Story of the Sun: Our goal was to let users explore development of solar photovoltaic projects state-by-state across the US.

* Data collection: Minimum level of granularity of data available from all sources was annual; thus, we treated it as a static dataset to be manipulated in Pandas.
  + EIA: Scraped electricity data from the table on the page
  + NREL: Downloaded large CSV of data
  + CDC: Downloaded a CSV
* Data source notes:
  + Solar resource is a very granular metric; while an apparently obvious correlation, there is not a reasonable method to aggregate via publicly available data sources
  + Skin cancer incidence rate as an interesting potential proxy for solar resource across a state… (plenty of confounding variables)
    - Think about how
  + Why size?
    - Smaller systems typically residential
    - Medium systems large residential or commercial
    - Large systems large commercial, ranging up to utility scale
* Used Pandas to pull all data into a single SQLite database with multiple tables
* Used SQLAlchemy to manipulate the database
* Combination of flask and