COVID-19 Data Analysis with R - Worldwide*

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

19 April 2020

Contents

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')
```

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

```
raw.data.recovered <- read.csv('./data/time_series_covid19_recovered_global.csv')

dim(raw.data.confirmed)
```

```
## [1] 264  92
```

Each dataset has 264 rows, corresponding to country/region/province/state. It has 92 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.

```
raw.data.confirmed[1:10, 1:10] %>%
  kable('latex', booktabs=T, caption='Raw Data (Confirmed, First 10 Columns only)') %>%
  kable_styling(font_size=5, 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" "2020-04-18"
```

```
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') %>% paste('UTC')
```

It shows that the data was last updated on 18 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)
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 185 countries with confirmed COVID-19 cases, as of 18 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.

```
## 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)
```

```r
## 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.

```r
## 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
```

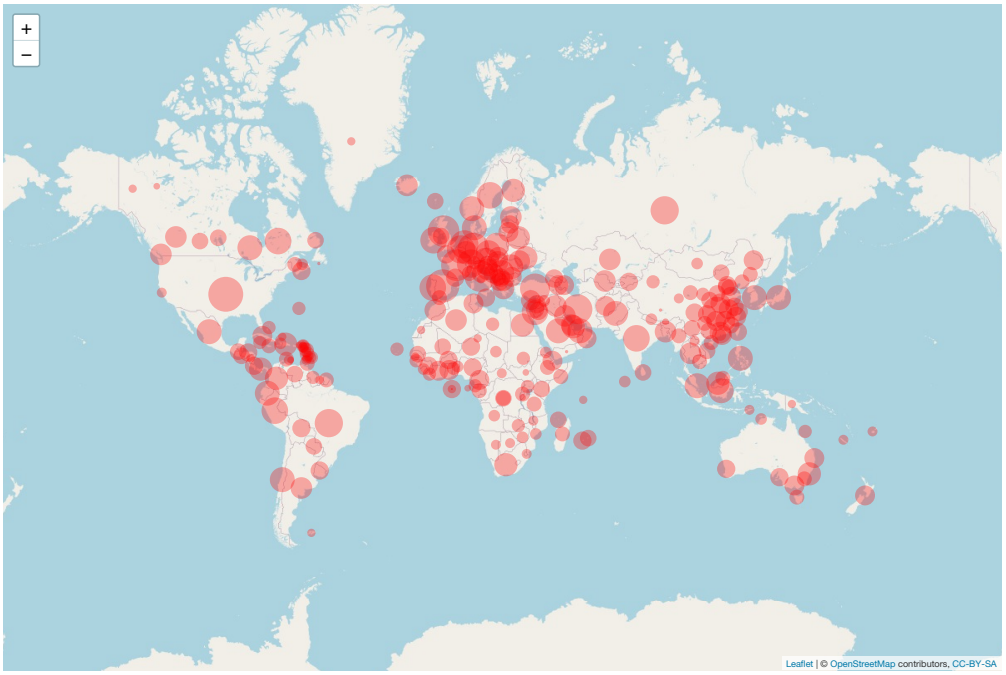


Figure 1: World Map

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=7),
        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=7),
        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')
## show two plots side by side
grid.arrange(plot1, plot2, ncol=2)
```

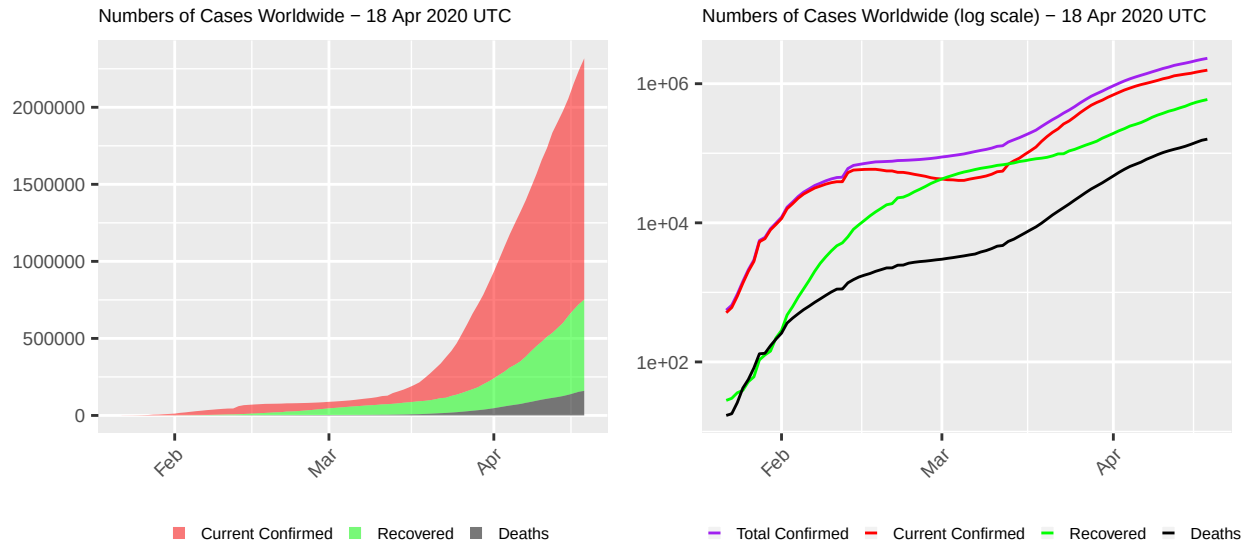


Figure 2: COVID-19 Cases Worldwide

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)
```

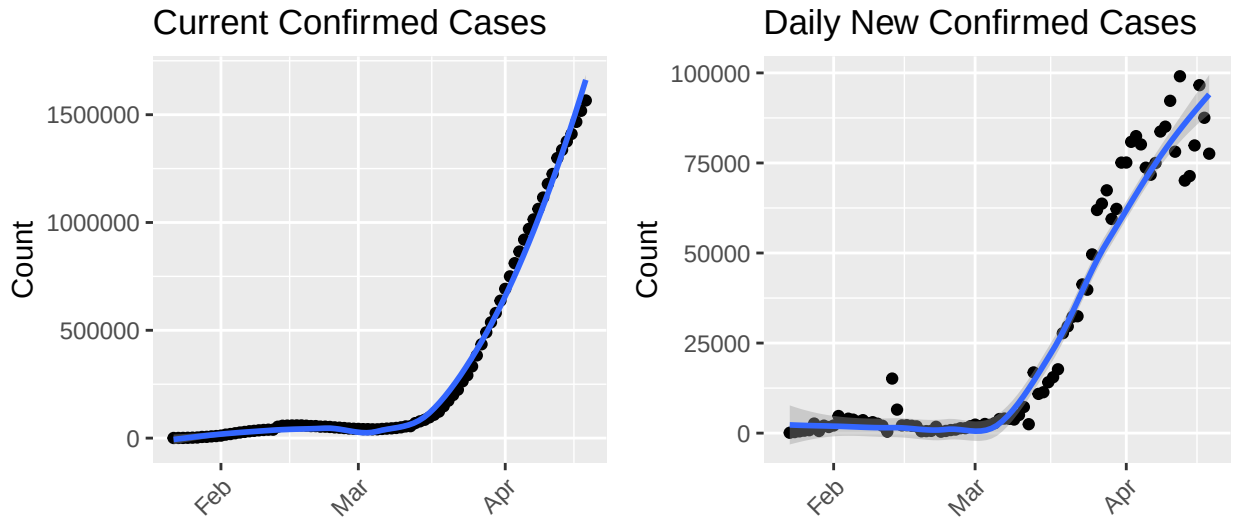


Figure 3: Current Confirmed Cases

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)
```

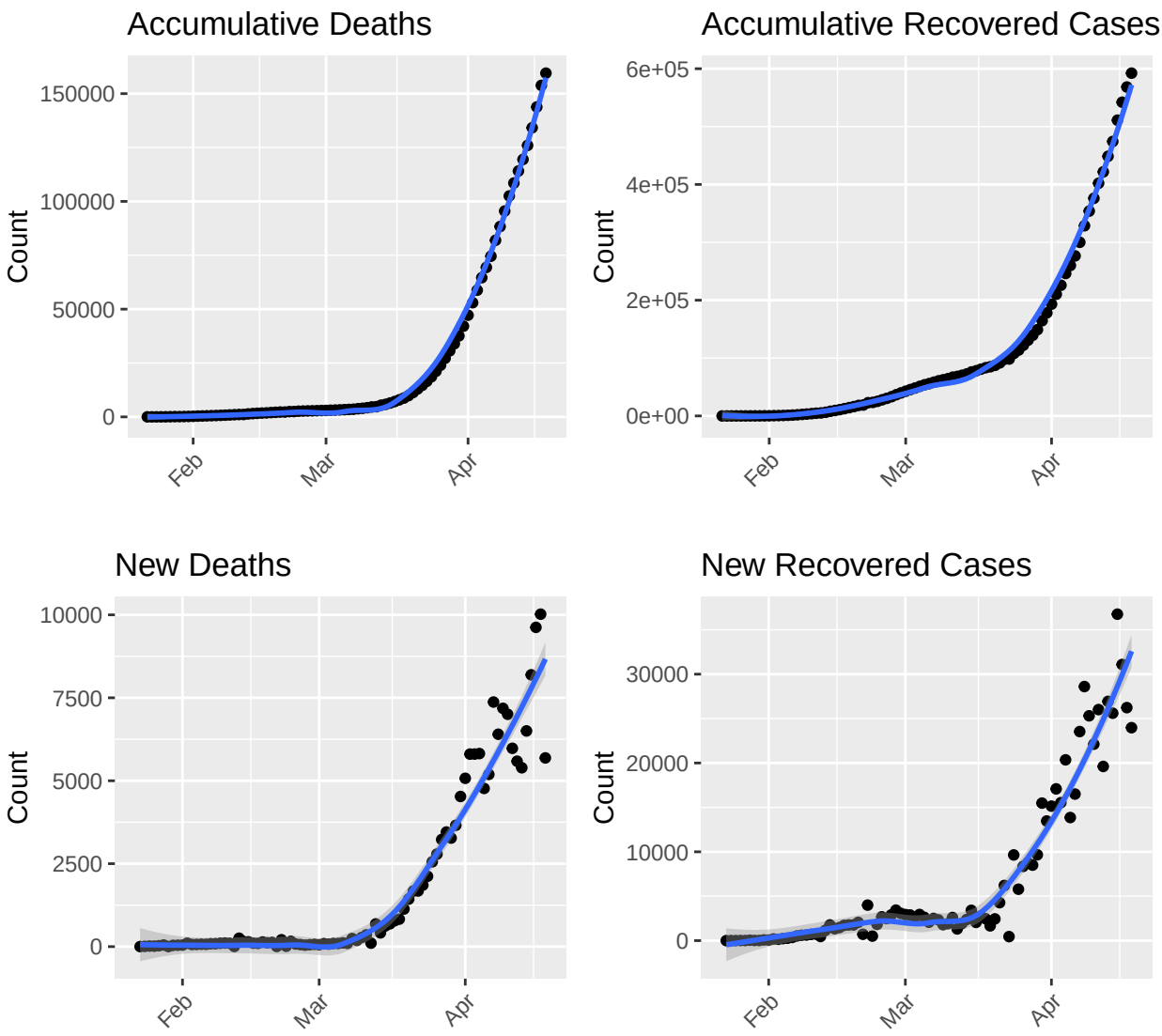


Figure 4: Deaths and Recovered Cases

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 18 Apr 2020 UTC 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 18 Apr 2020 UTC UTC, it will be between 6.9% and 21.2%.

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 outbreak of coronavirus in Iran, European and US.

```
## 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)
```

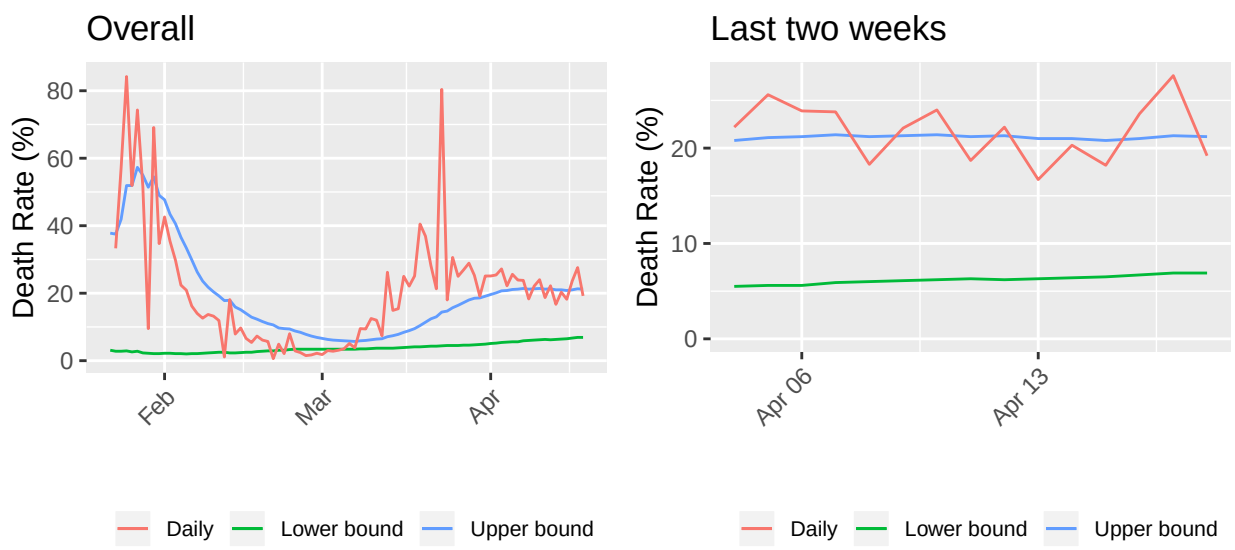


Figure 5: Death Rate

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)))
```

```r
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"        "United Kingdom" "China"          "Turkey"
##  [9] "Iran"           "Belgium"        "Russia"         "Brazil"
## [13] "Canada"         "Netherlands"    "Switzerland"    "Portugal"
## [17] "India"          "Ireland"        "Austria"        "Peru"
```

```r
## 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))
```

```r
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'))
```

```r
## 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'))
```

```r
## 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 - 18 Apr 2020 UTC. 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	2,317,758	159,509	6.9%	77,568	5,688	1,565,930
2	US	732,197	38,664	5.3%	32,491	1,891	628,693
3	Spain	191,726	20,043	10.5%	887	41	96,886
4	Italy	175,925	23,227	13.2%	3,491	482	107,771
5	France	149,149	19,345	13.0%	19	642	93,217
6	Germany	143,342	4,459	3.1%	1,945	107	53,483
7	United Kingdom	115,314	15,498	13.4%	5,545	891	99,402
8	China	83,787	4,636	5.5%	27	0	1,537
9	Turkey	82,329	1,890	2.3%	3,783	121	69,986
10	Iran	80,868	5,031	6.2%	1,374	73	19,850
11	Belgium	37,183	5,453	14.7%	1,045	290	23,382
12	Russia	36,793	313	0.9%	4,785	40	33,423
13	Brazil	36,658	2,354	6.4%	2,976	213	20,278
14	Canada	34,355	1,399	4.1%	1,542	45	21,992
15	Netherlands	31,766	3,613	11.4%	1,147	142	27,836
16	Switzerland	27,404	1,368	5.0%	326	41	8,936
17	Portugal	19,685	687	3.5%	663	30	18,388
18	India	15,722	521	3.3%	1,370	35	12,738
19	Ireland	14,758	571	3.9%	778	41	14,110
20	Austria	14,671	443	3.0%	76	12	4,014
21	Peru	14,420	348	2.4%	931	48	7,388
22	Others	279,706	9,646	3.4%	12,368	503	202,620

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

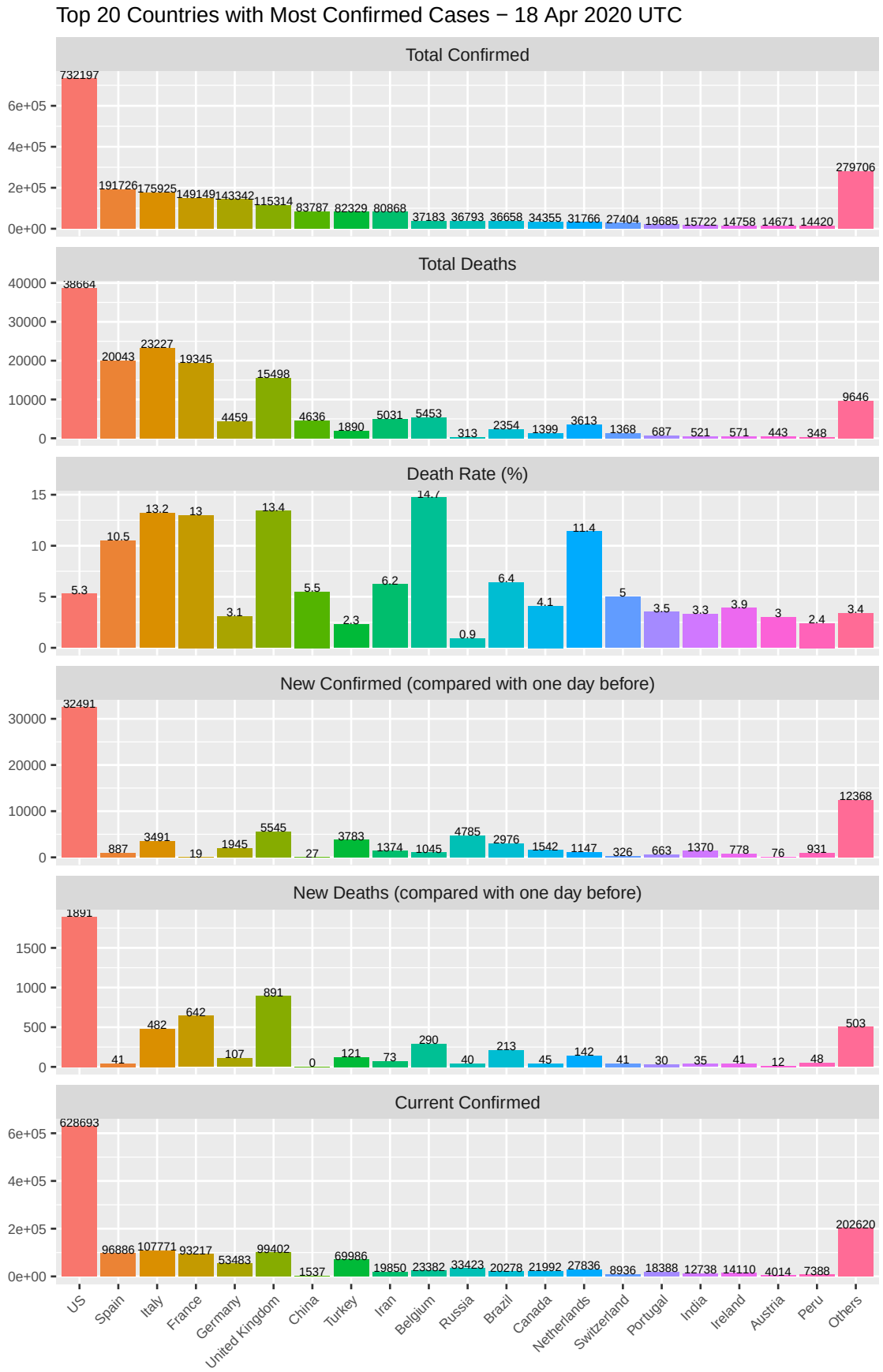
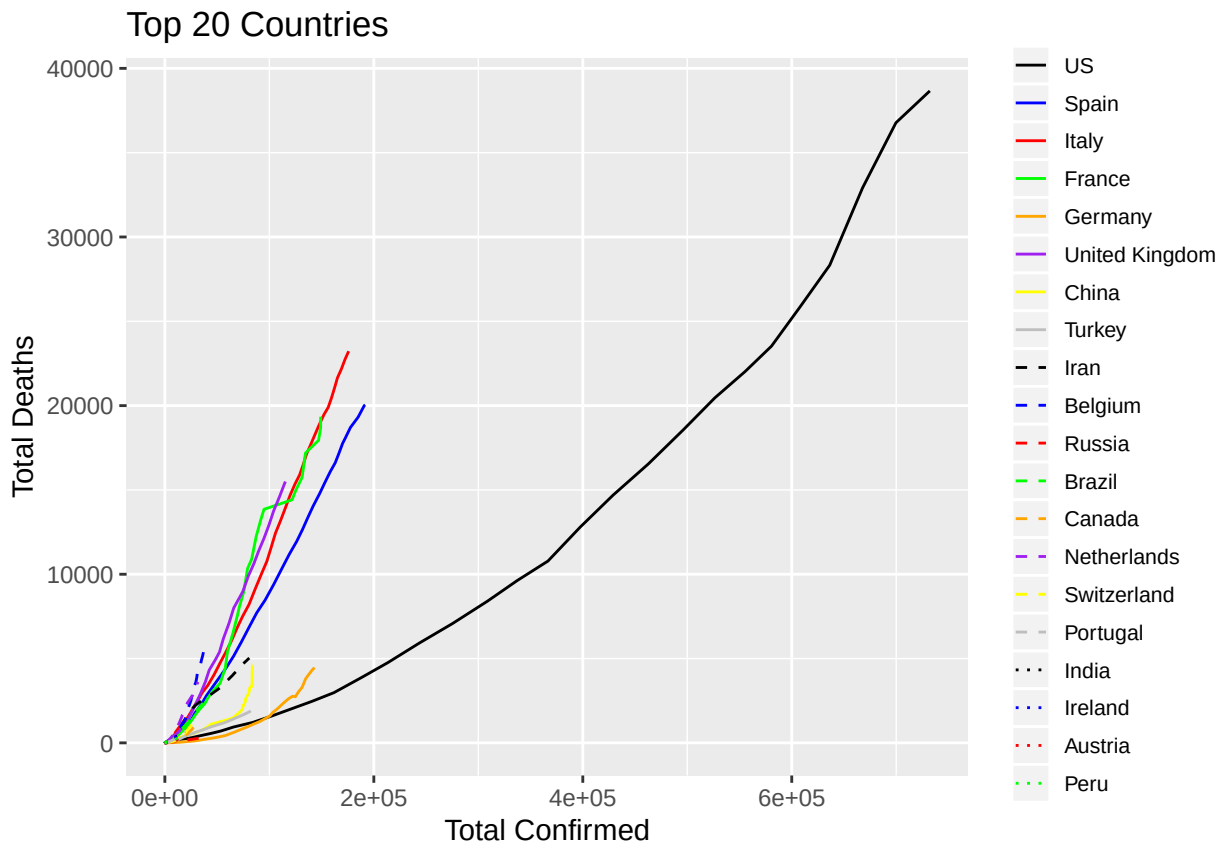


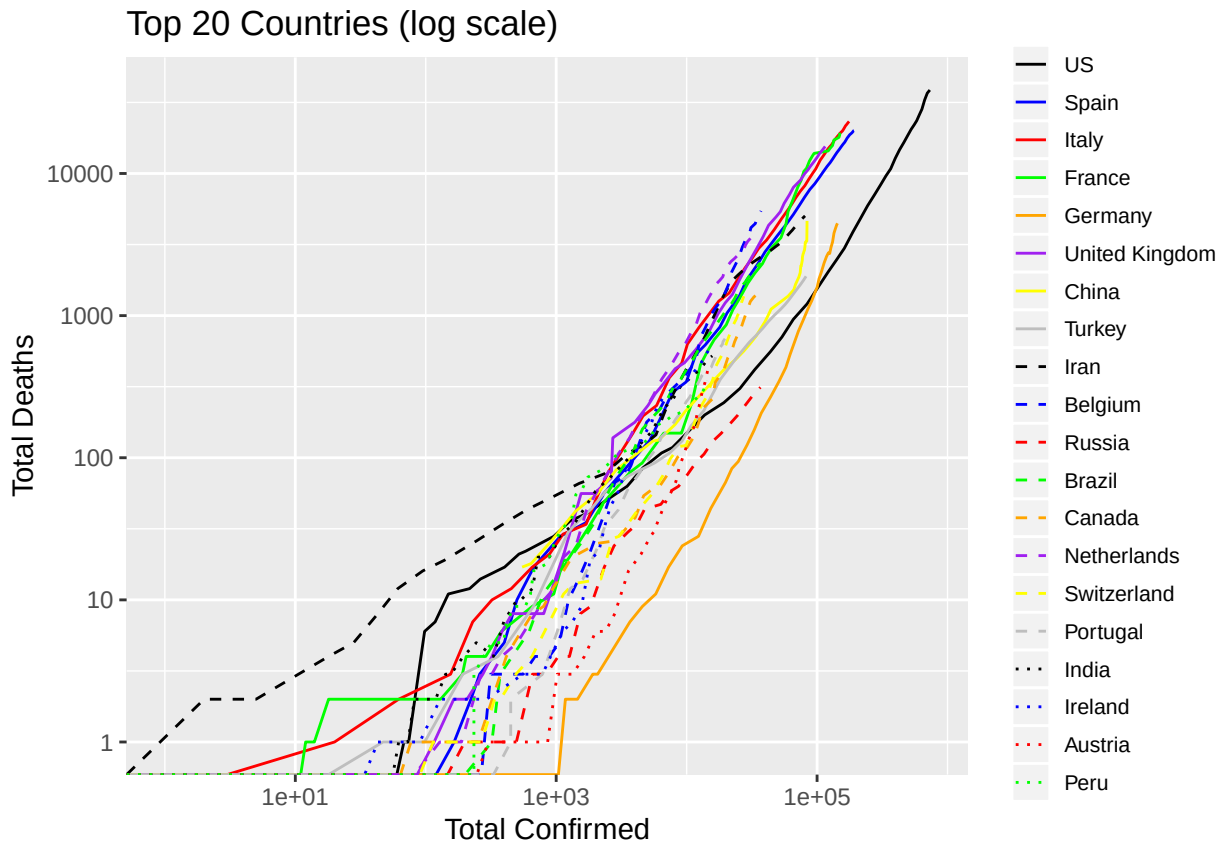
Figure 6: Top 20 Countries with Most Confirmed Cases

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 + labs(title=paste0('Top 20 Countries'))
```




```
p + scale_x_log10() + scale_y_log10() +
  labs(title=paste0('Top 20 Countries (log scale)'))
```



The two figures below show the numbers of confirmed cases and deaths of top 20 countries, as well as the death rates up to 18 Apr 2020 UTC.

```
df <- data.latest %>% filter(country %in% setdiff(top.countries, 'World'))
## breaks for circle size in legend; needs to be adjusted accordingly when the number of total confirme
breaks.confirmed <- c(5e3, 1e4, 2e4, 5e4, 1e5, 2e5, 5e5, 1e6, 2e6, 5e6, 1e7)

plot1 <- df %>% ggplot(aes(x=confirmed, y=deaths, col=death.rate, size=current.confirmed)) +
  scale_size(name='Current Confirmed', trans='log2', breaks=breaks.confirmed) +
  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() +
  labs(title=paste0('Top 20 Countries - Confirmed vs Deaths (log scale)'))

plot2 <- df %>% ggplot(aes(x=new.confirmed, y=new.deaths, col=death.rate, size=current.confirmed)) +
  scale_size(name='Current Confirmed', trans='log2', breaks=breaks.confirmed) +
  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() +
  labs(title=paste0('Top 20 Countries - New Confirmed vs New Deaths (log scale)'))
```

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

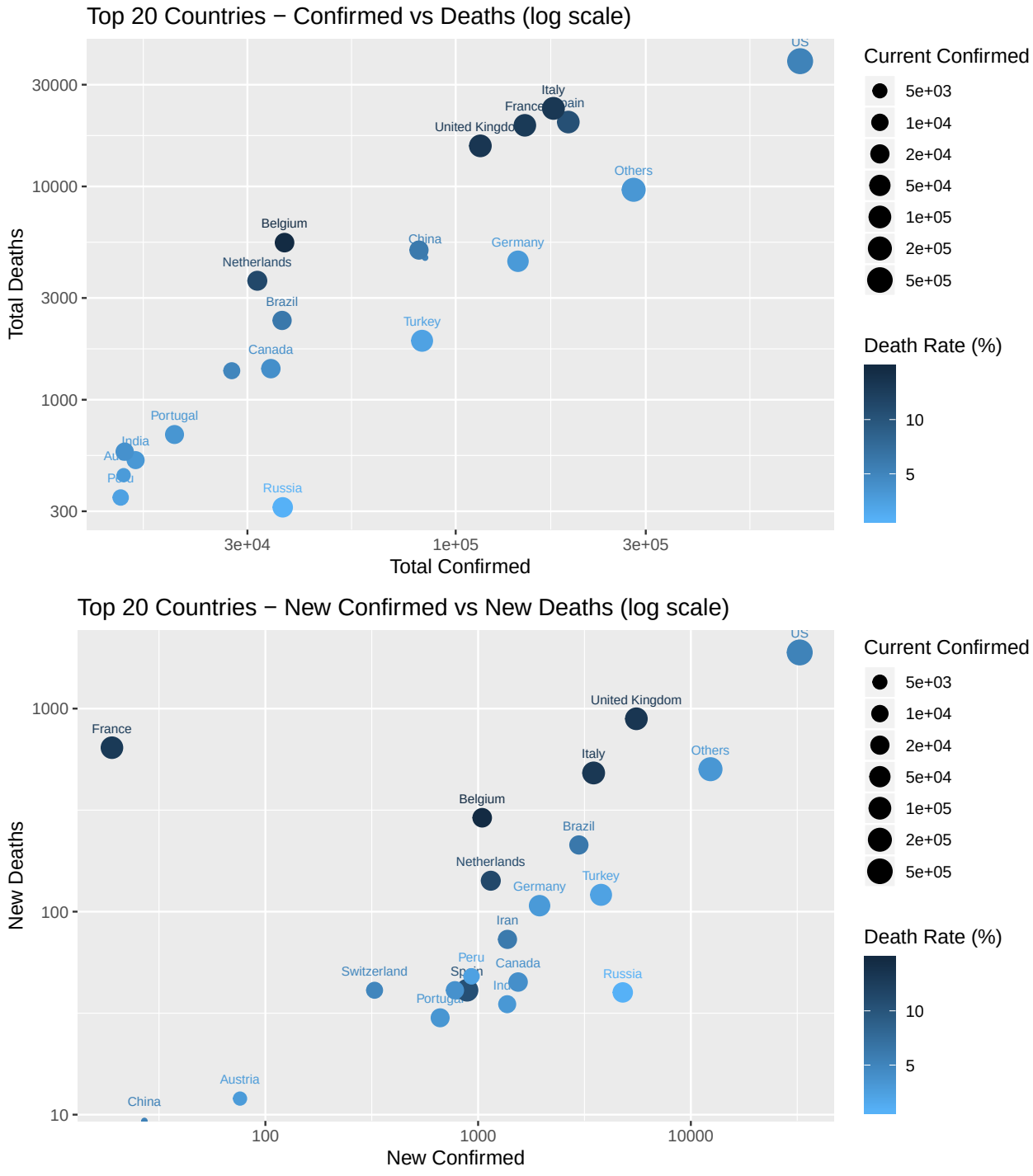


Figure 7: Top 20 Countries

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.

```
## 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)
```

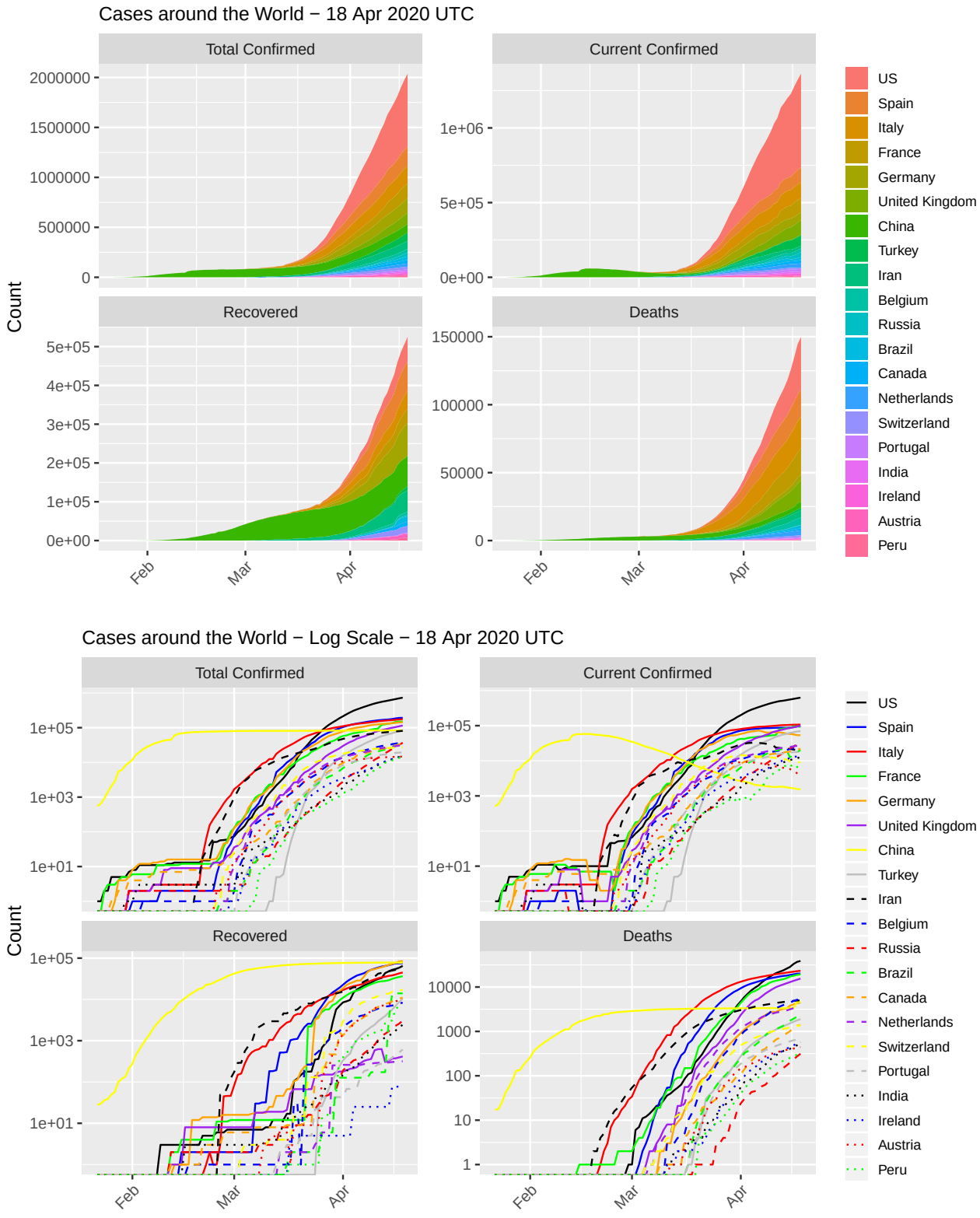


Figure 8: Cases around the World

```
## 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))
```

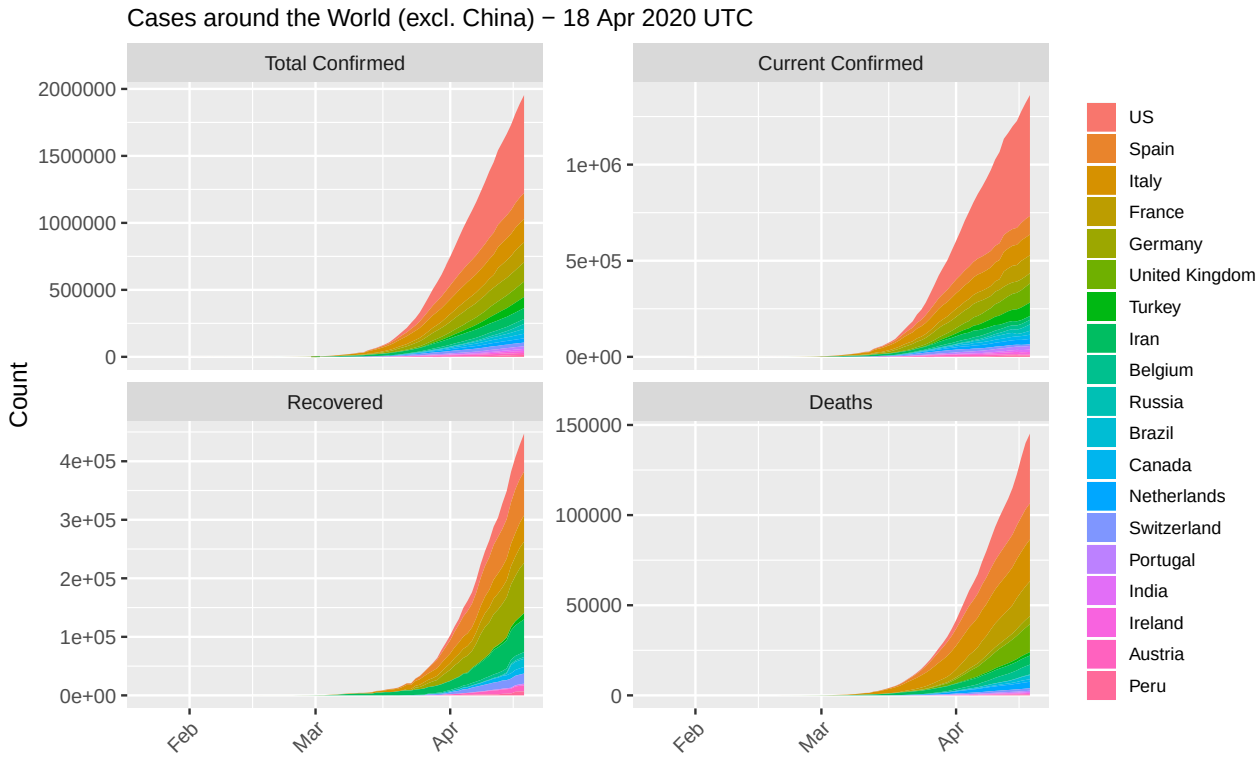


Figure 9: Cases around the World (excl. China)

```
## 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')
```

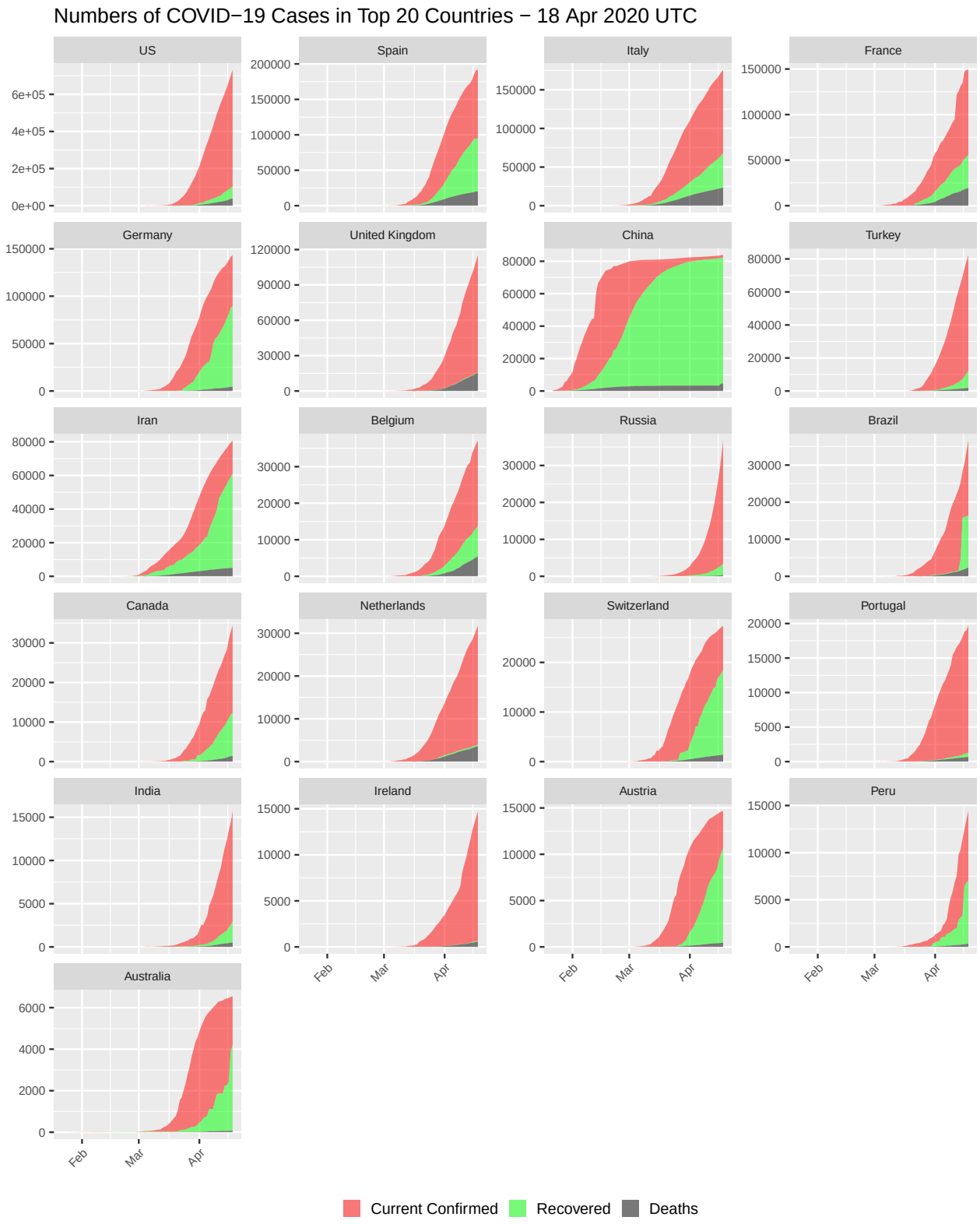


Figure 10: COVID-19 Cases in Top 20 Countries. Ordered descendingly by number of confirmed cases.

```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')
```

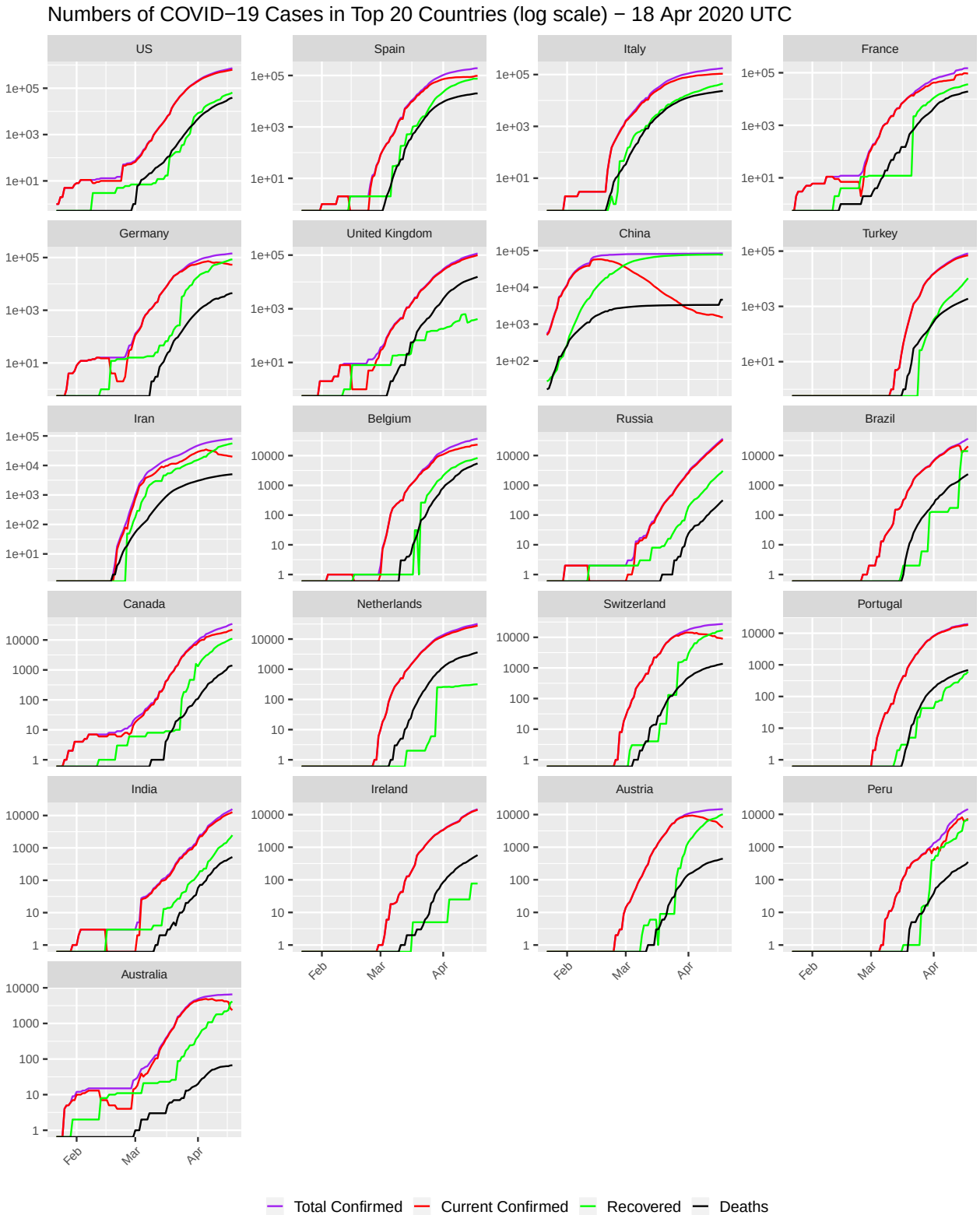


Figure 11: COVID-19 Cases Top 20 Countries (log scale). Ordered descendingly by number of confirmed cases.

```
## 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)
```

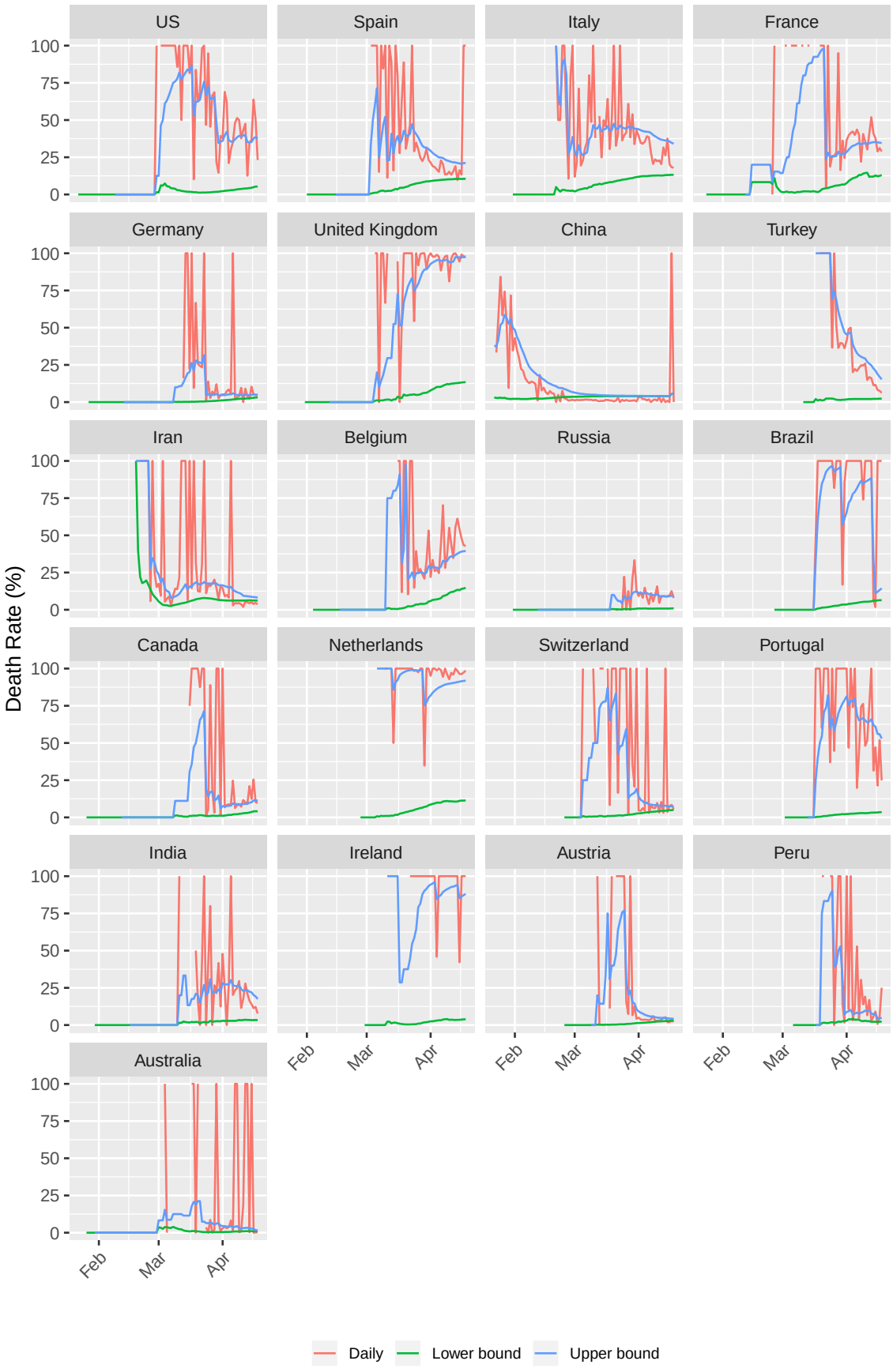


Figure 12: Death Rates

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 - 18 Apr 2020 UTC

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	Belgium	37,183	1,045	23,382	8,348	5,453	290	14.7%
2	Algeria	2,534	116	1,273	894	367	3	14.5%
3	United Kingdom	115,314	5,545	99,402	414	15,498	891	13.4%
4	Italy	175,925	3,491	107,771	44,927	23,227	482	13.2%
5	France	149,149	19	93,217	36,587	19,345	642	13.0%
6	Netherlands	31,766	1,147	27,836	317	3,613	142	11.4%
7	Sweden	13,822	606	11,761	550	1,511	111	10.9%
8	Spain	191,726	887	96,886	74,797	20,043	41	10.5%
9	Hungary	1,834	71	1,431	231	172	16	9.4%
10	Indonesia	6,248	325	5,082	631	535	15	8.6%
11	Mexico	6,875	578	4,204	2,125	546	60	7.9%
12	Egypt	3,032	188	2,107	701	224	19	7.4%
13	Philippines	6,087	209	5,174	516	397	10	6.5%
14	Brazil	36,658	2,976	20,278	14,026	2,354	213	6.4%
15	Iran	80,868	1,374	19,850	55,987	5,031	73	6.2%
16	China	83,787	27	1,537	77,614	4,636	0	5.5%
17	Iraq	1,513	31	478	953	82	1	5.4%
18	Slovenia	1,317	13	1,057	190	70	4	5.3%
19	US	732,197	32,491	628,693	64,840	38,664	1,891	5.3%
20	Ecuador	9,022	572	7,558	1,008	456	35	5.1%

6 Conclusions

As of 18 Apr 2020 UTC, there are 185 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 6.9% and 21.2%, 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)) %>%
```

```r
  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 %>%
  mutate(rate.upper = rate.upper %>% format(nsmall=1) %>% paste0('\\%'),
         rate.lower = rate.lower %>% format(nsmall=1) %>% paste0('\\%'),
         rate.daily = rate.daily %>% format(nsmall=1) %>% paste0('\\%')) %>%
  kable('latex', escape=F, booktabs=T, longtable=T,
        caption='Cases in the Whole World',
        format.args=list(big.mark=','),
        align=c('l', rep('r', 10))) %>%
  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
2020-04-18	2,317,758	159,509	592,319	1,565,930	77,568	5,688	23,976	6.9%	21.2%	19.2%
2020-04-17	2,240,190	153,821	568,343	1,518,026	87,544	10,021	26,236	6.9%	21.3%	27.6%
2020-04-16	2,152,646	143,800	542,107	1,466,739	96,592	9,624	31,088	6.7%	21.0%	23.6%
2020-04-15	2,056,054	134,176	511,019	1,410,859	79,863	8,193	36,758	6.5%	20.8%	18.2%
2020-04-14	1,976,191	125,983	474,261	1,375,947	71,353	6,502	25,606	6.4%	21.0%	20.3%
2020-04-13	1,904,838	119,481	448,655	1,336,702	70,117	5,391	26,933	6.3%	21.0%	16.7%
2020-04-12	1,834,721	114,090	421,722	1,298,909	99,071	5,588	19,612	6.2%	21.3%	22.2%
2020-04-11	1,735,650	108,502	402,110	1,225,038	78,124	5,977	26,014	6.3%	21.2%	18.7%
2020-04-10	1,657,526	102,525	376,096	1,178,905	92,248	7,004	22,121	6.2%	21.4%	24.0%
2020-04-09	1,565,278	95,521	353,975	1,115,782	85,076	7,183	25,314	6.1%	21.3%	22.1%
2020-04-08	1,480,202	88,338	328,661	1,063,203	83,726	6,401	28,607	6.0%	21.2%	18.3%
2020-04-07	1,396,476	81,937	300,054	1,014,485	74,995	7,372	23,539	5.9%	21.4%	23.8%
2020-04-06	1,321,481	74,565	276,515	970,401	71,727	5,191	16,503	5.6%	21.2%	23.9%
2020-04-05	1,249,754	69,374	260,012	920,368	73,694	4,768	13,860	5.6%	21.1%	25.6%
2020-04-04	1,176,060	64,606	246,152	865,302	80,143	5,819	20,356	5.5%	20.8%	22.2%
2020-04-03	1,095,917	58,787	225,796	811,334	82,451	5,804	15,533	5.4%	20.7%	27.2%
2020-04-02	1,013,466	52,983	210,263	750,220	80,861	5,803	17,086	5.2%	20.1%	25.4%
2020-04-01	932,605	47,180	193,177	692,248	75,118	5,073	15,143	5.1%	19.6%	25.1%
2020-03-31	857,487	42,107	178,034	637,346	75,098	4,525	13,468	4.9%	19.1%	25.1%
2020-03-30	782,389	37,582	164,566	580,241	62,249	3,657	15,484	4.8%	18.6%	19.1%
2020-03-29	720,140	33,925	149,082	537,133	59,447	3,273	9,667	4.7%	18.5%	25.3%
2020-03-28	660,693	30,652	139,415	490,626	67,402	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	49,612	2,556	5,787	4.5%	15.7%	30.6%
2020-03-24	418,041	18,625	108,000	291,416	39,810	2,120	9,649	4.5%	14.7%	18.0%
2020-03-23	378,231	16,505	98,351	263,375	41,278	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%

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
2020-02-12	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%	NA%

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=ifelse(confirmed >= 1000 & death.rate >= 5,
                           cell_spec(format(death.rate, big.mark=',') %>% paste0('%'),
                                     "latex", color="red", bold=T),
                           cell_spec(format(death.rate, big.mark=',') %>% paste0('%'),
                                     "latex", color="black", bold=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 (18 Apr 2020 UTC)

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	World	2,317,758	77,568	1,565,930	592,319	159,509	5,688	**6.9%**
2	US	732,197	32,491	628,693	64,840	38,664	1,891	**5.3%**
3	Spain	191,726	887	96,886	74,797	20,043	41	**10.5%**
4	Italy	175,925	3,491	107,771	44,927	23,227	482	**13.2%**
5	France	149,149	19	93,217	36,587	19,345	642	**13.0%**
6	Germany	143,342	1,945	53,483	85,400	4,459	107	3.1%
7	United Kingdom	115,314	5,545	99,402	414	15,498	891	**13.4%**
8	China	83,787	27	1,537	77,614	4,636	0	**5.5%**
9	Turkey	82,329	3,783	69,986	10,453	1,890	121	2.3%
10	Iran	80,868	1,374	19,850	55,987	5,031	73	**6.2%**
11	Belgium	37,183	1,045	23,382	8,348	5,453	290	**14.7%**
12	Russia	36,793	4,785	33,423	3,057	313	40	0.9%
13	Brazil	36,658	2,976	20,278	14,026	2,354	213	**6.4%**
14	Canada	34,355	1,542	21,992	10,964	1,399	45	4.1%
15	Netherlands	31,766	1,147	27,836	317	3,613	142	**11.4%**
16	Switzerland	27,404	326	8,936	17,100	1,368	41	**5.0%**
17	Portugal	19,685	663	18,388	610	687	30	3.5%
18	India	15,722	1,370	12,738	2,463	521	35	3.3%
19	Ireland	14,758	778	14,110	77	571	41	3.9%
20	Austria	14,671	76	4,014	10,214	443	12	3.0%
21	Peru	14,420	931	7,388	6,684	348	48	2.4%
22	Sweden	13,822	606	11,761	550	1,511	111	**10.9%**
23	Israel	13,265	283	9,645	3,456	164	13	1.2%
24	Korea, South	10,653	18	2,484	7,937	232	2	2.2%
25	Japan	10,296	509	9,005	1,069	222	32	2.2%
26	Chile	9,730	478	5,569	4,035	126	10	1.3%

Table 6: Cases by Country (18 Apr 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
27	Ecuador	9,022	572	7,558	1,008	456	35	**5.1%**
28	Poland	8,742	363	7,414	981	347	15	4.0%
29	Romania	8,418	351	6,267	1,730	421	10	**5.0%**
30	Saudi Arabia	8,274	1,132	6,853	1,329	92	5	1.1%
31	Pakistan	7,638	613	5,663	1,832	143	8	1.9%
32	Denmark	7,437	169	3,060	4,031	346	10	4.7%
33	Norway	7,036	99	6,840	32	164	3	2.3%
34	Mexico	6,875	578	4,204	2,125	546	60	**7.9%**
35	Czechia	6,606	57	5,198	1,227	181	8	2.7%
36	Australia	6,547	25	2,356	4,124	67	1	1.0%
37	United Arab Emirates	6,302	0	5,077	1,188	37	0	0.6%
38	Indonesia	6,248	325	5,082	631	535	15	**8.6%**
39	Philippines	6,087	209	5,174	516	397	10	**6.5%**
40	Serbia	5,994	304	5,240	637	117	7	2.0%
41	Singapore	5,992	942	5,241	740	11	0	0.2%
42	Malaysia	5,305	54	2,115	3,102	88	2	1.7%
43	Ukraine	5,106	444	4,698	275	133	8	2.6%
44	Qatar	5,008	345	4,490	510	8	1	0.2%
45	Belarus	4,779	0	4,392	342	45	3	0.9%
46	Dominican Republic	4,335	209	3,806	312	217	17	**5.0%**
47	Panama	4,210	194	3,972	122	116	7	2.8%
48	Finland	3,681	192	1,891	1,700	90	8	2.4%
49	Luxembourg	3,537	57	2,864	601	72	0	2.0%
50	Colombia	3,439	0	2,652	634	153	0	4.4%
51	South Africa	3,034	251	2,079	903	52	2	1.7%
52	Egypt	3,032	188	2,107	701	224	19	**7.4%**
53	Argentina	2,758	89	1,944	685	129	6	4.7%
54	Thailand	2,733	33	899	1,787	47	0	1.7%
55	Morocco	2,685	121	2,234	314	137	2	**5.1%**
56	Algeria	2,534	116	1,273	894	367	3	**14.5%**
57	Moldova	2,378	114	1,930	391	57	1	2.4%
58	Greece	2,235	11	1,856	269	110	2	4.9%
59	Bangladesh	2,144	306	1,994	66	84	9	3.9%
60	Hungary	1,834	71	1,431	231	172	16	**9.4%**
61	Croatia	1,832	18	1,178	615	39	3	2.1%
62	Bahrain	1,773	33	1,011	755	7	0	0.4%
63	Iceland	1,760	6	460	1,291	9	0	0.5%
64	Kuwait	1,751	93	1,465	280	6	1	0.3%
65	Kazakhstan	1,615	69	1,221	377	17	0	1.1%
66	Iraq	1,513	31	478	953	82	1	**5.4%**
67	Estonia	1,512	53	1,312	162	38	0	2.5%
68	Uzbekistan	1,490	85	1,291	194	5	1	0.3%
69	New Zealand	1,422	13	544	867	11	0	0.8%
70	Azerbaijan	1,373	33	765	590	18	3	1.3%
71	Slovenia	1,317	13	1,057	190	70	4	**5.3%**
72	Bosnia and Herzegovina	1,268	54	883	338	47	1	3.7%
73	Armenia	1,248	47	705	523	20	1	1.6%
74	Lithuania	1,239	90	978	228	33	0	2.7%
75	Oman	1,180	111	998	176	6	0	0.5%
76	North Macedonia	1,170	53	957	164	49	0	4.2%
77	Slovakia	1,089	40	865	213	11	2	1.0%
78	Cameroon	1,017	21	818	177	22	0	2.2%
79	Cuba	986	63	727	227	32	1	3.2%
80	Afghanistan	933	27	791	112	30	0	3.2%
81	Bulgaria	878	32	684	153	41	0	4.7%
82	Tunisia	864	0	784	43	37	0	4.3%
83	Ghana	834	193	726	99	9	1	1.1%
84	Cote d'Ivoire	801	113	554	239	8	2	1.0%
85	Cyprus	761	11	670	79	12	0	1.6%
86	Djibouti	732	0	654	76	2	0	0.3%
87	Diamond Princess	712	0	55	644	13	0	1.8%
88	Latvia	712	30	619	88	5	0	0.7%
89	Andorra	704	8	464	205	35	0	5.0%
90	Lebanon	672	4	552	99	21	0	3.1%
91	Costa Rica	655	6	554	97	4	0	0.6%
92	Niger	639	12	507	113	19	1	3.0%
93	Burkina Faso	565	8	208	321	36	1	6.4%
94	Albania	548	9	220	302	26	0	4.7%

Table 6: Cases by Country (18 Apr 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
95	Nigeria	542	49	357	166	19	2	3.5%
96	Guinea	518	41	450	65	3	0	0.6%
97	Kosovo	510	30	405	93	12	0	2.4%
98	Uruguay	508	6	205	294	9	0	1.8%
99	Kyrgyzstan	506	17	371	130	5	0	1.0%
100	Bolivia	493	28	431	31	31	0	6.3%
101	Honduras	457	15	401	10	46	5	10.1%
102	San Marino	455	20	356	60	39	0	8.6%
103	Malta	426	4	324	99	3	0	0.7%
104	West Bank and Gaza	418	16	347	69	2	0	0.5%
105	Jordan	413	6	137	269	7	0	1.7%
106	Taiwan*	398	3	214	178	6	0	1.5%
107	Georgia	388	18	298	86	4	1	1.0%
108	Senegal	350	8	136	211	3	1	0.9%
109	Mauritius	325	1	136	180	9	0	2.8%
110	Congo (Kinshasa)	307	20	256	26	25	2	8.1%
111	Montenegro	307	4	247	55	5	0	1.6%
112	Vietnam	268	0	67	201	0	0	0.0%
113	Kenya	262	16	190	60	12	1	4.6%
114	Sri Lanka	254	10	161	86	7	0	2.8%
115	Guatemala	235	21	207	21	7	0	3.0%
116	Venezuela	227	23	105	113	9	0	4.0%
117	Mali	216	45	162	41	13	0	6.0%
118	Paraguay	202	3	159	35	8	0	4.0%
119	El Salvador	190	13	140	43	7	0	3.7%
120	Jamaica	163	20	133	25	5	0	3.1%
121	Tanzania	147	0	131	11	5	0	3.4%
122	Rwanda	144	1	75	69	0	0	0.0%
123	Congo (Brazzaville)	143	0	126	11	6	0	4.2%
124	Brunei	137	1	23	113	1	0	0.7%
125	Somalia	135	19	126	2	7	1	5.2%
126	Cambodia	122	0	19	103	0	0	0.0%
127	Madagascar	120	3	85	35	0	0	0.0%
128	Trinidad and Tobago	114	0	85	21	8	0	7.0%
129	Gabon	108	0	100	7	1	0	0.9%
130	Ethiopia	105	9	86	16	3	0	2.9%
131	Burma	98	10	88	5	5	1	5.1%
132	Monaco	94	0	69	22	3	0	3.2%
133	Togo	84	1	30	49	5	0	6.0%
134	Equatorial Guinea	79	0	75	4	0	0	0.0%
135	Liechtenstein	79	0	23	55	1	0	1.3%
136	Liberia	76	0	62	7	7	0	9.2%
137	Barbados	75	0	53	17	5	0	6.7%
138	Sudan	66	33	50	6	10	4	15.2%
139	Guyana	63	0	48	9	6	0	9.5%
140	Cabo Verde	58	2	56	1	1	0	1.7%
141	Zambia	57	5	22	33	2	0	3.5%
142	Bahamas	55	1	36	10	9	0	16.4%
143	Uganda	55	0	33	22	0	0	0.0%
144	Libya	49	0	37	11	1	0	2.0%
145	Guinea-Bissau	46	3	46	0	0	0	0.0%
146	Haiti	44	1	41	0	3	0	6.8%
147	Eritrea	39	4	36	3	0	0	0.0%
148	Syria	38	0	31	5	2	0	5.3%
149	Benin	35	0	16	18	1	0	2.9%
150	Maldives	35	7	19	16	0	0	0.0%
151	Mozambique	35	1	31	4	0	0	0.0%
152	Chad	33	6	25	8	0	0	0.0%
153	Mongolia	31	0	26	5	0	0	0.0%
154	Nepal	31	1	29	2	0	0	0.0%
155	Sierra Leone	30	4	30	0	0	0	0.0%
156	Zimbabwe	25	1	20	2	3	0	12.0%
157	Angola	24	5	16	6	2	0	8.3%
158	Antigua and Barbuda	23	0	17	3	3	0	13.0%
159	Eswatini	22	6	13	8	1	0	4.5%
160	Laos	19	0	17	2	0	0	0.0%
161	Belize	18	0	16	0	2	0	11.1%
162	Timor-Leste	18	0	17	1	0	0	0.0%

Table 6: Cases by Country (18 Apr 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
163	Fiji	17	0	17	0	0	0	0.0%
164	Malawi	17	0	12	3	2	0	11.8%
165	Dominica	16	0	8	8	0	0	0.0%
166	Namibia	16	0	10	6	0	0	0.0%
167	Botswana	15	0	14	0	1	0	6.7%
168	Saint Lucia	15	0	4	11	0	0	0.0%
169	Grenada	14	0	8	6	0	0	0.0%
170	Saint Kitts and Nevis	14	0	14	0	0	0	0.0%
171	Central African Republic	12	0	8	4	0	0	0.0%
172	Saint Vincent and the Grenadines	12	0	11	1	0	0	0.0%
173	Seychelles	11	0	6	5	0	0	0.0%
174	Suriname	10	0	3	6	1	0	10.0%
175	Gambia	9	0	6	2	1	0	11.1%
176	MS Zaandam	9	0	7	0	2	0	22.2%
177	Nicaragua	9	0	1	6	2	1	22.2%
178	Holy See	8	0	6	2	0	0	0.0%
179	Mauritania	7	0	4	2	1	0	14.3%
180	Papua New Guinea	7	0	7	0	0	0	0.0%
181	Western Sahara	6	0	6	0	0	0	0.0%
182	Bhutan	5	0	3	2	0	0	0.0%
183	Burundi	5	0	4	0	1	0	20.0%
184	Sao Tome and Principe	4	0	4	0	0	0	0.0%
185	South Sudan	4	0	4	0	0	0	0.0%
186	Yemen	1	0	1	0	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!