

# Telling a Story with Data: Basic Principles and Best Practices



Melissa Cantrell, Scholarly Communication Librarian, University Libraries  
[melissa.cantrell@colorado.edu](mailto:melissa.cantrell@colorado.edu)

Ellery Galvin, Statistics Specialist, Applied Mathematics  
[ellery.galvin@colorado.edu](mailto:ellery.galvin@colorado.edu)



Center for Research Data & Digital Scholarship  
UNIVERSITY OF COLORADO **BOULDER**

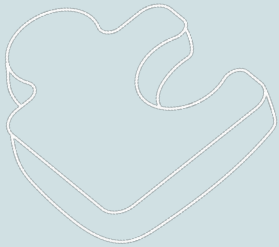


# Agenda

1. The Humanity of Design
  - a. What and Why
  - b. Story vs. message
2. Elements of Design
  - a. The right fit
  - b. Aesthetics
  - c. Legibility
3. Applying Principles of Design



# 1. The Humanity in Design



What is data visualization, and when is it needed?

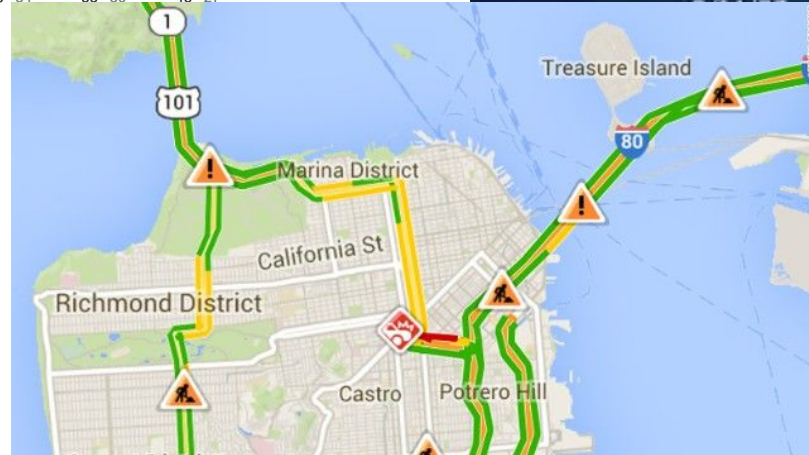
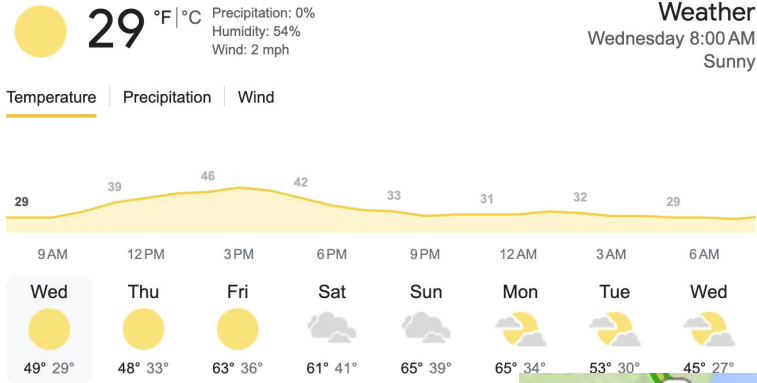
## What?

Data visualization is a **graphical display of information** for two purposes: sense-making (i.e. data analysis) and communication.

## Why?

Visual displays **give form to information** which before had none. Human visual perception and cognition is particularly strong.

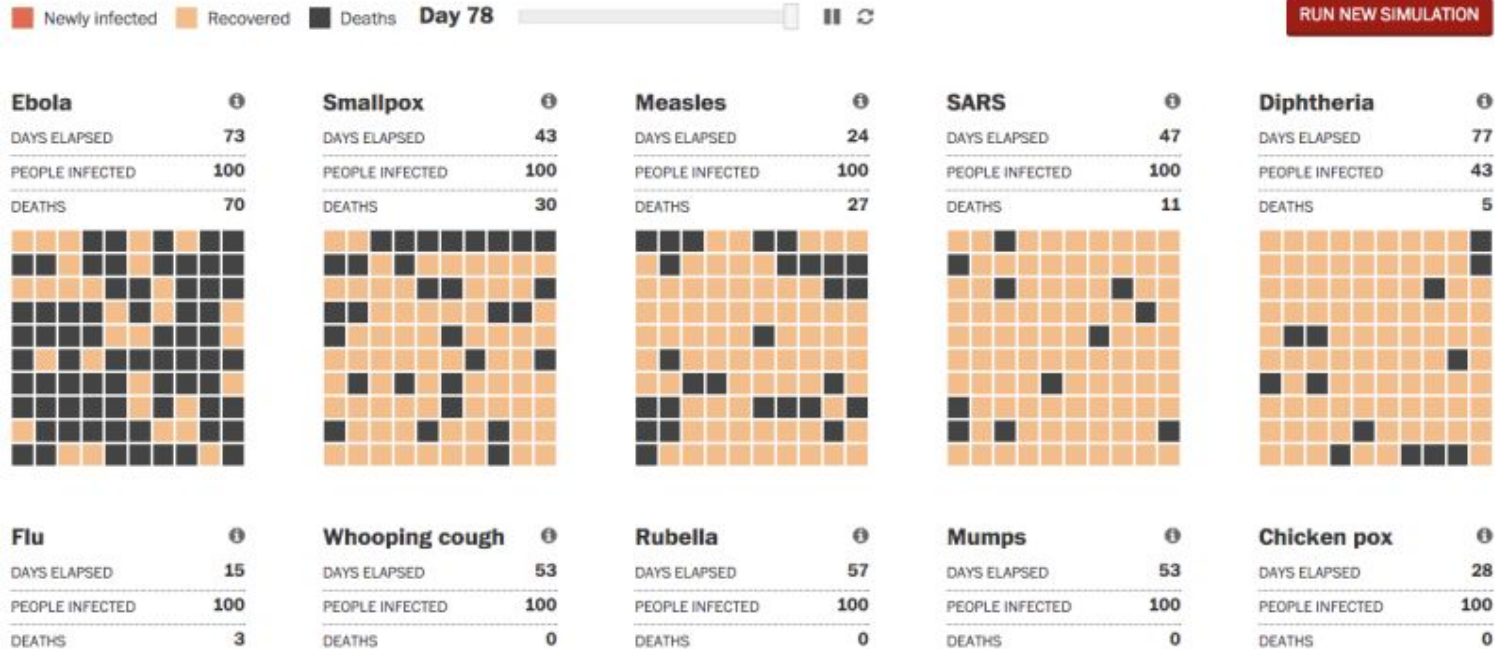
# We read/process data visualizations all the time



# Do you want to convey a specific message?

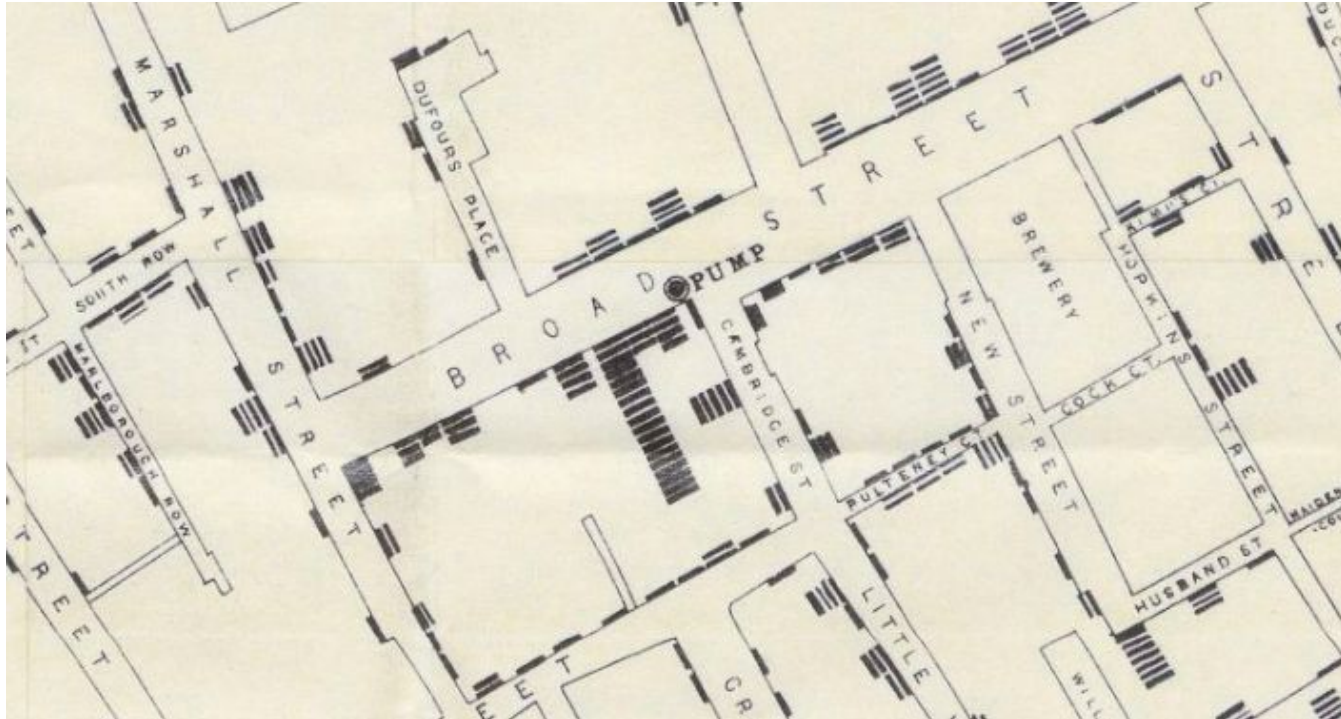
## Ebola spreads slower, kills more than other diseases

This simulation shows how quickly 10 diseases, from more fatal to less fatal, could **spread from one person to 100 unvaccinated people**.





Do you want to explore your data to get new insights?

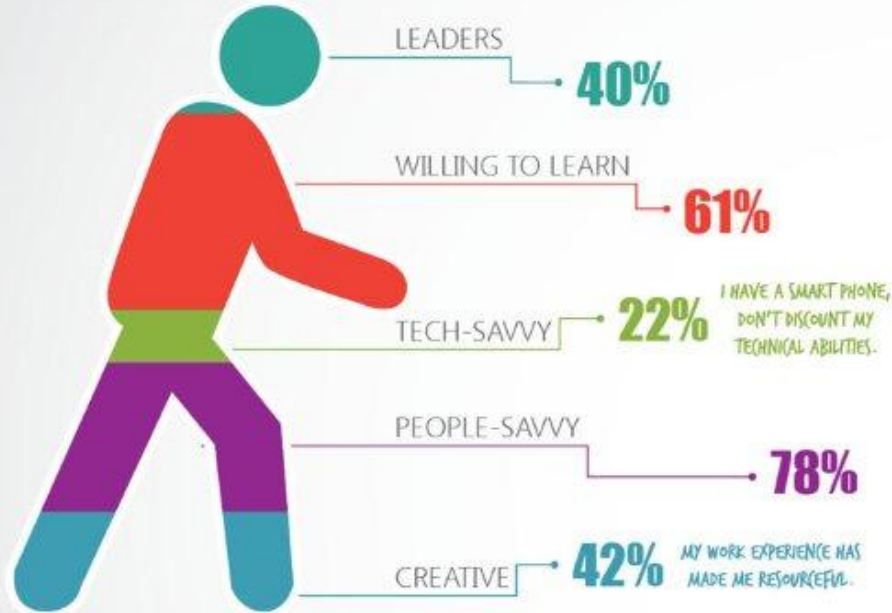


Dr. John Snow,  
*The Cholera  
Epidemic in  
London, 1854*



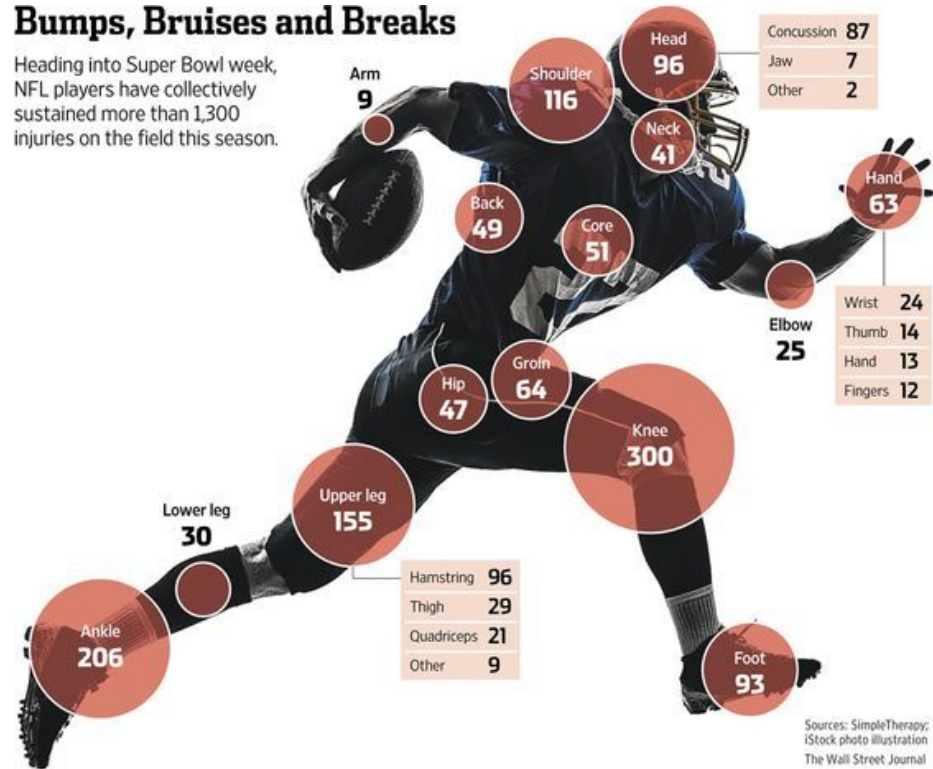
# Visualizing data can illuminate or obfuscate

## HOW BABY BOOMERS DESCRIBE THEMSELVES



## Bumps, Bruises and Breaks

Heading into Super Bowl week, NFL players have collectively sustained more than 1,300 injuries on the field this season.



Sources: SimpleTherapy;  
iStock photo illustration  
The Wall Street Journal

# Story vs. Message

# Why tell stories with data?

- Storytelling has been central to human learning for millenia
- Storytelling makes information more persuasive and memorable
- Storytelling adds texture and form to make information more easily digestible

**Just as important to your narrative is understanding competing stories and narratives**



The **message** is the key piece of your story that turns **insight into action**.

Effectively turning insight into action through your message hinges on **knowing and understanding your audience**.



Understanding your audiences

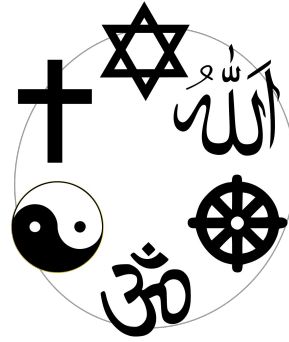




Who needs to hear  
your message, and  
who can act on it?



Needs



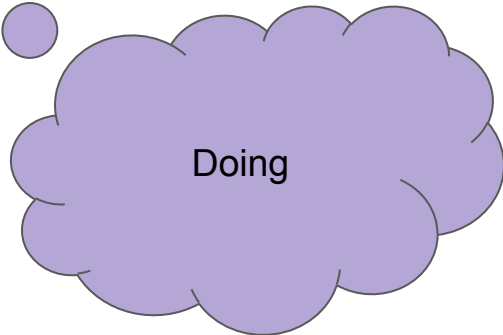
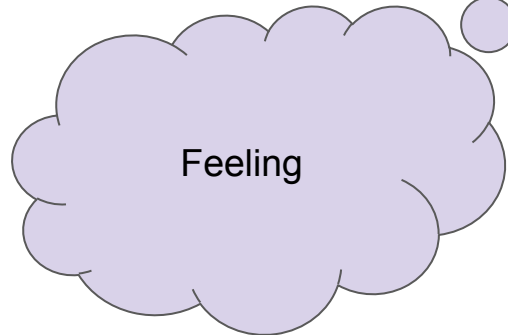
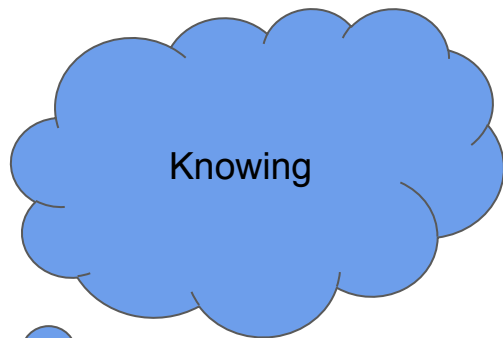
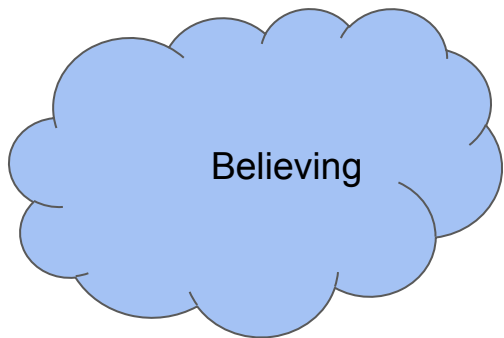
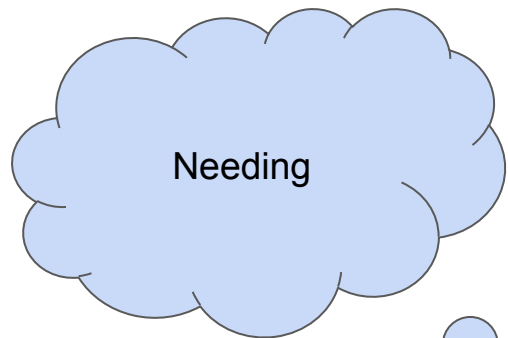
Beliefs



Knowledge



zoom in on the viewing moment



The Viewer

# Viewing Context

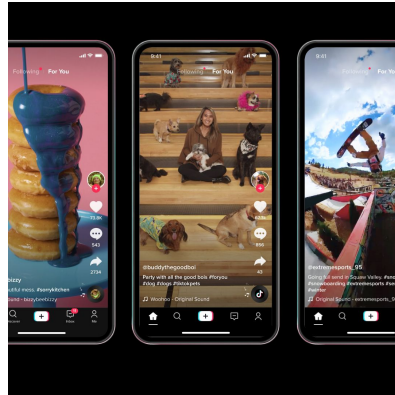
# Location



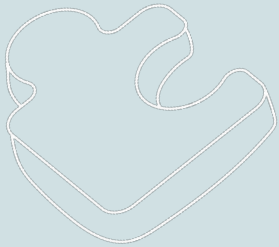
# Medium



# Time



## 2. Elements of Design and Best Practices



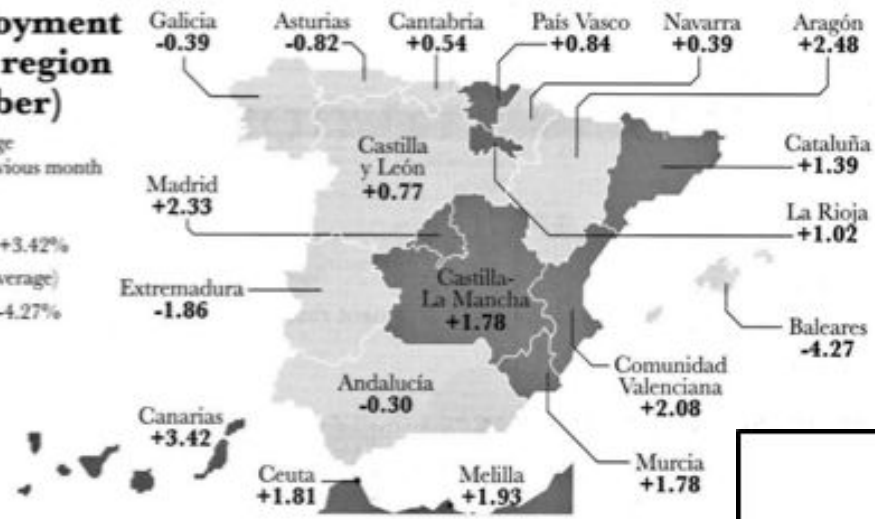


Finding the right fit

**Unemployment rates by region (in October)**

Percentage change compared to previous month

- +0.83% a +3.42%
- +0.82% (average)
- +0.82% a -4.27%



Both of these visualizations use the same set of data on unemployment rates for October in the various regions of Spain



# Seven Major Types of Data Visualizations

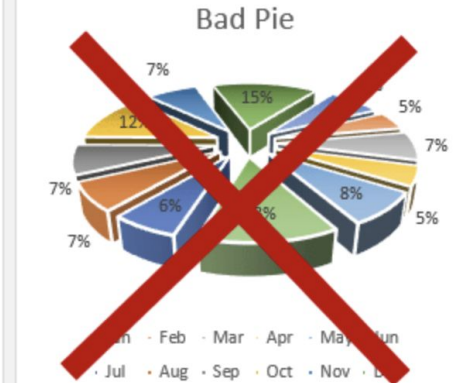
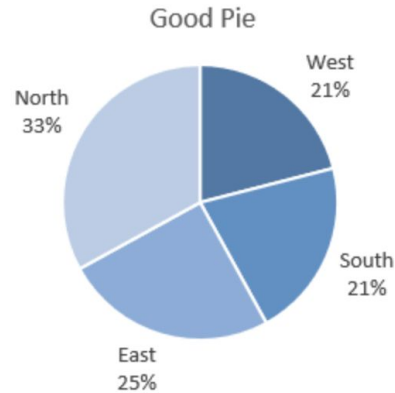
1. Composition
2. Comparative
3. Distribution
4. Relational
5. Temporal
6. Geospatial
7. Word/Topic Clouds

# Composition

Used to show **parts of a whole**.

- **Pie Charts**
- Bar / Column Charts
- Area Charts

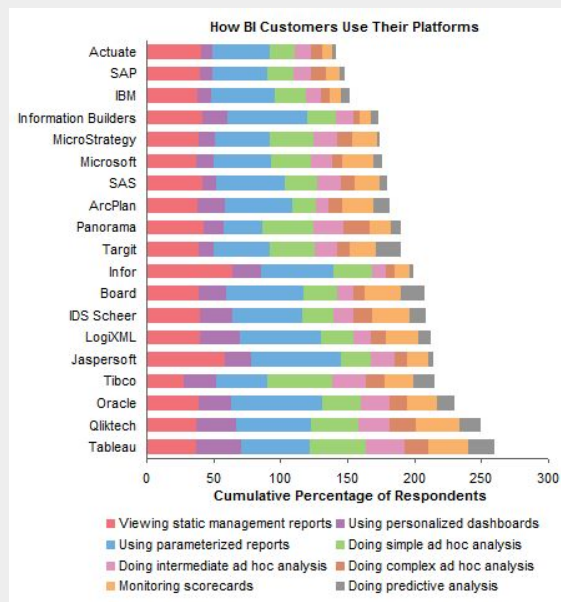
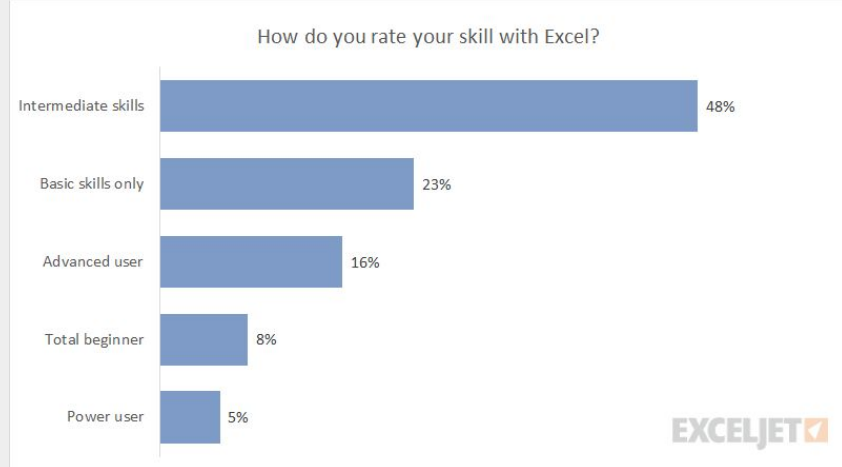
## Best Practices for Using Pie Charts

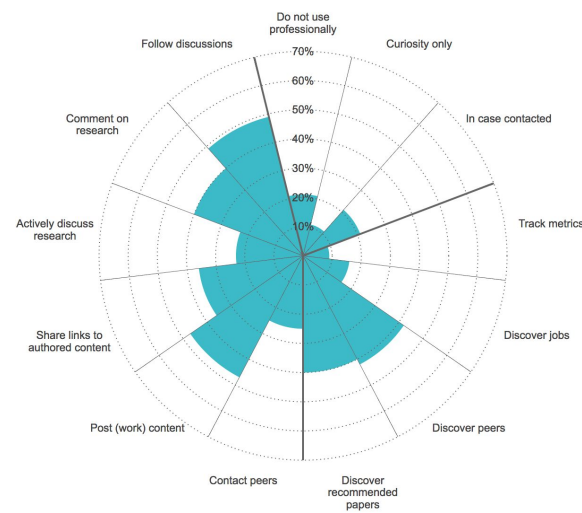


# Composition

Used to show **parts of a whole**.

- Pie Charts
- **Bar / Column Charts**
- Area Charts

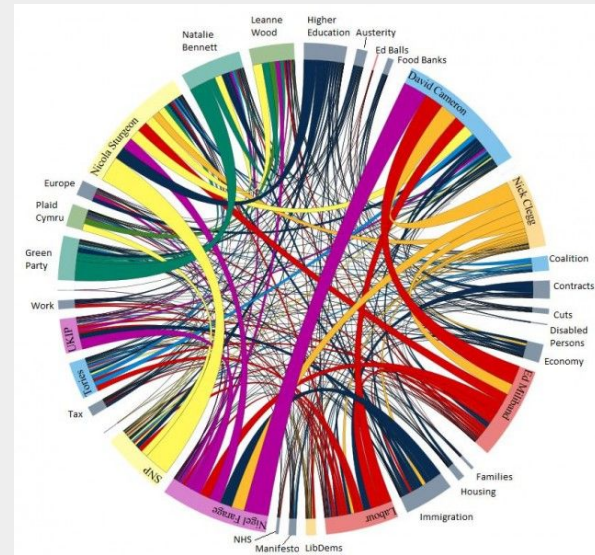




# Comparative

Use for comparing, ranking, contrasting data.

- Bar / Column Charts
- **Circular Area Charts**
- Bubble Charts
- Line Charts



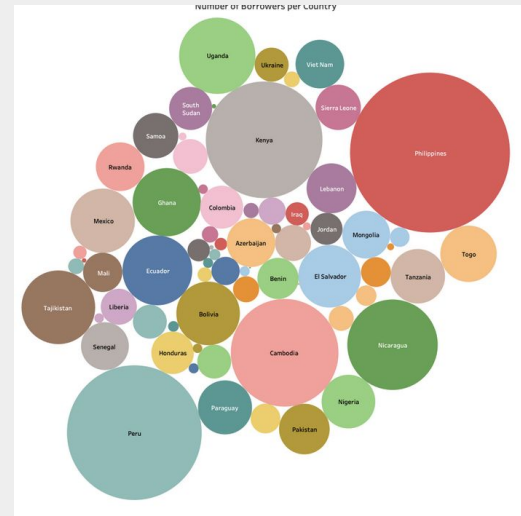
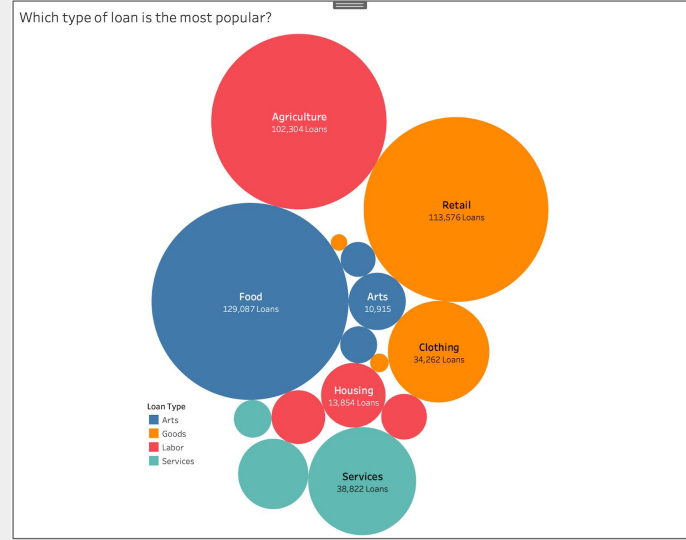


# Comparative

Use for comparing, ranking, contrasting data.

- Bar / Column Charts
- Circular Area Charts
- **Bubble Charts**
- Line Charts

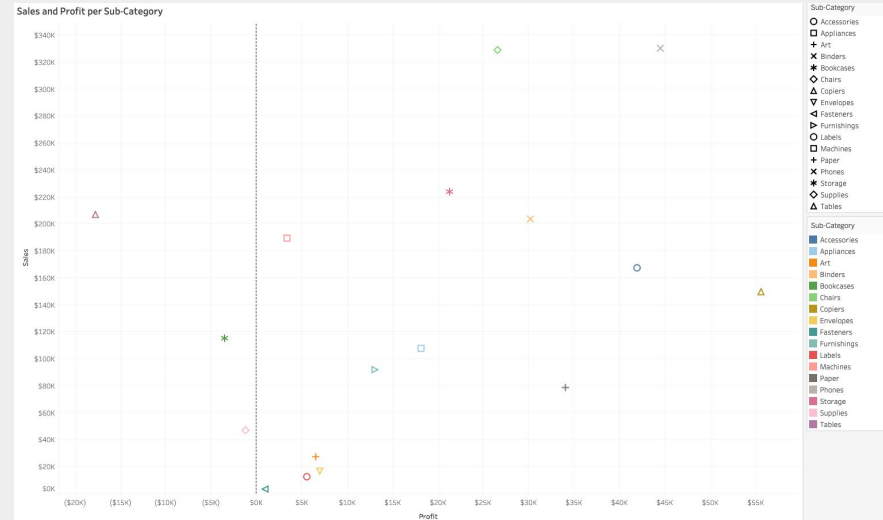
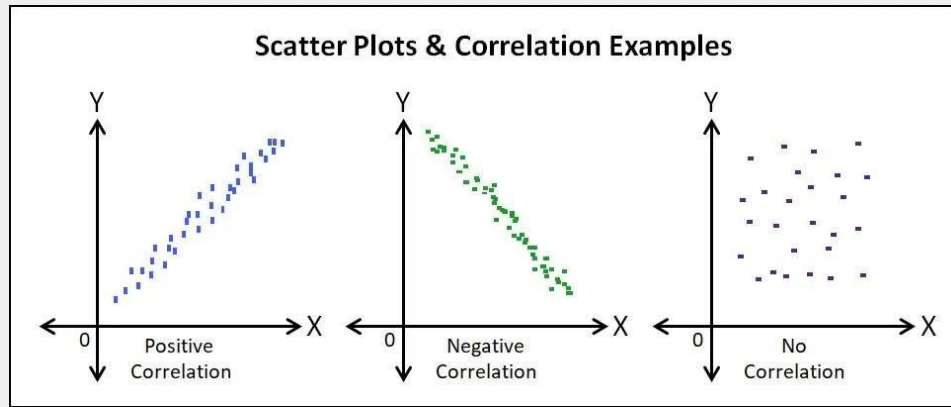
<https://www.tableau.com/chart/what-is-bubble-chart>



# Distribution

Use to visualize correlations and interactions between values.

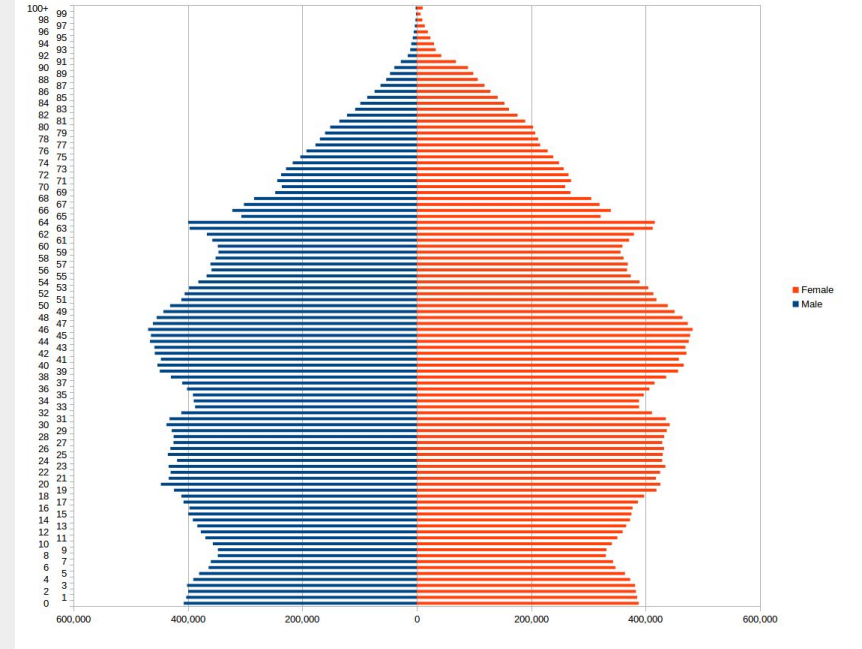
- **Scatter Plots**
- Histograms
- Alluvial Diagrams
- Geospatial
- Heat Maps



# Distribution

Use to visualize correlations and interactions between values.

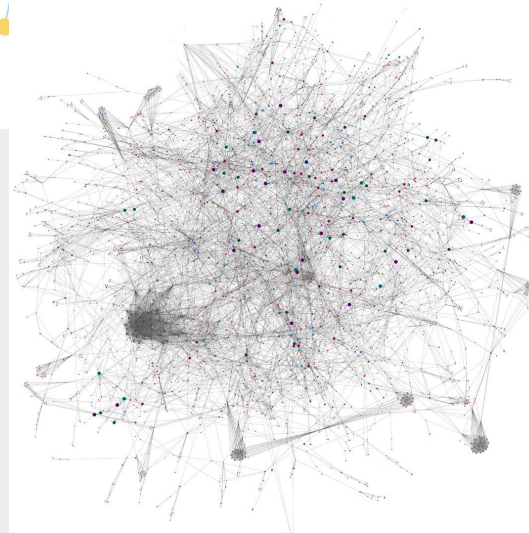
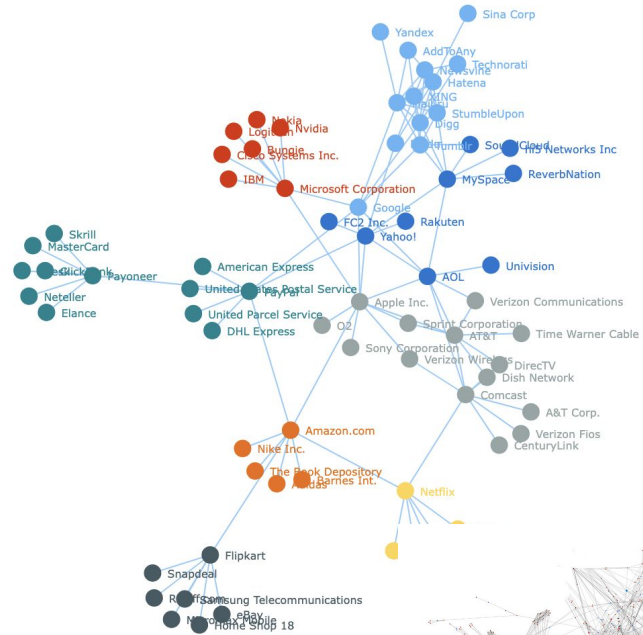
- Scatter Plots
- **Histograms**
- Alluvial Diagrams
- Geospatial
- Heat Maps



# Relational

Use for visualizing and tracing relationships and dependencies.

- **Network Graphs**
- Alluvial Diagrams
- Tree Graphs
- Treemaps



# Relational

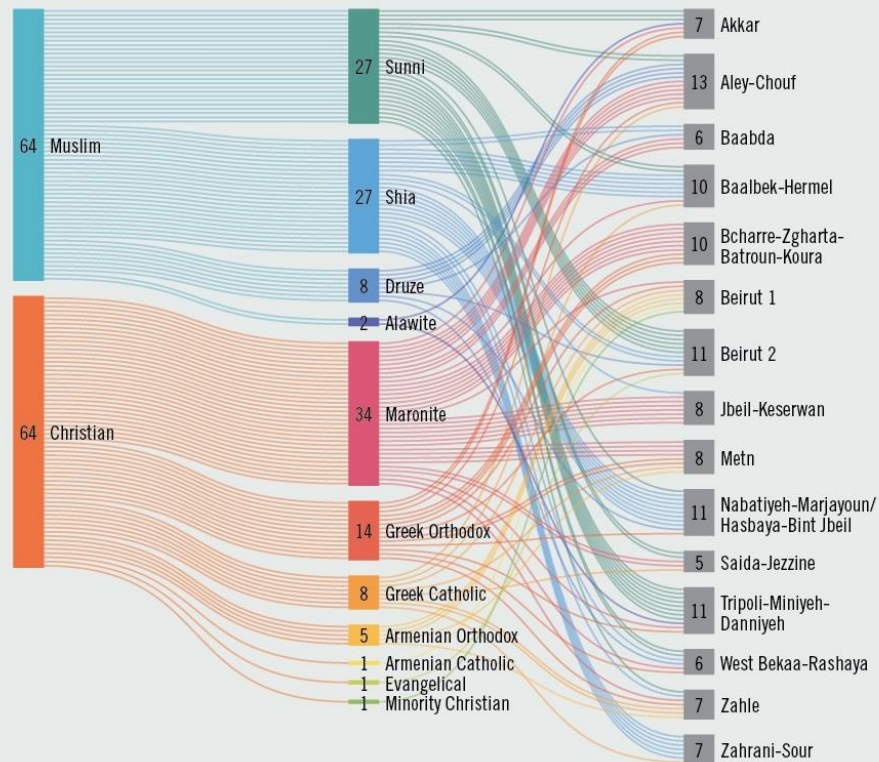
Use for visualizing and tracing relationships and dependencies.

- Network Graphs
- **Alluvial Diagrams**
- Tree Graphs
- Treemaps

## 2017 LEBANESE ELECTION LAW

### 2018 SEATS BY CONFESSION

Distribution of seats by confession and district



Source: 2017 Election Law; Ministry of Interior and Municipalities.

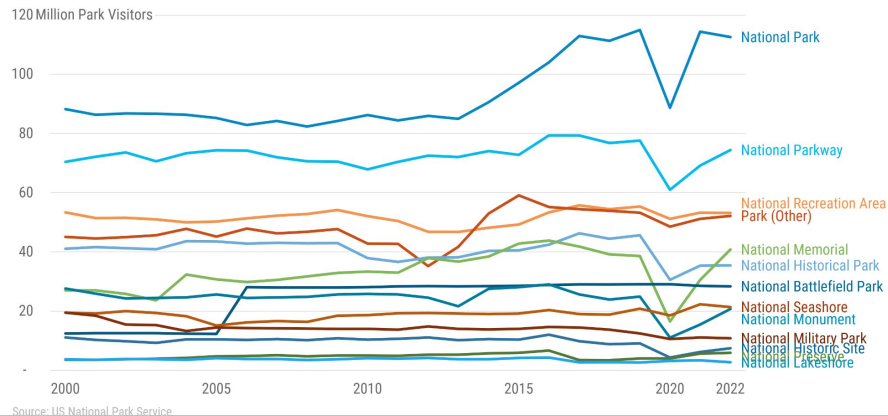
© 2017 Executive.com.lb

# Temporal

Use to show change over time.

- **Line, Bar, Column Charts**
- Heat Maps
- Gantt Charts

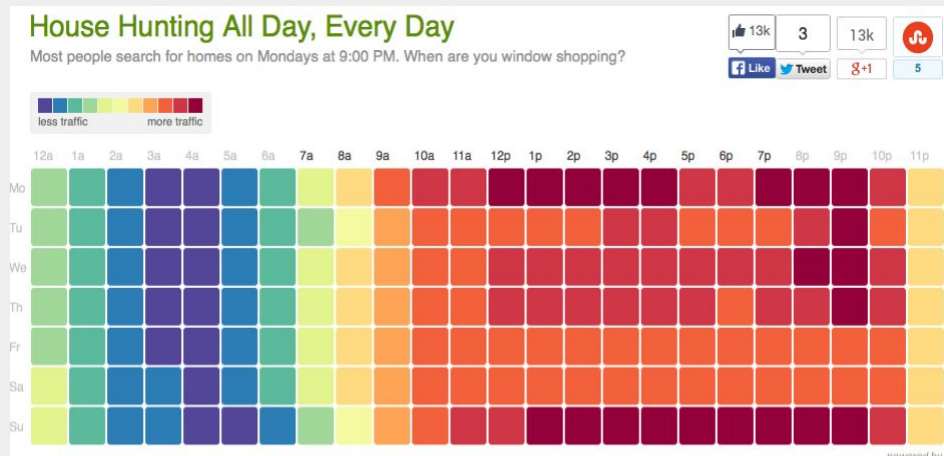
US national park visitation by designated park type.



# Temporal

Use to show change over time.

- Line, Bar, Column Charts
- **Heat Maps**
- Gantt Charts

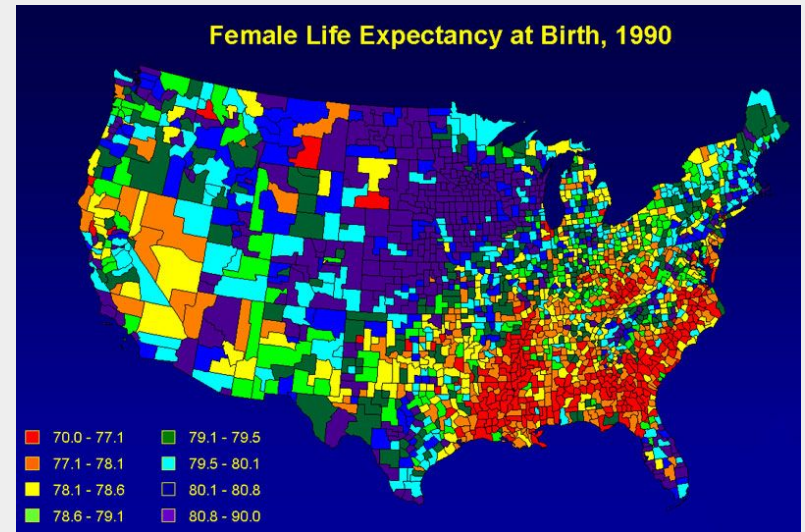
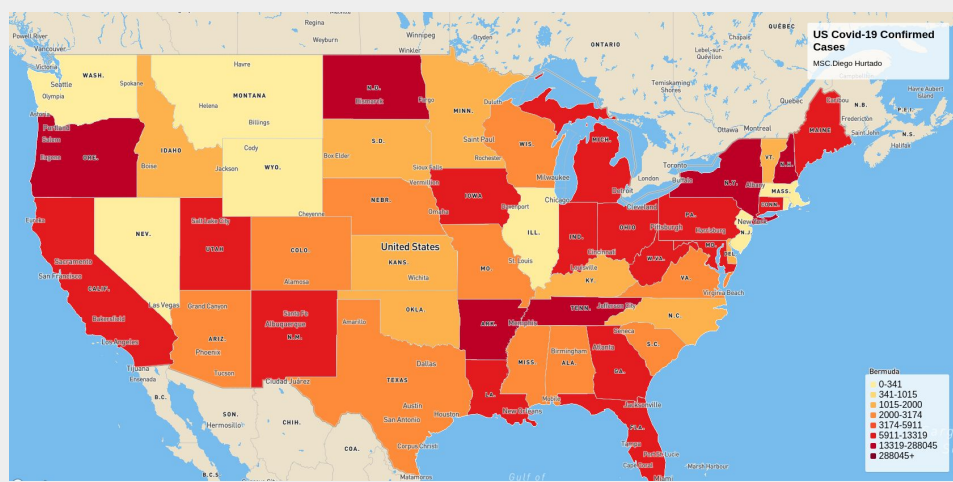




# Geospatial

Use to illustrate spatial relationships and distributions.

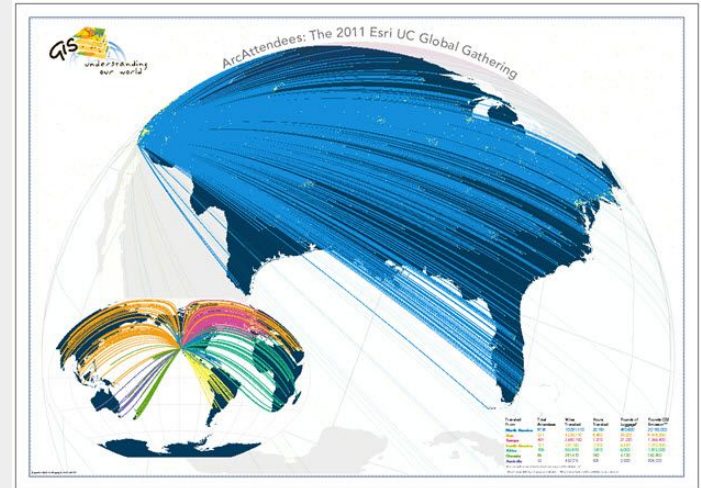
- **Choropleth**
- Flow Maps
- Dot and Bubble Maps
- Network Maps



# Geospatial

Use to illustrate spatial relationships and distributions.

- Choropleth
- **Flow Maps**
- Dot and Bubble Maps
- Network Maps



# Word/Topic Clouds

Use to show frequency  
and co-occurrence of terms.



# Elements of design for aesthetics

# Use of Space

## IR Offices: Types of Staff

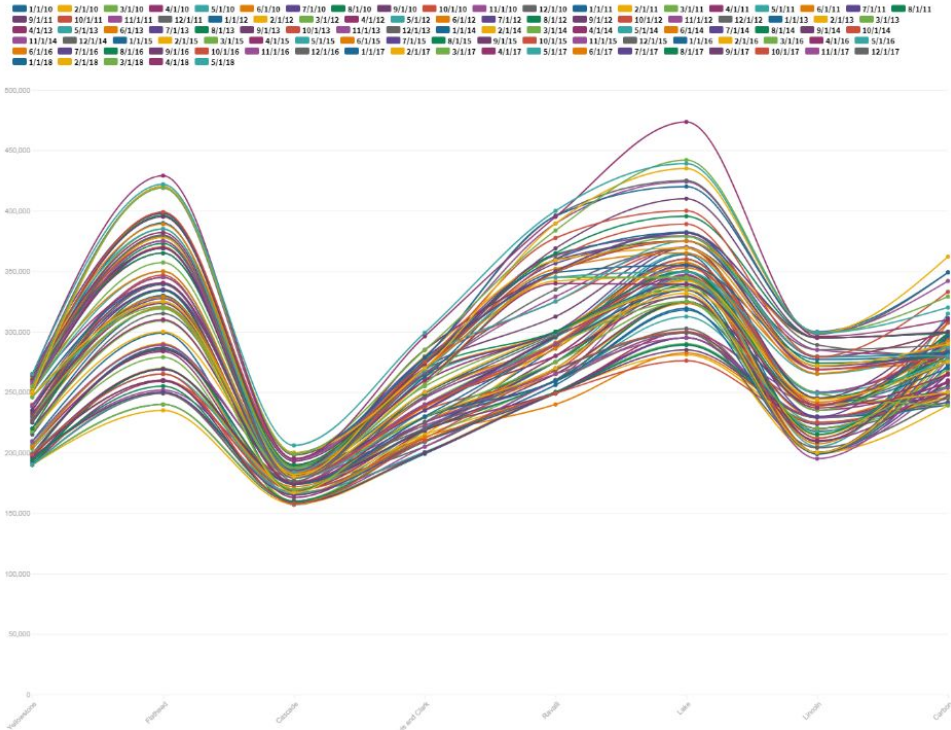
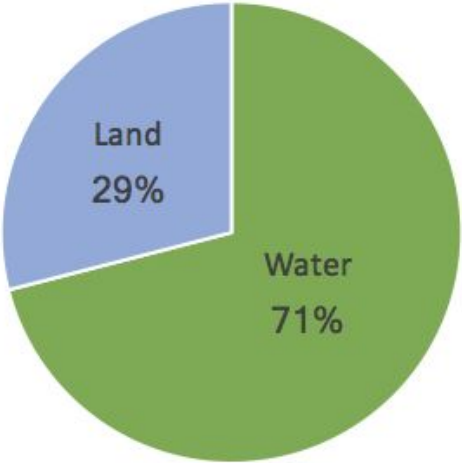


## WEDDING BUDGET BREAKDOWN



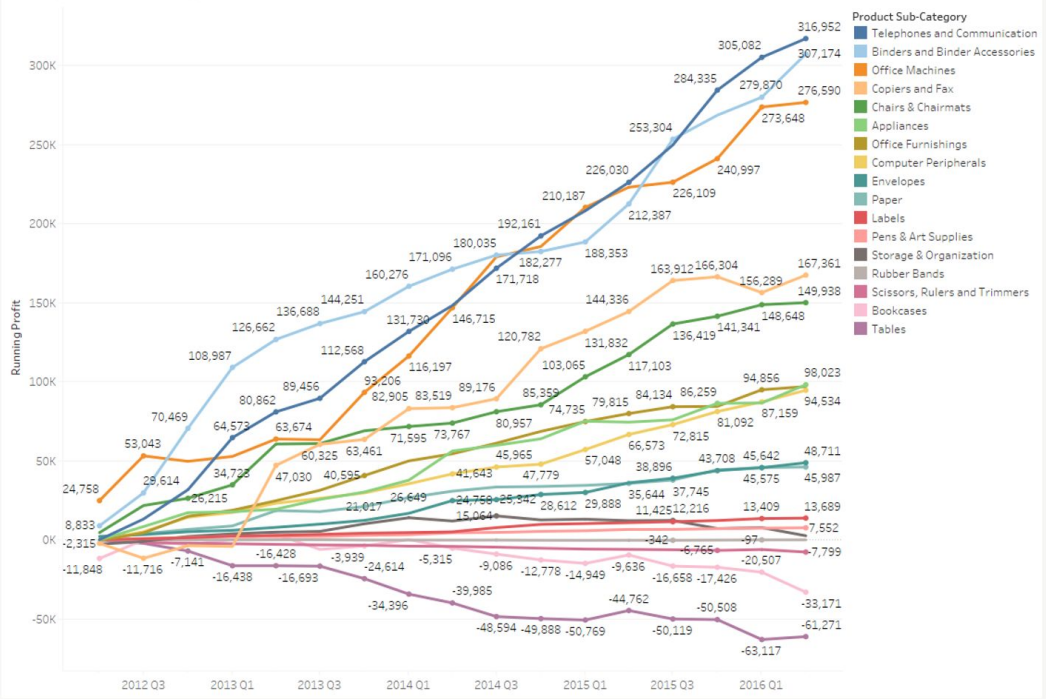
# Color

Land to Water Ratio on Planet Earth

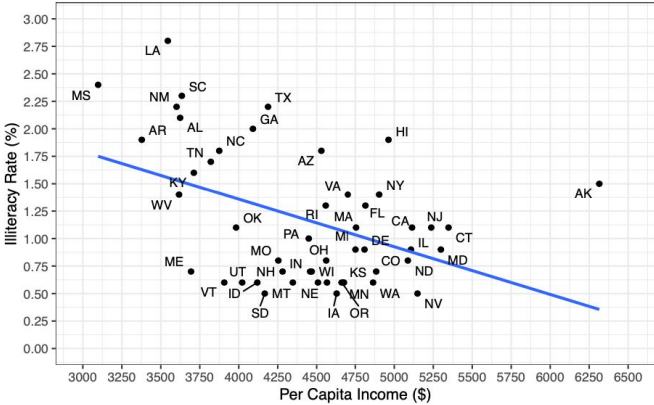




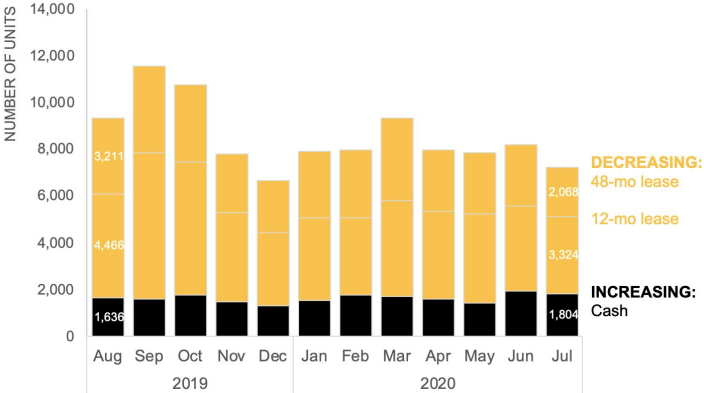
Text



US State Facts (1977)



Sales over time by purchase type





# Text

## Deviation

Emphasize variations ( $\pm$ ) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average. Can also be used to show sentiment (positive/negative).

**Example FT uses**  
Trade surplus/deficit, climate change



**Diverging stacked bar**  
Perfect for presenting survey results which include sentiment (eg disagree/neutral/agree).



**Surplus/deficit filled line**  
The shaded area of these charts allows a balance to be shown – either against a baseline or between two series.

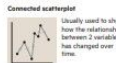
## Correlation

Show the relationship between two or more variables. Be careful that unless you tell them otherwise, many readers will assume the relationship you show them to be causal (i.e. one causes the other).

**Example FT uses**  
Inflation and unemployment, income and life expectancy



**Columns + line timeline**  
A good way of showing the relationship between two continuous variables, each of which has its own axis.



**Bubble**  
Like a scatterplot, but adds additional detail by using the circle according to a third variable.

**XY heatmap**  
A good way of showing the patterns between 2 categories of data, less effective at showing fine differences in amounts.



**Bump**  
Effective for showing changing settings across multiple data. For large datasets, consider grouping lines using colour.

## Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest or equality in the data.

**Example FT uses**  
Wealth, deprivation, league tables, constituency election results



**Ordered column**  
See above.

**Ordered proportional symbol**  
Use when there are big variations between values and/or seeing fine differences between data is not so important.

**Dot strip plot**  
Data placed in order on a strip are a space efficient method of laying out ranks across multiple categories.

**Slope**  
Perfect for showing how rates have changed over time or vary between categories.



**Bump**  
Effective for showing changing settings across multiple data. For large datasets, consider grouping lines using colour.

## Distribution

Show values in a dataset and how often they occur. The chart for 'bins' of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data.

**Example FT uses**  
Income distribution, population, Gephared distribution, reading inequality



**Dot plot**  
A simple way of showing the change or range (interval) of an amount (columns) across multiple categories.

**Dot strip plot**  
Good for showing individual values in a distribution, seeing a problem when too many data have the same value.

**Barnard plot**  
Like dot strip plots, good for displaying all the data in a table. They work best when highlighting individual values.

**Boxplot**  
Summarise multiple distributions by showing the median (central) and range of the data.

**Violin plot**  
Similar to a box plot but more effective with complex distributions (data that cannot be summarised with simple average).

**Population pyramid**  
A standard way for showing the age and sex composition of a population distribution effectively, back to back histograms.

## Change over Time

Give emphasis to changing trends. These can be short (days) or long (decades) or extended series (moving averages or smoothed). Showing the correct time period is important to provide suitable context for the reader.

**Example FT uses**  
Share price movements, economic time series, sector or changes in a market



**Column**  
Columns work well for showing change over time, but usually best with only one series of data at a time.

**Column + line timeline**  
A good way of showing the relationship over time between an amount (columns) and a rate (line).

**Slope**  
Good for showing change over time as long as the data can be simplified into 2 or 3 points without missing a key part of story.

**Area chart**  
Use with care – these are good at showing changes to totals, but using change in components can be very difficult.

**Candlestick**  
Usually focused on day-to-day activity, these charts show opening/closing and high/low points of each day.

**Fan chart (projections)**  
Use to show the uncertainty in future projections – usually this grows the further forward to projection.

## Magnitude

Show size comparisons. These can be relative (just being able to see larger/smaller) or absolute (need to see fine differences). Usually these show a 'counted' number (for example, barrels, dollars or people) rather than a calculated rate or per cent.

**Example FT uses**  
Commodity production, market capitalisation, volumes in general



**Bar**  
See above. Good when the data are not time series and labels have long category names.

**Paired column**  
As per standard column but allows for multiple series. Can become tricky to read with more than 2 series.

**Paired bar**  
See above.

**Heatmap**  
A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.

**Proportional symbol**  
Use when there are big variations between values and/or seeing fine differences between data is not so important.

**Isotype (person)**  
Excellent solution in some instances – use only with small numbers (do not slice off or aim to represent a decimal).

## Part-to-whole

Show how a single entity can be broken down into its component elements, if the reader's interest is solely in the size of the components, consider using a magnitude type chart instead.

**Example FT uses**  
Fossil budgets, company structures, national election results



**Marimark**  
A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.

**Pie**  
A common way of showing part-to-whole data – but be aware that it's difficult to accurately compare the size of the segments.

**Donut**  
Similar to a pie chart – but the centre can be a good area of realising space to include more information about the data (eg total).

**Tree map**  
Use for hierarchical data – showing the size and proportion of data at the same time – as long as the data are not too complicated.

**Vertical**  
A way of turning points into areas – any point within each area is closer to the central point than any other element.

**Arc**  
A heuristic often used for visualising composition by number of wells.

## Spatial

Avoid from locator maps only used when precise location or geographical patterns in data are more important to the reader than anything else.

**Example FT uses**  
Population density, natural resource locations, natural disaster risk/impact, settlement areas, variation in election results

**Basic choropleth (rate/ratio)**  
The standard approach for putting data on a map – should always be used rather than a map – use a sensible base geography.

**Proportional symbol (count/magnitude)**  
Use for totals rather than rates – but may be used if small differences in data will be hard to see.

**Flow map**  
For showing unambiguous movement across a map.

**Contour map**  
For showing areas of equal value on a map. Can use deviation colour schemes for showing  $\pm$  values.

**Equalized cartogram**  
Converting each unit on a map to a regular and equally-sized shape – good for representing voting regions with equal value.

**Scaled cartogram (value)**  
Stretching and absorbing a map so that each area is sized according to a particular value.

**Dot density**  
Used to show the location of individual events/locations – make sure to annotate any patterns the reader should use.

## Flow

Show the reader volumes or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations.

**Example FT uses**  
Movement of funds, trade, migrants, lawsuits, information, relationship graphs.

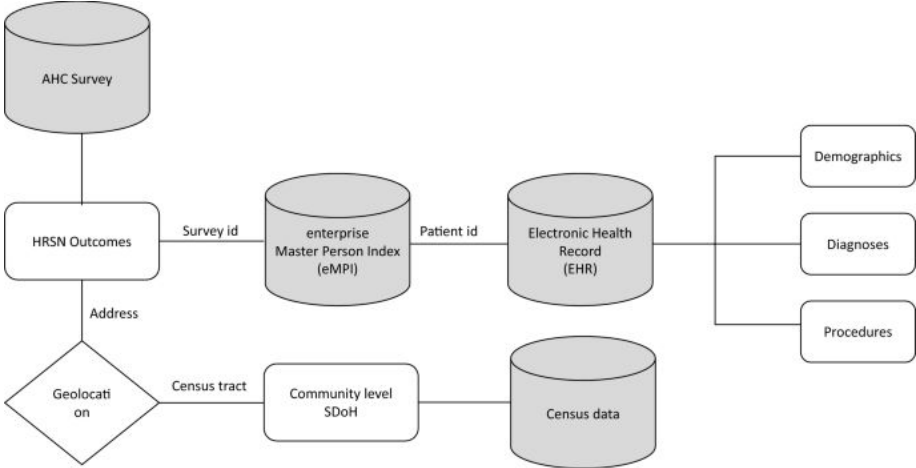
**Sankey**  
Shows changes in flows from one condition to at least one other, good for seeing the eventual outcome of a complex process.

**Waterfall**  
Designed to show the sequencing of data through a flow process. Can include  $\pm$  components.

**Chord**  
A complex but powerful diagram which can illustrate 2-way flows (and net scores) in a matrix.

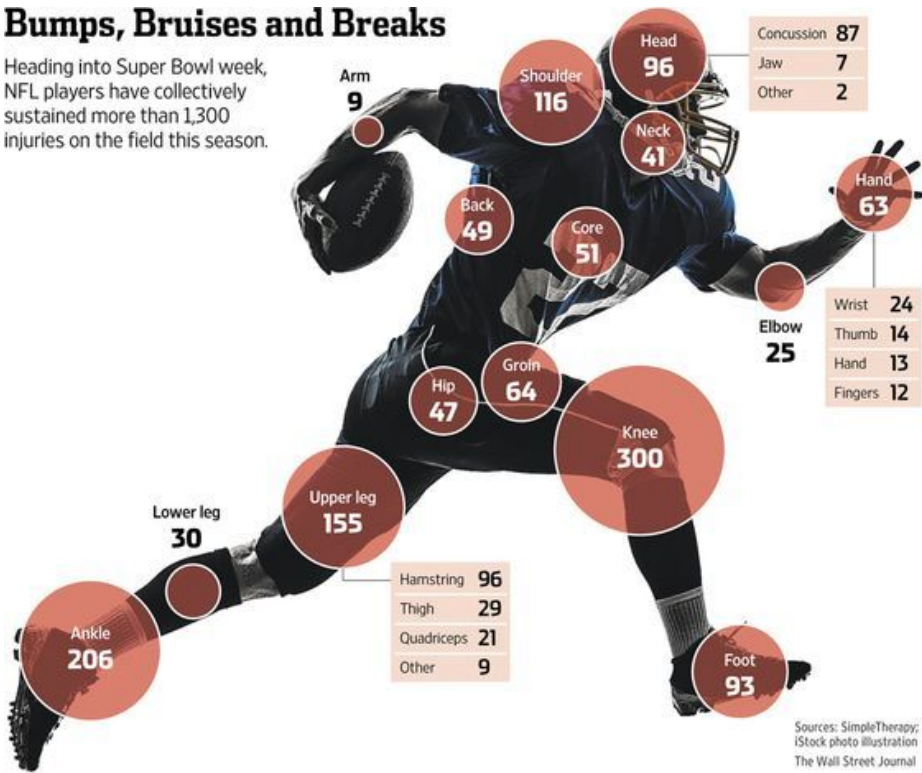
**Network**  
Used for showing the strength and inter-connections of relationships of varying types.

# Image and Diagram

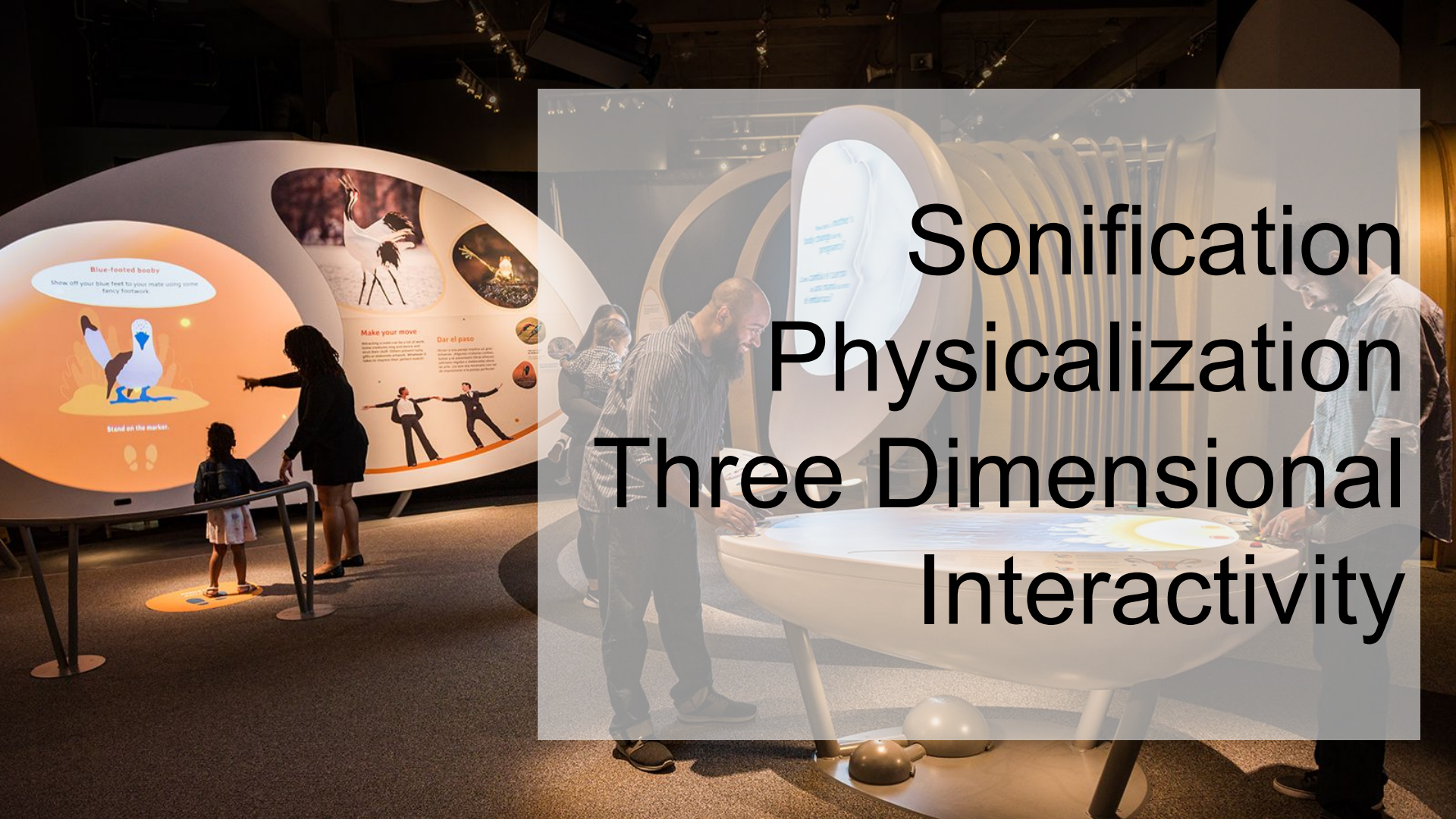


## Bumps, Bruises and Breaks

Heading into Super Bowl week, NFL players have collectively sustained more than 1,300 injuries on the field this season.



Sources: SimpleTherapy;  
iStock photo illustration  
The Wall Street Journal

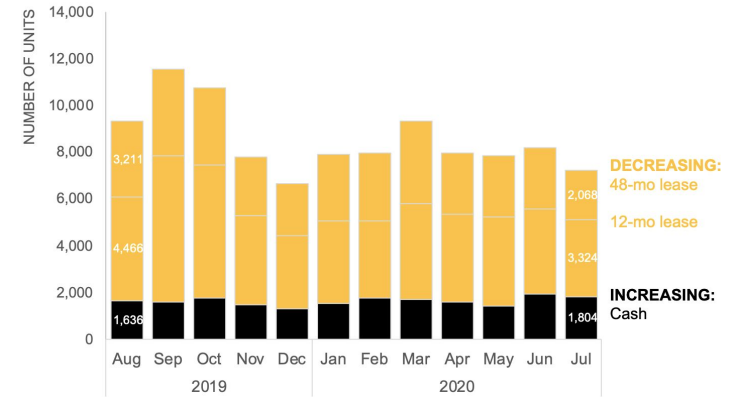


# Sonification Physicalization Three Dimensional Interactivity

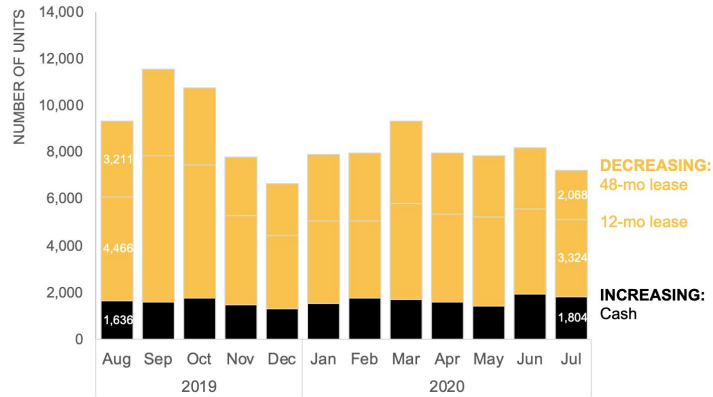
# Best Practices for Legibility

## Leasing falls while cash sales rise

Sales over time by purchase type



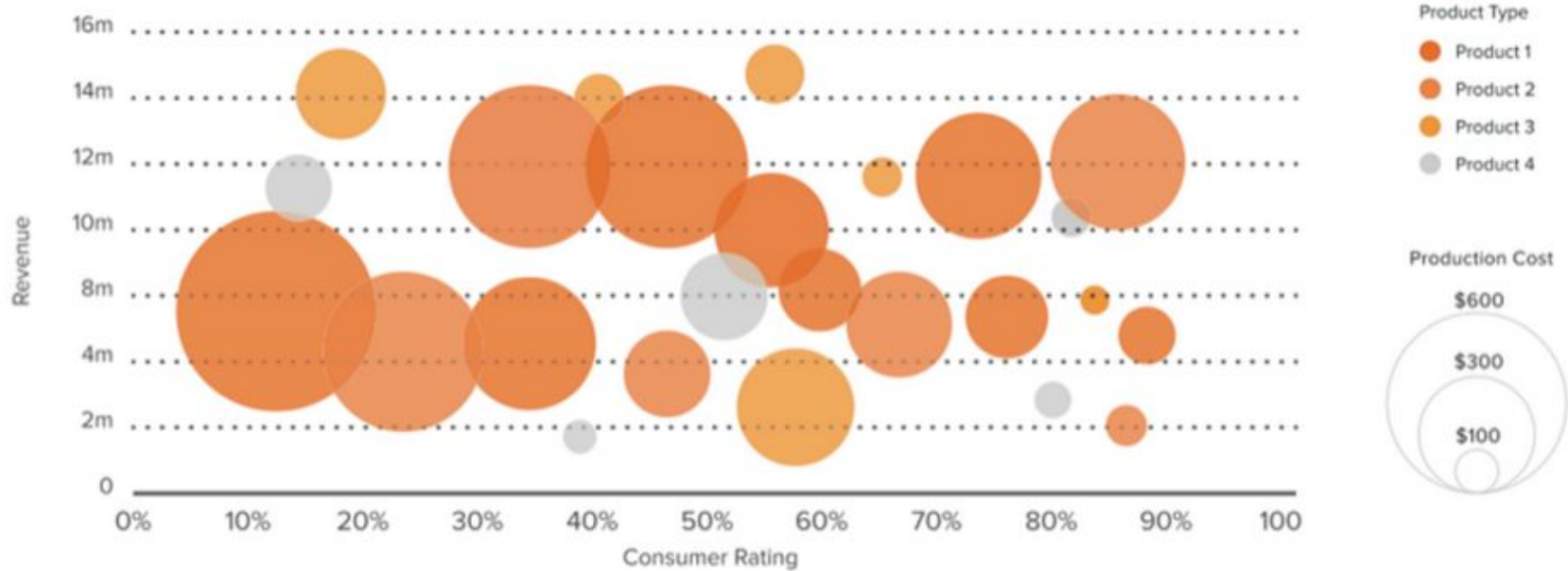
Sales over time by purchase type



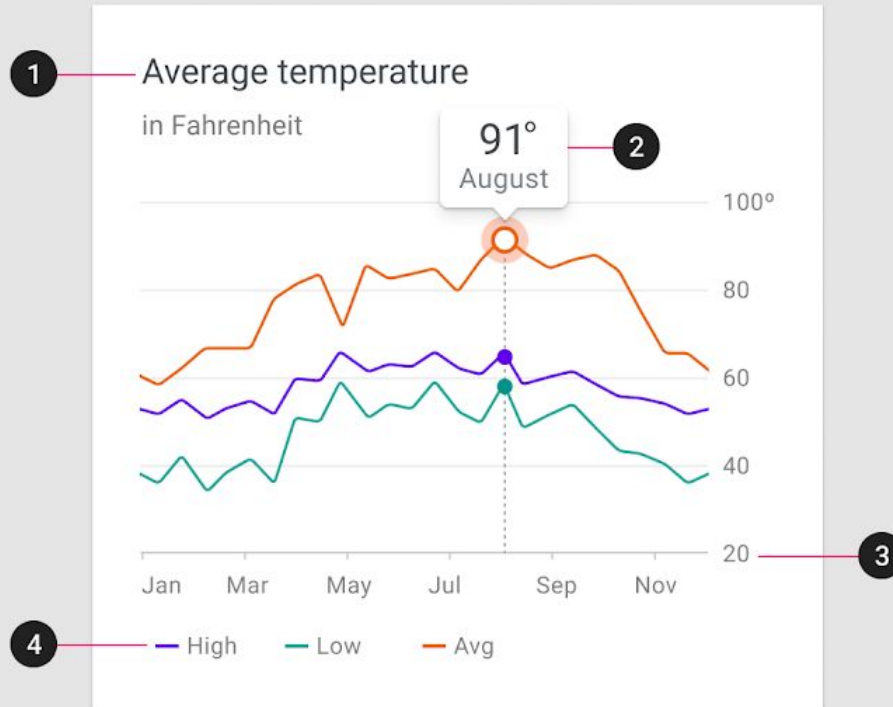
Titles should contain the insight not the description



## REVENUE VS. RATING

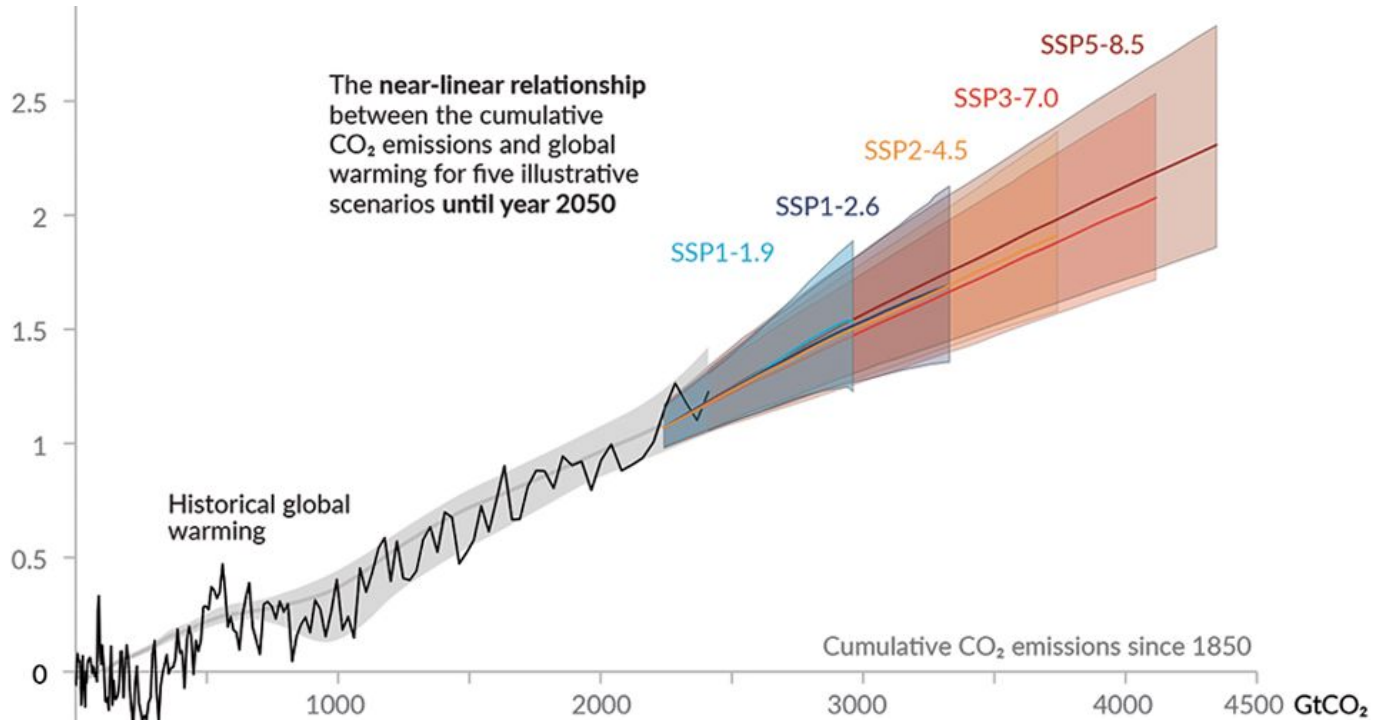


Legends associate visual aesthetics with simple labels.

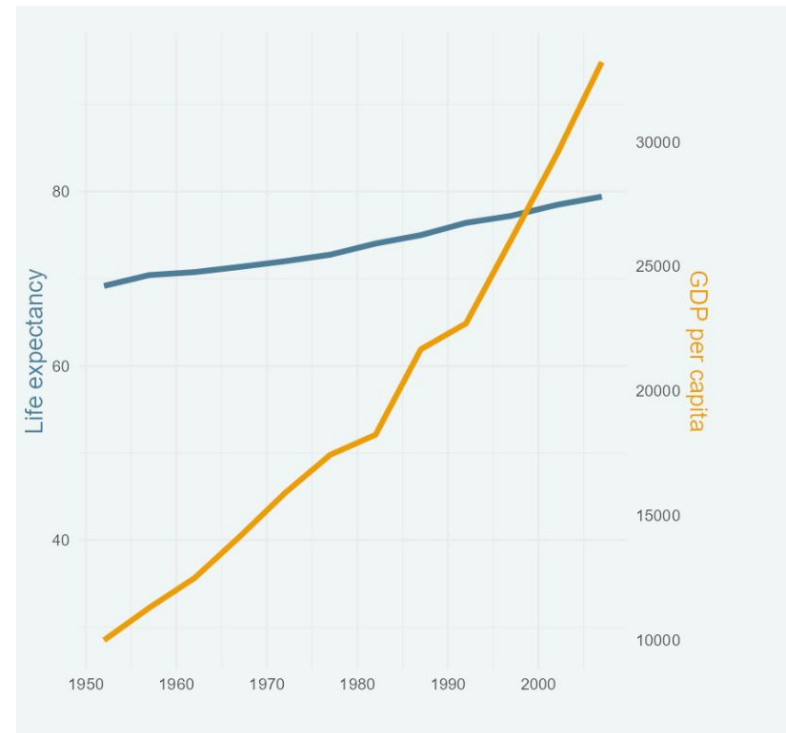
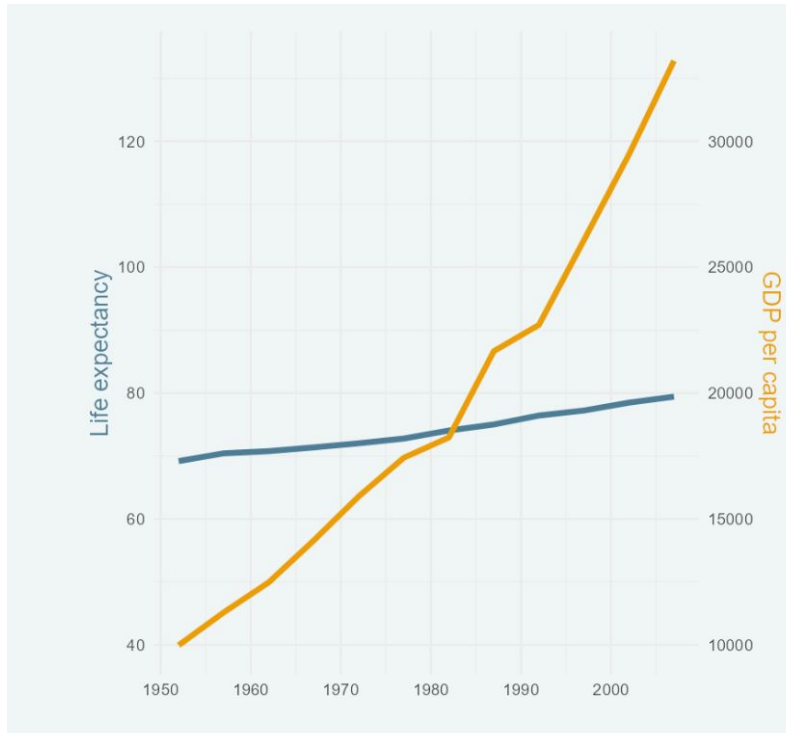


Proofread legends, and try to minimize eye jumps





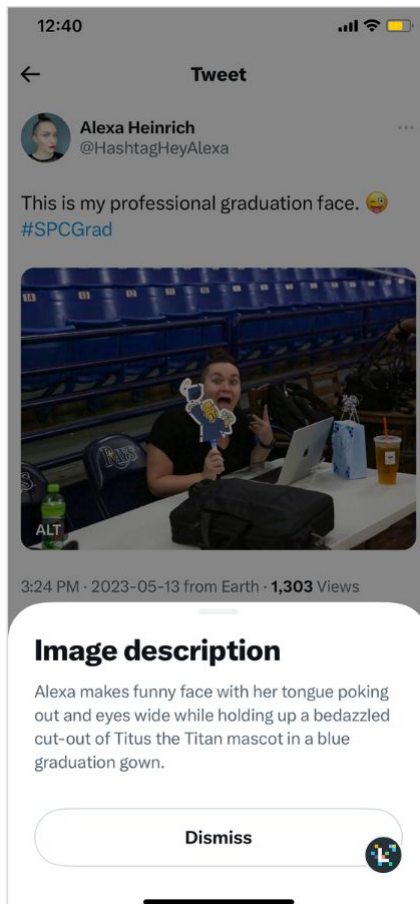
Axes labels describe the data and its units, on chart labels keep eyes close.



Choice of axes, scales, and proportions manipulates the story.

# Accessibility

- Use Alt-Text for images online.
- Choose color-blind friendly color scales.
- Use universal file types that anyone can open such as .png or .jpg.



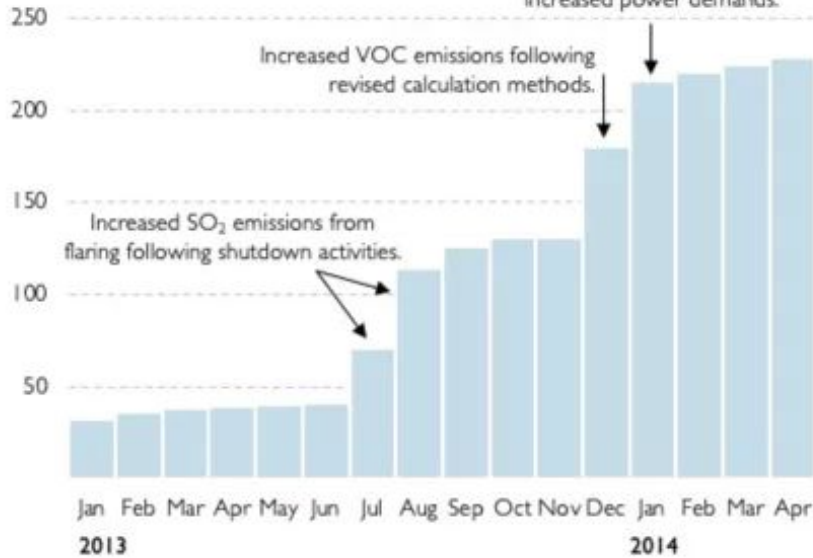
For divergent schemes try red to blue or purple to green:



# Data-ink ratio

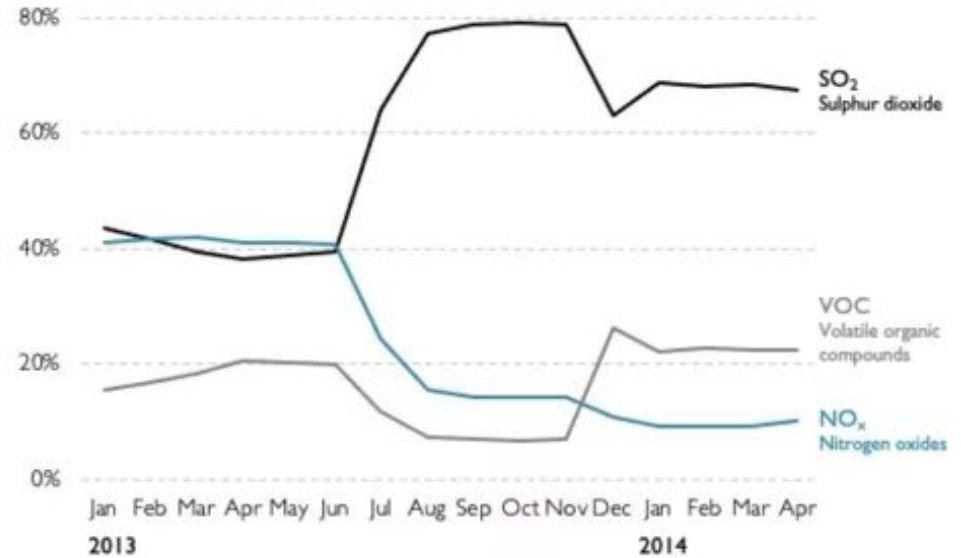
## Emissions of Green House Gases

'000 Tonnes of SO<sub>2</sub>/NO<sub>x</sub>/VOC

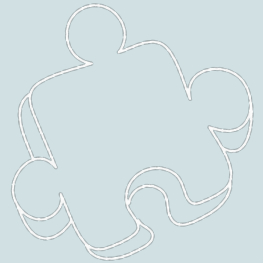


## Emissions of Green House Gases

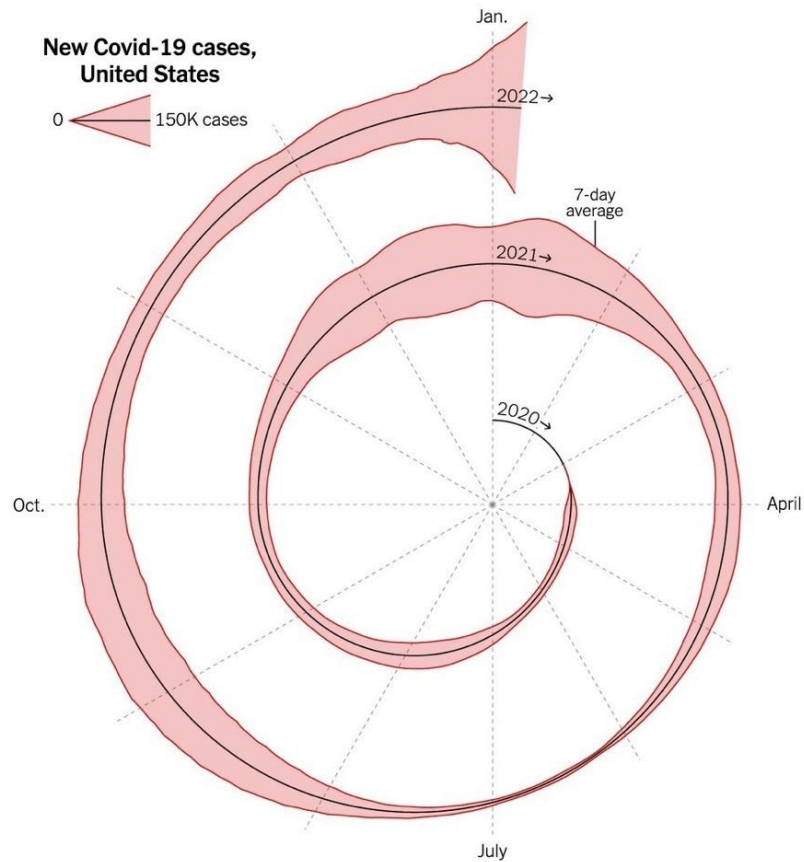
Share of total, by gas type



### 3. Applying the Principles of Design



Jan. 6, 2022



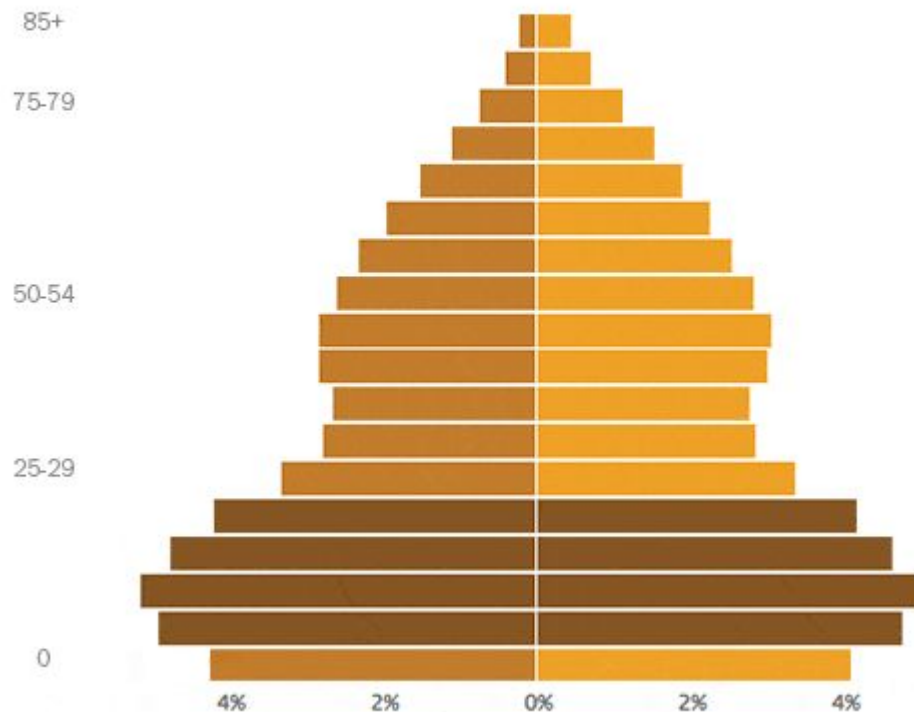
*Percent of U.S. Population by Age Group, 1950-2060*

■ Baby Boomers

MALE

1970

FEMALE





# Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui ont été en Russie, le noir ceux qui en sont sortis. Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Thiers, de Séguir, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davoust, qui avaient été détachés sur Minsk et Mohilow et qui rejoignent Orscha et Wilna, avaient toujours marché avec l'armée.

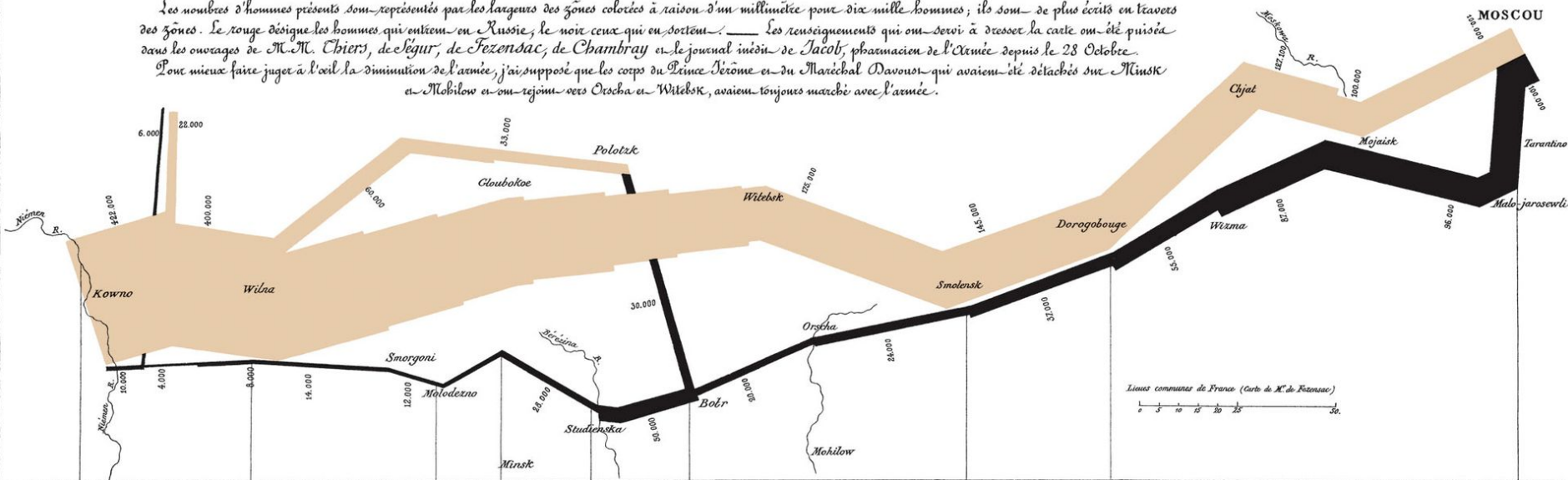
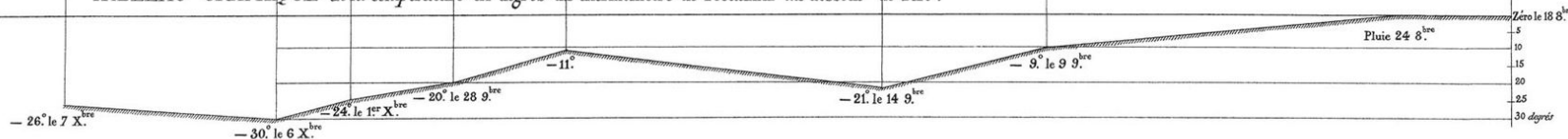


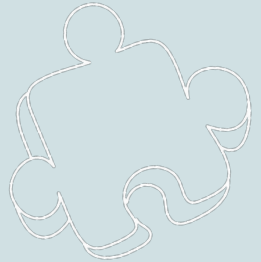
TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les Cosaques passent au galop le Niémen, gelé.



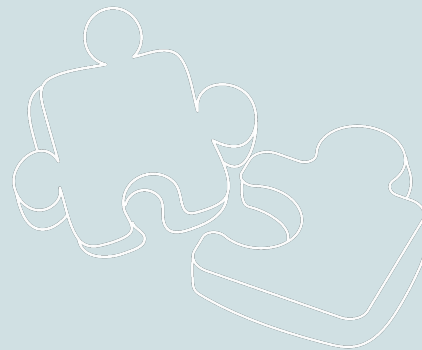
# Buffalization Data Visualization Challenge

<https://buff.link/2>



# Thank you!

Comments? Questions?



Reach out to us with questions:

[crdds@colorado.edu](mailto:crdds@colorado.edu)

Melissa Cantrell & Ellery Galvin



Center for Research Data & Digital Scholarship

UNIVERSITY OF COLORADO **BOULDER**