Examining the State of NFL Kickoffs

Andre Adkins

August 1, 2024

```r
install.packages("tidyverse", repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
##  /var/folders/g5/27tb8m595rs918z5723khcth0000gn/T//RtmpYYCC2g/downloaded_packages
```

```r
install.packages("knitr", repos = "http://cran.us.r-project.org")
```

```
##
##   There is a binary version available but the source version is later:
##       binary source needs_compilation
## knitr   1.47   1.48             FALSE
```

```
## installing the source package 'knitr'
```

```r
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.4     ✔ readr     2.1.5
## ✔ forcats   1.0.0     ✔ stringr   1.5.1
## ✔ ggplot2   3.5.1     ✔ tibble    3.2.1
## ✔ lubridate 1.9.3     ✔ tidyr     1.3.1
## ✔ purrr     1.0.2
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(knitr)
```

Reading the Data into R

```r
NFLKickoffData <- read_csv("/Users/andreproctor/Documents/Data Analysis Projects/January - NFL Kickoffs/kickoff_data_by_year.csv")
```

```
## Rows: 19 Columns: 9
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## dbl (9): Year, Avg_Yds_Per_Return, Avg_KRet_Yds_Per_Season, Total_KRet_TD, T...
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Data Transformation

The code below creates 5 new columns and calculates percent change from year to year for the following categories; `Avg_Ret_Allowed`, `Total_Touchbacks`, `Total_20_Yd_Returns`, `Total_40_Yd_Returns`, and `Avg_KRet_Yds_Per_Season`.

```r
NFLKickoffData <- NFLKickoffData    %>% mutate("AvgReturnsAllowed%_to_LY" = 0, .after = Avg_Ret_Allowed) %>% mutate("TotalTouchbacks%_to_LY" = 0, .after = Total_Touchbacks) %>% mutate("20YdReturns%_to_LY" = 0, .after = Total_20_Yd_Returns) %>% mutate("40YdReturns%_to_LY" = 0, .after = Total_40_Yd_Returns) %>% mutate("KRetYds%_to_LY" = 0, .after = Avg_KRet_Yds_Per_Season)


for(i in 2:nrow(NFLKickoffData))
{
  NFLKickoffData$"KRetYds%_to_LY"[i] <- round(((NFLKickoffData$Avg_KRet_Yds_Per_Season[i] - NFLKickoffData$Avg_KRet_Yds_Per_Season[i-1])/NFLKickoffData$Avg_KRet_Yds_Per_Season[i-1]) * 100, digits = 2)
  NFLKickoffData$"TotalTouchbacks%_to_LY"[i] <- round(((NFLKickoffData$Total_Touchbacks[i] - NFLKickoffData$Total_Touchbacks[i-1])/NFLKickoffData$Total_Touchbacks[i-1]) * 100, digits = 2)
  NFLKickoffData$"20YdReturns%_to_LY"[i] <- round(((NFLKickoffData$Total_20_Yd_Returns[i] - NFLKickoffData$Total_20_Yd_Returns[i-1])/NFLKickoffData$Total_20_Yd_Returns[i-1]) * 100, digits = 2)
  NFLKickoffData$"40YdReturns%_to_LY"[i] <- round(((NFLKickoffData$Total_40_Yd_Returns[i] - NFLKickoffData$Total_40_Yd_Returns[i-1])/NFLKickoffData$Total_40_Yd_Returns[i-1]) * 100, digits = 2)
  NFLKickoffData$"AvgReturnsAllowed%_to_LY"[i] <- round(((NFLKickoffData$Avg_Ret_Allowed[i] - NFLKickoffData$Avg_Ret_Allowed[i-1])/NFLKickoffData$Avg_Ret_Allowed[i-1]) * 100, digits = 2)

}

NFLKickoffData
```

```
## # A tibble: 19 × 14
##     Year Avg_Yds_Per_Return Avg_KRet_Yds_Per_Season `KRetYds%_to_LY`
##    <dbl>              <dbl>                   <dbl>            <dbl>
##  1  2005               22.2                   1488.             0
##  2  2006               22.6                   1435.            -3.61
##  3  2007               22.6                   1467.             2.24
##  4  2008               22.8                   1508.             2.8
##  5  2009               22.6                   1417.            -6.03
##  6  2010               22.4                   1419.             0.17
##  7  2011               23.8                   1021.           -28.0
##  8  2012               23.5                   1028.             0.65
##  9  2013               23.2                    941.            -8.43
## 10  2014               23.6                    910.            -3.3
## 11  2015               23.4                    795.           -12.6
## 12  2016               21.7                    709            -10.8
## 13  2017               21.6                    700.            -1.29
## 14  2018               22.9                    695.            -0.73
## 15  2019               22.3                    659.            -5.08
## 16  2020               22.4                    699.             5.96
## 17  2021               21.9                    763.             9.23
## 18  2022               22.7                    722.            -5.4
## 19  2023               22.6                    423.           -41.4
## # ℹ 10 more variables: Total_KRet_TD <dbl>, Total_20_Yd_Returns <dbl>,
## #   `20YdReturns%_to_LY` <dbl>, Total_40_Yd_Returns <dbl>,
## #   `40YdReturns%_to_LY` <dbl>, Total_Fair_Catches <dbl>,
## #   Total_Touchbacks <dbl>, `TotalTouchbacks%_to_LY` <dbl>,
## #   Avg_Ret_Allowed <dbl>, `AvgReturnsAllowed%_to_LY` <dbl>
```

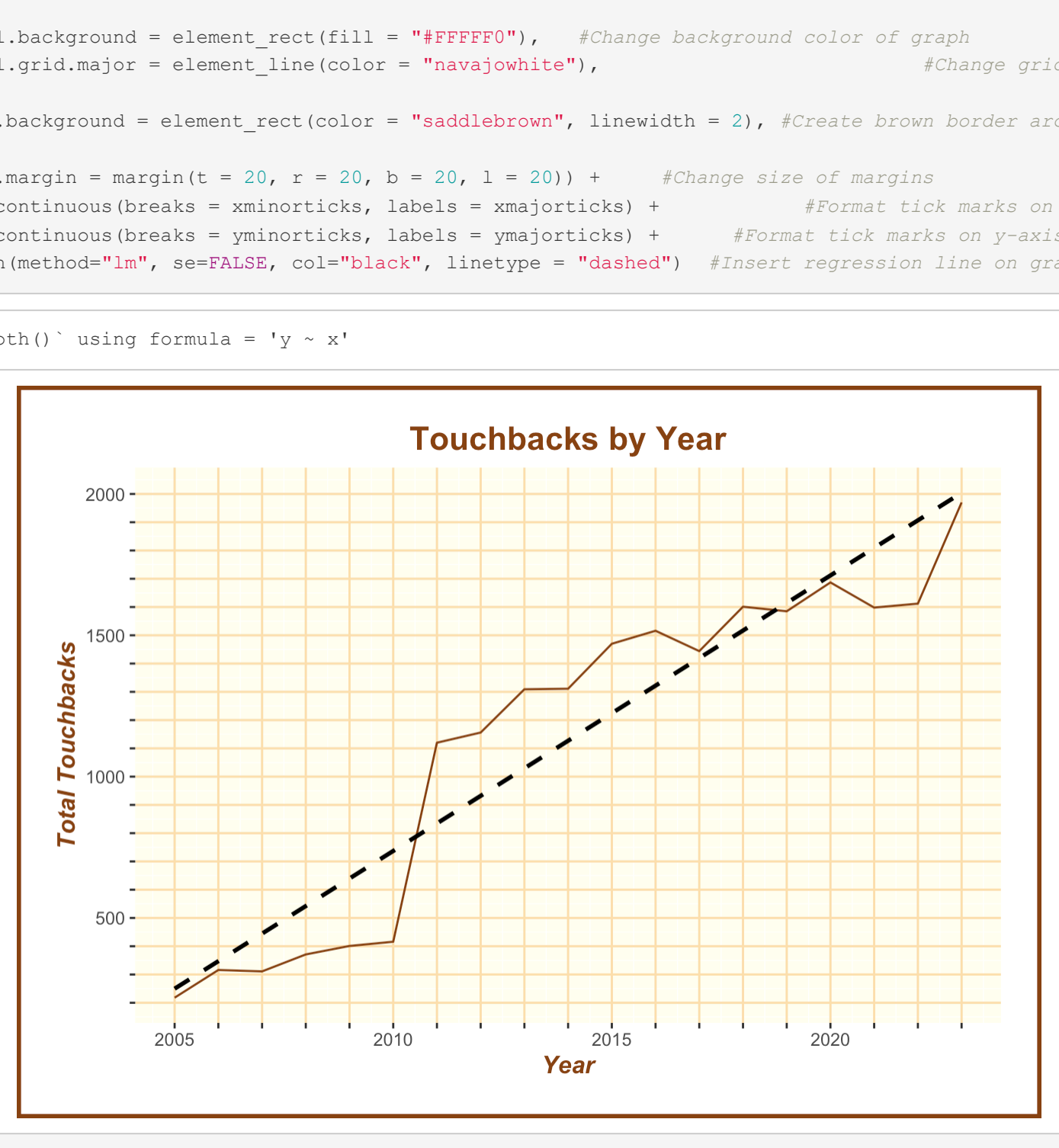
Data Visualizations

```r
xminorticks <- NFLKickoffData$Year  #Create list for minor tick marks on the x-axis
xmajorticks <- xminorticks           #Create list for major tick marks on the x-axis
xmajorticks[c(2005:2023) %% 5 != 0] <- " "  #Make all entries in xmajorticks not divisible by 5 blank

#Linear regression (x=Year, y=Total_Touchbacks)
year_touchbacks_lm <- lm(Total_Touchbacks ~ Year, data = NFLKickoffData)
summary(year_touchbacks_lm)
```

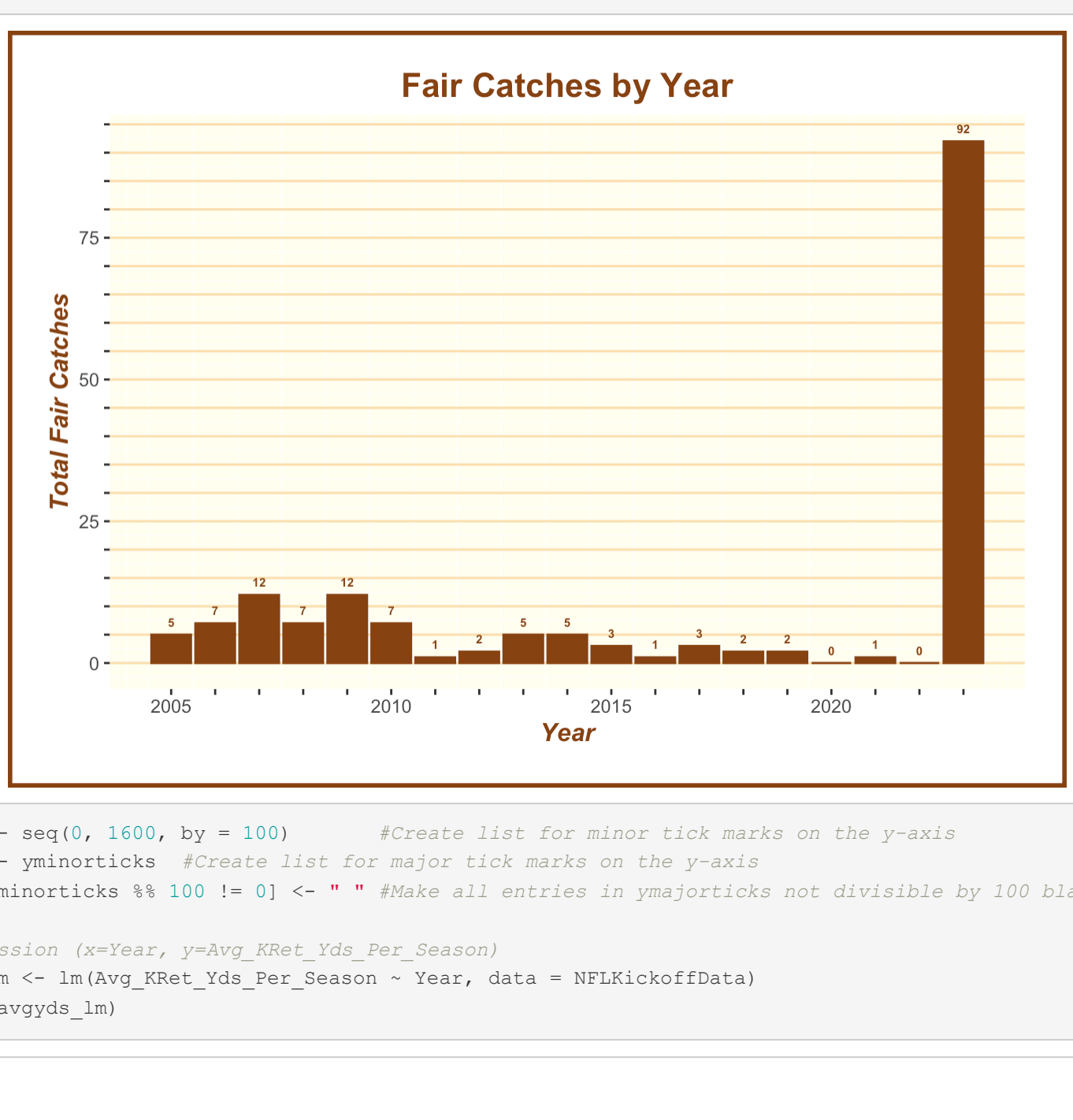
```
##
## Call:
## lm(formula = Total_Touchbacks ~ Year, data = NFLKickoffData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -321.09 -152.44  -29.26  189.09  285.44
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.952e+05  1.703e+04  -11.46 2.02e-09 ***
## Year         9.746e+01  8.454e+00   11.53 1.85e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 201.8 on 17 degrees of freedom
## Multiple R-squared:  0.8866, Adjusted R-squared:  0.8799
## F-statistic: 132.9 on 1 and 17 DF,  p-value: 1.854e-09
```

```r
yminorticks <- seq(0, 2000, by = 100)       #Create list for minor tick marks on the y-axis
ymajorticks <- yminorticks                    #Create list for major tick marks on the y-axis
ymajorticks[yminorticks %% 500 != 0] <- " " #Make all entries in ymajorticks not divisible by 500 blank

ggplot(NFLKickoffData, aes(Year, Total_Touchbacks)) +       #Create plot using Year and Total_Touchbacks fields
    geom_line(color = "saddlebrown") +                      #Plot line graph with brown line
    ggtitle("Touchbacks by Year") +                         #Create title
    labs(x = "Year", y = "Total Touchbacks") +              #Create x and y axis labels
    theme(axis.title = element_text(size = 12, color = "saddlebrown", face = "bold.italic"),   #Format x and y axis labels
        plot.title = element_text(size = 16, color = "saddlebrown", face = "bold", hjust=0.5),   #Format main title
        panel.background = element_rect(fill = "#FFFFF0"),    #Change background color of graph
        panel.grid.major = element_line(color = "navajowhite"),                  #Change grid color of graph
        plot.background = element_rect(color = "saddlebrown", linewidth = 2), #Create brown border around entire chart
        plot.margin = margin(t = 20, r = 20, b = 20, l = 20)) +      #Change size of margins
    scale_x_continuous(breaks = xminorticks, labels = xmajorticks) +           #Format tick marks on x-axis
    scale_y_continuous(breaks = yminorticks, labels = ymajorticks) +      #Format tick marks on y-axis
  geom_smooth(method="lm", se=FALSE, col="black", linetype = "dashed")  #Insert regression line on graph
```

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




```r
yminorticks <- seq(0, 100, by = 5)              #Reset list for minor tick marks on the y-axis
ymajorticks <- yminorticks                        #Reset list for major tick marks on the y-axis
ymajorticks[yminorticks %% 25 != 0] <- " "  #Make all entries in ymajorticks not divisible by 25 blank


#Create plot using Year and Total_Fair_Catches fields
ggplot(NFLKickoffData, aes(Year, Total_Fair_Catches)) +
    #Create bar graph with brown bars
  geom_bar(stat = "identity", color = "saddlebrown", fill = "saddlebrown") +
    #Insert data labels above each bar
  geom_text(aes(label = Total_Fair_Catches),size = 2, fontface = "bold", vjust = -1, colour = "saddlebrown")    +
    ggtitle("Fair Catches by Year") +
    labs(x = "Year", y = "Total Fair Catches") +
    theme(axis.title = element_text(size = 12, color = "saddlebrown", face = "bold.italic"),
        plot.title = element_text(size = 16, color = "saddlebrown", face = "bold", hjust=0.5),
        panel.background = element_rect(fill = "#FFFFF0"),
        panel.grid.major.y = element_line(color = "navajowhite"),
        panel.grid.major.x = element_blank(),
        plot.background = element_rect(color = "saddlebrown", linewidth = 2),
        plot.margin = margin(t = 20, r = 20, b = 20, l = 20)) +
    scale_x_continuous(breaks = xminorticks, labels = xmajorticks) +
    scale_y_continuous(breaks = yminorticks, labels = ymajorticks)
```

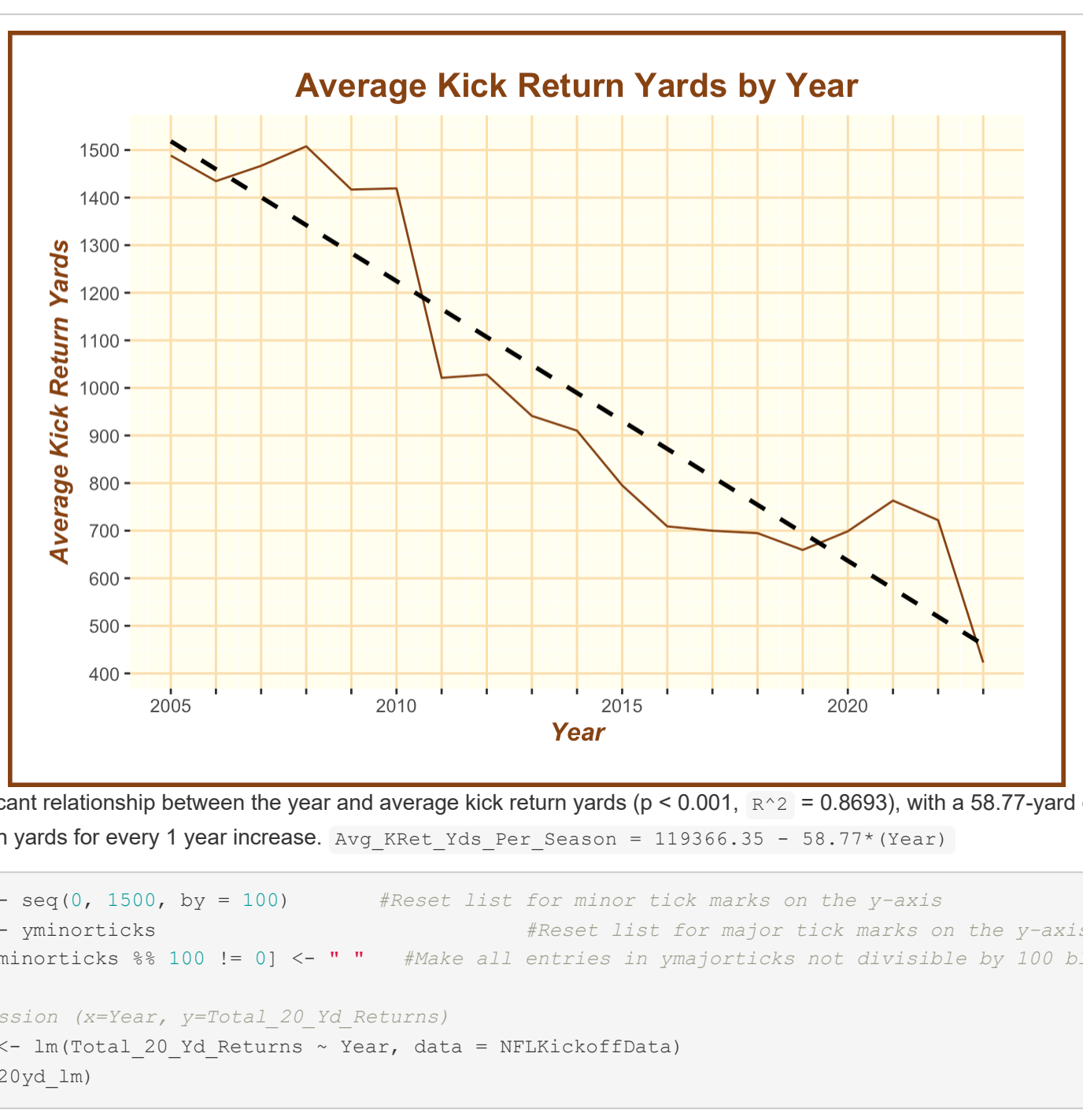



```r
yminorticks <- seq(0, 1600, by = 100)       #Create list for minor tick marks on the y-axis
ymajorticks <- yminorticks  #Create list for major tick marks on the y-axis
ymajorticks[yminorticks %% 100 != 0] <- " " #Make all entries in ymajorticks not divisible by 100 blank

#Linear regression (x=Year, y=Avg_KRet_Yds_Per_Season)
year_avgyds_lm <- lm(Avg_KRet_Yds_Per_Season ~ Year, data = NFLKickoffData)
summary(year_avgyds_lm)
```

```
##
## Call:
## lm(formula = Avg_KRet_Yds_Per_Season ~ Year, data = NFLKickoffData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -162.88  -93.17  -36.08   99.74  202.81
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 119366.352  10773.467   11.08 3.38e-09 ***
## Year           -58.777      5.349  -10.99 3.83e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 127.7 on 17 degrees of freedom
## Multiple R-squared:  0.8766, Adjusted R-squared:  0.8693
## F-statistic: 120.7 on 1 and 17 DF,  p-value: 3.826e-09
```

```r
#Create plot using Year and Avg_KRet_Yds_Per_Season fields
ggplot(NFLKickoffData, aes(Year, Avg_KRet_Yds_Per_Season)) +
  geom_line(color = "saddlebrown") +
  ggtitle("Average Kick Return Yards by Year") +
  labs(x = "Year", y = "Average Kick Return Yards") +
  theme(axis.title = element_text(size = 12, color = "saddlebrown", face = "bold.italic"),
        plot.title = element_text(size = 16, color = "saddlebrown", face = "bold", hjust=0.5),
        panel.background = element_rect(fill = "#FFFFF0"),
        panel.grid.major = element_line(color = "navajowhite"),
        plot.background = element_rect(color = "saddlebrown", linewidth = 2),
        plot.margin = margin(t = 20, r = 20, b = 20, l = 20)) +
  scale_x_continuous(breaks = xminorticks, labels = xmajorticks) +
  scale_y_continuous(breaks = yminorticks, labels = ymajorticks) +
  geom_smooth(method="lm", se=FALSE, col="black", linetype = "dashed")
```

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



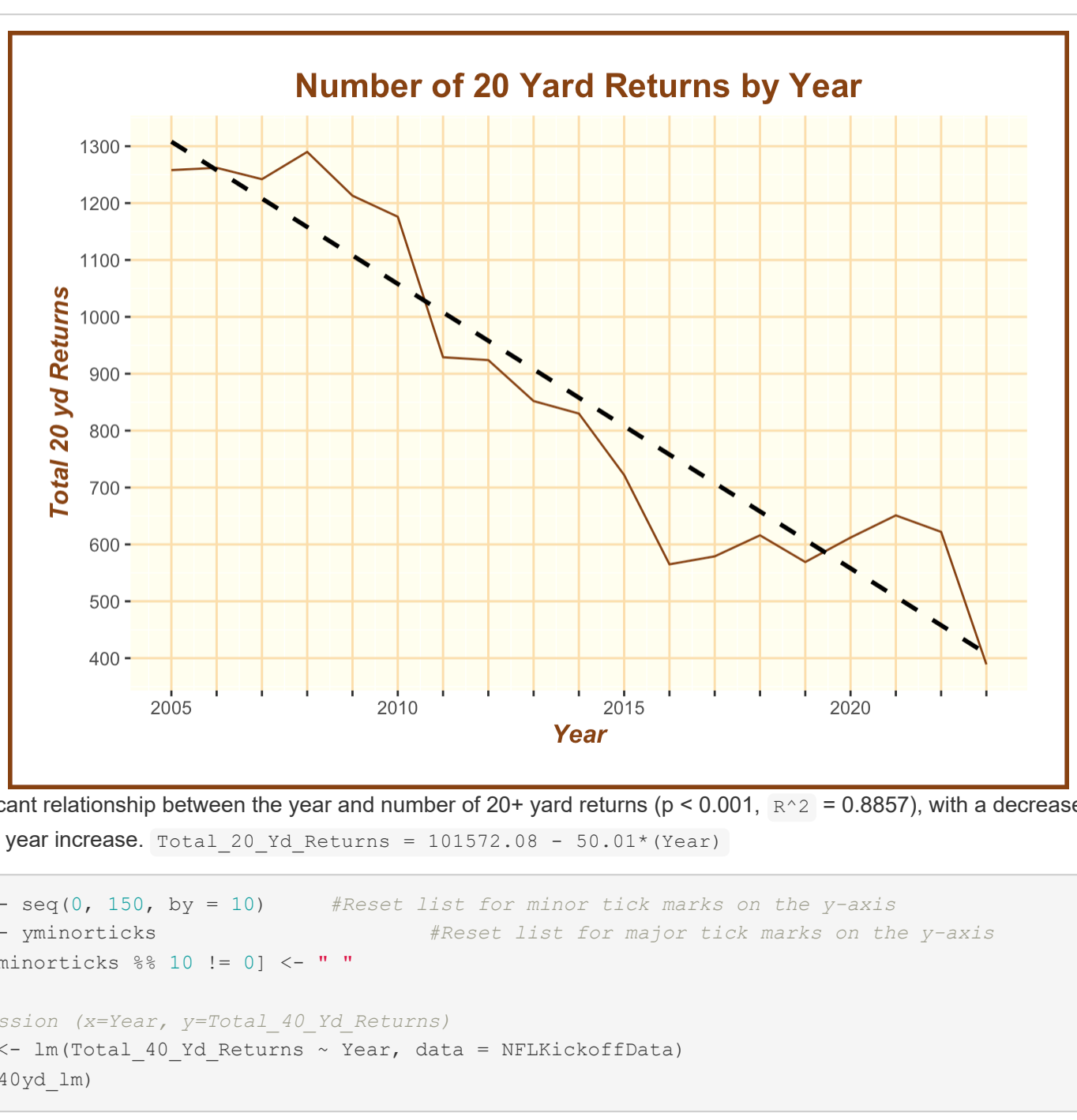
We found a significant relationship between the year and average kick return yards (p < 0.001, `R^2` = 0.8693), with a 58.77-yard decrease in average kick return yards for every 1 year increase. `Avg_KRet_Yds_Per_Season = 119366.35 - 58.77*(Year)`

```r
yminorticks <- seq(0, 1500, by = 100)       #Reset list for minor tick marks on the y-axis
ymajorticks <- yminorticks                   #Reset list for major tick marks on the y-axis
ymajorticks[yminorticks %% 100 != 0] <- " "   #Make all entries in ymajorticks not divisible by 100 blank

#Linear regression (x=Year, y=Total_20_Yd_Returns)
year_20yd_lm <- lm(Total_20_Yd_Returns ~ Year, data = NFLKickoffData)
summary(year_20yd_lm)
```

```
##
## Call:
## lm(formula = Total_20_Yd_Returns ~ Year, data = NFLKickoffData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -192.93  -52.98  -27.95   79.56  164.11
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 101572.08    8498.90   11.95 1.07e-09 ***
## Year           -50.01       4.22  -11.85 1.22e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 100.7 on 17 degrees of freedom
## Multiple R-squared:  0.892,  Adjusted R-squared:  0.8857
## F-statistic: 140.4 on 1 and 17 DF,  p-value: 1.219e-09
```

```r
#Create plot using Year and Total_20_Yd_Returns fields
ggplot(NFLKickoffData, aes(Year, Total_20_Yd_Returns)) +
  geom_line(color = "saddlebrown") +
  ggtitle("Number of 20 Yard Returns by Year") +
  labs(x = "Year", y = "Total 20 yd Returns") +
  theme(axis.title = element_text(size = 12, color = "saddlebrown", face = "bold.italic"),
        plot.title = element_text(size = 16, color = "saddlebrown", face = "bold", hjust=0.5),
        panel.background = element_rect(fill = "#FFFFF0"),
        panel.grid.major = element_line(color = "navajowhite"),
        plot.background = element_rect(color = "saddlebrown", linewidth = 2),
        plot.margin = margin(t = 20, r = 20, b = 20, l = 20)) +
  scale_x_continuous(breaks = xminorticks, labels = xmajorticks) +
  scale_y_continuous(breaks = yminorticks, labels = ymajorticks) +
  geom_smooth(method="lm", se=FALSE, col="black", linetype = "dashed")
```

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



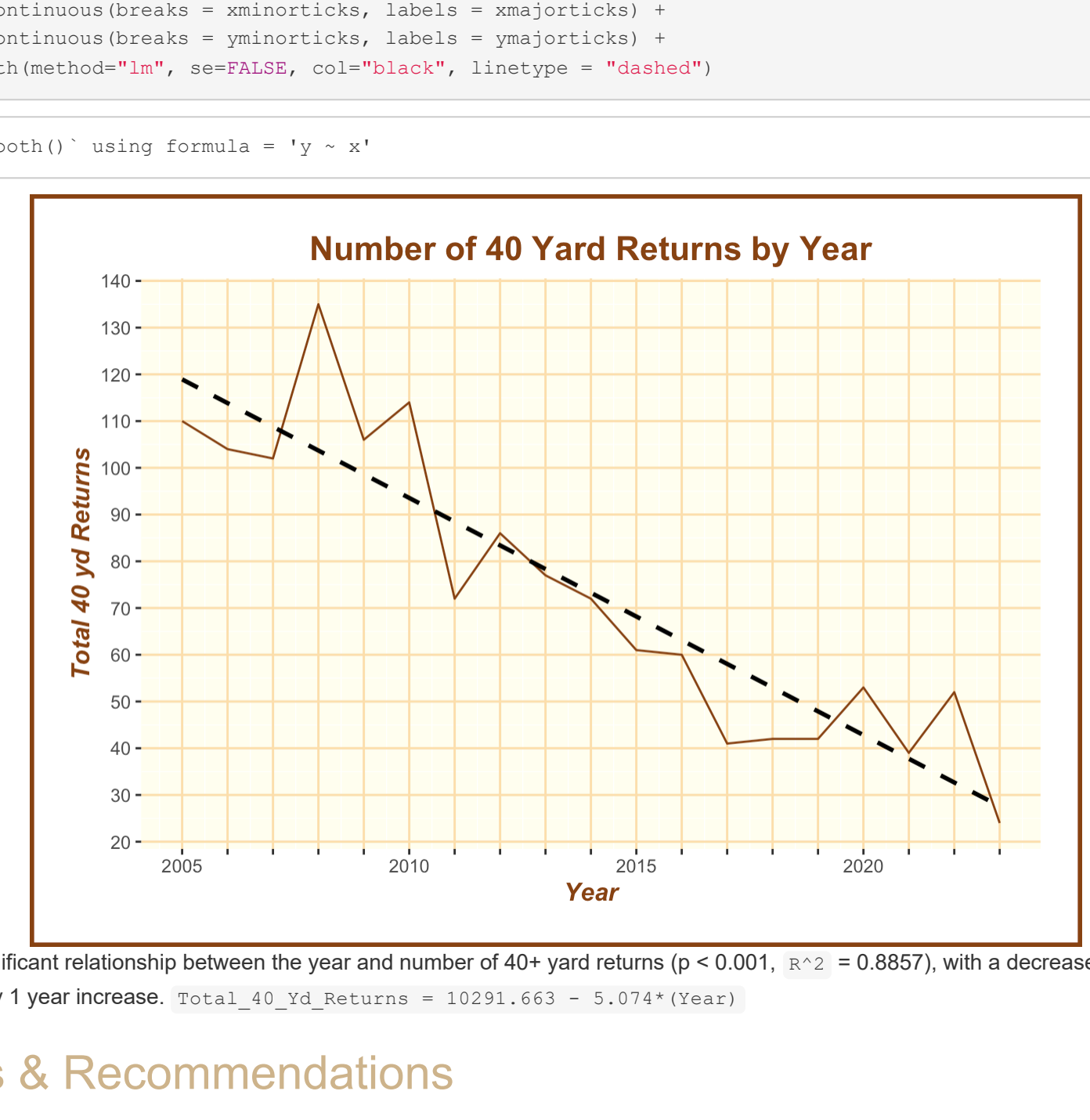
We found a significant relationship between the year and number of 20+ yard returns (p < 0.001, `R^2` = 0.8857), with a decrease of 50 20-yard returns for every 1 year increase. `Total_20_Yd_Returns = 101572.08 - 50.01*(Year)`

```r
yminorticks <- seq(0, 150, by = 10)      #Reset list for minor tick marks on the y-axis
ymajorticks <- yminorticks                   #Reset list for major tick marks on the y-axis
ymajorticks[yminorticks %% 10 != 0] <- " "

#Linear regression (x=Year, y=Total_40_Yd_Returns)
year_40yd_lm <- lm(Total_40_Yd_Returns ~ Year, data = NFLKickoffData)
summary(year_40yd_lm)
```

```
##
## Call:
## lm(formula = Total_40_Yd_Returns ~ Year, data = NFLKickoffData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.042  -8.058  -3.116   4.979  31.295
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10291.6632  1112.1100   9.254 4.76e-08 ***
## Year           -5.0737     0.5522  -9.188 5.28e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.18 on 17 degrees of freedom
## Multiple R-squared:  0.8324, Adjusted R-squared:  0.8225
## F-statistic: 84.43 on 1 and 17 DF,  p-value: 5.275e-08
```

```r
#Create plot using Year and Total_20_Yd_Returns fields
ggplot(NFLKickoffData, aes(Year, Total_40_Yd_Returns)) +
  geom_line(color = "saddlebrown") +
  ggtitle("Number of 40 Yard Returns by Year") +
  labs(x = "Year", y = "Total 40 yd Returns") +
  theme(axis.title = element_text(size = 12, color = "saddlebrown", face = "bold.italic"),
        plot.title = element_text(size = 16, color = "saddlebrown", face = "bold", hjust=0.5),
        panel.background = element_rect(fill = "#FFFFF0"),
        panel.grid.major = element_line(color = "navajowhite"),
        plot.background = element_rect(color = "saddlebrown", linewidth = 2),
        plot.margin = margin(t = 20, r = 20, b = 20, l = 20)) +
  scale_x_continuous(breaks = xminorticks, labels = xmajorticks) +
  scale_y_continuous(breaks = yminorticks, labels = ymajorticks) +
  geom_smooth(method="lm", se=FALSE, col="black", linetype = "dashed")
```

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



We found a significant relationship between the year and number of 40+ yard returns (p < 0.001, `R^2` = 0.8857), with a decrease of ~5 40-yard returns for every 1 year increase. `Total_40_Yd_Returns = 10291.663 - 5.074*(Year)`

Insights & Recommendations

Several key data points emerge upon analyzing the percentage changes within the dataset. Notably, the years 2011 and 2023 stand out as exhibiting the most significant increases/decreases.

In 2011, there was a notable decline in the average number of kick return yards by approximately 400 yards compared to the previous year, representing a 28% decrease. Moreover, the total number of 20+ yard returns decreased by around 250 (a 21% decline), while the total number of 40+ yard returns dropped by 42 (a 36.8% decrease). The number of touchbacks surged dramatically from 416 to 1120 in this season, marking a substantial 169% increase due to the relocation of the kickoff to the 35-yard line. Additionally, 2011 witnessed the first instance where the average number of returns per team fell below 60, declining from 64 in 2010 to 43 (a 32.8% decrease).

In the most recent season (2023-24), there were significantly more fair catches than the cumulative total of the preceding 18 seasons. The league experienced a reduction of 300 kickoff return yards in 2023 compared to 2022, reflecting a 41.4% decrease. Touchbacks saw an increase of approximately 360 (a 22% rise). Notably, the number of 20+ and 40+ yard returns experienced substantial declines in 2023, decreasing by 37.5% and 53.9%, respectively. Furthermore, this season marked the first instance where the average number of kick returns dropped below 20.

According to the NFL, the implemented rule changes have significantly reduced injuries during kickoffs. However, it is crucial to strike a balance between permitting kickoffs while mitigating potential injuries. One proposed solution involves revising the kickoff setup. Additionally, it is recommended that a touchback or fair catch results in the ball being placed at the 20-yard line instead of the current 25-yard line. Presently, there is minimal incentive to field a kickoff given that the average kickoff yardage remains around 22-23 yards. Opting for a fair catch or touchback automatically grants the receiving team an additional 2-3 yards. The league could consider adopting strategies akin to those seen in the XFL or UFL to encourage more kickoffs while reducing injury risks. Failure to address these concerns may lead to the potential extinction of one of the NFL's most dynamic and thrilling plays within the next 5 years.