

1 Introduction

Install the `tidycensus` and `tigris` packages for R `install.packages(c("tidycensus", "tigris"))`. These packages include tools for collecting and mapping data from the US Census.

2 Collecting Data

2.1 Electric Vehicle (EV) Charging Stations

You can download data about electronic vehicle (EV) charging stations in the United States [here](#).

We are interested in New York, so we can remove any observations outside of New York State (`State == "NY"`). These data will be merged with Census data, so we will want to map these observations to Census-level geographies. To do this we will use the `call_geolocator_latlon()` function from the `tigris` library. This function takes latitude and longitude entries and returns a number that describes the Census locations. You may find an observation is missing a latitude or longitude entry. If necessary, you can manually edit the data by looking up the information using Google Maps. Make sure to fully document any changes.

The 15 digits returned by the `call_geolocator_latlon()` function represent the following characteristics.

- Geography ID (GEOID): Characters 1-11
- State: Characters 1-2
- County: Characters 3-5
- Tract: Characters 6-11
- Block: Characters 12-15

Use the `str_split()` function from the `stringr` package or the base `substring()` base function to create these variables in the data frame.

Finally, we will focus our analysis on New York City. We can separate that data out by keeping observations with a county corresponding to one of the five Boroughs.

- Bronx County = 005
- Kings County (Brooklyn) = 047
- New York County (Manhattan) = 061
- Richmond County (Staten Island) = 085
- Queens County = 081

2.2 Census Data

Ask for help about the `census_api_key()` function. Note there are instructions for getting your own census api key, which you shouldn't share. When you write up your lab report you can dedicate a chunk of code to defining your key, but hide it with the `echo=FALSE` option.

```
> library(tidycensus)
> library(tigris)
> ?census_api_key
```

Once you have your census key, make sure to specify it using the `census_api_key()` function.

Use the help for `get_acs()` and resources from the web to pull American Community Survey data from the US Census at the tract level. Specifically, you will want to pull the following variables:

- Median Income (B06011_001E)
- Mean Income (S1902_C03_001E)
- Total Population (B02001_001E)
- Population that identifies as Black or African American Alone (B02001_003E)

Preview the data and note the strange setup. Each tract has multiple rows – one for each variable – and the variable is listed in the `variable` column, while the `estimate` and margin of error (`moe`) are listed in the following columns. To get the data into a style we are more familiar with we have to reshape it. Assuming you've named your data frame `acs.dat` you can reshape your data as follows.

```
> acs.dat <- acs.dat %>%
+   select(-moe) %>%      # remove the margin of error column
+   pivot_wider(values_from = "estimate", names_from = "variable") # reshape file
```

View the data and notice the changes. Why is this type of reshaping necessary? See Hadley Wickhams chapter about tidy data principles [here](#).

Now, create a new column that contains the percent of the population that identifies as Black or African American.

3 Merge Data

First, we need to make the EV charging station data 'compatible' with the Census data. To do this, we need to summarize the data at the tract level. Use `tidyverse` to `group_by()` the `GEOID` and `summarize()` the number of EV charging stations (note you can ask for the number using `n` in the `summarize()` operation). Save the summary into a data frame object.

Use the `merge()` function from `tidyverse` to merge the EV charging station counts and the census data. Make sure to merge by `GEOID`, and that you retain all of the census data. That is, we don't want to drop observations in our census data simply because there are no EV charging stations there. You will likely find one of the `all.x` or `all.y` arguments to the `merge()` function helpful. Use `?merge` for helpful documentation.

4 Summarize Data

The stakeholders doing this research are particularly interested in how EV charging stations are distributed across New York City with respect to income and race. Create some tables (in \LaTeX , of course) and some plots (that are scaled, with labels and captions) that can help the stakeholders answer their questions. Don't forget the Sweave Cheatsheet on Moodle!

As you write your lab report, you may find popular press articles like `citeAxios` and academic articles like `citekhan2021` helpful. I found both by searching "disparity in EV charging stations" on Google for popular press articles and Google Scholar for academic articles.

If you'd like a challenge, you can include a map like those in `citekhan2021` in the appendix of your lab submission.