

DataEng: Data Integration Activity

Genevieve LaLonde
Data Engineering Winter 2021
Bruce Irvin
27 Feb 2021

A. Aggregate Census Data to County Level

Your integration will use two different dimensions: location (as indicated by state and county) and time. You should greatly simplify your processing and reduce your time by pre-processing your data along each of these dimensions.

The ACS data is separated into “Census Tracts” which are regions within counties that correspond to groups of approximately 4000 people. The Census Bureau defines these to help organize the actual job of collecting census data, but this grouping can make your Data Engineering job more more challenging. This level of detail is not needed for your county-level analysis, and you can greatly decrease your efforts by aggregating per-tract data to the county level.

Create a python program that produces a one-row-per-county version of the ACS data set. To do this you will need to think about how to properly aggregate Census Tract-level data into County-level summaries.

In this step you can also eliminate unneeded columns from the ACS data.

Question: Show your aggregated county-level data rows for the following counties: Loudon County Virginia, Washington County Oregon, Harlan County Kentucky, Malheur County Oregon

From the Data Storage data from last week for years 2015 and 2017:

```
year | state | county | totalpop | men | women | hispanic | white | black | native |
asian | pacific | citizen | income | incomeerr | incomepercap | incomepercaperr | poverty
| childpoverty | professional | service | office | construction | production | drive | carpool |
transit | walk | othertransp | workathome | meancommute | employed | privatework |
publicwork | selfemployed | familywork | unemployment
```

```
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
```

[illegible]

6.5	19.4	10824	73.1	17.5	8.8	0.6
8.3						
Oregon	Washington	572071	282381	289690	16.4	67.1 1.9
0.3	9.7	0.4	384659	78199	10188	35369 4079 10.3
13.4		44.9	15.5	23.0	6.9	9.7 73.1 10.3 6.3 2.4
1.9		6.1		25.0	292979	84.5 9.6 5.7
0.1		5.6				
Virginia	Washington	54457	26857	27600	1.5	95.4 1.4 0.0
0.6	0.0	43820	46234	7057	25640	3322 14.8 20.6
	33.7	18.8	24.4	7.0	16.2	83.1 9.1 0.2 1.5 1.2
4.9		24.0	23980	78.6	15.5	5.8 0.1
6.4						
(5 rows)						