**Files Assumptions:**

An additional py file is needed (I did this real quick with Excel Power Query to reshape date) to transform “regression\_input.csv” to “final\_state.csv” and “final\_county.csv”

**Data structure for both file:**

final\_state.csv: id (first 2 character of county), Disease, Year, Summed\_Cost (summed cost group by state, year, disease), Weighted\_Prevalence (sum of (prevalence \* population) / sum of population, group by state, year, disease)

final\_county.csv: id, County, Cost, Disease, Year, Prevalence, Children\_with\_Disease (population)

**Graph Assumptions:**

1. I used weighted\_Prevalence for color scale.
2. Tooltip shows top 5 counties per state sorted by prevalence

**Other Questions:**

1. I did not see a “regression\_output.csv”, but we can just plug in the new files later with same columns
2. Some states are missing records (no records at all), some states are “missing data”, especially some have cost but not prevalence in “regression\_input.csv” such as California. However I did find prevalence data in the 100 raw data files, Alex could you look into this?