The dataset that I am using for my final project is one I personally scraped from the NFL Database ([www.footballdb.com](http://www.footballdb.com)). It collects the following stats from all 985 quarterbacks that have ever played in the NFL:

* player name
* first year in league
* last year in league
* draft position
* draft team
* league
* games played
* games started
* rushing attempts
* total yards
* yards per attempt
* rushing touchdowns
* yards per game

I do a bit of data cleaning/preparation and exploratory data analysis before I begin the linear regression portion of this project. That is because, as previously mentioned, this project is in response to a personal question about whether quarterbacks who scramble more have overall shorter careers. In order to provide some context to the answer, I think it is important to explain to the reader why I have made certain decisions in the data analysis.

As an example, I remove any quarterbacks who played less than one year and started less than 10 games. Occasionally, quarterbacks who are wildly successful in college come to the NFL and are complete flops. Being extremely athletic can be enough to push a quarterback into elite status in college, because they can simply run past the defense. But once they join the NFL, that athleticism doesn’t always translate to success. They may start a few games, have some big running plays, but they don’t last because they aren’t actually good at football – they are simply athletic. I aim to avoid some of these outliers by making certain cutoffs.