ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Data Visualisation Process Book

Pierre Fouché, Matthias Leroy, Alexandre Poussard
December 22, 2017

Abstract

This project is part of the Data Visualisation course at EPFL taught by Dr. Kirell. The goal is to create a nice, interactive and sufficiently complex d3.js data visualisation. Our project focus on world climate evolution during the past century with the objective to find breakthrough temperature changes.

1 - Context

1.1 - Motivation & Target Audience

The climate-change risks that have emerged with the mass industrialisation indicate the need for a transition to sustainable energy, but attempts to encourage people to adopt environmental behaviour often achieve only limited success. Indeed, despite evidences that an energy transition is critical to our survival, people in general are reluctant to change their energy related decisions and behaviours.

As Data Science students feeling concerned by this situation, we wanted to provide a nice visualisation tool that could help people realise that global warming is real and that measures should be taken.

Nevertheless, because we are not scientists with a climatologist background, our project is not professional oriented but built for skeptical and less skeptical people that might be interested by this topic without looking for a pure rigorous scientific proof.

1.2 - Related work

Our project is authentic but inspired by various sources on Internet. Nevertheless the main inspiration comes from this work by the NASA Agency: Five-Year Global Temperature Anomalies which displays nicely and efficiently global temperatures' animations from 1880 to 2016 since we can clearly see a variation through the years.

Our goal here is to show an animated world map displaying land temperatures through the past 150 years approximatively.

2 - Dataset

2.1 - Dataset Processing

We chose to visualise the climate changes over the years. To do so, we used a dataset found on Kaggle: Climate Change: Earth Surface Temperature Data.

This dataset contains 5 different files. We had the temperatures (in Celsius degrees) for the cities, major cities, states, countries and the global land from 1750 to 2012. At this point of the project we did not know if we would use all the years or if we would target a more precise period. Moreover, it will also depend of the quality of the set for the older data.

Thus, we have a lot of data points to help us with the visualisation.

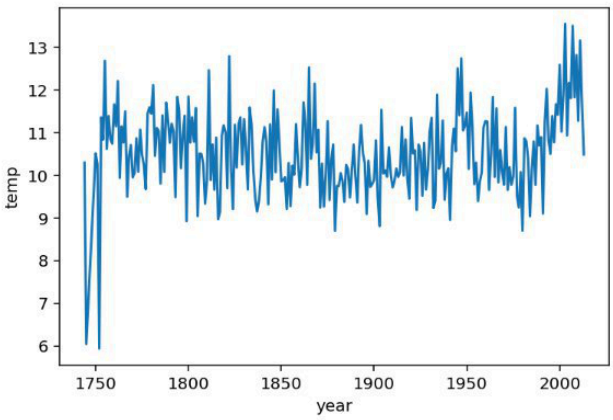
For the processing we decide to use the Pandas Library in Python, because it allows us to manage quickly and easily big datasets.
One of our first cleaning step was to reduce our dataset's size because the temperatures were given for each month. So we have decided to divide the dataset by seasons because we think that temperature evolution is more relevant when we compare seasons over the years.

Moreover, the temperature had too much decimals that would not be visually attracting and relevant for our visualisation, that is why we decided to keep only the first 3 decimals.

As explained in the next section we early decided to make a choropleth map. Thus, we had to use a GeoJSON file with d3.js and do more processing on the dataset. Actually, when we tried to display the values, we quickly realised that not all of our countries appeared. Indeed, their names in the dataset and in the GeoJSON file were different. In order to solve this issue we decided to use the 3-letters IDs which define each country in the GeoJSON. Thus, we found a dictionary on the web that link each country to its id in order to complete our dataset.

2.2 - Exploratory data analysis

During the processing and cleaning steps, we had the possibility to plot and explore the data. First of all we used line charts and scatter plots in order to have a brief insight of the evolution of our data. Then we started to plot the evolution of the temperatures over the years and seasons for some countries as we can see on the following plot.

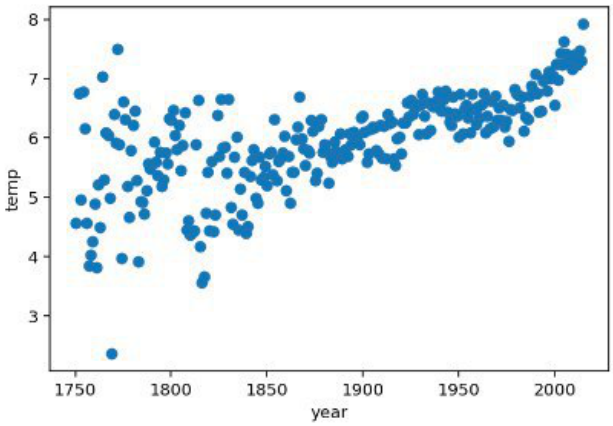


Switzerland fall season evolution

This step allowed us to notice than for most of the countries the temperatures of the 18th century were not consistent and complete. That's why we have decided to start our visualisation from 1850. We actually think that this is a good idea for our project

since this coincides with the beginning of industrialisation and therefore with the major evolution of climate.

Moreover, we have also plotted the global land temperatures over the years in order to know if our guesses were right and if we could show something meaningful for our visualisation. As we can see in the next figure the global land temperature is clearly increasing over the years. Thus, it would be relevant to display it.



Global land fall season evolution

3 - Original Idea and Map draft

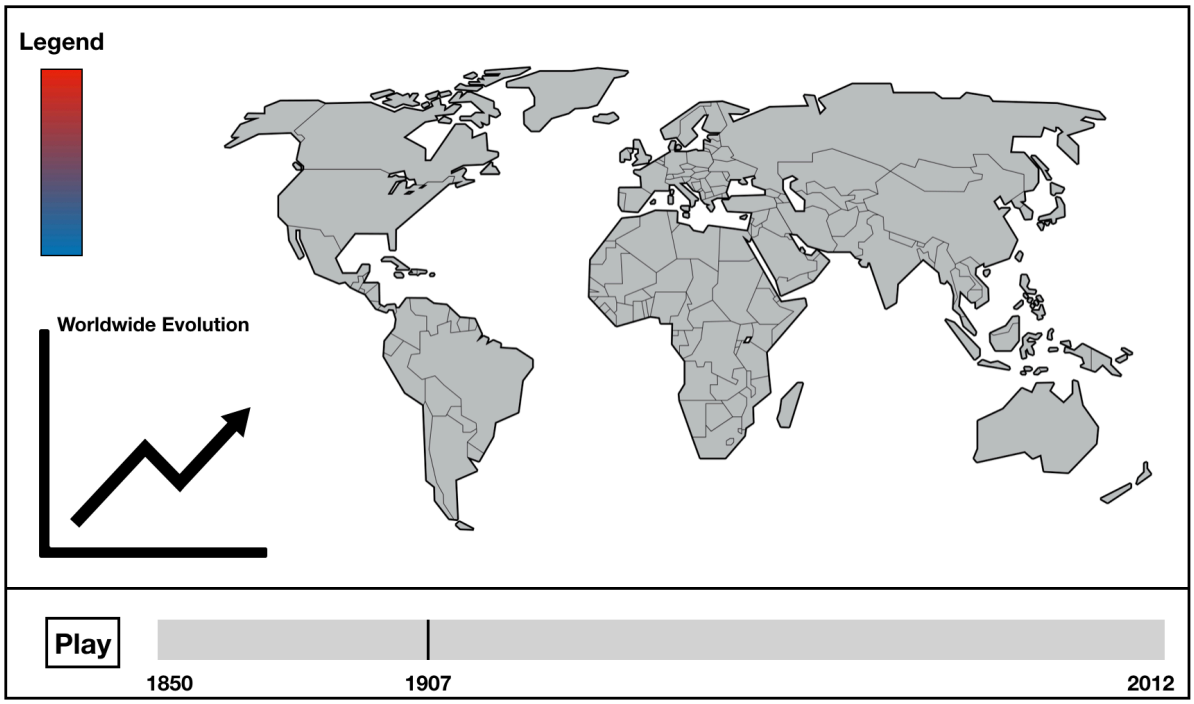
3.1 - Visualisation Ideas according to the dataset

After the dataset's analysis we decided that a choropleth map would be the best visualisation as we wanted to show the climate's evolution over the world.

As we have seen in class, choropleth maps are commonly used when we want to visualise values over a geographical area, which can show variation or patterns across the display location.

However, we also thought of other plots in order to support our goal. Indeed, a choropleth map isn't enough to see breakthrough events and that's the reason why we decided to show the global land temperatures over the years in a distinct area in order to see and understand more clearly the climate's evolution.

One of the objective of our visualisation is still to keep something clean without to much information.



First sketch for the visualisation

4 - Final Concept

We didn't really deviate from our initial proposal but we decided to add some features when we started the implementation.

Indeed, we thought that it would be a good idea to display the temperatures of a country over the years and for each distinct season when it is selected (by clicking on the map). The user can also select a season and travel through the years with the slider. With this interactive way, the user have the possibility to base its reflexion on the charts and also thanks to the animated map.

Finally, after implementation of the above features we realised that the visualisation was interesting but not really determining. We were a bit disappointed and therefore wanted to add another way of interpreting the dataset.
Indeed, we decided to show, on top of that, the difference of temperatures for each country within a 4 years interval. As an example, if we display the choropleth map for 1950, the colour will represent the difference between the temperatures in 1950 and in 1946 in each distinct country and season. This computation was handled in Python with Panda where we added an extra-column in our dataset.

This way of seeing the animation is interesting because it kind of normalises the data which makes more sense for a country temperature evolution.

4.1 - Implementation

4.1.1 - Choropleth Map

First of all, the main part of our visualisation is the map. We have decided to us a Mercator projection because after several tests it seems to be the best projection for us: every countries have decent sizes, every colours are nicely rendered. Additionally, we have also decided to use the colour palette "RdYlBu" because it has a good representation of the temperature (blue for cold, yellow for intermediate temperatures and red for hot). When we don't have the data for a country (inconsistent dataset) it appears in light gray. Furthermore, when we pass the mouse over a country we can see the name of the country and the current temperature for the selected year and season.

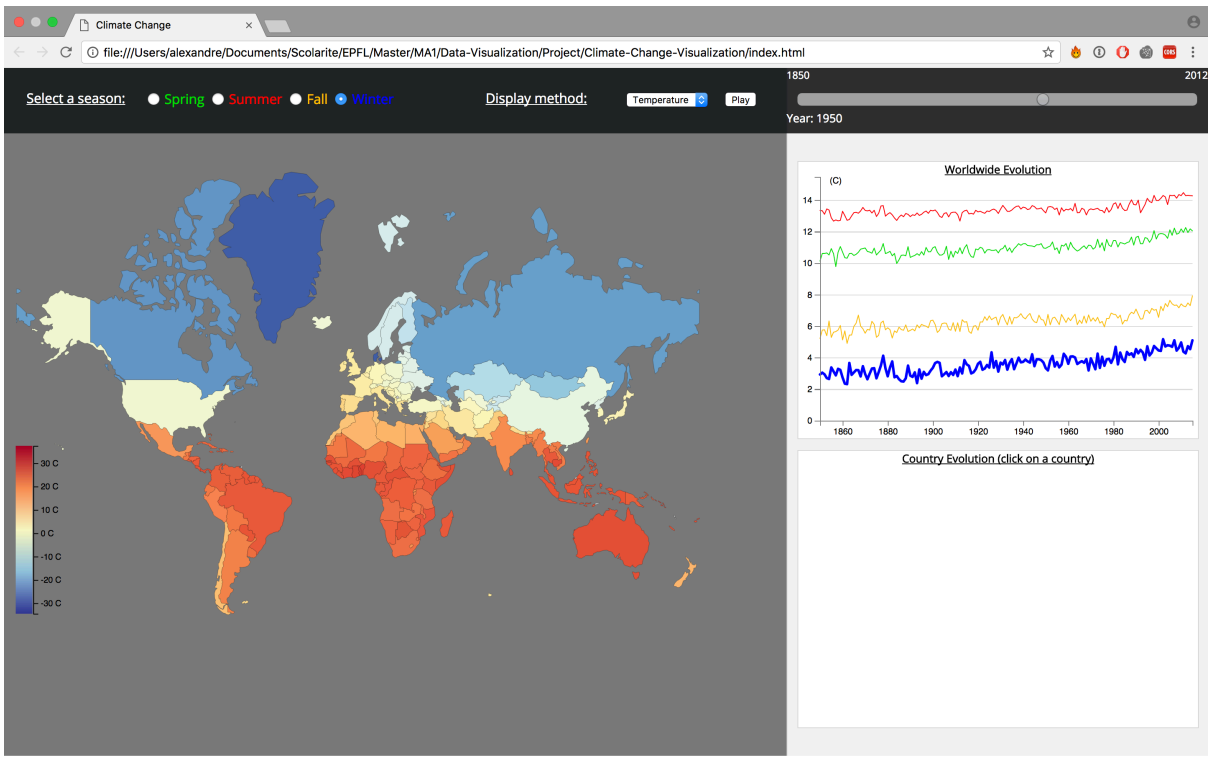
Finally, when we click on a country, the map zooms on it, in order to focus the attention with its colours' variations and a specific plot appears in the right corner of the screen as explain in the next section. If we want to set the global view again, we can click once again on the same country or on a water space (ocean, sea ...).

4.1.2 - Plot Section

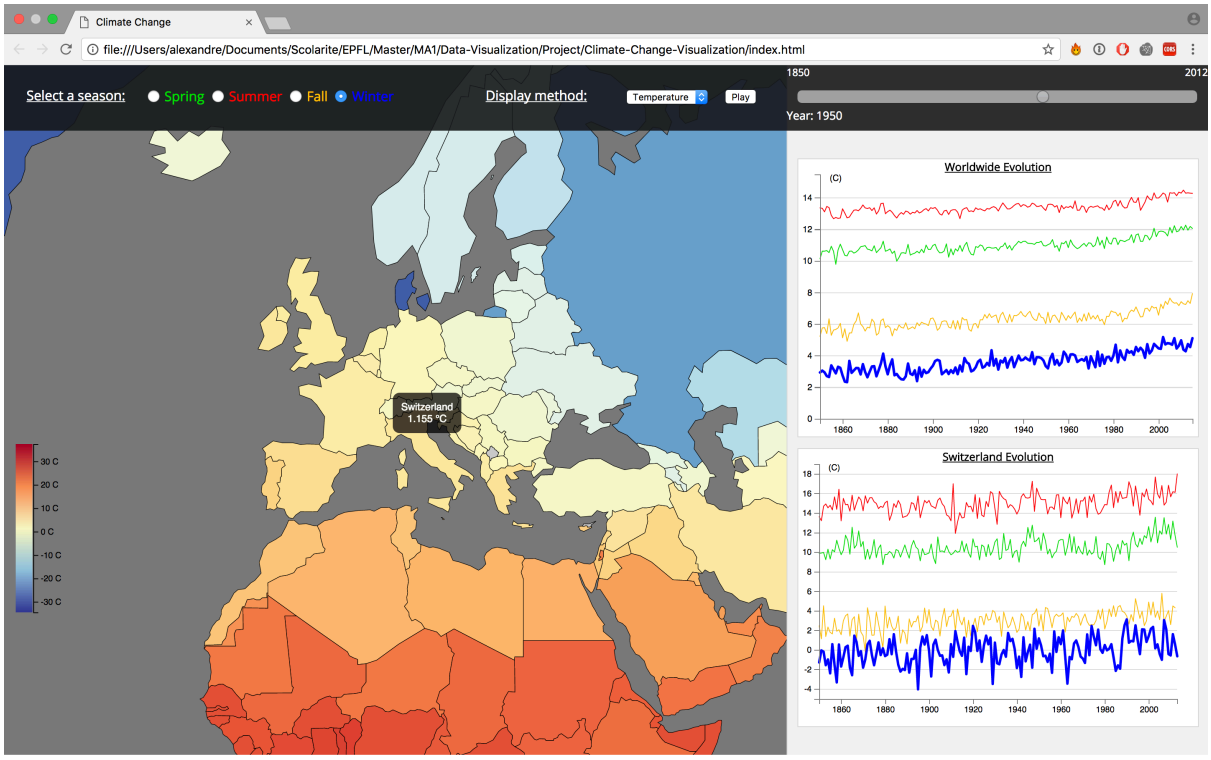
On the right of the screen we have implemented a plot section. The first one is the worldwide temperature evolution over the years (from 1850 to 2012). This plot shows 4 different curves, one for each season (spring in green, summer in red, fall in orange and winter in blue); the current selected one is highlighted.
Moreover, when we select a country by clicking on it, a specific plot appears with the same display as the worldwide one but for the country. When we unselected it the plot disappears until we choose another country.

4.1.3 - Control Section

On the upper side of the screen we have the control section, which allows the user to interact with the map and the plots.
Indeed, the user can change seasons, this will obviously update the map with the new temperatures, but also highlight the selected season curve on the plots.
Then, with a slider the user can choose the year he wants to observe the temperature.
Moreover, there is also a Play button which start the animation. When it is started, we can pause it with the Pause button.
Furthermore, the user can change the visualisation method between "Temperature" which display the temperature of each country for a given year and season, and "Difference" which display the difference of temperature between the current year and 4 years before as explained in part 4.



Winter season, temperature colour mode in 1950 (not zoomed)

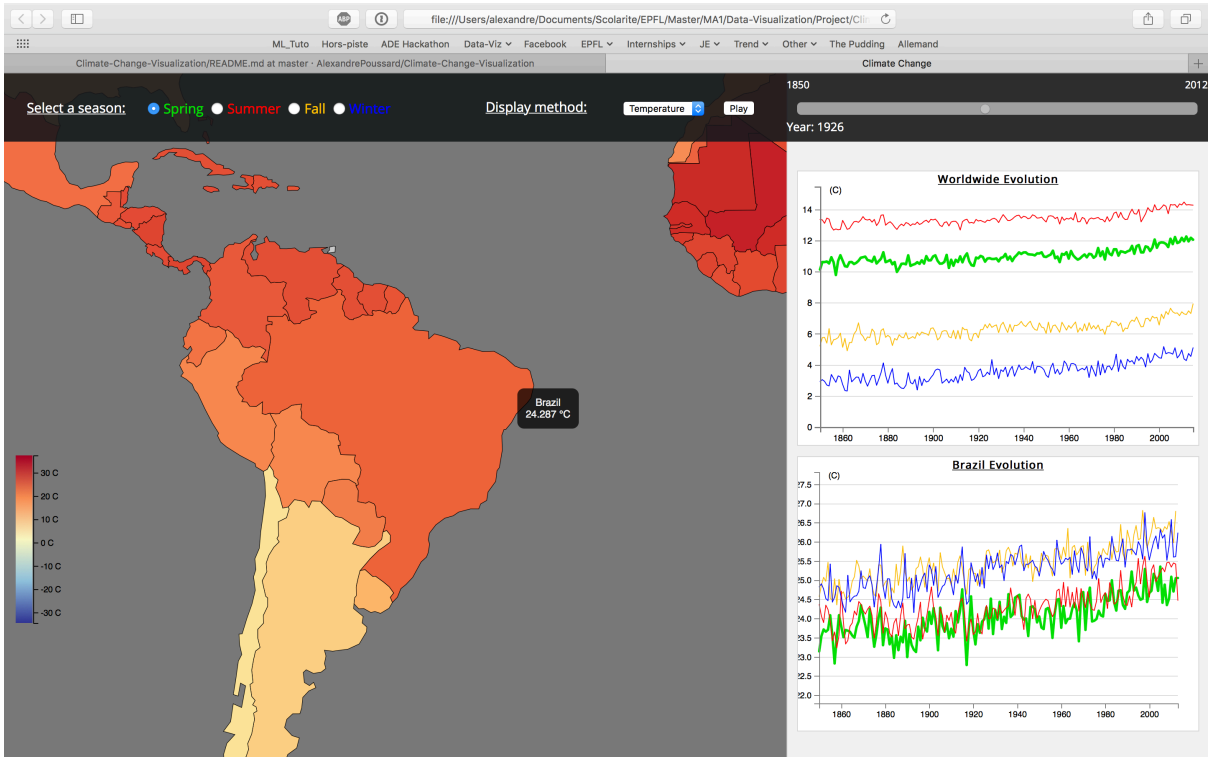


Switzerland focus, still in winter 1950 with temperature colour mode

4.2 - Evaluation

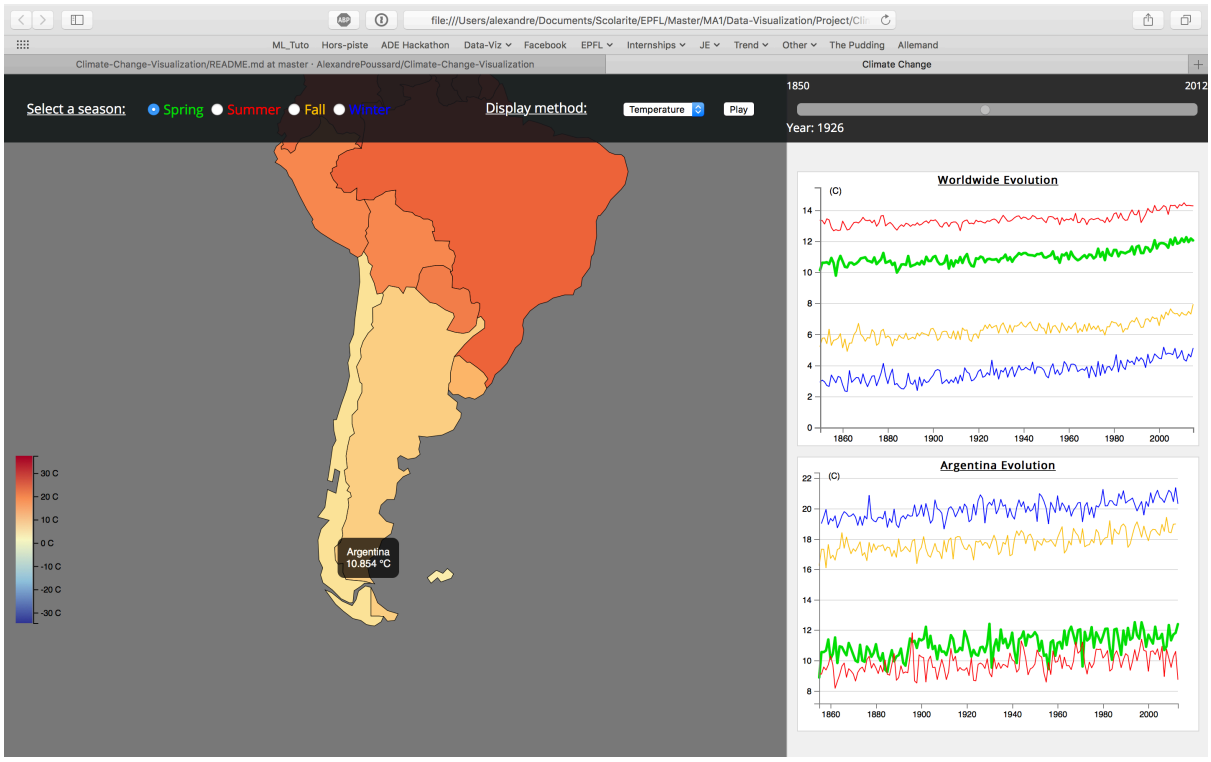
4.2.1 - Data learning

As expected when we have selected this subject, we have clearly observe the climate evolution over the years, especially with the help of the two plots. For example, we can see that temperature grows slowly but surely since the twenties, but it really increases faster since the sixties which is a bit frightening since even with the different treaties we don't have an improvement or a stabilisation.
What is really interesting is to see that some countries (especially in south hemisphere) tend to have a bigger augmentation than other countries as we can see on the picture below. We could not use this as a proof but we are pretty sure that this effect can be generalised.



Observe the fast augmentation on the bottom right plot for Brazil

This visualisation is also interesting in the sense that you can clearly see which countries have clean separate season and which don't. This was not the goal of this project but it is a fun feature. See picture below.



Argentina bottom plot, we can see that this country seems to have 2 really different seasons compare to countries like in Europe for exemple.

4.2.2 - Project evaluation

We are glad of our visualisation because everything works, however it is not adapted to every browsers or every screen resolutions and this could be a nice improvement. An other improvement could be to make the plot animated with the slider (in sync).
Moreover, we have implemented everything from scratch and maybe it would have been smarter to use some external libraries for time saving. The good side is that we have learned a lot on web technologies that none of us had practiced before.

5 - Peer Assessment

Pierre Fouché:

- Preparation: Yes
- Contribution: Yes
- Respect for other's ideas: Yes
- Flexibility: Yes

Matthias Leroy:

- Preparation: Yes
- Contribution: Yes
- Respect for other's ideas: Yes
- Flexibility: Yes

Alexandre Poussard:

- Preparation: Yes
- Contribution: Yes
- Respect for other's ideas: Yes
- Flexibility: Yes