Max Wagner - Data 608 Final Project Proposal 2

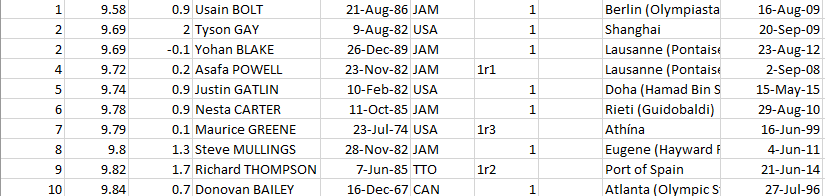
Track and field record progression has hit a relative ceiling in recent years. The IAAF (International Association of Athletics Federations) has archives of meet performances reaching back to the 1960’s. The relevance of the data set has to do with the effect of both technology and drug use in athletics.

Ideally I would like to be able to visualize the times of athletes over time interactively with shiny. There are multiple possibilities for filtering the data; event, year, nationality, and venue. I would like to include as many possibilities as I can without cluttering the shiny app. The data set itself is fairly simple, but will require some cleaning and combining, which can be done with R using csv files. Additional platforms such as Python, Google Charts, d3, and Javascript may be used if the need arises.

The data will come from the IAAF website. One example of where it will be pulled from:  
<https://www.iaaf.org/records/toplists/sprints/100-metres/outdoor/men/senior>

Hi Max, this is interesting but I have some questions. How are you going to download the data: is there a bulk download or are you going to do it manually? Also, how are you going to calculate the records over time? that is probably calculable but not explicit. Do you think the records change enough over 15 years? I know there are some T&F records standing for generations. Email me your thoughts

As far as collecting goes it seems the easiest way is to actually just manually download it through excel. It actually comes out fairly clean without any work. Looks like they formatted with tables in the webpage. It looks like this with no cleaning:



The 15 year problem I had thought about, and there is a fairly extensive listing for records leading back to the 1800’s available. It doesn’t format quite as well, but extracts fairly easily as well. There will definitely need to be cleaning done on it, but it should be a simple job.



There are records from before this as well, but not officially taken through the IAAF, which means the accuracy is questionable, and I am not sure they will be relevant to the project in any case.