Geography 176A

Lab 02: COVID-19 Pandemic

[Abigail Porter](https://ucsbaporter.github.io/UCSBaporterW1/first-webpage/index.html)

## COVID-19 Data

### Data

https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv

### Start

library(tidyverse)  
library(knitr)  
library(readxl)  
library(zoo)  
  
url = 'https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv'  
home = read\_csv(url)  
  
read\_excel = 'data/PopulationEstimates.xls'  
  
pop = readxl::read\_excel("data/PopulationEstimates.xls", skip = 2) %>%   
 select(pop2019 = POP\_ESTIMATE\_2019, fips = FIPStxt, State)  
  
CA\_pop\_2019 = pop %>%  
 filter(State == "CA")

## Question 1: Covid-19 California Cases

library(tidyverse)  
  
url = 'https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv'  
  
home = read\_csv(url)

## Parsed with column specification:  
## cols(  
## date = col\_date(format = ""),  
## county = col\_character(),  
## state = col\_character(),  
## fips = col\_character(),  
## cases = col\_double(),  
## deaths = col\_double()  
## )

#Step 2  
  
steptwo = home %>% filter(state == "California") %>% group\_by(county) %>%   
 mutate(newCase = cases - lag(cases)) %>% arrange(-newCase) %>% ungroup()  
  
#Step 3  
  
#Table 1: 5 counties with most cumulative cases  
  
most\_cumulative\_cases = steptwo %>% filter(date == max(date)) %>% group\_by(county) %>%  
 summarize(sum\_cases = sum(cases)) %>% arrange(-sum\_cases) %>%  
 ungroup() %>% slice\_max(sum\_cases, n = 5)

## `summarise()` ungrouping output (override with `.groups` argument)

knitr::kable(most\_cumulative\_cases, caption = c("Cumulative Case Counts: Top 5 CA counties"),  
 col.names = c("County", "Cumulative Cases"))

Cumulative Case Counts: Top 5 CA counties

|  |  |
| --- | --- |
| County | Cumulative Cases |
| Los Angeles | 223131 |
| Riverside | 47669 |
| Orange | 43925 |
| San Bernardino | 41724 |
| San Diego | 35023 |

#Step 2  
  
steptwo = home %>% filter(state == "California") %>% group\_by(county) %>%   
 mutate(newCase = cases - lag(cases)) %>% arrange(-newCase) %>% ungroup()  
  
#Step 3  
  
#Table 2: 5 counties with most new cases  
most\_new\_cases = steptwo %>% filter(date == max(date)) %>% slice\_max(newCase, n=5) %>% select (county, newCase)  
  
knitr::kable(most\_new\_cases, caption = "New Case Counts: Top 5 CA counties", col.names = c("County", "New Cases"))

New Case Counts: Top 5 CA counties

|  |  |
| --- | --- |
| County | New Cases |
| Los Angeles | 1181 |
| Riverside | 949 |
| San Bernardino | 493 |
| Fresno | 407 |
| Sacramento | 308 |

pop = readxl::read\_excel("data/PopulationEstimates.xls", skip = 2) %>%  
 select(pop2019 = "POP\_ESTIMATE\_2019", fips = "FIPStxt", state = "State")  
  
CA\_pop\_covid = right\_join(pop, home, by ="fips") %>%   
 filter(date >= max(date, na.rm = TRUE)-13, state.x == "CA") %>%   
 group\_by(county) %>%  
 mutate(newCase = cases - lag(cases),  
 newCase\_pc = newCase / pop2019,  
 cumulative\_pc = cases / pop2019) %>%   
 ungroup()  
  
today = CA\_pop\_covid %>%   
 filter(date == max(date))  
  
most\_per\_cap\_today = today %>%   
 slice\_max(newCase\_pc, n = 5)  
  
most\_per\_cap\_all\_time = today %>%   
 slice\_max(cumulative\_pc, n = 5) %>%   
 select(county = county, cumulative\_pc)  
  
  
#Table 1  
pop\_most\_cumulative\_cases = CA\_pop\_covid %>%   
 filter(date == max(date)) %>%   
 slice\_max(cases, n=5) %>%   
 select(county, cumulative\_pc)  
  
knitr::kable(pop\_most\_cumulative\_cases, caption = c("Pop Cumulative Case Counts: Top 5 CA counties"),  
 col.names = c("County", "Cumulative Cases with Pop (100,000)"))

Pop Cumulative Case Counts: Top 5 CA counties

|  |  |
| --- | --- |
| County | Cumulative Cases with Pop (100,000) |
| Los Angeles | 0.0222262 |
| Riverside | 0.0192949 |
| Orange | 0.0138316 |
| San Bernardino | 0.0191387 |
| San Diego | 0.0104912 |

#Table 2  
  
pop\_most\_new\_cases = CA\_pop\_covid %>% filter(date == max(date)) %>% slice\_max(newCase\_pc, n=5) %>%   
 select(county, newCase\_pc)   
  
  
#%>% arrange(-most\_new\_cases) %>% head(20)  
  
knitr::kable(pop\_most\_new\_cases, caption = c("Pop New Case Counts: Top 5 CA counties"),  
 col.names = c("County", "New Cases with Pop (100,000)"))

Pop New Case Counts: Top 5 CA counties

|  |  |
| --- | --- |
| County | New Cases with Pop (100,000) |
| Madera | 0.0011378 |
| Merced | 0.0009507 |
| Fresno | 0.0004074 |
| Tulare | 0.0004054 |
| Riverside | 0.0003841 |

#Table 3: Safe counties  
  
pop\_safe\_county = CA\_pop\_covid %>%   
 group\_by(county) %>%   
 summarize(totNew = sum(newCase, na.rm= TRUE) / (max(pop2019) /100000)) %>%   
 filter(totNew <= 100)  
  
  
knitr::kable(pop\_safe\_county, caption = c("Pop Lowest Case Counts: CA counties"),  
 col.names = c("County", "Lowest New Cases with Pop"))

Pop Lowest Case Counts: CA counties

|  |  |
| --- | --- |
| County | Lowest New Cases with Pop |
| Alpine | 0.00000 |
| Calaveras | 82.77965 |
| Del Norte | 64.72026 |
| El Dorado | 93.85873 |
| Humboldt | 55.32687 |
| Mariposa | 29.06470 |
| Modoc | 11.31094 |
| Nevada | 69.16947 |
| Plumas | 42.53735 |
| Santa Cruz | 94.43182 |
| Shasta | 61.63927 |
| Siskiyou | 59.71658 |
| Trinity | 24.42002 |
| Tuolumne | 38.54767 |

### Results: Covid-19 cumulative case counts are the highest in the following counties: Los Angeles, Riverside, Orange, San Bernardino, and San Diego. The highest new cases over the past 14 days are in the following counties: Los Angeles, San Bernardino, Fresno, Sacramento, and Orange. Despite the covid-19 increases throughout the state the following counties have shown zero new cases over the past 14 days: Alpine, Calaveras, Colusa, Del Norte, and Humboldt. Based on these results you can see that population density plays a large role in the spread of covid-19 and more remote counties in Northern California have less current spread.

## Question 2: Covid-19 New York, California, Louisiana, and Florida

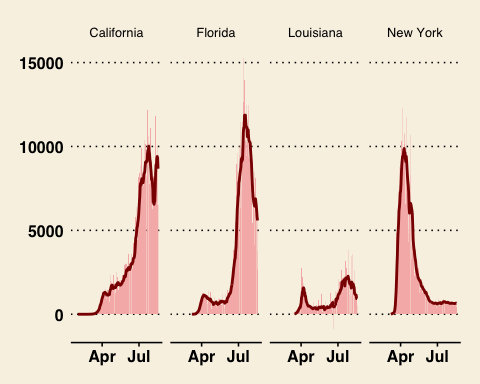
Q2 = home %>%  
 filter(state %in% c("New York","California", "Louisiana", "Florida")) %>%  
 group\_by(state, date) %>% summarise(cases = sum(cases)) %>%   
 mutate(newCases = cases - lag(cases),  
 roll7 = zoo::rollmean(newCases, 7, fill = NA, align = 'right')) %>% ungroup()

## `summarise()` regrouping output by 'state' (override with `.groups` argument)

Q2 %>% ggplot(aes(x = date, y = newCases)) + geom\_col(aes(y = newCases), col = NA, fill = "#F5B8B5") +  
 geom\_line(aes(y = roll7), col = "darkred", size = 1) + facet\_grid(~state, scale = "free\_y") +  
 ggthemes::theme\_wsj() + theme(legend.position = "right")

## Warning: Removed 4 rows containing missing values (position\_stack).

## Warning: Removed 7 row(s) containing missing values (geom\_path).



labs(title = paste("Daily Cases in NY, CA, LA, FL")) +  
 theme(plot.background = element\_rect(fill = "white"),  
 panel.background = element\_rect(fill = "white"),  
 plot.title = element\_text(size = 10, face = 'bold')) +  
 theme(aspect.ratio = .5)

## NULL