

## Input Data

The input data is **ca\_csi\_2020.pkl.zip**, a zipped pickle file. It contains one record for every CSI application, and it was last updated in early 2020. The full database contains 124 variables but for this exercise it has been trimmed down to thirteen: **"incentive"**, the dollar value of the incentive payments provided on the project; **"total\_cost"**, the total cost of the project; **"nameplate"**, the project's rated DC power output in kW; **"app\_status"**, the status of the application; **"sector"**, the host sector where the project was built (residential, commercial, etc.); **"county"**; **"state"**; **"zip"**; **"completed"**, the date the final incentive payment was sent; **"third\_party"**, a string indicating whether the owner of the solar array is not the host customer; **"inst\_status"**, the status of the system; **"type"**, a variable indicating whether the array tracks the sun or is fixed; and **"year"**, the year from the **"completed"** field. **firstlines.csv** provides a few lines from the file.

The script, **figures.py**, generates a range of visualizations to analyze certain aspects of the California Solar Initiative.