



CDS 6324

DATA VISUALIZATION

Lecture 5: More on Visualization Design and Chart Types

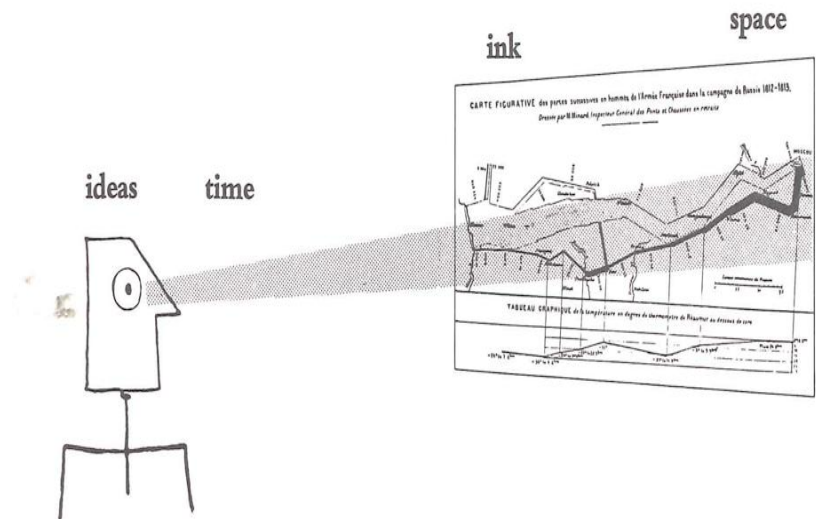


Revision: Graphics Integrity, Visualization Design & Tools

Graphics Excellence

Tufte's Principles:

- Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space



Graphics Integrity

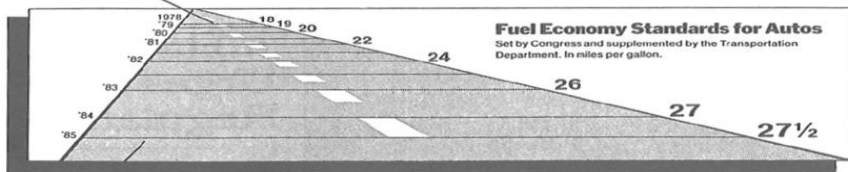
Graphical Integrity Principles:

- Avoid distortion and ambiguity



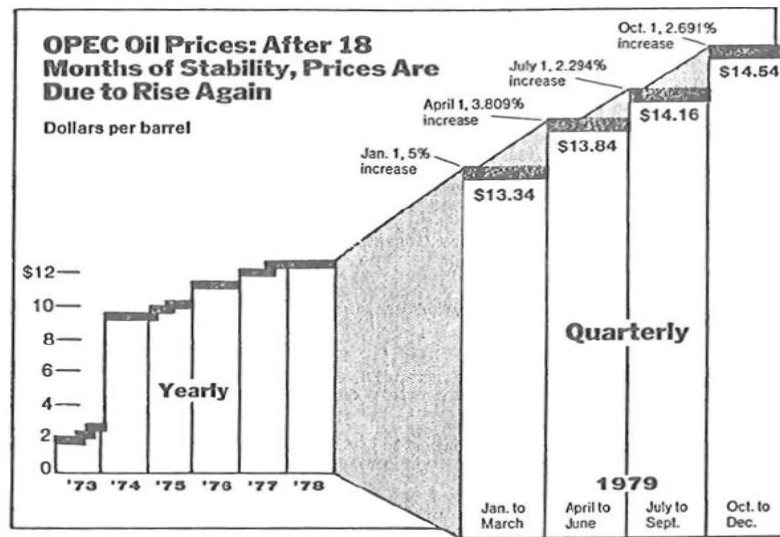
Pittsburgh Civic Commission, *Report on Expenditures of the Department of Charities* (Pittsburgh, 1913), p. 7.

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

Lie Factor = 1.0

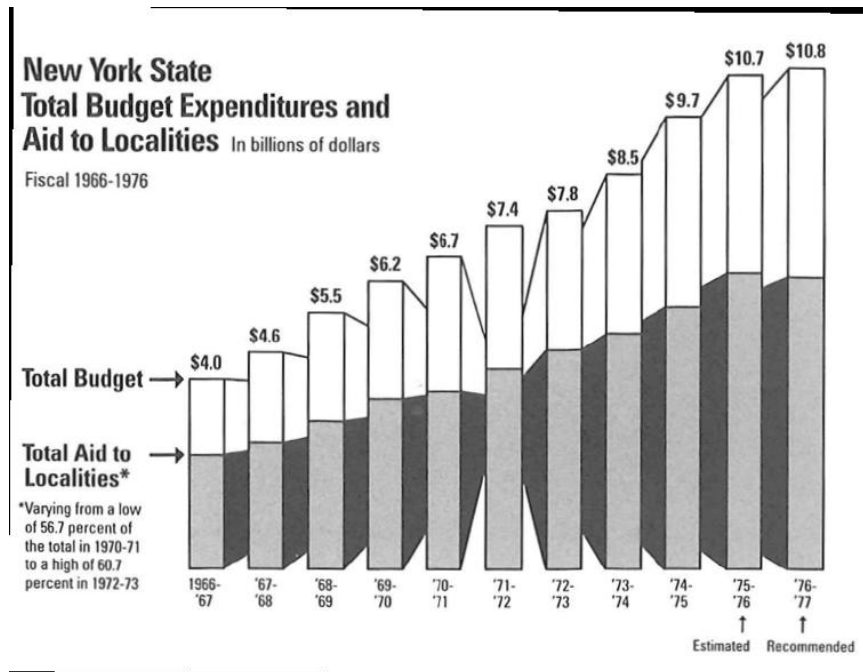


The New York Times / Dec. 19, 1978

Graphics Integrity

Graphical Integrity Principles

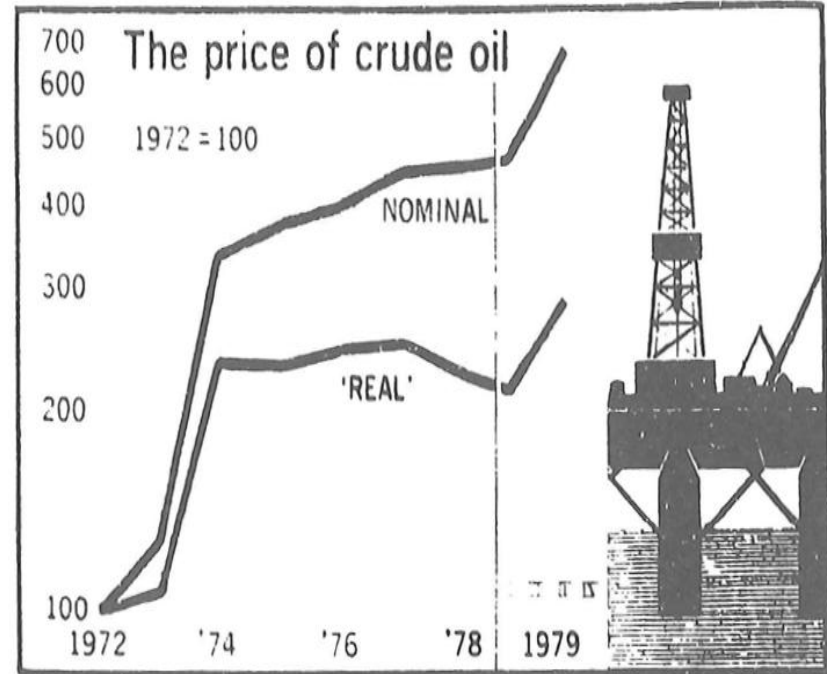
- **Avoid distortion** and ambiguity
- **Show data variation**, not design variation (avoid fancy chart junk / tricks!)



Graphics Integrity

Graphical Integrity Principles

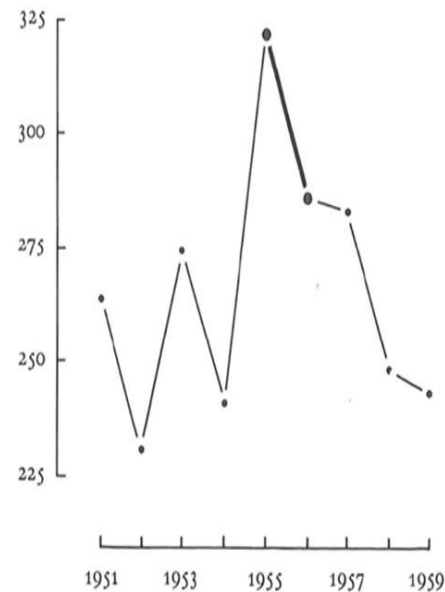
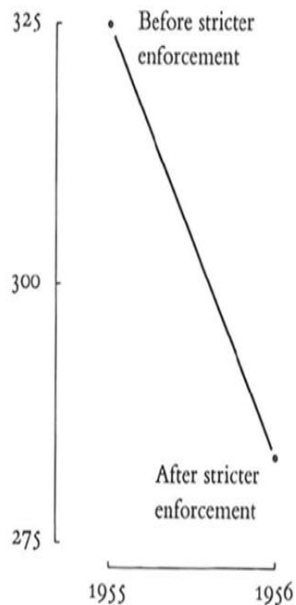
- **Avoid distortion** and ambiguity
- **Show data variation**, not design variation (avoid fancy chart junk / tricks!)
- Account for **Inflation**



Graphics Integrity

Graphical Integrity Principles

- **Avoid distortion** and ambiguity
- **Show data variation**, not design variation (avoid fancy chart junk / tricks!)
- Account for **Inflation**
- Graphics should be put in **context**.





Tufte's Design Principles

- ❑ Data-Ink Maximization and Graphical Redesign
- ❑ Remove Chartjunk
- ❑ Multi-functioning Graphical Elements
- ❑ Data Density
- ❑ Small Multiples and Parallel Sequencing



More Visualization Design Principles

Adapted from: Knaflic, Cole. *Storytelling with Data: A Data Visualization Guide for Business Professionals*, Wiley, © 2015, 2025.

Cole Nussbaumer's Design Concepts

Form Follows Function

- What function do we want our audience to do?
- Create a form that will allow the audience to do the function with ease.



Cole Nussbaumer's Design Concepts

4 important concepts in design:

- Affordances
- Accessibility
- Aesthetics
- Acceptance

Cole Nussbaumer's Design Concepts


Affordances

- Make it obvious how to use
- Indicate to the audience how to use and interact with visualizations
 - Highlight the important
 - Eliminate distractions



Cole Nussbaumer's Design Concepts

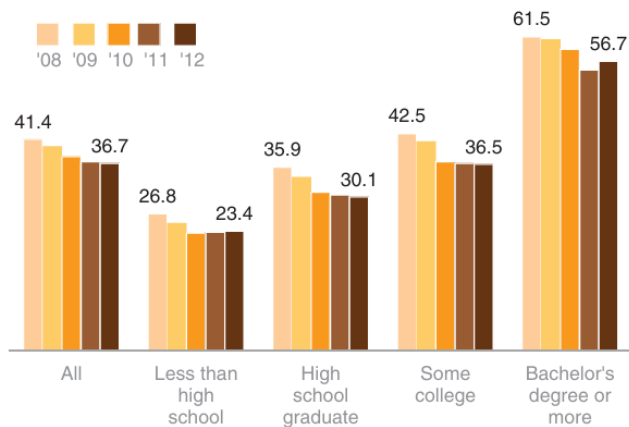
Affordances

- Highlight the important  Graphic design affordances
 - Bold, italics, or underlining
 - CASE and typeface
 - Size and color

Cole Nussbaumer's Design Concepts

New Marriage Rate by Education

Number of newly married adults per 1,000 marriage eligible adults



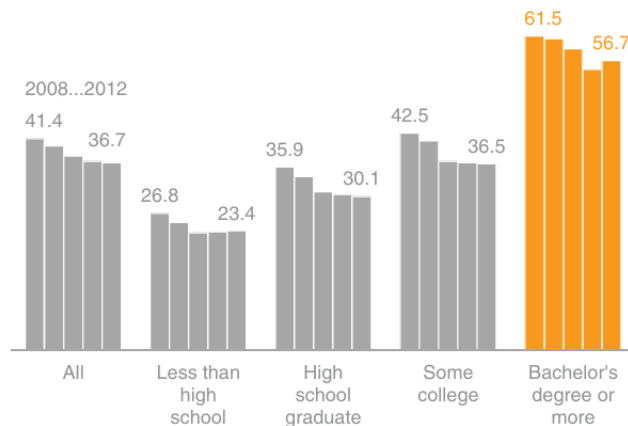
Note: Marriage eligible includes the newly married plus those widowed, divorced, or never married at interview.

Source: U.S. Census

Adapted from PEW RESEARCH CENTER

New Marriage Rate by Education

Number of newly married adults per 1,000 marriage eligible adults



Note: Marriage eligible includes the newly married plus those widowed, divorced, or never married at interview.

Source: U.S. Census

Adapted from PEW RESEARCH CENTER

Main message to audience: the increase in new marriages from 2011 to 2012 was mainly due to people with a bachelor's degree or more.

Cole Nussbaumer's Design Concepts

Affordances

- Eliminate distractions
 - Clutter
 - Context

Not all data are equally important

When detail isn't needed, summarize

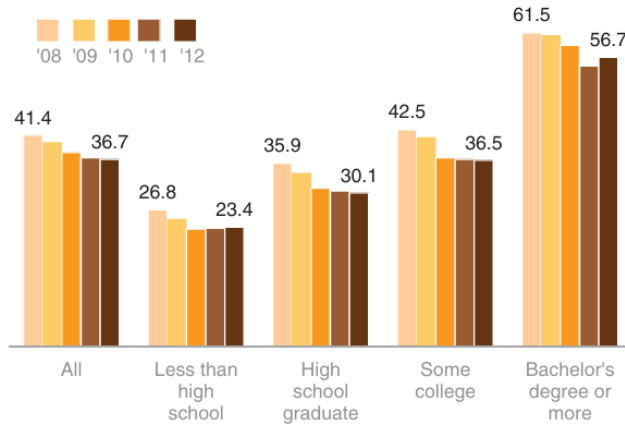
Ask yourself: would eliminating this change anything?

Keep necessary but non-critical elements in the background.

Identify the design changes made to eliminate distractions

New Marriage Rate by Education

Number of newly married adults per 1,000 marriage eligible adults



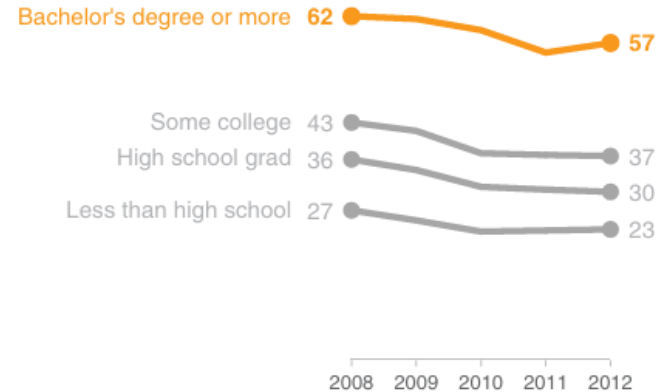
Note: Marriage eligible includes the newly married plus those widowed, divorced, or never married at interview.

Source: U.S. Census

Adapted from PEW RESEARCH CENTER

New marriage rate by education

Number of newly married adults per 1,000 marriage eligible adults



Note: Marriage eligible includes the newly married plus those widowed, divorced, or never married at interview.

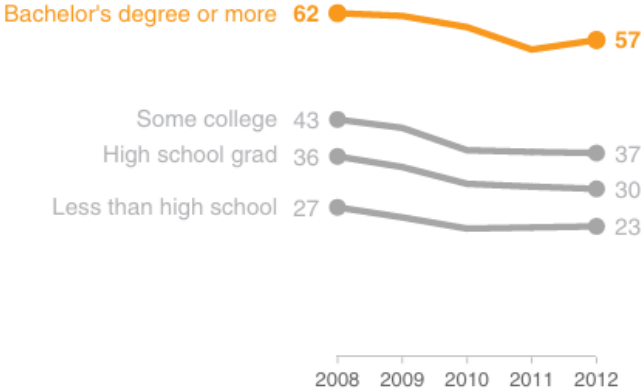
Source: U.S. Census

Adapted from PEW RESEARCH CENTER

Cole Nussbaumer's Design Concepts

New marriage rate by education

Number of newly married adults per 1,000 marriage eligible adults



Note: Marriage eligible includes the newly married plus those widowed, divorced, or never married at interview.

Source: U.S. Census

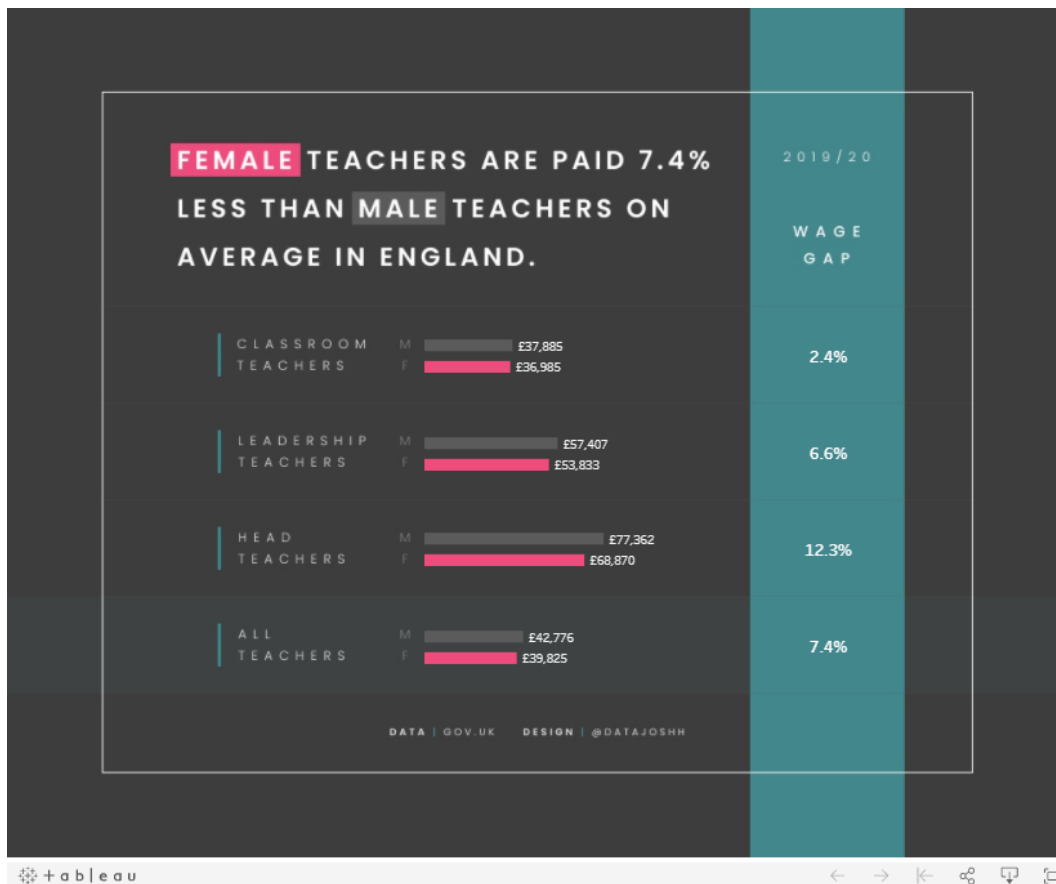
Adapted from PEW RESEARCH CENTER

Cole Nussbaumer's Design Concepts

Accessibility

Usable by people of widely varying technical skills

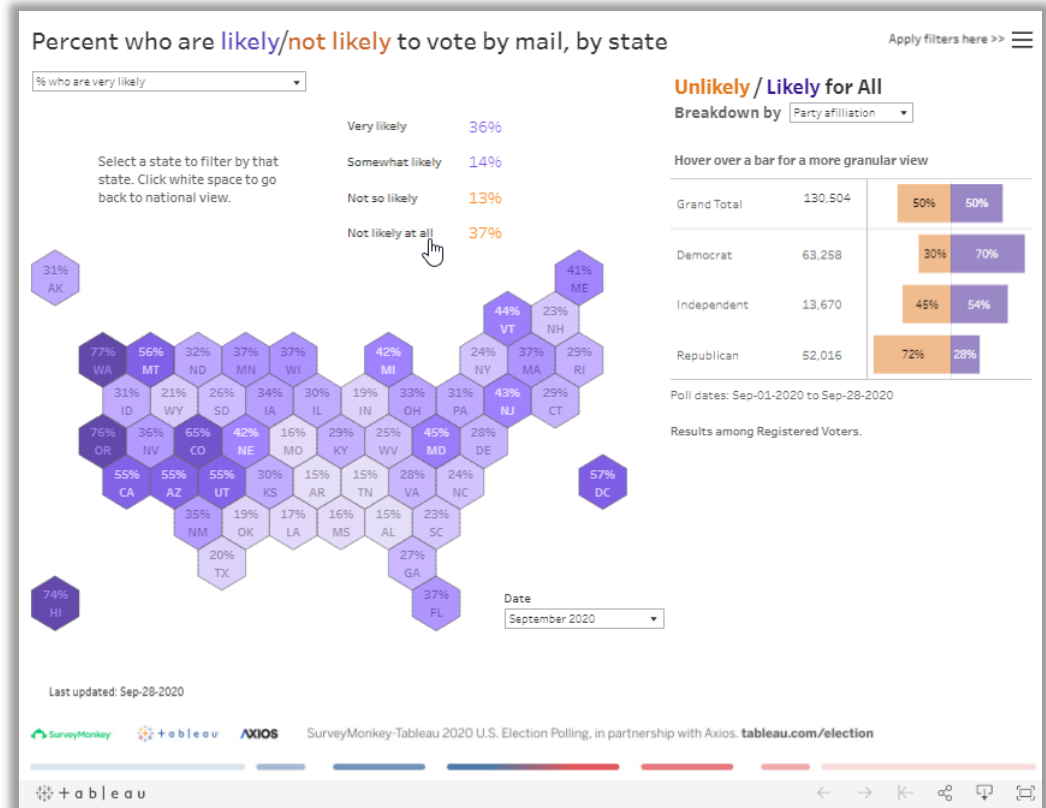
- Make it legible
- Keep it clean
- Use plain language
- Remove unnecessary complexity



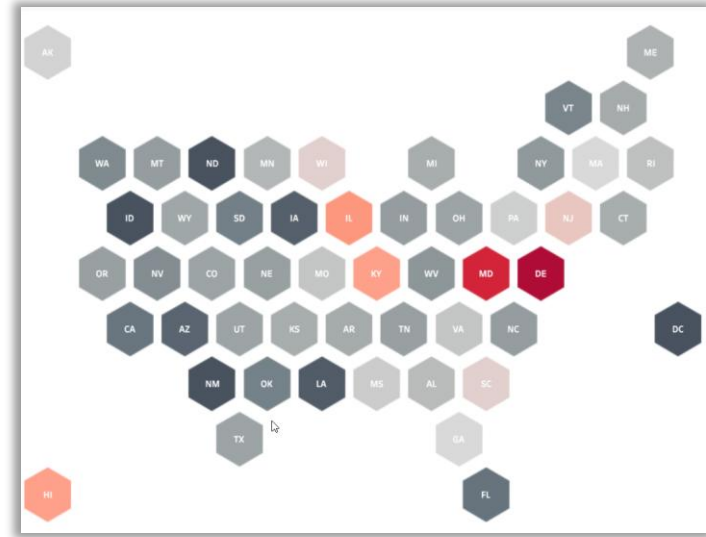
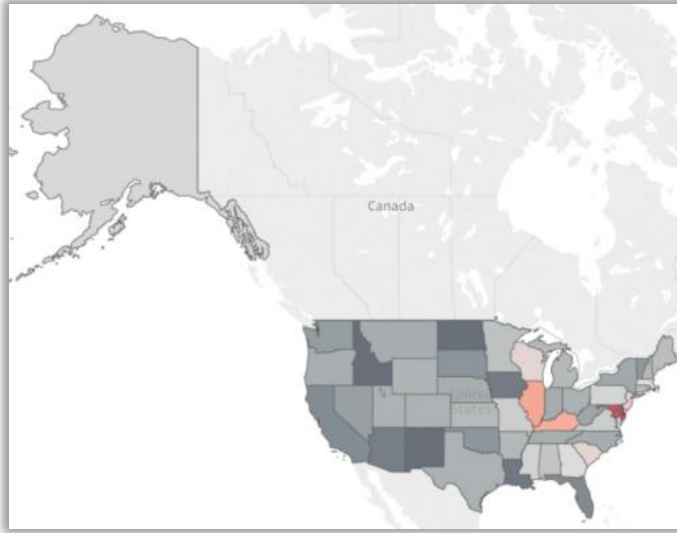
Cole Nussbaumer's Design Concepts

Aesthetics

- Be smart with color
- Pay attention to alignment
- Leverage white space



Cole Nussbaumer's Design Concepts



Acceptance

- Articulate benefits
- Show the side-by-side
- Provide multiple options or seek input
- Get a vocal audience member on board



The Shaffer 4 C's of Data Visualization

Clear - easily seen; sharply defined

- who's the audience? what's the message?
- clarity more important than aesthetics

Clean - thorough; complete; unadulterated

- labels, axis, gridlines, formatting, right chart type, color choice, etc.

Concise - brief but comprehensive

- **not minimalist** but not verbose

Captivating - to attract and hold by beauty or excellence

- does it capture attention? is it interesting? does it tell the story?

Guided Independent



Learn more about Shaffer's 4Cs of Data Visualization
to deepen your understanding



<http://www.storytellingwithdata.com/blog/2012/09/some-finer-points-of-data-visualization>

Chart Types

[Adapted from: Fundamentals of Data Visualization](#)

Visualization by Analytical Task

Amount

Distribution

Proportions

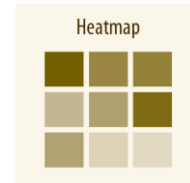
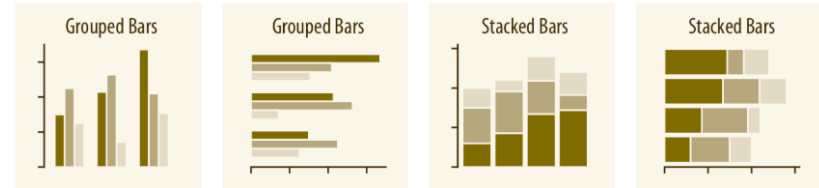
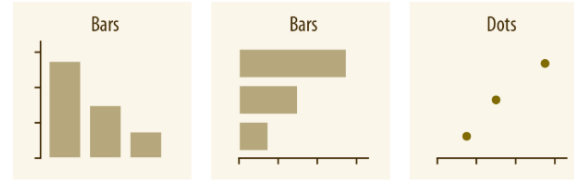
X-Y
Relationships

Geospatial
Data

Uncertainty

Amounts

- Visualizing amounts – compare magnitude across categories
- Categorical variable + quantitative value
- Visual goal: make size differences easy to see



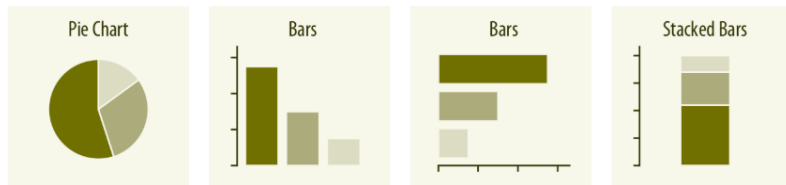
Distributions

- Distribution – how values are spread within a dataset
- Most common:
 - Histogram
 - Density plot
 - Cumulative density
 - Q-Q plots

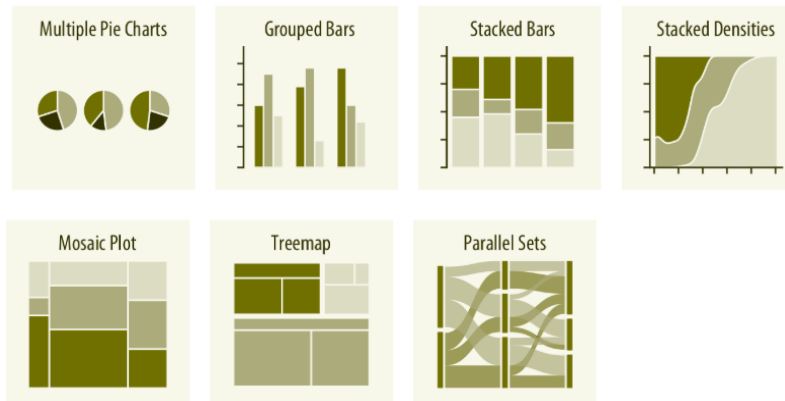


Proportions

- Common charts
 - Pie charts
 - Side-by-side (grouped) bar charts
 - Stacked bar charts

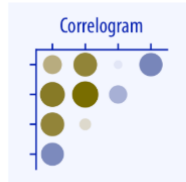
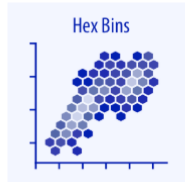
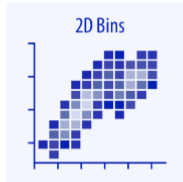
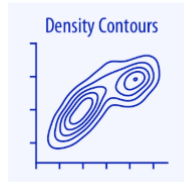
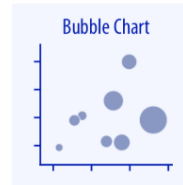


- Changing proportions over a continuous variable
- Multiple grouping variables



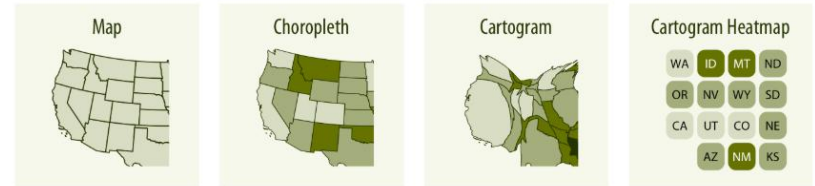
Relationships

- Scatterplots - to show how one quantitative variable changes relative to another
- Bubble chart – 3 quantitative variables
- Paired data
- Large dataset
- Time or ordered axis



Geospatial Data

- Map – convert coordinates to a 2D representation
- Choropleth maps – coloured regions based on data value
- Cartograms – regions are resized based on other quantity



Uncertainty

- Show how confident we are in a value
- Error bars / confidence bands
 - Show a likely range
 - Simple but limited
- Distribution-based plots
 - Show the shape of uncertainty
 - More informative, harder to read

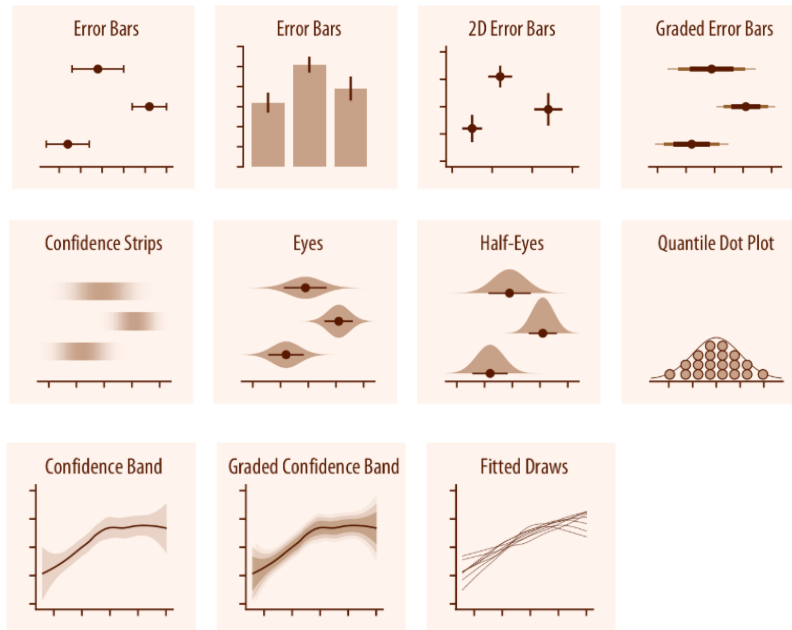




Chart Types : Common Charts

Bar



Bar Chart

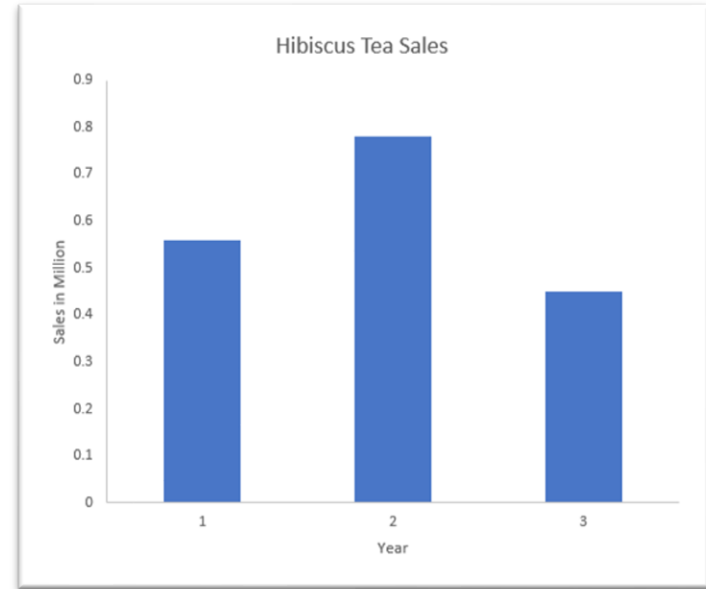
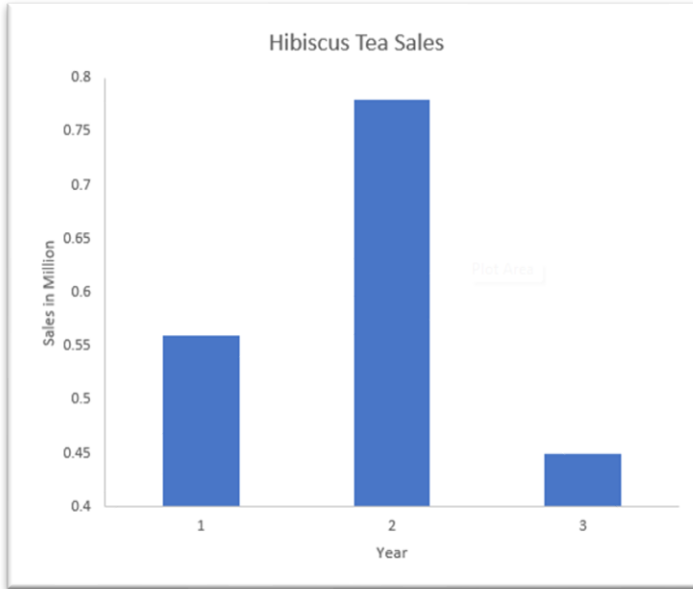
- Good bar chart design

- Bars start at zero
- Axes are clearly labeled
- Colors are consistent and meaningful
- Avoid showing too many bars

- Avoid

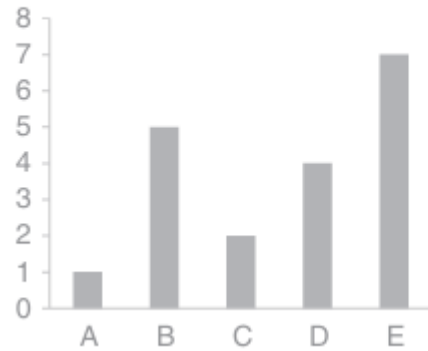
- Bars with different widths
- Too many subcategories in one chart
- Missing or unclear axis labels

Zero Baseline

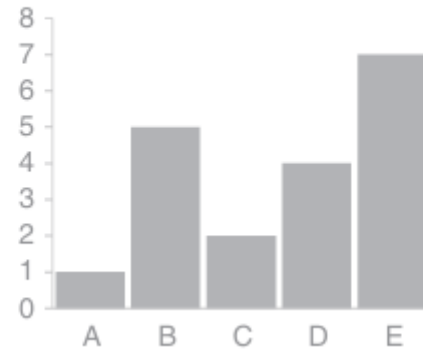


Bar Width

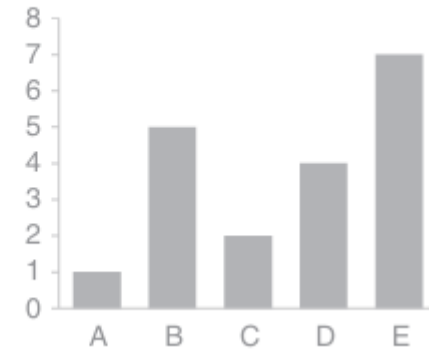
Too thin



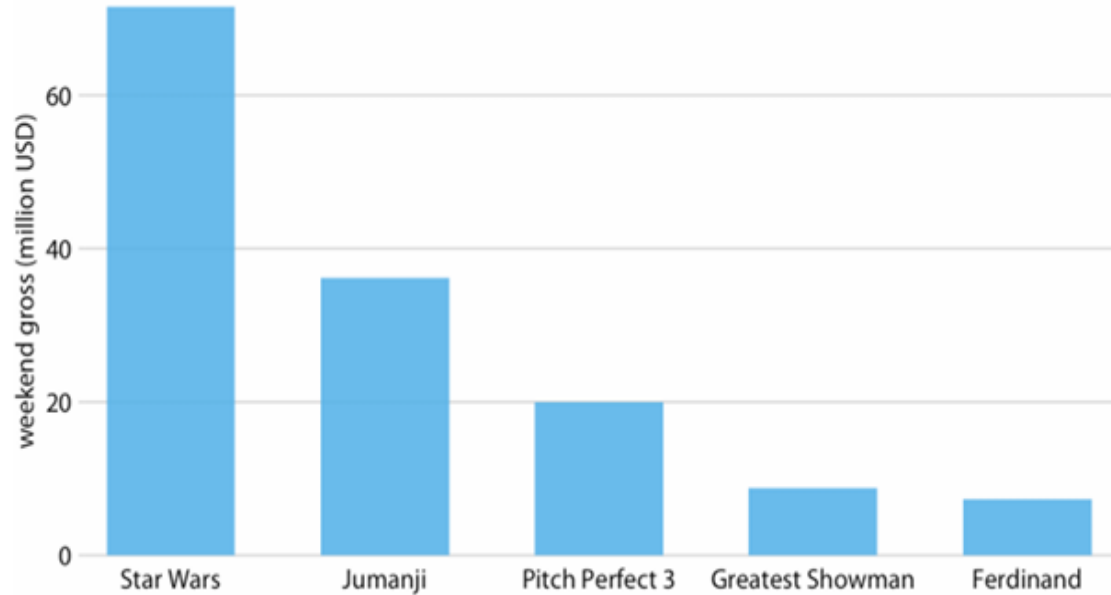
Too thick



Just right

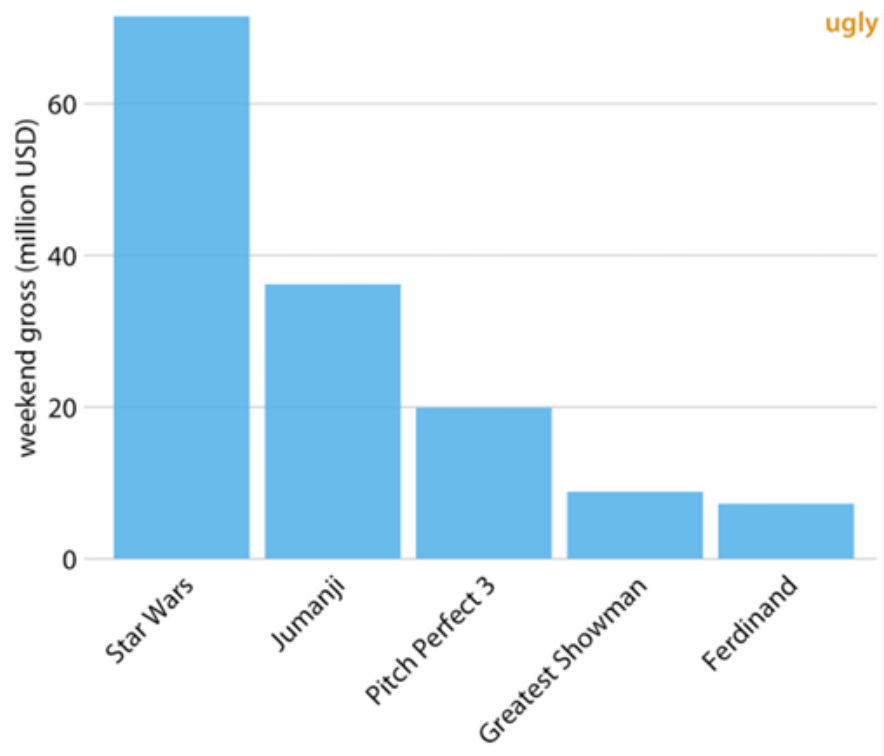


Horizontal or vertical?

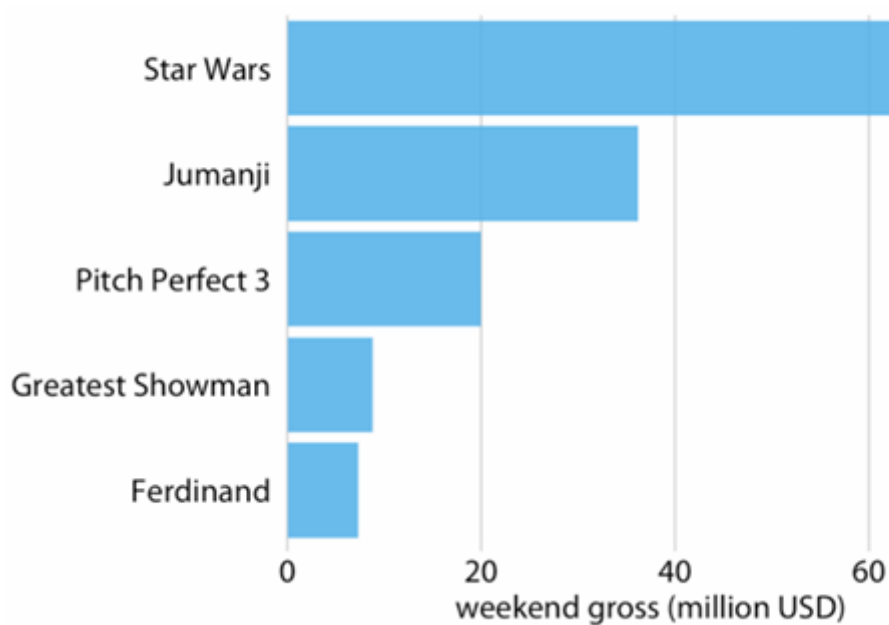


Each bar starts at zero and ends at the dollar value of the movie's weekend earnings.

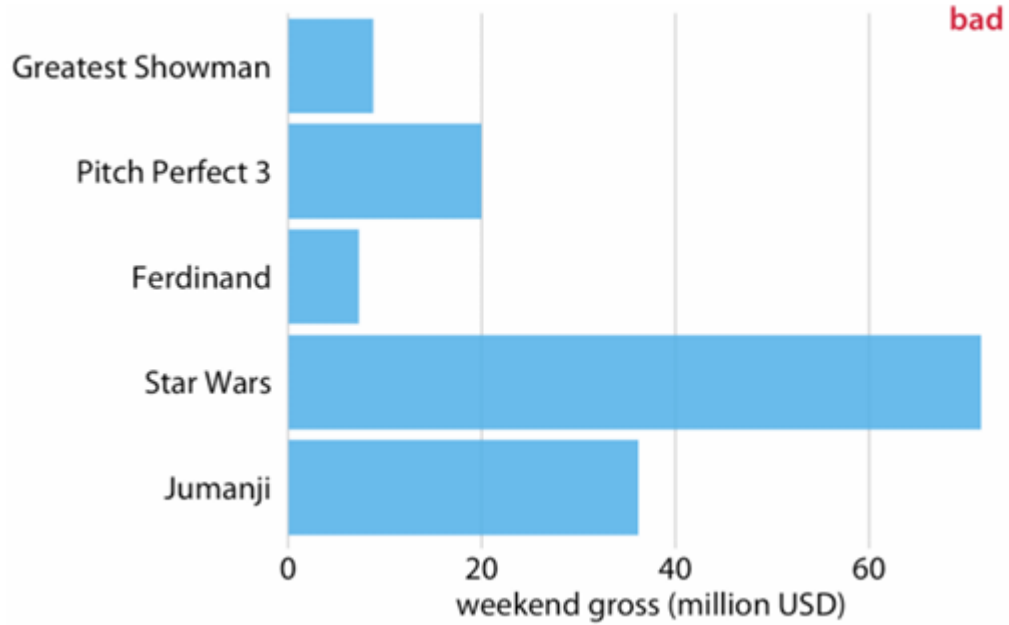
Horizontal or vertical?



Horizontal or vertical?

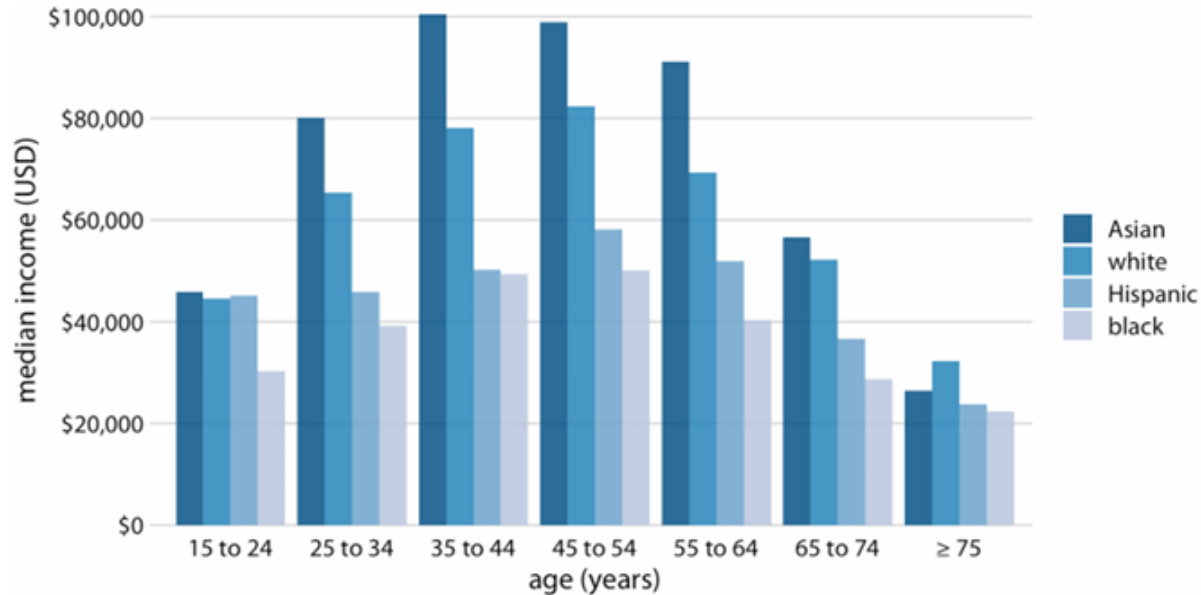


Ordering



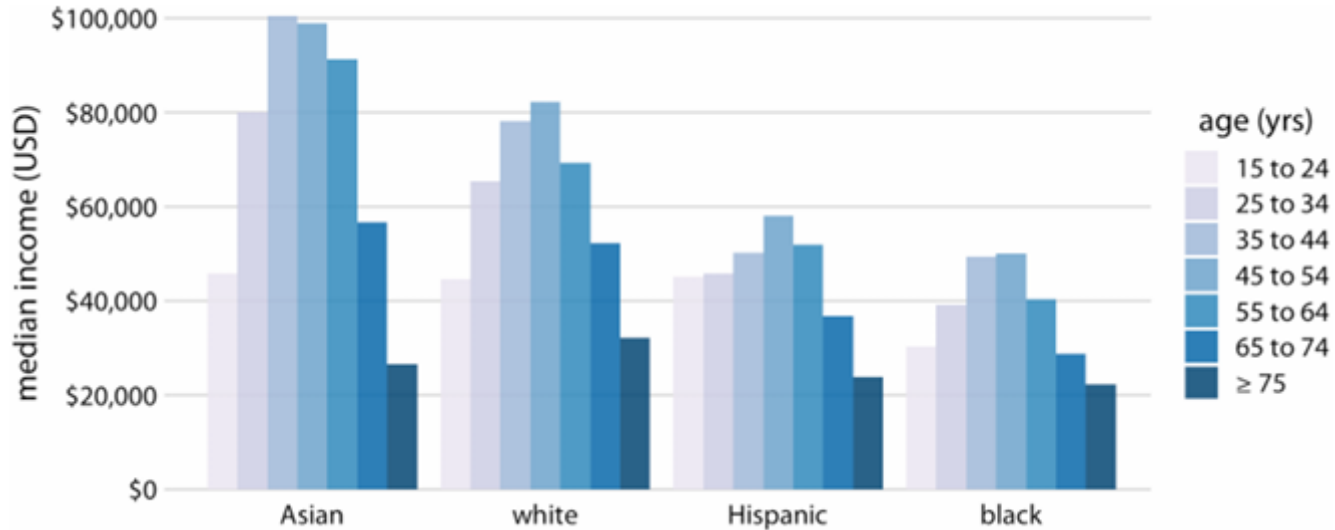
Rearrange bars only
when categories have
no natural order.

Grouped and Stacked Bars



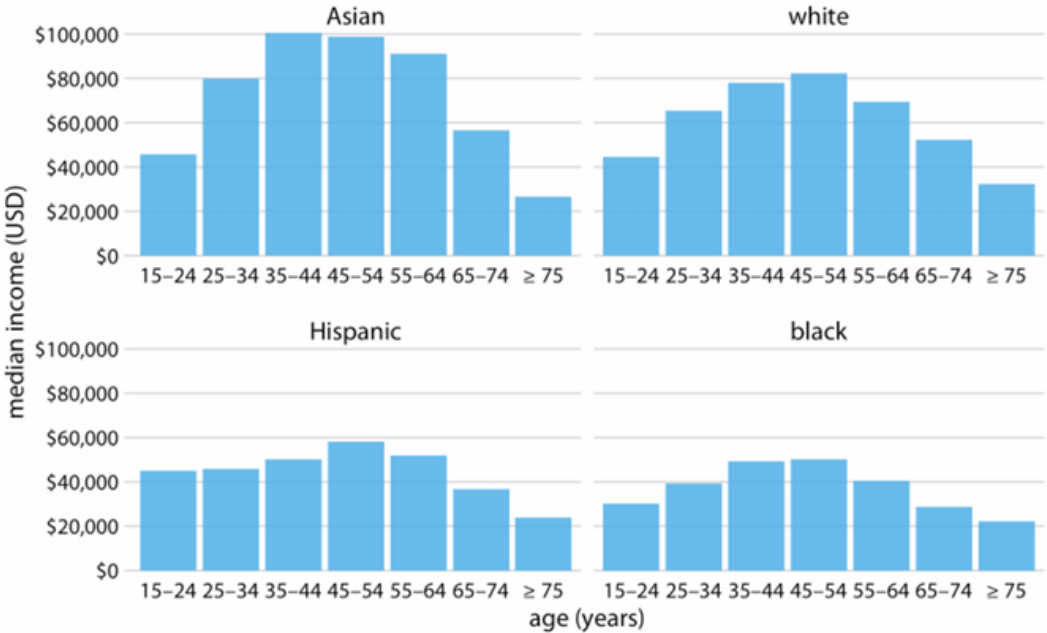
compare income differences among racial groups within specific age groups ✓

Grouped and Stacked Bars



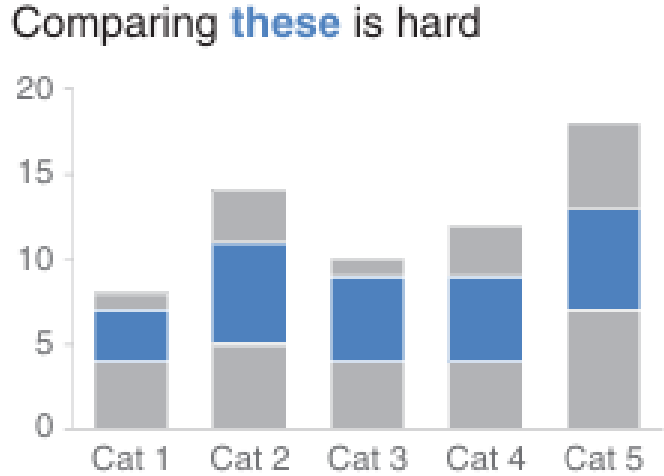
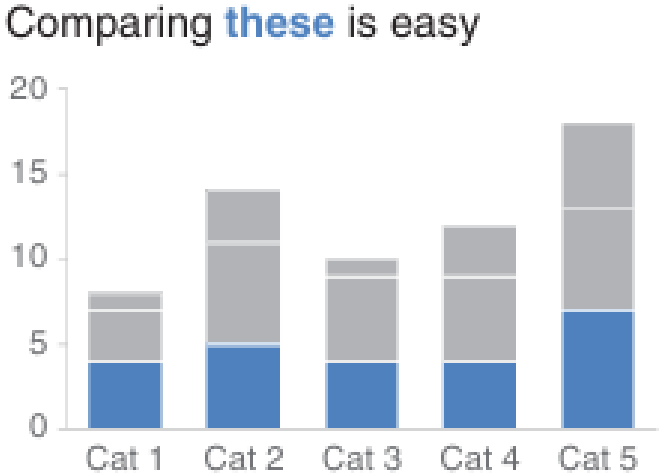
compare median incomes across age groups for the same racial group ✓

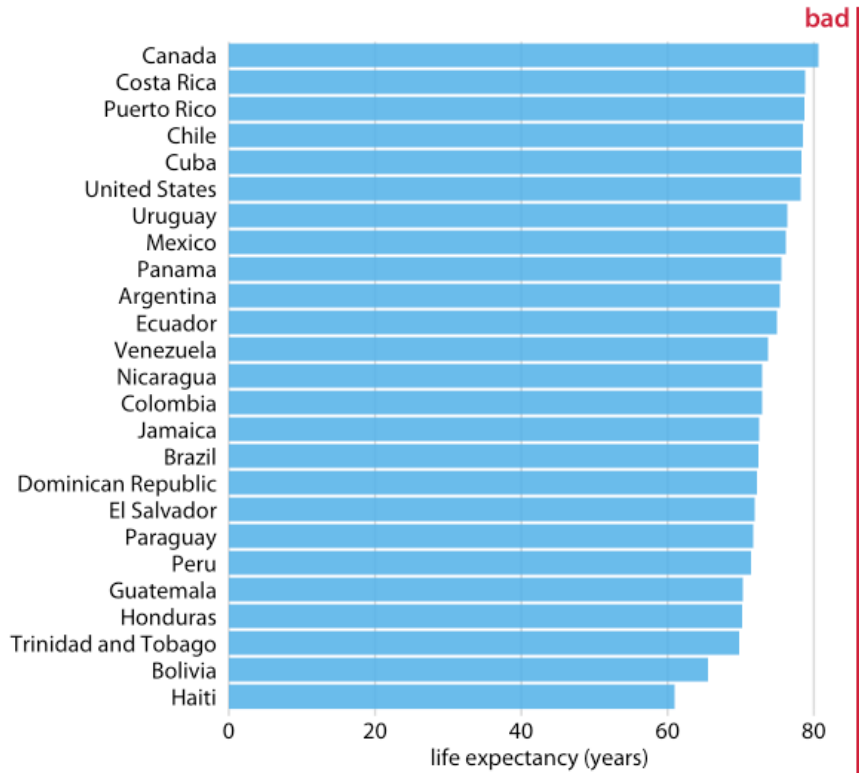
Grouped and Stacked Bars



Less mental effort ✓

Grouped and Stacked Bars





Bar chart is not ideal.
The long bars draw attention away from the small differences in life expectancy.

Life expectancies in 25 countries in the Americas

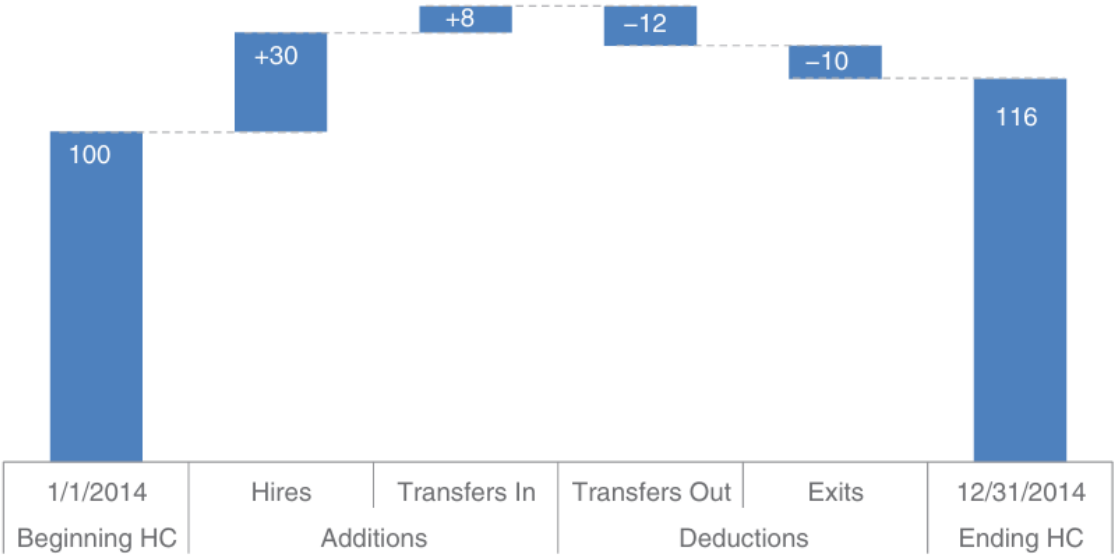
Waterfall Chart

- Show cumulative effect of positive and negative changes on a variable of interest.
- Changes are reported for a series of categories (e.g. time period)
- Magnitude of change is represented by a column anchored at the cumulative height of the changes in the preceding categories.

Waterfall Chart

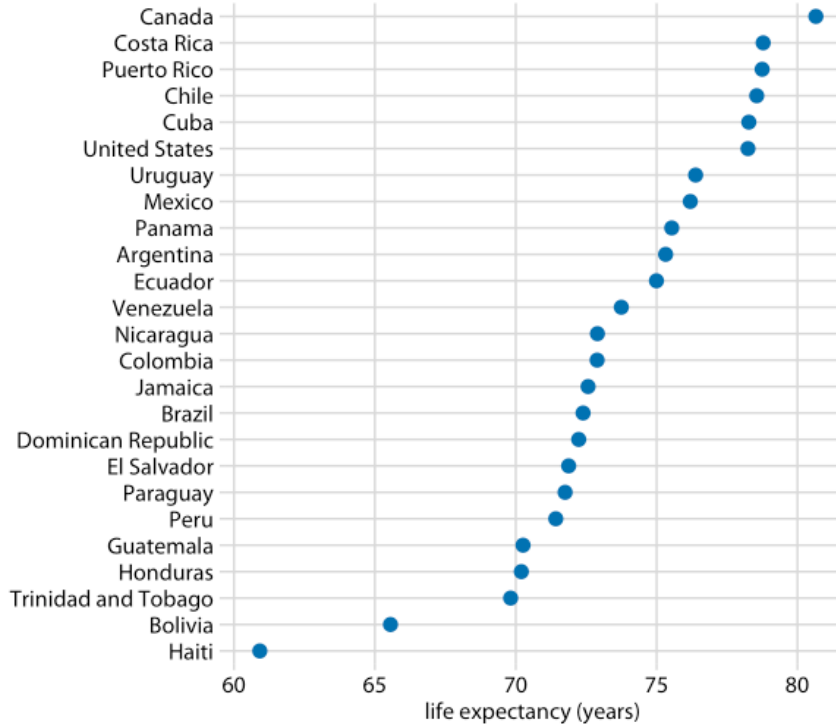
2014 Headcount math

Though more employees transferred out of the team than transferred in, aggressive hiring means overall headcount (HC) increased 16% over the course of the year.



Dot and Heatmap

Dot and Heatmap



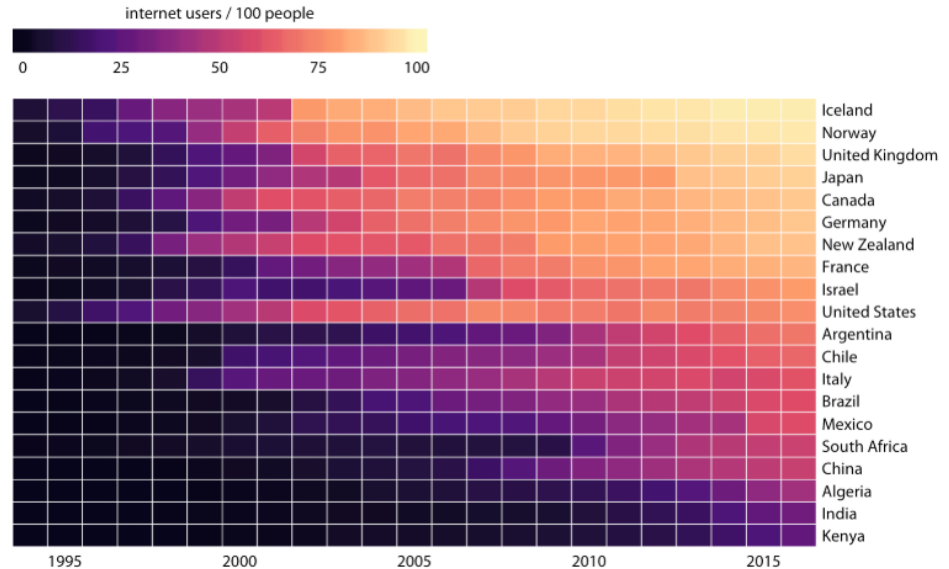
- Dot plots

- Show values using position
- Allow axes that do not need to start at zero
- Often clearer than bars for highlighting small differences

Dot and Heatmap

- Heatmaps

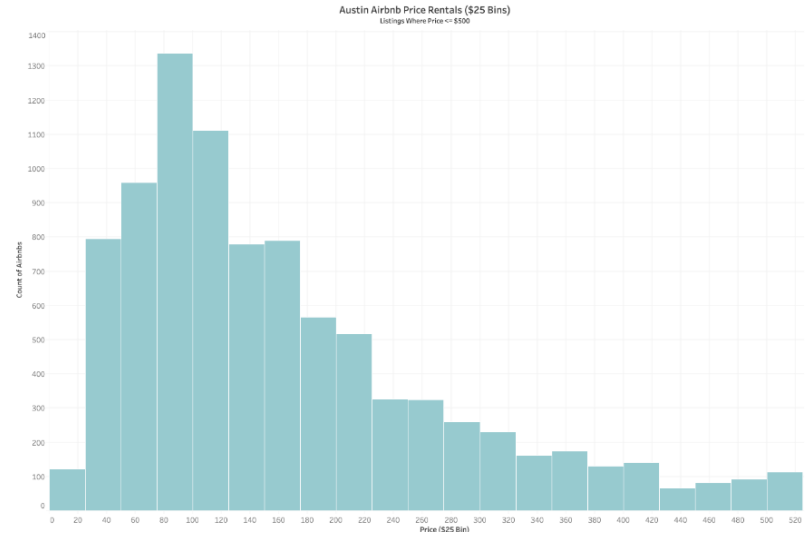
- Map data values onto colours
- Suitable for larger datasets
- Highlighting broader trends



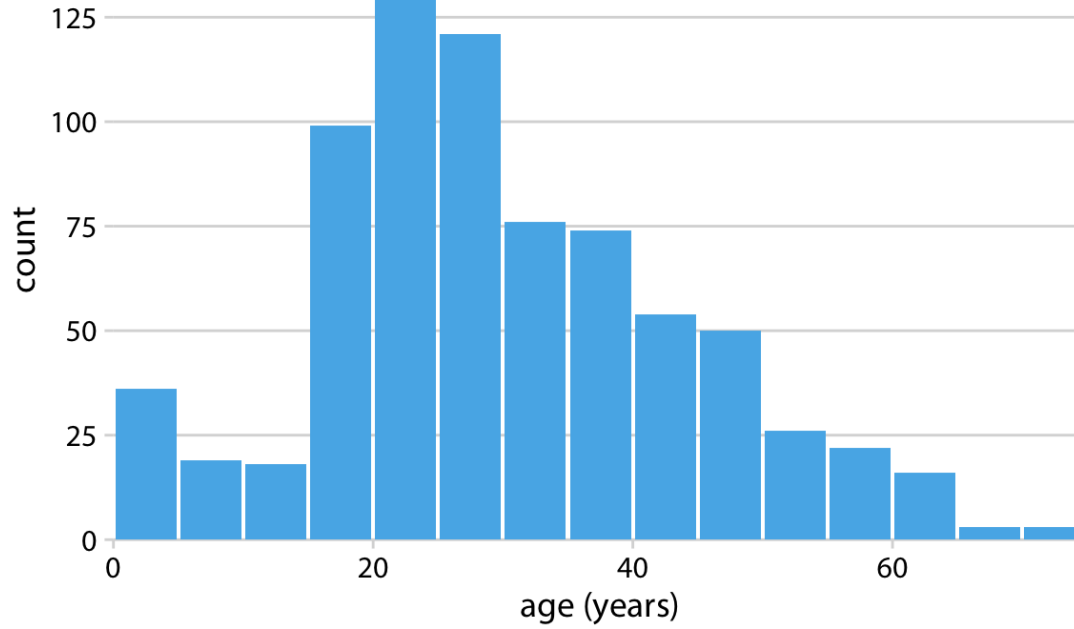
Histogram

Histogram

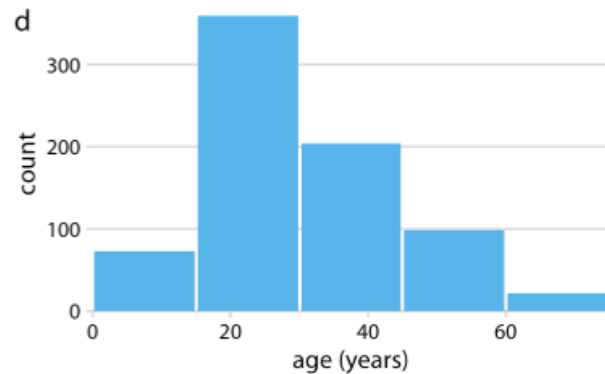
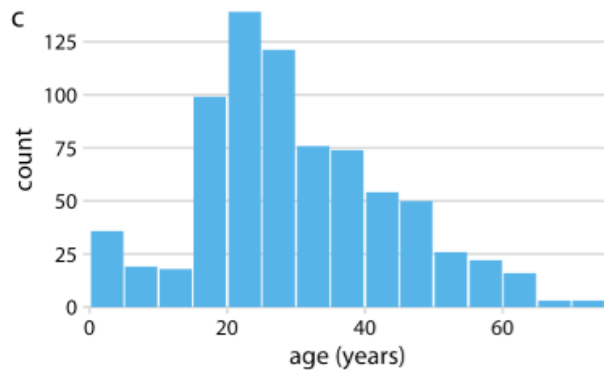
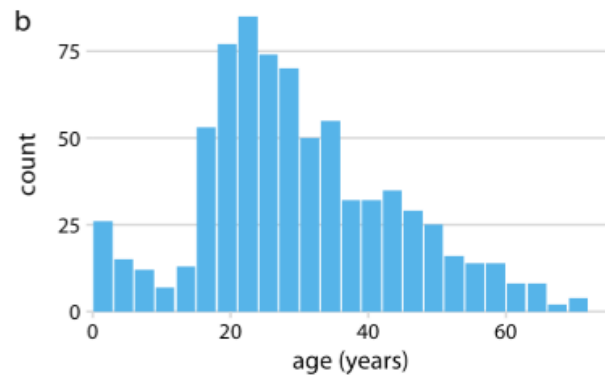
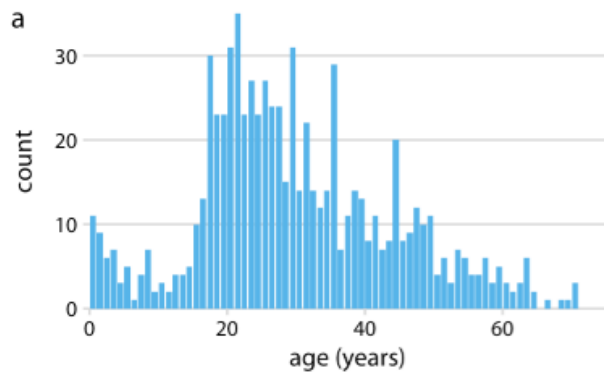
- A variation of a bar chart used to show data distributions
- Displays one numerical variable that can be split into bins
- Best for continuous numerical data
- Not suitable for comparing multiple categories



Ages of Titanic Passengers

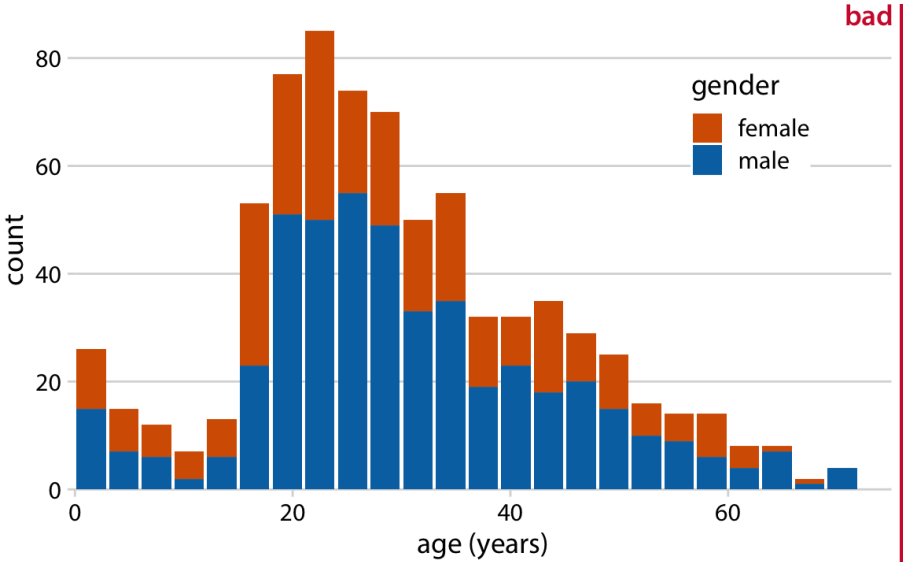


Bin Size



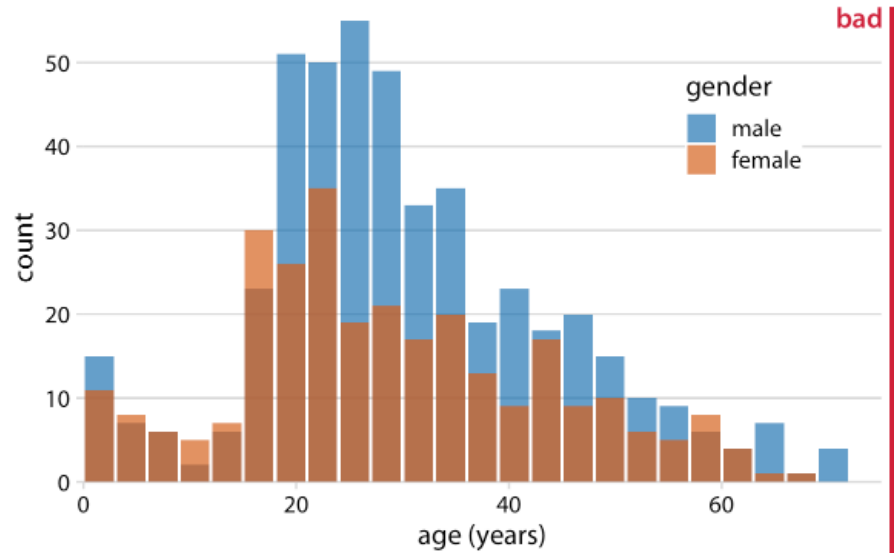
Stacked Histogram

- How are the ages of Titanic passengers distributed between men and women?
- Bars starts at different heights



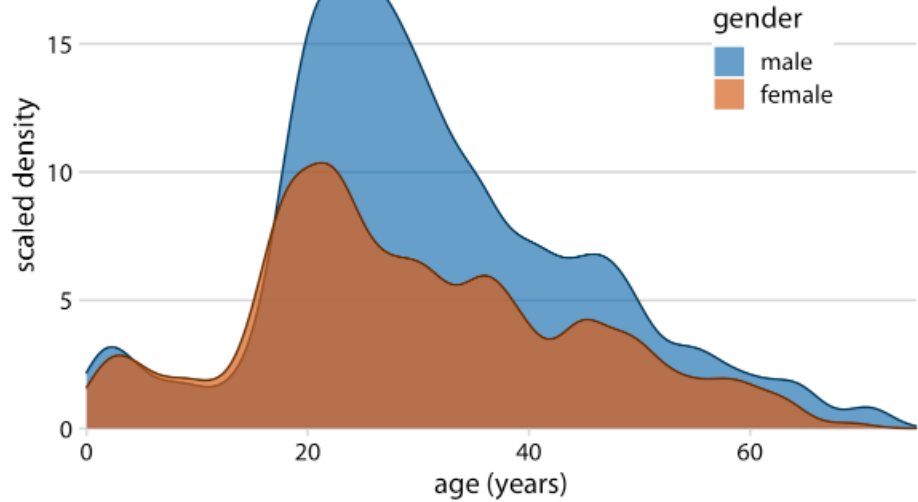
Stacked Histogram

- Overlapping histograms
- 2 or 3 classes?
- Blue bars start at 0?



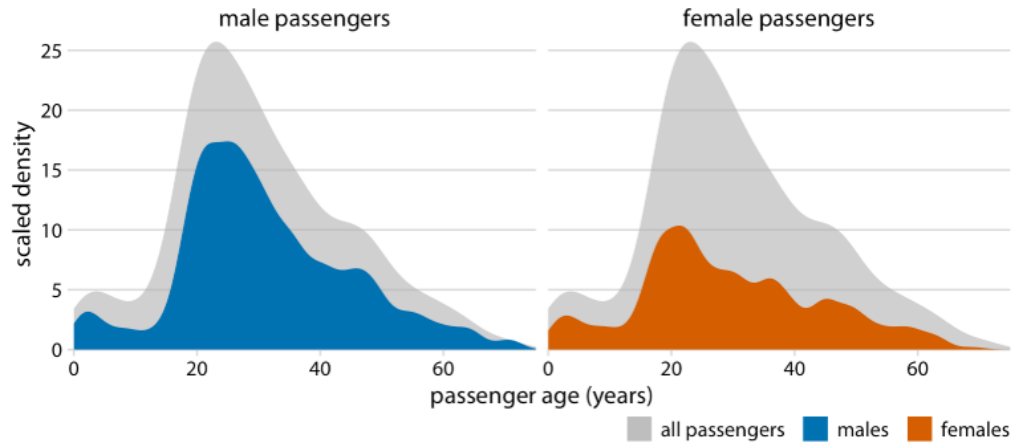
Density Plot

- Overlapping density plots



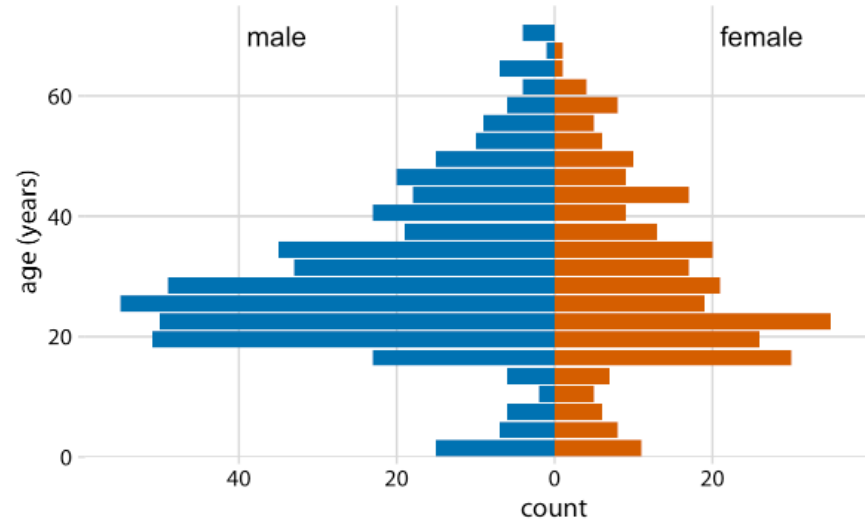
Density Plot

- Show male and female age distributions separately, as proportion of the passenger total.



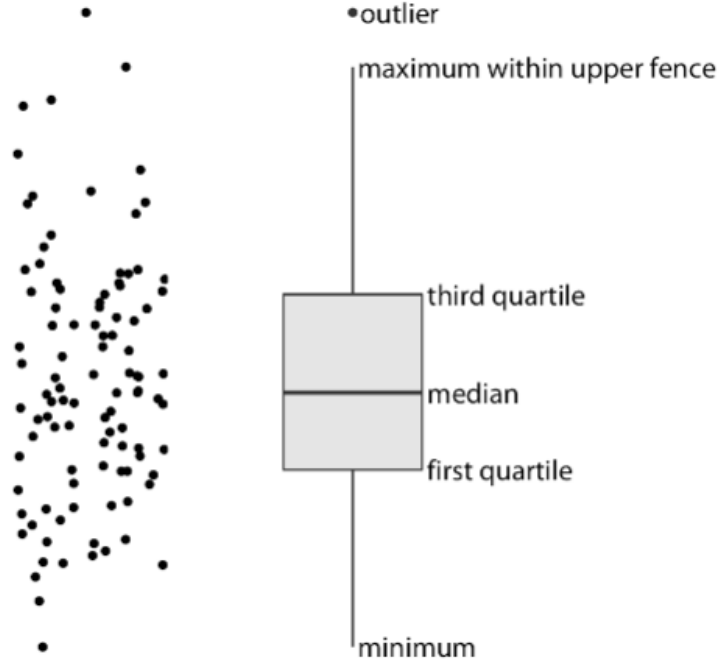
Pyramid Chart

- Visualize two distributions



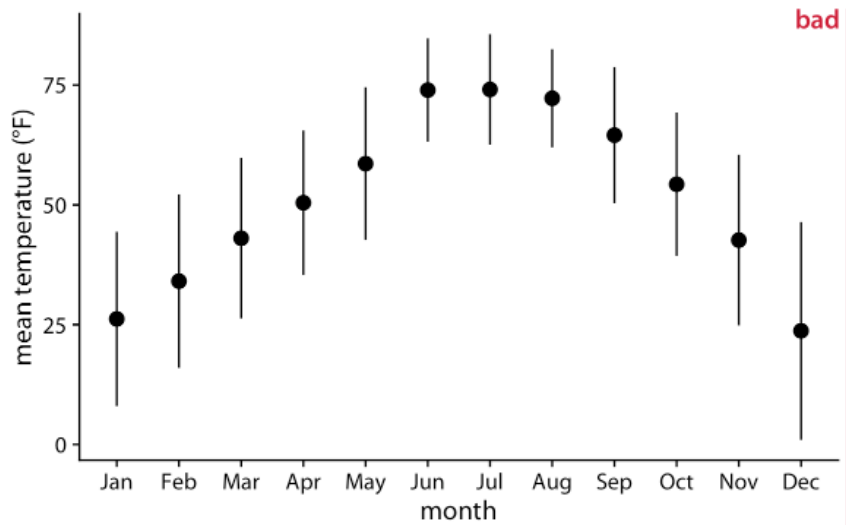
Boxplot

- Summarizes a distribution using five key values
- Comparing multiple distributions

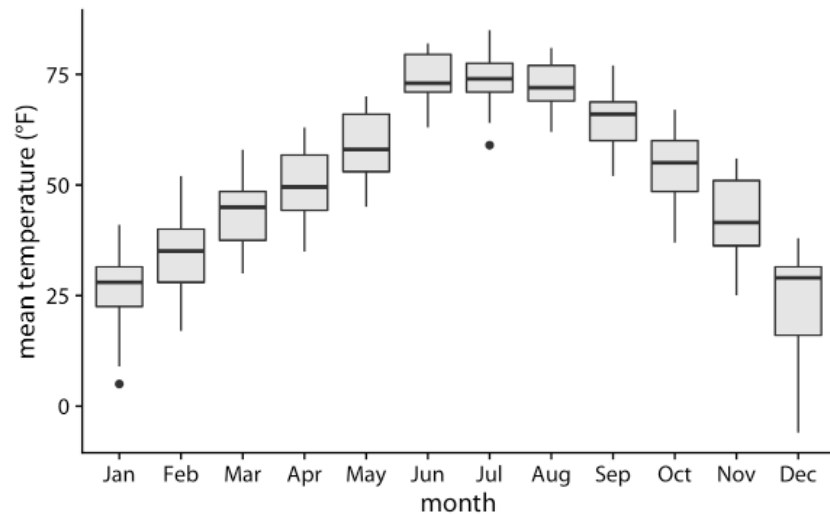




Error Plot



Box Plot



Mean daily temperatures in Lincoln, Nebraska in 2016

Pie

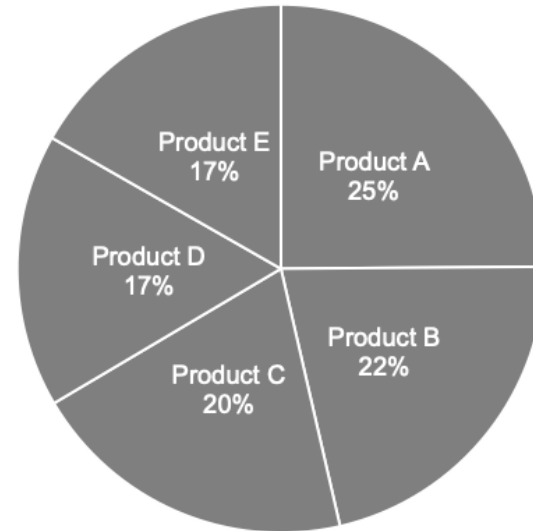


Pie Chart

- Sum of all slices must be 100%
- Component of each slice:
 - Area
 - Arc length
 - Central angle

Sales by product

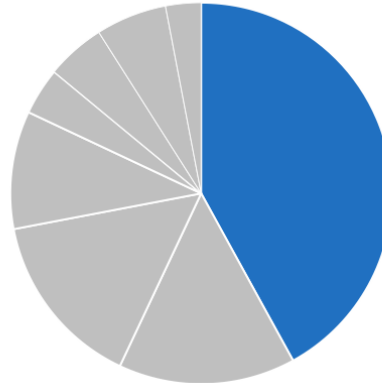
TOTAL = \$75.2M



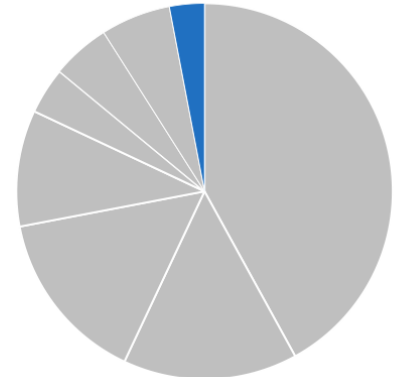
Pie Chart

- Use pie chart:
 - to show a general part-to-whole relationship
 - when precise comparison is not critical

This slice is **large**
% OF TOTAL



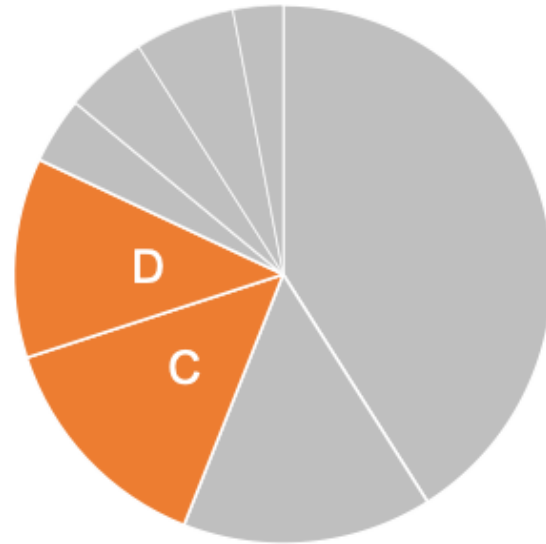
This slice is **small**
% OF TOTAL



Pie Chart

- Avoid when:
 - Comparing size of categories
 - Make precise comparison

Is C or D larger, and by how much?
% OF TOTAL

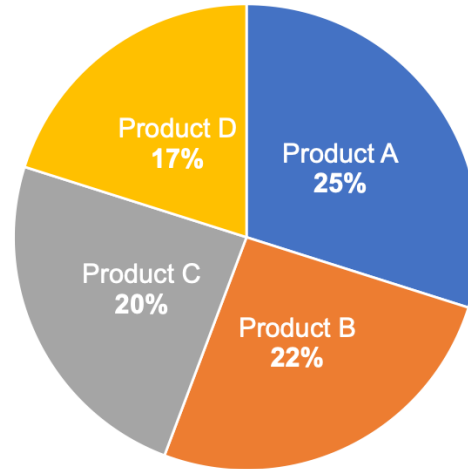


Pie Chart

- Avoid when:
 - Comparing size of categories
 - Make precise comparison
 - Data does not sum to 100%

This **does not** make sense!

Sales by product
% OF TOTAL

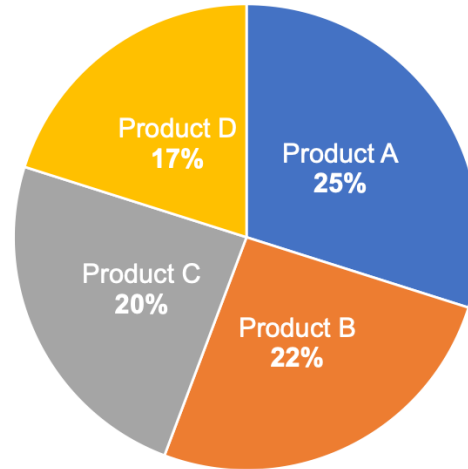


Pie Chart

- Avoid when:
 - Comparing size of categories
 - Make precise comparison
 - Data does not sum to 100%

This **does not** make sense!

Sales by product
% OF TOTAL

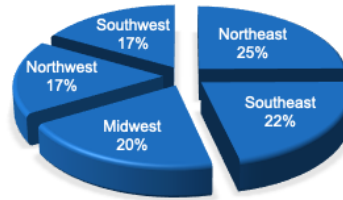


Pie Chart : Avoid 3D

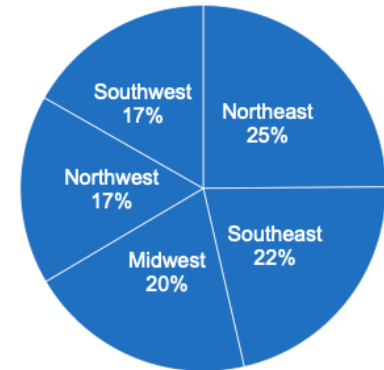
3D Pie Charts

- Distort perceived areas
- Increase error in interpretation

Call breakdown by region
% OF TOTAL



Call breakdown by region
% OF TOTAL

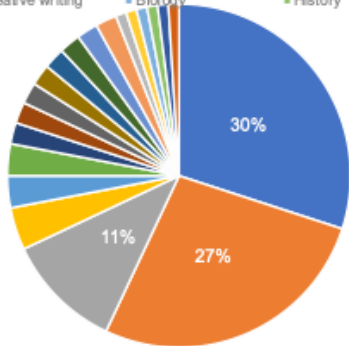


Pie Chart : Many Slices

Too many slices

Declared majors

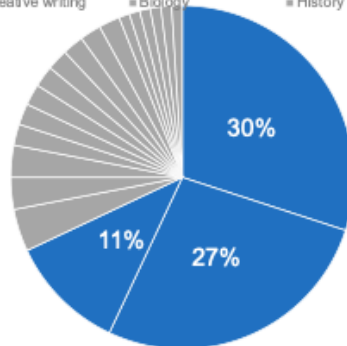
% OF ANDERSON HALL RESIDENTS



De-emphasize

Declared majors

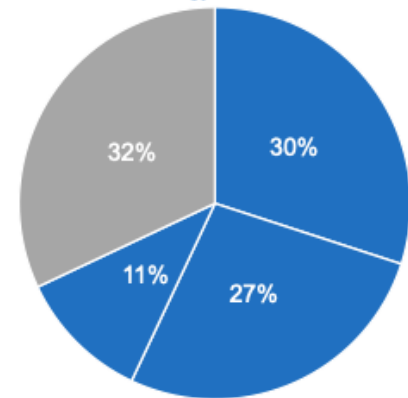
% OF ANDERSON HALL RESIDENTS



Aggregate

Declared majors

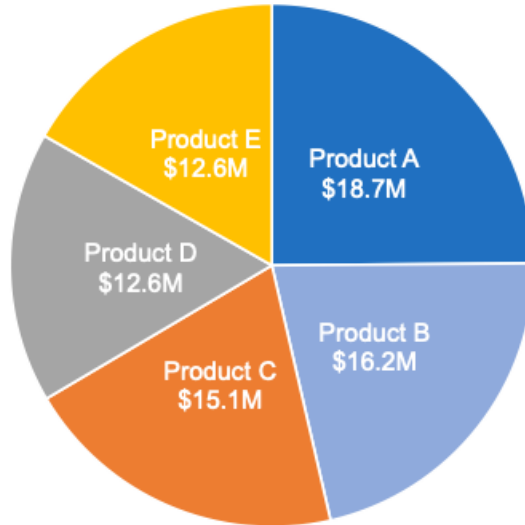
% OF ANDERSON HALL RESIDENTS



Pie Chart : Color

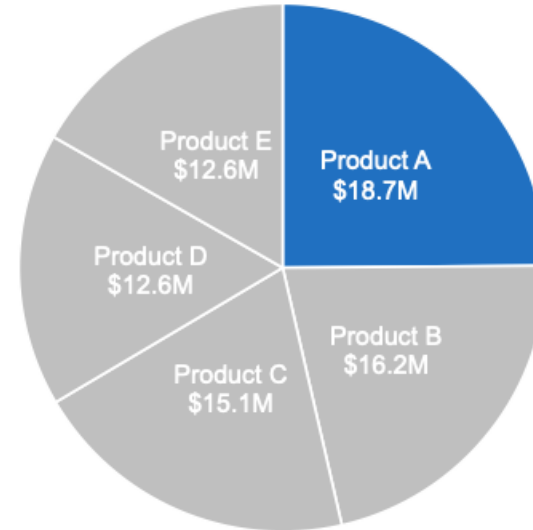
Sales by product

TOTAL = \$75.2M



Sales by product

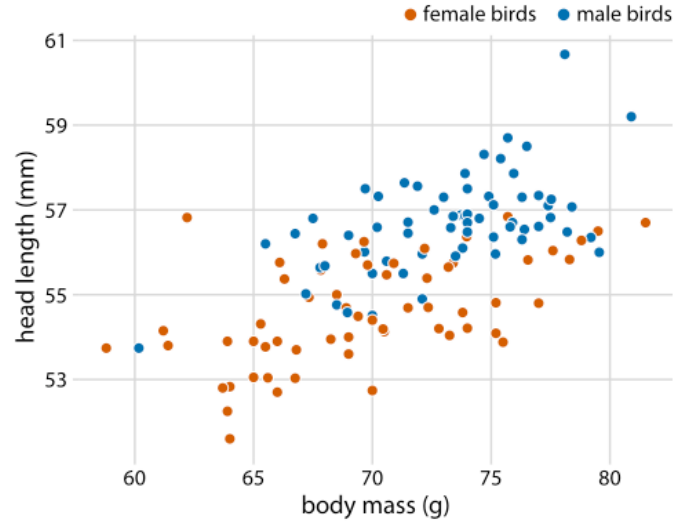
TOTAL = \$75.2M



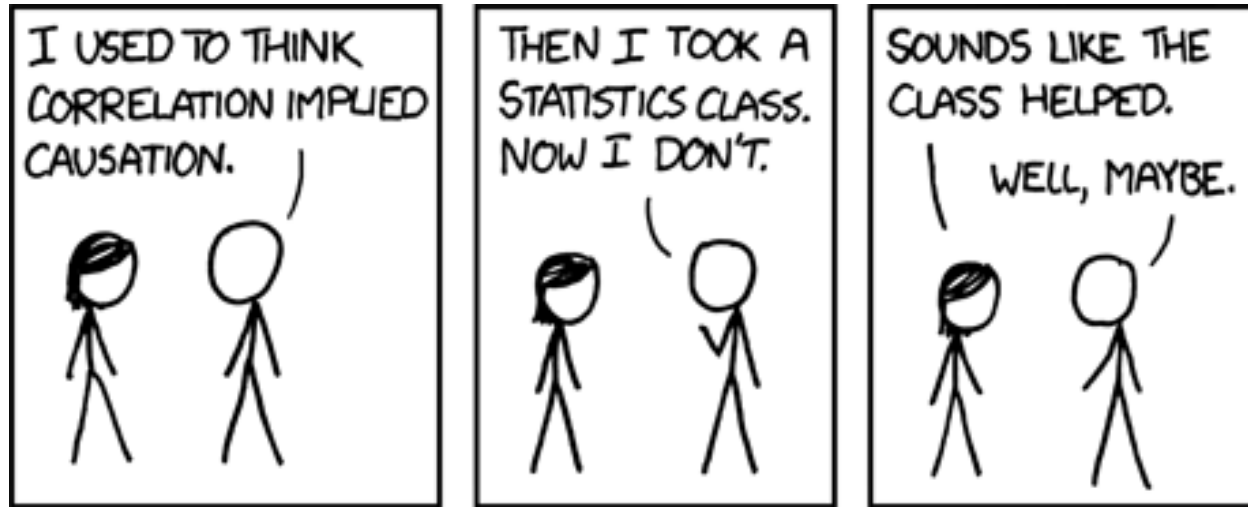
Scatter and Bubble

Scatter Plot

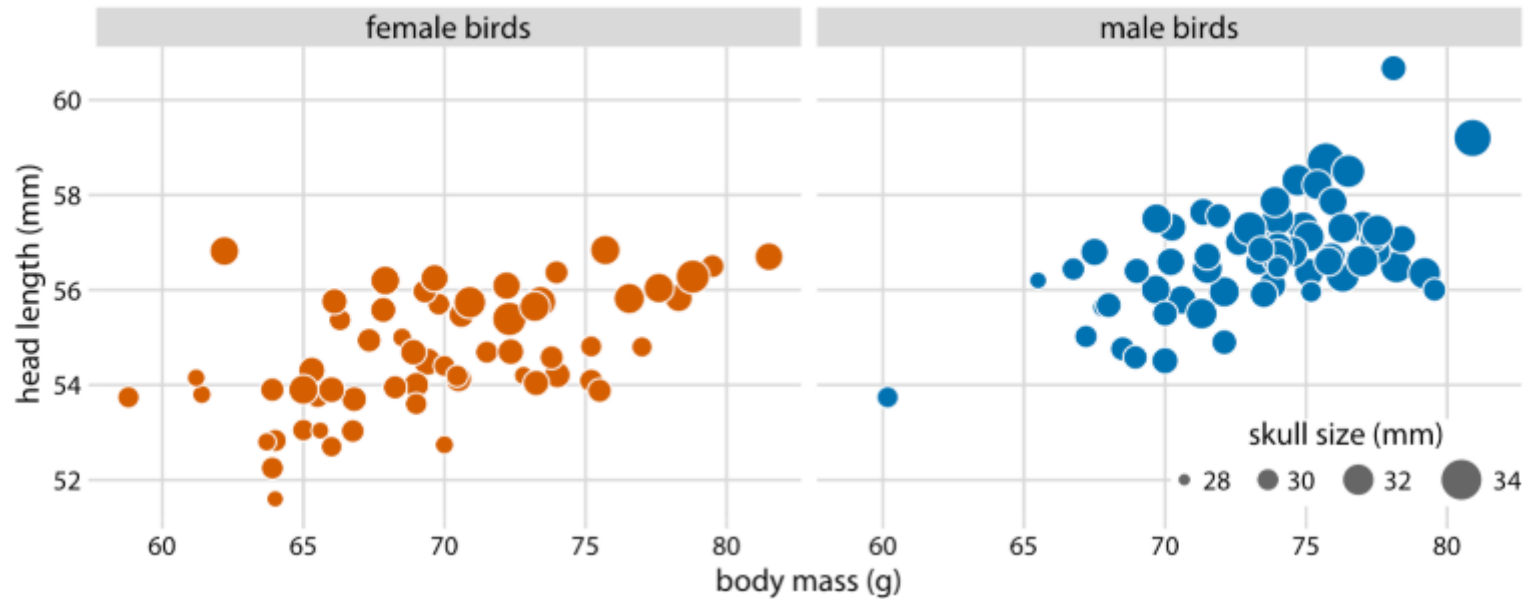
- Show relationship between two numerical variables
- Identify
 - Trend
 - Relationship strengths
 - Clusters or outliers



Association does not imply causation



Bubble Plot



Line



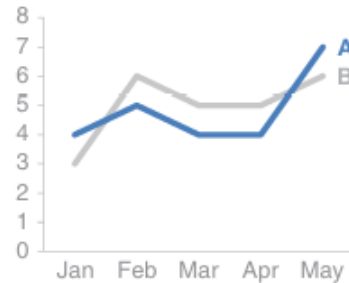
Line Chart

- Connect series of continuous data points using a line – some unit of time or data with natural order
- Track trends and patterns
- Can display both single or many lines

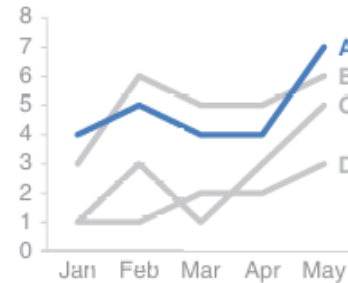
Single series



Two series

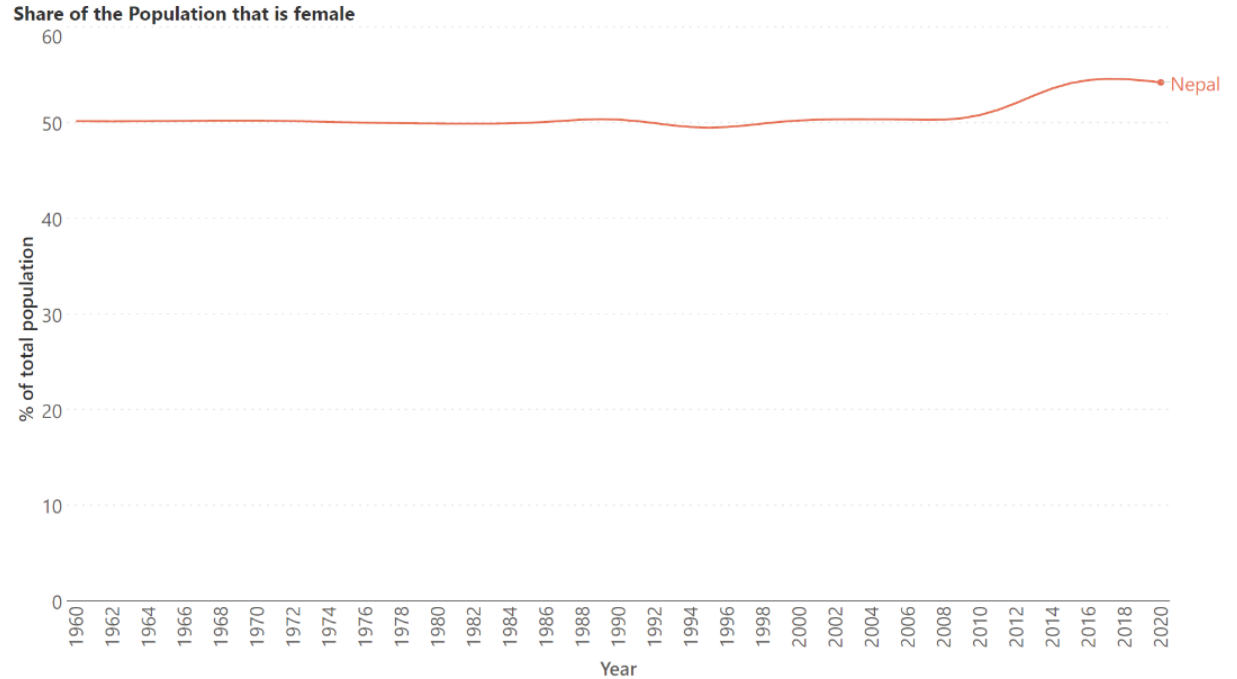


Multiple series



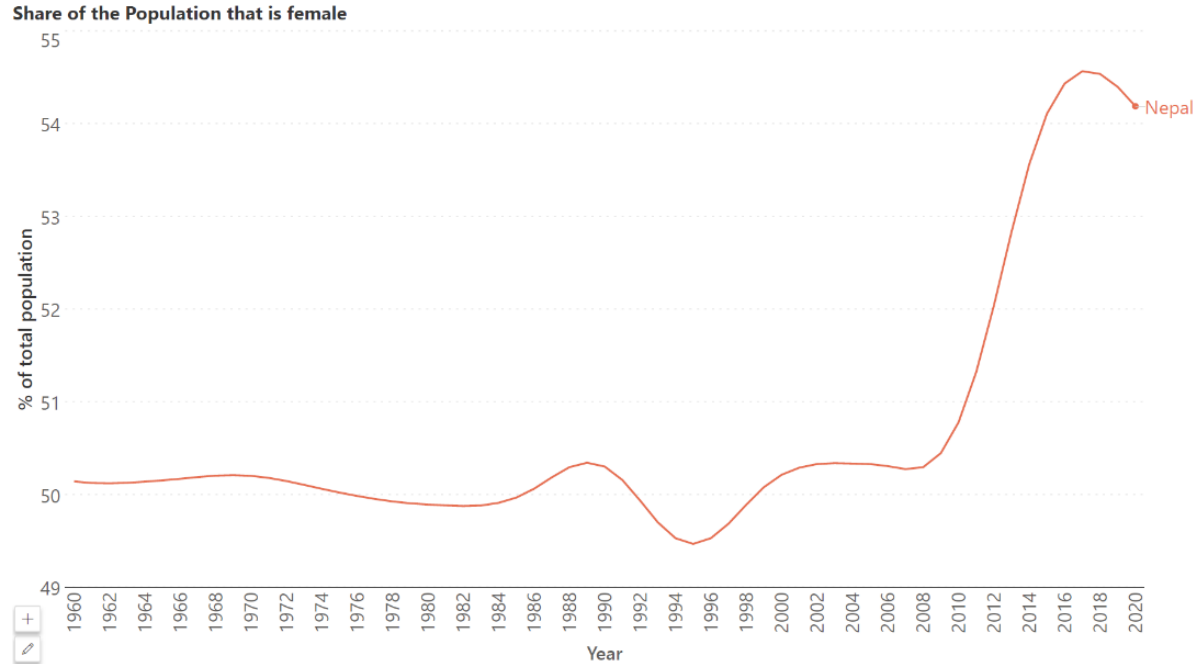
Line Chart

Baselines do not have to start at zero



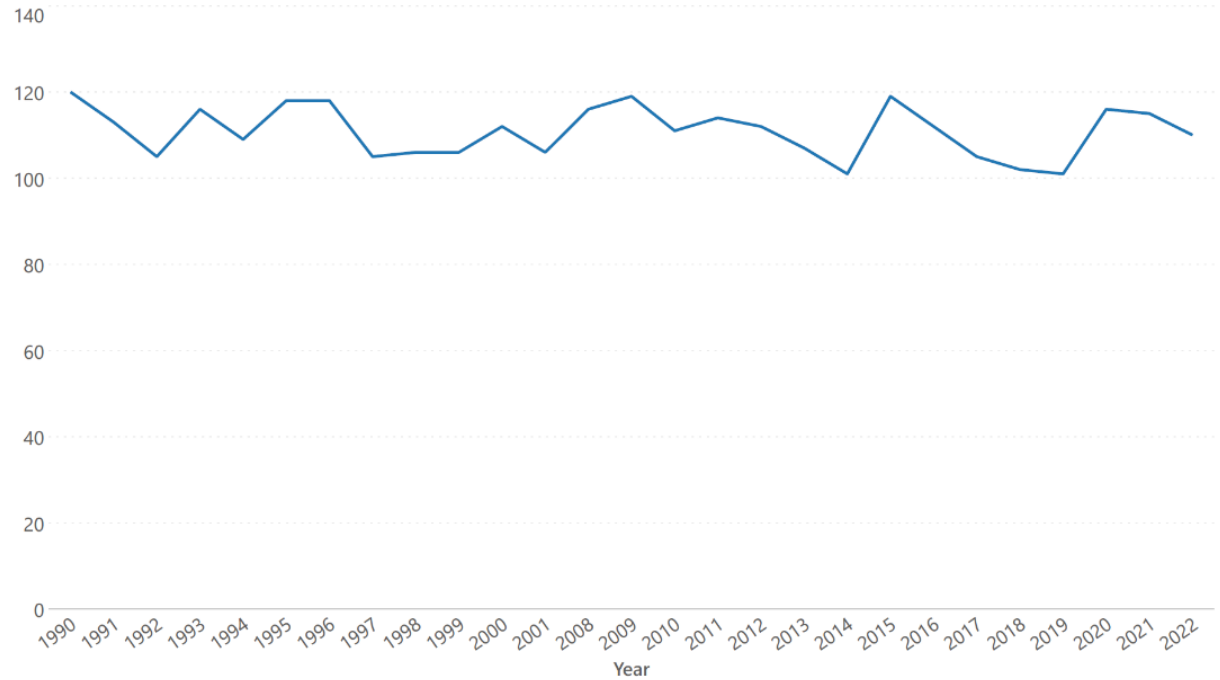
Line Chart

Baselines do not have to start at zero



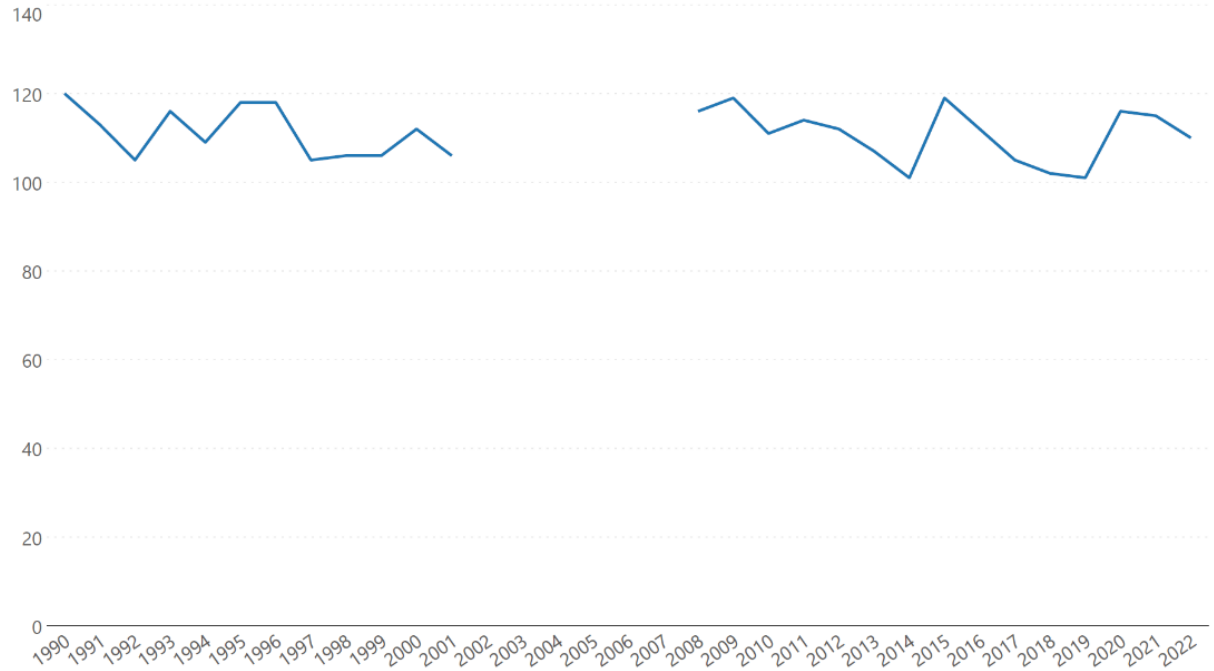
Missing Data

Identify what is wrong with this graph?



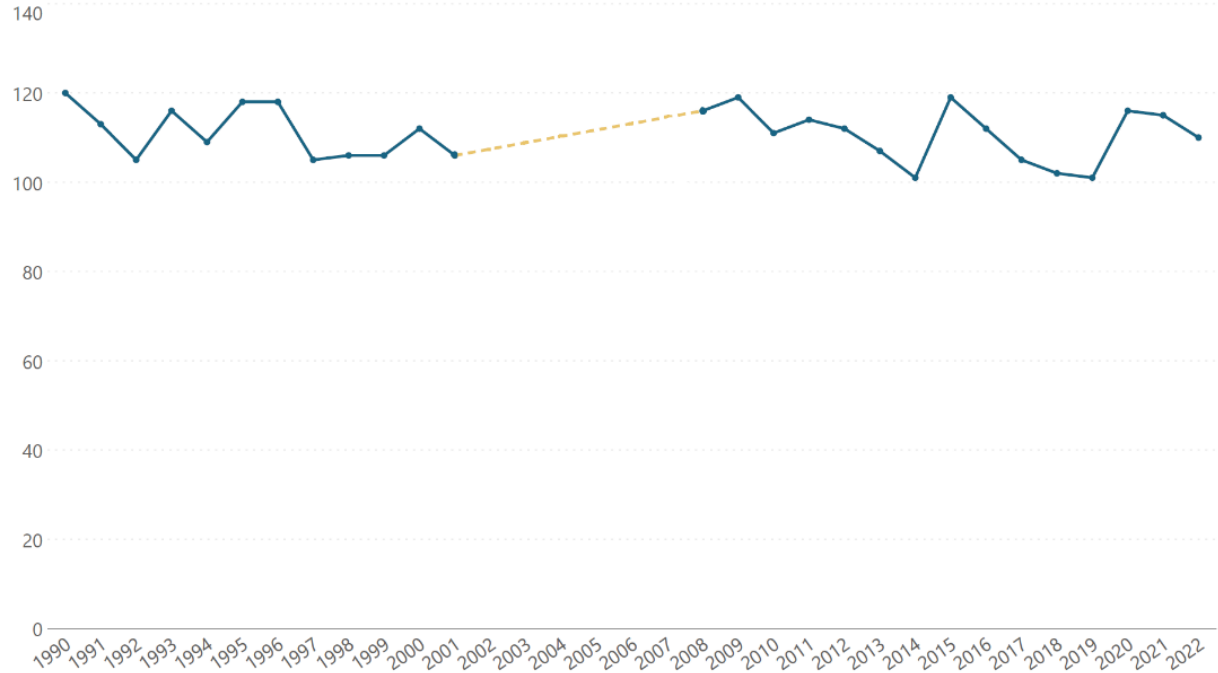
Missing Data

Missing data

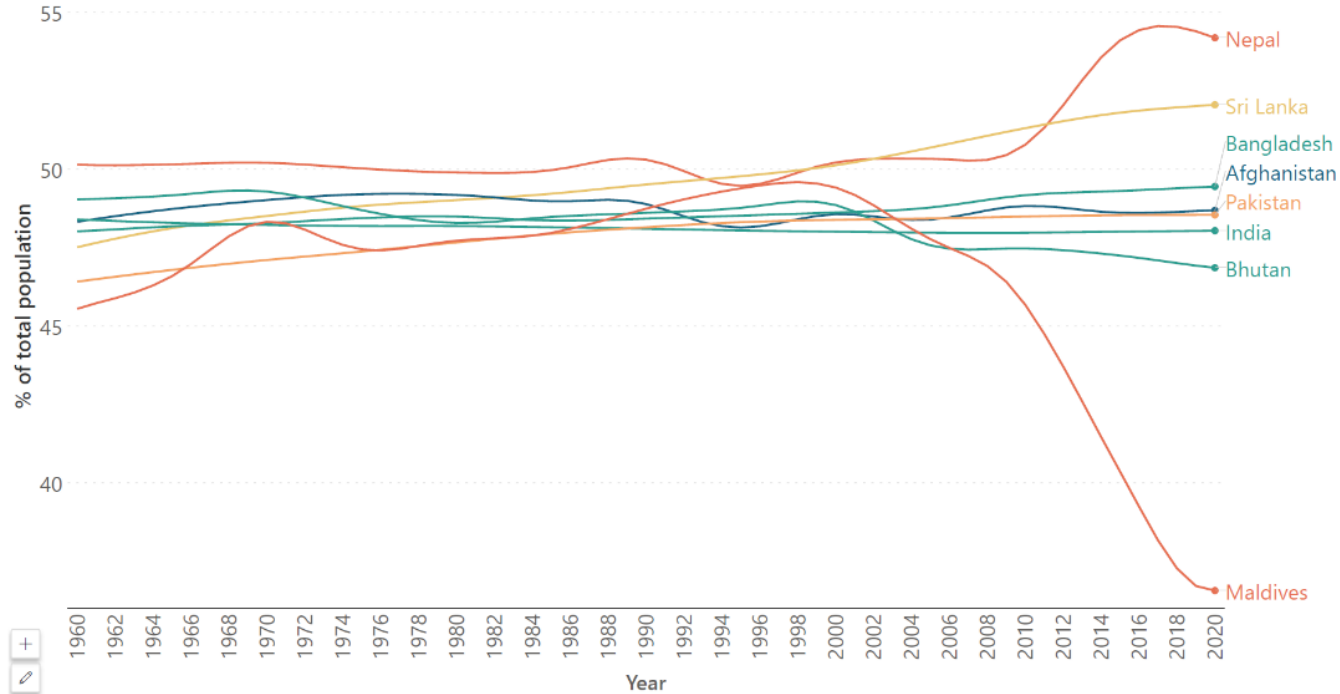


Missing Data

Use dotted or dashed line to indicate missing data

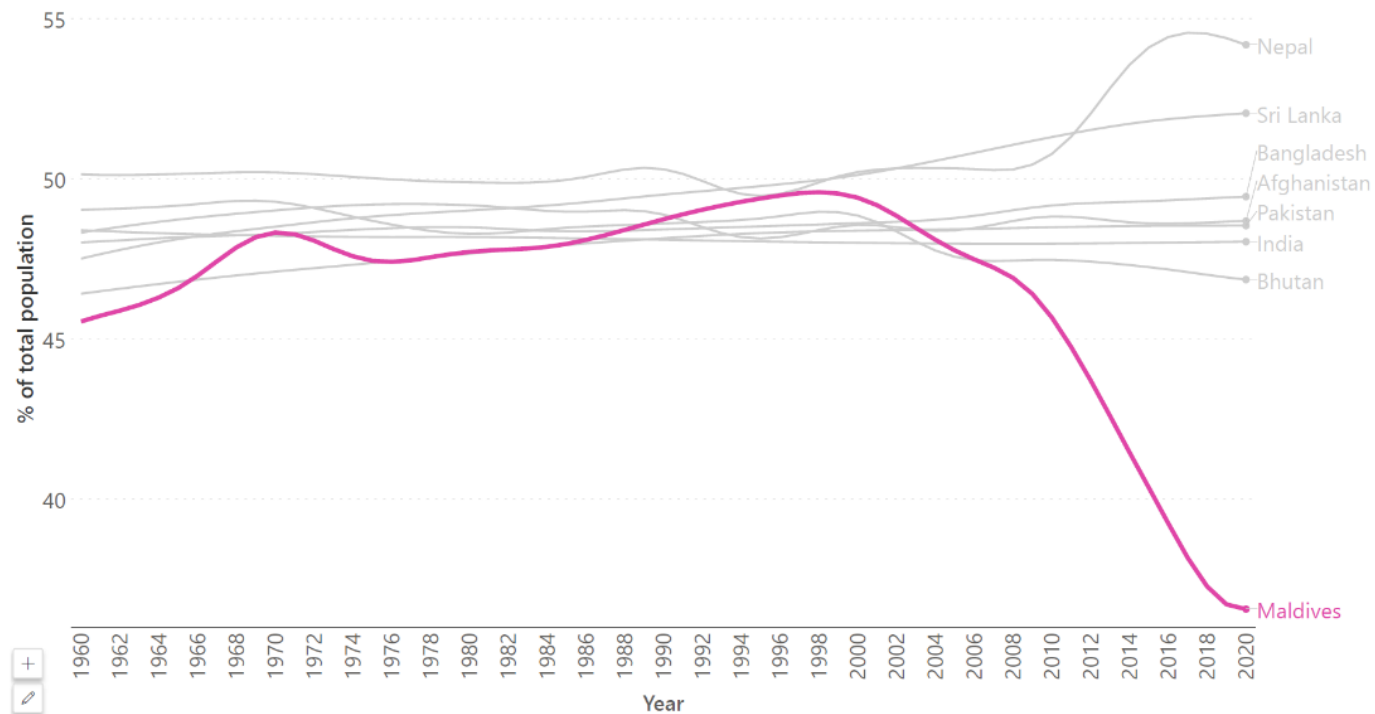


Share of the Population that is female

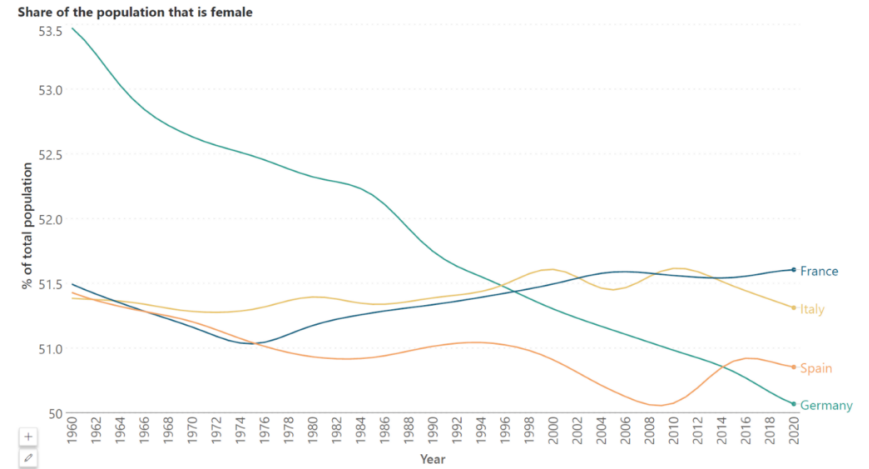
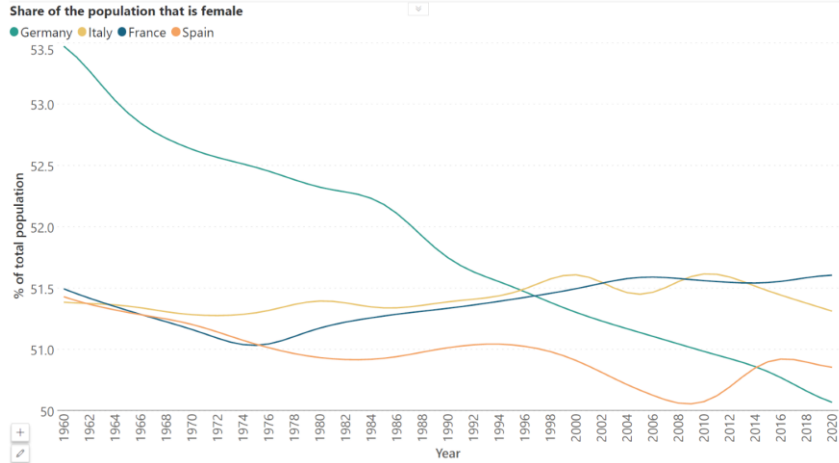


What are the issues with this line chart?

Share of the Population that is female



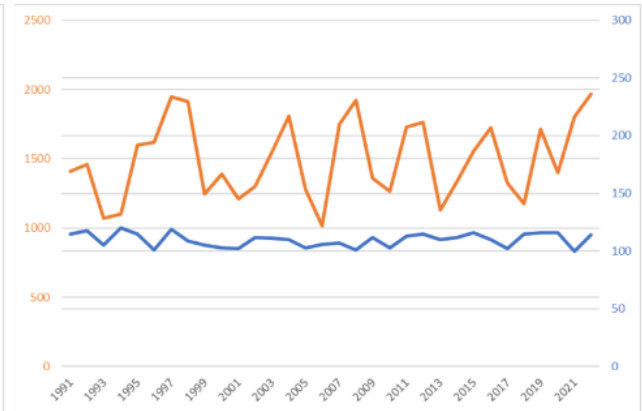
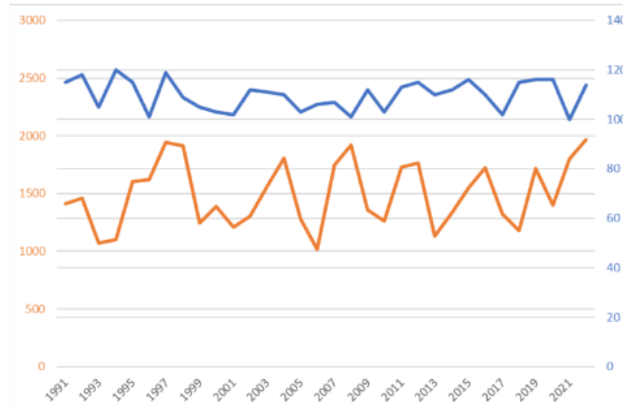
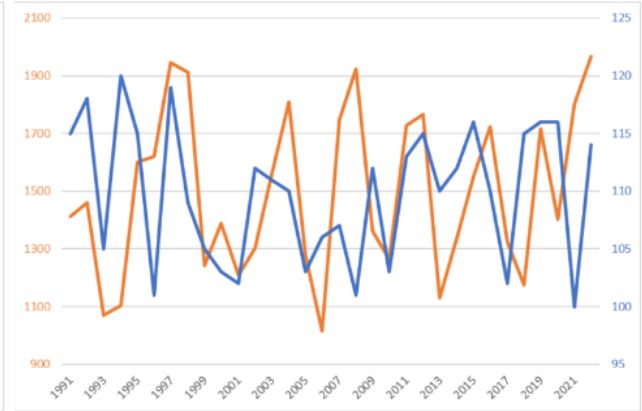
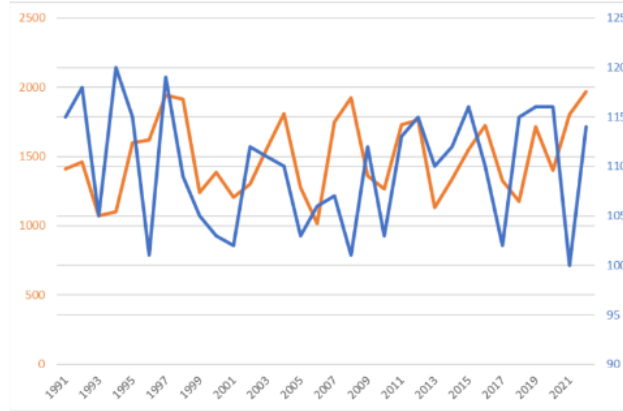
Label



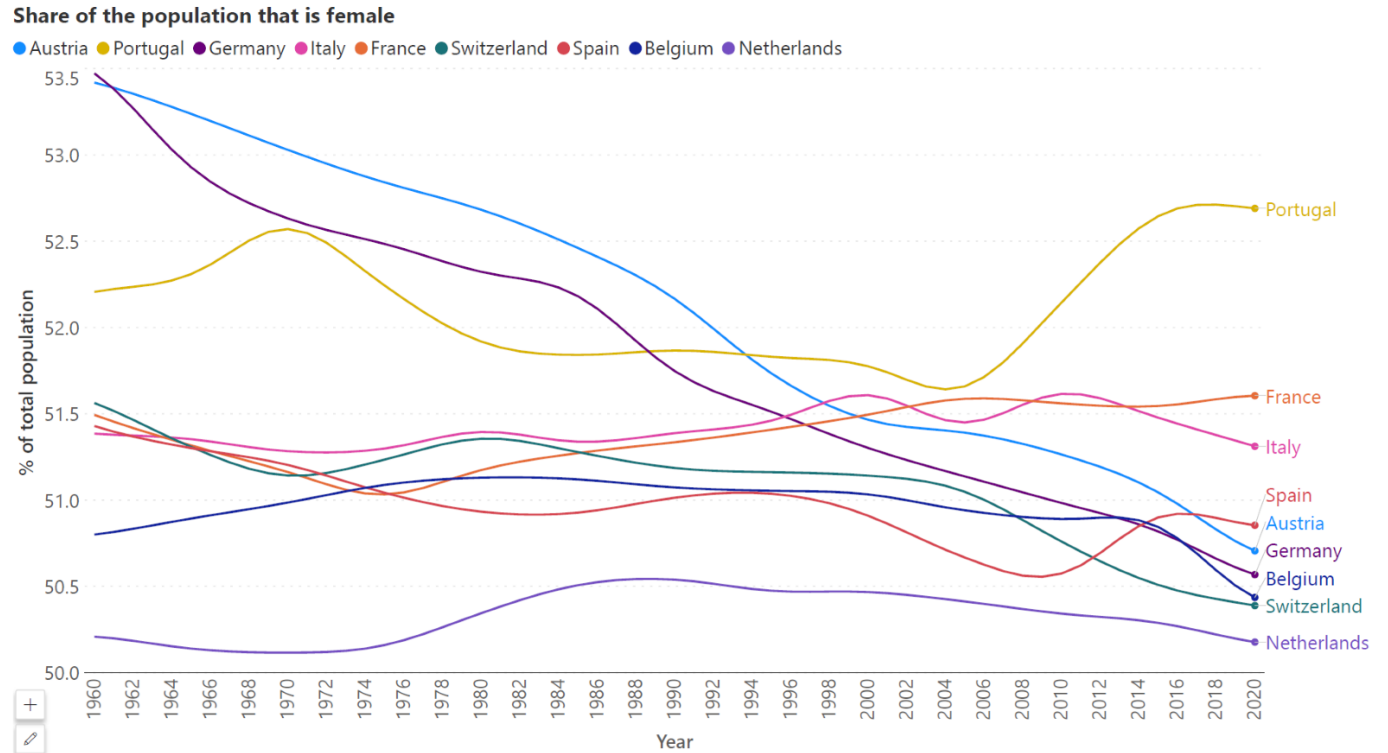
Avoid Dual Axis

Check the y-axis scale and starting point.

What's wrong here?

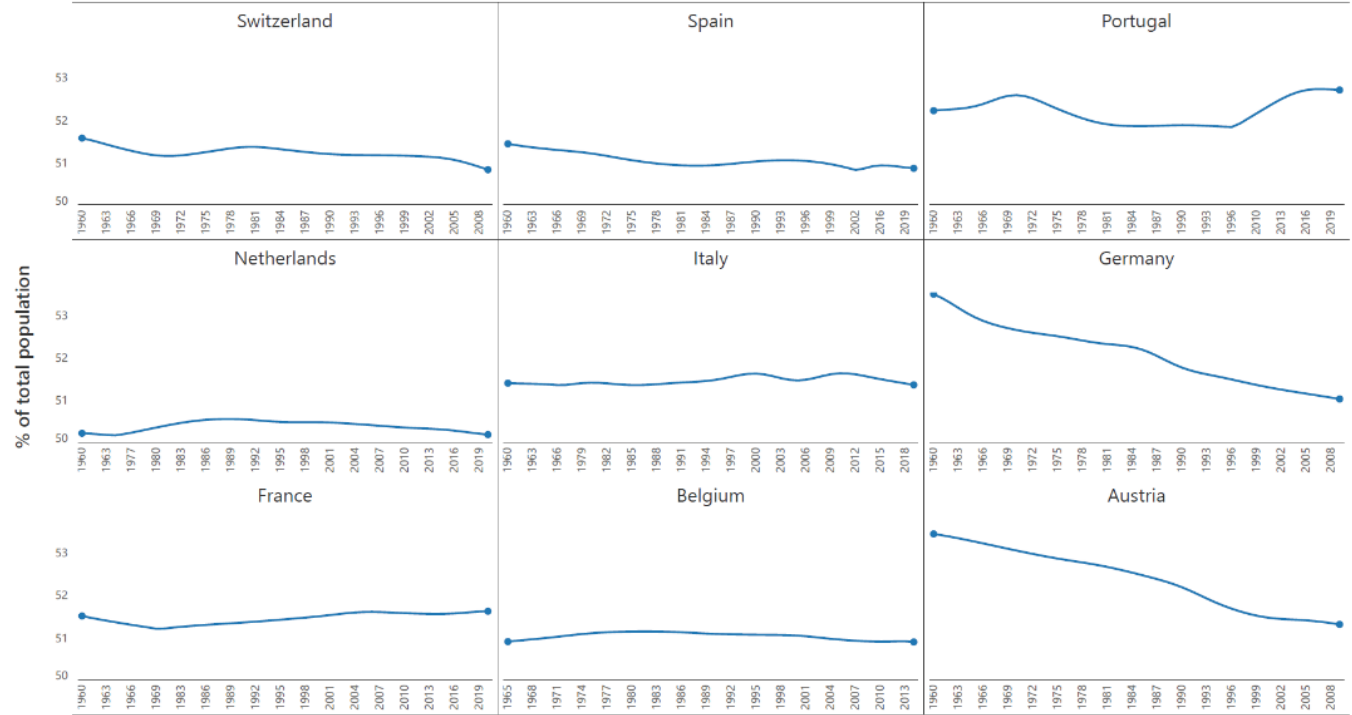


Spaghetti charts



Spaghetti charts

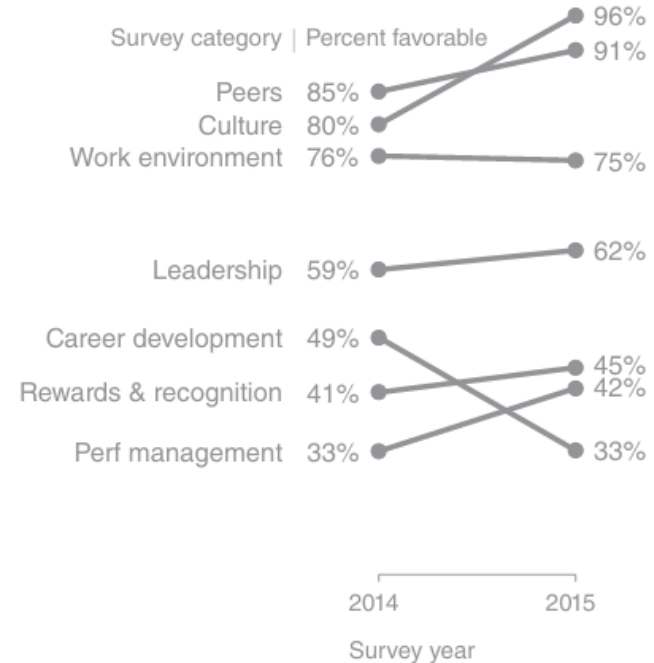
Share of the Population that is female



Slope Graph

- Useful when having two time periods or points of comparison.
- Show relative increases or decreases across categories

Employee feedback over time

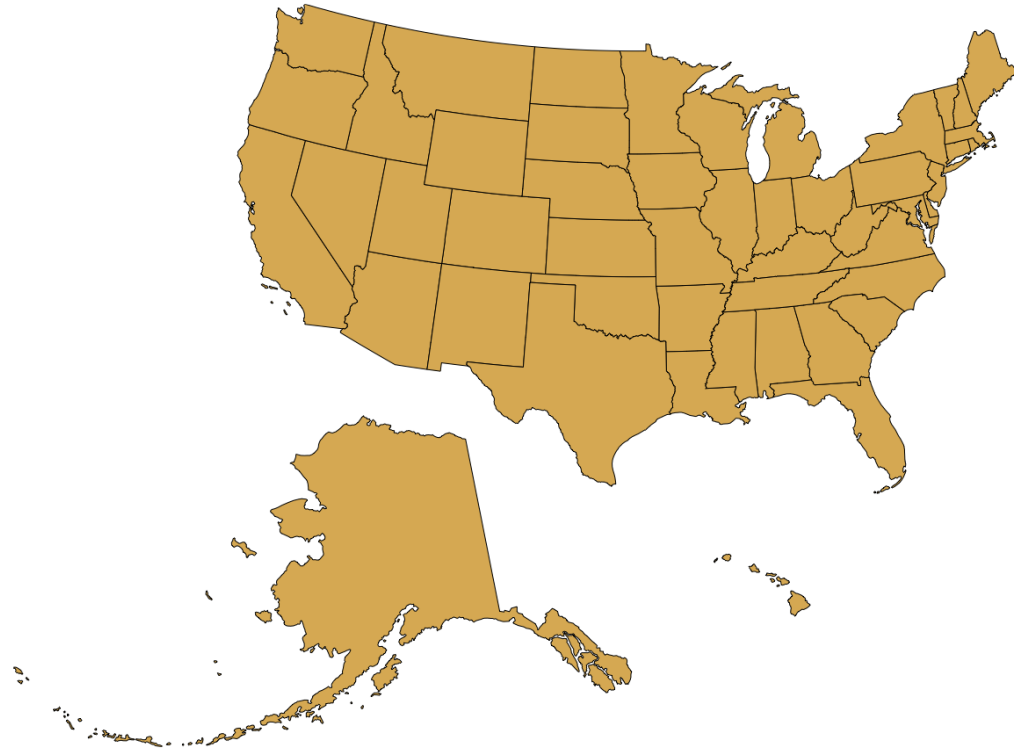


Geospatial Data

Maps

- **Geographic** map – charts that shows characteristics and the arrangement of the geography of our physical reality.
- **Choropleth** map – geographic map that uses shades of a color, unique colors or symbols to indicate quantitative or categorical variables by geographic region or area.

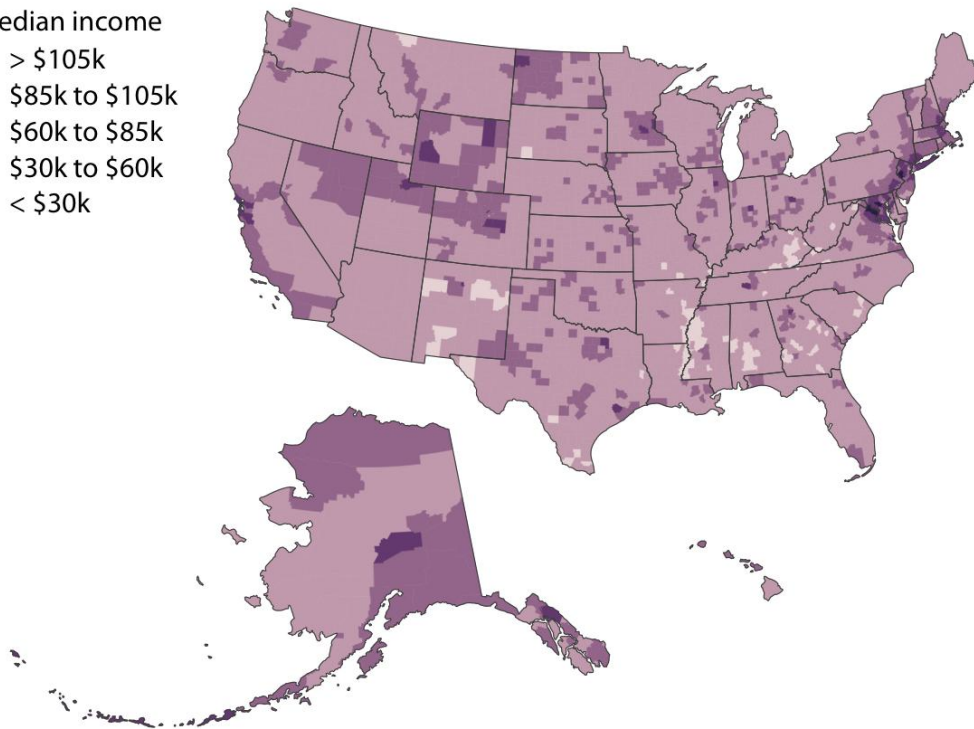
Geographic



Choropleth

median income

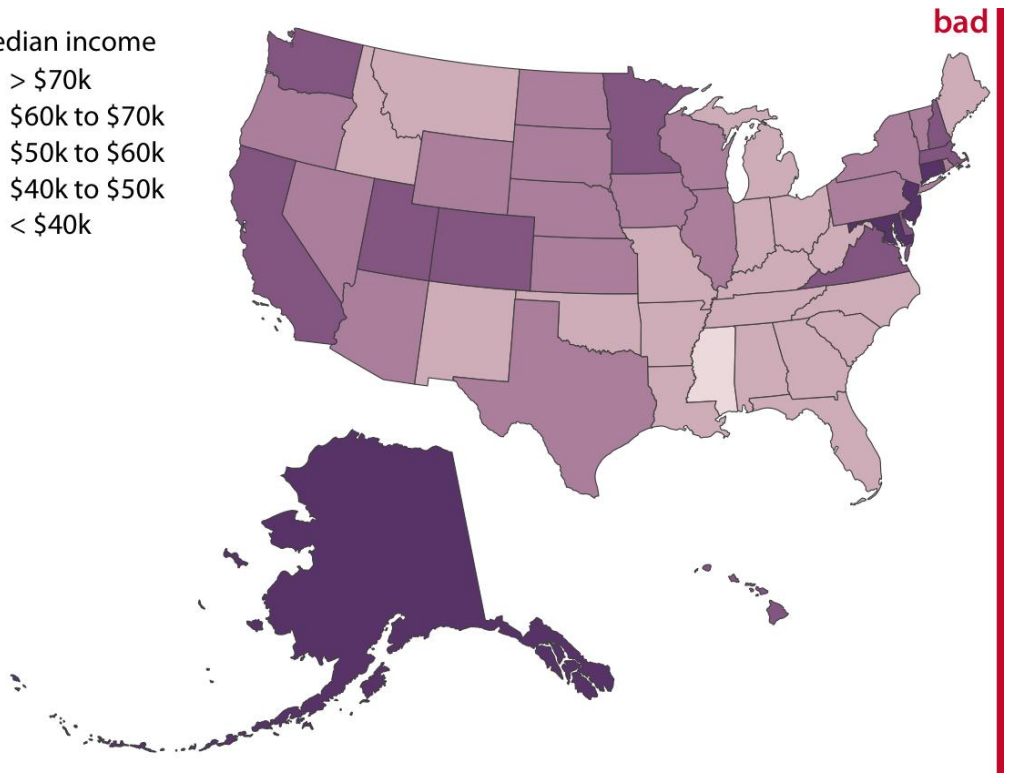
- > \$105k
- \$85k to \$105k
- \$60k to \$85k
- \$30k to \$60k
- < \$30k



Choropleth

median income

- > \$70k
- \$60k to \$70k
- \$50k to \$60k
- \$40k to \$50k
- < \$40k



Cartogram Heatmap

median income

- > \$70k
- \$60k to \$70k
- \$50k to \$60k
- \$40k to \$50k
- < \$40k

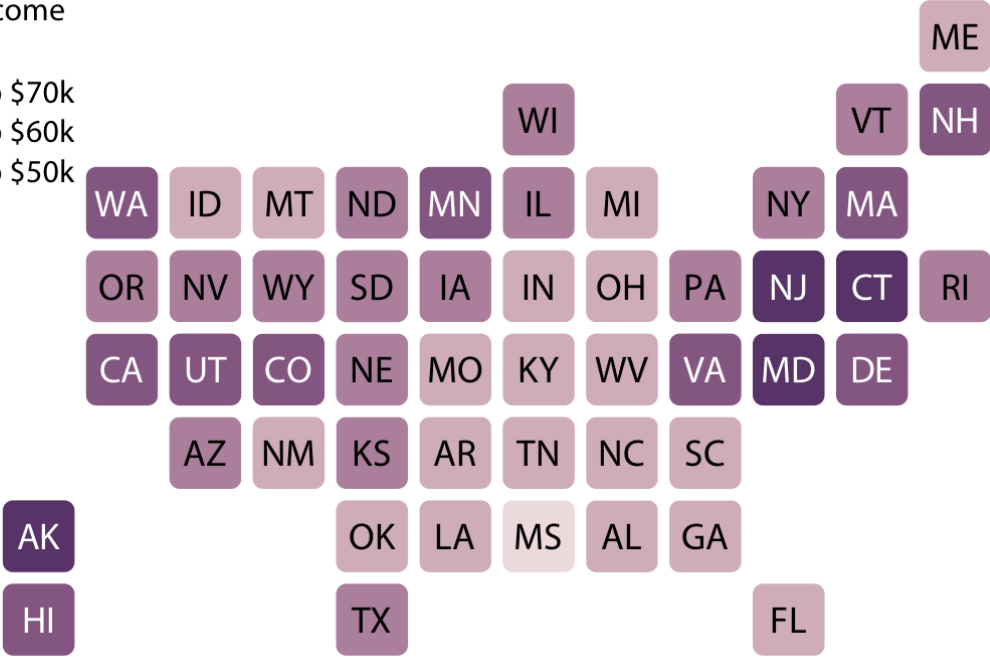
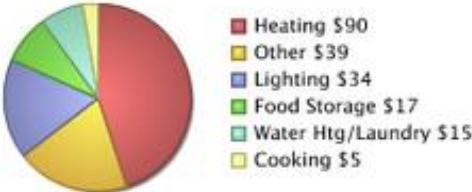


Chart Re-Design

A Pie Chart Redesign

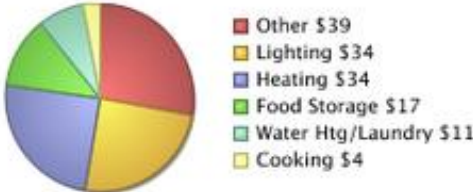
Original Version by Duke Energy

Total Costs 2/25/2014 to 3/25/2014



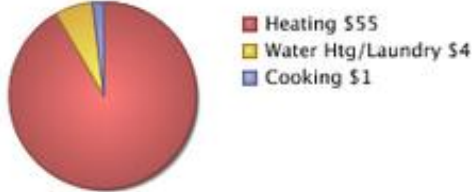
Total Electricity Gas

Electricity Costs 2/25/2014 to 3/25/2014



Total Electricity Gas

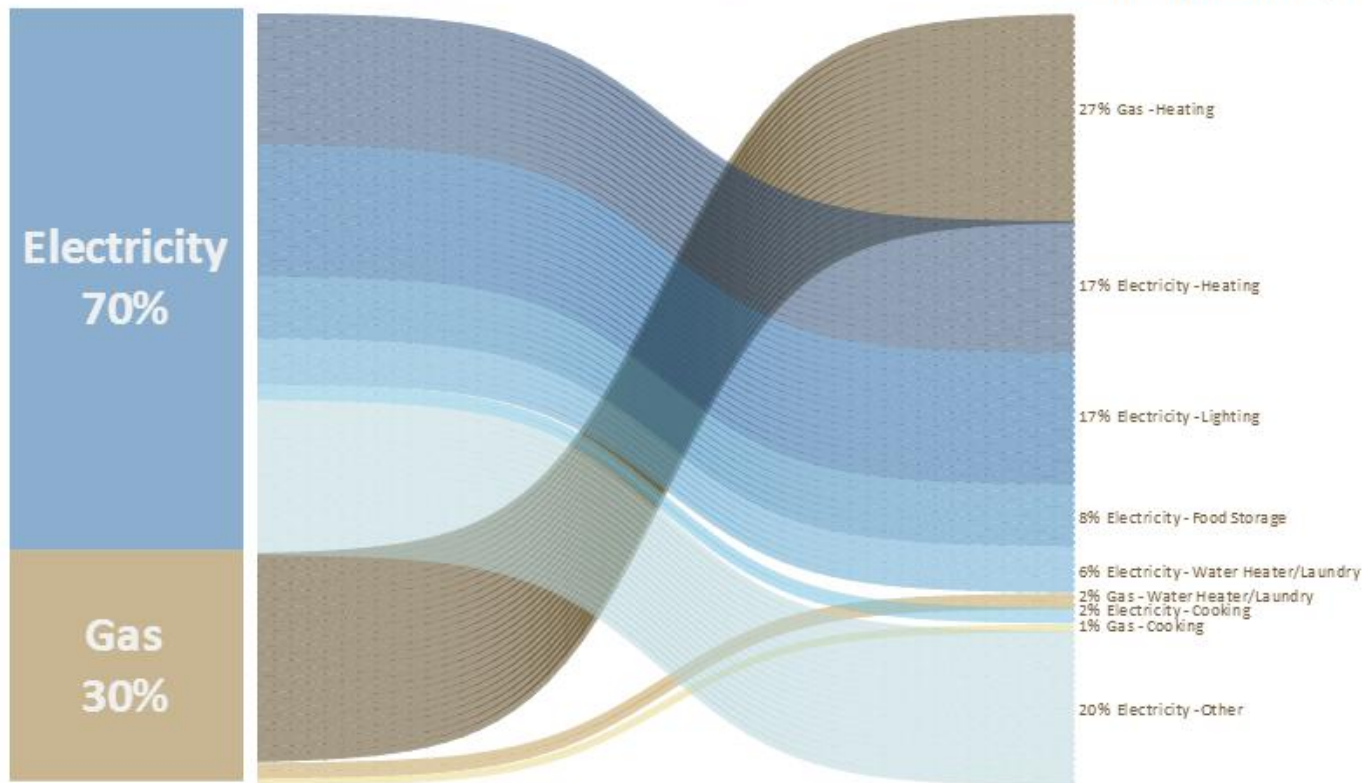
Gas Costs 2/25/2014 to 3/25/2014



Total Electricity Gas

My Monthly Energy Bill Redesigned

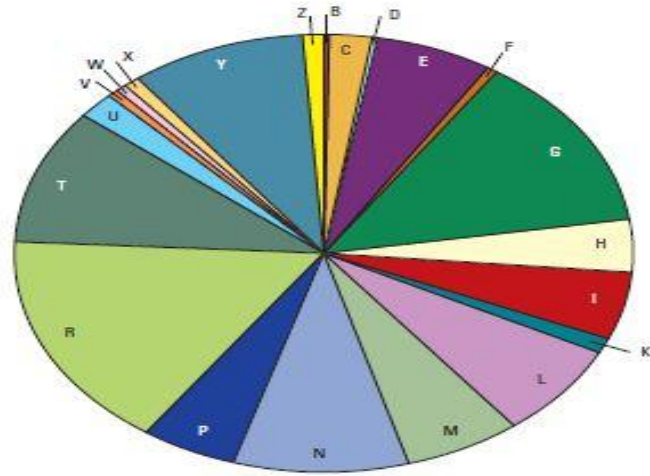
data + science
= transforming data to insight



Redesigned Using a Sankey Diagram

**Agency
Operations
Information**

If you indicated your debt collection agency used collection software, which type?

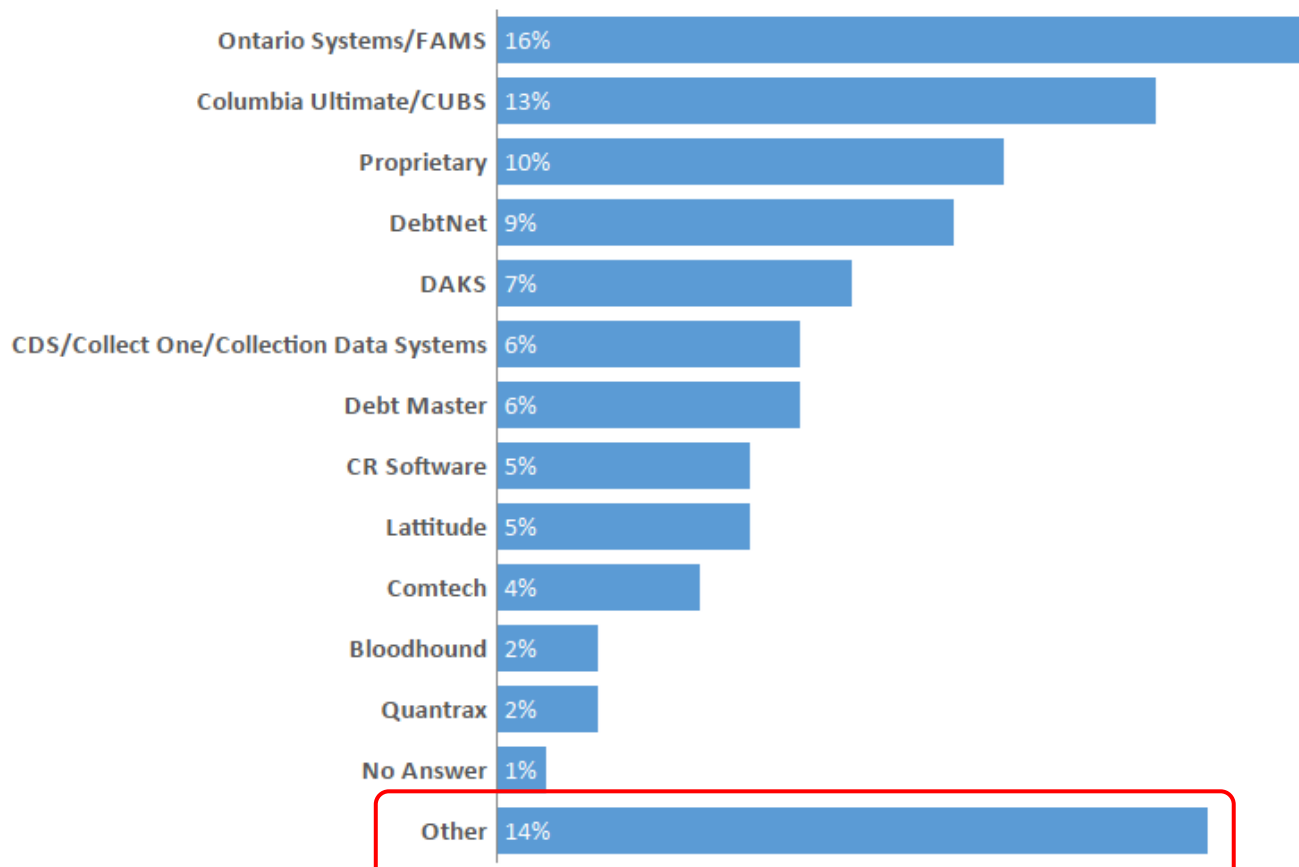


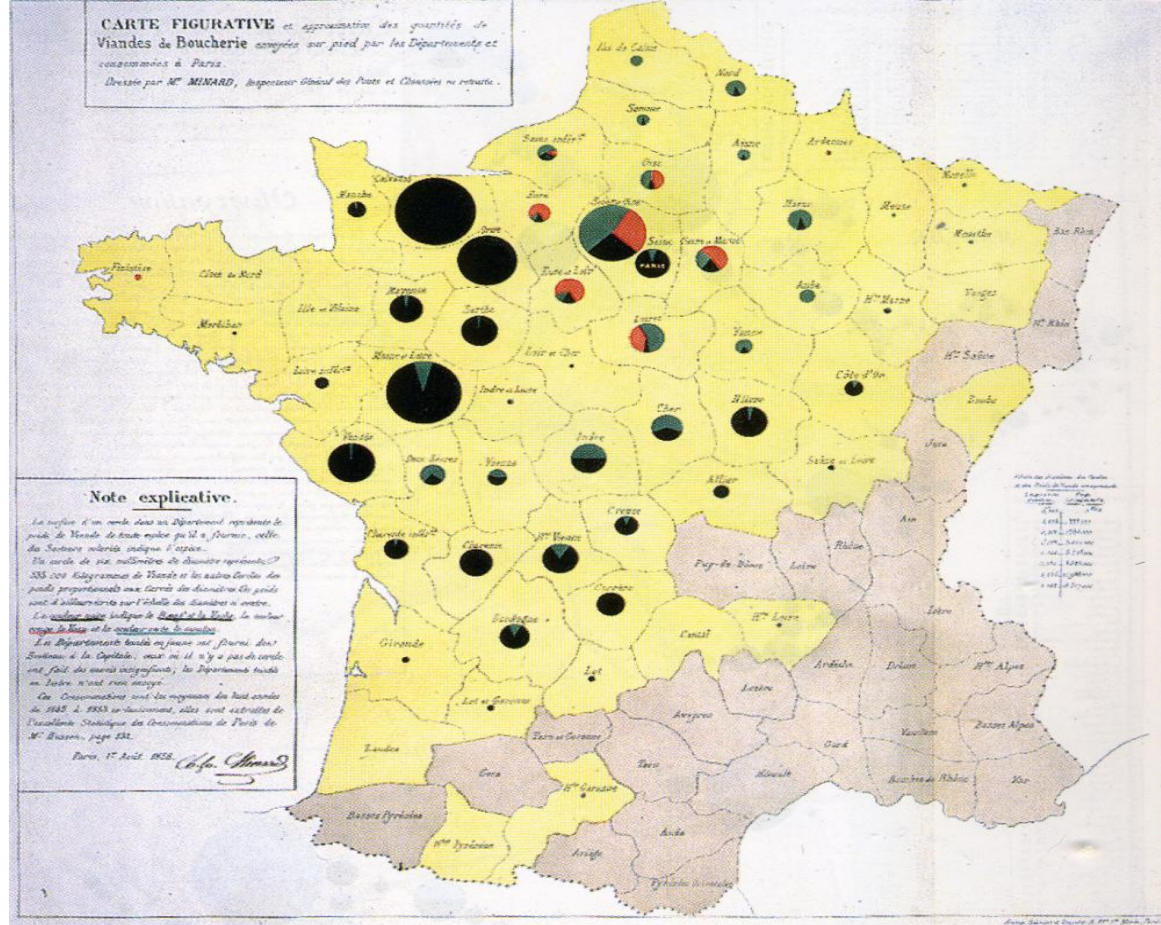
Base: 214 respondents

A	ADS	0 %	N	DebtNet	9 %
B	Advantage	<1 %	O	OTC	0 %
C	Bloodhound	2 %	P	Latitude	5 %
D	C.A.M.S.	<1 %	Q	Northridge Loan System	0 %
E	CDS/CollectOne/ Collection Data Systems	6 %	R	Ontario Systems/FAMS	16 %
F	Collection Master	<1 %	S	PaPlus	0 %
G	Columbia Ultimate/CUBS	13 %	T	Proprietary	10 %
H	Comtech	4 %	U	Quantrax	2 %
I	CR Software	5 %	V	RZ Designs	<1 %
J	Credit Soft	0 %	W	Unisource	<1 %
K	Cyclone	1 %	X	WinDebt	1 %
L	DAKCS	7 %	Y	Other	9 %
M	DebtMaster	6 %	Z	No Answer	1 %

Source:
American
Collectors
Association

If you indicated your debt collection agency used collection software, which type?

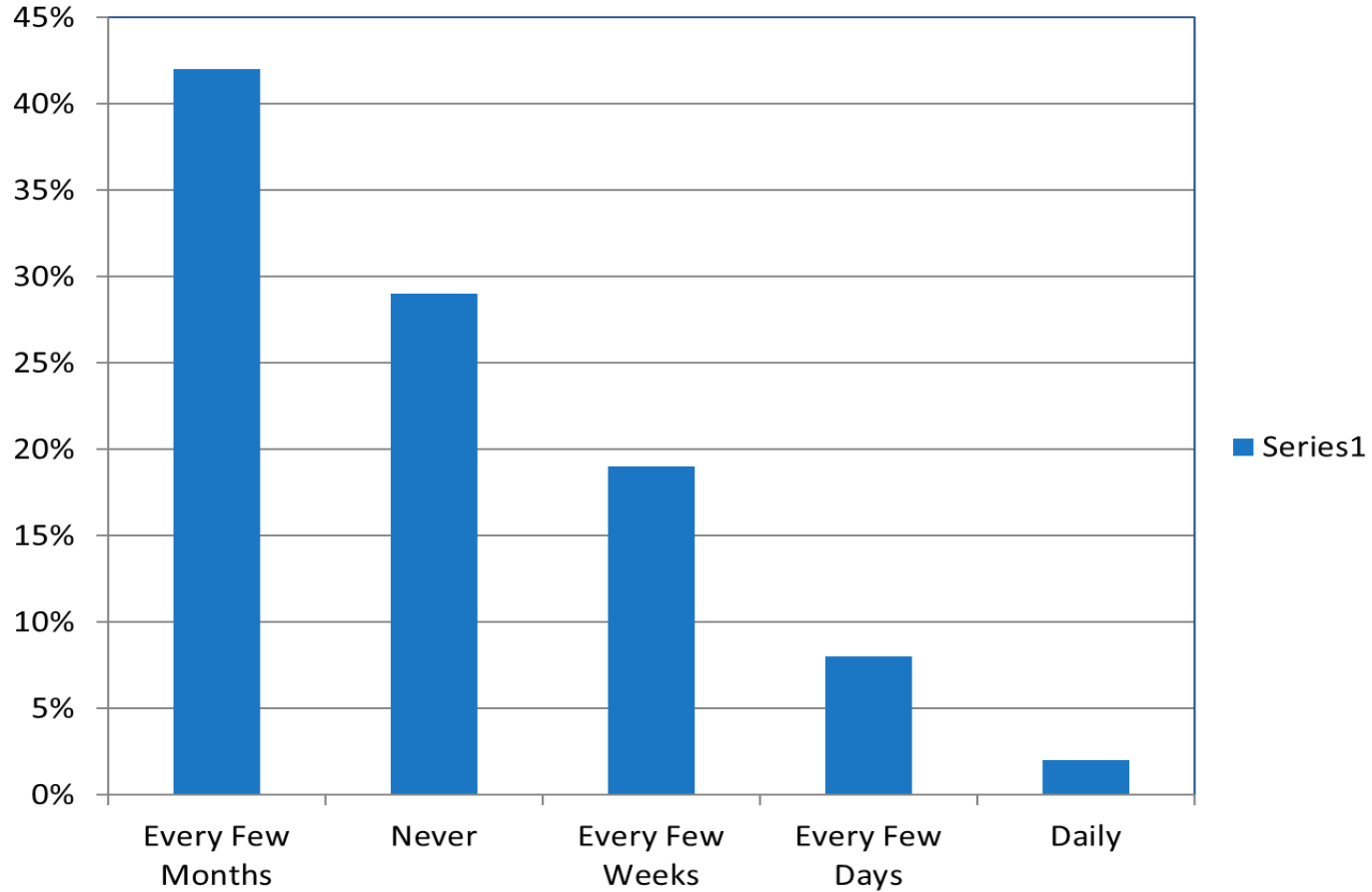




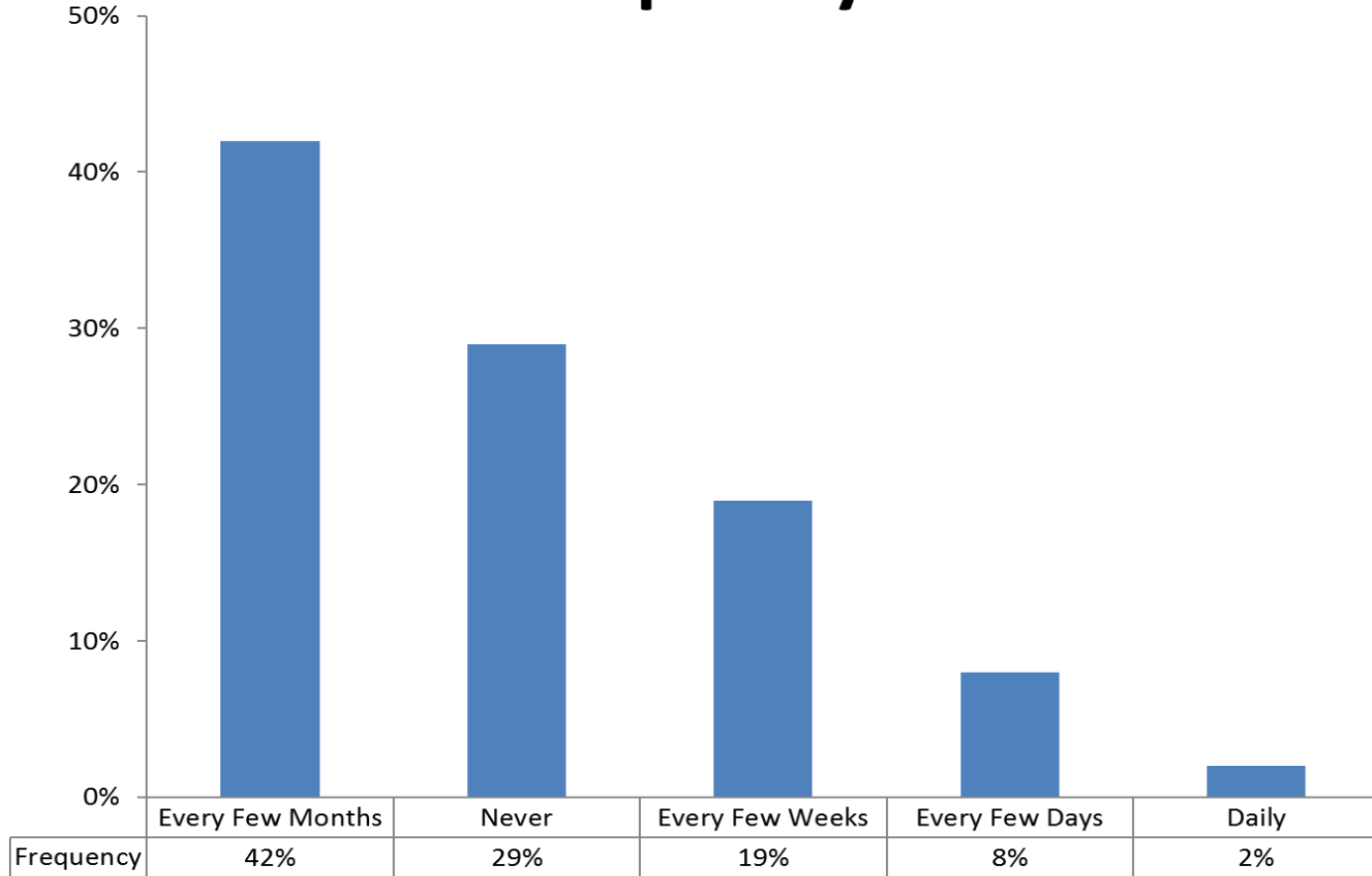
Pie Charts on a map work!
 It's good for visualizing a part-to-whole relationship

Supply of meat to Paris
 - Charles Joseph Minard, 1858

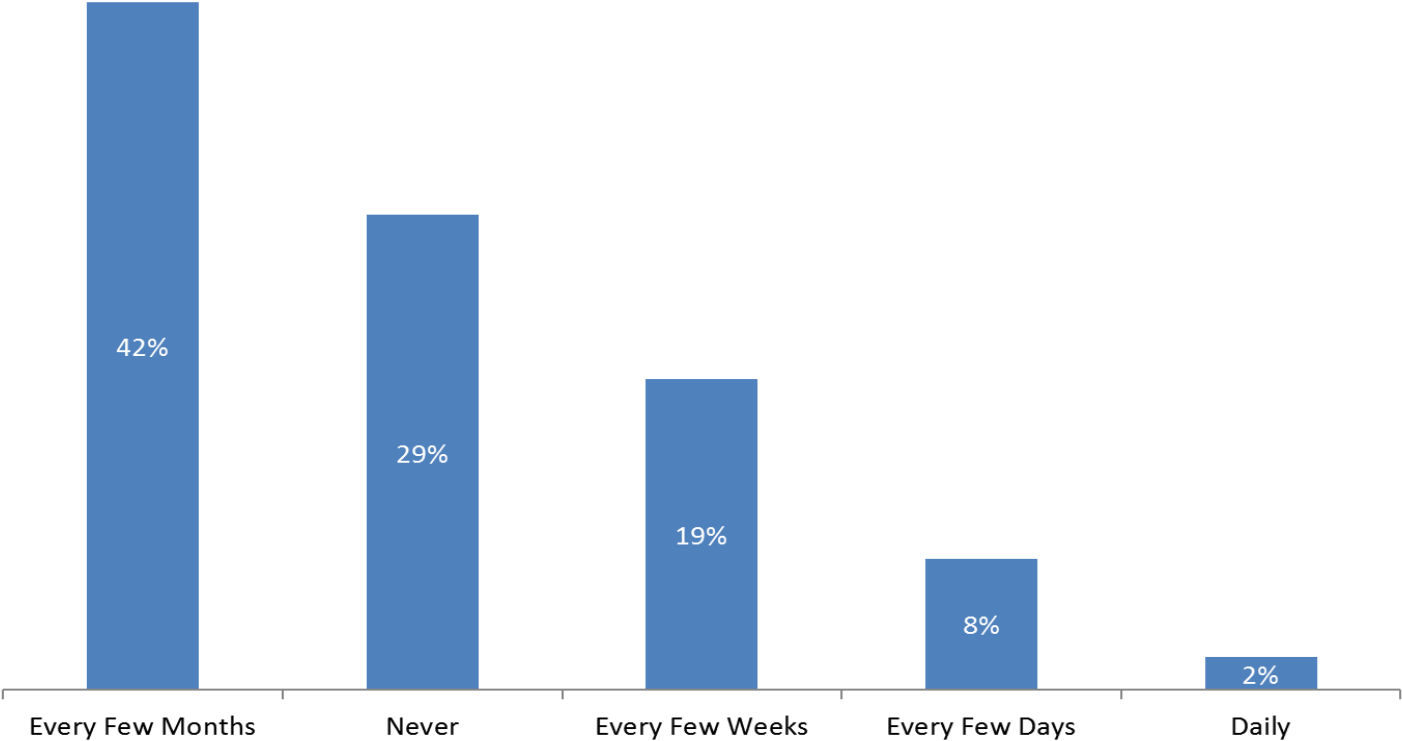
Default in Excel



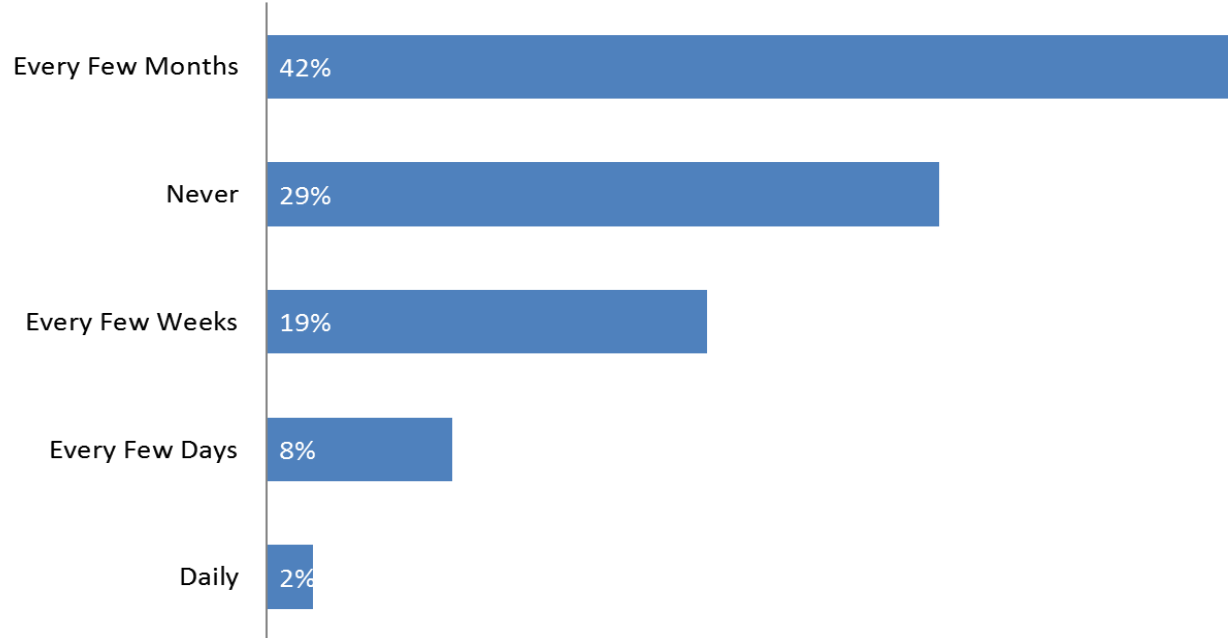
Frequency



Frequency

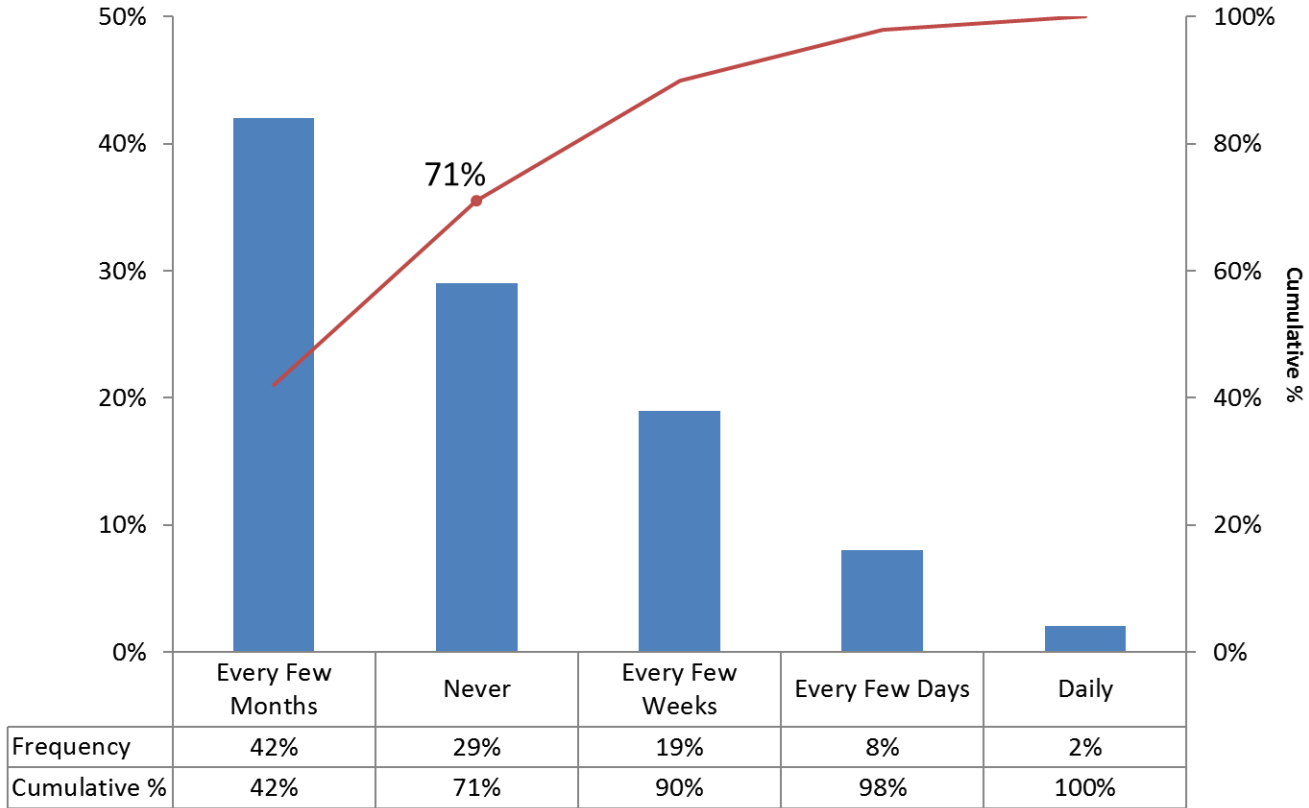


Frequency



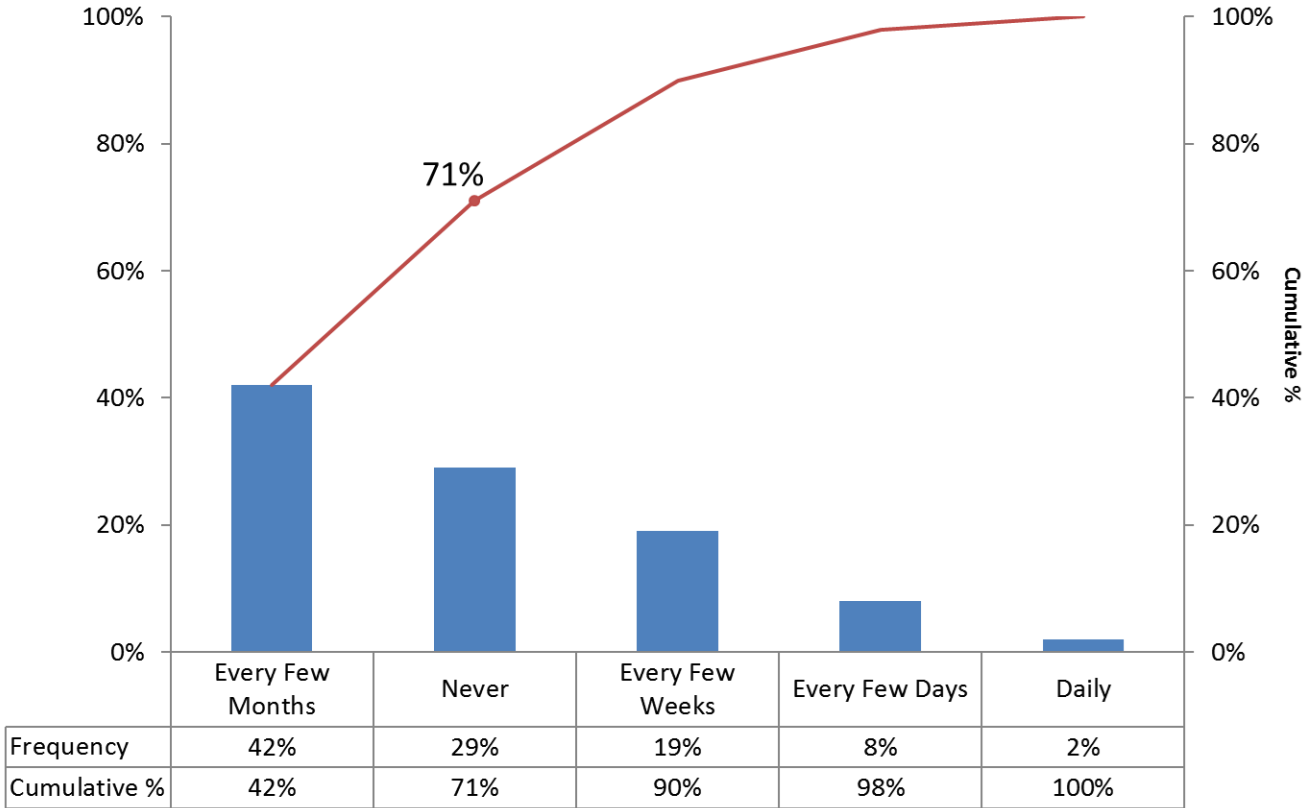
Frequency

71% of occurrences are caused by low frequency



Frequency

71% of occurrences are caused by low frequency



End of Lecture