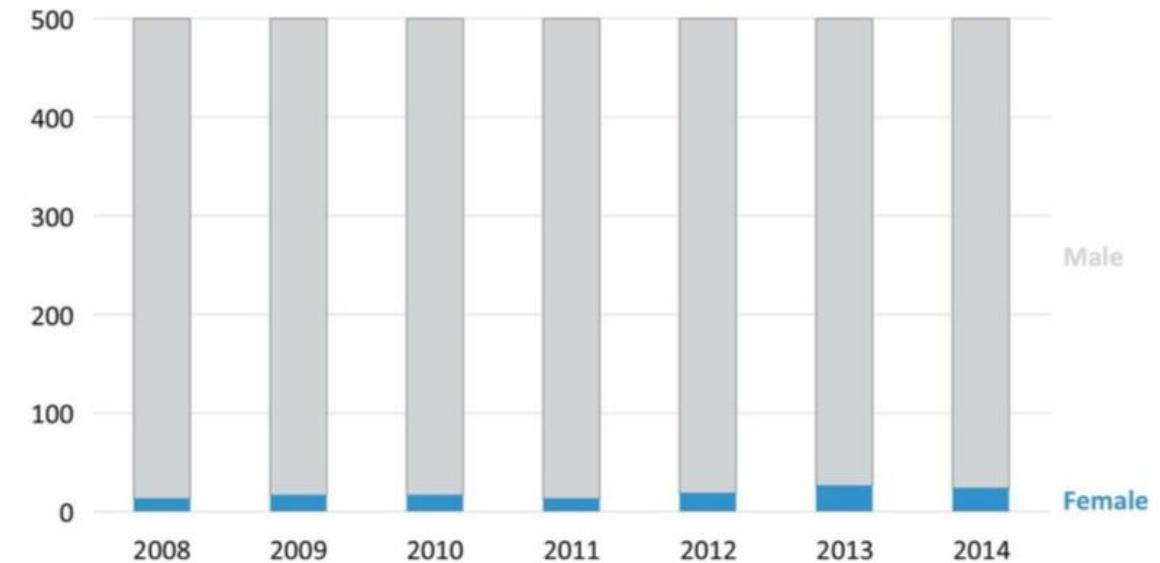


Stacked Columns

While increasing slightly over the years, female CEOs for Fortune 500s are still few in number.



While increasing slightly over the years, female CEOs for Fortune 500s are still few in number.

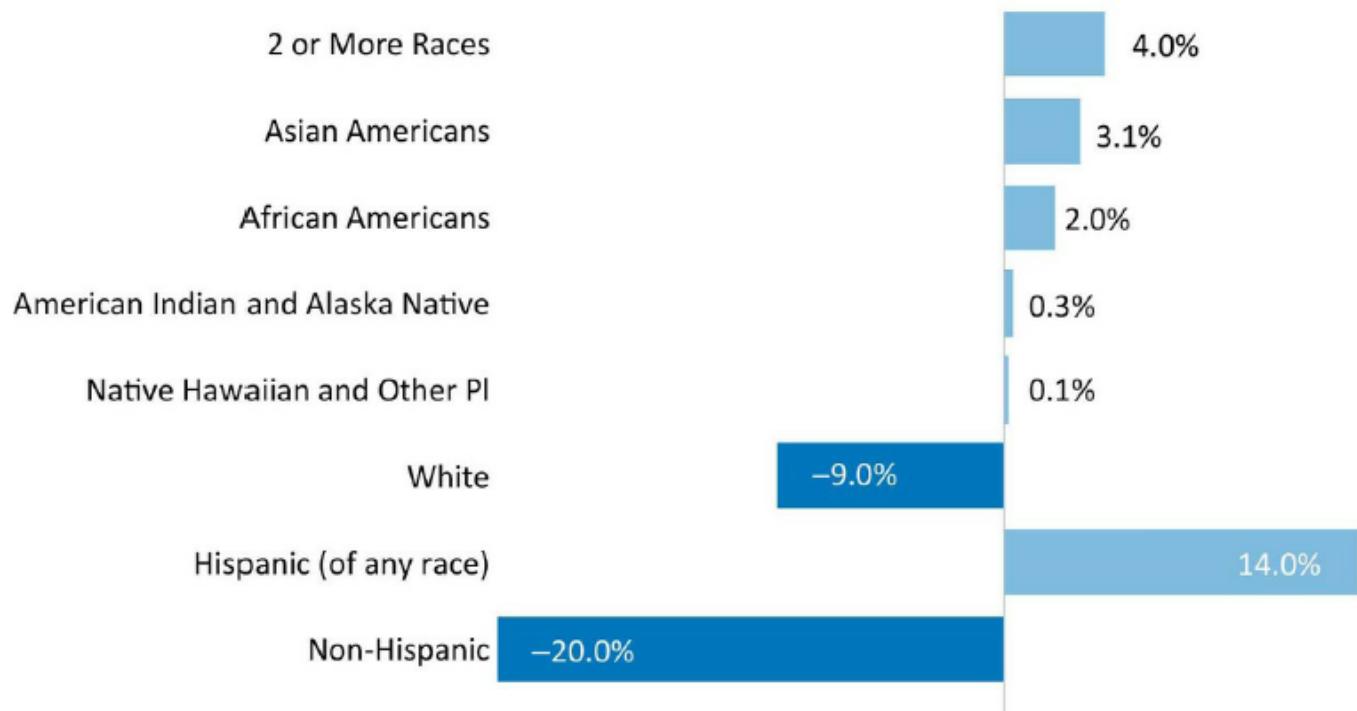


Categories are parts of a whole AND one category is really tiny

Deviation Graph

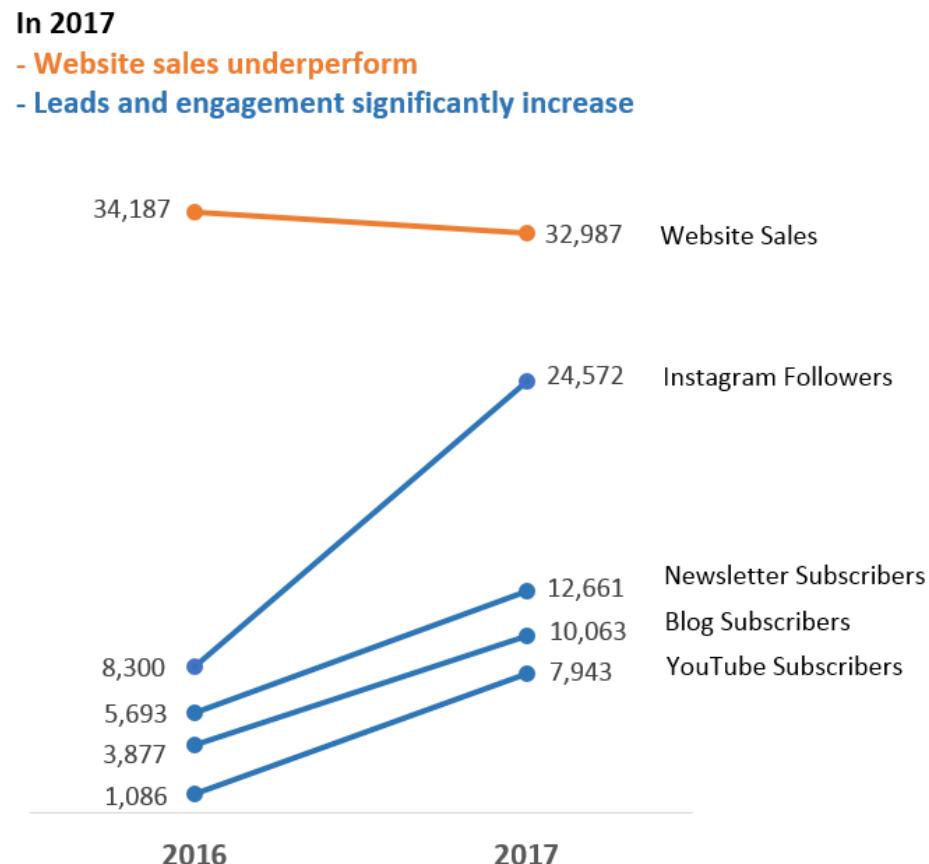
- Situation: focus on the **how much change** has happened since the beginning

From 2012 to 2060, the US Census predicts increases for all races except White.



Slope Graphs

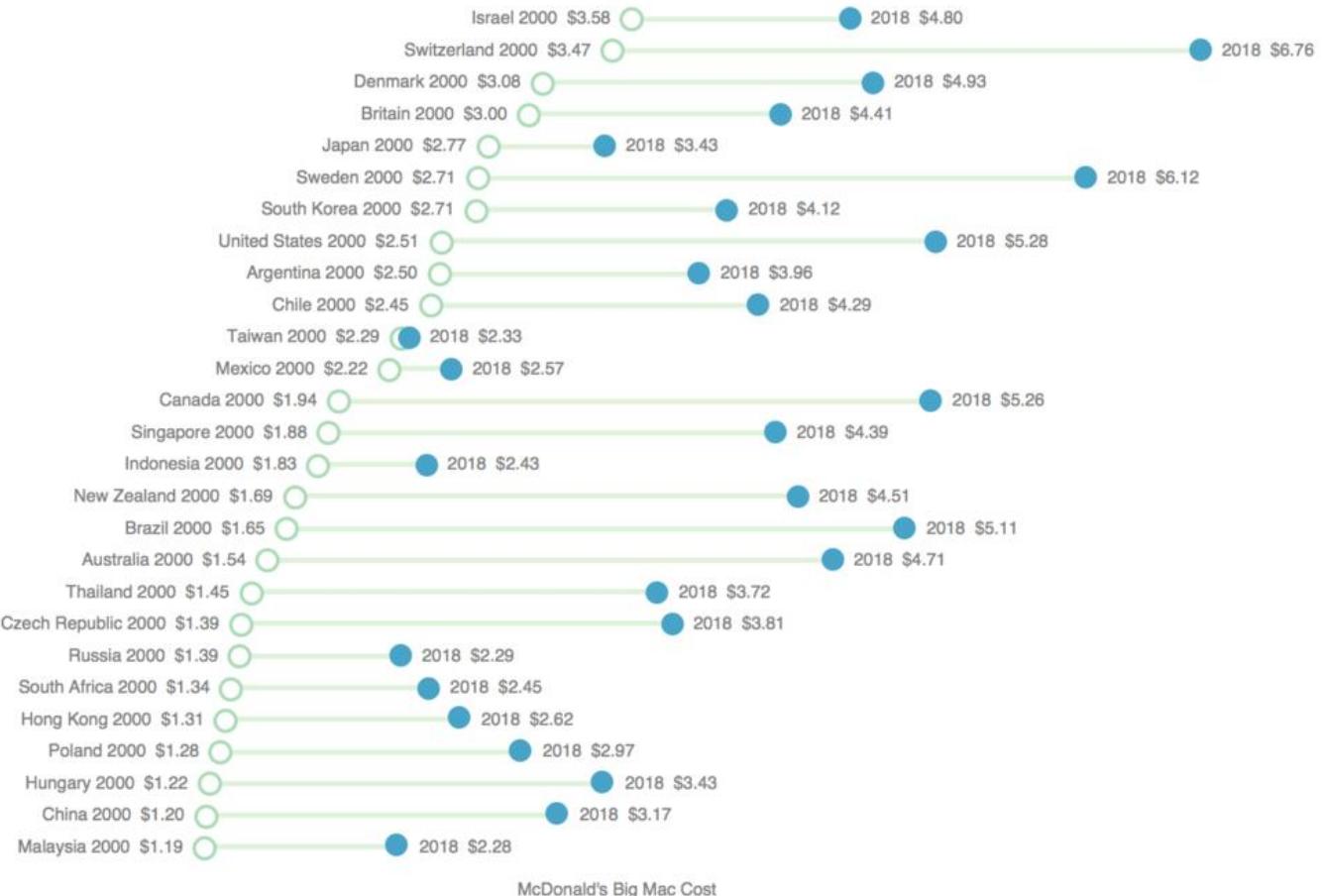
- Slopegraphs can also be used to show change over time.



Dot Plots

- Dot Plots can be used to show change over time as well

From 2000-2018 the cost of a Big Mac has increased at a different rate for 27 tracked nations
The Big Mac Index is described as a "lighthearted guide to whether currencies are at their correct level." This graph does not focus on the currencies but on the price change (in USA dollars) of a Big Mac from 2000 to 2018.



SOURCE: <https://www.economist.com/news/2018/07/11/the-big-mac-index>

NOTE: Prices will vary depending on where you live, and this graph may not depict the exact price of a Big Mac at your local McDonalds.

Outline

- **Concepts of Data Visualization**
 - Value, goals
 - History
- **Design of Visualization**
 - Workflow of data visualization
 - Color, Size, Text, Titles, Labels
 - Choosing the right chart
- **Case study: COVID-19**
 - Mask wearing trends
 - People concerned about the impact of COVID-19

Case study - COVID 19

- What are the mask wearing trends?
 - This data viz shows the prevalence of mask wearing in different scenarios: while grocery shopping, while exercising outside, while visiting with family and friends, and while working at the workplace. See what behaviors have changed over time as vaccinations have rolled out and restrictions lifted.
- How worried are people about the impact of COVID-19?
 - As we enter into a "new normal", how worried are people about the impact of COVID-19 on personal finances and the economy? How does this break down by party affiliation, race, education, gender and age?

Outbreaks Near Me | Momentive Data:



Where the world comes for answers

How likely are you to wear a mask when...
(data from the last week)

Industry
All
Age
Race
Gender
Education
Income
Party
Vaccinated?
Which Vaccine?
Industry
Transportation & Delivery
Government
Finance & Financial Services
Construction

Data from 937,577 respondents since June 1, 2020

How likely are you to wear a mask when...
(data from the last week)

Very likely
Somewhat likely
Not so likely
Not likely at all

Industry



78% of Healthcare say they are
Very likely to wear a mask
when At Workplace

Party
Vaccinated?
Which Vaccine?
Industry
Transportation & Delivery
Government
Finance & Financial Services
Construction

Trend:

100%

80%

Select a demographic to
see trends over time

Select a date range:

6/1/2020

8/8/2021

60%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%

0%

100%

80%

60%

40%

20%</

Percent of people in the U.S. who are very worried...

Apply filters here >> ≡

the coronavirus outbreak will have a negative economic effect on the global economy

N = 2,561,090

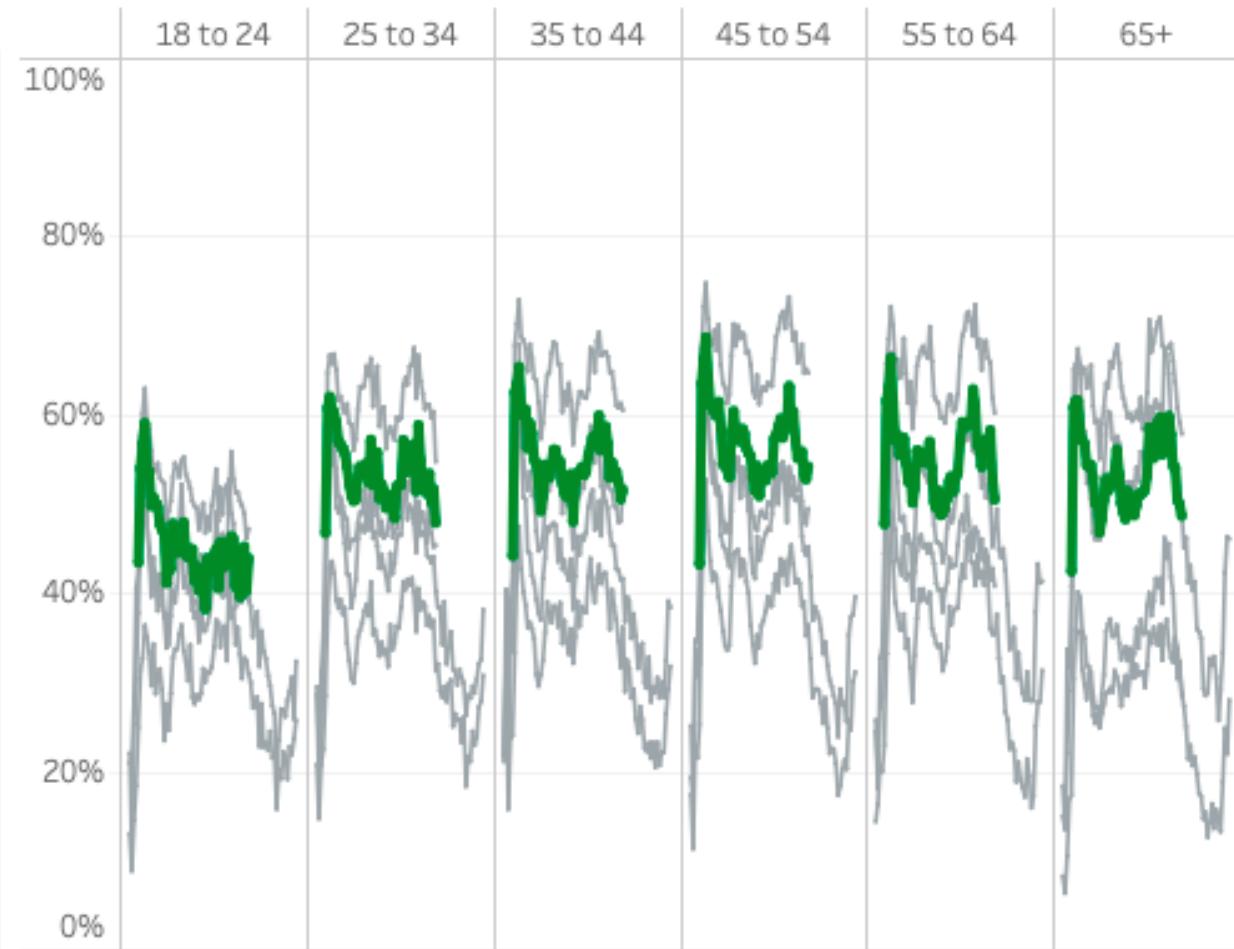
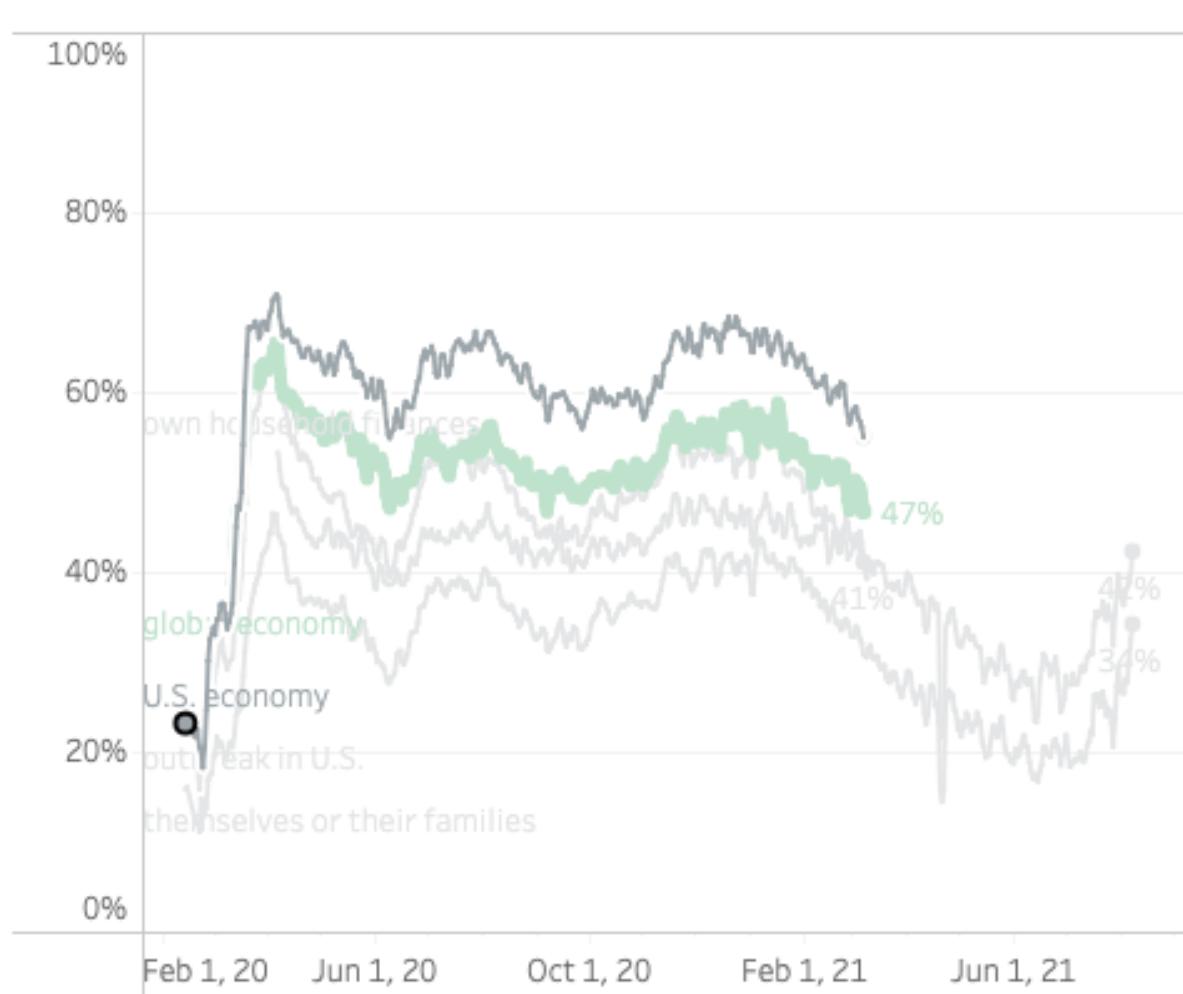
Degree of worry

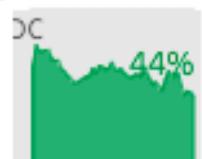
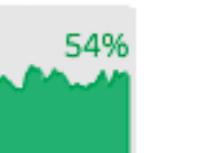
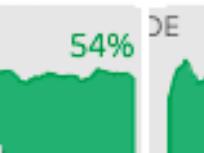
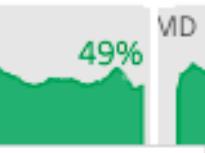
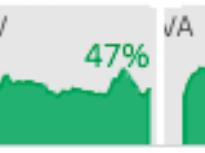
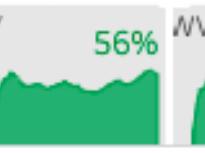
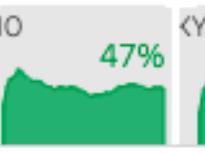
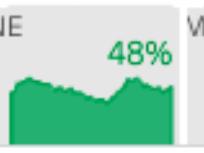
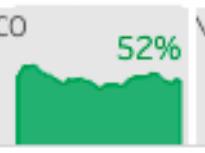
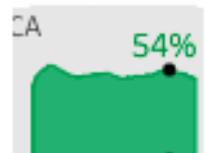
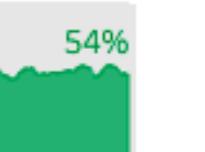
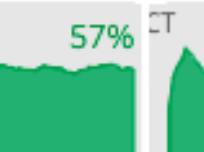
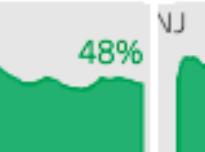
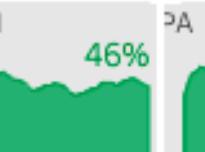
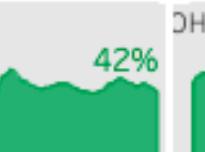
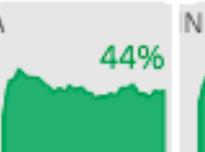
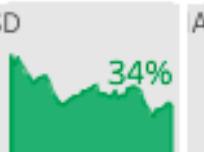
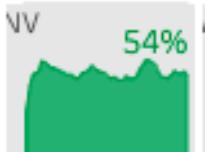
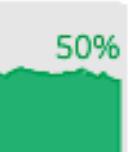
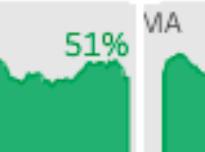
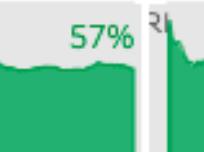
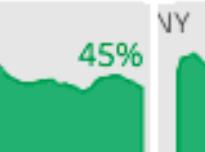
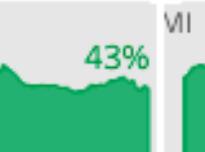
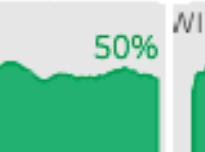
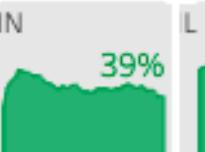
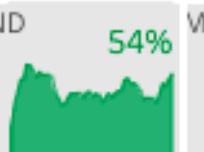
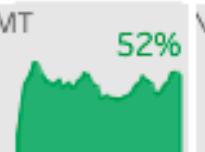
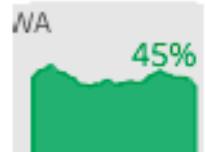
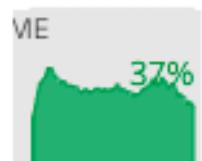
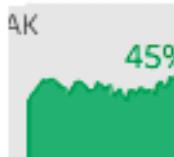
very worried

Breakdown by

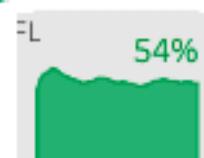
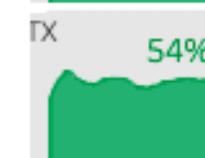
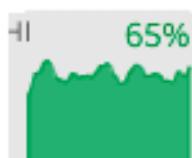
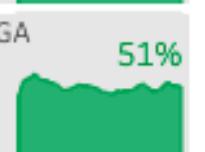
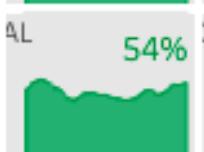
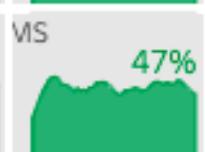
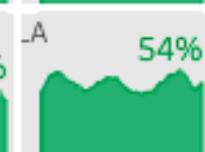
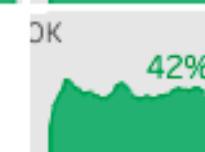
Age

Includes polls from: Tuesday, February 11, 2020 to Sunday, August 8, 2021





Monthly rolling average for California in the week beginning January 4, 2021: 59%



Dashboard layout

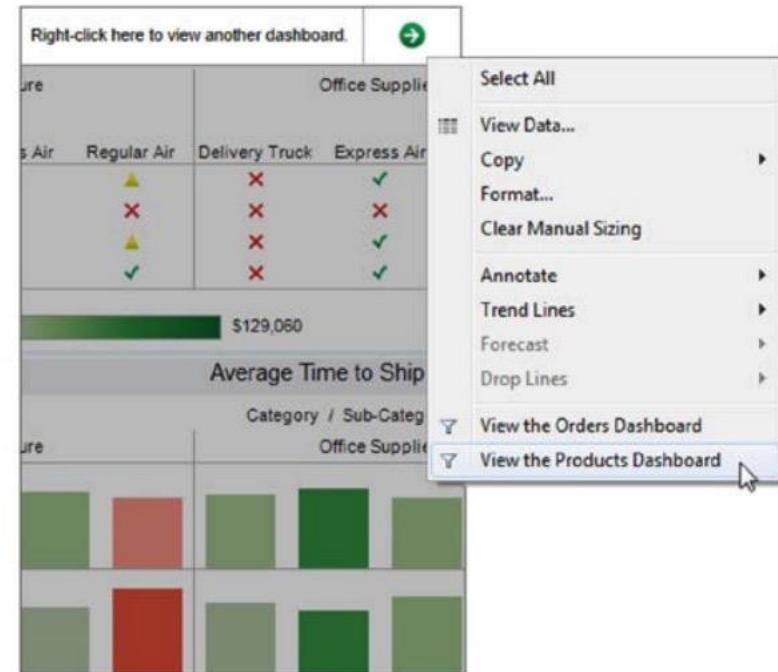
Your dashboard's purpose is to help guide the reader's eye through more than one visualization, **tell the story of each insight**, and reveal how they're **connected**.

The more you employ better dashboard design, your users will discover what's happening, why and what's most important. Take into account how you're guiding their eyes across the dashboard.

Guide the user

Don't leave people high and dry without guidance on how to use a visualization.

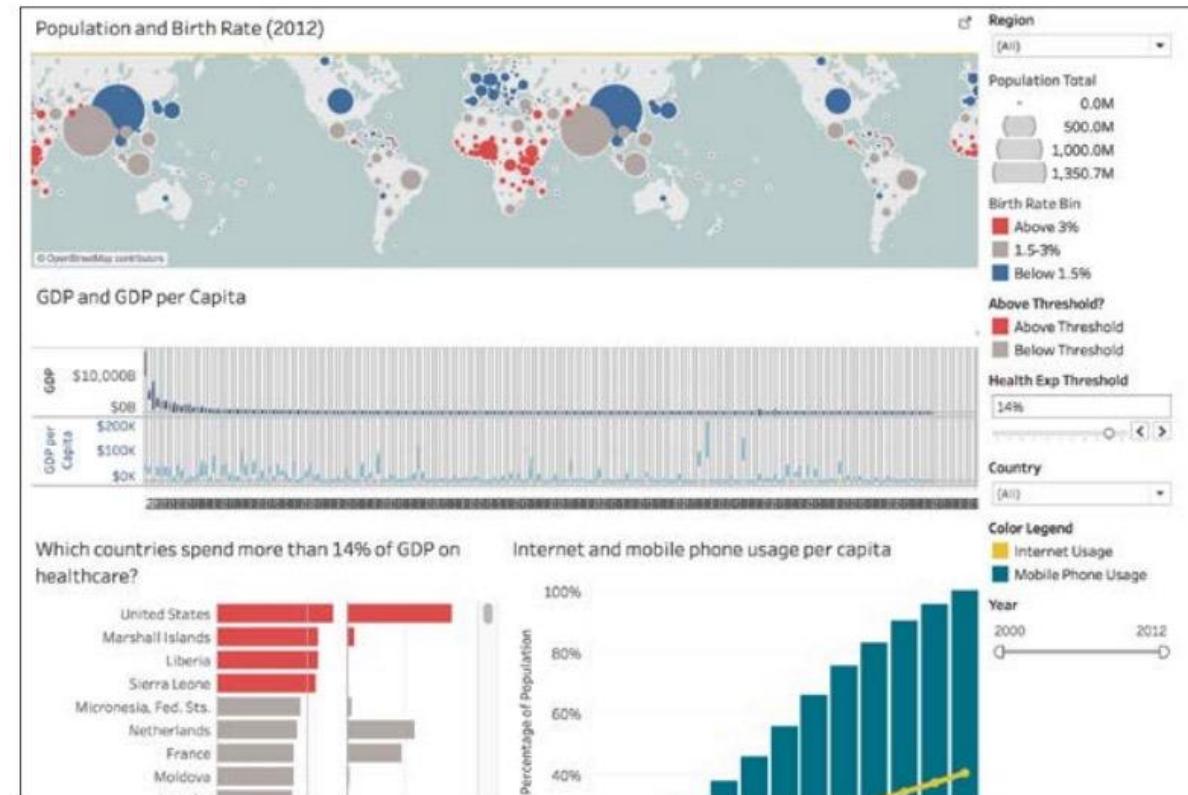
Try swapping a filter title with **explicit language** directions about how to navigate.



Rule of three

Don't make a lot of important information compete for attention.

Most of the time, more than **three visualizations** on one dashboard is too many.



Tell a story

If you need more than one visualization, or are preparing for a presentation, **connect the different visualizations with story points.**

Tell the **narrative of your data with visuals** that build on each other, highlight specific insights, and provide additional context, all in one seamless presentation.



Great visualizations will not only help you understand more about your data, they'll offer **faster, more meaningful answers**, and even **inspire** others to ask and answer new questions.

Resources

- Tufte, E. R., & Graves-Morris, P. R. (1983). *The visual display of quantitative information* (Vol. 2, No. 9). Cheshire, CT: Graphics press.
- Tufte, E. R. (1991). Envisioning information. *Optometry & Vision Science*, 68(4), 322-324.
- Tufte, E. R. (2006). Beautiful evidence. New York.
- Tufte, E. R., & Weise Moeller, E. (1997). *Visual explanations: images and quantities, evidence and narrative* (Vol. 36). Cheshire, CT: Graphics Press.
- Börner, K. (2014). Atlas of knowledge.
- Steele, J., & Iliinsky, N. (2010). *Beautiful visualization: looking at data through the eyes of experts*. " O'Reilly Media, Inc.".
- Fry, B. (2007). *Visualizing data: exploring and explaining data with the Processing environment*. " O'Reilly Media, Inc.".
- Good enough to great - A quick guide for better data visualizations,
<https://www.tableau.com/good-to-great>