Corona Analysis

Samuel Knapp - samuel.k@gmx.de

Download data

Data were downloaded from the github repository of the Johns Hopkins University. These are the same data, from which the famous GIS world map is created. See: https://github.com/CSSEGISandData/COVID-19

```
base_url = "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master"
cases_url = paste(
  base_url,
  "/csse_covid_19_data/csse_covid_19_time_series",
  "/time_series_covid19_confirmed_global.csv", sep="")
cases <- fread(cases_url)
# turn into long format
datecols <- names(cases)[-c(1:4)]
idcols <- names(cases)[c(1:4)]
cases <- melt(cases,id.vars=idcols,measure.vars=datecols,variable.name="date")
cases$action<-"confirmed"

#add death
death_data_url = paste(
  base_url,
  "/csse_covid_19_data/csse_covid_19_time_series",
  "/time_series_covid19_deaths_global.csv", sep="")
death <- fread(death_data_url)
datecols <- names(death)[-c(1:4)]
idcols <- names(death)[c(1:4)]
death <- melt(death,id.vars=idcols,measure.vars=datecols,variable.name="date")
death$action<-"death"
# bind
cases <- rbind(cases,death)

##################################
# some renaming
setnames(cases,"Country/Region","country")
setnames(cases,"Province/State","province")
setnames(cases,"value","number")

# format date
cases[,date:=as.Date(date,tryFormats = c("%m/%d/%Y"))]
# number of days after first date in table
cases[,days:=as.numeric(date-min(date))]



# as Hong Kong is listed as country China, take out Hong Kong and set as country
```

```
cases[province=="Hong Kong",country:="Hong Kong"]


# sum over provinces for China
chinadat <- cases[country=="China"]
chinadat <- chinadat[,.(number=sum(number)),.(date,days,action,country)]
cut <- cases[!country=="China"]
cases <- rbind(cut,chinadat,fill=T)

# remove cruise ships
cases <- cases[!(country%in%c("Diamond Princess","MS Zaandam"))]

# remove * in Taiwan*
cases[country=="Taiwan*",country:="Taiwan"]

# some countries have outside provinces, mainland is identified by empty province

counts <- c("France","United Kingdom","Denmark","Netherlands","Canada")
for (counti in counts)
{
  changecases <- cases[country==counti& province==""]
  cases <- cases[country!=counti]
  cases <- rbind(cases,changecases)
}


# add population from
# https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population
pop <- fread("pop.csv")
# unique(cases$country)[!(unique(cases$country) %in% pop$country)]
cases <- merge(cases,pop,by="country",all.x=T)

# check how many countries and population
# contsum <- cases[,.(pop=unique(population)),country]
# nrow(contsum)
# sum(contsum$pop,na.rm=T)
```

Newest date

```
max(cases$date)
```

```
## [1] "0020-05-26"
```

Selected countries

```
# number of countries to be plotted
nplot <- 20

# countries that shall definetly be selected
countadd <- c("Germany","Switzerland","Hong Kong",
              "Singapore","Sweden","Austria","Greece")
# set actioni to confirmed or death
actioni <- "confirmed"
```

```
# table with most cases for given action
countover <- cases[action==actioni,.(maxnumber=max(number)),country]
countmost <- countover[order(-maxnumber),country]
# remove chosen countries
countmost <- countmost[!(countmost %in% countadd)]
countsel <- c(countadd,countmost[1:(nplot-length(countadd))])
#countsel
countsel <- sort(countsel)
```

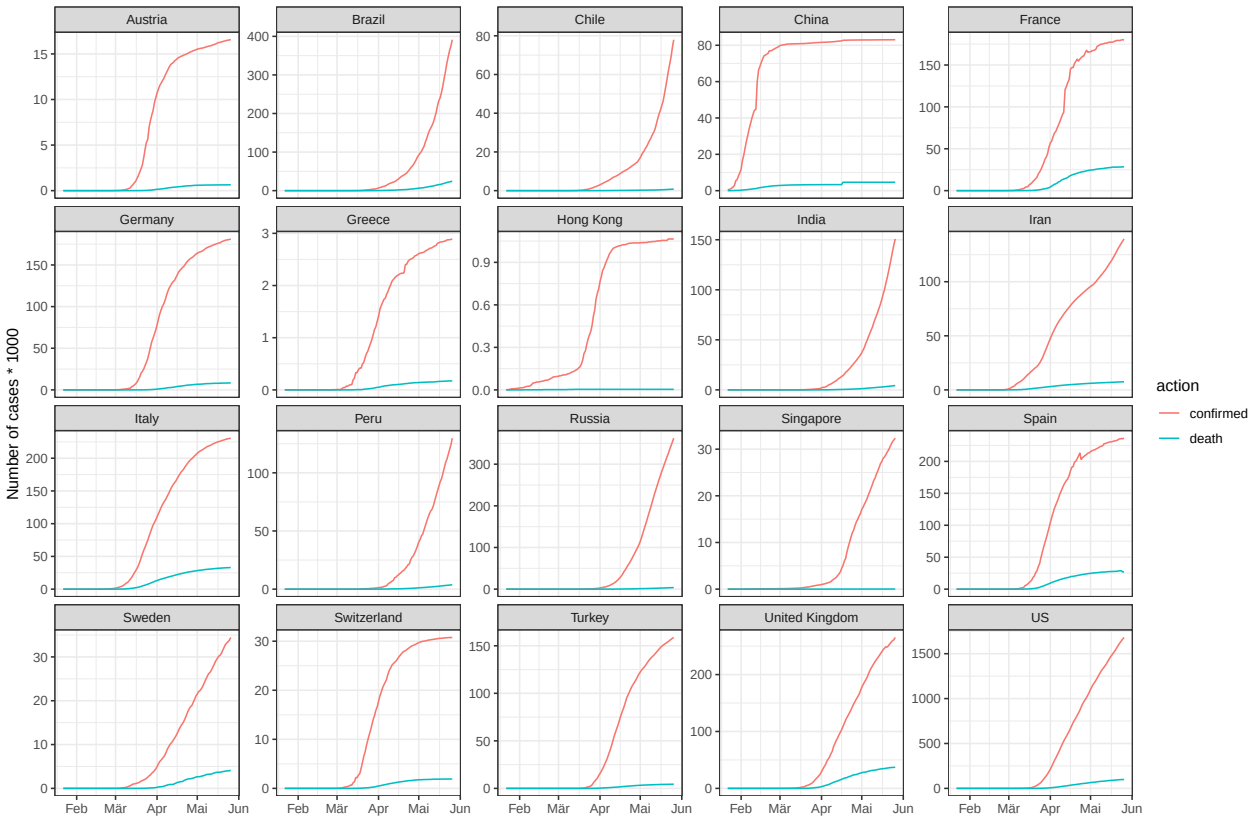
The following countries were set to be included: Germany, Switzerland, Hong Kong, Singapore, Sweden, Austria, Greece. Additionally, 13 countries with the highest number of confirmed cases were added.

Actual numbers

The number of confirmed and death cases for each day.

Hong Kong and Singapore both show two phases of linear growth. South Korea first had an exponential growth and then turned into linear growth.

```
# # cases per country, both confirmed and death
ggplot(cases[country%in%countsel],aes(date,number/1000,colour=action))+
  facet_wrap(vars(country),scales="free_y")+
  labs(x="",y="Number of cases * 1000")+
  geom_line()+
  theme_bw()
```



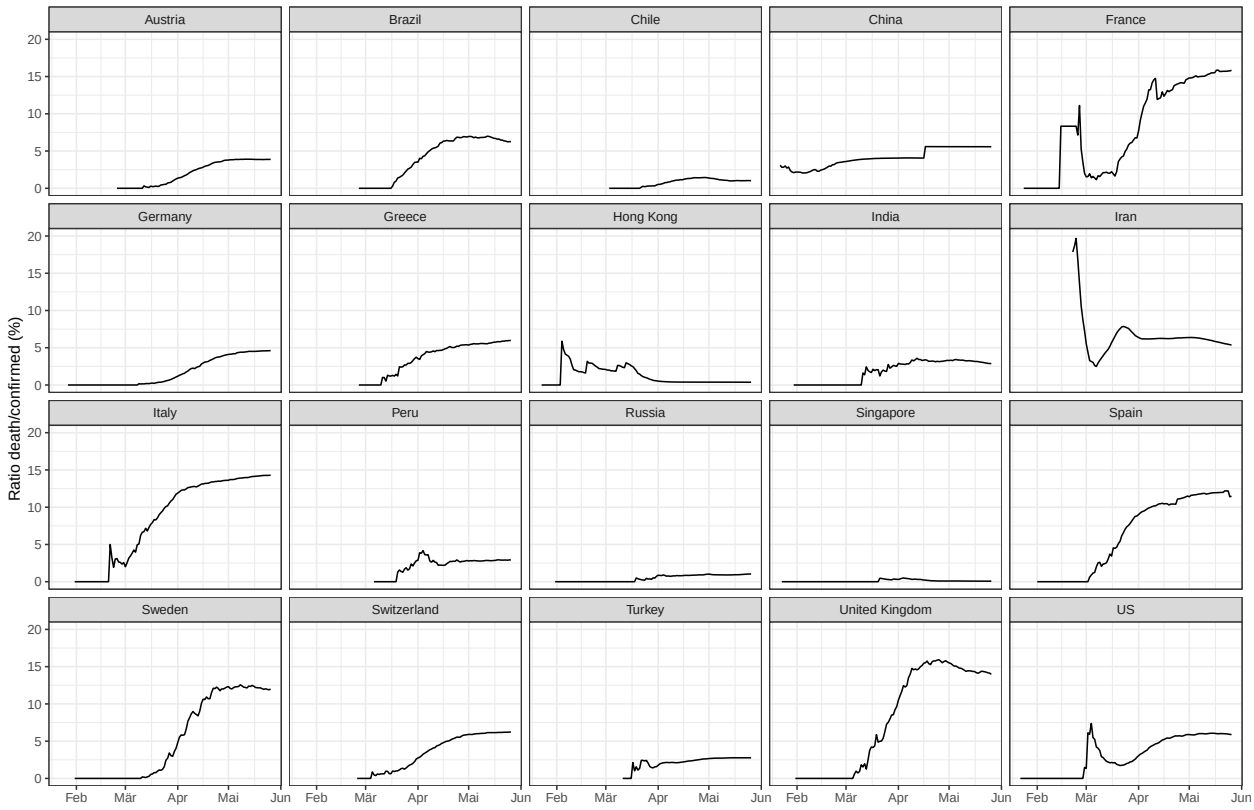
Ratio of death to infected

Simply the ratio of reported deaths divided by number of confirmed cases for each day. Interesting to see that this ratio increases i most countries. A particularly sharp increase can be observed for countries that start to struggle: Italy, Spain, and Belgium. However, this calculation is probably too simple, as it does not take account of recovered cases.

Note, that in Italy last points are not in plot anymore.

```
#
# ################
# # use wide form
# wide form with columns for confirmed and death
casw <- dcast(cases,country+province+Lat+Long+date~action,value.var="number")

casw[,ratio:=death/confirmed]
ggplot(casw[country%in%countsel],aes(date,ratio*100))+
  facet_wrap(vars(country),#scales="free_y"
             )+
  geom_line()+
  labs(x="",y="Ratio death/confirmed (%)")+
  lims(y=c(0,20))+
  theme_bw()
```




```
#
#
```

New cases

Simply the daily increase of confirmed cases.

```
cases[,newcases:=number-shift(number),country]
cases[,relnewcases:=number/shift(number),country]
# remove ones
cases <- cases[relnewcases!=1]
```

Relative increase per day

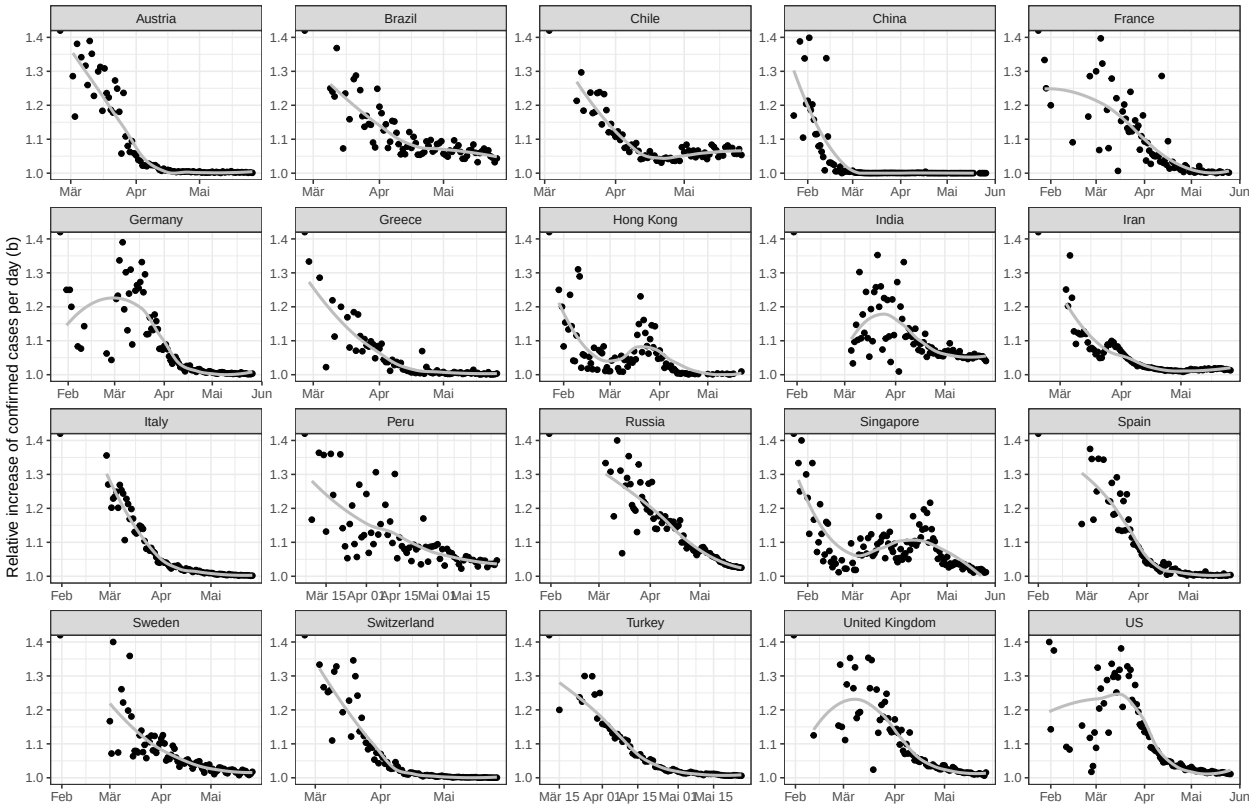
A relative increase of e.g. b=1.2 indicates that the number of confirmed cases increases by 20% in one day, e.g. from 1000 to 1200. This number (b) can be related to the number of days needed for doubling the number of confirmed cases by $b^x = 2$, with x as the number of days. The following shows the relation of b to x. The sometimes mentioned aim of a doubling time of ten days thus corresponds to $b \approx 1.07$.

```
b=seq(1.05,1.4,0.05)
tab <- data.frame(b=b,
                  NumberOfDays=log(2)/log(b))
kable(tab,digits=2)
```

b	NumberOfDays
1.05	14.21
1.10	7.27
1.15	4.96
1.20	3.80
1.25	3.11
1.30	2.64
1.35	2.31
1.40	2.06

While the relative increase was at around $b \approx 1.3$ to $b \approx 1.4$ (meaning a doubling of confirmed cases every 2 to 2.6 days), this rate has dropped to around $b \approx 1.1$ in most countries. This might be most probably due to the imposed measures.

```
ggplot(cases[country%in%countsel&action=="confirmed"],aes(date,relnewcases))+
  facet_wrap(vars(country),scales = "free")+
  geom_point()+
  geom_smooth(col="grey",se=F)+
  #geom_smooth(col="grey",se=F,method="lm",formula = y ~ x + I(x^2))+
  #scale_y_continuous(expand = expand_scale(mult = c(0.0001, .2))) +
  labs(x="",y="Relative increase of confirmed cases per day (b)")+
  lims(y=c(1,1.4))+
  theme_bw()
```



Absolute increase

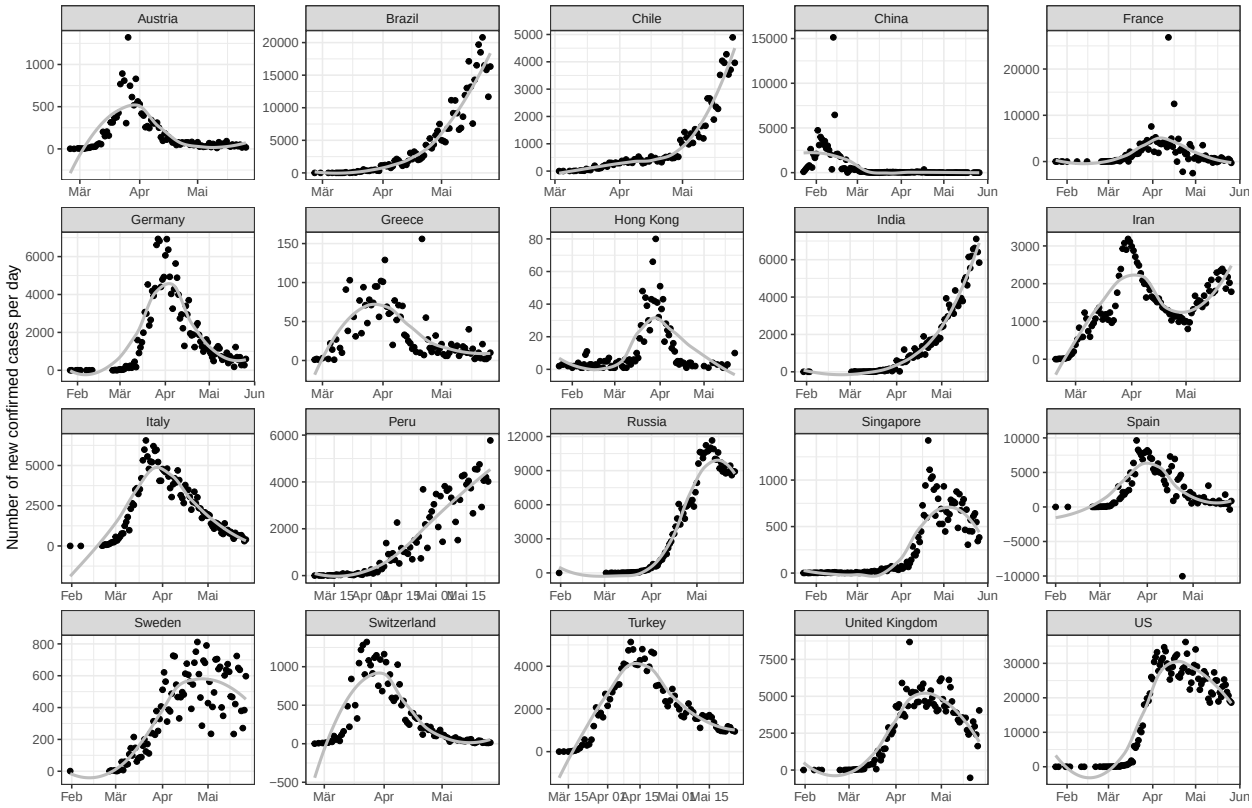
Absolute numbers

For the capacity of the health systems, it is more important to look at the absolute numbers of new confirmed cases. The aim should be to get a constant number of new cases at a niveau which can be handled by the health system.

Austria and Switzerland have managed to drop the increase to a constant level. In many other countries (also Germany) the daily increases are still increasing.

In South Korea it can be nicely see how the exponential growth was lowered to a linear growth. This could/should be the aim...

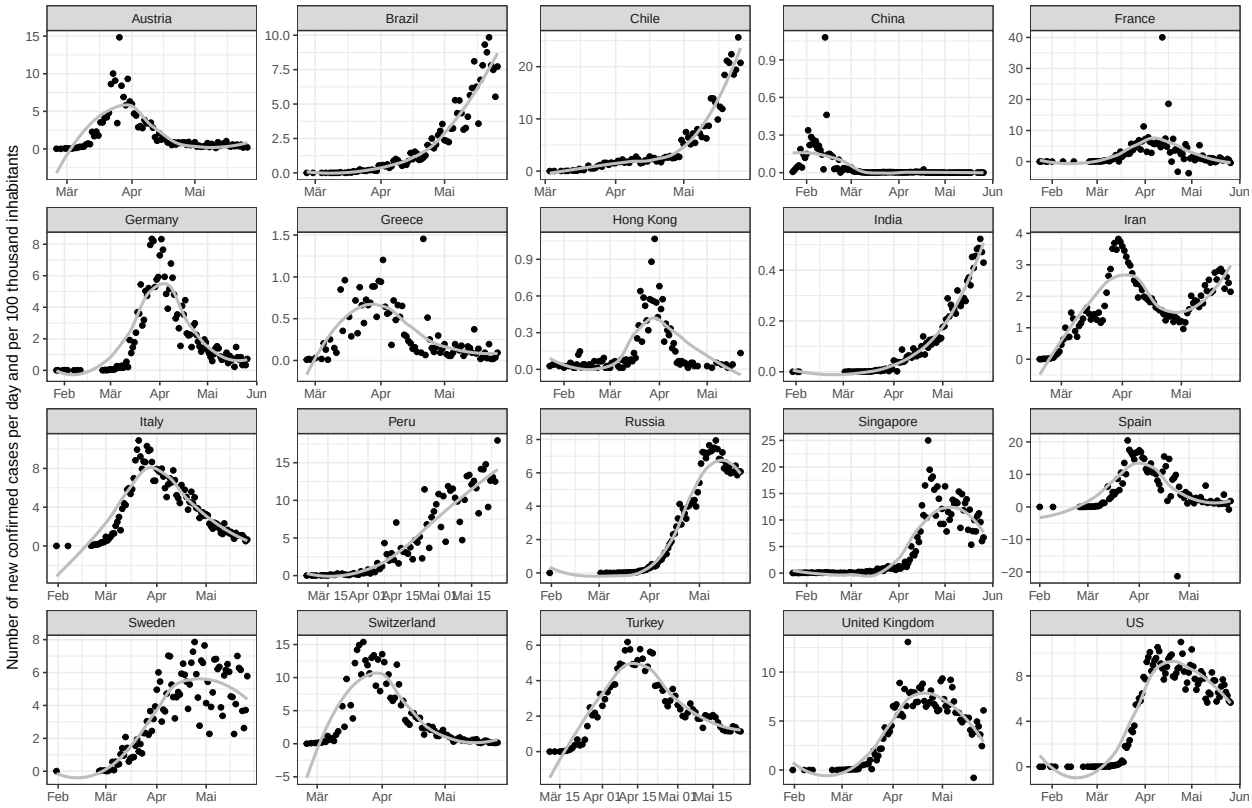
```
ggplot(cases[country%in%countsel&action=="confirmed"],aes(date,newcases))+
  facet_wrap(vars(country),scales = "free")+
  #geom_hline(aes(yintercept=newvent))+
  geom_point()+
  geom_smooth(col="grey",se=F)+
  labs(x="",y="Number of new confirmed cases per day")  +
  #lims(y=c(1,1.5))+
  theme_bw()
```



As number per 100 thousand inhabitants

Relating the absolute number of new cases to the total population per country. All in similar range, but still different. Not sure about the interpretation.

```
cases[,newcases_pop:=newcases/population]
ggplot(cases[country%in%countsel&action=="confirmed"],aes(date,newcases_pop*100000))+
  facet_wrap(vars(country),scales = "free")+
  #geom_hline(aes(yintercept=newvent))+
  geom_point()+
  geom_smooth(col="grey",se=F)+
  #scale_y_continuous(expand = expand_scale(mult = c(0.0001, .2))) +
  labs(x="",y="Number of new confirmed cases per day and per 100 thousand inhabitants")  +
  #lims(y=c(1,1.5))+
  theme_bw()
```



Percentage of population

Number of confirmed cases (most recent day) divided by the total population.

```
ratab <- cases[action=="confirmed",.(RatioPercent=max(number)/max(population)*100),.(country)]
kable(head(ratab[order(-RatioPercent)],50),digits=3)
```

country	RatioPercent
San Marino	1.984
Qatar	1.718
Holy See	1.502
Andorra	0.984
Luxembourg	0.651
Bahrain	0.607
Singapore	0.567
Kuwait	0.511
US	0.510
Ireland	0.503
Spain	0.502
Belgium	0.499
Iceland	0.495
Chile	0.408
Belarus	0.404
Peru	0.404
United Kingdom	0.399
Maldives	0.384
Italy	0.383

country	RatioPercent
Switzerland	0.358
Sweden	0.333
United Arab Emirates	0.314
Portugal	0.302
Moldova	0.272
Panama	0.271
France	0.269
Netherlands	0.261
Monaco	0.256
Armenia	0.250
Russia	0.247
Djibouti	0.229
Saudi Arabia	0.224
Germany	0.218
Ecuador	0.214
Liechtenstein	0.212
Denmark	0.196
Turkey	0.191
Austria	0.186
Brazil	0.185
Israel	0.183
Oman	0.174
Iran	0.167
Serbia	0.161
Norway	0.156
Dominican Republic	0.147
Estonia	0.138
Malta	0.124
Finland	0.120
Cyprus	0.107
Gabon	0.103