Homework 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 = 20), 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)
```

HOMEWORK EXERCISES

Exercise 1

Conduct exploratory analyses for the question posed in exercise 3 from the lab. The steps and outputs should mirror what you did for exercise 1, but adjusted for the question posed and the kind of analysis done. For example, think through what you are comparing and then decide how best to structure your table of descriptive stats, graphs, and tests of normality. Determine what to use for the visual assessment of normal distribution referring to the lecture and readings for this week. Write a summary statement about your observations, including whether assumptions of the t-test you conducted in exercise 3 were met. Include all pertinent information in your summary statement.

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

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

ct_bt <- vigilance %>%
    filter(species ==
        "CT" | species ==
        "BT")

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")) %>%
    mutate(control_differences = BT_control -
        CT_control)

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")) %>%
    mutate(urban_differences = `BT_urban noise` -
        `CT_urban noise`)

combined <- as.data.frame(c(control %>%
    select(site, control_differences),
    urban %>%
        select(urban_differences))) %>%
    pivot_longer(cols = control_differences:urban_differences,
        names_to = "treatment",
        values_to = "diffs")
```

```r
# Produce summary
# table. Be sure to
# group by species
# and treatment.
summary_table <- ct_bt %>%
    group_by(species,
        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),
        se = sd_vigilance/sqrt(n()),
        CI_low = mean_vigilance -
            1.96 * se,
        CI_high = mean_vigilance +
            1.96 * se)
```

```
## `summarise()` has grouped output by 'species'. You can override using the
## `.groups` argument.
```

```r
summary_table
```

```
## # A tibble: 4 x 10
## # Groups:   species [2]
##   species treatment   mean_vigilance sd_vigilance median_vigilance   iqr   var
##   <chr>   <chr>                <dbl>        <dbl>            <dbl> <dbl> <dbl>
## 1 BT      control              11.9          5.83            11.3   6.65  34.0
## 2 BT      urban noise          15.2          7.48            14.4   7.56  55.9
## 3 CT      control               8.38         8.87             7.14  7.25  78.6
## 4 CT      urban noise           9.42         7.00             9.14  9.10  49.0
## # i 3 more variables: se <dbl>, CI_low <dbl>, CI_high <dbl>
```

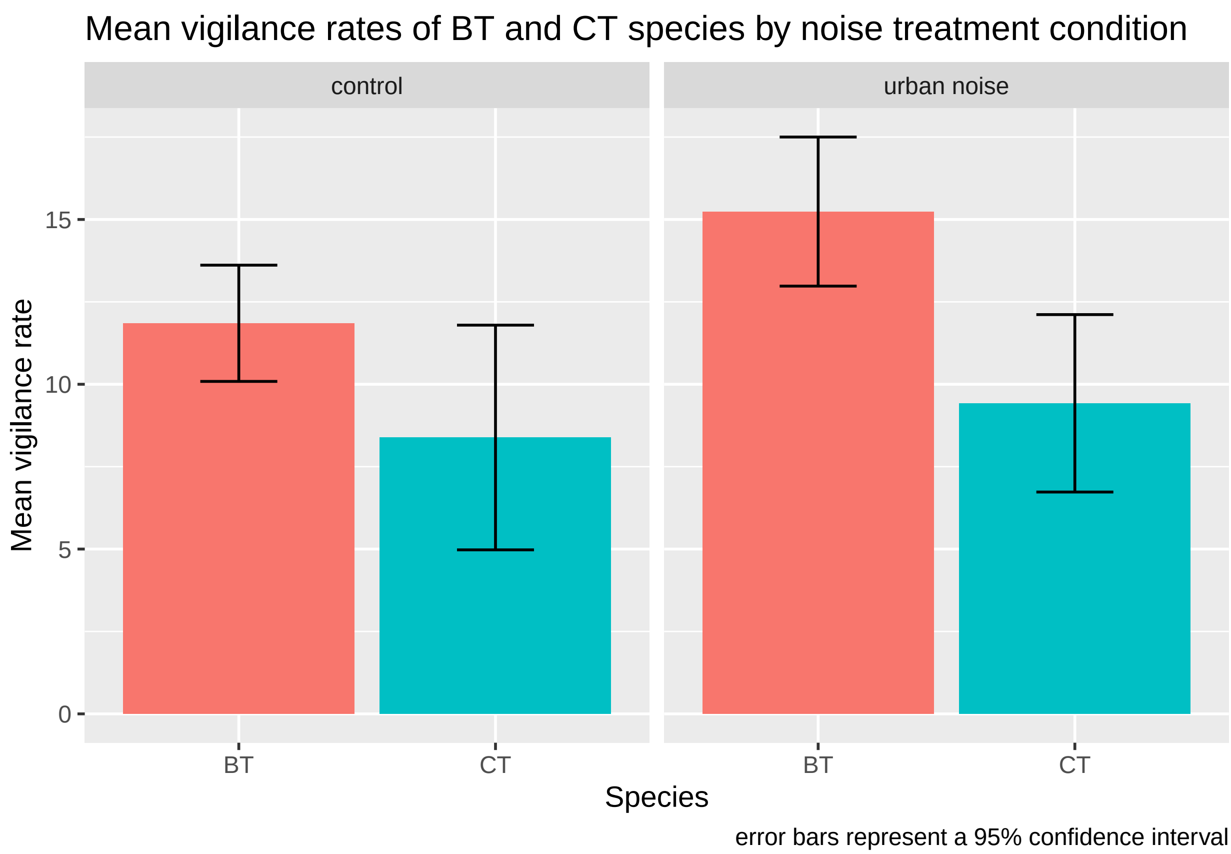
```r
# Make nice table
# for printing.
# Note that your
# primary goal is
# to compare
# species within
# each condition,
# which is
# different from
# Exercise 1. Think
# about how to best
# arrange the
# summary table to
# make this
# comparison as
# easy as possible.
summarized <- summary_table %>%
    pivot_longer(cols = "mean_vigilance":"CI_high",
        names_to = "stat",
```

```r
        values_to = "values") %>%
    pivot_wider(names_from = species,
        values_from = values)

pretty_summary <- kable(summarized,
    caption = "Summary statistics by treatment and species")
# Print table
pretty_summary
```

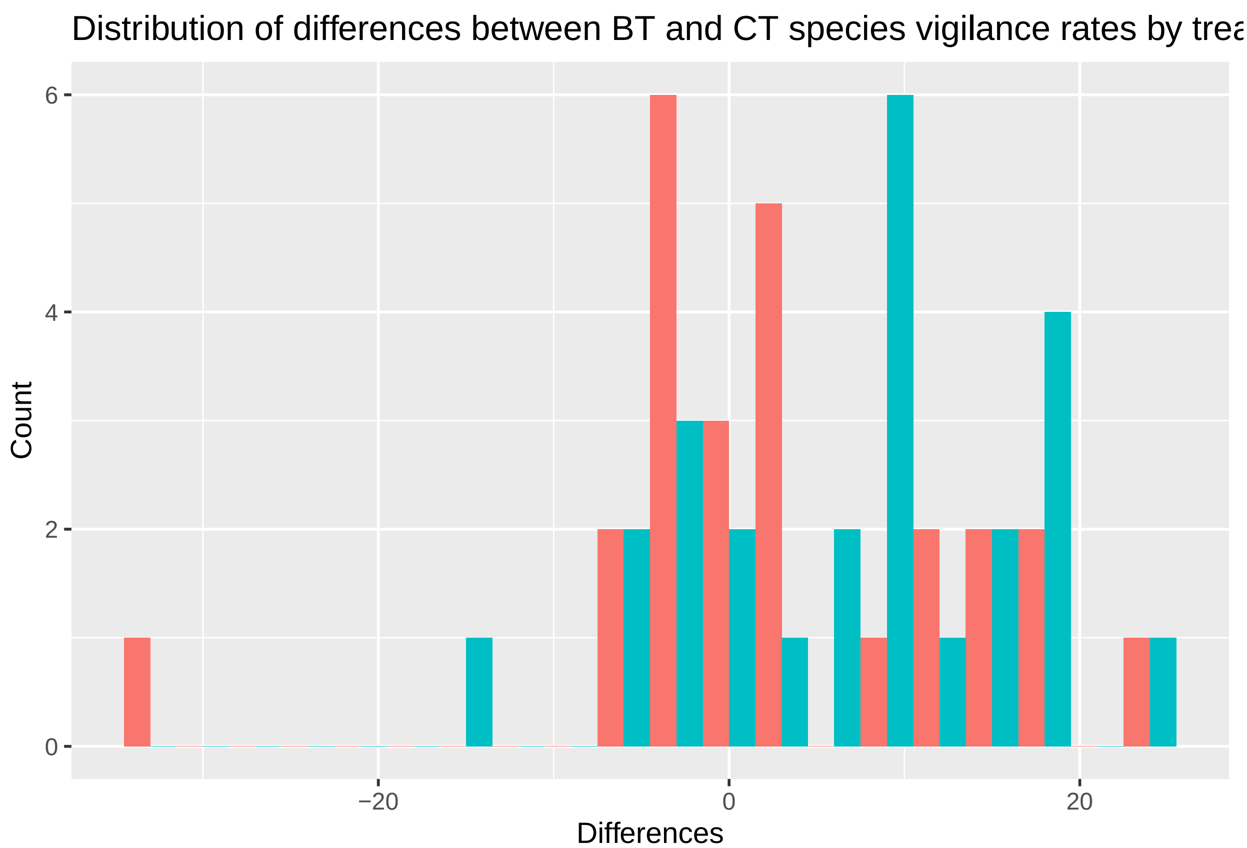
Table 1: Summary statistics by treatment and species

treatment	stat	BT	CT
control	mean_vigilance	11.8502381	8.385000
control	sd_vigilance	5.8319107	8.867963
control	median_vigilance	11.3450000	7.145000
control	iqr	6.6525000	7.250000
control	var	34.0111829	78.640762
control	se	0.8998834	1.739151
control	CI_low	10.0864667	4.976265
control	CI_high	13.6140095	11.793735
urban noise	mean_vigilance	15.2397619	9.421923
urban noise	sd_vigilance	7.4768630	7.001872
urban noise	median_vigilance	14.4300000	9.140000
urban noise	iqr	7.5625000	9.095000
urban noise	var	55.9034804	49.026208
urban noise	se	1.1537050	1.373180
urban noise	CI_low	12.9785001	6.730490
urban noise	CI_high	17.5010237	12.113356

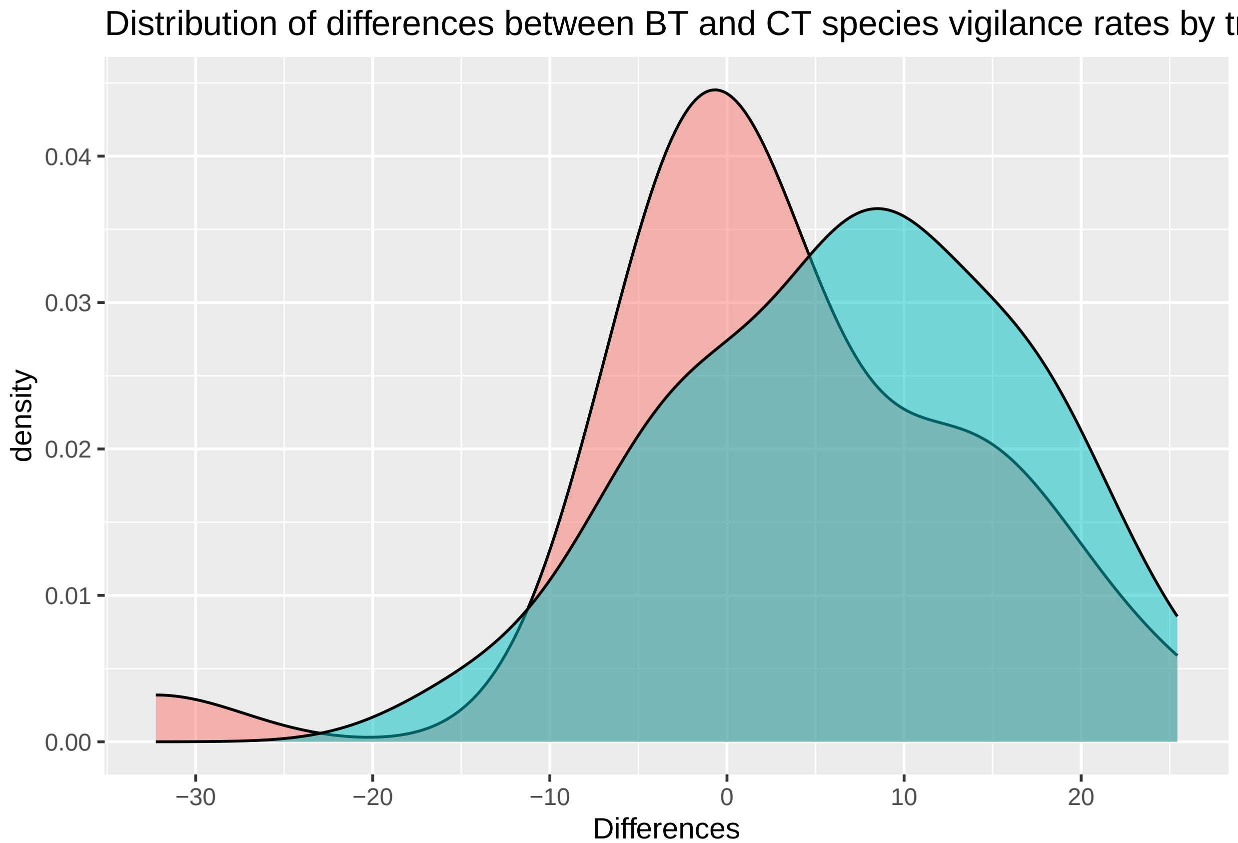
```r
# Figure comparing
# the means
summary_table %>%
    ggplot(aes(x = species,
        y = mean_vigilance,
        fill = species)) +
    geom_col(show.legend = FALSE) +
    geom_errorbar(aes(ymin = CI_low,
        ymax = CI_high),
        width = 0.3,
        color = "black") +
    facet_wrap(~treatment) +
    labs(title = "Mean vigilance rates of BT and CT species by noise treatment condition",
        x = "Species",
        y = "Mean vigilance rate",
        caption = "error bars represent a 95% confidence interval")
```



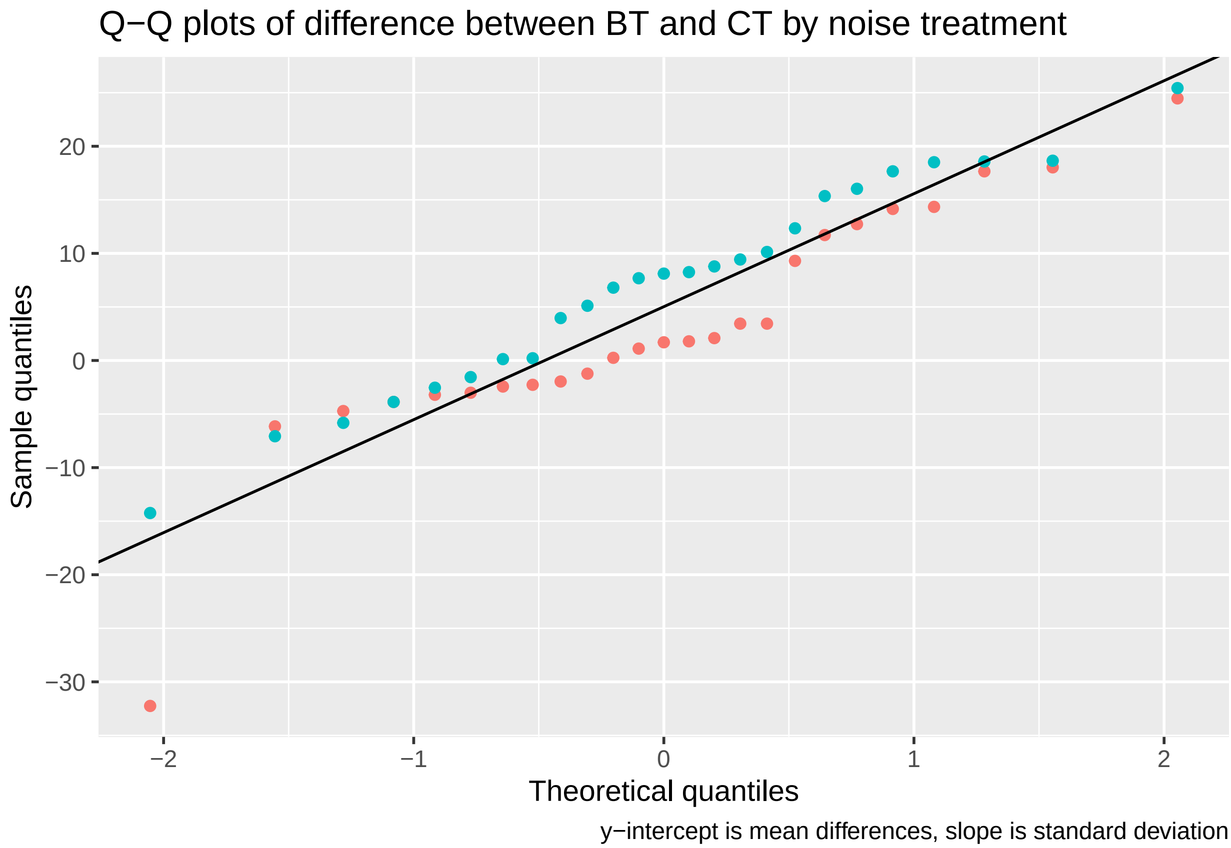

```r
# Histograms Here
# what matters is a
# histogram of the
# differences. You
# can add to the
# pivoted wide
# table created
# above.
ggplot(combined, aes(x = diffs,
    fill = treatment)) +
    geom_histogram(binwidth = 3,
        position = "dodge",
        na.rm = TRUE,
        show.legend = FALSE) +
    labs(title = "Distribution of differences between BT and CT species vigilance rates by treatment",
        x = "Differences",
        y = "Count")
```




```r
ggplot(combined, aes(x = diffs,
    fill = treatment)) +
    geom_density(alpha = 0.5,
        na.rm = TRUE,
        show.legend = FALSE) +
    labs(title = "Distribution of differences between BT and CT species vigilance rates by treatment",
        x = "Differences")
```




```
# Q-Q plots Be sure
# to do these for
# the differences
# as well
ggplot(na.omit(combined),
    aes(sample = diffs,
        color = treatment)) +
    stat_qq(show.legend = FALSE) +
    geom_abline(aes(intercept = mean(x = diffs),
        slope = sd(x = diffs))) +
    labs(title = "Q-Q plots of difference between BT and CT by noise treatment",
        x = "Theoretical quantiles",
        y = "Sample quantiles",
        caption = "y-intercept is mean differences, slope is standard deviation")
```



E1 Summary:

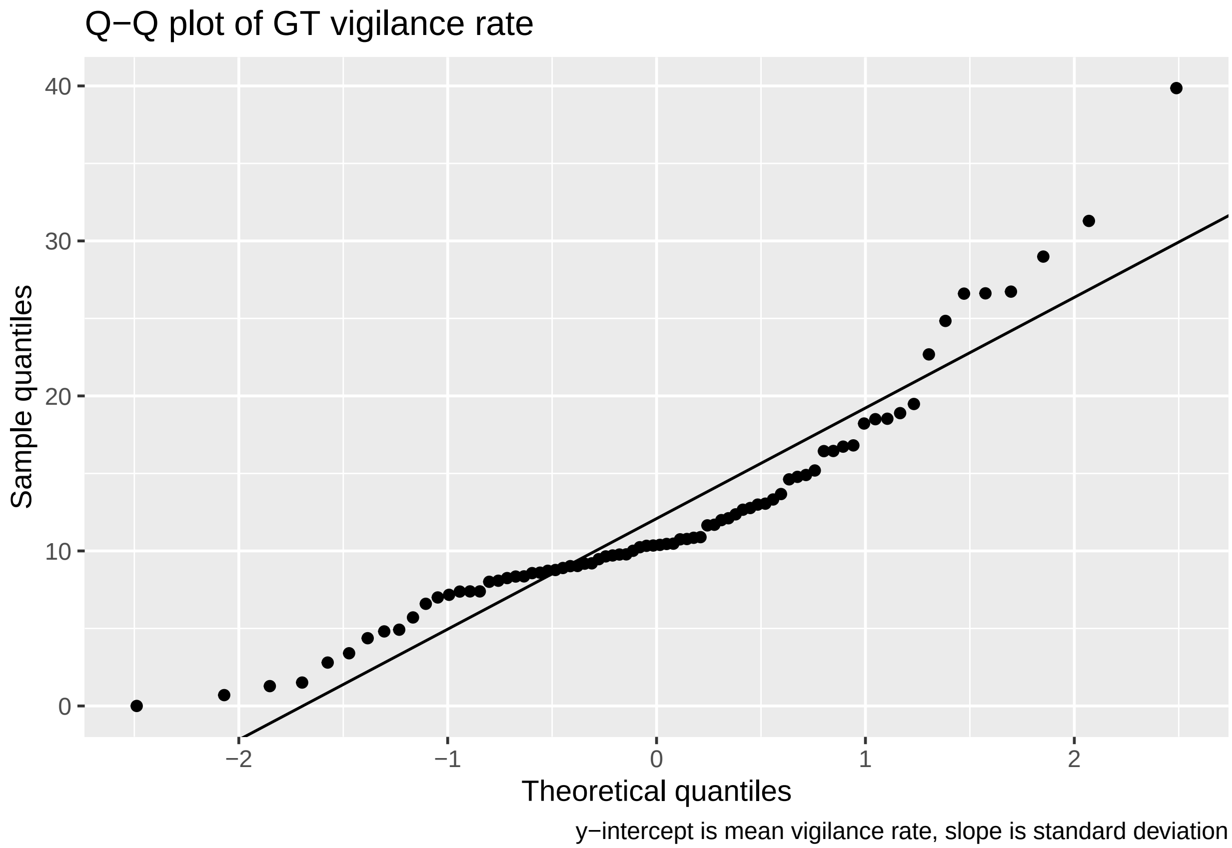
Based on our graph comparing the means of each species for each noise condition, we can see that there is a larger difference between the Blue Tits and Coal Tits in the urban noise condition. Based on our histograms and Q-Q plots, we can see that the differences in the vigilance rates between the species is roughly normally distributed in each noise condition. Based on our paired t-tests (run in the lab exercise 3), we found that the differences between the two species' mean vigilance rates are not statistically significant [t(24) = -1.35, p = 0.19] in the control condition. However in the urban noise condition, the two species exhibit statistically different vigilance rates [t(24) = -3.63, p <0.01]. Blue Tits were likely to have vigilance rates 3.04 to 11.05 lower than Coal Tits (95% CI), with a mean difference 7.04 lower. We were able to run the paired t-tests because our data consists of continuous variables, our observations are independent while the groups are dependent, and the differences are roughly normally distributed.

Exercise 2

Similar to exercise 1 above, compare the vigilance behavior between control and noise treatment sites for all other species present in the data. Here, simply check whether assumption for the t-test are met, then conduct the t-tests, and report on the results. No need to produce tables of descriptive statistics for each species. In which species is the difference the greatest and smallest? Be sure to report the difference in mean vigilance rates and their 95% confidence.

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

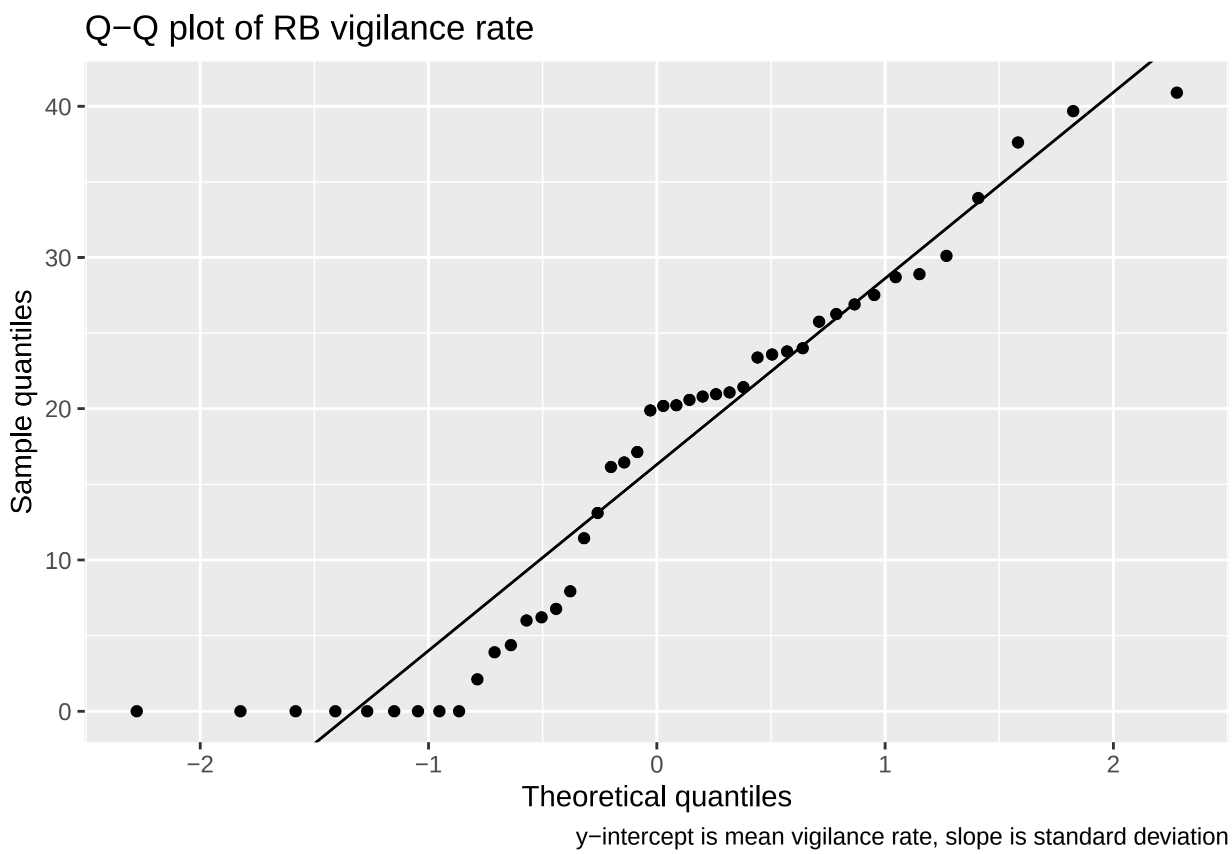
```
    select(site, species,
        treatment, vigilance.rate)

ggplot(GT, aes(sample = vigilance.rate)) +
    stat_qq() + geom_abline(aes(intercept = mean(x = vigilance.rate),
    slope = sd(x = vigilance.rate))) +
    labs(title = "Q-Q plot of GT vigilance rate",
        x = "Theoretical quantiles",
        y = "Sample quantiles",
        caption = "y-intercept is mean vigilance rate, slope is standard deviation")
```




```
t.test(formula = GT$vigilance.rate ~
    GT$treatment)
```

```
##
##  Welch Two Sample t-test
##
## data:  GT$vigilance.rate by GT$treatment
## t = -1.6674, df = 71.77, p-value = 0.09979
## alternative hypothesis: true difference in means between group control and group urban noise is not e
## 95 percent confidence interval:
##  -5.8487335  0.5210412
## sample estimates:
##     mean in group control mean in group urban noise
##                  10.75462                  13.41846
```

```r
RB <- vigilance %>%
    filter(species ==
        "RB") %>%
    select(site, species,
        treatment, vigilance.rate)

ggplot(RB, aes(sample = vigilance.rate)) +
    stat_qq() + geom_abline(aes(intercept = mean(x = vigilance.rate),
    slope = sd(x = vigilance.rate))) +
    labs(title = "Q-Q plot of RB vigilance rate",
        x = "Theoretical quantiles",
        y = "Sample quantiles",
        caption = "y-intercept is mean vigilance rate, slope is standard deviation")
```

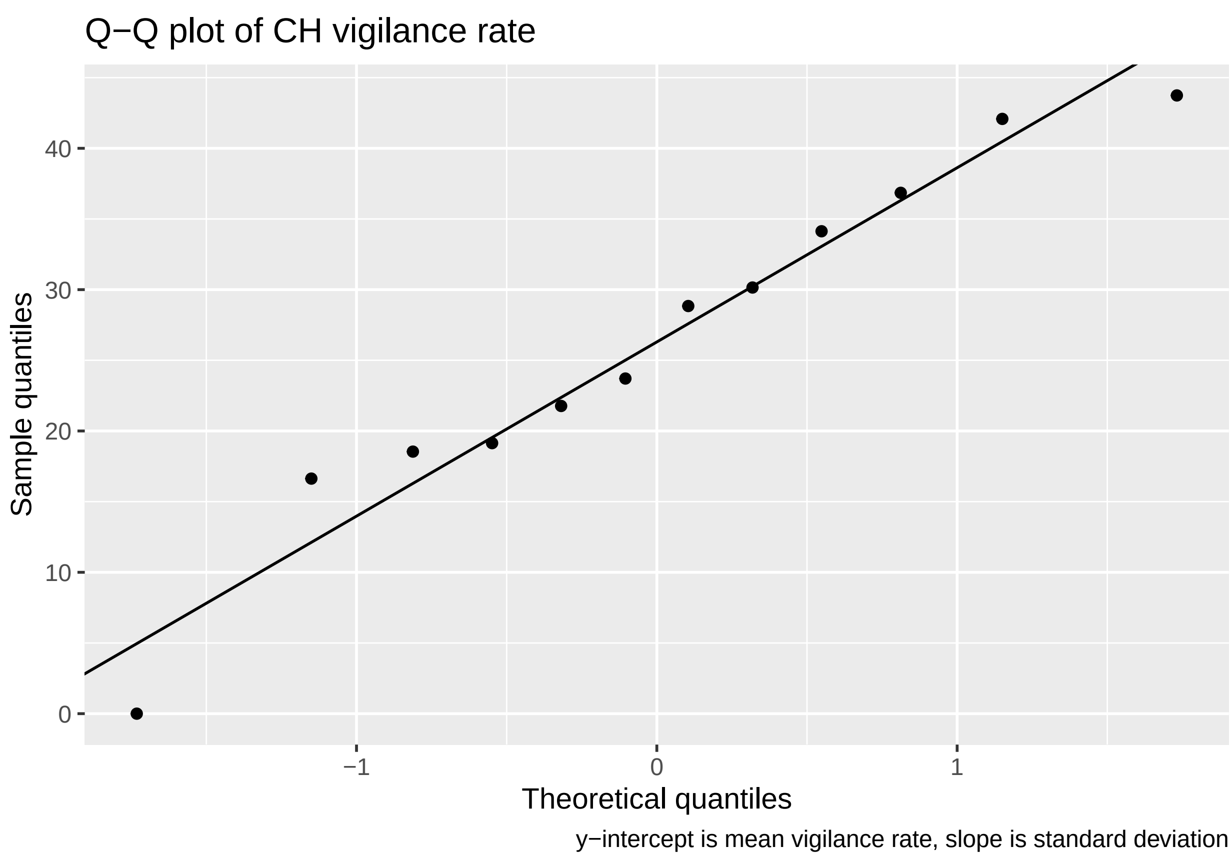



```r
t.test(formula = RB$vigilance.rate ~
    RB$treatment)
```

```
##
##  Welch Two Sample t-test
##
## data:  RB$vigilance.rate by RB$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 e
## 95 percent confidence interval:
##  -9.392586  5.716222
```

```
## sample estimates:
##     mean in group control mean in group urban noise
##                  15.39455                  17.23273
```

```
CH <- vigilance %>%
    filter(species ==
        "CH") %>%
    select(site, species,
        treatment, vigilance.rate)

ggplot(CH, aes(sample = vigilance.rate)) +
    stat_qq() + geom_abline(aes(intercept = mean(x = vigilance.rate),
    slope = sd(x = vigilance.rate))) +
    labs(title = "Q-Q plot of CH vigilance rate",
        x = "Theoretical quantiles",
        y = "Sample quantiles",
        caption = "y-intercept is mean vigilance rate, slope is standard deviation")
```

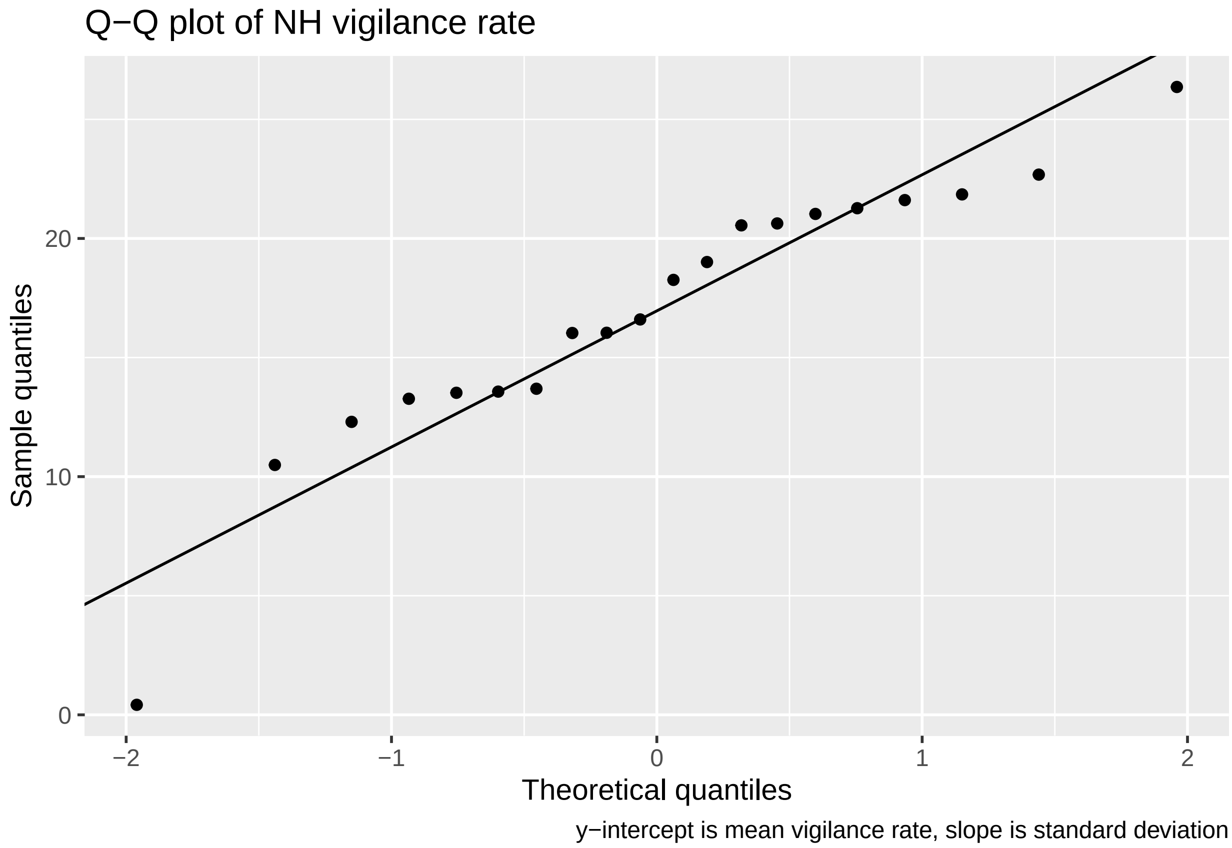



```
t.test(formula = CH$vigilance.rate ~
    CH$treatment)
```

```
##
##  Welch Two Sample t-test
##
## data:  CH$vigilance.rate by CH$treatment
```

```
## t = -1.3198, df = 7.8586, p-value = 0.224
## alternative hypothesis: true difference in means between group control and group urban noise is not e
## 95 percent confidence interval:
##  -25.02218   6.84218
## sample estimates:
##     mean in group control mean in group urban noise
##                  21.75333                  30.84333
```

```
NH <- vigilance %>%
    filter(species ==
        "NH") %>%
    select(site, species,
        treatment, vigilance.rate)

ggplot(NH, aes(sample = vigilance.rate)) +
    stat_qq() + geom_abline(aes(intercept = mean(x = vigilance.rate),
    slope = sd(x = vigilance.rate))) +
    labs(title = "Q-Q plot of NH vigilance rate",
        x = "Theoretical quantiles",
        y = "Sample quantiles",
        caption = "y-intercept is mean vigilance rate, slope is standard deviation")
```




```
t.test(formula = NH$vigilance.rate ~
    NH$treatment)
```

```
##
```

```
##  Welch Two Sample t-test
##
## data:  NH$vigilance.rate by NH$treatment
## t = -2.9156, df = 14.572, p-value = 0.01091
## alternative hypothesis: true difference in means between group control and group urban noise is not e
## 95 percent confidence interval:
##  -10.934749  -1.685251
## sample estimates:
##     mean in group control mean in group urban noise
##                    13.804                    20.114
```

E2 Summary:

We are able to run independent t-tests here because our data consists of continuous variables, our observations are independent along with each group being independent, and the rates are roughly normally distributed. Based on our independent t-tests, we can conclude the following: - There is not a statistically significant difference between vigilance rates in the control and noise treatments for the GT bird species [t(71.77)) = -1.67, p = 0.09]. The mean vigilance rate is likely to vary by 10 to 13 (95% CI). - There is not a statistically significant difference between vigilance rates in the control and noise treatments for European Robins [t(41.97) = -0.49, p = 0.63]. The mean vigilance rate is likely to vary by -9 to 5 (95% CI). This species has the smallest difference between the mean vigilance rates between treatment sites, compared to the other species. - There is not a statistically significant difference between vigilance rates in the control and noise treatments for the CH bird species [t(7.85) = -1.31, p = 0.22]. The mean vigilance rate is likely to vary by -25 to 6 (95% CI). This species has the greatest difference between the mean vigilance rates between treatment sites, compared to the other species. - There is a statistically significant difference between vigilance rates in the control and noise treatments for the NH bird species [t(14.57) = -2.91, p = 0.01]. The mean vigilance rate is likely to vary by -10 to -1 (95% CI).

Exercise 3

Assess whether there is a difference in wind speed, temperature and hours from sunrise between control and noise treatment sites. This could of course influence the outcome of the study by influencing vigilance behavior of any birds that are being observed. For each of these three continuous variables, conduct descriptive analyses and summary, a visual assessment of normal distribution, conduct an appropriate t-test, and report your results in a summary statement.

```r
## WIND SPEED
wind_speed <- vigilance %>%
    select(site, species,
        treatment, wind.speed)

wind_summary <- wind_speed %>%
    group_by(treatment) %>%
    summarize(mean_wind = mean(wind.speed),
        sd_wind = sd(wind.speed),
        median_wind = median(wind.speed),
        iqr = IQR(wind.speed),
        var = var(wind.speed),
        se = sd_wind/sqrt(n()),
        CI_low = mean_wind -
            1.96 * se,
        CI_high = mean_wind +
```

```
            1.96 * se)
wind_summary
```

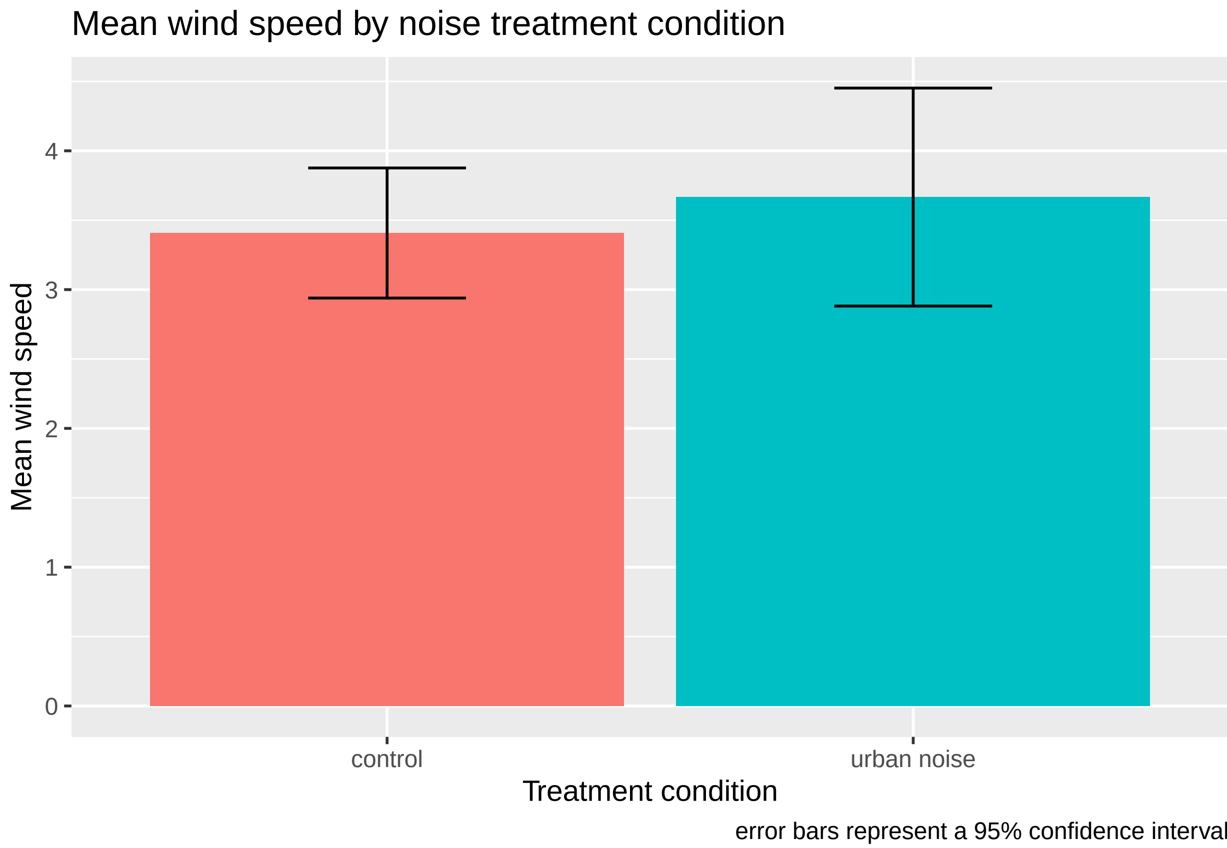
```
## # A tibble: 2 x 9
##   treatment    mean_wind sd_wind median_wind   iqr   var    se CI_low CI_high
##   <chr>            <dbl>   <dbl>       <dbl> <dbl> <dbl> <dbl>  <dbl>   <dbl>
## 1 control           3.41    2.88        2.95  3.55  8.29 0.239   2.94    3.88
## 2 urban noise       3.67    4.82        2.2   3.8  23.3  0.401   2.88    4.45
```

```
summarized_wind <- wind_summary %>%
    pivot_longer(cols = "mean_wind":"CI_high",
        names_to = "stat",
        values_to = "values") %>%
    pivot_wider(names_from = treatment,
        values_from = values)

pretty_summary_wind <- kable(summarized_wind,
    caption = "Summary statistics by treatment for wind speed")
pretty_summary_wind
```

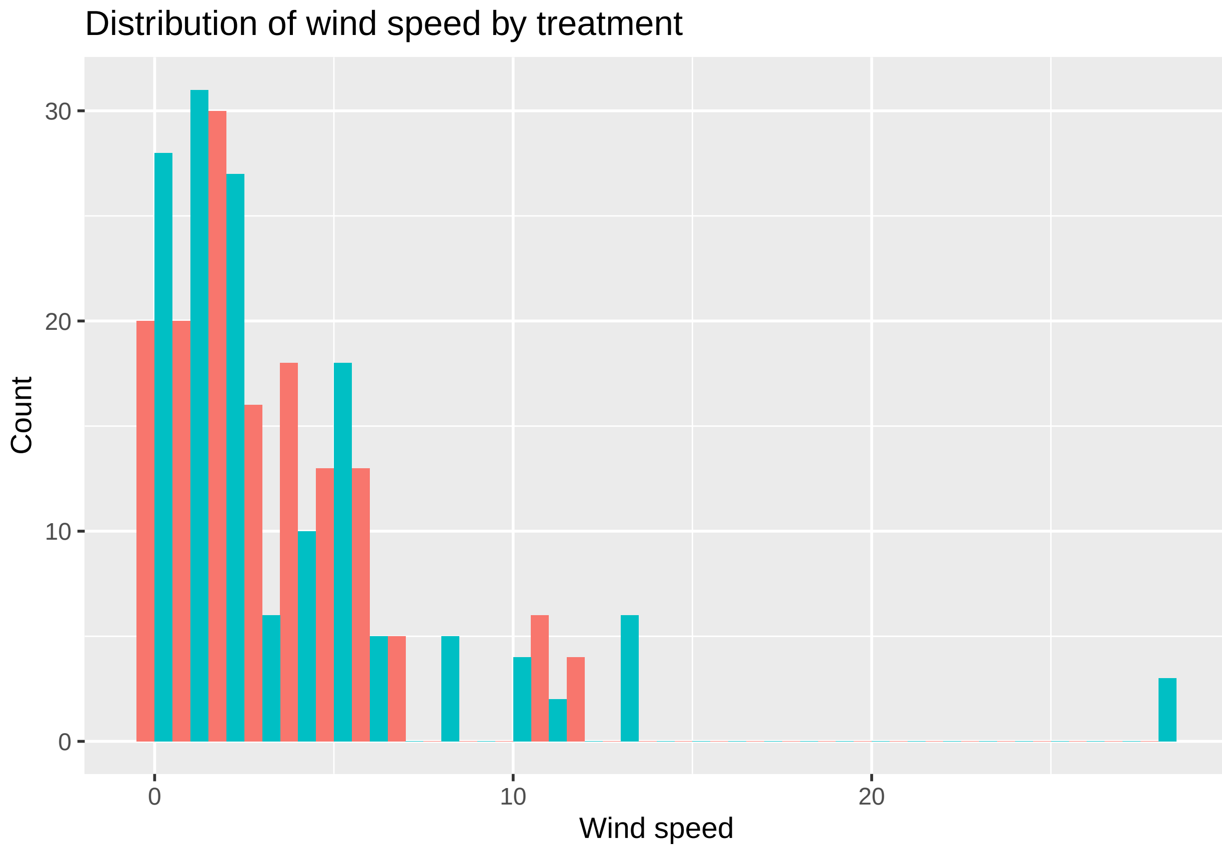
Table 2: Summary statistics by treatment for wind speed

stat	control	urban noise
mean_wind	3.4076552	3.6666897
sd_wind	2.8796055	4.8244903
median_wind	2.9500000	2.2000000
iqr	3.5500000	3.8000000
var	8.2921278	23.2757070
se	0.2391382	0.4006521
CI_low	2.9389443	2.8814115
CI_high	3.8763661	4.4519678

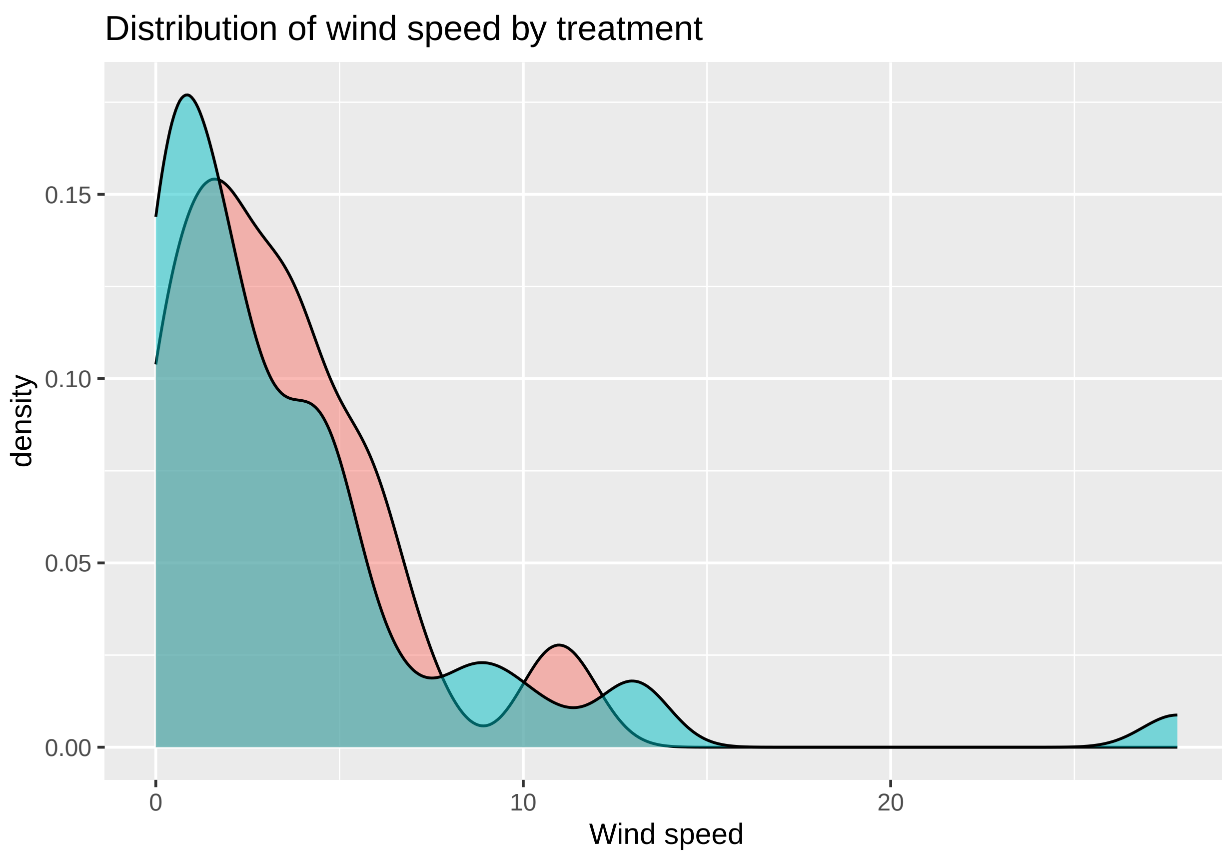
```
# Figure comparing
# the means
wind_summary %>%
    ggplot(aes(x = treatment,
        y = mean_wind,
        fill = treatment)) +
    geom_col(show.legend = FALSE) +
    geom_errorbar(aes(ymin = CI_low,
        ymax = CI_high),
        width = 0.3,
        color = "black") +
    labs(title = "Mean wind speed by noise treatment condition",
        x = "Treatment condition",
        y = "Mean wind speed",
        caption = "error bars represent a 95% confidence interval")
```



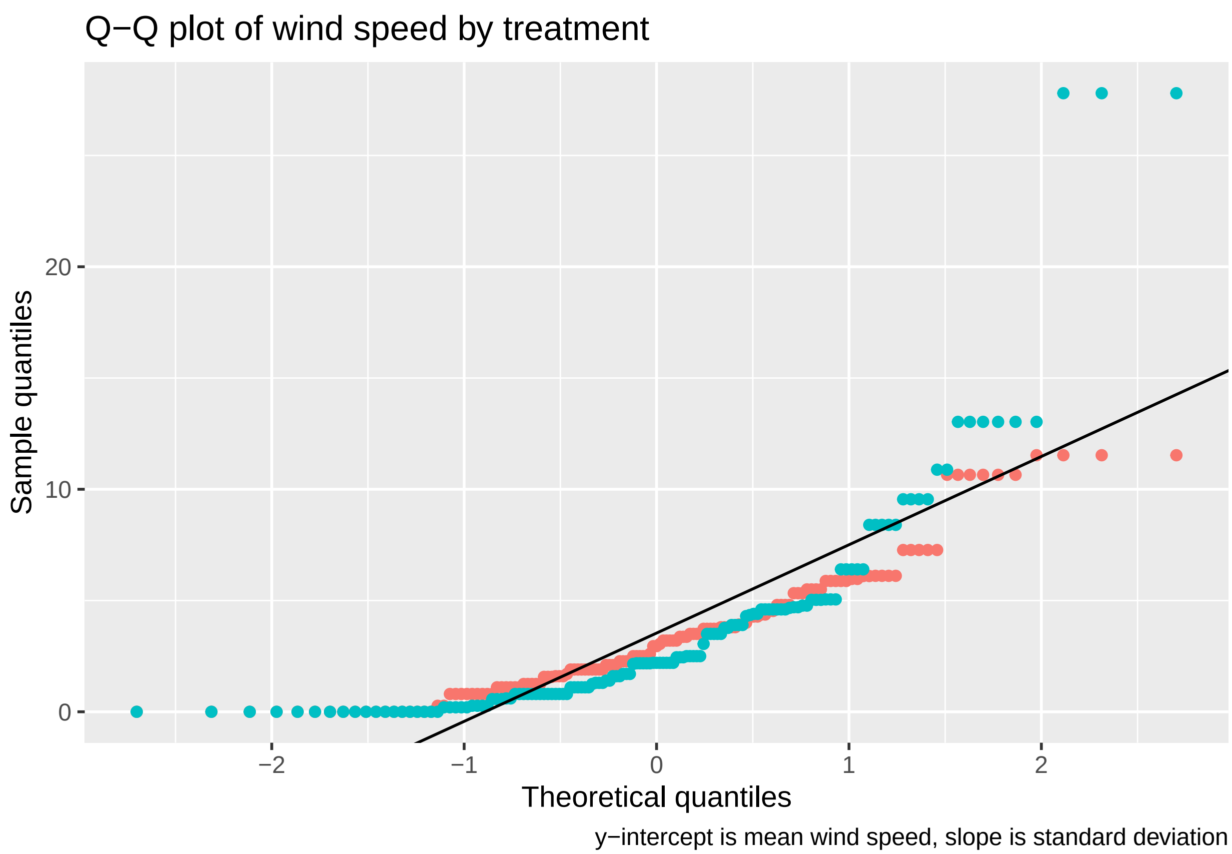

```
# Histograms
ggplot(wind_speed, aes(x = wind.speed,
    fill = treatment)) +
    geom_histogram(binwidth = 1,
        position = "dodge",
        na.rm = TRUE,
        show.legend = FALSE) +
    labs(title = "Distribution of wind speed by treatment",
        x = "Wind speed",
        y = "Count")
```




```r
ggplot(wind_speed, aes(x = wind.speed,
    fill = treatment)) +
    geom_density(alpha = 0.5,
        na.rm = TRUE,
        show.legend = FALSE) +
    labs(title = "Distribution of wind speed by treatment",
        x = "Wind speed")
```




```r
# Q-Q plots
ggplot(wind_speed, aes(sample = wind.speed,
    color = treatment)) +
    stat_qq(show.legend = FALSE) +
    geom_abline(aes(intercept = mean(x = wind.speed),
        slope = sd(x = wind.speed))) +
    labs(title = "Q-Q plot of wind speed by treatment",
        x = "Theoretical quantiles",
        y = "Sample quantiles",
        caption = "y-intercept is mean wind speed, slope is standard deviation")
```




```r
t.test(formula = wind_speed$wind.speed ~
    wind_speed$treatment)
```

```
##
##  Welch Two Sample t-test
##
## data:  wind_speed$wind.speed by wind_speed$treatment
## t = -0.55516, df = 235.05, p-value = 0.5793
## alternative hypothesis: true difference in means between group control and group urban noise is not e
## 95 percent confidence interval:
##  -1.1782735  0.6602045
## sample estimates:
##     mean in group control mean in group urban noise
##                  3.407655                  3.666690
```

E3 Wind Speed Summary:

Based on our graph comparing the means of the wind speeds by treatment condition, we can see that there isn't a large difference between the means for each condition. Based on our histograms and Q-Q plots, we can see that both treatment conditions are heavily right-skewed. Though the data are skewed, we ran independent t-tests here because our data consists of continuous variables and our observations are independent along with each group being independent. Based on our independent t-tests, there is not a statistically significant difference between vigilance rates in the control and noise treatments for the GT bird species [t(235)) = -0.55, p = 0.05]. The mean vigilance rate is likely to vary by -1.1 to 0.6 (95% CI).

```r
## TEMPERATURE
temperature <- vigilance %>%
    select(site, species,
        treatment, temperature)

temperature_summary <- temperature %>%
    group_by(treatment) %>%
    summarize(mean_temp = mean(temperature),
        sd_temp = sd(temperature),
        median_temp = median(temperature),
        iqr = IQR(temperature),
        var = var(temperature),
        se = sd_temp/sqrt(n()),
        CI_low = mean_temp -
            1.96 * se,
        CI_high = mean_temp +
            1.96 * se)
temperature_summary
```

```
## # A tibble: 2 x 9
##   treatment   mean_temp sd_temp median_temp   iqr   var    se CI_low CI_high
##   <chr>           <dbl>   <dbl>       <dbl> <dbl> <dbl> <dbl>  <dbl>   <dbl>
## 1 control          9.56    3.26        9.5    4.1  10.7 0.271   9.03    10.1
## 2 urban noise      9.31    3.24        8.23   4.4  10.5 0.269   8.78    9.83
```

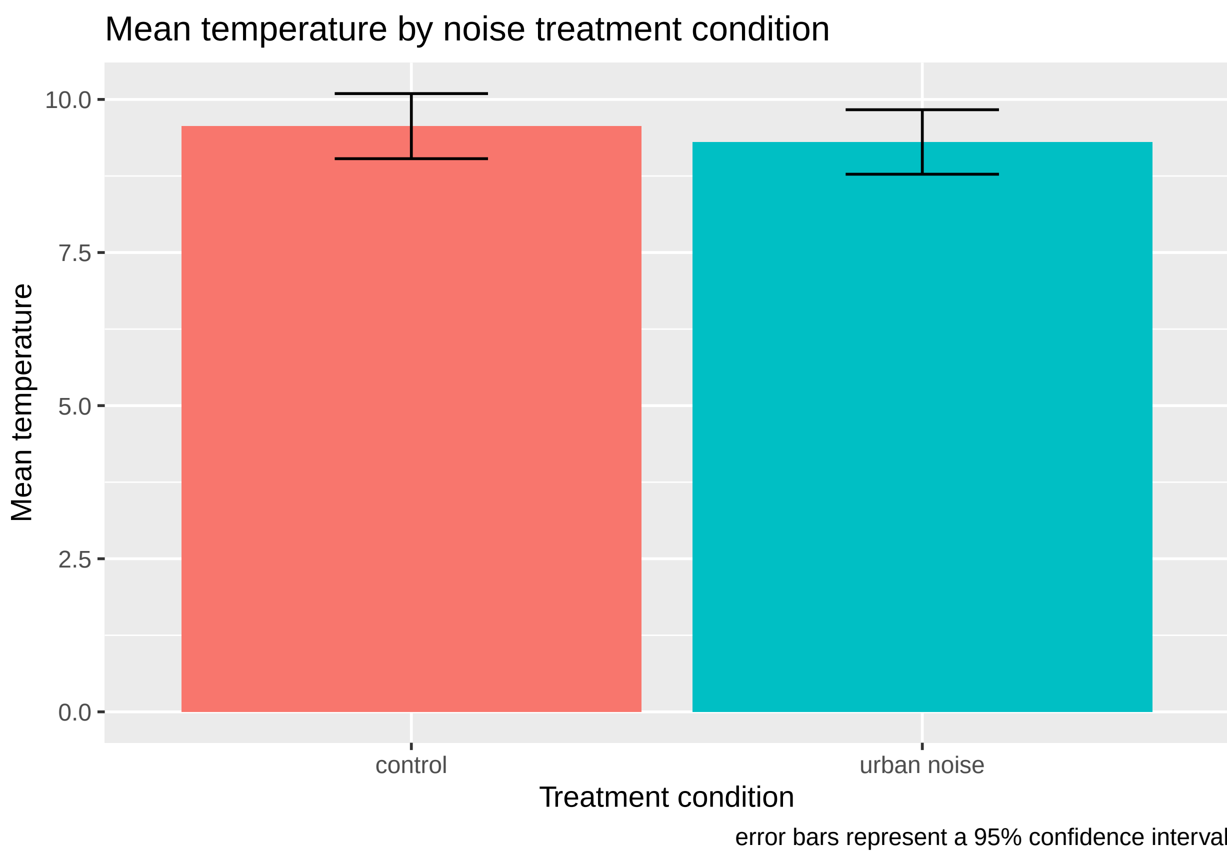
```r
summarized_temp <- temperature_summary %>%
    pivot_longer(cols = "mean_temp":"CI_high",
        names_to = "stat",
        values_to = "values") %>%
    pivot_wider(names_from = treatment,
        values_from = values)

pretty_summary_temp <- kable(summarized_temp,
    caption = "Summary statistics by treatment for temperature")
pretty_summary_temp
```

Table 3: Summary statistics by treatment for temperature

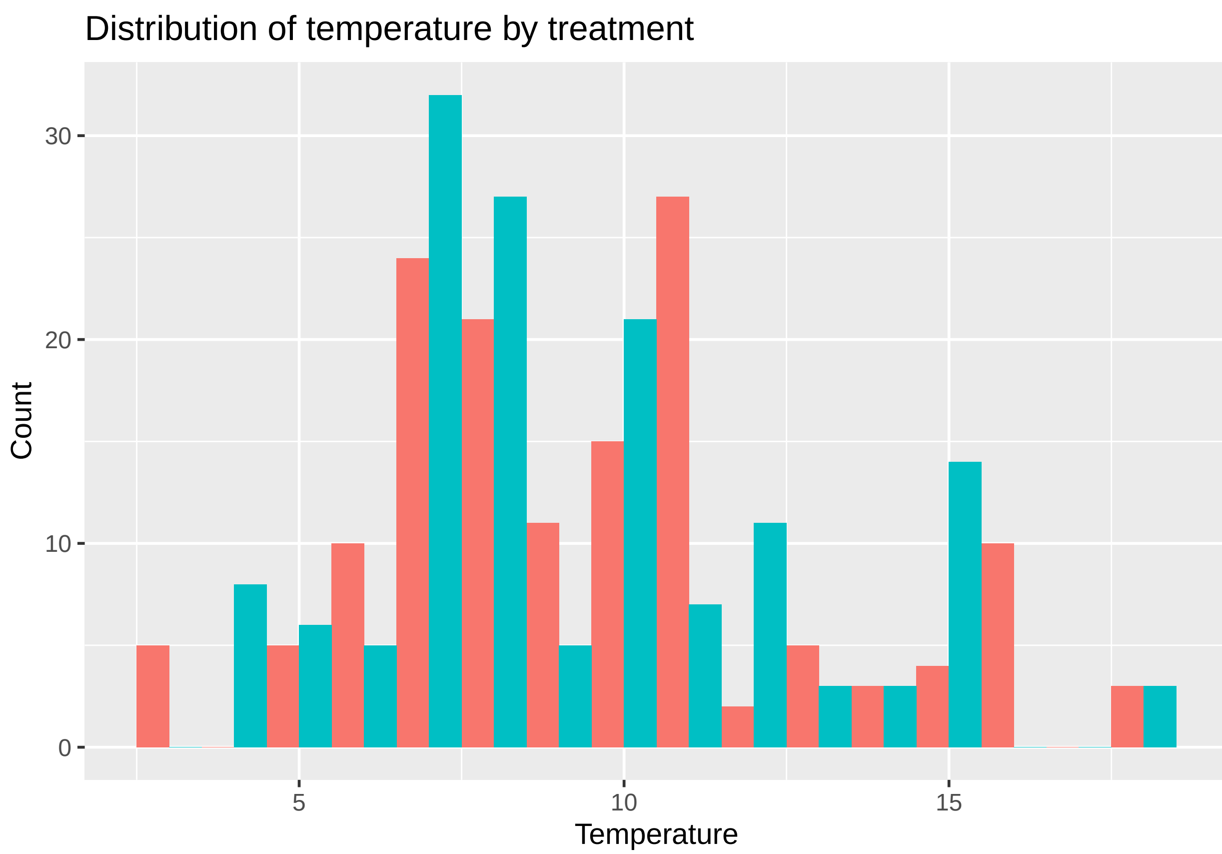
stat	control	urban noise
mean_temp	9.5640000	9.3054483
sd_temp	3.2640458	3.2364397
median_temp	9.5000000	8.2300000
iqr	4.1000000	4.4000000
var	10.6539950	10.4745416
se	0.2710642	0.2687717
CI_low	9.0327141	8.7786558
CI_high	10.0952859	9.8322408

```r
# Figure comparing
# the means
temperature_summary %>%
```

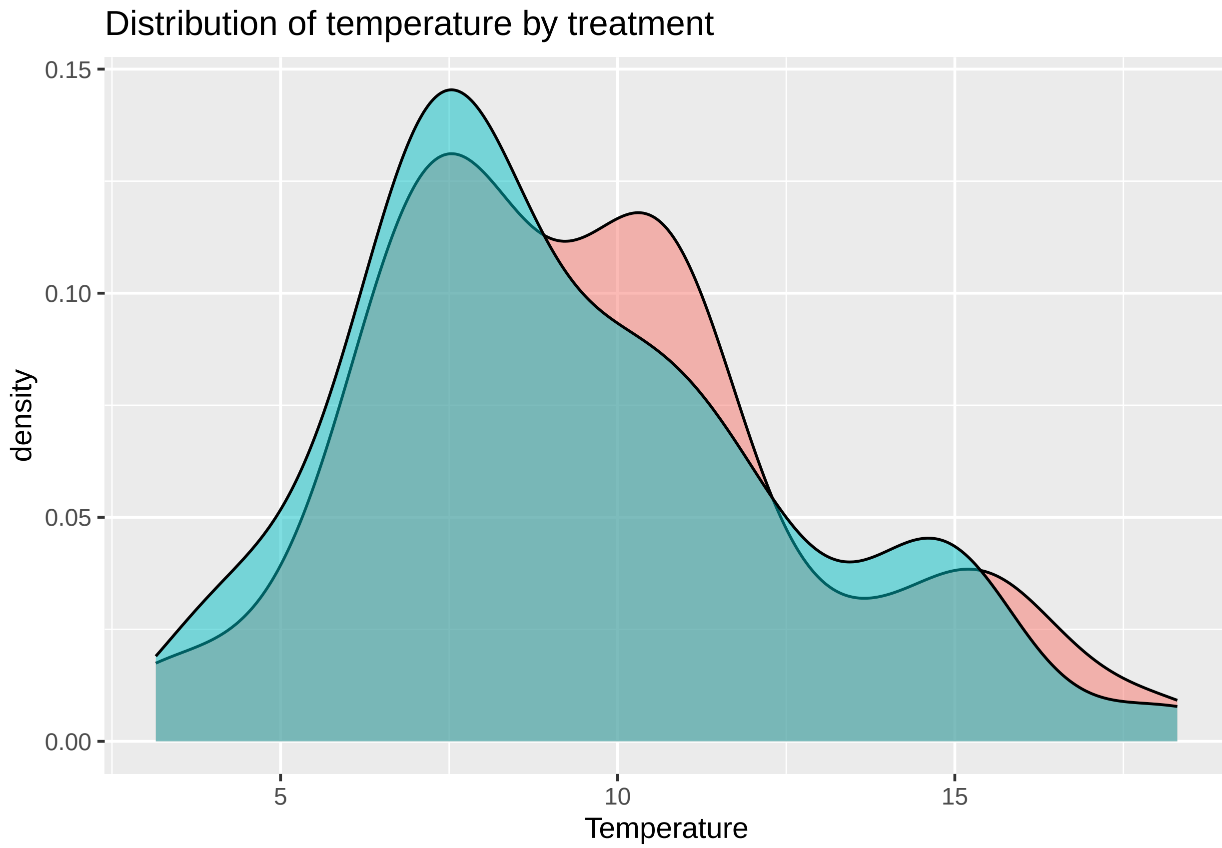
```r
    ggplot(aes(x = treatment,
        y = mean_temp,
        fill = treatment)) +
    geom_col(show.legend = FALSE) +
    geom_errorbar(aes(ymin = CI_low,
        ymax = CI_high),
        width = 0.3,
        color = "black") +
    labs(title = "Mean temperature by noise treatment condition",
        x = "Treatment condition",
        y = "Mean temperature",
        caption = "error bars represent a 95% confidence interval")
```



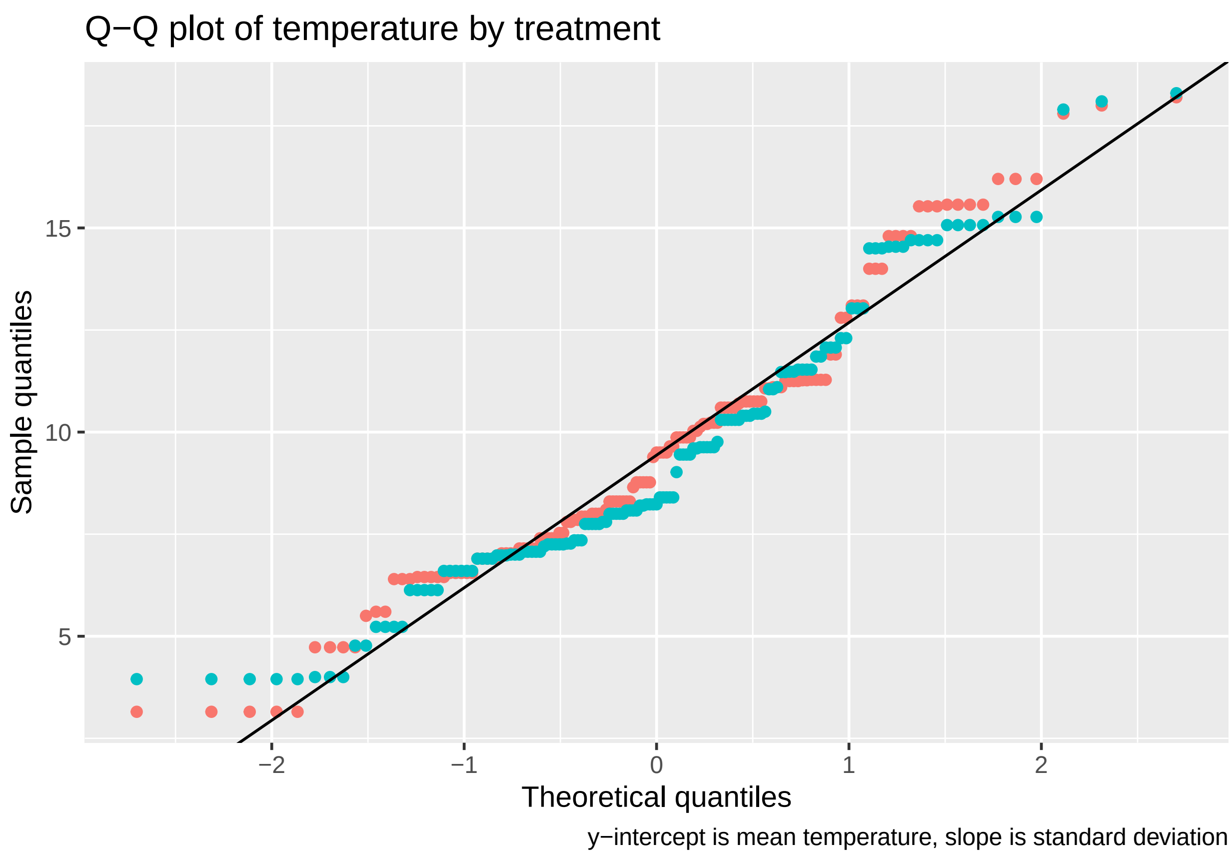

```r
# Histograms
ggplot(temperature, aes(x = temperature,
    fill = treatment)) +
    geom_histogram(binwidth = 1,
        position = "dodge",
        na.rm = TRUE,
        show.legend = FALSE) +
    labs(title = "Distribution of temperature by treatment",
        x = "Temperature",
        y = "Count")
```




```
ggplot(temperature, aes(x = temperature,
    fill = treatment)) +
    geom_density(alpha = 0.5,
        na.rm = TRUE,
        show.legend = FALSE) +
    labs(title = "Distribution of temperature by treatment",
        x = "Temperature")
```




```r
# Q-Q plots
ggplot(temperature, aes(sample = temperature,
    color = treatment)) +
    stat_qq(show.legend = FALSE) +
    geom_abline(aes(intercept = mean(x = temperature),
        slope = sd(x = temperature))) +
    labs(title = "Q-Q plot of temperature by treatment",
        x = "Theoretical quantiles",
        y = "Sample quantiles",
        caption = "y-intercept is mean temperature, slope is standard deviation")
```




```r
t.test(formula = temperature$temperature ~
    temperature$treatment)
```

```
##
##  Welch Two Sample t-test
##
## data:  temperature$temperature by temperature$treatment
## t = 0.67732, df = 287.98, p-value = 0.4987
## alternative hypothesis: true difference in means between group control and group urban noise is not e
## 95 percent confidence interval:
##  -0.4927733  1.0098767
## sample estimates:
##     mean in group control mean in group urban noise
##                  9.564000                  9.305448
```

E3 Temperature Summary:

Based on our graph comparing the means of the temperatures by treatment condition, we can see that the means for each condition are roughly the same. Based on our histograms and Q-Q plots, we can see that both treatment conditions are roughly normally distributed. We are able to run independent t-tests here because our data consists of continuous variables, our observations are independent along with each group being independent, and the data are roughly normally distributed. Based on our independent t-tests, there is not a statistically significant difference between vigilance rates in the control and noise treatments for the GT bird species [t(278)) = 0.67, p = 0.49]. The mean vigilance rate is likely to vary by -0.49 to 1.0 (95% CI).

```r
## HOURS FROM
## SUNRISE
sunrise_hours <- vigilance %>%
    select(site, species,
        treatment, hours.from.sunrise)

sun_hours_summary <- sunrise_hours %>%
    group_by(treatment) %>%
    summarize(mean_sun_hours = mean(hours.from.sunrise),
        sd_sun_hours = sd(hours.from.sunrise),
        median_sun_hours = median(hours.from.sunrise),
        iqr = IQR(hours.from.sunrise),
        var = var(hours.from.sunrise),
        se = sd_sun_hours/sqrt(n()),
        CI_low = mean_sun_hours -
            1.96 * se,
        CI_high = mean_sun_hours +
            1.96 * se)
sun_hours_summary
```

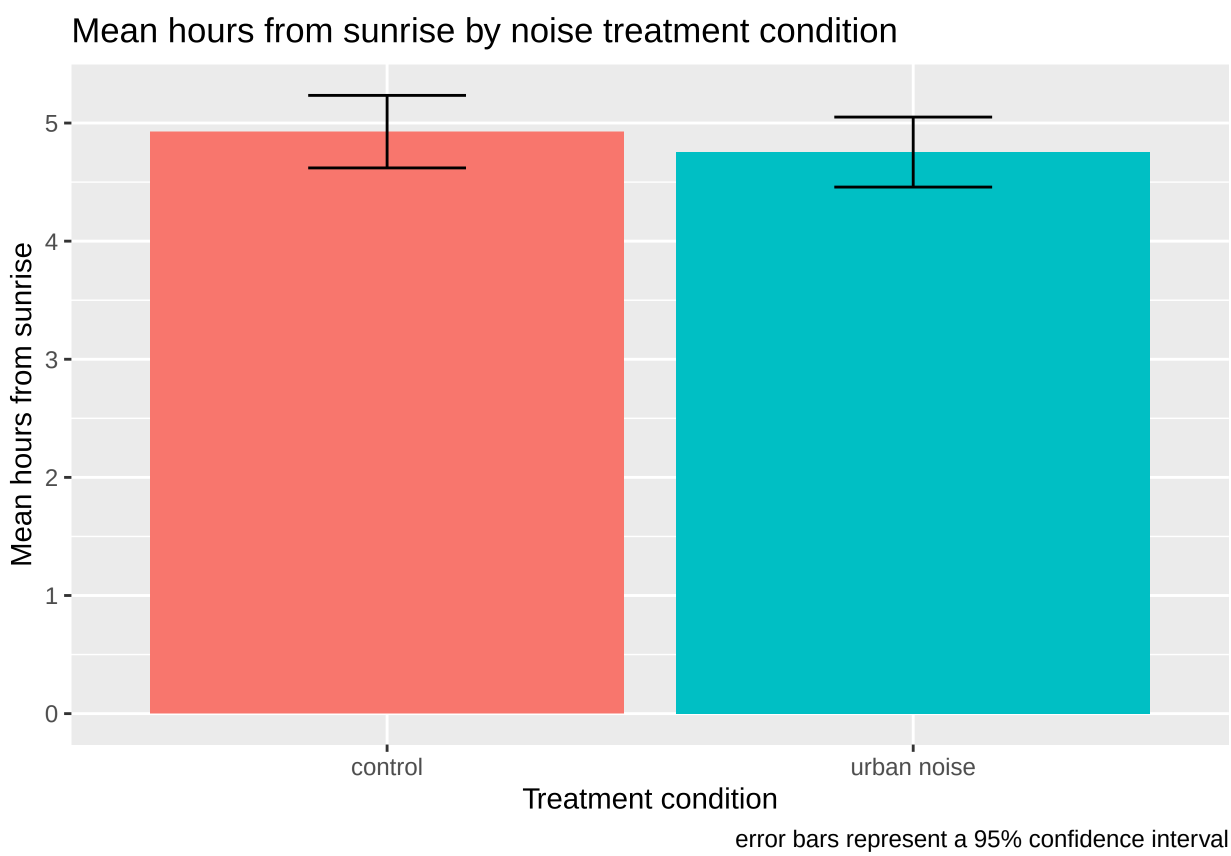
```
## # A tibble: 2 x 9
##   treatment   mean_sun_hours sd_sun_hours median_sun_hours   iqr   var    se
##   <chr>                <dbl>        <dbl>            <dbl> <dbl> <dbl> <dbl>
## 1 control               4.93         1.89             5.45  3.22  3.56 0.157
## 2 urban noise           4.75         1.82             4.78  2.6   3.32 0.151
## # i 2 more variables: CI_low <dbl>, CI_high <dbl>
```

```r
summarized_sun_hours <- sun_hours_summary %>%
    pivot_longer(cols = "mean_sun_hours":"CI_high",
        names_to = "stat",
        values_to = "values") %>%
    pivot_wider(names_from = treatment,
        values_from = values)

pretty_summary_sun <- kable(summarized_sun_hours,
    caption = "Summary statistics by treatment for hours from sunrise")
pretty_summary_sun
```

Table 4: Summary statistics by treatment for hours from sunrise

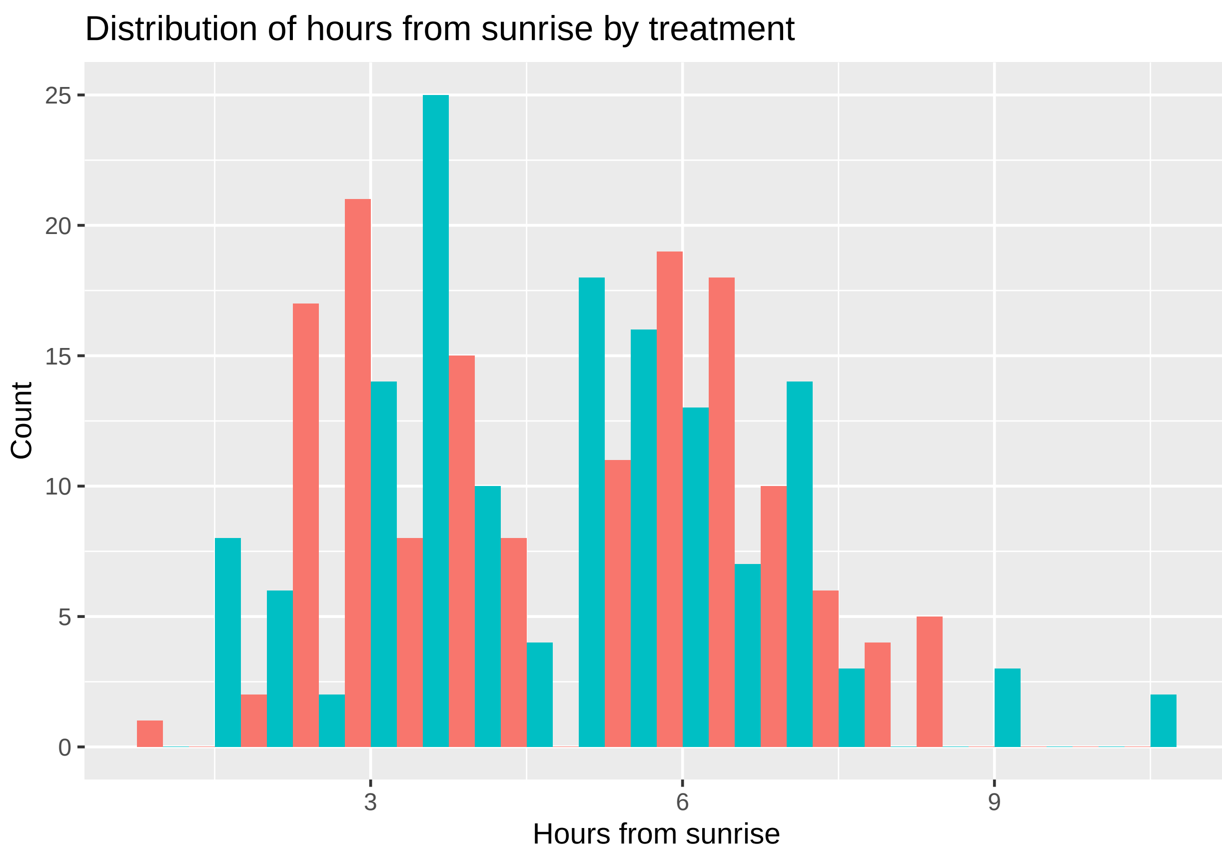
stat	control	urban noise
mean_sun_hours	4.9269655	4.7542069
sd_sun_hours	1.8860417	1.8209444
median_sun_hours	5.4500000	4.7800000
iqr	3.2200000	2.6000000
var	3.5571532	3.3158384
se	0.1566272	0.1512212
CI_low	4.6199761	4.4578133
CI_high	5.2339549	5.0506004

```r
# Figure comparing
# the means
sun_hours_summary %>%
    ggplot(aes(x = treatment,
        y = mean_sun_hours,
        fill = treatment)) +
    geom_col(show.legend = FALSE) +
    geom_errorbar(aes(ymin = CI_low,
        ymax = CI_high),
        width = 0.3,
        color = "black") +
    labs(title = "Mean hours from sunrise by noise treatment condition",
        x = "Treatment condition",
        y = "Mean hours from sunrise",
        caption = "error bars represent a 95% confidence interval")
```

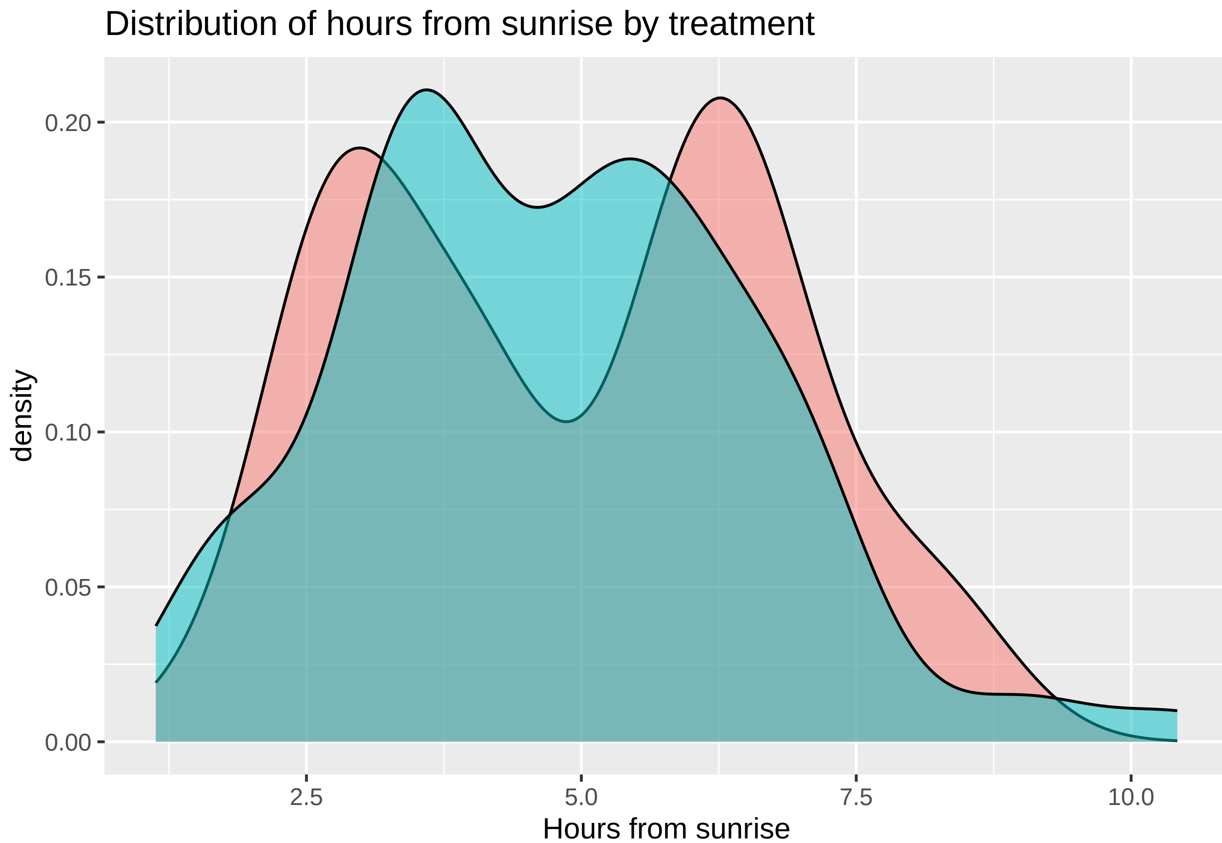



```r
# Histograms
ggplot(sunrise_hours,
    aes(x = hours.from.sunrise,
        fill = treatment)) +
    geom_histogram(binwidth = 0.5,
        position = "dodge",
        na.rm = TRUE,
        show.legend = FALSE) +
    labs(title = "Distribution of hours from sunrise by treatment",
```

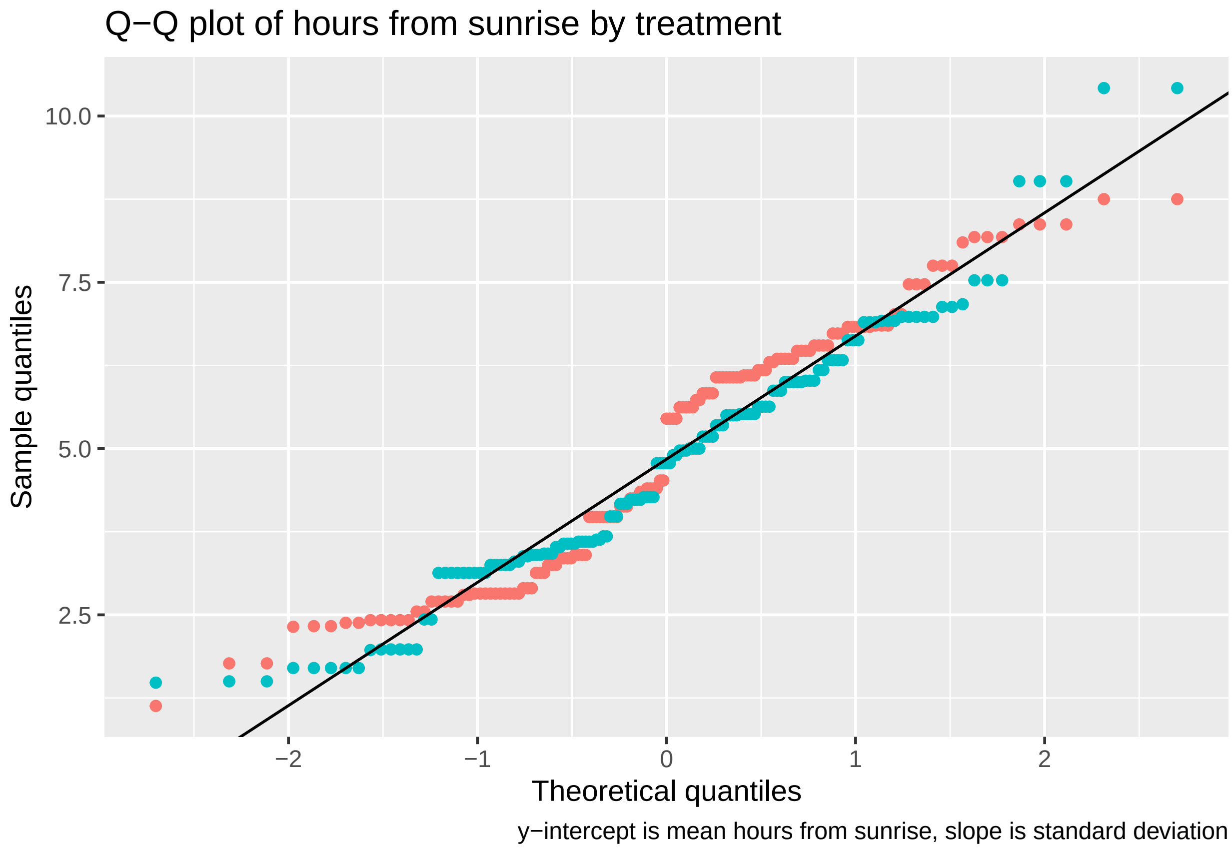
```
        x = "Hours from sunrise",
        y = "Count")
```




```
ggplot(sunrise_hours,
    aes(x = hours.from.sunrise,
        fill = treatment)) +
    geom_density(alpha = 0.5,
        na.rm = TRUE,
        show.legend = FALSE) +
    labs(title = "Distribution of hours from sunrise by treatment",
        x = "Hours from sunrise")
```




```
# Q-Q plots
ggplot(sunrise_hours,
    aes(sample = hours.from.sunrise,
        color = treatment)) +
    stat_qq(show.legend = FALSE) +
    geom_abline(aes(intercept = mean(x = hours.from.sunrise),
        slope = sd(x = hours.from.sunrise))) +
    labs(title = "Q-Q plot of hours from sunrise by treatment",
        x = "Theoretical quantiles",
        y = "Sample quantiles",
        caption = "y-intercept is mean hours from sunrise, slope is standard deviation")
```




```
t.test(formula = sunrise_hours$hours.from.sunrise ~
    sunrise_hours$treatment)
```

```
##
##  Welch Two Sample t-test
##
## data:  sunrise_hours$hours.from.sunrise by sunrise_hours$treatment
## t = 0.79351, df = 287.65, p-value = 0.4281
## alternative hypothesis: true difference in means between group control and group urban noise is not
## 95 percent confidence interval:
##  -0.2557585  0.6012757
## sample estimates:
##     mean in group control mean in group urban noise
##                  4.926966                  4.754207
```

E3 Hours fromn Sunrise Summary:

Based on our graph comparing the means of the hours after sunrise by treatment condition, we can see that the means for each condition are roughly the same. Based on our histograms and Q-Q plots, we can see that the control condition has two peaks, while the urban noise condition is roughly normally distributed. Although the shape of the data isn't perfectly normal for the control condition, we ran independent t-tests here because our data consists of continuous variables, our observations are independent along with each group being independent. Based on our independent t-tests, we can see that there is not a statistically significant difference between vigilance rates in the control and noise treatments for the GT bird species [t(287)) = 0.79, p = 0.42]. The mean vigilance rate is likely to vary by -0.2 to 0.6 (95% CI).