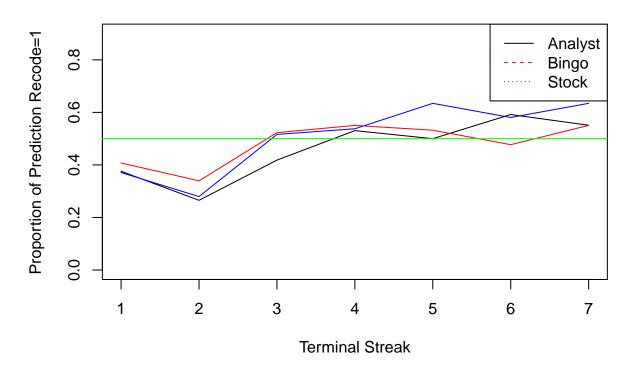```r
library(readxl)
data <- read_excel("../Data/PredictingOutcomes_ParticipantPredictions.xlsx", sheet = "Study 3B")

# divide the data based on the generator
data1 <- data[data$generator == "analyst",]
data2 <- data[data$generator == "bingo",]
data3 <- data[data$generator == "stock",]
```

give count of entries in all three data

```r
nrow(data1)
```

```
## [1] 1764
```

```r
nrow(data2)
```

```
## [1] 1962
```

```r
nrow(data3)
```

```
## [1] 1674
```

calculate the proportion of participants who predicted the prediction_recode=1 for each terminal_streak_length from 1 to 7

```r
prop1 <- aggregate(data1$prediction_recode, by = list(data1$terminal_streak_length), FUN = mean)
prop2 <- aggregate(data2$prediction_recode, by = list(data2$terminal_streak_length), FUN = mean)
prop3 <- aggregate(data3$prediction_recode, by = list(data3$terminal_streak_length), FUN = mean)
```

```r
prop1
```

```
##   Group.1         x
## 1       1 0.3758503
## 2       2 0.2653061
## 3       3 0.4183673
## 4       4 0.5306122
## 5       5 0.5000000
## 6       6 0.5918367
## 7       7 0.5510204
```

```r
prop2
```

```
##   Group.1         x
## 1       1 0.4074924
## 2       2 0.3394495
## 3       3 0.5229358
## 4       4 0.5504587
## 5       5 0.5321101
## 6       6 0.4770642
## 7       7 0.5504587
```

```r
prop3
```

```
##   Group.1         x
## 1       1 0.3709677
## 2       2 0.2795699
## 3       3 0.5161290
## 4       4 0.5376344
## 5       5 0.6344086
```

```
## 6        6 0.5806452
## 7        7 0.6344086
```

```
plot(prop1$Group.1,prop1$x, type = "l",ylim=c(0.0,0.9), xlab = "Terminal Streak", ylab = "Proportion of
lines(prop2$Group.1,prop2$x, col = "red")
lines(prop3$Group.1,prop3$x,  col = "blue")
abline(h = 0.5, col = "green")
legend("topright", legend = c("Analyst", "Bingo", "Stock"), col = c("black", "red", "blue"), lty = 1:3)
```

Proportion of Prediction Recode=1 for each Terminal Streak