Spring 2020 Assignment

Data Mining

- Assignment

- Programming (coding) almost always re-enforces concepts (for this class) better than paper exercises
- There is a very, very large amount of code publicly available for many concepts, rewriting the same functionality is "***problematical***"
- There are also systems, such as R and R studio (free)
- So understanding and using R and public code seems a reasonable component

Data Mining

- Assignment

- Will be in several parts (pieces)

- Part 1
  - Given a dataset, find some "characteristics"
  - Using K-NN and K-Means
  - More to follow

Data Mining

- You may use R, R-Studio, Weka, Sage or other public software tools
- You may use any *public* code… **if** you site it.
- You must site (give credit) to any place you get code, whether you use it as a "reference", modify it, or use the code as it is.

Data Mining

- You should implement this incrementally, so that you can show progress in the next few weeks
- You should be complete by the first week of May (2020) (the end of the semester)

Data Mining

- Questions?