Assignment 3: Dangerous Foul Balls

Phil Graff

10/8/2020

Introduction

In the summer of 2019, I planned to take my partner, father, and four-year old daughter out to a baseball game. Earlier that spring, I was horrified to watch a video clip of a batter hitting a foul ball into the stands where it struck a two-year old child (Albert Almora batting against the Houston Astros). I knew my priority would be to find seats for my family behind protective netting. The Great American Ballpark in Cincinnati had recently installed netting down the first- and third-base lines, so we were easily able to find safe seats with a great view. Because of this experience, this article and corresponding dataset on FiveThirtyEight stood out to me.

https://fivethirtyeight.com/features/we-watched-906-foul-balls-to-find-out-where-the-most-dangerous-ones-land/ (https://fivethirtyeight.com/features/we-watched-906-foul-balls-to-find-out-where-the-most-dangerous-ones-land/)

Methodology

Because data on foul ball trajectories were not complete or readily available, FiveThirtyEight analysts set out to compile their own data. Their sampling methodology was to find the ten ballparks with the most foul balls in the 2019 season through June 5, and then analyze the game day for each ballpark with the most foul balls. Foul ball type, location, and speed were compiled from StatCast data and television camera feeds into a .csv and made available to the public.

Narrative

Through the use of two charts, FiveThirtyEight argued that although Major League ballparks had uniformly installed netting that covered behind home plate and the dugouts (zones 1, 2, and 3 in their example), a significant number of foul balls landed in zones 4 and 5. In particular, *dangerous* foul balls (those being hit at speeds >= 90 mph) were landing in these zones, which did not have protective netting in all MLB parks.

Chart 1: Percentage of Foul Balls

```
# Import foul-balls.csv downloaded from Five Thirty-Eight's GitHub
# https://github.com/fivethirtyeight/data/blob/master/foul-balls/foul-balls.csv
foulballs <- read.csv("foul-balls.csv")
foulcounts <- table(foulballs$used_zone)
foulpct <- foulcounts / nrow(foulballs)
```

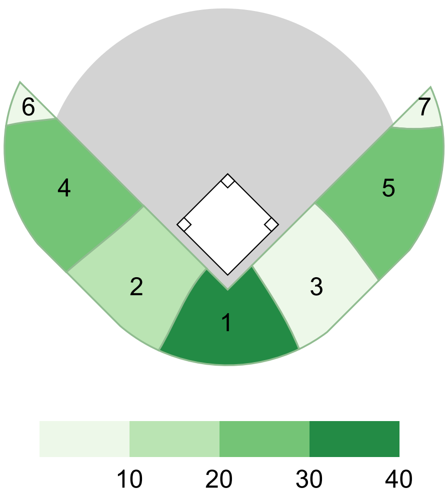
```
# Assign colors
library(RColorBrewer)

# derive number of color bins
bins <- ceiling(max(foulpct*10))

# Create palette for the number of bins
palette <- brewer.pal(n = bins, name = "Greens")

# Assign palette value to each stadium region
hex_vector <- palette[ceiling(foulpct*10)]

# Pass palette vector into D3 script to render field and color stands
library(r2d3)
r2d3(data=hex_vector, script = "baseballd3.js")
```



Percentage of Foul Balls by Zone (1-7)

Chart 1 Discussion

The reader can easily understand the foul ball distribution data when it is laid on top of a baseball field graphic, as FiveThirtyEight chose to do. Zone 1 receives the most foul balls, which makes sense because that is in the direction the ball is heading after the pitcher throws it. Zones 4 and 5, which often do not have protective netting, are the next most frequent landing zones for foul balls.

My replication of this graph did not improve on FiveThirtyEight's in any manner. In fact, due to my weak Adobe Illustrator skills, it appears less polished, but did allow for me to explore SVG creation and the D3 package introduced in class.

Chart 2: Location of Dangerous Foul Balls

```
# Add classification variable based on exit velocity: over90, under90, unknown
foulballs$velocity_class[is.na(foulballs$exit_velocity)] <- "unknown"
foulballs$velocity_class[foulballs$exit_velocity < 90] <- "under90"
foulballs$velocity_class[foulballs$exit_velocity >= 90] <- "over90"

# Create summary table by zone/velocity class
velocity_table <- as.data.frame(table(foulballs$used_zone, foulballs$velocity_class))

# Add column names
colnames(velocity_table) <- c("zone","speed","freq")
```

```
library(ggplot2)

p <- ggplot(velocity_table, aes(x = freq, y = zone)) +
  geom_col(aes(fill = speed), width = 0.8, position = position_stack(reverse = TRUE))
p + theme_minimal() +
  scale_fill_manual(values=c("#dd6666", "#66dd66", "#bbbbbb")) +
  labs(title = "Foul Ball Velocities by Zone",
       caption = "Data from FiveThirtyEight/Baseball Savant",
       x = "Frequency",
       y = "Zone")
```

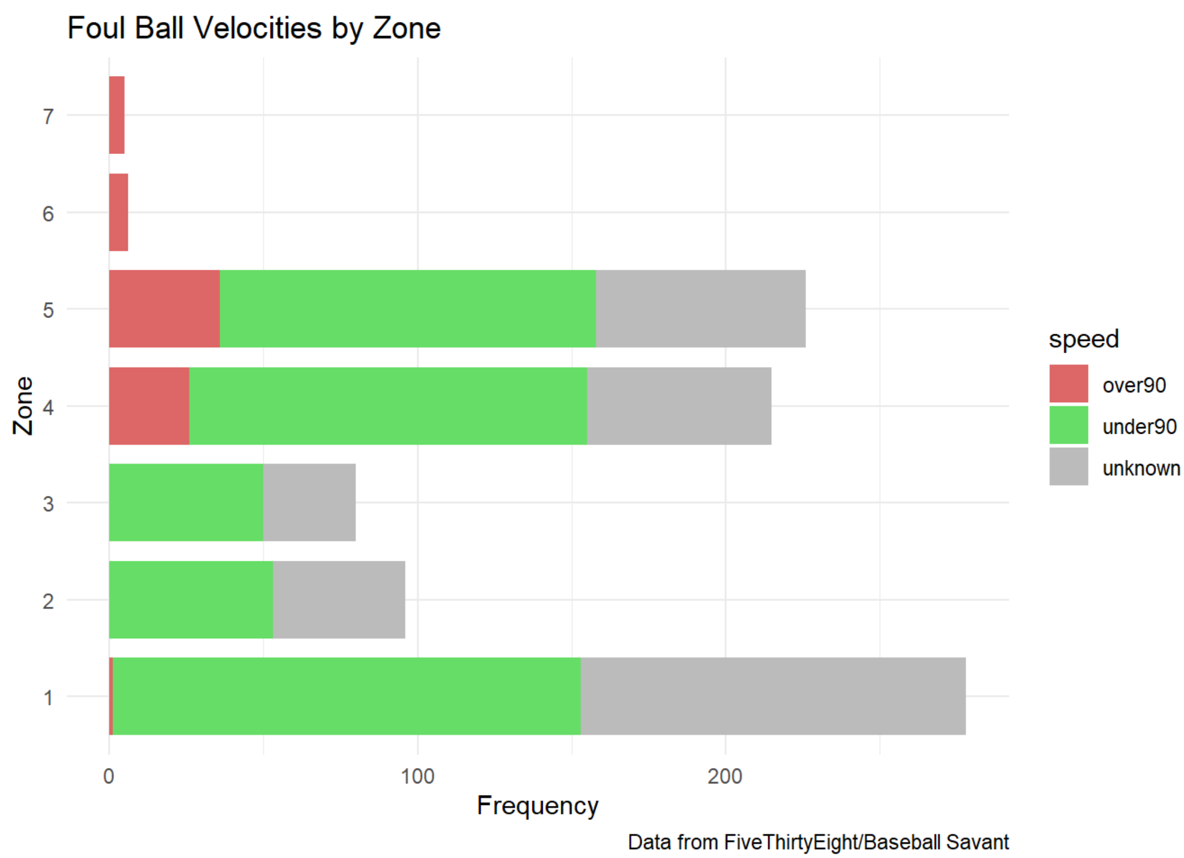


Chart 2 Discussion

I made two changes from the FiveThirtyEight chart:

First, I aligned the dangerous foul balls (90+ mph) to 0 on the x axis, because that makes it easier to compare them. Additionally, this creates a hierarchy reading left to right from most serious to least serious to unknown.

Second, I changed the color of the dangerous foul balls to red, I concede that red/green can create accessibility issues, but it taps into the common usage of red for danger or warning.

Critique

FiveThirtyEight supported these graphs and data with a number of recent examples of people being injured by foul balls. I agree with their assertion that zones 4 and 5 (from the ends of the dugouts to the foul poles) receive a significant percentage of foul balls, and a high number of dangerous foul balls. Based on these data, it seems to be in the best interest of Major League Baseball to protect the safety and well-being of its fans by extending the netting from the dugouts to the foul poles.

However, FiveThirtyEight also noted an incident where a fan in zone 1 – behind netting – was fatally struck by a foul ball. It is true that netting will not provide 100% safety, but I think closing with this example detracted from their argument in support of protective netting.

Final note: my GitHub for this project is available at https://github.com/mapsquatch/datastory_assn3 (https://github.com/mapsquatch/datastory_assn3). It contains all the files necessary to build this project.