COVID-19 Data Analysis with R - Worldwide*

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

08 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] 266 111
```

Each dataset has 266 rows, corresponding to country/region/province/state. It has 111 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" "2020-05-07"
```

```
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 07 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 187 countries with confirmed COVID-19 cases, as of 07 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
```

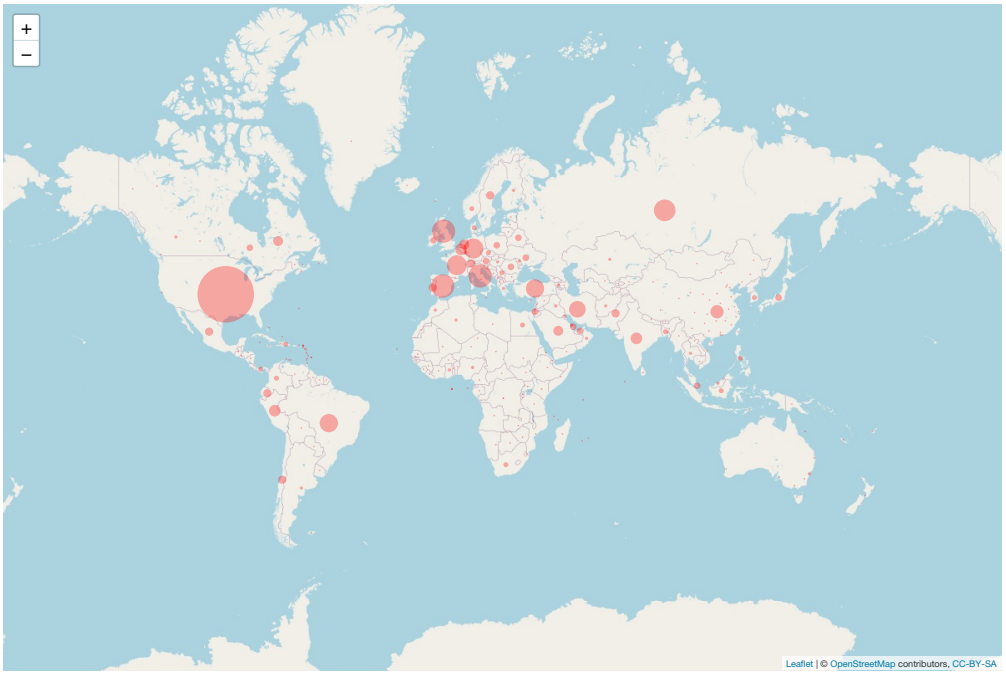


Figure 1: World Map

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)
```

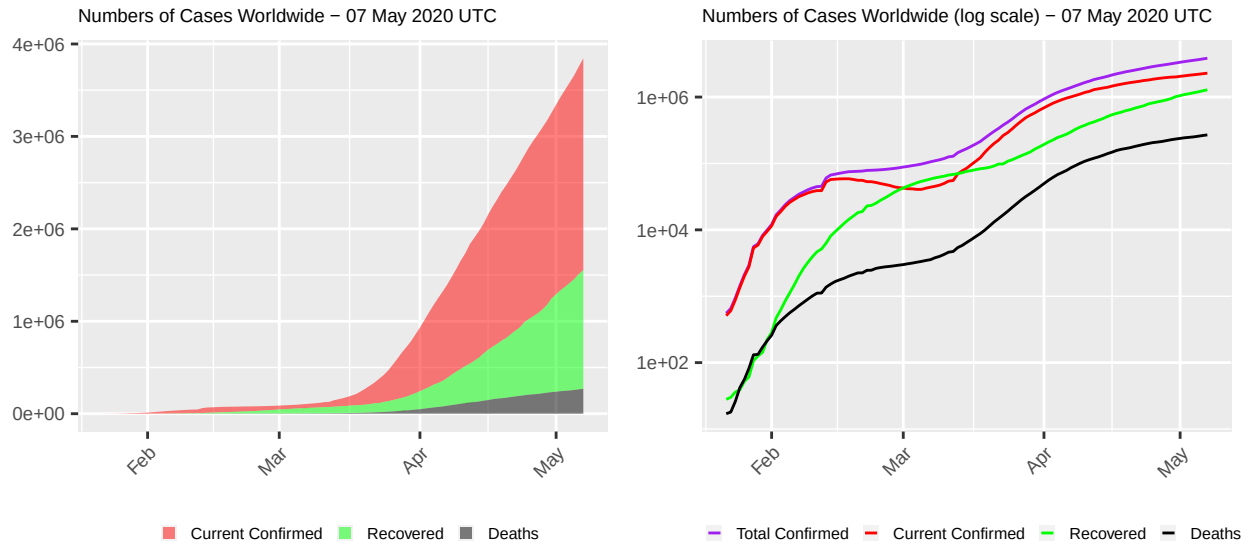


Figure 2: COVID-19 Cases Worldwide

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)
```

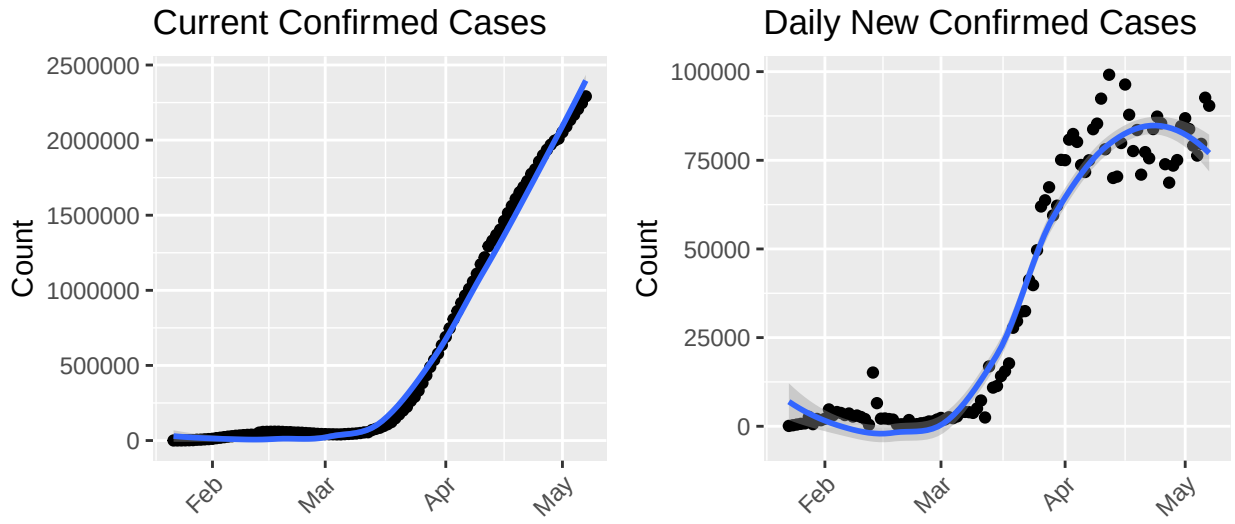


Figure 3: Current Confirmed Cases

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)
```

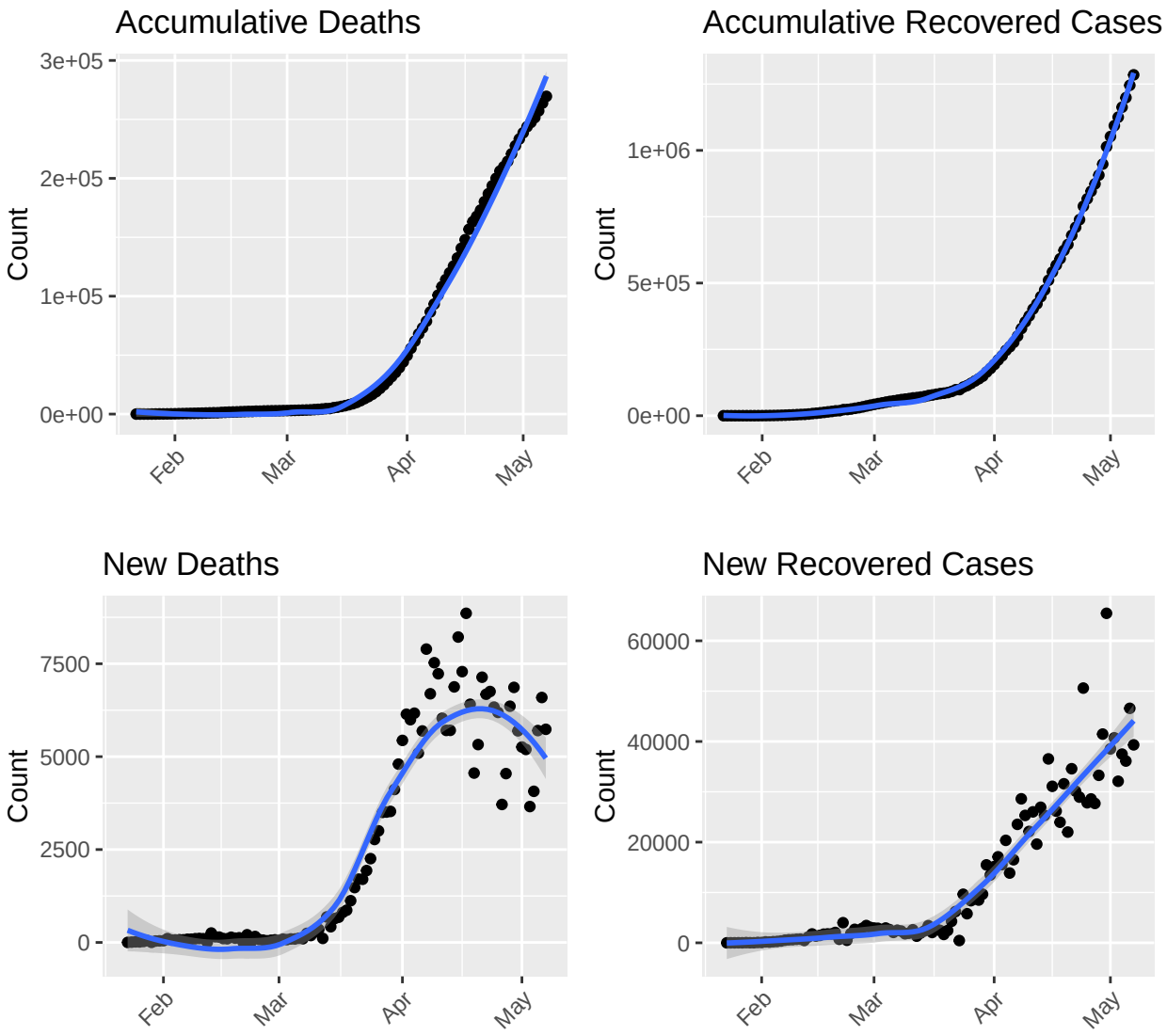


Figure 4: Deaths and Recovered Cases

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 07 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 07 May 2020 UTC UTC, it will be between 7% and 17.3%.

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

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

```
  geom_line(aes(y=rate.upper, colour='Upper bound')) +
  geom_line(aes(y=rate.lower, colour='Lower bound')) +
  geom_line(aes(y=rate.daily, colour='Daily')) +
  xlab('') + ylab('Death Rate (%)') + labs(title='Overall') +
  theme(legend.position='bottom', legend.title=element_blank(),
        legend.text=element_text(size=8),
        legend.key.size=unit(0.5, 'cm'),
        axis.text.x=element_text(angle=45, hjust=1))
## focusing on last 2 weeks
y.max <- data.world[n-(14:0), ] %>% select(rate.upper, rate.lower, rate.daily) %>% max()
plot2 <- ggplot(data.world[n-(14:0),], aes(x=date)) +
  geom_line(aes(y=rate.upper, colour='Upper bound')) +
  geom_line(aes(y=rate.lower, colour='Lower bound')) +
  geom_line(aes(y=rate.daily, colour='Daily')) +
  xlab('') + ylab('Death Rate (%)') + labs(title='Last two weeks') +
  theme(legend.position='bottom', legend.title=element_blank(),
        legend.text=element_text(size=8),
        legend.key.size=unit(0.5, 'cm'),
        axis.text.x=element_text(angle=45, hjust=1)) +
  ylim(c(0, y.max))
grid.arrange(plot1, plot2, ncol=2)
```

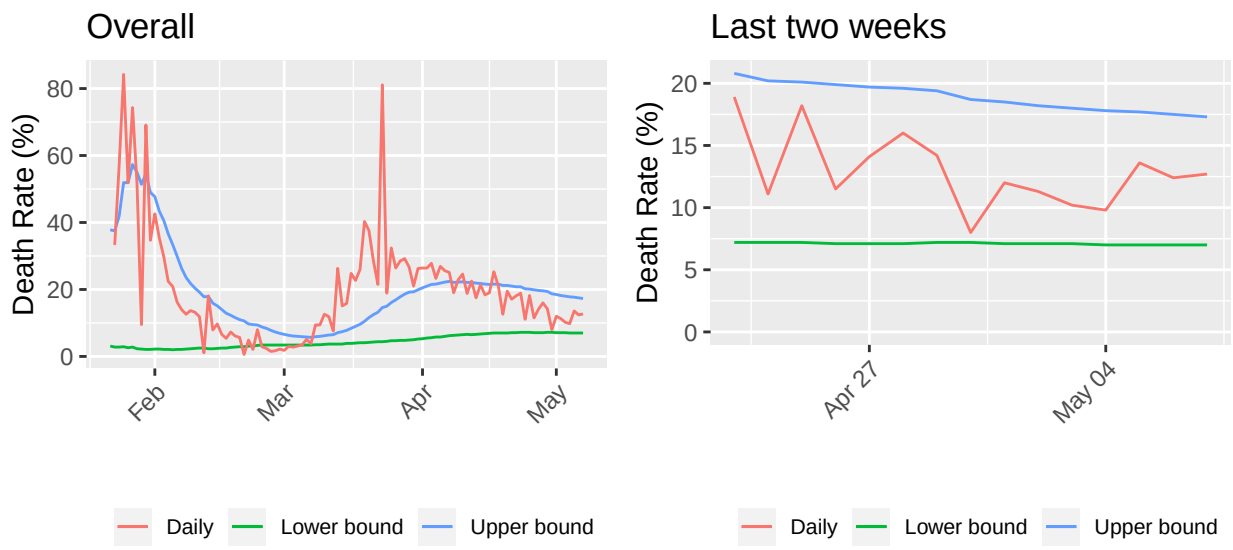


Figure 5: Death Rate

5 Top Twenty Countries

Next, we will have a look at the top 20 countries in total confirmed cases.

```
## ranking by confirmed cases
data.latest.all <- data %>% filter(date == max(date)) %>%
  select(country, date,
         confirmed, new.confirmed, current.confirmed,
         recovered, deaths, new.deaths, death.rate=rate.lower) %>%
  mutate(ranking = dense_rank(desc(confirmed)))
```

```
k <- 20
## top 20 countries: 21 incl. 'World'
top.countries <- data.latest.all %>% filter(ranking <= k + 1) %>%
  arrange(ranking) %>% pull(country) %>% as.character()
top.countries %>% setdiff('World') %>% print()
```

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

```
## add 'Others'
top.countries %<>% c('Others')
## put all others in a single group of 'Others'
data.latest <- data.latest.all %>% filter(!is.na(country)) %>%
  mutate(country=ifelse(ranking <= k + 1, as.character(country), 'Others')) %>%
  mutate(country=country %>% factor(levels=c(top.countries)))
data.latest %<>% group_by(country) %>%
  summarise(confirmed=sum(confirmed), new.confirmed=sum(new.confirmed),
            current.confirmed=sum(current.confirmed),
            recovered=sum(recovered), deaths=sum(deaths), new.deaths=sum(new.deaths)) %>%
  mutate(death.rate=(100 * deaths/confirmed) %>% round(1))
data.latest %<>% select(c(country, confirmed, deaths, death.rate,
                 new.confirmed, new.deaths, current.confirmed))
```

```
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 - 07 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	3,845,718	269,567	7.0%	90,377	5,736	2,291,410
2	US	1,257,023	75,662	6.0%	28,420	2,231	986,325
3	Spain	221,447	26,070	11.8%	1,122	213	66,866
4	Italy	215,858	29,958	13.9%	1,401	274	89,624
5	United Kingdom	207,977	30,689	14.8%	5,618	539	176,318
6	Russia	177,160	1,625	0.9%	11,231	88	151,732
7	France	174,918	25,990	14.9%	694	178	93,737
8	Germany	169,430	7,392	4.4%	1,268	117	20,338
9	Brazil	135,773	9,190	6.8%	9,162	602	71,233
10	Turkey	133,721	3,641	2.7%	1,977	57	47,096
11	Iran	103,135	6,486	6.3%	1,485	68	13,905
12	China	83,975	4,637	5.5%	5	0	361
13	Canada	66,201	4,541	6.9%	1,507	175	32,400
14	Peru	58,526	1,627	2.8%	3,709	94	38,511
15	India	56,351	1,889	3.4%	3,364	104	37,686
16	Belgium	51,420	8,415	16.4%	639	76	30,025
17	Netherlands	41,973	5,306	12.6%	455	85	36,520
18	Saudi Arabia	33,731	219	0.6%	1,793	10	25,714
19	Ecuador	30,298	1,654	5.5%	0	36	25,211
20	Switzerland	30,126	1,810	6.0%	66	5	2,416
21	Mexico	29,616	2,961	10.0%	1,982	257	8,874
22	Others	567,059	19,805	3.5%	16,062	527	336,518

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

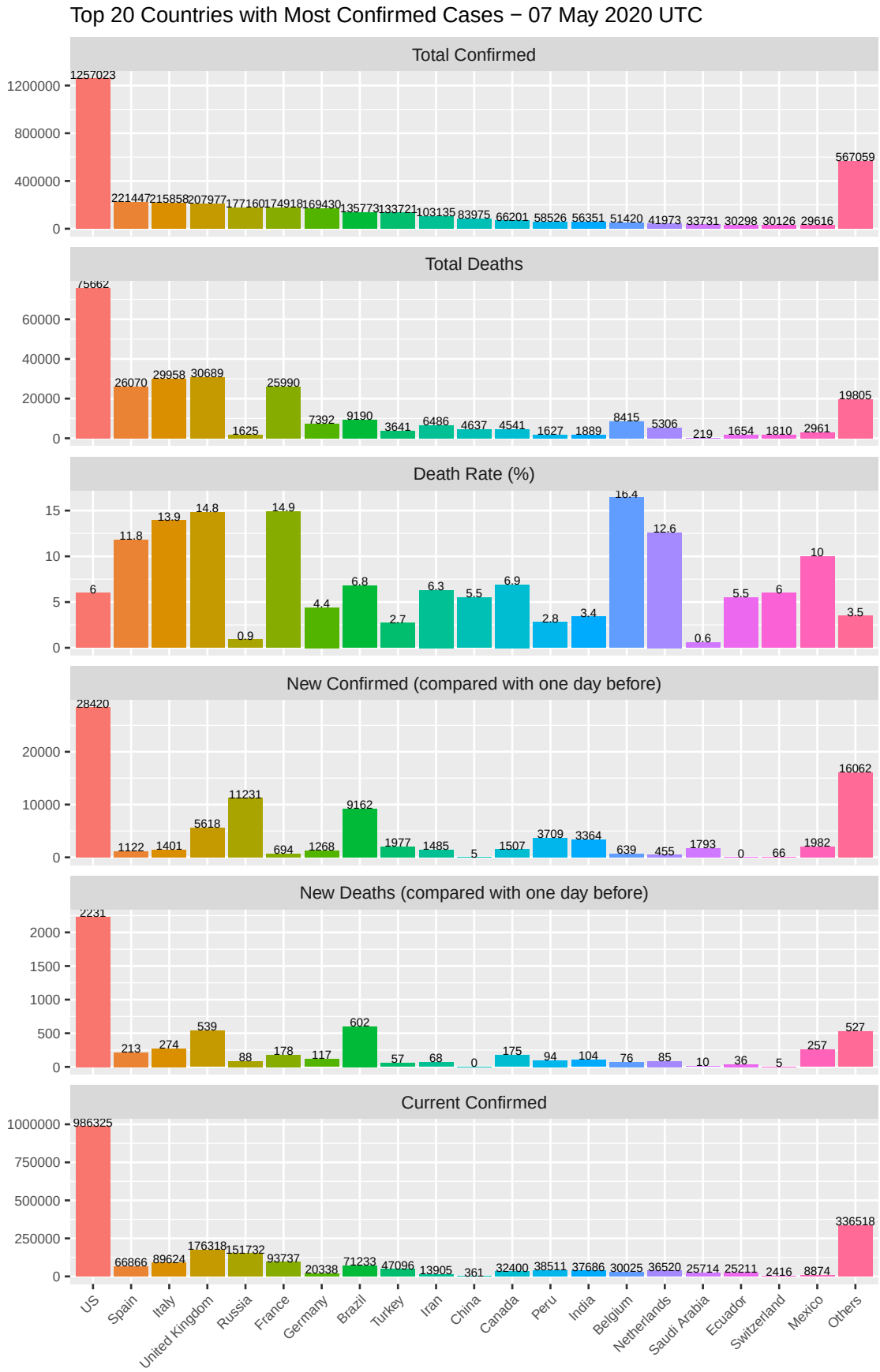
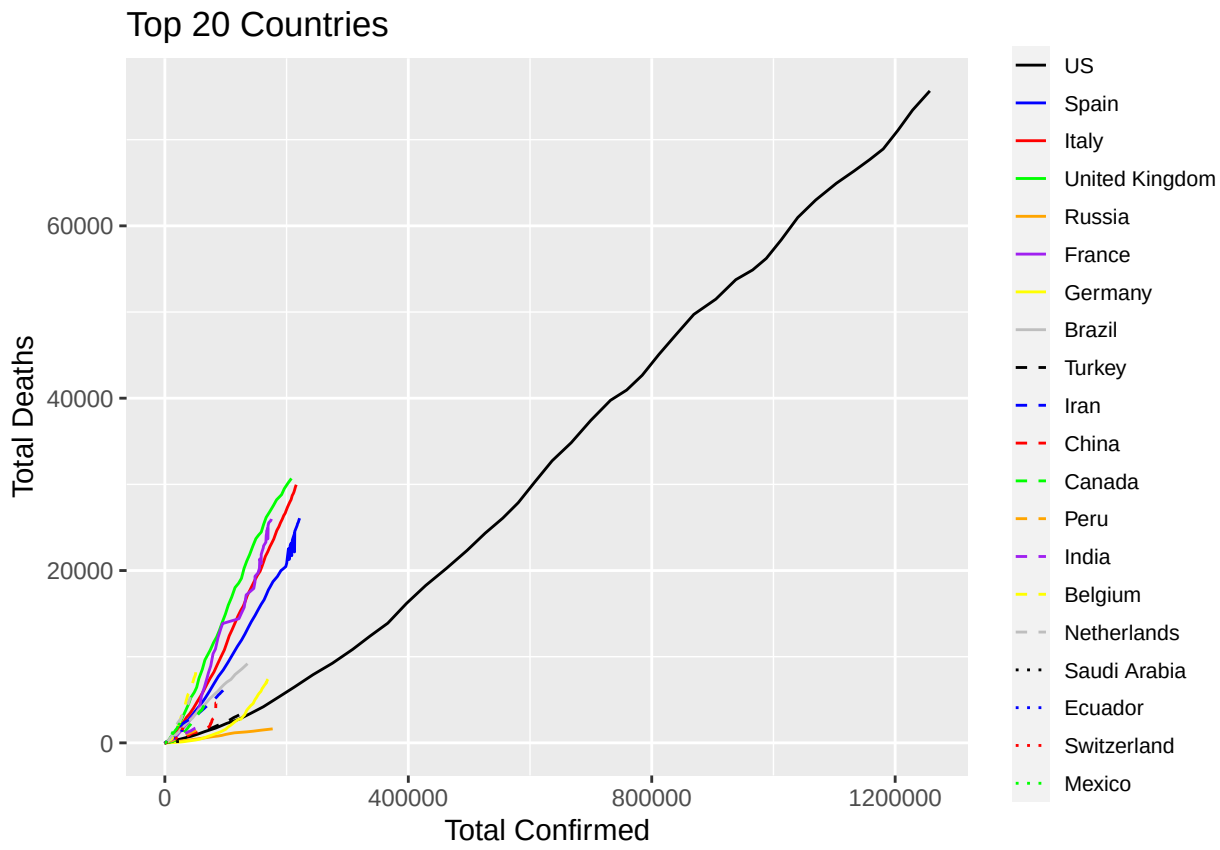


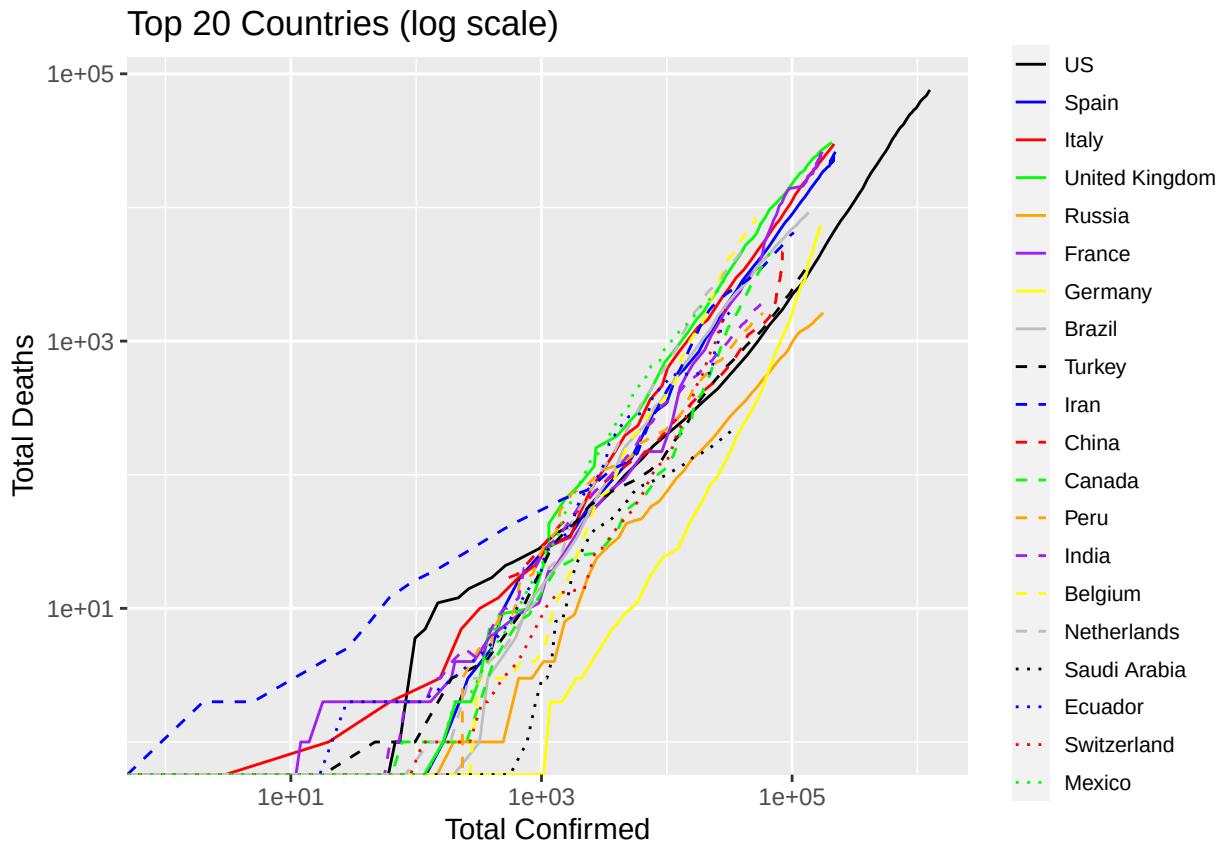
Figure 6: Top 20 Countries with Most Confirmed Cases

5.1 Confirmed vs Deaths

```
# linetypes <- rep(c("dotted", "dashed", "solid"), each=8)
# colors <- rep(c('grey', 'yellow', 'purple', 'orange', 'green', 'red', 'blue', 'black'), 3)
linetypes <- rep(c("solid", "dashed", "dotted"), each=8)
colors <- rep(c('black', 'blue', 'red', 'green', 'orange', 'purple', 'yellow', 'grey'), 3)
df <- data %>% filter(country %in% setdiff(top.countries, c('World', 'Others'))) %>%
  mutate(country=country %>% factor(levels=c(top.countries)))
p <- df %>% ggplot(aes(x=confirmed, y=deaths, group=country)) +
  geom_line(aes(color=country, linetype=country)) +
  xlab('Total Confirmed') + ylab('Total Deaths') +
  scale_linetype_manual(values=linetypes) +
  scale_color_manual(values=colors) +
  theme(legend.title=element_blank(),
        legend.text=element_text(size=8),
        legend.key.size=unit(0.5, 'cm'))
p + labs(title=paste0('Top 20 Countries'))
```




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



The two figures below show the numbers of confirmed cases and deaths of top 20 countries, as well as the death rates up to 07 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)
```

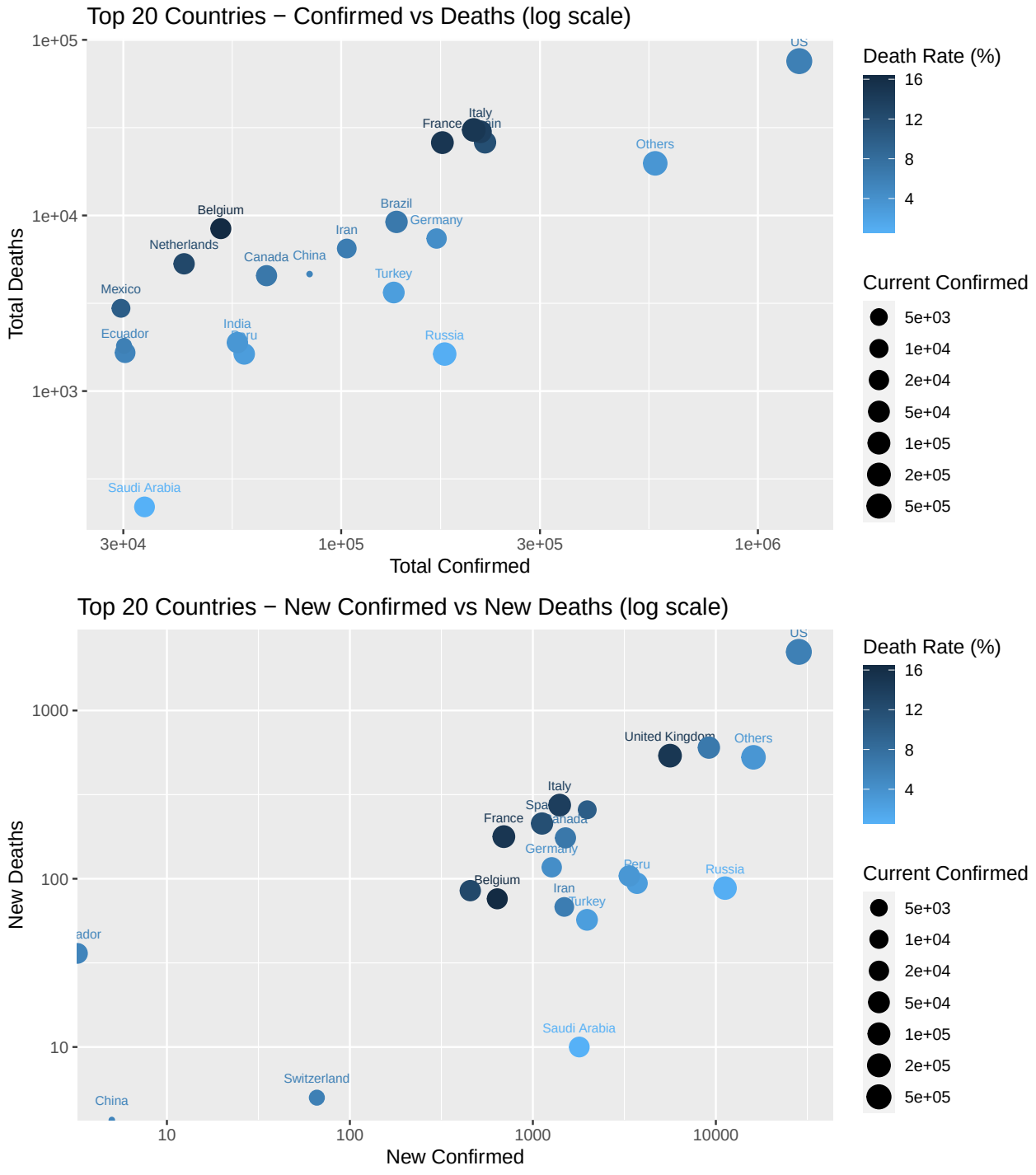


Figure 7: Top 20 Countries

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)
```

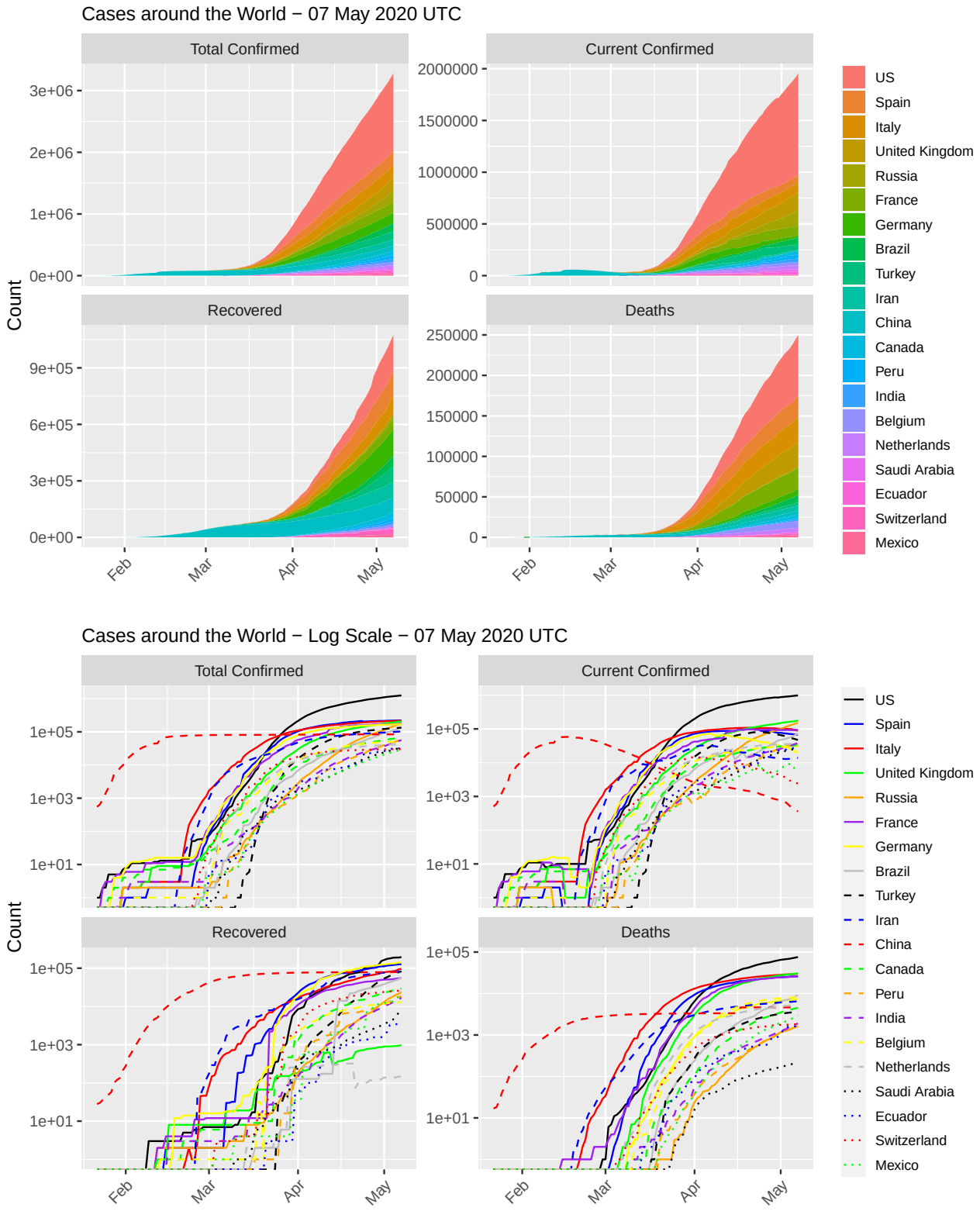


Figure 8: Cases around the World

```
## 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))
```

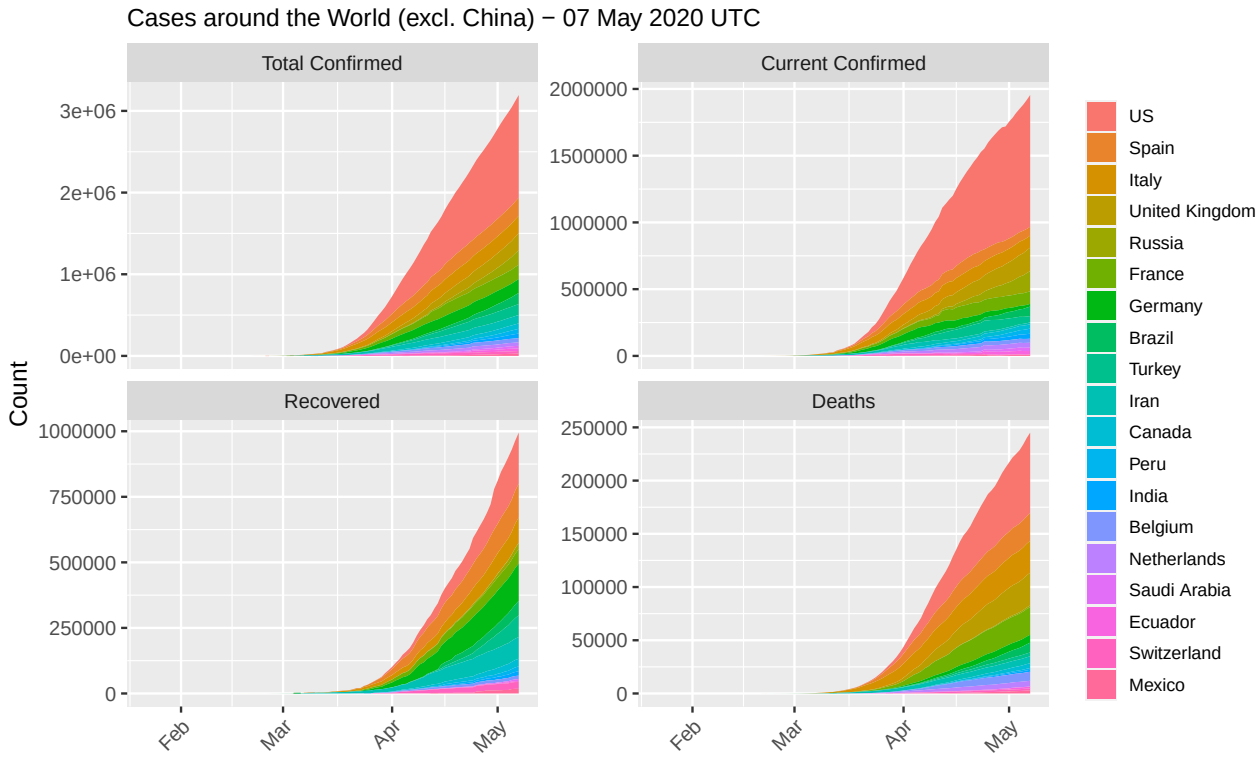


Figure 9: Cases around the World (excl. China)

```
## 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')
```

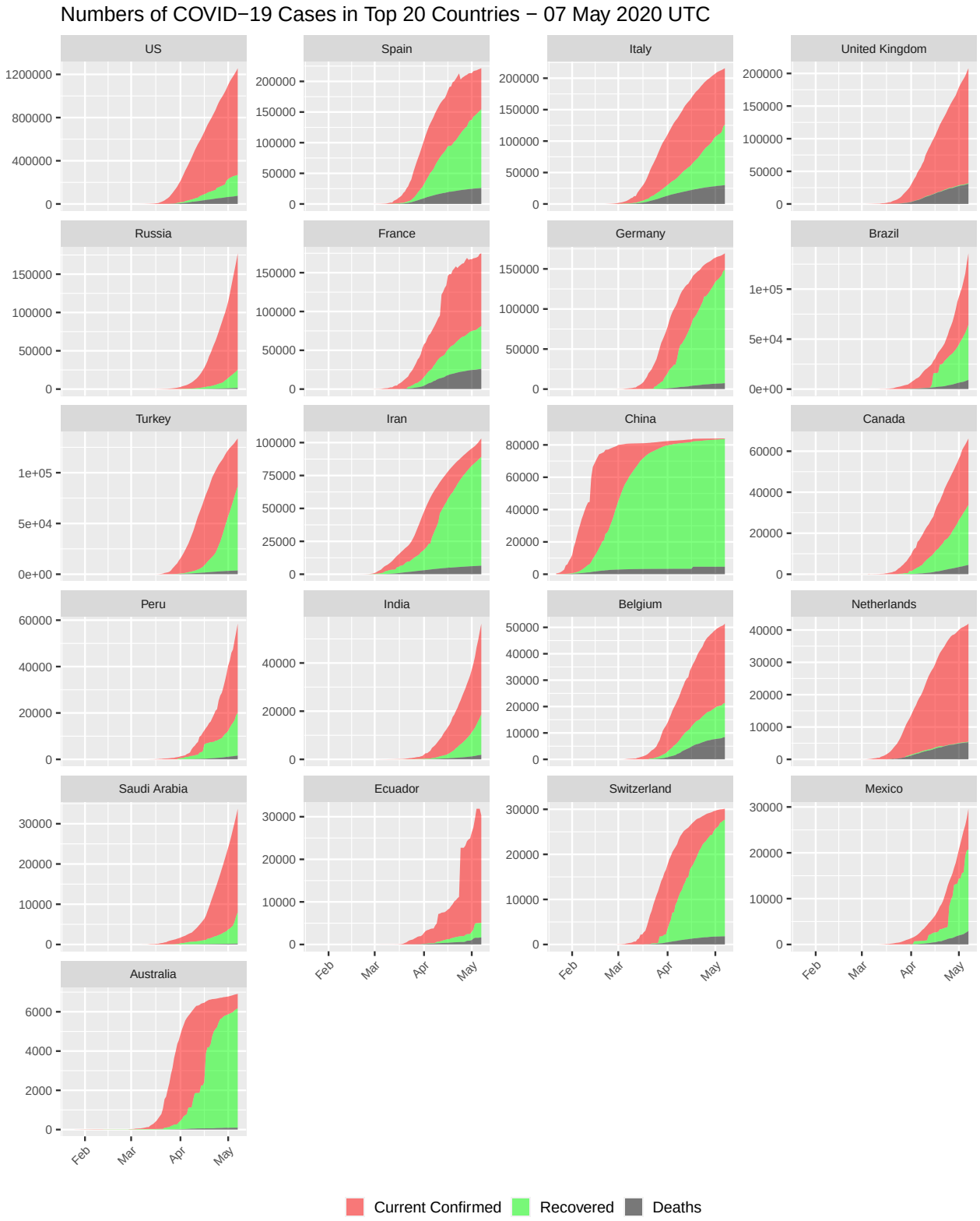


Figure 10: COVID-19 Cases in Top 20 Countries. Ordered descendingly by number of confirmed cases.

```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')
```

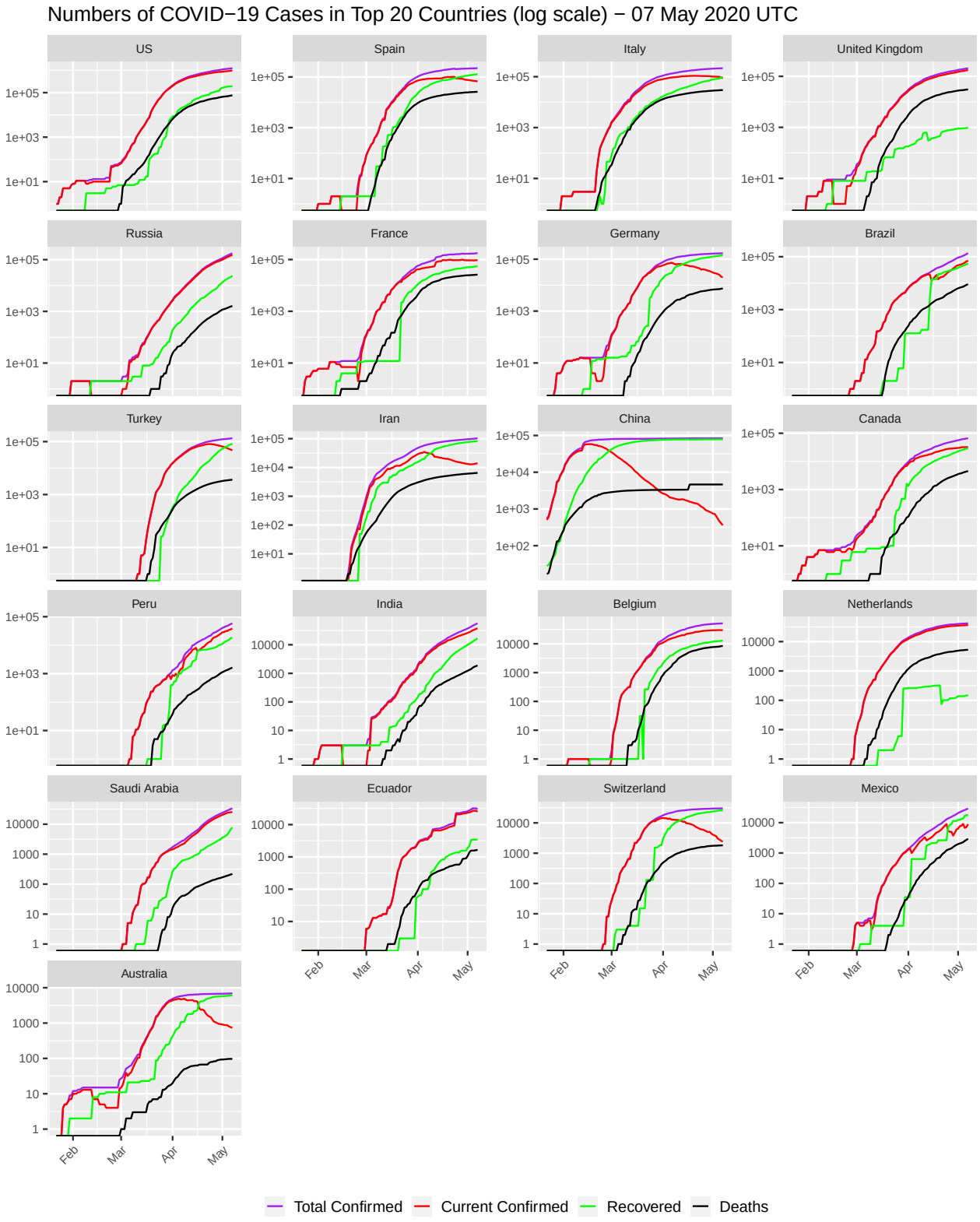


Figure 11: COVID-19 Cases Top 20 Countries (log scale). Ordered descendingly by number of confirmed cases.

```
## 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, 100)) +
  facet_wrap(~country, ncol=4)
```

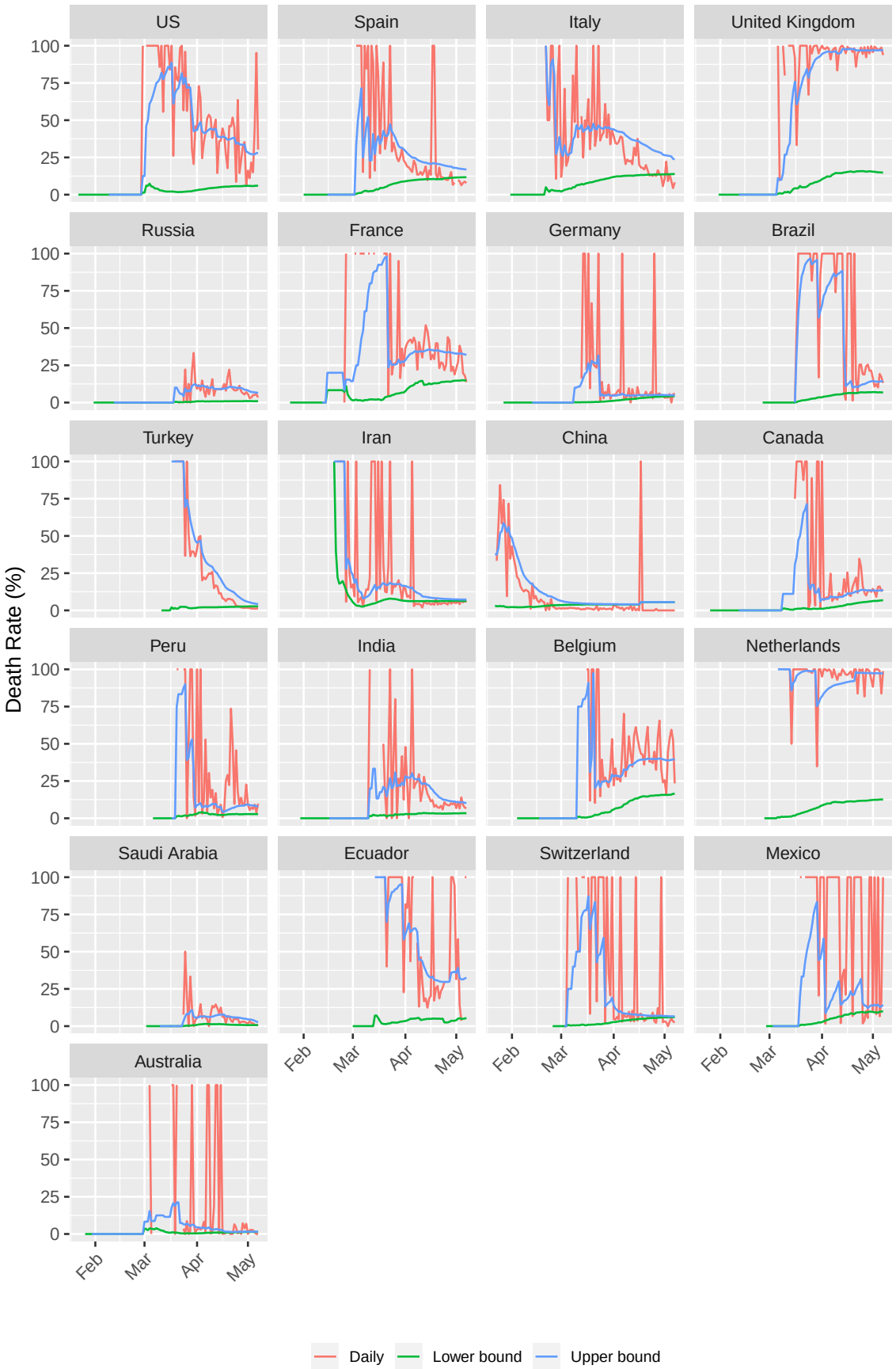


Figure 12: Death Rates

5.4 Countries with Highest Death Rates

Below are a list of top 20 countries with the highest death rates out of countires having 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 - 07 May 2020 UTC

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	Belgium	51,420	639	30,025	12,980	8,415	76	16.4%
2	France	174,918	694	93,737	55,191	25,990	178	14.9%
3	United Kingdom	207,977	5,618	176,318	970	30,689	539	14.8%
4	Italy	215,858	1,401	89,624	96,276	29,958	274	13.9%
5	Netherlands	41,973	455	36,520	147	5,306	85	12.6%
6	Sweden	24,623	705	16,612	4,971	3,040	99	12.3%
7	Hungary	3,150	39	1,966	801	383	10	12.2%
8	Spain	221,447	1,122	66,866	128,511	26,070	213	11.8%
9	Mexico	29,616	1,982	8,874	17,781	2,961	257	10.0%
10	Algeria	5,182	185	2,376	2,323	483	7	9.3%
11	Indonesia	12,776	338	9,465	2,381	930	35	7.3%
12	Canada	66,201	1,507	32,400	29,260	4,541	175	6.9%
13	Brazil	135,773	9,162	71,233	55,350	9,190	602	6.8%
14	Philippines	10,343	339	8,040	1,618	685	27	6.6%
15	Iran	103,135	1,485	13,905	82,744	6,486	68	6.3%
16	Ireland	22,385	137	3,872	17,110	1,403	28	6.3%
17	Romania	14,499	392	7,467	6,144	888	24	6.1%
18	Egypt	7,981	393	5,612	1,887	482	13	6.0%
19	Switzerland	30,126	66	2,416	25,900	1,810	5	6.0%
20	US	1,257,023	28,420	986,325	195,036	75,662	2,231	6.0%

6 Conclusions

As of 07 May 2020 UTC, there are 187 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 7% and 17.3%, 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
2020-05-07	3,845,718	269,567	1,284,741	2,291,410	90,377	5,736	39,328	7.0%	17.3%	12.7%
2020-05-06	3,755,341	263,831	1,245,413	2,246,097	92,650	6,592	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	83,807	5,194	40,722	7.1%	18.2%	11.3%
2020-05-01	3,343,777	238,619	1,052,415	2,052,743	86,867	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	83,801	6,753	28,947	7.2%	20.8%	18.9%
2020-04-22	2,624,602	186,912	710,033	1,727,657	75,564	6,677	30,139	7.1%	20.8%	18.1%
2020-04-21	2,549,038	180,235	679,894	1,688,909	77,311	7,138	34,586	7.1%	21.0%	17.1%
2020-04-20	2,471,727	173,097	645,308	1,653,322	70,951	5,325	22,001	7.0%	21.2%	19.5%
2020-04-19	2,400,776	167,772	623,307	1,609,697	83,542	4,558	31,592	7.0%	21.2%	12.6%
2020-04-18	2,317,234	163,214	591,715	1,562,305	77,606	6,410	23,958	7.0%	21.6%	21.1%
2020-04-17	2,239,628	156,804	567,757	1,515,067	87,836	8,858	26,165	7.0%	21.6%	25.3%
2020-04-16	2,151,792	147,946	541,592	1,462,254	96,369	7,288	31,088	6.9%	21.5%	19.0%
2020-04-15	2,055,423	140,658	510,504	1,404,261	79,866	8,219	36,536	6.8%	21.6%	18.4%
2020-04-14	1,975,557	132,439	473,968	1,369,150	70,400	6,878	25,313	6.7%	21.8%	21.4%
2020-04-13	1,905,157	125,561	448,655	1,330,941	70,012	5,708	26,933	6.6%	21.9%	17.5%
2020-04-12	1,835,145	119,853	421,722	1,293,570	99,120	5,707	19,612	6.5%	22.1%	22.5%
2020-04-11	1,736,025	114,146	402,110	1,219,769	78,096	6,033	26,014	6.6%	22.1%	18.8%
2020-04-10	1,657,929	108,113	376,096	1,173,720	92,391	7,231	22,121	6.5%	22.3%	24.6%
2020-04-09	1,565,538	100,882	353,975	1,110,681	85,338	7,528	25,314	6.4%	22.2%	22.9%
2020-04-08	1,480,200	93,354	328,661	1,058,185	83,762	6,692	28,607	6.3%	22.1%	19.0%
2020-04-07	1,396,438	86,662	300,054	1,009,722	75,011	7,895	23,539	6.2%	22.4%	25.1%
2020-04-06	1,321,427	78,767	276,515	966,145	71,690	5,691	16,503	6.0%	22.2%	25.6%
2020-04-05	1,249,737	73,076	260,012	916,649	73,678	5,095	13,860	5.8%	21.9%	26.9%
2020-04-04	1,176,059	67,981	246,152	861,926	80,183	6,169	20,356	5.8%	21.6%	23.3%
2020-04-03	1,095,876	61,812	225,796	808,268	82,418	5,995	15,533	5.6%	21.5%	27.8%
2020-04-02	1,013,458	55,817	210,263	747,378	80,820	6,142	17,086	5.5%	21.0%	26.4%
2020-04-01	932,638	49,675	193,177	689,786	75,030	5,437	15,143	5.3%	20.5%	26.4%
2020-03-31	857,608	44,238	178,034	635,336	75,118	4,799	13,468	5.2%	19.9%	26.3%
2020-03-30	782,490	39,439	164,566	578,485	62,205	4,116	15,484	5.0%	19.3%	21.0%
2020-03-29	720,285	35,323	149,082	535,880	59,461	3,526	9,667	4.9%	19.2%	26.7%
2020-03-28	660,824	31,797	139,415	489,612	67,401	3,509	8,500	4.8%	18.6%	29.2%
2020-03-27	593,423	28,288	130,915	434,220	63,722	3,500	8,765	4.8%	17.8%	28.5%
2020-03-26	529,701	24,788	122,150	382,763	61,978	3,003	8,363	4.7%	16.9%	26.4%
2020-03-25	467,723	21,785	113,787	332,151	49,644	2,771	5,787	4.7%	16.1%	32.4%
2020-03-24	418,079	19,014	108,000	291,065	39,797	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	32,463	1,700	6,207	4.4%	13.2%	21.5%
2020-03-21	304,555	13,125	91,692	199,738	32,308	1,703	4,272	4.3%	12.5%	28.5%
2020-03-20	272,247	11,422	87,420	173,405	29,631	1,476	2,445	4.2%	11.6%	37.6%
2020-03-19	242,616	9,946	84,975	147,695	27,770	1,123	1,663	4.1%	10.5%	40.3%
2020-03-18	214,846	8,823	83,312	122,711	17,733	867	2,472	4.1%	9.6%	26.0%
2020-03-17	197,113	7,956	80,840	108,317	15,510	806	2,752	4.0%	9.0%	22.7%
2020-03-16	181,603	7,150	78,088	96,365	14,137	678	2,054	3.9%	8.4%	24.8%
2020-03-15	167,466	6,472	76,034	84,960	11,350	642	3,410	3.9%	7.8%	15.8%
2020-03-14	156,116	5,830	72,624	77,662	10,897	422	2,373	3.7%	7.4%	15.1%
2020-03-13	145,219	5,408	70,251	69,560	16,867	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%

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
2020-03-02	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 (07 May 2020 UTC)

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	World	3,845,718	90,377	2,291,410	1,284,741	269,567	5,736	**7.0%**
2	US	1,257,023	28,420	986,325	195,036	75,662	2,231	**6.0%**
3	Spain	221,447	1,122	66,866	128,511	26,070	213	**11.8%**
4	Italy	215,858	1,401	89,624	96,276	29,958	274	**13.9%**
5	United Kingdom	207,977	5,618	176,318	970	30,689	539	**14.8%**
6	Russia	177,160	11,231	151,732	23,803	1,625	88	0.9%
7	France	174,918	694	93,737	55,191	25,990	178	**14.9%**
8	Germany	169,430	1,268	20,338	141,700	7,392	117	4.4%
9	Brazil	135,773	9,162	71,233	55,350	9,190	602	**6.8%**
10	Turkey	133,721	1,977	47,096	82,984	3,641	57	2.7%
11	Iran	103,135	1,485	13,905	82,744	6,486	68	**6.3%**

Table 6: Cases by Country (07 May 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
12	China	83,975	5	361	78,977	4,637	0	**5.5%**
13	Canada	66,201	1,507	32,400	29,260	4,541	175	**6.9%**
14	Peru	58,526	3,709	38,511	18,388	1,627	94	2.8%
15	India	56,351	3,364	37,686	16,776	1,889	104	3.4%
16	Belgium	51,420	639	30,025	12,980	8,415	76	**16.4%**
17	Netherlands	41,973	455	36,520	147	5,306	85	**12.6%**
18	Saudi Arabia	33,731	1,793	25,714	7,798	219	10	0.6%
19	Ecuador	30,298	0	25,211	3,433	1,654	36	**5.5%**
20	Switzerland	30,126	66	2,416	25,900	1,810	5	**6.0%**
21	Mexico	29,616	1,982	8,874	17,781	2,961	257	**10.0%**
22	Portugal	26,715	533	23,352	2,258	1,105	16	4.1%
23	Pakistan	24,644	571	17,595	6,464	585	21	2.4%
24	Sweden	24,623	705	16,612	4,971	3,040	99	**12.3%**
25	Chile	24,581	1,533	12,632	11,664	285	4	1.2%
26	Ireland	22,385	137	3,872	17,110	1,403	28	**6.3%**
27	Singapore	20,939	741	19,207	1,712	20	0	0.1%
28	Belarus	20,168	913	14,985	5,067	116	4	0.6%
29	Qatar	18,890	918	16,592	2,286	12	0	0.1%
30	Israel	16,381	71	5,268	10,873	240	1	1.5%
31	United Arab Emirates	16,240	502	12,503	3,572	165	8	1.0%
32	Austria	15,752	68	1,445	13,698	609	1	3.9%
33	Japan	15,477	224	9,982	4,918	577	21	3.7%
34	Poland	15,047	307	9,430	4,862	755	22	**5.0%**
35	Romania	14,499	392	7,467	6,144	888	24	**6.1%**
36	Ukraine	13,691	507	10,955	2,396	340	13	2.5%
37	Indonesia	12,776	338	9,465	2,381	930	35	**7.3%**
38	Bangladesh	12,425	706	10,316	1,910	199	13	1.6%
39	Korea, South	10,822	12	1,082	9,484	256	0	2.4%
40	Philippines	10,343	339	8,040	1,618	685	27	**6.6%**
41	Denmark	10,281	145	1,860	7,907	514	8	**5.0%**
42	Serbia	9,848	57	7,482	2,160	206	3	2.1%
43	Colombia	9,456	497	6,749	2,300	407	10	4.3%
44	Dominican Republic	9,095	288	6,658	2,064	373	11	4.1%
45	South Africa	8,232	424	4,918	3,153	161	8	2.0%
46	Norway	8,034	38	7,785	32	217	1	2.7%
47	Czechia	8,031	57	3,390	4,371	270	8	3.4%
48	Egypt	7,981	393	5,612	1,887	482	13	**6.0%**
49	Panama	7,868	137	6,757	886	225	7	2.9%
50	Australia	6,913	19	738	6,078	97	0	1.4%
51	Kuwait	6,567	278	4,142	2,381	44	2	0.7%
52	Malaysia	6,467	39	1,584	4,776	107	0	1.7%
53	Finland	5,673	100	1,918	3,500	255	3	4.5%
54	Morocco	5,548	140	3,186	2,179	183	0	3.3%
55	Argentina	5,371	163	3,488	1,601	282	9	**5.3%**
56	Algeria	5,182	185	2,376	2,323	483	7	**9.3%**
57	Moldova	4,605	129	2,713	1,747	145	2	3.1%
58	Kazakhstan	4,578	156	3,030	1,518	30	0	0.7%
59	Bahrain	4,199	265	2,191	2,000	8	0	0.2%
60	Luxembourg	3,859	8	254	3,505	100	2	2.6%
61	Afghanistan	3,563	171	2,989	468	106	2	3.0%
62	Nigeria	3,526	381	2,818	601	107	4	3.0%
63	Hungary	3,150	39	1,966	801	383	10	**12.2%**
64	Ghana	3,091	0	2,770	303	18	0	0.6%
65	Thailand	2,992	3	165	2,772	55	0	1.8%
66	Oman	2,958	55	1,963	980	15	2	0.5%
67	Armenia	2,884	102	1,657	1,185	42	2	1.5%
68	Greece	2,678	15	1,156	1,374	148	1	**5.5%**
69	Iraq	2,543	63	815	1,626	102	0	4.0%
70	Uzbekistan	2,298	65	632	1,656	10	0	0.4%
71	Cameroon	2,267	2	1,157	1,002	108	0	4.8%
72	Azerbaijan	2,204	77	625	1,551	28	0	1.3%
73	Croatia	2,125	6	398	1,641	86	1	4.0%
74	Bolivia	2,081	195	1,760	219	102	11	4.9%
75	Bosnia and Herzegovina	2,027	40	983	954	90	4	4.4%
76	Guinea	1,927	71	1,287	629	11	0	0.6%
77	Bulgaria	1,829	51	1,361	384	84	0	4.6%
78	Iceland	1,801	2	36	1,755	10	0	0.6%
79	Cuba	1,729	26	625	1,031	73	4	4.2%

Table 6: Cases by Country (07 May 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
80	Estonia	1,720	7	1,391	273	56	1	3.3%
81	Honduras	1,685	224	1,426	154	105	6	6.2%
82	North Macedonia	1,572	33	404	1,079	89	1	5.7%
83	Cote d'Ivoire	1,571	55	809	742	20	2	1.3%
84	Senegal	1,492	59	917	562	13	1	0.9%
85	New Zealand	1,490	1	122	1,347	21	0	1.4%
86	Slovenia	1,449	1	1,103	247	99	0	6.8%
87	Slovakia	1,445	16	613	806	26	1	1.8%
88	Lithuania	1,433	5	645	739	49	1	3.4%
89	Djibouti	1,133	9	331	799	3	0	0.3%
90	Tunisia	1,026	1	382	600	44	1	4.3%
91	Sudan	930	78	786	92	52	3	5.6%
92	Somalia	928	55	778	106	44	5	4.7%
93	Latvia	909	9	427	464	18	1	2.0%
94	Kyrgyzstan	895	24	246	637	12	0	1.3%
95	Cyprus	889	6	474	400	15	0	1.7%
96	Congo (Kinshasa)	863	66	724	103	36	1	4.2%
97	Kosovo	861	5	272	562	27	1	3.1%
98	Albania	842	10	206	605	31	0	3.7%
99	Guatemala	832	34	719	90	23	2	2.8%
100	Sri Lanka	824	27	583	232	9	0	1.1%
101	Lebanon	784	34	539	220	25	0	3.2%
102	Niger	781	11	153	586	42	4	5.4%
103	Costa Rica	765	4	314	445	6	0	0.8%
104	Andorra	752	1	179	526	47	1	6.2%
105	Burkina Faso	736	7	126	562	48	0	6.5%
106	Diamond Princess	712	0	54	645	13	0	1.8%
107	El Salvador	695	62	428	252	15	0	2.2%
108	Uruguay	684	11	175	492	17	0	2.5%
109	Mali	650	19	347	271	32	0	4.9%
110	Maldives	648	31	625	20	3	1	0.5%
111	San Marino	622	14	475	106	41	0	6.6%
112	Georgia	615	5	331	275	9	0	1.5%
113	Kenya	607	25	381	197	29	3	4.8%
114	Guinea-Bissau	564	89	537	25	2	0	0.4%
115	Gabon	504	107	386	110	8	2	1.6%
116	Jordan	494	21	104	381	9	0	1.8%
117	Jamaica	488	10	421	58	9	0	1.8%
118	Malta	486	2	68	413	5	0	1.0%
119	Tanzania	480	0	297	167	16	0	3.3%
120	Paraguay	462	22	304	148	10	0	2.2%
121	Tajikistan	461	82	449	0	12	4	2.6%
122	Taiwan*	440	1	87	347	6	0	1.4%
123	Equatorial Guinea	439	0	422	13	4	0	0.9%
124	Venezuela	381	2	186	185	10	0	2.6%
125	West Bank and Gaza	375	1	197	176	2	0	0.5%
126	Mauritius	332	0	2	320	10	0	3.0%
127	Montenegro	324	0	51	265	8	0	2.5%
128	Vietnam	288	17	55	233	0	0	0.0%
129	Rwanda	271	3	138	133	0	0	0.0%
130	Congo (Brazzaville)	264	0	224	30	10	0	3.8%
131	Chad	253	83	176	50	27	10	10.7%
132	Sierra Leone	231	6	161	54	16	2	6.9%
133	Cabo Verde	218	27	178	38	2	0	0.9%
134	Madagascar	193	35	92	101	0	0	0.0%
135	Ethiopia	191	29	94	93	4	0	2.1%
136	Liberia	189	11	90	79	20	0	10.6%
137	Sao Tome and Principe	187	13	179	4	4	1	2.1%
138	Burma	176	15	108	62	6	0	3.4%
139	Eswatini	153	30	139	12	2	0	1.3%
140	Zambia	153	7	46	103	4	0	2.6%
141	Brunei	141	2	9	131	1	0	0.7%
142	Benin	140	44	85	53	2	0	1.4%
143	Togo	135	7	41	85	9	0	6.7%
144	Haiti	129	28	101	16	12	0	9.3%
145	Cambodia	122	0	2	120	0	0	0.0%
146	Trinidad and Tobago	116	0	5	103	8	0	6.9%
147	Nepal	101	2	79	22	0	0	0.0%

Table 6: Cases by Country (07 May 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
148	Uganda	101	1	46	55	0	0	0.0%
149	Monaco	95	0	9	82	4	0	4.2%
150	Central African Republic	94	0	84	10	0	0	0.0%
151	Guyana	93	0	56	27	10	0	10.8%
152	Bahamas	92	0	55	26	11	0	12.0%
153	Barbados	82	0	22	53	7	0	8.5%
154	Liechtenstein	82	0	26	55	1	0	1.2%
155	Mozambique	81	0	57	24	0	0	0.0%
156	South Sudan	74	16	74	0	0	0	0.0%
157	Libya	64	0	37	24	3	0	4.7%
158	Syria	45	0	15	27	3	0	6.7%
159	Malawi	43	0	26	14	3	0	7.0%
160	Mongolia	41	0	28	13	0	0	0.0%
161	Eritrea	39	0	9	30	0	0	0.0%
162	Angola	36	0	23	11	2	0	5.6%
163	Zimbabwe	34	0	25	5	4	0	11.8%
164	Antigua and Barbuda	25	0	6	16	3	0	12.0%
165	Yemen	25	0	19	1	5	0	20.0%
166	Timor-Leste	24	0	3	21	0	0	0.0%
167	Botswana	23	0	13	9	1	0	4.3%
168	Grenada	21	0	8	13	0	0	0.0%
169	Laos	19	0	10	9	0	0	0.0%
170	Belize	18	0	0	16	2	0	11.1%
171	Fiji	18	0	4	14	0	0	0.0%
172	Gambia	18	1	8	9	1	0	5.6%
173	Saint Lucia	18	0	1	17	0	0	0.0%
174	Saint Vincent and the Grenadines	17	0	8	9	0	0	0.0%
175	Dominica	16	0	2	14	0	0	0.0%
176	Namibia	16	0	7	9	0	0	0.0%
177	Nicaragua	16	0	4	7	5	0	31.2%
178	Burundi	15	0	7	7	1	0	6.7%
179	Saint Kitts and Nevis	15	0	3	12	0	0	0.0%
180	Holy See	12	0	10	2	0	0	0.0%
181	Seychelles	11	0	3	8	0	0	0.0%
182	Suriname	10	0	0	9	1	0	10.0%
183	MS Zaandam	9	0	7	0	2	0	22.2%
184	Comoros	8	0	7	0	1	0	12.5%
185	Mauritania	8	0	1	6	1	0	12.5%
186	Papua New Guinea	8	0	0	8	0	0	0.0%
187	Bhutan	7	0	2	5	0	0	0.0%
188	Western Sahara	6	0	1	5	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!