In today's world of Major League Baseball we tend to see pitchers with higher velocities then what we were used to ten years ago. Could this be from an increase in focus on biomechanics? Could it be from even more specialized training than before? Or, are our memories deceiving us and increased velocity isn't real?

It would seem that MLB's focus on increased velocities stems from increased focus on both biomechanics and specialized training. This is proven with programs such as Driveline Baseball which has a significant roll in using biomechanics and data analytics to implement their training programs. However, could this increased focus on velocity be having a negative effect on control? Could increased velocity lead to more hit by pitches? More base on balls? More losses than wins?

The data collecting, cleaning, and transforming were all fairly simple. I downloaded the CSV from https: //www.fangraphs.com/. From the homepage of fangraphs I clicked on Team Stats. From there I went down to the Multiple Seasons and entered 2007 to 2020. I then went down to the Custom Leaderboards and created my own Leaderboard by selecting the columns Season, Team, Win, Loss, ERA, Base_on_Balls, HBP, WP, Balls, Strikes, K/9, BB/9, K/BB, and Velo. After looking at the data I realized that I wanted to get rid of the 2020 season where there were only 60 games played. I chose to do this because only 60 games played is a outlier to the regular 162 game season. In doing this I chose to only use odd numbered years in the data starting with 2009 and ending with 2021.

'''{r} library(tidyverse) library(knitr) library(readxl) install.packages("reshape2") library(reshape2) fullstats <- read_excel("full_stats07-21.xlsx")

fb21 <- fullstats %>% filter(Season=="2021") fb20 <- fullstats %>% filter(Season=="2020") fb19 <- fullstats %>% filter(Season=="2019") fb18 <- fullstats %>% filter(Season=="2018") fb17 <- fullstats %>% filter(Season=="2017") fb16 <- fullstats %>% filter(Season=="2016") fb15 <- fullstats %>% filter(Season=="2015") fb14 <- fullstats %>% filter(Season=="2014") fb13 <- fullstats %>% filter(Season=="2013") fb12 <- fullstats %>% filter(Season=="2012") fb11 <- fullstats %>% filter(Season=="2011") fb10 <- fullstats %>% filter(Season=="2010") fb09<- fullstats %>% filter(Season=="2009") fb08 <- fullstats %>% filter(Season=="2008") fb07 <- fullstats %>% filter(Season=="2007")

mean <- round(c(mean(fb21$velo), mean(fb19$velo),mean(fb17$velo), mean(fb15$velo),mean(fb13$velo), mean(fb11$velo), mean(fb09$velo)),2) season <- c("'21","'19","'17","'15","'13","'11","'09") mean_df <- data.frame(season,mean)

hbp_mean <- round(c(mean(fb21$HBP), mean(fb19HBP), mean(fb17$HBP), mean(fb15HBP),mean(fb13$HBP), mean(fb11 mean(fb09$HBP)),2) season <- c("'21","'19","'17","'15","'13","'11","'09") hbpmean_df <- data.frame(season, hbp_mean)

ggplot(mean_df, aes(season,mean, group=1))+ geom_point(colour="orange")+geom_line(size=1.4, colour="orange")+ xlab("Season")+ylab("Velocity")+ ggtitle("Avg. Fastball Velocity per Season")+theme_bw()

ggplot(hbpmean_df, aes(season, hbp_mean, group=1))+ geom_point(colour="blue")+geom_line(size=1.4, colour="blue")+ xlab("Season")+ylab("Avg. Hit by Pitch")+ ggtitle("Avg. Hit by Pitch per Season")+theme_bw()

linear <- lm(mean~ hbp_mean) summary(linear) '''