Exploring NFL data with nflscrapR

Ron Yurko (@Stat_Ron)

Department of Statistics & Data Science

Carnegie Mellon University
Suppose it’s 4th down with 4 yards to go from the 40 yard line…

You have three options:

1. **Punt** - sacrifice possession, but gain (some) field position

2. **Attempt a field goal** - sacrifice possession but (possibly) gain three points

3. **GO FOR IT** - try to advance the ball four yards and maintain possession

… what do you do???
How do we value a play?

- **Expected Points** - historically how many points have teams scored in similar situations?

- **Win Probability** - have teams in similar situations won the game?

- **Expected points added (EPA)** & **Win probability added (WPA)**
Brief history of football analytics

• *Operations Research on Football* (Carter and Machol, 1971)

• *The Hidden Game of Football* (Carroll, et al 1988)

• Aaron Schatz, Football Outsiders

• Brian Burke, Advanced Football Analytics

• numberFire, Pro Football Focus, ESPN, etc.
NOT ALL YARDS ARE CREATED EQUAL!
Where’s the data???

Recent work in football analytics is not easily reproducible:

- Reliance on proprietary & costly data sources
- Data quality potentially biased by human judgement

Prevents the growth of a football analytics community!
Enter nflscrapR

- There is publicly available play-by-play data!
- NFL API structures its data using JSON for games going back to 2009
- In python, can use `nflgame` and `nfldb` by the unsung hero `burntsushi` (no longer maintained…)
- `nflscrapR` retrieves and organizes the API data into easy-to-use data frames, with more!
Expected Points

![Graph showing expected points based on yards from opponent's end zone. The graph includes lines for different models such as nflscrapR - 1st down, nflscrapR - 3rd down, Carter, nflscrapR - 2nd down, nflscrapR - 4th down, and Hidden Game of Football.](image-url)
Win Probability

Super Bowl LII Win Probability Chart
Philadelphia Eagles vs. New England Patriots

Time Remaining (seconds)

Win Probability

Data from nflscrapR
air and YAC EPA/WPA
nflWAR: A Reproducible Method for Offensive Player Evaluation in Football
(Extended Edition)

Ronald Yurko, Samuel Ventura, and Maksim Horowitz
Department of Statistics & Data Science, Carnegie Mellon University

Timeline of the Pittsburgh Steelers Killer B’s with Wins Above Replacement (WAR), 2013-17
Comparison of Ben, Brown, and Bell to Positional Averages

Data courtesy of rftscrapR (Ron Yurko, Max Horowitz, Sam Ventura)
Department of Statistics, Carnegie Mellon University
IT’S WORKING
IT’S WORKING

For A Passing League, The NFL Still Doesn’t Pass Enough

by Josh Herzog
Filed under NFL

First and 10? Time to pass

NFL teams’ expected points added per play and success rate when running vs. dropping back for a pass on first and 10 facing seven to nine men in the box, 2017

<table>
<thead>
<tr>
<th>TEAM</th>
<th>EPA/PLAY</th>
<th>SUCCESS RATE</th>
<th>EPA/PLAY</th>
<th>SUCCESS RATE</th>
<th>DIFF IN SUCCESS RATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tampa Bay</td>
<td>-0.19</td>
<td>28%</td>
<td>+0.37</td>
<td>84%</td>
<td>+38</td>
</tr>
<tr>
<td>Atlanta</td>
<td>-0.26</td>
<td>30%</td>
<td>+0.31</td>
<td>85%</td>
<td>+35</td>
</tr>
<tr>
<td>Houston</td>
<td>-0.20</td>
<td>25%</td>
<td>+0.28</td>
<td>57%</td>
<td>+32</td>
</tr>
<tr>
<td>Washington</td>
<td>-0.22</td>
<td>24%</td>
<td>+0.28</td>
<td>54%</td>
<td>+30</td>
</tr>
</tbody>
</table>

It’s stat geeks like these that are ruining sports. They aren’t athletic at all and need to find a way to make themselves relevant. Anyone can make up a stat and algorithms to fit their agenda. Both burkhead and gilisiele were injured as wel (let’s not forget the #1 most important stat here...IF a player played a whole season then their stats would be the best). Again, these stat geeks do not contemplate injuries and other, ya know, real life stuff.

Steelers fans should worry about Patriots RB Rex Burkhead. These CMU statisticians tell you why.
Let’s look at some data

(\texttt{ggplot2} football field code courtesy of Michael Lopez
Oct 19th: Football Analytics Workshop

Featuring a Q&A session with
Michael Lopez, NFL Director of Data & Analytics

Oct 20th: Carnegie Mellon Sports Analytics Conference


#CMSAC18