# **The goal of this project is to correlate child abuse deaths with median household income by state with ETL.**

## Extraction:

1. Child Abuse Prevention and Treatment Act Reports, 2011-2015 found at <https://www.propublica.org/datastore/dataset/child-abuse-prevention-and-treatment-act-reports-2011-2015>. It ends up as 20191212-CAPTA-reports-2011-2015.xlsx. This is the raw dataset of what kids where killed, where, and why.
2. Median income by state comes from <https://data.world/garyhoov/household-income-by-state> and ends up as household\_median\_income\_2017.csv
3. Dataset 1 has state abbreviations and dataset 2 has state names, so we bridge the gap with a lookup table found at [https://abbreviations.yourdictionary.com/articles/state-abbrev.html](https://abbreviations.yourdictionary.com/articles/state-abbrev.html%20) and was turned into a csv file manually as states.csv
4. It is obvious that more populated states will have more deaths, so we need to find the population from each state. This was gathered from <https://www.census.gov/data/datasets/time-series/demo/popest/2010s-state-total.html>. It was messy, so was fixed a bit with Excel and Textpad to end up as nst-est2019-01.csv.

Transform:

* The data only overlapped from 2011 to 2015, so subsetting was required.
* Data had to be transposed and renamed to be able to merge.
* Dataset one had only state abbreviations and dataset two had only state names, so a lookup and merge was necessary.
* Joins were done from all four datasets.
* Averages were done over the years in question, for income and population. Changing datatypes was necessary so that calculations could be done.
* There was messy data. Years were formatted as numbers, dashes, and ‘FY year’.

Load:

* Postgres was chosen because the data was fairly table-like, and it could be useful for future joins.
* The state\_abbreviation table was saved for further use in the states table
* The final results were saved in the child\_deaths table

## Conclusion – you do not want to move your kids to Arkansas:

