How to answer energy questions with EIA data and Python.

How old are our power plants?

I started there. This is something I’ve thought about before, although the idea never demanded enough of my attention to make the time-investment of a graph. I approach all of these small, hobby, energy-nerd questions with a graph as the finish line. I do inherently enjoy most data visualizations, but more importantly, reaching that point is an effective form of personal enrichment. My methods for any topic are vertically integrated. I find or create the data, figure out the calculations I might need, write code to get it all done, decide on a visualization, write code to make that visualization happen, then figure out the best way to present the results, and possibly talk about it. I don’t think this process is particularly unique in the realm of people-writing-on-the-internet, but because my hobby and work ambitions overlap I’m able to casually gain experience that could be helpful professionally. So, how would I approach that question?

Here’s my current workflow:

1. Ask a question.
2. Define the question
3. Acquire the data
4. Sketch out the workflow and processes
5. Write code
6. Decide on a visualization
7. Write code
8. Refine
9. Present

I’ll take this step-by-step.

1. I already have my question: How old are our power plants?

2. Now I need to define that question. I separate this out because most of my ideas get written down and stop at step 1. Defining the question is turning it around, thinking about its viability, and making sure it encompasses what I’m actually thinking. In this case, the idea evolved in 3 steps: it gets more specific, it gets more complicated, then it gets more expansive.

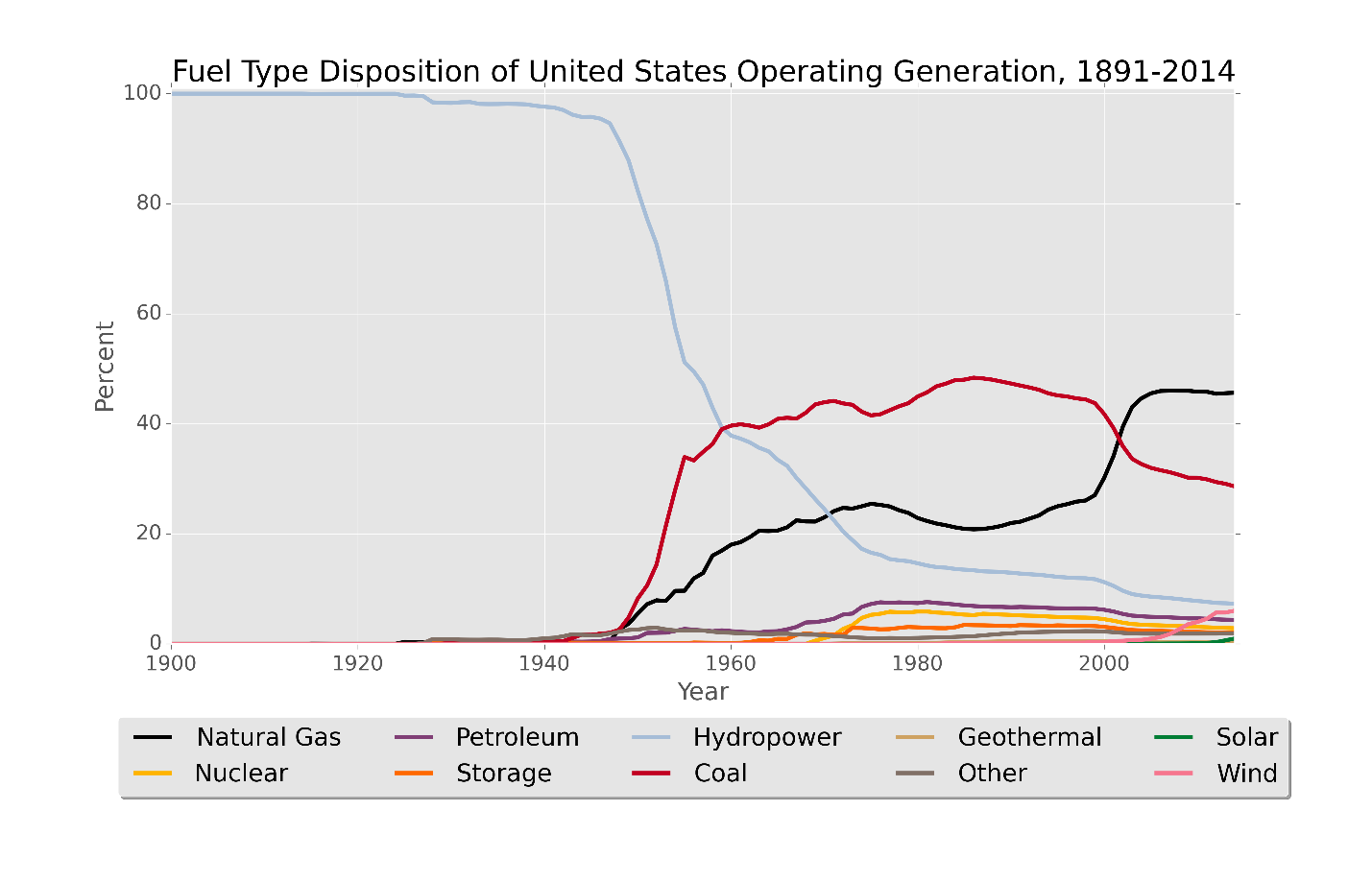
First, I realized I wanted to think in terms of generators and not power plants. It’s possible for a plant to retire partially over time, and rather than tracking MW loss at the powerplant level I thought it would be easier to look at specific generators, because those are typically the units being brought into and out of operation (plus, I already knew that EIA Form-860 has generator data, step 3). So now the question is a bit more specific: How old are our generators?

Second, I complicated the question by weighting everything by capacity. This is important because there can be a large number of individual generators in any single year (skewing the overall fleet age), and in terms of electricity supply, overall capacity is more important than a raw number of generators. Raw generator numbers can be useful for predicting maintenance and other costs that might scale based on the number of units in service, but for the case of age a single 1,500 MW coal-fired steam plant is more important than 4 250 MW natural gas turbines. The question is now: What is the capacity-weighted age of our generators?

Finally, time. Form 860’s first operating generators appear in 1891 (hydropower in Portage, Wisconsin), and a trend is way more intriguing than the single point of today it adds context and you ultimately get to make a more interesting graph as well. Now I’m looking at a line plot, or something similar (already, and always thinking about step 6). This expands the question, and also complicates it. Each year needs to consider every year that came before, adding extra steps to the to-be-determined process. The final (unwieldy plot-title-ish) question: What is the capacity-weighted age of US generators in every year from 1891-2014?

3. Easy. I already had EIA Form 860 in mind as I defined the question, and that’s all I needed in this case.

4. The part that I tend to put off and mentally grumble about the most. Sometimes this is totally minimal (just calculating capacity factor or something else that’s basically just arithmetic), but in any case where I need to work with numbers in a loop as they’re being determined I probably need to sketch a workflow to keep track of things. This question totally fits that, and I couldn’t intuitively determine how to precisely solve my question for the data I wanted, so a write-up was required. In this case, I actually got distracted as soon as I started this phase. I solved for a different question than what I actually wanted and ended up with this graph.



I like the way it looks (and I discovered a new set of colors to use, via this publication), but it’s not only not what I wanted, it’s also wrong. I didn’t include any plant that had ever retired, only ones that had always been in operation since they entered service. I could have also totally done the calculations wrong. For example, at a glance, nuclear seems far too low.