```r
library(readxl)
library(lme4)
```

```
## Loading required package: Matrix
```

```r
library(lmerTest)
```

```
##
## Attaching package: 'lmerTest'
```

```
## The following object is masked from 'package:lme4':
##
##     lmer
```

```
## The following object is masked from 'package:stats':
##
##     step
```

```r
library(car)
```

```
## Loading required package: carData
```

```r
library(MuMIn)
```

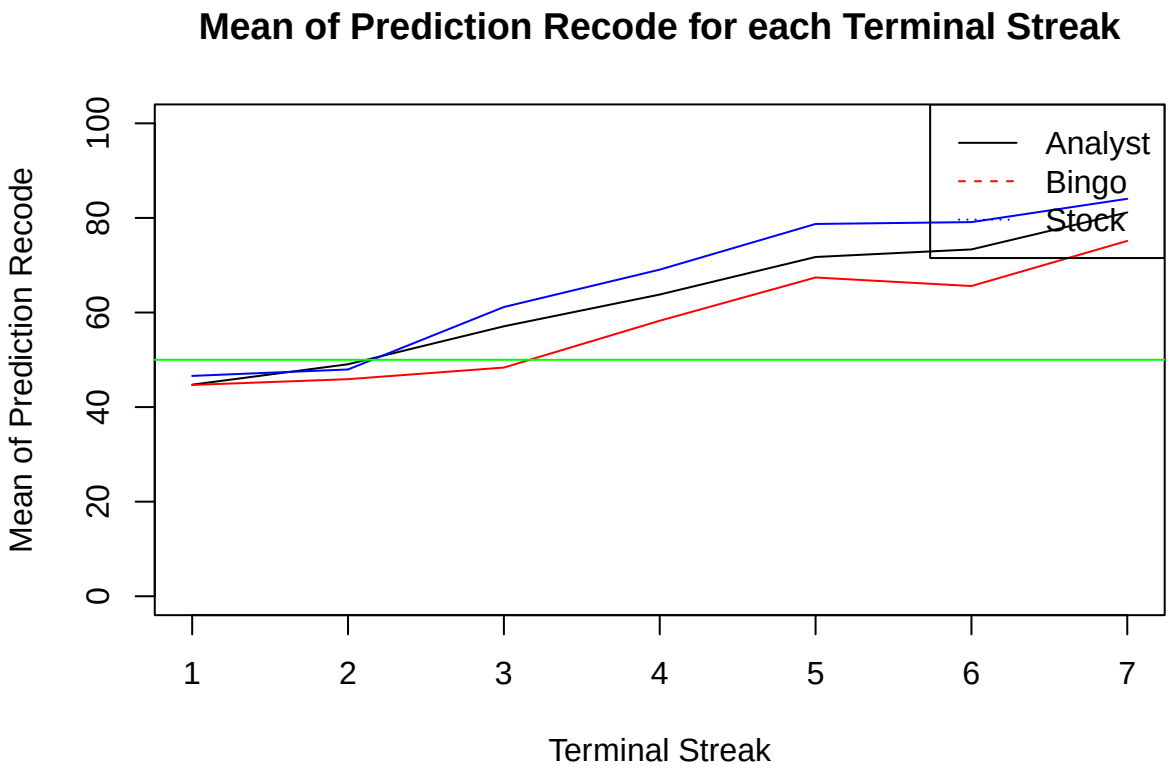
```r
data <- read_excel("../Data/PredictingOutcomes_ParticipantPredictions.xlsx", sheet = "Study 1A")

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

calculate the mean of prediciotn_recode for each terminwal streak from 1 to 7

```r
mean1 <- aggregate(data1$prediction_recode, by = list(data1$terminal_streak_length), FUN = mean)
mean2 <- aggregate(data2$prediction_recode, by = list(data2$terminal_streak_length), FUN = mean)
mean3 <- aggregate(data3$prediction_recode, by = list(data3$terminal_streak_length), FUN = mean)
```

```r
plot(mean1$Group.1,mean1$x, type = "l",ylim=c(0,100), xlab = "Terminal Streak", ylab = "Mean of Predicti
lines(mean2$Group.1,mean2$x, col = "red")
lines(mean3$Group.1,mean3$x,  col = "blue")
abline(h = 50, col = "green")
legend("topright", legend = c("Analyst", "Bingo", "Stock"), col = c("black", "red", "blue"), lty = 1:3)
```



calculate the mean of prediciotn_recode for each participant_id

```
mean1_id <- aggregate(data1$prediction_recode, by = list(data1$participant_id), FUN = median)
mean2_id <- aggregate(data2$prediction_recode, by = list(data2$participant_id), FUN = median)
mean3_id <- aggregate(data3$prediction_recode, by = list(data3$participant_id), FUN = median)
```

apply kruskal wais test so that the p value is 0.018

```
kruskal.test(mean1_id$x, mean2_id$x, mean3_id$x)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  mean1_id$x and mean2_id$x
## Kruskal-Wallis chi-squared = 28.016, df = 32, p-value = 0.6686
```

apply a one way mixed anova to test the effect of condition and one within streak length on the rating of probability that the terminal streak would repeat

```
model <- lmer(prediction_recode ~ generator + terminal_streak_length + (1|participant_id), data = data)
summary(model)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: prediction_recode ~ generator + terminal_streak_length + (1 |
##     participant_id)
##    Data: data
##
## REML criterion at convergence: 23923
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -2.9610 -0.7164  0.1112  0.6704  2.5875
##
## Random effects:
##  Groups         Name        Variance Std.Dev.
##  participant_id (Intercept)  22.7     4.764
##  Residual                   581.5    24.114
## Number of obs: 2592, groups:  participant_id, 144
##
## Fixed effects:
##                          Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)               39.1414     1.1769  222.6852  33.258   <2e-16 ***
## generatorbingo            -2.0111     1.4832  141.0000  -1.356   0.1773
## generatorstock             2.5634     1.5330  141.0000   1.672   0.0967 .
## terminal_streak_length     5.8588     0.2464 2447.0000  23.776   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) gnrtrb gnrtrs
## generatrbng -0.630
## genertrstck -0.610  0.484
## trmnl_strk_ -0.454  0.000  0.000
```

tell the p value

```
anova(model)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##                         Sum Sq Mean Sq NumDF DenDF  F value  Pr(>F)
## generator                 5184    2592     2   141   4.4574 0.01327 *
## terminal_streak_length 328707  328707     1  2447 565.2955 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

calculate the p-value

```
p <- anova(model)$`Pr(>F)`
p
```

```
## [1]  1.327064e-02 1.357107e-112
```