

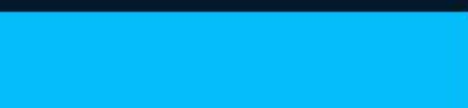
# 8 rules for better data storytelling



# 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 & CentraleSupélec
- 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**

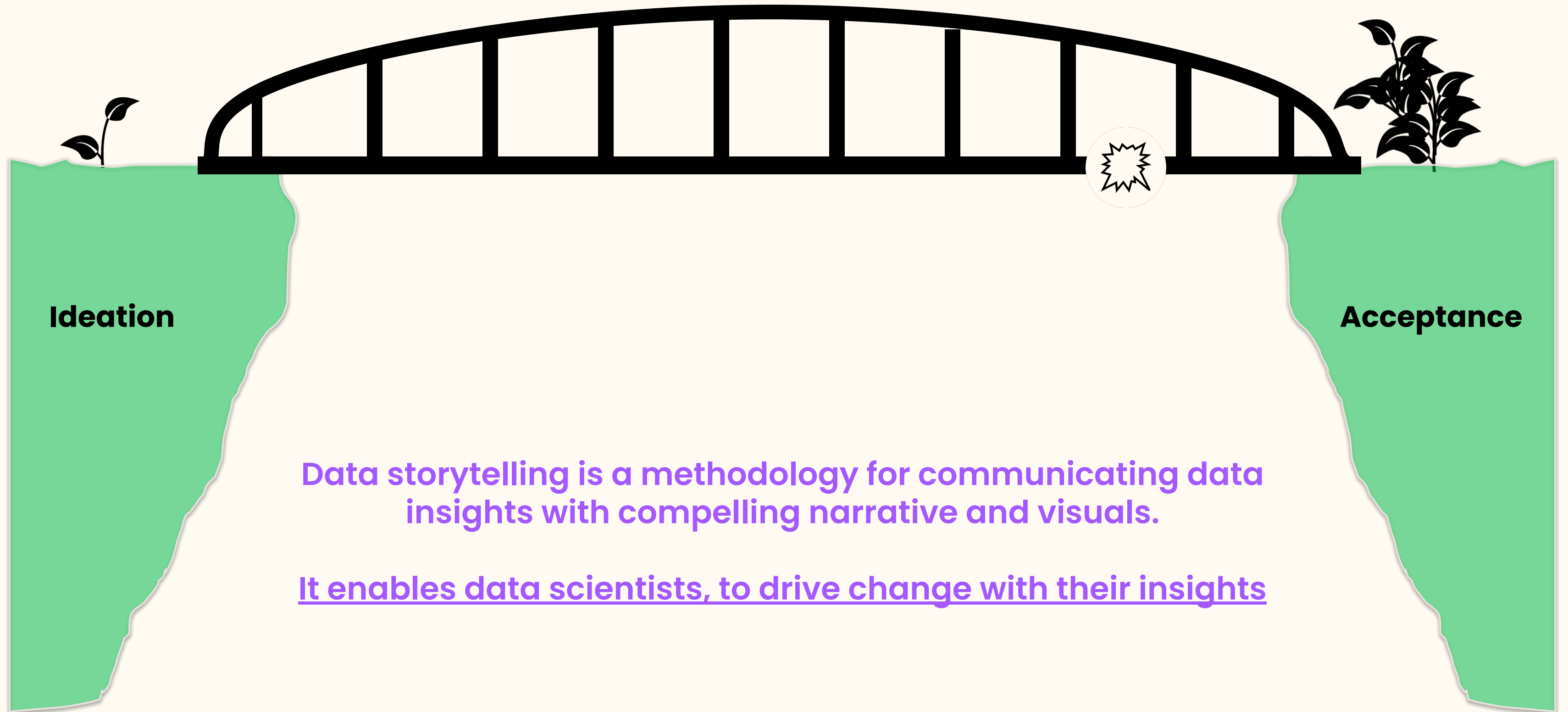


*"To be impactful with data science, organizations need data scientists with stories "* — **Gert de Geyter, Machine Learning Lead at Deloitte**

*"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?



**Ideation**

**Acceptance**

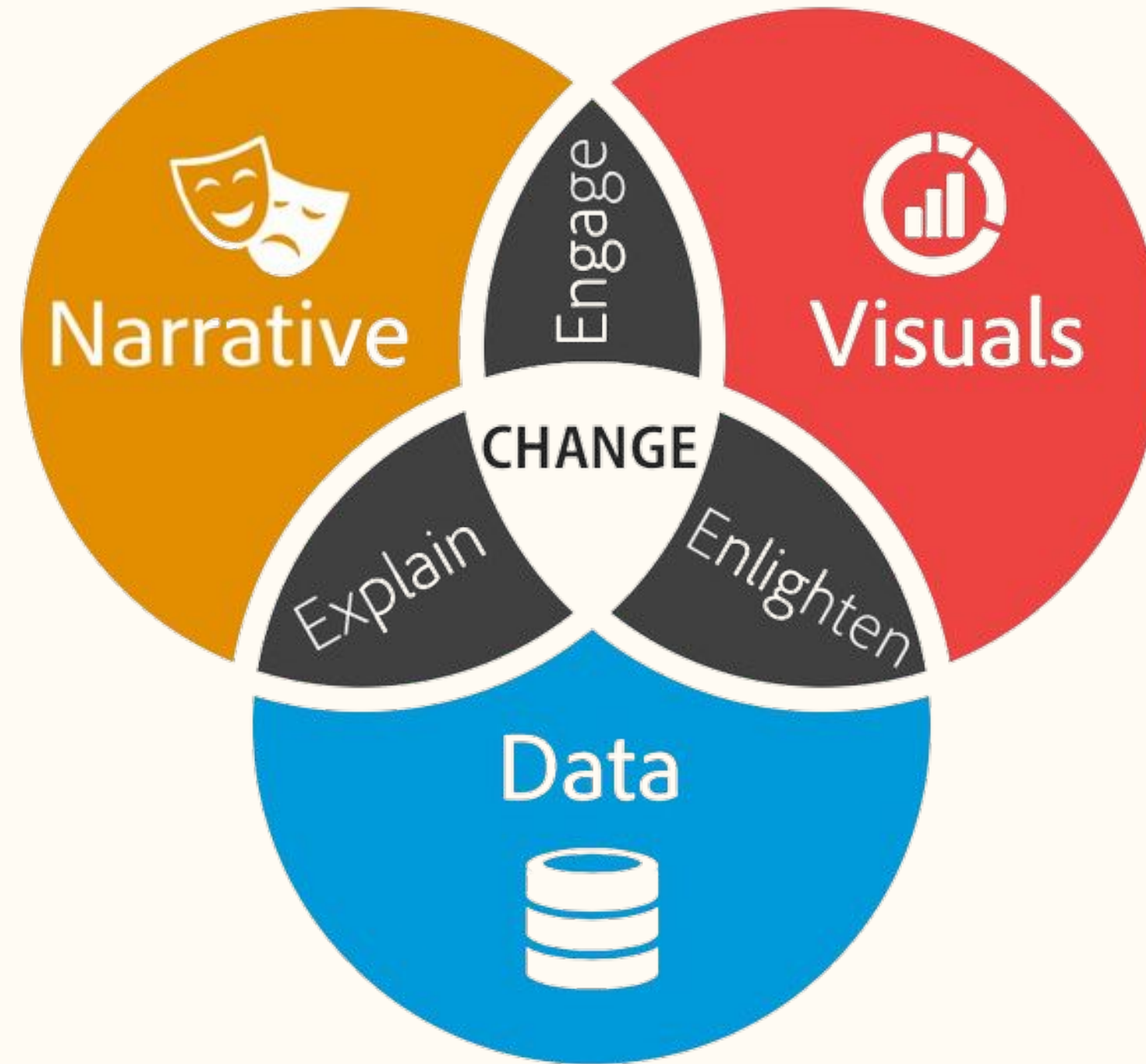
Data storytelling is a methodology for communicating data insights with compelling narrative and visuals.

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

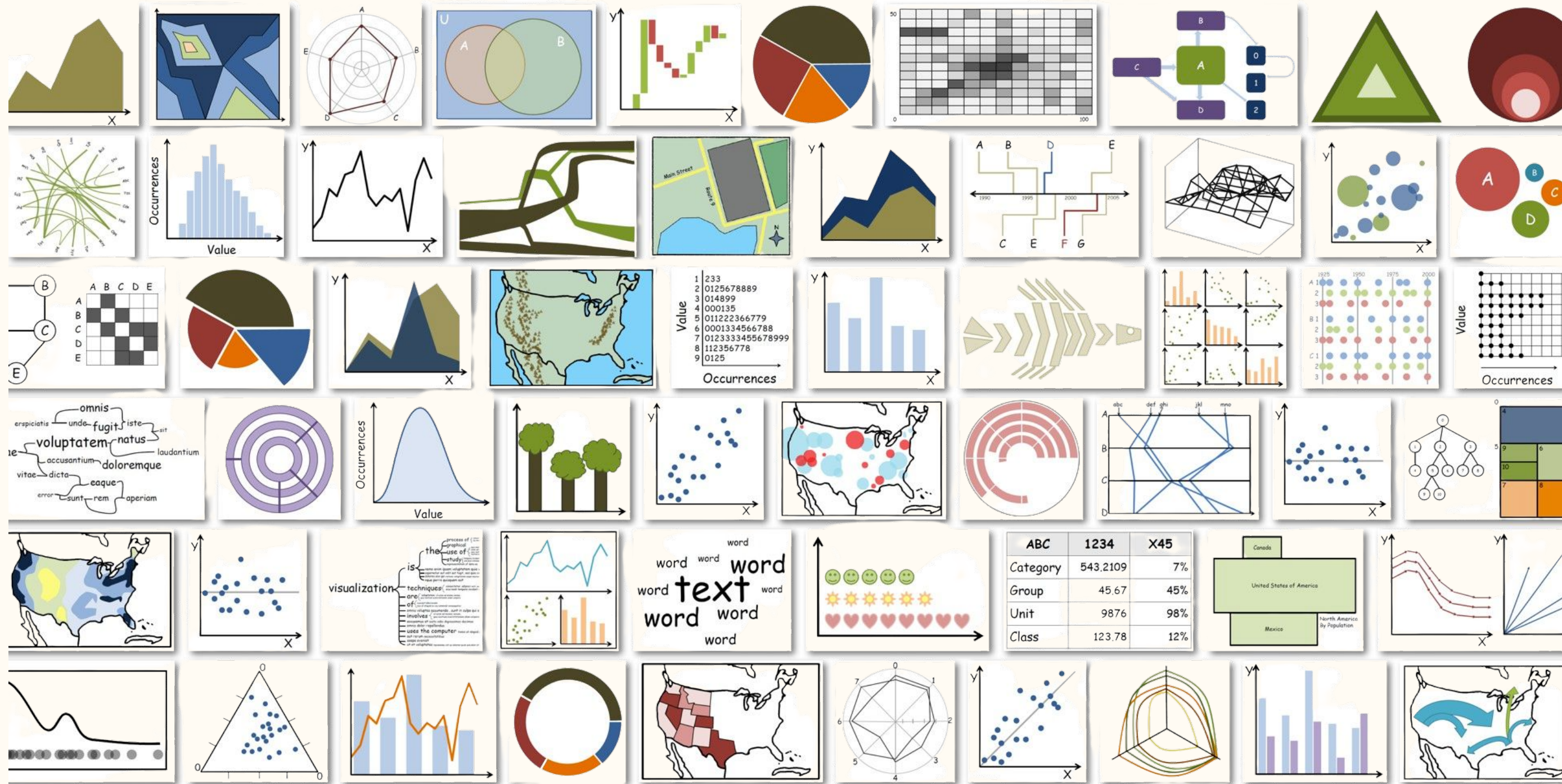
## *Rules for better data visualizations*



# 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

---

- Banking products and associated customers
  - Branch cost data broken down into different verticals
  - Visualizing different customer segments
- 
- 
- 
- 
- 

## PROBLEM DESCRIPTION

---

Showing comparisons of different categories

---

---

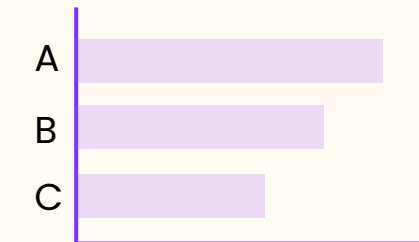
---

---

---

## MOST USEFUL VISUALIZATION

---



**BAR CHARTS**

---

---

---

---

---



# Common uses of bar charts

Number of customers

Visualizing number of customers by loan product



Business Loans

Loan Against Property

Commercial Vehicle Financing

Construction Equipment Loan

Farm Equipment Loan

Loan Products





# Always work back from your data

## COMMONLY USED DATA

---

- Banking products and associated customers
  - Branch cost data broken down into different verticals
  - Visualizing different customer segments
- 

- Stock price change over time
  - Number of app users over time
  - Number of customer support tickets over time
- 

## PROBLEM DESCRIPTION

---

Showing comparisons of different categories

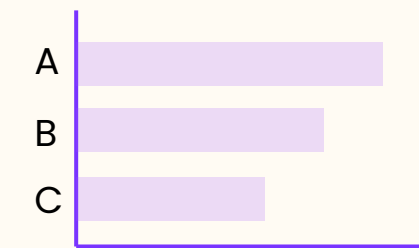
---

Showing changes of a variable over time

---

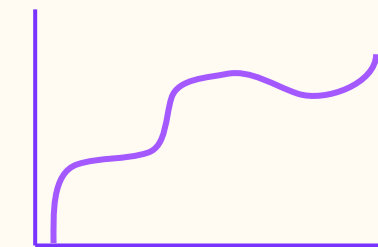
## MOST USEFUL VISUALIZATION

---



**BAR CHARTS**

---



**LINE CHARTS**

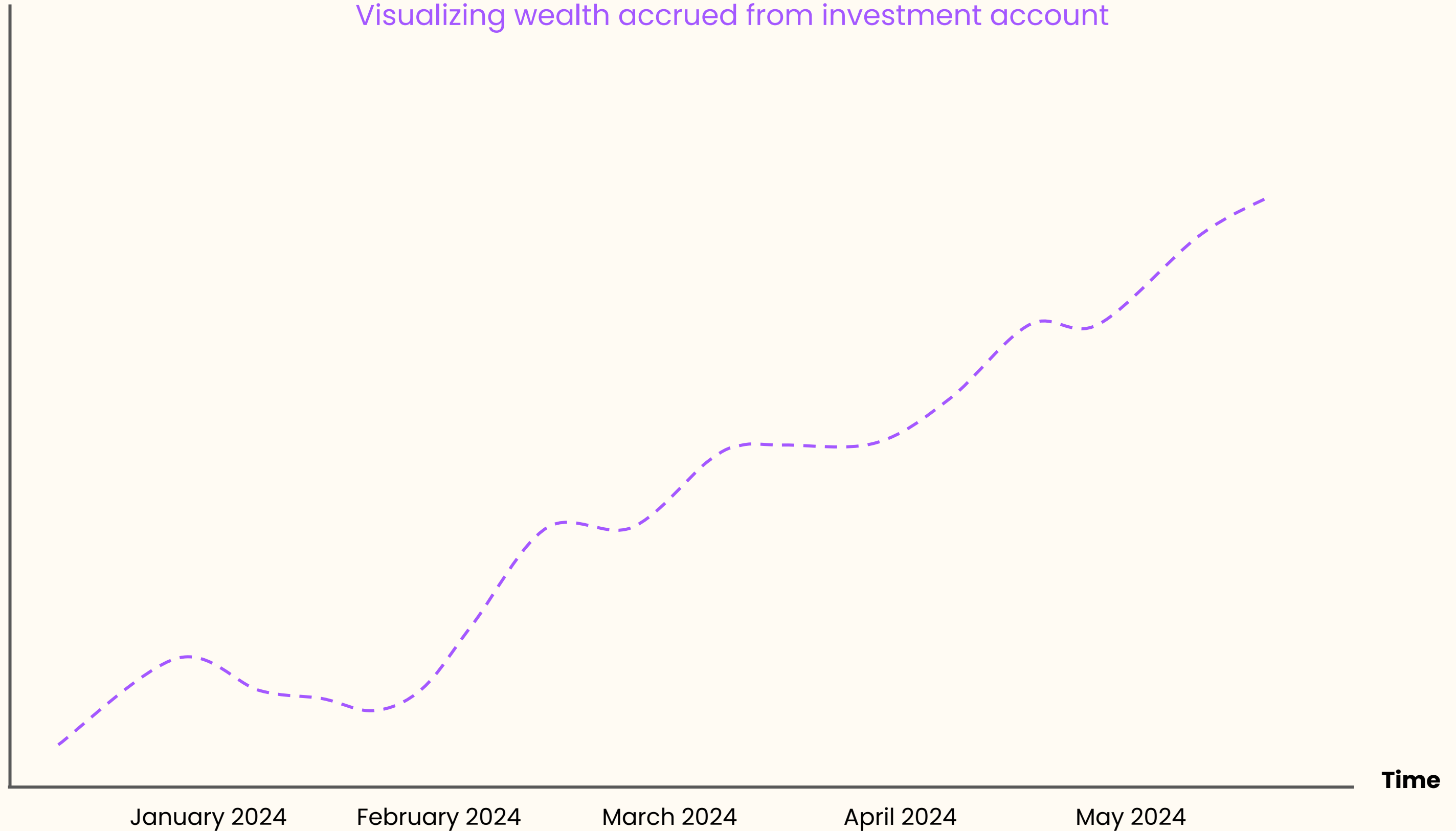
---



# Common uses of line charts

Visualizing wealth accrued from investment account

**Investment  
Account Gains  
over time**



# Always work back from your data

## COMMONLY USED DATA

---

- Banking products and associated customers
  - Branch cost data broken down into different verticals
  - Visualizing different customer segments
- 

- Stock price change over time
  - Number of banking app users over time
  - Number of customer support tickets over time
- 

- The distribution of savings for customers in savings accounts
  - App opens for all customers in a given time period
- 

## PROBLEM DESCRIPTION

---

Showing comparisons of different categories

---

Showing changes of a variable over time

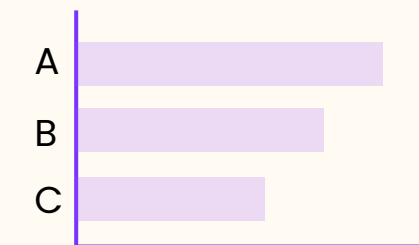
---

Show the distribution of a variable over time

---

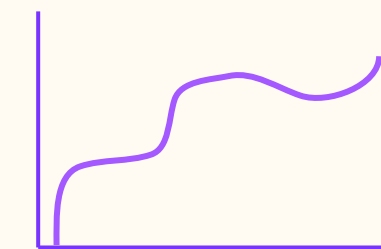
## MOST USEFUL VISUALIZATION

---



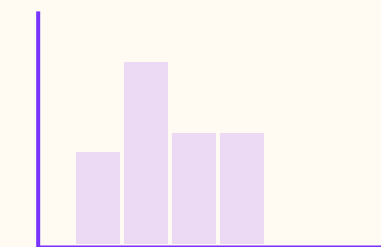
**BAR CHARTS**

---



**LINE CHARTS**

---



**HISTOGRAMS**

---

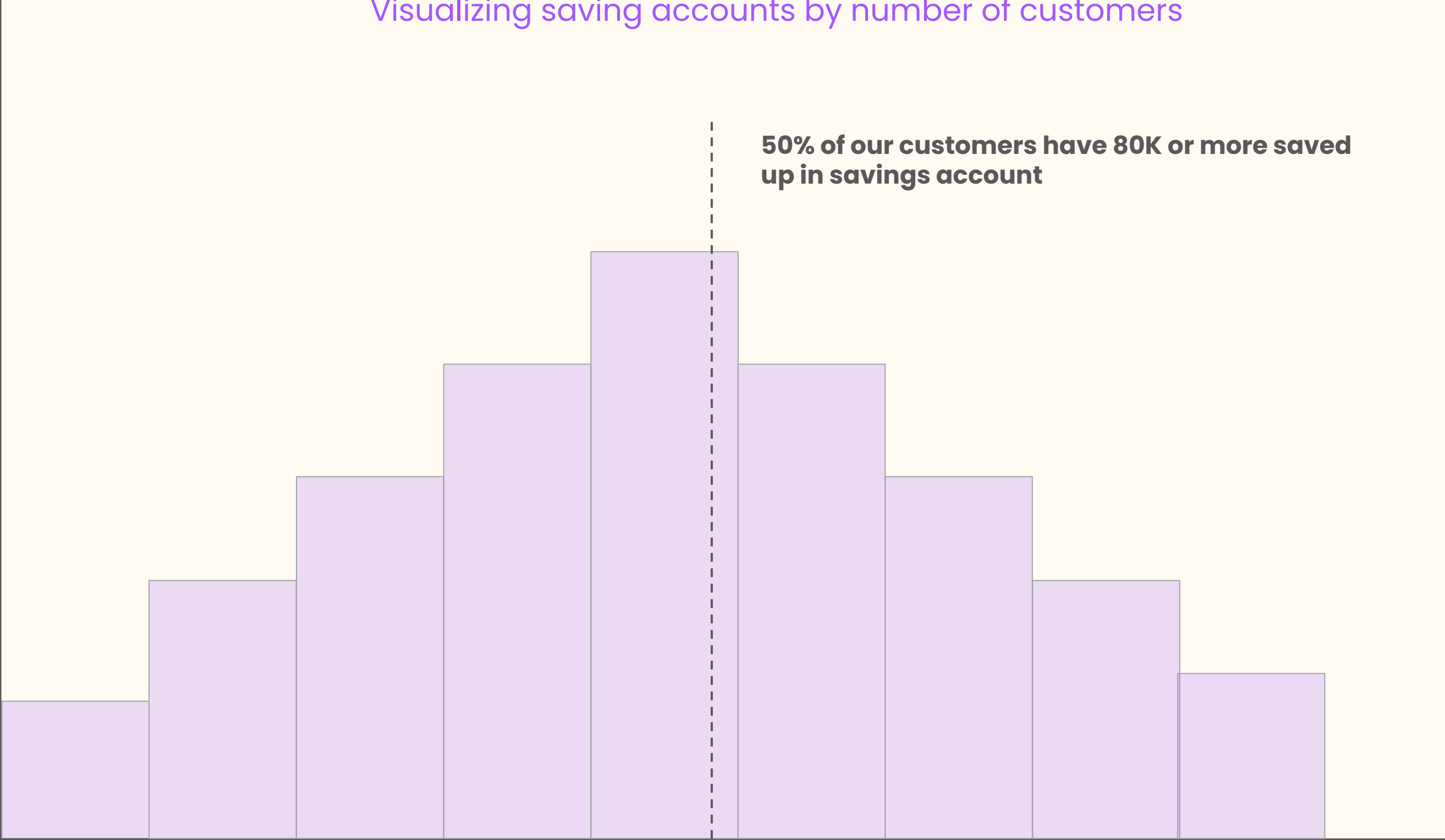


# Common uses of histograms

Number of customers

Visualizing saving accounts by number of customers

50% of our customers have 80K or more saved up in savings account



Savings account \$



# Always work back from your data

## COMMONLY USED DATA

---

- Banking products and associated customers
  - Branch cost data broken down into different verticals
  - Visualizing different customer segments
- 

- Stock price change over time
  - Number of banking app users over time
  - Number of customer support tickets over time
- 

- The distribution of savings for customers in savings accounts
  - Banking app opens for all customers in a given time period
- 

- The relationship between historical credit scores and number of loans taken
  - The relationship between customer lifetime value and number of products purchased
- 

## PROBLEM DESCRIPTION

---

Showing comparisons of different categories

---

Showing changes of a variable over time

---

Show the distribution of a variable over time

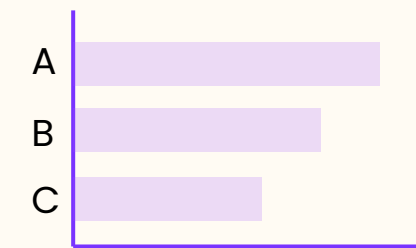
---

Show the relationship between two variables

---

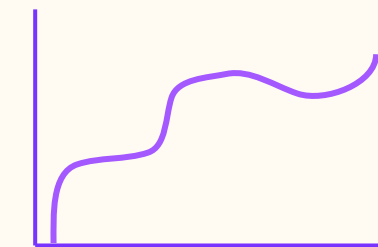
## MOST USEFUL VISUALIZATION

---



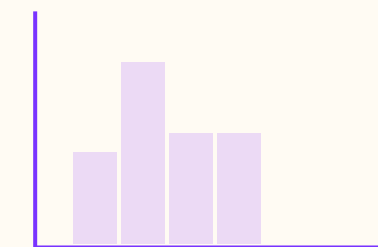
**BAR CHARTS**

---



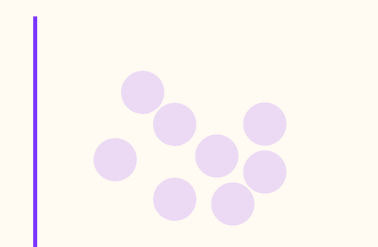
**LINE CHARTS**

---



**HISTOGRAMS**

---



**SCATTER PLOTS**

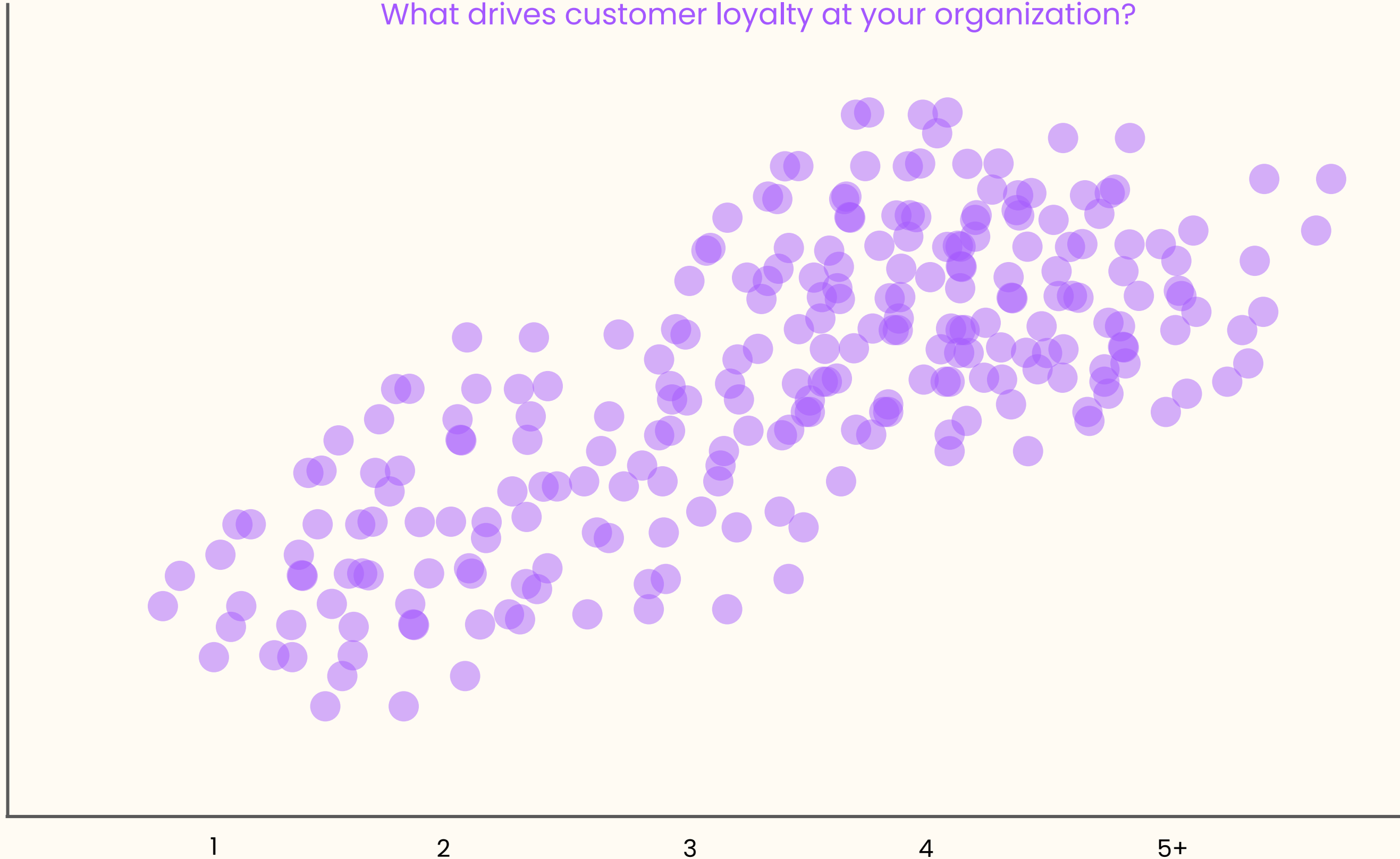
---



# Common uses of scatter plots

**Customer  
lifetime value**

What drives customer loyalty at your organization?



**Products  
purchased**



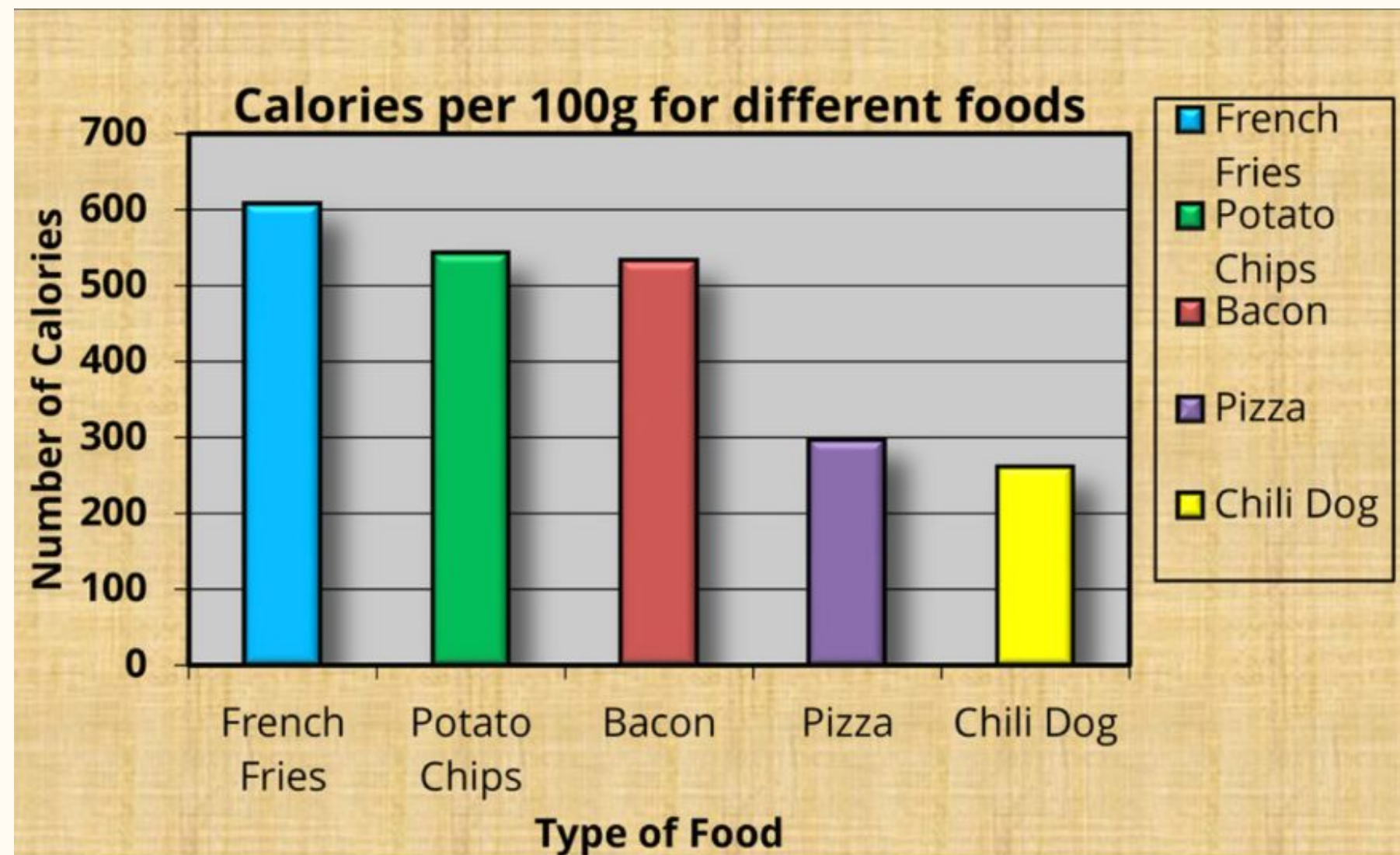


## Rule #2

*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](#)

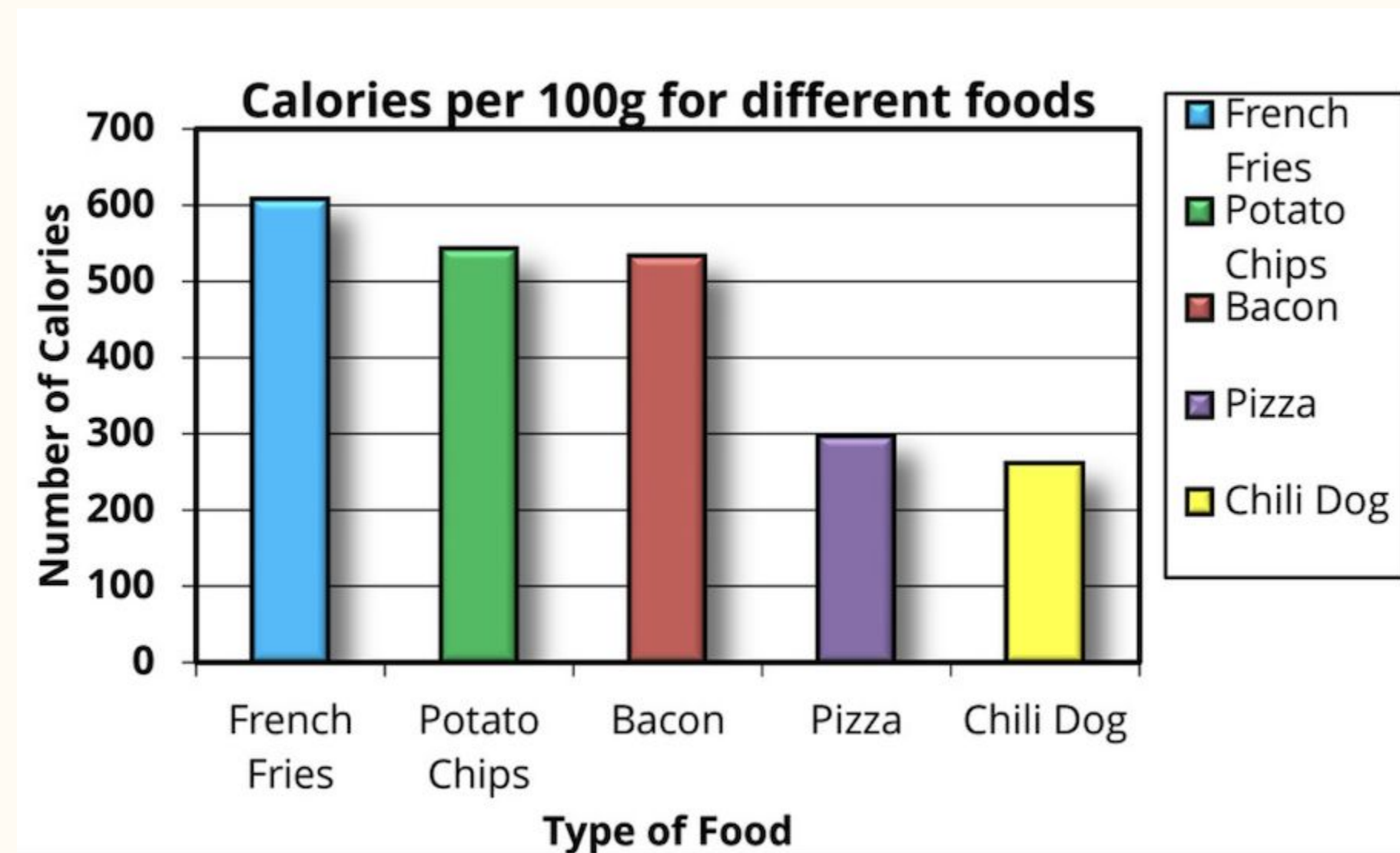




# The cognitive load and effectiveness tradeoff

*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



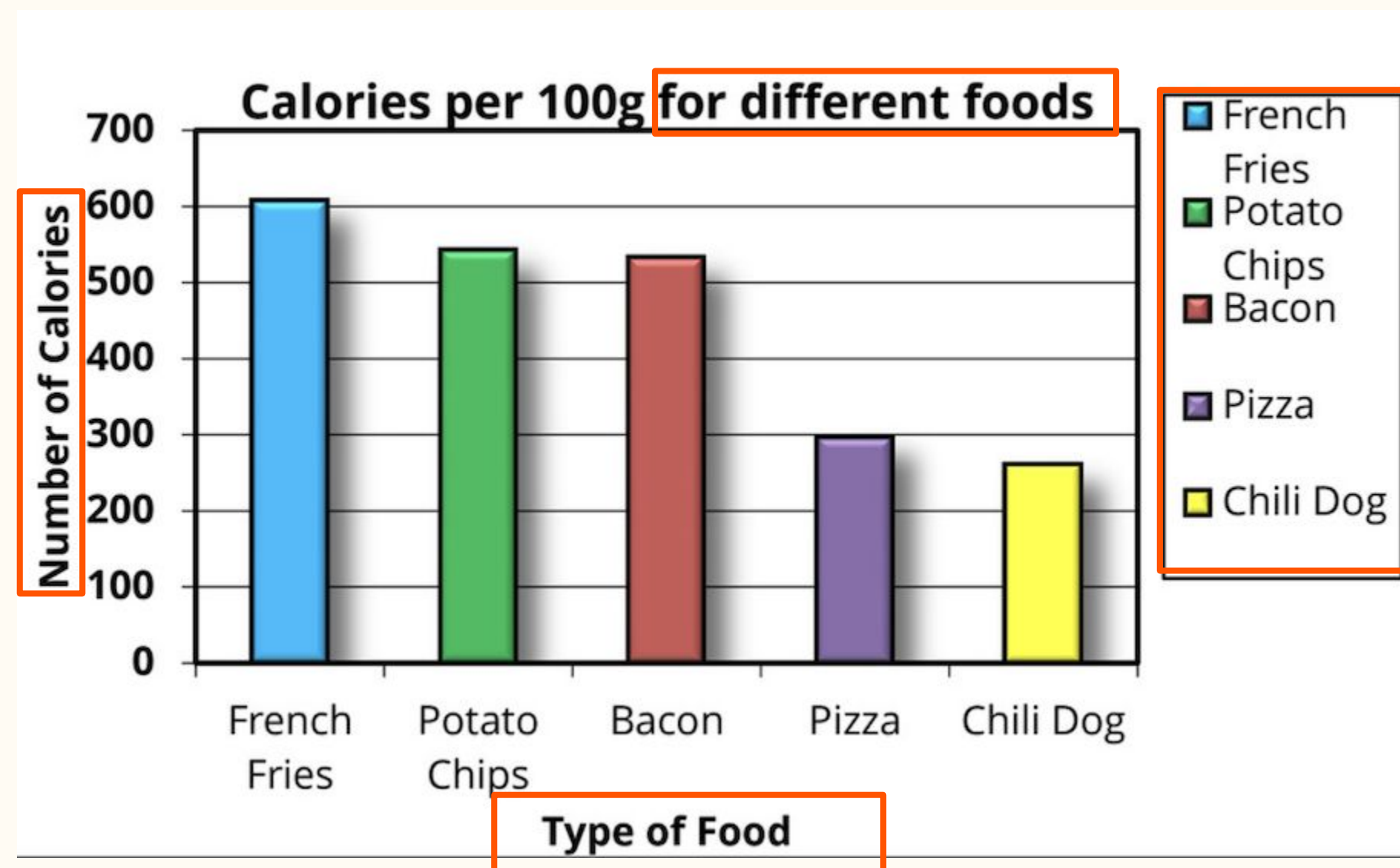
[Source: Darkhorse Analytics](#)



# 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



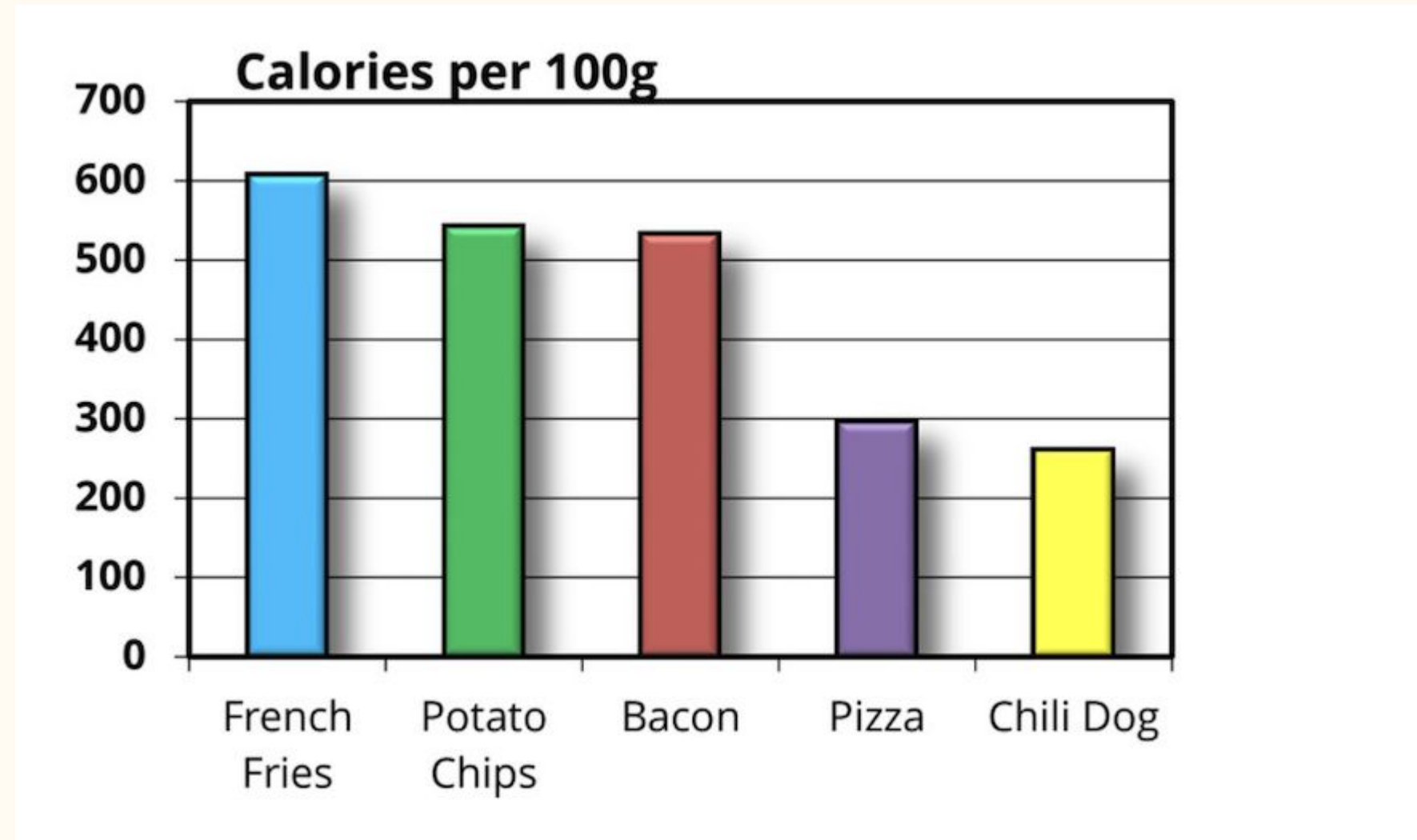
[Source: Darkhorse Analytics](#)



# 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



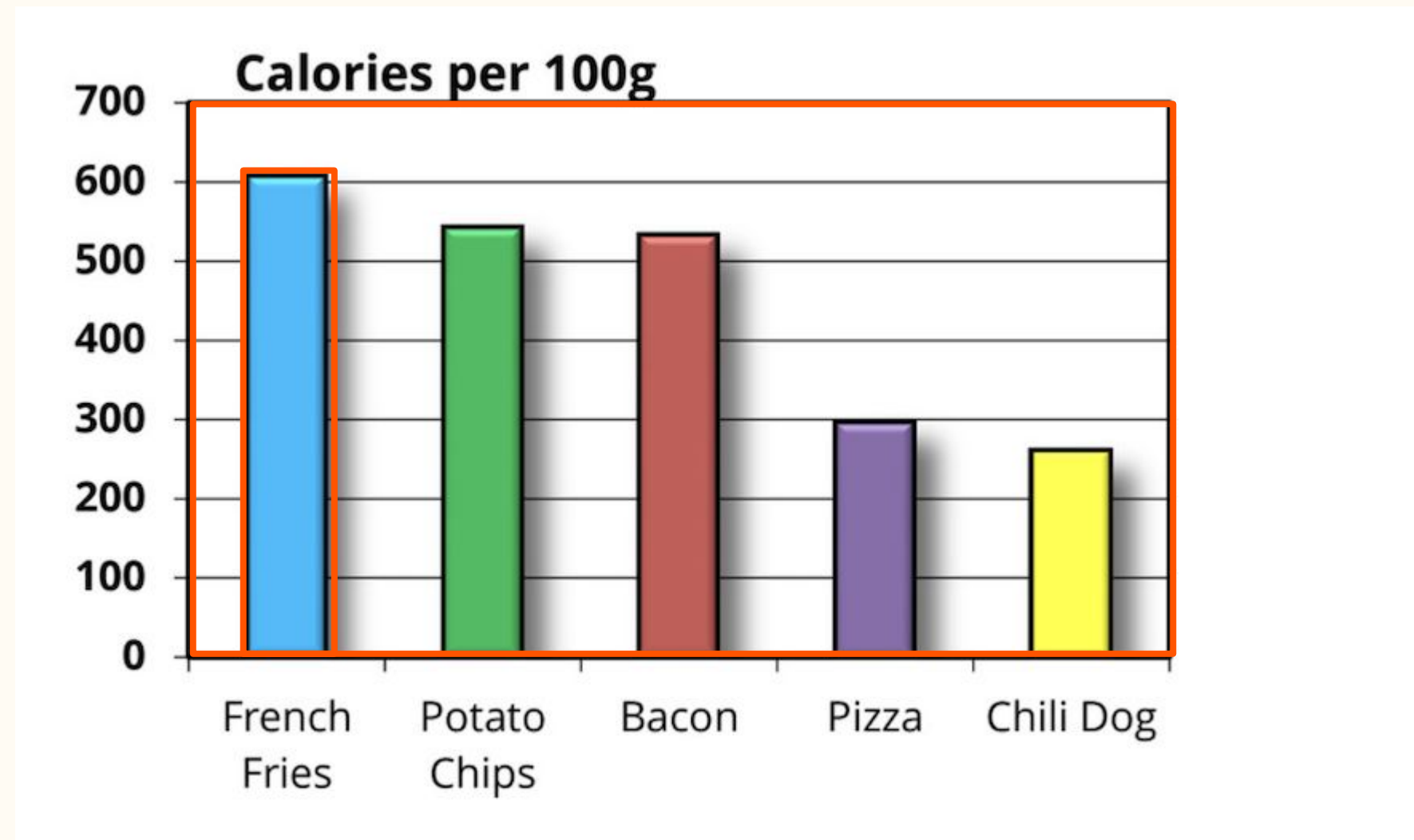
[Source: Darkhorse Analytics](#)



# 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



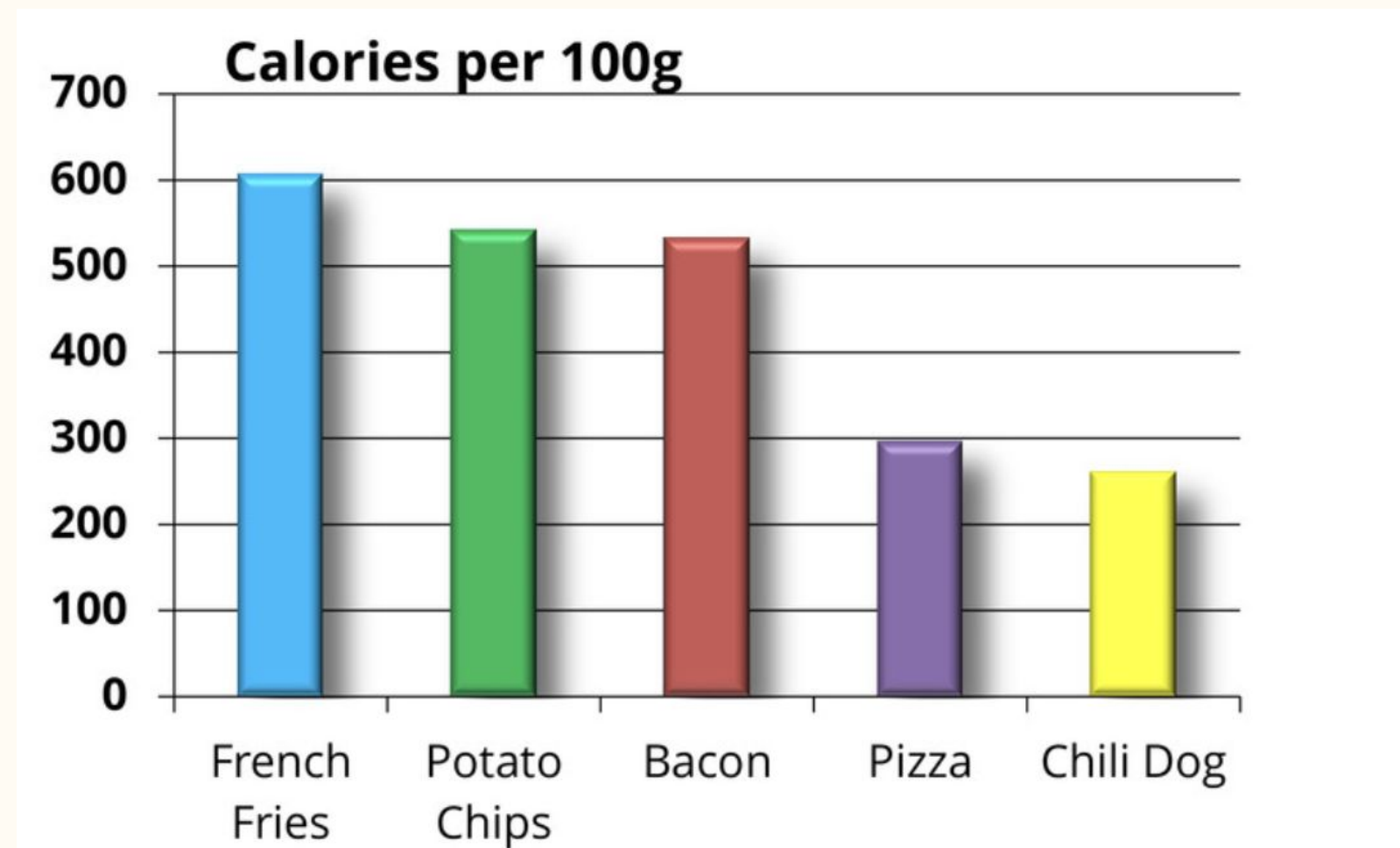
[Source: Darkhorse Analytics](#)



# 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



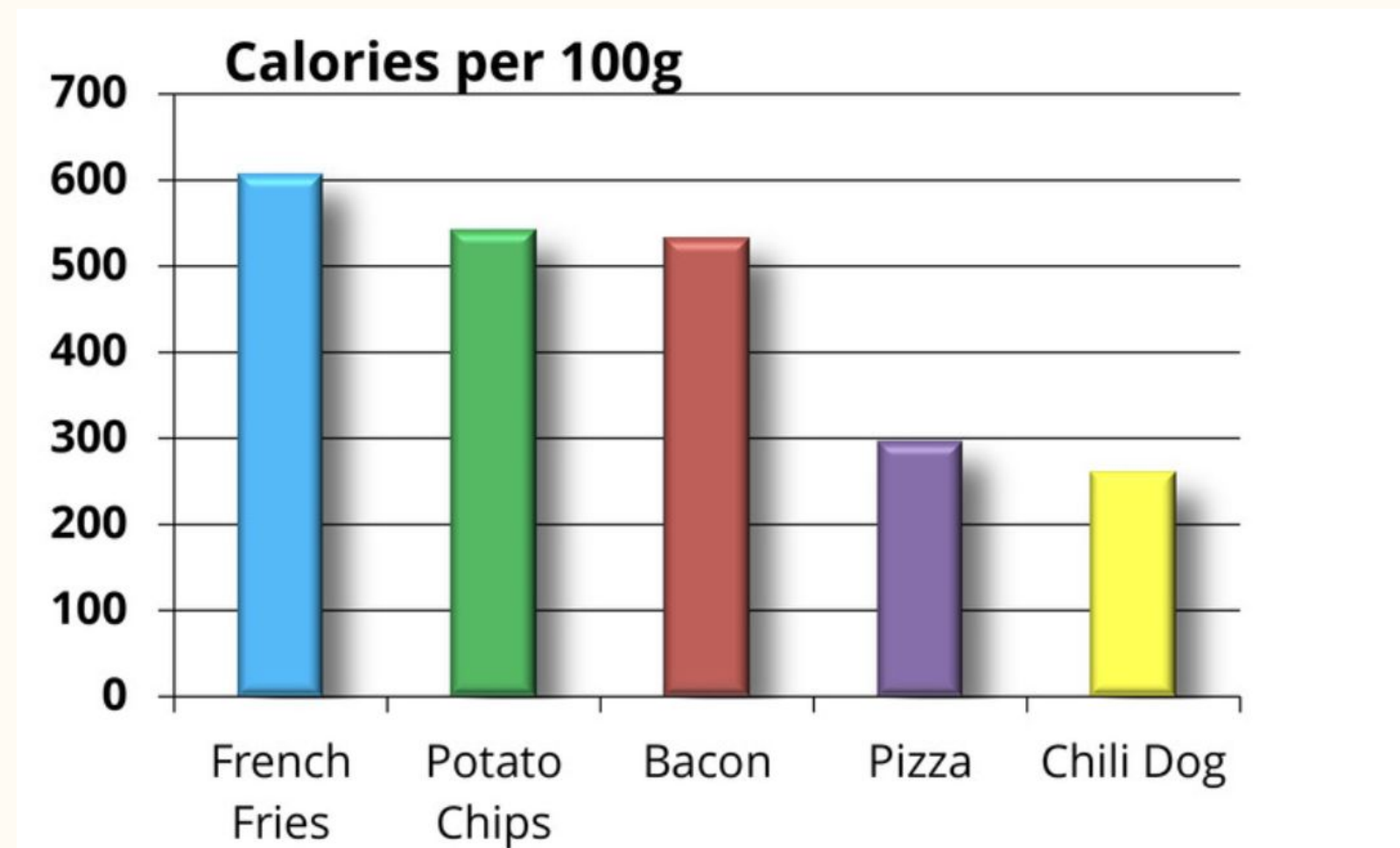
[Source: Darkhorse Analytics](#)



# 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



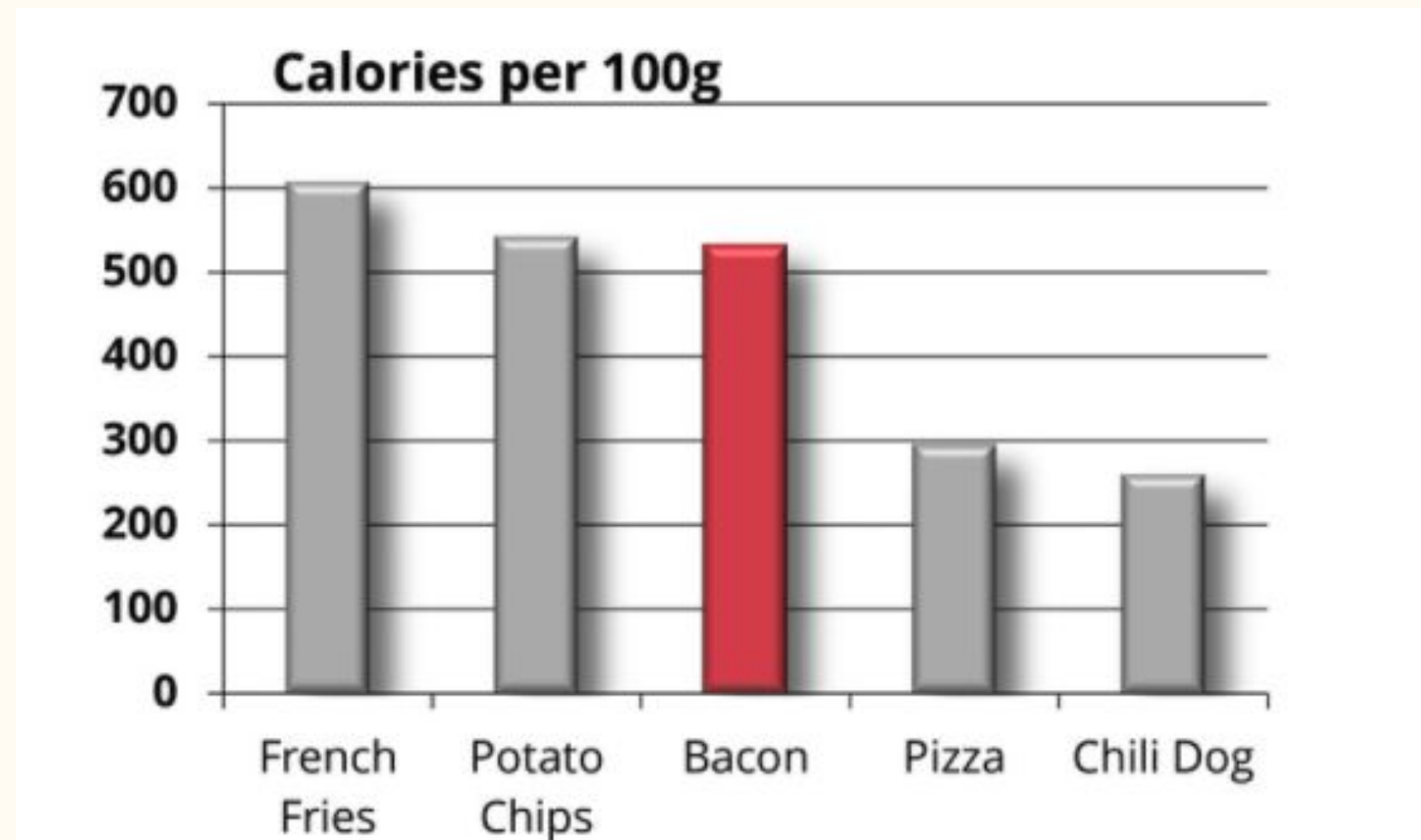
[Source: Darkhorse Analytics](#)



# 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



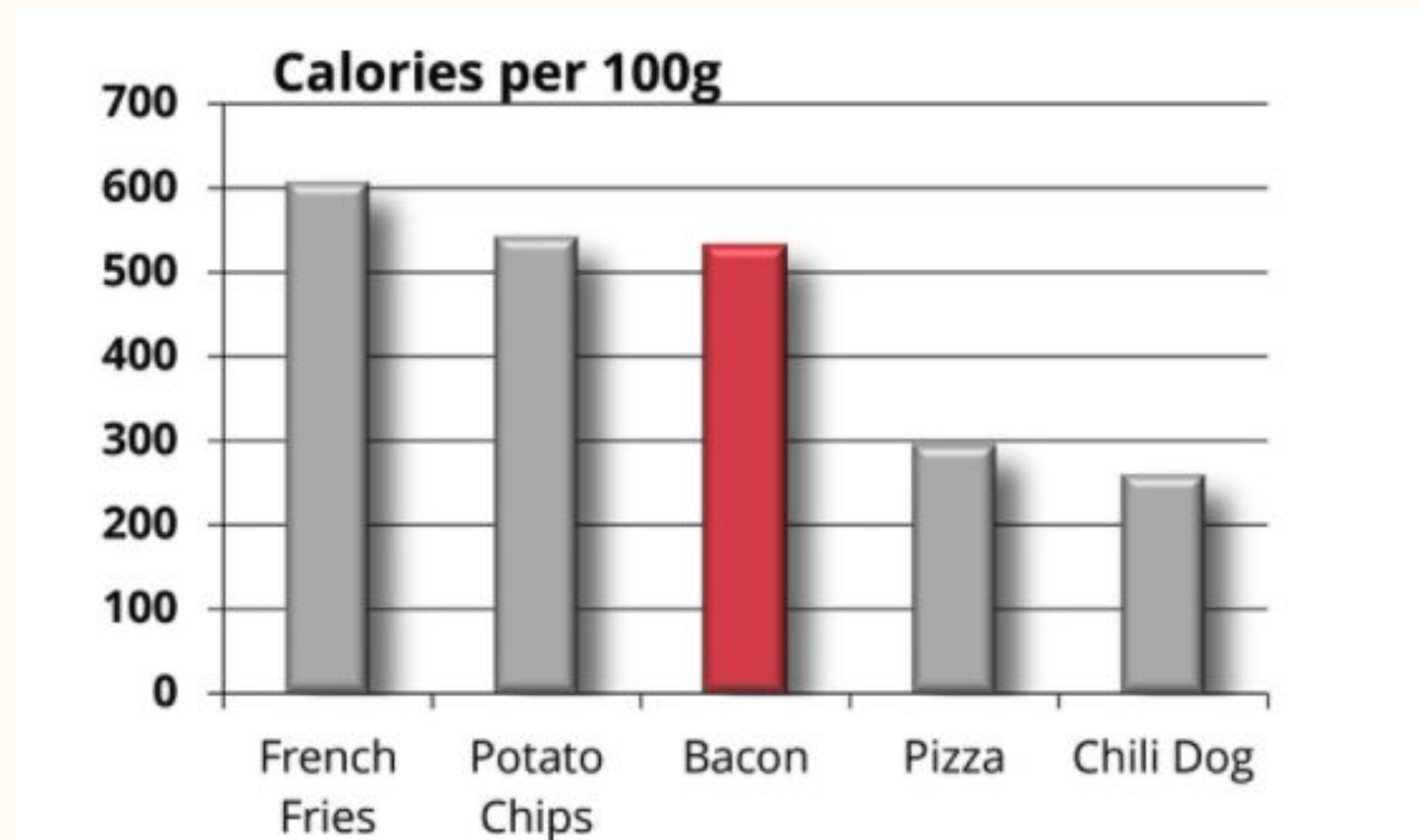
[Source: Darkhorse Analytics](#)



# 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](#)

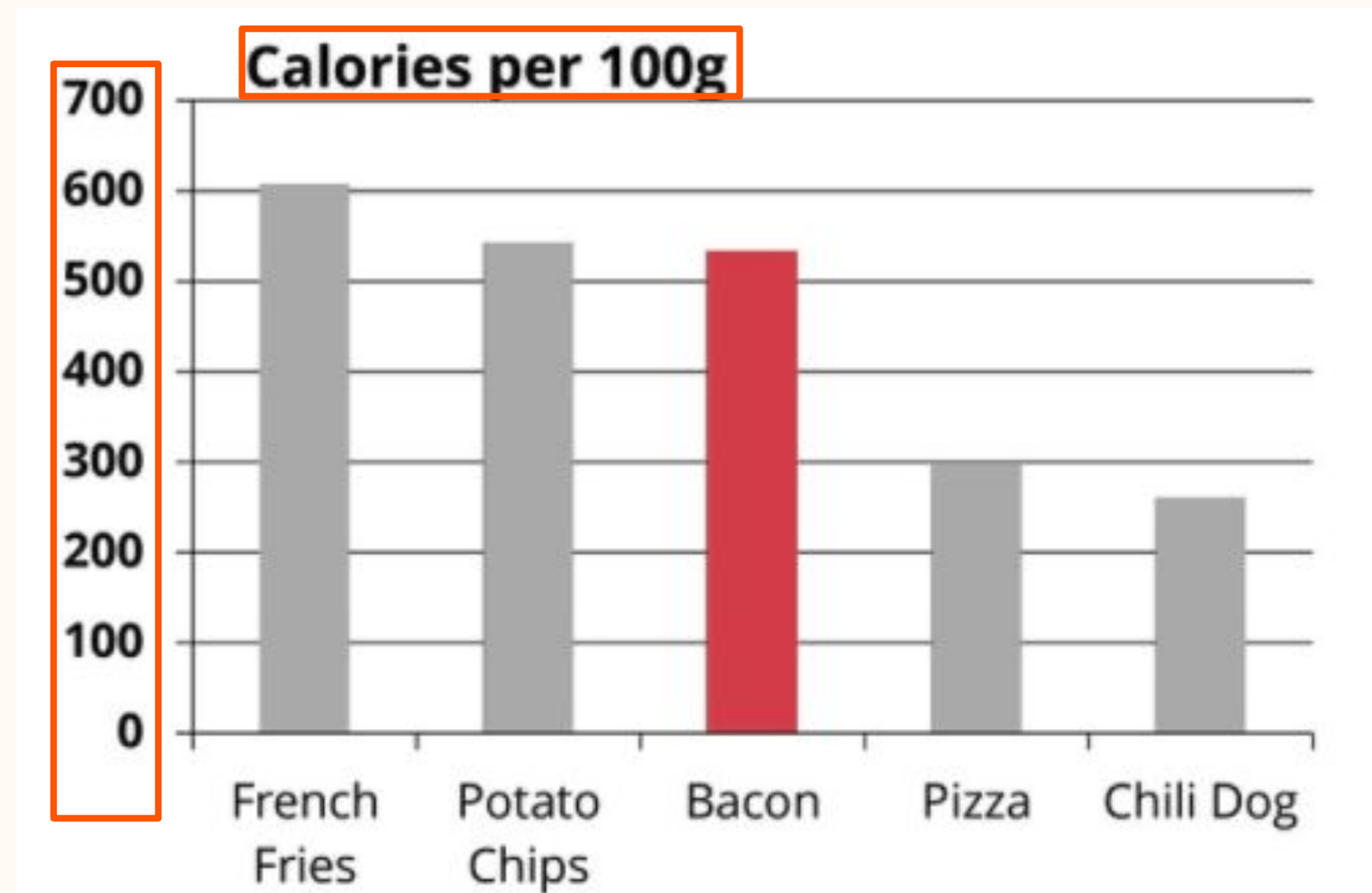




# The cognitive load and effectiveness tradeoff

*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



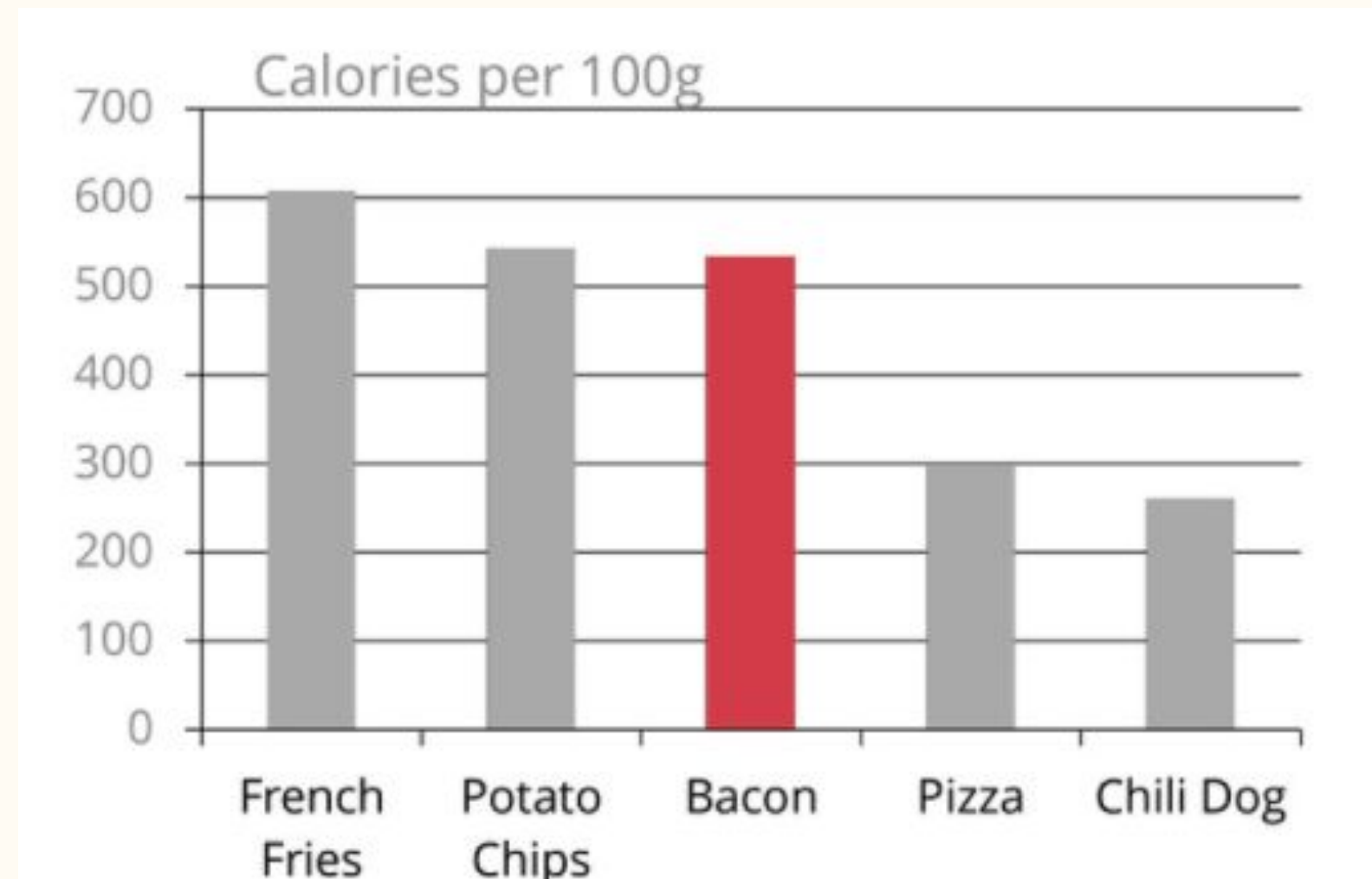
[Source: Darkhorse Analytics](#)



# 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



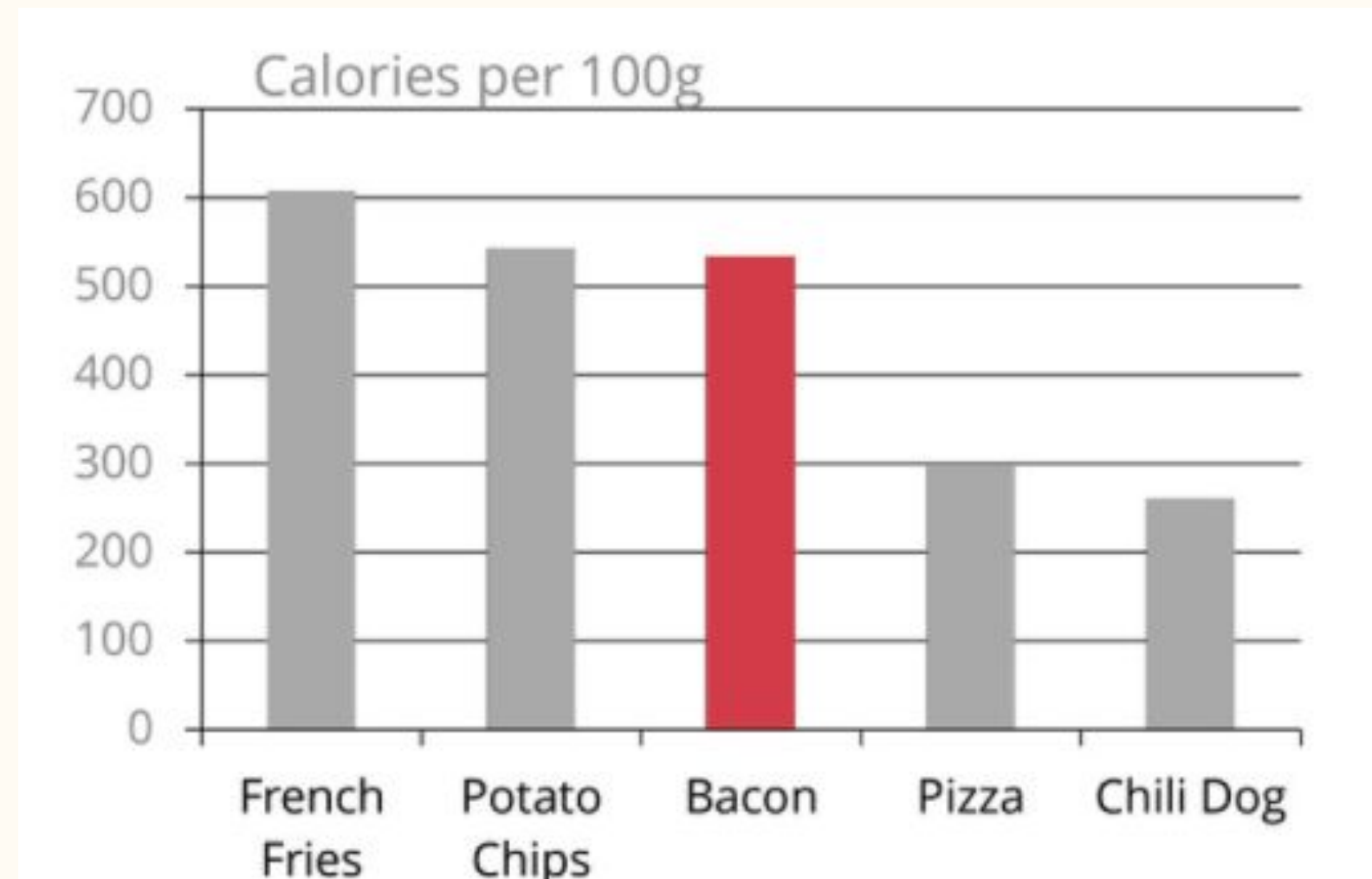
[Source: Darkhorse Analytics](#)



# 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



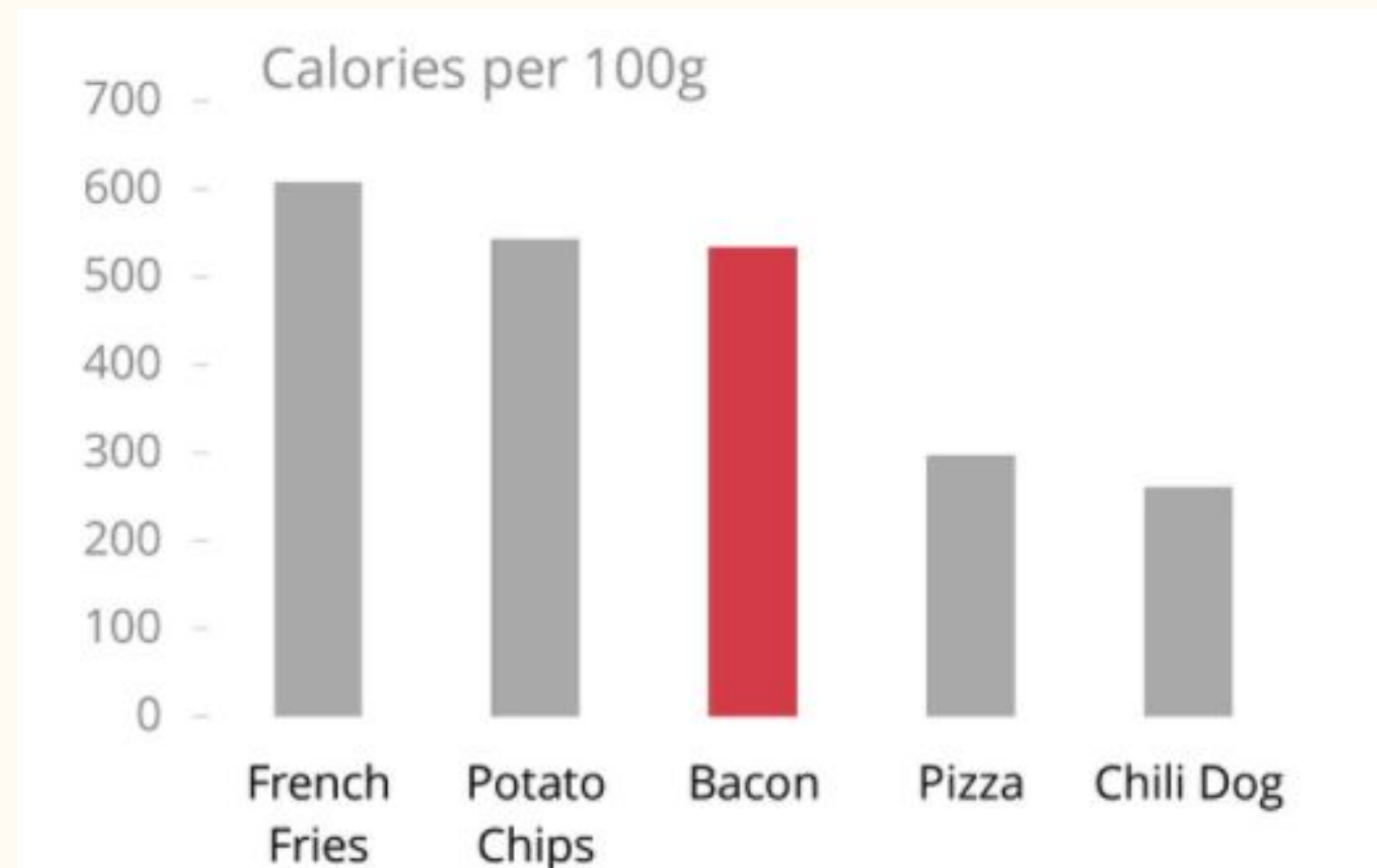
[Source: Darkhorse Analytics](#)



# 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



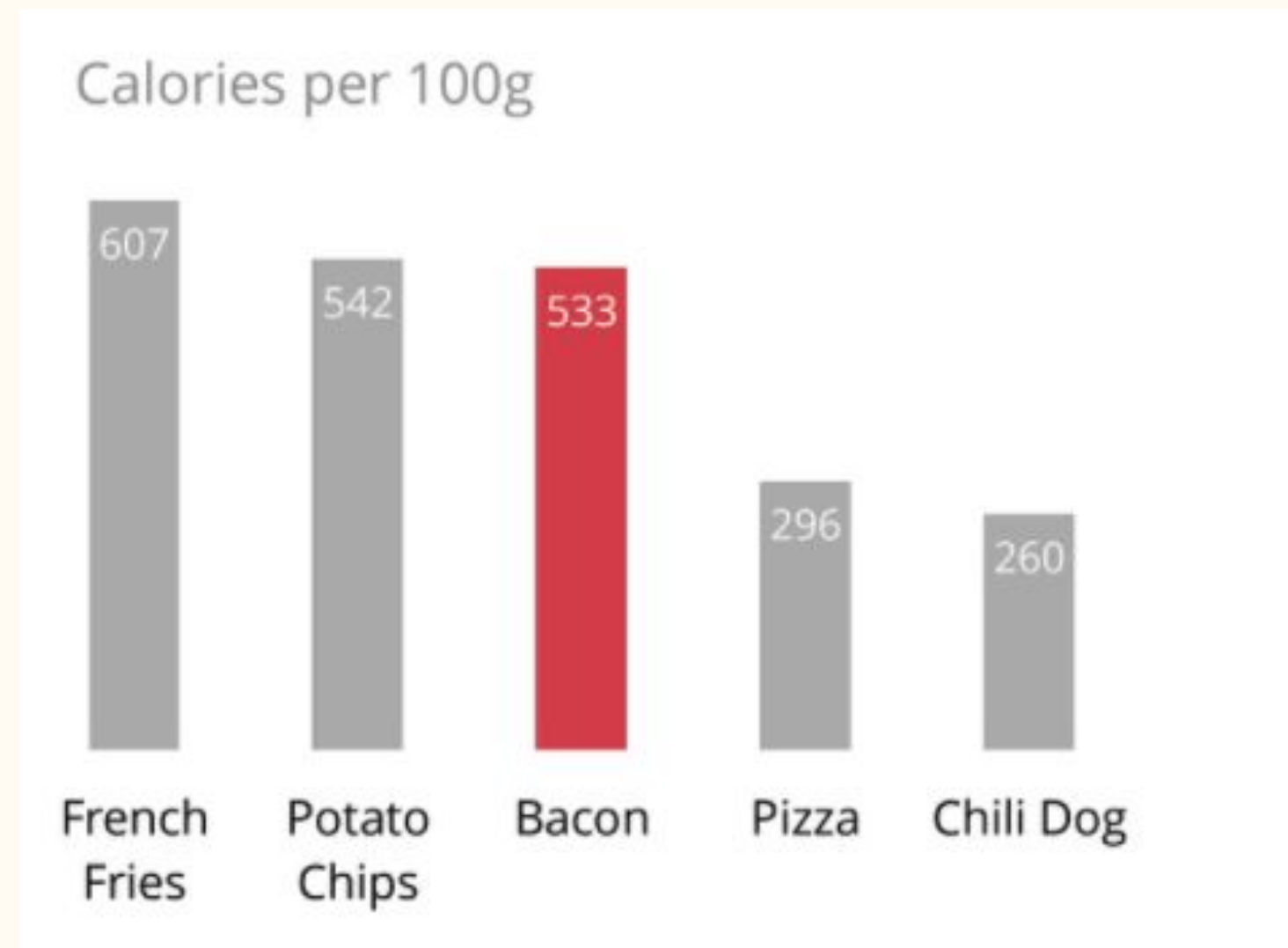
[Source: Darkhorse Analytics](#)



# 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](#)



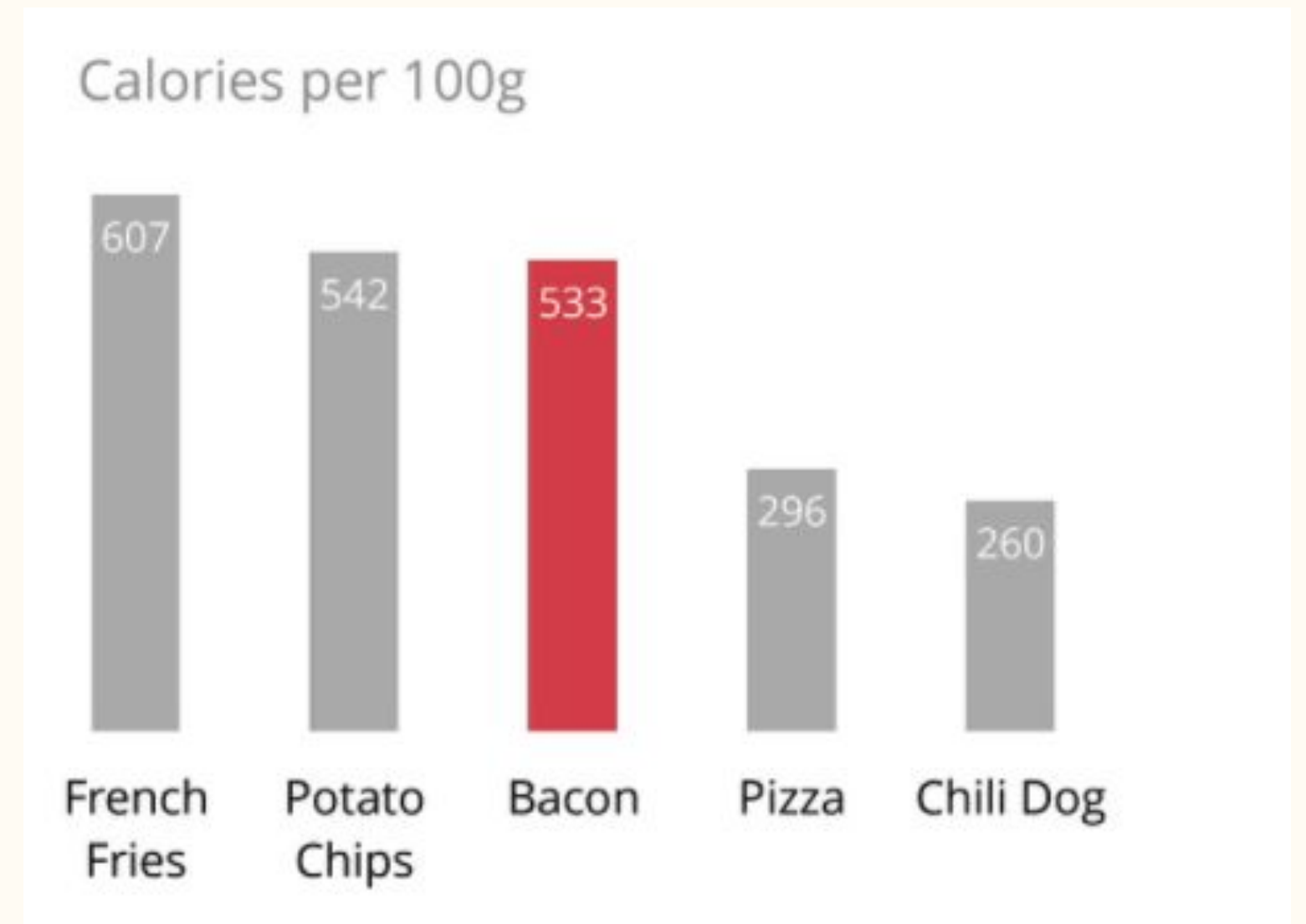
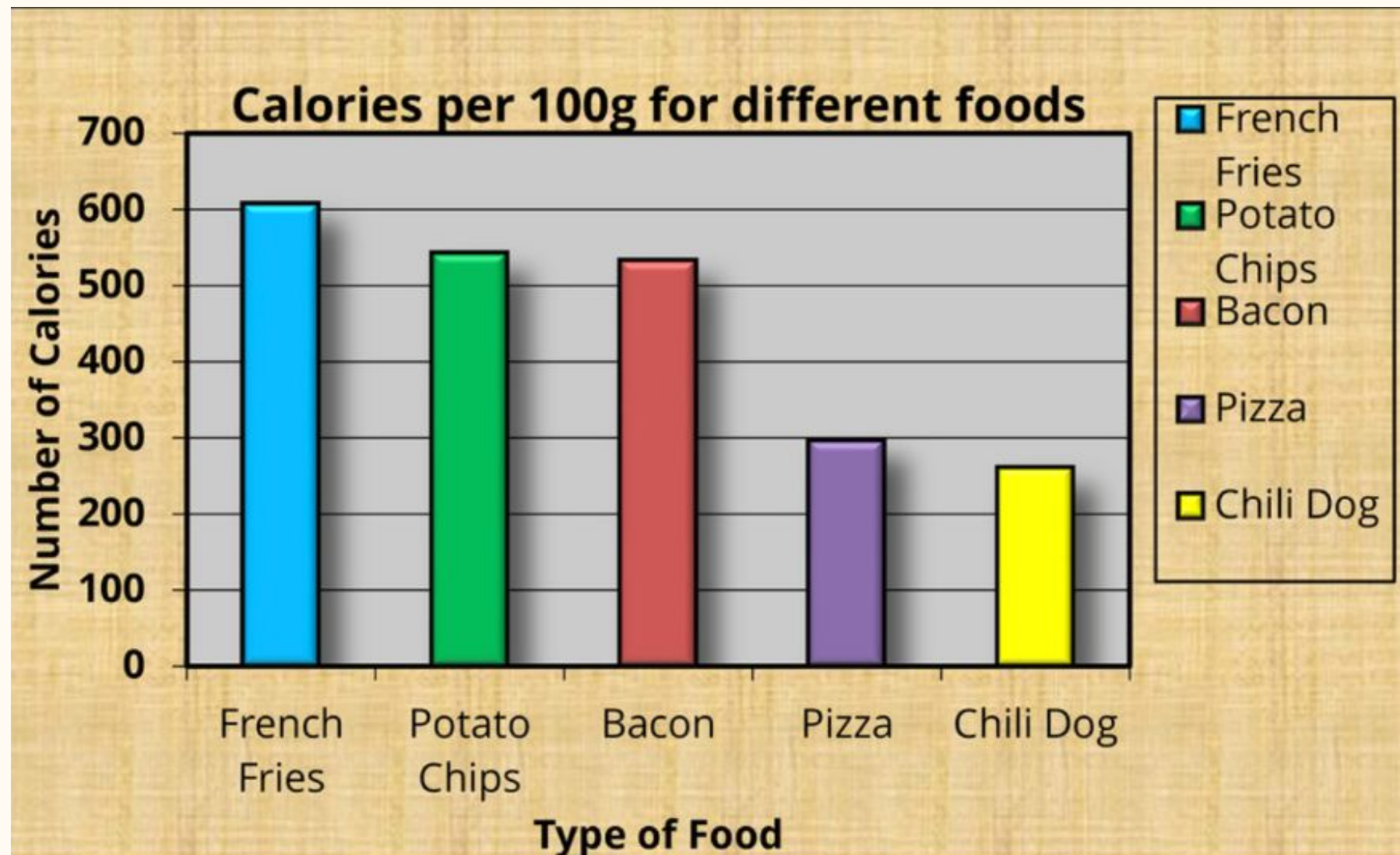
# Decluttering techniques at your disposal

- ✓ 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

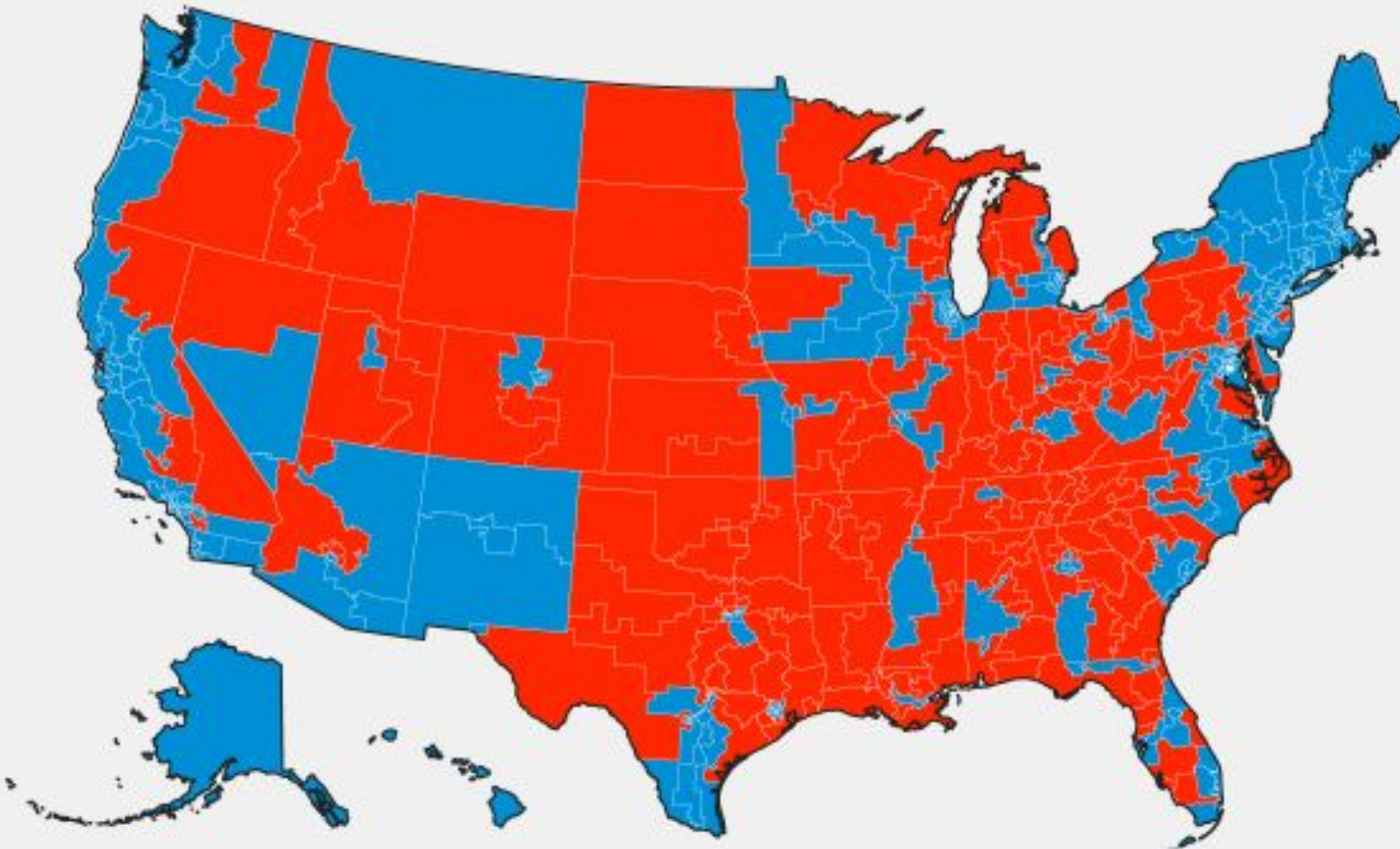
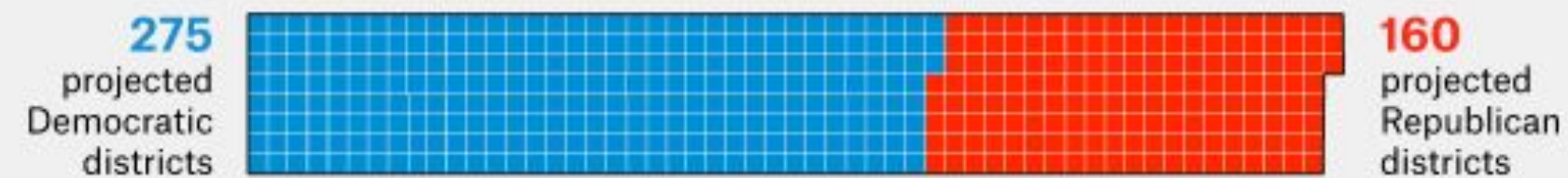
*Use color creatively*



# Colors to distinguish between groups

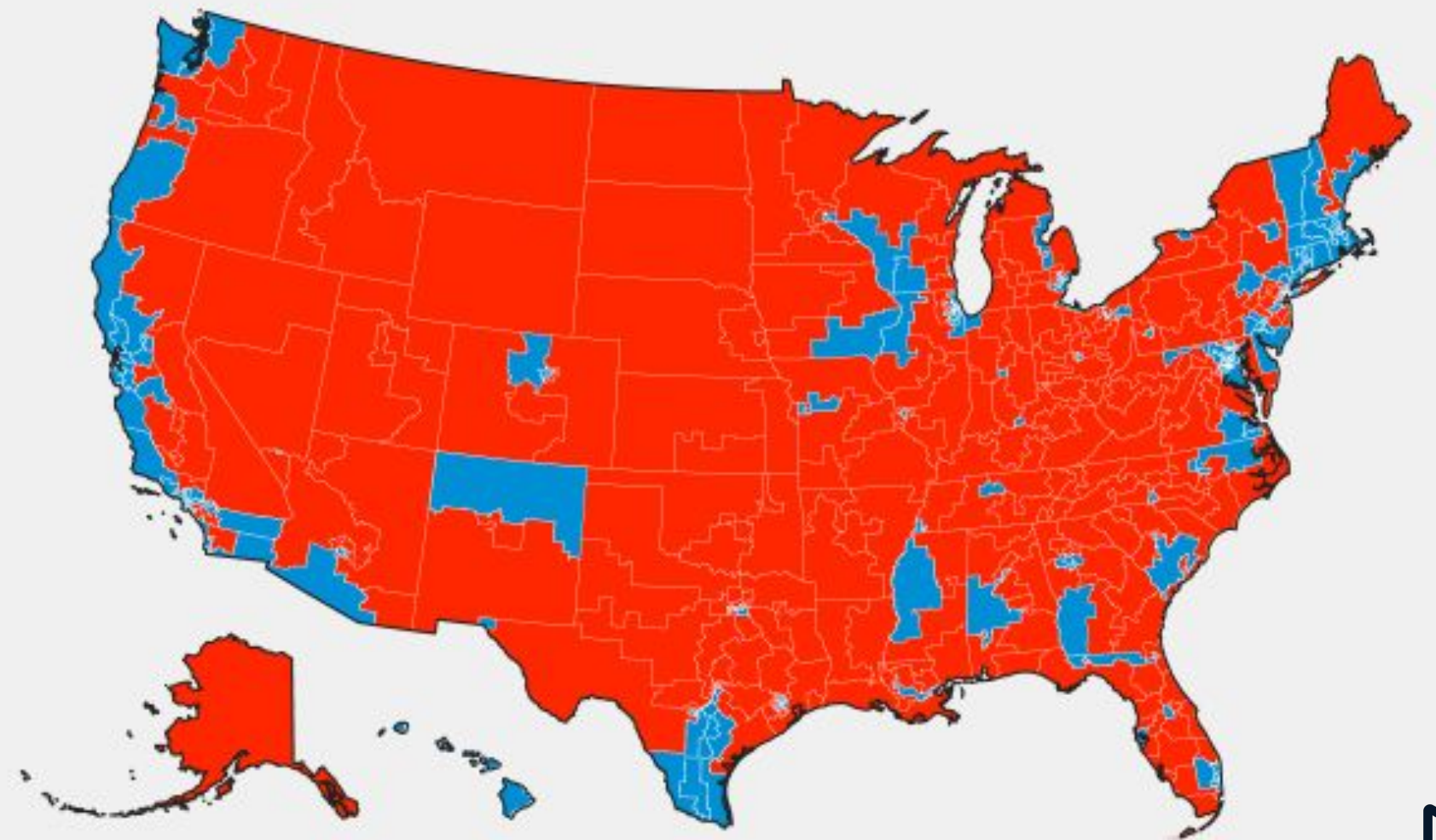
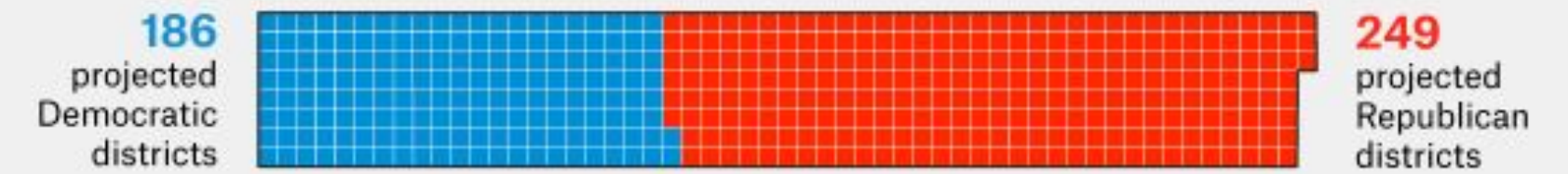
## What if only women voted?

Projected results for the 2018 midterms based on polling patterns and FiveThirtyEight's Lite forecast on Oct. 24



## What if only men voted?

Projected results for the 2018 midterms based on polling patterns and FiveThirtyEight's Lite forecast on Oct. 24



# How this can be used in “normal” datasets

**Customer  
lifetime value**

What drives customer loyalty at your organization?



**Products  
purchased**

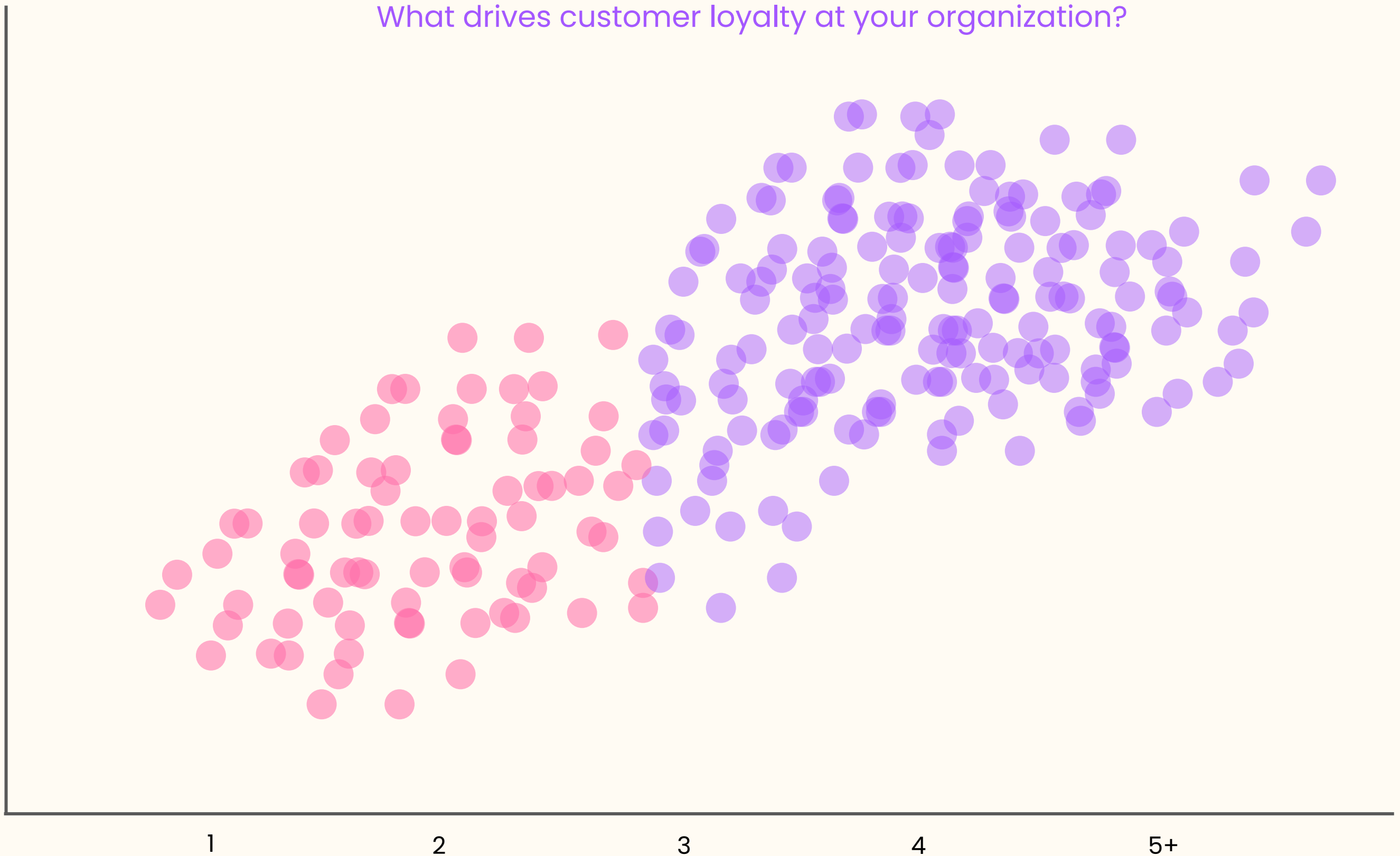


# How this can be used in “normal” datasets

**Customer  
lifetime value**

What drives customer loyalty at your organization?

**Segment A**  
**Segment B**



1

2

3

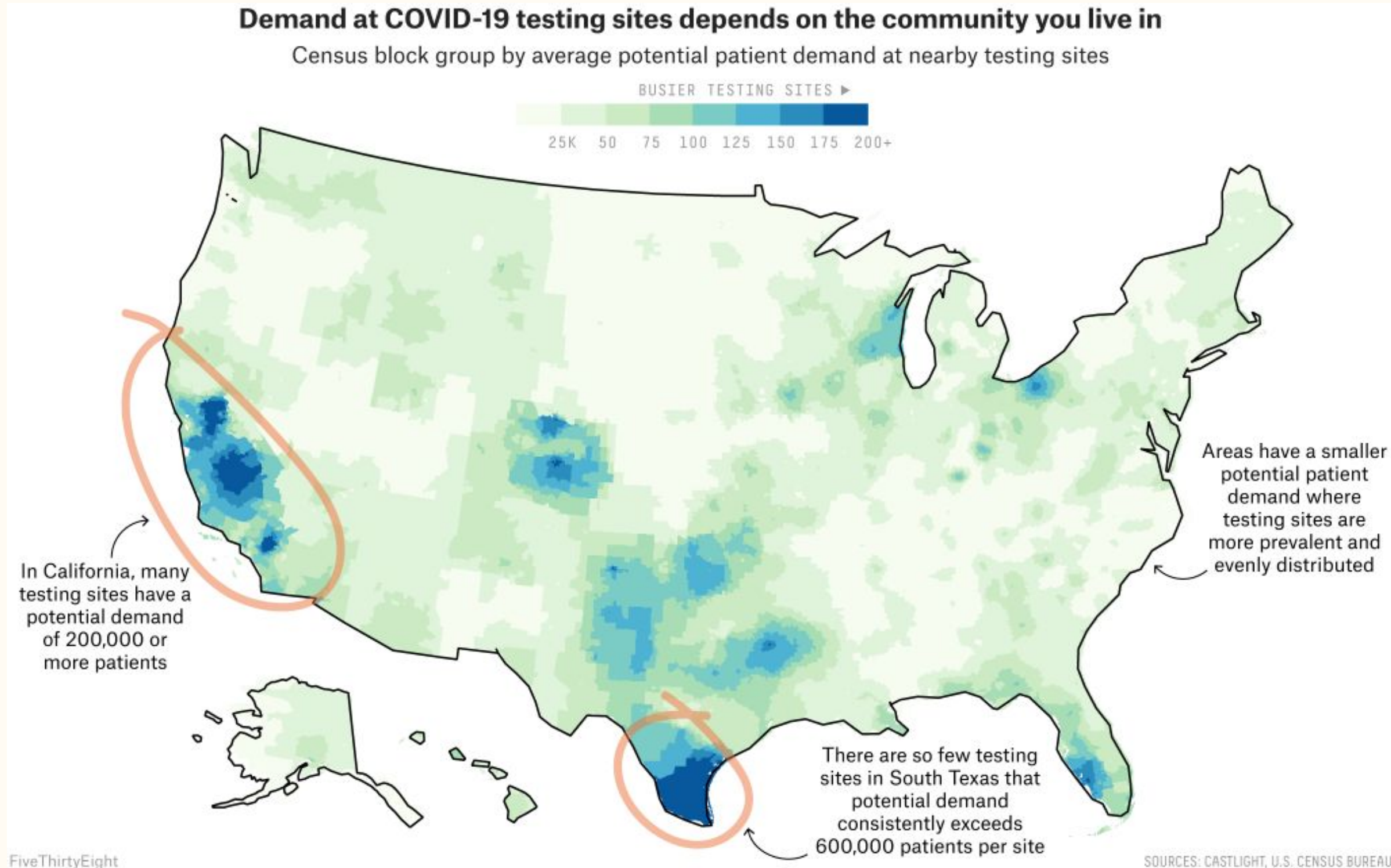
4

5+

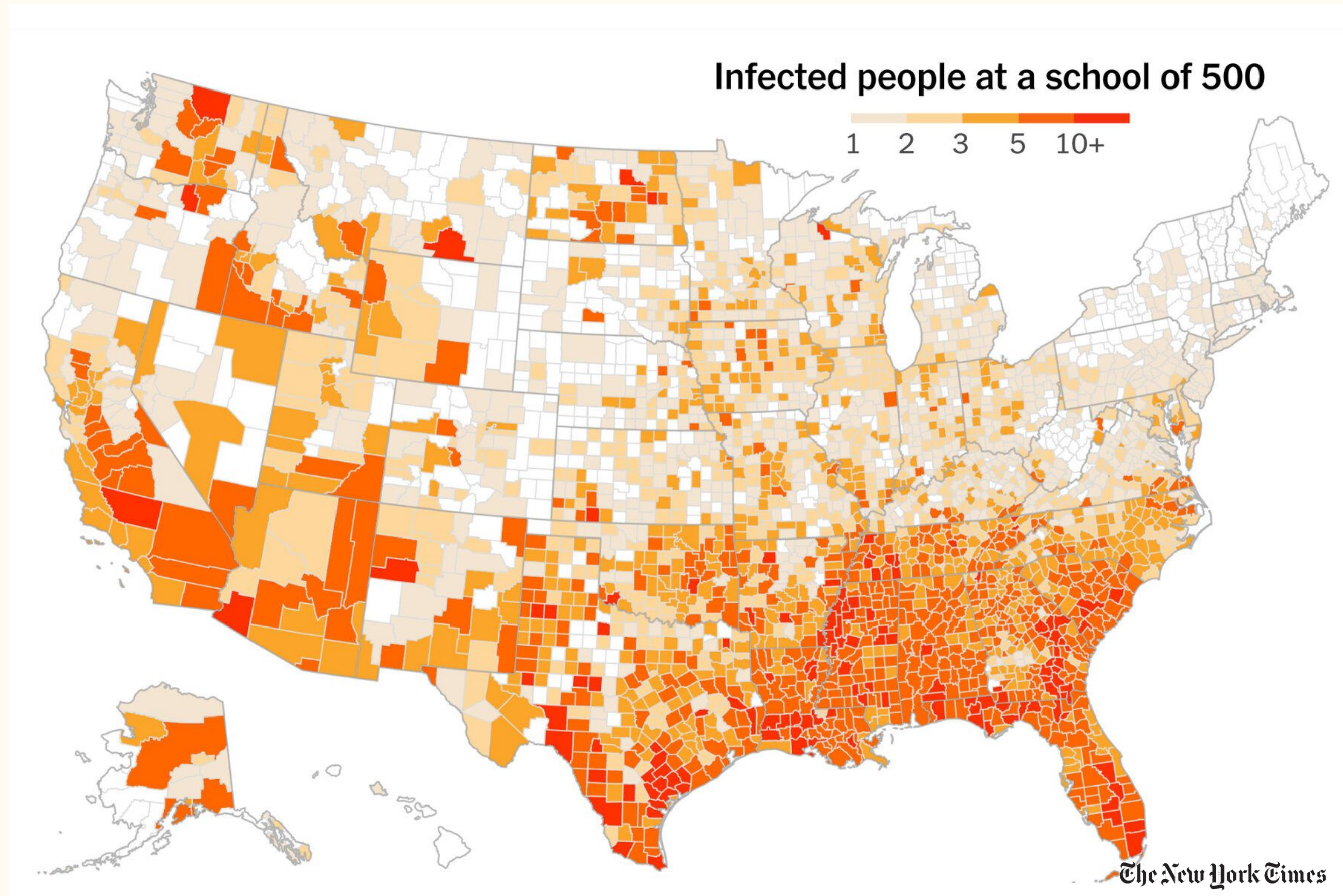
**Products  
purchased**



# Colors to highlight intensity



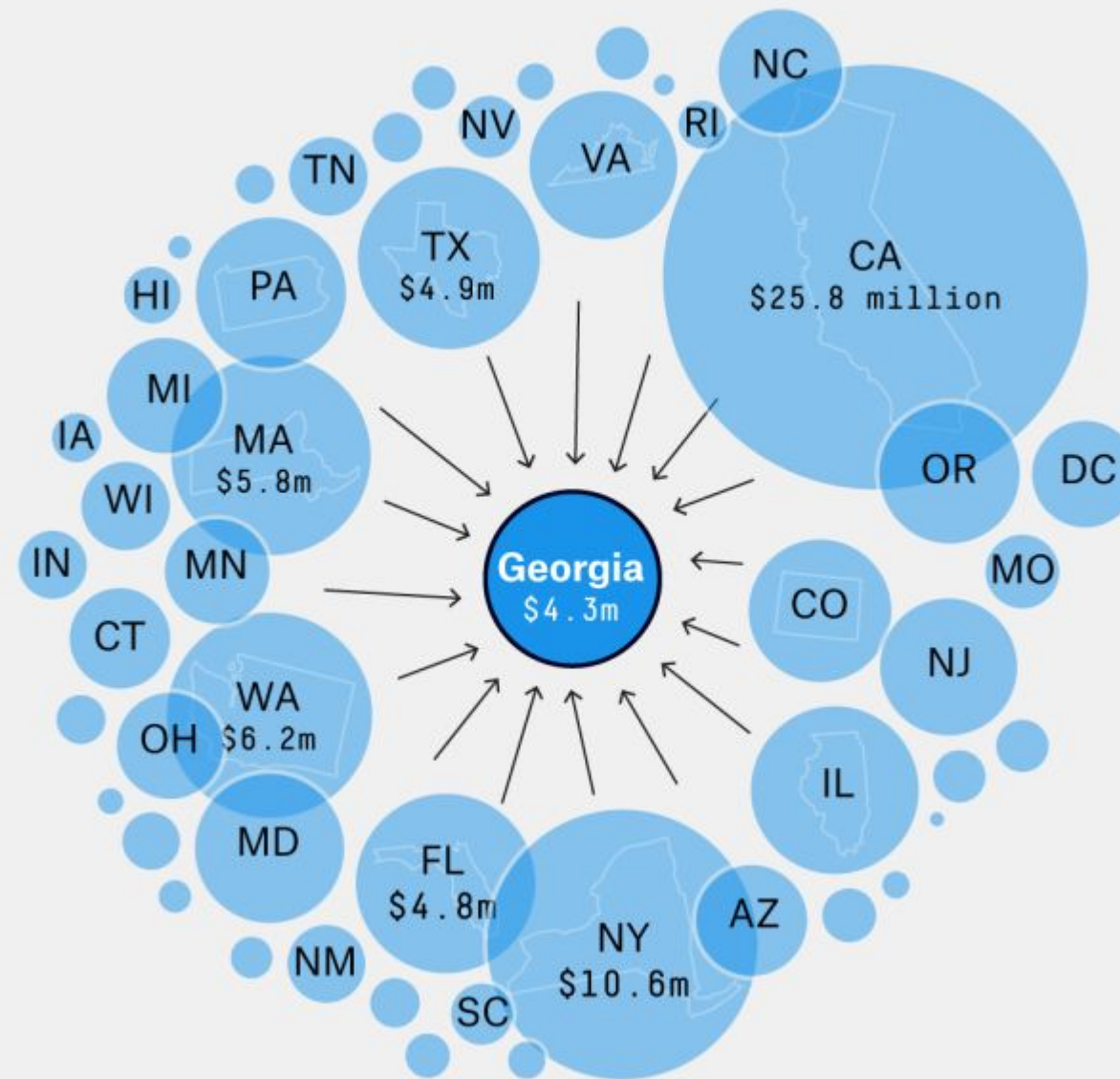
# Visualizing covid infection hotspots



# Colors to distinguish between groups

## Democrats from every state are donating to Georgia

All donations sent to Democratic candidates in Georgia's Senate runoffs via ActBlue, from Nov. 4 through Nov. 23



FiveThirtyEight

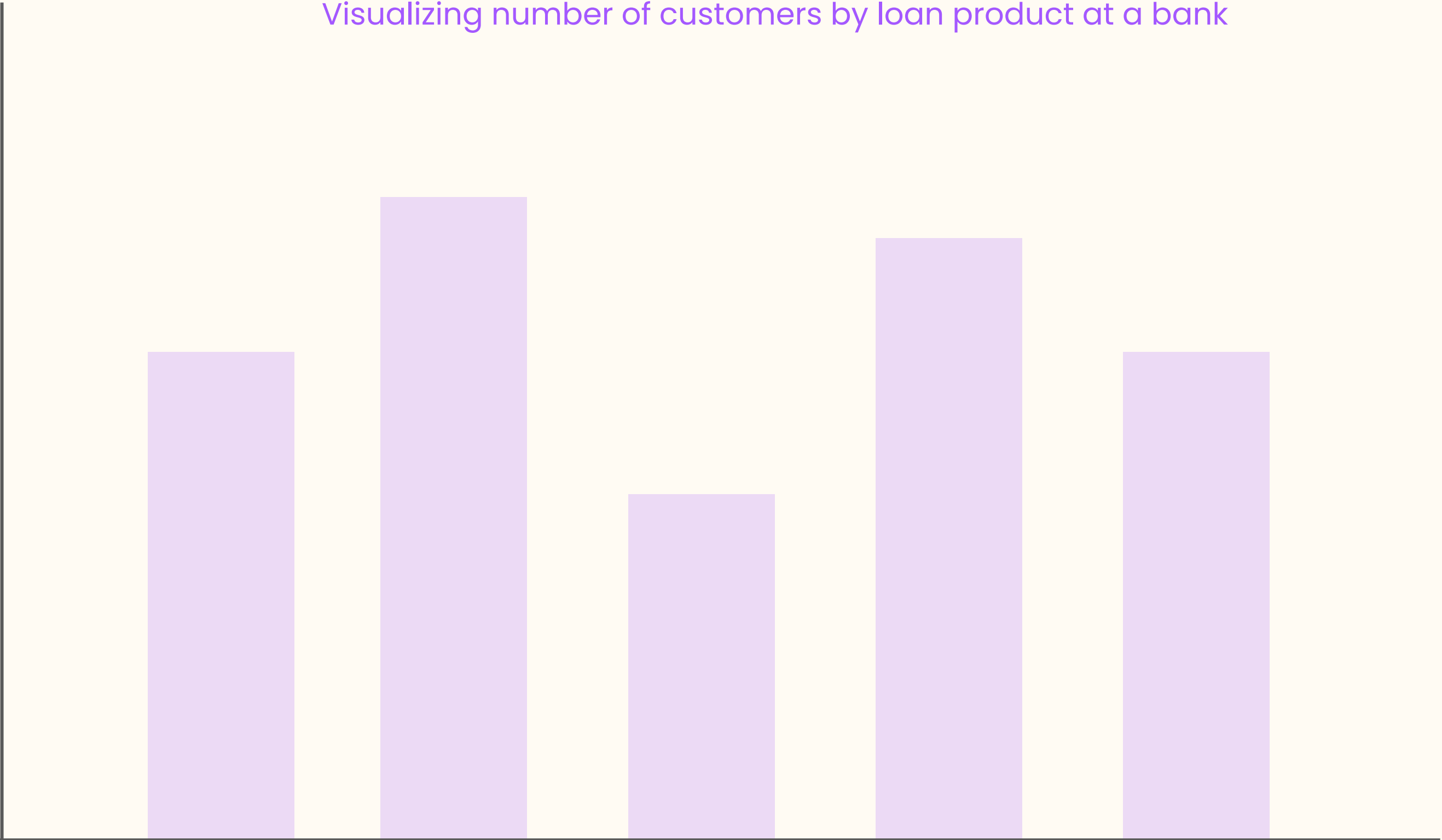
SOURCE: FEDERAL ELECTION COMMISSION



# How this can be used in “normal” datasets

Number of customers

Visualizing number of customers by loan product at a bank



Business Loans

Loan Against Property

Commercial Vehicle Financing

Construction Equipment Loan

Farm Equipment Loan

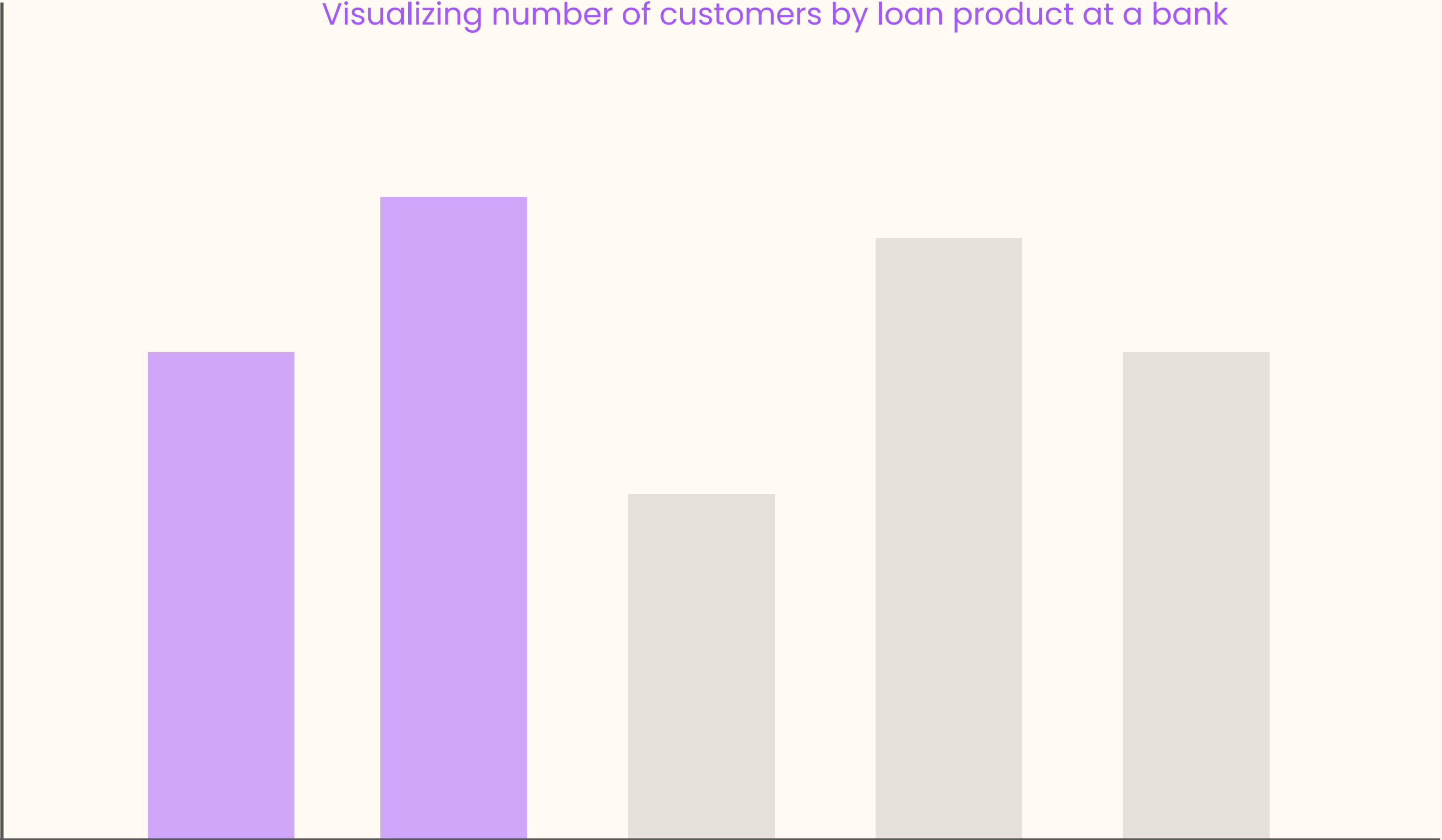
Loan Products



# How this can be used in “normal” datasets

Number of customers

Visualizing number of customers by loan product at a bank



Business Loans

Loan Against Property

Commercial Vehicle Financing

Construction Equipment Loan

Farm Equipment Loan

Loan Products



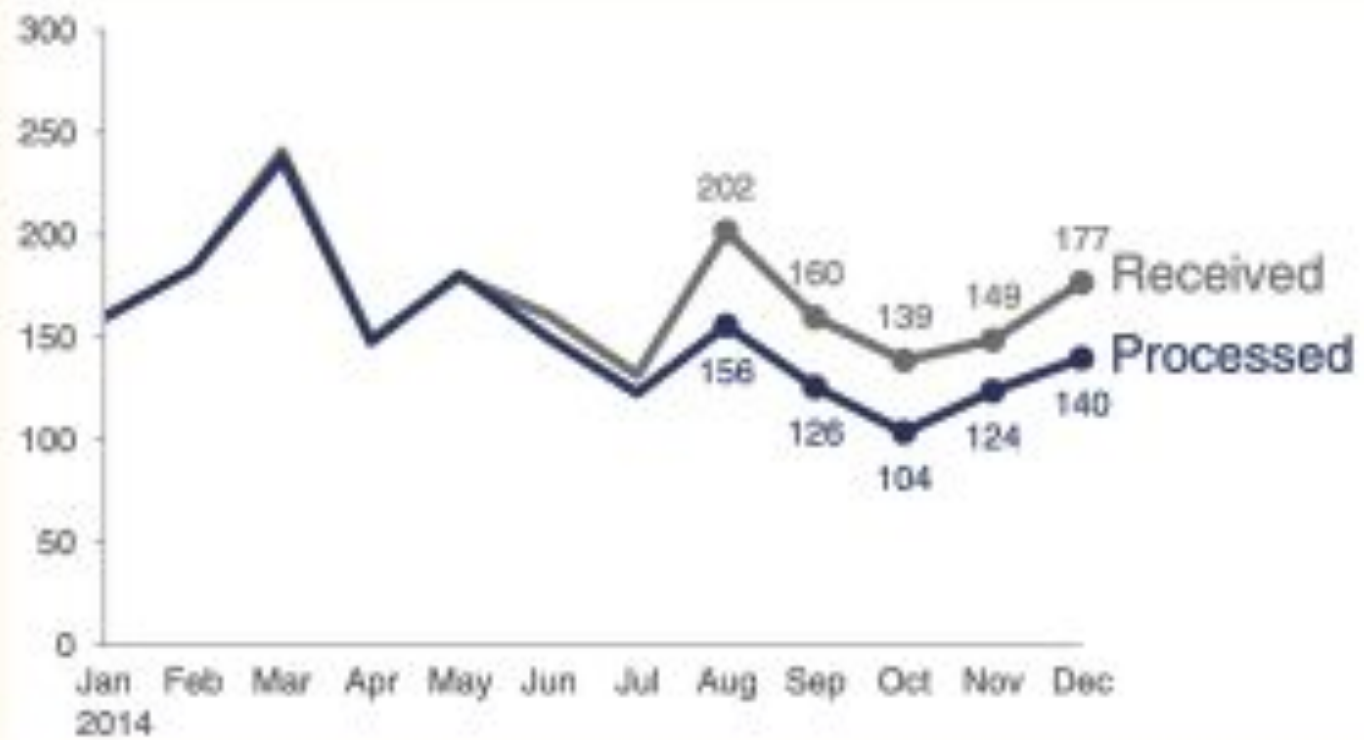
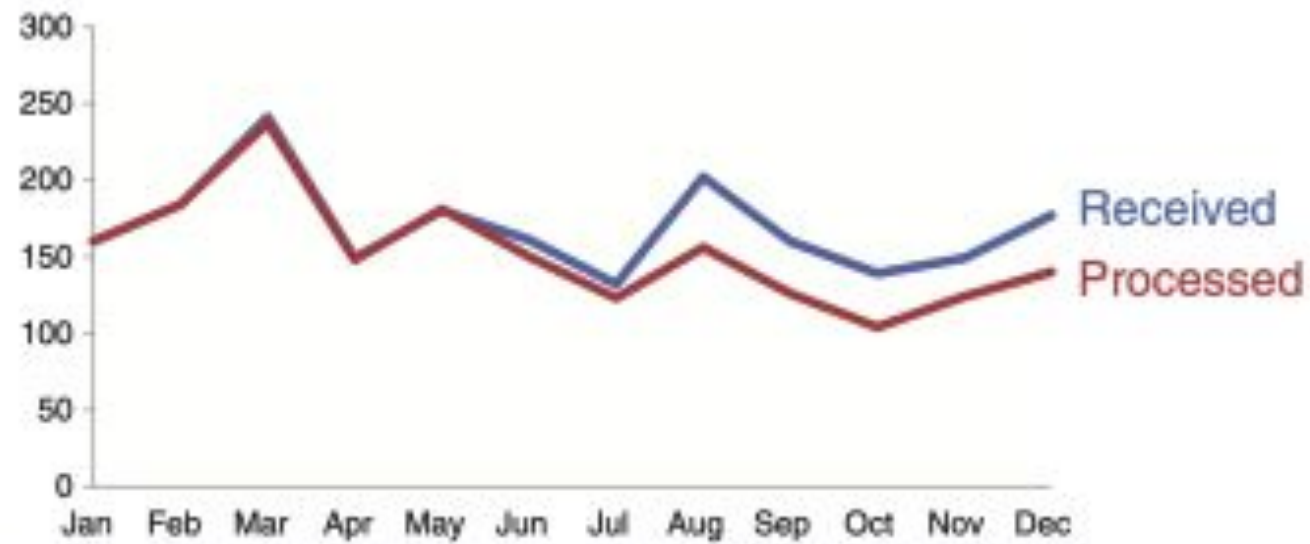




## Rule #4

*Use texts appropriately*

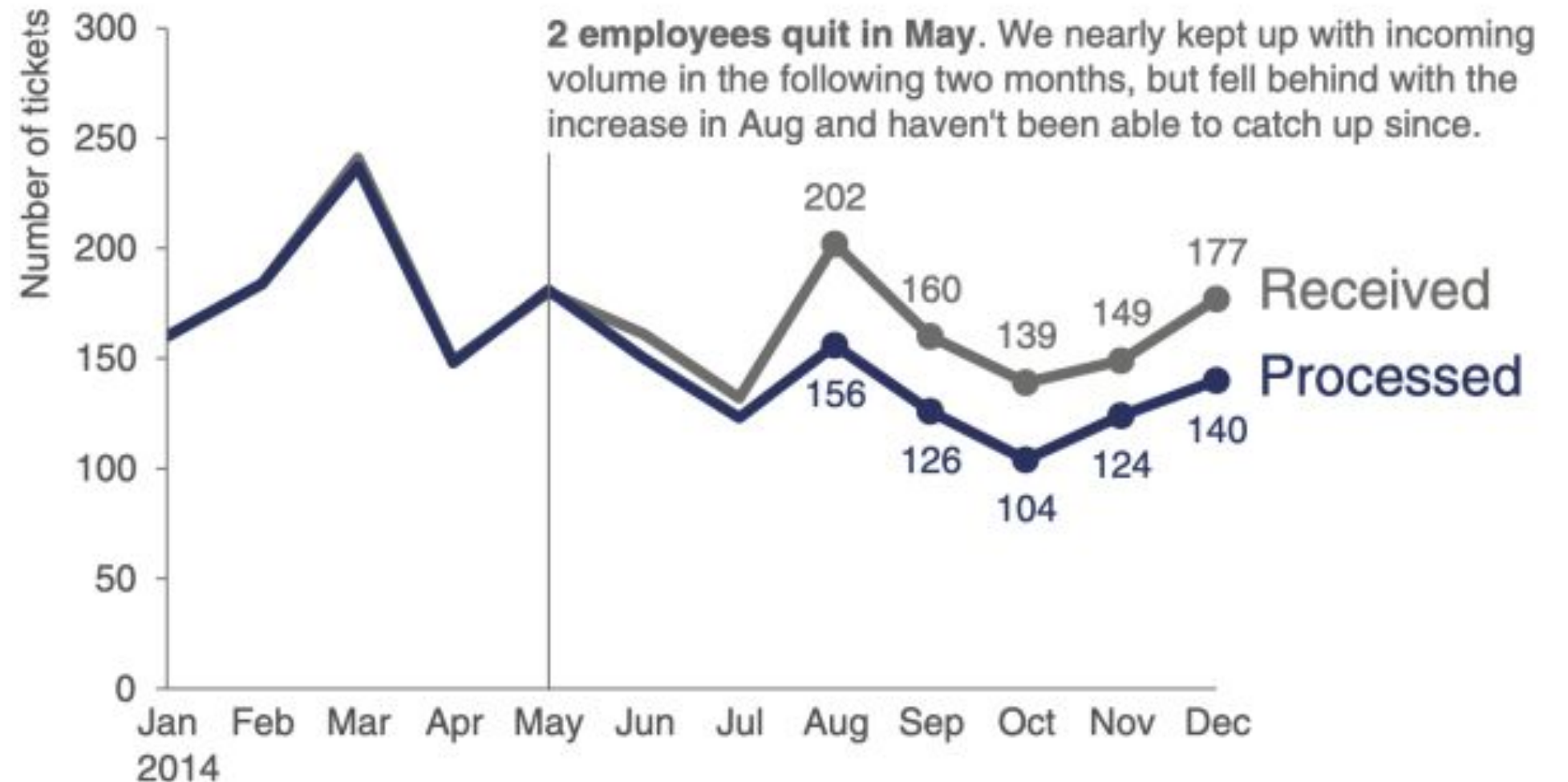
# Labels can be extremely effective at highlighting insights



## Please approve the hire of 2 FTEs

to backfill those who quit in the past year

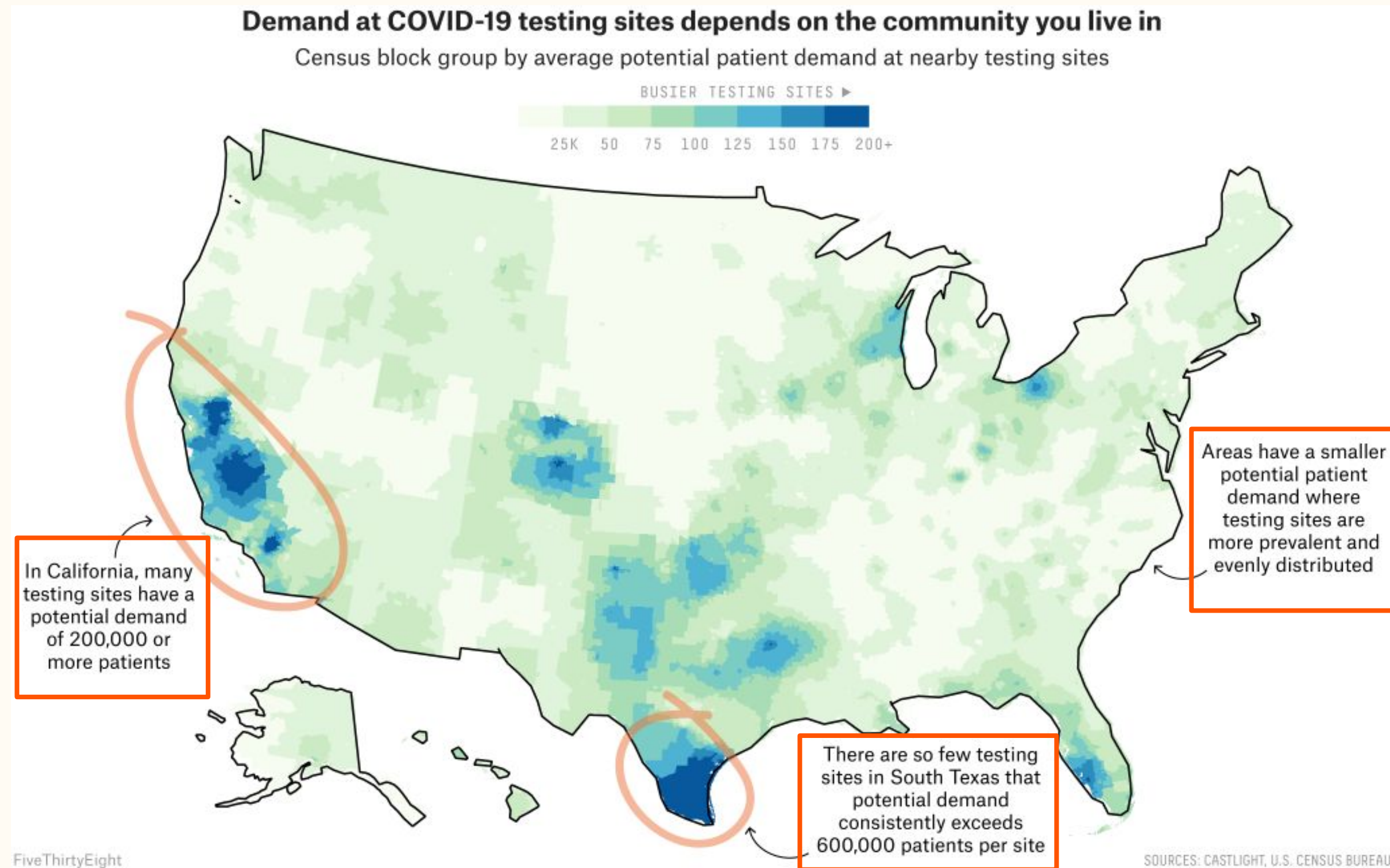
### Ticket volume over time



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*

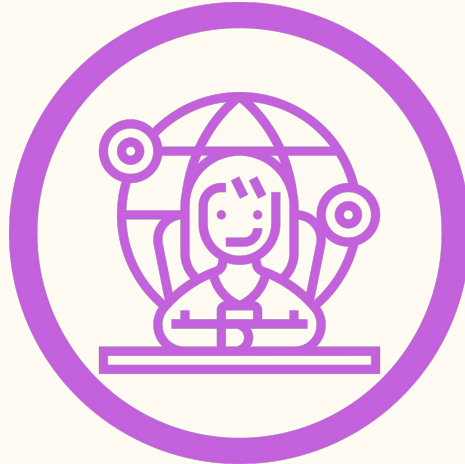


# Rule #1

*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**



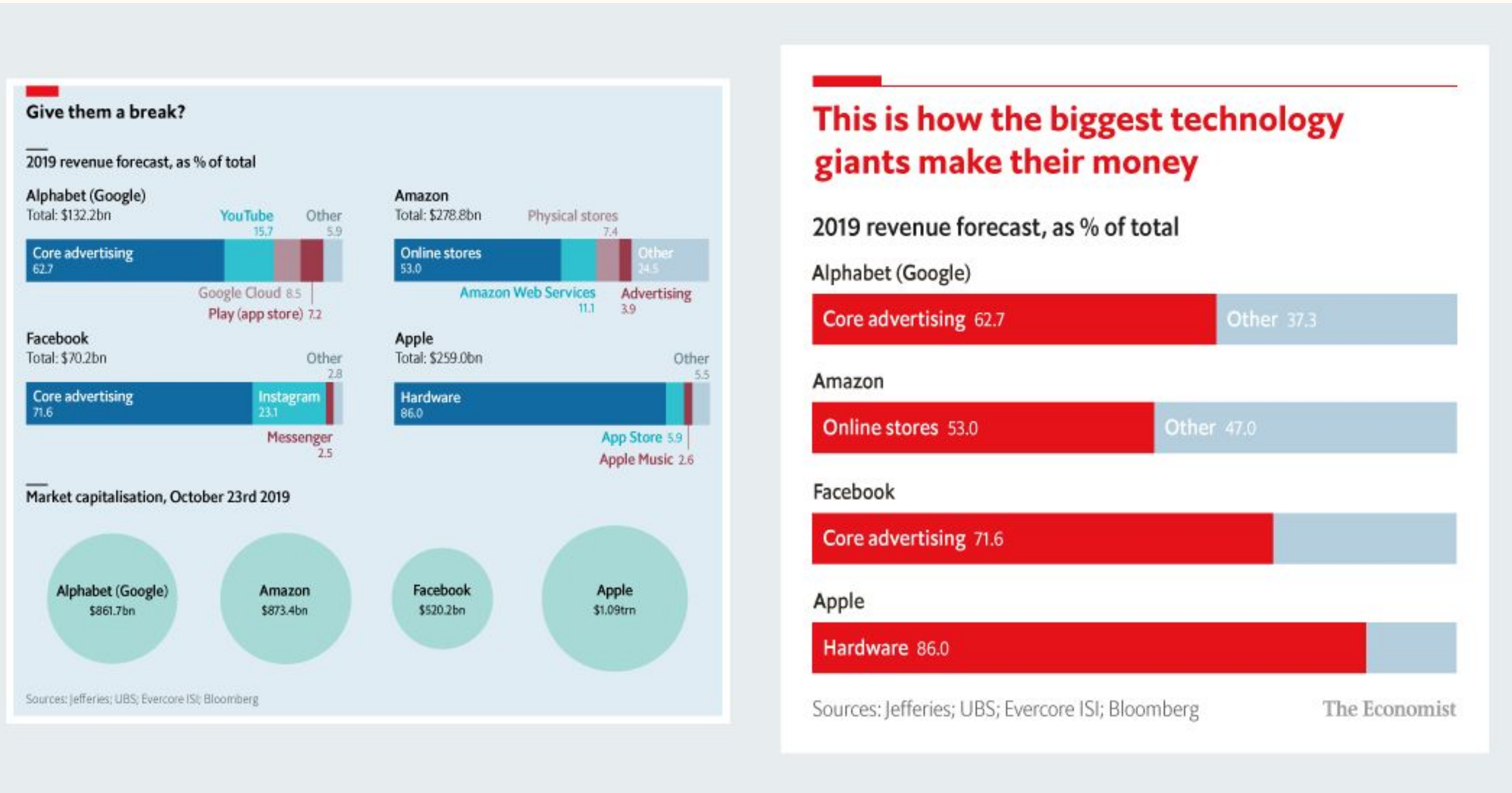


# Practice empathy with your audience

- ✓ 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**

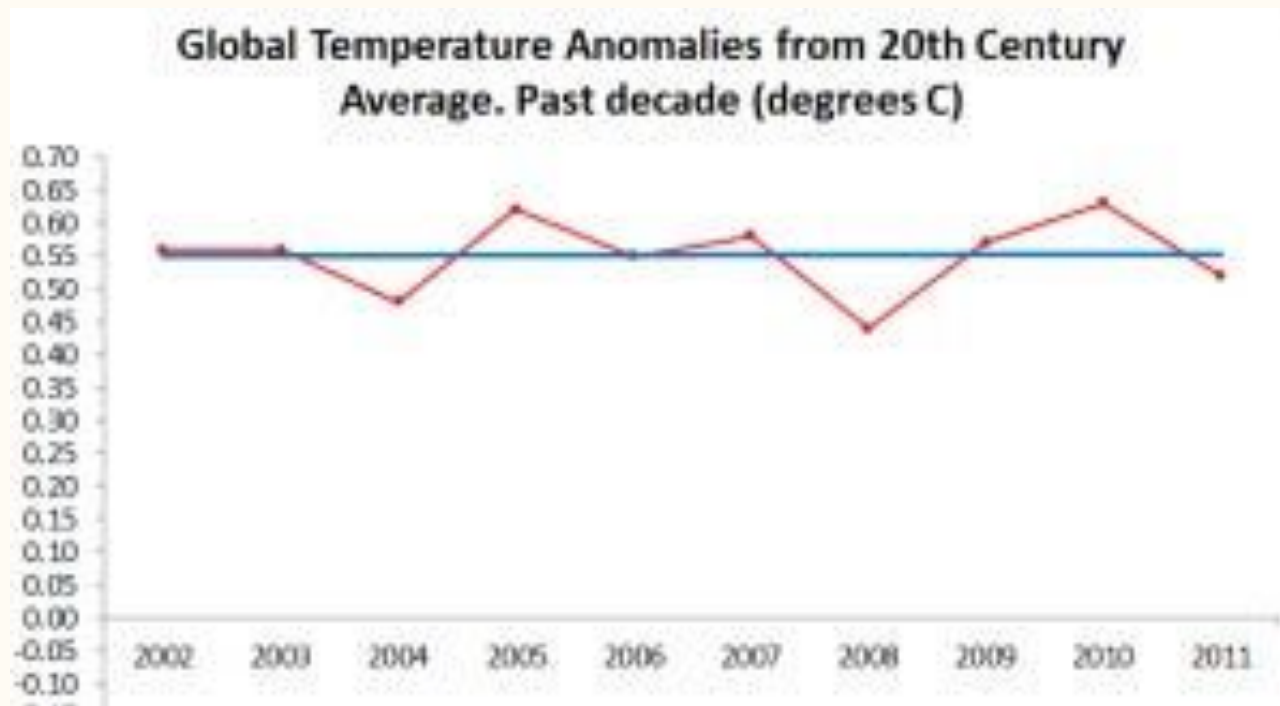




## Rule #3

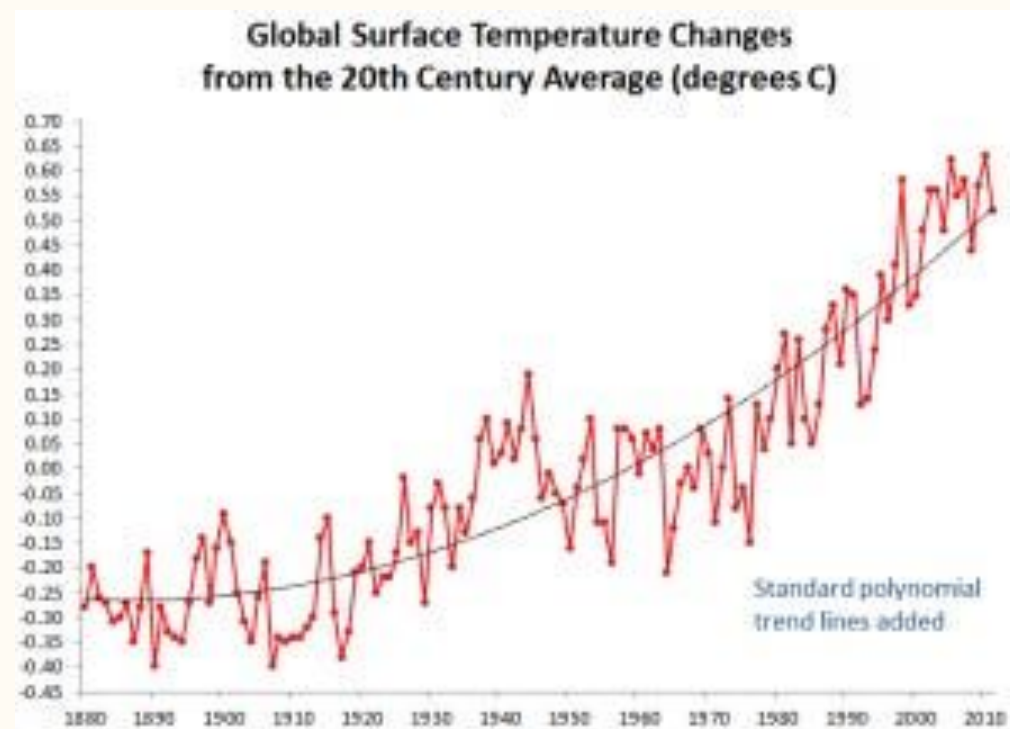
*Do not mislead with data stories*

# Avoiding the fastest way to lose credibility



## Pitfall to avoid #1

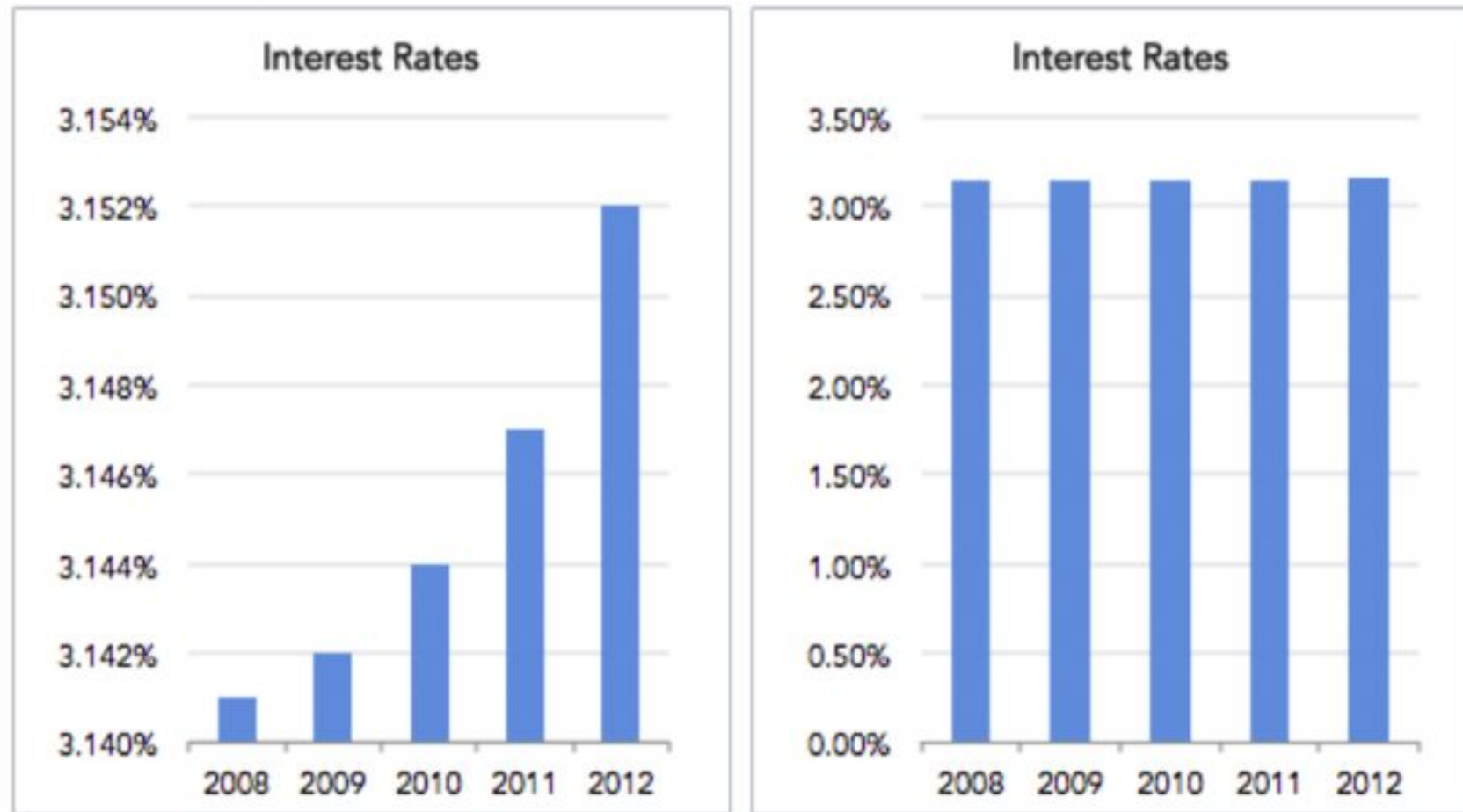
*Always make sure your time horizons make sense given the data you're treating*





# Avoiding the fastest way to lose credibility

Same Data, Different Y-Axis



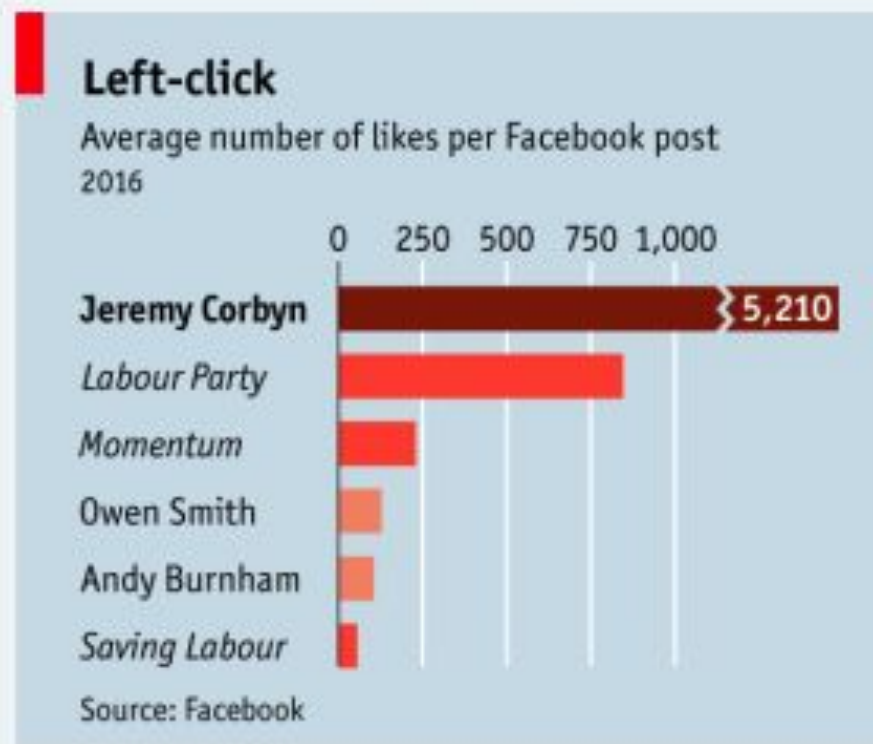
Misleading Graph No. 1

**Pitfall to avoid #2**  
*Ensure that axes start with zero*

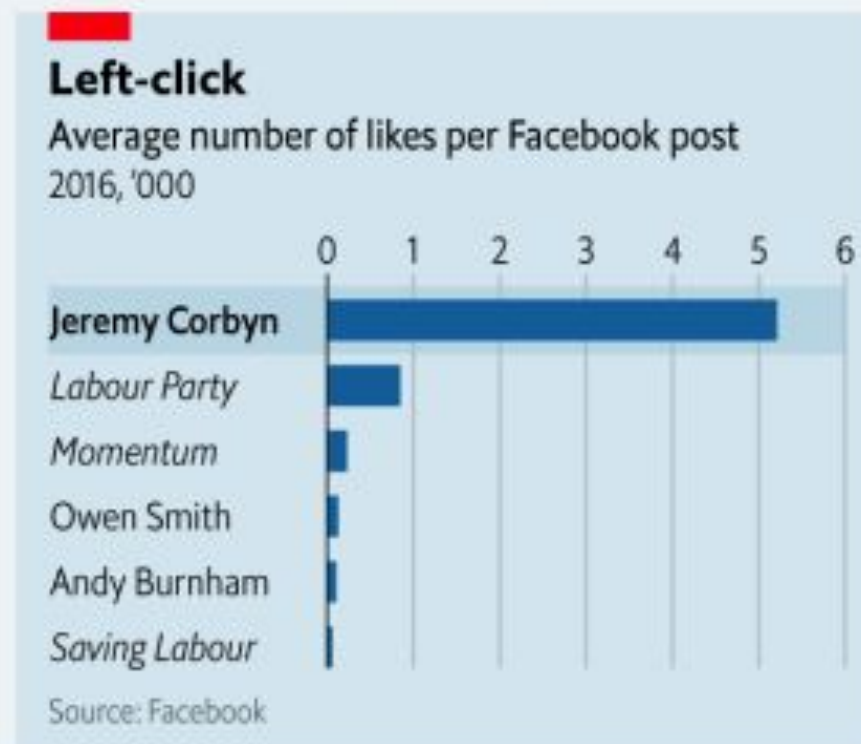


# Avoiding the fastest way to lose credibility

Original



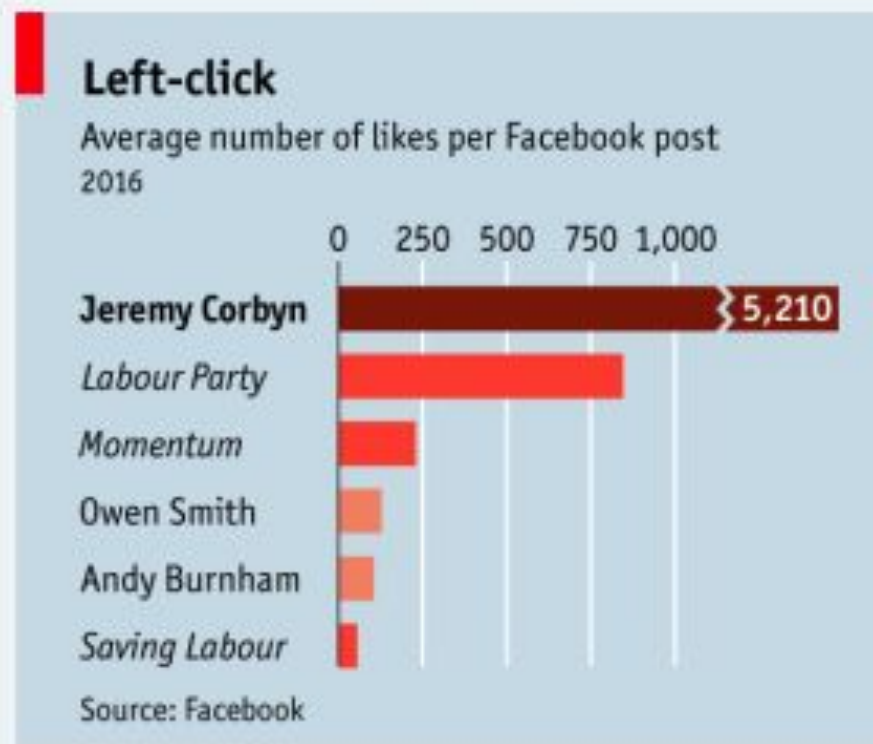
Better



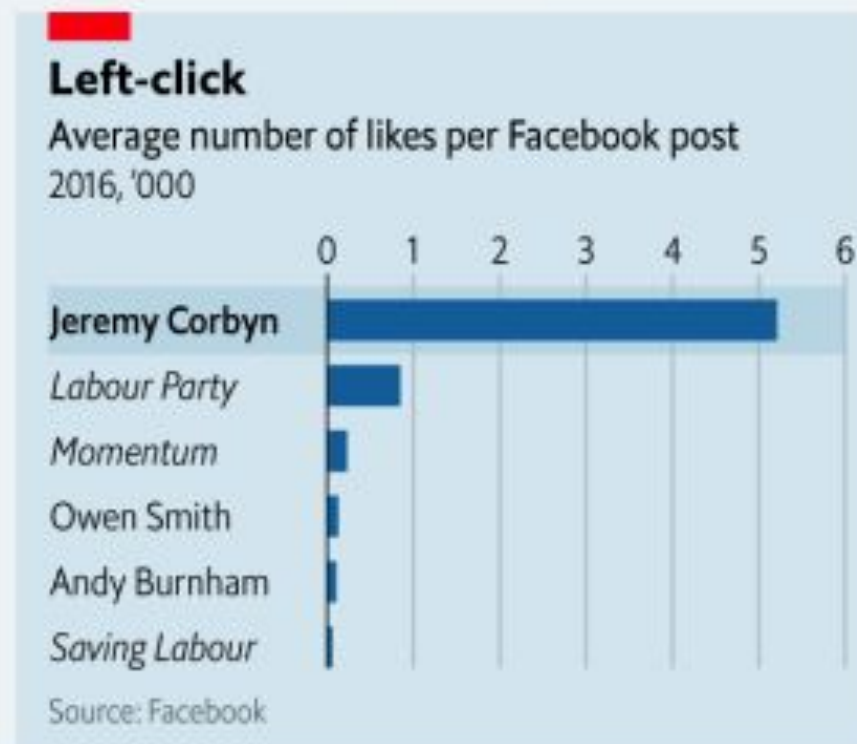
**Pitfall to avoid #3**  
*Ensure that axes scales are appropriate*

# Avoiding the fastest way to lose credibility

Original



Better



**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*

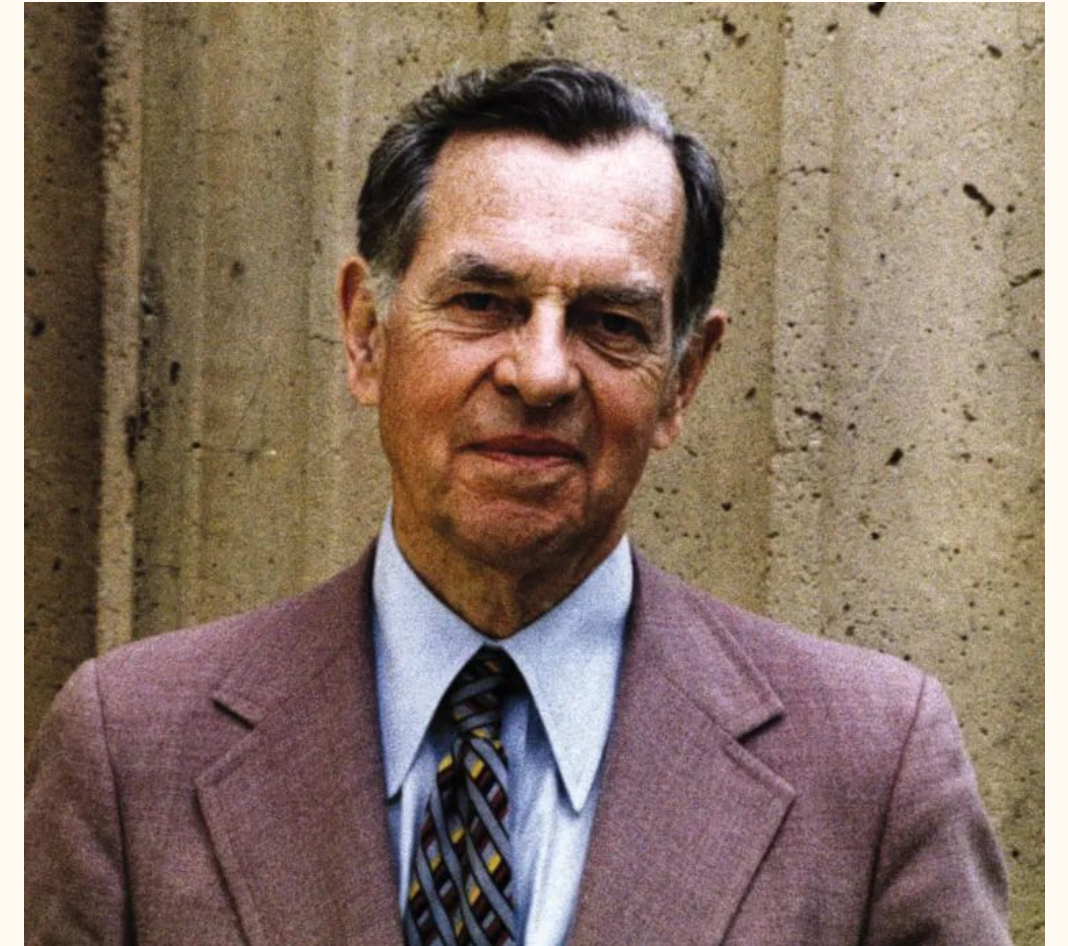
# Different narrative structures to choose from



Aristotle's Tragedy Structure



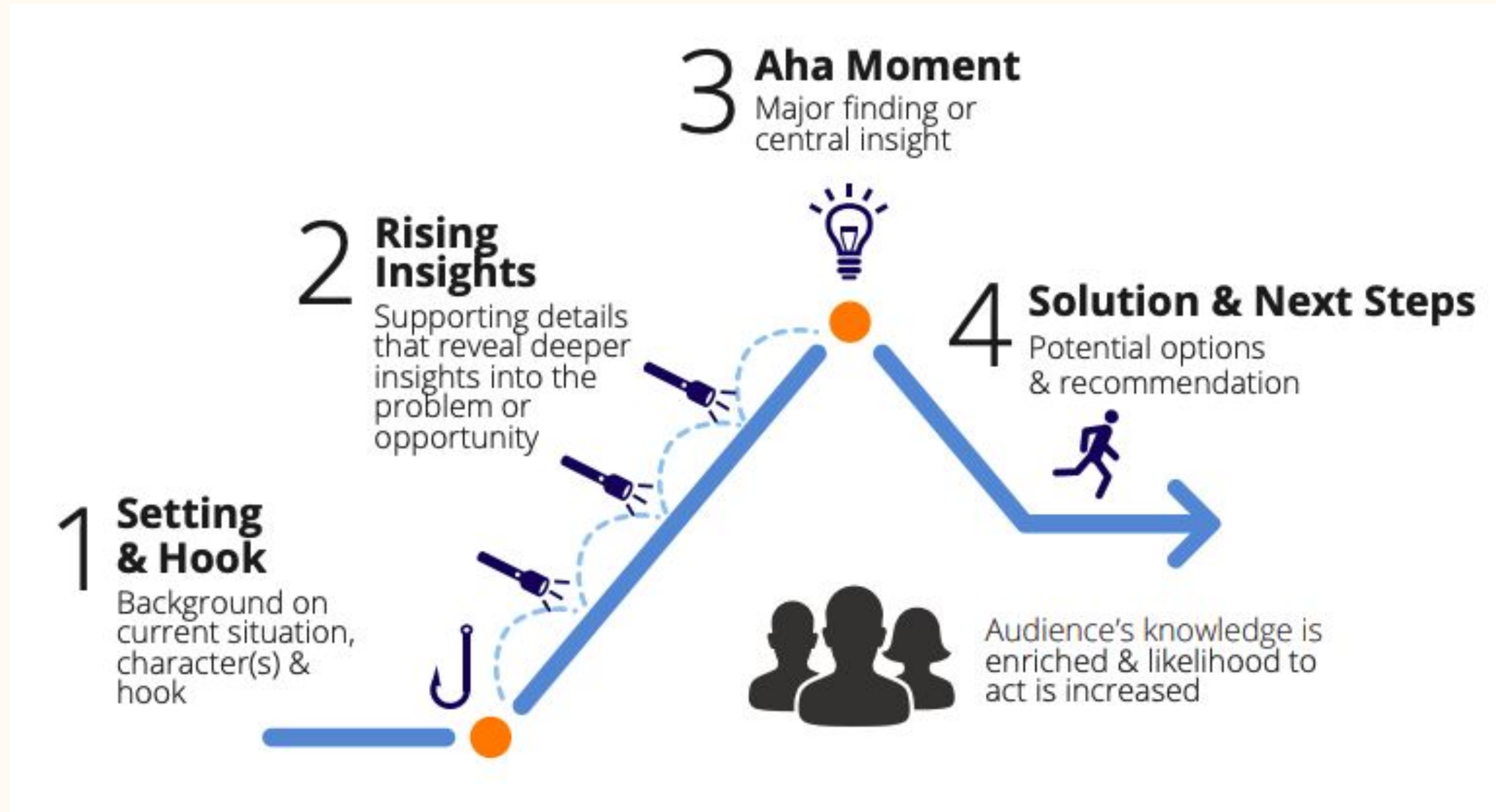
Freytag's Pyramid



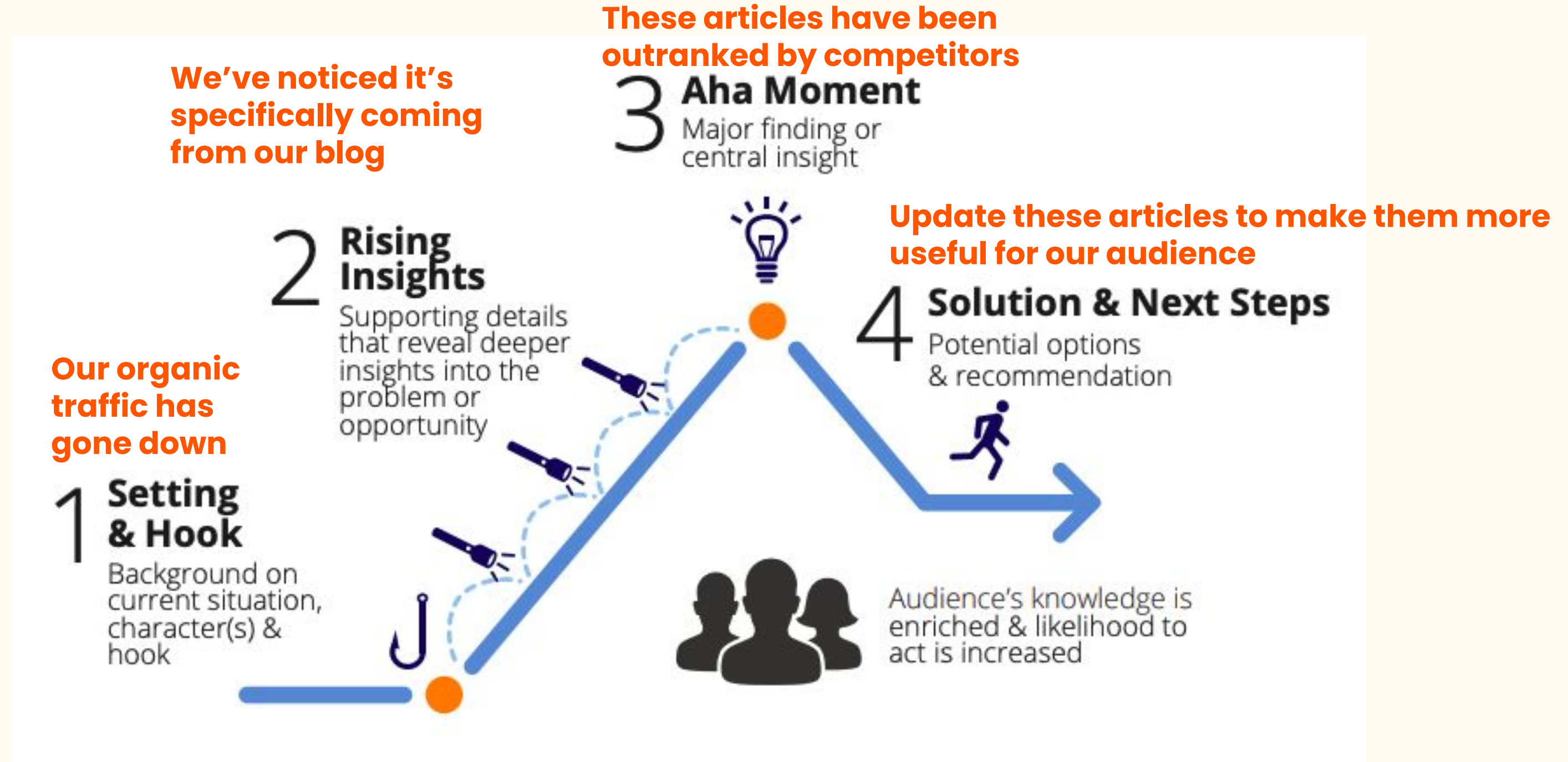
Campbell's Hero Journey



# Different narrative structures to choose from



# 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](#)



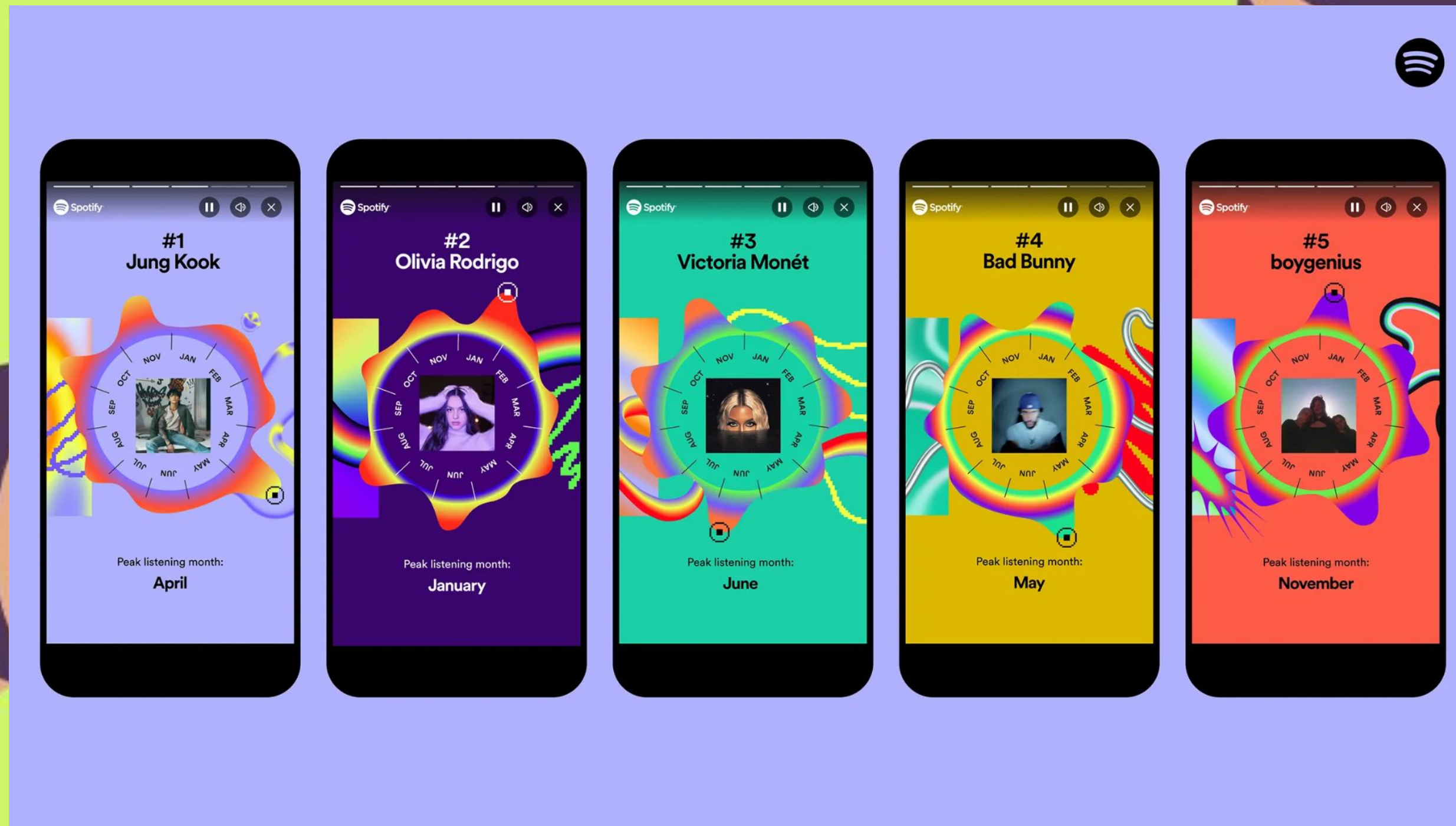


3

Everyone will become a data storyteller

*Resources for your data  
visualization skills*

# Data storytelling is table stakes now



# Data storytelling is table stakes now

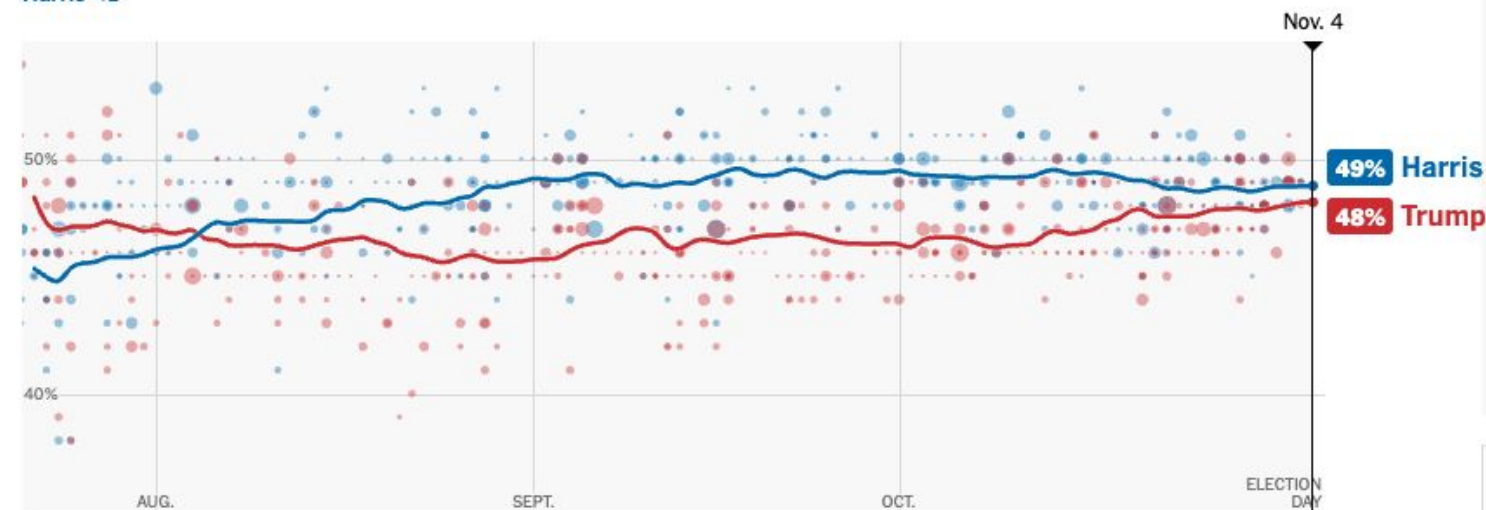
## Election 2024 Polls: Harris vs. Trump

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



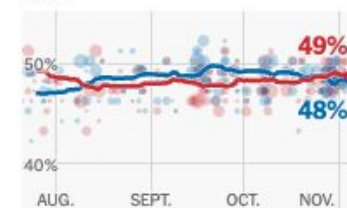
**Nate Cohn**  
Chief political analyst

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*

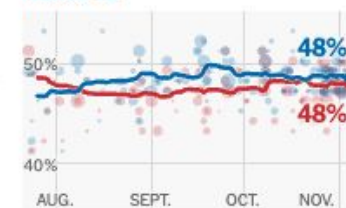


Enjoy open access to the election hub in The Times app.  
Download The Times app to explore the hub, for a limited time.

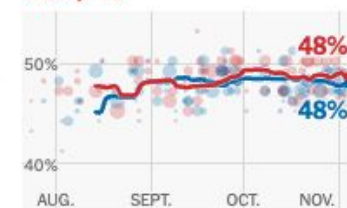
**Pennsylvania** >  
Even



**Michigan** >  
Harris <1



**North Carolina** >  
Trump <1



**Nevada** >  
Trump <1



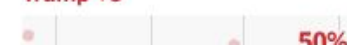
**Wisconsin** >  
Harris <1



**Georgia** >  
Trump +1



**Arizona** >  
Trump +3



datacamp

← Course Outline → Daily XP

Exercise

### Customizing heatmaps

Seaborn supports several types of additional customizations to improve the output of a heatmap. For this exercise, we will continue to use the Daily Show data that is stored in the `df` variable but we will customize the output.

Instructions 100 XP

- Create a crosstab table of `Group` and `YEAR`
- Create a heatmap of the data using the `BuGn` palette
- Disable the `cbar` and increase the `linewidth` to 0.3

Take Hint (-30 XP)

```
script.py
1 # Create the crosstab DataFrame
2 pd_crosstab = pd.crosstab(df["Group"], df["YEAR"])
3
4 # Plot a heatmap of the table with no color bar and using the BuGn palette
5 sns.heatmap(pd_crosstab, cbar=False, cmap="BuGn", linewidths=0.3)
6
7 # Rotate tick marks for visibility
8 plt.yticks(rotation=0)
9 plt.xticks(rotation=90)
10
11 # Show the plot
12 plt.show()
13 plt.clf()
```

Run Code Submit Answer

IPython Shell Slides

In [1]:

# Level up your data visualization skills with DataCamp

Visualize intensity with Python using heatmaps

[View Track](#)



Learn / Courses / Introduction to Tableau

Tableau Public - 1\_3\_new\_york\_neighborhood\_prices

File Data Worksheet Dashboard Story Analysis Map Format Server Window Help

new\_york

Columns: Neighborhood

Rows: Neighborhood

Sheet 1

| Neighborhood       | Price |
|--------------------|-------|
| Allerton           | Abc   |
| Alphabet City      | Abc   |
| Annadale           | Abc   |
| Astoria            | Abc   |
| Bath Beach         | Abc   |
| Battery Park City  | Abc   |
| Bay Ridge          | Abc   |
| Baychester         | Abc   |
| Bayside            | Abc   |
| Bedford Park       | Abc   |
| Bedford-Stuyvesant | Abc   |
| Bensonhurst        | Abc   |
| Bergen Beach       | Abc   |
| Boerum Hill        | Abc   |
| Borough Park       | Abc   |
| Brighton Beach     | Abc   |
| Bronxdale          | Abc   |
| Brooklyn           | Abc   |
| Brooklyn Heights   | Abc   |
| Brooklyn Navy Yard | Abc   |
| Brownsville        | Abc   |
| Bushwick           | Abc   |
| Canarsie           | Abc   |

Context Instructions

1 2 3 4

Drag the Price field to the Text Marks card.

Hint Retwatch Video Next

# Level up your data visualization skills with DataCamp

Tableau Fundamentals

[View Track](#)



datacamp Home Learn Certification Groups DataLab Search

Progress My Library Leaderboard Assignments

LEARN

- Tracks
- Courses
- Practice
- Assessments
- Tutorials

APPLY

- Real World Projects
- Code Alongs
- Competitions
- Popular Topics **NEW**

Getting Started (3/4)

### SKILL TRACK

## Design in Power BI

Continue Track

Power BI, Theory 4 hours 2 courses 448 participants

TRACK COMPLETION 0% BONUS MATERIAL - 0 OF 2

#### Track Description

This short track will take your dashboarding and reporting skills to the next level.

#### COURSE

### 1 Dashboard Design Concepts

0%

Learn the skills needed to create impactful dashboards. Understand dashboard design fundamentals, visual analytics components, and dashboard types.

Chapters:

|                                   |         |
|-----------------------------------|---------|
| The Building Blocks               | 650 XP  |
| Visual Analytics Components       | 1000 XP |
| Types of Dashboards               | 900 XP  |
| Dashboard Success and Integration | 750 XP  |

#### INSTRUCTORS

- Olga Scrivner**  
President, Scrivner Solutions Inc
- Maarten Van den Broeck**  
Senior Content Developer at DataCamp

# Level up your data visualization skills with DataCamp

## Design in Power BI

[> View Track](#)



datacamp Home Learn Certification Groups DataLab Search

Progress  
My Library  
Leaderboard  
Assignments

LEARN  
Tracks  
Courses  
Practice  
Assessments  
Tutorials

APPLY  
Real World Projects  
Code Alongs  
Competitions  
Popular Topics **NEW**

Getting Started (3/4)

SKILL TRACK  
**Data Storytelling**  
Start Track

Theory 6 hours 4 courses 7,496 participants

TRACK COMPLETION 0%

**Track Description**  
Discover the art of data storytelling. Transform raw information into memorable narratives.

**COURSE**  
1 **Communicating Data Insights** 0%

Data-driven organizations consistently rely on insights to inspire action and drive change.

**Chapters:**

|                           |         |
|---------------------------|---------|
| Communicating Information | 750 XP  |
| Effective Data Visuals    | 800 XP  |
| Storytelling with Data    | 1050 XP |

View Course Continue

**INSTRUCTORS**

- Joe Franklin**  
Associate Data Literacy and Essentials Manager, DataCamp
- Leondra Gonzalez**  
Sr. Data & Applied Scientist, Microsoft
- Camilo Martinez**  
Data analysis and public policy consultant

# Level up your data visualization skills with DataCamp

## Data Storytelling

[> View Track](#)







# What questions can I answer **for you?**

## *Additional Resources*



[Take DataCamp's Data Maturity Assessment](#)



[Learn more about DataCamp for Business](#)



[WHITE PAPER: Your Organization's Guide to Data Maturity](#)



[ON-DEMAND: Storytelling for more impactful data science](#)



[Register for one of our upcoming webinars](#)



[ON-DEMAND: How Data Governance Enables Scalable Data Science](#)





# Thank you

Adel Nehme  
VP of Media  
[adel@datacamp.com](mailto:adel@datacamp.com)