Data manipulation - Exercises

Use the dataset `covid-ita-regions.csv` to answer the following questions.

Lubridate

1. Are all the data recorded at 8:00?
2. What is the local time in New York when each entry is recorded?
3. How old is each recorded information? Express the value in days.

Tidyr

1. Create a covid_long dataset, where all the statistics related to coronavirus are in long form (all names are in a column, all values are in another)
2. Use the previous dataset to derive a plot of your interest. Hint: use the filter() function to select the stats that you want to represent/compare
3. Compute the number of tests done (tamponi) by each region in each weekday (monday, tuesday etc). Present this table with regions as observations, one column for each weekday. Are the number of test done each weekday stable?

Stringr

1. Create the new variable `record_id` that collapse together the record date (format yyyymmdd, without time and without separators) and `region_code`, separated by "_". Be sure that `region_code` is always a 2 figures number. Example: Suppose the date is"2020-02-24 18:00:00" and the `region_code` is "9", then `record_id` must be "20200224_09"
2. Turn in CamelCase the `stats` names stored in your `covid_long` dataframe