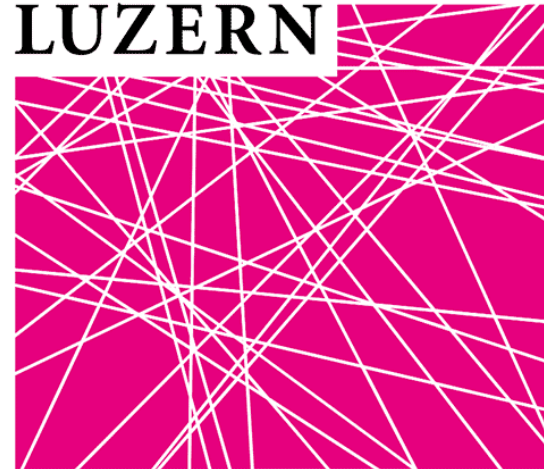


Big Data Analytics: Overview



UNIVERSITÄT
LUZERN

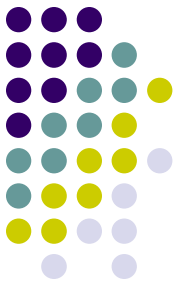


Boring (but needed) information



- Four days (10/11/18/19 March)
- Since 9:15 am to 4:45 pm
- We will take one short break during the morning, one short break during the afternoon
- Lunch break: 12-13 pm
- Office Hour: plz write me in advance to fix an appointment!
- Email: luigi.curini@unimi.it

Boring (but needed) information



- Each class will have a theoretical discussion followed by a lab class, both in the morning as well as in the afternoon
- All the slides, scripts and datasets that we employ during our classes (for my part) will be made available the day before each lecture at the following URL:

<http://www.luigicurini.com/big-data-analytics-lumacss.html>

Boring (but needed) information



This course is aimed to:

- ✓ Introduce you some of the new methods developed within the literature in the last years to analyze texts
- ✓ Offer you guidelines on how to effectively (and practically) use text methods for social scientific research

Boring (but needed) information



Plan of the course (in the ideal world...):

1. An introduction to text analytics. When? 10 March, morning
2. From words to positions: scaling algorithms (unsupervised & supervised). When? 10 March, afternoon
3. From words to issues: unsupervised classification algorithms. When? 11 March, morning
4. From words to issues: dictionary & semi-supervised classification algorithms. When? 11 March, afternoon
5. From words to issues: supervised algorithms (first part). When? 18 March, morning
6. From words to issues: supervised algorithms (second part). When? 18 March, afternoon
7. From words to issues: supervised algorithms & cross-validation. When? 19 March, morning
8. From (bag of) words to word embeddings. When? 19 March, afternoon

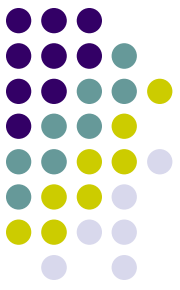
Boring (but needed) information



How to evaluate you???

90% Course grades will be based on 1 home-assignment

10% Course grades will be based on on-line class-participation



Before our first Lab class

Install the following packages by running these lines:

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

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

```
install.packages('quanteda.textstats', repos='http://cran.us.r-project.org')
```

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

```
devtools::install_github("quanteda/quanteda.corpora")
```

```
devtools::install_github("quanteda/quanteda.textmodels")
```

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

Before our first Lab class



```
install.packages('ggplot2', repos='http://cran.us.r-project.org')
```

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

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

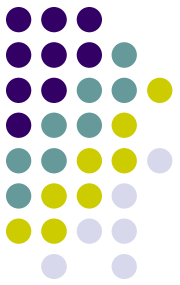
```
install.packages("PerformanceAnalytics",  
repos='http://cran.us.r-project.org')
```

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

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

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

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

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/>)
- 3) 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.

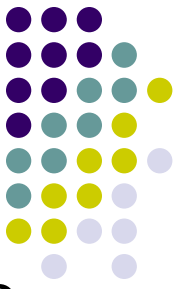
4 After finishing download, click "Finder" and then "download."
Double click “Xcode 12”. It may take a while to open this file

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



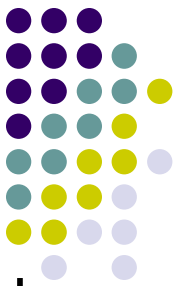
Before our first Lab class

We will use since the next week the `rtweet` package: so start to install it as well!

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

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

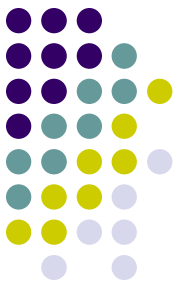
```
install.packages("ggmap", repos='http://cran.us.r-project.org')
```



Before our first Lab class

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])
```



Optional

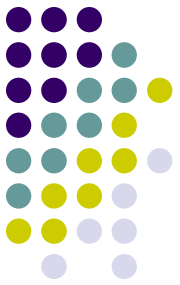
Before we can start geocoding data, we need to obtain an [API key from Google](#). Go to the registration page, and [follow the instructions](#) (select all mapping options)

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"))
```

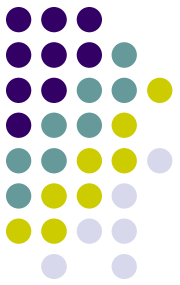
You should get this result back:

```
# 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](#)





**Please check that
everything is ok with
your R before **the beginning
of the course!****