COVID-19 Data Analysis with R - Worldwide*

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

17 June 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] 266 150
```

Each dataset has 266 rows, corresponding to country/region/province/state. It has 150 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" "2020-06-15"
```

```
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 15 Jun 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 188 countries with confirmed COVID-19 cases, as of 15 Jun 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$confirmed),
                        radius=0.03*sqrt(x$confirmed),
                        stroke=F,
                        color='red', fillOpacity=0.3,
                        popup=x$txt)

# world
m
```

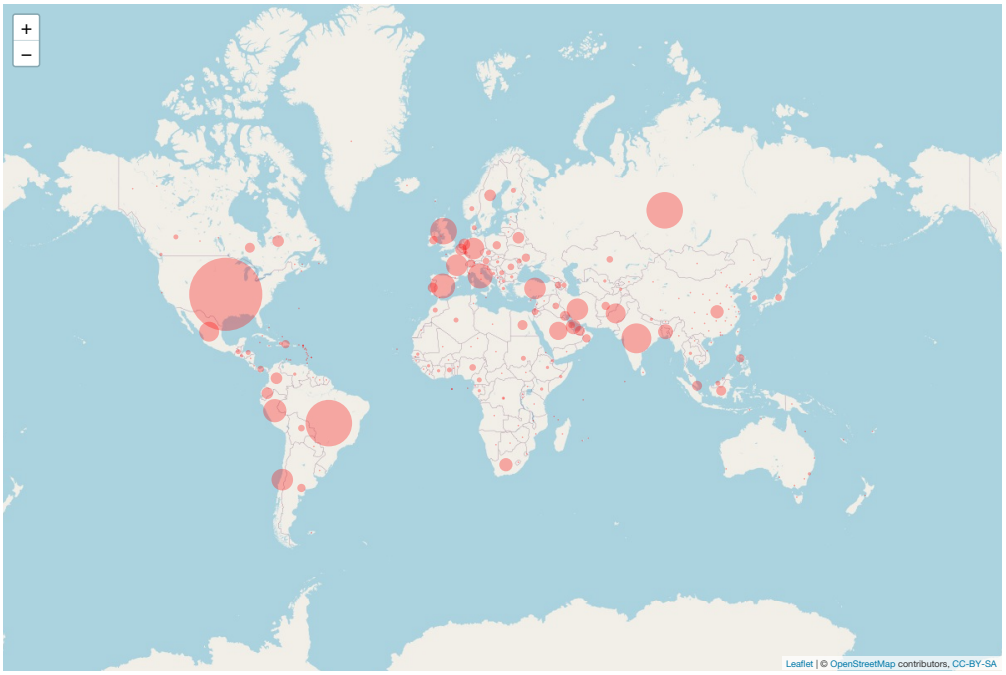


Figure 1: World Map

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)
```

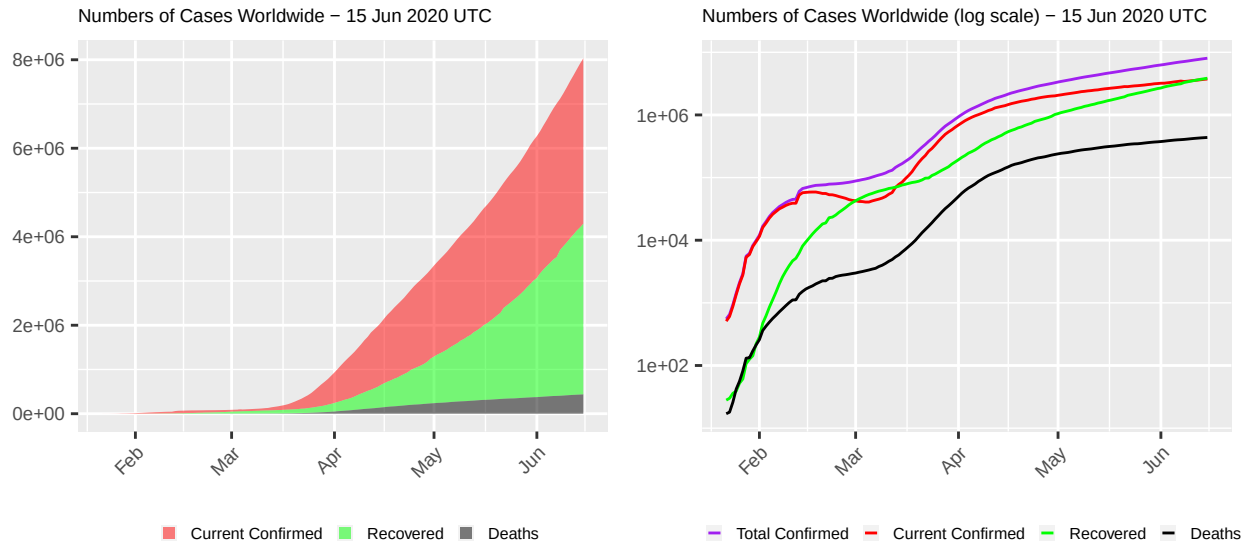


Figure 2: COVID-19 Cases Worldwide

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)
```

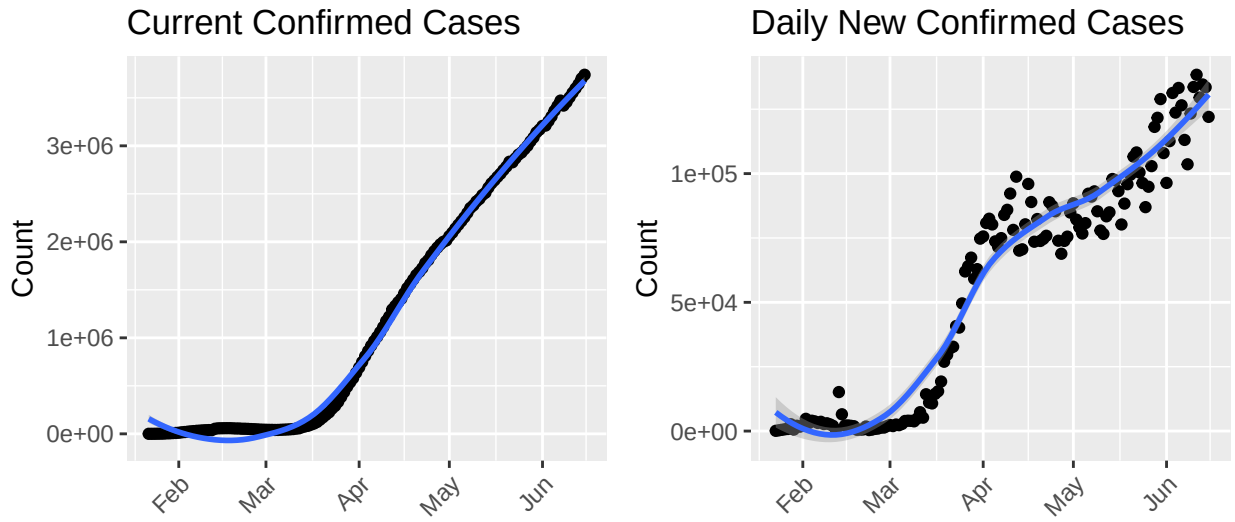


Figure 3: Current Confirmed Cases

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)
```

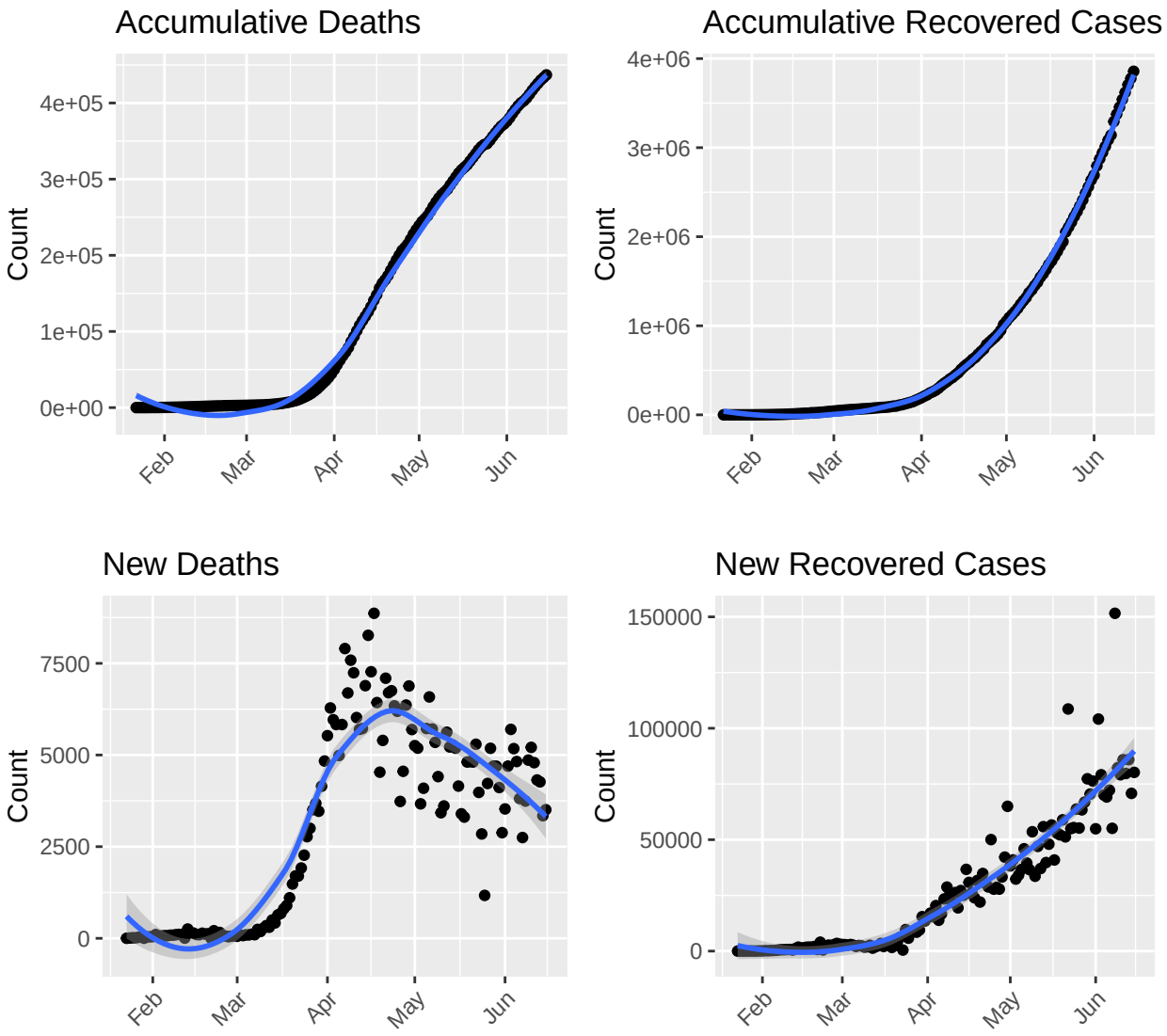


Figure 4: Deaths and Recovered Cases

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 15 Jun 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 15 Jun 2020 UTC UTC, it will be between 5.4% and 10.2%.

A surge in the daily death rate (in red) in late March 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, Europe 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)) +
  ylim(c(0, 99))
## 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, 20))
grid.arrange(plot1, plot2, ncol=2)
```

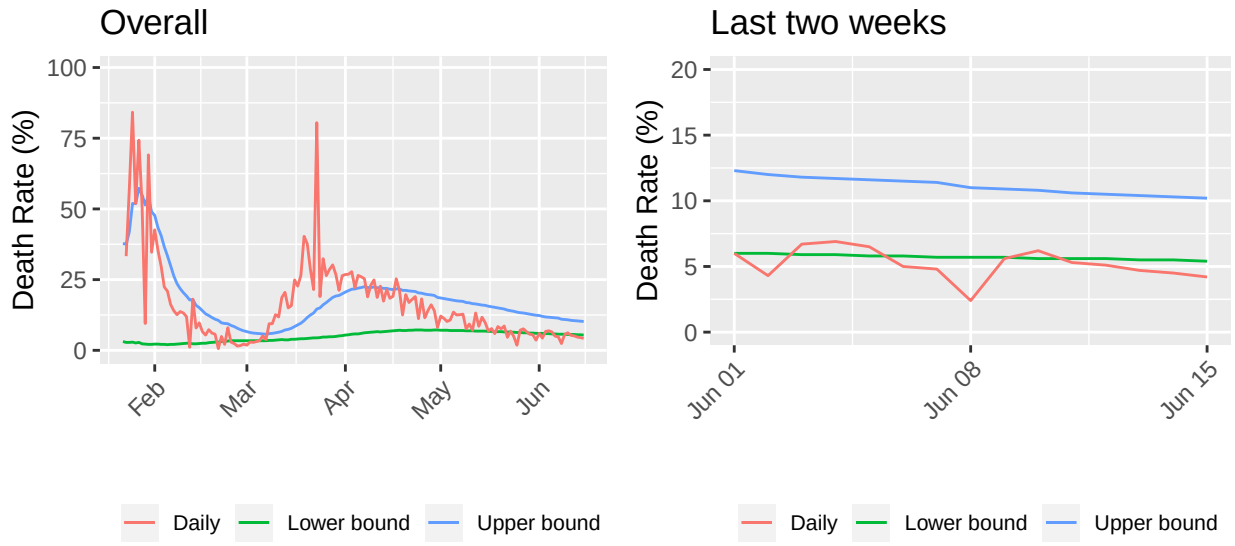


Figure 5: Death Rate

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"             "Brazil"         "Russia"         "India"
##  [5] "United Kingdom" "Spain"          "Italy"          "Peru"
##  [9] "France"         "Iran"           "Germany"        "Turkey"
## [13] "Chile"          "Mexico"         "Pakistan"       "Saudi Arabia"
## [17] "Canada"         "Bangladesh"     "China"          "Qatar"
```

```
## 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, 'Others')))
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 - 15 Jun 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	8,034,461	436,899	5.4%	122,035	3,508	3,740,224
2	US	2,114,026	116,127	5.5%	19,968	395	1,421,565
3	Brazil	888,271	43,959	4.9%	20,647	627	366,603
4	Russia	536,484	7,081	1.3%	8,217	143	245,382
5	India	343,091	9,900	2.9%	10,667	380	153,178
6	United Kingdom	298,315	41,821	14.0%	973	38	255,210
7	Spain	244,109	27,136	11.1%	181	0	66,597
8	Italy	237,290	34,371	14.5%	301	26	25,909
9	Peru	232,992	6,860	2.9%	3,256	172	106,723
10	France	194,305	29,439	15.2%	152	29	91,698
11	Iran	189,876	8,950	4.7%	2,449	113	30,336
12	Germany	187,682	8,807	4.7%	164	6	6,183
13	Turkey	179,831	4,825	2.7%	1,592	18	22,642
14	Chile	179,436	3,362	1.9%	5,143	39	27,282
15	Mexico	150,264	17,580	11.7%	3,427	439	20,392
16	Pakistan	148,921	2,839	1.9%	4,443	110	89,692
17	Saudi Arabia	132,048	1,011	0.8%	4,507	39	43,147
18	Canada	100,763	8,228	8.2%	359	10	31,069
19	Bangladesh	90,619	1,209	1.3%	3,099	38	70,679
20	China	84,378	4,638	5.5%	43	0	251
21	Qatar	80,876	76	0.1%	1,274	3	22,119
22	Others	1,420,884	58,680	4.1%	31,173	883	643,567

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

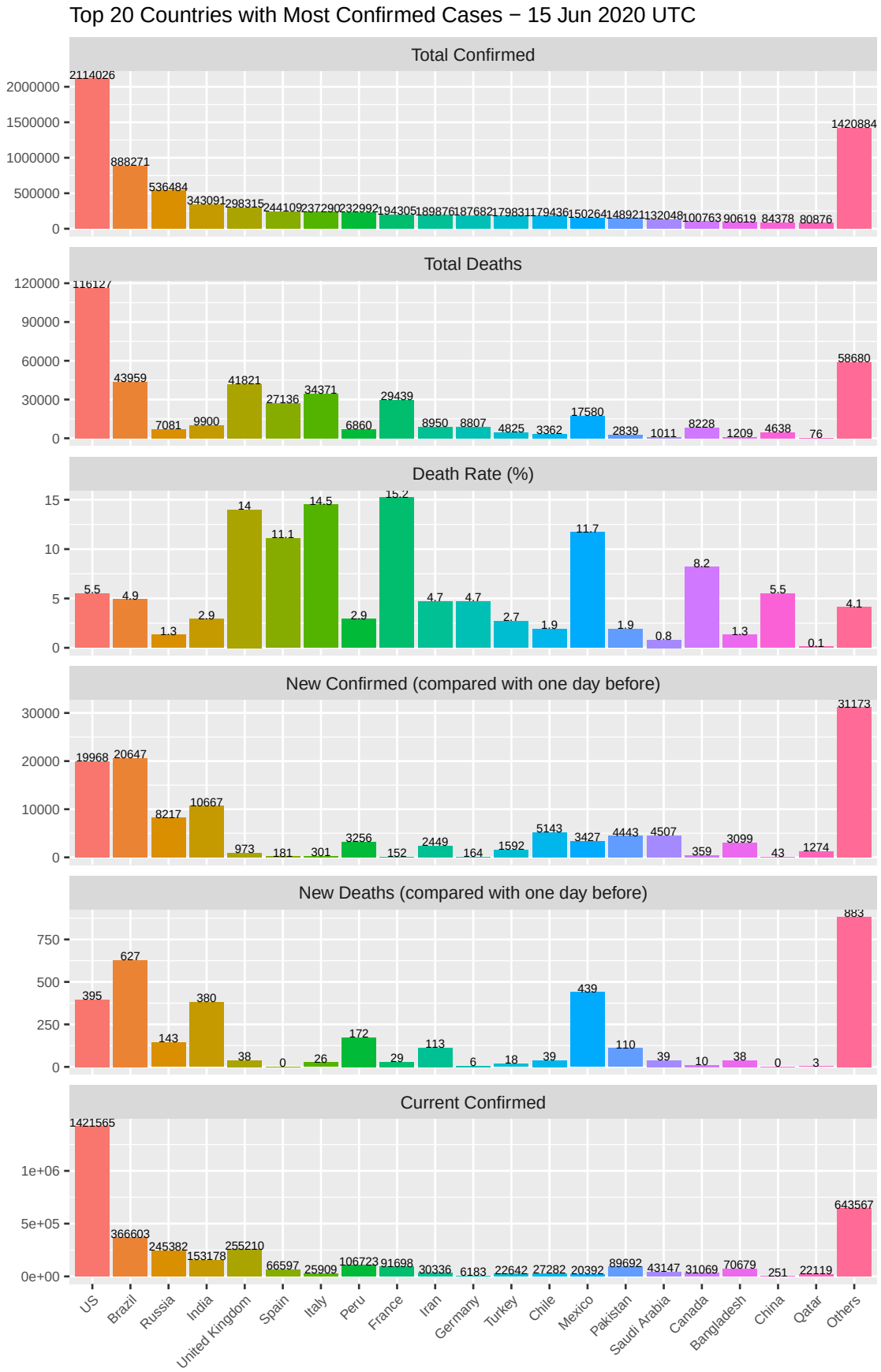
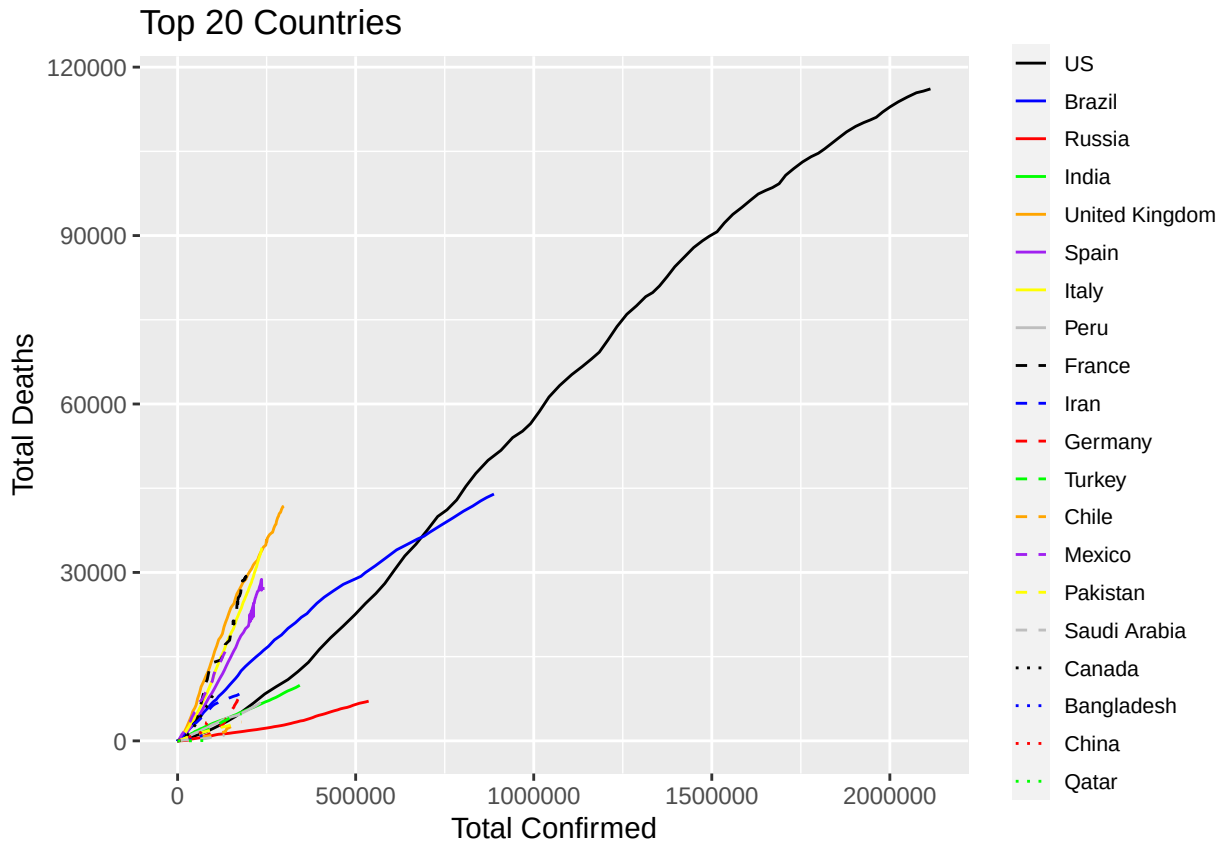


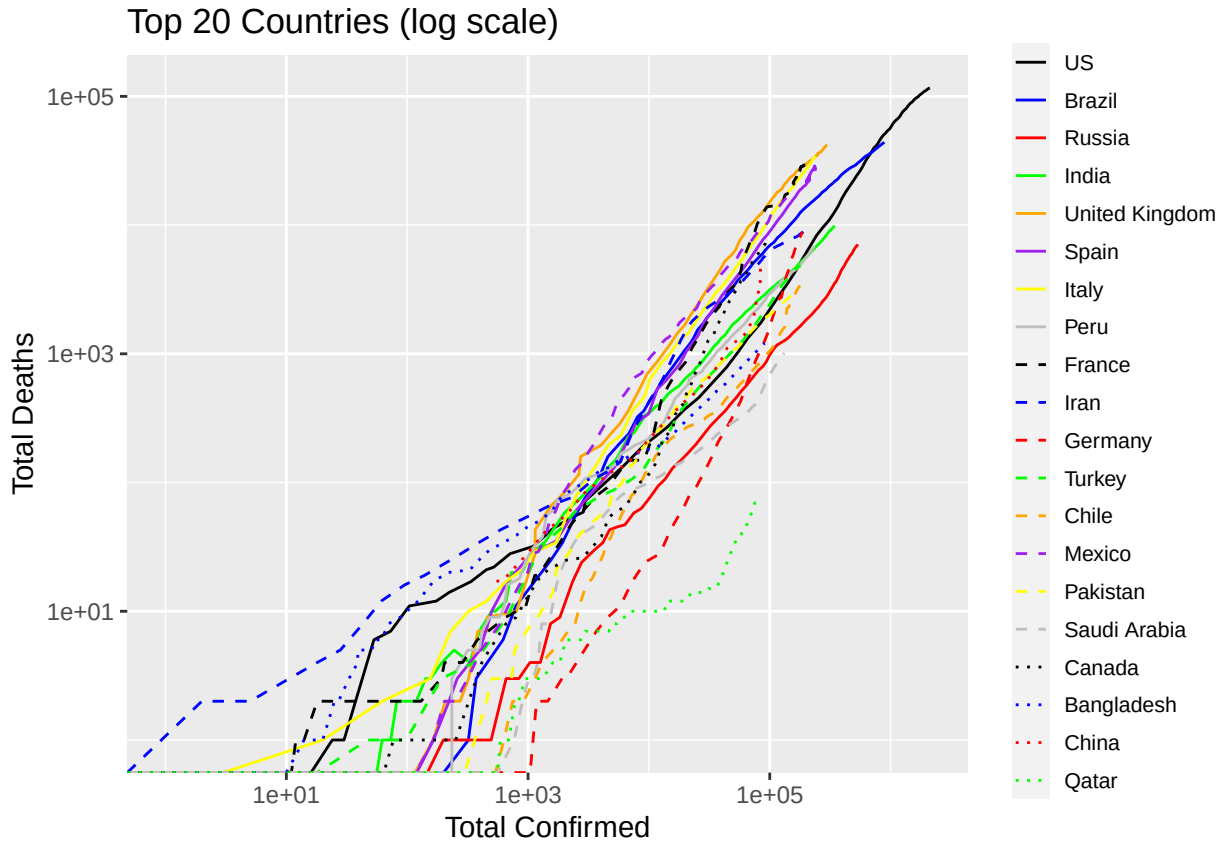
Figure 6: Top 20 Countries with Most Confirmed Cases

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'))) %>%
  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 15 Jun 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)
```

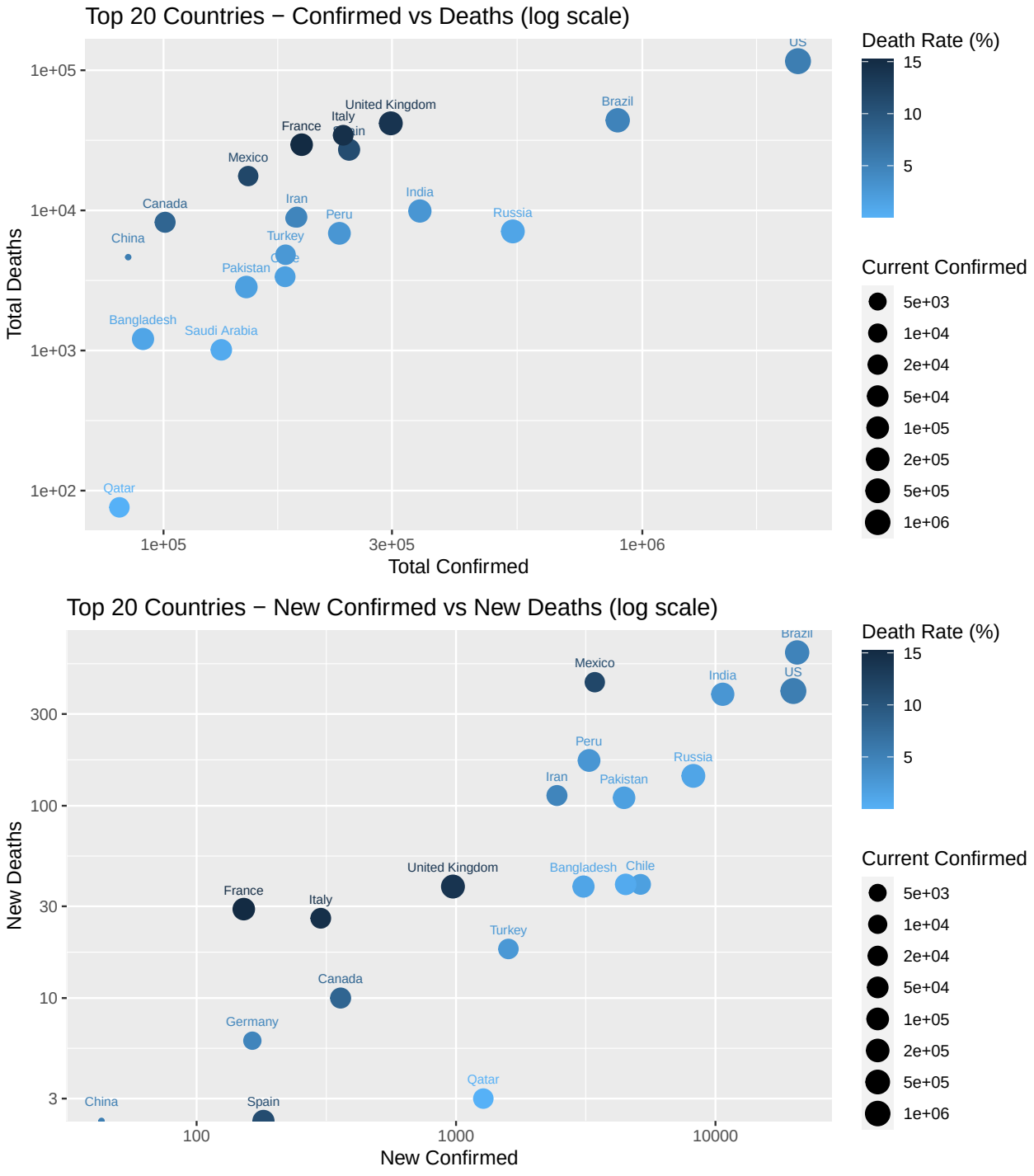


Figure 7: Top 20 Countries

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)
```

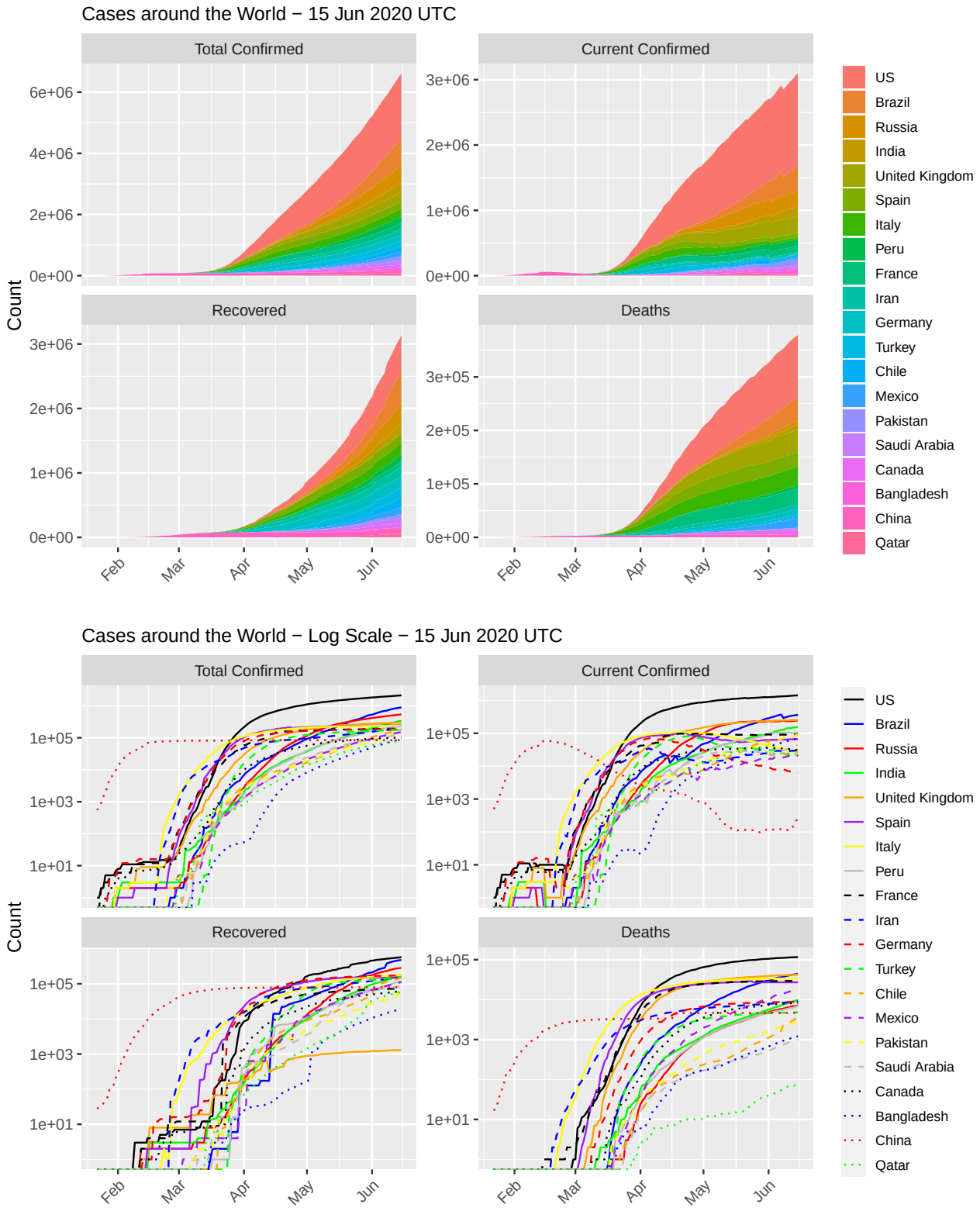


Figure 8: Cases around the World

```
## 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))
```

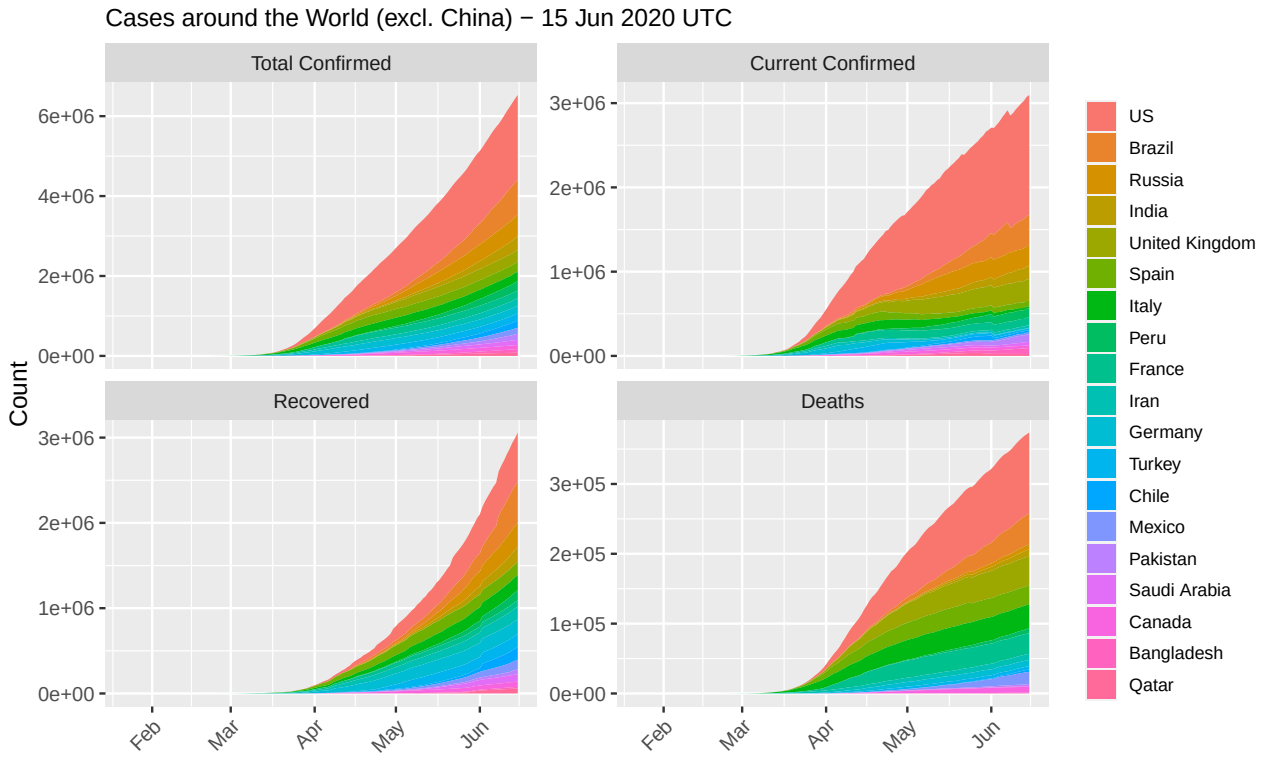


Figure 9: Cases around the World (excl. China)

```
## 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')
```

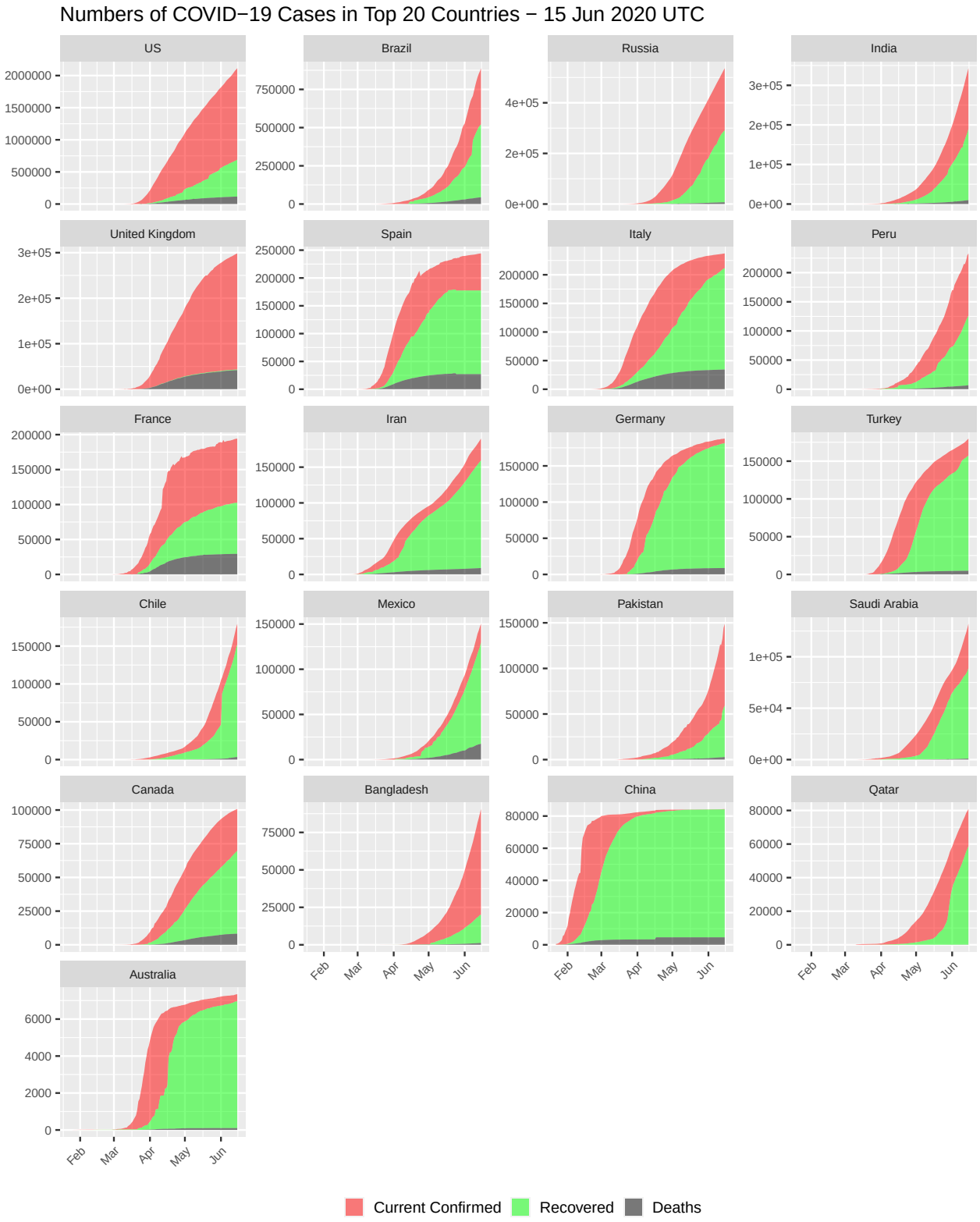


Figure 10: COVID-19 Cases in Top 20 Countries. Ordered descendingly by number of confirmed cases.

```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')
```

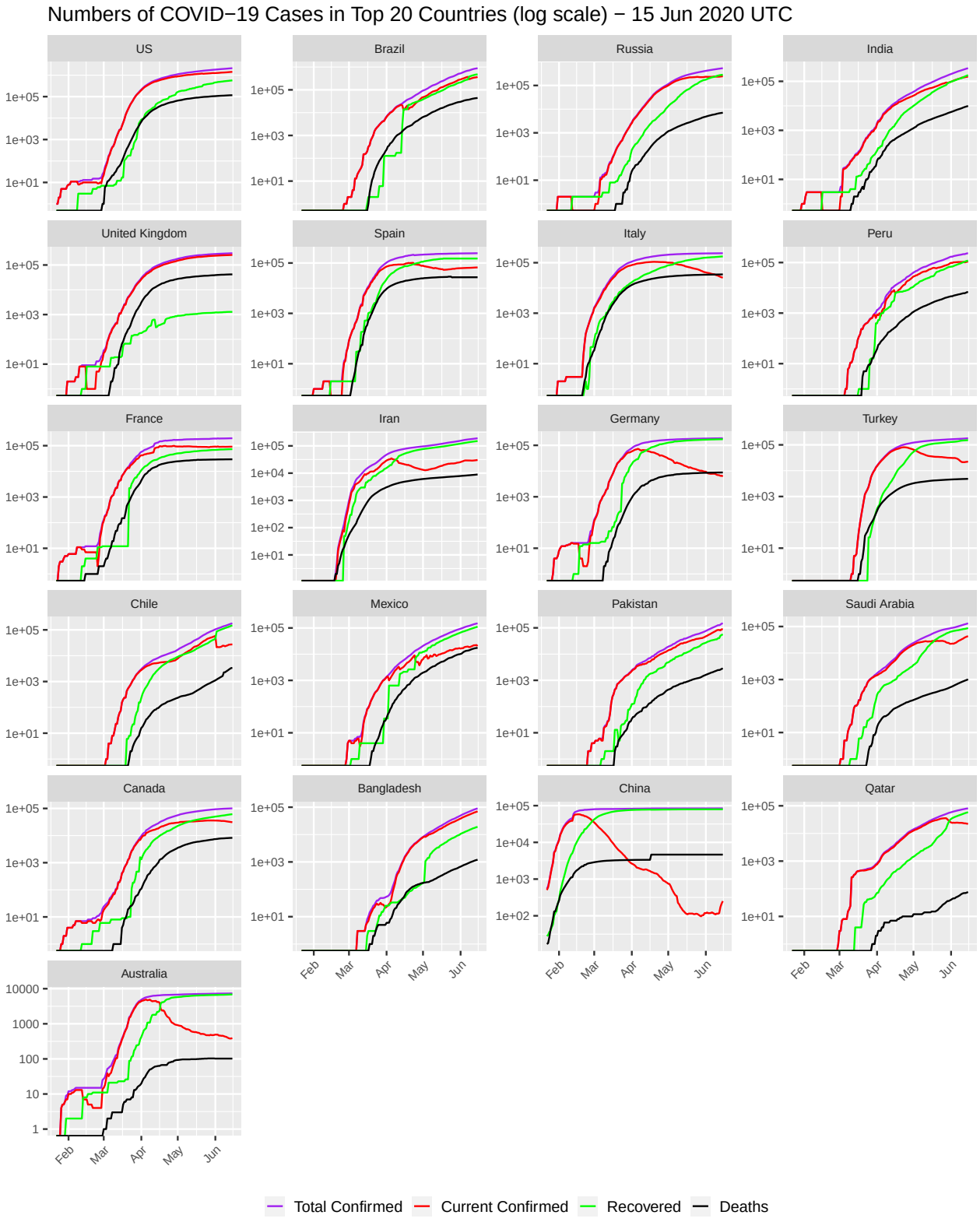


Figure 11: COVID-19 Cases Top 20 Countries (log scale). Ordered descendingly by number of confirmed cases.

```
## 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 China has entered a post-epidemic phase, followed by Australia and Germany, with an increase of recovered cases (in green) every day and a shrinking of the current confirmed cases (in red). In contrast, there are sharp surges in Russia, South America (incl. Brazil, Peru, Chile and Mexico) and West/South Asia (incl. Saudi Arabia, India and Pakistan), which suggests that the virus spread is accerelating 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, 99)) +
  facet_wrap(~country, ncol=4)
```

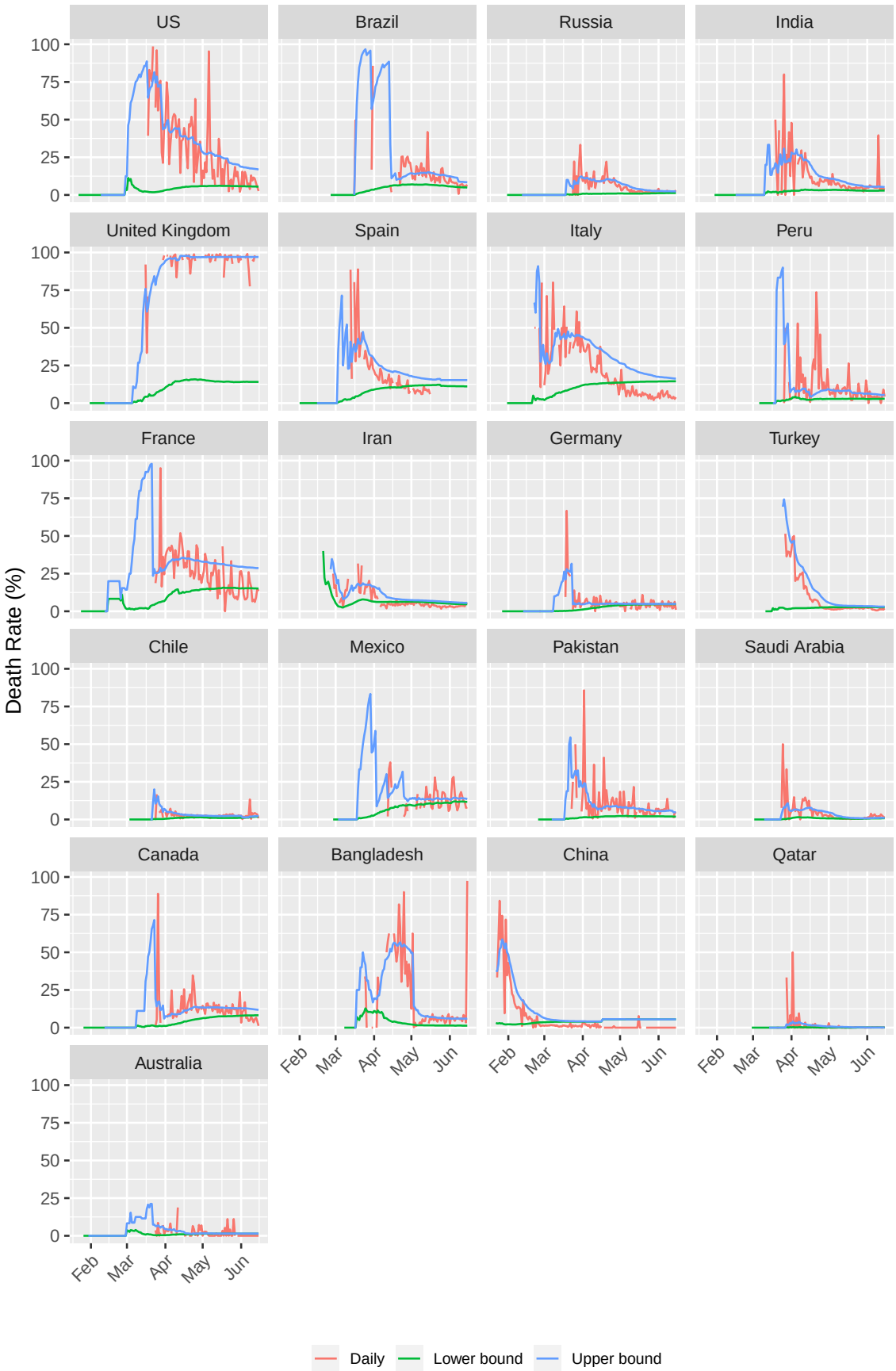


Figure 12: Death Rates

5.4 Countries with Highest Death Rates

Below are a list of top 20 countries with the highest death rates out of countires having 2000+ confirmed cases.

```
## sort the latest data by death rate, and if tie, by confirmed
df <- data %>% filter(date == max(date) & country != 'World' & confirmed >= 2000) %>%
  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 - 15 Jun 2020 UTC

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	Belgium	60,100	71	33,829	16,610	9,661	6	16.1%
2	France	194,305	152	91,698	73,168	29,439	29	15.2%
3	Italy	237,290	301	25,909	177,010	34,371	26	14.5%
4	United Kingdom	298,315	973	255,210	1,284	41,821	38	14.0%
5	Hungary	4,076	7	1,028	2,485	563	1	13.8%
6	Netherlands	49,155	165	42,885	186	6,084	6	12.4%
7	Mexico	150,264	3,427	20,392	112,292	17,580	439	11.7%
8	Spain	244,109	181	66,597	150,376	27,136	0	11.1%
9	Sweden	52,383	769	47,492	0	4,891	17	9.3%
10	Ecuador	47,322	571	20,044	23,349	3,929	33	8.3%
11	Canada	100,763	359	31,069	61,466	8,228	10	8.2%
12	Algeria	11,031	112	2,519	7,735	777	10	7.0%
13	Ireland	25,321	18	917	22,698	1,706	0	6.7%
14	Romania	22,165	166	4,921	15,817	1,427	17	6.4%
15	Sudan	7,435	215	4,247	2,720	468	9	6.3%
16	Switzerland	31,131	14	292	28,900	1,939	1	6.2%
17	Greece	3,134	13	1,576	1,374	184	1	5.9%
18	Indonesia	39,294	1,017	21,973	15,123	2,198	64	5.6%
19	China	84,378	43	251	79,489	4,638	0	5.5%
20	US	2,114,026	19,968	1,421,565	576,334	116,127	395	5.5%

6 Conclusions

As of 15 Jun 2020 UTC, there are 188 countries with confirmed COVID-19 cases. It seems to be contained in China, but starts to break out in rest of the world. The current death rate is in between 5.4% and 10.2%, 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
2020-06-15	8,034,461	436,899	3,857,338	3,740,224	122,035	3,508	80,207	5.4%	10.2%	4.2%
2020-06-14	7,912,426	433,391	3,777,131	3,701,904	133,545	3,344	70,778	5.5%	10.3%	4.5%
2020-06-13	7,778,881	430,047	3,706,353	3,642,481	134,621	4,267	85,941	5.5%	10.4%	4.7%
2020-06-12	7,644,260	425,780	3,620,412	3,598,068	129,536	4,319	79,716	5.6%	10.5%	5.1%
2020-06-11	7,514,724	421,461	3,540,696	3,552,567	138,391	4,791	85,889	5.6%	10.6%	5.3%
2020-06-10	7,376,333	416,670	3,454,807	3,504,856	133,641	5,209	79,139	5.6%	10.8%	6.2%
2020-06-09	7,242,692	411,461	3,375,668	3,455,563	123,337	4,861	82,260	5.7%	10.9%	5.6%
2020-06-08	7,119,355	406,600	3,293,408	3,419,347	103,616	3,744	151,598	5.7%	11.0%	2.4%
2020-06-07	7,015,739	402,856	3,141,810	3,471,073	113,089	2,749	55,093	5.7%	11.4%	4.8%
2020-06-06	6,902,650	400,107	3,086,717	3,415,826	126,524	3,813	72,202	5.8%	11.5%	5.0%
2020-06-05	6,776,126	396,294	3,014,515	3,365,317	133,273	4,822	69,156	5.8%	11.6%	6.5%
2020-06-04	6,642,853	391,472	2,945,359	3,306,022	123,689	5,174	70,029	5.9%	11.7%	6.9%
2020-06-03	6,519,164	386,298	2,875,330	3,257,536	131,315	5,699	79,142	5.9%	11.8%	6.7%
2020-06-02	6,387,849	380,599	2,796,188	3,211,062	112,603	4,697	104,123	6.0%	12.0%	4.3%
2020-06-01	6,275,246	375,902	2,692,065	3,207,279	96,386	3,529	54,895	6.0%	12.3%	6.0%
2020-05-31	6,178,860	372,373	2,637,170	3,169,317	107,976	2,881	76,308	6.0%	12.4%	3.6%
2020-05-30	6,070,884	369,492	2,560,862	3,140,530	128,946	4,112	70,472	6.1%	12.6%	5.5%
2020-05-29	5,941,938	365,380	2,490,390	3,086,168	121,685	4,694	77,327	6.1%	12.8%	5.7%
2020-05-28	5,820,253	360,686	2,413,063	3,046,504	118,140	4,696	66,857	6.2%	13.0%	6.6%
2020-05-27	5,702,113	355,990	2,346,206	2,999,917	102,897	5,183	63,393	6.2%	13.2%	7.6%
2020-05-26	5,599,216	350,807	2,282,813	2,965,596	94,892	4,224	55,214	6.3%	13.3%	7.1%
2020-05-25	5,504,324	346,583	2,227,599	2,930,142	86,970	1,171	63,723	6.3%	13.5%	1.8%
2020-05-24	5,417,354	345,412	2,163,876	2,908,066	96,332	2,847	55,440	6.4%	13.8%	4.9%
2020-05-23	5,321,022	342,565	2,108,436	2,870,021	100,437	3,980	54,971	6.4%	14.0%	6.8%
2020-05-22	5,220,585	338,585	2,053,465	2,828,535	108,219	5,293	108,651	6.5%	14.2%	4.6%
2020-05-21	5,112,366	333,292	1,944,814	2,834,260	106,605	4,809	51,265	6.5%	14.6%	8.6%
2020-05-20	5,005,761	328,483	1,893,549	2,783,729	99,568	4,821	58,913	6.6%	14.8%	7.6%
2020-05-19	4,906,193	323,662	1,834,636	2,747,895	95,878	4,809	52,165	6.6%	15.0%	8.4%
2020-05-18	4,810,315	318,853	1,782,471	2,708,991	88,347	3,307	52,898	6.6%	15.2%	5.9%
2020-05-17	4,721,968	315,546	1,729,573	2,676,849	80,241	3,396	40,886	6.7%	15.4%	7.7%
2020-05-16	4,641,727	312,150	1,688,687	2,640,890	93,178	4,152	56,591	6.7%	15.6%	6.8%
2020-05-15	4,548,549	307,998	1,632,096	2,608,455	97,423	5,185	47,989	6.8%	15.9%	9.8%
2020-05-14	4,451,126	302,813	1,584,107	2,564,206	97,915	5,274	39,735	6.8%	16.0%	11.7%
2020-05-13	4,353,211	297,539	1,544,372	2,511,300	84,964	5,220	55,842	6.8%	16.2%	8.5%
2020-05-12	4,268,247	292,319	1,488,530	2,487,398	83,409	5,622	37,035	6.8%	16.4%	13.2%
2020-05-11	4,184,838	286,697	1,451,495	2,446,646	76,568	3,611	46,994	6.9%	16.5%	7.1%
2020-05-10	4,108,270	283,086	1,404,501	2,420,683	77,907	3,425	33,594	6.9%	16.8%	9.3%
2020-05-09	4,030,363	279,661	1,370,907	2,379,795	85,361	4,411	53,549	6.9%	16.9%	7.6%
2020-05-08	3,945,002	275,250	1,317,358	2,352,394	93,107	5,345	36,522	7.0%	17.3%	12.8%
2020-05-07	3,851,895	269,905	1,280,836	2,301,154	91,059	5,709	39,497	7.0%	17.4%	12.6%
2020-05-06	3,760,836	264,196	1,241,339	2,255,301	92,201	6,584	45,918	7.0%	17.5%	12.5%
2020-05-05	3,668,635	257,612	1,195,421	2,215,602	80,761	5,722	36,605	7.0%	17.7%	13.5%
2020-05-04	3,587,874	251,890	1,158,816	2,177,168	76,717	4,093	34,064	7.0%	17.9%	10.7%
2020-05-03	3,511,157	247,797	1,124,752	2,138,608	79,107	3,668	32,323	7.1%	18.1%	10.2%
2020-05-02	3,432,050	244,129	1,092,429	2,095,492	82,135	5,187	40,917	7.1%	18.3%	11.3%
2020-05-01	3,349,915	238,942	1,051,512	2,059,461	88,465	5,255	38,231	7.1%	18.5%	12.1%
2020-04-30	3,261,450	233,687	1,013,281	2,014,482	84,854	5,695	64,971	7.2%	18.7%	8.1%
2020-04-29	3,176,596	227,992	948,310	2,000,294	75,518	6,883	42,168	7.2%	19.4%	14.0%
2020-04-28	3,101,078	221,109	906,142	1,973,827	73,863	6,362	33,266	7.1%	19.6%	16.1%
2020-04-27	3,027,215	214,747	872,876	1,939,592	68,863	4,555	27,803	7.1%	19.7%	14.1%
2020-04-26	2,958,352	210,192	845,073	1,903,087	73,932	3,733	28,603	7.1%	19.9%	11.5%
2020-04-25	2,884,420	206,459	816,470	1,861,491	85,356	6,193	27,779	7.2%	20.2%	18.2%
2020-04-24	2,799,064	200,266	788,691	1,810,107	87,429	6,340	50,033	7.2%	20.3%	11.2%
2020-04-23	2,711,635	193,926	738,658	1,779,051	88,885	6,752	28,791	7.2%	20.8%	19.0%
2020-04-22	2,622,750	187,174	709,867	1,725,709	75,845	6,699	30,429	7.1%	20.9%	18.0%
2020-04-21	2,546,905	180,475	679,438	1,686,992	74,641	7,094	34,827	7.1%	21.0%	16.9%
2020-04-20	2,472,264	173,381	644,611	1,654,272	73,841	5,398	22,002	7.0%	21.2%	19.7%
2020-04-19	2,398,423	167,983	622,609	1,607,831	82,350	4,531	31,654	7.0%	21.2%	12.5%
2020-04-18	2,316,073	163,452	590,955	1,561,666	73,536	6,430	23,923	7.1%	21.7%	21.2%
2020-04-17	2,242,537	157,022	567,032	1,518,483	88,959	8,865	26,120	7.0%	21.7%	25.3%
2020-04-16	2,153,578	148,157	540,912	1,464,509	95,994	7,271	30,828	6.9%	21.5%	19.1%
2020-04-15	2,057,584	140,886	510,084	1,406,614	80,297	8,265	36,667	6.8%	21.6%	18.4%
2020-04-14	1,977,287	132,621	473,417	1,371,249	70,595	6,891	25,093	6.7%	21.9%	21.5%
2020-04-13	1,906,692	125,730	448,324	1,332,638	70,077	5,723	27,166	6.6%	21.9%	17.4%
2020-04-12	1,836,615	120,007	421,158	1,295,450	98,802	5,700	19,403	6.5%	22.2%	22.7%
2020-04-11	1,737,813	114,307	401,755	1,221,751	78,139	6,021	26,247	6.6%	22.1%	18.7%

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
2020-04-10	1,659,674	108,286	375,508	1,175,880	92,251	7,243	21,819	6.5%	22.4%	24.9%
2020-04-09	1,567,423	101,043	353,689	1,112,691	85,933	7,586	25,336	6.4%	22.2%	23.0%
2020-04-08	1,481,490	93,457	328,353	1,059,680	83,953	6,692	28,716	6.3%	22.2%	18.9%
2020-04-07	1,397,537	86,765	299,637	1,011,135	74,939	7,903	23,388	6.2%	22.5%	25.3%
2020-04-06	1,322,598	78,862	276,249	967,487	71,475	5,831	16,633	6.0%	22.2%	26.0%
2020-04-05	1,251,123	73,031	259,616	918,476	73,676	4,987	13,839	5.8%	22.0%	26.5%
2020-04-04	1,177,447	68,044	245,777	863,626	80,254	5,831	20,413	5.8%	21.7%	22.2%
2020-04-03	1,097,193	62,213	225,364	809,616	82,480	5,962	15,447	5.7%	21.6%	27.8%
2020-04-02	1,014,713	56,251	209,917	748,545	80,808	6,283	17,041	5.5%	21.1%	26.9%
2020-04-01	933,905	49,968	192,876	691,061	75,588	5,528	15,090	5.4%	20.6%	26.8%
2020-03-31	858,317	44,440	177,786	636,091	74,737	4,836	13,486	5.2%	20.0%	26.4%
2020-03-30	783,580	39,604	164,300	579,676	62,885	4,148	15,437	5.1%	19.4%	21.2%
2020-03-29	720,695	35,456	148,863	536,376	59,151	3,466	9,467	4.9%	19.2%	26.8%
2020-03-28	661,544	31,990	139,396	490,158	67,366	3,682	8,494	4.8%	18.7%	30.2%
2020-03-27	594,178	28,308	130,902	434,968	64,040	3,508	8,769	4.8%	17.8%	28.6%
2020-03-26	530,138	24,800	122,133	383,205	61,983	3,001	8,365	4.7%	16.9%	26.4%
2020-03-25	468,155	21,799	113,768	332,588	49,586	2,773	5,783	4.7%	16.1%	32.4%
2020-03-24	418,569	19,026	107,985	291,558	40,188	2,268	9,639	4.5%	15.0%	19.0%
2020-03-23	378,381	16,758	98,346	263,277	40,784	1,918	465	4.4%	14.6%	80.5%
2020-03-22	337,597	14,840	97,881	224,876	32,753	1,699	6,215	4.4%	13.2%	21.5%
2020-03-21	304,844	13,141	91,666	200,037	32,146	1,702	4,264	4.3%	12.5%	28.5%
2020-03-20	272,698	11,439	87,402	173,857	29,614	1,481	2,445	4.2%	11.6%	37.7%
2020-03-19	243,084	9,958	84,957	148,169	26,923	1,106	1,637	4.1%	10.5%	40.3%
2020-03-18	216,161	8,852	83,320	123,989	19,244	895	2,483	4.1%	9.6%	26.5%
2020-03-17	196,917	7,957	80,837	108,123	15,465	806	2,752	4.0%	9.0%	22.7%
2020-03-16	181,452	7,151	78,085	96,216	14,430	680	2,054	3.9%	8.4%	24.9%
2020-03-15	167,022	6,471	76,031	84,520	10,739	640	3,410	3.9%	7.8%	15.8%
2020-03-14	156,283	5,831	72,621	77,831	11,079	420	2,371	3.7%	7.4%	15.0%
2020-03-13	145,204	5,411	70,250	69,543	14,295	498	1,927	3.7%	7.2%	20.5%
2020-03-12	130,909	4,913	68,323	57,673	5,205	303	1,321	3.8%	6.7%	18.7%
2020-03-11	125,704	4,610	67,002	54,092	7,329	347	2,598	3.7%	6.4%	11.8%
2020-03-10	118,375	4,263	64,404	49,708	4,839	276	1,911	3.6%	6.2%	12.6%
2020-03-09	113,536	3,987	62,493	47,056	3,782	186	1,799	3.5%	6.0%	9.4%
2020-03-08	109,754	3,801	60,694	45,259	3,972	243	2,335	3.5%	5.9%	9.4%
2020-03-07	105,782	3,558	58,359	43,865	4,021	99	2,494	3.4%	5.7%	3.8%
2020-03-06	101,761	3,459	55,865	42,437	3,917	112	2,069	3.4%	5.8%	5.1%
2020-03-05	97,844	3,347	53,796	40,701	2,769	93	2,626	3.4%	5.9%	3.4%
2020-03-04	95,075	3,254	51,170	40,651	2,280	94	2,942	3.4%	6.0%	3.1%
2020-03-03	92,795	3,160	48,228	41,407	2,533	75	2,626	3.4%	6.1%	2.8%
2020-03-02	90,262	3,085	45,602	41,575	1,937	89	2,886	3.4%	6.3%	3.0%
2020-03-01	88,325	2,996	42,716	42,613	2,358	55	2,934	3.4%	6.6%	1.8%
2020-02-29	85,967	2,941	39,782	43,244	1,897	69	3,071	3.4%	6.9%	2.2%
2020-02-28	84,070	2,872	36,711	44,487	1,366	58	3,434	3.4%	7.3%	1.7%
2020-02-27	82,704	2,814	33,277	46,613	1,358	44	2,893	3.4%	7.8%	1.5%
2020-02-26	81,346	2,770	30,384	48,192	974	62	2,479	3.4%	8.4%	2.4%
2020-02-25	80,372	2,708	27,905	49,759	847	79	2,678	3.4%	8.8%	2.9%
2020-02-24	79,525	2,629	25,227	51,669	567	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%	NA%

Appendix A.2 Latest Cases by Country

```
## hightlight high death rates (if >= 5%) for those countries with 2000+ confirmed cases
data.latest.all %>% arrange(desc(confirmed)) %>% select(-c(date, ranking)) %>%
  mutate(death.rate=ifelse(confirmed >= 2000 & 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 (15 Jun 2020 UTC)

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	World	8,034,461	122,035	3,740,224	3,857,338	436,899	3,508	**5.4%**
2	US	2,114,026	19,968	1,421,565	576,334	116,127	395	**5.5%**
3	Brazil	888,271	20,647	366,603	477,709	43,959	627	4.9%
4	Russia	536,484	8,217	245,382	284,021	7,081	143	1.3%
5	India	343,091	10,667	153,178	180,013	9,900	380	2.9%
6	United Kingdom	298,315	973	255,210	1,284	41,821	38	**14.0%**
7	Spain	244,109	181	66,597	150,376	27,136	0	**11.1%**
8	Italy	237,290	301	25,909	177,010	34,371	26	**14.5%**
9	Peru	232,992	3,256	106,723	119,409	6,860	172	2.9%
10	France	194,305	152	91,698	73,168	29,439	29	**15.2%**
11	Iran	189,876	2,449	30,336	150,590	8,950	113	4.7%
12	Germany	187,682	164	6,183	172,692	8,807	6	4.7%
13	Turkey	179,831	1,592	22,642	152,364	4,825	18	2.7%
14	Chile	179,436	5,143	27,282	148,792	3,362	39	1.9%
15	Mexico	150,264	3,427	20,392	112,292	17,580	439	**11.7%**
16	Pakistan	148,921	4,443	89,692	56,390	2,839	110	1.9%
17	Saudi Arabia	132,048	4,507	43,147	87,890	1,011	39	0.8%
18	Canada	100,763	359	31,069	61,466	8,228	10	**8.2%**
19	Bangladesh	90,619	3,099	70,679	18,731	1,209	38	1.3%
20	China	84,378	43	251	79,489	4,638	0	**5.5%**
21	Qatar	80,876	1,274	22,119	58,681	76	3	0.1%
22	South Africa	73,533	3,495	32,098	39,867	1,568	88	2.1%
23	Belgium	60,100	71	33,829	16,610	9,661	6	**16.1%**
24	Belarus	54,680	707	23,948	30,420	312	4	0.6%
25	Colombia	53,168	4,272	31,375	19,985	1,808	138	3.4%
26	Sweden	52,383	769	47,492	0	4,891	17	**9.3%**
27	Netherlands	49,155	165	42,885	186	6,084	6	**12.4%**
28	Ecuador	47,322	571	20,044	23,349	3,929	33	**8.3%**
29	Egypt	46,289	1,691	32,288	12,329	1,672	97	3.6%
30	United Arab Emirates	42,636	342	14,216	28,129	291	2	0.7%
31	Singapore	40,818	214	10,426	30,366	26	0	0.1%
32	Indonesia	39,294	1,017	21,973	15,123	2,198	64	**5.6%**
33	Portugal	37,036	346	12,664	22,852	1,520	3	4.1%
34	Kuwait	36,431	511	8,602	27,531	298	2	0.8%
35	Argentina	32,785	1,208	22,040	9,891	854	21	2.6%
36	Ukraine	32,536	685	16,854	14,771	911	12	2.8%
37	Switzerland	31,131	14	292	28,900	1,939	1	**6.2%**
38	Poland	29,788	396	14,149	14,383	1,256	9	4.2%
39	Philippines	26,420	490	19,070	6,252	1,098	10	4.2%
40	Afghanistan	25,527	761	19,885	5,164	478	7	1.9%
41	Ireland	25,321	18	917	22,698	1,706	0	**6.7%**
42	Oman	24,524	1,043	14,883	9,533	108	4	0.4%
43	Dominican Republic	23,271	309	8,641	14,025	605	13	2.6%
44	Romania	22,165	166	4,921	15,817	1,427	17	**6.4%**
45	Panama	21,422	4	7,208	13,766	448	11	2.1%
46	Iraq	21,315	1,106	11,392	9,271	652	45	3.1%
47	Israel	19,237	182	3,520	15,415	302	2	1.6%
48	Bolivia	19,073	614	15,011	3,430	632	21	3.3%
49	Bahrain	19,013	786	5,700	13,267	46	4	0.2%
50	Japan	17,439	70	943	15,567	929	2	**5.3%**

Table 6: Cases by Country (15 Jun 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
51	Austria	17,135	26	391	16,066	678	1	4.0%
52	Armenia	17,064	397	10,503	6,276	285	16	1.7%
53	Nigeria	16,658	573	10,885	5,349	424	4	2.5%
54	Kazakhstan	15,192	696	5,723	9,388	81	4	0.5%
55	Denmark	12,417	24	529	11,290	598	1	4.8%
56	Serbia	12,367	57	551	11,561	255	1	2.1%
57	Korea, South	12,155	34	1,117	10,760	278	1	2.3%
58	Ghana	11,964	0	7,652	4,258	54	0	0.5%
59	Moldova	11,879	139	4,674	6,794	411	5	3.5%
60	Algeria	11,031	112	2,519	7,735	777	10	**7.0%**
61	Azerbaijan	10,324	367	4,463	5,739	122	3	1.2%
62	Guatemala	10,272	427	7,907	1,966	399	15	3.9%
63	Czechia	10,064	40	2,438	7,296	330	1	3.3%
64	Cameroon	9,864	1,183	4,018	5,570	276	64	2.8%
65	Honduras	9,178	320	7,831	1,025	322	10	3.5%
66	Morocco	8,885	92	845	7,828	212	0	2.4%
67	Norway	8,647	16	267	8,138	242	0	2.8%
68	Malaysia	8,494	41	973	7,400	121	0	1.4%
69	Sudan	7,435	215	4,247	2,720	468	9	**6.3%**
70	Australia	7,347	12	389	6,856	102	0	1.4%
71	Finland	7,108	4	582	6,200	326	0	4.6%
72	Nepal	6,211	451	5,151	1,041	19	0	0.3%
73	Cote d'Ivoire	5,439	355	2,803	2,590	46	1	0.8%
74	Uzbekistan	5,263	183	1,225	4,019	19	0	0.4%
75	Senegal	5,173	83	1,685	3,424	64	4	1.2%
76	Tajikistan	5,097	62	1,544	3,503	50	0	1.0%
77	Congo (Kinshasa)	4,837	59	4,112	613	112	5	2.3%
78	Guinea	4,572	40	1,287	3,259	26	1	0.6%
79	Djibouti	4,501	36	1,275	3,183	43	0	1.0%
80	Haiti	4,441	276	4,341	24	76	6	1.7%
81	North Macedonia	4,157	100	2,241	1,723	193	5	4.6%
82	Hungary	4,076	7	1,028	2,485	563	1	**13.8%**
83	Luxembourg	4,072	2	31	3,931	110	0	2.7%
84	Gabon	4,033	570	2,672	1,334	27	4	0.7%
85	El Salvador	3,826	106	1,826	1,926	74	0	1.9%
86	Kenya	3,727	133	2,337	1,286	104	1	2.8%
87	Ethiopia	3,521	176	2,841	620	60	3	1.7%
88	Bulgaria	3,341	51	1,381	1,784	176	2	**5.3%**
89	Thailand	3,135	0	90	2,987	58	0	1.9%
90	Greece	3,134	13	1,576	1,374	184	1	**5.9%**
91	Venezuela	3,062	84	2,201	835	26	1	0.8%
92	Bosnia and Herzegovina	3,040	147	713	2,162	165	2	**5.4%**
93	Somalia	2,642	24	1,932	622	88	0	3.3%
94	Kyrgyzstan	2,472	187	597	1,847	28	1	1.1%
95	Cuba	2,262	14	213	1,965	84	0	3.7%
96	Croatia	2,254	2	7	2,140	107	0	4.7%
97	Central African Republic	2,222	165	1,846	369	7	0	0.3%
98	Maldives	2,065	30	517	1,540	8	0	0.4%
99	Estonia	1,974	1	188	1,717	69	0	3.5%
100	Sri Lanka	1,905	16	552	1,342	11	0	0.6%
101	Mauritania	1,887	104	1,436	360	91	4	4.8%
102	Mali	1,860	51	631	1,125	104	0	5.6%
103	Iceland	1,810	0	4	1,796	10	0	0.6%
104	Lithuania	1,773	5	268	1,429	76	1	4.3%
105	Costa Rica	1,744	29	961	771	12	0	0.7%
106	South Sudan	1,693	0	1,617	49	27	0	1.6%
107	Albania	1,590	69	499	1,055	36	0	2.3%
108	Slovakia	1,552	4	114	1,410	28	0	1.8%
109	New Zealand	1,506	2	2	1,482	22	0	1.5%
110	Slovenia	1,496	1	28	1,359	109	0	7.3%
111	Guinea-Bissau	1,492	32	1,324	153	15	0	1.0%
112	Kosovo	1,486	49	500	953	33	1	2.2%
113	Lebanon	1,464	18	557	875	32	0	2.2%
114	Nicaragua	1,464	0	456	953	55	0	3.8%
115	Zambia	1,382	24	229	1,142	11	0	0.8%
116	Equatorial Guinea	1,306	0	1,094	200	12	0	0.9%
117	Paraguay	1,296	7	611	673	12	1	0.9%
118	Madagascar	1,290	18	896	384	10	0	0.8%

Table 6: Cases by Country (15 Jun 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
119	Sierra Leone	1,176	7	442	683	51	0	4.3%
120	Tunisia	1,110	14	62	999	49	0	4.4%
121	Latvia	1,097	0	224	845	28	0	2.6%
122	Cyprus	985	2	160	807	18	0	1.8%
123	Niger	980	0	29	885	66	0	6.7%
124	Jordan	979	18	278	692	9	0	0.9%
125	Burkina Faso	894	0	37	804	53	0	5.9%
126	Congo (Brazzaville)	883	155	465	391	27	3	3.1%
127	Georgia	879	15	161	704	14	0	1.6%
128	Andorra	853	0	13	789	51	0	6.0%
129	Chad	850	0	57	720	73	0	8.6%
130	Uruguay	848	0	33	792	23	0	2.7%
131	Yemen	844	116	557	79	208	44	24.6%
132	Cabo Verde	760	10	413	340	7	1	0.9%
133	Diamond Princess	712	0	48	651	13	0	1.8%
134	Uganda	705	9	406	299	0	0	0.0%
135	San Marino	694	0	77	575	42	0	6.1%
136	Sao Tome and Principe	662	1	473	177	12	0	1.8%
137	Malta	650	1	38	603	9	0	1.4%
138	Jamaica	621	4	181	430	10	0	1.6%
139	Rwanda	612	30	272	338	2	0	0.3%
140	Mozambique	609	26	449	157	3	0	0.5%
141	Malawi	555	8	480	69	6	0	1.1%
142	Togo	531	1	219	299	13	0	2.4%
143	Tanzania	509	0	305	183	21	0	4.1%
144	Eswatini	506	16	253	249	4	0	0.8%
145	West Bank and Gaza	505	13	87	415	3	0	0.6%
146	Liberia	498	40	244	221	33	1	6.6%
147	Benin	483	41	242	232	9	3	1.9%
148	Libya	467	13	387	70	10	0	2.1%
149	Taiwan*	445	2	5	433	7	0	1.6%
150	Zimbabwe	387	4	329	54	4	0	1.0%
151	Mauritius	337	0	2	325	10	0	3.0%
152	Vietnam	334	0	11	323	0	0	0.0%
153	Montenegro	326	1	2	315	9	0	2.8%
154	Burma	262	1	81	175	6	0	2.3%
155	Suriname	229	21	176	48	5	2	2.2%
156	Mongolia	197	0	88	109	0	0	0.0%
157	Syria	177	0	93	78	6	0	3.4%
158	Comoros	176	0	60	114	2	0	1.1%
159	Guyana	159	0	48	99	12	0	7.5%
160	Angola	142	2	72	64	6	0	4.2%
161	Brunei	141	0	1	138	2	0	1.4%
162	Cambodia	128	0	3	125	0	0	0.0%
163	Trinidad and Tobago	123	0	6	109	8	0	6.5%
164	Eritrea	109	13	70	39	0	0	0.0%
165	Bahamas	103	0	24	68	11	0	10.7%
166	Monaco	99	0	2	93	4	0	4.0%
167	Barbados	97	1	7	83	7	0	7.2%
168	Burundi	85	0	39	45	1	0	1.2%
169	Liechtenstein	82	0	26	55	1	0	1.2%
170	Bhutan	67	1	45	22	0	0	0.0%
171	Botswana	60	0	35	24	1	0	1.7%
172	Namibia	32	0	15	17	0	0	0.0%
173	Gambia	30	2	5	24	1	0	3.3%
174	Saint Vincent and the Grenadines	27	0	2	25	0	0	0.0%
175	Antigua and Barbuda	26	0	3	20	3	0	11.5%
176	Timor-Leste	24	0	0	24	0	0	0.0%
177	Grenada	23	0	1	22	0	0	0.0%
178	Belize	21	1	3	16	2	0	9.5%
179	Laos	19	0	0	19	0	0	0.0%
180	Saint Lucia	19	0	1	18	0	0	0.0%
181	Dominica	18	0	2	16	0	0	0.0%
182	Fiji	18	0	0	18	0	0	0.0%
183	Saint Kitts and Nevis	15	0	0	15	0	0	0.0%
184	Holy See	12	0	0	12	0	0	0.0%
185	Seychelles	11	0	0	11	0	0	0.0%
186	MS Zaandam	9	0	7	0	2	0	22.2%

Table 6: Cases by Country (15 Jun 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
187	Western Sahara	9	0	0	8	1	0	11.1%
188	Papua New Guinea	8	0	0	8	0	0	0.0%
189	Lesotho	4	0	2	2	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!