Lab Exercises Week 08

Alison Lawyer

2024-11-20

```r
# keep this chunk in all your RMarkdown scripts
knitr::opts_chunk$set(echo = TRUE)
knitr::opts_chunk$set(tidy.opts = list(width.cutoff = 60), tidy = TRUE)
```

```r
# List required packages
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(openxlsx)
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'
##
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```r
library(ggplot2)
```

LAB EXERCISES

The dataset "vigilance.xlsx" contains data on vigilant behavior of several European song bird species (expressed as a rater per hour). The study aimed to address the question of whether birds show higher vigilance behavior in urban areas with more noise, which could make it more difficult to use auditory cues for avoiding predators. For each species, data were collected at several sites (cities), each having an "urban noise" and "control" condition. Some environmental factors are also part of this dataset but we will ignore those for now.

Exercise 1

Import the data and clean your column names (remove special characters and shorten for easier use in your code). Filter the dataset to European Robin (RB).

```r
# Load the data
vigilance <- read.xlsx("vigilance.xlsx")

# Clean column names
colnames(vigilance) <- sub(pattern = "\\.\\((.*?)\\)", replacement = "",
    colnames(vigilance))

european_robin <- vigilance %>%
    filter(species == "RB")
european_robin
```

```
##    site urbanisation.intensity day temperature wind.speed   treatment
## 1     3              1.9398800  51        9.45       0.80 urban noise
## 2     3              1.9398800  51        9.50       2.27     control
## 3     4              4.2293000  56        8.00       0.00 urban noise
## 4     4              4.2293000  56       10.60       0.00     control
## 5     6              2.5961000  52       14.70       0.80 urban noise
## 6     6              2.5961000  52       15.57       0.80     control
## 7    11              2.7075900  72        6.90       3.90 urban noise
## 8    11              2.7075900  72        7.03       3.50     control
## 9    13              0.0619194  76        7.07       2.50 urban noise
## 10   13              0.0619194  76        6.45       2.50     control
## 11   14              4.4214000  81       13.03       1.10 urban noise
## 12   14              4.4214000  81       13.10       1.90     control
## 13   15              4.4056700  81       11.48      10.88 urban noise
## 14   15              4.4056700  81       12.80      10.65     control
## 15   20              0.5740110  92        7.35       1.60 urban noise
## 16   20              0.5740110  92        8.30       1.60     control
## 17   22             -0.6133470  50        7.80       2.20 urban noise
## 18   22             -0.6133470  50        7.80       1.90     control
## 19   23             -1.4429500  51       10.45       1.70 urban noise
## 20   23             -1.4429500  51       11.10       1.10     control
## 21   25              0.1584160  56       11.53       0.80 urban noise
## 22   25              0.1584160  56       11.28       6.11     control
## 23   27              0.8755660  60        8.40       0.00 urban noise
## 24   27              0.8755660  60        8.77       0.00     control
## 25   29              1.3138900  77        6.98       5.03 urban noise
## 26   29              1.3138900  77        7.93       5.88     control
## 27   31              2.3775300  82        8.08       4.60 urban noise
## 28   31              2.3775300  82        8.30       4.80     control
## 29   32             -0.1811550  81       10.30      13.03 urban noise
## 30   32             -0.1811550  81       11.25      11.53     control
## 31   39             -0.0499581  93        4.00       3.50 urban noise
## 32   39             -0.0499581  93        6.40       1.90     control
## 33   41             -1.4745500  86        9.63       2.18 urban noise
## 34   41             -1.4745500  86        9.87       3.80     control
## 35   44              0.4263110  93        9.60       3.50 urban noise
## 36   44              0.4263110  93        7.85       2.95     control
```

```
## 37   51             -2.1759600  59        6.13       0.20 urban noise
## 38   51             -2.1759600  59        7.15       0.00     control
## 39   52             -2.1285700  67        6.60       6.40 urban noise
## 40   52             -2.1285700  67        6.90       7.27     control
## 41   54             -2.9754700  70        3.95       2.20 urban noise
## 42   54             -2.9754700  70        3.15       3.20     control
## 43   55             -2.9922400  77        7.25       0.00 urban noise
## 44   55             -2.9922400  77        6.55       1.25     control
##    hours.from.sunrise species vigilance.rate
## 1                4.27      RB          30.11
## 2                3.35      RB           0.00
## 3                1.70      RB          23.79
## 4                2.70      RB          28.90
## 5                5.63      RB           4.37
## 6                6.47      RB          28.70
## 7                3.57      RB          21.08
## 8                4.40      RB          25.76
## 9                6.98      RB           0.00
## 10               6.07      RB           6.21
## 11               9.02      RB           0.00
## 12               8.18      RB           0.00
## 13               6.18      RB           0.00
## 14               7.02      RB           0.00
## 15               6.63      RB          37.61
## 16               7.47      RB          20.23
## 17               4.90      RB           0.00
## 18               5.73      RB          11.44
## 19               6.02      RB          13.11
## 20               6.85      RB           0.00
## 21               6.33      RB          26.90
## 22               5.45      RB          33.93
## 23               5.52      RB          19.89
## 24               6.35      RB          21.43
## 25               5.35      RB           2.11
## 26               6.18      RB           0.00
## 27               3.13      RB          23.59
## 28               3.97      RB          26.26
## 29               5.18      RB          27.52
## 30               6.10      RB          20.96
## 31               3.42      RB          20.59
## 32               4.25      RB          40.90
## 33               3.13      RB           7.93
## 34               3.97      RB           6.77
## 35              10.42      RB          23.39
## 36               8.75      RB           3.90
## 37               1.98      RB          20.81
## 38               2.82      RB          24.00
## 39               4.78      RB          20.19
## 40               5.62      RB          17.14
## 41               3.25      RB          39.68
## 42               2.42      RB          16.15
## 43               3.60      RB          16.45
## 44               2.82      RB           6.00
```

Your first task is to conduct exploratory analyses on this species, assessing - measures of central tendency, - variability and - distribution of vigilance rates for control and noise conditions.

At the end, you should have - a summary table with all relevant measures, - a figure comparing the means in both conditions, and - figures to assess the distribution and potential skew.

For your table, make it so that your statistics are in rows and the values for control and noise condition are in columns. This produces a table that is easier to read. Format this table so that it is printed nicely in your markdown (tip: the function kable() from the kableExtra package helps).

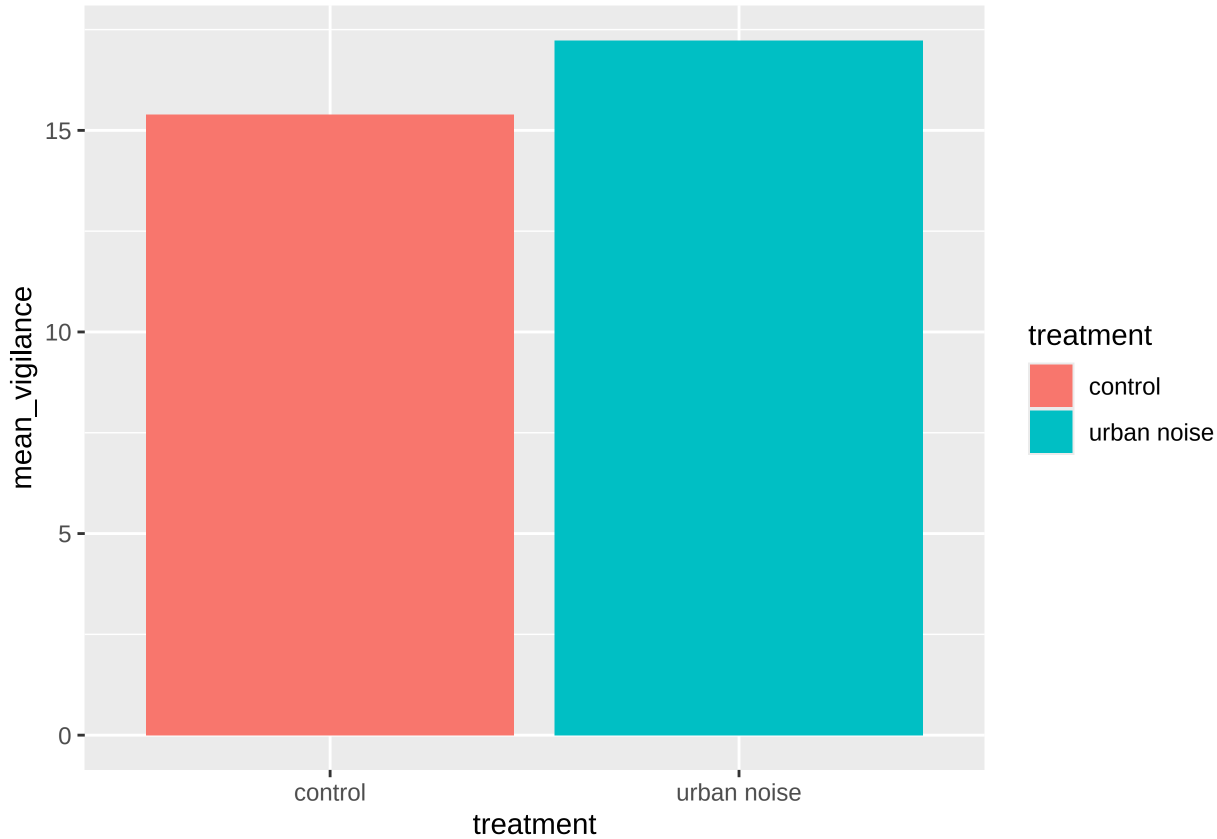
Inspect this table and discuss your findings without any formal hypothesis testing.

```
# Produce summary table
summary_table <- european_robin %>%
    group_by(treatment) %>%
    summarize(mean_vigilance = mean(vigilance.rate), sd_vigilance = sd(vigilance.rate),
        median_vigilance = median(vigilance.rate), iqr = IQR(vigilance.rate),
        var = var(vigilance.rate))

# Make nice table for printing
summarized <- summary_table %>%
    pivot_longer(cols = "mean_vigilance":"var", names_to = "stat",
        values_to = "values") %>%
    pivot_wider(names_from = treatment, values_from = values)

transposed <- t(summary_table[, 2:6])
transposed <- as.data.frame(transposed)
colnames(transposed) <- c("control", "urban noise")

summary <- kable(summarized, caption = "Summary statistics by treatment")

# Print table
summary
```

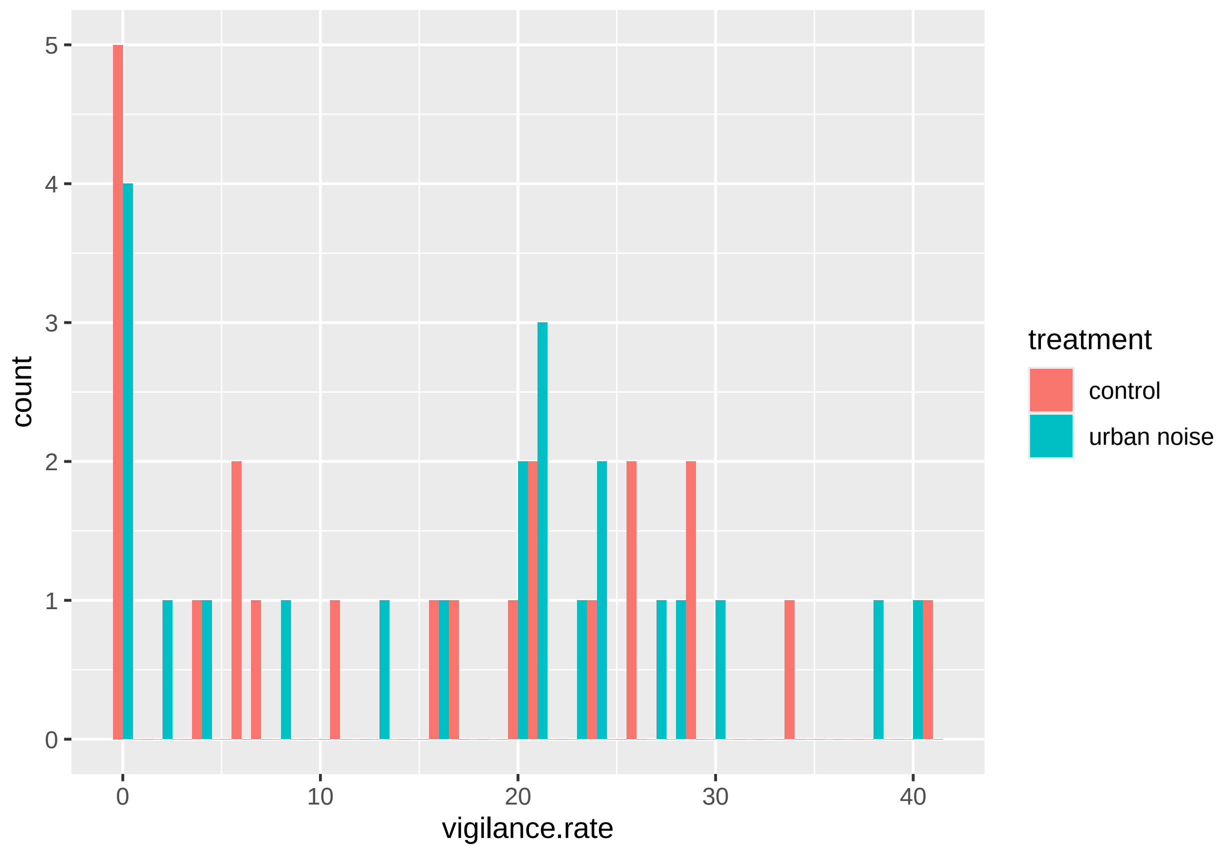
Table 1: Summary statistics by treatment

stat	control	urban noise
mean_vigilance	15.39455	17.23273
sd_vigilance	12.58721	12.24042
median_vigilance	16.64500	20.39000
iqr	20.89500	18.48000
var	158.43798	149.82795

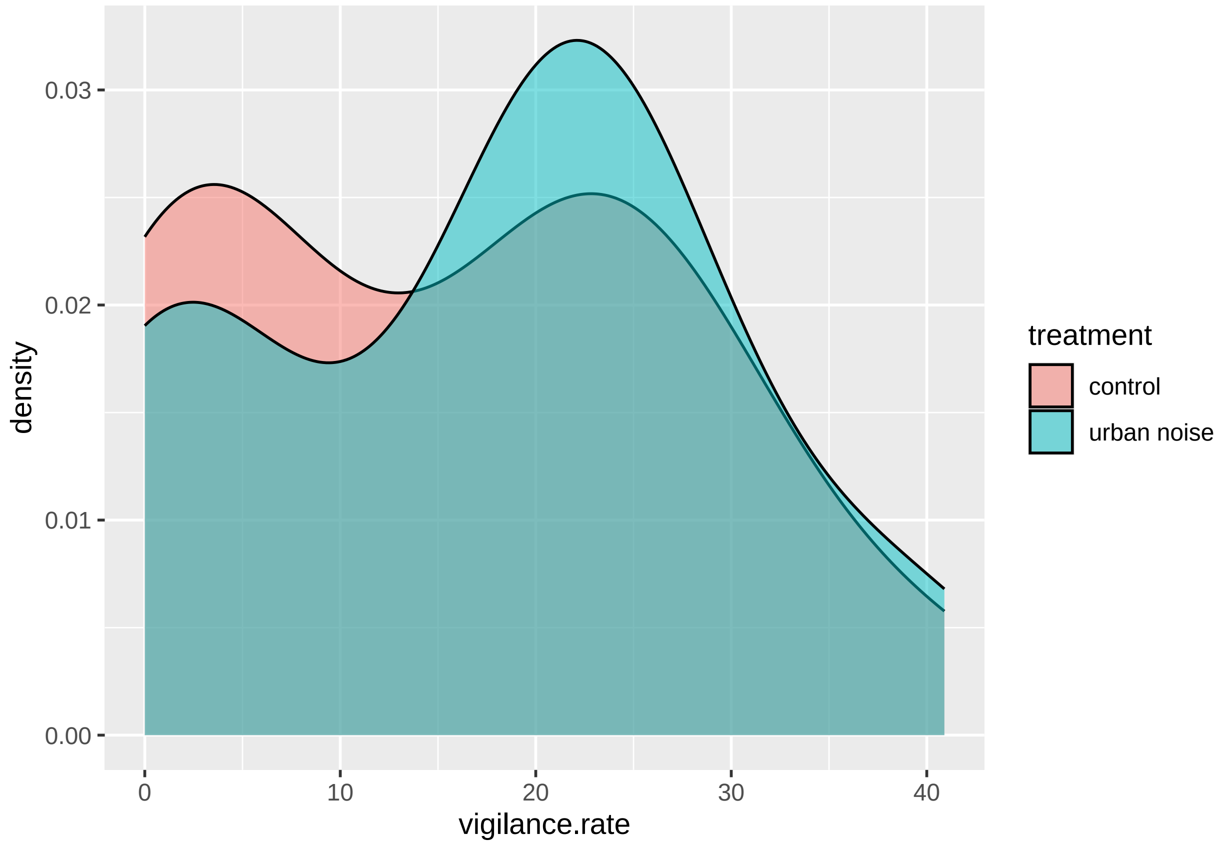
```
# Figure comparing the means
ggplot(summary_table, aes(x = treatment, y = mean_vigilance,
    fill = treatment)) + geom_col()
```



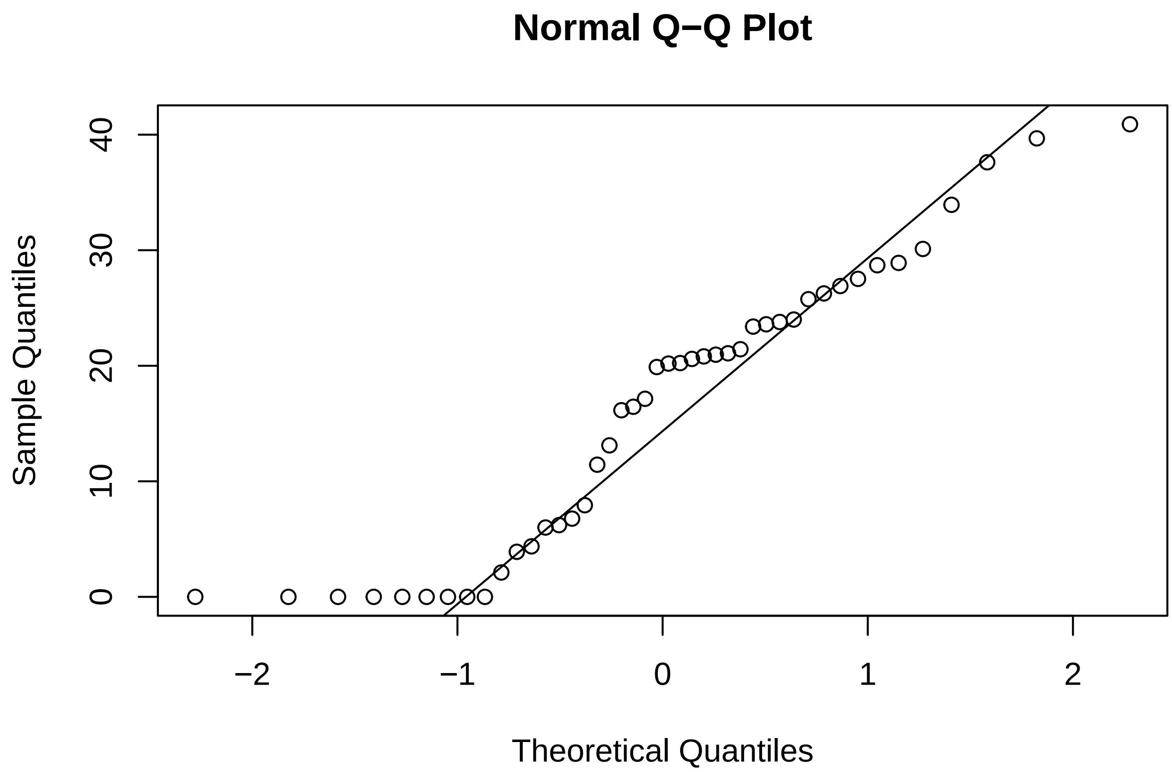

```
# Histograms
ggplot(european_robin, aes(x = vigilance.rate, fill = treatment)) +
    geom_histogram(binwidth = 1, position = "dodge")
```



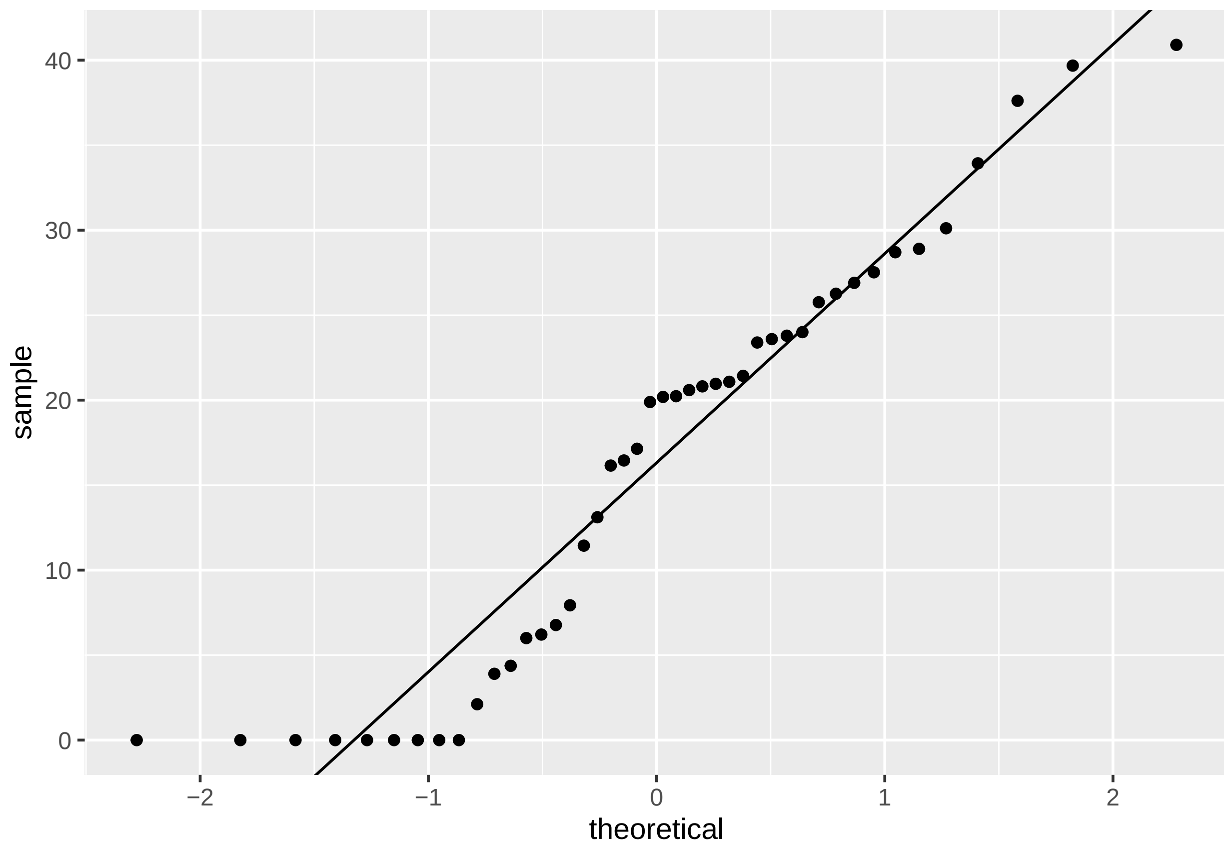

```
ggplot(european_robin, aes(x = vigilance.rate, fill = treatment)) +
    geom_density(alpha = 0.5)
```




```r
# Q-Q plots
qqnorm(european_robin$vigilance.rate)
qqline(european_robin$vigilance.rate)
```




```r
ggplot(european_robin, aes(sample = vigilance.rate)) + stat_qq() +
    geom_abline(aes(intercept = mean(x = vigilance.rate), slope = sd(x = vigilance.rate)))
```



Answer:

Exercise 2

Conduct an appropriate t-test to test the null hypothesis that vigilance rates do not differ between noise and control conditions for the European Robin. Explain your choice of test. Should any assumptions be violated, discuss why but continue to conduct the test anyways for the sake of the exercise. Interpret the output and write a summary statement.

```r
# Conduct the t-test
t_test <- t.test(european_robin$vigilance.rate ~ european_robin$treatment)

# Print the result
t_test
```

```
##
##  Welch Two Sample t-test
##
## data:  european_robin$vigilance.rate by european_robin$treatment
## t = -0.49106, df = 41.967, p-value = 0.6259
## alternative hypothesis: true difference in means between group control and group urban noise is not
## 95 percent confidence interval:
##  -9.392586  5.716222
## sample estimates:
```

```
##     mean in group control mean in group urban noise
##                  15.39455                  17.23273
```

Answer:

Exercise 3

Now compare the difference in vigilance rates between coal tits (CT) and blue tits (BT) across all sites. Do this separately for the control and noise condition. In other words, compare CT and BT in all control sites, and CT and BT in all treatment (noise) sites. To do this, you will need to think carefully about how to arrange your data. You will conduct exploratory analyses for this problem, similar to what you did above, as a homework exercise.

What is the most appropriate t-tests to use for testing your hypothesis? Explain your choice. Interpret your output and write a summary statement of results.

```r
# run a pair sample t-test

# Reshaping the data for the t-tests
control <- vigilance %>%
    filter(treatment == "control", species == "CT" | species ==
        "BT") %>%
    select(site, species, treatment, vigilance.rate) %>%
    pivot_wider(values_from = "vigilance.rate", names_from = c("species",
        "treatment"))

urban <- vigilance %>%
    filter(treatment == "urban noise", species == "CT" | species ==
        "BT") %>%
    select(site, species, treatment, vigilance.rate) %>%
    pivot_wider(values_from = "vigilance.rate", names_from = c("species",
        "treatment"))

# Run the tests control noise
t.test(x = control$CT_control, y = control$BT_control, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  control$CT_control and control$BT_control
## t = -1.3465, df = 24, p-value = 0.1907
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -7.610438  1.600838
## sample estimates:
## mean difference
##         -3.0048
```

```r
# urban noise
t.test(x = urban$`CT_urban noise`, y = urban$`BT_urban noise`,
    paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  urban$`CT_urban noise` and urban$`BT_urban noise`
## t = -3.6289, df = 24, p-value = 0.001338
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -11.047011  -3.036989
## sample estimates:
## mean difference
##          -7.042
```

Answer: