Project Data Report

Lorenz Menendez

# Introduction

Isaac and I separated the data wrangling into two equal parts. Isaac worked on cleaning and wrangling election result data, which he will describe in his data report. I worked on wrangling county boundary data with demographic and social data. Here are the input datasets I worked with.

# Input Datasets

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Name*** | *Description* | *Variables* | *Temporal Resolution* | *Spatial Resolution* | *File Format* |
| **County Facts** | Demographic and social data for each county in the U.S. | FIPS code, Name, State, Population, Ages, Sex, Race | 2010-2014 | County | .csv |
| **County Boundaries** | Multipolygon GIS data representing county boundaries | State & County FIPS code, GEOID, Name, Land Area, Water Area | 2016 | County | .shp |

I developed an R script to automatically to download, process and import clean data into R. You can view it [here](https://github.com/LorenzMenendez/GIS3-Final/blob/master/Data%20Wrangling/Kaggle%20and%20Census%20Data%20-%3E%20Single%20Dataset.R). The R script takes the input datasets and joins them together based on FIPS codes to create a master dataset called ‘electnData16’.

I tried my best to have R import from URLs so that it’s very reproducible. Unfortunately, using a simple URL is not working for the County Facts dataset. I think I’ll have to eventually setup a Kaggle API to automatically import that data into R.

# Output Datasets

# Sources

**County Facts**

Hamner, Ben. “2016 US Election.” Kaggle, 1 July 2016, www.kaggle.com/benhamner/2016-us-election.

**County Boundaries**

US Census Bureau. “Cartographic Boundary Files - Shapefile.” Cartographic Boundary Files - Shapefile, 2 May 2019, www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html. Downloaded: cb\_2018\_us\_county\_500k.zip