Scott’s slides outline

Slide 1:

We were interested in environmental policy and after exploring a few different ideas, decided to do a project on renewable energy in California. We were aware that California had a set a goal of greatly increased renewable energy production by 2020 and wanted to explore what happened after this goal was set into law in 2011.

Slide 2:

With the state’s goal of 33% of energy coming from renewable resources, the central question was what progress was made toward that goal. We figured that in order to best understand what happened, we’d need to understand the contributions of different kinds of renewable energy. Because renewable energy is sometimes criticized for being at the whims of nature, we also wanted to look at how variations in time of day and year were impacting California’s renewable energy production. Finally, we wanted to also look at how much renewables contributed when there was maximum demand on the system.

Slide 3:

Renewable energy data is available in a few different places online, but not particularly easy to find in large chunks. CASIO provides the bulk of energy in California, but aside from limited snapshots, little was available on their website.

Fortunately, a Kaggle user had scrapped data from CAISO’s site and had an aggregate dataset that allowed us to really dive into the level of detail we were after.

We also turned to the USEIA for historic demand data specific to CAISO.

Slide 4:

In terms of challenges, I mentioned the trouble getting CAISO data. CAISO has an API that might have had what we were looking for, but neither Graham nor I were able to get approved for access despite online and phone attempts. It appears that they don’t want just anyone, or students at least, to gain access.

This in turn impacted the data we had to work with. The Kaggle set only had completed years of data for 2011-2017. Fortunately, our 8 years of data still contained much of the critical information we were looking for.

We had challenges with git repo management. There was nothing we could not resolve, but diverging branches slowed us down. The Jupyter checkpoint files seemed to be the source of our woes.

In terms of coding, CAISO changed the way they documented solar power in December 2012, when they went from tracking just “solar” to tracking solar thermal and solar photovoltaic separately. We filled in blank spots with zeros and created a new column to serve as a total combined value for all years.

Since time and date was key to our analysis, converting the data’s timestamp information into easy to call columns for hours, months, and years was key to making things smooth and we extensively used a groupby approach utilizing the broken out time categories. It also helped us merge datasets that had different timestamp formats.

Slide 10:

By 2017, California was getting closer to the goal of 33% renewable energy production, up from 13% in 2011 to about 25% in 2017. Renewables made their greatest contributions in the sunny spring months of April and May. While not shown here, the state proclaimed that it met its goal of 33% the following year primarily due to even greater contributions from solar.

Slide 11:

Our final question was how do renewables do during periods of peak demand and this slide shows the highest demand day in our dataset – September 1, 2017. The blue here shows the demand not met by renewables and so there is lots of room for renewables to make additional contributions during these periods.