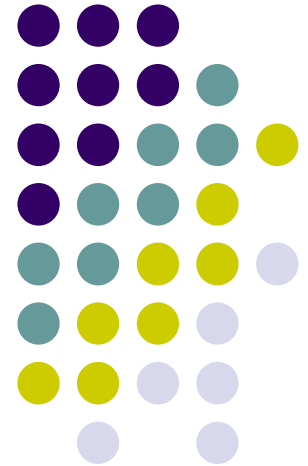
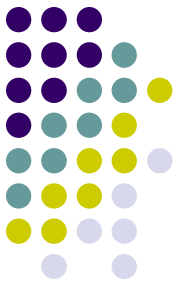


Applied Scaling & Classification Techniques in Political Science

Lecture 1 (Lab part)



Quanteda approach

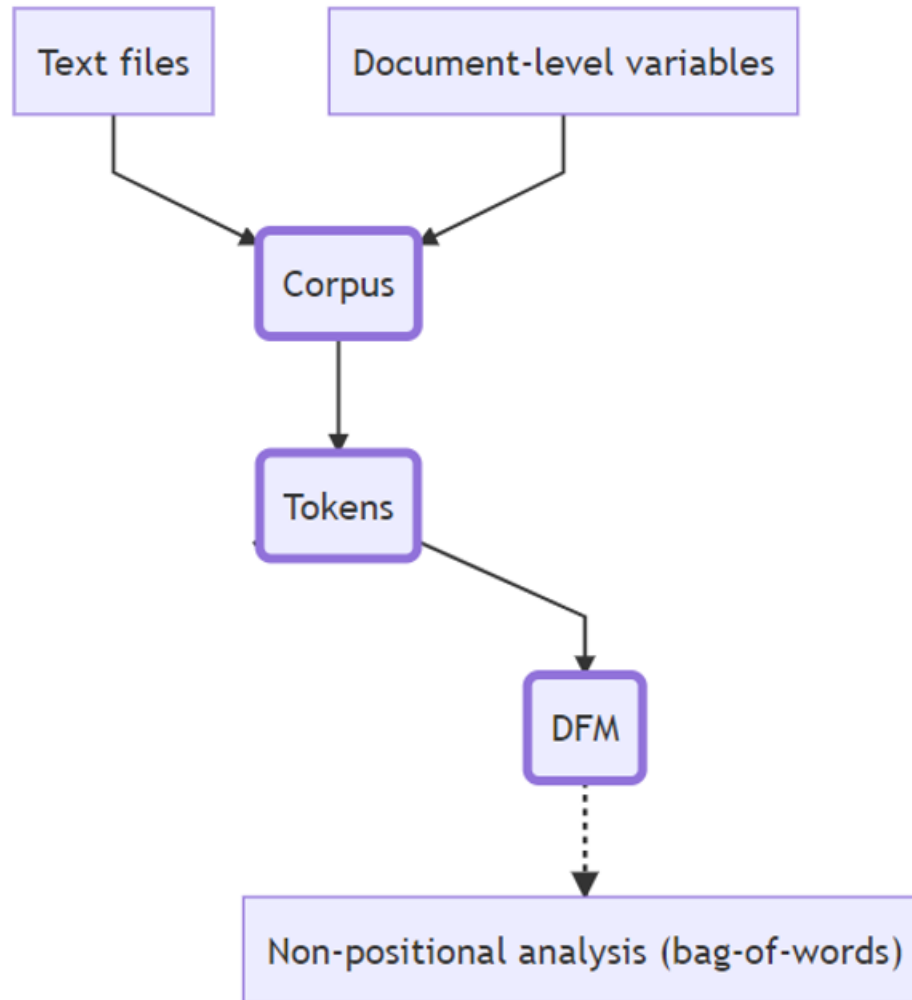


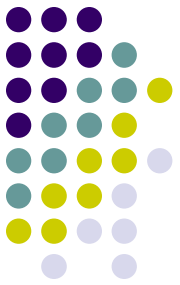
Quanteda package has three basic types of objects:

- *Corpus*: it saves character strings and variables in a data frame, by also combining texts with document-level variables (where available)
- *Tokens*: it stores tokens in a list of vectors (in a more efficient way than character strings), while still preserving positions of words. At this stage, you can apply pre-processing
- *Document-feature matrix (dfm)*: it represents frequencies of features in documents in a matrix. By doing it, it does not preserve information on positions of words within each text

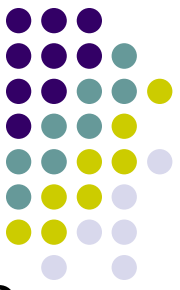
Text analysis with Quanteda via bag-of-words goes always through all those three types of objects

Quanteda approach





IMPORTANT!!!



Before using rtweet

We will use since the next week 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')
```

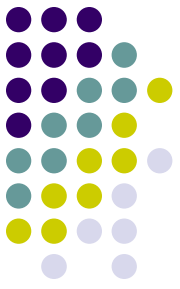
P.S. you need to have a Twitter account! Open it: it is for free! And you do not need to post it there anything if you do not want!



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>

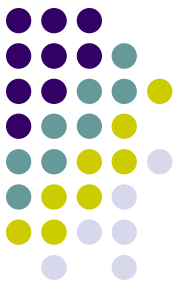
And learn how to retrieve the information required to create your own token

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:

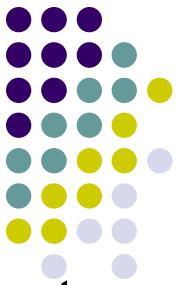
```
library(ggmap)
register_google(key = "NUMBER OF YOUR GOOGLE API!")
geocode(c("White House", "Uluru"))
```

You should get this result back:

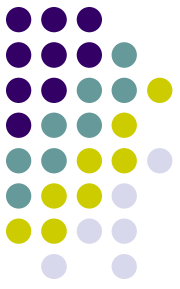
```
# A tibble: 2 x 2
  lon   lat
  <dbl> <dbl>
1 -77.0  38.9
2 131.  -25.3
```

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](#)

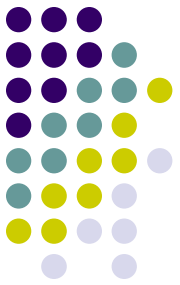


About social media data



When dealing with social-media data, you should be always very careful about privacy:

1. what kind of information can be ethically gathered about the users (public information)
2. how published data should look like to comply with privacy regulations (like the GDPR)
3. and what consequences violating the social network's terms of service may entail for the researcher



About social media data

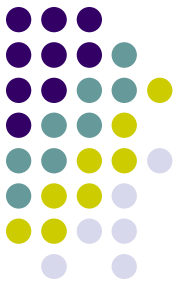
Some good readings about these points:

[Computational Research in the Post-API Age](#)

[What You Can Scrape and What Is Right to Scrape: A Proposal for a Tool to Collect Public Facebook Data](#)

([Supplementary Material](#) with a great code in R to use with Facebook!)

An interesting paper with a review of nine different free-of-charge and low-cost software tools for studying Twitter:
[“Free and Low-Cost Twitter Research Software Tools for Social Science”](#)



**Please check that
everything is ok with
rtweet before **next class!****