COVID-19 Data Analysis with R - Worldwide*

Yanchang Zhao

yanchang@RDataMining.com

http://RDataMining.com

24 June 2020

Contents

---

*©2020 Yanchang Zhao, RDataMining.com.

1 Introduction

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

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

1.1 Data Source

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

1.2 R Packages

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.

```r
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.

```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] 266 157

Each dataset has 266 rows, corresponding to country/region/province/state. It has 157 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" "2020-06-22"

```
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 22 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())
```

```r
  ## 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 22 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.

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

data %<>% rbind(data.world)
```

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

3.3 Daily Increases and Death Rates

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

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

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

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

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

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

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

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

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


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

```
                                                deaths='Deaths'))


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

4   Worldwide Cases

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

4.1   World Map

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

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

m <- leaflet(width=1200, height=800) %>% addTiles()
# circle marker (units in pixels)
m %<>% addCircleMarkers(x$Long, x$Lat,
                        # radius=2+log2(x$confirmed),
                        radius=0.03*sqrt(x$confirmed),
                        stroke=F,
                        color='red', fillOpacity=0.3,
                        popup=x$txt)

# world
m
```
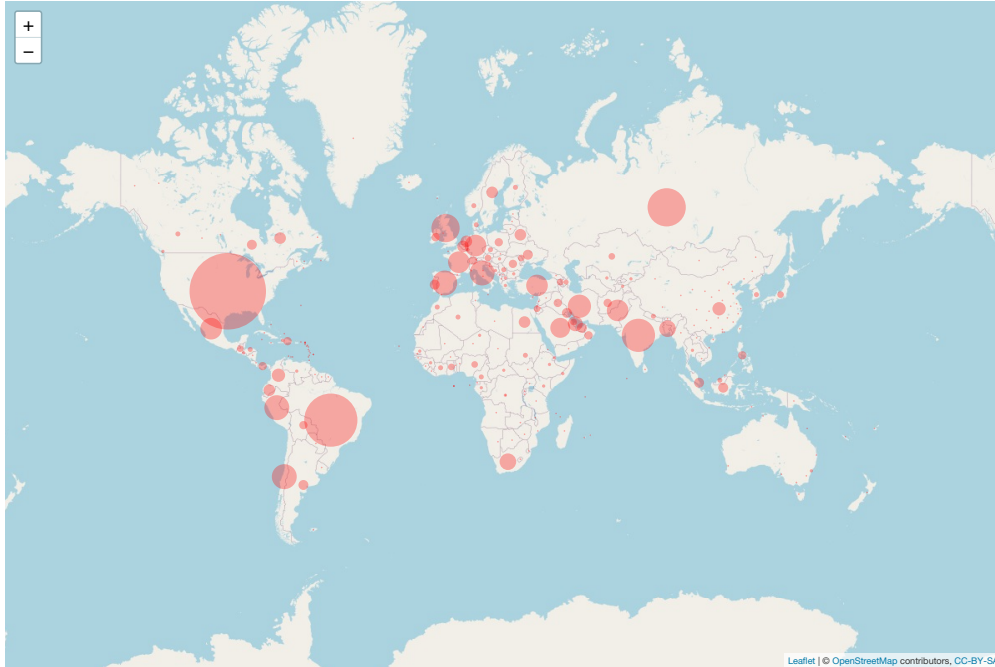
Figure 1: World Map

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 focus on the cases worldwide. Similar analysis for a single country can be done by filter the data with the corresponding country name.

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

## cases - area plot
plot1 <- world.long %>% filter(type != 'Total Confirmed') %>%
  ggplot(aes(x=date, y=count)) +
  geom_area(aes(fill=type), alpha=0.5) +
  labs(title=paste0('Numbers of Cases Worldwide - ', max.date.txt)) +
  scale_fill_manual(values=c('red', 'green', 'black')) +
  theme(legend.title=element_blank(), legend.position='bottom',
        plot.title = element_text(size=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)
```
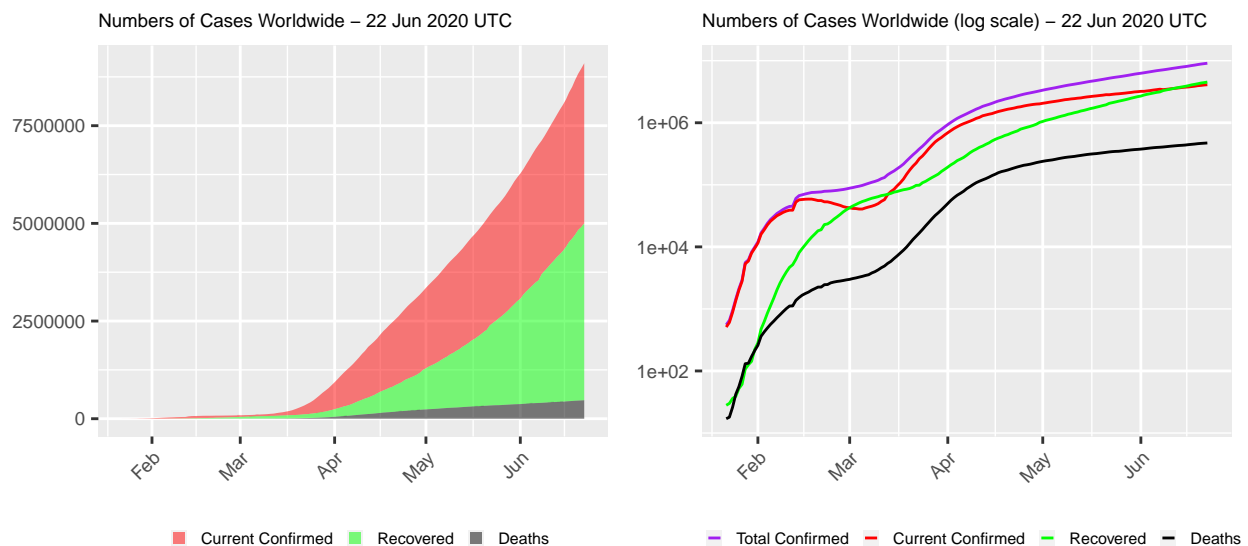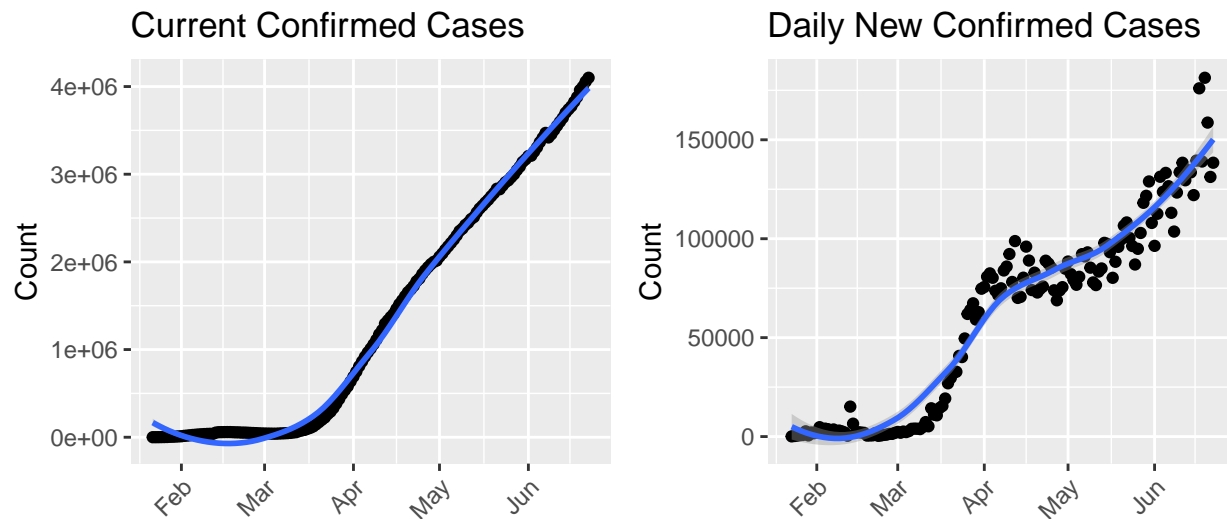

Figure 3: Current Confirmed Cases

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)
```
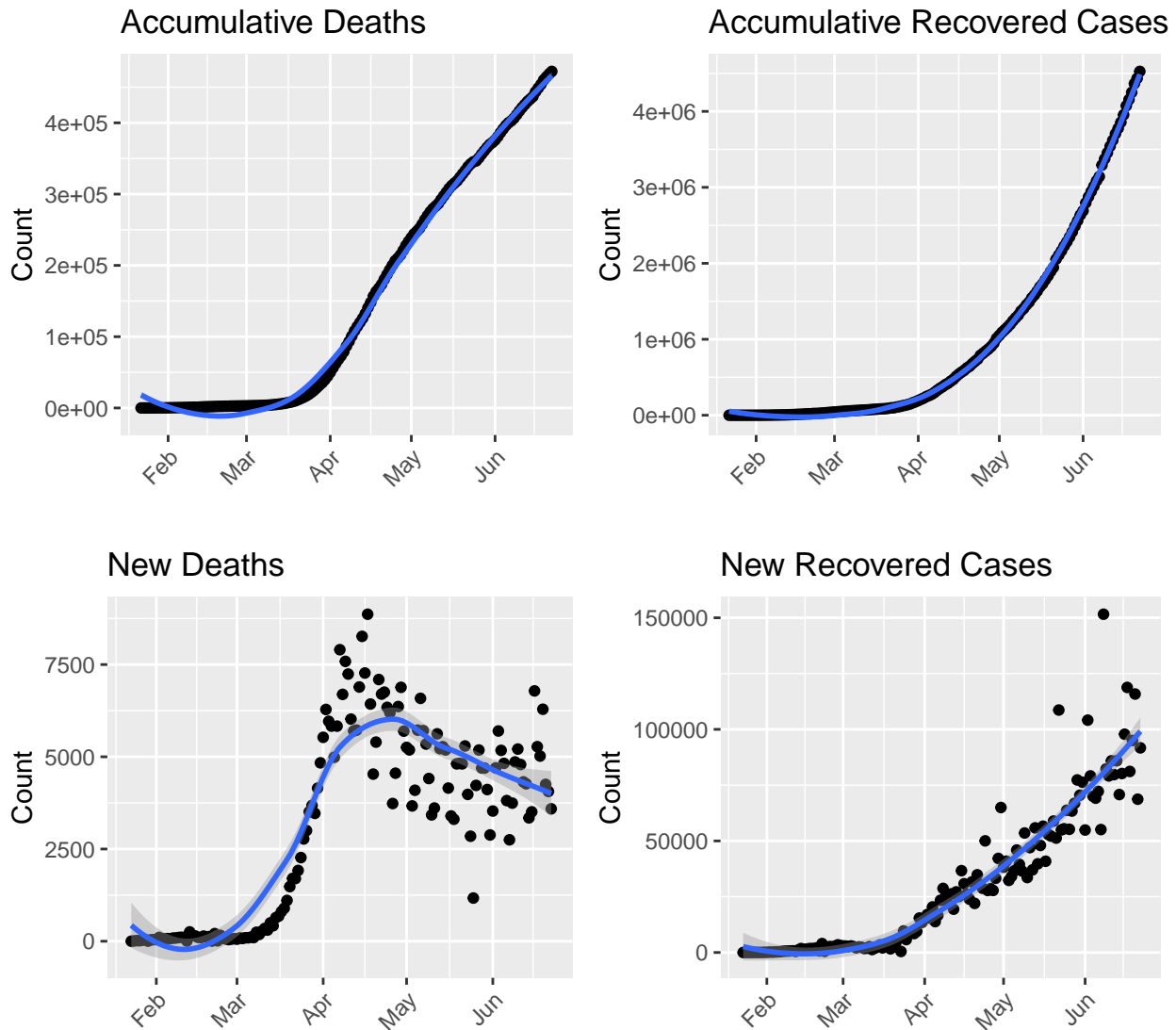
Figure 4: Deaths and Recovered Cases

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 22 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 22 Jun 2020 UTC UTC, it will be between 5.2% and 9.4%.

A surge in the daily death rate (in red) in late March suggests that the situation is changing dramatically (actually, getting worse) and that above lower/upper bounds are likely to increase shortly. A likely reason of that surge is the 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)))
```
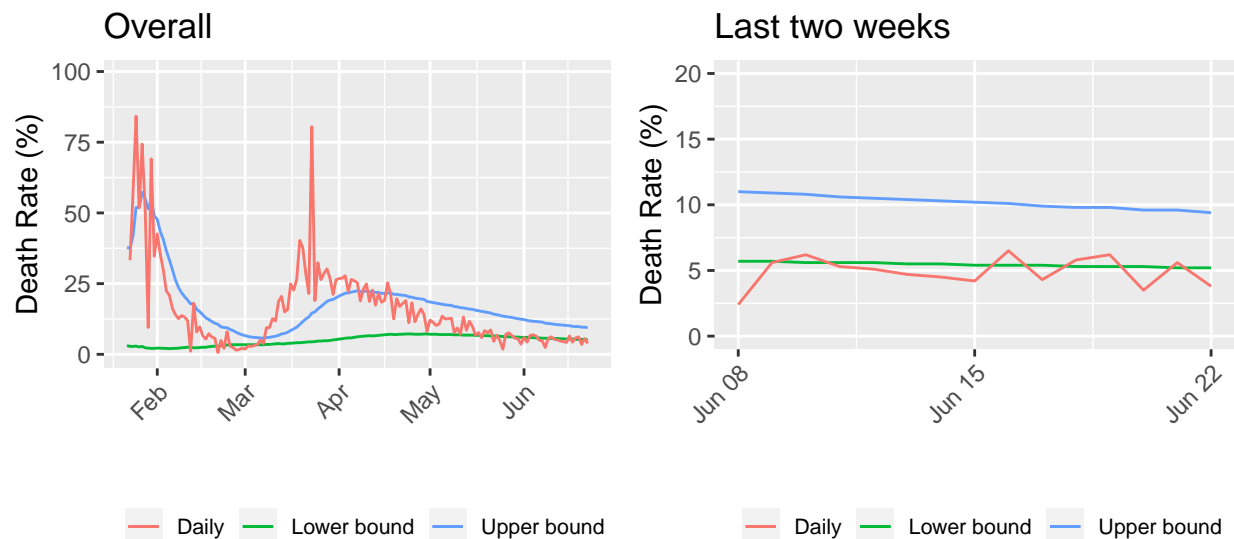
```r
k <- 20
## top 20 countries: 21 incl. 'World'
top.countries <- data.latest.all %>% filter(ranking <= k + 1) %>%
  arrange(ranking) %>% pull(country) %>% as.character()
top.countries %>% setdiff('World') %>% print()
```

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

```r
## add 'Others'
# top.countries %<>% c('Others')
## put all others in a single group of 'Others'
data.latest <- data.latest.all %>% filter(!is.na(country)) %>%
  mutate(country=ifelse(ranking <= k + 1, as.character(country), 'Others')) %>%
  mutate(country=country %>% factor(levels=c(top.countries, '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 - 22 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	9,098,643	472,171	5.2%	138,401	3,590	4,100,139
2	US	2,312,302	120,402	5.2%	31,012	425	1,551,702
3	Brazil	1,106,470	51,271	4.6%	23,129	680	453,463
4	Russia	591,465	8,196	1.4%	7,586	95	239,422
5	India	440,215	14,011	3.2%	14,933	312	178,014
6	United Kingdom	306,761	42,731	13.9%	958	14	262,708
7	Peru	257,447	8,223	3.2%	2,511	178	103,904
8	Chile	246,963	4,502	1.8%	4,608	23	37,064
9	Spain	246,504	28,324	11.5%	232	1	67,804
10	Italy	238,720	34,657	14.5%	221	23	20,637
11	Iran	207,525	9,742	4.7%	2,573	119	31,356
12	France	197,381	29,666	15.0%	373	23	92,979
13	Germany	191,768	8,899	4.6%	496	4	7,726
14	Turkey	188,897	4,974	2.6%	1,212	24	22,390
15	Mexico	185,122	22,584	12.2%	4,577	759	23,155
16	Pakistan	185,034	3,695	2.0%	3,946	105	107,868
17	Saudi Arabia	161,005	1,307	0.8%	3,393	40	54,523
18	Bangladesh	115,786	1,502	1.3%	3,480	38	67,529
19	Canada	103,418	8,494	8.2%	340	12	29,203
20	South Africa	101,590	1,991	2.0%	4,288	61	46,155
21	Qatar	88,403	99	0.1%	1,034	1	18,348
22	Others	1,625,867	66,901	4.1%	27,499	653	684,189

```
      axis.text=element_text(size=7),
      axis.text.x=element_text(angle=45, hjust=1)) +
facet_wrap(~type, ncol=1, scales='free_y')
```
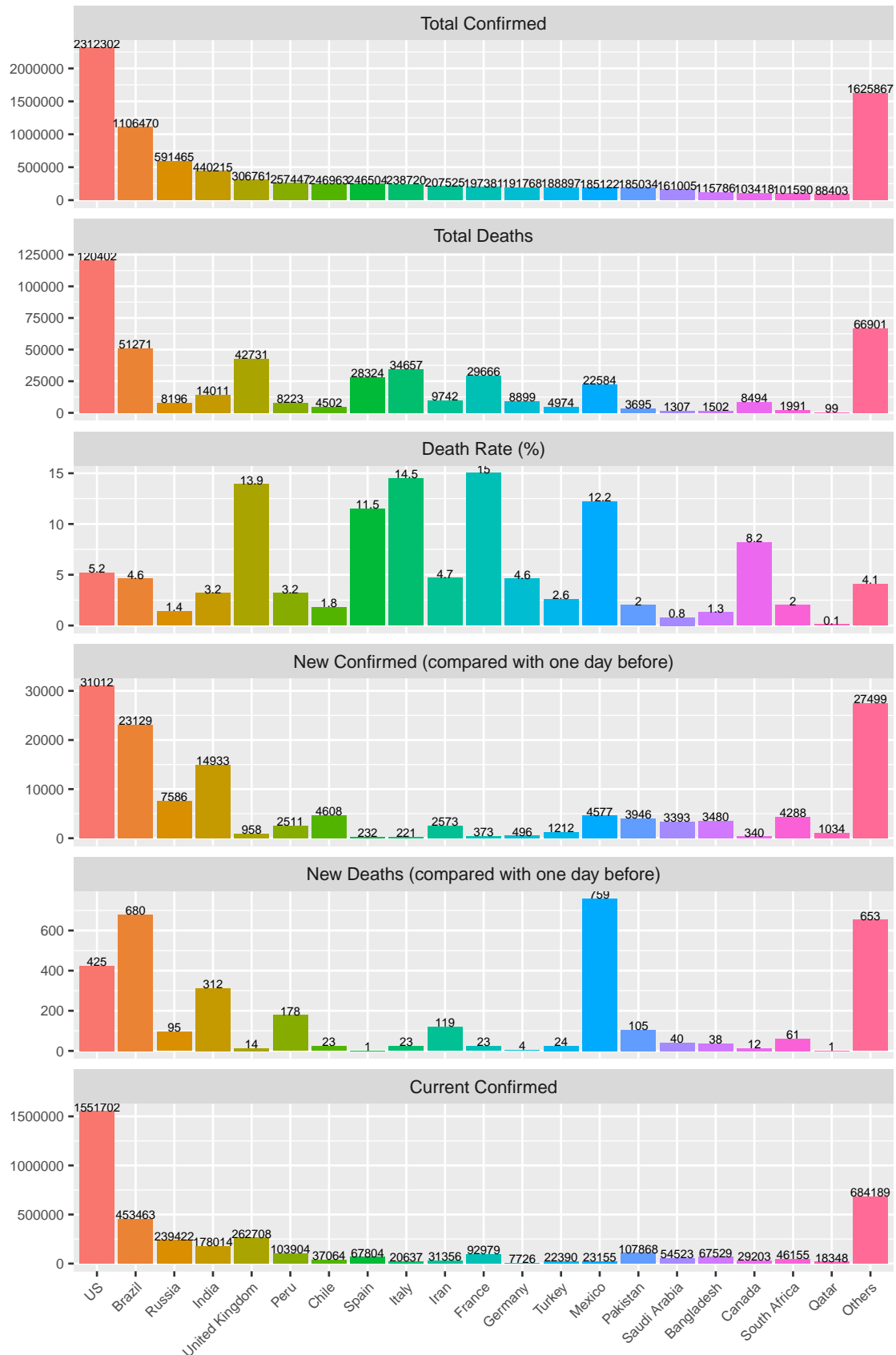
Figure 6: Top 20 Countries with Most Confirmed Cases

5.1 Confirmed vs Deaths

```r
# 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'))
```



```r
p + scale_x_log10() + scale_y_log10() +
  labs(title=paste0('Top 20 Countries (log scale)'))
```
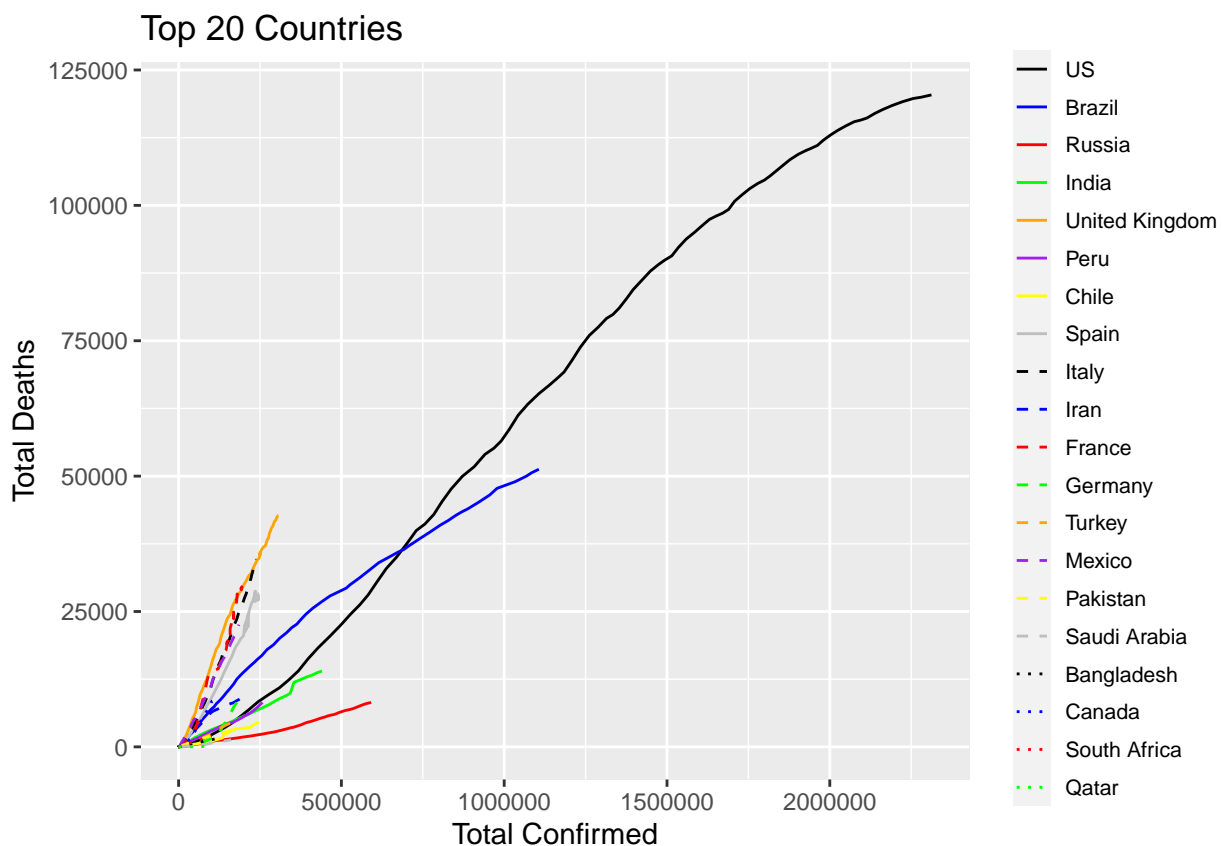
Top 20 Countries (log scale)
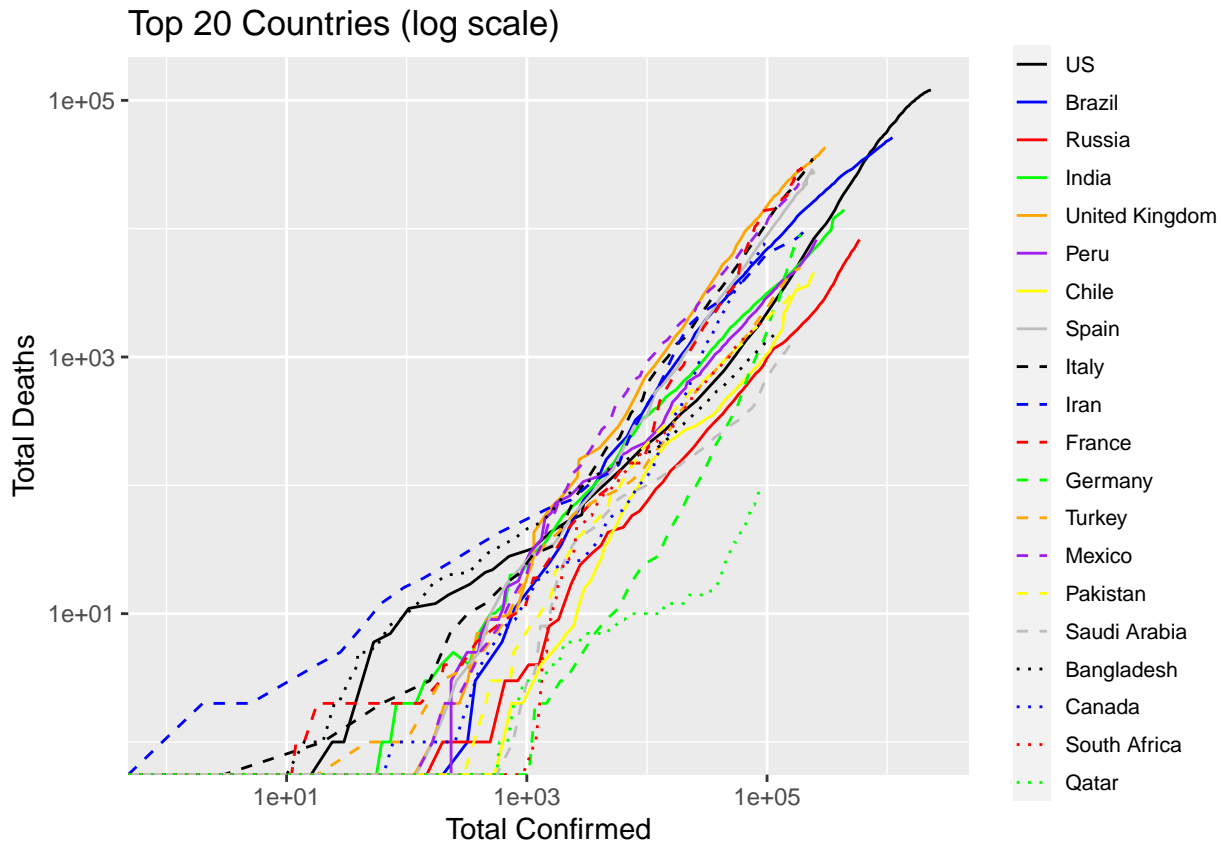
The two figures below show the numbers of confirmed cases and deaths of top 20 countries, as well as the death rates up to 22 Jun 2020 UTC.

```r
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)
```
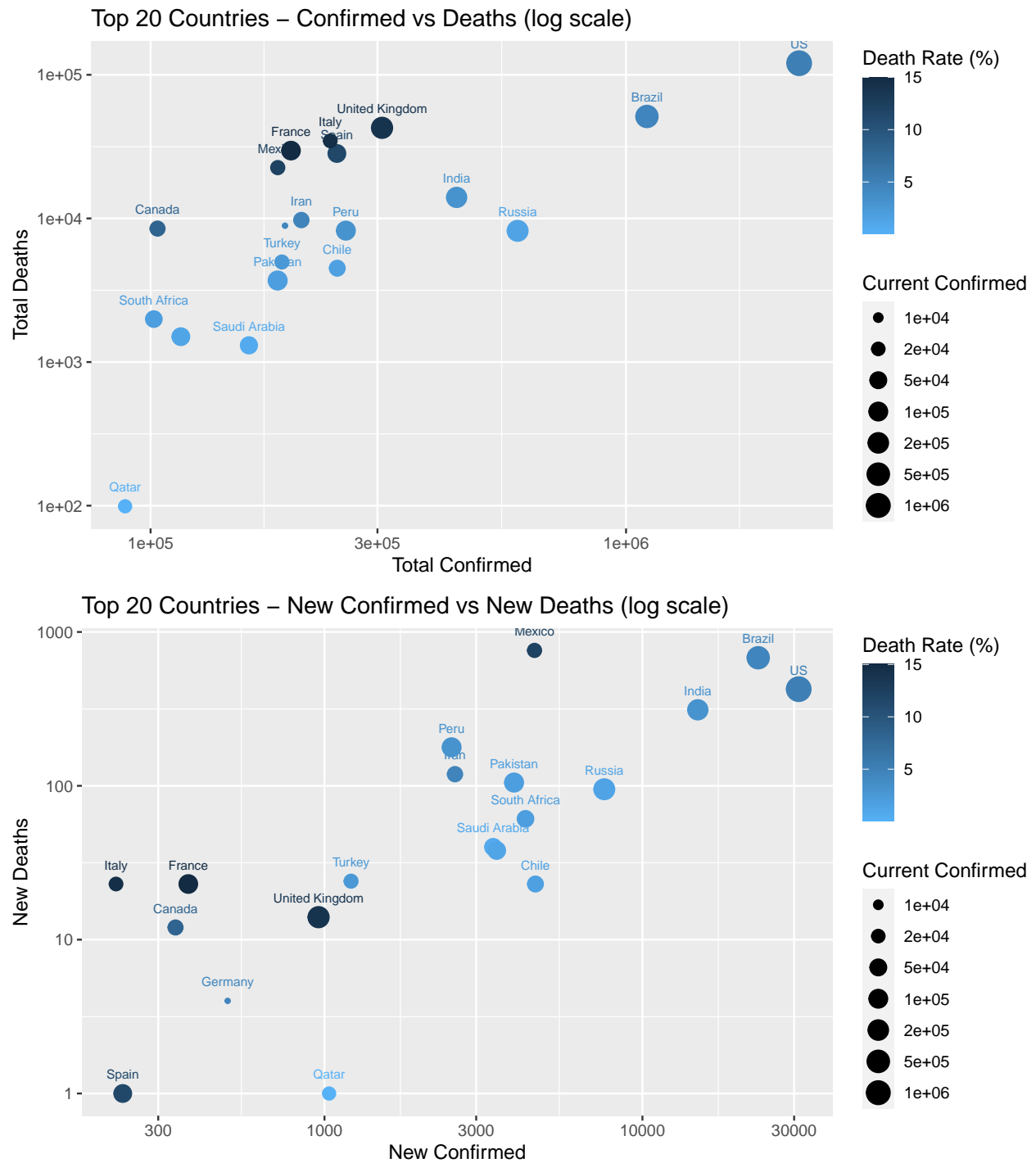

Figure 7: Top 20 Countries

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 confimed,

recovered and dead.

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

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

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

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

grid.arrange(plot1, plot2, ncol=1)
```
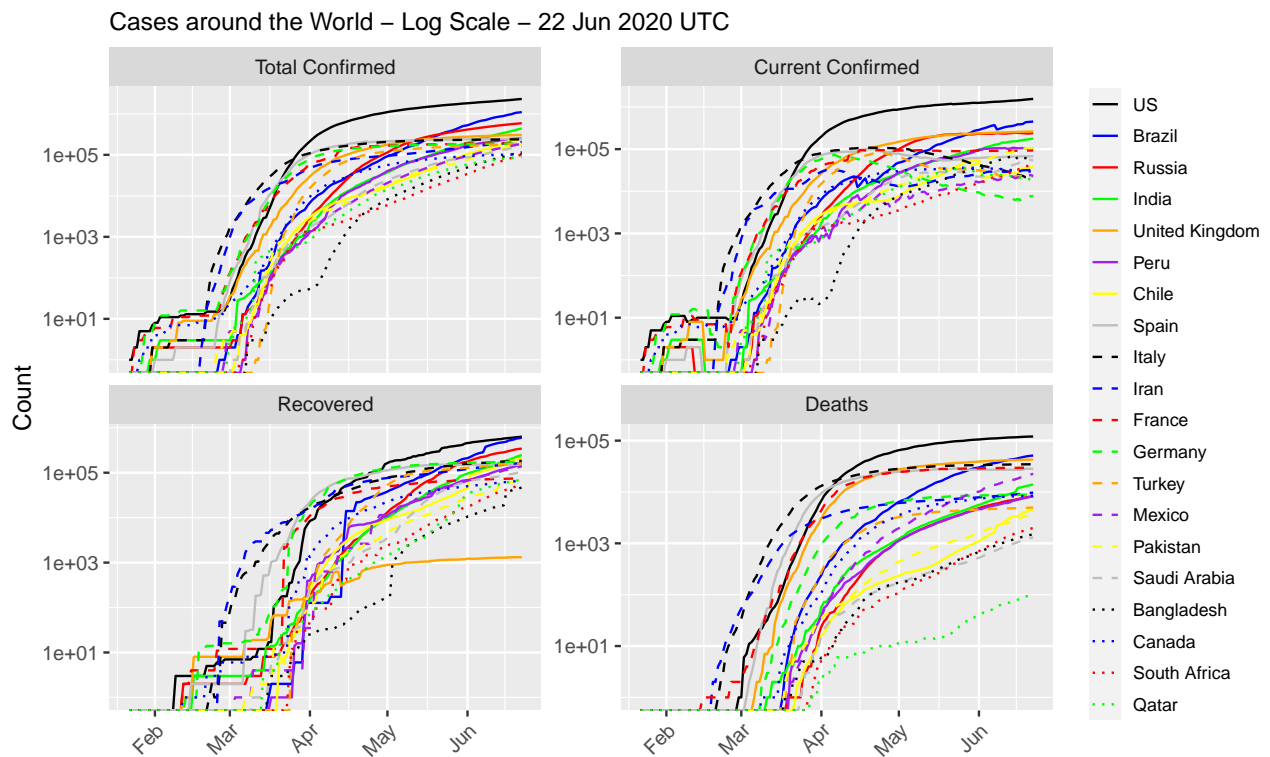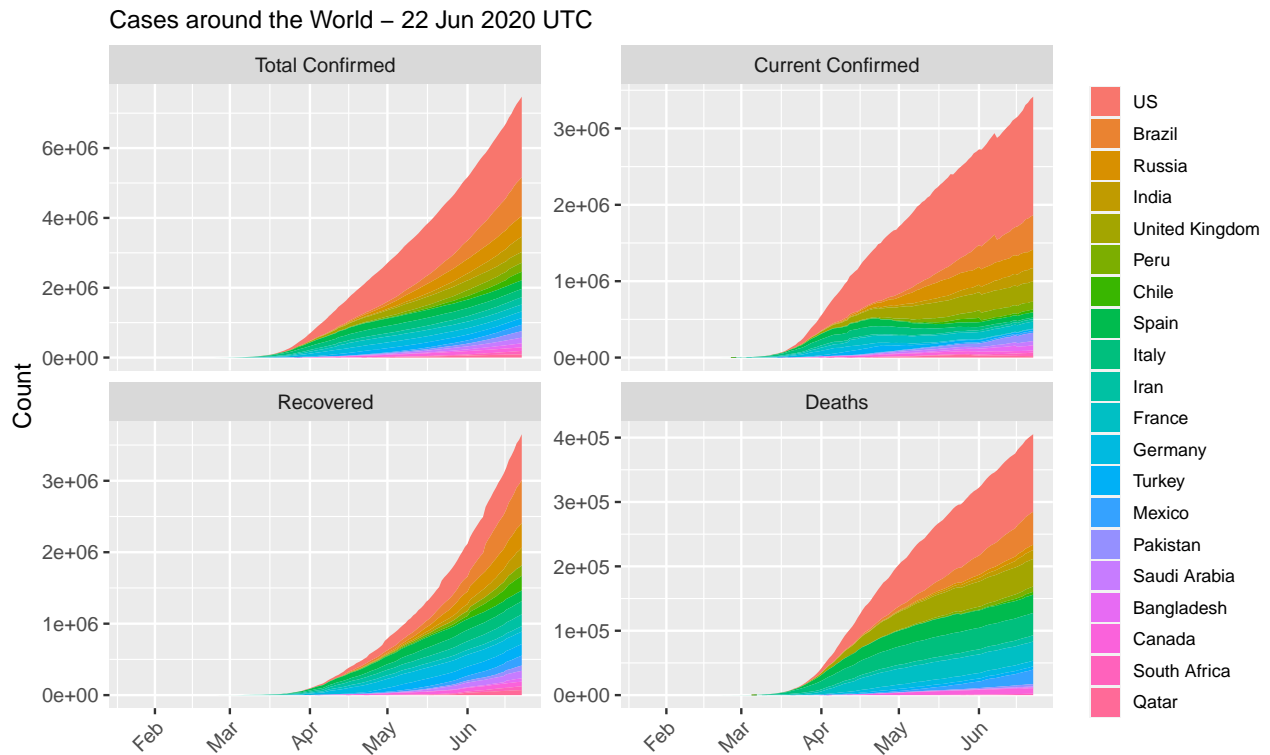
Figure 8: Cases around the World

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

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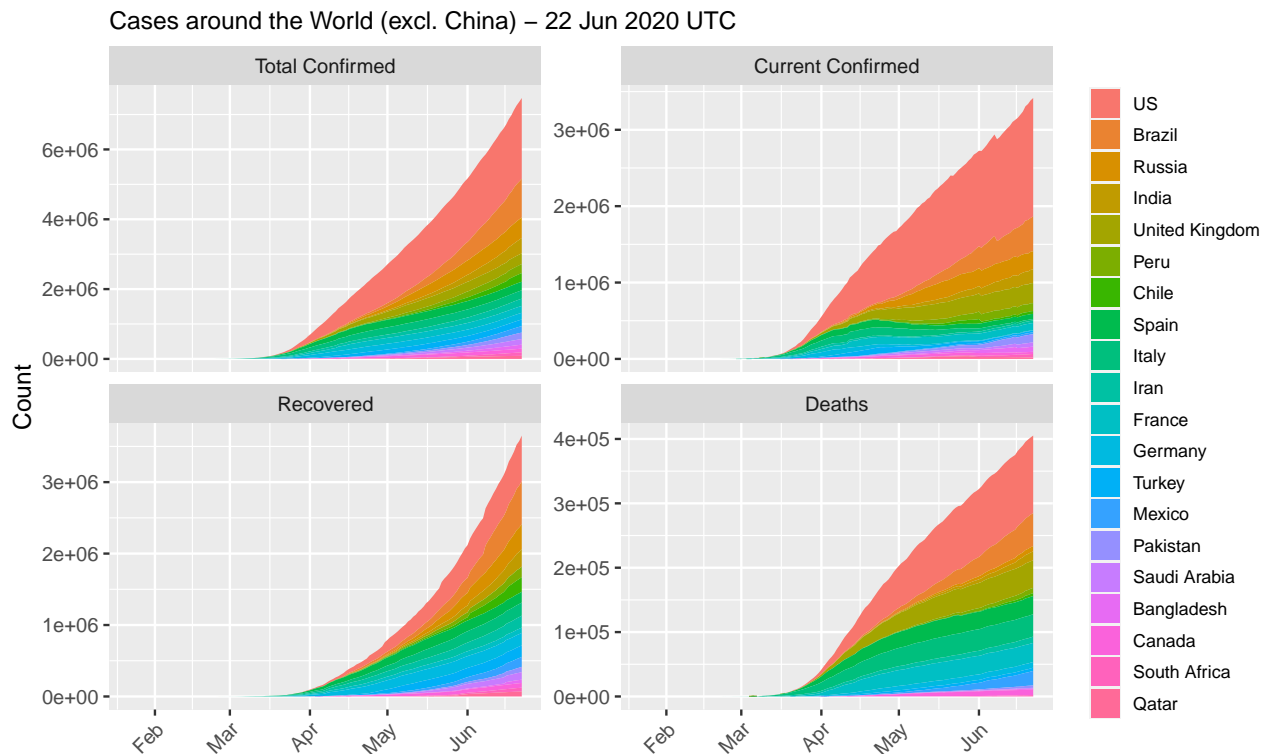

Figure 9: Cases around the World (excl. China)

```
## remove 'Others'
top.countries %<>% setdiff('Others')
## if China or Australia not in top 20, add them in
if(!('China' %in% top.countries)) {
  top.countries %<>% c('China')
}
if(!('Australia' %in% top.countries)) {
  top.countries %<>% 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) +
```

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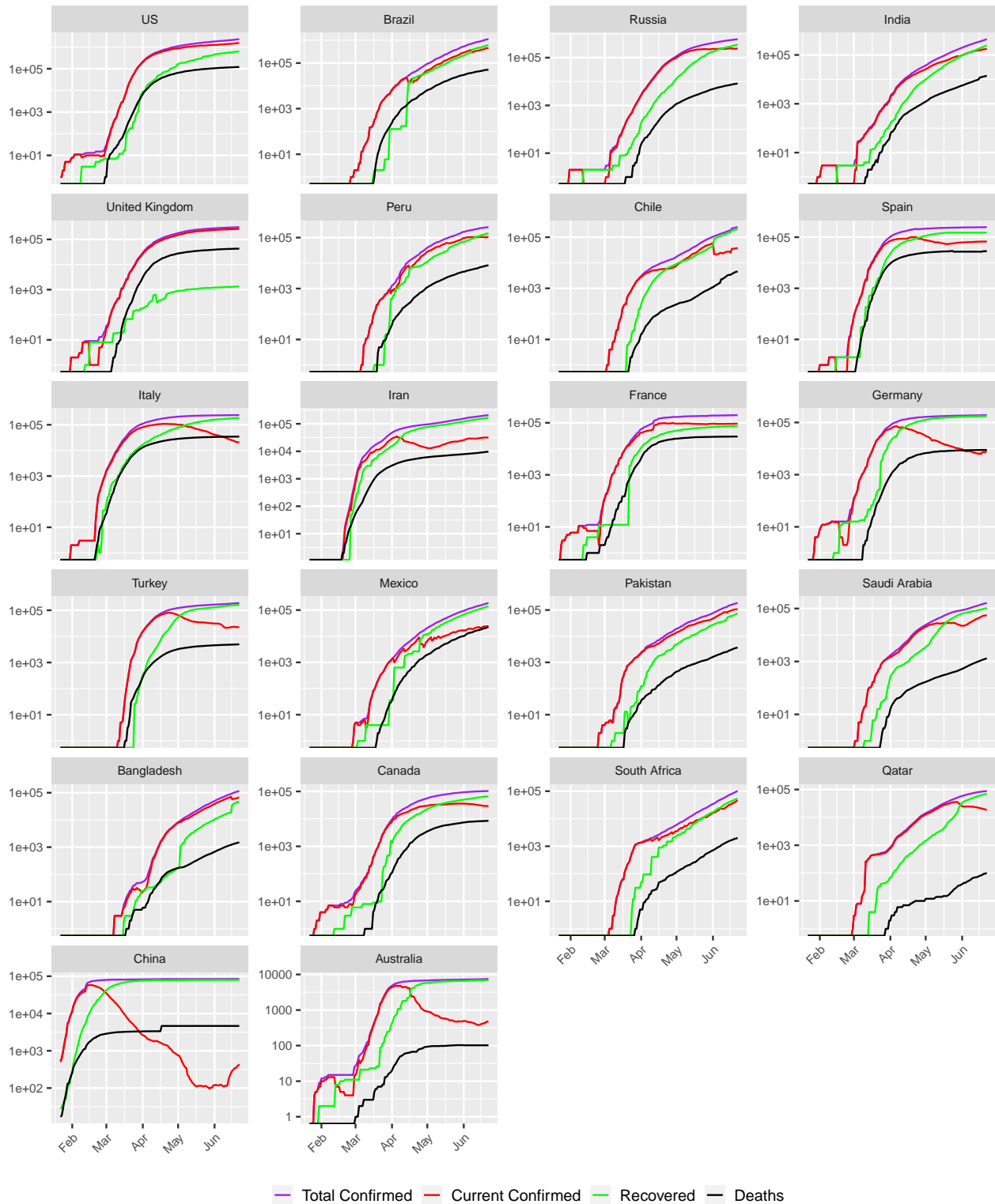
Figure 11: COVID-19 Cases Top 20 Countries (log scale). Ordered descendingly by number of confirmed cases.

```
## 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)
```
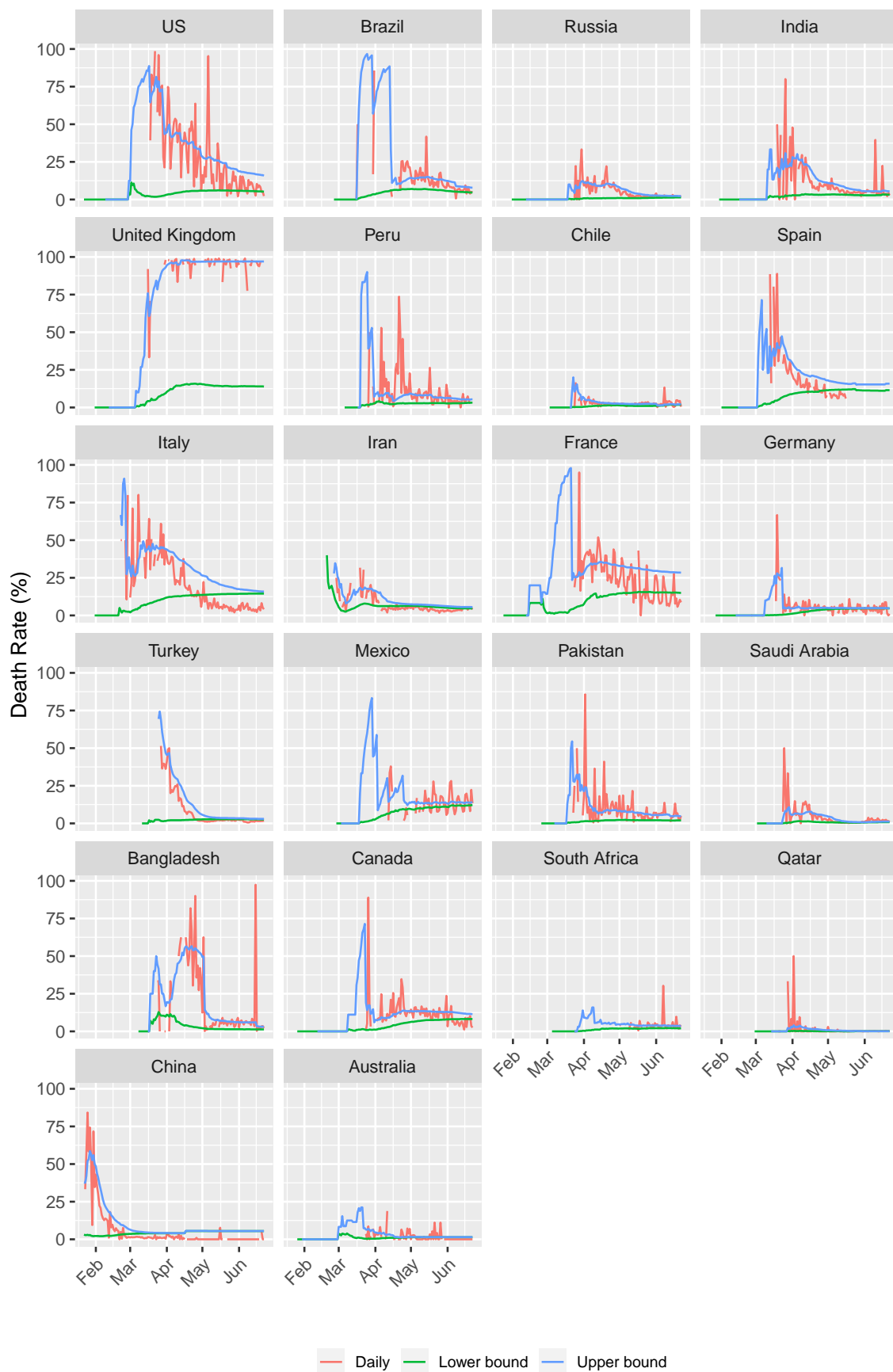
Figure 12: Death Rates

26

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 - 22 Jun 2020 UTC

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	Belgium	60,550	0	34,083	16,771	9,696	0	16.0%
2	France	197,381	373	92,979	74,736	29,666	23	15.0%
3	Italy	238,720	221	20,637	183,426	34,657	23	14.5%
4	Hungary	4,102	8	940	2,590	572	2	13.9%
5	United Kingdom	306,761	958	262,708	1,322	42,731	14	13.9%
6	Netherlands	49,866	65	43,571	186	6,109	0	12.3%
7	Mexico	185,122	4,577	23,155	139,383	22,584	759	12.2%
8	Spain	246,504	232	67,804	150,376	28,324	1	11.5%
9	Sweden	58,932	84	53,810	0	5,122	11	8.7%
10	Ecuador	50,640	0	21,426	24,991	4,223	0	8.3%
11	Canada	103,418	340	29,203	65,721	8,494	12	8.2%
12	Algeria	11,920	149	2,509	8,559	852	7	7.1%
13	Ireland	25,383	4	968	22,698	1,717	2	6.8%
14	Romania	24,291	246	5,737	17,031	1,523	11	6.3%
15	Switzerland	31,310	18	354	29,000	1,956	0	6.2%
16	Sudan	8,698	118	4,705	3,460	533	12	6.1%
17	Greece	3,287	21	1,723	1,374	190	0	5.8%
18	China	84,624	52	438	79,547	4,639	0	5.5%
19	Japan	17,820	40	908	15,957	955	0	5.4%
20	Indonesia	46,845	954	25,610	18,735	2,500	35	5.3%

6   Conclusions

As of 22 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 5.2% and 9.4%, but it is likely to change dramatically with the breakout in many countries, such as European countries.

Appendix A. Processed Data

Blow is the processed data for this analysis.

Appendix A.1 COVID-19 Cases Worldwide

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

```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
2020-06-22	9,098,643	472,171	4,526,333	4,100,139	138,401	3,590	91,705	5.2%	9.4%	3.8%
2020-06-21	8,960,242	468,581	4,434,628	4,057,033	131,234	4,060	68,696	5.2%	9.6%	5.6%
2020-06-20	8,829,008	464,521	4,365,932	3,998,555	158,685	4,253	115,825	5.3%	9.6%	3.5%
2020-06-19	8,670,323	460,268	4,250,107	3,959,948	181,347	6,289	95,008	5.3%	9.8%	6.2%
2020-06-18	8,488,976	453,979	4,155,099	3,879,898	139,026	5,020	81,144	5.3%	9.8%	5.8%
2020-06-17	8,349,950	448,959	4,073,955	3,827,036	176,010	5,274	118,786	5.4%	9.9%	4.3%
2020-06-16	8,173,940	443,685	3,955,169	3,775,086	139,479	6,786	97,831	5.4%	10.1%	6.5%
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	72,813	5,398	22,002	7.0%	21.2%	19.7%
2020-04-19	2,399,451	167,983	622,609	1,608,859	82,860	4,531	31,654	7.0%	21.2%	12.5%
2020-04-18	2,316,591	163,452	590,955	1,562,184	74,054	6,430	23,923	7.1%	21.7%	21.2%

date	confirmed	deaths	recovered	current.confirmed	new.confirmed	new.deaths	new.recovered	rate.lower	rate.upper	rate.daily
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%
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 (22 Jun 2020 UTC)

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	World	9,098,643	138,401	4,100,139	4,526,333	472,171	3,590	**5.2%**
2	US	2,312,302	31,012	1,551,702	640,198	120,402	425	**5.2%**
3	Brazil	1,106,470	23,129	453,463	601,736	51,271	680	4.6%
4	Russia	591,465	7,586	239,422	343,847	8,196	95	1.4%
5	India	440,215	14,933	178,014	248,190	14,011	312	3.2%
6	United Kingdom	306,761	958	262,708	1,322	42,731	14	**13.9%**
7	Peru	257,447	2,511	103,904	145,320	8,223	178	3.2%
8	Chile	246,963	4,608	37,064	205,397	4,502	23	1.8%
9	Spain	246,504	232	67,804	150,376	28,324	1	**11.5%**
10	Italy	238,720	221	20,637	183,426	34,657	23	**14.5%**
11	Iran	207,525	2,573	31,356	166,427	9,742	119	4.7%
12	France	197,381	373	92,979	74,736	29,666	23	**15.0%**
13	Germany	191,768	496	7,726	175,143	8,899	4	4.6%
14	Turkey	188,897	1,212	22,390	161,533	4,974	24	2.6%
15	Mexico	185,122	4,577	23,155	139,383	22,584	759	**12.2%**
16	Pakistan	185,034	3,946	107,868	73,471	3,695	105	2.0%
17	Saudi Arabia	161,005	3,393	54,523	105,175	1,307	40	0.8%
18	Bangladesh	115,786	3,480	67,529	46,755	1,502	38	1.3%
19	Canada	103,418	340	29,203	65,721	8,494	12	**8.2%**
20	South Africa	101,590	4,288	46,155	53,444	1,991	61	2.0%
21	Qatar	88,403	1,034	18,348	69,956	99	1	0.1%
22	China	84,624	52	438	79,547	4,639	0	**5.5%**
23	Colombia	71,367	2,531	39,917	29,024	2,426	73	3.4%
24	Belgium	60,550	0	34,083	16,771	9,696	0	**16.0%**
25	Belarus	59,023	518	20,749	37,923	351	5	0.6%
26	Sweden	58,932	84	53,810	0	5,122	11	**8.7%**
27	Egypt	56,809	1,576	39,398	15,133	2,278	85	4.0%
28	Ecuador	50,640	0	21,426	24,991	4,223	0	**8.3%**
29	Netherlands	49,866	65	43,571	186	6,109	0	**12.3%**
30	Indonesia	46,845	954	25,610	18,735	2,500	35	**5.3%**
31	United Arab Emirates	45,303	378	11,954	33,046	303	1	0.7%
32	Argentina	44,931	2,146	30,735	13,153	1,043	32	2.3%
33	Singapore	42,313	218	6,697	35,590	26	0	0.1%
34	Kuwait	40,291	641	8,191	31,770	330	4	0.8%
35	Portugal	39,392	259	12,310	25,548	1,534	4	3.9%
36	Ukraine	38,056	695	19,823	17,211	1,022	10	2.7%
37	Iraq	32,676	1,808	16,724	14,785	1,167	67	3.6%
38	Poland	32,227	296	13,792	17,076	1,359	3	4.2%
39	Switzerland	31,310	18	354	29,000	1,956	0	**6.2%**
40	Oman	31,076	1,605	14,531	16,408	137	6	0.4%
41	Philippines	30,682	630	21,362	8,143	1,177	8	3.8%
42	Afghanistan	29,157	324	19,718	8,841	598	17	2.1%
43	Dominican Republic	27,370	693	11,363	15,338	669	7	2.4%
44	Panama	26,752	722	11,567	14,664	521	20	1.9%
45	Bolivia	25,493	1,105	18,816	5,857	820	47	3.2%
46	Ireland	25,383	4	968	22,698	1,717	2	**6.8%**
47	Romania	24,291	246	5,737	17,031	1,523	11	**6.3%**
48	Bahrain	22,407	643	5,480	16,862	65	2	0.3%
49	Israel	21,082	304	5,014	15,761	307	1	1.5%

Table 6: Cases by Country (22 Jun 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
50	Nigeria	20,919	675	13,285	7,109	525	7	2.5%
51	Armenia	20,588	320	11,097	9,131	360	10	1.7%
52	Kazakhstan	18,231	499	6,946	11,158	127	7	0.7%
53	Japan	17,820	40	908	15,957	955	0	**5.4%**
54	Austria	17,380	39	449	16,241	690	0	4.0%
55	Moldova	14,363	163	5,864	8,019	480	7	3.3%
56	Ghana	14,154	0	3,596	10,473	85	0	0.6%
57	Guatemala	13,769	624	10,404	2,818	547	16	4.0%
58	Honduras	13,356	587	11,599	1,362	395	32	3.0%
59	Azerbaijan	13,207	478	5,878	7,168	161	7	1.2%
60	Serbia	12,990	96	731	11,997	262	1	2.0%
61	Denmark	12,727	136	578	11,547	602	2	4.7%
62	Korea, South	12,484	46	1,295	10,908	281	1	2.3%
63	Cameroon	12,041	149	3,993	7,740	308	5	2.6%
64	Algeria	11,920	149	2,509	8,559	852	7	**7.1%**
65	Czechia	10,523	25	2,650	7,537	336	0	3.2%
66	Morocco	10,172	195	1,592	8,366	214	0	2.1%
67	Nepal	9,561	535	7,390	2,148	23	0	0.2%
68	Norway	8,751	6	365	8,138	248	4	2.8%
69	Sudan	8,698	118	4,705	3,460	533	12	**6.1%**
70	Malaysia	8,587	15	289	8,177	121	0	1.4%
71	Cote d'Ivoire	7,677	185	4,128	3,493	56	2	0.7%
72	Australia	7,492	18	475	6,915	102	0	1.4%
73	Finland	7,144	1	417	6,400	327	1	4.6%
74	Uzbekistan	6,461	146	1,992	4,450	19	0	0.3%
75	Senegal	5,970	82	1,931	3,953	86	2	1.4%
76	Congo (Kinshasa)	5,924	98	4,933	856	135	5	2.3%
77	Tajikistan	5,513	56	1,422	4,039	52	0	0.9%
78	Haiti	5,211	134	5,099	24	88	0	1.7%
79	North Macedonia	5,196	90	2,975	1,974	247	9	4.8%
80	Guinea	4,988	0	1,292	3,669	27	0	0.5%
81	El Salvador	4,808	182	2,046	2,655	107	9	2.2%
82	Kenya	4,797	59	2,992	1,680	125	2	2.6%
83	Gabon	4,739	311	2,698	2,002	39	5	0.8%
84	Ethiopia	4,663	131	3,291	1,297	75	1	1.6%
85	Djibouti	4,599	17	599	3,952	48	3	1.0%
86	Luxembourg	4,121	1	52	3,959	110	0	2.7%
87	Hungary	4,102	8	940	2,590	572	2	**13.9%**
88	Venezuela	4,048	131	2,686	1,327	35	2	0.9%
89	Bulgaria	3,984	79	1,606	2,171	207	8	**5.2%**
90	Bosnia and Herzegovina	3,525	252	1,084	2,270	171	2	4.9%
91	Kyrgyzstan	3,356	0	1,295	2,021	40	0	1.2%
92	Greece	3,287	21	1,723	1,374	190	0	**5.8%**
93	Thailand	3,151	3	71	3,022	58	0	1.8%
94	Mauritania	3,121	137	2,104	905	112	1	3.6%
95	Central African Republic	2,963	155	2,438	495	30	7	1.0%
96	Somalia	2,812	33	1,904	818	90	0	3.2%
97	Croatia	2,336	19	87	2,142	107	0	4.6%
98	Cuba	2,315	3	117	2,113	85	0	3.7%
99	Costa Rica	2,277	64	1,222	1,043	12	0	0.5%
100	Maldives	2,217	14	396	1,813	8	0	0.4%
101	Kosovo	2,169	683	1,085	1,047	37	4	1.7%
102	Albania	1,995	33	792	1,159	44	0	2.2%
103	Estonia	1,981	0	147	1,765	69	0	3.5%
104	Mali	1,961	28	584	1,266	111	2	5.7%
105	Sri Lanka	1,951	1	414	1,526	11	0	0.6%
106	South Sudan	1,916	24	1,691	190	35	1	1.8%
107	Iceland	1,823	0	8	1,805	10	0	0.5%
108	Nicaragua	1,823	0	521	1,238	64	0	3.5%
109	Lithuania	1,801	3	250	1,475	76	0	4.2%
110	Equatorial Guinea	1,664	0	1,117	515	32	0	1.9%
111	Madagascar	1,640	44	933	692	15	1	0.9%
112	Lebanon	1,603	16	494	1,077	32	0	2.0%
113	Slovakia	1,588	1	113	1,447	28	0	1.8%
114	Guinea-Bissau	1,556	15	1,346	191	19	2	1.2%
115	Slovenia	1,521	1	36	1,376	109	0	7.2%
116	New Zealand	1,515	2	10	1,483	22	0	1.5%
117	Zambia	1,430	0	225	1,194	11	0	0.8%

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
118	Paraguay	1,392	13	476	903	13	0	0.9%
119	Sierra Leone	1,340	13	497	788	55	0	4.1%
120	Tunisia	1,159	2	89	1,020	50	0	4.3%
121	Latvia	1,111	0	178	903	30	0	2.7%
122	Congo (Brazzaville)	1,087	204	594	456	37	10	3.4%
123	Niger	1,046	10	66	913	67	0	6.4%
124	Jordan	1,042	9	282	751	9	0	0.9%
125	West Bank and Gaza	1,001	168	556	442	3	0	0.3%
126	Cyprus	988	2	145	824	19	0	1.9%
127	Yemen	967	26	360	350	257	1	26.6%
128	Cabo Verde	944	54	517	419	8	0	0.8%
129	Georgia	908	2	133	761	14	0	1.5%
130	Burkina Faso	903	0	36	814	53	0	5.9%
131	Uruguay	882	6	42	815	25	0	2.8%
132	Chad	858	0	29	755	74	0	8.6%
133	Andorra	855	0	7	796	52	0	6.1%
134	Benin	807	42	541	253	13	0	1.6%
135	Rwanda	787	59	415	370	2	0	0.3%
136	Uganda	774	4	143	631	0	0	0.0%
137	Malawi	749	19	480	258	11	0	1.5%
138	Mozambique	737	4	551	181	5	0	0.7%
139	Diamond Princess	712	0	48	651	13	0	1.8%
140	Sao Tome and Principe	702	4	487	203	12	0	1.7%
141	San Marino	697	1	28	627	42	0	6.0%
142	Jamaica	665	6	139	516	10	0	1.5%
143	Malta	665	0	39	617	9	0	1.4%
144	Liberia	650	24	356	260	34	0	5.2%
145	Eswatini	643	8	346	291	6	1	0.9%
146	Libya	595	24	469	116	10	0	1.7%
147	Togo	569	0	176	380	13	0	2.3%
148	Zimbabwe	512	23	442	64	6	0	1.2%
149	Tanzania	509	0	305	183	21	0	4.1%
150	Taiwan*	446	0	4	435	7	0	1.6%
151	Montenegro	367	5	43	315	9	0	2.5%
152	Vietnam	349	0	21	328	0	0	0.0%
153	Mauritius	340	3	4	326	10	0	2.9%
154	Suriname	319	5	179	132	8	0	2.5%
155	Burma	291	1	85	200	6	0	2.1%
156	Comoros	247	0	83	159	5	0	2.0%
157	Syria	219	15	129	83	7	0	3.2%
158	Mongolia	215	2	57	158	0	0	0.0%
159	Guyana	205	21	90	103	12	0	5.9%
160	Angola	186	3	99	77	10	1	5.4%
161	Burundi	144	0	50	93	1	0	0.7%
162	Eritrea	143	0	104	39	0	0	0.0%
163	Brunei	141	0	0	138	3	0	2.1%
164	Cambodia	130	1	3	127	0	0	0.0%
165	Trinidad and Tobago	123	0	6	109	8	0	6.5%
166	Bahamas	104	0	16	77	11	0	10.6%
167	Monaco	101	1	2	95	4	0	4.0%
168	Barbados	97	0	5	85	7	0	7.2%
169	Botswana	89	0	63	25	1	0	1.1%
170	Liechtenstein	83	1	13	69	1	0	1.2%
171	Bhutan	68	0	36	32	0	0	0.0%
172	Namibia	63	8	42	21	0	0	0.0%
173	Gambia	41	4	13	26	2	0	4.9%
174	Saint Vincent and the Grenadines	29	0	3	26	0	0	0.0%
175	Antigua and Barbuda	26	0	1	22	3	0	11.5%
176	Timor-Leste	24	0	0	24	0	0	0.0%
177	Belize	23	1	4	17	2	0	8.7%
178	Grenada	23	0	0	23	0	0	0.0%
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	0	18	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	Lesotho	12	8	10	2	0	0	0.0%

Table 6: Cases by Country (22 Jun 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
186	Seychelles	11	0	0	11	0	0	0.0%
187	Western Sahara	10	1	1	8	1	0	10.0%
188	MS Zaandam	9	0	7	0	2	0	22.2%
189	Papua New Guinea	9	1	1	8	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!