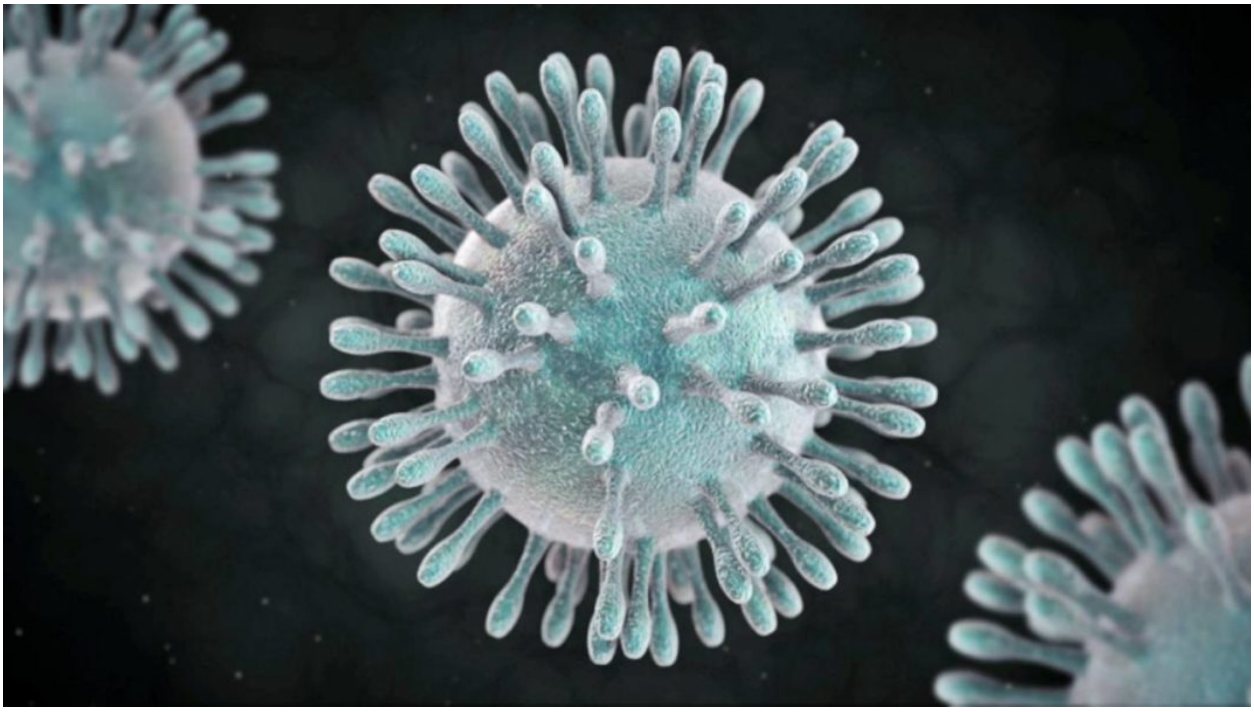
COVID-19 'Flattening The Curve' Graphs

Hrvoje Krpan



Description

v.1.

- In this workbook, I will show how to (re)create a graph that was presented by a Croatian health minister **Vili Beroš** by using ggplot package
- I saw graph originally at Slobodna Dalmacija portal. Link to original article.
- Each point on a red line is representing an average of 5 previous days. For instance, the red line over "ožu 30" (Mar 30 in Croatian) is the mean of cases that happened between 26.- 30. March.
- With a little adjustment of moving average, this plot can be used in salse purposes as well.

v.1.1.

- Sweden is a country that has taken a different approach to fighting the coronavirus. That was the reason I wanted to see how it compares to Croatia.
- After gathering the data for Sweden, the moving average was calculated (for 7 days) and plotted beside Croatia on one graph. I have used a logarithmic scale.

Install Packages

```
#devtools::install_github("covid19r/coronavirus")
#install.packages("tidyverse")
#install.packages("RcppRoll")
```

Load packages

```
library(coronavirus)
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v ggplot2 3.3.0     v purrr   0.3.3
## v tibble  3.0.0     v dplyr   0.8.5
## v tidyr   1.0.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

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

```
library(RcppRoll)
theme_set(theme_bw())
```

Getting the data

```
data("coronavirus")
tail(coronavirus)
```

```
##       Province.State Country.Region     Lat     Long       date cases      type
## 64569       Zhejiang          China 29.1832 120.0934 2020-04-08     2 recovered
## 64570       Zhejiang          China 29.1832 120.0934 2020-04-09     3 recovered
## 64571       Zhejiang          China 29.1832 120.0934 2020-04-10     0 recovered
## 64572       Zhejiang          China 29.1832 120.0934 2020-04-11     1 recovered
## 64573       Zhejiang          China 29.1832 120.0934 2020-04-12     2 recovered
## 64574       Zhejiang          China 29.1832 120.0934 2020-04-13     1 recovered
```

```
swe_covid19_basic <- coronavirus %>%
  filter(Country.Region == "Sweden")

tail(swe_covid19_basic)
```

```
##     Province.State Country.Region Lat Long       date cases      type
## 244                        Sweden  63   16 2020-04-08     0 recovered
## 245                        Sweden  63   16 2020-04-09     0 recovered
```

```
## 246                         Sweden  63   16 2020-04-10   176 recovered
## 247                         Sweden  63   16 2020-04-11     0 recovered
## 248                         Sweden  63   16 2020-04-12     0 recovered
## 249                         Sweden  63   16 2020-04-13     0 recovered
```

```
cro_covid19_basic <- coronavirus %>%
  filter(Country.Region == "Croatia")

tail(cro_covid19_basic)
```

```
##     Province.State Country.Region  Lat Long       date cases      type
## 244                       Croatia 45.1 15.2 2020-04-08    12 recovered
## 245                       Croatia 45.1 15.2 2020-04-09    40 recovered
## 246                       Croatia 45.1 15.2 2020-04-10    12 recovered
## 247                       Croatia 45.1 15.2 2020-04-11    92 recovered
## 248                       Croatia 45.1 15.2 2020-04-12    50 recovered
## 249                       Croatia 45.1 15.2 2020-04-13    27 recovered
```

```
cro_covid19 <- coronavirus %>%
    select(Country.Region, cases, type, date) %>%
    group_by(type) %>%
    filter(Country.Region=="Croatia") %>%
    pivot_wider(names_from = type, values_from = cases) %>%
    rename(country=Country.Region) %>%
    mutate(rollavg = roll_meanr(confirmed, n = 7)) %>%
    filter(confirmed > 15) %>%
    arrange(desc(date))

head(cro_covid19)
```

```
## # A tibble: 6 x 6
##   country date       confirmed death recovered rollavg
##   <chr>   <date>         <int> <int>     <int>   <dbl>
## 1 Croatia 2020-04-13        50     2        27    61.1
## 2 Croatia 2020-04-12        66     2        50    59.7
## 3 Croatia 2020-04-11        39     0        92    58.3
## 4 Croatia 2020-04-10        88     1        12    59.4
## 5 Croatia 2020-04-09        64     1        40    56.6
## 6 Croatia 2020-04-08        61     1        12    54.3
```

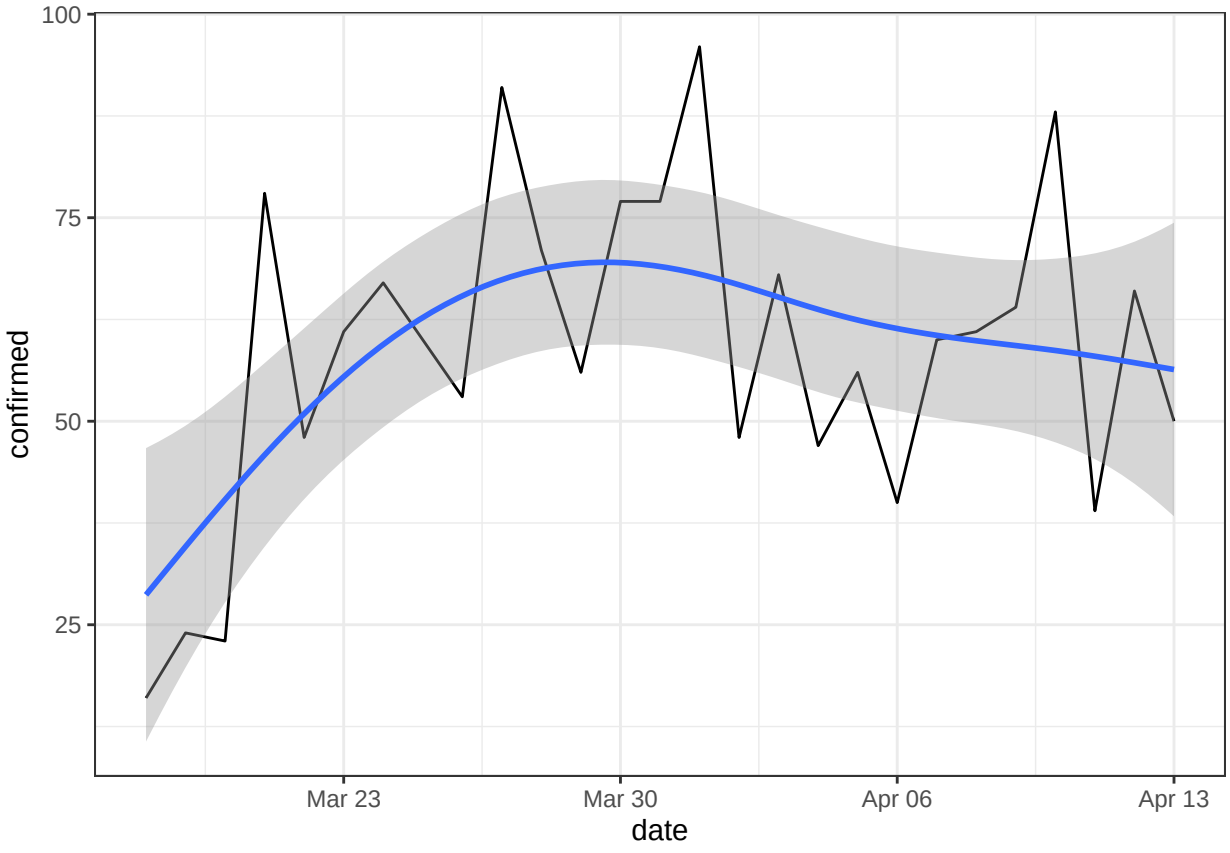
```
swe_covid19 <- coronavirus %>%
    select(Country.Region, cases, type, date) %>%
    group_by(type) %>%
    filter(Country.Region=="Sweden") %>%
    pivot_wider(names_from = type, values_from = cases) %>%
    rename(country=Country.Region) %>%
    mutate(rollavg = roll_meanr(confirmed, n = 7)) %>%
    filter(confirmed > 15) %>%
    arrange(desc(date))

head(swe_covid19)
```

```
## # A tibble: 6 x 6
##   country date       confirmed death recovered rollavg
##   <chr>   <date>         <int> <int>     <int>   <dbl>
## 1 Sweden  2020-04-13       465    20         0    535.
## 2 Sweden  2020-04-12       332    12         0    522.
## 3 Sweden  2020-04-11       466    17         0    530.
## 4 Sweden  2020-04-10       544    77       176    508.
## 5 Sweden  2020-04-09       722   106         0    510.
## 6 Sweden  2020-04-08       726    96         0    496
```

Data Visualization

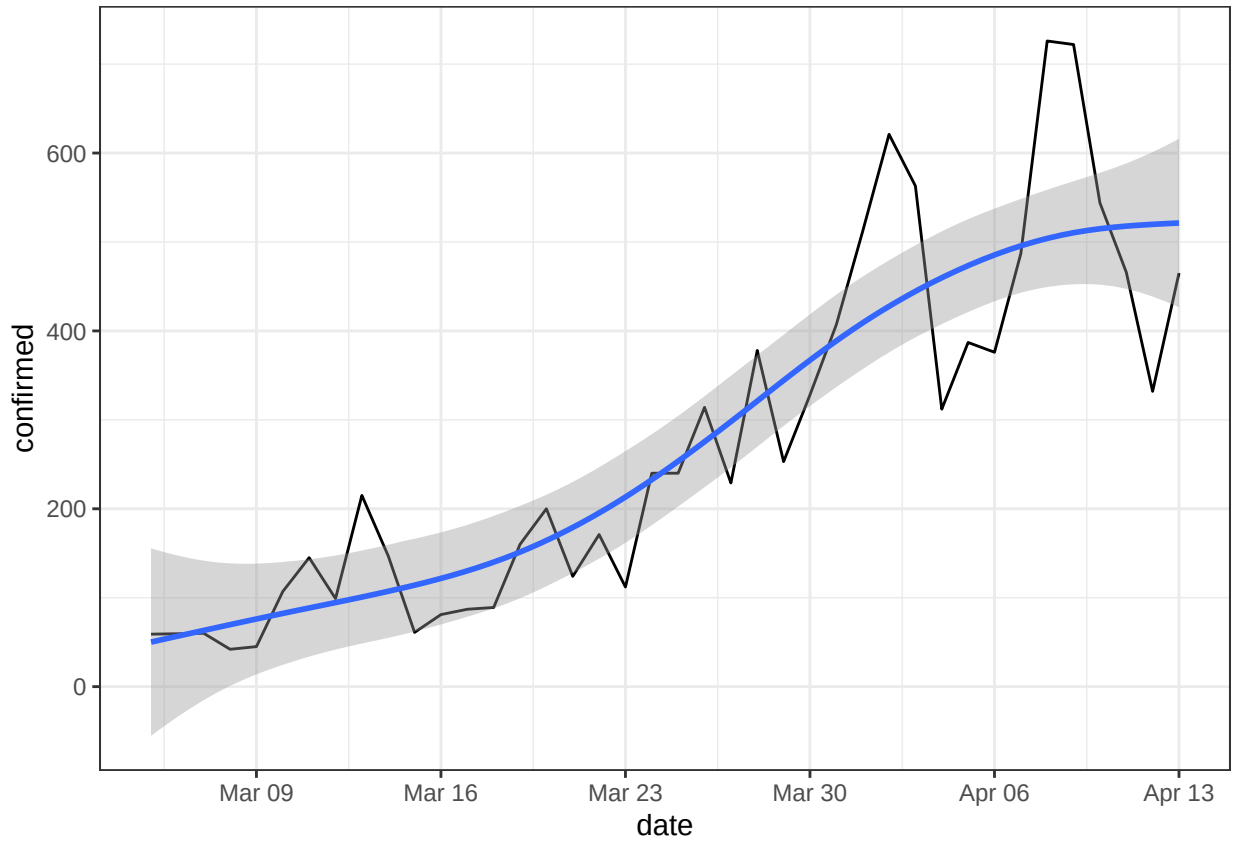
```
cro_covid19 %>%
  ggplot(aes(x = date, y = confirmed)) +
  geom_line() +
  geom_smooth(method = "gam")
```

```
## `geom_smooth()` using formula 'y ~ s(x, bs = "cs")'
```

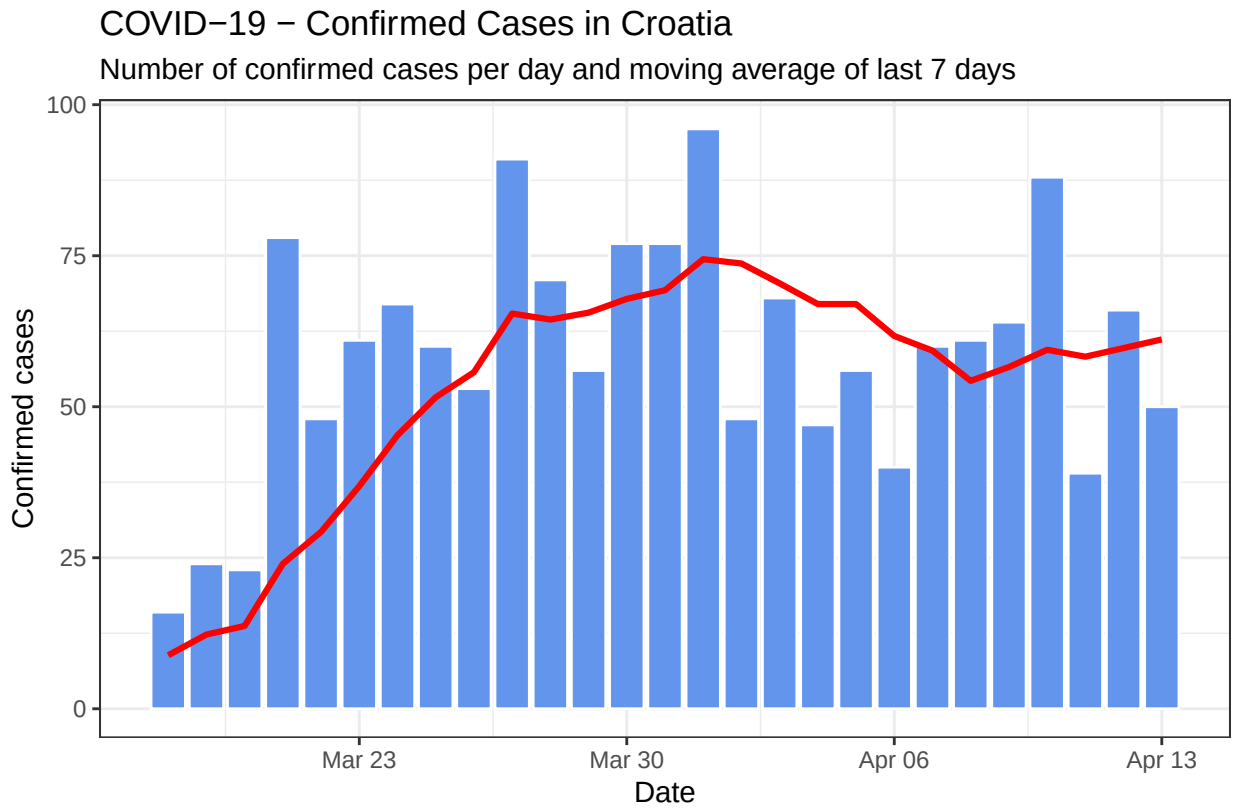



```
swe_covid19 %>%
  ggplot(aes(x = date, y = confirmed)) +
  geom_line() +
  geom_smooth(method = "gam")
```

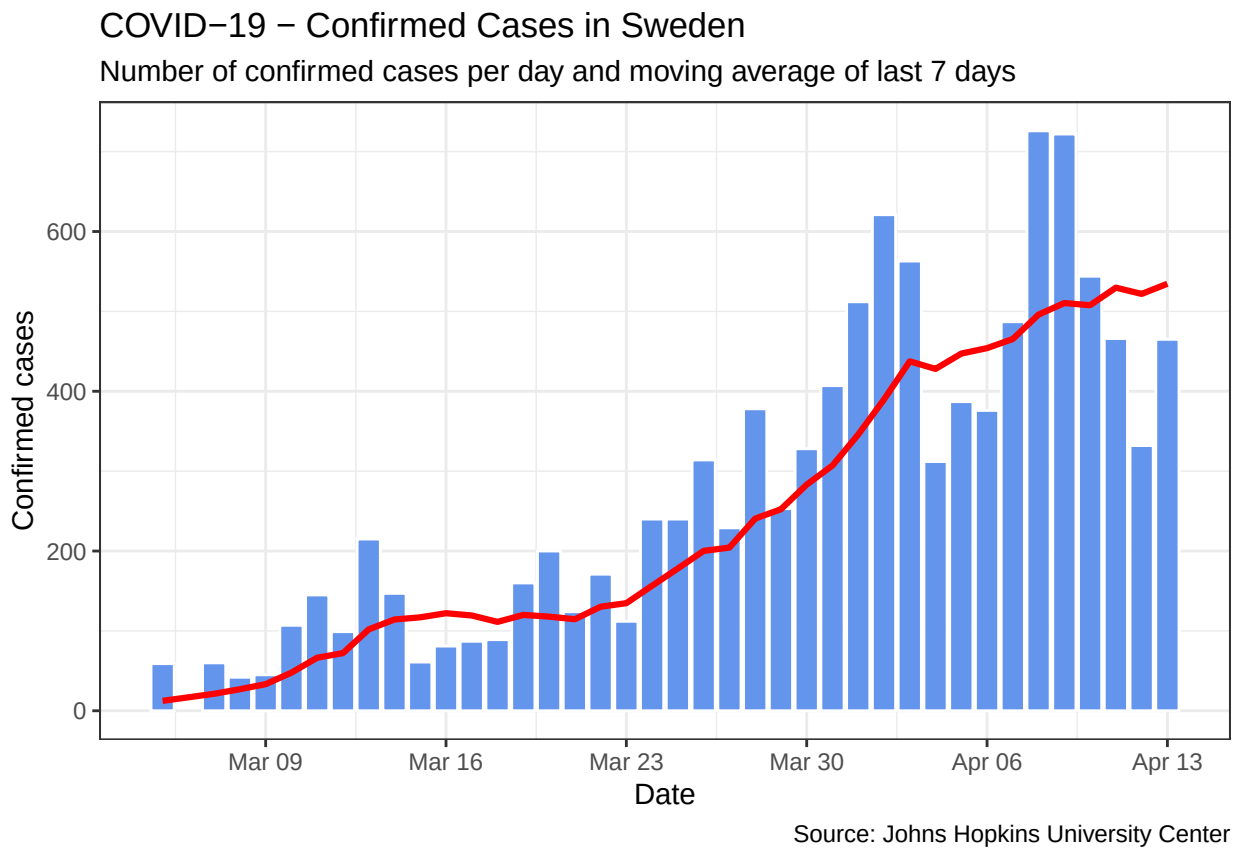
```
## `geom_smooth()` using formula 'y ~ s(x, bs = "cs")'
```




```
cro_covid19 %>%
  ggplot(aes(x = date, y = confirmed)) +
  geom_bar(stat = "identity",fill = "cornflowerblue", color = "white")+
  geom_line(aes(x = date, y = rollavg),color = "red", size = 1.1) +
  labs(subtitle="Number of confirmed cases per day and moving average of last 7 days",
       y="Confirmed cases",
       x="Date",
       title="COVID-19 - Confirmed Cases in Croatia",
       caption="Source: Johns Hopkins University Center") +
  scale_x_date(date_breaks = '1 week', date_labels = "%b %d")
```




```
swe_covid19 %>%
  ggplot(aes(x = date, y = confirmed)) +
  geom_bar(stat = "identity",fill = "cornflowerblue", color = "white")+
  geom_line(aes(x = date, y = rollavg),color = "red", size = 1.1) +
  labs(subtitle="Number of confirmed cases per day and moving average of last 7 days",
       y="Confirmed cases",
       x="Date",
       title="COVID-19 - Confirmed Cases in Sweden",
       caption="Source: Johns Hopkins University Center") +
  scale_x_date(date_breaks = '1 week', date_labels = "%b %d")
```




```
require(scales)
```

```
## Loading required package: scales
```

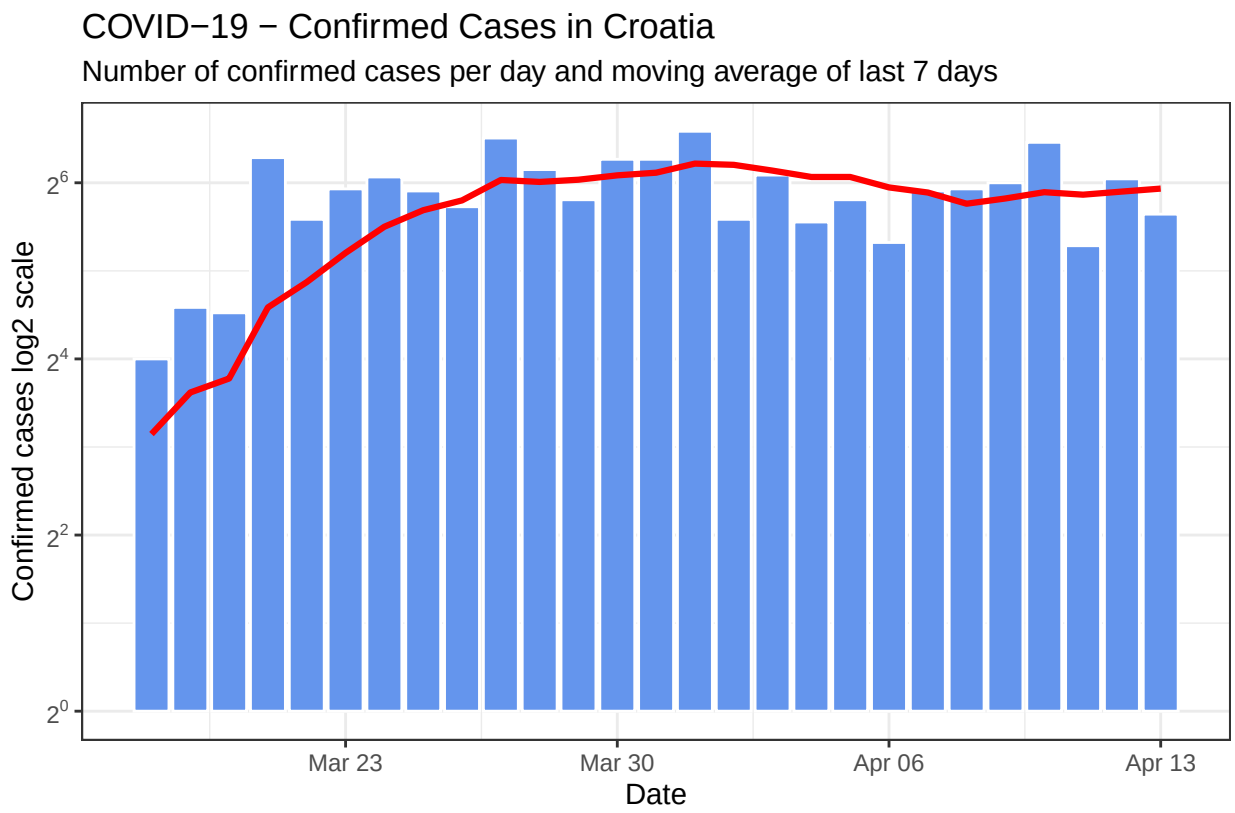
```
##
## Attaching package: 'scales'
```

```
## The following object is masked from 'package:purrr':
##
##     discard
```

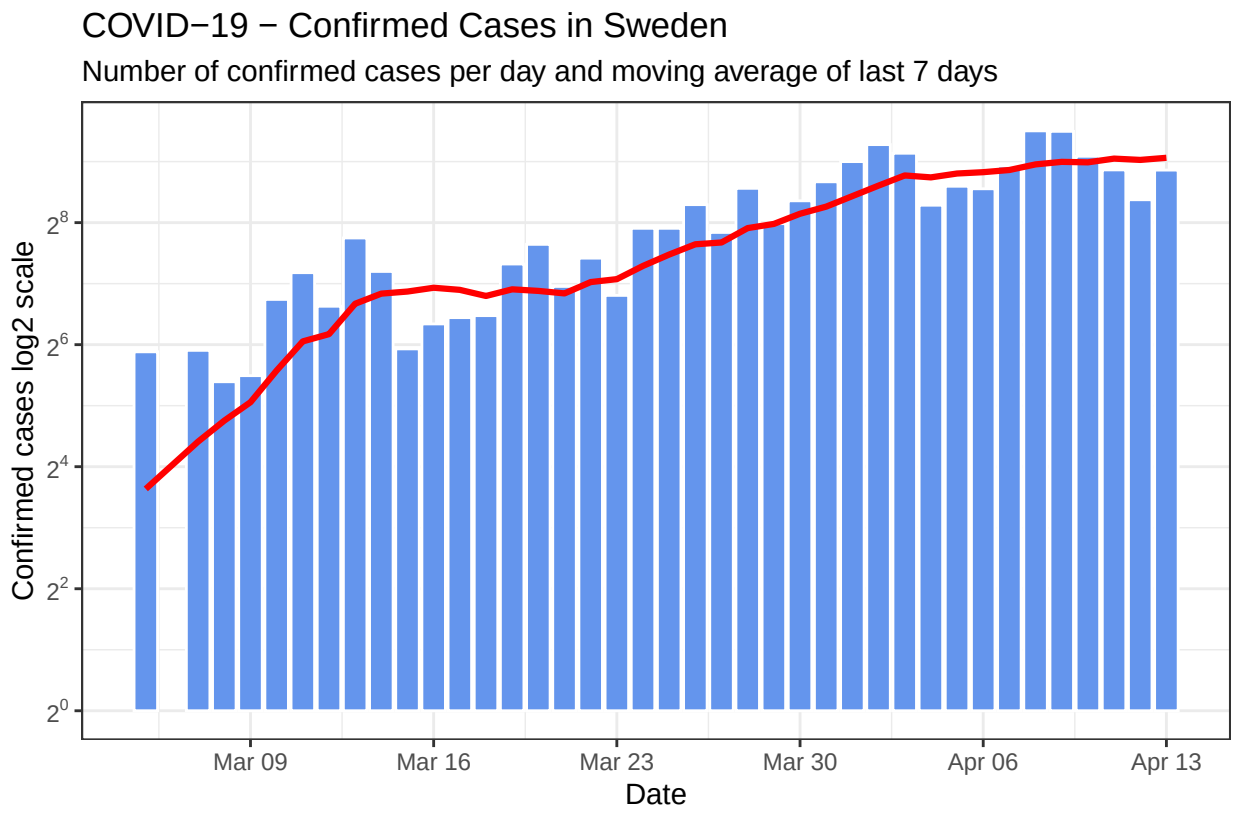
```
## The following object is masked from 'package:readr':
##
##     col_factor
```

```
cro_covid19 %>%
  ggplot(aes(x = date, y = confirmed)) +
  geom_bar(stat = "identity",fill = "cornflowerblue", color = "white")+
  geom_line(aes(x = date, y = rollavg),color = "red", size = 1.1) +
  labs(subtitle="Number of confirmed cases per day and moving average of last 7 days",
       y="Confirmed cases log2 scale",
       x="Date",
       title="COVID-19 - Confirmed Cases in Croatia",
       caption="Source: Johns Hopkins University Center") +
```

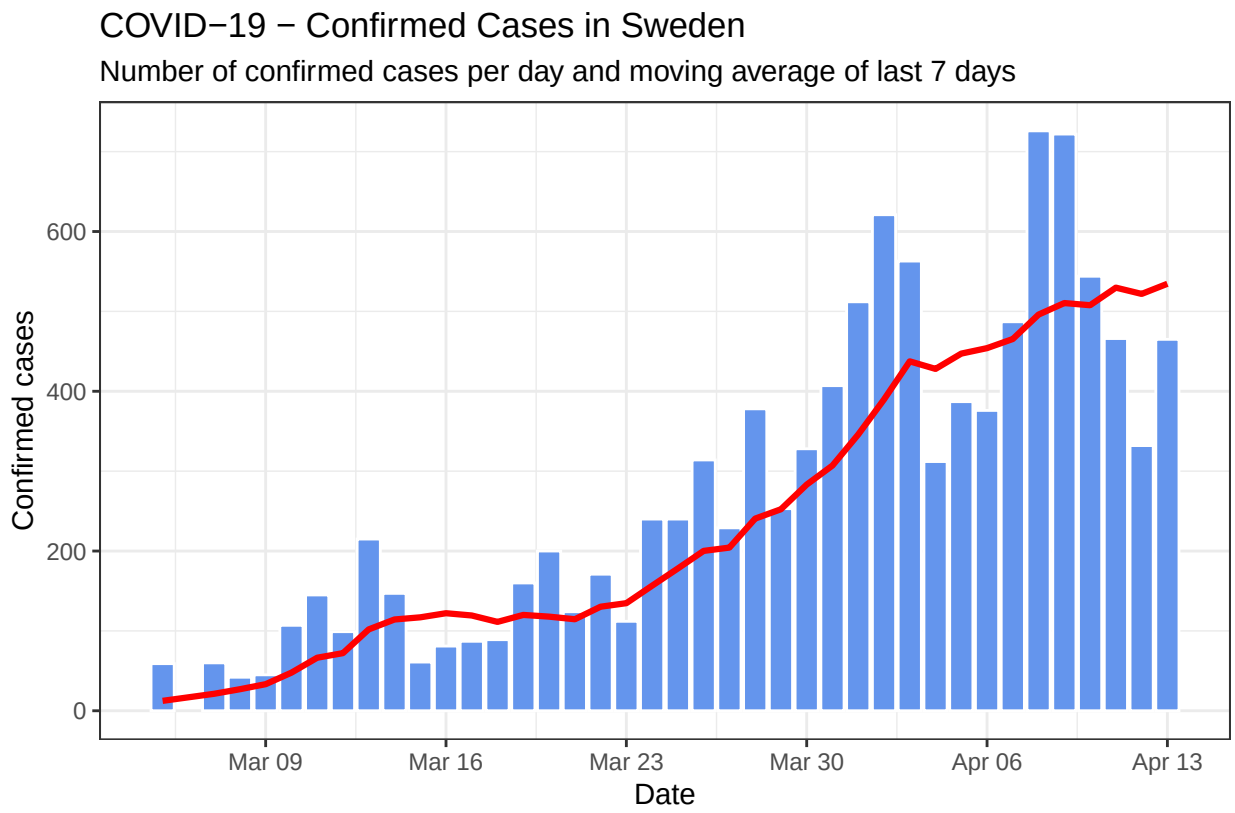
```
  scale_x_date(date_breaks = '1 week', date_labels = "%b %d") +
  scale_y_continuous(trans = log2_trans(),
    breaks = trans_breaks("log2", function(x) 2^x),
    labels = trans_format("log2", math_format(2^.x)))
```




```
swe_covid19 %>%
  ggplot(aes(x = date, y = confirmed)) +
  geom_bar(stat = "identity",fill = "cornflowerblue", color = "white")+
  geom_line(aes(x = date, y = rollavg),color = "red", size = 1.1) +
  labs(subtitle="Number of confirmed cases per day and moving average of last 7 days",
       y="Confirmed cases log2 scale",
       x="Date",
       title="COVID-19 - Confirmed Cases in Sweden",
       caption="Source: Johns Hopkins University Center") +
  scale_x_date(date_breaks = '1 week', date_labels = "%b %d")+
  scale_y_continuous(trans = log2_trans(),
    breaks = trans_breaks("log2", function(x) 2^x),
    labels = trans_format("log2", math_format(2^.x)))
```




```
swe_covid19 %>%
  ggplot(aes(x = date, y = confirmed)) +
  geom_bar(stat = "identity",fill = "cornflowerblue", color = "white")+
  geom_line(aes(x = date, y = rollavg),color = "red", size = 1.1) +
  labs(subtitle="Number of confirmed cases per day and moving average of last 7 days",
       y="Confirmed cases",
       x="Date",
       title="COVID-19 - Confirmed Cases in Sweden",
       caption="Source: Johns Hopkins University Center") +
  scale_x_date(date_breaks = '1 week', date_labels = "%b %d")
```




```
swecro_join <- right_join(cro_covid19, swe_covid19, by="date")

head(swecro_join)
```

```
## # A tibble: 6 x 11
##   country.x date       confirmed.x death.x recovered.x rollavg.x country.y
##   <chr>     <date>           <int>   <int>       <int>     <dbl> <chr>
## 1 Croatia   2020-04-13          50       2          27      61.1 Sweden
## 2 Croatia   2020-04-12          66       2          50      59.7 Sweden
## 3 Croatia   2020-04-11          39       0          92      58.3 Sweden
## 4 Croatia   2020-04-10          88       1          12      59.4 Sweden
## 5 Croatia   2020-04-09          64       1          40      56.6 Sweden
## 6 Croatia   2020-04-08          61       1          12      54.3 Sweden
## # ... with 4 more variables: confirmed.y <int>, death.y <int>,
## #   recovered.y <int>, rollavg.y <dbl>
```

```
swecro_join %>% ggplot(aes(x=date)) +
  geom_line(aes(y=rollavg.x, color = "Croatia"),size = 1.1) +
  geom_line(aes(y=rollavg.y, color = "Sweden"), size = 1.1) +
  scale_y_continuous(trans = log2_trans(),
    breaks = trans_breaks("log2", function(x) 2^x),
    labels = trans_format("log2", math_format(2^.x))) +
  labs(subtitle="Moving average of last 7 days for Sweden and Croatia",
       y="Confirmed cases (log2 scale)",
       x="Date",
```

```
      title="COVID-19 - Confirmed Cases in Sweden and Croatia",
      caption="Source: Johns Hopkins University Center") +
scale_colour_manual("",
                    breaks = c("Croatia", "Sweden"),
                    values = c("blue", "yellow")) +
scale_x_date(date_breaks = '1 week', date_labels = "%b %d")
```