Visualizing COVID-19

Rae

Load the readr, ggplot2, and dplyr packages

```
library(readr)
```

```
## Registered S3 methods overwritten by 'tibble':
##   method     from
##   format.tbl pillar
##   print.tbl  pillar
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

1. From epidemic to pandemic

```
# Read datasets/confirmed_cases_worldwide.csv into confirmed_cases_worldwide
confirmed_cases_worldwide <- read_csv("datasets/confirmed_cases_worldwide.csv")
```

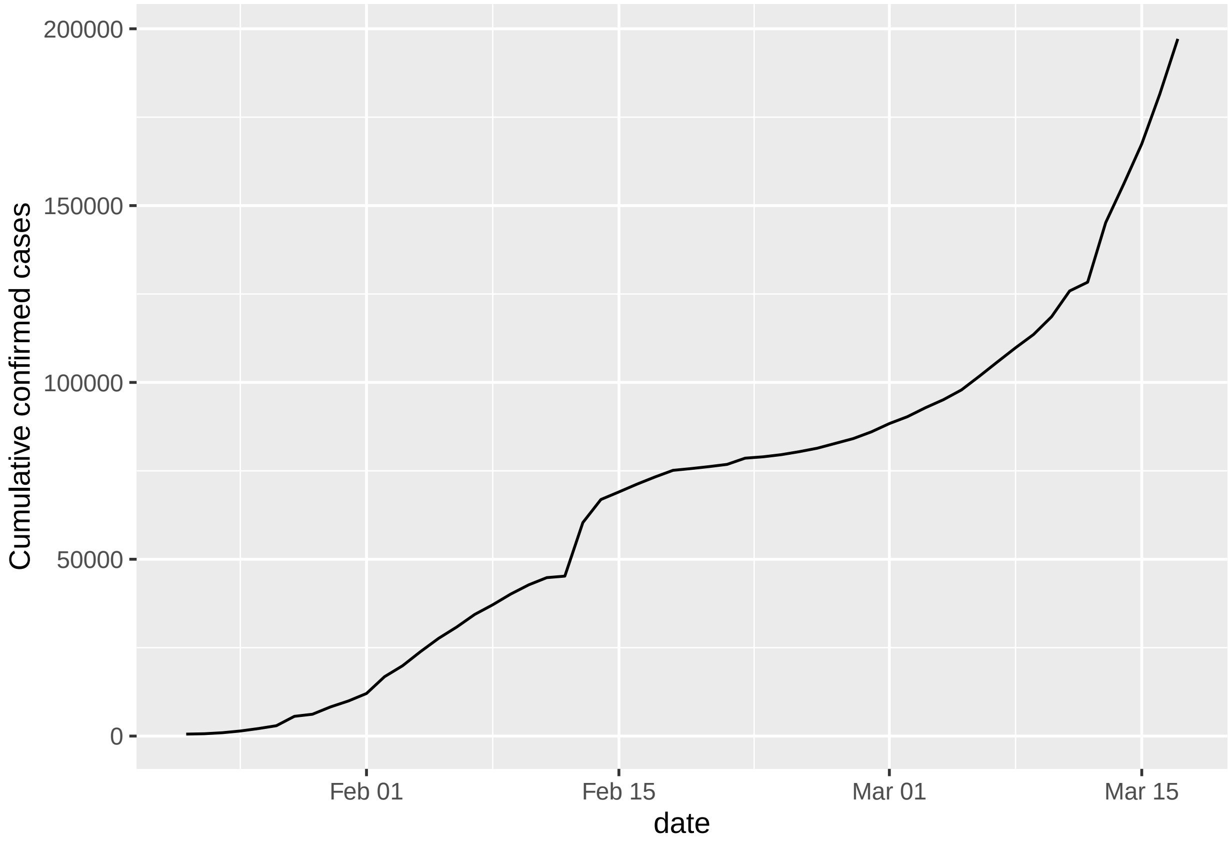
```
## Parsed with column specification:
## cols(
##   date = col_date(format = ""),
##   cum_cases = col_double()
## )
```

```
# See the result
confirmed_cases_worldwide
```

```
## # A tibble: 56 x 2
##    date       cum_cases
##    <date>         <dbl>
##  1 2020-01-22       555
##  2 2020-01-23       653
##  3 2020-01-24       941
##  4 2020-01-25      1434
##  5 2020-01-26      2118
##  6 2020-01-27      2927
##  7 2020-01-28      5578
##  8 2020-01-29      6166
##  9 2020-01-30      8234
## 10 2020-01-31      9927
## # ... with 46 more rows
```

2. Confirmed cases throughout the world

```
# Draw a line plot of cumulative cases vs. date
# Label the y-axis
ggplot(confirmed_cases_worldwide, aes(x = date, y = cum_cases)) +
  geom_line() +
  labs(y= "Cumulative confirmed cases")
```



3. China compared to the rest of the world

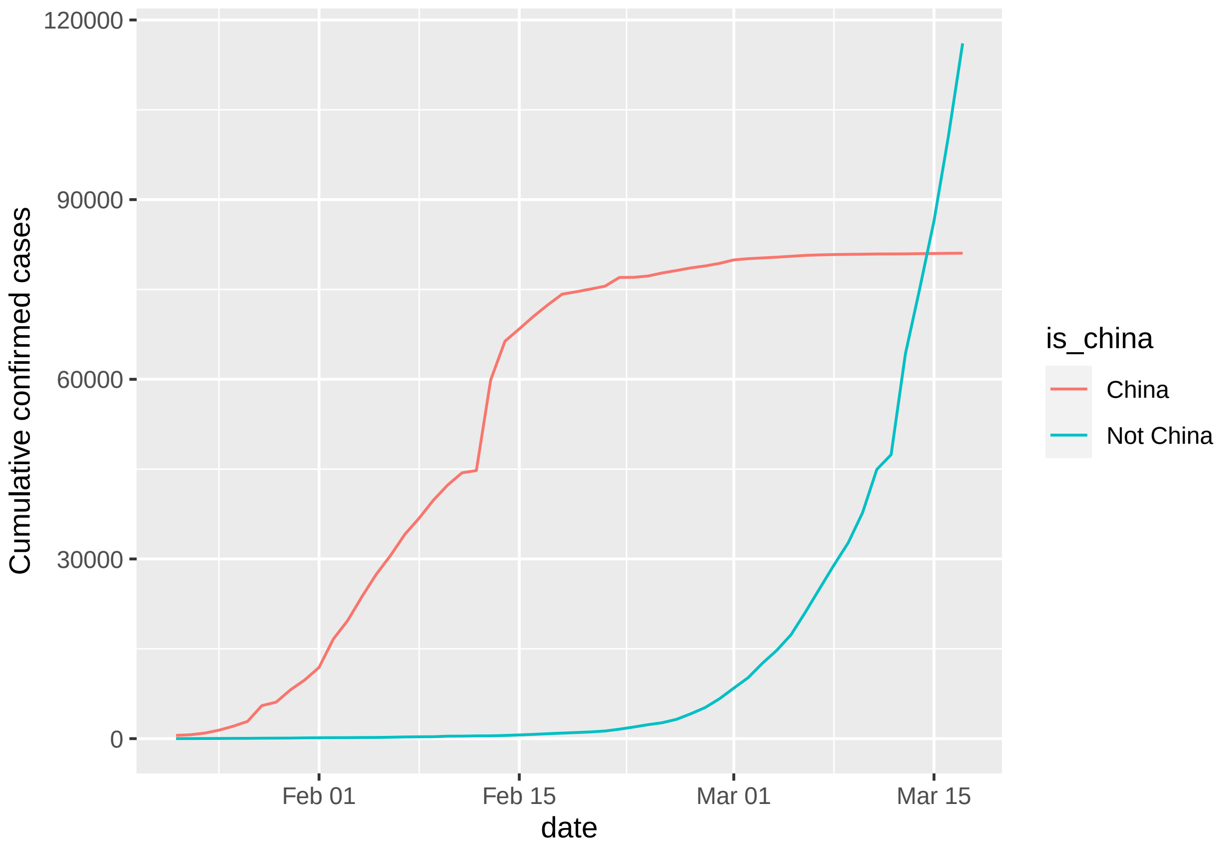
```r
# Read in datasets/confirmed_cases_china_vs_world.csv
confirmed_cases_china_vs_world <- read_csv("datasets/confirmed_cases_china_vs_world.csv")
```

```
## Parsed with column specification:
## cols(
##   is_china = col_character(),
##   date = col_date(format = ""),
##   cases = col_double(),
##   cum_cases = col_double()
## )
```

```r
# See the result
confirmed_cases_china_vs_world
```

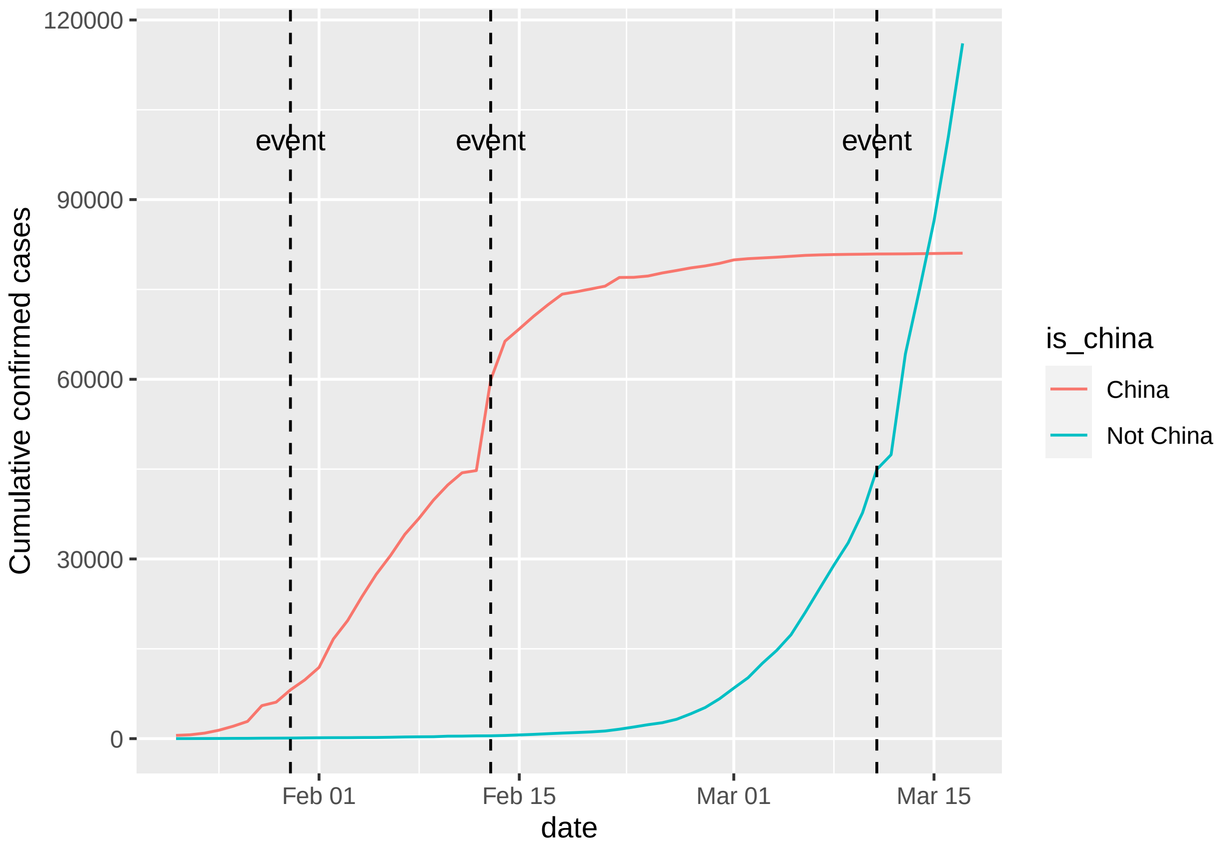
```
## # A tibble: 112 x 4
##    is_china date       cases cum_cases
##    <chr>    <date>     <dbl>     <dbl>
##  1 China    2020-01-22   548       548
##  2 China    2020-01-23    95       643
##  3 China    2020-01-24   277       920
##  4 China    2020-01-25   486      1406
##  5 China    2020-01-26   669      2075
##  6 China    2020-01-27   802      2877
##  7 China    2020-01-28  2632      5509
##  8 China    2020-01-29   578      6087
##  9 China    2020-01-30  2054      8141
## 10 China    2020-01-31  1661      9802
## # ... with 102 more rows
```

```r
# Draw a line plot of cumulative cases vs. date, colored by is_china
# Define aesthetics within the line geom
plt_cum_confirmed_cases_china_vs_world <- ggplot(confirmed_cases_china_vs_world) +
  geom_line(aes(x = date, y = cum_cases, color = is_china)) +
  ylab("Cumulative confirmed cases")

# See the plot
plt_cum_confirmed_cases_china_vs_world
```



4. Let's annotate

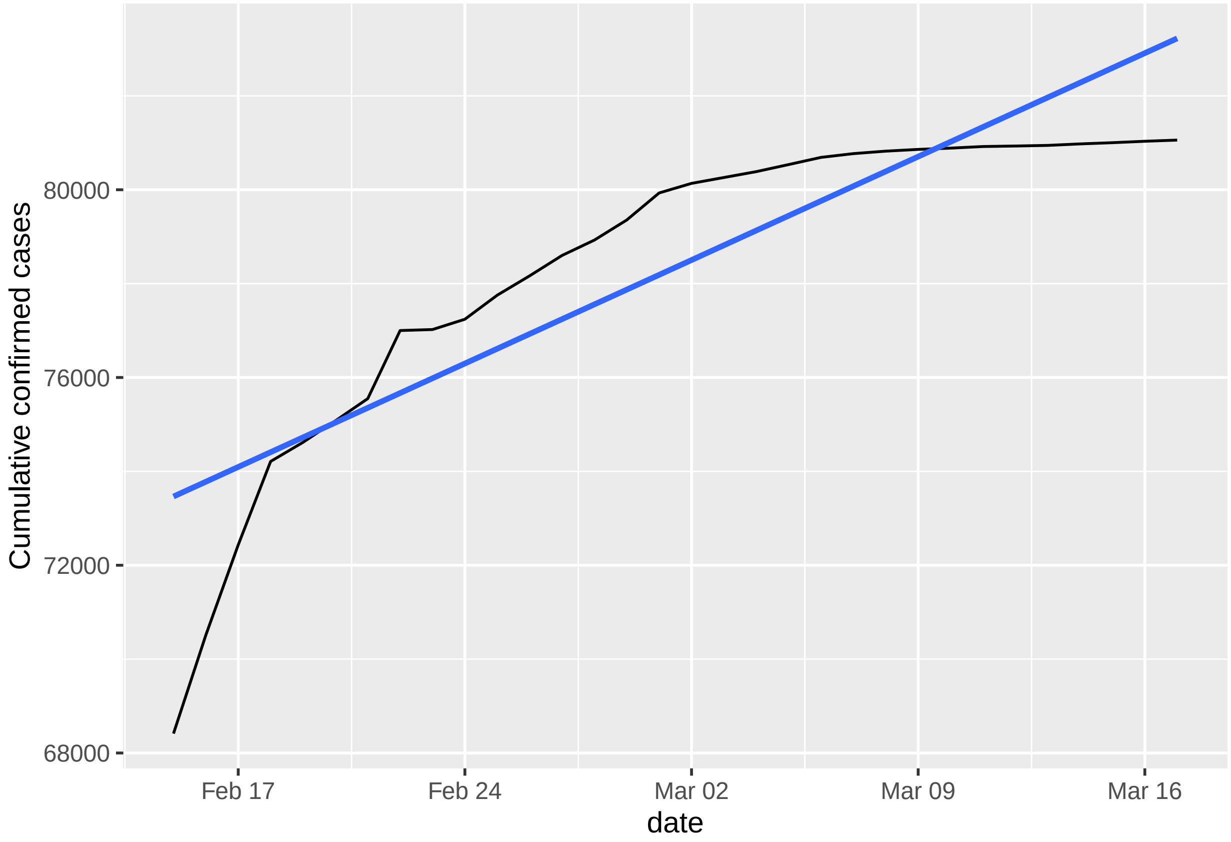
```
who_events <- tribble(
  ~ date, ~ event,
  "2020-01-30", "Global health\nemergency declared",
  "2020-03-11", "Pandemic\ndeclared",
  "2020-02-13", "China reporting\nchange"
) %>%
  mutate(date = as.Date(date))

# Using who_events, add vertical dashed lines with an xintercept at date
# and text at date, labeled by event, and at 100000 on the y-axis
plt_cum_confirmed_cases_china_vs_world +
  geom_vline(aes(xintercept = date), data = who_events, linetype = "dashed") +
  geom_text(aes(x = date, label = "event"), data = who_events, y = 1e5)
```



5. Adding a trend line to China

```
# Filter for China, from Feb 15
china_after_feb15 <- confirmed_cases_china_vs_world %>%
  filter(is_china == "China", date >= "2020-02-15")
# Using china_after_feb15, draw a line plot cum_cases vs. date
# Add a smooth trend line using linear regression, no error bars
ggplot(china_after_feb15, aes(x = date, y = cum_cases)) +
  geom_line() +
  geom_smooth(method = "lm", se = FALSE) +
  ylab("Cumulative confirmed cases")
```

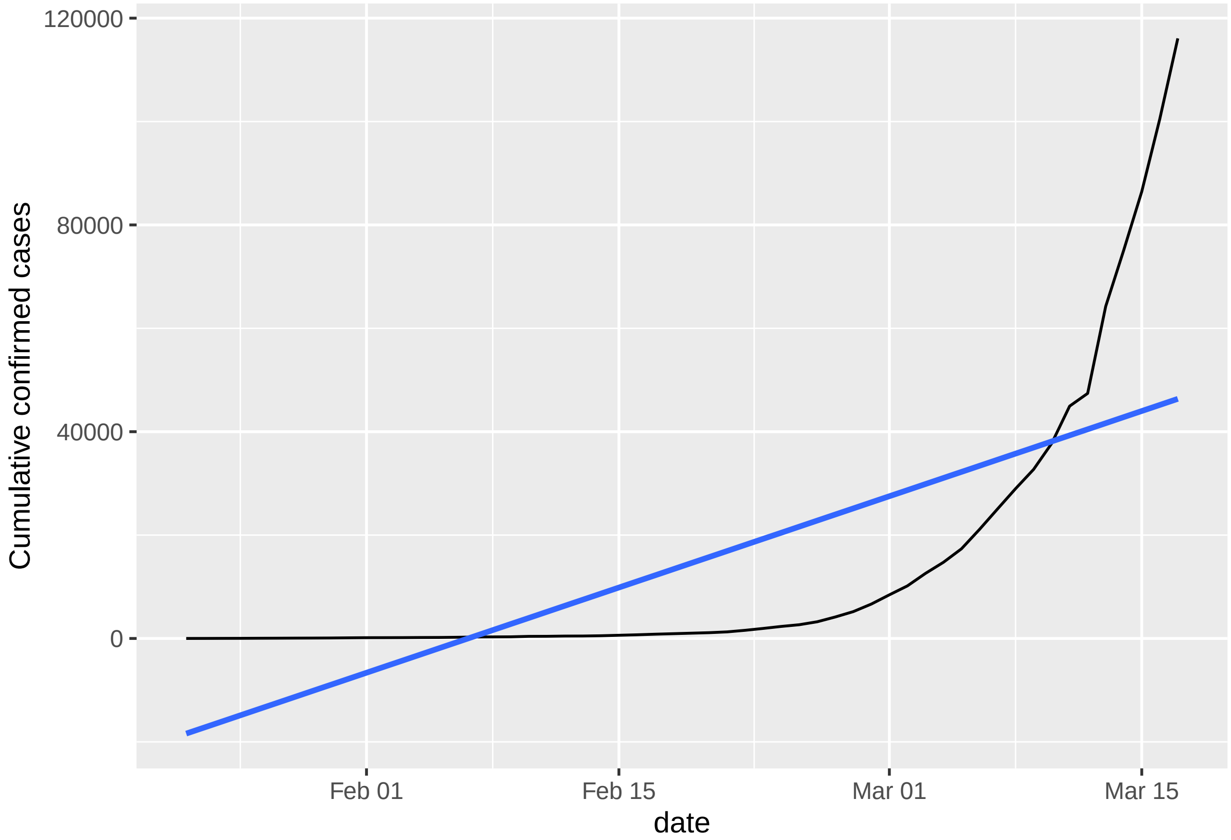
```
## ‘geom_smooth()‘ using formula ’y ~ x’
```



6. And the rest of the world

```
# Filter confirmed_cases_china_vs_world for not China
not_china <- confirmed_cases_china_vs_world %>%
  filter(is_china == "Not China")

# Using not_china, draw a line plot cum_cases vs. date
# Add a smooth trend line using linear regression, no error bars
plt_not_china_trend_lin <- ggplot(not_china, aes(x = date, y = cum_cases)) +
  geom_line() +
  geom_smooth(method = "lm", se = FALSE) +
  ylab("Cumulative confirmed cases")
# See the result
plt_not_china_trend_lin
```

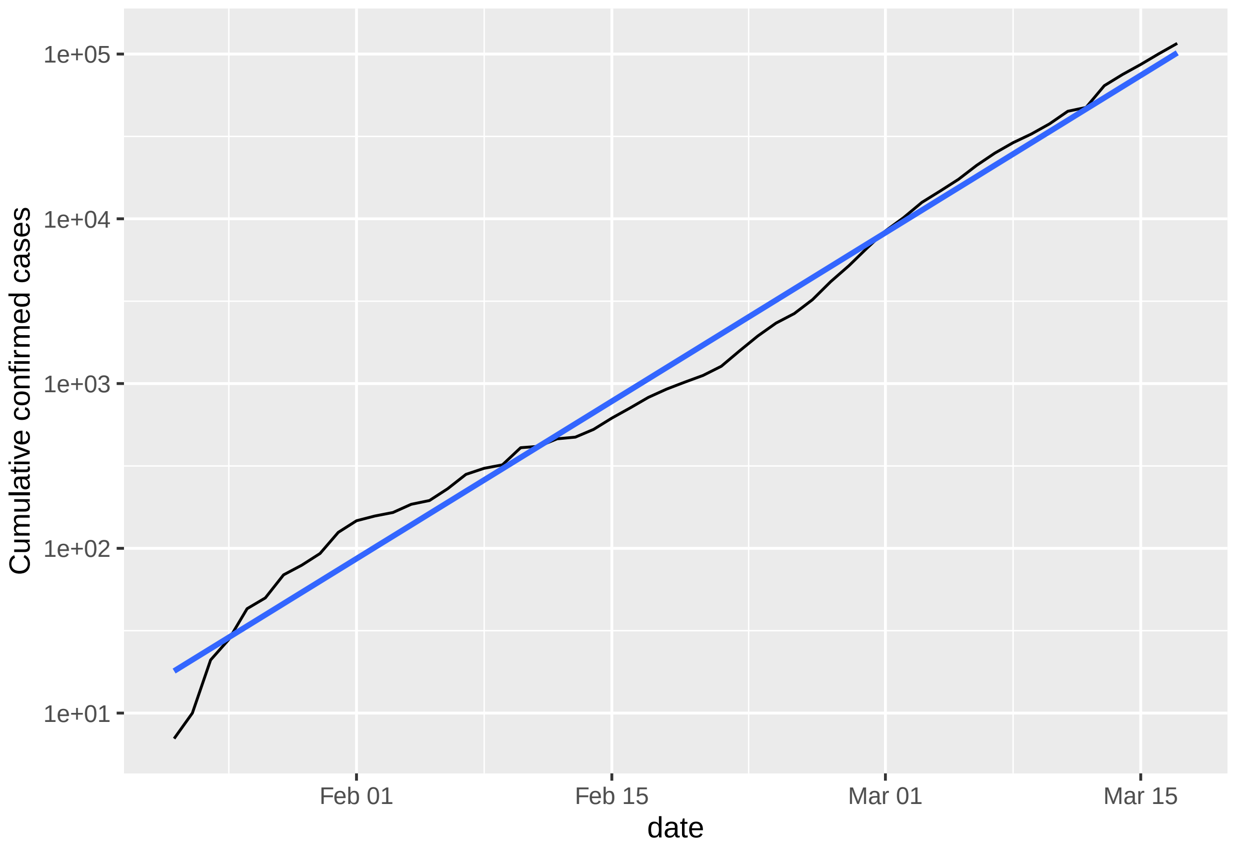
```
## ‘geom_smooth()‘ using formula ’y ~ x’
```



7. Adding a logarithmic scale

```
# Modify the plot to use a logarithmic scale on the y-axis
plt_not_china_trend_lin +
  scale_y_log10()
```

```
## ‘geom_smooth()‘ using formula 'y ~ x'
```



8. Which countries outside of China have been hit hardest

```
# Run this to get the data for each country
confirmed_cases_by_country <- read_csv("datasets/confirmed_cases_by_country.csv")
```

```
## Parsed with column specification:
## cols(
##   country = col_character(),
##   province = col_character(),
##   date = col_date(format = ""),
##   cases = col_double(),
##   cum_cases = col_double()
## )
```

```
glimpse(confirmed_cases_by_country)
```

```
## Rows: 13,272
## Columns: 5
## $ country   <chr> "Afghanistan", "Albania", "Algeria", "Andorra", "Antigua and~
## $ province  <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ date      <date> 2020-01-22, 2020-01-22, 2020-01-22, 2020-01-22, 2020-01-22,~
## $ cases     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ cum_cases <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
```

```
# Group by country, summarize to calculate total cases, find the top 7
top_countries_by_total_cases <- confirmed_cases_by_country %>%
  group_by(country) %>%
  summarize(total_cases = max(cum_cases)) %>%
  top_n(7, total_cases)

# See the result
top_countries_by_total_cases
```

```
## # A tibble: 7 x 2
##   country      total_cases
##   <chr>              <dbl>
## 1 France              7699
## 2 Germany             9257
## 3 Iran               16169
## 4 Italy              31506
## 5 Korea, South        8320
## 6 Spain              11748
## 7 US                  6421
```

9. Plotting hardest hit countries as of Mid-March 2020

```
# Read in the dataset from datasets/confirmed_cases_top7_outside_china.csv
confirmed_cases_top7_outside_china <- read_csv("datasets/confirmed_cases_top7_outside_china.csv")
```

```
## Parsed with column specification:
## cols(
##   country = col_character(),
##   date = col_date(format = ""),
##   cum_cases = col_double()
## )
```

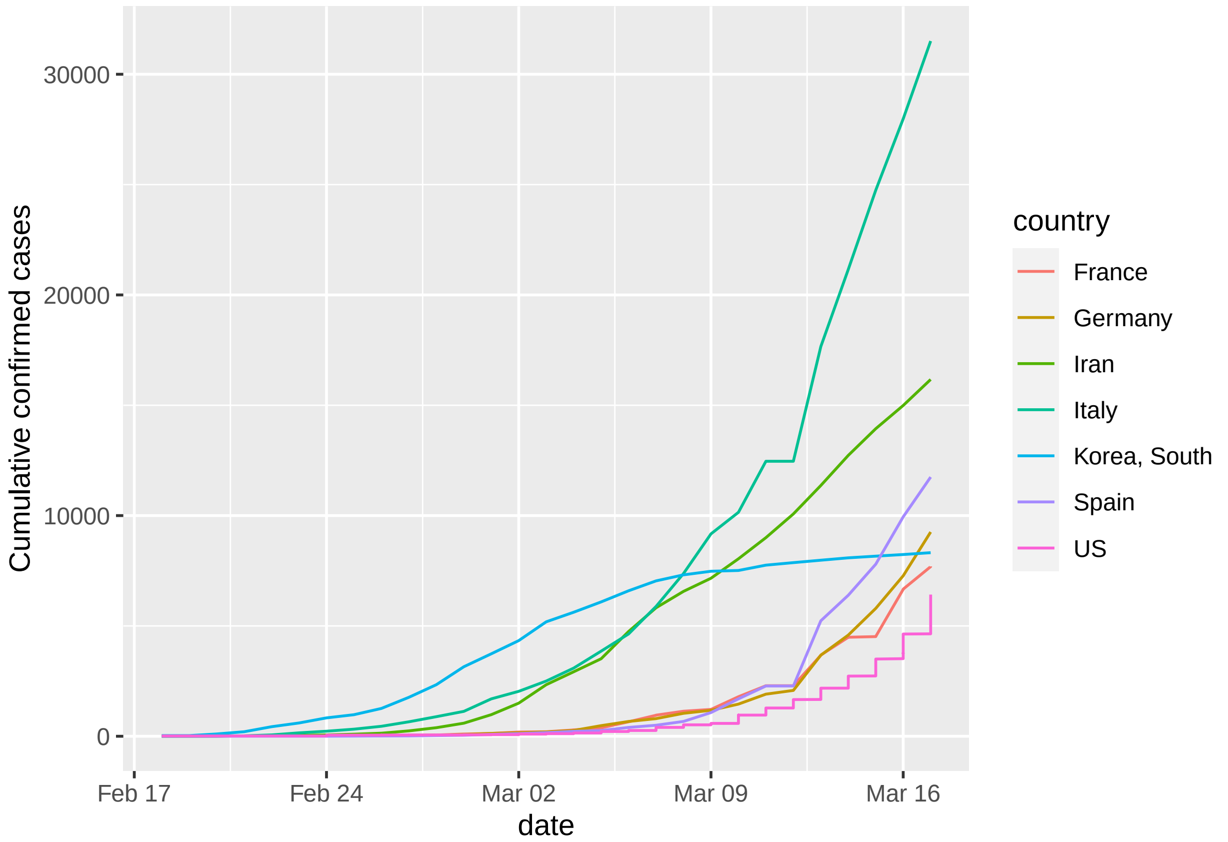
```
confirmed_cases_top7_outside_china
```

```
## # A tibble: 2,030 x 3
##    country      date       cum_cases
##    <chr>        <date>         <dbl>
##  1 Germany      2020-02-18        16
##  2 Iran         2020-02-18         0
##  3 Italy        2020-02-18         3
##  4 Korea, South 2020-02-18        31
##  5 Spain        2020-02-18         2
##  6 US           2020-02-18        13
##  7 US           2020-02-18        13
##  8 US           2020-02-18        13
##  9 US           2020-02-18        13
## 10 US           2020-02-18        13
## # ... with 2,020 more rows
```

```r
# Glimpse at the contents of confirmed_cases_top7_outside_china
glimpse(confirmed_cases_top7_outside_china)
```

```
## Rows: 2,030
## Columns: 3
## $ country   <chr> "Germany", "Iran", "Italy", "Korea, South", "Spain", "US", "~
## $ date      <date> 2020-02-18, 2020-02-18, 2020-02-18, 2020-02-18, 2020-02-18,~
## $ cum_cases <dbl> 16, 0, 3, 31, 2, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13,~
```

```r
# Using confirmed_cases_top7_outside_china, draw a line plot of
# cum_cases vs. date, colored by country
ggplot(confirmed_cases_top7_outside_china, aes(x = date, y = cum_cases, color = country)) +
  geom_line() +
  ylab("Cumulative confirmed cases")
```


Cumulative confirmed cases
Feb 17
Feb 24
Mar 02
Mar 09
Mar 16
date
country
France
Germany
Iran
Italy
Korea, South
Spain
US