Internet Access in Illinois

Danya Sherbini

2022-08-30

```
library(tidyverse)
library(tidycensus)
library(viridis)
```

Introduction

Using Tidycensus

I'm going to examine internet access data from the U.S. Census. Below I activate the tidycensus API by using my own census key. Once activated, I identified some variables of interest related to household internet access, such as demographic and income data.

```
census_api_key("338b89135dc1066f5241fa4e16804fc7345d1822",
               install = TRUE,
               overwrite = TRUE)
```

```
## [1] "338b89135dc1066f5241fa4e16804fc7345d1822"
```

```
readRenviron("~/.Renviron")

v1 <- load_variables(2020, "acs5") #here I loaded variables for the 2020 ACS
```

Loading Internet Access Data

I'm loading data from the 2020 American Community Survey and filtering by Illinois counties.

```
internet_data_raw <- get_acs( #calling data from the 5-year ACS
  geography = "county", #geo by county
  variables = c(with_subscription = "B28002_002", #HHs with internet sub
                without_subscription = "B28002_012", #HHs w internet but no sub
                no_internet = "B28002_013", #HHs w/o internet
                hhs_total = "B28002_001", #total households
                pop_total = "B01003_001", #total population
                num_white = "B02001_002", #number white
                num_black = "B02001_003", #number Black
                num_asian = "B02001_005", #number Asian
                num_latinx = "B03003_003",#number Latinx
                median_hhs_income = "B19013_001"), #Median household income
```

```
  year = 2020, #filtering by year 2020
  state = "IL" #filtering by IL
)
```

Examining the Dataframe

```
nrow(internet_data_raw)
```

```
## [1] 1020
```

```
ncol(internet_data_raw)
```

```
## [1] 5
```

```
names(internet_data_raw)
```

```
## [1] "GEOID"    "NAME"     "variable" "estimate" "moe"
```

Data Cleaning & Manipulation

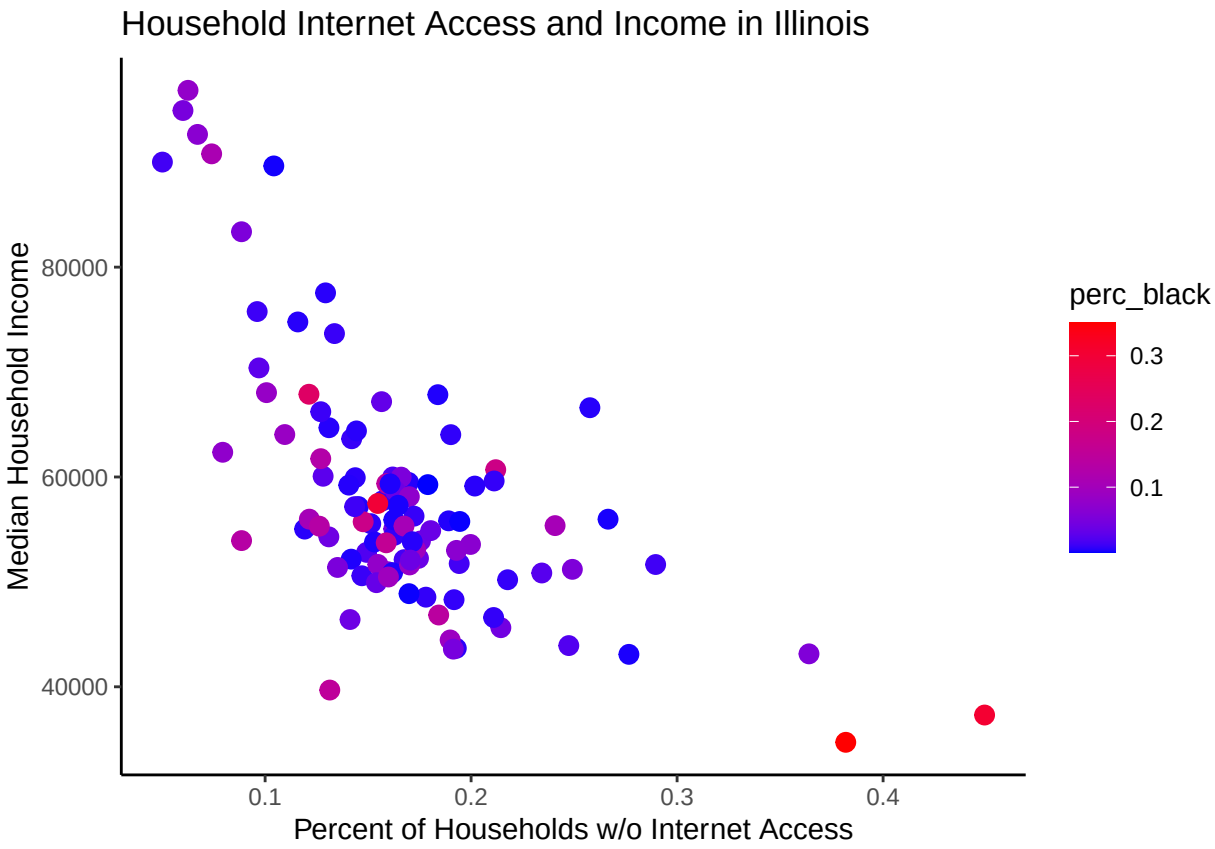
```
internet_data <- internet_data_raw %>%
  select(-moe) %>% #removing moe column
  spread(variable, estimate) #transforming these variables from rows to columns


#adding new columns for %
internet_data <- internet_data %>%
  mutate(perc_nointernet = no_internet / hhs_total,
         perc_withsubscription = with_subscription / hhs_total,
         perc_wo_subscription = without_subscription / hhs_total,
         perc_asian = num_asian / pop_total,
         perc_black = num_black / pop_total,
         perc_latinx = num_latinx / pop_total,
         perc_white = num_white / pop_total)
```

Data Visualization

The data viz below looks at the correlation between household median income, percent of households without internet access, and percent of Black residents by county.

```
internet_data %>%
  ggplot(aes(x = perc_nointernet, y = median_hhs_income,
             color = perc_black)) +
  geom_point(size = 3) +
  scale_color_gradient(low="blue", high="red") +
  labs(title="Household Internet Access and Income in Illinois", x="Percent of Households w/o Internet A
  theme_classic()
```


Household Internet Access and Income in Illinois
Median Household Income
80000
60000
40000
0.1
0.2
0.3
0.4
Percent of Households w/o Internet Access
perc_black
0.3
0.2
0.1