Tips on Editing Data Stories

**While editing a story:**

* Think: Can this be quantified?
* A data point can often explain a trend better than a reporter or a source.
* Take extra care with the lede and beware of “records,” unless we’re sure it really is a record.
* Stop when you see a number.

**Data Googling tips:**

* Use the word “historical” when searching for older data.
* Use the word “data” in your search.
* Keep your search broad (crime data NYC) to see the sourcing options, then explore the different sources (FBI, NYPD, real estate sites, etc.).
* Use an advanced Google search for .xls or .xlsx files or government websites.
* Census data:
  + Census Reporter: Easy to use, less information.
  + Census Factfinder: More information, harder to use.
  + Social Explorer: Historical data, mashed with other databases.
* Google scholar is a great source for research.

**Common data pitfalls on first edit:**

* Cherry picking: Picking out a data range that supports a point of view, while ignoring the larger trend.
* Bad sample: Saying something about a larger group based on a non-representative sample.
* Reporting trends without proper context, especially when other factors may account for them. Using percentages to create a false comparison between a small number to a large number.
* A related problem: Using percentage change for small numbers; this is misleading.
* Correlation doesn’t equal causation. Even if we don’t say something is causing something else, putting two trends next to each other encourages readers to draw that conclusion.
* Using a quote when a data point would be better: Quoting a source when we should be using data or quoting a source stating a data point as fact that they couldn’t possibly know
* Watch for data-like words such as “arguably more,” “more and more” or “increasingly more.” Do we actually know this? If yes, say so.
* Pay special attention to data in your lede and headline. Is it written as accurately and transparently as possible?
* Credit and link (if possible) to the source of the data; ask where reporters got data.
* If the reporter analyzed data themselves, consider a “nerd box.”

**Common data pitfalls on second edit:**

* Median and average/mean: Use medians if outliers are present. Otherwise, you can use average.
* Beware the “illion” problem: Do a search on “illion” to ensure that every “illion” that should start with a “m,” “b” or “tr” does in fact start with the right letter.
* In describing increases or decreases, be sure to use a to/from formulation -- “Net income rose **to** $68 million **from** $54 million.” (Otherwise it sounds like a range.)
* Taking a percentage past the decimal point -- 12.2 percent -- can be deceptive if one number in the calculation is an estimate or padded with many zeros.
* Use figures for ages of people, but for inanimate objects spell out one through nine and use figures for 10 and above.
* Avoid “times less” or “time more” formulations, as in “five times less.”
* When in doubt, adjust!
  + For inflation: <https://data.bls.gov/cgi-bin/cpicalc.pl>
  + Seasonally: Use numbers that are already seasonally adjusted or compare the same month in different years to avoid this issue altogether.
  + Population: Use rates instead of real numbers in a comparison if the population could have changed. The world is always growing!
* Don’t confuse the percentage point difference (40 percent - 30 percent = 10 percentage points) with percent change (40 percent to 30 percent is a 25 percent decrease). In general, take caution when using percent change with values that are already percentages. This can introduce other errors.
  + Percent change calculator: <http://percent-change.com/>

**Headlines on data stories:**

* If there’s a crucial number, try to get it into the headline.
* If the data at the center of the story won’t be readily grasped in the headline, promise an explanation.
* If there are charts, make that clear in the headline.