COVID-19 Data Analysis with R - Worldwide*

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

20 May 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] 266 123
```

Each dataset has 266 rows, corresponding to country/region/province/state. It has 123 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" "2020-05-19"
```

```
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 19 May 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 19 May 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
```

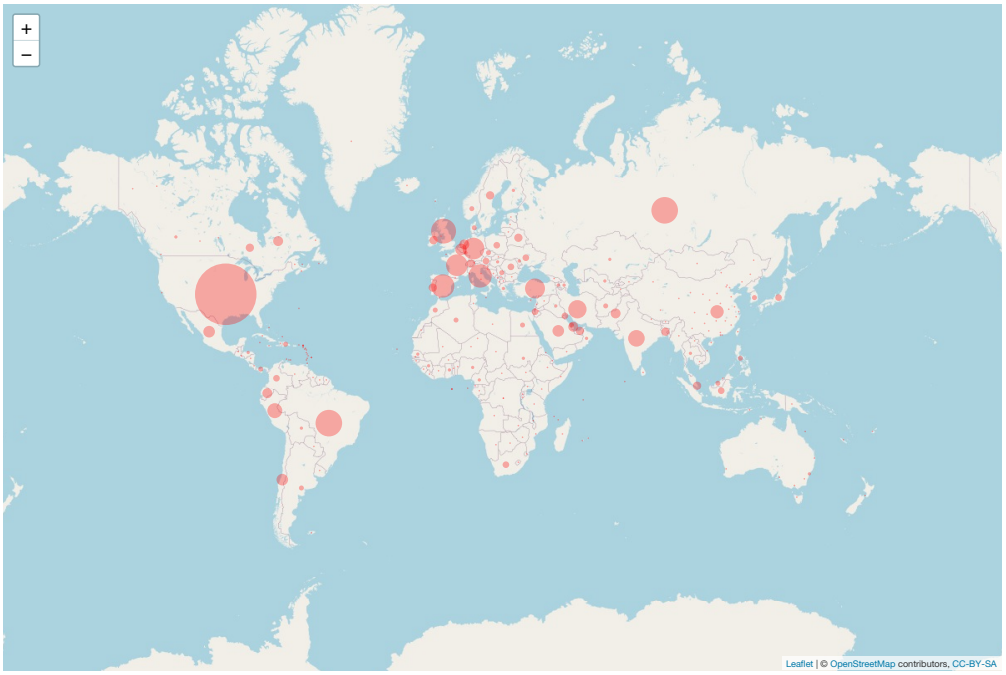


Figure 1: World Map

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)
```

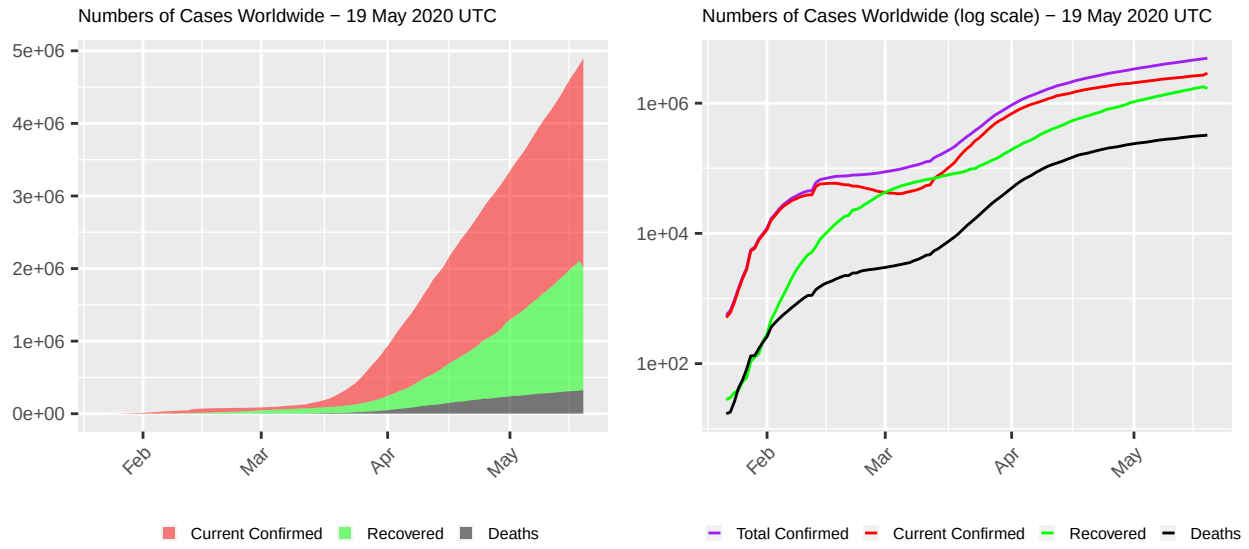


Figure 2: COVID-19 Cases Worldwide

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)
```

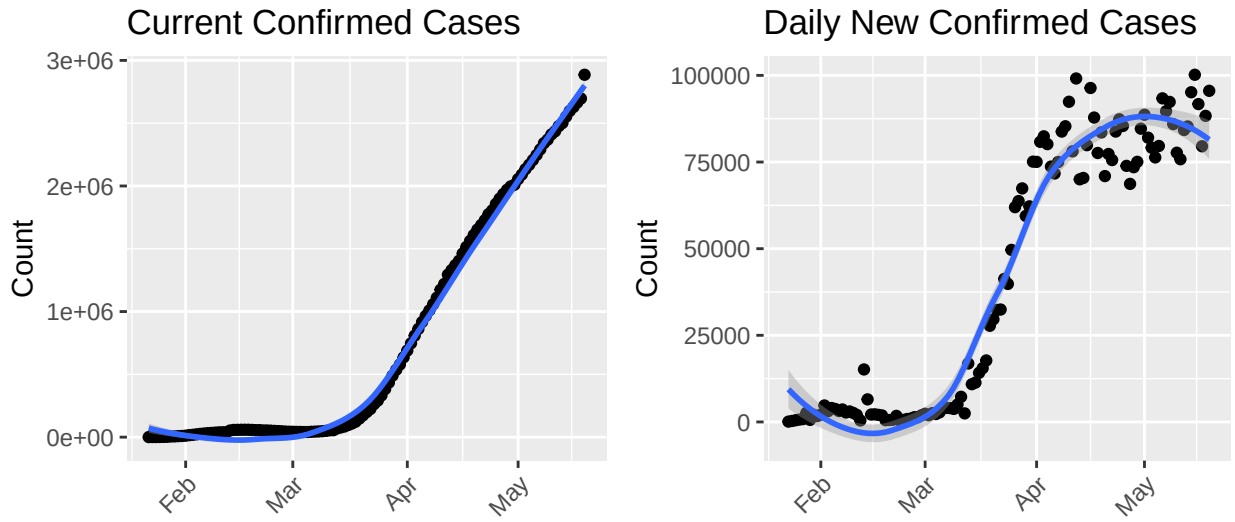


Figure 3: Current Confirmed Cases

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)
```

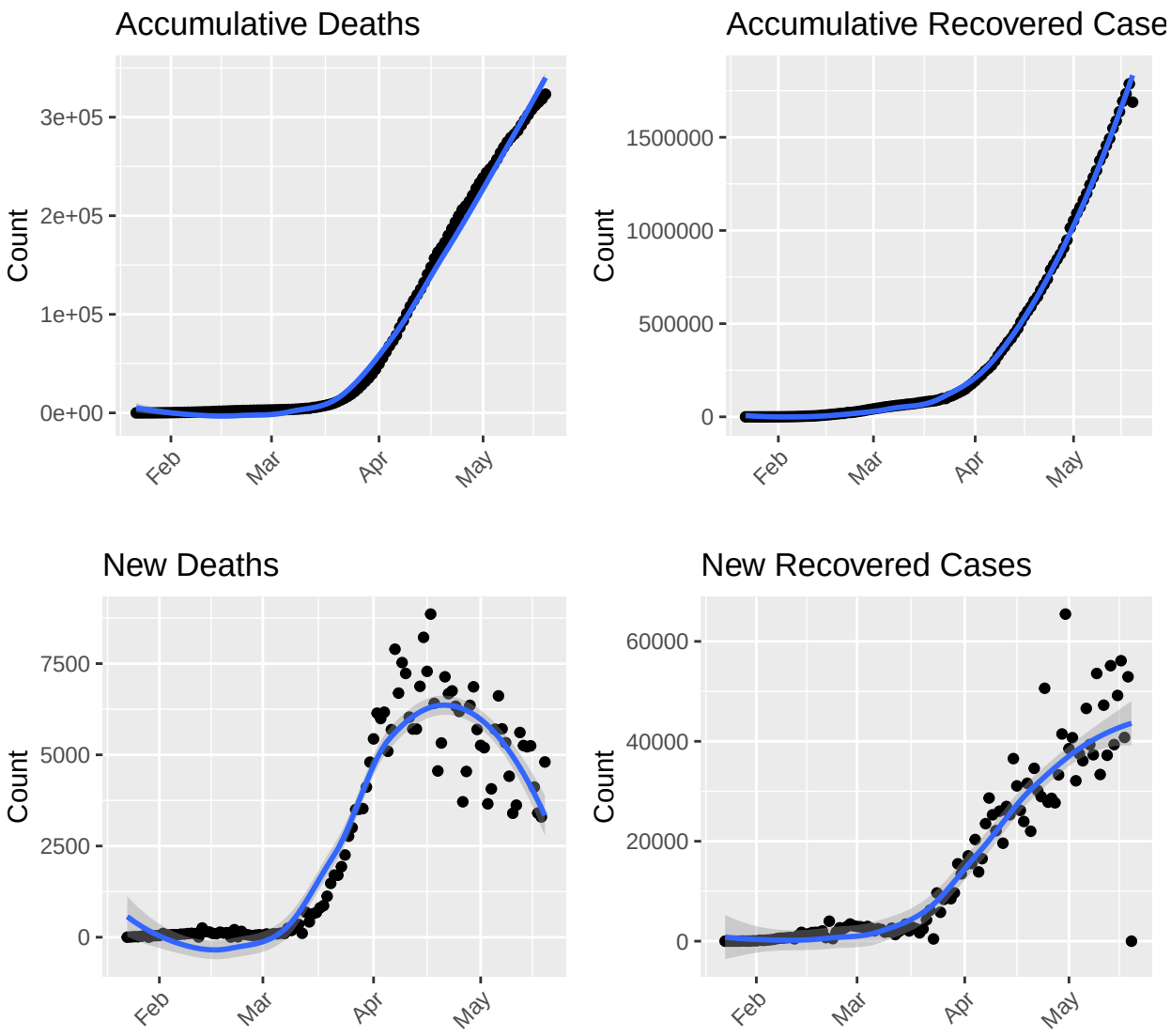


Figure 4: Deaths and Recovered Cases

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 19 May 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 19 May 2020 UTC UTC, it will be between 6.6% and 16.1%.

A surge in the daily death rate (in red) suggests that the situlation is changing dramatically (actually, getting worse) and that above lower/upper bounds are likely to increase shortly. A likely reason of that surge is the outbreak of coronavirus in Iran, European and US.

```
## three death rates
plot1 <- ggplot(data.world, aes(x=date)) +
```

```
  geom_line(aes(y=rate.upper, colour='Upper bound')) +
  geom_line(aes(y=rate.lower, colour='Lower bound')) +
  geom_line(aes(y=rate.daily, colour='Daily')) +
  xlab('') + ylab('Death Rate (%)') + labs(title='Overall') +
  theme(legend.position='bottom', legend.title=element_blank(),
        legend.text=element_text(size=8),
        legend.key.size=unit(0.5, 'cm'),
        axis.text.x=element_text(angle=45, hjust=1)) +
  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, 25))
grid.arrange(plot1, plot2, ncol=2)
```

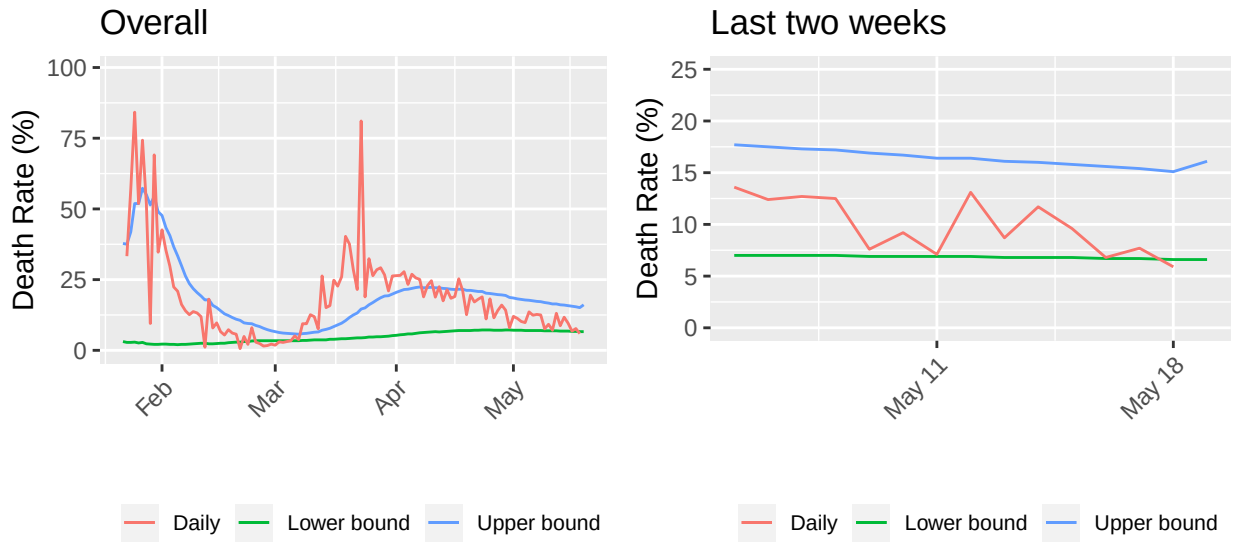


Figure 5: Death Rate

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

```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))
```

```r
data.latest %>% mutate(death.rate=death.rate %>% format(nsmall=1) %>% paste0('%')) %>%
  kable('latex', booktabs=T, row.names=T, align=c('l', rep('r', 6)),
        caption=paste0('Cases in Top 20 Countries - ', max.date.txt,
                       '. See a complete list of all infected countries at the end of this report.'),
        format.args=list(big.mark=',')) %>%
  kable_styling(font_size=7, latex_options=c('striped', 'hold_position', 'repeat_header'))
```

```r
## convert from wide to long format, for drawing area plots
data.latest.long <- data.latest %>% filter(country!='World') %>%
  gather(key=type, value=count, -country)
## set factor levels to show them with proper text and in a desirable order
data.latest.long %<>% mutate(type=recode_factor(type,
                                    confirmed='Total Confirmed',
                                    deaths='Total Deaths',
                                    death.rate='Death Rate (%)',
                                    new.confirmed='New Confirmed (compared with one day before)',
                                    new.deaths='New Deaths (compared with one day before)',
                                    current.confirmed='Current Confirmed'))
```

```r
## bar chart
data.latest.long %>% ggplot(aes(x=country, y=count, fill=country, group=country)) +
  geom_bar(stat='identity') +
  geom_text(aes(label=count, y=count), size=2, vjust=0) +
  xlab('') + ylab('') +
  labs(title=paste0('Top 20 Countries with Most Confirmed Cases - ', max.date.txt)) +
  scale_fill_discrete(name='Country', labels=aes(count)) +
  theme(legend.title=element_blank(),
        legend.position='none',
        plot.title=element_text(size=11),
```

Table 3: Cases in Top 20 Countries - 19 May 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	4,897,492	323,285	6.6%	95,549	4,804	2,885,588
2	US	1,528,568	91,921	6.0%	20,260	1,574	1,147,255
3	Russia	299,941	2,837	0.9%	9,263	115	220,974
4	Brazil	271,885	17,983	6.6%	16,517	1,130	147,108
5	United Kingdom	250,138	35,422	14.2%	2,429	546	213,617
6	Spain	232,037	27,778	12.0%	431	69	204,259
7	Italy	226,699	32,169	14.2%	813	162	65,129
8	France	180,933	28,025	15.5%	882	0	90,230
9	Germany	177,778	8,081	4.5%	1,227	78	14,016
10	Turkey	151,615	4,199	2.8%	1,022	28	34,521
11	Iran	124,603	7,119	5.7%	2,111	62	20,311
12	India	106,475	3,302	3.1%	6,147	146	60,864
13	Peru	99,483	2,914	2.9%	4,550	125	60,045
14	China	84,063	4,638	5.5%	0	0	115
15	Canada	80,493	6,028	7.5%	1,082	68	34,396
16	Saudi Arabia	59,854	329	0.5%	2,509	9	27,891
17	Belgium	55,791	9,108	16.3%	232	28	31,996
18	Mexico	54,346	5,666	10.4%	2,713	334	11,355
19	Chile	49,579	509	1.0%	3,520	31	27,563
20	Netherlands	44,449	5,734	12.9%	108	21	38,548
21	Pakistan	43,966	939	2.1%	1,841	36	30,538
22	Others	774,796	28,584	3.7%	18,101	459	404,857

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

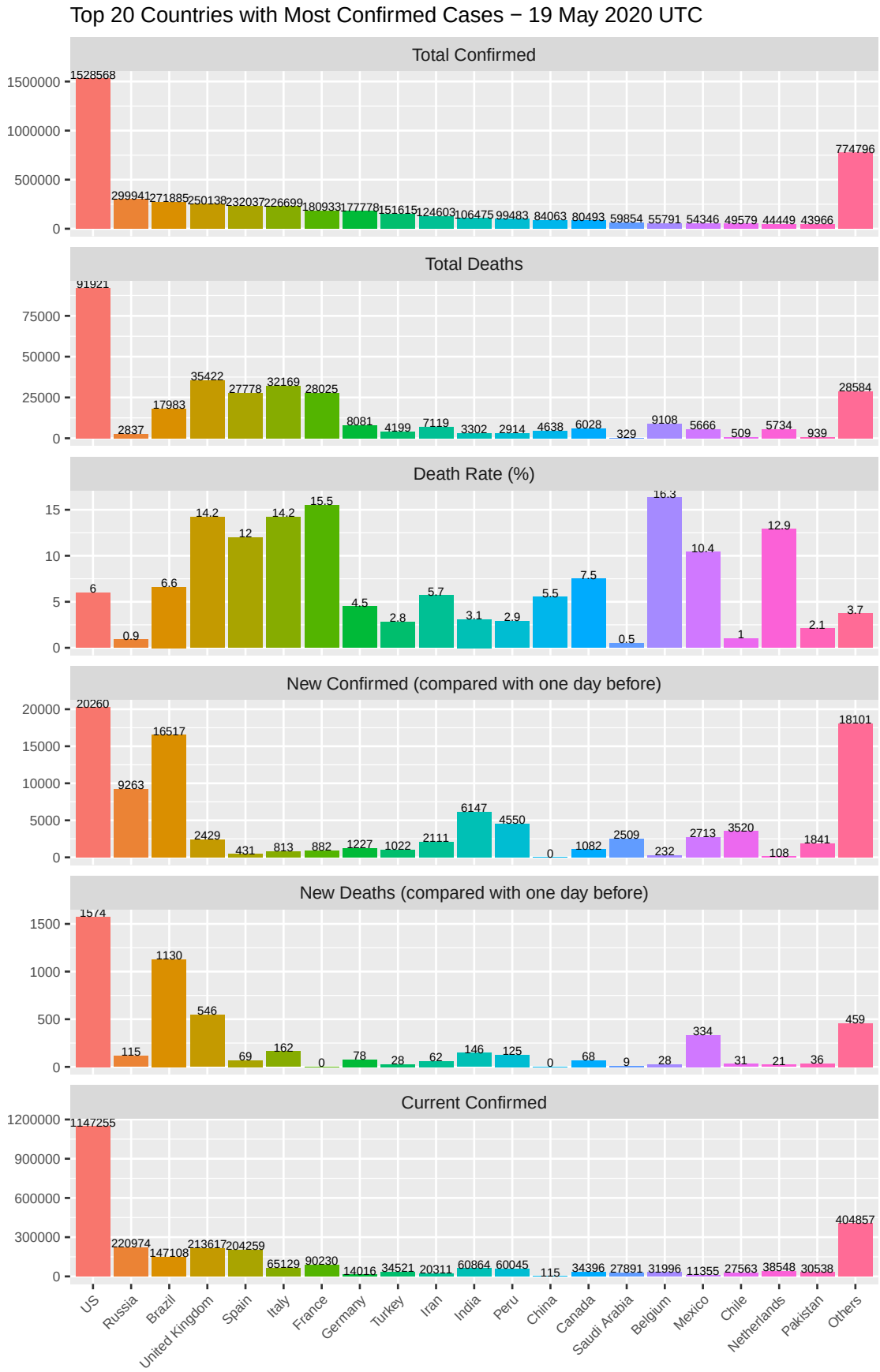
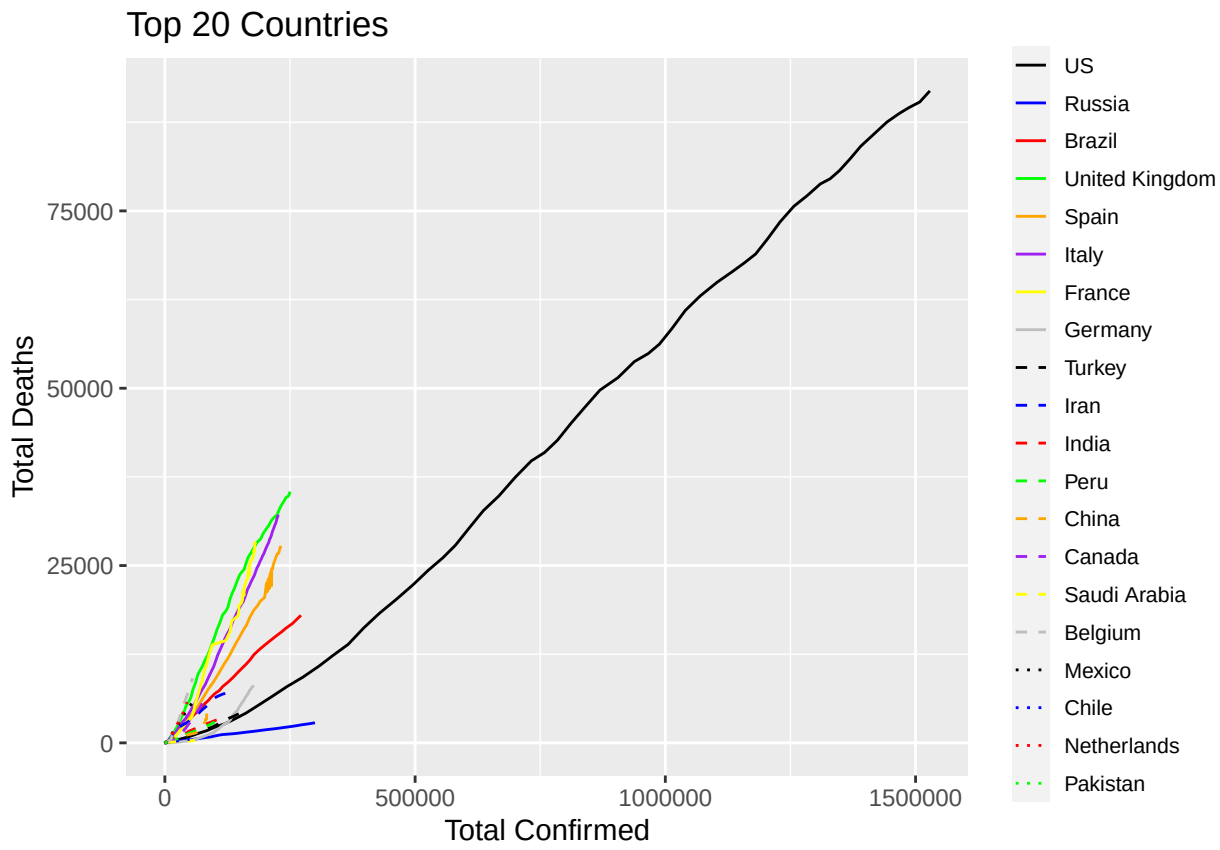


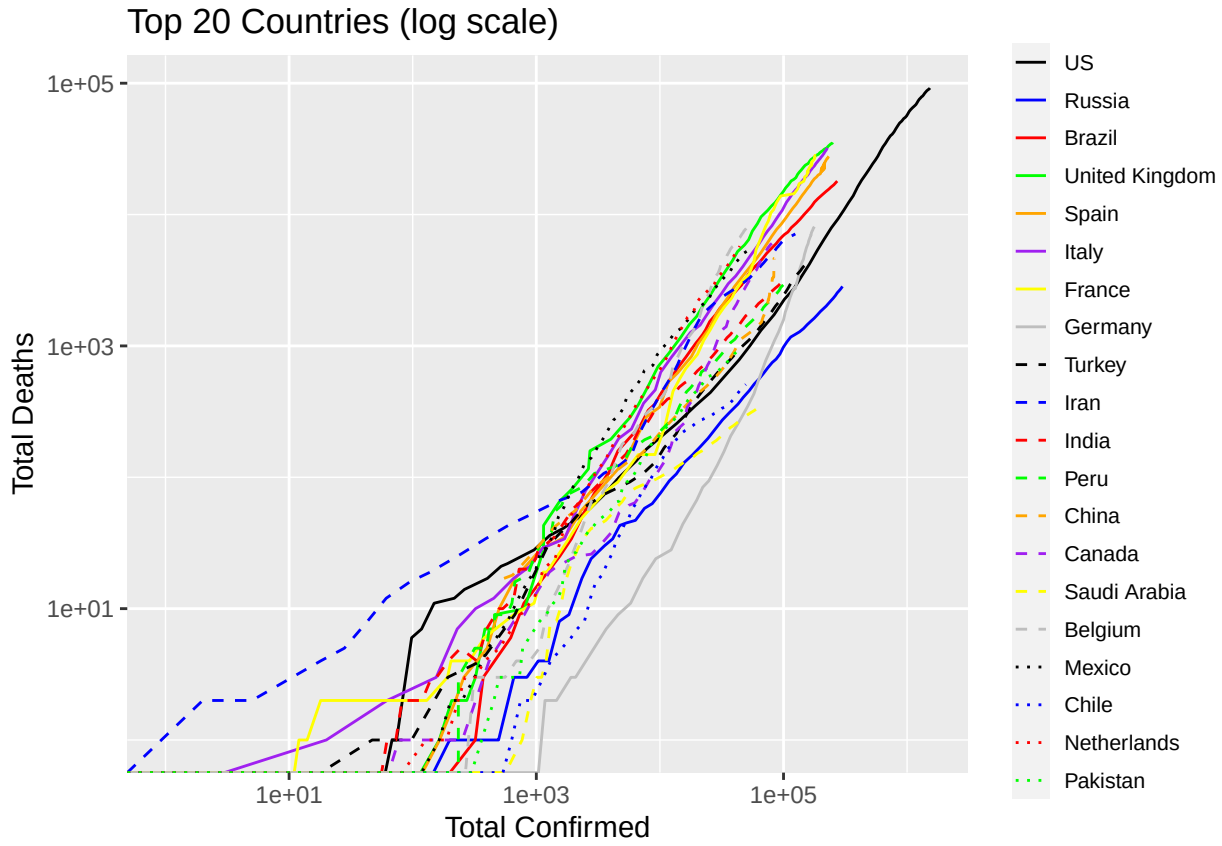
Figure 6: Top 20 Countries with Most Confirmed Cases

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 19 May 2020 UTC.

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

plot1 <- df %>% ggplot(aes(x=confirmed, y=deaths, col=death.rate, size=current.confirmed)) +
  scale_size(name='Current Confirmed', trans='log2', breaks=breaks.confirmed) +
  geom_text(aes(label=country), size=2.5, check_overlap=T, vjust=-1.6) +
  geom_point() +
  xlab('Total Confirmed') + ylab('Total Deaths') +
  labs(col="Death Rate (%)") +
  scale_color_gradient(low='#56B1F7', high='#132B43') +
  scale_x_log10() + scale_y_log10() +
  labs(title=paste0('Top 20 Countries - Confirmed vs Deaths (log scale)'))

plot2 <- df %>% ggplot(aes(x=new.confirmed, y=new.deaths, col=death.rate, size=current.confirmed)) +
  scale_size(name='Current Confirmed', trans='log2', breaks=breaks.confirmed) +
  geom_text(aes(label=country), size=2.5, check_overlap=T, vjust=-1.6) +
  geom_point() +
  xlab('New Confirmed') + ylab('New Deaths') +
  labs(col="Death Rate (%)") +
  scale_color_gradient(low='#56B1F7', high='#132B43') +
  scale_x_log10() + scale_y_log10() +
  labs(title=paste0('Top 20 Countries - New Confirmed vs New Deaths (log scale)'))
```

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

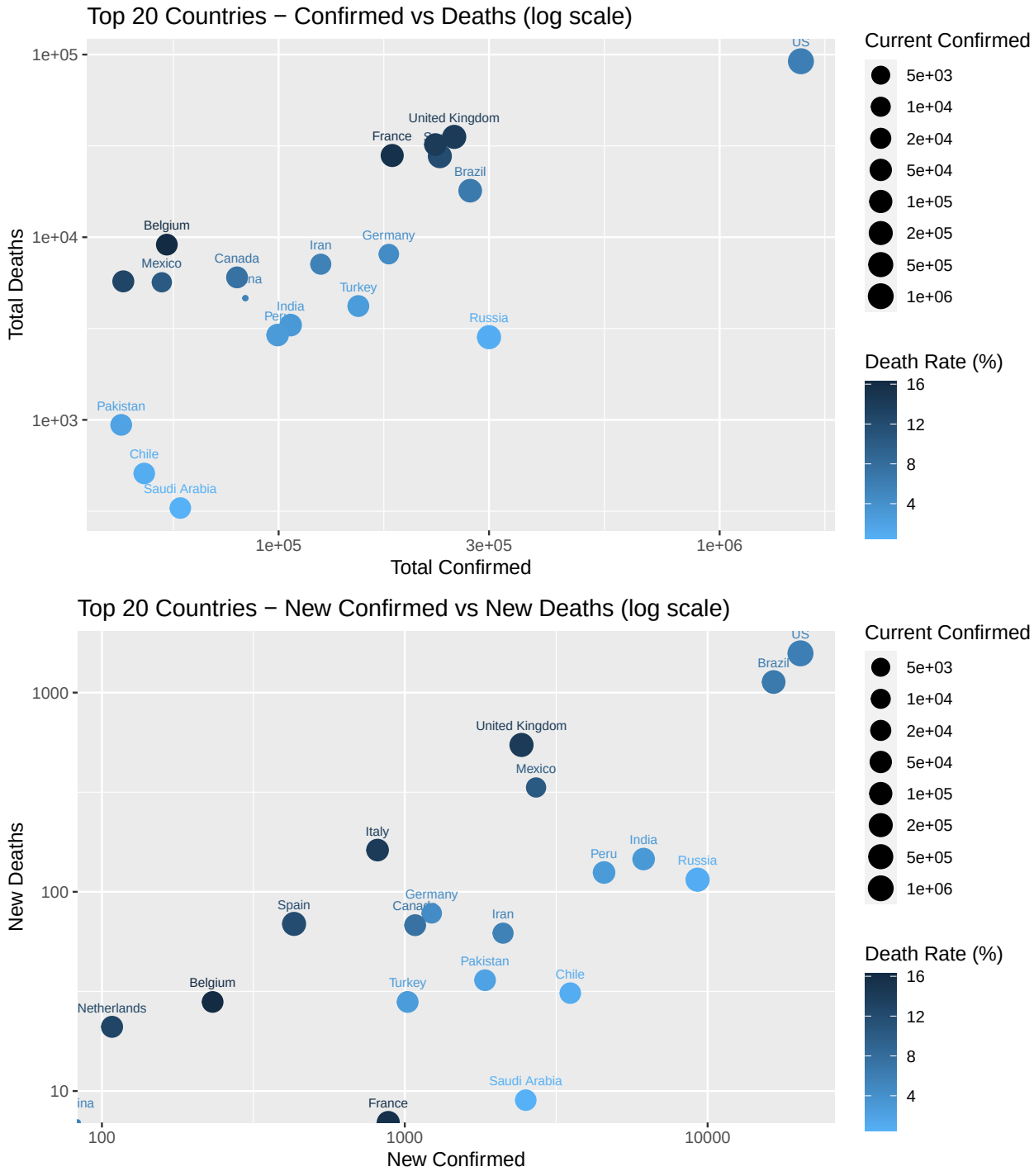


Figure 7: Top 20 Countries

5.2 Comparison across Countries

The area plots blow show the numbers of dead, recovered, total and current confimed cases. Note that, in the area plot, the total number of total confirmed cases is represented by the total areas of current confirmed,

recovered and dead.

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

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

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

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

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

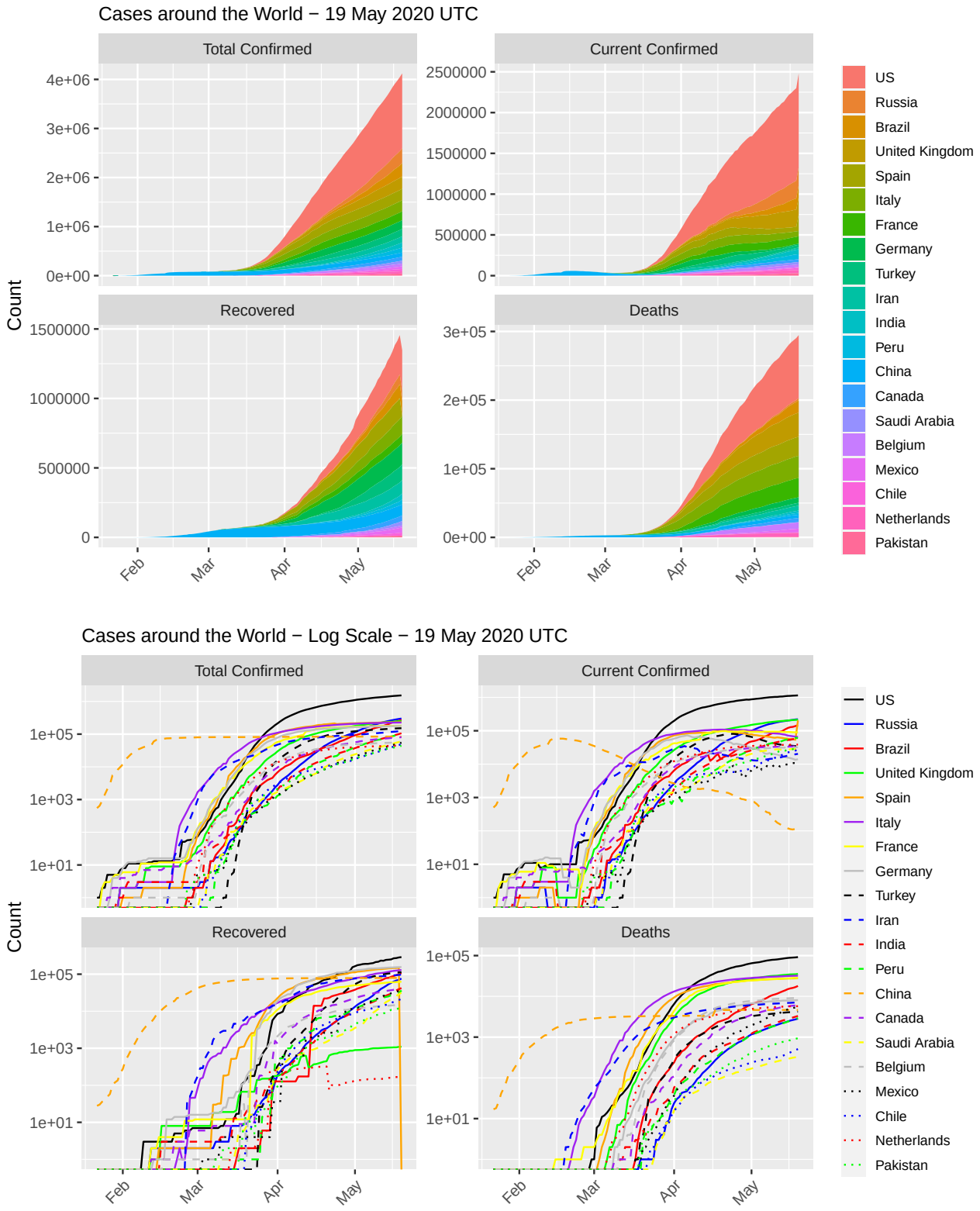


Figure 8: Cases around the World

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

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

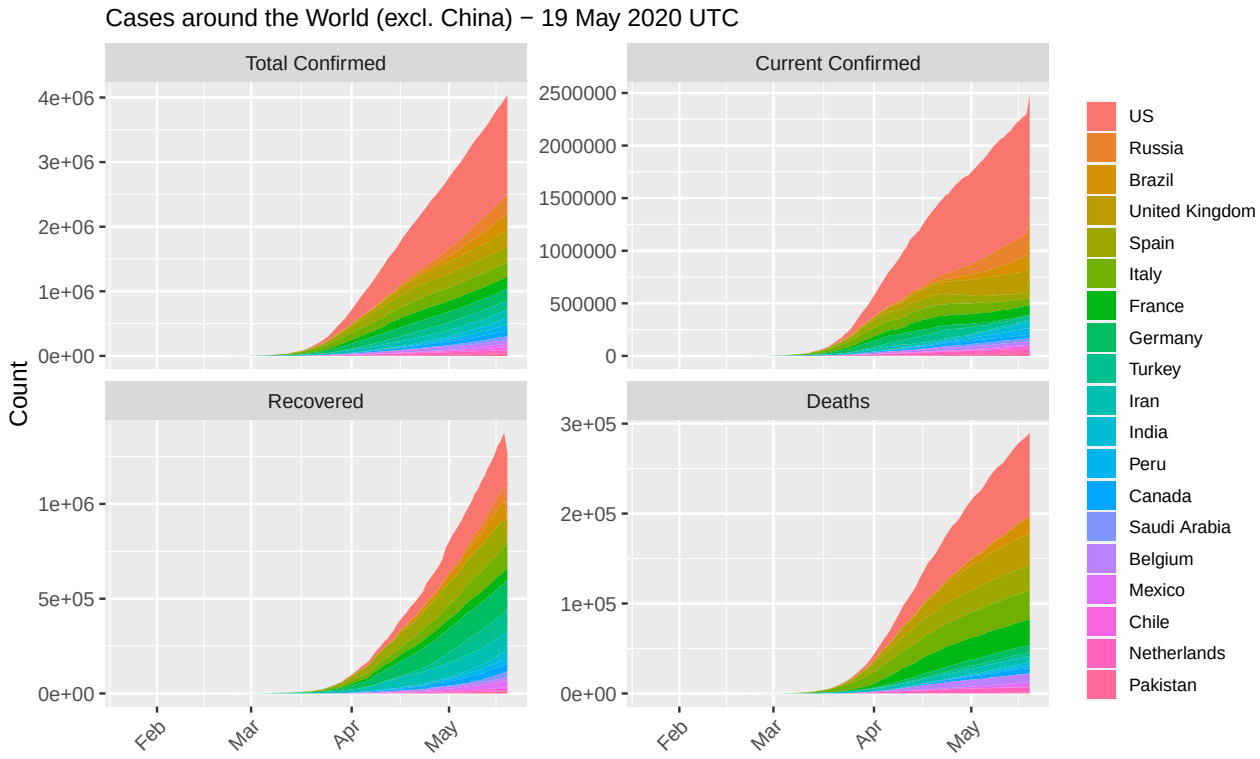


Figure 9: Cases around the World (excl. China)

```
## if Australia in not in top 20, add it in and remove 'Others'
if(!('Australia' %in% top.countries)) {
  top.countries %<>% setdiff('Others') %>% c('Australia')
  df <- data.long %>% filter(country %in% top.countries) %<>%
    mutate(country=country %>% factor(levels=c(top.countries)))
}

## cases by country - area plot
df %>% filter(country != 'World' & type != 'Total Confirmed') %>%
  ggplot(aes(x=date, y=count, fill=type)) +
  geom_area(alpha=0.5) +
  # xlab('') + ylab('') +
  labs(title=paste0('Numbers of COVID-19 Cases in Top 20 Countries - ',
                    max.date.txt)) +
  scale_fill_manual(values=c('red', 'green', 'black')) +
  theme(legend.title=element_blank(), legend.position='bottom',
```

```
      plot.title = element_text(size=12),
      axis.title.x=element_blank(),
      axis.title.y=element_blank(),
      legend.key.size=unit(0.4, 'cm'),
      # legend.text=element_text(size=7),
      strip.text.x=element_text(size=7),
      axis.text=element_text(size=7),
      axis.text.x=element_text(angle=45, hjust=1)) +
facet_wrap(~country, ncol=4, scales='free_y')
```

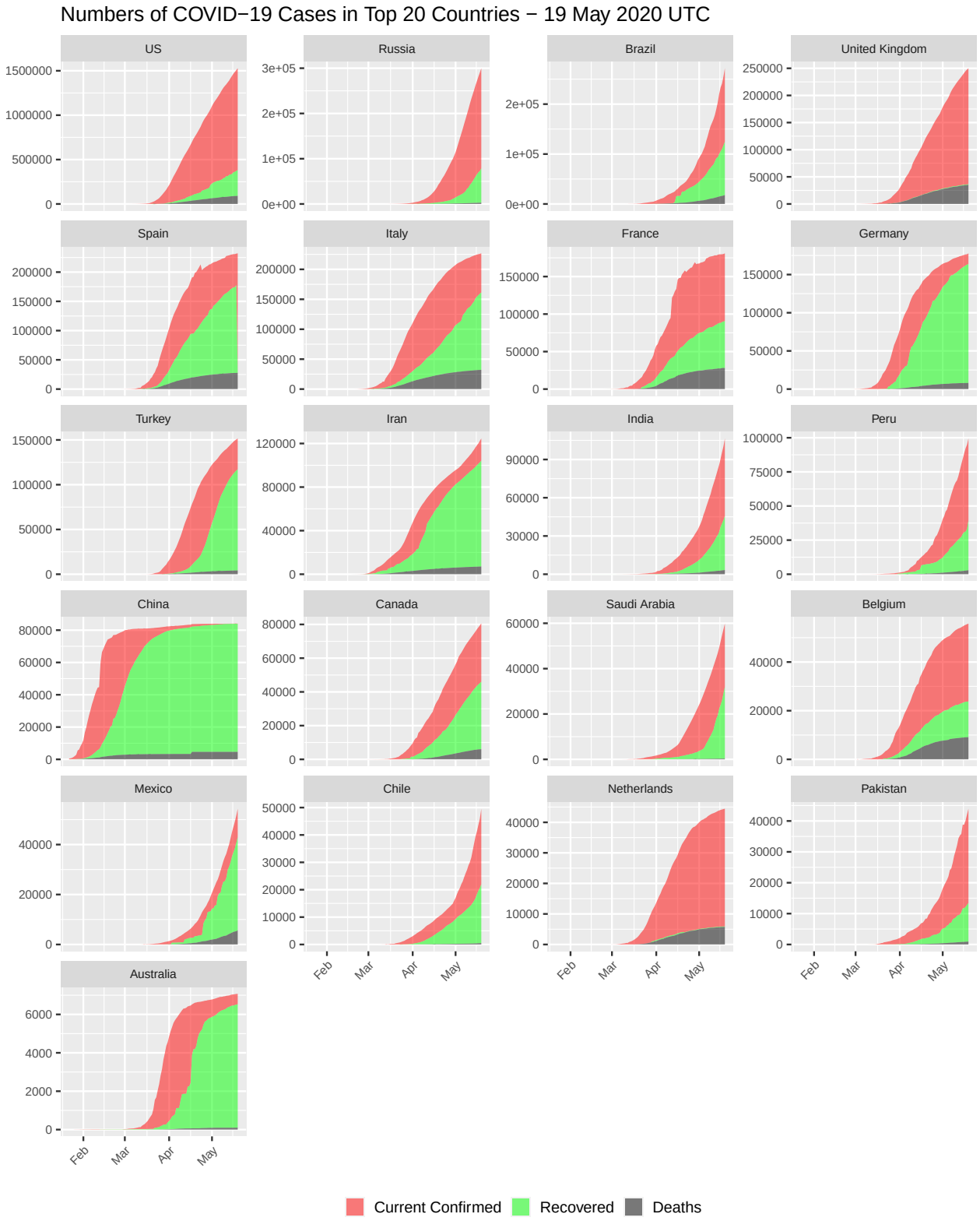


Figure 10: COVID-19 Cases in Top 20 Countries. Ordered descendingly by number of confirmed cases.

```r
## cases by country - line plot - log scale
p <- df %>% filter(country != 'World') %>%
  ggplot(aes(x=date, y=count, color=type)) +
  geom_line() +
  labs(title=paste0('Numbers of COVID-19 Cases in Top 20 Countries (log scale) - ',
                    max.date.txt)) +
  scale_color_manual(values=c('purple', 'red', 'green', 'black')) +
  theme(legend.title=element_blank(), legend.position='bottom',
        plot.title = element_text(size=12),
        axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        legend.key.size=unit(0.4, 'cm'),
        # legend.text=element_text(size=7),
        strip.text.x=element_text(size=7),
        axis.text=element_text(size=7),
        axis.text.x=element_text(angle=45, hjust=1)) +
  scale_y_continuous(trans='log10')
p + facet_wrap(~country, ncol=4, scales='free_y')
```

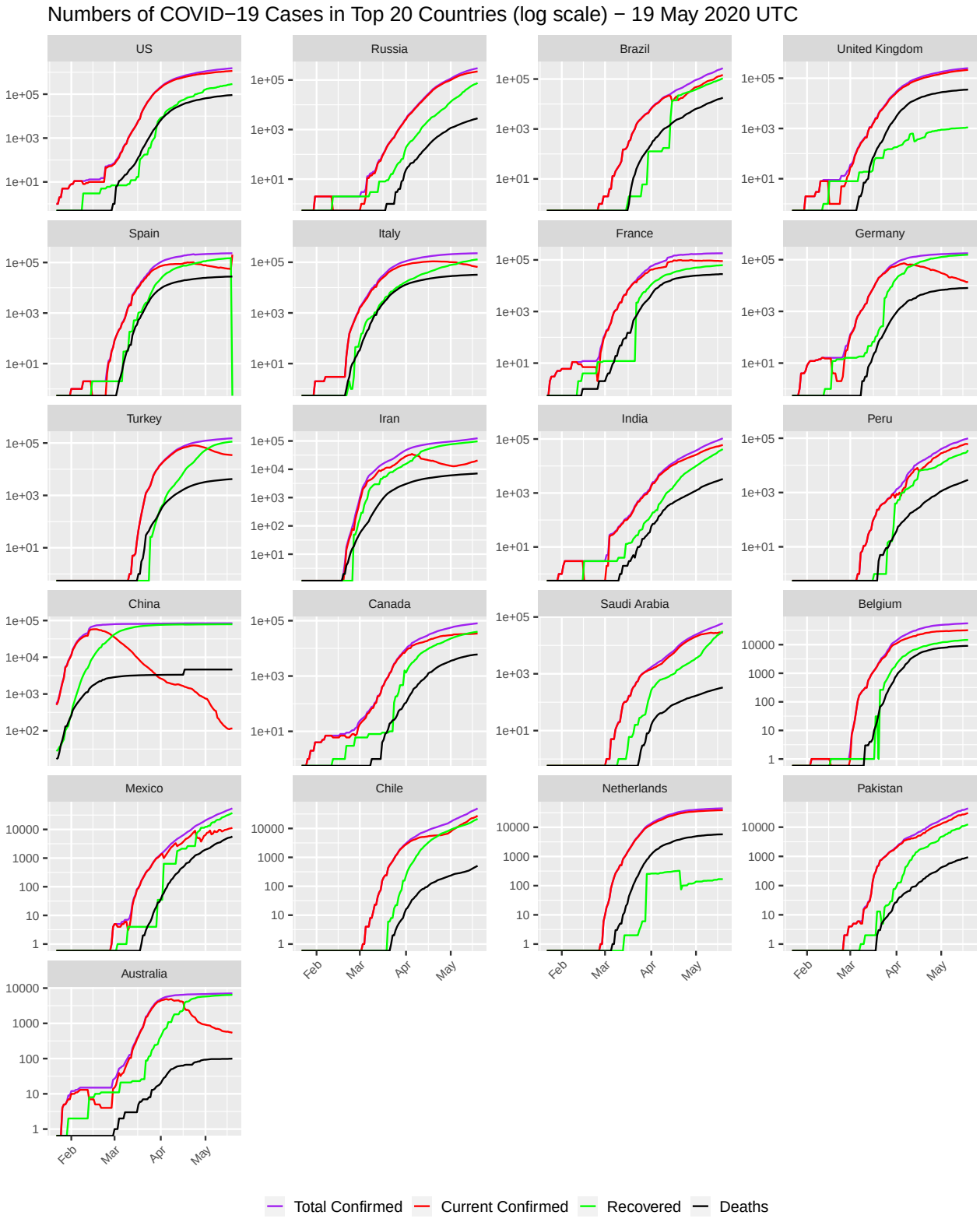


Figure 11: COVID-19 Cases Top 20 Countries (log scale). Ordered descendingly by number of confirmed cases.

```
## plot over multiple pages
# p + facet_wrap_paginate(~country, nrow=4, ncol=3, page=1, scales='free_y')
# p + facet_wrap_paginate(~country, nrow=4, ncol=3, page=2, scales='free_y')
```

Figures 10 and 11 show that the coronavirus seems to be under control in China, with an increase of recovered cases (in green) every day and a shrinking of the current confrimed cases (in red). However, in the rest of the world (i.e., outside of China), the confirmed cases are surging up in many other countries, which suggests that the virus has broken out there.

5.3 Death Rates

```
## three death rates
rate.max <- rates.long$count %>% max(na.rm=T)
df <- rates.long %>% filter(country %in% setdiff(top.countries, 'World')) %>%
  mutate(country=factor(country, levels=top.countries))
df %>% ggplot(aes(x=date, y=count, color=type)) +
  geom_line() +
  xlab('') + ylab('Death Rate (%)') +
  theme(legend.position='bottom', legend.title=element_blank(),
        legend.text=element_text(size=8),
        legend.key.size=unit(0.5, 'cm'),
        axis.text.x=element_text(angle=45, hjust=1)) +
  ylim(c(0, 99)) +
  facet_wrap(~country, ncol=4)
```

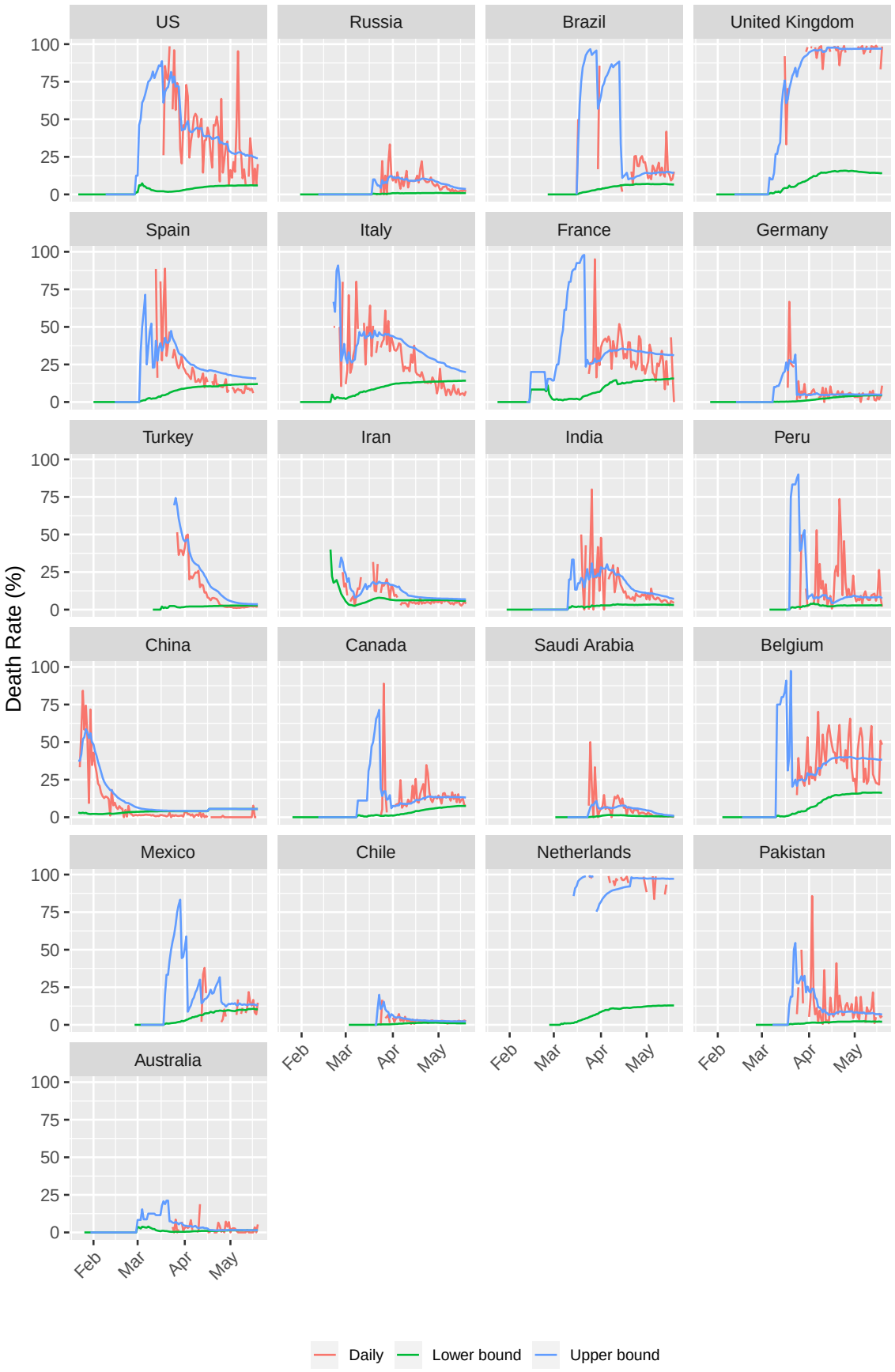


Figure 12: Death Rates

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 - 19 May 2020 UTC

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	Belgium	55,791	232	31,996	14,687	9,108	28	16.3%
2	France	180,933	882	90,230	62,678	28,025	0	15.5%
3	Italy	226,699	813	65,129	129,401	32,169	162	14.2%
4	United Kingdom	250,138	2,429	213,617	1,099	35,422	546	14.2%
5	Hungary	3,556	21	1,677	1,412	467	5	13.1%
6	Netherlands	44,449	108	38,548	167	5,734	21	12.9%
7	Sweden	30,799	422	22,085	4,971	3,743	45	12.2%
8	Spain	232,037	431	204,259	0	27,778	69	12.0%
9	Mexico	54,346	2,713	11,355	37,325	5,666	334	10.4%
10	Ecuador	34,151	569	27,855	3,457	2,839	40	8.3%
11	Algeria	7,377	176	3,070	3,746	561	6	7.6%
12	Canada	80,493	1,082	34,396	40,069	6,028	68	7.5%
13	Brazil	271,885	16,517	147,108	106,794	17,983	1,130	6.6%
14	Indonesia	18,496	486	12,808	4,467	1,221	30	6.6%
15	Romania	17,191	155	5,888	10,166	1,137	17	6.6%
16	Philippines	12,942	224	9,262	2,843	837	6	6.5%
17	Ireland	24,251	51	3,220	19,470	1,561	14	6.4%
18	Switzerland	30,618	21	1,027	27,700	1,891	5	6.2%
19	US	1,528,568	20,260	1,147,255	289,392	91,921	1,574	6.0%
20	Bosnia and Herzegovina	2,321	17	665	1,522	134	1	5.8%

6 Conclusions

As of 19 May 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 6.6% and 16.1%, 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
2020-05-19	4,897,492	323,285	1,688,619	2,885,588	95,549	4,804	0	6.6%	16.1%	100.0%
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	77,319	7,139	34,597	7.1%	21.0%	17.1%
2020-04-20	2,471,727	173,097	645,308	1,653,322	70,940	5,325	22,001	7.0%	21.2%	19.5%
2020-04-19	2,400,787	167,772	623,307	1,609,708	83,544	4,558	31,588	7.0%	21.2%	12.6%
2020-04-18	2,317,243	163,214	591,719	1,562,310	77,609	6,410	23,954	7.0%	21.6%	21.1%
2020-04-17	2,239,634	156,804	567,765	1,515,065	87,842	8,858	26,173	7.0%	21.6%	25.3%
2020-04-16	2,151,792	147,946	541,592	1,462,254	96,368	7,287	31,076	6.9%	21.5%	19.0%
2020-04-15	2,055,424	140,659	510,516	1,404,249	79,858	8,220	36,536	6.8%	21.6%	18.4%
2020-04-14	1,975,566	132,439	473,980	1,369,147	70,406	6,878	25,308	6.7%	21.8%	21.4%
2020-04-13	1,905,160	125,561	448,672	1,330,927	70,015	5,708	26,950	6.6%	21.9%	17.5%
2020-04-12	1,835,145	119,853	421,722	1,293,570	99,113	5,707	19,602	6.5%	22.1%	22.5%
2020-04-11	1,736,032	114,146	402,120	1,219,766	78,082	6,033	26,016	6.6%	22.1%	18.8%
2020-04-10	1,657,950	108,113	376,104	1,173,733	92,395	7,231	22,115	6.5%	22.3%	24.6%
2020-04-09	1,565,555	100,882	353,989	1,110,684	85,323	7,528	25,286	6.4%	22.2%	22.9%
2020-04-08	1,480,232	93,354	328,703	1,058,175	83,794	6,692	28,649	6.3%	22.1%	18.9%
2020-04-07	1,396,438	86,662	300,054	1,009,722	75,002	7,895	23,520	6.2%	22.4%	25.1%
2020-04-06	1,321,436	78,767	276,534	966,135	71,693	5,690	16,511	6.0%	22.2%	25.6%
2020-04-05	1,249,743	73,077	260,023	916,643	73,684	5,096	13,871	5.8%	21.9%	26.9%
2020-04-04	1,176,059	67,981	246,152	861,926	80,166	6,169	20,338	5.8%	21.6%	23.3%
2020-04-03	1,095,893	61,812	225,814	808,267	82,416	5,995	15,545	5.6%	21.5%	27.8%
2020-04-02	1,013,477	55,817	210,269	747,391	80,827	6,142	17,071	5.5%	21.0%	26.5%
2020-04-01	932,650	49,675	193,198	689,777	75,042	5,437	15,164	5.3%	20.5%	26.4%
2020-03-31	857,608	44,238	178,034	635,336	75,108	4,799	13,468	5.2%	19.9%	26.3%
2020-03-30	782,500	39,439	164,566	578,495	62,210	4,116	15,484	5.0%	19.3%	21.0%
2020-03-29	720,290	35,323	149,082	535,885	59,437	3,525	9,667	4.9%	19.2%	26.7%
2020-03-28	660,853	31,798	139,415	489,640	67,394	3,510	8,500	4.8%	18.6%	29.2%
2020-03-27	593,459	28,288	130,915	434,256	63,737	3,500	8,765	4.8%	17.8%	28.5%
2020-03-26	529,722	24,788	122,150	382,784	61,971	3,003	8,363	4.7%	16.9%	26.4%
2020-03-25	467,751	21,785	113,787	332,179	49,645	2,771	5,787	4.7%	16.1%	32.4%
2020-03-24	418,106	19,014	108,000	291,092	39,824	2,255	9,649	4.5%	15.0%	18.9%
2020-03-23	378,282	16,759	98,351	263,172	41,264	1,934	452	4.4%	14.6%	81.1%
2020-03-22	337,018	14,825	97,899	224,294	32,438	1,700	6,207	4.4%	13.2%	21.5%
2020-03-21	304,580	13,125	91,692	199,763	32,317	1,703	4,272	4.3%	12.5%	28.5%
2020-03-20	272,263	11,422	87,420	173,421	29,632	1,476	2,445	4.2%	11.6%	37.6%
2020-03-19	242,631	9,946	84,975	147,710	27,756	1,123	1,663	4.1%	10.5%	40.3%
2020-03-18	214,875	8,823	83,312	122,740	17,741	867	2,472	4.1%	9.6%	26.0%
2020-03-17	197,134	7,956	80,840	108,338	15,510	806	2,752	4.0%	9.0%	22.7%
2020-03-16	181,624	7,150	78,088	96,386	14,154	678	2,054	3.9%	8.4%	24.8%
2020-03-15	167,470	6,472	76,034	84,964	11,352	642	3,410	3.9%	7.8%	15.8%

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
2020-03-14	156,118	5,830	72,624	77,664	10,895	422	2,373	3.7%	7.4%	15.1%
2020-03-13	145,223	5,408	70,251	69,564	16,871	686	1,927	3.7%	7.1%	26.3%
2020-03-12	128,352	4,722	68,324	55,306	2,477	108	1,321	3.7%	6.5%	7.6%
2020-03-11	125,875	4,614	67,003	54,258	7,255	351	2,599	3.7%	6.4%	11.9%
2020-03-10	118,620	4,263	64,404	49,953	5,030	276	1,910	3.6%	6.2%	12.6%
2020-03-09	113,590	3,987	62,494	47,109	3,769	186	1,800	3.5%	6.0%	9.4%
2020-03-08	109,821	3,801	60,694	45,326	3,974	243	2,336	3.5%	5.9%	9.4%
2020-03-07	105,847	3,558	58,358	43,931	4,046	99	2,493	3.4%	5.7%	3.8%
2020-03-06	101,801	3,459	55,865	42,477	3,915	112	2,069	3.4%	5.8%	5.1%
2020-03-05	97,886	3,347	53,796	40,743	2,766	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 (19 May 2020 UTC)

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	World	4,897,492	95,549	2,885,588	1,688,619	323,285	4,804	**6.6%**
2	US	1,528,568	20,260	1,147,255	289,392	91,921	1,574	**6.0%**

Table 6: Cases by Country (19 May 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
3	Russia	299,941	9,263	220,974	76,130	2,837	115	0.9%
4	Brazil	271,885	16,517	147,108	106,794	17,983	1,130	**6.6%**
5	United Kingdom	250,138	2,429	213,617	1,099	35,422	546	**14.2%**
6	Spain	232,037	431	204,259	0	27,778	69	**12.0%**
7	Italy	226,699	813	65,129	129,401	32,169	162	**14.2%**
8	France	180,933	882	90,230	62,678	28,025	0	**15.5%**
9	Germany	177,778	1,227	14,016	155,681	8,081	78	4.5%
10	Turkey	151,615	1,022	34,521	112,895	4,199	28	2.8%
11	Iran	124,603	2,111	20,311	97,173	7,119	62	**5.7%**
12	India	106,475	6,147	60,864	42,309	3,302	146	3.1%
13	Peru	99,483	4,550	60,045	36,524	2,914	125	2.9%
14	China	84,063	0	115	79,310	4,638	0	**5.5%**
15	Canada	80,493	1,082	34,396	40,069	6,028	68	**7.5%**
16	Saudi Arabia	59,854	2,509	27,891	31,634	329	9	0.5%
17	Belgium	55,791	232	31,996	14,687	9,108	28	**16.3%**
18	Mexico	54,346	2,713	11,355	37,325	5,666	334	**10.4%**
19	Chile	49,579	3,520	27,563	21,507	509	31	1.0%
20	Netherlands	44,449	108	38,548	167	5,734	21	**12.9%**
21	Pakistan	43,966	1,841	30,538	12,489	939	36	2.1%
22	Qatar	35,606	1,637	29,957	5,634	15	0	0.0%
23	Ecuador	34,151	569	27,855	3,457	2,839	40	**8.3%**
24	Belarus	31,508	936	20,713	10,620	175	4	0.6%
25	Sweden	30,799	422	22,085	4,971	3,743	45	**12.2%**
26	Switzerland	30,618	21	1,027	27,700	1,891	5	**6.2%**
27	Portugal	29,432	223	21,754	6,431	1,247	16	4.2%
28	Singapore	28,794	451	18,407	10,365	22	0	0.1%
29	Bangladesh	25,121	1,251	19,758	4,993	370	21	1.5%
30	United Arab Emirates	25,063	873	14,045	10,791	227	3	0.9%
31	Ireland	24,251	51	3,220	19,470	1,561	14	**6.4%**
32	Poland	19,268	383	10,417	7,903	948	12	4.9%
33	Ukraine	18,876	260	12,696	5,632	548	13	2.9%
34	Indonesia	18,496	486	12,808	4,467	1,221	30	**6.6%**
35	South Africa	17,200	767	8,928	7,960	312	26	1.8%
36	Romania	17,191	155	5,888	10,166	1,137	17	**6.6%**
37	Colombia	16,935	640	12,272	4,050	613	21	3.6%
38	Kuwait	16,764	1,073	11,962	4,681	121	3	0.7%
39	Israel	16,659	16	2,946	13,435	278	2	1.7%
40	Japan	16,367	62	4,035	11,564	768	19	4.7%
41	Austria	16,321	52	1,011	14,678	632	3	3.9%
42	Egypt	13,484	720	9,083	3,742	659	14	4.9%
43	Dominican Republic	13,223	498	6,169	6,613	441	7	3.3%
44	Philippines	12,942	224	9,262	2,843	837	6	**6.5%**
45	Denmark	11,242	76	1,077	9,614	551	3	4.9%
46	Korea, South	11,110	32	781	10,066	263	0	2.4%
47	Serbia	10,733	34	5,595	4,904	234	3	2.2%
48	Panama	9,867	141	3,392	6,194	281	2	2.8%
49	Argentina	8,809	438	5,544	2,872	393	11	4.5%
50	Czechia	8,647	61	2,619	5,726	302	5	3.5%
51	Norway	8,267	10	8,002	32	233	0	2.8%
52	Afghanistan	7,653	581	6,625	850	178	5	2.3%
53	Bahrain	7,532	348	4,568	2,952	12	0	0.2%
54	Algeria	7,377	176	3,070	3,746	561	6	**7.6%**
55	Australia	7,072	4	541	6,431	100	1	1.4%
56	Morocco	7,023	71	2,929	3,901	193	1	2.7%
57	Malaysia	6,978	37	1,218	5,646	114	1	1.6%
58	Kazakhstan	6,751	0	3,118	3,598	35	0	0.5%
59	Nigeria	6,401	226	4,475	1,734	192	1	3.0%
60	Finland	6,399	19	1,098	5,000	301	1	4.7%
61	Moldova	6,340	202	3,611	2,508	221	4	3.5%
62	Ghana	6,096	361	4,292	1,773	31	2	0.5%
63	Oman	5,671	292	4,070	1,574	27	2	0.5%
64	Armenia	5,041	218	2,813	2,164	64	3	1.3%
65	Bolivia	4,481	218	3,759	533	189	15	4.2%
66	Luxembourg	3,958	11	131	3,718	109	2	2.8%
67	Iraq	3,611	57	1,114	2,366	131	4	3.6%
68	Hungary	3,556	21	1,677	1,412	467	5	**13.1%**
69	Cameroon	3,529	0	1,822	1,567	140	0	4.0%
70	Azerbaijan	3,518	131	1,279	2,198	41	1	1.2%

Table 6: Cases by Country (19 May 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
71	Thailand	3,033	2	120	2,857	56	0	1.8%
72	Honduras	2,955	157	2,459	349	147	1	**5.0%**
73	Guinea	2,863	67	1,320	1,525	18	2	0.6%
74	Uzbekistan	2,855	64	504	2,338	13	0	0.5%
75	Greece	2,840	4	1,301	1,374	165	0	**5.8%**
76	Sudan	2,728	137	2,331	286	111	6	4.1%
77	Senegal	2,617	73	1,454	1,133	30	4	1.1%
78	Bosnia and Herzegovina	2,321	17	665	1,522	134	1	**5.8%**
79	Bulgaria	2,259	24	1,501	646	112	2	**5.0%**
80	Croatia	2,232	4	169	1,967	96	1	4.3%
81	Cote d'Ivoire	2,153	34	1,075	1,050	28	0	1.3%
82	Guatemala	2,133	221	1,935	155	43	8	2.0%
83	Tajikistan	1,936	207	1,254	641	41	0	2.1%
84	Cuba	1,887	6	270	1,538	79	0	4.2%
85	North Macedonia	1,839	22	382	1,351	106	2	5.8%
86	Iceland	1,802	0	3	1,789	10	0	0.6%
87	Estonia	1,791	7	789	938	64	0	3.6%
88	Congo (Kinshasa)	1,629	91	1,278	290	61	0	3.7%
89	Djibouti	1,618	100	578	1,033	7	0	0.4%
90	Lithuania	1,562	15	477	1,025	60	1	3.8%
91	New Zealand	1,503	4	35	1,447	21	0	1.4%
92	Gabon	1,502	70	1,172	318	12	1	0.8%
93	Somalia	1,502	47	1,265	178	59	2	3.9%
94	El Salvador	1,498	85	966	502	30	0	2.0%
95	Slovakia	1,495	0	275	1,192	28	0	1.9%
96	Slovenia	1,467	1	25	1,338	104	0	7.1%
97	Kyrgyzstan	1,243	27	331	898	14	0	1.1%
98	Maldives	1,143	37	1,048	91	4	0	0.3%
99	Tunisia	1,044	1	171	826	47	1	4.5%
100	Guinea-Bissau	1,038	6	990	42	6	2	0.6%
101	Sri Lanka	1,027	35	449	569	9	0	0.9%
102	Latvia	1,012	3	297	694	21	2	2.1%
103	Kosovo	989	34	191	769	29	0	2.9%
104	Kenya	963	51	555	358	50	0	5.2%
105	Lebanon	954	23	677	251	26	0	2.7%
106	Albania	949	1	176	742	31	0	3.3%
107	Cyprus	918	1	386	515	17	0	1.9%
108	Niger	914	5	125	734	55	0	6.0%
109	Mali	901	27	319	529	53	1	5.9%
110	Costa Rica	882	16	295	577	10	0	1.1%
111	Paraguay	829	41	588	230	11	0	1.3%
112	Equatorial Guinea	825	106	796	22	7	0	0.8%
113	Burkina Faso	796	0	93	652	51	0	6.4%
114	Zambia	772	11	573	192	7	0	0.9%
115	Andorra	761	0	82	628	51	0	6.7%
116	Venezuela	749	131	486	253	10	0	1.3%
117	Uruguay	738	1	139	579	20	0	2.7%
118	Diamond Princess	712	0	48	651	13	0	1.8%
119	Georgia	707	6	239	456	12	0	1.7%
120	San Marino	655	1	403	211	41	0	6.3%
121	Jordan	649	20	223	417	9	0	1.4%
122	Malta	569	11	103	460	6	0	1.1%
123	Chad	545	26	350	139	56	3	10.3%
124	Sierra Leone	534	15	334	167	33	0	6.2%
125	Haiti	533	0	491	21	21	0	3.9%
126	Jamaica	520	0	366	145	9	0	1.7%
127	Tanzania	509	0	305	183	21	0	4.1%
128	Taiwan*	440	0	35	398	7	0	1.6%
129	Congo (Brazzaville)	420	8	273	132	15	0	3.6%
130	Nepal	402	27	363	37	2	0	0.5%
131	West Bank and Gaza	391	3	43	346	2	0	0.5%
132	Central African Republic	366	39	348	18	0	0	0.0%
133	Ethiopia	365	13	240	120	5	0	1.4%
134	Togo	338	8	219	107	12	0	3.6%
135	Cabo Verde	335	7	247	85	3	0	0.9%
136	Mauritius	332	0	0	322	10	0	3.0%
137	Madagascar	326	4	205	119	2	1	0.6%
138	Montenegro	324	0	3	312	9	0	2.8%

Table 6: Cases by Country (19 May 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
139	Vietnam	324	0	61	263	0	0	0.0%
140	Rwanda	308	11	99	209	0	0	0.0%
141	South Sudan	290	0	282	4	4	0	1.4%
142	Uganda	260	12	197	63	0	0	0.0%
143	Nicaragua	254	229	38	199	17	9	6.7%
144	Sao Tome and Principe	251	5	239	4	8	1	3.2%
145	Liberia	233	4	85	125	23	1	9.9%
146	Eswatini	208	3	119	87	2	0	1.0%
147	Burma	193	2	83	104	6	0	3.1%
148	Yemen	167	37	134	5	28	8	16.8%
149	Mozambique	146	1	102	44	0	0	0.0%
150	Brunei	141	0	4	136	1	0	0.7%
151	Mongolia	140	0	114	26	0	0	0.0%
152	Mauritania	131	50	120	7	4	0	3.1%
153	Benin	130	0	45	83	2	0	1.5%
154	Guyana	125	1	69	46	10	0	8.0%
155	Cambodia	122	0	0	122	0	0	0.0%
156	Trinidad and Tobago	116	0	1	107	8	0	6.9%
157	Monaco	97	0	6	87	4	0	4.1%
158	Bahamas	96	0	42	43	11	0	11.5%
159	Barbados	90	2	15	68	7	0	7.8%
160	Liechtenstein	82	0	26	55	1	0	1.2%
161	Malawi	70	0	40	27	3	0	4.3%
162	Libya	68	3	30	35	3	0	4.4%
163	Syria	58	0	19	36	3	0	5.2%
164	Angola	52	2	32	17	3	0	5.8%
165	Zimbabwe	46	0	24	18	4	0	8.7%
166	Burundi	42	0	21	20	1	0	2.4%
167	Eritrea	39	0	0	39	0	0	0.0%
168	Antigua and Barbuda	25	0	3	19	3	0	12.0%
169	Botswana	25	0	7	17	1	0	4.0%
170	Gambia	24	0	10	13	1	0	4.2%
171	Timor-Leste	24	0	0	24	0	0	0.0%
172	Grenada	22	0	8	14	0	0	0.0%
173	Bhutan	21	0	16	5	0	0	0.0%
174	Laos	19	0	5	14	0	0	0.0%
175	Belize	18	0	0	16	2	0	11.1%
176	Fiji	18	0	3	15	0	0	0.0%
177	Saint Lucia	18	0	0	18	0	0	0.0%
178	Saint Vincent and the Grenadines	17	0	3	14	0	0	0.0%
179	Dominica	16	0	0	16	0	0	0.0%
180	Namibia	16	0	3	13	0	0	0.0%
181	Saint Kitts and Nevis	15	0	0	15	0	0	0.0%
182	Holy See	12	0	10	2	0	0	0.0%
183	Comoros	11	0	7	3	1	0	9.1%
184	Seychelles	11	0	0	11	0	0	0.0%
185	Suriname	11	0	1	9	1	0	9.1%
186	MS Zaandam	9	0	7	0	2	0	22.2%
187	Papua New Guinea	8	0	0	8	0	0	0.0%
188	Western Sahara	6	0	0	6	0	0	0.0%
189	Lesotho	1	0	1	0	0	0	0.0%

Appendix B. How to Cite This Work

Citation

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

BibTex

@techreport{Zhao2020Covid19world,
Author = {Yanchang Zhao},
Institution = {RDataMining.com},
Title = {COVID-19 Data Analysis with R – Worldwide},

Url = {http://www.rdatamining.com/docs/Coronavirus-data-analysis-world.pdf},
Year = {2020}}

Appendix C. Contact

Contact:

Dr. Yanchang Zhao

Email: yanchang@RDataMining.com

Twitter: @RDataMining

LinkedIn: http://group.rdatamining.com

Comments and suggestions and welcome. Thanks!