COVID-19 Data Analysis with R - Worldwide*

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

02 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] 266 136
```

Each dataset has 266 rows, corresponding to country/region/province/state. It has 136 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" "2020-06-01"
```

```
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 01 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 01 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
```

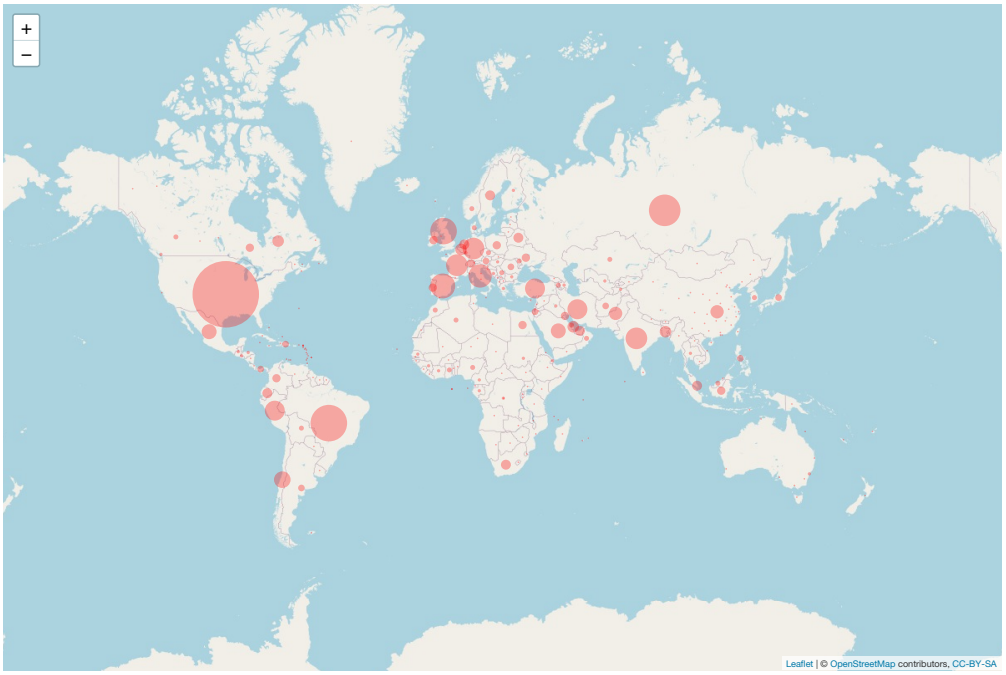


Figure 1: World Map

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)
```

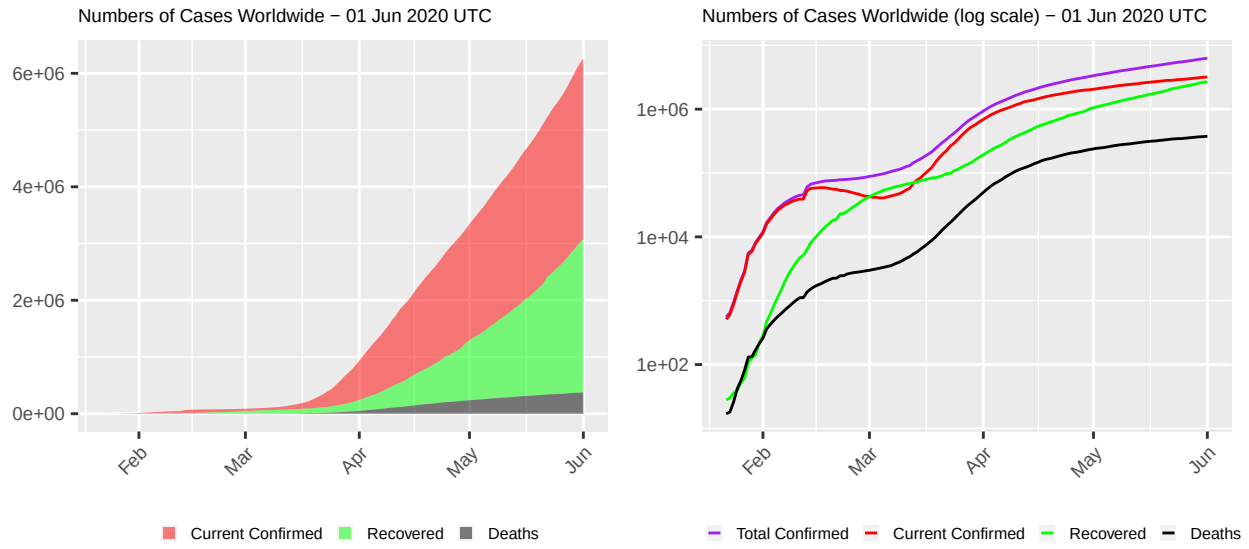


Figure 2: COVID-19 Cases Worldwide

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)
```

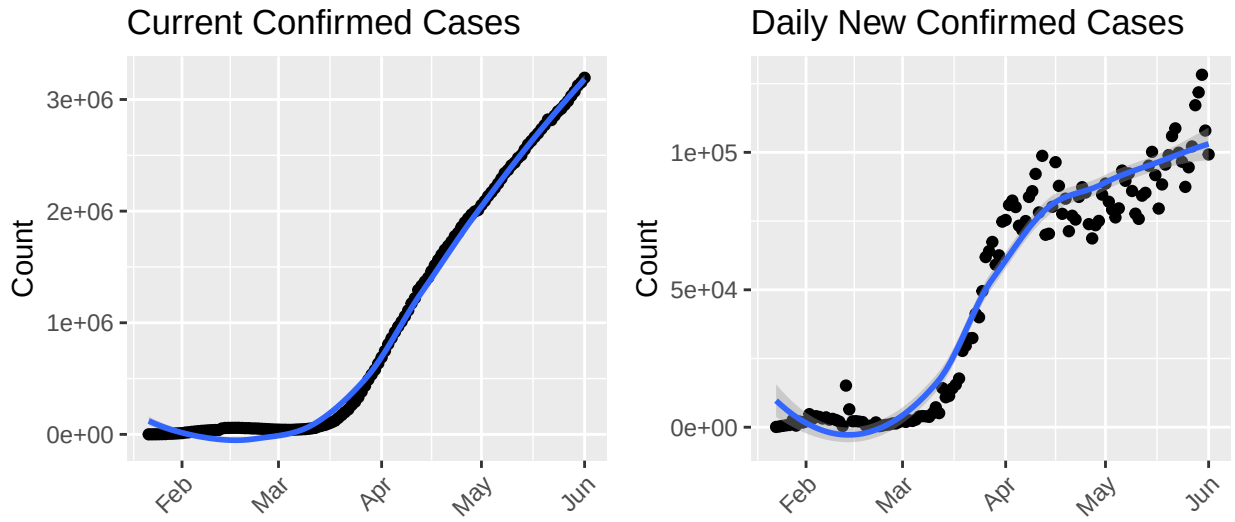


Figure 3: Current Confirmed Cases

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)
```

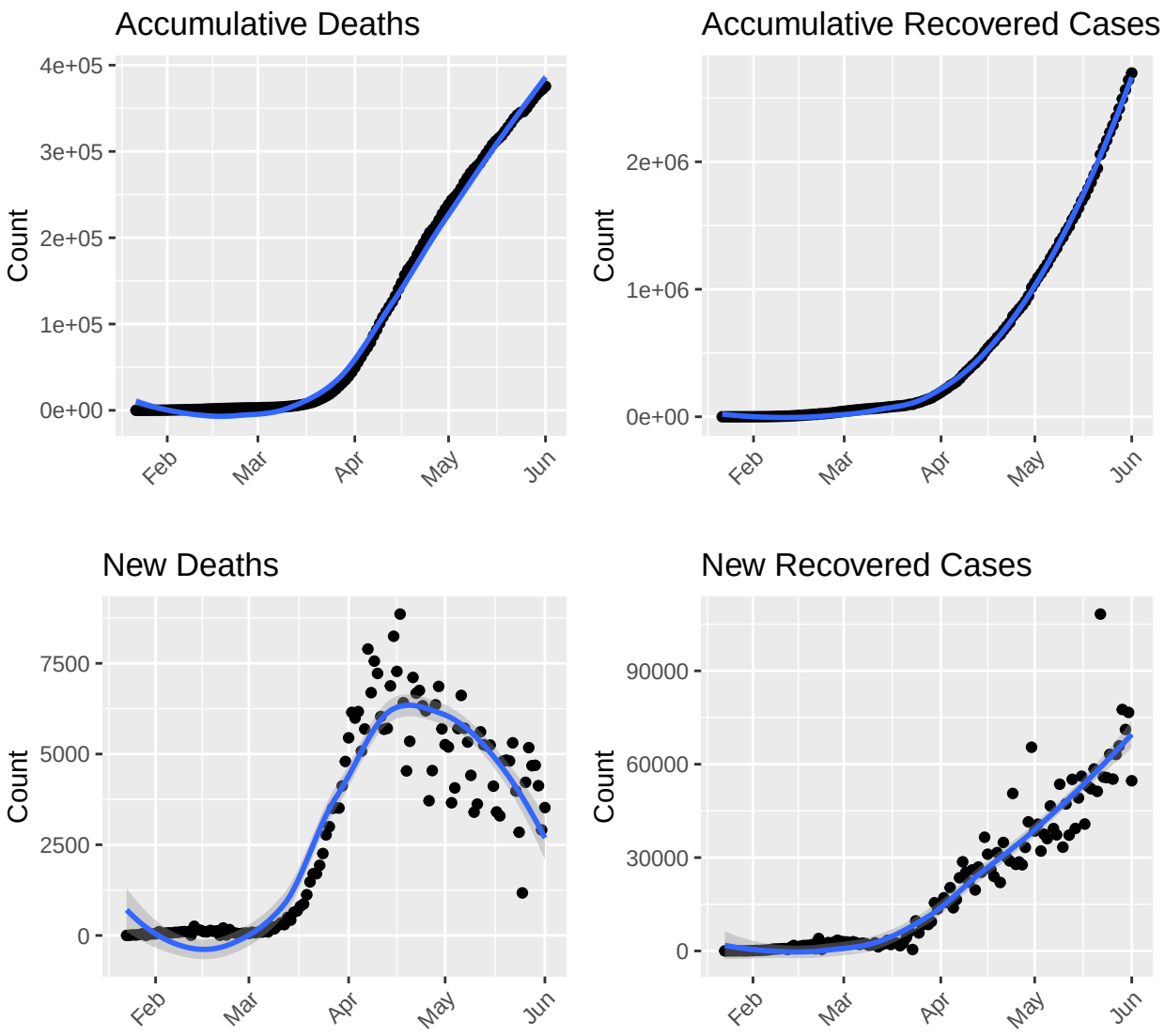


Figure 4: Deaths and Recovered Cases

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 01 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 01 Jun 2020 UTC UTC, it will be between 6% and 12.2%.

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)
```

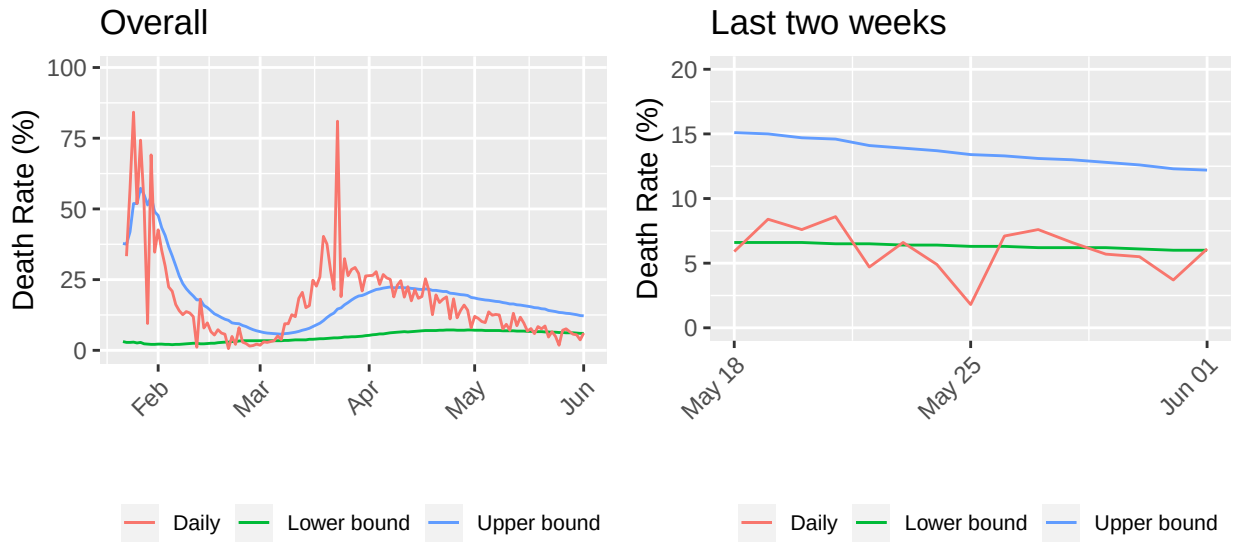


Figure 5: Death Rate

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"         "United Kingdom"
##  [5] "Spain"          "Italy"          "India"          "France"
##  [9] "Germany"        "Peru"           "Turkey"         "Iran"
## [13] "Chile"          "Mexico"         "Canada"         "Saudi Arabia"
## [17] "China"          "Pakistan"       "Belgium"        "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 - 01 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	6,266,192	375,559	6.0%	99,246	3,524	3,194,624
2	US	1,811,360	105,165	5.8%	21,188	784	1,247,964
3	Brazil	526,447	29,937	5.7%	11,598	623	285,430
4	Russia	414,328	4,849	1.2%	8,485	156	233,965
5	United Kingdom	277,736	39,127	14.1%	1,580	556	237,388
6	Spain	239,638	27,127	11.3%	159	0	62,135
7	Italy	233,197	33,475	14.4%	200	60	41,367
8	India	198,370	5,608	2.8%	7,761	200	97,008
9	France	189,348	28,836	15.2%	339	31	91,954
10	Germany	183,594	8,555	4.7%	184	15	9,407
11	Peru	170,039	4,634	2.7%	5,563	128	96,898
12	Turkey	164,769	4,563	2.8%	827	23	31,259
13	Iran	154,445	7,878	5.1%	2,979	81	25,563
14	Chile	105,158	1,113	1.1%	5,470	59	59,099
15	Mexico	93,435	10,167	10.9%	2,771	237	16,303
16	Canada	93,288	7,404	7.9%	809	30	35,793
17	Saudi Arabia	87,142	525	0.6%	1,881	22	22,311
18	China	84,154	4,638	5.5%	8	0	118
19	Pakistan	72,460	1,543	2.1%	2,964	60	44,834
20	Belgium	58,517	9,486	16.2%	136	19	33,112
21	Qatar	58,433	40	0.1%	1,523	2	24,956
22	Others	1,050,334	40,889	3.9%	22,821	443	497,760

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

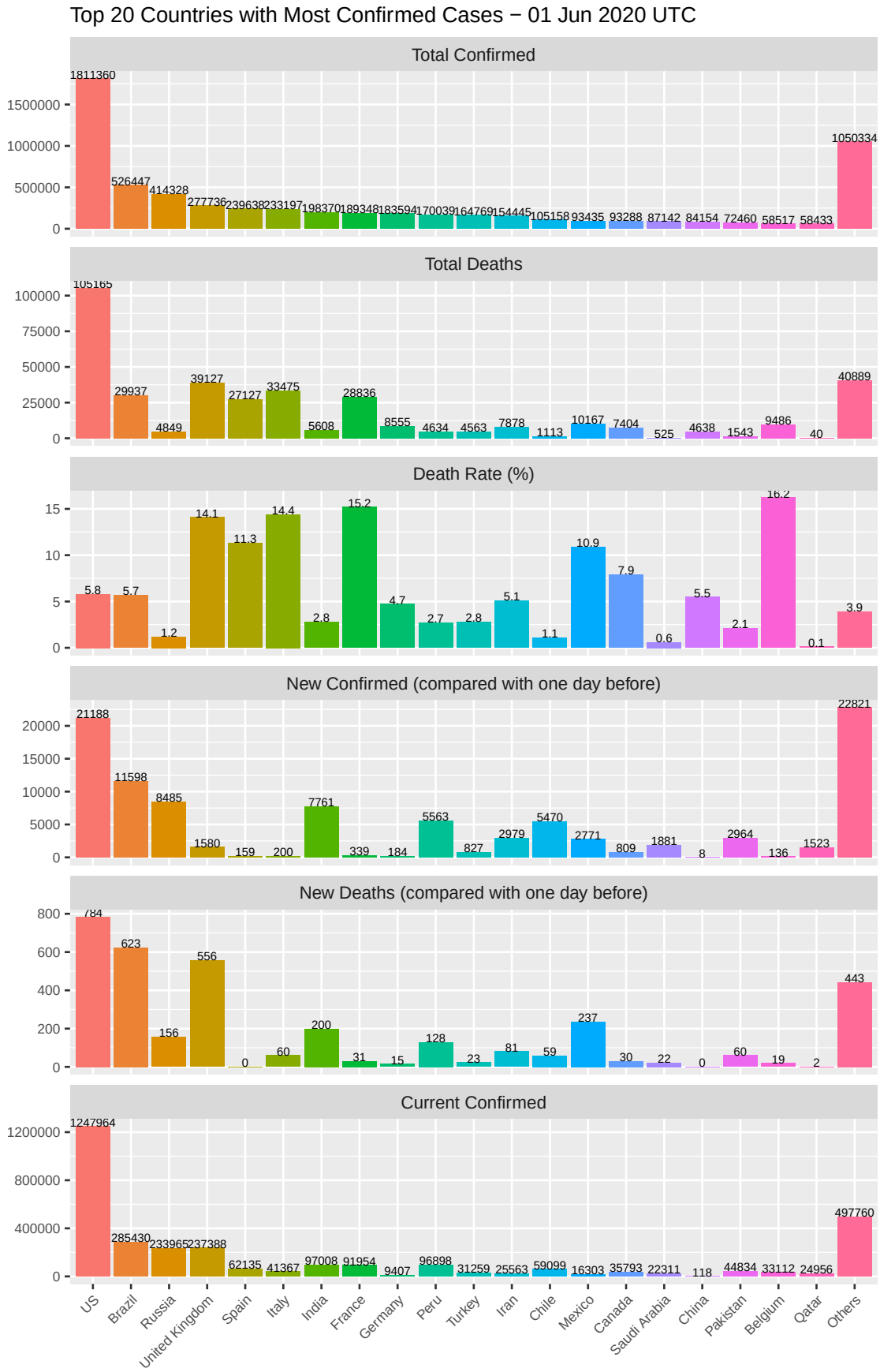
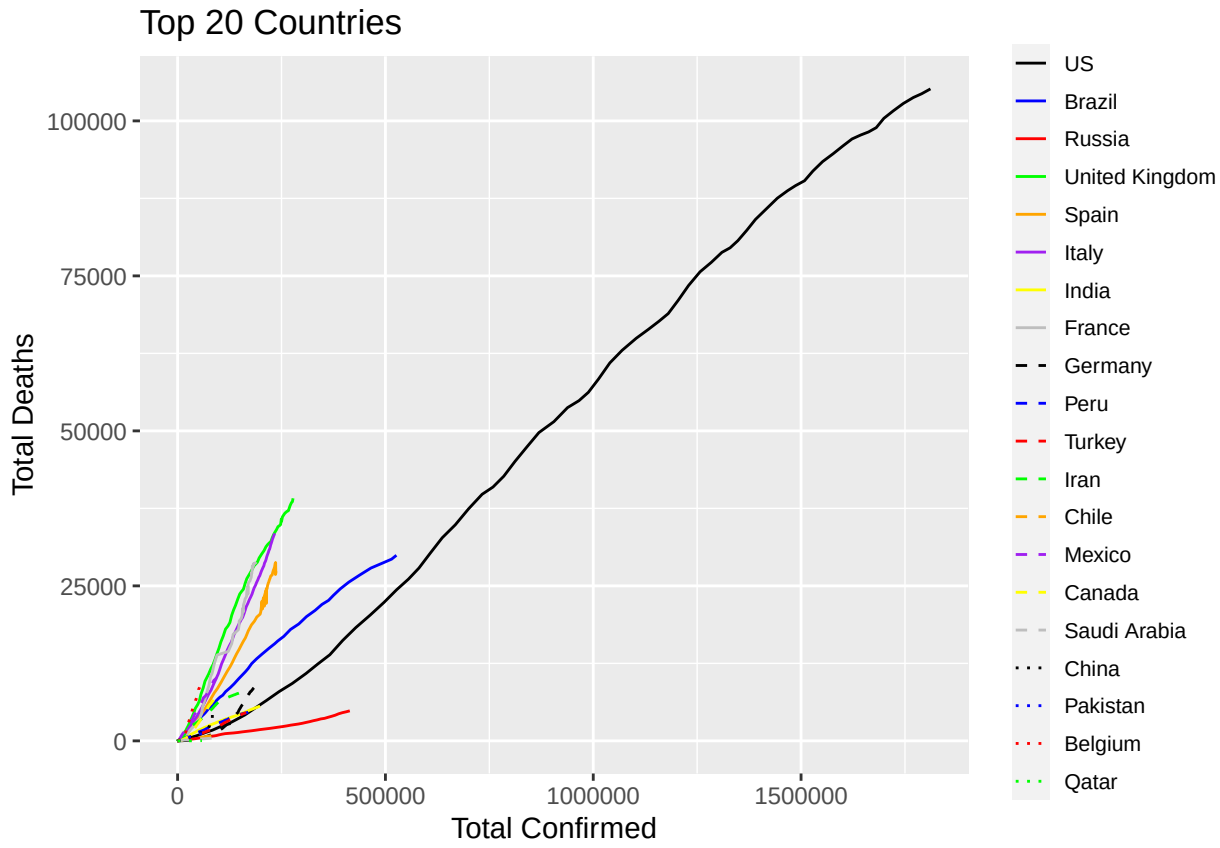


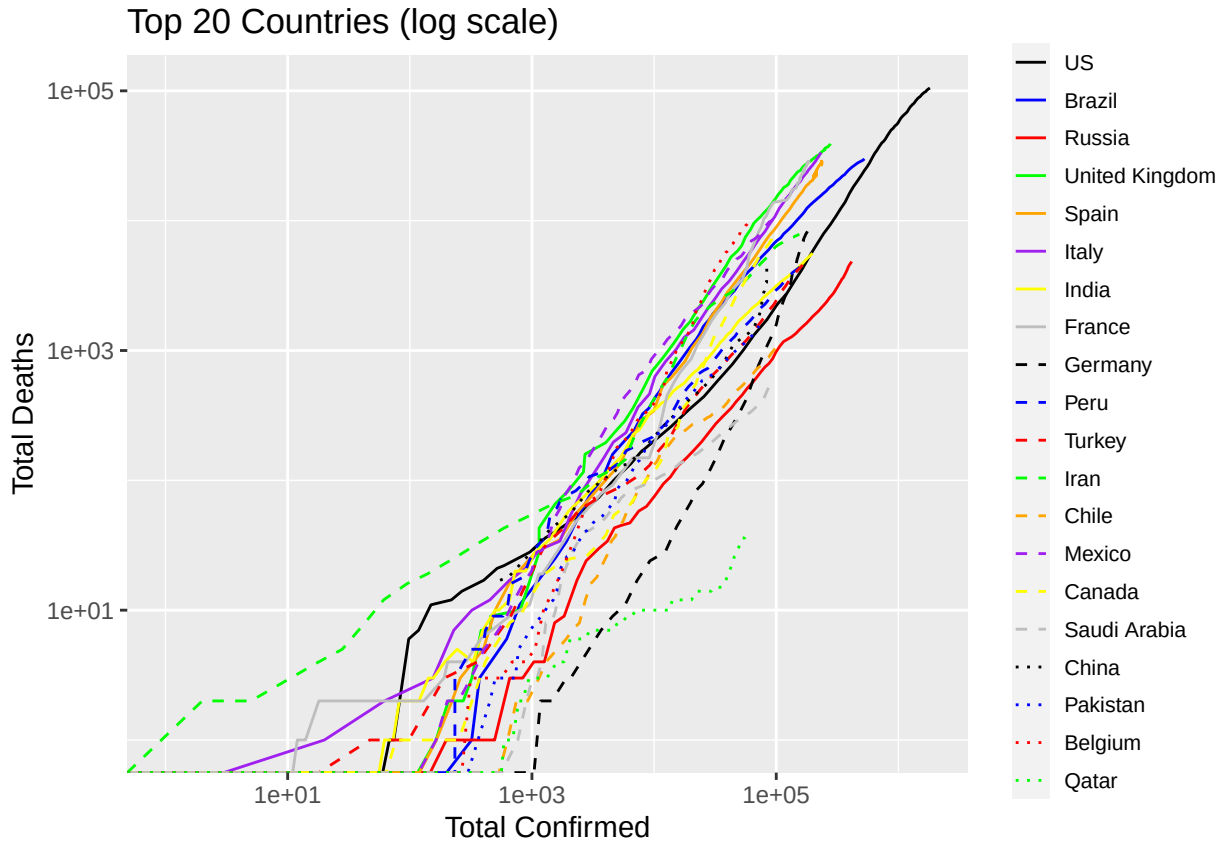
Figure 6: Top 20 Countries with Most Confirmed Cases

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 01 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)
```

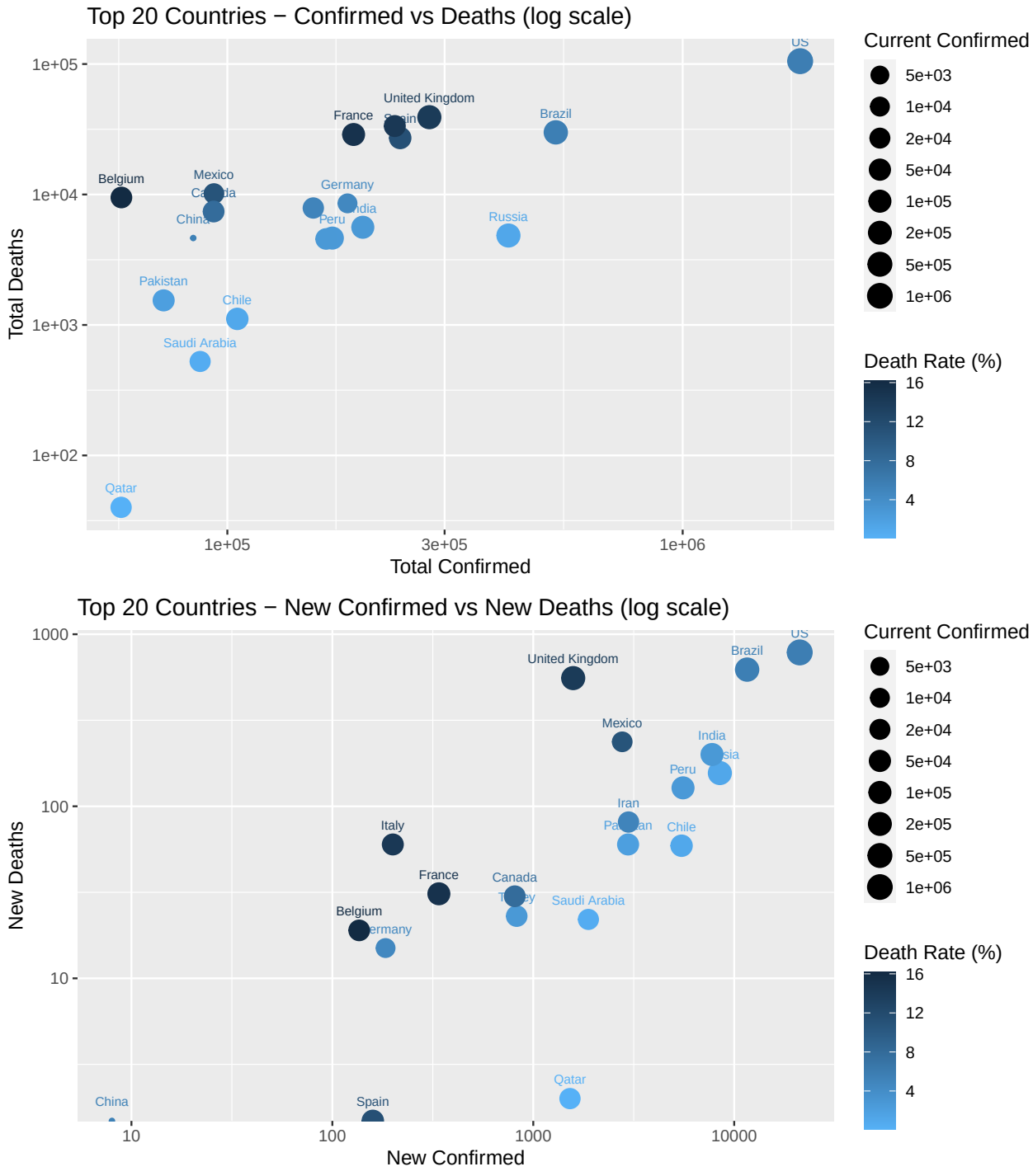


Figure 7: Top 20 Countries

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)
```

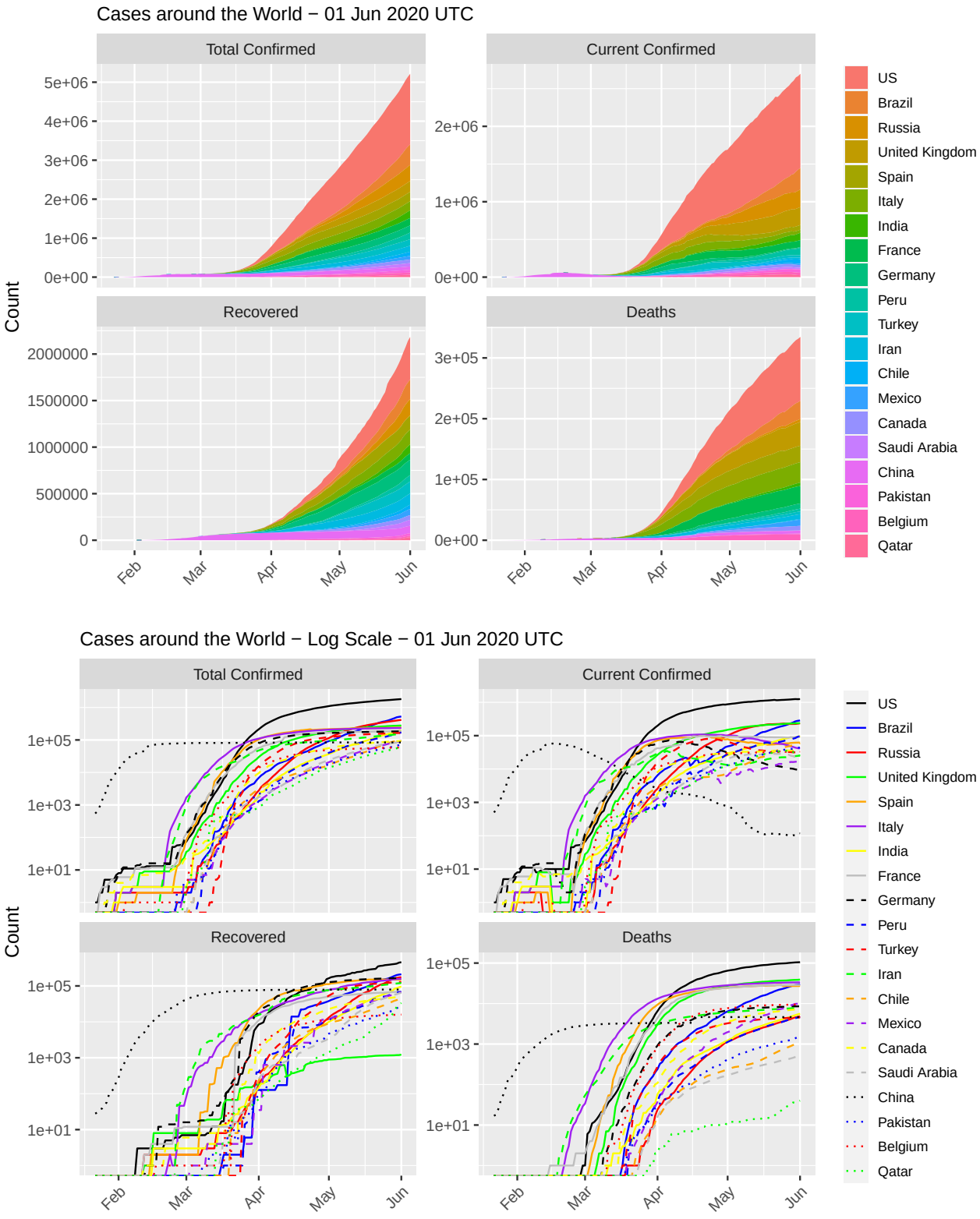


Figure 8: Cases around the World

```
## 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))
```

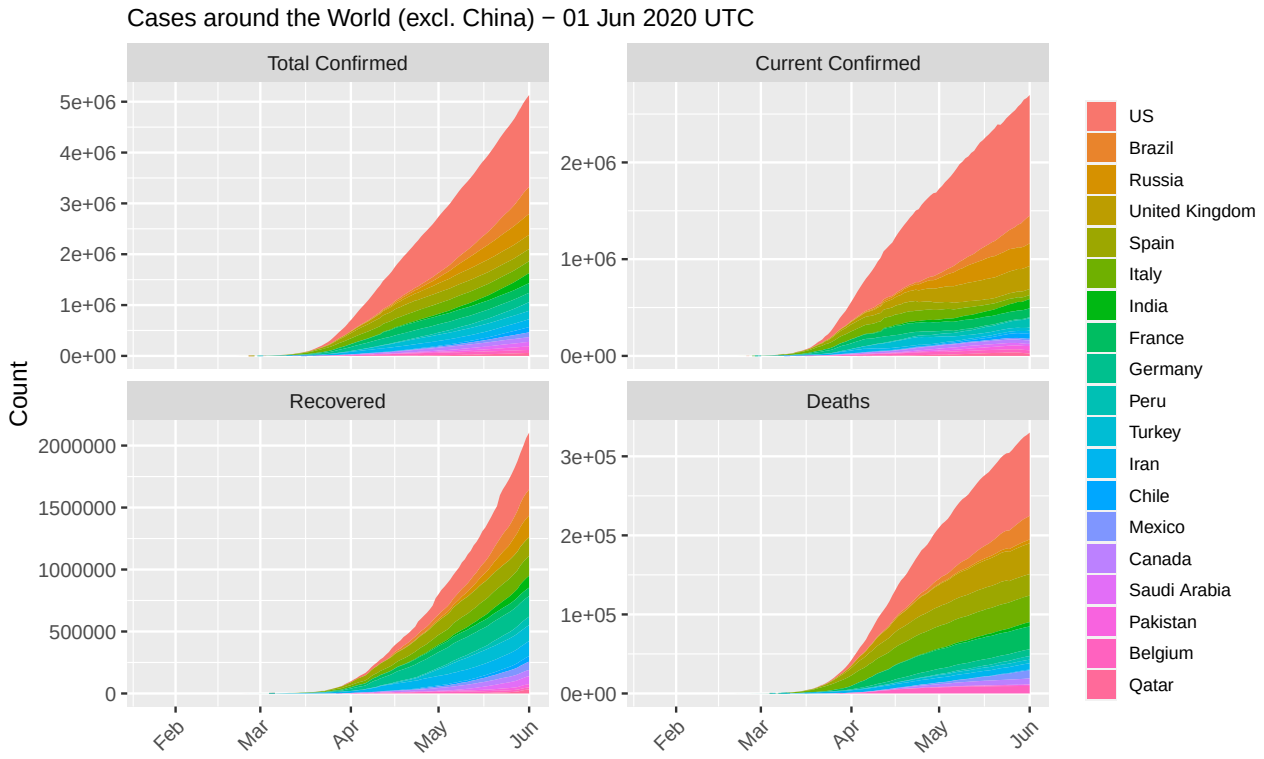


Figure 9: Cases around the World (excl. China)

```
## 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')
```

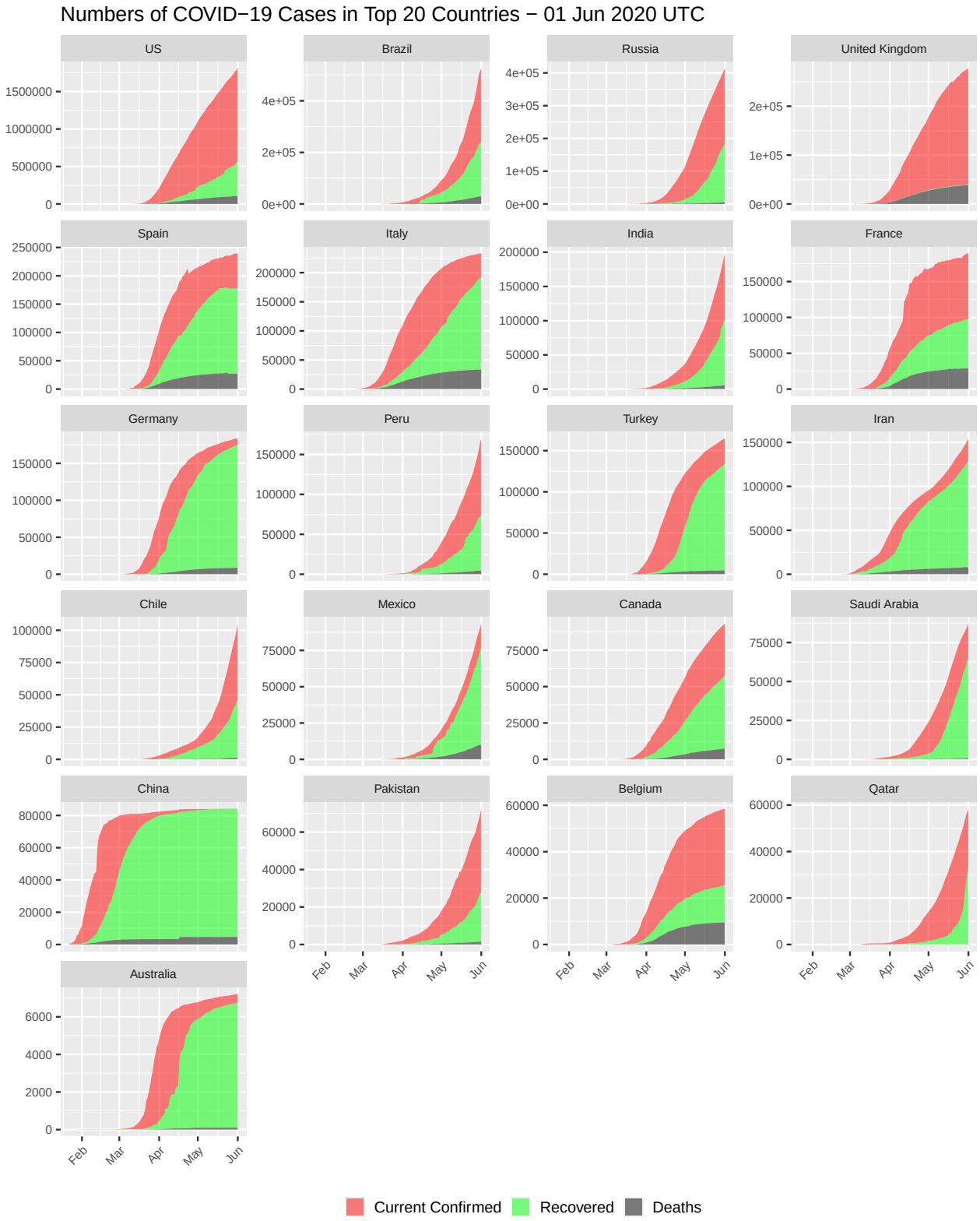


Figure 10: COVID-19 Cases in Top 20 Countries. Ordered descendingly by number of confirmed cases.

```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')
```

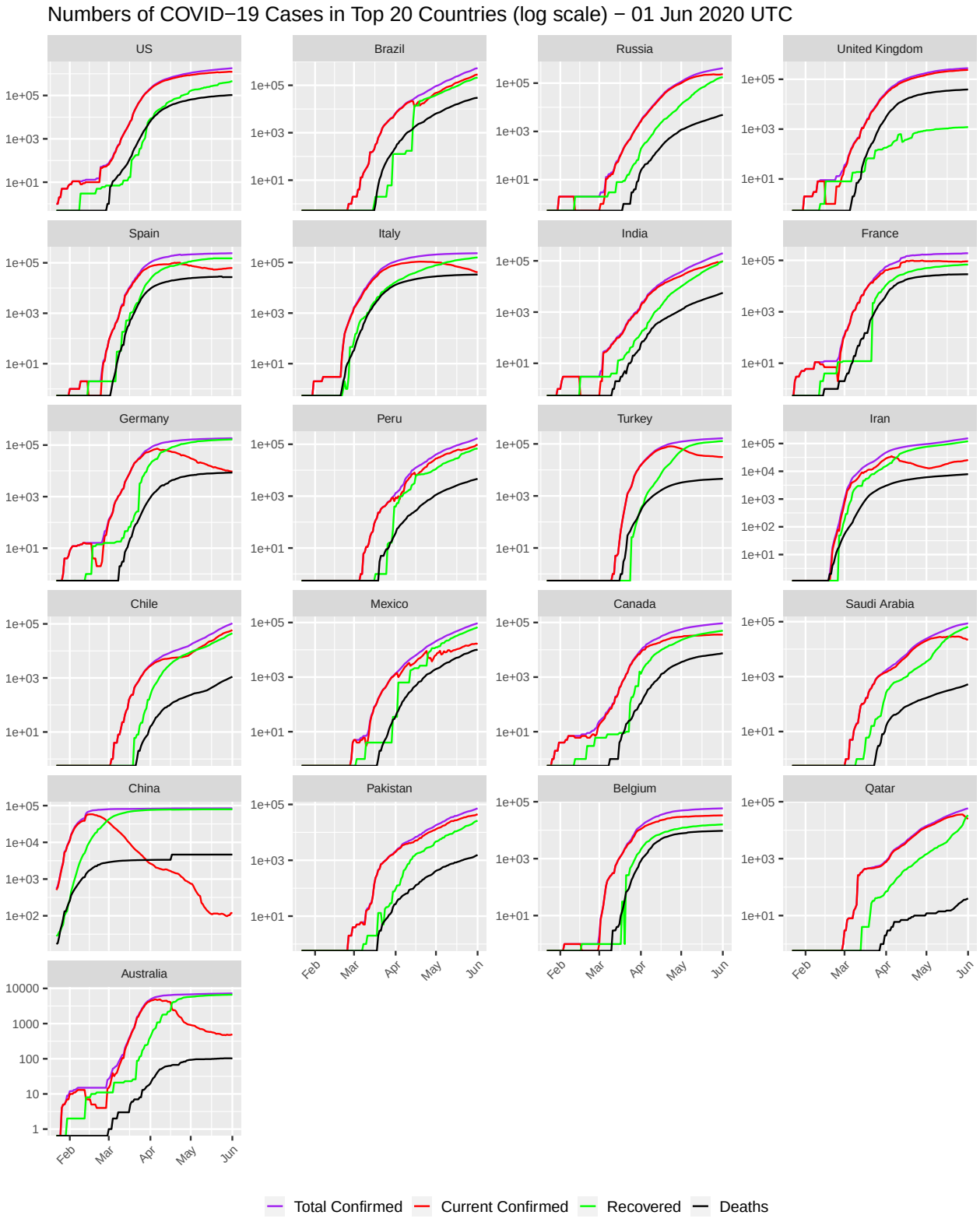


Figure 11: COVID-19 Cases Top 20 Countries (log scale). Ordered descendingly by number of confirmed cases.

```
## 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)
```

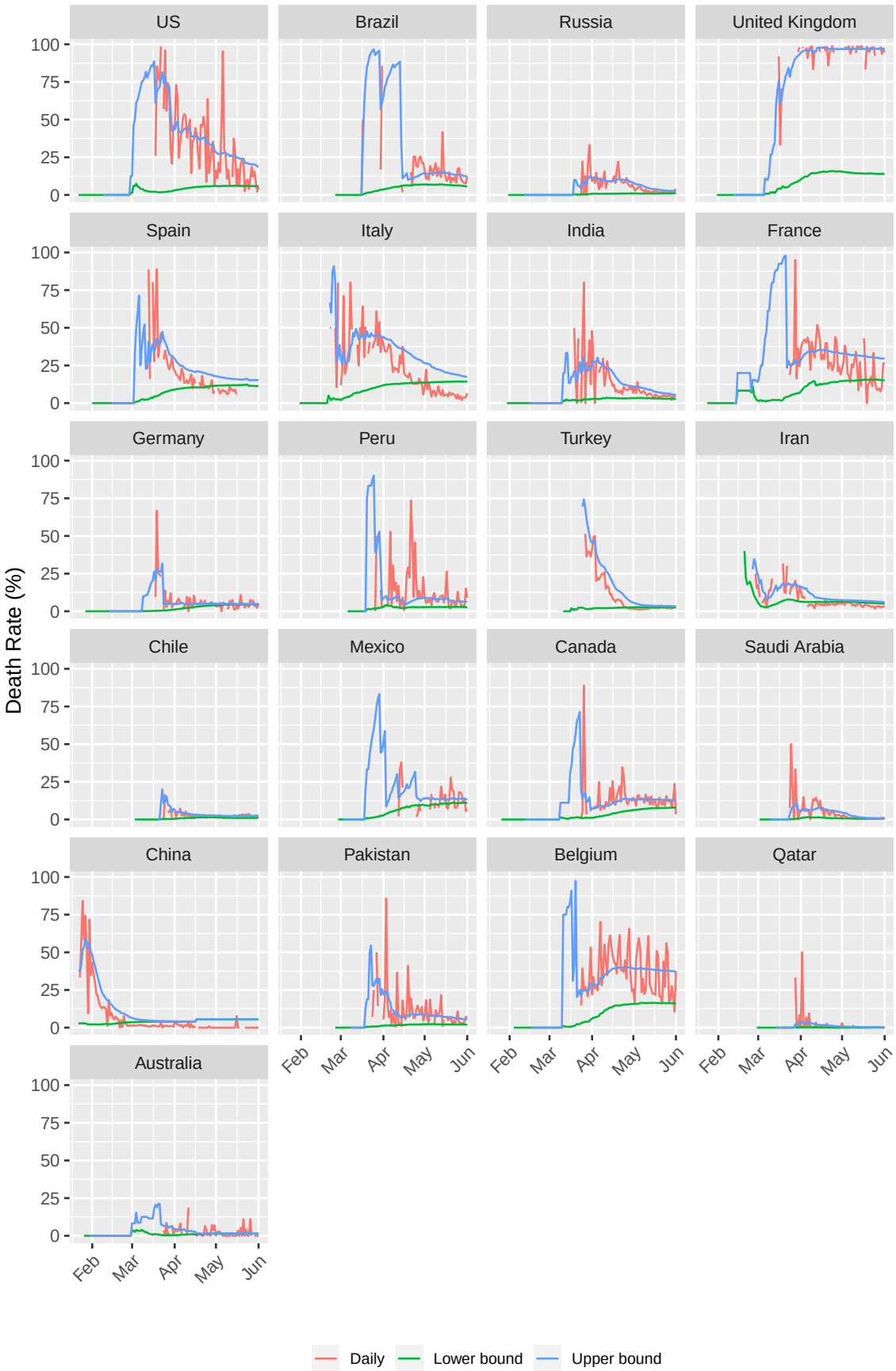


Figure 12: Death Rates

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 - 01 Jun 2020 UTC

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	Belgium	58,517	136	33,112	15,919	9,486	19	16.2%
2	France	189,348	339	91,954	68,558	28,836	31	15.2%
3	Italy	233,197	200	41,367	158,355	33,475	60	14.4%
4	United Kingdom	277,736	1,580	237,388	1,221	39,127	556	14.1%
5	Hungary	3,892	16	1,209	2,156	527	1	13.5%
6	Netherlands	46,749	104	40,589	179	5,981	6	12.8%
7	Sweden	37,814	272	28,440	4,971	4,403	8	11.6%
8	Spain	239,638	159	62,135	150,376	27,127	0	11.3%
9	Mexico	93,435	2,771	16,303	66,965	10,167	237	10.9%
10	Ecuador	39,098	0	16,148	19,592	3,358	0	8.6%
11	Canada	93,288	809	35,793	50,091	7,404	30	7.9%
12	Algeria	9,513	119	2,958	5,894	661	8	6.9%
13	Ireland	25,062	72	1,323	22,089	1,650	0	6.6%
14	Romania	19,398	141	4,696	13,426	1,276	10	6.6%
15	Switzerland	30,871	9	451	28,500	1,920	0	6.2%
16	Bosnia and Herzegovina	2,524	14	482	1,888	154	1	6.1%
17	Greece	2,918	1	1,365	1,374	179	4	6.1%
18	Indonesia	26,940	467	17,662	7,637	1,641	28	6.1%
19	North Macedonia	2,315	89	606	1,569	140	7	6.0%
20	Sudan	5,173	147	3,353	1,522	298	12	5.8%

6 Conclusions

As of 01 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 6% and 12.2%, 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)) %>%
```

```
  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
2020-06-01	6,266,192	375,559	2,696,009	3,194,624	99,246	3,524	54,680	6.0%	12.2%	6.1%
2020-05-31	6,166,946	372,035	2,641,329	3,153,582	107,929	2,909	76,636	6.0%	12.3%	3.7%
2020-05-30	6,059,017	369,126	2,564,693	3,125,198	128,236	4,128	71,158	6.1%	12.6%	5.5%
2020-05-29	5,930,781	364,998	2,493,535	3,072,248	121,835	4,690	77,575	6.2%	12.8%	5.7%
2020-05-28	5,808,946	360,308	2,415,960	3,032,678	117,156	4,680	65,872	6.2%	13.0%	6.6%
2020-05-27	5,691,790	355,628	2,350,088	2,986,074	102,164	5,176	63,132	6.2%	13.1%	7.6%
2020-05-26	5,589,626	350,452	2,286,956	2,952,218	94,565	4,221	55,218	6.3%	13.3%	7.1%
2020-05-25	5,495,061	346,231	2,231,738	2,917,092	87,448	1,173	63,175	6.3%	13.4%	1.8%
2020-05-24	5,407,613	345,058	2,168,563	2,893,992	96,593	2,845	55,701	6.4%	13.7%	4.9%
2020-05-23	5,311,020	342,213	2,112,862	2,855,945	99,864	3,980	55,878	6.4%	13.9%	6.6%
2020-05-22	5,211,156	338,233	2,056,984	2,815,939	108,732	5,309	108,245	6.5%	14.1%	4.7%
2020-05-21	5,102,424	332,924	1,948,739	2,820,761	105,952	4,809	51,273	6.5%	14.6%	8.6%
2020-05-20	4,996,472	328,115	1,897,466	2,770,891	98,980	4,830	58,471	6.6%	14.7%	7.6%
2020-05-19	4,897,492	323,285	1,838,995	2,735,212	95,549	4,804	52,120	6.6%	15.0%	8.4%
2020-05-18	4,801,943	318,481	1,786,875	2,696,587	88,323	3,296	52,912	6.6%	15.1%	5.9%
2020-05-17	4,713,620	315,185	1,733,963	2,664,472	79,552	3,404	40,766	6.7%	15.4%	7.7%
2020-05-16	4,634,068	311,781	1,693,197	2,629,090	91,721	4,115	56,130	6.7%	15.6%	6.8%
2020-05-15	4,542,347	307,666	1,637,067	2,597,614	100,184	5,248	49,174	6.8%	15.8%	9.6%
2020-05-14	4,442,163	302,418	1,587,893	2,551,852	95,145	5,221	39,346	6.8%	16.0%	11.7%
2020-05-13	4,347,018	297,197	1,548,547	2,501,274	85,271	5,255	55,133	6.8%	16.1%	8.7%
2020-05-12	4,261,747	291,942	1,493,414	2,476,391	84,245	5,612	37,205	6.9%	16.4%	13.1%
2020-05-11	4,177,502	286,330	1,456,209	2,434,963	75,803	3,621	47,229	6.9%	16.4%	7.1%
2020-05-10	4,101,699	282,709	1,408,980	2,410,010	77,690	3,398	33,356	6.9%	16.7%	9.2%
2020-05-09	4,024,009	279,311	1,375,624	2,369,074	85,945	4,413	53,574	6.9%	16.9%	7.6%
2020-05-08	3,938,064	274,898	1,322,050	2,341,116	92,346	5,331	37,309	7.0%	17.2%	12.5%
2020-05-07	3,845,718	269,567	1,284,741	2,291,410	89,649	5,712	39,328	7.0%	17.3%	12.7%
2020-05-06	3,756,069	263,855	1,245,413	2,246,801	93,378	6,616	46,581	7.0%	17.5%	12.4%
2020-05-05	3,662,691	257,239	1,198,832	2,206,620	79,636	5,702	36,108	7.0%	17.7%	13.6%
2020-05-04	3,583,055	251,537	1,162,724	2,168,794	76,326	4,067	37,488	7.0%	17.8%	9.8%
2020-05-03	3,506,729	247,470	1,125,236	2,134,023	79,145	3,657	32,099	7.1%	18.0%	10.2%
2020-05-02	3,427,584	243,813	1,093,137	2,090,634	82,026	5,194	40,722	7.1%	18.2%	11.3%
2020-05-01	3,345,558	238,619	1,052,415	2,054,524	88,648	5,259	38,529	7.1%	18.5%	12.0%
2020-04-30	3,256,910	233,360	1,013,886	2,009,664	84,623	5,695	65,461	7.2%	18.7%	8.0%
2020-04-29	3,172,287	227,665	948,425	1,996,197	75,058	6,864	41,470	7.2%	19.4%	14.2%
2020-04-28	3,097,229	220,801	906,955	1,969,473	73,507	6,357	33,278	7.1%	19.6%	16.0%
2020-04-27	3,023,722	214,444	873,677	1,935,601	68,689	4,544	27,692	7.1%	19.7%	14.1%
2020-04-26	2,955,033	209,900	845,985	1,899,148	73,893	3,713	28,580	7.1%	19.9%	11.5%
2020-04-25	2,881,140	206,187	817,405	1,857,548	85,409	6,190	27,820	7.2%	20.1%	18.2%
2020-04-24	2,795,731	199,997	789,585	1,806,149	87,328	6,332	50,605	7.2%	20.2%	11.1%
2020-04-23	2,708,403	193,665	738,980	1,775,758	83,795	6,753	28,934	7.2%	20.8%	18.9%
2020-04-22	2,624,608	186,912	710,046	1,727,650	75,562	6,676	30,141	7.1%	20.8%	18.1%
2020-04-21	2,549,046	180,236	679,905	1,688,905	76,937	7,112	34,847	7.1%	21.0%	16.9%
2020-04-20	2,472,109	173,124	645,058	1,653,927	71,322	5,352	22,001	7.0%	21.2%	19.6%
2020-04-19	2,400,787	167,772	623,057	1,609,958	83,151	4,536	31,588	7.0%	21.2%	12.6%
2020-04-18	2,317,636	163,236	591,469	1,562,931	77,646	6,415	23,954	7.0%	21.6%	21.1%
2020-04-17	2,239,990	156,821	567,515	1,515,654	87,809	8,858	26,173	7.0%	21.7%	25.3%
2020-04-16	2,152,181	147,963	541,342	1,462,876	96,433	7,278	31,076	6.9%	21.5%	19.0%
2020-04-15	2,055,748	140,685	510,266	1,404,797	80,182	8,246	36,536	6.8%	21.6%	18.4%
2020-04-14	1,975,566	132,439	473,730	1,369,397	70,401	6,878	25,308	6.7%	21.8%	21.4%
2020-04-13	1,905,165	125,561	448,422	1,331,182	70,020	5,708	26,950	6.6%	21.9%	17.5%
2020-04-12	1,835,145	119,853	421,472	1,293,820	98,733	5,683	19,602	6.5%	22.1%	22.5%
2020-04-11	1,736,412	114,170	401,870	1,220,372	78,151	6,033	26,016	6.6%	22.1%	18.8%
2020-04-10	1,658,261	108,137	375,854	1,174,270	92,159	7,222	22,115	6.5%	22.3%	24.6%
2020-04-09	1,566,102	100,915	353,739	1,111,448	85,870	7,561	25,286	6.4%	22.2%	23.0%
2020-04-08	1,480,232	93,354	328,453	1,058,425	83,794	6,692	28,649	6.3%	22.1%	18.9%
2020-04-07	1,396,438	86,662	299,804	1,009,972	75,002	7,895	23,520	6.2%	22.4%	25.1%
2020-04-06	1,321,436	78,767	276,284	966,385	71,693	5,690	16,511	6.0%	22.2%	25.6%
2020-04-05	1,249,743	73,077	259,773	916,893	73,307	5,081	13,871	5.8%	22.0%	26.8%
2020-04-04	1,176,436	67,996	245,902	862,538	80,112	6,170	20,338	5.8%	21.7%	23.3%
2020-04-03	1,096,324	61,826	225,564	808,934	82,461	5,992	15,545	5.6%	21.5%	27.8%
2020-04-02	1,013,863	55,834	210,019	748,010	80,853	6,149	17,071	5.5%	21.0%	26.5%
2020-04-01	933,010	49,685	192,948	690,377	75,402	5,447	15,164	5.3%	20.5%	26.4%
2020-03-31	857,608	44,238	177,784	635,586	74,792	4,795	13,468	5.2%	19.9%	26.3%
2020-03-30	782,816	39,443	164,316	579,057	62,525	4,120	15,484	5.0%	19.4%	21.0%
2020-03-29	720,291	35,323	148,832	536,136	59,117	3,512	9,420	4.9%	19.2%	27.2%
2020-03-28	661,174	31,811	139,412	489,951	67,410	3,519	8,500	4.8%	18.6%	29.3%

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

date	confirmed	deaths	recovered	current.confirmed	new.confirmed	new.deaths	new.recovered	rate.lower	rate.upper	rate.daily
2020-03-27	593,764	28,292	130,912	434,560	64,042	3,504	8,765	4.8%	17.8%	28.6%
2020-03-26	529,722	24,788	122,147	382,787	61,905	2,999	8,363	4.7%	16.9%	26.4%
2020-03-25	467,817	21,789	113,784	332,244	49,522	2,772	5,786	4.7%	16.1%	32.4%
2020-03-24	418,295	19,017	107,998	291,280	39,994	2,258	9,649	4.5%	15.0%	19.0%
2020-03-23	378,301	16,759	98,349	263,193	41,279	1,933	452	4.4%	14.6%	81.0%
2020-03-22	337,022	14,826	97,897	224,299	32,442	1,701	6,207	4.4%	13.2%	21.5%
2020-03-21	304,580	13,125	91,690	199,765	32,316	1,702	4,272	4.3%	12.5%	28.5%
2020-03-20	272,264	11,423	87,418	173,423	29,632	1,476	2,445	4.2%	11.6%	37.6%
2020-03-19	242,632	9,947	84,973	147,712	27,759	1,123	1,663	4.1%	10.5%	40.3%
2020-03-18	214,873	8,824	83,310	122,739	17,723	867	2,472	4.1%	9.6%	26.0%
2020-03-17	197,150	7,957	80,838	108,355	15,497	806	2,752	4.0%	9.0%	22.7%
2020-03-16	181,653	7,151	78,086	96,416	14,146	678	2,054	3.9%	8.4%	24.8%
2020-03-15	167,507	6,473	76,032	85,002	11,387	641	3,410	3.9%	7.8%	15.8%
2020-03-14	156,120	5,832	72,622	77,666	10,895	423	2,371	3.7%	7.4%	15.1%
2020-03-13	145,225	5,409	70,251	69,565	14,218	497	1,927	3.7%	7.1%	20.5%
2020-03-12	131,007	4,912	68,324	57,771	5,128	298	1,321	3.7%	6.7%	18.4%
2020-03-11	125,879	4,614	67,003	54,262	7,256	351	2,599	3.7%	6.4%	11.9%
2020-03-10	118,623	4,263	64,404	49,956	5,028	276	1,910	3.6%	6.2%	12.6%
2020-03-09	113,595	3,987	62,494	47,114	3,773	186	1,800	3.5%	6.0%	9.4%
2020-03-08	109,822	3,801	60,694	45,327	3,974	243	2,336	3.5%	5.9%	9.4%
2020-03-07	105,848	3,558	58,358	43,932	4,042	99	2,493	3.4%	5.7%	3.8%
2020-03-06	101,806	3,459	55,865	42,482	3,918	112	2,069	3.4%	5.8%	5.1%
2020-03-05	97,888	3,347	53,796	40,745	2,768	93	2,626	3.4%	5.9%	3.4%
2020-03-04	95,120	3,254	51,170	40,696	2,280	94	2,942	3.4%	6.0%	3.1%
2020-03-03	92,840	3,160	48,228	41,452	2,534	75	2,626	3.4%	6.1%	2.8%
2020-03-02	90,306	3,085	45,602	41,619	1,937	89	2,886	3.4%	6.3%	3.0%
2020-03-01	88,369	2,996	42,716	42,657	2,358	55	2,934	3.4%	6.6%	1.8%
2020-02-29	86,011	2,941	39,782	43,288	1,899	69	3,071	3.4%	6.9%	2.2%
2020-02-28	84,112	2,872	36,711	44,529	1,366	58	3,434	3.4%	7.3%	1.7%
2020-02-27	82,746	2,814	33,277	46,655	1,358	44	2,893	3.4%	7.8%	1.5%
2020-02-26	81,388	2,770	30,384	48,234	982	62	2,479	3.4%	8.4%	2.4%
2020-02-25	80,406	2,708	27,905	49,793	845	79	2,678	3.4%	8.8%	2.9%
2020-02-24	79,561	2,629	25,227	51,705	603	160	1,833	3.3%	9.4%	8.0%
2020-02-23	78,958	2,469	23,394	53,095	386	11	508	3.1%	9.5%	2.1%
2020-02-22	78,572	2,458	22,886	53,228	1,753	207	3,996	3.1%	9.7%	4.9%
2020-02-21	76,819	2,251	18,890	55,678	622	4	713	2.9%	10.6%	0.6%
2020-02-20	76,197	2,247	18,177	55,773	558	125	2,056	2.9%	11.0%	5.7%
2020-02-19	75,639	2,122	16,121	57,396	503	115	1,769	2.8%	11.6%	6.1%
2020-02-18	75,136	2,007	14,352	58,777	1,878	139	1,769	2.7%	12.3%	7.3%
2020-02-17	73,258	1,868	12,583	58,807	2,034	98	1,718	2.5%	12.9%	5.4%
2020-02-16	71,224	1,770	10,865	58,589	2,194	104	1,470	2.5%	14.0%	6.6%
2020-02-15	69,030	1,666	9,395	57,969	2,145	143	1,337	2.4%	15.1%	9.7%
2020-02-14	66,885	1,523	8,058	57,304	6,517	152	1,763	2.3%	15.9%	7.9%
2020-02-13	60,368	1,371	6,295	52,702	15,147	253	1,145	2.3%	17.9%	18.1%
2020-02-12	45,221	1,118	5,150	38,953	419	5	467	2.5%	17.8%	1.1%
2020-02-11	44,802	1,113	4,683	39,006	2,040	100	737	2.5%	19.2%	11.9%
2020-02-10	42,762	1,013	3,946	37,803	2,612	107	702	2.4%	20.4%	13.2%
2020-02-09	40,150	906	3,244	36,000	3,030	100	628	2.3%	21.8%	13.7%
2020-02-08	37,120	806	2,616	33,698	2,729	87	605	2.2%	23.6%	12.6%
2020-02-07	34,391	719	2,011	31,661	3,597	85	524	2.1%	26.3%	14.0%
2020-02-06	30,794	634	1,487	28,673	3,159	70	363	2.1%	29.9%	16.2%
2020-02-05	27,635	564	1,124	25,947	3,743	72	272	2.0%	33.4%	20.9%
2020-02-04	23,892	492	852	22,548	4,011	66	229	2.1%	36.6%	22.4%
2020-02-03	19,881	426	623	18,832	3,094	64	151	2.1%	40.6%	29.8%
2020-02-02	16,787	362	472	15,953	4,749	103	188	2.2%	43.4%	35.4%
2020-02-01	12,038	259	284	11,495	2,111	46	62	2.2%	47.7%	42.6%
2020-01-31	9,927	213	222	9,492	1,693	42	79	2.1%	49.0%	34.7%
2020-01-30	8,234	171	143	7,920	2,068	38	17	2.1%	54.5%	69.1%
2020-01-29	6,166	133	126	5,907	588	2	19	2.2%	51.4%	9.5%
2020-01-28	5,578	131	107	5,340	2,651	49	46	2.3%	55.0%	51.6%
2020-01-27	2,927	82	61	2,784	809	26	9	2.8%	57.3%	74.3%
2020-01-26	2,118	56	52	2,010	684	14	13	2.6%	51.9%	51.9%
2020-01-25	1,434	42	39	1,353	493	16	3	2.9%	51.9%	84.2%
2020-01-24	941	26	36	879	287	8	6	2.8%	41.9%	57.1%
2020-01-23	654	18	30	606	99	1	2	2.8%	37.5%	33.3%
2020-01-22	555	17	28	510				3.1%	37.8%	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 (01 Jun 2020 UTC)

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	World	6,266,192	99,246	3,194,624	2,696,009	375,559	3,524	**6.0%**
2	US	1,811,360	21,188	1,247,964	458,231	105,165	784	**5.8%**
3	Brazil	526,447	11,598	285,430	211,080	29,937	623	**5.7%**
4	Russia	414,328	8,485	233,965	175,514	4,849	156	1.2%
5	United Kingdom	277,736	1,580	237,388	1,221	39,127	556	**14.1%**
6	Spain	239,638	159	62,135	150,376	27,127	0	**11.3%**
7	Italy	233,197	200	41,367	158,355	33,475	60	**14.4%**
8	India	198,370	7,761	97,008	95,754	5,608	200	2.8%
9	France	189,348	339	91,954	68,558	28,836	31	**15.2%**
10	Germany	183,594	184	9,407	165,632	8,555	15	4.7%
11	Peru	170,039	5,563	96,898	68,507	4,634	128	2.7%
12	Turkey	164,769	827	31,259	128,947	4,563	23	2.8%
13	Iran	154,445	2,979	25,563	121,004	7,878	81	**5.1%**
14	Chile	105,158	5,470	59,099	44,946	1,113	59	1.1%
15	Mexico	93,435	2,771	16,303	66,965	10,167	237	**10.9%**
16	Canada	93,288	809	35,793	50,091	7,404	30	**7.9%**
17	Saudi Arabia	87,142	1,881	22,311	64,306	525	22	0.6%
18	China	84,154	8	118	79,398	4,638	0	**5.5%**
19	Pakistan	72,460	2,964	44,834	26,083	1,543	60	2.1%
20	Belgium	58,517	136	33,112	15,919	9,486	19	**16.2%**
21	Qatar	58,433	1,523	24,956	33,437	40	2	0.1%
22	Bangladesh	49,534	2,381	38,265	10,597	672	22	1.4%
23	Netherlands	46,749	104	40,589	179	5,981	6	**12.8%**
24	Belarus	43,403	847	24,387	18,776	240	5	0.6%
25	Ecuador	39,098	0	16,148	19,592	3,358	0	**8.6%**
26	Sweden	37,814	272	28,440	4,971	4,403	8	**11.6%**
27	Singapore	35,292	408	12,802	22,466	24	1	0.1%
28	United Arab Emirates	35,192	635	16,588	18,338	266	2	0.8%
29	South Africa	34,357	1,674	16,361	17,291	705	22	2.1%
30	Portugal	32,700	200	11,724	19,552	1,424	14	4.4%
31	Switzerland	30,871	9	451	28,500	1,920	0	**6.2%**
32	Colombia	29,384	2,165	20,037	8,384	963	47	3.3%
33	Kuwait	27,762	719	14,643	12,899	220	8	0.8%
34	Indonesia	26,940	467	17,662	7,637	1,641	28	**6.1%**
35	Egypt	26,384	1,399	18,932	6,447	1,005	46	3.8%
36	Ireland	25,062	72	1,323	22,089	1,650	0	**6.6%**
37	Ukraine	24,562	890	13,785	10,053	724	16	2.9%
38	Poland	24,165	379	11,642	11,449	1,074	10	4.4%
39	Romania	19,398	141	4,696	13,426	1,276	10	**6.6%**
40	Philippines	18,638	552	13,699	3,979	960	3	**5.2%**
41	Dominican Republic	17,572	287	6,177	10,893	502	0	2.9%
42	Argentina	17,415	564	11,338	5,521	556	17	3.2%
43	Israel	17,169	98	2,006	14,878	285	0	1.7%
44	Japan	16,787	36	1,425	14,463	899	1	**5.4%**
45	Austria	16,733	2	469	15,596	668	0	4.0%
46	Afghanistan	15,750	545	14,057	1,428	265	8	1.7%
47	Panama	13,837	374	3,979	9,514	344	8	2.5%
48	Oman	12,223	786	9,491	2,682	50	1	0.4%
49	Denmark	11,899	30	713	10,610	576	2	4.8%
50	Bahrain	11,871	473	4,782	7,070	19	0	0.2%
51	Korea, South	11,541	38	823	10,446	272	1	2.4%
52	Serbia	11,430	18	4,460	6,726	244	1	2.1%
53	Kazakhstan	11,308	450	5,680	5,587	41	1	0.4%
54	Nigeria	10,578	416	7,157	3,122	299	12	2.8%
55	Bolivia	10,531	549	9,051	1,137	343	30	3.3%
56	Algeria	9,513	119	2,958	5,894	661	8	**6.9%**
57	Armenia	9,492	210	5,951	3,402	139	8	1.5%
58	Czechia	9,302	34	2,339	6,642	321	1	3.5%
59	Norway	8,446	6	483	7,727	236	0	2.8%
60	Moldova	8,360	109	3,433	4,622	305	10	3.6%
61	Ghana	8,070	0	5,087	2,947	36	0	0.4%
62	Malaysia	7,857	38	1,338	6,404	115	0	1.5%
63	Morocco	7,833	26	1,735	5,893	205	0	2.6%

Table 6: Cases by Country (01 Jun 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
64	Australia	7,221	19	493	6,626	102	0	1.4%
65	Finland	6,885	26	1,067	5,500	318	0	4.6%
66	Iraq	6,868	429	3,378	3,275	215	10	3.1%
67	Cameroon	6,397	493	2,569	3,629	199	8	3.1%
68	Azerbaijan	5,662	168	2,086	3,508	68	5	1.2%
69	Honduras	5,362	160	4,596	549	217	5	4.0%
70	Guatemala	5,336	249	4,425	795	116	8	2.2%
71	Sudan	5,173	147	3,353	1,522	298	12	**5.8%**
72	Luxembourg	4,019	1	64	3,845	110	0	2.7%
73	Tajikistan	4,013	83	1,877	2,089	47	0	1.2%
74	Hungary	3,892	16	1,209	2,156	527	1	**13.5%**
75	Guinea	3,844	138	1,686	2,135	23	0	0.6%
76	Senegal	3,739	94	1,839	1,858	42	0	1.1%
77	Uzbekistan	3,702	79	828	2,859	15	0	0.4%
78	Djibouti	3,569	215	2,024	1,521	24	0	0.7%
79	Congo (Kinshasa)	3,195	125	2,669	454	72	0	2.3%
80	Thailand	3,082	1	60	2,965	57	0	1.8%
81	Cote d'Ivoire	2,951	118	1,451	1,467	33	0	1.1%
82	Greece	2,918	1	1,365	1,374	179	4	**6.1%**
83	Gabon	2,655	0	1,916	722	17	0	0.6%
84	El Salvador	2,582	65	1,473	1,063	46	0	1.8%
85	Bosnia and Herzegovina	2,524	14	482	1,888	154	1	**6.1%**
86	Bulgaria	2,519	6	1,289	1,090	140	0	**5.6%**
87	North Macedonia	2,315	89	606	1,569	140	7	**6.0%**
88	Croatia	2,246	0	66	2,077	103	0	4.6%
89	Haiti	2,226	102	2,157	24	45	1	2.0%
90	Cuba	2,083	38	174	1,826	83	0	4.0%
91	Somalia	2,023	47	1,583	361	79	1	3.9%
92	Kenya	2,021	59	1,470	482	69	5	3.4%
93	Estonia	1,870	1	177	1,625	68	0	3.6%
94	Maldives	1,829	56	1,335	488	6	1	0.3%
95	Kyrgyzstan	1,817	69	620	1,181	16	0	0.9%
96	Nepal	1,811	239	1,582	221	8	0	0.4%
97	Iceland	1,806	0	2	1,794	10	0	0.6%
98	Lithuania	1,678	3	372	1,236	70	0	4.2%
99	Venezuela	1,662	152	1,343	302	17	3	1.0%
100	Sri Lanka	1,643	10	821	811	11	1	0.7%
101	Slovakia	1,522	1	126	1,368	28	0	1.8%
102	New Zealand	1,504	0	1	1,481	22	0	1.5%
103	Slovenia	1,473	0	6	1,358	109	1	7.4%
104	Guinea-Bissau	1,339	83	1,278	53	8	0	0.6%
105	Mali	1,315	50	493	744	78	1	5.9%
106	Equatorial Guinea	1,306	0	1,094	200	12	0	0.9%
107	Ethiopia	1,257	85	1,028	217	12	1	1.0%
108	Lebanon	1,233	13	491	715	27	0	2.2%
109	Albania	1,143	6	233	877	33	0	2.9%
110	Zambia	1,089	32	170	912	7	0	0.6%
111	Costa Rica	1,084	28	398	676	10	0	0.9%
112	Tunisia	1,084	7	72	964	48	0	4.4%
113	Central African Republic	1,069	58	1,042	23	4	2	0.4%
114	Latvia	1,066	0	297	745	24	0	2.3%
115	Kosovo	1,064	0	205	829	30	0	2.8%
116	Paraguay	995	9	496	488	11	0	1.1%
117	South Sudan	994	0	978	6	10	0	1.0%
118	Niger	958	0	49	844	65	1	6.8%
119	Cyprus	949	5	142	790	17	0	1.8%
120	Sierra Leone	865	4	344	475	46	0	5.3%
121	Burkina Faso	847	0	74	720	53	0	6.3%
122	Madagascar	826	55	646	174	6	0	0.7%
123	Uruguay	825	2	113	689	23	1	2.8%
124	Georgia	794	11	158	624	12	0	1.5%
125	Chad	790	12	185	539	66	1	8.4%
126	Andorra	765	1	16	698	51	0	6.7%
127	Nicaragua	759	0	354	370	35	0	4.6%
128	Jordan	746	7	202	535	9	0	1.2%
129	Diamond Princess	712	0	48	651	13	0	1.8%
130	San Marino	671	0	270	359	42	0	6.3%
131	Malta	619	1	73	537	9	0	1.5%

Table 6: Cases by Country (01 Jun 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
132	Congo (Brazzaville)	611	0	412	179	20	0	3.3%
133	Jamaica	588	2	257	322	9	0	1.5%
134	Mauritania	588	58	538	27	23	0	3.9%
135	Tanzania	509	0	305	183	21	0	4.1%
136	Sao Tome and Principe	484	1	404	68	12	0	2.5%
137	Cabo Verde	458	23	261	193	4	0	0.9%
138	Uganda	457	40	385	72	0	0	0.0%
139	West Bank and Gaza	449	1	74	372	3	0	0.7%
140	Taiwan*	443	1	9	427	7	0	1.6%
141	Togo	443	1	215	215	13	0	2.9%
142	Rwanda	377	7	114	262	1	0	0.3%
143	Yemen	354	31	256	14	84	4	23.7%
144	Malawi	336	52	290	42	4	0	1.2%
145	Mauritius	335	0	3	322	10	0	3.0%
146	Vietnam	328	0	35	293	0	0	0.0%
147	Montenegro	324	0	0	315	9	0	2.8%
148	Liberia	296	8	110	159	27	0	9.1%
149	Eswatini	293	8	96	194	3	1	1.0%
150	Mozambique	254	0	155	97	2	0	0.8%
151	Benin	243	11	93	147	3	0	1.2%
152	Burma	228	4	84	138	6	0	2.6%
153	Zimbabwe	203	25	170	29	4	0	2.0%
154	Mongolia	185	6	141	44	0	0	0.0%
155	Libya	168	12	111	52	5	0	3.0%
156	Guyana	153	0	71	70	12	0	7.8%
157	Brunei	141	0	1	138	2	0	1.4%
158	Cambodia	125	0	2	123	0	0	0.0%
159	Syria	123	1	72	46	5	0	4.1%
160	Trinidad and Tobago	117	0	1	108	8	0	6.8%
161	Comoros	106	0	78	26	2	0	1.9%
162	Bahamas	102	0	42	49	11	0	10.8%
163	Monaco	99	0	5	90	4	0	4.0%
164	Barbados	92	0	9	76	7	0	7.6%
165	Angola	86	0	64	18	4	0	4.7%
166	Liechtenstein	82	0	26	55	1	0	1.2%
167	Burundi	63	0	29	33	1	0	1.6%
168	Suriname	44	21	34	9	1	0	2.3%
169	Bhutan	43	0	37	6	0	0	0.0%
170	Eritrea	39	0	0	39	0	0	0.0%
171	Botswana	38	3	17	20	1	0	2.6%
172	Antigua and Barbuda	26	0	4	19	3	0	11.5%
173	Saint Vincent and the Grenadines	26	0	11	15	0	0	0.0%
174	Gambia	25	0	4	20	1	0	4.0%
175	Namibia	25	1	9	16	0	0	0.0%
176	Timor-Leste	24	0	0	24	0	0	0.0%
177	Grenada	23	0	5	18	0	0	0.0%
178	Laos	19	0	3	16	0	0	0.0%
179	Belize	18	0	0	16	2	0	11.1%
180	Fiji	18	0	3	15	0	0	0.0%
181	Saint Lucia	18	0	0	18	0	0	0.0%
182	Dominica	16	0	0	16	0	0	0.0%
183	Saint Kitts and Nevis	15	0	0	15	0	0	0.0%
184	Holy See	12	0	10	2	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%
187	Western Sahara	9	0	2	6	1	0	11.1%
188	Papua New Guinea	8	0	0	8	0	0	0.0%
189	Lesotho	2	0	1	1	0	0	0.0%

Appendix B. How to Cite This Work

Citation

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

BibTex

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

Appendix C. Contact

Contact:

Dr. Yanchang Zhao

Email: yanchang@RDataMining.com

Twitter: @RDataMining

LinkedIn: http://group.rdatamining.com

Comments and suggestions and welcome. Thanks!