NFL 2002-2021 Margin of Victory

Kenizzer

Data and Summary This data set was generated by Reddit user **gigantoir** https://www.reddit.com /r/NFLstatheads/comments/q73yd0/nfl_scores_20172020/ I added the 2021 data that was scrapped from https://www.footballdb.com/games/index.html and 2002-2016 data from Reddit user **yuxbni76** https://www.reddit.com/user/yuxbni76

```
Scores <- read.csv("NFL_SCORES_2002-2021.txt", header=TRUE, sep= "\t")
Scores$Score_differential <- abs(Scores$Score_away - Scores$Score_home)
Scores$Week <- as.factor(Scores$Week)
summary(Scores)
```

```
##      Year          Week          Home               Away
##  Min.   :2002   16     : 336   Length:5136        Length:5136
##  1st Qu.:2007   2      : 323   Class :character   Class :character
##  Median :2012   15     : 320   Mode  :character   Mode  :character
##  Mean   :2012   17     : 320
##  3rd Qu.:2017   14     : 318
##  Max.   :2021   13     : 316
##                 (Other):3203
##    Score_home      Score_away     Score_differential
##  Min.   : 0.00   Min.   : 0.00   Min.   : 0.00
##  1st Qu.:16.00   1st Qu.:14.00   1st Qu.: 4.00
##  Median :23.00   Median :21.00   Median : 8.00
##  Mean   :23.19   Mean   :21.28   Mean   :11.71
##  3rd Qu.:30.00   3rd Qu.:28.00   3rd Qu.:17.00
##  Max.   :62.00   Max.   :59.00   Max.   :59.00
##
```

Team colors

Team colors were extracted from https://teamcolorcodes.com, I took the first primary color for each team and created a list that will be for later use. For the Browns and Titans I took the secondary color as it seemed more *appropriate*.

```
Team_colors <- c("SF"="#AA0000",
                 "CHI"="#0B162A",
                 "CIN"="#FB4F14",
                 "BUF"="#00338D",
                 "DEN"="#FB4F14",
                 "CLE"="#FF3C00",
                 "TB"="#D50A0A",
                 "ARI"="#97233F",
                 "LAC"="#0080C6",
                 "KC"="#E31837",
                 "IND"="#002C5F",
                 "DAL"="#041E42",
```

```
                "MIA"="#008E97",
                "PHI"="#004C54",
                "ATL"="#A71930",
                "NYG"="#0B2265",
                "JAX"="#006778",
                "NYJ"="#125740",
                "DET"="#0076B6",
                "GB"="#203731",
                "CAR"="#0085CA",
                "NE"="#002244",
                "LV"="#000000",
                "LA"="#003594",
                "BAL"="#241773",
                "WAS"="#773141",
                "NO"="#D3BC8D",
                "SEA"="#002244",
                "PIT"="#FFB612",
                "HOU"="#03202F",
                "TEN"="#4B92DB",
                "MIN"="#4F2683")
```

All teams, trend line

```
ggplot(Scores, aes(x=Week, y=Score_differential)) +
  geom_boxplot() +
  scale_x_discrete(breaks = seq(1,17,1), guide = guide_axis(n.dodge=2)) +
  facet_wrap(~Year) +
  ylab("Absolute Score Differential")
```
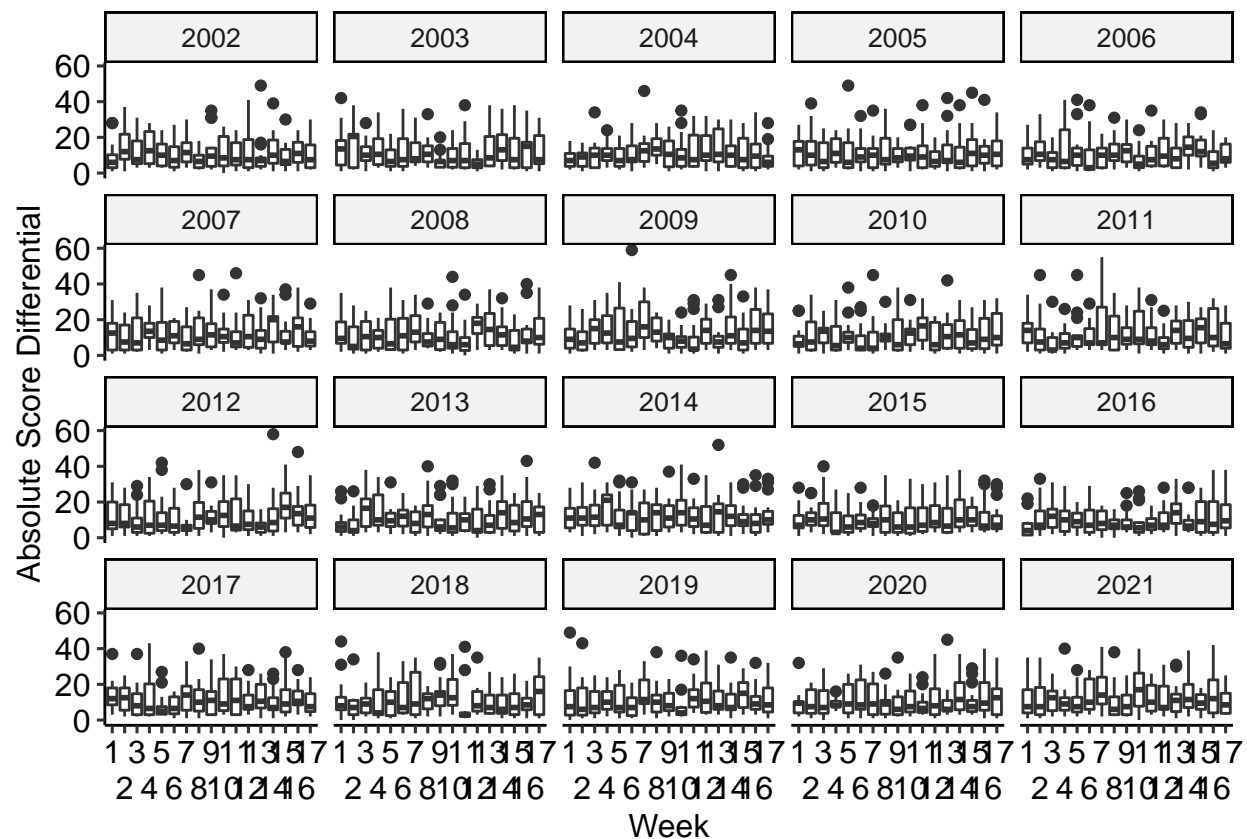
```
x <- ggplot(Scores, aes(x=Week, y=Score_differential)) +
  geom_smooth(aes(as.numeric(Week), Score_differential), method = "loess") +
  scale_x_continuous(breaks = c(1,9,17)) +
  scale_y_continuous(breaks = seq(7,21,7)) +
  facet_wrap(~Year) +
  ylab("Absolute Score Differential")
x

## `geom_smooth()` using formula 'y ~ x'
```
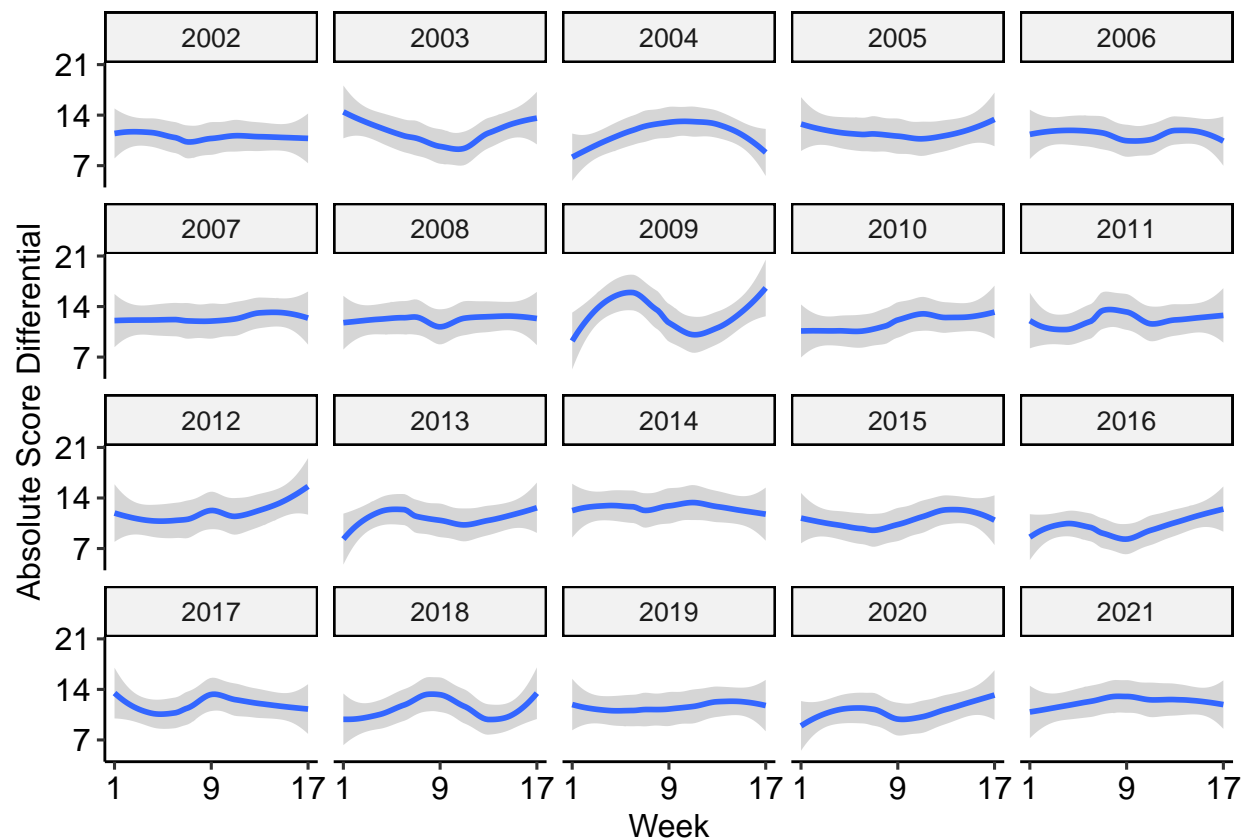
```
## `geom_smooth()` using formula 'y ~ x'
```

Variance plot

```r
Score_variance <- c(var(Scores[Scores$Year == 2002, ]$Score_differential),
                    var(Scores[Scores$Year == 2003, ]$Score_differential),
                    var(Scores[Scores$Year == 2004, ]$Score_differential),
                    var(Scores[Scores$Year == 2005, ]$Score_differential),
                    var(Scores[Scores$Year == 2006, ]$Score_differential),
                    var(Scores[Scores$Year == 2007, ]$Score_differential),
                    var(Scores[Scores$Year == 2008, ]$Score_differential),
                    var(Scores[Scores$Year == 2009, ]$Score_differential),
                    var(Scores[Scores$Year == 2010, ]$Score_differential),
                    var(Scores[Scores$Year == 2011, ]$Score_differential),
                    var(Scores[Scores$Year == 2012, ]$Score_differential),
                    var(Scores[Scores$Year == 2013, ]$Score_differential),
                    var(Scores[Scores$Year == 2014, ]$Score_differential),
                    var(Scores[Scores$Year == 2015, ]$Score_differential),
                    var(Scores[Scores$Year == 2016, ]$Score_differential),
                    var(Scores[Scores$Year == 2017, ]$Score_differential),
                    var(Scores[Scores$Year == 2018, ]$Score_differential),
                    var(Scores[Scores$Year == 2019, ]$Score_differential),
                    var(Scores[Scores$Year == 2020, ]$Score_differential),
                    var(Scores[Scores$Year == 2021, ]$Score_differential))
Score_variance <- as.data.frame(Score_variance)
```

```
Score_variance$Year <- c(2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,
                         2012,2013,2014,2015,2016,2017,2018,2019,2020,2021)
Score_variance$Year <- as.factor(Score_variance$Year)


ggplot(Score_variance, aes(x=Year, y=Score_variance)) +
  geom_point(size = 3) +
  geom_line(aes(x=as.numeric(Year))) +
  theme(axis.text.x = element_text(angle = 45, hjust=1)) +
  ylab("Margin of Victory variance")
```
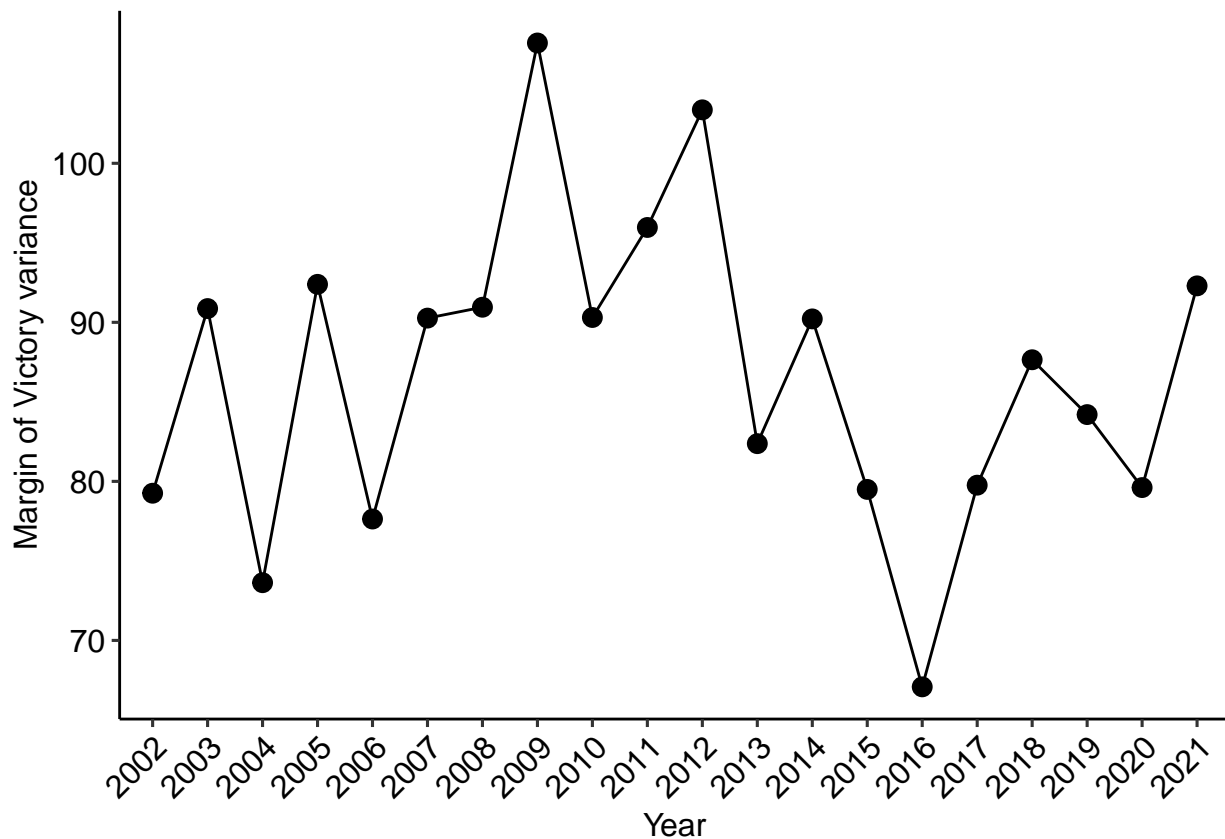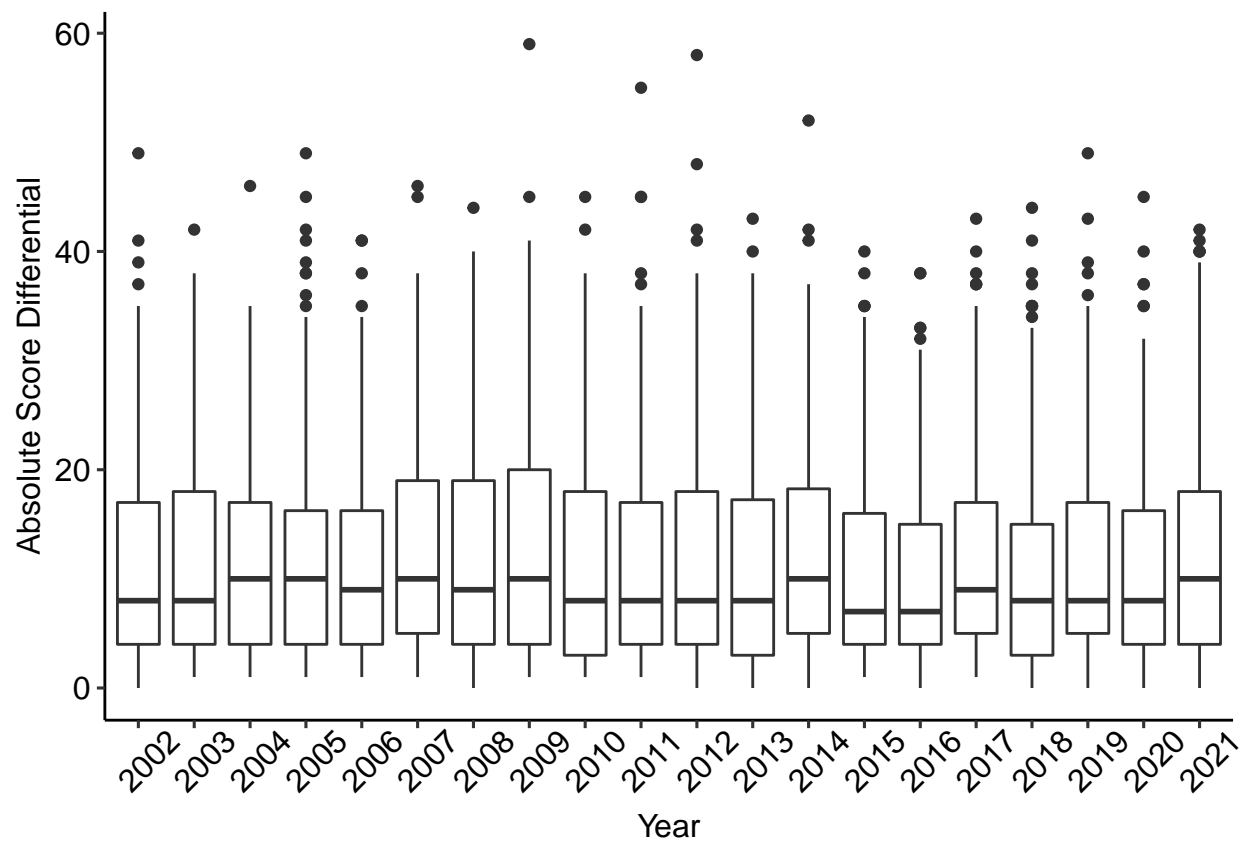

Average Margin of Victory by season, boxplot

```
Scores$Year <- as.factor(Scores$Year)

ggplot(Scores, aes(x=Year, y=Score_differential)) +
  geom_boxplot() +
  theme(axis.text.x = element_text(angle = 45, vjust = 0.5, hjust=0.3)) +
  ylab("Absolute Score Differential")
```

```
with(Scores,tapply(Score_differential, Year, mean))
```

```
##     2002     2003     2004     2005     2006     2007     2008     2009
## 11.10547 11.89062 11.36719 11.68750 11.42578 12.46875 12.22266 12.97266
##     2010     2011     2012     2013     2014     2015     2016     2017
## 11.75391 12.05469 12.15234 11.29297 12.66797 11.06250 10.23047 11.81250
##     2018     2019     2020     2021
## 11.09375 11.64062 11.07031 12.17647
```