Coronavirus Second Wave

Tyler DeGroff

6/26/2020

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

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

data <- data %>%
  filter(county != "Unknown") %>%
  mutate(date = as.Date(date))

data <- within(data, fips[county == "New York City"] <- 99999) # treats NYC as a county

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

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

# county-level cases and deaths per capita or per 100K data only valid for New York State
# county-level population data was not imported for any other state. NAs result.

data <- data %>%
  mutate(
    cases.percap = cases / population,
    cases.per100k = cases / (population / 100000),
    deaths.percap = deaths / population,
    deaths.per100k = deaths / (population / 100000)
  )
data <- data[order(data$date, data$state, data$county), ]
states <- sort(unique(data$state))
temp <- data.frame()

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

  counties <- sort(unique(data %>% filter(state == states[i]) %>% select(county))[, 1])

  for (j in 1:length(counties)) {

    temp2 <- data %>%
      filter(
        state == states[i],
        county == counties[j]
      ) %>%
      mutate(
        cases.new = c(cases[1], diff(cases)),
        deaths.new = c(deaths[1], diff(deaths))
      )

    temp <- rbind(temp, temp2)

    print(paste(states[i], "->", counties[j]))

  }

}

data <- temp
#rm(list = c("i", "j", ls(pattern = "temp")))

write_csv(data, "data.csv")
```