Dataset
Find a dataset (or multiple) that you will explore. Assess the quality of the data it contains and how much preprocessing / data-cleaning it will require before tackling visualization. We recommend using a standard dataset as this course is not about scraping nor data processing.
Hint: some good pointers for finding quality publicly available datasets (Google dataset search, Kaggle, OpenSwissData, SNAP and FiveThirtyEight)

We have decided to choose the energy reporter dataset which is available on 'opendata.swiss'. It was made by geoimpact AG , WWF Switzerland, and EnergySwiss using data from swisstopo. This dataset contains government data that describes how advanced each swiss municipality is in terms of renewable energy. The dataset has 3 indicators:
- Renewable_heating_share: share of buildings with renewable heat technologies
- electric_car_share: share of electric cars
- solar_potential_usage: share of roof area that is economically and technologically exploitable that already has solar panels installed

Every canton and municipality has a time series of these indicators from march 2021 till now. We'd like to combine this dataset with the swisstopo 'swissBOUNDARIES3D' dataset that provides the borders of each canton and municipality. We would use this to create an interactive map where the user can click on a part of the map and we would provide statistics of the indicators on the canton or commune in question.

Additionally, for each canton and municipality, we would provide its population using a dataset provided by the Swiss government ([link](#)). We hope to check if the population has an impact on these indicators.

Problematic

Frame the general topic of your visualization and the main axis that you want to develop.
• What am I trying to show with my visualization?
• Think of an overview of the project, your motivation, and the target audience.

- Visual exploration of sustainability in Switzerland by geography.
- Interactive map of cantons, click when zoom in on canton, click on communes to further zoom. Display country, canton, commune statistics. We could show how communes or cantons compare to the average or another canton!
- The motivation is to provide a visually appealing view for anyone to explore sustainability in Switzerland. Possibly to encourage various cantons and communes to improve their performance. See how…

Exploratory Data Analysis

SWISSBOUNDARIES3D:

As mentioned previously, the swiss boundaries dataset contains the borders of all swiss cantons and municipalities. The data is in a ESRI Shapefile format. Therefore the preprocessing to be done is to reorganise the dataset in a JSON format. The JSON format is more suitable for drawing the map. We have successfully preprocessed the data in the "swissBoundaries3D.ipynb" notebook. We were then able to provide an initial visualization of our cantons map and our communes map using matplotlib:

Figure 1. Borders of swiss cantons. Extracted from swissboundaries3d dataset
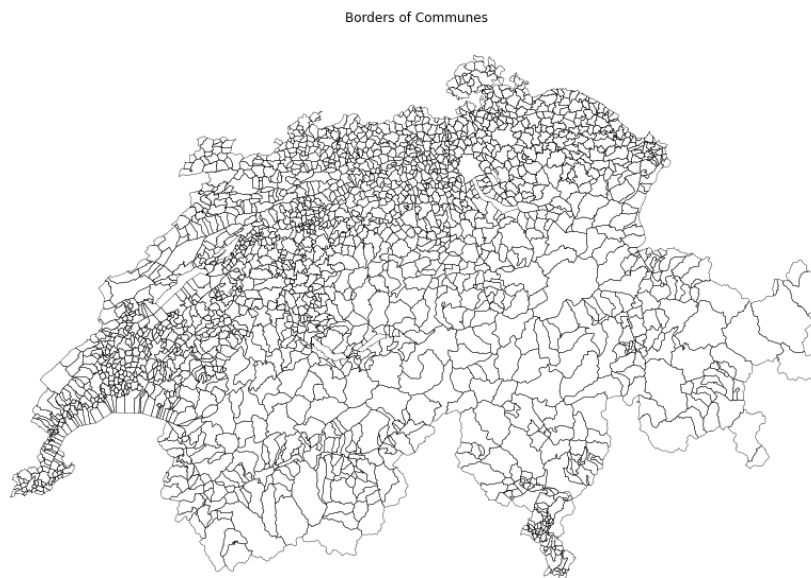
Figure 2. Borders of swiss communes. Extracted from swissboundaries3d dataset

EnergyReporter:

This dataset has 3 CSV files. The first one is **Energyreporter_municipality_historized.csv** which contains data on the 3 indicators for every commune/municipality. It's basic statistics are the following:

	electric_car_share	solar_potential_usage	renewable_heating_share
count	28110.000000	28110.000000	28110.000000
mean	0.013291	0.050500	0.370185
std	0.008106	0.037145	0.161631
min	0.000000	0.000000	0.024606
25%	0.008197	0.028344	0.246654
50%	0.012165	0.043247	0.366748
75%	0.017167	0.064151	0.475167
max	0.097674	0.689336	0.933333

Figure 3. Basic statistics of Energyreporter_municipality_historized

Then comes **Energyreporter_canton_historized.csv** which contains data on the 3 indicators for every canton. Its basic statistics are the following:

	electric_car_share	solar_potential_usage	renewable_heating_share
count	338.000000	338.000000	338.000000
mean	0.013417	0.051807	0.344861
std	0.004304	0.013983	0.102019
min	0.005665	0.021381	0.100473
25%	0.010837	0.041232	0.290011
50%	0.012814	0.050669	0.330031
75%	0.015056	0.059562	0.405243
max	0.035403	0.087705	0.549165

Figure 4. Energyreporter_canton_historized.csv

Finally, there's **Energyreporter_national_historized.csv:** which contains data on the 3 indicators for switzerland as a whole. It's basic statistics are the following:

	electric_car_share	solar_potential_usage	renewable_heating_share
count	13.000000	13.000000	13.000000
mean	0.014157	0.050598	0.314318
std	0.002467	0.004217	0.004802
min	0.009759	0.042506	0.308651
25%	0.012357	0.049528	0.310601
50%	0.013854	0.051056	0.313651
75%	0.015900	0.053285	0.316616
max	0.017949	0.055920	0.326350

Figure 5. Energyreporter_national_historized.csv

This dataset is very clean and doesn't contain any outliers. One in interesting part of this data is that the indicators have a time series for every commune and canton. Our analysis seems to indicate that all 3 indicators have tended to increase this last year:
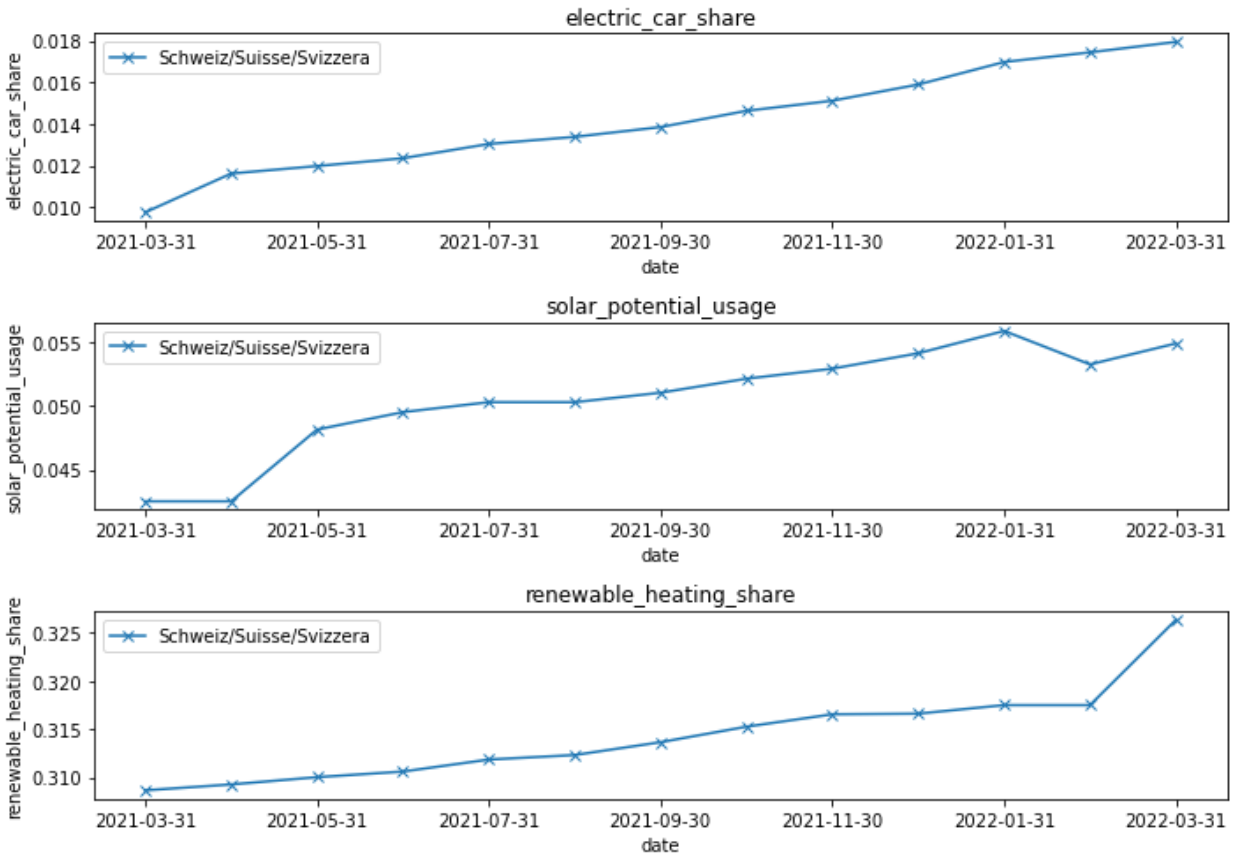
Figure 6. Time-Series of 3 Indicators of Energy Reported dataset for switzerland as a whole

For more plots, see the 'energy_reporter.ipynb' notebook.

Swiss Population:

Preprocessing has to be done for this dataset because the layout of the file is made to be read in an excel spread sheet. This preprocessing is done in the 'swiss_population.ipynb' notebook (see preprocessing section). This dataset is only used to retrieve the population of cantons and municipalities so there is not much analysis to be done. Nevertheless, the basic statistics are the following:

	year	Population on 1 January
count	284520.000000	2.845200e+05
mean	2000.500000	8.213470e+03
std	11.543417	1.115277e+05
min	1981.000000	0.000000e+00
25%	1990.750000	3.830000e+02
50%	2000.500000	9.340000e+02
75%	2010.250000	2.522000e+03
max	2020.000000	8.606033e+06

Figure 7. Basic Statistics of Swiss Population datatset

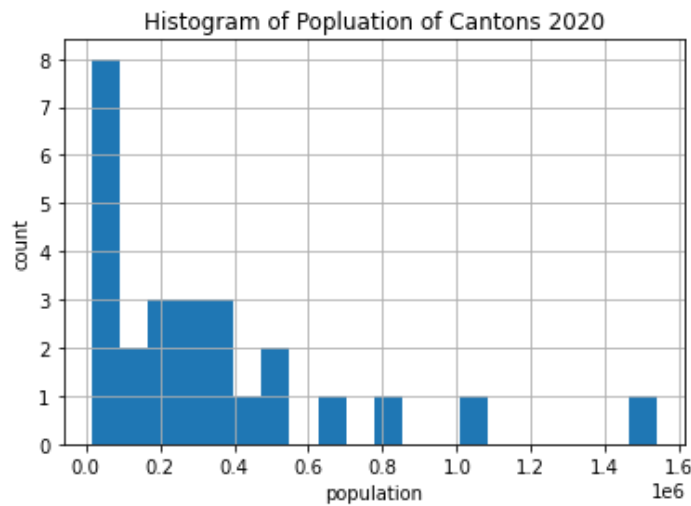We also plotted the distribution of population per canton and per commune for the year 2020:

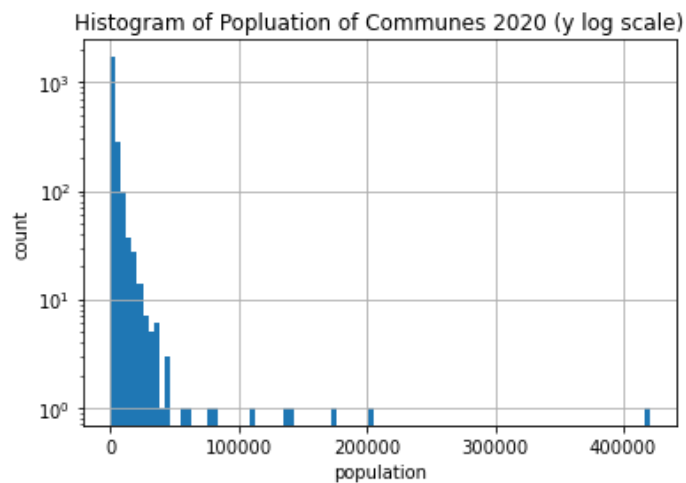*Figure 8. Histogram of population per canton for the year 2020*

Figure 9. Histogram of population per commune for the year 2020

As we can see (and also as expected), the population per commune and per canton iis right skewed. For the next milestones, we hope to potentially find a correlation among the population and the indicators.

Related work

EnergySchweiz website has created an 'Energie reporter' (https://www.energieschweiz.ch/tools/energiereporter/) which is a simple visualization tool where it is possible to type a canton, to which information considering the three main percentages (electric cars, solar and renewable heating) are displayed as coloured bars. Since we have noticed that it shows a map that is not interactive, from this we realized that an interactive map would cover the lack that the website provided. The goal is to get an overall view of the current situation in Switzerland regarding the three components in a map form, which is easy to be

compared, and add the possibility to zoom in by looking not only at the single canton but also more in detail at every single municipality. Along with the energy information, there will be what we know as metadata including information such as population number and GPA per capita, to give a broader view that considers the population characteristics and helps for comparison purposes.

As inspiration, we have looked at interactive maps in which it is possible to zoom in with a double click, which also gave the idea to include a color gradient depending on the intensity and percentages of three main factors.(https://bl.ocks.org/andybarefoot/raw/765c937c8599ef540e1e0b394ca89dc5/,https://coronavirus.data.gov.uk/details/interactive-map/cases).