InSeason_Projections_TE

Jack DePree

2024-03-6

Within Season Fantasy Projections for Tight Ends

Contextualization

This is a within season analysis of TE performance and predicted year-end total fantasy points. Each NFL week adds more data that can be used to predict the future; it is assumed that the more data we have, then the more accurate our predictions. Some variables will be unstable and others will be stable. Is there a point where a small sample of data is able to predict the future just as well as a subset of more data would? If not, then does that mean there is a way we can leverage the unpredictability to outsmart other bettors?

Reasoning

We can ignore variables that are unstable even though the community would still factor them into their predictions (incorrectly). There should also be a point within a season where the amount of usable data passes the threshold of profitability, meaning that we finally have enough data to accurately predict the future: maybe two or three weeks. Furthermore, using different sample sizes (set of weeks) will influence the predictive power of each variable.

Scientiic Questions

1. Which variables are highly associated with fantasy points? (Feature Association)
2. Are there variables that account for more variance in the dataset than others? (PCA)
3. How do variables change - in terms of predictability - with different subsets of weekly data? (Backwards Stepwise Regression Modeling)
4. Are there models that are more predictive than others, given different subsets of weekly data? (Predictive Analysis)

Methodology

1. Collect, assign, and transform data. Compose dataframes to where each record is a player's output in a given subset of weeks. The dataframe will have redundant players but no redundant records. This methodology has an advantage in creating more samples, but limited in the fact that each sample is not explicitly far away from the year end fantasy points in terms of time.

2. Perform a PCA analysis. Calculate the contribution of each variable in the principal components that account for more than 80% of the covariance in the data. The advantage is that the PCA will uncover which variables can be reduced or removed, but it is limited because the PCA does not know that we basically only care about covariance with fantasy points.
3. Use visualization tools and summary statistics to show how the variables change over different subsets of weeks. Of course, there is the dimensional curse and it is impossible to make enough plots to visualize all the variables in every condition.
4. Use regression modeling, stepwise reduction and anova tools to compare variables and models to best predict the future based on different subsets of weeks.

Data Cleaning

It is important to note that I removed all records of TEs when they rush for more than 2 times. This essentially removes Taysom Hill from the data. The data was extremely clean and required nearly no changes.

Data Transformation

```
## Rows: 1,727
## Columns: 49
## $ recTarg        <dbl> 1.0, 1.5, 3.0, 3.5, 2.0, 2.0, 2.0, 3.0, 2.0, 3.0, 2.0,~
## $ recRec         <dbl> 1.0, 1.5, 1.5, 2.5, 2.0, 1.0, 1.5, 2.5, 1.0, 2.0, 2.0,~
## $ recRec40s      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ recYds         <dbl> 7.0, 20.5, 21.5, 15.0, 24.0, 4.5, 12.5, 32.0, 26.0, 29~
## $ recYds100Games <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ recTds         <dbl> 0.5, 0.5, 0.5, 0.5, 0.0, 0.0, 0.5, 0.5, 0.5, 0.5, 0.0,~
## $ recTd40s       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ recDrops       <dbl> 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,~
## $ catch          <dbl> 100.00, 100.00, 50.00, 65.00, 100.00, 50.00, 83.35, 87~
## $ depth          <dbl> 8.50, 12.25, 7.50, 8.80, 8.50, 12.00, 6.50, 8.25, 20.6~
## $ ypt            <dbl> 7.00, 12.25, 7.16, 3.60, 12.00, 2.25, 6.16, 10.00, 15.~
## $ ypr            <dbl> 7.00, 12.25, 16.75, 5.25, 12.00, 4.50, 7.75, 12.00, 26~
## $ rac            <dbl> 0.50, 2.00, 14.50, 3.50, 3.50, 4.50, 2.50, 3.00, 1.00,~
## $ rzRecTarg      <dbl> 0.5, 0.5, 1.5, 1.5, 0.0, 0.0, 1.0, 1.0, 1.0, 1.0, 0.5,~
## $ rzRecRec       <dbl> 0.5, 0.5, 0.5, 1.0, 0.0, 0.0, 0.5, 0.5, 0.5, 0.5, 0.5,~
## $ rzRecTds       <dbl> 0.5, 0.5, 0.5, 0.5, 0.0, 0.0, 0.5, 0.5, 0.5, 0.5, 0.0,~
## $ rzRecTargPct   <dbl> 50.00, 50.00, 50.00, 30.00, 0.00, 0.00, 33.35, 25.00, ~
## $ rzRecRecPct    <dbl> 50.00, 50.00, 25.00, 25.00, 0.00, 0.00, 25.00, 16.65, ~
## $ rzRecTdPct     <dbl> 50, 50, 50, 50, 0, 0, 50, 50, 50, 50, 0, 0, 0, 0, 0, 0~
## $ ezRecTarg      <dbl> 0.5, 0.5, 1.5, 0.5, 0.0, 0.0, 1.0, 0.0, 0.5, 0.0, 0.0,~
## $ ezRecTds       <dbl> 0.5, 0.5, 0.5, 0.5, 0.0, 0.0, 0.5, 0.0, 0.5, 0.0, 0.0,~
## $ ezRecTargPct   <dbl> 50.00, 50.00, 50.00, 10.00, 0.00, 0.00, 33.35, 0.00, 1~
```

```
## $ ezRecRecPct     <dbl> 50.0, 50.0, 25.0, 12.5, 0.0, 0.0, 25.0, 0.0, 50.0, 0.0~
## $ ezRecTdPct      <dbl> 50, 50, 50, 50, 0, 0, 50, 0, 50, 0, 0, 0, 0, 0, 0, ~
## $ rushCarries     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ rush40s         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ rushYds         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ rushYds100Games <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ rushTds         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ rushTd40s       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ ypc             <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ yac             <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ rushTa          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ tat             <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ fumbles         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ rzRushCarries   <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ rzRushTds       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ rzRushPct       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ rzRushTdPct     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ i5RushCarries   <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ i5RushTds       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ i5RushPct       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ i5RushTdPct     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ patConversions  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ patAttempts     <dbl> 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5,~
## $ fantasyPts.x    <dbl> 4.20, 5.80, 5.90, 5.75, 3.40, 0.95, 5.00, 7.45, 6.10, ~
## $ ptsPerSnap      <dbl> 0.07, 0.11, 0.15, 0.11, 0.08, 0.03, 0.11, 0.26, 0.18, ~
## $ ptsPerTouch     <dbl> 4.20, 4.72, 3.67, 1.77, 1.70, 0.95, 2.78, 2.70, 6.10, ~
## $ fantasyPts.y    <dbl> 46.8, 46.8, 46.8, 46.8, 46.8, 46.8, 46.8, 31.2, 31.2, ~

## Rows: 1,175
## Columns: 49
## $ recTarg         <dbl> 1.33, 3.00, 1.67, 2.33, 1.67, 2.67, 2.00, 1.33, 2.00, ~
## $ recRec          <dbl> 1.33, 2.00, 1.33, 1.33, 1.67, 2.00, 2.00, 1.33, 1.67, ~
## $ recRec40s       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ recYds          <dbl> 15.67, 11.67, 11.00, 16.33, 18.67, 25.00, 16.00, 12.00~
## $ recYds100Games  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ recTds          <dbl> 0.33, 0.33, 0.67, 0.33, 0.33, 0.33, 0.00, 0.00, 0.00, ~
## $ recTd40s        <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ recDrops        <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, ~
## $ catch           <dbl> 100.00, 60.00, 88.90, 66.67, 100.00, 75.00, 100.00, 10~
## $ depth           <dbl> 9.83, 10.53, 8.33, 6.67, 9.67, 8.33, 2.00, 1.83, 2.67,~
## $ ypt             <dbl> 10.17, 3.23, 6.78, 6.78, 10.67, 8.50, 8.00, 8.83, 6.83~
## $ ypr             <dbl> 10.17, 5.17, 7.83, 13.17, 10.67, 11.67, 8.00, 8.83, 8.~
## $ rac             <dbl> 1.67, 3.67, 1.67, 10.00, 2.33, 4.00, 6.00, 7.00, 6.00,~
## $ rzRecTarg       <dbl> 0.33, 1.00, 1.00, 1.00, 0.33, 0.67, 0.33, 0.00, 0.33, ~
## $ rzRecRec        <dbl> 0.33, 0.67, 0.67, 0.33, 0.33, 0.33, 0.33, 0.00, 0.33, ~
## $ rzRecTds        <dbl> 0.33, 0.33, 0.67, 0.33, 0.33, 0.33, 0.00, 0.00, 0.00, ~
## $ rzRecTargPct    <dbl> 33.33, 20.00, 55.57, 33.33, 33.33, 16.67, 16.67, 0.00,~
## $ rzRecRecPct     <dbl> 33.33, 16.67, 50.00, 16.67, 33.33, 11.10, 16.67, 0.00,~
## $ rzRecTdPct      <dbl> 33.33, 33.33, 66.67, 33.33, 33.33, 33.33, 0.00, 0.00, ~
## $ ezRecTarg       <dbl> 0.33, 0.33, 1.00, 1.00, 0.33, 0.00, 0.00, 0.00, 0.00, ~
## $ ezRecTds        <dbl> 0.33, 0.33, 0.67, 0.33, 0.33, 0.00, 0.00, 0.00, 0.00, ~
## $ ezRecTargPct    <dbl> 33.33, 6.67, 55.57, 33.33, 33.33, 0.00, 0.00, 0.00, 0.~
## $ ezRecRecPct     <dbl> 33.33, 8.33, 50.00, 16.67, 33.33, 0.00, 0.00, 0.00, 0.~
## $ ezRecTdPct      <dbl> 33.33, 33.33, 66.67, 33.33, 33.33, 0.00, 0.00, 0.00, 0~
```

```
## $ rushCarries     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ rush40s         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ rushYds         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ rushYds100Games <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ rushTds         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ rushTd40s       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ ypc             <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ yac             <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ rushTa          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ tat             <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ fumbles         <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, ~
## $ rzRushCarries   <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ rzRushTds       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ rzRushPct       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ rzRushTdPct     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ i5RushCarries   <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ i5RushTds       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ i5RushPct       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ i5RushTdPct     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ patConversions  <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, ~
## $ patAttempts     <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.33, 0.00, 0.00, ~
## $ fantasyPts.x    <dbl> 4.23, 4.17, 5.77, 4.30, 4.70, 5.50, 2.60, 1.87, 2.20, ~
## $ ptsPerSnap      <dbl> 0.08, 0.08, 0.11, 0.11, 0.09, 0.19, 0.10, 0.13, 0.07, ~
## $ ptsPerTouch     <dbl> 3.52, 1.52, 4.28, 2.82, 3.57, 2.33, 1.30, 1.38, 1.37, ~
## $ fantasyPts.y    <dbl> 46.8, 46.8, 46.8, 46.8, 46.8, 31.2, 30.0, 30.0, 30.0, ~
```

```
## Rows: 69
## Columns: 49
## $ recTarg         <dbl> 5.14, 5.71, 5.88, 6.25, 5.38, 3.83, 7.38, 5.25, 3.57, ~
## $ recRec          <dbl> 3.86, 5.14, 4.12, 4.50, 4.00, 2.00, 6.38, 4.00, 3.00, ~
## $ recRec40s       <dbl> 0.00, 0.00, 0.00, 0.00, 0.12, 0.00, 0.00, 0.00, 0.00, ~
## $ recYds          <dbl> 32.71, 44.71, 43.75, 48.00, 39.88, 12.83, 54.25, 55.38~
## $ recYds100Games  <dbl> 0.00, 0.00, 0.12, 0.00, 0.00, 0.00, 0.00, 0.12, 0.00, ~
## $ recTds          <dbl> 0.43, 0.14, 0.50, 0.12, 0.25, 0.17, 0.00, 0.38, 0.29, ~
## $ recTd40s        <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, ~
## $ recDrops        <dbl> 0.00, 0.00, 0.50, 0.00, 0.00, 0.33, 0.12, 0.50, 0.14, ~
## $ catch           <dbl> 72.63, 92.23, 67.72, 71.28, 76.71, 57.22, 86.01, 78.88~
## $ depth           <dbl> 7.07, 4.35, 7.47, 7.44, 3.01, 7.18, 4.16, 9.79, 3.17, ~
## $ ypt             <dbl> 6.21, 7.28, 6.63, 7.02, 7.46, 4.75, 7.31, 9.96, 7.27, ~
## $ ypr             <dbl> 8.98, 8.12, 9.19, 9.46, 10.15, 7.94, 8.50, 11.71, 8.38~
## $ rac             <dbl> 4.81, 3.81, 4.10, 3.51, 8.77, 5.78, 5.66, 4.71, 6.03, ~
## $ rzRecTarg       <dbl> 1.29, 1.29, 1.62, 0.75, 1.00, 1.33, 0.50, 1.12, 0.71, ~
## $ rzRecRec        <dbl> 0.57, 1.00, 0.75, 0.38, 0.38, 0.50, 0.12, 0.75, 0.57, ~
## $ rzRecTds        <dbl> 0.43, 0.14, 0.50, 0.12, 0.25, 0.17, 0.00, 0.25, 0.29, ~
## $ rzRecTargPct    <dbl> 26.59, 17.66, 27.85, 9.20, 16.81, 32.22, 6.69, 22.86, ~
## $ rzRecRecPct     <dbl> 16.66, 16.07, 14.70, 5.95, 9.38, 27.77, 1.25, 14.69, 1~
## $ rzRecTdPct      <dbl> 28.57, 14.29, 50.00, 12.50, 25.00, 16.67, 0.00, 8.34, ~
## $ ezRecTarg       <dbl> 0.29, 0.14, 1.00, 0.25, 0.38, 0.50, 0.00, 0.38, 0.43, ~
## $ ezRecTds        <dbl> 0.14, 0.00, 0.50, 0.12, 0.12, 0.17, 0.00, 0.25, 0.29, ~
## $ ezRecTargPct    <dbl> 8.73, 2.39, 19.30, 3.35, 5.44, 15.00, 0.00, 10.43, 10.~
## $ ezRecRecPct     <dbl> 2.39, 0.00, 10.32, 1.79, 3.12, 5.55, 0.00, 8.34, 9.51,~
## $ ezRecTdPct      <dbl> 7.14, 0.00, 50.00, 12.50, 12.50, 16.67, 0.00, 8.34, 28~
## $ rushCarries     <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.12, 0.29, ~
## $ rush40s         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
```

```
## $ rushYds         <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.25, 0.57, ~
## $ rushYds100Games <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ rushTds         <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, ~
## $ rushTd40s       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ ypc             <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.25, 0.57, ~
## $ yac             <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.29, ~
## $ rushTa          <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, ~
## $ tat             <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, ~
## $ fumbles         <dbl> 0.14, 0.14, 0.12, 0.00, 0.12, 0.00, 0.25, 0.00, 0.29, ~
## $ rzRushCarries   <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, ~
## $ rzRushTds       <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, ~
## $ rzRushPct       <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, ~
## $ rzRushTdPct     <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, ~
## $ i5RushCarries   <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, ~
## $ i5RushTds       <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, ~
## $ i5RushPct       <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, ~
## $ i5RushTdPct     <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, ~
## $ patConversions  <dbl> 0.14, 0.00, 0.00, 0.00, 0.00, 0.17, 0.00, 0.00, 0.00, ~
## $ patAttempts     <dbl> 0.14, 0.00, 0.12, 0.00, 0.00, 0.17, 0.00, 0.00, 0.00, ~
## $ fantasyPts.x    <dbl> 7.77, 7.90, 9.44, 7.80, 7.49, 3.28, 8.61, 9.81, 5.66, ~
## $ ptsPerSnap      <dbl> 0.11, 0.17, 0.22, 0.14, 0.13, 0.07, 0.17, 0.17, 0.16, ~
## $ ptsPerTouch     <dbl> 1.97, 1.48, 2.04, 1.55, 1.89, 1.63, 1.35, 2.38, 1.75, ~
## $ fantasyPts.y    <dbl> 89.1, 103.9, 116.3, 80.2, 164.7, 32.4, 158.4, 170.7, 8~
```

I vertically joined the weekly data sets so that I can compile all data into one dataframe. This process adds more records to the dataframe, but does not accurately represent the idea of predicting future outcomes. In the real world, we can compile all of the past weeks and use them to predict he future. However, the way I set up the dataframes makes it so that week 16 (in the one-week sample size) is trying to predict the year end points, even though in the real world we already have weeks 1-15 to predict the year end points.

Data Analysis

```
##                   PC1      PC2     PC3    PC4     sum
## fantasyPts.y 1100.222 1737.020  54.144 54.144 953.971
## rzRecTdPct    691.488  253.494 365.032 43.218 382.126
## ezRecTdPct    133.664   65.894 158.930 36.692  90.383
## rzRecRecPct    44.501   53.948 162.834 16.735  52.047
## rzRecTargPct   27.098   28.955 126.550  0.064  32.862
## recYds         14.480   18.457   2.660 95.058  17.338
## ezRecRecPct     4.094    4.752   1.852  3.440   3.413
## ezRecTargPct    3.451    3.494   0.192  7.541   2.864
## fantasyPts.x    0.348    0.020   0.106  0.573   0.186
## depth           0.038    0.001   0.013  0.011   0.017
## recTarg         0.016    0.021   0.001  0.002   0.013
## recRec          0.007    0.008   0.002  0.024   0.007
## ptsPerTouch     0.004    0.003   0.000  0.001   0.002
## rzRecTarg       0.002    0.000   0.001  0.000   0.001
## ypr             0.000    0.001   0.000  0.015   0.001
```

```
## rzRecRec        0.000    0.000    0.000  0.000   0.000
## recTds          0.000    0.000    0.000  0.000   0.000
## rzRecTds        0.000    0.000    0.000  0.000   0.000
## ezRecTarg       0.000    0.000    0.000  0.000   0.000
## ezRecTds        0.000    0.000    0.000  0.000   0.000
```

These first four principal components account for more than 85% of the covariance in the data (single-week sample). The numbers for each variable describe the contribution towards each principal component, meaning that those variables explain a substantial amount of the rest of the data's variation.

The PCA reports that year-end fantasy points are the most responsible for explaining covariance in the data. Redzone receiving touchdowns percentage is the second most responsible, followed by other percentage based variables and receiving yards. These percentage based variables attempt to explain where a TE get targeted on the field: with a high percentage, then their receptions or targets were mostly in the redzone or endzone.

```
##                   sum
## fantasyPts.x  13.262
## recTds        11.612
## rzRecTds      11.189
## rzRecTdPct    10.854
## ptsPerTouch   10.723
## rzRecRec      10.407
## ezRecTds      10.177
## rzRecTarg      9.821
## ezRecTdPct     9.660
## recYds         9.440
## ezRecTarg      9.090
## ptsPerSnap     8.710
## ezRecRecPct    8.018
## rushTds        7.838
## rzRushTds      7.838
## rzRushTdPct    7.838
## i5RushCarries  7.755
## i5RushPct      7.755
## rzRushCarries  7.711
## rzRushPct      7.711
```

This is extremely similar to the PCA. This table shows the sum of all correlations within the correlation matrix. Each number for each variable represents the gross correlation it has with all other variables. The difference between this and the PCA is that the PCA uses covariance instead of correlation, covariance is dependent upon units whereas correlation is a unitless and scaled metric.

```
##                    colnames cor.x varco  skew
## fantasyPts.x     fantasyPts.x 1.000 0.920 0.868
## recYds                 recYds 0.834 0.898 0.842
## recRec                 recRec 0.778 0.721 0.882
## recTds                 recTds 0.756 2.291 0.436
## rzRecTds             rzRecTds 0.722 2.329 0.429
## recTarg               recTarg 0.708 0.720 1.010
## rzRecTdPct         rzRecTdPct 0.657 2.203 0.454
## rzRecRec             rzRecRec 0.648 1.509 0.663
## rzRecTarg           rzRecTarg 0.596 1.268 0.788
## fantasyPts.y     fantasyPts.y 0.544 0.751 0.401
## ezRecTds             ezRecTds 0.537 2.916 0.343
```

```
## ptsPerTouch       ptsPerTouch 0.494 0.713 0.555
## ptsPerSnap         ptsPerSnap 0.478 1.232 0.608
## ezRecTdPct         ezRecTdPct 0.472 2.855 0.350
## ezRecTarg           ezRecTarg 0.453 2.064 0.484
## recYds100Games recYds100Games 0.449 6.894 0.145
## ypr                       ypr 0.281 0.608 0.521
## ezRecRecPct       ezRecRecPct 0.276 3.503 0.285
## recRec40s           recRec40s 0.269 6.182 0.162
## depth                   depth 0.262 0.791 0.278
```

colnames: variables in the data cor.x: correlation between the variable and the given week's fantasy points varco: the relative variance of each variable skew: skewness of each variable (positive number = positive skew)

In the single-week subset, the variables most associated with the current week's fantasy points are: yards, receptions, touchdowns and redzone receiving touchdowns. The variables with the highest variance are of course the rushing variables, stats that simply do not apply to most TEs. There is a negative correlation (-0.3167) between the correlation of the year-end points and the variance of each variable; this makes sense because if something is less stable it would be assumed that it would also provide less predictability. Almost all variables are skewed positively, with the exception of catchrate.

```
##                      colnames cor.x varco  skew
## fantasyPts.x     fantasyPts.x 1.000 0.685 0.749
## recYds                 recYds 0.864 0.670 0.451
## recRec                 recRec 0.814 0.561 0.922
## recTarg               recTarg 0.763 0.562 0.718
## recTds                 recTds 0.754 1.588 0.630
## rzRecTds             rzRecTds 0.720 1.602 0.624
## rzRecRec             rzRecRec 0.690 1.066 0.938
## fantasyPts.y     fantasyPts.y 0.660 0.636 0.457
## rzRecTarg           rzRecTarg 0.658 0.904 0.503
## rzRecTdPct         rzRecTdPct 0.654 1.498 0.668
## ptsPerSnap         ptsPerSnap 0.609 0.705 0.518
## ezRecTds             ezRecTds 0.537 2.008 0.498
## recYds100Games recYds100Games 0.531 4.411 0.227
## ezRecTarg           ezRecTarg 0.492 1.403 0.713
## ezRecTdPct         ezRecTdPct 0.464 1.959 0.511
## ptsPerTouch       ptsPerTouch 0.456 0.467 0.688
## recRec40s           recRec40s 0.311 4.081 0.245
## depth                   depth 0.276 0.555 0.597
## ypr                       ypr 0.274 0.413 0.457
## recTd40s             recTd40s 0.249 7.739 0.129
```

```
##                      colnames cor.x varco  skew
## fantasyPts.x     fantasyPts.x 1.000 0.586 0.645
## recYds                 recYds 0.885 0.577 0.938
## recRec                 recRec 0.831 0.491 0.614
## recTarg               recTarg 0.792 0.490 0.842
## recTds                 recTds 0.760 1.266 0.790
## fantasyPts.y     fantasyPts.y 0.730 0.571 0.849
## rzRecTds             rzRecTds 0.724 1.272 0.786
## rzRecRec             rzRecRec 0.715 0.865 0.545
## rzRecTarg           rzRecTarg 0.688 0.741 0.015
## rzRecTdPct         rzRecTdPct 0.649 1.184 0.845
```

```
## ptsPerSnap         ptsPerSnap 0.633 0.577 0.540
## recYds100Games recYds100Games 0.581 3.564 0.281
## ezRecTds             ezRecTds 0.538 1.608 0.622
## ezRecTarg           ezRecTarg 0.508 1.138 0.879
## ezRecTdPct         ezRecTdPct 0.459 1.574 0.635
## ptsPerTouch       ptsPerTouch 0.437 0.380 0.840
## recRec40s           recRec40s 0.342 3.288 0.304
## depth                   depth 0.287 0.451 0.717
## ypr                       ypr 0.285 0.333 0.077
## recTd40s             recTd40s 0.259 5.585 0.179
```

```
##                      colnames cor.x varco   skew
## fantasyPts.x     fantasyPts.x 1.000 0.399 -0.039
## recYds                 recYds 0.935 0.414 -0.798
## ptsPerSnap         ptsPerSnap 0.890 0.313 -0.187
## recRec                 recRec 0.886 0.345 -0.092
## recTarg               recTarg 0.873 0.338 -0.548
## fantasyPts.y     fantasyPts.y 0.828 0.440  1.094
## rzRecRec             rzRecRec 0.794 0.571  0.501
## rzRecTarg           rzRecTarg 0.786 0.479  0.310
## recTds                 recTds 0.757 0.740  0.094
## rzRecTds             rzRecTds 0.738 0.755  0.050
## recYds100Games recYds100Games 0.698 2.226  0.449
## rzRecTdPct         rzRecTdPct 0.658 0.668 -0.148
## ezRecTarg           ezRecTarg 0.530 0.715  1.398
## ezRecTds             ezRecTds 0.509 0.968  1.033
## ptsPerTouch       ptsPerTouch 0.500 0.205  0.066
## ypt                       ypt 0.477 0.205 -1.081
## ypr                       ypr 0.450 0.197  0.507
## ezRecTdPct         ezRecTdPct 0.431 0.936  1.069
## recDrops             recDrops 0.407 0.837  1.195
## patAttempts       patAttempts 0.339 2.269  0.441
```

Here are the variables most positively associated with the average fantasy points per set of weeks:

One-Week	Two-Weeks	Three-Weeks	First-Half
Yards	Yards	Yards	Yards
Rec	Rec	Rec	Rec
Tds	Targets	Targets	Targets
rzRecTds	Tds	Tds	rzRec

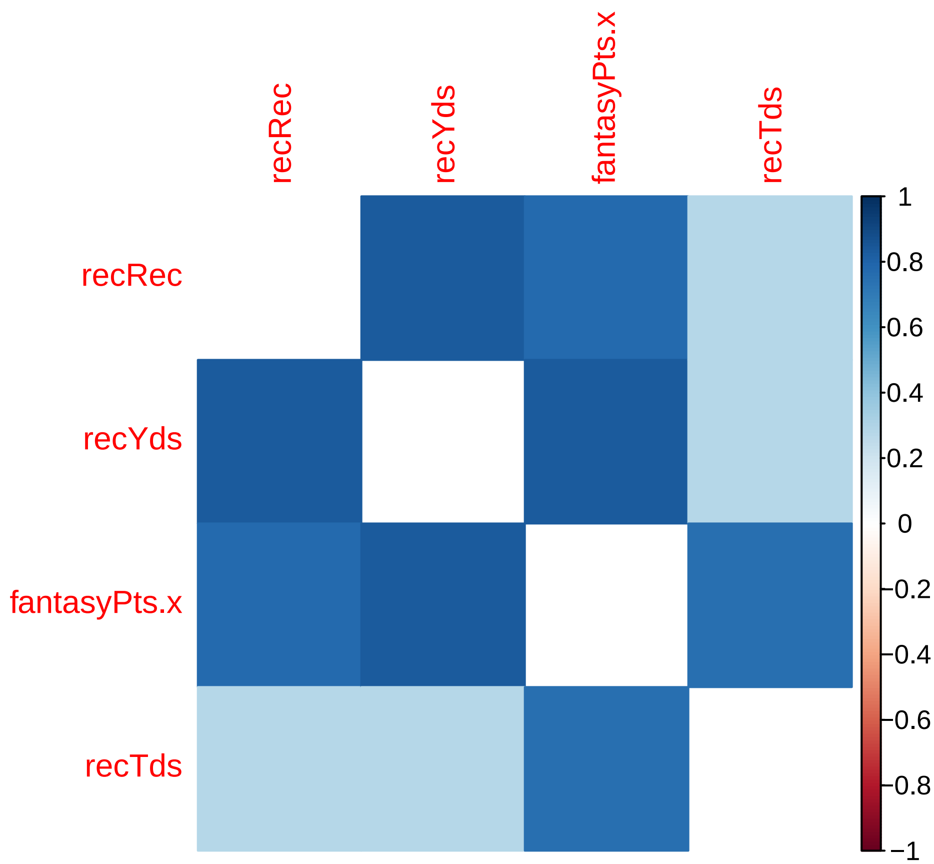
As the samples use more and more weeks, the highest associated variables with fantasy points become more and more stable and predictive.

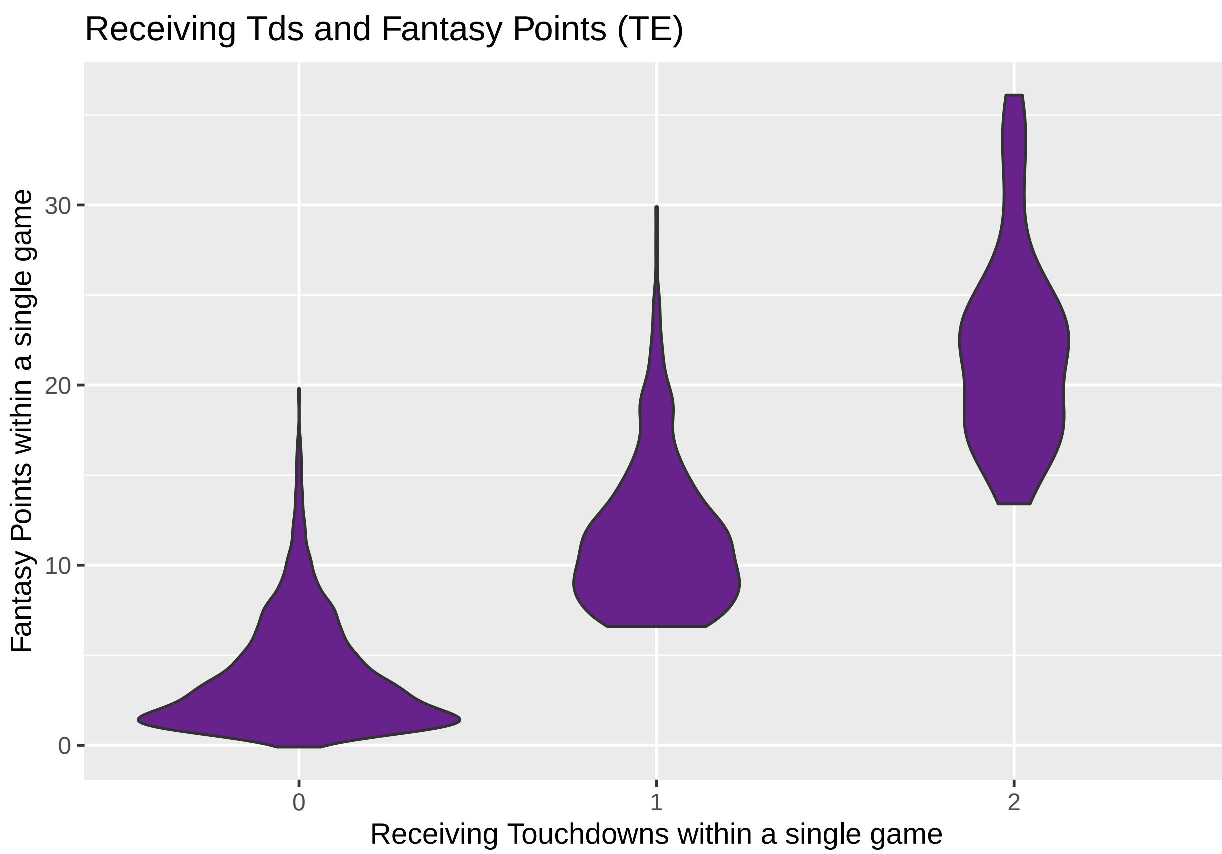
```
##          colnames   One   Two Three First_Half average
## 1          recYds 0.583 0.690 0.746      0.813 0.70800
## 2         recTarg 0.661 0.722 0.753      0.693 0.70725
## 3          recRec 0.614 0.691 0.728      0.730 0.69075
## 4    fantasyPts.x 0.544 0.660 0.730      0.828 0.69050
## 5       rzRecTarg 0.408 0.500 0.563      0.588 0.51475
## 6        rzRecRec 0.310 0.403 0.468      0.614 0.44875
## 7  recYds100Games 0.288 0.396 0.462      0.640 0.44650
```

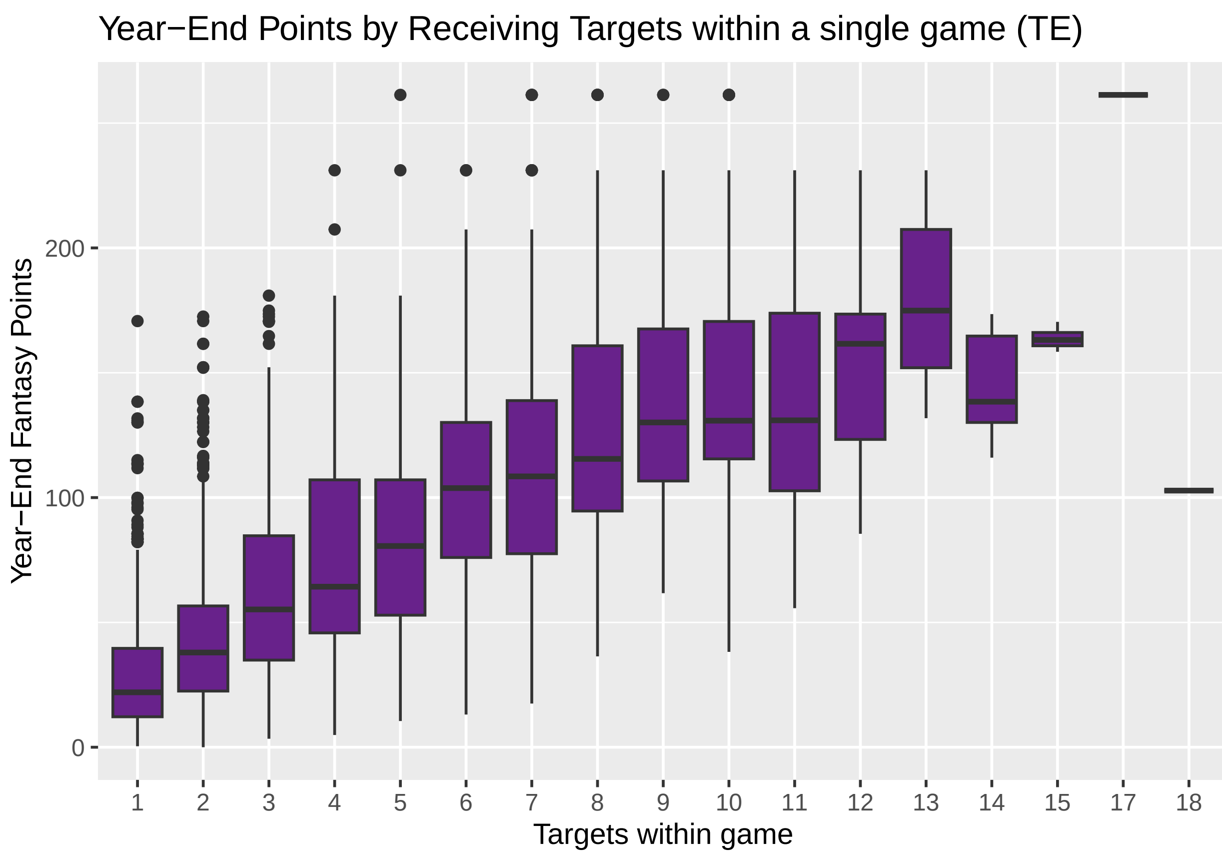
```
## 8          recTds 0.225 0.320 0.400      0.573 0.37950
## 9        rzRecTds 0.205 0.290 0.362      0.551 0.35200
## 10     ptsPerSnap 0.078 0.230 0.323      0.704 0.33375
## 11      ezRecTarg 0.241 0.314 0.371      0.377 0.32575
## 12     rzRecTdPct 0.168 0.238 0.291      0.444 0.28525
## 13       ezRecTds 0.146 0.210 0.271      0.394 0.25525
## 14      recRec40s 0.135 0.206 0.250      0.369 0.24000
## 15            ypr 0.120 0.168 0.207      0.446 0.23525
## 16          depth 0.195 0.231 0.267      0.224 0.22925
## 17       recDrops 0.185 0.227 0.267      0.228 0.22675
## 18    patAttempts 0.118 0.158 0.182      0.388 0.21150
## 19     ezRecTdPct 0.118 0.169 0.216      0.309 0.20300
## 20    ptsPerTouch 0.050 0.106 0.146      0.400 0.17550
```

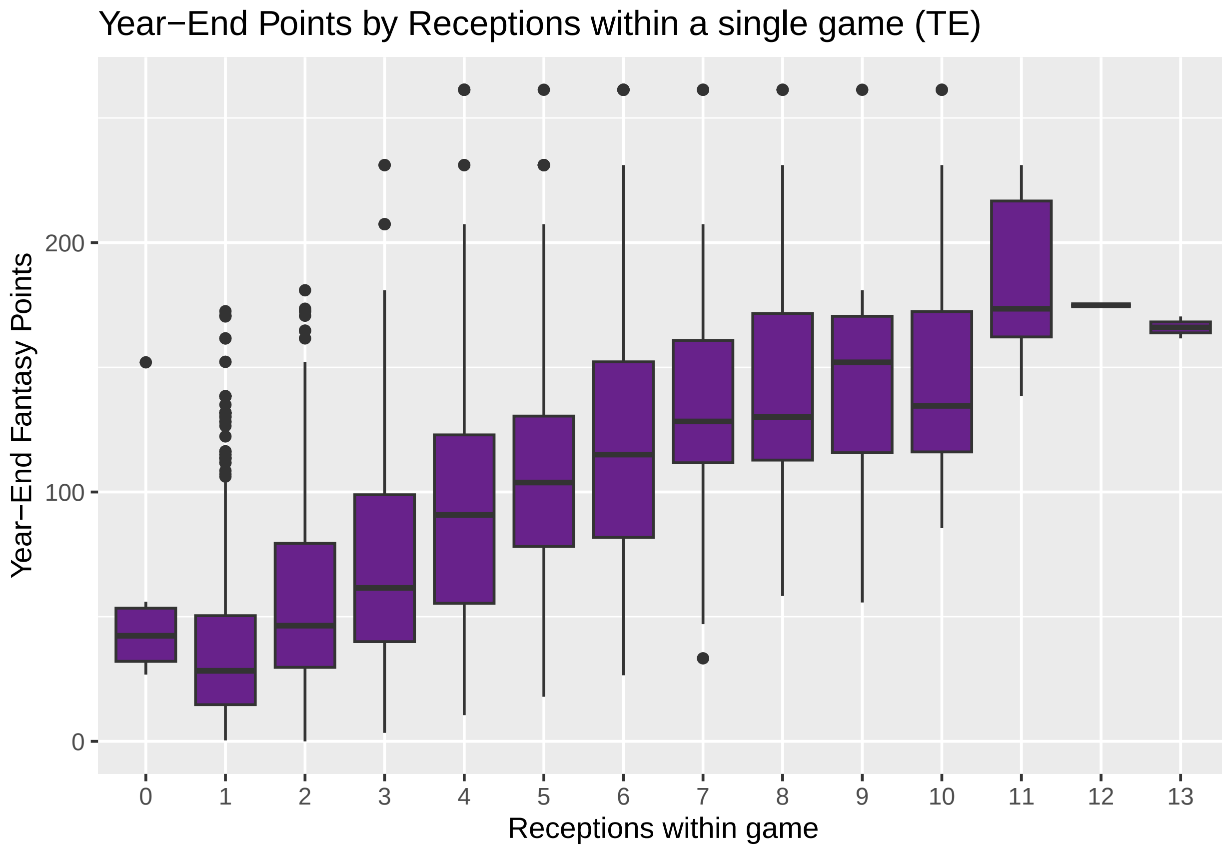
This is a table that shows the association between each variable and the year-end fantasy points. The most intuitive pattern is the increase in correlation as the sample size increases. Some variables actually decrease (targets, ADOT, drops); even though the variable is stable, the sample size becomes so large (first half of season) that the variables that contribute more to actual fantasy points become relatively more predictive (yards, touchdowns).

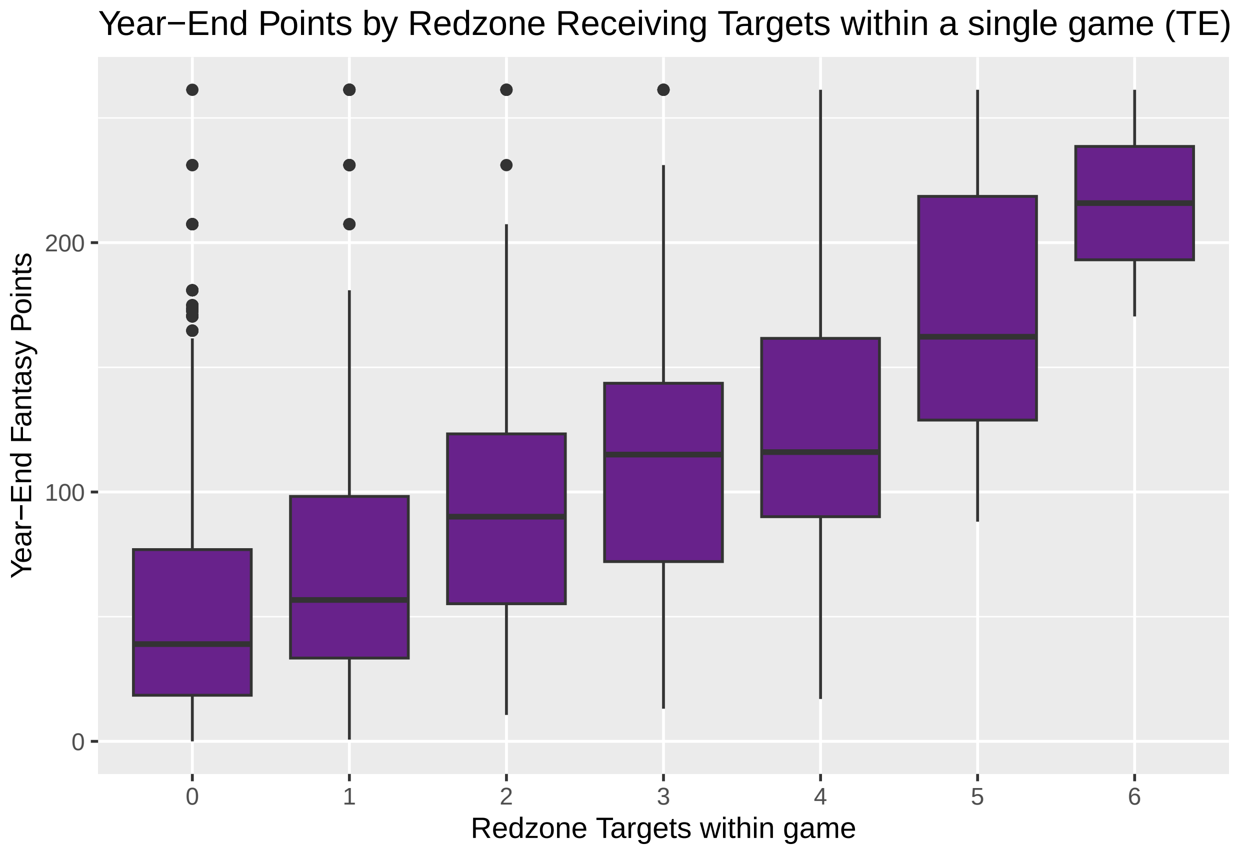
Data Visualization

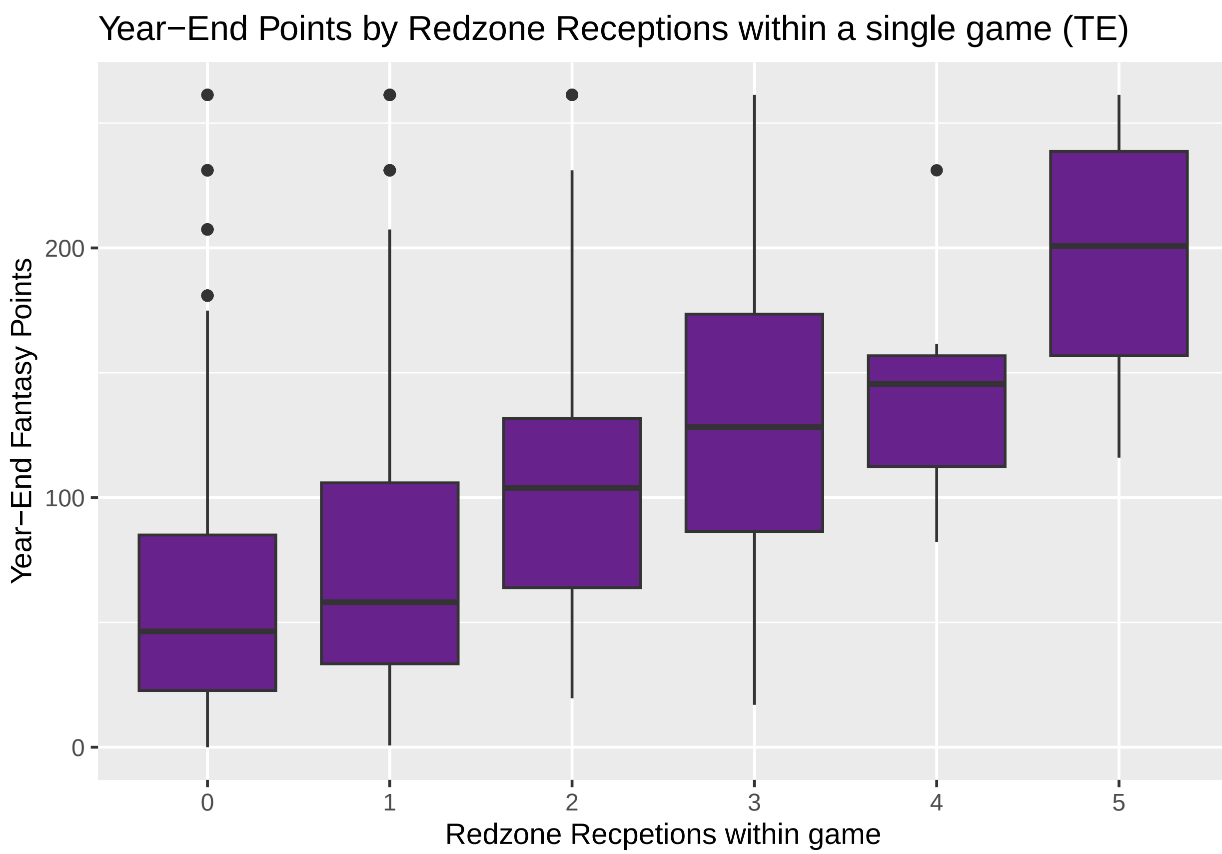


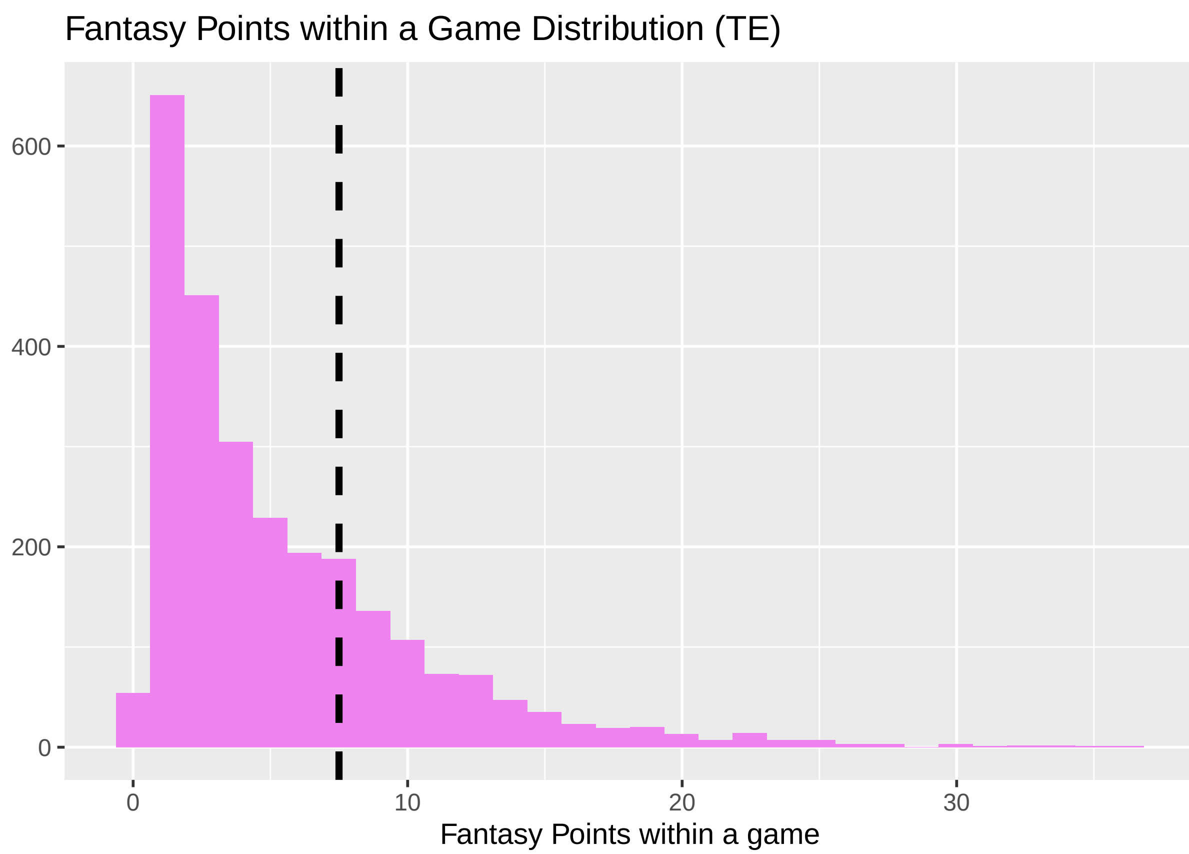

Receiving Tds and Fantasy Points (TE)
Fantasy Points within a single game
30
20
10
0
0
1
2
Receiving Touchdowns within a single game


Year−End Points by Receiving Targets within a single game (TE)
Year−End Fantasy Points
0
100
200
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
17
18
Targets within game

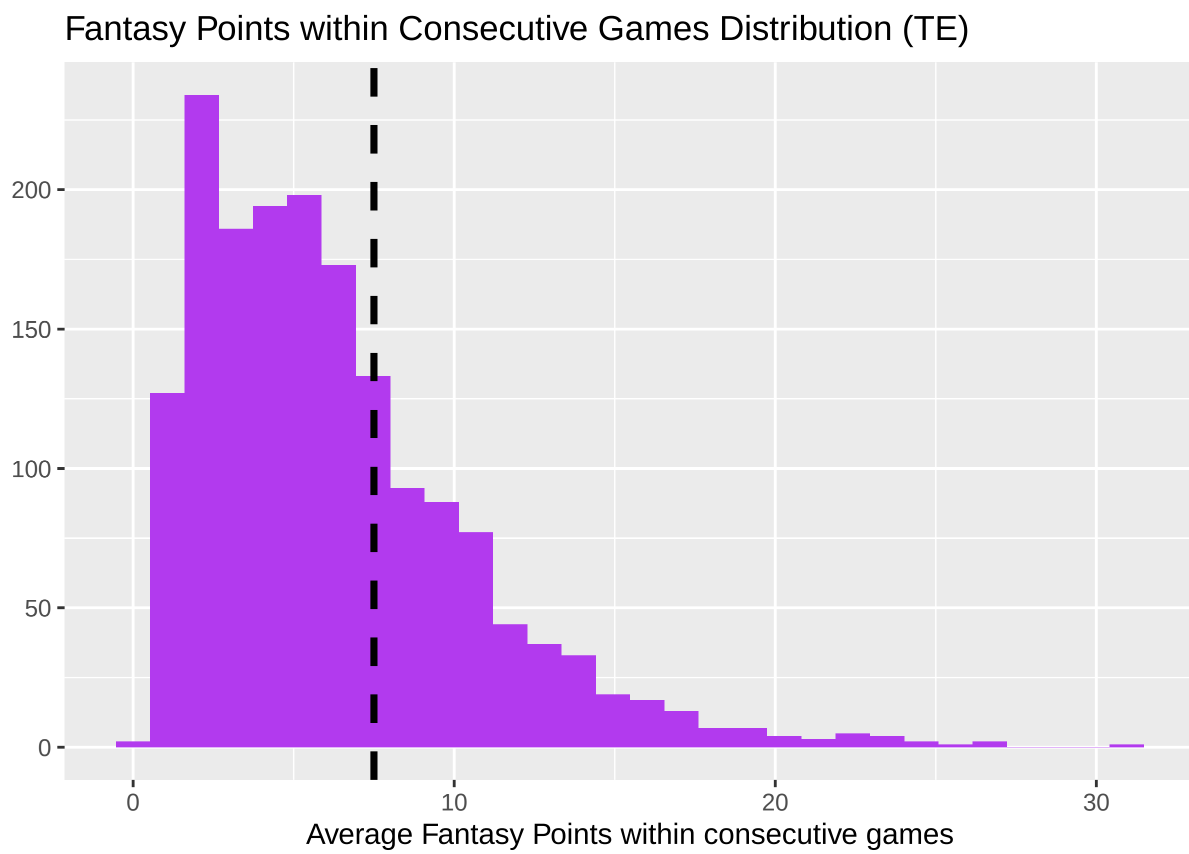

Year−End Points by Receptions within a single game (TE)
Year−End Fantasy Points
0
100
200
0
1
2
3
4
5
6
7
8
9
10
11
12
13
Receptions within game

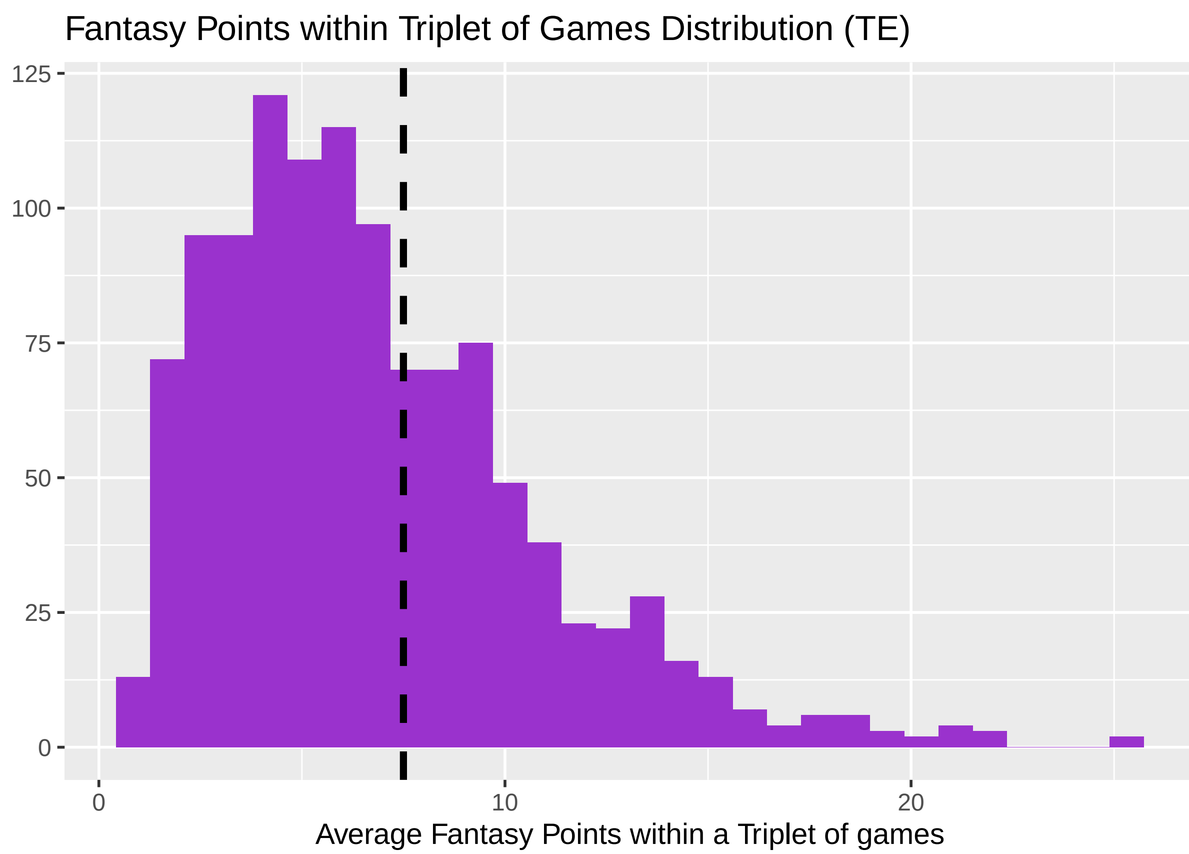

Year−End Points by Redzone Receiving Targets within a single game (TE)
Year−End Fantasy Points
200
100
0
0
1
2
3
4
5
6
Redzone Targets within game

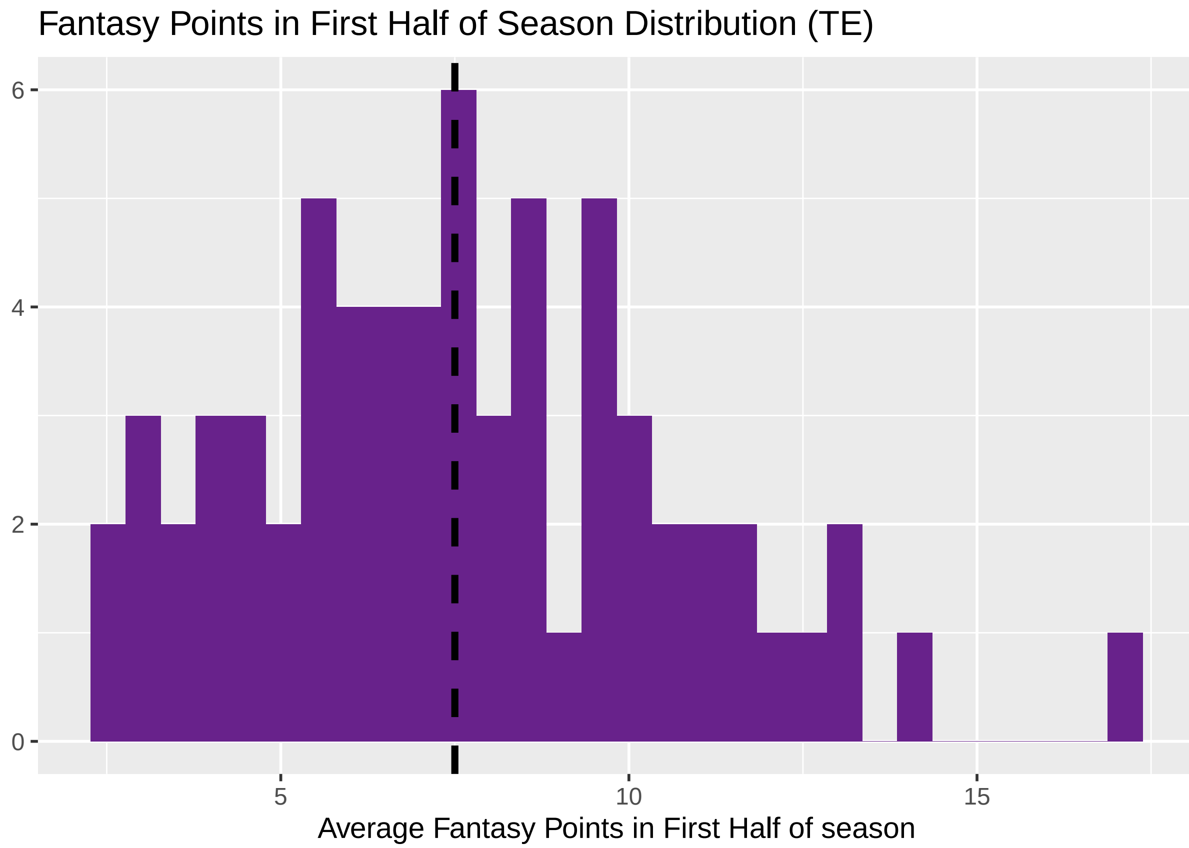

Year−End Points by Redzone Receptions within a single game (TE)
Year−End Fantasy Points
0
100
200
0
1
2
3
4
5
Redzone Recpetions within game



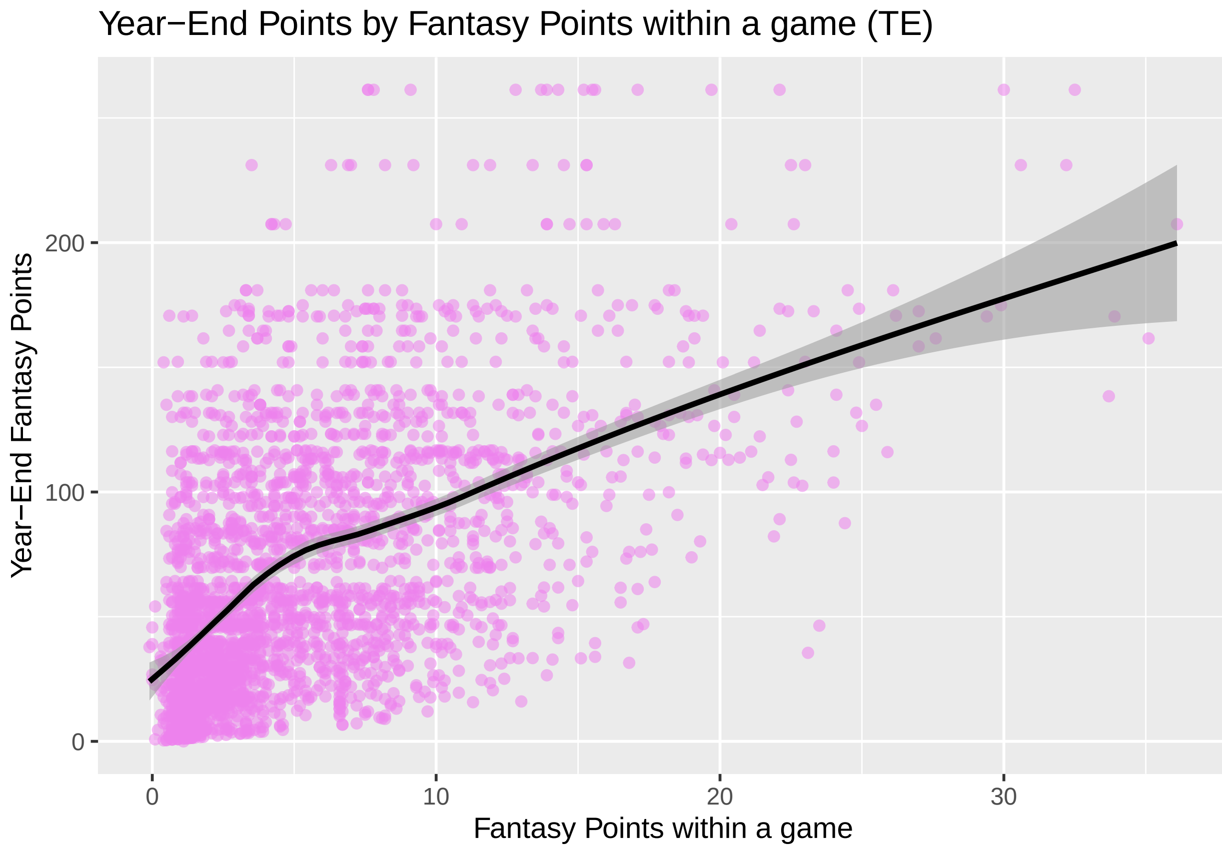
Fantasy Points within a Game Distribution (TE)
600
400
200
0
0
10
20
30
Fantasy Points within a game

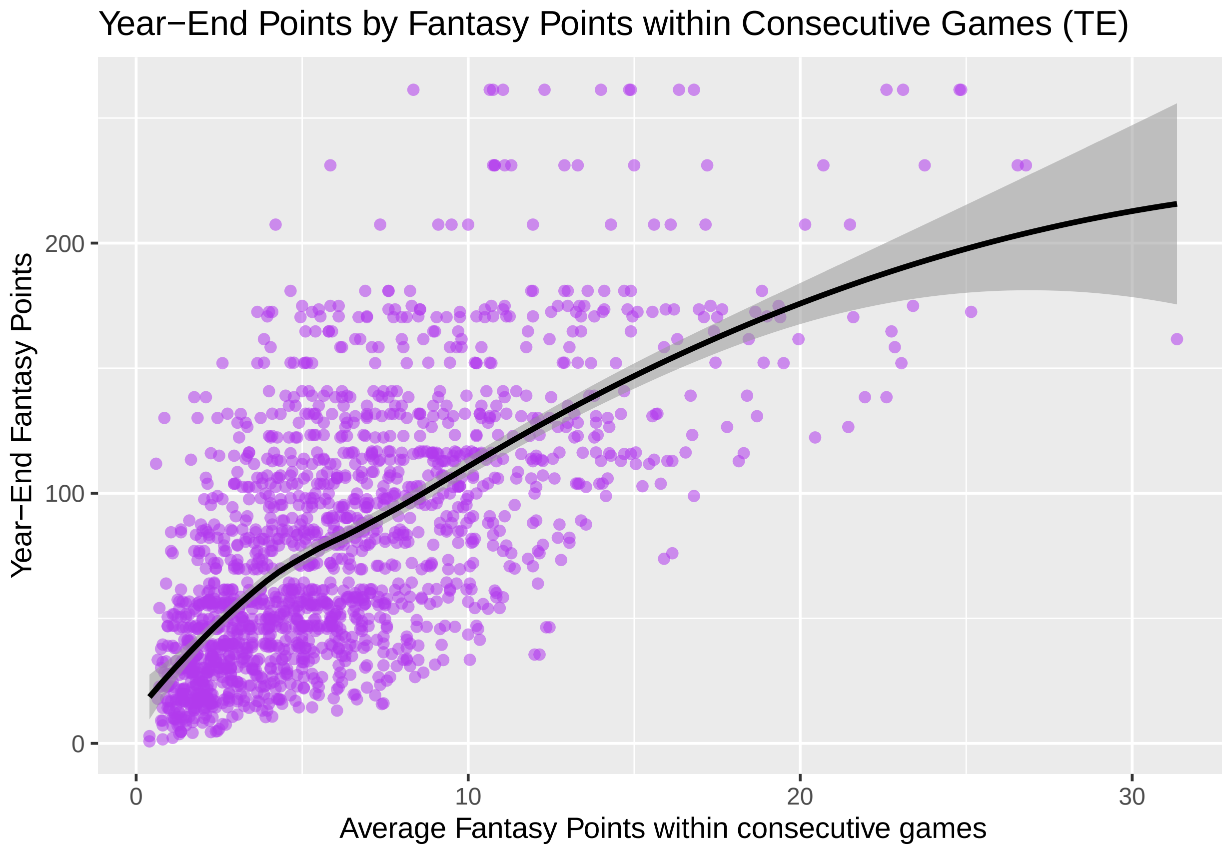

Fantasy Points within Consecutive Games Distribution (TE)
200
150
100
50
0
0
10
20
30
Average Fantasy Points within consecutive games

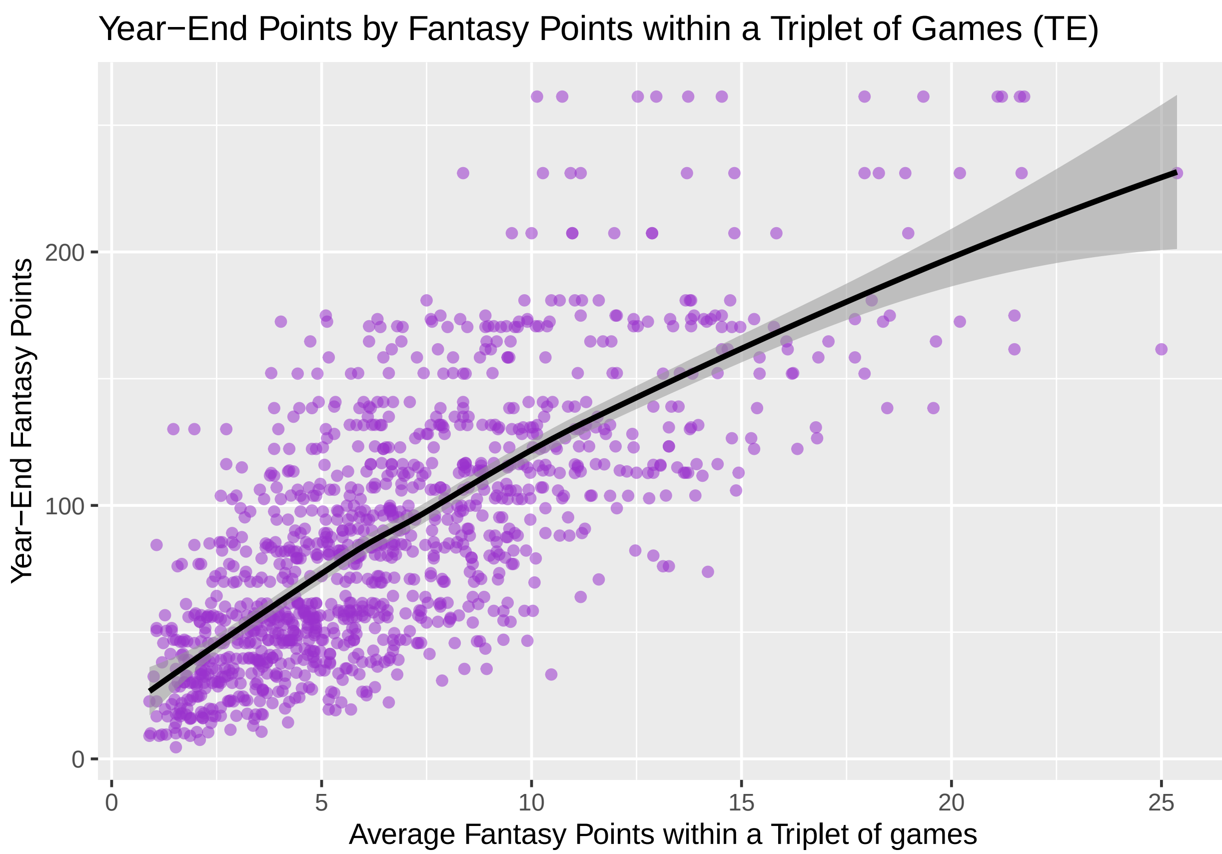

Fantasy Points within Triplet of Games Distribution (TE)
125
100
75
50
25
0
0
10
20
Average Fantasy Points within a Triplet of games

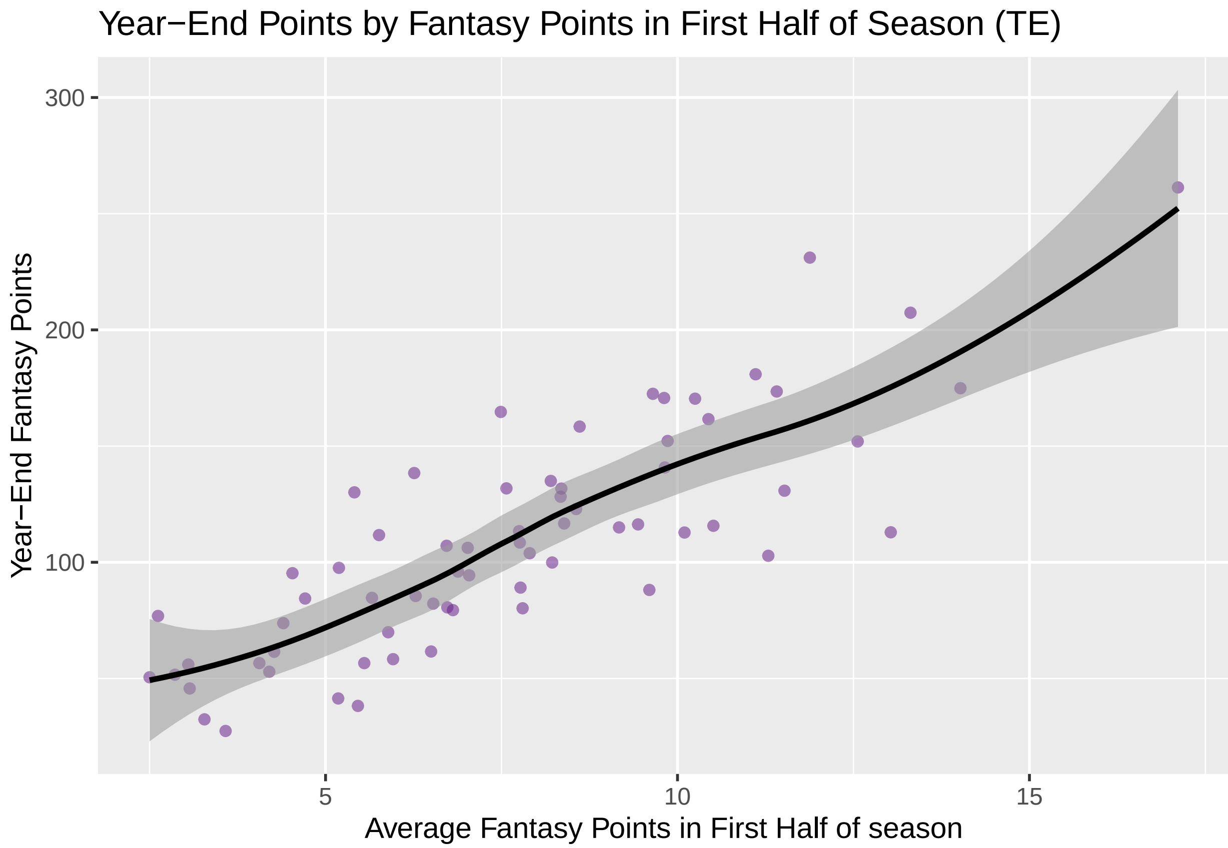


These are distribution plots of fantasy points over the different sample sizes. The dotted line is set at 7.5 fantasy points, it acts as a reference point and represents the points at which the last starting TE (12 team) would average over a given year.

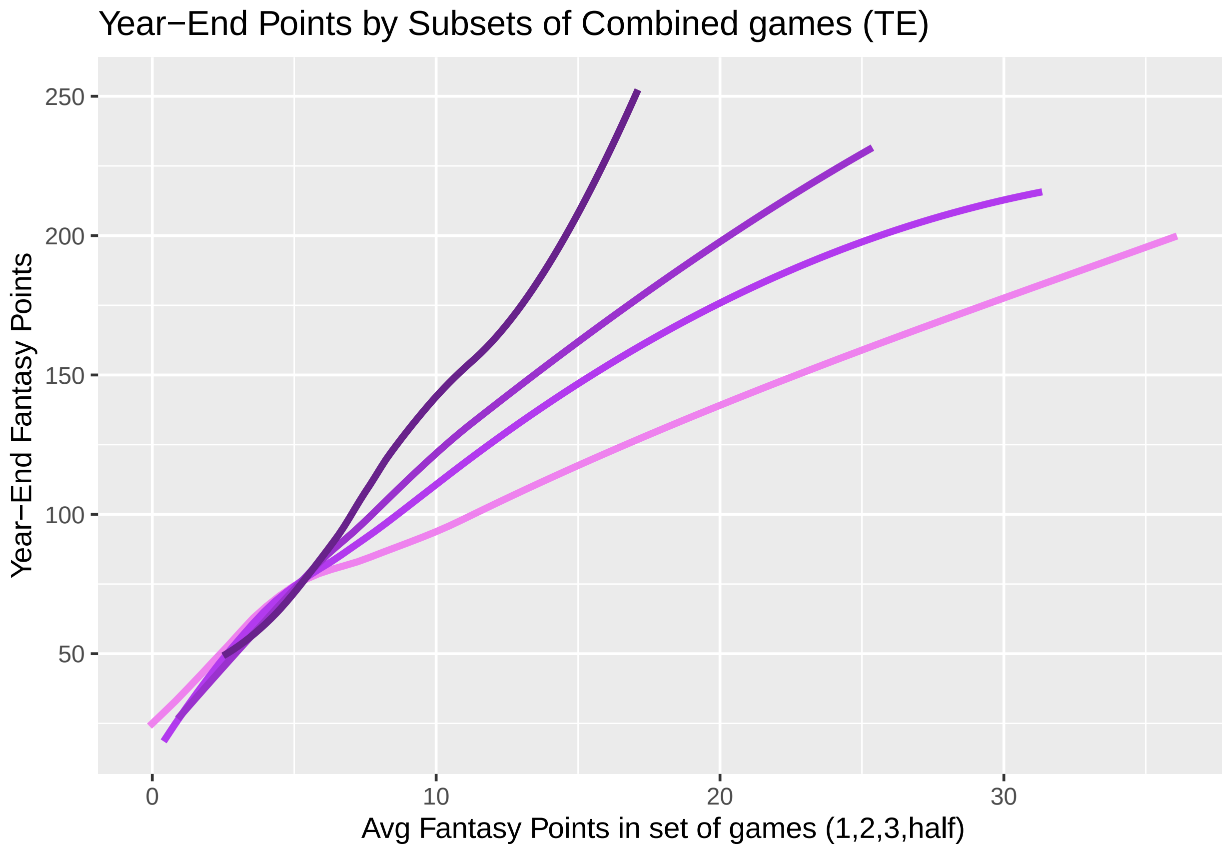

Year−End Points by Fantasy Points within a game (TE)
Year−End Fantasy Points
Fantasy Points within a game
0
100
200
0
10
20
30


Year−End Points by Fantasy Points within Consecutive Games (TE)
Year−End Fantasy Points
0
100
200
0
10
20
30
Average Fantasy Points within consecutive games

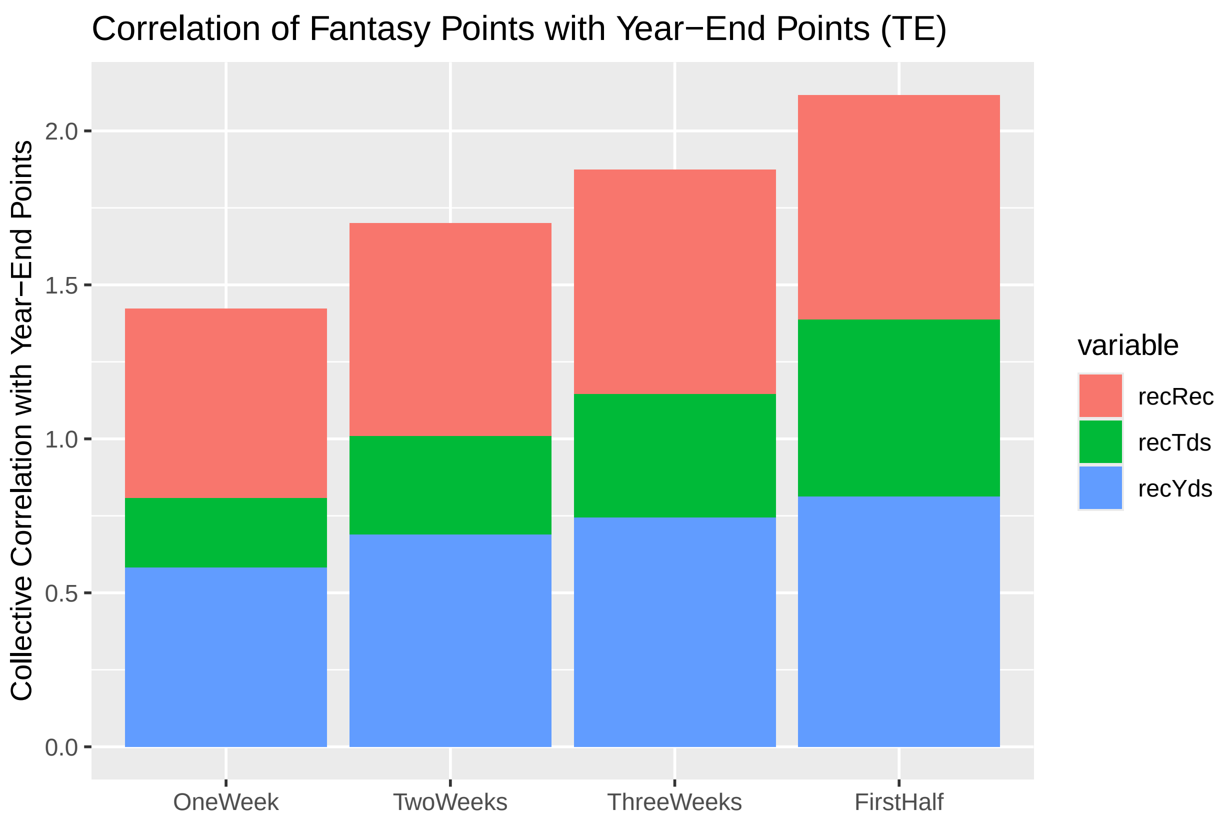

Year−End Points by Fantasy Points within a Triplet of Games (TE)
Year−End Fantasy Points
200
100
0
0
5
10
15
20
25
Average Fantasy Points within a Triplet of games



These plots are similar to the last, but also show the year-end fantasy points. As the sample size increases, the rolling average function line becomes more linear. There also seems to be no threshold (fantasy points) that characterizes a gross increase in predictive power.

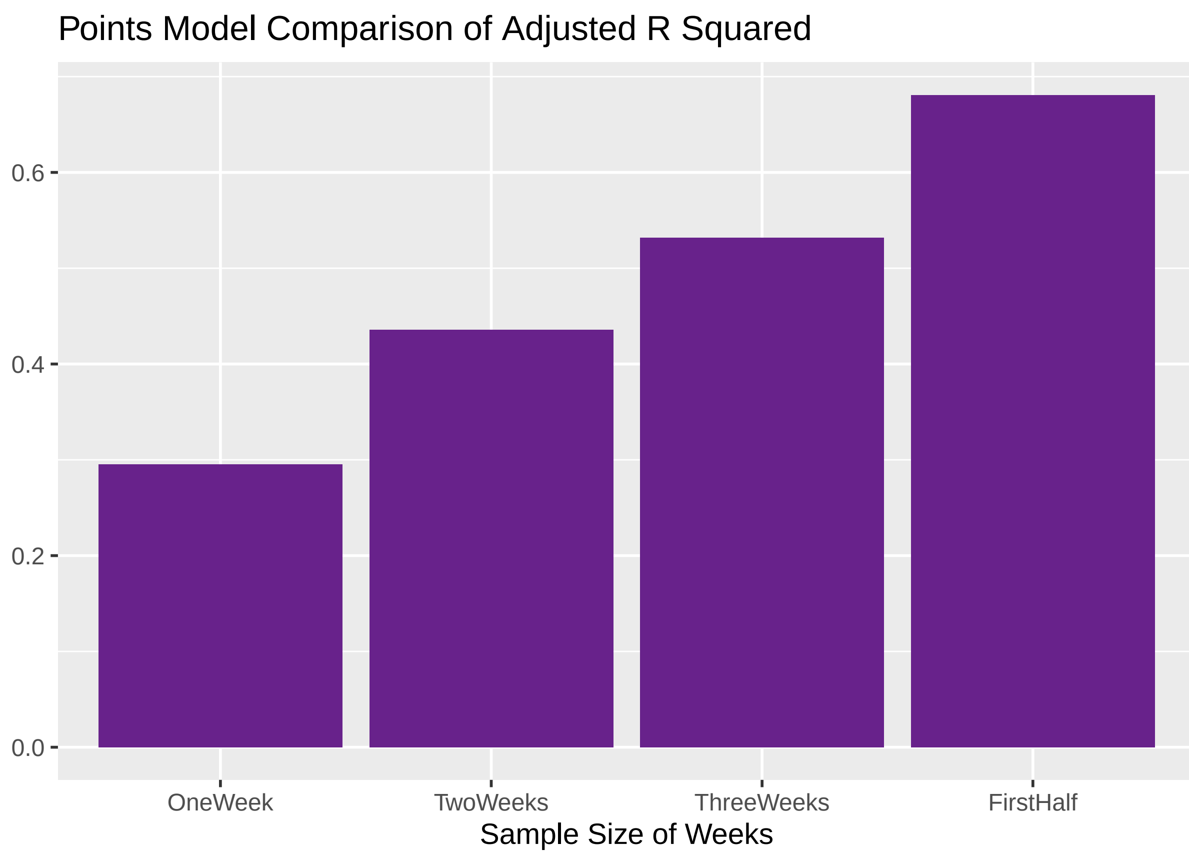


This plots the last four function lines together. It is clear that with increasing sample size there is also an increase in predictive power, however the increase seems to get progressively smaller.

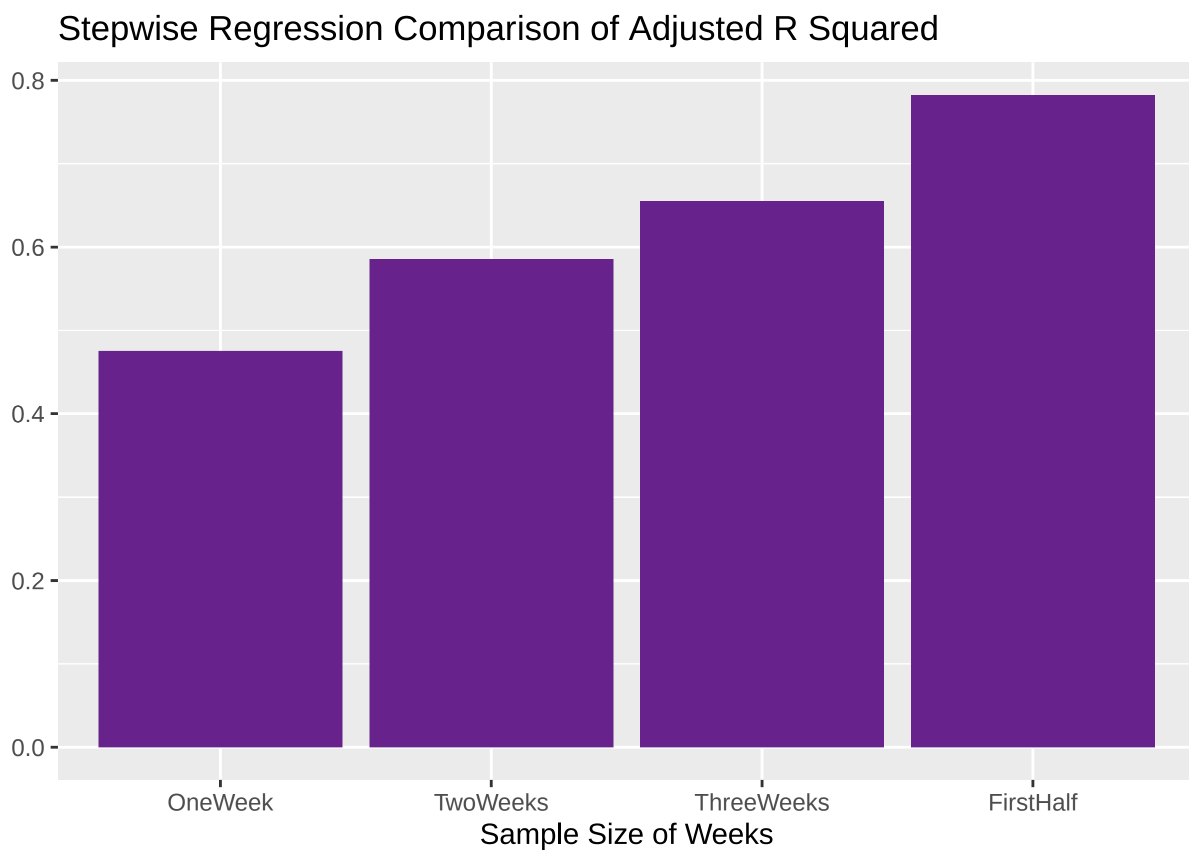


Of course, the cumulative correlation and individual correlations increase with sample size. However, the largest change in predictive power is with receiving touchdowns. This because touchdowns are infrequent and unpredictable in any single week, but are more frequent and more predictable over the course of multiple weeks.

Regression Analysis



These are the adjusted r squared of each model, which represents the percent of data that the model can predict adjusted for the complexity of the model. These models only use average fantasy points to predict year-end points. Obviously, the predictive power of each model gets better with larger sample sizes, however it also appears that the same pattern of decreasing rate of return exists. One Week does not surpass 50% variance explained, which is an intuitive threshold.



Thankfully, the more complex models (backwards stepwise regression) that use more variables are also significantly different in predictive power (all under 0.0001 p values) compared to the models that only use average fantasy points. This means that digging deeper into the data has revealed that there is an advantage in more complex models; not just looking at fantasy points.

- Number of Variables Used in Each Stepwise Model:
  - One Week: 16
  - Two Weeks: 22
  - Three Weeks: 21
  - First Half: 19

Next, I used the most highly associated variables from each of the data analysis techniques to compile a 'reduced' set of 12 variables. Using the anova test, there is a significant difference between all of the reduced models and their stepwise counterparts, meaning that the stepwise models are better. Here, I was hoping that the simplified models would be just as good as the stepwise models. The stepwise models have an average of 23 variables whereas the reduced models have only 12 variables.

It is also important to note that the margin of gain between reduced models (all except for the first half sample size) and the fantasy points models is larger than the margin of gain between the stepwise models and the reduced models.

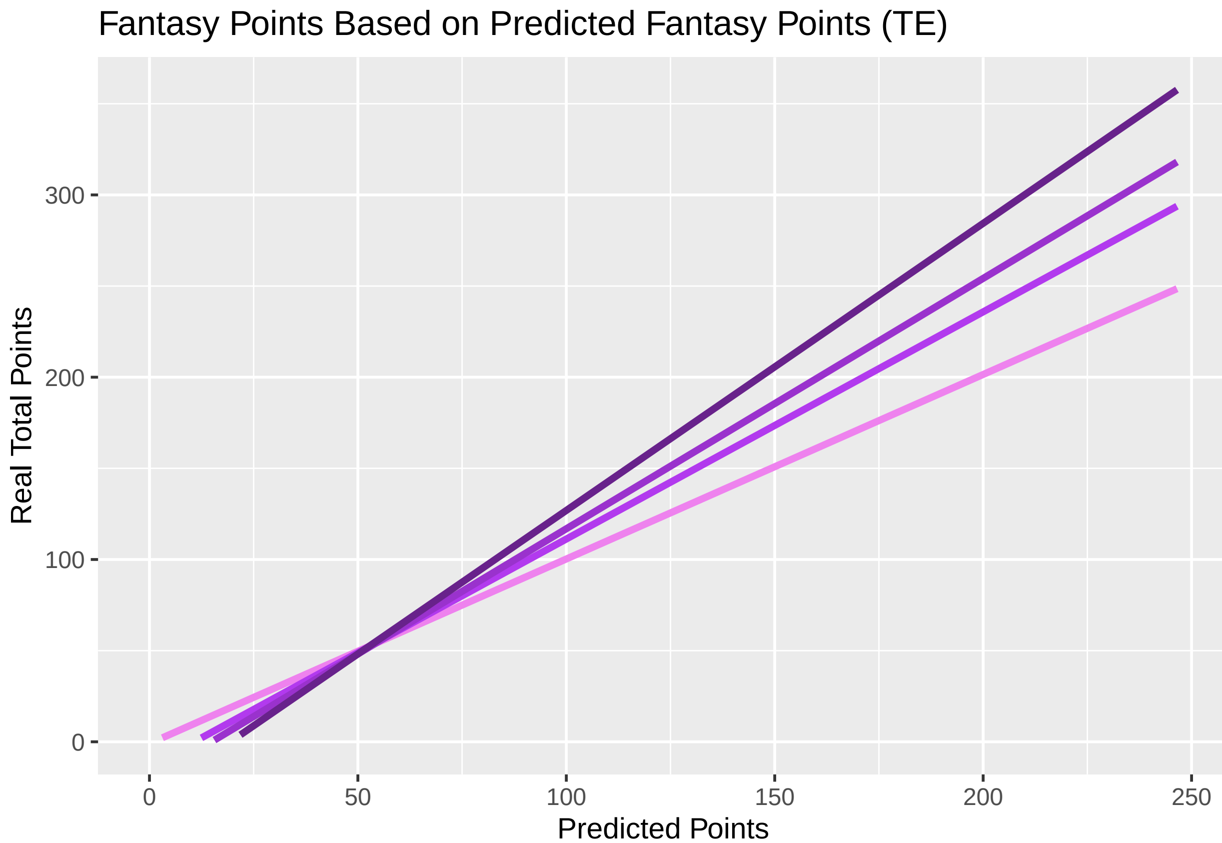
```
##
## Call:
## lm(formula = fantasyPts.y ~ recTarg + recYds + recTds + recTd40s +
##     recDrops + catch + depth + ypt + rzRecTdPct + ezRecTds +
##     ezRecTdPct + rushCarries + rushTa + i5RushCarries + patAttempts +
```

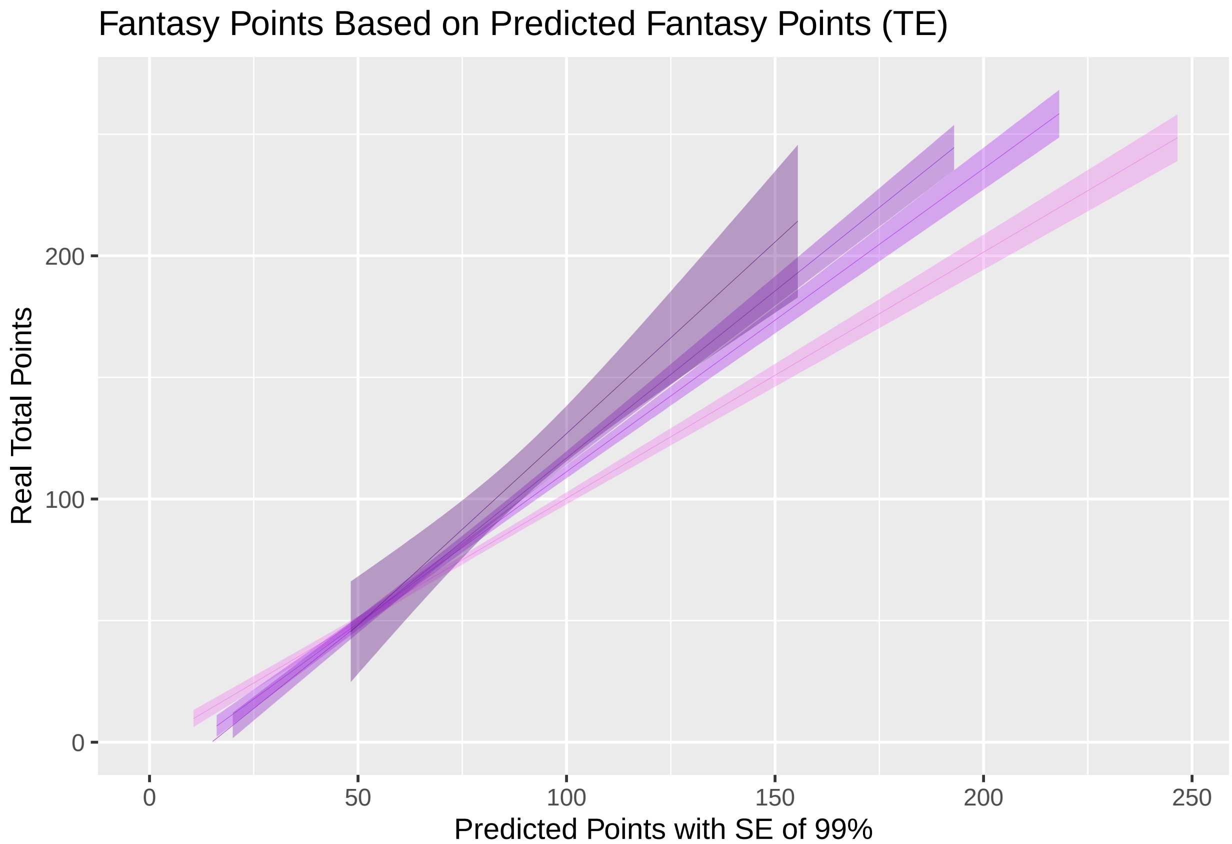
```
##     ptsPerSnap, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -117.951  -23.703   -7.028   18.261  172.700
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    31.97267    4.61060   6.935 5.10e-12 ***
## recTarg         8.64943    0.66925  12.924  < 2e-16 ***
## recYds          0.47095    0.07199   6.542 7.26e-11 ***
## recTds         33.71214    6.02761   5.593 2.46e-08 ***
## recTd40s      -15.90494   10.18153  -1.562  0.11838
## recDrops       -5.29975    2.02795  -2.613  0.00902 **
## catch          -0.09396    0.04651  -2.020  0.04348 *
## depth           0.61172    0.19144   3.195  0.00141 **
## ypt            -0.43328    0.22761  -1.904  0.05707 .
## rzRecTdPct     -0.30280    0.07231  -4.188 2.91e-05 ***
## ezRecTds      -31.21487   10.44806  -2.988  0.00284 **
## ezRecTdPct      0.36960    0.11762   3.142  0.00170 **
## rushCarries    10.99432    4.96613   2.214  0.02692 *
## rushTa         20.11546   10.32640   1.948  0.05152 .
## i5RushCarries  16.71036   10.93486   1.528  0.12659
## patAttempts    13.83740    4.64209   2.981  0.00290 **
## ptsPerSnap    -28.13930    4.38246  -6.421 1.60e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 36.58 on 2651 degrees of freedom
## Multiple R-squared:  0.4786, Adjusted R-squared:  0.4755
## F-statistic: 152.1 on 16 and 2651 DF,  p-value: < 2.2e-16
```

Here is the summary for the one week sample size stepwise reduced model. I believe that this is the model that should be used to predict future fantasy points, because it can be applied to the other sample sizes better than they can be applied to the one week sample size. Plus, this model uses the least amount of variables.

The "estimate" is the coefficient for each variable, meaning that for every one unit increase in the given variable there is a corresponding change in the predicted value of total fantasy points, regardless of the other variables. The "t value" and p value" both show the chance that these variables provide unique information on the predicted value of fantasy points more than that of the other variable (while assuming there should not be any uniqueness). There is high multicollinearity between the variables in the model, so the coefficients and significance measures are less accurate. These coefficients are also based on the one-week sample, so they may be different when applied to other sample sizes. However, these variables are the most important variables for predicting the year-end total fantasy points of a TE.

Predictive Analysis


Fantasy Points Based on Predicted Fantasy Points (TE)
Real Total Points
300
200
100
0
0
50
100
150
200
250
Predicted Points



The first plot shows the change in slope of real fantasy points scored and predicted fantasy points scored, based on the specified model. The slope slowly increases towards 1 with every increase in sample size. There still seems to be a diminishing rate of return for every increase in sample size.

The second plot shows the confidence intervals for the average real fantasy points based on predicted fantasy points at the 99% level. Based on the standard errors of the linear trend lines, the one-week sample is clearly different and worse than the others. However, the other samples all touch each other, so it is reasonable to assume that there is limited difference between the projections.

Data from of PFF