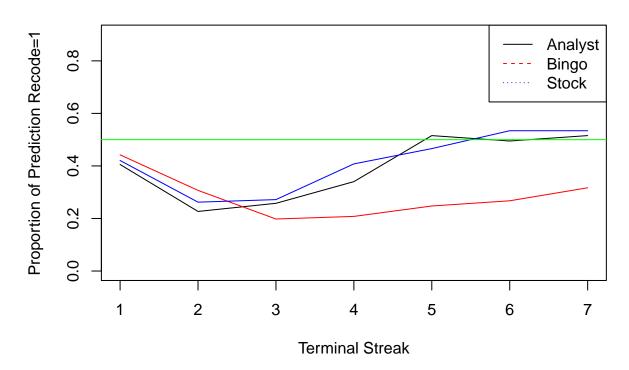```r
library(readxl)
data <- read_excel("../Data/PredictingOutcomes_ParticipantPredictions.xlsx", sheet = "Study 2B")

# 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] 1746
```

```r
nrow(data2)
```

```
## [1] 1818
```

```r
nrow(data3)
```

```
## [1] 1854
```

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 0.4054983
## 2       2 0.2268041
## 3       3 0.2577320
## 4       4 0.3402062
## 5       5 0.5154639
## 6       6 0.4948454
## 7       7 0.5154639
```

```r
prop2
```

```
##   Group.1         x
## 1       1 0.4422442
## 2       2 0.3069307
## 3       3 0.1980198
## 4       4 0.2079208
## 5       5 0.2475248
## 6       6 0.2673267
## 7       7 0.3168317
```

```r
prop3
```

```
##   Group.1         x
## 1       1 0.4207120
## 2       2 0.2621359
## 3       3 0.2718447
## 4       4 0.4077670
## 5       5 0.4660194
```

```
## 6        6 0.5339806
## 7        7 0.5339806
```

```r
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