COVID-19 Data Analysis with R - Worldwide

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

12 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')
raw.data.recovered <- read.csv('./data/time_series_covid19_recovered_global.csv')

dim(raw.data.confirmed)
```

[1] https://github.com/CSSEGISandData/COVID-19

```
## [1] 264  84
```

Each dataset has 264 rows, corresponding to country/region/province/state. It has 84 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=6, 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" "2020-04-10"
```

```
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 10 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 10 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)

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

3.3 Daily Increases and Death Rates

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

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

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

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

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

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

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

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

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


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


## convert from wide to long format, for drawing area plots
rates.long <- data %>%
```

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

4 Worldwide Cases

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

4.1 World Map

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

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

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

# world
m
```

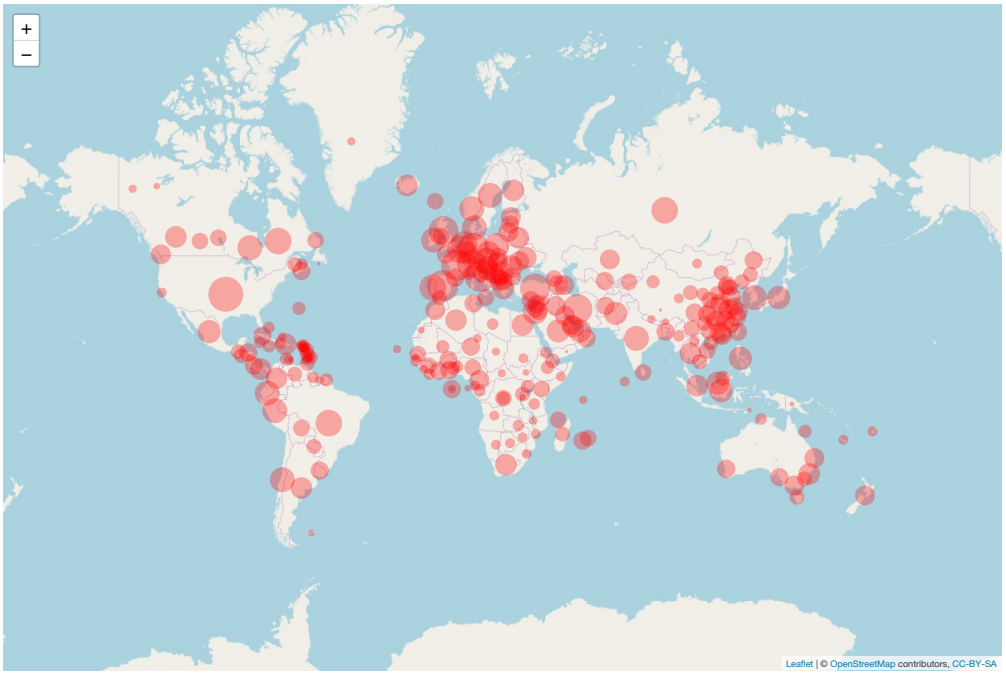


Figure 1: World Map

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)
```

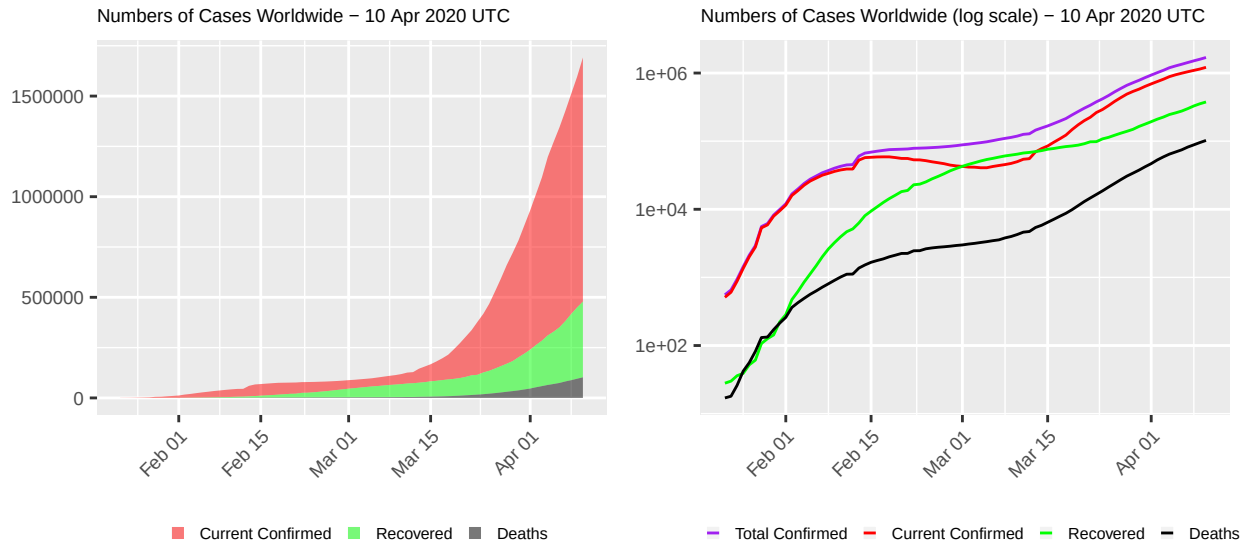


Figure 2: COVID-19 Cases Worldwide

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)
```

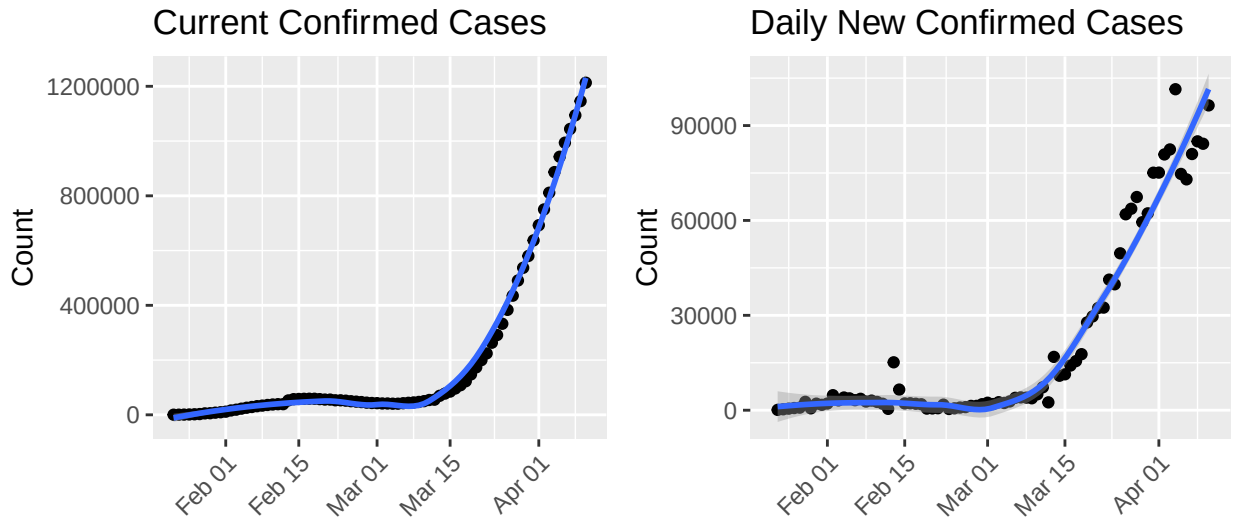


Figure 3: Current Confirmed Cases

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)
```

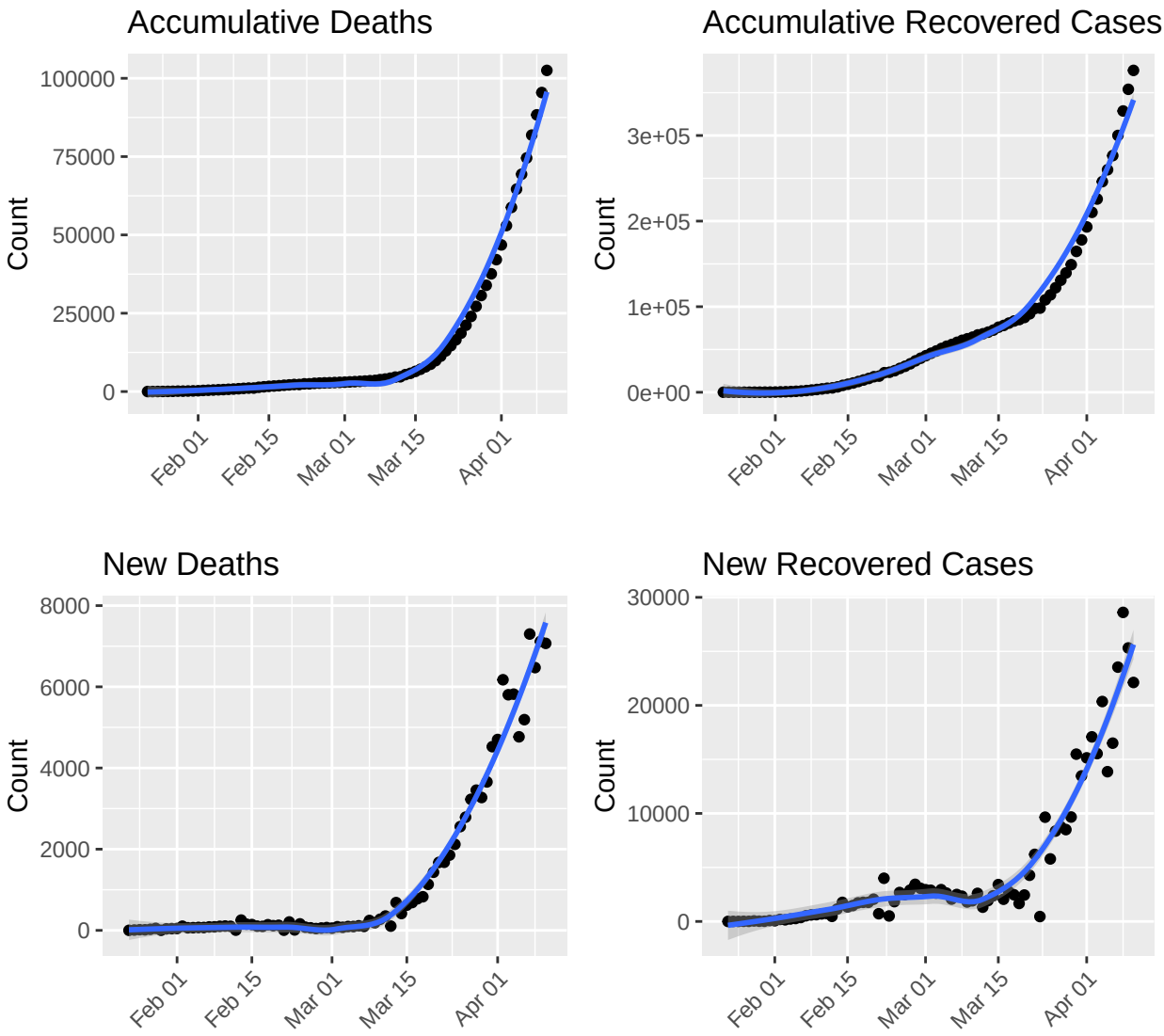


Figure 4: Deaths and Recovered Cases

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 10 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 10 Apr 2020 UTC UTC, it will be between 6.1% and 21.4%.

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 recent outbreak of coronavirus in Italy, Iran and some other European countries.

```
## 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)
```

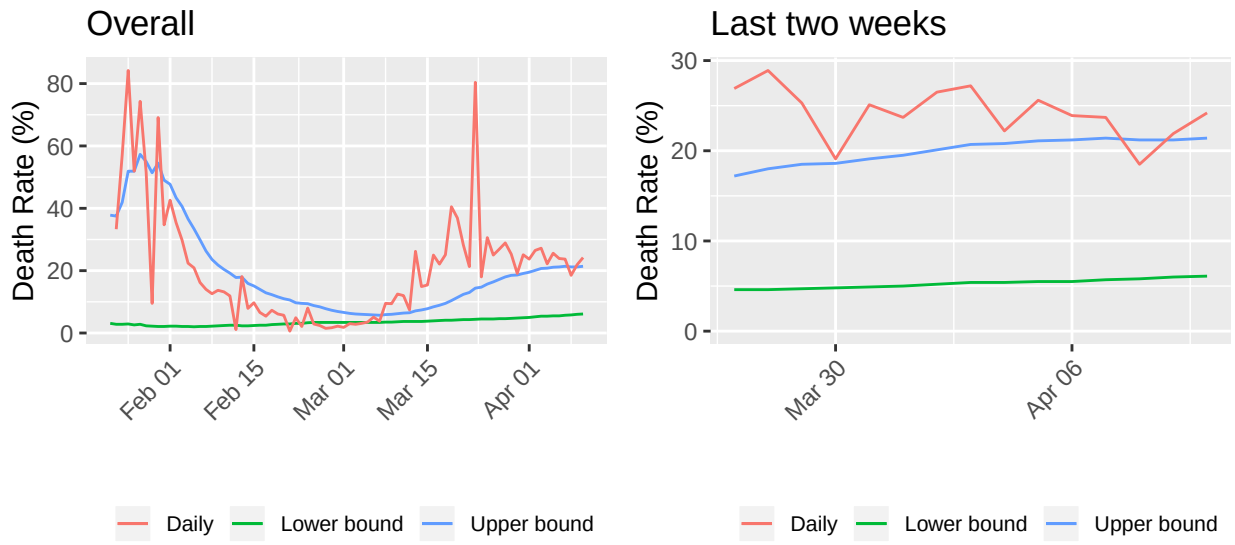


Figure 5: Death Rate

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] "Germany"        "China"          "United Kingdom" "Iran"
##  [9] "Turkey"         "Belgium"        "Switzerland"    "Netherlands"
## [13] "Canada"         "Brazil"         "Portugal"       "Austria"
## [17] "Russia"         "Korea, South"   "Israel"         "Sweden"
```

```
## 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 - 10 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	1,691,719	102,525	6.1%	96,369	7,070	1,213,098
2	US	496,535	18,586	3.7%	35,098	2,108	449,159
3	Spain	158,273	16,081	10.2%	5,051	634	86,524
4	Italy	147,577	18,849	12.8%	3,951	570	98,273
5	France	125,931	13,215	10.5%	7,150	987	87,521
6	Germany	122,171	2,767	2.3%	3,990	160	65,491
7	China	82,941	3,340	4.0%	58	1	1,810
8	United Kingdom	74,605	8,974	12.0%	8,733	981	65,043
9	Iran	68,192	4,232	6.2%	1,972	122	28,495
10	Turkey	47,029	1,006	2.1%	4,747	98	43,600
11	Belgium	26,667	3,019	11.3%	1,684	496	18,080
12	Switzerland	24,551	1,002	4.1%	500	54	12,449
13	Netherlands	23,249	2,520	10.8%	1,346	117	20,442
14	Canada	22,059	557	2.5%	1,405	54	15,647
15	Brazil	19,638	1,057	5.4%	1,546	107	18,408
16	Portugal	15,472	435	2.8%	1,516	26	14,804
17	Austria	13,555	319	2.4%	311	24	7,172
18	Russia	11,917	94	0.8%	1,786	18	11,028
19	Korea, South	10,450	208	2.0%	27	4	3,125
20	Israel	10,408	95	0.9%	440	9	9,130
21	Sweden	9,685	870	9.0%	544	77	8,434
22	Others	180,814	5,299	2.9%	14,514	423	148,463

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

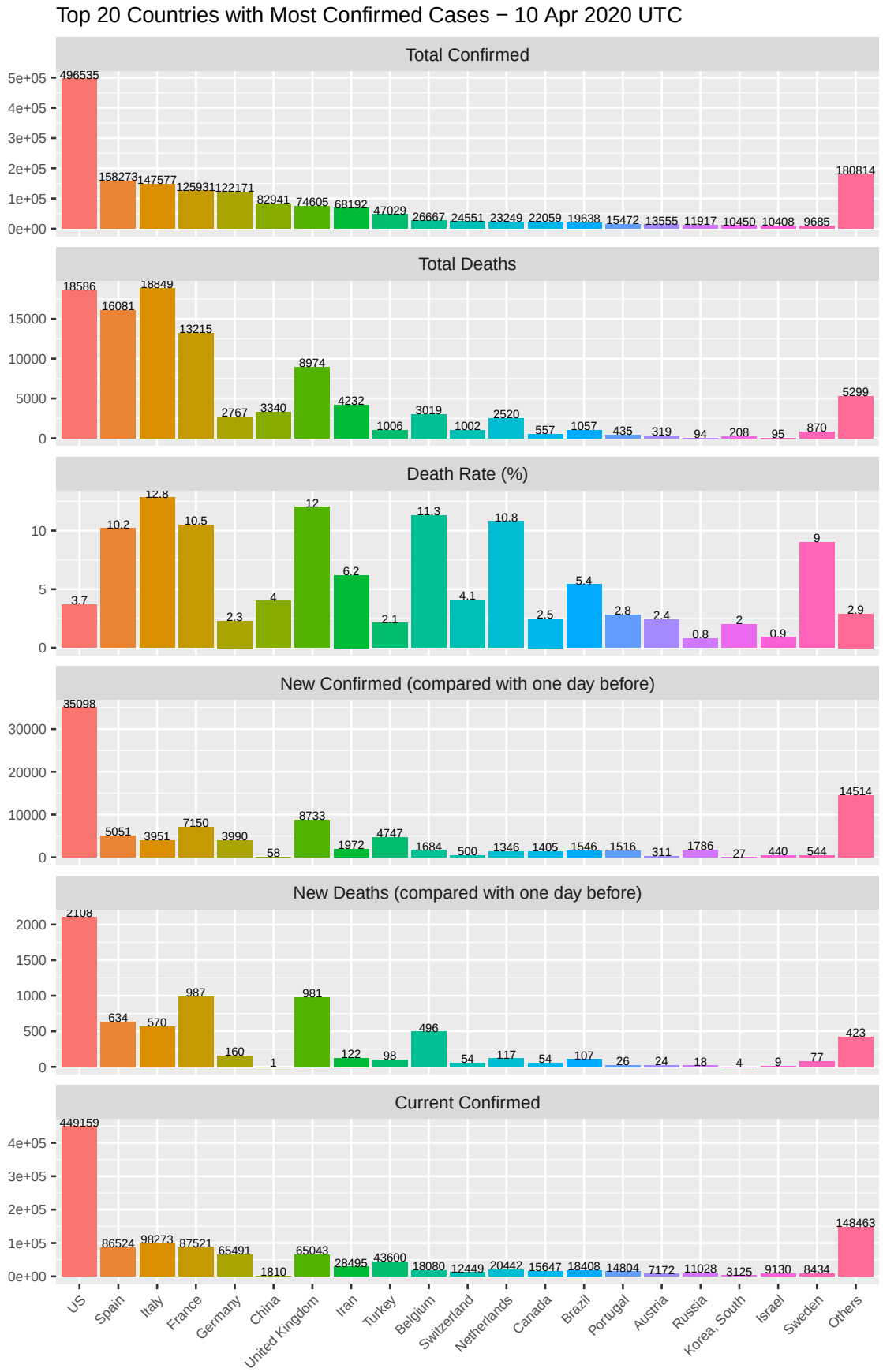
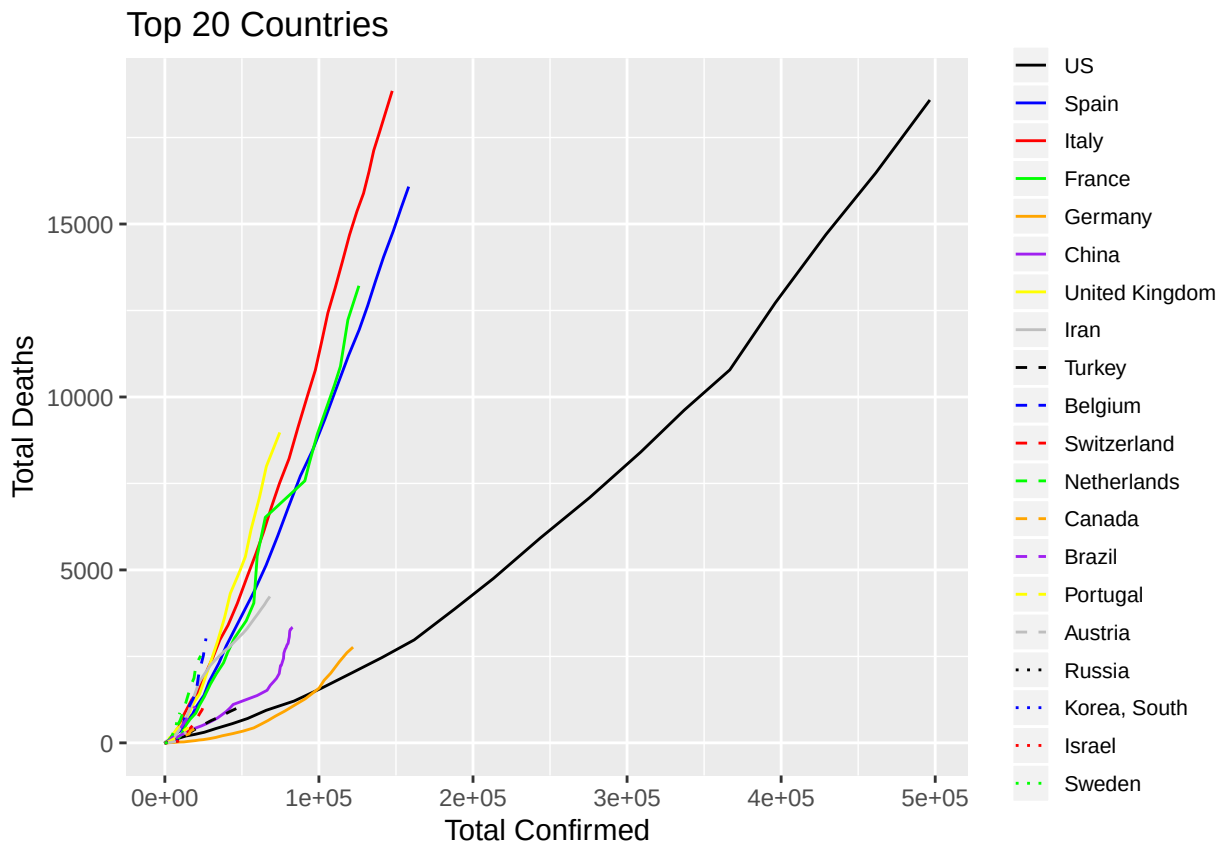


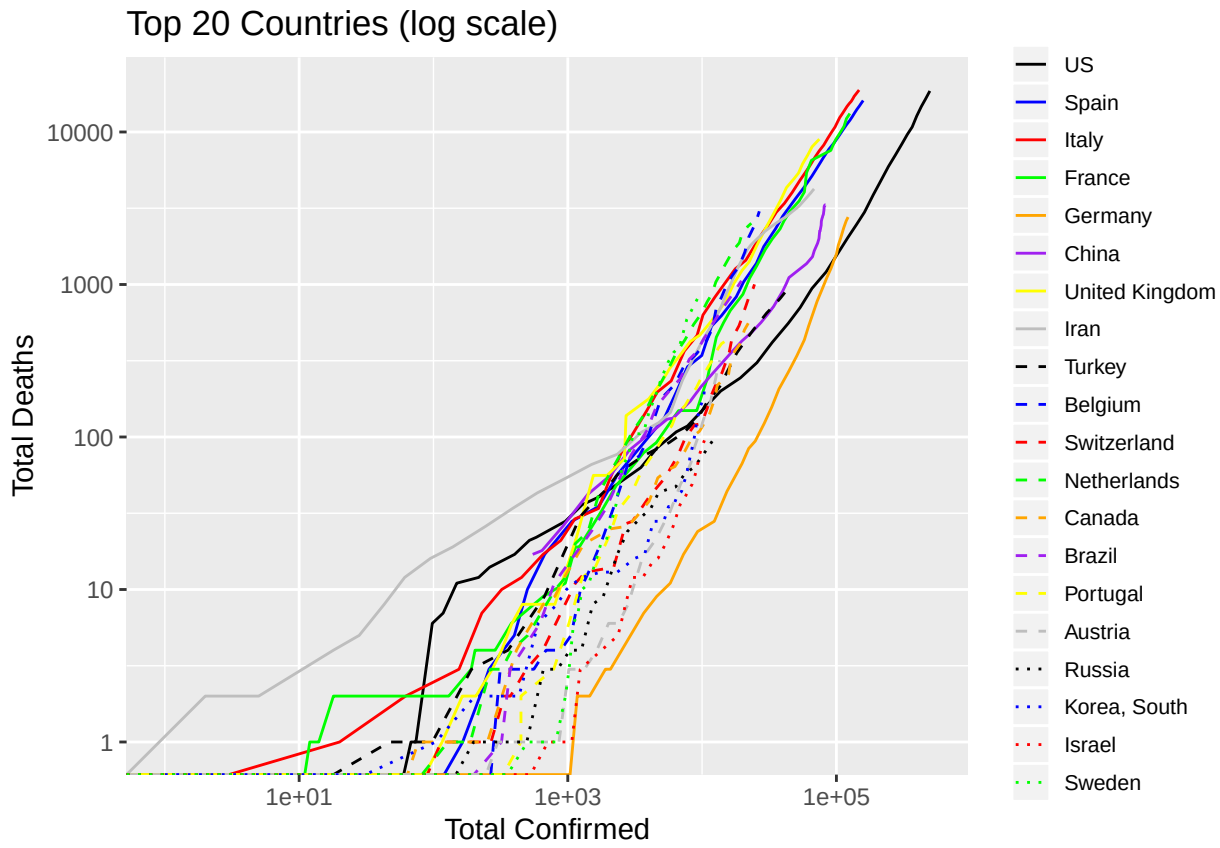
Figure 6: Top 20 Countries with Most Confirmed Cases

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 10 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, 4e5, 1e6)

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)
```

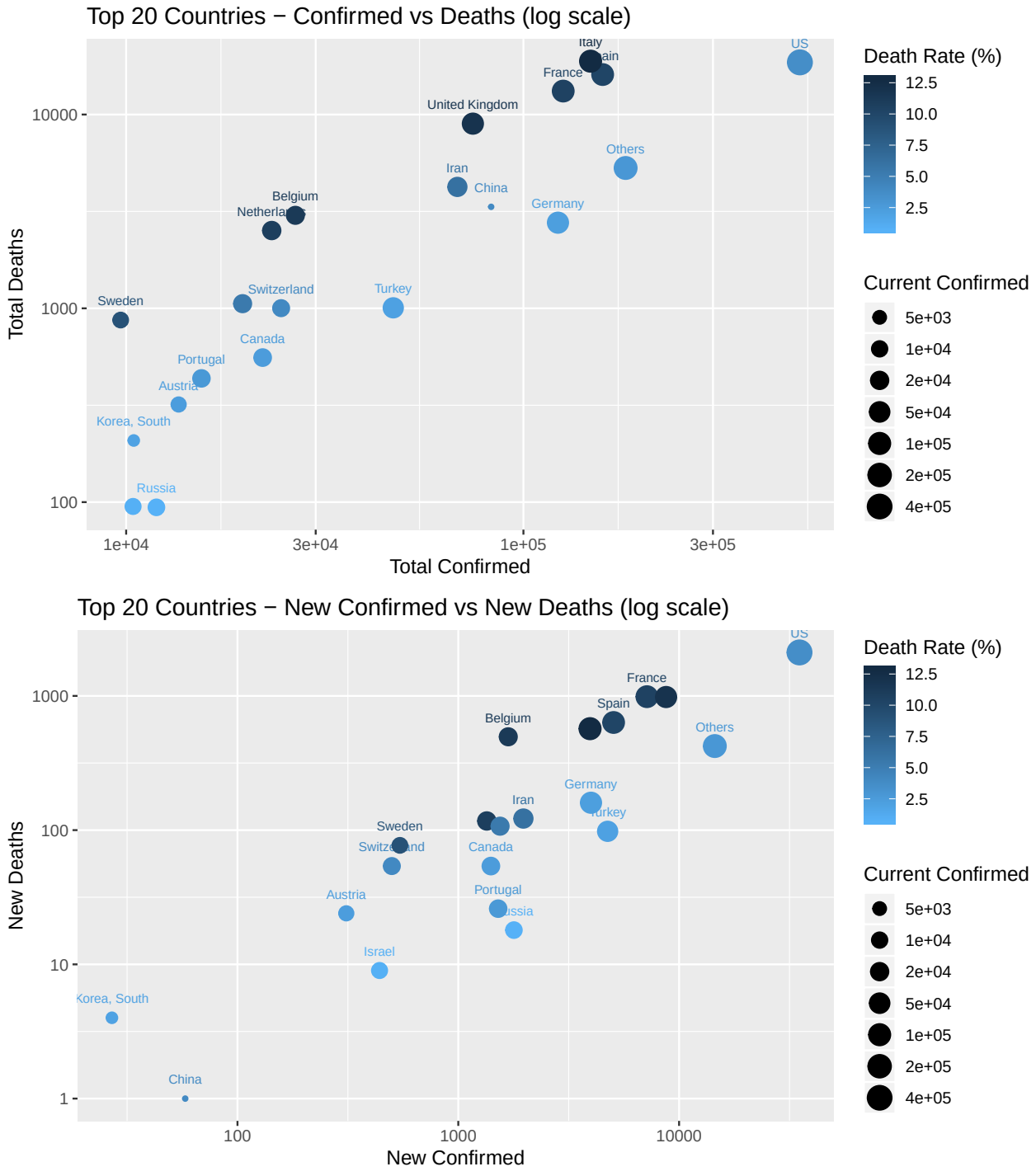


Figure 7: Top 20 Countries

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)
```

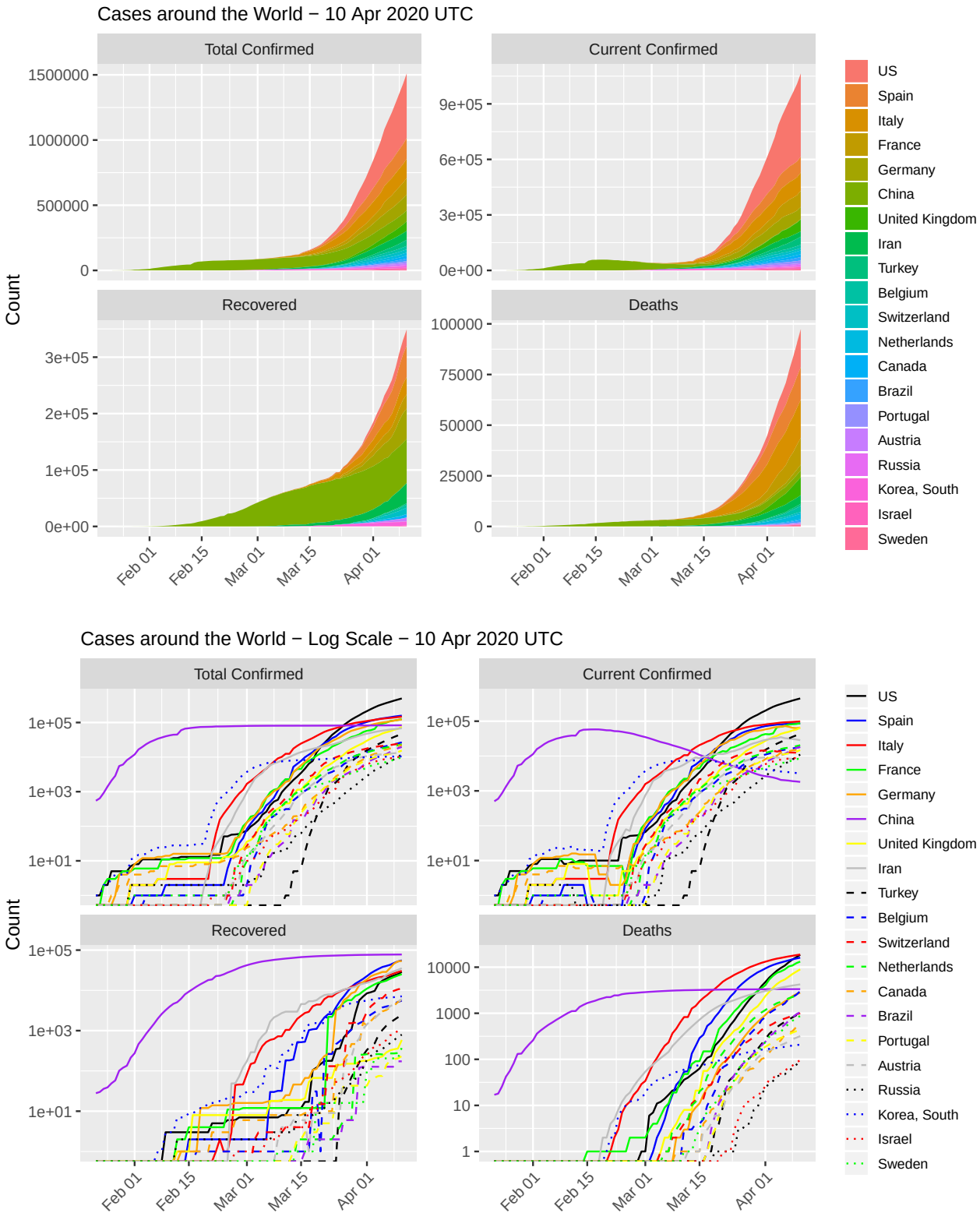


Figure 8: Cases around the World

```
## 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))
```

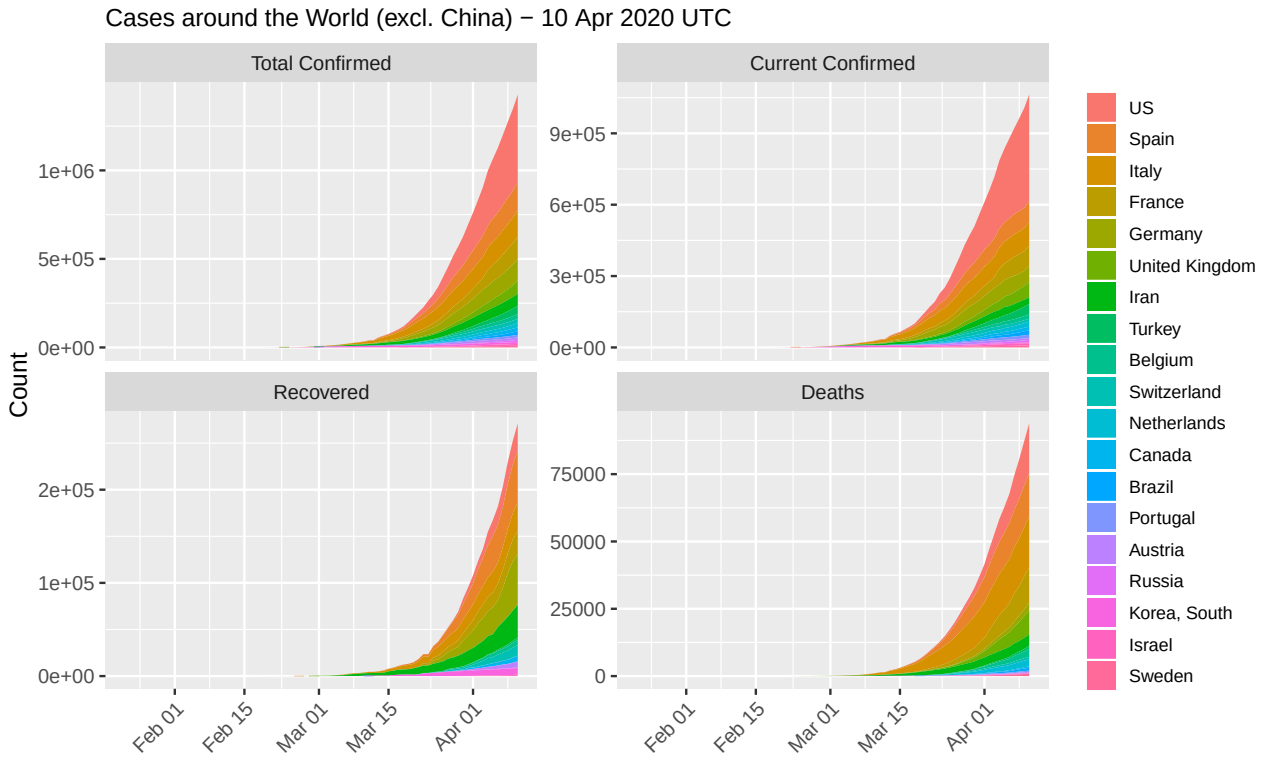


Figure 9: Cases around the World (excl. China)

```
## 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')
```

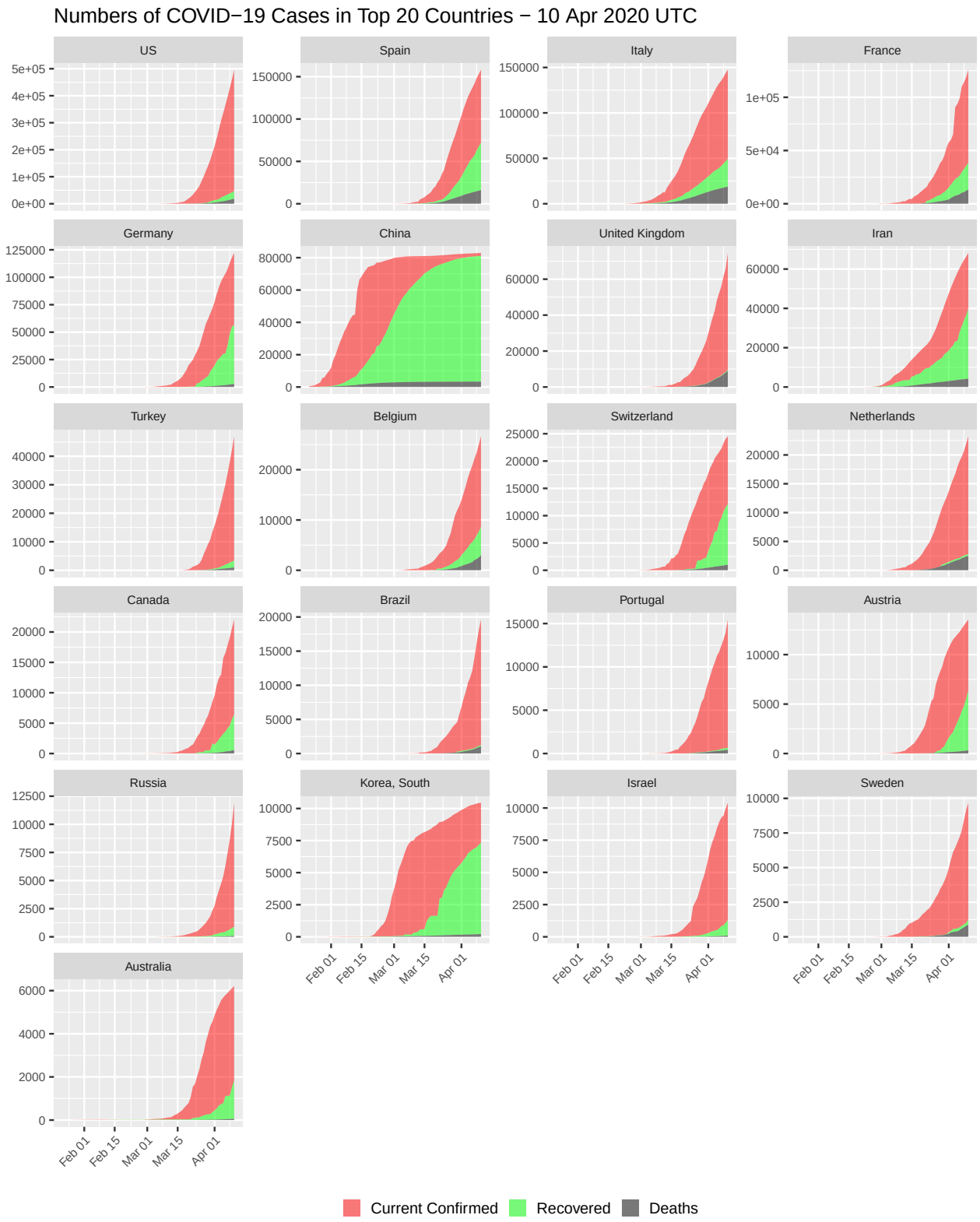


Figure 10: COVID-19 Cases in Top 20 Countries. Ordered descendingly by number of confirmed cases.

```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')
```

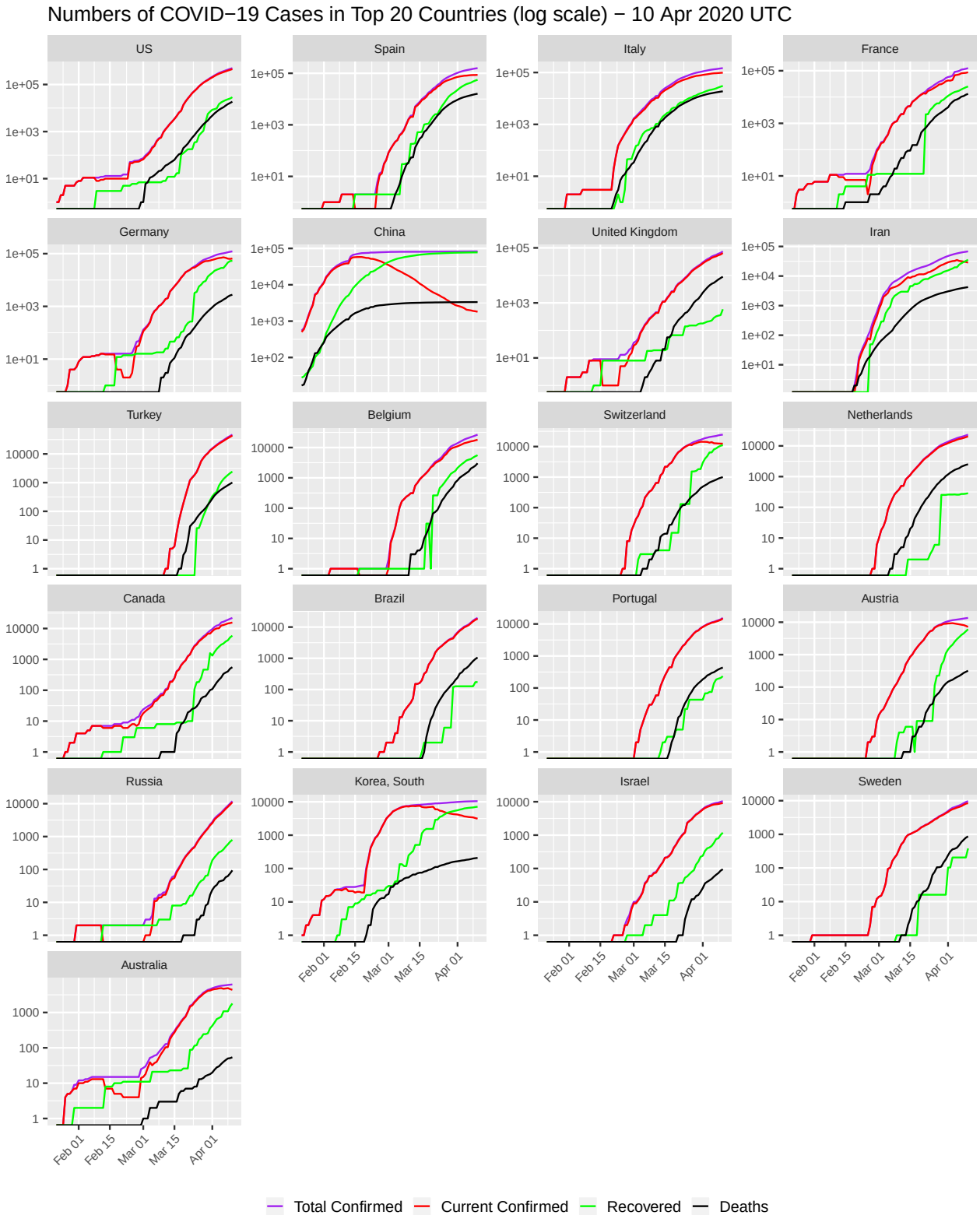


Figure 11: COVID-19 Cases Top 20 Countries (log scale). Ordered descendingly by number of confirmed cases.

```
## 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)
```

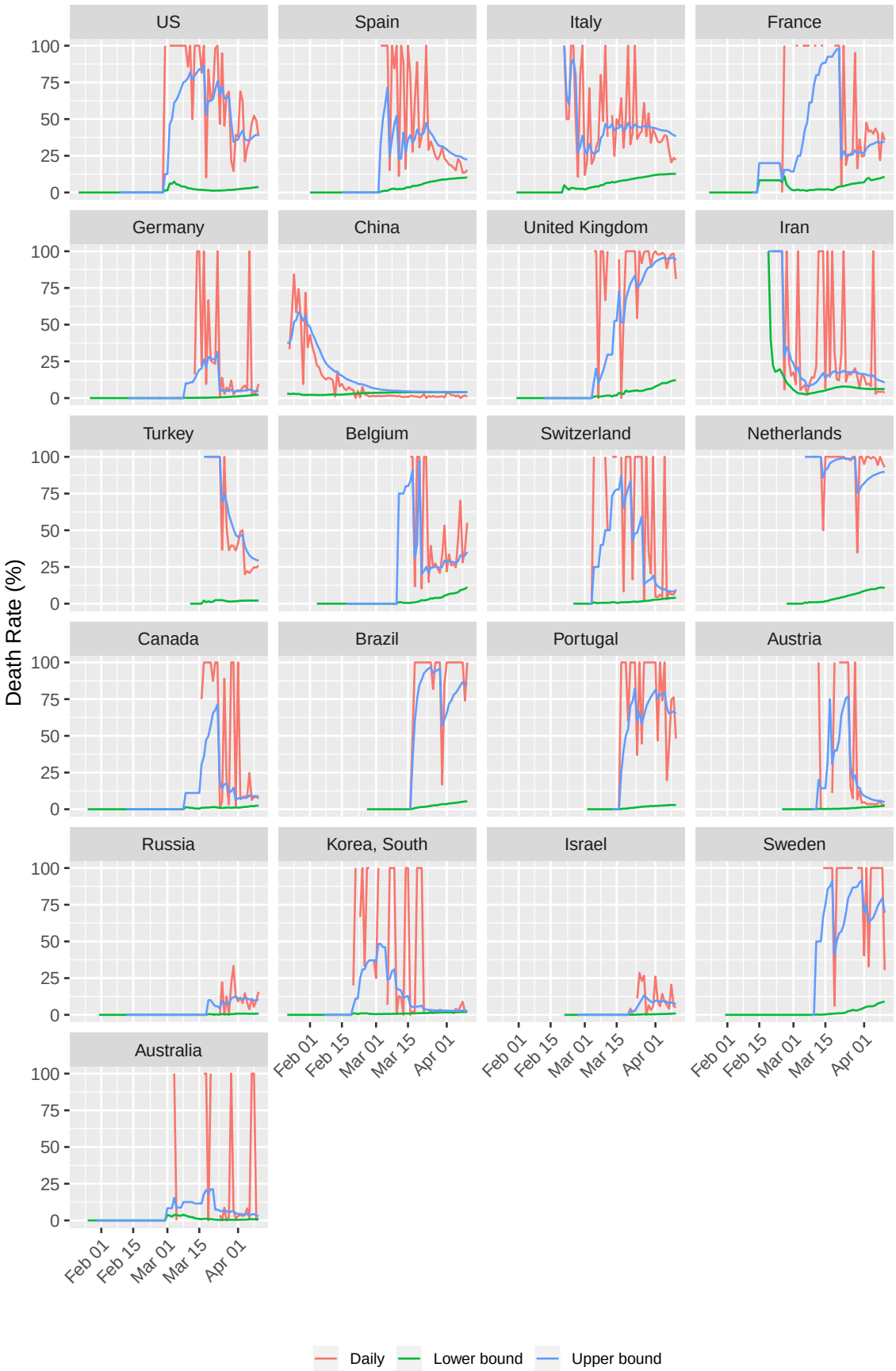


Figure 12: Death Rates

5.4 Countries with Highest Death Rates

Below are a list of top 20 countries with the highest death rates out of countires having 1000+ confirmed cases.

```
## sort the latest data by death rate, and if tie, by confirmed
df <- data %>% filter(date == max(date) & country != 'World' & confirmed >= 1000) %>%
  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 - 10 Apr 2020 UTC

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	Algeria	1,761	95	1,100	405	256	21	14.5%
2	Italy	147,577	3,951	98,273	30,455	18,849	570	12.8%
3	United Kingdom	74,605	8,733	65,043	588	8,974	981	12.0%
4	Belgium	26,667	1,684	18,080	5,568	3,019	496	11.3%
5	Netherlands	23,249	1,346	20,442	287	2,520	117	10.8%
6	France	125,931	7,150	87,521	25,195	13,215	987	10.5%
7	Spain	158,273	5,051	86,524	55,668	16,081	634	10.2%
8	Sweden	9,685	544	8,434	381	870	77	9.0%
9	Indonesia	3,512	219	2,924	282	306	26	8.7%
10	Egypt	1,794	95	1,275	384	135	17	7.5%
11	Morocco	1,448	74	1,219	122	107	10	7.4%
12	Hungary	1,190	210	1,001	112	77	11	6.5%
13	Iran	68,192	1,972	28,495	35,465	4,232	122	6.2%
14	Mexico	3,441	260	2,614	633	194	20	5.6%
15	Iraq	1,279	47	659	550	70	1	5.5%
16	Brazil	19,638	1,546	18,408	173	1,057	107	5.4%
17	Philippines	4,195	119	3,834	140	221	18	5.3%
18	Romania	5,467	265	4,468	729	270	22	4.9%
19	Dominican Republic	2,620	271	2,396	98	126	8	4.8%
20	Greece	2,011	56	1,650	269	92	5	4.6%

6 Conclusions

As of 10 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 6.1% and 21.4%, but it is likely to change dramatically with the breakout in many countries, such as European countries.

Appendix A. Processed Data

Blow is the processed data for this analysis.

Appendix A.1 COVID-19 Cases Worldwide

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

```
  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 %>% kable('latex', booktabs=T, longtable=T, caption='Cases in the Whole World',
               format.args=list(big.mark=',')) %>%
    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
2020-04-10	1,691,719	102,525	376,096	1,213,098	96,369	7,070	22,121	6.1	21.4	24.2
2020-04-09	1,595,350	95,455	353,975	1,145,920	84,246	7,117	25,314	6.0	21.2	21.9
2020-04-08	1,511,104	88,338	328,661	1,094,105	85,008	6,473	28,607	5.8	21.2	18.5
2020-04-07	1,426,096	81,865	300,054	1,044,177	80,995	7,300	23,539	5.7	21.4	23.7
2020-04-06	1,345,101	74,565	276,515	994,021	72,986	5,191	16,503	5.5	21.2	23.9
2020-04-05	1,272,115	69,374	260,012	942,729	74,707	4,768	13,860	5.5	21.1	25.6
2020-04-04	1,197,408	64,606	246,152	886,650	101,491	5,819	20,356	5.4	20.8	22.2
2020-04-03	1,095,917	58,787	225,796	811,334	82,451	5,804	15,533	5.4	20.7	27.2
2020-04-02	1,013,466	52,983	210,263	750,220	80,861	6,174	17,086	5.2	20.1	26.5
2020-04-01	932,605	46,809	193,177	692,619	75,118	4,702	15,143	5.0	19.5	23.7
2020-03-31	857,487	42,107	178,034	637,346	75,098	4,525	13,468	4.9	19.1	25.1
2020-03-30	782,389	37,582	164,566	580,241	62,249	3,657	15,484	4.8	18.6	19.1
2020-03-29	720,140	33,925	149,082	537,133	59,447	3,273	9,667	4.7	18.5	25.3
2020-03-28	660,693	30,652	139,415	490,626	67,402	3,454	8,500	4.6	18.0	28.9
2020-03-27	593,291	27,198	130,915	435,178	63,700	3,228	8,765	4.6	17.2	26.9
2020-03-26	529,591	23,970	122,150	383,471	61,938	2,789	8,363	4.5	16.4	25.0
2020-03-25	467,653	21,181	113,787	332,685	49,608	2,556	5,787	4.5	15.7	30.6
2020-03-24	418,045	18,625	108,000	291,420	39,810	2,120	9,649	4.5	14.7	18.0
2020-03-23	378,235	16,505	98,351	263,379	41,282	1,854	452	4.4	14.4	80.4
2020-03-22	336,953	14,651	97,899	224,403	32,446	1,678	6,207	4.3	13.0	21.3
2020-03-21	304,507	12,973	91,692	199,842	32,299	1,674	4,272	4.3	12.4	28.2
2020-03-20	272,208	11,299	87,420	173,489	29,638	1,432	2,445	4.2	11.4	36.9
2020-03-19	242,570	9,867	84,975	147,728	27,749	1,134	1,663	4.1	10.4	40.5
2020-03-18	214,821	8,733	83,312	122,776	17,719	828	2,472	4.1	9.5	25.1
2020-03-17	197,102	7,905	80,840	108,357	15,528	779	2,752	4.0	8.9	22.1
2020-03-16	181,574	7,126	78,088	96,360	14,120	686	2,054	3.9	8.4	25.0
2020-03-15	167,454	6,440	76,034	84,980	11,353	621	3,410	3.8	7.8	15.4
2020-03-14	156,101	5,819	72,624	77,658	10,896	415	2,373	3.7	7.4	14.9
2020-03-13	145,205	5,404	70,251	69,550	16,853	684	1,927	3.7	7.1	26.2
2020-03-12	128,352	4,720	68,324	55,308	2,477	105	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
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

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
2020-01-25	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	

Appendix A.2 Latest Cases by Country

```
## hightlight high death rates (if >= 5%) for those countries with 1000+ confirmed cases
data.latest.all %>% arrange(desc(confirmed)) %>% select(-c(date, ranking)) %>%
  mutate(death.rate = cell_spec(death.rate, "latex",
                                color = ifelse(confirmed >= 1000 & death.rate >= 5, "red", "black"),
                                bold = ifelse(confirmed >= 1000 & death.rate >= 5, T, 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 (10 Apr 2020 UTC)

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
1	World	1,691,719	96,369	1,213,098	376,096	102,525	7,070	**6.1**
2	US	496,535	35,098	449,159	28,790	18,586	2,108	3.7
3	Spain	158,273	5,051	86,524	55,668	16,081	634	**10.2**
4	Italy	147,577	3,951	98,273	30,455	18,849	570	**12.8**
5	France	125,931	7,150	87,521	25,195	13,215	987	**10.5**
6	Germany	122,171	3,990	65,491	53,913	2,767	160	2.3
7	China	82,941	58	1,810	77,791	3,340	1	4
8	United Kingdom	74,605	8,733	65,043	588	8,974	981	**12**
9	Iran	68,192	1,972	28,495	35,465	4,232	122	**6.2**
10	Turkey	47,029	4,747	43,600	2,423	1,006	98	2.1
11	Belgium	26,667	1,684	18,080	5,568	3,019	496	**11.3**
12	Switzerland	24,551	500	12,449	11,100	1,002	54	4.1
13	Netherlands	23,249	1,346	20,442	287	2,520	117	**10.8**
14	Canada	22,059	1,405	15,647	5,855	557	54	2.5
15	Brazil	19,638	1,546	18,408	173	1,057	107	**5.4**
16	Portugal	15,472	1,516	14,804	233	435	26	2.8
17	Austria	13,555	311	7,172	6,064	319	24	2.4
18	Russia	11,917	1,786	11,028	795	94	18	0.8
19	Korea, South	10,450	27	3,125	7,117	208	4	2
20	Israel	10,408	440	9,130	1,183	95	9	0.9
21	Sweden	9,685	544	8,434	381	870	77	**9**
22	Ireland	8,089	1,515	7,777	25	287	24	3.5
23	India	7,598	873	6,578	774	246	20	3.2
24	Ecuador	7,161	2,196	6,496	368	297	25	4.1
25	Chile	6,501	529	4,865	1,571	65	8	1
26	Norway	6,314	103	6,169	32	113	5	1.8
27	Australia	6,215	107	4,368	1,793	54	3	0.9
28	Denmark	6,014	184	3,838	1,929	247	10	4.1
29	Poland	5,955	380	5,456	318	181	7	3
30	Peru	5,897	641	4,159	1,569	169	31	2.9
31	Czechia	5,732	163	5,267	346	119	7	2.1
32	Japan	5,530	863	4,746	685	99	5	1.8
33	Romania	5,467	265	4,468	729	270	22	4.9
34	Pakistan	4,695	206	3,902	727	66	1	1.4
35	Malaysia	4,346	118	2,446	1,830	70	3	1.6
36	Philippines	4,195	119	3,834	140	221	18	**5.3**
37	Saudi Arabia	3,651	364	2,919	685	47	3	1.3
38	Indonesia	3,512	219	2,924	282	306	26	**8.7**
39	Mexico	3,441	260	2,614	633	194	20	**5.6**
40	United Arab Emirates	3,360	370	2,926	418	16	2	0.5
41	Luxembourg	3,223	108	2,669	500	54	2	1.7
42	Serbia	3,105	238	3,034	0	71	5	2.3
43	Finland	2,769	164	2,421	300	48	6	1.7
44	Panama	2,752	224	2,670	16	66	3	2.4

Table 6: Cases by Country (10 Apr 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
45	Dominican Republic	2,620	271	2,396	98	126	8	4.8
46	Qatar	2,512	136	2,279	227	6	0	0.2
47	Colombia	2,473	250	2,196	197	80	11	3.2
48	Thailand	2,473	50	1,427	1,013	33	1	1.3
49	Ukraine	2,203	311	2,073	61	69	12	3.1
50	Singapore	2,108	198	1,609	492	7	1	0.3
51	Greece	2,011	56	1,650	269	92	5	4.6
52	South Africa	2,003	69	1,569	410	24	6	1.2
53	Belarus	1,981	495	1,793	169	19	3	1
54	Argentina	1,975	180	1,518	375	82	10	4.2
55	Egypt	1,794	95	1,275	384	135	17	**7.5**
56	Algeria	1,761	95	1,100	405	256	21	**14.5**
57	Iceland	1,675	27	917	751	7	1	0.4
58	Croatia	1,495	88	1,243	231	21	1	1.4
59	Morocco	1,448	74	1,219	122	107	10	**7.4**
60	Moldova	1,438	149	1,353	56	29	0	2
61	New Zealand	1,283	44	908	373	2	1	0.2
62	Iraq	1,279	47	659	550	70	1	**5.5**
63	Estonia	1,258	51	1,141	93	24	0	1.9
64	Hungary	1,190	210	1,001	112	77	11	**6.5**
65	Slovenia	1,160	36	978	137	45	2	3.9
66	Lithuania	999	44	923	54	22	6	2.2
67	Kuwait	993	83	869	123	1	0	0.1
68	Azerbaijan	991	65	822	159	10	1	1
69	Armenia	937	16	776	149	12	2	1.3
70	Bahrain	925	38	380	539	6	1	0.6
71	Bosnia and Herzegovina	901	43	736	129	36	1	4
72	Cameroon	820	90	710	98	12	2	1.5
73	Kazakhstan	812	31	738	64	10	2	1.2
74	Slovakia	715	14	690	23	2	0	0.3
75	Diamond Princess	712	0	82	619	11	0	1.5
76	North Macedonia	711	48	638	41	32	2	4.5
77	Tunisia	671	28	621	25	25	0	3.7
78	Bulgaria	635	17	556	54	25	1	3.9
79	Uzbekistan	624	42	579	42	3	0	0.5
80	Latvia	612	23	593	16	3	0	0.5
81	Lebanon	609	27	513	76	20	1	3.3
82	Andorra	601	18	504	71	26	1	4.3
83	Cyprus	595	31	527	58	10	0	1.7
84	Cuba	564	49	498	51	15	0	2.7
85	Costa Rica	558	19	513	42	3	0	0.5
86	Afghanistan	521	37	474	32	15	0	2.9
87	Oman	484	27	372	109	3	0	0.6
88	Uruguay	473	17	260	206	7	0	1.5
89	Cote d'Ivoire	444	0	389	52	3	0	0.7
90	Burkina Faso	443	0	273	146	24	0	5.4
91	Niger	438	28	386	41	11	0	2.5
92	Bangladesh	424	94	364	33	27	6	6.4
93	Albania	416	7	211	182	23	0	5.5
94	Honduras	382	39	352	7	23	0	6
95	Taiwan*	382	2	285	91	6	1	1.6
96	Ghana	378	0	368	4	6	0	1.6
97	Jordan	372	0	195	170	7	0	1.9
98	Malta	350	13	332	16	2	0	0.6
99	San Marino	344	11	260	50	34	0	9.9
100	Mauritius	318	4	286	23	9	2	2.8
101	Nigeria	305	17	240	58	7	0	2.3
102	Kyrgyzstan	298	18	258	35	5	1	1.7
103	Bolivia	268	4	247	2	19	1	7.1
104	West Bank and Gaza	267	4	220	45	2	1	0.7
105	Senegal	265	15	126	137	2	0	0.8
106	Vietnam	257	2	113	144	0	0	0
107	Montenegro	255	3	249	4	2	0	0.8
108	Kosovo	250	66	191	52	7	2	2.8
109	Georgia	234	16	177	54	3	0	1.3
110	Congo (Kinshasa)	215	35	182	13	20	2	9.3
111	Guinea	212	18	197	15	0	0	0
112	Sri Lanka	190	0	129	54	7	0	3.7

Table 6: Cases by Country (10 Apr 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
113	Kenya	189	5	160	22	7	0	3.7
114	Venezuela	171	0	78	84	9	0	5.3
115	Djibouti	150	15	113	36	1	1	0.7
116	Brunei	136	1	36	99	1	0	0.7
117	Paraguay	129	5	105	18	6	1	4.7
118	Guatemala	126	31	106	17	3	0	2.4
119	Cambodia	119	0	47	72	0	0	0
120	Rwanda	118	8	111	7	0	0	0
121	El Salvador	117	14	96	15	6	0	5.1
122	Trinidad and Tobago	109	0	100	1	8	0	7.3
123	Madagascar	93	0	82	11	0	0	0
124	Monaco	90	6	84	5	1	0	1.1
125	Mali	87	13	58	22	7	0	8
126	Liechtenstein	79	1	23	55	1	0	1.3
127	Togo	76	3	48	25	3	0	3.9
128	Barbados	67	1	52	11	4	1	6
129	Ethiopia	65	9	58	4	3	1	4.6
130	Jamaica	63	0	46	13	4	0	6.3
131	Congo (Brazzaville)	60	0	50	5	5	0	8.3
132	Uganda	53	0	53	0	0	0	0
133	Gabon	44	0	42	1	1	0	2.3
134	Bahamas	42	1	29	5	8	0	19
135	Zambia	40	1	13	25	2	1	5
136	Guyana	37	0	23	8	6	0	16.2
137	Liberia	37	6	29	3	5	1	13.5
138	Guinea-Bissau	36	0	36	0	0	0	0
139	Benin	35	9	29	5	1	0	2.9
140	Eritrea	34	1	34	0	0	0	0
141	Tanzania	32	7	24	5	3	2	9.4
142	Haiti	31	1	29	0	2	0	6.5
143	Burma	27	4	22	2	3	0	11.1
144	Libya	24	0	15	8	1	0	4.2
145	Somalia	21	9	19	1	1	0	4.8
146	Mozambique	20	3	18	2	0	0	0
147	Angola	19	0	15	2	2	0	10.5
148	Antigua and Barbuda	19	0	17	0	2	0	10.5
149	Maldives	19	0	6	13	0	0	0
150	Syria	19	0	13	4	2	0	10.5
151	Equatorial Guinea	18	0	15	3	0	0	0
152	Sudan	17	2	13	2	2	0	11.8
153	Dominica	16	1	11	5	0	0	0
154	Fiji	16	1	16	0	0	0	0
155	Laos	16	0	16	0	0	0	0
156	Mongolia	16	0	12	4	0	0	0
157	Namibia	16	0	13	3	0	0	0
158	Saint Lucia	15	1	14	1	0	0	0
159	Grenada	14	2	14	0	0	0	0
160	Botswana	13	0	12	0	1	0	7.7
161	Zimbabwe	13	2	10	0	3	0	23.1
162	Eswatini	12	0	5	7	0	0	0
163	Saint Kitts and Nevis	12	1	12	0	0	0	0
164	Saint Vincent and the Grenadines	12	0	11	1	0	0	0
165	Chad	11	0	9	2	0	0	0
166	Seychelles	11	0	11	0	0	0	0
167	Belize	10	1	8	0	2	1	20
168	Suriname	10	0	5	4	1	0	10
169	Malawi	9	1	8	0	1	0	11.1
170	MS Zaandam	9	0	7	0	2	0	22.2
171	Nepal	9	0	8	1	0	0	0
172	Central African Republic	8	0	8	0	0	0	0
173	Holy See	8	0	6	2	0	0	0
174	Sierra Leone	8	1	8	0	0	0	0
175	Cabo Verde	7	0	5	1	1	0	14.3
176	Mauritania	7	0	4	2	1	0	14.3
177	Nicaragua	7	0	6	0	1	0	14.3
178	Bhutan	5	0	3	2	0	0	0
179	Gambia	4	0	1	2	1	0	25
180	Sao Tome and Principe	4	0	4	0	0	0	0

Table 6: Cases by Country (10 Apr 2020 UTC) *(continued)*

	country	confirmed	new.confirmed	current.confirmed	recovered	deaths	new.deaths	death.rate
181	South Sudan	4	1	4	0	0	0	0
182	Western Sahara	4	0	4	0	0	0	0
183	Burundi	3	0	3	0	0	0	0
184	Papua New Guinea	2	0	2	0	0	0	0
185	Timor-Leste	2	1	1	1	0	0	0
186	Yemen	1	1	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!