Big Data Analytics

Before our first week

UNIVERSITÄT LUZERN

Before our first Lab class

If you **have a laptop** with you:

1) Install the latest version of R
2) For **Windows platforms**: install the latest version of Rtools (i.e., Rtools 4) from here (https://cran.r-project.org/bin/windows/Rtools/)

For **OS X**, do the following:

a) First try to install Quanteda directly
b) If you fail in doing that, install XCode from the App Store

Before our first Lab class

c) To install XCode, follow these simple rules:

1 Access to “Apple Developer”

https://developer.apple.com/download/more/

(You need Apple ID and password)

2 Insert “Xcode” in “Search Downloads” located on the left side of the page.

3 Choose “Xcode 12” and download.

4 After finishing download, click "Finder" and then "download." Double click “Xcode 12”. It may take a while to open this file

d) If you have problems to install the **latest version** of Xcode, **uses an earlier one**, such as Xcode 9!

Before our first Lab class

e) To make things even more complicated for Mac users: the latest R could not be compatible with the most recent Xcode. In that case, they the second most recent version of R

5 In case somebody encounters *still* some problems: you could need also to install the fortran compiler as described here: https://mac.r-project.org/tools/ (GNU Fortran compiler -> Apple silicon Macs) for the quanteda/quanteda.textmodels package

Before our first Lab class

Install the following packages by running these lines (1):

```
install.packages(‘quanteda', repos='http://cran.us.r-project.org')
install.packages(‘quanteda.textstats', repos='http://cran.us.r-project.org')
install.packages(‘quanteda.textplots', repos='http://cran.us.r-project.org')
install.packages(‘readtext', repos='http://cran.us.r-project.org')
install.packages(‘devtools', repos='http://cran.us.r-project.org')
devtools::install_github("quanteda/quanteda.corpora")
devtools::install_github("quanteda/quanteda.textmodels")
```

Before our first Lab class

Install the following packages by running these lines (2):

```
install.packages('ggplot2', repos='http://cran.us.r-project.org')
install.packages('SnowballC', repos='http://cran.us.r-project.org')
install.packages('corrplot', repos='http://cran.us.r-project.org’)
install.packages("psych", repos='http://cran.us.r-project.org')
install.packages("PerformanceAnalytics", repos='http://cran.us.r-project.org')
install.packages("stringr", repos='http://cran.us.r-project.org’)
install.packages("dplyr", repos='http://cran.us.r-project.org')
install.packages("gridExtra", repos='http://cran.us.r-project.org')
```

Before our first Lab class

```
install.packages("ldatuning", repos='http://cran.us.r-project.org')
install.packages("lubridate", repos='http://cran.us.r-project.org')
install.packages("tidytext", repos='http://cran.us.r-project.org’)
install.packages("stm", repos='http://cran.us.r-project.org’)
install.packages("topicdoc", repos='http://cran.us.r-project.org’)
install.packages("topicmodels", repos='http://cran.us.r-project.org')
```

If you have a Mac and you have problems to install the package «topicmodels», please read this:

https://kenbenoit.net/how-to-install-the-r-package-topicmodels-on-os-x/

Before our first Lab class

```
install.packages("igraph", repos='http://cran.us.r-project.org')
install.packages("servr", repos='http://cran.us.r-project.org')
library(devtools)
devtools::install_github("mroberts/stmBrowser",dependencies=TRUE)
install.packages("syuzhet", repos='http://cran.us.r-project.org')
install.packages ("reshape2", repos='http://cran.us.r-project.org')
install.packages ("plyr", repos='http://cran.us.r-project.org')
```

Before our first Lab class

```
install.packages("maps", repos='http://cran.us.r-project.org')
install.packages("leaflet", repos='http://cran.us.r-project.org')
devtools::install_github("hadley/emo")
```

IMPORTANT!!!

Before using rtweet

We will also use the `rtweet` package: so start to install it!

install.packages("rtweet", repos='http://cran.us.r-project.org')

install.packages("httpuv", repos='http://cran.us.r-project.org')

install.packages("ggmap", repos='http://cran.us.r-project.org’)

P.S. you need to have a Twitter account! Open it: it is for free! And you do not need to post it there anything if you do not want!

Before using rtweet

Then open an R session and type the following commands. Plz let me know if you are able (or not) to download the 10 tweets:

```
library(rtweet)

library(httpuv)

rt <- search_tweets( "#rstats", n = 10,
include_rts = FALSE)

print(rt$text[1:10])
```

Before using rtweet

Then read this link:

https://cran.r-project.org/web/packages/rtweet/vignettes/auth.html

And learn how to retrieve the information required to create your own token

Basically you need the following info:

```
app = [your_twitter_api_app]

consumer_key = [your_api_consumer_key]

consumer_secret = your_api_consumer_secret]
```

Optional

Before we can start geocoding data, we need to obtain an API key from Google. Go to the registration page, and follow the instructions (select all mapping options)

The **geocoding API** is a free service, but you nevertheless need to associate a credit card with the account.

Please note that the Google Maps API is not a free service. There is a free allowance of 40,000 calls to the geocoding API per month, and beyond that calls are $0.005 each

This implies that basically you have a monthly free limit of $200 (more than enough…)

To register you need to have: a) a gmail account; b) a credit card

Optional

After you finish the registration (if everything hopefully works fine!) Google gives you back an API number. Save it!

Then type:

```
library(ggmap)
register_google(key = “NUMBER OF YOUR GOOGLE API!")
geocode(c("White House", "Uluru"))
You should get this result back:
# A tibble: 2 x 2
    lon   lat
  <dbl> <dbl>
1 -77.0  38.9
2 131.  -25.3
```

Optional

If you are able to get the Google API, but GGMAP does not get any results back, enable the “geocoding app” in your console developer. Check how to enable GOOGLE API **here**