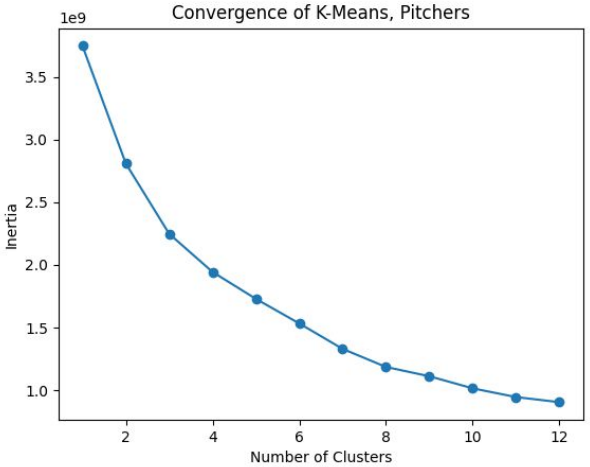
BEM/Ec/120 Update

Edward Speer

Research Question

What factors are the most important in determining when to pull a pitcher from a game BEFORE giving the offensive team an advantage?

Progress



- Clustering Success:
  - Unsupervised K-means converged at identified 10 classifications
  - 4 of these classifications I believed could be combined with others
    - 1: Power fastball with a high speed cutter or sinker (Ex: Joe Musgrove)
    - 2: Power fastball / power slider combo (Ex: Jacob Degrom)
    - 3: Lower velo, 4 pitch mix (Ex: Zack Greinke)
    - 4: Power breaker (Ex: Jose Berrios)
    - 5: Splitters / power sinkers (Ex: Ryne Stanek)
    - 6: No Fastballs (Ex: Diego Castillo)
- Regression on time series:
  - Scraping this data is proving difficult due to request caps on the baseball savant API
  - I have downloaded some large CSV's and a precursory look suggests that secondary pitch spin is a driving factor in declining performance, but I haven't yet acquired a large enough sample to really run stats on it

Next Steps:

- Parse the large CSV's of time series data into player data objects in software
- Replicate the methodology presented in the paper I mentioned during my proposal, but on this finer mesh of clusters, with additional data included
    - I have begun to set this up; in the process of understanding the simulated data technique used to unbias the sample in the paper
- Produce artifacts
    - So far I have only a series of scatterplots, and I need to start collecting more graphical representations of the data