Data Visualization, COM-480 (2022-2023)

Elisa Baux, Tibor Vaughan, Justine Stoll

Milestone 1 — April 7, 2023

Group Meteo

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.

For this first Milestone, we chose to work on a dataset provided by MeteoSwiss, containing precipitation data for Switzerland over a period of time.
Our data has the following structure. We have a four dimensional grid, with a realization of one variable for each point on this grid. The variable is a prediction of the total amount of precipitations over one hour, in units of kg/m2. Our grid is defined as follows. We have two dimensions, x and y, covering Switzerland with a grid of points every 2km in both directions. A time dimension, with points every hour covers a period of five days. The fourth dimension, of size 21, corresponds to the 21 different predictions made by the model. This means that for every point in space and time, we have 21 different agents making a prediction.
The dataset is of high quality and needs very little preprocessing or cleaning. In fact we found almost no missing values. However, before visualizing the data, we will most likely have to combine the 21 different predictions, in order to have one per point in space and time.

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 for the project, your motivation, and the target audience.

Milestone 1 — April 7, 2023

Our project aims to provide an informative and engaging visualization of the weather patterns in Switzerland over a certain period. Our primary objective is to find a way to represent the probabilistic information of multiple ensemble members rather than just one, and to show the variability and uncertainty in precipitation predictions, which would provide a more complete and nuanced understanding of the weather forecast.
We were inspired to work on a real-world visualization project, and we were fortunate to receive help from Lea Beusch who works in analysis and forecasting at Meteo Swiss. This opportunity excited us since working with a company adds a layer to the educational setting of this class.
In the long run, our visualization could target the general public, scientists, or farmers interested in understanding the patterns and trends of the weather in Switzerland (for example in a warning relevant case). For now, our target audience is the team working at Meteo Swiss, as we aim to provide them with an idea of how this data could be visualized.

Exploratory Data Analysis Pre-processing of the data set you chose

Given that the dataset we are working with does not need particular cleaning, we can proceed with basic operations that will allow for meaningful visualizations. We have one value per four dimensional point, which means that we might need to fix some of the dimensions, or perform some sort of aggregation.
First we consider one arbitrary time point, and aggregate the 21 predictions for each position. The intuitive approach is to compute the mean of the 21 agents, and to visualize the precipitation prediction over x and y coordinates. A basic example of such a visualization is shown in figure 1, where we also show the Swiss boarder for visual reference.
Another intuitive way to explore this data is to set the x and y coor-

dinates to a location of interest. Since the coordinates in our data are given in the Swiss projection LV95, we easily extract the data corresponding to Lausanne. In addition, we can repeat a similar aggregation of the 21 realizations for each point in time, where we compute the mean and standard deviation of the predictions. Figure 2 shows the mean prediction versus time, in dark blue, and the standard deviation, as a shaded surface around the mean. The idea behind this computation is that instead of simply predicting the mean of predictions, we also obtain a measure of confidence. Note, however, that computing simply the mean and standard deviation, is probably not appropriate. In fact, the standard deviation grows importantly with the mean.
At this point we notice that our dataset might lack some complexity to tackle our problematic. It could be interesting to work with an additional variable, such as a prediction of wind gusts over space and time. In addition, we plan on working with data from a longer period of time for the rest of this project. In this way, our visualizations will be more meaningful.

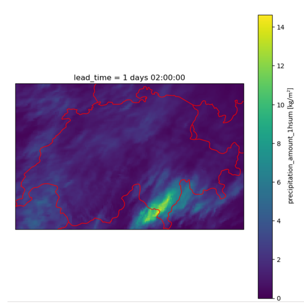

Figure 1: Visualization of precipitation predictions over Switzerland. Time dimension is fixed, and realizations are aggregated over the 21 agents.

Related work

- What others have already done with the data?
- Why is your approach original?

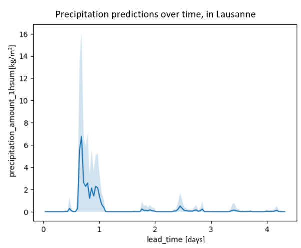

Figure 2: Visualization of precipitation predictions over time, in Lausanne. The two position dimensions are fixed, and realizations are aggregated over the 21 agents.

- What source of inspiration do you take? Visualizations that you found on other websites or magazines (might be unrelated to your data).

The most well-known and extensive use of this dataset is probably the extensive visualizations that MeteoSwiss already provides as part of their website [1] *and mobile app. Others also provide similar weather visualisations for different parts of the globe, however they mostly consist of simple pictograms on top of a map, often without any kind of animation, such graphics can be found on the German* [2] *and French* [3] *weather institutes. Weather.com* [4] *provides in-depth visualisations for extreme weather in the US, it combines a variety of shapes and colors in order to effectively communicate a large amount of data and the risk for the population, however as this kind of extreme weather is less common in Switzerland it does not present the same information as we want. Another good example is the visualisation provided by the BBC* [5]*, they manage to effectively communicate multiple variables on a single visualisation by using multiple shapes and patterns that are relatable to a common natural phenomenon. For instance precipitations are denoted by using either a shade from light blue to green with small, blurry and vertical strokes on top for rain while snow is represented*

[1] https://www.meteosuisse.admin.ch/
[2] Deutscher Wetterdienst
[3] Météo France
[4] weather.com for Zürich
[5] BBC Weather

with a gradient from transparent with small white dots to fully white with big white dots, on top of that they also represent clouds and fog on the same map by using a cloud like, fuzzy and greyish pattern for clouds or one with some kind of horizontal movement for fog. The map also shows pressure with lines drawn on top.

Despite all these services showcasing similar data we see that there is not a lot of variety and even the two most complex (BBC and MeteoSwiss) fail to communicate the uncertainties of the previsions on the map (the mobile app from MeteoSwiss does however display the uncertainties on the bar plot).

Our approach would allow to visualize the predictions of MeteoSwiss with multiple variables on a single graphic, a method only used by the BBC now, and also to display prevision uncertainties, which we have never seen on a map.