HW5 Hints

COSC6323/Spring 2024

2024-02-23

Contents

The template for homework

Summary Statistics
Data Visualization
Check for Outliers (no extreme outliers found)
Test - ANOVA
Post-hoc Tests
Visualization of Full Analysis

400 Sample from each day

```
## 'data.frame':    14400 obs. of  4 variables:
##  $ Day    : Factor w/ 3 levels "non-mission",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ HR     : int  73 72 78 73 61 87 66 87 81 72 ...
##  $ Speed  : num  4.193 15.276 1.103 1.648 0.535 ...
##  $ Cadence: int  55 0 47 0 0 0 0 61 0 0 ...
```

ANOVA HR

```
library(rstatix) #
```

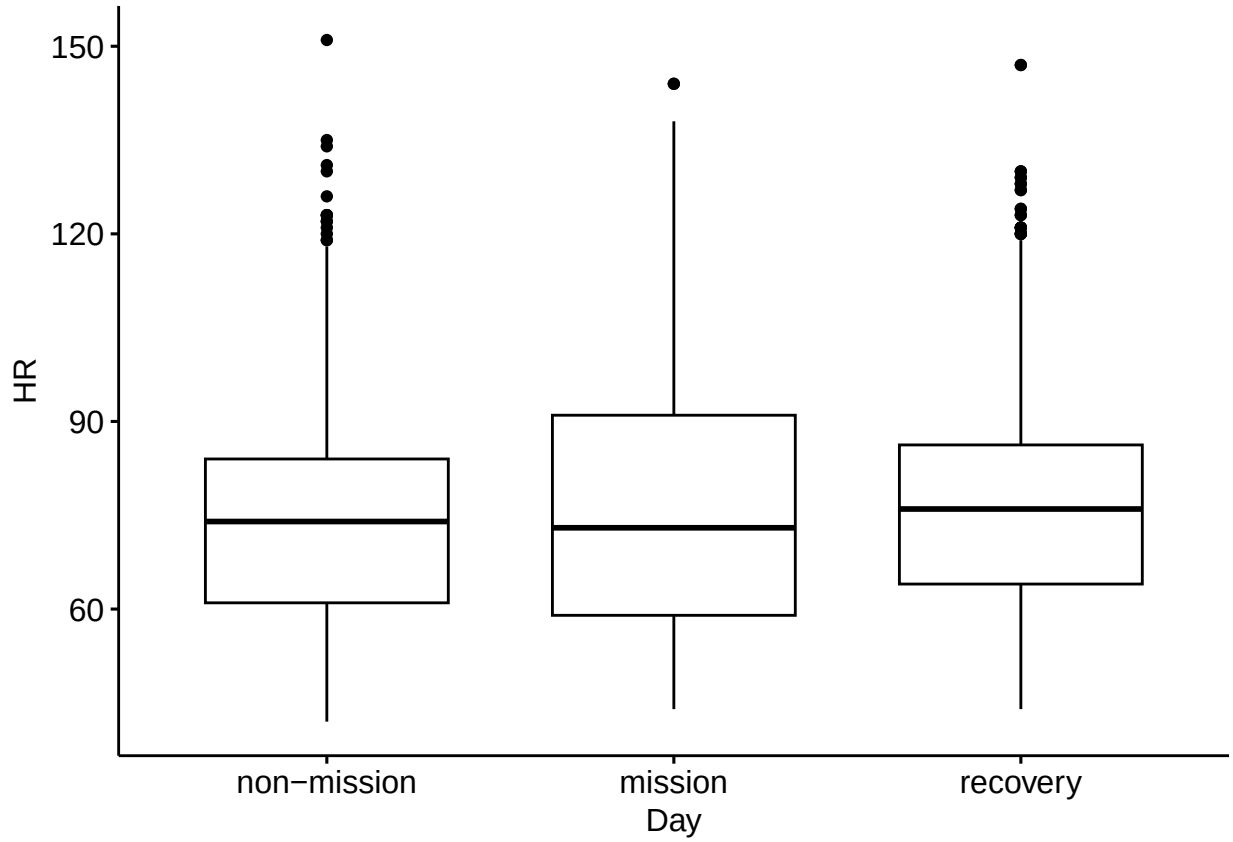
```
##
## Attaching package: 'rstatix'
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
#Summary Statistics
data %>%
  group_by(Day) %>%
  get_summary_stats(HR)
```

```
## # A tibble: 3 x 14
##   Day      variable     n   min   max median    q1    q3   iqr   mad  mean    sd
##   <fct>    <fct>    <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 non-mis~ HR        4800    42   151     74    61  84    23    16.3  73.4  15.2
## 2 mission  HR        4800    44   144     73    59  91    32    23.7  75.2  18.9
## 3 recovery HR        4800    44   147     76    64  86.2  22.2  16.3  76.1  15.3
## # i 2 more variables: se <dbl>, ci <dbl>
```

```
#Data Visualization

data %>%
  group_by(Day) %>%
  ggboxplot( x = "Day", y = "HR")
```




```
#Check for Outliers (no extreme outliers found)
data %>%
  group_by(Day) %>%
  identify_outliers(HR)
```

```
## # A tibble: 39 x 6
##    Day            HR Speed Cadence is.outlier is.extreme
##    <fct>       <int> <dbl>   <int> <lgl>      <lgl>
##  1 non-mission   120  2.65      47 TRUE       FALSE
##  2 non-mission   122  3.23      33 TRUE       FALSE
##  3 non-mission   131  3.86      55 TRUE       FALSE
##  4 non-mission   130  3.64      50 TRUE       FALSE
##  5 non-mission   151  4.05      47 TRUE       FALSE
##  6 non-mission   135  4.28      48 TRUE       FALSE
##  7 non-mission   119  5.72      61 TRUE       FALSE
##  8 non-mission   123 13.2       62 TRUE       FALSE
##  9 non-mission   123  3.02       0 TRUE       FALSE
## 10 non-mission   134  5.26      48 TRUE       FALSE
## # i 29 more rows
```

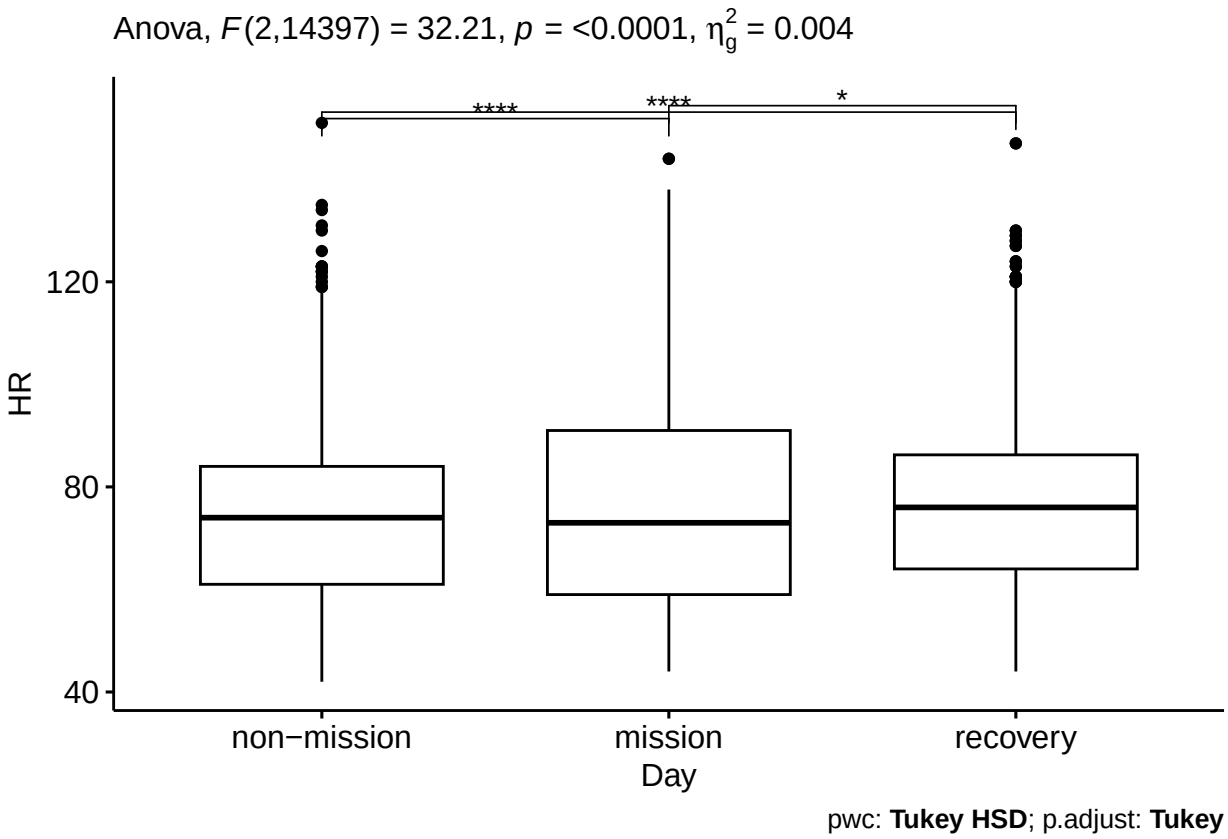
```
?anova_test()

# Independent measures ANOVA:

result.aov_HR <- anova_test(HR ~ Day, data = data)
result.aov_HR
```

```
## ANOVA Table (type II tests)
##
##   Effect DFn   DFd      F        p p<.05   ges
## 1    Day   2 14397 32.206 1.11e-14     * 0.004
```

```
# H0: The means of the groups are equal
# H1: At least one group mean is different

# Now we can see that the p-value is less than 0.05, so we can reject the null hypothesis and conclude

# Then run a post-hoc test to determine which groups are different from each other.

#Post-hoc Tests
OWA_pwc_HR <- data %>% tukey_hsd(HR ~ Day)
OWA_pwc_HR
```

```
## # A tibble: 3 x 9
##   term  group1      group2   null.value estimate conf.low conf.high        p.adj
## * <chr> <chr>       <chr>         <dbl>    <dbl>    <dbl>     <dbl>        <dbl>
## 1 Day   non-mission mission           0    1.80    1.00        2.59      3.23e-7
## 2 Day   non-mission recovery          0    2.66    1.86        3.45      6.07e-9
## 3 Day   mission     recovery          0    0.860   0.0680      1.65      2.94e-2
## # i 1 more variable: p.adj.signif <chr>
```

```
#Visualization of Full Analysis
OWA_pwc_HR <- OWA_pwc_HR %>% add_xy_position(x = "Day")
ggboxplot(data, x = "Day", y = "HR") +
  stat_pvalue_manual(OWA_pwc_HR, hide.ns = T) +
  labs(subtitle = get_test_label(result.aov_HR, detailed = TRUE),
       caption = get_pwc_label(OWA_pwc_HR))
```



What is the issue here with samples?

The issue here is that data is not normally distributed.

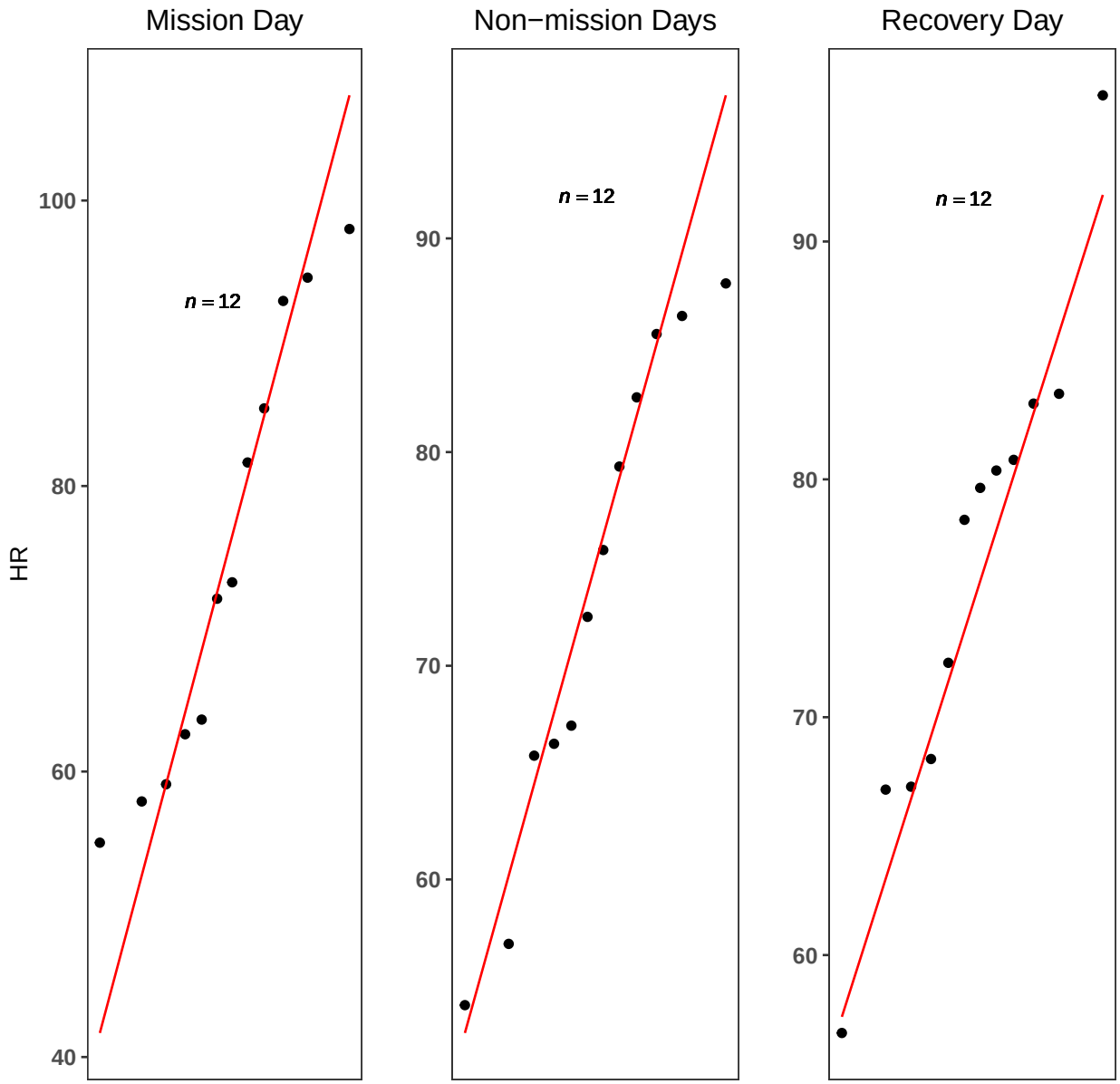
The p-value of the Shapiro-Wilk test is less than 0.05, so we can reject the null hypothesis and conclude that the data is not normally distributed. This is a problem because the ANOVA test assumes that the data is normally distributed.

What we'll try to make the data normally distributed is to take the log of the data and then run the ANOVA test again.

What is the best way to handle this issue?

Q1 - Normality test and Q-Q Plots with Means

HR normality test



Normality Check with bestNormalize library on

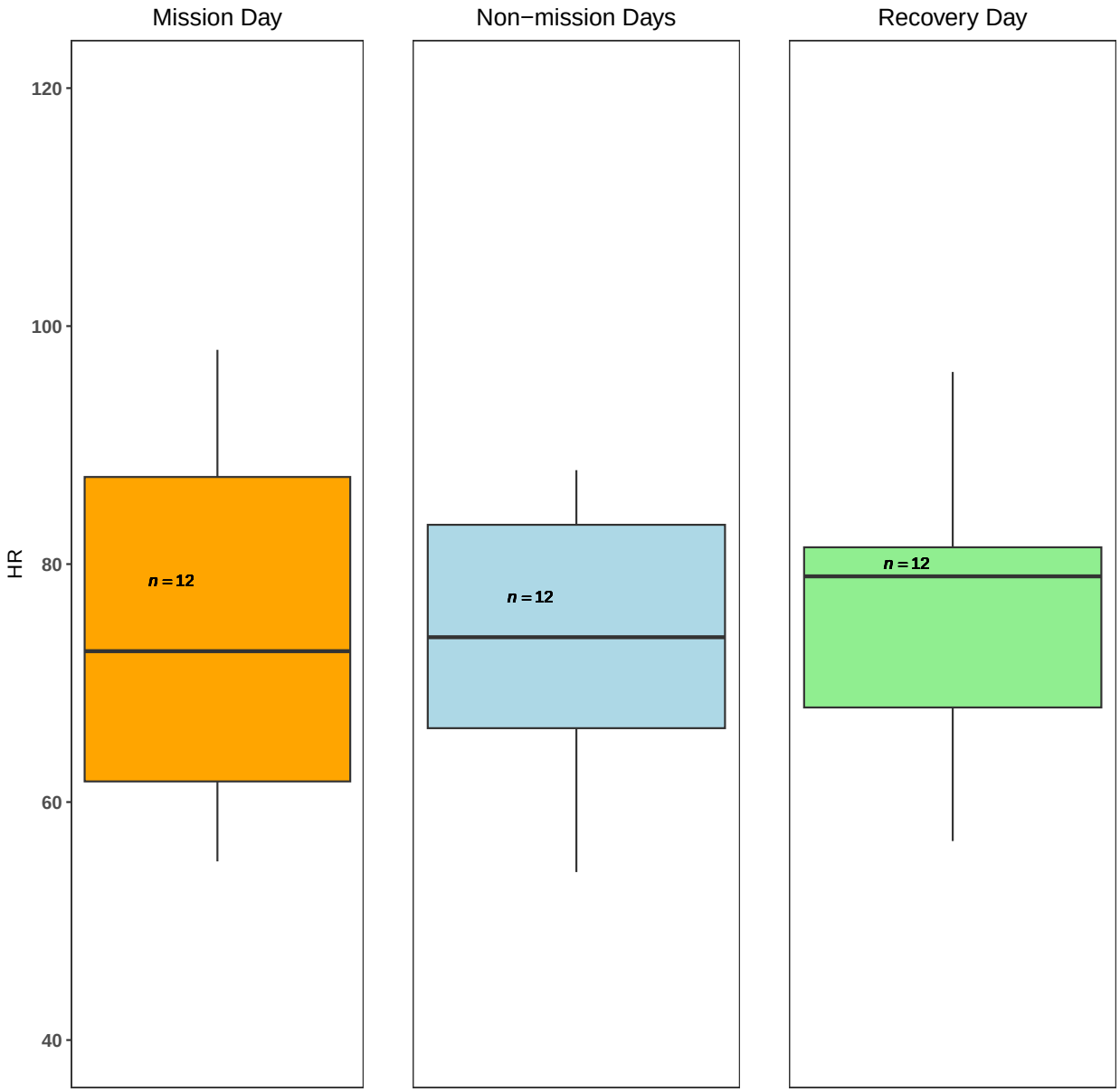
The best optimally transformed values

```
##
##  Shapiro-Wilk normality test
##
## data:  mis_bn_HR$x.t
## W = 0.99806, p-value = 1
```

```
##
##  Shapiro-Wilk normality test
##
## data:  non_bn_HR$x.t
## W = 0.99806, p-value = 1
```

```
##
##  Shapiro-Wilk normality test
##
## data:  rec_bn_HR$x.t
## W = 0.99806, p-value = 1
```

Box plots on sample




```
## 'data.frame':    36 obs. of  4 variables:
##  $ ID    : Factor w/ 6 levels "T004","T005",..: 1 1 2 2 3 3 4 4 5 5 ...
##  $ Period: chr  "Afternoon" "Morning" "Afternoon" "Morning" ...
##  $ Day   : Factor w/ 5 levels "DC1","DC2","DC3",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ HR    : num  73.3 63.6 98 93 85.4 ...
```

```
## [1] "DC1" "DC2" "DC3" "DC4" "DNM"
```

```
## [1] "DC2" "DC4" "DNM"
```

```
## 'data.frame':    36 obs. of  4 variables:
##  $ ID    : Factor w/ 6 levels "T004","T005",..: 1 1 2 2 3 3 4 4 5 5 ...
##  $ Period: Factor w/ 2 levels "Afternoon","Morning": 1 2 1 2 1 2 1 2 1 2 ...
##  $ Day   : Factor w/ 3 levels "mission","recovery",..: 1 1 1 1 1 1 1 1 1 1 ...
```

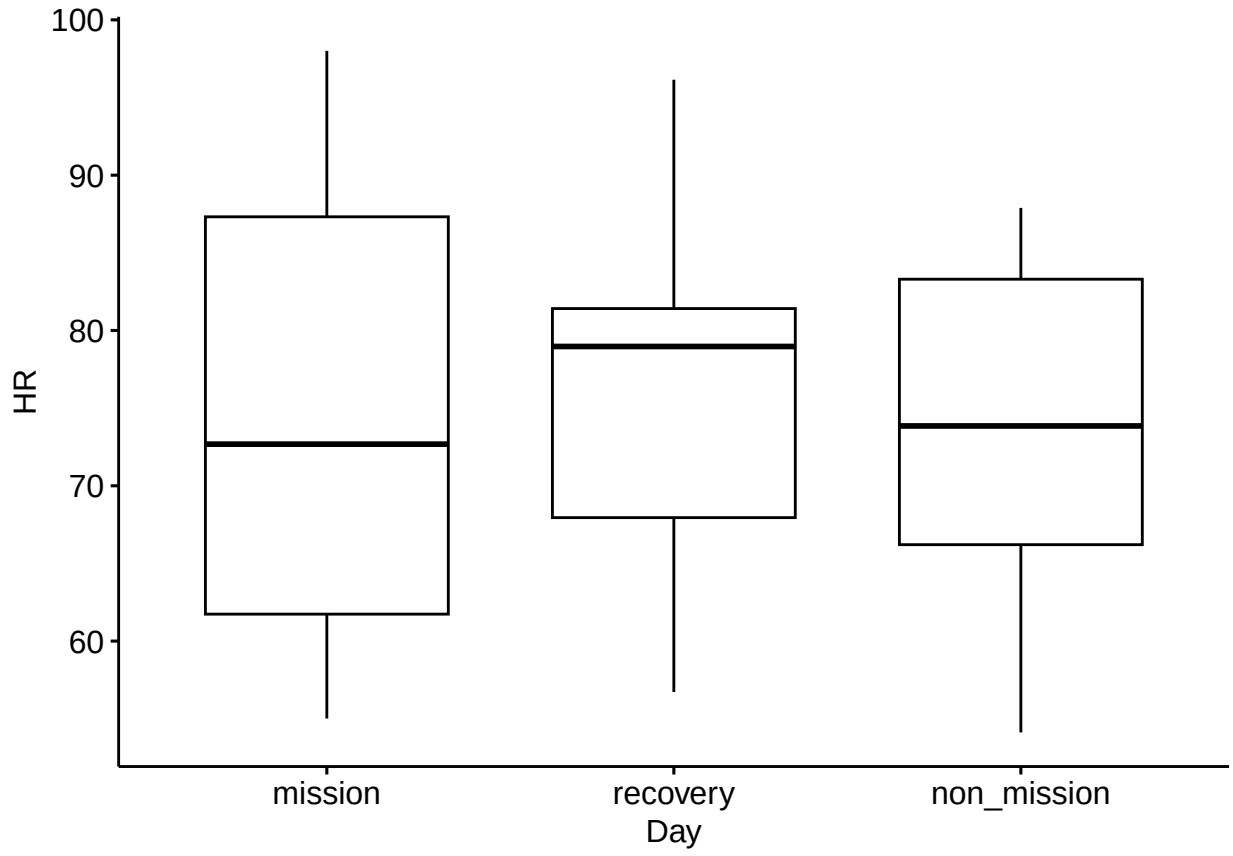
```
##  $ HR     : num  73.3 63.6 98 93 85.4 ...
```

ANOVA HR

```
#Summary Statistics
data %>%
  group_by(Day) %>%
  get_summary_stats(HR)
```

```
## # A tibble: 3 x 14
##   Day       variable     n   min   max median    q1    q3   iqr   mad  mean    sd
##   <fct>     <fct>    <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 mission   HR          12  55.0  98.0   72.7  61.7  87.3  25.6 19.5   74.7  15.5
## 2 recovery HR           12  56.7  96.1   79.0  68.0  81.4  13.5  8.38  76.1  10.4
## 3 non_mis~ HR           12  54.1  87.9   73.9  66.2  83.3  17.1 12.4   73.3  11.4
## # i 2 more variables: se <dbl>, ci <dbl>
```

```
#Data Visualization

data %>%
  group_by(Day) %>%
  ggboxplot( x = "Day", y = "HR")
```




```
#Check for Outliers (no extreme outliers found)
data %>%
  group_by(Day) %>%
  identify_outliers(HR)
```

```
## [1] Day        ID         Period     HR         is.outlier is.extreme
## <0 rows> (or 0-length row.names)
```

```r
#ANOVA
result.aov_HR <- anova_test(HR ~ Day, data = data)
result.aov_HR
```

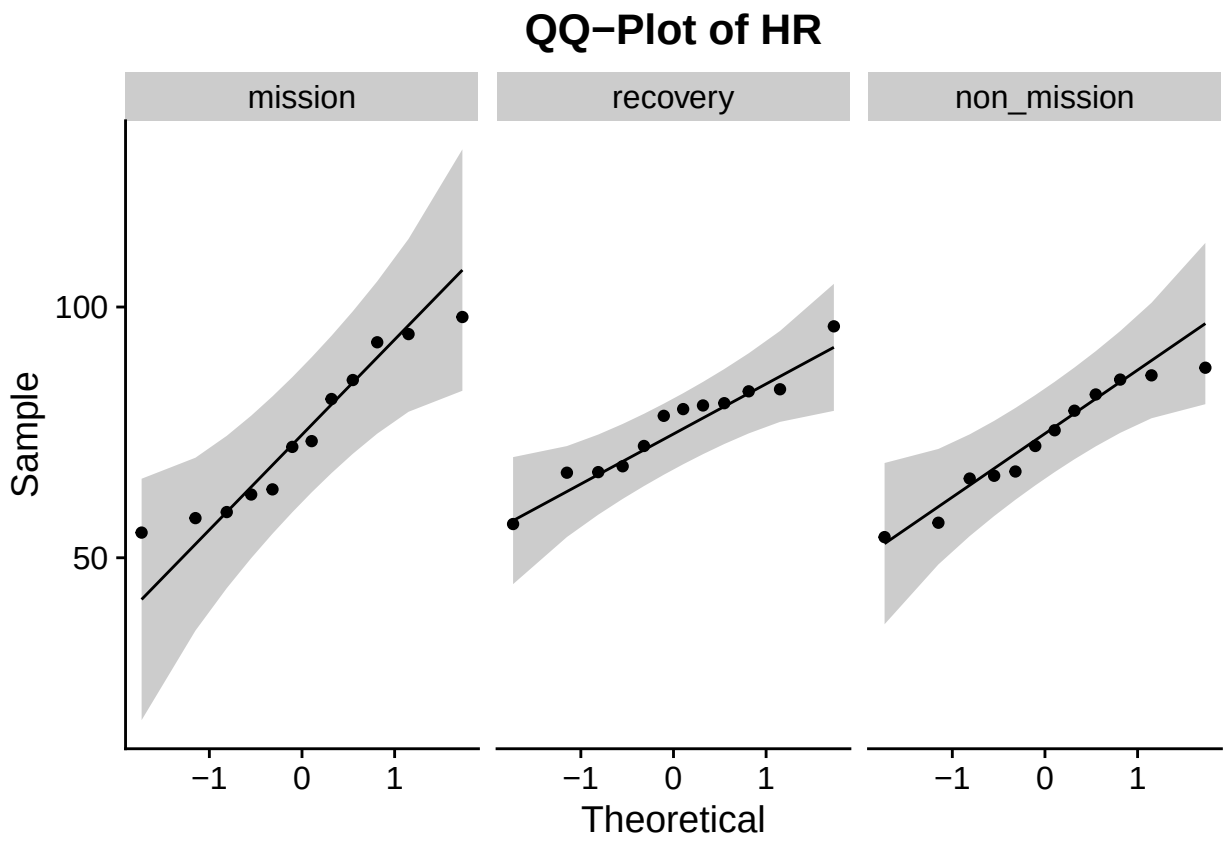
```
## ANOVA Table (type II tests)
##
##   Effect DFn DFd     F     p p<.05   ges
## 1    Day   2  33 0.147 0.864       0.009
```

```r
# result.aov_HR <- anova_test(HR ~ Period, data = data)
# result.aov_HR


# There is no differences between the groups (p = 0.05),
# no need to perform post-hoc tests to see which one is bigger :)
```

Q2 -Repeated measures ANOVA: within-Subjects designs

QQ-Plot

```
ggqqplot(data, "HR", facet.by = "Day", title = "QQ-Plot of HR") +
  theme_cowplot() + theme(plot.title = element_text(hjust = 0.5))
```




```
 #theme_classic()
```

```
str(data)
```

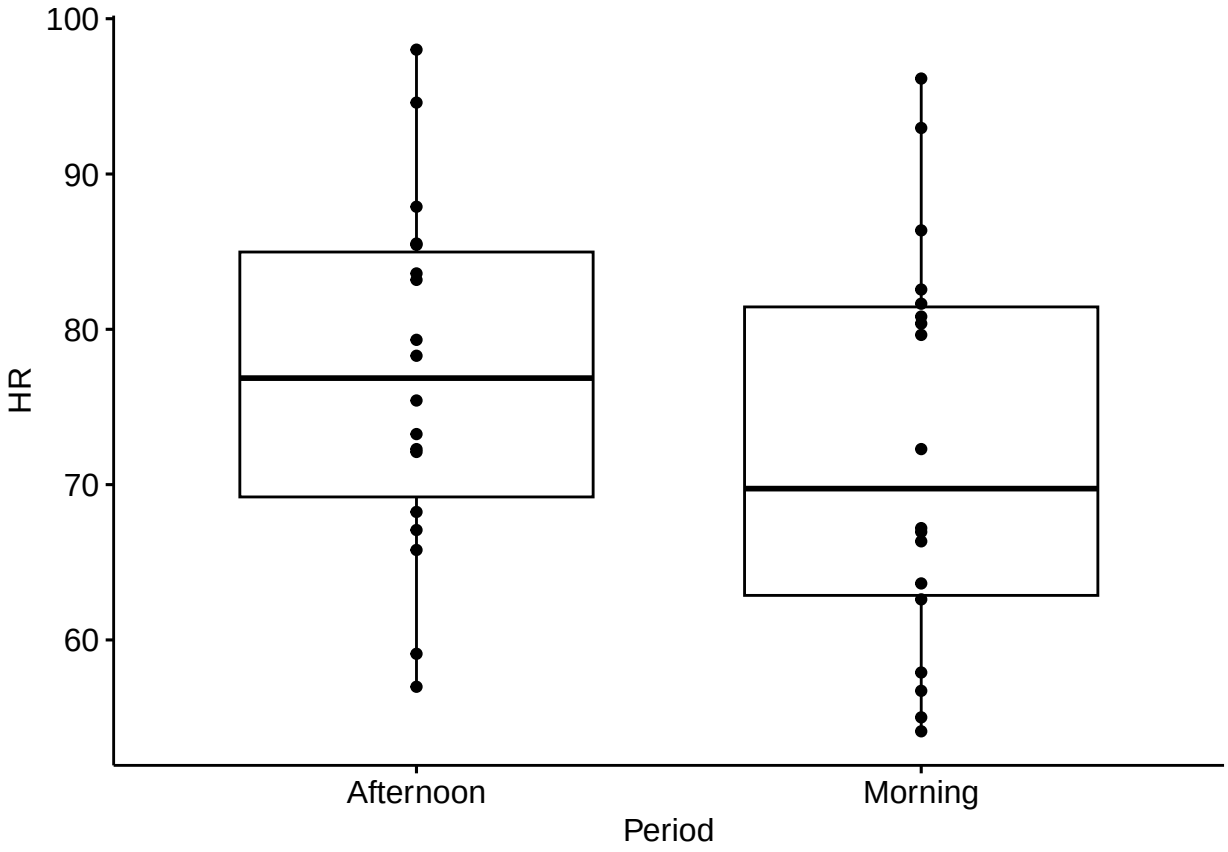
```
## 'data.frame':    36 obs. of  4 variables:
##  $ ID    : Factor w/ 6 levels "T004","T005",..: 1 1 2 2 3 3 4 4 5 5 ...
##  $ Period: Factor w/ 2 levels "Afternoon","Morning": 1 2 1 2 1 2 1 2 1 2 ...
##  $ Day   : Factor w/ 3 levels "mission","recovery",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ HR    : num  73.3 63.6 98 93 85.4 ...
```

```
#Summary Statistics
data %>%
  group_by(Day, Period) %>%
  get_summary_stats(HR)
```

```
## # A tibble: 6 x 15
##   Period   Day   variable     n   min   max median    q1    q3   iqr   mad  mean
##   <fct>    <fct> <fct>    <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Afterno~ miss~ HR           6  59.1  98.0   79.3  72.4  92.3  19.9 16.7   80.4
## 2 Morning  miss~ HR           6  55.0  93.0   63.1  59.1  77.1  18.1  9.88  69.0
## 3 Afterno~ reco~ HR           6  67.1  83.6   75.3  69.3  82.0  12.7 11.1   75.4
```

```
## 4 Morning  reco~ HR           6  56.7  96.1   80.0  70.1  80.7  10.6 10.3   76.8
## 5 Afterno~ non_~ HR           6  57.0  87.9   77.4  68.2  84.0  15.8 13.8   75.2
## 6 Morning  non_~ HR           6  54.1  86.4   69.7  66.6  80.0  13.4 12.0   71.5
## # i 3 more variables: sd <dbl>, se <dbl>, ci <dbl>
```

```
#Data Visualization

RM_bxp = data %>%
  #group_by(Day, Period) %>%
  group_by(Day) %>%
  ggboxplot( x = "Period", y = "HR" , add = "point")

RM_bxp
```




```
#Check for Outliers (no extreme outliers found)
data %>%
  group_by(Day, Period) %>%
  identify_outliers(HR)
```

```
## [1] Period     Day        ID         HR         is.outlier is.extreme
## <0 rows> (or 0-length row.names)
```

```
#ANOVA
?anova_test()


# 2  way to do it
```

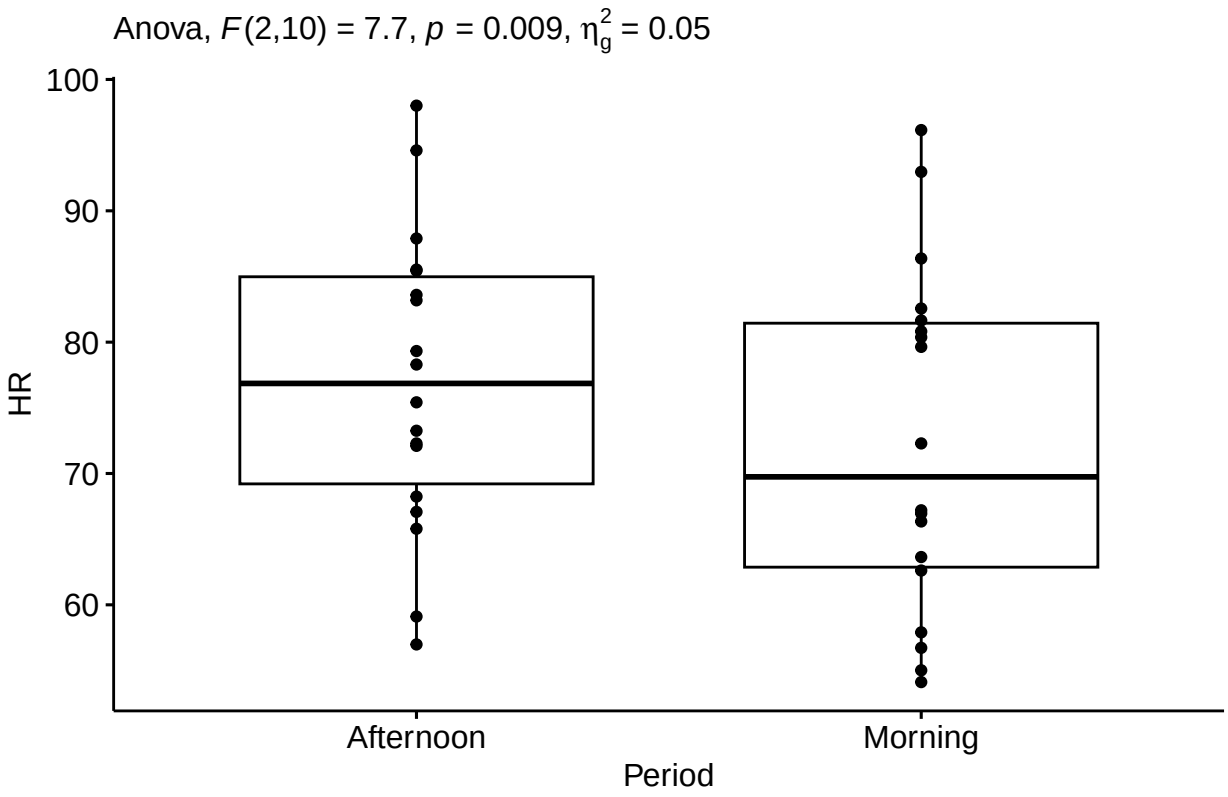
```r
result.aov_HR <- data %>%
  anova_test(dv = HR, wid = ID, within = c(Day, Period))


# since Main effect of Day is not significant,
# we can ignore the interaction between Day and Period
# no need to perform post-hoc tests to see which one is bigger :)
```

```r
# BUT if you like

#Post-hoc Tests
RM_pwc_HR <- data %>%
  pairwise_t_test(
    HR ~ Day, paired = TRUE,
    p.adjust.method = "bonferroni"
  )
RM_pwc_HR
```

```
## # A tibble: 3 x 10
##   .y.   group1   group2        n1    n2 statistic    df     p p.adj p.adj.signif
## * <chr> <chr>    <chr>      <int> <int>     <dbl> <dbl> <dbl> <dbl> <chr>
## 1 HR    mission  recovery      12    12    -0.400    11 0.697     1 ns
## 2 HR    mission  non_missi~    12    12     0.653    11 0.527     1 ns
## 3 HR    recovery non_missi~    12    12     0.905    11 0.385     1 ns
```

```r
#Visualization of Full Analysis
RM_pwc_HR <- RM_pwc_HR %>% add_xy_position(x = "time")
RM_bxp +
  stat_pvalue_manual(RM_pwc_HR) +
  labs(subtitle = get_test_label(result.aov_HR, detailed = TRUE),
    caption = get_pwc_label(RM_pwc_HR)
  )
```

```
## Warning: Removed 3 rows containing non-finite values (`stat_bracket()`).
```



pwc: **T test**; p.adjust: **Bonferroni**

Testing

```
## # A tibble: 3 x 8
##   Day         Effect   DFn   DFd      F     p `p<.05`   ges
## * <fct>       <chr>  <dbl> <dbl>  <dbl> <dbl> <chr>   <dbl>
## 1 mission     Period     1     5 16.6   0.01  "*"     0.15
## 2 recovery    Period     1     5  0.152 0.712 ""      0.004
## 3 non_mission Period     1     5  5.63  0.064 ""      0.028
```

```
## ANOVA Table (type II tests)
##
##   Effect DFn DFd     F     p p<.05   ges
## 1 Period   1  34 1.269 0.268       0.036
```