Interpreting the Tables

In problem 2 we are providing you with data from the American Community Survey (ACS) of 2021. Specifically, data pertaining to North Carolina.

The competition data is composed of 14 files:

- “Table_Index.xlxs” - gives descriptions of data variables,
- “NC_Pharmacies.xlxs” - gives the names and locations of pharmacies in NC
- “spatial.csv” - gives locations of census tracts, and
- 11 tables of census data labeled Table 1 - Table 11

Below are a few notes on how this data should be interpreted.
For more information on how the ACS data is recorded see:
https://www2.census.gov/programs-surveys/acs/tech_docs/subject_definitions/2022_ACSSubjectDefinitions.pdf

GEOID

We are providing data on a census tract level. When the ACS is conducted responses are grouped at multiple levels: including, state, county and, census tract. Each census tract has a unique identifier called a *GEOID.* This ID number has the form *ssccctttttt* where *ss* = state ID, *ccc* = county ID and *tttttt* = tract ID.

For example: 37001020100.

The first column of all tables is GEOID.

Spatial.csv

This table gives Latitude and Longitude coordinates of all census tracts.

Census Tables

Census tables should be interpreted as histograms of the population against either two or three characteristics. For example, take Table 1 showing the distributions of Sex by Age.

Each row of the table is for a specific census tract identified by its GEOID. The columns represent specific variables collected in the ACS, which are described in ‘Table_Index.xlxs’. Table 1 denotes the number of people in each sex and age category, as well as their respective totals. This can be summarized as:

Age	Male	Female
Below 5	B01001_003E	B01001_027E
5-9	B01001_004E	B01001_028E
10-14	B01001_005E	B01001_029E
15-17	B01001_006E	B01001_030E
18-19	B01001_007E	B01001_031E
20	B01001_008E	B01001_032E
21	B01001_009E	B01001_033E
22-24	B01001_010E	B01001_034E
25-29	B01001_011E	B01001_035E
30-34	B01001_012E	B01001_036E
35-39	B01001_013E	B01001_037E
40-44	B01001_014E	B01001_038E
45-49	B01001_015E	B01001_039E
50-54	B01001_016E	B01001_040E
55-59	B01001_017E	B01001_041E
60-61	B01001_018E	B01001_042E
62-64	B01001_019E	B01001_043E
65-66	B01001_020E	B01001_044E
67-69	B01001_021E	B01001_045E
70-74	B01001_022E	B01001_046E
75-79	B01001_023E	B01001_047E
80-84	B01001_024E	B01001_048E
85 and above	B01001_025E	B01001_049E
Total	B01001_002E	B01001_026E

Note the final row are totals, meaning the sum of all previous entries in the column. Also the variable B010001_001E is the total number of individuals in the table, and is given by the sum of B01001_002E and B01001_026E.

Other tables are written in a similar way, although some are more complicated.

Please note that we **do not** expect you to use all the available data.

A note on how the data was assembled…

For anyone who is interested, I have included a Jupyter notebook that uses the ACS and Census API to generate these tables. Along with this I have included a *conda* environment file with all the dependencies needed to run the Jupyter notebook on my laptop.

(Note if you want to use the spatial data: I also had to install *rust* and then use pip to install the python *us* library.)

You have been given plenty of data to work on this problem already. It is **NOT** expected that you retrieve additional data.

A full example of how to use the API can be found here:
https://pygis.io/docs/d_access_census.html