Baby name lab notes

* How are we feeling? How did that go? What was difficult? What was unexpected? Was it harder to find a story or plot the data?
* ALWAYS LABEL YOUR AXES
* Tables are good for tabular data. They are very bad at telling stories. They are very good for exploring complex data, or for helping the user find his own interest.
  + Can you tell how much rainier the #1 rainiest day in Seattle was? <http://www.seattleweatherblog.com/wp-content/uploads/2013/01/SeattleWettest.png>
  + Example: What is the #1 rainiest month in Seattle? <http://theweatherprediction.com/weatherpapers/118/6.jpg>
  + Example: <http://apps.chicagotribune.com/news/local/red-light-camera-tickets/>
* Data viz should use raw numbers, not rankings. Rankings are an inaccurate way to plot data, because rankings are dependent upon the other data in the dataset. Rank depends on the universe of data.
  + Rankings should always have context. The ranking is determined by the universe of data and can be influenced by very complex factors.
  + In the case of baby names, the Social Security Administration creates their popularity rankings this way: *We count the number of occurrences of each name during the decade, and then rank them in order of decreasing frequency.* Rank 1 is assigned to the name with the highest number of occurrences, rank 2 to the name with the next highest, and so on.
  + Another tricky thing about rankings: For instance, say you have a very boring ice cream cart that only sells chocolate and vanilla. if only 5 people bought ice cream in February, the #1 flavor only has to be sold a max of 3 times. If 1,000 people buy ice cream in August, you’d have to sell 501 of the same flavor for a #1 rank. But if you plot the rank, you won’t see any change at all between February and August.
* The visualization itself will sometimes **show** the ranking, but it needs raw numbers of the entire applicable universe (<http://stats.stackexchange.com/questions/30995/how-would-one-graph-the-results-of-subjective-rank-order>) OR A DEFINED SUBSET THEREOF (<http://flowingdata.com/2014/05/15/alcohol-consumption-per-drinker/>) to calculate that rank.
* Examples of baby name data viz

Final project prep and homework for next class

You should have done:

* *Find a simple, clean dataset you'd like to explore for the final project. (Please email me if you're at a loss finding one.)*
* *Find a story in this data.*
* *Identify at least two goals for telling this story, and come up with an initial idea for how to visualize it.*
* *Develop this idea into a small pitch (100 words) and accompanying sketch.*

I will be reading your pitches and giving you feedback on them in Canvas in the next few days.

Any questions or confusion?

For next week:

1. *Settle on a group for your final project, if you're choosing to collaborate. Together, come up with five different design options for your project. This means type of data display, as well as layout on the page. Consider throwing them together with a readymade tool like Google, or you can always sketch them. Just make sure that your depiction of the data is as accurate as possible with the tool you choose.*
2. *Come up with a color scheme and proposed type hierarchy—not five, but one that could work across your data designs. Identify why you've chosen these colors and types.*
3. *Identify at least two primary goals for your graphic. (Example goal: Easily compare data across states.)*
4. *Identify two task questions or imperatives to help test these goals. (Example task: Which has a higher rate of incarceration, Alabama or Georgia?)*

Submit all of this here in a compiled PDF or folder, and bring **hard copies to class** on Thursday for a round of rapid testing and iteration. We will also go one-on-one to help form ideas and spend most of class in project development time.