INSTRUCTIONS

The experiments was a lab study, where participants had to look at both a tree and a Venn diagram presenting a Bayesian problem and they had to answer to the problem, requiring them to make an inference based on the data provided about weather prediction.

The file `dataBayesian.csv` contains the variables and results about performance collected from the experiments. This was one of the input files of the R script `stringExtractor.R`

The file `bayes2ET.csv` contains the raw data as extracted from the eye tracker Tobii Studio 3.2. The data represents events as ordered fixations on AOIs by participant. This data set is fed into the R script file `datasetCreation.R`. This script returns a cleaned data file called study2ET.csv. This is the second input file of the R script `stringExtractor.R`.

The file `stringExtractor.R`. takes in both csv files and process the data extracting the scanpaths for each participant, for each condition (Tree and Venn) and for each group (Correct and Incorrect). This returns two outputs `bayeScanpaths.csv` and `bayeScanpathsrep.csv`. Both files contain the scanpaths of participants with all the metadata. The first one has the complete scanpaths, the second one has the reduced scanpaths (without repetitions e.g. AA).

These files are used in `bayes2Analysis.R`, which produces descriptive statistics, dwell time for fixations, regression analysis, and the permutation tests, on transition matrices, for the Jenson-Shannon, the Hellinger and the Bhattacharyya. It has the code to convert the dataframes into transition matrices and to add a Bayesian prior for the Jenson-Shannon distance. It also compute the average deviation analog (ADA), which is a measure of dispersion for nominal distribution. This is perfomed on transition matrices. This also output a data file called `bayesScanpathNEW.csv`.

This last file together with `study2ET.csv`, are the inputs of the R script `bayesLCS.R`. This script performs a different analysis. It produces the descriptive statistics on fixations frequencies, on the number of transitions, fixation times, long fixations, ADA for fixations, ADA for 2-grams frequency; permutation tests using the Hellinger distance for 2-grams frequency, and find the optimal 2-grams and the difference in 2-grams between groups.