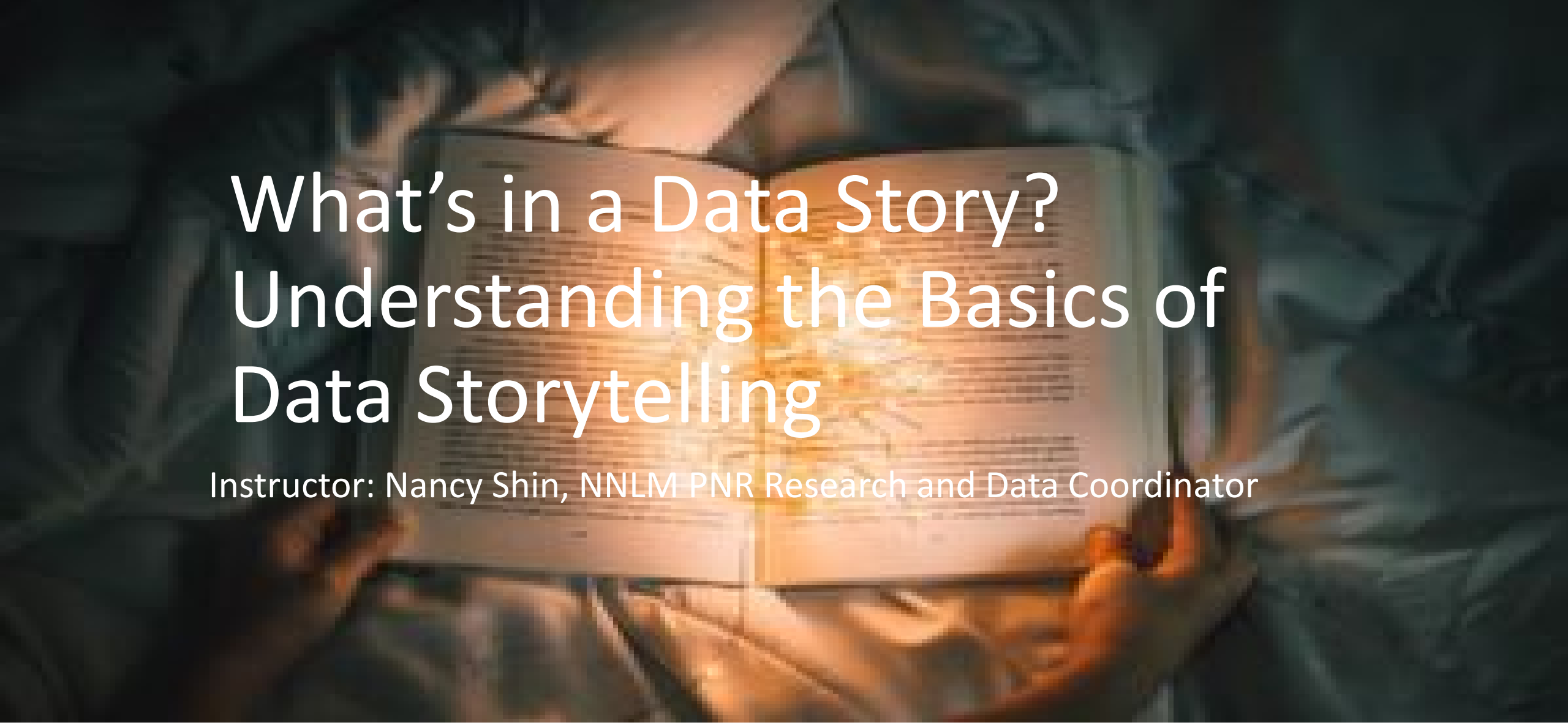


What's in a Data Story? Understanding the Basics of Data Storytelling

Instructor: Nancy Shin, NNLM PNR Research and Data Coordinator

National Library of Medicine
Network of the National Library of Medicine

Agenda

1. What is a data story? Data storytelling?
2. What is the purpose of data storytelling?
3. What are data stories made up of?
4. 5 steps to good data storytelling
5. The 7 types of data stories
6. A data story exemplar
7. The good, the bad, and the ugly

What is a data story? Data storytelling?

- A *data story* is exactly it; it's a story about data that connects with people on all levels – practical, personal, and emotional.
  - Some other terms to describe data stories include: quantitative stories and statistical narratives.

- *Data storytelling* is the process of transforming/translating data analyses into common and understandable terms or information in order to affect a particular decision or action.

What is the purpose of data storytelling?

Being able to tell stories and visualize data with it is an essential process of turning it from a pile of numbers into useable information that can be used to drive better decision making.

What are data stories made up of?

- Data stories feature two types of data:
  - Quantitative
  - Categorical
- Quantitative values measure things.
  - e.g. Number of library reference transactions
- Categories divide information into useful groups and the items that make up each category identify the things that are measured.
  - e.g. Type of institution (i.e. public library, academic library, hospital library etc.)

Good data storytelling vs. bad storytelling?

5 steps for good data storytelling:

1. Know your audience
2. Choose an effective visual display for your data story
3. Properly arrange the visuals
4. Check for design consistency based on graphic design best practices
5. Tell a story

Know your audience

Who is your target audience? Narrow your audience down; sometimes, this means creating different communications for different audiences.

- e.g. Audience is a group of librarians versus audience is a group of medical students for a story about public library reference transactions

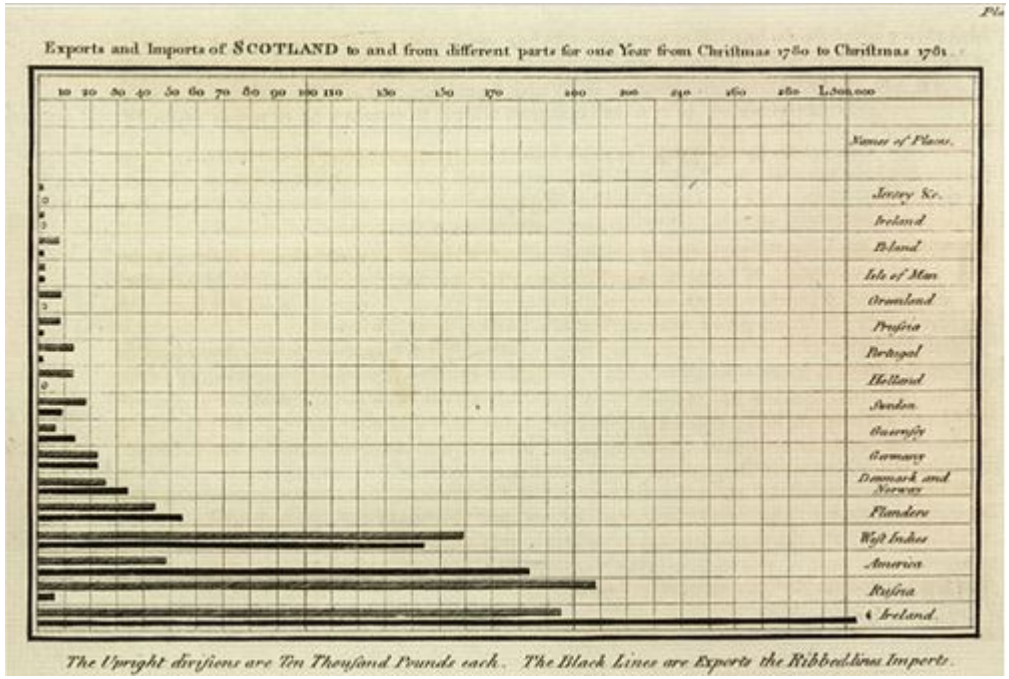
Choose an effective visual display for your data story

Before we choose a visual display, lets look at what communication platform we are going to use to convey our story:

- Live presentation
- Written document or email

Tables and graphs are the two fundamental vehicles for presenting quantitative information.

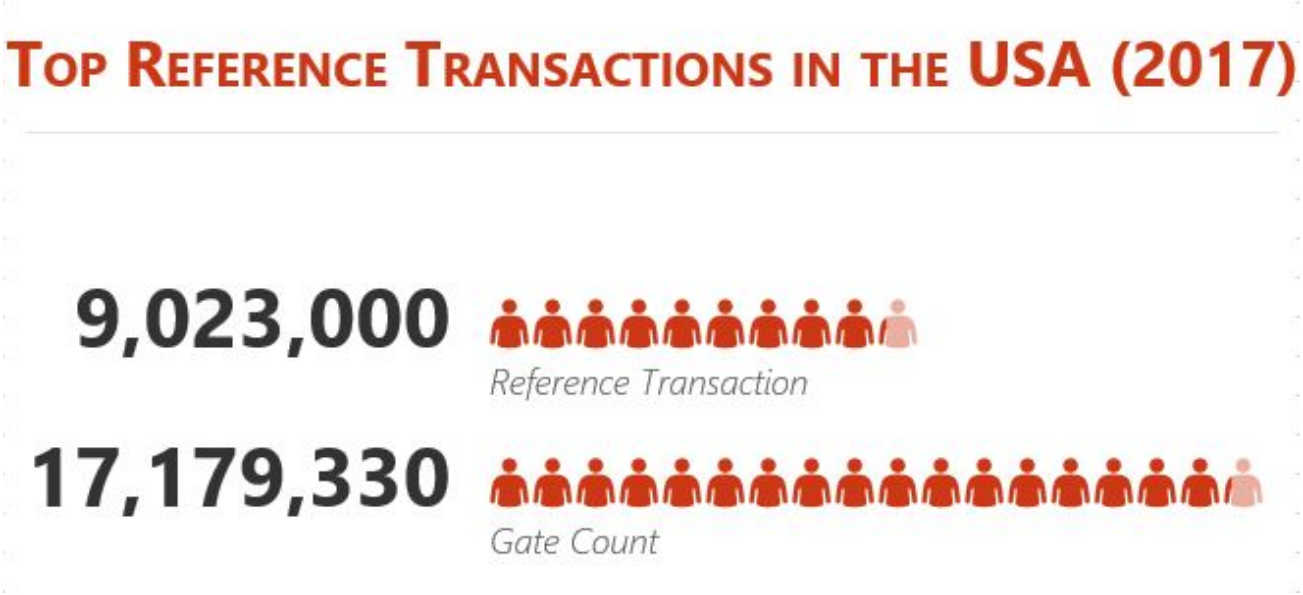
Historically, it wasn't until the 18th century that the use of graphs to present numbers became popular. William Playfair, a Scottish social scientist, invented the bar graph.



Seven common and credible types of visual display

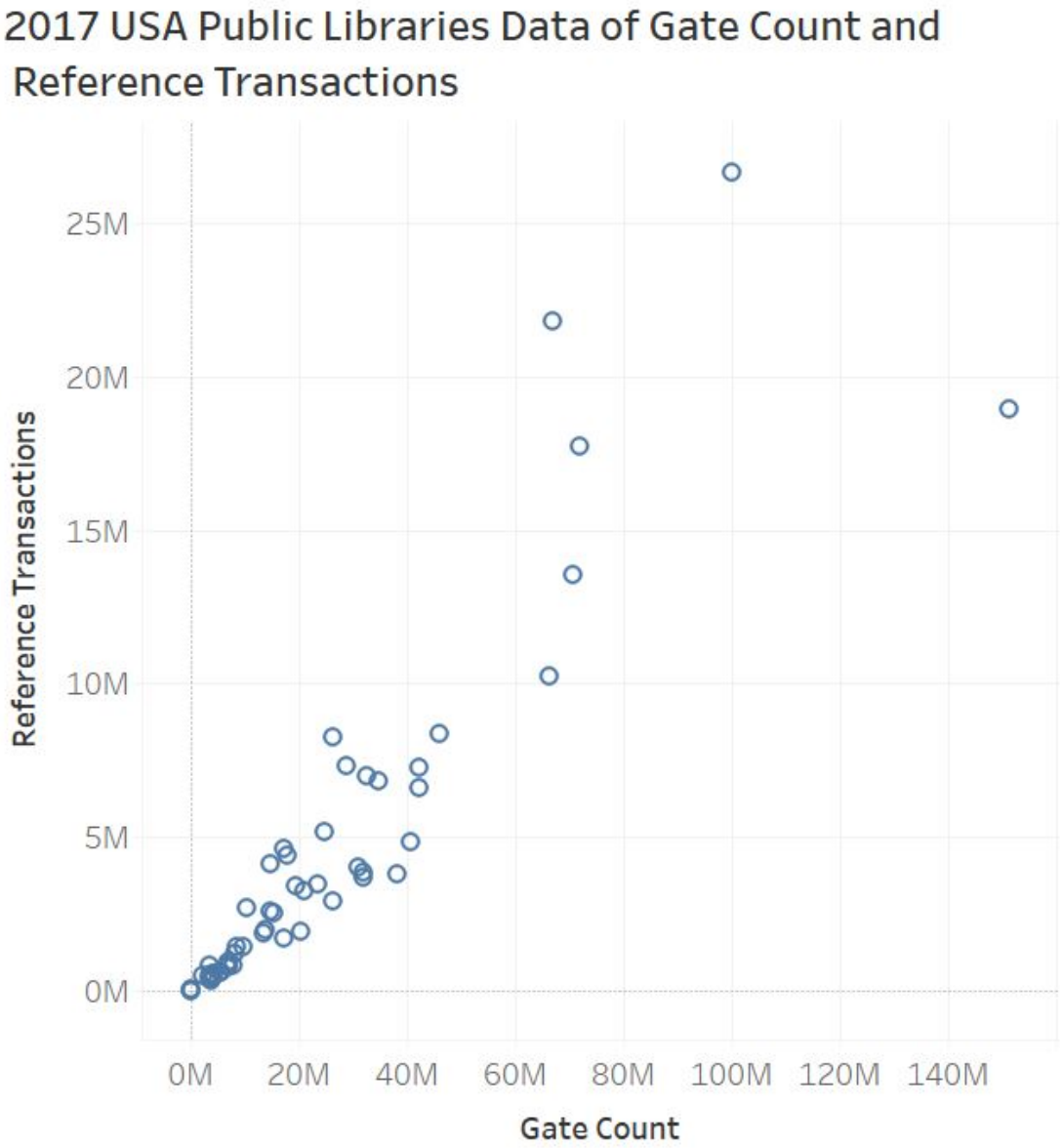
Simple Text

When you have just a number or two to share, simple text
is a great way to communicate



Scatterplot

Useful for showing relationship between two things because they allow you to encode data on a horizontal x-axis and vertical y-axis to see whether and what relationships exist



Tables

Useful for individual lookups or comparisons and not recommended for use in live presentations

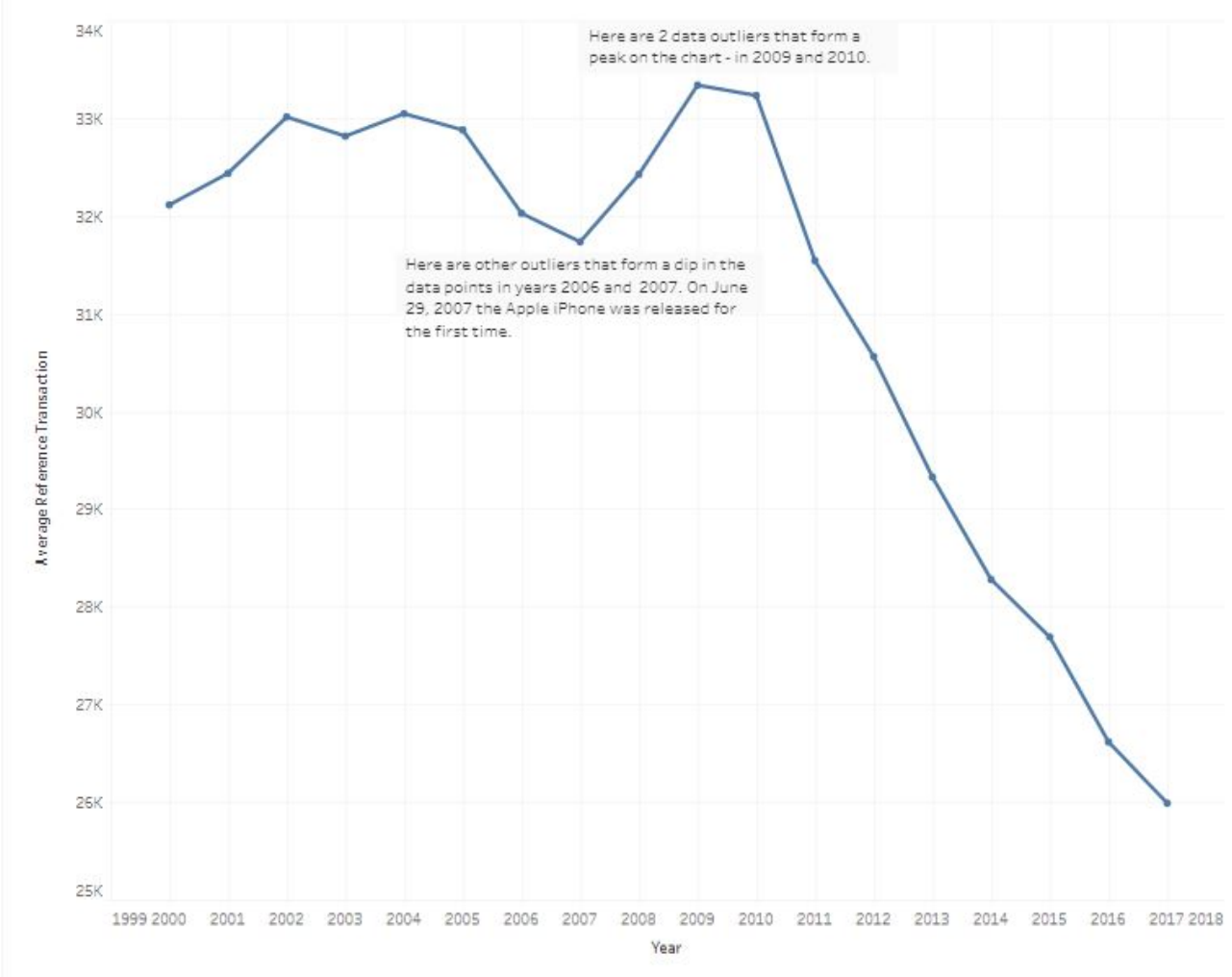
2017 USA Public Librares:

Top 10 States with most reference transactions

State	Gate Count	Reference Transactions
NY	100,012,193	26,644,064
FL	66,753,349	21,779,507
CA	151,058,515	18,953,587
OH	71,895,854	17,735,110
TX	70,521,174	13,528,641
IL	66,174,629	10,261,002
MI	46,052,561	8,384,564
MD	26,089,963	8,234,789
GA	28,816,233	7,328,518
PA	42,251,845	7,275,836

Line chart

Mostly used to plot continuous
data (i.e. data that is measurable and not categorical)



Heatmap

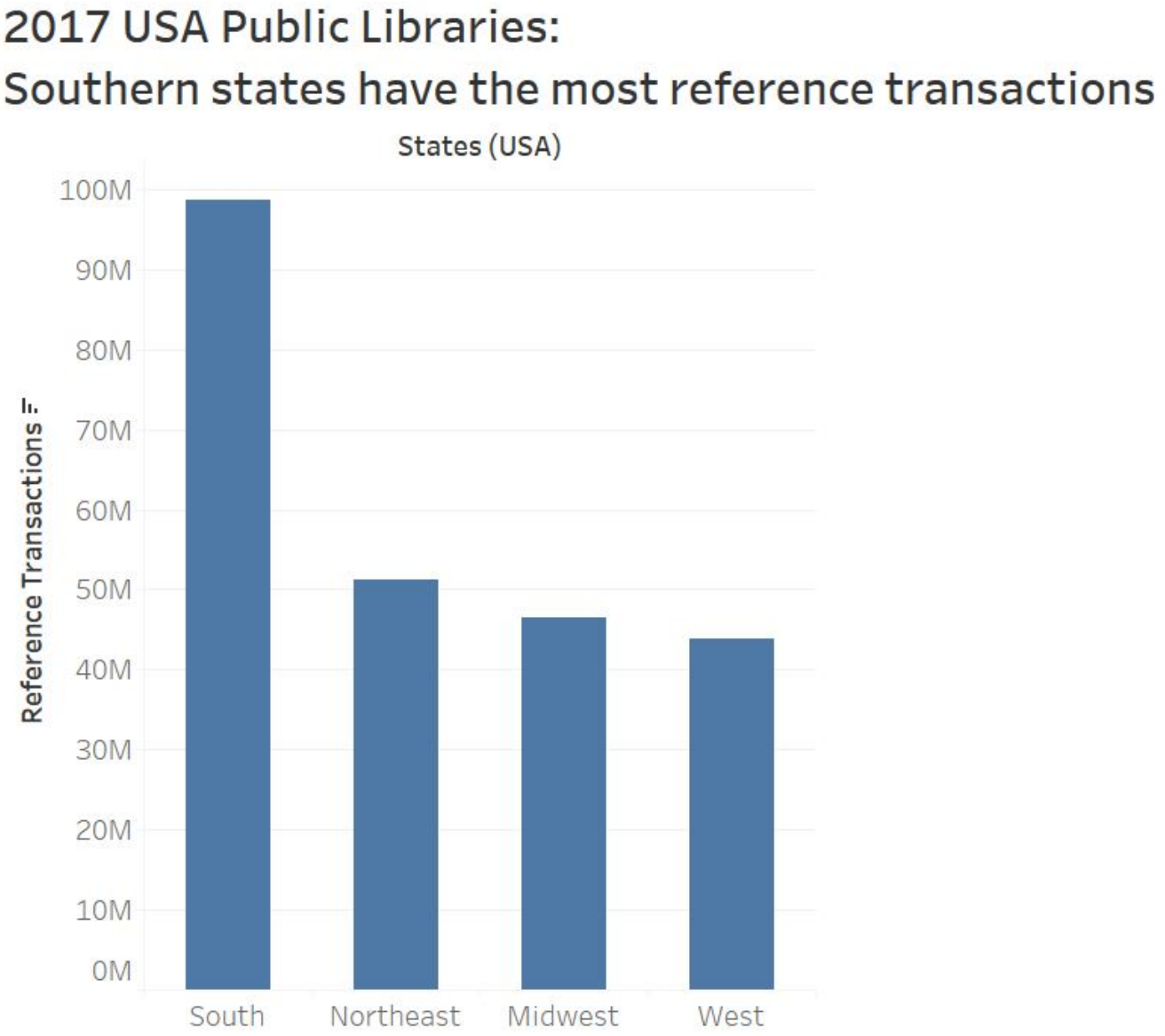
Used to visualize data in tabular format,
where in place of numbers you have colored cells that communicate the relative magnitude of the numbers

Heatmap of the top 10 States in the USA with the greatest reference transactions

State	Reference Transactions	Gate Count
New York	100,012,193	26,644,064
Florida	66,753,349	21,779,507
California	151,058,515	18,954,587
Ohio	71,895,854	17,735,110
Texas	70,521,174	13,528,641
Illinois	66,174,629	10,261,002
Michigan	46,052,561	8,384,564
Maryland	26,089,963	8,234,789
Georgia	28,816,233	7,328,518
Pennsylvania	42,251,845	7,275,836

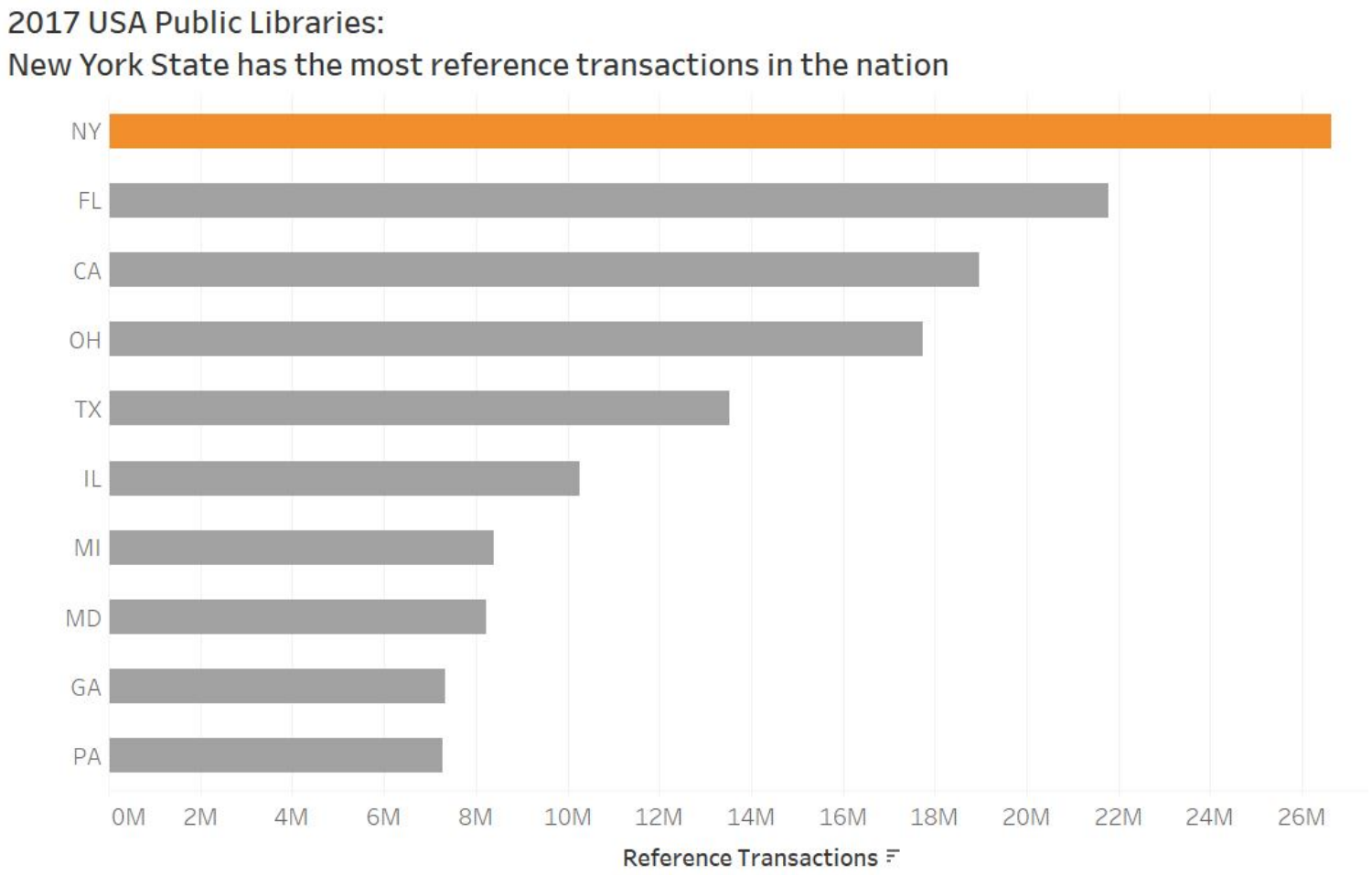
Vertical Bar Graph/Stacked Vertical Bar Graph

Bar charts are easy on the eyes because you are comparing the end points of bars so it is easy to see quickly which category is the biggest and smallest



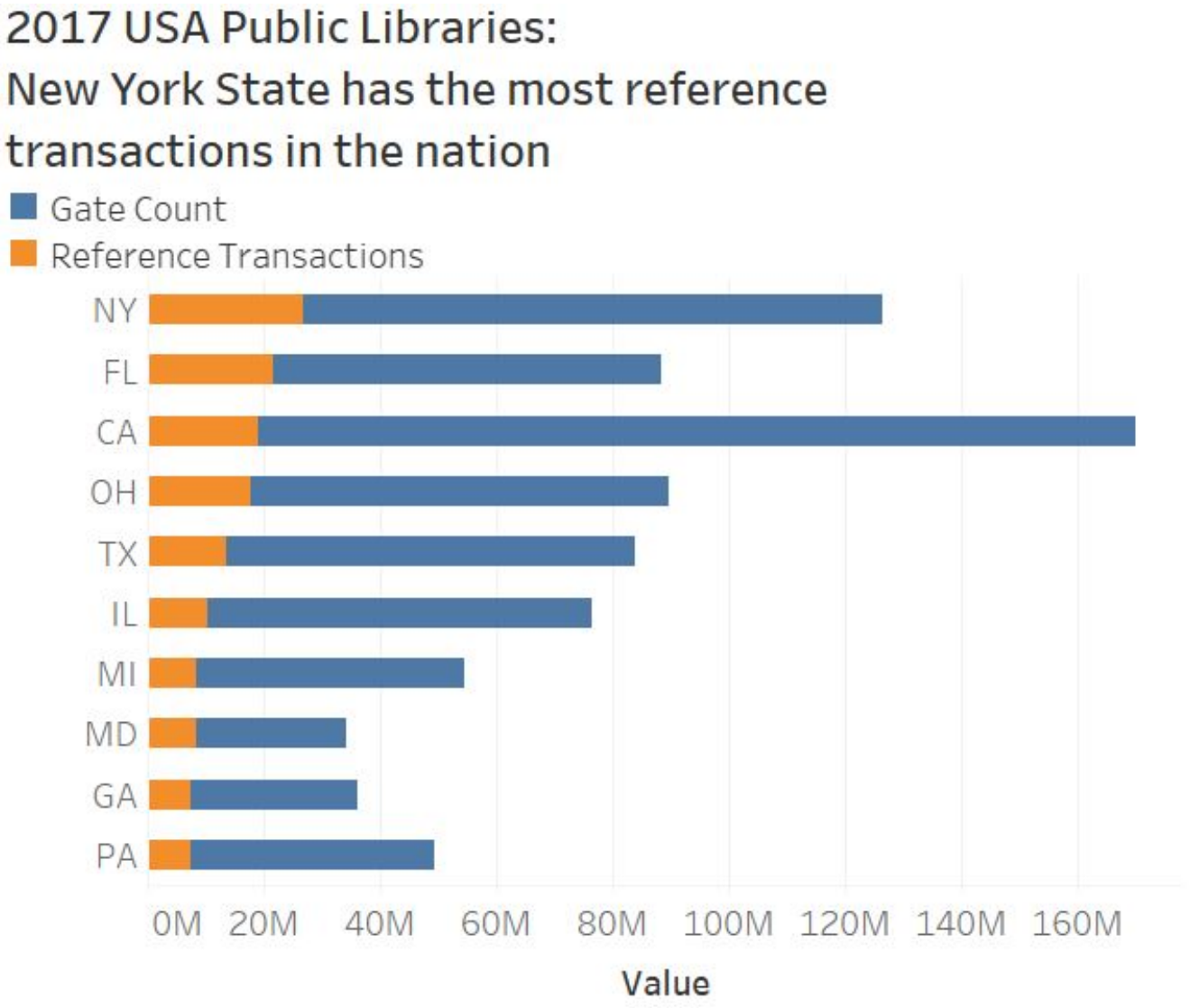
Horizontal Bar Graph

Similar to the Vertical Bar Graph, comparing the end points of bars so it is easy to see quickly which category is the biggest and smallest.



Stacked Horizontal Bar Graph

Similar to the Vertical Bar Graph, comparing the end points of bars so it is easy to see quickly which category is the biggest and smallest.



Properly arrange the visuals

You can order your data story in one of four ways:

1. Chronologically
   - Identify a problem
   - Gather data to understand the problem
   - Analyze the data
   - Emerge with a solution or finding
   - Recommend an action

2. Begin with the ending
   - Begin with the call to action – what your audience need to know or do.
   - Back up with the critical pieces of the story that support this call to action.

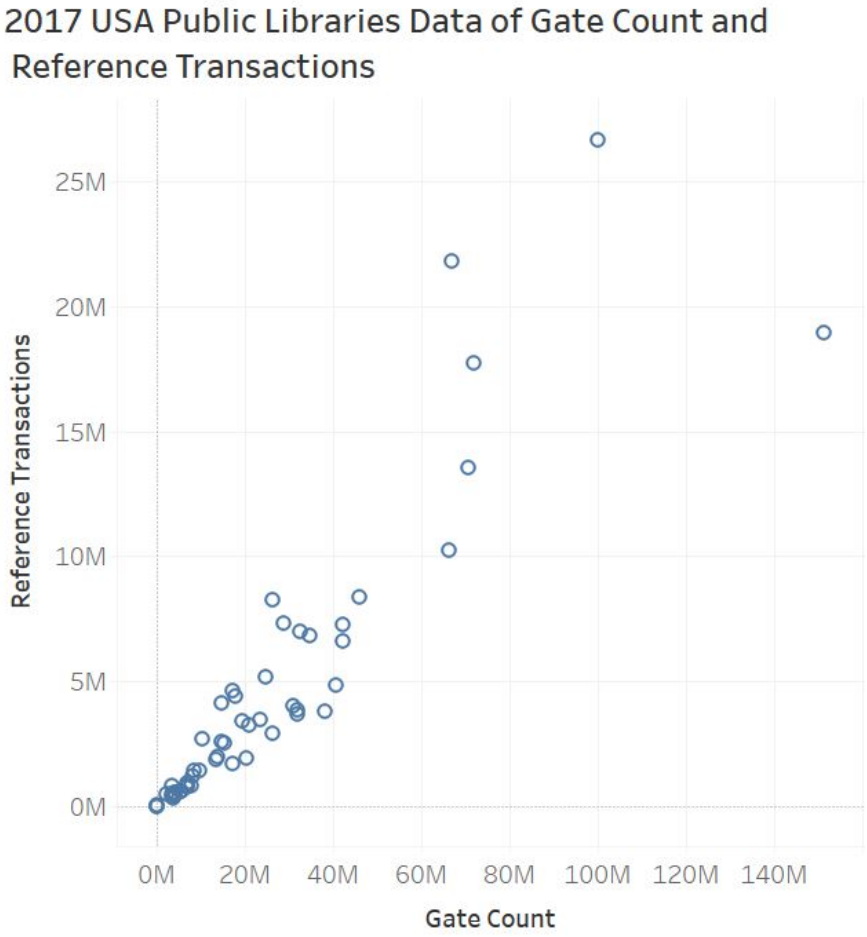
Check for design consistency based on graphic design best practices

Graphic design best practices

Primary Data Components, Part I

Points on a graph

When points overlap, have transparent interiors vs. filled ones so we can more readily see when points overlap.

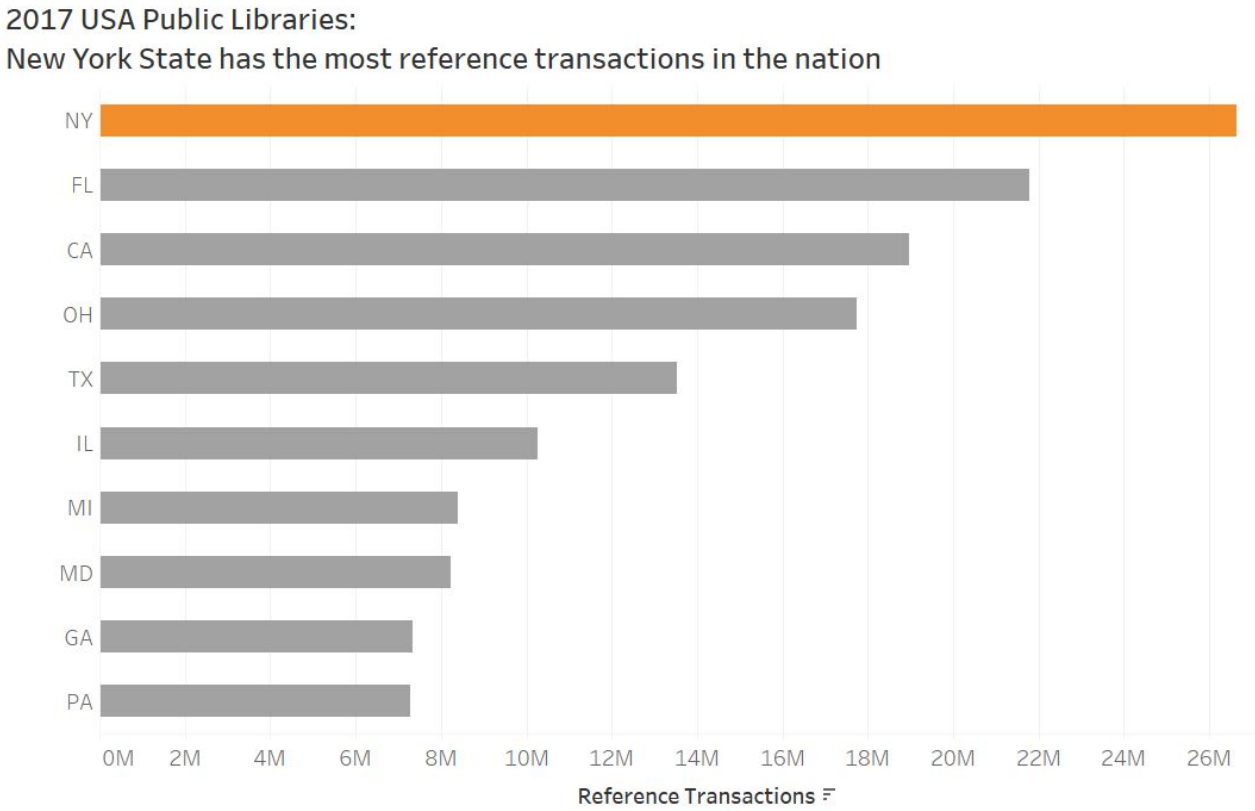


Data Source: PLS 2017

Primary Data Components, Part II

Bars on a graph

Ratio of approximately
1:1 for how closely bars should be placed to one another. Use fill colors for bars that are fairly balanced in intensity for data sets that are of equal importance; use fill colors that are more intense than others when you wish to highlight particular values above the others.

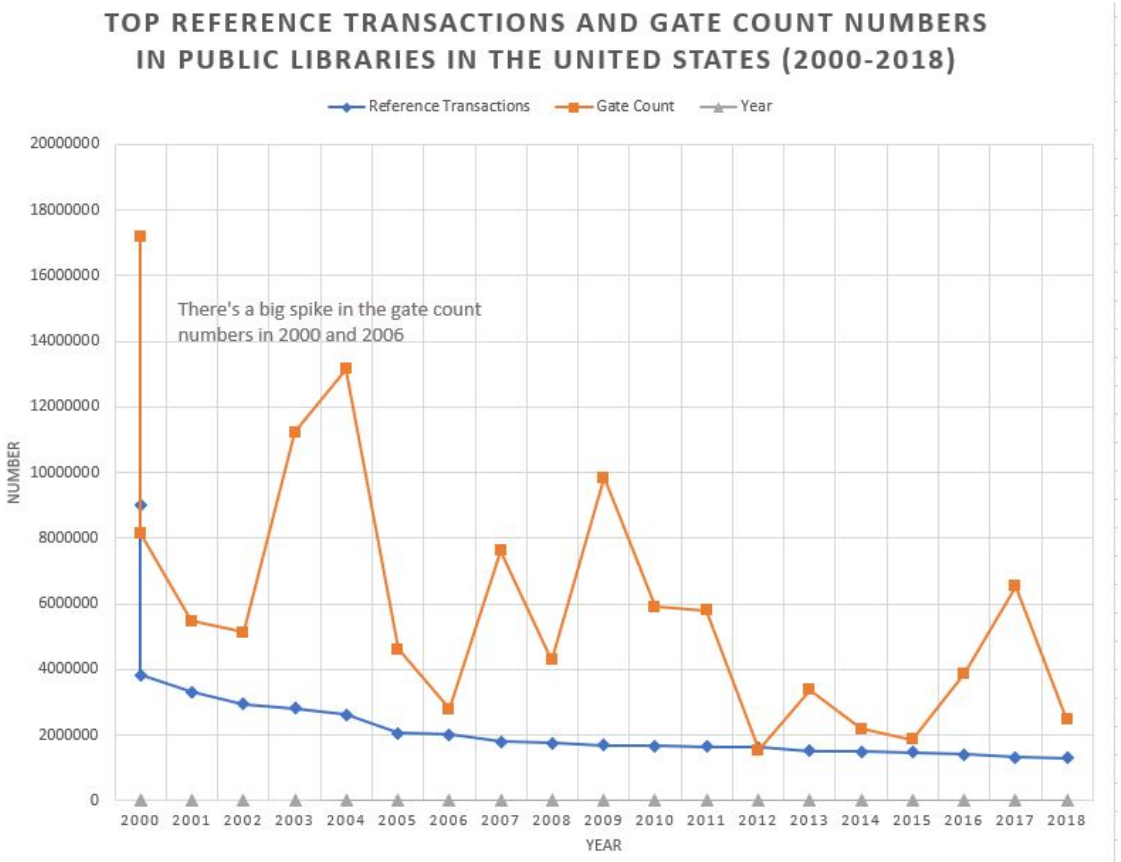


Data Source: PLS 2017

Primary Data Components, Part III

Lines on a graph

When a graph
contains multiple sets of values, each encoded as a line, you must take care to make them visually distinct – hue works more effectively than color intensities (i.e. gray shades).



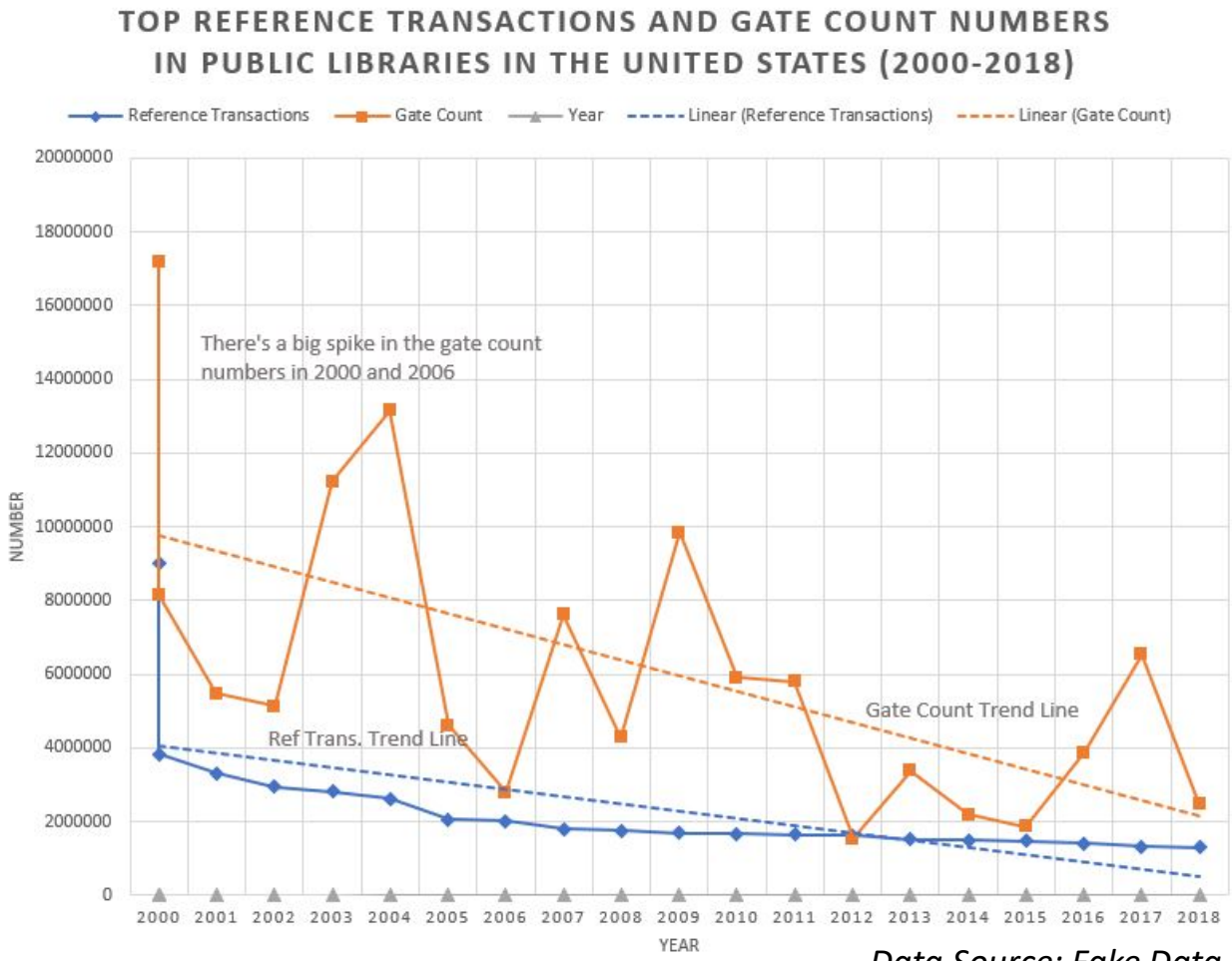
Data Source: Fake Data

Secondary Data Components, Part I

- **Trend lines** – trend line should be deemphasized relative to the line of actual values to make the latter stand out as more important (i.e. lighter or dashed).

- **Reference lines** – should look different from lines that encode the primary data (i.e. lighter or dashed).

- **Annotations** – placing annotations in the data region is useful to make them inconspicuous to prevent clutter; reduce the intensity of the text from black to gray to prevent distraction.



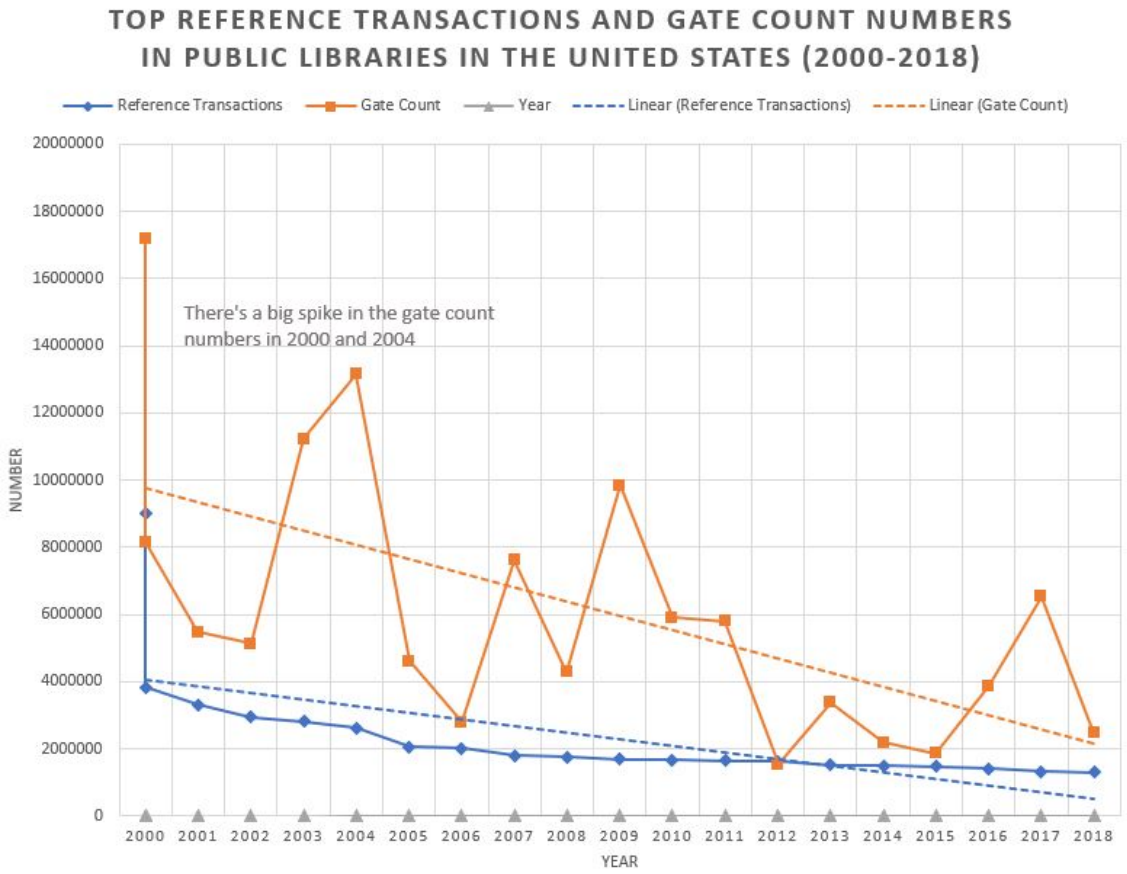
Data Source: Fake Data

Secondary Data Components, Part II

- **Legends** – legends could be replaced by labelling the lines directly and if they are used they should be made less prominent than the actual data and arranged horizontally below the title which prevents the graph from being wider than necessary.

- **Aspect ratio** – never the manipulate the ratio of the graph's width to its height (aspect ratio) to exaggerate or downplay the degree of change.

- **Data region** – white is normally the best background but there are times when other light colors such as light gray or yellow are useful.



Data Source: Fake Data

Tell a story

Like real narrative stories, movie scripts, or novels, a data story has three parts:

The beginning is known as the plot of the story

The beginning of any data story begins with setting the context:

- Setting – when and where does the data story take place
- Main Character – who is driving the action?
- Imbalance/Problem – what has changed and why is it necessary?
- Balance – what do you want to see happen?
- Solution/Desired Outcome – how will you bring about the changes?

The middle of the data story is known for having twists/conflict in it

This is key to grabbing and maintaining your audience's attention. Some ideas to help you build out your story and to convince your audience to buy into your suggested solution:

- Further develop the problem by covering relevant background information by giving external context or comparison points.
- Include data that demonstrate the problem.
- Articulate what will happen if no action is taken or no change is made.
- Illustrate the benefits of your recommended solution.

The end of a data story is ultimately the call to action from your audience.

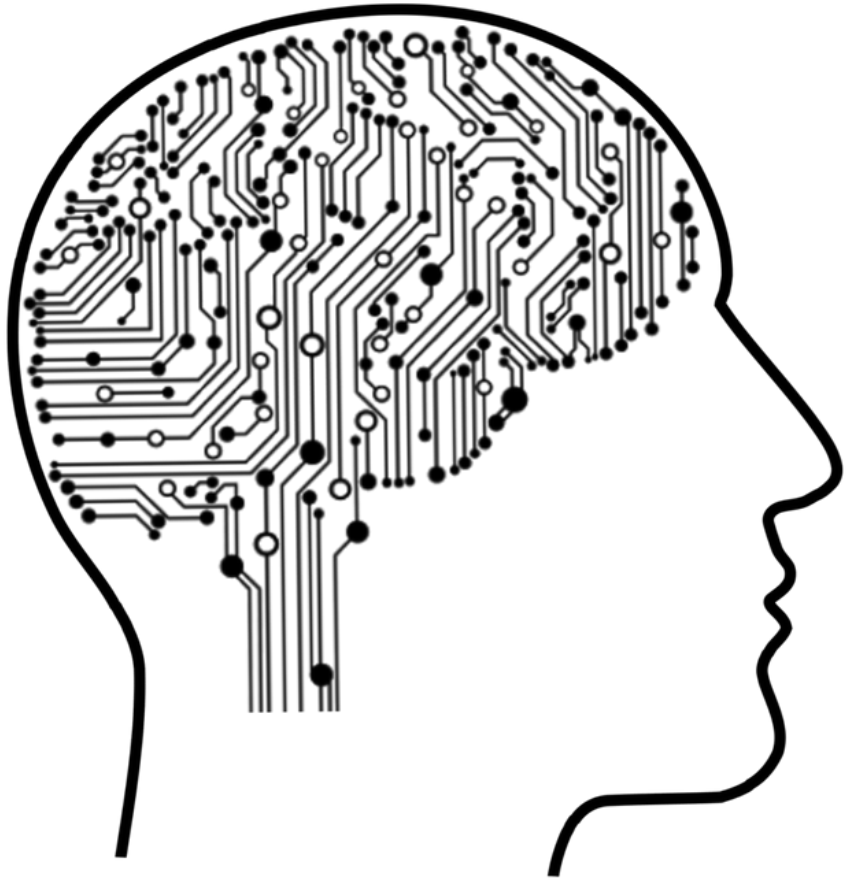
Every data story must have an end in place; generally, data stories end with a call to action.

- To wrap up, you can think about recapping the problem and the resulting need for action.
- Make it totally clear to your audience what you want them to do with the new knowledge that you've shared with them.

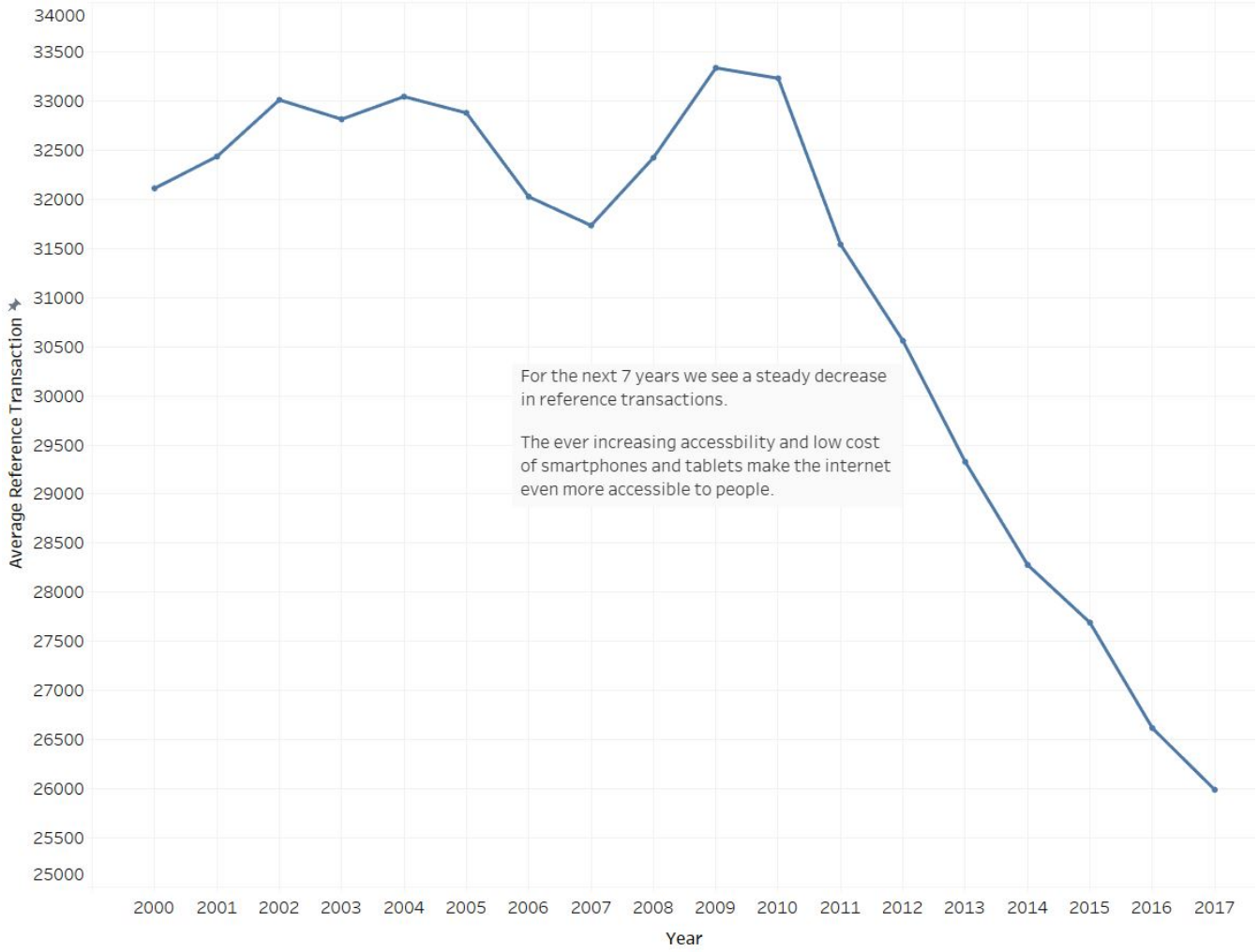
7 Types of Data Stories

There are 7 types of data stories:

1. Change over time
2. Drill down
3. Zoom out
4. Contrast
5. Intersections
6. Factors
7. Outliers

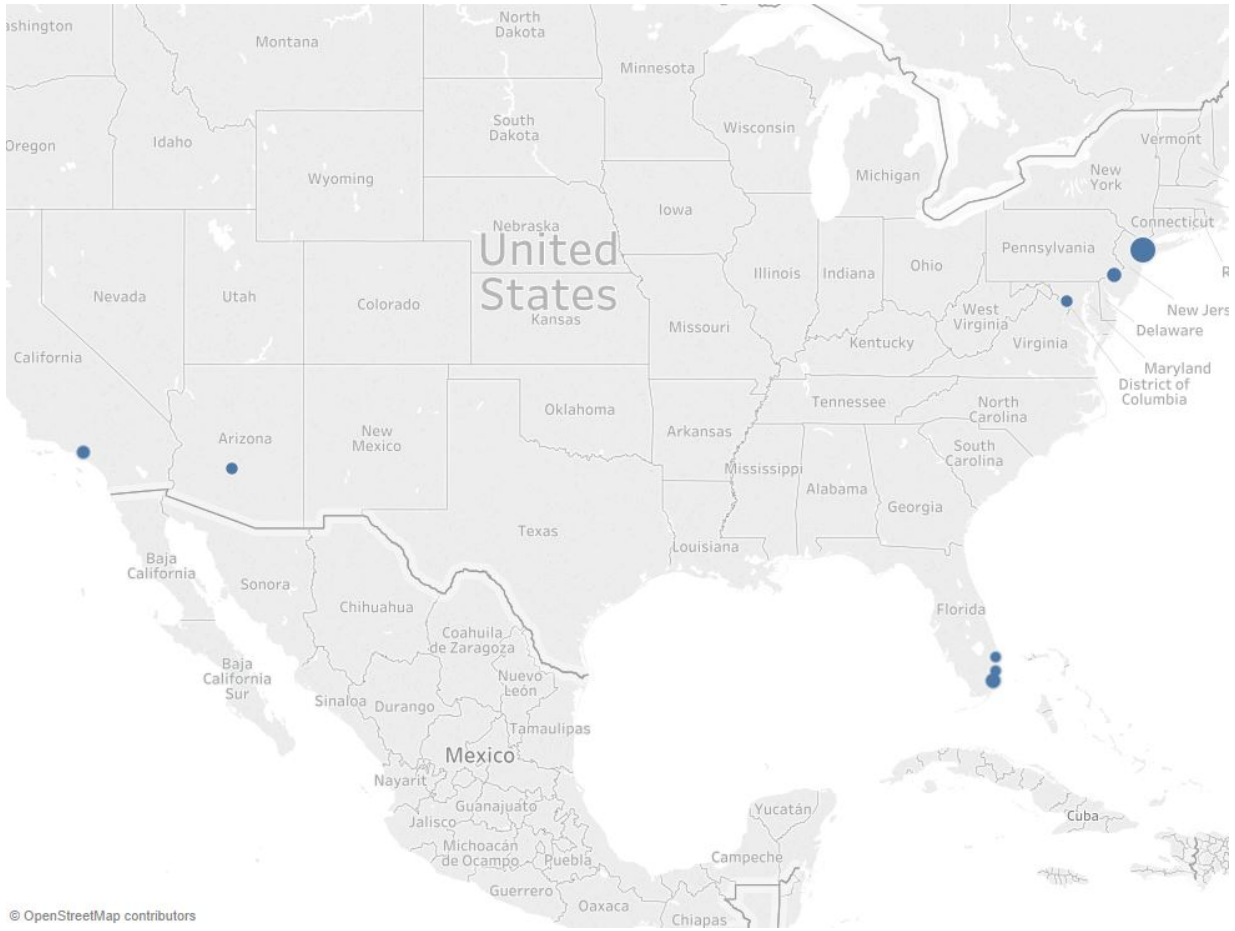


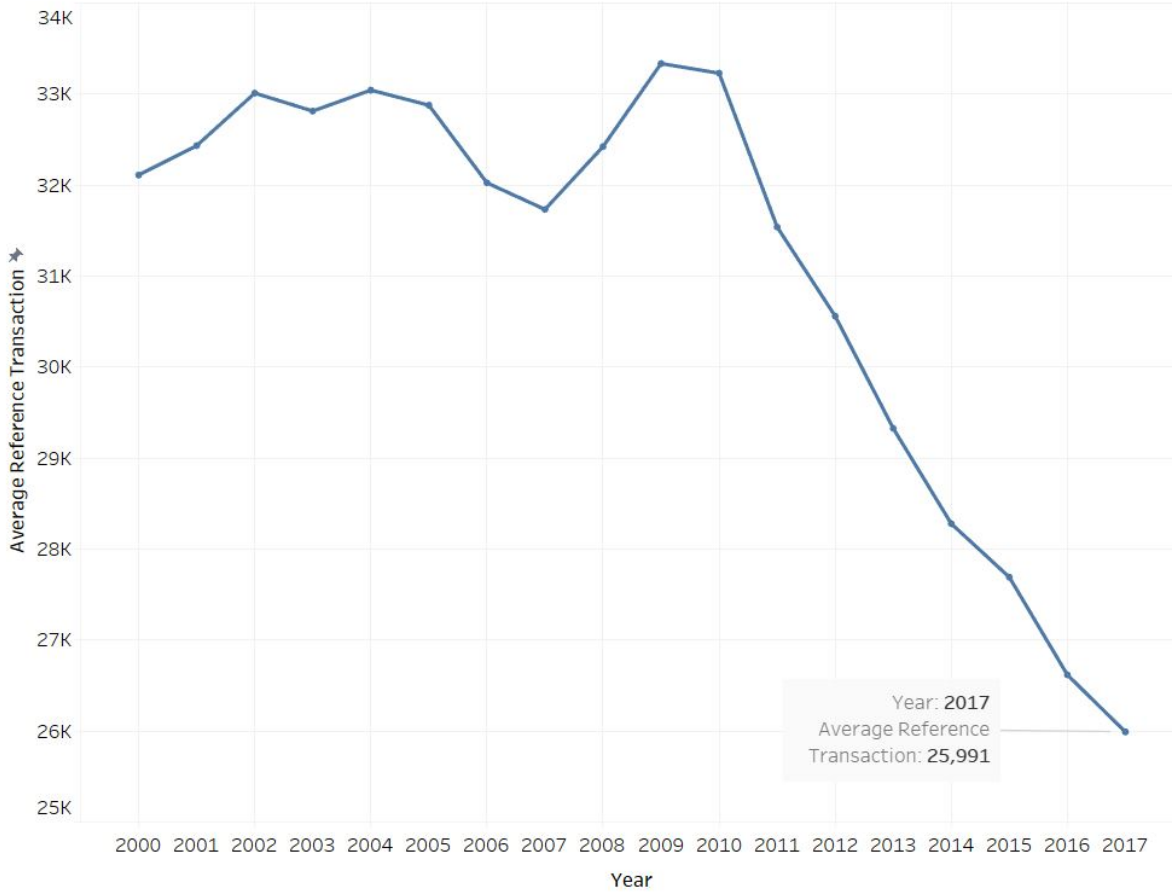
Story #1 - Change over time
National Public Library Average Reference Transactions (2000-2017)



Data Source: PLS 2000 - 2017

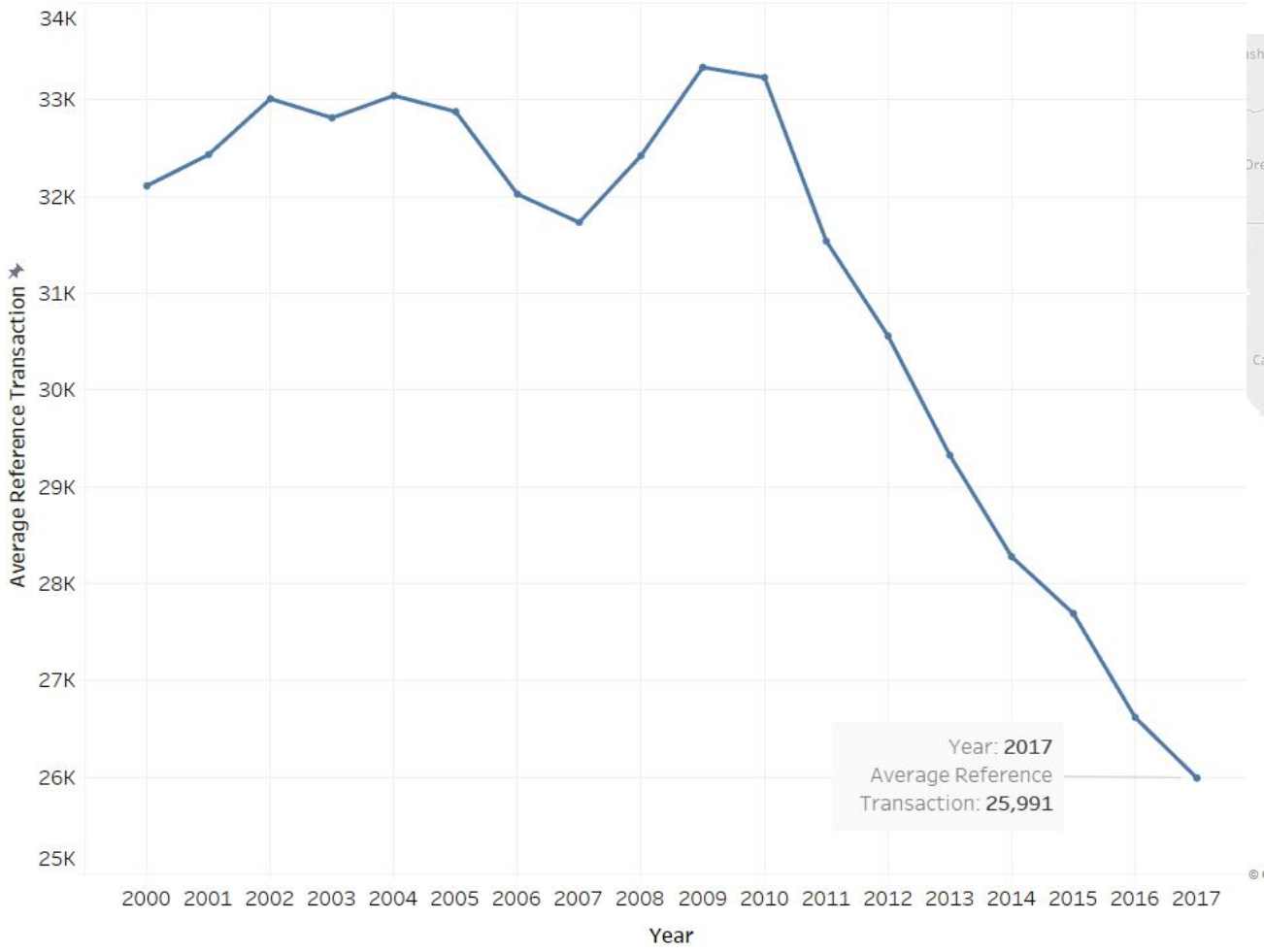
Story #2 – Zoom out
Top 10 Reference Transactions in the United States for 2017

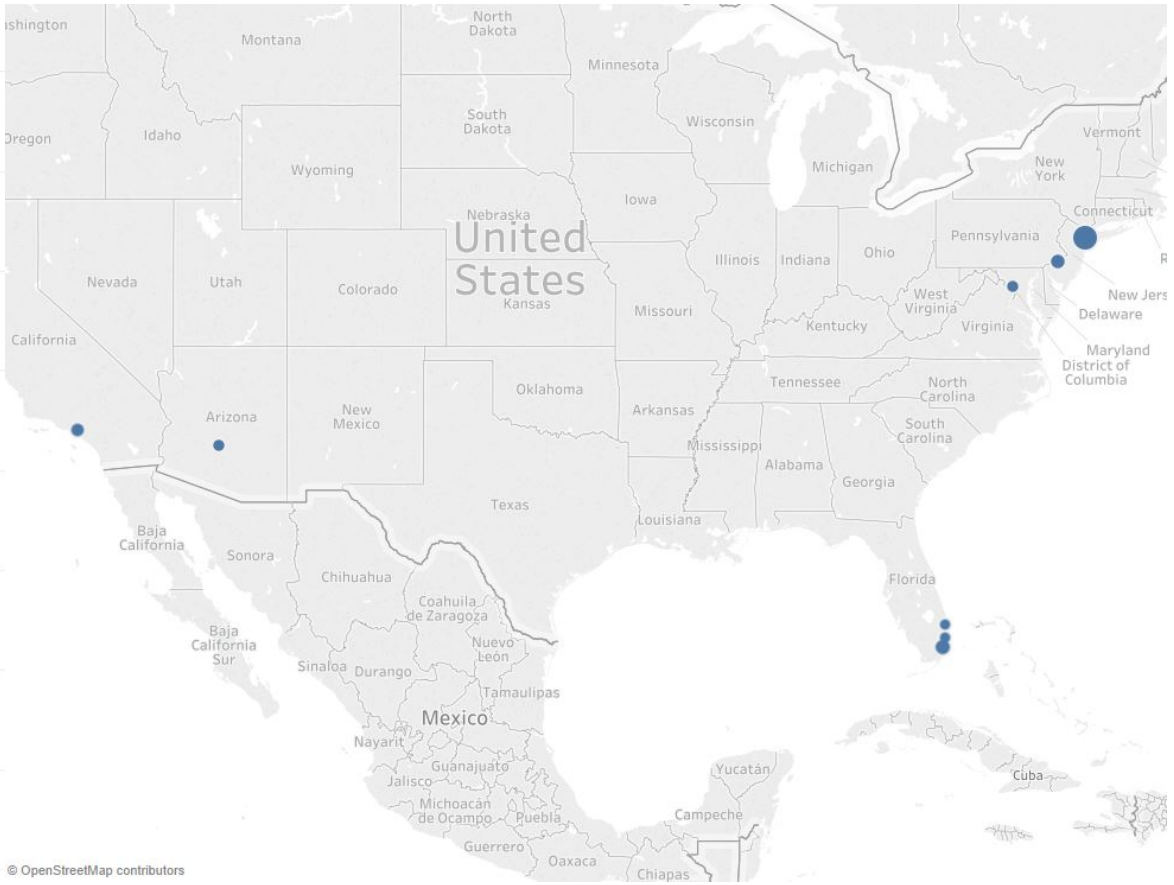




Data Source: PLS 2000 - 2017

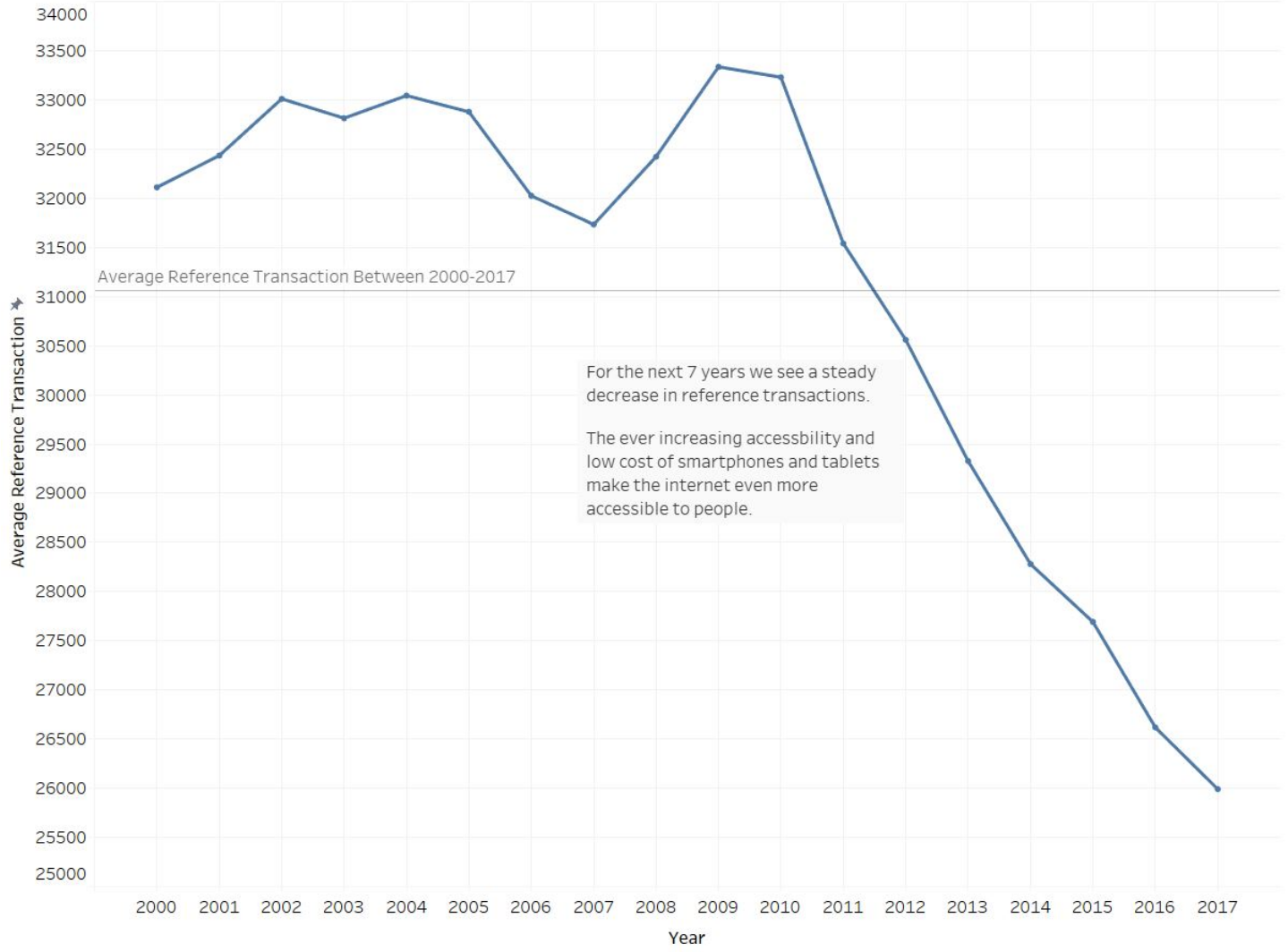
Story #3 – Zoom in/Drill Down
Top 10 Reference Transactions in the United States for 2017





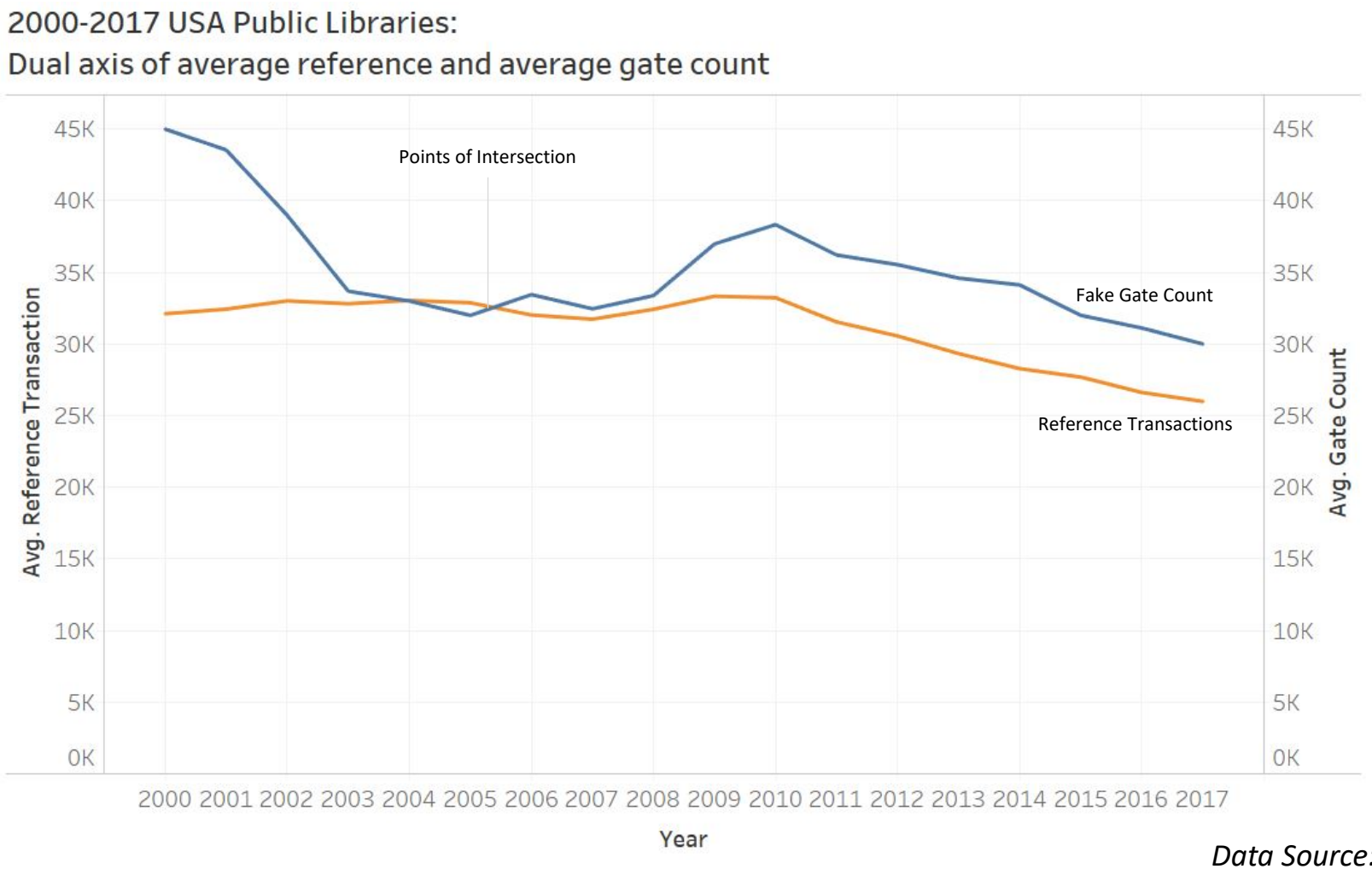
Data Source: PLS 2000 - 2017

Story #4 – Contrast
National Public Library Average Reference Transactions (2000-2017)



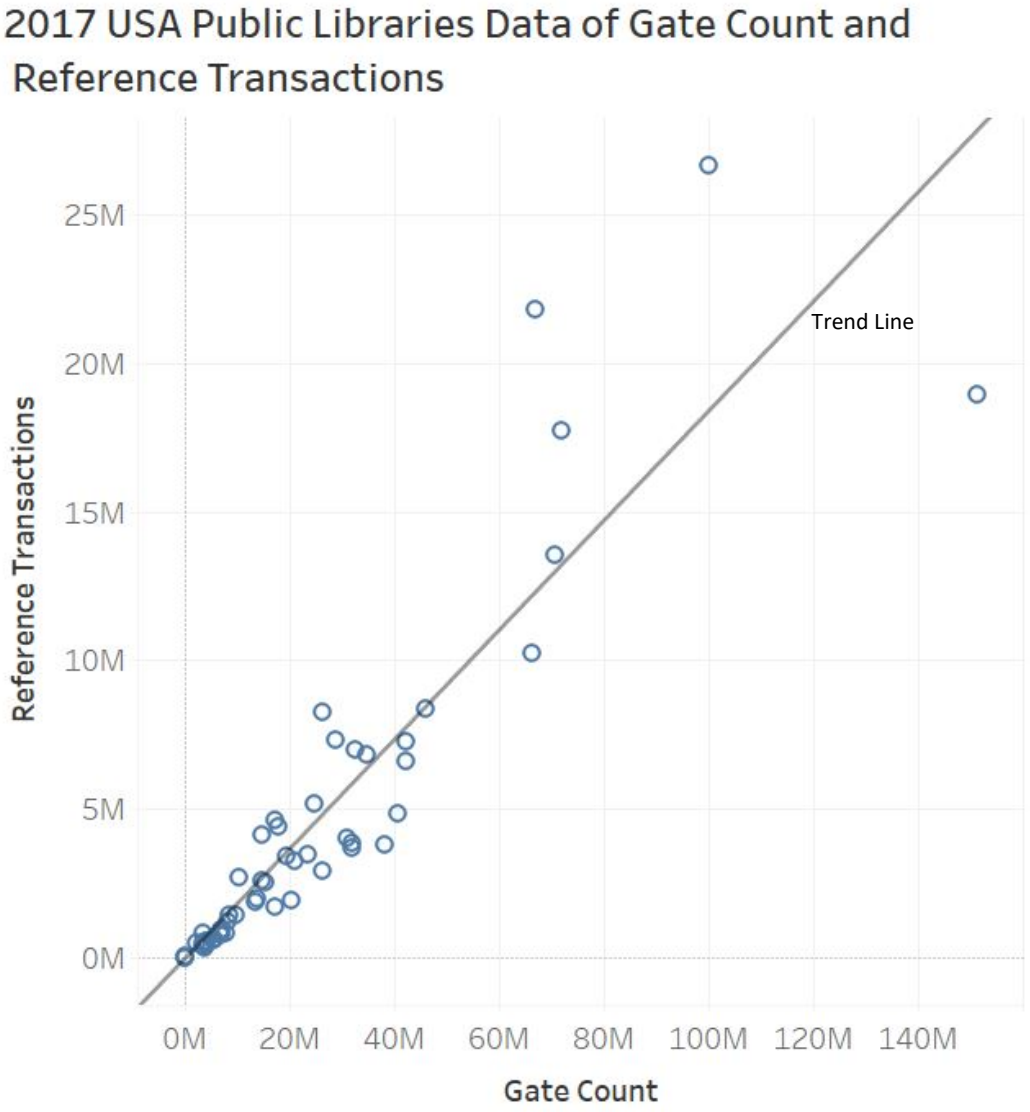
Data Source: PLS 2000 - 2017

Story #5 – Intersections



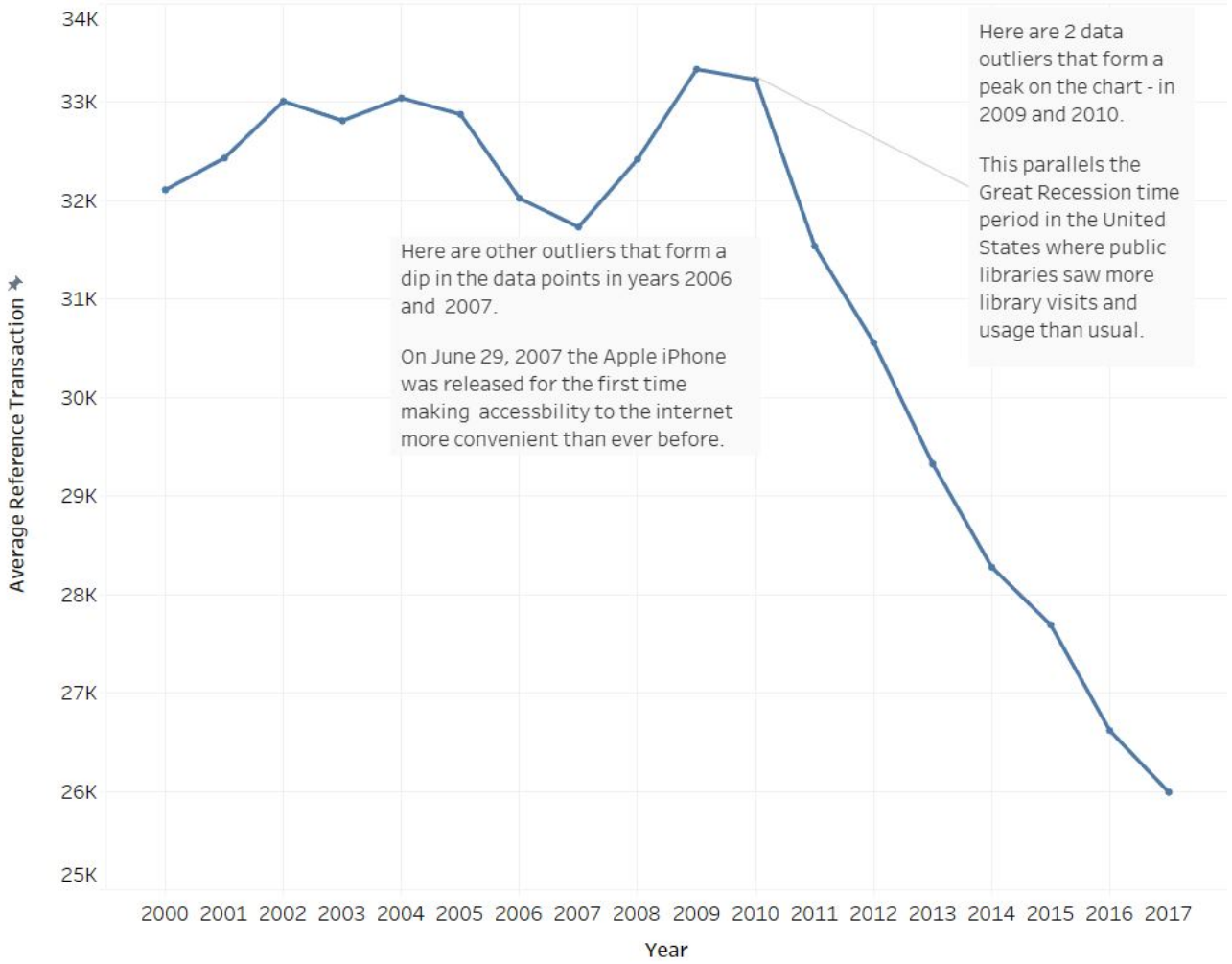
Data Source: Fake Data & PLS 2000-2017

Story #6 – Factors



Data Source: PLS 2017

Story #7 - Outliers
National Public Library Average Number of Reference Transactions (2000-2017)



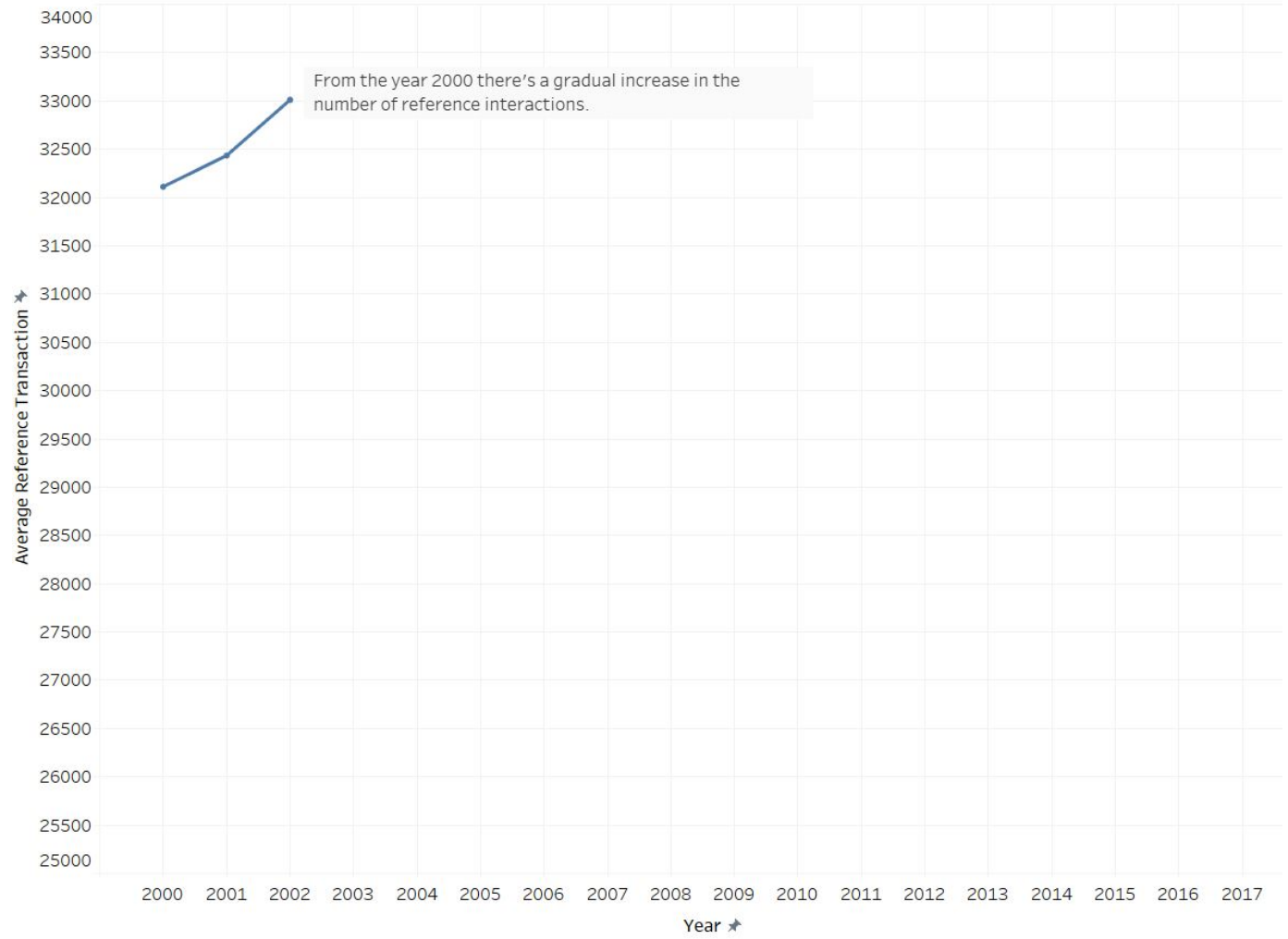
Data Source: PLS 2000 - 2017

Data Storytelling Example: Data Context

Here's some background information on the data that I'll use to demonstrate a data storytelling example:

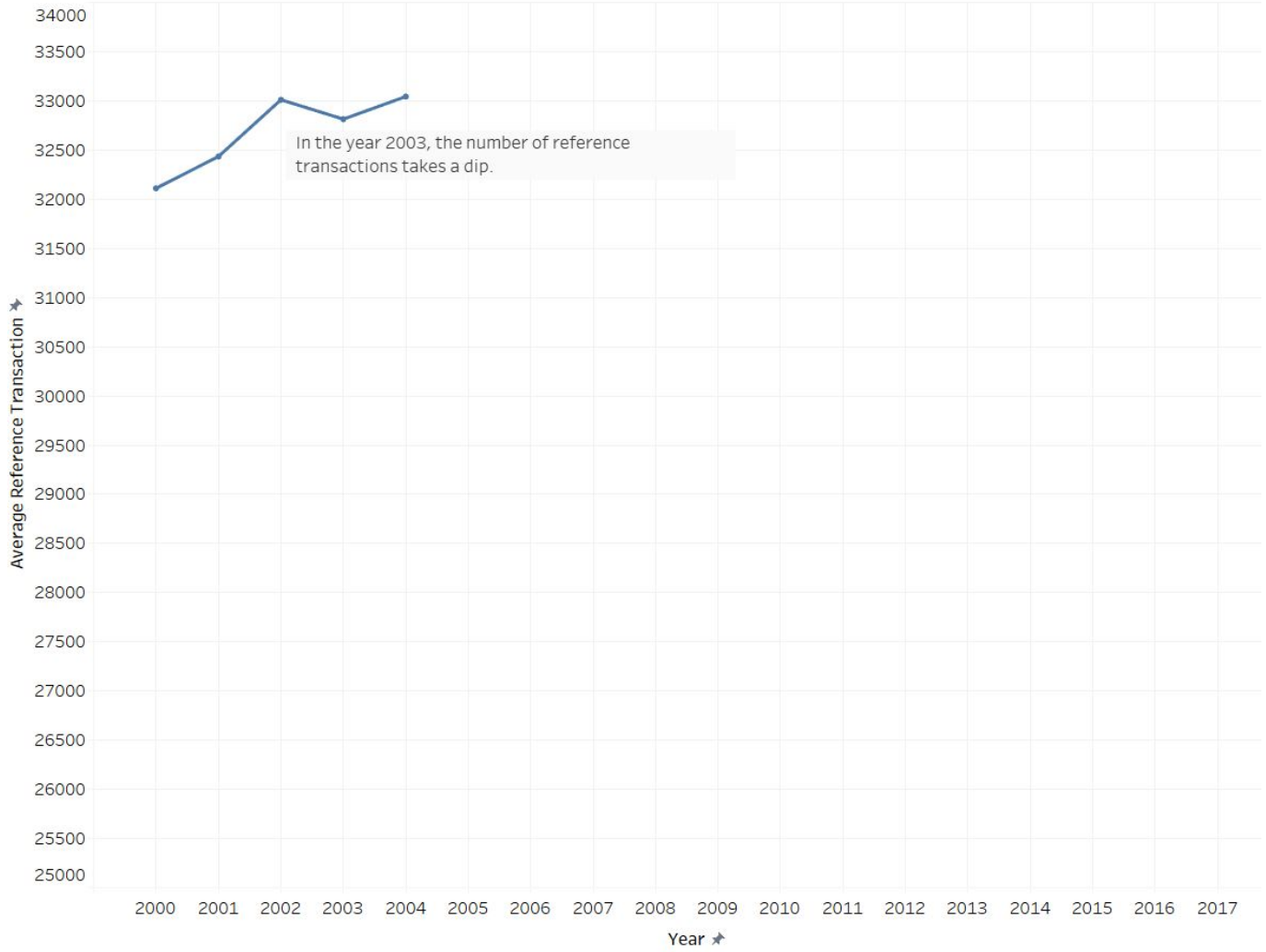
- **Data source**
  - the national Public Libraries Survey (PLS) is available on the Institute of Museum and Library Services website. Deidentified data from ARL on reference transactions and on occasion fake data, when used, will be indicated.
- **Purpose**
  - the PLS provides statistics on the status of public libraries in the United States.
- **Coverage**
  - the data from the PLS are collected from about 9,000 public libraries with about 17,000 individual public library outlets (i.e. main libraries, branches etc) in the 50 states.
- **Content**
  - PLS data includes information about a wide variety of things ranging from number of reference transactions to gate counts.
- **Frequency**
  - PLS collected annually since 1988. For our purposes, we'll look at data from the millennium onwards to 2017.

Story #1 – Change over time, Part 1
National Public Library Average Reference Transactions (2000-2017)



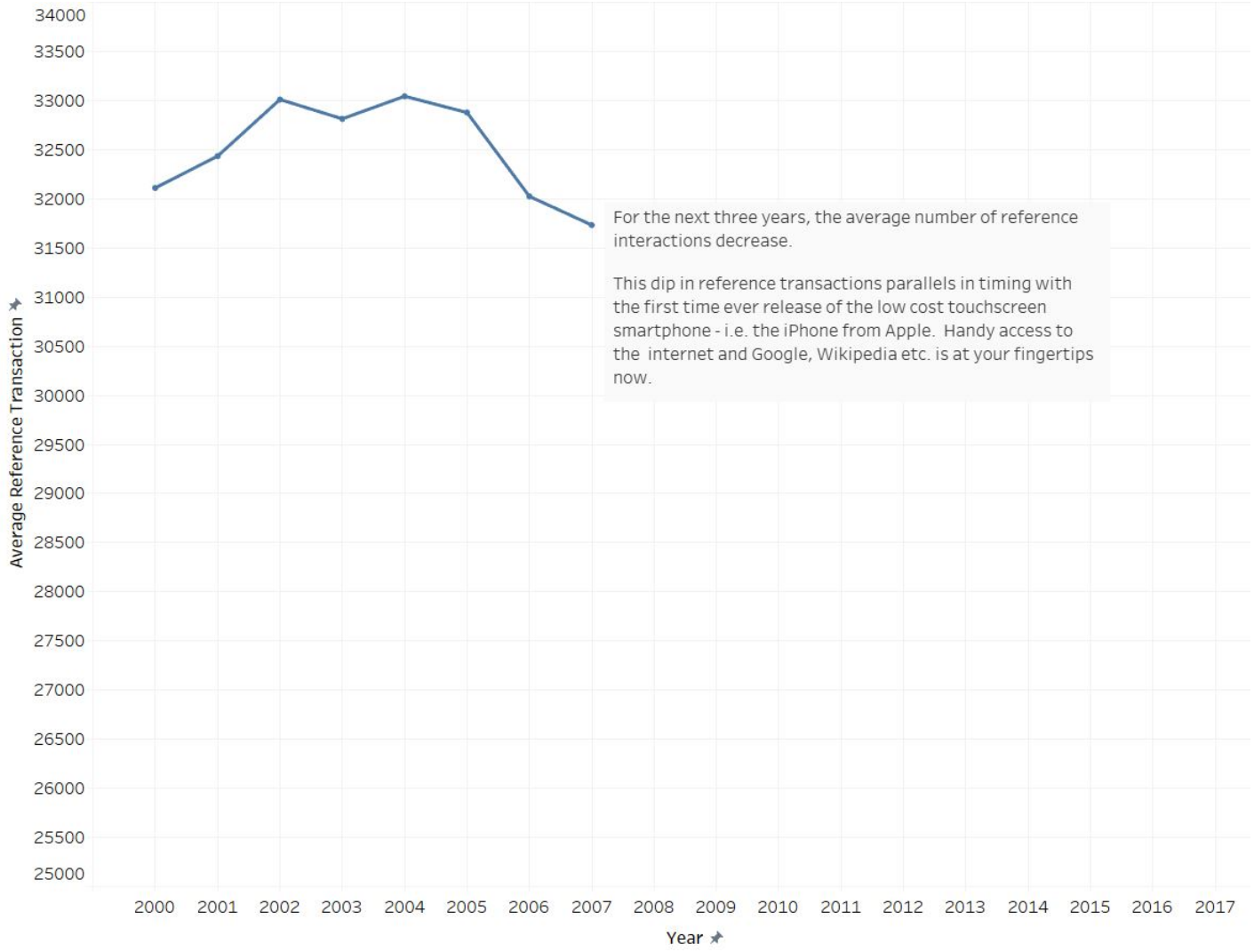
Data Source: PLS 2000 - 2017

Story #1 – Change over time, Part 2
National Public Library Average Reference Transactions (2000-2017)



Data Source: PLS 2000 - 2017

Story #1 – Change over time, Part 3
National Public Library Average Reference Transactions (2000-2017)

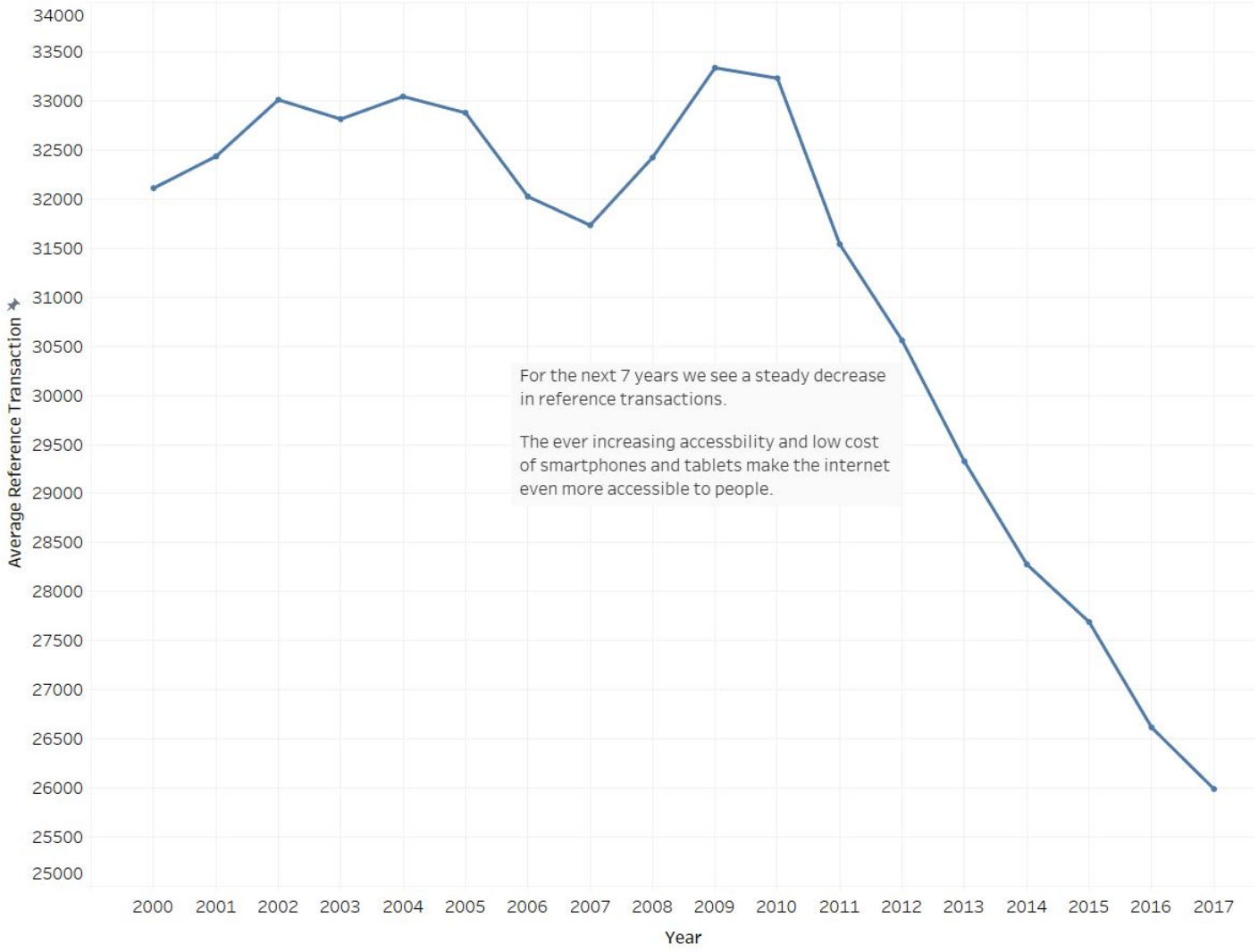


Data Source: PLS 2000 - 2017

Story #1 – Change over time, Part 4
National Public Library Average Reference Transactions (2000-2017)

Data Source: PLS 2000 - 2017

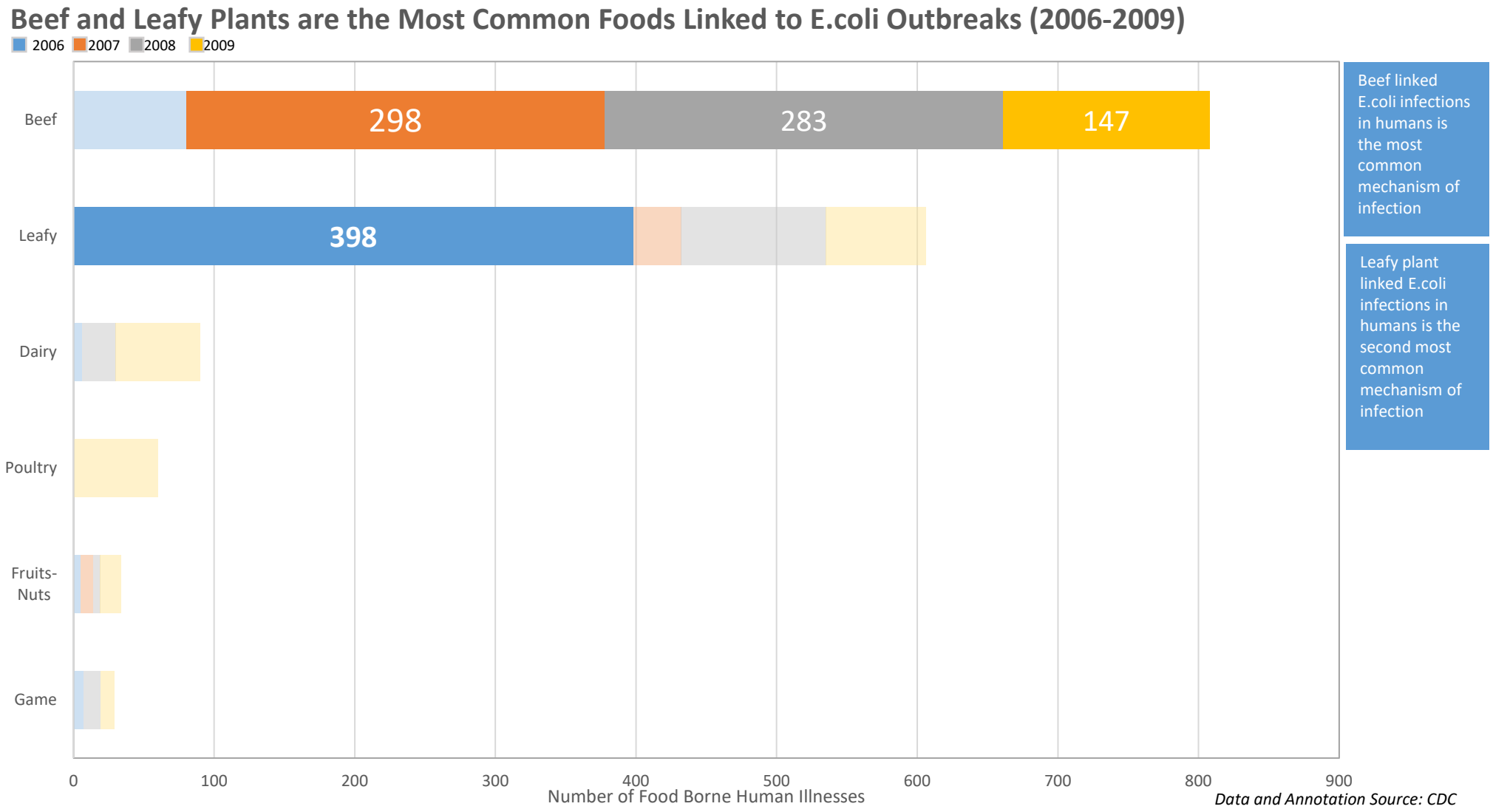
Story #1 – Change over time, Part 5
National Public Library Average Reference Transactions (2000-2017)



Data Source: PLS 2000 - 2017

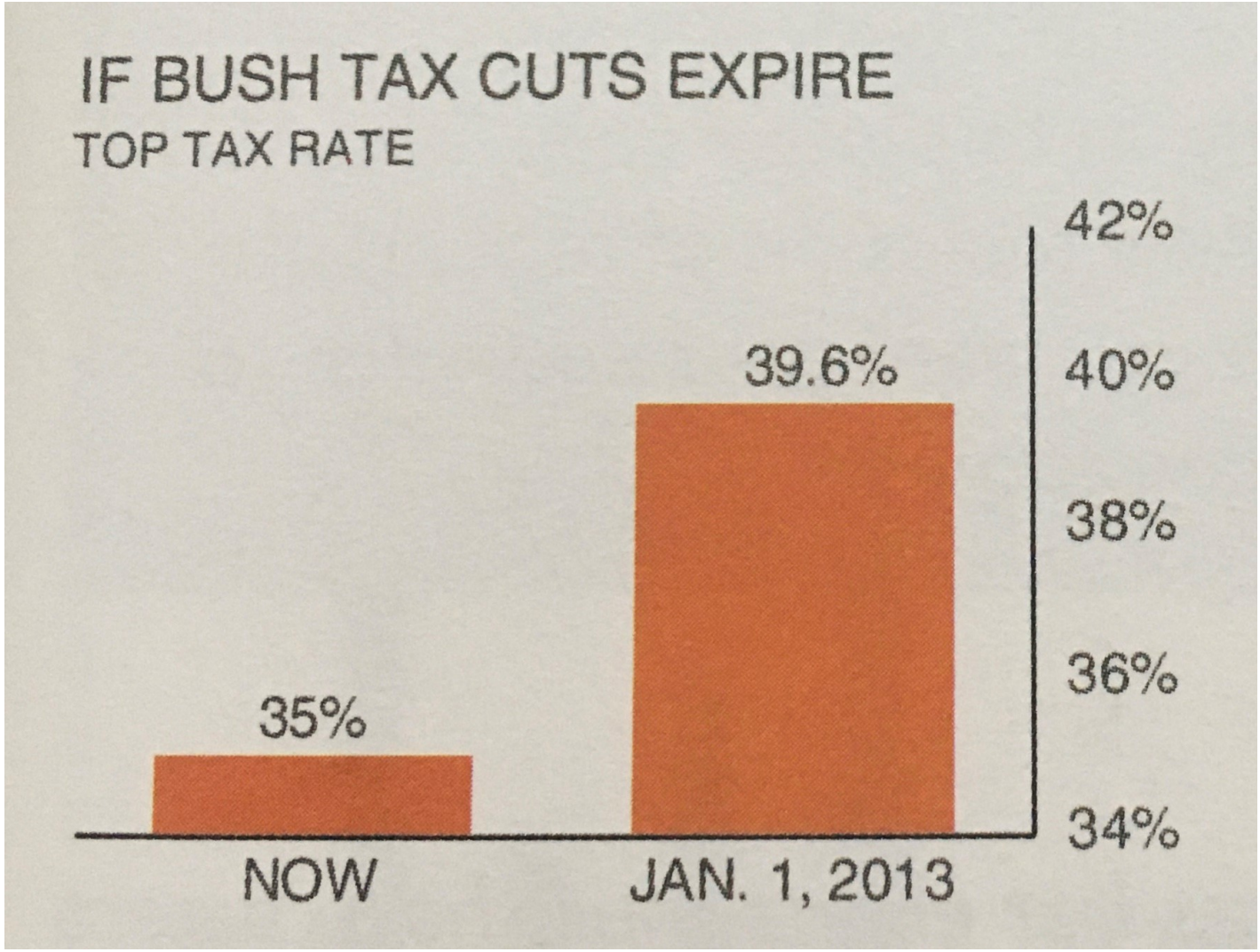
Group Discussion

Good Example of a Data Visualization



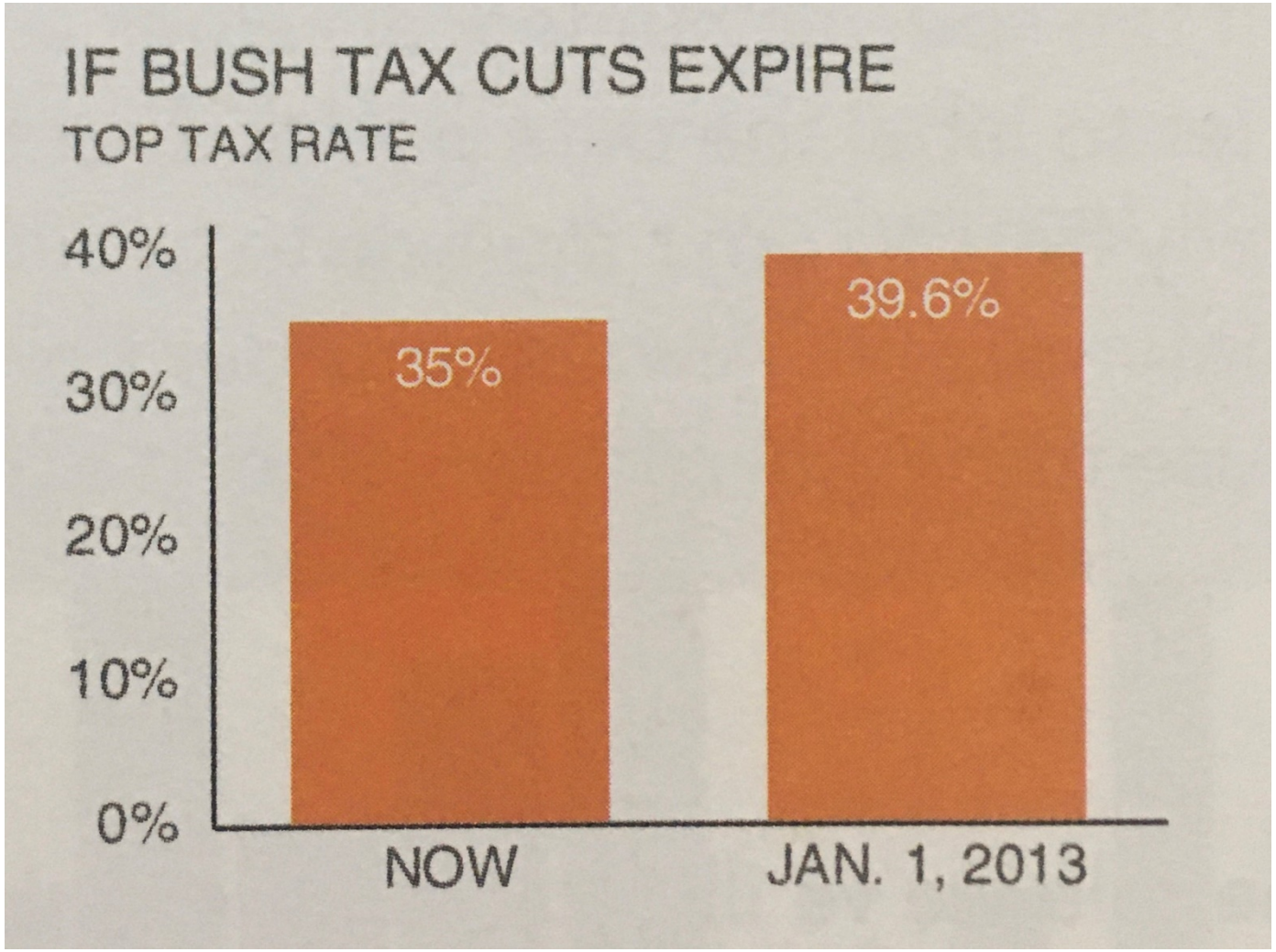
Data Source: CDC

Bad Example of a Data Visualization



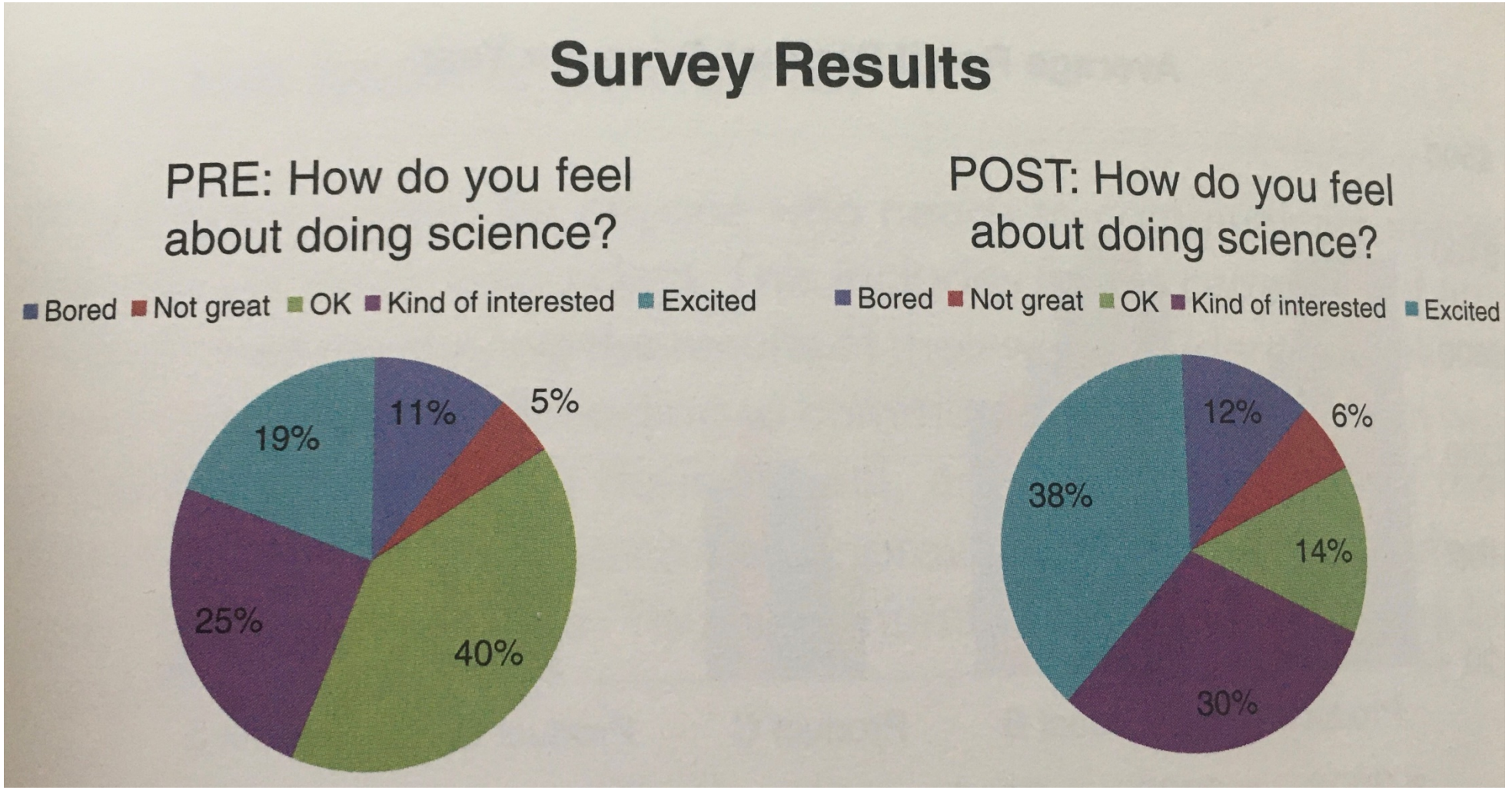
Data Source:
Cole Nusbaumer Knaflic 2015

Bad Example of a Data Visualization Fixed



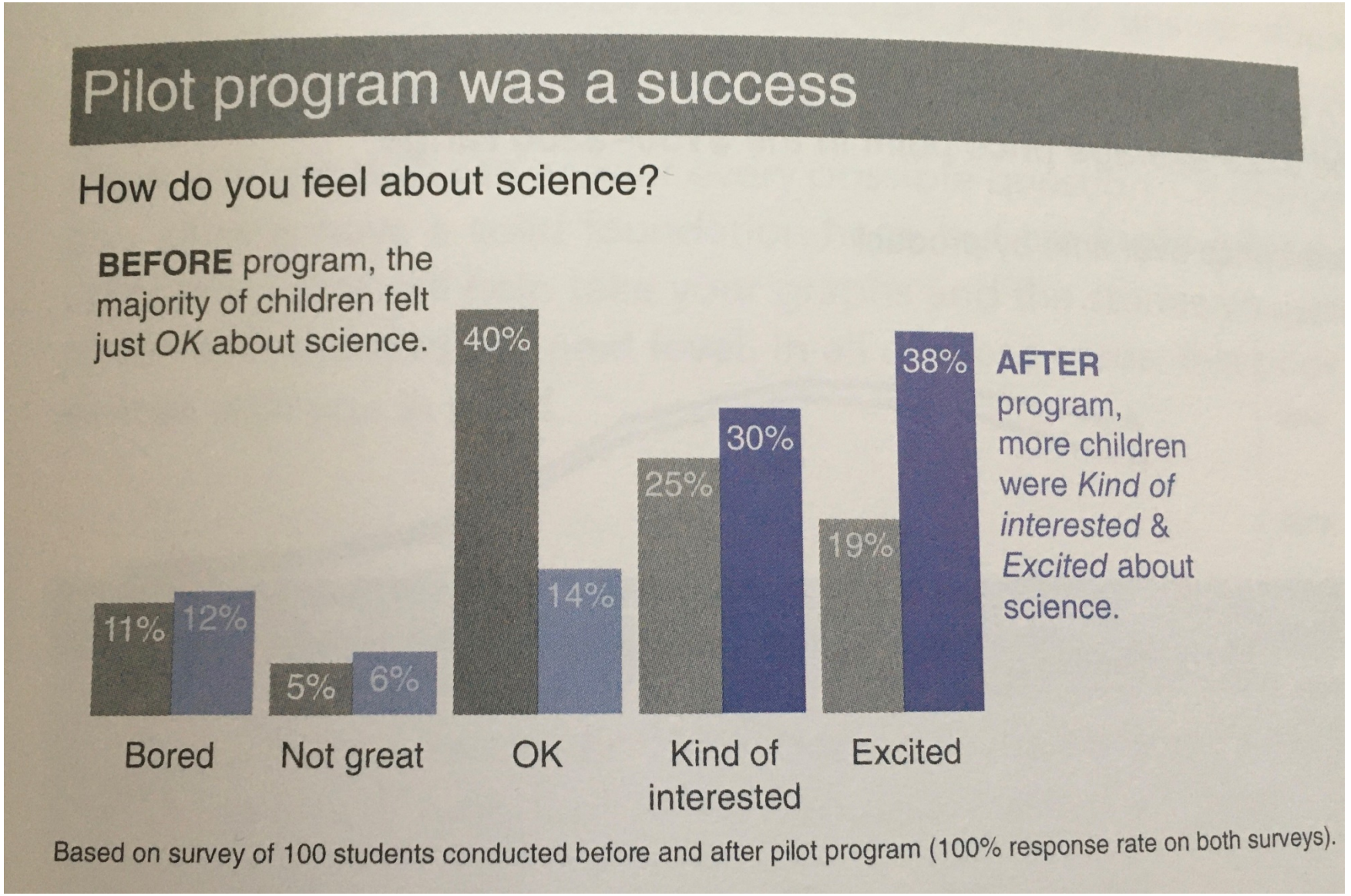
Data Source:
Cole Nusbaumer Knaflic 2015

Ugly Example of a Data Visualization



Data Source:
Cole Nusbaumer Knaflic 2015

Ugly Example of a Data Visualization Fixed



Data Source:
Cole Nusbaumer Knaflic 2015

Thank You!!!

For further questions, please contact:

Nancy Shin, MLIS, B.Sc.

nkshin1@uw.edu

https://nnlm.gov/pnr

References, Part 1

Images:

All images used in this presentation are from Bing.com under a Public Domain license

Books:

Few, S. (2012). Show me the numbers : Designing tables and graphs to enlighten (2nd ed.). Burlingame, Calif.: Analytics Press.

Kelly, M. (2017). Data Visualization: A Guide to Visual Storytelling for Libraries. Journal of Web Librarianship, 11(2), 143-144.

Knaflic, C. (2015). Storytelling with data : A data visualization guide for business professionals. Hoboken, New Jersey: Wiley.Magnuson

Data:

MMWR Reports 1982-2015, Centers for Disease Control and Prevention (CDC) – 2006-2009

Public Library Survey (PLS) – 2000-2017, Institute of Museum and Library Services

References, Part 2

Websites:

Tableau's 7 Data Stories

Tableau, Best Practices for Telling Great Stories

YouTube Videos

Data Visualization and Storytelling with Alberto Cairo & Microsoft Power BI

Making data mean more through storytelling | Ben Wellington | TEDxBroadway

PNR Rendezvous: Tips and Tricks for Learning Data Visualization

Talks at Google: Cole Nussbaumer Knaflic – Storytelling with Data

References, Part 3

Articles:

The American Library Association (ALA) released its annual report "The State of America's Libraries, 2011". (2011). Information Today, 28(6), 10.

Anne Kennan, M., Cole, F., Willard, P., Wilson, C., & Marion, L. (2006). Changing workplace demands: What job ads tell us. Aslib Proceedings, 58(3), 179-196.

Johnson, W. (2011). The Evolution of the Reference Librarian. Community & Junior College Libraries, 17(2), 91-103.

Lebeau, C. (2018). Librarians as Target. Reference & User Services Quarterly, 57(3), 158-161.

Thompson, J. (2015). Changing needs, changing roles: How public libraries are expanding traditional service models to best serve their communities. Reference & User Services Quarterly, 54(3), 2-5.

Vavrek, B. (2000). Downsizing the Public Library. Public Library Quarterly, 18(1), 31-34.