

Tips to find/analyze/present data

Aleszu Bajak, USA TODAY

1 FIND

- Ideation/Inspiration
 - Maybe a competing newsroom's investigation or data visualization has you thinking you could find and present another dataset in a similar way.
- Research/Interviews/FOIAs
 - Think, "Wouldn't it be cool if I had, by state and by year and by program, a breakdown of funding for X?"
 - Dream/request big; settle for good enough.
 - FOIAs: Design a simple version of the spreadsheet you're looking for and make a very specific records request. If you see in a FOIA log that a similar dataset was granted to someone else, request *that* or that they update that.
- Literature/Coverage review
 - I wonder who studies X?
 - I wonder how others have covered this?
 - These expert sources end up being "rabbis" for your story, helping review your analysis and presentation.
- Collect/Structure
 - Structure the data, inspiration and literature you find
 - Note the vintage, source, keywords to help you organize/call up later, include a note on why you found it worthy of saving.

2 ANALYZE

- Exploratory data analysis
 - Find the minimum, maximum, average
 - Look for "missingness"
 - Build some pie charts and histograms (I present these to experts and colleagues)
- Recipe
 - Write down a "recipe" of every analytical step you take on a dataset.
 - i.e. First I filtered for X, then I found the mean number of "barrels of oil" and then I did Y.

- Ideally, save each time you've transformed the dataset into its own spreadsheet so you can retrace your steps.
- Merge
 - Brainstorm what would be cool to merge with this dataset. Think, "Wouldn't it make sense to try and see if this correlates with race or household income or amount of federal funding or fall migration?"
 - Great data stories and insights lie at these intersections

3 BULLETPROOF

- Run your analysis by an expert or five!
- Dummy check
 - make sure your conclusions make common sense, lines up with previous reporting, or has been concluded in empirical literature.
 - Are you sure you're not making a dangerous assumption?
 - Run it by colleagues, friends.
- Beware making causal conclusions
- Push back on editors that want to change careful language

4 PRESENT

- Make sure you translate the numbers/conclusions into easy-to-digest takeaways.
 - "If you lined X up end to end, how many times would it circle the globe?"
- Point out caveats and limitations to editors, colleagues and readers
- Be clear with readers
 - Write a "nerd box" with methods and assumptions/limitations.
 - Increasingly there is a push to make data journalism replicable. Short of posting your code and data to GitHub, what can you tell readers and researchers about your methods?

Aleszu Bajak
Senior data reporter
USA TODAY, Investigations
Cell: 216.644.1952
Email: abajak@usatoday.com