Making Sense of Census Data

Don Perkus
12/14/16
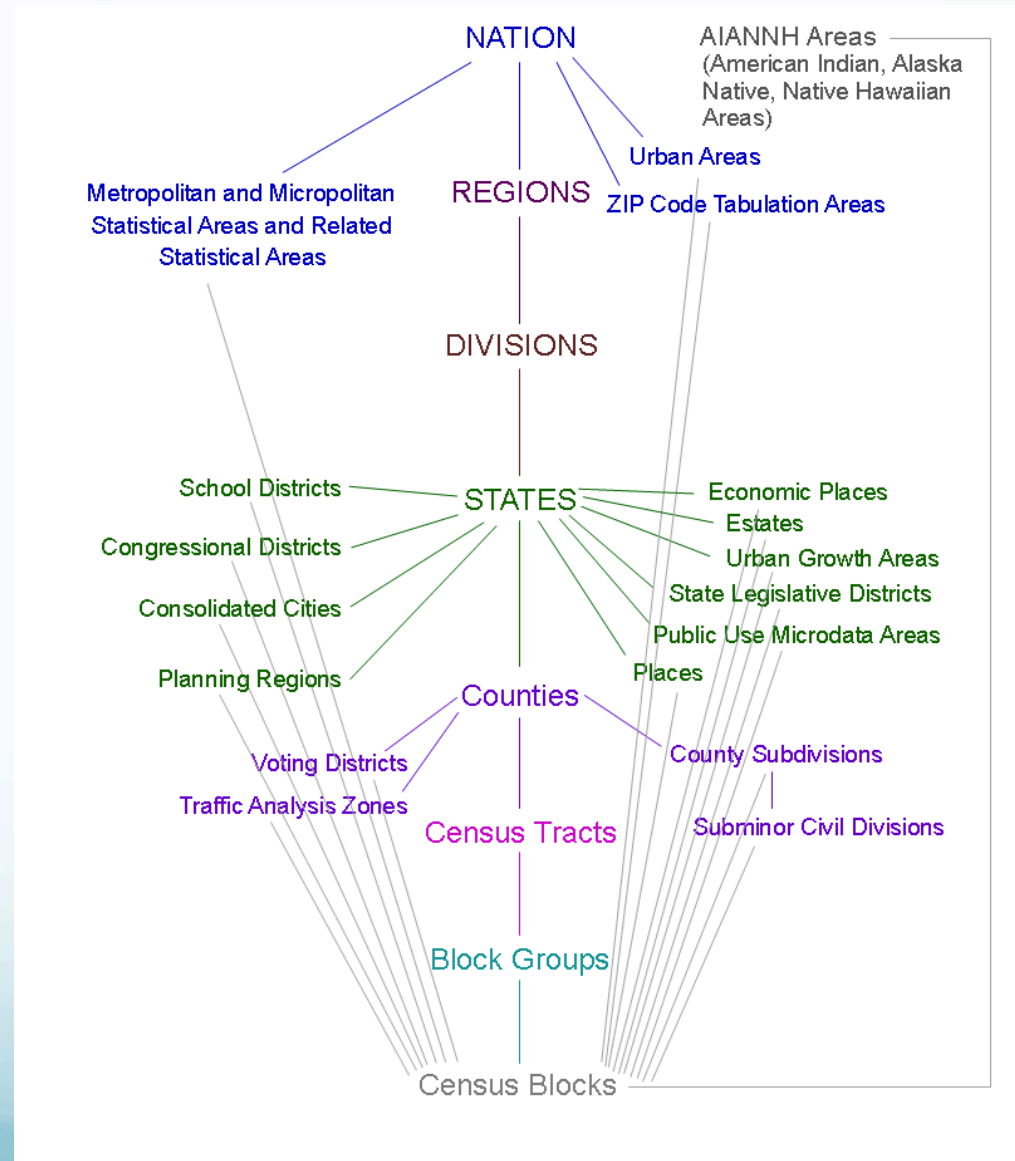
Intro

- Clustering model of our web site users

- Needed demographic data

- Very little demographic data available in-house.

- US Census == "Problem Solved" ???

- HUGE amount of data available
  - Various methods, APIs, permutations

- It is not easy to figure out how to locate and obtain the "right" set of data.

Lots of data

- over 130 different surveys and programs - thousands of data variables

- Lots of different crosstabs
  - Example
    - Education high school degree, college / grad degree
    - By poverty Rate
    - By Earnings
    - By ... etc ...
    - And always by gender

Census Geographic Hierarchy

4

American Fact Finder
"Popular Tables" to the rescue!

- SELECTED SOCIAL CHARACTERISTICS - DP02
  - Language and place of birth

- SELECTED ECONOMIC CHARACTERISTICS - DP03
  - Employment and Income

- SELECTED HOUSING CHARACTERISTICS - DP04
  - Renters and home owners

- DEMOGRAPHIC AND HOUSING ESTIMATES - DP05
  - Ethnicity

- EDUCATIONAL ATTAINMENTS – S1501
  - High school and college grads
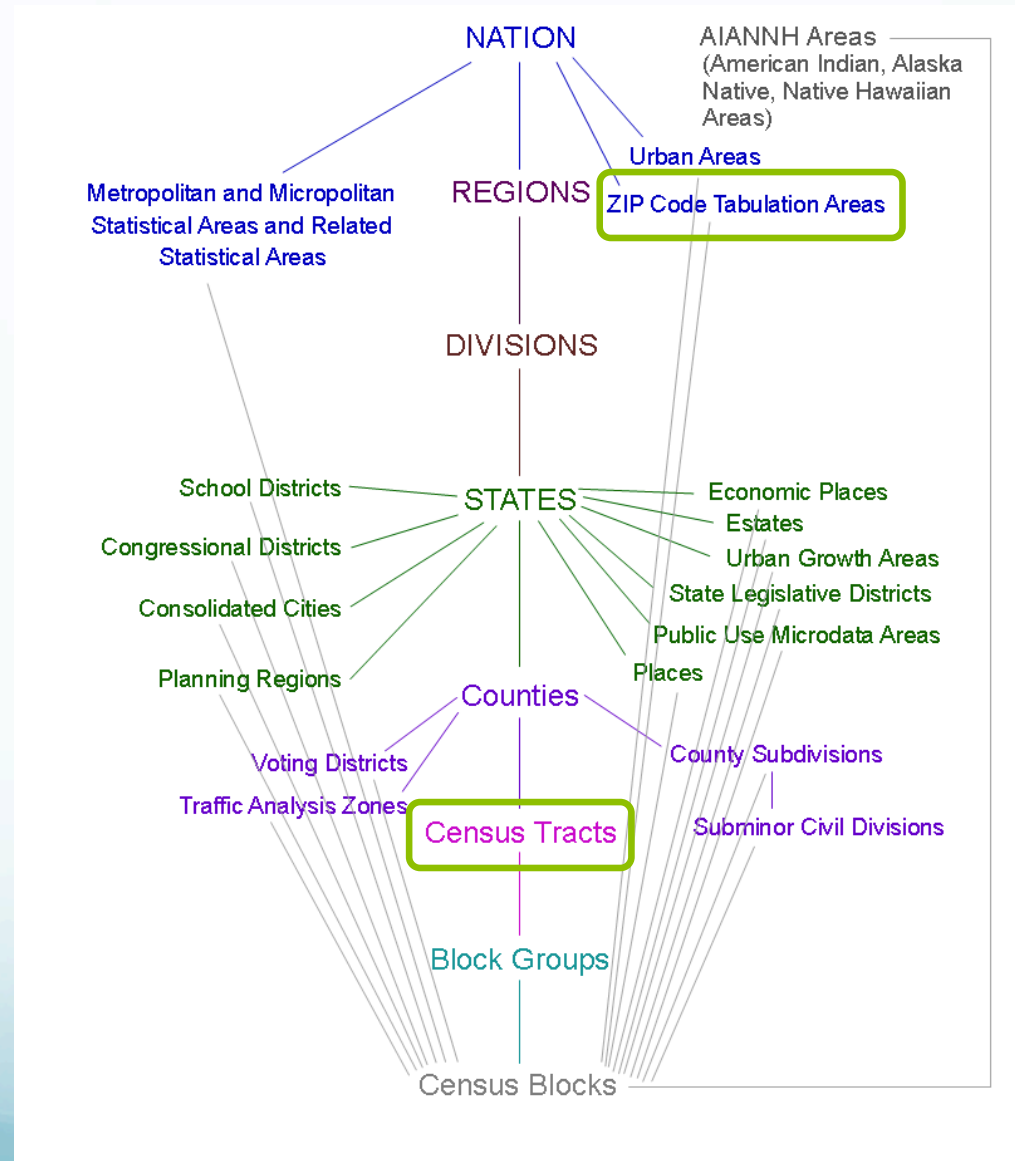
Educational Attainment – Table S1501

Versions of this table are available for the following years:

2015
2014
2013
2012
2011
2010
2009

	United States											
	Total		**Percent**		**Males**		**Percent Males**		**Females**		**Percent Females**	
Subject	**Estimate**	**Margin of Error**	**Estimate**	**Margin of Error**	**Estimate**	**Margin of Error**	**Estimate**	**Margin of Error**	**Estimate**	**Margin of Error**	**Estimate**	**Margin of Error**
Population 18 to 24 years	31,368,674	+/-10,442	(X)	(X)	16,071,993	+/-6,048	(X)	(X)	15,296,681	+/-5,798	(X)	(X)
Less than high school graduate	4,503,448	+/-20,111	14.4%	+/-0.1	2,647,327	+/-12,420	16.5%	+/-0.1	1,856,121	+/-11,156	12.1%	+/-0.1
High school graduate (includes equivalency)	9,321,843	+/-21,965	29.7%	+/-0.1	5,250,852	+/-15,435	32.7%	+/-0.1	4,070,991	+/-14,853	26.6%	+/-0.1
Some college or associate's degree	14,459,475	+/-25,050	46.1%	+/-0.1	6,894,271	+/-16,940	42.9%	+/-0.1	7,565,204	+/-17,623	49.5%	+/-0.1
Bachelor's degree or higher	3,083,908	+/-22,885	9.8%	+/-0.1	1,279,543	+/-11,448	8.0%	+/-0.1	1,804,365	+/-14,738	11.8%	+/-0.1
Population 25 years and over	211,462,522	+/-16,135	(X)	(X)	102,007,511	+/-8,681	(X)	(X)	109,455,011	+/-8,870	(X)	(X)
Less than 9th grade	12,093,869	+/-57,253	5.7%	+/-0.1	6,003,728	+/-30,011	5.9%	+/-0.1	6,090,141	+/-29,965	5.6%	+/-0.1
9th to 12th grade, no diploma	16,135,225	+/-65,476	7.6%	+/-0.1	8,318,466	+/-34,541	8.2%	+/-0.1	7,816,759	+/-35,633	7.1%	+/-0.1
High school graduate (includes equivalency)	58,722,528	+/-173,641	27.8%	+/-0.1	28,952,834	+/-94,591	28.4%	+/-0.1	29,769,694	+/-84,797	27.2%	+/-0.1
Some college, no degree	44,529,161	+/-49,403	21.1%	+/-0.1	21,120,934	+/-34,153	20.7%	+/-0.1	23,408,227	+/-34,185	21.4%	+/-0.1
Associate's degree	17,029,467	+/-40,247	8.1%	+/-0.1	7,300,637	+/-22,146	7.2%	+/-0.1	9,728,830	+/-26,800	8.9%	+/-0.1
Bachelor's degree	39,166,047	+/-136,480	18.5%	+/-0.1	18,721,504	+/-66,712	18.4%	+/-0.1	20,444,543	+/-72,897	18.7%	+/-0.1
Graduate or professional degree	23,786,225	+/-147,082	11.2%	+/-0.1	11,589,408	+/-72,475	11.4%	+/-0.1	12,196,817	+/-76,750	11.1%	+/-0.1
Percent high school graduate or higher	(X)	(X)	86.7%	+/-0.1	(X)	(X)	86.0%	+/-0.1	(X)	(X)	87.3%	+/-0.1
Percent bachelor's degree or higher	(X)	(X)	29.8%	+/-0.1	(X)	(X)	29.7%	+/-0.1	(X)	(X)	29.8%	+/-0.1

Census Geographic Hierarchy

From https://www.census.gov/geo/reference/webatlas/

Geocoding

- Address => geographic identifier

- Address to Zip Code is easy

- Census Tract provides finer grain

- But, Address to census tract is not so easy
  - Many free sources are volume limited
  - Our vendor could not provide "sub-tracts"
    - 12345.~~01~~

Extract and Transform

- S1501 Table for Zip codes
  - 100Mb = 800 columns by 33,000 rows

- S1501 Table for Census Tracts
  - 225Mb = 800 columns by 70,000 rows

- Only a few dozen are of interest

- SMOP*
  - 5 Tables
  - Extract useful columns
  - Handle missing data, clean up

* Simple Matter of Programming ☺

9

Resources available on my Github

- https://github.com/dperkus/CensusMakingSense

- Instructions for downloading tables

- My list of selected data variables

- Pseudocode

- Thank you!