Northwestern Medicine®
Feinberg School of Medicine

# Principles of Effective Data Visualization

Karen Gutzman
Head, Research Assessment and Communications
Galter Health Sciences Library & Learning Center
Feinberg School of Medicine
Northwestern University
karen.gutzman@northwestern.edu

# Agenda

- **Know your audience -** Tailor your story to who's listening
- **Foundation first -** Clean data structure principles
- **Visual perception -** How our brains decode information
- **Chart encyclopedia -** Choosing the right visualization for your story
- **Design principles -** Creating compelling information experiences
- **Ethics matter -** Responsibility and truth in visual communication
- **Tools of the trade -** From Excel to advanced platforms
- **From principles to practice -** Building your expertise

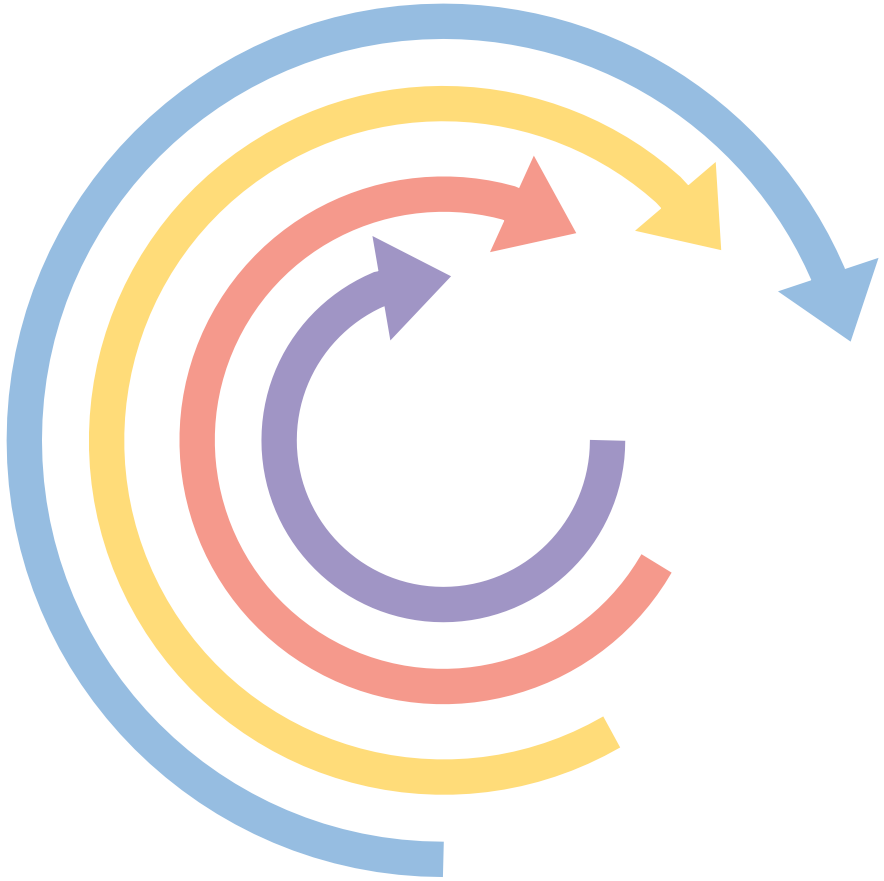
# Why Data Visualizations Matters

1. **Faster decision-making**
   Spot trends and outliers quickly in any dataset.

2. **Clear communication**
   Present complex information in an accessible format

3. **Team alignment**
   Ensure everyone interprets data consistently

4. **Quality assurance**
   Visual patterns reveal data issues and anomalies



# Know Your Audience, Shape Your Story

Effective visualization starts with understanding who you're communicating with

## The Executive (5-minute attention span)

- **What they want:** Bottom line impact, key trends, actionable insights
- **Best practices:** Lead with conclusions, use big numbers and simple comparisons
- **Chart choices:** Clean bar charts, trend lines, executive dashboards
- **Avoid:** Complex methodologies, detailed breakdowns, technical jargon

## The Analyst (Detail-oriented)

- **What they want:** Methodology, data sources, statistical significance
- **Best practices:** Show your work, include confidence intervals, provide data tables
- **Chart choices:** Scatter plots, box plots, detailed multi-series charts
- **Avoid:** Oversimplified summaries, hiding uncertainty

## The General Public (Diverse backgrounds)

- **What they want:** Clear takeaways, relatable context, engaging stories
- **Best practices:** Use familiar references, progressive disclosure, clear annotations
- **Chart choices:** Simple comparisons, infographics, interactive elements
- **Avoid:** Technical terms, assumption of statistical literacy

## The Technical Team (Implementation focus)

- **What they want:** Specifications, processes, operational metrics
- **Best practices:** Show workflows, highlight bottlenecks, include benchmarks
- **Chart choices:** Process flows, performance dashboards, diagnostic charts
- **Avoid:** High-level summaries without operational details

# Data Structure Matters

| | ✓ Consistent |
|---|---|
| **General Examples** | Same formats, naming conventions, time periods. *Example: 01/02/2025 not a mix of "Jan 2025", "Human Resources" vs. "HR"* |
| **Excel** | **TEXT()** to standardize dates: =TEXT(A1,"MM/DD/YYYY")<br>**SUBSTITUTE()** to standardize names: =SUBSTITUTE(A1,"HR","Human Resources") |
| **Python**<br>Pandas, Numpy, Scipy | pd.to_datetime(df['date'], format='%m/%d/%Y') to standardize dates |
| **R**<br>Dplyr, tidyr, stringr | str_replace_all(department, "HR", "Human Resources") to clean text |

# Data Structure Matters

| | ✓ Consistent | ✓ Complete |
|---|---|---|
| **General Examples** | Same formats, naming conventions, time periods. *Example: 01/02/2025 not a mix of "Jan 2025", "Human Resources" vs. "HR"* | No missing values that could mislead. *Example: Show "0" or "No Data" rather than blank cells Excel might skip* |
| **Excel** | **TEXT()** to standardize dates: =TEXT(A1,"MM/DD/YYYY") **SUBSTITUTE()** to standardize names: =SUBSTITUTE(A1,"HR","Human Resources") | IF(ISBLANK(A1),"No Data",A1) to replace blank cells that Excel might skip |
| **Python** Pandas, Numpy, Scipy | pd.to_datetime(df['date'], format='%m/%d/%Y') to standardize dates | df['compliance_rate'].fillna('No Data') instead of leaving NaN values |
| **R** Dplyr, tidyr, stringr | str_replace_all(department, "HR", "Human Resources") to clean text | replace_na(compliance_data, list(score = "No Data")) |

# Data Structure Matters

| | ✓ Consistent | ✓ Complete | ✓ Validated |
|---|---|---|---|
| **General Examples** | Same formats, naming conventions, time periods. *Example: 01/02/2025 not a mix of "Jan 2025", "Human Resources" vs. "HR"* | No missing values that could mislead. *Example: Show "0" or "No Data" rather than blank cells Excel might skip* | Cross-checked against source systems. *Example: Use dropdown lists for consistent entry (High/Medium/Low risk levels)* |
| **Excel** | **TEXT()** to standardize dates: =TEXT(A1,"MM/DD/YYYY") **SUBSTITUTE()** to standardize names: =SUBSTITUTE(A1,"HR","Human Resources") | IF(ISBLANK(A1),"No Data",A1) to replace blank cells that Excel might skip | Data → Data Validation → List → Source: High,Medium,Low for dropdown controls |
| **Python** Pandas, Numpy, Scipy | pd.to_datetime(df['date'], format='%m/%d/%Y') to standardize dates | df['compliance_rate'].fillna('No Data') instead of leaving NaN values | pd.Categorical(df['risk_level'], categories=['Low', 'Medium', 'High']) |
| **R** Dplyr, tidyr, stringr | str_replace_all(department, "HR", "Human Resources") to clean text | replace_na(compliance_data, list(score = "No Data")) | factor(risk_level, levels = c("Low", "Medium", "High"), ordered = TRUE) |

# Data Structure Matters

| | ✓ Consistent | ✓ Complete | ✓ Validated | ✓ Documented |
|---|---|---|---|---|
| **General Examples** | Same formats, naming conventions, time periods. *Example: 01/02/2025 not a mix of "Jan 2025", "Human Resources" vs. "HR"* | No missing values that could mislead. *Example: Show "0" or "No Data" rather than blank cells Excel might skip* | Cross-checked against source systems. *Example: Use dropdown lists for consistent entry (High/Medium/Low risk levels)* | Clear definitions of what each metric means *Example: "Compliance Rate" = (Passed Audits / Total Audits) × 100* |
| **Excel** | **TEXT()** to standardize dates: =TEXT(A1,"MM/DD/YYYY") **SUBSTITUTE()** to standardize names: =SUBSTITUTE(A1,"HR","Human Resources") | IF(ISBLANK(A1),"No Data",A1) to replace blank cells that Excel might skip | Data → Data Validation → List → Source: High,Medium,Low for dropdown controls | Insert → Comment or create "Notes" tab with formulas like =B2/C2*100 |
| **Python** Pandas, Numpy, Scipy | pd.to_datetime(df['date'], format='%m/%d/%Y') to standardize dates | df['compliance_rate'].fillna('No Data') instead of leaving NaN values | pd.Categorical(df['risk_level'], categories=['Low', 'Medium', 'High']) | df.attrs['compliance_rate'] = 'Passed Audits / Total Audits * 100' |
| **R** Dplyr, tidyr, stringr | str_replace_all(department, "HR", "Human Resources") to clean text | replace_na(compliance_data, list(score = "No Data")) | factor(risk_level, levels = c("Low", "Medium", "High"), ordered = TRUE) | attr(df$compliance_rate, "label") <- "Passed Audits / Total Audits * 100" |

# Power Query: Your Data Cleaning Superhero

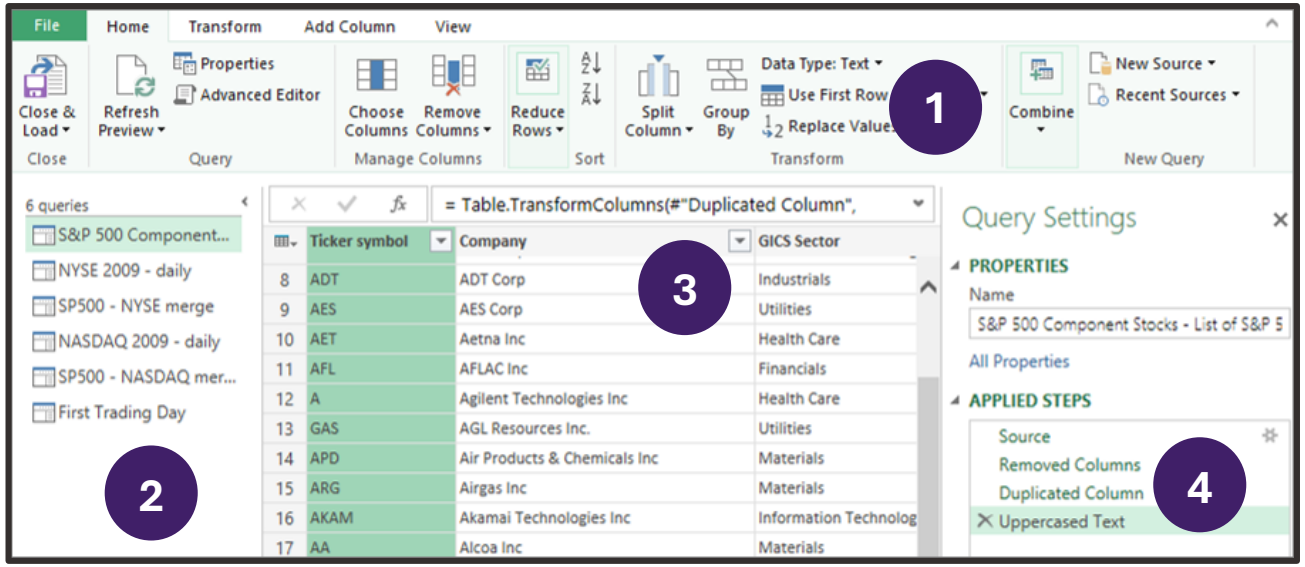
Excel's built-in tool for cleaning and transforming data without changing your source files

## Why use Power Query:

- **Works on copies** - never changes source files
- **Reusable recipes** - save and replay cleaning steps
- **One-click refresh** - auto-updates with new data
- **Team sharing** - export/import processes
- **Multi-source** - combine different folders
- **Step-by-step undo** - see before/after
- **Error handling** - skips bad rows

## How to find Power Query:

In Excel's data ribbon, click on Data → Get Data




1. **Ribbon** - tabs for transform and query options
2. **Queries pane** - view all available queries
3. **Current view** - data preview (main working area)
4. **Query settings** - query name, steps, and indicators

*Image:* The Power Query user interface. Available at: https://learn.microsoft.com/en-us/power-query/power-query-ui

# Understanding Human Perception

How our brains decode visual information

> The eye and the visual cortex of the brain form a remarkably sophisticated image processing system.

Colin Ware

*Information Visualization Expert*

# Types of data

**Qualitative: data whose values describe some characteristic or category**

**Nominal:** examples a

- Country,
- Gender,
- Race,
- Hair color
- Zip Codes

Category

**Ordinal:** examples come with a level of order, such as

- Level of customer experience (good, neutral, bad),
- Socioeconomic status (low, middle, high),
- Education level (high school, BS, MS, PHD),
- Income level (less than 50K, 50
- Motion picture ratings(G, PG, P

Quantity

**Quantitative: data tha**

**Discrete**: countable measurements such as:

- Results from the roll of dice,
- Typing speed per minute,
- Goals in a soccer game,
- Number of books in a library

# Visual Cues

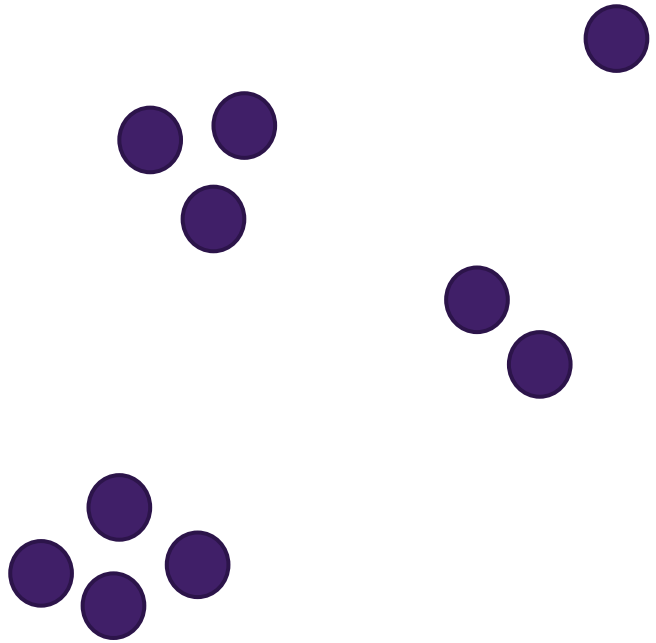
More accurate

Less accurate

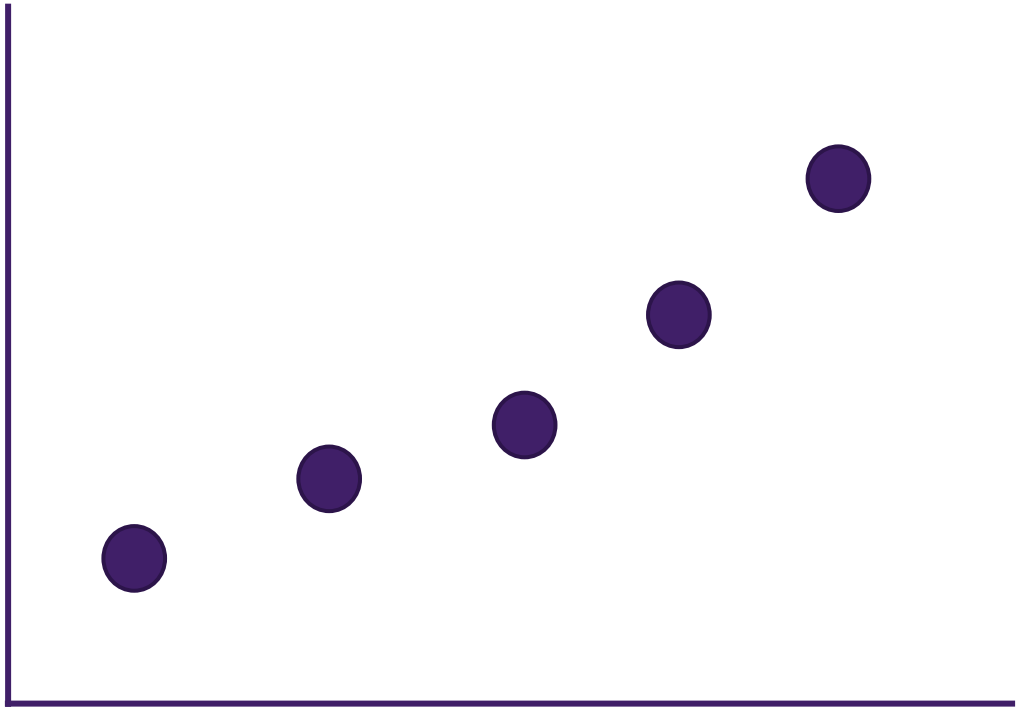
| | Category | Quantity | Text |
|---|---|---|---|
| 1. | Position | Position | Size |
| 2. | Color: Hue | Length | Color: Hue |
| 3. | Shape | Angle | Color: Intensity |
| 4. | Texture | Size | Space |
| 5. | [none] | Color: Intensity | Styling (italics) |

# Visual Cue: Position

Category

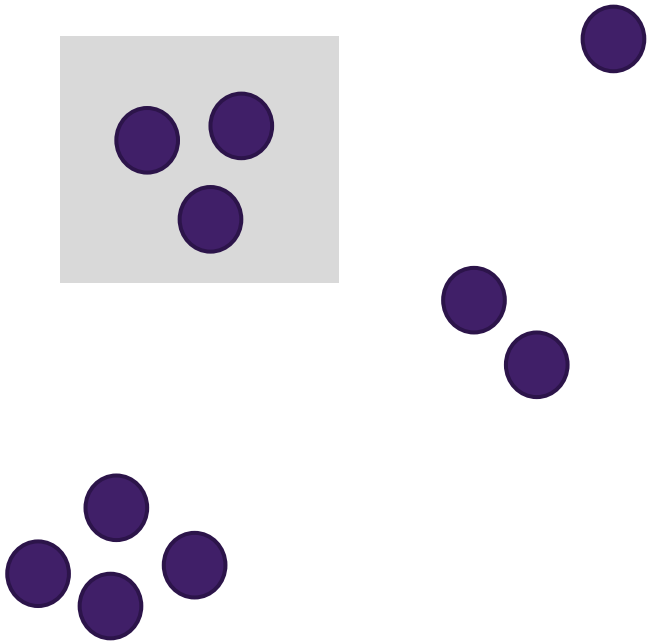


Quantity

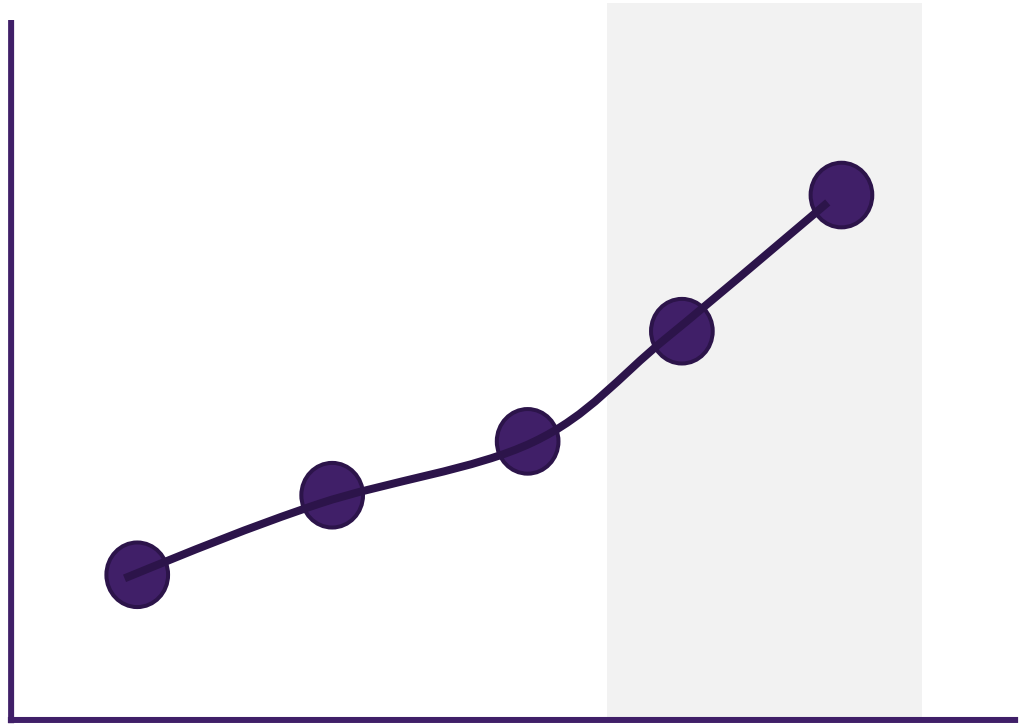


# Visual Cue: Position with enclosures

## Category



## Quantity



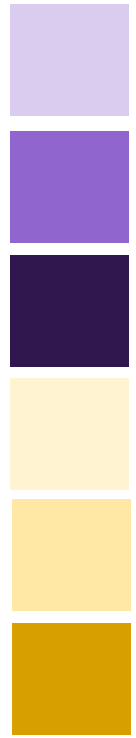
# Visual Cue: Color

## Category

different hues

## Category or Quantity

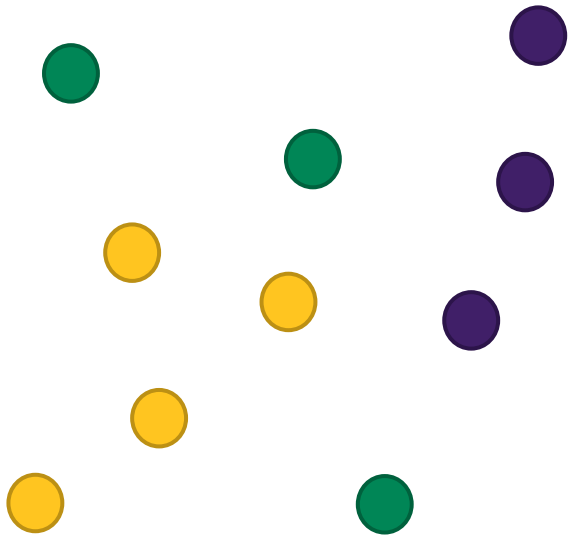
intensity



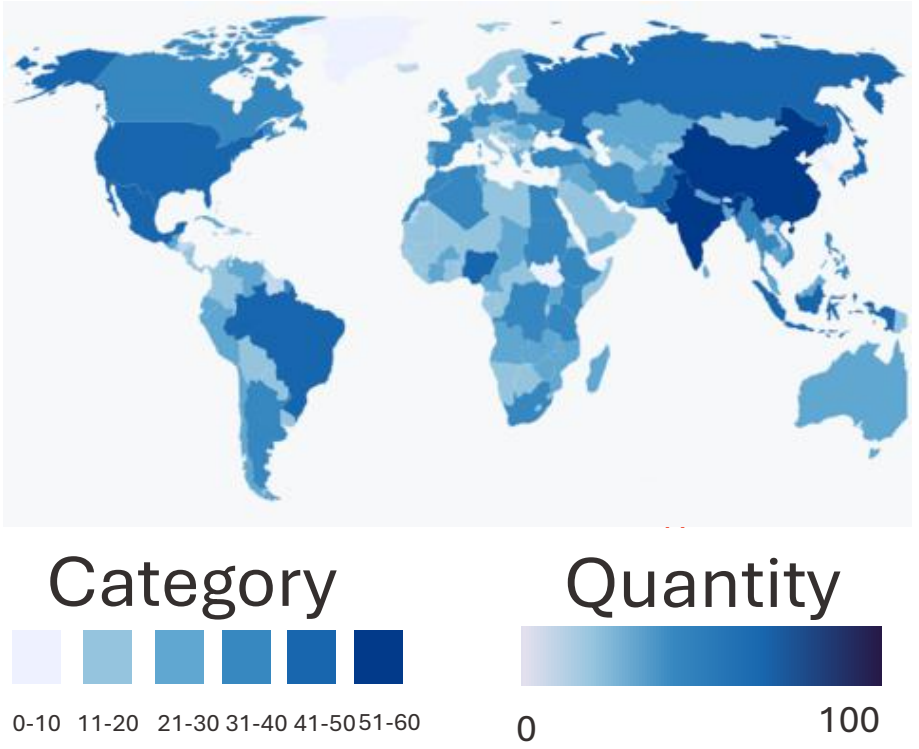
intensity & different hues

# Visual Cue: Color

## Category



## Category or Quantity



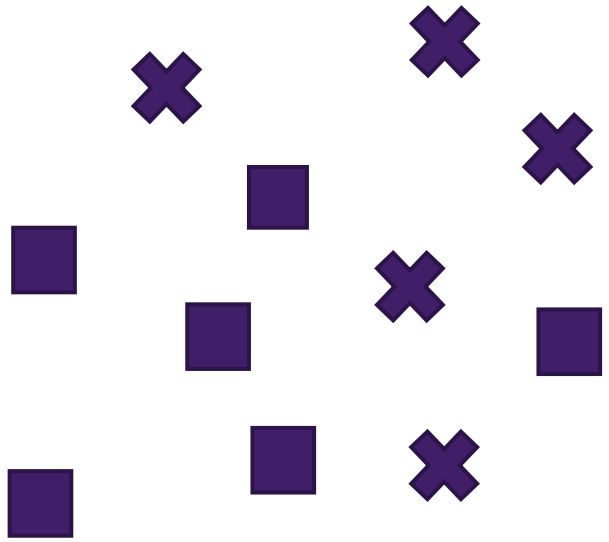
# Visual Cue: Color

## Text

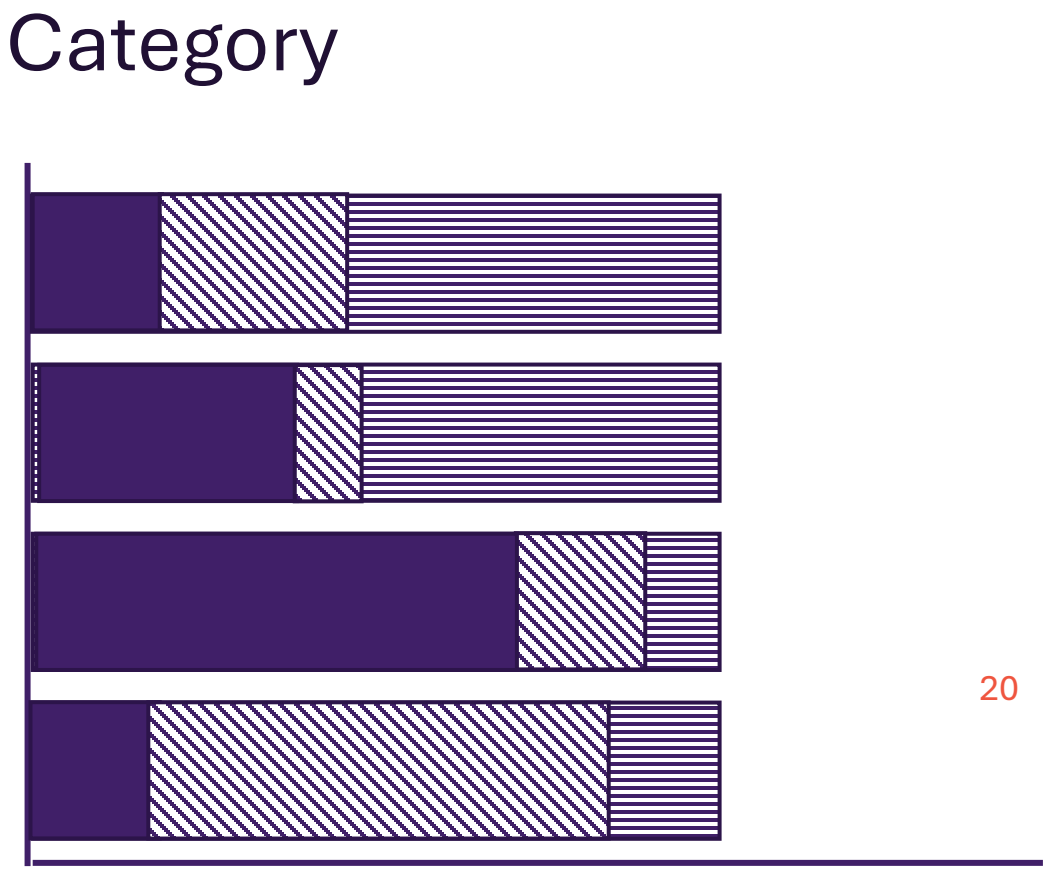
**Lorem Ipsum is simply dummy text** of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. **It was popularized in the 1960s with the release of Letraset sheets** containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum

# Visual Cue: Shape

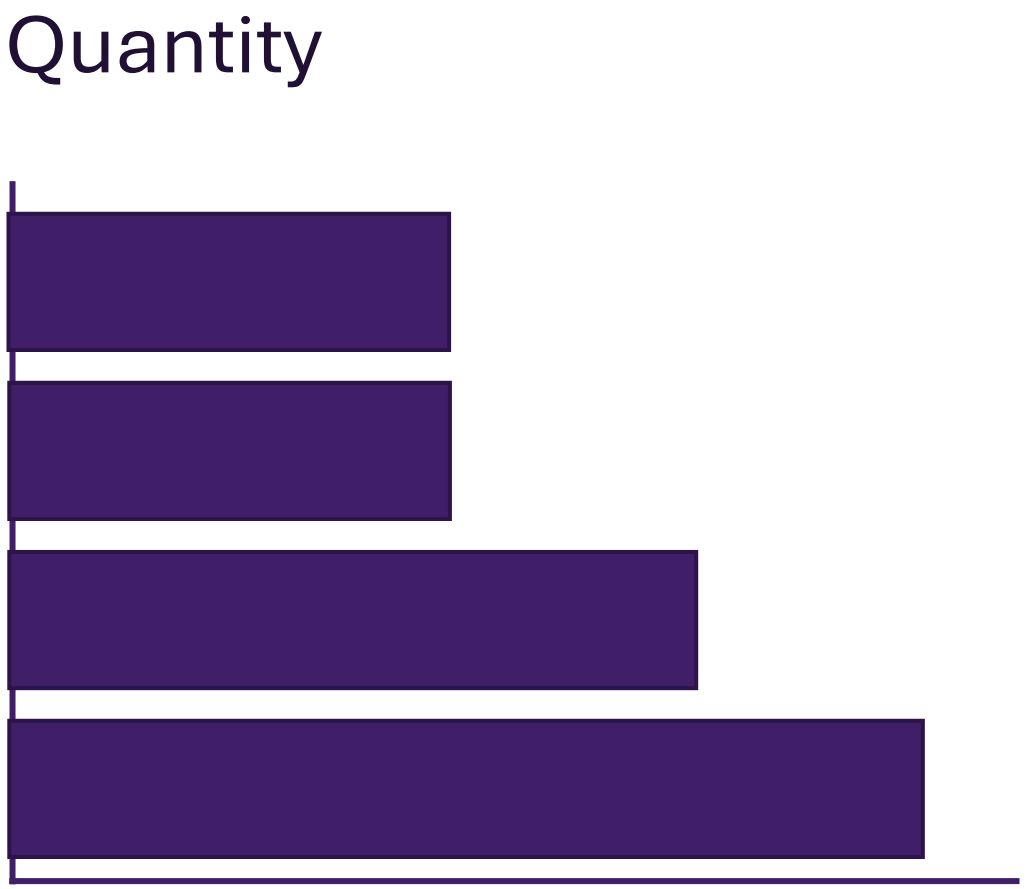
Category



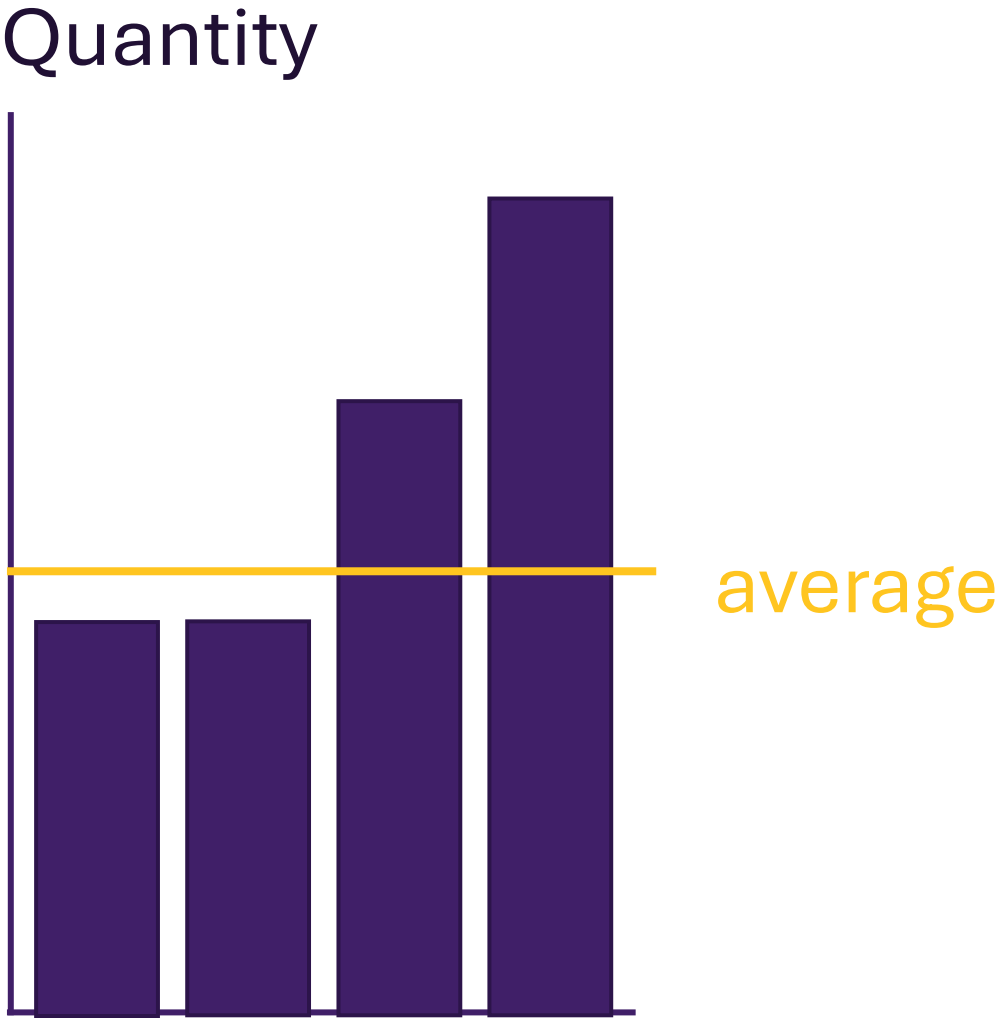
# Visual Cue: Texture



# Visual Cue: Length or Height

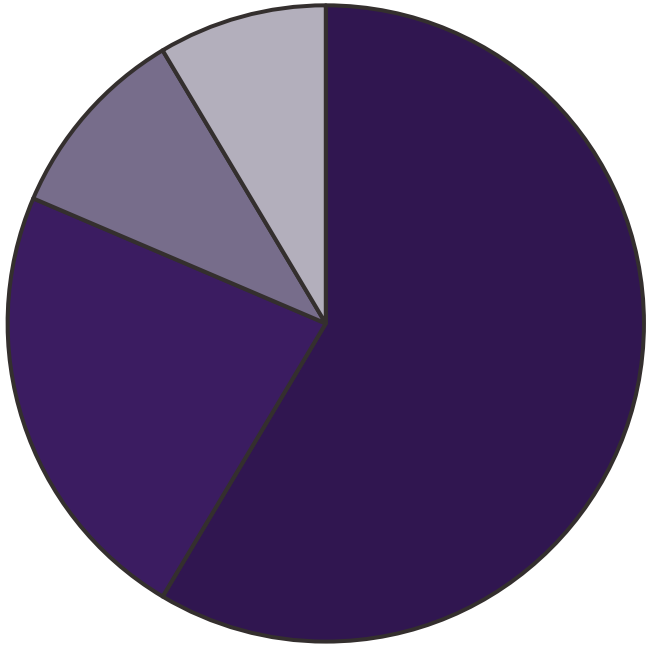


# Visual Cue: Length or Height with baselines

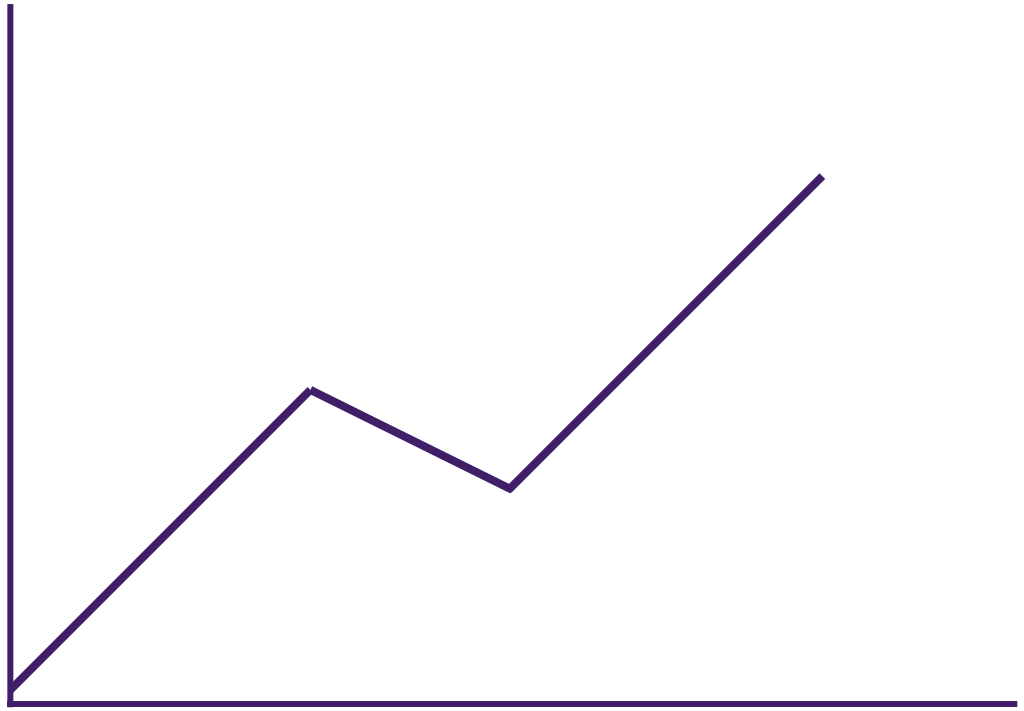


# Visual Cue: Angle

Quantity

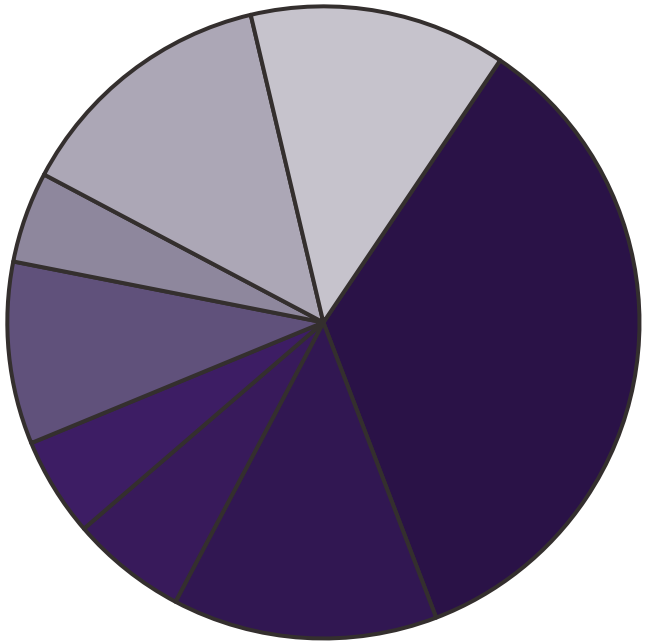


Quantity

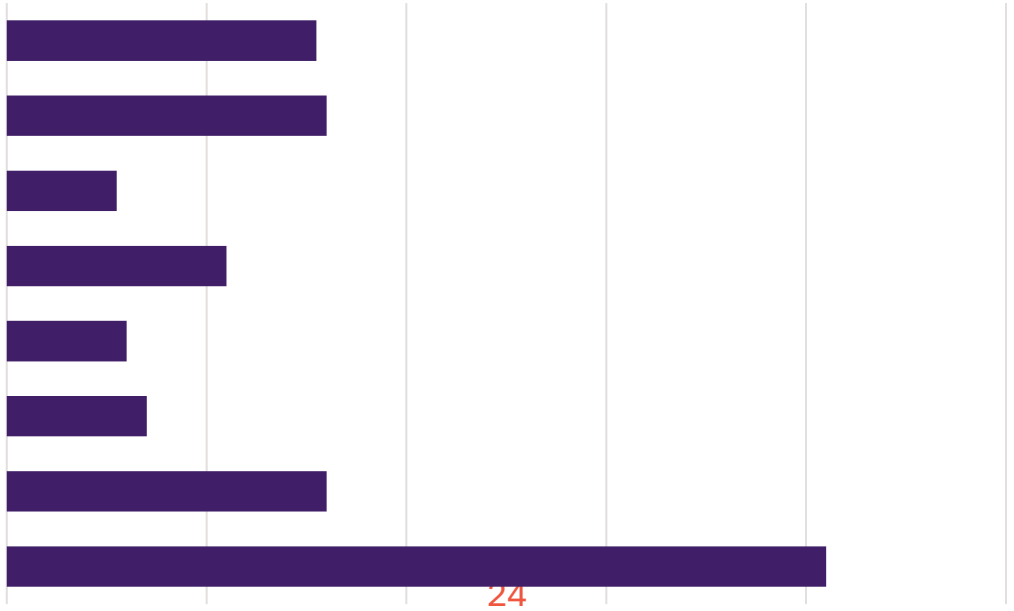


# Visual Cue: Angle

Quantity

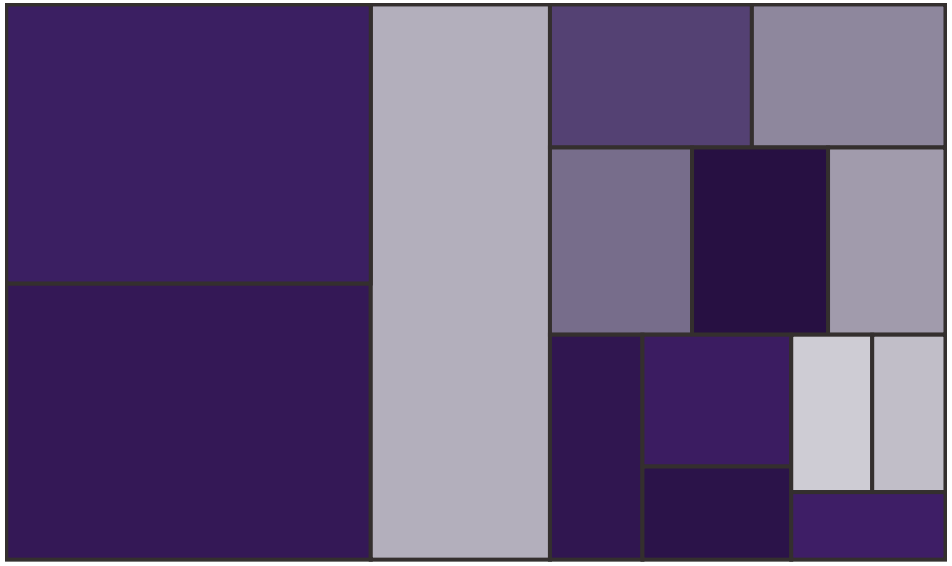


Quantity




# Visual Cue: Size

Quantity



Quantity

# Visual Cue: Size, Spacing, Styling for Text

## Text

*Lorem Ipsum is simply dummy text* of *the printing and typesetting industry*. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book.

> It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged.

It was popularized in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum

# Chart Chooser

**Juice Analytics Chart Chooser: http://labs.juiceanalytics.com/chartchooser/index.html**

**Data Visualization Catalog: https://datavizcatalogue.com/**

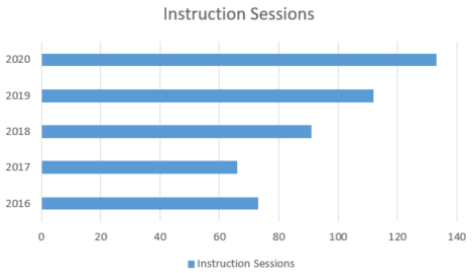
# Sketching Visualizations

> The greatest value of a picture is when it forces us to notice what we never expected to see.
>
> John Tukey
>
> *Statistician*

# Chart Encyclopedia

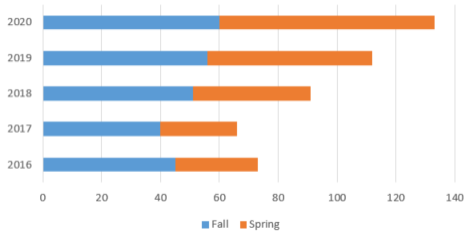
Choosing the right visualization for your story

# Bar Charts



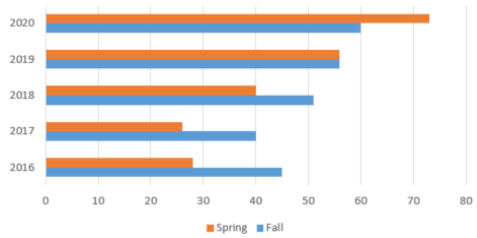

## Bar Chart

Shows change in a single data type over time. Perfect for tracking trends in instruction sessions across multiple years.



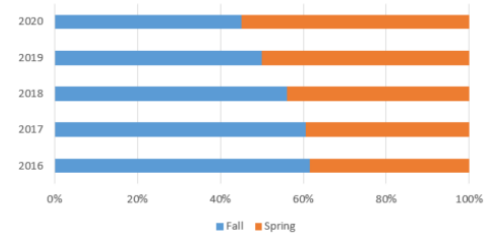

## Stacked Bar Chart

Displays multiple related data types within a single bar. Ideal for showing Fall and Spring sessions together while maintaining visibility of the total.




## Clustered Bar Chart

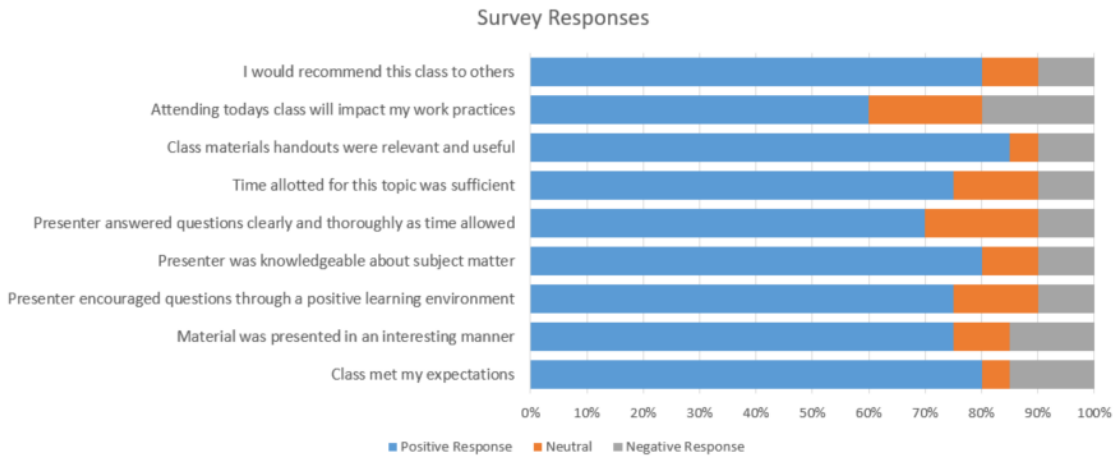
Places related bars side-by-side for direct comparison. Best when comparing Fall vs. Spring sessions across multiple years.




## 100% Stacked Bar Chart

Converts data to percentages (out of 100%), highlighting proportional relationships rather than absolute values.

# Bar Charts continued

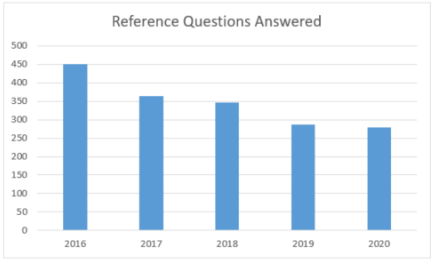



## 100% Stacked Bar Chart

Show change in more than one related data type over time as percent of 100.

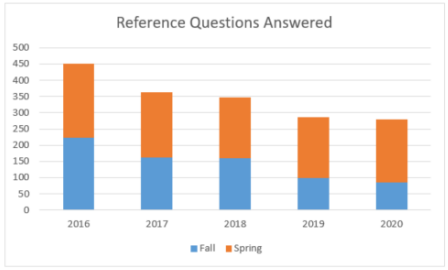
A common use for a 100% Stacked Bar Chart is to show survey responses from a Likert Questions where the categories show positive, neutral, or negative responses.

# Column Charts



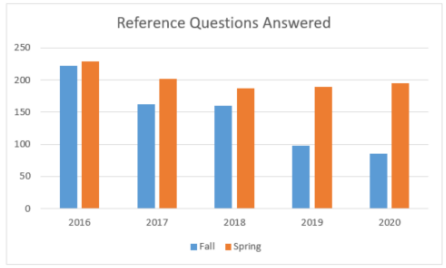

## Column Chart

Used to compare values across one or more categories. Now showing count of reference questions answered each year from 2016-2020



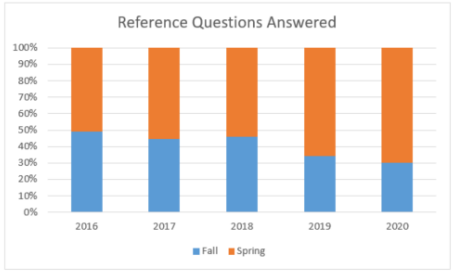

## Stacked Column Chart

Used to compare parts of a whole. Use to show how segments  of a whole change over time. Now showing Count of reference questions answered in Fall and Spring each year from 2016-2020.




## Clustered Column Chart

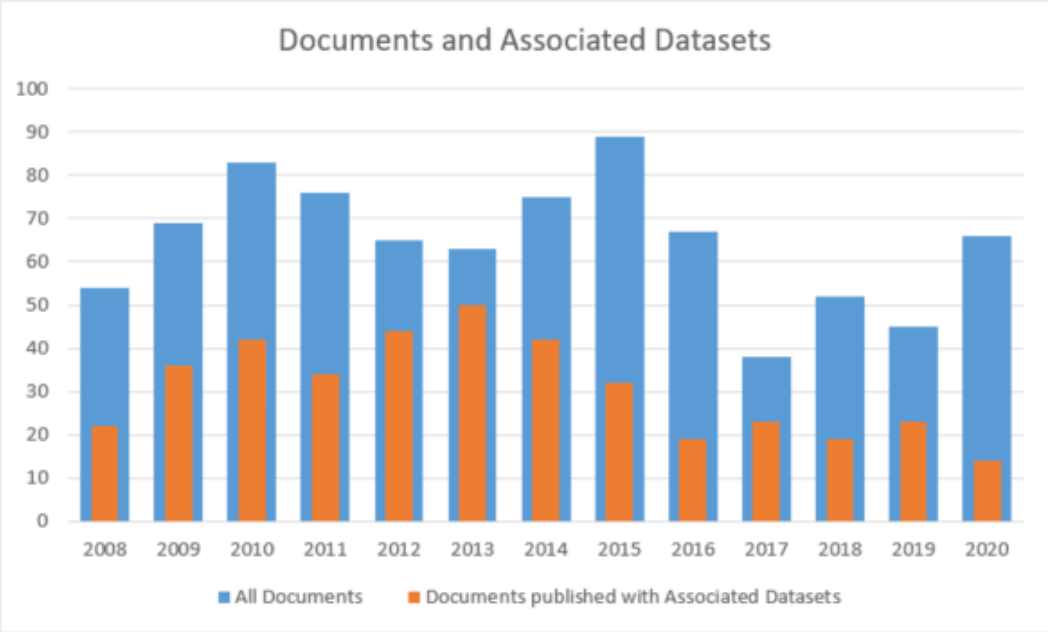
Used to compare values across a few categories. Use when order of categories is not important. Now showing comparison of reference questions answered in Fall and Spring each year from 2016-2020.




## 100% Stacked Column Chart

Used to compare percentage that each value contributes to the total. Use to show how the percentage of each value contributes changes over time. Now showing Percent of reference questions answered in Fall and Spring each year from 2016-2020.

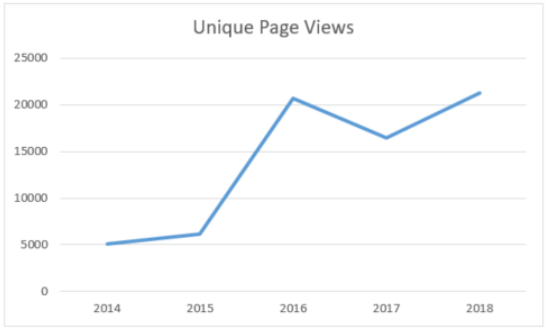
# Column Charts continued

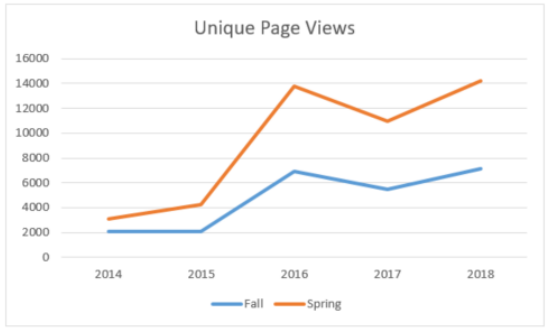


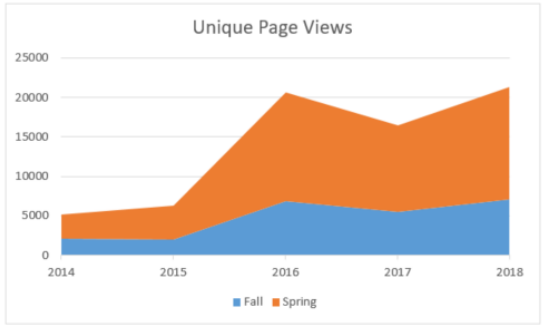

## Overlapping Column Chart

**Used for comparing one set of columns to an underlying or nested set of columns**

# Line Charts




## Line Chart

Use line charts to show trends over time (years, months, days) or when the sequence of categories is important. Ideal for displaying continuous data with many points.

*Example: Unique library website page views each year from 2014 to 2018.*

## Multiple Line Chart

Use when comparing trends across two or more groups over time or across ordered categories. Each line represents a different group, making it easy to see patterns, similarities, and differences in how values change.

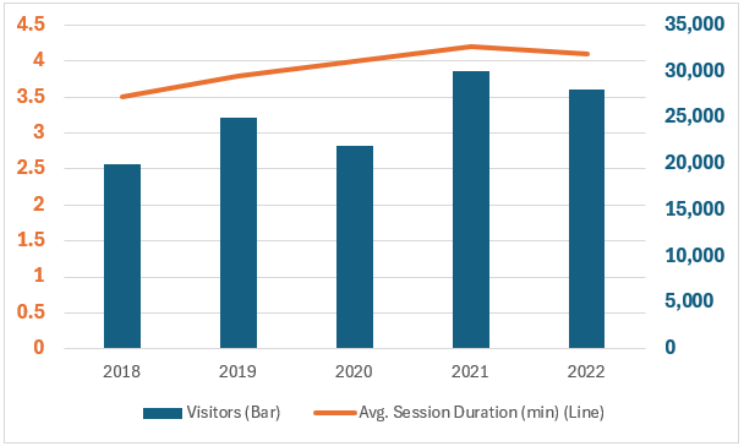
*Example: Comparing unique page views for Fall and Spring on the library website from 2014 to 2018.*

## Stacked Area Chart

Use stacked area charts to show how parts contribute to a whole over time or across ordered categories. Great for highlighting both individual category trends and the overall total.

*Example: Unique library website page views from Fall to Spring, 2014–2018.*

# Combining Chart Types

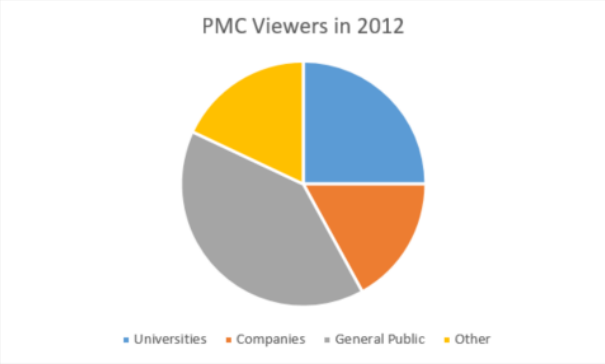



## Combo Chart (Bar + Line)

Used when you want to show two types of information on the same chart—often with different scales or units.
Use the bars to display quantities (e.g., total visitors) and the line to show a trend or rate (e.g., bounce rate or percentage change). Great for showing how one variable relates to another over time or categories.

*Example: Total number of library visitors to the library website by year (bars) compared with average session duration (line) from 2018–2022.*

# Pie and Donut Charts



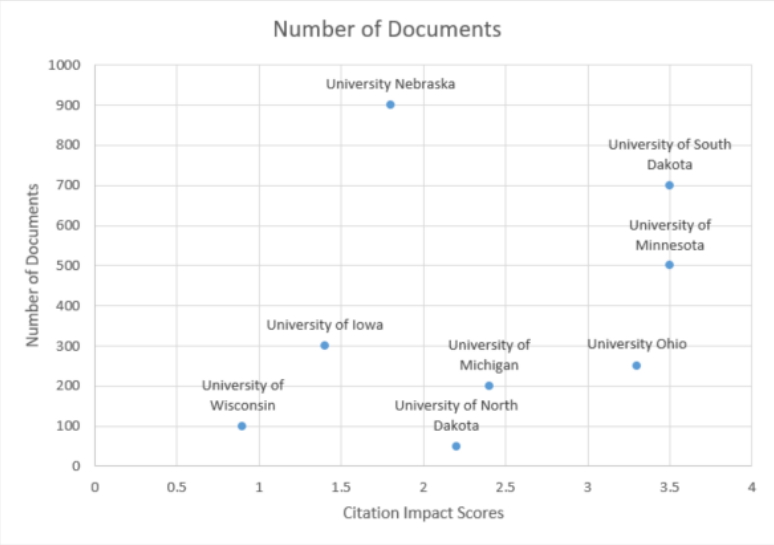

## Pie Chart

Used to show proportions of a whole. Use to show numbers that related to a larger sum and always equal 100%. Do not use this chart if it contains many slices as angles are hard to estimate. Now showing type of viewers of PMC in 2012.


## Donut Chart

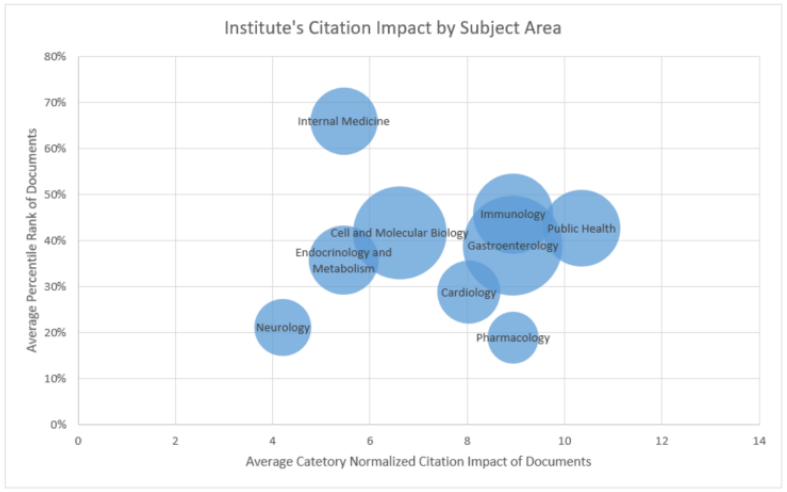
Like a pie chart, a doughnut chart shows the relationship of parts to a whole, but a doughnut chart can contain more than one data series. Each data series that you plot in a doughnut chart adds a ring to the chart. But if you want to compare the data points side by side, you should use a stacked column or stacked bar chart instead. Now showing category of clinical questions answered in 2018.

# Scatter Plot and Bubble Charts




## Scatter Plot

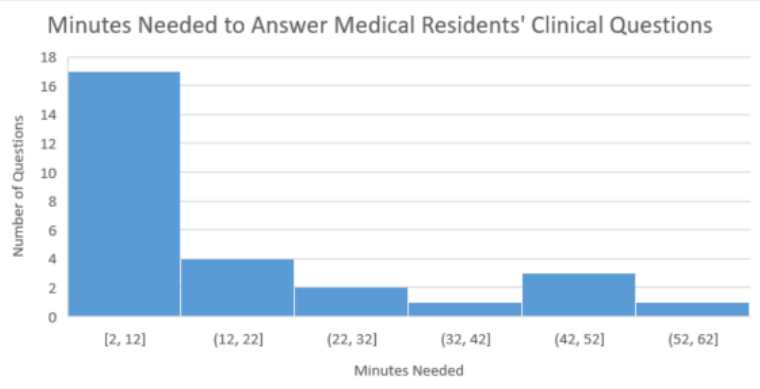
Used to plot numerical data on both the x and y axis. Use when you have two variables that pair well together. Now showing the number of documents published and the average citation impact scores of those documents for several universities.




## Bubble Chart

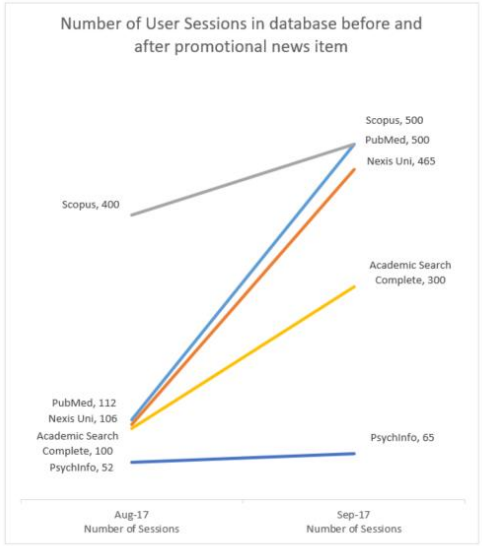
Similar to a scatter plot but allows for a third variable to be charted as the size of the bubble. Now showing the average percentile rank of documents vs. the average normalized category impact of documents, with bubbles sized by the number of documents published.

# Histogram and Slope Chart




## Histograms

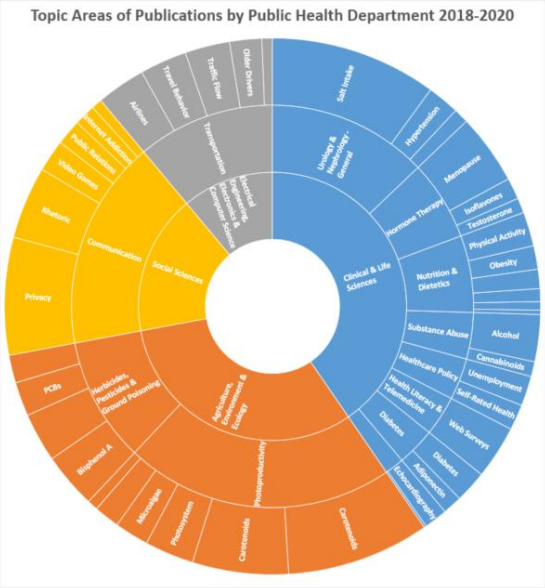
Used to group data into ranges. Tall bars show that more data falls into that range. Now showing minutes needed to answer the questions of medical residents. There were 17 questions that took between 2 and 12 minutes to answer.




## Slope Chart

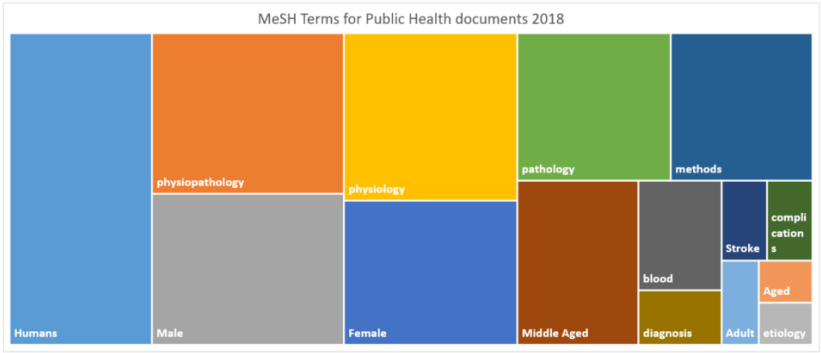
Used to show developments between two points, e.g. between two dates. Now showing the increase in number of user sessions before and after a promotional news item about the database.

# Sunburst and Tree Map




## Sunburst

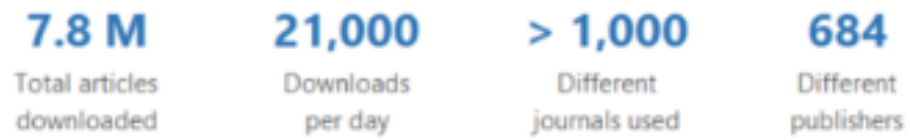
Used to display a hierarchical structure data in circular form. Just like a doughnut chart, Sunburst Chart is also used to display a part of the whole data and compare relative sizes. But it can also show the relationships in the hierarchy. Now showing Topic Areas (macro, meso, micro) of publications by Public Health Department from 2018 -2020.

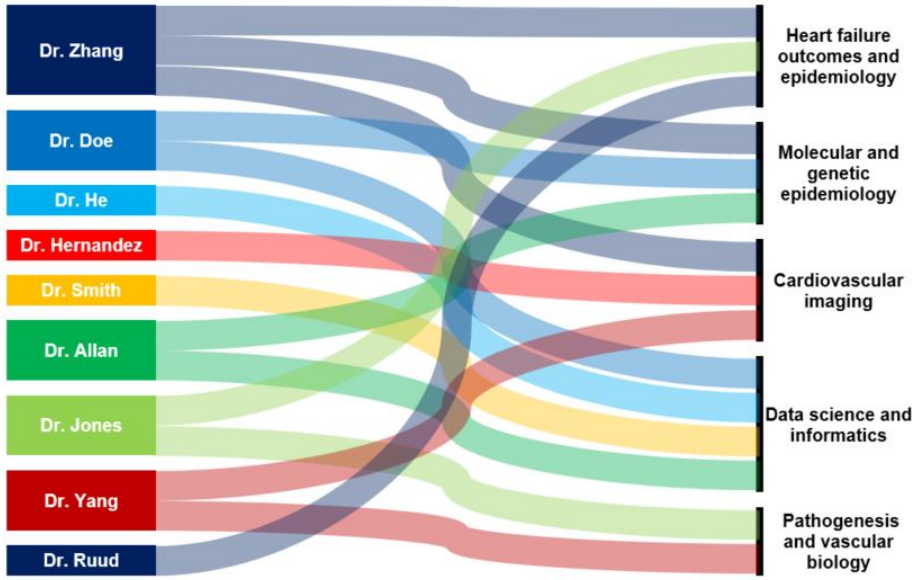



## Tree Map

Used to display a hierarchical view of data and makes it easy to spot patterns. The tree map chart displays categories by color and proximity and can easily show lots of data which would be difficult with other chart types. Now showing MeSH terms for Public Health Documents in 2018

# Big Number and Sankey



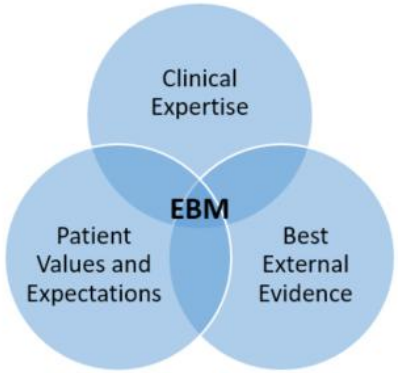

## Big Number

The "Big Number" is about emphasizing the magnitude of a number. You can use a Text Box and Font Size and Color to create a "Big Number".

## Sankey Diagram

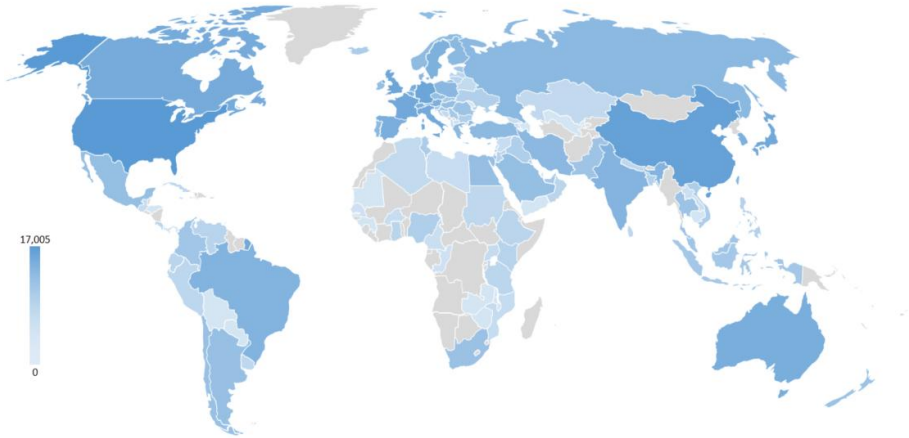
Sankey diagrams emphasize the major transfers or flows within a system. They help locate the most important contributions to a flow. Now showing faculty expertise in various clinical research areas.

# Venn Diagram and Choropleth Map




## Venn Diagram

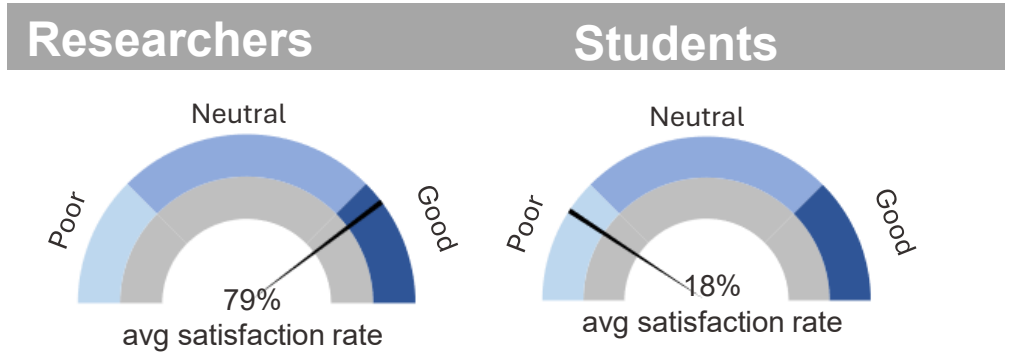
Uses overlapping circles to illustrate the similarities, differences, and relationships between concepts, ideas, categories, or groups.




## Choropleth Map

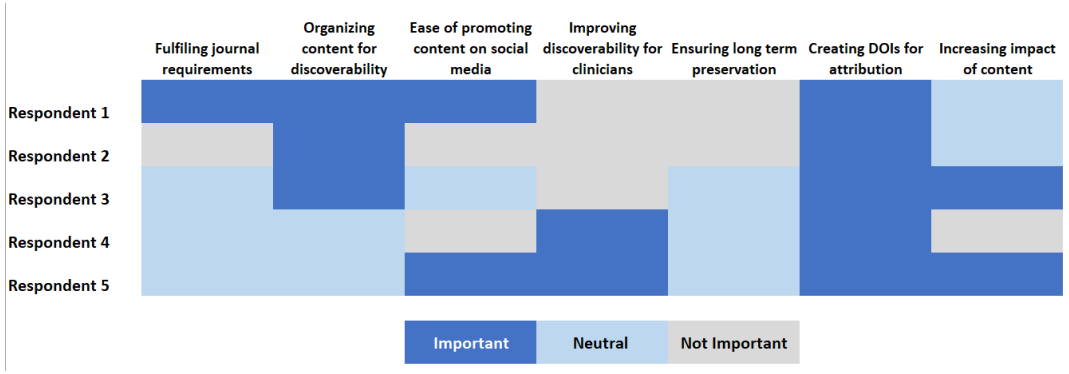
Displays divided geographical areas or regions that are colored, shaded or patterned in relation to a data variable. This provides a way to visualize values over a geographical area, which can show variation or patterns across the displayed location.

# Gauge Chart and Heat Map




## Gauge Charts

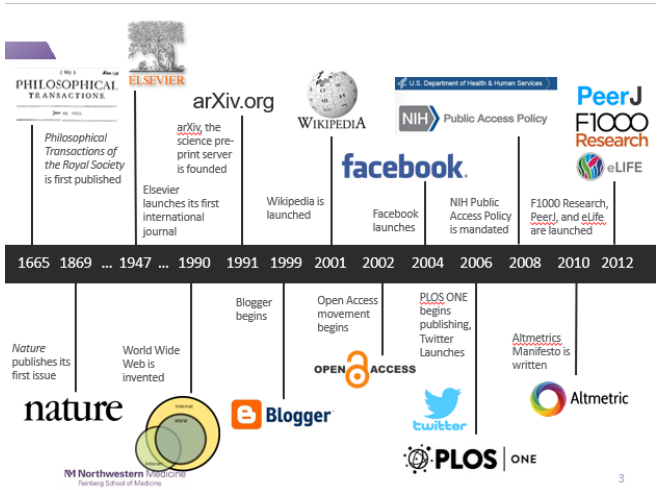
Defies rules for good data visualization but provide a big visual to balance out text for qualitative data. Now showing average satisfaction rate for repository services.

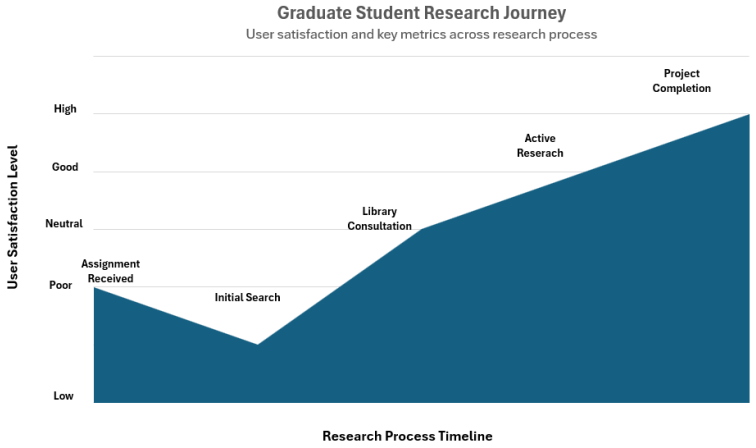



## Heat Map

Graphical representation of data showing relationship between two values using color. Now showing themes from interviews about repository use.

# Timelines and Journey Map



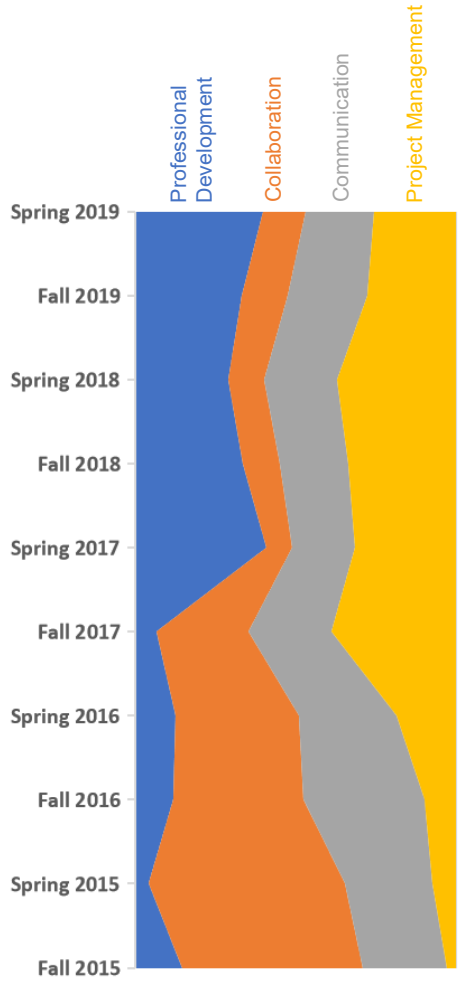

## Timeline

Displays list of events in chronological order, often in a story format. Now showing the history of scholarly communication.

## Journey Map

Shows how a person moves through an experience with your organization, and can highlight areas of strength and weakness. Now showing the average rating of user experience with library classes and services.

# Histomaps and Call Out Boxes




## Histomaps/Stream Maps

Display qualitative data over time. Similar to stacked area graphs. Now showing mentions of major themes from staff survey over 5-year period.

## Call Out Boxes/Quotes

Display quotes that emphasize important data. Now showing quotes from the 2019 Library Staff Survey.

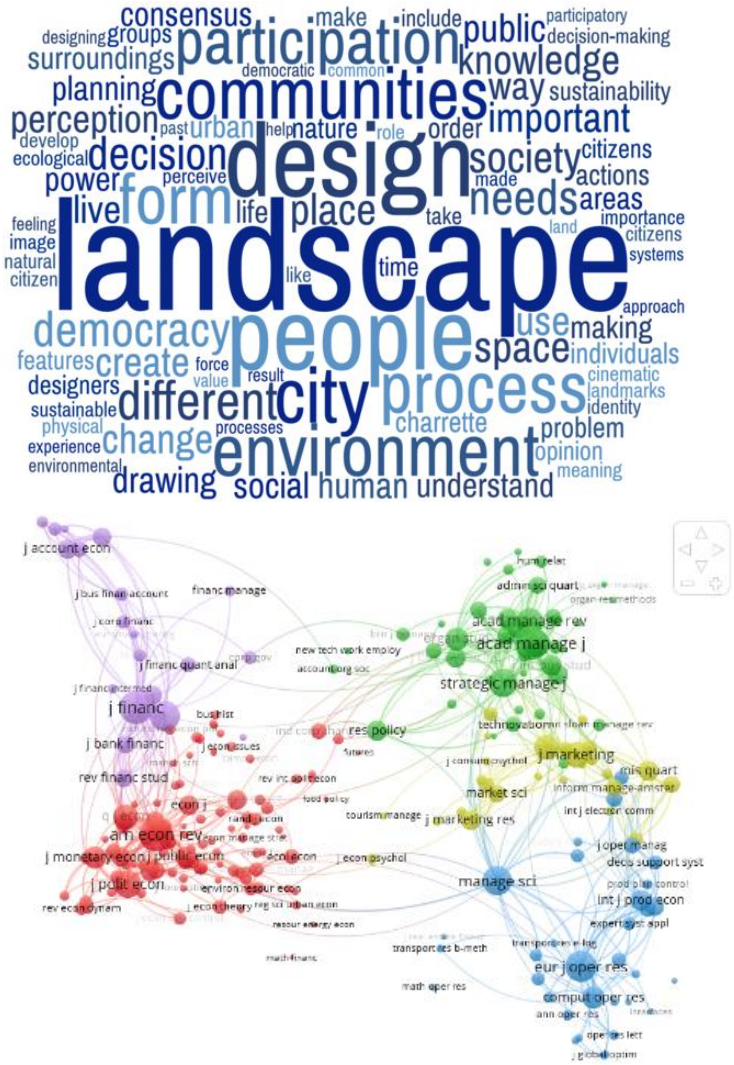
"We need to focus more on improving our project management skills so that we make better use of everyone's time."

"I enjoy all of the opportunities I have to collaborate with others in the library. "

"I want to improve my skills but don't feel like my manager is supportive of me."

*Library Staff Survey 2019*

# Word Cloud and Network Map




## Word Cloud

Displays how frequently words appear in a body of text, by making the size of each word proportional to its frequency. Now showing keywords from documents from a search on democracy.

## Network Map

Displays relationships between entities. Entities are displayed as round nodes and lines (also called edges) show the relationships between them. Now showing relationships between journals publishing economics themed articles.

# Tables

- Title indicating purpose and content
- Emphasis on relationships within the data
- At least 3 columns
- Clear headings
- No blank cells
- Limited decimals (up to 2)
- Decimal alignment
- Careful use of footnotes
- Indication of the number of cases in the table (*n*=)
    - *(Wallace& Van Fleet, 2012)*

# Resources

**Basic and Advanced Excel:**

- Gutzman, K. E., Mendoza, A. R., Norton, C., Belter, C., & Soulakis, M. (2024). Charting a New Course Excel Workbook (v9.0.0). DigitalHub. Galter Health Sciences Library & Learning Center. https://doi.org/10.18131/1nan8-mzk41

**Practice data and examples of charts**

- Gutzman, K. E., Mendoza, A. R., Norton, C., & Belter, C. (2021). Hands On Creating Basic Charts (v3.0.0).DigitalHub. Galter Health Sciences Library & Learning Center. https://doi.org/10.18131/g3-69ds-eq61

> Clutter and confusion are failures of design, not attributes of information.
>
> Edward Tufte
>
> *Statistician and Professor*

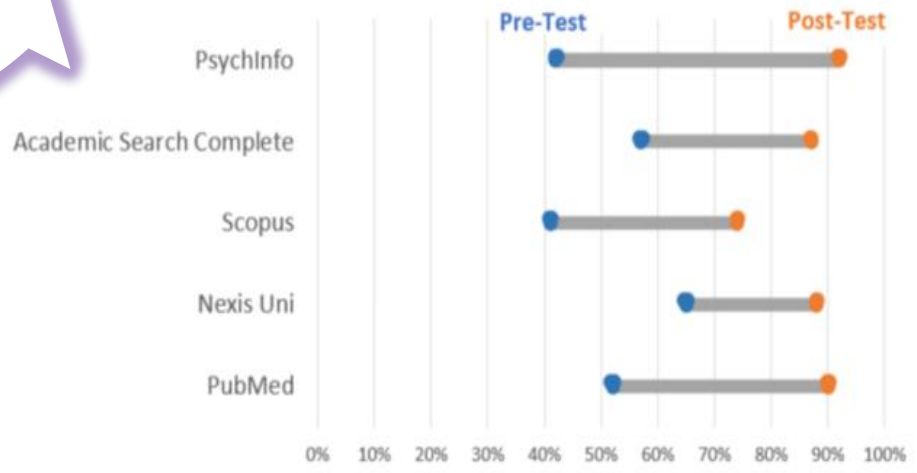
# Designing Information Experiences

Using charts, graphs, and visuals to tell compelling stories with your data

# Lead with Meaning

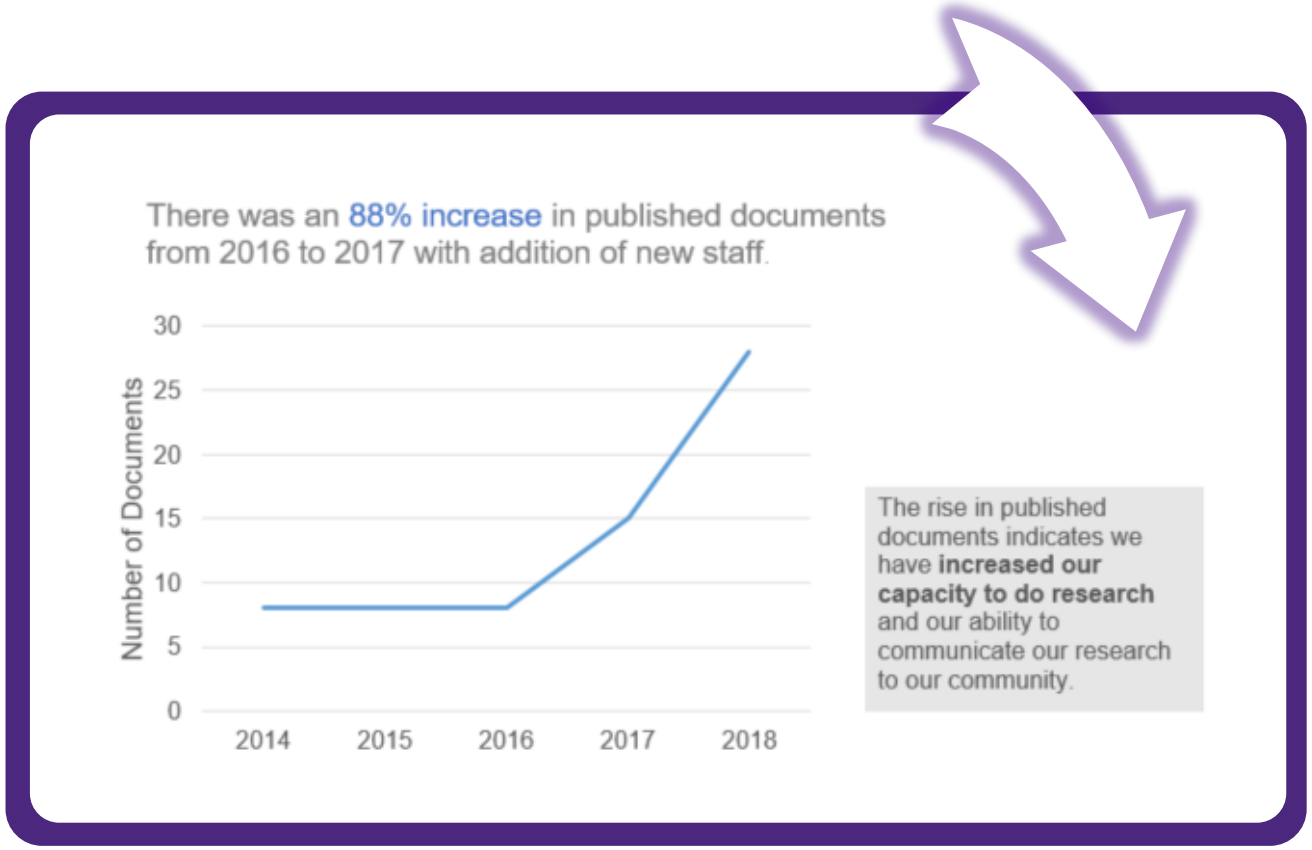
## Our teaching is effective.

Attendees **scores increase by 72%** on average from **pre-test** to **post-test** in our databases information sessions.



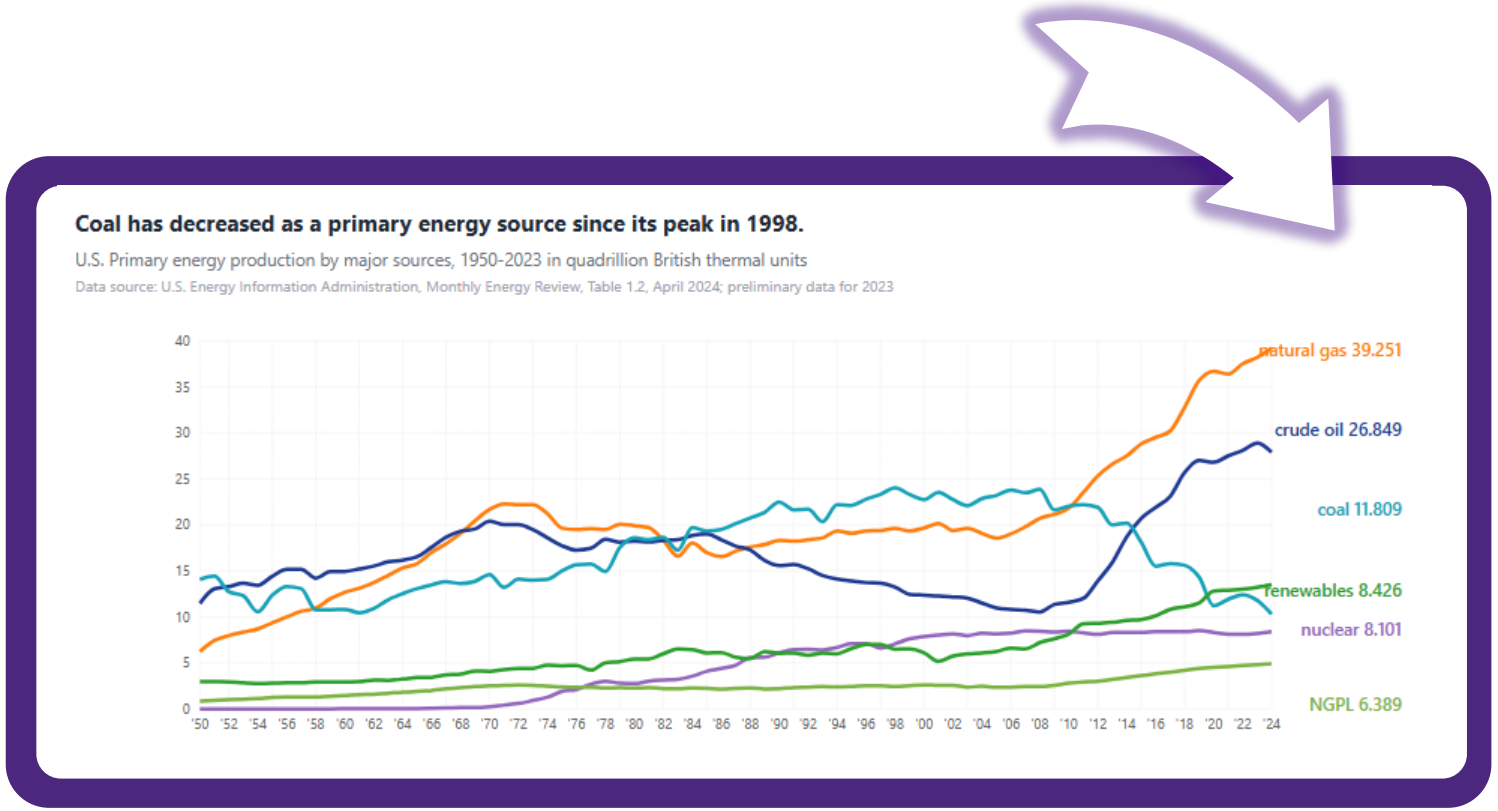

Inspired by Stephanie Evergreen So What? Available at: https://stephanieevergreen.com/so-what/

# Use Call-Out boxes & Side Bars



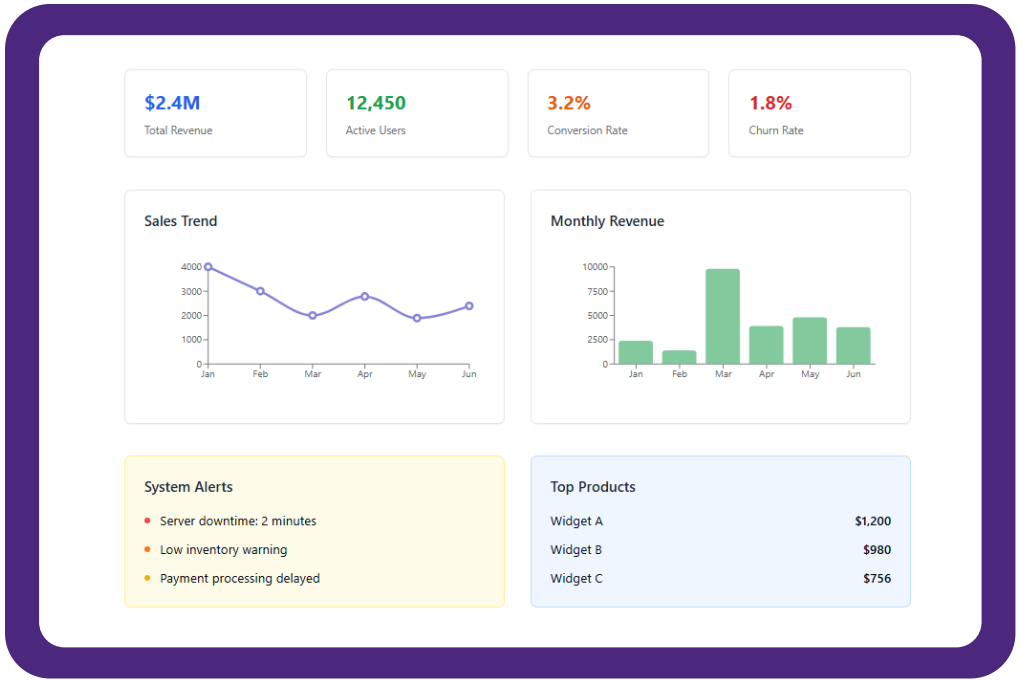

Inspired by Stephanie Evergreen. Sidebars are Your Friend. Available at: https://stephanieevergreen.com/sidebars-are-your-friend/

# Use Informative Labels and Legends

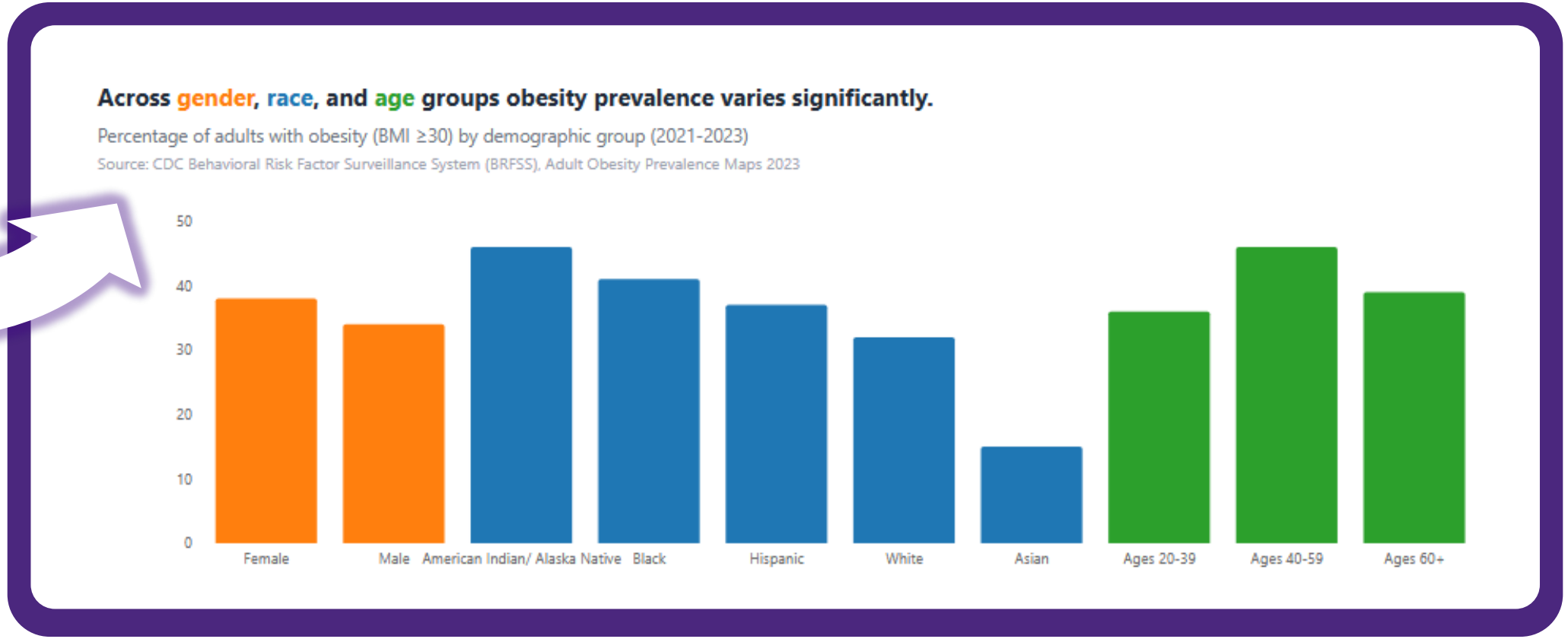



Inspired by Stephanie Evergreen. Use Labels Sparingly. Available at: https://stephanieevergreen.com/labels-are-used-sparingly/

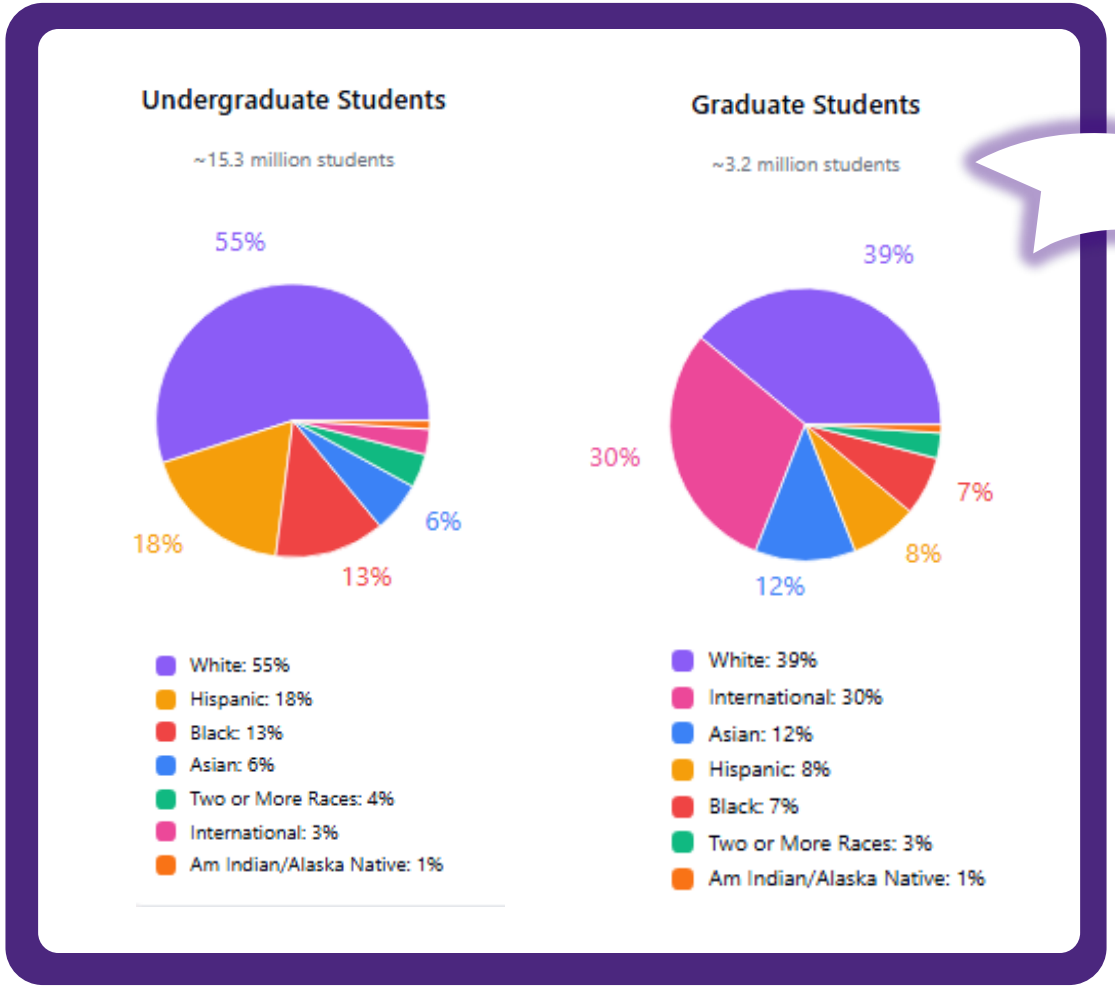
# Empty Space Isn't Wasted Space

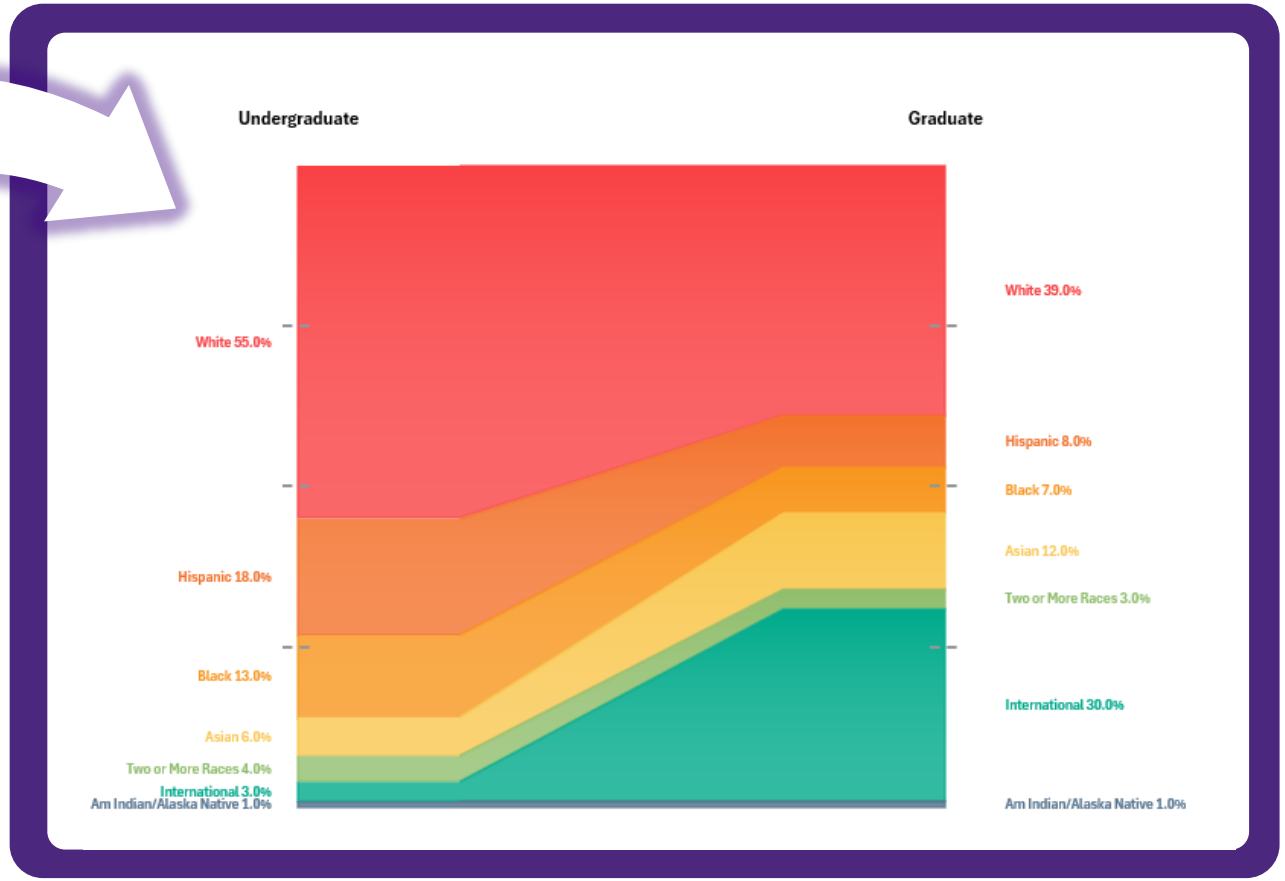



# Colors Should Guide, Not Overwhelm



# Try different kinds of charts




Inspired by Stephanie Evergreen. You Just Need More Chart Choices. Available at: https://stephanieevergreen.com/you-just-need-more-chart-choices/

# Transform Raw Numbers into Actionable Insights

## When Raw Numbers Feel Abstract

- **Convert to percentages or ratios** - *"45% market share" vs. "450,000 customers"*
- **Show change over time instead of snapshots** - *"Sales up 15%" vs. "Sales = $2.3M"*
- **Use familiar units** - *"$12 per day" vs. "$4,380 per year"*

## When Comparisons Are Difficult

- **Create rankings or categories** - "Top performer" vs. "scored 847 points"
- **Show relative to benchmark** - "20% above average" vs. "3.2 rating"
- **Normalize by population** - "5 per 1,000 residents" vs. "500 total cases"

## When Precision Overwhelms

- **Round to meaningful digits -** "About $50K" vs. "$49,847.23"
- **Group into categories -** "High/Medium/Low risk" vs. "Risk scores 1-100"
- **Focus on top performers + "other" -** "Show top 5 + Other" vs. "Show all 47 items"

# Example of Transforming Numbers

| | Relative survival rate, % (SE) | | | |
|---|---|---|---|---|
| | 5 years | 10 years | 15 years | 20 years |
| **Cancer site** | | | | |
| Oral cavity and pharynx | 56·7 (1·3) | 44·2 (1·4) | 37·5 (1·6) | 33·0 (1·8) |
| Oesophagus | 14·2 (1·4) | 7·9 (1·3) | 7·7 (1·6) | 5·4 (2·0) |
| Stomach | 23·8 (1·3) | 19·4 (1·4) | 19·0 (1·7) | 14·9 (1·9) |
| Colon | 61·7 (0·8) | 55·4 (1·0) | 53·9 (1·2) | 52·3 (1·6) |
| Rectum | 62·6 (1·2) | 55·2 (1·4) | 51·8 (1·8) | 49·2 (2·3) |
| Liver and intrahepatic bile duct | 7·5 (1·1) | 5·8 (1·2) | 6·3 (1·5) | 7·6 (2·0) |
| Pancreas | 4·0 (0·5) | 3·0 (0·5) | 2·7 (0·6) | 2·7 (0·8) |
| Larynx | 68·8 (2·1) | 56·7 (2·5) | 45·8 (2·8) | 37·8 (3·1) |
| Lung and bronchus | 15·0 (0·4) | 10·6 (0·4) | 8·1 (0·4) | 6·5 (0·4) |
| Melanomas | 89·0 (0·8) | 86·7 (1·1) | 83·5 (1·5) | 82·8 (1·9) |
| Breast | 86·4 (0·4) | 78·3 (0·6) | 71·3 (0·7) | 65·0 (1·0) |
| Cervix uteri | 70·5 (1·6) | 64·1 (1·8) | 62·8 (2·1) | 60·0 (2·4) |
| Corpus uteri and uterus, NOS | 84·3 (1·0) | 83·2 (1·3) | 80·8 (1·7) | 79·2 (2·0) |
| Ovary | 55·0 (1·3) | 49·3 (1·6) | 49·9 (1·9) | 49·6 (2·4) |
| Prostate | 98·8 (0·4) | 95·2 (0·9) | 87·1 (1·7) | 81·1 (3·0) |
| Testis | 94·7 (1·1) | 94·0 (1·3) | 91·1 (1·8) | 88·2 (2·3) |
| Urinary bladder | 82·1 (1·0) | 76·2 (1·4) | 70·3 (1·9) | 67·9 (2·4) |
| Kidney and renal pelvis | 61·8 (1·3) | 54·4 (1·6) | 49·8 (2·0) | 47·3 (2·6) |
| Brain and other nervous system | 32·0 (1·4) | 29·2 (1·5) | 27·6 (1·6) | 26·1 (1·9) |
| Thyroid | 96·0 (0·8) | 95·8 (1·2) | 94·0 (1·6) | 95·4 (2·1) |
| Hodgkin's disease | 85·1 (1·7) | 79·8 (2·0) | 73·8 (2·4) | 67·1 (2·8) |
| Non-Hodgkin lymphomas | 57·8 (1·0) | 46·3 (1·2) | 38·3 (1·4) | 34·3 (1·7) |
| Multiple myeloma | 29·5 (1·6) | 12·7 (1·5) | 7·0 (1·3) | 4·8 (1·5) |
| Leukaemias | 42·5 (1·2) | 32·4 (1·3) | 29·7 (1·5) | 26·2 (1·7) |

Rates derived from SEER 1973–98 database (both sexes, all ethnic groups).[12]
NOS=not otherwise specified.

Table 4: **Most recent period estimates of relative survival rates, by cancer site**

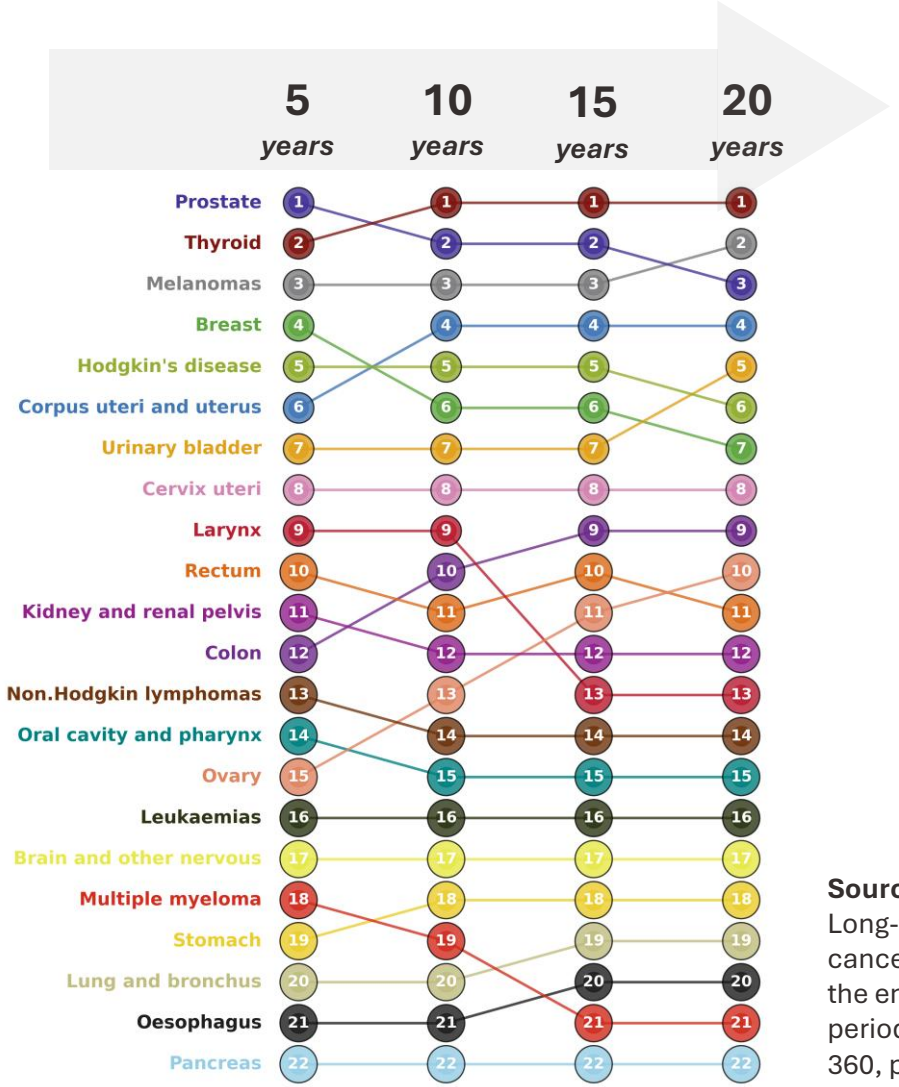
**Source:** Brenner, H. (2002). Long-term survival rates of cancer patients achieved by the end of the 20th century: a period analysis. The Lancet, 360, pp. 1131-1135.

# Example of Transforming Numbers

**From Diagnosis to 20 Years Later:**
*How Cancer Survival Rankings Evolve*

**Insights:**

- **Success stories**. Prostate and thyroid cancers stand out with exceptional survival rates - prostate at 98.8% at 5 years and thyroid at 96%.
- **The Pancreatic Challenge**. Pancreas cancer shows devastatingly low survival rates (4% at 5 years, dropping to 2.7% at 20 years), highlighting one of oncology's most urgent challenges.
- **Brain Cancers**. Show consistently low survival rates (32% at 5 years), highlighting the challenges of treating central nervous system malignancies.




**Source:** Brenner, H. (2002). Long-term survival rates of cancer patients achieved by the end of the 20th century: a period analysis. The Lancet, 360, pp. 1131-1135.

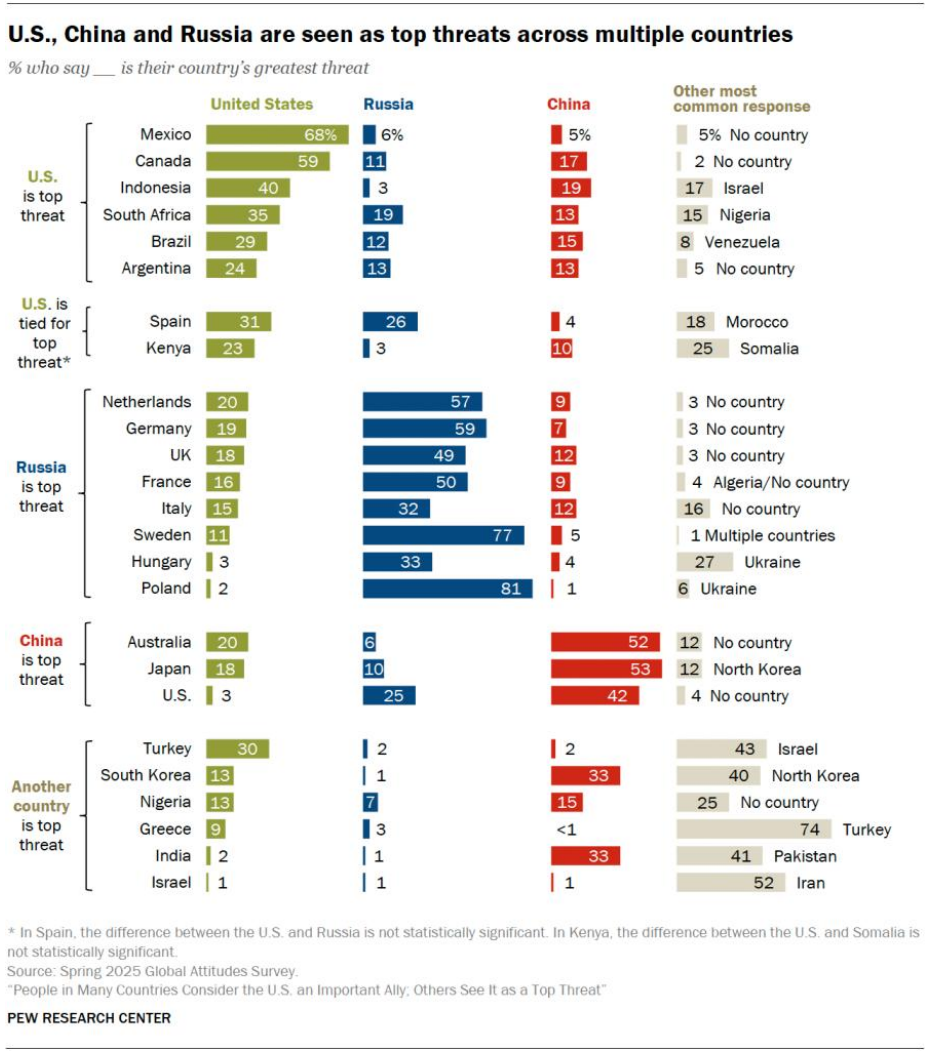
# Present One Big Idea at a Time

**Focus on Insights:**

- Every visualization should answer one primary question or make one key point.
- Your goal isn't to show all your data, it's to communicate insights that drive action.

**Strategies include:**

- **Use multiple simple charts** instead of one complex chart.
- **Gray out supporting data.** Keep context but don't let it compete for attention.




See: https://www.pewresearch.org/global/2025/07/08/people-in-many-countries-consider-the-u-s-an-important-ally-others-see-it-as-a-top-threat/pg_2025-07-08_allies-threats_0_03/

# US, China, Russia seen as top threats across multiple countries

*% who say _____ is their country's greatest threat.*

**Not shown:** Regional rivalries and internal tensions often eclipse global powers in perceived threat assessments. For example, Turkey (74%) is seen as the top threat by its own citizens, reflecting internal divisions. South Korea names Pakistan. Iran, North Korea, and Israel also appear in regional contexts.

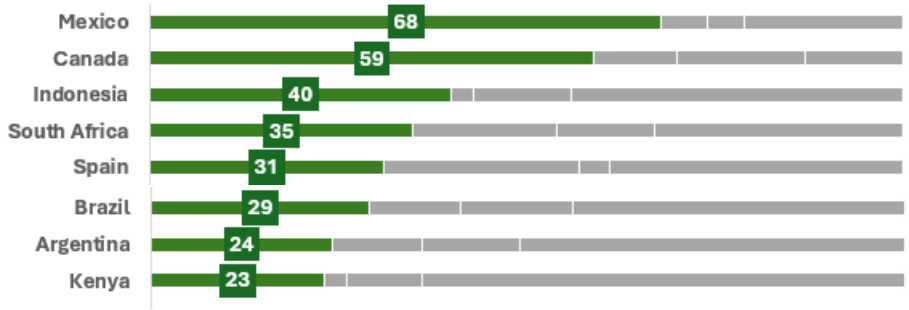
*2025 Pew Research Center*

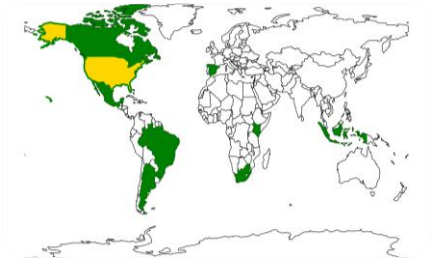
## Threats

## Patterns

## Insights

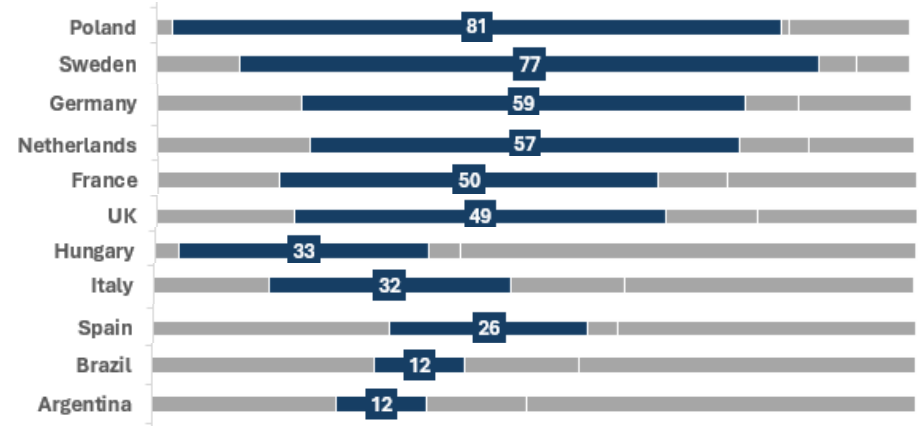
**United States** is perceived as the greatest threat for **8** countries



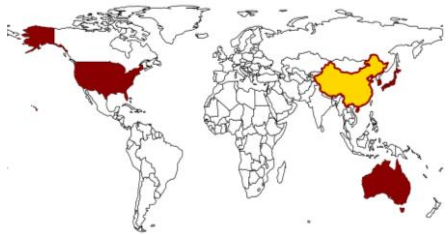
The **U.S.** is seen as a threat that spans diverse regions and are not limited to adversaries but includes allies and neutral countries as well.

**Russia** is perceived as the greatest threat for **11** countries




**Russia** is seen as a threat with regional concentration, suggesting influence related to proximity, conflict history, and energy ties.

**China** is perceived as the greatest threat for **3** countries

**China** is seen as a strategic rival in both the Indo-Pacific and the U.S., aligning with tensions over trade, military buildup, and regional influence.

# Data Visualization Checklist

## Data Visualization Checklist

by Stephanie Evergreen & Ann K. Emery

This checklist is meant to be used as a guide for the development of high impact data visualizations. Rate each aspect of the data visualization by circling the most appropriate number, where 2 points means the guideline was fully met, 1 means it was partially met, and 0 means it was not met at all. n/a should not be used frequently, but reserved for when the guideline truly does not apply. For example, a pie chart has no axes lines or tick marks to rate. If the guideline has been broken intentionally to make a point, rate it n/a and deduct those points from the total possible. Refer to the Data Visualization Anatomy Chart on the last page for guidance on vocabulary and the Resources at the end for more details.

| | Guideline | Rating |
|---|---|---|
| **Text**<br><br>Graphs don't contain much text, so existing text must encapsulate your message and pack a punch. | **6-12 word descriptive title is left-justified in upper left corner**<br>Short titles enable readers to comprehend takeaway messages even while quickly skimming the graph. Rather than a generic phrase, use a descriptive sentence that encapsulates the graph's finding or "so what?" Western cultures start reading in the upper left, so locate the title there. | 2 1 0 n/a |
| | **Subtitle and/or annotations provide additional information**<br>Subtitles and annotations (call-out text within the graph) can add explanatory and interpretive power to a graph. Use them to answer questions a viewer might have or to highlight specific data points. | 2 1 0 n/a |
| | **Text size is hierarchical and readable**<br>Titles are in a larger size than subtitles or annotations, which are larger than labels, which are larger than axis labels, which are larger than source information. The smallest text - axis labels - are at least 9 point font size on paper, at least 20 on screen. | 2 1 0 n/a |
| | **Text is horizontal**<br>Titles, subtitles, annotations, and data labels are horizontal (not vertical or diagonal). Line labels and axis labels can deviate from this rule and still receive full points. Consider switching graph orientation (e.g., from column to bar chart) to make text horizontal. | 2 1 0 n/a |
| | **Data are labeled directly**<br>Position data labels near the data rather than in a separate legend (e.g., on top of or next to bars and next to lines). Eliminate/embed legends when possible because eye movement back and forth between the legend and the data can interrupt the brain's attempts to interpret the graph. | 2 1 0 n/a |
| | **Labels are used sparingly**<br>Focus attention by removing the redundancy. For example, in line charts, label every other year on an axis. Do not add numeric labels "and" use a y-axis scale, since this is redundant. | 2 1 0 n/a |

EVERGREENDATA
INTENTIONAL REPORTING & DATA VISUALIZATION

**Source:** Evergreen, Stephanie, Data Visualization Checklist. Available at: https://stephanieevergreen.com/wp-content/uploads/2020/12/EvergreenDataVizChecklist.pdf

# Ethics in Data Visualization

Responsibility, Truth, and Trust in Visual Communication

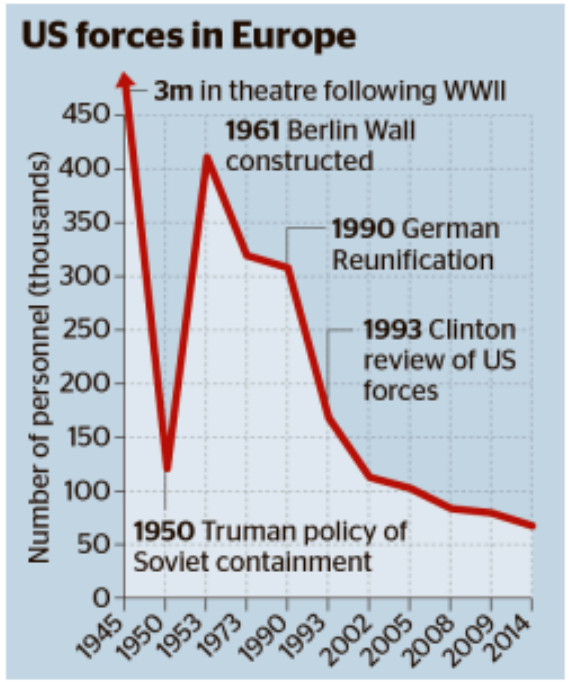
> The responsibility of an ethical data visualizer is to create the most good while doing the least harm.
>
> Alberto Cairo
>
> *Visual Journalist*

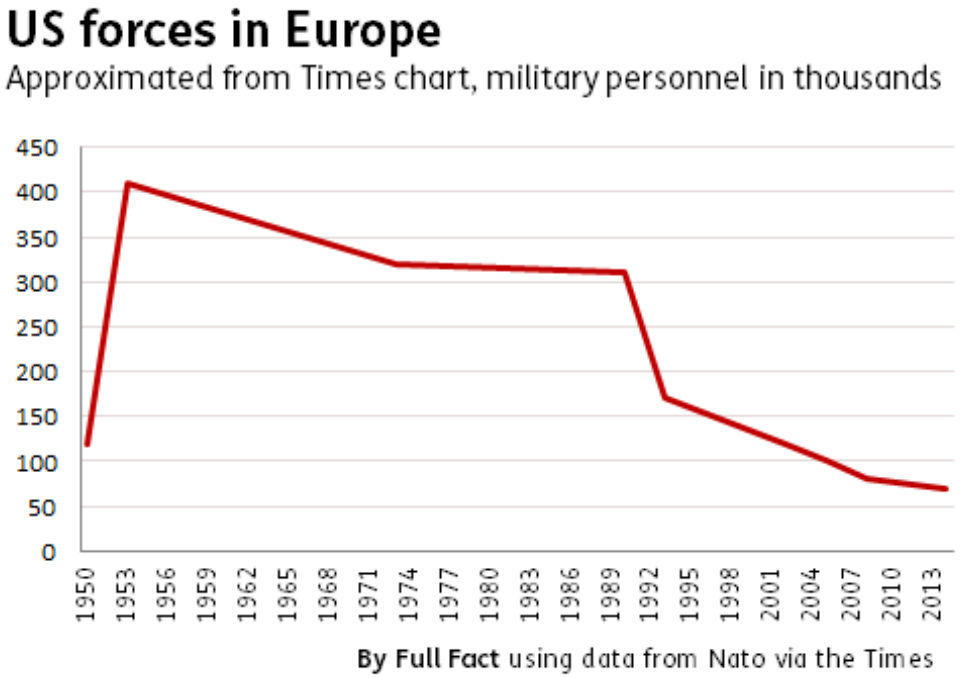
# The Ethics of Data Visualization

"I shall not use visualization to intentionally hide or confuse the truth which it is intended to portray. I will **respect the great power visualization has** in garnering wisdom and misleading the uninformed. I accept this responsibility willfully and without reservation, and promise to defend this oath against all enemies, both domestic and foreign."

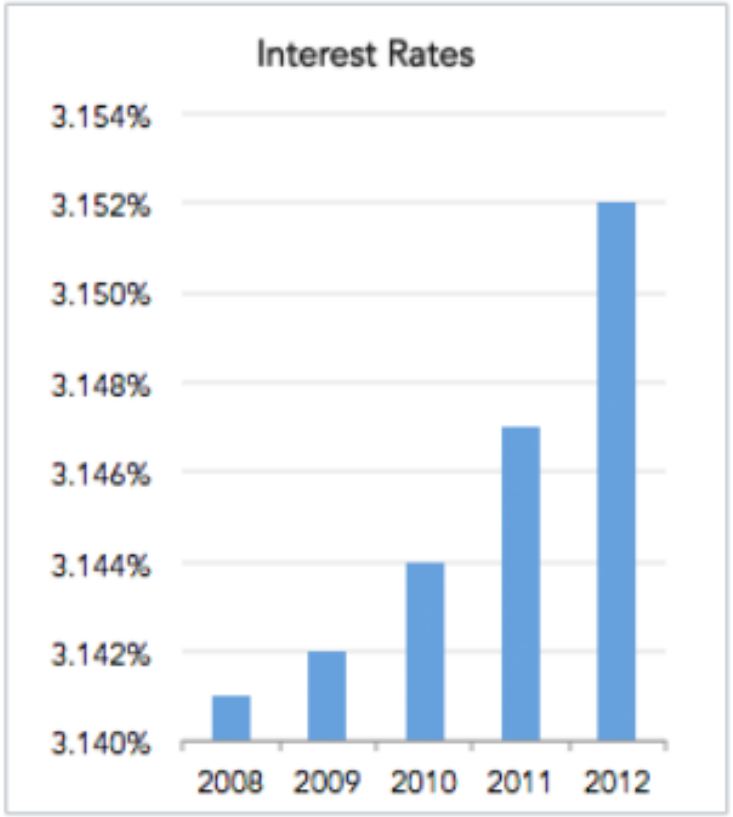
- Jason Moore of the Air Force Research Lab as quoted during a presentation at the 2011 VisWeek conference

# Manipulating an axis



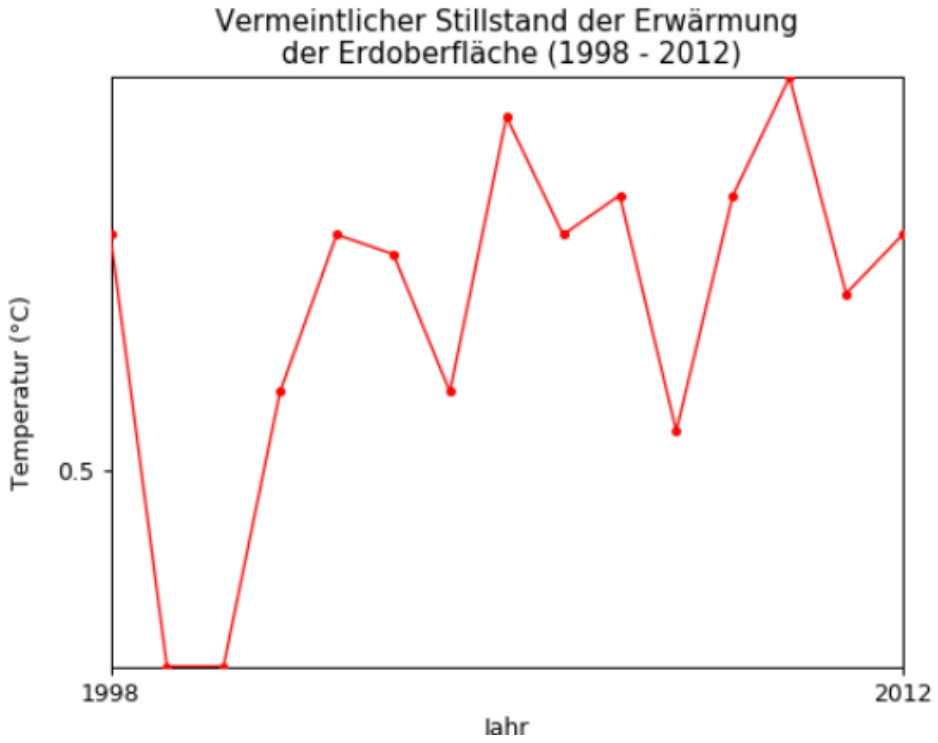



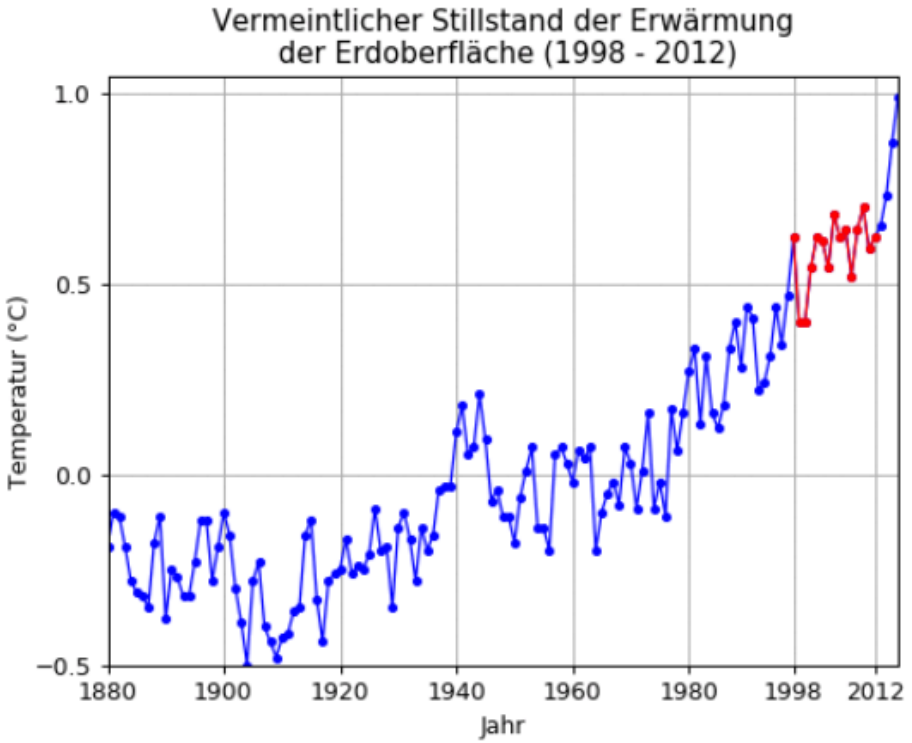
# Omitting the baseline



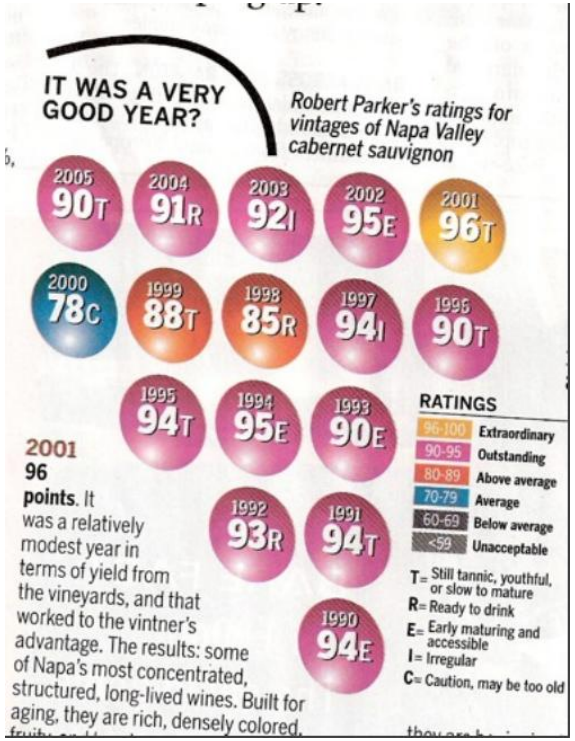


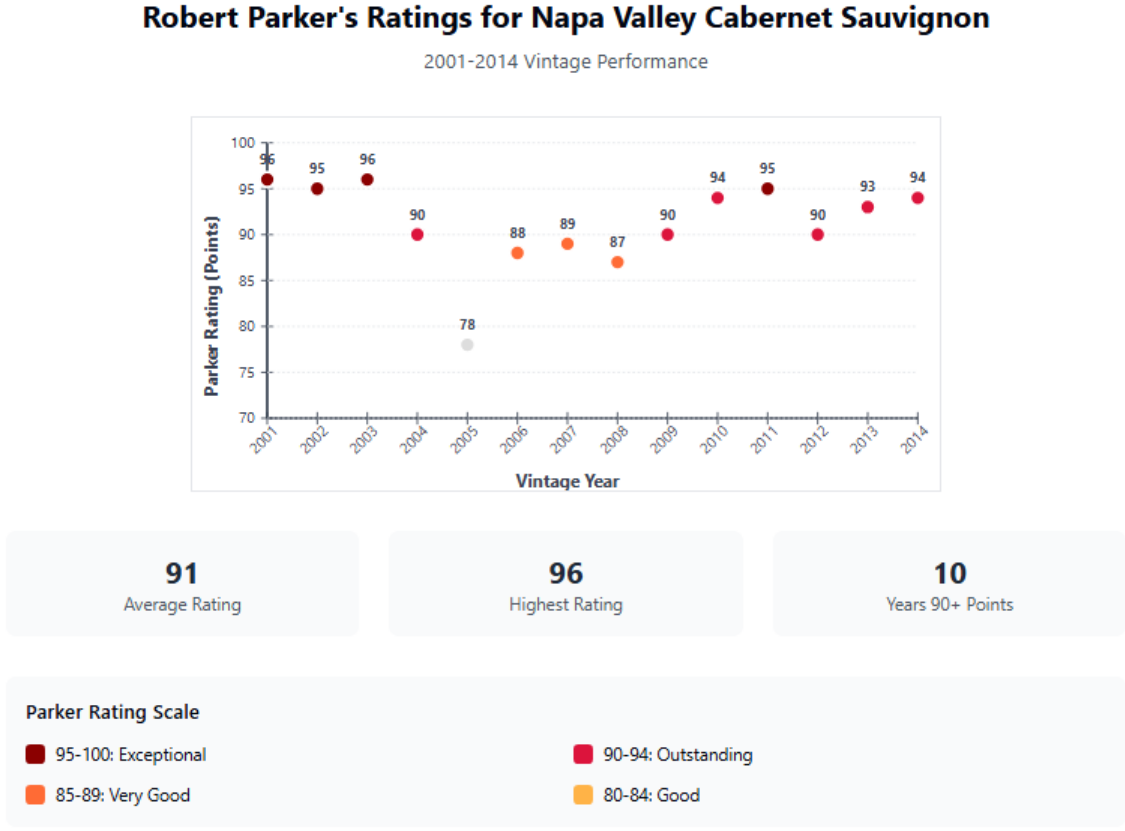
# Cherry Picking the Data



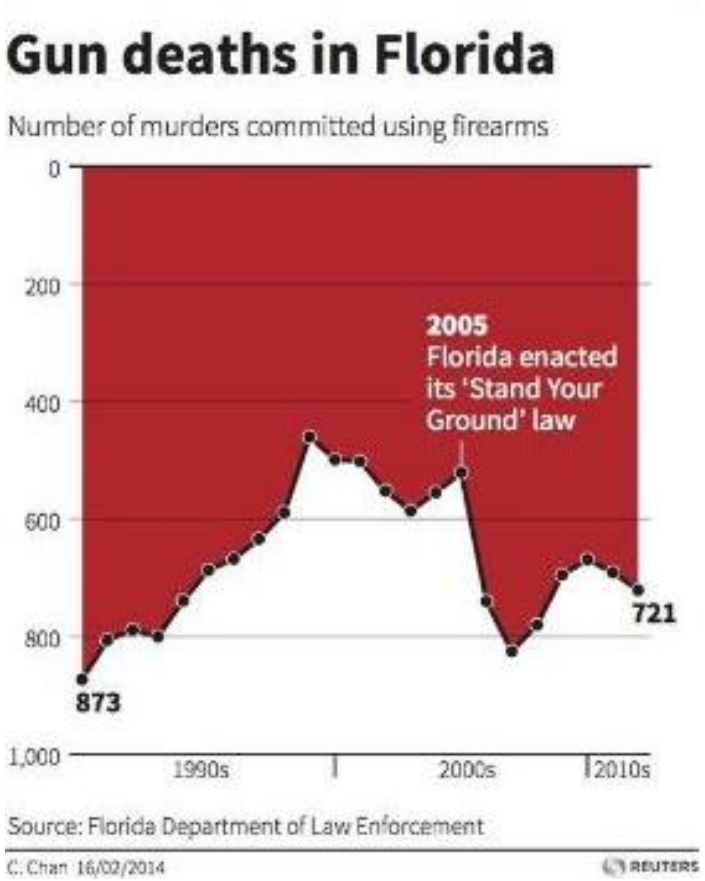



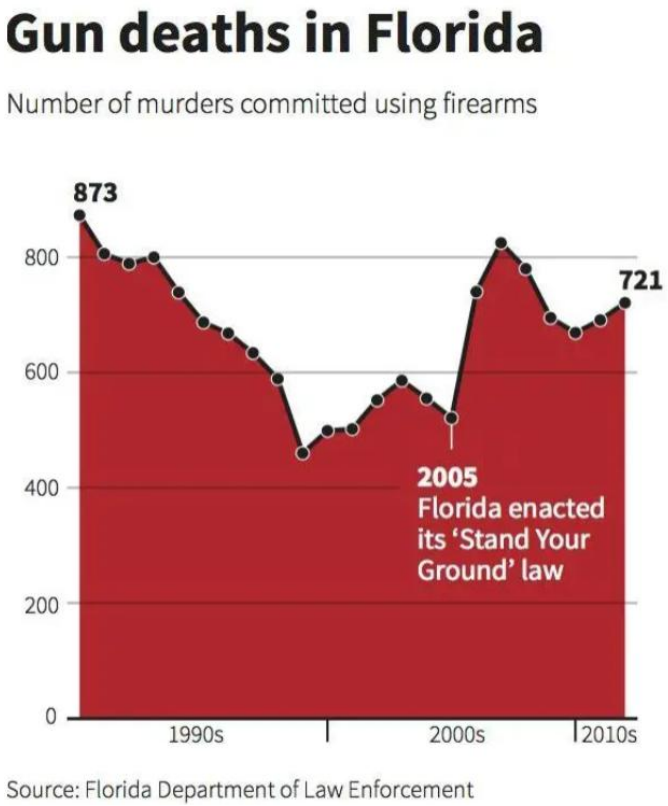
# Using nonsensical representation



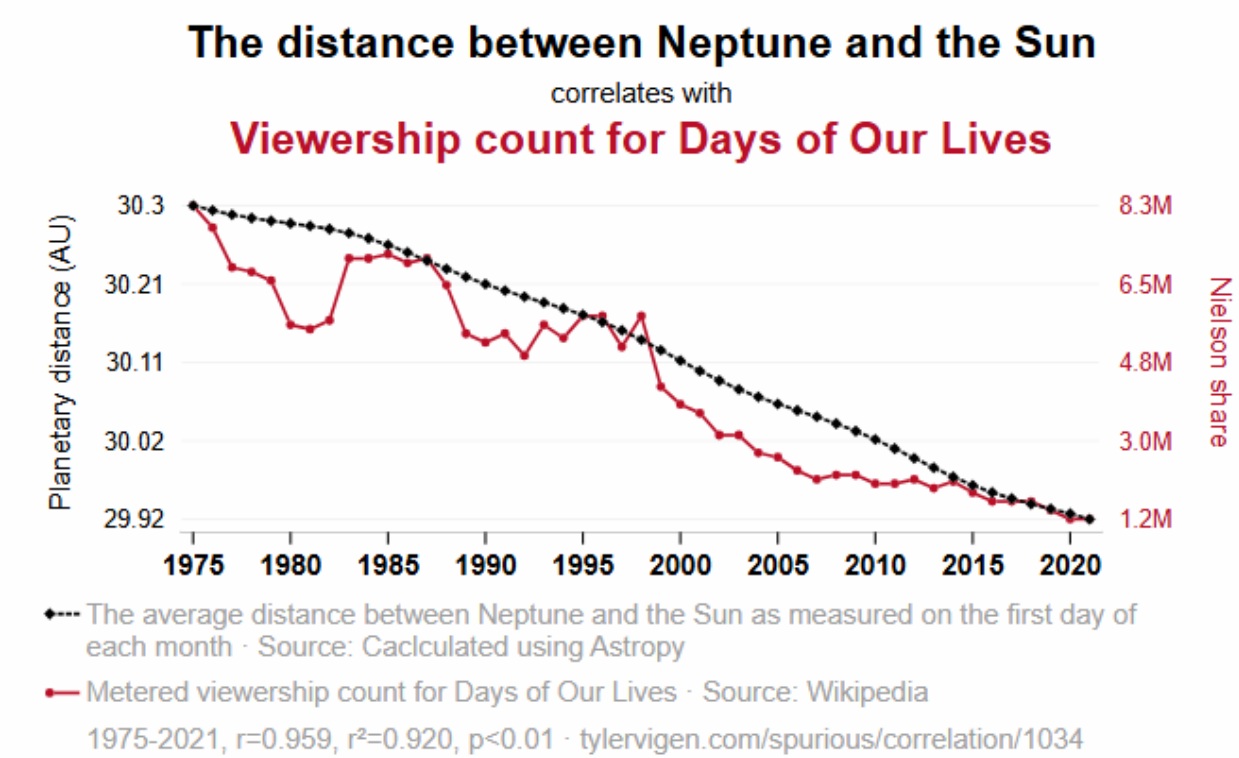



# Going against convention




# Make Spurious Correlations




# Tools of the Trade

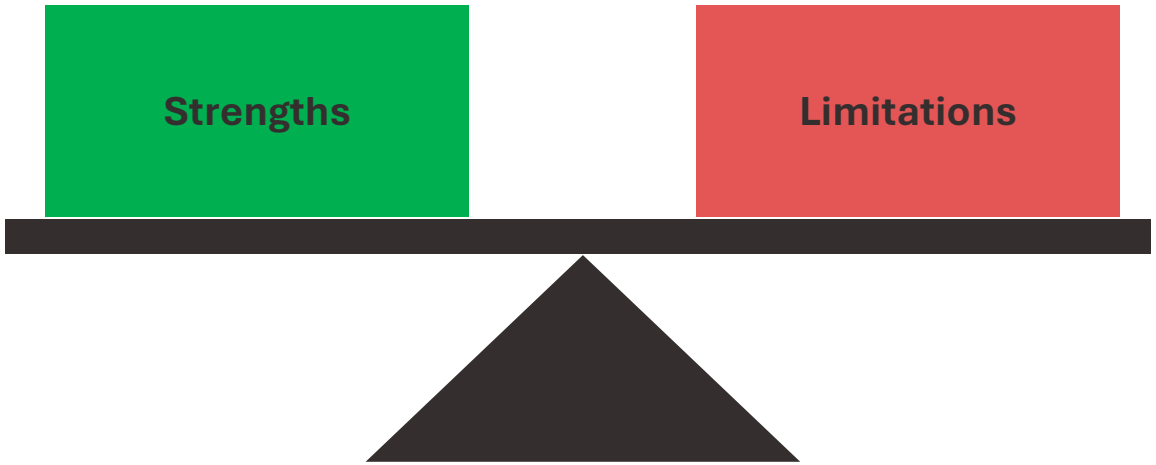
From Excel to advanced platforms

# Visualization Software

| | Advantages | Challenges | Support |
|---|---|---|---|
| Excel | • High familiarity,<br>• Compatible w/ Microsoft products,<br>• Easy to edit/analyze underlying data | • No report portal<br>• VBA & Macros have low familiarity<br>• May need add-ons for some chart types | • How to Build Data Visualizations in Excel: https://stephanieevergreen.com/how-to/ |
| Google Charts | • Free<br>• Interactive charts<br>• Compatible with Google Sheets | • Some familiarity with JavaScript<br>• Difficult to edit or analyze underlying data | • Google Charts: https://developers.google.com/chart |
| Python or R | • Free<br>• Highly customizable charts<br>• Powerful data analysis tool<br>• Interactive charts | • Requires familiarity with the language<br>• Ability to access Jupyter Lab or R Studio | • JupyterLab: https://jupyter.org/<br>• R Studio: https://www.rstudio.com/ |
| Tableau | • Customizable and interactive charts<br>• Integrates with Python or R | • Only Tableau Public is free making your underlying data and visualizations open to the public | • Tableau Public Support: https://www.tableau.com/support/public |

# The Excel Reality Check

Understanding the strengths and limitations for data visualization




## Strengths to Leverage

- Ubiquitous across organizations
- Built-in chart types for common needs
- Real-time chart updates from data changes

## Limitations to Work Around

- Limited advanced formatting options
- Difficult to create complex visualizations
- Version control challenges

# Excel Macros: Automate Your Repetitive Tasks

**What are Macros?**

Recorded sequences of actions that Excel can replay automatically - like having a robot perform your repetitive tasks

## Common Use Cases

**Monthly Report Formatting**
Format headers → Apply conditional formatting → Create charts → Apply branding

**Data Validation Setup**
Create dropdown lists → Set validation rules → Apply protection settings. Apply same rules across all department reports.

**Data Import Cleanup**
Remove blank rows → Standardize formats → Apply validation → Create summary

## More Advanced Workflows

**Chart Creation & Styling**
Build standard compliance dashboard charts with consistent colors, fonts, and layout

**Multi-Sheet Report Generation** Create pivot tables → Generate charts → Format consistently → Compile summary

# How to Get Started with Excel Macros

## Recording Your First Macro

1. Enable Developer Tab:
   File → Options → Customize Ribbon → Check "Developer"
2. Click Developer Tab → Record Macro
3. Give it a descriptive name: FormatMonthlyReport
4. Perform your tasks (Excel watches and remembers everything)
5. Click "Stop Recording" when finished
6. Run anytime with a button or keyboard shortcut

### 🔒 Security & Saving

- Save as .xlsm (macro-enabled) to keep macros
- Enable macros via yellow security bar for trusted files
- Only run macros from trusted sources

### ⚠️ What Macros Can't Do

- Think or make decisions (just replay exact steps)
- Handle varying data layouts well
- Work across different file structures easily

💡 For complex data transformation, use Power Query instead

# “Vibe Coding” for Excel

Let AI write your VBA code - no programming experience required!

## 🤖 Step 1: Get Your Macro

1. Open **Claude.ai** or **ChatGPT**
2. Create free account if needed
3. Copy and paste this exact prompt:

Create an Excel VBA macro that formats charts with Northwestern University brand colors. The macro should check if charts exist, format the first chart found, and apply these specific colors to data series in order: Series 1: RGB(72, 36, 118), Series 2: RGB(0, 134, 86), Series 3: RGB(13, 45, 108), Series 4: RGB(255, 197, 32), Series 5: RGB(80, 145, 205). Also format the title with Northwestern purple, remove gridlines, and include error handling.

# “Vibe Coding” for Excel

Let AI write your VBA code - no programming experience required!

## 🤖 Step 1: Get Your Macro

1. Open **Claude.ai** or **ChatGPT**
2. Create free account if needed
3. Copy and paste this exact prompt:

Create an Excel VBA macro that formats charts with Northwestern University brand colors. The macro should check if charts exist, format the first chart found, and apply these specific colors to data series in order: Series 1: RGB(72, 36, 118), Series 2: RGB(0, 134, 86), Series 3: RGB(13, 45, 108), Series 4: RGB(255, 197, 32), Series 5: RGB(80, 145, 205). Also format the title with Northwestern purple, remove gridlines, and include error handling.

## 📊 Step 2: Create Chart

1. Open Excel, create new workbook and save as .xlsm
2. Copy this data into cells A1:F5:

```
Quarter,Enrollment,Revenue,Research,Satisfaction,Alumni
Q1,15420,28.5,12.3,8.2,5.1
Q2,16890,31.2,14.7,8.7,5.8
Q3,17250,33.8,15.9,9.1,6.2
Q4,18100,35.4,17.2,9.4,6.7
```

3. Select all data (A1:F5)
4. Insert → Charts → Line Chart
5. Add chart title: **"Northwestern Metrics"**

# “Vibe Coding” for Excel

Let AI write your VBA code - no programming experience required!

## 🤖 Step 1: Get Your Macro

1. Open **Claude.ai** or **ChatGPT**
2. Create free account if needed
3. Copy and paste this exact prompt:

Create an Excel VBA macro that formats charts with Northwestern University brand colors. The macro should check if charts exist, format the first chart found, and apply these specific colors to data series in order: Series 1: RGB(72, 36, 118), Series 2: RGB(0, 134, 86), Series 3: RGB(13, 45, 108), Series 4: RGB(255, 197, 32), Series 5: RGB(80, 145, 205). Also format the title with Northwestern purple, remove gridlines, and include error handling.

## 📊 Step 2: Create Chart

1. Open Excel, create new workbook and save as .xlsm
2. Copy this data into cells A1:F5:

```
Quarter,Enrollment,Revenue,Research,Satisfaction,Alumni
Q1,15420,28.5,12.3,8.2,5.1
Q2,16890,31.2,14.7,8.7,5.8
Q3,17250,33.8,15.9,9.1,6.2
Q4,18100,35.4,17.2,9.4,6.7
```

3. Select all data (A1:F5)
4. Insert → Charts → Line Chart
5. Add chart title: **"Northwestern Metrics"**

## ⚡ Step 3: Run the Macro

1. Press Alt + F11 (opens VBA editor)
2. Insert → Module
3. Paste the AI-generated code
4. Press F5 to run macro
5. Return to Excel (Alt + F11)
6. Watch your chart transform!

💡 *If macro doesn't work, enable macros in Excel security settings or save as .xlsm file*

> A journey of a thousand miles begins with a single step.
>
> Lao Tzu

# From Principles to Practice

Building your data visualization expertise

# Action Items

**Next 30 days:**

Focus on data structure. Clean up one recurring report using the principles we discussed. Create templates for consistent formatting.

**Next 60 days:**

Experiment with three new chart types from our encyclopedia. Replace at least one table with a visualization.

**Next 90 days:**

Apply the design principles to redesign your most important dashboard or report.

# Tools and Resources

- "Resonate" by Nancy Duarte - Audience-focused presentation structure
- Stephanie Evergreen's Blog (stephanieevergreen.com) - Data visualization excellence

# Share! Upload Reports to a Repository








- Creates a permanent link (DOI).
- Content is findable by search engines such as Google.

*Note: The content is a permanent deposit into the repository. All authors listed on poster or presentation should provided their approval.*

# Questions and Comments

# Thank you

Northwestern Medicine®
Feinberg School of Medicine