

PS4 *Gillingham*

Andrew Gillingham

February 2019

1 Data Sets

To be honest the a vast majority of the data that I want to collect is sports stuff from Pitchf/x and Baseball-Savant. I think most of them have an API that I can use, assuming I know how to use an API. Twitter data would be cool in theory but I just don't think I would find much use for it.

For question 6.4 my df1 was class data.frame while df was sparkR. However, when I moved to look at the sepal length, I got two totally different results. The df data frame gave me a clean answer while df1 looks like it had a null value. I have no clue how that happened because I thought I loaded both dataframes correctly. When filtering df all of the sepal lengths over 5.5 were shown correctly but for df1 there was an error in the "method for function".