

Clustering NBA Players How-to

The program is written in C++

Running the program:

```
./run.sh <K> <large/medium> <V> <<x.csv>
```

K - clusters to be used, optional but not including will run the clustering 180 times with 18 different K values to determine the best silhouette score

Large/medium – denotes the dimensions and years for analysis (large for the largest dataset 1952-2017, 9-D) (medium 1978-2017, 16-D)

V – optional flag that if present will output the names and years of every player for every group

Example: `./run.sh 5 large V < Seasons_Stats.csv`

5 clusters from 1952-2017 on the entire dataset, outputs the names of all players for every cluster

Accepts the provided .csv files and (probably) others so long as all the dimensions are in the correct order

Initial seeding depends on random value from system clock

Run multiple times to insure a good silhouette value that is suitable for analysis