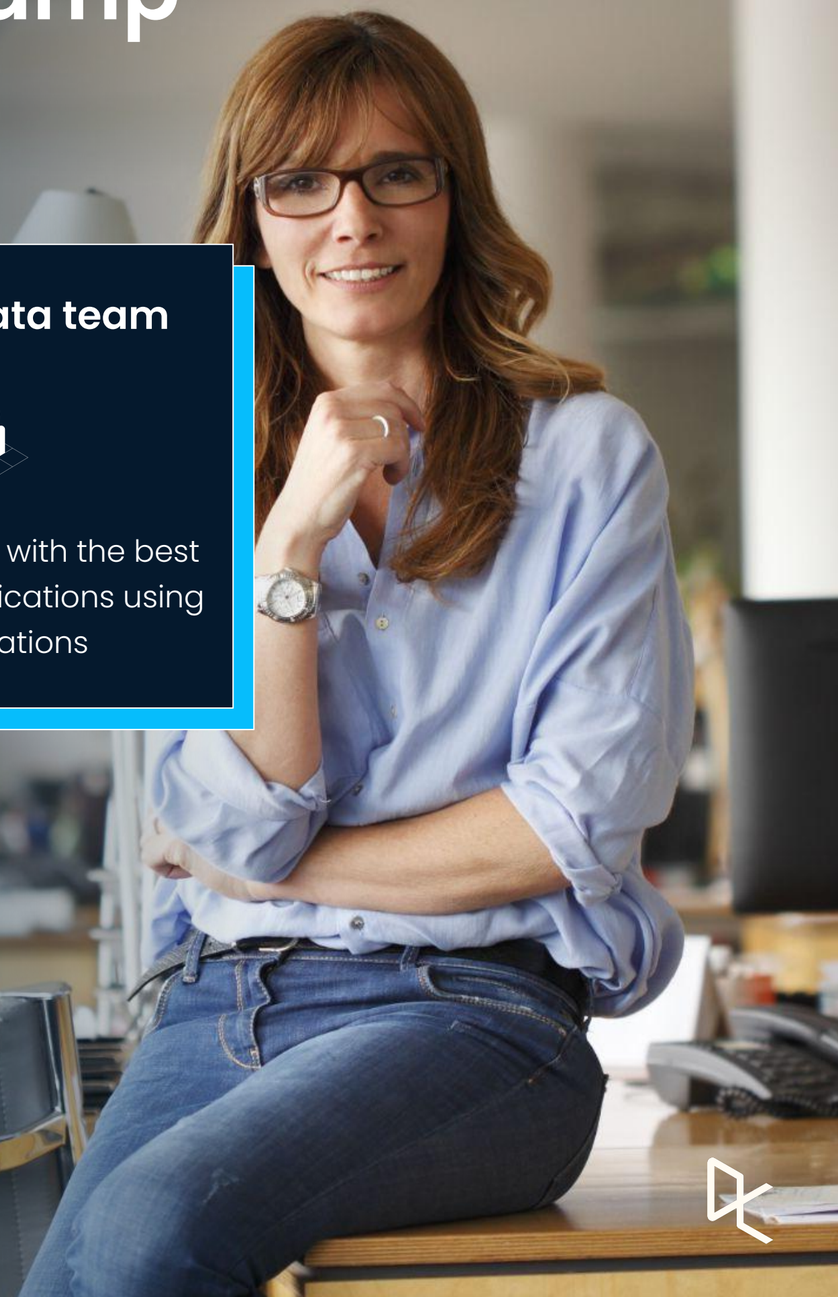
8 Rules for Better Data Storytelling

Our Mission

Our mission is to democratize data & AI skills for everyone

Future-proof your skills with DataCamp



Best in class learning

Give your employees access to market-leading training with DataCamp Learn

Build work ready skills

Apply your skills in a risk-free online coding environment with DataCamp Datalab

Grow your data team

Certify your people with the best in-class data certifications using DataCamp Certifications

Trusted by more than 14 million learners and 5,000 data-driven companies

Google Microsoft ebay HSBC COLGATE-PALMOLIVE T··Mobile···

PayPal Uber Deloitte. CREDIT SUISSE Mercedes-Benz BNP PARIBAS

About me

Adel Nehme

VP of Media

DataCamp

adelnehme

- Graduate in Economics from the American University of Beirut
- MSc in Business Analytics & Data Science from ESSEC Business School & CentraleSupelec
- Data Science Educator & Evangelist @ DataCamp
- 🎤 Host of the DataFramed Podcast 🎤

Agenda

1. Data storytelling: The last mile of analytics
2. 8 rules for better data storytelling
    - *4 rules for better data visualizations*
    - *4 rules for better narrative*
3. Become a better data storyteller

1

Data Storytelling

The last mile of analytics

What is data storytelling?

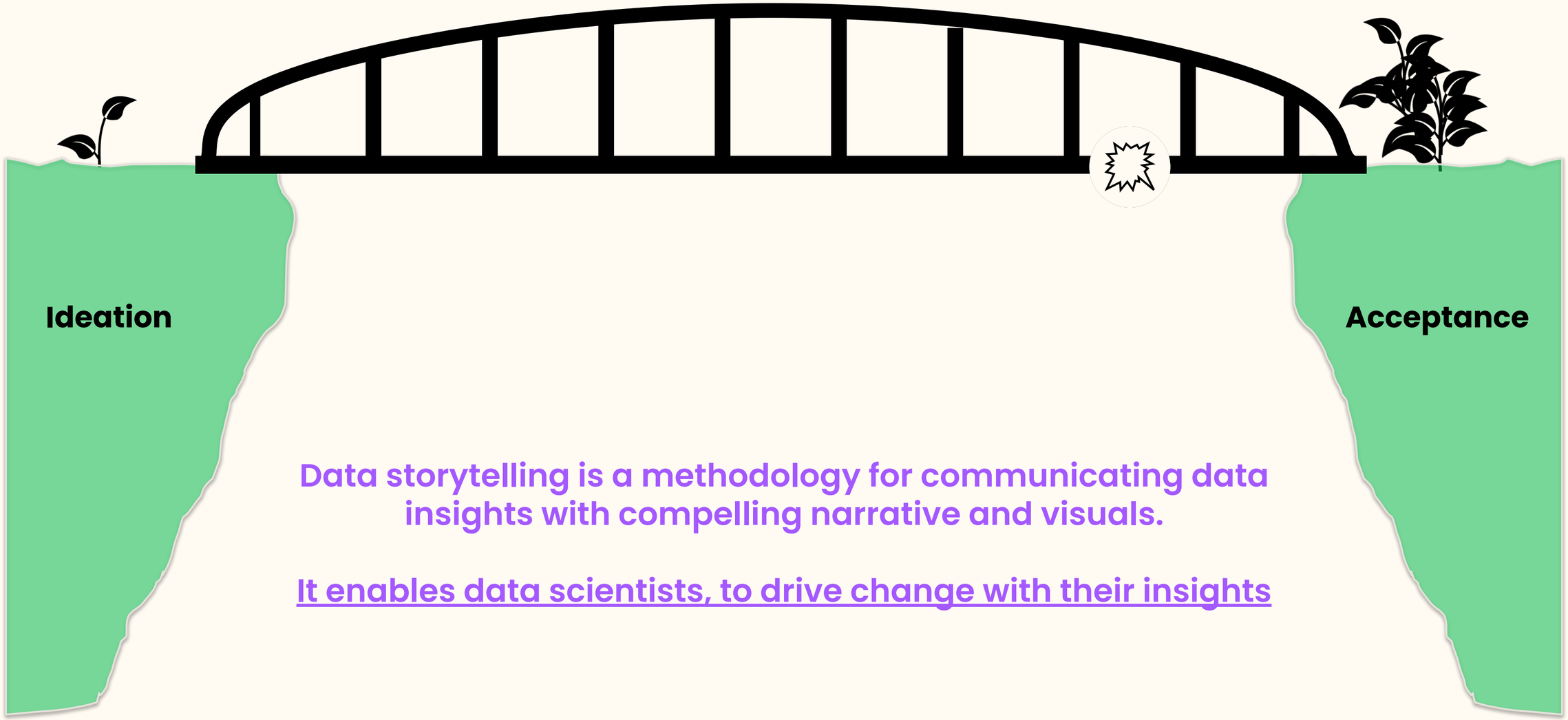
*“Having all the information in the world at our fingertips doesn’t make it easier to communicate: it makes it harder. The more information you’re dealing with, the more difficult it is to filter” — **Cole Nussbaumer Knaflic, Author of Storytelling with Data: A Data Visualization Guide for Business Professionals***

*“Data storytelling enables data teams to wield the power to frame arguments and persuade with data responsibly and deliberately ” — **Andy Cotgreave, Technical Evangelist at Tableau***

*“To be impactful with data science, organizations need data scientists with stories ” — **Gert de Geyter, Machine Learning Lead at Deloitte***

*“Data stories are powerful vehicles for sharing data insights to influence and drive change within an organization — **Brent Dykes, Author of Effective Data Storytelling***

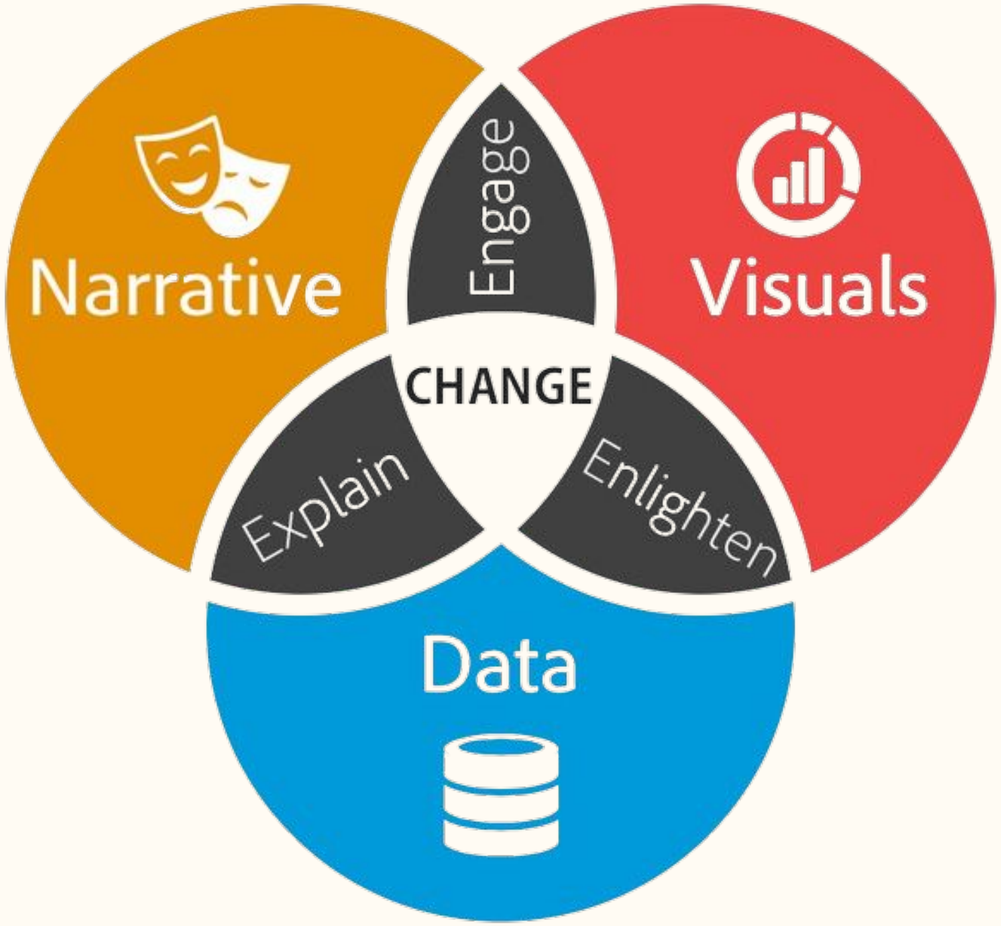
What is data storytelling?



Data storytelling is a methodology for communicating data insights with compelling narrative and visuals.

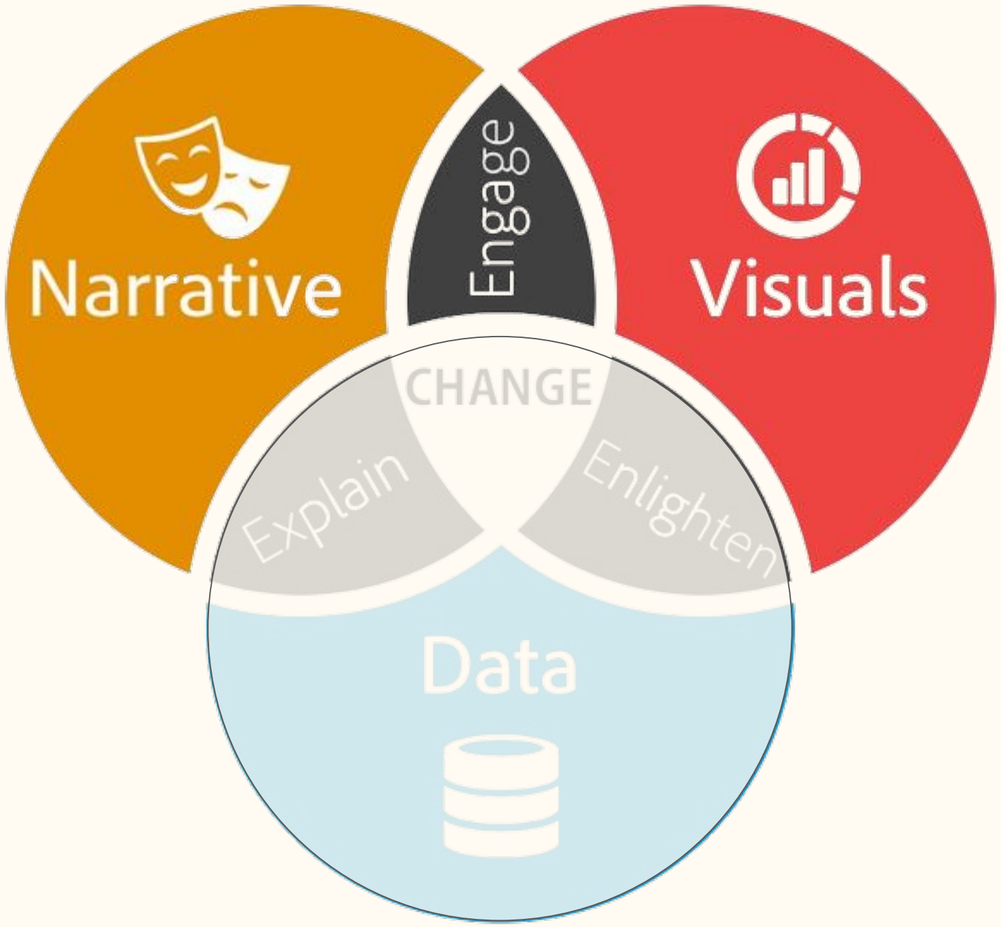
It enables data scientists, to drive change with their insights

What is data storytelling?



Source: Effective Data Storytelling: How to Drive Change with Data, Narrative, and Visuals

What is data storytelling?



Source: Effective Data Storytelling: How to Drive Change with Data, Narrative, and Visuals

2

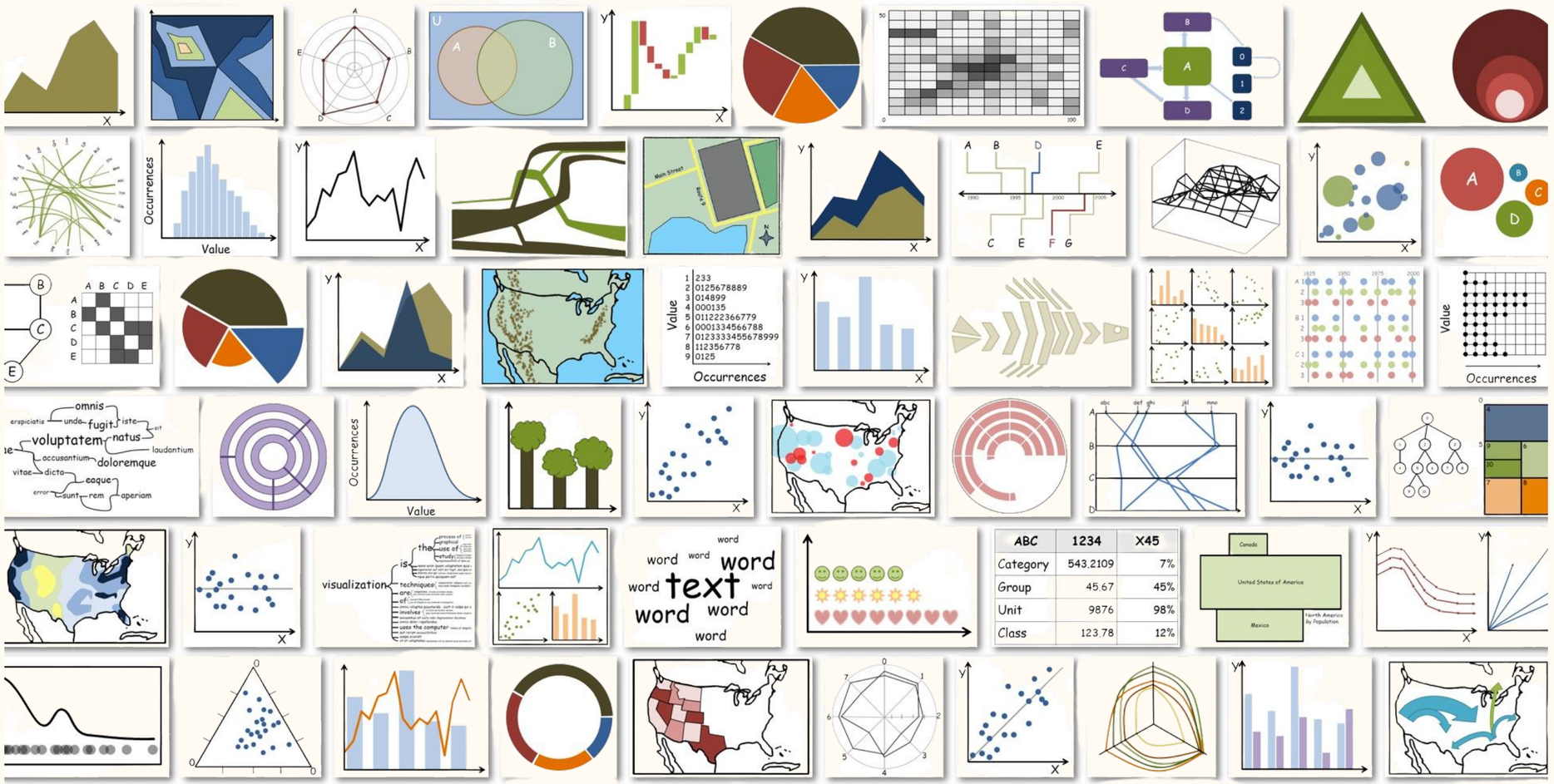
8 rules for better data storytelling

Rules for better data visualizations

Rule #1

Choose the best visualization for your story

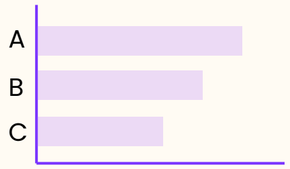
So many ways to cut a cake



Always work back from your data

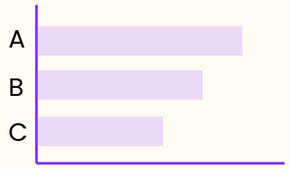
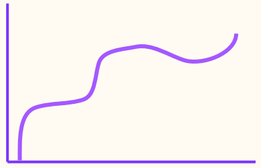
COMMONLY USED DATA	PROBLEM DESCRIPTION	MOST USEFUL VISUALIZATION

Always work back from your data

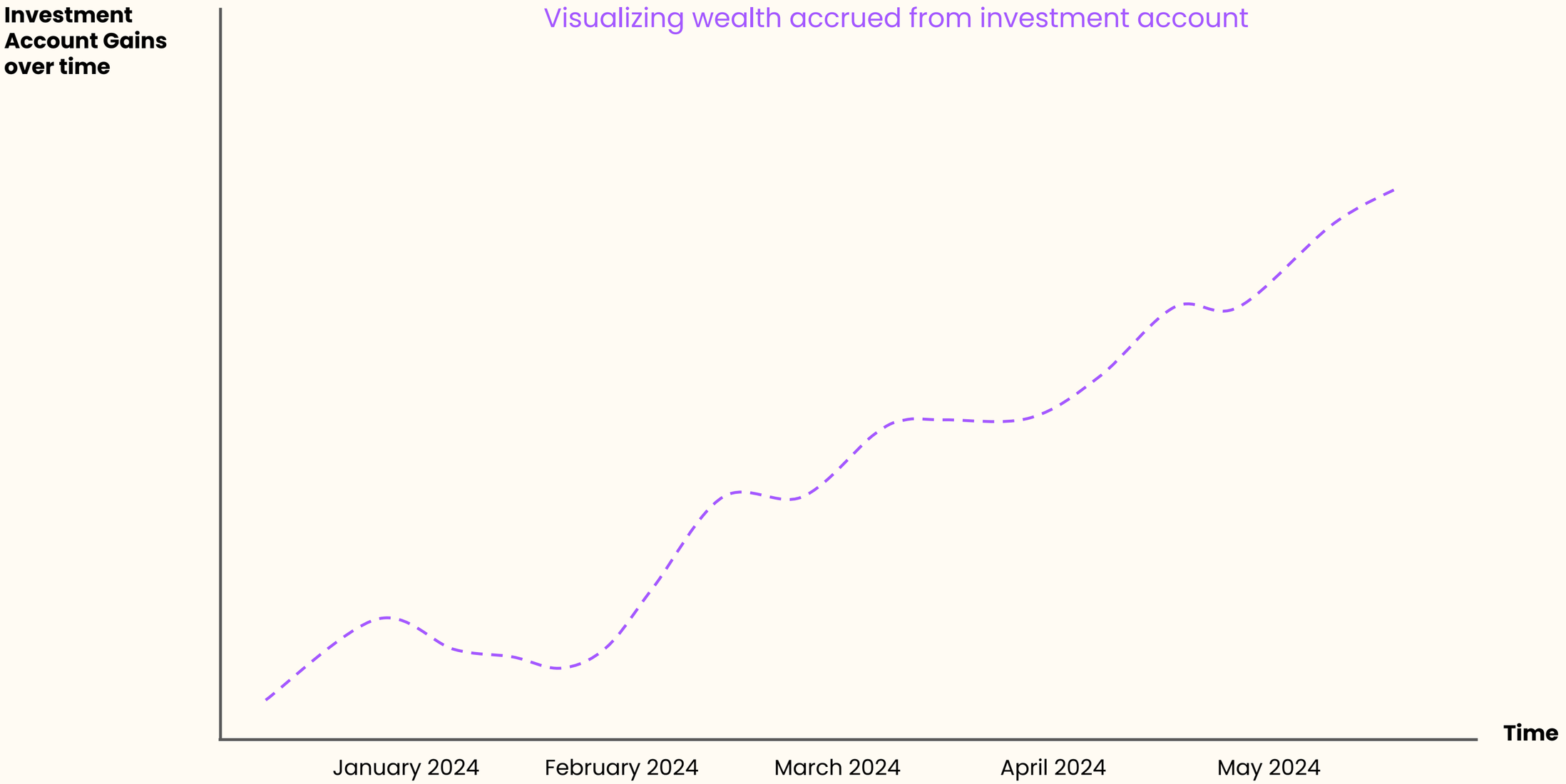
COMMONLY USED DATA	PROBLEM DESCRIPTION	MOST USEFUL VISUALIZATION
• Banking products and associated customers • Branch cost data broken down into different verticals • Visualizing different customer segments	Showing comparisons of different categories	 **BAR CHARTS**

Common uses of bar charts

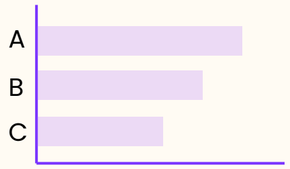
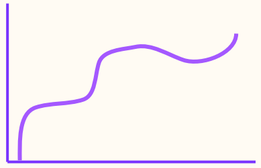
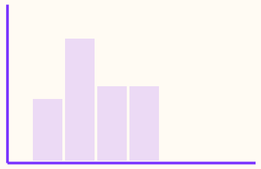
Always work back from your data

COMMONLY USED DATA	PROBLEM DESCRIPTION	MOST USEFUL VISUALIZATION
• Banking products and associated customers • Branch cost data broken down into different verticals • Visualizing different customer segments	Showing comparisons of different categories	 **BAR CHARTS**
• Stock price change over time • Number of app users over time • Number of customer support tickets over time	Showing changes of a variable over time	 **LINE CHARTS**

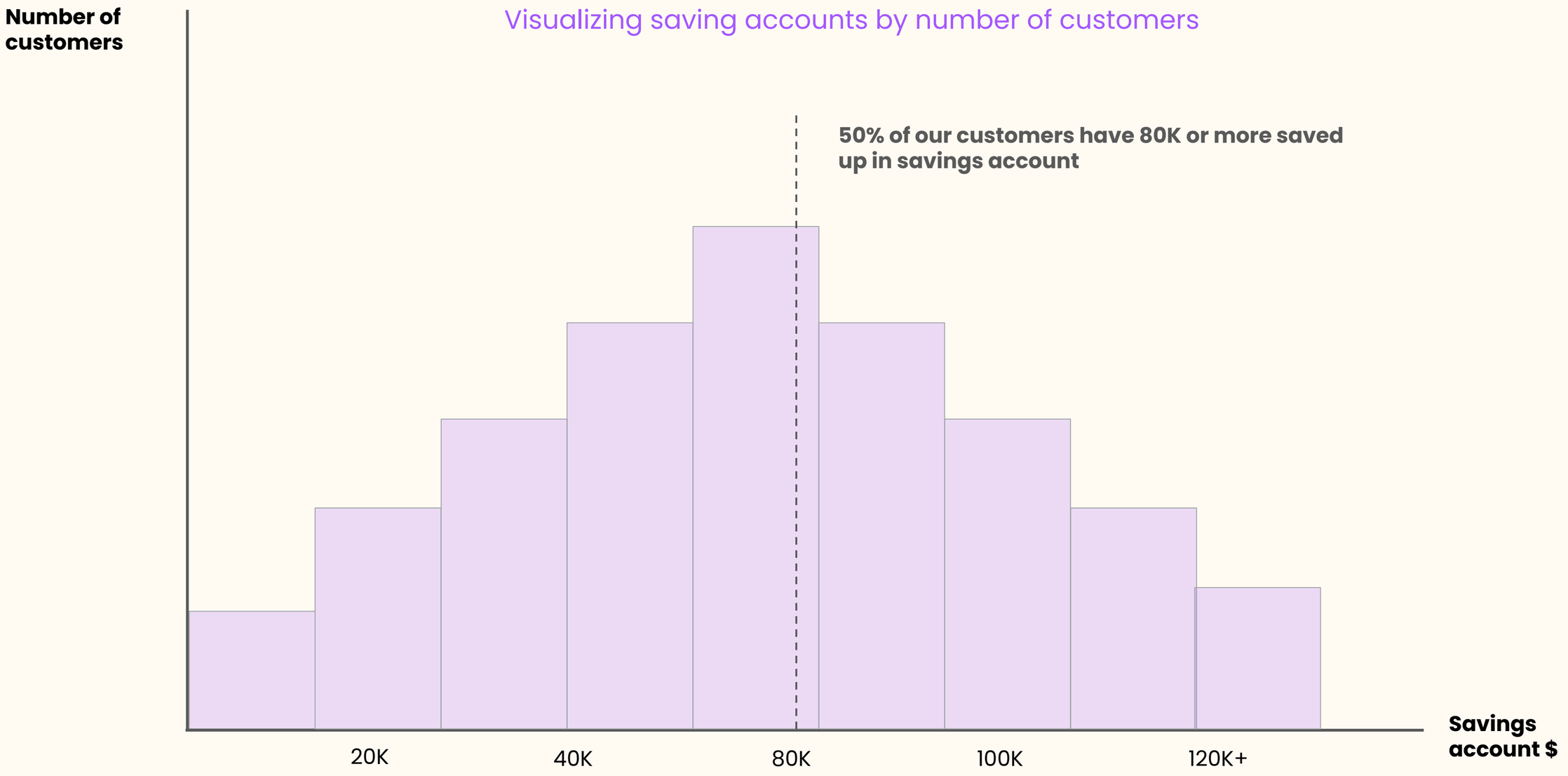
Common uses of line charts



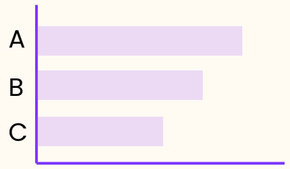
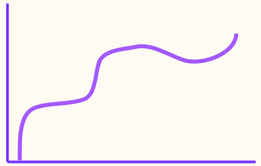
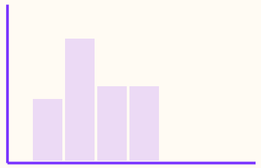
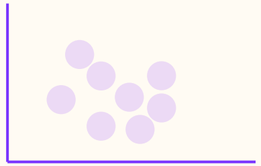
Always work back from your data

COMMONLY USED DATA	PROBLEM DESCRIPTION	MOST USEFUL VISUALIZATION
• Banking products and associated customers • Branch cost data broken down into different verticals • Visualizing different customer segments	Showing comparisons of different categories	 **BAR CHARTS**
• Stock price change over time • Number of banking app users over time • Number of customer support tickets over time	Showing changes of a variable over time	 **LINE CHARTS**
• The distribution of savings for customers in savings accounts • App opens for all customers in a given time period	Show the distribution of a variable over time	 **HISTOGRAMS**

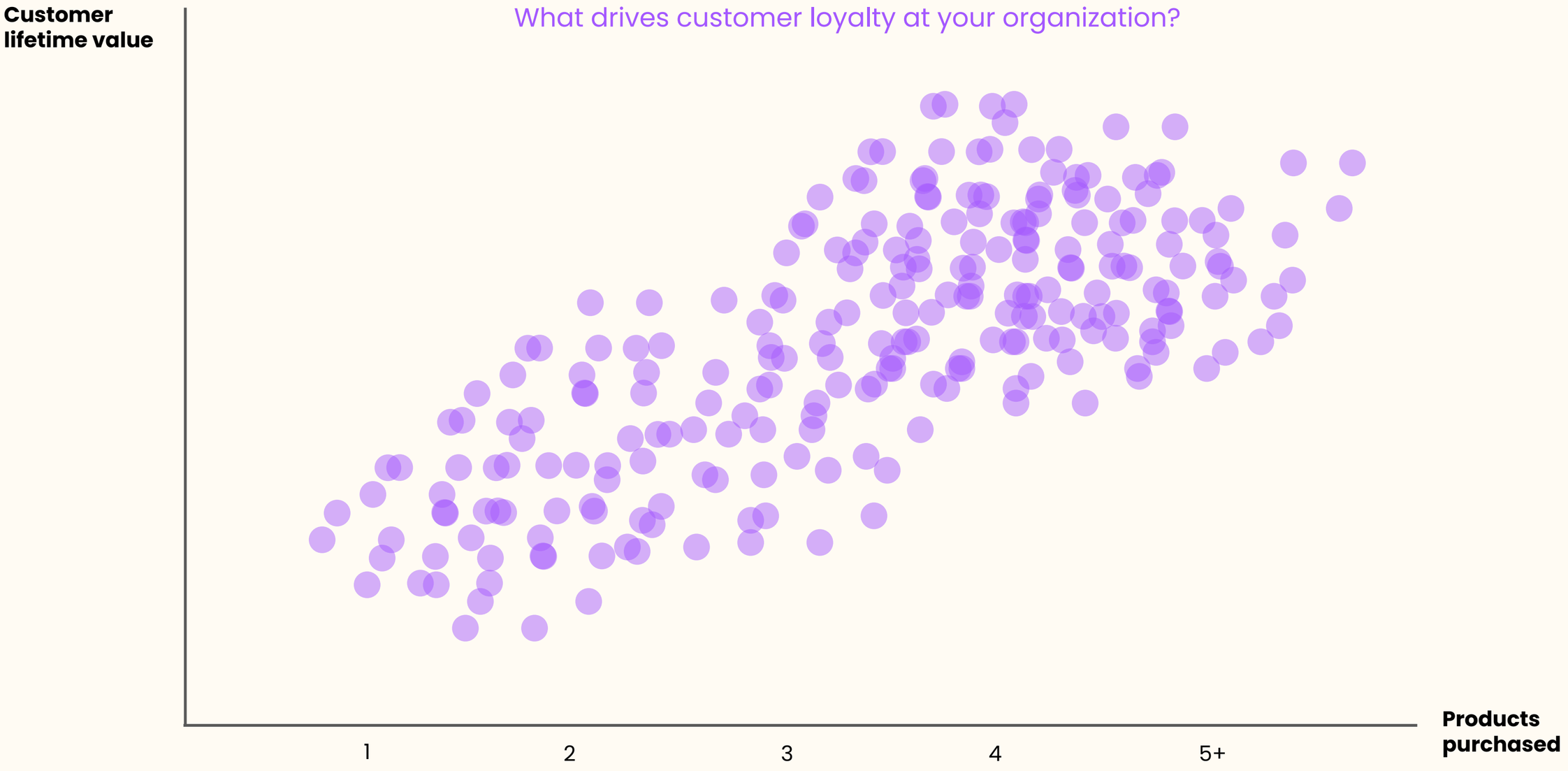
Common uses of histograms



Always work back from your data

COMMONLY USED DATA	PROBLEM DESCRIPTION	MOST USEFUL VISUALIZATION
• Banking products and associated customers • Branch cost data broken down into different verticals • Visualizing different customer segments	Showing comparisons of different categories	 **BAR CHARTS**
• Stock price change over time • Number of banking app users over time • Number of customer support tickets over time	Showing changes of a variable over time	 **LINE CHARTS**
• The distribution of savings for customers in savings accounts • Banking app opens for all customers in a given time period	Show the distribution of a variable over time	 **HISTOGRAMS**
• The relationship between historical credit scores and number of loans taken • The relationship between customer lifetime value and number of products purchased	Show the relationship between two variables	 **SCATTER PLOTS**

Common uses of scatter plots

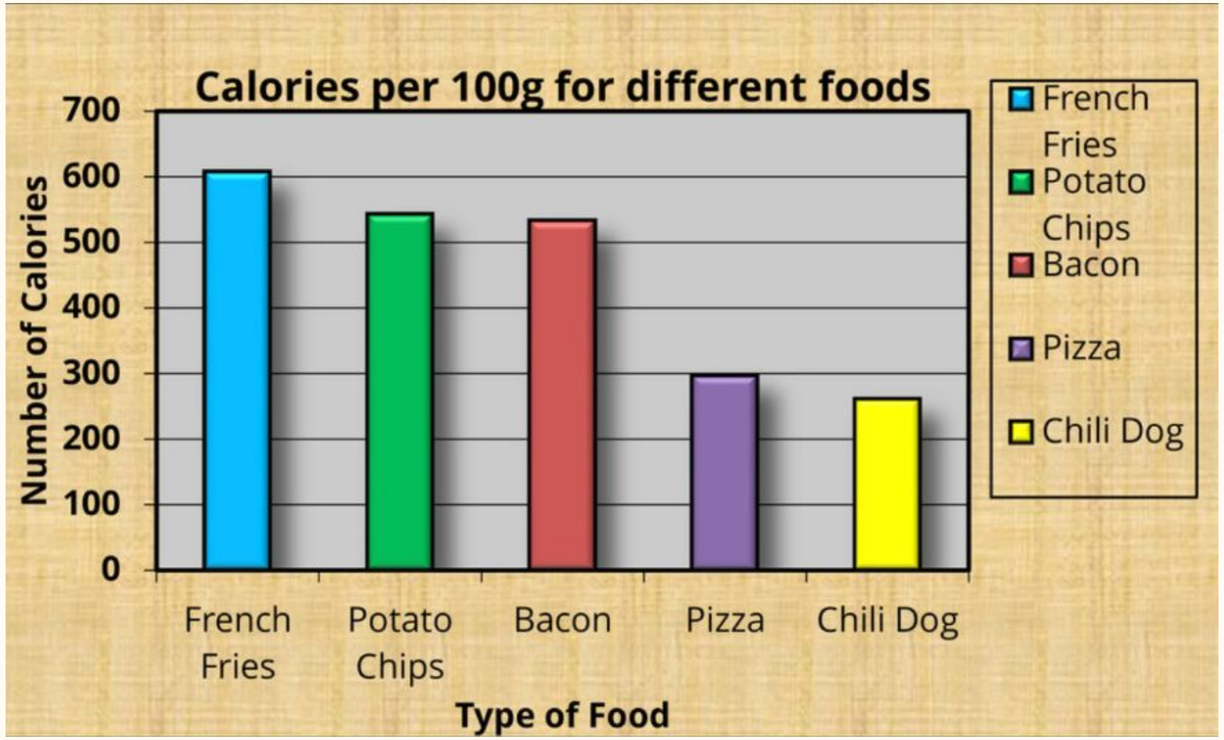


Rule #2

Keep visualizations minimal and avoid clutter

The cognitive load and effectiveness tradeoff

Each visualization serves a purpose. As a rule of thumb, remove all the elements of your visualization that doesn't serve a purpose.

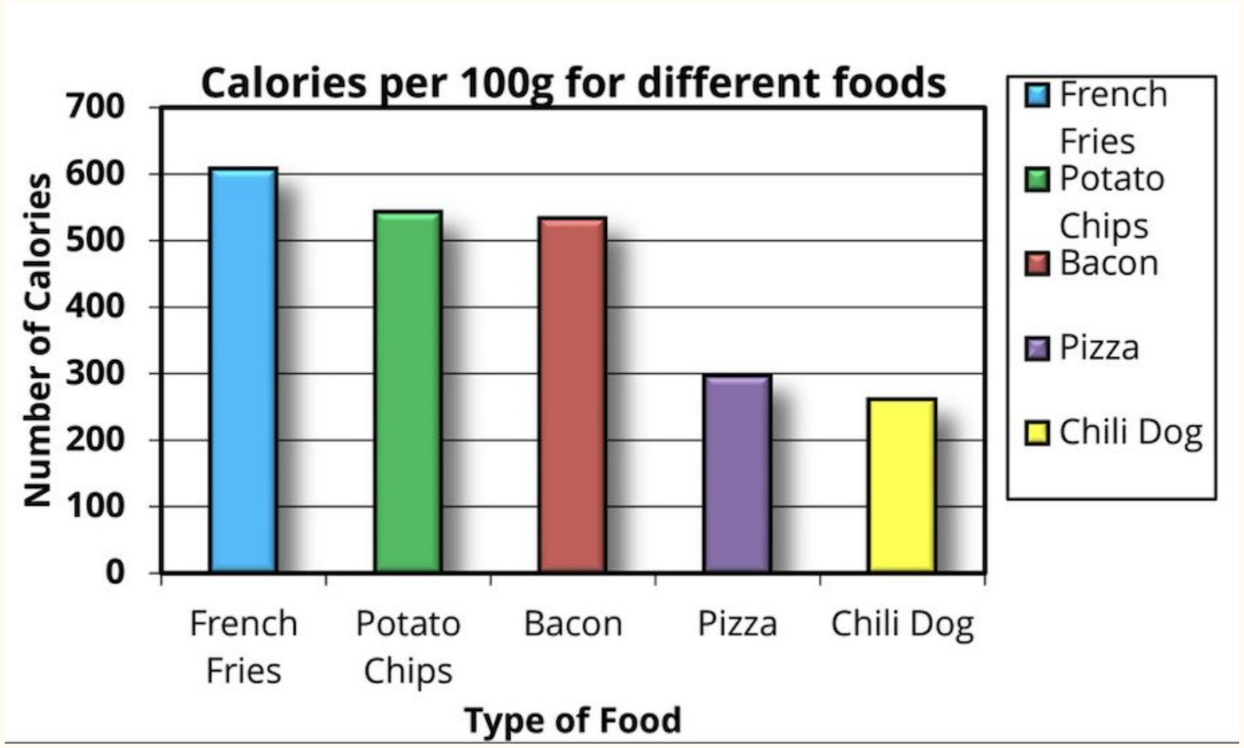


Source: Darkhorse Analytics

The cognitive load and effectiveness tradeoff

Each visualization serves a purpose. As a rule of thumb, remove all the elements of your visualization that doesn't serve a purpose.

Clutter removal technique #1 — Remove backgrounds

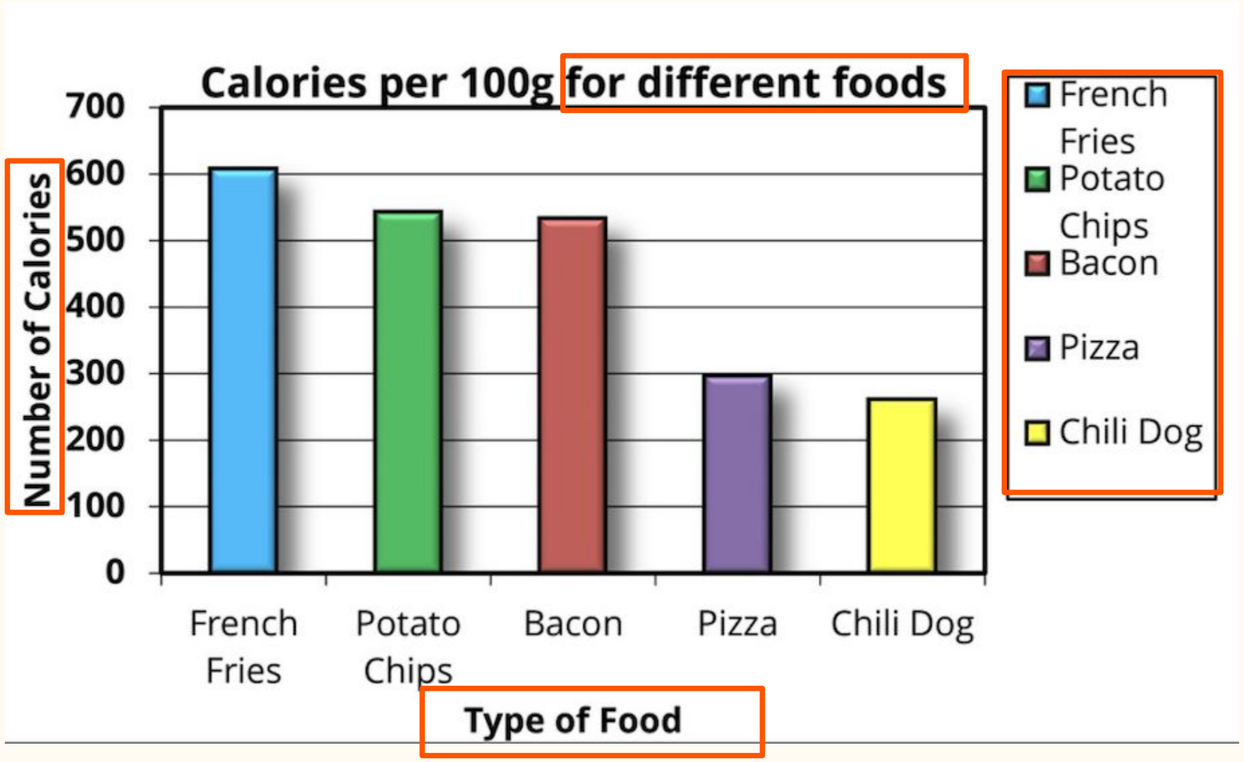


Source: Darkhorse Analytics

The cognitive load and effectiveness tradeoff

Each visualization serves a purpose. As a rule of thumb, remove all the elements of your visualization that doesn't serve a purpose.

Clutter removal technique #2 — Remove redundant labels

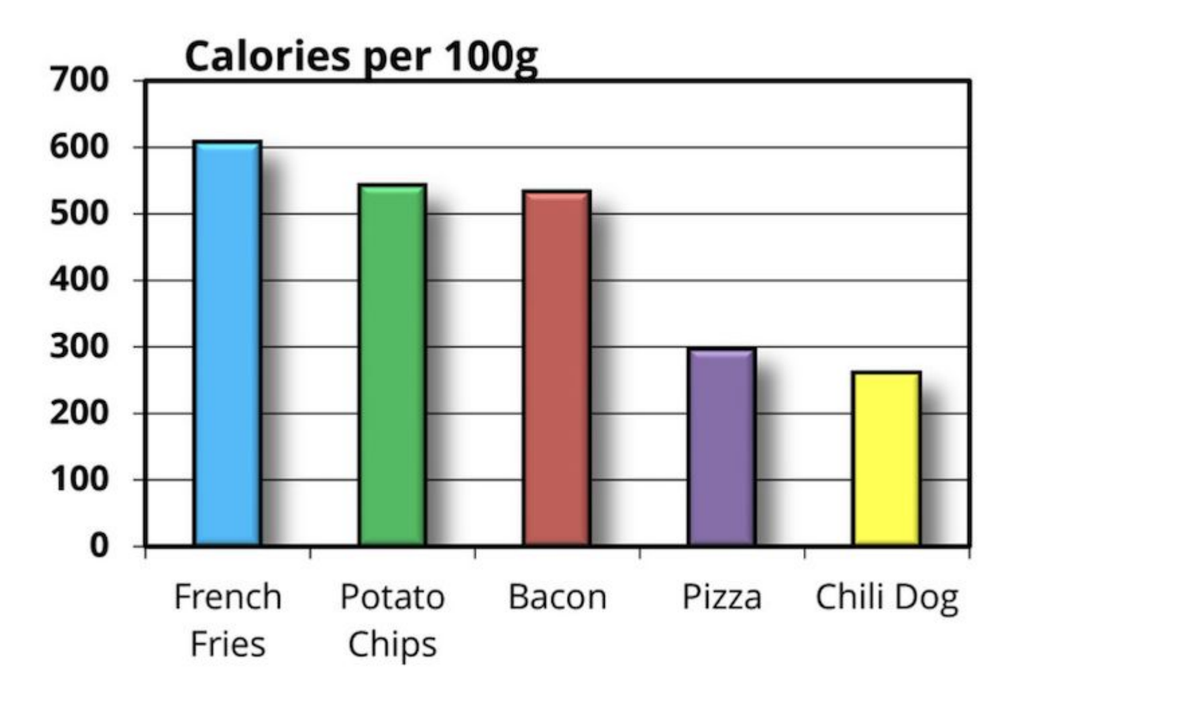


Source: Darkhorse Analytics

The cognitive load and effectiveness tradeoff

Each visualization serves a purpose. As a rule of thumb, remove all the elements of your visualization that doesn't serve a purpose.

Clutter removal technique #2 — Remove redundant labels

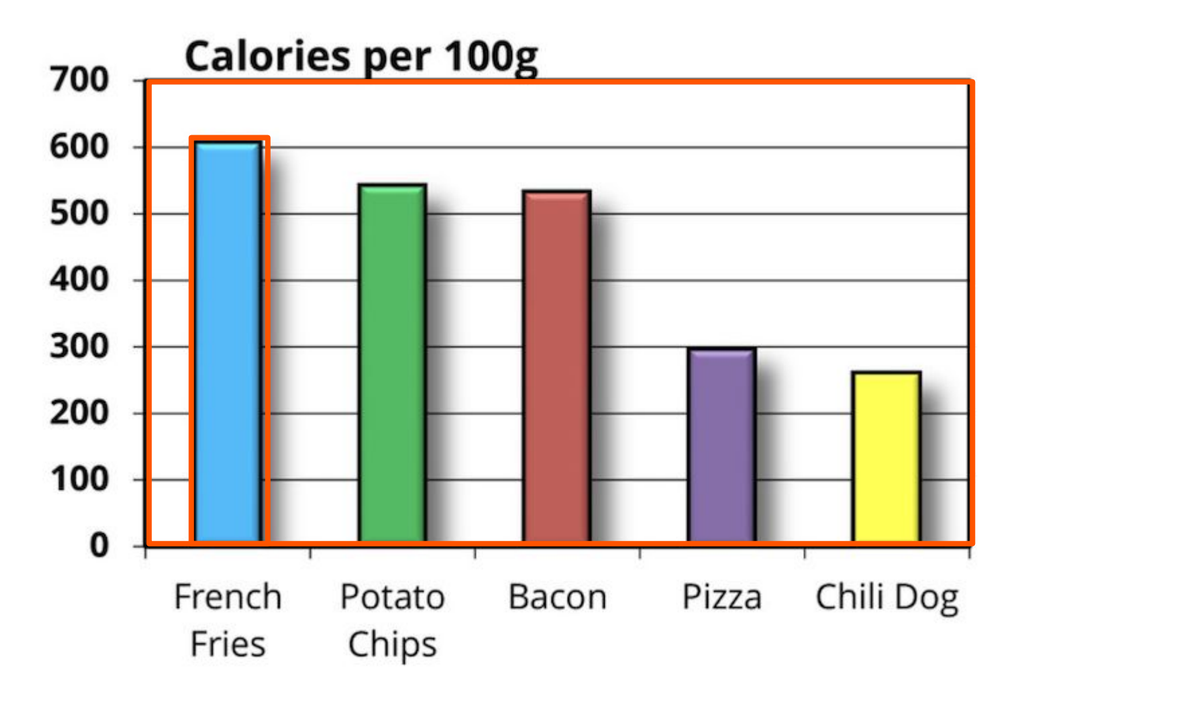


Source: Darkhorse Analytics

The cognitive load and effectiveness tradeoff

Each visualization serves a purpose. As a rule of thumb, remove all the elements of your visualization that doesn't serve a purpose.

Clutter removal technique #3 — Remove borders

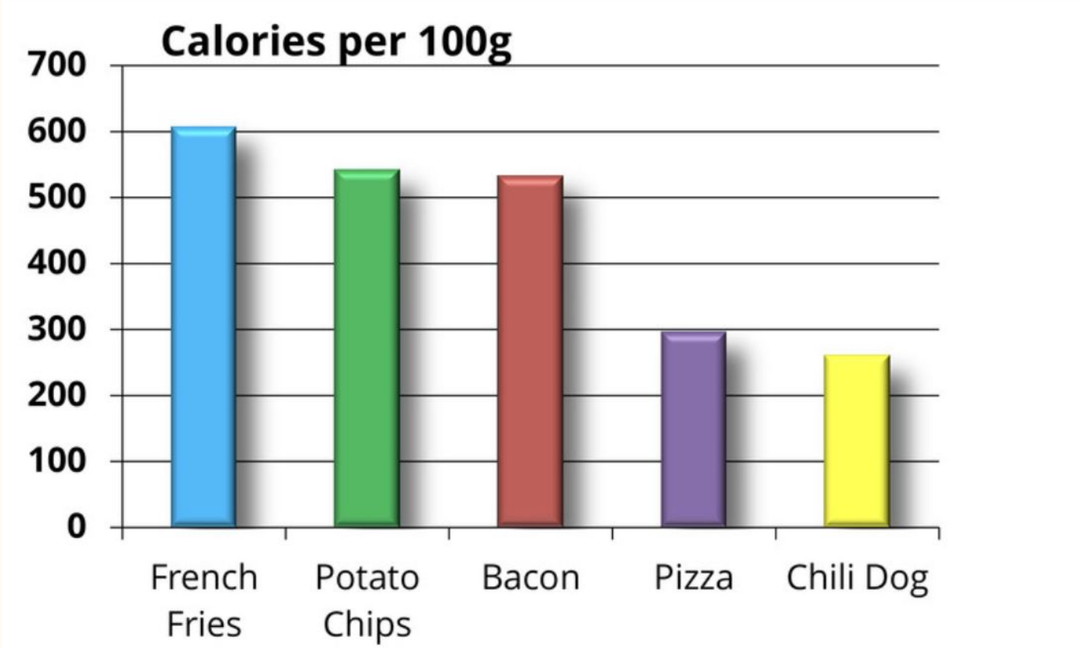


Source: Darkhorse Analytics

The cognitive load and effectiveness tradeoff

Each visualization serves a purpose. As a rule of thumb, remove all the elements of your visualization that doesn't serve a purpose.

Clutter removal technique #3 — Remove borders

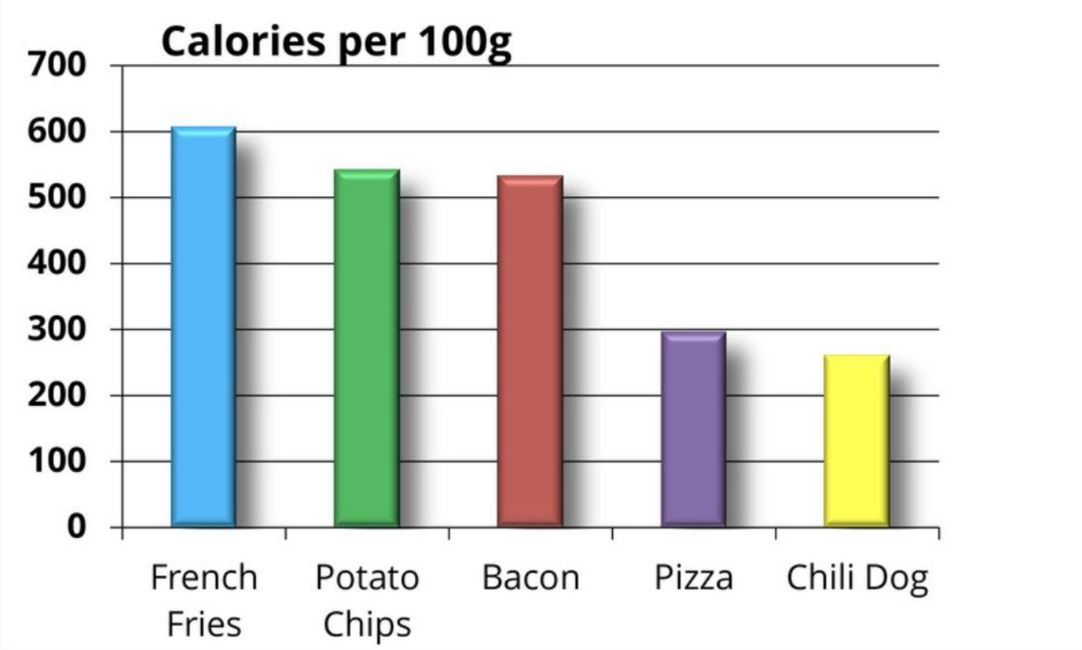


Source: Darkhorse Analytics

The cognitive load and effectiveness tradeoff

Each visualization serves a purpose. As a rule of thumb, remove all the elements of your visualization that doesn't serve a purpose.

Clutter removal technique #4 — Reduce colors when it doesn't matter

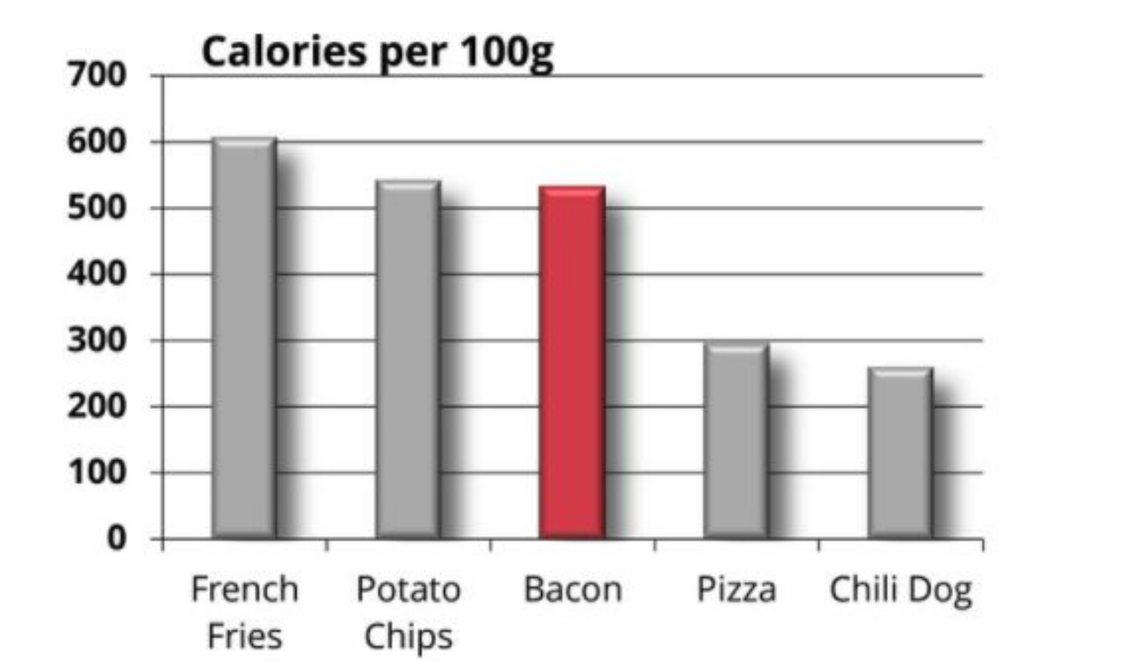


Source: Darkhorse Analytics

The cognitive load and effectiveness tradeoff

Each visualization serves a purpose. As a rule of thumb, remove all the elements of your visualization that doesn't serve a purpose.

Clutter removal technique #4 — Reduce colors when it doesn't matter

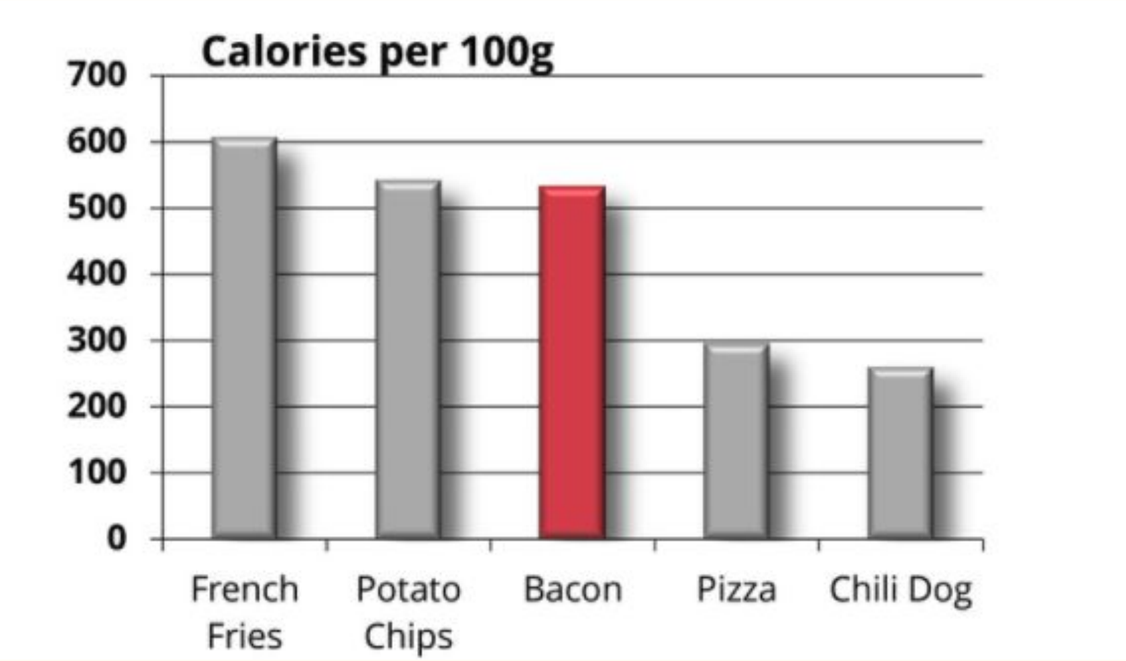


Source: Darkhorse Analytics

The cognitive load and effectiveness tradeoff

Each visualization serves a purpose. As a rule of thumb, remove all the elements of your visualization that doesn't serve a purpose.

Clutter removal technique #5 — Remove needless effects

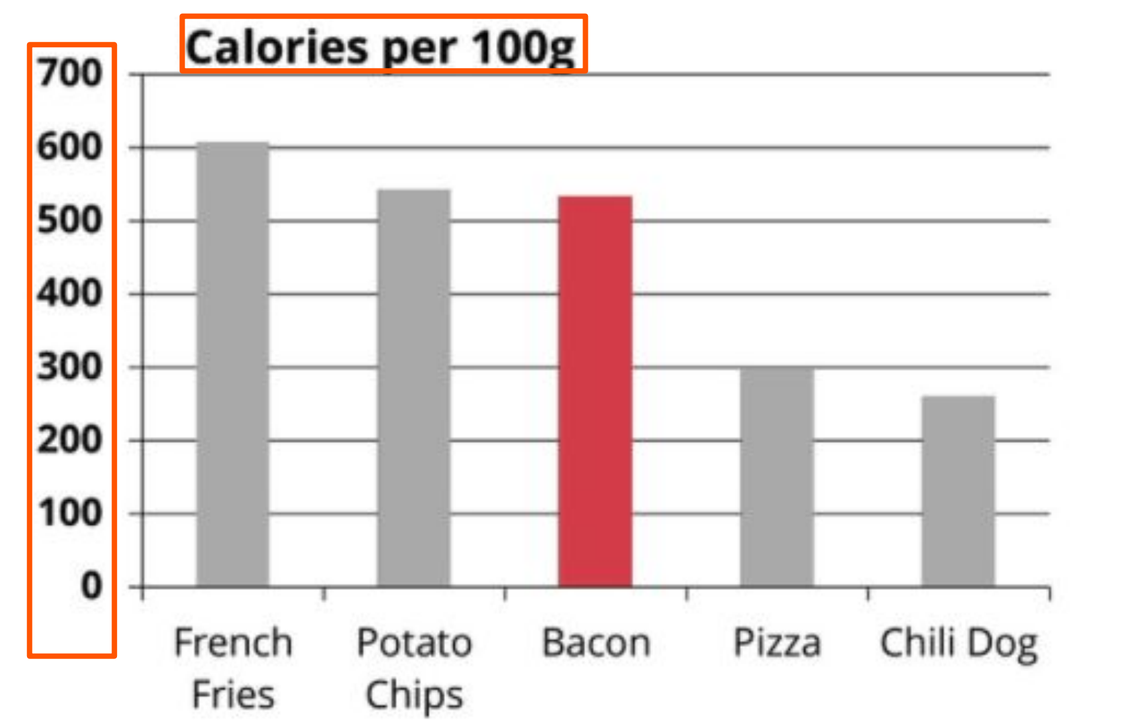


Source: Darkhorse Analytics

The cognitive load and effectiveness tradeoff

Each visualization serves a purpose. As a rule of thumb, remove all the elements of your visualization that doesn't serve a purpose.

Clutter removal technique #5 — Remove needless effects

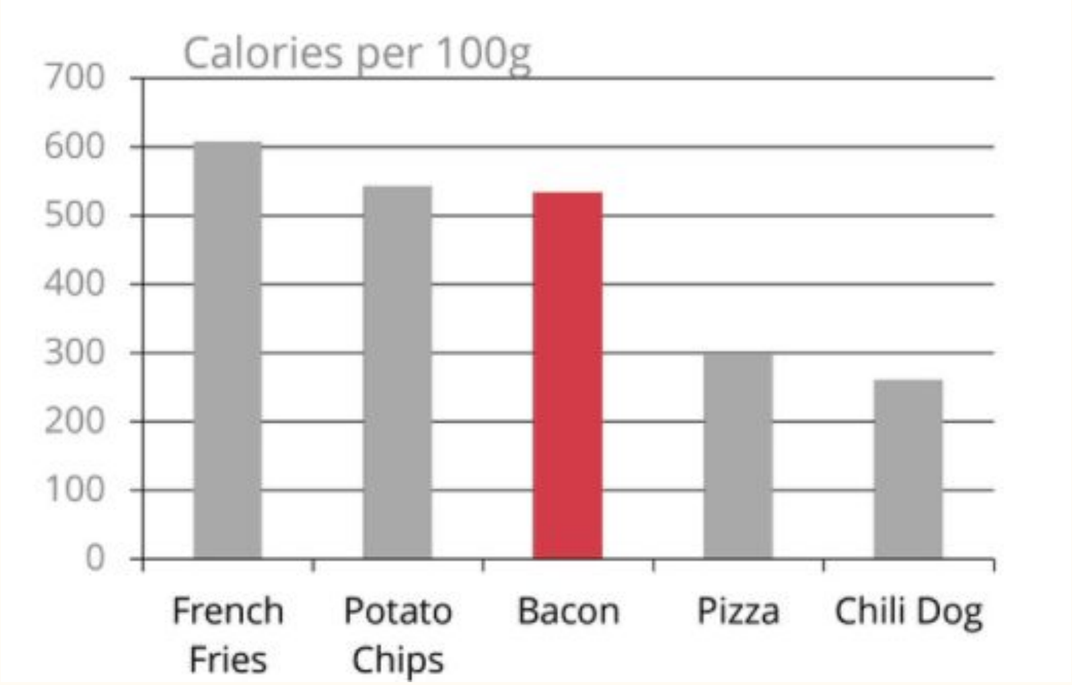


Source: Darkhorse Analytics

The cognitive load and effectiveness tradeoff

Each visualization serves a purpose. As a rule of thumb, remove all the elements of your visualization that doesn't serve a purpose.

Clutter removal technique #5 — Remove needless effects

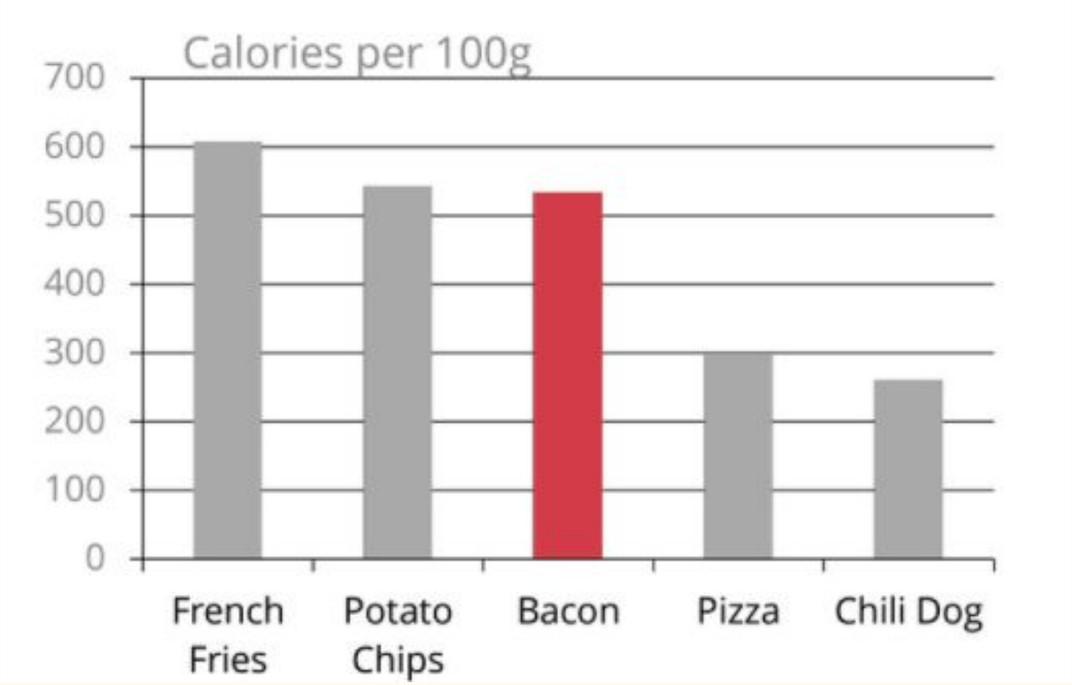


Source: Darkhorse Analytics

The cognitive load and effectiveness tradeoff

Each visualization serves a purpose. As a rule of thumb, remove all the elements of your visualization that doesn't serve a purpose.

Clutter removal technique #6 — Remove axis lines when not needed

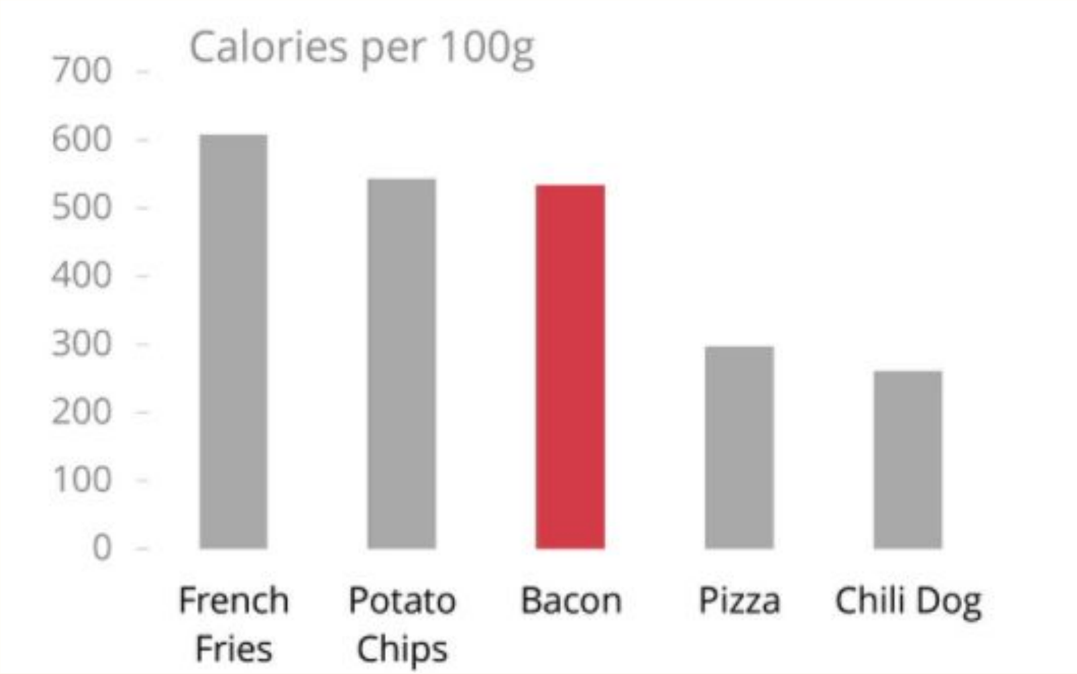


Source: Darkhorse Analytics

The cognitive load and effectiveness tradeoff

Each visualization serves a purpose. As a rule of thumb, remove all the elements of your visualization that doesn't serve a purpose.

Clutter removal technique #6 — Remove axis lines when not needed

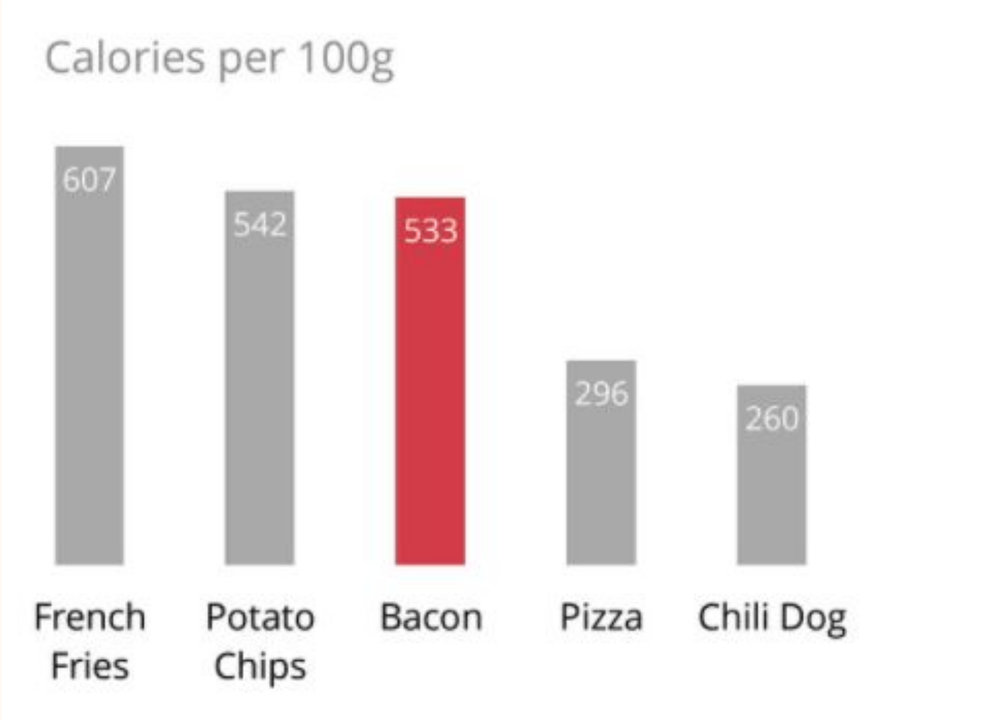


Source: Darkhorse Analytics

The cognitive load and effectiveness tradeoff

Each visualization serves a purpose. As a rule of thumb, remove all the elements of your visualization that doesn't serve a purpose.

Clutter removal technique #7 — Add labels directly on the plot



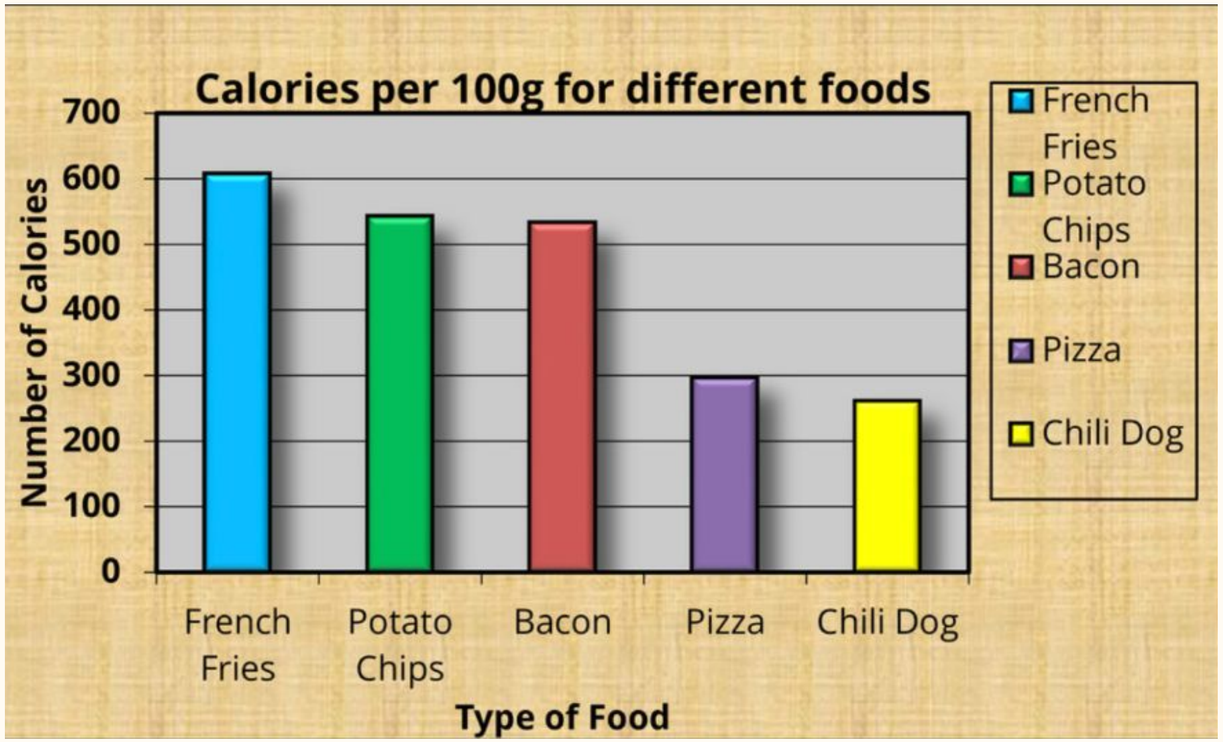
Source: Darkhorse Analytics

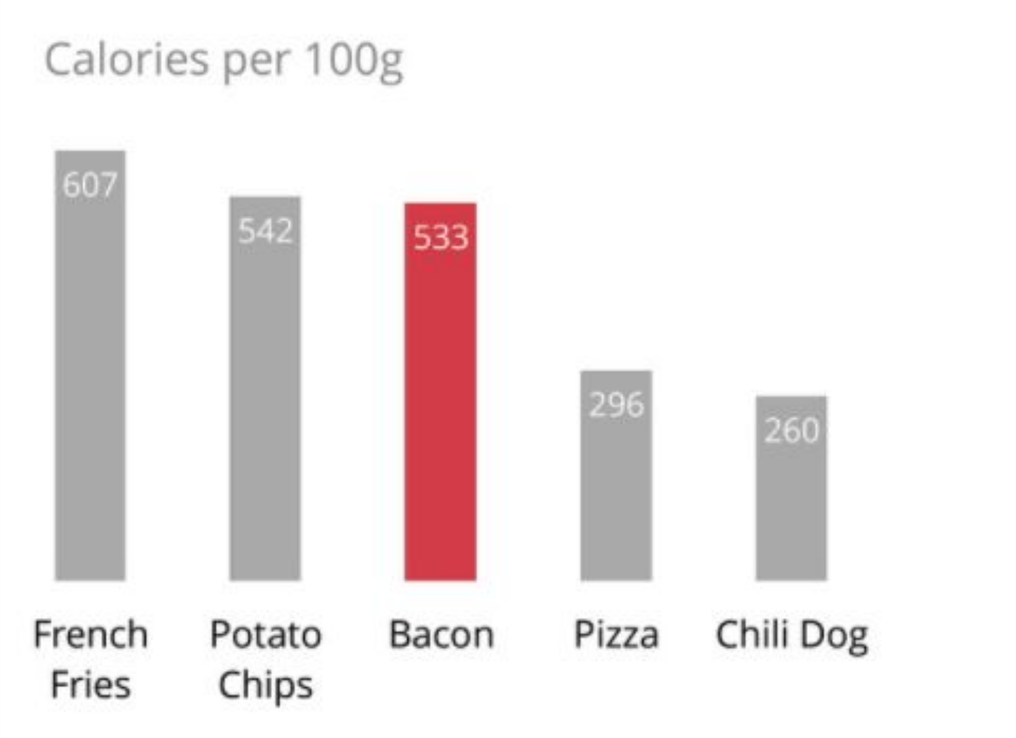
Decluttering techniques at your disposal

- ✓ Use white spaces
- ✓ Remove chart borders
- ✓ Remove gridlines or axes
- ✓ Clean up axis labels
- ✓ Label data directly (as opposed to using a legend)
- ✓ Remove data markers
- ✓ Use special effects (bold, underline, italics, shadows) sparingly

The cognitive load and effectiveness tradeoff

Each visualization serves a purpose. As a rule of thumb, remove all the elements of your visualization that doesn't serve a purpose.



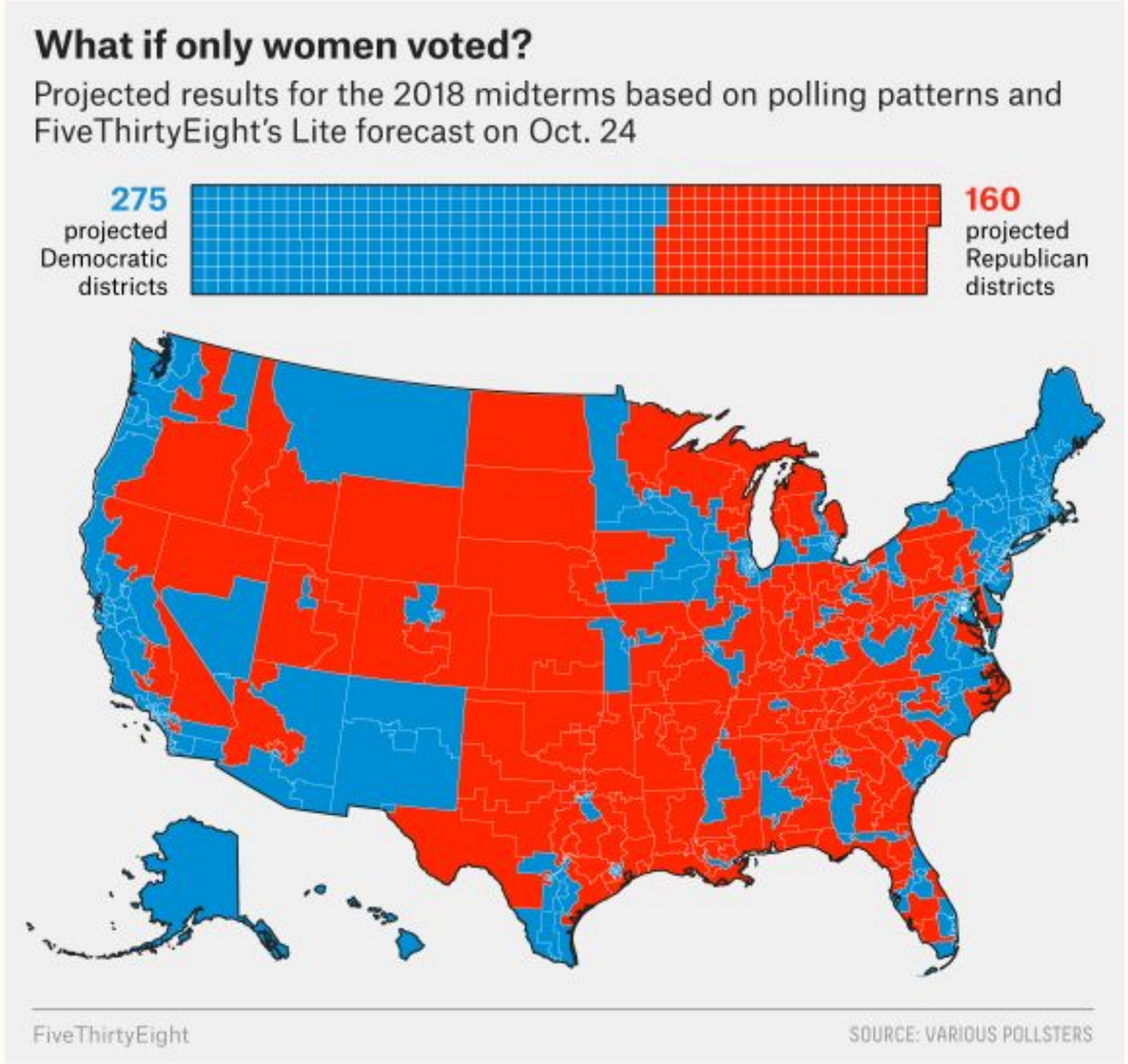


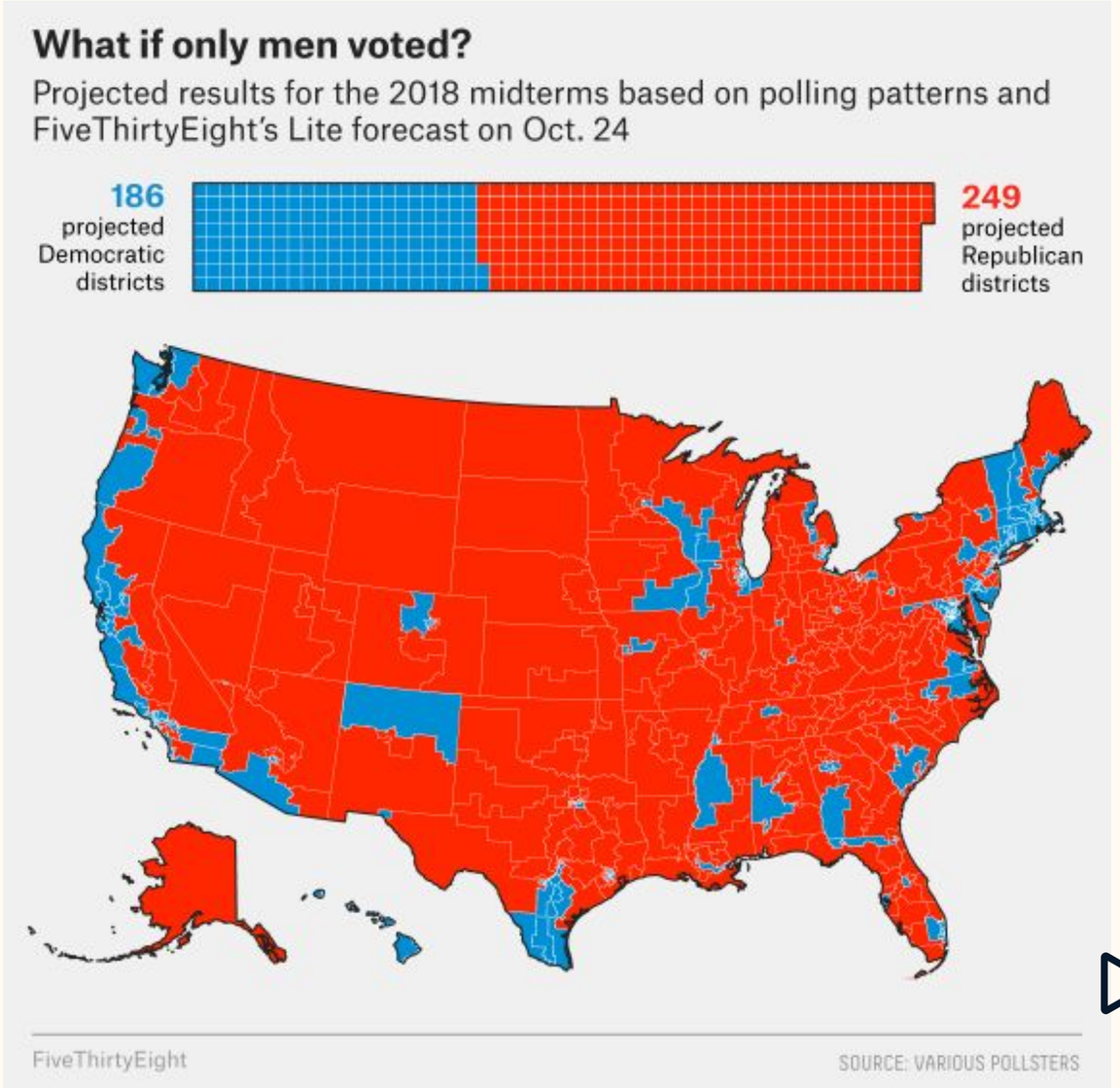
Source: Darkhorse Analytics

Rule #3

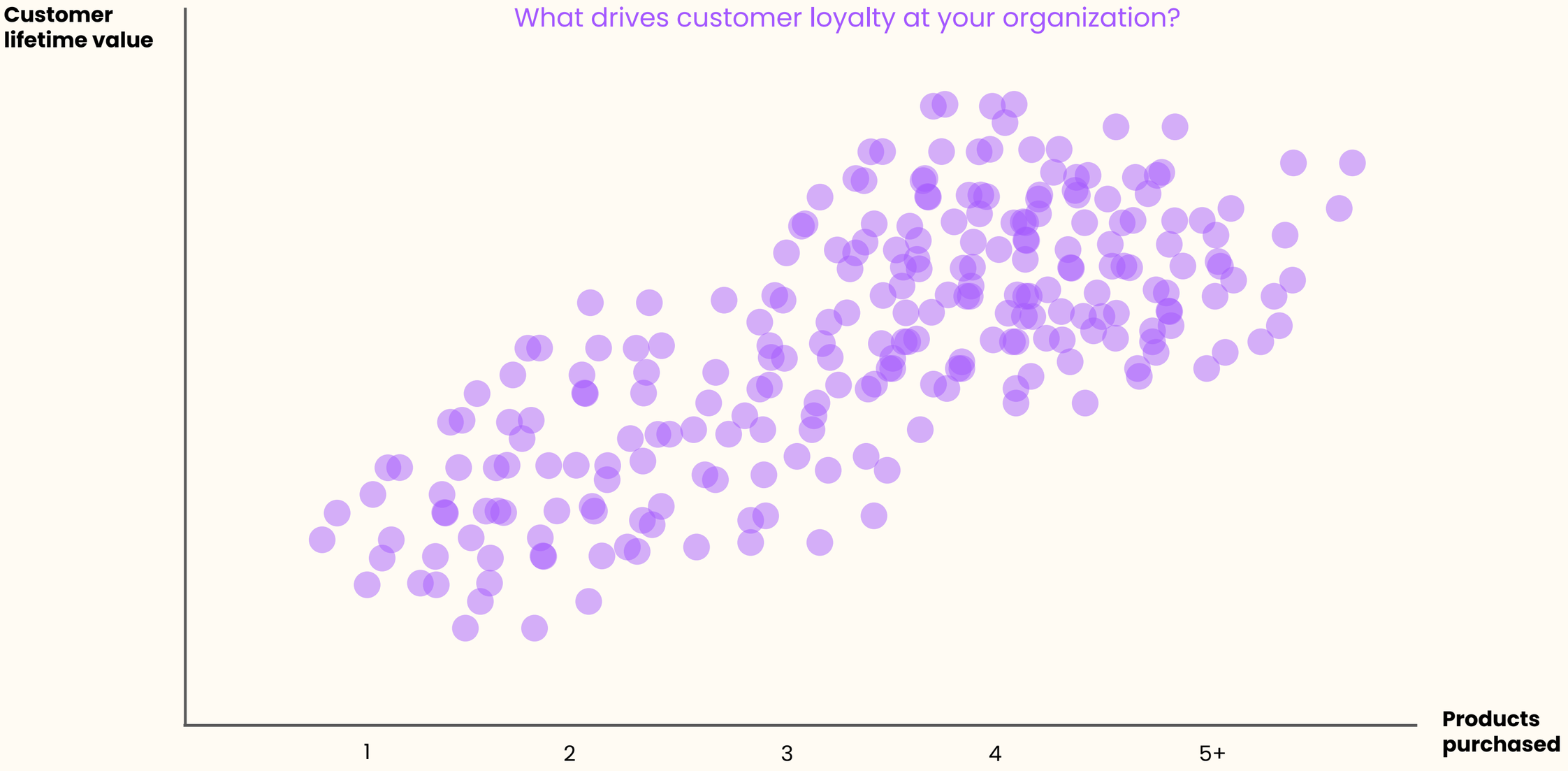
Use color creatively

Colors to distinguish between groups

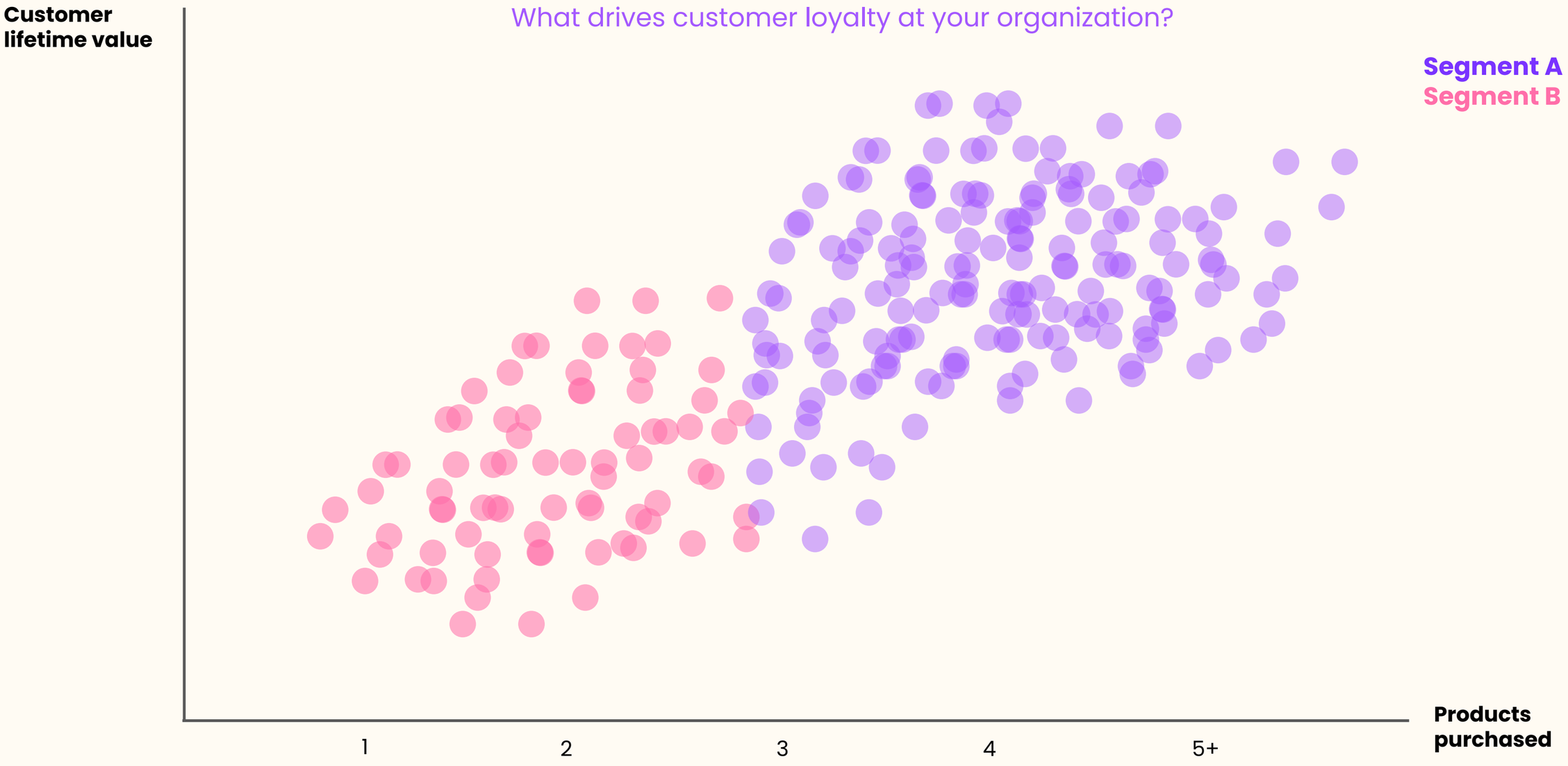




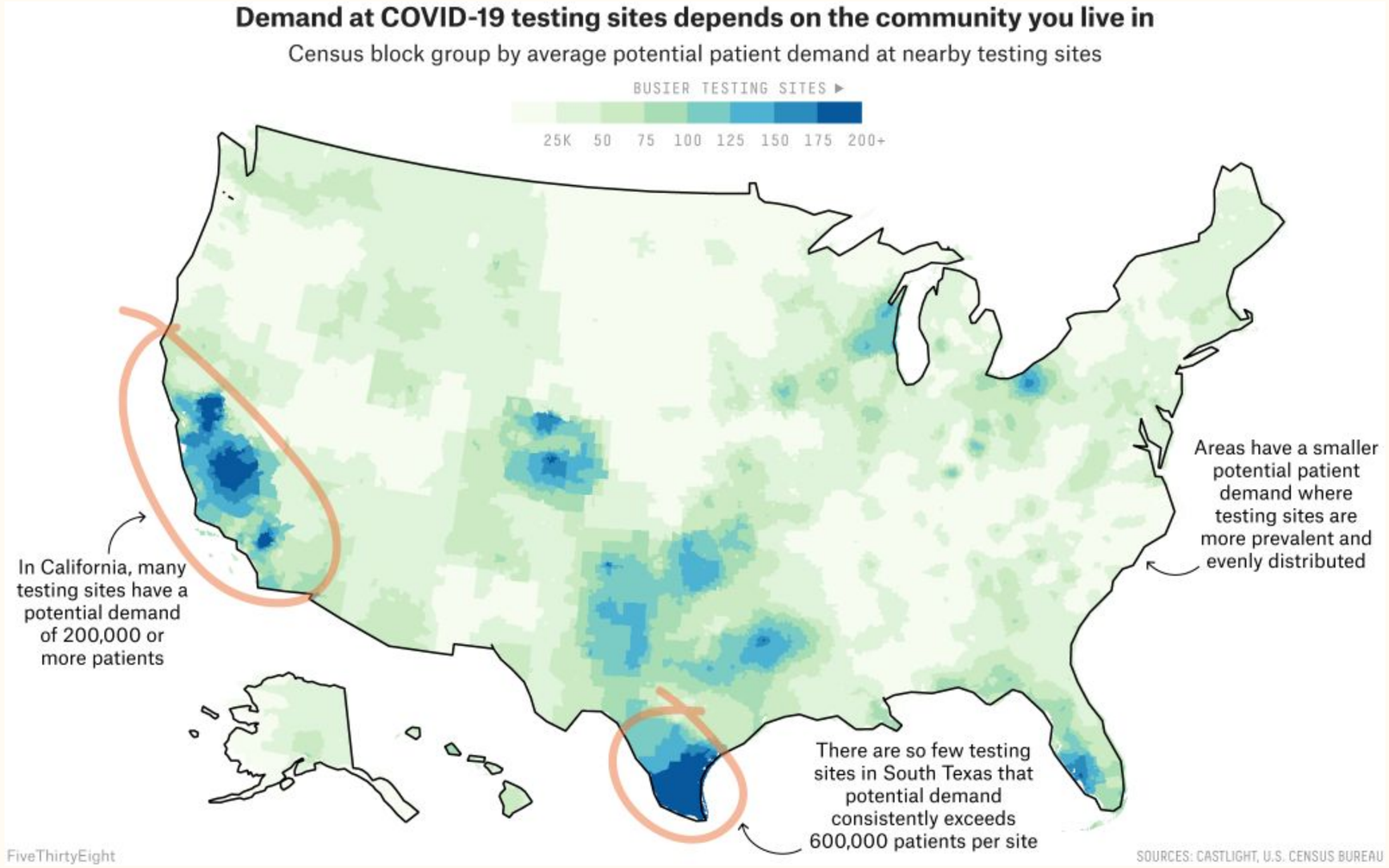
How this can be used in “normal” datasets



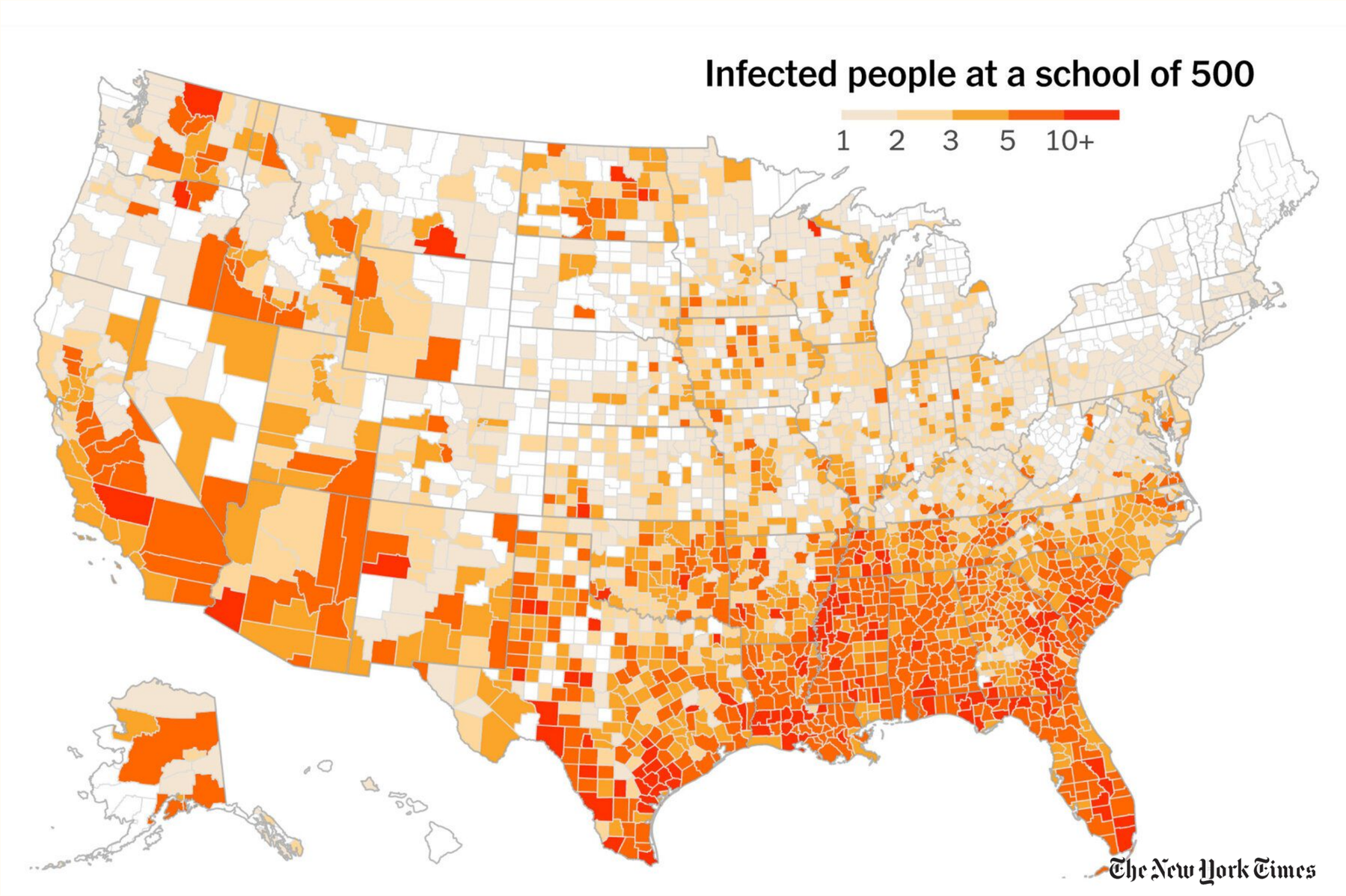
How this can be used in “normal” datasets



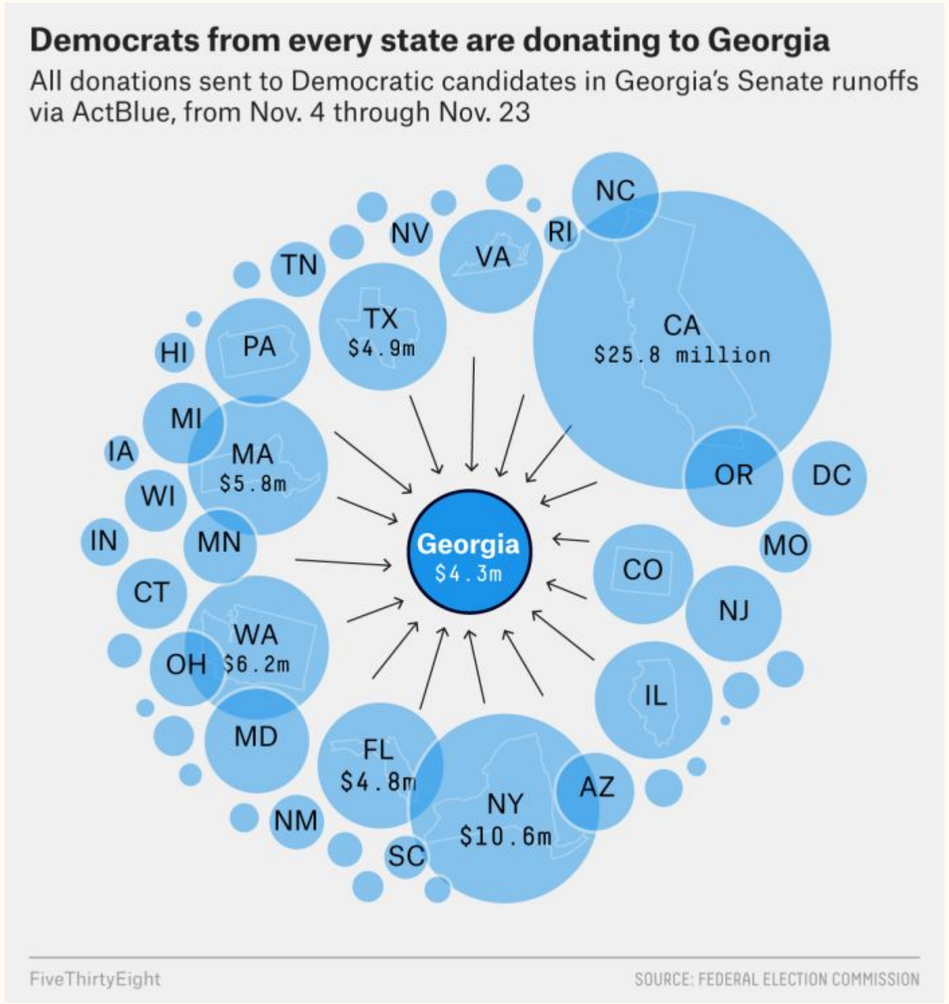
Colors to highlight intensity



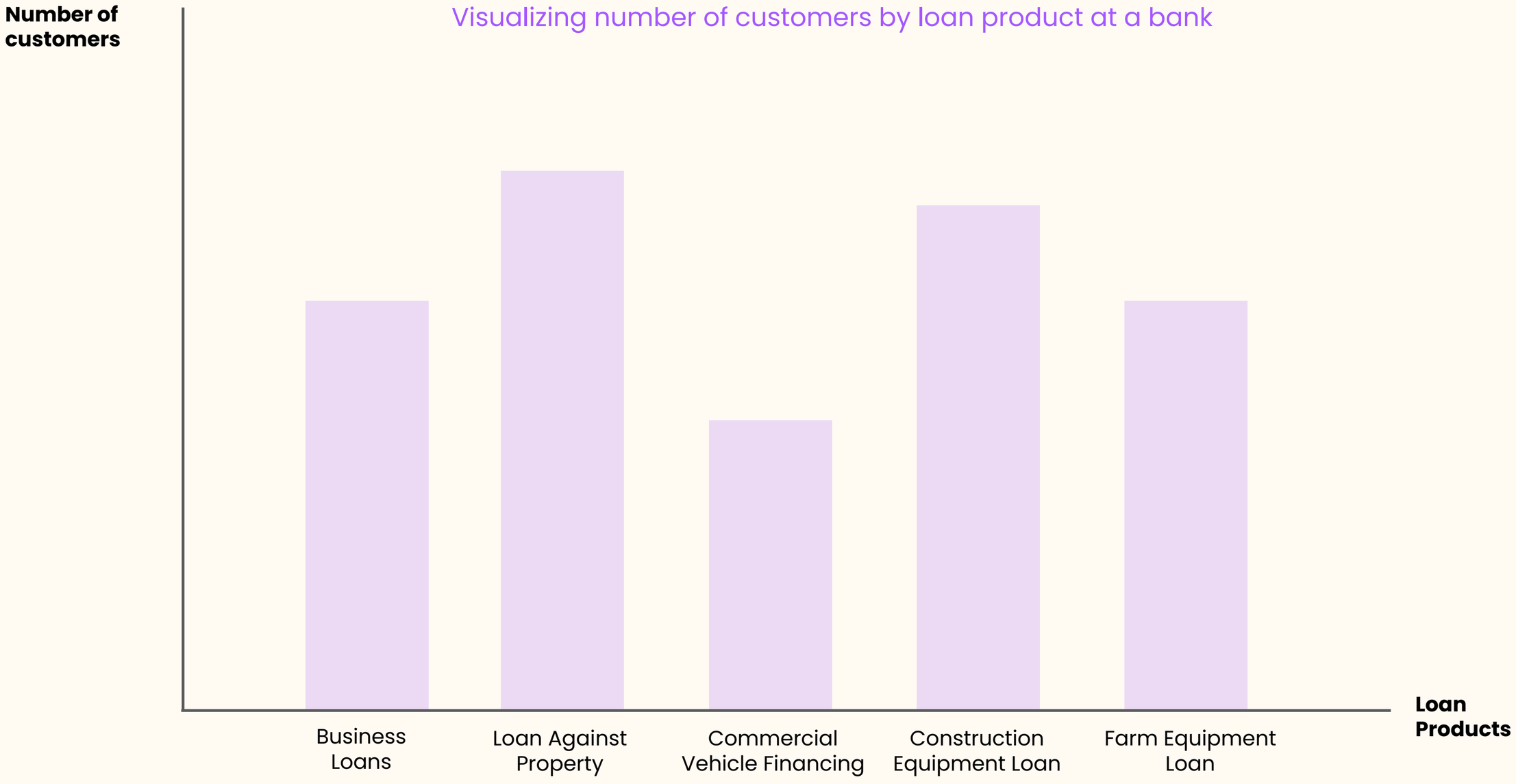
Visualizing covid infection hotspots



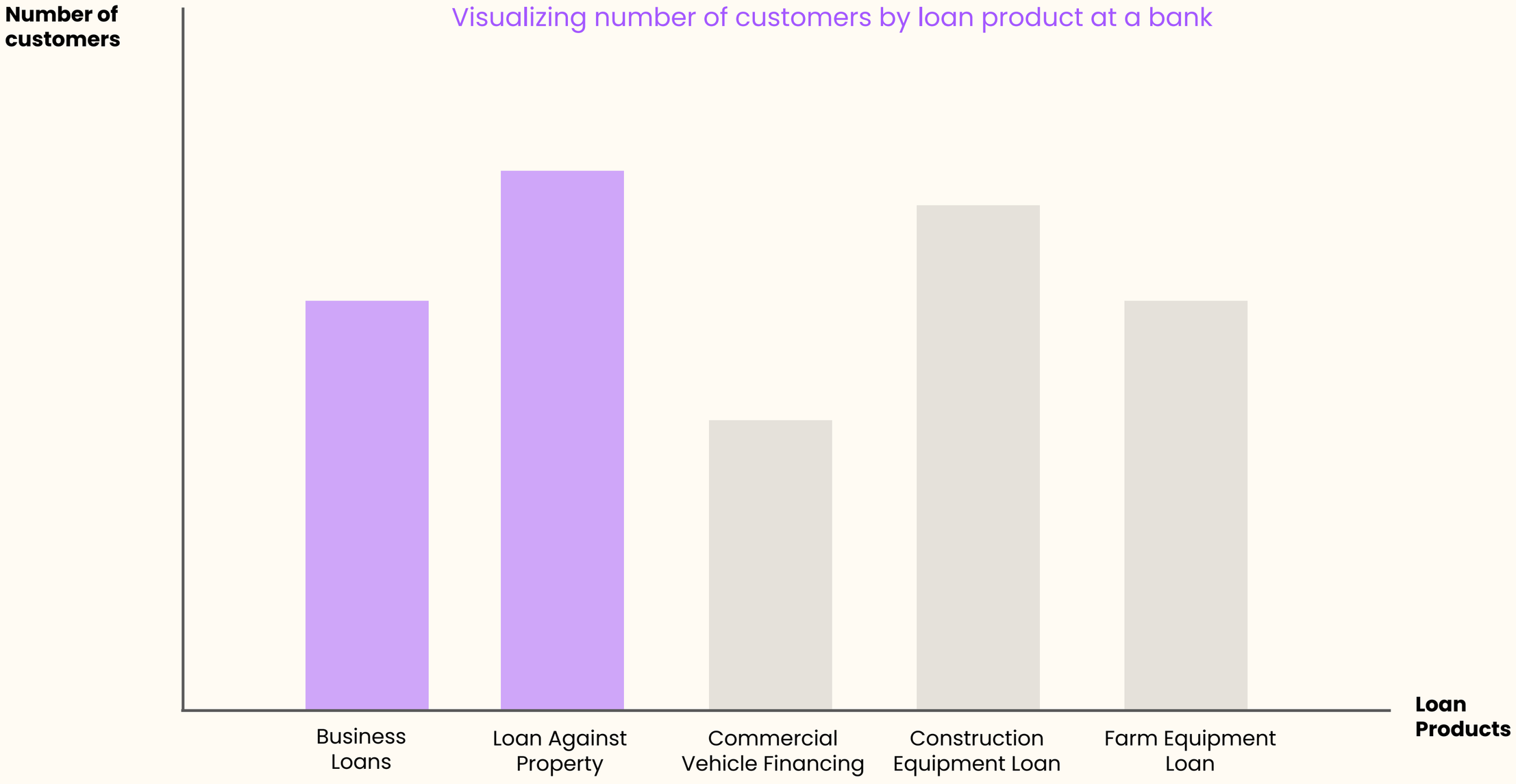
Colors to distinguish between groups



How this can be used in “normal” datasets



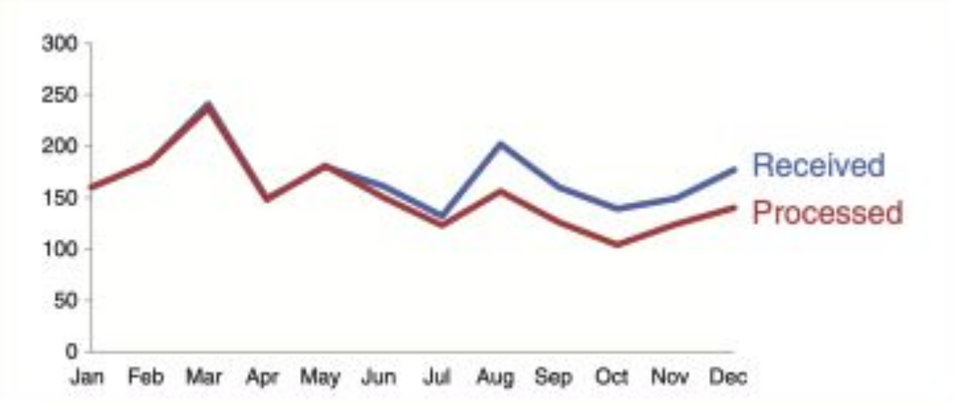
How this can be used in “normal” datasets



Rule #4

Use texts appropriately

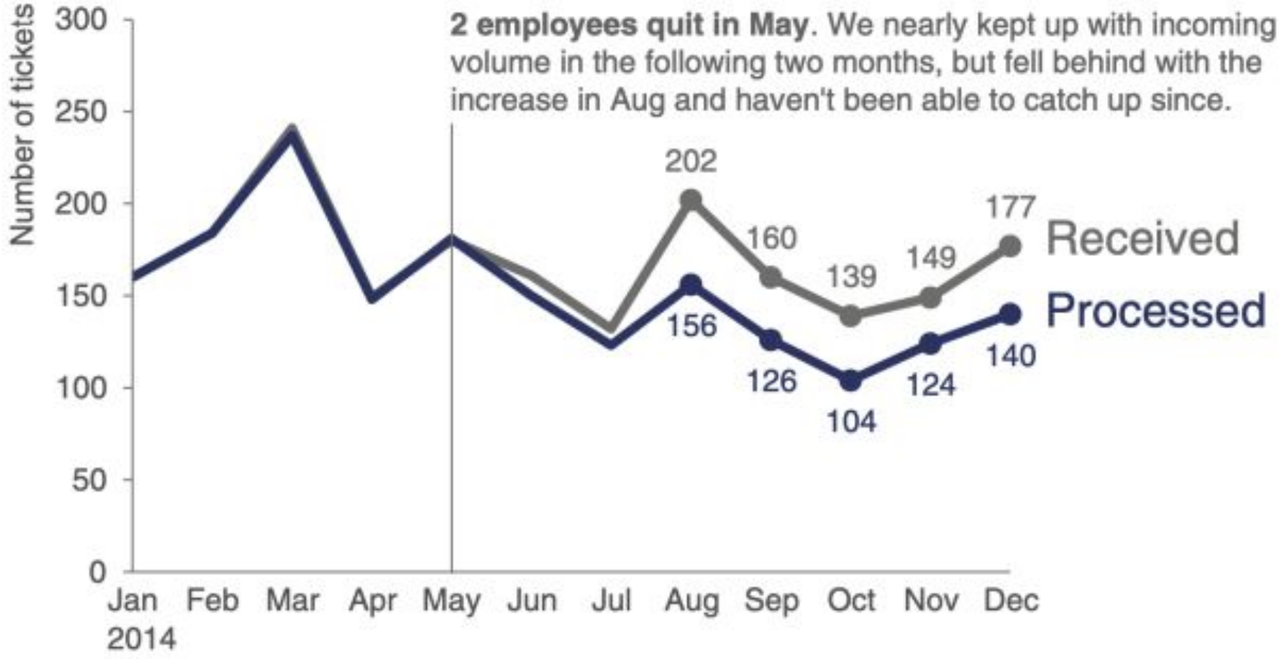
Labels can be extremely effective at highlighting insights



Please approve the hire of 2 FTEs

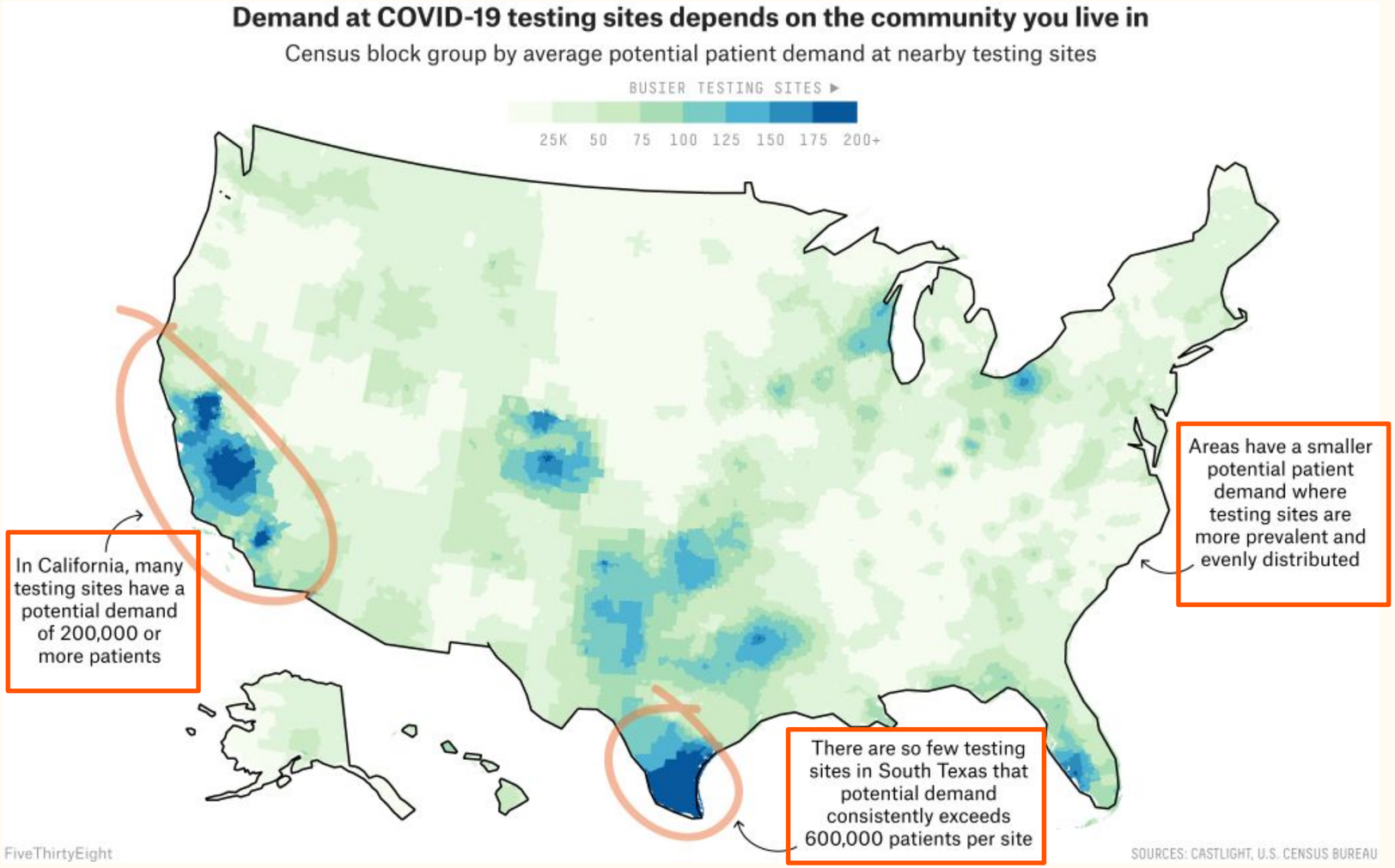
to backfill those who quit in the past year

Ticket volume over time



Data source: XYZ Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.

Labels can be extremely effective at showing context



Rules of thumb when using labels

- ✓ Label axes and titles for clarity
- ✓ Label data points when necessary
- ✓ Play around with font sizes when highlighting specific message
- ✓ Common audience questions should go into labels

2

8 rules for better data storytelling

Rules for better narrative

Rule #1

Know your audience, know your format

We have different stakeholders, and different formats

AUDIENCE TYPES

Executive
Low data literacy
Cares about outcomes and decisions

Data Science Leader or Partner
Data expert
Cares about rigour and insights

Business Partner
High data literacy
Cares about tactical next steps

FORMAT TYPES

Presentation

Coding Notebook

Written Report

Know the priority of the audience

AUDIENCE TYPES

Executive
Low data literacy
Cares about outcomes and decisions

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

Data Science Leader or Partner
Data expert
Cares about rigour and insights

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

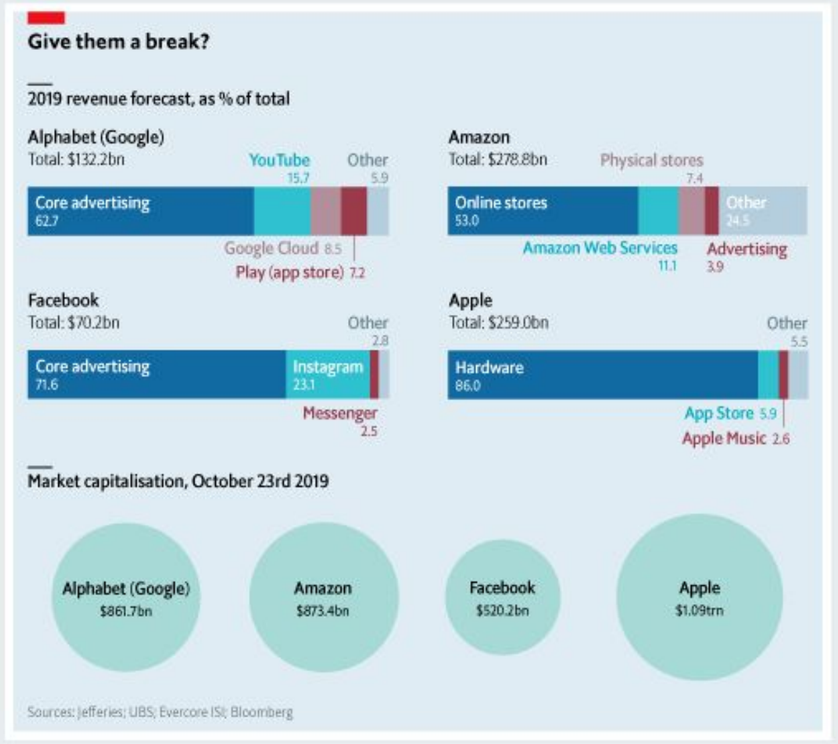
Business Partner
High data literacy
Cares about tactical next steps

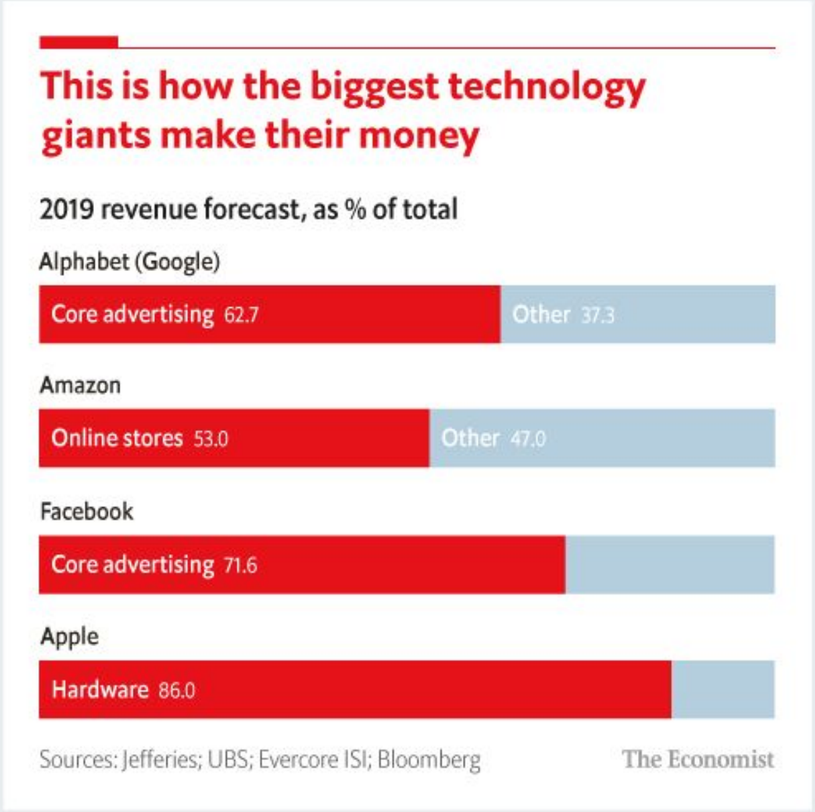
Cares much more about how your analysis impacts their workflow, and what should be their main takeaway from the data story

Practice empathy with your audience

- ✓ Does the audience have the necessary prerequisite knowledge to understand a particular metric?
- ✓ How much time does the audience have to consume this data story?
- ✓ What is the medium of presentation (written/oral) that the audience prefers?

Know their level of data literacy (or subject matter expertise)





[How the economist altered its charts on instagram for a younger audience]

Rule #2

Begin with the goal in mind

It’s very tempting to throw everything you got at a data story

Start with the goal in mind

Who is the audience?
The more specific the audience, the easier it is to build a data story that resonates

What should the audience know?
Know your recommendations before you craft the data story

What data can you use to convey your point?
Identify the data that is relevant for this data story

Goal
Convince management of investing in additional resource on support team

Start with the goal in mind

Who is the audience?
The more specific the audience, the easier it is to build a data story that resonates

What should the audience know?
Know your recommendations before you craft the data story

What data can you use to convey your point?
Identify the data that is relevant for this data story

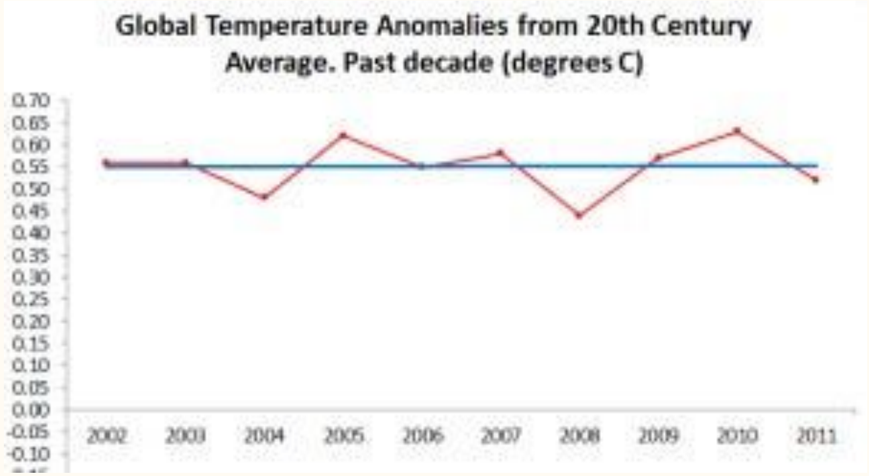
Goal
Convince management of investing in additional resource on support team

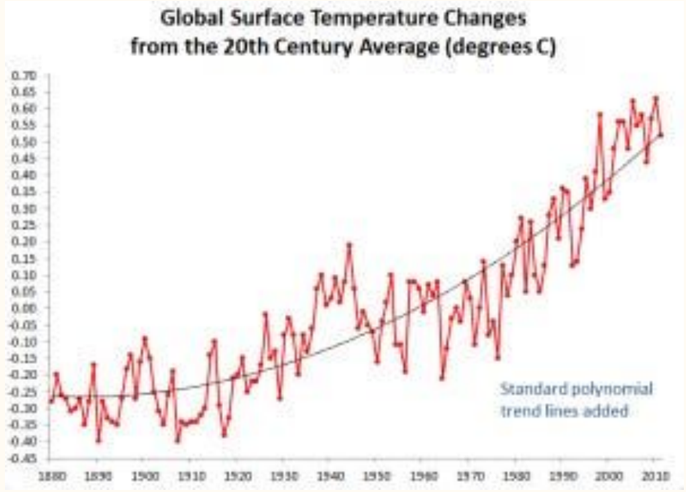
However, this doesn't mean that goals should determine the data story — what the data is telling you always takes precedence

Rule #3

Do not mislead with data stories

Avoiding the fastest way to lose credibility

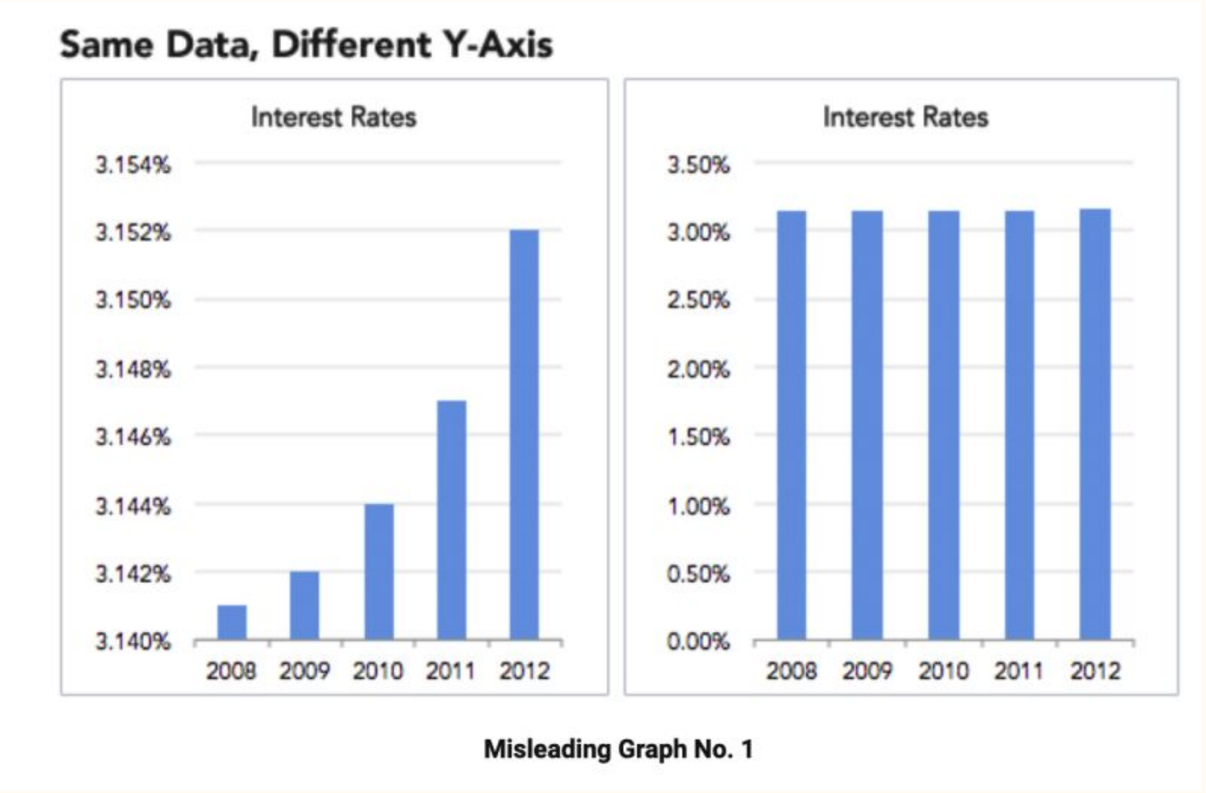




Pitfall to avoid #1

Always make sure your time horizons make sense given the data you're treating

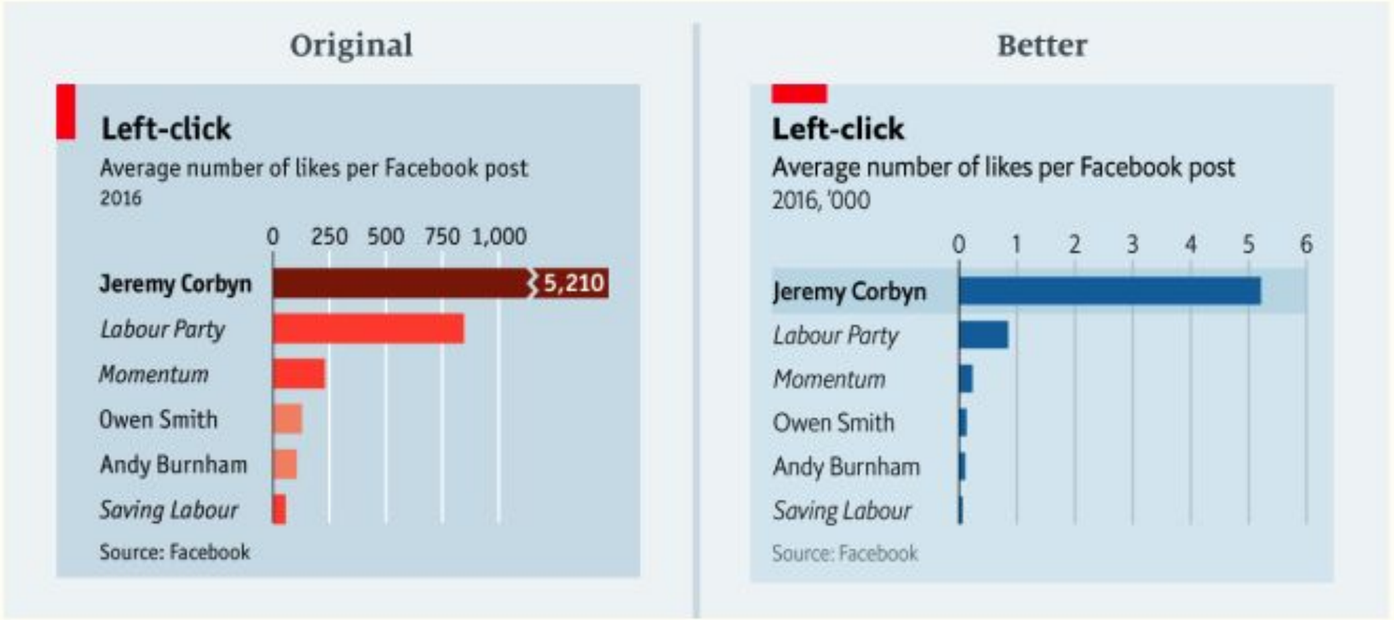
Avoiding the fastest way to lose credibility



Pitfall to avoid #2

Ensure that axes start with zero

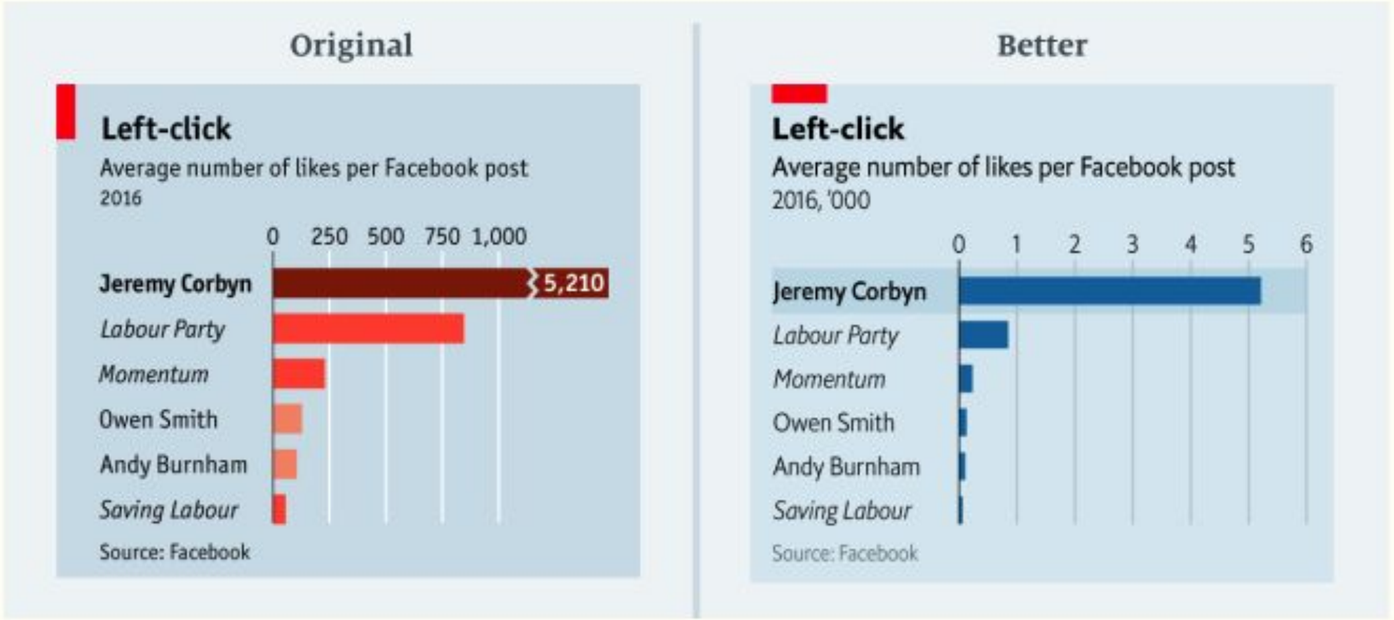
Avoiding the fastest way to lose credibility



Pitfall to avoid #3

Ensure that axes scales are appropriate

Avoiding the fastest way to lose credibility



Pitfall to avoid #3

Ensure that axes scales are appropriate

Other best practices to ensure data stories are not misleading

- ✓ If you’re sampling data, make sure sample is representative of population
- ✓ Use centrality measures (median, mean, etc…) to ensure context around a population is taken into account

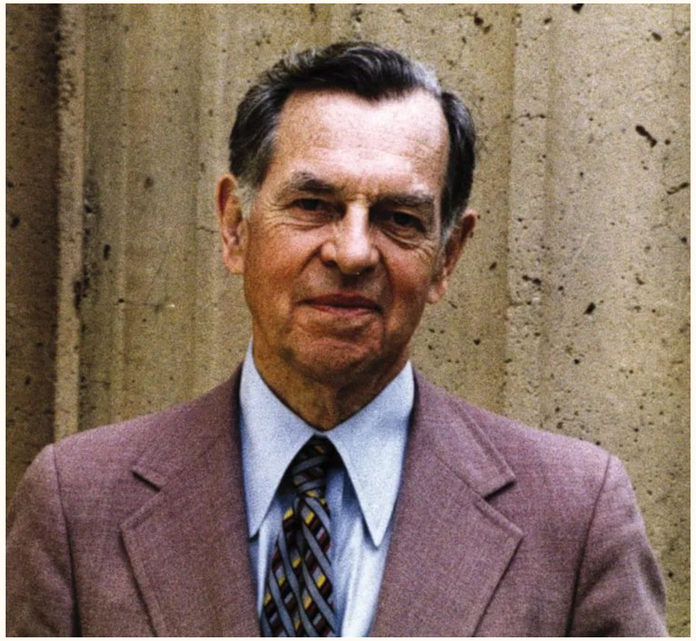
Rule #4

Develop a narrative around your data

Different narrative structures to choose from

Aristotle’s Tragedy Structure

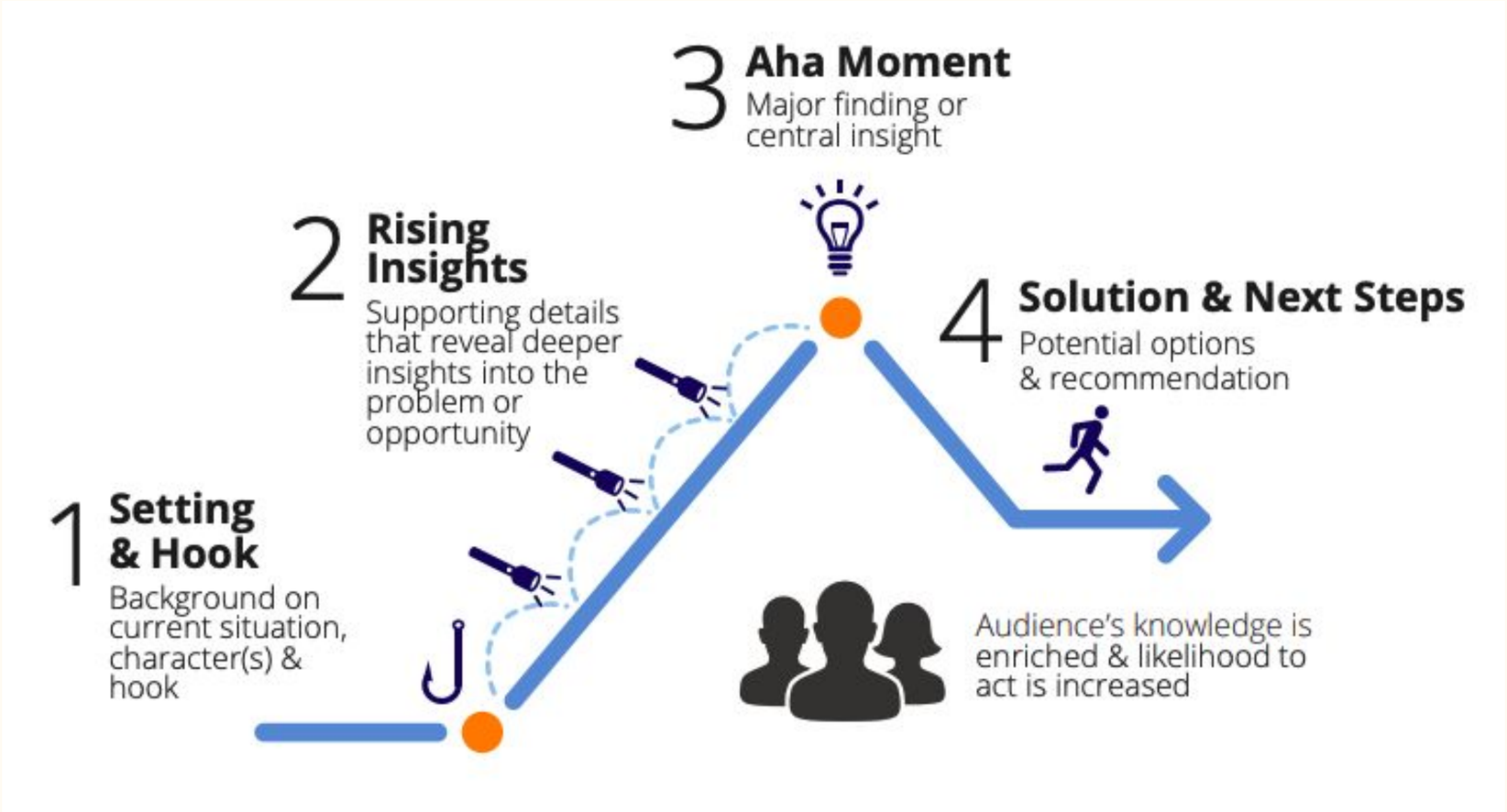
Freytag’s Pyramid



Campbell’s Hero Journey

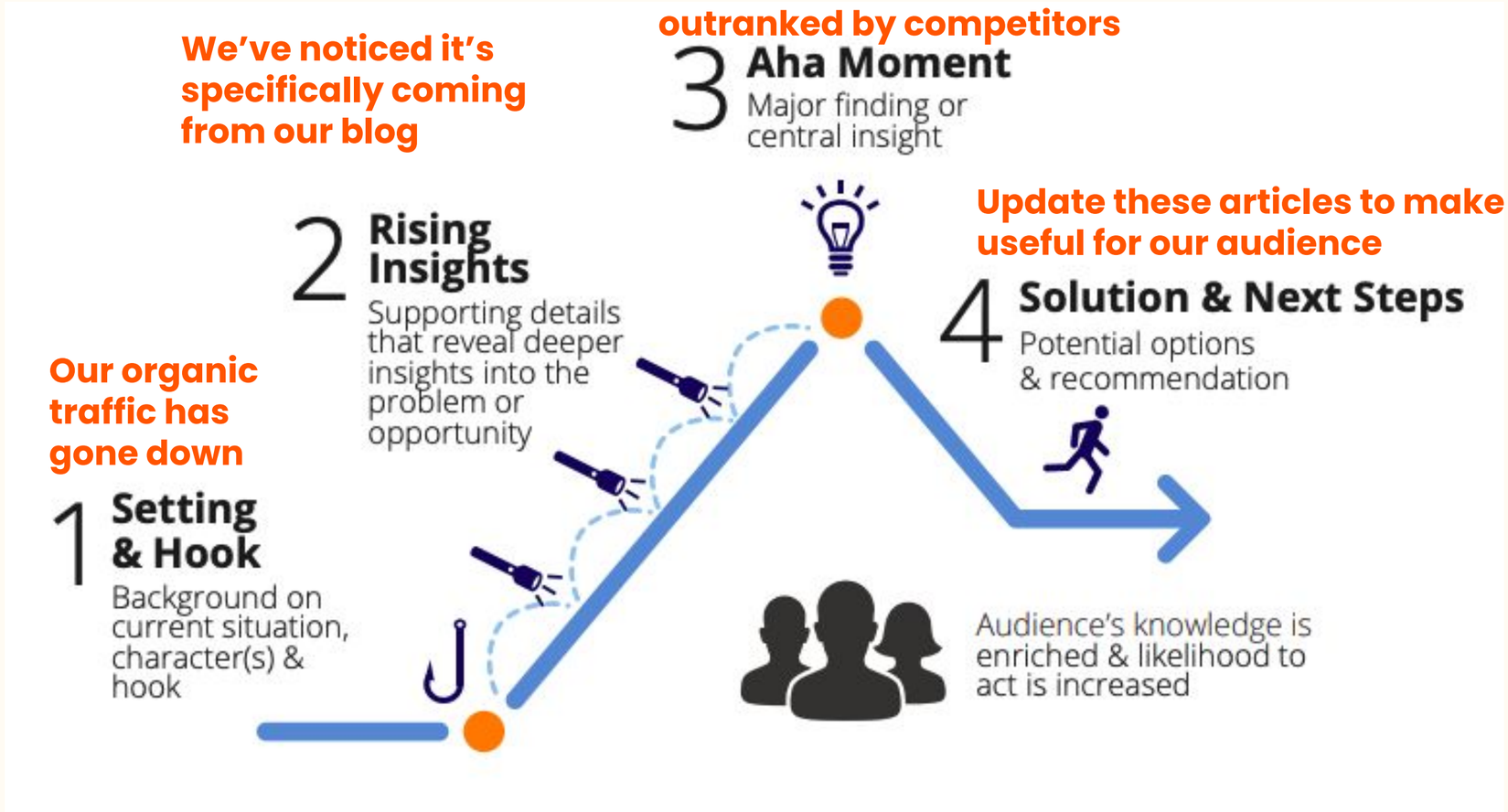
Source: Effective Data Storytelling: How to Drive Change with Data, Narrative, and Visuals

Different narrative structures to choose from



Source: Effective Data Storytelling: How to Drive Change with Data, Narrative, and Visuals

Different narrative structures to choose from



Source: Effective Data Storytelling: How to Drive Change with Data, Narrative, and Visuals

More resources on narrative

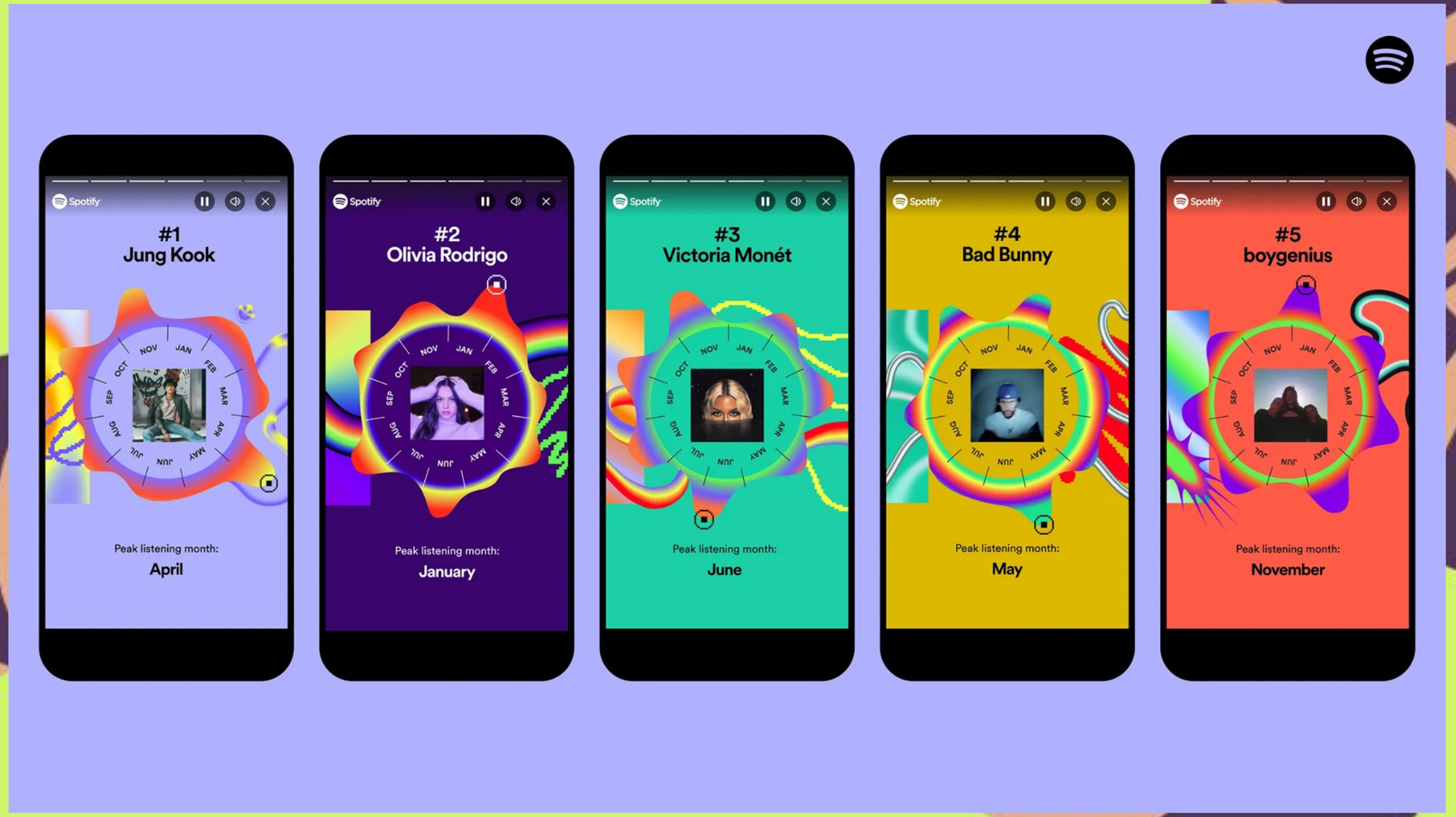
- ✓ Tableau's 5 best practices for telling great stories with data
- ✓ Brent Dyke's Effective Data Storytelling: How to Drive Change with Data, Narrative, and Visuals
- ✓ Storytelling for more impactful data science by Gert de Geyter
- ✓ The data storytelling skills teams need with Andy Cotgreave
- ✓ Cole Nussbaumer Knaflic's Storytelling with Data

3

Everyone will become a data storyteller

Resources for your data visualization skills

Data storytelling is table stakes now



Data storytelling is table stakes now

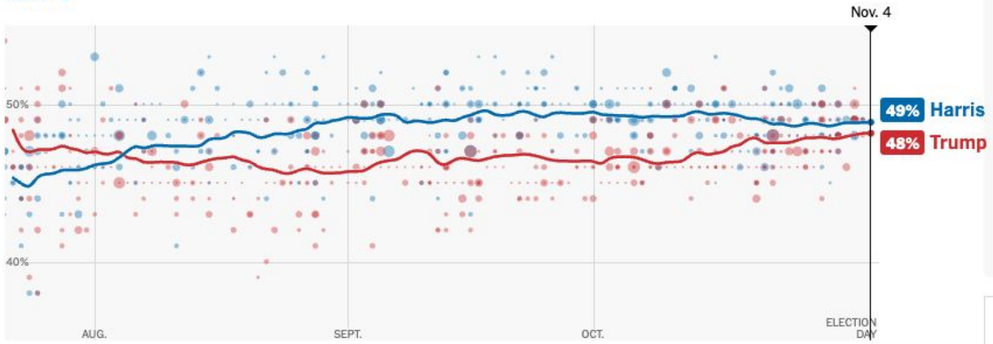
Election 2024 Polls: Harris vs. Trump

Updated Nov. 4, 2024 Leer en español

See the final Times/Siena polls of Arizona, Georgia, Michigan, Nevada, North Carolina, Pennsylvania and Wisconsin ›

Who's leading the polls?

National polling average
Harris <1

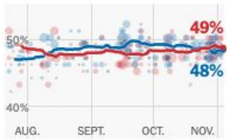


Nate Cohn
Chief political analyst

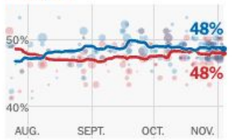
While the overall result of our final Times/Siena polls is largely unchanged since our previous wave of battleground polls, there were some notable shifts. Surprisingly, the longstanding gap between the Rust Belt and Sun Belt battlegrounds narrowed considerably. The overall effect of these swings is somewhat contradictory — Harris's position in the Electoral College isn't necessarily improved.
Updated Nov. 3

Enjoy open access to the election hub in The Times app.
Download The Times app to explore the hub, for a limited time.

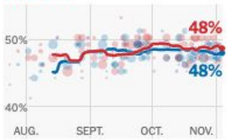
Pennsylvania ›
Even



Michigan ›
Harris <1



North Carolina ›
Trump <1

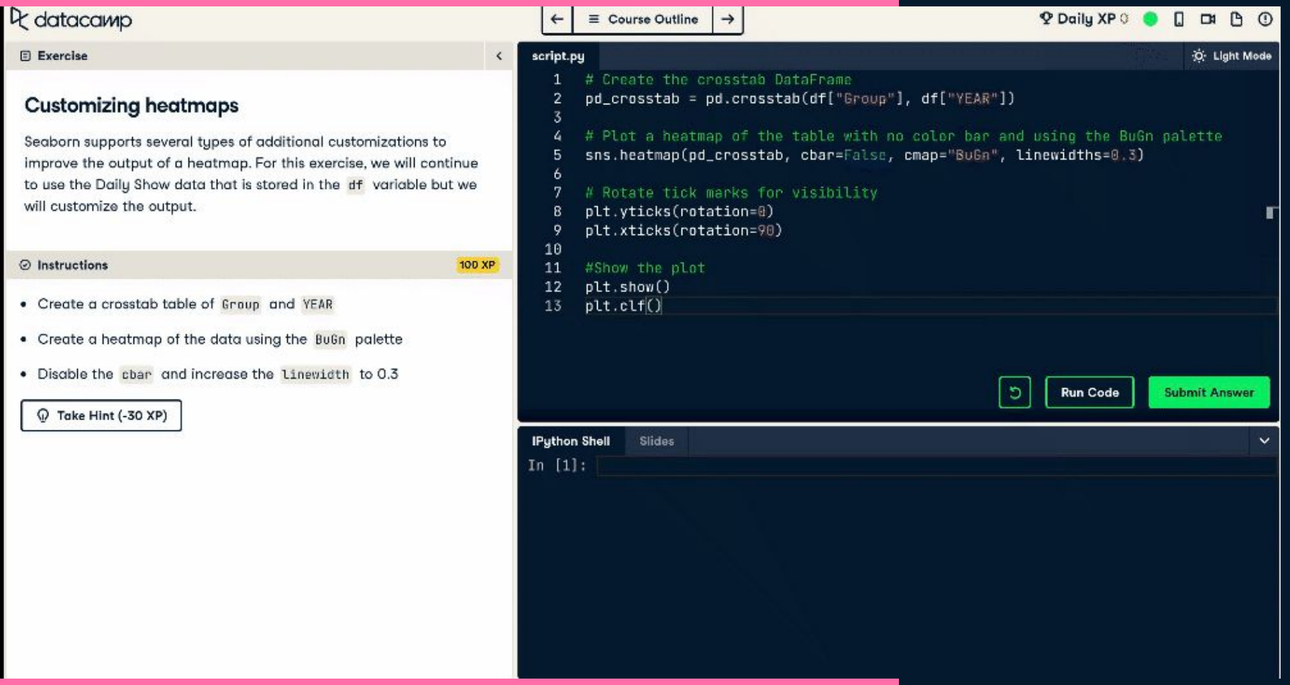


Nevada ›
Trump <1

Wisconsin ›
Harris <1

Georgia ›
Trump +1

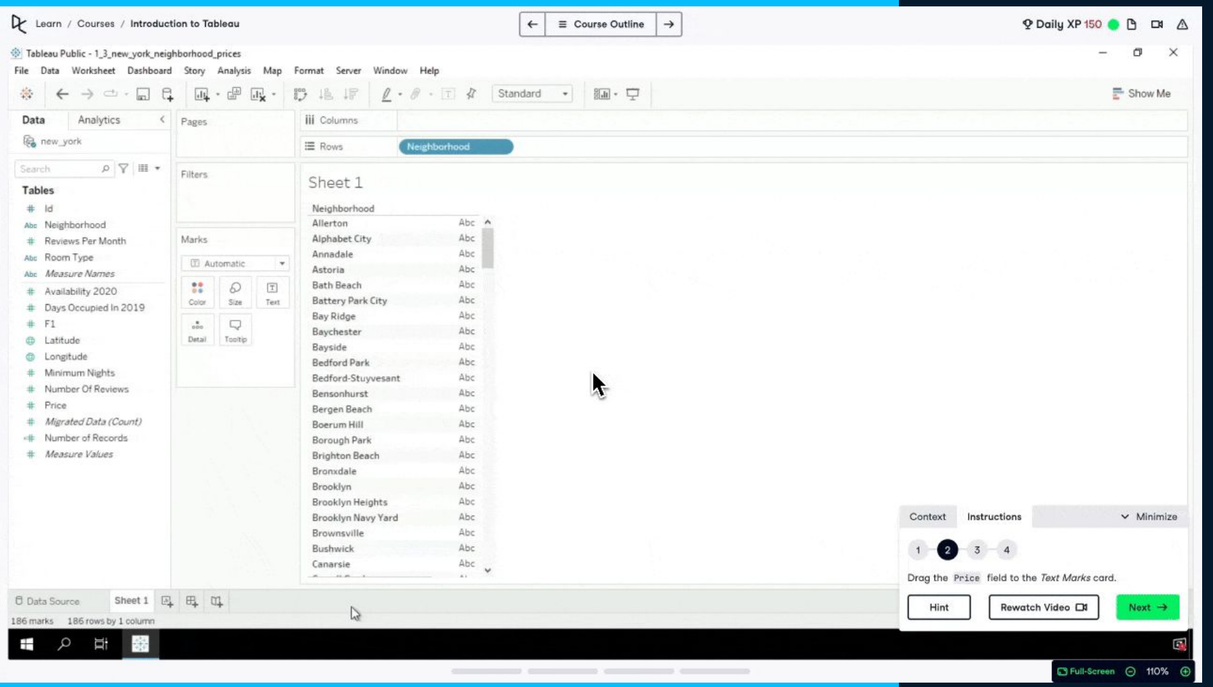
Arizona ›
Trump +3



Level up your data visualization skills with DataCamp

Visualize intensity with Python using heatmaps

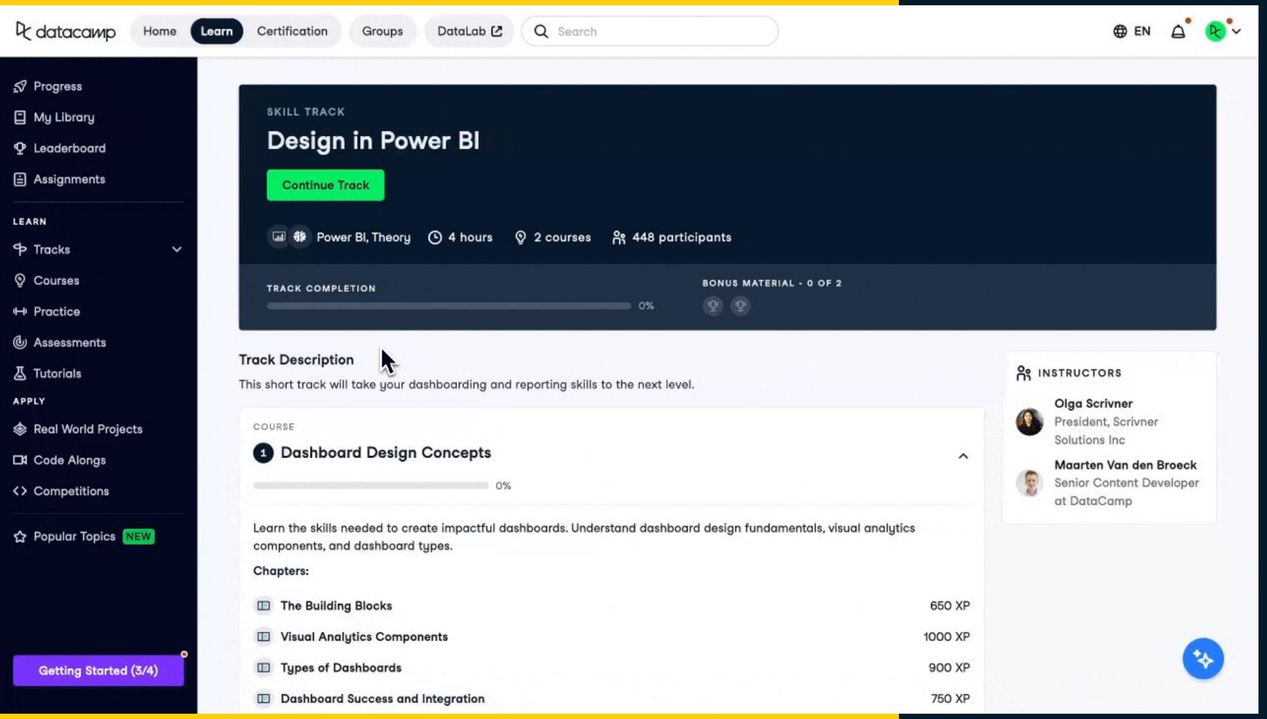
› View Track



Level up your data visualization skills with DataCamp

Introduction to Tableau

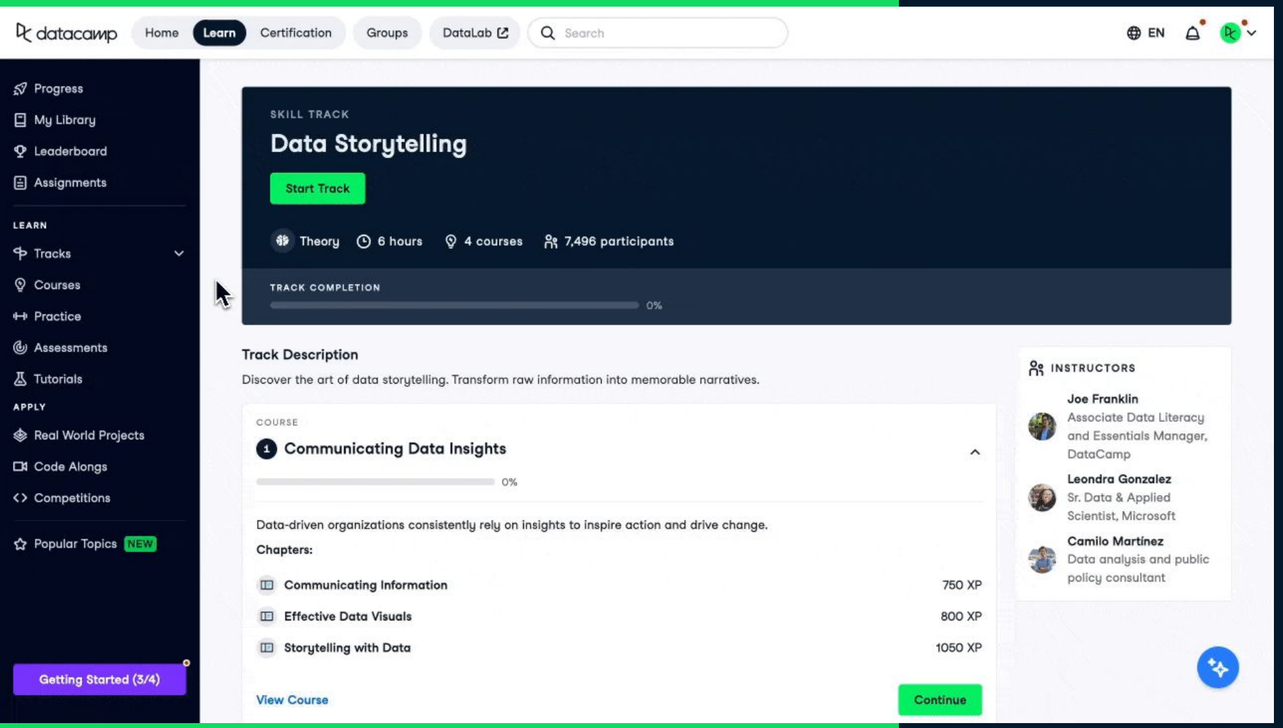
[› View Course]



Level up your data visualization skills with DataCamp

User-Oriented Design in Power BI

› View Course



Level up your data visualization skills with DataCamp

Data Storytelling Concepts

› View Course

What questions can I answer for you?

Additional Resources

- Take DataCamp's Data Maturity Assessment
- WHITE PAPER: Your Organization's Guide to Data Maturity
- Register for one of our upcoming webinars
- Learn more about DataCamp for Business
- ON-DEMAND: Storytelling for more impactful data science
- ON-DEMAND: How Data Governance Enables Scalable Data Science

Thank you

Adel Nehme
VP of Media
adel@datacamp.com