



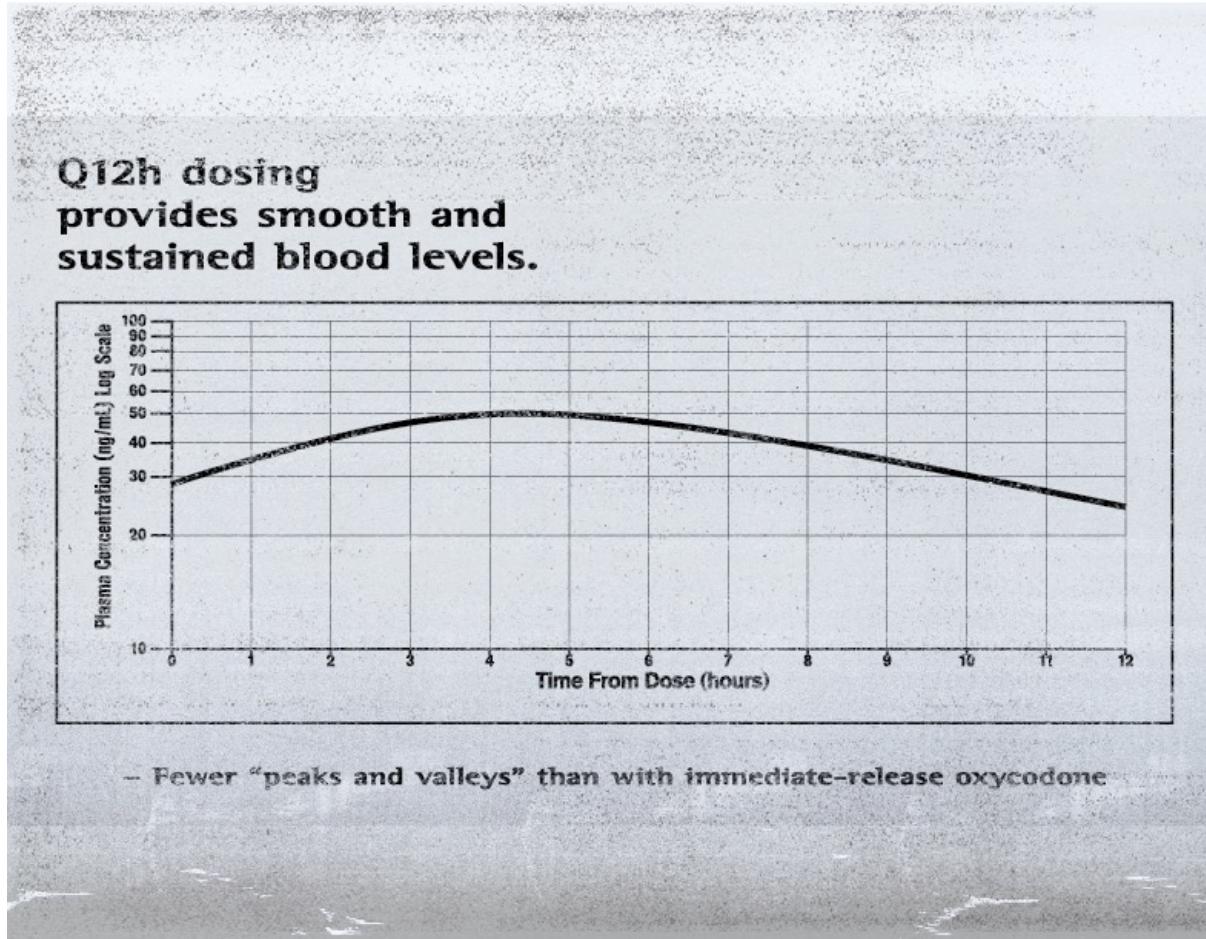
香港浸會大學  
HONG KONG BAPTIST UNIVERSITY

# 数据可视化

---

Instructor: Dr. Xiaoyi Fu

# Anything wrong in this chart?



<https://forms.gle/ykNLJzf1rHp9xXbq5>



# Find materials here!

- <https://github.com/xiaoyifu777/DataVizSummerCourse>



# Who am I

Xiaoyi Fu, PhD

IT Professional – Computer Science Researcher

- Lecturer
  - Department of Journalism- Hong Kong Baptist University



- Background
  - Bachelor of Engineering in Computer Science & Technology, Zhejiang University
  - PhD in Computer Science, HKBU

# Outline

- Introduction to Data Visualization
- Dos and Don'ts
- Hands-on Practice

# A perfect example of data analysis and visualization

- 200 years in 4 minutes <https://www.youtube.com/watch?v=Z8t4k0Q8e8Y>

# What is Data Visualization

- In 1987
  - The National Science Foundation (of the U.S.) started “Visualization in scientific computing” as a new discipline, and a panel of the ACM coined the term “scientific visualization”
  - Scientific visualization, briefly defined: The use of computer graphics for the analysis and presentation of computed or measured scientific data.
- Oxford Engl. Dict., 1989
  - To form a mental vision, image, or picture of (something not visible or present to the sight, or of an abstraction); to make visible to the mind or imagination
- Visualization transforms data into images that effectively and accurately represent information about the data.
  - Schroeder et al. The Visualization Toolkit, 2nd ed. 1998

**Tool to enable a User insight into Data**

# What Does Visualization Do?

- Three types of goals for visualization
  - ... to **explore**
    - Nothing is known,
    - For data exploration
  - ... to **analyze**
    - There are hypotheses,
    - For Verification or Falsification
  - ... to **present**
    - “everything” known about the data,
    - For Communication of Results

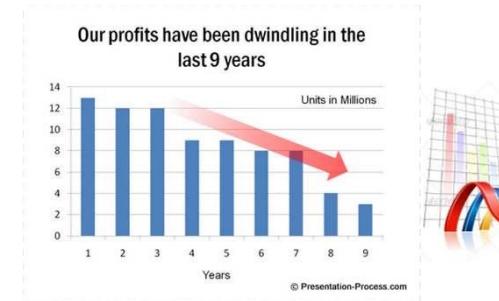
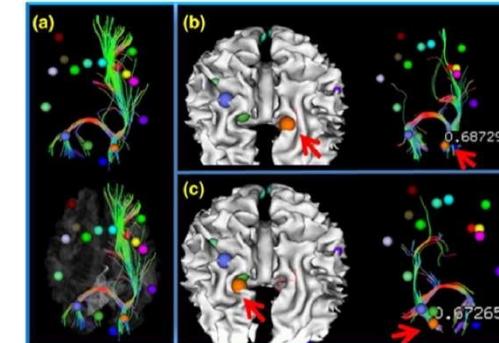
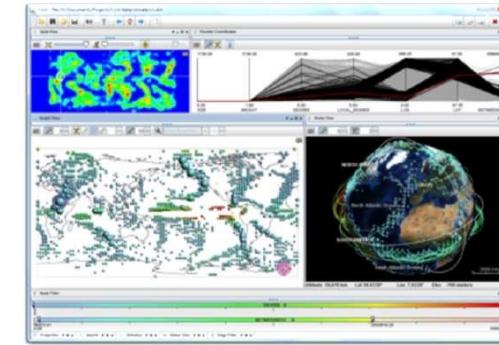


Image source: Google images

The greatest value of a picture is when it forces us to notice what we never expected to see.

—John W. Tukey, *Exploratory Data Analysis*

# Data Visualization Example

- Students applying for graduate study in CS at UIUC
- Students select three faculty members they might want to work with
- Faculty presented with a list of three faculty student selects as a potential advisor

Please select up to three faculty members you would be interested in working with as your graduate thesis advisor...

Faculty #1

E. Gunter

I. Gupta

J. Han

S. Har-Peled

J. Hart

J. Hockenmeyer

Faculty #2

J. Han

S. Har-Peled

J. Hart

J. Hockenmeyer

Derek Hoiem

S. Jacobson

Faculty #3

J. Torrella

M. Viswanathan

T. Warnow

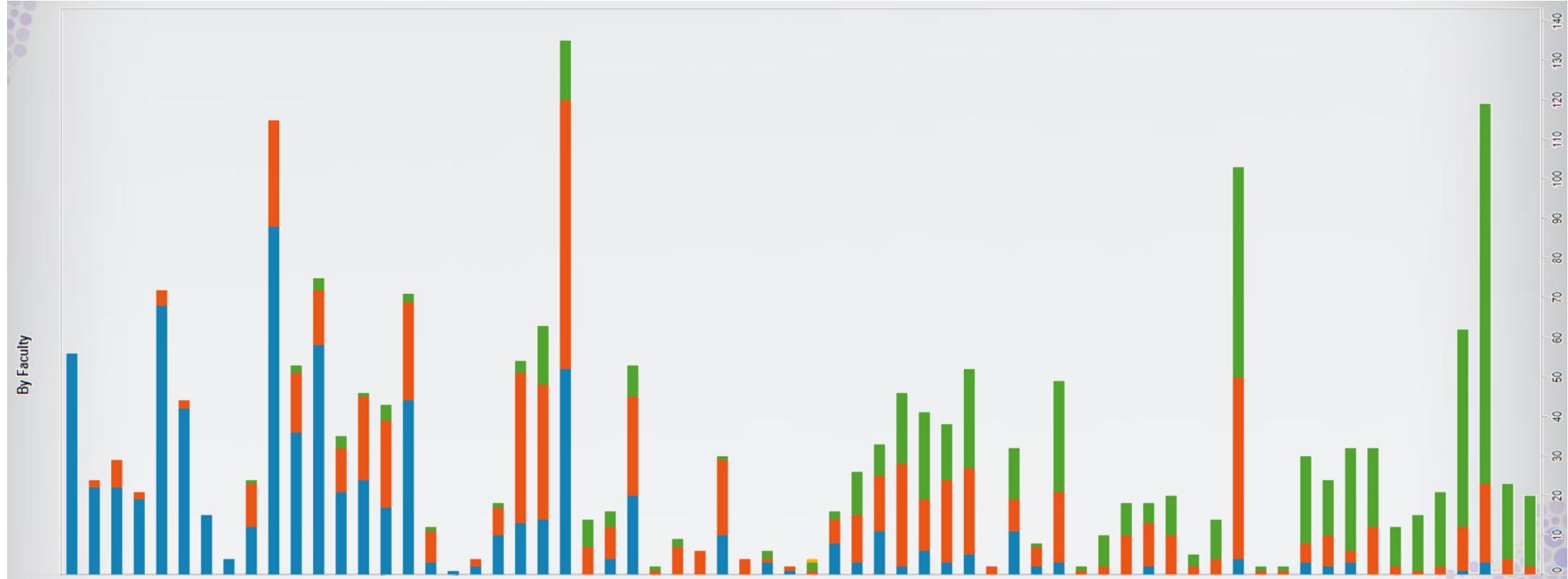
T. Xie

C. Zhai

C. Zilles

Faculty Interest: 1. J. Han, 2. J. Hart, 3. C. Zhai

# Data Visualization Example



1<sup>st</sup> option: blue, 2<sup>nd</sup> option: orange, 3<sup>rd</sup> option: green

# Data Visualization Example

- Students selecting advisors in **alphabetical order**
- Given three sorted pulldown menus of the same option, easiest to pick first option in first menu
- Faculty misunderstood selection order as ranking order

Please select up to three faculty members you would be interested in working with as your graduate thesis advisor...

Faculty #1

- E. Gunter
- I. Gupta
- J. Han**
- S. Har-Peled
- J. Hart
- J. Hockenmeyer

Faculty #2

- J. Han
- S. Har-Peled
- J. Hart**
- J. Hockenmeyer
- Derek Hoiem
- S. Jacobson

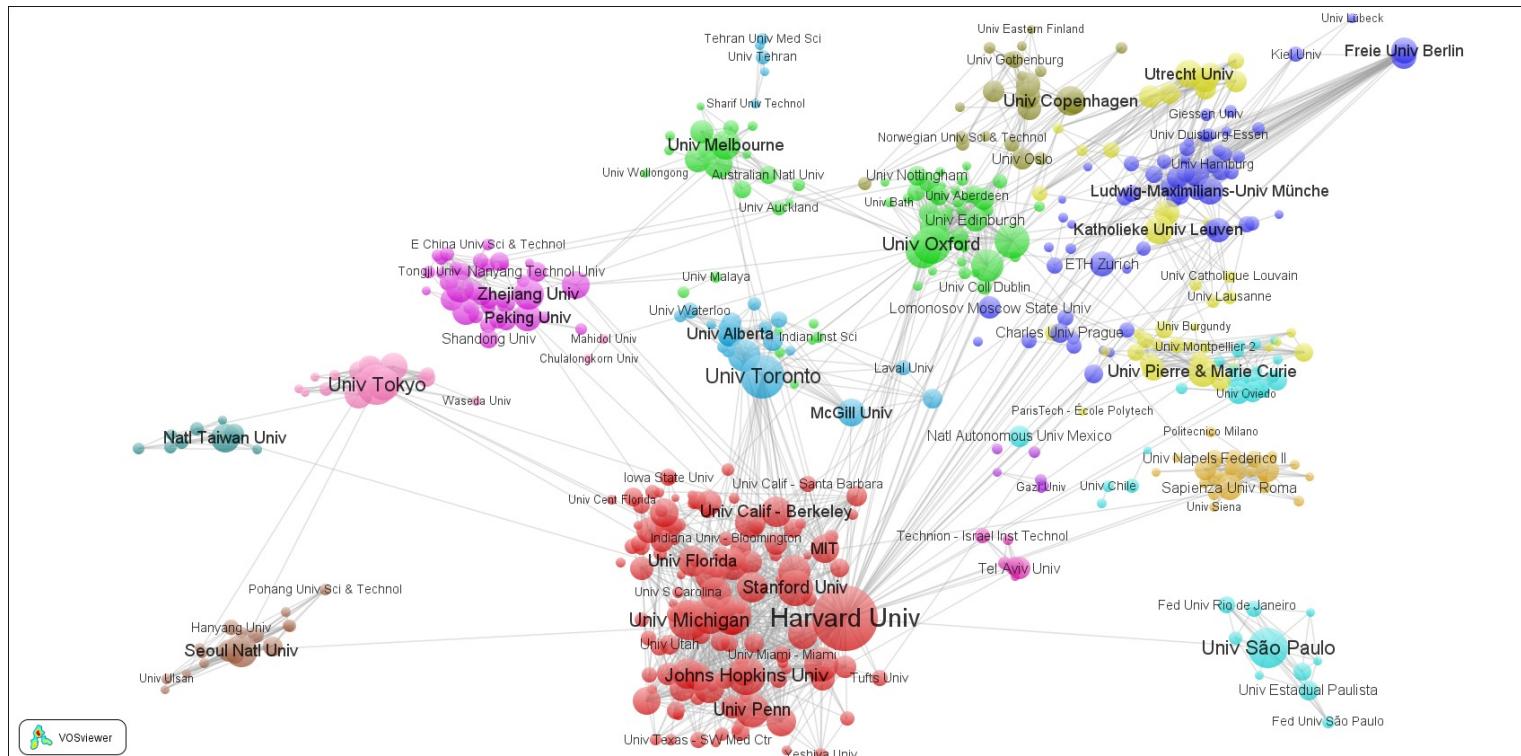
Faculty #3

- J. Torrelas
- M. Viswanathan
- T. Warnow
- T. Xie
- C. Zhai**
- C. Zilles

Faculty Interest: 1. J. Han, 2. J. Hart, 3. C. Zhai

# Information Visualization

- Visualization of more abstract, non-coordinate data
- Relies more heavily on processing abstract data into a more concrete form that can be more effectively perceived by an observer
- [Data to Viz](#)



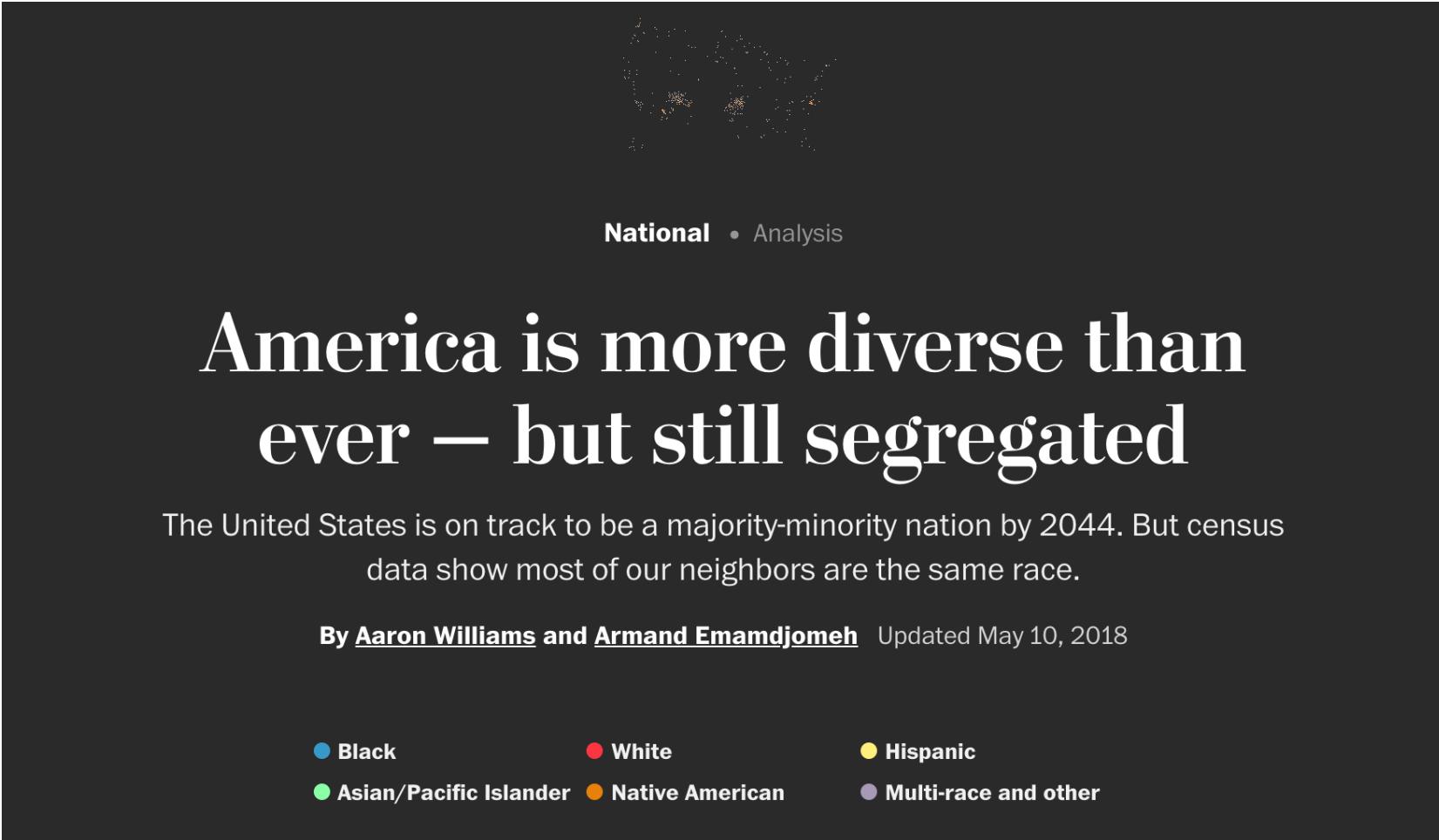
Co-author& university  
collaboration network

# Visualization in news media

The collage displays ten different news media websites, each featuring a unique data visualization:

- The New York Times:** A grid of small images showing women who have accused Larry Nassar of sexual abuse.
- The Washington Post:** A map titled "The insane news cycle of Trump's presidency in 1 chart" showing search interest across the US.
- PROPUBLICA:** A map of the Red Cross building with an inset showing a man in a suit.
- Standard:** A video thumbnail of a drone flying over water.
- FiveThirtyEight:** A choropleth map of the US showing average search interest for "Trump's sex scandal".
- Vox:** A line graph titled "How Popular Is Donald Trump?" showing popularity from January 2017 to April 2018.
- El Confidencial:** A large image of a blue dump truck at night.
- Berliner Morgenpost:** A line graph titled "Meistgelesene Artikel" showing article popularity over time.
- ТЕКСТИ.ORG.UA:** A line graph titled "Динаміка балансу додаткової недоробки" with a large headline in Ukrainian.
- ZEIT ONLINE:** A large image of Martin Schulz speaking.

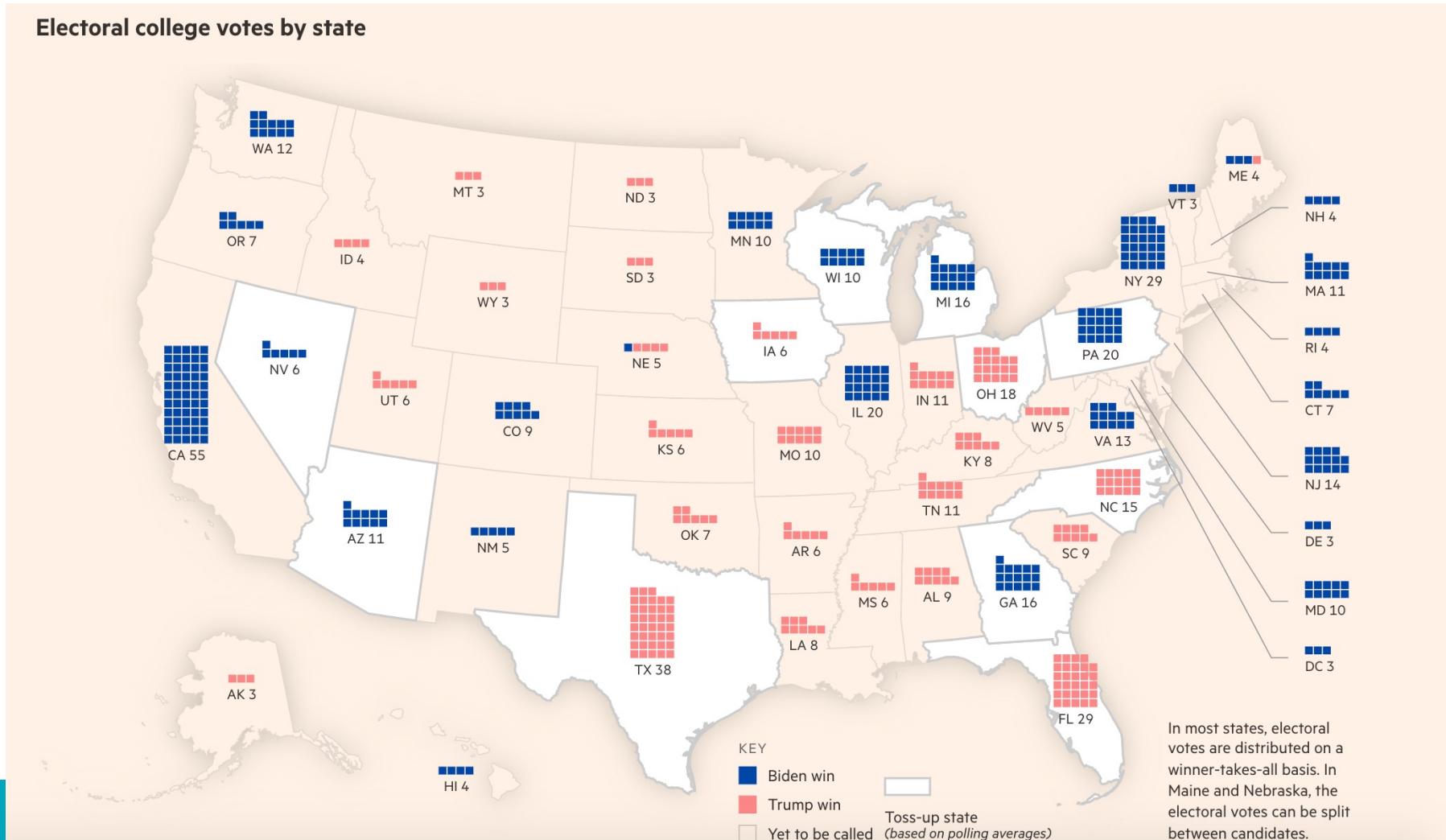
# Visualization in news media



[https://www.washingtonpost.com/graphics/2018/national/segregation-us-cities/?tid==sm\\_pg&utm\\_term=.6785ebcaf520](https://www.washingtonpost.com/graphics/2018/national/segregation-us-cities/?tid==sm_pg&utm_term=.6785ebcaf520)

# Visualization in news media

- Financial Times: Biden vs Trump: live results 2020 [link](#)



# Visualization in news media

<https://m.mp.oeeee.com/a/BAAFRD000020221118741087.html>



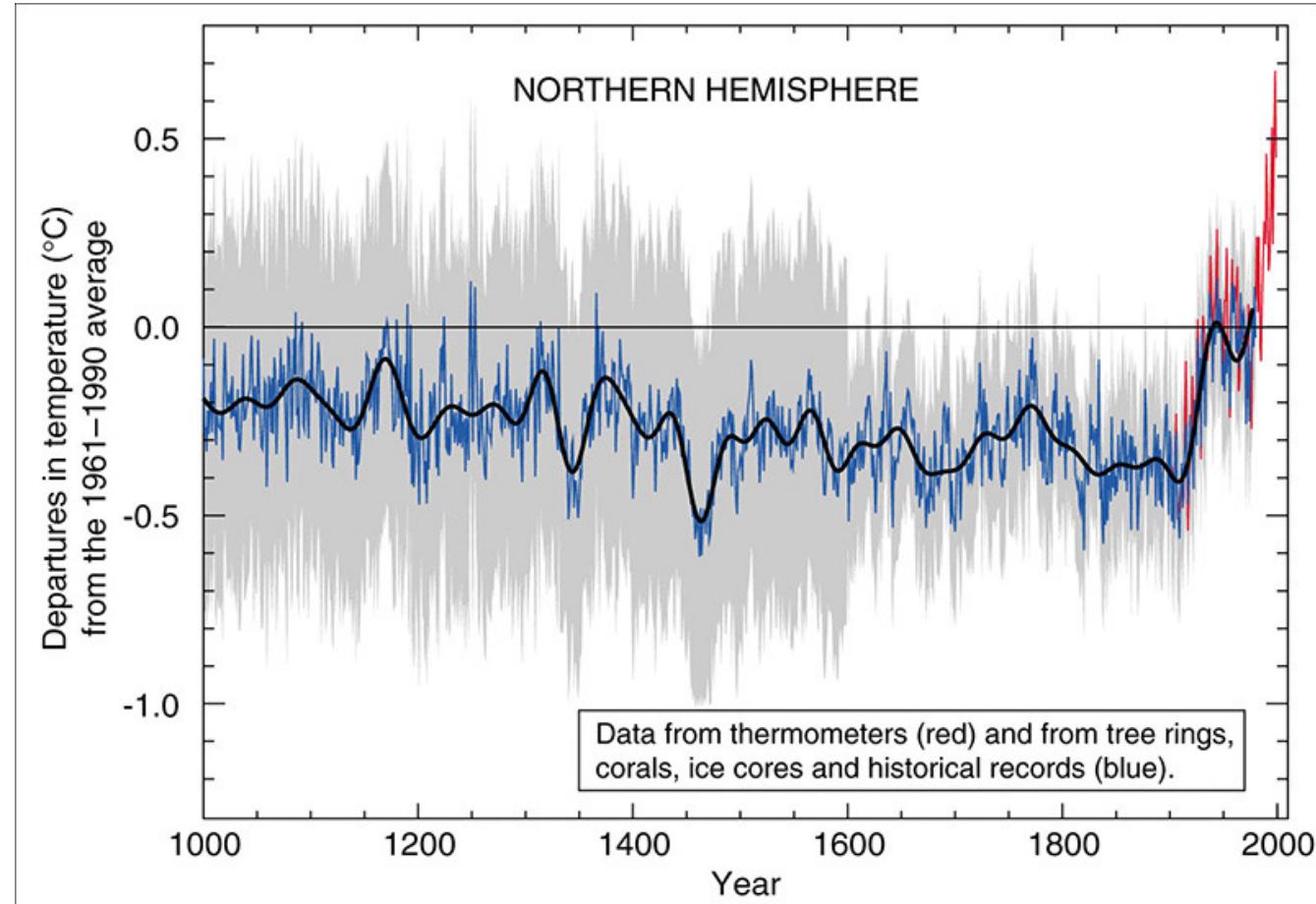
# The Five Qualities of Great Visualizations

---

I see a lot of visualization around me now, and I am extremely excited about it. Yet, are we making any real difference? I mean, are we having any real impact on people's lives other than telling them beautiful stories? Yes, I know, impact could be defined in a million different ways, and it may be hard to capture. But why do I never stumble on to an article or blog post showing, I don't know, for instance, how a visualization helped a group of doctors do something remarkable? Is it just because this stuff does not get reported or what?

— Enrico Bertini, NYU, “Where are the data visualization success stories?”

# Graphics that changed public understanding



The hockey stick chart.

# Graphics that changed public understanding

- In 2001, in Intergovernmental Panel on Climate Change (IPCC) assessment report, it showcased the bomb of a chart
  - That graphic shows temperature variation measured in degrees Celsius in comparison to the 1981 to 1990 average
  - As the systematic record of temperatures began only in the seventeenth century, the researchers had to look into multiple sources of data to draw a complete picture of temperatures in the past millennium. They used proxy variables such as tree-ring growth rates and changes in ice cores, and corals.

# Graphics that changed public understanding

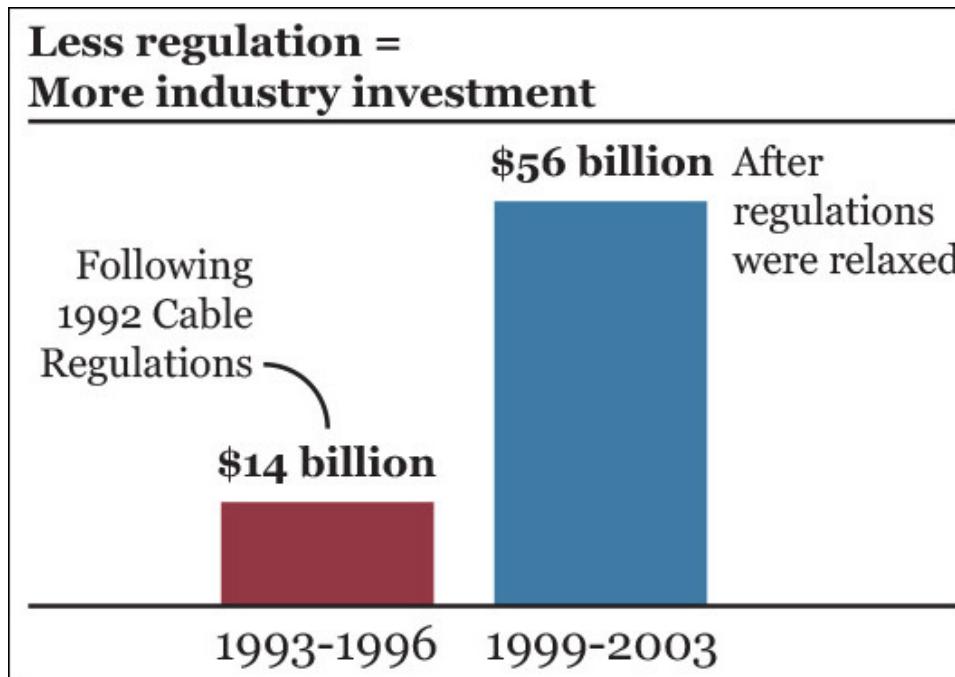
- The message of the chart is **unambiguous**: in the first years of the 20th century, temperatures experienced a sharp rise. This was a time when emissions of human-made greenhouse gases, like carbon dioxide from fossil fuels, increased rapidly.
- After that, the hockey stick became one of the most famous, influential, and disputed graphics ever.

# The Five Qualities of Great Visualizations

- The hockey stick chart is one of the most iconic and persuasive visualizations ever created.
- It's a success story, in Enrico Bertini's words, because it has certain qualities:
  1. It is **truthful**, as it's based on thorough and honest research.
  2. It is **functional**, as it constitutes an accurate depiction of the data, and it's built in a way that lets people do meaningful operations based on it (seeing change in time).
  3. It is **beautiful**, in the sense of being attractive, intriguing, and even aesthetically pleasing for its intended audience—scientists, in the first place, but the general public, too.
  4. It is **insightful**, as it reveals evidence that we would have a hard time seeing otherwise.
  5. It is **enlightening** because if we grasp and accept the evidence it depicts, it will change our minds for the better.

# Truthful

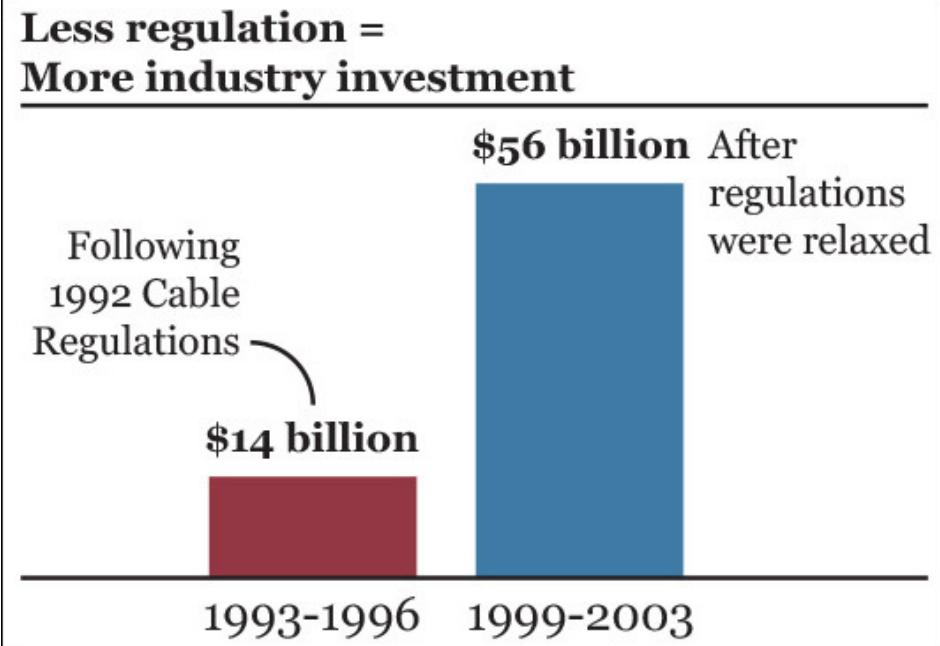
- Should clarity be the prime value of visualization design?



- Here's what the chart suggests: As a consequence of U.S. government regulations in 1992, cable companies invested little money in infrastructure. After regulations were relaxed, investment boomed.

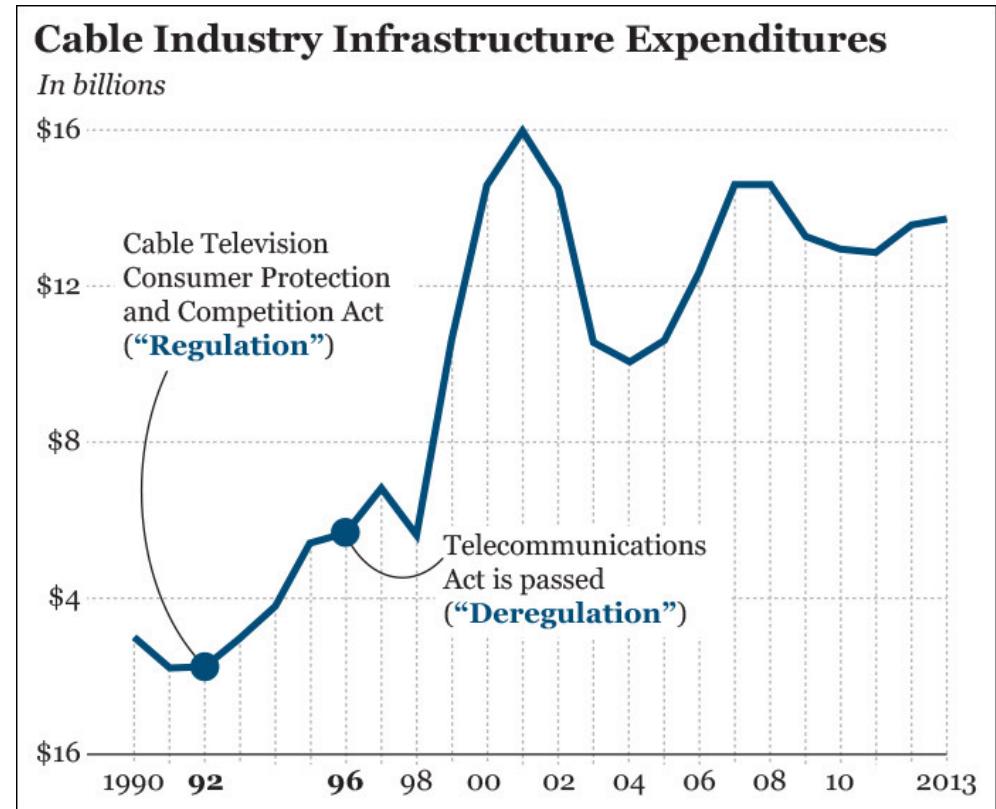
# Truthful

- Not just seeing or looking at visualizations but to reading them.
- What happened between 1996 and 1999?
- There are four years between 1993 and 1996 (first bar) and *five* between 1999 and 2003.
- If someone hides data from you, it's probably because he has something to hide.



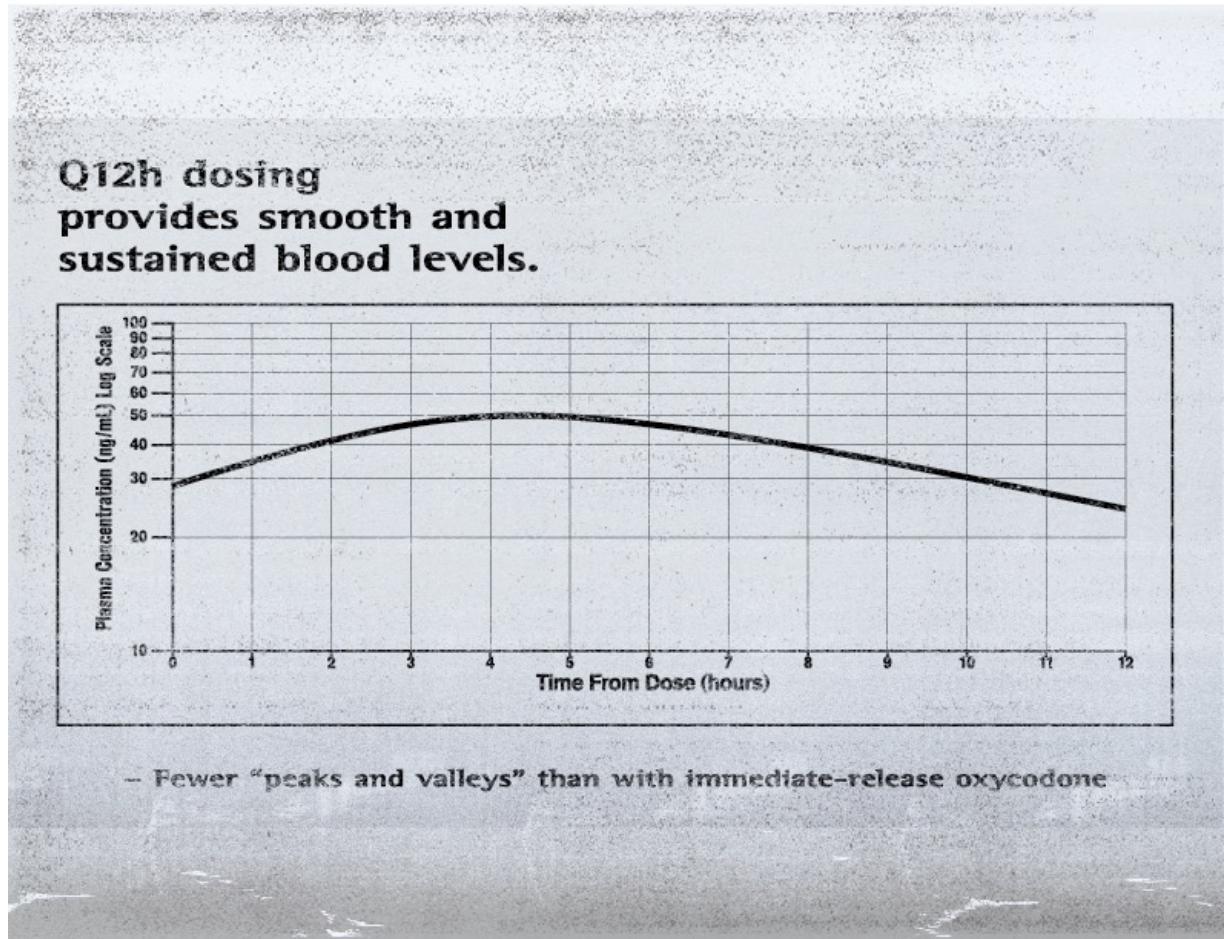
# Truthful

- Following the deregulatory of 1996, investment fell, possibly because of the financial crisis of 1997–98.
- The rapid increase could correspond to the rapid expansion of Internet access and the dot-com bubble between 1997 and 2000. In those years, an increasing amount of people got interested in accessing online content.
  - The spike in the line chart was likely just a reaction to high demand.
- The first graphic is not a complete lie. But the second one is much **truer**.

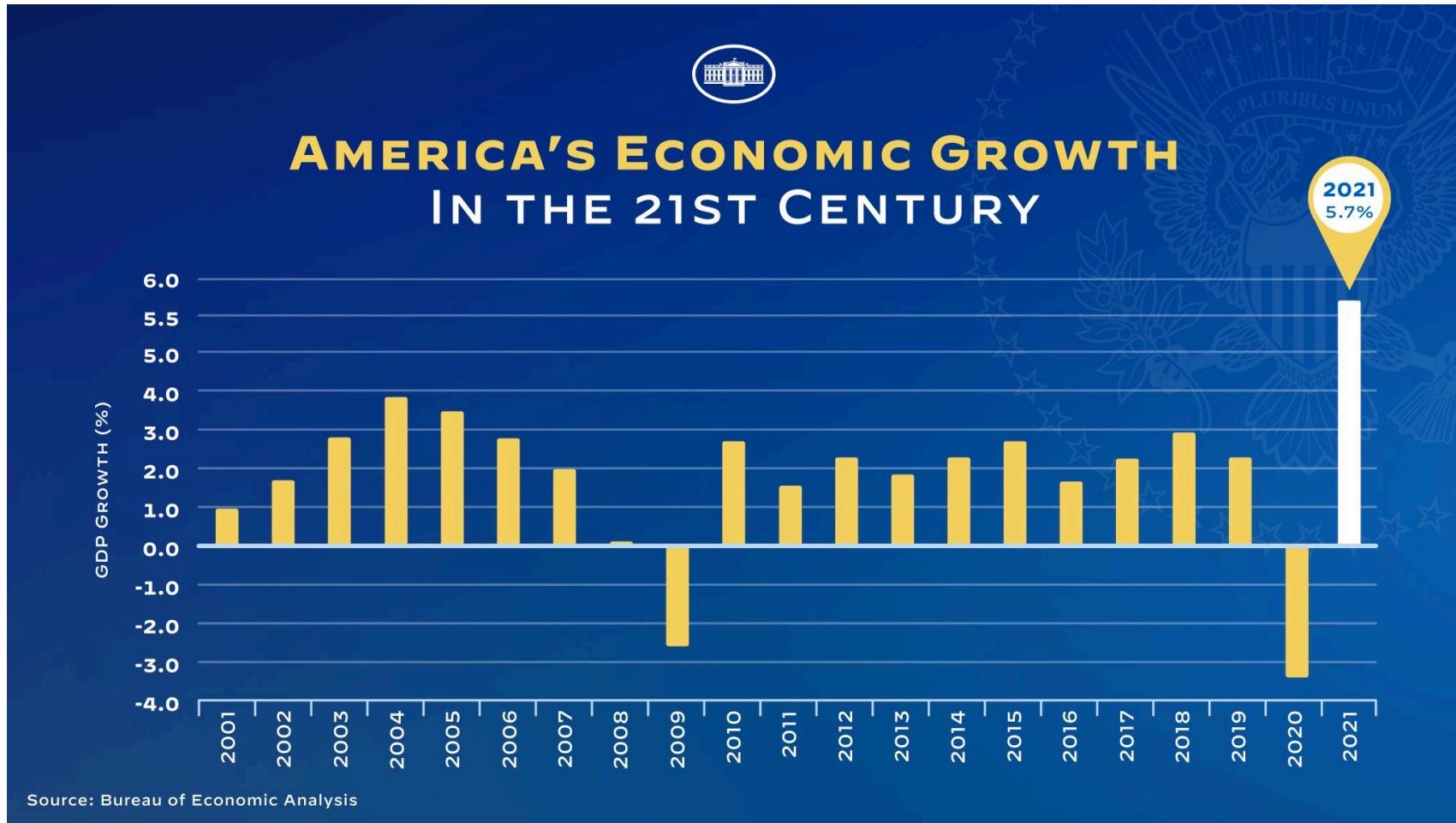


# Truthful

- In TV series *Dopesick*



Source: [https://twitter.com/benjamin\\_rubin/status/1456752559155302412/photo/1](https://twitter.com/benjamin_rubin/status/1456752559155302412/photo/1)

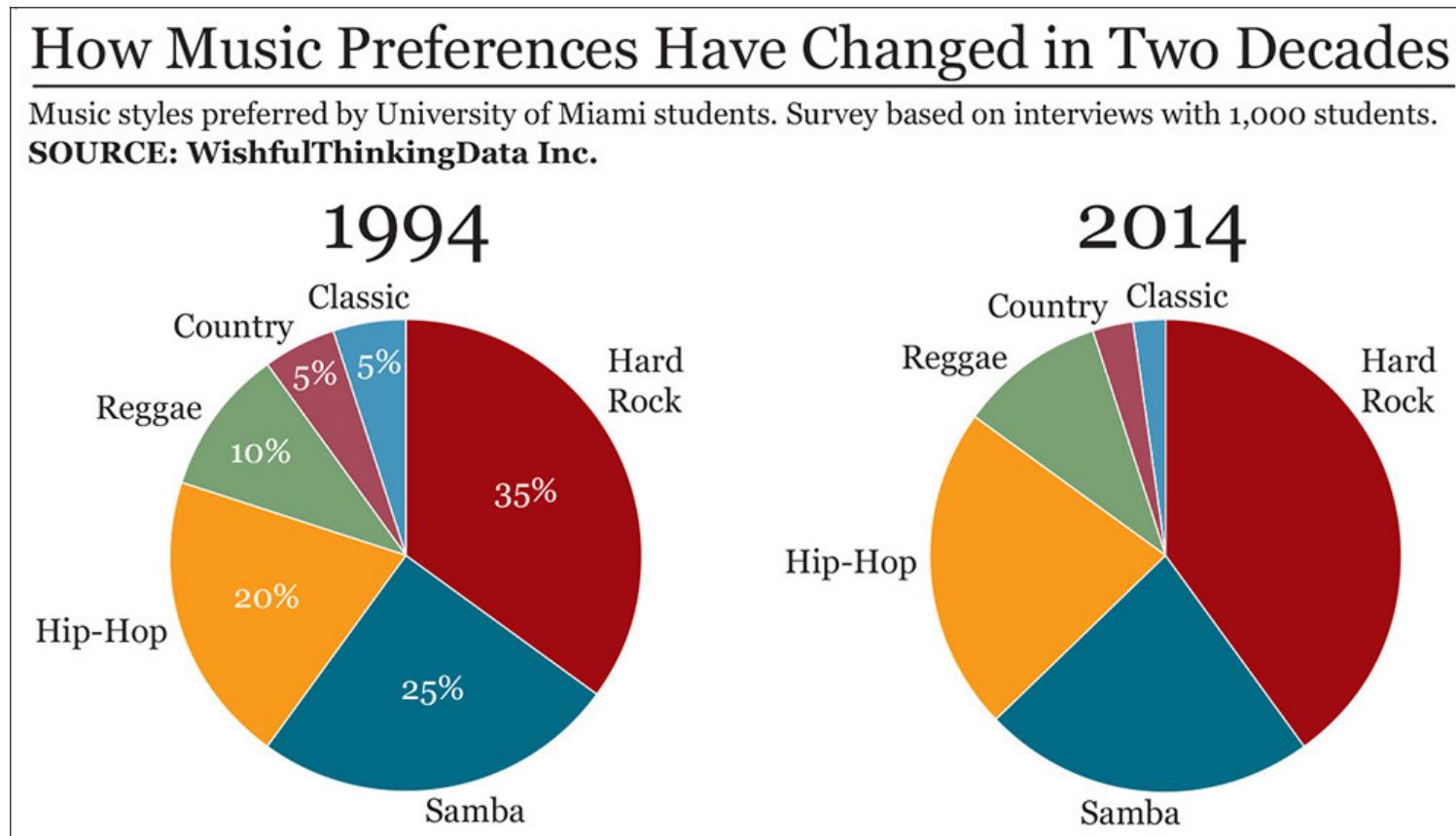


I don't think that clarity should be the prime goal of visualization design.  
Getting our information as right as possible comes first.

-- Alberto Cairo. *The Truthful Art*

# Functional

- If getting your information right is the most important step in creating any visualization, the second one is [helping the audience interpret it correctly](#).



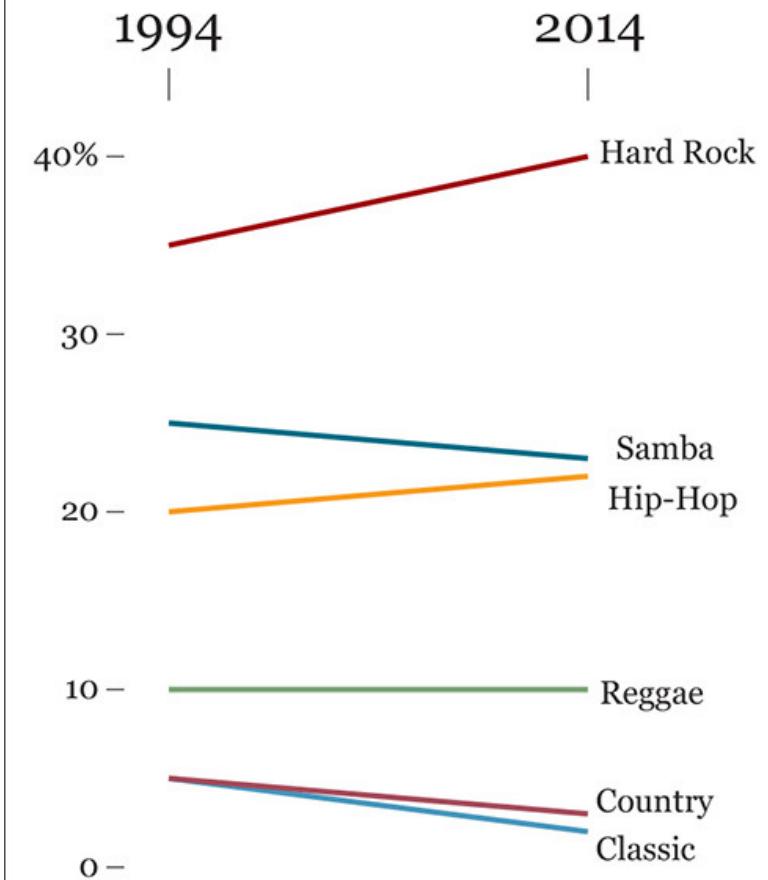
# Functional

- Display exactly the same data as a [slope chart](#)
- What grows bigger goes up, and what becomes smaller goes down.
- This is what functional visualization means: choose graphic forms according to the tasks you wish to enable.
- The purpose of your graphics should somehow guide your decision of how to shape the information.

## How Music Preferences Have Changed in Two Decades

Music styles preferred by University of Miami students.  
Survey based on interviews with 1,000 students.

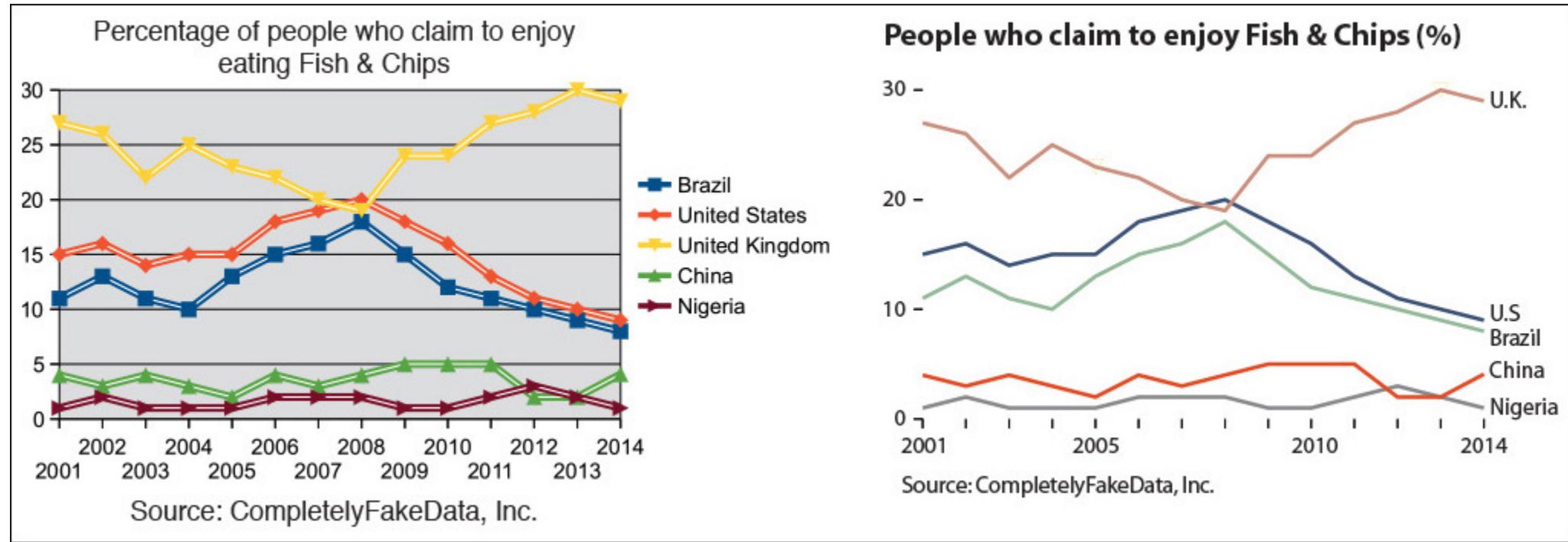
SOURCE: WishfulThinkingData Inc.



# Beautiful

- “Art moves us because it is beautiful, and it is beautiful because it means something. It can be meaningful without being beautiful; but to be beautiful it must be meaningful.”
  - philosopher Roger Scruton
- What matters isn’t if the objects of our creation are beautiful or not, but if they are experienced as beautiful by as many people as possible.

# Beautiful



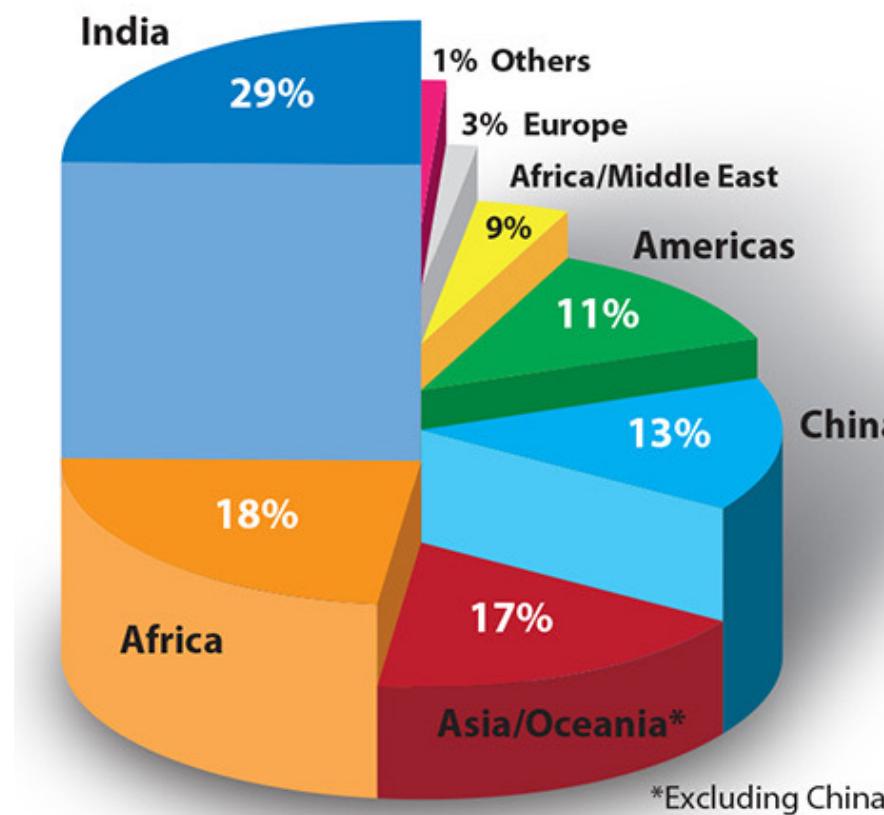
- Both of them present exactly the same content, but the second one does it in a much simpler, clearer, and more elegant way.

# Beautiful

- Beauty is a measure of the emotional experience of awe, wonder or pleasure.
- The very word “aesthetics” comes from the Greek *aísthēsis*, which means sensation or feeling.
- Some things can be beautiful just by virtue of being efficient: they deliver a lot using a minimum amount of elements.
  - The importance of simplicity
- Data visualization isn't the same as data decoration.

# Beautiful

**Share of worldwide urban population growth 2010-2050**

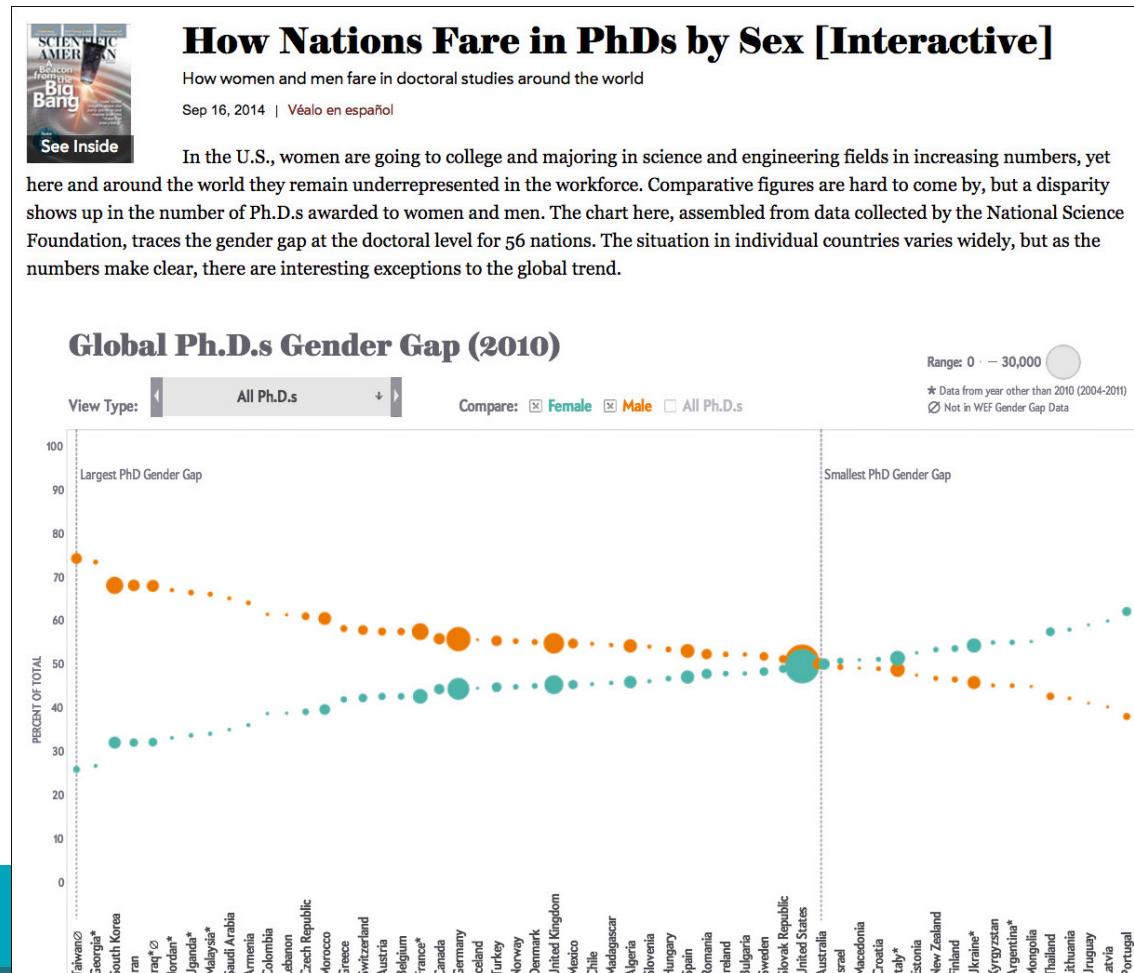


# Insightful

- The purpose of visualization is insight, not pictures
- Insight
  - spontaneous insight, is equivalent to an “a-ha” moment. It’s sudden, surprising, and unexpected.
  - knowledge-building insight, is based on a gradual and deliberate process of exploration of the information that doesn’t necessarily lead to “wow” moments.
- The hockey stick chart is an example of spontaneous insight.
  - Meaningful patterns and trends turned into something they couldn’t ignore.

# Insightful

- Knowledge-building insight is much more common in interactive visualizations.
- Example: How Nations Fare in PhDs by Sex [[link](#)]



# Enlightening

- Great visualizations change people's minds for the better. They are enlightening.
- An enlightening graphic is a consequence of paying attention to the previous four qualities.
  - A graphic that is truthful, functional, beautiful, and insightful has the potential of being enlightening as well.
- But there's something else to consider at this point: the **topic** of the visualization.
  - Choosing topics ethically and wisely—casting light over relevant issues—matters a lot.

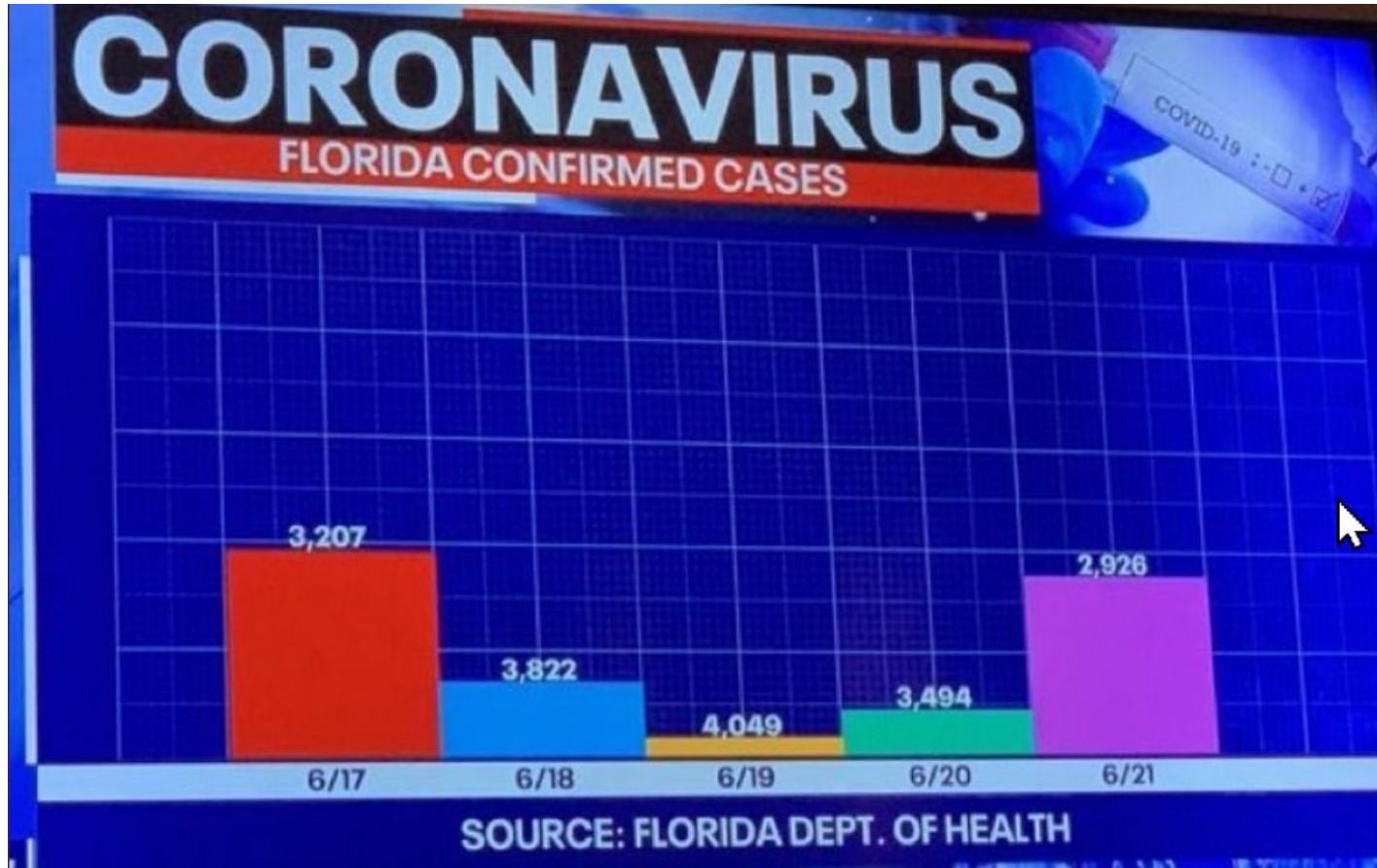
# Bad Examples

---

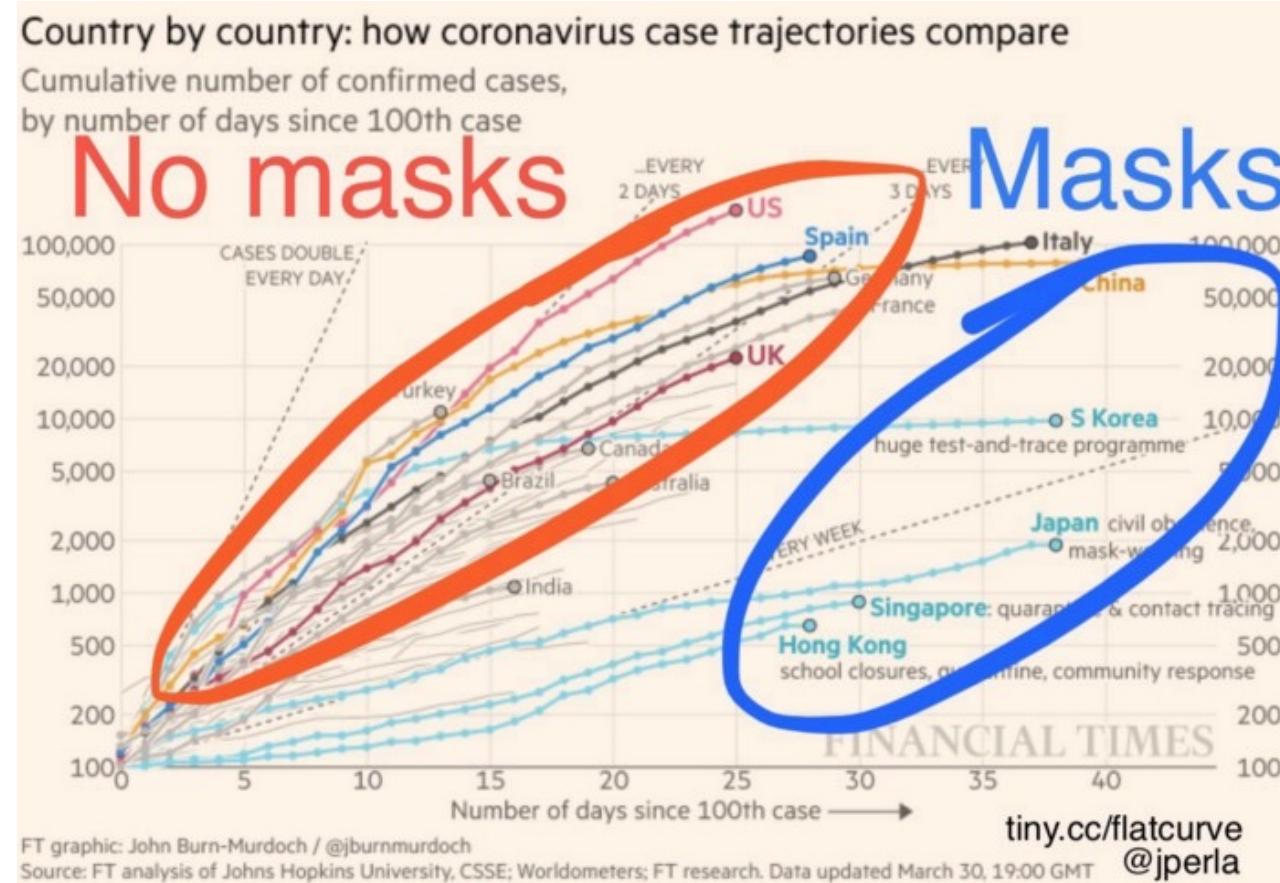
# Bad example



# Bad example

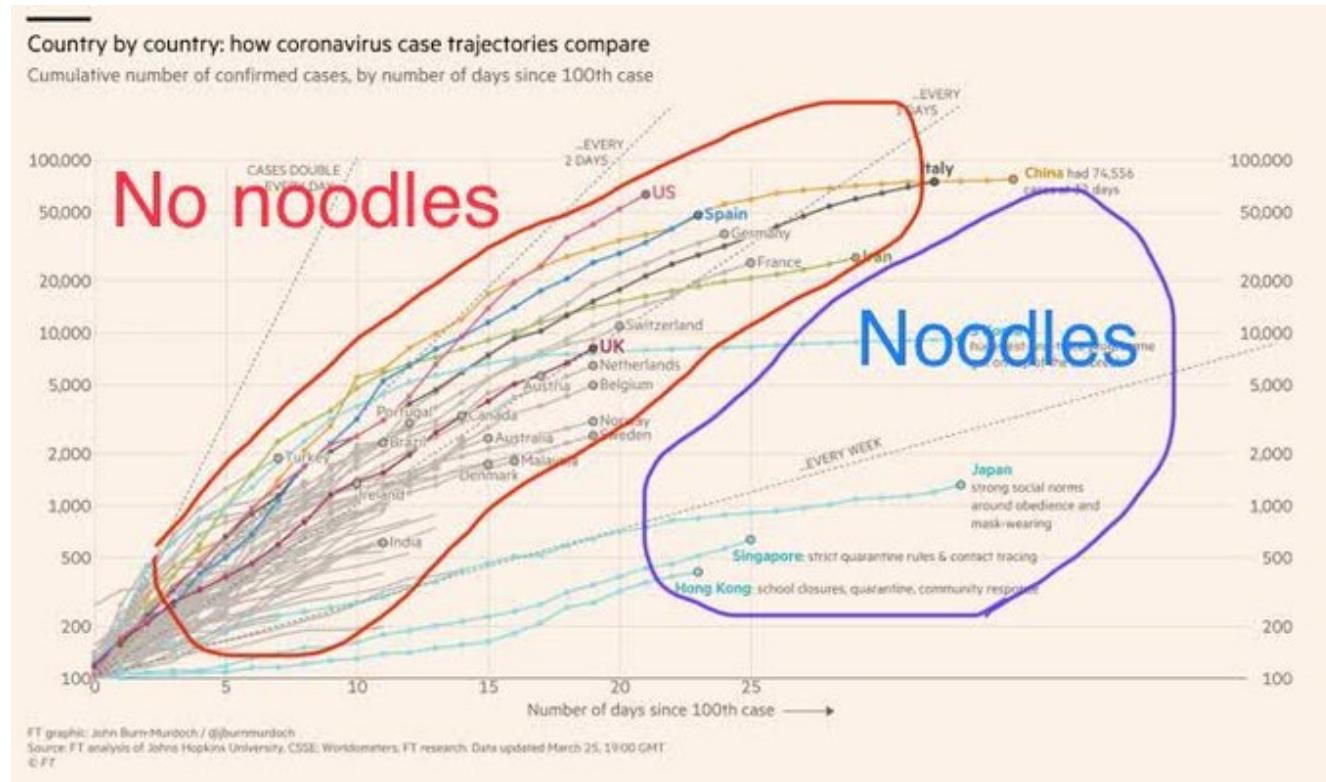


# Bad example

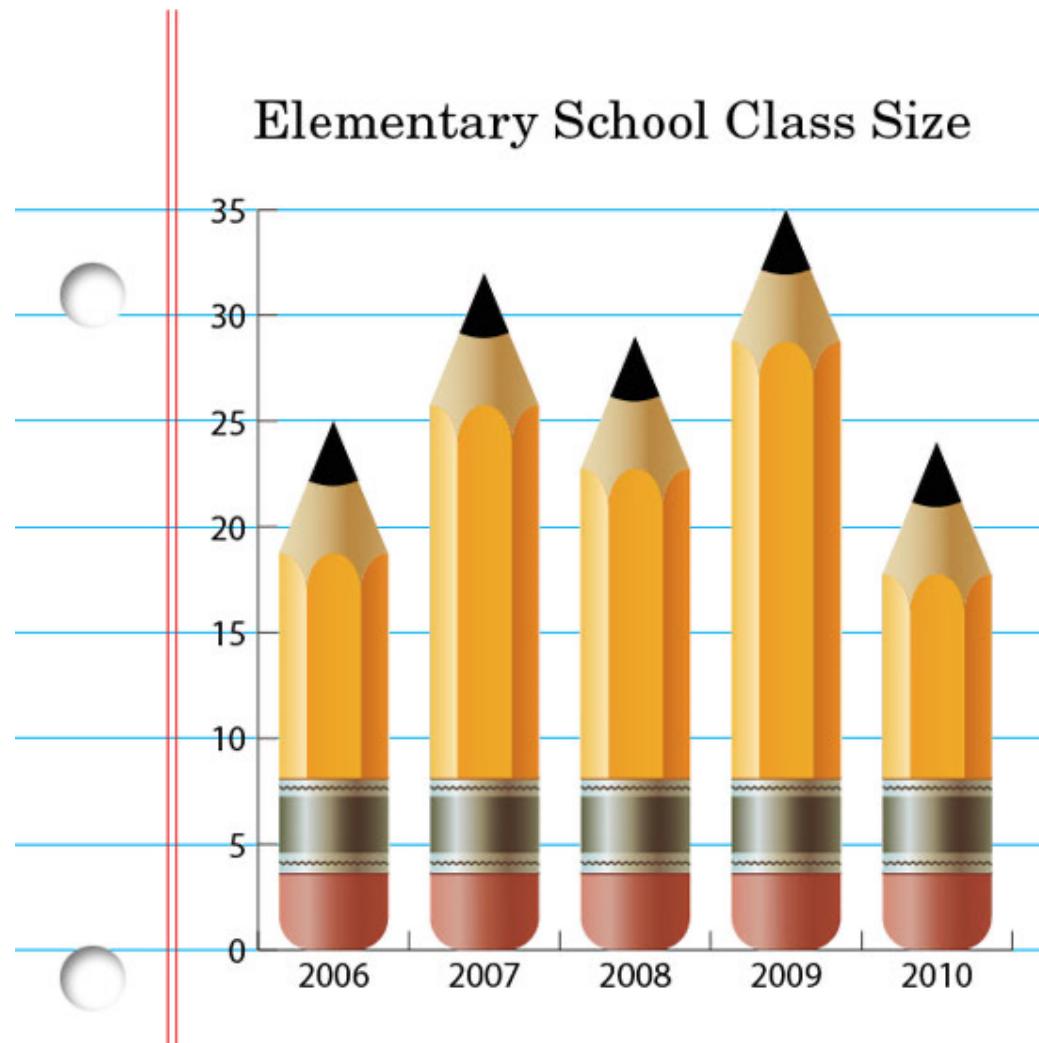


- Whether we agree with the argument or not, this graphic **conflates correlation with causation**

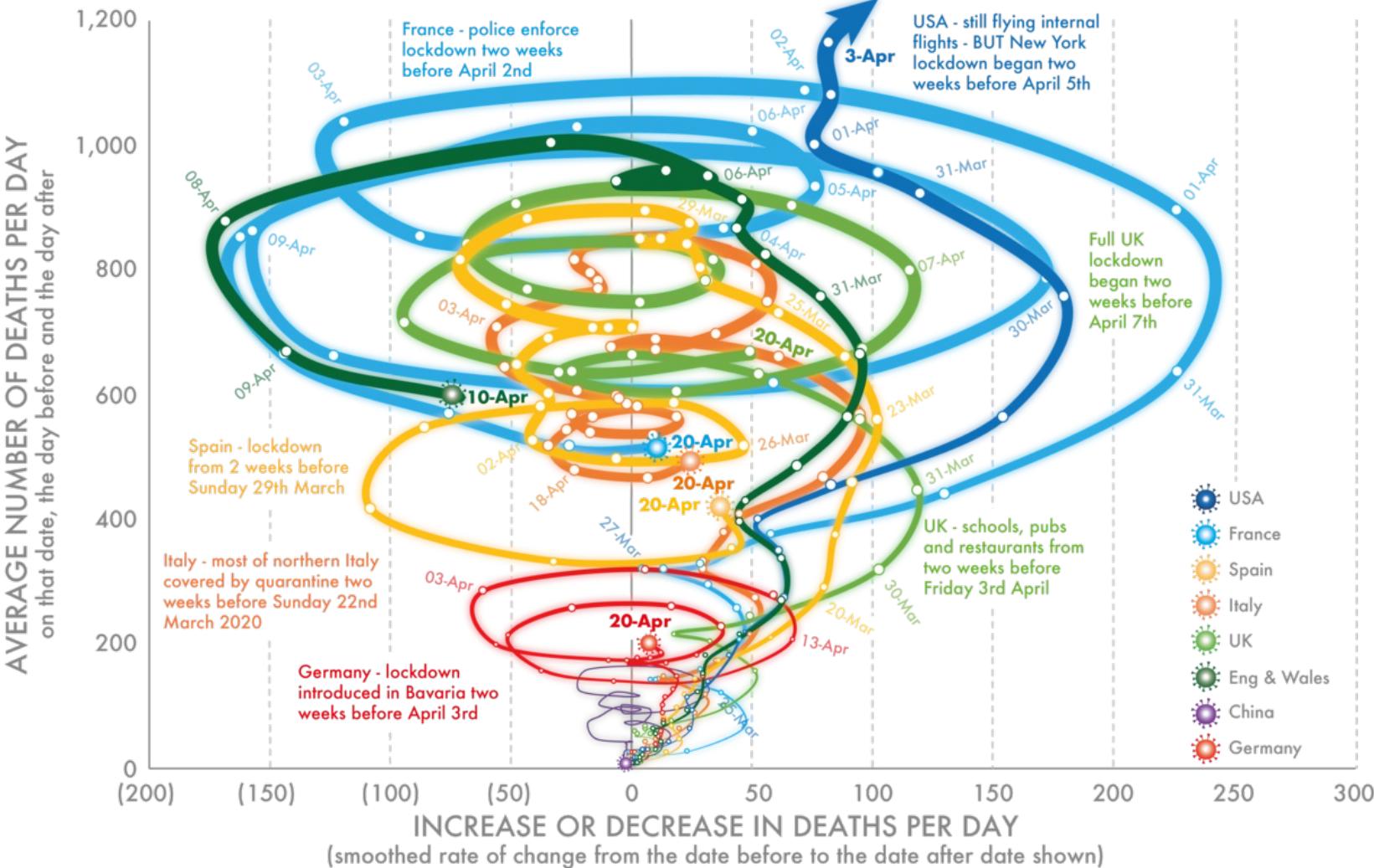
# Bad example



# Bad example



# Bad example



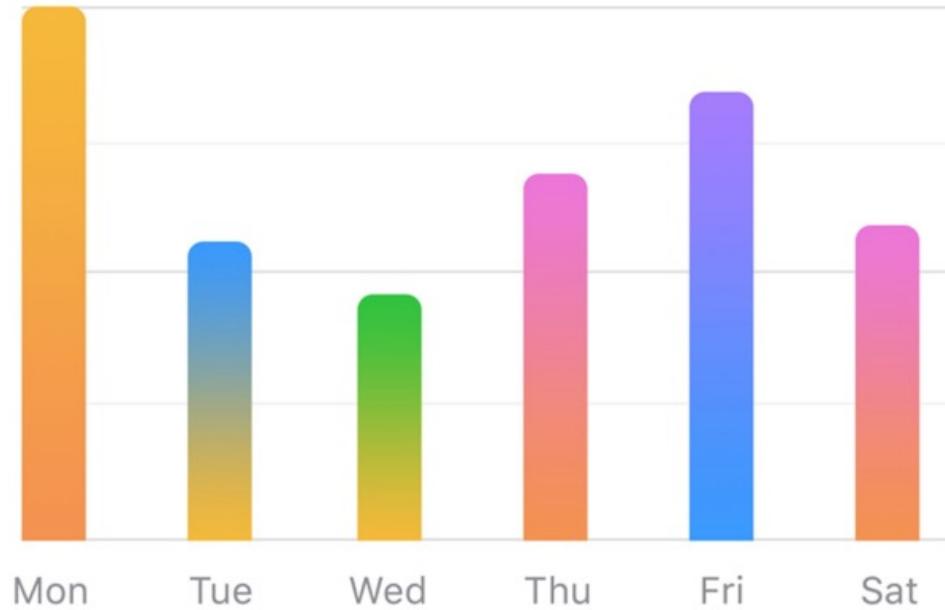
- Please minimize clutter in your infographic

# Bad example

**Mar 18 – 24**

Total Spending

**\$506.72**  8%

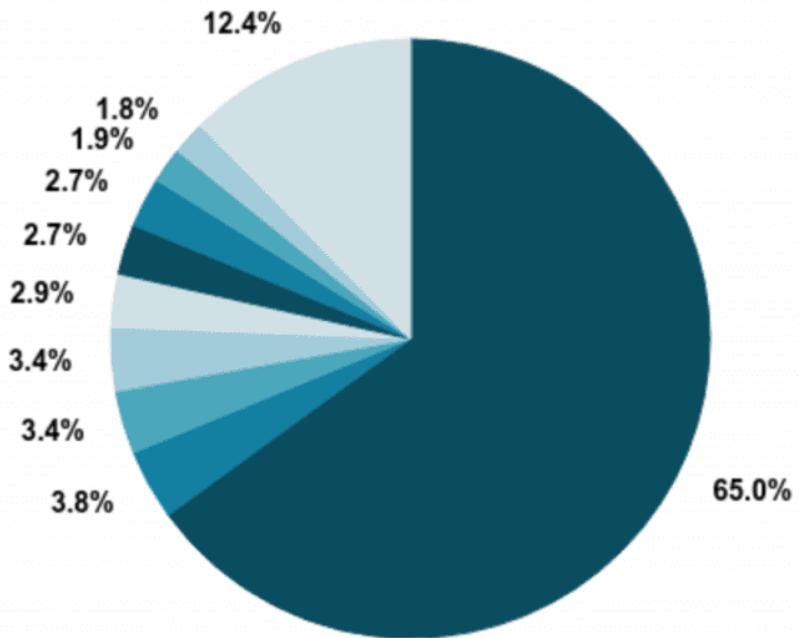


- Use color intentionally
- The gradient colored bars could distract the reader from what they might learn if they focused on the data instead, which is not clearly labeled either

# Bad example

**Foreign investment in China (2018)**

Hong Kong	Singapore	British Virgin Islands	South Korea
Cayman Islands	Japan	Germany	USA
UK	Others		

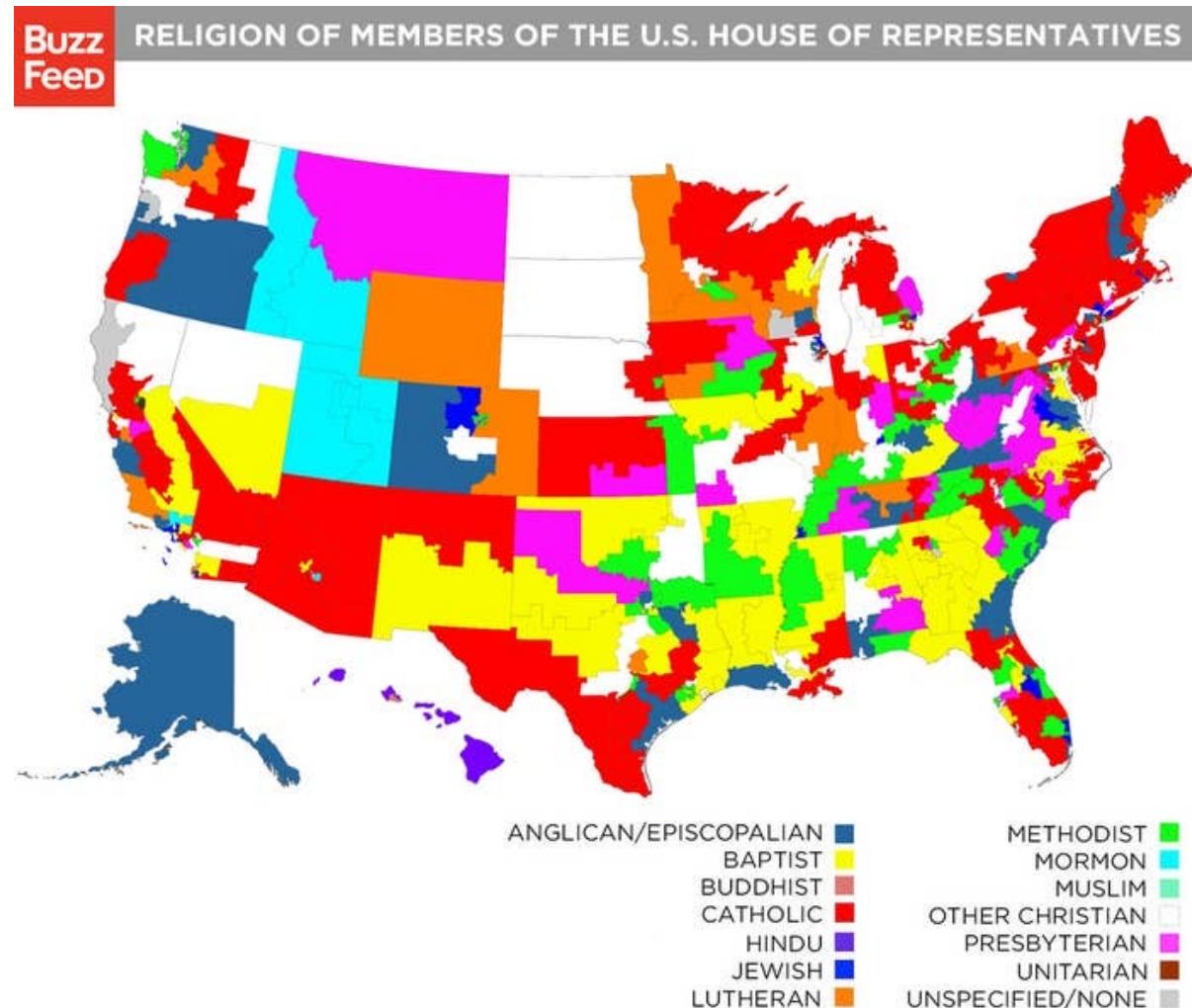


Source: Invest in China

BBC

- Use the same color multiple times to represent different data points.

# Bad example



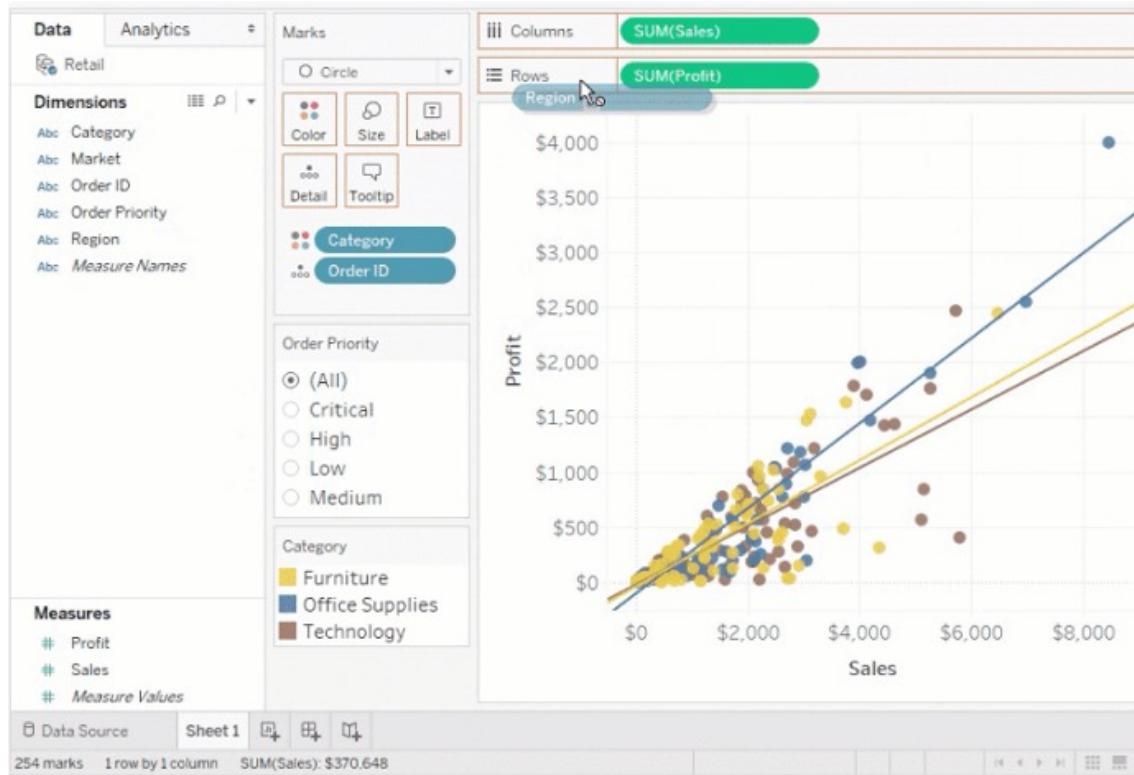
- Too many clashing colors.

# Tableau

---

# Tableau

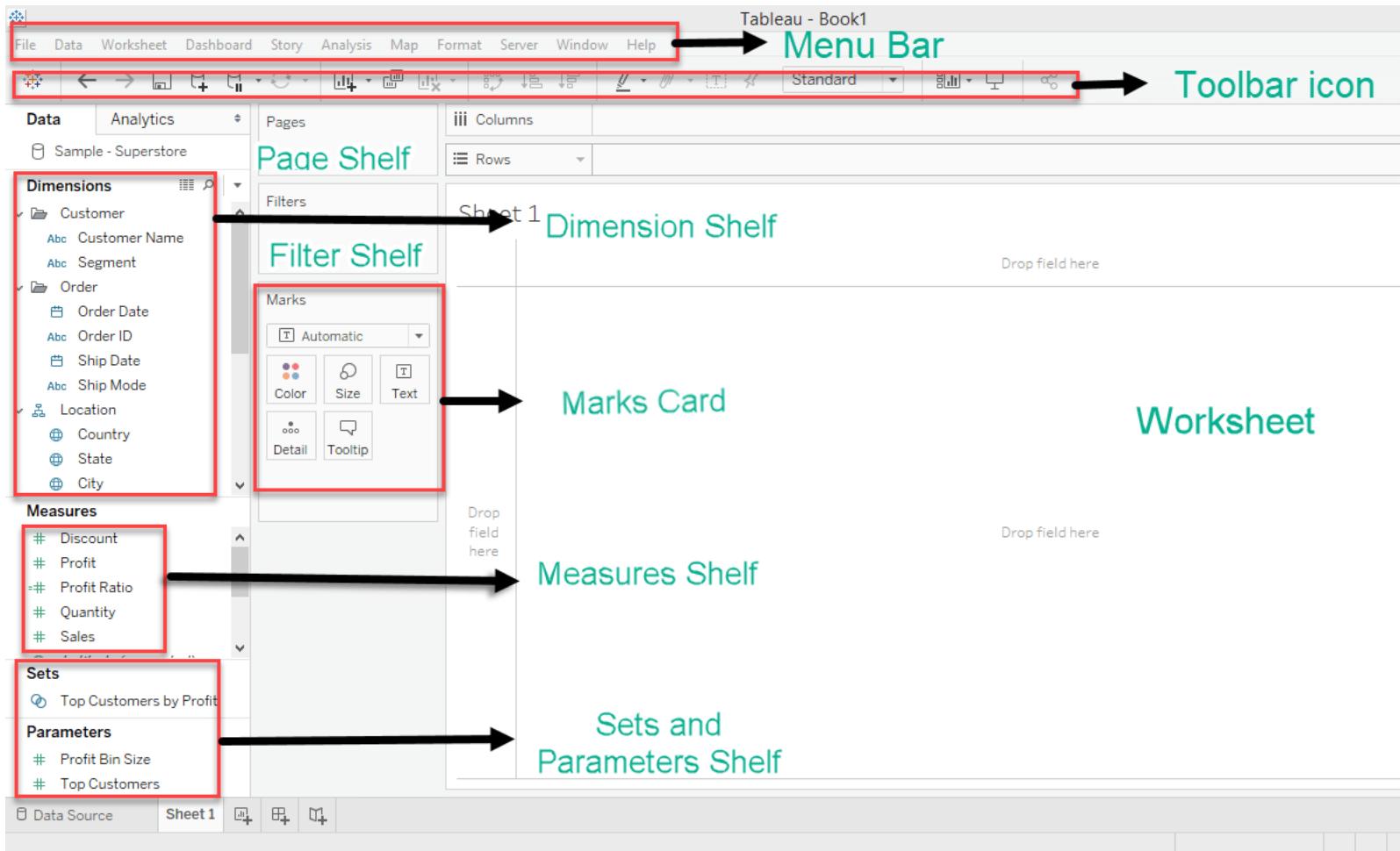
- Tableau Software is an **interactive** data visualization software.



# Tableau

- Tableau Software is an **interactive** data visualization software.





# Viz of the Day

- Every weekday, the Tableau Public team selects one data visualization to be featured as Viz of the Day (VOTD).
- Tableau Public's Viz Wrap: [link](#)
  - Interesting Data Visualizations from 2022

# References

- Stuart Card, Jock Mackinlay, Ben Schneiderman: Readings in Information Visualization
  - Curated collection of research papers
- Colin Ware: Information Visualization: Perception for Design
  - Perceptual principles applied to data visualization
  - Focus on user interaction
- Chuck Hansen, Chris Johnson: The Visualization Handbook
  - Scientific Visualization
- Tamara Munzner: Visualization Analysis & Design
  - Covers both scientific and information visualization
  - Includes task analysis
- Randy Lao: A Beginner's Guide to the Data Science Pipeline

# References

- Alberto Cairo. 2016. The Truthful Art: Data, Charts, and Maps for Communication (1st. ed.). New Riders Publishing, USA.
- <https://www.kaggle.com/datasets/tunguz/data-on-covid19-coronavirus?resource=download>
- Free Public Data Sets For Analysis:  
<https://www.tableau.com/learn/articles/free-public-data-sets>

# Thank You

---

