

The data is stored as an R file: "bscProjectData.RData" and can be downloaded here:
<https://www.dropbox.com/scl/fi/ujnws9c6alsar3p8f522c/bscProjectData.RData?rlkey=v84s12ir3t92pax23jbaxfln9&dl=0>

To work on the data you must load it into R (using the `load(path)` function, where `path` is the location of the file on your computer. The Data file is a list containing 2 data frames: 1) "mainData" and 2) "articleFeatures".

1) Main Data:

- 1) Each row of main data refers to one observation. The first 7 columns contain the timestamp, the 5 user features and whether the user clicked on the featured article (a 1 if they clicked and a 0 zero if they did not click). The other columns refer to the articles such that a column is 2 if the article was featured for the user, a 1 if the article could have been featured (i.e. was available for featuring for that specific user) and a 0 if the article could not have been shown to the user.

2) Article Features

- 1) Each article corresponds to 5 contextual features. Note that the article ordering in this data frame is not the same as in the main data. Furthermore, note that in some cases multiple articles have the exact same contextual information, this is part of the data. Whether it is a bug or a feature only Yahoo! knows.