
NFL-Go-For-It!

Unknown Author

April 28, 2014

NFL Go For It! by Mike Ghirardo and Thomas McCann In football there are many decisions a team needs to make in order to win the game. In this project we focus on the decision that needs to be made on the 4th down of any given play. There are three decisions to be made on the fourth down. 1. Punt the ball 2. Kick a field goal 3. Go for a first down. In this project we try to determine which decision should be made under certain conditions. The following are the conditions which we take into account in determining the decision. 1. Offensive and defensive rank of the offensive team. 2. Offensive and defensive rank of the defensive team. 3. The number of yards to convert for a first down. 4. The field position started from. With this information from the data we were able to estimate the expected points scored for each of the three decisions. Finally, with this information a decision can be made.

Populating the interactive namespace from numpy and matplotlib

Importing NFL play-by-play data from years 2002 to 2012.

Out [2]: 0

Joining data sets together to get one long dataset of play-by-play data across all 11 years. Taking out post-season games to get a more accurate ranking of teams. The following retrieves the first and last plays of each game. This will be useful in helping us determine team ranks by how many points scored per game and how many points let go per game. The following sums points gained and points let go per game per team per season. The means by which the teams are ranked offensively and defensively is taking the total number of points scored and total number of points let go and adding them. Creating a two matrices of the total points scored and total points let go with the season as the column and the team as the row, and then ranking them to get the offensive and defensive ranks. Here we bring in the team rankings into the main data frame. Here we create dummy variables and factor variables concerning the ranking of teams. This will help us run a logistic regression using team ranking as covariates.

The following gives drive by drive information. This information is useful in helping us know the expected number of points the offensive team will score given they started the first play of the drive on a specific yard line. The following code quantifies the consequences of certain events occurring. We create a vector of the number of points scored at the end of a team's drive. We assume that touchdown plays automatically get seven points, which means we assume the team gets the extra point given they score a touchdown. Then we run a multinomial logistic regression to determine the likelihood of these events based on certain factors, such as team rank and field position. With both pieces of information we determine the expected number of points of a team given they convert a first down, their ranking and their field position.

```
Optimization terminated successfully.  
Current function value: 0.872453  
Iterations 12
```

```
Out [13]: <class 'statsmodels.iolib.summary.Summary'>
```

```
"""
```

```
MNLogit Regression Results
```

```
=====
```

```
=====
```

```
Dep. Variable:                score    No. Observations:
```

```

65105
Model:                      MNLogit    Df Residuals:
65081
Method:                      MLE       Df Model:
20
Date:                        Mon, 28 Apr 2014    Pseudo R-squ.:
0.05169
Time:                        22:44:08    Log-Likelihood:
-56801.
converged:                    True       LL-Null:
-59897.
                                LLR p-value:
0.000
=====
=====

```

```

      score=DefTD      coef      std err          z      P>|z|      [95.0%
Conf. Int.]
-----

```

```

-----
intercept      17.1021      1.159      14.759      0.000
14.831      19.373
ydline      -0.1892      0.012     -15.208      0.000
-0.214      -0.165
offrankMid      0.0065      0.010       0.620      0.535
-0.014       0.027
offrank31t32      0.0238      0.347       0.069      0.945
-0.656       0.704
defrankMid      0.0103      0.010       1.018      0.309
-0.010       0.030
defrank31t32      0.4581      0.395       1.161      0.246
-0.315       1.231
-----

```

```

      score=FG      coef      std err          z      P>|z|      [95.0%
Conf. Int.]
-----

```

```

-----
intercept      23.3837      1.125      20.785      0.000      21.179
25.589
ydline      -0.2381      0.012     -19.834      0.000      -0.262
-0.215
offrankMid      -0.0052      0.009      -0.565      0.572      -0.023
0.013
offrank31t32      -0.7413      0.308      -2.410      0.016      -1.344
-0.139
defrankMid      0.0217      0.009       2.428      0.015       0.004
0.039
defrank31t32      0.3214      0.357       0.901      0.368      -0.378
1.021
-----

```

```

      score=NoPoints      coef      std err          z      P>|z|      [95.0%
Conf. Int.]
-----
-----

```


0.03018

Time:

22:44:11

Log-Likelihood:

-23149.

converged:

True

LL-Null:

-23869.

LLR p-value:

2.247e-293

=====

score=DefTD
[95.0% Conf. Int.]

	coef	std err	z	P> z
intercept	-2.5052	0.567	-4.415	0.000
-3.617 -1.393				
removelast_ydline	0.0502	0.009	5.741	0.000
0.033 0.067				
offrankMid	0.0073	0.014	0.520	0.603
-0.020 0.035				
offrank31t32	-0.1668	0.431	-0.387	0.699
-1.011 0.678				
defrankMid	0.0189	0.013	1.431	0.153
-0.007 0.045				
defrank31t32	0.2559	0.592	0.432	0.666
-0.905 1.417				

=====

score=FG
[95.0% Conf. Int.]

	coef	std err	z	P> z
intercept	-2.0875	0.461	-4.532	0.000
-2.990 -1.185				
removelast_ydline	0.0874	0.007	11.719	0.000
0.073 0.102				
offrankMid	-0.0018	0.012	-0.154	0.878
-0.024 0.021				
offrank31t32	-0.9428	0.348	-2.712	0.007
-1.624 -0.262				
defrankMid	0.0284	0.011	2.621	0.009
0.007 0.050				
defrank31t32	0.6024	0.488	1.235	0.217
-0.354 1.558				

=====

score=NoPoints
[95.0% Conf. Int.]

	coef	std err	z	P> z
intercept	1.3069	0.451	2.897	0.004
0.423 2.191				
removelast_ydline	0.0607	0.007	8.239	0.000
0.046 0.075				
offrankMid	0.0132	0.011	1.160	0.246
-0.009 0.036				

```

offrank31t32      -0.3978      0.338      -1.176      0.240
-1.061      0.265
defrankMid      0.0149      0.011      1.395      0.163
-0.006      0.036
defrank31t32      0.2583      0.482      0.536      0.592
-0.687      1.204
-----
-----
score=TD      coef      std err      z      P>|z|
[95.0% Conf. Int.]
-----
-----
intercept      -1.1274      0.457      -2.469      0.014
-2.022      -0.233
removelast_ydline      0.0802      0.007      10.814      0.000
0.066      0.095
offrankMid      -0.0164      0.011      -1.428      0.153
-0.039      0.006
offrank31t32      -1.4860      0.346      -4.295      0.000
-2.164      -0.808
defrankMid      0.0381      0.011      3.529      0.000
0.017      0.059
defrank31t32      1.1201      0.485      2.311      0.021
0.170      2.070
=====
=====
"""

```

The following code is used to pull out fourth down plays from the data. This is important since we'll use plays from these downs to find the expected number of points given field goal attempt, punt attempt, or go for it attempt. Pulling out field goal attempt data and running a logistic regression to determine the likelihood of converting depending on the yardline the field goal is attempted from.

```

Optimization terminated successfully.
Current function value: 0.410096
Iterations 7

```

```

Out [18]: <class 'statsmodels.iolib.summary.Summary'>
"""
Logit Regression Results
=====
Dep. Variable:      converted      No. Observations:
9650
Model:      Logit      Df Residuals:
9648
Method:      MLE      Df Model:
1
Date:      Mon, 28 Apr 2014      Pseudo R-squ.:
0.1206
Time:      22:44:11      Log-Likelihood:
-3957.4
converged:      True      LL-Null:
-4500.1
LLR p-value:

```

```

5.246e-238
=====
=====
              coef      std err          z      P>|z|      [95.0%
Conf. Int.]
-----
intercept      3.6030      0.083      43.532      0.000      3.441
3.765
ydline        -0.0981      0.003     -29.699      0.000     -0.105
-0.092
=====
=====
"""

```

In the following cell we create binary columns in the fourth down play data set to indicate what decision was made on the fourth down. The following pulls out the rankings of teams and yards to go to convert on the fourth down. We also perform a logistic regression to determine the likelihood of converting given the rankings and yards to go. Third down plays are used in the logistic regression since using the fourth down plays gives a bias in the estimate.

```

Optimization terminated successfully.
      Current function value: 0.639043
      Iterations 5

```

```

Out [21]: <class 'statsmodels.iolib.summary.Summary'>
"""
                                Logit Regression Results
=====
Dep. Variable:                  converted   No. Observations:
72787
Model:                          Logit      Df Residuals:
72776
Method:                          MLE      Df Model:
10
Date:                Mon, 28 Apr 2014   Pseudo R-squ.:
0.06244
Time:                22:44:13   Log-Likelihood:
-46514.
converged:                True   LL-Null:
-49612.
                                LLR p-value:
0.000
=====
=====
              coef      std err          z      P>|z|      [95.0%
Conf. Int.]
-----
intercept      0.3322      0.035      9.441      0.000      0.263
0.401
togo          -0.1417      0.002     -68.676      0.000     -0.146
-0.138
ydline10t30    0.2740      0.035      7.912      0.000      0.206
0.342

```

ydline30t50	0.4415	0.033	13.261	0.000	0.376
0.507					
ydline50t70	0.4682	0.032	14.554	0.000	0.405
0.531					
ydline70t90	0.4394	0.033	13.149	0.000	0.374
0.505					
ydline90t100	0.3284	0.063	5.242	0.000	0.206
0.451					
offrankMid	-0.0136	0.001	-15.401	0.000	-0.015
-0.012					
offrank31t32	-0.4547	0.036	-12.807	0.000	-0.524
-0.385					
defrankMid	0.0062	0.001	7.073	0.000	0.005
0.008					
defrank31t32	0.2090	0.035	5.922	0.000	0.140
0.278					

=====

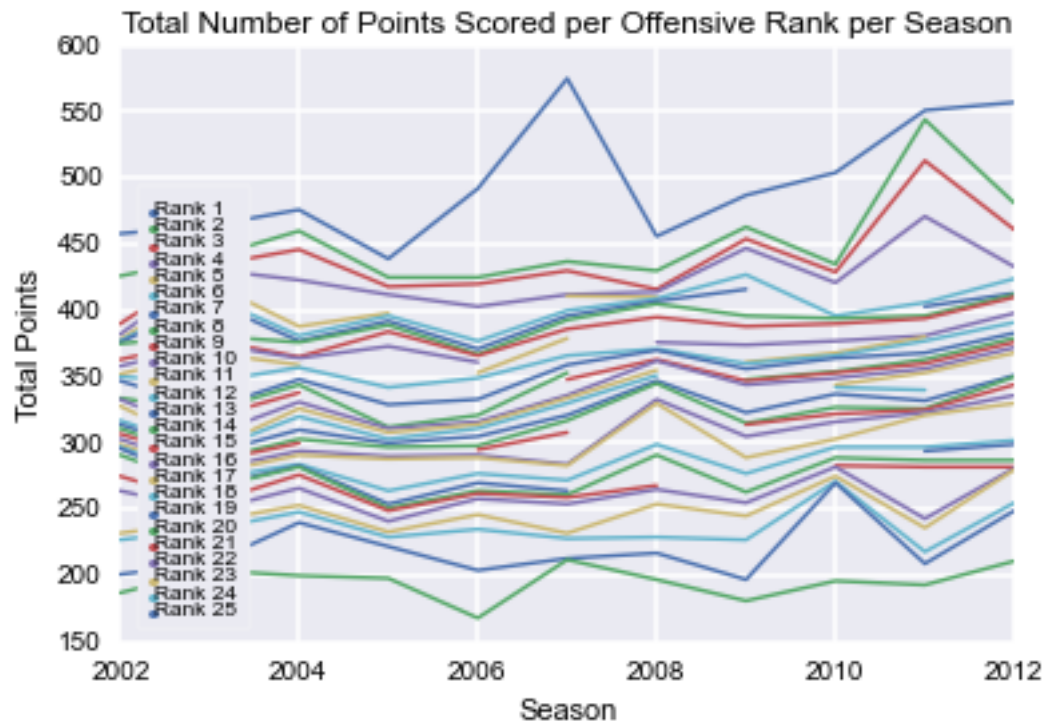
=====

"" "

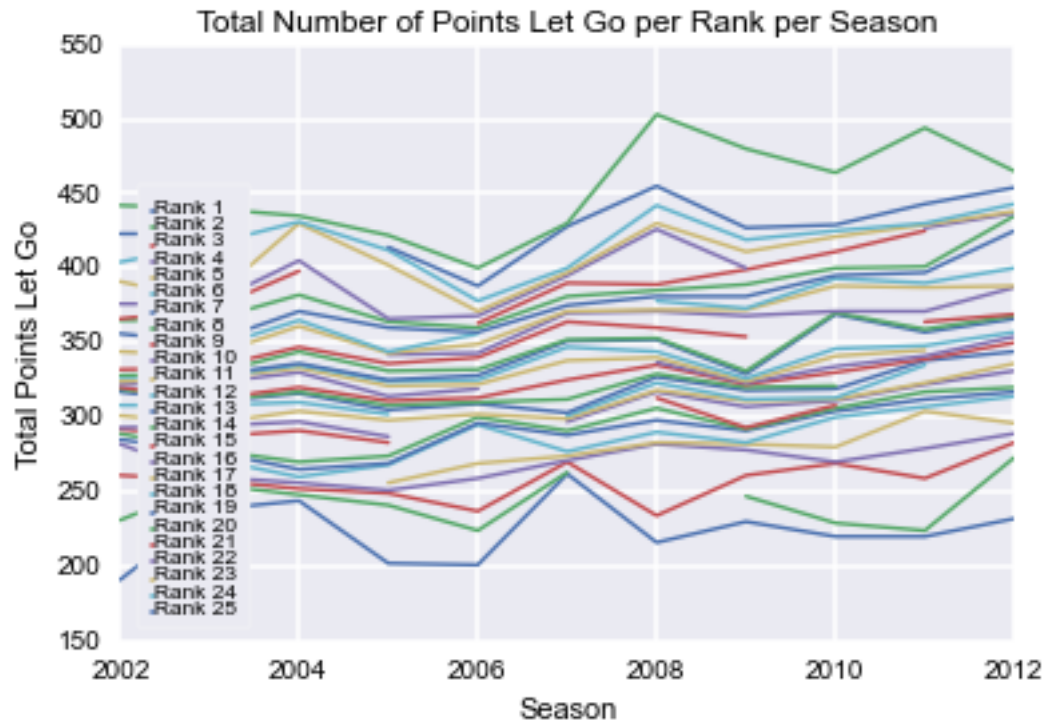
The following functions will be used to determine the choice to be made given yard line, yards to go, and both offensive and defensive rankings of the team making the decision as well as the other team.

convert_prob is a three dimensional matrix that gives the probability of the offensive team converting a first down given how many yards to go for the first down, the offensive rank of the offense and the defensive rank of the defense.

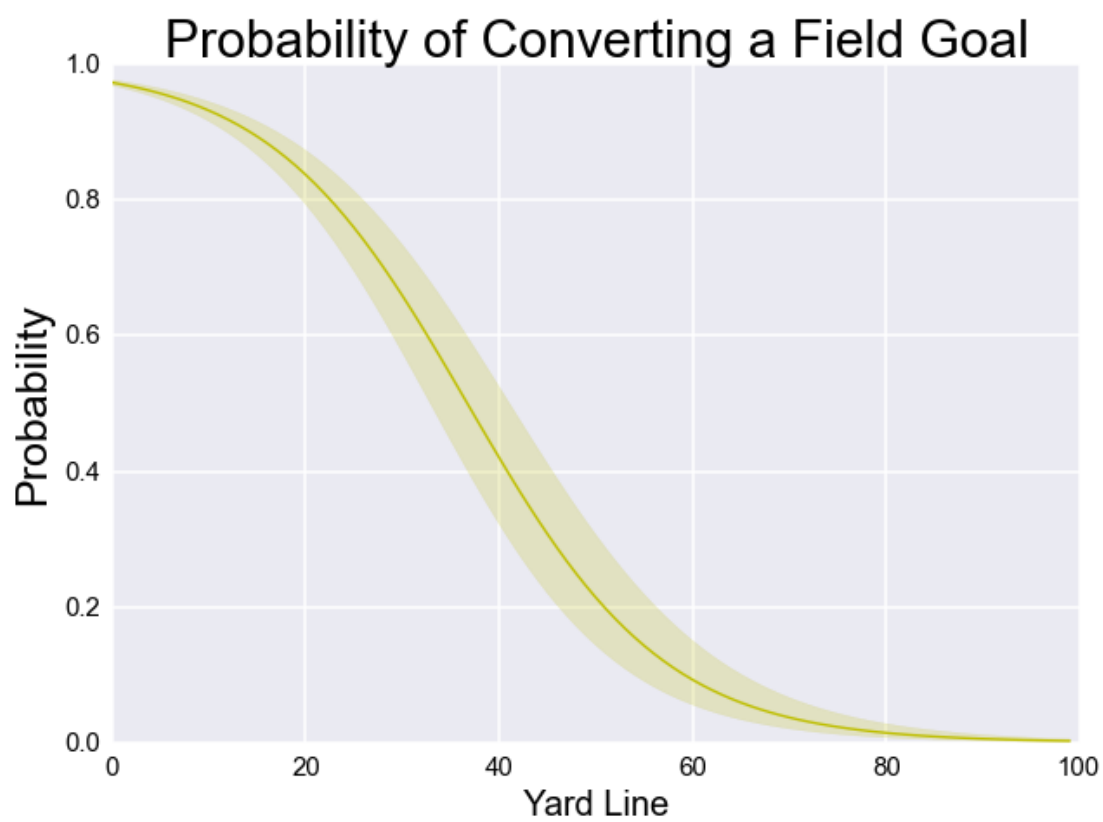
The following is a plot of total points scored on offense per season per ranked team. Each color represents a different rank.



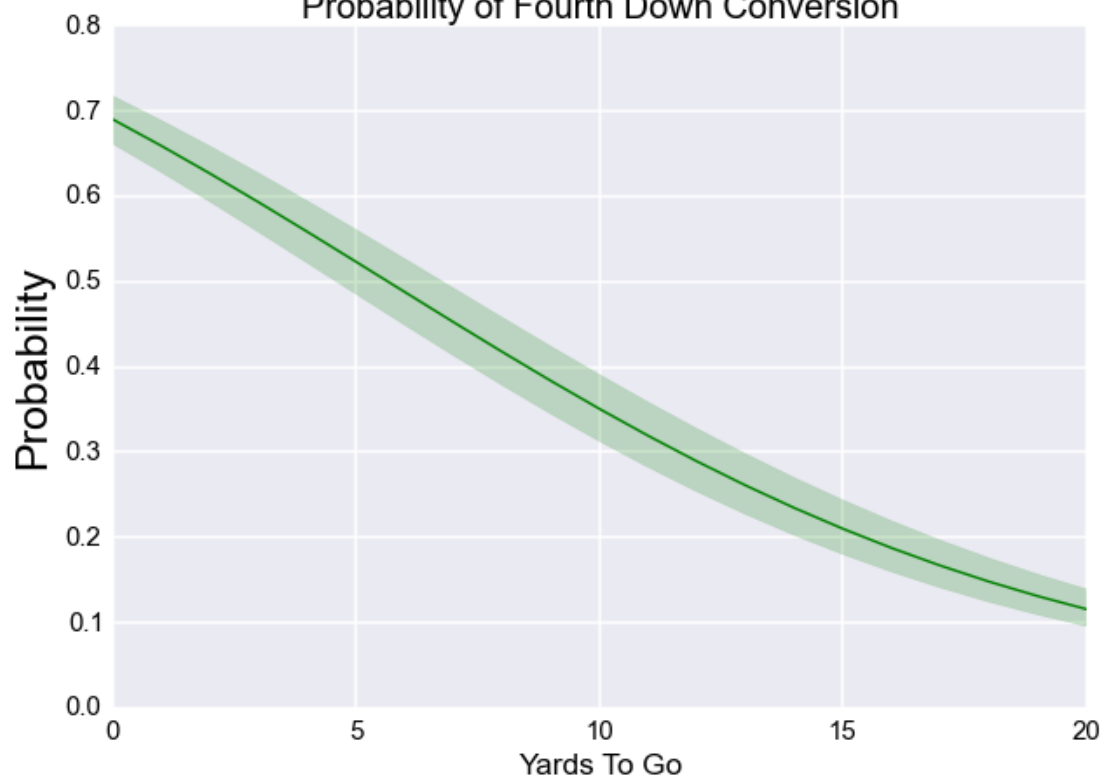
The following is a plot of total points scored on a defense per season per ranked team. Each color represents a different rank.



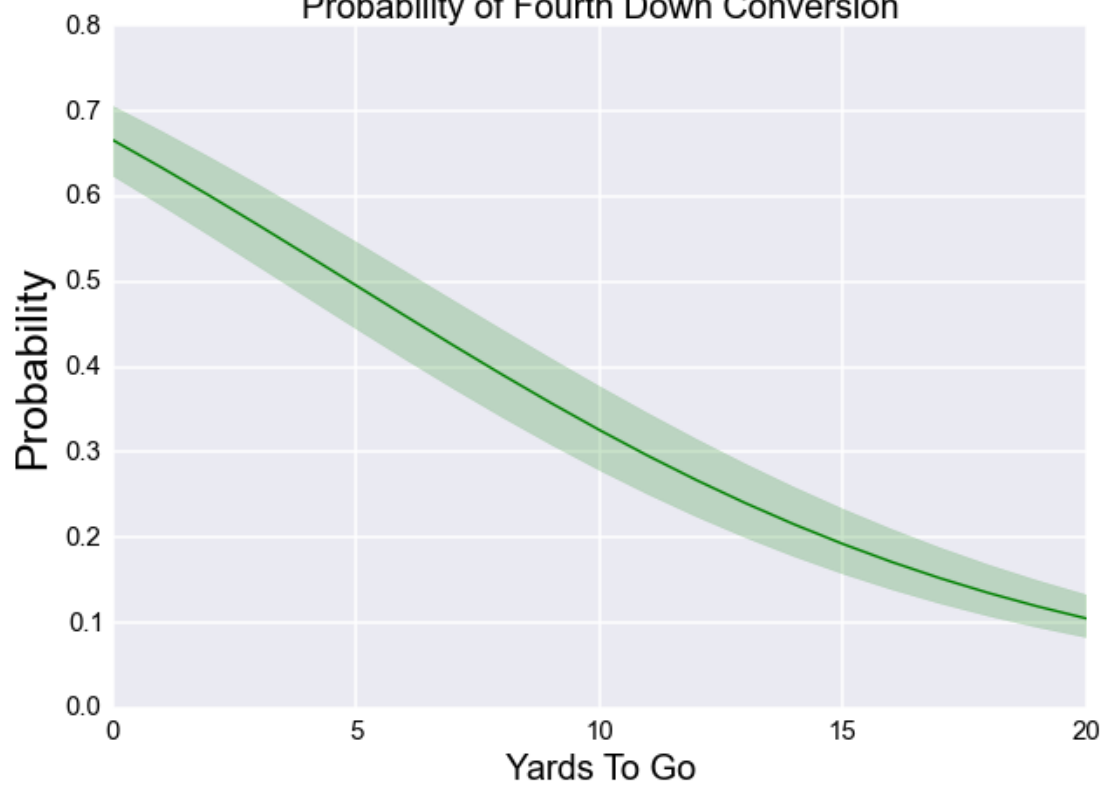
The following graph has probabilities of converting field goals on the y-axis and the yardline the field goal is attempted from. In the midterm presentation the shading on the following graph was purely aesthetic. After creating vectors of the standard errors the following graph now contains the actual confidence interval.

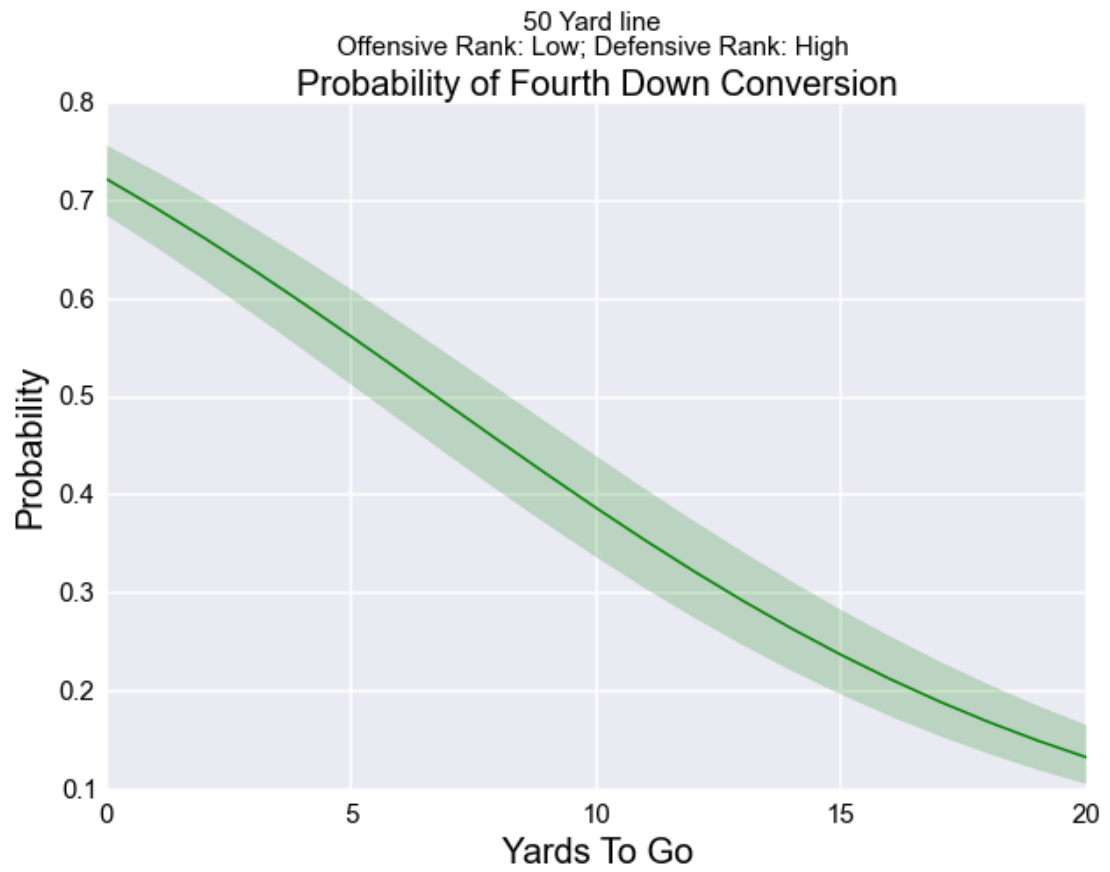


50 Yard line
Offensive Rank: High; Defensive Rank: High
Probability of Fourth Down Conversion

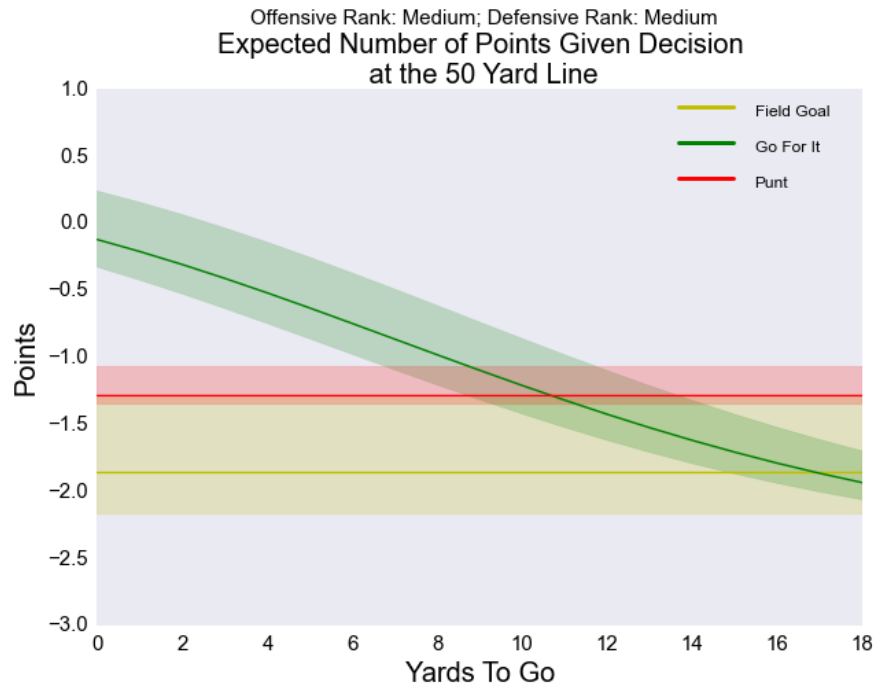


50 Yard line
Offensive Rank: Medium; Defensive Rank: Medium
Probability of Fourth Down Conversion

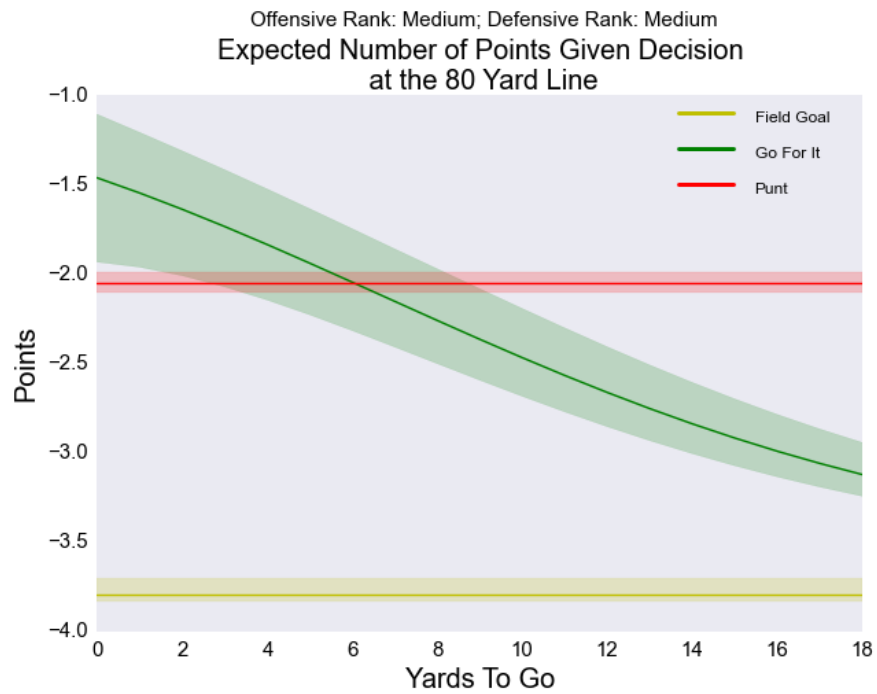




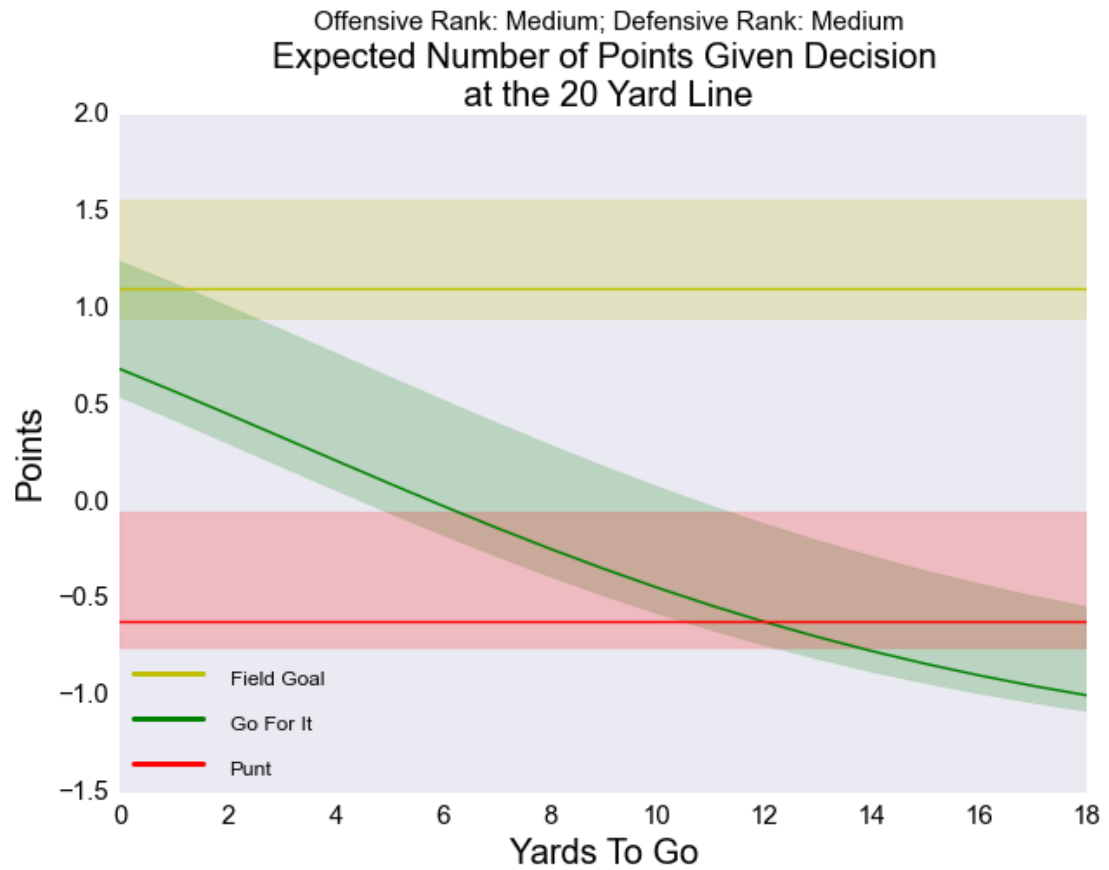
The following is the plot of the decision that should be made from the 30 yard line, where the offensive team has an outstanding ranking and the defensive team is ranked poorly.



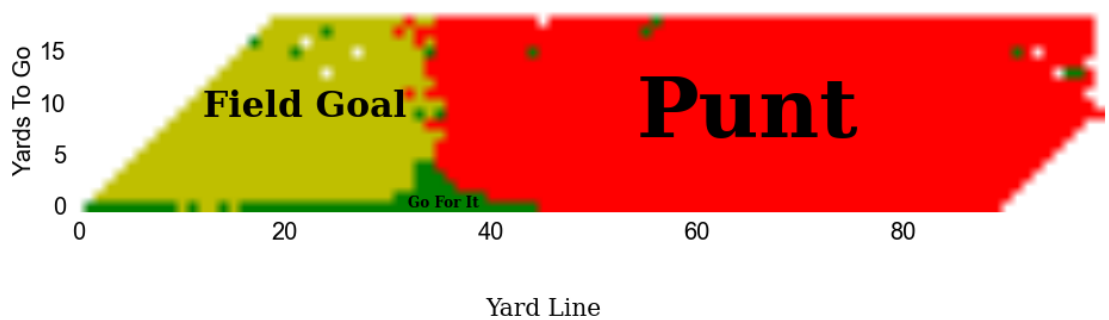
The following is the plot of the decision that should be made from the 30 yard line, where the offensive team has an mediocre ranking and the defensive team is ranked mediocre.



The following is the plot of the decision that should be made from the 30 yard line, where the offensive team has a poor ranking and the defensive team is ranked highly.



Current NFL Decision Map



The following gives the decision to be made given high ranking of the offensive team and poor ranking of the defensive team.

Optimal Decision Map

Offense Rank: High; Defense Rank: Poor



The following gives the decision to be made given mediocre ranking of the offensive team and mediocre ranking of the defensive team.

Optimal Decision Map

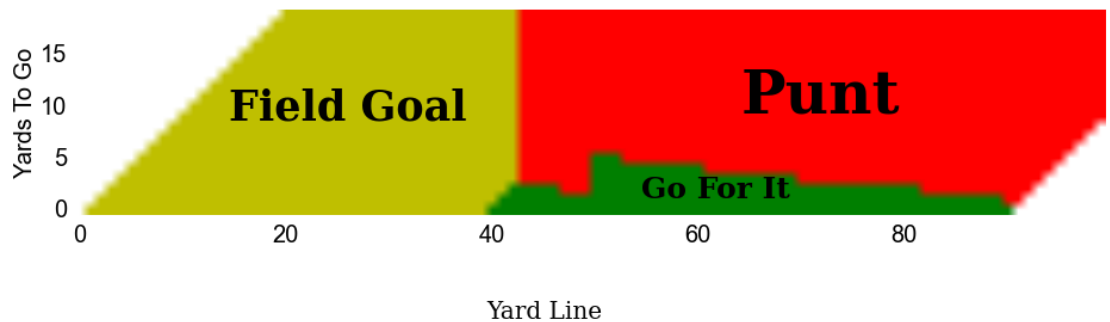
Offense Rank: Medium; Defense Rank: Medium



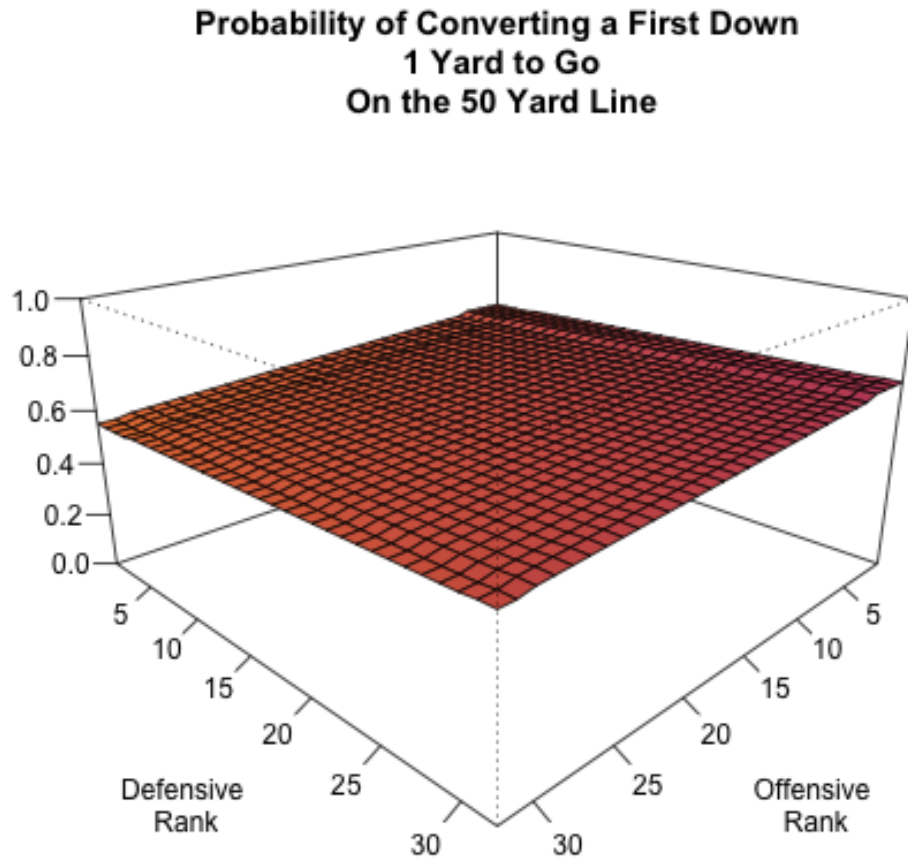
The following gives the decision to be made given poor ranking of the offensive team and high ranking of the defensive team.

Optimal Decision Map

Offense Rank: Poor; Defense Rank: High

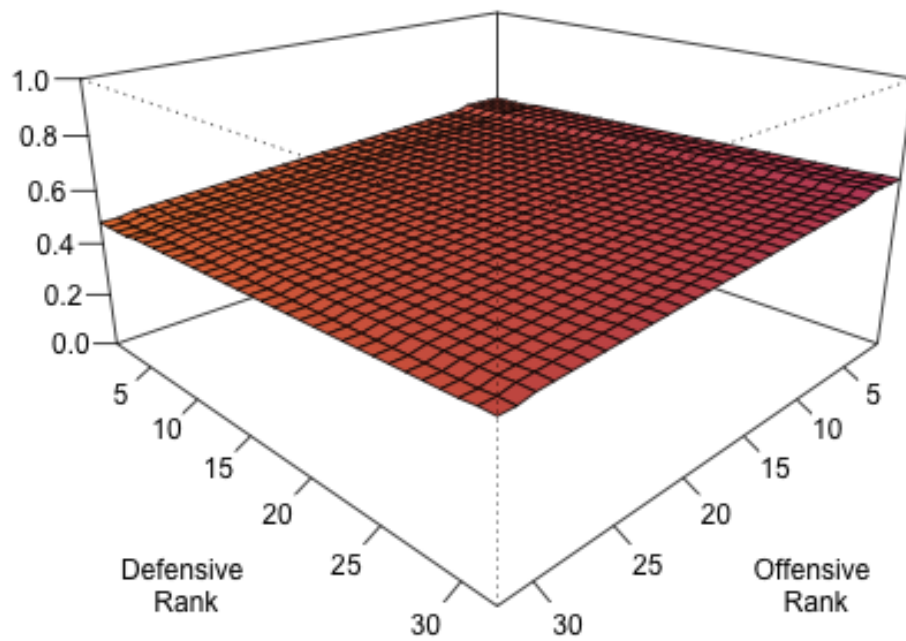


The following shows the probability of conversion of first down per offensive and defensive rankings with 1 yard to go.



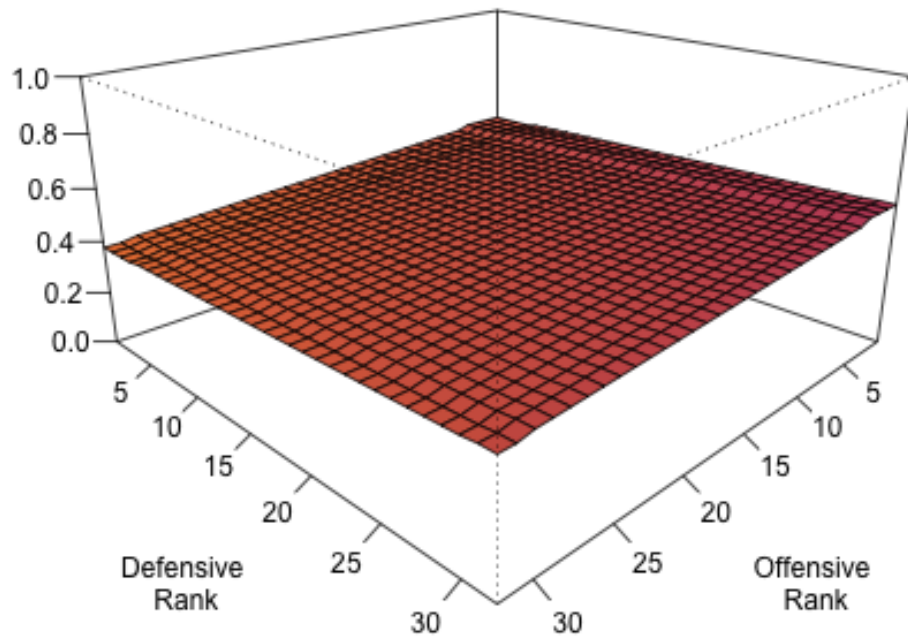
The following shows the probability of conversion of first down per offensive and defensive rankings with 3 yard to go.

**Probability of Converting a First Down
3 Yards to Go
On the 50 Yard Line**



The following shows the probability of conversion of first down per offensive and defensive rankings with 6 yards to go.

**Probability of Converting a First Down
6 Yards to Go
On the 50 Yard Line**



The following shows the probability of conversion of first down per offensive and defensive rankings with 9 yards to go.

**Probability of Converting a First Down
9 Yards to Go
On the 50 Yard Line**

