Assignment 1 Example

Carole Voulgaris

10/16/2021

Research question

How well do density and location relative to a metropolitan center predict county-level case-rates and death-rates from COVID-19 in the United States?

Prior research

Wong and Li (2020) find that population density is an effective predictor of county-level cumulative COVID-19 case rates, particularly after the earliest stages of the pandemic. In contrast, Hamidi, Sabouri, and Ewing (2020) find that, while total metropolitan population is a significant predictor of county-level COVID case rates, county population density is not.

Data

The sample population for this study is the full set of all counties in the United States. The analysis will include the following variables:

- Cumulative number of COVID-19 cases to date, as of October 15, 2021 (Dong et al. 2020)
- Cumulative number of COVID deaths to data, as of October 15, 2021 (Dong et al. 2020)
- Percent Republican in last presidential election (MIT Election Data and Science Lab, 2018)
- Majority vote in last presidential election (MIT Election Data and Science Lab, 2018)
- People per occupied housing unit (United States Census Bureau 2020)
- People per square mile (United States Census Bureau 2020)
- Median age (United States Census Bureau 2019)
- Urban-Rural County classification (National Center for Health Statistics 2013)

Load data

I'll be using the following libraries for this exercise:

```
library(tidyverse)
library(tidycensus)
library(readxl)
library(knitr)
```

First, I'll load the total population and total number of housing units, along with the county boundaries, from the decennial census, using the `tidycensus` package (Walker 2021).

```
census <- get_decennial(geography = "county",
                        year = 2020,
                        variables = c(pop = "P1_001N",
                                      HUs = "H1_002N"),
                        output = "wide",
                        geometry = TRUE)
```

I'll also use the same package to get the median age for each county from the 2019 American Community Survey.

```
acs_age <- get_acs(geography = "county",
                   variables = c(med_age_ = "B01002_001"),
                   output = "wide")
```

I'm loading COVID data directly from the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (Dong et al. 2020).

```
covid <- read_csv("https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_daily_reports/01-02-2021.csv") %>%
  filter(Country_Region == "US" &
           !is.na(Admin2)) %>%
  mutate(GEOID = case_when(str_length(as.character(FIPS)) == 5 ~
                             as.character(FIPS),
                           str_length(as.character(FIPS)) == 4 ~
                             paste("0", FIPS, sep=""),
                           TRUE ~ "not a county")) %>%
  filter(GEOID != "not a county") %>%
  select(Confirmed, Deaths, GEOID)
```

I've downloaded a spreadsheet with the urban-rural classifications from https://www.cdc.gov/nchs/data_access/urban_rural.htm (https://www.cdc.gov/nchs/data_access/urban_rural.htm) and saved it to the `data` file in my project folder. I'll load it here.

```
CO_type <- read_xlsx(path = "data/NCHSURCodes2013.xlsx",
                     sheet = "NCHSURCodes2013") %>%
  mutate(GEOID = case_when(str_length(as.character(`FIPS code`)) == 5 ~
                             as.character(`FIPS code`),
                           str_length(as.character(`FIPS code`)) == 4 ~
                             paste("0", `FIPS code`, sep=""),
                           TRUE ~ "unknown")) %>%
  mutate(type = case_when(`2013 code` == 1 ~ "Large central metro",
                          `2013 code` == 2 ~ "Large fringe metro",
                          `2013 code` == 3 ~ "Medium metro",
                          `2013 code` == 4 ~ "Small metro",
                          `2013 code` == 5 ~ "Micropolitan",
                          `2013 code` == 6 ~ "Non-core",
                          TRUE ~ "unknown")) %>%
  select(GEOID, type)
```

I've downloaded a csv file with election results from the Harvard Dataverse (https://doi.org/10.7910/DVN/VOQCHQ (https://doi.org/10.7910/DVN/VOQCHQ)) and saved it to the data file in my project folder. I'll load it here.

```
election <- read_csv('data/countypres_2000-2020.csv') %>%
  filter(year == 2020) %>%
  filter(party == "REPUBLICAN") %>%
  rename(GEOID = county_fips) %>%
  group_by(GEOID) %>%
  summarize(candidatevotes = sum(candidatevotes),
            totalvotes = first(totalvotes)) %>%
  mutate(pct_GOP = candidatevotes / totalvotes) %>%
  mutate(majority_vote = ifelse(pct_GOP > 0.5, "Republican", "Democrat")) %>%
  select(GEOID, pct_GOP, majority_vote)
```

Now that I have all my datasets loaded, I can join them all together and display the first few rows

```
data <- left_join(census, acs_age) %>%
  left_join(election) %>%
  left_join(CO_type) %>%
  left_join(covid)

kable(head(data))
```

GEOID	NAME	pop	HUs	geometry	med_age_E	med_age_M	pct_GOP	majority_vote	type	Confirmed	Deatl
21141	Logan County, Kentucky	27432	11000	MULTIPOLYGON (((-87.06037 3...	40.7	0.6	0.7344079	Republican	Non-core	1737	3
36081	Queens County, New York	2405464	847210	MULTIPOLYGON (((-73.96262 4...	39.0	0.2	0.2676722	Democrat	Large central metro	128917	758
34017	Hudson County, New Jersey	724854	289408	MULTIPOLYGON (((-74.0422 40...	35.3	0.1	0.2629498	Democrat	Large central metro	48440	17

GEOID	NAME	pop	HUs	geometry	med_age_E	med_age_M	pct_GOP	majority_vote	type	Confirmed	Deatl
34019	Hunterdon County, New Jersey	128947	48978	MULTIPOLYGON (((-75.19511 4...	46.3	0.1	0.5122504	Republican	Large fringe metro	4251	1:
21147	McCreary County, Kentucky	16888	6058	MULTIPOLYGON (((-84.77845 3...	39.4	1.4	0.8801865	Republican	Non-core	1121	
21195	Pike County, Kentucky	58669	24534	MULTIPOLYGON (((-82.73269 3...	42.5	0.4	0.7993929	Republican	Non-core	3098	:

The dataset includes 3221 counties.

References

Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Inf Dis. 20(5):533-534. doi: 10.1016/S1473-3099(20)30120-1

Hamidi, Shima, Sadegh Sabouri, and Reid Ewing. “Does density aggravate the COVID-19 pandemic? Early findings and lessons for planners.” Journal of the American Planning Association 86, no. 4 (2020): 495-509.

MIT Election Data and Science Lab, 2018, “County Presidential Election Returns 2000-2020”, https://doi.org/10.7910/DVN/VOQCHQ (https://doi.org/10.7910/DVN/VOQCHQ), Harvard Dataverse, V9, UNF:6:qSwUYo7FKxI6vd/3Xev2Ng== [fileUNF]

National Center for Health Statistics. “NCHS Urban-Rural Classification Scheme for Counties” 2013. https://www.cdc.gov/nchs/data_access/urban_rural.htm (https://www.cdc.gov/nchs/data_access/urban_rural.htm)

United States Census Bureau. American Community Survey, 5-year estimates. 2019.

United States Census Bureau. Redistricting Data. 2020.

Walker, Kyle, and Matt Herman (2021). tidycensus: Load US Census Boundary and Attribute Data as ‘tidyverse’ and ‘sf’-Ready Data Frames. R package version 1.1. https://CRAN.R-project.org/package=tidycensus (https://CRAN.R-project.org/package=tidycensus)

Wong, David WS, and Yun Li. “Spreading of COVID-19: Density matters.” Plos one 15, no. 12 (2020): e0242398.