

# Correlating Alcohol Consumption and UFO Sightings in the USA

This experiment aims to explore the connection between the alcohol consumption per capita and the number of ufo sightings in the USA.

## Prerequisites

Prior to running the experiment make sure that the following folders exist:

- `data/raw` - Folder to store the external datasets
- `data/processed` - Folder to store the intermediate dataset in
- `reports/figures` - Target folder for generated correlation plot

## Data Sources

- Ufo Sightings: Sigmond Axel. (2014). ufo-reports (Version commit-c0915f18186e5e2227083702049a838258001a2a) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.1205624>
- Alcohol Consumption: OECD (2018), Alcohol consumption (indicator). doi: 10.1787/e6895909-en (Accessed on 22 March 2018) via <https://data.oecd.org/healthrisk/alcohol-consumption.htm>

The cited datasources have already been added to this repository.

Follow these instructions if you want to use updated versions of these datasource:

1. Download CSV files to folder `data/raw`
2. Set paths to CSV files in notebook `01_data-preprocessing.ipynb` by changing the values of `UFO_SIGHTINGS` and `ALC_CONSUMPTION`

## Running the code

To run the code in this repository you will need to have access to a machine running `python` (at least version `3.5`) and `pip`.

Run `pip install -r requirements.txt` to install the required dependencies.

Once the dependencies have been installed, start the jupyter notebook server via `jupyter notebook` and open `http://localhost:8888`.

In the `notebooks` folder you'll find the following notebooks:

### 01\_data-preprocessing.ipynb

Running this notebook generates a dataset consisting of the number of ufo sightings and the alcohol consumption in the usa per year by preprocessing and accumulating the data provided by the datasources mentioned above.

The resulting dataset is located at `data/processed/ufo_alcohol.csv`

## 02\_visualization.ipynb

This notebook takes the data generated by running `01_data-preprocessing.ipynb` as input and generates a plot to visualize correlations between the data points.

The resulting plot is stored at `reports/figures/correlation.png`

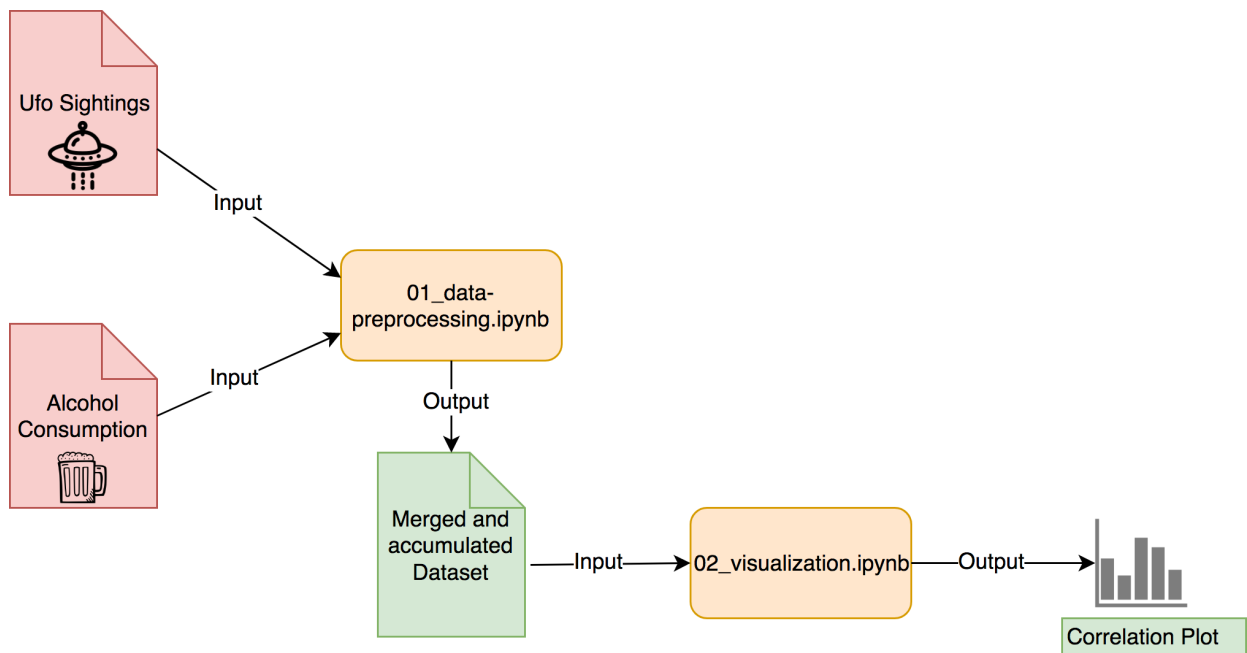
## Docker

Run `docker build .` to create a docker image of this repository. The resulting image exposes the jupyter notebook on port `8888`.

Boot a docker container via `docker run -i -p 8888:8888 <IMAGE_ID>` to start a jupyter instance. The resulting console output will show the url you can open in your browser to take a look at the code, e.g.

Copy/paste this URL into your browser when you connect for the first time,  
to login with a token:  
`http://0.0.0.0:8888/?token=<SOME_TOKEN>`

## Architecture



*System Architecture Diagram*