VISUALIZING DATA
Veracity, Accuracy, Accountability

Alberto Cairo
Dialogues in Research Ethics
University of Miami, 2020
We are living through a Golden Age of visualization
We are living through a Golden Age of visualization
The most-viewed piece ever published by The Washington Post online

Why outbreaks like coronavirus spread exponentially, and how to “flatten the curve”

By Harry Stevens March 14, 2020

PLEASE NOTE
The Washington Post is providing this story for free so that all readers have access to this important information about the coronavirus. For more free stories, sign up for our daily Coronavirus Updates newsletter.

https://www.washingtonpost.com/graphics/2020/world/corona-simulator/
Common misconceptions when talking about visualization:

1. “A picture is worth a thousand words”
2. “Visualization is intuitive”
3. “The data should speak for itself”
4. “Show, don’t tell!”
Visualizations can’t be designed based just on our personal preferences—although these are important.

**Visualization is a bit like writing:** beyond some conventions and constraints regarding symbols, visual grammar, perception, and cognition, visualization can’t be based on “rules” that are set in stone.

Instead, when designing visualizations, we need to be guided by reasoned, justifiable choices.
“Facts give us reasons [...] when they count in favor of our having some belief or desire, or acting in some way.”

Derek Parfit, *On What Matters*
Reasoning about visualization.

Key questions:

1. Why to visualize?
2. What to visualize?
3. Who to visualize for?
4. How much to visualize?
5. How to visualize it?
6. What style to use?

Note: this is just a tentative list of questions
Why should my visualization exist?

Do the potential benefits of designing my visualization outweigh the possible harm it might cause?
"Where are the gun permits in your neighborhood?" That's the question posed by The Journal News, a New York newspaper that published a Google map on Sunday that shows the names and addresses of pistol or revolver permits in Westchester and Rockland counties."

“We felt sharing information about gun permits in our area was important in the aftermath of the Newtown shootings.”

Janet Hasson, president and publisher of the Journal News Media Group

Published Dec. 23, 2012 (the Sandy Hook Elementary School shooting was on Dec. 14)
Why should this data be made public?
Why should it be made public through a map?
Why should it be this type of map?
Even if we decided that this data is worth publishing, wouldn’t a different map be better?

What are the potential consequences of my decisions?
Are the benefits worth the risk of harm?
2. **What to visualize?**

Do I understand my data, its limitations, uncertainty, or glitches? What or who is being measured (*or not being measured,*) and why?
In Florida more than 71,000 students are homeless. During the last decade, this population rocketed as a result of the recession and how hard it has become for the poorest families to find affordable housing.
Kansas is the nation’s porn capital, according to Pornhub

Blue states watch more porn. But what’s the matter with Kansas?

According to Pornhub Insights, Kansas leads the nation in porn pageviews per capita at roughly 194. They don't specify what interval this is over (monthly, weekly, etc), but the state-by-state comparison is nonetheless interesting.

Plotting Obama vote share in 2012 versus porn consumption, it looks like blue states consume more porn per capita than red ones. Aside from Kansas - a clear outlier - and Georgia, the remaining top ten per-capita porn consumers are all blue. Similarly, New Mexico and Maine are the only blue states in the bottom ten per-capita porn consumers.

https://wonkviz.tumblr.com/post/82488570278/kansas-is-the-nations-porn-capital-according-to
Kansas is the nation’s porn capital, according to Pornhub

Blue states watch more porn. But what’s the matter with Kansas?

According to Pornhub Insights, Kansas leads the nation in porn pageviews per capita at roughly 194. They don’t specify what interval this is over (monthly, weekly, etc), but the state-by-state comparison is nonetheless interesting.

Plotting Obama vote share in 2012 versus porn consumption, it looks like blue states consume more porn per capita than red ones. Aside from Kansas - a clear outlier - and Georgia, the remaining top ten per-capita porn consumers are all blue. Similarly, New Mexico and Maine are the only blue states in the bottom ten per-capita porn consumers.

https://wonkviz.tumblr.com/post/82488570278/kansas-is-the-nations-porn-capital-according-to
Kansas is the nation’s porn capital, according to Pornhub

Blue states watch more porn. But what’s the matter with Kansas?

According to Pornhub Insights, Kansas leads the nation in porn pageviews per capita at roughly 194. They don’t specify what interval this is over (monthly, weekly, etc), but the state-by-state comparison is nonetheless interesting.

Plotting Obama vote share in 2012 versus porn consumption, it looks like blue states consume more porn per capita than red ones. Aside from Kansas - a clear outlier - and Georgia, the remaining top ten per-capita porn consumers are all blue. Similarly, New Mexico and Maine are the only blue states in the bottom ten per-capita porn consumers.

https://wonkviz.tumblr.com/post/82488570278/kansas-is-the-nations-porn-capital-according-to
Disclosing limitations and uncertainty

Uncertainty and graphicacy
How should statisticians, journalists, and designers reveal uncertainty in graphics for public consumption?

https://graphics.cs.wisc.edu/Papers/2014/CG14/Preprint.pdf

Collection of papers about visualizing uncertainty:
https://www.dropbox.com/sh/jk4ginxyai6ylqu/AABvqdyTIhJtyFN9nKNHxY9Ba?dl=0

Alberto Cairo • University of Miami • www.thefunctionalart.com • Twitter: @albertocairo
3. Who to visualize for?

Have I thought about how my intended audience will access my graphic? Will they understand it? Can I explain it better?
Good journalism isn't just showing charts. It's also about explaining them: twitter.com/brianstelter/s...

This infographic is worth a thousand words – showing why "social distancing" and other protective measures helps to slow an outbreak. Hat tips to CDC, @theeconomist, @drewaharris, and @CT_Bergstrom

https://twitter.com/AlbertoCairo/status/1236773377865658370
Show AND tell

Hans Rosling, www.gapminder.org
Show **AND** tell

Hans Rosling, *The Joy of Stats*
“I and my colleagues here at the FT, we really do think one of the most valuable things we can do as data visualization practitioners is add this expert annotation layer.”

John Burn-Murdoch
Financial Times

“Design secrets behind the FT’s best charts of the year”
https://www.ft.com/content/4743ce96-e4bf-11e7-97e2-916d4fbac0da
Every outcome in our simulations

All possible Electoral College outcomes for each candidate, with higher bars showing outcomes that appeared more often in our 40,000 simulations

https://projects.fivethirtyeight.com/2020-election-forecast/
4. How much to visualize?

Am I showing too little?
Am I showing too much?
Most places are pretty safe, and have likely remained down here (these aren’t real data points).
The danger of aggregating data too much, and presenting just averages and other statistical summaries.
5. How to visualize it?
What types of charts or maps should I use?
What is the best way to organize the visualization?
Figure 2 - Main nationalities of arriving migrants – 2016

Greece
Figure 2 - Main nationalities of arriving migrants – 2016

Greece

- Syria 47%
- Afghanistan 24%
- Iraq 15%
- Pakistan 5%
- Iran 3%
- Palestine 1%
- Algeria 1%
- Others 4%
Syria 47%
Afghanistan 24%
Iraq 15%
Pakistan 5%
Iran 3%
Palestine 1%
Algeria 1%
Others 4%

Books to make design choices

- Fundamentals of Data Visualization
  A Primer on Making Informative and Compelling Figures
  Claus O. Wilke

- Storytelling with Data
  A data visualization guide for business professionals
  Cole Nussbaumer Knaflic

- The Truthful Art
  Data, Charts, and Maps for Communication
  Alberto Cairo

Draft available online:
https://serialmentor.com/dataviz/
6. What visual style to use?

Not all visualizations need to be minimalist. Not all visualizations need to be flashy and innovative, either.
Standard visualizations
Appropriate for graphics we use all the time

**Total Deficits and Surpluses**

Over the next decade, federal spending on health care is projected to take up an increasing share of gross domestic product.

Source: Congressional Budget Office (as of August 2018).

- CHIP = Children's Health Insurance Program.
- a. Net Medicare spending includes offsets of receipts from premiums paid by beneficiaries, recoveries of overpayments made to providers, and amounts paid by states from savings on Medicare's prescription drug costs.
- b. Spending to subsidize health insurance purchased in the marketplaces established under the Affordable Care Act and provided through the Basic Health Program and spending to stabilize premiums for health insurance purchased by individuals and small employers.
https://jaimeserra-archivos.blogspot.com/
Fully customized visualizations:
Appropriate for one-time use when we want to provoke curiosity, surprise—or simply a smile

http://visualoop.com/28792/portfolio-of-the-week-jaime-serra

Alberto Cairo • University of Miami • www.thefunctionalart.com • Twitter: @albertocairo
What if all covid-19 deaths in the United States had happened in your neighborhood?

Find out what would happen if your neighborhood was the epicenter of the coronavirus pandemic in the United States.

The purpose of visualization isn’t visualization per se. The purpose of visualization is to help people make sense of the world through a combination of visuals and words.
The End.

www.thefunctionalart.com, www.albertocairo.com, alberto.cairo@gmail.com