COVID-19 Data Analysis with R - Worldwide*

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

05 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] 266 107
```

Each dataset has 266 rows, corresponding to country/region/province/state. It has 107 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" "2020-05-03"
```

```
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 03 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 03 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$confirmed), stroke=F,
                        color='red', fillOpacity=0.3,
                        popup=x$txt)

# world
m
```

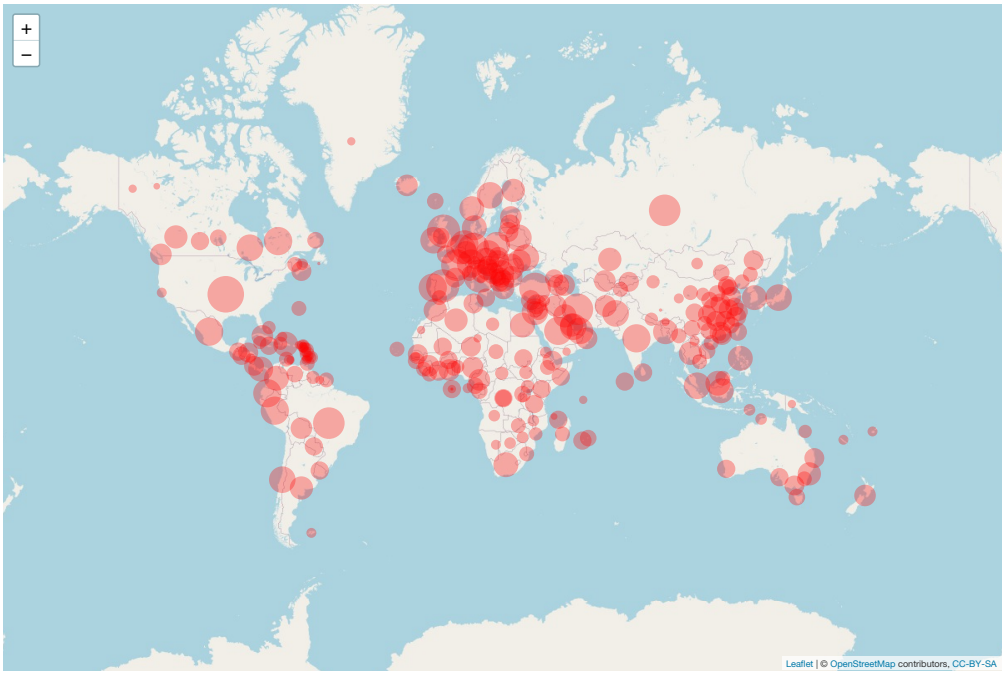


Figure 1: World Map

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)
```

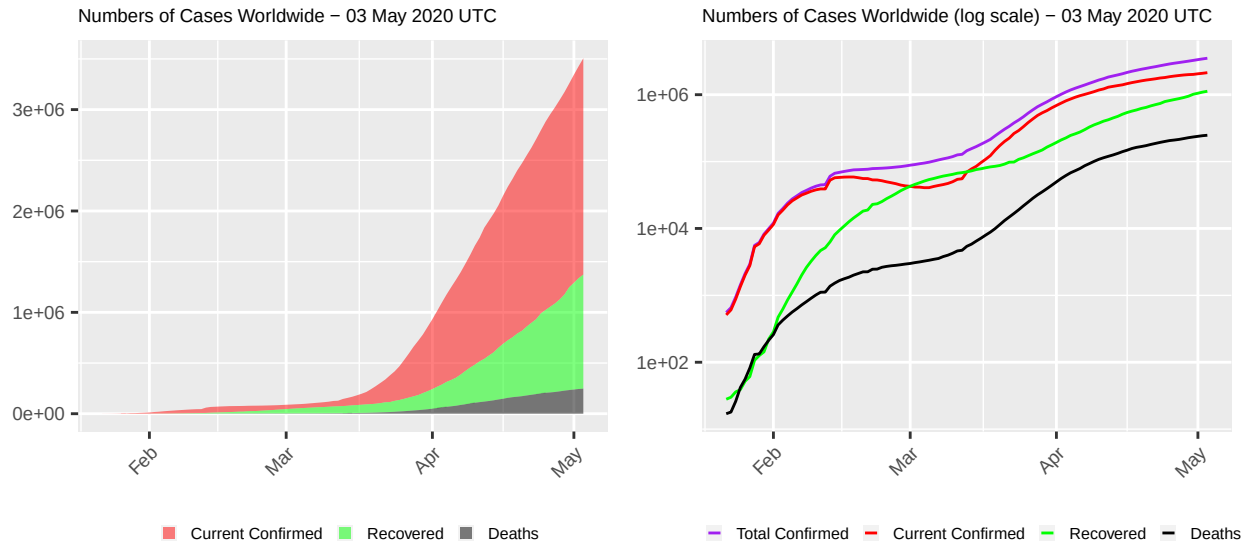


Figure 2: COVID-19 Cases Worldwide

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)
```

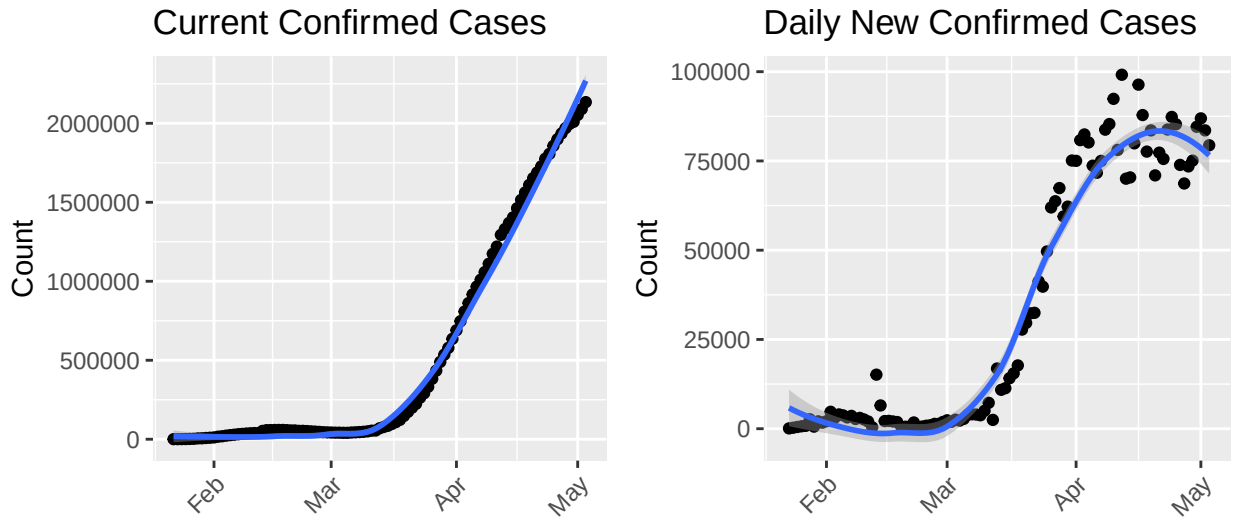


Figure 3: Current Confirmed Cases

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)
```

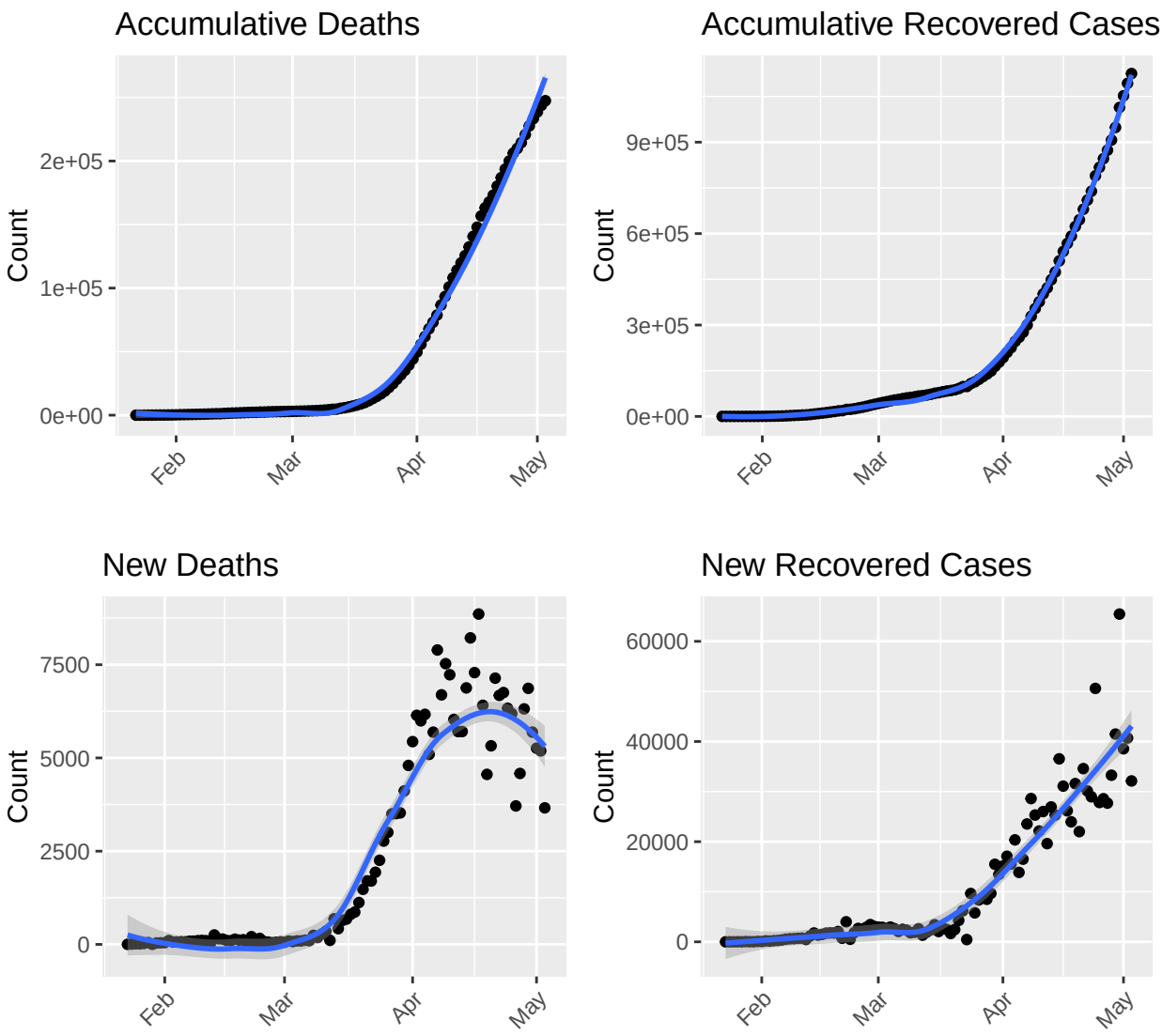


Figure 4: Deaths and Recovered Cases

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 03 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 03 May 2020 UTC UTC, it will be between 7.1% and 18%.

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)
```

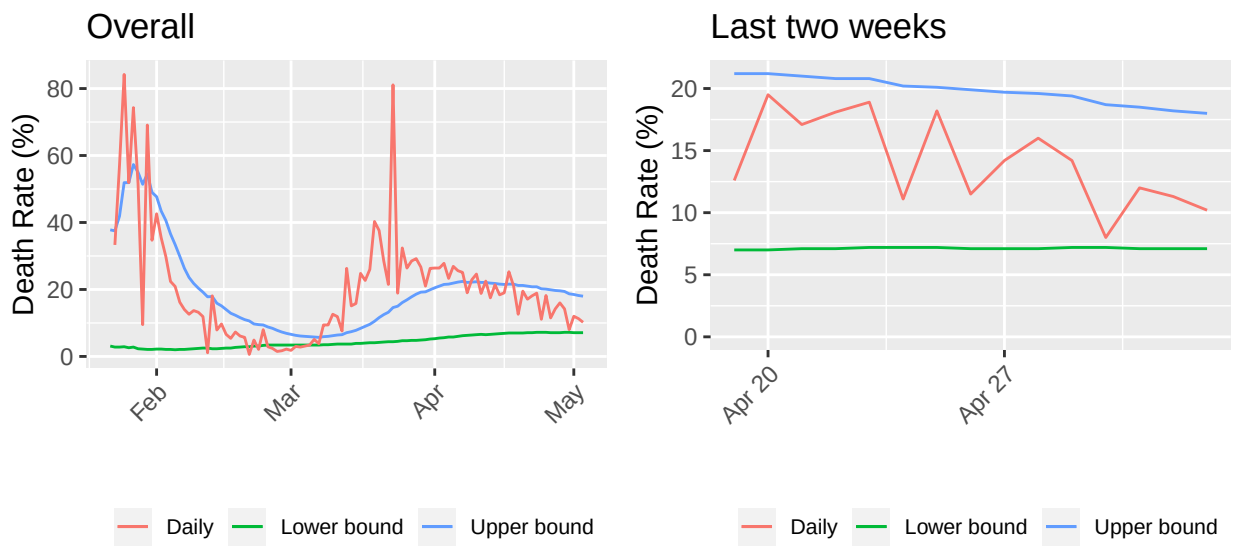


Figure 5: Death Rate

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)))
```

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

```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)))
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 - 03 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	3,506,729	247,470	7.1%	79,386	3,662	2,134,023
2	US	1,158,040	67,682	5.8%	25,501	1,313	910,206
3	Spain	217,466	25,264	11.6%	884	164	73,300
4	Italy	210,717	28,884	13.7%	1,389	174	100,179
5	United Kingdom	187,842	28,520	15.2%	4,342	315	158,421
6	France	168,925	24,900	14.7%	407	137	93,140
7	Germany	165,664	6,866	4.1%	697	54	28,198
8	Russia	134,687	1,280	1.0%	10,633	58	116,768
9	Turkey	126,045	3,397	2.7%	1,670	61	59,497
10	Brazil	101,826	7,051	6.9%	4,726	290	51,784
11	Iran	97,424	6,203	6.4%	976	47	12,799
12	China	83,964	4,637	5.5%	5	0	643
13	Canada	60,504	3,795	6.3%	2,578	111	31,788
14	Belgium	49,906	7,844	15.7%	389	79	29,753
15	Peru	45,928	1,286	2.8%	3,394	86	31,092
16	India	42,505	1,391	3.3%	2,806	68	29,339
17	Netherlands	40,769	5,072	12.4%	335	69	35,559
18	Switzerland	29,905	1,762	5.9%	88	0	3,643
19	Ecuador	29,538	1,564	5.3%	2,074	193	24,674
20	Saudi Arabia	27,011	184	0.7%	1,552	8	22,693
21	Portugal	25,282	1,043	4.1%	92	20	22,550
22	Others	502,781	18,845	3.7%	14,848	415	297,997

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

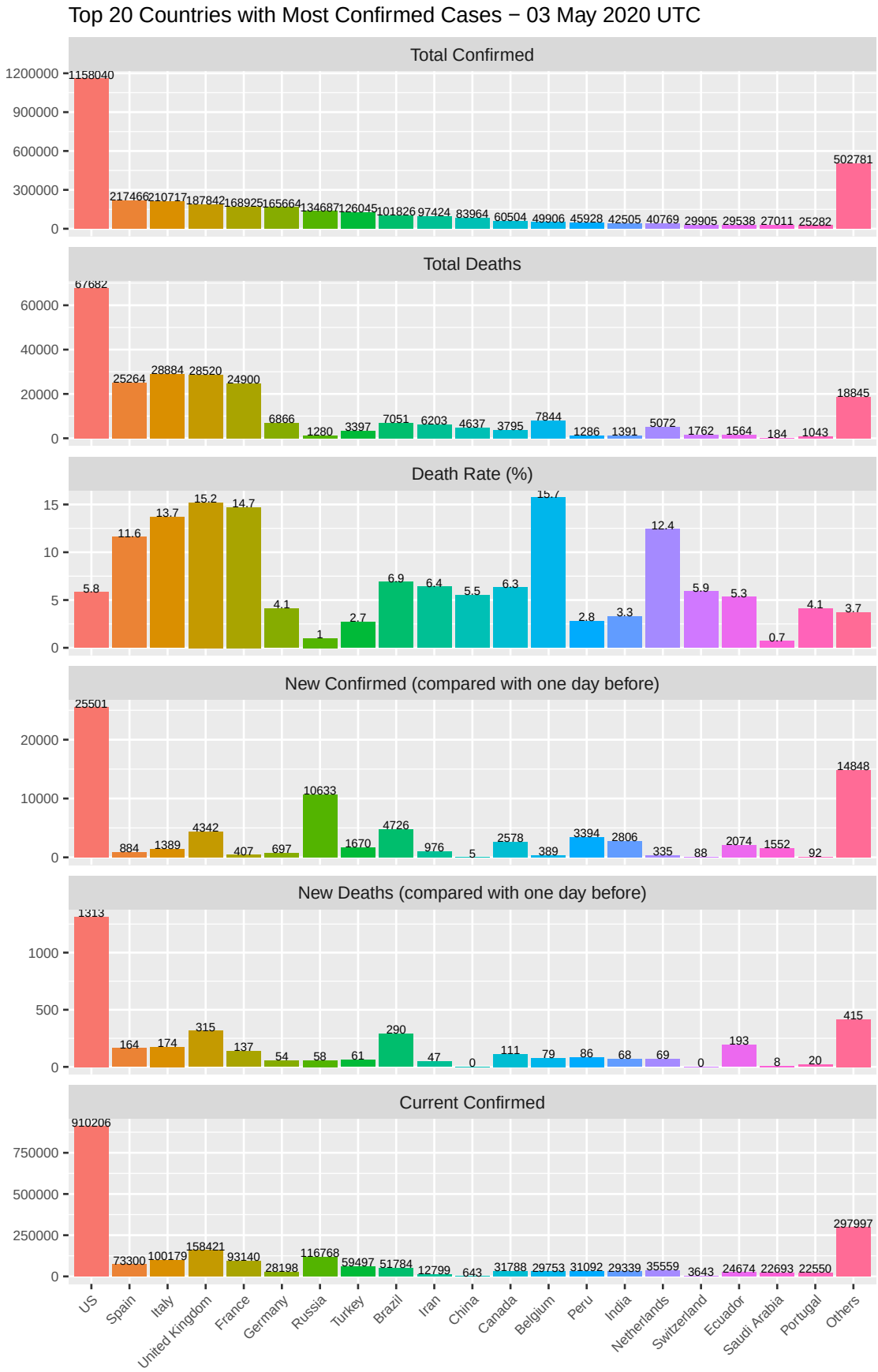
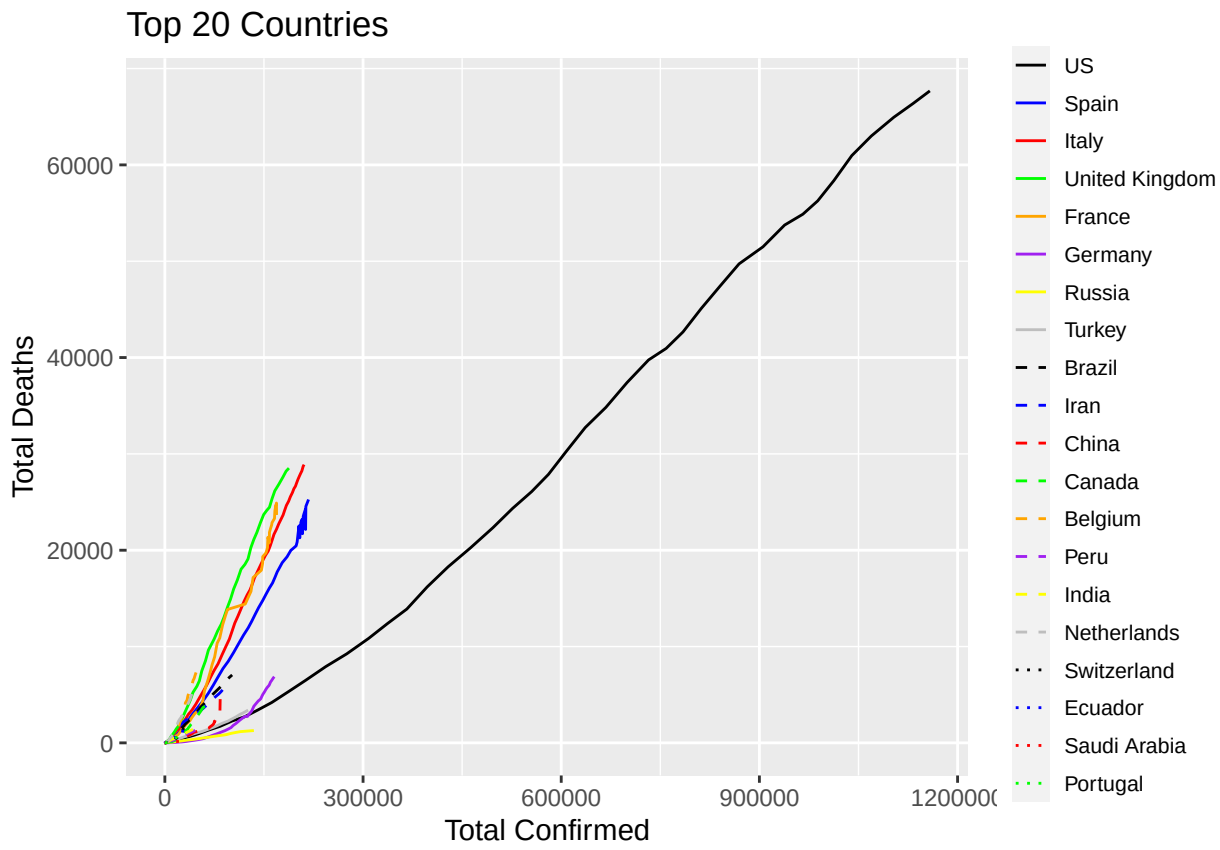


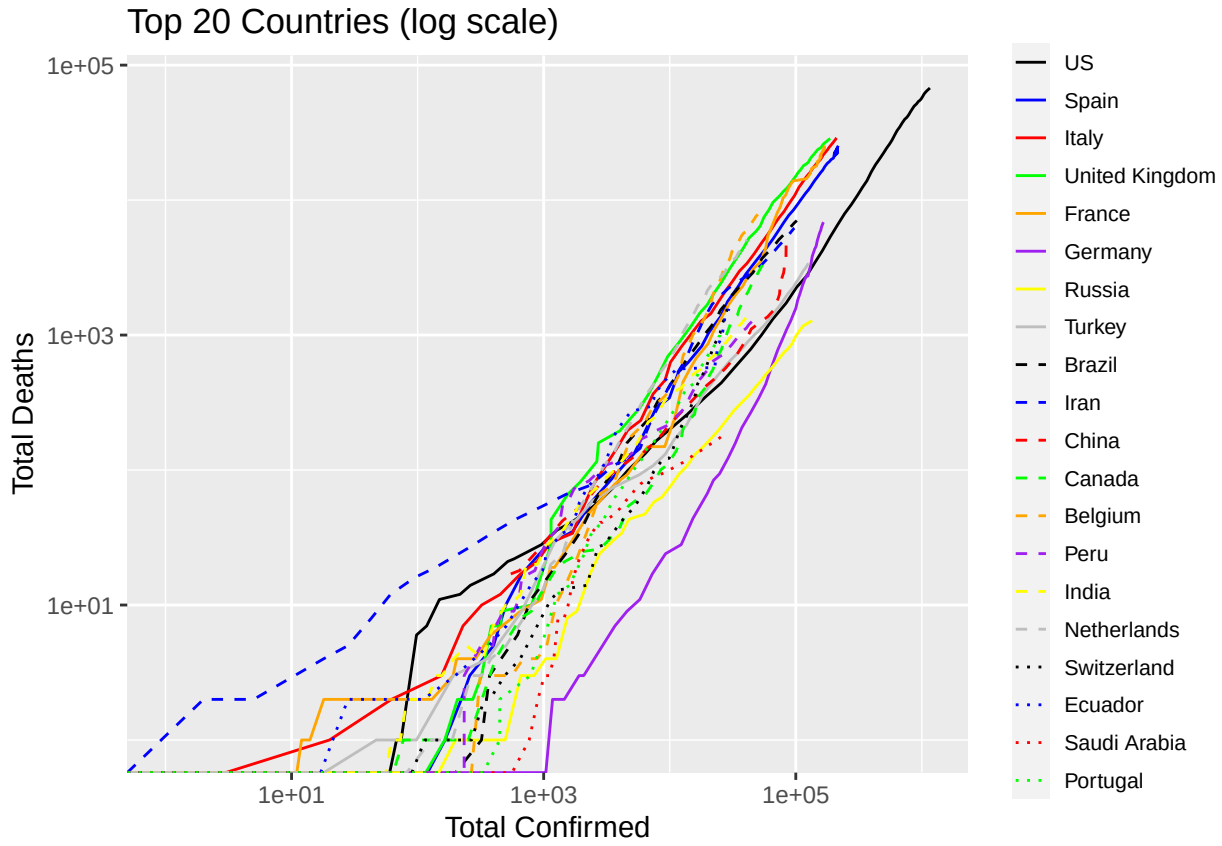
Figure 6: Top 20 Countries with Most Confirmed Cases

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 03 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)
```

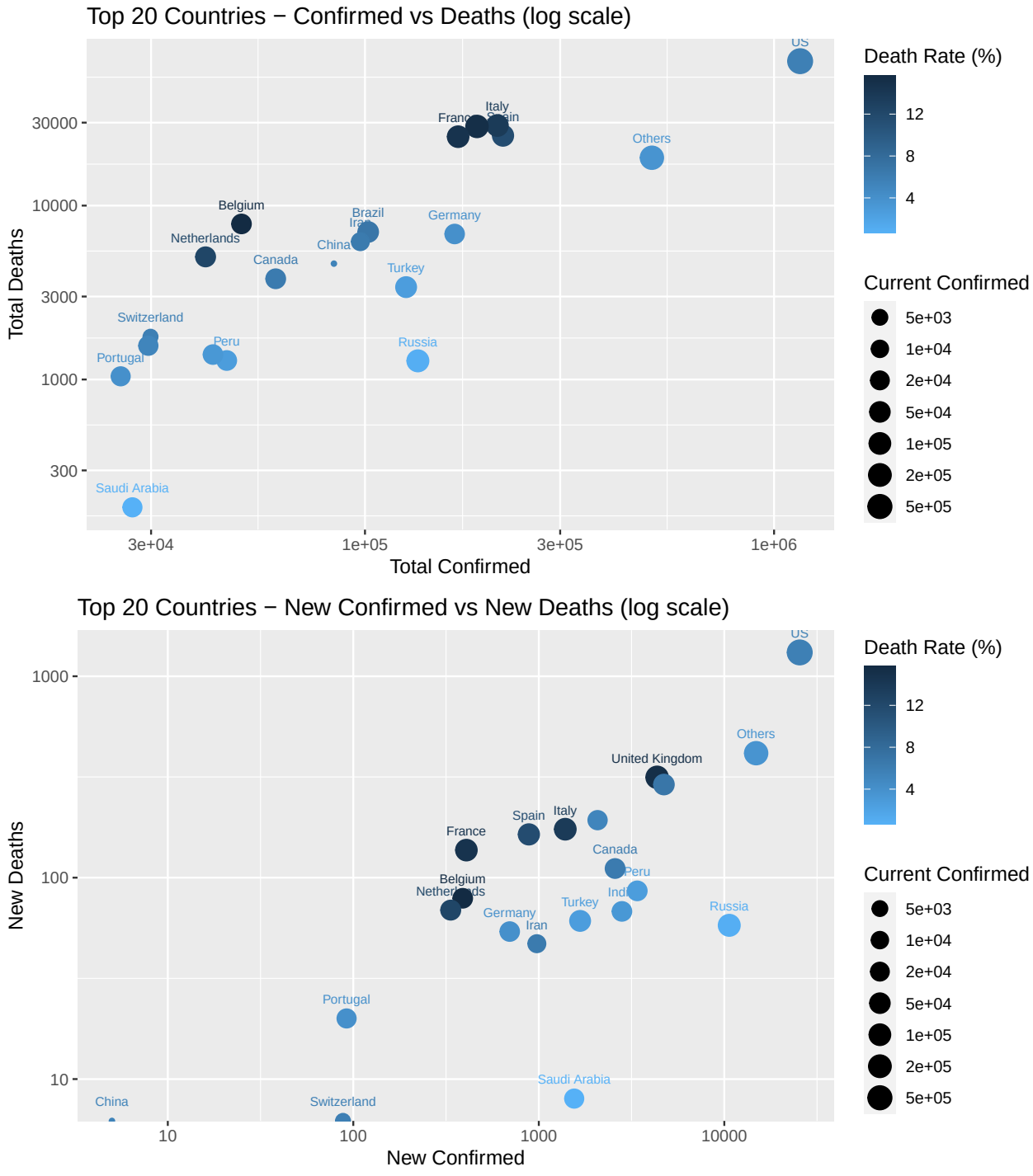


Figure 7: Top 20 Countries

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)
```

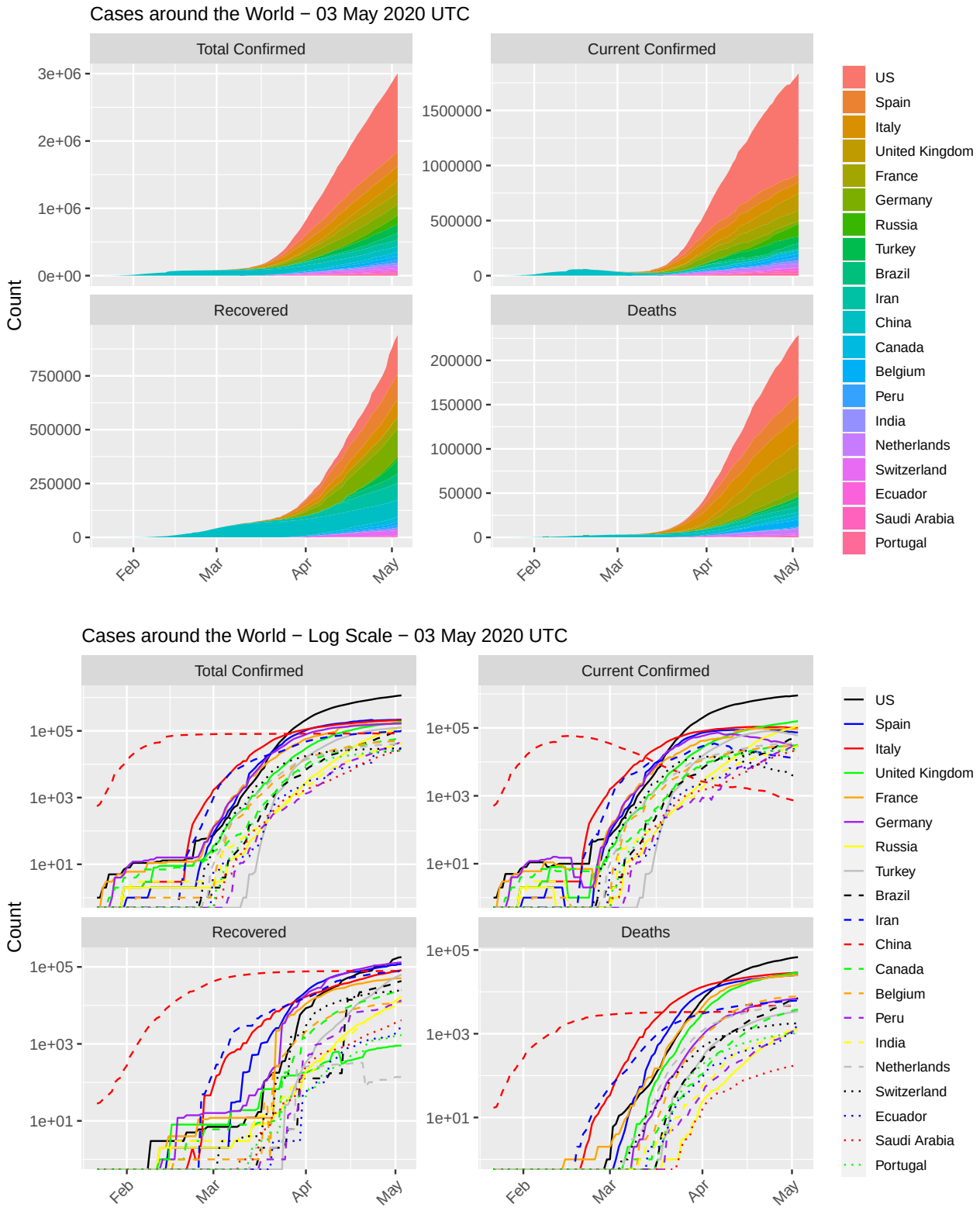


Figure 8: Cases around the World

```
## 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))
```

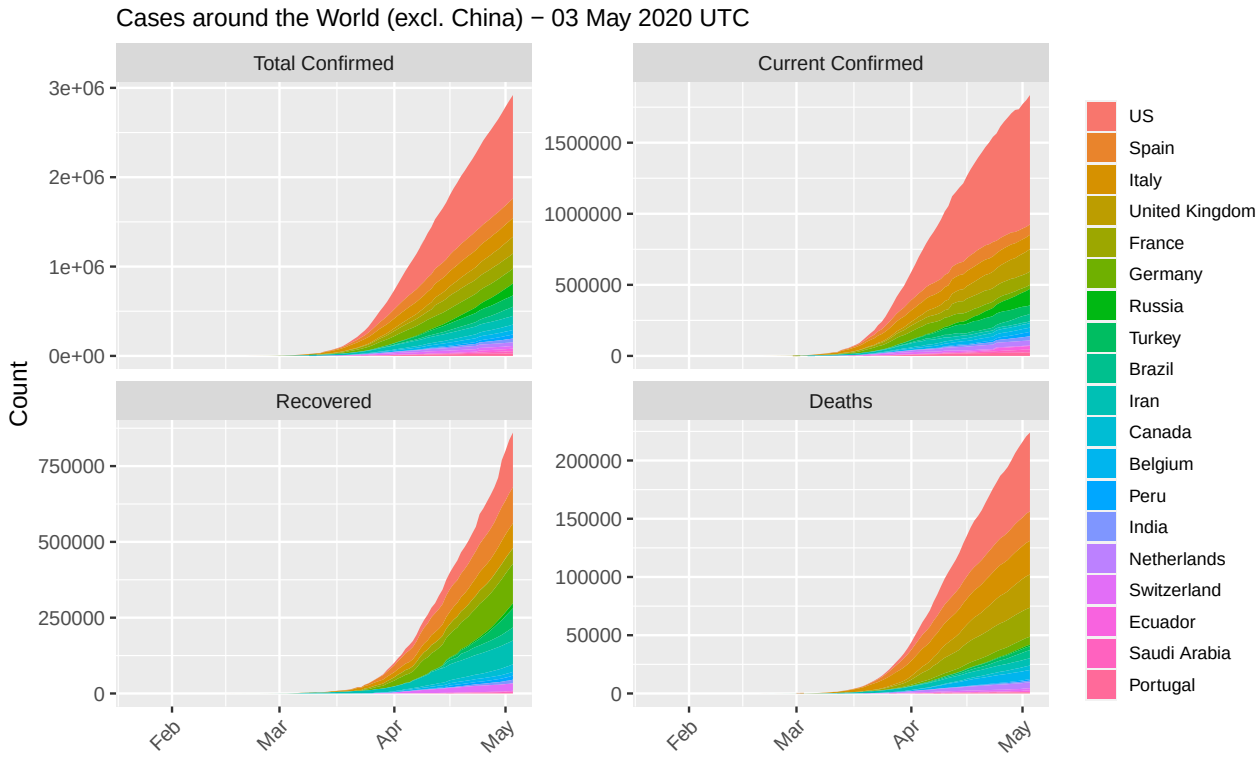


Figure 9: Cases around the World (excl. China)

```
## 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')
```

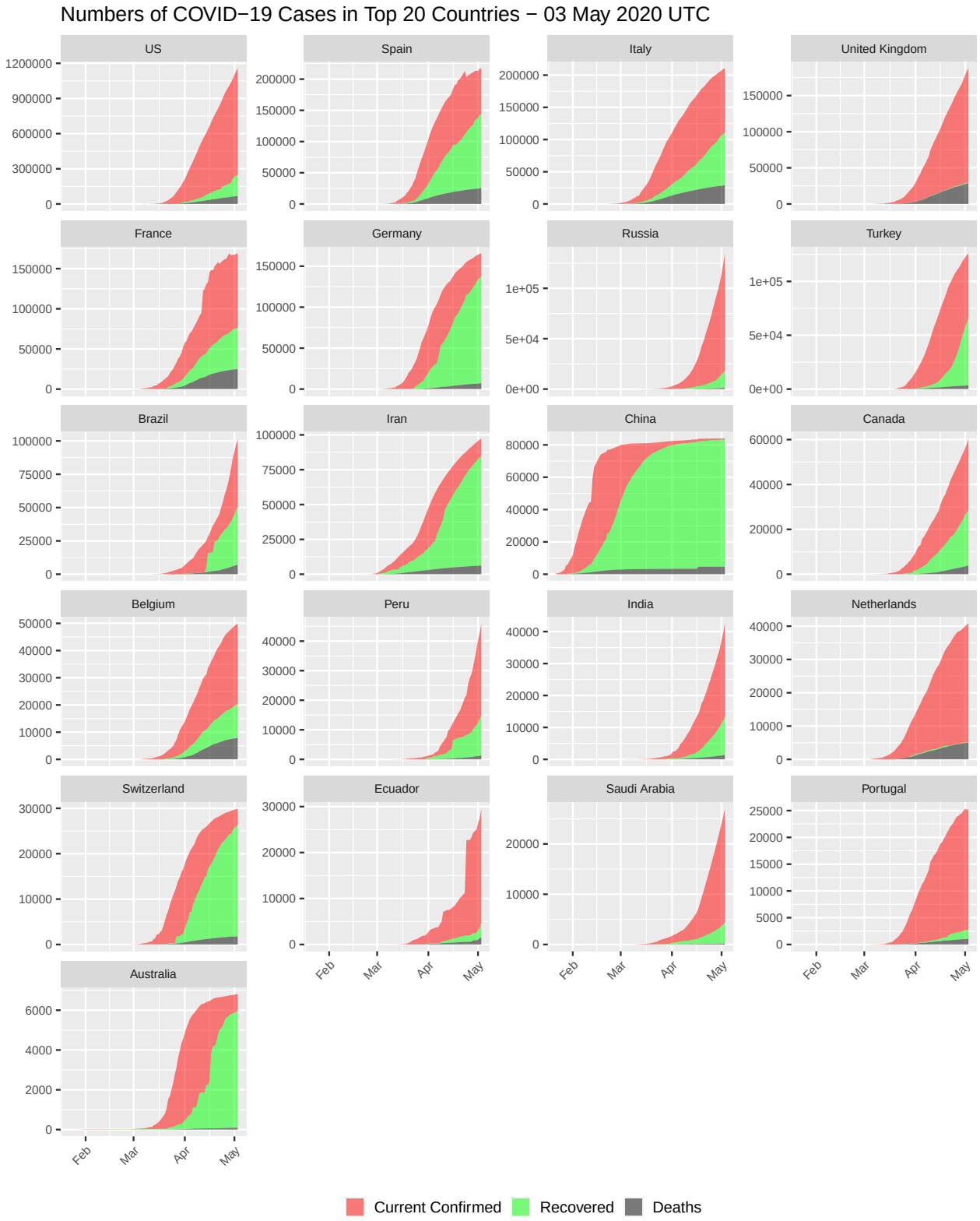


Figure 10: COVID-19 Cases in Top 20 Countries. Ordered descendingly by number of confirmed cases.

```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')
```

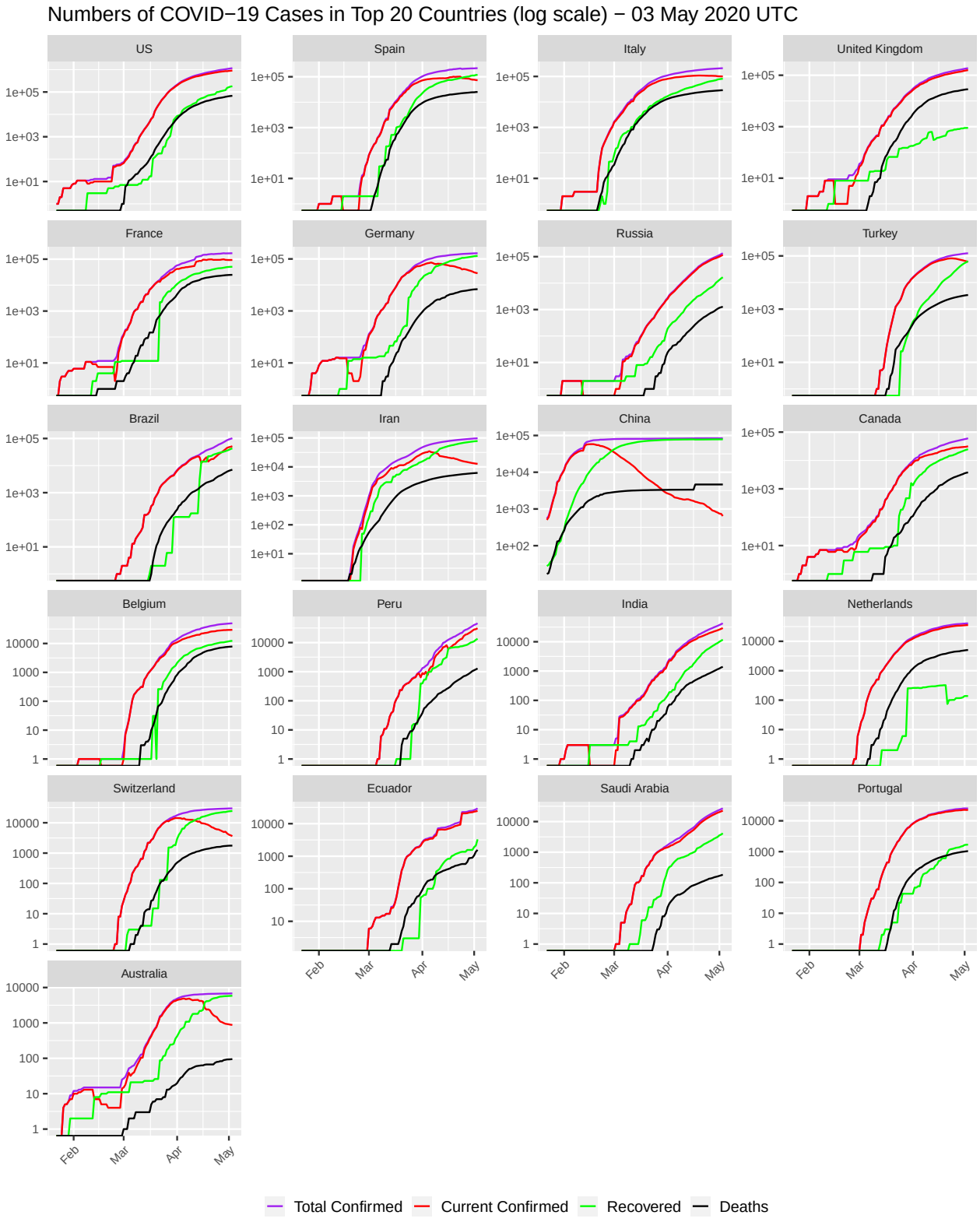


Figure 11: COVID-19 Cases Top 20 Countries (log scale). Ordered descendingly by number of confirmed cases.

```
## 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)
```

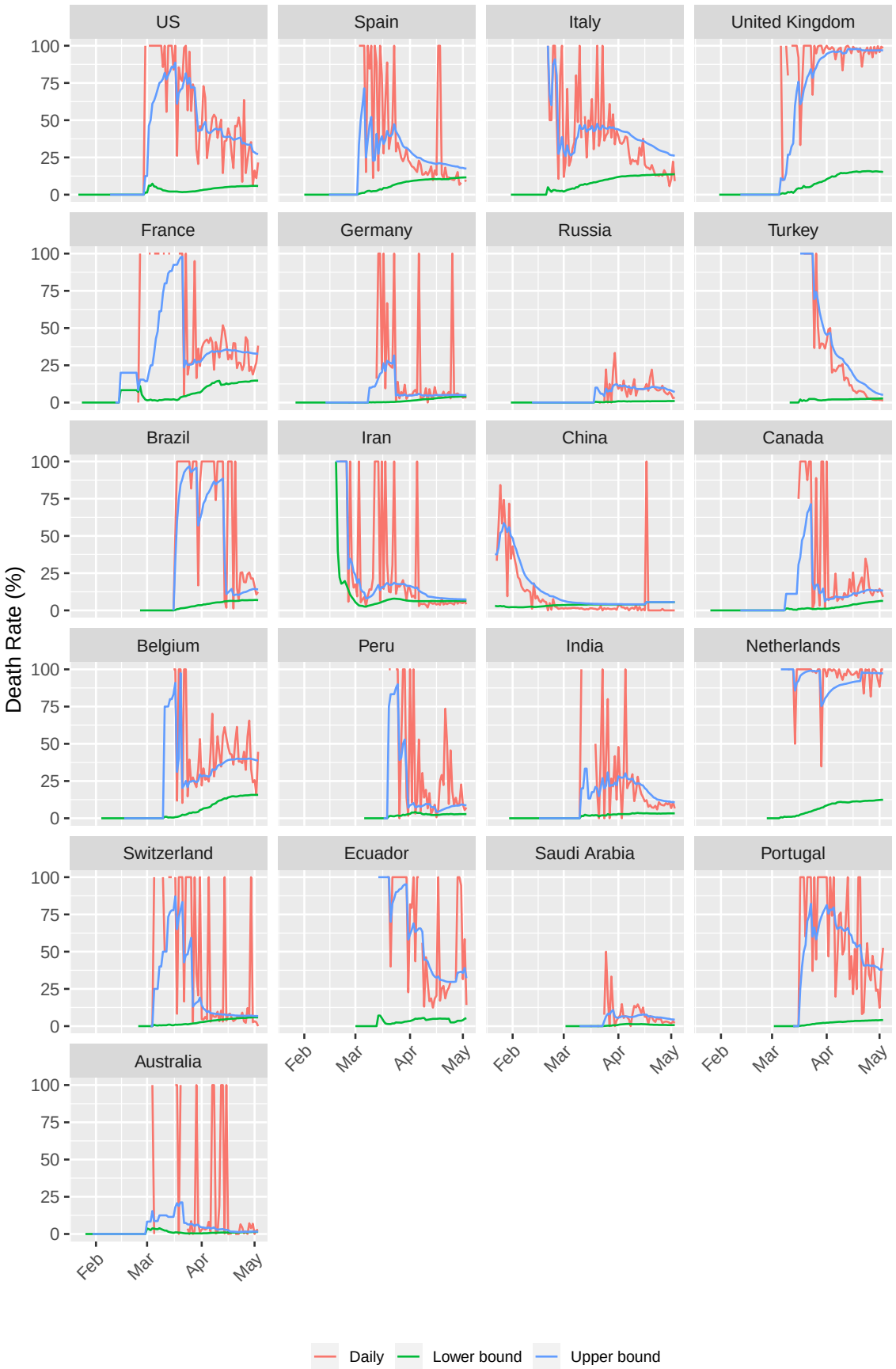


Figure 12: Death Rates

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 - 03 May 2020 UTC

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	Belgium	49,906	389	29,753	12,309	7,844	79	15.7%
2	United Kingdom	187,842	4,342	158,421	901	28,520	315	15.2%
3	France	168,925	407	93,140	50,885	24,900	137	14.7%
4	Italy	210,717	1,389	100,179	81,654	28,884	174	13.7%
5	Netherlands	40,769	335	35,559	138	5,072	69	12.4%
6	Sweden	22,317	235	18,633	1,005	2,679	10	12.0%
7	Spain	217,466	884	73,300	118,902	25,264	164	11.6%
8	Hungary	2,998	56	2,029	629	340	5	11.3%
9	Algeria	4,474	179	2,075	1,936	463	4	10.3%
10	Mexico	23,471	1,383	7,870	13,447	2,154	93	9.2%
11	Indonesia	11,192	349	8,471	1,876	845	14	7.6%
12	Brazil	101,826	4,726	51,784	42,991	7,051	290	6.9%
13	Egypt	6,465	272	4,474	1,562	429	14	6.6%
14	Philippines	9,223	295	7,402	1,214	607	4	6.6%
15	Iran	97,424	976	12,799	78,422	6,203	47	6.4%
16	Canada	60,504	2,578	31,788	24,921	3,795	111	6.3%
17	Ireland	21,506	330	6,817	13,386	1,303	17	6.1%
18	Romania	13,163	431	7,504	4,869	790	19	6.0%
19	Switzerland	29,905	88	3,643	24,500	1,762	0	5.9%
20	US	1,158,040	25,501	910,206	180,152	67,682	1,313	5.8%

6 Conclusions

As of 03 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 7.1% and 18%, but it is likely to change dramatically with the breakout in many countries, such as European countries.

Appendix A. Processed Data

Blow is the processed data for this analysis.

Appendix A.1 COVID-19 Cases Worldwide

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

```
  select(c(date, confirmed, deaths, recovered, current.confirmed,
           new.confirmed, new.deaths, new.recovered, rate.lower, rate.upper, rate.daily))
## output as a table
data.world %>%
  mutate(rate.upper = rate.upper %>% format(nsmall=1) %>% paste0('\\%'),
         rate.lower = rate.lower %>% format(nsmall=1) %>% paste0('\\%'),
         rate.daily = rate.daily %>% format(nsmall=1) %>% paste0('\\%')) %>%
  kable('latex', escape=F, booktabs=T, longtable=T,
        caption='Cases in the Whole World',
        format.args=list(big.mark=','),
        align=c('l', rep('r', 10))) %>%
  kable_styling(font_size=4, latex_options=c('striped', 'hold_position', 'repeat_header'))
```

Table 5: Cases in the Whole World

date	confirmed	deaths	recovered	current.confirmed	new.confirmed	new.deaths	new.recovered	rate.lower	rate.upper	rate.daily
2020-05-03	3,506,729	247,470	1,125,236	2,134,023	79,386	3,662	32,124	7.1%	18.0%	10.2%
2020-05-02	3,427,343	243,808	1,093,112	2,090,423	83,566	5,189	40,697	7.1%	18.2%	11.3%
2020-05-01	3,343,777	238,619	1,052,415	2,052,743	86,924	5,262	38,544	7.1%	18.5%	12.0%
2020-04-30	3,256,853	233,357	1,013,871	2,009,625	84,566	5,692	65,446	7.2%	18.7%	8.0%
2020-04-29	3,172,287	227,665	948,425	1,996,197	75,097	6,866	41,482	7.2%	19.4%	14.2%
2020-04-28	3,097,190	220,799	906,943	1,969,448	73,468	6,315	33,266	7.1%	19.6%	16.0%
2020-04-27	3,023,722	214,484	873,677	1,935,561	68,689	4,584	27,692	7.1%	19.7%	14.2%
2020-04-26	2,955,033	209,900	845,985	1,899,148	73,893	3,713	28,571	7.1%	19.9%	11.5%
2020-04-25	2,881,140	206,187	817,414	1,857,539	85,265	6,188	27,818	7.2%	20.1%	18.2%
2020-04-24	2,795,875	199,999	789,596	1,806,280	87,328	6,332	50,598	7.2%	20.2%	11.1%
2020-04-23	2,708,547	193,667	738,998	1,775,882	83,806	6,753	28,965	7.2%	20.8%	18.9%
2020-04-22	2,624,741	186,914	710,033	1,727,794	75,566	6,677	30,139	7.1%	20.8%	18.1%
2020-04-21	2,549,175	180,237	679,894	1,689,044	77,328	7,139	34,586	7.1%	21.0%	17.1%
2020-04-20	2,471,847	173,098	645,308	1,653,441	70,953	5,325	22,001	7.0%	21.2%	19.5%
2020-04-19	2,400,894	167,773	623,307	1,609,814	83,555	4,559	31,592	7.0%	21.2%	12.6%
2020-04-18	2,317,339	163,214	591,715	1,562,410	77,616	6,410	23,958	7.0%	21.6%	21.1%
2020-04-17	2,239,723	156,804	567,757	1,515,162	87,851	8,858	26,165	7.0%	21.6%	25.3%
2020-04-16	2,151,872	147,946	541,592	1,462,334	96,366	7,288	31,088	6.9%	21.5%	19.0%
2020-04-15	2,055,506	140,658	510,504	1,404,344	79,925	8,219	36,536	6.8%	21.6%	18.4%
2020-04-14	1,975,581	132,439	473,968	1,369,174	70,389	6,878	25,313	6.7%	21.8%	21.4%
2020-04-13	1,905,192	125,561	448,655	1,330,976	70,028	5,708	26,933	6.6%	21.9%	17.5%
2020-04-12	1,835,164	119,853	421,722	1,293,589	99,139	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%
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%

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
2020-02-27	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 (03 May 2020 UTC)

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	World	3,506,729	79,386	2,134,023	1,125,236	247,470	3,662	**7.1%**
2	US	1,158,040	25,501	910,206	180,152	67,682	1,313	**5.8%**
3	Spain	217,466	884	73,300	118,902	25,264	164	**11.6%**
4	Italy	210,717	1,389	100,179	81,654	28,884	174	**13.7%**
5	United Kingdom	187,842	4,342	158,421	901	28,520	315	**15.2%**
6	France	168,925	407	93,140	50,885	24,900	137	**14.7%**
7	Germany	165,664	697	28,198	130,600	6,866	54	4.1%
8	Russia	134,687	10,633	116,768	16,639	1,280	58	1.0%
9	Turkey	126,045	1,670	59,497	63,151	3,397	61	2.7%
10	Brazil	101,826	4,726	51,784	42,991	7,051	290	**6.9%**
11	Iran	97,424	976	12,799	78,422	6,203	47	**6.4%**
12	China	83,964	5	643	78,684	4,637	0	**5.5%**
13	Canada	60,504	2,578	31,788	24,921	3,795	111	**6.3%**
14	Belgium	49,906	389	29,753	12,309	7,844	79	**15.7%**
15	Peru	45,928	3,394	31,092	13,550	1,286	86	2.8%

Table 6: Cases by Country (03 May 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
16	India	42,505	2,806	29,339	11,775	1,391	68	3.3%
17	Netherlands	40,769	335	35,559	138	5,072	69	**12.4%**
18	Switzerland	29,905	88	3,643	24,500	1,762	0	**5.9%**
19	Ecuador	29,538	2,074	24,674	3,300	1,564	193	**5.3%**
20	Saudi Arabia	27,011	1,552	22,693	4,134	184	8	0.7%
21	Portugal	25,282	92	22,550	1,689	1,043	20	4.1%
22	Mexico	23,471	1,383	7,870	13,447	2,154	93	**9.2%**
23	Sweden	22,317	235	18,633	1,005	2,679	10	**12.0%**
24	Ireland	21,506	330	6,817	13,386	1,303	17	**6.1%**
25	Pakistan	20,084	981	14,513	5,114	457	17	2.3%
26	Chile	19,663	1,228	9,362	10,041	260	13	1.3%
27	Singapore	18,205	657	16,779	1,408	18	1	0.1%
28	Belarus	16,705	877	13,410	3,196	99	2	0.6%
29	Israel	16,208	23	6,227	9,749	232	3	1.4%
30	Austria	15,597	39	1,771	13,228	598	2	3.8%
31	Qatar	15,551	679	13,875	1,664	12	0	0.1%
32	Japan	14,877	306	10,409	3,981	487	13	3.3%
33	United Arab Emirates	14,163	564	11,274	2,763	126	7	0.9%
34	Poland	13,693	318	9,070	3,945	678	14	**5.0%**
35	Romania	13,163	431	7,504	4,869	790	19	**6.0%**
36	Ukraine	11,913	502	10,077	1,548	288	9	2.4%
37	Indonesia	11,192	349	8,471	1,876	845	14	**7.6%**
38	Korea, South	10,801	8	1,332	9,217	252	2	2.3%
39	Denmark	9,721	116	2,054	7,183	484	9	**5.0%**
40	Serbia	9,464	102	7,720	1,551	193	4	2.0%
41	Bangladesh	9,455	665	8,215	1,063	177	2	1.9%
42	Philippines	9,223	295	7,402	1,214	607	4	**6.6%**
43	Dominican Republic	7,954	376	6,015	1,606	333	7	4.2%
44	Norway	7,847	38	7,604	32	211	0	2.7%
45	Czechia	7,781	26	3,946	3,587	248	3	3.2%
46	Colombia	7,668	383	5,606	1,722	340	16	4.4%
47	Panama	7,090	0	6,252	641	197	0	2.8%
48	Australia	6,822	23	878	5,849	95	1	1.4%
49	South Africa	6,783	447	4,103	2,549	131	8	1.9%
50	Egypt	6,465	272	4,474	1,562	429	14	**6.6%**
51	Malaysia	6,298	122	1,780	4,413	105	2	1.7%
52	Finland	5,254	78	2,024	3,000	230	10	4.4%
53	Kuwait	4,983	364	3,169	1,776	38	5	0.8%
54	Morocco	4,903	174	3,291	1,438	174	1	3.5%
55	Argentina	4,783	102	3,183	1,354	246	9	**5.1%**
56	Algeria	4,474	179	2,075	1,936	463	4	**10.3%**
57	Moldova	4,121	69	2,614	1,382	125	1	3.0%
58	Kazakhstan	3,920	63	2,809	1,084	27	2	0.7%
59	Luxembourg	3,824	12	349	3,379	96	4	2.5%
60	Bahrain	3,383	99	1,657	1,718	8	0	0.2%
61	Hungary	2,998	56	2,029	629	340	5	**11.3%**
62	Thailand	2,969	3	176	2,739	54	0	1.8%
63	Afghanistan	2,704	235	2,274	345	85	13	3.1%
64	Greece	2,626	6	1,108	1,374	144	1	**5.5%**
65	Oman	2,568	85	1,806	750	12	0	0.5%
66	Nigeria	2,558	170	2,071	400	87	2	3.4%
67	Armenia	2,386	113	1,316	1,035	35	2	1.5%
68	Iraq	2,296	77	709	1,490	97	2	4.2%
69	Ghana	2,169	0	1,922	229	18	0	0.8%
70	Uzbekistan	2,149	31	820	1,319	10	1	0.5%
71	Croatia	2,096	8	528	1,489	79	2	3.8%
72	Cameroon	2,077	0	1,060	953	64	0	3.1%
73	Azerbaijan	1,932	38	466	1,441	25	0	1.3%
74	Bosnia and Herzegovina	1,857	18	955	825	77	5	4.1%
75	Iceland	1,799	1	72	1,717	10	0	0.6%
76	Estonia	1,700	1	1,386	259	55	2	3.2%
77	Cuba	1,649	38	755	827	67	1	4.1%
78	Bulgaria	1,618	24	1,237	308	73	1	4.5%
79	Bolivia	1,594	365	1,352	166	76	10	4.8%
80	Guinea	1,586	0	1,174	405	7	0	0.4%
81	North Macedonia	1,511	5	482	945	84	2	5.6%
82	New Zealand	1,487	0	191	1,276	20	0	1.3%
83	Slovenia	1,439	0	1,102	241	96	2	6.7%

Table 6: Cases by Country (03 May 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
84	Lithuania	1,410	4	729	635	46	0	3.3%
85	Slovakia	1,408	1	765	619	24	0	1.7%
86	Cote d'Ivoire	1,398	36	728	653	17	2	1.2%
87	Senegal	1,182	67	801	372	9	0	0.8%
88	Djibouti	1,112	0	424	686	2	0	0.2%
89	Honduras	1,055	45	855	118	82	6	7.8%
90	Tunisia	1,013	4	643	328	42	0	4.1%
91	Latvia	879	8	515	348	16	0	1.8%
92	Cyprus	872	8	561	296	15	0	1.7%
93	Kosovo	851	28	448	381	22	0	2.6%
94	Albania	795	6	233	531	31	0	3.9%
95	Kyrgyzstan	795	26	221	564	10	2	1.3%
96	Niger	750	14	196	518	36	1	4.8%
97	Andorra	748	1	210	493	45	1	6.0%
98	Costa Rica	739	6	347	386	6	0	0.8%
99	Lebanon	737	4	512	200	25	0	3.4%
100	Somalia	722	51	646	44	32	1	4.4%
101	Sri Lanka	718	13	527	184	7	0	1.0%
102	Diamond Princess	712	0	54	645	13	0	1.8%
103	Guatemala	703	15	614	72	17	0	2.4%
104	Congo (Kinshasa)	674	0	566	75	33	0	4.9%
105	Burkina Faso	662	10	77	540	45	1	6.8%
106	Uruguay	655	3	196	442	17	0	2.6%
107	Sudan	592	0	499	52	41	0	6.9%
108	Georgia	589	7	357	223	9	1	1.5%
109	San Marino	582	2	455	86	41	0	7.0%
110	Mali	563	19	323	213	27	1	4.8%
111	Maldives	527	8	508	18	1	0	0.2%
112	El Salvador	490	44	325	154	11	0	2.2%
113	Tanzania	480	0	297	167	16	0	3.3%
114	Malta	477	9	81	392	4	0	0.8%
115	Jamaica	469	6	422	38	9	1	1.9%
116	Kenya	465	30	274	167	24	2	5.2%
117	Jordan	461	1	85	367	9	0	2.0%
118	Taiwan*	436	4	98	332	6	0	1.4%
119	Paraguay	396	26	260	126	10	0	2.5%
120	Venezuela	357	12	189	158	10	0	2.8%
121	West Bank and Gaza	353	0	274	77	2	0	0.6%
122	Gabon	335	0	245	85	5	0	1.5%
123	Mauritius	332	0	7	315	10	0	3.0%
124	Montenegro	322	0	65	249	8	0	2.5%
125	Equatorial Guinea	315	0	305	9	1	0	0.3%
126	Vietnam	271	1	52	219	0	0	0.0%
127	Rwanda	259	4	135	124	0	0	0.0%
128	Guinea-Bissau	257	0	237	19	1	0	0.4%
129	Congo (Brazzaville)	229	0	195	25	9	0	3.9%
130	Sierra Leone	166	11	129	29	8	0	4.8%
131	Cabo Verde	165	13	130	33	2	0	1.2%
132	Liberia	158	4	82	58	18	0	11.4%
133	Burma	155	4	106	43	6	0	3.9%
134	Madagascar	149	14	51	98	0	0	0.0%
135	Brunei	138	0	9	128	1	0	0.7%
136	Ethiopia	135	2	57	75	3	0	2.2%
137	Tajikistan	128	52	126	0	2	0	1.6%
138	Togo	124	1	48	67	9	0	7.3%
139	Zambia	124	5	43	78	3	0	2.4%
140	Cambodia	122	0	2	120	0	0	0.0%
141	Chad	117	0	68	39	10	0	8.5%
142	Trinidad and Tobago	116	0	15	93	8	0	6.9%
143	Eswatini	112	4	99	12	1	0	0.9%
144	Monaco	95	0	13	78	4	0	4.2%
145	Benin	90	0	46	42	2	0	2.2%
146	Uganda	89	1	37	52	0	0	0.0%
147	Haiti	88	3	69	10	9	1	10.2%
148	Bahamas	83	0	48	24	11	0	13.3%
149	Barbados	82	1	31	44	7	0	8.5%
150	Guyana	82	0	51	22	9	0	11.0%
151	Liechtenstein	82	0	26	55	1	0	1.2%

Table 6: Cases by Country (03 May 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
152	Mozambique	80	1	61	19	0	0	0.0%
153	Nepal	75	16	59	16	0	0	0.0%
154	Central African Republic	72	0	62	10	0	0	0.0%
155	Libya	63	0	38	22	3	0	4.8%
156	South Sudan	46	1	46	0	0	0	0.0%
157	Syria	44	0	14	27	3	0	6.8%
158	Eritrea	39	0	13	26	0	0	0.0%
159	Malawi	39	1	27	9	3	0	7.7%
160	Mongolia	39	0	29	10	0	0	0.0%
161	Angola	35	0	22	11	2	0	5.7%
162	Zimbabwe	34	0	25	5	4	0	11.8%
163	Antigua and Barbuda	25	0	7	15	3	0	12.0%
164	Timor-Leste	24	0	8	16	0	0	0.0%
165	Botswana	23	0	14	8	1	0	4.3%
166	Grenada	21	0	8	13	0	0	0.0%
167	Laos	19	0	10	9	0	0	0.0%
168	Belize	18	0	3	13	2	0	11.1%
169	Fiji	18	0	4	14	0	0	0.0%
170	Saint Lucia	18	1	3	15	0	0	0.0%
171	Gambia	17	0	7	9	1	0	5.9%
172	Dominica	16	0	3	13	0	0	0.0%
173	Namibia	16	0	8	8	0	0	0.0%
174	Saint Vincent and the Grenadines	16	0	8	8	0	0	0.0%
175	Sao Tome and Principe	16	0	11	4	1	0	6.2%
176	Burundi	15	0	7	7	1	0	6.7%
177	Nicaragua	15	1	3	7	5	2	33.3%
178	Saint Kitts and Nevis	15	0	7	8	0	0	0.0%
179	Holy See	11	0	9	2	0	0	0.0%
180	Seychelles	11	0	5	6	0	0	0.0%
181	Suriname	10	0	0	9	1	0	10.0%
182	Yemen	10	0	7	1	2	0	20.0%
183	MS Zaandam	9	0	7	0	2	0	22.2%
184	Mauritania	8	0	1	6	1	0	12.5%
185	Papua New Guinea	8	0	8	0	0	0	0.0%
186	Bhutan	7	0	2	5	0	0	0.0%
187	Western Sahara	6	0	1	5	0	0	0.0%
188	Comoros	3	0	3	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!