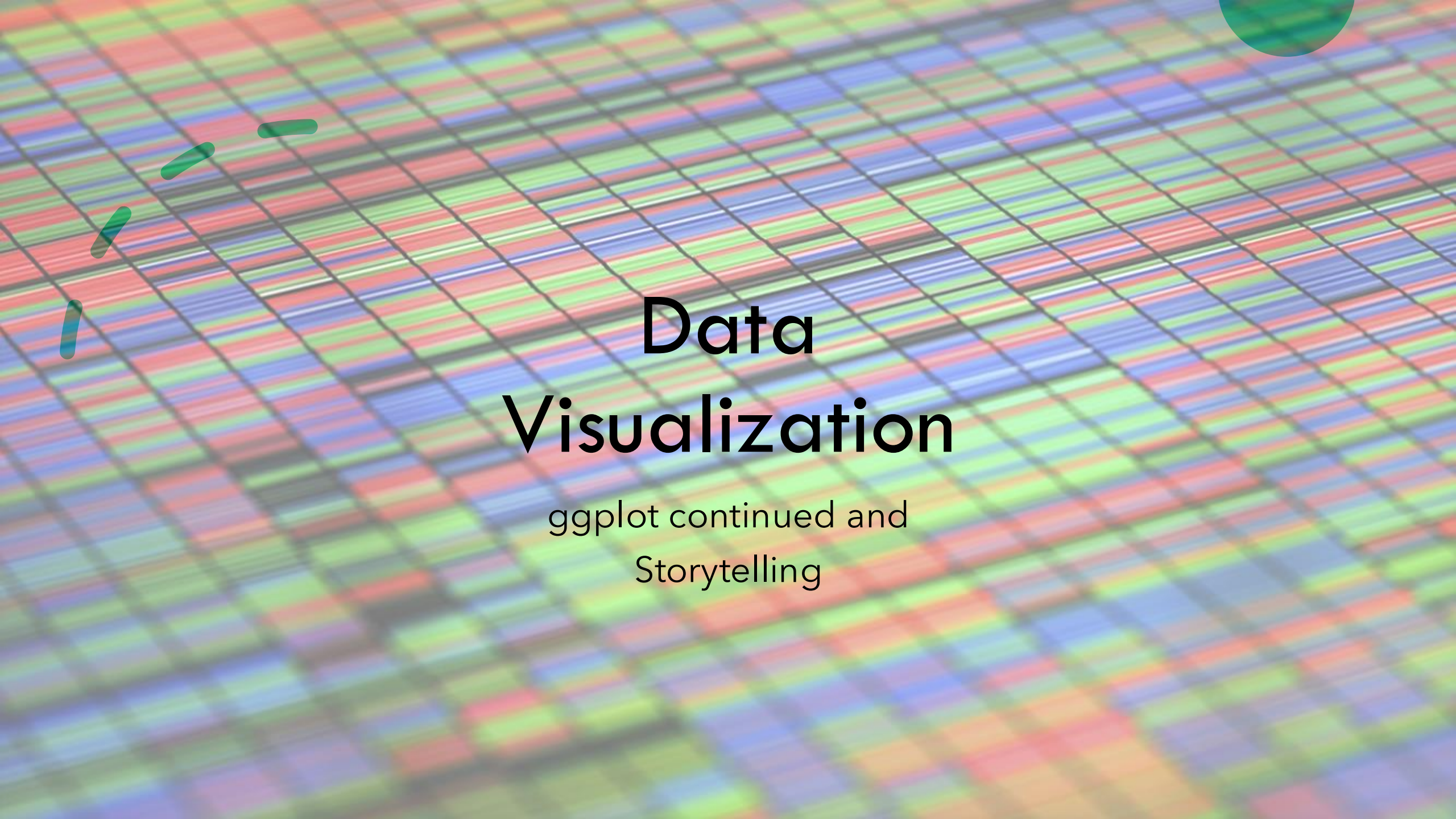


Data Visualization

ggplot continued and Storytelling

What is coming due

https://uzk-evaluation.uni-koeln.de/evasys/online.php?pswd=Z5ND2

Agenda

<u>Today (Tuesday 27 January)</u>

- Some more advanced ggplot examples
- Story Telling

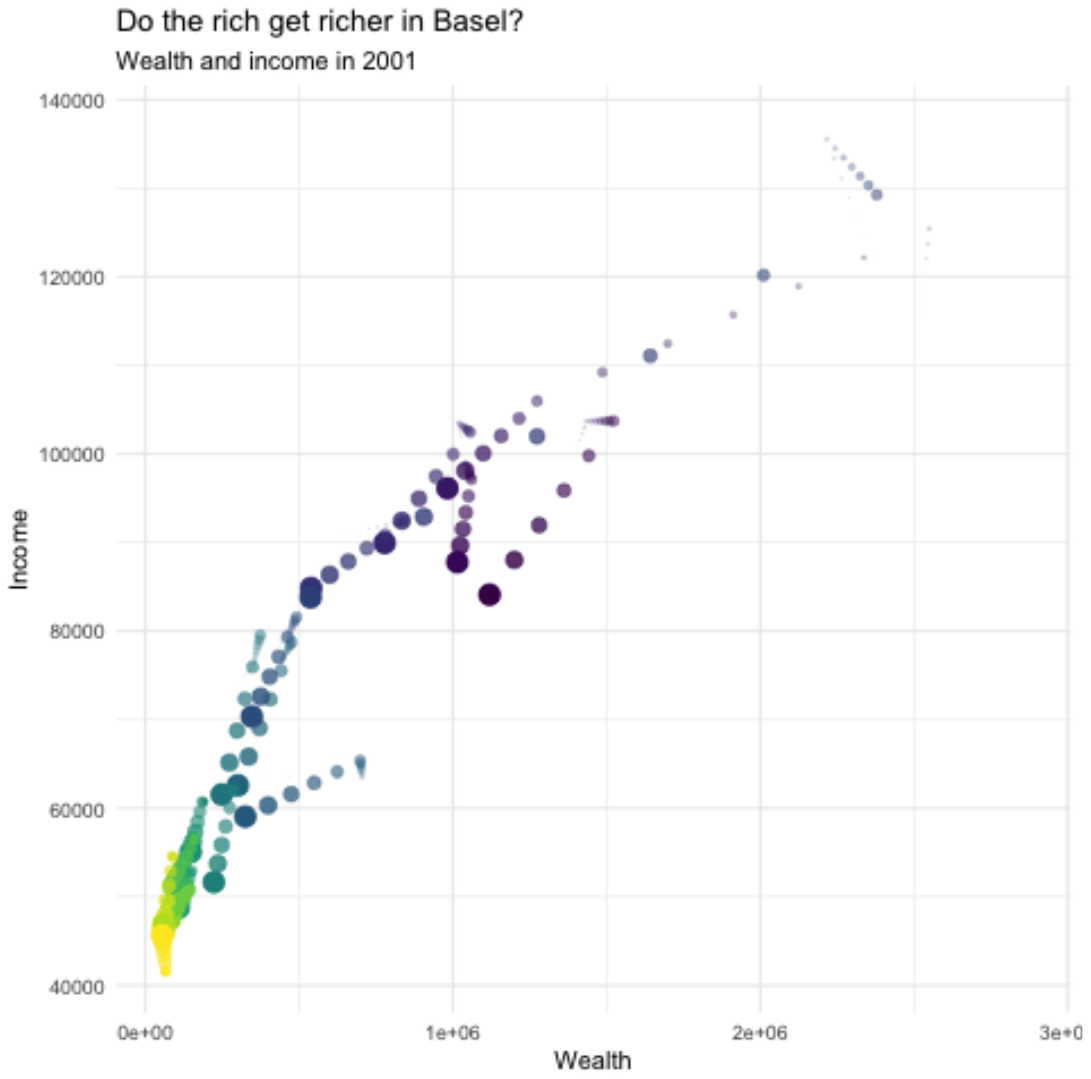
Tomorrow (Wednesday 28 January)

- Story Telling Exercise (Introduction to data & assignment)

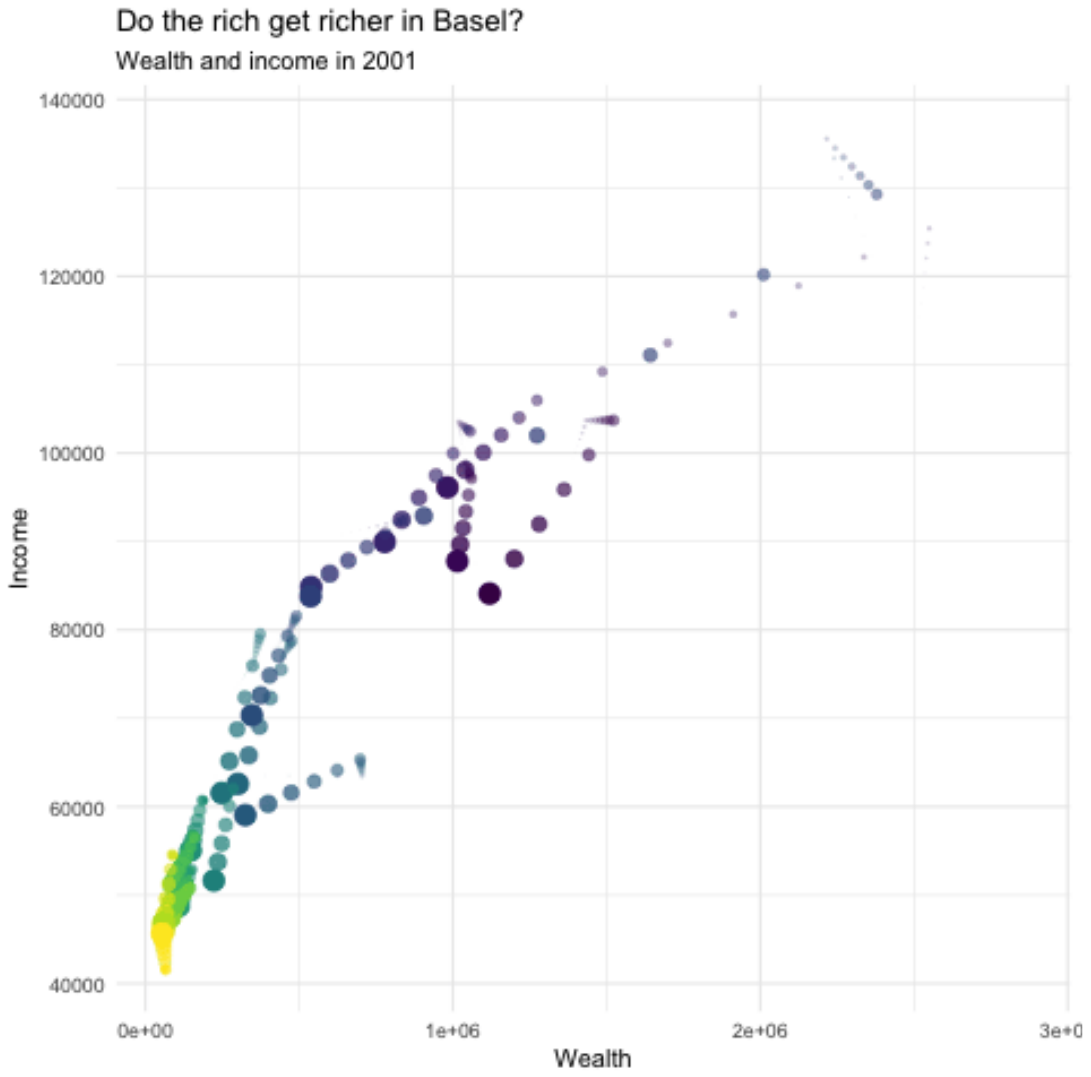
Animations in R

taxation.csv

animate_class_solution.R



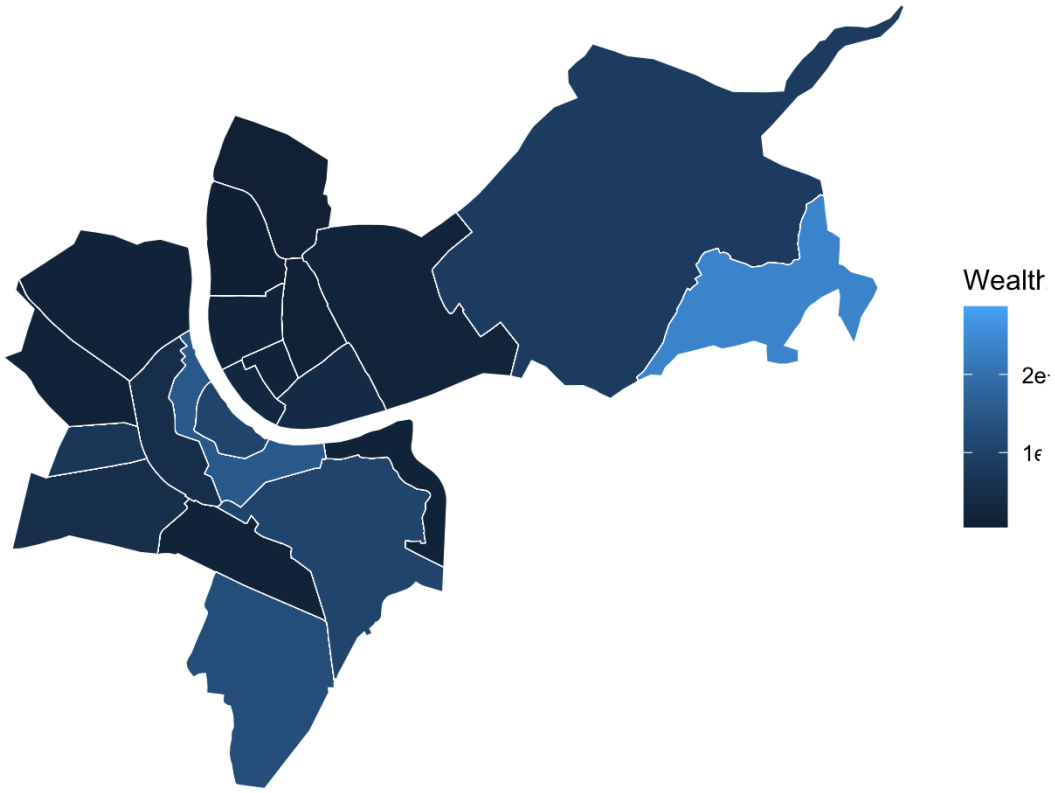
Class Example



Maps

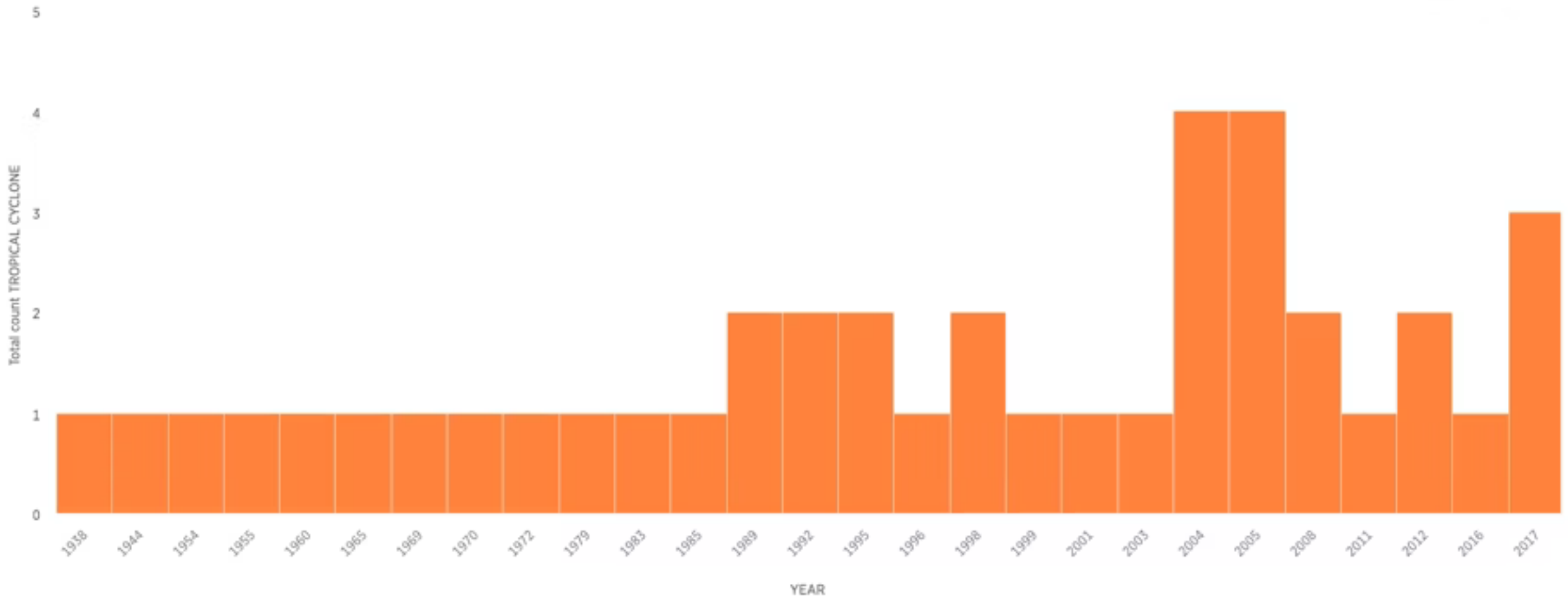
Class Example

taxation.csv
quarters.zip
maps_and_nets_class_solution.R



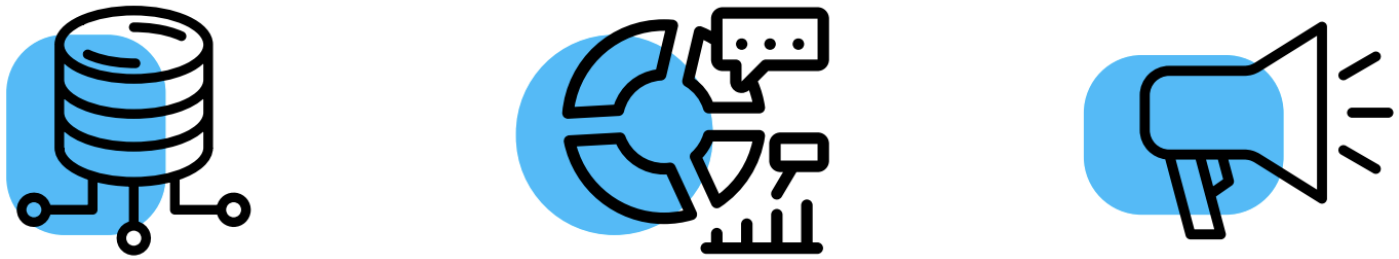

Storytelling

Number of high intensity hurricanes by year



What is data storytelling?

Data storytelling is often called the last mile of analytics. Sound communication skills, allows data professionals to drive action out of their insights. According to Brent Dykes, Author of Effective Data Storytelling: How to Drive Change with Data, Narrative, and Visuals—Data Storytelling is a combination of data, visuals, and narrative.



Data **Visuals** **Narrative**

The three elements of data storytelling
(Source: Effective Data Storytelling: How to Drive Change with Data, Narrative, and Visuals by Brent Dykes)

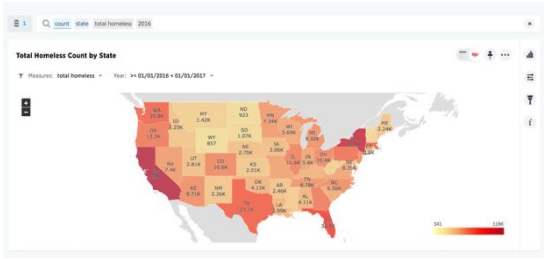
Examples thoughtspot.com

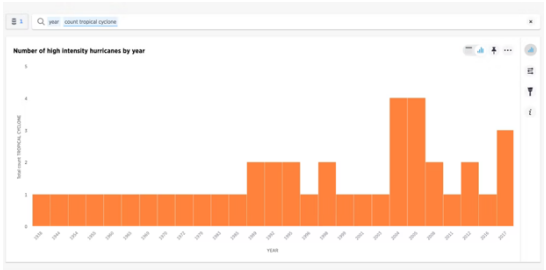
- https://www.thoughtspot.com/thoughtspot-blog/homelessness-america-story-10-charts

- https://www.thoughtspot.com/thoughtspot-blog/are-hurricanes-getting-worse-story-told-7-hurricane-strength-charts-thoughtspot

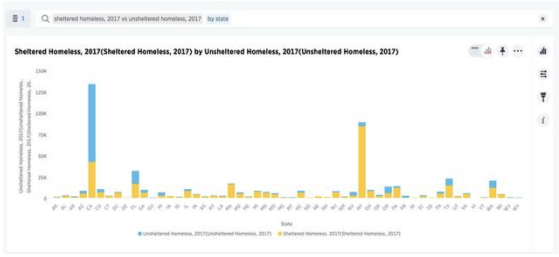
Good data stories follow a script that resembles traditional stories

1. **Setup:** establish the problem

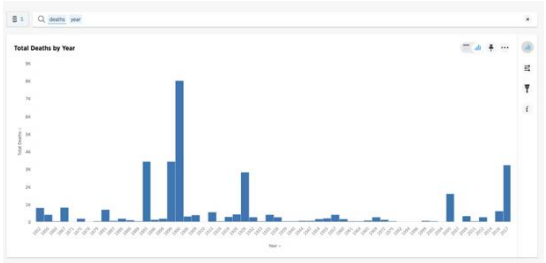




2. **Rising Action:** exploring complexities, nuances and causes



3. **Resolution:** Provide solutions, show what works, and inspire action

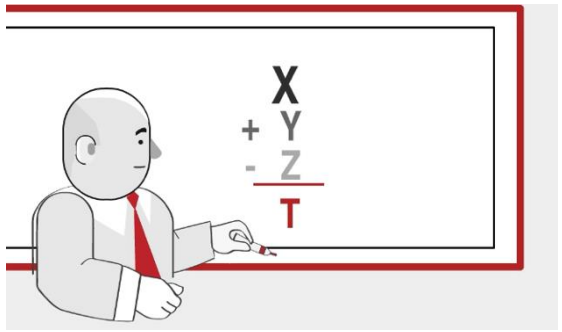


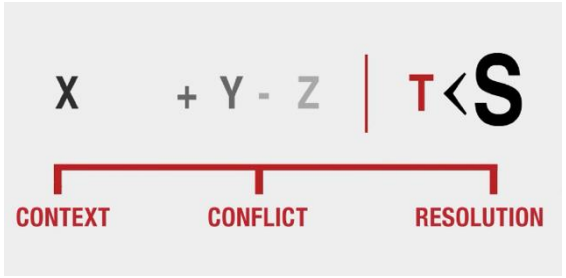
Examples

Story Telling: *The How We Got Here Method*

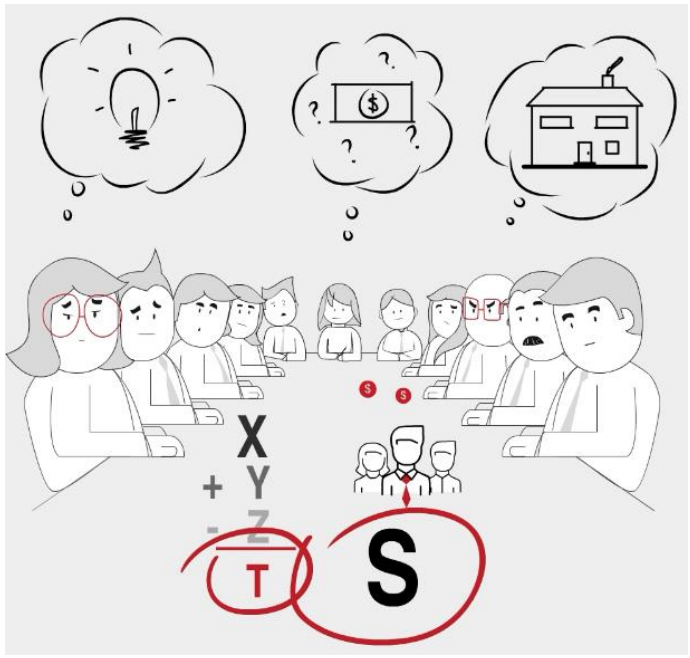
Story Telling: *The How We Got Here Method*

Instead of giving the answer, walk the audience through

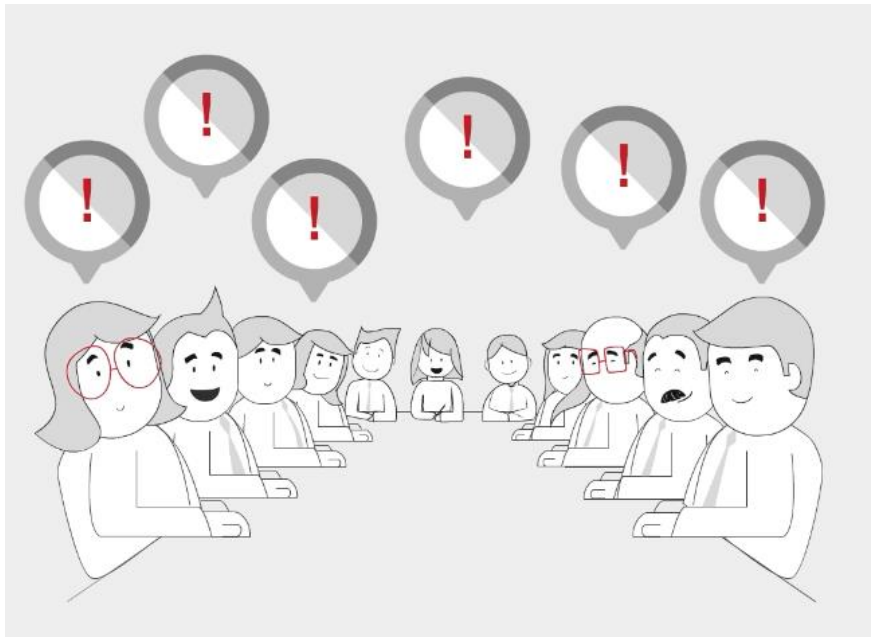




Demonstrating instead of telling, let the audience figure out the problem



Let the audience come up with their own suggestions



Story Telling: *The Discovery Journey Method*

Story Telling: *The Discovery Journey Method*

HOW THIS DATA BECAME A STORY

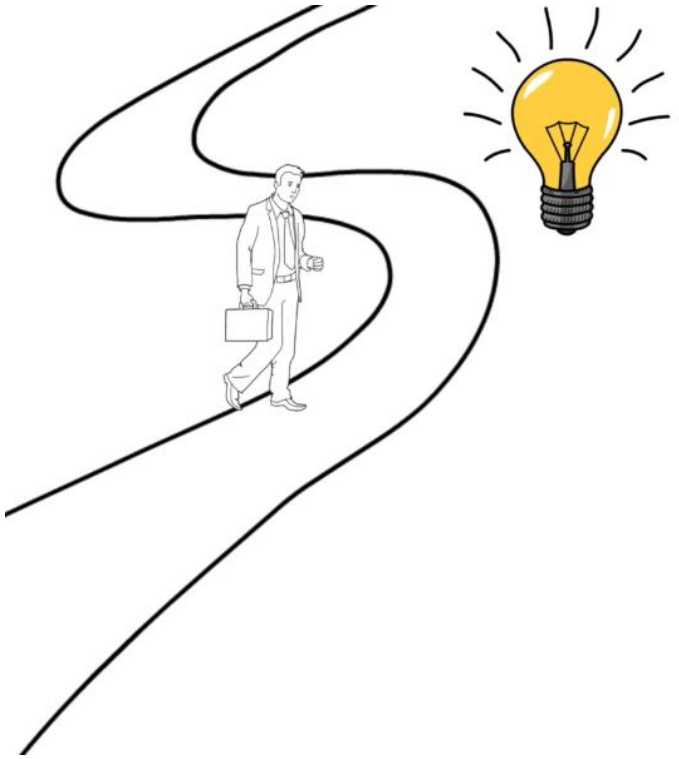
1) STRUCTURED LIKE A STORY, NOT A PRESENTATION

2) LET AUDIENCE STRUGGLE WITH DATA TO FIND SOLUTION

3) EMOTIONAL IMPACT OF SHOCKING DATA

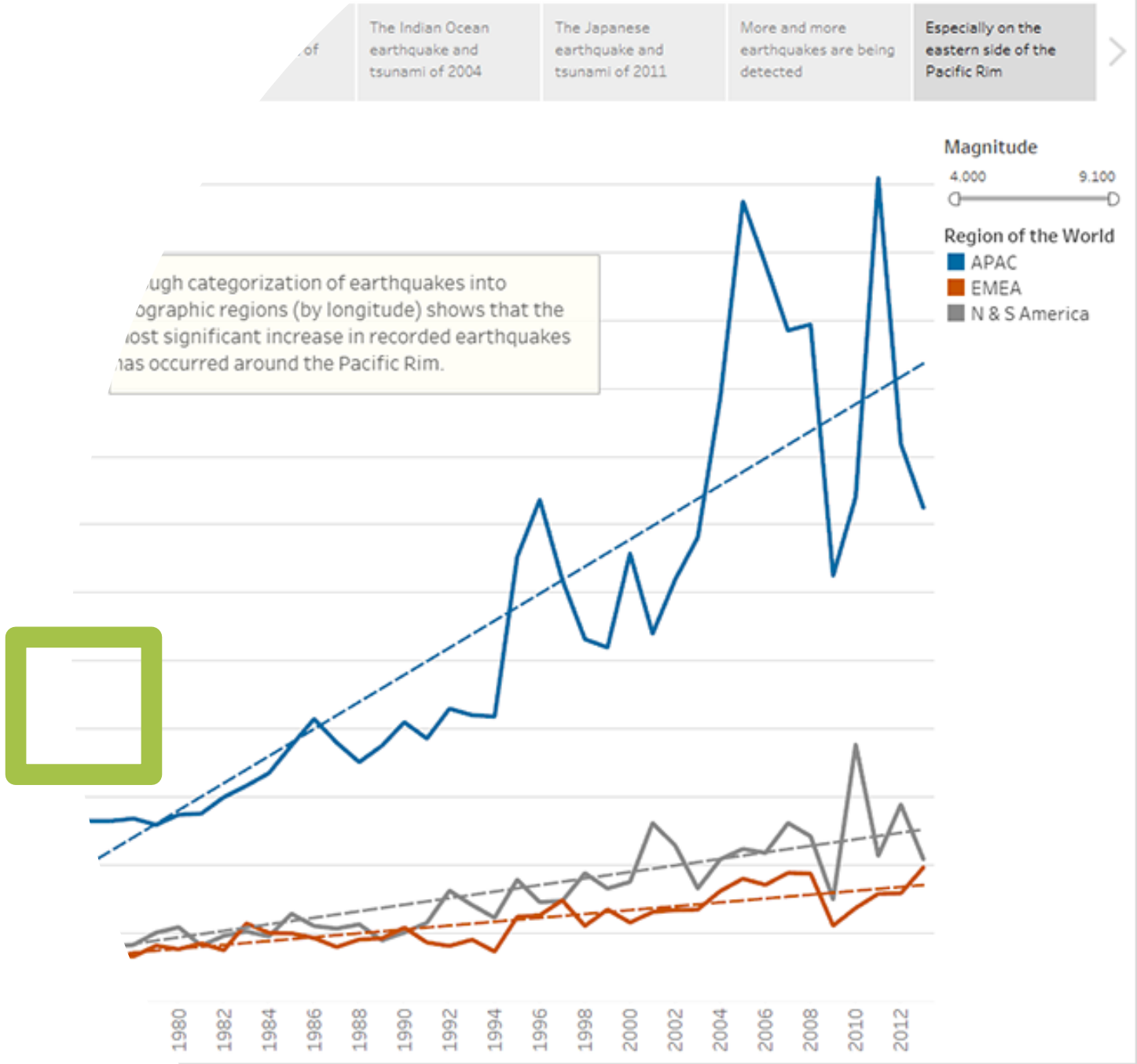
4) SURPRISE OF SOLVING THE MYSTERY

5) LET AUDIENCE DRAW OWN CONCLUSIONS AND RECO



Tools for Story Telling

Tableau Stories Example



R Markdown

Rmarkdown documents

````
---
title: "Example Analysis with RMarkdown"
author: "Your Name"
date: "January 12, 2025"
output: html_document
---

# Introduction

This document demonstrates how to combine **text** and **R code** in an RMarkdown file. We will:
1. Load a dataset.
2. Explore the data with basic statistics.
3. Create a simple plot.

The dataset we'll use is the built-in `cars` dataset, which contains information about cars, including miles per gallon (mpg), horsepower (hp), and weight (wt).

# Load Libraries

First, we load the necessary library.

```{r}
library(tidyverse)
````
````

Data Storytelling & Communication

Cheat Sheet

Learn more online at www.DataCamp.com

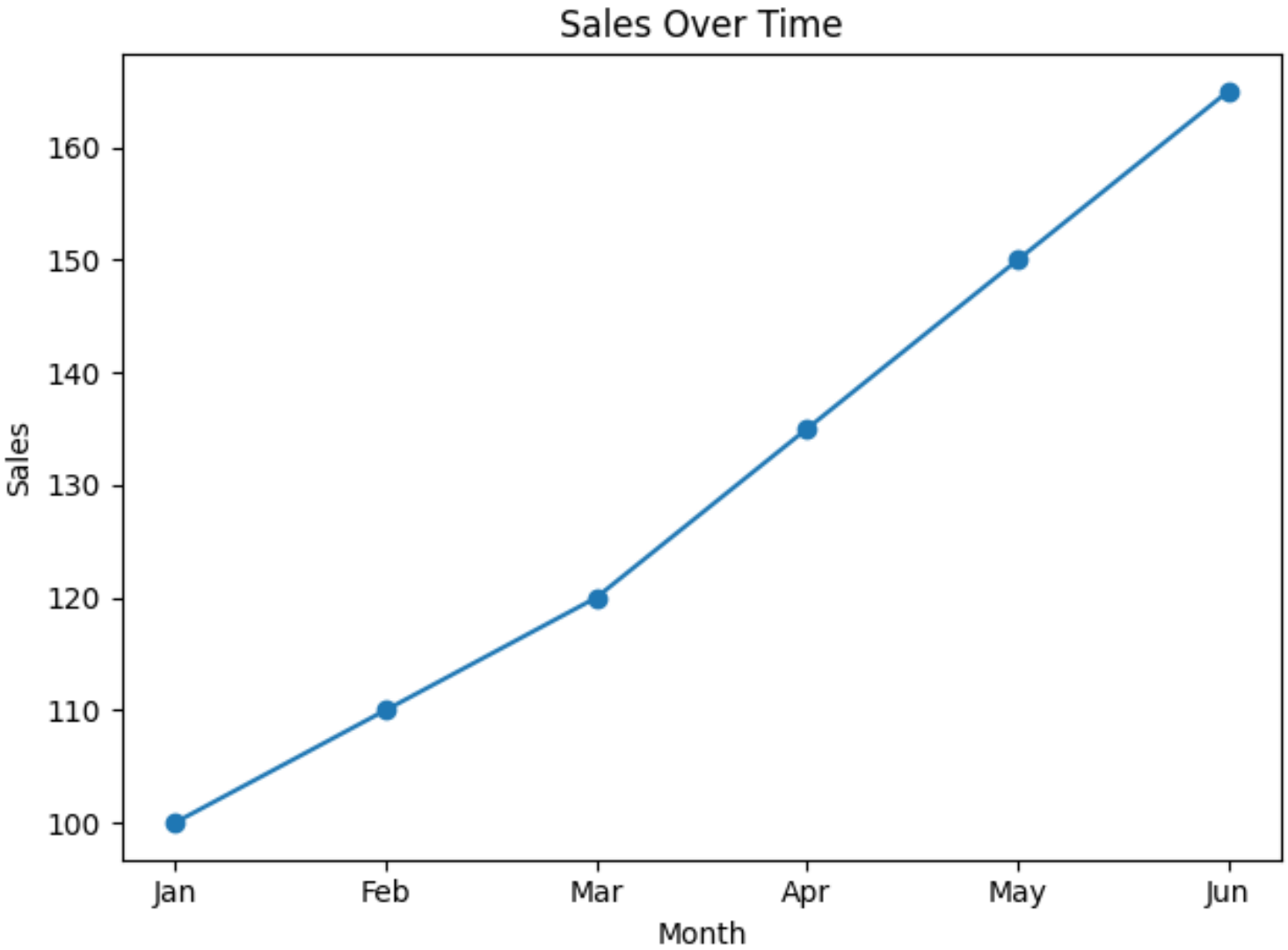
Best Practices for Story Telling

1. Find the story within the data

Before constructing a data story, you must first determine what the data is actually telling you. For instance, a correlation or causal link in a dataset may provide an interesting insight that can inform a more meaningful narrative. This provides a starting point into building a data story that can be conveyed to a larger and potentially technically diverse audience.

https://www.thoughtspot.com/data-trends/best-practices/data-storytelling

1. Find the story within the data



> Crafting effective visuals

Choose the best visualization for your story

Each plot type is suited for communicating specific things about specific types of data. Start by choosing an appropriate plot type.

Line plot

Show changes in numeric values over time.

Bar plot

Visualizes numeric values by categories. It can be ranked or unranked

Scatter plot

Show the relationship between two numeric values.

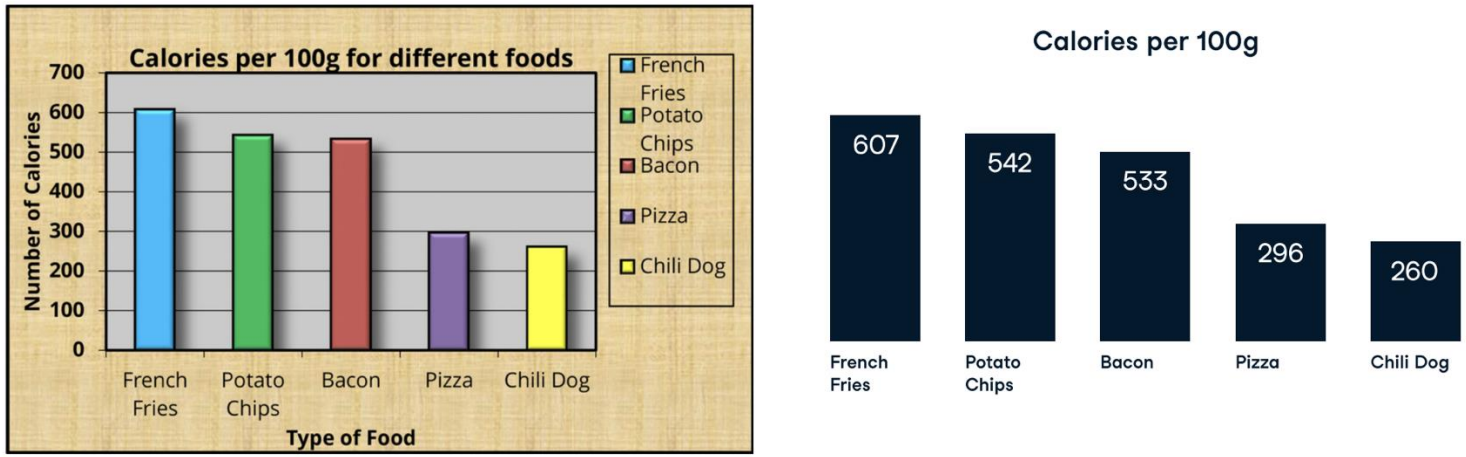
Histogram

Show the distribution of numeric values.

To learn about all the types of visualizations you can use, check out our **Data Visualization Cheat Sheet.**

Keep visualizations minimal and avoid clutter

Ruthlessly edit your plots to remove or minimize elements that distract from the message of the plot. In particular, make non-data elements (parts of the plot that don't directly represent a data value, like the grid lines) less distracting. A great example comes from Darkhorse Analytics, which showcases exactly the value of decluttering visualizations.



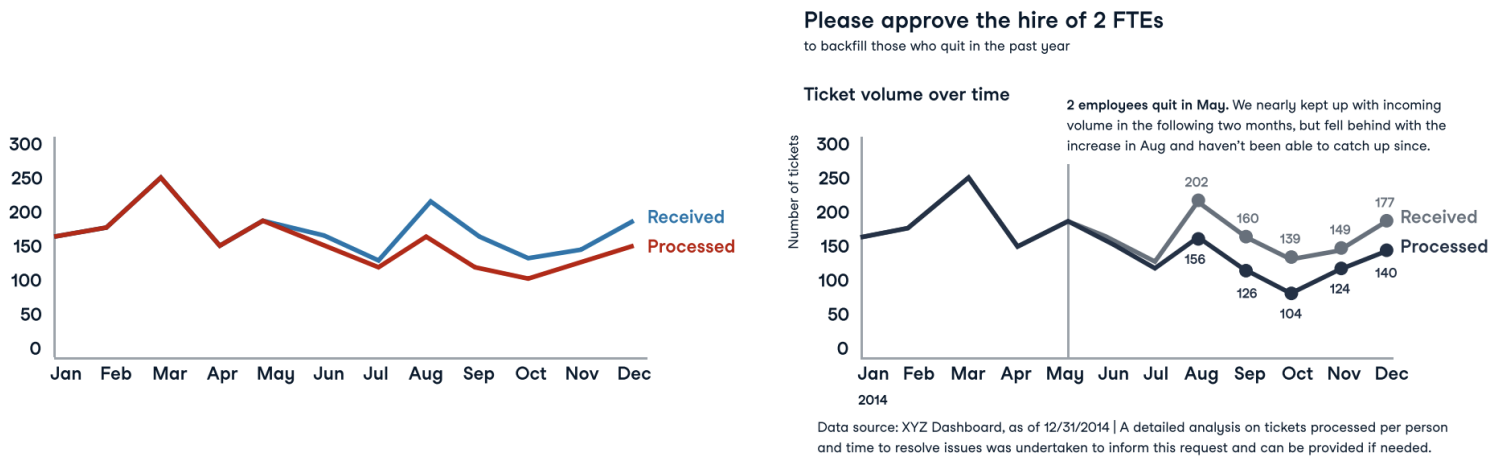
Decluttering a visualization in action
(Source: Darkhorse Analytics)

Data visualization decluttering best practices

- Use just enough white space to keep the visualization from looking busy
- Remove chart borders when applicable
- Remove or minimize gridlines or axes when applicable
- Clean up axis labels when applicable
- Label data directly (as opposed to using a legend)
- Remove data markers when applicable
- Use special effects (**bold**, underline, *italics*, shadows) sparingly

Use text appropriately

While too much text can add clutter, text can also be an extremely effective tool at highlighting insights within your visualizations. Cole Nussbaumer Knaflic, Author of Storytelling with Data, provides an excellent example with the following visualization.



How text can be a useful visual tool when crafting effective visuals
(Source: Storytelling with Data by Cole Nussbaumer Knaflic)

Using text in data visualizations

- When applicable, label axes and titles for clarity
- Label important data points when necessary
- Provide useful context around insights within the title or subtitle
- Adjust font size when highlighting specific messages within your labels
- When applicable, try to answer common audience questions with labels

Use colors effectively

The fundamentals of color theory in data visualization

Color is one of the most powerful tools available for emphasizing different aspects of your data visualization. Here are different properties to keep in mind when choosing an appropriate color palette for your visualization.

- **Hue** represents the range of possible colors, from red, through orange, green and blue, to purple and back to red.
- **Chroma** is the intensity of the color, from grey to a bright color.
- **Luminance** is the brightness of the color, from black to white.

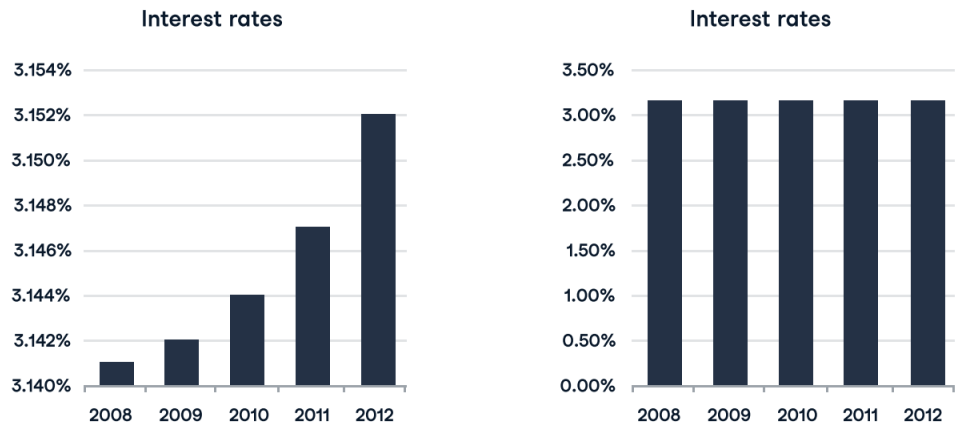
There are three common types of color palettes, that depend on these dimensions.

Type	Purpose	What to vary	Example
Qualitative	Distinguish unordered categories	Hue 	A bar chart of 2022 smartphone sales for different smartphone manufacturers
Sequential	Showcase intensity of a single variable	Chroma or luminance 	A map showcasing Covid-19 vaccination prevalence
Diverging	Compare between two groups	Chroma or luminance with two hues 	Voter registration prevalence by political party in the USA

Do not mislead with data stories

The fastest way to lose credibility when presenting data stories is to inadvertently (or intentionally) mislead with your data insights. Here are top best practices to avoid misleading with data stories.

Same Data, Different Y-Axis



Starting the y-axis at the smallest value or at zero dramatically changes the story told by the plot

Best practices to avoid misleading with data stories

- If you are visualizing times series data, make sure your time horizons are large enough to truly represent the data
- If the relative size of each value is important, then ensure that your axes start with zero
- Ensure that axes scales are appropriate given the data you're treating
- If you are sampling data for descriptive purposes, make sure the sample is representative of the broader population
- Use centrality measures such as mean or median to provide context around your data

> Crafting effective narratives with data

Know the audience

To communicate effectively, you need to know who your audience is, and what their priorities are. There is a range of possible audiences you may encounter when presenting, and crafting an audience specific message will be important. Examples of audiences you may present to are:

Executive

Basic data literacy skills

Prioritizes outcomes & decisions

Cares much more about business impact than a 1% incremental gain in a machine learning model accuracy or a new technique you're using

Data Leader

Data expert

Prioritizes rigour & insights

Cares much more about how your arrived at your insights and to battle test them for rigour

Business Partner

Advanced data literacy skills

Prioritizes tactical next steps

Cares much more about how your analysis impacts their workflow, and what should be their main takeaway from the data story

Considerations when crafting audience specific messaging

Aspect	What do you need to consider?
Prior knowledge 🎓	▪ What context do they have about the problem? ▪ What is their level of data literacy?
Priorities 🎯	▪ What does the audience care about? ▪ How does your message relate to their goals? ▪ Who is driving decision-making within your audience?
Constraints 🌟	▪ What is the audience's preferred format? ▪ How much time does an audience have to consume a data story?

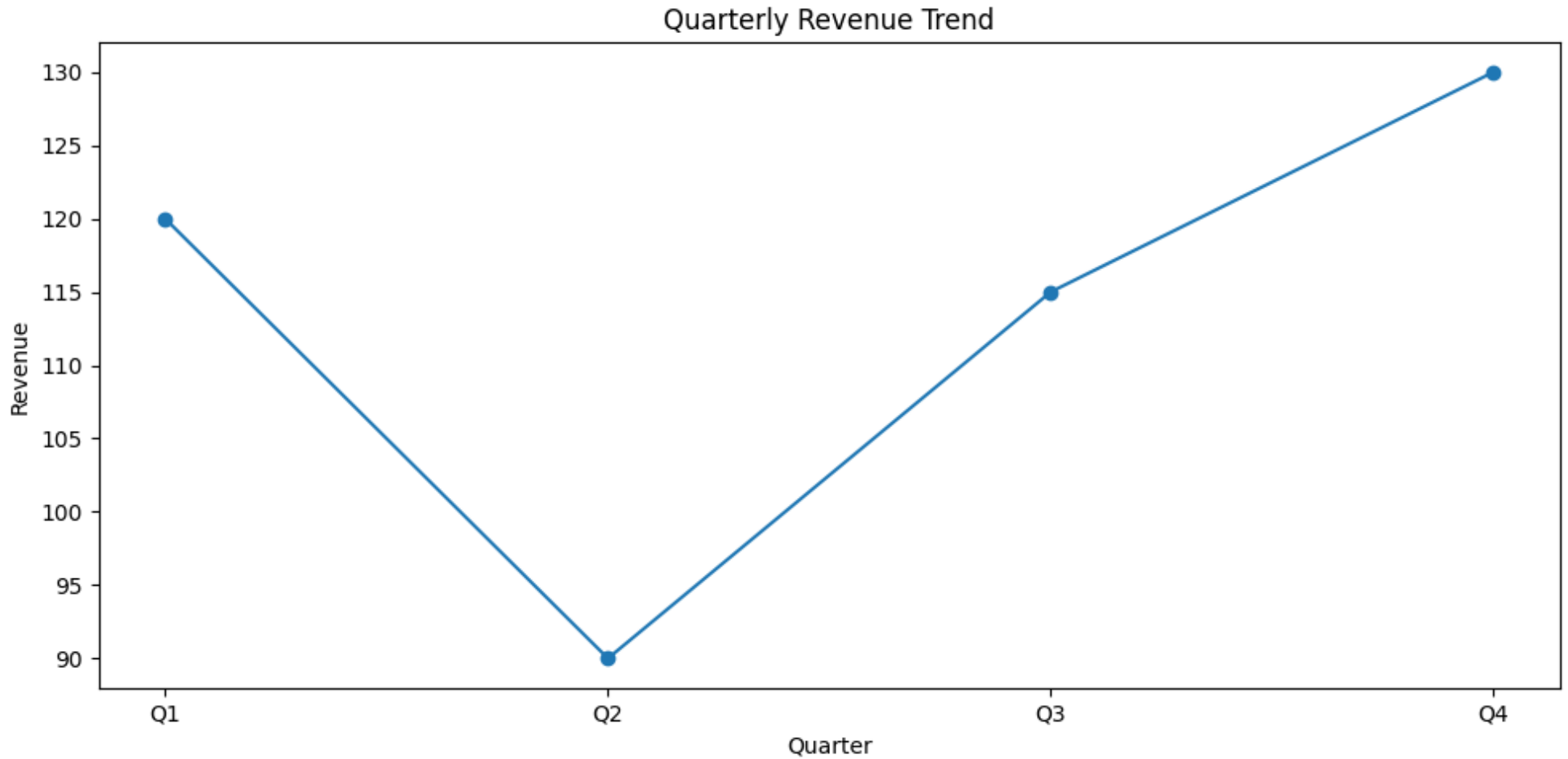
6. Provide context

Data on its own is not enough to create an optimal data story. The domain expertise around the topic is essential to crafting the right narrative, and making it understandable to the audience. Therefore, it's critical to weave context and data insights together. This doesn't just mean providing an overview at the beginning before diving into data visualizations. It means framing the data insights with the business reality and any other relevant information that your audience may not otherwise be familiar with throughout the entire story.

Perhaps a correlation is caused by something that is very specific to an industry or business segment, or maybe there's an anomaly in the data that is impacted by an external event. Pointing these out will be important to provide sufficient context to understand the broader narrative.

https://www.thoughtspot.com/data-trends/best-practices/data-storytelling

Provide context



7. Structure your story

Telling a traditional story often includes providing an introduction, a rising action that builds tension through problems, a climax that provides a crucial moment or insight, a falling action that resolves the problems, and a final resolution and retrospection.

This type of flow is often still valuable in telling a data story. Leveraging a structure that people are familiar with will help capture their attention, create a framework for information consumption they're already familiar with, and allow them to better understand the relevance and importance of the story that the data is telling.

https://www.thoughtspot.com/data-trends/best-practices/data-storytelling

8. Edit until the story is clear and concise

When presenting a data story, regardless of the topic, it's important that the story is clean and to the point. As with other writing and presentation, you should continuously review and edit the narrative until it's extremely clear. As Blaise Pascal famously said "If I had more time, I would have written a shorter letter." The same is true for data stories. Avoiding fluff that distracts from the main points is critical to ensure the audience is focused on what really matters.

https://www.thoughtspot.com/data-trends/best-practices/data-storytelling

Choose the best medium to share your story

There are different ways you can deliver a data story. The importance of each is different depending on the audience of your data story and the setting you're delivering your data story in.

Type	Important considerations
Presentation 💬	▪ Ensure the length of your presentation is appropriate ▪ Leave any highly technical details to the appendix ▪ Ensure there is a narrative arc to your presentation
Long-form report ✍️	▪ Be extra diligent about providing useful context around data visualizations and insights ▪ Leave any highly technical details to the appendix
Notebook 📊	▪ Ensure that you provide useful context on how you arrived at a certain conclusion
Dashboard 🙌	▪ Make use of the dashboard grid layout ▪ Organize data insights from left to right, top to bottom ▪ Provide useful summary text of key visualizations in your dashboard