ProjectAnalysis

Enyu Li

2025-02-18

Explanation Website: https://www.retrosheet.org/downloads/csvcontents.html

```
df <- read.csv("stat405baseball.csv",row.names = NULL)
df <- df[,-1]
```

Make the Histogram of the Pitch counts

```
par(tck = 0.02,
    mgp = c(1.5,0.4,0),
    mar = c(2.5, 2.5, 2.1, 0.5))
h <- hist(df$count, breaks = 20,
     ylim = c(0, 400000),
     xlab = "Pitch Count",
     ylab = "Frequency",
     main = "Histogram of the Pitch Count",
     axes = FALSE)

# manually add the axes
my_x_axis <- seq(0,max(df$count),by = 2)
my_y_axis <- seq(0,400000, by = 20000)
axis(1, at = my_x_axis,labels=as.character(my_x_axis), cex.axis=0.8)
axis(2, at = my_y_axis) #labels = my_axis

# add the number of counts on top of each bin
no_zero_indices <- h$counts > 0
text(h$mids[no_zero_indices],
     h$counts[no_zero_indices],
     labels=h$counts[no_zero_indices], pos=3, cex=0.8, col="blue",adj=c(0.5, -0.5))
```

Histogram of the Pitch Count