Texas Covid Data Project

Nathen Byford, Connor Bryson, Miguel Iglesias

11/11/2021

Part 1: Data manipulation

```
tb <- read_xlsx(".\\data\\data.xlsx", skip = 2)

cases <- tb |>
  rename(county = "County Name") |>
  gather(-1, key = "date", value = "cases") |>
  mutate(date = str_extract(date, "[0-9]{2}-[0-9]{2}-[0-9]{4}"),
         date = mdy(date))

cases
```

```
## # A tibble: 160,420 x 3
##    county    date       cases
##    <chr>     <date>     <dbl>
##  1 Anderson  2020-03-04     0
##  2 Andrews   2020-03-04     0
##  3 Angelina  2020-03-04     0
##  4 Aransas   2020-03-04     0
##  5 Archer    2020-03-04     0
##  6 Armstrong 2020-03-04     0
##  7 Atascosa  2020-03-04     0
##  8 Austin    2020-03-04     0
##  9 Bailey    2020-03-04     0
## 10 Bandera   2020-03-04     0
## # ... with 160,410 more rows
```

Part 2: Data merging

```
pop_dat <- read_csv(".\\data/county-populations.csv")
```

```
## Rows: 255 Columns: 2
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): county
## dbl (1): population
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
cases <- left_join(cases, pop_dat)
```

```
## Joining, by = "county"
```

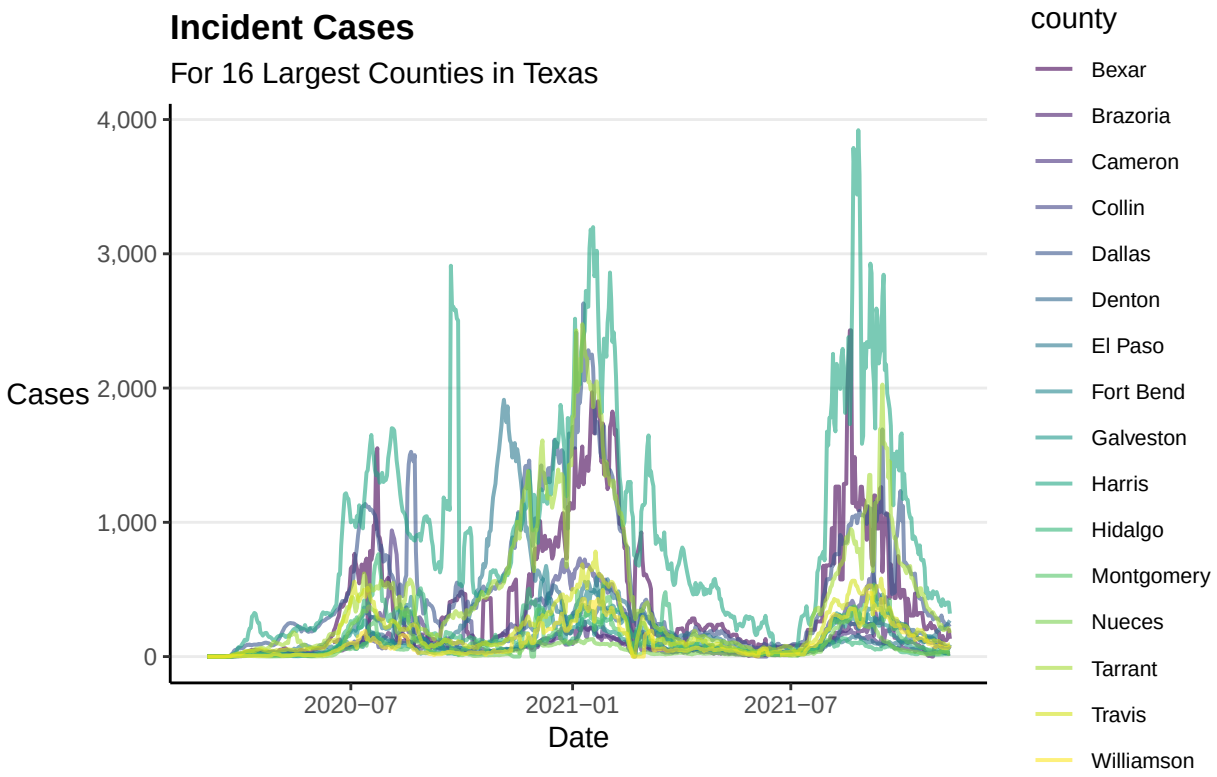
```
cases
```

```
## # A tibble: 160,420 x 4
##    county    date       cases population
##    <chr>     <date>     <dbl>      <dbl>
##  1 Anderson  2020-03-04     0      58199
##  2 Andrews   2020-03-04     0      22269
##  3 Angelina  2020-03-04     0      90437
##  4 Aransas   2020-03-04     0      27699
##  5 Archer    2020-03-04     0       8344
##  6 Armstrong 2020-03-04     0       1948
##  7 Atascosa  2020-03-04     0      51831
##  8 Austin    2020-03-04     0      30402
##  9 Bailey    2020-03-04     0       7692
## 10 Bandera   2020-03-04     0      21246
## # ... with 160,410 more rows
```

part 3: Data visualization

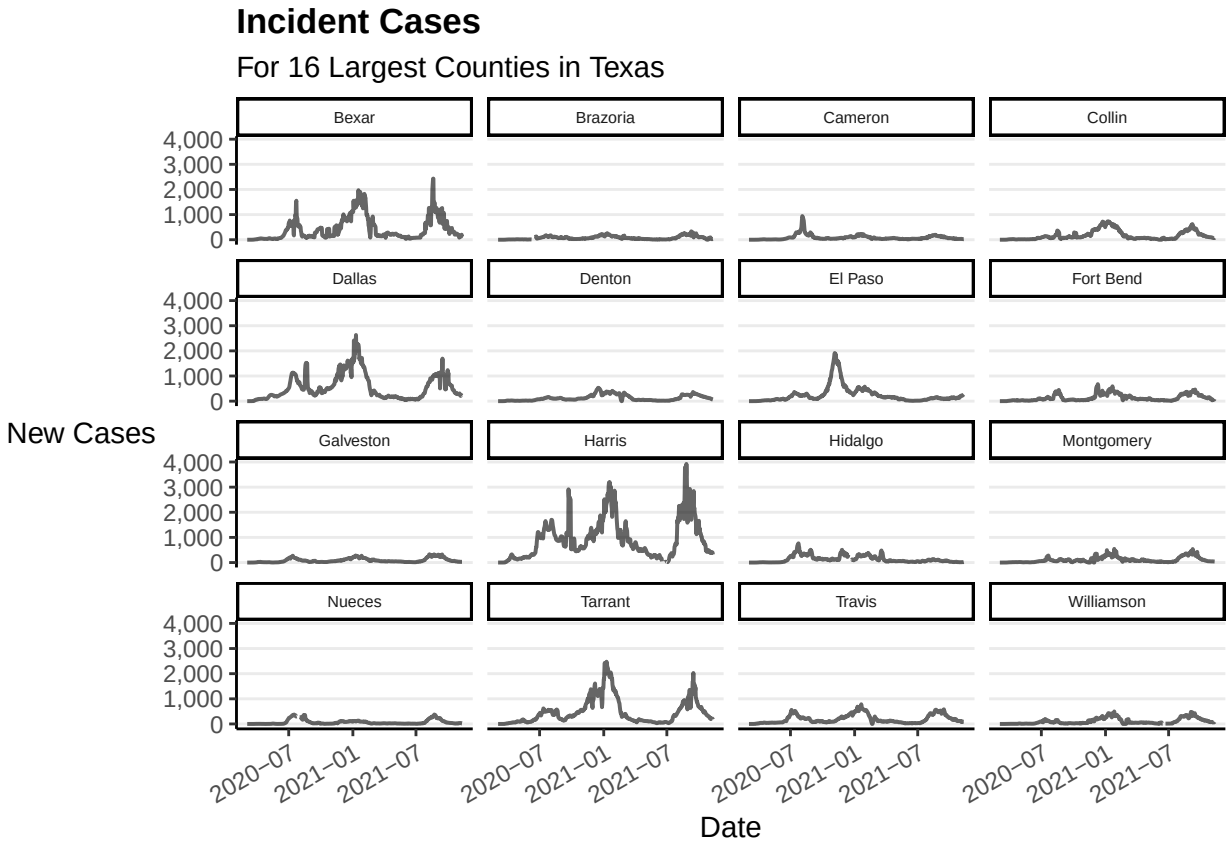
```
# Find top 16 counties, and total.
top_count <- tail(pop_dat[order(pop_dat$population), ], n = 17) |> pull(county)

cases |> filter(county %in% top_count, county != "Total") |>
  group_by(county) |>
  mutate(new_cases = c(cases[1], diff(cases)),
         new_cases = slide_dbl(new_cases, mean, .before = 6)) |>
  ggplot(aes(x = date, y = new_cases, group = county)) +
  geom_line(aes(color = county), size = .7, alpha = .6) +
  scale_y_continuous(labels = scales::comma, limits = c(0, NA)) +
  scale_color_viridis_d() +
  labs(title = "Incident Cases",
       subtitle = "For 16 Largest Counties in Texas",
       y = "Cases",
       x = "Date")
```



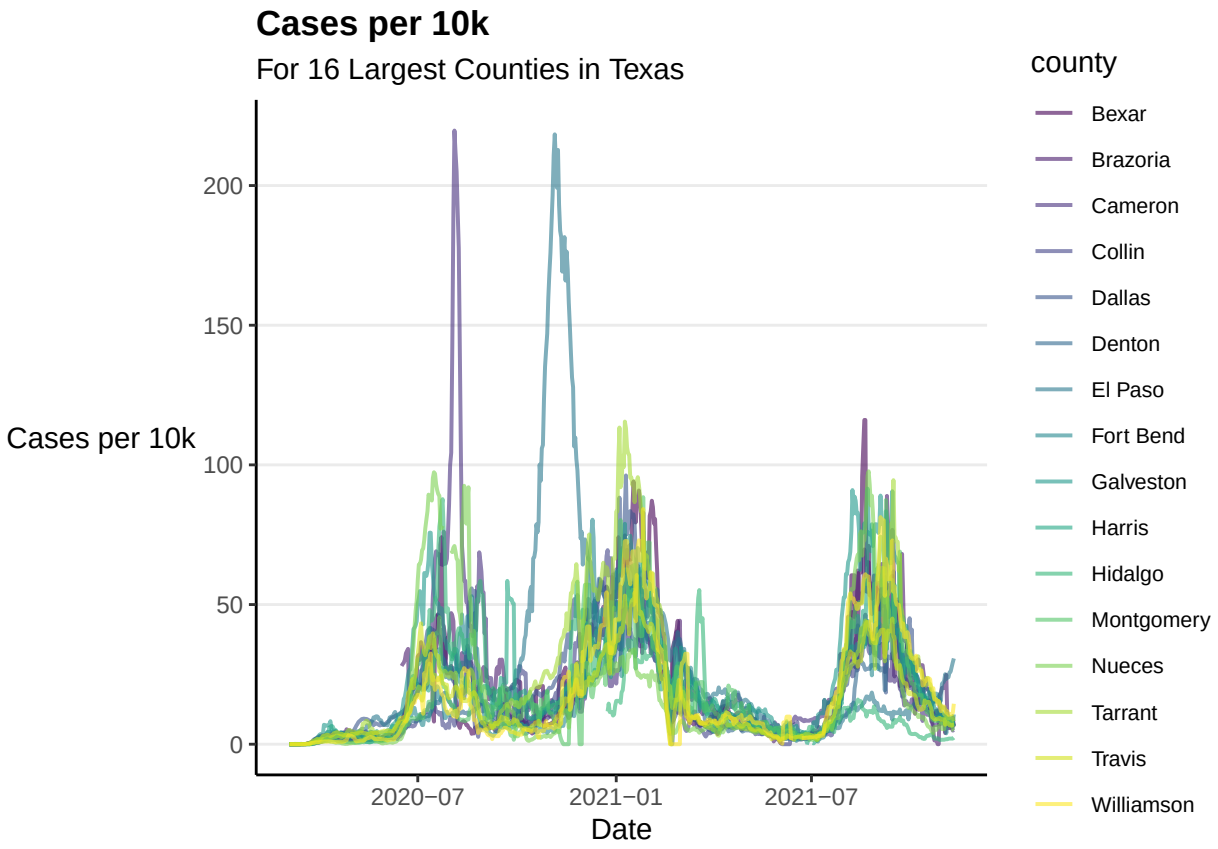
Part 4: Facet Graphics

```
cases |> filter(county %in% top_count, county != "Total") |>
  group_by(county) |>
  mutate(new_cases = c(cases[1], diff(cases)),
         new_cases = slide_dbl(new_cases, mean, .before = 6)) |>
  ggplot(aes(x = date, y = new_cases, group = county)) +
  geom_line(size = 0.7, alpha = .6) +
  scale_y_continuous(labels = scales::comma, limits = c(0,NA)) +
  labs(title = "Incident Cases",
       subtitle = "For 16 Largest Counties in Texas",
       x = "Date",
       y = "New Cases") +
  theme(legend.position = "none",
        axis.text.x = element_text(angle = 30, vjust = 1, hjust = 1),
        strip.text = element_text(size = 6))+
  facet_wrap(~county, nrow = 4)
```

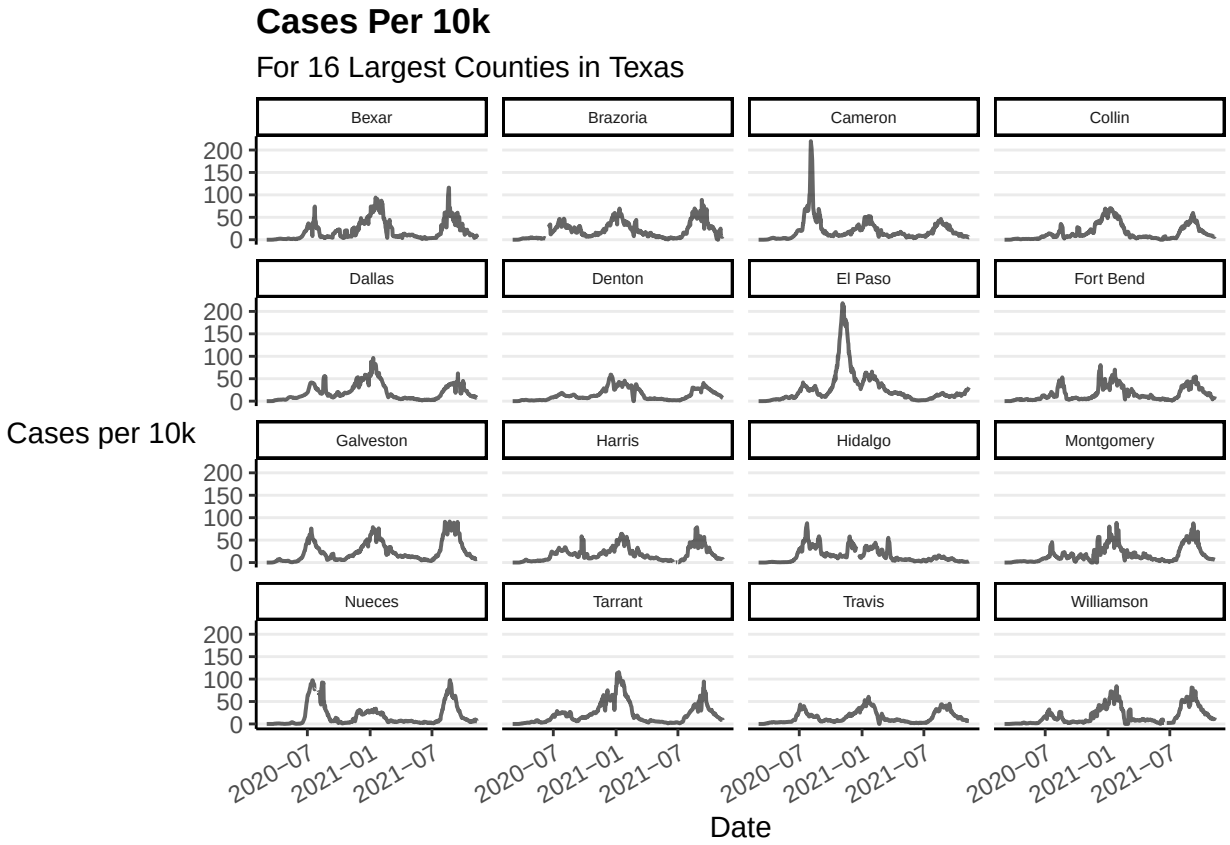


Part 5: Cases per 10k

```
cases |> filter(county %in% top_count, county != "Total") |>
  group_by(county) |>
  mutate(new_cases = c(cases[1], diff(cases)),
         new_cases = slide_dbl(new_cases, mean, .before = 6),
         cases_10k = (new_cases/population)*1e5) |>
  ggplot(aes(x = date, y = cases_10k, group = county)) +
  geom_line(aes(color = county), size = .7, alpha = .6) +
  scale_y_continuous(labels = scales::comma, limits = c(0, NA)) +
  scale_color_viridis_d() +
  labs(title = "Cases per 10k",
       subtitle = "For 16 Largest Counties in Texas",
       y = "Cases per 10k",
       x = "Date")
```




```
cases |> filter(county %in% top_count, county != "Total") |>
  group_by(county) |>
  mutate(new_cases = c(cases[1], diff(cases)),
         new_cases = slide_dbl(new_cases, mean, .before = 6),
         cases_10k = (new_cases/population)*1e5) |>
  ggplot(aes(x = date, y = cases_10k, group = county)) +
  geom_line(size = 0.7, alpha = .6) +
  scale_y_continuous(labels = scales::comma, limits = c(0,NA)) +
  labs(title = "Cases Per 10k",
       subtitle = "For 16 Largest Counties in Texas",
       x = "Date",
       y = "Cases per 10k") +
  theme(legend.position = "none",
        axis.text.x = element_text(angle = 30, vjust = 1, hjust = 1),
        strip.text = element_text(size = 6))+
  facet_wrap(~county, nrow = 4)
```


Cases Per 10k
For 16 Largest Counties in Texas
Bexar
Brazoria
Cameron
Collin
Dallas
Denton
El Paso
Fort Bend
Galveston
Harris
Hidalgo
Montgomery
Nueces
Tarrant
Travis
Williamson
200
150
100
50
0
2020–07
2021–01
2021–07
Cases per 10k
Date