king_dav_ps03

Dav King

2/24/2022

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(socviz)
library(slider)
library(covdata)
```

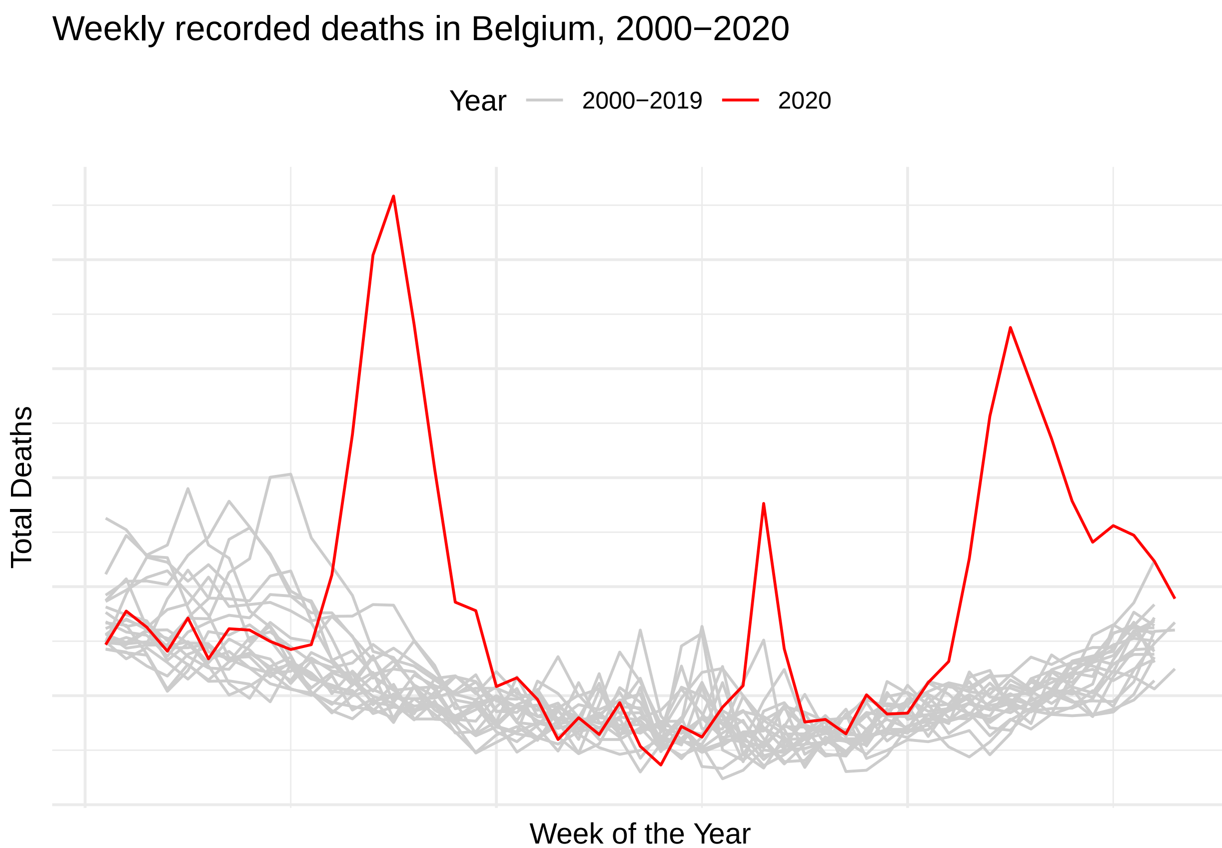
```
##
## Attaching package: 'covdata'

## The following object is masked from 'package:socviz':
##
##     %nin%

## The following object is masked from 'package:datasets':
##
##     uspop
```

```
library(cowplot)
```

Section A

1

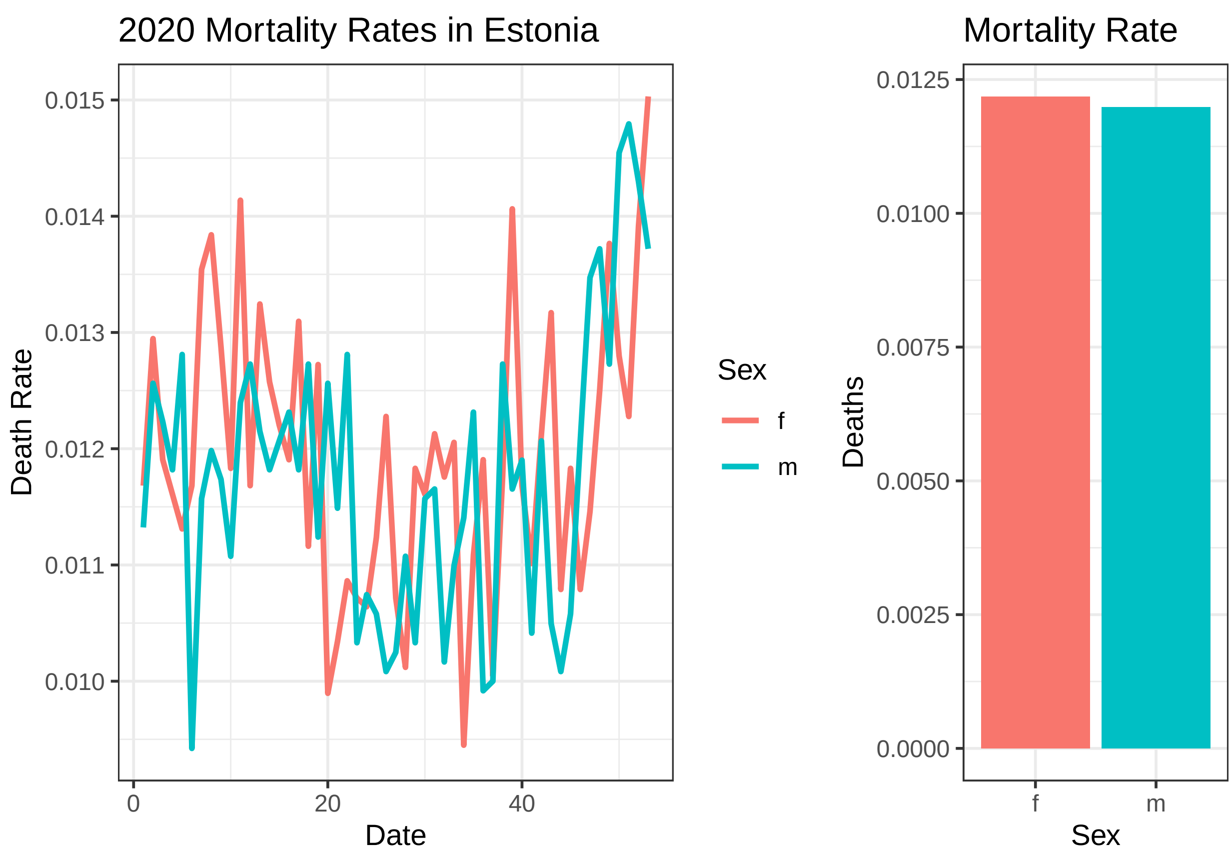
```r
stmf %>%
  filter(cname == "Belgium") %>%
  filter(year != 2021) %>%
  mutate(Year = if_else(year == 2020, "2020", "2000-2019")) %>%
  group_by(year, week) %>%
  mutate(deaths = sum(deaths_total)) %>%
  ggplot(aes(x = week, y = deaths, fill = factor(year), color = Year)) +
  geom_line(size = 0.5) +
  scale_color_manual(values=c("grey80", "red")) +
  labs(title = "Weekly recorded deaths in Belgium, 2000-2020",
       x = "Week of the Year", y = "Total Deaths") +
  theme_minimal() +
  theme(legend.position = "top", axis.text.x=element_blank(),
        axis.text.y=element_blank())
```



2

```r
a <- stmf %>%
  filter(cname == "Estonia" & year == "2020" & sex != "b") %>%
  ggplot(aes(x = week, y = rate_total, color = sex)) +
  geom_line(size = 1) +
  labs(title = "2020 Mortality Rates in Estonia",
       x = "Date", y = "Death Rate", color = "Sex") +
```

```
  theme_bw()

b <- stmf %>%
  filter(cname == "Estonia" & year == "2020" & sex != "b") %>%
  group_by(sex) %>%
  summarize(cases = sum(rate_total)/(52*5)) %>%
  ggplot(aes(x = sex, y = cases, fill = sex)) +
  geom_col() +
  labs(title = "Mortality Rate", x = "Sex", y = "Deaths") +
  guides(fill = "none") +
  theme_bw()

plot_grid(a, b, rel_widths = c(2, 1))
```

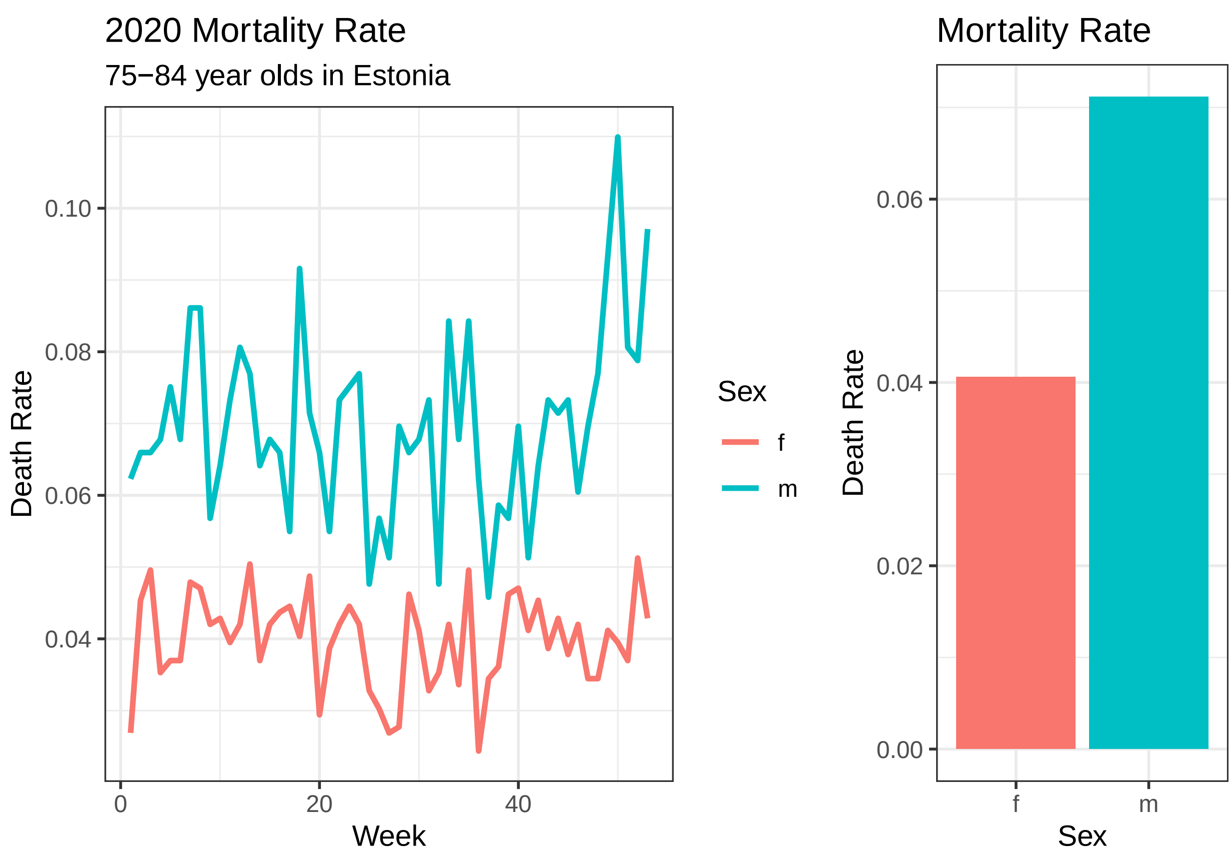


While there are some fluctuations in the mortality rates of Estonian men and women across 2020, ultimately their mortality rates were almost identical.

3

```
c <- stmf %>%
  filter(cname == "Estonia" & year == "2020" &
           age_group == "75-84" & sex != "b") %>%
  group_by(sex) %>%
  summarize(dRate = sum(death_rate)/52) %>%
```

```
  ggplot(aes(x = sex, y = dRate, fill = sex)) +
  geom_col() +
  guides(fill = "none") +
  labs(title = "Mortality Rate",
       x = "Sex", y = "Death Rate") +
  theme_bw()

d <- stmf %>%
  filter(cname == "Estonia" & year == "2020" &
           age_group == "75-84" & sex != "b") %>%
  ggplot(aes(x = week, y = death_rate, color = sex)) +
  geom_line(size = 1) +
  labs(title = "2020 Mortality Rate", subtitle = "75-84 year olds in Estonia",
       x = "Week", y = "Death Rate", color = "Sex") +
  theme_bw()

plot_grid(d, c, rel_widths = c(2, 1))
```



Here, there is a clear difference - older men died at much higher rates than older women (nearly 2x as many!)

Section B

```r
nytcovstate
```

```
## # A tibble: 32,766 x 5
##    date       state      fips  cases deaths
##    <date>     <chr>      <chr> <dbl>  <dbl>
##  1 2020-01-21 Washington 53        1      0
##  2 2020-01-22 Washington 53        1      0
##  3 2020-01-23 Washington 53        1      0
##  4 2020-01-24 Illinois   17        1      0
##  5 2020-01-24 Washington 53        1      0
##  6 2020-01-25 California 06        1      0
##  7 2020-01-25 Illinois   17        1      0
##  8 2020-01-25 Washington 53        1      0
##  9 2020-01-26 Arizona    04        1      0
## 10 2020-01-26 California 06        2      0
## # ... with 32,756 more rows
```

```r
nytcovstate <- nytcovstate %>%
  group_by(state) %>%
  mutate(daily_cases = cases - lag(cases, order_by = date),
         daily_deaths = deaths - lag(deaths, order_by = date))

state_pops <- uspop %>%
  filter(sex_id == "totsex", hisp_id == "tothisp") %>%
  select(state_abbr, statefips, pop, state) %>%
  rename(name = state, state = state_abbr, fips = statefips) %>%
  mutate(state = replace(state, fips == "11", "DC"))
```

1

```r
nytcovstate <- nytcovstate %>%
  left_join(state_pops, by = "fips")
```

2

```r
nytcovstate %>%
  group_by(state.x) %>%
  summarize(nCase = sum(daily_cases, na.rm = T),
            nDeath = sum(daily_deaths, na.rm = T),
            popul = sum(pop, na.rm = T)) %>%
  summarize(state.x, case_rate = (nCase/popul) * 100000,
            death_rate = (nDeath/popul) * 100000)
```

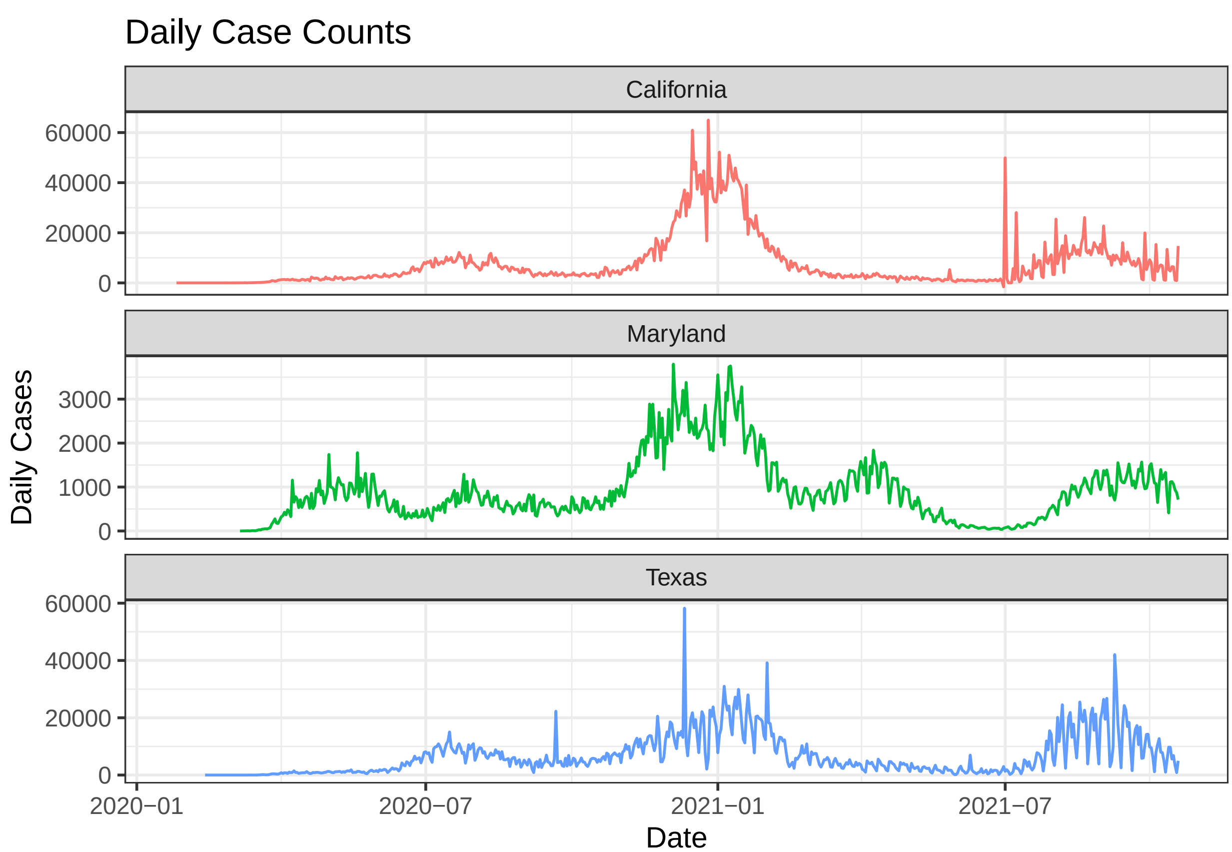
```
## # A tibble: 56 x 3
##    state.x              case_rate death_rate
##    <chr>                    <dbl>      <dbl>
##  1 Alabama                   28.5      0.531
```

```
##   2 Alaska                     29.4      0.139
##   3 American Samoa            NaN       NaN
##   4 Arizona                    25.0      0.452
##   5 Arkansas                   28.7      0.464
##   6 California                 19.3      0.283
##   7 Colorado                   21.1      0.240
##   8 Connecticut                18.9      0.414
##   9 Delaware                   24.7      0.358
##  10 District of Columbia       15.2      0.285
## # ... with 46 more rows
```

3

```r
nytcovstate %>%
  filter(state.y == "MD" | state.y == "TX" | state.y == "CA") %>%
  ggplot(aes(x = date, y = daily_cases, color = state.x)) +
  geom_line() +
  facet_wrap(~ state.x, ncol = 1, scales = "free_y") +
  labs(title = "Daily Case Counts", x = "Date", y = "Daily Cases") +
  guides(color = "none") +
  theme_bw()
```

```
## Warning: Removed 3 row(s) containing missing values (geom_path).
```

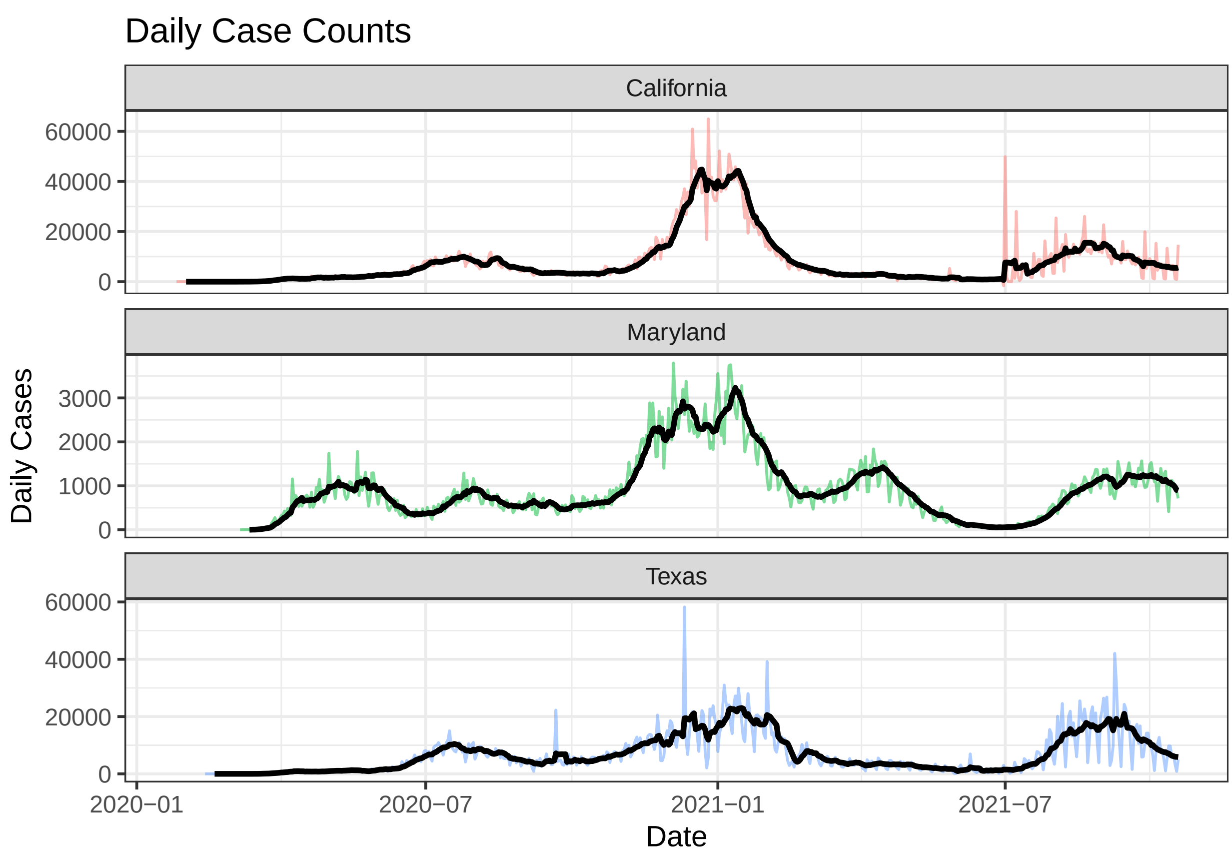


4

```r
nytcovstate %>%
  filter(state.y == "MD" | state.y == "TX" | state.y == "CA") %>%
  group_by(state.x) %>%
  mutate(avg = slide_dbl(daily_cases, mean, .before = 6)) %>%
  ggplot(aes(x = date, color = state.x)) +
  geom_line(aes(y = daily_cases), alpha = 0.5) +
  geom_line(aes(y = avg), color = "black", size = 1) +
  facet_wrap(~ state.x, ncol = 1, scales = "free_y") +
  labs(title = "Daily Case Counts", x = "Date", y = "Daily Cases") +
  guides(color = "none") +
  theme_bw()
```

```
## Warning: Removed 3 row(s) containing missing values (geom_path).
```

```
## Warning: Removed 7 row(s) containing missing values (geom_path).
```



Section C

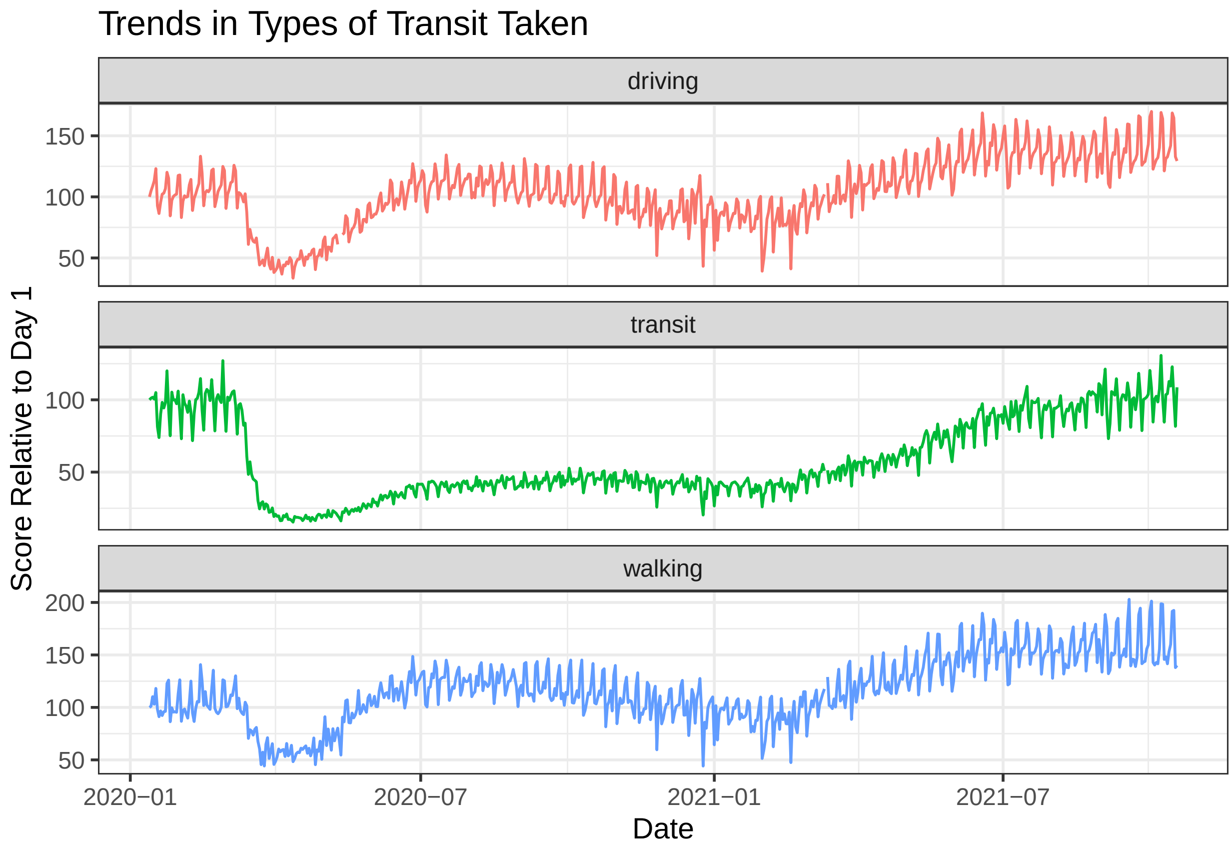
1

```r
apple_mobility %>%
  filter(region == "Montgomery County" & sub_region == "Maryland")
```

```
## # A tibble: 1,935 x 8
##    geo_type region      transportation_ty~ alternative_name sub_region country
##    <chr>    <chr>       <chr>              <chr>            <chr>      <chr>
##  1 county   Montgomery~ driving            <NA>             Maryland   United S~
##  2 county   Montgomery~ driving            <NA>             Maryland   United S~
##  3 county   Montgomery~ driving            <NA>             Maryland   United S~
##  4 county   Montgomery~ driving            <NA>             Maryland   United S~
##  5 county   Montgomery~ driving            <NA>             Maryland   United S~
##  6 county   Montgomery~ driving            <NA>             Maryland   United S~
##  7 county   Montgomery~ driving            <NA>             Maryland   United S~
##  8 county   Montgomery~ driving            <NA>             Maryland   United S~
##  9 county   Montgomery~ driving            <NA>             Maryland   United S~
## 10 county   Montgomery~ driving            <NA>             Maryland   United S~
## # ... with 1,925 more rows, and 2 more variables: date <date>, score <dbl>
```

2

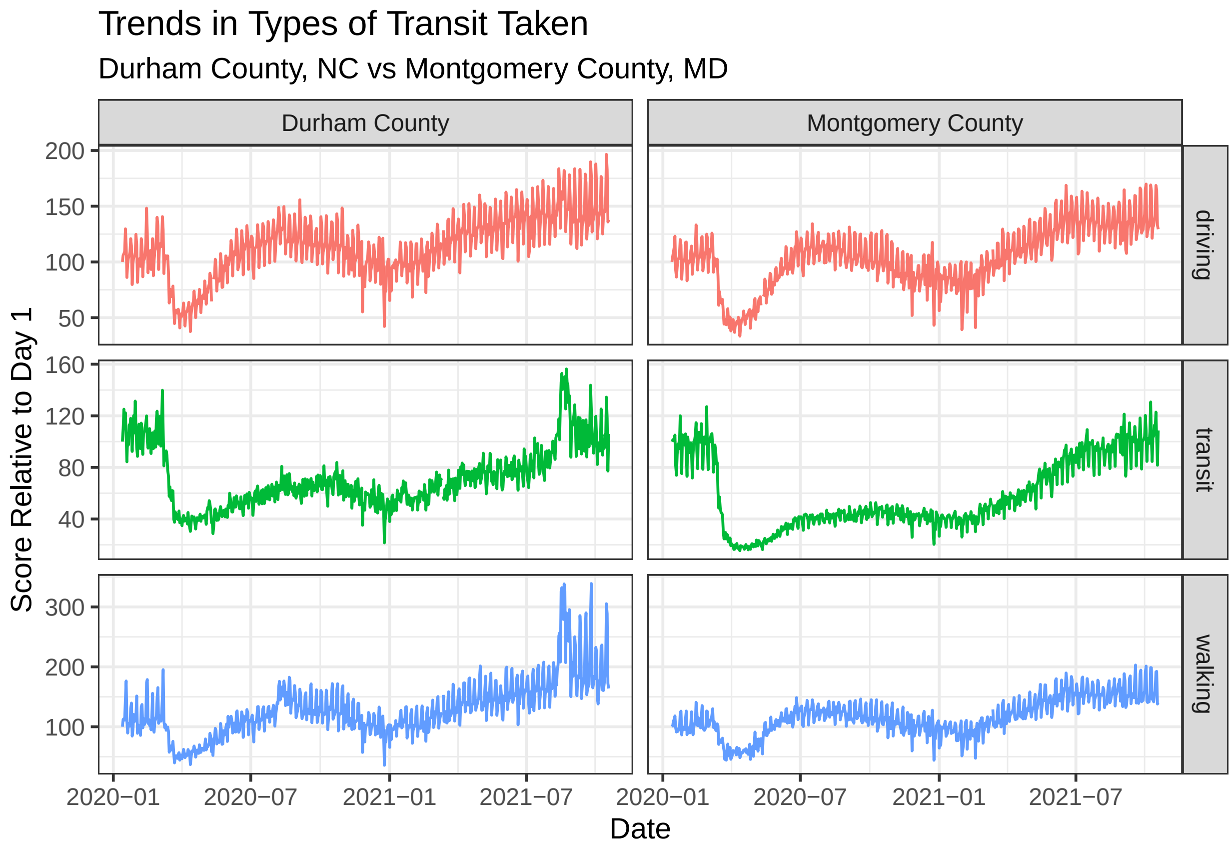
```r
apple_mobility %>%
  filter(region == "Montgomery County" & sub_region == "Maryland") %>%
  ggplot(aes(x = date, y = score, color = transportation_type)) +
  geom_line() +
  facet_wrap(~ transportation_type, ncol = 1, scales = "free_y") +
  labs(title = "Trends in Types of Transit Taken", x = "Date",
       y = "Score Relative to Day 1") +
  guides(color = "none") +
  theme_bw()
```



3

This pattern really follows (inversely) the trend in Covid cases over time, such that all types of transit dropped heavily in March 2020 as Covid hit, but that walking and driving (more isolated, personal forms of transit) recovered much more quickly than public transit did. On the most basic level, it tells us about quite explicitly what forms of transit people were taken. If we wanted to compare two or more places (see below), we would likely see similar trends - though they would depend on the makeup of each location (a county that does not already have much public transit, for example, would likely not have much to lose; a country which had Covid spikes at different times would show slightly different trends).

```
apple_mobility %>%
  filter(sub_region == "Maryland" & region == "Montgomery County" |
         sub_region == "North Carolina" & region == "Durham County") %>%
  ggplot(aes(x = date, y = score, color = transportation_type)) +
  geom_line() +
  facet_grid(transportation_type ~ region, scales = "free_y") +
  labs(title = "Trends in Types of Transit Taken",
       subtitle = "Durham County, NC vs Montgomery County, MD",
       x = "Date", y = "Score Relative to Day 1") +
  guides(color = "none") +
  theme_bw()
```


Trends in Types of Transit Taken
Durham County, NC vs Montgomery County, MD
Durham County
Montgomery County
driving
transit
walking
Score Relative to Day 1
200
150
100
50
160
120
80
40
300
200
100
2020−01
2020−07
2021−01
2021−07
Date