Coronavirus Second Wave - Synthetic New York City

Tyler DeGroff

6/23/2020

```r
rm(list = ls())
library(tidyverse)

data <- read.csv(
  'https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv'
)

data$date <- as.Date(data$date)

data <- data %>%
  filter(
    state == "New York",
    county != "Unknown") %>%
  select(-state)

data <- within(data, fips[county == "New York City"] <- 99999)

population <- read.csv(
  "https://data.ny.gov/api/views/krt9-ym2k/rows.csv?accessType=DOWNLOAD&sorting=true"
)

population <- population %>%
  filter(
    Year == 2019,
    Geography != "New York State"
  ) %>%
    select(
    -Year,
    -Program.Type
  ) %>%
  rename(
    fips = FIPS.Code,
    county = Geography,
    population = Population
  )

population.nyc <- population %>%
  filter(
    fips == 36005 | # Bronx County or Bronx
    fips == 36047 | # Kings County or Brooklyn
    fips == 36061 | # New York County or Manhattan
    fips == 36081 | # Queens County or Queens
    fips == 36085   # Richmond County or Staten Island
  )

population <- rbind(
  population,
```

```r
  data.frame(
    fips = 99999,
    county = "New York City",
    population = sum(population.nyc$population)
  )
)

data <- merge(
  x = data,
  y = population %>% select(fips, population),
  by = "fips",
  all.x = TRUE
)

synthetic <- data.frame()
date_unique <- unique(data$date)

for (i in 1:length(date_unique)) {

  temp <- data %>% filter(
    date == date_unique[i],
    fips == 99999 # where county is "New York City"
  )

  synthetic <- rbind(
    synthetic,
    data.frame(
      fips = c(36005, 36047, 36061, 36081,36085),
      date = temp$date[1],
      county = c(
        "Bronx County",     # Bronx
        "Kings County",     # Brooklyn
        "New York County",  # Manhattan
        "Queens County",    # Queens
        "Richmond County"   # Staten Island
      ),
      cases = rep(temp$cases[1], 5),
      deaths = rep(temp$deaths[1], 5),
      population = rep(temp$population[1], 5)
    )
  )

}

data <- rbind(data, synthetic) %>% filter(county != "New York City")

write.csv(data, "data.syntheticNYC.csv")
```