```
url = "https://raw.githubusercontent.com/COVID19Tracking/covid-tracking-data/master/data/
states_daily_4pm_et.csv"
download.file(url,destfile="covid.csv")
base = read.csv("covid.csv")
```

Unfortunately, there is no information about the population. That we can find on wikipedia. But in that table, the state is given by its full name (and the symbol in the previous dataset). So we new also to match the two datasets properly,

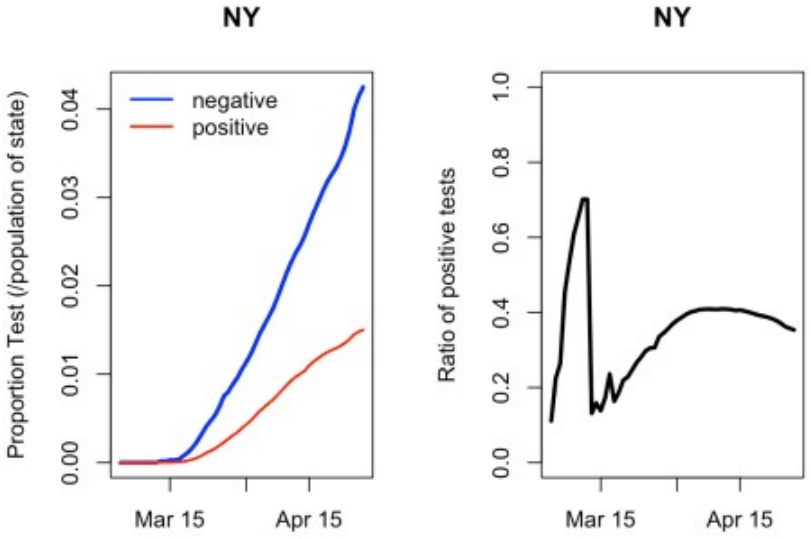
```
url="https://en.wikipedia.org/wiki/List_of_states_and_territories_of_the_United_States_by_population"
download.file(url,destfile = "popUS.html")
#pas contaminé 2/3 R=3
library(XML)
tables=readHTMLTable("popUS.html")
T=tables[[1]][3:54,c("V3","V4")]
names(T)=c("state","pop")
url="https://en.wikipedia.org/wiki/List_of_U.S._state_abbreviations"
download.file(url,destfile = "nameUS.html")
tables=readHTMLTable("nameUS.html")
T2=tables[[1]][13:63,c(1,4)]
names(T2)=c("state","symbol")
T=merge(T,T2)
T$population = as.numeric(gsub(",", "", T$pop, fixed = TRUE))
names(base)[2]="symbol"
base = merge(base,T[,c("symbol","population")])
```

Now our dataset is fine… and we can get a function to plot the number of people tested in the U.S. (cumulated). Here, we distinguish between the positive and the negative,

```
drawing = function(st ="NY"){
sbase=base[base$symbol==st,c("date","positive","negative","population")]
sbase$DATE = as.Date(as.character(sbase$date),"%Y%m%d")
sbase=sbase[order(sbase$DATE),]
par(mfrow=c(1,2))
plot(sbase$DATE,(sbase$positive+sbase$negative)/sbase$population,ylab="Proportion
Test (/population of state)",type="l",xlab="",col="blue",lwd=3)
lines(sbase$DATE,sbase$positive/sbase$population,col="red",lwd=2)
legend("topleft",c("negative","positive"),lwd=2,col=c("blue","red"),bty="n")
title(st)
plot(sbase$DATE,sbase$positive/(sbase$positive+sbase$negative),ylab="Ratio of
positive tests",ylim=c(0,1),type="l",xlab="",col="black",lwd=3)
title(st)}
```

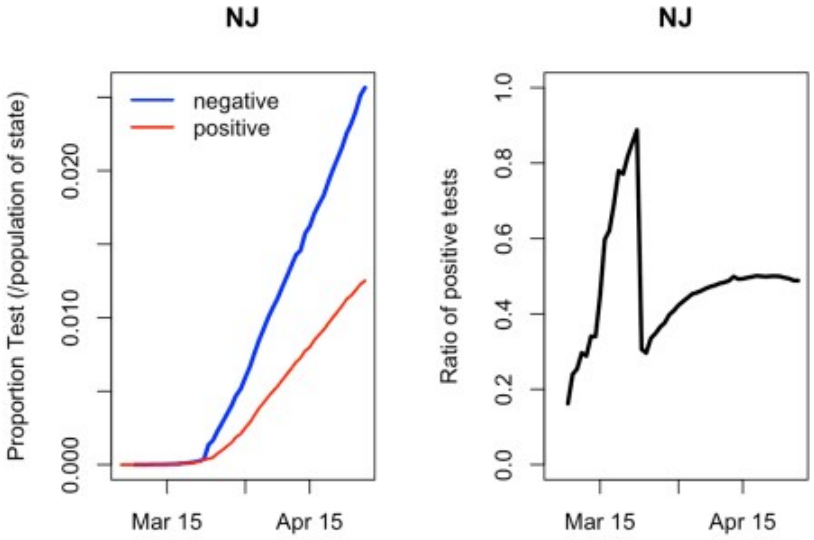
Let us start with New York

```
drawing("NY")
```



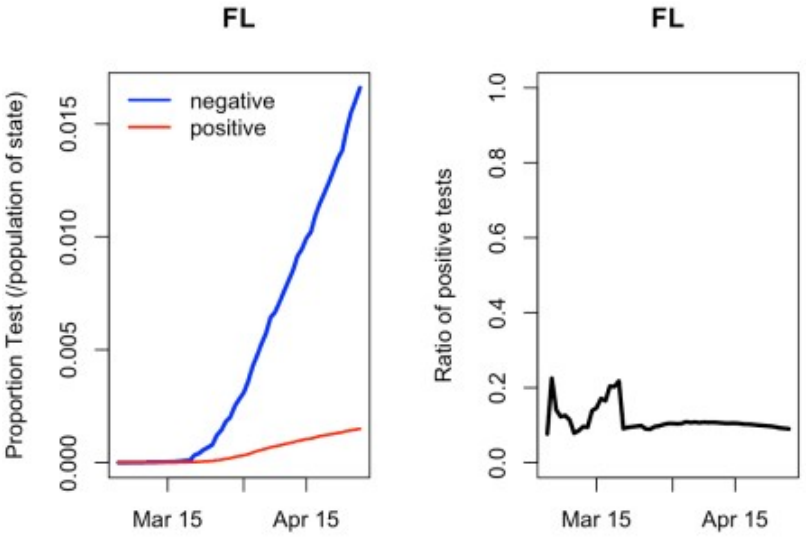
As at now, 4% of the entiere population got tested… over 6 weeks…. The graph on the right is the proportion of people who tested positive… I won't get back on that one here today, I keep it for our work. In New Jersey, we got about 2.5% of the entiere population tested, overall,

```
1 drawing("NJ")
```



Let us try a last one, Florida

```
1 drawing("FL")
```



As at today, it is 1.5% of the population, over 6 weeks. Overall, in the U.S. less than 0.1% people are tested, on a daily basis. Which is far from the 1.5% in the U.K. scenarios. Now, here come the second question,

- **what are we actually testing for ?**