

Visualization (Vis)

**Storytelling with
Interactive Data Visualizations**



Lecture 11

—
The Dark Side Of Vis





Visualization

The Dark Side Of Vis

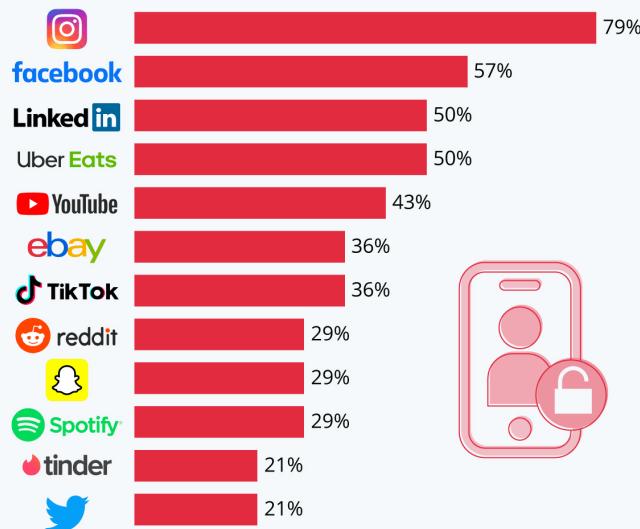
1.  Data Ethics
2. From bad visualizations...
3. ...to actively misleading visualizations



Current & Potential Future Data Ethics Landscape

Personal Data: Instagram Is a Real Tattletale

Percentage of personal data categories shared with third parties by selected iOS apps*



* Based on privacy labels in the App Store, which group user data into 14 categories and inform users what data a given app collects and how it is used.

Source: pCloud



statista

Source: <https://www.statista.com/chart/24495/apps-sharing-personal-information-with-third-parties/>

Visualization

© M. Breunig, TH Rosenheim

The Dark Side Of Vis

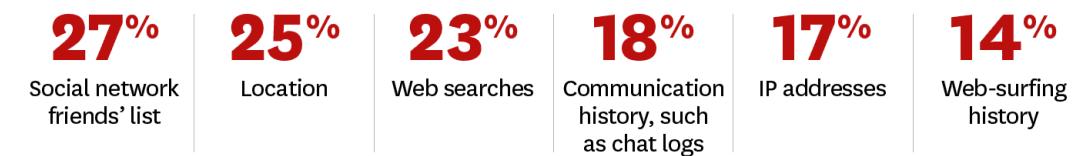
Source: <https://www.datacamp.com/blog/introduction-to-data-ethics>

3

In the Dark About Data

While most people are broadly aware that companies collect data on them, they're surprisingly uninformed about the specific types of data they give up when they go online.

Percentage of people who realize they're sharing their:



SOURCE TIMOTHY MOREY, THEODORE "THEO" FORBATH, AND ALLISON SCHOOP FROM "CUSTOMER DATA: DESIGNING FOR TRANSPARENCY AND TRUST," MAY 2015

© HBR.ORG



Data Ethics

- ◆ **Data ethics** covers the
 - ethical and moral implications of
 - collecting, sharing, and using data,
 - especially personally identifiable information
- ◆ Aspects
 - regulatory compliance: legislation around personal information (GDPR / CCPA)
 - covers employee/company conduct around how data is collected, shared and used
- ◆ Example
 - It may be legal to use freely-given consumer data to segment your customer base, but unethical if it is then used to deliver different levels of service dependent on factors such as race, sex or location.



Sources: <https://www.opendatasoft.com/en/glossary/data-ethics>
<https://online.hbs.edu/blog/post/data-ethics>



Principles of Data Ethics

- ◆ **Transparency:** is it clear what data is being used for, how/where it is stored. Is this information freely available to all?
- ◆ **Accountability:** is there strong oversight and management of data within the project to ensure it is used ethically?
- ◆ **Fairness:** will your use of data have discriminatory consequences for particular groups, even if this is unintended?
- ◆ **Ownership:** individuals have ownership over their data and how it is used. Have you received their informed consent for its use in your particular context?
- ◆ **Privacy:** is personal data being protected and kept secure, so that any identifiable information is not available to unauthorized users? Has it been anonymized to further protect privacy?
- ◆ **Intention:** do you have a clear reason for using particular datasets? Is all the information you have collected relevant, necessary, and appropriate to your stated intention?
- ◆ **Outcomes:** have you investigated any potentially negative, harmful or discriminatory outcomes from your use of particular data? It is vital to address this at the beginning of a project, not when any impacts have occurred.

Source: <https://www.opendatasoft.com/en/glossary/data-ethics>



Advantages of taking an ethical approach

- ◆ Greater customer trust
- ◆ Higher revenues
- ◆ Meet legislative requirements
- ◆ Protect organizational reputation





Visualization

The Dark Side Of Vis

1. Data Ethics
2.  From bad visualizations...
3. ...to actively misleading visualizations



Bad Vis and Misleading Vis

- **Bad visualizations**

- Distort data by visual means
- Are hard to interpret
- May or may no lead the audience to the wrong conclusions
- Have no underlying intention to actively mislead

- **Actively misleading visualization**

- Distort data by visual means
- Are “easy” to interpret
- Create a difference between “what your eyes see” and “what the data really say”
- Have an underlying intention to actively mislead

➔ Distinction not easy and often unclear/blurry



Example 1



- ◆ This donut chart shows the number of outbound airport passengers, comparing the values from 2019 and 2024.
- ◆ Issues
 - No legend → impossible to interpret
 - Wrong chart type → hard to compare
 - No logical conclusions possible

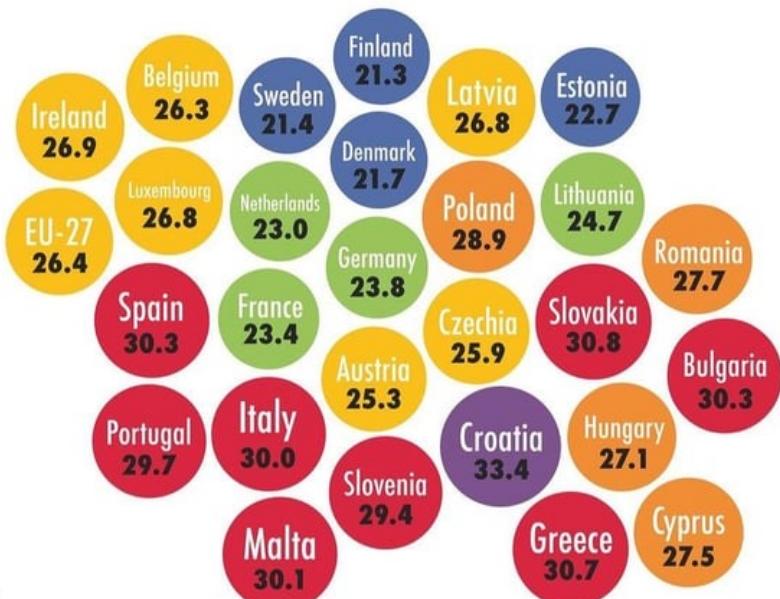


Example 2

When Europeans fly nest

Average age at which young people leave the parental household, 2022

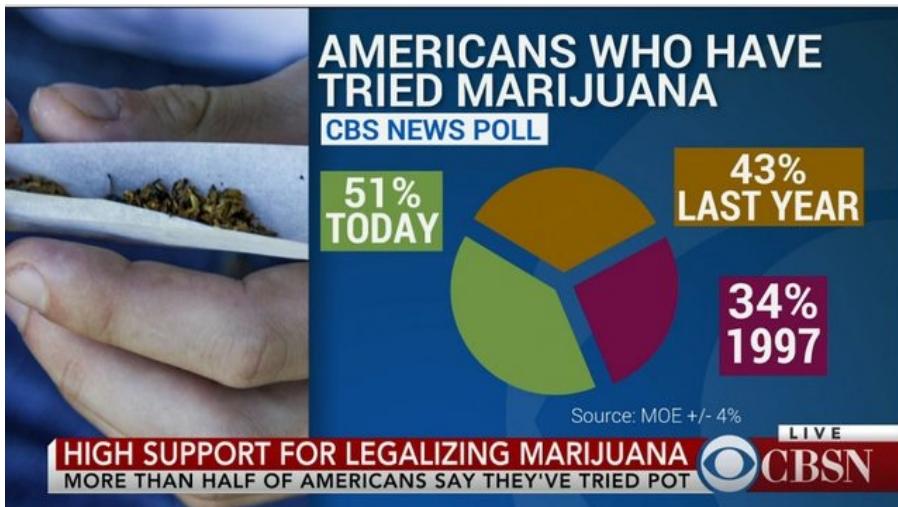
■ below 23 ■ 23-25 ■ 25-27 ■ 27-29 ■ 29-31 ■ over 31



Source: Eurostat



Example 3



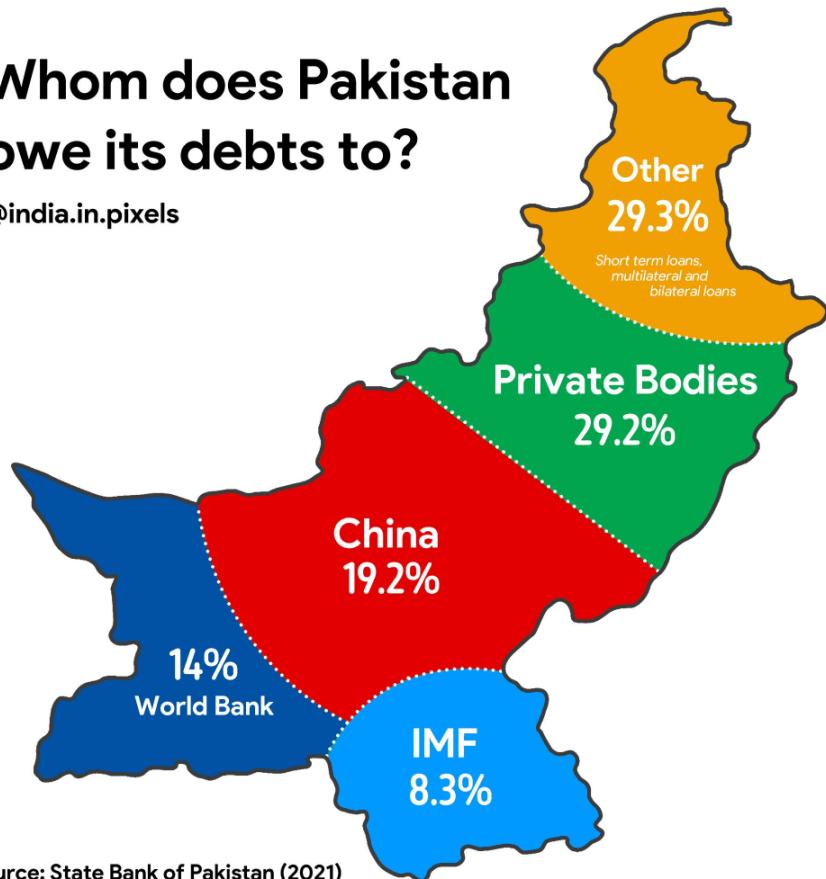
- ◆ Pie chart shows the percentage of Americans who have tried marijuana in three different years.
- ◆ Issues
 - Wrong chart – pie chart used to show percentages over time (should show percentages of a whole)



Example 4

Whom does Pakistan owe its debts to?

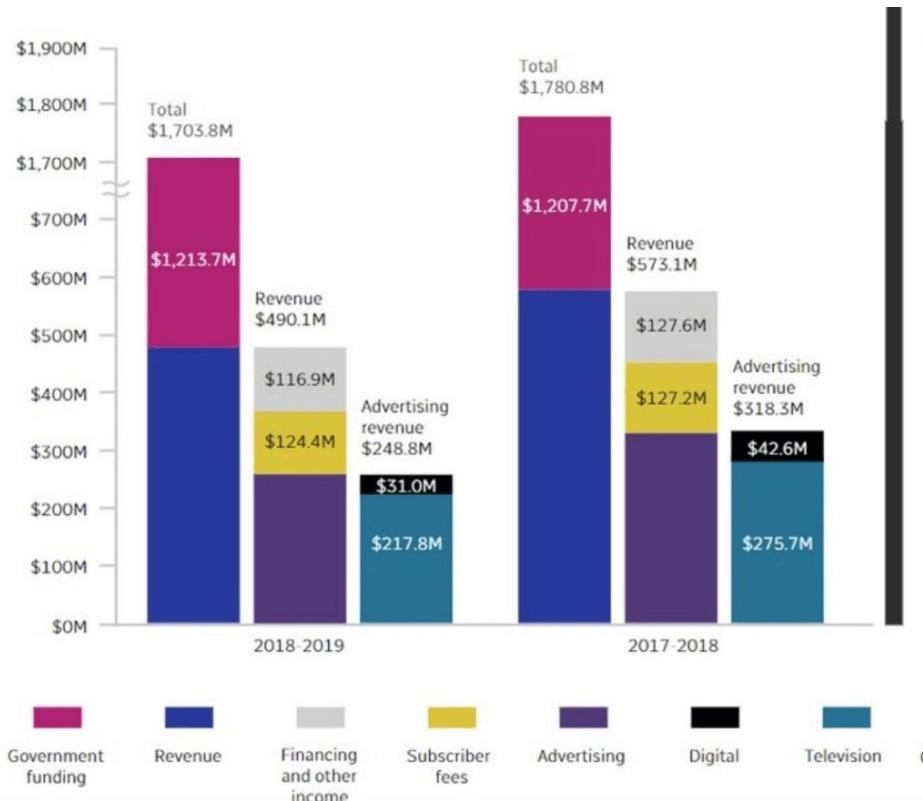
@india.in.pixels



- ◆ Charts shows the percentage of debts owed by Pakistan to different institutions.
- ◆ Issues
 - Are not proportional to the dept owed



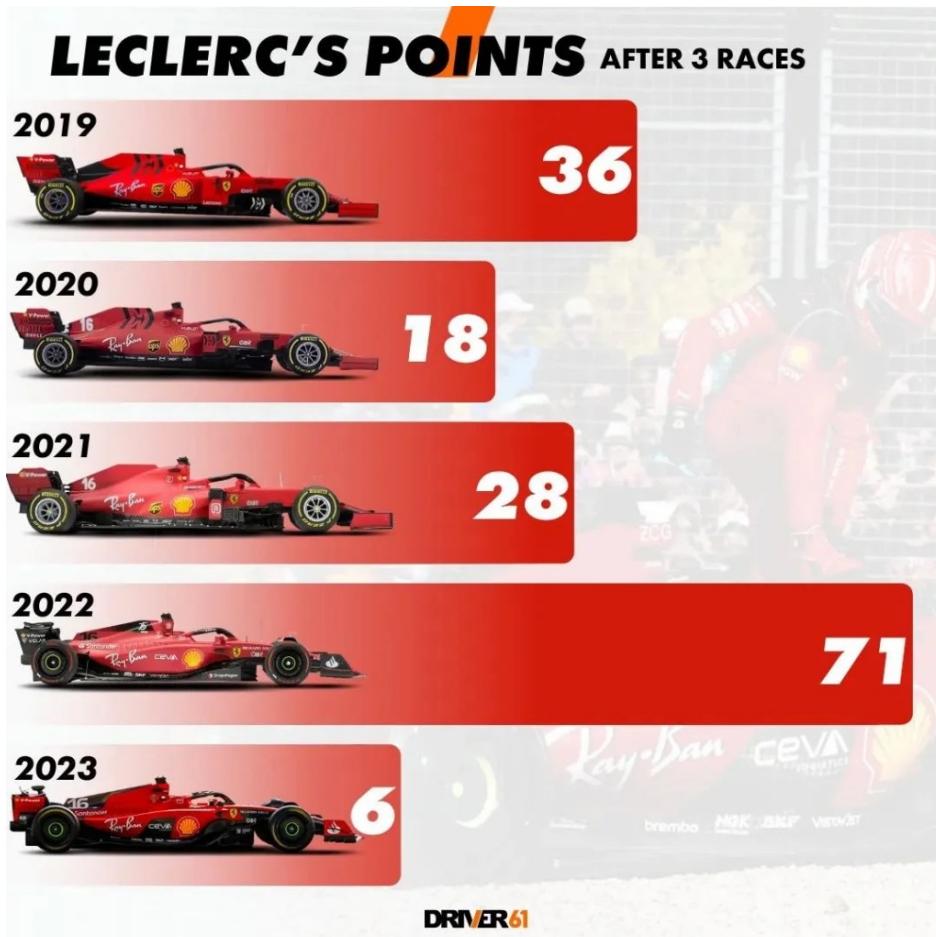
Example 5



- ◆ This chart shows the sources of funding in two financial years for the Canadian National Broadcast Company.
- ◆ Issues
 - "break" in y-Axis – 1213 Mio looks smaller than 490 Mio
 - Television looks equal to Government funding
 - Three bars should not be separate: they show the progressive breakdown of the previous bar



Example 6



- ◆ This chart shows the points obtained by Charles Leclerc after three races in different years.
- ◆ Issues
 - Length of bars seem non-proportional to number
 - However, it is proportional when considered starting at the right-end of the cars!

Source: <https://www.codeconquest.com/blog/12-bad-data-visualization-examples-explained/>



Example 7

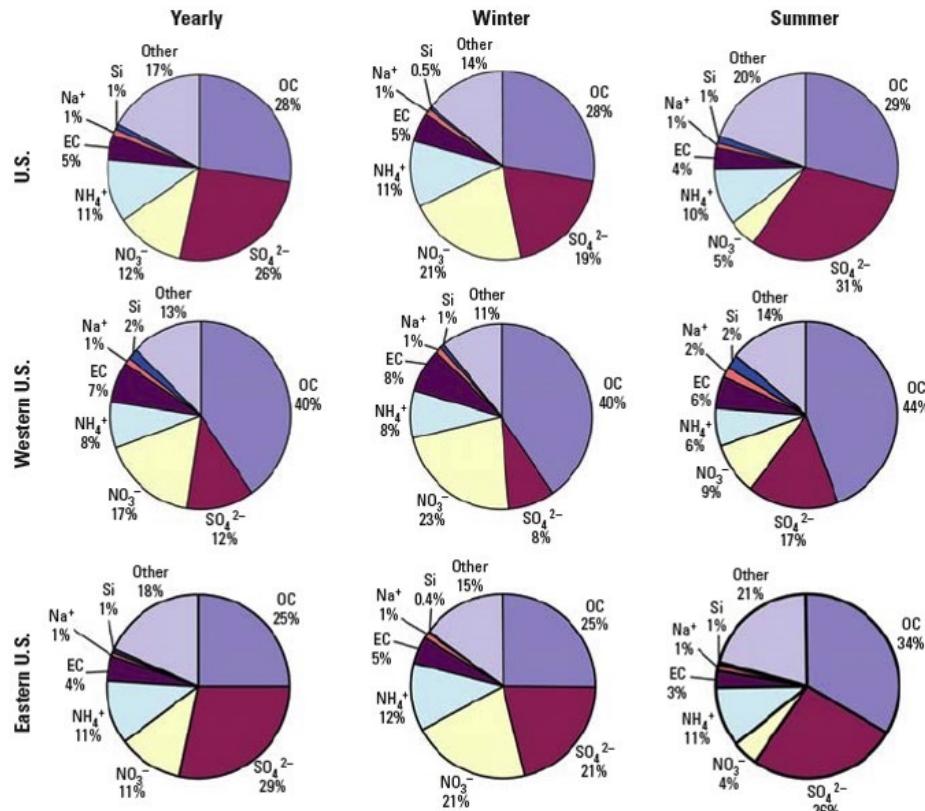
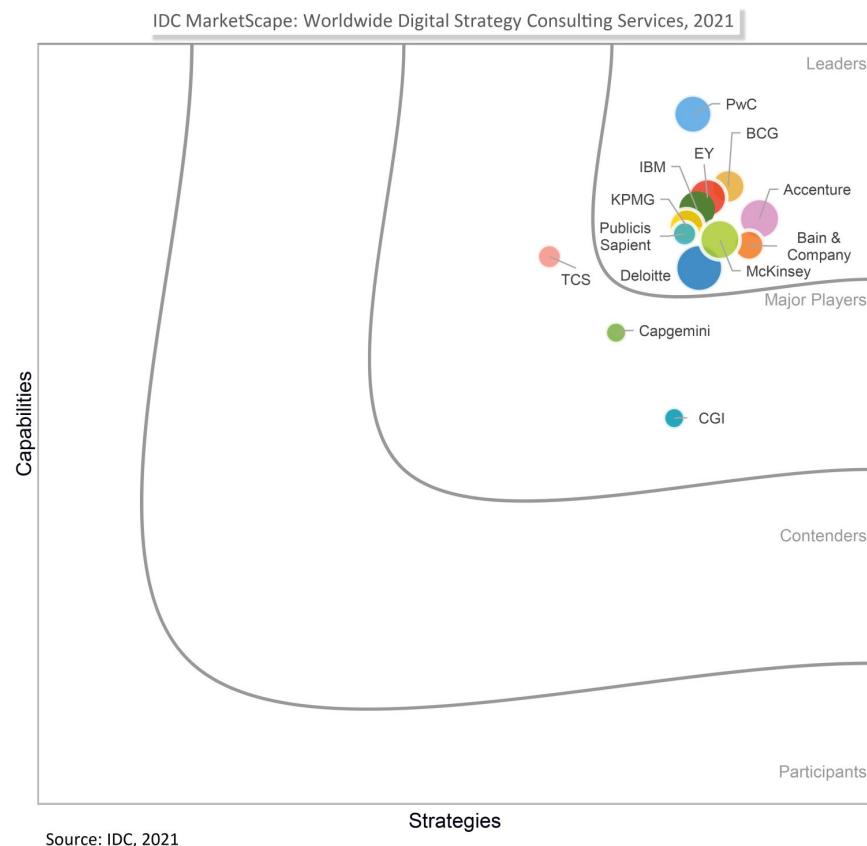


Figure 3. Percent of PM_{2.5} composition by component for yearly, winter, and summer averages, by region.

- ◆ This chart shows the percentages different PM_{2.5} (fine particulate matter = “Feinstaub”) in different areas.
- ◆ Issues
 - Too much information, overloading the viewer!



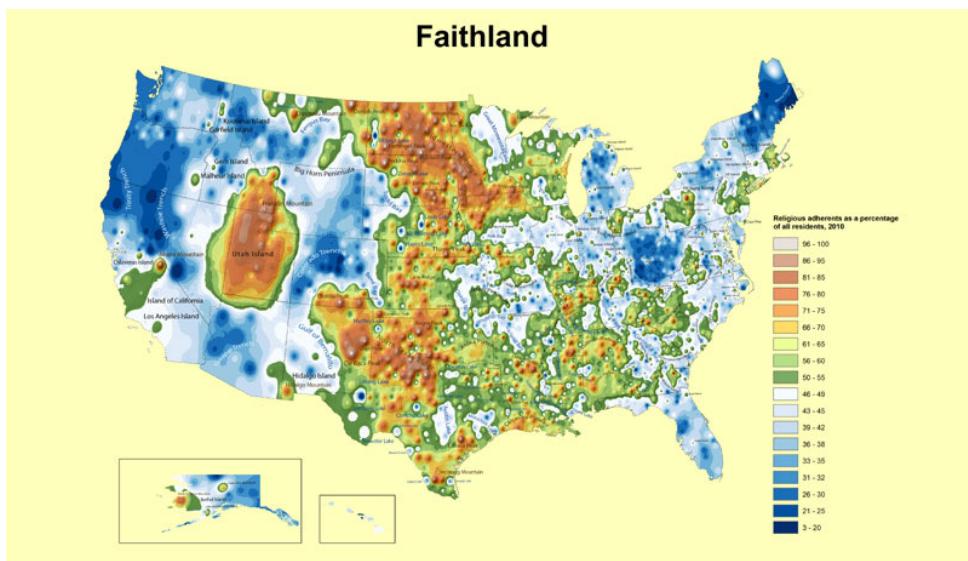
Example 8



- ◆ Scatter plot comparing different digital strategy consulting services..
- ◆ Issues
 - Tell's us nothing: when everyone is a leader, no one is a leader



Example 9



- ◆ Map shows religious adherents as percentage of residents.
- ◆ Issues
 - Very unconventional colour choice.



Visualization

The Dark Side Of Vis

1. Data Ethics
2. From bad visualizations...
3. ...to actively misleading visualizations



What are misleading visualizations?

- ◆ Characterized by **Distortion** and **Manipulation**
- ◆ **Intentional** or **Unintentional**
- ◆ Affects **Credibility** and **Integrity** of data and message
- ◆ **Causes** for misleading visualizations
 - using inappropriate or misleading visual elements (e.g., scales, axes, ranges, proportions), that alter or exaggerate the data representation
 - using visual techniques (e.g., framing, highlighting, hiding), that influence/bias the data interpretation
 - omitting, cherry-picking, or fabricating data



Common Techniques of Misleading Visualizations

I. Y-Axis Manipulated

II. X-Axis Manipulated

III. Arbitrary dual Y-Axis

IV. Improper Scaling of Icons

V. Unusual Colouring

VI. 3-dim Distortions/Occlusions

X. Correlation and Causation



VII. Cherry-Picking

VIII. Cumulative vs. Annual Data

IX. Out-Of-Context Data

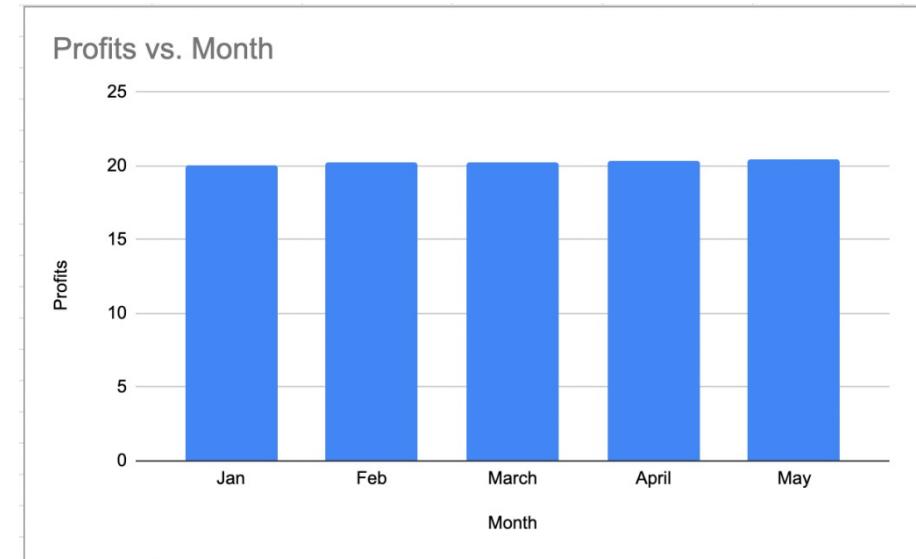
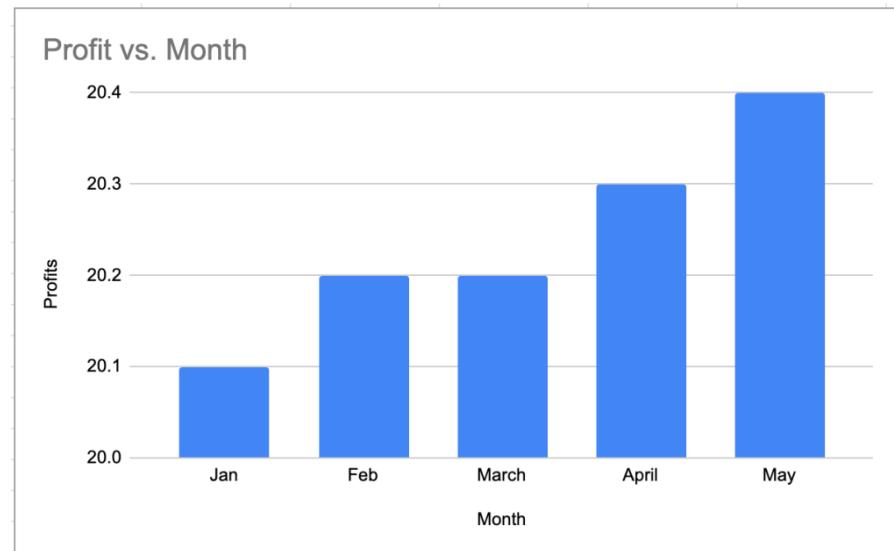
XI. Data not adding up

XII. Wrong Chart Type



I. Y-Axis Manipulated

- ◆ **Y-Axis Manipulated:** not starting at 0, skipping numbers, extended, linear vs. log

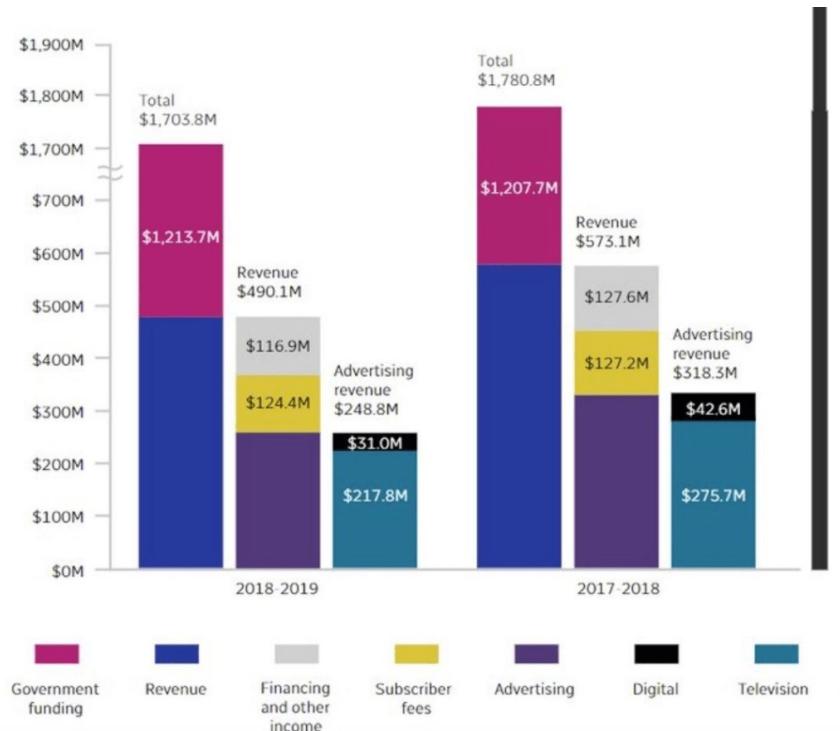


Source: <https://blog.coupler.io/misleading-data-visualization-examples/>



I. Y-Axis Manipulated

- ◆ **Y-Axis Manipulated:** not starting at 0, **skipping numbers**, extended, linear vs. log

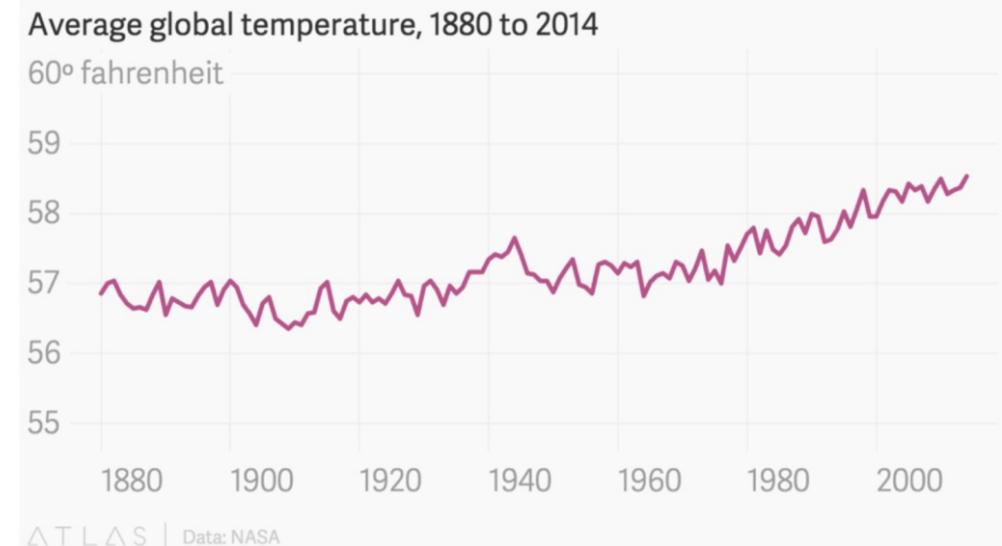
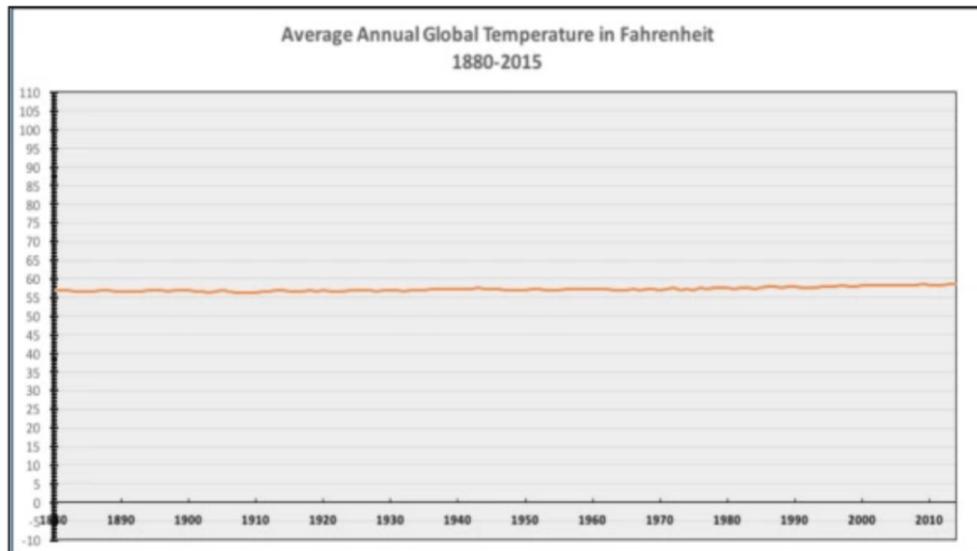


Source: <https://www.codeconquest.com/blog/12-bad-data-visualization-examples-explained/>



I. Y-Axis Manipulated

- ◆ **Y-Axis Manipulated:** not starting at 0, skipping numbers, **extended**, linear vs. log

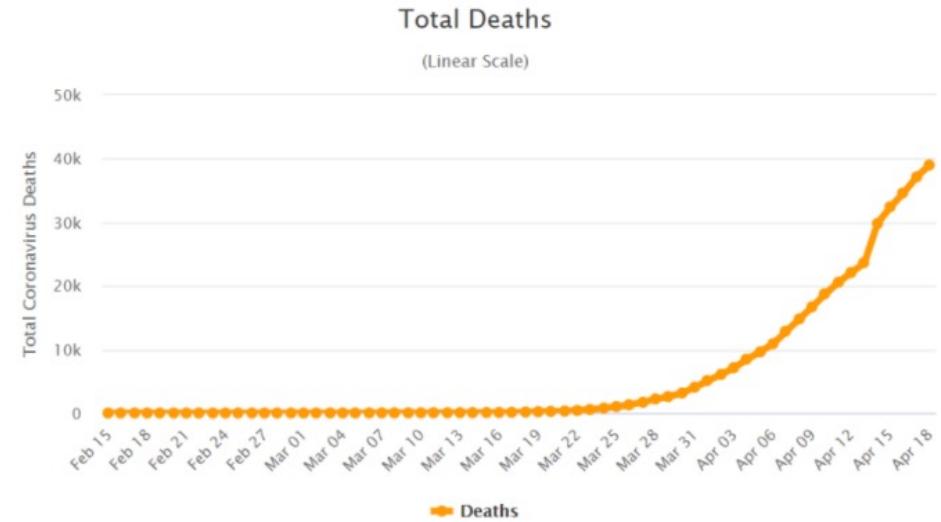
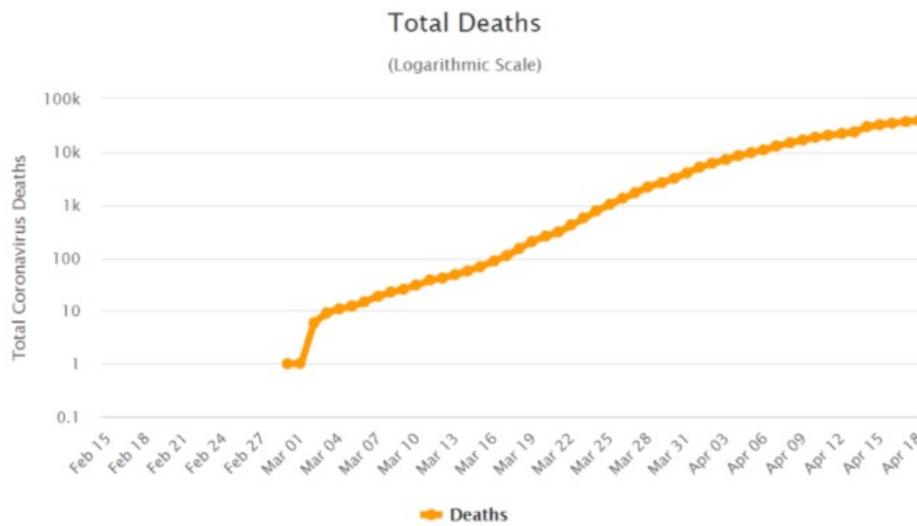


Source: <https://blog.coupler.io/misleading-data-visualization-examples/>



I. Y-Axis Manipulated

- ◆ **Y-Axis Manipulated:** not starting at 0, skipping numbers, extended, linear vs. log



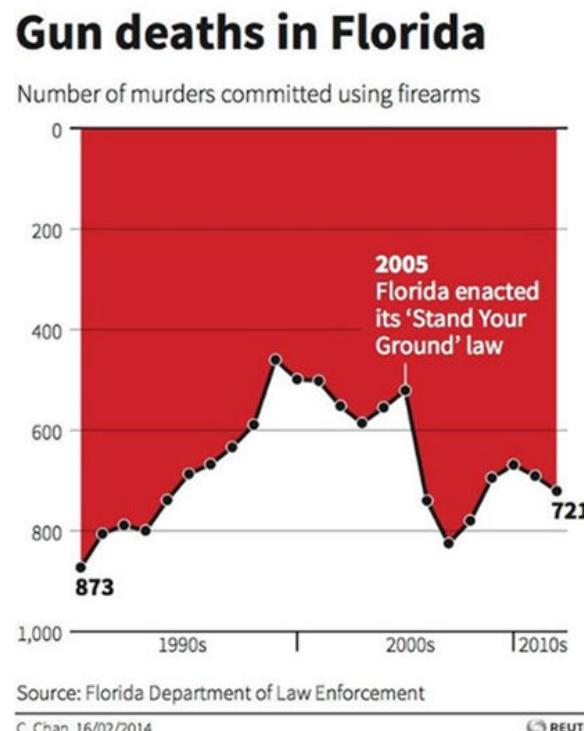
Source: <https://blogs.lse.ac.uk/covid19/2020/05/19/the-public-doesnt-understand-logarithmic-graphs-often-used-to-portray-covid-19/>





I. Y-Axis Manipulated

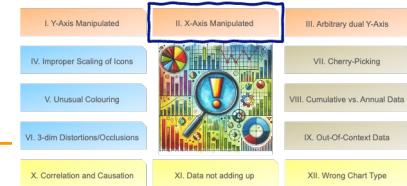
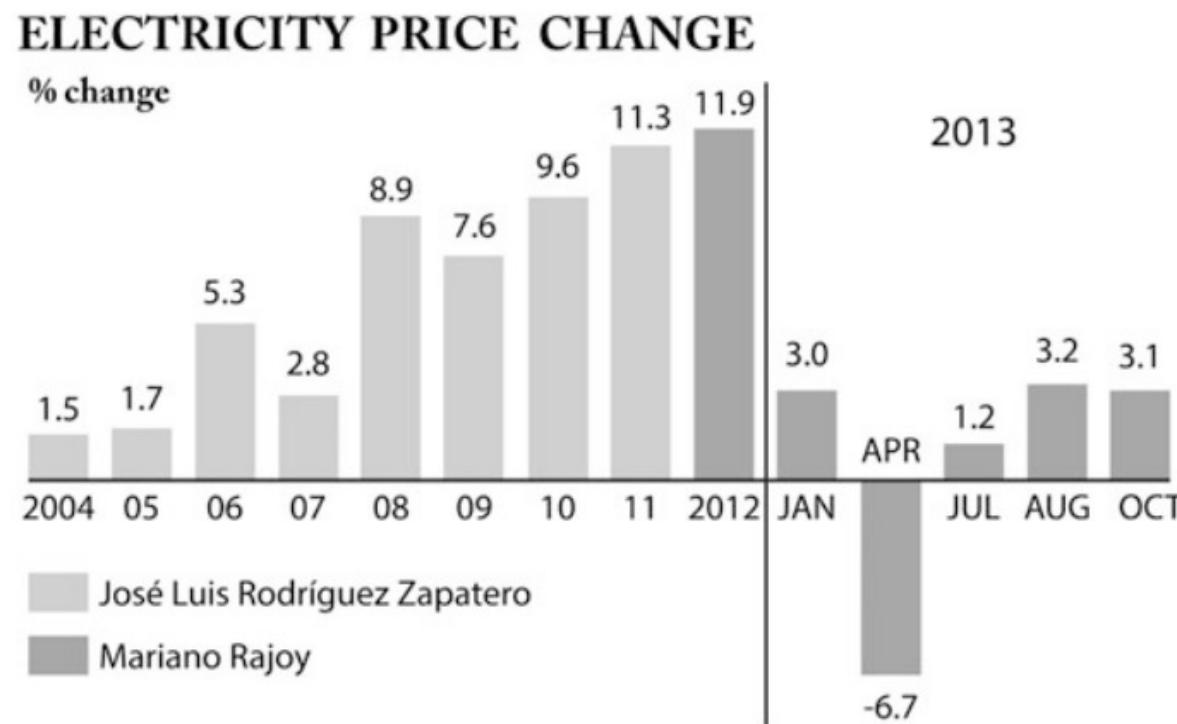
- ◆ **Y-Axis Manipulated:** bonus: upside-down axis!





II. X-Axis Manipulated

◆ X-Axis Manipulated



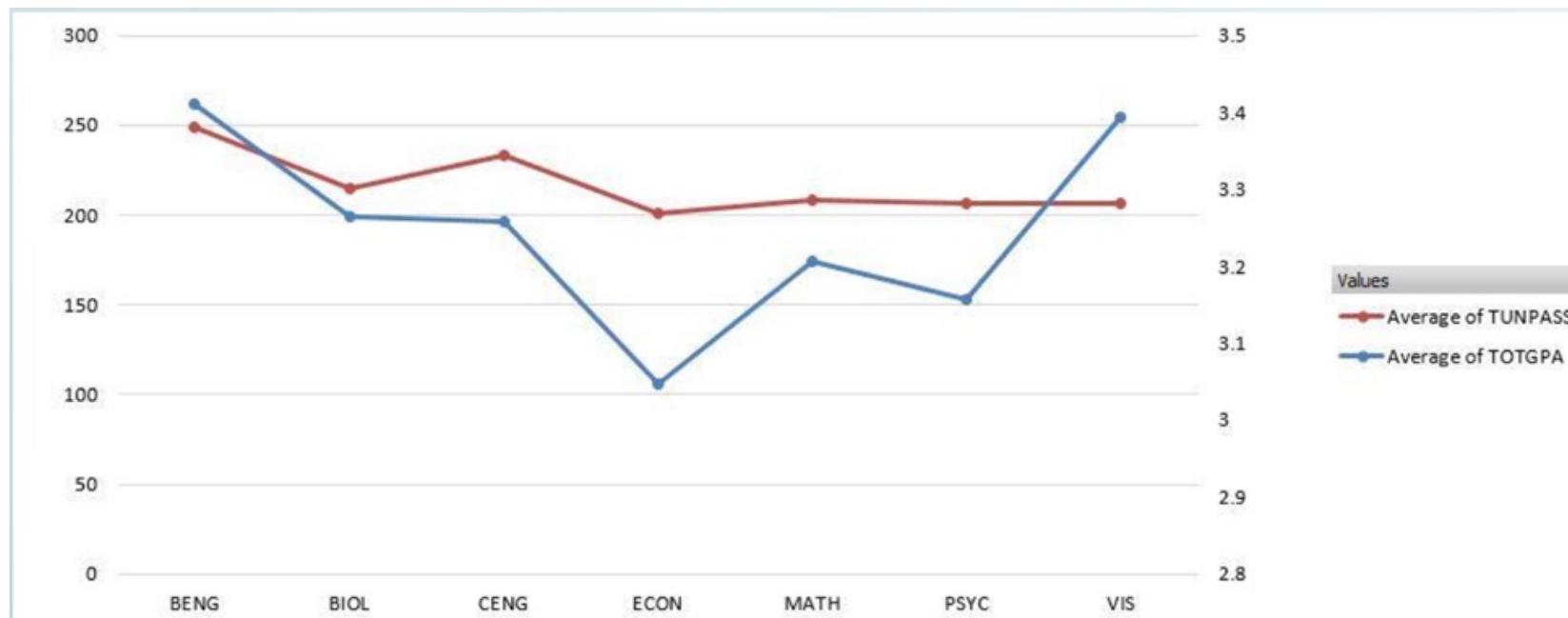


III. Arbitrary dual Y-Axis



Arbitrary dual Y-Axis

- The scales for these axes can be set arbitrarily, which may (inadvertently or deliberately) mislead readers about the relationship between two data series.



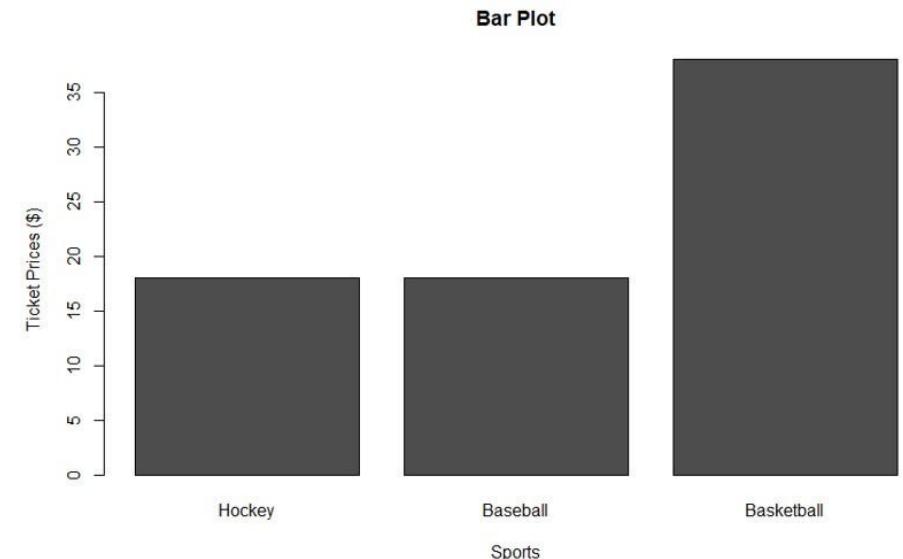
Source: <https://www.eisquare.co.uk/blogs/ethical-data-visualisation>



IV. Improper Scaling of Icons (Example 1)



- ◆ Improper Scaling of Icons
 - The sizes of images used in Pictographs differ for the categories being graphed.



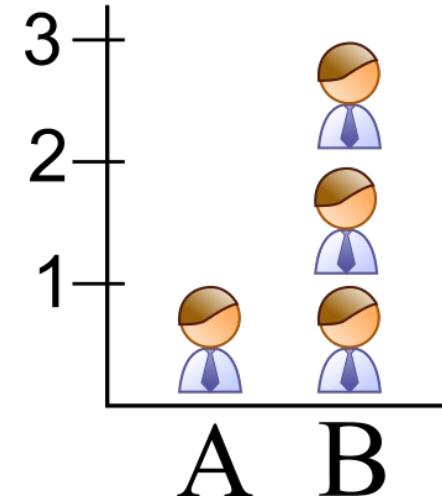
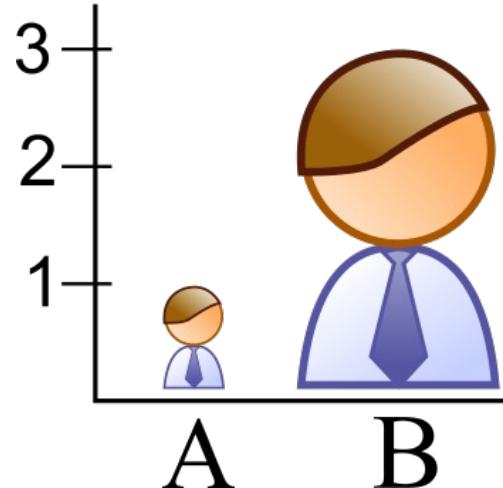
Source: https://medium.com/@Ana_kin/graphs-gone-wrong-misleading-data-visualizations-d4805d1c4700



IV. Improper Scaling of Icons (Example 2)

I. Y-Axis Manipulated	II. X-Axis Manipulated	III. Arbitrary dual Y-Axis
IV. Improper Scaling of Icons	V. Unusual Colouring	VII. Cherry-Picking
VI. 3-dim Distortions/Occlusions	VIII. Cumulative vs. Annual Data	VIII. Cumulative vs. Annual Data
X. Correlation and Causation	XI. Data not adding up	XII. Wrong Chart Type

- Improper Scaling of Icons
 - The sizes of images used in Pictographs differ for the categories being graphed.



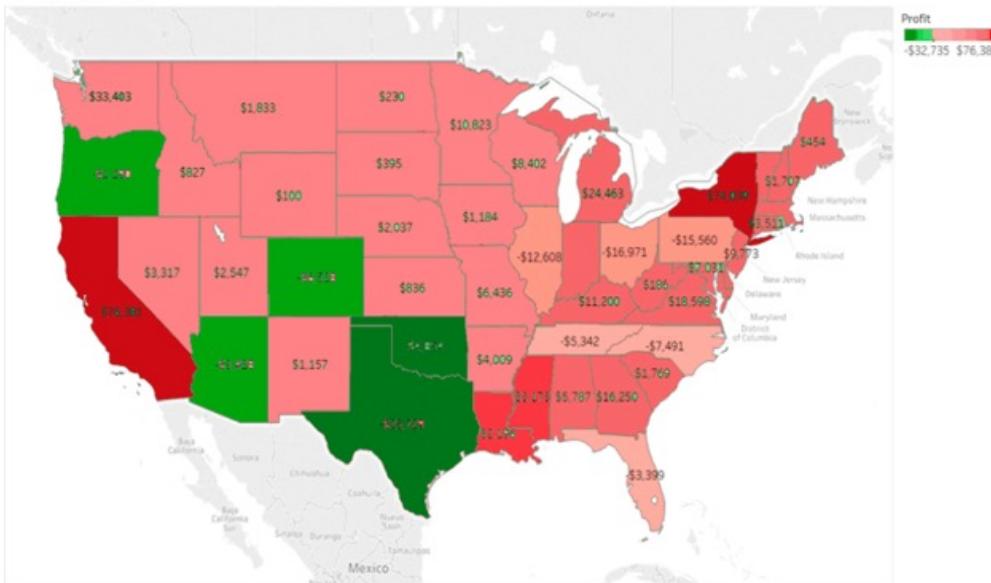


V. Unusual Colouring

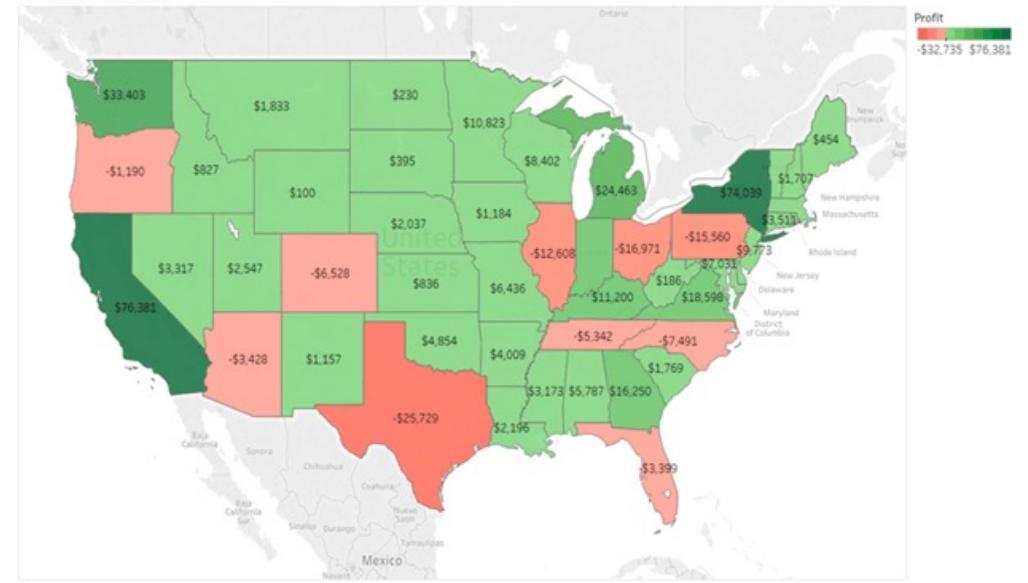


- ◆ Unusual Colouring: going against the norm

50 states profits



50 states profits

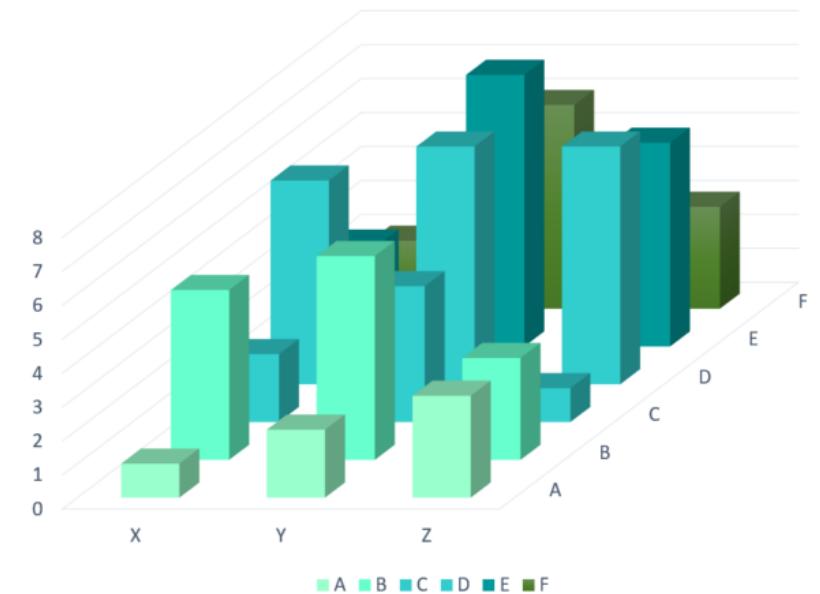
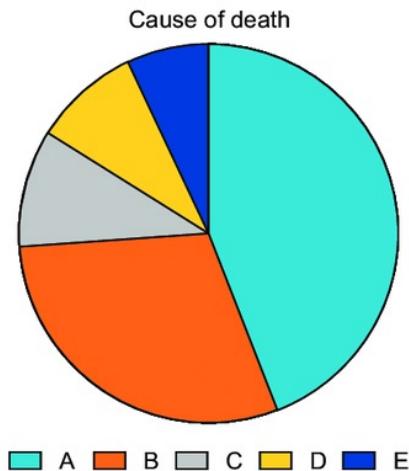
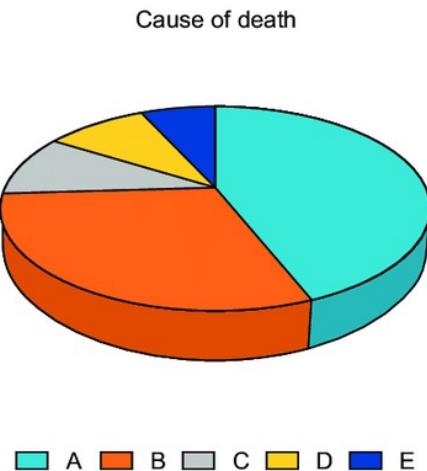




VI. 3-dim Distortions/Occlusions



◆ 3-dim Distortions/Occlusions



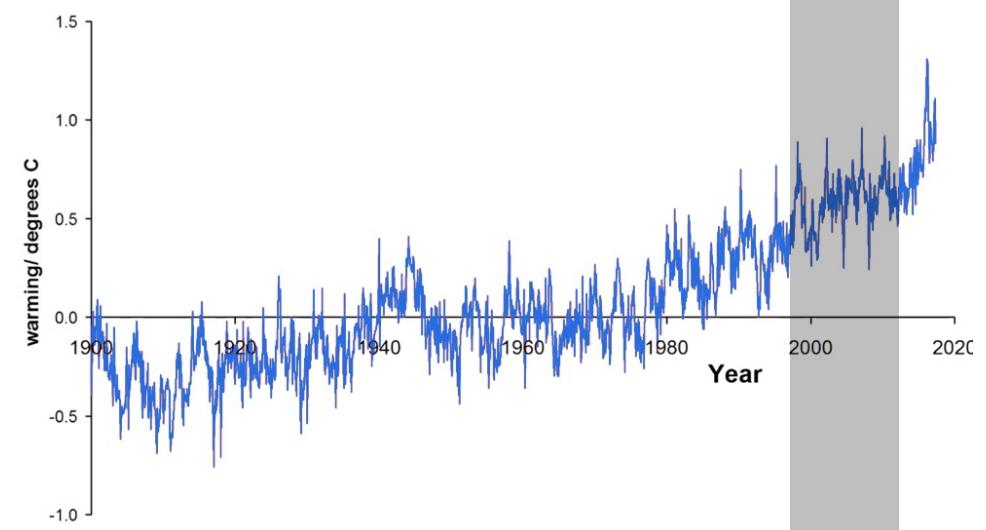
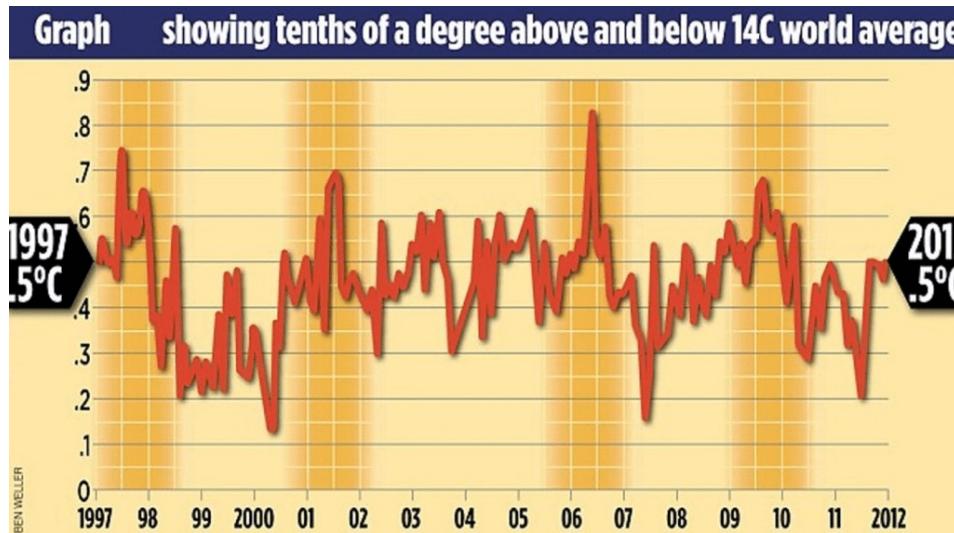
Source: <https://blog.coupler.io/misleading-data-visualization-examples/> , <https://towardsdatascience.com/when-to-use-3d-in-data-visualisation-f3739e320da7>



VII. Cherry-Picking



- ◆ Cherry-Picking
 - selecting certain pieces of information to back up an argument rather than the whole body of evidence



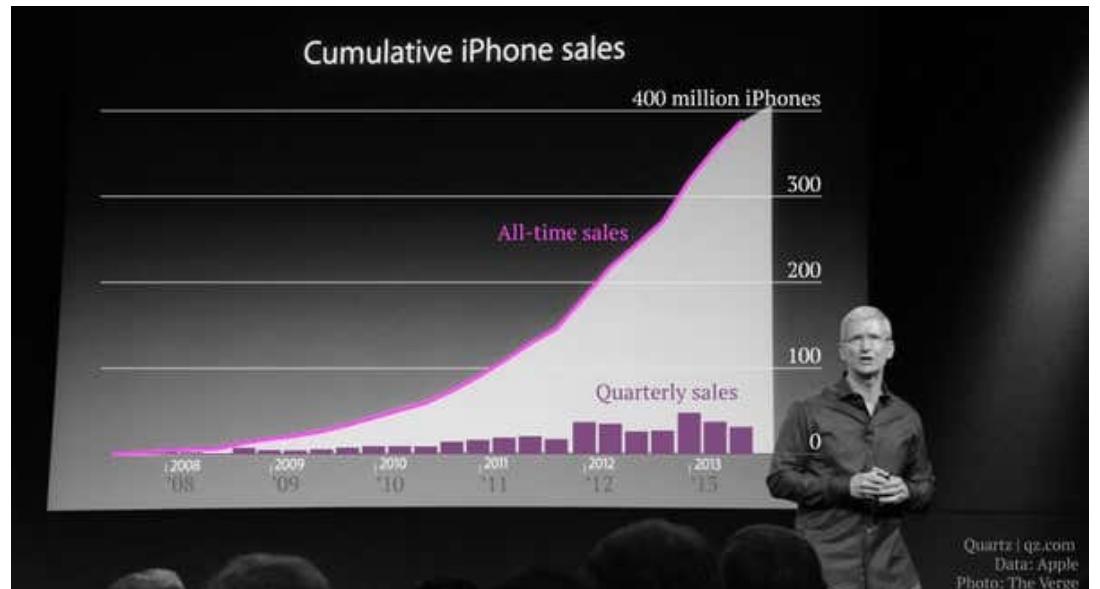
Source: <https://blog.coupler.io/misleading-data-visualization-examples/>



VIII. Cumulative vs. Annual Data



- ◆ Cumulative vs. Annual Data
 - Hide actual data





IX. Out-Of-Context Data

Out-Of-Context Data

- Compare apples to oranges
- Very challenging to spot!

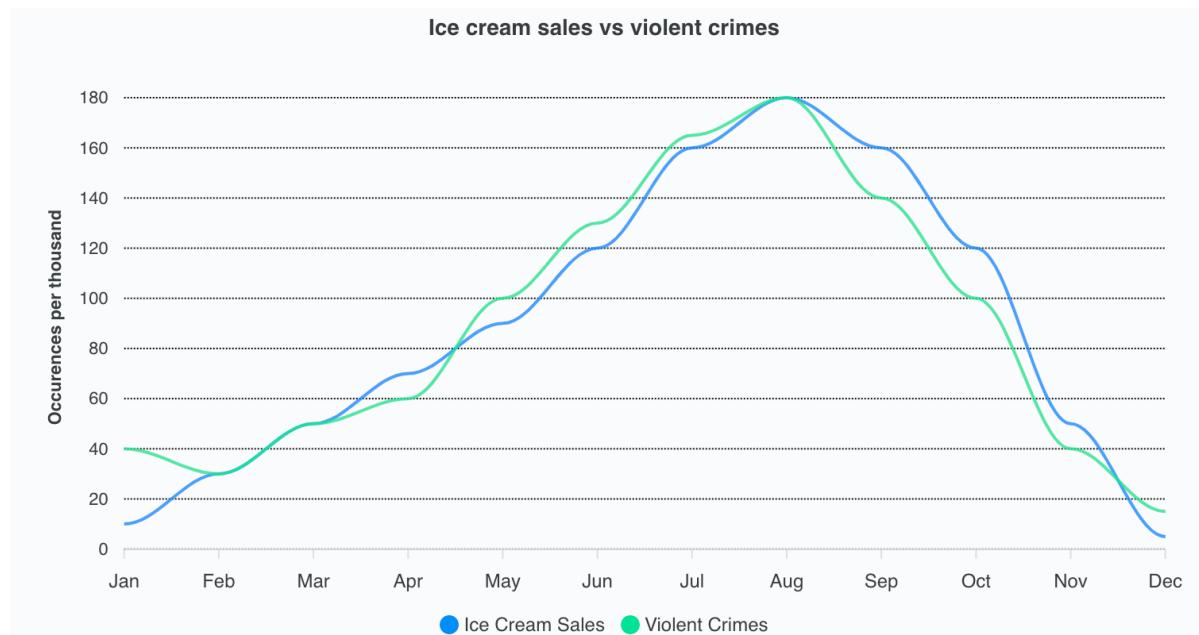


Sources: <https://www.luzmo.com/blog/bad-data-visualization>, https://medium.com/@Ana_kin/graphs-gone-wrong-misleading-data-visualizations-d4805d1c4700



X. Correlation and Causation

- ◆ Correlation and Causation
 - Implying a causation by showing a correlation



Source: <https://wpdatatables.com/misleading-data-visualization-examples>

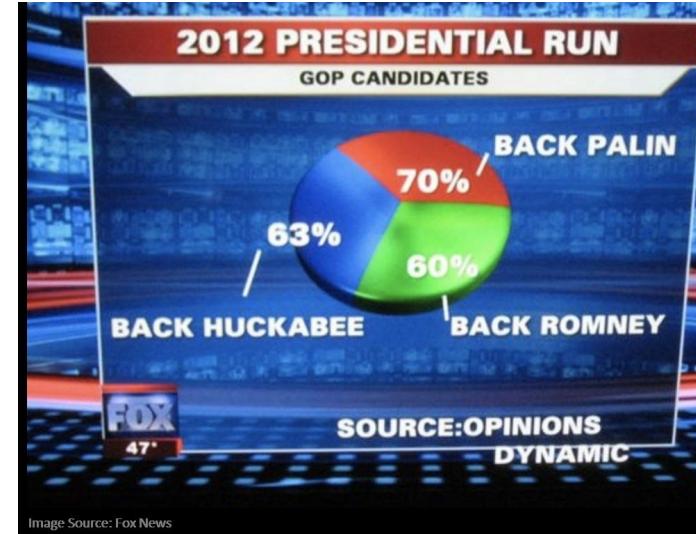
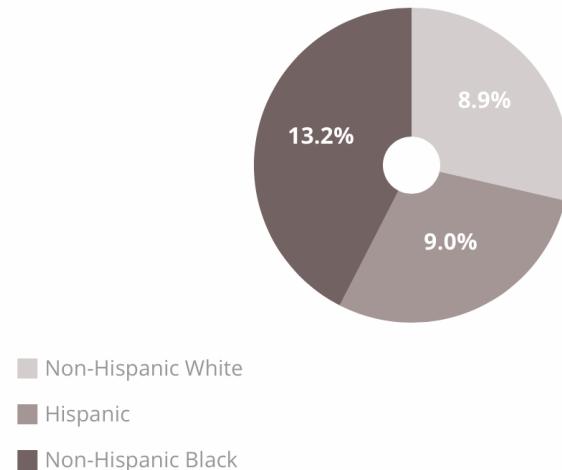


XI. Data not adding up

- ◆ Data not adding up
 - In Charts showing “percentages of a whole” (e.g. Bar Charts, Donut Charts)
 - Numbers should always add up to 100%
- ➔ Do not use such Chart for e.g. survey questions where multiple answers are possible



PRETERM BIRTH BY RACE & ETHNICITY





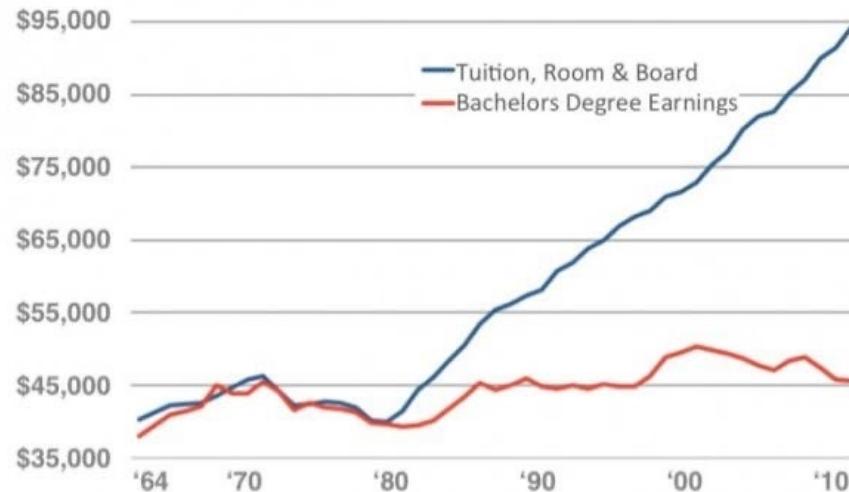
XII. Wrong Chart Type

Wrong Chart Type

- Issues caused by using a non-optimal chart
- (a bit of a “catch-all-remaining-issues”)

The diminishing financial return of higher education

Costs of 4-yr degree vs. earnings of 4-yr degree



Source: Source: U.S. Census Data & NCES Table 345.

Notes: All figures have been adjusted to 2010 dollars using the Consumer Price Index from the BLS.

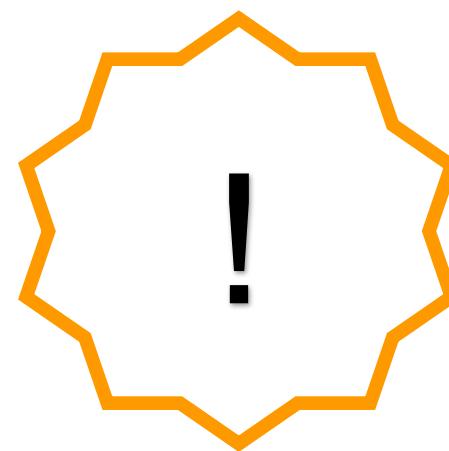




Your Turn !

Exercise 1

Identifying Techniques in Misleading Visualizations





Key Takeaways

- ◆ **Data Ethics**
 - Ethical and moral implications of collecting, sharing, and using data
- ◆ **Misleading Visualizations:** Distortion and Manipulation
 - Intentional or Unintentional
 - Affect Credibility and Integrity of data and message
- ◆ **Common Techniques of Misleading Visualizations**

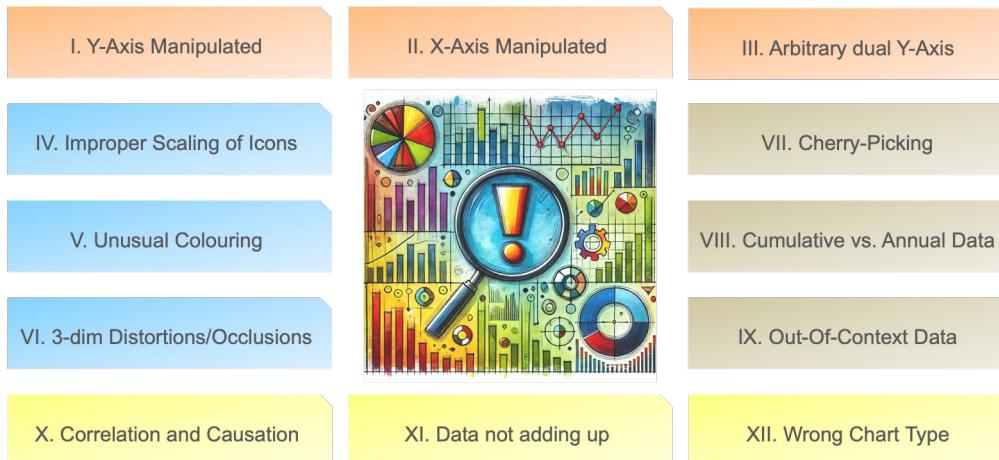


Photo by Dragonfly Ave on Unsplash