Session 5: Contagion and Diffusion

Welcome!

TA: Jennifer Lin

MMSS 211: Institutions, Rules, & Models in Social Science

2022-04-25

Goals

1. Increase comfort with working with large datasets
2. Become familiar with date objects in R for plots
3. Become more advanced with `ggplot2`, learning different ways to enhance plotting skills

Diffusion Models

Data

How do protest events spread across the country?

The Armed Conflict Location & Event Data Project (ACLED) tracks protests and other political violence instances around the world

Lin and Lunz-Trujillo (2022, Working Paper) pulled and geocoded data for the United States in 2020.

```
load("2020Prot.RData")
```

Modeling Protest Activity Spread

For today, we model the spread of protest events in 2020, specifically focused on

- Black Lives Matter
- Coronavirus

Get Protests by Topic

```
keywords <- c(
  "coronavirus",
  "Black Lives Matter")

m <- sapply(keywords, grepl, protest20$notes)

protest20$keywords <- apply(
  m, 1,
  function(y) paste0(colnames(m)[y], collapse=","))

protest20 <- protest20  %>%
  separate(
    col = "keywords",
    into = c("k1", "k2"),
    sep = ",",
    remove = FALSE
  ) %>%
  filter(event_type == "Protests") %>%
  mutate(
    date = dmy(event_date)
  )
```

Spread of Protests

For all protests, we assume that the county is the unit of analysis and that

- **Susceptible**: All counties have the chance to be host to a protest
- **Infected**: Counties are "infected" when they have hosted a protest
- **Recovery**: Post-Protest feelings that may lead to another protest in the same/similar area

Black Lives Matter

How have BLM protests sprung up and have they been occurring at the same rates throughout the course of 2020?

```
blm_prot_time <- protest20 %>%
  filter(k1 == "Black Lives Matter") %>%
  group_by(date) %>%
  summarise(
    n = n()
  ) %>%
  mutate(
    total = cumsum(n),
    days = mdy("01-01-2020") %--% date,
    day_of_year = as.duration(days) / ddays(1)
  )
```

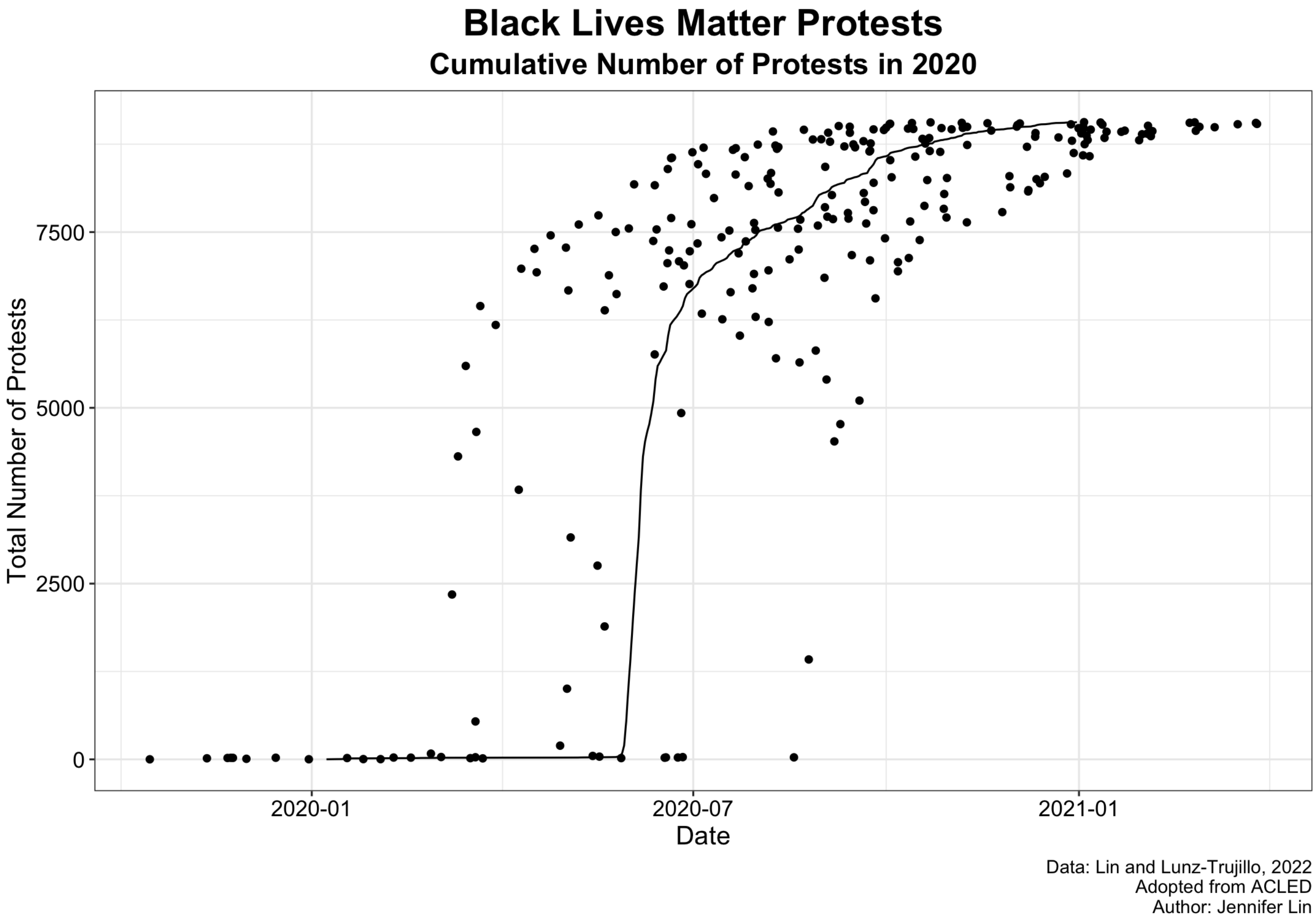
Before Plotting...

```r
theme_contagion <- function() {
  theme_bw()+
  theme(
    plot.title          = element_text(
      hjust = 0.5, size = 20, colour="black", face = "bold"),
    plot.subtitle       = element_text(
      hjust = 0.5, size = 16, colour="black", face = "bold"),
    legend.title        = element_text(
      hjust = 0.5, size = 14, colour="black", face = "bold"),
    plot.caption        = element_text(size = 10, colour="black"),
    axis.title          = element_text(size = 14, colour="black"),
    axis.text.x         = element_text(
      size = 12, colour="black"),
    axis.text.y         = element_text(size = 12, colour="black"),
    legend.position     = 'bottom',
    legend.direction    = "horizontal",
    legend.text         = element_text(size = 12, colour="black")
  )
}
```

Black Lives Matter

Plotting the Trends

```
ggplot(blm_prot_time, aes(x = date, y = total, group=1))+
  geom_line()+
  geom_point(position = position_jitter(100))+
  labs(
    title = "Black Lives Matter Protests",
    subtitle = "Cumulative Number of Protests in 2020",
    caption = "Data: Lin and Lunz-Trujillo, 2022
    Adopted from ACLED
    Author: Jennifer Lin"
  )+
  xlab("Date")+
  ylab("Total Number of Protests")+
  theme_contagion()
```


Black Lives Matter Protests
Cumulative Number of Protests in 2020
Total Number of Protests
7500
5000
2500
0
2020-01
2020-07
2021-01
Date
Data: Lin and Lunz-Trujillo, 2022
Adopted from ACLED
Author: Jennifer Lin

Black Lives Matter

Spread Across the Country

As we will recall, BLM protests did not all spring up at once. Events were held in various counties across time, mostly in the late-May to mid-June time span.

Let's model this using a SIR model -- to see the rate of "infection" of BLM protests.

To do this, we first need a grand dataset of counties. All we need to know is the total number of susceptible counties, so let's just use the canned data from `tidycensus`

```
data("fips_codes")
```

Black Lives Matter

Generate Adoption by County

```
adopt_blm <- protest20 %>%
  filter(k1 == "Black Lives Matter") %>%
  group_by(county) %>%
  filter(date == min(date)) %>%
  slice(1) %>%
  ungroup() %>%
  arrange(date) %>%
  group_by(date) %>%
  summarise(
    n = n()) %>%
  mutate(
    total = cumsum(n),
    pct_adopt = total/nrow(fips_codes)) %>%
  filter(date >= "2020-05-01" & date <= "2020-08-01") %>%
  mutate(
    days = mdy("05-01-2020") %--% date,
    day_of_year = as.duration(days) / ddays(1))
```

SIR Model in Theory

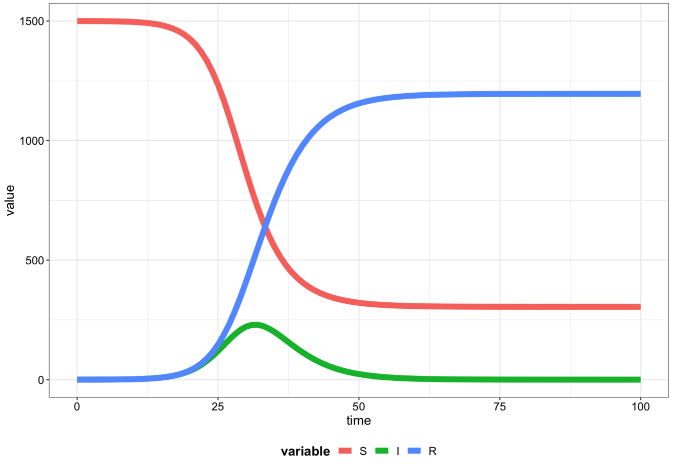
```
get_sir <- function(beta, gamma, S0, I0, R0, times) {
  # Equation for Calculating SIR
  sir_equations <- function(time, variables, parameters) {
    with(as.list(c(variables, parameters)), {
      N = S+I+R
      lambda = beta*(I/N)
      dS = -lambda*S
      dI = lambda*S-gamma*I
      dR = gamma*I
      return(list(c(dS, dI, dR)))
    })}
  # Parameter Values
  parameters_values <- c(beta = beta, gamma = gamma)
  # Initial Values
  initial_values <- c(S = S0, I = I0, R = R0)
  # Generate the model
  out <- ode(initial_values, times,
             sir_equations, parameters_values)
  # Return
  as.data.frame(out)
}
```

Black Lives Matter

Generate SIR Model

```
model_BLM <- get_sir(
  beta = 0.6, gamma = 0.3, S0 = 1500, I0 = .1,
  R0 = 0, times = seq(0, 100, 1)) %>%
  reshape2::melt(id = "time")
```

```
ggplot(
  model_BLM,
  aes(x = time, y = value,
      group = variable,
      color = variable))+
  geom_line(size = 3)+
  theme_contagion()
```



Black Lives Matter

Adopting Protests

We define "infected" as counties that host a protest.

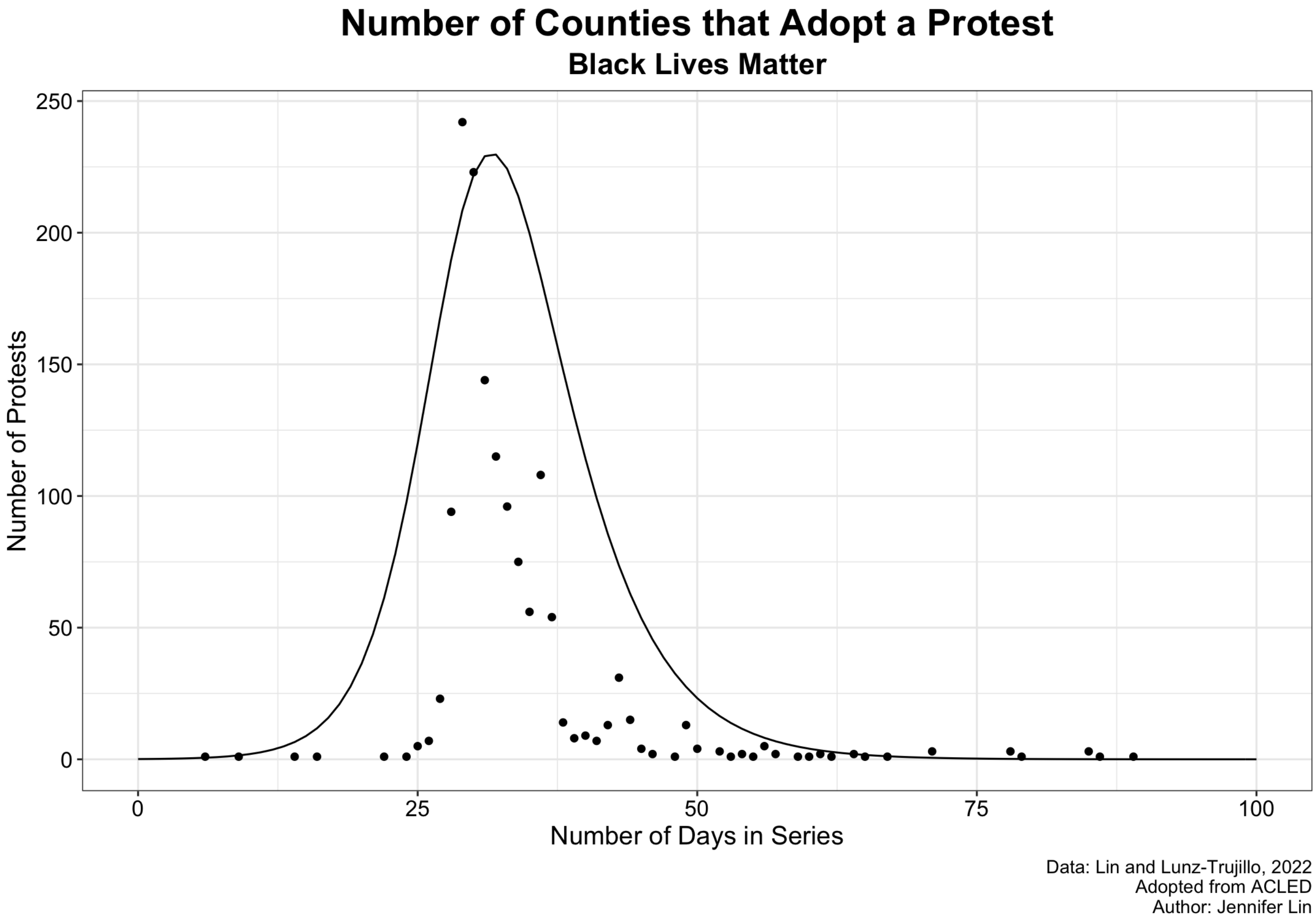
Therefore, our plot for protest adoption concerns the "infected" category. We get those predictions.

```
I_BLM <- model_BLM %>%
  filter(variable == "I")
```

Black Lives Matter

Adopting Protests

```
ggplot(adopt_blm, aes(x = day_of_year, y = n, group=1))+
  geom_point()+
  geom_line(I_BLM, mapping = aes(x = time, y = value, group = 1))+
  labs(
    title = "Number of Counties that Adopt a Protest",
    subtitle = "Black Lives Matter",
    caption = "Data: Lin and Lunz-Trujillo, 2022
    Adopted from ACLED
    Author: Jennifer Lin"
  )+
  xlab("Number of Days in Series")+
  ylab("Number of Protests")+
  theme_contagion()
```

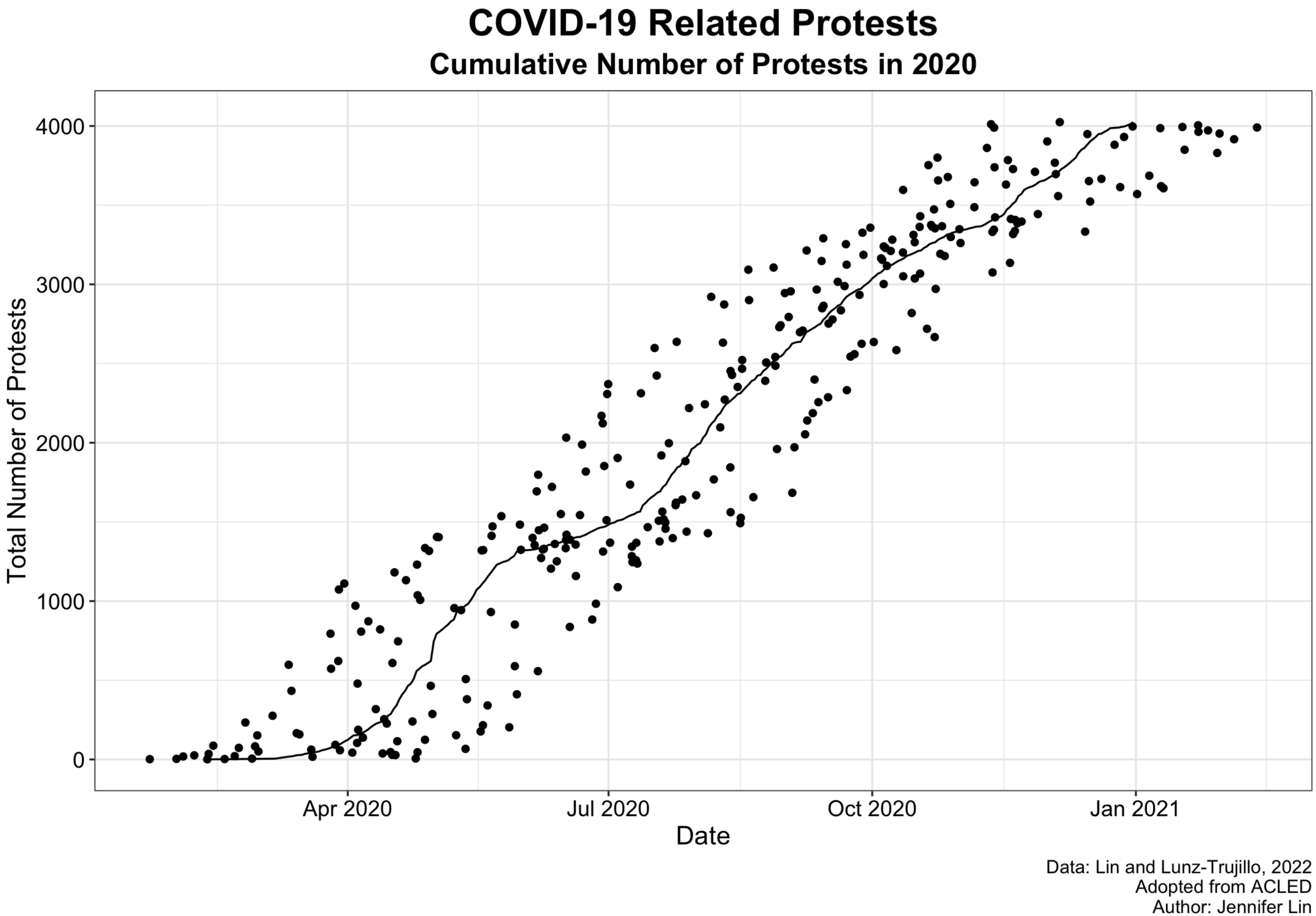

Number of Counties that Adopt a Protest
Black Lives Matter
250
200
150
100
50
0
Number of Protests
0
25
50
75
100
Number of Days in Series
Data: Lin and Lunz-Trujillo, 2022
Adopted from ACLED
Author: Jennifer Lin

COVID-19

Now, let's repeat the process to explore the adoption of COVID-19 related protests.

Here, I am rather agnostic to which position the protest takes.

```
covid_prot_time <- protest20 %>%
  filter(k1 == "coronavirus") %>%
  group_by(date) %>%
  summarise(
    n = n()
  ) %>%
  mutate(
    total = cumsum(n)
  )
```


COVID-19 Related Protests
Cumulative Number of Protests in 2020
Total Number of Protests
0
1000
2000
3000
4000
Apr 2020
Jul 2020
Oct 2020
Jan 2021
Date
Data: Lin and Lunz-Trujillo, 2022
Adopted from ACLED
Author: Jennifer Lin

COVID-19

Adoptions by County

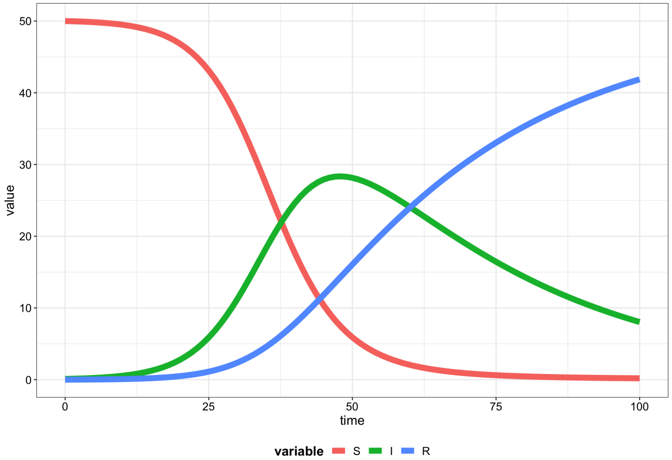
```
adopt_covid <- protest20 %>%
  filter(k1 == "coronavirus") %>%
  group_by(county) %>%
  filter(date == min(date)) %>%
  slice(1) %>%
  ungroup() %>%
  arrange(date) %>%
  group_by(date) %>%
  summarise(
    n = n()) %>%
  mutate(
    total = cumsum(n),
    pct_adopt = total/nrow(fips_codes)) %>%
  filter(date >= "2020-03-01" & date <= "2020-06-15") %>%
  mutate(
    days = mdy("03-01-2020") %--% date,
    day_of_year = as.duration(days) / ddays(1))
```

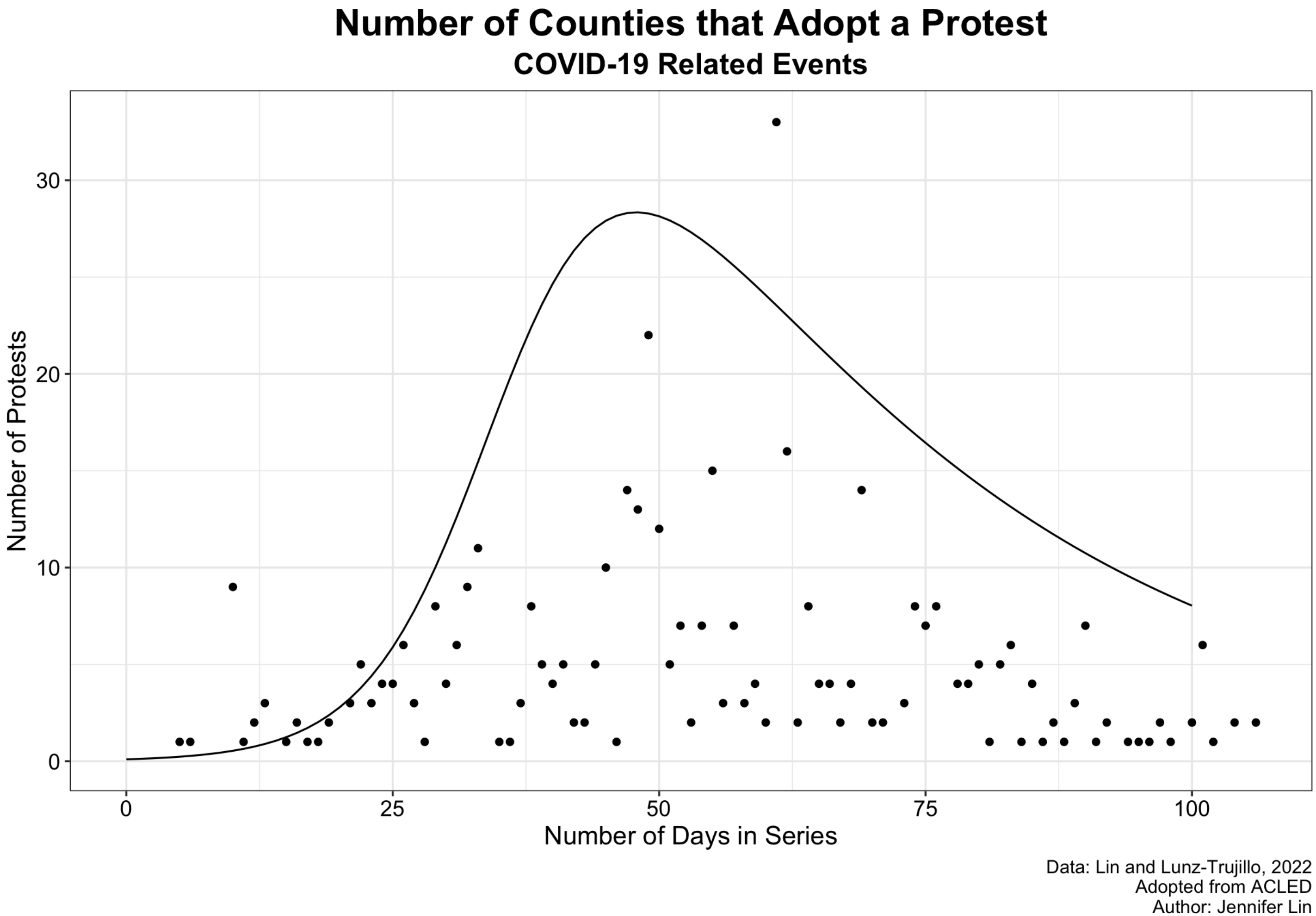
COVID-19

Generate SIR Model

```
model_covid <- get_sir(
  beta = 0.2, gamma = 0.03, S0 = 50, I0 = .1,
  R0 = 0, times = seq(0, 100, 1)) %>%
  reshape2::melt(id = "time")
```

```
ggplot(
  model_covid,
  aes(x = time, y = value,
      group = variable, color =
  geom_line(size = 3)+
  theme_contagion()
```




Number of Counties that Adopt a Protest
COVID-19 Related Events
Number of Protests
30
20
10
0
0
25
50
75
100
Number of Days in Series
Data: Lin and Lunz-Trujillo, 2022
Adopted from ACLED
Author: Jennifer Lin

Contagion Models

Data

Google Search Trends for various health related issues during the COVID-19 Pandemic -- reflects how often people searched for a medical term.

What are the most searched medical conditions during the pandemic? Did they ebb and flow across time? How do these vary by state?

We will read the data from source:

```
root <- "https://storage.googleapis.com/"
file <- "covid19-open-data/v3/google-search-trends.csv"
trends <- paste0(root, file)
trends_data <- rio::import(trends)
```

This will take a minute -- If it fails, read and use the CSV provided

Top Medical Condition Searches

Are certain medical search terms more popular during the COVID-19 Pandemic and do they persist?

Are people searching more for COVID-19 related symptoms compared to non-related symptoms?

```
get_top_search <- function(key, position){
  loc = as.character(key)
  pos = as.numeric(position)

  most_searched <- trends_data %>%
    filter(location_key == loc) %>%
    gather(key, value, 4:425) %>%
    group_by(date) %>%
    arrange(desc(value)) %>%
    filter(key != "location_key") %>%
    slice(pos)

  return(most_searched)
}
```

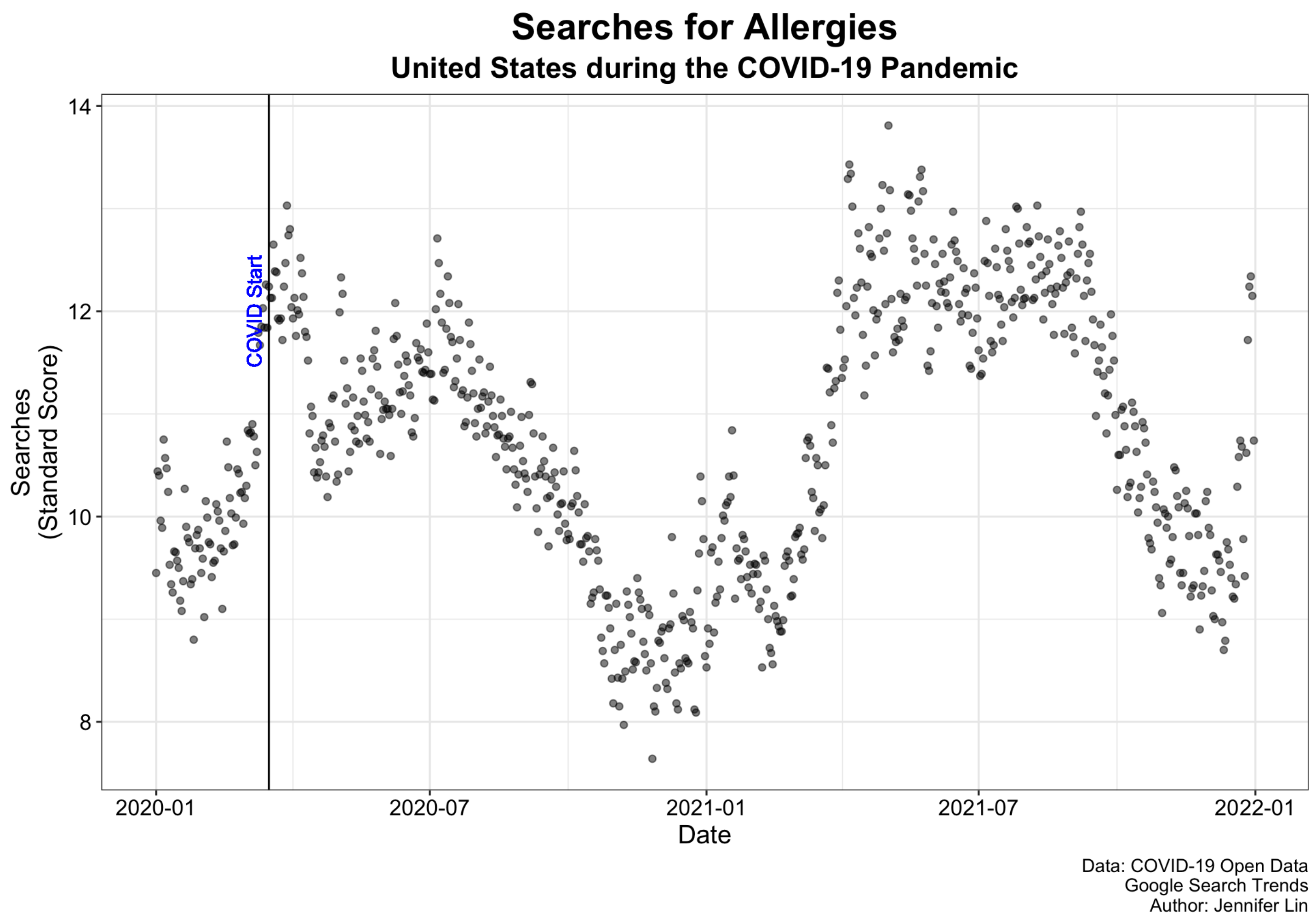
Look at Most Searched Medical Conditions

Looking at the top searches, it seems that the most searched are:

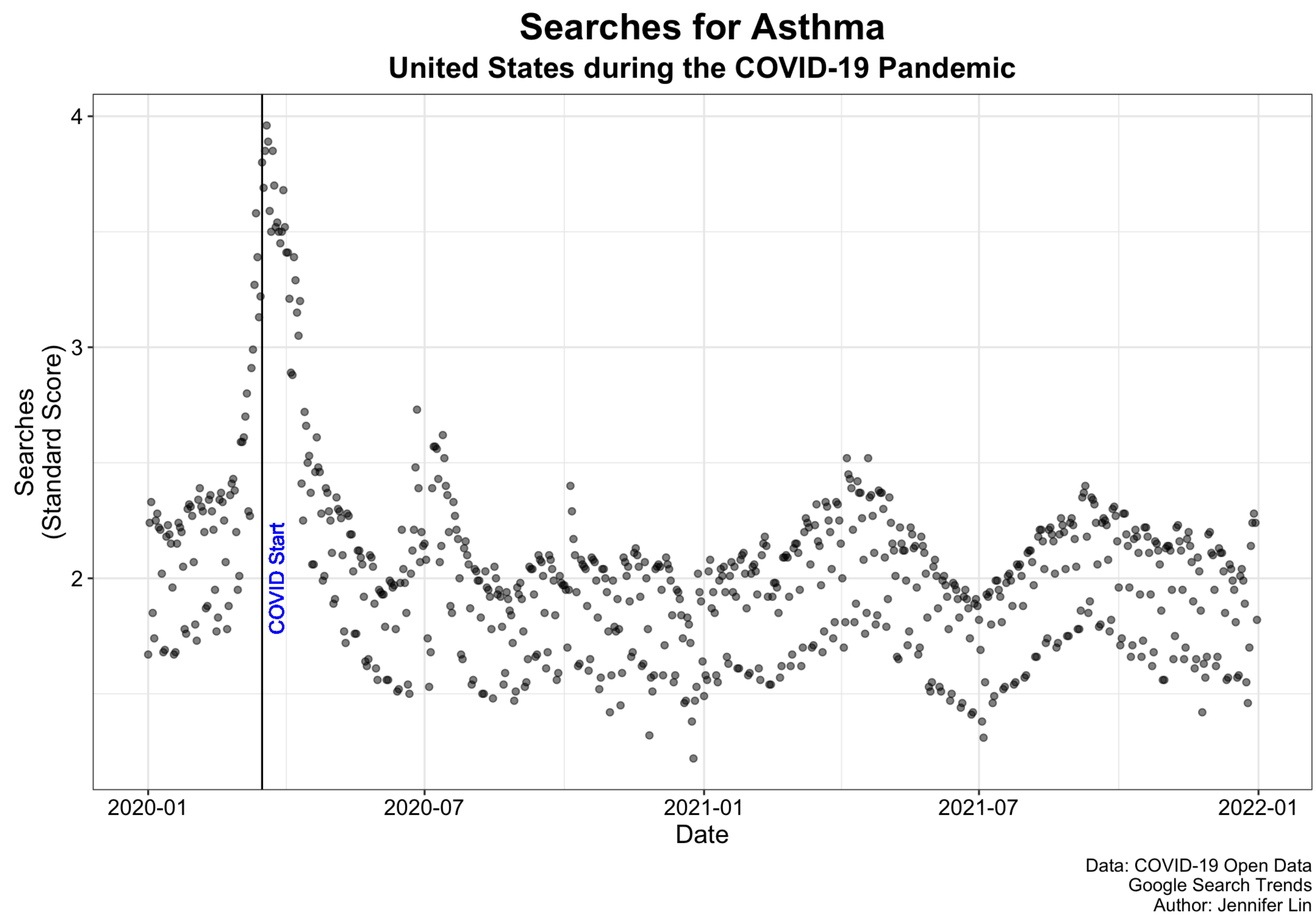
- `search_trends_infection`
- `search_trends_pain`
- `search_trends_allergy`

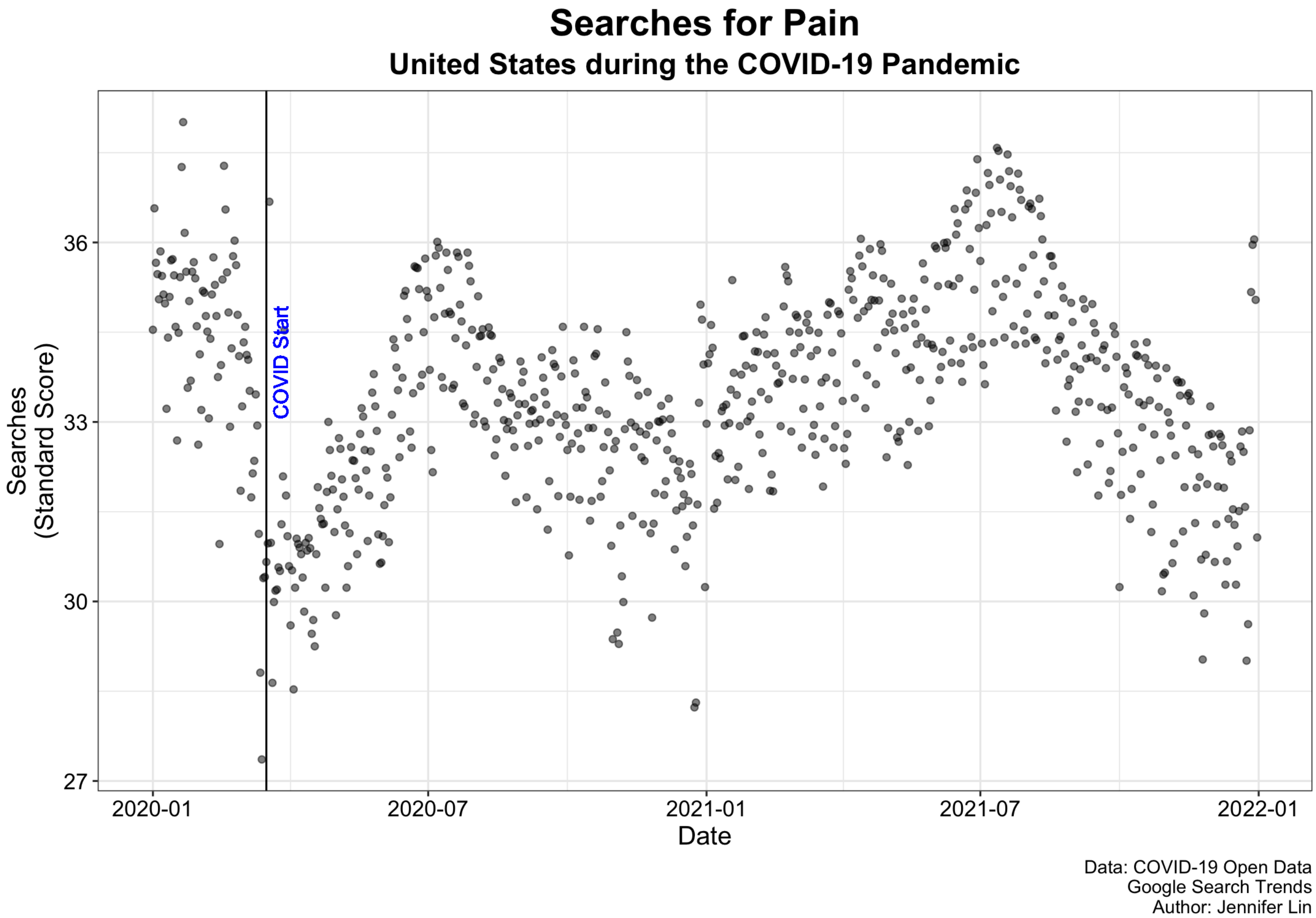
```
most_searched <- get_top_search("US", 1)
```

```r
trends_data %>%
  filter(location_key == "US") %>%
  ggplot()+
  geom_point(
    aes(x = date, y = search_trends_allergy), alpha = 0.5)+
  geom_vline(xintercept = ymd("2020-03-16"))+
  geom_text(
    aes(
      x=ymd("2020-03-16"),
      label="COVID Start\n", y = 12),
    colour="blue", angle=90) +
  labs(
    title = "Searches for Allergies",
    subtitle = "United States during the COVID-19 Pandemic",
    caption = "Data: COVID-19 Open Data
    Google Search Trends
    Author: Jennifer Lin")+
  xlab("Date")+
  ylab("Searches\n(Standard Score)")+
  theme_contagion()
```



Searches for Allergies
United States during the COVID-19 Pandemic
Searches
(Standard Score)
14
12
10
8
COVID Start
2020-01
2020-07
2021-01
2021-07
2022-01
Date
Data: COVID-19 Open Data
Google Search Trends
Author: Jennifer Lin


Searches for Asthma
United States during the COVID-19 Pandemic
Searches
(Standard Score)
4
3
2
COVID Start
2020-01
2020-07
2021-01
2021-07
2022-01
Date
Data: COVID-19 Open Data
Google Search Trends
Author: Jennifer Lin


Searches for Pain
United States during the COVID-19 Pandemic
Searches
(Standard Score)
36
33
30
27
COVID Start
2020-01
2020-07
2021-01
2021-07
2022-01
Date
Data: COVID-19 Open Data
Google Search Trends
Author: Jennifer Lin

Exercise -- Pick One

1. Using the Protests Data, rerun the fuzzy match code and generate a new SIR model for protests related to "election", "school board", "Iran", "fracking", "abortion", "gun violence", or anything else that might be interesting.
2. Find other medical symptoms that people googled in the US, Australia, New Zealand or the UK, and plot the tends. You can use the entire country, a state/province, or the the US, a county (cross reference with FIPS codes data for the county identifier) Try to see if you can identify key dates that might explain the ebbs and flows of the trends.