

Assignment 4: Becoming an Independent Data Scientist

Jesús Ros

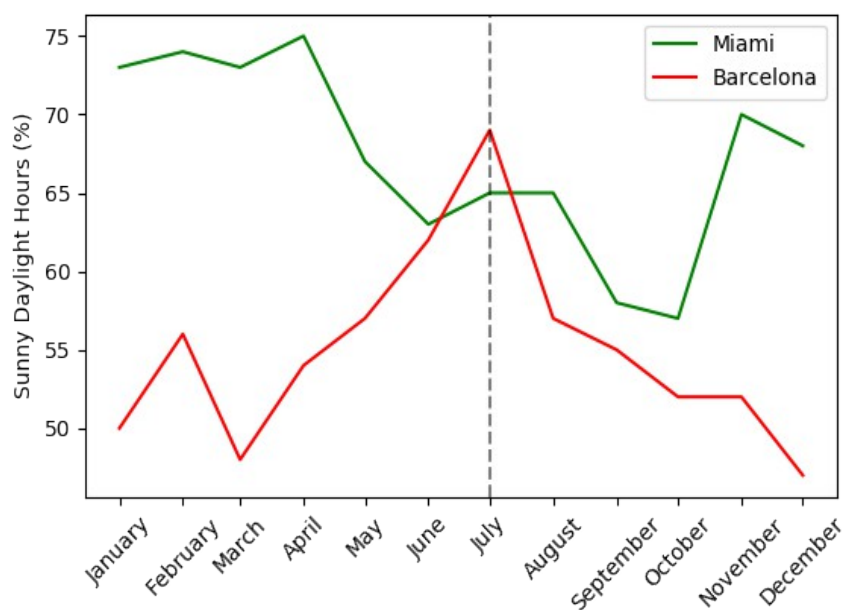
Introduction

For this assignment we have to answer a question about a *weather phenomena* for the region of *Barcelona*.

The precise question we aim to answer is: Is Barcelona sunnier than Miami at any month of the year?

We are going to show the differences in *mean sunny daylight hours* between the cities of *Barcelona* and *Miami* along the year. To this end we have obtained two datasets on sunshine and daylight hours for both cities: [Miami Dataset](#) and [Barcelona Dataset](#).

Results



The visualization was concerned with answering the question of whether Barcelona is sunnier than Miami at any month of the year. Data from the sources cited above was obtained and cleaned so as to plot the corresponding time series for both cities for the variable *Sunny Daylight Hours (%)*. This variable represents the total amount of sunny hours divided by the total amount of daylight hours.

From the plot we can see that even though Miami has a greater percentage of sunny daylight hours than Barcelona for most months (it is commonly referred as the sunshine state for a reason), on July Barcelona has a greater percentage of sunny daylight hours.

Relation to Cairo's principles

1. **Truthfulness:** a legend is provided to support identification of each time series.
2. **Beauty:** line colours were chosen so as to be easily identifiable and the placement of the axis labels and legend considered so as to not difficult the visualization of the elements. The vertical line has an alpha value and a duller colours so as to highlight the data lines. No x-label or title are provided to increase the ink ratio, since adding *Month* as x-label would be redundant and the figure does not need a title to be understandable.
3. **Functionality:** we have chosen a line graph as it highlights the tendencies of both time series better, allowing for better comparison.
4. **Insightfulness:** a vertical line was drawn at the point of interest to focus the attention of the reader to the most interesting part of the plot.