

Measuring The Effects of Age

```
knitr::opts_chunk$set(echo = TRUE)
require(dplyr)

## Loading required package: dplyr

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

require(Lahman)

## Loading required package: Lahman

require(ggplot2)

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.2.4
```

Introduction

In our group analysis of the role of aging in baseball performance, we decided to focus on stolen bases and caught stealing; both of these are commonly regarded as basic yet widely accepted measures of speed and agility.

Method

Our mission to operationalize the comparison of base stealing ability among players of different ages brought us to the use of two statistics, these being Stolen Base Percentage (SBP) and Stolen Base Ratio (SBR). SBP accounts for stolen bases over total steal attempts ($SB/(SB+CS)$), while SBR is simply Stolen Bases over Caught Stealing instances. In order to get a reasonable amount of data, we limited our data range to the years between 1970 and 2015.

```
data("Batting")
data("Master")
Batting.Age <- left_join(Batting,Master)

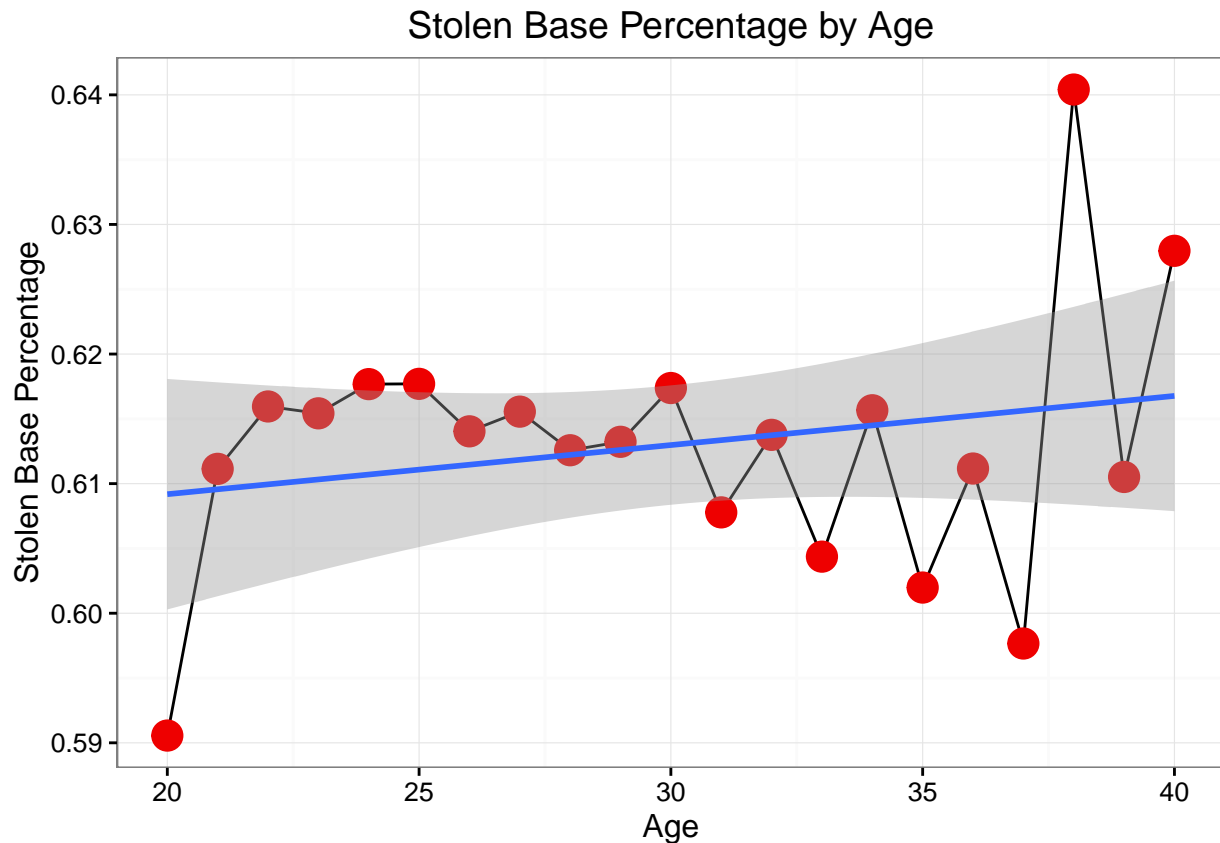
## Joining by: "playerID"
```

```

Batting.Age$age.corr <- ifelse(Batting.Age$birthMonth<7,0,-1)

Batting.Age %>% filter(yearID>1970) %>%
  filter(SB>0,CS>0) %>%
  mutate(age=yearID-birthYear+age.corr) %>%
  mutate(StealPct=((SB)/(SB+CS))) %>%
  select(SB,CS,nameFirst,nameLast,age,StealPct) %>%
  filter(age>19) %>% filter(age<41) %>%
  group_by(age) %>% summarize(mean.StealPct=mean(StealPct)) %>%
  ggplot(., aes(age, mean.StealPct)) + geom_line() + geom_point(col = "red2", cex=5) + xlab("Age") + ylab("Stolen Base Percentage")

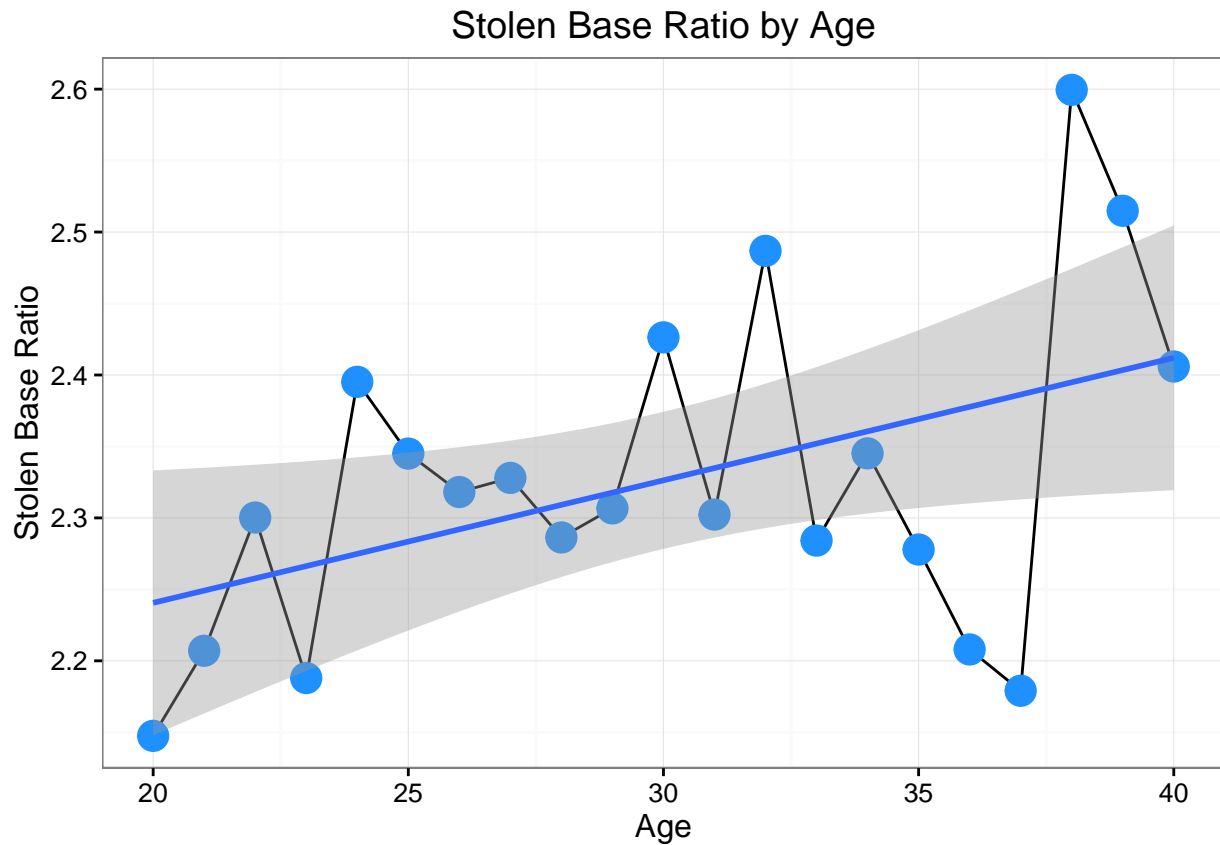
```



```

Batting.Age %>% filter(yearID>1970) %>%
  filter(SB>0,CS>0) %>%
  mutate(age=yearID-birthYear+age.corr) %>%
  mutate(StealRatio=((SB)/(CS))) %>%
  select(SB,CS,nameFirst,nameLast,age,StealRatio) %>%
  filter(age>19) %>% filter(age<41) %>%
  group_by(age) %>% summarize(mean.StealRatio=mean(StealRatio)) %>%
  ggplot(., aes(age, mean.StealRatio)) + geom_line() + geom_point(col = "dodgerblue", cex=5) + xlab("Age") + ylab("Stolen Base Ratio") + ggtitle("Stolen Base Ratio by Age") + theme_bw() + stat_smooth(method="lm")

```



Analysis and Results

Stolen Base Percentage and Stolen Base Ratio both show a general decline in the ages constituting the average Major League career, from 24 to 37. The notable peaks around age 38 in both SBP and SBR are the result of two causation factors of base stealing: older players are most likely going to steal either if the opportunity for a steal will encounter little challenge by the defence and therefore be successful, or if the baserunner is a proven base stealer with speed and agility despite their relatively advanced age. Other than these two caveates, the general decline in SBP and SBR as players age clearly shows that base stealing as a part of the game does not “age well.”