COVID-19 Data Analysis with R - Worldwide

Yanchang Zhao
yanchang@RDataMining.com
http://RDataMining.com

09 April 2020

Contents

1 Introduction **1**
1.1 Data Source . . . 1
1.2 R Packages . . . 1

2 Loading Data **1**

3 Data Preparation **3**
3.1 Data Cleaning . . . 3
3.2 Worldwide Cases . . . 4
3.3 Daily Increases and Death Rates . . . 4

4 Worldwide Cases **5**
4.1 World Map . . . 5
4.2 Number of Cases . . . 6
4.3 Current Confirmed Cases . . . 7
4.4 Deaths and Recovered Cases . . . 8
4.5 Death Rates . . . 9

5 Top Twenty Countries **10**
5.1 Confirmed vs Deaths . . . 14
5.2 Comparison across Countries . . . 16
5.3 Death Rates . . . 24
5.4 Countries with Highest Death Rates . . . 26

6 Conclusions **26**

Appendix A. Processed Data **26**
Appendix A.1 COVID-19 Cases Worldwide . . . 26
Appendix A.2 Latest Cases by Country . . . 28

Appendix B. How to Cite This Work **31**

Appendix C. Contact **31**

1 Introduction

This is an analysis report of the Novel Coronavirus (COVID-19) around the world, to demonstrate data processing and visualisation with R, *tidyverse* and *ggplot2*. This report will be updated from time to time, with new data and more analysis. Please find its latest version at http://www.rdatamining.com/docs/Coronavirus-data-analysis-world.pdf.

A similar COVID-19 analysis report for China is available at http://www.rdatamining.com/docs/Coronavirus-data-analysis-china.pdf, if you are particually interested what has happened in China.

1.1 Data Source

The data source used for this analysis is *the 2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository*[1] built by the Center for Systems Science and Engineering, Johns Hopkins University.

1.2 R Packages

Blow is a list of R packages used for this analysis. Package *magrittr* is for pipe operations like `%>%` and `%<>%` and *lubridate* for date operations. Package *tidyverse* is a collection of R packages for data science, including *dplyr* and *tidyr* for data processing and *ggplot2* for graphics. Package *gridExtra* is for arranging multiple grid-based plots on a page and *kableExtra* works together with `kable()` from *knitr* to build complex HTML or LaTeX tables.

```
library(magrittr) # pipe operations
library(lubridate) # date operations
library(tidyverse) # ggplot2, tidyr, dplyr...
library(gridExtra) # multiple grid-based plots on a page
library(ggforce) # accelerating ggplot2
library(kableExtra) # complex tables
library(leaflet) # map
```

2 Loading Data

At first, the datasets, which are three CSV files, are downloaded and saved as local files and then are loaded into R.

```
## source data files
filenames <- c('time_series_covid19_confirmed_global.csv',
               'time_series_covid19_deaths_global.csv',
               'time_series_covid19_recovered_global.csv')
url.path <- paste0('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/',
                   'master/csse_covid_19_data/csse_covid_19_time_series/')

## download files to local
download <- function(filename) {
  url <- file.path(url.path, filename)
  dest <- file.path('./data', filename)
  download.file(url, dest)
}
bin <- lapply(filenames, download)

## load data into R
raw.data.confirmed <- read.csv('./data/time_series_covid19_confirmed_global.csv')
raw.data.deaths <- read.csv('./data/time_series_covid19_deaths_global.csv')
raw.data.recovered <- read.csv('./data/time_series_covid19_recovered_global.csv')

dim(raw.data.confirmed)
```

```
## [1] 263  81
```

Each dataset has 263 rows, corresponding to country/region/province/state. It has 81 columns. Starting from column 5, each column corresponds to a single day. Here we have a look at the first 10 rows and the first 10 columns.

[1] https://github.com/CSSEGISandData/COVID-19

```
raw.data.confirmed[1:10, 1:10] %>%
  kable('latex', booktabs=T, caption='Raw Data (Confirmed, First 10 Columns only)') %>%
  kable_styling(font_size=6, latex_options = c('striped', 'hold_position', 'repeat_header'))
```

Table 1: Raw Data (Confirmed, First 10 Columns only)

Province.State	Country.Region	Lat	Long	X1.22.20	X1.23.20	X1.24.20	X1.25.20	X1.26.20	X1.27.20
	Afghanistan	33.0000	65.0000	0	0	0	0	0	0
	Albania	41.1533	20.1683	0	0	0	0	0	0
	Algeria	28.0339	1.6596	0	0	0	0	0	0
	Andorra	42.5063	1.5218	0	0	0	0	0	0
	Angola	-11.2027	17.8739	0	0	0	0	0	0
	Antigua and Barbuda	17.0608	-61.7964	0	0	0	0	0	0
	Argentina	-38.4161	-63.6167	0	0	0	0	0	0
	Armenia	40.0691	45.0382	0	0	0	0	0	0
Australian Capital Territory	Australia	-35.4735	149.0124	0	0	0	0	0	0
New South Wales	Australia	-33.8688	151.2093	0	0	0	0	3	4

Below we check the time frame of the data.

```
n.col <- ncol(raw.data.confirmed)
## get dates from column names
dates <- names(raw.data.confirmed)[5:n.col] %>% substr(2,8) %>% mdy()
range(dates)
```

```
## [1] "2020-01-22" "2020-04-07"
```

```
min.date <- min(dates)
max.date <- max(dates)
min.date.txt <- min.date %>% format('%d %b %Y')
max.date.txt <- max.date %>% format('%d %b %Y')
```

It shows that the data was last updated on 07 Apr 2020 UTC and all the stats and charts in this report are based on that data.

3 Data Preparation

3.1 Data Cleaning

The three datesets are converted from wide to long format and then are aggregated by country. After that, they are merged into one single dataset.

```
## data cleaning and transformation
cleanData <- function(data) {
  ## remove some columns
  data %<>% select(-c(Province.State, Lat, Long)) %>% rename(country=Country.Region)
  ## convert from wide to long format
  data %<>% gather(key=date, value=count, -country)
  ## convert from character to date
  data %<>% mutate(date = date %>% substr(2,8) %>% mdy())
  ## aggregate by country
  data %<>% group_by(country, date) %>% summarise(count=sum(count, na.rm=T)) %>% as.data.frame()
  return(data)
}

## clean the three datasets
data.confirmed <- raw.data.confirmed %>% cleanData() %>% rename(confirmed=count)
data.deaths <- raw.data.deaths %>% cleanData() %>% rename(deaths=count)
```

```r
data.recovered <- raw.data.recovered %>% cleanData() %>% rename(recovered=count)

## merge above 3 datasets into one, by country and date
data <- data.confirmed %>% merge(data.deaths, all=T) %>% merge(data.recovered, all=T)
# data %<>% mutate(recovered = ifelse(is.na(recovered), lag(recovered, 1), recovered))

## countries/regions with confirmed cases, excl. cruise ships
countries <- data %>% pull(country) %>% setdiff('Cruise Ship')

## first 10 records when it first broke out in China
data %>% filter(country=='China') %>% head(10) %>%
  kable('latex', booktabs=T, caption='Raw Data (with first 10 Columns Only)',
        format.args=list(big.mark=',')) %>%
  kable_styling(latex_options = c('striped', 'hold_position', 'repeat_header'))
```

Table 2: Raw Data (with first 10 Columns Only)

country	date	confirmed	deaths	recovered
China	2020-01-22	548	17	28
China	2020-01-23	643	18	30
China	2020-01-24	920	26	36
China	2020-01-25	1,406	42	39
China	2020-01-26	2,075	56	49
China	2020-01-27	2,877	82	58
China	2020-01-28	5,509	131	101
China	2020-01-29	6,087	133	120
China	2020-01-30	8,141	171	135
China	2020-01-31	9,802	213	214

There are 184 countries with confirmed COVID-19 cases, as of 07 Apr 2020 UTC.

3.2 Worldwide Cases

The raw data provide the daily number of cases in every country. They are aggregated below to derive the daily stats of the whole world.

```r
## counts for the whole world
data.world <- data %>% group_by(date) %>%
  summarise(country='World',
            confirmed = sum(confirmed, na.rm=T),
            deaths = sum(deaths, na.rm=T),
            recovered = sum(recovered, na.rm=T))

data %<>% rbind(data.world)

## current confirmed cases
data %<>% mutate(current.confirmed = confirmed - deaths - recovered)
```

3.3 Daily Increases and Death Rates

After that, the daily increases of death and recovered cases and the death rates are calculated.

`rate.upper` is caculated with the total dead and recovered cases. It is the upper bound of death rate and the reasons are

1) there were much more deaths than recovered cases when the coronavirus broke out and when it was not contained, and
2) the daily number of death will decrease and that of recovered will increase as it becomes contained and more effective measures and treatments are used.

`rate.lower` is caculated with total dead and confirmed cases. It is a lower bound of death rate, because there are and will be new deaths from the current confirmed cases. The final death rate is expected to be in between of the above two rates.

`rate.daily` is caculated with the daily dead and recovered cases and therefore is more volatile than the above two. However, it can give us a clue of the current situlation: whether it is very serious or is getting better.

```
## sort by country and date
data %<>% arrange(country, date)

## daily increases of deaths and recovered cases
## set NA to the increases on day1
n <- nrow(data)
day1 <- min(data$date)
data %<>% mutate(new.confirmed = ifelse(date == day1, NA, confirmed - lag(confirmed, n=1)),
                 new.deaths = ifelse(date == day1, NA, deaths - lag(deaths, n=1)),
                 new.recovered = ifelse(date == day1, NA, recovered - lag(recovered, n=1)))

## change negative number of new cases to zero
data %<>% mutate(new.confirmed = ifelse(new.confirmed < 0, 0, new.confirmed),
                 new.deaths = ifelse(new.deaths < 0, 0, new.deaths),
                 new.recovered = ifelse(new.recovered < 0, 0, new.recovered))

## death rate based on total deaths and recovered cases
data %<>% mutate(rate.upper = (100 * deaths / (deaths + recovered)) %>% round(1))
## lower bound: death rate based on total confirmed cases
data %<>% mutate(rate.lower = (100 * deaths / confirmed) %>% round(1))
## death rate based on the number of death/recovered on every single day
data %<>% mutate(rate.daily = (100 * new.deaths / (new.deaths + new.recovered)) %>% round(1))


## convert from wide to long format, for drawing area plots
data.long <- data %>%
  select(c(country, date, confirmed, current.confirmed, recovered, deaths)) %>%
  gather(key=type, value=count, -c(country, date))
## set factor levels to show them in a desirable order
data.long %<>% mutate(type=recode_factor(type, confirmed='Total Confirmed',
                                         current.confirmed='Current Confirmed',
                                         recovered='Recovered',
                                         deaths='Deaths'))


## convert from wide to long format, for drawing area plots
rates.long <- data %>%
  # filter(country %in% top.countries) %>%
  select(c(country, date, rate.upper, rate.lower, rate.daily)) %>%
  # mutate(country=factor(country, levels=top.countries)) %>%
  gather(key=type, value=count, -c(country, date))
# set factor levels to show them in a desirable order
```

```
rates.long %<>% mutate(type=recode_factor(type, rate.daily='Daily',
                                          rate.lower='Lower bound',
                                          rate.upper='Upper bound'))
```

4 Worldwide Cases

After tidying up the data, we visualise it with various charts.

4.1 World Map

Below is a world map of vconfirmed cases. An interactive map can be created if running the code in R or RStudio, or knitting it into a HTML file.

```
## select last column, which is the number of latest confirmed cases
x <- raw.data.confirmed
x$confirmed <- x[, ncol(x)]
x %<>% select(c(Country.Region, Province.State, Lat, Long, confirmed)) %>%
  mutate(txt=paste0(Country.Region, ' - ', Province.State, ': ', confirmed))

m <- leaflet(width=1200, height=800) %>% addTiles()
# circle marker (units in pixels)
m %<>% addCircleMarkers(x$Long, x$Lat,
                        radius=2+log2(x$confirmed), stroke=F,
                        color='red', fillOpacity=0.3,
                        popup=x$txt)

# world
m
```

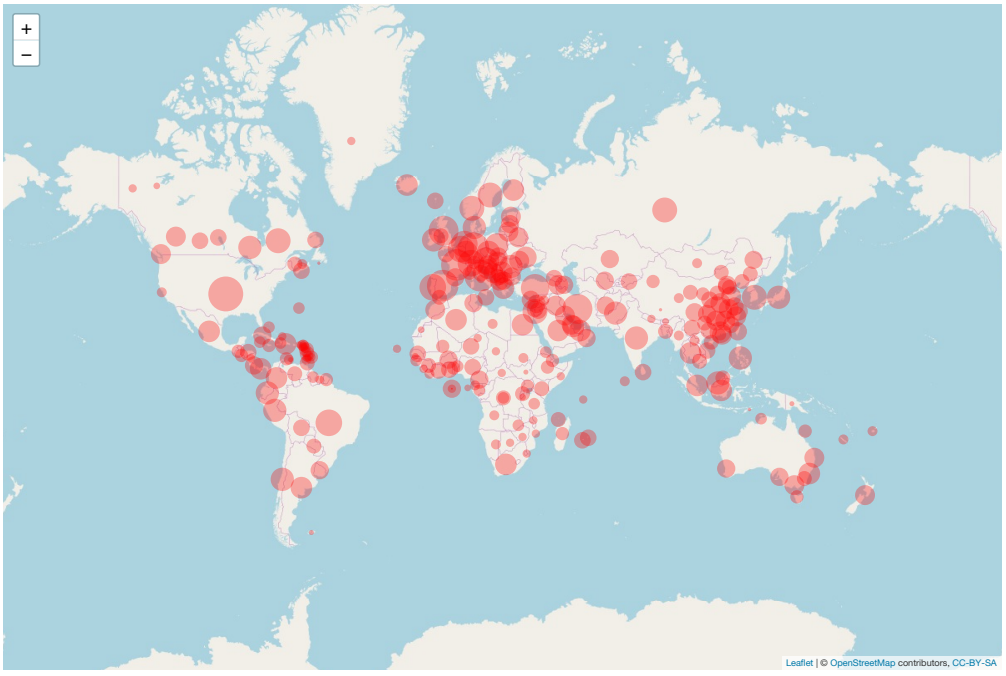


Figure 1: World Map

Views of some specific countries or regions can be produced with the script below.

```
## China
m %>% setView(95, 35, zoom=4)
## Australia and New Zealand
m %>% setView(135, -27, zoom=4)
## US and Canada
m %>% setView(-105, 40, zoom=4)
## Europe
m %>% setView(10, 50, zoom=4)
```

4.2 Number of Cases

In the rest of this section, we will focuse on the cases worldwide. Similar analysis for a single country can be done by filter the data with the corresponding country name.

```
# data %<>% filter(country=='China')
# data %<>% filter(country=='Australia')
world.long <- data.long %>% filter(country == 'World')

## cases - area plot
plot1 <- world.long %>% filter(type != 'Total Confirmed') %>%
  ggplot(aes(x=date, y=count)) +
  geom_area(aes(fill=type), alpha=0.5) +
  labs(title=paste0('Numbers of Cases Worldwide - ', max.date.txt)) +
  scale_fill_manual(values=c('red', 'green', 'black')) +
  theme(legend.title=element_blank(), legend.position='bottom',
        plot.title = element_text(size=8),
        axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        legend.key.size=unit(0.2, 'cm'),
        legend.text=element_text(size=6),
        axis.text=element_text(size=7),
        axis.text.x=element_text(angle=45, hjust=1))

plot2 <- world.long %>%
  ggplot(aes(x=date, y=count)) +
  geom_line(aes(color=type)) +
  labs(title=paste0('Numbers of Cases Worldwide (log scale) - ', max.date.txt)) +
  scale_color_manual(values=c('purple', 'red', 'green', 'black')) +
  theme(legend.title=element_blank(), legend.position='bottom',
        plot.title = element_text(size=8),
        axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        legend.key.size=unit(0.2, 'cm'),
        legend.text=element_text(size=6),
        axis.text=element_text(size=7),
        axis.text.x=element_text(angle=45, hjust=1)) +
  scale_y_continuous(trans='log10')

grid.arrange(plot1, plot2, ncol=2)
```

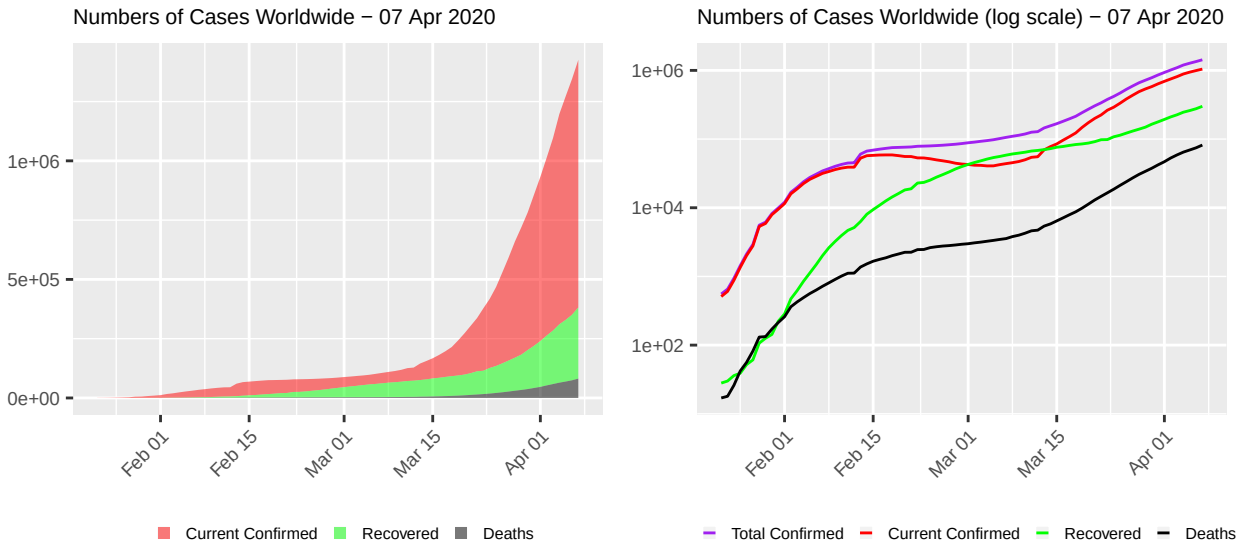


Figure 2: COVID-19 Cases Worldwide

4.3 Current Confirmed Cases

```
data.world <- data %>% filter(country=='World')
n <- nrow(data.world)
```

```
## current confirmed and daily new confirmed
plot1 <- ggplot(data.world, aes(x=date, y=current.confirmed)) +
  geom_point() + geom_smooth() +
  xlab('') + ylab('Count') + labs(title='Current Confirmed Cases') +
  theme(axis.text.x=element_text(angle=45, hjust=1))
plot2 <- ggplot(data.world, aes(x=date, y=new.confirmed)) +
  geom_point() + geom_smooth() +
  xlab('') + ylab('Count') + labs(title='Daily New Confirmed Cases') +
  theme(axis.text.x=element_text(angle=45, hjust=1))
## show two plots side by side
grid.arrange(plot1, plot2, ncol=2)
```

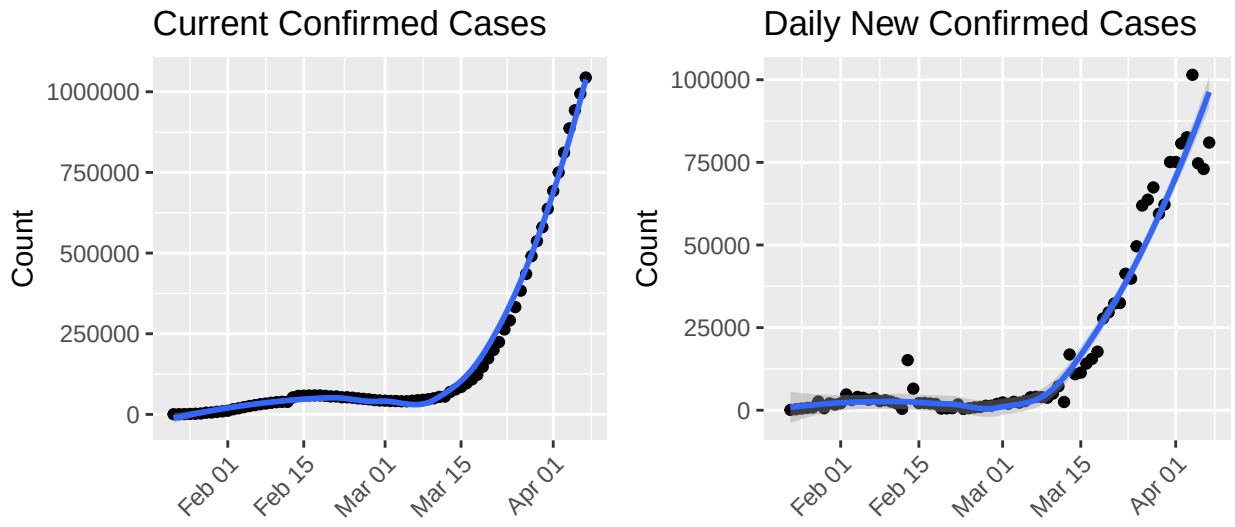


Figure 3: Current Confirmed Cases

Figure 3 shows the numbers of current (see left chart) and new (see right chart) confirmed cases. The blue lines are smoothed conditional means and the grey band around them show the 95% confidence interval.

4.4 Deaths and Recovered Cases

```
## a scatter plot with a smoothed line and vertical x-axis labels
plot1 <- ggplot(data.world, aes(x=date, y=deaths)) +
  geom_point() + geom_smooth() +
  xlab('') + ylab('Count') + labs(title='Accumulative Deaths') +
  theme(axis.text.x=element_text(angle=45, hjust=1))
plot2 <- ggplot(data.world, aes(x=date, y=recovered)) +
  geom_point() + geom_smooth() +
  xlab('') + ylab('Count') + labs(title='Accumulative Recovered Cases') +
  theme(axis.text.x=element_text(angle=45, hjust=1))
plot3 <- ggplot(data.world, aes(x=date, y=new.deaths))  +
  geom_point() + geom_smooth()  +
  xlab('') + ylab('Count') + labs(title='New Deaths') +
  theme(axis.text.x=element_text(angle=45, hjust=1))
plot4 <- ggplot(data.world, aes(x=date, y=new.recovered)) +
  geom_point() + geom_smooth()  +
  xlab('') + ylab('Count') + labs(title='New Recovered Cases') +
  theme(axis.text.x=element_text(angle=45, hjust=1))
## show four plots together, with 2 plots in each row
grid.arrange(plot1, plot2, plot3, plot4, nrow=2)
```

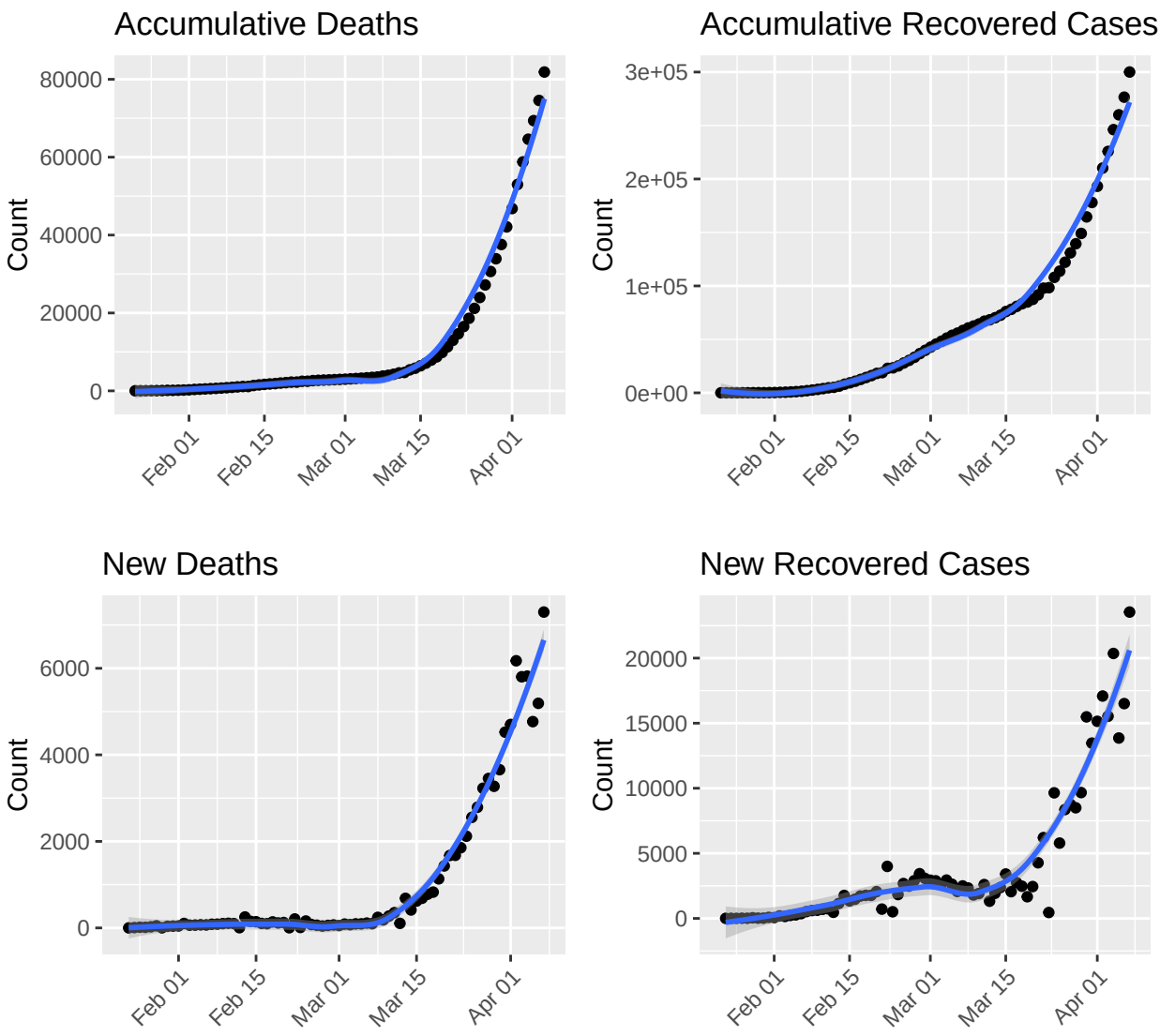


Figure 4: Deaths and Recovered Cases

4.5 Death Rates

Figure 5 shows death rates caculated in three different ways (see Section 3.3 for details). The left chart shows the death rates from 22 Jan 2020 to 07 Apr 2020 and the right one is a zoom-in view of the rates in last two weeks.

In the right chart, the upper bound (in blue) is decreasing, as there will be more recovered cases and fewer dead ones daily as time goes on. However, the lower bound (in green) keeps going up, as there are and will be new deaths from the current confirmed cases. Therefore, the final death rate is expected to be in-between of those two rates, and based on the latest data retrieved as of 07 Apr 2020 UTC, it will be between 5.7% and 21.4%.

A surge in the daily death rate (in red) suggests that the situlation is changing dramatically (actually, getting worse) and that above lower/upper bounds are likely to increase shortly. A likely reason of that surge is the recent outbreak of coronavirus in Italy, Iran and some other European countries.

```
## three death rates
plot1 <- ggplot(data.world, aes(x=date)) +
```

```
  geom_line(aes(y=rate.upper, colour='Upper bound')) +
  geom_line(aes(y=rate.lower, colour='Lower bound')) +
  geom_line(aes(y=rate.daily, colour='Daily')) +
  xlab('') + ylab('Death Rate (%)') + labs(title='Overall') +
  theme(legend.position='bottom', legend.title=element_blank(),
        legend.text=element_text(size=8),
        legend.key.size=unit(0.5, 'cm'),
        axis.text.x=element_text(angle=45, hjust=1))
## focusing on last 2 weeks
y.max <- data.world[n-(14:0), ] %>% select(rate.upper, rate.lower, rate.daily) %>% max()
plot2 <- ggplot(data.world[n-(14:0),], aes(x=date)) +
  geom_line(aes(y=rate.upper, colour='Upper bound')) +
  geom_line(aes(y=rate.lower, colour='Lower bound')) +
  geom_line(aes(y=rate.daily, colour='Daily')) +
  xlab('') + ylab('Death Rate (%)') + labs(title='Last two weeks') +
  theme(legend.position='bottom', legend.title=element_blank(),
        legend.text=element_text(size=8),
        legend.key.size=unit(0.5, 'cm'),
        axis.text.x=element_text(angle=45, hjust=1)) +
  ylim(c(0, y.max))
grid.arrange(plot1, plot2, ncol=2)
```

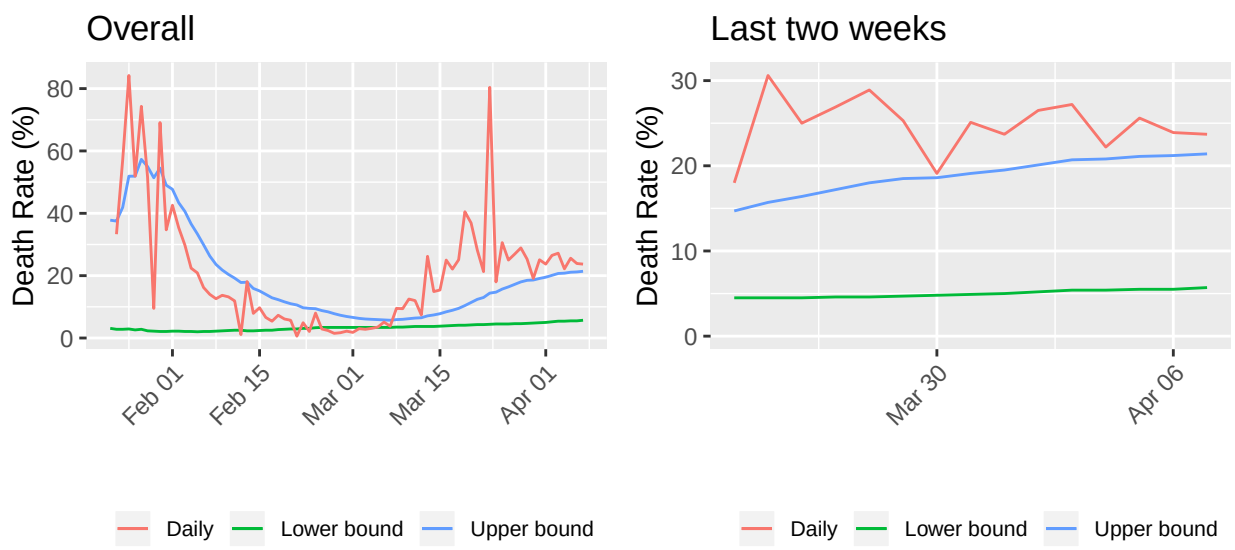


Figure 5: Death Rate

5 Top Twenty Countries

Next, we will have a look at the top 20 countries in total confirmed cases.

```
## ranking by confirmed cases
data.latest.all <- data %>% filter(date == max(date)) %>%
  select(country, date,
         confirmed, new.confirmed, current.confirmed,
         recovered, deaths, new.deaths, death.rate=rate.lower) %>%
  mutate(ranking = dense_rank(desc(confirmed)))
```

```
k <- 20
## top 20 countries: 21 incl. 'World'
top.countries <- data.latest.all %>% filter(ranking <= k + 1) %>%
  arrange(ranking) %>% pull(country) %>% as.character()
top.countries %>% setdiff('World') %>% print()
```

```
##  [1] "US"             "Spain"          "Italy"          "France"
##  [5] "Germany"        "China"          "Iran"           "United Kingdom"
##  [9] "Turkey"         "Switzerland"    "Belgium"        "Netherlands"
## [13] "Canada"         "Brazil"         "Austria"        "Portugal"
## [17] "Korea, South"   "Israel"         "Sweden"         "Russia"
```

```
## add 'Others'
top.countries %<>% c('Others')
## put all others in a single group of 'Others'
data.latest <- data.latest.all %>% filter(!is.na(country)) %>%
  mutate(country=ifelse(ranking <= k + 1, as.character(country), 'Others')) %>%
  mutate(country=country %>% factor(levels=c(top.countries)))
data.latest %<>% group_by(country) %>%
  summarise(confirmed=sum(confirmed), new.confirmed=sum(new.confirmed),
            current.confirmed=sum(current.confirmed),
            recovered=sum(recovered), deaths=sum(deaths), new.deaths=sum(new.deaths)) %>%
  mutate(death.rate=(100 * deaths/confirmed) %>% round(1))
data.latest %<>% select(c(country, confirmed, deaths, death.rate,
                 new.confirmed, new.deaths, current.confirmed))
```

```
data.latest %>% mutate(death.rate=death.rate %>% format(nsmall=1) %>% paste0('%')) %>%
  kable('latex', booktabs=T, row.names=T, align=c('l', rep('r', 6)),
        caption=paste0('Cases in Top 20 Countries - ', max.date.txt,
                       '. See a complete list of all infected countries at the end of this report.'),
        format.args=list(big.mark=',')) %>%
  kable_styling(font_size=7, latex_options=c('striped', 'hold_position', 'repeat_header'))
```

```
## convert from wide to long format, for drawing area plots
data.latest.long <- data.latest %>% filter(country!='World') %>%
  gather(key=type, value=count, -country)
## set factor levels to show them with proper text and in a desirable order
data.latest.long %<>% mutate(type=recode_factor(type,
                                    confirmed='Total Confirmed',
                                    deaths='Total Deaths',
                                    death.rate='Death Rate (%)',
                                    new.confirmed='New Confirmed (compared with one day before)',
                                    new.deaths='New Deaths (compared with one day before)',
                                    current.confirmed='Current Confirmed'))
```

```
## bar chart
data.latest.long %>% ggplot(aes(x=country, y=count, fill=country, group=country)) +
  geom_bar(stat='identity') +
  geom_text(aes(label=count, y=count), size=2, vjust=0) +
  xlab('') + ylab('') +
  labs(title=paste0('Top 20 Countries with Most Confirmed Cases - ', max.date.txt)) +
  scale_fill_discrete(name='Country', labels=aes(count)) +
  theme(legend.title=element_blank(),
        legend.position='none',
        plot.title=element_text(size=11),
```

Table 3: Cases in Top 20 Countries - 07 Apr 2020. See a complete list of all infected countries at the end of this report.

	country	confirmed	deaths	death.rate	new.confirmed	new.deaths	current.confirmed
1	World	1,426,096	81,865	5.7%	80,995	7,300	1,044,177
2	US	396,223	12,722	3.2%	29,556	1,939	361,738
3	Spain	141,942	14,045	9.9%	5,267	704	84,689
4	Italy	135,586	17,127	12.6%	3,039	604	94,067
5	France	110,065	10,343	9.4%	11,102	1,417	80,199
6	Germany	107,663	2,016	1.9%	4,289	206	69,566
7	China	82,718	3,335	4.0%	53	0	1,973
8	Iran	62,589	3,872	6.2%	2,089	133	31,678
9	United Kingdom	55,949	6,171	11.0%	3,670	786	49,453
10	Turkey	34,109	725	2.1%	3,892	76	31,802
11	Switzerland	22,253	821	3.7%	596	56	12,728
12	Belgium	22,194	2,035	9.2%	1,380	403	16,002
13	Netherlands	19,709	2,108	10.7%	783	234	17,329
14	Canada	17,872	375	2.1%	1,309	36	13,706
15	Brazil	14,034	686	4.9%	1,873	122	13,221
16	Austria	12,639	243	1.9%	342	23	8,350
17	Portugal	12,442	345	2.8%	712	34	11,913
18	Korea, South	10,331	192	1.9%	47	6	3,445
19	Israel	9,248	65	0.7%	344	8	8,413
20	Sweden	7,693	591	7.7%	487	114	6,897
21	Russia	7,497	58	0.8%	1,154	11	6,945
22	Others	143,340	3,990	2.8%	9,011	388	120,063

```
      axis.text=element_text(size=7),
      axis.text.x=element_text(angle=45, hjust=1)) +
facet_wrap(~type, ncol=1, scales='free_y')
```

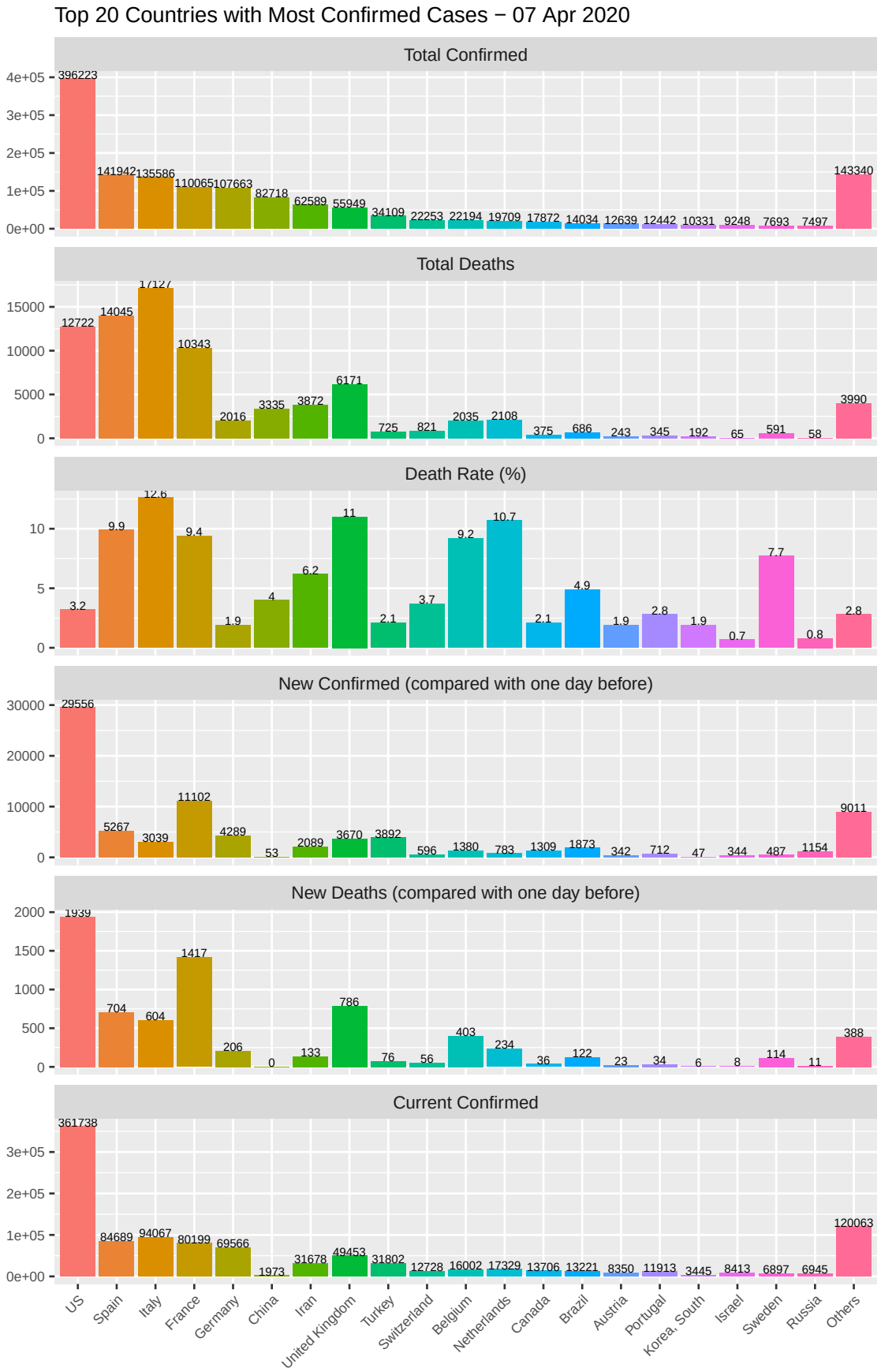
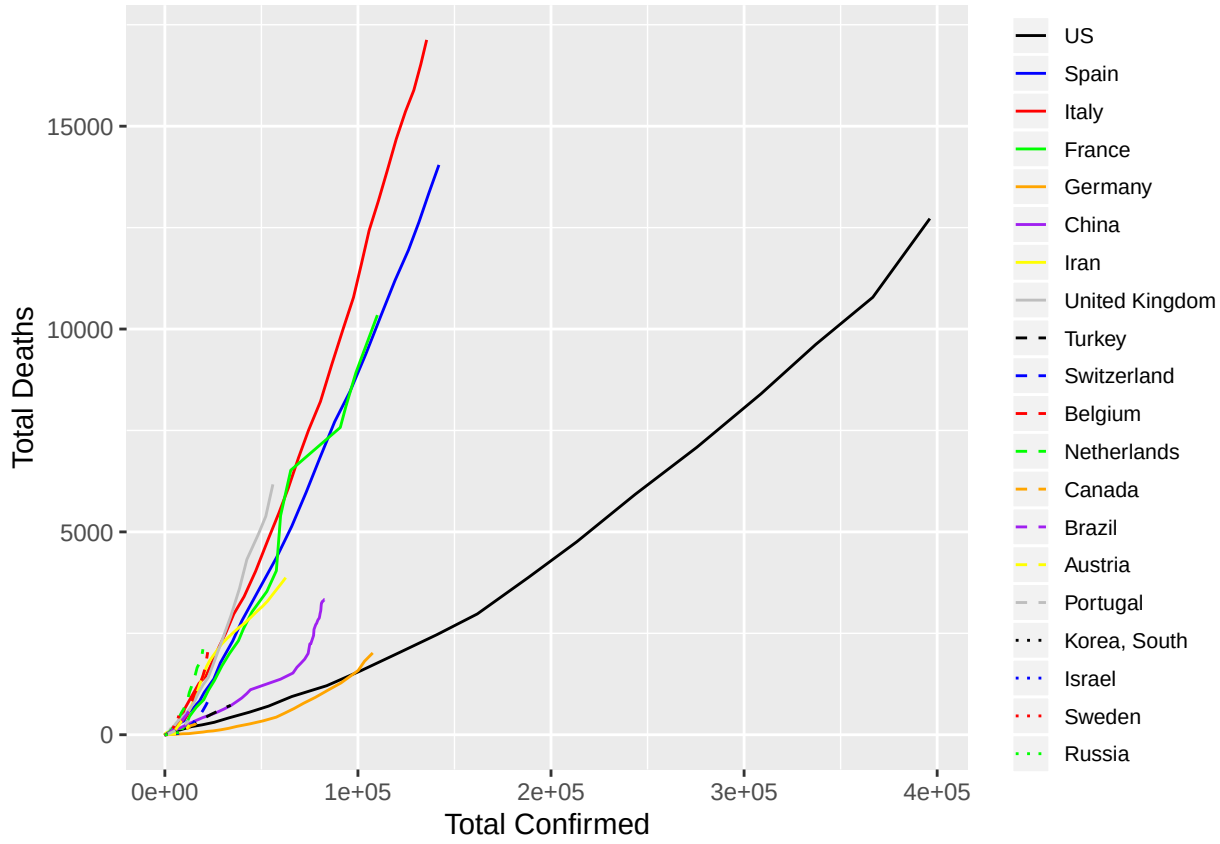


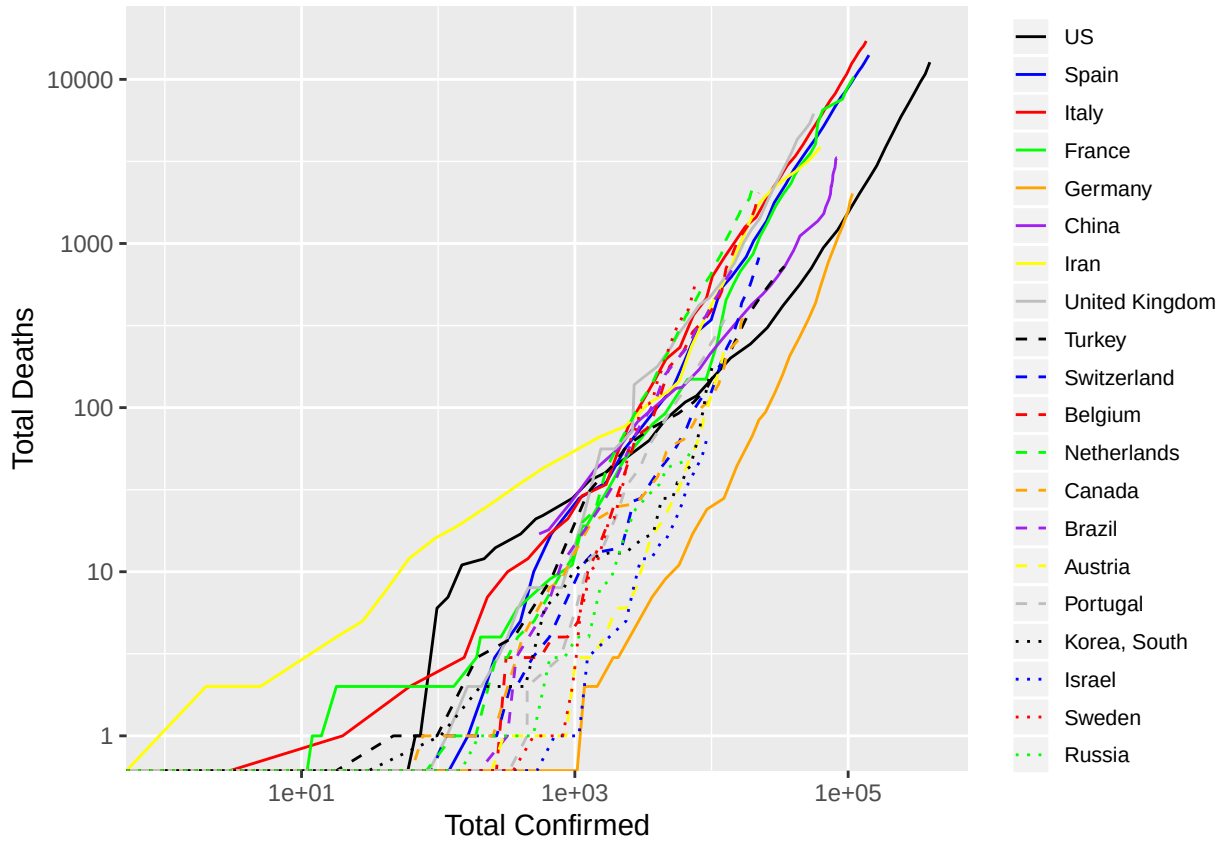
Figure 6: Top 20 Countries with Most Confirmed Cases

5.1 Confirmed vs Deaths

```
# linetypes <- rep(c("dotted", "dashed", "solid"), each=8)
# colors <- rep(c('grey', 'yellow', 'purple', 'orange', 'green', 'red', 'blue', 'black'), 3)
linetypes <- rep(c("solid", "dashed", "dotted"), each=8)
colors <- rep(c('black', 'blue', 'red', 'green', 'orange', 'purple', 'yellow', 'grey'), 3)
df <- data %>% filter(country %in% setdiff(top.countries, c('World', 'Others'))) %>%
  mutate(country=country %>% factor(levels=c(top.countries)))
p <- df %>% ggplot(aes(x=confirmed, y=deaths, group=country)) +
  geom_line(aes(color=country, linetype=country)) +
  xlab('Total Confirmed') + ylab('Total Deaths') +
  scale_linetype_manual(values=linetypes) +
  scale_color_manual(values=colors) +
  theme(legend.title=element_blank(),
        legend.text=element_text(size=8),
        legend.key.size=unit(0.5, 'cm'))
p
```




```
p + scale_x_log10() + scale_y_log10()
```



The two figures below show the numbers of confirmed cases and deaths of top 20 countries, as well as the death rates up to 07 Apr 2020.

```
df <- data.latest %>% filter(country %in% setdiff(top.countries, 'World'))

plot1 <- df %>% ggplot(aes(x=confirmed, y=deaths, col=death.rate, size=current.confirmed)) +
  scale_size(name='Current Confirmed', trans='log2', breaks=c(1e3, 2e3, 5e3, 1e4, 2e4, 4e4)) +
  geom_text(aes(label=country), size=2.5, check_overlap=T, vjust=-1.6) +
  geom_point() +
  xlab('Total Confirmed') + ylab('Total Deaths') +
  labs(col="Death Rate (%)") +
  scale_color_gradient(low='#56B1F7', high='#132B43') +
  scale_x_log10() + scale_y_log10()

plot2 <- df %>% ggplot(aes(x=new.confirmed, y=new.deaths, col=death.rate, size=current.confirmed)) +
  scale_size(name='Current Confirmed', trans='log2', breaks=c(1e3, 2e3, 5e3, 1e4, 2e4, 4e4)) +
  geom_text(aes(label=country), size=2.5, check_overlap=T, vjust=-1.6) +
  geom_point() +
  xlab('New Confirmed') + ylab('New Deaths') +
  labs(col="Death Rate (%)") +
  scale_color_gradient(low='#56B1F7', high='#132B43') +
  scale_x_log10() + scale_y_log10()

grid.arrange(plot1, plot2, ncol=1)
```

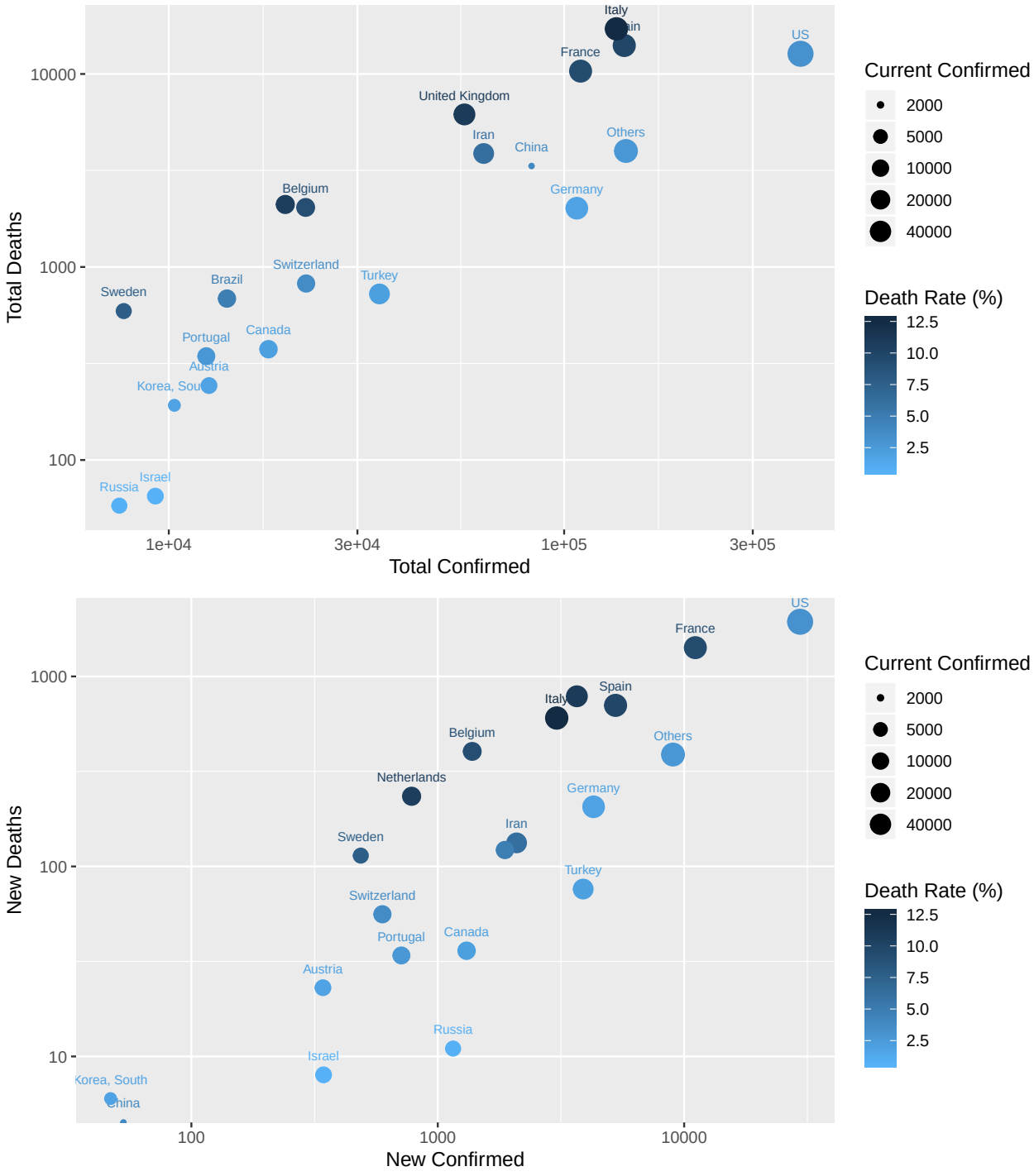


Figure 7: Top 20 Countries

5.2 Comparison across Countries

The area plots blow show the numbers of dead, recovered, total and current confimed cases. Note that, in the area plot, the total number of total confirmed cases is represented by the total areas of current confirmed, recovered and dead.

```r
## plot: cases by type
df <- data.long %>% filter(country %in% top.countries) %<>%
  mutate(country=country %>% factor(levels=c(top.countries)))

p <- df %>% filter(country != 'World') %>%
  ggplot(aes(x=date, y=count)) + xlab('') + ylab('Count')  +
  theme(legend.title=element_blank(),
        legend.text=element_text(size=8),
        legend.key.size=unit(0.5, 'cm'),
        plot.title=element_text(size=11),
        axis.text.x=element_text(angle=45, hjust=1)) +
  facet_wrap(~type, ncol=2, scales='free_y')

## area plot
plot1 <- p + geom_area(aes(fill=country)) +
  labs(title=paste0('Cases around the World - ', max.date.txt))

## line plot and in log scale
# linetypes <- rep(c("solid", "dashed", "dotted"), each=8)
# colors <- rep(c('black', 'blue', 'red', 'green', 'orange', 'purple', 'yellow', 'grey'), 3)
plot2 <- p + geom_line(aes(color=country, linetype=country)) +
  scale_linetype_manual(values=linetypes) +
  scale_color_manual(values=colors) +
  labs(title=paste0('Cases around the World - Log Scale - ', max.date.txt)) +
  scale_y_continuous(trans='log10')

grid.arrange(plot1, plot2, ncol=1)
```

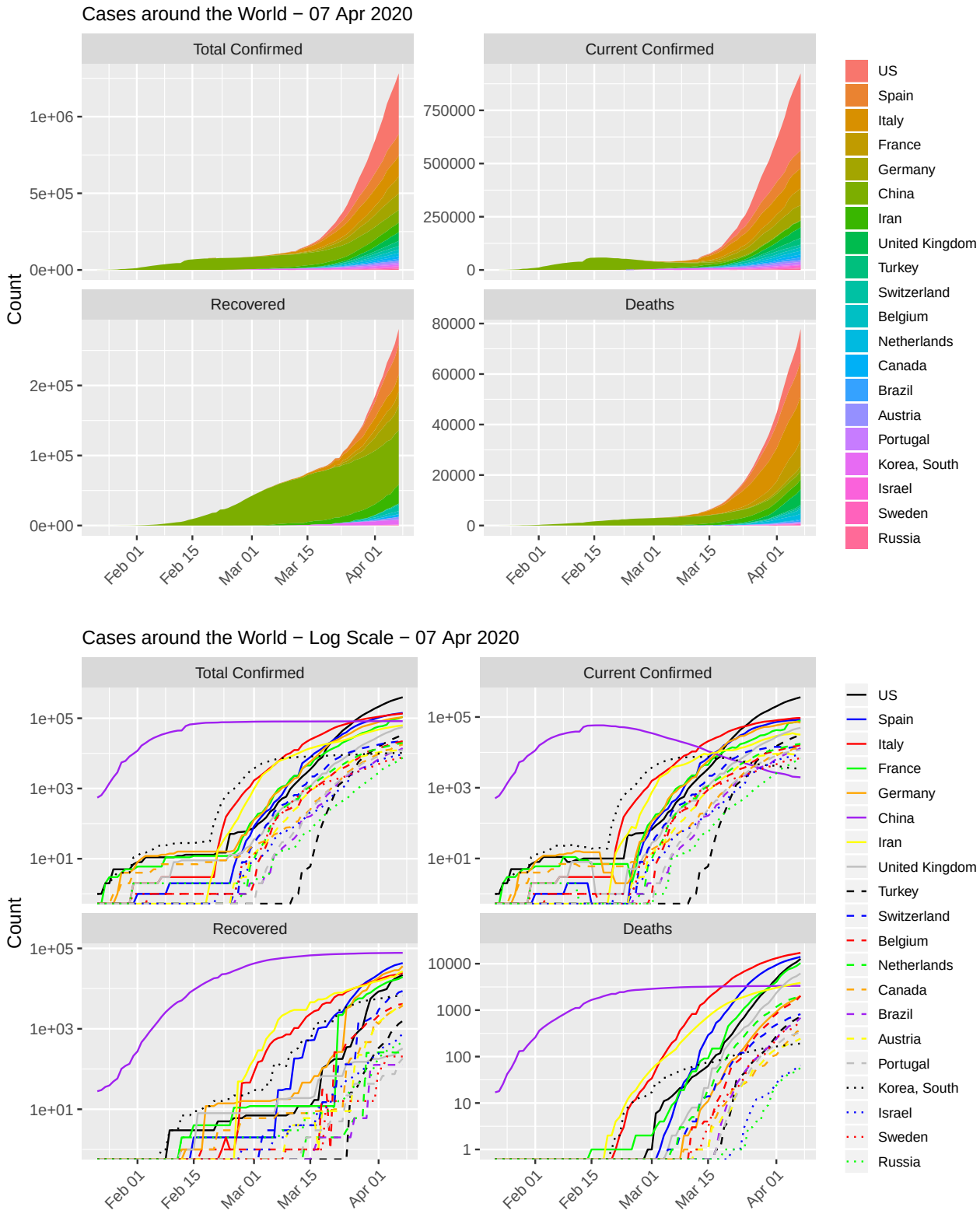


Figure 8: Cases around the World

```
## plot: excluding China
p <- df %>% filter(!(country %in% c('World', 'China'))) %>%
```

```
  ggplot(aes(x=date, y=count)) + xlab('') + ylab('Count') +
  theme(legend.title=element_blank(),
        legend.text=element_text(size=8),
        legend.key.size=unit(0.5, 'cm'),
        plot.title=element_text(size=11),
        axis.text.x=element_text(angle=45, hjust=1)) +
  facet_wrap(~type, ncol=2, scales='free_y')
p + geom_area(aes(fill=country)) +
  labs(title=paste0('Cases around the World (excl. China) - ', max.date.txt))
```

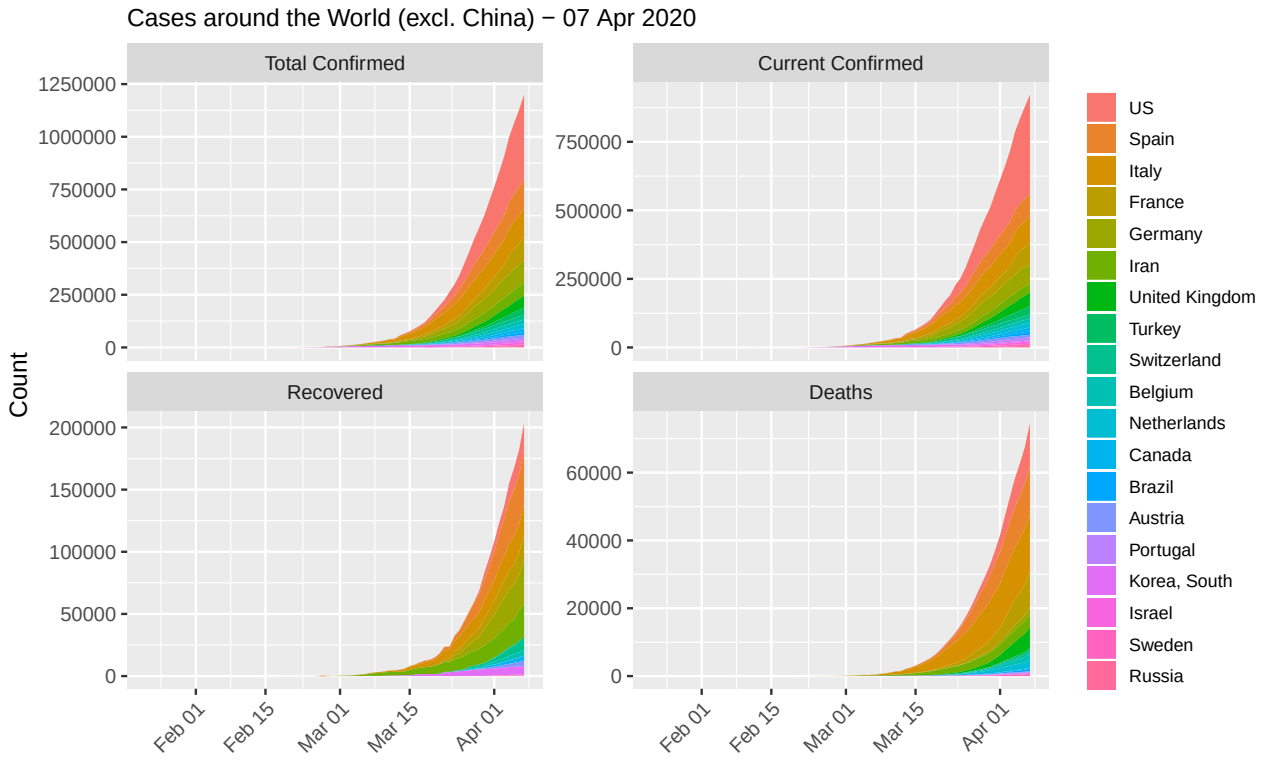


Figure 9: Cases around the World (excl. China)

```
## if Australia in not in top 20, add it in and remove 'Others'
if(!('Australia' %in% top.countries)) {
  top.countries %<>% setdiff('Others') %>% c('Australia')
  df <- data.long %>% filter(country %in% top.countries) %<>%
    mutate(country=country %>% factor(levels=c(top.countries)))
}

## cases by country - area plot
df %>% filter(country != 'World' & type != 'Total Confirmed') %>%
  ggplot(aes(x=date, y=count, fill=type)) +
  geom_area(alpha=0.5) +
  # xlab('') + ylab('') +
  labs(title=paste0('Numbers of COVID-19 Cases in Top 20 Countries - ',
                    max.date.txt)) +
  scale_fill_manual(values=c('red', 'green', 'black')) +
  theme(legend.title=element_blank(), legend.position='bottom',
```

```
      plot.title = element_text(size=12),
      axis.title.x=element_blank(),
      axis.title.y=element_blank(),
      legend.key.size=unit(0.4, 'cm'),
      # legend.text=element_text(size=7),
      strip.text.x=element_text(size=7),
      axis.text=element_text(size=7),
      axis.text.x=element_text(angle=45, hjust=1)) +
facet_wrap(~country, ncol=4, scales='free_y')
```

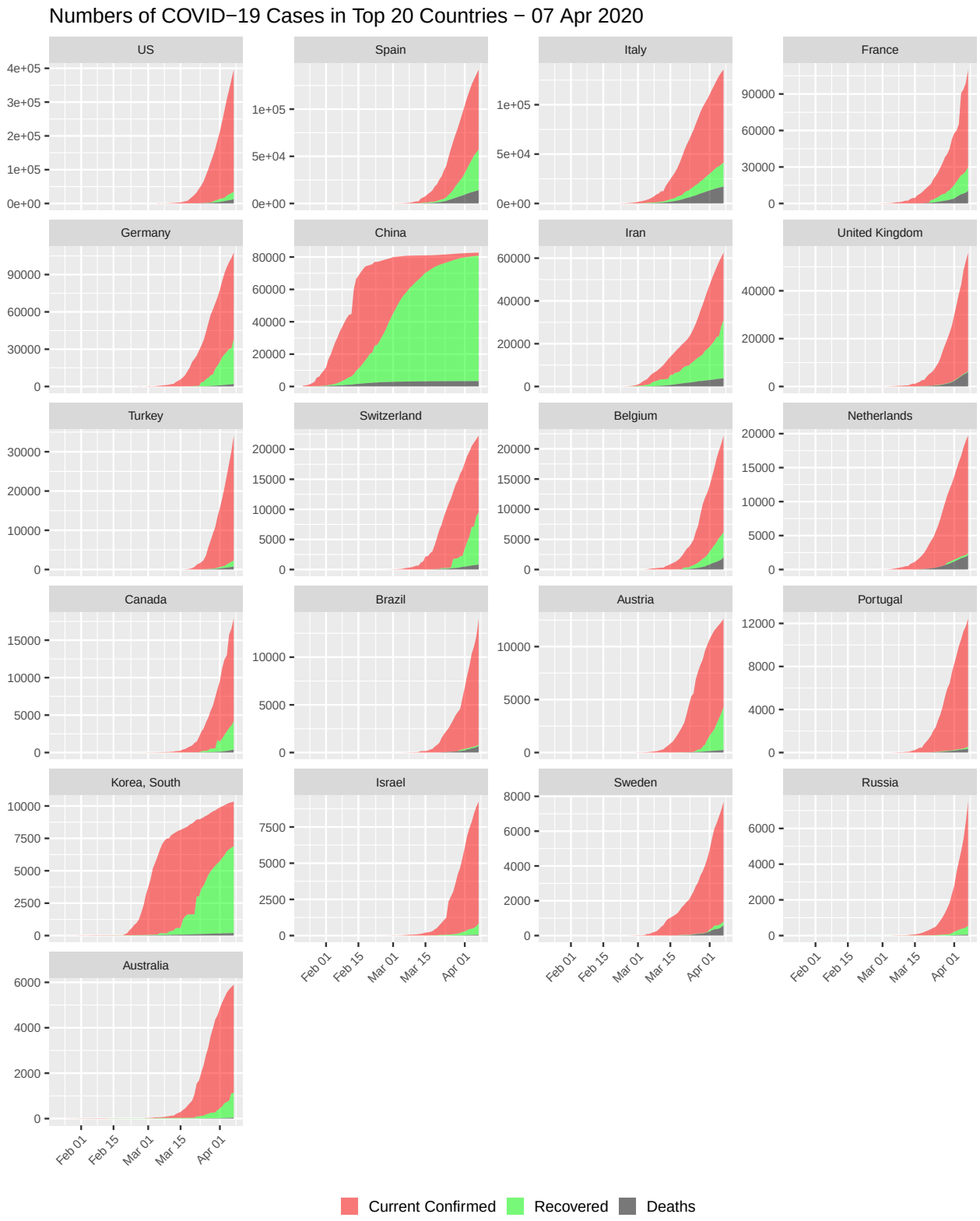


Figure 10: COVID-19 Cases in Top 20 Countries. Ordered descendingly by number of confirmed cases.

```r
## cases by country - line plot - log scale
p <- df %>% filter(country != 'World') %>%
  ggplot(aes(x=date, y=count, color=type)) +
  geom_line() +
  labs(title=paste0('Numbers of COVID-19 Cases in Top 20 Countries (log scale) - ',
                    max.date.txt)) +
  scale_color_manual(values=c('purple', 'red', 'green', 'black')) +
  theme(legend.title=element_blank(), legend.position='bottom',
        plot.title = element_text(size=12),
        axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        legend.key.size=unit(0.4, 'cm'),
        # legend.text=element_text(size=7),
        strip.text.x=element_text(size=7),
        axis.text=element_text(size=7),
        axis.text.x=element_text(angle=45, hjust=1)) +
  scale_y_continuous(trans='log10')
p + facet_wrap(~country, ncol=4, scales='free_y')
```

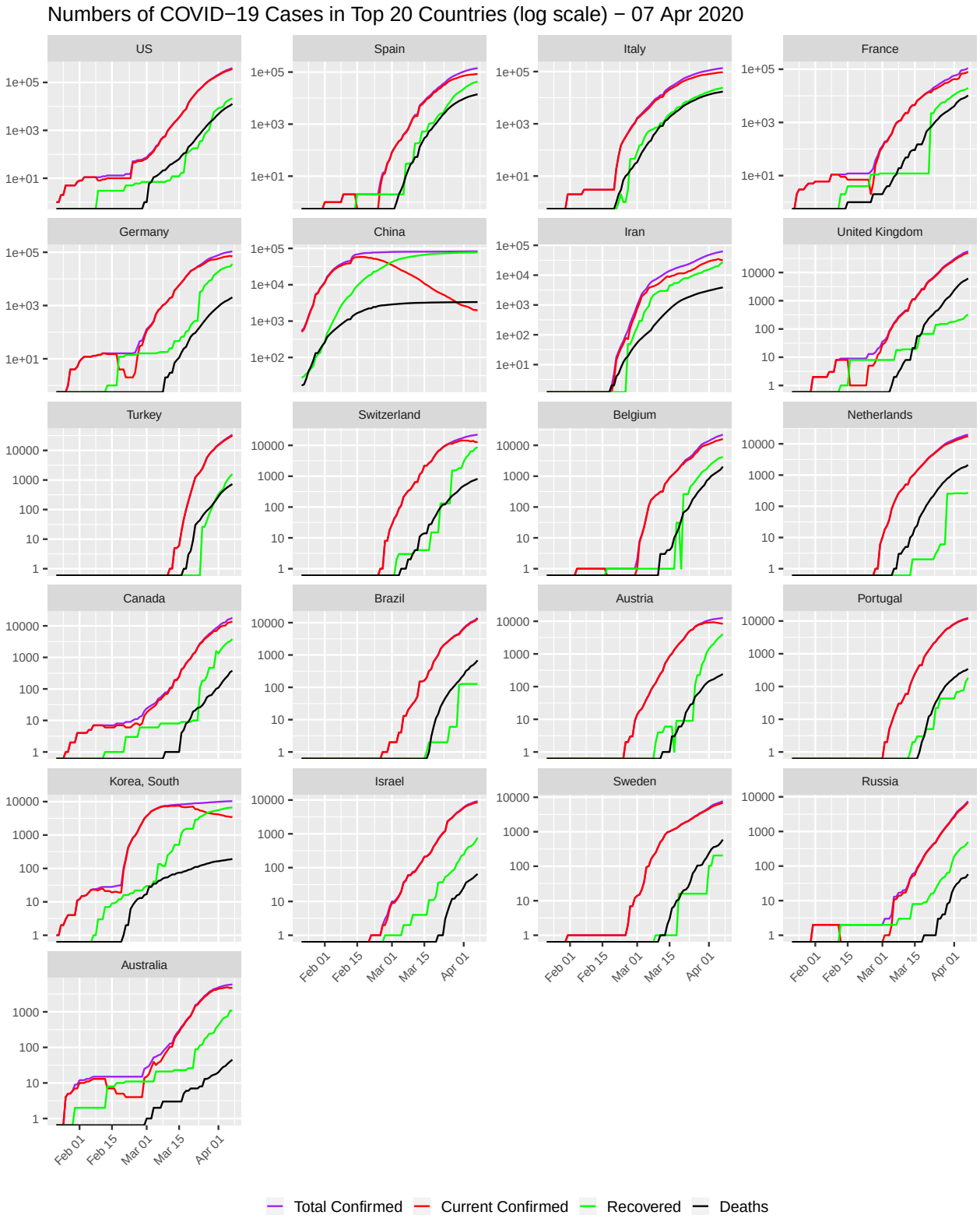


Figure 11: COVID-19 Cases Top 20 Countries (log scale). Ordered descendingly by number of confirmed cases.

```
## plot over multiple pages
# p + facet_wrap_paginate(~country, nrow=4, ncol=3, page=1, scales='free_y')
# p + facet_wrap_paginate(~country, nrow=4, ncol=3, page=2, scales='free_y')
```

Figures 10 and 11 show that the coronavirus seems to be under control in China, with an increase of recovered cases (in green) every day and a shrinking of the current confrimed cases (in red). However, in the rest of the world (i.e., outside of China), the confirmed cases are surging up in many other countries, which suggests that the virus has broken out there.

5.3 Death Rates

```
## three death rates
rate.max <- rates.long$count %>% max(na.rm=T)
df <- rates.long %>% filter(country %in% setdiff(top.countries, 'World')) %>%
  mutate(country=factor(country, levels=top.countries))
df %>% ggplot(aes(x=date, y=count, color=type)) +
  geom_line() +
  xlab('') + ylab('Death Rate (%)') +
  theme(legend.position='bottom', legend.title=element_blank(),
        legend.text=element_text(size=8),
        legend.key.size=unit(0.5, 'cm'),
        axis.text.x=element_text(angle=45, hjust=1)) +
  ylim(c(0, 100)) +
  facet_wrap(~country, ncol=4)
```

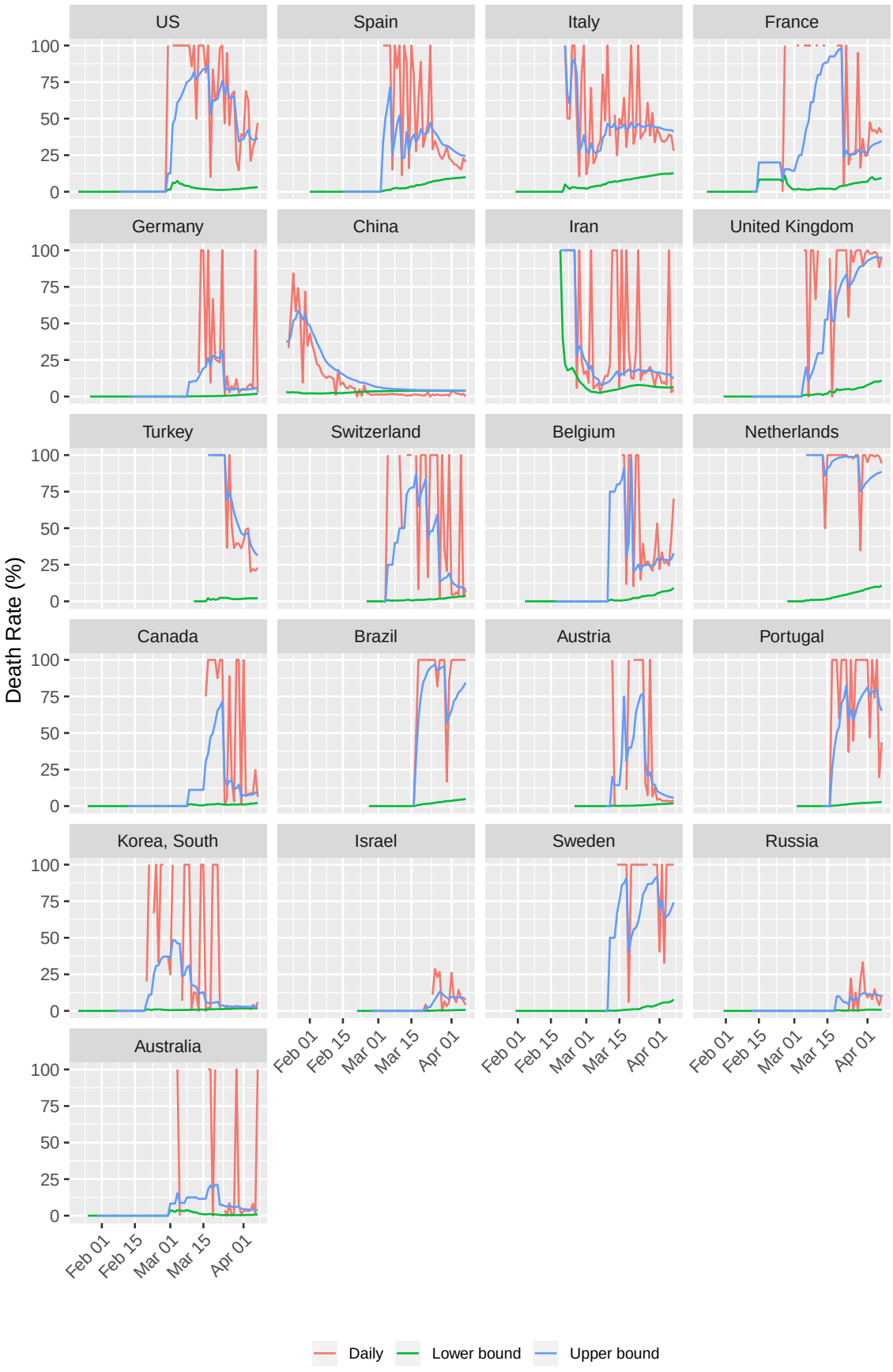


Figure 12: Death Rates

5.4 Countries with Highest Death Rates

Below are a list of top 20 countries with the highest death rates out of countires having 1000+ confirmed cases.

```
## sort the latest data by death rate, and if tie, by confirmed
df <- data %>% filter(date == max(date) & country != 'World' & confirmed >= 1000) %>%
  select(country, confirmed, new.confirmed, current.confirmed,
         recovered, deaths, new.deaths, death.rate=rate.lower) %>%
  arrange(desc(death.rate, confirmed))

df %>% head(20) %>%
  mutate(death.rate=death.rate %>% format(nsmall=1) %>% paste0('%')) %>%
  kable('latex', booktabs=T, row.names=T, align=c('l', rep('r', 7)),
        caption=paste0('Top 20 Countries with Highest Death Rates - ', max.date.txt),
        format.args=list(big.mark=',')) %>%
  kable_styling(font_size=7, latex_options=c('striped', 'hold_position', 'repeat_header'))
```

Table 4: Top 20 Countries with Highest Death Rates - 07 Apr 2020

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	Algeria	1,468	45	1,162	113	193	20	13.1%
2	Italy	135,586	3,039	94,067	24,392	17,127	604	12.6%
3	United Kingdom	55,949	3,670	49,453	325	6,171	786	11.0%
4	Netherlands	19,709	783	17,329	272	2,108	234	10.7%
5	Spain	141,942	5,267	84,689	43,208	14,045	704	9.9%
6	France	110,065	11,102	80,199	19,523	10,343	1,417	9.4%
7	Belgium	22,194	1,380	16,002	4,157	2,035	403	9.2%
8	Indonesia	2,738	247	2,313	204	221	12	8.1%
9	Sweden	7,693	487	6,897	205	591	114	7.7%
10	Morocco	1,184	64	1,001	93	90	10	7.6%
11	Egypt	1,450	128	1,080	276	94	9	6.5%
12	Iran	62,589	2,089	31,678	27,039	3,872	133	6.2%
13	Iraq	1,122	91	684	373	65	1	5.8%
14	Ecuador	3,747	0	3,456	100	191	0	5.1%
15	Mexico	2,439	296	1,681	633	125	31	5.1%
16	Dominican Republic	1,956	128	1,822	36	98	12	5.0%
17	Brazil	14,034	1,873	13,221	127	686	122	4.9%
18	Philippines	3,764	104	3,503	84	177	14	4.7%
19	Romania	4,417	360	3,760	460	197	21	4.5%
20	Greece	1,832	77	1,482	269	81	2	4.4%

6 Conclusions

As of 07 Apr 2020, there are 184 countries with confirmed COVID-19 cases. It seems to be contained in China, but starts to break out in rest of the world. The current death rate is in between 5.7% and 21.4%, but it is likely to change dramatically with the breakout in many countries, such as European countries.

Appendix A. Processed Data

Blow is the processed data for this analysis.

Appendix A.1 COVID-19 Cases Worldwide

```
## sort by date descendingly and re-order columns
data.world %<>% arrange(desc(date)) %>%
```

```
  select(c(date, confirmed, deaths, recovered, current.confirmed,
           new.confirmed, new.deaths, new.recovered, rate.lower, rate.upper, rate.daily))
## output as a table
data.world %>% kable('latex', booktabs=T, longtable=T, caption='Cases in the Whole World',
               format.args=list(big.mark=',')) %>%
    kable_styling(font_size=4, latex_options=c('striped', 'hold_position', 'repeat_header'))
```

Table 5: Cases in the Whole World

date	confirmed	deaths	recovered	current.confirmed	new.confirmed	new.deaths	new.recovered	rate.lower	rate.upper	rate.daily
2020-04-07	1,426,096	81,865	300,054	1,044,177	80,995	7,300	23,539	5.7	21.4	23.7
2020-04-06	1,345,101	74,565	276,515	994,021	72,986	5,191	16,503	5.5	21.2	23.9
2020-04-05	1,272,115	69,374	260,012	942,729	74,710	4,768	13,860	5.5	21.1	25.6
2020-04-04	1,197,405	64,606	246,152	886,647	101,488	5,819	20,356	5.4	20.8	22.2
2020-04-03	1,095,917	58,787	225,796	811,334	82,597	5,804	15,533	5.4	20.7	27.2
2020-04-02	1,013,320	52,983	210,263	750,074	80,715	6,174	17,086	5.2	20.1	26.5
2020-04-01	932,605	46,809	193,177	692,619	75,118	4,702	15,143	5.0	19.5	23.7
2020-03-31	857,487	42,107	178,034	637,346	75,092	4,525	13,468	4.9	19.1	25.1
2020-03-30	782,395	37,582	164,566	580,247	62,255	3,657	15,484	4.8	18.6	19.1
2020-03-29	720,140	33,925	149,082	537,133	59,434	3,273	9,667	4.7	18.5	25.3
2020-03-28	660,706	30,652	139,415	490,639	67,415	3,454	8,500	4.6	18.0	28.9
2020-03-27	593,291	27,198	130,915	435,178	63,700	3,228	8,765	4.6	17.2	26.9
2020-03-26	529,591	23,970	122,150	383,471	61,938	2,789	8,363	4.5	16.4	25.0
2020-03-25	467,653	21,181	113,787	332,685	49,608	2,556	5,787	4.5	15.7	30.6
2020-03-24	418,045	18,625	108,000	291,420	39,810	2,120	9,649	4.5	14.7	18.0
2020-03-23	378,235	16,505	98,351	263,379	41,282	1,854	452	4.4	14.4	80.4
2020-03-22	336,953	14,651	97,899	224,403	32,446	1,678	6,207	4.3	13.0	21.3
2020-03-21	304,507	12,973	91,692	199,842	32,299	1,674	4,272	4.3	12.4	28.2
2020-03-20	272,208	11,299	87,420	173,489	29,638	1,432	2,445	4.2	11.4	36.9
2020-03-19	242,570	9,867	84,975	147,728	27,749	1,134	1,663	4.1	10.4	40.5
2020-03-18	214,821	8,733	83,312	122,776	17,719	828	2,472	4.1	9.5	25.1
2020-03-17	197,102	7,905	80,840	108,357	15,528	779	2,752	4.0	8.9	22.1
2020-03-16	181,574	7,126	78,088	96,360	14,120	686	2,054	3.9	8.4	25.0
2020-03-15	167,454	6,440	76,034	84,980	11,353	621	3,410	3.8	7.8	15.4
2020-03-14	156,101	5,819	72,624	77,658	10,896	415	2,373	3.7	7.4	14.9
2020-03-13	145,205	5,404	70,251	69,550	16,853	684	1,927	3.7	7.1	26.2
2020-03-12	128,352	4,720	68,324	55,308	2,477	105	1,321	3.7	6.5	7.4
2020-03-11	125,875	4,615	67,003	54,257	7,255	353	2,599	3.7	6.4	12.0
2020-03-10	118,620	4,262	64,404	49,954	5,030	274	1,910	3.6	6.2	12.5
2020-03-09	113,590	3,988	62,494	47,108	3,769	186	1,800	3.5	6.0	9.4
2020-03-08	109,821	3,802	60,694	45,325	3,974	244	2,336	3.5	5.9	9.5
2020-03-07	105,847	3,558	58,358	43,931	4,046	98	2,493	3.4	5.7	3.8
2020-03-06	101,801	3,460	55,865	42,476	3,915	112	2,069	3.4	5.8	5.1
2020-03-05	97,886	3,348	53,796	40,742	2,766	94	2,626	3.4	5.9	3.5
2020-03-04	95,120	3,254	51,170	40,696	2,280	94	2,942	3.4	6.0	3.1
2020-03-03	92,840	3,160	48,228	41,452	2,534	75	2,626	3.4	6.1	2.8
2020-03-02	90,306	3,085	45,602	41,619	1,937	89	2,886	3.4	6.3	3.0
2020-03-01	88,369	2,996	42,716	42,657	2,358	55	2,934	3.4	6.6	1.8
2020-02-29	86,011	2,941	39,782	43,288	1,899	69	3,071	3.4	6.9	2.2
2020-02-28	84,112	2,872	36,711	44,529	1,366	58	3,434	3.4	7.3	1.7
2020-02-27	82,746	2,814	33,277	46,655	1,358	44	2,893	3.4	7.8	1.5
2020-02-26	81,388	2,770	30,384	48,234	982	62	2,479	3.4	8.4	2.4
2020-02-25	80,406	2,708	27,905	49,793	845	79	2,678	3.4	8.8	2.9
2020-02-24	79,561	2,629	25,227	51,705	603	160	1,833	3.3	9.4	8.0
2020-02-23	78,958	2,469	23,394	53,095	386	11	508	3.1	9.5	2.1
2020-02-22	78,572	2,458	22,886	53,228	1,753	207	3,996	3.1	9.7	4.9
2020-02-21	76,819	2,251	18,890	55,678	622	4	713	2.9	10.6	0.6
2020-02-20	76,197	2,247	18,177	55,773	558	125	2,056	2.9	11.0	5.7
2020-02-19	75,639	2,122	16,121	57,396	503	115	1,769	2.8	11.6	6.1
2020-02-18	75,136	2,007	14,352	58,777	1,878	139	1,769	2.7	12.3	7.3
2020-02-17	73,258	1,868	12,583	58,807	2,034	98	1,718	2.5	12.9	5.4
2020-02-16	71,224	1,770	10,865	58,589	2,194	104	1,470	2.5	14.0	6.6
2020-02-15	69,030	1,666	9,395	57,969	2,145	143	1,337	2.4	15.1	9.7
2020-02-14	66,885	1,523	8,058	57,304	6,517	152	1,763	2.3	15.9	7.9
2020-02-13	60,368	1,371	6,295	52,702	15,147	253	1,145	2.3	17.9	18.1
2020-02-12	45,221	1,118	5,150	38,953	419	5	467	2.5	17.8	1.1
2020-02-11	44,802	1,113	4,683	39,006	2,040	100	737	2.5	19.2	11.9
2020-02-10	42,762	1,013	3,946	37,803	2,612	107	702	2.4	20.4	13.2
2020-02-09	40,150	906	3,244	36,000	3,030	100	628	2.3	21.8	13.7
2020-02-08	37,120	806	2,616	33,698	2,729	87	605	2.2	23.6	12.6
2020-02-07	34,391	719	2,011	31,661	3,597	85	524	2.1	26.3	14.0
2020-02-06	30,794	634	1,487	28,673	3,159	70	363	2.1	29.9	16.2
2020-02-05	27,635	564	1,124	25,947	3,743	72	272	2.0	33.4	20.9
2020-02-04	23,892	492	852	22,548	4,011	66	229	2.1	36.6	22.4
2020-02-03	19,881	426	623	18,832	3,094	64	151	2.1	40.6	29.8
2020-02-02	16,787	362	472	15,953	4,749	103	188	2.2	43.4	35.4
2020-02-01	12,038	259	284	11,495	2,111	46	62	2.2	47.7	42.6
2020-01-31	9,927	213	222	9,492	1,693	42	79	2.1	49.0	34.7
2020-01-30	8,234	171	143	7,920	2,068	38	17	2.1	54.5	69.1
2020-01-29	6,166	133	126	5,907	588	2	19	2.2	51.4	9.5
2020-01-28	5,578	131	107	5,340	2,651	49	46	2.3	55.0	51.6
2020-01-27	2,927	82	61	2,784	809	26	9	2.8	57.3	74.3
2020-01-26	2,118	56	52	2,010	684	14	13	2.6	51.9	51.9
2020-01-25	1,434	42	39	1,353	493	16	3	2.9	51.9	84.2
2020-01-24	941	26	36	879	287	8	6	2.8	41.9	57.1
2020-01-23	654	18	30	606	99	1	2	2.8	37.5	33.3

Table 5: Cases in the Whole World *(continued)*

date	confirmed	deaths	recovered	current.confirmed	new.confirmed	new.deaths	new.recovered	rate.lower	rate.upper	rate.daily
2020-01-22	555	17	28	510				3.1	37.8	

Appendix A.2 Latest Cases by Country

```
## hightlight high death rates (if >= 5%) for those countries with 1000+ confirmed cases
data.latest.all %>% arrange(desc(confirmed)) %>% select(-c(date, ranking)) %>%
  mutate(death.rate = cell_spec(death.rate, "latex",
                                color = ifelse(confirmed >= 1000 & death.rate >= 5, "red", "black"),
                                bold = ifelse(confirmed >= 1000 & death.rate >= 5, T, F))) %>%
  kable(format='latex', escape=F, booktabs=T, longtable=T, row.names=T,
        caption=paste0('Cases by Country (', max.date.txt, ')'),
        format.args=list(big.mark=','), align=c('l', rep('r', 7))) %>%
  kable_styling(font_size=6, latex_options=c('striped', 'hold_position', 'repeat_header'))
```

Table 6: Cases by Country (07 Apr 2020)

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	World	1,426,096	80,995	1,044,177	300,054	81,865	7,300	**5.7**
2	US	396,223	29,556	361,738	21,763	12,722	1,939	3.2
3	Spain	141,942	5,267	84,689	43,208	14,045	704	**9.9**
4	Italy	135,586	3,039	94,067	24,392	17,127	604	**12.6**
5	France	110,065	11,102	80,199	19,523	10,343	1,417	**9.4**
6	Germany	107,663	4,289	69,566	36,081	2,016	206	1.9
7	China	82,718	53	1,973	77,410	3,335	0	4
8	Iran	62,589	2,089	31,678	27,039	3,872	133	**6.2**
9	United Kingdom	55,949	3,670	49,453	325	6,171	786	**11**
10	Turkey	34,109	3,892	31,802	1,582	725	76	2.1
11	Switzerland	22,253	596	12,728	8,704	821	56	3.7
12	Belgium	22,194	1,380	16,002	4,157	2,035	403	**9.2**
13	Netherlands	19,709	783	17,329	272	2,108	234	**10.7**
14	Canada	17,872	1,309	13,706	3,791	375	36	2.1
15	Brazil	14,034	1,873	13,221	127	686	122	4.9
16	Austria	12,639	342	8,350	4,046	243	23	1.9
17	Portugal	12,442	712	11,913	184	345	34	2.8
18	Korea, South	10,331	47	3,445	6,694	192	6	1.9
19	Israel	9,248	344	8,413	770	65	8	0.7
20	Sweden	7,693	487	6,897	205	591	114	**7.7**
21	Russia	7,497	1,154	6,945	494	58	11	0.8
22	Norway	6,086	221	5,965	32	89	13	1.5
23	Australia	5,895	98	4,770	1,080	45	5	0.8
24	Ireland	5,709	345	5,474	25	210	36	3.7
25	India	5,311	533	4,740	421	150	14	2.8
26	Denmark	5,266	391	3,442	1,621	203	16	3.9
27	Chile	5,116	301	4,175	898	43	6	0.8
28	Czechia	5,017	195	4,757	172	88	10	1.8
29	Poland	4,848	435	4,528	191	129	22	2.7
30	Romania	4,417	360	3,760	460	197	21	4.5
31	Pakistan	4,035	269	3,549	429	57	4	1.4
32	Malaysia	3,963	170	2,579	1,321	63	1	1.6
33	Japan	3,906	252	3,222	592	92	7	2.4
34	Philippines	3,764	104	3,503	84	177	14	4.7
35	Ecuador	3,747	0	3,456	100	191	0	**5.1**
36	Luxembourg	2,970	127	2,426	500	44	3	1.5
37	Peru	2,954	393	1,546	1,301	107	15	3.6
38	Saudi Arabia	2,795	190	2,139	615	41	3	1.5
39	Indonesia	2,738	247	2,313	204	221	12	**8.1**
40	Serbia	2,447	247	2,386	0	61	3	2.5
41	Mexico	2,439	296	1,681	633	125	31	**5.1**
42	United Arab Emirates	2,359	283	2,161	186	12	1	0.5
43	Finland	2,308	132	1,974	300	34	7	1.5
44	Thailand	2,258	38	1,343	888	27	1	1.2
45	Panama	2,100	112	2,031	14	55	1	2.6
46	Qatar	2,057	225	1,901	150	6	2	0.3

Table 6: Cases by Country (07 Apr 2020) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
47	Dominican Republic	1,956	128	1,822	36	98	12	**5**
48	Greece	1,832	77	1,482	269	81	2	4.4
49	Colombia	1,780	201	1,630	100	50	4	2.8
50	South Africa	1,749	63	1,641	95	13	1	0.7
51	Argentina	1,628	74	1,234	338	56	8	3.4
52	Iceland	1,586	24	1,021	559	6	0	0.4
53	Singapore	1,481	106	1,098	377	6	0	0.4
54	Algeria	1,468	45	1,162	113	193	20	**13.1**
55	Ukraine	1,462	143	1,389	28	45	7	3.1
56	Egypt	1,450	128	1,080	276	94	9	**6.5**
57	Croatia	1,282	60	1,097	167	18	2	1.4
58	Morocco	1,184	64	1,001	93	90	10	**7.6**
59	New Zealand	1,160	54	918	241	1	0	0.1
60	Estonia	1,149	41	1,059	69	21	2	1.8
61	Iraq	1,122	91	684	373	65	1	**5.8**
62	Slovenia	1,059	38	921	102	36	6	3.4
63	Moldova	1,056	91	994	40	22	3	2.1
64	Lithuania	880	37	857	8	15	0	1.7
65	Belarus	861	161	794	54	13	0	1.5
66	Armenia	853	20	758	87	8	0	0.9
67	Hungary	817	73	699	71	47	9	5.8
68	Bahrain	811	55	348	458	5	1	0.6
69	Bosnia and Herzegovina	764	90	663	68	33	4	4.3
70	Kuwait	743	78	637	105	1	0	0.1
71	Azerbaijan	717	76	665	44	8	1	1.1
72	Diamond Princess	712	0	82	619	11	0	1.5
73	Kazakhstan	697	35	640	51	6	0	0.9
74	Cameroon	658	0	606	43	9	0	1.4
75	Tunisia	623	27	575	25	23	1	3.7
76	North Macedonia	599	29	543	30	26	3	4.3
77	Slovakia	581	47	566	13	2	0	0.3
78	Bulgaria	577	28	512	42	23	1	4
79	Latvia	548	6	530	16	2	1	0.4
80	Lebanon	548	7	467	62	19	0	3.5
81	Andorra	545	20	484	39	22	1	4
82	Uzbekistan	520	63	488	30	2	0	0.4
83	Cyprus	494	29	438	47	9	0	1.8
84	Costa Rica	483	16	457	24	2	0	0.4
85	Uruguay	424	18	267	150	7	1	1.7
86	Afghanistan	423	56	391	18	14	3	3.3
87	Cuba	396	46	358	27	11	2	2.8
88	Burkina Faso	384	20	238	127	19	1	4.9
89	Albania	383	6	230	131	22	1	5.7
90	Taiwan*	376	3	314	57	5	0	1.3
91	Oman	371	40	302	67	2	0	0.5
92	Jordan	353	4	209	138	6	0	1.7
93	Cote d'Ivoire	349	26	305	41	3	0	0.9
94	Honduras	305	7	277	6	22	0	7.2
95	Malta	293	52	288	5	0	0	0
96	Ghana	287	73	251	31	5	0	1.7
97	San Marino	279	13	205	40	34	2	12.2
98	Niger	278	25	241	26	11	1	4
99	Mauritius	268	24	253	8	7	0	2.6
100	West Bank and Gaza	261	7	218	42	1	0	0.4
101	Nigeria	254	16	204	44	6	1	2.4
102	Vietnam	249	4	126	123	0	0	0
103	Montenegro	241	8	235	4	2	0	0.8
104	Senegal	237	11	130	105	2	0	0.8
105	Kyrgyzstan	228	12	191	33	4	0	1.8
106	Georgia	196	8	147	46	3	1	1.5
107	Bolivia	194	11	178	2	14	3	7.2
108	Sri Lanka	185	7	137	42	6	1	3.2
109	Congo (Kinshasa)	180	19	153	9	18	0	10
110	Kenya	172	14	159	7	6	0	3.5
111	Kosovo	170	25	142	24	4	3	2.4
112	Venezuela	165	0	93	65	7	0	4.2
113	Bangladesh	164	41	114	33	17	5	10.4
114	Guinea	144	16	139	5	0	0	0

Table 6: Cases by Country (07 Apr 2020) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
115	Brunei	135	0	49	85	1	0	0.7
116	Cambodia	115	1	57	58	0	0	0
117	Paraguay	115	2	95	15	5	0	4.3
118	Trinidad and Tobago	107	2	98	1	8	0	7.5
119	Rwanda	105	0	98	7	0	0	0
120	Djibouti	90	0	81	9	0	0	0
121	Madagascar	88	6	81	7	0	0	0
122	Monaco	79	2	74	4	1	0	1.3
123	El Salvador	78	9	69	5	4	0	5.1
124	Liechtenstein	78	1	22	55	1	0	1.3
125	Guatemala	77	7	57	17	3	0	3.9
126	Togo	65	7	39	23	3	0	4.6
127	Barbados	63	3	54	6	3	1	4.8
128	Jamaica	63	5	52	8	3	0	4.8
129	Mali	56	9	39	12	5	0	8.9
130	Ethiopia	52	8	46	4	2	0	3.8
131	Uganda	52	0	52	0	0	0	0
132	Congo (Brazzaville)	45	0	38	2	5	0	11.1
133	Zambia	39	0	31	7	1	0	2.6
134	Bahamas	33	4	22	5	6	1	18.2
135	Guinea-Bissau	33	15	33	0	0	0	0
136	Guyana	33	2	20	8	5	1	15.2
137	Eritrea	31	0	31	0	0	0	0
138	Gabon	30	6	28	1	1	0	3.3
139	Benin	26	0	20	5	1	0	3.8
140	Haiti	25	1	24	0	1	0	4
141	Tanzania	24	0	18	5	1	0	4.2
142	Burma	22	0	21	0	1	0	4.5
143	Libya	20	1	18	1	1	0	5
144	Antigua and Barbuda	19	4	18	0	1	1	5.3
145	Maldives	19	0	6	13	0	0	0
146	Syria	19	0	14	3	2	0	10.5
147	Angola	17	1	13	2	2	0	11.8
148	Equatorial Guinea	16	0	13	3	0	0	0
149	Namibia	16	0	13	3	0	0	0
150	Dominica	15	0	14	1	0	0	0
151	Fiji	15	1	15	0	0	0	0
152	Mongolia	15	0	11	4	0	0	0
153	Laos	14	2	14	0	0	0	0
154	Liberia	14	0	8	3	3	0	21.4
155	Saint Lucia	14	0	13	1	0	0	0
156	Sudan	14	2	10	2	2	0	14.3
157	Grenada	12	0	12	0	0	0	0
158	Saint Kitts and Nevis	11	1	11	0	0	0	0
159	Seychelles	11	0	11	0	0	0	0
160	Zimbabwe	11	1	9	0	2	1	18.2
161	Chad	10	1	8	2	0	0	0
162	Eswatini	10	0	6	4	0	0	0
163	Mozambique	10	0	9	1	0	0	0
164	Suriname	10	0	9	0	1	0	10
165	MS Zaandam	9	0	7	0	2	0	22.2
166	Nepal	9	0	8	1	0	0	0
167	Central African Republic	8	0	8	0	0	0	0
168	Malawi	8	3	7	0	1	1	12.5
169	Saint Vincent and the Grenadines	8	1	7	1	0	0	0
170	Somalia	8	1	7	1	0	0	0
171	Belize	7	0	6	0	1	0	14.3
172	Cabo Verde	7	0	5	1	1	0	14.3
173	Holy See	7	0	7	0	0	0	0
174	Botswana	6	0	5	0	1	0	16.7
175	Mauritania	6	0	3	2	1	0	16.7
176	Nicaragua	6	0	5	0	1	0	16.7
177	Sierra Leone	6	0	6	0	0	0	0
178	Bhutan	5	0	3	2	0	0	0
179	Gambia	4	0	1	2	1	0	25
180	Sao Tome and Principe	4	0	4	0	0	0	0
181	Western Sahara	4	0	4	0	0	0	0
182	Burundi	3	0	3	0	0	0	0

Table 6: Cases by Country (07 Apr 2020) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
183	Papua New Guinea	2	0	2	0	0	0	0
184	South Sudan	2	1	2	0	0	0	0
185	Timor-Leste	1	0	1	0	0	0	0

Appendix B. How to Cite This Work

Citation

Yanchang Zhao, COVID-19 Data Analysis with R – Worldwide. RDataMining.com, 2020. URL: http://www.rdatamining.com/docs/Coronavirus-data-analysis-world.pdf.

BibTex

@techreport{Zhao2020Covid19world,
Author = {Yanchang Zhao},
Institution = {RDataMining.com},
Title = {COVID-19 Data Analysis with R – Worldwide},
Url = {http://www.rdatamining.com/docs/Coronavirus-data-analysis-world.pdf},
Year = {2020}}

Appendix C. Contact

Contact:

Dr. Yanchang Zhao

Email: yanchang@RDataMining.com

Twitter: @RDataMining

LinkedIn: http://group.rdatamining.com

Comments and suggestions and welcome. Thanks!