Fish Data Analysis

Victor Chien

4/26/2020

Our group consists of: Victor Chien, Nat Jennings, Deniz Sevilmis, Jeremiah Wells, Doran Kaplan, Daniel Moses

```
#Acquire Dataset
#install.packages("tidyverse")
library(tidyverse)
library("readxl")

#MODIFY PATH - SHEET 2 IS TIDIED FOR R
fishdata <- read_excel("C:\\Users\\Victor Chien\\Documents\\University of Texas at Austin\\Junior\\BIO 3
```

```
#fishSB is just using shuttlebox assay data
fishSB <- fishdata %>% select(-white_scoto,-cross_scoto,-corner_scoto) %>% na.omit() %>%
  mutate(learner2 = case_when(learner=="0" ~ "nonlearner", learner=="1" ~ "learner"),
  trt2 = case_when(trt=="LS"|trt=="INT"|trt=="SS"~ "experienced", trt=="FF"|trt=="LL" ~ "naive"),
  trt3 = case_when(trt2=="naive" ~ 0, trt2=="experienced" ~ 1))

t.test(data=fishSB, corner_shuttle ~ trt2)
```

```
##
##  Welch Two Sample t-test
##
## data:  corner_shuttle by trt2
## t = -0.87733, df = 30.116, p-value = 0.3873
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -38.51514  15.36514
## sample estimates:
## mean in group experienced       mean in group naive
##                   553.800                   565.375
```

```
t.test(data=fishSB, corner_shuttle ~ learner2)
```

```
##
##  Welch Two Sample t-test
##
## data:  corner_shuttle by learner2
## t = 0.81162, df = 29.431, p-value = 0.4235
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##  -16.95049  39.27825
## sample estimates:
##    mean in group learner mean in group nonlearner
##                 566.0769                 554.9130
```

Because there were many videos missing from the scototaxis assay video dataset, we will first analyze the data just using data from the shuttlebox assay. Using Welch's two sample t-test, we see that there is no significant difference in time spent in corners between coercive experience and coercive naive nor is there one between learners and nonlearners. This suggests that coercion experience does not lead to more time in the corner/make the fish more anxious (t = -0.87733, df = 30.116, p-value = 0.3873) and that time spent in corner does not predict the ability to learn (t = 0.81162, df = 29.431, p-value = 0.4235).

```
#fish is the tidy data with all variables (NAs omitted)
fish <- fishdata %>% na.omit %>%
  mutate(learner2 = case_when(learner=="0" ~ "nonlearner", learner=="1" ~ "learner"),
  trt2 = case_when(trt=="LS"|trt=="INT"|trt=="SS"~ "experienced", trt=="FF"|trt=="LL" ~ "naive"),
  trt3 = case_when(trt2=="naive" ~ 0, trt2=="experienced" ~ 1))

head(fish)
```

```
## # A tibble: 6 x 10
##   name  corner_shuttle learner white_scoto corner_scoto cross_scoto trt
##   <chr>          <dbl>   <dbl>       <dbl>        <dbl>       <dbl> <chr>
## 1 Adele            534       0           0          474           0 LS
## 2 Alli~            558       1         138          444          18 FF
## 3 Amy              548       0         184          538          14 FF
## 4 Aria~            528       0         116          542          12 LS
## 5 Audr~            576       0          11          542           3 FF
## 6 Bern~            584       1         102          521           9 FF
## # ... with 3 more variables: learner2 <chr>, trt2 <chr>, trt3 <dbl>
```

```
#Summary Statistics

#grouped by treatment group
fish %>% group_by(trt) %>% summarize(n=n(),
                                     learners = sum(learner),
                                     mean_corner_shuttle = mean(corner_shuttle),
                                     mean_white_scoto = mean(white_scoto),
                                     mean_corner_scoto = mean(corner_scoto),
                                     mean_cross_scoto = mean(cross_scoto))
```

```
## # A tibble: 4 x 7
##   trt       n learners mean_corner_shu~ mean_white_scoto mean_corner_sco~
##   <chr> <int>    <dbl>            <dbl>            <dbl>            <dbl>
## 1 FF        4        2             566.             109.             511.
## 2 INT       3        1             536.             122.             538
## 3 LL        5        2             564.              79.8            496.
## 4 LS        8        3             561.             113              524.
## # ... with 1 more variable: mean_cross_scoto <dbl>
```

```
#grouped by coercive experienced and naive
fish %>% group_by(trt2) %>% summarize(n=n(),
```

```
                                        learners = sum(learner),
                                        mean_corner_shuttle = mean(corner_shuttle),
                                        mean_white_scoto = mean(white_scoto),
                                        mean_corner_scoto = mean(corner_scoto),
                                        mean_cross_scoto = mean(cross_scoto))
```

```
## # A tibble: 2 x 7
##   trt2      n learners mean_corner_shu~ mean_white_scoto mean_corner_sco~
##   <chr> <int>    <dbl>            <dbl>            <dbl>            <dbl>
## 1 expe~    11        4              554             116.              527.
## 2 naive     9        4              565.             92.7             503.
## # ... with 1 more variable: mean_cross_scoto <dbl>
```

```
#grouped by learners
fish %>% group_by(learner2) %>% summarize(n=n(),
                                        mean_corner_shuttle = mean(corner_shuttle),
                                        mean_white_scoto = mean(white_scoto),
                                        mean_corner_scoto = mean(corner_scoto),
                                        mean_cross_scoto = mean(cross_scoto),
                                        experienced=sum(trt2=="experienced")
                                        )
```

```
## # A tibble: 2 x 7
##   learner2     n mean_corner_shu~ mean_white_scoto mean_corner_sco~
##   <chr>    <int>            <dbl>            <dbl>            <dbl>
## 1 learner      8              569              123              523.
## 2 nonlear~    12              552.              93.4            512.
## # ... with 2 more variables: mean_cross_scoto <dbl>, experienced <int>
```

Summary statistics have been run on the smaller fish dataset using both shuttlebox and scototaxis assays. The question we are attempting to answer is: Are bolder fish better at learning? It is important to note that there are no SS treatment fish in this datasest, which could be more extreme in anxiety levels.

```
#Correlation
fish %>% select(corner_shuttle:cross_scoto) %>% cor() %>% round(4)
```

```
##                corner_shuttle learner white_scoto corner_scoto cross_scoto
## corner_shuttle         1.0000  0.1931     -0.0572      -0.0096     -0.0621
## learner                0.1931  1.0000      0.1946       0.1417      0.0795
## white_scoto           -0.0572  0.1946      1.0000       0.3856      0.7286
## corner_scoto          -0.0096  0.1417      0.3856       1.0000      0.1845
## cross_scoto           -0.0621  0.0795      0.7286       0.1845      1.0000
```

```
cor.test(fish$cross_scoto,fish$white_scoto) #significant!
```

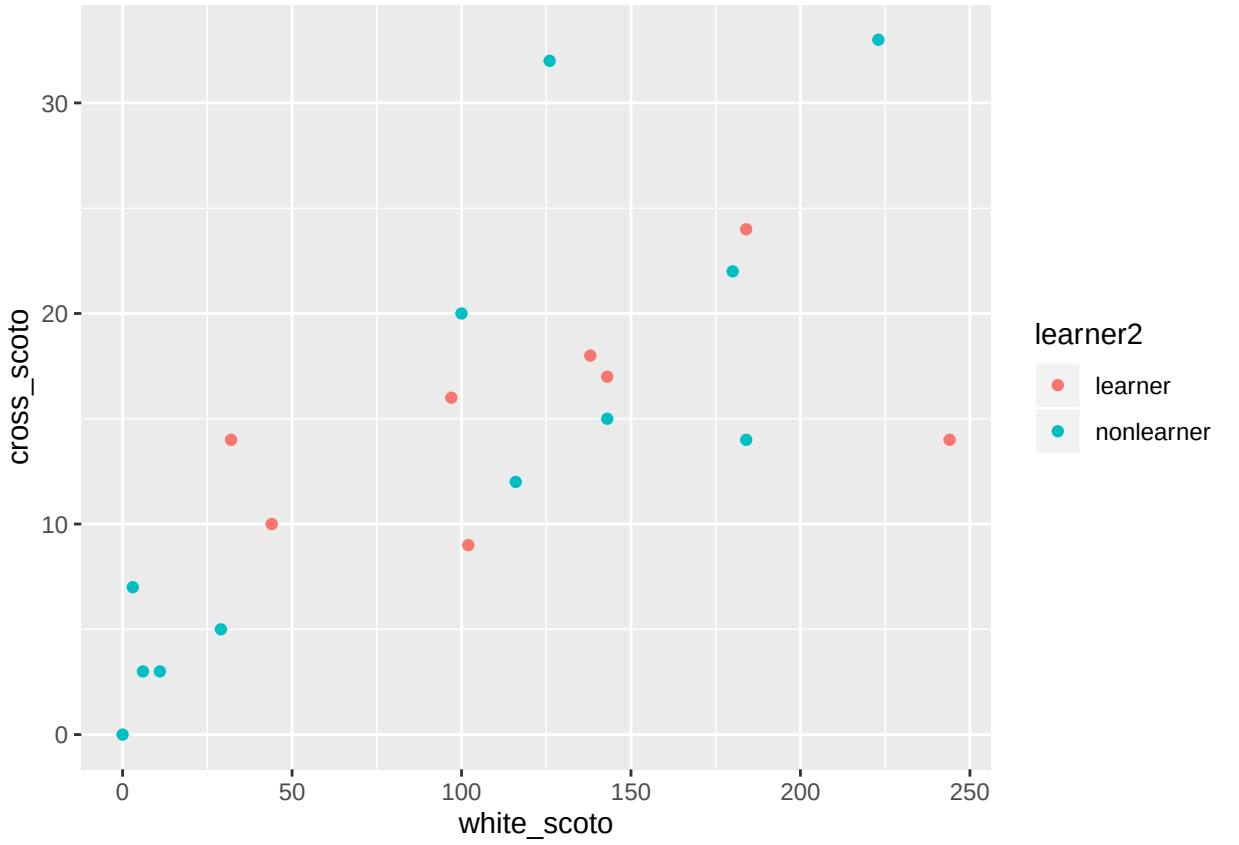
```
##
##  Pearson's product-moment correlation
##
## data:  fish$cross_scoto and fish$white_scoto
## t = 4.5133, df = 18, p-value = 0.000269
## alternative hypothesis: true correlation is not equal to 0
```

```
## 95 percent confidence interval:
##  0.4222419 0.8855977
## sample estimates:
##       cor
## 0.7286195
```

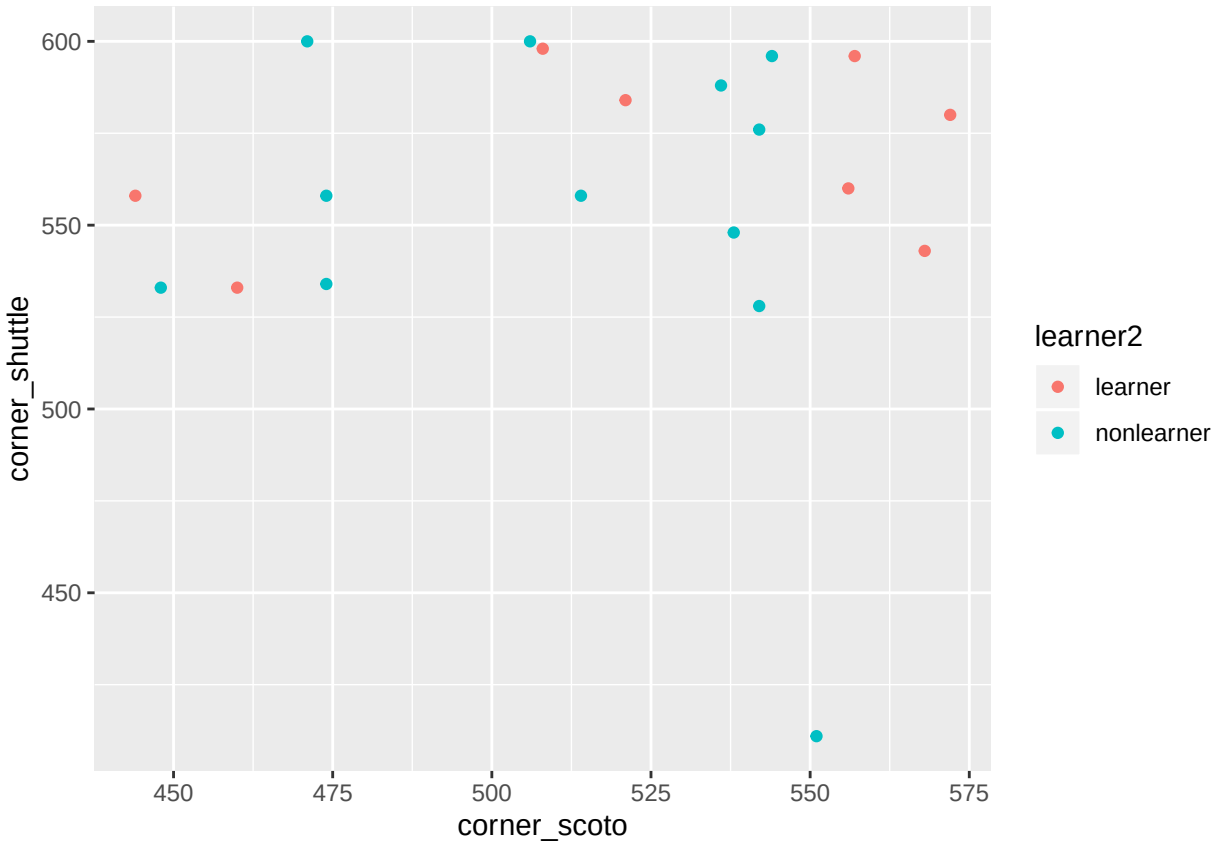
We found one high correlation between the time spent in white in the scototaxis assay and number of crosses into white in the scototaxis assay (r = 0.7286). There are no significant linear relationships between any of the other variables. We find it interesting that time in the corner in scototaxis and time in corner in shuttlebox has almost no relationship (r = -0.0096).

From cor.test, We found one significant correlation between the time spent in white in scototaxis and number of crosses into white in scototaxis (t = 4.5133, df = 18, p-value = 0.000269). This suggests that fish who crossed over into white more also tended to spend more time in white.

```
#cross_scoto and white_scoto are correlated
fish %>% ggplot(aes(x=white_scoto, y=cross_scoto, color = learner2)) +
  geom_point()
```




```
fish %>% ggplot(aes(x=corner_scoto, y=corner_shuttle, color = learner2)) +
  geom_point()
```



A plot of this significant correlation and a plot of the surprisingly uncorrelated thigmotaxis variables.

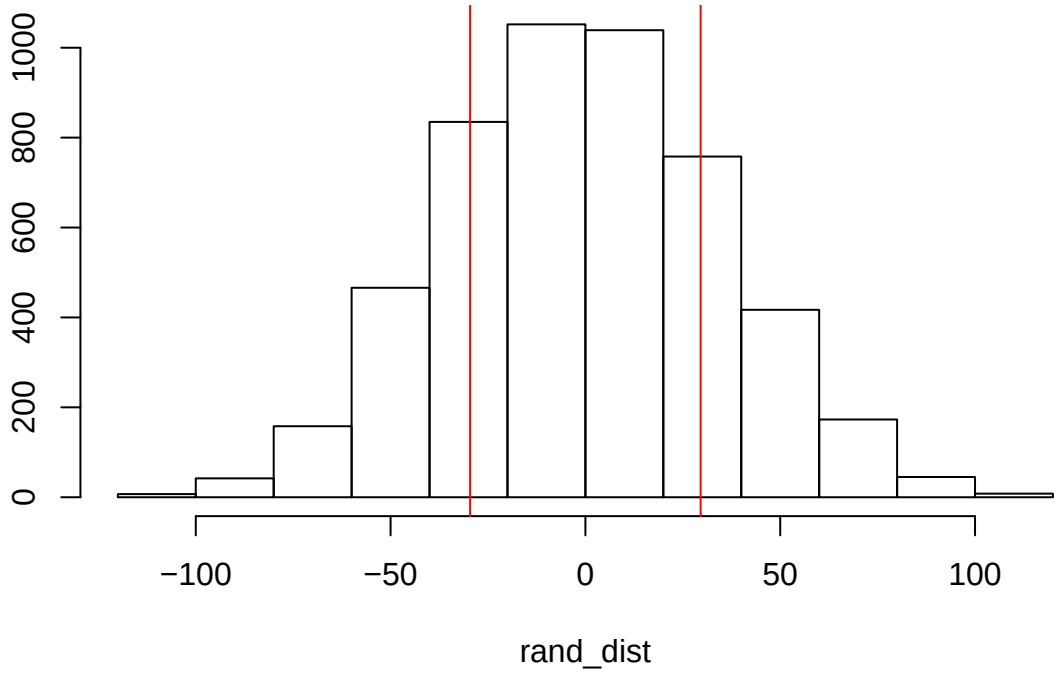
```
#Randomization Test
rand_dist<-vector()
cric<-data.frame(condition=c(rep("learner",8),rep("nonlearner",12)),time=fish$white_scoto)

for(i in 1:5000){
new<-data.frame(time=sample(cric$time),condition=cric$condition)
rand_dist[i]<-mean(new[new$condition=="learner",]$time)-
mean(new[new$condition=="nonlearner",]$time)}

fish %>% group_by(learner2) %>% summarize(mean(white_scoto)) #actual mean difference
```

```
## # A tibble: 2 x 2
##   learner2   `mean(white_scoto)`
##   <chr>                    <dbl>
## 1 learner                  123
## 2 nonlearner                93.4
```

```
hist(rand_dist,main="",ylab=""); abline(v = 29.5833, col="red"); abline(v = -29.5833, col="red")
```




```
mean(rand_dist>29.5833 | rand_dist< -29.5833) #two-tailed p-value
```

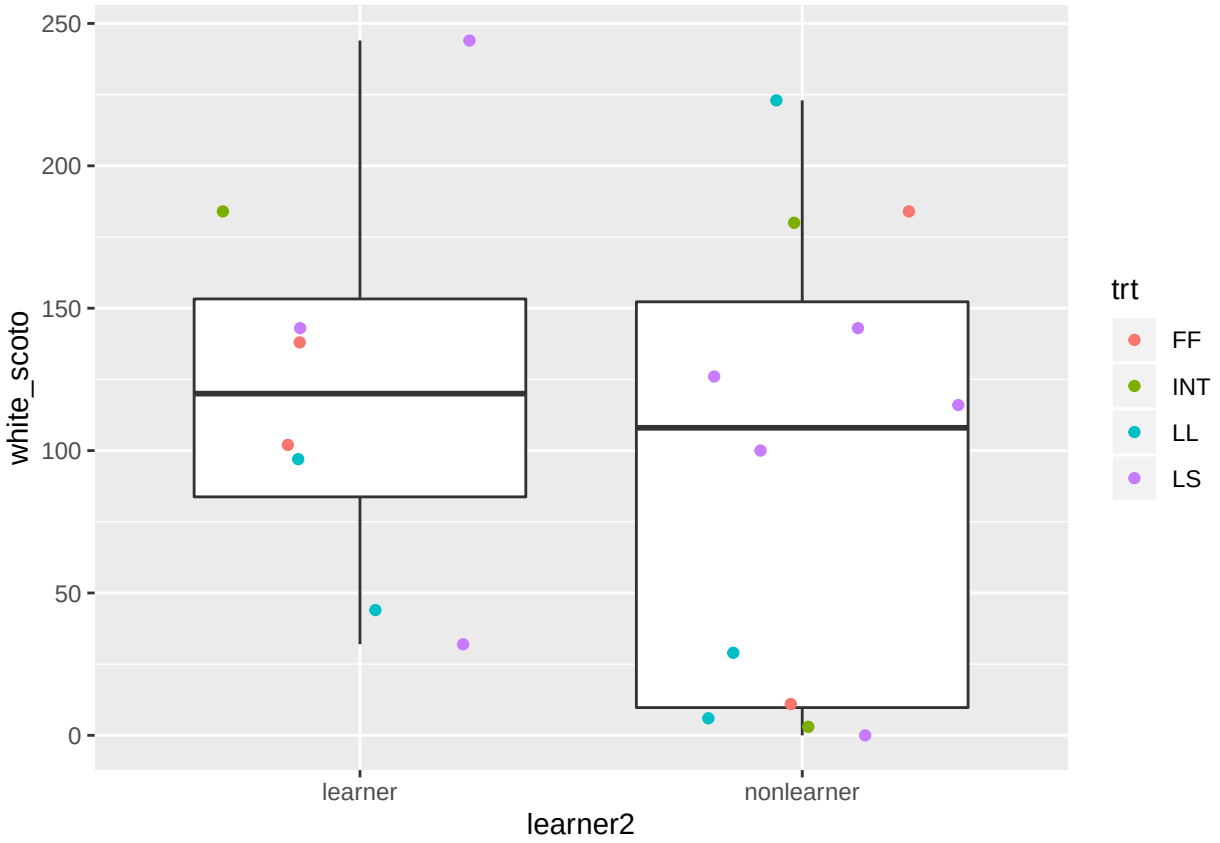
```
## [1] 0.4146
```

Since we have a small sample size, a randomization t test has been conducted to create a new null distribution. From the randomization test used above, we see that there is not a significant effect of any variable between learners and non learners (the code for only white_scoto is shown). White_scoto seems to have the lowest p-value (0.395 using randomization t test and 0.3983 using welch's two sample t test below).

```
#Two Sample t-test
t.test(data = fish, white_scoto ~ learner2) #between learners and non learners
```

```
##
##  Welch Two Sample t-test
##
## data:  white_scoto by learner2
## t = 0.86704, df = 16.588, p-value = 0.3983
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -42.54011 101.70678
## sample estimates:
##    mean in group learner mean in group nonlearner
##                123.00000                 93.41667
```

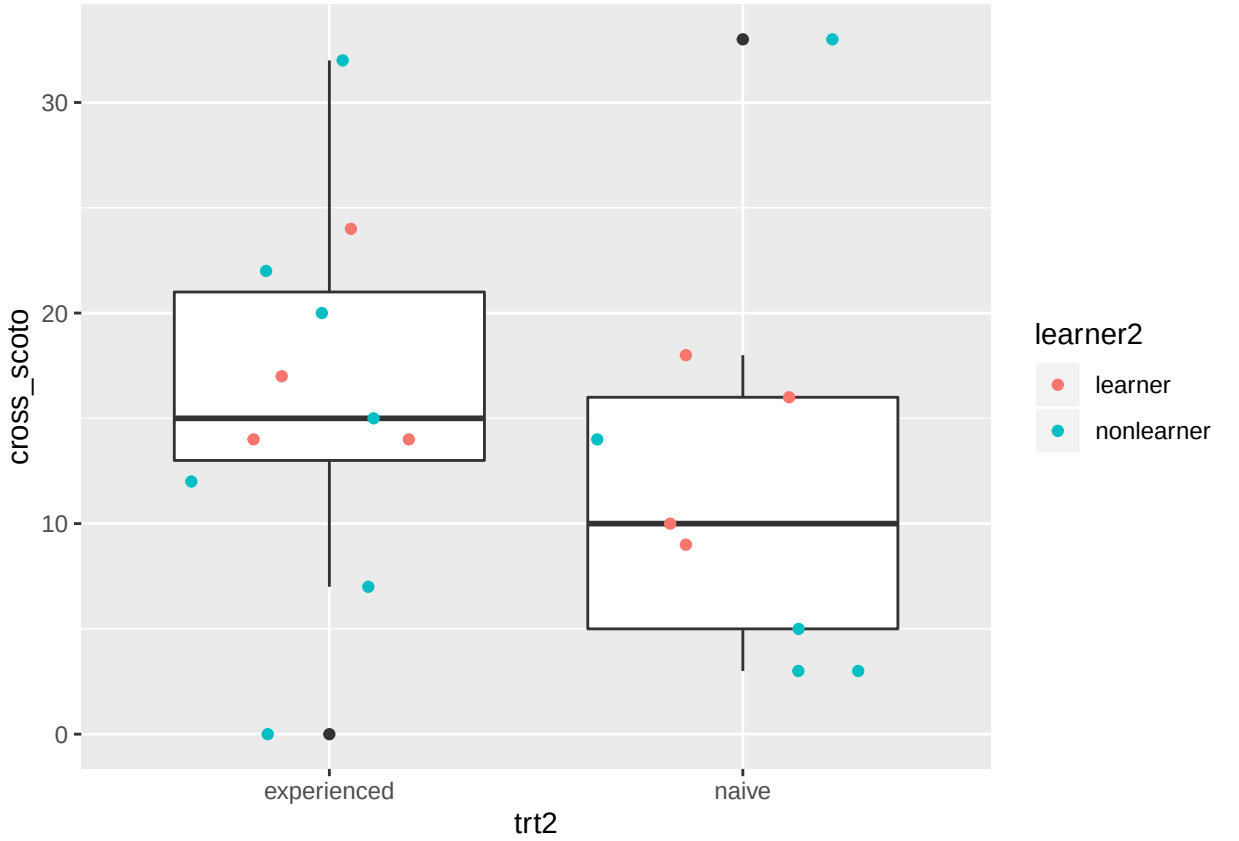
```r
#plot
fish %>% ggplot(aes(x=learner2,y=white_scoto)) +
  geom_boxplot() +
  geom_jitter(height=0, aes(color=trt))
```




```r
t.test(data = fish, cross_scoto ~ trt2) #between coercive experienced and coercive naive
```

```
##
##  Welch Two Sample t-test
##
## data:  cross_scoto by trt2
## t = 0.92002, df = 16.403, p-value = 0.3709
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.883358 12.398510
## sample estimates:
## mean in group experienced       mean in group naive
##                  16.09091                  12.33333
```

```r
#plot
fish %>% ggplot(aes(x=trt2,y=cross_scoto)) +
  geom_boxplot() +
  geom_jitter(height=0, aes(color=learner2))
```



We found no significant difference in means between time spent in white in the scototaxis assay between learners and non learners (t = 0.86704, df = 16.588, p-value = 0.3983). This suggests that there may be no relationship, or our sample size was too small to find a relationship.

We also found no significant difference in means for the number of crosses into the white area between coercion experienced and coercion naive. (t = 0.92002, df = 16.403, p-value = 0.3709). Crosses into white can be seen as a measure of boldness, meaning that coercion experience does not affect boldness, at least in this variable.

```
#MANOVA
man<-manova(cbind(corner_shuttle,white_scoto,corner_scoto,cross_scoto)~learner2, data=fish)
summary(man)
```

```
##           Df   Pillai approx F num Df den Df Pr(>F)
## learner2   1 0.090235  0.37194      4     15  0.825
## Residuals 18
```

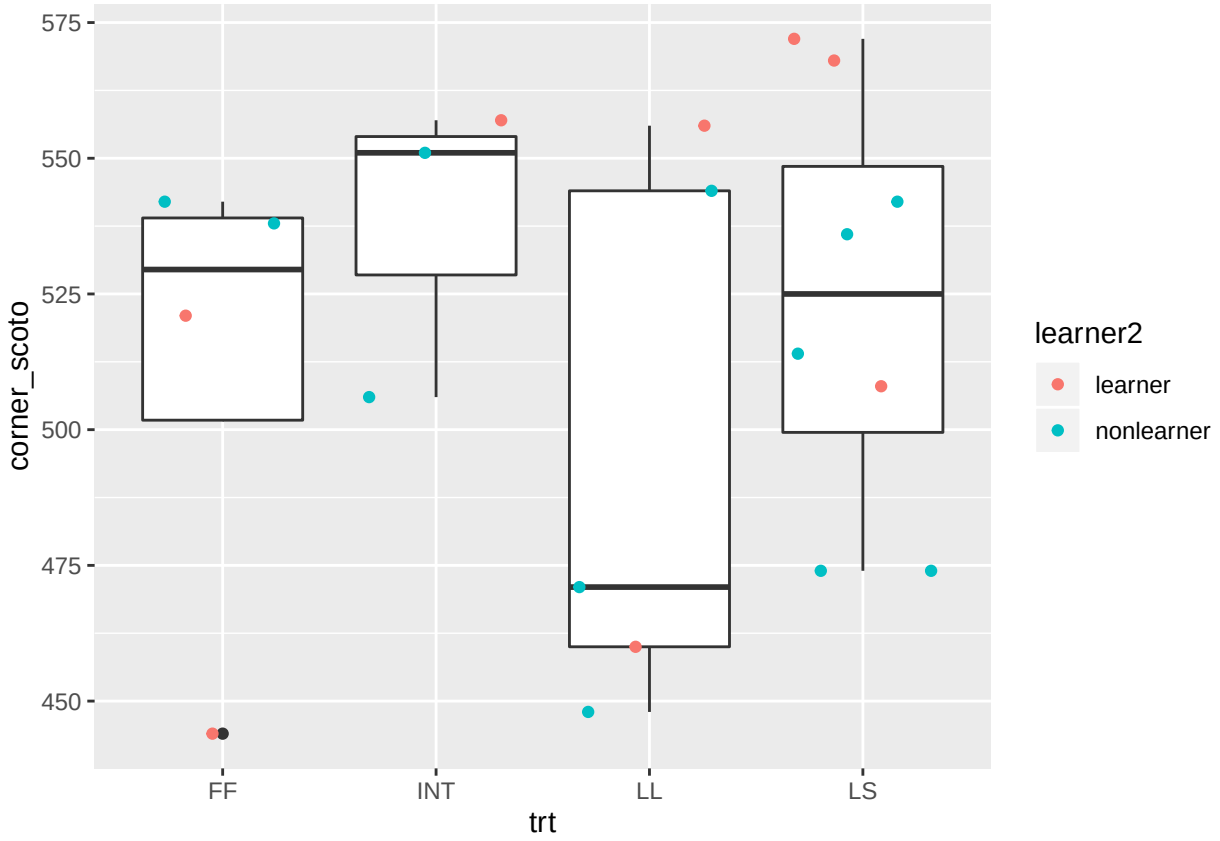
```
man2<-manova(cbind(corner_shuttle,white_scoto,corner_scoto,cross_scoto)~trt2, data=fish)
summary(man2)
```

```
##           Df  Pillai approx F num Df den Df Pr(>F)
## trt2       1 0.14916  0.65741      4     15 0.6309
## Residuals 18
```

```r
#ANOVA
summary(aov(corner_scoto~trt,data=fish)) #observed F-stat - no response variable is signifcant
```

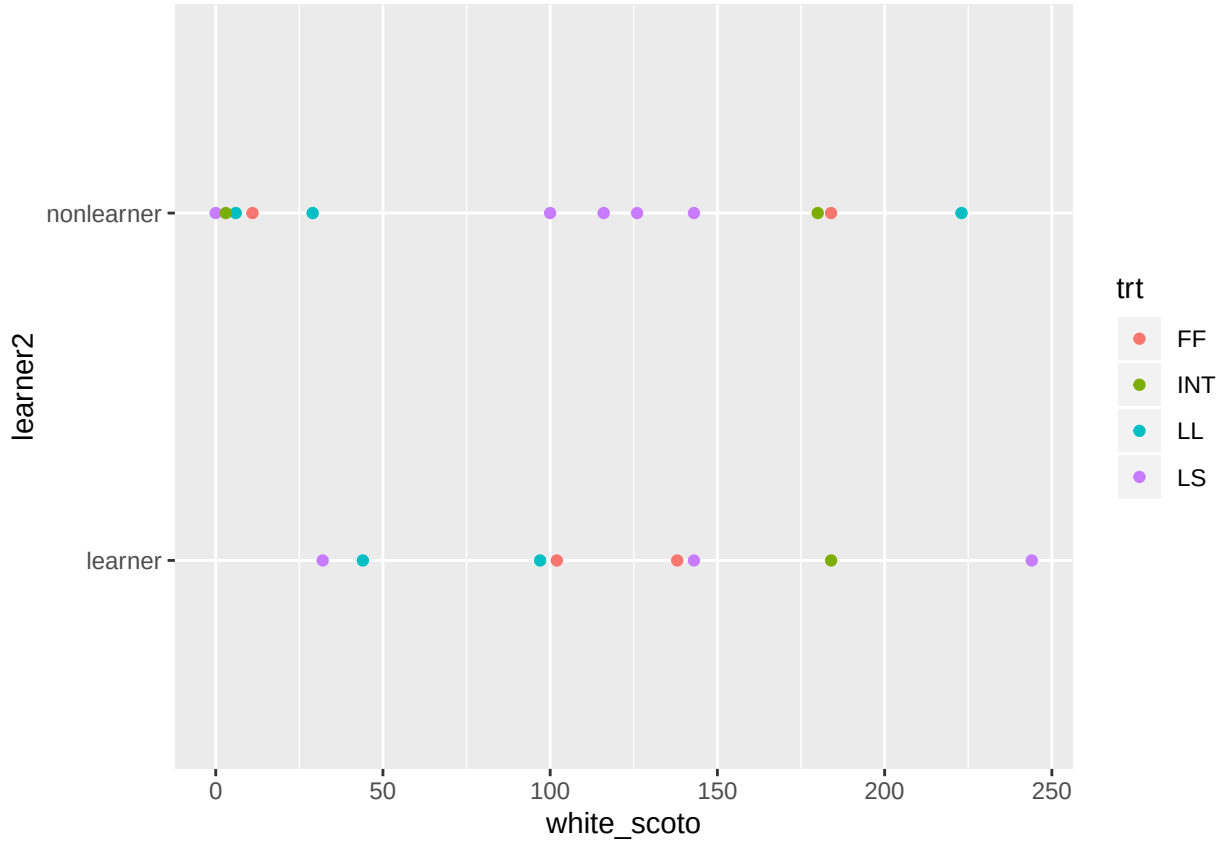
```
##              Df Sum Sq Mean Sq F value Pr(>F)
## trt           3   4031    1344   0.767  0.529
## Residuals    16  28024    1752
```

```r
#plot
fish %>% ggplot(aes(x=trt,y=corner_scoto)) +
  geom_boxplot() +
  geom_jitter(height = 0, aes(color=learner2))
```



A multivariate ANOVA (MANOVA) has been run on both learner2 and trt2 to see if there is a difference in means for any of the predictor variables. None of these variables are significantly different between learners and nonlearners as well as coercive naive and coercive experienced. An ANOVA has been run to compare the mean time in corner in the scototaxis assay between the 4 treatment groups. There is not a significant difference between any group.

```r
#binary plot
fish %>% ggplot(aes(x=white_scoto, y=learner2, color = trt)) +
  geom_point()
```




```
#logistic regression
fit <- glm(learner ~ white_scoto*cross_scoto, data=fish, family=binomial(link="logit"))
summary(fit)
```

```
##
## Call:
## glm(formula = learner ~ white_scoto * cross_scoto, family = binomial(link = "logit"),
##     data = fish)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4738  -0.9042  -0.3449   1.0331   1.7051
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -3.278575   2.075484  -1.580   0.1142
## white_scoto              0.035648   0.020878   1.707   0.0877 .
## cross_scoto              0.237210   0.185911   1.276   0.2020
## white_scoto:cross_scoto -0.002301   0.001509  -1.525   0.1273
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 26.920  on 19  degrees of freedom
```

```
## Residual deviance: 21.913  on 16  degrees of freedom
## AIC: 29.913
##
## Number of Fisher Scoring iterations: 5
```

Because our predictor variable (learner) is a binary variable, a logistic regression has been conducted instead of a linear regression. Coefficient estimates are log odds. Fitting the interaction between white_scoto and cross_scoto, we still see that neither are significant predictors of learning.

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.