Adrian Millan

Abel Ortiz

Hector Jimenez

Alwin Alwin Joseph

CS 4950

Data Check

Our data comes from the UCI machine learning repository in the form of a zip file. Upon downloading and decompressing, the folder holds eight comma separated value files all under 20 kilobytes of data. These files are separated into four major tennis tournaments in the UK, Australia, France, and the US. Furthermore, there are also separate files for female and male divisions. All files contain over 100 rows of data and over 20 columns which we believe in sufficient data to draw meaningful insights. After reading through all files, there appears to be no missing values or other formatting issues. The first row gives the title of the variables so we will need to account for that when we commence our work.

As stated previously, there are over 20 columns of variables. The UCI repository gives a full list of the variables and their meaning. I have never played tennis at a competitive level so we will need to conduct further research on the meaning behind all the variables. Other than that, the data seems ready to be worked on.