Chapter 1 Notes

Table of contents

# Chapter 1: The United States Census and the R programming language

## 1.1 Census data: an overview

* Census - complete enumeration of the population
  + in the US, completed every 10 years, per Constitution
    - Intended (in the US) for *apportionment* - ensures representation in House of Representatives
* Since 2010, **American Community Survey** used to get more detailed demographic info
  + ACS sent to ~3% of population annually
  + Released in two public formats:
    - 1-year ACS (covers populations of 65,000 or greater)
    - 5-year ACS (moving average of data, covering down to the block group)
  + ACS is an estimate, with margins of error
* Census conducts other surveys including, but not limited to:
  + economic/business surveys
  + housing surveys
  + international data
  + population estimates/projections

## 1.2 Census hierarchies

* *Enumeration units*: geographies at which data are tabulated
  + *legal entities* (states and counties)
  + *statistical entities* standardized geographies for tabulation
  + Smallest unit for census is the *block*
  + Smallest unit for **ACS** is *block group* - a collection of blocks
* *Enumeration units* are nested:
  + Block groups are **fully** composed of blocks
  + Census tracts are **fully** composed of block groups
  + Some units don’t nest into a parent geography smaller than the nation as a whole - this means they can cross the boundaries of units (ZIP Code Tabulation Areas can cross city, county, even State boundaries)

## 1.3 How to find US Census Data

1. Data downloads from [Census Bureau](https://data.census.gov)
2. Download of bulk data through the [Census FTP site](https://www2.census.gov/programs-surveys/)
3. Utilize the Census API, available through <https://www.census.gov/data/developers/data-sets.html>
   1. Census APIs have an API endpoint - a reference url for a specific data set
   2. A query is appended to the end of the url, following a ?
   * Example API endpoint for 2010 Decennial Census: https://api.census.gov/data/2010/dec/sf1
   * Example API query: ?get=P001001,NAME&for=county:\*&in=state:06
   * This is completed as https://api.census.gov/data/2010/dec/sf1?get=P001001,NAME&for=county:\*&in=state:06 which returns a JSON-formatted stream of the data
4. Third-party data distributors such as:
   1. The NHGIS
   2. Census Reporter from Northwestern University’s Knight Lab
   3. Social Explorer (commercial product)

## 1.4 What is R?

## 1.5 Analyses using R and US Census data

### Census data packages in R

1. **totalcensus:** download and load bulk Census data
2. **PL94171:** process and load PL-94171 redistricting files
3. **acs: *“****Download, Manipulate, and Present American Community Survey and*

\*Decennial Data from the US Census"\*

1. **choroplethr:** map production with census data
2. **censusapi:** Retrieve Data from the Census APIs
3. **tidycensus:** focus of the book; focused on retrieving data from specific key API endpoints to return spatially-contextualized census data
4. **tigris:** focus of the book;downloads TIGER/Line shapefiles from Census Bureau for spatial analysis in R
5. **ipumsr:** access third-party Census data from Minnesota Population Center datasets such as the National Historical GIS

### Health resource access

* Hannah Recht (developer of **censusapi**) produced a study of stroke care access in the Mississippi Delta and in Appalachia; code for the analysis available [here](https://www.github.com/khnews/2021-delta-appalachia-stroke-access)

### COVID-19 and pandemic response

* Matt Worthington at University of Texas LBJ School of Public Affairs built the **Texas COVID-19 Vaccine Tracker**
  + The application was built using several r tools, including Shiny, to produce an interactive dashboard
  + Source code is available [here](https://github.com/utexas-lbjp-data/t_vaccine_site)
  + Application can be used [here](https://texasvaccinetracker.com)
    - The interactive elements seem to be mostly non-functional at this state

### Politics and gerrymandering

* Harvard’s [ALARM project](https://alarm-redist.org) develops numerous r packages for redistricting analysis
  + **PL94171** pulls 2020 redistricting data
  + **geomander** prepares data for redistricting analysis
  + **redist** generates and evaluates potential redistricting solutions

### Social equity research

* Jerry Shannon published a study in 2020 using R and Census data to study dollar store geography
  + the study finds dollar stores are more likely to be sited in Black and Latino neighborhoods even when controlling for other structural and economic drivers, due to continuing legacies of racial segregation.
  + The study uses NHGIS data
  + analytic code is available [here](https://github.com/jshannon75/metrodollars)
  + the paper is available [here](https://doi.org/10.1080/24694452.2020.1775544)

### Census data visualization

* the book author, Kyle Walker, developed and published [Mapping Immigrant America](personal.tcu.edu/kylewalker/immigrant-america/)
  + The back-end data processing was done entirely in R
  + The author published the data through Mapbox Studio
  + Source code for data processing is [here](hithub.com/walkerke/mb-immigrants)