**Data Log**

1. Generating Initial Particle data

* The data for this project is all simulated, so luckily most of it comes in a standard form that is already “cleaned”. The tricky part is to generate the information in a way that makes sense.
* First, I generated a particle process involving 2 top quarks and 2 other particles. The allowed other particles are up, down, strange, and charm quarks, as well as gluons.
* Generation was done using MadGraph, which output 10000 events produced in current LHC running conditions.
* The particles generated from this sample then need to be showered in order to simulate hadronization of the initially generated particles
* Showering is handled with PYTHIA 8, which produces a very large and complicated file containing information of all the intermediate steps in the hadronization process

1. Finding and Matching the Jets

* For this study, all I was interested in was the final state of the showering, so I cut out all particle information that was not in the final state
* Then, I used FastJet packages in order to perform a clustering algorithm on the final state particles. This gave me all relevant kinematics of the jet itself and its constituents.
* Finally, I needed to match each jet with one of the 4 initial particles made using MadGraph. For each jet, I looped over the initial particles, and only kept the jets that were a small angular distance away from a parton (initial particle). Any jets that did not match to exactly 1 parton were discarded. This left about 4 jets per event (as expected).

1. Analyzing the Jets

* In R, I separated the jets into groups corresponding to their flavor. A jet’s flavor describes the type of particle that caused the jet
* The flavors I looked at were light( up, down, strange quarks), top, and gluon
* I cut out jets with a low number of constituents and only selected jets within a certain transverse momentum range