Assignments Romain

ıine.ch romain.claret@unine.ch joelle.strautmann@unine.ch jonathan.besse@unine.ch AH KB

Computational Thinking Project

Author: Romain Claret

University: Neuchâtel, Switzerland

Class: Computational Thinking

Type: Project

Github repo: https://github.com/RomainClaret/msc.ct.project

▾ Hide

Instructions

- Formulate a problem from any domain that can be solved using a program (with constraints below)
- Describe the problem using concepts of computational thinking
  - Abstraction
  - Patterns
  - Decomposition
- Provide an algorithmic solution to your problem that:
  - Uses a data from a file as input
  - Creates custom objects from this data

Problem

As of early 2021, information about covid became usual. Indeed, datasets and graphic visualization became common to find, but often doesn't explicitly show what we are looking for. Here was are aiming at building a tool to extract information from multiple datasets with the goal of finding and showing explicitly what we are looking for, e.g., trying to find the correlations between the new Covid cases and the Vaccination campaign in Switzerland, based on the official datasets provided by the Swiss Government.

Solution

Our approach is to provide a class for the Regions (canton) of Switzerland, containing all the information related to Covid based on a specific date. Region objects self-updates when the date is modified, all information for a specific date can be displayed as a graph or a table, and exported as a dataframe or a list.

Similar projects

The following projects are all using data from the sars-cov 2 and display information but none is providing a tool to return all the current data available in Switzerland or even let the user play with the data as they are pleased.

- https://corona-data.ch/
- https://www.covid19-data.ch/vaccinations
- https://www.corona-immunitas.ch/fr/
- https://sciencetaskforce.ch/en/current-situation/

▾ Hide

Datasets found

During the research phase, we found the following datasets are CSV files containing the information about Covid in Switzerland

- https://www.bag.admin.ch/bag/fr/home/krankheiten/a...
- https://www.bag.admin.ch/bag/fr/home/krankheiten/a...
- https://github.com/openZH/covid_19
- https://github.com/rsalzer/COVID_19_VACC
- https://github.com/doerfli/foph-covid19-data
- https://github.com/daenuprobst/covid19-cases-switz...
- https://github.com/covid-19-Re/dailyRe-Data

▾ Hide

Abstractions

As I am no expert in the medical field, I will try to gather as much information as I find officially and publicly available (preferably by date and by canton):

- number of tests done for the date and in total until this date
- number of positive tests for the date and in total until this date
- number of hospitalizations for the date and in total until this date
- number of deaths for the date and in total until this date
- number of vaccinations done for the date and in total until this date
- number of vaccinations completed (seconde dose) for the date and in total until this date
- percentage of vaccination for the date and in total until this date
- reproductive number Re for the date and in total until this date

Additionally, I will be focusing on maintained (up to date) datasets:

- Use of auto-maintained datasets (which use scheduled scrapers)
- I will replace the missing information with 0 if necessary

▾ Hide

Patterns

The following are the part of the project using patterns:

- show graphs (based on a template)

Hide

Decompositions

The following are the decomposed steps to build the algorithm

- use of pandas' dataframes for the data manipulations
- gather the official raw datasets from sources
  - focusing on csv files ready to use (summaries)
  - make a scraper for csv files unready to use as it is
  - build a custom dataset for our project
  - standardize the cantons columns names
  - take into consideration of unmatching name for Switzerland (CH <-> CHE
- standardize value type
- calculate additional data
  - use the datetime object to substract days and load information previous day information
- display values in tables
- use of different charts (and maps to highlight key information)
  - plotting charts
  - (Switzerland map ->not implemented: geoposition libraries like geopandas, shapefile are not present in graasp)

Step 1: Gather data

- It came to the conclusion that it would not be necessary to build a scraper, as multiple public projects are already gathering and building datasets.
- After some research, we could find all the features required, but no dataset was aggregating all the required features at once.
- It was then undertaken to build a custom dataset by merging the minimum amount of datasets to cumulate our required features.
- 5 datasets were selected:
  - (Vaccins) foph-covid19-data: https://github.com/doerfli/foph-covid19-data/blob/...
  - (Positive Tests) covid19-cases-switzerland: https://github.com/daenuprobst/covid19-cases-switz...
  - (Deaths) covid19-cases-switzerland: https://github.com/daenuprobst/covid19-cases-switz...
  - (Hospitalizations) covid19-cases-switzerland: https://github.com/daenuprobst/covid19-cases-switz...
  - (Re step-wise) dailyRe-Data: https://github.com/covid-19-Re/dailyRe-Data/blob/m...

Hide

(Vaccins) foph-covid19-data: https://github.com/doerfli/foph-covid19-data/blob/...

deliveredTotal	sum of vaccines delivered to the canton till the date	Number	21000
deliveredPer100	sum of vaccines delivered to the canton till the date per 100 population	Number	11.9
administeredTotal	number of vaccines vaccinated in the canton till the date	Number	15955
administeredPer100	number of vaccines vaccinated in the canton till the date per 100 population	Number	9.04
fullyVaccinatedTotal	number of people who are fully vaccinated (2 doses at the moment)	Number	2572
fullyVaccinatedPer100	number of people who are fully vaccinated (2 doses at the moment) per 100 population	Number	1.46

Hide

(Positive Tests) covid19-cases-switzerland: https://github.com/daenuprobst/covid19-cases-switz...

Field Name	Description	Format	Example
Date	Date of notification	YYYY-MM-DD	2020-02-26
Cantons Abbreviations [AG to ZH and CH]	Total of Confirmed cases until this date	Value	11.0
Canton_diff [AG to ZH and CH]	Current date new cases	Value	5.0
Canton_pc [AG to ZH and CH]	Percent cases of the canton population	Value	1.2878435761777038e-06
Canton_diff_pc [AG to ZH and CH]	Current date percent cases of the canton population	Value	5.85383443717138e-07

Hide

(Hospitalizations) covid19-cases-switzerland: https://github.com/daenuprobst/covid19-cases-switz...

Field Name	Description	Format	Example
Date	Date of notification	YYYY-MM-DD	2020-02-26
region [AG to ZH and CH]	Total of Hospitalizations until this date	Value	0.0
Canton_diff [AG to ZH and CH]	Current date new Hospitalizations	Value	0.0
Canton_pc [AG to ZH and CH]	Percent Hospitalization of the canton population	Value	0.0
Canton_diff_pc [AG to ZH and CH]	Current date percent Hospitalization of the canton population	Value	0.0

Hide

(Deaths) covid19-cases-switzerland: https://github.com/daenuprobst/covid19-cases-switz...

Field Name	Description	Format	Example
Date	Date of notification	YYYY-MM-DD	2020-02-26
Cantons Abbreviations [AG to ZH and CH]	Total of Deaths until this date	Value	0.0
Canton_diff [AG to ZH and CH]	Current date new deaths	Value	0.0
Canton_pc [AG to ZH and CH]	Percent deaths of the canton population	Value	0.0
Canton_diff_pc [AG to ZH and CH]	Current date percent deaths of the canton population	Value	0.0

Hide

(Re step-wise) dailyRe-Data: https://github.com/covid-19-Re/dailyRe-Data/blob/m...

date	Date of notification	YYYY-MM-DD	2020-02-09
median_R_mean	Median R_e	Value	1.68
median_R_highHPD	High Median R_e	Value	3.32
median_R_lowHPD	Low Median R_e	Value	0.59

▾ Hide

(Side note): Building a custom dataset

This part will not be used as the project specifications require the use of classes. And this part of the project seemed to be compatible with that specification. Since I was not planning initially to use classes, this is how it would have been done using a dataframe.

- As we have multiple datasets, they will be merged into 1 custom dataset
- 6 datasets were gathered (from openZH, foph-covid19-data, and dailyRe-Data), they have 2 features in common: the Date and the Canton. The merge will based on those 2 features.
- As the datasets are coming from different sources, the labeling are not always the same, it must be unified
  - As the name of the Canton column is not always the same, it will be unified as "region"
  - As the name of the Dates column is not always the same, it will be unified as "date"
  - The region value for Switzerland, will be unified as CH
  - Unify the usage of the region column region
  - It will be only kept the rows present in all datasets for the 2 features
- Some features are meaningless for our project, so we will drop them.

▾ Hide

tested	Total of Tests until this date	Value	11.0
tested_diff	Current date new tests	Value	5.0
tested_pc	Percent tests of the canton population	Value	1.2878435761777038e-06
tested_diff_pc	Current date percent tests of the canton population	Value	5.85383443717138e-07
cases	Total of Confirmed positive cases until this date	Value	11.0
cases_diff	Current date new cases	Value	5.0
cases_pc	Percent cases of the canton population	Value	1.2878435761777038e-06
cases_diff_pc	Current date percent cases of the canton population	Value	5.85383443717138e-07
deaths	Total of Deaths until this date	Value	0.0
deaths_diff	Current date new deaths	Value	0.0
deaths_pc	Percent deaths of the canton population	Value	0.0
deaths_diff_pc	Current date percent deaths of the canton population	Value	0.0
hospitalizations	Total of Hospitalizations until this date	Value	0.0
hospitalizations_diff	Current date new Hospitalizations	Value	0.0
hospitalizations_pc	Percent Hospitalization of the canton population	Value	0.0
hospitalizations_diff_pc	Current date percent Hospitalization of the canton population	Value	0.0
median_R_mean	Median R_e	Value	1.68
median_R_highHPD	High Median R_e	Value	3.32
median_R_lowHPD	Low Median R_e	Value	0.59
deliveredTotal	sum of vaccines delivered to the canton till the date	Number	21000
deliveredPer100	sum of vaccines delivered to the canton till the date per 100 population	Number	11.9
administeredTotal	number of vaccines vaccinated in the canton till the date	Number	15955
administeredPer100	number of vaccines vaccinated in the canton till the date per 100 population	Number	9.04
fullyVaccinatedTotal	number of people who are fully vaccinated (2 doses at the moment)	Number	2572
fullyVaccinatedPer100	number of people who are fully vaccinated (2 doses at the moment) per 100 population	Number	1.46

Making a Class for Regions

I would not recommend this approach excepted in a pedagological context, such as this one, as it's clearly not optimized, and kinda meaningless in our project...

- Each Region object is initialized with a canton abreviation and a date
- It uses dataframes as global values
- It has only one set method, which is for the date, which will call an internal update method for all internal values
- It has a get method for all internal values
- It has a method returning the list of all internal values in a specific order
- The delivered_diff, administered_diff, fullyVaccinated_diff attributes are calculated by the class

Region

+str name
+str date-pandas.DataFrame df
-float cases-float cases_diff
-float cases_pc
-float cases_diff_pc
-float deaths
-float deaths_diff

Assignments Romain

ine.ch romain.claret@unine.ch joelle.strautmann@unine.ch jonathan.besse@unine.ch AH KB

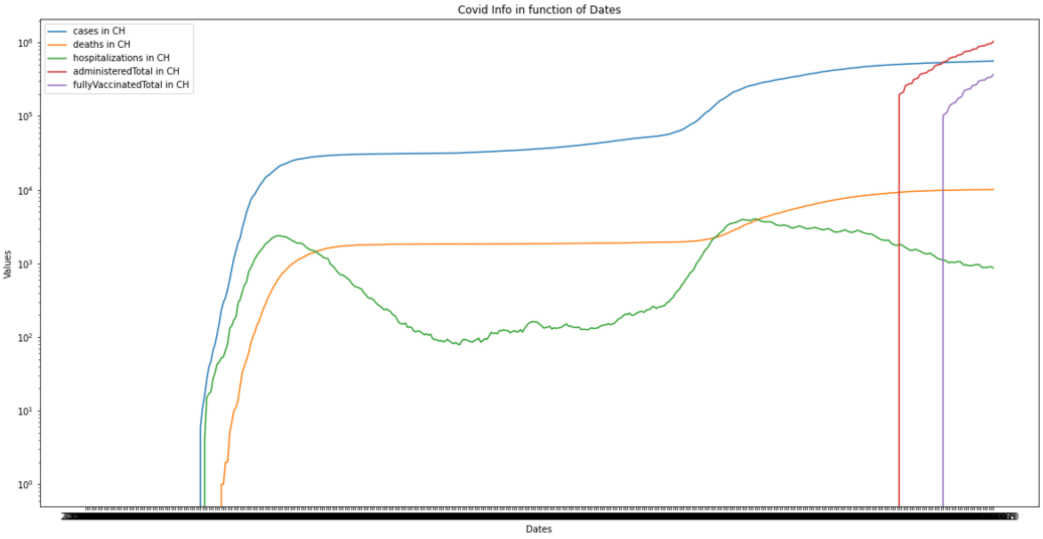
-float hospitalizations_diff
-float hospitalizations_pc
-float hospitalizations_diff_pc
-float median_R_mean
-float median_R_highHPD
-float median_R_lowHPD
-float delivered_diff
-float deliveredTotal
-float deliveredPer100
-float administered_diff
-float administeredTotal
-float administeredPer100
-float fullyVaccinated_diff
-float fullyVaccinatedTotal
-float fullyVaccinatedPer100

-update_InternalValues(): void
+get_all(): List
+get_df(): pandas.DataFrame
+print_data(): void
+get_name(): str
+get_date(): str
+get_cases(): float
+get_cases_diff(): float
+get_cases_pc(): float
+get_cases_diff_pc(): float
+get_deaths(): float
+get_deaths_diff(): float
+get_deaths_pc(): float
+get_deaths_diff_pc(): float
+get_hospitalizations(): float
+get_hospitalizations_diff(): float
+get_hospitalizations_pc(): float
+get_hospitalizations_diff_pc(): float
+get_median_R_mean(): float
+get_median_R_highHPD(): float
+get_median_R_lowHPD(): float
+get_delivered_diff(): float
+get_deliveredTotal(): float
+get_deliveredPer100(): float
+get_administered_diff(): float
+get_administeredTotal(): float
+get_administeredPer100(): float
+get_fullyVaccinated_diff(): float
+get_fullyVaccinatedTotal(): float
+get_fullyVaccinatedPer100(): float

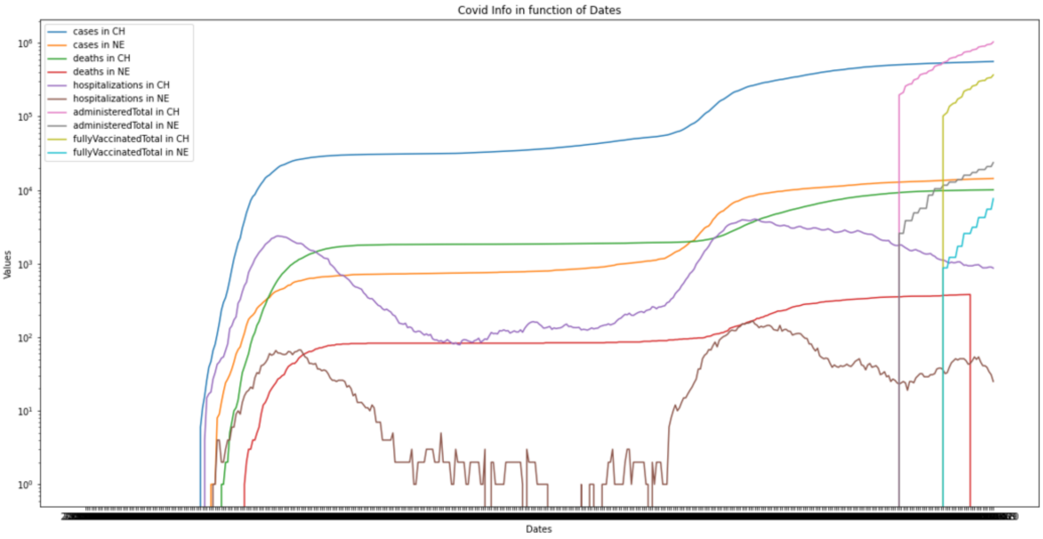
Step 4: Show data

We now have a tool to aggregate covid information per day and for a time span, we would like to use it to show something.

```
plot_data(build_dataset(["CH"],"2020-01-01","2021-03-10"),
["cases","deaths","hospitalizations","administeredTotal","fullyVaccinatedTotal"])
```



```
plot_data(build_dataset(["CH","NE"],"2020-01-01","2021-03-10"),
["cases","deaths","hospitalizations","administeredTotal","fullyVaccinatedTotal"])
```



Assignments

Romain

ine.ch romain.claret@unine.ch joelle.strautmann@unine.ch jonathan.besse@unine.ch AH KB