

## VE572 —Methods and tools for big data

### Assignment 5

Jing and Manuel — UM-JI (Summer 2018)

#### Reminders

- Write in a neat and legible handwriting or use L<sup>A</sup>T<sub>E</sub>X
- Clearly explain the reasoning process
- Write in a complete style (subject, verb, and object)
- Be critical on your results

To complete this assignment download the `weather.tar` archive from the course server, using the same credentials as for the project.

#### Ex. 1 — Holidays!

After a long semester of hard work you want to go on holidays in any place, as long as the weather is nice. Therefore you download the weather data from the year 2017 and decide to analyse it using the knowledge from VE572.

1. Define what “perfect weather” according to you. Express it in terms of *precipitations*, *average temperature*, and *daily temperature amplitude*.
2. Using Spark or Drill, with or without R, determine the perfect location of your next holidays.

*Note:* the dataset being small Drill and/or Spark can be run in standalone mode (no need for a Hadoop cluster).

#### Ex. 2 — Data visualisation

Use either Drill or Spark to retrieve the information, then plot the results in R.

1. Plot the temperature variation for each continent.
2. Plot the average temperature for each continent.

*Note:* the file `country_continent.txt` contains a list of the countries with the continent they belong to.