

Final Project Data Scraping and Cleansing

1. I am thinking about doing a study on the variability of zone and man coverage in pass defense for each of the 32 NFL teams. By calculating standardized entropy for the major types of pass coverage options for each team, I can attempt to find any potential relationships with pass defense outcomes.
2. First, I need to obtain data relating to each team's zone coverage rate, man coverage rate, and rates for all other significant types of coverage schemes. Second, I need to find data for key pass defense outcomes, which could be common metrics such as completion percentage and pass yards allowed or underlying statistics such as the distance of the closest defender to the targeted receiver.
3. It would be very beneficial to talk to an NFL (or college) defensive coordinator, defensive backs coach, or pass game coordinator. It would be very helpful to narrow down exactly what the different types of pass coverage schemes are, so my formulas for entropy and standardized entropy would be correct. Additionally, they could help me determine which pass defense outcomes I should use and what years I should focus on. For example, NFL pass defenses are in a late 2010s-present era with two safeties playing very high (called two high safeties).
4. In order to identify more relevant data, I need to first figure out what effective pass defense outcome metrics would be. Second, I need to find metrics that control for team pass coverage skills in order to isolate the potential effect of the variability on success.