

Getting Started with Data Visualization

Oluwakemi Ola, PhD
Associate Professor of Teaching
Computer Science Department

<https://tinyurl.com/ashesi-viz>



**PRACTICE
ISN'T THE THING YOU DO
ONCE YOU'RE GOOD.
IT'S THE THING YOU DO
THAT MAKES YOU GOOD.**

Malcolm Gladwell

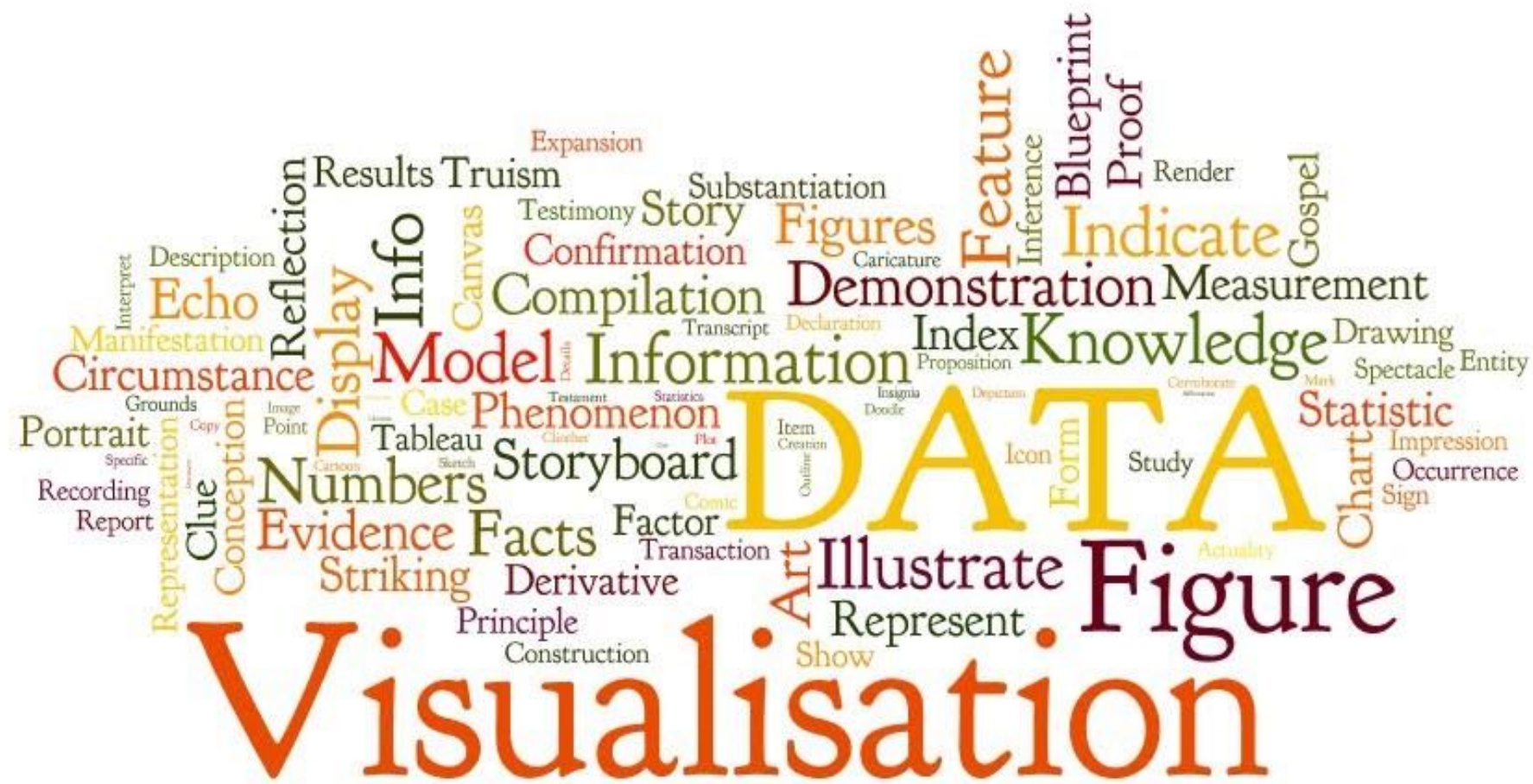
Let's get to coding.

<https://tinyurl.com/ashesi-viz>

This is the repo for all the files for the workshop

Today's file is Ashesi_WK3S.ipynb





Approaches to Supporting the Discourse with Large Datasets

- Faceting – split data into multiple views
- Interaction – manipulate the external view
- Reduce
 - embed additional data into smaller glyphs
 - reduce amount of attributes visualized (e.g. maps reduces from 3D to 2D)



All Spending

Types of Spending

Changes

Department Totals

How \$3.7 Trillion Is Spent

Mr. Obama's budget proposal includes \$3.7 trillion in spending in 2013, and forecasts a \$901 billion deficit.

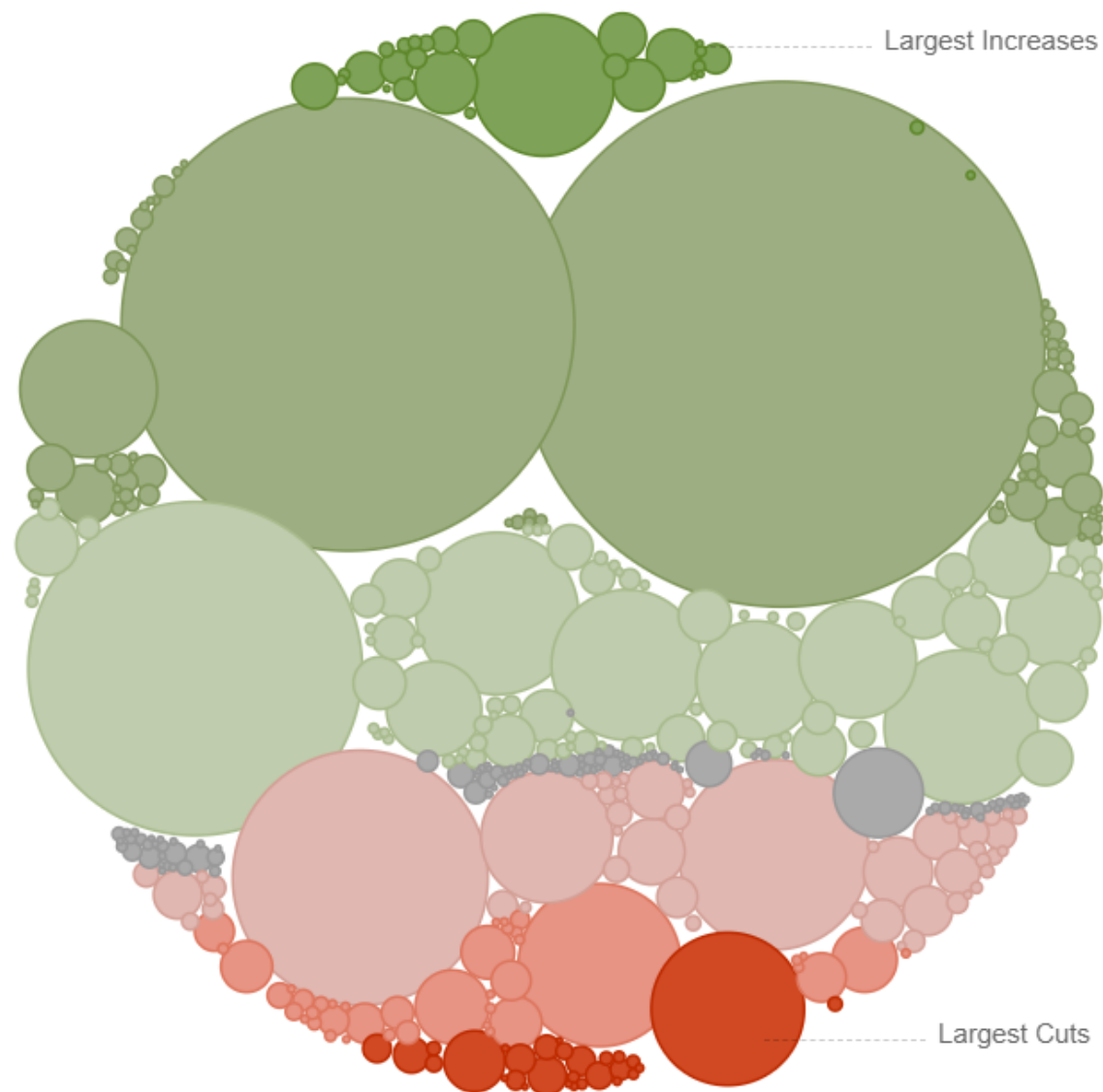
Circles are sized according to the proposed spending.



Color shows amount of cut or increase from 2012.

Legend for color showing amount of cut or increase from 2012:

-25% -5% 0 +5% +25%



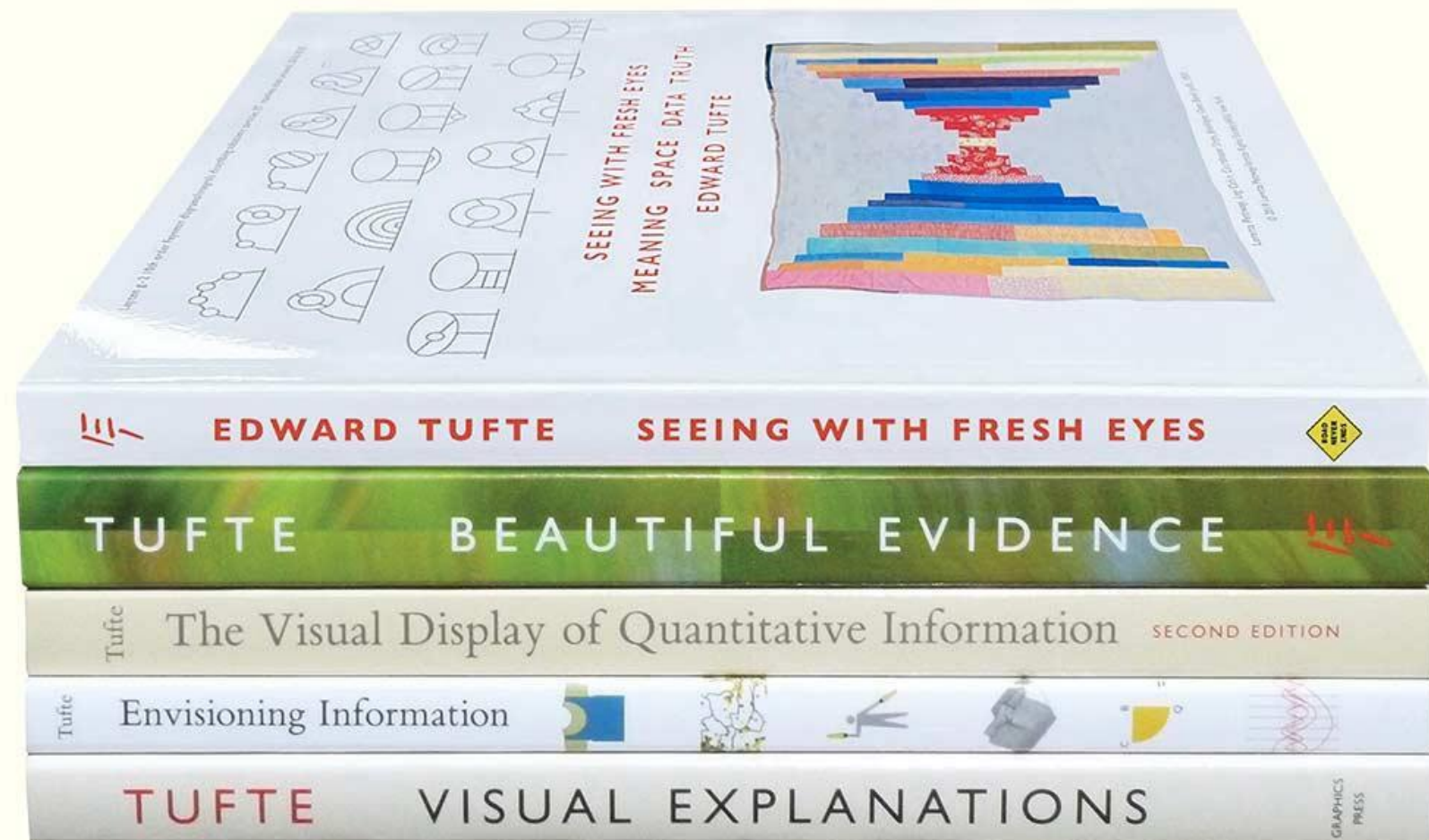
Ethical Considerations

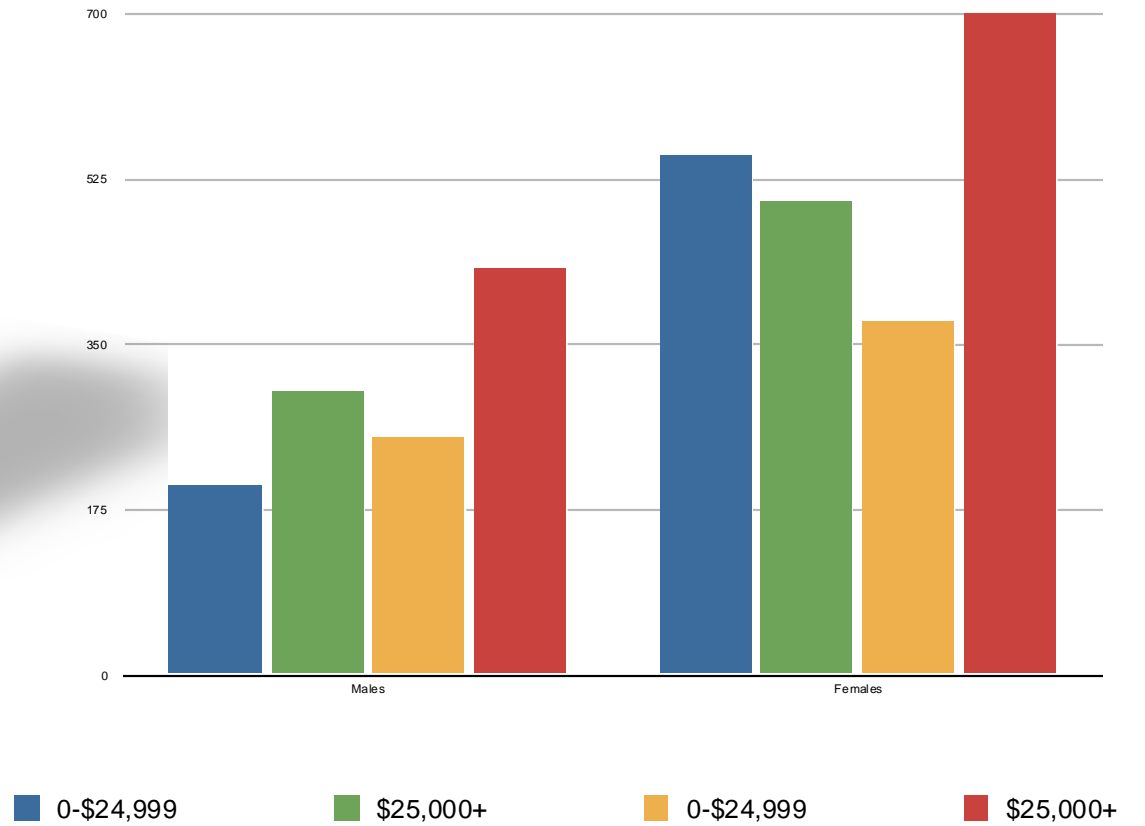
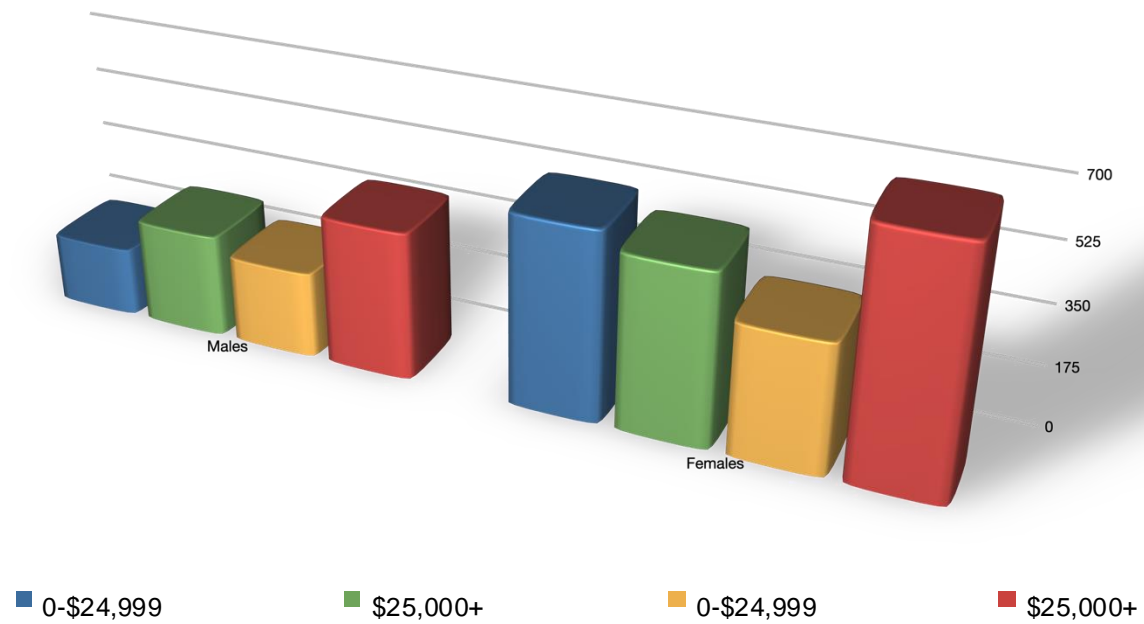
Visualizations are not **fact**

A visualization's accuracy is dependent upon

- The quality of the data used
- The manner in which the data is conveyed (i.e., its representation)
- The objectivity of its creator

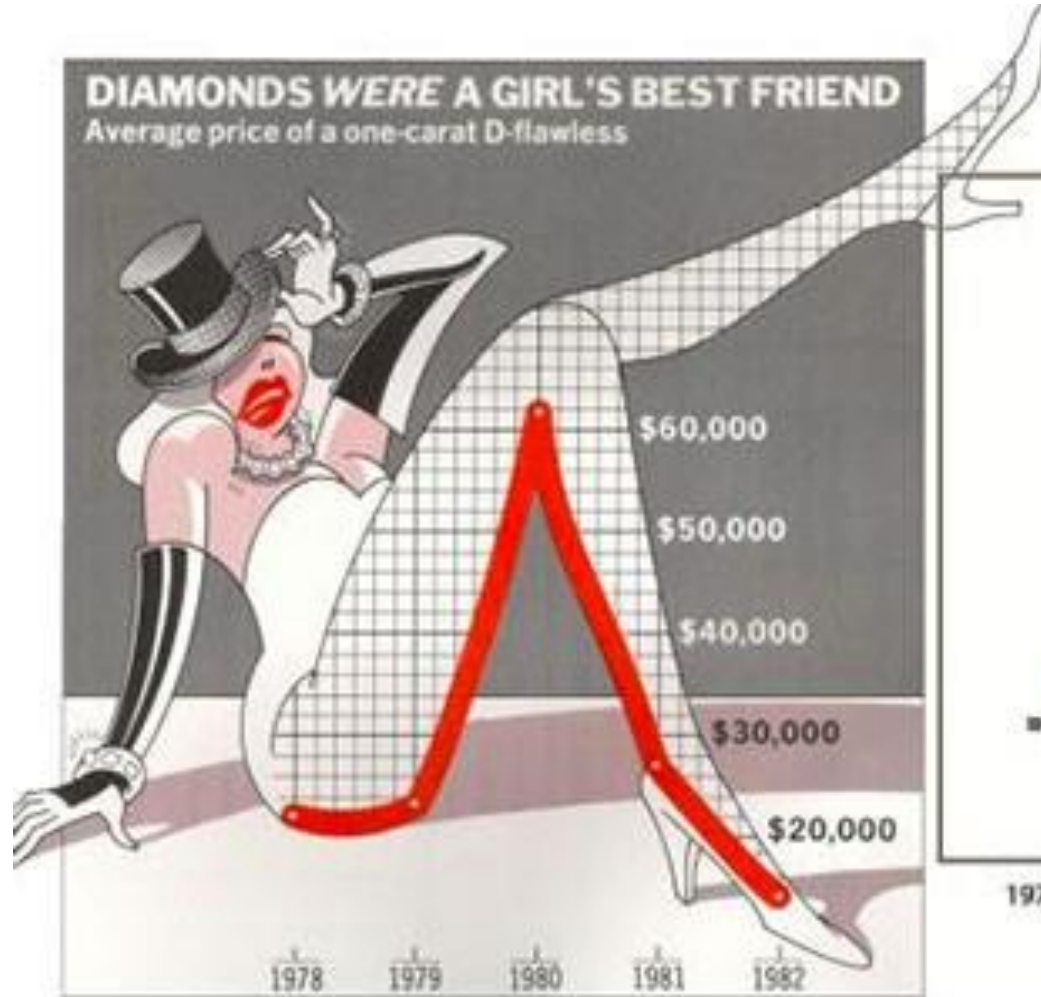
Edward Tufte



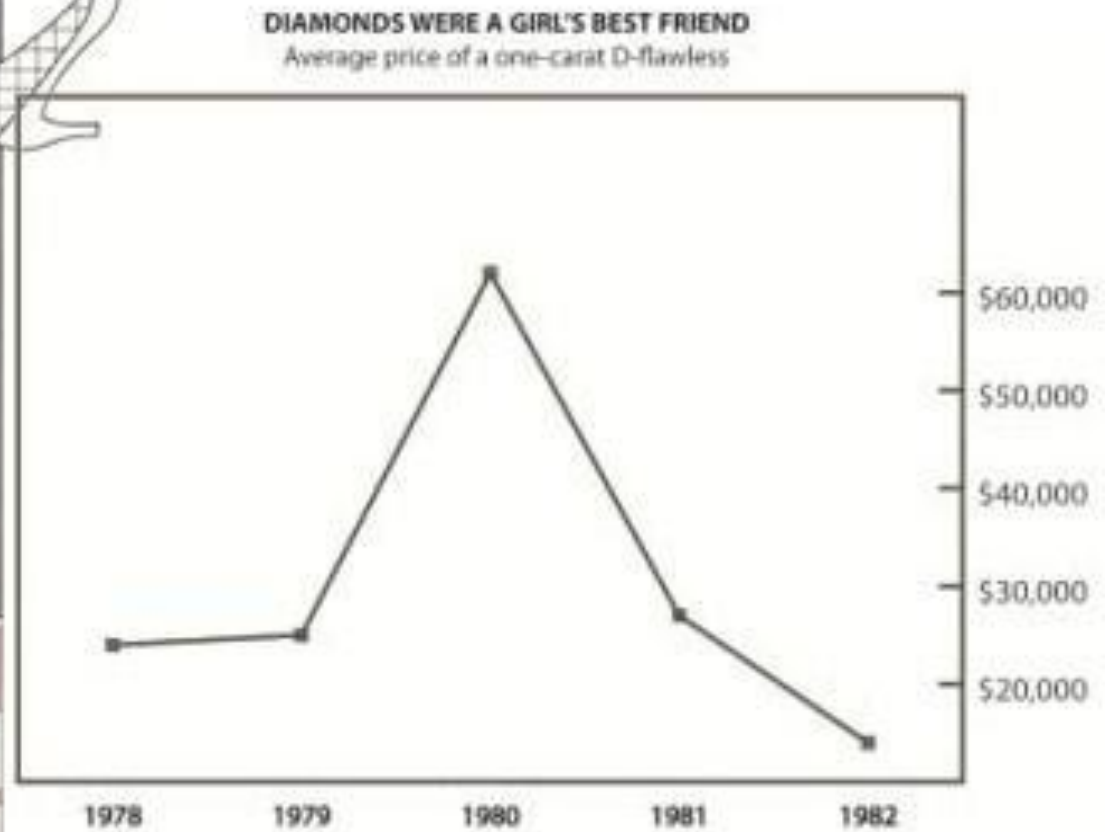


Which is better? Why?

A



B



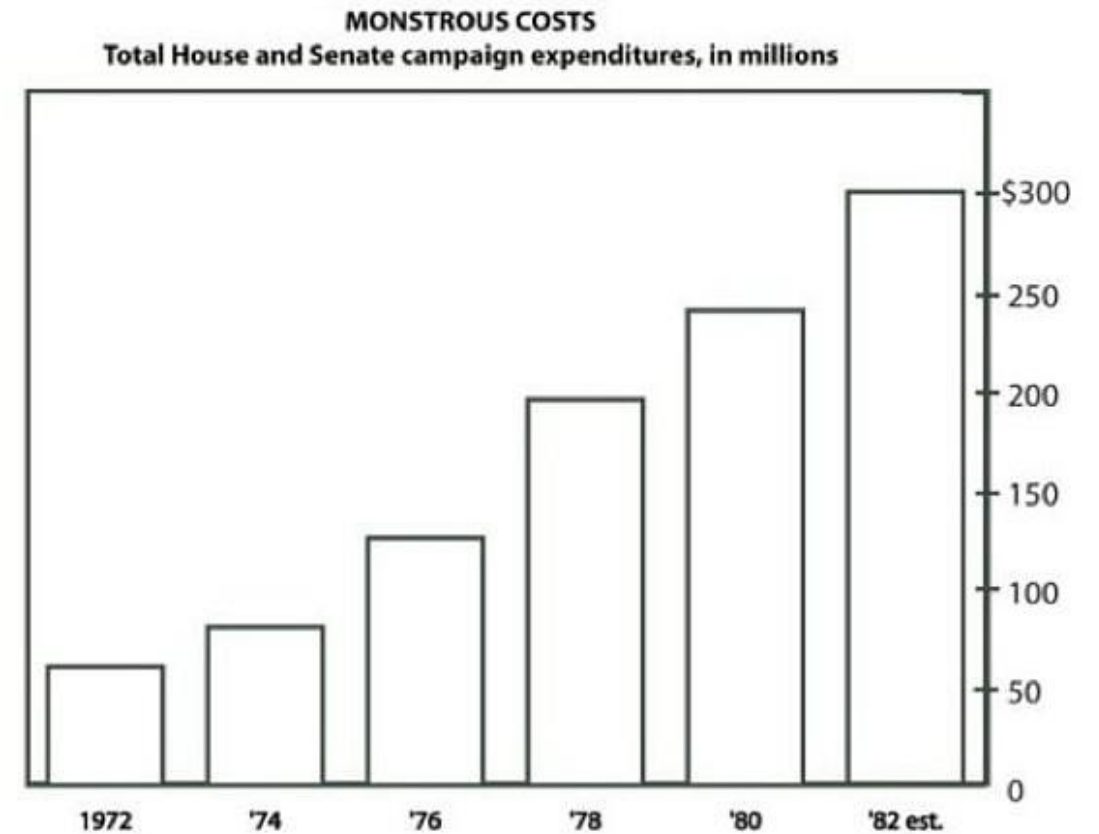
[Bateman et al. 2010]

Which is better? Why?

A



B



Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts

Scott Bateman, Regan L. Mandryk, Carl Gutwin,
Aaron Genest, David McDine, Christopher Brooks

Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
scott.bateman@usask.ca, regan@cs.usask.ca, gutwin@cs.usask.ca,
aaron.genest@usask.ca, dam085@mail.usask.ca, cab938@mail.usask.ca

ABSTRACT

Guidelines for designing information charts often state that the presentation should reduce ‘chart junk’ – visual embellishments that are not essential to understanding the data. In contrast, some popular chart designers wrap the presented data in detailed and elaborate imagery, raising the questions of whether this imagery is really as detrimental to understanding as has been proposed, and whether the visual embellishment may have other benefits. To investigate these issues, we conducted an experiment that compared embellished charts with plain ones, and measured both interpretation accuracy and long-term recall. We found that people’s accuracy in describing the embellished charts was no worse than for plain charts, and that their recall after a two-to-three-week gap was significantly better. Although we are cautious about recommending that all charts be produced in this style, our results question some of the premises of the minimalist approach to chart design.

Author Keywords

Charts, information visualization, imagery, memorability.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

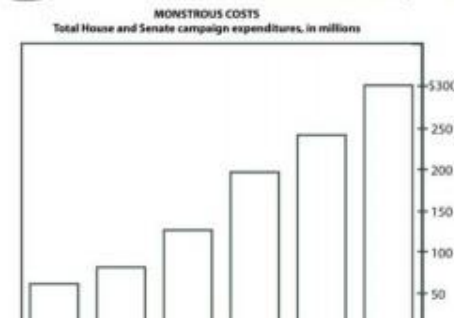
Design, Human Factors

INTRODUCTION

Many experts in the area of chart design, such as Edward Tufte, criticize the inclusion of visual embellishment in charts and graphs; their guidelines for good chart design often suggest that the addition of *chart junk*, decorations and other kinds of non-essential imagery, to a chart can make interpretation more difficult and can distract readers from the data [22]. This *minimalist* perspective advocates

data-ink – or the ink in the chart used to represent data.

Despite these minimalist guidelines, many designers include a wide variety of visual embellishments in their charts, from small decorations to large images and visual backgrounds. One well-known proponent of visual embellishment in charts is the graphic artist Nigel Holmes, whose work regularly incorporates strong visual imagery into the fabric of the chart [7] (e.g., Figure 1).



Experimental results

1. No difference for **interpretation accuracy**
2. No difference in **recall accuracy after a five-minute gap**
3. Significantly **better recall for Holmes charts** of both the chart topic and the details (categories and trend) **after long-term gap** (2-3 weeks).
4. Participants **saw value messages** in the Holmes charts significantly more often than in the plain charts.
5. Participants found the Holmes charts **more attractive, most enjoyed** them, and found that they were **easiest and fastest to remember**.

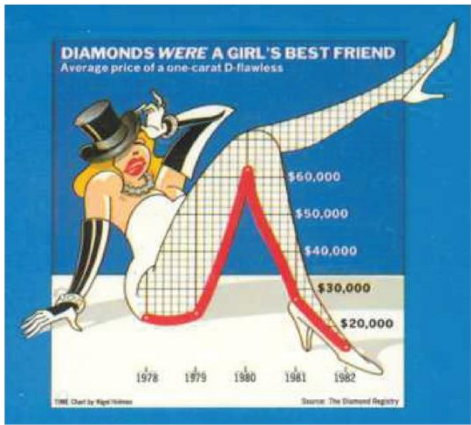
“Chart Embellishments Debate”

Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts



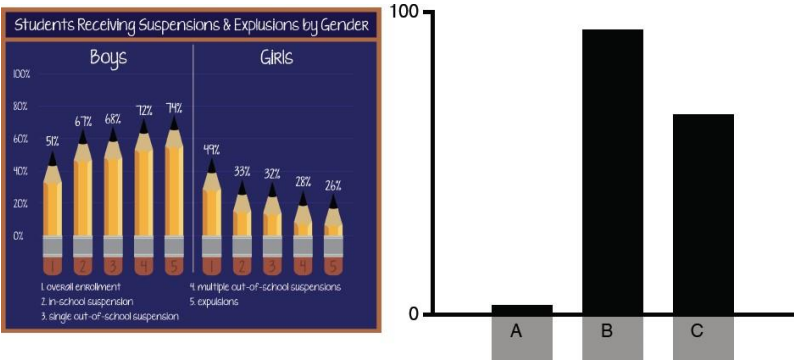
Bateman, et al. (2010)

Benefitting InfoVis with Visual Difficulties



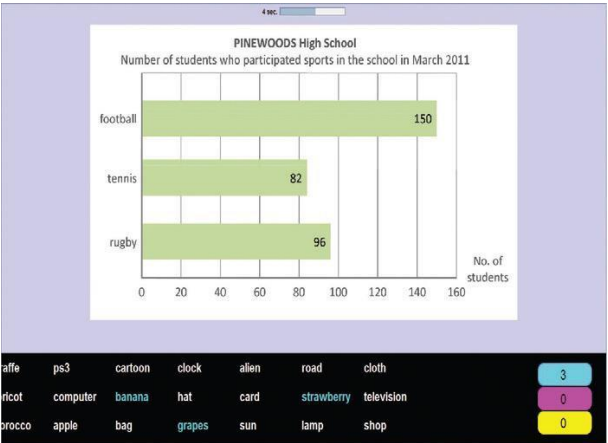
Hullman, et al. (2011)

An Evaluation of the Impact of Visual Embellishments in Bar Charts



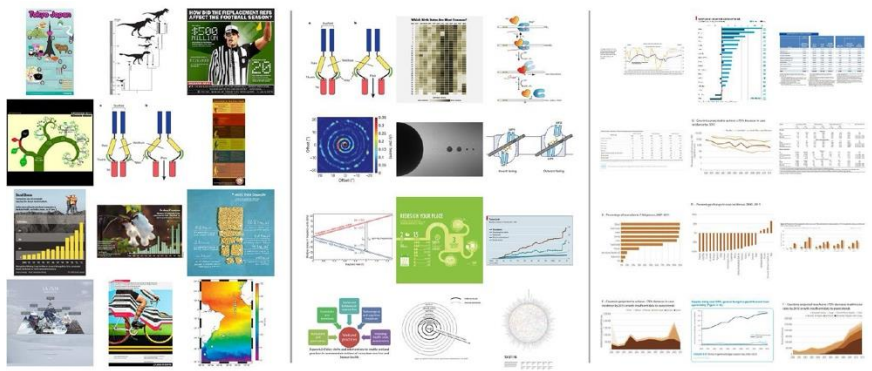
Skau, et al. (2015)

An Empirical Study on Using Visual Embellishments in Visualization



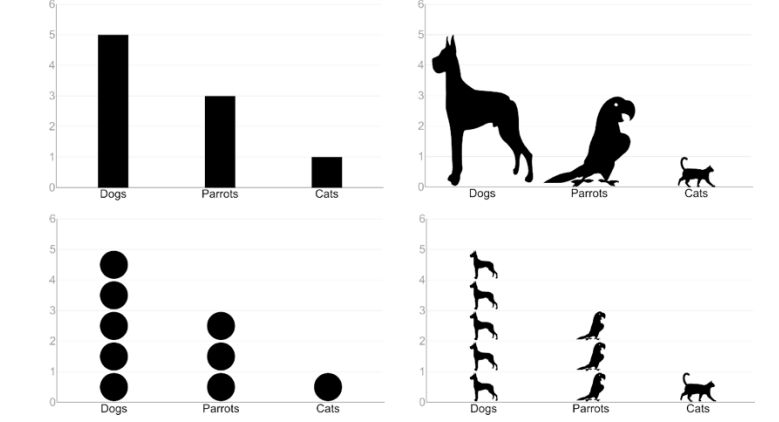
Borgo, et al. (2012)

What makes a visualization memorable?



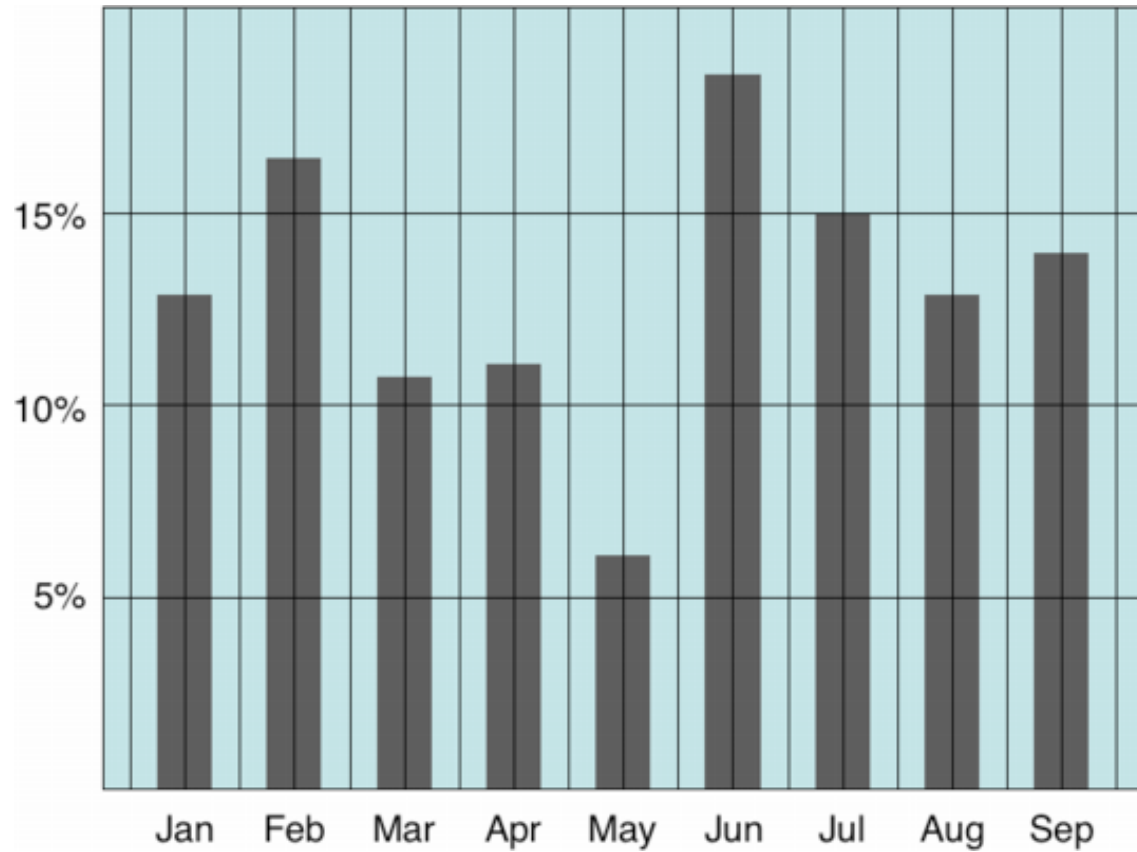
Borkin, et al. (2013)
Borkin, et al. (2015)

ISOTYPE Visualization – Working Memory, Performance, and Engagement with Pictographs



Haroz, et al. (2015)¹₃

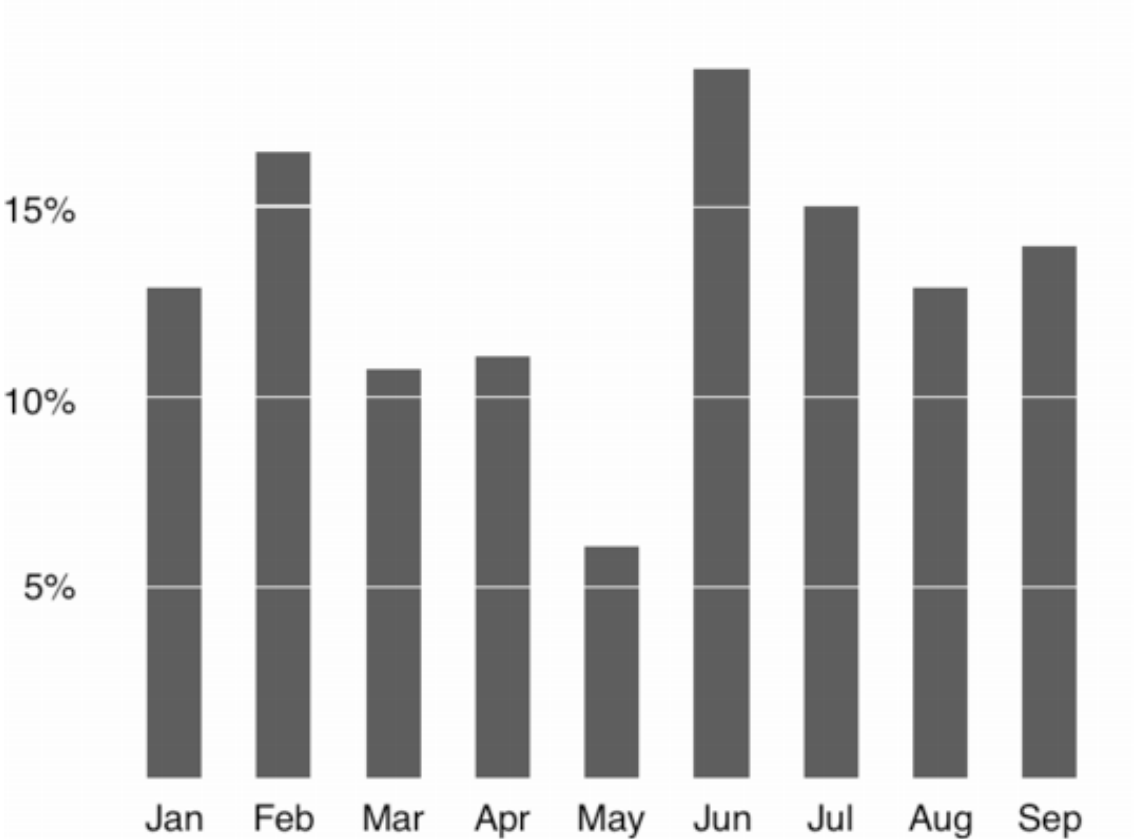
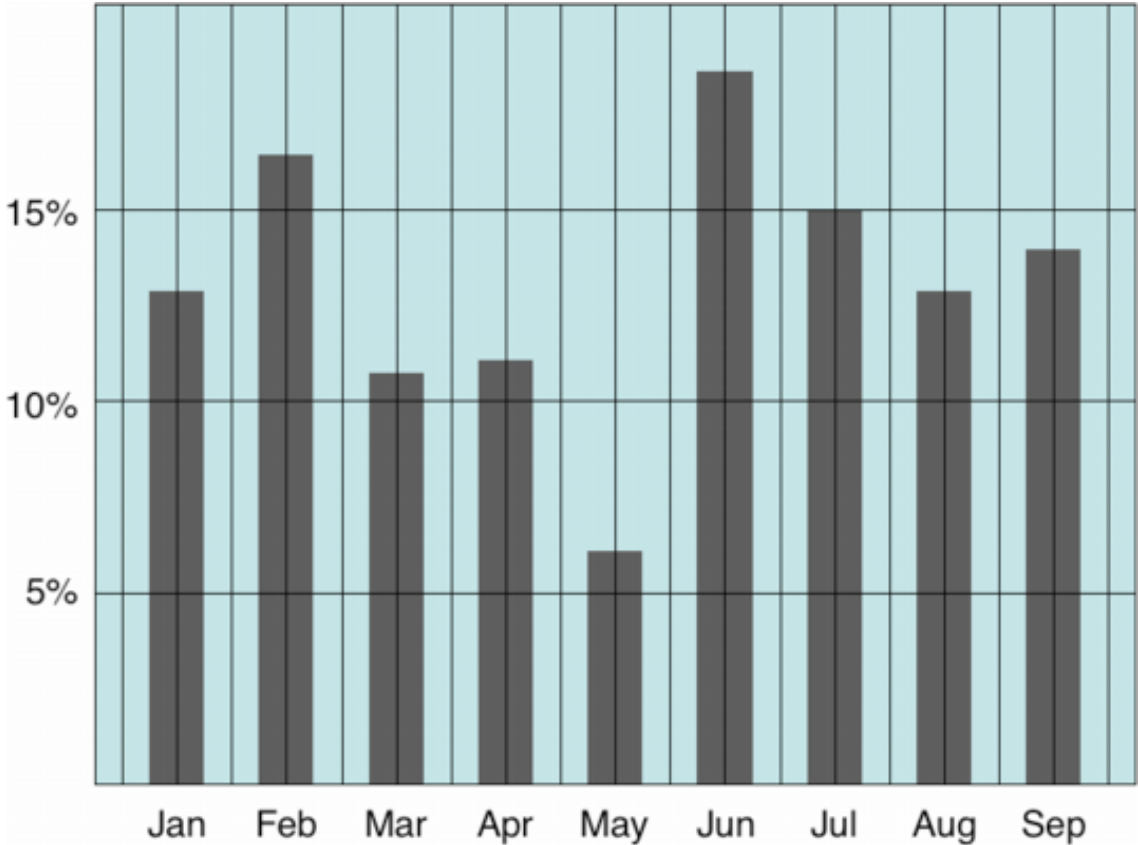
Data Ink Maximization



Activity – (10 minutes)

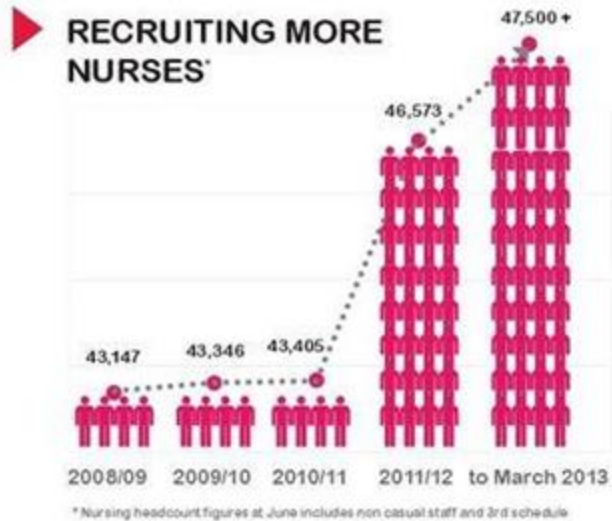
Redraw the chart removing the necessary ink that is present.

Avoid Chart Junk – Data Ink Maximization Example



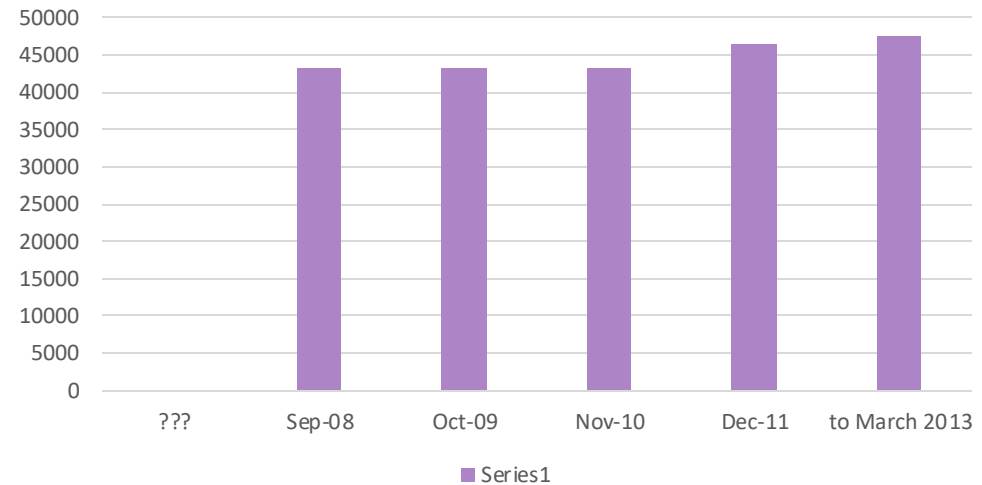
Is this visualizations misleading? Why or why not?

The NSW Health system is...



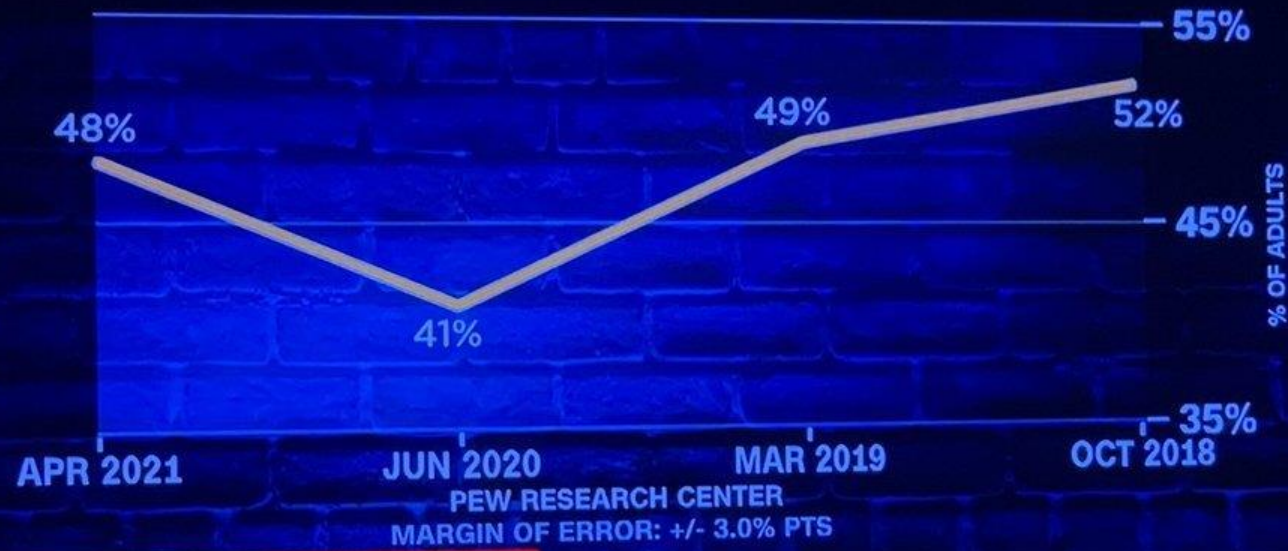
NSW Ministry of Health March 2013

Same data with zero shown



VIOLENT CRIME IS A VERY BIG PROBLEM

ADULTS



THE WIZARD OF ODDS

WHITE HOUSE PREPARES TO ADDRESS SURGE IN VIOLENT CRIME

LIVE
CNN
6:32 PM PT

LESS EAGER TO GET THE SHOT," JEFF ZIENTS SAYS ► BIDEN WANTED 70 CUOMO PRIME TIME

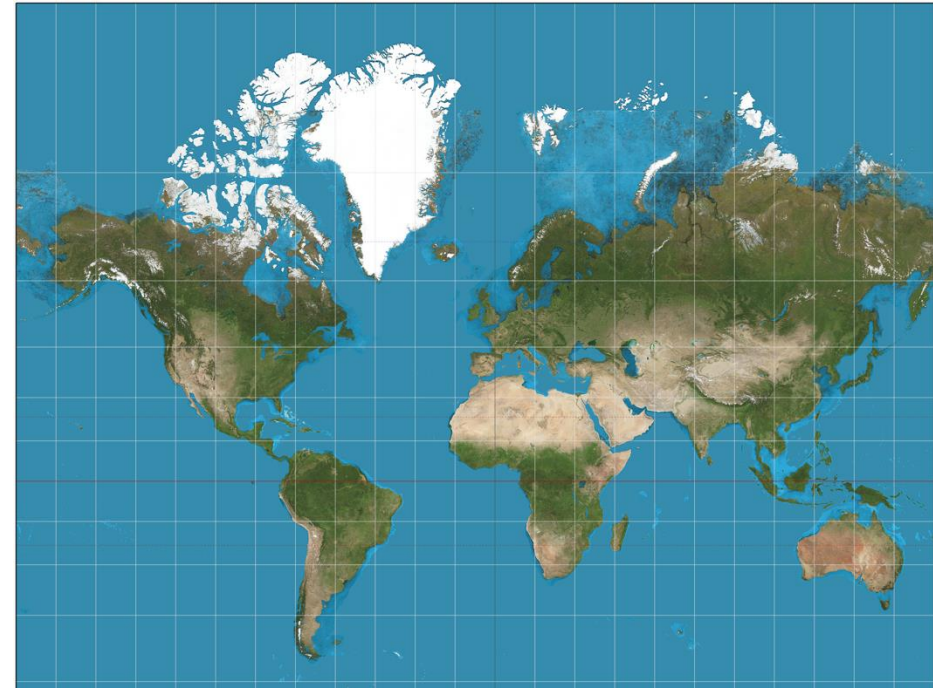
Question

If Greenland and Africa were compared by size, which would be true?

- A) Greenland is nearly twice the size of Africa.
- B) Greenland and Africa are roughly the same size.
- C) Africa is significantly larger than Greenland.
- D) Greenland is slightly larger than Africa.

If Greenland and India were compared by size, which would be true?

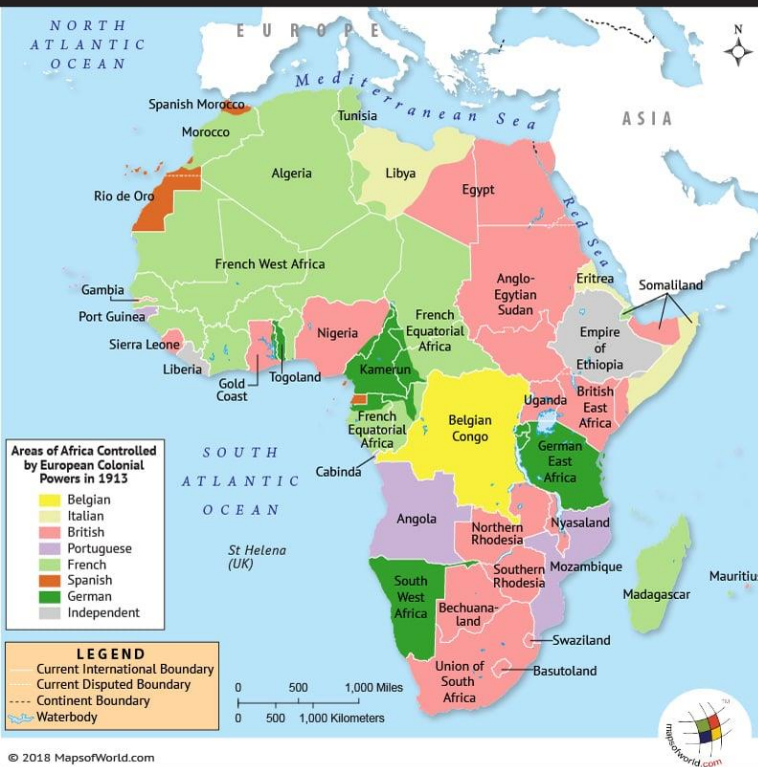
- A) Greenland is nearly twice the size of India.
- B) Greenland and India are roughly the same size.
- C) India is significantly larger than Greenland.
- D) Greenland is slightly larger than India.



Activity I: Why Mercator ~~was~~ is Problematic

Visit <https://thetruesize.com>, and start with a clear map. Use the search bar to find at least two countries and place them side by side at the equator. Observe if this aligns with your understanding of their shape and size. Next, move these countries closer to the poles and note how their shape and size appear to change.

COLONIZATION OF AFRICA



The True Size of Africa

A small contribution in the fight against rampant *innapapancy*, by Kai Krause

In addition to the well known social issues of illiteracy and innumeracy, there also should be such a concept as "*innapapancy*", meaning *in-sufficient geographical knowledge*.

A survey with random American schoolkids let them guess the population and land area of their country. Not entirely unexpected, but still rather unsettling, the majority chose "*~2 billion*" and "*largest in the world*", respectively. Even with Asian and European college students, geographical estimates were often off by factors of 2-3. This is partly due to the highly distorted nature of the predominantly used mapping projections (such as *Mercator*).

A particularly extreme example is the worldwide misjudgement of the true size of Africa. This single image tries to embody the massive scale, which is larger than the *USA*, *China*, *India*, *Japan* and *all of Europe* - combined!

COUNTRY	AREA x 1000 km²
USA	9.629
China	9.573
India	3.287
Mexico	1.964
Peru	1.285
France	633
Spain	506
Papua New Guinea	462
Sweden	441
Japan	378
Germany	357
Norway	324
Italy	301
New Zealand	270
United Kingdom	243
Nepal	147
Bangladesh	144
Greece	132
TOTAL	30.102
AFRICA	30.221
Just for Reference: The Surface of the MOON	37.530

Please note:

The graphical layout of this map is meant purely as a visualization to illustrate the fact: Africa is much larger than almost everyone assumes! Even totally blurred outlines could have been used to make that point, however the table at left is very accurate, citing:

http://en.wikipedia.org/wiki/List_of_countries_and_outlying_territories_by_total_area

Note for instance that the figure in the table for the USA does include Alaska and Hawaii, but they are not even used in the map, as are a handful of other entries (such as Norway and Sweden).

The reason for this is that the map purposely uses the familiar shapes, as if you are moving pieces in Google Maps! Because the mathematical exact depiction, using equal area scaling, would be even more drastic, but would appear highly distorted. I chose to retain the commonly known outlines and proportions to tell the story, even if this conservative size has 'leftover parts'.

The small maps on the right are again the singular message: see some of the countries in direct relation to Africa, a view that is quite unfamiliar and rarely seen.

It is worth looking at Bucky Fullers maps or the Peters equal area proposals, among many other beautiful attempts to display geographical information. Numerous other side-by-side comparisons have been made, this is by far not the first and hopefully not the last such map: someone should find the best fit of all puzzle pieces in a neutral projection.

Until then, please do not take it all too literal! ("where is Ibiza??") and simply take that one impression with you: Africa... is immense.

© creative commons

No Rights Reserved

This work is placed in the Public Domain



Incorrect Attribute to Channel Mapping or Intentionally Misleading

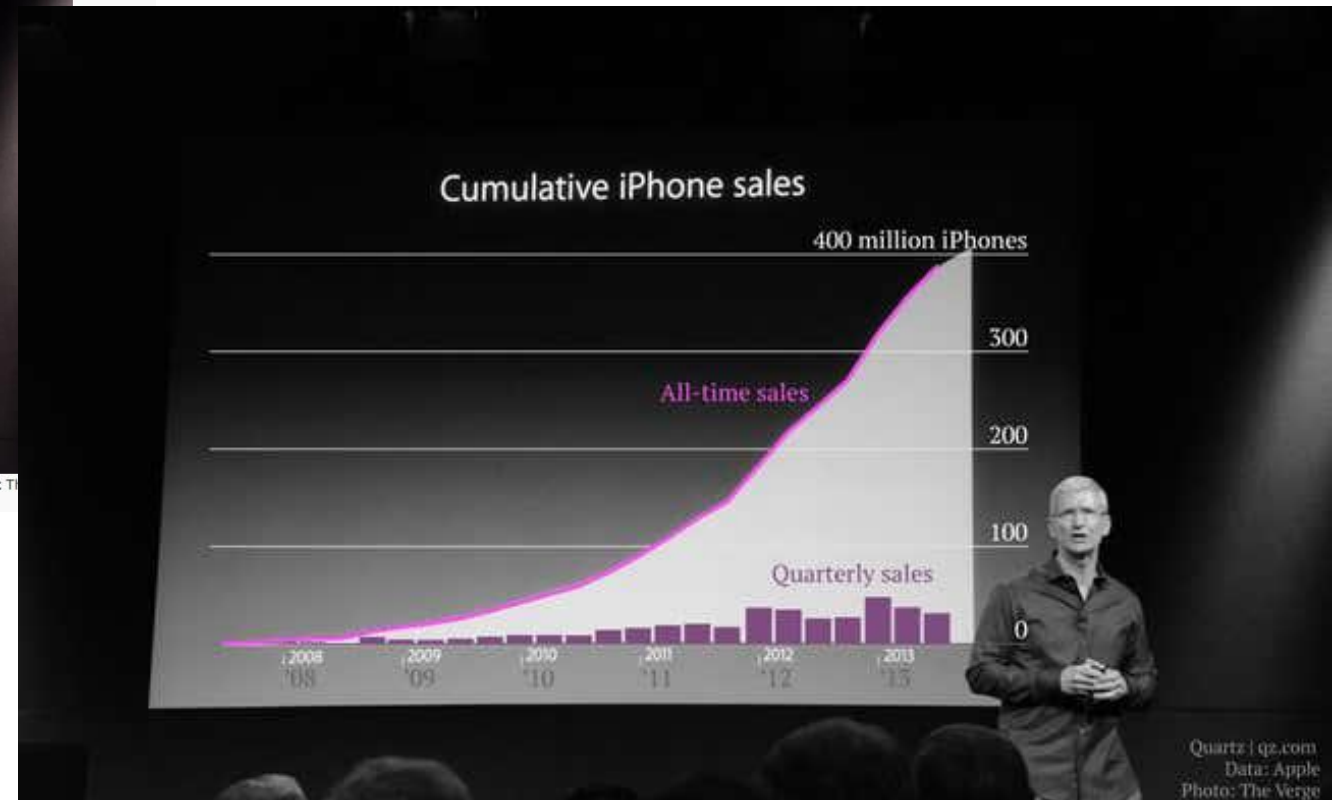


Image: The Verge

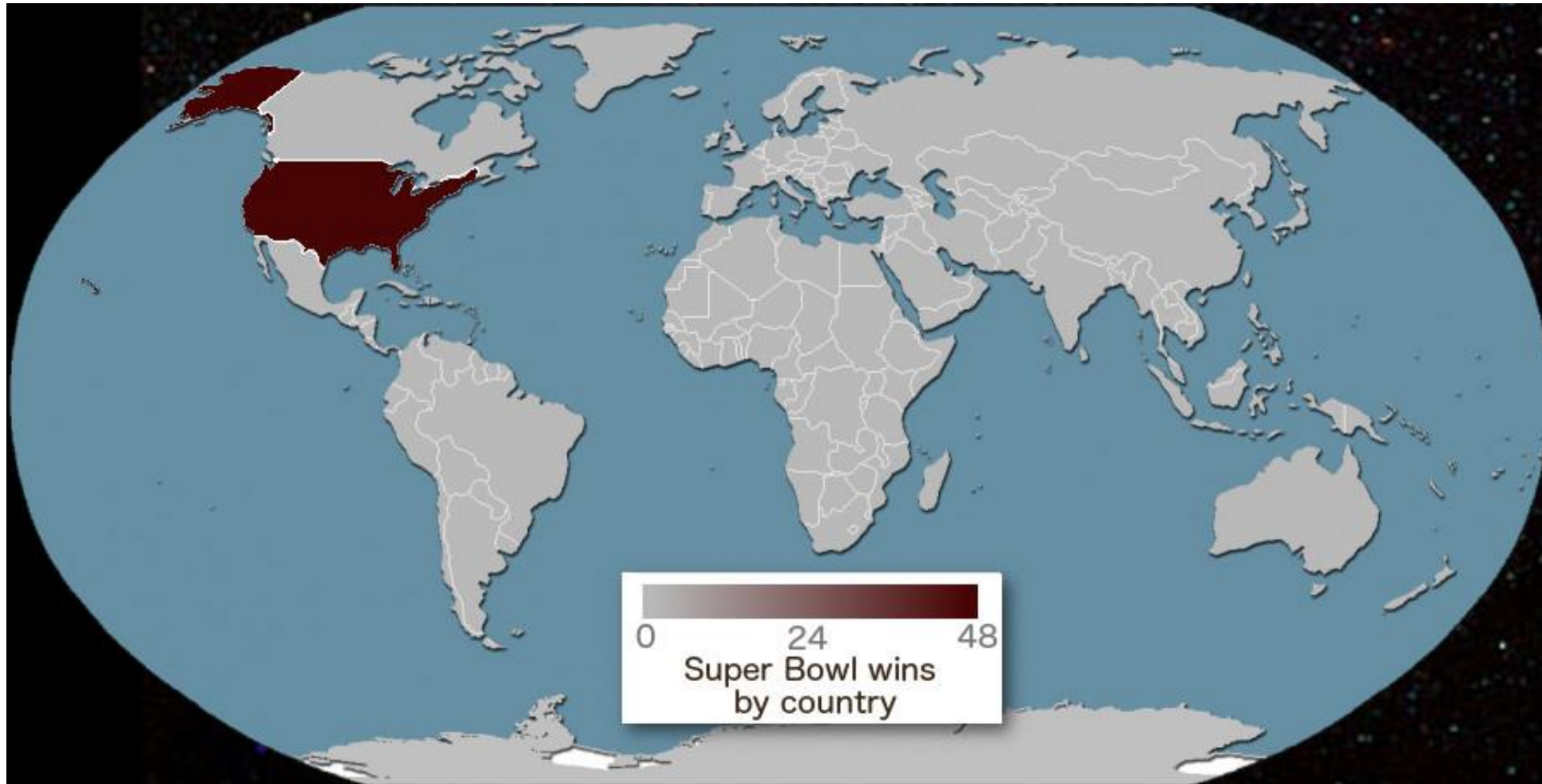
Incorrect Attribute to Channel Mapping or Intentionally Misleading



Image: T



Intentionally Misleading Viz.



The purpose of computing is insight, not numbers

Richard Hamming (1973)

The purpose of visualization is insight, not pictures

Card, Mackinlay, Shneiderman (1999)

Looking Forward

Wow, as mentioned, in 3 days (roughly 5 hours) you were exposed to content that typically takes students months to learn. Wow.

So what is next

1. Don't stop here. Now that you know the “language” of viz, keep using it so you don't lose it
2. Learn more about the theoretical underpinnings of VIZ. There are many research studies that explore the dance between visual perception and cognition.

Common Ways Forward

Now that I **know** Altair, what's next

- In Python, you can explore similar libraries like [Seaborn](#)
- In JavaScript
 - Learn [Vega](#), the JavaScript superset on which Altair is built
 - Learn [D3.js](#), the JavaScript superset on which Vega is built

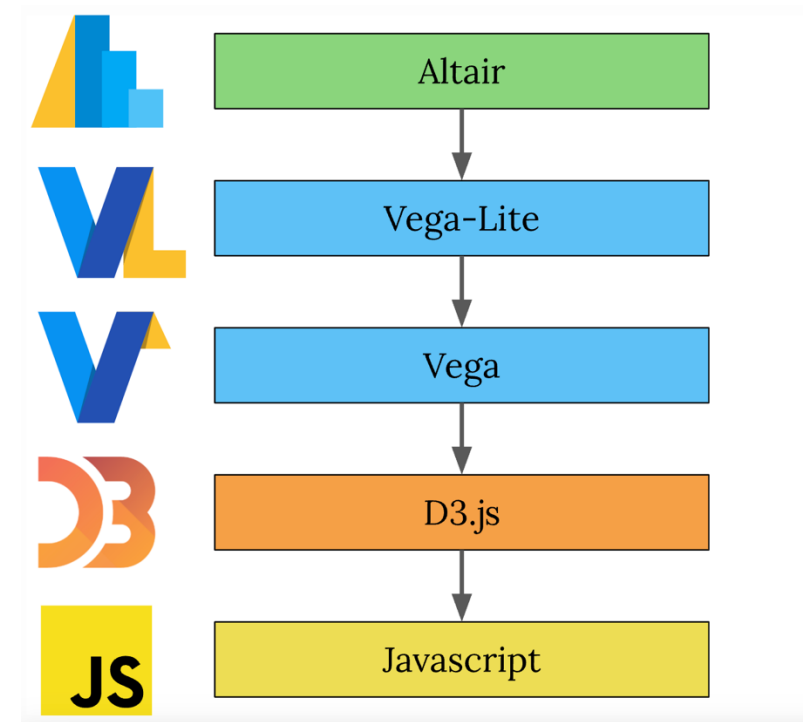
Now if you feel you still don't know Altair, work through the [textbook](#) created by the creators of Altair. It has additional examples. Note that this book is not updated and uses Altair v.4 and not v.5. For updated examples consult the Altair API

Join weekly design competitions

- <https://makeovermonday.co.uk/>

For self-paced projects, tidy clean data can be found @

- <https://ourworldindata.org/>
- <https://www.gapminder.org/>
- <https://archive.ics.uci.edu/>
- <https://data.worldbank.org/>



<https://eitanlees.github.io/altair-stack/>

you
can
do it



Stay in touch

- I would love to hear about how you are using Data Visualizations.
- My email address is kemiola@cs.ubc.ca