Final Report

Iman Byndloss, Leah Li, and Rachel Williams

12/09/2024

Contents

Contents

List of Tables

List of Figures

Rationale and Research Questions

The upcoming analyses seek to answer the following research questions: Does the overall activity (daily

To describe the motivation behind this research, we are interested in understanding how different facto

\newpage

Dataset Information

The datasets we used were collected from wildlife insights and the Nashville open data portal (https://

``` r
names(Daily_Species_Counts)

## [1] "Date"          "CountRaw"      "CountAdjusted" "Species"
## [5] "ParkName"
```

```r
summary(Daily_Species_Counts)
```

```
##          Date          CountRaw        CountAdjusted
##  2021-07-20:   37   Min.   : 0.0000   Min.   : 0.0000
##  2021-07-21:   37   1st Qu.: 0.0000   1st Qu.: 0.0000
##  2021-07-22:   37   Median : 0.0000   Median : 0.0000
##  2021-07-23:   37   Mean   : 0.4754   Mean   : 0.1611
##  2021-07-24:   37   3rd Qu.: 0.0000   3rd Qu.: 0.0000
##  2021-07-25:   37   Max.   :43.0000   Max.   :21.5000
##  (Other)   :10064
##                       Species                     ParkName
##  Coyote                   : 834   Alvin G. Beaman Park:3892
##  Eastern Chipmunk         : 834   Bells Bend Park     :3336
##  Eastern Cottontail       : 834   Mill Creek Park     :3058
##  Eastern Gray Squirrel    : 834
##  North American River Otter: 834
##  Northern Raccoon         : 834
##  (Other)                  :5282
```

Exploratory Analysis

To answer our second question, we needed to know which species were most abundant at the target parks. To do this, we aggregated the data to add up the adjusted counts for each species. Once we did this, we sorted it in descending order and looked at which species were the two most abundant.

```
aggregated_data <- Daily_Species_Counts %>%
  group_by(Species) %>%
  summarise(total_count = sum(CountAdjusted, na.rm = TRUE))

sorted_data <- aggregated_data %>%
  arrange(desc(total_count))

most_abundant_species <- head(sorted_data, 5)

kable(most_abundant_species, caption = "Top 5 Most Abundant Species")
```

Table 1: Top 5 Most Abundant Species

Species	total_count
Eastern Gray Squirrel	618.43333
White-tailed Deer	523.21667
Northern Raccoon	225.73333
Nine-banded Armadillo	84.06667
Virginia Opossum	71.66667

According to these results, White-tailed deer and Eastern Gray Squirrels were the most abundant species.

Analysis

<Question 1: The first question we wanted to answer is whether daily and monthly species activity differs between the target parks. We looked at Beaman, Bell's Bend, and Mill's Creek Park.>

Does species' daily and monthly activity differ between parks?

#Daily Analysis

```
# Perform the two-way ANOVA with interaction
daily_anova_result <- aov(CountAdjusted ~ ParkName * Species, data = Daily_Species_Counts)

# View the ANOVA summary
summary(daily_anova_result)
```

```
##                      Df Sum Sq Mean Sq F value Pr(>F)
## ParkName              2   55.9   27.96   99.91 <2e-16 ***
## Species              15  599.6   39.97  142.86 <2e-16 ***
## ParkName:Species     19  676.2   35.59  127.19 <2e-16 ***
## Residuals         10249 2867.8    0.28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# If the interaction term is significant, perform a post-hoc test
Daily_Species_Counts$ParkName <- as.factor(Daily_Species_Counts$ParkName)
Daily_Species_Counts$Species <- as.factor(Daily_Species_Counts$Species)

# Tukey's HSD test for the interaction term (ParkName * Species)
library(agricolae)
daily_tukey_result <- TukeyHSD(daily_anova_result, "ParkName:Species")

# Extract significant results
daily_tukey_df <- as.data.frame(daily_tukey_result$`ParkName:Species`)
daily_tukey_df$Comparison <- rownames(daily_tukey_result$`ParkName:Species`)
rownames(daily_tukey_df) <- NULL

# Filter significant comparisons
daily_significant_tukey <- subset(daily_tukey_df, `p adj` < 0.05)

# Print the top 10 significant results
head(daily_significant_tukey, 10)
```

```
##          diff         lwr       upr        p adj
## 61  0.2113309 0.033001641 0.3896602 2.405810e-03
## 62  0.1829137 0.004584374 0.3612430 3.447013e-02
## 63  1.8195444 1.641215070 1.9978737 0.000000e+00
## 68  0.2880096 0.109680298 0.4663389 1.598396e-07
## 74  0.2181655 0.039836173 0.3964948 1.168225e-03
## 75  0.4247602 0.246430897 0.6030895 0.000000e+00
## 88  0.3705036 0.192174303 0.5488329 0.000000e+00
## 89  1.3284173 1.150087972 1.5067466 0.000000e+00
## 106 0.2137290 0.035399722 0.3920583 1.873439e-03
## 107 0.1853118 0.006982456 0.3636410 2.818104e-02
##                                                                 Comparison
```

```
## 61  Alvin G. Beaman Park:Eastern Gray Squirrel-Bells Bend Park:American Beaver
## 62       Bells Bend Park:Eastern Gray Squirrel-Bells Bend Park:American Beaver
## 63       Mill Creek Park:Eastern Gray Squirrel-Bells Bend Park:American Beaver
## 68       Bells Bend Park:Nine-banded Armadillo-Bells Bend Park:American Beaver
## 74            Bells Bend Park:Northern Raccoon-Bells Bend Park:American Beaver
## 75            Mill Creek Park:Northern Raccoon-Bells Bend Park:American Beaver
## 88      Alvin G. Beaman Park:White-tailed Deer-Bells Bend Park:American Beaver
## 89           Bells Bend Park:White-tailed Deer-Bells Bend Park:American Beaver
## 106 Alvin G. Beaman Park:Eastern Gray Squirrel-Mill Creek Park:American Beaver
## 107      Bells Bend Park:Eastern Gray Squirrel-Mill Creek Park:American Beaver
```

#Monthly Analysis

```
# Perform the two-way ANOVA with interaction
monthly_anova_result <- aov(CountAdjusted ~ ParkName * Species, data = Monthly_Species_Counts)

# View the ANOVA summary
summary(monthly_anova_result)
```

```
##                   Df Sum Sq Mean Sq F value   Pr(>F)
## ParkName           2   1554   777.2   11.81 1.11e-05 ***
## Species           15  16669  1111.3   16.88  < 2e-16 ***
## ParkName:Species  19  18798   989.4   15.03  < 2e-16 ***
## Residuals        333  21922    65.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# If the interaction term is significant, perform a post-hoc test
Monthly_Species_Counts$ParkName <- as.factor(Monthly_Species_Counts$ParkName)
Monthly_Species_Counts$Species <- as.factor(Monthly_Species_Counts$Species)

# Tukey's HSD test for the interaction term (ParkName * Species)
library(agricolae)
monthly_tukey_result <- TukeyHSD(monthly_anova_result, "ParkName:Species")

# Extract significant results
monthly_tukey_df <- as.data.frame(monthly_tukey_result$`ParkName:Species`)
monthly_tukey_df$Comparison <- rownames(monthly_tukey_result$`ParkName:Species`)
rownames(monthly_tukey_df) <- NULL

# Filter significant comparisons
monthly_significant_tukey <- subset(monthly_tukey_df, `p adj` < 0.05)

# Print the top 10 significant results
head(monthly_significant_tukey, 10)
```

```
##         diff      lwr      upr        p adj
## 63  50.58333 36.02293 65.14373 3.911316e-13
## 89  36.93000 22.36960 51.49040 8.375522e-13
## 108 50.65000 36.08960 65.21040 3.911316e-13
## 134 36.99667 22.43627 51.55707 8.519851e-13
## 152 50.47500 35.91460 65.03540 3.911316e-13
## 178 36.82167 22.26127 51.38207 8.308909e-13
```

```
## 195 47.29833 32.73793 61.85873 3.911316e-13
## 221 33.64500 19.08460 48.20540 8.526513e-13
## 278 49.69167 35.13127 64.25207 3.911316e-13
## 304 36.03833 21.47793 50.59873 8.282264e-13
##                                                                     Comparison
## 63  Mill Creek Park:Eastern Gray Squirrel-Bells Bend Park:American Beaver
## 89      Bells Bend Park:White-tailed Deer-Bells Bend Park:American Beaver
## 108 Mill Creek Park:Eastern Gray Squirrel-Mill Creek Park:American Beaver
## 134     Bells Bend Park:White-tailed Deer-Mill Creek Park:American Beaver
## 152     Mill Creek Park:Eastern Gray Squirrel-Alvin G. Beaman Park:Bobcat
## 178         Bells Bend Park:White-tailed Deer-Alvin G. Beaman Park:Bobcat
## 195          Mill Creek Park:Eastern Gray Squirrel-Bells Bend Park:Bobcat
## 221              Bells Bend Park:White-tailed Deer-Bells Bend Park:Bobcat
## 278     Mill Creek Park:Eastern Gray Squirrel-Alvin G. Beaman Park:Coyote
## 304         Bells Bend Park:White-tailed Deer-Alvin G. Beaman Park:Coyote
```

Question 2: The second question we wanted to answer is what factors might be causing different activity levels between parks. This study focused on the size, classification, and disruptive activities of each park. We used Deer and squirrels for this analysis because they were the most abundant species at all three parks.

What factors might be causing different activity levels between parks for WHite-tailed Deer and Grey Squirrels? Does classification, size, or disruptive activities at each park cause changes in activity levels?

#Creating Daily and Monthly merged sets for Deer and Squirrels

```r
dmSquirrel <- Daily_Merge %>%
  filter(Species == "Eastern Gray Squirrel")

dmDeer <- Daily_Merge %>%
  filter(Species == "White-tailed Deer")

mmSquirrel <- Monthly_Merge %>%
  filter(Species == "Eastern Gray Squirrel")

mmDeer <- Monthly_Merge %>%
  filter(Species == "White-tailed Deer")
```

#Daily t-tests

```r
#Classification t-tests:
SQ_t_test_class <- t.test(CountAdjusted ~ Classification, data = dmSquirrel)
SQ_t_test_class
```

```
##
##  Welch Two Sample t-test
##
## data:  CountAdjusted by Classification
## t = 14.034, df = 286.34, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group Community and group Regional is not e
## 95 percent confidence interval:
##  1.394874 1.849970
## sample estimates:
## mean in group Community  mean in group Regional
##               1.8231415               0.2007194
```

```r
D_t_test_class <- t.test(CountAdjusted ~ Classification, data = dmDeer)
D_t_test_class
```

```
##
##  Welch Two Sample t-test
##
## data:  CountAdjusted by Classification
## t = -8.5867, df = 822.47, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group Community and group Regional is not e
## 95 percent confidence interval:
##  -0.8318784 -0.5223183
## sample estimates:
## mean in group Community  mean in group Regional
##               0.1759592               0.8530576
```

```r
#Park Size t-tests
SQ_t_test_size <- t.test(CountAdjusted ~ Size, data = dmSquirrel)
SQ_t_test_size
```

```
##
##  Welch Two Sample t-test
##
## data:  CountAdjusted by Size
## t = 11.149, df = 649.39, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group <1000ac and group >1000ac is not equal
## 95 percent confidence interval:
##  0.6507735 0.9290227
## sample estimates:
## mean in group <1000ac mean in group >1000ac
##             1.0048261             0.2149281
```

```r
D_t_test_size <- t.test(CountAdjusted ~ Size, data = dmDeer)
D_t_test_size
```

```
##
##  Welch Two Sample t-test
##
## data:  CountAdjusted by Size
## t = 5.0213, df = 811.35, p-value = 6.312e-07
## alternative hypothesis: true difference in means between group <1000ac and group >1000ac is not equal
## 95 percent confidence interval:
##  0.2313844 0.5283877
## sample estimates:
## mean in group <1000ac mean in group >1000ac
##             0.7539868             0.3741007
```

```r
#Disruption t-tests
SQ_t_test_disrupt <- t.test(CountAdjusted ~ DisruptiveActivities, data = dmSquirrel)
SQ_t_test_disrupt
```

```
##
```

```
##  Welch Two Sample t-test
##
## data:  CountAdjusted by DisruptiveActivities
## t = 11.756, df = 660.68, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group No and group Yes is not equal to 0
## 95 percent confidence interval:
##  0.6934655 0.9715825
## sample estimates:
##  mean in group No mean in group Yes
##         1.0190348         0.1865108
```

```
D_t_test_disrupt <- t.test(CountAdjusted ~ DisruptiveActivities, data = dmDeer)
D_t_test_disrupt
```

```
##
##  Welch Two Sample t-test
##
## data:  CountAdjusted by DisruptiveActivities
## t = -9.1974, df = 322.21, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group No and group Yes is not equal to 0
## 95 percent confidence interval:
##  -1.2830766 -0.8308922
## sample estimates:
##  mean in group No mean in group Yes
##          0.275030          1.332014
```

#Monthly t-tests

```
#Classification t-test
SQ_t_test_class2 <- t.test(CountAdjusted ~ Classification, data = mmSquirrel)
SQ_t_test_class2
```

```
##
##  Welch Two Sample t-test
##
## data:  CountAdjusted by Classification
## t = 3.7485, df = 9.0843, p-value = 0.004491
## alternative hypothesis: true difference in means between group Community and group Regional is not e
## 95 percent confidence interval:
##  17.92248 72.28418
## sample estimates:
## mean in group Community  mean in group Regional
##                50.68333                 5.58000
```

```
D_t_test_class2 <- t.test(CountAdjusted ~ Classification, data = mmDeer)
D_t_test_class2
```

```
##
##  Welch Two Sample t-test
##
## data:  CountAdjusted by Classification
## t = -3.7042, df = 25.158, p-value = 0.001046
```

```
## alternative hypothesis: true difference in means between group Community and group Regional is not e
## 95 percent confidence interval:
##  -29.285658  -8.361009
## sample estimates:
## mean in group Community  mean in group Regional
##                4.891667               23.715000
```

```
#Park size t-tests
SQ_t_test_size2 <- t.test(CountAdjusted ~ Size, data = mmSquirrel)
SQ_t_test_size2
```

```
##
##  Welch Two Sample t-test
##
## data:  CountAdjusted by Size
## t = 2.7673, df = 19.745, p-value = 0.01198
## alternative hypothesis: true difference in means between group <1000ac and group >1000ac is not equa
## 95 percent confidence interval:
##   5.392609 38.525724
## sample estimates:
## mean in group <1000ac mean in group >1000ac
##              27.93417               5.97500
```

```
D_t_test_size2 <- t.test(CountAdjusted ~ Size, data = mmDeer)
D_t_test_size2
```

```
##
##  Welch Two Sample t-test
##
## data:  CountAdjusted by Size
## t = 1.8233, df = 27.553, p-value = 0.07912
## alternative hypothesis: true difference in means between group <1000ac and group >1000ac is not equa
## 95 percent confidence interval:
##  -1.312252 22.433918
## sample estimates:
## mean in group <1000ac mean in group >1000ac
##              20.96083              10.40000
```

```
#Disruption t-tests
SQ_t_test_disrupt2 <- t.test(CountAdjusted ~ DisruptiveActivities, data = mmSquirrel)
SQ_t_test_disrupt2
```

```
##
##  Welch Two Sample t-test
##
## data:  CountAdjusted by DisruptiveActivities
## t = 2.932, df = 19.977, p-value = 0.008249
## alternative hypothesis: true difference in means between group No and group Yes is not equal to 0
## 95 percent confidence interval:
##   6.677357 39.610976
## sample estimates:
##  mean in group No mean in group Yes
##          28.32917           5.18500
```

```r
D_t_test_disrupt2 <- t.test(CountAdjusted ~ DisruptiveActivities, data = mmDeer)
D_t_test_disrupt2
```

```
##
##  Welch Two Sample t-test
##
## data:  CountAdjusted by DisruptiveActivities
## t = -4.3139, df = 10.489, p-value = 0.001368
## alternative hypothesis: true difference in means between group No and group Yes is not equal to 0
## 95 percent confidence interval:
##  -44.46569 -14.30264
## sample estimates:
##  mean in group No mean in group Yes
##          7.645833         37.030000
```

Summary and Conclusions

References

<add references here if relevant, otherwise delete this section>