Introduction

AlphaTwirl is a python library that summarizes large event data into a set of (multi-dimensional) categorical data, which can be loaded as data frames in R and pandas. AlphaTwirl is used in multiple analyses to summarize event data in Delphes trees, Heppy trees, and CMS MiniAOD and NanoAOD. AlphaTwirl is released under the BSD license and is available on GitHub at https://github.com/alphatwirl/alphatwirl.

Data frames

- a 2-dimensional tabular data structure: columns for variables, rows for entries.
- implemented as a data type in R and pandas, providing a rich set of data operations: sort by values, concatenate, merge by keys, group by values, reshape between long and wide formats.
- can express multi-dimensional categorical data.

Data frames can be considered as extensions of histograms

Why does AlphaTwirl summarize event data?

- As data frames can express event data, it is in principle possible to convert the data type.
- However, event data are often too large for R and pandas, which load all data into memory.
- Furthermore, the next step after the conversion in typical binned analyses would be to make histograms or, in more general terms, summarize event data.