



ODYCCEUS Summer School
18th of September, 2019

Computational Tools for Analysing the Content and Structure of Social Interactions

Robin Lamarche-Perrin (CNRS / SU)

Eckehard Olbrich (MPG)

Yiannis Siglidis (SU)

Objectives of this lecture

During the first days of school, you

- Built huge corpora of data: maybe thousands, or even millions of documents?
- With many dimensions: individuals interacting with each others, sharing texts or opinions, through time and space?
- Began to explore the data?

These tools will help you

- Pursue data exploration
- Ask research questions...
- Begin to answer them!

Corpus of tweets about the migrant crisis of 2015

- Collected with DMI-TCAT
- Request: “migrant OR immigrant OR emigrant”
- From January 2015 to July 2016
- 30M documents: tweets, replies, retweets
- Meta-data: dates, hashtags, mentions, pictures, urls (and others)



okeh @fotopak · 11 Sep

[#SeaWatch3](#) | Update. "Sea-Watch 3 is currently detained in the Italian port of Lucata."

Captain who broke Italy's migrant ban says ready to rescue more people at sea, raps EU: [@Reuters](#)

data.cube: An R Library for the Exploration of Multidimensional Data

Robin Lamarche-Perrin

CNRS / Sorbonne Université

Mail: `Robin.Lamarche-Perrin@lip6.com`

Sources: <https://github.com/Lamarche-Perrin/data.cube>

Webpage: <https://www-complexnetworks.lip6.fr/~lamarche/>

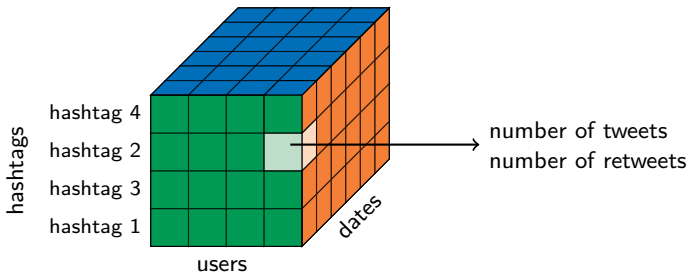
Explore multidimensional data...

...to discover statistical outliers

...to find macroscopic trends

...to make first hypotheses

data.cube of 3 dimensions
and 2 variables

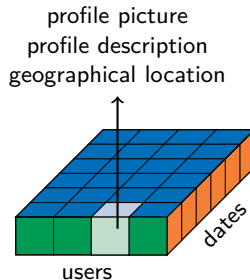
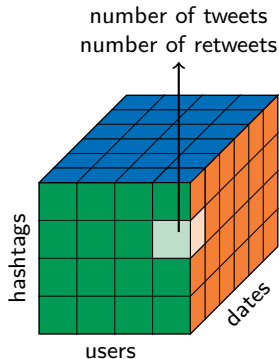


Explore multidimensional data...

...to discover statistical outliers

...to find macroscopic trends

...to make first hypotheses

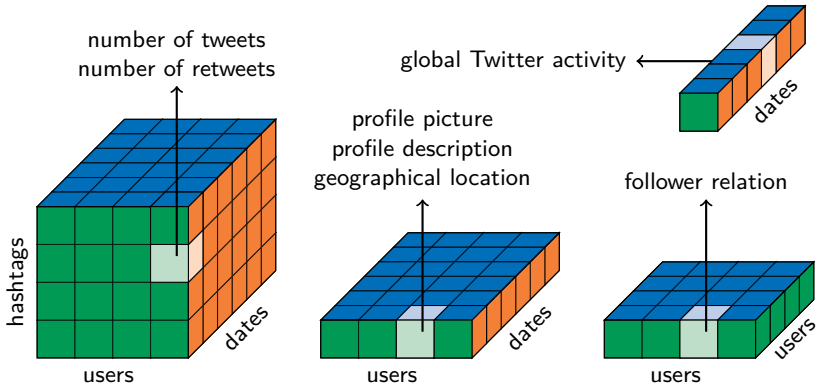


Explore multidimensional data...

...to discover statistical outliers

...to find macroscopic trends

...to make first hypotheses

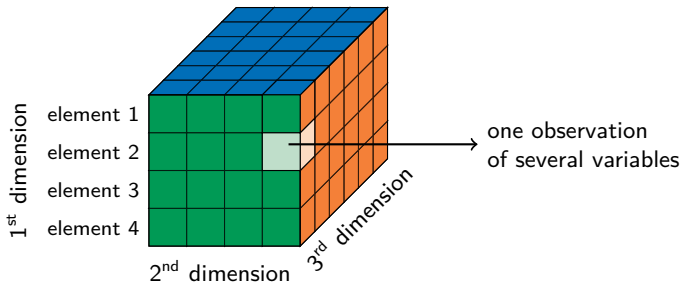


Explore multidimensional data...

...to discover statistical outliers

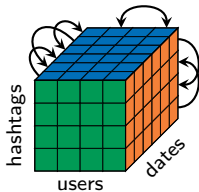
...to find macroscopic trends

...to make first hypotheses

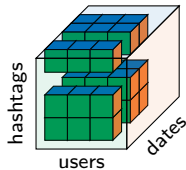


Operations on elements

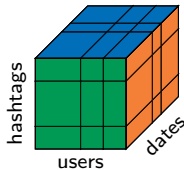
`arrange.elm`



`filter.elm`

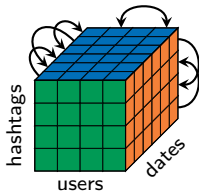


`group.elm`

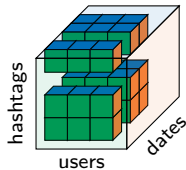


Operations on elements

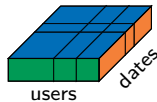
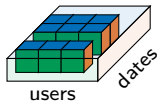
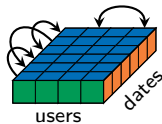
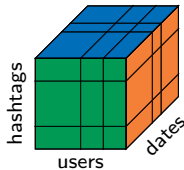
`arrange.elm`



`filter.elm`

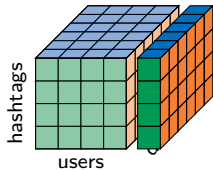


`group.elm`

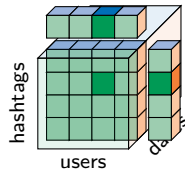


Operations on dimension and variables

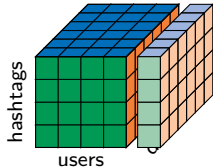
`select.dim`



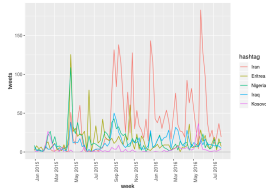
`mutate.var.model`



`spread.dim`

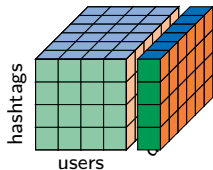


`ggplot.var`

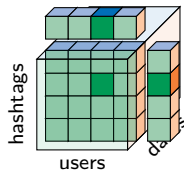


Operations on dimension and variables

`select.dim`



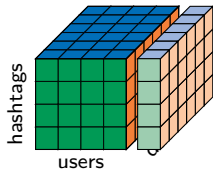
`mutate.var.model`



Example of script

```
data.cube %>%  
  select.dim (users) %>%  
  filter.elm (users, followers > 80)  
  arrange.elm (users, age) %>%  
  ggplot.var (tweet_number)
```

`spread.dim`



`ggplot.var`

