Food Production in 2050

# 2050 : A new society for a new food production system

Process Book

JOULOT Eliott PELLETIER Kevin PETITPIERRE Maxence

Introduction

This process book will detail the visualization projects from the brainstorming to the technical implementation. It will then cover our different merged ideas, and the thoughts behind each step following the good practices and implementation linked to the Data Visualization. Indeed, this project is a realization aiming the course Data Visualization (COM-480) at EPFL during the 2018 Fall Semester in collaboration with the Stanford University and especially the Natural Capital project. This team is developing practical tools and approaches to account for nature’s contributions to society. Their studies will then help important actors within companies, countries and organizations to make smarter decision for a more sustainable future.

# Overview

We are currently, in 2018, more than 7.5 billions humans on earth. In 2050, this population number will increase up to more than 10 billions. It will then mean to feed this increased population. The Natural Capital team analyzed and produced data and prediction regarding the food production environment for 2050. Hence, this data presents many attributes, as the calories produced, the temperature, the production and others, each described in the world map with precise point map with the associated attributes. This study followed 5 different behavioral model following the global society organization linked to a greenhouse gas concentration value. This allow the people to see the possible food production results for a sustainable implicated society with a low greenhouse gas concentration, compared to a nationalist society, possibly leading to wars and less concerned about the ecology. Our project will then explore, analyze and visualize this data in a spatial manner. Currently, 40% of the world terrestrial lands are used for agriculture. This could then be interesting to see how the world map could evolve according to these attributes.

# Motivation

The subject concerning predictions about food production in 2050 possess obviously an ecological aspect. First, this part of the project really interested us, as we are sensitive to the ecology. Indeed, in another EPFL course named Applied Data Analysis we’re doing a project analyzing and predicting ecofriendly trends using the Amazon dataset. Moreover, the fact that our work could help to understand the different tradeoffs and possibilities is a motivational aspect. We wished to learn about the type of prediction scenarios related to the food production, and to apply our knowledge into a map visualization. This was an opportunity to work on a concrete and impactful research.

# Target Audience

Our target audience is anyone interested or involved in the ecological aspects. We implemented our visualization to allow non-specific users, as we were, to understand the different scenarios which are properly explained, and the different values shown in the map. We hope to give a better understanding of the scenarios involved, and maybe a self-thought about the behavior of each person to choose the best possible future in 2050.

Concept

# Dataset

Our given data corresponds to 20 csv of approximately 1 million samples described by 20 features each. Indeed, there is 5 SSP (Shared Socio Economic Pathways) which are then split in 4 different RCP (Representative Concentration Pathways). To resume the SSP models will integrate how the society will evolve and behave (Sustainability, Middle of the Road, Regional Rivalry, Inequality and Fossil Fuel Development are the 5 different pathways). The RCP in each model will give the results using a possible value of the concentrated gas emission / radiative forcing (for example if we get to +4.5 W/m²).

# Exploratory Data Analysis - Processing

First, we visualized the data using Python and especially pandas/numpy. We provided in the github our Jupyter Notebook allowing the loading of a data csv, and the total process of cleaning and extraction. We first of all focused our efforts to handle the SSP1cc model data. We then visualized the data using some description methods to see the ranges, means, max/min of the values and how the points were organized. For the organization the points are ordered by the latitude feature, and in each discrete latitude values, the corresponding longitude values are ordered.

To have a first overview of the map, we decided to extract and process the “calories” feature with the associated position (latitude/longitude). We then cleaned the new dataframe by removing all the invalid values, and “standardized” the data, by taking the log and a coefficient of the calories, and then round the values in order to have a reduced length of string. Indeed we will then convert it to geojson so the string values have to be optimized.

As it is not really efficient to plot all the data points in a map (1 million), we decided to create a data compression. For this we implemented an algorithm that will iterate through the different discrete values of latitude (north to south) and for each latitude value, iterate through the longitude west to east. Then we will cluster every 5 (parameter) points of following longitude to compute the mean of each feature and store it in a single point, if and only if the distance between each point is less than a threshold (parameter). This avoid then to compute the mean of some points starting in the east cost of the USA with some points in the west cost of Europe, leading to an averaged point a bit meaningless and with a position in the ocean. This will then allow to have a data size divided by 5 (parameter) with a really small loss of precision.

Finally we decided to transform the dataframe into a geojson file, which is well implemented by our map API (cf Design) and which reduce the size memory.

Using this parameterizable python notebook we will then be able to load, clean and extract all the different csv models that we have, and select the wanted feature (calories, production, population etc..).

//Create schema of map clustering method

# Designs