EDS241: Assignment 1

Joe DeCesaro

01/17/2022

1 Assignment 1

The data for this assignment come from CalEnviroScreen 4.0, a mapping and data tool produced by the California Office of Environmental Health Hazards Assessment (OEHHA). The data are compiled and constructed from a variety of sources and cover all 8,035 census tracts in California. Source: https://oehha.ca.gov/calenviroscreen/report/calenviroscreen-40

The full data are contained in the file CES4.xls, which is available on Gauchospace (note that the Excel file has three "tabs" or "sheets"). The data is in the tab "CES4.0FINAL_results" and "Data Dictionary" contains the definition of the variables.

For the assignment, you will need the following variables: **CensusTract, TotalPopulation, CaliforniaCounty (the county where the census tract is located), LowBirthWeight (percent of census tract births with weight less than 2500g), PM25 (ambient concentrations of PM2.5 in the census tract, in micrograms per cubic meters), and Poverty (percent of population in the census tract living below twice the federal poverty line)**.

1.1 Clean data

The following code loads and cleans the data.

```
# Read in the first sheet and clean up
data_sheet1 <- read_xlsx(here("CES4.xlsx"), sheet = 1) %>%
  clean_names() %>%
  select(census_tract, total_population, california_county,  low_birth_weight, pm2_5, poverty)
```

1.2 a) What is the average concentration of PM2.5 across all census tracts in California?

The average concentration of PM2.5 across all census tracts in California is 10.15