COVID-19 Data Analysis with R - Worldwide*

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

29 April 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] 264 102
```

Each dataset has 264 rows, corresponding to country/region/province/state. It has 102 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" "2020-04-28"
```

```
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 28 Apr 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 185 countries with confirmed COVID-19 cases, as of 28 Apr 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
```

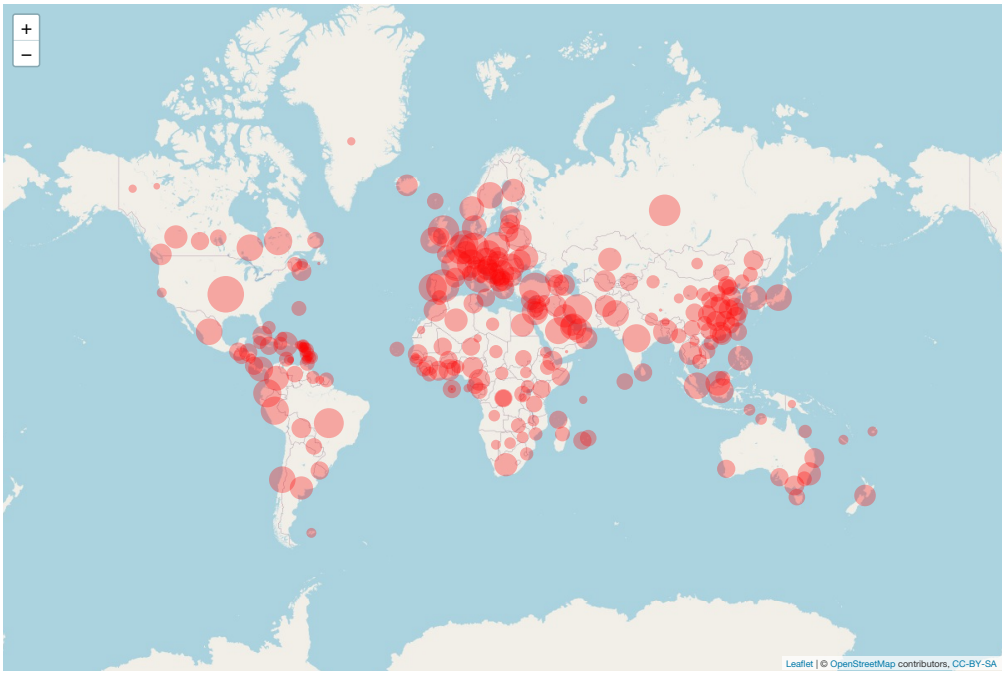


Figure 1: World Map

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)
```

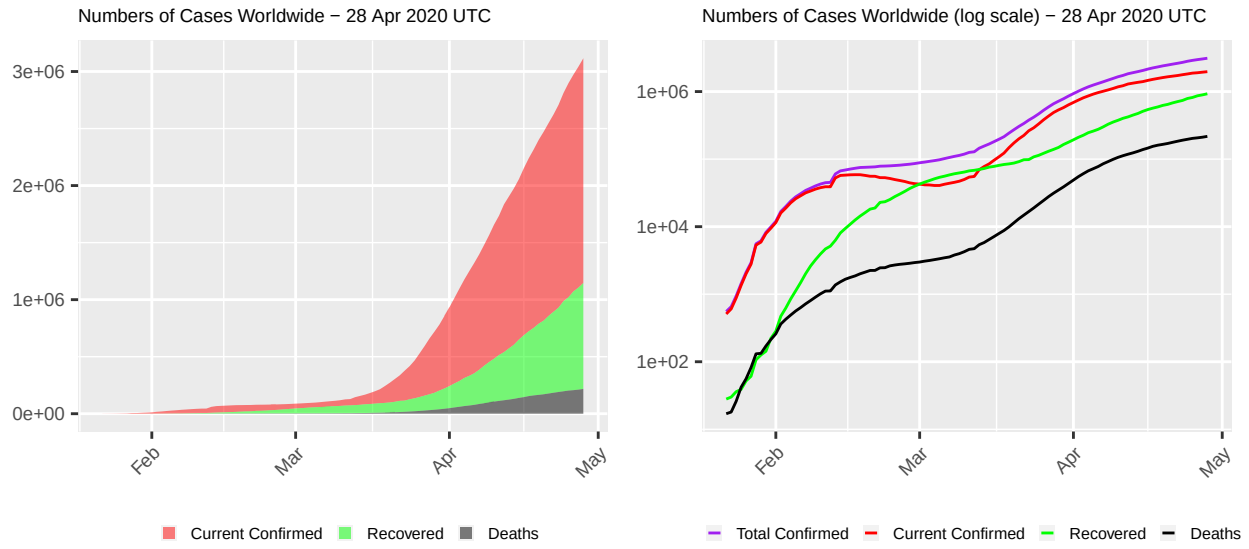


Figure 2: COVID-19 Cases Worldwide

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)
```

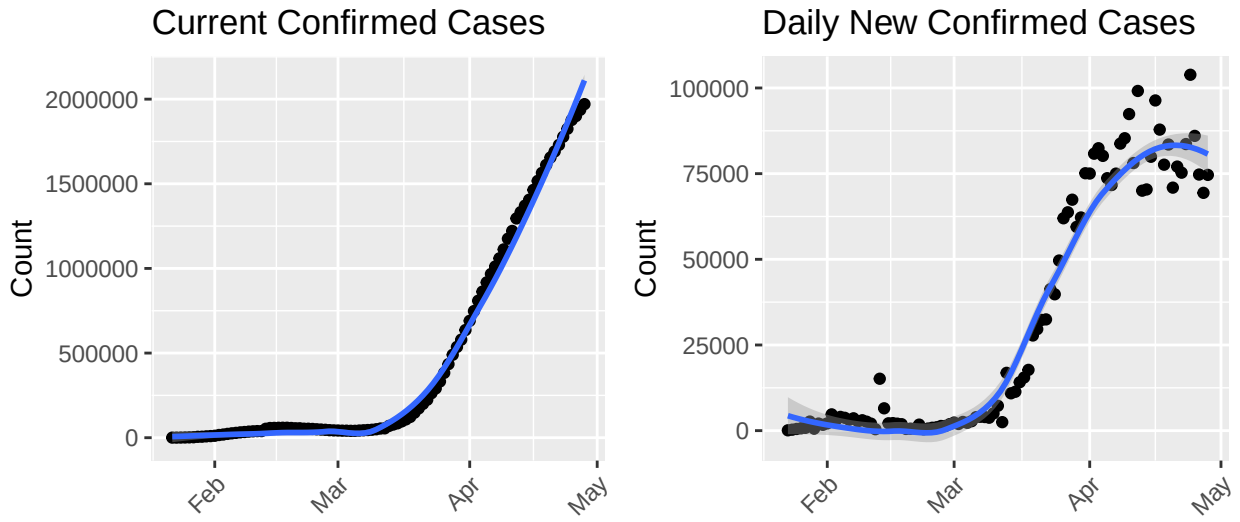


Figure 3: Current Confirmed Cases

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)
```

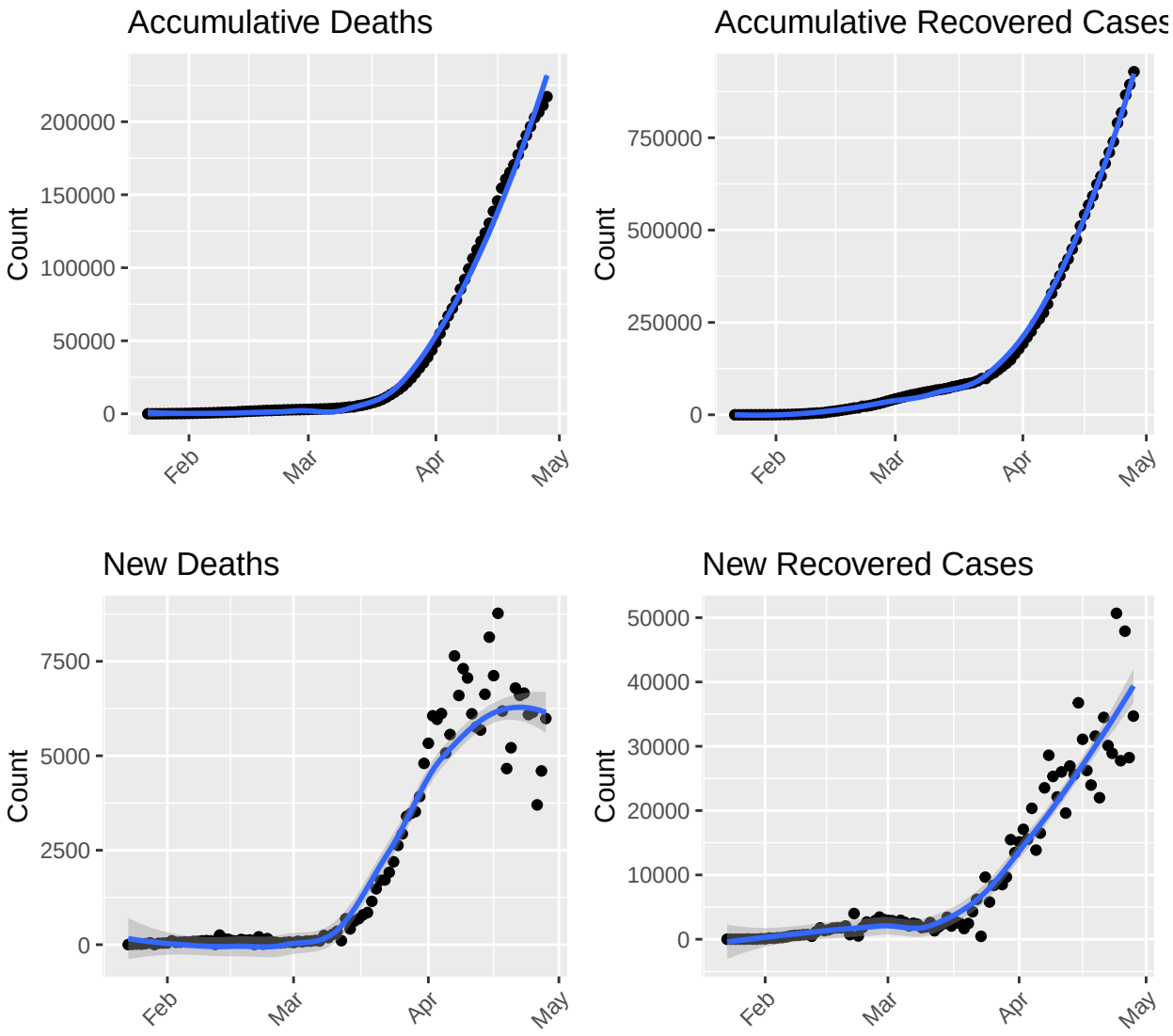


Figure 4: Deaths and Recovered Cases

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 28 Apr 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 28 Apr 2020 UTC UTC, it will be between 7% and 19%.

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

```
## 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 - 28 Apr 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	3,116,398	217,153	7.0%	74,634	5,986	1,970,587
2	US	1,012,582	58,355	5.8%	24,385	2,096	838,291
3	Spain	232,128	23,822	10.3%	2,706	301	84,403
4	Italy	201,505	27,359	13.6%	2,091	382	105,205
5	France	169,053	23,694	14.0%	3,090	367	97,584
6	United Kingdom	162,350	21,745	13.4%	4,002	588	139,792
7	Germany	159,912	6,314	3.9%	1,154	188	36,198
8	Turkey	114,653	2,992	2.6%	2,392	92	72,852
9	Russia	93,558	867	0.9%	6,411	73	84,235
10	Iran	92,584	5,877	6.3%	1,112	71	14,268
11	China	83,940	4,637	5.5%	22	0	881
12	Brazil	73,235	5,083	6.9%	5,789	480	35,608
13	Canada	51,150	2,983	5.8%	1,534	142	28,936
14	Belgium	47,334	7,331	15.5%	647	124	29,060
15	Netherlands	38,612	4,582	11.9%	172	48	33,913
16	India	31,324	1,008	3.2%	1,873	69	22,569
17	Peru	31,190	854	2.7%	2,491	72	21,157
18	Switzerland	29,264	1,699	5.8%	100	34	4,965
19	Portugal	24,322	948	3.9%	295	20	21,985
20	Ecuador	24,258	871	3.6%	1,018	208	21,830
21	Saudi Arabia	20,077	152	0.8%	1,266	8	17,141
22	Others	423,367	15,980	3.8%	12,606	623	259,714

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

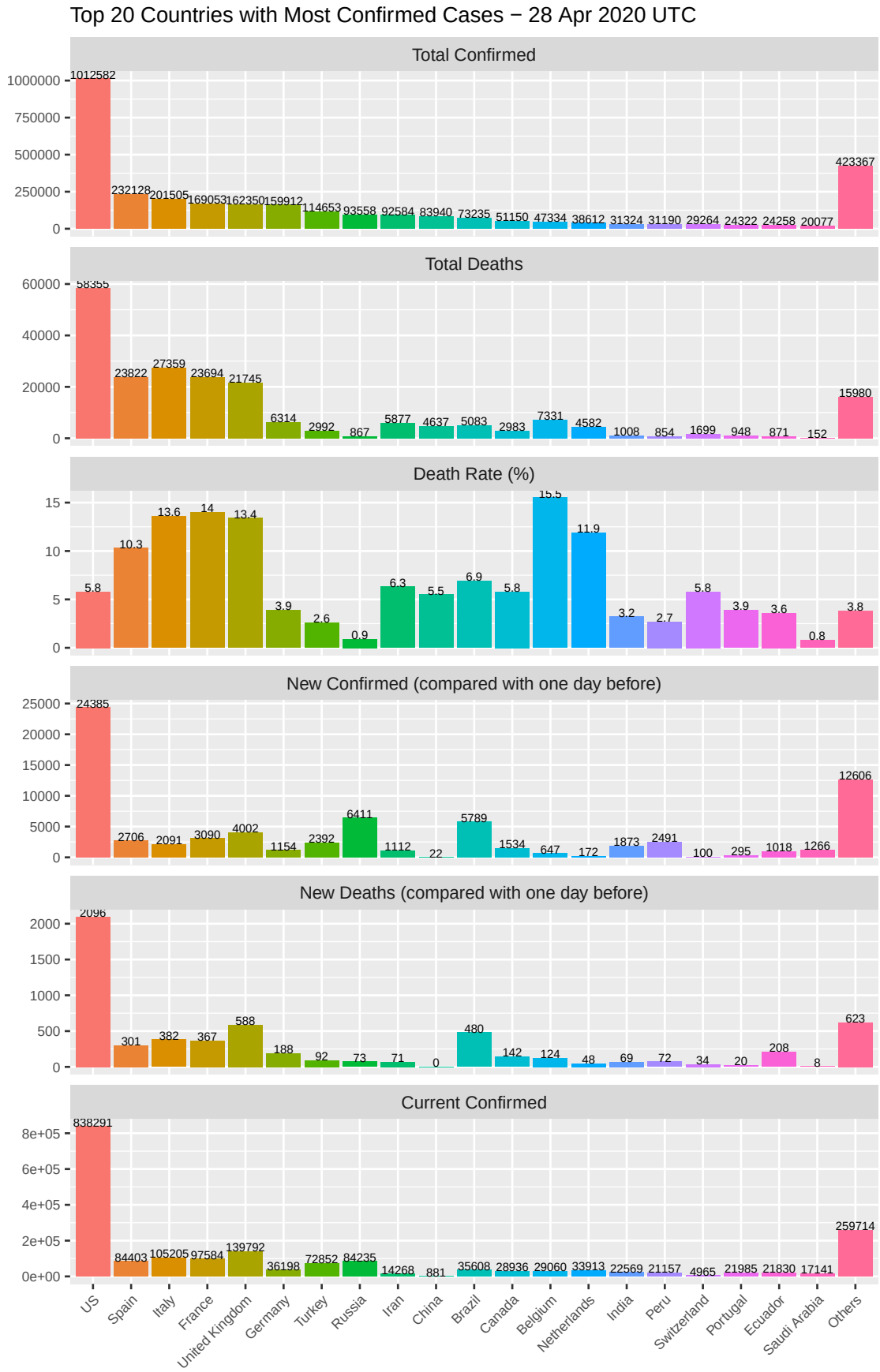
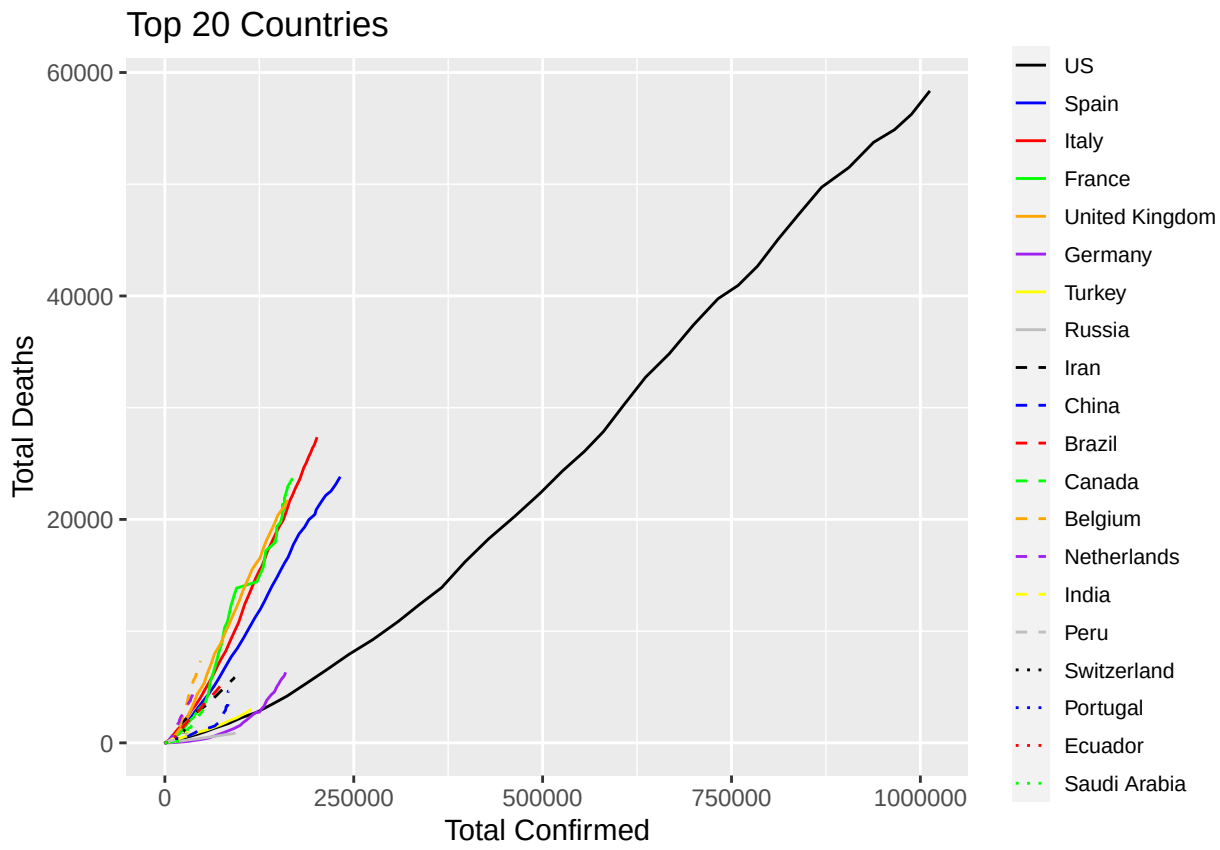


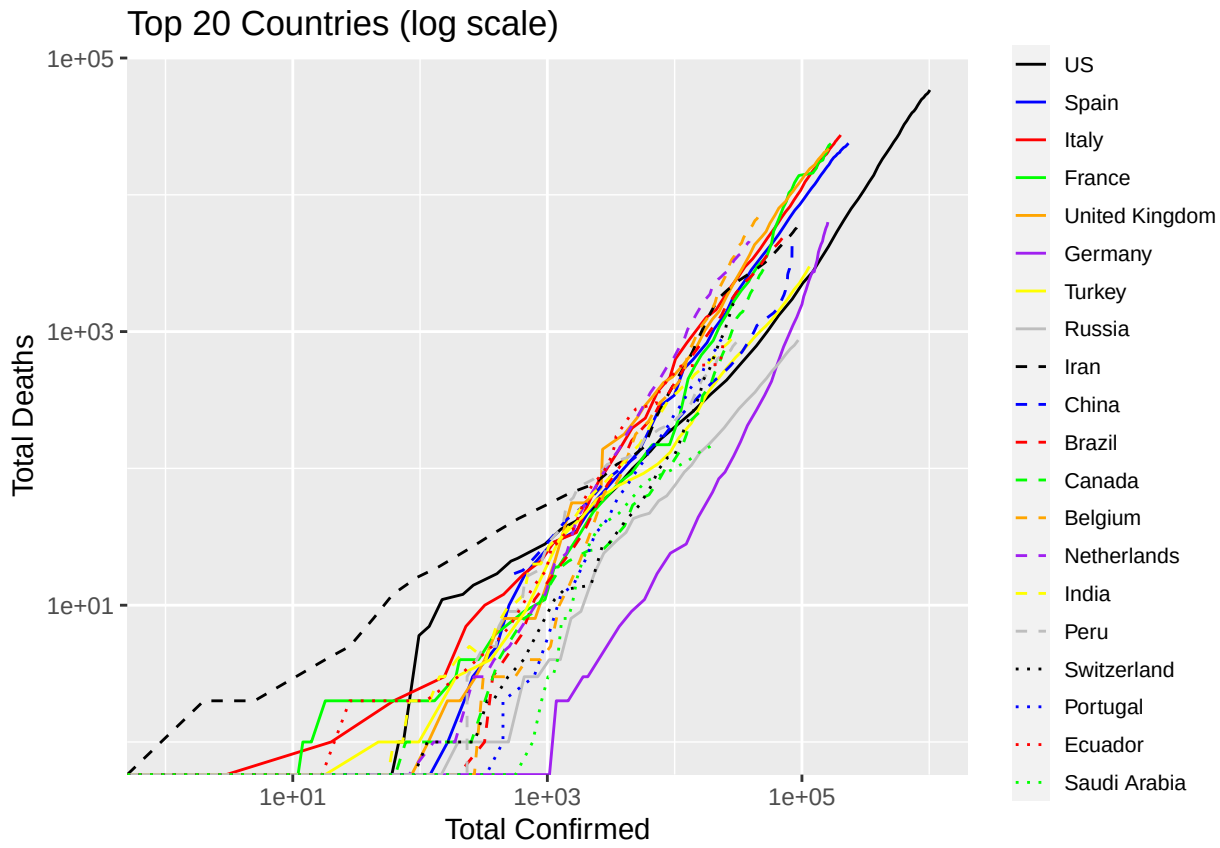
Figure 6: Top 20 Countries with Most Confirmed Cases

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 28 Apr 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)
```

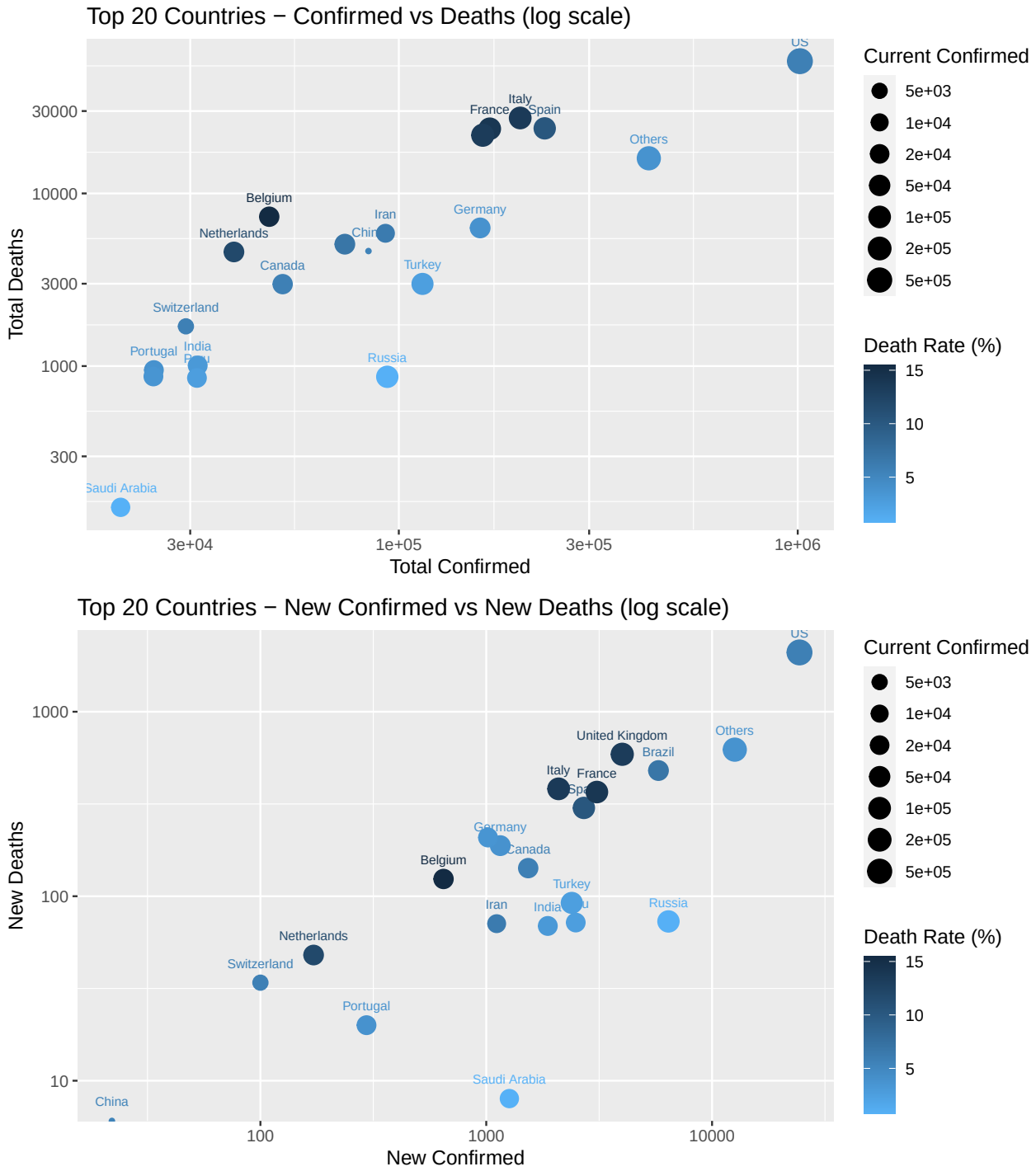


Figure 7: Top 20 Countries

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)
```

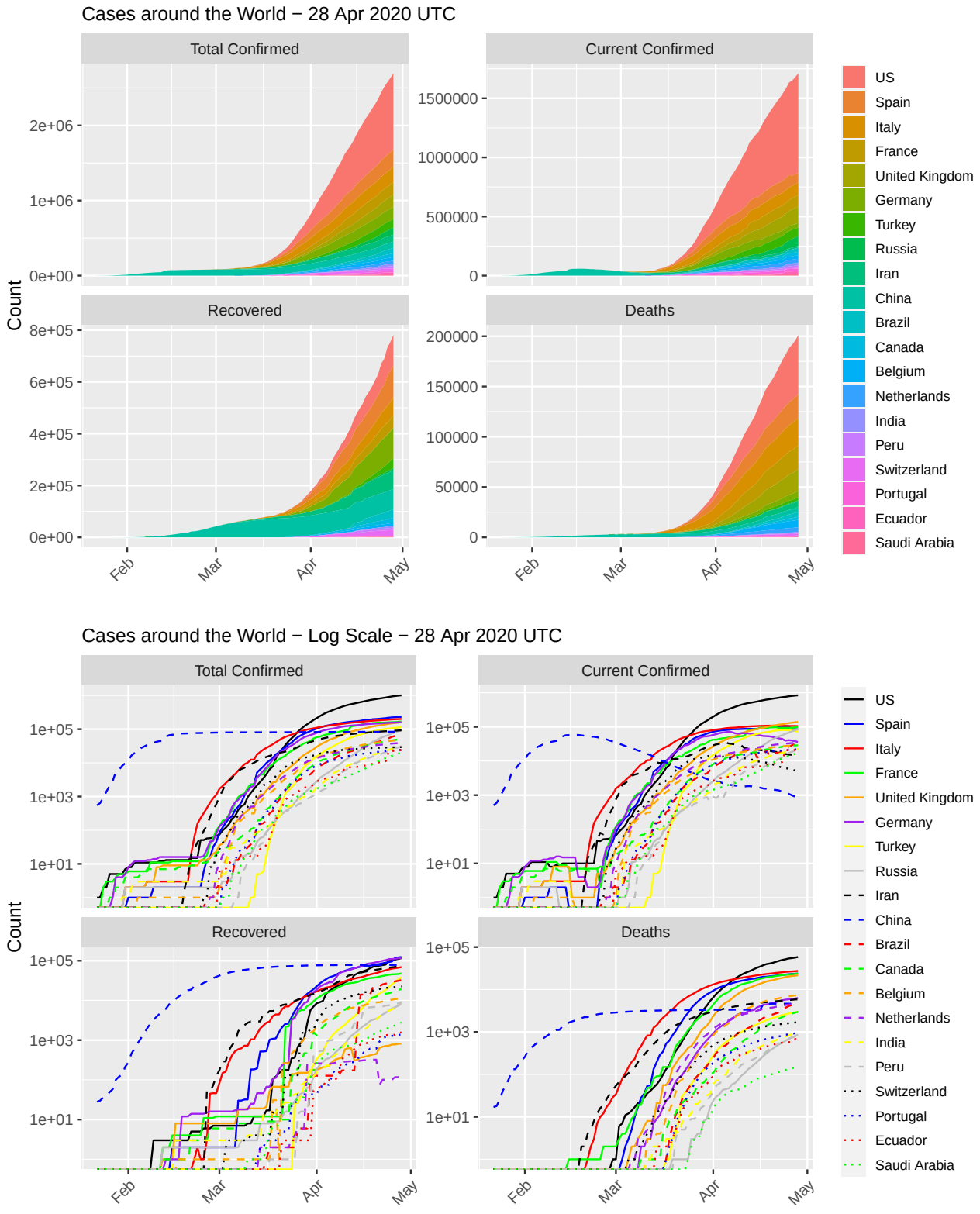


Figure 8: Cases around the World

```
## 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))
```

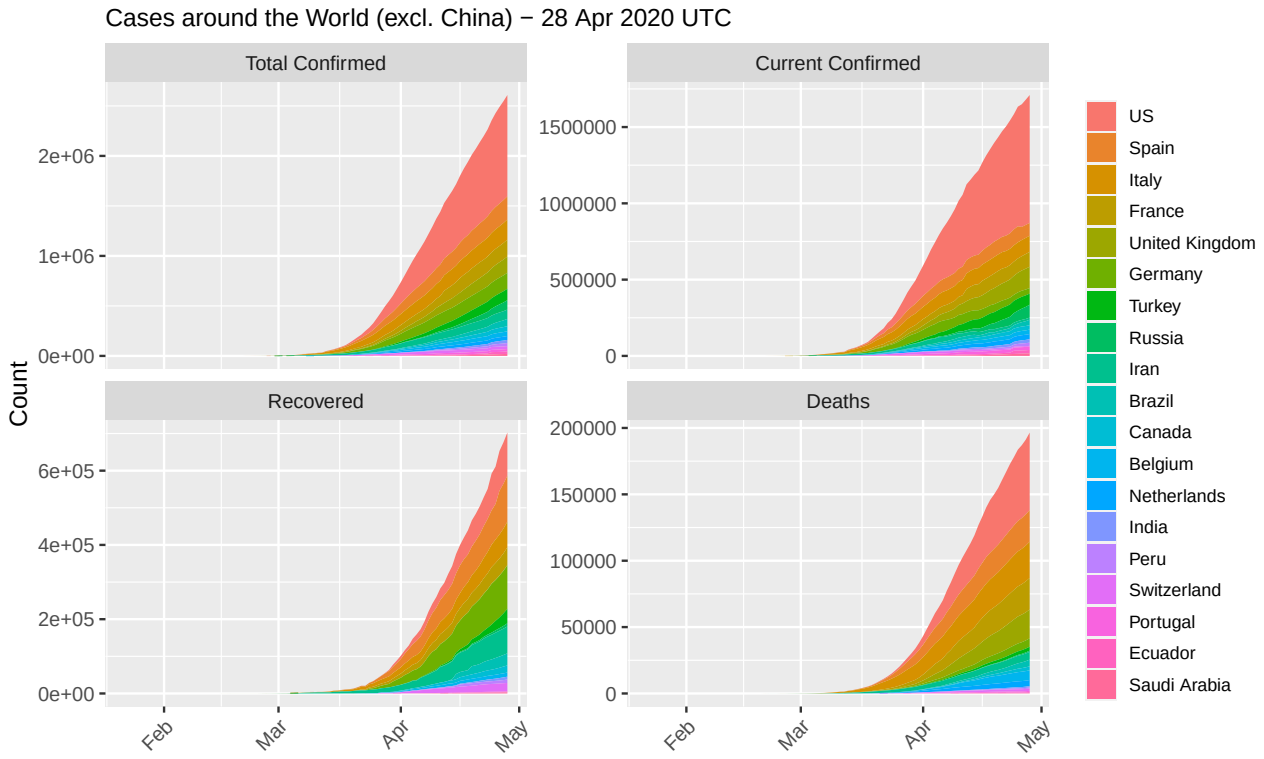


Figure 9: Cases around the World (excl. China)

```
## 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')
```

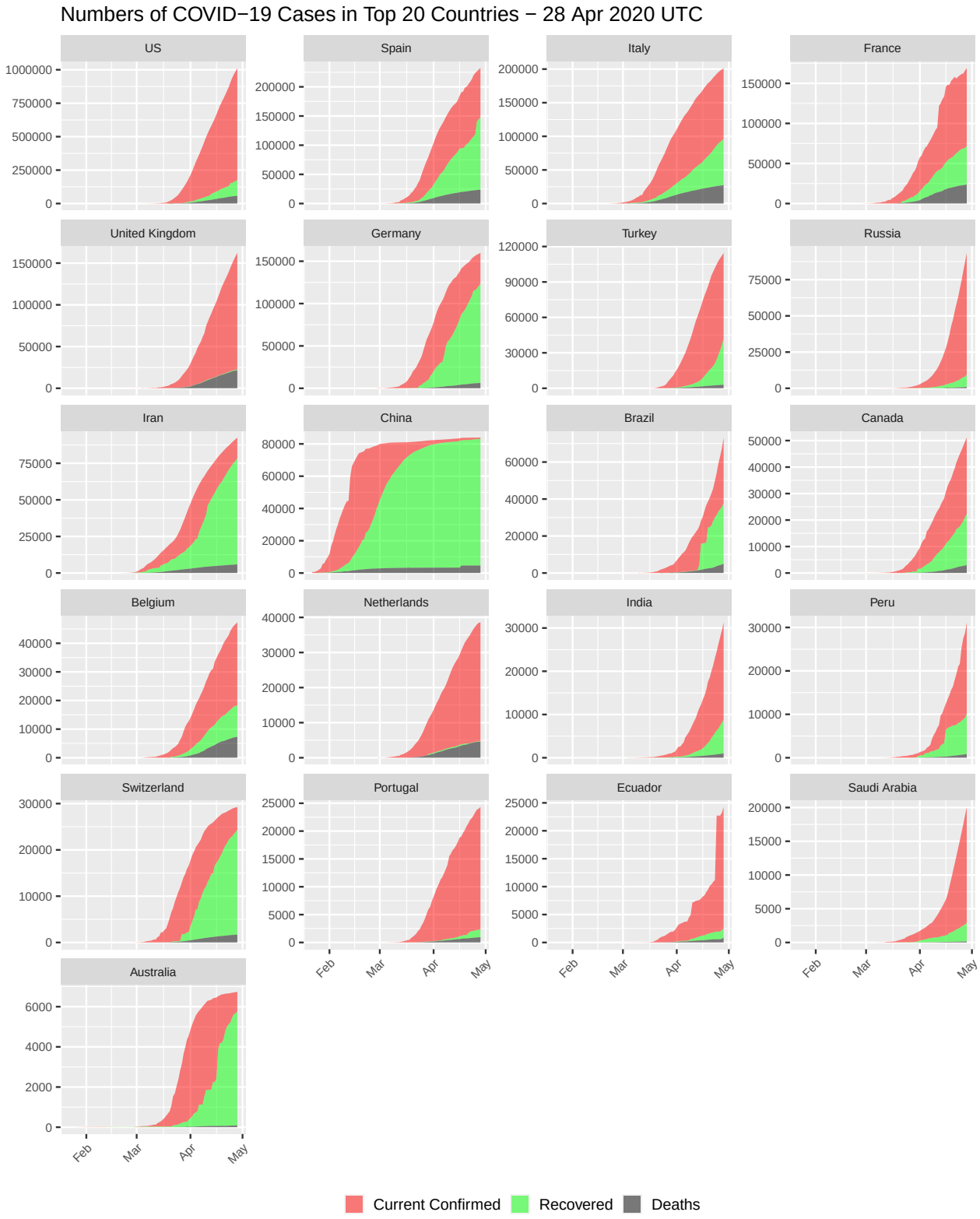


Figure 10: COVID-19 Cases in Top 20 Countries. Ordered descendingly by number of confirmed cases.

```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')
```

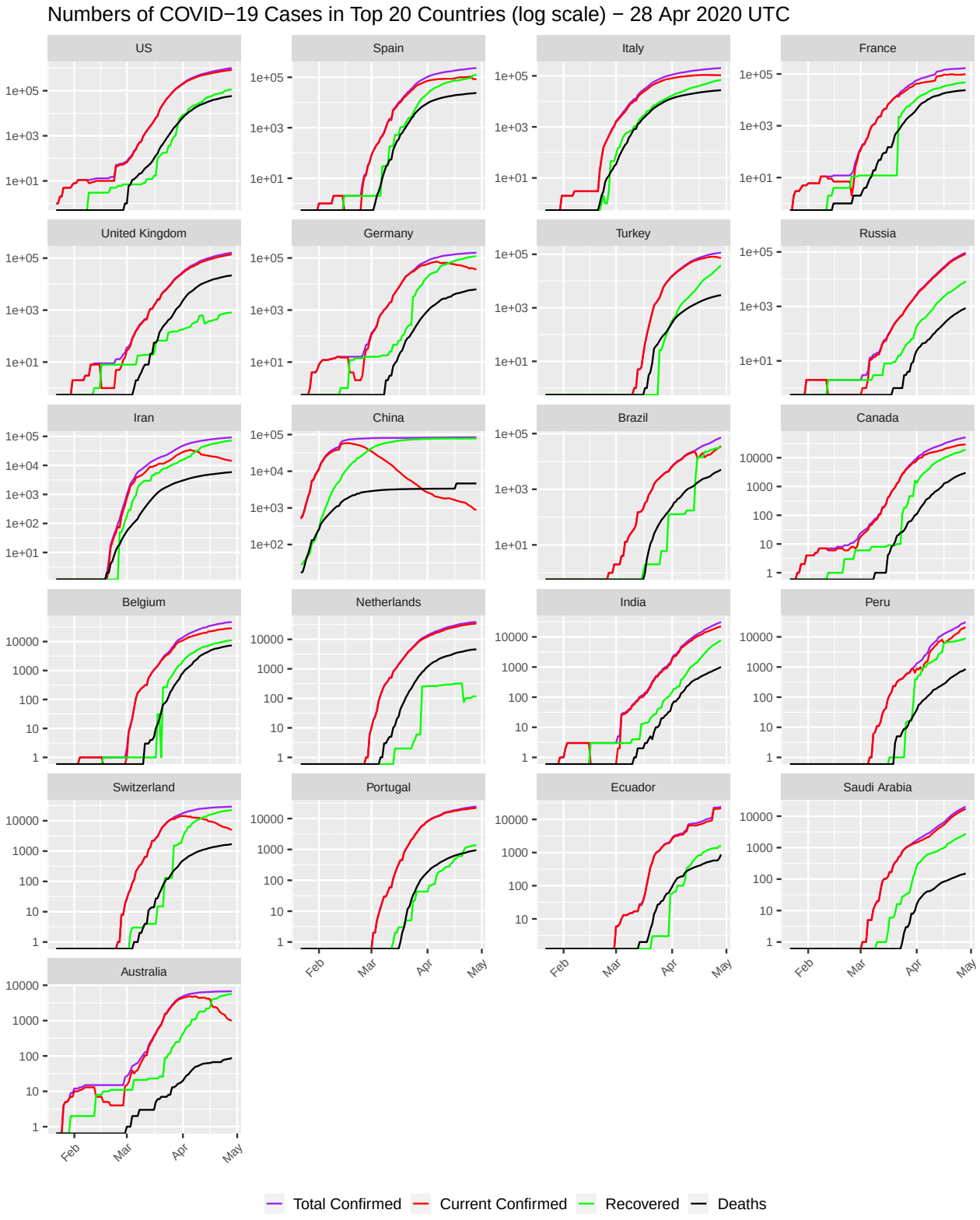


Figure 11: COVID-19 Cases Top 20 Countries (log scale). Ordered descendingly by number of confirmed cases.

```
## 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)
```

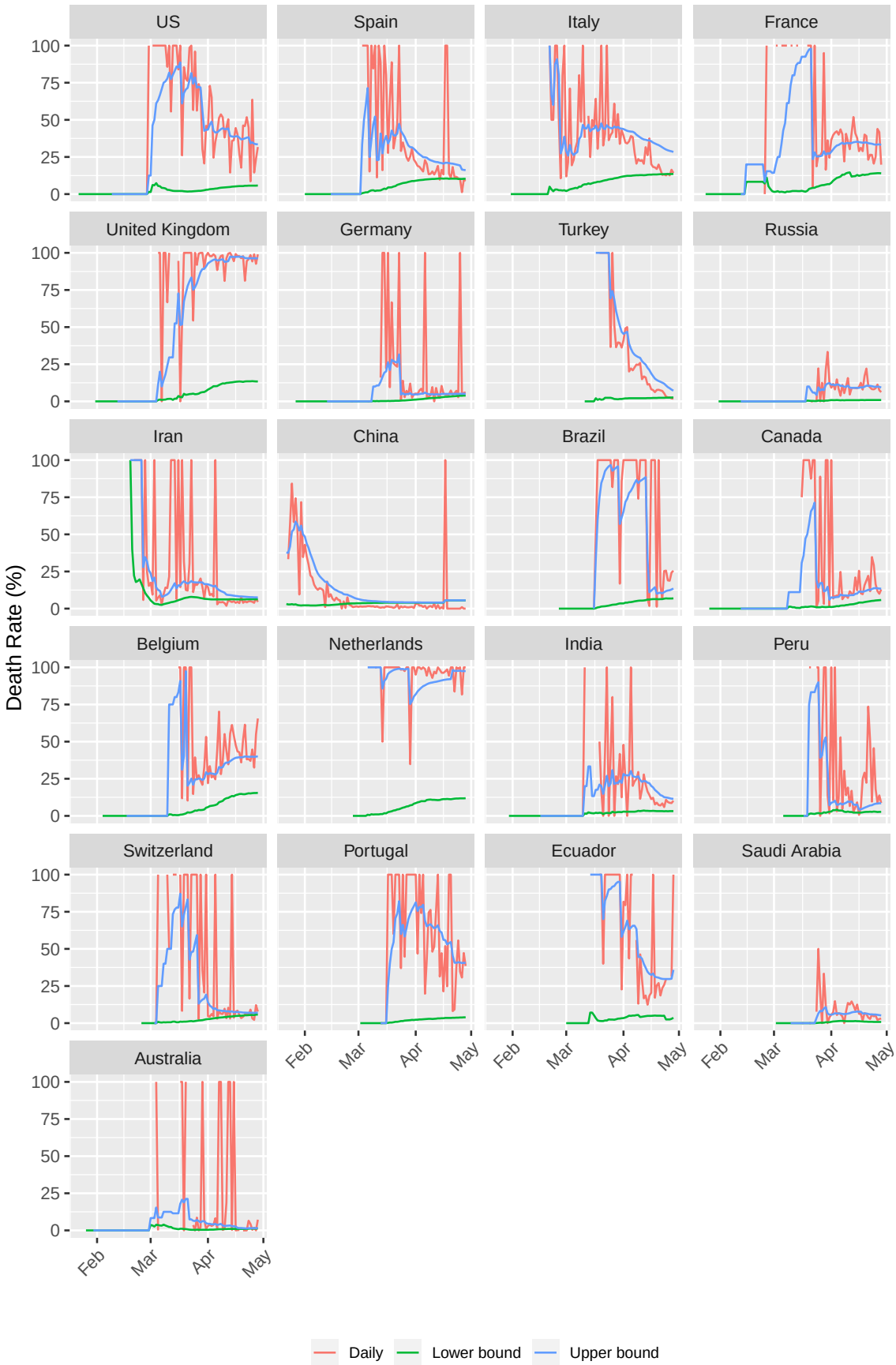


Figure 12: Death Rates

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 - 28 Apr 2020 UTC

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	Belgium	47,334	647	29,060	10,943	7,331	124	15.5%
2	France	169,053	3,090	97,584	47,775	23,694	367	14.0%
3	Italy	201,505	2,091	105,205	68,941	27,359	382	13.6%
4	United Kingdom	162,350	4,002	139,792	813	21,745	588	13.4%
5	Algeria	3,649	132	1,561	1,651	437	5	12.0%
6	Sweden	19,621	695	16,261	1,005	2,355	81	12.0%
7	Netherlands	38,612	172	33,913	117	4,582	48	11.9%
8	Hungary	2,649	66	1,842	516	291	11	11.0%
9	Spain	232,128	2,706	84,403	123,903	23,822	301	10.3%
10	Mexico	16,752	1,223	3,760	11,423	1,569	135	9.4%
11	Indonesia	9,511	415	7,484	1,254	773	8	8.1%
12	Egypt	5,042	260	3,379	1,304	359	22	7.1%
13	Brazil	73,235	5,789	35,608	32,544	5,083	480	6.9%
14	Philippines	7,958	181	6,453	975	530	19	6.7%
15	Iran	92,584	1,112	14,268	72,439	5,877	71	6.3%
16	Canada	51,150	1,534	28,936	19,231	2,983	142	5.8%
17	Ireland	19,877	229	9,485	9,233	1,159	57	5.8%
18	Switzerland	29,264	100	4,965	22,600	1,699	34	5.8%
19	US	1,012,582	24,385	838,291	115,936	58,355	2,096	5.8%
20	Romania	11,616	277	7,549	3,404	663	22	5.7%

6 Conclusions

As of 28 Apr 2020 UTC, there are 185 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 19%, 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
2020-04-28	3,116,398	217,153	928,658	1,970,587	74,634	5,986	34,691	7.0%	19.0%	14.7%
2020-04-27	3,041,764	211,167	893,967	1,936,630	69,401	4,599	28,234	6.9%	19.1%	14.0%
2020-04-26	2,972,363	206,568	865,733	1,900,062	74,739	3,700	47,895	6.9%	19.3%	7.2%
2020-04-25	2,897,624	202,868	817,838	1,876,918	86,021	6,150	27,757	7.0%	19.9%	18.1%
2020-04-24	2,811,603	196,718	790,081	1,824,804	103,861	6,090	50,672	7.0%	19.9%	10.7%
2020-04-23	2,707,742	190,628	739,409	1,777,705	83,635	6,658	28,907	7.0%	20.5%	18.7%
2020-04-22	2,624,107	183,970	710,502	1,729,635	75,286	6,599	30,112	7.0%	20.6%	18.0%
2020-04-21	2,548,821	177,371	680,390	1,691,060	77,062	6,793	34,485	7.0%	20.7%	16.5%
2020-04-20	2,471,759	170,578	645,905	1,655,276	70,916	5,213	22,002	6.9%	20.9%	19.2%
2020-04-19	2,400,843	165,365	623,903	1,611,575	83,504	4,661	31,584	6.9%	21.0%	12.9%
2020-04-18	2,317,339	160,704	592,319	1,564,316	77,616	6,183	23,976	6.9%	21.3%	20.5%
2020-04-17	2,239,723	154,521	568,343	1,516,859	87,851	8,771	26,236	6.9%	21.4%	25.1%
2020-04-16	2,151,872	145,750	542,107	1,464,015	96,366	7,120	31,088	6.8%	21.2%	18.6%
2020-04-15	2,055,506	138,630	511,019	1,405,857	79,925	8,140	36,758	6.7%	21.3%	18.1%
2020-04-14	1,975,581	130,490	474,261	1,370,830	70,389	6,629	25,606	6.6%	21.6%	20.6%
2020-04-13	1,905,192	123,861	448,655	1,332,676	70,028	5,681	26,933	6.5%	21.6%	17.4%
2020-04-12	1,835,164	118,180	421,722	1,295,262	99,139	5,758	19,612	6.4%	21.9%	22.7%
2020-04-11	1,736,025	112,422	402,110	1,221,493	78,096	6,111	26,014	6.5%	21.8%	19.0%
2020-04-10	1,657,929	106,311	376,096	1,175,522	92,391	7,059	22,121	6.4%	22.0%	24.2%
2020-04-09	1,565,538	99,252	353,975	1,112,311	85,338	7,306	25,314	6.3%	21.9%	22.4%
2020-04-08	1,480,200	91,946	328,661	1,059,593	83,762	6,596	28,607	6.2%	21.9%	18.7%
2020-04-07	1,396,438	85,350	300,054	1,011,034	75,011	7,643	23,539	6.1%	22.1%	24.5%
2020-04-06	1,321,427	77,707	276,515	967,205	71,690	5,562	16,503	5.9%	21.9%	25.2%
2020-04-05	1,249,737	72,145	260,012	917,580	73,678	5,072	13,860	5.8%	21.7%	26.8%
2020-04-04	1,176,059	67,073	246,152	862,834	80,183	6,117	20,356	5.7%	21.4%	23.1%
2020-04-03	1,095,876	60,956	225,796	809,124	82,418	5,965	15,533	5.6%	21.3%	27.7%
2020-04-02	1,013,458	54,991	210,263	748,204	80,820	6,059	17,086	5.4%	20.7%	26.2%
2020-04-01	932,638	48,932	193,177	690,529	75,030	5,330	15,143	5.2%	20.2%	26.0%
2020-03-31	857,608	43,602	178,034	635,972	75,118	4,798	13,468	5.1%	19.7%	26.3%
2020-03-30	782,490	38,804	164,566	579,120	62,205	3,922	15,484	5.0%	19.1%	20.2%
2020-03-29	720,285	34,882	149,082	536,321	59,461	3,521	9,667	4.8%	19.0%	26.7%
2020-03-28	660,824	31,361	139,415	490,048	67,401	3,475	8,500	4.7%	18.4%	29.0%
2020-03-27	593,423	27,886	130,915	434,622	63,722	3,397	8,765	4.7%	17.6%	27.9%
2020-03-26	529,701	24,489	122,150	383,062	61,978	2,933	8,363	4.6%	16.7%	26.0%
2020-03-25	467,723	21,556	113,787	332,380	49,644	2,628	5,787	4.6%	15.9%	31.2%
2020-03-24	418,079	18,928	108,000	291,151	39,797	2,193	9,649	4.5%	14.9%	18.5%
2020-03-23	378,282	16,735	98,351	263,196	41,264	1,914	452	4.4%	14.5%	80.9%
2020-03-22	337,018	14,821	97,899	224,298	32,463	1,713	6,207	4.4%	13.1%	21.6%
2020-03-21	304,555	13,108	91,692	199,755	32,308	1,703	4,272	4.3%	12.5%	28.5%
2020-03-20	272,247	11,405	87,420	173,422	29,631	1,480	2,445	4.2%	11.5%	37.7%
2020-03-19	242,616	9,925	84,975	147,716	27,770	1,146	1,663	4.1%	10.5%	40.8%
2020-03-18	214,846	8,779	83,312	122,755	17,733	849	2,472	4.1%	9.5%	25.6%
2020-03-17	197,113	7,930	80,840	108,343	15,510	790	2,752	4.0%	8.9%	22.3%
2020-03-16	181,603	7,140	78,088	96,375	14,137	690	2,054	3.9%	8.4%	25.1%
2020-03-15	167,466	6,450	76,034	84,982	11,350	627	3,410	3.9%	7.8%	15.5%
2020-03-14	156,116	5,823	72,624	77,669	10,897	417	2,373	3.7%	7.4%	14.9%
2020-03-13	145,219	5,406	70,251	69,562	16,867	685	1,927	3.7%	7.1%	26.2%
2020-03-12	128,352	4,721	68,324	55,307	2,477	106	1,321	3.7%	6.5%	7.4%
2020-03-11	125,875	4,615	67,003	54,257	7,255	353	2,599	3.7%	6.4%	12.0%
2020-03-10	118,620	4,262	64,404	49,954	5,030	274	1,910	3.6%	6.2%	12.5%
2020-03-09	113,590	3,988	62,494	47,108	3,769	186	1,800	3.5%	6.0%	9.4%
2020-03-08	109,821	3,802	60,694	45,325	3,974	244	2,336	3.5%	5.9%	9.5%
2020-03-07	105,847	3,558	58,358	43,931	4,046	98	2,493	3.4%	5.7%	3.8%
2020-03-06	101,801	3,460	55,865	42,476	3,915	112	2,069	3.4%	5.8%	5.1%
2020-03-05	97,886	3,348	53,796	40,742	2,766	94	2,626	3.4%	5.9%	3.5%
2020-03-04	95,120	3,254	51,170	40,696	2,280	94	2,942	3.4%	6.0%	3.1%
2020-03-03	92,840	3,160	48,228	41,452	2,534	75	2,626	3.4%	6.1%	2.8%
2020-03-02	90,306	3,085	45,602	41,619	1,937	89	2,886	3.4%	6.3%	3.0%
2020-03-01	88,369	2,996	42,716	42,657	2,358	55	2,934	3.4%	6.6%	1.8%
2020-02-29	86,011	2,941	39,782	43,288	1,899	69	3,071	3.4%	6.9%	2.2%
2020-02-28	84,112	2,872	36,711	44,529	1,366	58	3,434	3.4%	7.3%	1.7%
2020-02-27	82,746	2,814	33,277	46,655	1,358	44	2,893	3.4%	7.8%	1.5%
2020-02-26	81,388	2,770	30,384	48,234	982	62	2,479	3.4%	8.4%	2.4%
2020-02-25	80,406	2,708	27,905	49,793	845	79	2,678	3.4%	8.8%	2.9%
2020-02-24	79,561	2,629	25,227	51,705	603	160	1,833	3.3%	9.4%	8.0%
2020-02-23	78,958	2,469	23,394	53,095	386	11	508	3.1%	9.5%	2.1%

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
2020-02-22	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 (28 Apr 2020 UTC)

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	World	3,116,398	74,634	1,970,587	928,658	217,153	5,986	**7.0%**
2	US	1,012,582	24,385	838,291	115,936	58,355	2,096	**5.8%**
3	Spain	232,128	2,706	84,403	123,903	23,822	301	**10.3%**
4	Italy	201,505	2,091	105,205	68,941	27,359	382	**13.6%**
5	France	169,053	3,090	97,584	47,775	23,694	367	**14.0%**
6	United Kingdom	162,350	4,002	139,792	813	21,745	588	**13.4%**
7	Germany	159,912	1,154	36,198	117,400	6,314	188	3.9%
8	Turkey	114,653	2,392	72,852	38,809	2,992	92	2.6%
9	Russia	93,558	6,411	84,235	8,456	867	73	0.9%
10	Iran	92,584	1,112	14,268	72,439	5,877	71	**6.3%**
11	China	83,940	22	881	78,422	4,637	0	**5.5%**
12	Brazil	73,235	5,789	35,608	32,544	5,083	480	**6.9%**
13	Canada	51,150	1,534	28,936	19,231	2,983	142	**5.8%**
14	Belgium	47,334	647	29,060	10,943	7,331	124	**15.5%**
15	Netherlands	38,612	172	33,913	117	4,582	48	**11.9%**
16	India	31,324	1,873	22,569	7,747	1,008	69	3.2%
17	Peru	31,190	2,491	21,157	9,179	854	72	2.7%
18	Switzerland	29,264	100	4,965	22,600	1,699	34	**5.8%**

Table 6: Cases by Country (28 Apr 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
19	Portugal	24,322	295	21,985	1,389	948	20	3.9%
20	Ecuador	24,258	1,018	21,830	1,557	871	208	3.6%
21	Saudi Arabia	20,077	1,266	17,141	2,784	152	8	0.8%
22	Ireland	19,877	229	9,485	9,233	1,159	57	**5.8%**
23	Sweden	19,621	695	16,261	1,005	2,355	81	**12.0%**
24	Mexico	16,752	1,223	3,760	11,423	1,569	135	**9.4%**
25	Israel	15,728	173	7,772	7,746	210	6	1.3%
26	Austria	15,357	83	2,208	12,580	569	20	3.7%
27	Singapore	14,951	528	13,809	1,128	14	0	0.1%
28	Pakistan	14,612	697	11,067	3,233	312	20	2.1%
29	Chile	14,365	552	6,448	7,710	207	9	1.4%
30	Japan	13,736	0	11,443	1,899	394	9	2.9%
31	Poland	12,218	316	8,967	2,655	596	34	4.9%
32	Belarus	12,208	919	10,136	1,993	79	4	0.6%
33	Qatar	11,921	677	10,777	1,134	10	0	0.1%
34	Romania	11,616	277	7,549	3,404	663	22	**5.7%**
35	United Arab Emirates	11,380	541	9,110	2,181	89	7	0.8%
36	Korea, South	10,761	9	1,593	8,922	246	2	2.3%
37	Indonesia	9,511	415	7,484	1,254	773	8	**8.1%**
38	Ukraine	9,410	401	8,179	992	239	19	2.5%
39	Denmark	9,049	153	2,302	6,313	434	7	4.8%
40	Philippines	7,958	181	6,453	975	530	19	**6.7%**
41	Norway	7,660	61	7,422	32	206	1	2.7%
42	Czechia	7,504	59	4,329	2,948	227	4	3.0%
43	Australia	6,744	23	990	5,665	89	6	1.3%
44	Serbia	6,630	0	5,635	870	125	0	1.9%
45	Bangladesh	6,462	549	6,168	139	155	3	2.4%
46	Dominican Republic	6,416	123	4,965	1,165	286	4	4.5%
47	Panama	6,021	0	5,399	455	167	0	2.8%
48	Colombia	5,949	352	4,412	1,268	269	16	4.5%
49	Malaysia	5,851	31	1,719	4,032	100	1	1.7%
50	Egypt	5,042	260	3,379	1,304	359	22	**7.1%**
51	South Africa	4,996	203	2,830	2,073	93	3	1.9%
52	Finland	4,740	45	1,741	2,800	199	6	4.2%
53	Morocco	4,252	132	3,309	778	165	3	3.9%
54	Argentina	4,127	124	2,758	1,162	207	10	**5.0%**
55	Luxembourg	3,741	12	529	3,123	89	1	2.4%
56	Algeria	3,649	132	1,561	1,651	437	5	**12.0%**
57	Moldova	3,638	157	2,560	975	103	1	2.8%
58	Kuwait	3,440	152	2,241	1,176	23	1	0.7%
59	Kazakhstan	3,027	192	2,228	774	25	0	0.8%
60	Thailand	2,938	7	232	2,652	54	2	1.8%
61	Bahrain	2,811	88	1,493	1,310	8	0	0.3%
62	Hungary	2,649	66	1,842	516	291	11	**11.0%**
63	Greece	2,566	32	1,851	577	138	2	**5.4%**
64	Oman	2,131	82	1,757	364	10	0	0.5%
65	Croatia	2,047	8	752	1,232	63	4	3.1%
66	Uzbekistan	1,939	35	939	992	8	0	0.4%
67	Iraq	1,928	81	519	1,319	90	2	4.7%
68	Armenia	1,867	59	971	866	30	1	1.6%
69	Afghanistan	1,828	125	1,542	228	58	1	3.2%
70	Iceland	1,795	3	149	1,636	10	0	0.6%
71	Azerbaijan	1,717	39	474	1,221	22	0	1.3%
72	Cameroon	1,705	0	732	915	58	0	3.4%
73	Ghana	1,671	121	1,467	188	16	5	1.0%
74	Estonia	1,660	13	1,370	240	50	0	3.0%
75	Bosnia and Herzegovina	1,585	20	840	682	63	3	4.0%
76	Nigeria	1,532	195	1,233	255	44	4	2.9%
77	New Zealand	1,474	2	226	1,229	19	0	1.3%
78	Cuba	1,437	48	804	575	58	2	4.0%
79	North Macedonia	1,421	22	761	589	71	6	5.0%
80	Slovenia	1,408	6	1,099	223	86	3	6.1%
81	Bulgaria	1,399	36	1,119	222	58	0	4.1%
82	Slovakia	1,384	3	941	423	20	2	1.4%
83	Lithuania	1,344	0	764	536	44	3	3.3%
84	Guinea	1,240	77	964	269	7	0	0.6%
85	Cote d'Ivoire	1,183	19	644	525	14	0	1.2%
86	Djibouti	1,072	37	572	498	2	0	0.2%

Table 6: Cases by Country (28 Apr 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
87	Bolivia	1,014	0	863	98	53	0	5.2%
88	Tunisia	975	8	656	279	40	1	4.1%
89	Cyprus	837	15	674	148	15	0	1.8%
90	Latvia	836	18	556	267	13	0	1.6%
91	Senegal	823	87	518	296	9	0	1.1%
92	Albania	750	14	289	431	30	2	4.0%
93	Andorra	743	0	304	398	41	1	5.5%
94	Lebanon	717	7	548	145	24	0	3.3%
95	Diamond Princess	712	0	54	645	13	0	1.8%
96	Niger	709	8	275	403	31	2	4.4%
97	Kyrgyzstan	708	13	284	416	8	0	1.1%
98	Costa Rica	705	8	393	306	6	0	0.9%
99	Honduras	702	41	559	79	64	3	9.1%
100	Burkina Faso	638	3	120	476	42	0	6.6%
101	Uruguay	625	5	216	394	15	0	2.4%
102	Sri Lanka	619	31	478	134	7	0	1.1%
103	San Marino	553	15	448	64	41	0	7.4%
104	Guatemala	530	0	466	49	15	0	2.8%
105	Somalia	528	48	481	19	28	2	5.3%
106	Georgia	511	14	337	168	6	0	1.2%
107	Kosovo	510	0	405	93	12	0	2.4%
108	Congo (Kinshasa)	471	12	385	56	30	2	6.4%
109	Malta	458	8	151	303	4	0	0.9%
110	Jordan	449	0	93	348	8	1	1.8%
111	Taiwan*	429	0	116	307	6	0	1.4%
112	Mali	424	16	278	122	24	1	5.7%
113	Kenya	374	11	236	124	14	0	3.7%
114	Jamaica	364	0	328	29	7	0	1.9%
115	El Salvador	345	22	240	97	8	0	2.3%
116	West Bank and Gaza	343	1	270	71	2	0	0.6%
117	Mauritius	334	0	21	303	10	0	3.0%
118	Venezuela	329	0	177	142	10	0	3.0%
119	Montenegro	321	0	115	199	7	0	2.2%
120	Sudan	318	43	262	31	25	3	7.9%
121	Equatorial Guinea	315	57	305	9	1	0	0.3%
122	Tanzania	299	0	241	48	10	0	3.3%
123	Vietnam	270	0	48	222	0	0	0.0%
124	Maldives	250	24	233	17	0	0	0.0%
125	Paraguay	239	11	128	102	9	0	3.8%
126	Gabon	238	27	182	53	3	0	1.3%
127	Rwanda	212	5	117	95	0	0	0.0%
128	Congo (Brazzaville)	207	7	180	19	8	2	3.9%
129	Burma	150	4	129	16	5	0	3.3%
130	Liberia	141	17	80	45	16	4	11.3%
131	Brunei	138	0	13	124	1	0	0.7%
132	Madagascar	128	0	46	82	0	0	0.0%
133	Ethiopia	126	2	73	50	3	0	2.4%
134	Cambodia	122	0	3	119	0	0	0.0%
135	Trinidad and Tobago	116	0	49	59	8	0	6.9%
136	Cabo Verde	114	5	111	2	1	0	0.9%
137	Sierra Leone	104	11	88	12	4	0	3.8%
138	Togo	99	1	30	63	6	0	6.1%
139	Monaco	95	0	41	50	4	0	4.2%
140	Zambia	95	7	50	42	3	0	3.2%
141	Liechtenstein	82	0	26	55	1	0	1.2%
142	Bahamas	80	0	46	23	11	0	13.8%
143	Barbados	80	0	35	39	6	0	7.5%
144	Uganda	79	0	27	52	0	0	0.0%
145	Haiti	76	0	62	8	6	0	7.9%
146	Mozambique	76	0	64	12	0	0	0.0%
147	Guyana	74	0	51	15	8	0	10.8%
148	Guinea-Bissau	73	0	54	18	1	0	1.4%
149	Eswatini	71	6	60	10	1	0	1.4%
150	Benin	64	0	30	33	1	0	1.6%
151	Libya	61	0	41	18	2	0	3.3%
152	Nepal	54	2	38	16	0	0	0.0%
153	Chad	52	6	31	19	2	2	3.8%
154	Central African Republic	50	31	40	10	0	0	0.0%

Table 6: Cases by Country (28 Apr 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
155	Syria	43	0	19	21	3	0	7.0%
156	Eritrea	39	0	20	19	0	0	0.0%
157	Mongolia	38	0	28	10	0	0	0.0%
158	Malawi	36	0	28	5	3	0	8.3%
159	South Sudan	34	28	34	0	0	0	0.0%
160	Zimbabwe	32	0	23	5	4	0	12.5%
161	Angola	27	0	19	6	2	0	7.4%
162	Antigua and Barbuda	24	0	10	11	3	0	12.5%
163	Timor-Leste	24	0	18	6	0	0	0.0%
164	Botswana	23	1	22	0	1	0	4.3%
165	Grenada	19	1	9	10	0	0	0.0%
166	Laos	19	0	12	7	0	0	0.0%
167	Belize	18	0	7	9	2	0	11.1%
168	Fiji	18	0	6	12	0	0	0.0%
169	Dominica	16	0	3	13	0	0	0.0%
170	Namibia	16	0	8	8	0	0	0.0%
171	Saint Kitts and Nevis	15	0	11	4	0	0	0.0%
172	Saint Lucia	15	0	0	15	0	0	0.0%
173	Saint Vincent and the Grenadines	15	0	7	8	0	0	0.0%
174	Nicaragua	13	0	3	7	3	0	23.1%
175	Burundi	11	0	6	4	1	0	9.1%
176	Seychelles	11	0	5	6	0	0	0.0%
177	Gambia	10	0	1	8	1	0	10.0%
178	Holy See	10	1	8	2	0	0	0.0%
179	Suriname	10	0	2	7	1	0	10.0%
180	MS Zaandam	9	0	7	0	2	0	22.2%
181	Papua New Guinea	8	0	8	0	0	0	0.0%
182	Sao Tome and Principe	8	4	4	4	0	0	0.0%
183	Bhutan	7	0	2	5	0	0	0.0%
184	Mauritania	7	0	0	6	1	0	14.3%
185	Western Sahara	6	0	1	5	0	0	0.0%
186	Yemen	1	0	0	1	0	0	0.0%

Appendix B. How to Cite This Work

Citation

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

BibTex

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

Appendix C. Contact

Contact:

Dr. Yanchang Zhao

Email: yanchang@RDataMining.com

Twitter: @RDataMining

LinkedIn: http://group.rdatamining.com

Comments and suggestions and welcome. Thanks!