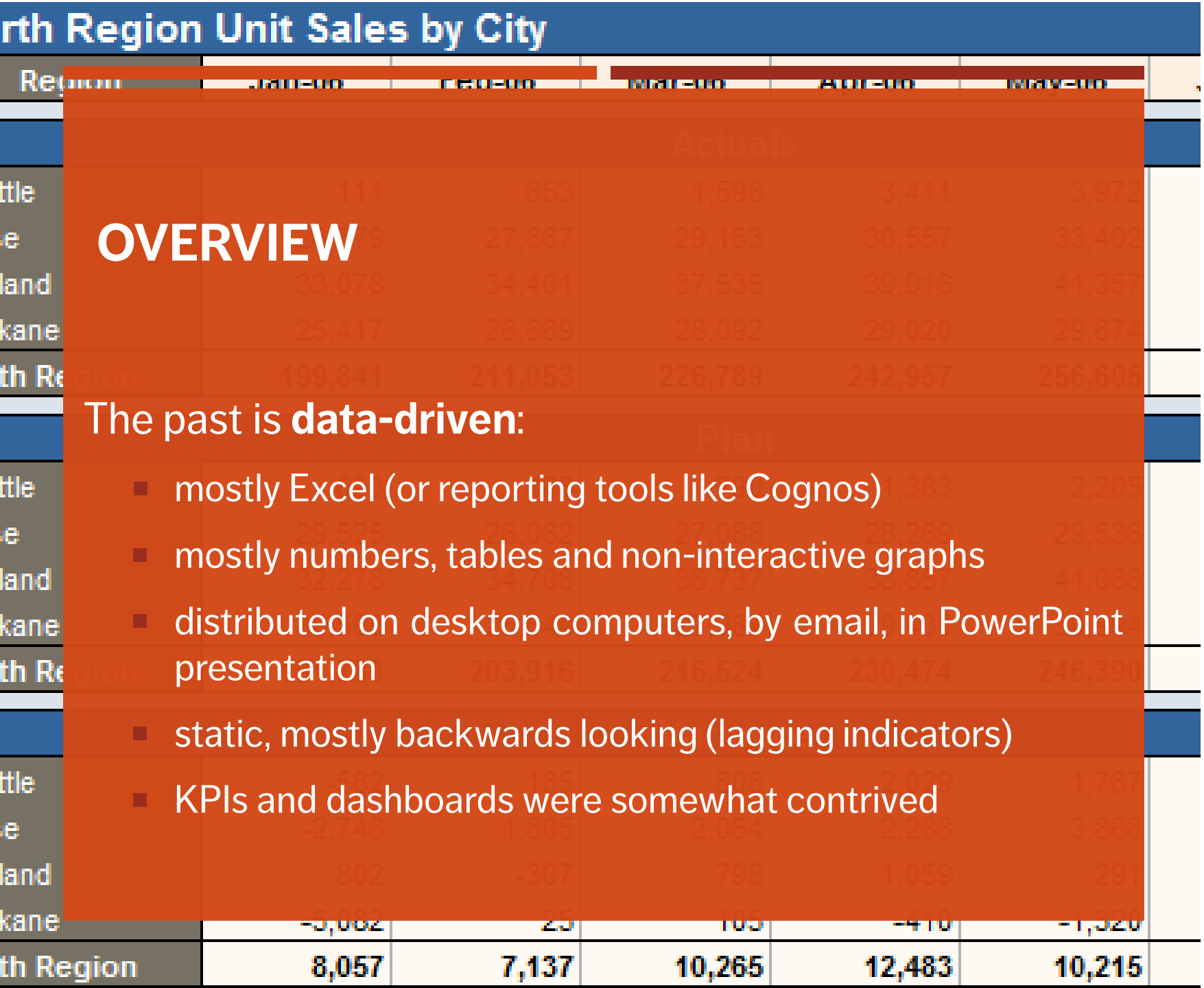
MODULE 3: DATA ANALYSIS AND VISUAL STORYTELLING

CT ACADEMY | DATA ACTION LAB

9. STORYTELLING AND VISUALIZATION

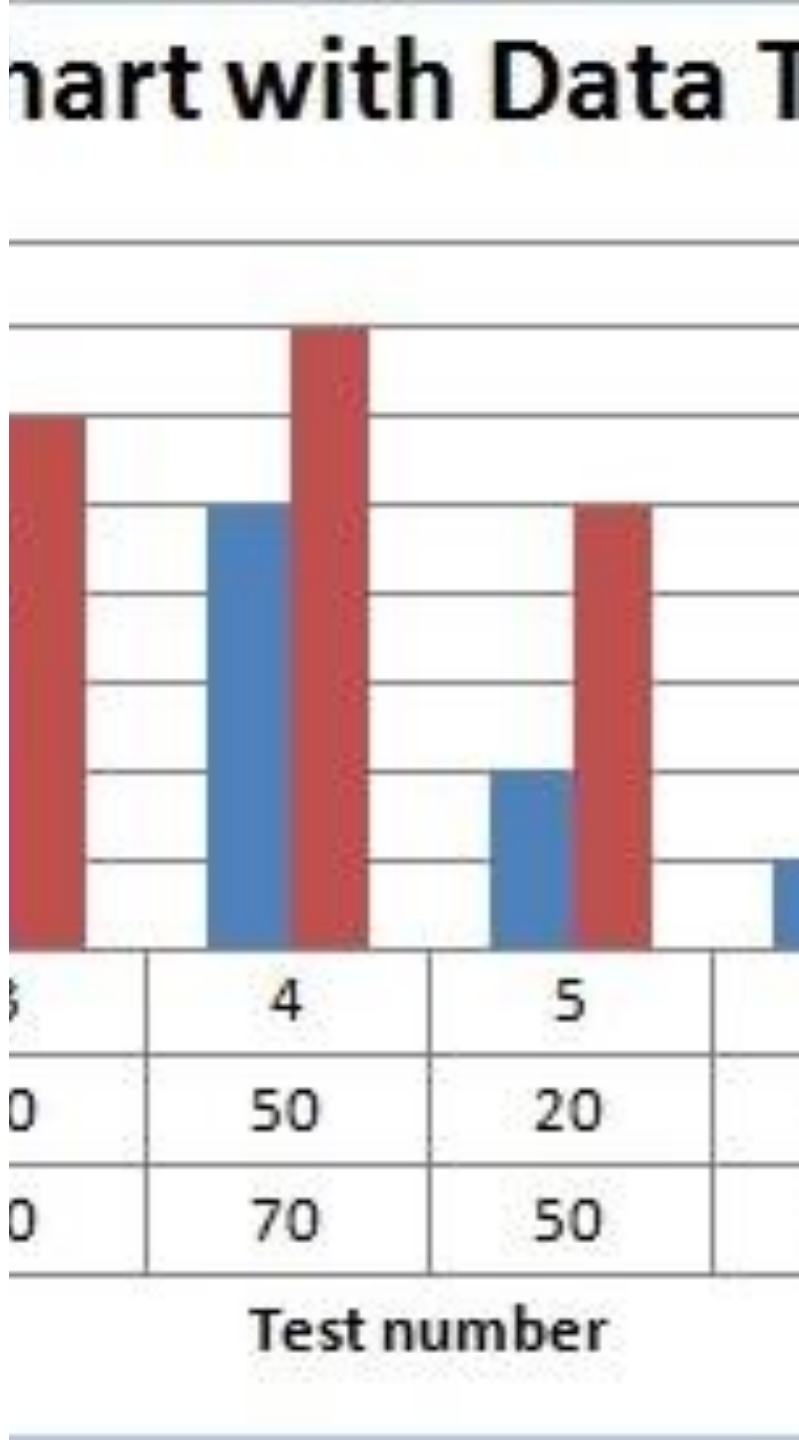
DATA ANALYSIS AND VISUAL STORYTELLING



OVERVIEW

The past is **data-driven**:

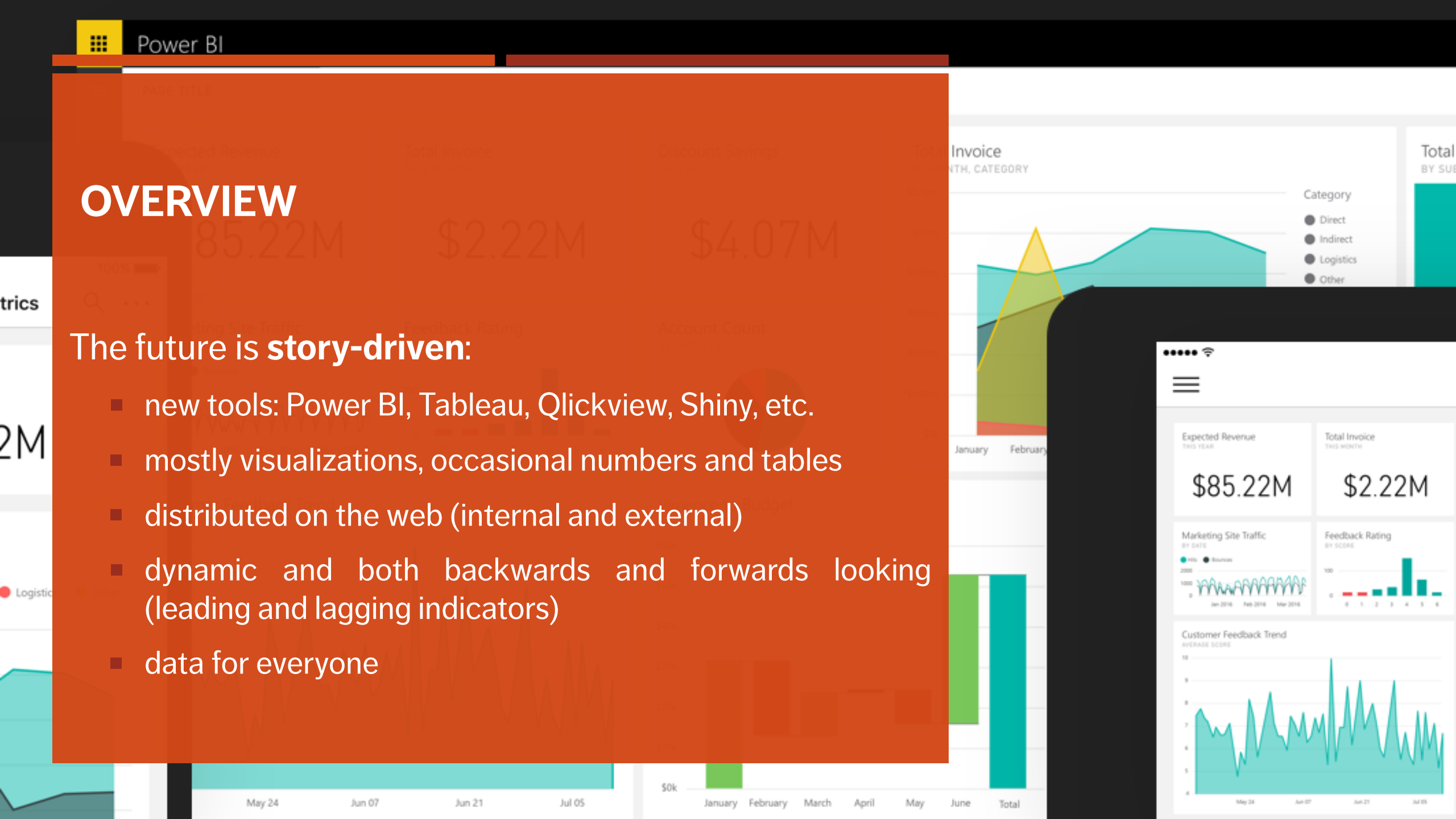
- mostly Excel (or reporting tools like Cognos)
- mostly numbers, tables and non-interactive graphs
- distributed on desktop computers, by email, in PowerPoint presentation
- static, mostly backwards looking (lagging indicators)
- KPIs and dashboards were somewhat contrived



OVERVIEW

The future is **story-driven**:

- new tools: Power BI, Tableau, Qlickview, Shiny, etc.
- mostly visualizations, occasional numbers and tables
- distributed on the web (internal and external)
- dynamic and both backwards and forwards looking (leading and lagging indicators)
- data for everyone

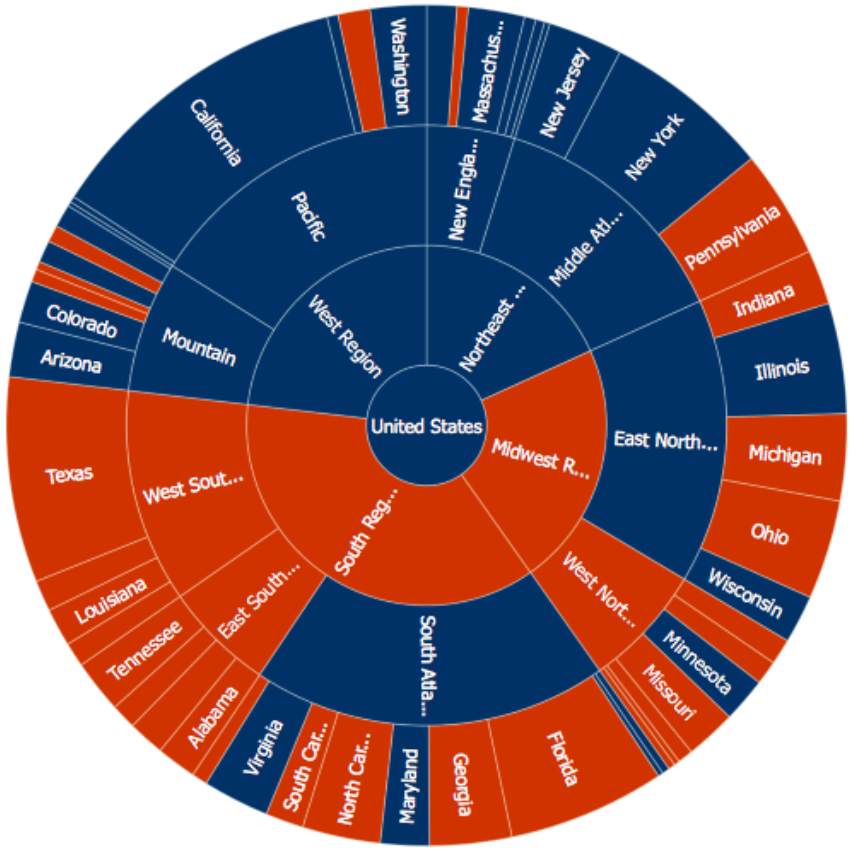


[Author unknown]

DATA VISUALIZATION VS. INFOGRAPHICS

Data Visualization

- A **method**, as well as an item (**objective**)
- Typically focuses on the **quantifiable**
- Used to make sense of the data or to make it **accessible** (datasets can be massive and unwieldy)
- May be generated **automatically**
- The look and feel are less important than the **insights conveyed by the data**



data-action-lab.com

[J.P. Blackard]

DATA VISUALIZATION VS. INFOGRAPHICS

Infographics

- Created for **story-telling** purposes (**subjective**)
- Intended for a **specific** audience
- **Self-contained** and discrete
- **Graphic design** aspect is key
- **Cannot** usually be re-used with other data
- Can incorporate **unquantifiable** information

data-action-lab.com

HISTORICAL CHARTS

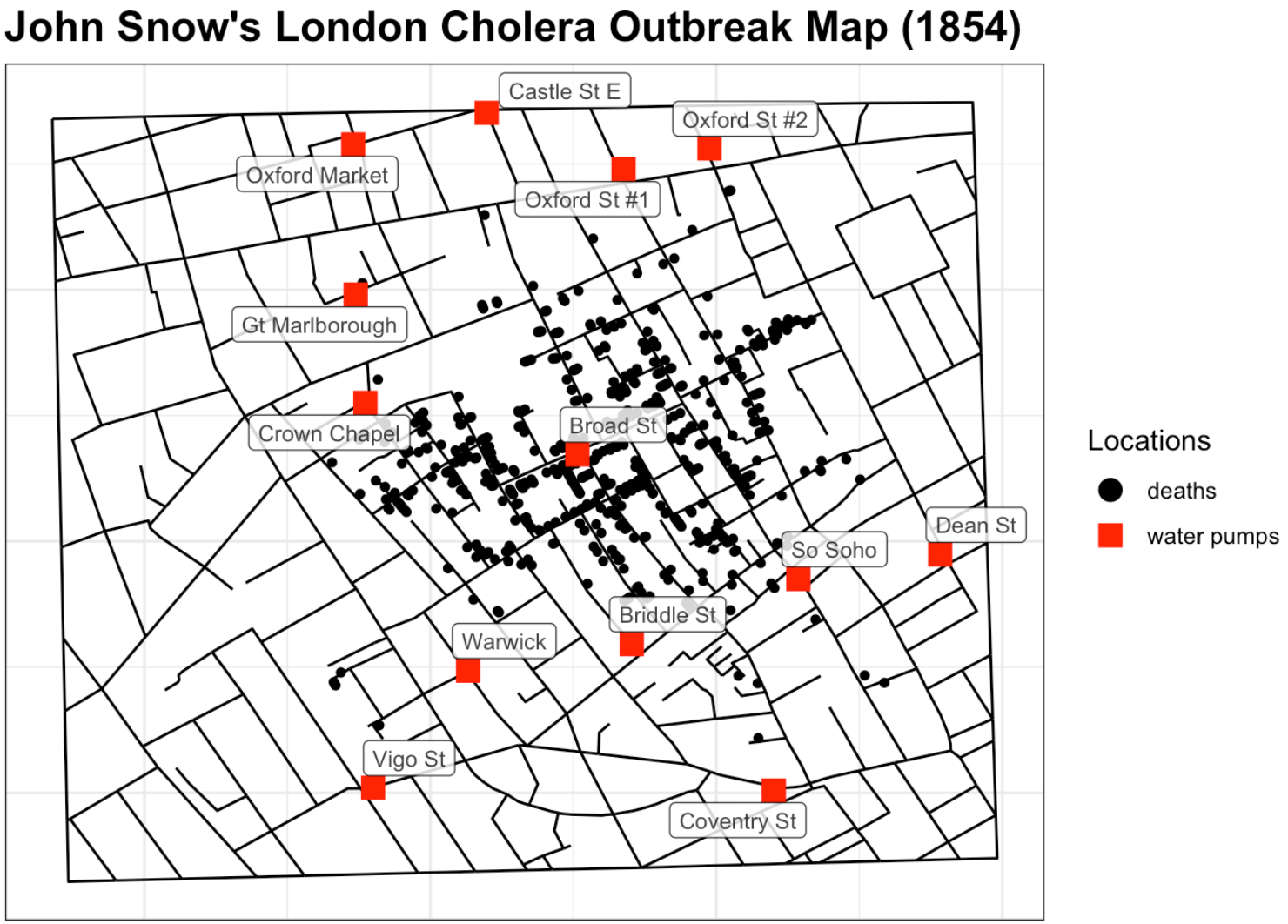
Data visualization is not confined to the recent past: charts have been used for many years to help **communicate information** and **tell stories**.

Due to the absence of technical tools, a lot of thought had to go into the design and creation of these visualizations.

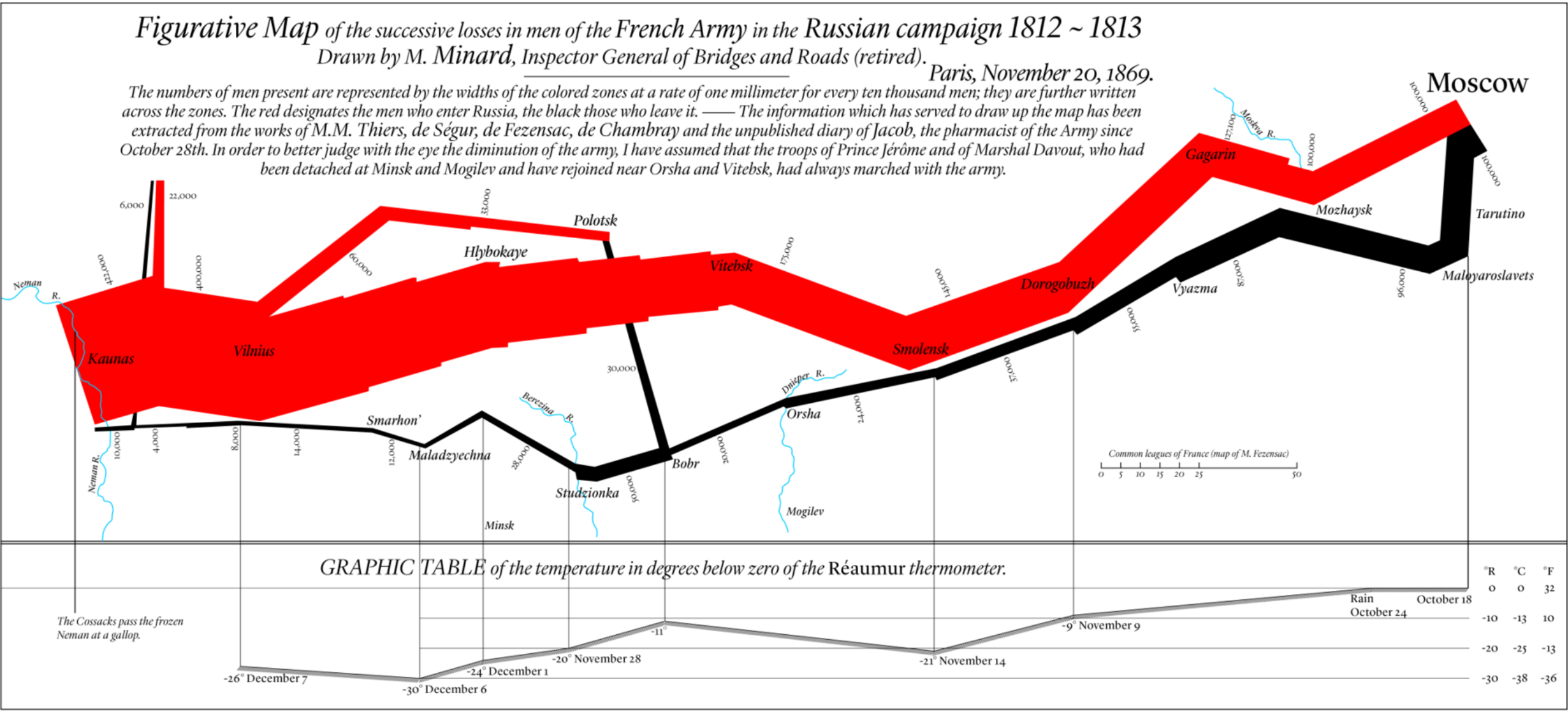
Consequently, there is a lot we can (and **should**) learn to bring into the development of charts from a **design and storytelling perspective**.

London's Cholera Outbreak of 1854

Physician John Snow links the outbreak to a contaminated well by plotting number of cases on a map, jump-starting the science of epidemiology.



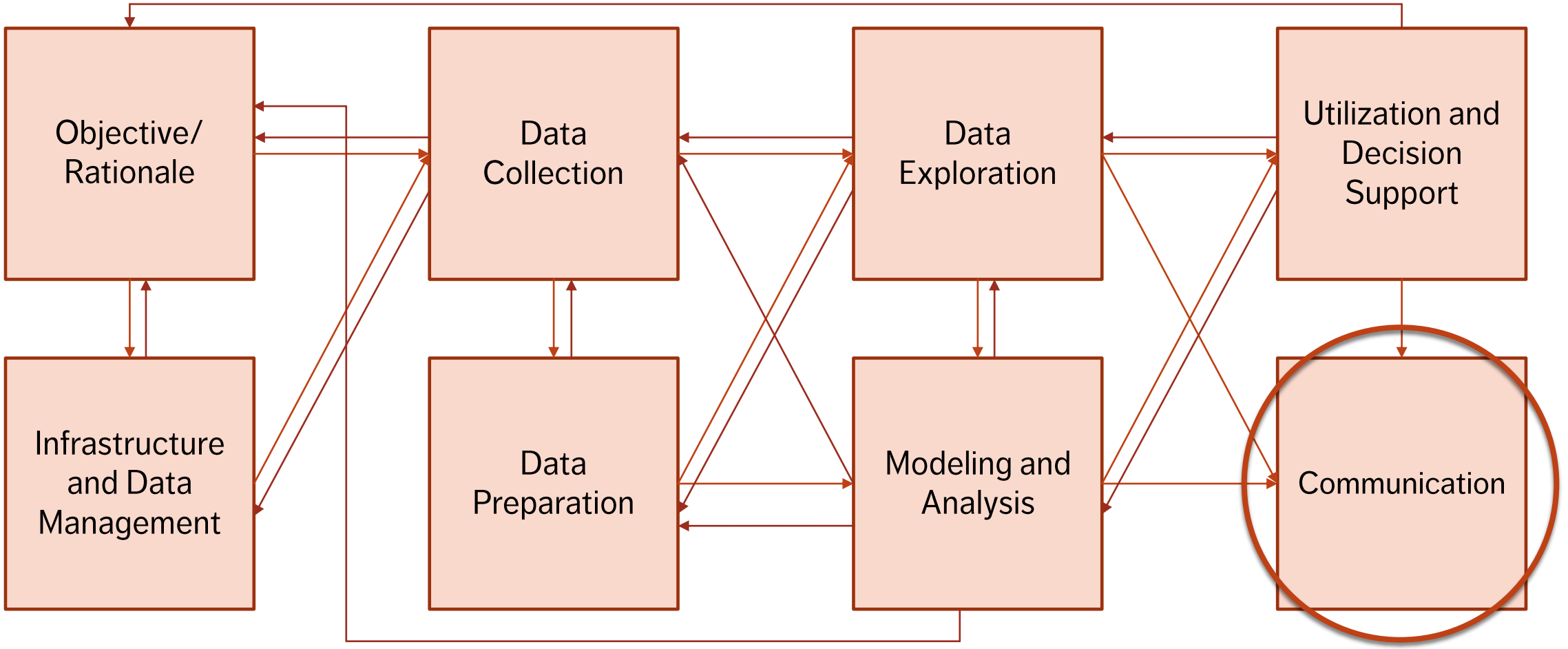
[DAL Podcast #3: https://youtu.be/tS_2_IQ5gRw]



Minard's March to Moscow

data-action-lab.com

THE (MESSY) ANALYSIS PROCESS



PRACTICAL DEFINITION OF A STORY

To paraphrase U.S. judge Potter Stewart: “I may not be able to define what a story is, but I know one when I see one”.

We could say that a **story** consists of:

- context,
- series of events, and
- outcome, result, consequence, or resolution.

STORYTELLING GOALS

Cultural Stories

- entertain, inform, teach, explore, shock

Data (Scientific) Stories

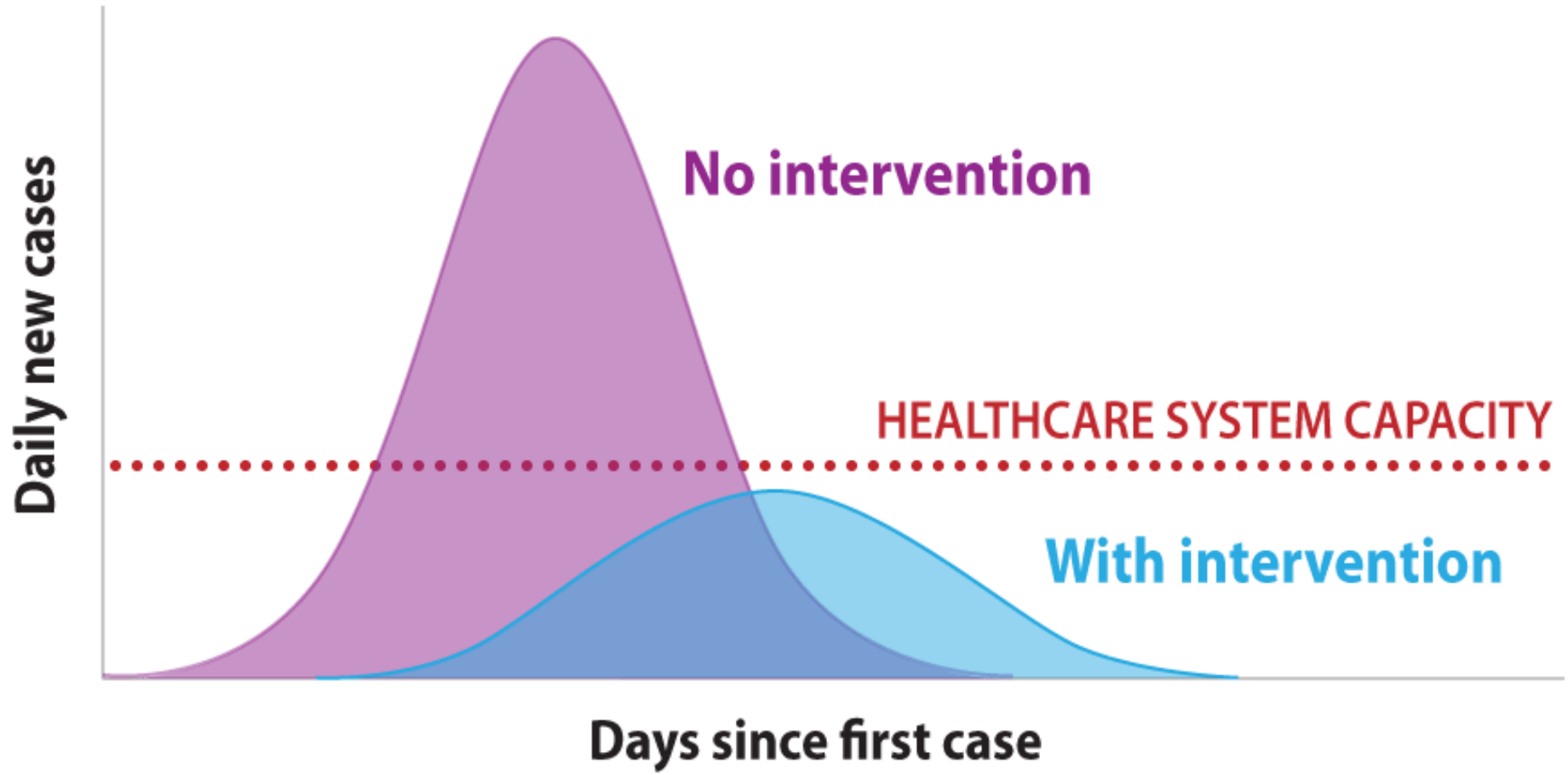
- describe, diagnose, predict, prescribe, persuade

Any overlap?

Anything missing?

FLATTENING THE CURVE

A look at the importance of slowing the spread of a virus, so that the rate of infection doesn't outpace the resources to fight against it.



SOURCE: CDC

THE CANADIAN PRESS

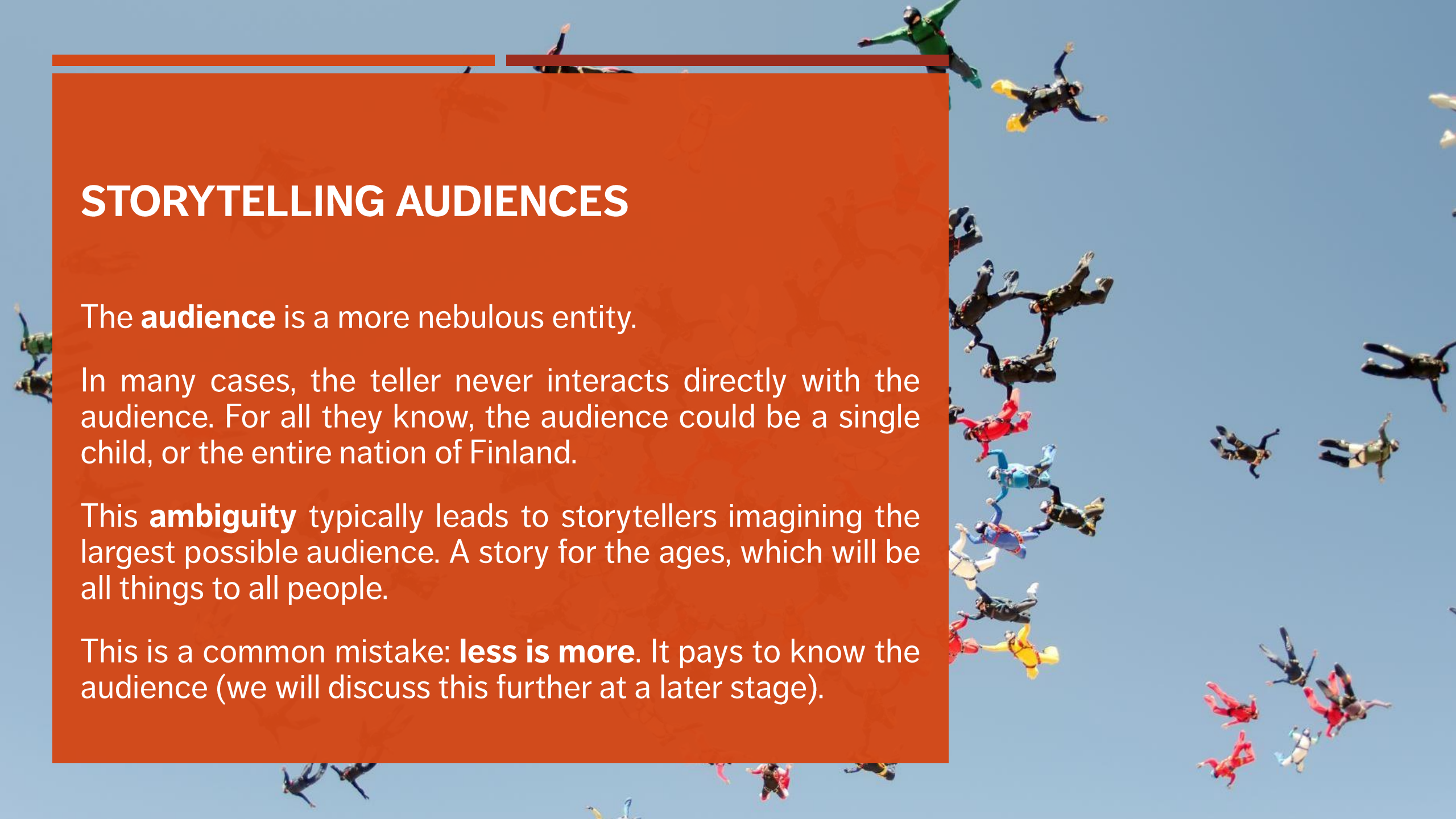
STORYTELLING AUDIENCES

Storytelling requires a **teller** and a **story**, but also an **audience.**

The **teller**'s job is to convince the audience to accept:

1. the premise ("I'm about to tell you a really interesting story, so listen up!")
2. the contents ("All these things happened, honest!")
3. the conclusion ("And that's why you should never put peanut butter in your laundry.")

The **story**'s must first and foremost not come in the way of the teller's job.



STORYTELLING AUDIENCES

The **audience** is a more nebulous entity.

In many cases, the teller never interacts directly with the audience. For all they know, the audience could be a single child, or the entire nation of Finland.

This **ambiguity** typically leads to storytellers imagining the largest possible audience. A story for the ages, which will be all things to all people.

This is a common mistake: **less is more**. It pays to know the audience (we will discuss this further at a later stage).

EXERCISE

Consider a data question of interest to you.

Identify the target audience and the goals for your storytelling dashboard.

STORYTELLING CONTEXT

A given action may be seen as positive or as negative by audiences with different pre-existing feelings/knowledge concerning the agent/situation.

- Would you be able to recognize nobility in a political enemy's actions?
- Could a fan of the Maple Leafs/Habs ever have something worthy to say about hockey?

Similarly, a story may have different **outcomes/impacts** in different contexts.

Trevor Greenway · Feb 9 · 5 min read

[https://www.lowdownonline.com/post/wakefield-nurse-fires-up-freedom-convoy]

Wakefield nurse fires up Freedom Convoy

Wakefield's Bethan Nodwell is known in the Gatineau Hills for many things: being the hospital's former head nurse, singing onstage at the Black Sheep Inn, and more recently, disseminating debatable facts and anti-vax sentiments on social media. Now she's running the main stage at the Freedom Convoy in downtown Ottawa, firing up the crowd as seen here Feb. 4. Trevor Greenway photo

Bethan Nodwell had thousands of demonstrators in Ottawa hanging onto her every word.

What might lead one to view the **subject** of this article in a positive light?

A negative light? A neutral light?

What might lead one to view the **author** of this article in a positive light?

A negative light? A neutral light?

[online.hbs.edu/blog/post/data-storytelling, C. Cote]

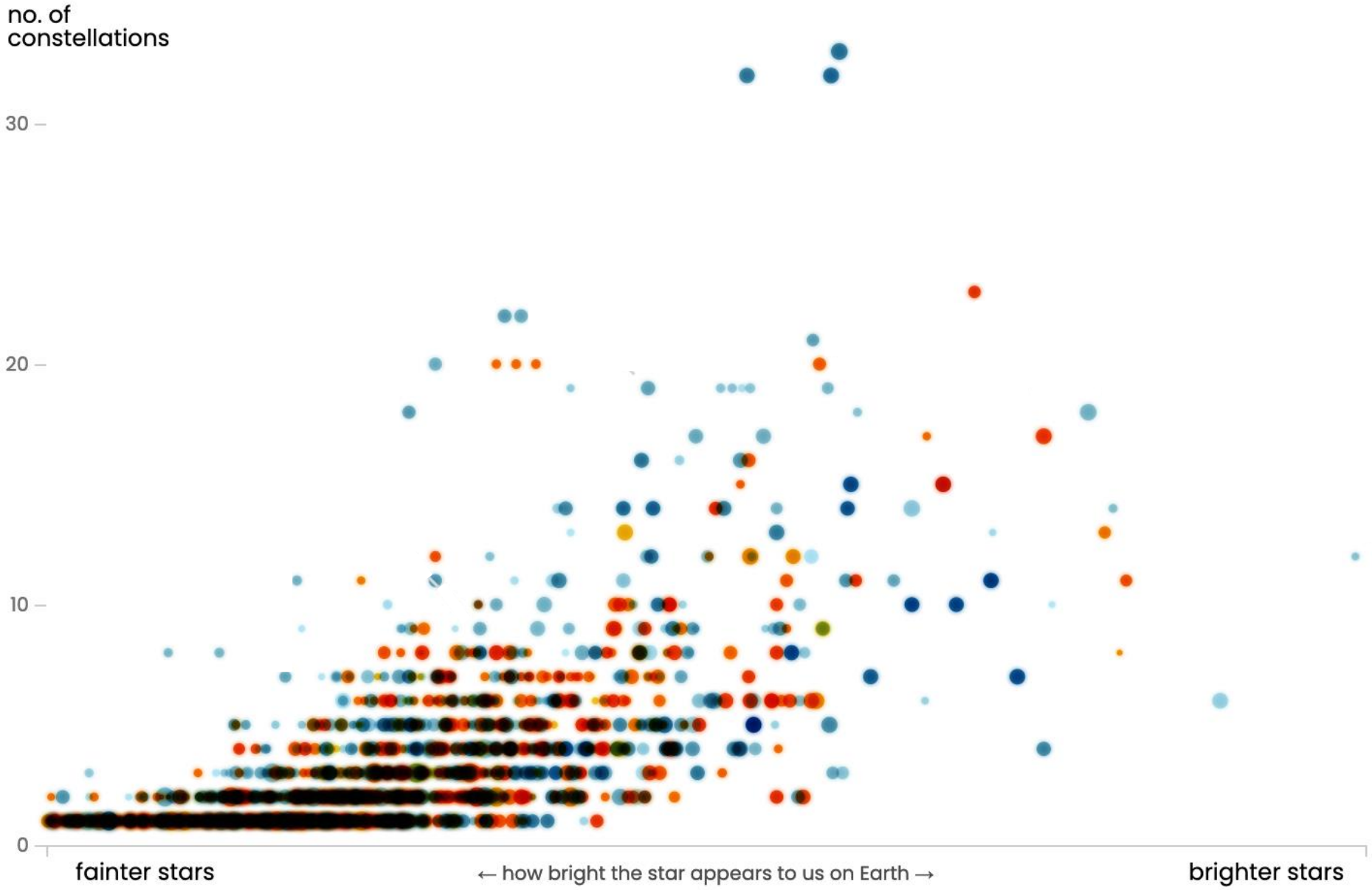
DATA STORIES

Data storytelling is the ability to effectively communicate insights from a dataset using narratives and visualizations. It can be used to put data insights into context for and inspire action from the audience.

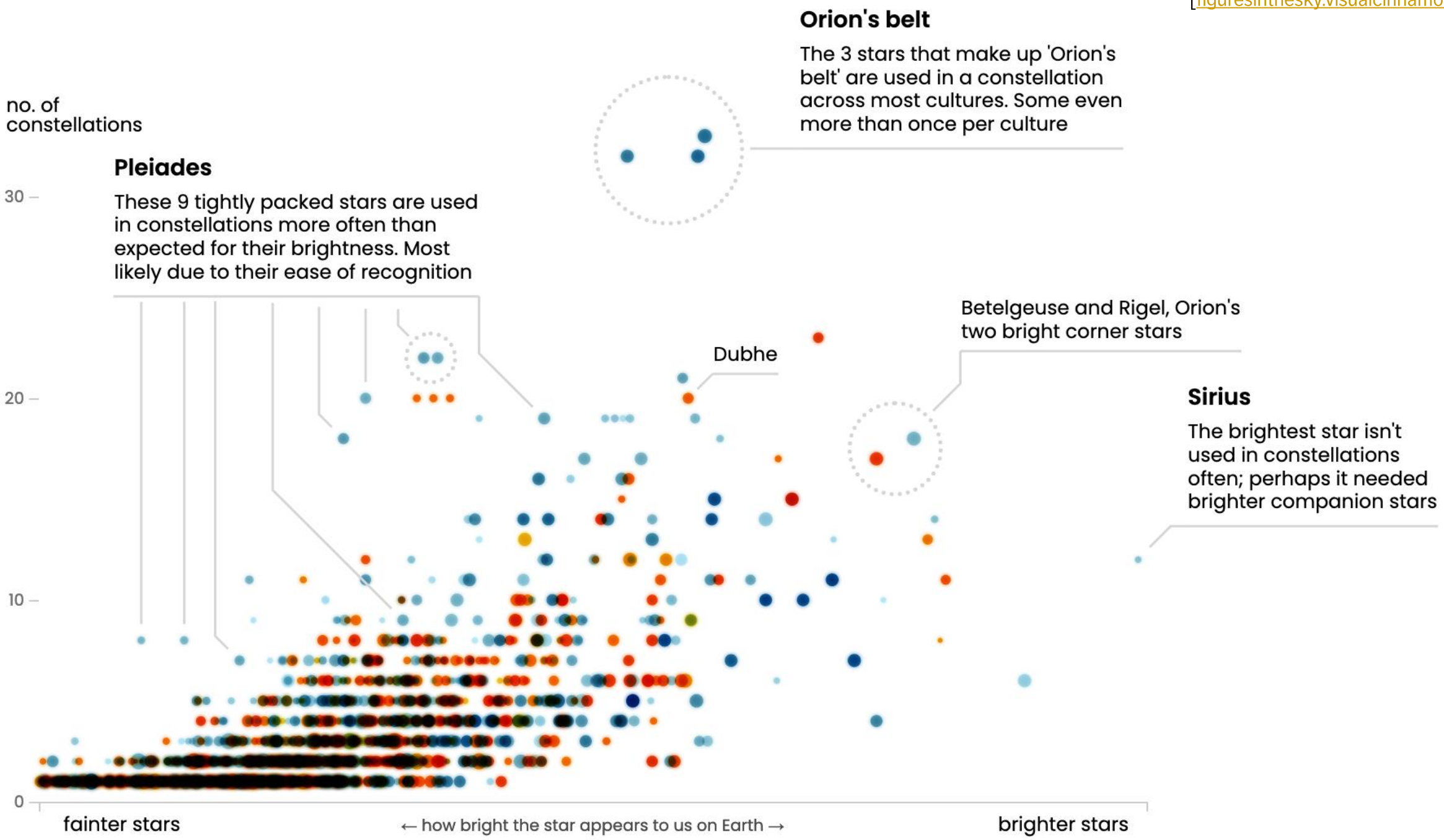
There are 3 key components:

1. **data:** foundation of data story (descriptive, diagnostic, predictive, prescriptive analysis)
2. **narrative:** storyline used to communicate the insights gleaned from data and context, and recommended actions
3. **visuals:** representations of data, analysis results, and narratives, which are used to communicate stories clearly and memorably (charts, graphs, diagrams, pictures, or videos)

[figuresinthesky.visualcinnamon.com]



[figuresinthesky.visualcinnamon.com]



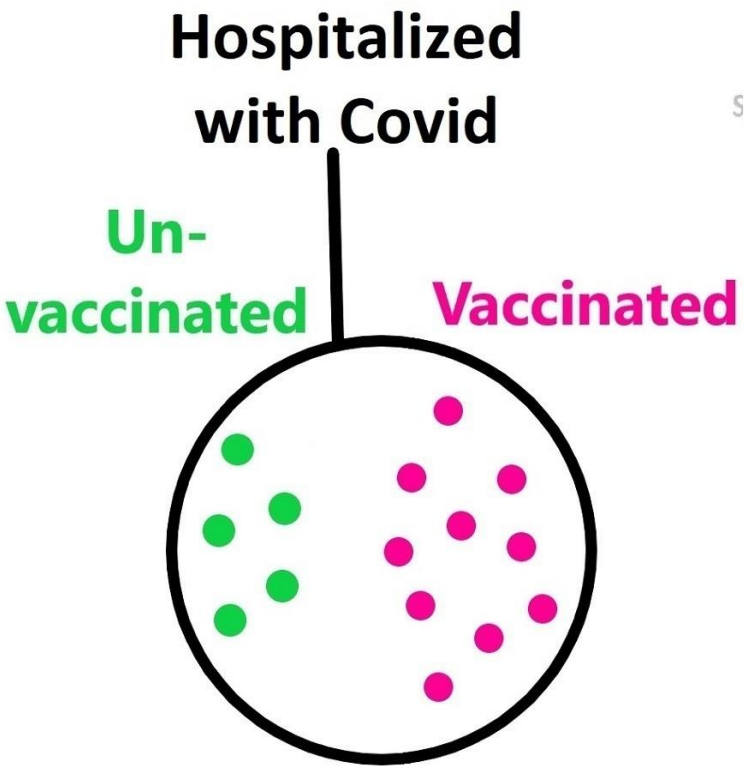
STORYTELLING RISKS

A good story can help shed insights on a situation, but storytelling requires **choices**; the outcome is affected by what is **included** and what is **omitted**.

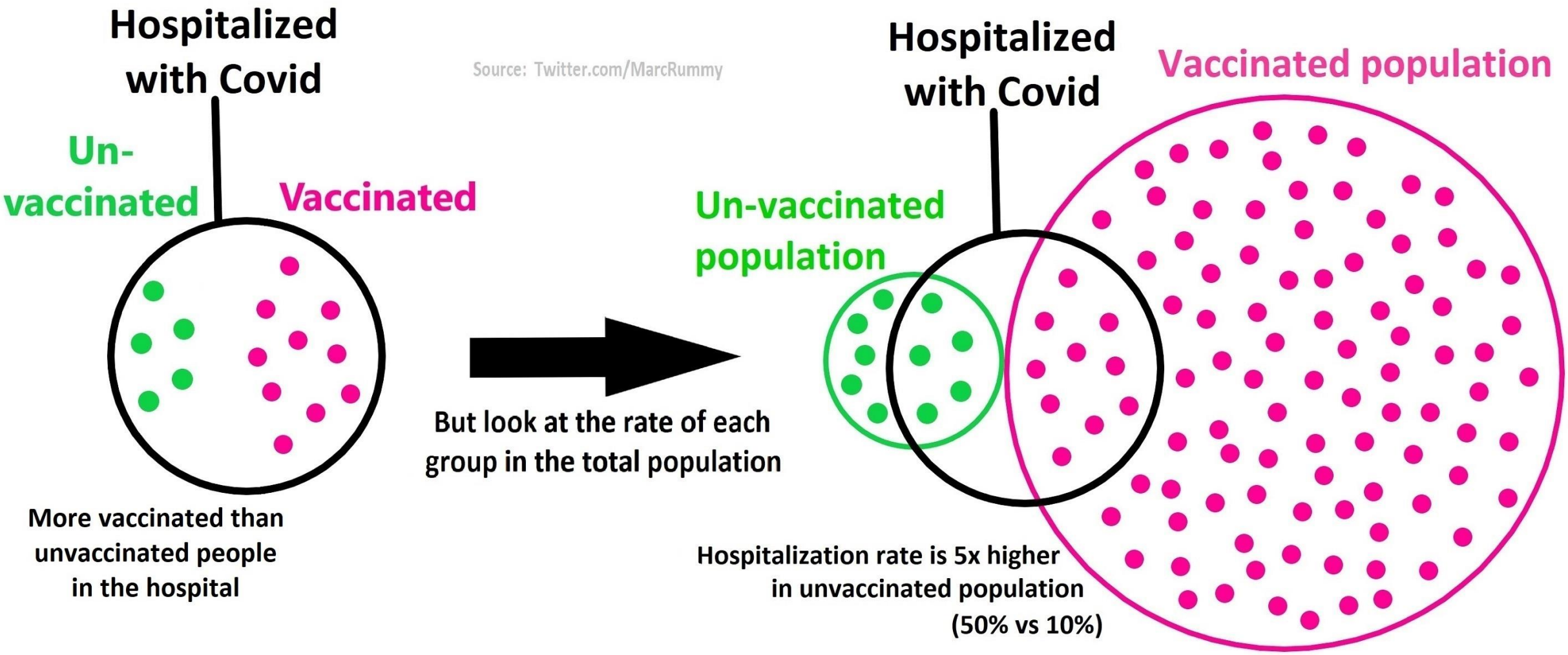
It is easy to mislead by **accident**; it is also easy to mislead by **design**.

With data stories, there is an additional complication: we usually only have access to the **available data**. The data that was not collected is, by definition, not available. Some of the data that was collected may also be unavailable for a variety of reasons.

This implicit bias can lead to compelling (yet **flawed**) data stories.



More vaccinated than unvaccinated people in the hospital



Note: The ratios presented are made to illustrate the concept of the base rate fallacy when the vaccination rate is high

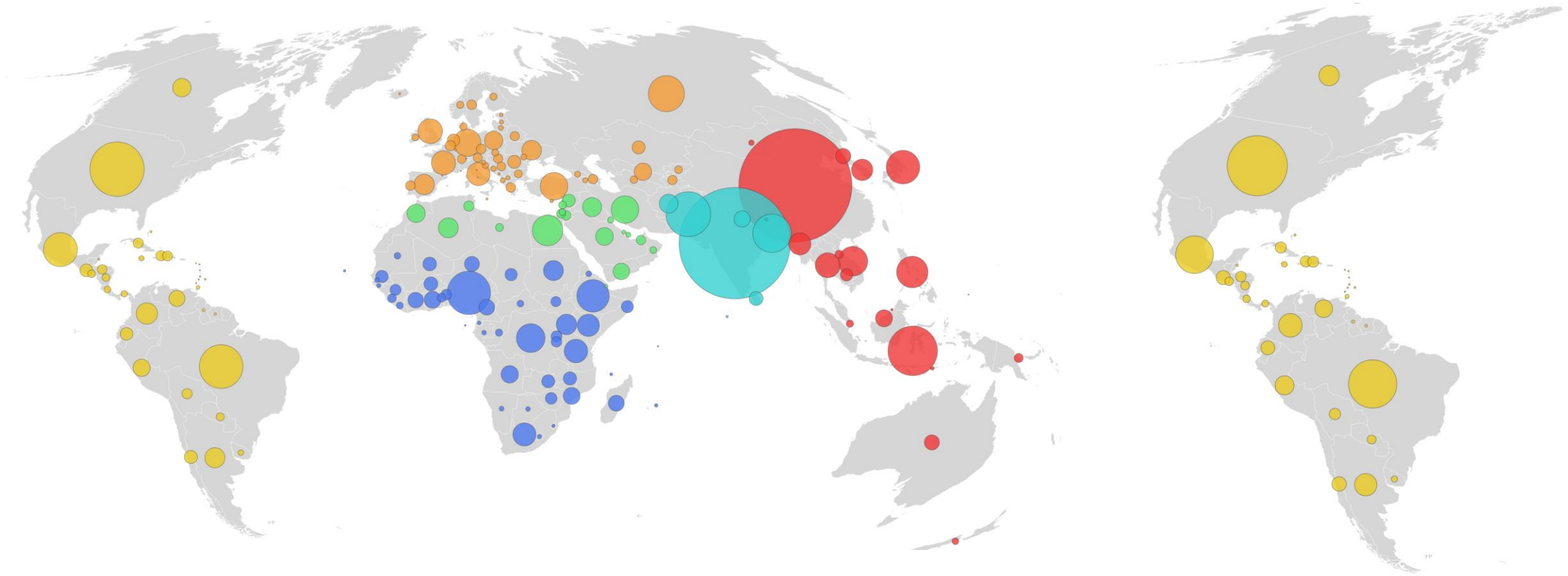
[based on an idea from S. McCloud, *Making Comics*]

VISUAL STORYTELLING CHOICES

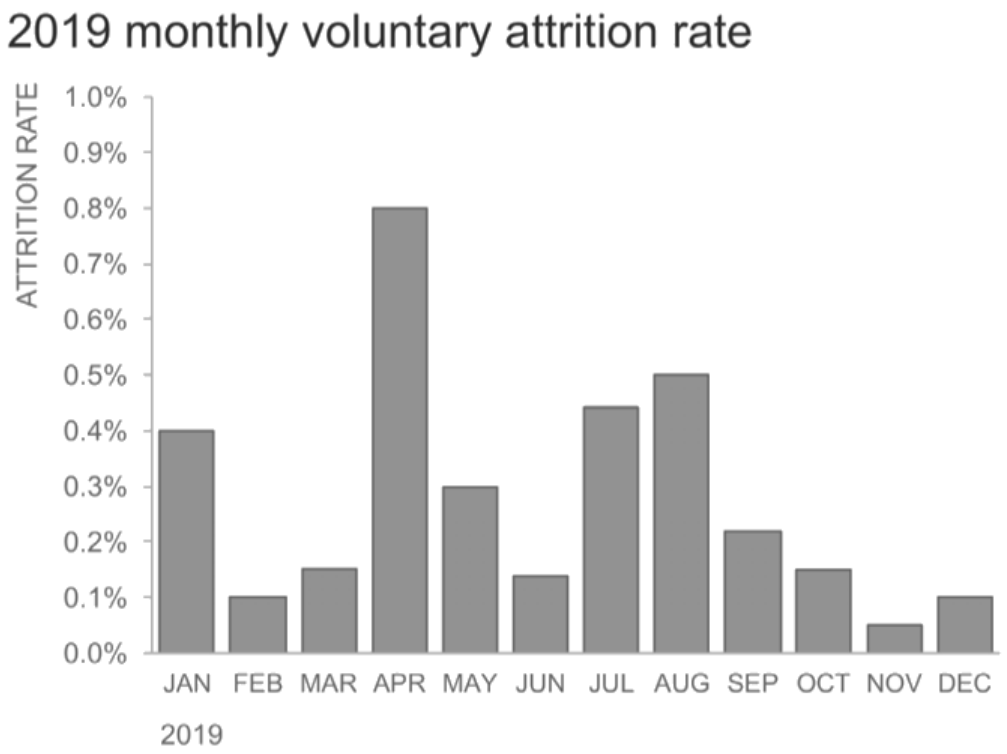
Communicating with **clarity** means that audience comprehension remains the **ultimate goal:**

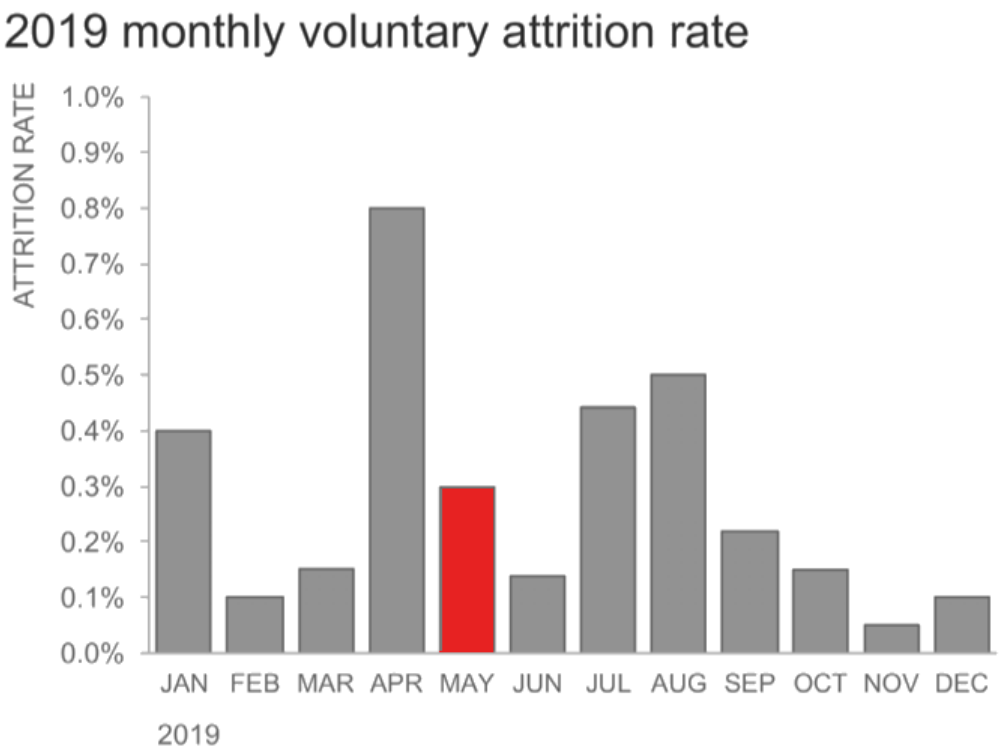
- choice of **moment** is 'connecting the dots', showing only what matters to the story;
- choice of **frame** is creating and directing the audience's focus;
- choice of **image** is selecting the right charts for the story, with emphasis on simplicity and ability to convey the message;
- choice of **word** is clearly and persuasively communicating ideas in seamless combination with the charts;
- choice of **flow** is guiding the audience from one chart to the next, from one page to the next, and creating a transparent and intuitive 'reading' experience, by arranging pages in a dashboard, charts on a page, and elements within charts intelligently.

CHOICE OF MOMENT

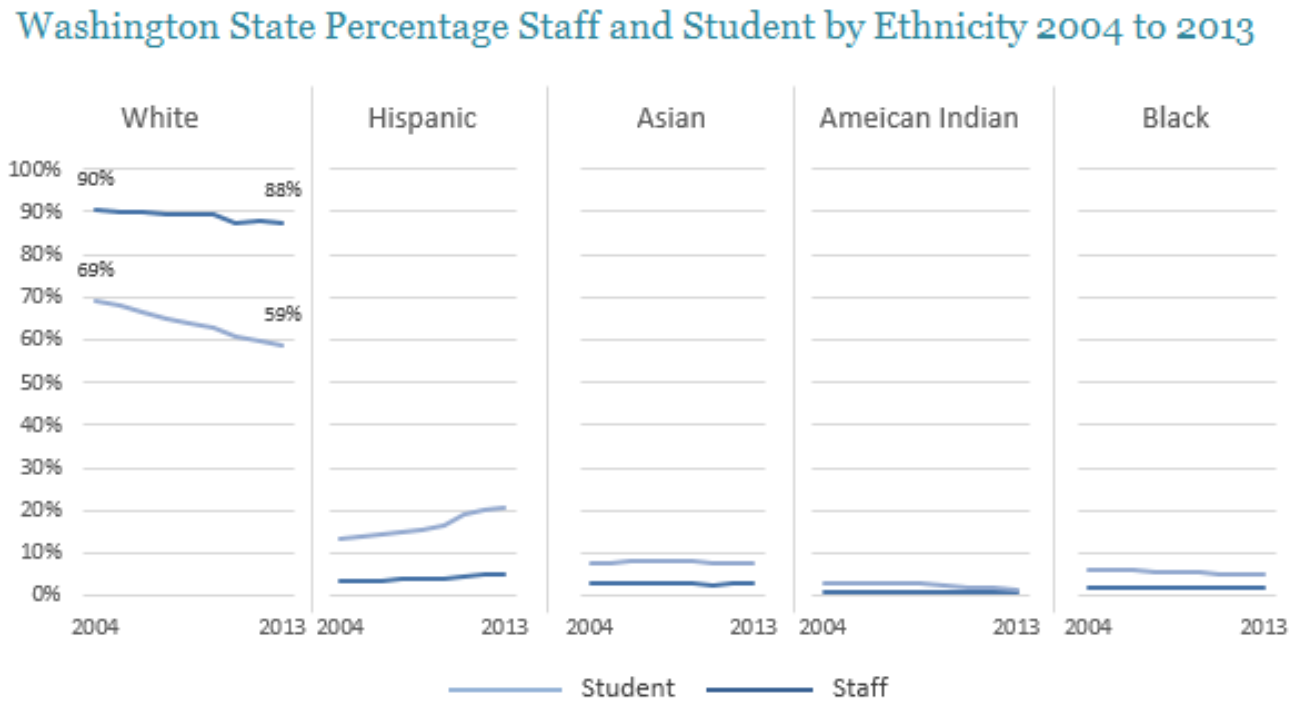


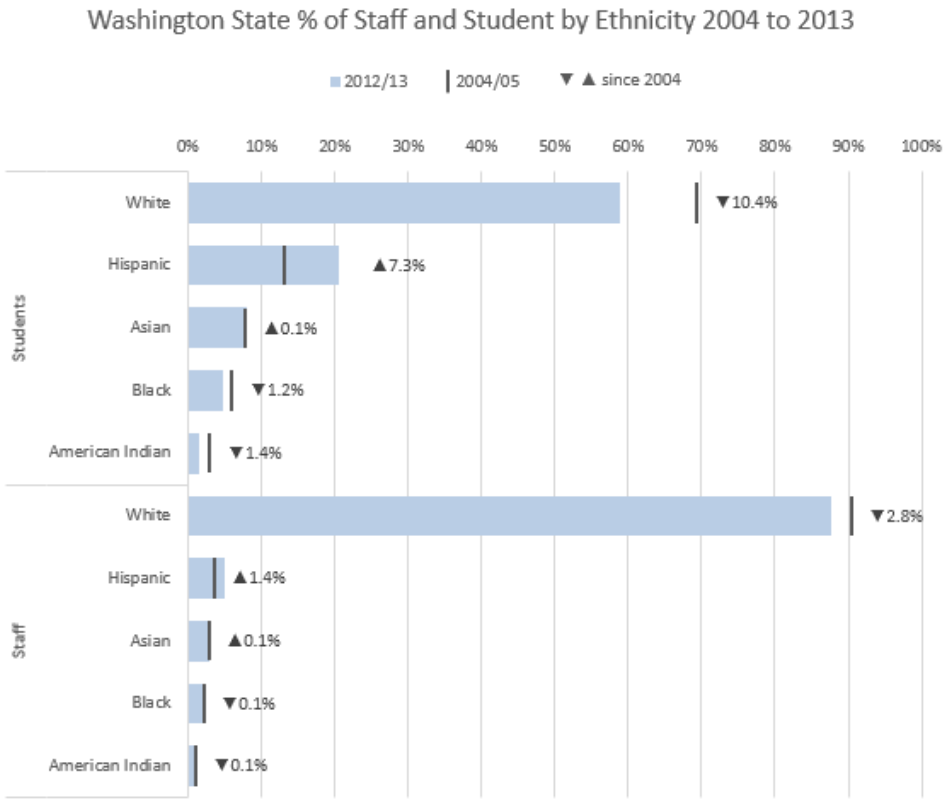
CHOICE OF FRAME



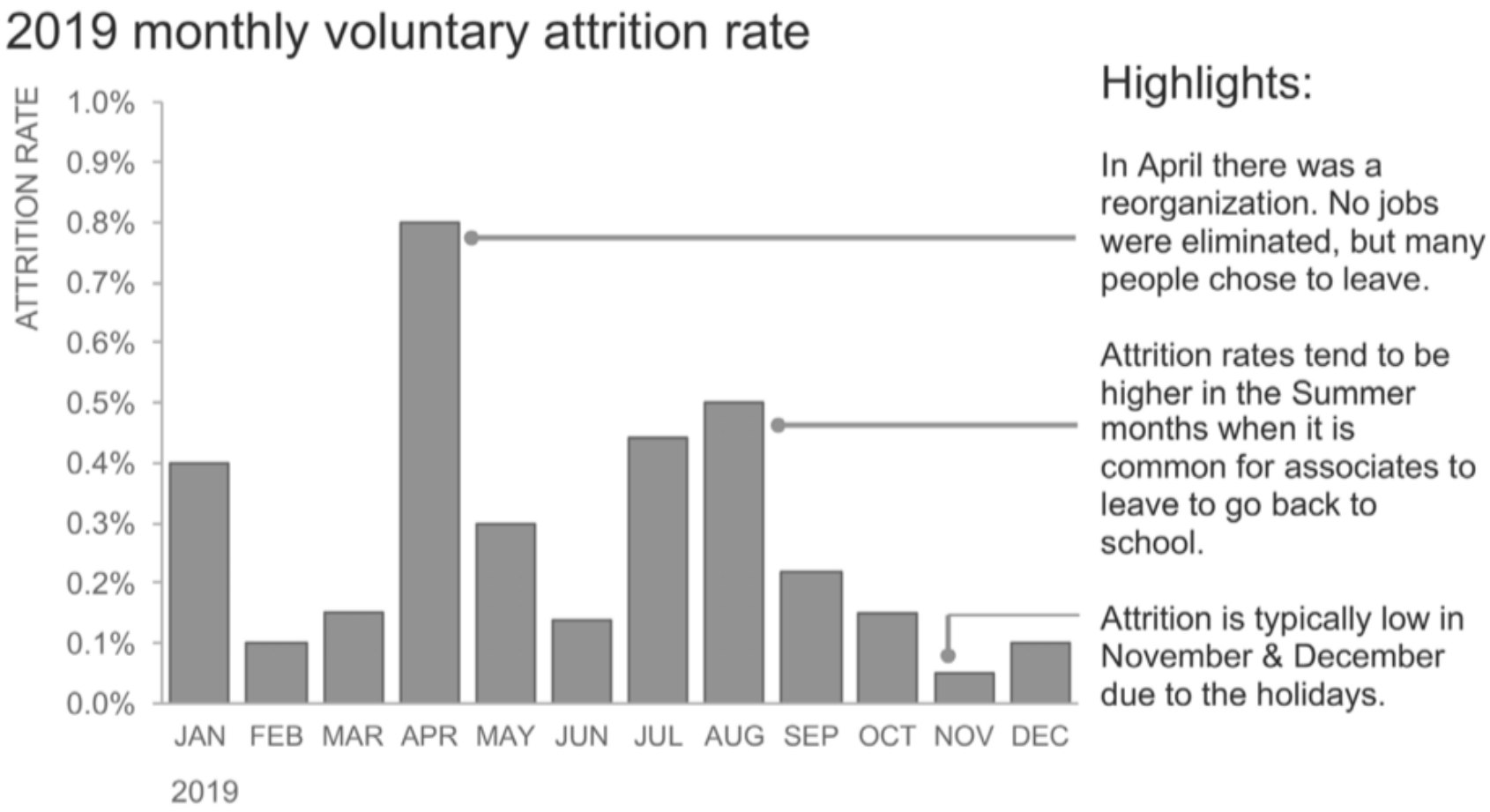


CHOICE OF IMAGE





CHOICE OF WORD



[based on an idea from S. McCloud, *Making Comics*]

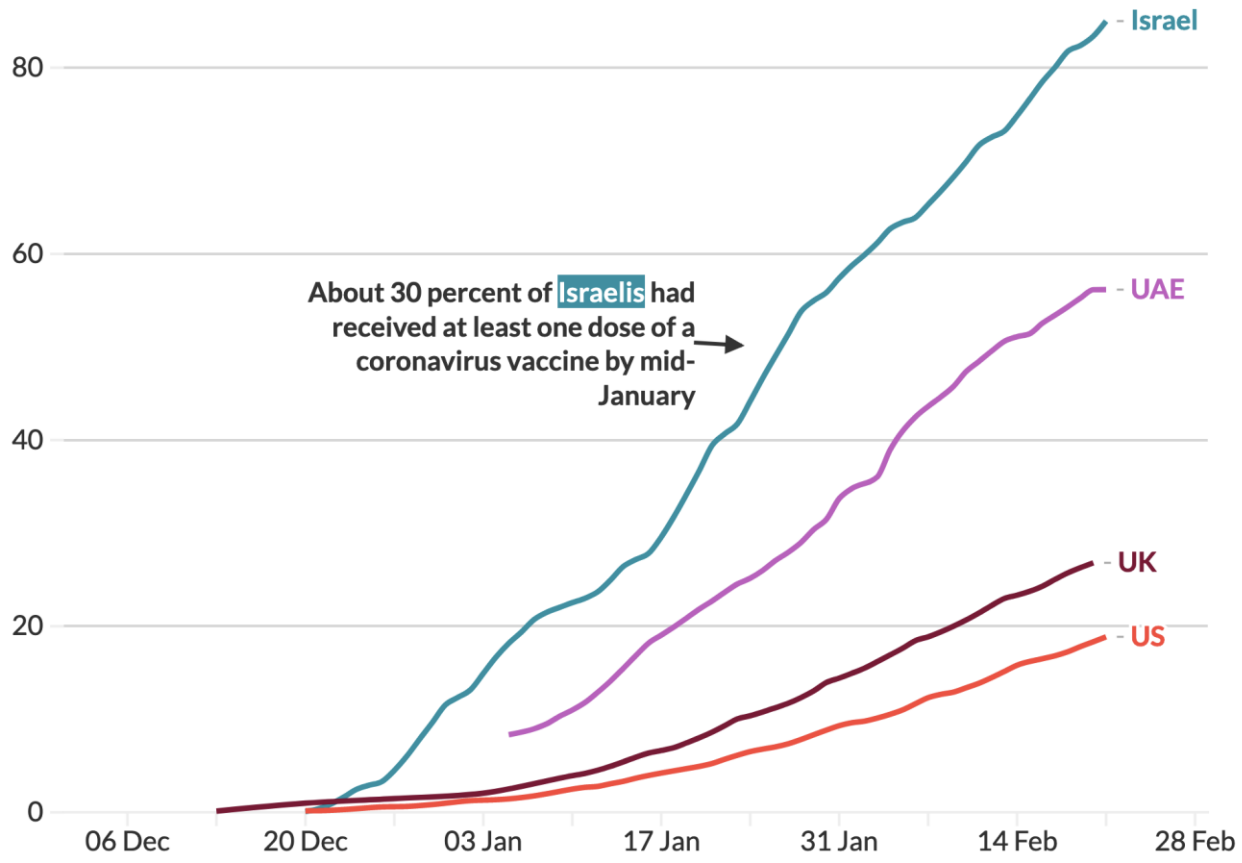
VISUAL STORYTELLING COMBINATIONS

- **text-specific**, where text provides all that is needed to know and the charts illustrate some aspects of the story that is described
- **chart-specific**, where the charts provide all that is needed to know and the text accentuates some aspects of the story that is shown
- **duo-specific**, where text and charts are both telling roughly the same story
- **intersecting**, where text and charts work together in some respects but also contribute to the story independently
- **interdependent**, where text and charts combine to convey an aspect of the story that neither could convey alone
- **parallel**, where words and charts follow seemingly different storylines, without intersecting

[https://flourish.studio/blog/accessible-chart-design]

Cumulative vaccination doses administered in Israel, UAE, UK and US

Cumulative doses administered per 100 residents • Data last updated 24 Feb



Source: ECDC/OWID • Graphic: Flourish • Embed this

[P. Tickell, The Story of the Impossible Train]

I have a story I'd like to tell you. It's about a train, and a group of people who live on that train and know of nothing else.

This train has been moving since anyone can remember. The people on the train can't imagine a time when the train wasn't moving, and when they were not on the train. Everyone works to keep the train moving. The train never stops.

It never stops. It cannot stop.

People on the train live in constant churn. The work to keep the train moving is hard, and inhumane. On the train, people are treated with cruelty and oppression. Some are treated worse than others. But nobody is truly living.

Sometimes they get breaks, but it is hard.

One day, a fire breaks out in one of the carriages of the train.

There is panic. The fire spreads throughout the whole train… Without getting off the train everyone is going to die.

Then the impossible happens.

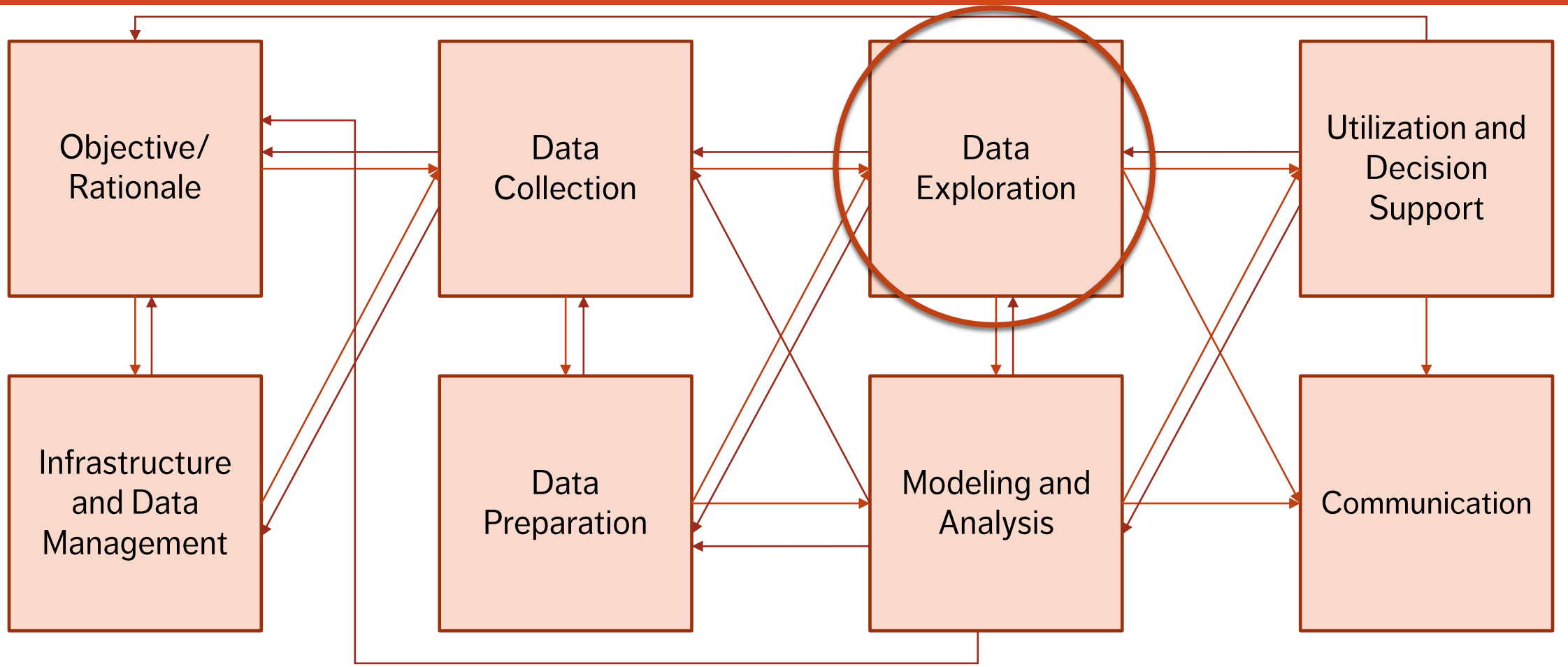
The brakes no-one believed existed start to work. In the emergency, no-one notices how extraordinary it is that the train is stopping. They're too focused on the fire. The old rules go out the window.

For years on the train, the "worker class" of people have been dying from the awful conditions of the work they have to do on the train. They sleep in the aisles and sometimes have nowhere to sleep at all.

Suddenly, there are orders to house them and treat their ailments.

The train stops, and people begin to get off. Apart from the sound of the fire, suddenly there is a great silence.

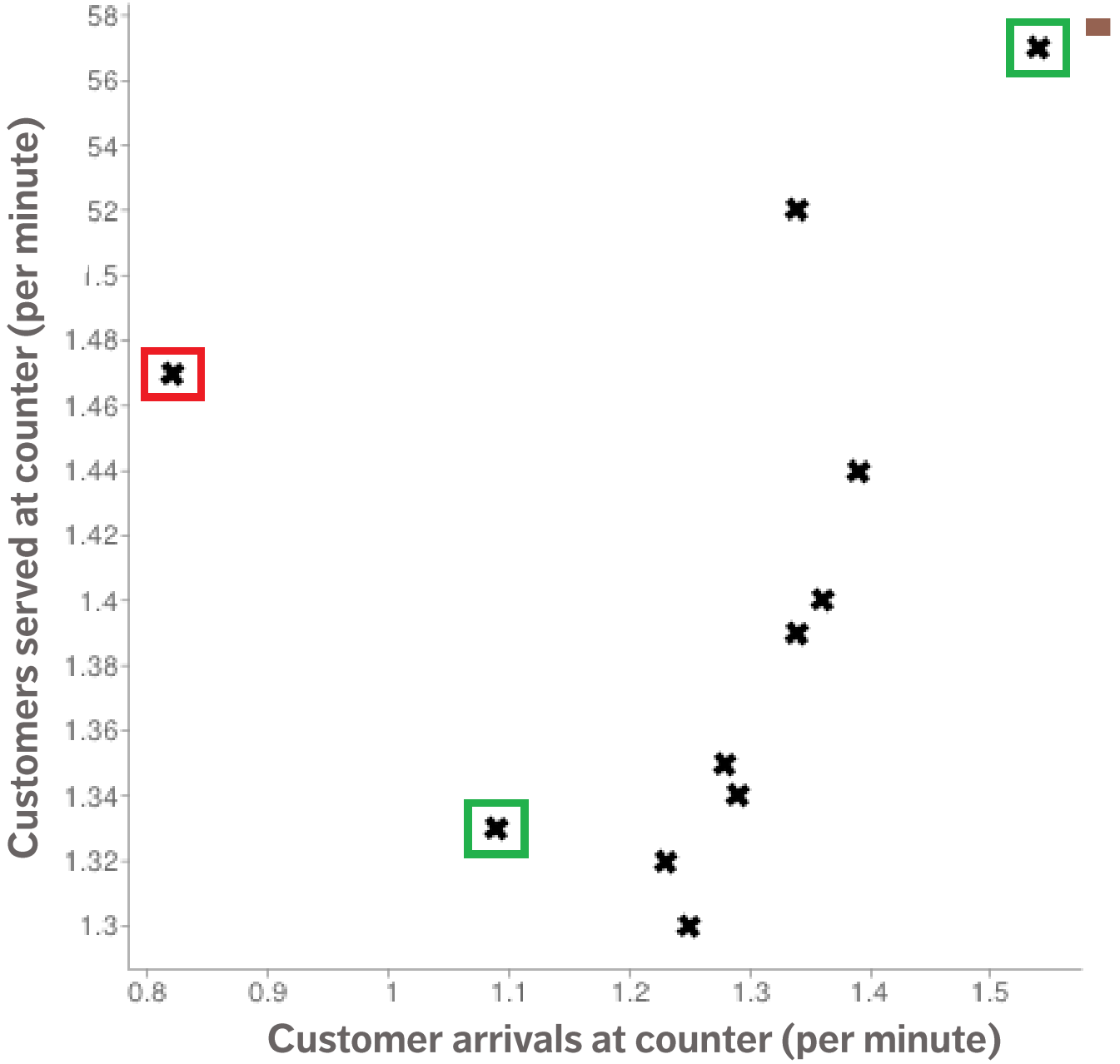
THE (MESSY) ANALYSIS PROCESS



PRE-ANALYSIS USE

Data visualization can be used to set the stage for analysis:

- **detecting anomalous entries**
  invalid entries, missing values, outliers
- **shaping the data transformations**
  binning, standardization, Box-Cox transformations, PCA-like transformations
- **getting a sense for the data**
  data analysis as an art form, exploratory analysis
- **identifying hidden data structure**
  clustering, associations, patterns informing the next stage of analysis

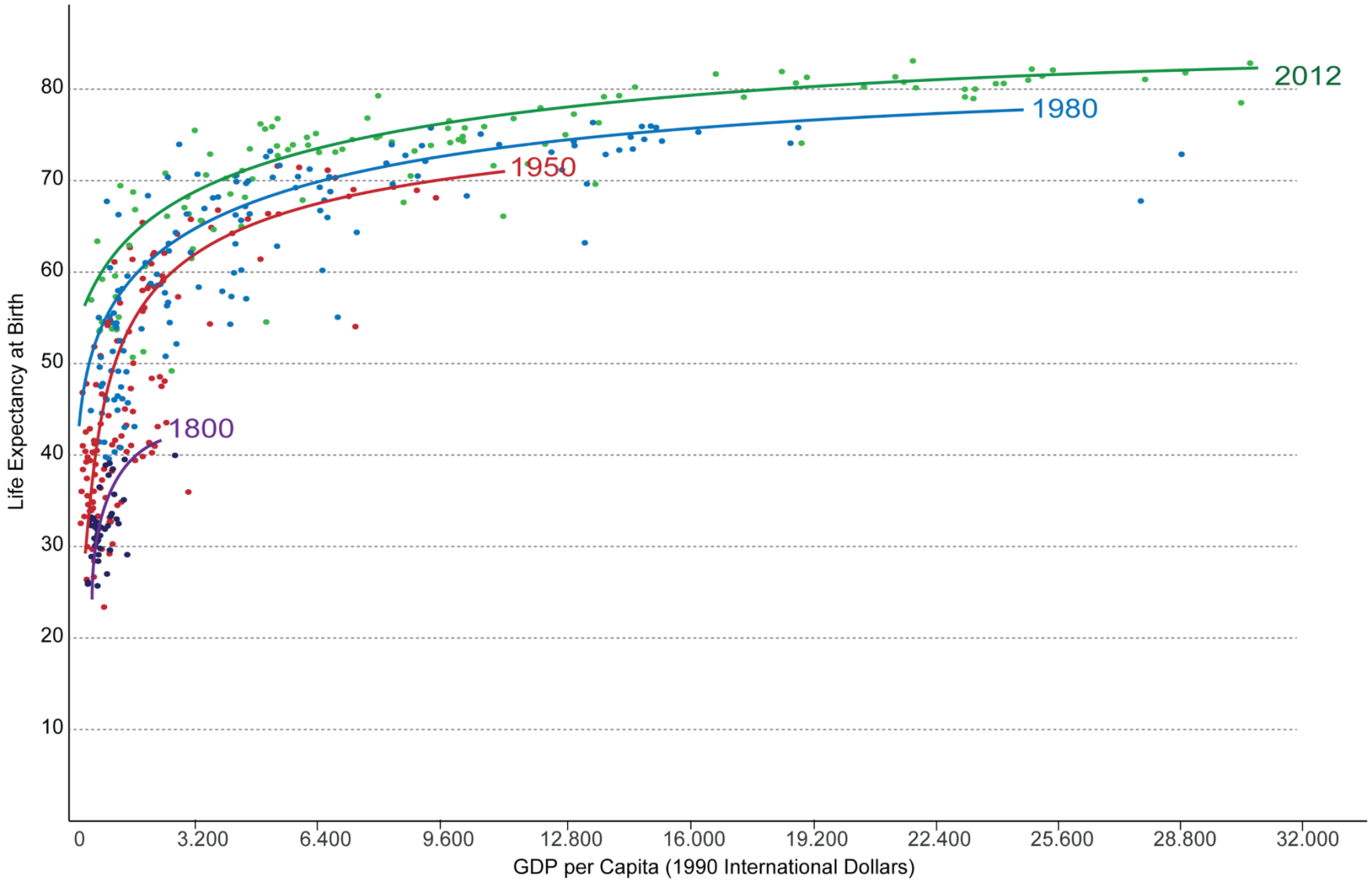

Customers served at counter (per minute)
58
56
54
52
1.5
1.48
1.46
1.44
1.42
1.4
1.38
1.36
1.34
1.32
1.3
0.8
0.9
1
1.1
1.2
1.3
1.4
1.5
Customer arrivals at counter (per minute)

[https://ourworldindata.org/life-expectancy]

Our World in Data

Life Expectancy vs. GDP per Capita from 1800 to 2012 – by Max Roser

GDP per capita is measured in International Dollars. This is a currency that would buy a comparable amount of goods and services a U.S. dollar would buy in the United States in 1990. Therefore incomes are comparable across countries and across time.



This graph displays the correlation between life expectancy and GDP per capita.

Countries with higher GDP have a higher life expectancy, in general.

The relationship seems to follow a logarithmic trend: the unit increase in life expectancy per unit increase in GDP decreases as GDP per capita increases.

Data sources: Data on life expectancy are from Gapminder.org; data on GDP per capita are from the 'New Maddison Project Database'.
The interactive data visualisation is available at OurWorldinData.org. There you find the raw data and more visualisations on this topic.

data-action-lab.com

IN PREPARATION FOR THE NEXT SESSION

1. Find examples of data presentations that you consider to be particularly insightful and/or powerful. Discuss their strengths/weaknesses.

2. Find examples of data presentations that you consider to be particularly misleading and/or useless. Discuss their strengths/weaknesses.

3. In teams or individually, identify a few data visualizations that appeal to you. What is the story being told by the visualization? What kind of data is needed to build these visualizations?

[E. Tufte, Beautiful Evidence]

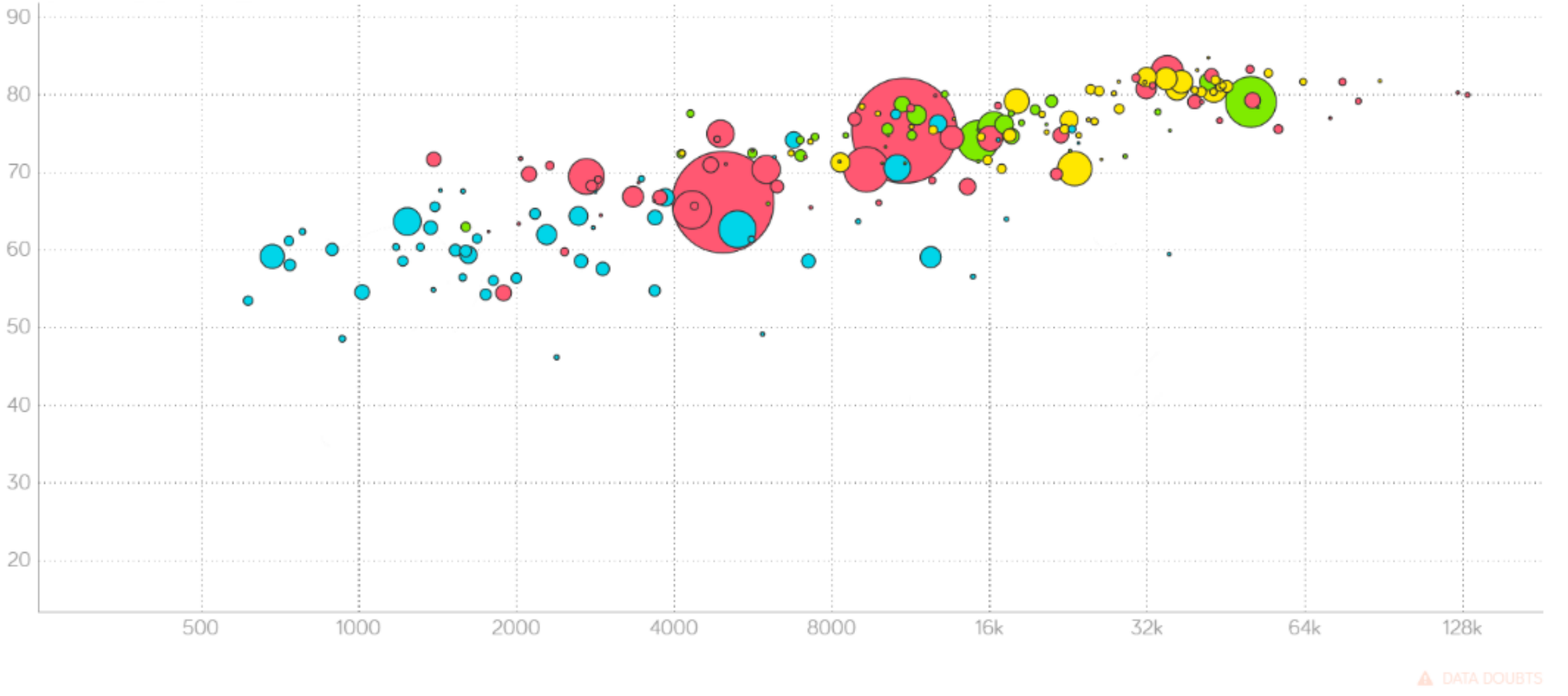
PRINCIPLES OF ANALYTICAL DESIGN

Reasoning and **communicating** our thoughts are intertwined with our lives in a causal and dynamic multivariate Universe.

Symmetry to visual displays of evidence: consumers should be seeking exactly what producers should be providing, namely:

- meaningful comparisons
- potential causal networks and underlying structure
- multivariate links
- integrated and relevant data
- honest documentation
- primary focus on content

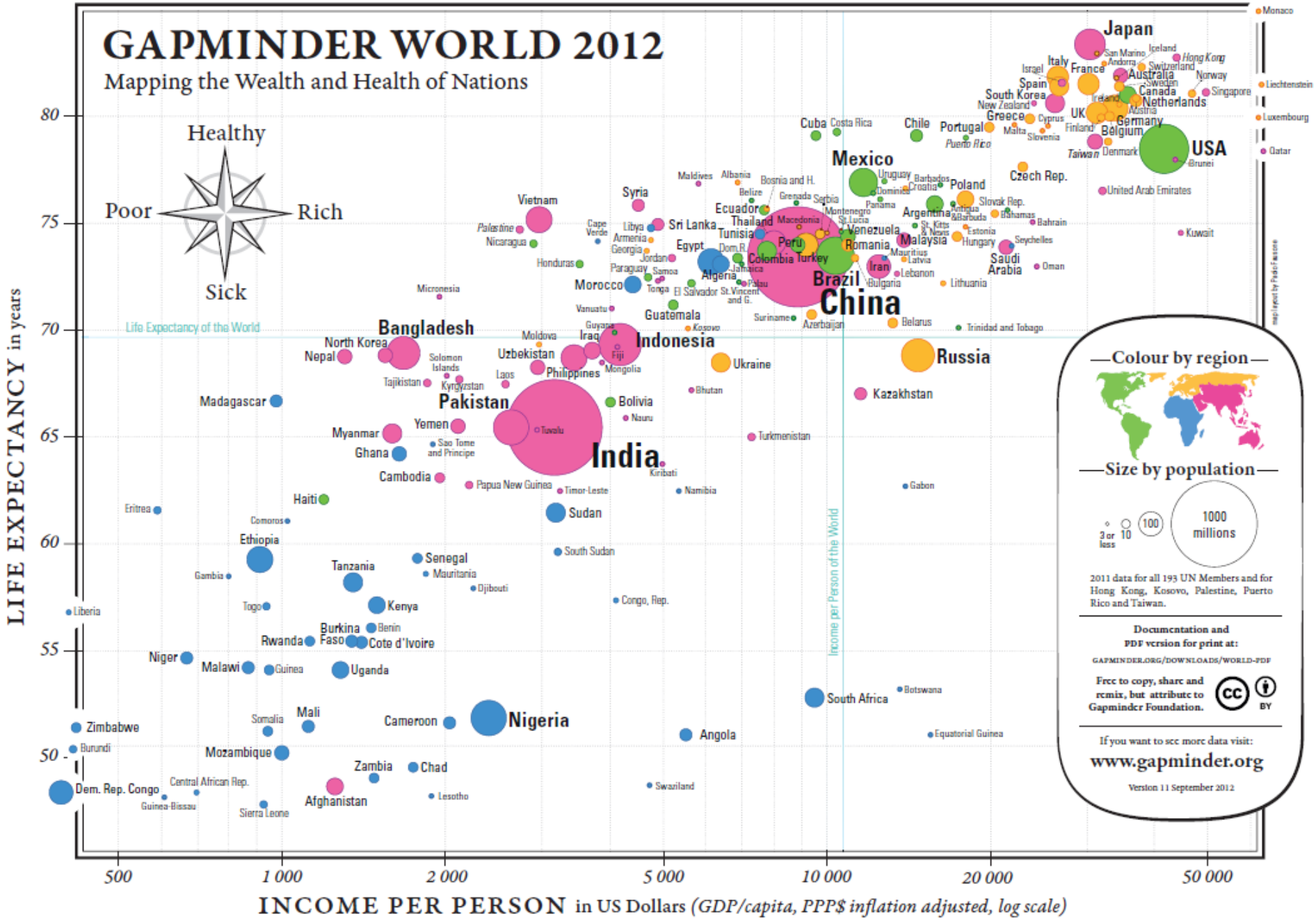
[Gapminder Foundation]



Non-Integrated Data

data-action-lab.com

[Gapminder Foundation, https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen]

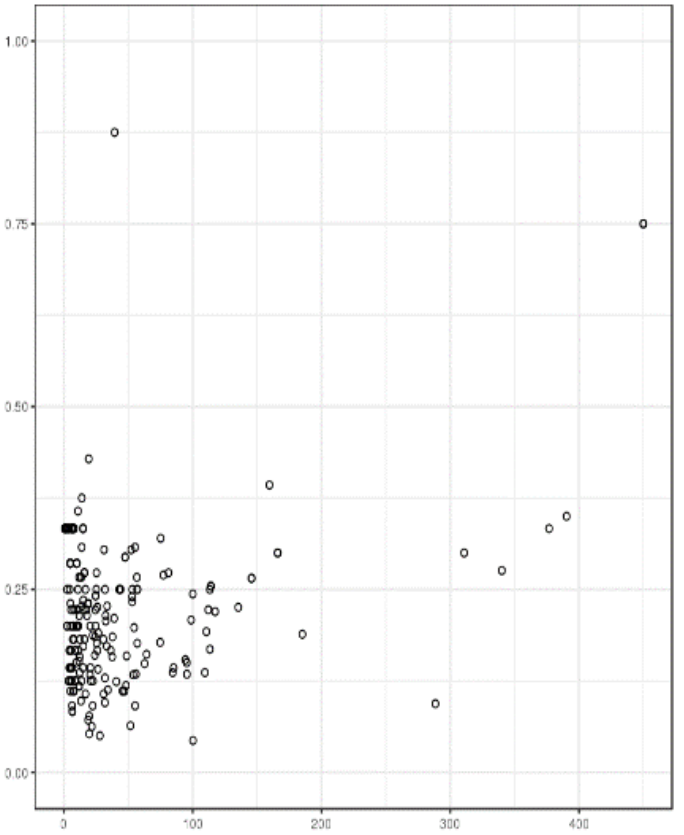


[http://promise.site.uottawa.ca/SERepository/datasets/cm1.arff]

REPRESENTING MULTIVARIATE DATA

2 variables can be represented by **position** in the plane.
Additional factors can be depicted with:

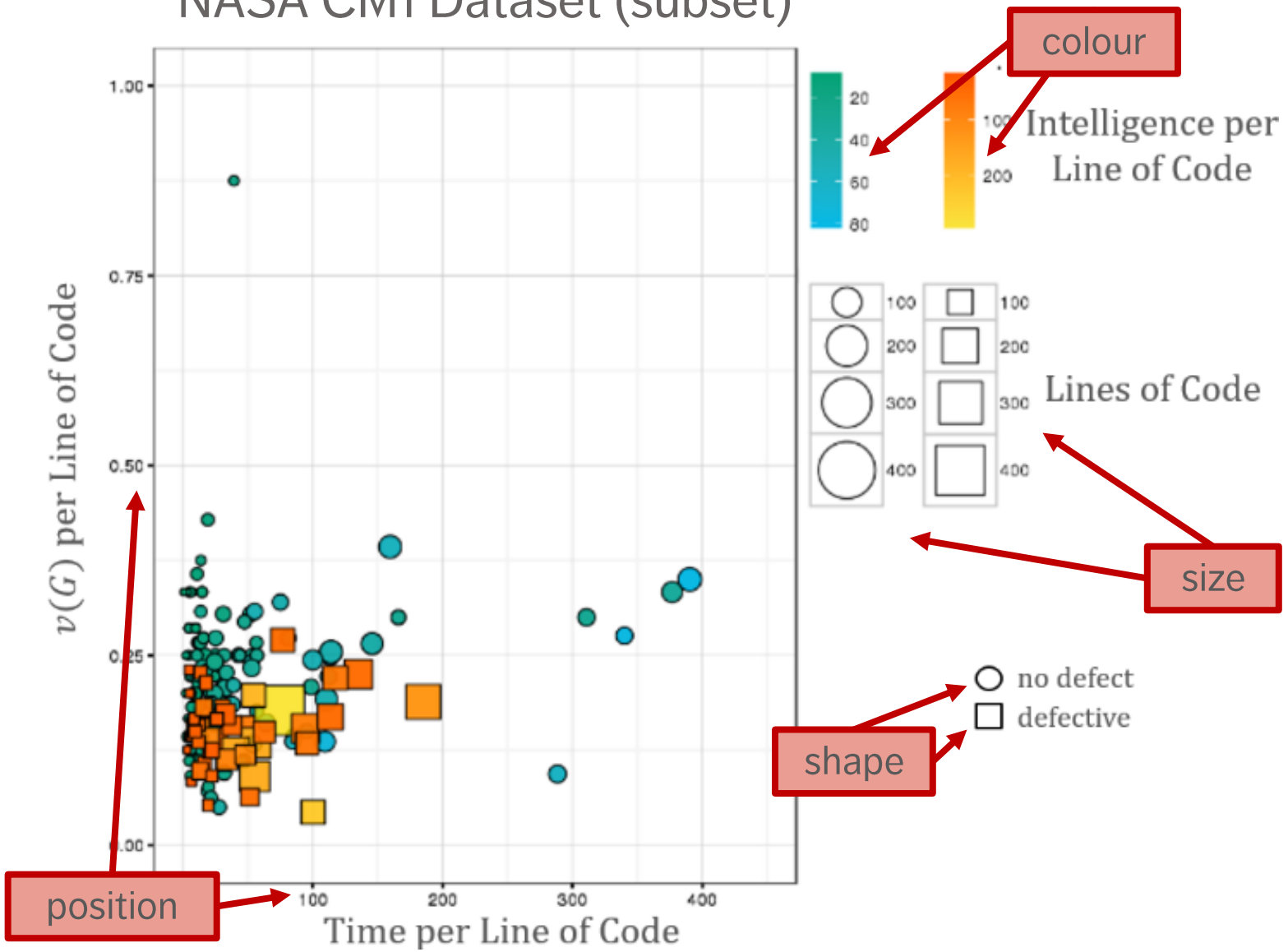
- size
- color
- value
- texture
- line orientation
- shape
- (motion?)



NASA CM1 Dataset (subset)

[http://promise.site.uottawa.ca/SERepository/datasets/cm1.arff]

NASA CM1 Dataset (subset)



VISUAL PROCESSING

Perception is fragmented – eyes are **ever scanning**.

Visual thinking seeks **patterns**

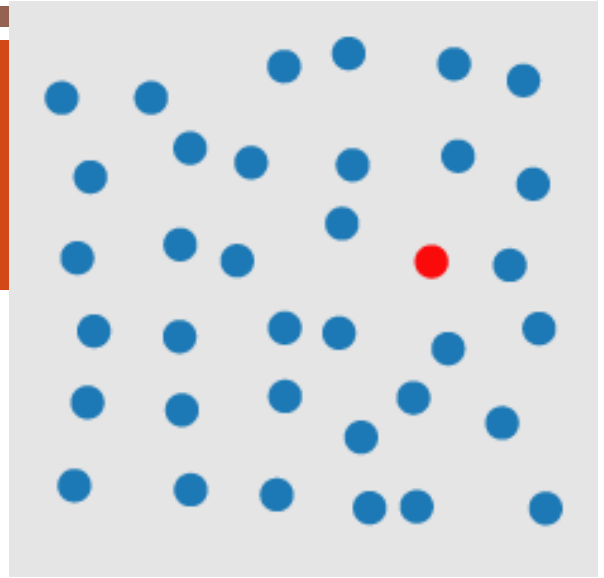
- **pre-attentive processes:** fast, instinctive, efficient, multitasking
  gather information and build patterns:

  features → patterns → objects

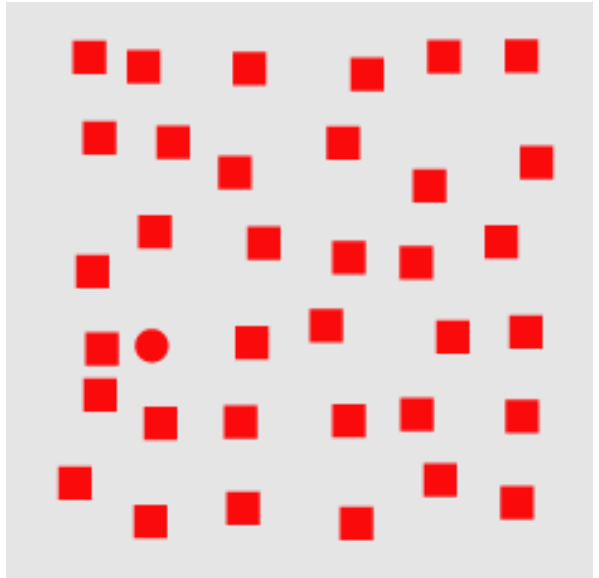
- **attentive process:** slow, deliberate, focused
  discover features in the patterns:

  objects → patterns → features

pre-attentive



attentive



[www.interaction-design.org/literature/topics/gestalt-principles]

GESTALT PRINCIPLES

The **Gestalt principles** are the “laws” of human perception.

They describe how humans group similar elements, recognize patterns and simplify complex images when they perceive objects.

Designers use them to organize content on charts, dashboards, websites, and other interfaces so that they be **aesthetically pleasing/easy to understand**.

[www.interaction-design.org/literature/topics/gestalt-principles]

GESTALT PRINCIPLES

"Gestalt" is German for "unified whole".

The first principles were devised in the 1920s by German psychologists Wertheimer, Koffka ("the whole is greater than the sum of the parts"), Kohler.

Aim: understand how we gain meaning from the chaotic stimuli around us.

The Gestalt principles are a set of "laws" which address the natural compulsion to find order in disorder. According to this, the mind "informs" what the eye sees by **perceiving a series of individual elements as a whole**.

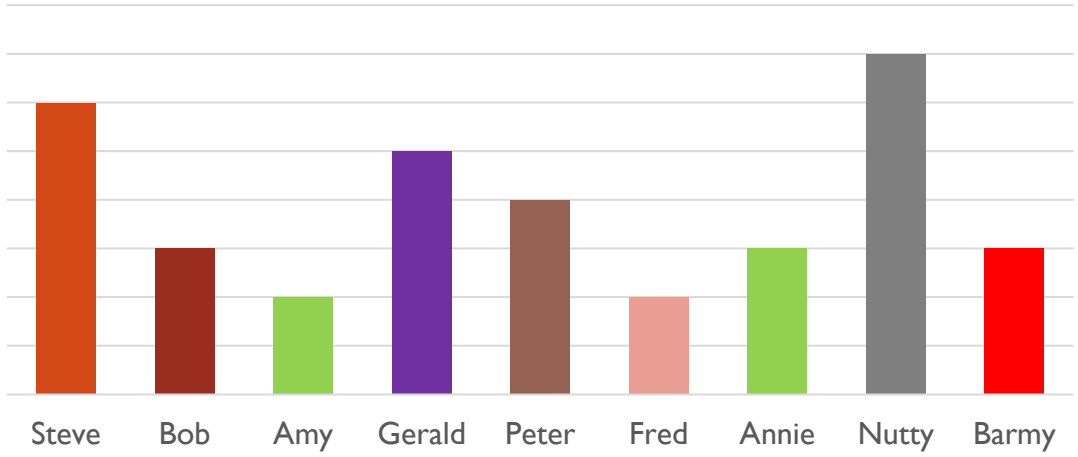
GESTALT PRINCIPLES

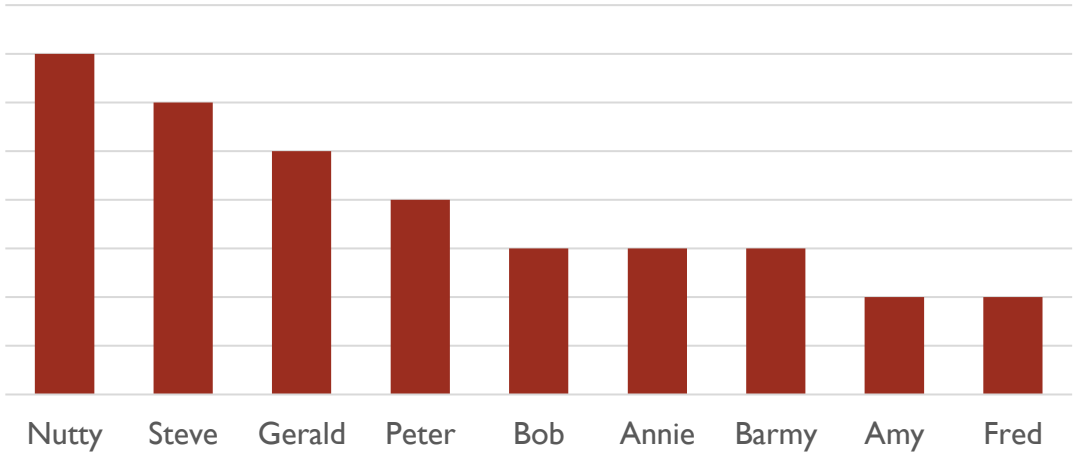
- **simplicity**
- continuation
- **proximity**
- **similarity (invariance)**
- **focal point**
- isomorphic correspondence
- **figure / ground duality**
- common fate
- closure
- uniform connectedness

GESTALT PRINCIPLES – SIMPLICITY

The brain has a preference for **simplicity** – it tends to process simple patterns faster than patterns that are more complex.

Lesson: arrange data simply and logically wherever possible.





GESTALT PRINCIPLES – PROXIMITY

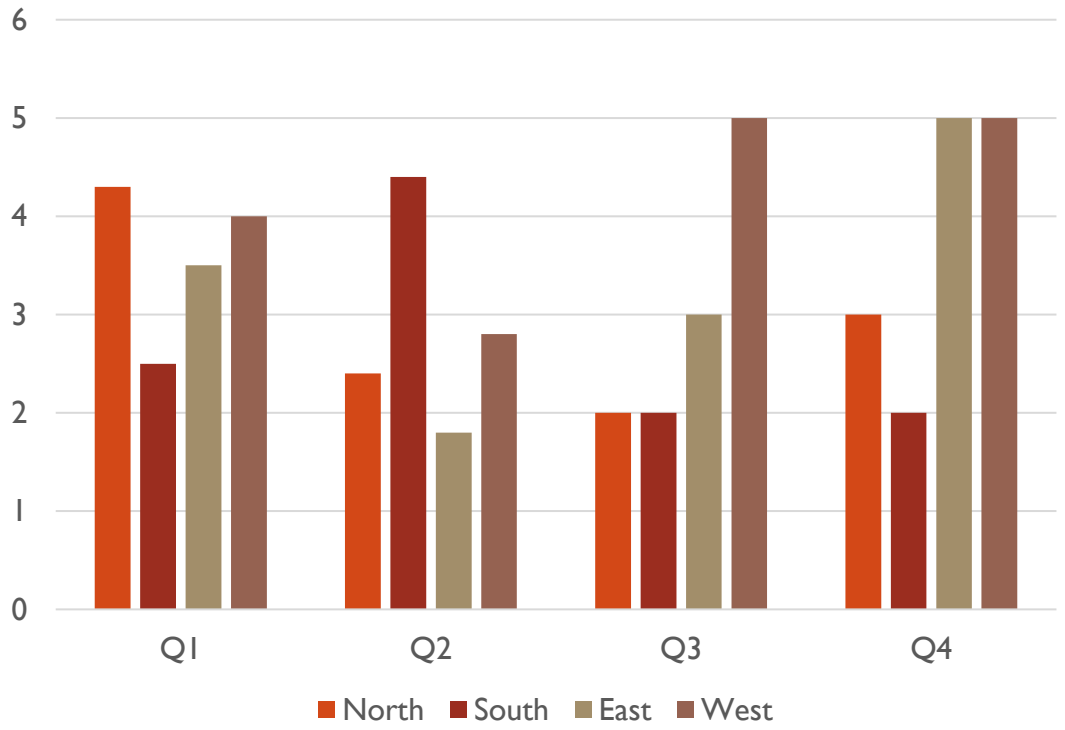
Objects/shapes that are in **proximity** (close) appear to form **groups**.

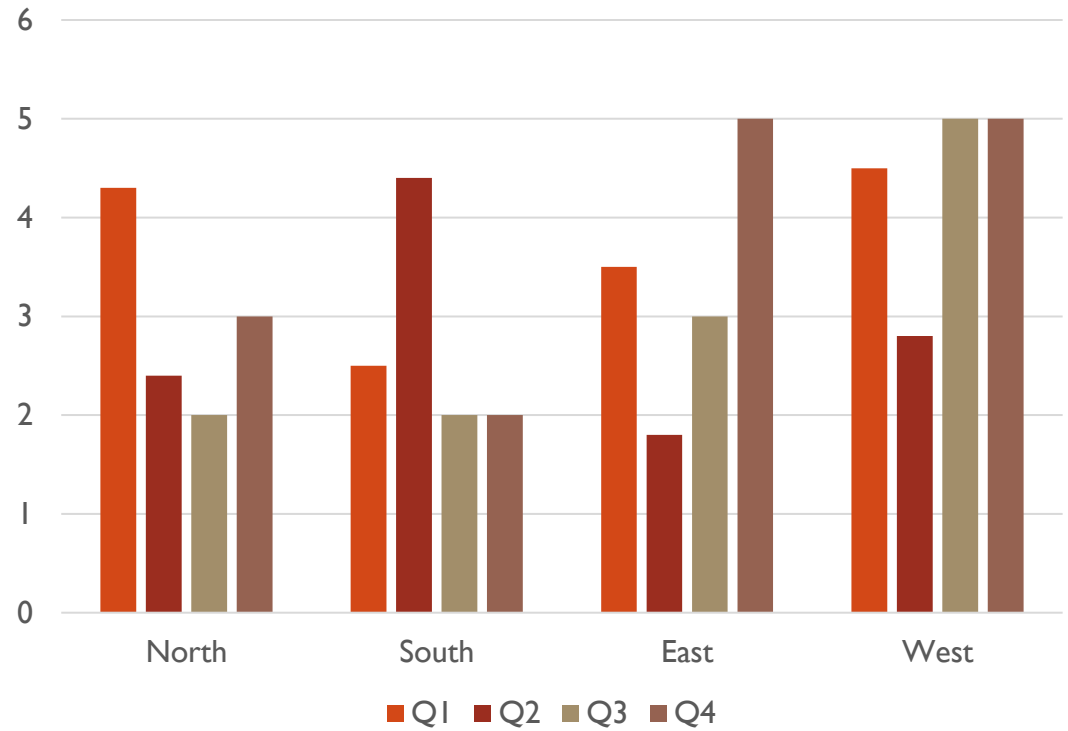
The effect generated by the collected group is more “powerful” than that generated by separate elements.

Elements which are grouped together create the **illusion** of shapes/planes in space, even if the elements are not touching.

Lesson: understand the chart’s priorities and create groupings through proximity that support those priorities.

GESTALT PRINCIPLES – PROXIMITY





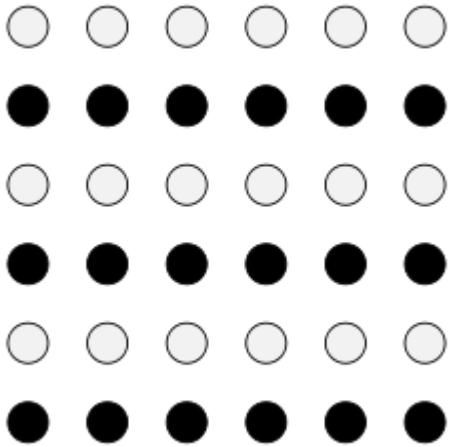
GESTALT PRINCIPLES – SIMILARITY

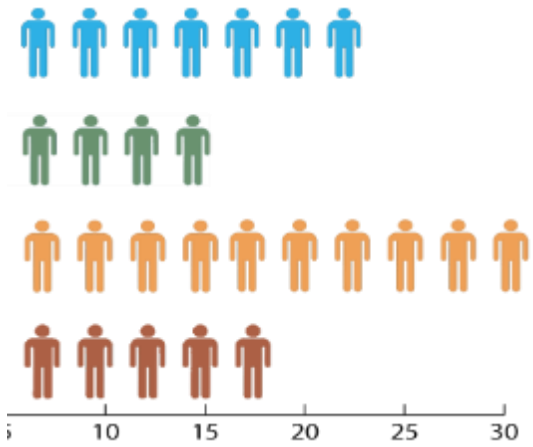
Stimuli that physically resemble each other are viewed as **part of the same object**; stimuli that don't are viewed as part of a different object.

Similarity and proximity often come together to form a **visual hierarchy**. Either principle can dominate the other, depending on their application and combination.

Lesson: use similar characteristics to establish relationships and to encourage groupings of objects.

GESTALT PRINCIPLES – SIMILARITY





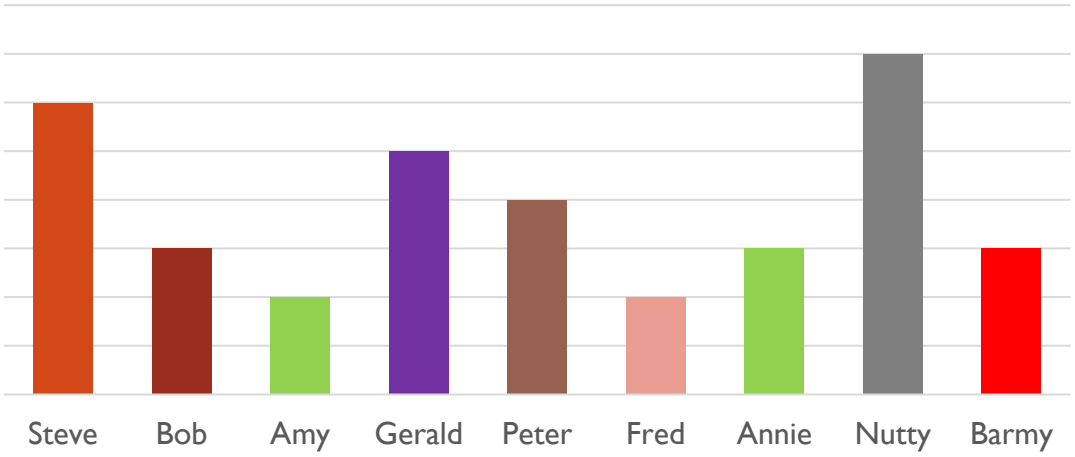
In these examples, similarity dominates over proximity: we see rows before we see columns.

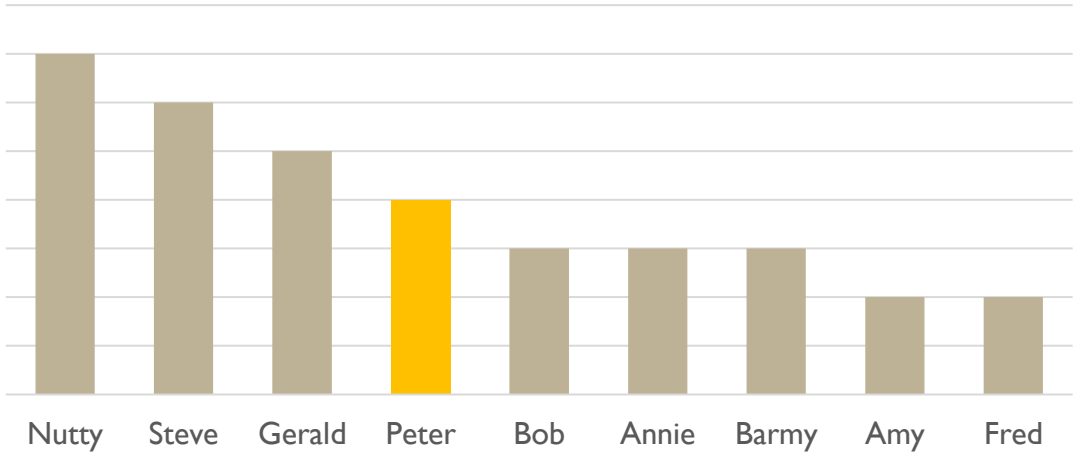
GESTALT PRINCIPLES – FOCAL POINT

In opposition to similarity, the **focal point** principle states that distinctive-looking objects can create a focal point.

To highlight one salesperson's performance, make their bar graph color different.

Lesson: use different characteristics to highlight and create focal points.





GESTALT PRINCIPLES – DUALITY

Chart elements are either perceived as **figures** (focus) or as (back)**ground**.

Foreground objects are **promoted** by the brain, background objects are **demoted**.

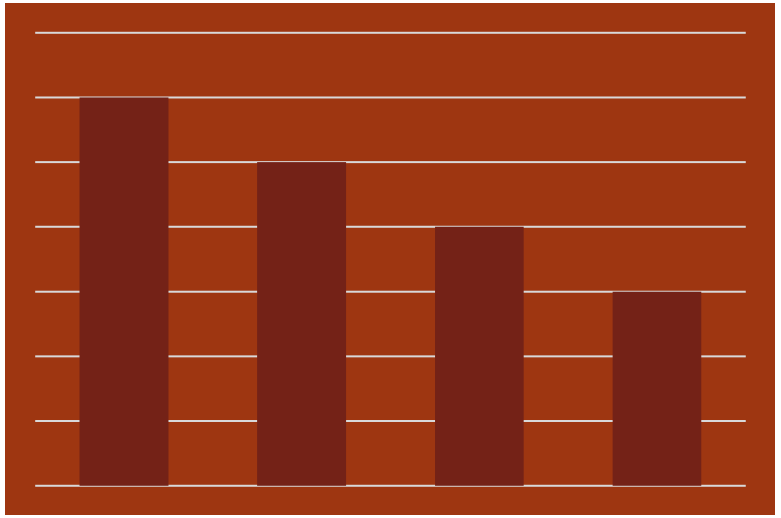
Strong contrast makes it easier to distinguish between the two types of objects.

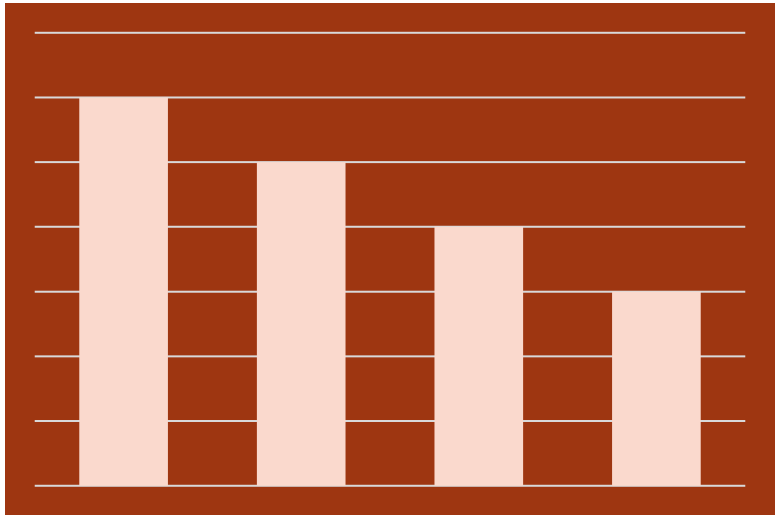
Lesson: ensure there is enough contrast between the chart foreground (figures) and their background.

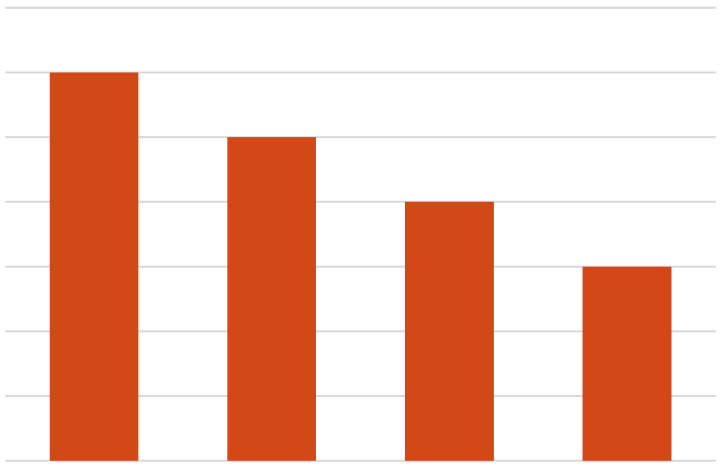
GESTALT PRINCIPLES – DUALITY

Because of the **low contrast** between the figure and background in the chart on the left, there is an **additional cognitive load**.

Increasing the contrast on the right improves readability.







DECLUTTERING

Clutter is the enemy!

Every element on a page adds cognitive load

- identify and remove anything that isn't adding value
- think of cognitive load as mental effort required to process information (lower is better)

Tufte refers to the **data-to-ink ratio** – "the larger the share of a graphic's ink devoted to data, the better"

In *Resonate*, Duarte refers to this as "**maximizing the signal-to-noise ratio**" where the signal is the information or the story we want to communicate.

DECLUTTERING

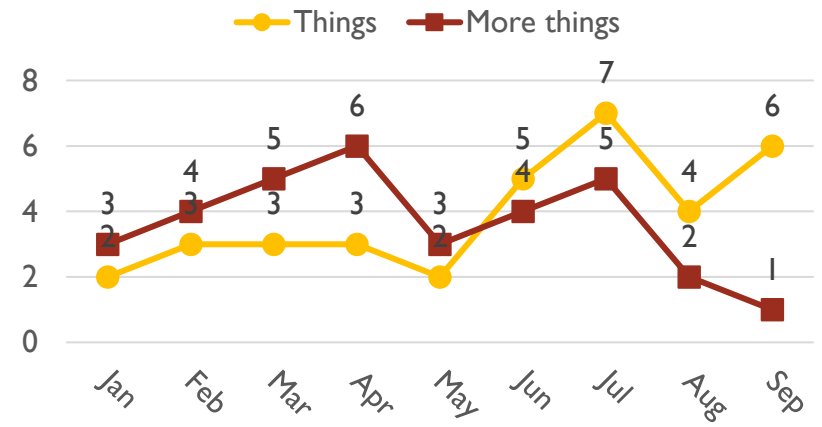
Use the Gestalt Principles to **organize/highlight** data in the chart.

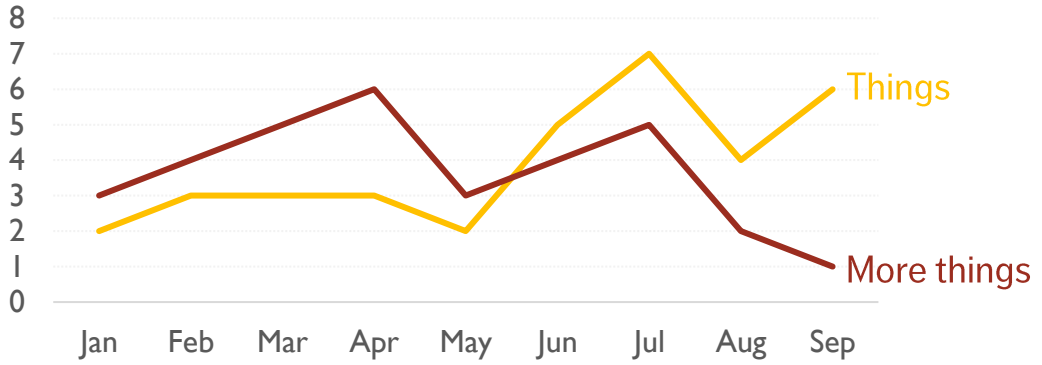
Align all elements (graphs, text, lines, etc.):

- don't rely on eye, use position boxes and values

Charts:

- remove border, gridlines, data markers
- clean up axis labels
- label data directly





[L.C. Muth, https://blog.datawrapper.de/fonts-for-data-visualization]

DECLUTTERING

Use **consistent** fonts, font size, colour and alignment.

Don't rotate text to anything other than 0 or 90 degrees (however: English/French incompatibility with vertical text).

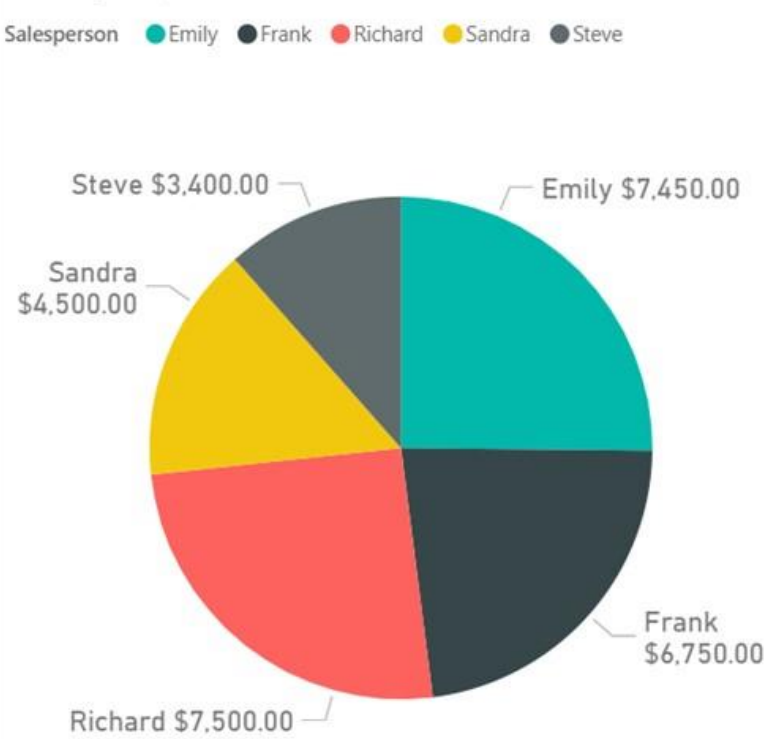
Use **white space**:

- margins should remain free of text and visuals
- don't stretch visuals to edge of page or too close to other visuals
- think of white space as a border

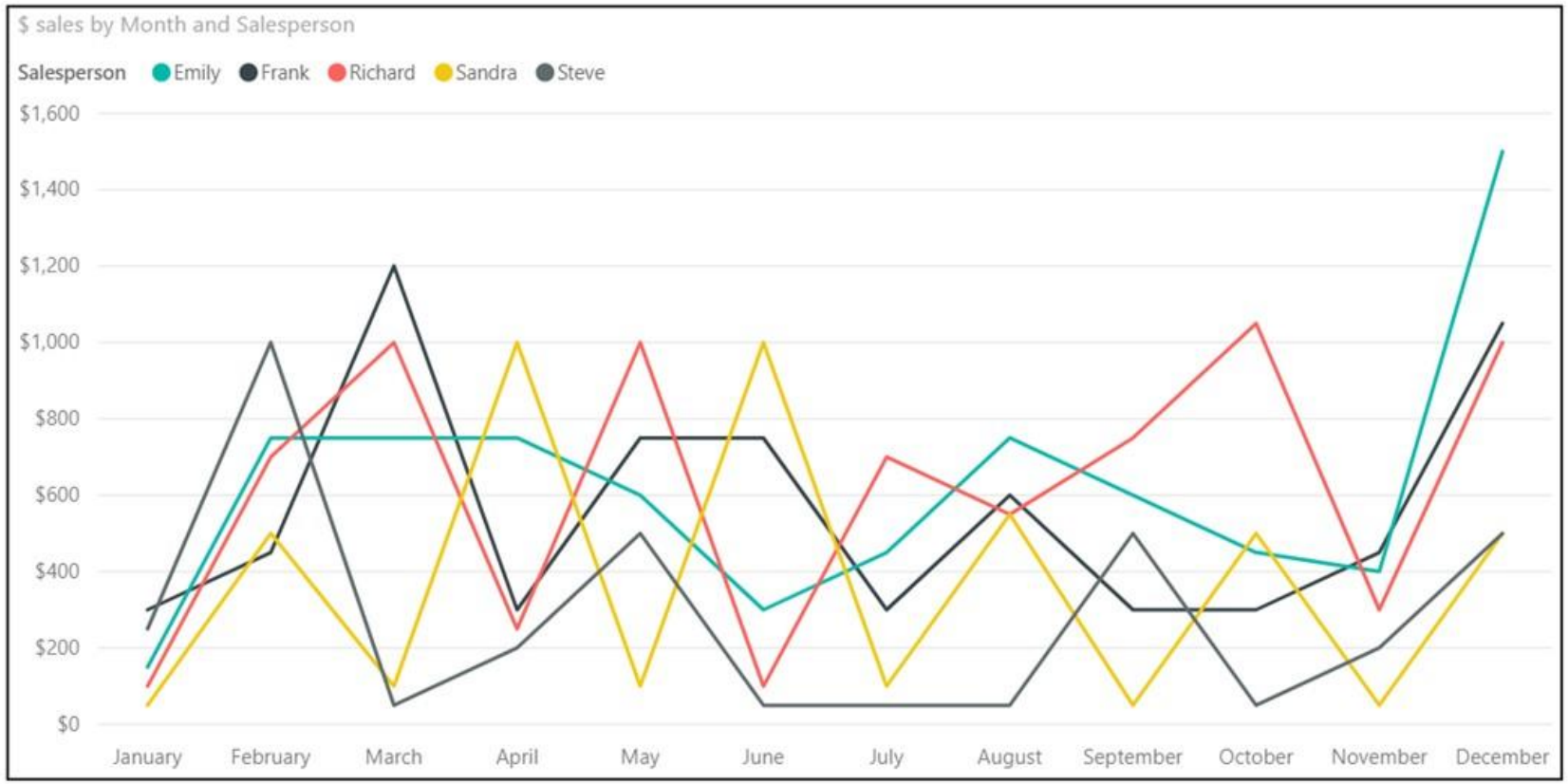
Sales Dashboard

$ sales

$ sales by Salesperson



$ sales by Month and Salesperson



$ sales by Product and Salesperson

Sales Dashboard

Annual Sales for 2017

Total Sales

$29.6K

A CLASSIFICATION OF CHART TYPES

Data comparison charts

Comparison

Bars

Dot plot

Bullet

1D Scatterplot

Heat map

Slope

Alert

Composition

Pie

Pareto

Multidimensional Pie

Distribution

Histogram

1D Scatterplot

Boxplot

Data reduction charts

Evolution

Line

Horizon

Step

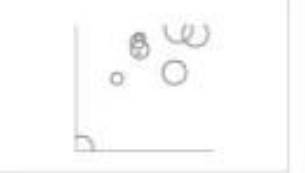
Connected Scatterplot

Relationship

Scatterplot

Connected Scatterplot

Bubble



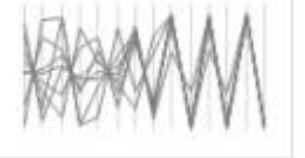
Profiling

Grouped bars

Cycle plot

Reorderable matrix

Parallel Plot



Scatterplot matrix

Horizon

Trellis

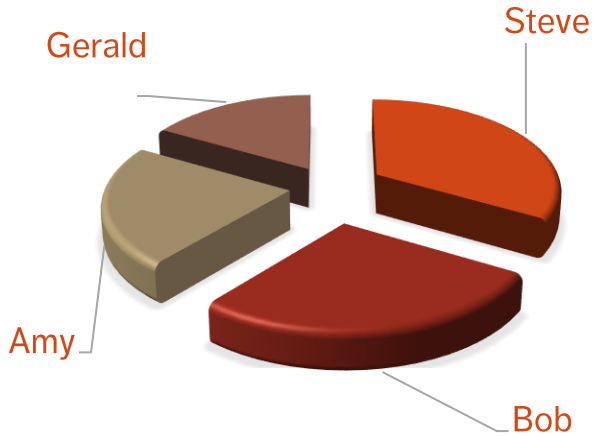
v 0.9

excelcharts.com

CHARTS TO AVOID

ANYTHING with an arc (except gauge)

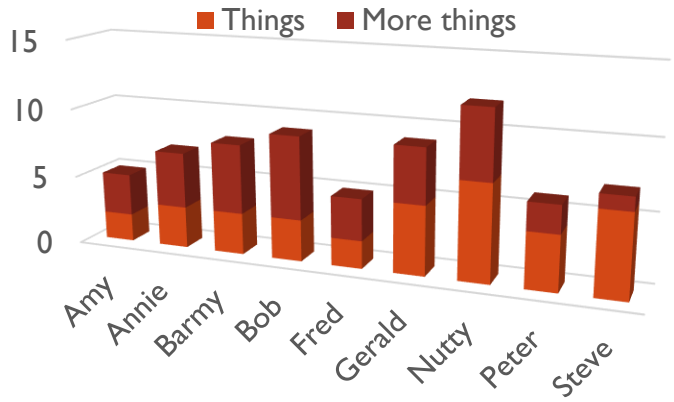
- pie
- donut



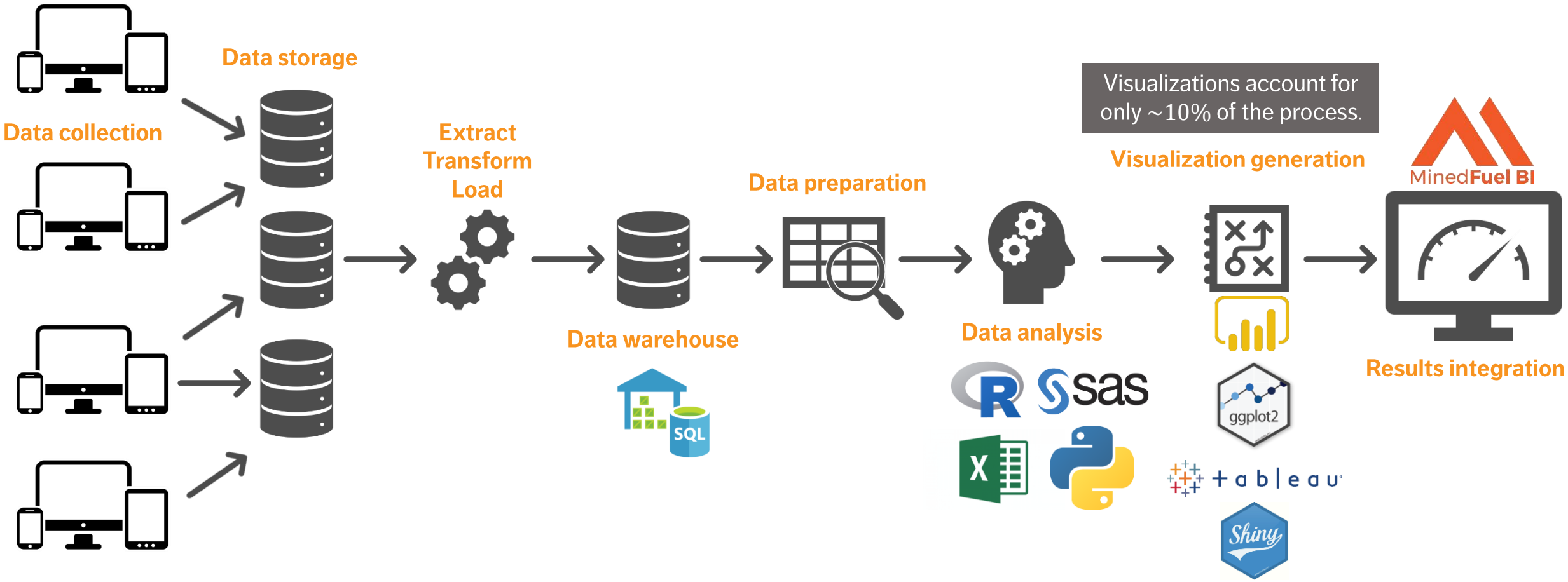
Brain cannot compare arcs and they can be misleading: how different are Steve & Bob in the pie chart?

ALL 3D IS EVIL!

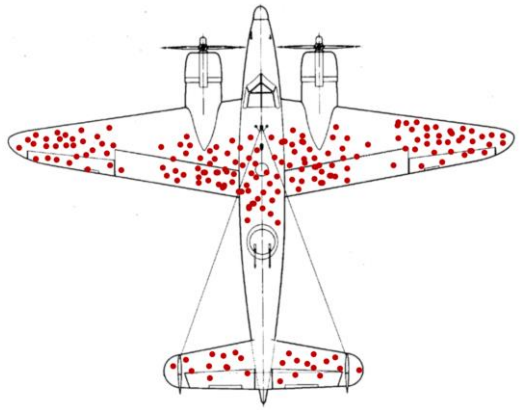
- as with arc, we cannot easily visually compare data series
- adds way too much clutter

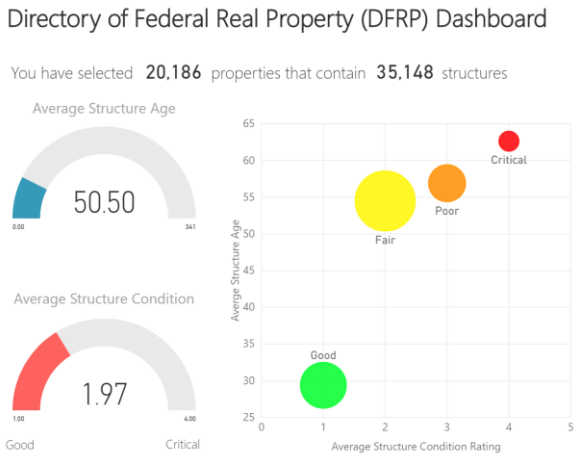


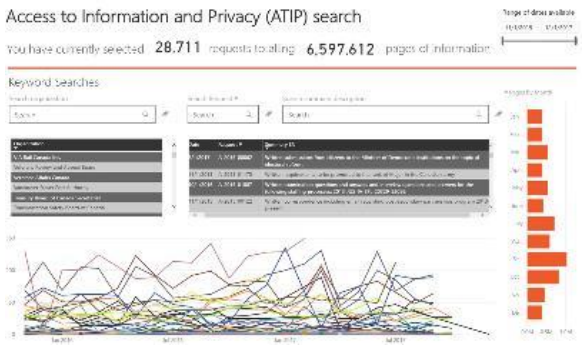
DATA ENVIRONMENT

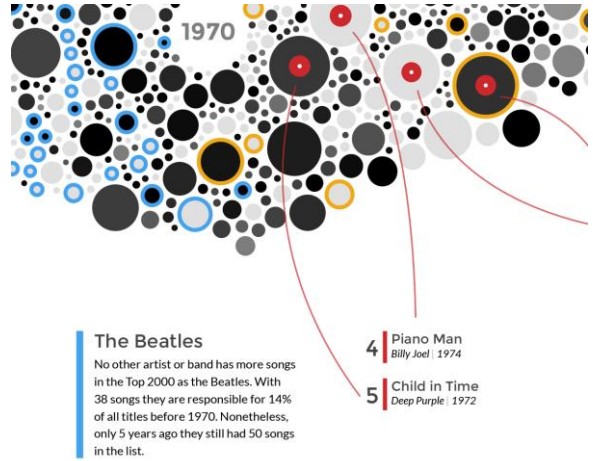


DEFINING CONTEXT









Seconds | **Minutes** | **Fraction of Hour** | **Hours**

← **Infographics/Data Viz** →

← **Dashboards** →

← **Reports and Exploration** →

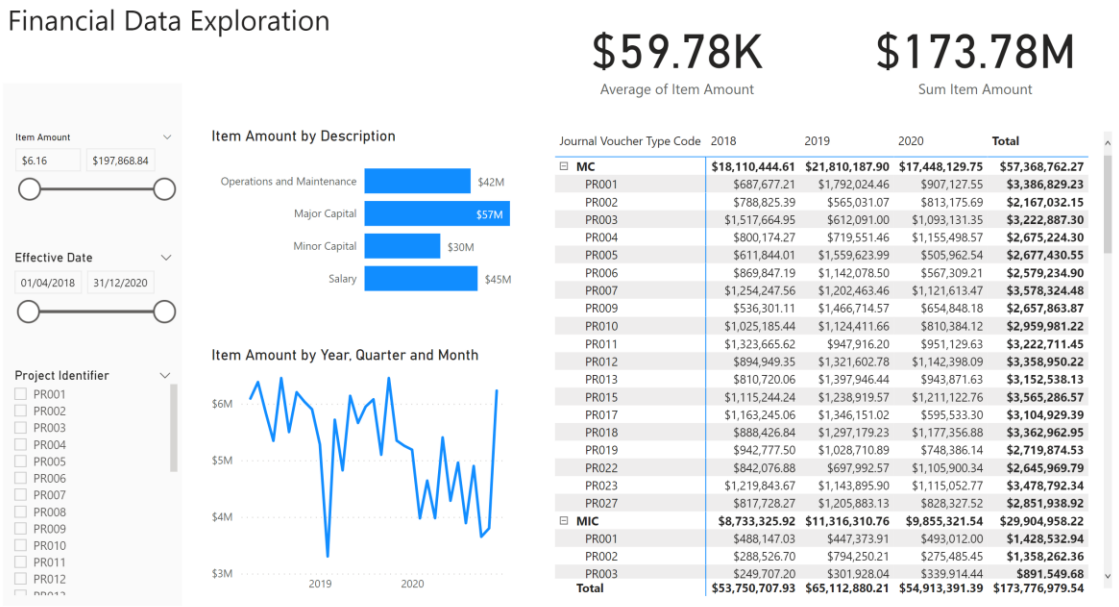
← **Data Art** →

data-action-lab.com

DASHBOARD TYPES

Exploration: using visualizations as a tool to explore/understand the data

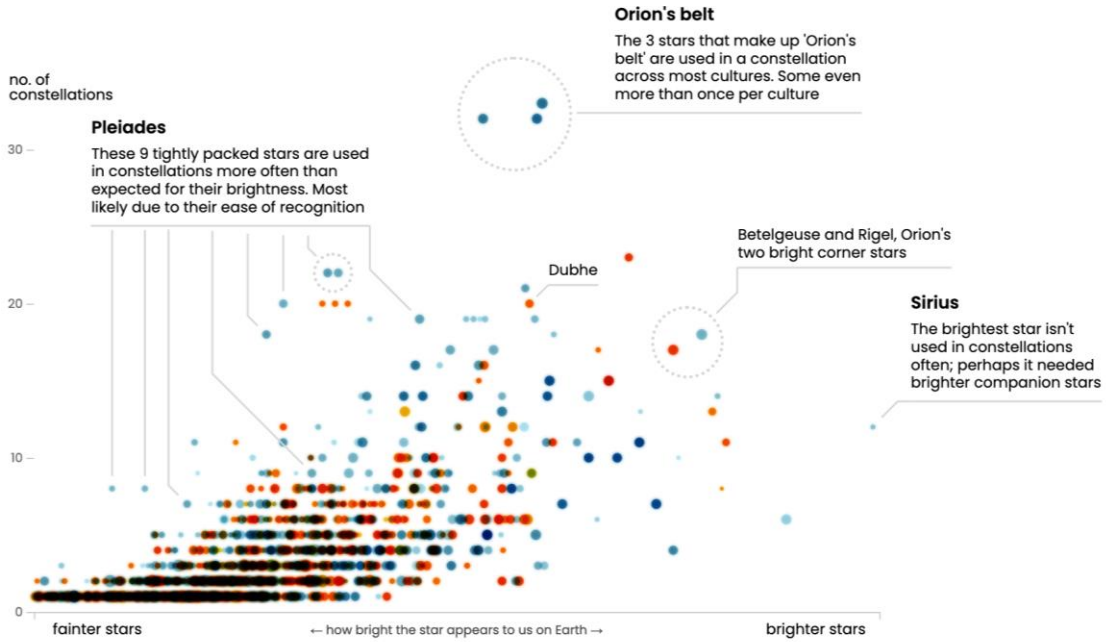
- high level of interactivity
- high level of detail
- **all** aspects of data should be represented (tables, columns, calculations etc.)
- no annotations or explanations required



DASHBOARD TYPES

Storybook: using visualizations as a tool to explain the data and communicate the story

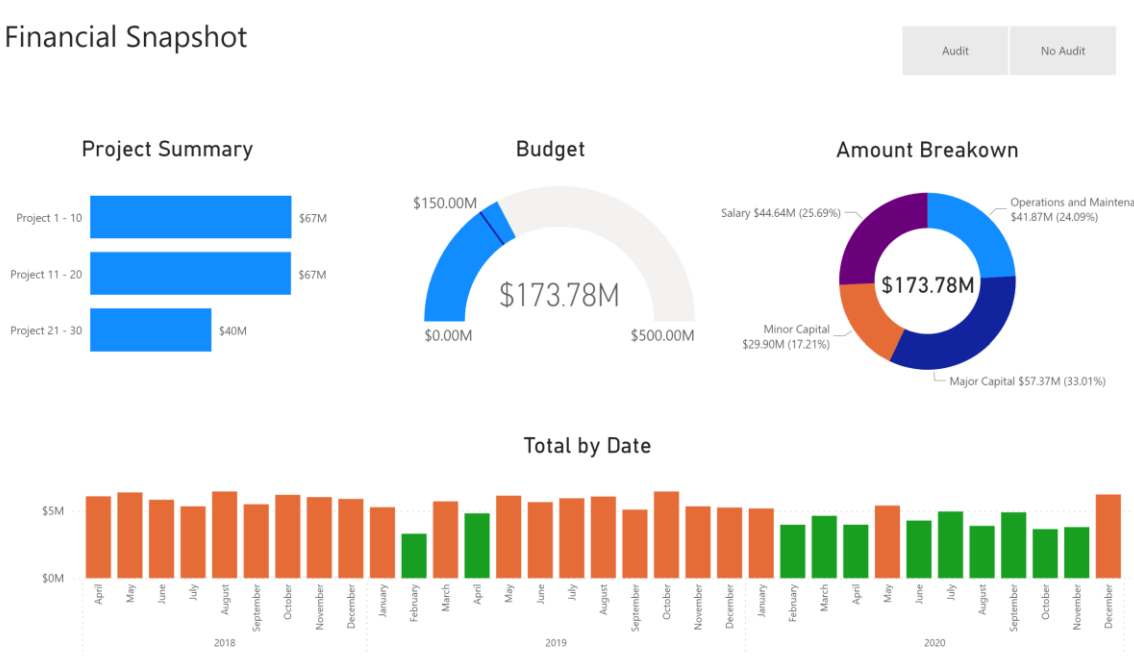
- low level of interactivity
- low level of detail
- key aspects of data should be represented
- annotations and explanations drive the “story”



DASHBOARD TYPES

Situational Awareness: using visualizations as a tool to provide a snapshot of the data

- medium level of interactivity
- not “scripted” but well organized (e.g., categorized)
- summary data should be represented
- anomalies are highlighted
- often used for internal presentations



EXERCISES

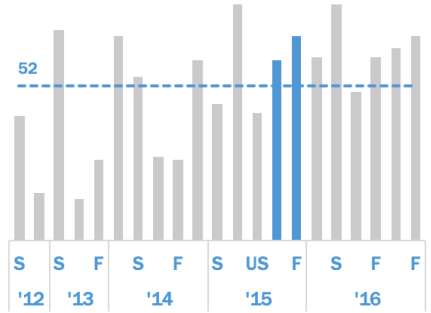
Consider the dashboards on the next 2 pages.

Can you figure out at a glance who their audience is? What are their types? Their strengths and limitations? How could you improve them?

Course Metrics

[https://bigbookofdashboards.com/dashboards.html]

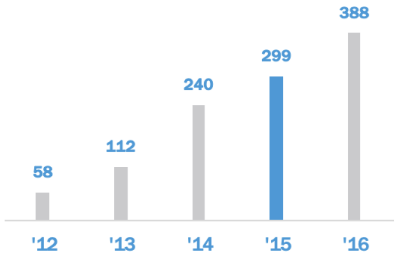
Students



1097

Total Students in five years

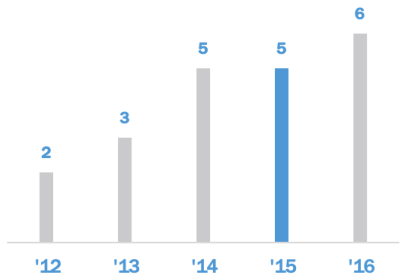
Enrollments



687

Total Students in 2015-2016

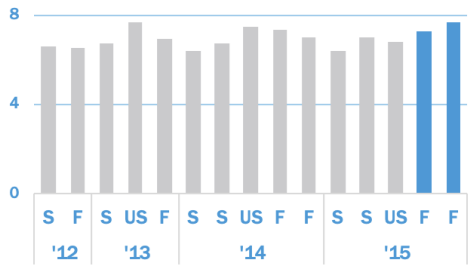
Classes



21

Total Classes in five years

Ratings



7.7 of 8

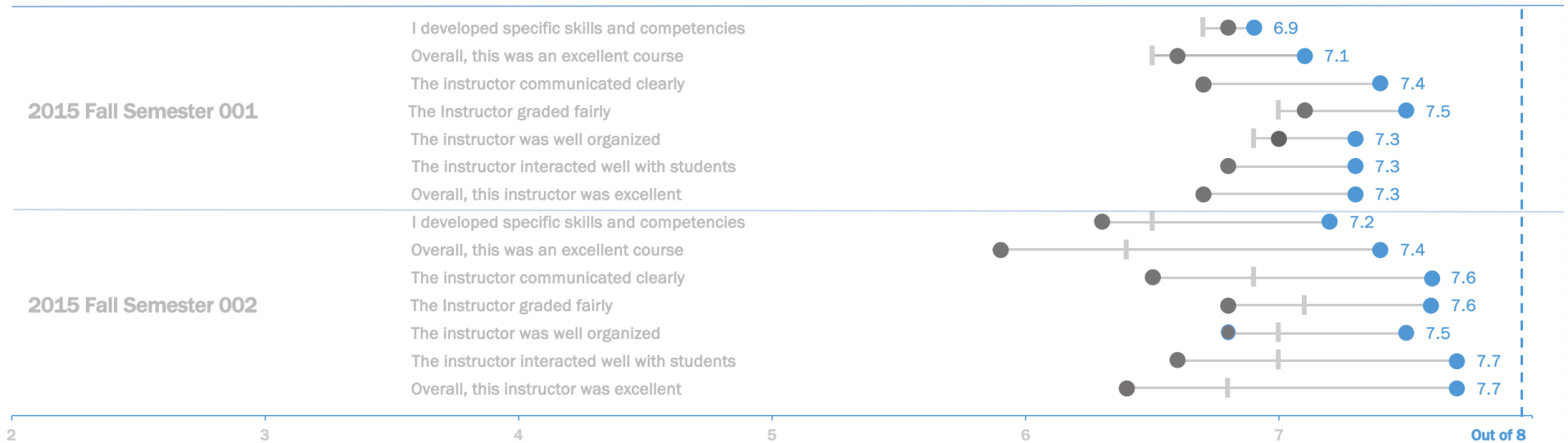
Most recent instructor rating (out of 8.0)

Semesters

Questions

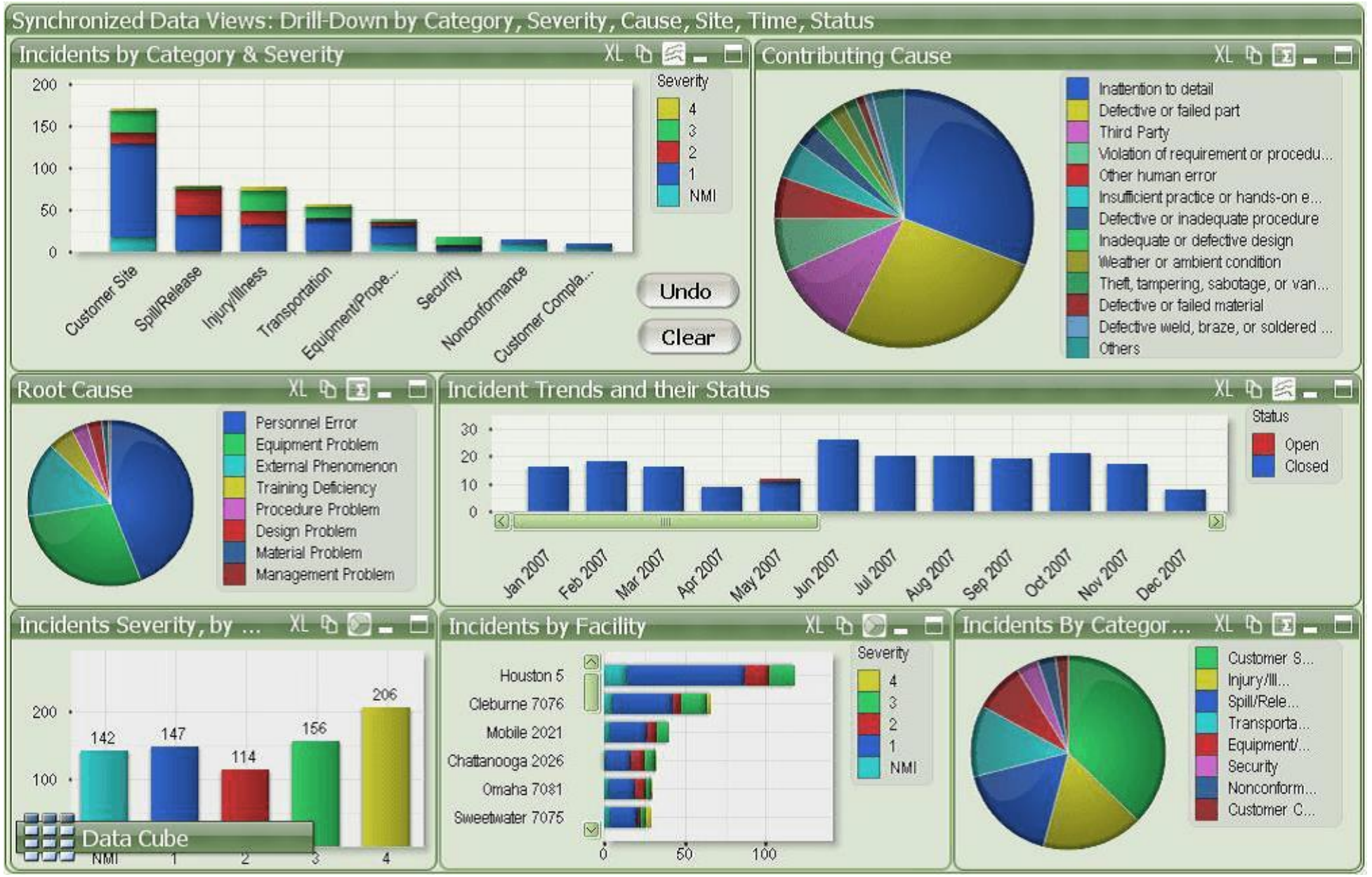
BANA College Shaffer

Ratings



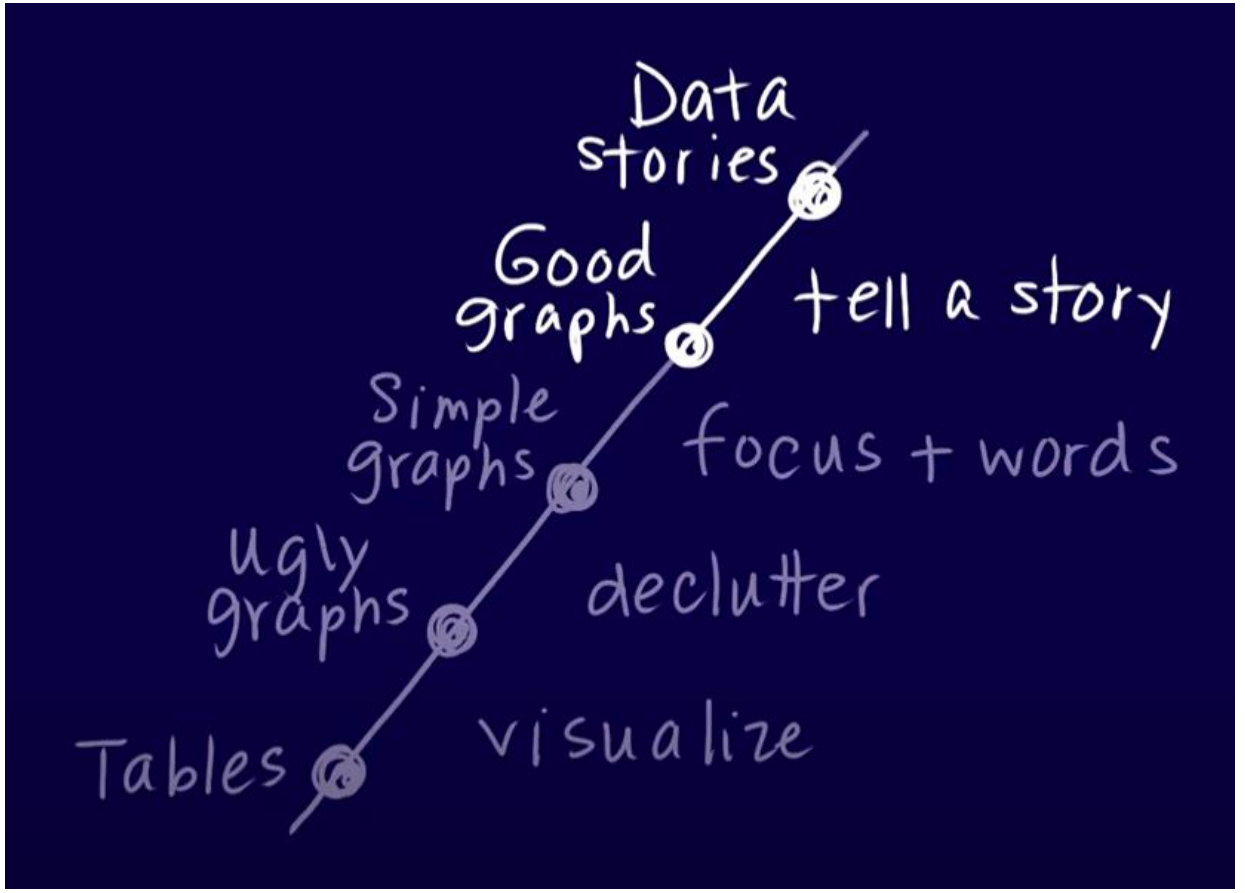
Course Metrics Dashboard created by Jeffrey A. Shaffer. Data from University of Cincinnati Course Evaluations. Blue indicates the 2 most recent rating periods.

[https://www.matillion.com/wp-content/uploads/2014/11/qlikview-poor-use-of-dashboard-software.png]



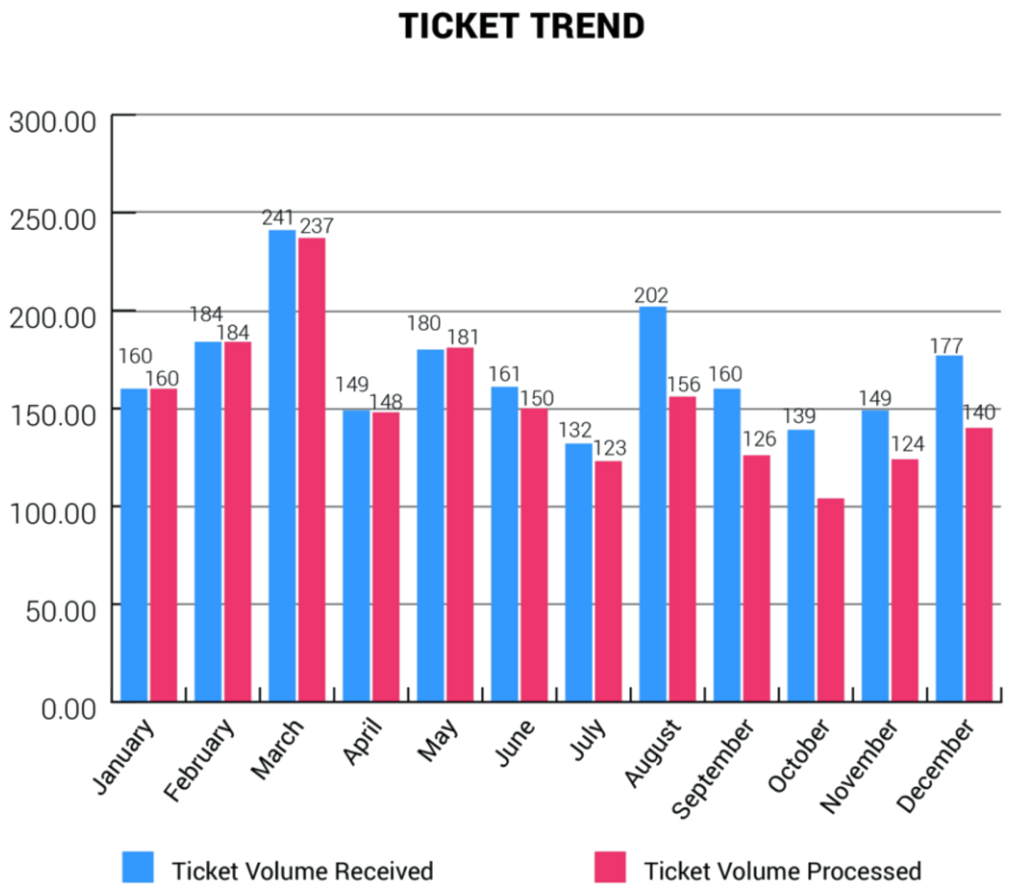
[C. Nussbaumer Knaflic, *Storytelling with Data*]

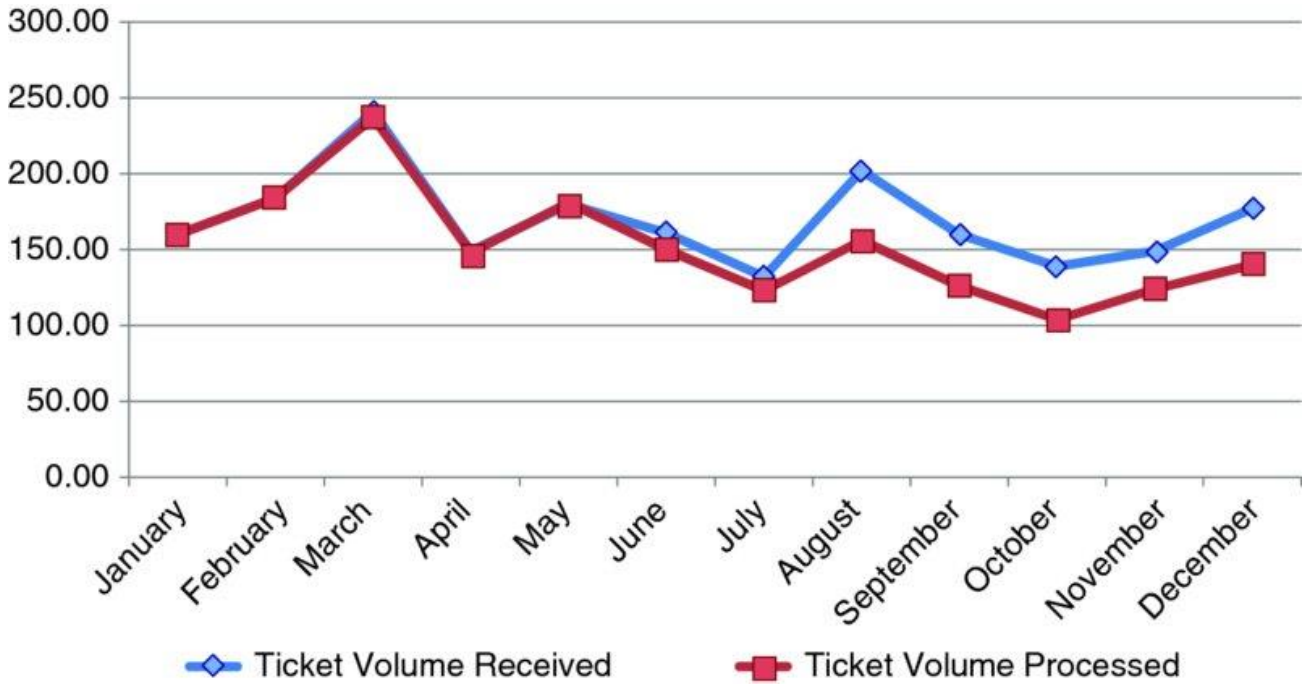
EVOLVING A VISUALIZATION



[C. Nussbaumer Knaflic, *Storytelling with Data*]

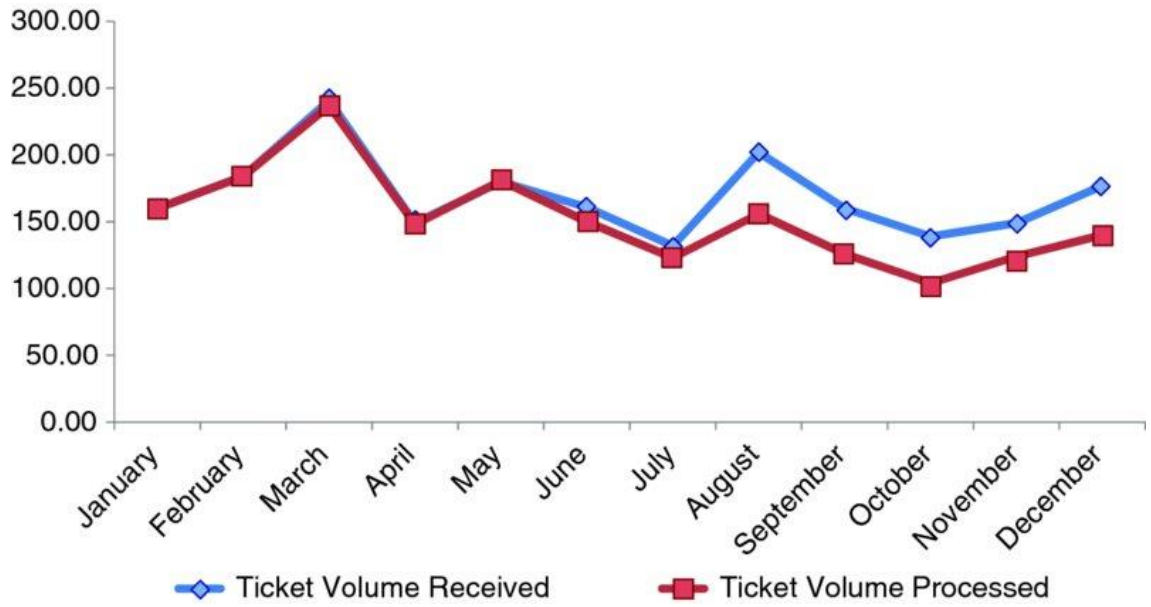
EXAMPLES



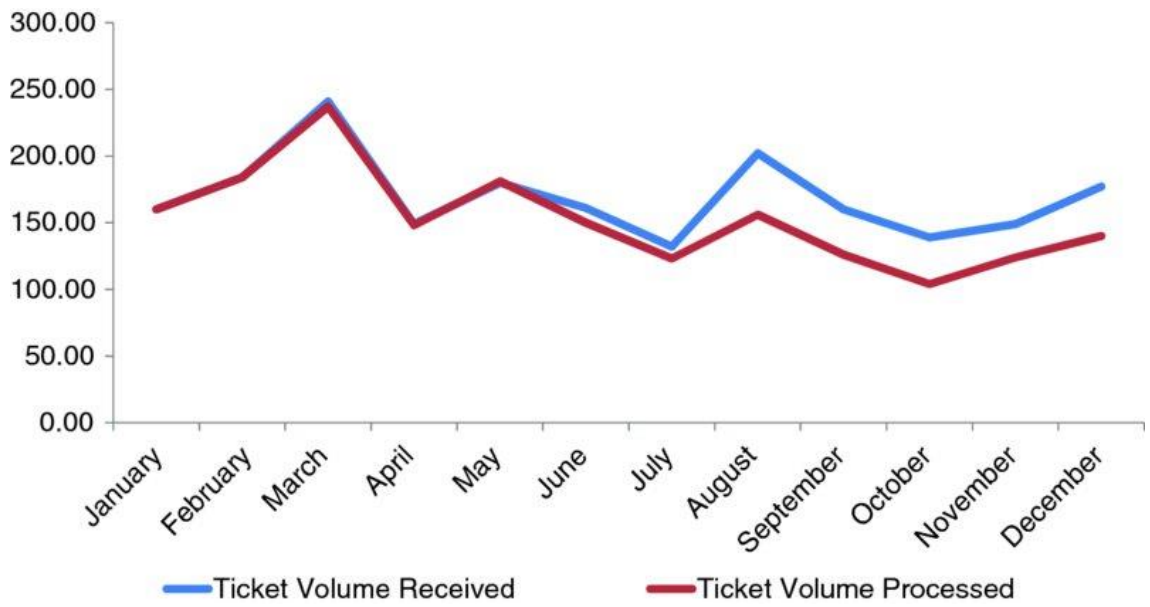


1. Change graph type (trope)

EXAMPLES

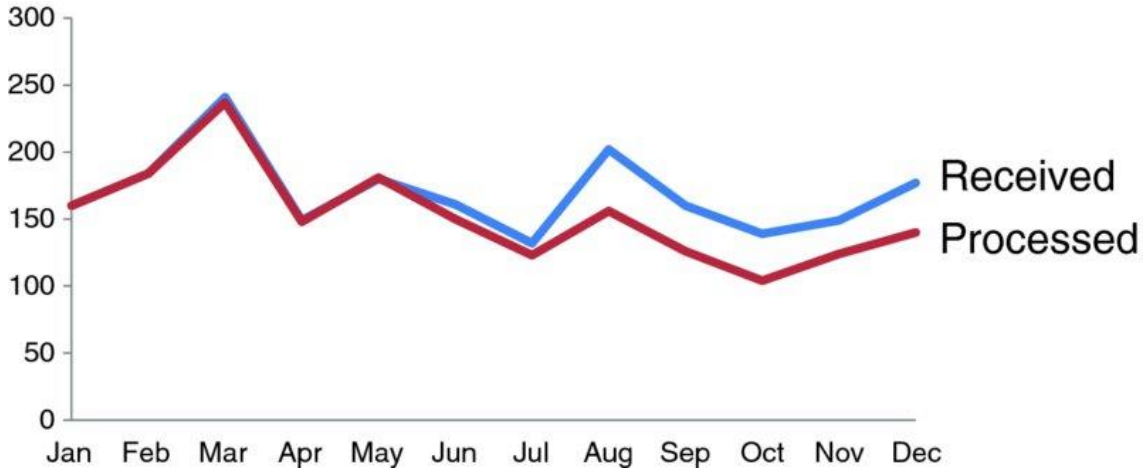


2. Remove border and gridlines

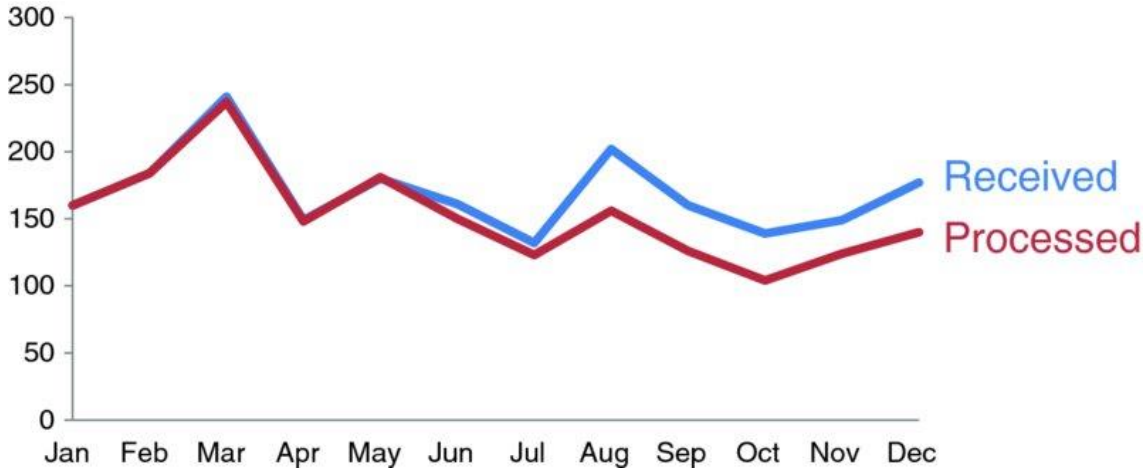


3. Remove markers

EXAMPLES



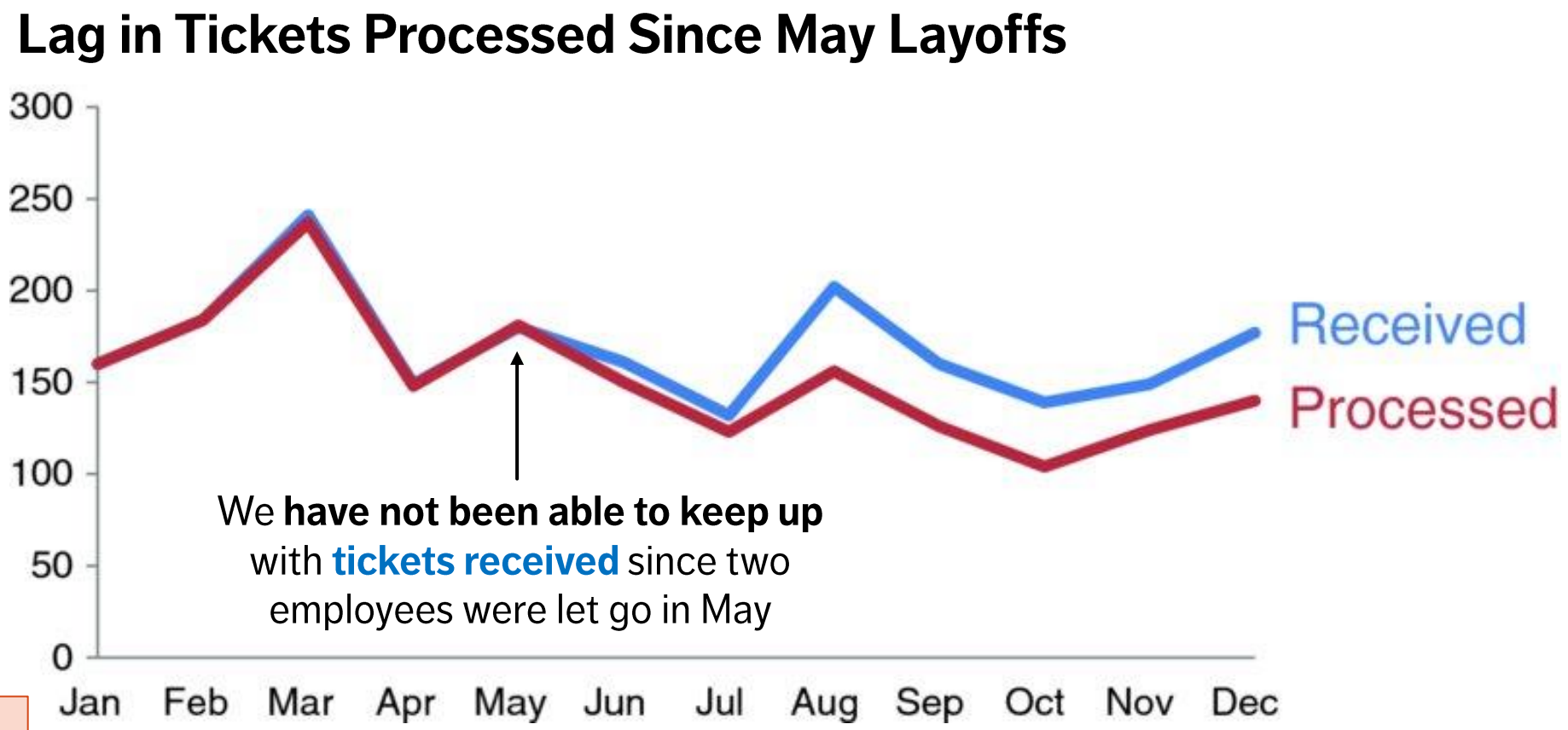
4. Clean-up axis labels and legend



5. Colour code the lines

[Based on C. Nussbaumer Knaflic, *Storytelling with Data*]

EXAMPLES



6. Tell the story

EXERCISE

Identify work scenarios for which data visualization could prove useful. What insight could be drawn from such visualizations?

Would such visualizations get a buy-in from your supervisors/employers?

How much work would be required to get from design to completion?

Are the obstacles mostly of a technical nature? Related to data procurement?

SUPPLEMENTAL MATERIAL

9. STORYTELLING AND VISUALIZATION

[Illustrations: V. Jaskina]

STORYTELLING UNIVERSALITY

There once was a shepherd boy who was bored as he sat on the hillside watching the village sheep. To amuse himself he took a great breath and sang out, "Wolf! Wolf! The Wolf is chasing the sheep!"

The villagers came running up the hill to help the boy drive the wolf away. But when they arrived at the top of the hill, they found no wolf. The boy laughed at the sight of their angry faces. "Don't cry 'wolf', shepherd boy," said the villagers, "when there's no wolf!" They went grumbling back down the hill.

Later, the boy sang out again, "Wolf! Wolf! The wolf is chasing the sheep!" To his naughty delight, he watched the villagers run up the hill to help him drive the wolf away.

When the villagers saw no wolf they sternly said, "Save your frightened song for when there is really something wrong! Don't cry 'wolf' when there is NO wolf!"

[Illustrations: V. Jaskina]

STORYTELLING UNIVERSALITY

But the boy just grinned and watched them go grumbling down the hill once more.

Later, he saw a REAL wolf prowling about his flock. Alarmed, he leaped to his feet and sang out as loudly as he could, "Wolf! Wolf!" But the villagers thought he was trying to fool them again, and so they didn't come.

At sunset, everyone wondered why the shepherd boy hadn't returned to the village with their sheep. They went up the hill to find the boy. They found him weeping.

"There really was a wolf here! The flock has scattered! I cried out, "Wolf!" Why didn't you come?"

An old man tried to comfort the boy as they walked back to the village. "We'll help you look for the lost sheep in the morning," he said, putting his arm around the youth, "Nobody believes a liar ... **even when they are telling the truth/so don't get caught telling the same lie twice**."

[based on an idea from S. McCloud, *Making Comics*]

WORDS AND IMAGES

A picture is worth a thousand words (vs. a picture is worth 1000 words).

Words bring an unparalleled level of **specificity**. There is no image so vague that words cannot lock it into a **desired meaning**.

Some concepts and names can only be clearly expressed **through words**.

[S. McCloud, *Making Comics*]

"Look, it's Kelly Donovan, twin brother of the Xander actor on *Buffy the Vampire Slayer*, plus Humphrey Bogart wearing a Freddy Mercury mask, and a robot duplicate of former U.N. Secretary General Boutros Boutros-Ghali!"

LINE CHART/RUG CHART

Gaps in the number line: **absence** of those numeric values in the data.

Remember: this is (possibly) different from the order that values appear in the dataset – since it is a number line, it shows where the values fall numerically.

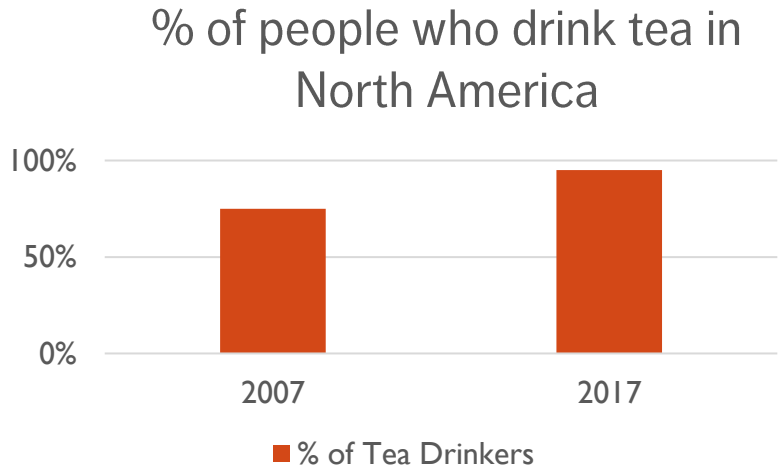
If some values are identical, they lie on top of each other (use **jitter**?).

SIMPLE TEXT

One or two numbers to focus on.

Good at “setting the scene”.

Draws focus to an area of the report.



95% of the population

drinks tea today compared to

75% in 2007

TABLES

Tables interact with our **verbal** system, which means we **read** them:

- used to compare values
- audiences will look for their rows

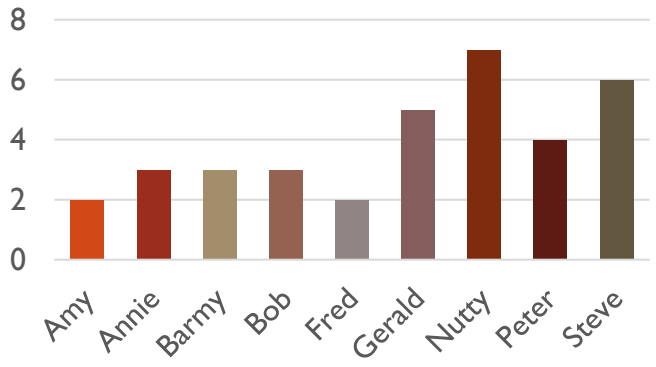
Table design needs to **blend** into background

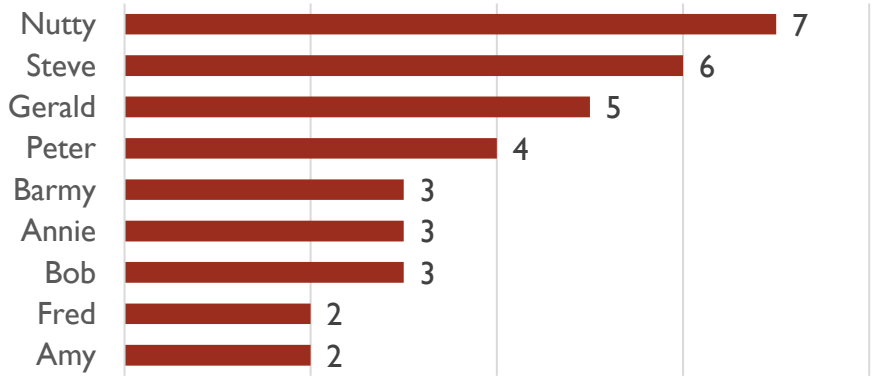
- the data should stand out, not the borders
- dense table/data: use alternating row colour

Name	Last Year	This Year
Bob	20	30
Fred	30	40
George	10	15

Name	Last Year	This Year
Bob	20	30
Fred	30	40
George	10	15

BAR CHARTS





Very versatile and useful.

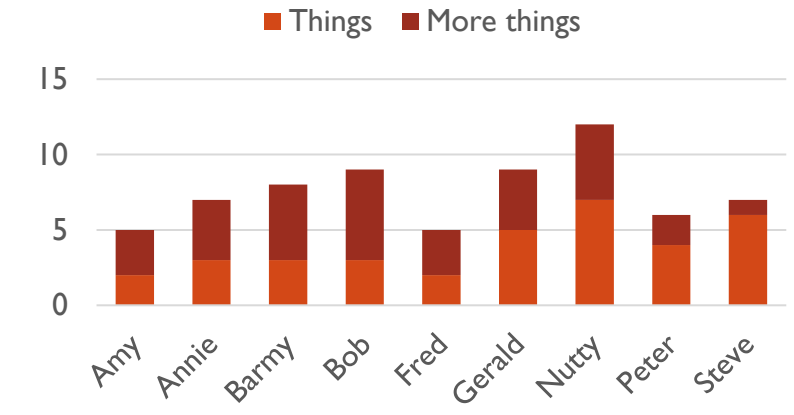
ALWAYS (?) have a zero baseline.

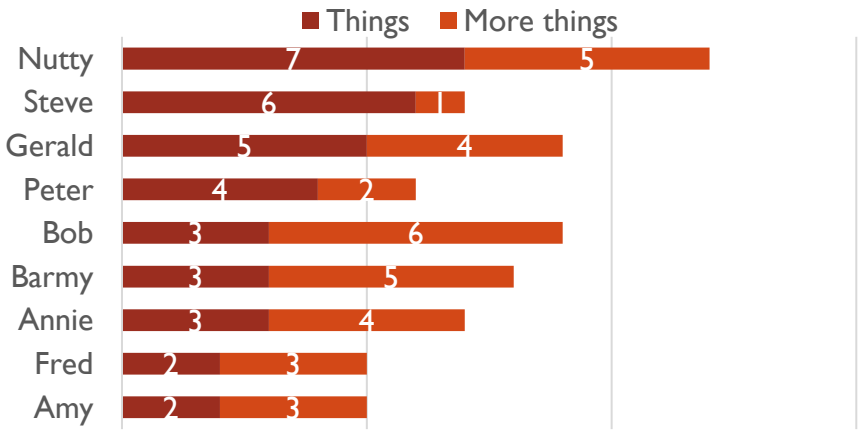
Use graph axis OR data labels. Axis for broad statements, data labels for more detail.

Horizontal charts are apparently **easier to read** (according to many studies).

Think about the ordering of categories.

STACKED BAR CHARTS





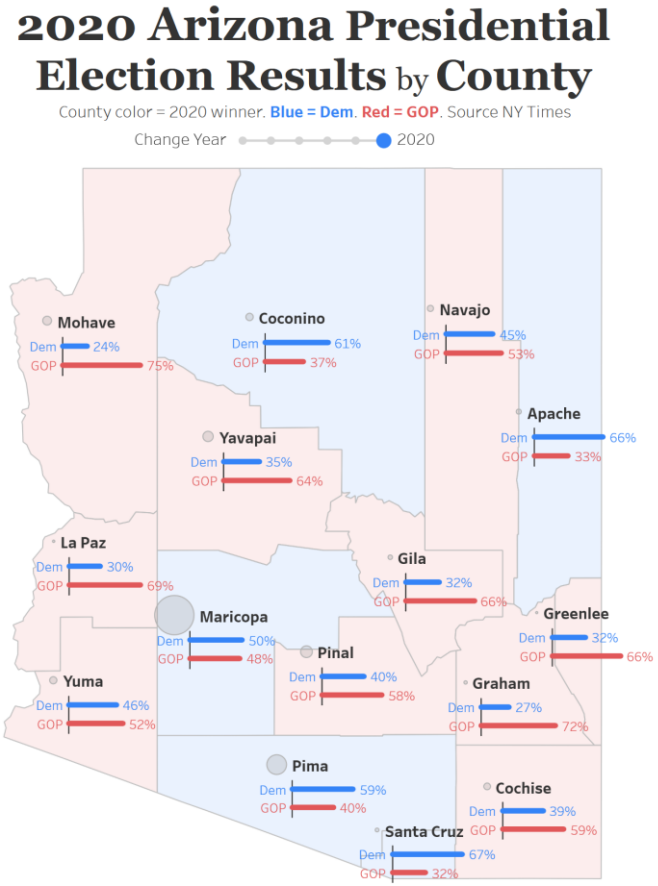
Designed for **comparing totals**, but can quickly become **overwhelming**.

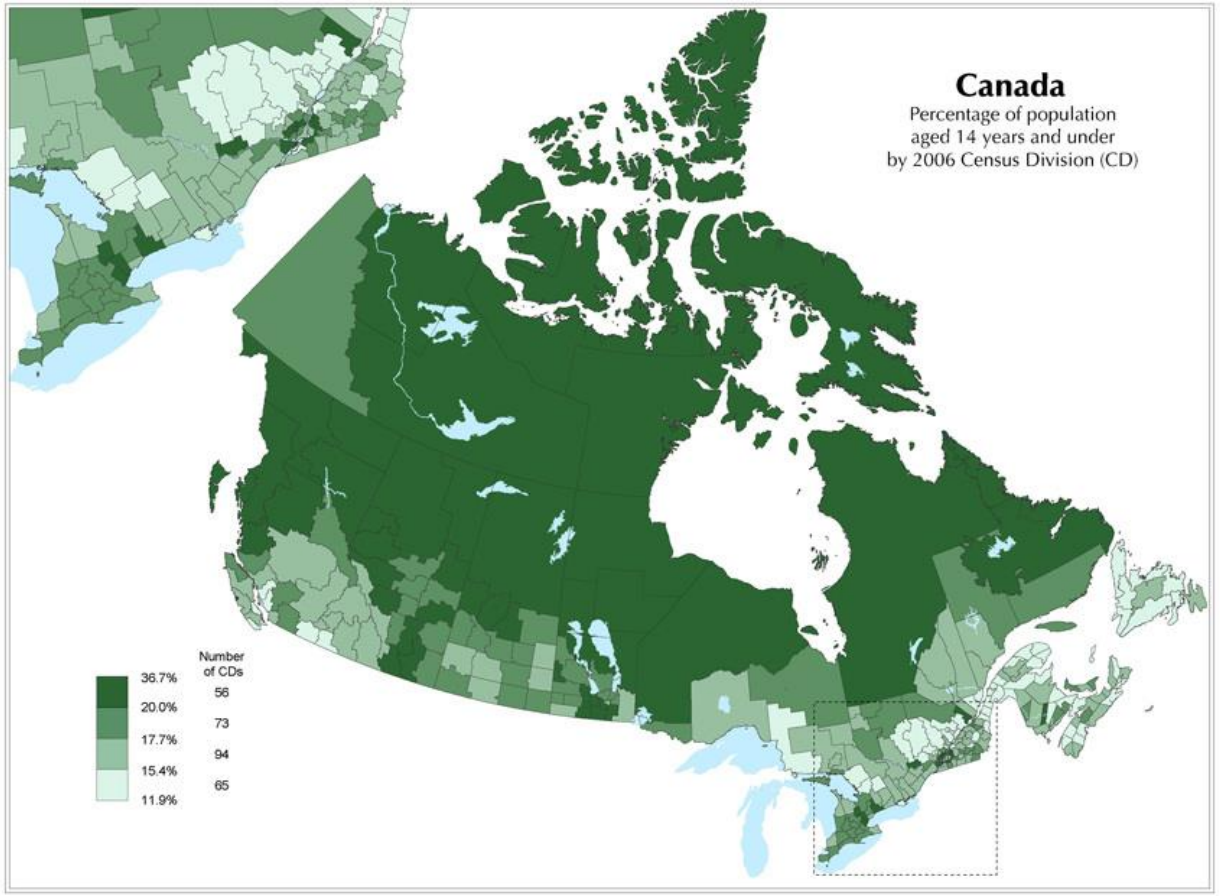
Hard to sort / order.

Filtering is complicated in Power BI (what do you click on & how the chart responds when filter is clicked on?)

[left: A.E. McCann, right: author unknown]

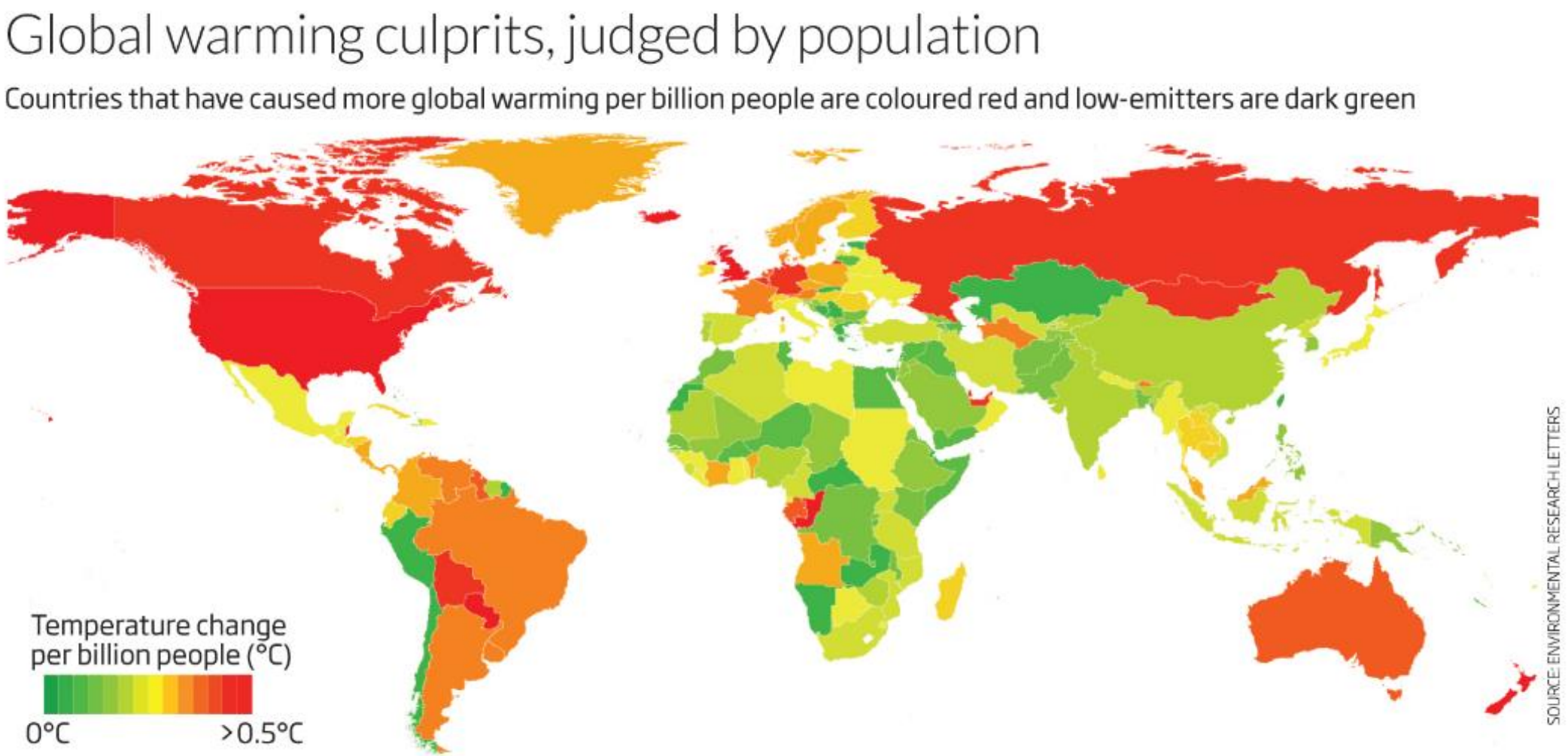
HEAT MAPS (CHOROPLETHS)





[New Scientist, Jan 2014]

GEOGRAPHICAL MAPS



[New Scientist, Jan 2014]

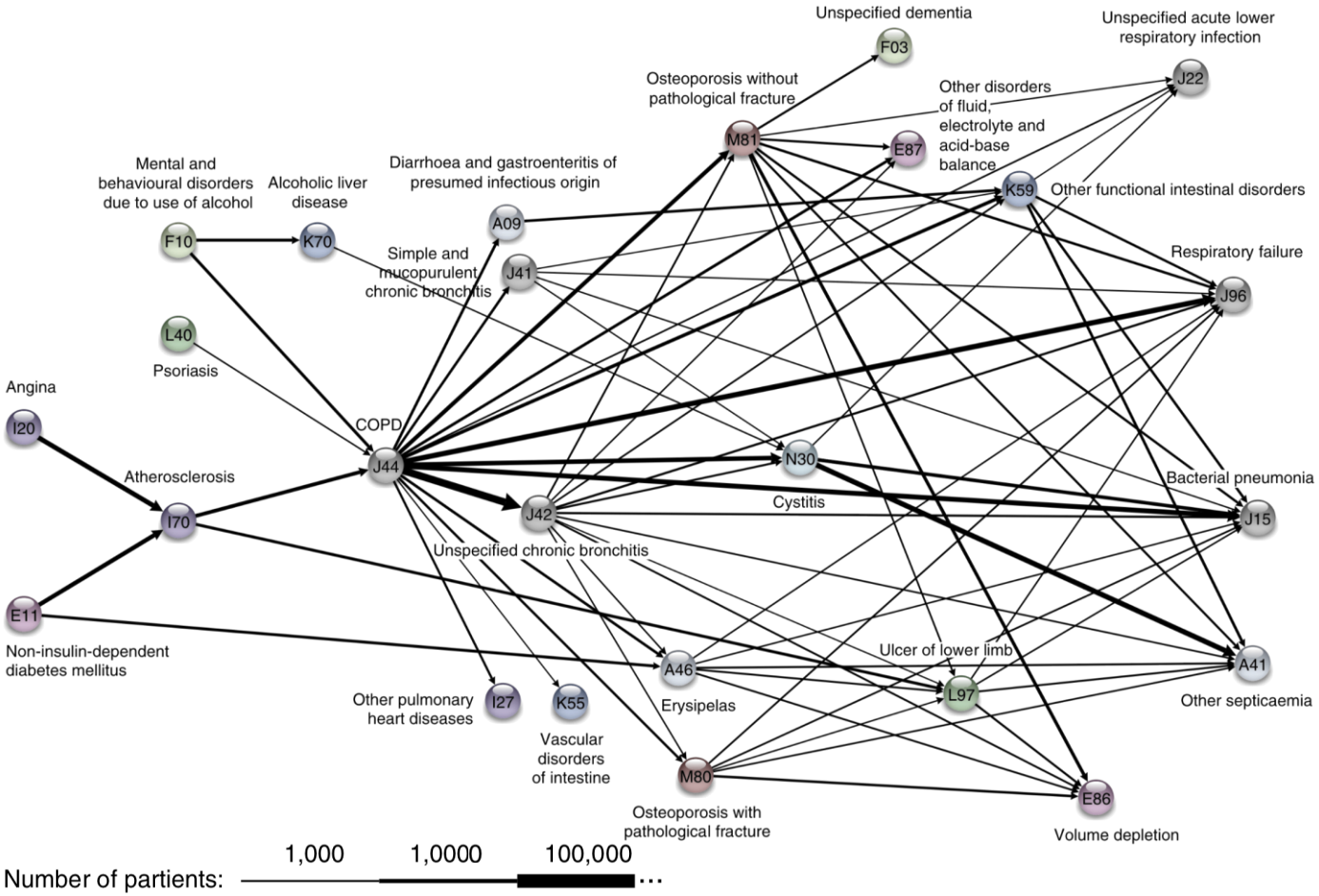
DISTORTED GEOGRAPHICAL MAPS

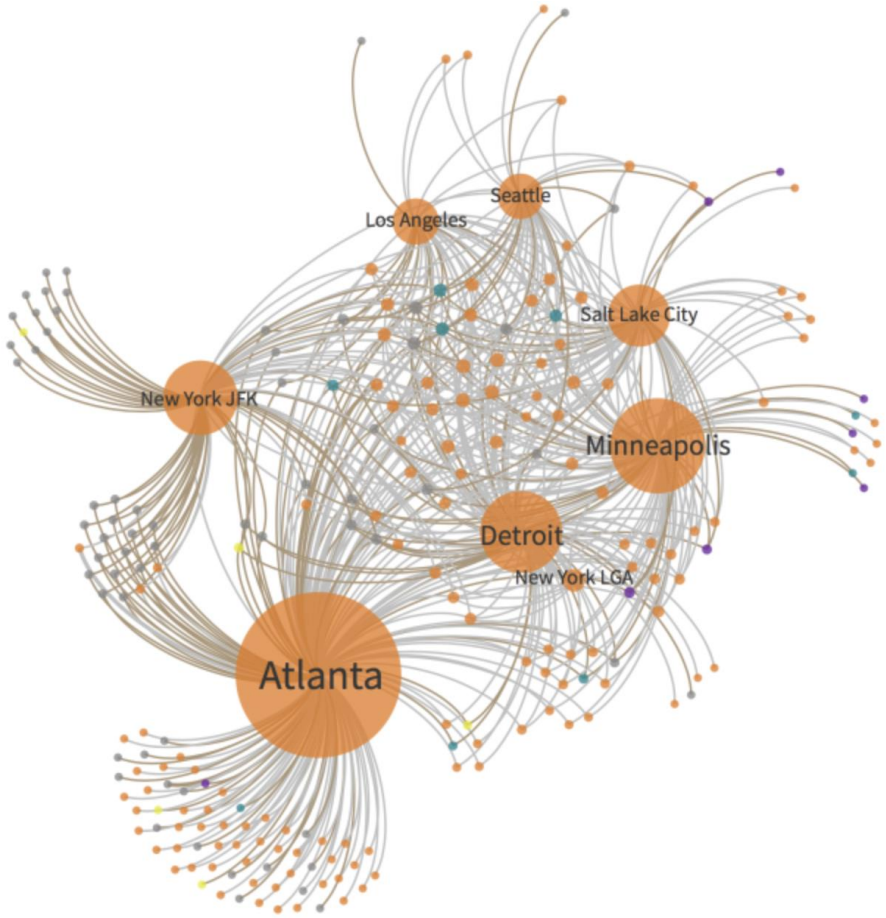
[A.B. Jensen et al., *Nature Communications*]

[Author unknown]

NETWORK DIAGRAMS

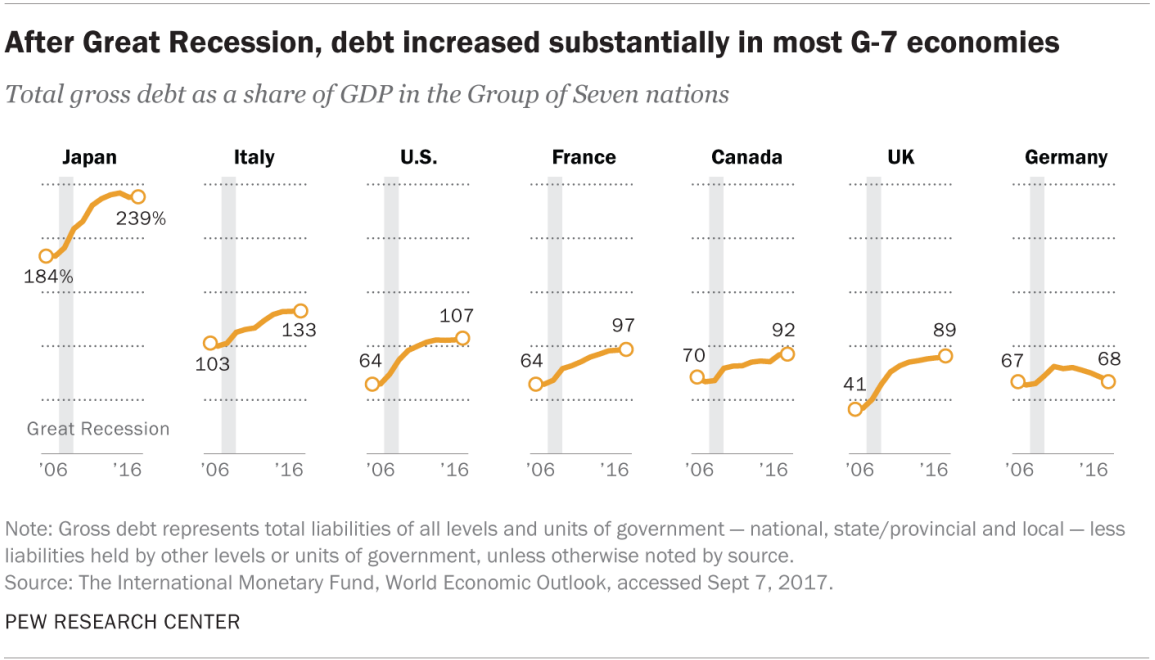
Diagnosis path around COPD in Danish population

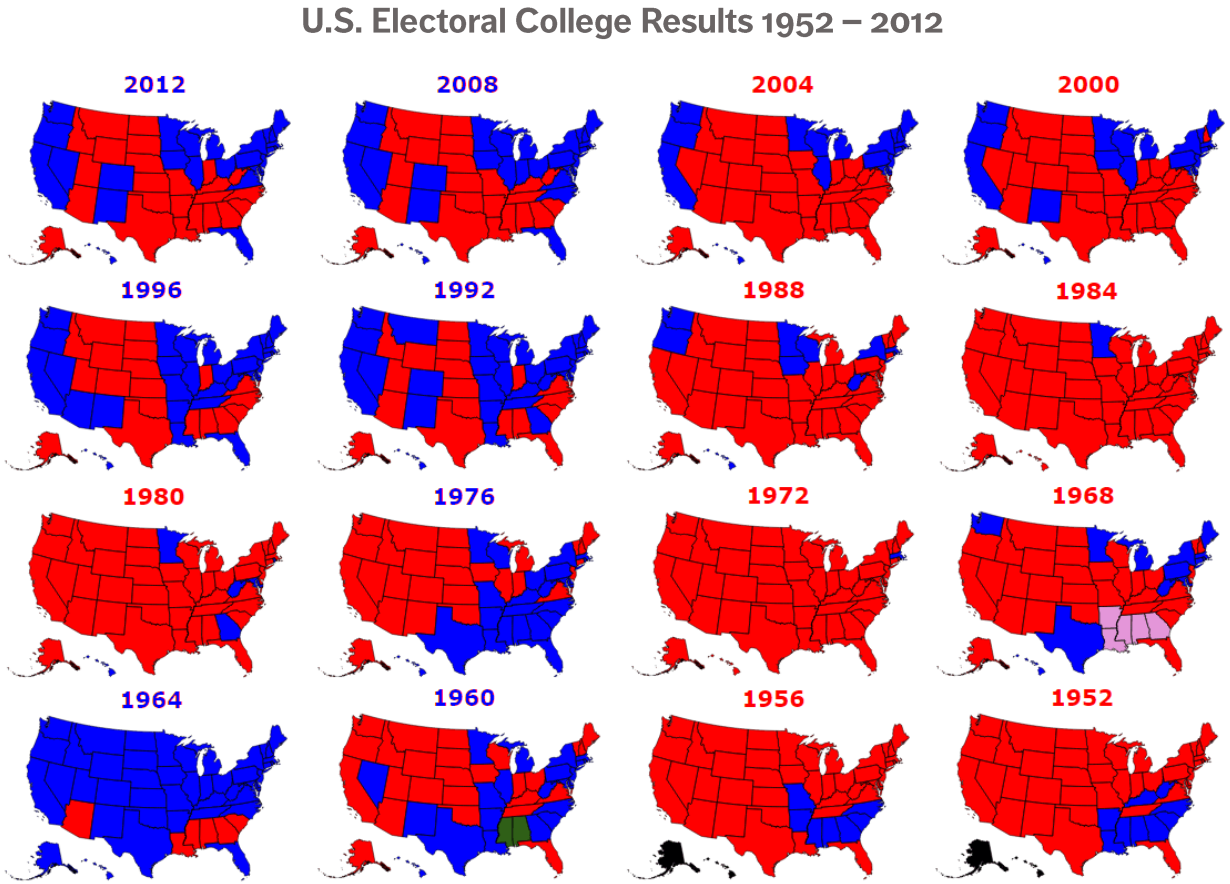




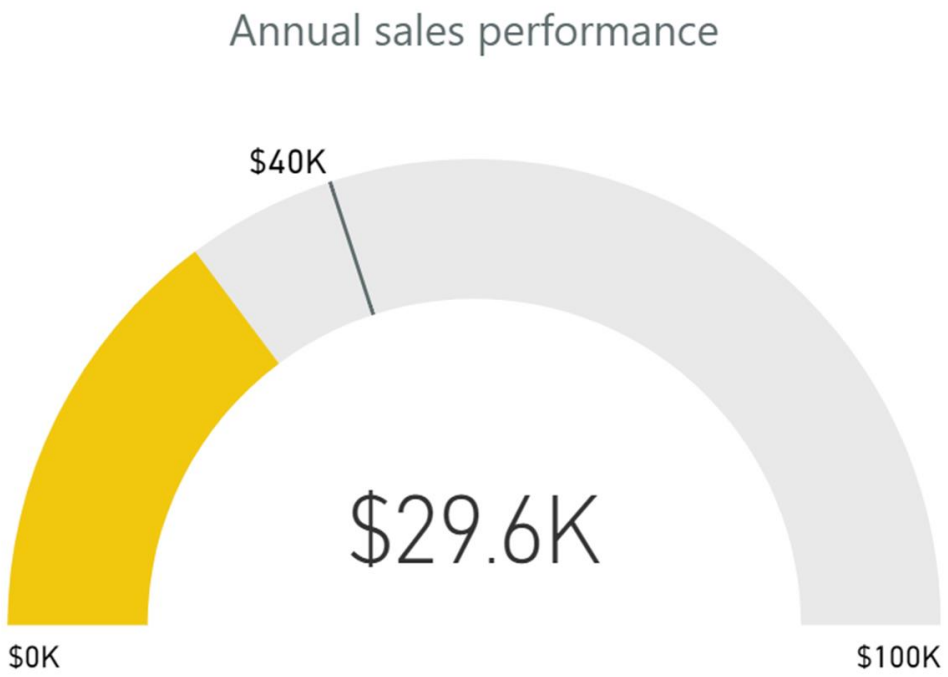
data-action-lab.com

SMALL MULTIPLES





GAUGE CHARTS



Often used as a dashboard component (with or without needle).

Displays single value measures towards goal / KPI.

Great to show progress (a bit of a management fad, though...)

Displays information that can be quickly **scanned** and **understood**.

INTERACTIVE & ANIMATED VISUALIZATIONS

"There is always a danger that if certain types of visualization techniques take over, the kinds of questions that are particularly well-suited to providing data for these techniques will come to dominate the landscape, which will then affect data collection techniques, data availability, future interest, and so on." (P. Boily)

Even when done well, 85% of users don't bother with interactive viz (NY Times).

Take-Away: explore the data and try different methods

COLOUR SCHEMES

Achromatic (colourless, using only blacks, whites and grays)

Monochromatic (1-colour schemes)

Complementary (colours directly across from each other on the colour wheel)

Split complementary (2 of the 3 colors are adjacent; 1 of the colours is opposite)

COLOUR TIPS

When it comes to colour, **less is more**: use it sparingly (graphic designers are taught to “get it right, in black and white”).

Based on the Gestalt Principles, **monochrome** schemes can be effective.

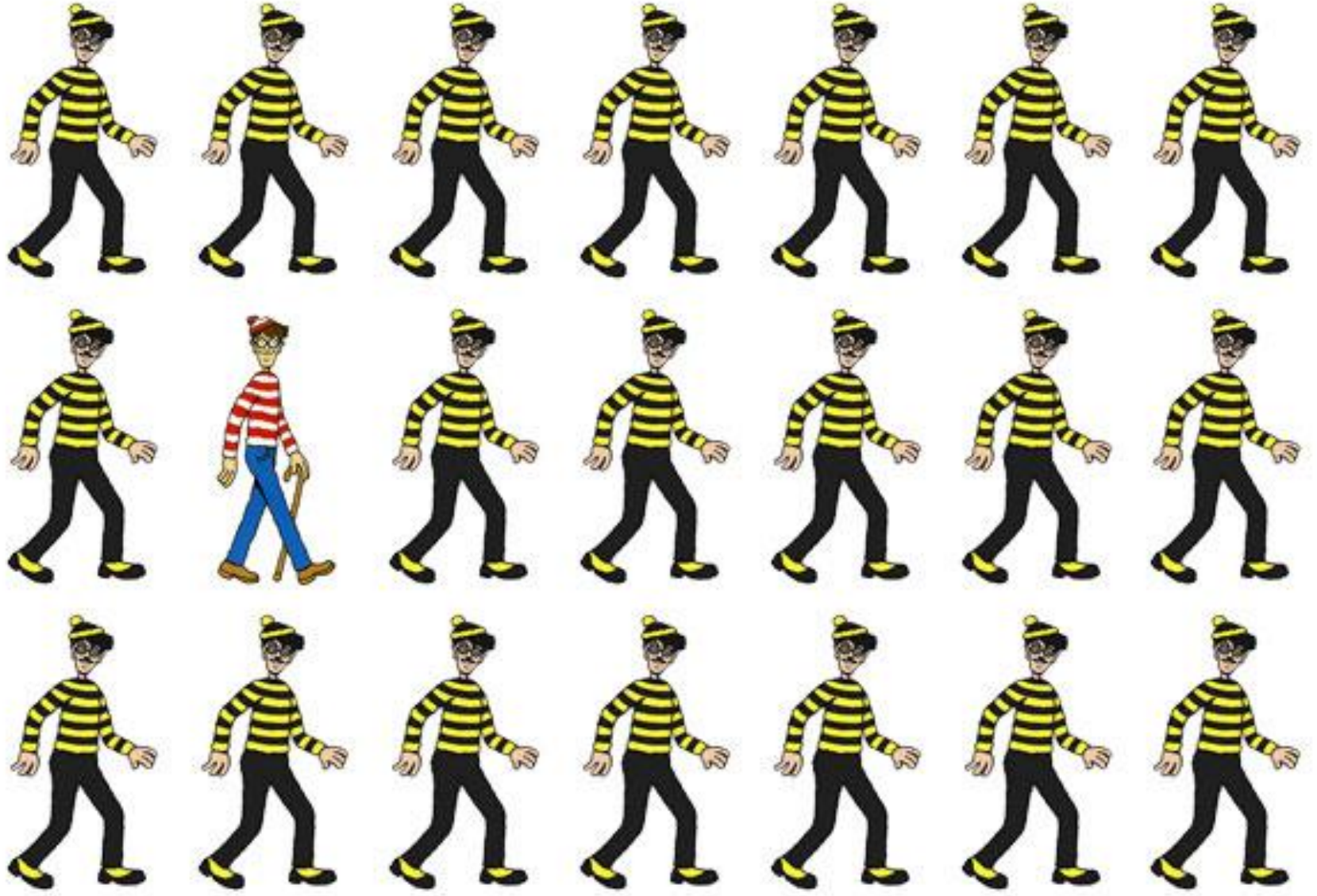
When appropriate, pick corporate identity scheme (this maximizes buy in).

Create a **template** (and stick to it).

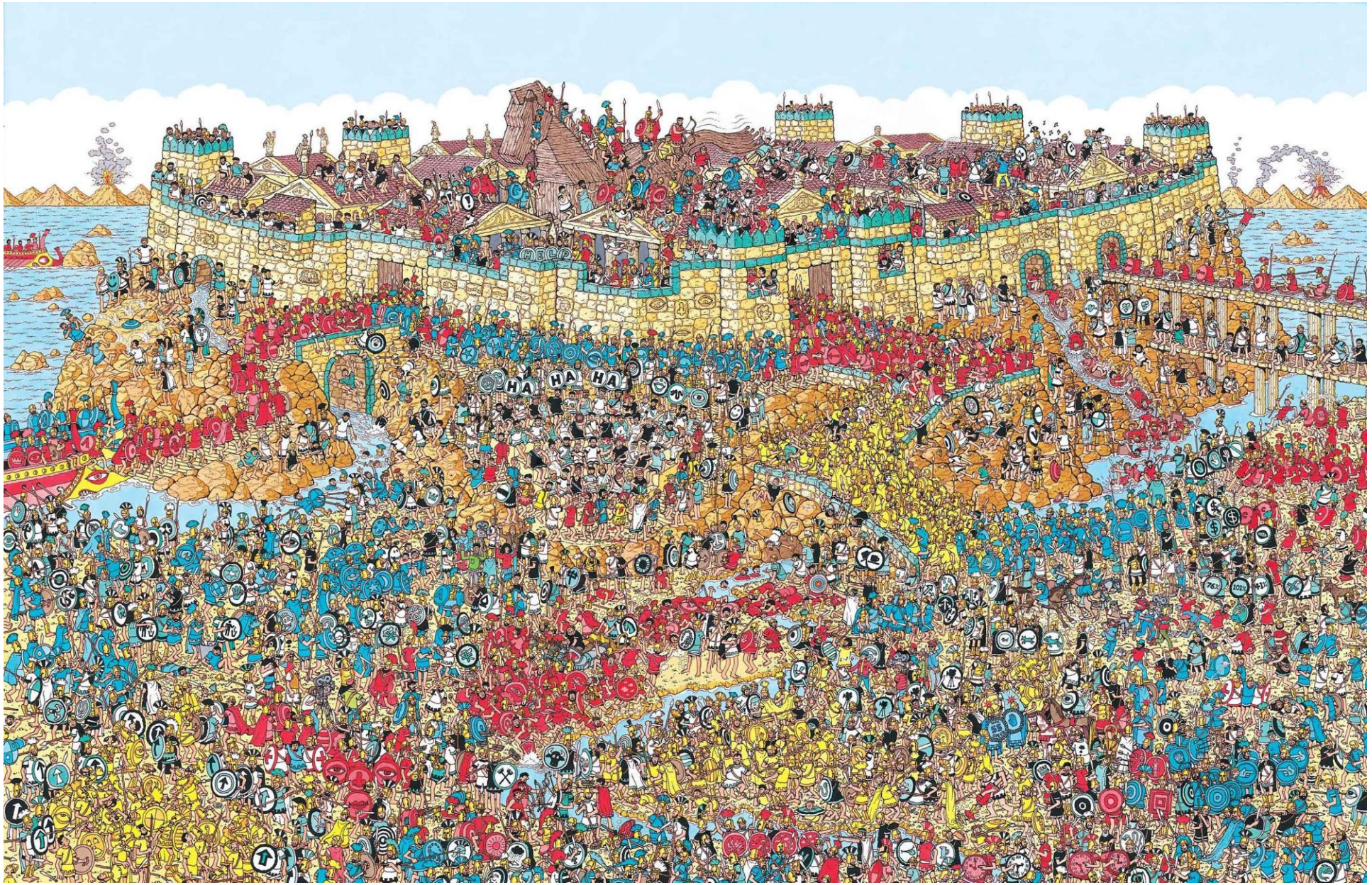
Upload images to see what charts look like for flavours of colourblindness:

- https://www.color-blindness.com/coblis-color-blindness-simulator (not the only tool)

[L. Groeger, A Big Article About Wee Things, ProPublica.com]



[L. Groeger, A Big Article About Wee Things, ProPublica.com]



A WORD ABOUT ACCESSIBILITY

A table can be translated to Braille, but that's not always possible for charts.

Describing the features and emerging structures in a visualization is a possible solution... **if they can be spotted**.

Analysts must produce clear and meaningful visualizations, but they must also describe their features in a fashion that allows all to "see" the insights.

But this requires them to have "seen" all the insights, which is not always necessarily the case (if at all possible).

A WORD ABOUT ACCESSIBILITY

Data Perception:

- texture-based representations
- text-to-speech
- sound/music
- odor-based representations (?)
- taste-based representations (?!?)

Sonifications:

- TRAPPIST Sounds : TRAPPIST-1 Planetary System Translated Directly Into Music
- Listening to data from the Large Hadron Collider, L. Asquith

DATA STORY TROPES

Some visualizations patterns are so familiar that they become **tropes:**

- a scatterplot with a trend line going straight up or straight down
- a cluster bar chart with two categories where one is always lower than the other
- a line chart with the two lines crossing in one place
- pie charts being used all over the place (to avoid)
- red for republican, blue for democrat (US); red for left-leaning, blue for right-leaning (ROW)
- using broken axes to exaggerate effects (sometimes justified…)
- etc.

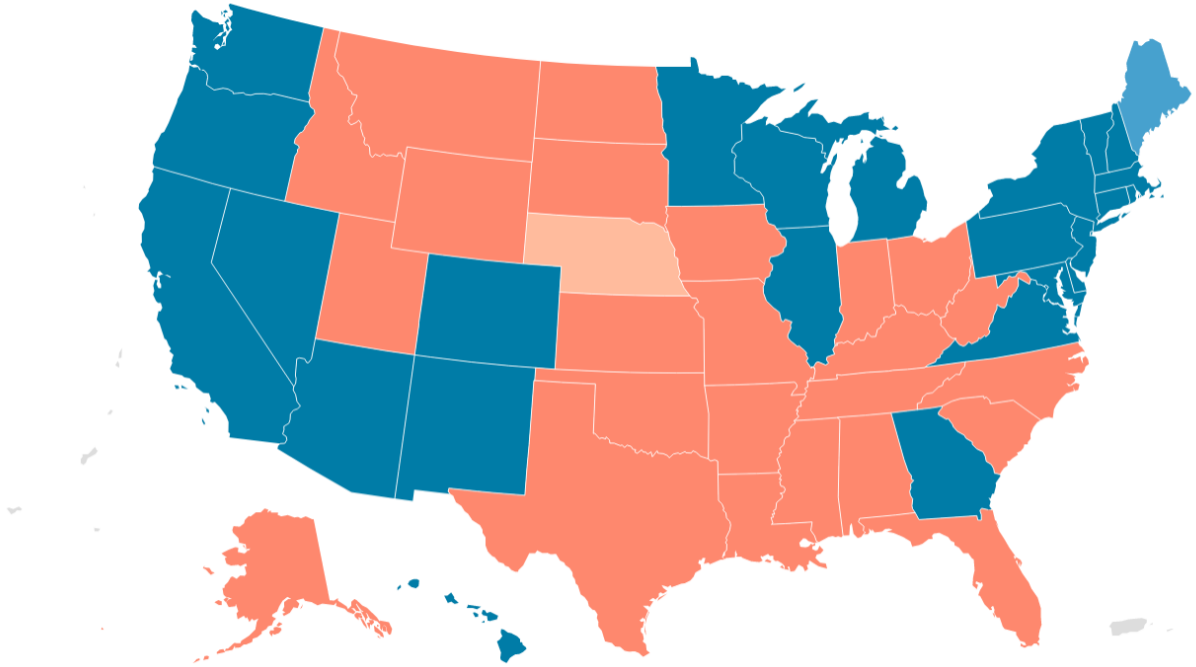
[https://handsondataviz.org/spatial-bias.html]

DATA STORYTELLING TROPES – EXAMPLES

Conventional Map of 2020 US Presidential Election Results

Maine and Nebraska allow some electoral votes to be split by district

Biden | Biden + Trump | Trump | Trump + Biden

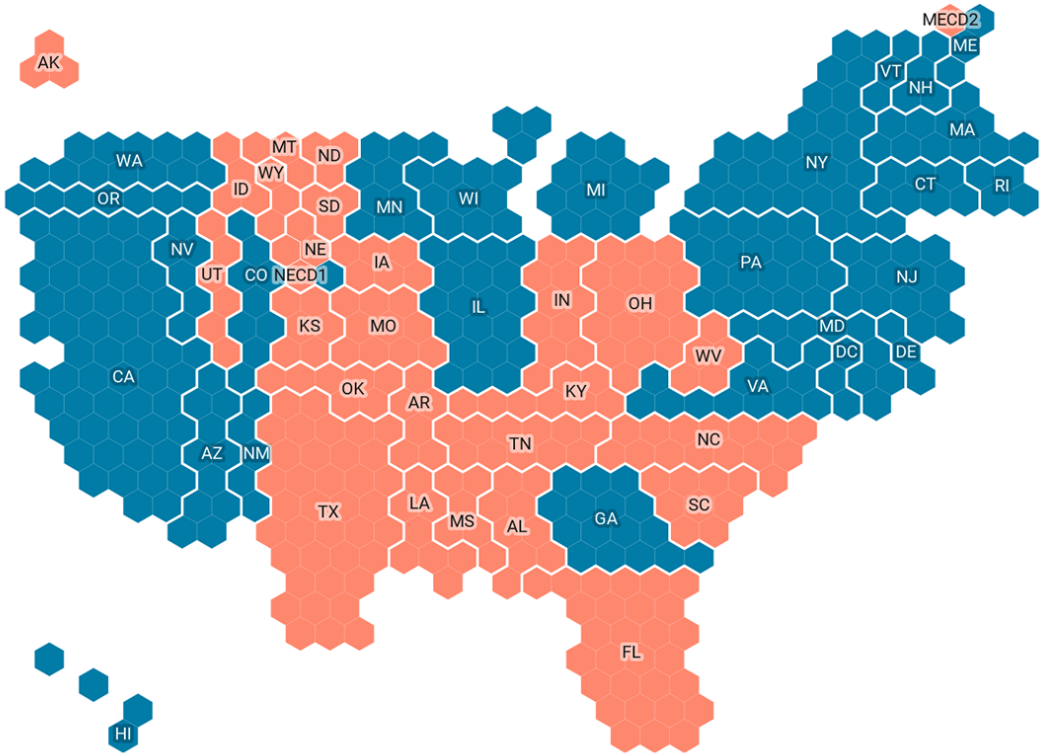


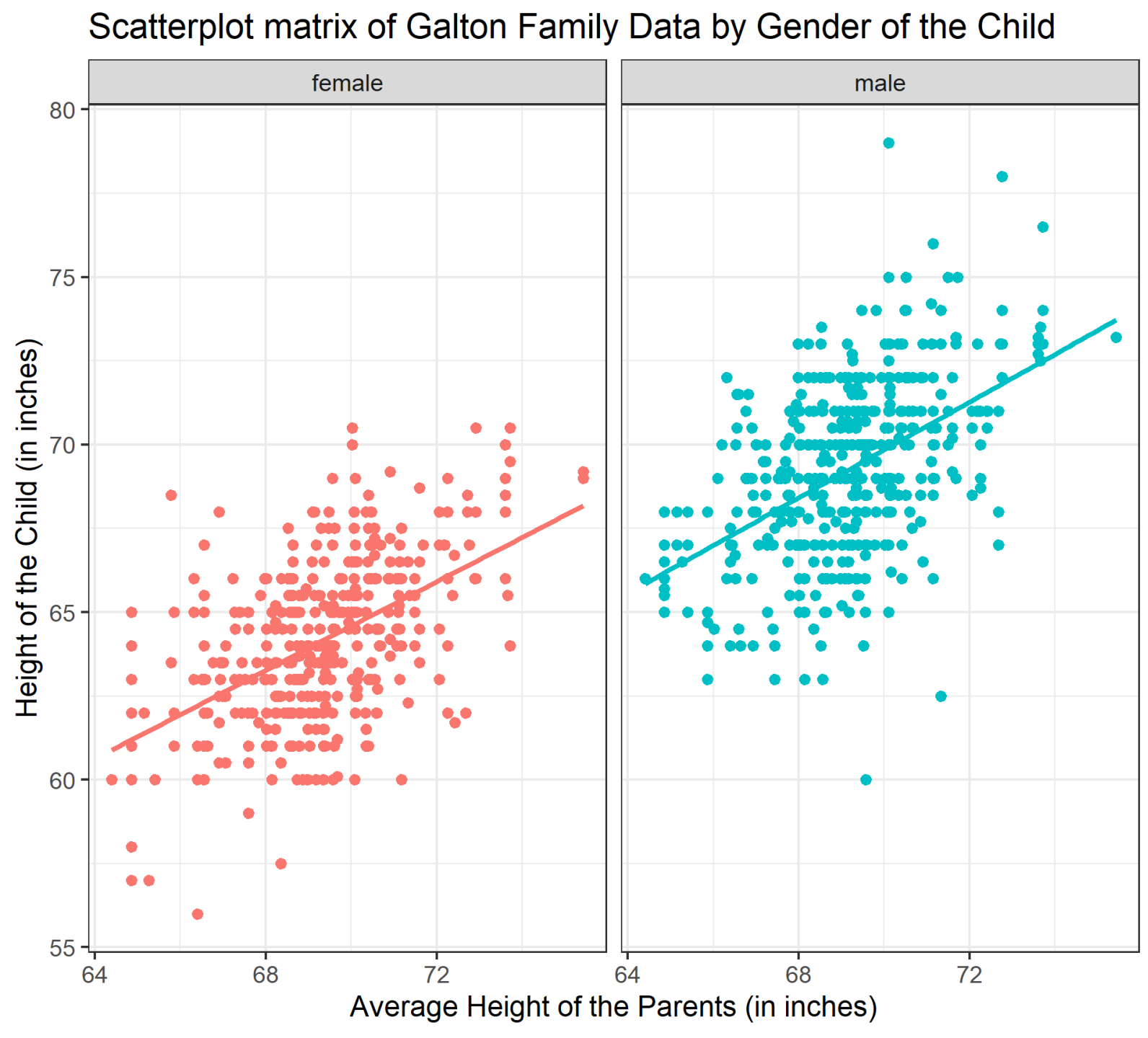
Created with Datawrapper

Cartogram of 2020 US Presidential Election Results

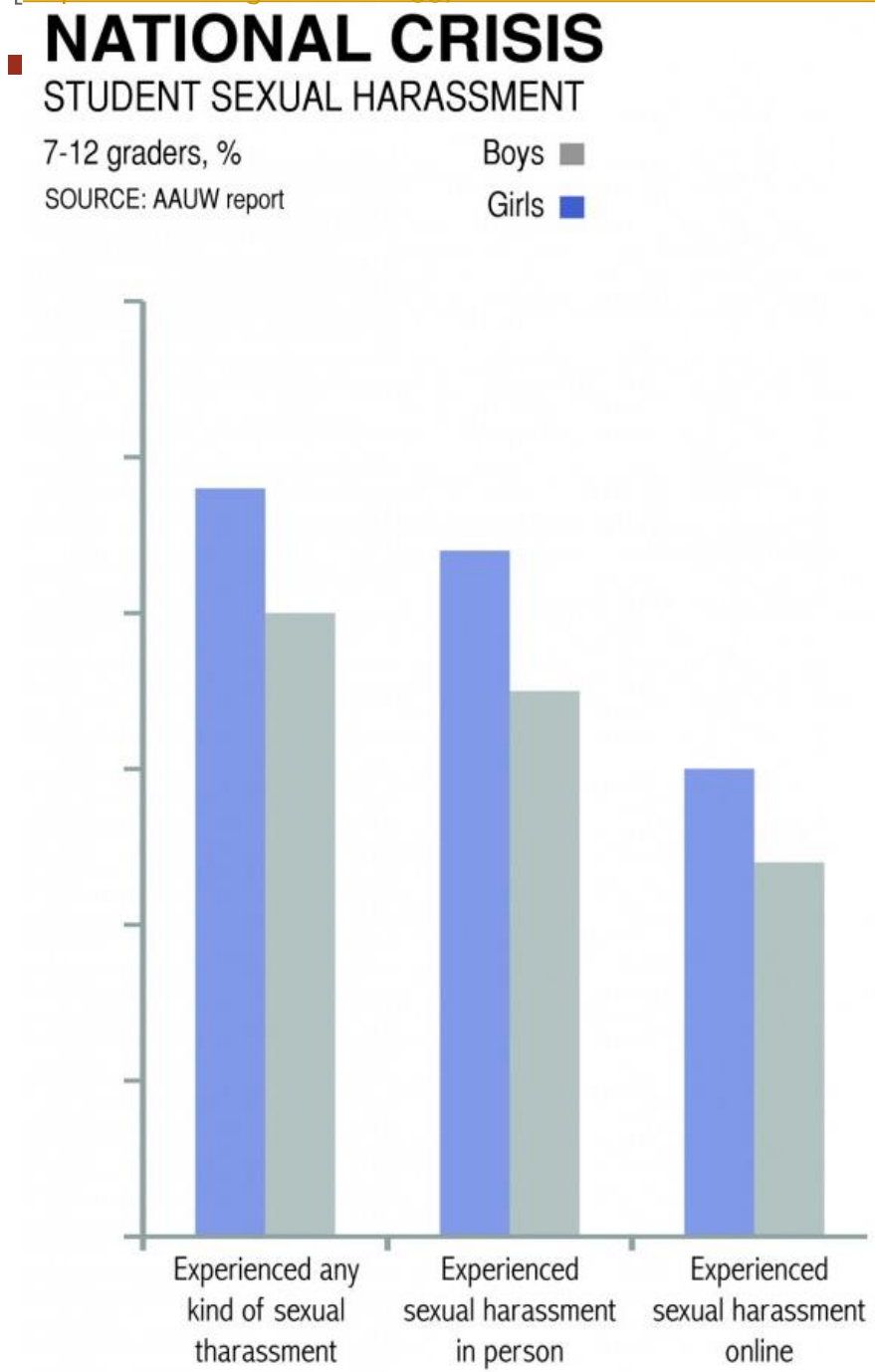
Each hexagon represents one electoral college vote

Biden | Trump

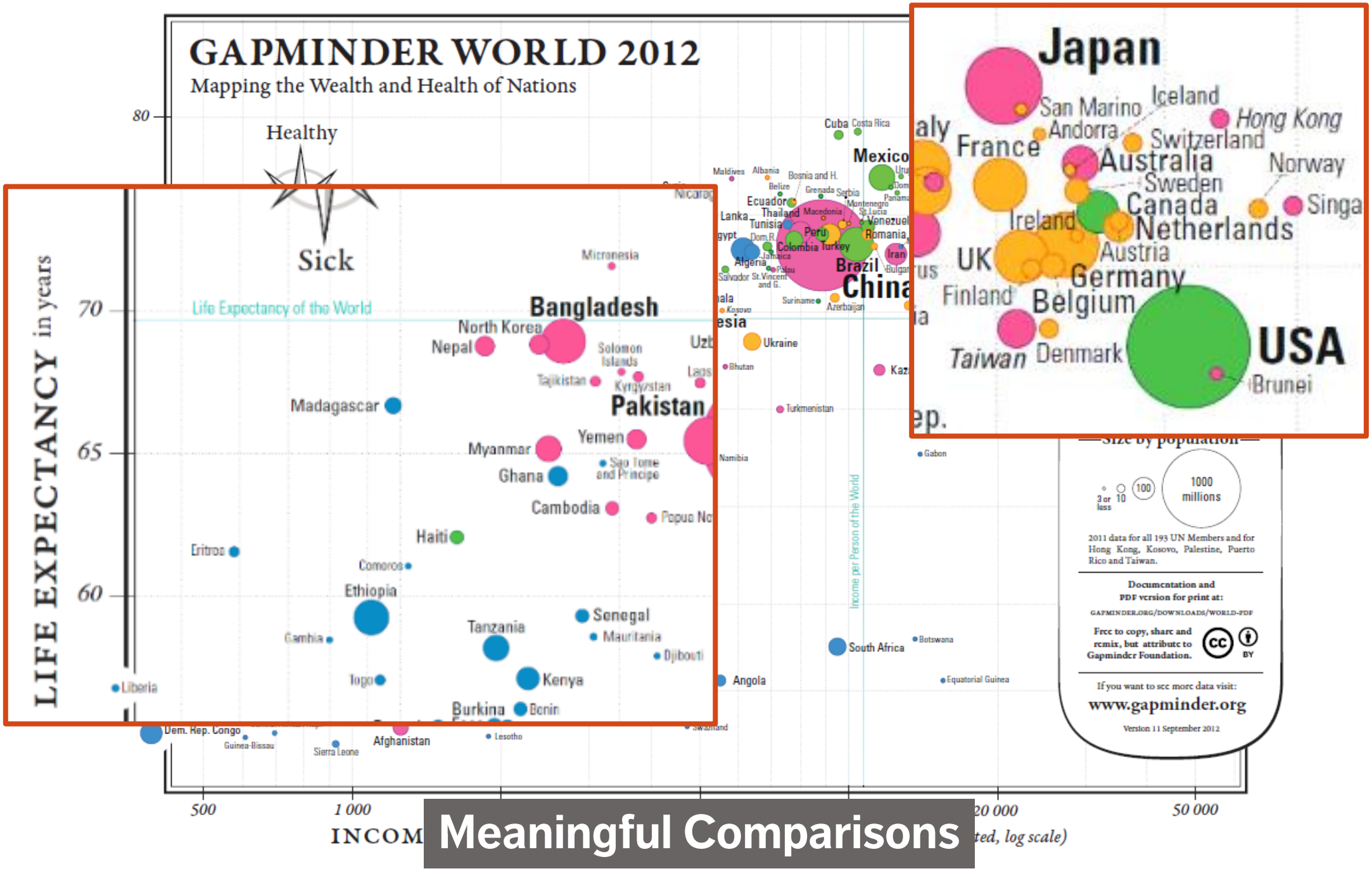




[https://www.chsglobe.com/13376/cover-stories/sexualharassment]

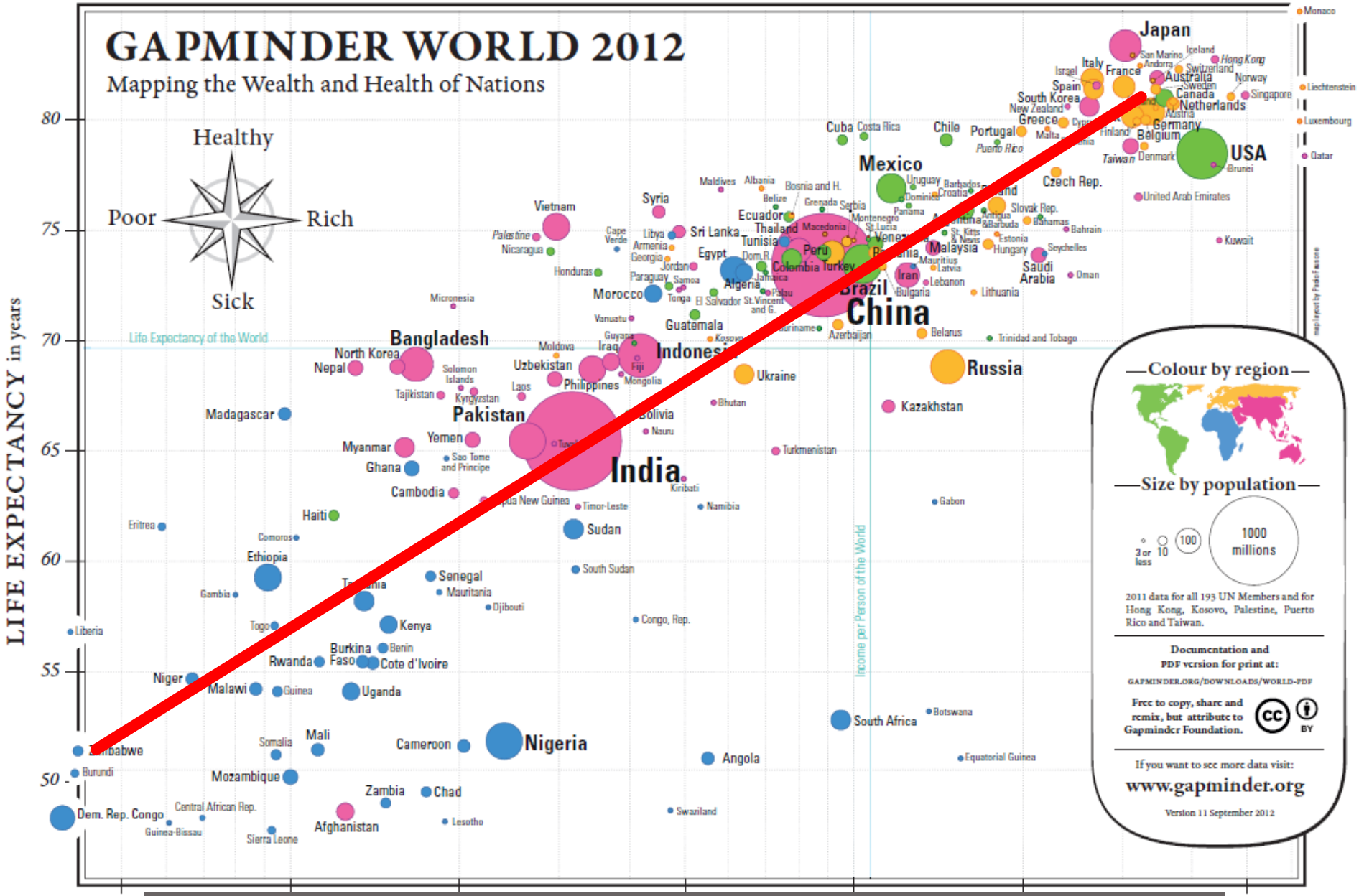


[Gapminder Foundation, https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen]



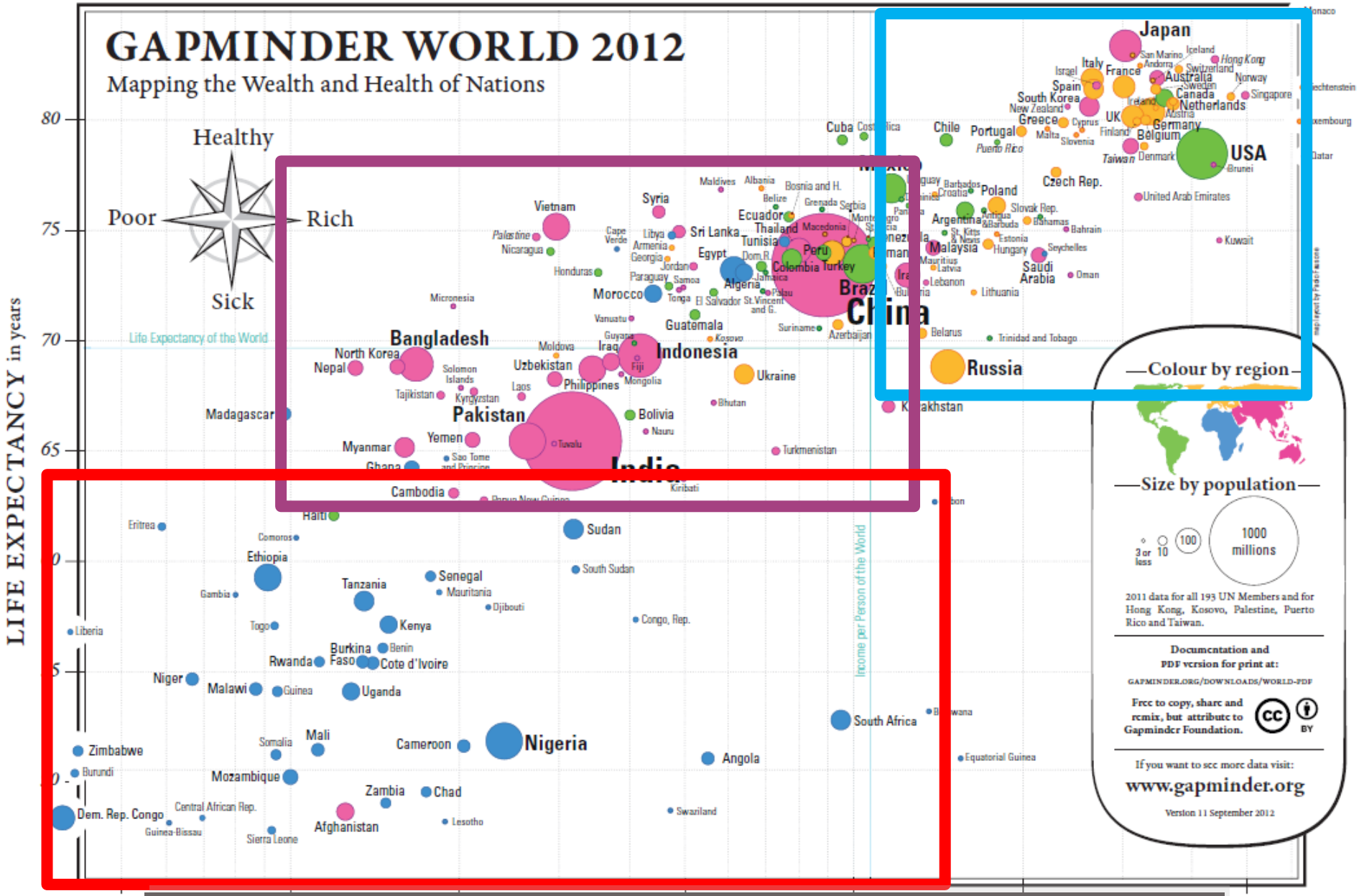
Meaningful Comparisons

[Gapminder Foundation, https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen]



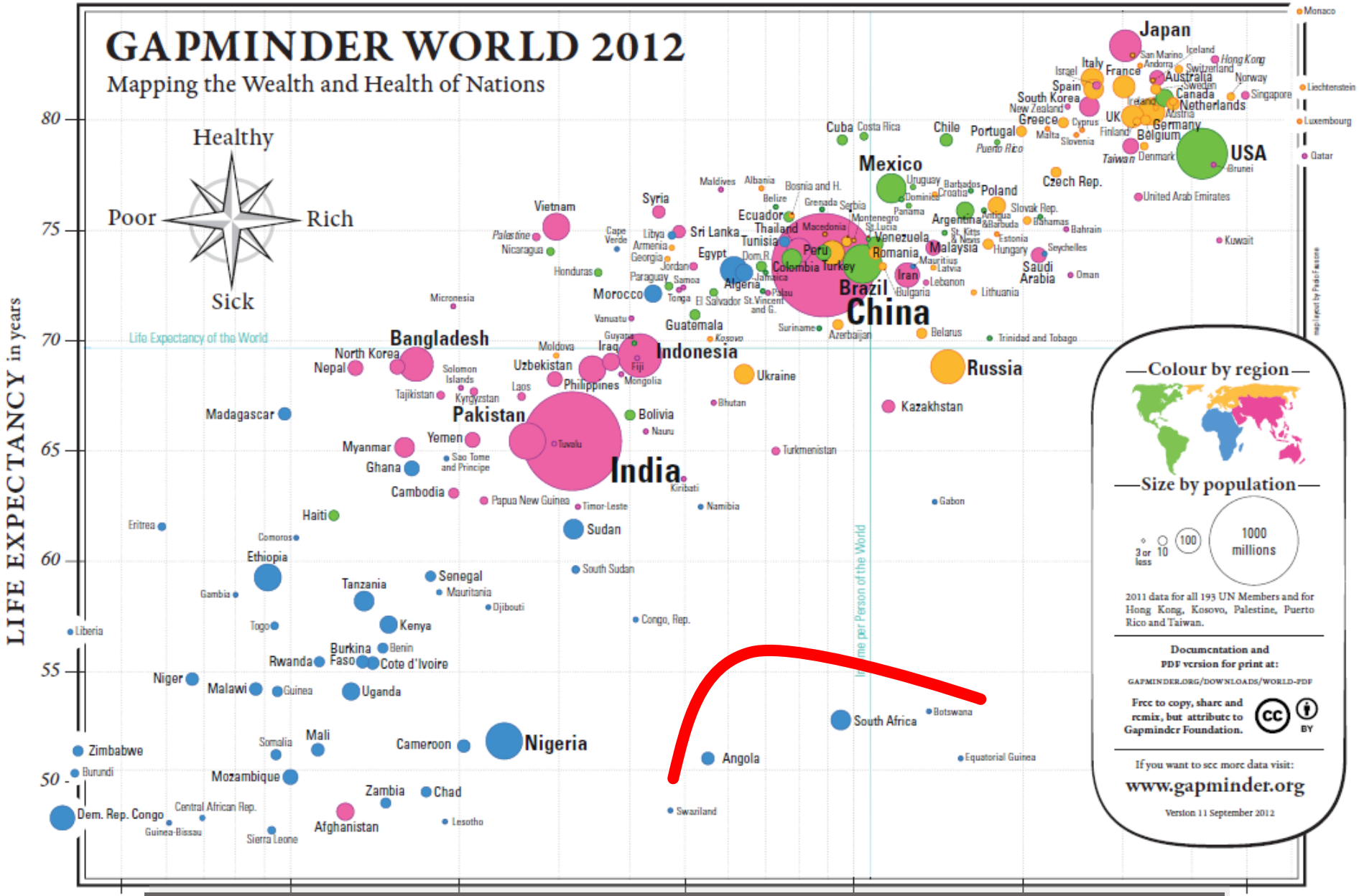
Underlying Structure and Multivariate Links

[Gapminder Foundation, https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen]



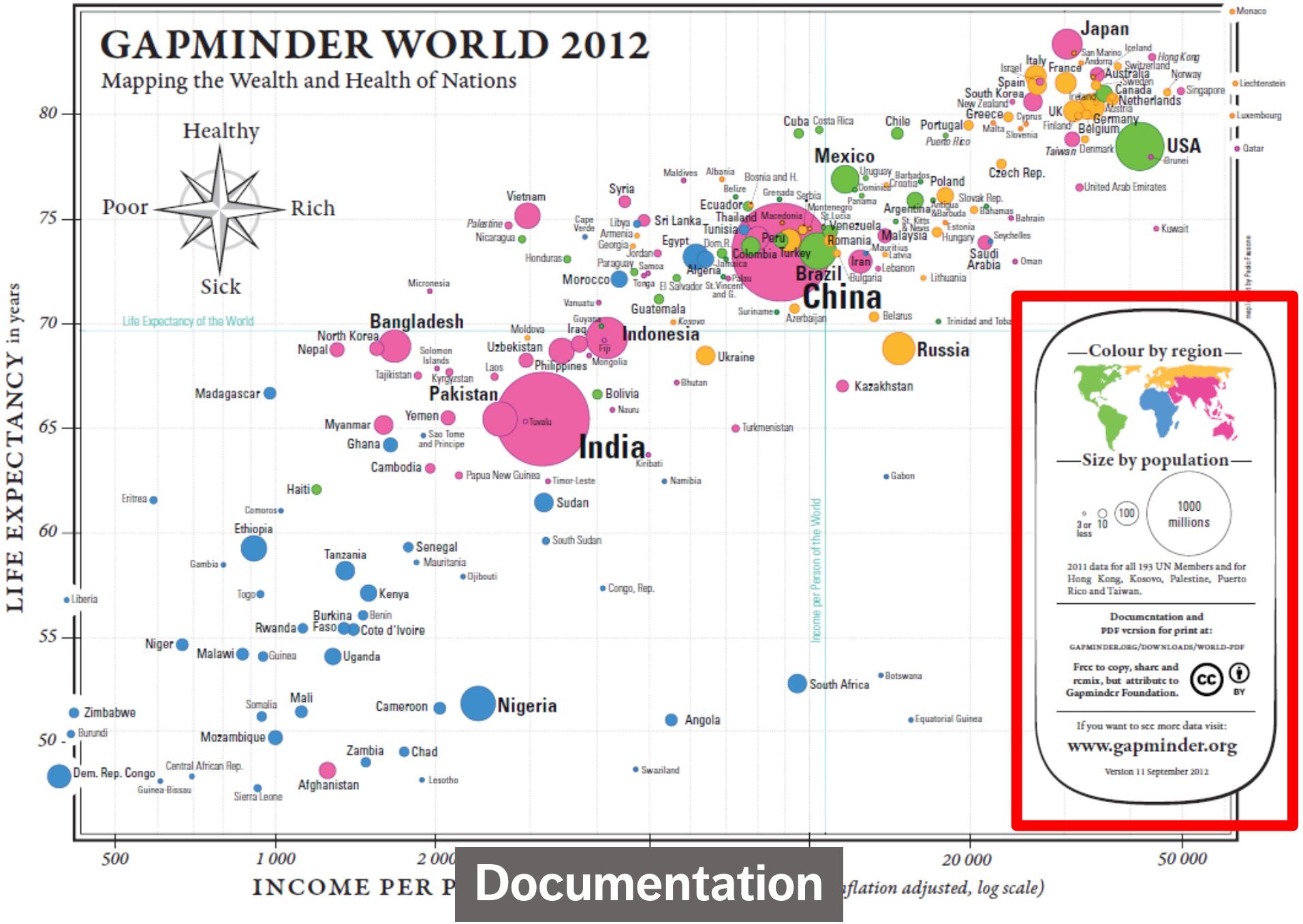
Underlying Structure and Multivariate Links

[Gapminder Foundation, https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen]



Underlying Structure and Multivariate Links

[Gapminder Foundation, https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen]



Documentation