1. I chose the College Football dataset because I have always been a fan of CFB. Growing up, I was a Georgia Bulldog fan coming from Atlanta. Although the team saw its ups and downs, I was always curious about what made a team so great. When you watch the tv, it’s easy to see the same uniforms and not pay attention to the stats and the changes that a team goes through year over year. I expected to find that sacks and first down conversions would be a good measurement for team success. This is because first down conversions are a sign of proper ball movement and offense. Sacks are a sign of successful defensive pressure which is a clear metric of a good defense. I hypothesized that a high value of these variables would lead to a successful team.
2. One problem I faced was cleaning the data. The data that I pulled from my source was quite messy and took a while to format properly into a dataset that I could actually work with. It took a lot of manual adjustments as well as discovering some helpful tidy verse packages that could streamline my data wrangling methods. In the future, I hope to web scrape using code to potentially ease data importation, as well as discover new R functions that could help with data wrangling.
3. Being a Georgia fan, sacks and pressure on the QB is an identity to the team. Georgia produces tons of defensive line and linebackers in the NFL which is likely why I leaned heavily on this variable being a strong indicator of success. Other individuals who are fans of different teams who have different schemes may value certain statistics more, thus changing the way they may approach their analysis.