

COVID-19 Data Analysis with R - Worldwide

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

31 March 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	30
	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 particularly 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

Below 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] 254 73
```

Each dataset has 254 rows, corresponding to country/region/province/state. It has 73 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" "2020-03-30"
```

```
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 30 Mar 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 datasets 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)
```

```

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 178 countries with confirmed COVID-19 cases, as of 30 Mar 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.

```

## 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 calculated 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 calculated 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 calculated 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 situation: 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
```

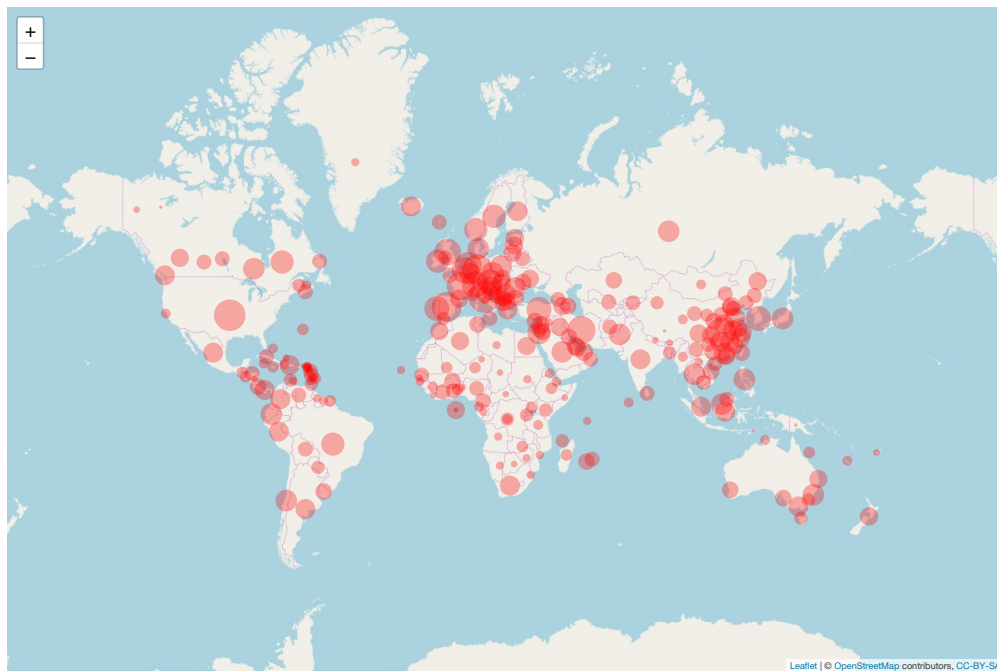


Figure 1: World Map

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 focus 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)
```

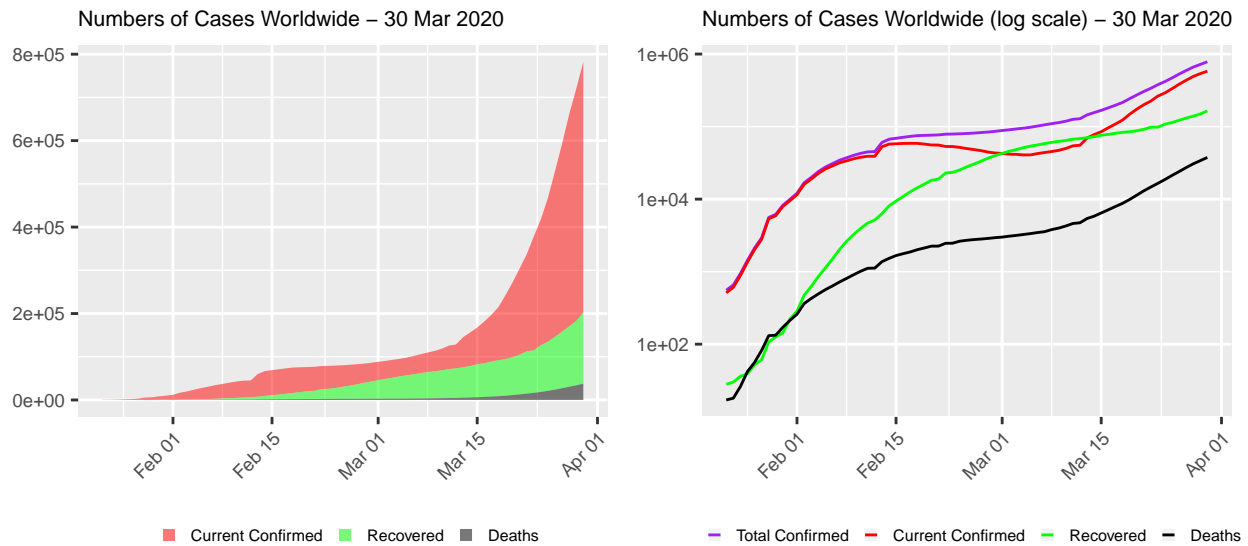


Figure 2: COVID-19 Cases Worldwide

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)
```

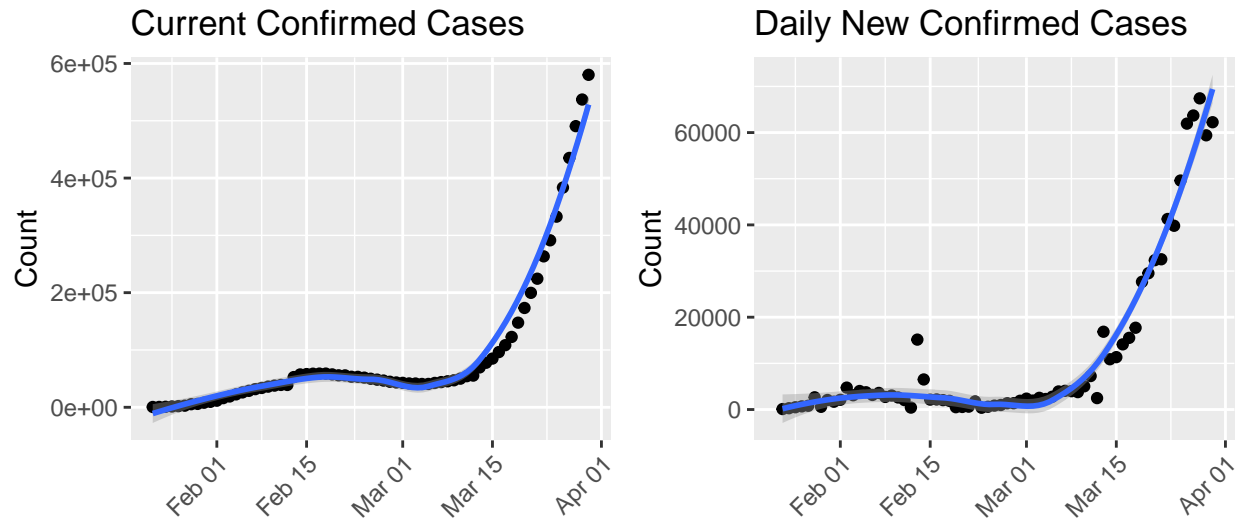



Figure 3: Current Confirmed Cases

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)
```

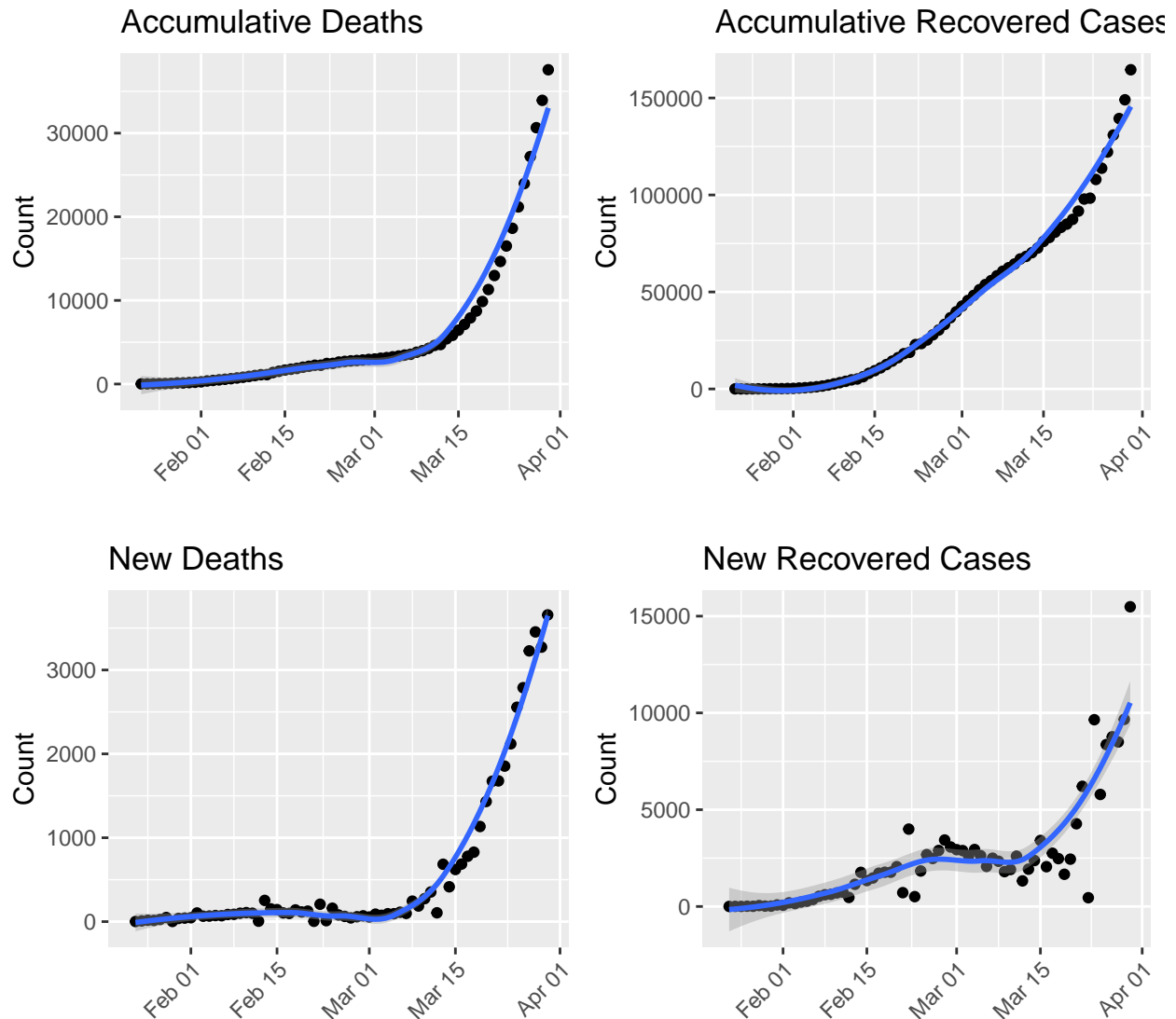


Figure 4: Deaths and Recovered Cases

4.5 Death Rates

Figure 5 shows death rates calculated in three different ways (see Section 3.3 for details). The left chart shows the death rates from 22 Jan 2020 to 30 Mar 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 30 Mar 2020 UTC, it will be between 4.8% and 18.6%.

A surge in the daily death rate (in red) suggests that the situation 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)

```

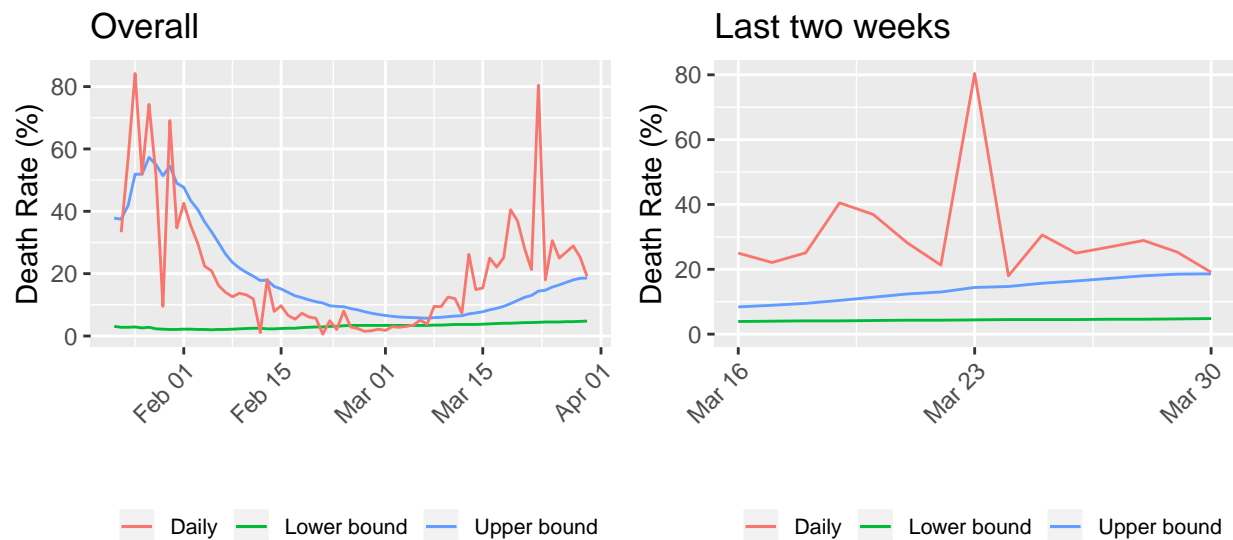


Figure 5: Death Rate

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

## 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 - 30 Mar 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	782,365	37,582	4.8%	62,248	3,657	580,217
2	US	161,807	2,978	1.8%	20,921	511	153,185
3	Italy	101,739	11,591	11.4%	4,050	812	75,528
4	Spain	87,956	7,716	8.8%	7,846	913	63,460
5	China	82,198	3,308	4.0%	76	4	2,967
6	Germany	66,885	645	1.0%	4,790	112	52,740
7	France	45,170	3,030	6.7%	4,462	419	34,176
8	Iran	41,495	2,757	6.6%	3,186	117	24,827
9	United Kingdom	22,453	1,411	6.3%	2,673	180	20,871
10	Switzerland	15,922	359	2.3%	1,093	59	13,740
11	Belgium	11,899	513	4.3%	1,063	82	9,859
12	Netherlands	11,817	865	7.3%	887	93	10,699
13	Turkey	10,827	168	1.6%	1,610	37	10,497
14	Korea, South	9,661	158	1.6%	78	6	4,275
15	Austria	9,618	108	1.1%	830	22	8,874
16	Canada	7,398	80	1.1%	1,118	16	6,852
17	Portugal	6,408	140	2.2%	446	21	6,225
18	Israel	4,695	16	0.3%	448	1	4,518
19	Brazil	4,579	159	3.5%	323	23	4,300
20	Norway	4,445	32	0.7%	161	7	4,401
21	Australia	4,361	17	0.4%	377	1	4,087
22	Others	71,032	1,531	2.2%	5,810	221	64,136

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

Top 20 Countries with Most Confirmed Cases – 30 Mar 2020

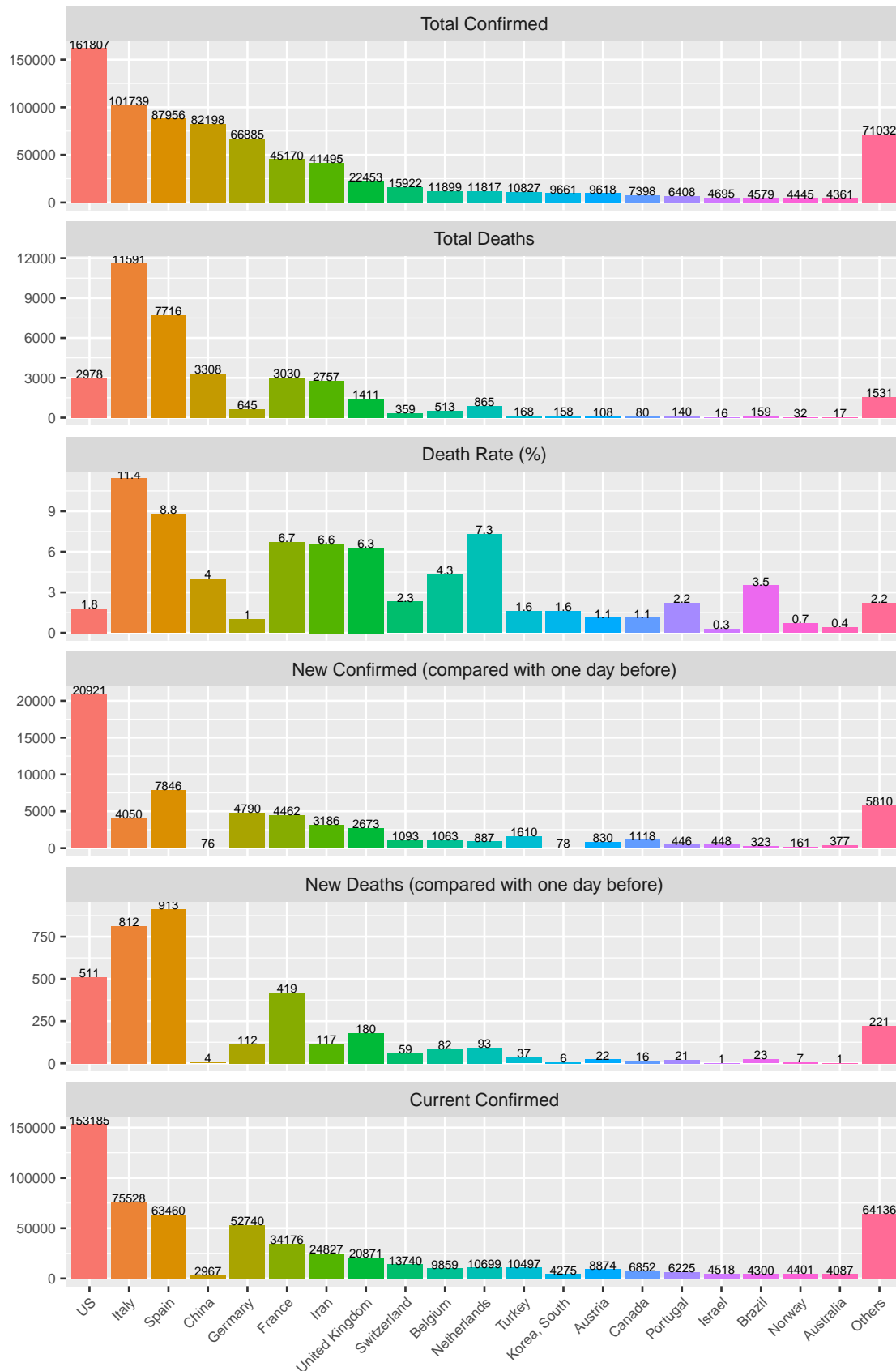
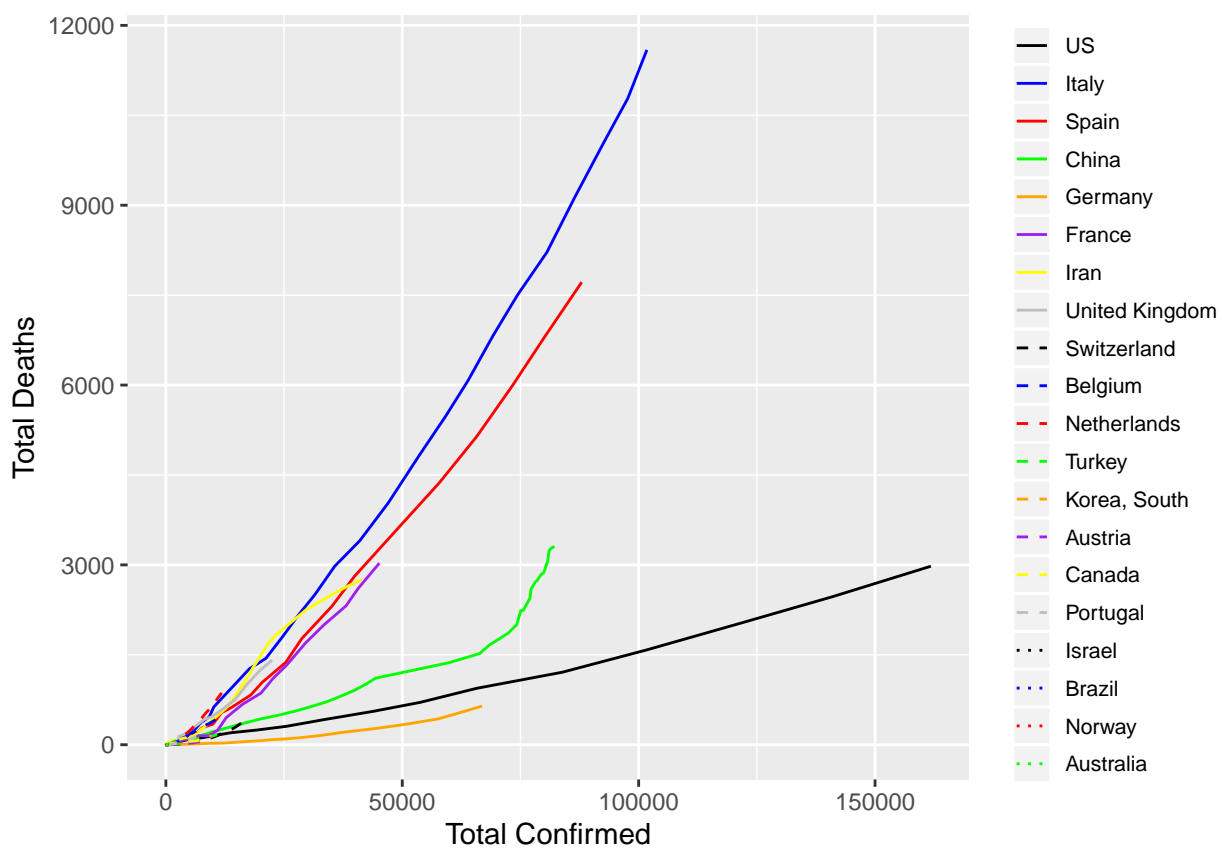


Figure 6: Top 20 Countries with Most Confirmed Cases

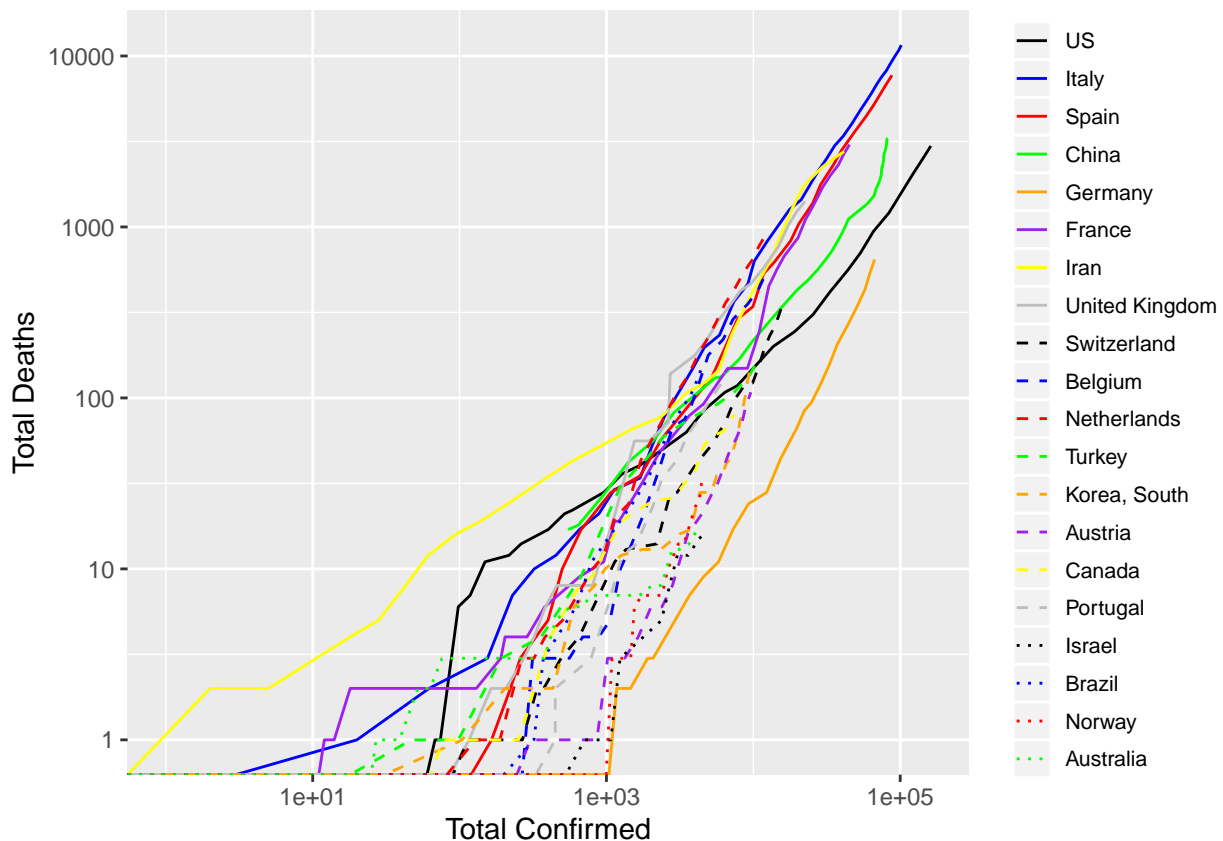
5.1 Confirmed vs Deaths

```
# 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 30 Mar 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)
```

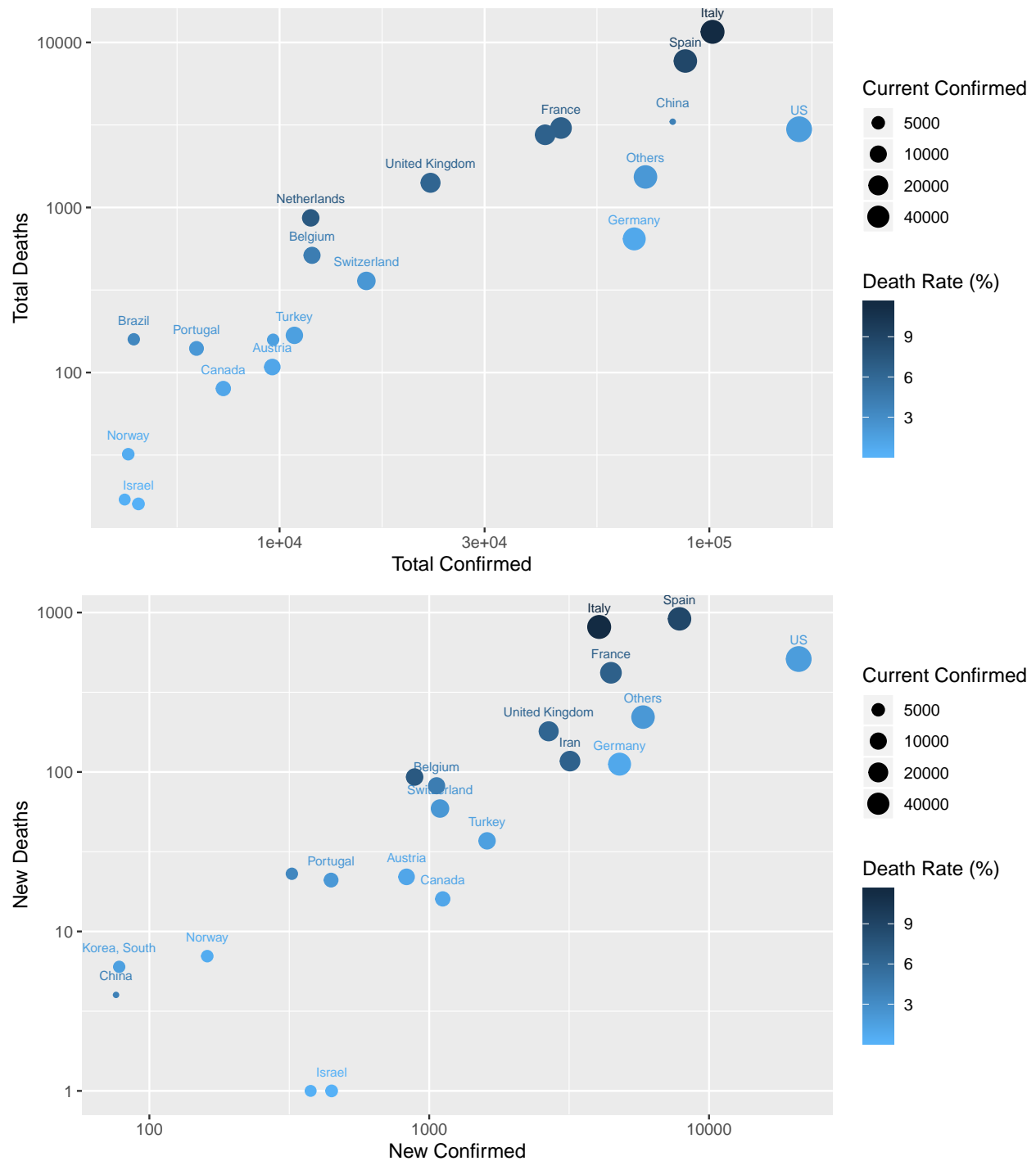



Figure 7: Top 20 Countries

5.2 Comparison across Countries

The area plots below show the numbers of dead, recovered, total and current confirmed 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.

```

## 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)

```

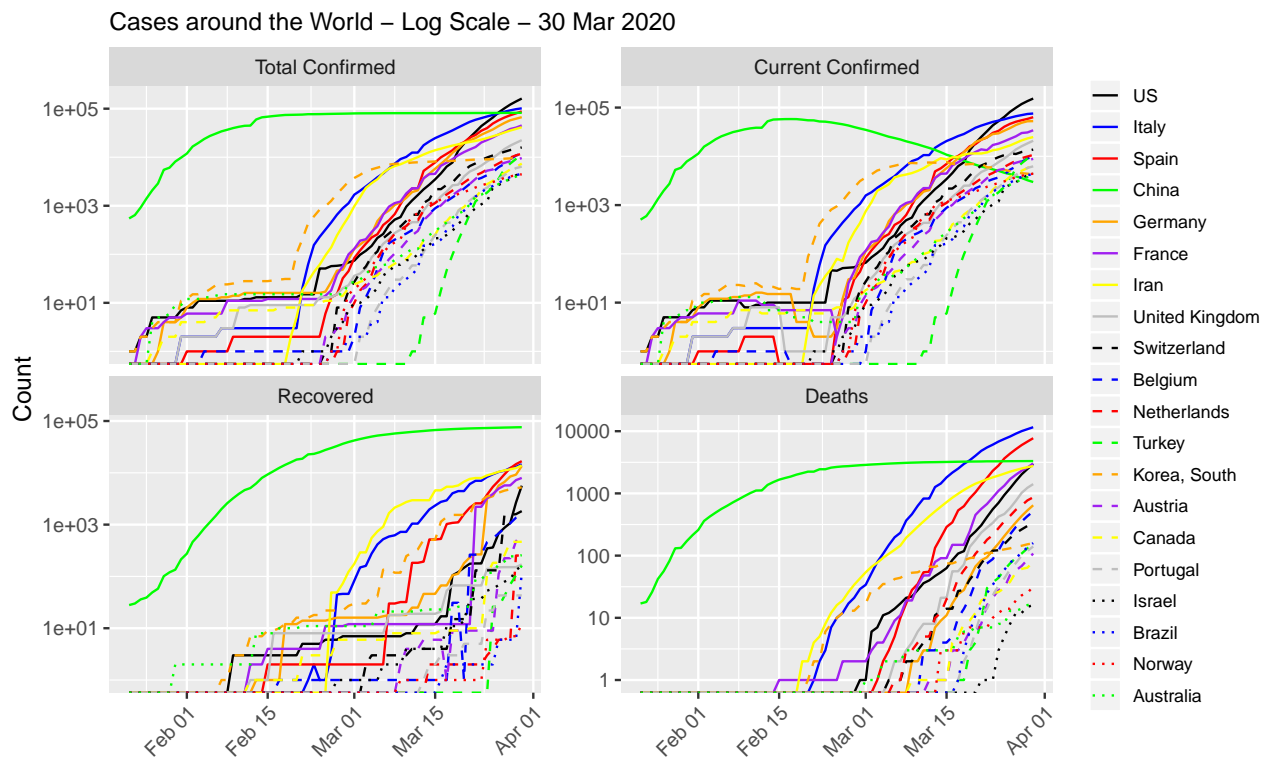
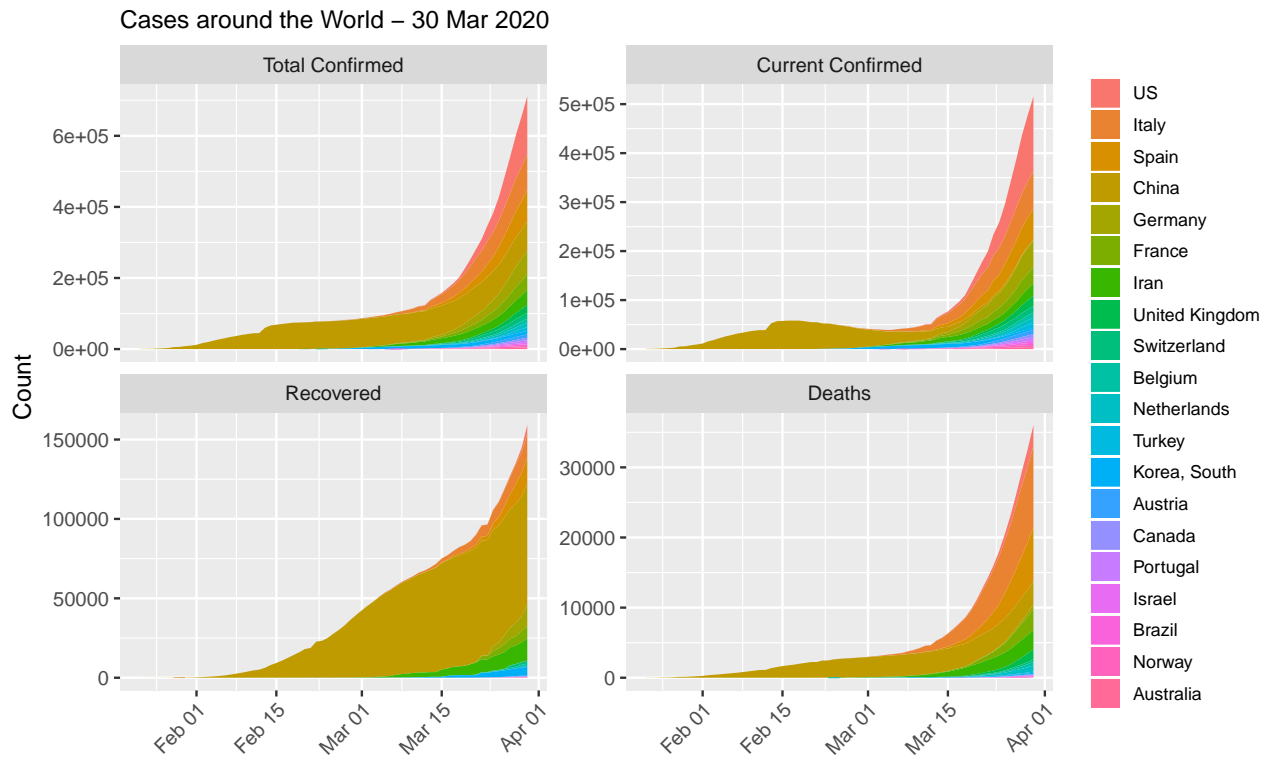


Figure 8: Cases around the World

```
## 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))

```

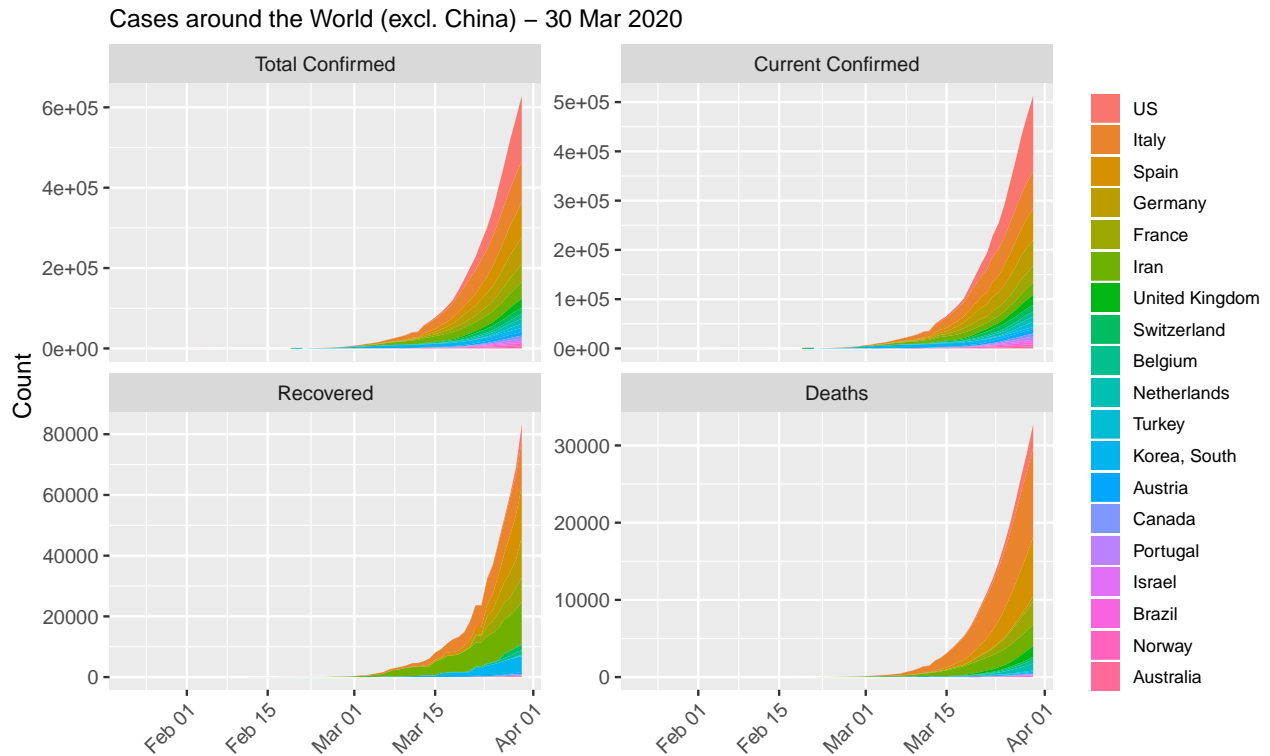


Figure 9: Cases around the World (excl. China)

```

## if Australia is 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')

```

Numbers of COVID-19 Cases in Top 20 Countries – 30 Mar 2020

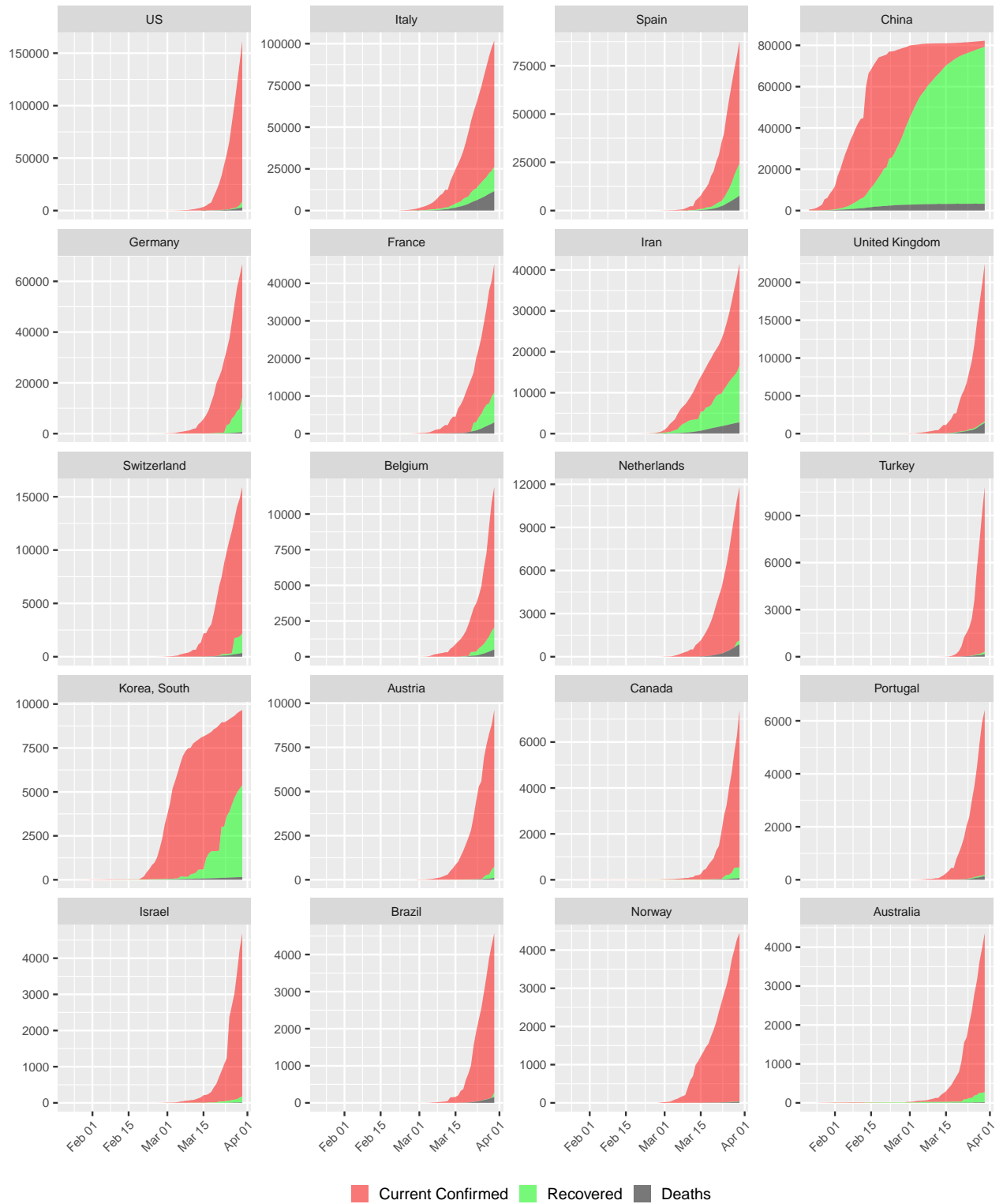


Figure 10: COVID-19 Cases in Top 20 Countries. Ordered descending by number of confirmed cases.

```

## 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')

```

Numbers of COVID-19 Cases in Top 20 Countries (log scale) – 30 Mar 2020

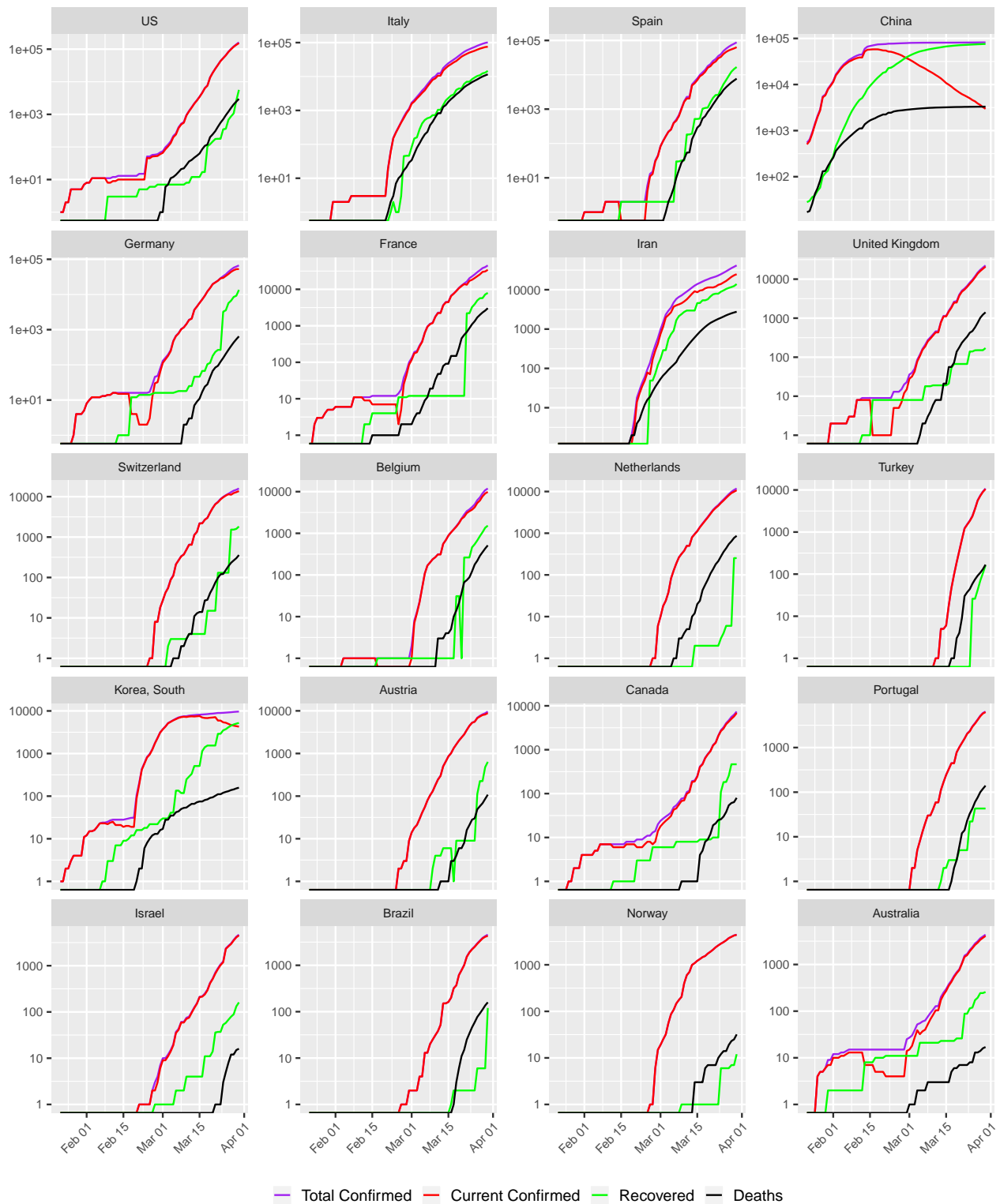


Figure 11: COVID-19 Cases Top 20 Countries (log scale). Ordered descendingly by number of confirmed cases.


```
## 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 confirmed 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)
```

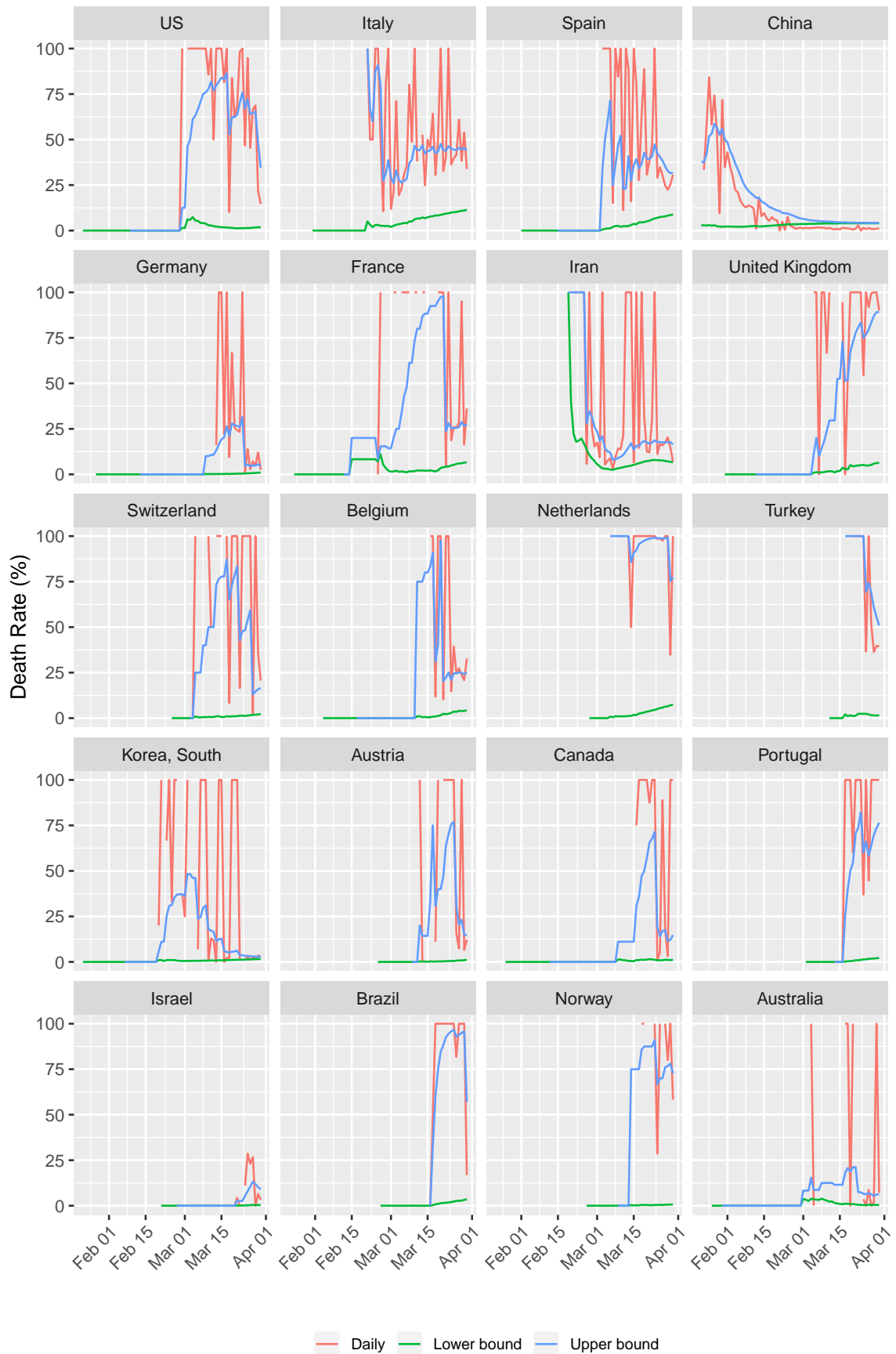


Figure 12: Death Rates
26

5.4 Countries with Highest Death Rates

Below are a list of top 20 countries with the highest death rates out of countries 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 - 30 Mar 2020

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	Italy	101,739	4,050	75,528	14,620	11,591	812	11.4%
2	Spain	87,956	7,846	63,460	16,780	7,716	913	8.8%
3	Indonesia	1,414	129	1,217	75	122	8	8.6%
4	Netherlands	11,817	887	10,699	253	865	93	7.3%
5	France	45,170	4,462	34,176	7,964	3,030	419	6.7%
6	Iran	41,495	3,186	24,827	13,911	2,757	117	6.6%
7	United Kingdom	22,453	2,673	20,871	171	1,411	180	6.3%
8	Philippines	1,546	128	1,426	42	78	7	5.0%
9	Belgium	11,899	1,063	9,859	1,527	513	82	4.3%
10	China	82,198	76	2,967	75,923	3,308	4	4.0%
11	Sweden	4,028	328	3,866	16	146	36	3.6%
12	Brazil	4,579	323	4,300	120	159	23	3.5%
13	Greece	1,212	56	1,117	52	43	5	3.5%
14	Ecuador	1,962	38	1,899	3	60	2	3.1%
15	Romania	2,109	294	1,835	209	65	22	3.1%
16	Japan	1,866	0	1,388	424	54	0	2.9%
17	Denmark	2,755	191	2,605	73	77	5	2.8%
18	India	1,251	227	1,117	102	32	5	2.6%
19	Switzerland	15,922	1,093	13,740	1,823	359	59	2.3%
20	Portugal	6,408	446	6,225	43	140	21	2.2%

6 Conclusions

As of 30 Mar 2020, there are 178 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 4.8% and 18.6%, but it is likely to change dramatically with the breakout in many countries, such as European countries.

Appendix A. Processed Data

Below 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
2020-03-30	782,365	37,582	164,566	580,217	62,248	3,657	15,484	4.8	18.6	19.1
2020-03-29	720,117	33,925	149,082	537,110	59,411	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	32,557	1,678	6,207	4.3	13.0	21.3
2020-03-21	304,396	12,973	91,692	199,731	32,361	1,674	4,272	4.3	12.4	28.2
2020-03-20	272,035	11,299	87,420	173,316	29,535	1,432	2,445	4.2	11.4	36.9
2020-03-19	242,500	9,867	84,975	147,658	27,679	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
2020-01-22	555	17	28	510				3.1	37.8	

Appendix A.2 Latest Cases by Country

```
## highlight 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=',')) %>%
  kable_styling(font_size=6, latex_options=c('striped', 'hold_position', 'repeat_header'))
```

Table 6: Cases by Country (30 Mar 2020)

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	World	782,365	62,248	580,217	164,566	37,582	3,657	4.8
2	US	161,807	20,921	153,185	5,644	2,978	511	1.8
3	Italy	101,739	4,050	75,528	14,620	11,591	812	11.4
4	Spain	87,956	7,846	63,460	16,780	7,716	913	8.8
5	China	82,198	76	2,967	75,923	3,308	4	4
6	Germany	66,885	4,790	52,740	13,500	645	112	1
7	France	45,170	4,462	34,176	7,964	3,030	419	6.7
8	Iran	41,495	3,186	24,827	13,911	2,757	117	6.6
9	United Kingdom	22,453	2,673	20,871	171	1,411	180	6.3
10	Switzerland	15,922	1,093	13,740	1,823	359	59	2.3
11	Belgium	11,899	1,063	9,859	1,527	513	82	4.3
12	Netherlands	11,817	887	10,699	253	865	93	7.3
13	Turkey	10,827	1,610	10,497	162	168	37	1.6
14	Korea, South	9,661	78	4,275	5,228	158	6	1.6
15	Austria	9,618	830	8,874	636	108	22	1.1
16	Canada	7,398	1,118	6,852	466	80	16	1.1
17	Portugal	6,408	446	6,225	43	140	21	2.2
18	Israel	4,695	448	4,518	161	16	1	0.3
19	Brazil	4,579	323	4,300	120	159	23	3.5
20	Norway	4,445	161	4,401	12	32	7	0.7
21	Australia	4,361	377	4,087	257	17	1	0.4
22	Sweden	4,028	328	3,866	16	146	36	3.6
23	Czechia	3,001	184	2,953	25	23	7	0.8
24	Ireland	2,910	295	2,851	5	54	8	1.9
25	Denmark	2,755	191	2,605	73	77	5	2.8
26	Malaysia	2,626	156	2,110	479	37	2	1.4
27	Chile	2,449	310	2,285	156	8	1	0.3
28	Romania	2,109	294	1,835	209	65	22	3.1
29	Poland	2,055	193	2,017	7	31	9	1.5
30	Luxembourg	1,988	38	1,926	40	22	1	1.1
31	Ecuador	1,962	38	1,899	3	60	2	3.1
32	Japan	1,866	0	1,388	424	54	0	2.9
33	Russia	1,836	302	1,761	66	9	1	0.5
34	Pakistan	1,717	120	1,620	76	21	7	1.2
35	Philippines	1,546	128	1,426	42	78	7	5
36	Thailand	1,524	136	1,286	229	9	2	0.6
37	Saudi Arabia	1,453	154	1,330	115	8	0	0.6
38	Indonesia	1,414	129	1,217	75	122	8	8.6
39	Finland	1,352	112	1,329	10	13	2	1
40	South Africa	1,326	46	1,292	31	3	1	0.2
41	India	1,251	227	1,117	102	32	5	2.6
42	Greece	1,212	56	1,117	52	43	5	3.5
43	Iceland	1,086	66	927	157	2	0	0.2
44	Mexico	993	145	938	35	20	4	2
45	Panama	989	88	961	4	24	7	2.4
46	Peru	950	98	873	53	24	6	2.5
47	Dominican Republic	901	42	855	4	42	3	4.7
48	Singapore	879	35	648	228	3	0	0.3
49	Argentina	820	75	569	228	23	4	2.8
50	Colombia	798	96	771	15	12	2	1.5
51	Croatia	790	77	717	67	6	0	0.8
52	Serbia	785	44	769	0	16	3	2

Table 6: Cases by Country (30 Mar 2020) (continued)

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
53	Slovenia	756	26	735	10	11	0	1.5
54	Estonia	715	36	692	20	3	0	0.4
55	Diamond Princess	712	0	99	603	10	0	1.4
56	Qatar	693	59	641	51	1	0	0.1
57	Egypt	656	47	465	150	41	1	6.2
58	Iraq	630	83	432	152	46	4	7.3
59	United Arab Emirates	611	41	545	61	5	2	0.8
60	New Zealand	589	75	525	63	1	0	0.2
61	Algeria	584	73	512	37	35	4	6
62	Morocco	556	77	508	15	33	7	5.9
63	Ukraine	548	73	527	8	13	3	2.4
64	Bahrain	515	16	232	279	4	0	0.8
65	Lithuania	491	31	477	7	7	0	1.4
66	Armenia	482	58	449	30	3	0	0.6
67	Hungary	447	39	398	34	15	2	3.4
68	Lebanon	446	8	400	35	11	1	2.5
69	Latvia	376	29	375	1	0	0	0
70	Andorra	370	36	352	10	8	2	2.2
71	Bosnia and Herzegovina	368	45	341	17	10	4	2.7
72	Bulgaria	359	13	334	17	8	0	2.2
73	Slovakia	336	22	329	7	0	0	0
74	Costa Rica	330	16	324	4	2	0	0.6
75	Tunisia	312	0	301	3	8	0	2.6
76	Uruguay	310	6	309	0	1	0	0.3
77	Taiwan*	306	8	262	39	5	3	1.6
78	Kazakhstan	302	18	280	21	1	0	0.3
79	Moldova	298	35	281	15	2	0	0.7
80	North Macedonia	285	26	266	12	7	1	2.5
81	Azerbaijan	273	64	243	26	4	0	1.5
82	Jordan	268	9	237	26	5	2	1.9
83	Kuwait	266	11	194	72	0	0	0
84	Burkina Faso	246	24	203	31	12	0	4.9
85	Cyprus	230	16	201	22	7	2	3
86	San Marino	230	6	192	13	25	3	10.9
87	Albania	223	11	168	44	11	1	4.9
88	Vietnam	203	15	148	55	0	0	0
89	Oman	179	12	150	29	0	0	0
90	Afghanistan	170	50	164	2	4	0	2.4
91	Cuba	170	31	162	4	4	1	2.4
92	Cote d'Ivoire	168	3	161	6	1	0	0.6
93	Senegal	162	20	135	27	0	0	0
94	Malta	156	5	154	2	0	0	0
95	Belarus	152	58	120	32	0	0	0
96	Ghana	152	0	145	2	5	0	3.3
97	Uzbekistan	149	5	140	7	2	0	1.3
98	Cameroon	139	0	128	5	6	0	4.3
99	Honduras	139	29	129	3	7	4	5
100	Venezuela	135	16	93	39	3	1	2.2
101	Nigeria	131	20	121	8	2	1	1.5
102	Mauritius	128	21	125	0	3	0	2.3
103	Brunei	127	1	88	38	1	0	0.8
104	Sri Lanka	122	5	105	15	2	1	1.6
105	West Bank and Gaza	116	7	97	18	1	0	0.9
106	Cambodia	107	4	86	21	0	0	0
107	Georgia	103	12	83	20	0	0	0
108	Bolivia	97	16	93	0	4	3	4.1
109	Kosovo	94	0	92	1	1	0	1.1
110	Kyrgyzstan	94	10	91	3	0	0	0
111	Montenegro	91	6	90	0	1	0	1.1
112	Trinidad and Tobago	82	4	78	1	3	0	3.7
113	Congo (Kinshasa)	81	16	71	2	8	2	9.9
114	Rwanda	70	0	70	0	0	0	0
115	Paraguay	64	5	60	1	3	0	4.7
116	Liechtenstein	62	6	62	0	0	0	0
117	Kenya	50	8	48	1	1	0	2
118	Bangladesh	49	1	25	19	5	0	10.2
119	Monaco	49	3	47	1	1	0	2
120	Madagascar	43	4	43	0	0	0	0

Table 6: Cases by Country (30 Mar 2020) (continued)

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
121	Guatemala	36	2	25	10	1	0	2.8
122	Jamaica	36	4	33	2	1	0	2.8
123	Zambia	35	6	35	0	0	0	0
124	Barbados	33	0	33	0	0	0	0
125	Uganda	33	0	33	0	0	0	0
126	El Salvador	30	6	30	0	0	0	0
127	Togo	30	5	28	1	1	0	3.3
128	Niger	27	9	24	0	3	2	11.1
129	Mali	25	7	23	0	2	1	8
130	Ethiopia	23	2	19	4	0	0	0
131	Guinea	22	6	22	0	0	0	0
132	Congo (Brazzaville)	19	0	19	0	0	0	0
133	Tanzania	19	5	18	1	0	0	0
134	Djibouti	18	0	18	0	0	0	0
135	Maldives	17	0	4	13	0	0	0
136	Haiti	15	0	14	1	0	0	0
137	Bahamas	14	3	13	1	0	0	0
138	Burma	14	4	14	0	0	0	0
139	Equatorial Guinea	12	0	12	0	0	0	0
140	Eritrea	12	0	12	0	0	0	0
141	Mongolia	12	0	10	2	0	0	0
142	Dominica	11	0	11	0	0	0	0
143	Namibia	11	0	9	2	0	0	0
144	Syria	10	1	8	0	2	1	20
145	Eswatini	9	0	9	0	0	0	0
146	Grenada	9	0	9	0	0	0	0
147	Saint Lucia	9	0	8	1	0	0	0
148	Guinea-Bissau	8	6	8	0	0	0	0
149	Guyana	8	0	7	0	1	0	12.5
150	Laos	8	0	8	0	0	0	0
151	Libya	8	0	8	0	0	0	0
152	Mozambique	8	0	8	0	0	0	0
153	Seychelles	8	0	8	0	0	0	0
154	Suriname	8	0	8	0	0	0	0
155	Angola	7	0	5	0	2	0	28.6
156	Antigua and Barbuda	7	0	7	0	0	0	0
157	Gabon	7	0	6	0	1	0	14.3
158	Saint Kitts and Nevis	7	5	7	0	0	0	0
159	Zimbabwe	7	0	6	0	1	0	14.3
160	Benin	6	0	6	0	0	0	0
161	Cabo Verde	6	0	5	0	1	0	16.7
162	Holy See	6	0	6	0	0	0	0
163	Sudan	6	0	4	0	2	1	33.3
164	Chad	5	2	5	0	0	0	0
165	Fiji	5	0	5	0	0	0	0
166	Mauritania	5	0	2	2	1	1	20
167	Nepal	5	0	4	1	0	0	0
168	Bhutan	4	0	4	0	0	0	0
169	Gambia	4	0	3	0	1	0	25
170	Nicaragua	4	0	3	0	1	0	25
171	Belize	3	1	3	0	0	0	0
172	Botswana	3	3	3	0	0	0	0
173	Central African Republic	3	0	3	0	0	0	0
174	Liberia	3	0	3	0	0	0	0
175	Somalia	3	0	3	0	0	0	0
176	MS Zaandam	2	0	2	0	0	0	0
177	Papua New Guinea	1	0	1	0	0	0	0
178	Saint Vincent and the Grenadines	1	0	0	1	0	0	0
179	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!