3 main data sources:

* Plant lvl hourly generation (US has ~1500 power plants) Continuous Emissions Monitoring System (CEMS) – Generation and Emissions
  + API to download hourly data
  + Gross generation: Total power generated, grid or pumped back into plant
    - Gross is problematic
  + Emissions data: CO2 (global pollutant), SO2, NOX (local pollutants)
  + Core Issue: Only for large enough plants – covers most plants in US but not all
  + Only Fossil fuel plants
* Balancing authority level hourly data generation by type
  + Balancing authorities can be multiple in one region
  + This data is net, not gross
  + Shows data from wind, solar, nuclear, hydro
  + Aggregate generation of fossil fuels (net)
* Residuals = 1-2 (net difference should be small)
* Weather data (control covariants)
  + Open Meteo
* Replicate data files using code from raw data
  + Learn how to streamline stuff in this regard