Throughout the COVID-19 pandemic, the main sources of information for case numbers in the State of Arkansas have been daily press conferences by the Governor's Office (until recently, when they moved to weekly) and the website arkansascovid.com. I haven't been particularly impressed with the visualizations used by either source. Today I'm sharing some code that I have been using throughout the pandemic to keep track of how Arkansas is doing with the pandemic.

We'll use several libraries, the purpose of which is indicated in the comments:

```
library(tidyverse)
library(lubridate) # Date wrangling
library(gganimate) # GIF production
library(tidycensus) # Population estimates
library(transformr) # used by gganimate
library(ggthemes) # map themes
library(viridis) # Heatmap color palette
library(scales) # Pretty axis labels
library(zoo) # rollapply

knitr::opts_chunk$set(
  message = F,
  echo = T,
  include = T
)

options( scipen = 10 ) # print full numbers, not scientific notation
```

We'll use the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, which is maintained at Github. The data can be read in a single line, although we'll reorganize the case counts into a long format for ease of further wrangling. Here's a snippet of the table:

```
covid_cases <- read_csv("https://raw.githubusercontent.com/
CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/
time_series_covid19_confirmed_US.csv")
covid_cases <- pivot_longer(covid_cases, 12:length(covid_cases),
names_to = "date", values_to = "cases") %>%
  mutate(date = lubridate::as_date(date, format = "%m/%d/%y")) %>%
  filter(Province_State == 'Arkansas') %>%
  arrange(date, Combined_Key)

tail(covid_cases %>% select(Combined_Key, date, cases))

## # A tibble: 6 x 3
##   Combined_Key             date       cases
##
## 1 Union, Arkansas, US      2020-09-28   894
## 2 Van Buren, Arkansas, US  2020-09-28   174
## 3 Washington, Arkansas, US 2020-09-28  9457
## 4 White, Arkansas, US      2020-09-28   849
## 5 Woodruff, Arkansas, US   2020-09-28    53
## 6 Yell, Arkansas, US       2020-09-28  1256
```

Because we'll be doing per-capita calculations, we need to load population estimates. Fortunately, the tidycensus package provides a convenient method of obtaining that information. Here's a snapshot of the population data:

```
population <- tidycensus::get_estimates(geography = "county",
"population") %>%
  mutate(GEOID = as.integer(GEOID)) %>%
  pivot_wider(
    names_from = variable,
    values_from = value
  ) %>%
  filter(grepl("Arkansas", NAME))

head(population)

## # A tibble: 6 x 4
##   NAME                      GEOID    POP DENSITY
##
## 1 Arkansas County, Arkansas  5001  17769    18.0
## 2 Ashley County, Arkansas    5003  20046    21.7
## 3 Baxter County, Arkansas    5005  41619    75.1
## 4 Benton County, Arkansas    5007 272608   322.
## 5 Boone County, Arkansas     5009  37480    63.5
## 6 Bradley County, Arkansas   5011  10897    16.8
```

The Governor's Office makes several design choices designed to spin statistics so that it looks like Arkansas is doing a good job with managing the crisis (as of the writing of this post, the statistics suggest otherwise). For example, the bar chart of rolling cases often splits out prison cases and community spread cases so the overall trend is obscured. Further, the use of bar charts rather than line charts also makes it harder to visualize the trend of new cases. We'll use a trendline of overall cases without spin:

```
ark_covid_cases <- covid_cases %>%
  filter(`Province_State` == 'Arkansas')

p <- ark_covid_cases %>%
  filter(cases > 0) %>%
  group_by(Province_State, date) %>%
  mutate(cases = sum(cases)) %>%
  ggplot(aes(x = date, y = cases)) +
  geom_line() +
  scale_x_date(breaks = scales::pretty_breaks()) +
  scale_y_continuous(labels = unit_format(unit = "k", sep = "",
big.mark = ",", scale = 1/1000)) +
  labs(
    title = "Total COVID-19 cases in Arkansas",
    x = "", y = "",
    caption = paste0("Image generated: ", Sys.time(), "\n", "Data
source: https://github.com/CSSEGISandData/COVID-19", "\n", "COVID-19 Data
Repository by CSSE at Johns Hopkins University")
  )
```

```
ggsave(filename = "images/ark_covid_total_cases.png", plot = p, height
= 3, width = 5.25)
p
```

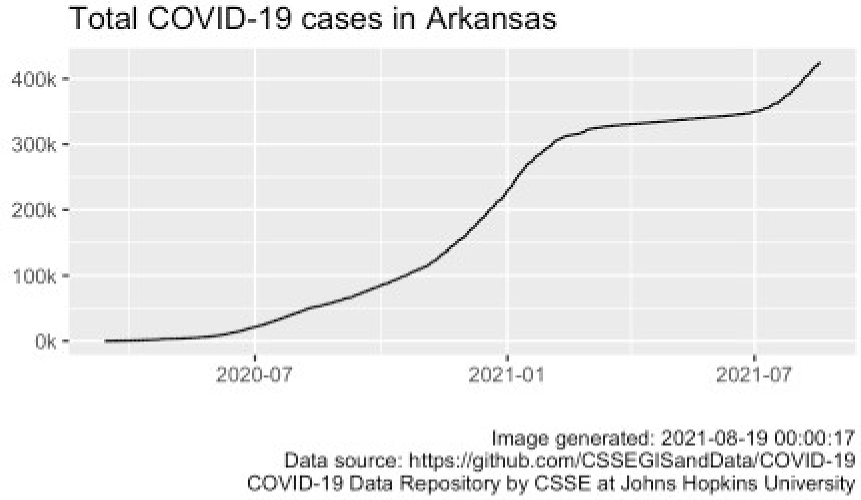


Image generated: 2021-08-19 00:00:17
Data source: https://github.com/CSSEGISandData/COVID-19
COVID-19 Data Repository by CSSE at Johns Hopkins University

This trendline shows no real signs of leveling off. As we'll see later on, the number of new cases isn't going down.

Both the Governor's Office and the website arkansascovid.com both use arbitrarily selected population metrics to depict per-capita cases (typically 10,000 residents). We'll use a different per-capita metric that is reasonably close to the median county size in the state. As such, for many counties, the per-capita number will be reasonably close to the actual population of the county. That number can be calculated from the state's population metrics:

```
per_capita <- population %>%
  filter(grepl("Arkansas", NAME)) %>%
  summarize(median = median(POP)) %>% # Get median county population
  unlist()

per_capita

## median
##  18188
```

Instead of using the actual median, we'll round it to the nearest 5,000 residents:

```
per_capita <- plyr::round_any(per_capita, 5e3) # Round population to
nearest 5,000
per_capita

## median
##  20000
```

Now that we have the population figure we want to use for the per-capita calculations, we will perform those using the lag function to calculate the new cases per day, and then using the rollapply function to smooth the number of daily cases over a sliding 1-week (7-day) window. The results look like this:

```
roll_ark_cases <- ark_covid_cases %>%
  arrange(date) %>%
  group_by(UID) %>%
  mutate(prev_count = lag(cases)) %>%
  mutate(prev_count = ifelse(is.na(prev_count), 0, prev_count)) %>%
  mutate(new_cases = cases - prev_count) %>%
  mutate(roll_cases = round(zoo::rollapply(new_cases, 7, mean, fill =
0, align = "right", na.rm = T)))%>%
  ungroup() %>%
  select(-prev_count) %>%
  left_join(
    population %>% select(-NAME),
    by = c("FIPS" = "GEOID")
  ) %>%
  mutate(
    cases_capita = round(cases / POP * per_capita), # cases per
per_capita residents
    new_capita = round(new_cases / POP * per_capita), # cases per
per_capita residents
    roll_capita = round(roll_cases / POP * per_capita) # rolling new
cases per per_capita residents
  )

tail(roll_ark_cases %>% select(date, Admin2, POP, cases, new_cases,
roll_cases, roll_capita))

## # A tibble: 6 x 7
##   date       Admin2        POP cases new_cases roll_cases
roll_capita
##
## 1 2020-09-28 Union       39126   894         2          8
4
## 2 2020-09-28 Van Buren   16603   174         0          1
1
## 3 2020-09-28 Washington 236961  9457        42         64
5
## 4 2020-09-28 White       78727   849         8         17
4
## 5 2020-09-28 Woodruff     6490    53         0          1
3
## 6 2020-09-28 Yell        21535  1256         2          3
3
```

We can summarize those results in order to get a total number of rolling cases for the entire state, which looks like this:

```
roll_agg_ark_cases <- roll_ark_cases %>%
  group_by(date) %>%
  summarize(roll_cases = sum(roll_cases))

tail(roll_agg_ark_cases)

## # A tibble: 6 x 2
##   date       roll_cases
```

```
##
## 1 2020-09-23        820
## 2 2020-09-24        835
## 3 2020-09-25        839
## 4 2020-09-26        797
## 5 2020-09-27        786
## 6 2020-09-28        812
```

We can then plot the aggregate number of rolling cases over time. We'll show a couple of different time points relevant to the spread of the coronavirus, including the Governor's mask/social distancing mandate and the reopening of public schools:

```
p <- roll_agg_ark_cases %>%
  ggplot(aes(date, roll_cases)) +
    geom_line() +
    geom_vline(xintercept = as.Date("2020-08-24"), color = "gray10",
linetype = "longdash") +
    annotate(geom ="text", label = "School\nstarts", x =
as.Date("2020-08-05"), y = 200, color = "gray10") +
    annotate(geom = "segment", y = 290, yend = 400, x =
as.Date("2020-08-05"), xend = as.Date("2020-08-24")) +
    geom_vline(xintercept = as.Date("2020-07-16"), color = "gray10",
linetype = "longdash") +
    annotate(geom ="text", label = "Mask\nmandate", x =
as.Date("2020-06-21"), y = 100, color = "gray10") +
    annotate(geom = "segment", y = 190, yend = 300, x =
as.Date("2020-06-21"), xend = as.Date("2020-07-16")) +
    geom_smooth(span = 1/5) +
    labs(
      title = "7-Day Rolling Average of New COVID-19 Cases in
Arkansas",
      x = "", y = "",
      caption = paste0("Image generated: ", Sys.time(), "\n", "Data
source: https://github.com/CSSEGISandData/COVID-19", "\n", "COVID-19 Data
Repository by CSSE at Johns Hopkins University")
    ) +
  theme(
    title = element_text(size = 10)
  )

ggsave(filename = "images/ark_covid_rolling_cases.png", plot = p,
height = 3, width = 5.25)
p
```

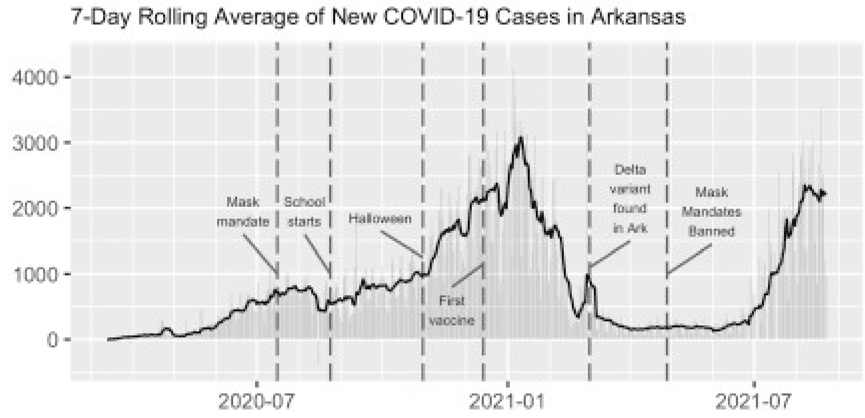


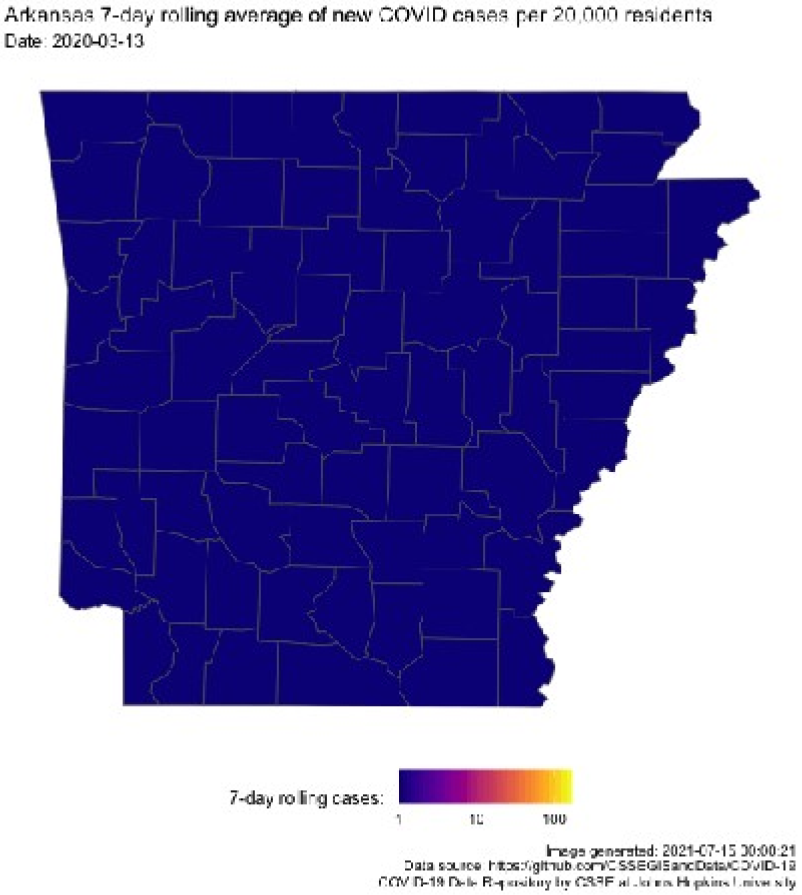
Image generated: 2021-08-24 00:00:17
Data source: https://github.com/CSSEGISandData/COVID-19
COVID-19 Data Repository by CSSE at Johns Hopkins University

From this plot, it appears that the mask mandate may have had a positive effect in leveling off the number of new COVID-19 cases. Conversely, it appears that the reopening of schools may have led to a rapid increase in the number of new cases. Of course, the rate of virus transmission has a multitude of causes, and the correlation here doesn't necessarily imply causation.

The website arkansascovid.com contains better visualizations than what the Governor's Office uses, but the default Tableau color scheme doesn't do a very good job of showing hotspots. Counties with a higher number of cases are depicted in dark blue (a color associated with cold), while counties with fewer cases are shown in pale green (a color without a heat association). In addition, there aren't visualizations that show changes at the county level over time. So, we'll use a county-level visualization that shows the number of rolling new cases over time with a color scheme that intuitively shows hot spots:

```
# Start when 7-day rolling cases in state > 0
first_date <- min({
  roll_ark_cases %>%
    group_by(date) %>%
    summarize(roll_cases = sum(roll_cases)) %>%
    ungroup() %>%
    filter(roll_cases > 0) %>%
    select(date)
}$date)

temp <- roll_ark_cases %>%
  filter(date >= first_date) %>%
  mutate(roll_capita = ifelse(roll_capita <= 0, 1, roll_capita)) %>% # 
log10 scale plot
  mutate(roll_cases = ifelse(roll_cases <= 0, 1, roll_cases)) # log10 
scale plot

# Prefer tigris projection for state map
temp_sf <- tigris::counties(cb = T, resolution = "20m") %>%
  mutate(GEOID = as.numeric(GEOID)) %>%
  inner_join(temp %>% select(FIPS, roll_cases, roll_capita, date), by = 
c("GEOID" = "FIPS")) %>%
  select(GEOID, roll_cases, roll_capita, date, geometry)
```

```
# tidycensus projection is skewed for state map
# data("county_laea")
# data("state_laea")
# temp_sf <- county_laea %>%
#   mutate(GEOID = as.numeric(GEOID)) %>%
#   inner_join(temp, by = c("GEOID" = "FIPS"))

days <- NROW(unique(temp$date))

p <- ggplot(temp_sf) +
  geom_sf(aes(fill = roll_capita), size = 0.25) +
  scale_fill_viridis(
    name = "7-day rolling cases: ",
    trans = "log10",
    option = "plasma",
  ) +
  ggthemes::theme_map() +
  theme(legend.position = "bottom", legend.justification = "center") +
  labs(
    title = paste0("Arkansas 7-day rolling average of new COVID cases
per ", scales::comma(per_capita), " residents"),
    subtitle = "Date: {frame_time}",
    caption = paste0("Image generated: ", Sys.time(), "\n", "Data
source: https://github.com/CSSEGISandData/COVID-19", "\n", "COVID-19 Data
Repository by CSSE at Johns Hopkins University")
  ) +
  transition_time(date)

Sys.time()
anim <- animate(
  p,
  nframes = days + 10 + 30,
  fps = 5,
  start_pause = 10,
  end_pause = 30,
  res = 96,
  width = 600,
  height = 600,
  units = "px"
)
Sys.time()

anim_save("images/ark_covid_rolling_cases_plasma.gif", animation =
anim)

# anim
```


Arkansas 7-day rolling average of new COVID cases per 20,000 residents
Date: 2020-03-13
7-day rolling cases:
1
10
100
Image generated: 2021-07-15 00:00:21
Data source: https://github.com/CSSEGISandData/COVID-19
COVID-19 Data Repository by CSSE at Johns Hopkins University

There are a couple of design choices here that are worth explaining. First, we’re animating the graphic over time, which shows where hotspots occur during the course of the pandemic.

Second, we’re using the plasma color palette from the viridis package. This palette goes from indigo on the low end to a hot yellow on the high end, so it intuitively shows hotspots.

Third, we’re using a log scale for the number of new cases – the idea here is that jumps of an order of magnitude or so are depicted in different colors (i.e., indigo, purple, red, orange, yellow) along the plasma palette. If we use a standard numerical scale for the number of new cases, jumps from 1-20 or so get washed out due to the large size of the worst outbreaks.

Conclusion

I hope you found my alternate visualizations for COVID-19 in Arkansas useful. The charts are set to update nightly, so these data should be current throughout the pandemic. If you have suggestions for improvements or notice that the figures aren’t updating, please comment! Thanks for reading.