NFL_Passing_Stats_Over_the_Years

Oscar Monroy

1/3/2022

```
# install.packages("ggpubr")
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 3.6.3
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
library(ggplot2)
tg <- read.csv("AvgPerTeamGame.csv")
ts <- read.csv("AvgPerTeamSeason.csv")
total <- read.csv("TotalPerSeason.csv")

sum(duplicated(tg))
```

```
## [1] 0
```

```
sum(duplicated(ts))
```

```
## [1] 0
```

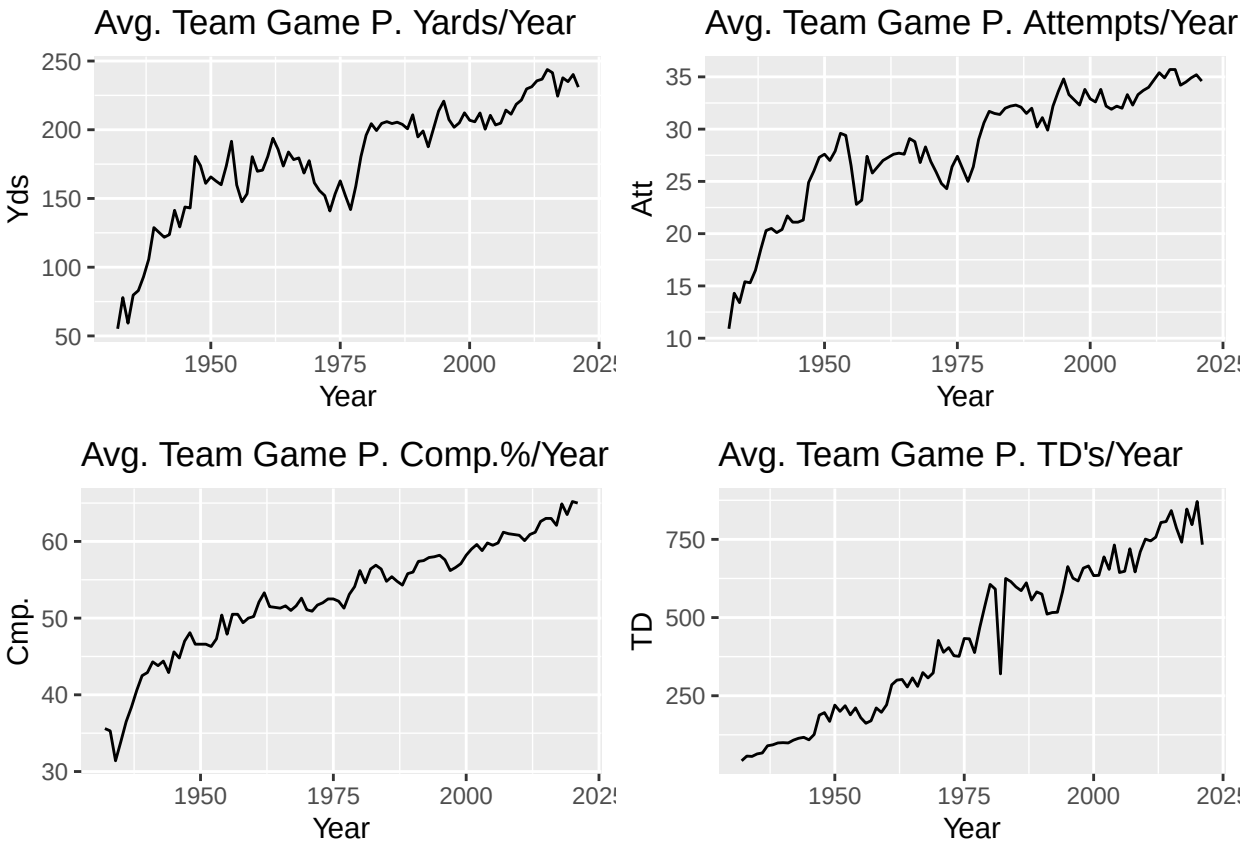
```
sum(duplicated(total))
```

```
## [1] 0
```

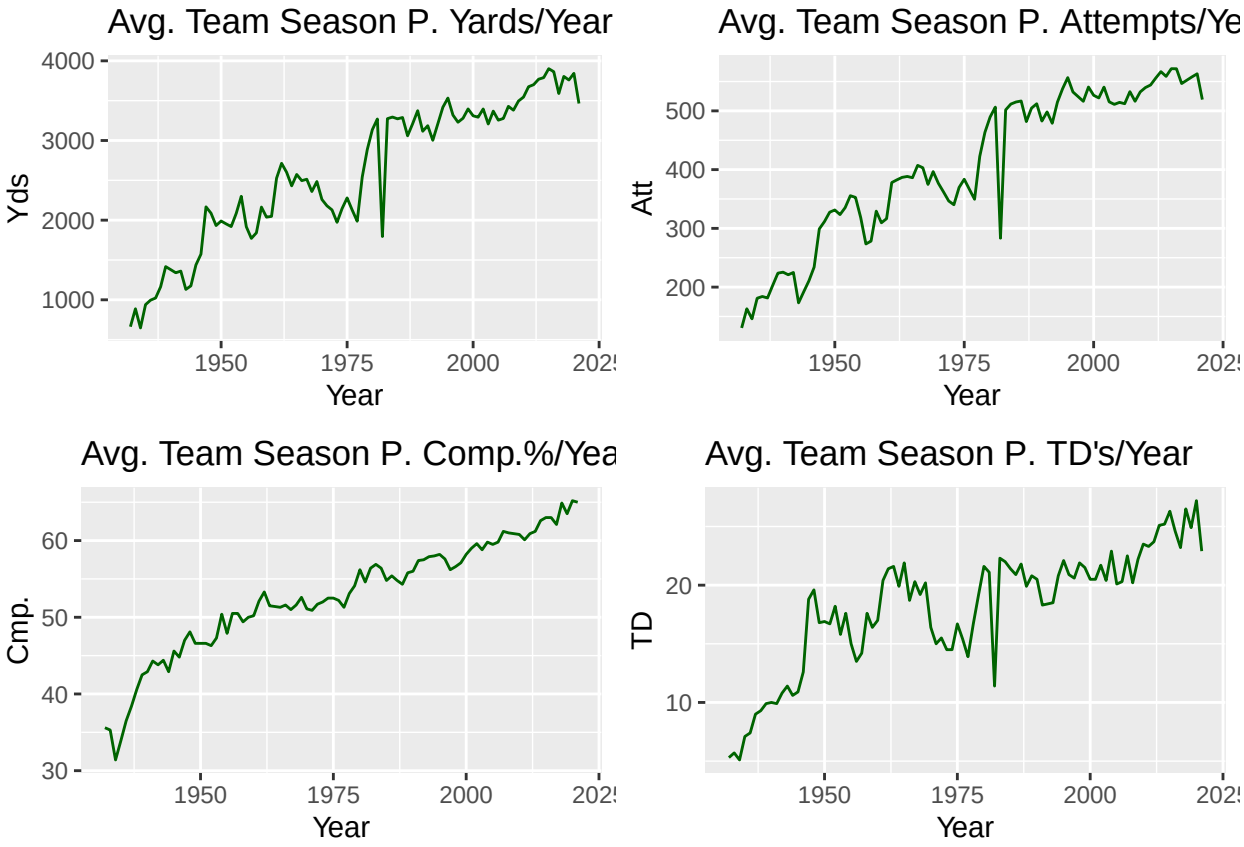
```
# summary(tg)
# summary(ts)
# summary(total)
```

```
g1 <- ggplot(tg, aes(Year, Yds)) +
  geom_line(color = "black") +
  ggtitle("Avg. Team Game P. Yards/Year")
g2 <- ggplot(tg, aes(Year, Att)) +
  geom_line(color = "black") +
  ggtitle("Avg. Team Game P. Attempts/Year")
g3 <- ggplot(tg, aes(Year, Cmp.)) +
```

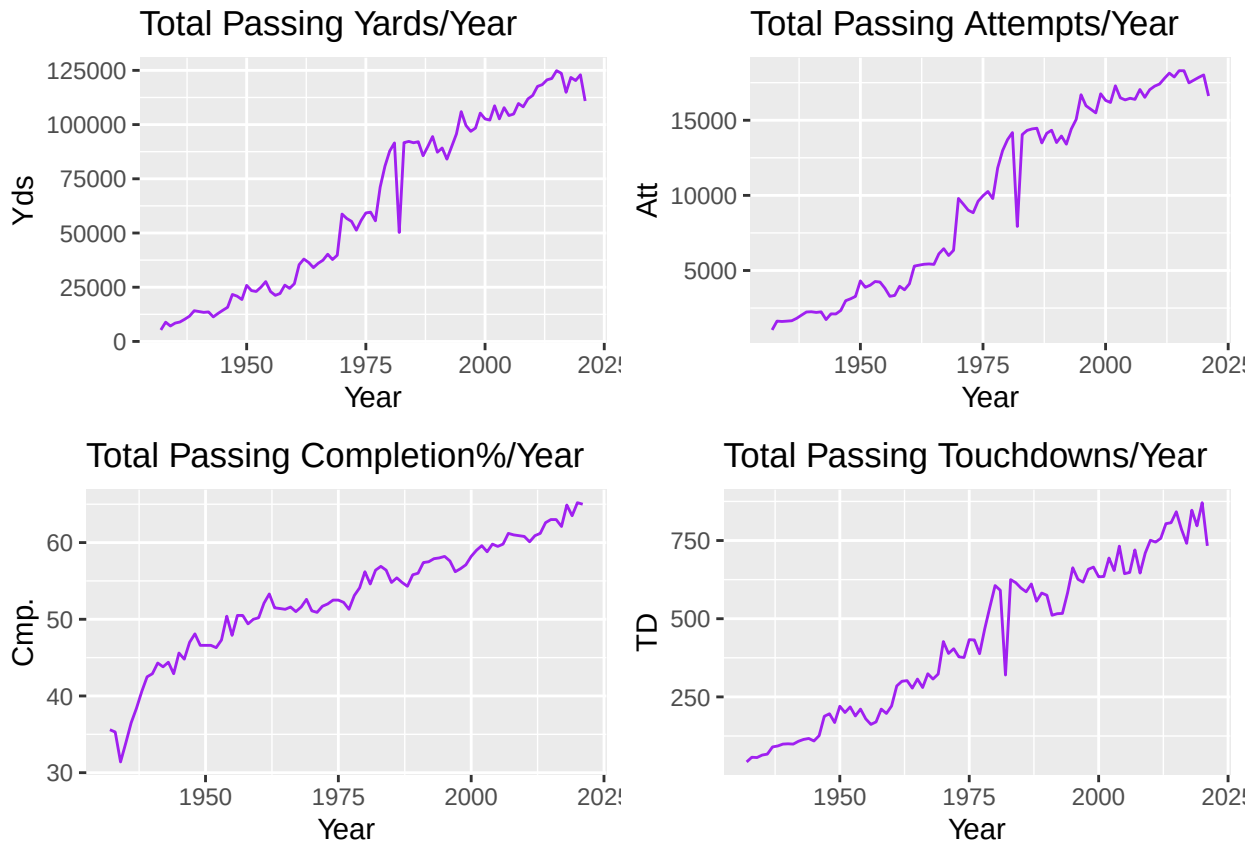
```
  geom_line(color = "black") +
  ggtitle("Avg. Team Game P. Comp.%/Year")
g4 <- ggplot(total, aes(Year, TD)) +
  geom_line(color = "black") +
  ggtitle("Avg. Team Game P. TD's/Year")
ggarrange(g1, g2, g3, g4,
                    ncol = 2, nrow = 2)
```




```
s1 <- ggplot(ts, aes(Year, Yds)) +
  geom_line(color = "darkgreen") +
  ggtitle("Avg. Team Season P. Yards/Year")
s2 <- ggplot(ts, aes(Year, Att)) +
  geom_line(color = "darkgreen") +
  ggtitle("Avg. Team Season P. Attempts/Year")
s3 <- ggplot(ts, aes(Year, Cmp.)) +
  geom_line(color = "darkgreen") +
  ggtitle("Avg. Team Season P. Comp.%/Year")
s4 <- ggplot(ts, aes(Year, TD)) +
  geom_line(color = "darkgreen") +
  ggtitle("Avg. Team Season P. TD's/Year")
ggarrange(s1, s2, s3, s4,
                    ncol = 2, nrow = 2)
```



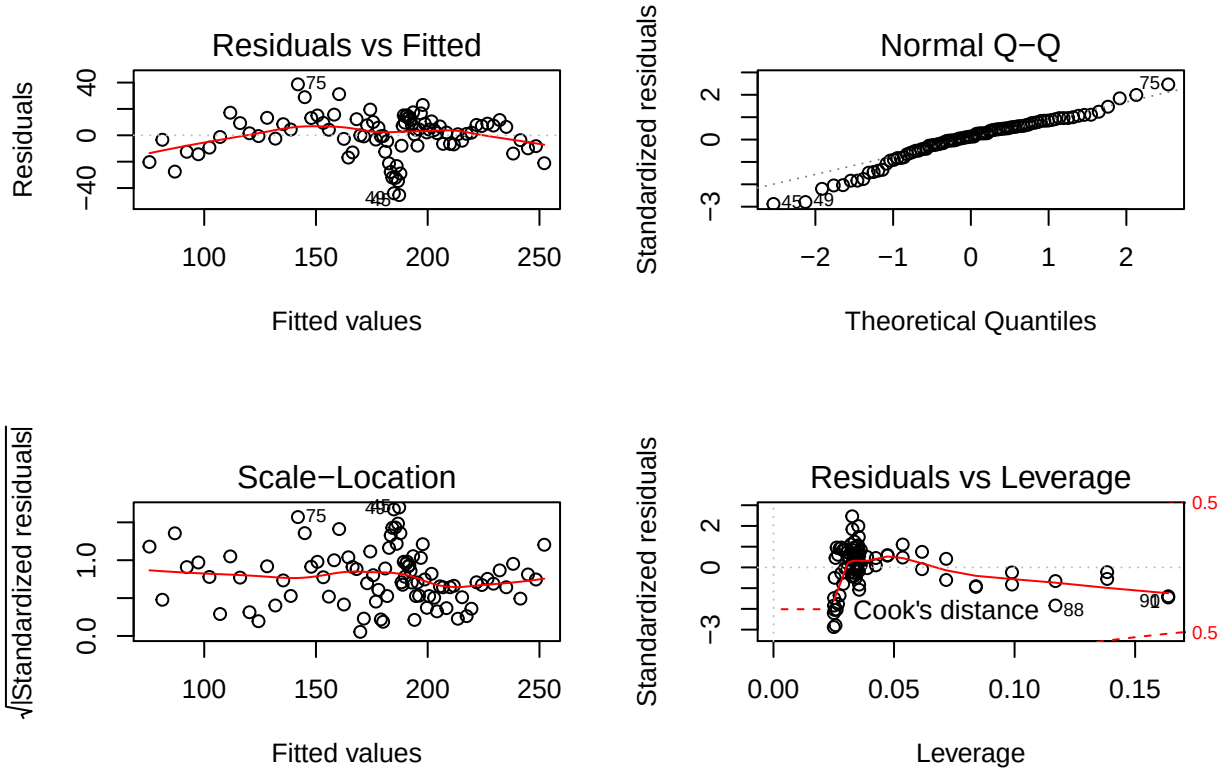

```
t1 <- ggplot(total, aes(Year, Yds)) +
  geom_line(color = "purple") +
  ggtitle("Total Passing Yards/Year")
t2 <- ggplot(total, aes(Year, Att)) +
  geom_line(color = "purple") +
  ggtitle("Total Passing Attempts/Year")
t3 <- ggplot(total, aes(Year, Cmp.)) +
  geom_line(color = "purple") +
  ggtitle("Total Passing Completion%/Year")
t4 <- ggplot(total, aes(Year, TD)) +
  geom_line(color = "purple") +
  ggtitle("Total Passing Touchdowns/Year")
ggarrange(t1, t2, t3, t4,
                    ncol = 2, nrow = 2)
```



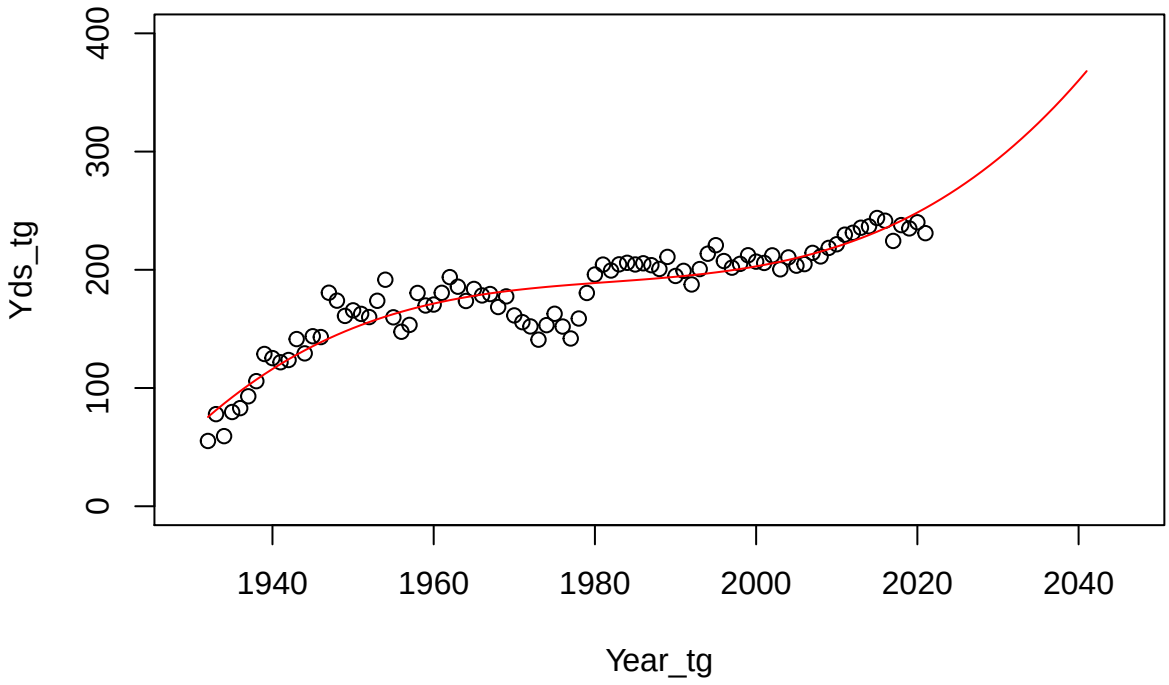

```
Yds_tg <- tg$Yds
Year_tg <- tg$Year
g_cbm <- lm(Yds_tg ~ poly(Year_tg, 3))
summary(g_cbm)
```

```
##
## Call:
## lm(formula = Yds_tg ~ poly(Year_tg, 3))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -45.353  -7.598   2.071   9.563  38.671
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        179.090      1.682 106.467  < 2e-16 ***
## poly(Year_tg, 3)1  352.768     15.958  22.106  < 2e-16 ***
## poly(Year_tg, 3)2  -66.931     15.958  -4.194 6.63e-05 ***
## poly(Year_tg, 3)3   94.426     15.958   5.917 6.50e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.96 on 86 degrees of freedom
## Multiple R-squared:  0.8629, Adjusted R-squared:  0.8581
## F-statistic: 180.4 on 3 and 86 DF,  p-value: < 2.2e-16
```

```r
par(mfrow=c(2,2))
plot(g_cbm) # Diagnositic plots shows that the model has no problems
```




```r
years_x1 <- seq(1932, 2041)
predictions_tg <- predict(g_cbm, newdata = data.frame(Year_tg = years_x1))
par(mfrow=c(1,1))
plot(Yds_tg ~ Year_tg, type = "p", ylim = c(0, 400), xlim = c(1930, 2046),
     main = "Avg. Team Season Passing Yards/Year Projections (Until 2041)")
lines(years_x1, predictions_tg, col = "red")
```



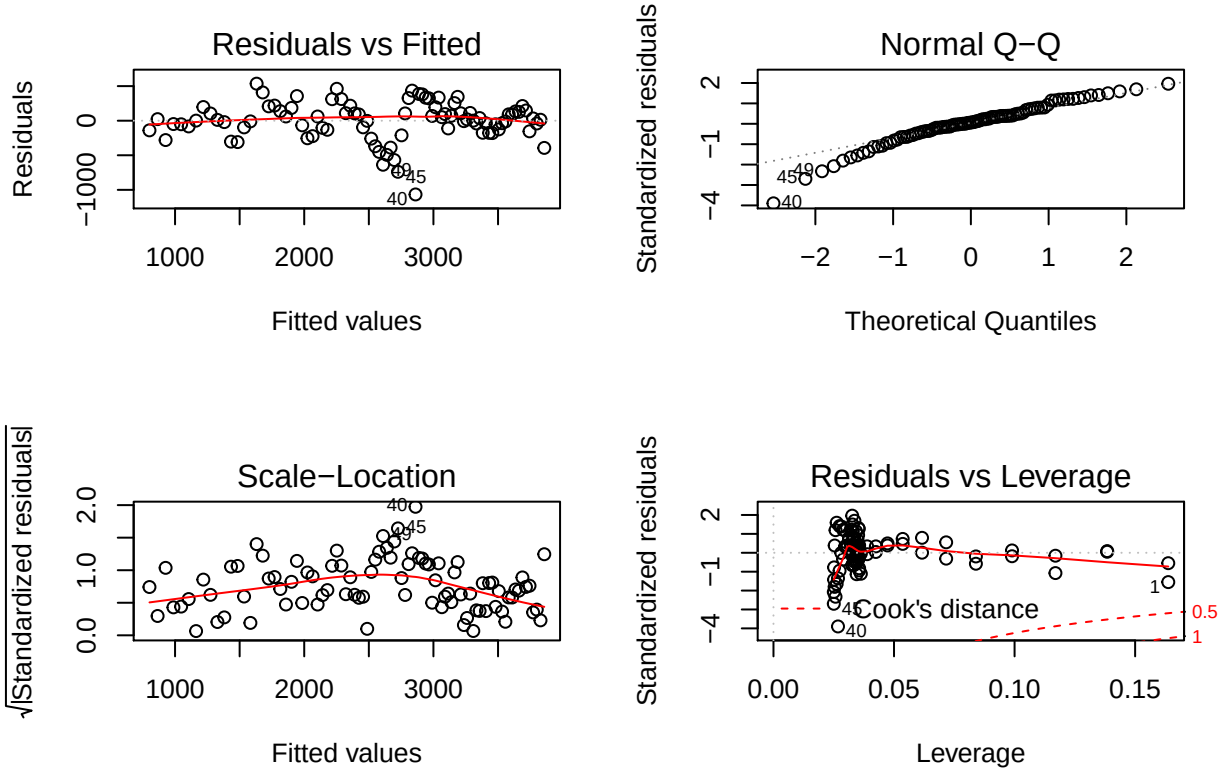

```
p_20yrs1 <- data.frame(Year = tail(years_x1, 20), Yds = tail(predictions_tg, 20))
p_20yrs1
```

```
##     Year      Yds
## 91  2022 256.0555
## 92  2023 260.1235
## 93  2024 264.3699
## 94  2025 268.7990
## 95  2026 273.4151
## 96  2027 278.2226
## 97  2028 283.2258
## 98  2029 288.4291
## 99  2030 293.8368
## 100 2031 299.4532
## 101 2032 305.2827
## 102 2033 311.3296
## 103 2034 317.5983
## 104 2035 324.0931
## 105 2036 330.8182
## 106 2037 337.7782
## 107 2038 344.9772
## 108 2039 352.4197
## 109 2040 360.1100
## 110 2041 368.0524
```

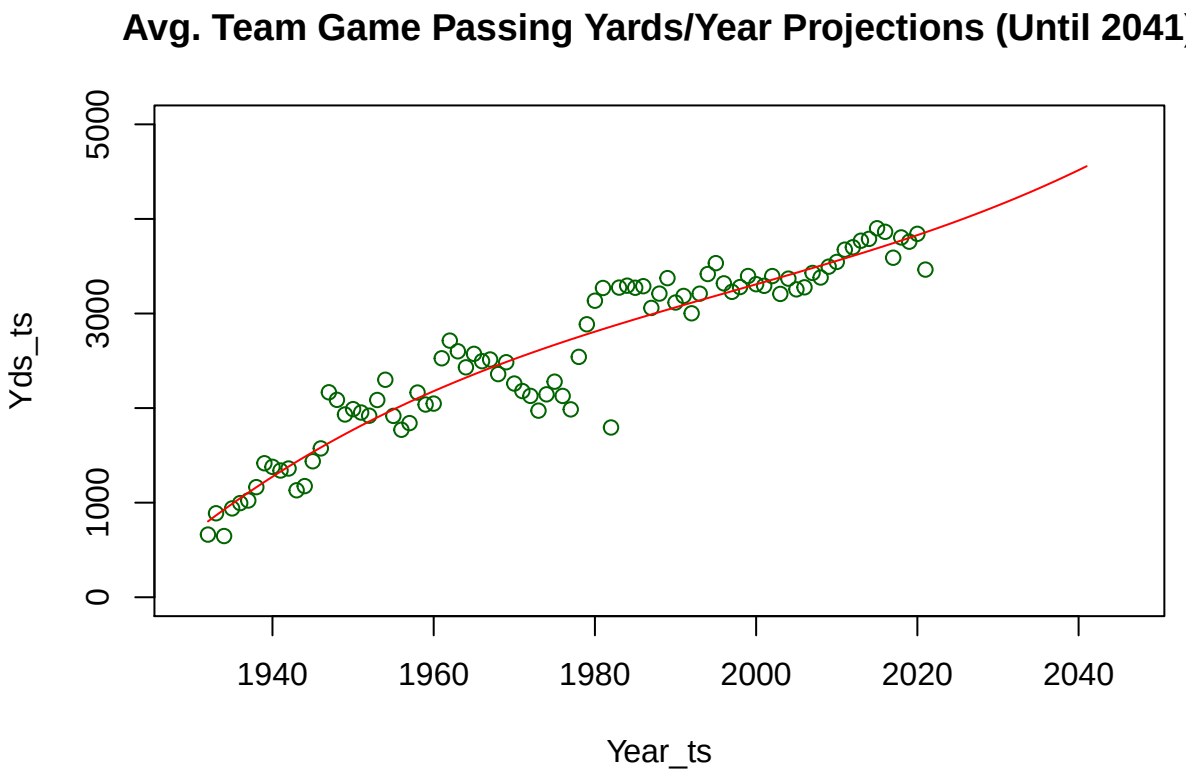
```r
# ------------------------------------------------------------------------------

Yds_ts <- ts$Yds
Year_ts <- ts$Year
s_cbm <- lm(Yds_ts ~ poly(Year_ts, 3))
summary(s_cbm)
```

```
##
## Call:
## lm(formula = Yds_ts ~ poly(Year_ts, 3))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1066.52  -122.25    16.45   145.17   535.41
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        2581.40      29.24  88.279  < 2e-16 ***
## poly(Year_ts, 3)1  7882.06     277.41  28.413  < 2e-16 ***
## poly(Year_ts, 3)2 -1102.72     277.41  -3.975 0.000146 ***
## poly(Year_ts, 3)3   401.64     277.41   1.448 0.151298
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 277.4 on 86 degrees of freedom
## Multiple R-squared:  0.9056, Adjusted R-squared:  0.9023
## F-statistic: 275.1 on 3 and 86 DF,  p-value: < 2.2e-16
```

```r
par(mfrow=c(2,2))
plot(s_cbm) # Diagnositic plots shows that the model has no problems
```




```
predictions_ts <- predict(s_cbm, newdata = data.frame(Year_ts = years_x1))
par(mfrow=c(1,1))
plot(Yds_ts ~ Year_ts, type = "p", ylim = c(0, 5000), xlim = c(1930, 2046),
     main = "Avg. Team Game Passing Yards/Year Projections (Until 2041)",
     col = "darkgreen")
lines(years_x1, predictions_ts, col = "red")
```



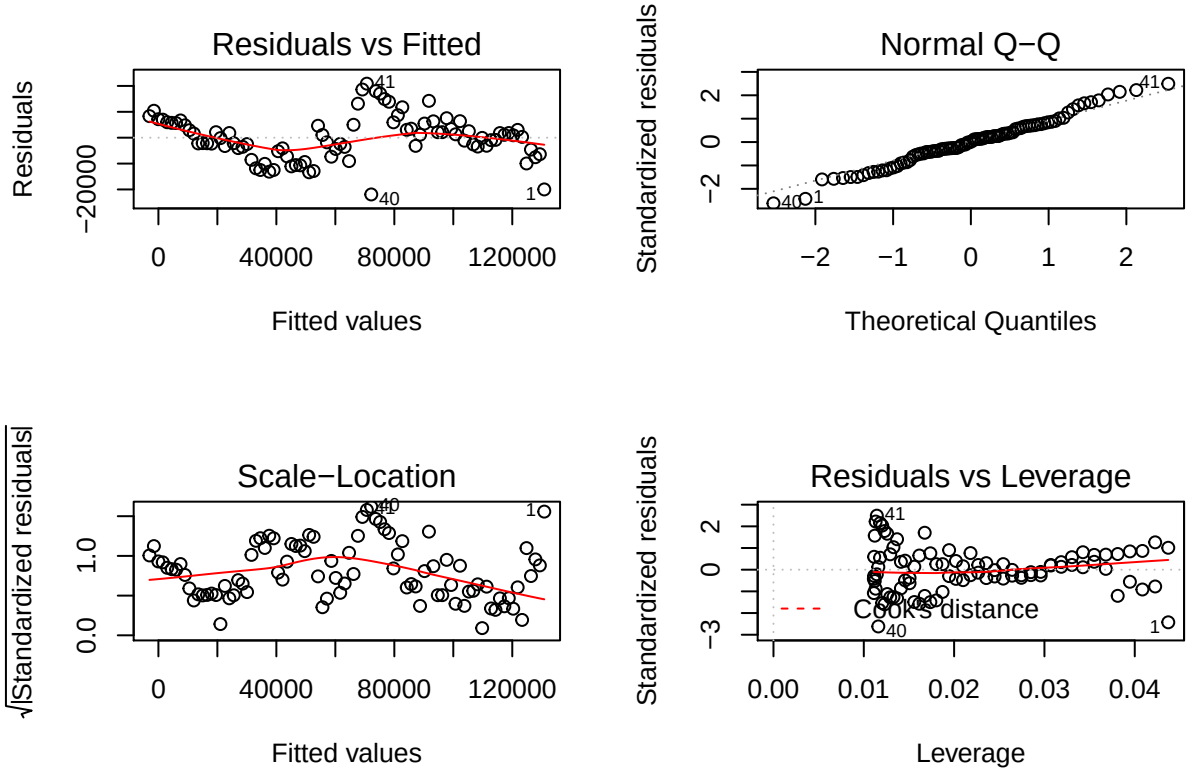

```
p_20yrs2 <- data.frame(Year = tail(years_x1, 20), Yds = tail(predictions_ts, 20))
p_20yrs2
```

```
##     Year      Yds
## 91  2022 3887.553
## 92  2023 3917.605
## 93  2024 3948.131
## 94  2025 3979.148
## 95  2026 4010.674
## 96  2027 4042.729
## 97  2028 4075.331
## 98  2029 4108.498
## 99  2030 4142.248
## 100 2031 4176.601
## 101 2032 4211.575
## 102 2033 4247.187
## 103 2034 4283.457
## 104 2035 4320.402
## 105 2036 4358.043
## 106 2037 4396.396
## 107 2038 4435.480
## 108 2039 4475.314
## 109 2040 4515.916
## 110 2041 4557.305
```

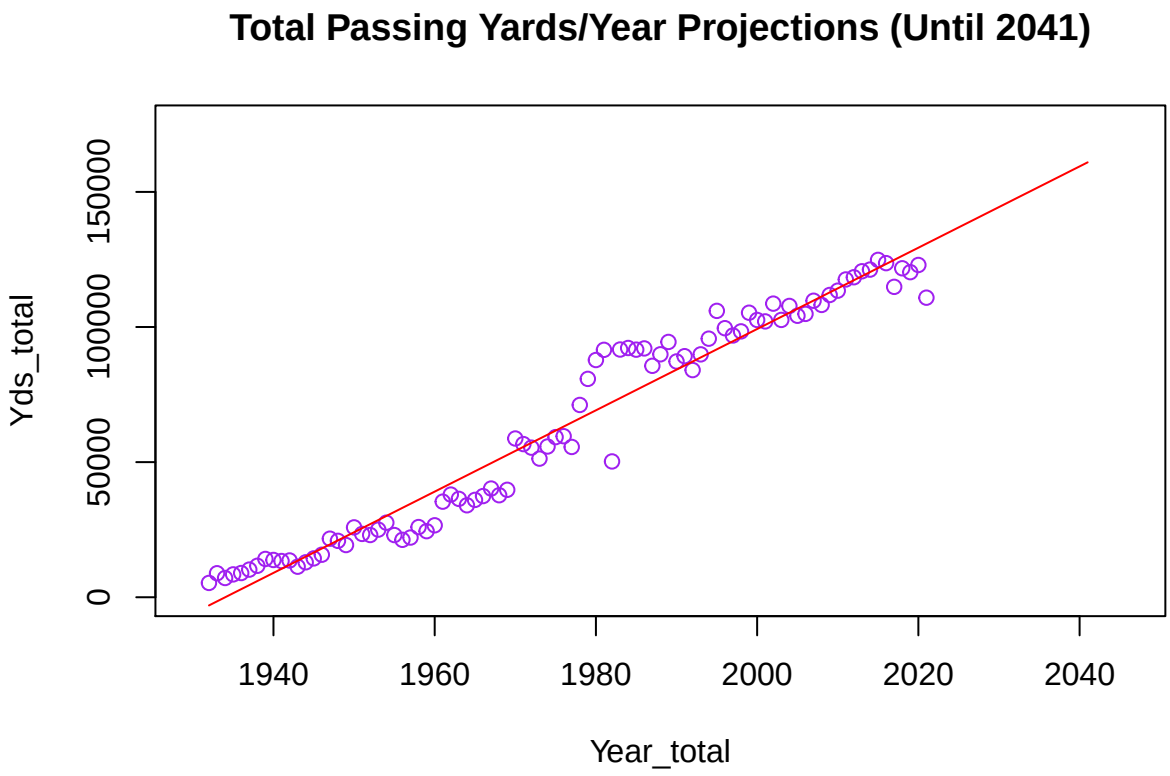
```r
# ------------------------------------------------------------------------------

# Since the plot using Yds from "total" shows a linear pattern compared to the
# other data sets, we'll use linear regression instead of
# cubic regression like the previous two prediction models.
Yds_total <- total$Yds
Year_total <- total$Year
t_lm <- lm(Yds_total ~ Year_total)
summary(t_lm)
```

```
##
## Call:
## lm(formula = Yds_total ~ Year_total)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21927.3  -4303.7    120.9   5342.5  20874.0
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.909e+06  6.756e+04  -43.06   <2e-16 ***
## Year_total   1.504e+03  3.418e+01   44.01   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8424 on 88 degrees of freedom
## Multiple R-squared:  0.9565, Adjusted R-squared:  0.956
## F-statistic:  1937 on 1 and 88 DF,  p-value: < 2.2e-16
```

```r
par(mfrow=c(2,2))
plot(t_lm) # Diagnositic plots shows that the model has no problems
```




```r
predictions_total <- predict(t_lm, newdata = data.frame(Year_total = years_x1))
par(mfrow=c(1,1))
plot(Yds_total ~ Year_total, type = "p", ylim = c(0, 175000), xlim = c(1930, 2046),
     main = "Total Passing Yards/Year Projections (Until 2041)",
     col = "purple")
lines(years_x1, predictions_total, col = "red")
```




```
p_20yrs3 <- data.frame(Year = tail(years_x1, 20), Yds = tail(predictions_total, 20))
p_20yrs3
```

```
##     Year      Yds
## 91  2022 132354.5
## 92  2023 133858.8
## 93  2024 135363.1
## 94  2025 136867.4
## 95  2026 138371.7
## 96  2027 139875.9
## 97  2028 141380.2
## 98  2029 142884.5
## 99  2030 144388.8
## 100 2031 145893.1
## 101 2032 147397.3
## 102 2033 148901.6
## 103 2034 150405.9
## 104 2035 151910.2
## 105 2036 153414.5
## 106 2037 154918.8
## 107 2038 156423.0
## 108 2039 157927.3
## 109 2040 159431.6
## 110 2041 160935.9
```