

# eda.Rmd

Brandon Leff

3/6/2020

Tyler start the backbone of the EDA file right here. The first chunk will be loading in the clean data we get at the end of the Data Management file. Since we don't have all DM done yet start with some univariate plots of variables and as DM continues the plots can just change with it.

## Read in Clean Data

```
nfl <- readRDS("~/MATH456/nfl.rds")
```

```
library(ggplot2)
library(pander)
```

## Univariate Analysis

```
#FirstDown
```

```
nfl$firstdown <- factor(nfl$FirstDown, labels = c("No First Down", "First Down"))
```

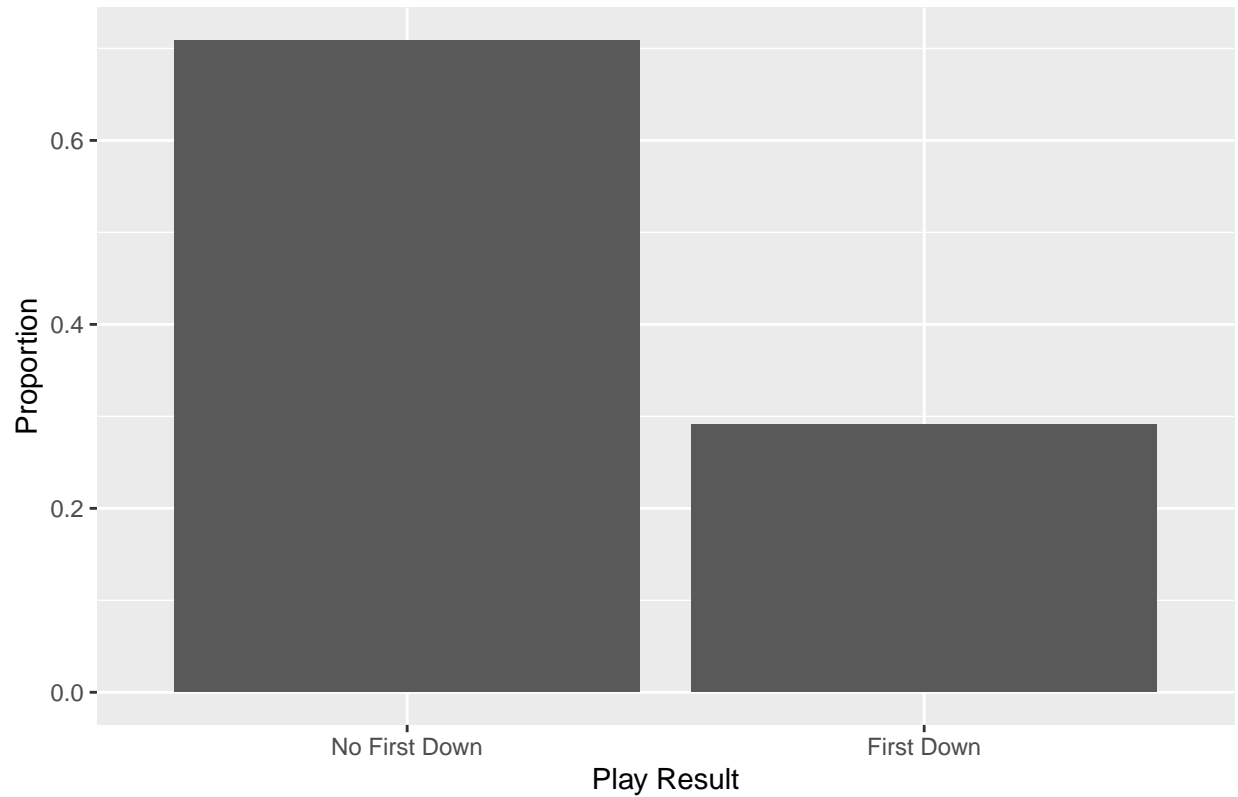
```
pander(summary(nfl$firstdown))
```

No First Down	First Down	NA's
238765	98085	25597

```
fd.props <- data.frame(prop.table(table(nfl$firstdown)))
```

```
ggplot(fd.props, aes(x=Var1, y=Freq)) + geom_col() +
  ylab("Proportion") + xlab("Play Result") +
  ggtitle("Proportion of Plays Resulting in a First Down")
```

Proportion of Plays Resulting in a First Down



#posteam

```
pander(summary(nfl$posteam))
```

Table 2: Table continues below

	ARI	ATL	BAL	BUF	CAR	CHI	CIN	CLE	DAL
21979	10585	10748	10885	10465	10548	10327	10718	10416	10489

Table 3: Table continues below

DEN	DET	GB	HOU	IND	JAC	JAX	KC	LA	MIA
11019	10982	10771	10944	10854	9253	1299	10509	1277	10398

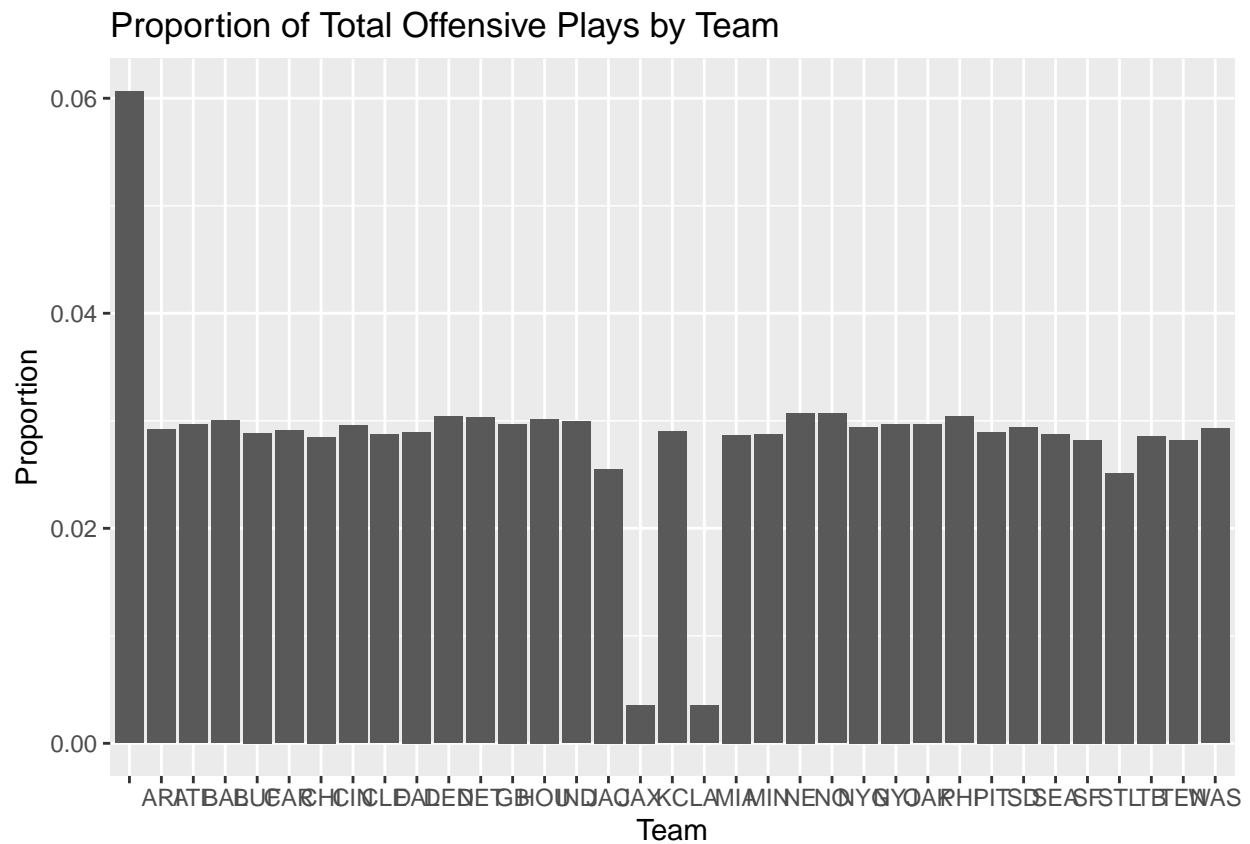
Table 4: Table continues below

MIN	NE	NO	NYG	NYJ	OAK	PHI	PIT	SD	SEA
10411	11131	11120	10666	10769	10748	11014	10499	10661	10426

SF	STL	TB	TEN	WAS
10225	9113	10357	10222	10619

```
pos.props <- data.frame(prop.table(table(nfl$posteam)))

ggplot(pos.props, aes(x=Var1, y=Freq)) + geom_col() +
  ylab("Proportion") + xlab("Team") +
  ggtitle("Proportion of Total Offensive Plays by Team")
```



```
#DefensiveTeam
pander(summary(nfl$DefensiveTeam))
```

Table 6: Table continues below

ARI	ATL	BAL	BUF	CAR	CHI	CIN	CLE	DAL	DEN
10894	10321	10671	10886	10404	10490	10545	10911	10475	10965

Table 7: Table continues below

DET	GB	HOU	IND	JAC	JAX	KC	LA	MIA	MIN
10609	10465	10513	10700	9401	1278	10661	1346	10813	10652

Table 8: Table continues below

NE	NO	NYG	NYJ	OAK	PHI	PIT	SD	SEA	SF
10759	10568	10740	10426	10879	11030	10264	10264	10459	10722

STL	TB	TEN	WAS	NA's
9290	10594	10947	10526	21979

```
defteam.props <- data.frame(prop.table(table(nfl$DefensiveTeam)))

ggplot(pos.props, aes(x=Var1, y=Freq)) + geom_col() +
  ylab("Proportion") + xlab("Team") +
  ggtitle("Proportion of Total Defensive Plays by Team")
```

