STAT 230 Homework 8

Ben Aoki-Sherwood

```
col_spec = cols(
  "HCFC-22 Production from HFC-23 Destruction" = col_double(),
  "Lead Production" = col_double(),
  "Petroleum and Natural Gas Systems - LNG Storage" = col_double(),
  "Phosphoric Acid Production" = col_double(),
  "Silicon Carbide Production" = col_double(),
  "Titanium Dioxide Production" = col_double()
)
states <- rbind(data.frame(name = state.name, abb = state.abb), c(name = "Puerto Rico", abb = "PR"))
```

```
state_unit <- Vectorize(function(state) {
  if (state == "LA") {
    "parish"
  } else if (state == "PR") {
    "municipio"
  } else {
    "county"
  }
})

special_tracts <- c("hopewell city", "anchorage municipality")

emissions <- read_csv("ghgp_data_2021.csv", skip=3, n_max = 6483, col_types = col_spec)
emissions <- emissions %>%
  drop_na(County) %>%
  rename_with(str_to_lower) %>%
  mutate(emissions.total = `total reported direct emissions`,
         industry.sector = `industry type (sectors)`,
         county = str_to_lower(county),
         county = ifelse((!str_detect(county, "county|municipio|parish") &
                          !(county %in% special_tracts)), paste(county, state_unit(state)), county)) %>%
  select(state, county, industry.sector, emissions.total)
```

get census data

```
census_data <- get_acs(geography = "county", variables = c("B02001_001E", "B19001_001E","B02001_002E"),
code_book <- rbind(c("B02001_001", "total_population"), c("B19001_001", "household_income"), c("B02001_
code_book <- as.data.frame(code_book)
colnames(code_book) <- c("variable", "var_name")
census_data <- left_join(census_data, code_book)

#format the data so there is a row for each census tract and column for every variable
acs_data <- maditr::dcast(census_data, NAME ~ var_name,
                                value.var = "estimate",
                                fun.aggregate = NULL) %>%
```

```
  rename(county = NAME) %>%
  separate(col = "county", into = c("county", "state"), sep = ", ") %>%
  left_join(states, by = c("state" = "name")) %>%
  mutate(state = abb, county = str_to_lower(county)) %>%
  select(-abb)
```

join data

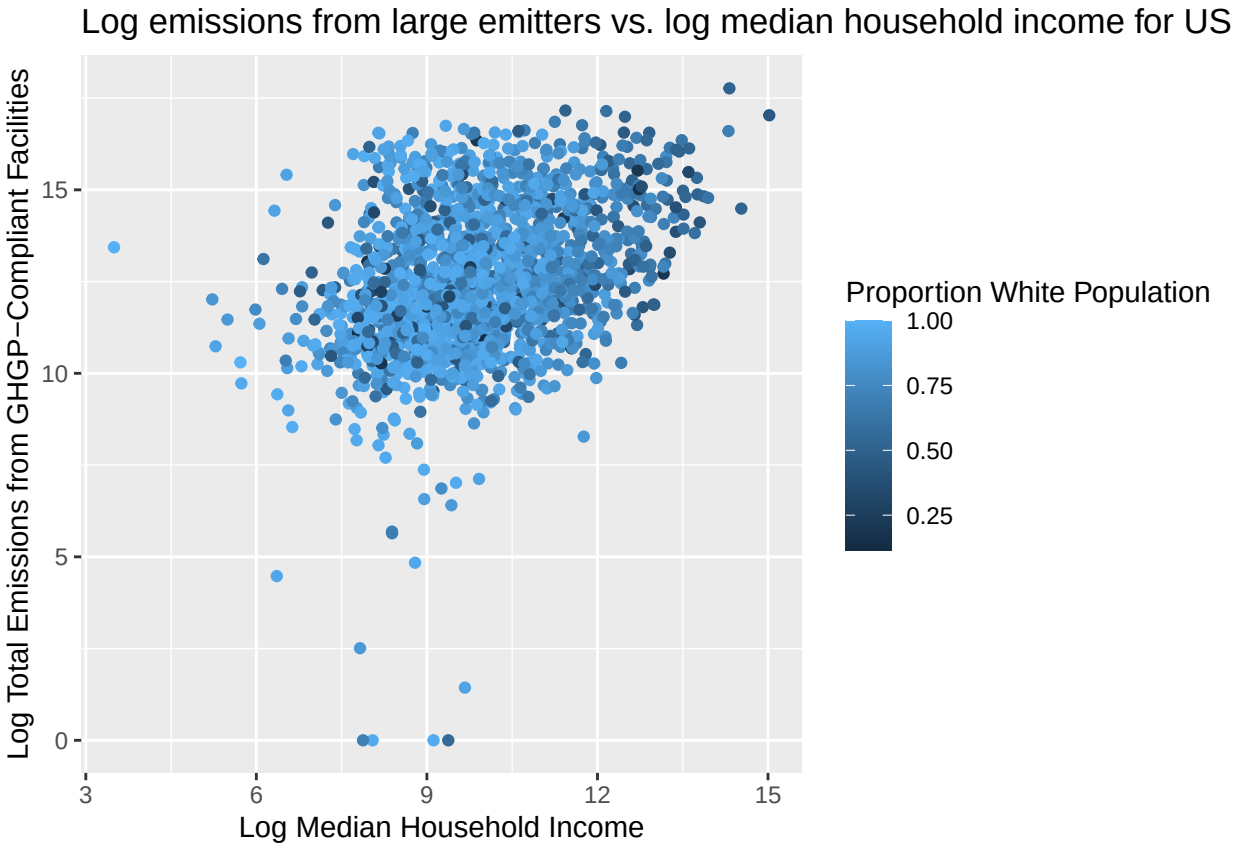
```
emissions.joined <- emissions %>% left_join(acs_data, by = c("state","county"))
head(emissions.joined)
```

```
## # A tibble: 6 x 7
##   state county        industry.sector emissions.total househol~1 total~2 total~3
##   <chr> <chr>         <chr>                     <dbl>      <dbl>   <dbl>   <dbl>
## 1 TX    collin county Waste                   314494.     369168 1039812  653729
## 2 KY    perry county  Other                   112349.      11252   28421   26988
## 3 NY    kings county  Power Plants             46082.     985108 2712360 1141959
## 4 IL    cook county   Waste                     7751.    2044658 5265398 2740032
## 5 TX    brown county  Minerals                 28243.      14651   38085   32042
## 6 MN    ramsey county Other                    82788.     215740  549377  349363
## # ... with abbreviated variable names 1: household_income, 2: total_population,
## #   3: total_white_population
```

```
county.emissions <- emissions.joined %>%
  group_by(county, state) %>%
  summarize(emissions = sum(emissions.total),
            household_income = first(household_income),
            total_population = first(total_population),
            total_white_population = first(total_white_population))
```

```
ggplot(county.emissions, aes(x = log(household_income), y = log1p(emissions), color = total_white_popula
  geom_point() +
  labs(x="Log Median Household Income", y="Log Total Emissions from GHGP-Compliant Facilities", color="
```

```
## Warning: Removed 55 rows containing missing values (`geom_point()`).
```


Log emissions from large emitters vs. log median household income for US
Log Total Emissions from GHGP−Compliant Facilities
0
5
10
15
3
6
9
12
15
Log Median Household Income
Proportion White Population
1.00
0.75
0.50
0.25