Jennifer Lu

J7462

Final Project Proposal

4/22/2016

I downloaded the NY Time’s version of the data from 2006 to 2014 in order to have a temporary working dataset that I can use for the visualization part of the project, and then swap out once I finish scraping the updated dataset.

I reformatted the date to yyyy-mm in Excel so I could use a pivot table to tally all the oilfield incidents reported by month (csv with mockup bar chart attached). I will then use the bar chart block example to graph # of incidents over time, with about 108 nominal categories on the x-axis. I suppose a line chart might be better, because 108 months are continuous enough and 108 nominal categories might get a little crowded.

Since this is fairly straightforward, I’m hoping that if I finish this quickly, I can do something a little more challenging, if not by the end of the semester, than over the summer. I’ve used pivot tables to tally the number of oilfield incidents by county and year, and temporarily charted themm as bar charts from 2006 to 2014. I’d like to make [choropleth maps](https://bl.ocks.org/mbostock/8ca036b3505121279daf) showing a heat map of incidents by county and year. The first step would be to build a choropleth maps for every year, and then stitch them together so you can see how the intensities change over time. Do you think it would be better to show year by year data or cumulative (ie. 2006 data, then 2006 + 2007 data, then 2006+2007+2008 data)?

4/15/2016

A lot has been written about the environmental toil resulting from increased drilling in the Bakken oil field, but I can’t find any information about whether this trend of increased environmental incidents in the oilfields has continued or reversed now that the oil boom has gone bust.

To find out, I’d like to tally the number of oil and gas-related environmental incidents by month and year, and plot a simple line chart or bar chart showing the number of incidents over time. Whether the number of incidents post-boom decreases, stays the same, or continues to increase, the reasoning behind the trend will be interesting.

Purely speculating, I’d expect the number of incidents to go down after the oil bust, because fewer wells are being drilled and fewer barrels of oil and gas are being produced. If the number of incidents declines, then that’s a happy story. Maybe the companies are being more responsible now that they’re not expanding at neck-breaking speeds. But if the numbers remain steady or go up, maybe they’re being more negligent because even though they have fewer wells to manage, they’re also not willing to spend the money for maintenance when they’re losing money during the bust.

North Dakota’s Department of Health and their Department of Mineral Resources' Oil and Gas Division of the North Dakota Industrial Commission keep track of all incidents reported to them on this website: <http://www.ndhealth.gov/ehs/spills/>

The format isn’t very user-friendly because the dataset is split between three webpages: on-site incidents within the past 12 months, off-site incidents within the past 12 months, and incidents older than 12 months. Also, the webpage shows only a certain number of rows of a table on the page at a time. I could scrape the data from each of the three pages and combine everything, but hopefully there’s an easier way. Tomorrow, I’m going to call the Department of Health to see if they’d be willing to give everything to me in one big csv file so I can tabulate it using a pivot table or SQL and work from there. I’ve also sent an email to Chase Davis, because [the New York Times did a very similar analysis during the oil boom](http://www.nytimes.com/interactive/2014/11/23/us/north-dakota-spill-database.html), and I’d like to know how they worked around the formatting to put together their database.

If this project sounds too ambitious or hinges on too many caveats, I will think of something else.

11/14/2016

Called and emailed Alison Ritter, PIO for the Department of Mineral Resources, about getting the raw data. Her response:

Hi Jennifer-

Our information is not compiled in excel or csv, but rather filed on an individual well by well basis.

You can access our information for free by using our Bismarck lobby computer- or by getting a basic website subscription for $50 year. Otherwise, the only other publically available information is through the Department of Health’s website.

Regards,

***Alison Ritter***

Public Information Officer

Department of Mineral Resources

1016 East Calgary Ave, Bismarck

Phone: [701-328-8036](tel:701-328-8036)

Fax: [701-328-8022](tel:701-328-8022)

[amritter@nd.gov](mailto:amritter@nd.gov)

[www.dmr.nd.gov/oilgas](http://www.dmr.nd.gov/oilgas)

I find the fact that the information exists in a database that can be recalled solely on an individual well by well basis hard to believe.