

# Covid19\_VS\_VisualAnalytics\_Timeseries\_Data

V.Srinivas

06/05/2020

```
# Load the required packages
```

```
library(ggplot2)
```

```
library(dplyr)
```

```
##
```

```
## 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
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v tibble 2.1.3      v purrr 0.3.3
```

```
## v tidyr 1.0.2      v stringr 1.4.0
```

```
## v readr 1.3.1      v forcats 0.5.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag() masks stats::lag()
```

```
library(caret)
```

```
## Loading required package: lattice
```

```
##
```

```
## Attaching package: 'caret'
```

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

```
##
```

```
## lift
```

```
library(scales)
```

```
##  
## Attaching package: 'scales'  
  
## The following object is masked from 'package:purrr':  
##  
##     discard  
  
## The following object is masked from 'package:readr':  
##  
##     col_factor
```

```
# Read datasets/confirmed_cases_worldwide.csv into confirmed_cases_worldwide  
covid_ds <- data.table::fread('https://raw.githubusercontent.com/RamiKrispin/coronavirus-csv/master/corona  
mutate (date=as.Date(date)) %>%  
rename(country=Country.Region) %>%  
select(-Lat, -Long)
```

```
# View info about the dataset  
glimpse(covid_ds)
```

```
## Observations: 79,968  
## Variables: 4  
## $ country <chr> "Afghanistan", "Afghanistan", "Afghanistan", "Afghanistan",...  
## $ date <date> 2020-01-22, 2020-01-23, 2020-01-24, 2020-01-25, 2020-01-26...  
## $ cases <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...  
## $ type <chr> "confirmed", "confirmed", "confirmed", "confirmed", "confir..."
```

```
str(covid_ds)
```

```
## 'data.frame': 79968 obs. of 4 variables:  
## $ country: chr "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...  
## $ date : Date, format: "2020-01-22" "2020-01-23" ...  
## $ cases : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ type : chr "confirmed" "confirmed" "confirmed" "confirmed" ...
```

```
tail(covid_ds,20)
```

```
##      country      date cases      type  
## 79949   China 2020-04-13     1 recovered  
## 79950   China 2020-04-14     3 recovered  
## 79951   China 2020-04-15     2 recovered  
## 79952   China 2020-04-16     0 recovered  
## 79953   China 2020-04-17     2 recovered  
## 79954   China 2020-04-18     0 recovered  
## 79955   China 2020-04-19     1 recovered  
## 79956   China 2020-04-20     2 recovered  
## 79957   China 2020-04-21     2 recovered  
## 79958   China 2020-04-22     1 recovered
```

```
## 79959    China 2020-04-23    4 recovered
## 79960    China 2020-04-24    0 recovered
## 79961    China 2020-04-25    1 recovered
## 79962    China 2020-04-26    1 recovered
## 79963    China 2020-04-27    3 recovered
## 79964    China 2020-04-28    1 recovered
## 79965    China 2020-04-29    1 recovered
## 79966    China 2020-04-30    1 recovered
## 79967    China 2020-05-01    1 recovered
## 79968    China 2020-05-02    0 recovered
```

```
dim(covid_ds)
```

```
## [1] 79968    4
```

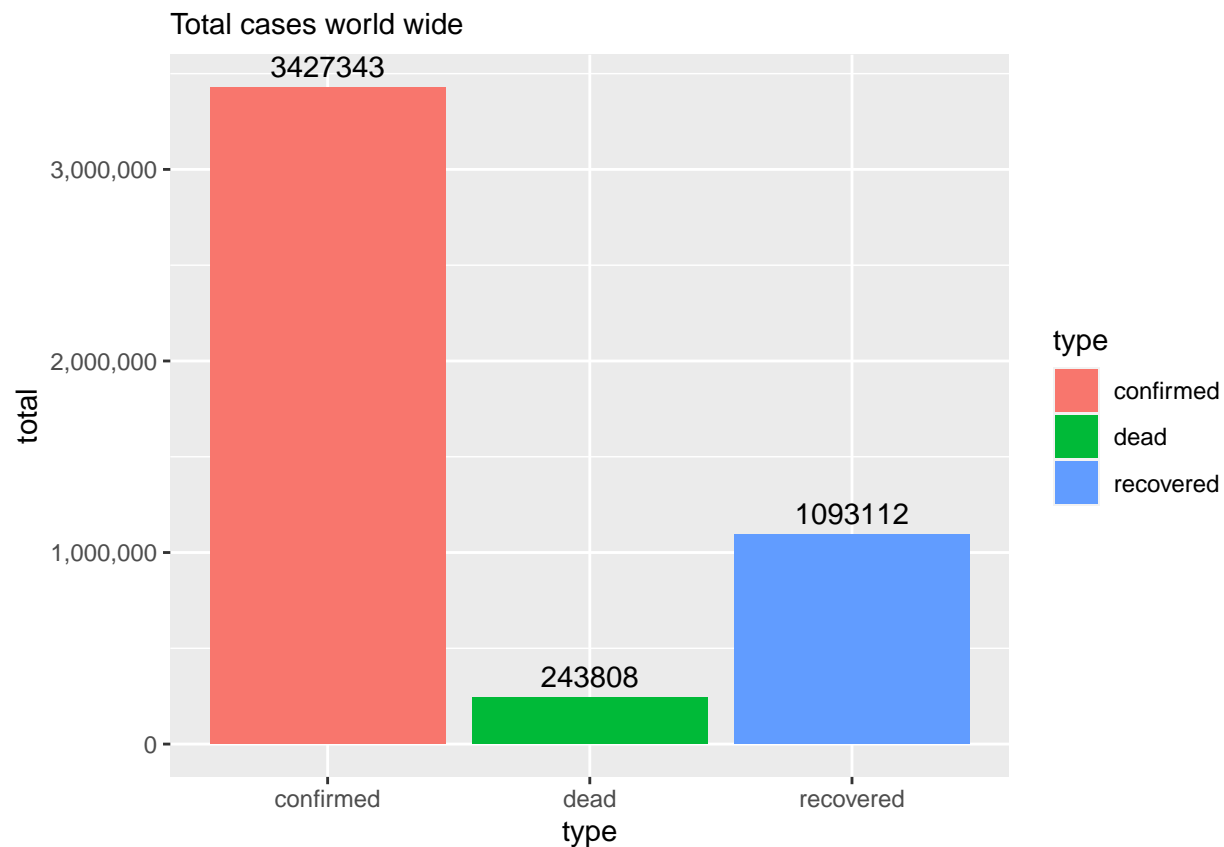
```
covid_ds$type[covid_ds$type == 'death'] <- 'dead'
```

```
# Convert negative cases to positive assuming it is typo error
#covid_ds$cases = abs(covid_ds$cases)
#covid_ds %>% filter(cases < 0) %>% select(country, cases, type)
```

```
# compute daily cases for all countries
daily_cases <- covid_ds %>%
  group_by(date, country, type) %>%
  summarize (daily_cases = sum(cases)) %>%
  ungroup()
```

```
# Totals so far worldwide
totals <- covid_ds %>%
  group_by(type) %>%
  summarize(total=sum(cases))

totals %>%
  ggplot(aes(type, total, fill=type)) +
  geom_col() +
  scale_y_continuous(labels=comma)+
  geom_text(aes(label=total), vjust=-.5) +
  labs(subtitle="Total cases world wide")
```



```
# Total daily cases across the globe
tot_daily_all_countries <- covid_ds %>%
  group_by(date, type) %>%
  mutate (total_daily_cases = sum(cases)) %>%
  arrange(desc(total_daily_cases)) %>%
  select(date, type, total_daily_cases)

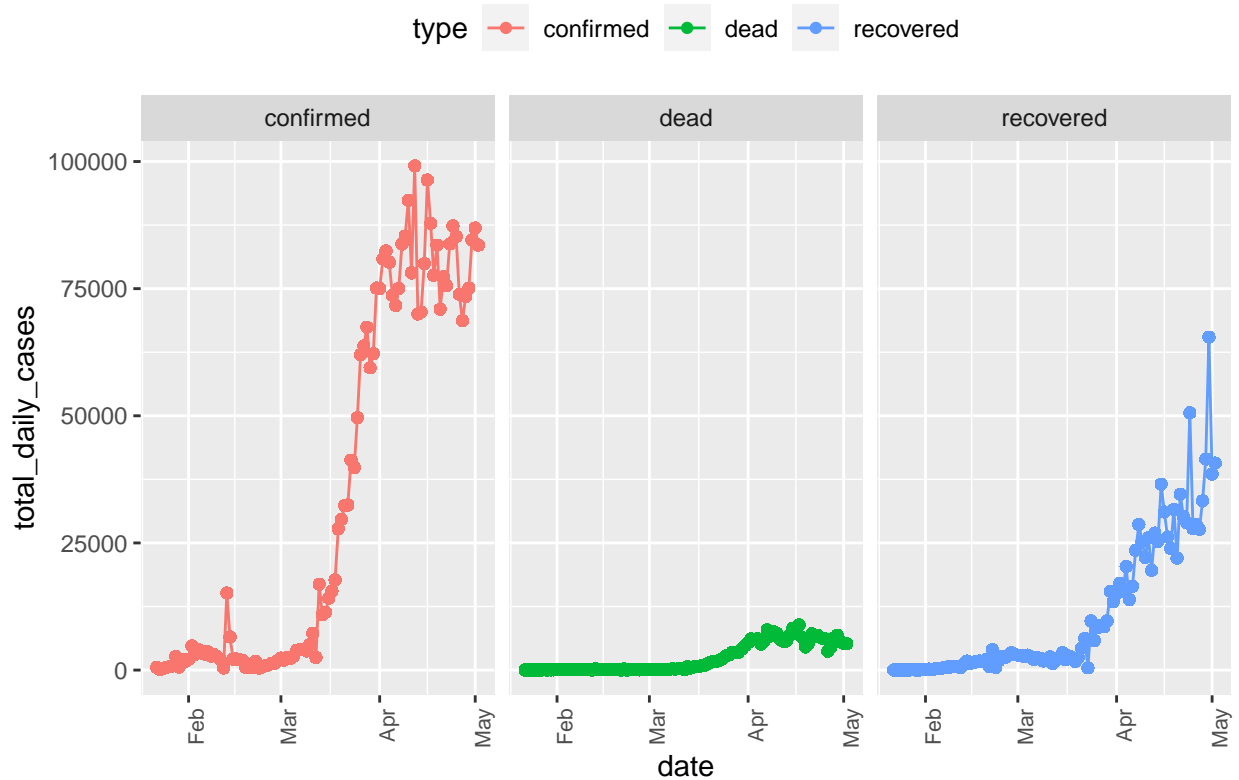
tot_daily_all_countries %>%
  filter(total_daily_cases == max(total_daily_cases)) %>%
  head(1)
```

```
## # A tibble: 1 x 3
## # Groups:   date, type [1]
##   date      type      total_daily_cases
##   <date>    <chr>          <int>
## 1 2020-04-12 confirmed          99139
```

```
tot_daily_all_countries %>% ggplot(aes(date, total_daily_cases, col=type)) +
  geom_point() +
  geom_line() +
  facet_wrap(~type) +
  theme(axis.text.x = element_text(
    angle = 90,
    size = 8,
    hjust = 1
```

```
),
legend.position = "top") +
labs(subtitle=" Total daily cases across all nations")
```

Total daily cases across all nations



```
# Confirmed cumulative cases country-wise
confirmed <- daily_cases %>%
  filter(type=='confirmed') %>%
  select(-type) %>%
  group_by(country) %>%
  mutate(cum_c_cases= cumsum(daily_cases)) %>%
  select(-daily_cases) %>%
  ungroup()
```

```
# Recovered cumulative cases
recovered <- daily_cases %>%
  filter(type=='recovered') %>%
  select(-type) %>%
  group_by(country) %>%
  mutate(cum_r_cases= cumsum(daily_cases)) %>%
  select(-daily_cases) %>%
  ungroup()
```

```
# dead cumulative cases
dead <- daily_cases %>%
```

```

filter(type=='dead') %>%
select(-type) %>%
group_by(country) %>%
mutate(cum_d_cases= cumsum(daily_cases)) %>%
select(-daily_cases) %>%
ungroup()

```

*# Combine cumulative cases of all 3 categories*

```

all_cases <- confirmed %>%
  inner_join(recovered, by=c('date', 'country')) %>%
  inner_join(dead, by=c('date', 'country')) %>%
  mutate(country=as.factor(country))

```

*# All cumulative cases, all countries on all dates with all 3 categories - Top 50*

```

all_cases %>% arrange(desc(date,cum_c_cases)) %>% head(50)

```

```

## # A tibble: 50 x 5
##   date      country      cum_c_cases cum_r_cases cum_d_cases
##   <date>    <fct>          <int>      <int>      <int>
## 1 2020-05-02 Afghanistan    2469        331         72
## 2 2020-05-02 Albania         789        519         31
## 3 2020-05-02 Algeria       4295       1872        459
## 4 2020-05-02 Andorra        747        472         44
## 5 2020-05-02 Angola          35         11          2
## 6 2020-05-02 Antigua and Barbuda    25         15          3
## 7 2020-05-02 Argentina     4681       1320        237
## 8 2020-05-02 Armenia       2273       1010         33
## 9 2020-05-02 Australia     6799       5814         94
## 10 2020-05-02 Austria      15558      13180        596
## # ... with 40 more rows

```

*# All cumulative cases of all countries on all dates*

```

countries_cases <- all_cases %>%
  rename(confirmed=cum_c_cases, recovered=cum_r_cases, dead=cum_d_cases) %>%
  gather(type, cum_cases, confirmed:dead) %>%
  mutate(type=as.factor(type)) %>% arrange(desc(date))

```

*# Select, major hit nations (top 30)*

```

top_20_countries <- covid_ds %>% group_by(country) %>% summarize(total_cases=sum(cases)) %>% arrange(desc(total_cases))
select_countries <- top_20_countries$country

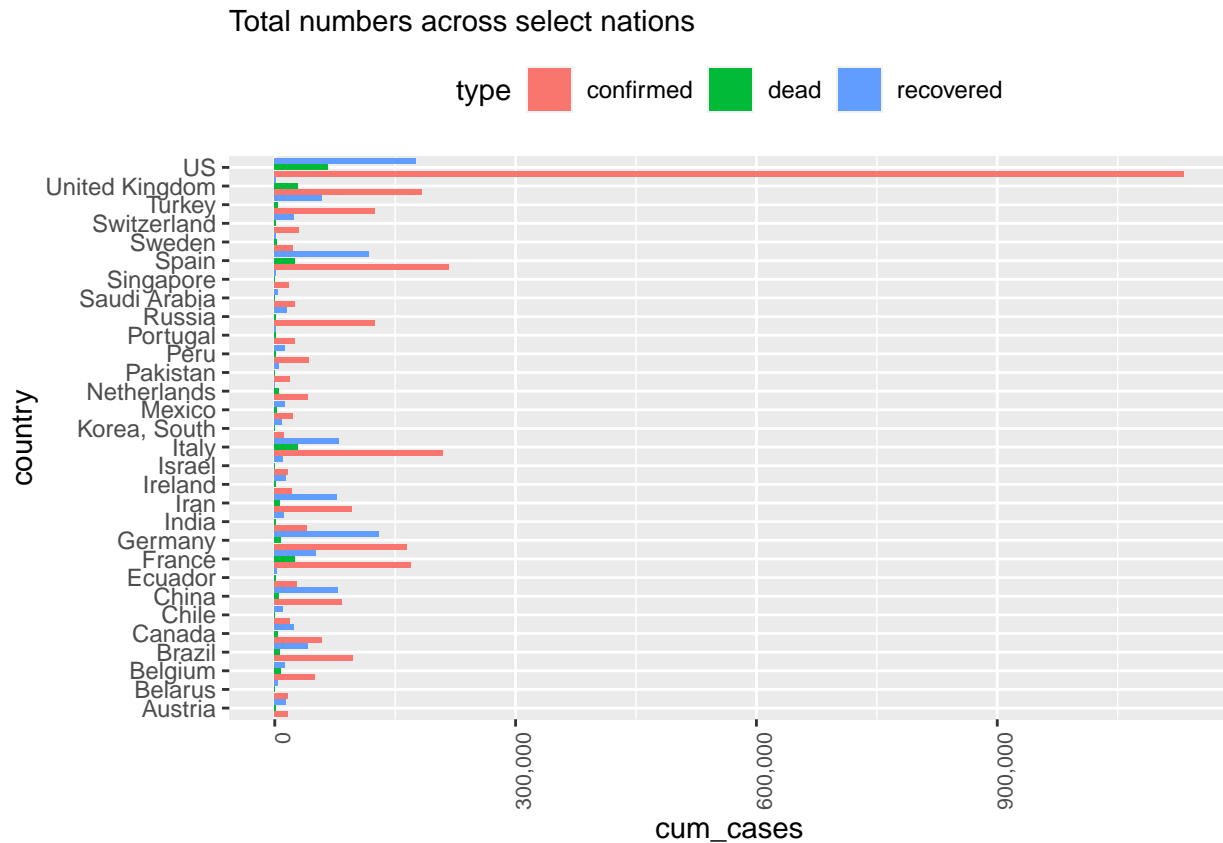
```

```

countries_cases %>%
  filter(country %in% select_countries) %>%
  ggplot(aes(country, cum_cases, fill=type)) +
  geom_col( position = position_dodge(1)) +
  scale_y_continuous(labels = comma)+
  coord_flip() +
  theme(axis.text.x = element_text(
    angle = 90,
    size = 8,
    hjust = 1
  )),

```

```
legend.position = "top") +
labs(subtitle="Total numbers across select nations")
```



```
# Cumulative cases by date and type
countries_cases_all <- countries_cases %>%
  group_by(date, type) %>%
  mutate(tot_cum_cases = sum(cum_cases)) %>%
  arrange(desc(tot_cum_cases))
```

```
# Cumulative cases by date and type - Top 50
cum_cases_all <- countries_cases_all %>%
  group_by(date, type) %>%
  top_n(1, wt = cum_cases) %>%
  ungroup()
```

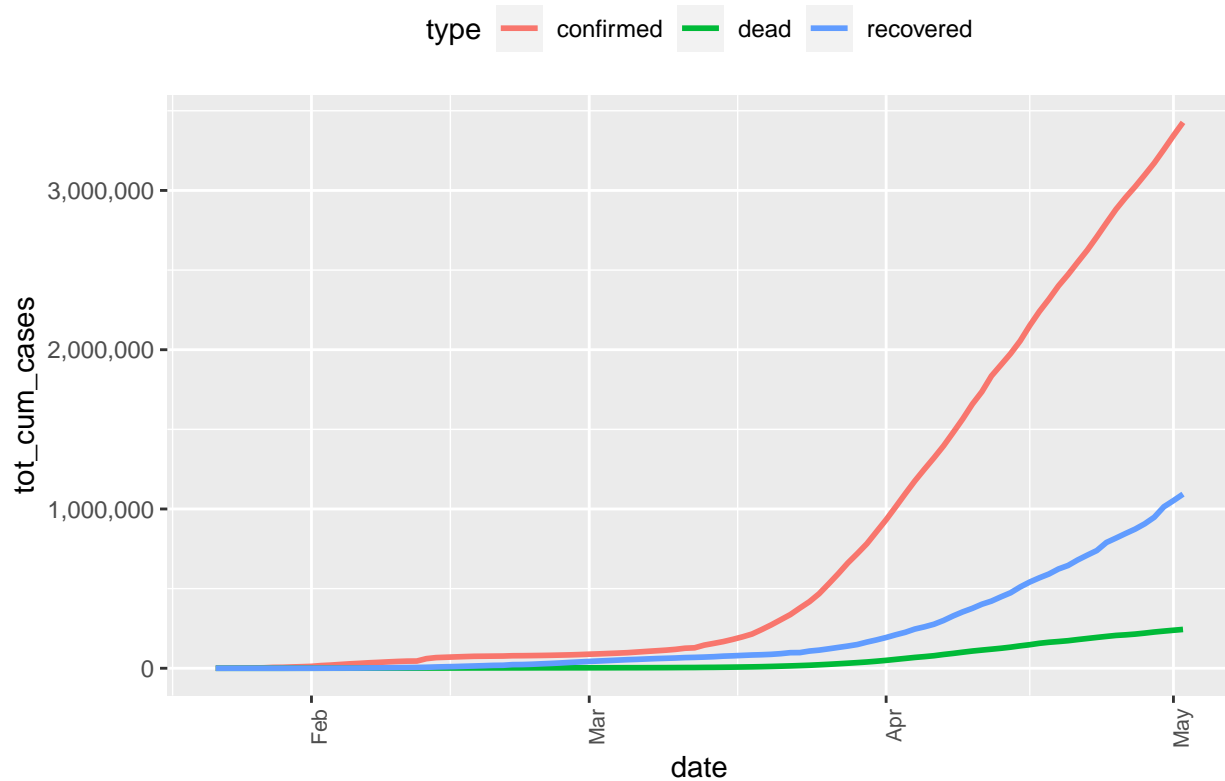
```
cum_cases_all %>%
  select(date, type, tot_cum_cases) %>%
  arrange(desc(date, type)) %>% head(50)
```

```
## # A tibble: 50 x 3
##   date      type      tot_cum_cases
##   <date>    <fct>         <int>
## 1 2020-05-02 confirmed     3427343
## 2 2020-05-02 recovered     1093112
```

```
## 3 2020-05-02 dead      243808
## 4 2020-05-01 confirmed 3343777
## 5 2020-05-01 recovered 1052415
## 6 2020-05-01 dead     238619
## 7 2020-04-30 confirmed 3256853
## 8 2020-04-30 recovered 1013871
## 9 2020-04-30 dead     233357
## 10 2020-04-29 confirmed 3172287
## # ... with 40 more rows
```

```
cum_cases_all %>%
  ggplot(aes(date, tot_cum_cases, col=type)) +
  geom_line(size=1) +
  scale_y_continuous(label=comma) +
  theme(axis.text.x = element_text(
    angle = 90,
    size = 8,
    hjust = 1
  ),
  legend.position = "top") +
  labs(subtitle=" Cumulative number progression across all nations")
```

Cumulative number progression across all nations

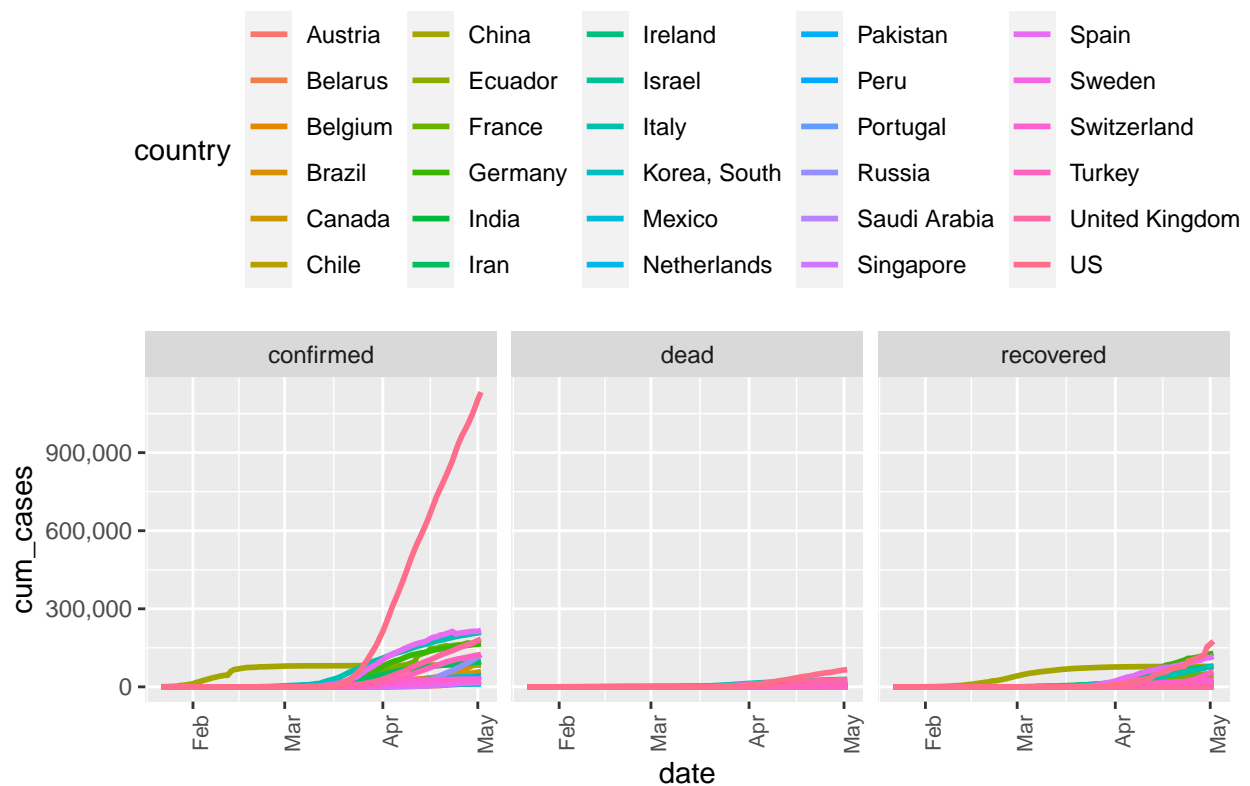


```
# Cases in select, major hot nations
select_cases <- countries_cases %>%
  filter(country %in% select_countries)
```



```
select_cases %>%
  ggplot(aes(date, cum_cases, col=country)) +
  geom_line(size=1) +
  scale_y_continuous(label=comma) +
  facet_grid(~type) +
  theme(axis.text.x = element_text(
    angle = 90,
    size = 8,
    hjust = 1
  ),
  legend.position = "top") +
  labs(subtitle=" Cumulative number progression across major hit nations")
```

Cumulative number progression across major hit nations



```
# Compare China and USA
china_vs_usa <- countries_cases %>%
  filter(country %in% c('China', 'US'))

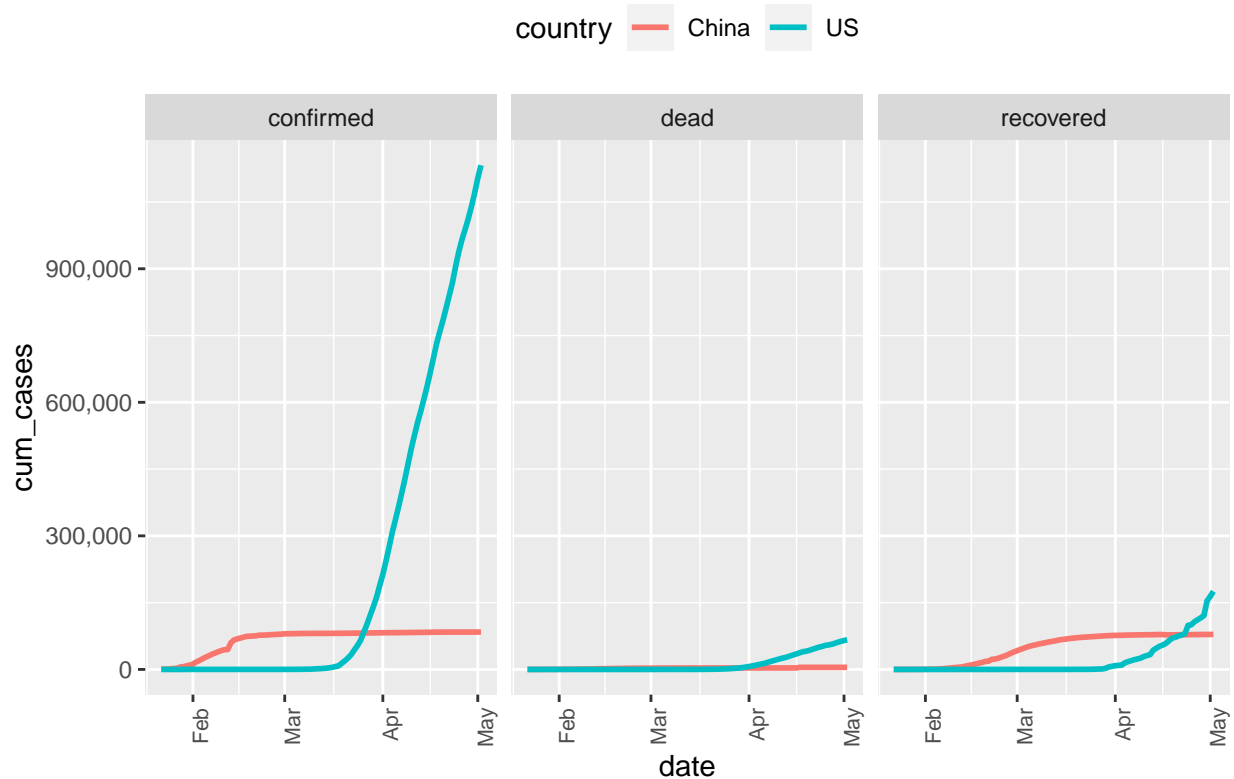
china_vs_usa %>%
  ggplot(aes(date, cum_cases, col=country)) +
  geom_line(size=1) +
  scale_y_continuous(label=comma) +
  facet_grid(~type) +
  theme(axis.text.x = element_text(
    angle = 90,
    size = 8,
```

```

    hjust = 1
  ),
  legend.position = "top") +
  labs(subtitle=" China vs USA cumulative progression")

```

China vs USA cumulative progression



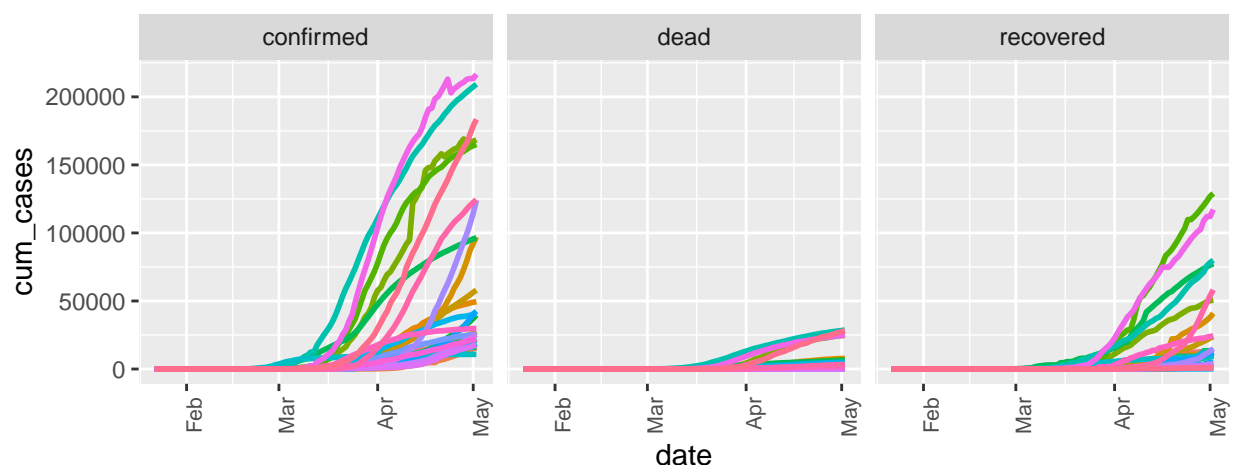
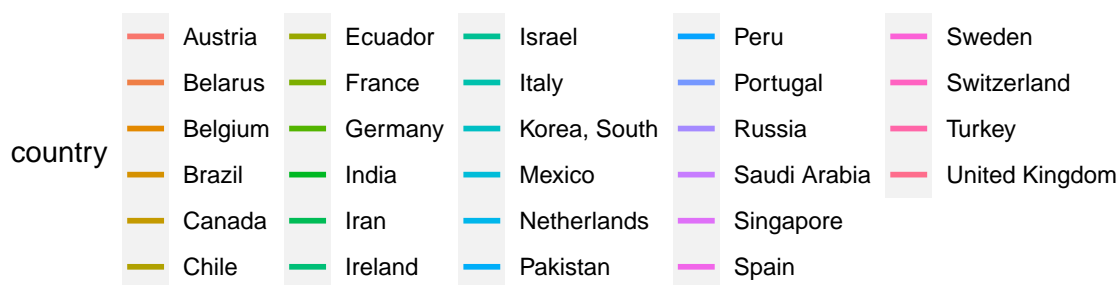
```

# non-China and non_USA cases (non-extreme case countries)
non_china_usa_cases <- countries_cases %>%
  filter(country %in% select_countries & country != 'China' & country != 'US')

non_china_usa_cases %>%
  ggplot(aes(date, cum_cases, col=country)) +
  geom_line(size=1) +
  facet_grid(~type) +
  theme(axis.text.x = element_text(
    angle = 90,
    size = 8,
    hjust = 1
  ),
  ),
  legend.position = "top") +
  labs(subtitle="Non_china, Non_USA cumulative number progression")

```

## Non\_china, Non\_USA cumulative number progression



```
# Cases in india
india_cases <- countries_cases %>%
  filter(country %in% c('India')) %>%
  select(-country)

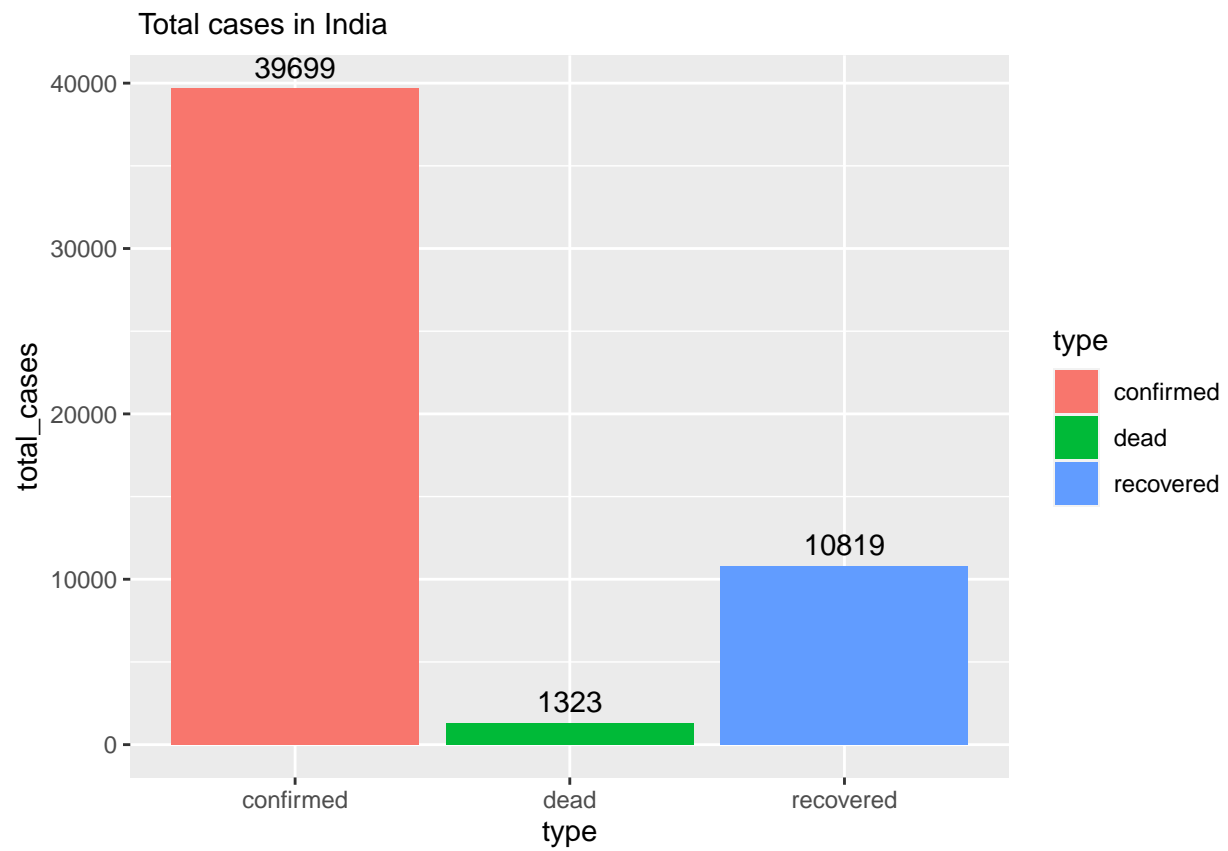
india_cases %>% arrange(desc(date)) %>% head(50)
```

```
## # A tibble: 50 x 3
##   date       type      cum_cases
##   <date>    <fct>      <int>
## 1 2020-05-02 confirmed    39699
## 2 2020-05-02 recovered   10819
## 3 2020-05-02 dead        1323
## 4 2020-05-01 confirmed    37257
## 5 2020-05-01 recovered   10007
## 6 2020-05-01 dead        1223
## 7 2020-04-30 confirmed    34863
## 8 2020-04-30 recovered    9068
## 9 2020-04-30 dead         1154
## 10 2020-04-29 confirmed    33062
## # ... with 40 more rows
```

```
india_cases %>% group_by(type) %>% top_n(1) %>% rename(total_cases = cum_cases) %>%
  ggplot(aes(type, total_cases, fill=type)) +
  geom_col() +
```

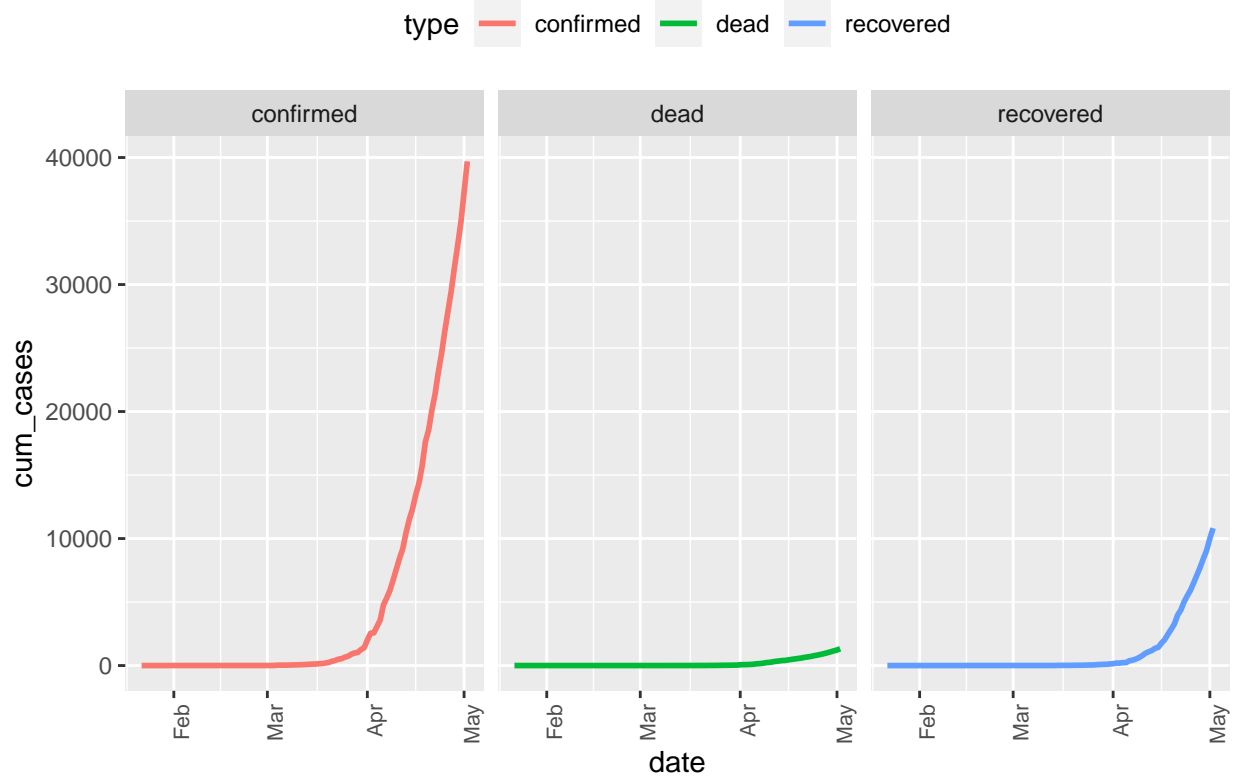
```
geom_text(aes(label = total_cases), vjust = -0.5) +
labs(subtitle=" Total cases in India")
```

```
## Selecting by cum_cases
```



```
india_cases %>%
  ggplot(aes(date, cum_cases, col=type)) +
  geom_line(size=1) +
  facet_grid(~ type) +
  theme(axis.text.x = element_text(
    angle = 90,
    size = 8,
    hjust = 1
  )),
  legend.position = "top") +
labs(subtitle=" Cumulative case progression in India")
```

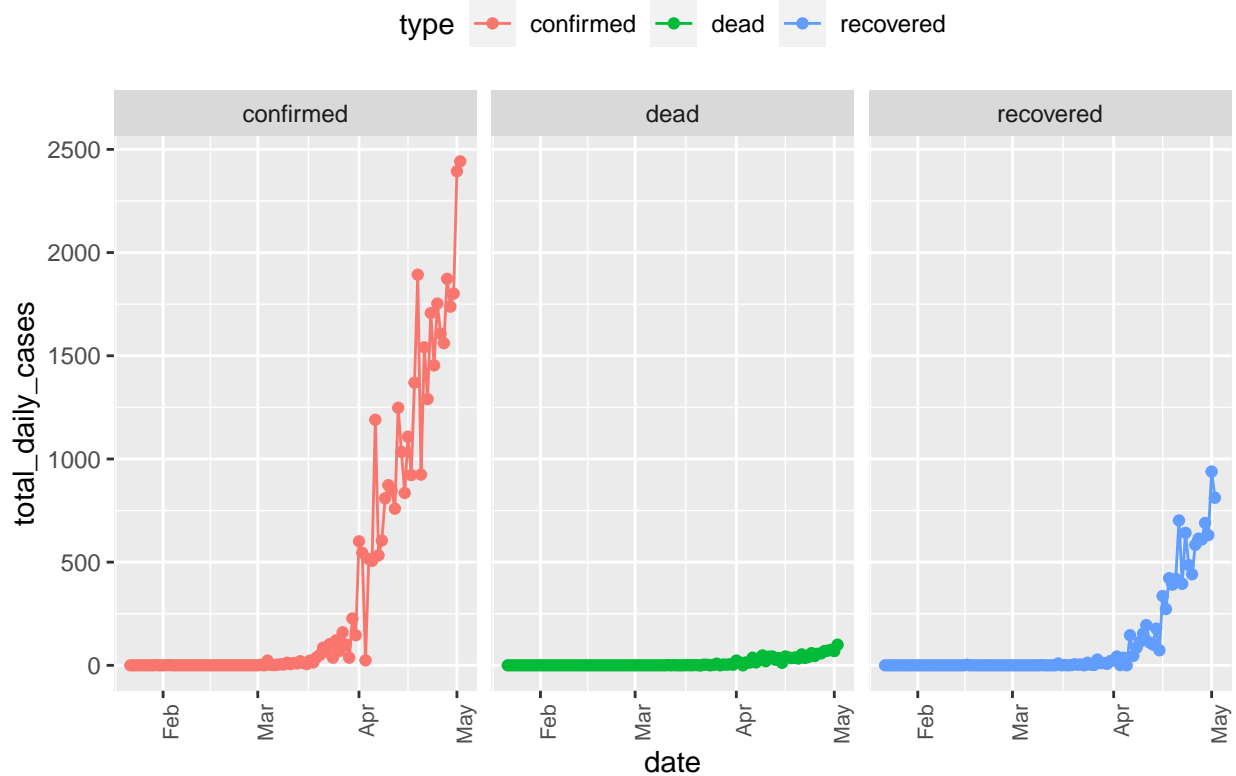
## Cumulative case progression in India



```
# Total daily cases in India
tot_d_india_cases <- covid_ds %>%
  filter(country=='India') %>%
  group_by(date, type) %>% mutate (total_daily_cases = sum(cases)) %>%
  arrange(desc(total_daily_cases))%>%
  select(date, type, total_daily_cases)

tot_d_india_cases %>% ggplot(aes(date,total_daily_cases, col=type)) +
  geom_point()+
  geom_line() +
  facet_wrap(~type)+
  theme(axis.text.x = element_text(
    angle = 90,
    size = 8,
    hjust = 1
  )),
  legend.position = "top") +
  labs(subtitle=" Total daily cases in India")
```

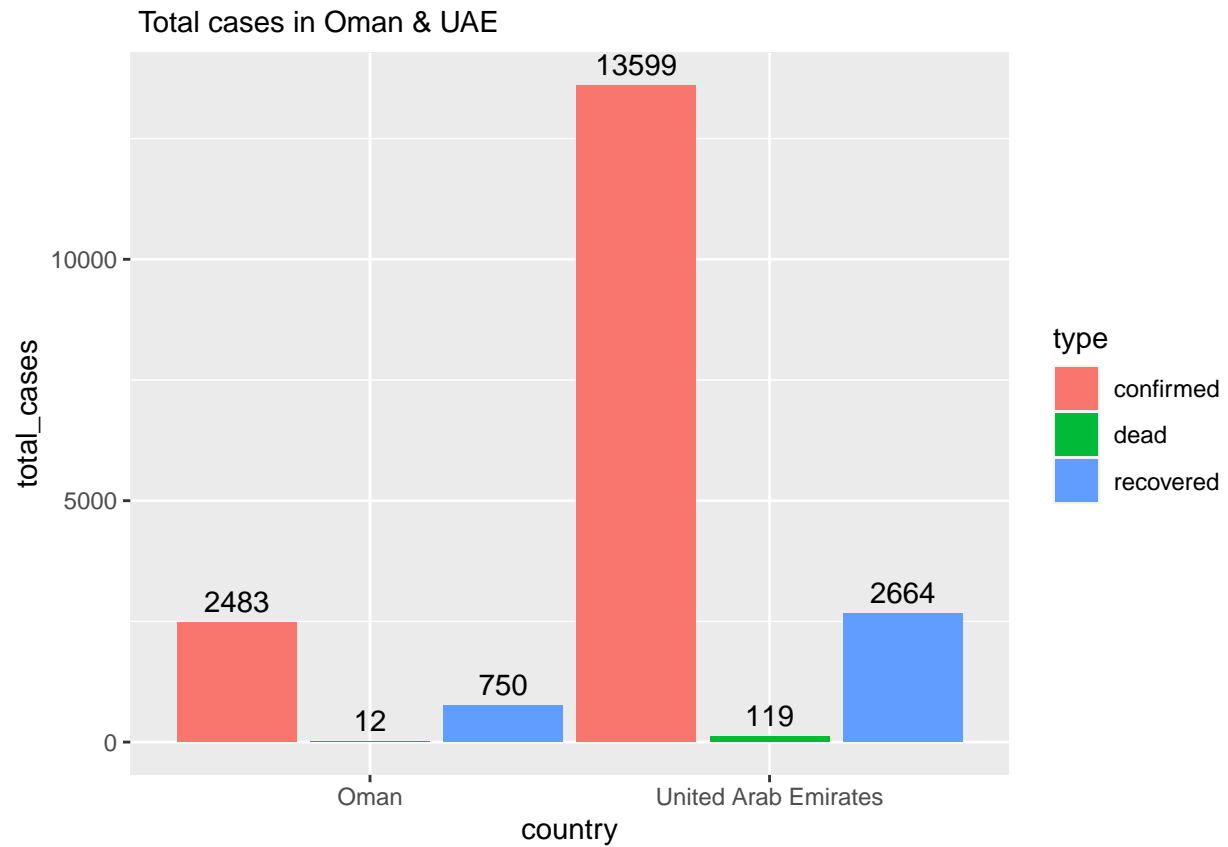
## Total daily cases in India



```
# Cases in Oman & UAE
oman_vs_uae <- countries_cases %>%
  filter(country %in% c('Oman', 'United Arab Emirates'))

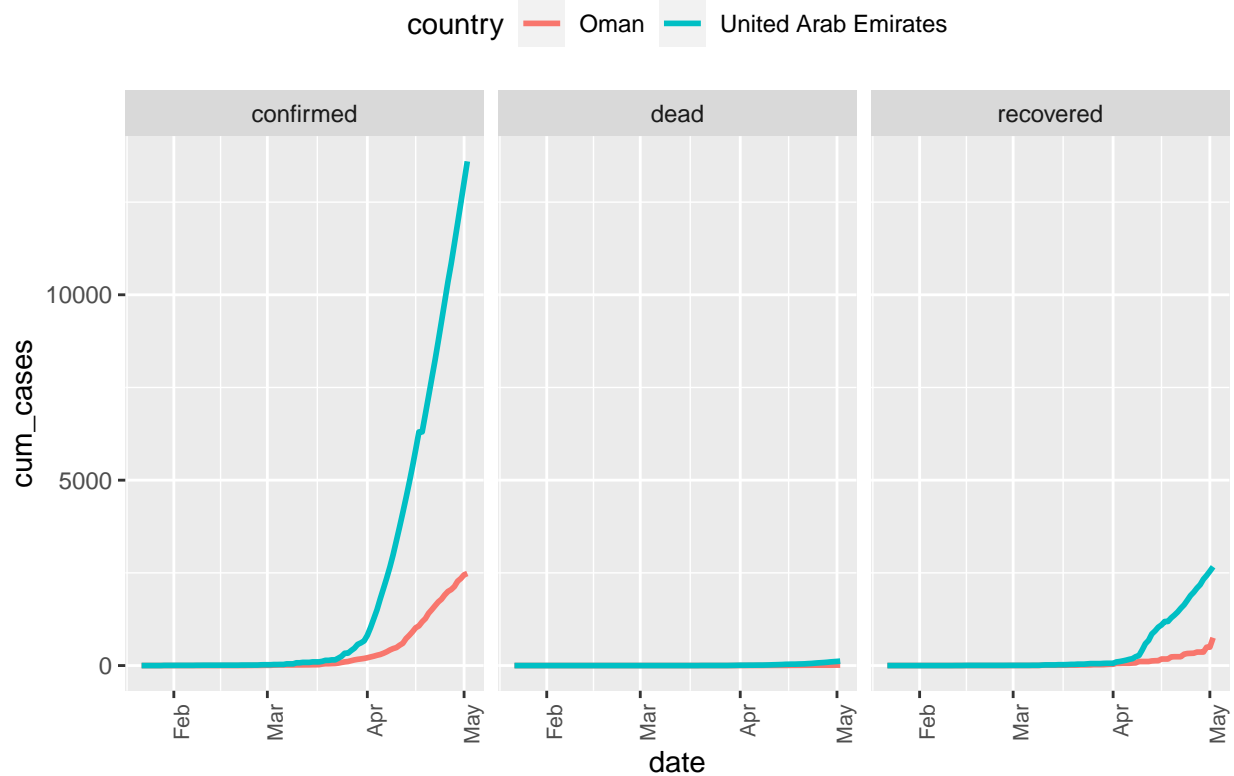
oman_vs_uae %>% group_by(country,type) %>% top_n(1) %>% rename(total_cases = cum_cases) %>%
  ggplot(aes(country, total_cases, fill=type)) +
  geom_col(position = position_dodge(1)) +
  geom_text(aes(label = total_cases),position = position_dodge(1), vjust = -0.5) +
  labs(subtitle=" Total cases in Oman & UAE")
```

```
## Selecting by cum_cases
```



```
oman_vs_uae %>% ggplot(aes(date, cum_cases, col=country)) +
  geom_line(size=1) +
  facet_grid(~type) +
  theme(axis.text.x = element_text(
    angle = 90,
    size = 8,
    hjust = 1
  ),
  legend.position = "top") +
  labs(subtitle=" Oman vs UAE cumulative progression")
```

## Oman vs UAE cumulative progression

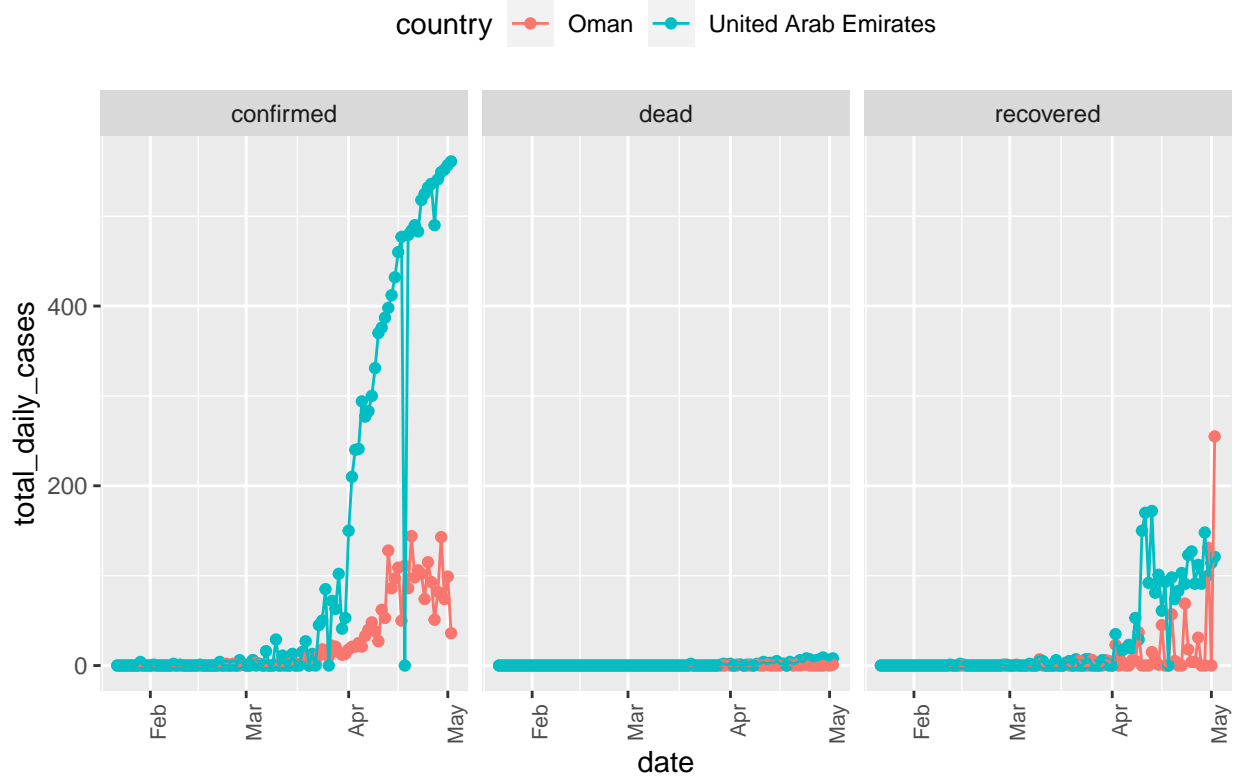


```
# Total daily cases across the globe
oman_vs_uae_daily <- covid_ds %>%
  filter(country=='Oman' | country == "United Arab Emirates") %>%
  group_by(date,country, type) %>% mutate (total_daily_cases = sum(cases)) %>%
  arrange(desc(total_daily_cases))%>%
  select(date, country, type, total_daily_cases)
```

```
oman_vs_uae_daily %>% ggplot(aes(date,total_daily_cases, col=country)) +
  geom_point()+
  geom_line() +
  facet_wrap(~type)+
  theme(axis.text.x = element_text(
    angle = 90,
    size = 8,
    hjust = 1
  ),
  legend.position = "top") +
  labs(subtitle=" Total daily cases in Oman & UAE")
```



## Total daily cases in Oman & UAE

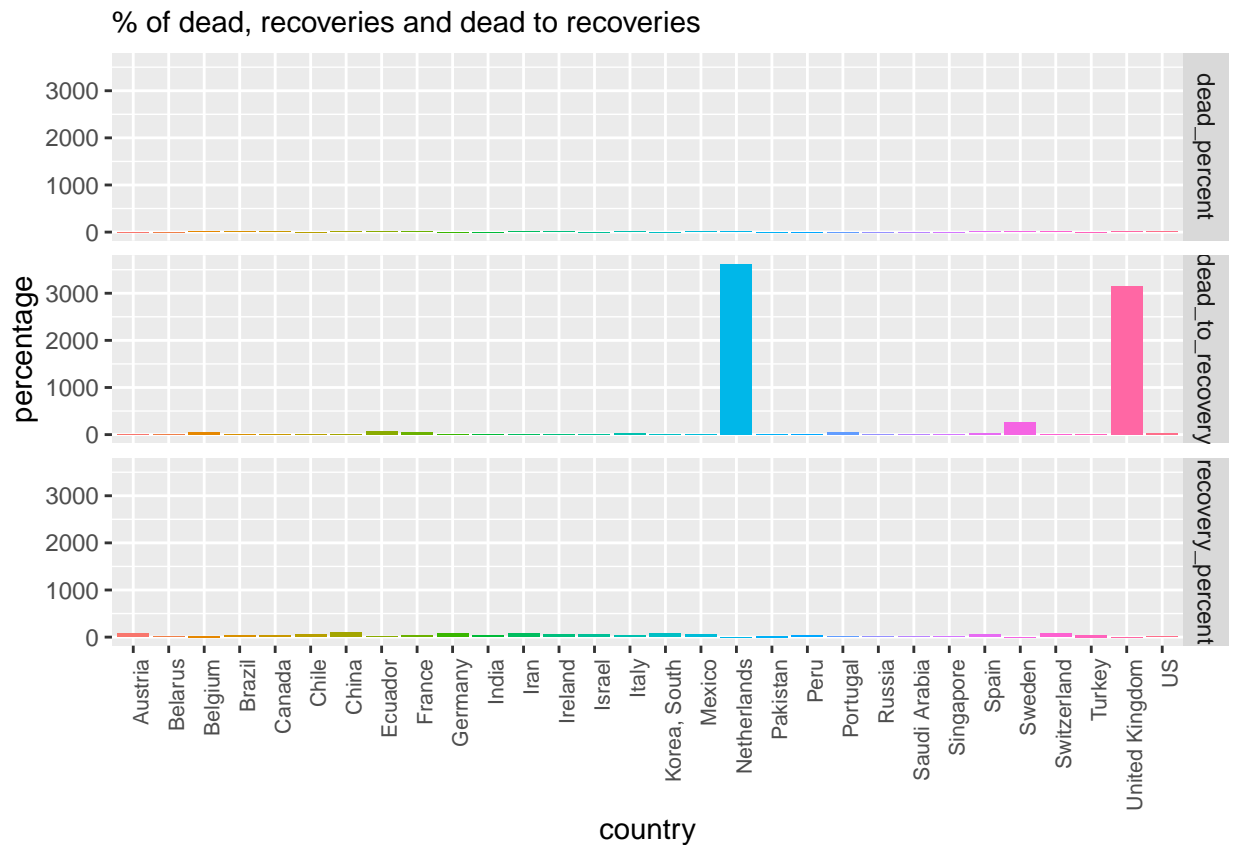


```
# Number of people dead for every 100 people recovered in the top 30 hits countries
percents <- covid_ds %>%
```

```
  select(country, type, cases) %>%
  group_by(country, type) %>%
  summarize(tot_cases= sum(cases)) %>%
  spread(type, tot_cases) %>%
  summarize(recovery_percent= 100*sum(recovered)/sum(confirmed),
            dead_percent= 100*sum(dead)/sum(confirmed),
            dead_to_recovery= 100*sum(dead)/sum(recovered)) %>%
  gather(type, percentage, recovery_percent:dead_to_recovery)
```

```
# Plot Number of people dead for every 100 people recovered in select countries
percents %>%
```

```
  filter(country %in% select_countries) %>%
  ggplot(aes(country, percentage, fill=country)) +
  geom_col()+
  facet_grid(type~.)+
  theme(axis.text.x = element_text(
    angle = 90,
    size = 8,
    hjust = 1
  )),
  legend.position = "none") +
  labs(subtitle="% of dead, recoveries and dead to recoveries")
```



# Plot Number of people dead for every 100 people recovered in select countries, excluding UK and Netherlands  
 percents %>%

```
filter(country %in% select_countries & country != 'United Kingdom' & country != 'Netherlands') %>%
ggplot(aes(country, percentage, fill=country)) +
geom_col()+
facet_grid(type~.)+
theme(axis.text.x = element_text(
  angle = 90,
  size = 8,
  hjust = 1
),
legend.position = "none") +
labs(subtitle="% of dead, recoveries and dead to recoveries")
```

