data journalism
semester 1, block 2
2021-2022

week 2: basic stats & numerical data analysis
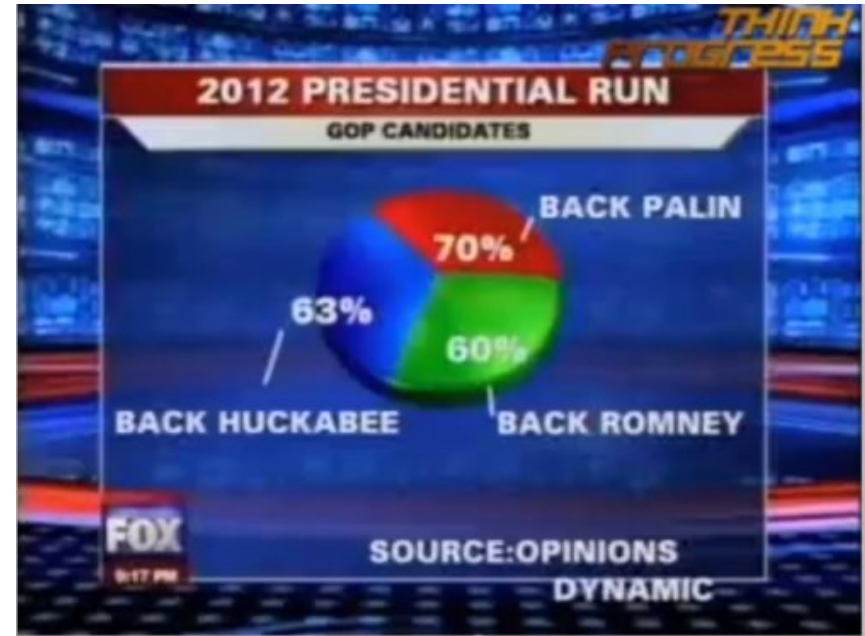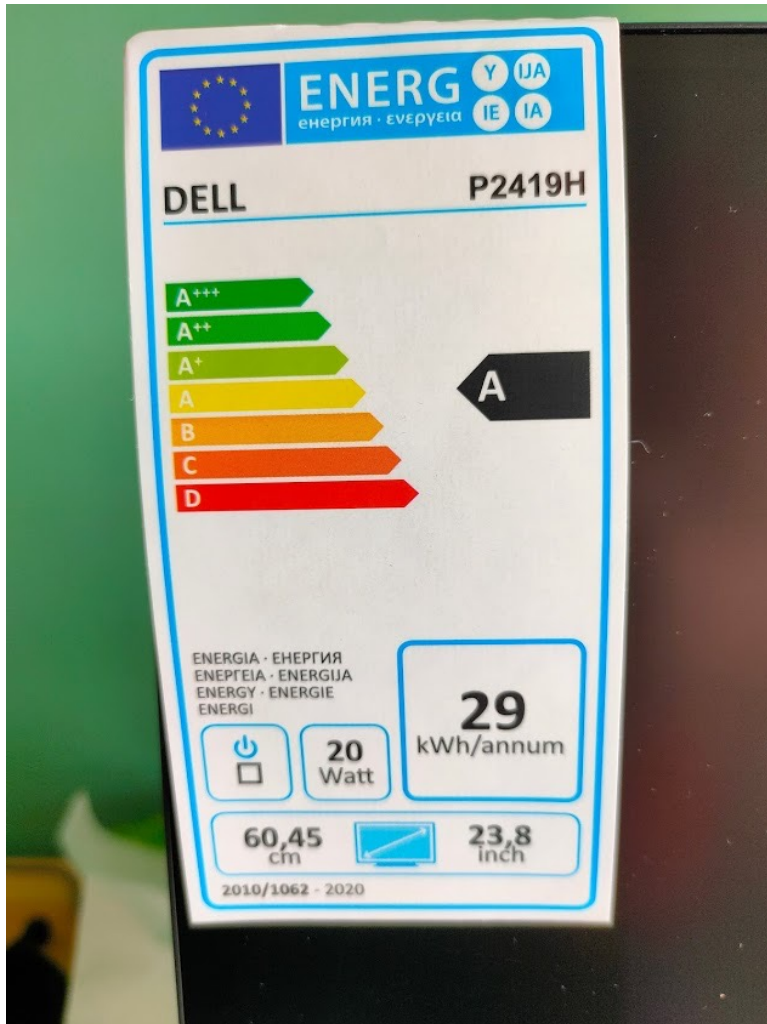
Picking up from last week…

- A lot of actual data journalism is (still) poorly done (Cushion et al., 2017)

- And: numbers are used misleadingly, ALL THE TIME.

- But if priorities of DJ include:

  - Transparency, (new, big) data, *interpretation* of info, (sometimes) interactivity for the reader

- …then accurate (statistical) analysis is essential

'Calling bullshit'

- Carl Bergstrom & Jevin West

- … did the calculations, and…

- strong coffee is also 99.9% caffeine free

- callingbullshit.org

Plenty of examples…

http://livingqlikview.com/the-9-worst-data-visualizations-ever-created/

Some considerations

- You: more training than average reader

- Consider:
    - Literacy of **audience** (DJ vs thesis)
    - Complexity of analyses
    - Your goals

- Bottom line: it doesn't have to be complex to tell a valuable story

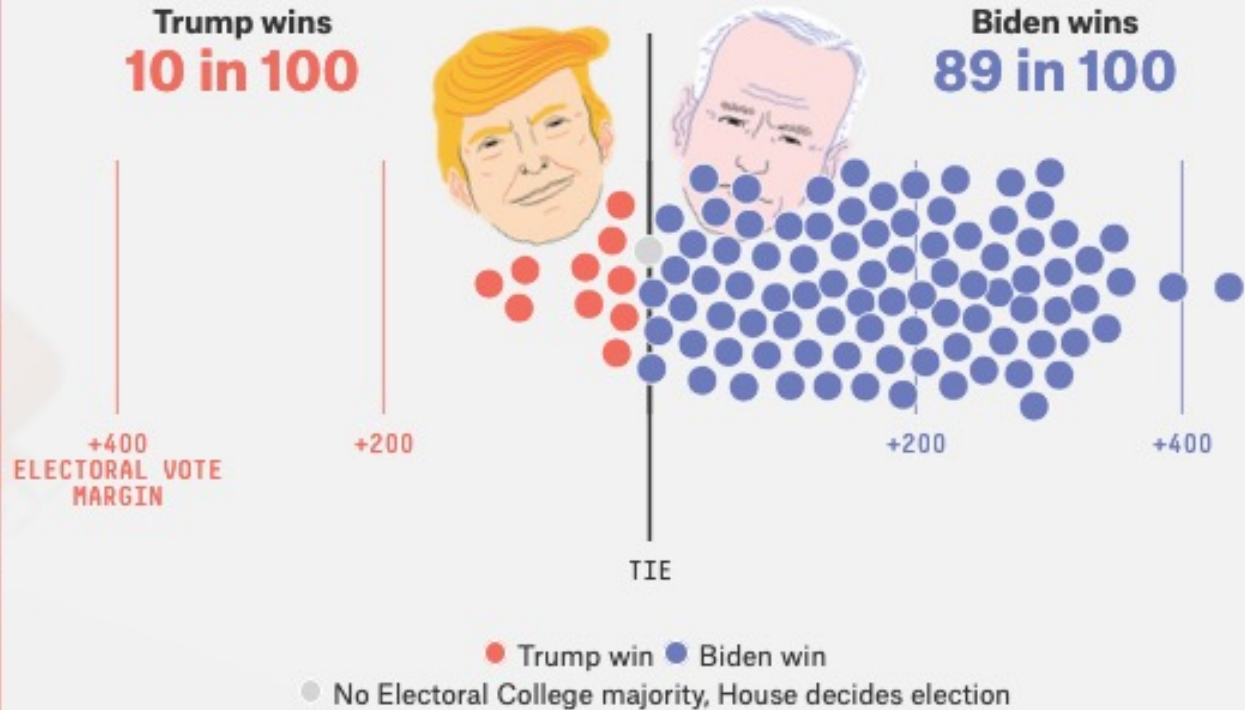- The more complex, the more explanation needed

Lessons learned since 2016…

Questions to ask of your data

- 1. Where is it from?

- 2. Is there an agenda associated with it, that might skew it?

- 3. How were the data collected/could it be better?
    - Sampling & operationalization issues!

- 4. What are they said to mean, and what do they actually mean?
    - Statistical versus practical significance
    - How will your reader be affected?
    - Risks versus harms

LIES, DAMN LIES AND STATISTICS

But perhaps first:

- Where the heck do I get some data?

Moving forward

- Thursday : work on finding your datasets for "Step 1" of your project. (Due Friday at noon!)

- Read through the assignment on Canvas

- Today: notebook 3
  - Exploring datasets (continued)
  - Getting descriptives, doing basic statistical comparisons

- But, tailored groups
  - 1. All clear, no big questions, ready to move on
  - 2. Some parts of notebook 2 were confusing, I'd like to review before moving on
  - 3. For those of you feeling more lost than (2), please join group 2, and set up an appt with Vlad.

- But also: all of you can use Vlad! Also for help on your projects.

For Thursday

- Bring rough ideas of your interests (topics, data)

- Start looking for datasets

- Accessing/opening them is critical

  - Can you describe them?

  - Do you know where they originated?

- Remember that this will lead to a feedback moment – you can pose questions for us, too, in your assignment.