STAT 217 Project 1

Ben Miller

Due: September 21, 2020 by 11:00pm in Gradescope

Dung odours signal sex, age, territorial and oestrous state in white rhinos

For this project you are encouraged to work in groups of up to four people. Save all files associated with the project in your STAT 217 folder. Questions should be answered in this Markdown file in bold text. All code and output must be included in final submission. Knit your completed Markdown file to Word, save this as a PDF, and submit to Gradescope by the deadline. When submitting your project, please indicate which page each question is on. Failure to do so may result in a loss of points.

Read the paper by Marneweck, Jurgens, and Shrader (2017) available at https://royalsocietypublishing.org/doi/10.1098/rspb.2016.2376 and posted on D2L. Their data set is posted in the Excel file "BEHAVIOUR DATA.xlsx". We will be focusing on the results related to sniffing time, which are in the "Sniff duration" sheet. Code is provided below to read the data set in directly from that sheet in the Excel file.

Citation:

- Marneweck, C., Jurgens, A., and Shrader, A. (2017) Dung odours signal sex, age, territorial and oestrous state in white rhinos. *Proc. R. Soc. B.* 284:20162376. https://royalsocietypublishing.org/doi/10.1098/rspb.2016.2376

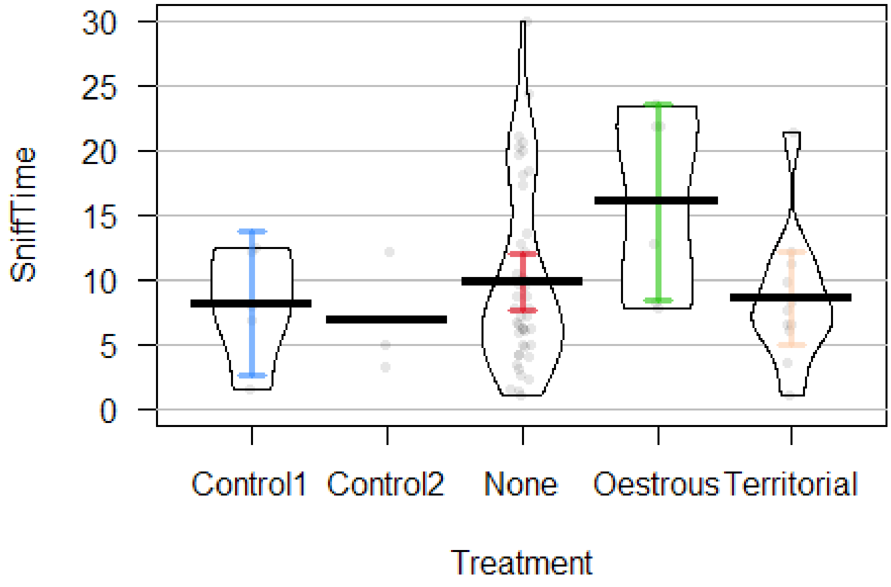
Note that your results will not match those of the authors. This is partially because of a difference in methods that they used and what we will do, but the means in your plots will also not match the results in the paper. They did not provide their code/work in a reproducible fashion so we are not able to trace the differences. It is possible the data set they posted differed from the one used in the analysis slightly.

Also note that the levels in `Treatment` of "Control1" and "Control2" correspond to "*We used two controls. First, the grass bolus (see below) soaked in water in case any VOCs were released by the addition of a liquid. Second, common plant-based herbivore dung odours (table 1), to test if the white rhinos would respond to any new, novel odour within the midden.*" (page 3 of the paper) and that the "None" level in the data set is what they labeled as their "Baseline" category in their results and plots.

1) Based on their Table 2, how many territorial males were included in the study? Briefly explain why the counts vary for the "number of treatment exposures" for each animal.

There were 10 territorial males included in the study. Because this was a behavioral study and the VOC's weren't effective after 3 days, the rhinos sometimes didn't visit the middens within the timeframe required and were not considered exposed.

```
library(readxl) #you'll need to install the readxl package

BEHAVIOUR_DATA <- read_excel("BEHAVIOUR DATA.xlsx",
    sheet = "Sniff duration")
BEHAVIOUR_DATA$Treatment <- factor(BEHAVIOUR_DATA$Treatment)
names(BEHAVIOUR_DATA)[3] <- "SniffTime"

library(yarrr)
pirateplot(SniffTime~Treatment, data=BEHAVIOUR_DATA, inf.method="ci",
inf.disp="line",
           pal="southpark", theme=2)
```

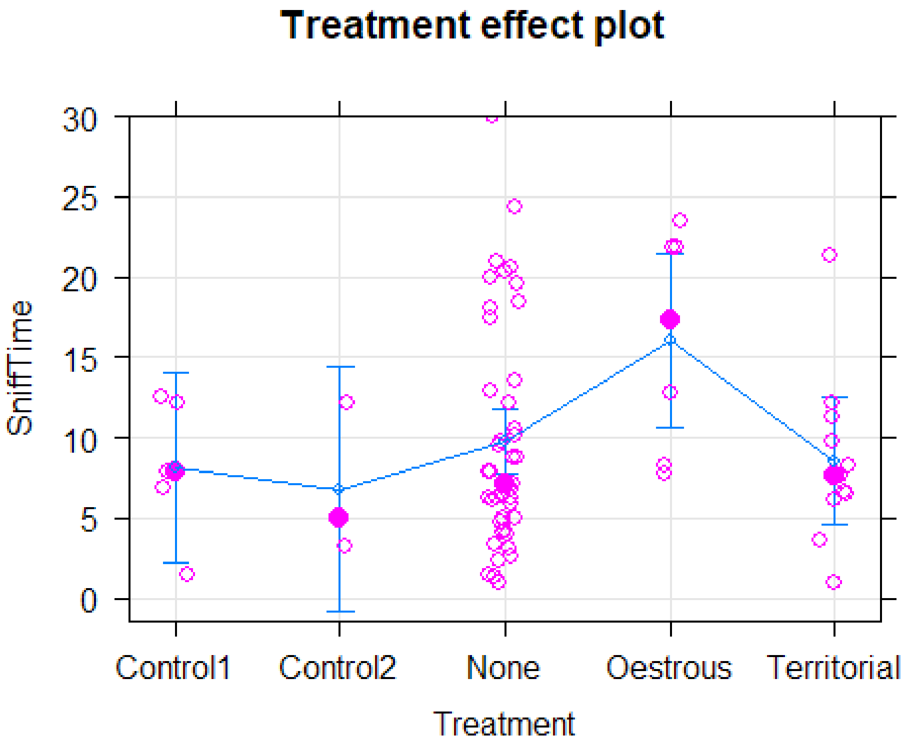


```
library(mosaic)
favstats(SniffTime~Treatment, data=BEHAVIOUR_DATA)

##     Treatment  min     Q1 median     Q3   max      mean       sd  n
missing
## 1    Control1 1.50 6.8800  7.880 12.120 12.51  8.178000 4.490960  5
0
## 2    Control2 3.25 4.1250  5.000  8.570 12.14  6.796667 4.709462  3
0
## 3        None 1.00 4.8750  7.120 13.185 30.00  9.820930 7.064641 43
0
```

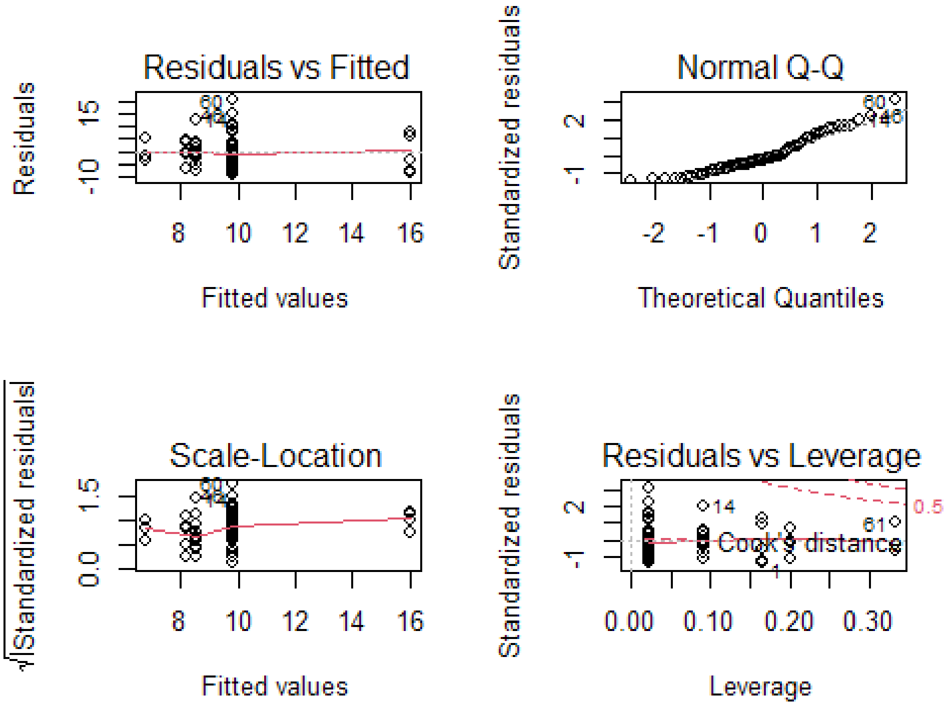
```
## 4    Oestrous 7.75 9.3675 17.315 21.880 23.50 16.000000 7.269682  6
0
## 5 Territorial 1.00 6.3150  7.620 10.495 21.38  8.568182 5.312396 11
0
```

2) For `lm1`, use the provided results to assess the equal variance and normality assumptions. Note that a histogram of the residuals is also provided to complement our regular diagnostic plot array and you should also consider it. Reference specific plots that you are using to make your assessments.

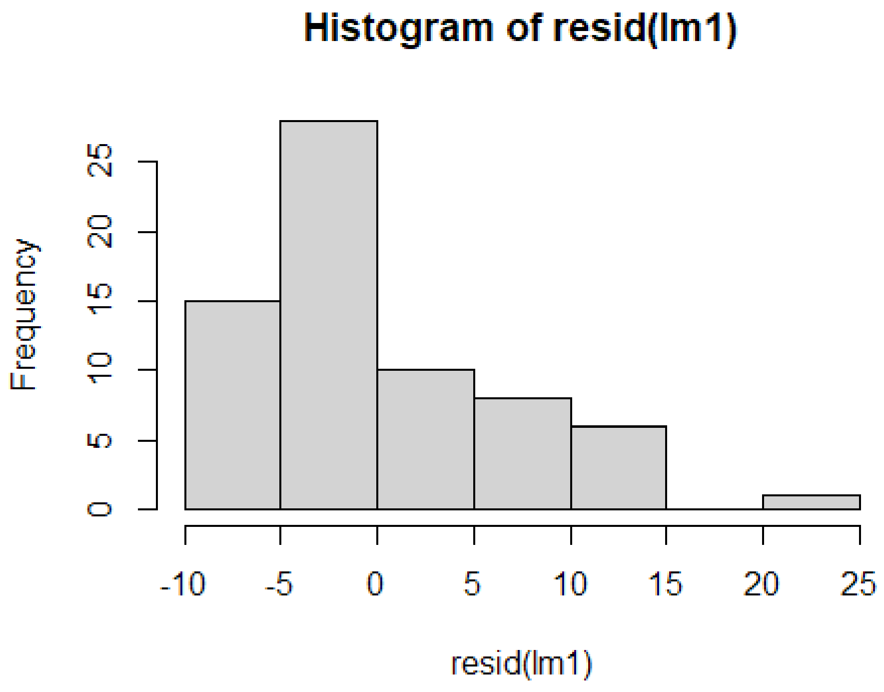
```
lm1 <- lm(SniffTime~Treatment, data=BEHAVIOUR_DATA)

library(car)
library(effects)
plot(allEffects(lm1, residuals=T), grid=T)
```



```
par(mfrow=c(2,2))
plot(lm1)
```



```r
par(mfrow=c(1,1))
hist(resid(lm1))
```



The Treatment effect plot shows shows a lot of unevenness in the residuals and the Residuals vs Fitted plot shows some large funneling and tunneling. The Scale-Location plot also has some evenness. This all shows fairly strong evidence against the assumption of equal variance.

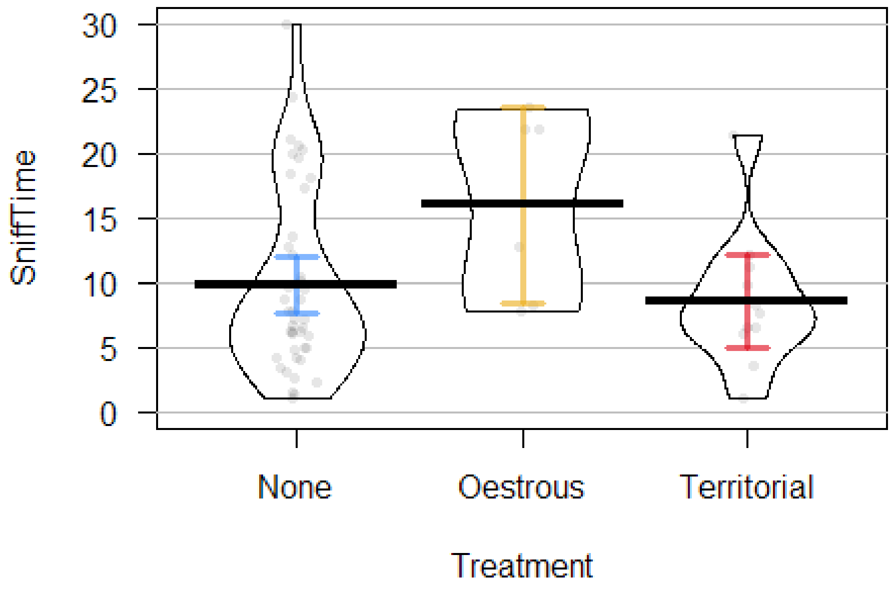
The Normal Q-Q plot shows that the data is slightly right skewed, but the histogram of residuals gives stronger evidence that the data is right skewed. I would say there is moderate evidence against the assumption of normality.

3) Use and provide `R` code to subset the data set to provide the three groups as in their Figure 2d (page 6 of the paper). Store these subsetted data in `BEHAVIOUR_DATA_SUB`. Recall the group they call "baseline" is "None" in the data set. Hint: Review the example in Chapter 2 on page 30 of the text or page 6 of the Week 1 Notes (Day 2).

```
BEHAVIOUR_DATA_SUB <- subset(BEHAVIOUR_DATA, Treatment %in% c("Territorial",
"Oestrous","None"))
BEHAVIOUR_DATA_SUB$Treatment <- factor(BEHAVIOUR_DATA_SUB$Treatment)

#favstats(SniffTime~Treatment, data=BEHAVIOUR_DATA_SUB)
```

4) Make a new pirate-plot with just the three groups. No discussion is necessary, just a plot for the three groups of interest is needed (be sure to include the appropriate `inf.method` and `inf.display` arguments, see code in question 1.)

```
pirateplot(SniffTime~Treatment, data=BEHAVIOUR_DATA_SUB, inf.method="ci",
inf.disp="line",
           pal="southpark", theme=2)
```



5) Fit a reference-coded model with the new subsetted data set using `Treatment` as an explanatory variable. Provide a summary of your fitted model. See code on page 89 of the text or page 20 in the Week 2 Notes (Day 6).

```
lm1 <- lm(SniffTime~Treatment, data=BEHAVIOUR_DATA_SUB)
summary(lm1)

##
## Call:
## lm(formula = SniffTime ~ Treatment, data = BEHAVIOUR_DATA_SUB)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.821 -4.853 -1.951  3.599 20.179
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)             9.821      1.038   9.458 2.78e-13
## TreatmentOestrous       6.179      2.967   2.082   0.0418
## TreatmentTerritorial   -1.253      2.301  -0.545   0.5882
##
## Residual standard error: 6.809 on 57 degrees of freedom
## Multiple R-squared:  0.0824, Adjusted R-squared:  0.0502
## F-statistic: 2.559 on 2 and 57 DF,  p-value: 0.08623
```

6) Use the estimated model coefficients from your summary output to re-construct the estimated mean for the "Oestrous" group. Show your work. How does your calculated value compare to the result in Figure 2d in the paper (page 6)?

The estimated mean for the Oestrous group can be found by $\alpha + \tau_o = 9.821 + 6.179 = 16.000$ **My calculated value is a bit higher than the result in the paper. In the paper, it appears that the result is about 13. However, my calculated value is within the range of standard error in the paper.**

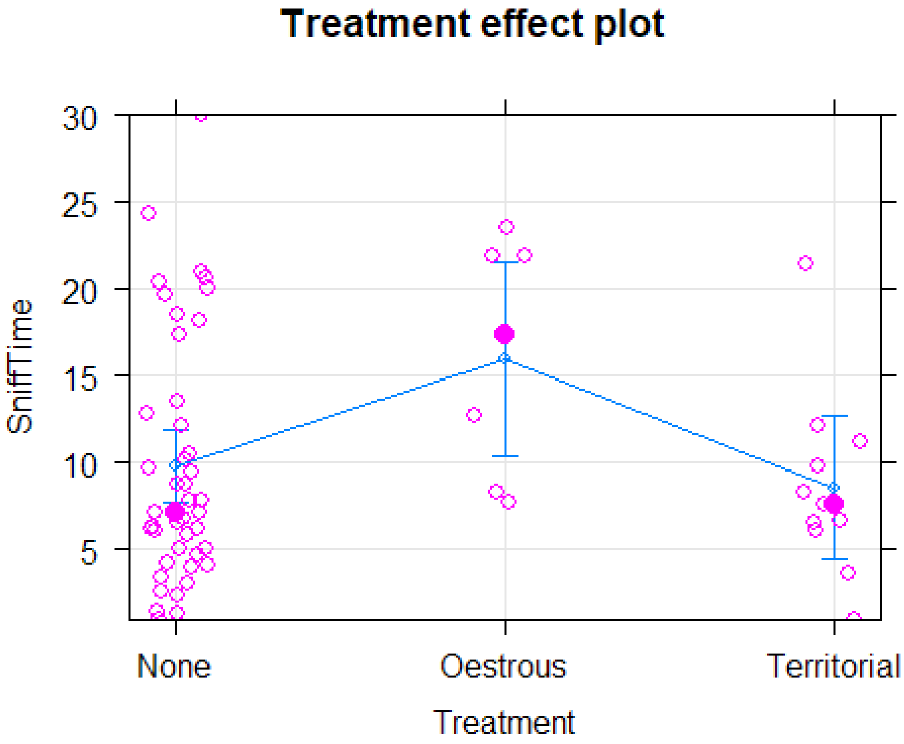
7) Generate an ANOVA table for this model and write a conclusion for `Treatment` in context. See page 93 of the text or page 7 of the Week 3 Notes (Day 8).

```
anova(lm1)

## Analysis of Variance Table
##
## Response: SniffTime
##           Df Sum Sq Mean Sq F value  Pr(>F)
## Treatment  2  237.3 118.652  2.5593 0.08623
## Residuals 57 2642.6  46.362
```

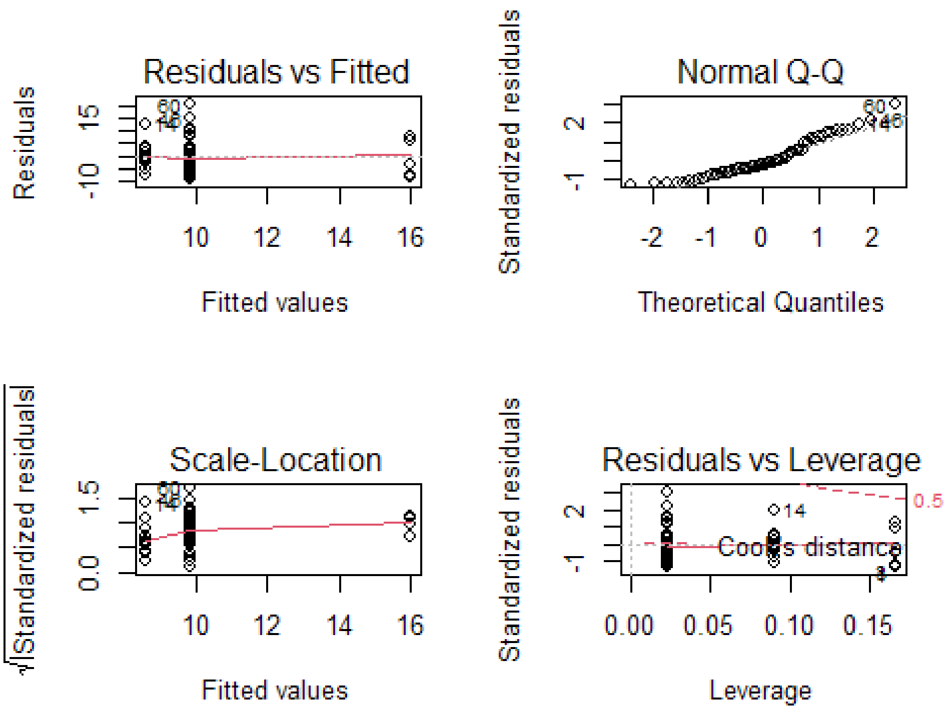
We have moderate to weak evidence against the null that the true mean sniff time among the three treatment types is the same (F(2,57) = 2.5593 with a p-value of 0.0862). Thus it is possible but unlikely that all treatment types yield the same average sniff time.

8) The authors state "As data were not normally distributed, we analyzed them using non-parametric…". Make a set of diagnostic plots for your new model with three groups (see code in question 2). Do you agree with their choice to perform non-parametric testing? Clearly state whether your agree or not and explain your reasoning by referencing your diagnostic plots.

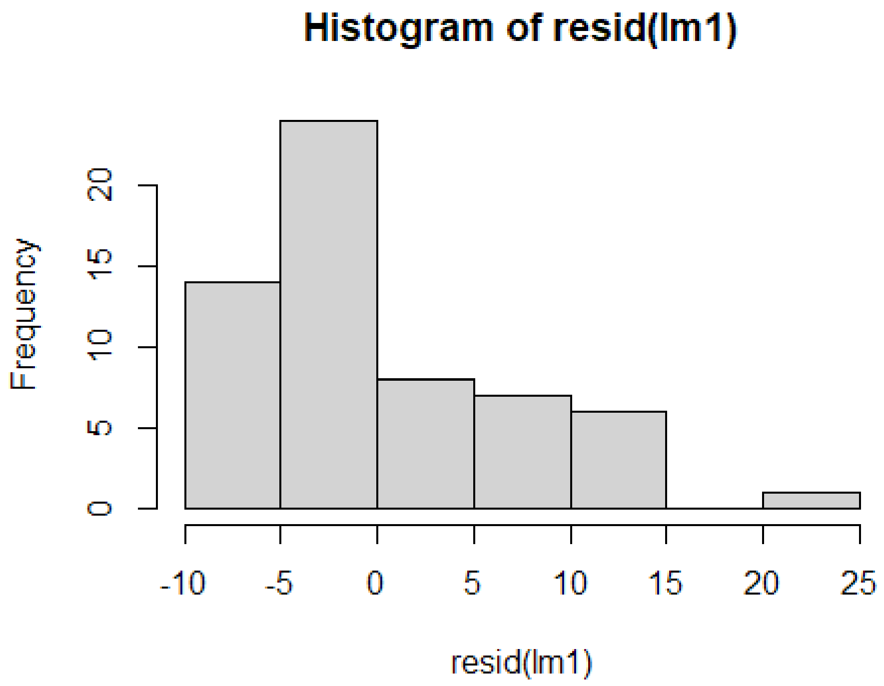
```
plot(allEffects(lm1, residuals=T), grid=T)
```



```
par(mfrow=c(2,2))
plot(lm1)
```



```
par(mfrow=c(1,1))
hist(resid(lm1))
```



I agree with the authors that non-parametric methods should be used. The RvF plot shows some funneling, the Scale-location plot is not very straight, and the effect plot shows an unequal spread of residuals, which gives moderate evidence against the assumption of equal variance. Additionally, the Normal Q-Q plot shows that the data is a bit right-skewed and the histogram of residuals a larger amount of right-skew, which gives strong evidence against the assumption of normality. The violation of these assumptions means we should use non-parametric methods.

9) Read the paper carefully and assess the independence assumption for these observed sniffing times. Make sure you consider their study design in answering this question and explain your reasoning completely.

The authors seem to have done a very good job maintaining independence of observation. They did every test during similar weather conditions, they used isolated rhinos, they used multiple control elements, and they randomized which treatment each rhino got each time. They also used very unobtrusive methods to record the rhinos' reactions. There is weak evidence against the assumption of independence.

10) Write a scope of inference for these results in the context of the study.

The authors randomly selected which treatment to use, but they were very careful about which rhinos they tested on. They only tested on territorial male rhinos. We can say that the treatment variable had a causal relationship with the male population of male territorial white rhinos.

11) Please list anyone that you worked on any part of this Project with. For tutors, give their name and affiliation (MLC, SmartyCats, etc.). For fellow students, give their name and section number. If you did not utilize any outside resources, simply enter "NA."

NA