Experimental Data Analysis

Savannah Weaver

July 2021

Contents

Packages

define not-in function:

```
`%nin%` = Negate(`%in%`)
```

Data

This data was collected in the Spring of 2021 in conjunction with a study carried out in Cal Poly's Herpetology class. Some lizards measured for that primary study were kept to observe physiological changes in response to different climate treatments. See (doi) for full details

Morphometrics & Hydration

Treatment Groups

variables: - individual lizard ID - temp_tmt_C = temperature treatment - humidity_tmt_percent = humidity treatment (high/low, not actually %) - trial_number = which set of lizards that individual was from - conclusion = how that individual's experiment ended (died, canceled, or complete)

```
tmts <- read.csv("./data/exp_tmt_assignment.csv")
```

Capture Data

variables: - date = date of capture & baseline measurements - individual lizard ID - mass_g = mass in grams - hematocrit_percent = % of blood sample that's red blood cells - osmolality_mmol_kg = concentration of solutes in blood plasma - type = when the measurements were taken along the course of the experiment (all on capture day)

```
capture_hydration <- read.csv("./exported_data/capture_hydration.csv",
                              na.strings=c("","NA") # fix empty cells
                              ) %>%
  mutate(# correctly format date-only variable
         date = as.Date(date, format = "%Y-%m-%d")
         ) %>%
  # select only relevant variables
  dplyr::select(date, individual_ID,
                mass_g, hematocrit_percent, osmolality_mmol_kg
                ) %>%
  dplyr::filter(individual_ID %in% tmts$individual_ID) %>%
  mutate(type = as.factor("capture"))
summary(capture_hydration)
```

```
##       date            individual_ID        mass_g      hematocrit_percent
##  Min.   :2021-04-19   Min.   : 31.00   Min.   : 8.20   Min.   :16.00
##  1st Qu.:2021-04-26   1st Qu.: 57.00   1st Qu.:11.10   1st Qu.:32.75
##  Median :2021-04-26   Median : 78.00   Median :12.65   Median :36.00
##  Mean   :2021-04-29   Mean   : 77.46   Mean   :12.18   Mean   :36.08
##  3rd Qu.:2021-05-03   3rd Qu.: 98.25   3rd Qu.:13.32   3rd Qu.:39.00
##  Max.   :2021-05-10   Max.   :122.00   Max.   :15.00   Max.   :54.00
##  osmolality_mmol_kg      type
##  Min.   :319.0      capture:52
##  1st Qu.:354.2
##  Median :373.0
##  Mean   :373.7
##  3rd Qu.:392.2
##  Max.   :423.0
```

extract SVL data separately from capture data:

```
SVL <- read.csv("./exported_data/capture_hydration.csv",
                              na.strings=c("","NA") # fix empty cells
                              ) %>%
  dplyr::select(individual_ID, SVL_mm) %>%
  dplyr::filter(individual_ID %in% tmts$individual_ID)
summary(SVL)
```

```
##  individual_ID        SVL_mm
##  Min.   : 31.00   Min.   :59.00
```

```
##  1st Qu.: 57.00   1st Qu.:65.00
##  Median : 78.00   Median :68.00
##  Mean   : 77.46   Mean   :67.62
##  3rd Qu.: 98.25   3rd Qu.:70.00
##  Max.   :122.00   Max.   :73.00
```

extract capture CEWL cloacal temperature separately:

```
cap_CT <- read.csv("./exported_data/capture_hydration.csv",
                           na.strings=c("","NA") # fix empty cells
                           ) %>%
  dplyr::select(individual_ID, cloacal_temp_C) %>%
  dplyr::filter(individual_ID %in% tmts$individual_ID)
summary(cap_CT)
```

```
##  individual_ID    cloacal_temp_C
##  Min.   : 31.00   Min.   :20.00
##  1st Qu.: 57.00   1st Qu.:22.00
##  Median : 78.00   Median :24.00
##  Mean   : 77.46   Mean   :23.68
##  3rd Qu.: 98.25   3rd Qu.:25.00
##  Max.   :122.00   Max.   :28.00
##                   NA's   :2
```

Experiment Data

variables: - date = date of measurements - individual lizard ID - mass_g = mass in grams - hematocrit_percent = % of blood sample that's red blood cells - osmolality_mmol_kg = concentration of solutes in blood plasma (mean of 1-3 replicates) - type = when the measurements were taken along the course of the experiment (either during experimental treatment or after rehab)

```
exp_dat <- read.csv("./data/experimental_data.csv",
                           na.strings=c("","NA") # fix empty cells
                           ) %>%
                # format date
  dplyr::mutate(date = as.Date(date, format = "%m/%d/%y"),
                type = as.factor(type)
                ) %>%
  # select only variables to be analyzed
  dplyr::select(date, individual_ID, mass_g,
                hematocrit_percent, type,
                osmolality_mmol_kg = osmolality_mmol_kg_replicate_mean)
summary(exp_dat)
```

```
##       date            individual_ID        mass_g       hematocrit_percent
##  Min.   :2021-04-21   Min.   : 31.00   Min.   : 6.700   Min.   :12.0
##  1st Qu.:2021-04-28   1st Qu.: 51.25   1st Qu.: 9.875   1st Qu.:23.0
##  Median :2021-05-07   Median : 87.50   Median :11.250   Median :28.0
##  Mean   :2021-05-06   Mean   : 77.85   Mean   :11.076   Mean   :27.8
##  3rd Qu.:2021-05-14   3rd Qu.:101.25   3rd Qu.:12.225   3rd Qu.:33.0
##  Max.   :2021-05-20   Max.   :122.00   Max.   :14.700   Max.   :43.0
##                                                         NA's   :19
##     type    osmolality_mmol_kg
##  exp  :98   Min.   :298.0
##  rehab:34   1st Qu.:342.0
##             Median :355.0
```

```
##              Mean   :360.1
##              3rd Qu.:374.8
##              Max.   :441.0
##              NA's   :22
```

Join Dataframes

Now, attach all the dataframes, only use individuals whose treatment was completed, and add a "day" variable for what day of treatment each lizard/observation was on. I also calculate SMI using the equation created in capture_analysis.

```
all_dat <- exp_dat %>%
  # join data
  rbind(capture_hydration) %>%
  # add tmt group info
  left_join(tmts, by = "individual_ID") %>%
  dplyr::select(-notes) %>%
  # add SVL value for each obs of each indiv.
  # for computing BCI and scaled mass indices
  left_join(SVL, by = "individual_ID") %>%
  # only use completed experiment runs
  dplyr::filter(conclusion == "complete") %>%
  group_by(individual_ID) %>%
  # reformat a lot of variables
  mutate(capture_date = min(date),
         day = as.numeric(date - capture_date),
         humidity_tmt_percent = as.factor(humidity_tmt_percent),
         individual_ID = as.factor(individual_ID),
         temp_tmt_C = as.factor(temp_tmt_C),
         trial_number = as.factor(trial_number),
         conclusion = as.factor(conclusion),
         SMI = mass_g * ((65.02158/SVL_mm) ^ (3.09059/sqrt(0.8944)))
         ) %>%
  # in the first trial we took measurements every 2 days
  # exclude those obs to make consistent across trials
  dplyr::filter(day %nin% c(2,6))

summary(all_dat)
```

```
##       date            individual_ID     mass_g      hematocrit_percent
##  Min.   :2021-04-19   37     :  4   Min.   : 6.70   Min.   :12.00
##  1st Qu.:2021-05-01   39     :  4   1st Qu.:10.20   1st Qu.:24.00
##  Median :2021-05-10   40     :  4   Median :11.50   Median :30.00
##  Mean   :2021-05-07   49     :  4   Mean   :11.35   Mean   :29.95
##  3rd Qu.:2021-05-14   52     :  4   3rd Qu.:12.70   3rd Qu.:35.00
##  Max.   :2021-05-20   54     :  4   Max.   :15.00   Max.   :54.00
##                       (Other):115                   NA's   :12
##       type    osmolality_mmol_kg temp_tmt_C humidity_tmt_percent trial_number
##  exp    :70   Min.   :298.0      25:139     dry  :68             1:23
##  rehab  :34   1st Qu.:342.0                 humid:71             2:24
##  capture:35   Median :358.5                                      3:44
##               Mean   :362.4                                      4:48
##               3rd Qu.:379.5
##               Max.   :441.0
##               NA's   :15
```

```
##     conclusion      SVL_mm       capture_date             day
##  complete:139   Min.   :59.00   Min.   :2021-04-19   Min.   : 0.000
##                 1st Qu.:66.00   1st Qu.:2021-04-26   1st Qu.: 2.000
##                 Median :68.00   Median :2021-05-03   Median : 4.000
##                 Mean   :67.66   Mean   :2021-05-01   Mean   : 5.547
##                 3rd Qu.:70.00   3rd Qu.:2021-05-10   3rd Qu.: 9.000
##                 Max.   :73.00   Max.   :2021-05-10   Max.   :11.000
##
##       SMI
##  Min.   : 7.343
##  1st Qu.: 8.976
##  Median : 9.977
##  Mean   : 9.958
##  3rd Qu.:10.751
##  Max.   :13.970
##
```

```
unique(all_dat$individual_ID)
```

```
##  [1] 47  49  39  52  37  40  80  66  54  61  74  73  92  91  95  88  93  96  98
## [20] 89  99  81  97  104 108 122 118 109 113 105 114 101 117 102 103
## 35 Levels: 37 39 40 47 49 52 54 61 66 73 74 80 81 88 89 91 92 93 95 96 ... 122
```

re-order some factors:

```
all_dat$humidity_tmt_percent <- factor(all_dat$humidity_tmt_percent,
                                       levels = c("humid", "dry"),
                                       labels = c("Humid", "Dry"))
```

```
all_dat$day <- factor(all_dat$day,
                      levels = c("0", "4", "8", "9", "10", "11"),
                      labels = c("Before Experiment",
                                 "Mid Experiment",
                                 "After Experiment",
                                 "After Experiment",
                                 "After Rehydration",
                                 "After Rehydration"))
summary(all_dat)
```

```
##       date             individual_ID     mass_g      hematocrit_percent
##  Min.   :2021-04-19   37     :  4   Min.   : 6.70   Min.   :12.00
##  1st Qu.:2021-05-01   39     :  4   1st Qu.:10.20   1st Qu.:24.00
##  Median :2021-05-10   40     :  4   Median :11.50   Median :30.00
##  Mean   :2021-05-07   49     :  4   Mean   :11.35   Mean   :29.95
##  3rd Qu.:2021-05-14   52     :  4   3rd Qu.:12.70   3rd Qu.:35.00
##  Max.   :2021-05-20   54     :  4   Max.   :15.00   Max.   :54.00
##                       (Other):115                   NA's   :12
##       type    osmolality_mmol_kg temp_tmt_C humidity_tmt_percent trial_number
##  exp    :70   Min.   :298.0      25:139     Humid:71             1:23
##  rehab  :34   1st Qu.:342.0                 Dry  :68             2:24
##  capture:35   Median :358.5                                      3:44
##               Mean   :362.4                                      4:48
##               3rd Qu.:379.5
##               Max.   :441.0
##               NA's   :15
##     conclusion      SVL_mm       capture_date                        day
```

```
##  complete:139   Min.   :59.00   Min.   :2021-04-19   Before Experiment:35
##                 1st Qu.:66.00   1st Qu.:2021-04-26   Mid Experiment   :35
##                 Median :68.00   Median :2021-05-03   After Experiment :35
##                 Mean   :67.66   Mean   :2021-05-01   After Rehydration:34
##                 3rd Qu.:70.00   3rd Qu.:2021-05-10
##                 Max.   :73.00   Max.   :2021-05-10
##
##       SMI
##  Min.   : 7.343
##  1st Qu.: 8.976
##  Median : 9.977
##  Mean   : 9.958
##  3rd Qu.:10.751
##  Max.   :13.970
##
```

make a sub-dataframe without rehab data to prevent any mix-ups:

```
all_dat_no_rehab <- all_dat %>%
  dplyr::filter(type != "rehab")
```

Checks

Dates:

```
# check that capture dates are valid
unique(all_dat$capture_date)
```

```
## [1] "2021-04-19" "2021-04-26" "2021-05-03" "2021-05-10"
```

Check that each lizard has an accurate number of measurements.

```
all_dat %>%
  group_by(individual_ID, type) %>%
  summarise(n = n()) %>%
  arrange(type)
```

```
## `summarise()` regrouping output by 'individual_ID' (override with `.groups` argument)
```

```
## # A tibble: 104 x 3
## # Groups:   individual_ID [35]
##    individual_ID type      n
##    <fct>         <fct> <int>
##  1 37            exp       2
##  2 39            exp       2
##  3 40            exp       2
##  4 47            exp       2
##  5 49            exp       2
##  6 52            exp       2
##  7 54            exp       2
##  8 61            exp       2
##  9 66            exp       2
## 10 73            exp       2
## # ... with 94 more rows
```

That all looks good, every lizard has 1 capture measurement, 2 experimental measurements, and 1 rehab measurement.

CEWL

Capture CEWL

variables: - date = date of capture & baseline measurements - individual lizard ID - region = which body area the measurement was taken from - TEWL_g_m2h = evaporative water loss - cloacal_temp_C = taken at measurement; influences CEWL

```
cap_CEWL <- read.csv("./exported_data/capture_CEWL.csv") %>%
  dplyr::select(date, individual_ID, region, TEWL_g_m2h) %>%
  mutate(#individual_ID = as.factor(individual_ID), # do later
         date = as.Date(date, format = "%Y-%m-%d"),
         region = as.factor(region),
         day = as.factor("before"),
         n_day = 0
         ) %>%
  dplyr::filter(individual_ID %in% all_dat$individual_ID) %>%
  left_join(cap_CT, by = 'individual_ID')
summary(cap_CEWL)
```

```
##       date            individual_ID     region     TEWL_g_m2h        day
##  Min.   :2021-04-19   Min.   : 37.00   dewl:32   Min.   : 7.48   before:163
##  1st Qu.:2021-04-26   1st Qu.: 73.00   dors:33   1st Qu.:20.54
##  Median :2021-05-03   Median : 95.00   head:33   Median :27.43
##  Mean   :2021-05-02   Mean   : 87.46   mite:32   Mean   :29.30
##  3rd Qu.:2021-05-10   3rd Qu.:104.00   vent:33   3rd Qu.:36.91
##  Max.   :2021-05-10   Max.   :122.00             Max.   :62.94
##      n_day   cloacal_temp_C
##  Min.   :0   Min.   :20.00
##  1st Qu.:0   1st Qu.:22.00
##  Median :0   Median :24.00
##  Mean   :0   Mean   :23.84
##  3rd Qu.:0   3rd Qu.:25.00
##  Max.   :0   Max.   :28.00
```

Post-Experiment CEWL

In the future, I could automate this like I did for the HOBO data.

Load in each of the post-rehab datafiles:

```
# trial 1
CEWL_t1 <- read.csv("./data/post_exp_CEWL/4-28-21-CEWL.csv", # filename
                          na.strings=c("","NA")) %>% # fix empty cells
  # rename and select the pertinent variables/cols
  # I have to do this for each one
  # so they all have the same number of columns for joining
  dplyr::select(date = Date,
                Status,
                ID = Comments,
                TEWL_g_m2h = TEWL..g..m2h.. # rename
                )

# trial 2
CEWL_t2 <- read.csv("./data/post_exp_CEWL/5-4-21-CEWL.csv",
                          na.strings=c("","NA")) %>%
  dplyr::select(date = Date,
```

```
                Status,
                ID = Comments,
                TEWL_g_m2h = TEWL..g..m2h..
                )

# trial 3
CEWL_t3 <- read.csv("./data/post_exp_CEWL/5-11-21-CEWL.csv",
                          na.strings=c("","NA")) %>%
  dplyr::select(date = Date,
                Status,
                ID = Comments,
                TEWL_g_m2h = TEWL..g..m2h..
                )

# trial 4
CEWL_t4 <- read.csv("./data/post_exp_CEWL/5-18-21-CEWL.csv",
                          na.strings=c("","NA")) %>%
  dplyr::select(date = Date,
                Status,
                ID = Comments,
                TEWL_g_m2h = TEWL..g..m2h..
                )
```

Load in cloacal temperatures:

```
exp_CT <- read.csv("./data/post_exp_CEWL_cloacal_temps.csv") %>%
  mutate(date = as.Date(date, format = "%Y/%m/%d")) %>%
  dplyr::select(-time)
summary(exp_CT)
```

```
##       date            individual_ID    cloacal_temp_C
##  Min.   :2021-04-28   Min.   : 37.00   Min.   :19.0
##  1st Qu.:2021-05-04   1st Qu.: 69.50   1st Qu.:21.0
##  Median :2021-05-11   Median : 93.00   Median :23.0
##  Mean   :2021-05-09   Mean   : 85.91   Mean   :22.4
##  3rd Qu.:2021-05-18   3rd Qu.:103.50   3rd Qu.:23.0
##  Max.   :2021-05-18   Max.   :122.00   Max.   :26.0
```

Join Dataframes

Merge all post-experiment CEWL, add cloacal temperature, add capture CEWL:

```
# merge all CEWL datafiles & reformat
CEWL <- CEWL_t1 %>% # trial 1
  rbind(., CEWL_t2, # trial 2
        CEWL_t3, # trial 3
        CEWL_t4 # trial 4
        ) %>%
  # remove any unsuccessful measurements
  dplyr::filter(Status == "Normal") %>%
  # extract individual_ID and region separately from the "ID" variable
  separate(ID, c("individual_ID", "region")) %>%
  # reformat data
  dplyr::mutate(# reformat date
                date = as.Date(date, format = "%m/%d/%y"),
```

```
                # format individual ID
                individual_ID = as.integer(individual_ID),
                # set body region as a factor variable after getting only the consistent characters due
                region = as.factor(substring(region, 1, 4)),
                # add when measurement taken
                day = as.factor("after"),
                n_day = 1 # technically day 8/9, just to help with figures
                ) %>%
  # remove cols not relevant to stats
  dplyr::select(-Status) %>%
  # remove any rows with missing values
  # none actually needed to be removed
  dplyr::filter(complete.cases(.)) %>%
  # add cloacal temperatures
  left_join(exp_CT, by = c("date", "individual_ID")) %>%
  # now matching dataframes, add capture CEWL data
  rbind(cap_CEWL) %>%
  # add tmt assignments
  left_join(tmts, by = "individual_ID") %>%
  mutate(humidity_tmt_percent = as.factor(humidity_tmt_percent),
         individual_ID = as.factor(individual_ID),
         conclusion = as.factor(conclusion),
         trial_number = as.factor(trial_number)
         ) %>%
  # lizards 49 & 80 are missing pre-exp CEWL, so remove them
  dplyr::filter((individual_ID %nin% c('49', '80')))
# every lizard should have 10 measurements
summary(CEWL)
```

```
##       date             individual_ID  region     TEWL_g_m2h         day
##  Min.   :2021-04-19   37     : 10   dewl:65   Min.   :  4.60   after :163
##  1st Qu.:2021-05-03   39     : 10   dors:65   1st Qu.: 20.09   before:163
##  Median :2021-05-10   40     : 10   head:66   Median : 27.18
##  Mean   :2021-05-06   47     : 10   mite:64   Mean   : 30.69
##  3rd Qu.:2021-05-11   52     : 10   vent:66   3rd Qu.: 38.72
##  Max.   :2021-05-18   54     : 10             Max.   :106.38
##                       (Other):266
##      n_day     cloacal_temp_C    temp_tmt_C humidity_tmt_percent trial_number
##  Min.   :0.0   Min.   :19.00   Min.   :25   dry  :158            1: 50
##  1st Qu.:0.0   1st Qu.:21.00   1st Qu.:25   humid:168            2: 48
##  Median :0.5   Median :23.00   Median :25                        3:110
##  Mean   :0.5   Mean   :23.11   Mean   :25                        4:118
##  3rd Qu.:1.0   3rd Qu.:24.75   3rd Qu.:25
##  Max.   :1.0   Max.   :28.00   Max.   :25
##
##     conclusion     notes
##  complete:326   Length:326
##                 Class :character
##                 Mode  :character
##
##
##
##
```

Check that data looks correct:

```
CEWL %>%
  group_by(individual_ID, day) %>%
  summarise(n = n()) %>%
  arrange(individual_ID, n)
```

```
## `summarise()` regrouping output by 'individual_ID' (override with `.groups` argument)
```

```
## # A tibble: 66 x 3
## # Groups:   individual_ID [33]
##    individual_ID day        n
##    <fct>         <fct>  <int>
##  1 37            after      5
##  2 37            before     5
##  3 39            after      5
##  4 39            before     5
##  5 40            after      5
##  6 40            before     5
##  7 47            after      5
##  8 47            before     5
##  9 52            after      5
## 10 52            before     5
## # ... with 56 more rows
```

Everything looks great! (after removing the observations for the two lizards with missing pre-experiment CEWL measurements.)

Before/after aren't perfectly even because sometimes we were unable to get the AquaFlux to equilibrate and take a measurement.

Finally, make a small edit so the regions are spelled out completely. This requires reordering factor levels:

```
CEWL$region <- factor(CEWL$region,
                      levels = c("dors", "vent", "head", "dewl", "mite"),
                      labels = c("Dorsum", "Ventrum", "Head",
                                 "Dewlap", "Mite Patch")
                    )
CEWL$humidity_tmt_percent <- factor(CEWL$humidity_tmt_percent,
                                    levels = c("humid", "dry"),
                                    labels = c("Humid", "Dry"))
CEWL$day <- factor(CEWL$day,
                                    levels = c("before", "after"),
                                    labels = c("Before", "After"))
summary(CEWL)
```

```
##       date             individual_ID        region     TEWL_g_m2h
##  Min.   :2021-04-19   37     : 10   Dorsum    :65   Min.   :  4.60
##  1st Qu.:2021-05-03   39     : 10   Ventrum   :66   1st Qu.: 20.09
##  Median :2021-05-10   40     : 10   Head      :66   Median : 27.18
##  Mean   :2021-05-06   47     : 10   Dewlap    :65   Mean   : 30.69
##  3rd Qu.:2021-05-11   52     : 10   Mite Patch:64   3rd Qu.: 38.72
##  Max.   :2021-05-18   54     : 10                   Max.   :106.38
##                       (Other):266
##      day          n_day     cloacal_temp_C    temp_tmt_C humidity_tmt_percent
##  Before:163   Min.   :0.0   Min.   :19.00   Min.   :25   Humid:168
##  After :163   1st Qu.:0.0   1st Qu.:21.00   1st Qu.:25   Dry  :158
```

```
##               Median :0.5   Median :23.00   Median :25
##               Mean   :0.5   Mean   :23.11   Mean   :25
##               3rd Qu.:1.0   3rd Qu.:24.75   3rd Qu.:25
##               Max.   :1.0   Max.   :28.00   Max.   :25
##
##  trial_number    conclusion     notes
##  1: 50        complete:326   Length:326
##  2: 48                       Class :character
##  3:110                       Mode  :character
##  4:118
##
##
##
```

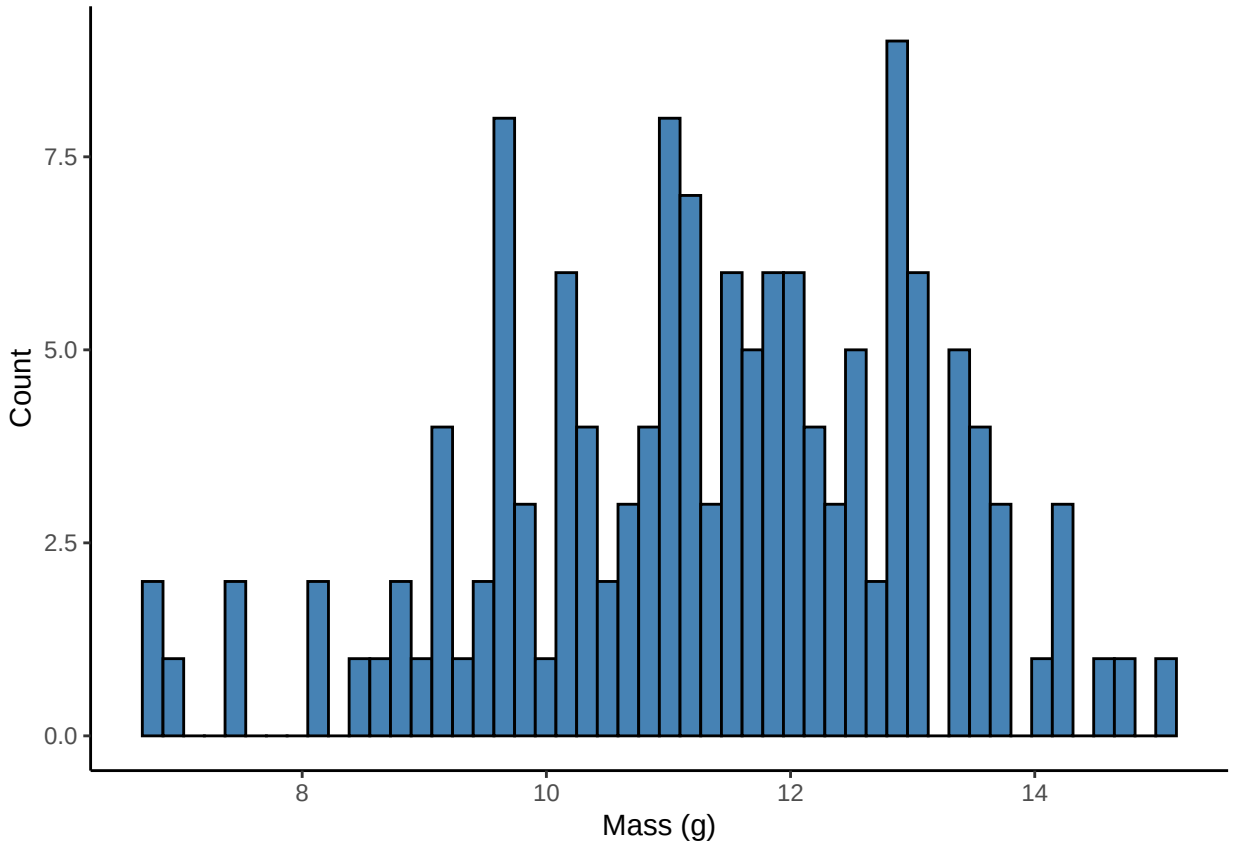
Export Data Frames for Power Analyses

```
#write.csv(all_dat, "exported_data/exp_effects_hydration.csv")
#write.csv(CEWL, "exported_data/exp_effects_CEWL.csv")
```

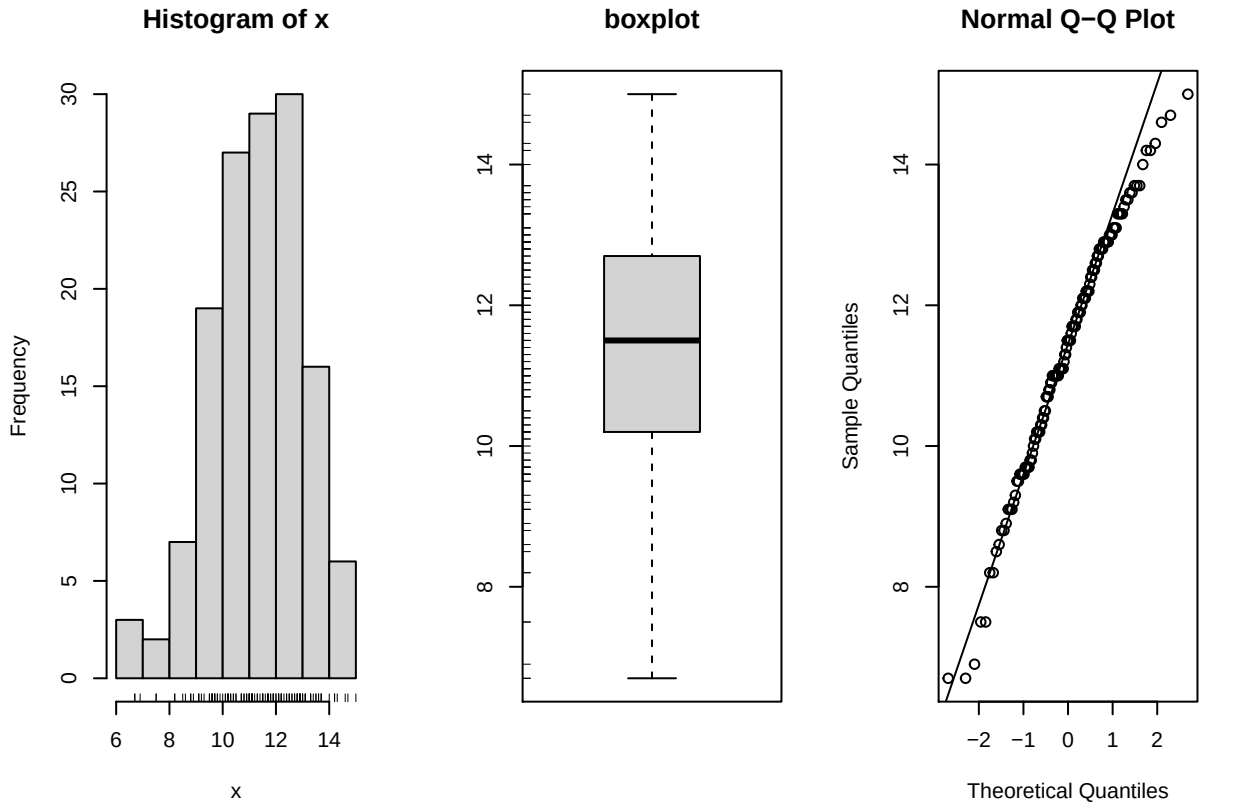
Data Distributions

Mass

```
all_dat %>%
  ggplot(., aes(x = mass_g)) +
  geom_histogram(color = "black", fill="steelblue", bins=50) +
  theme_classic() +
  xlab("Mass (g)") +
  ylab("Count")
```




```
simple.eda(all_dat$mass_g)
```



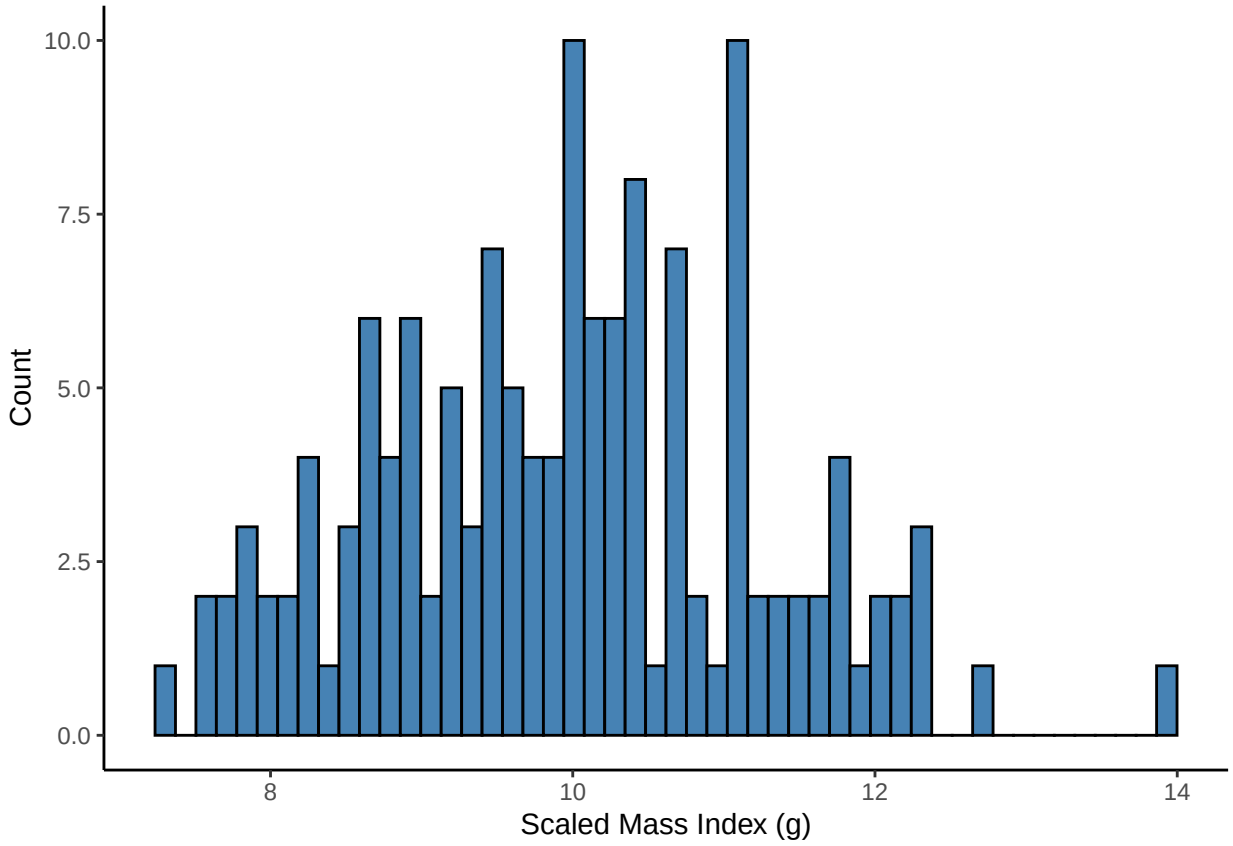
```
shapiro.test(all_dat$mass_g)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  all_dat$mass_g
## W = 0.98437, p-value = 0.1144
```

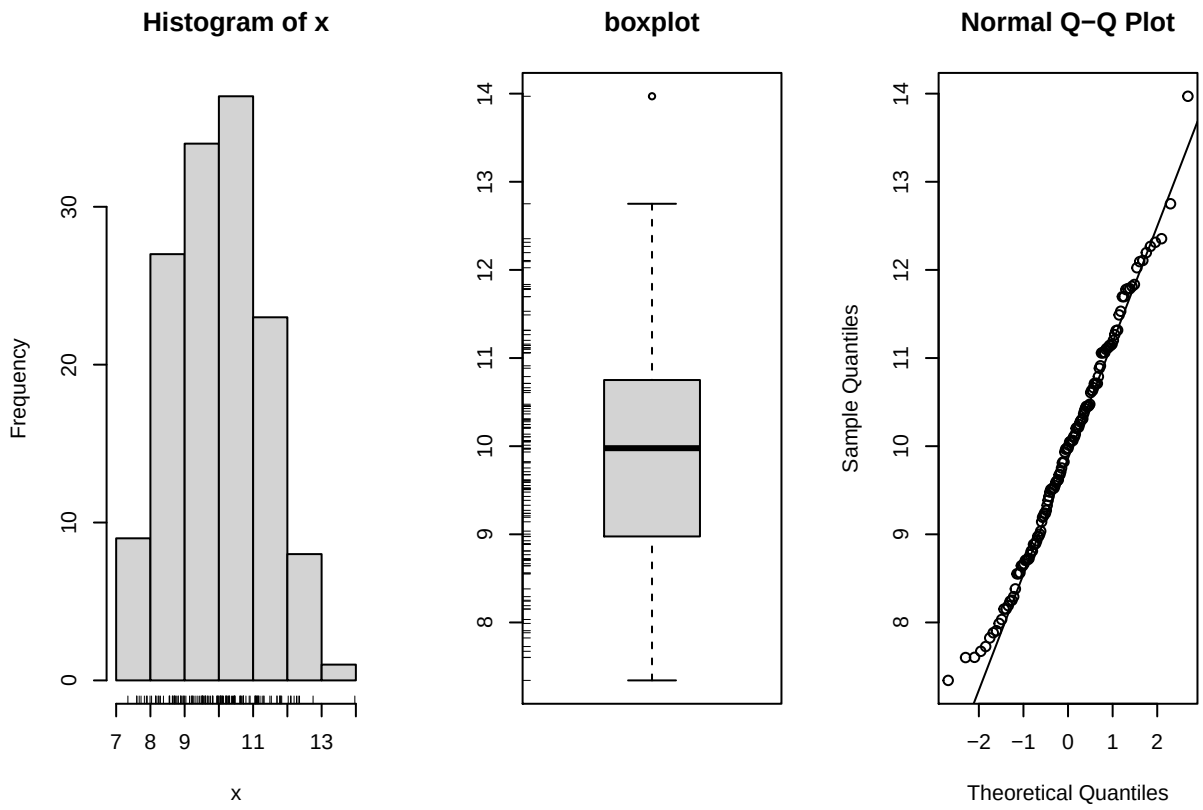
Mass distribution not normal, skewed to the left.

Scaled Mass Index

```
all_dat %>%
  ggplot(., aes(x = SMI)) +
  geom_histogram(color = "black", fill="steelblue", bins=50) +
  theme_classic() +
  xlab("Scaled Mass Index (g)") +
  ylab("Count")
```




```
simple.eda(all_dat$SMI)
```



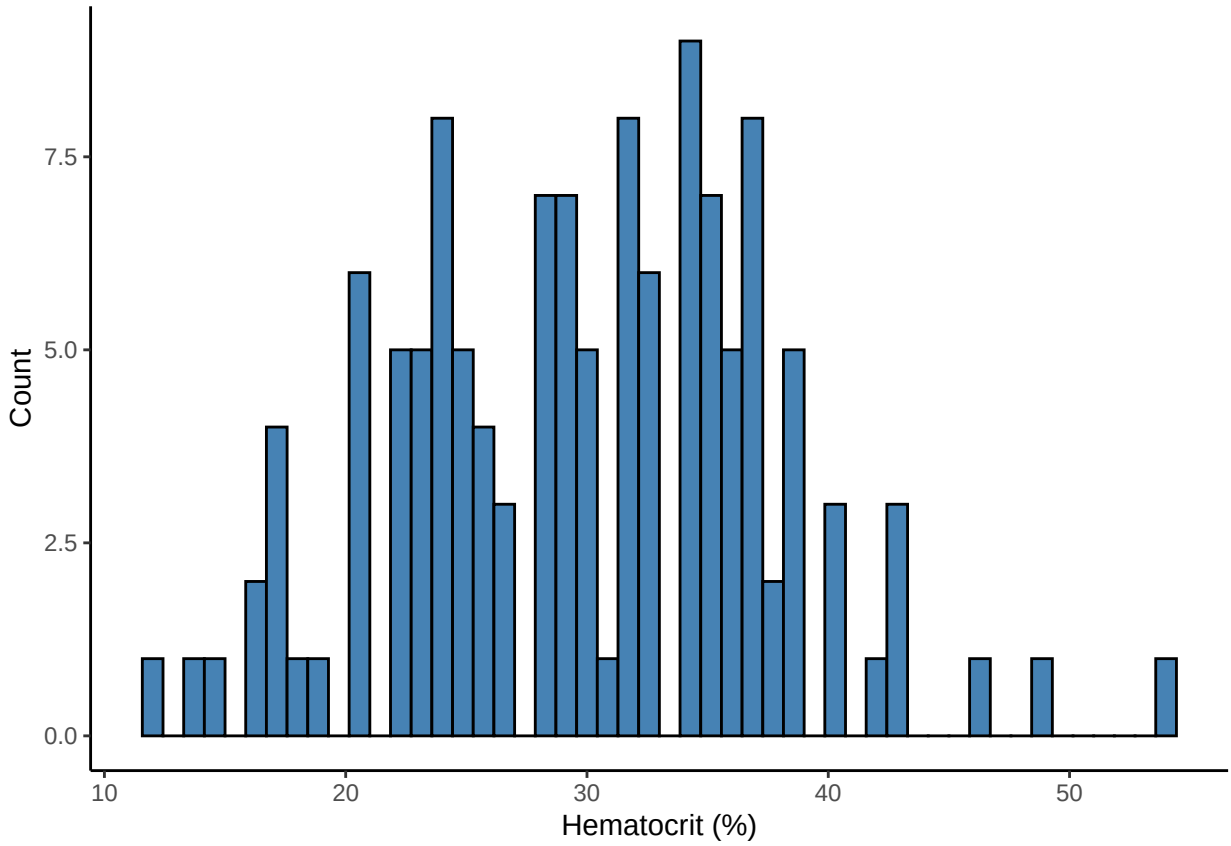

```
shapiro.test(all_dat$SMI)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  all_dat$SMI
## W = 0.98975, p-value = 0.4014
```

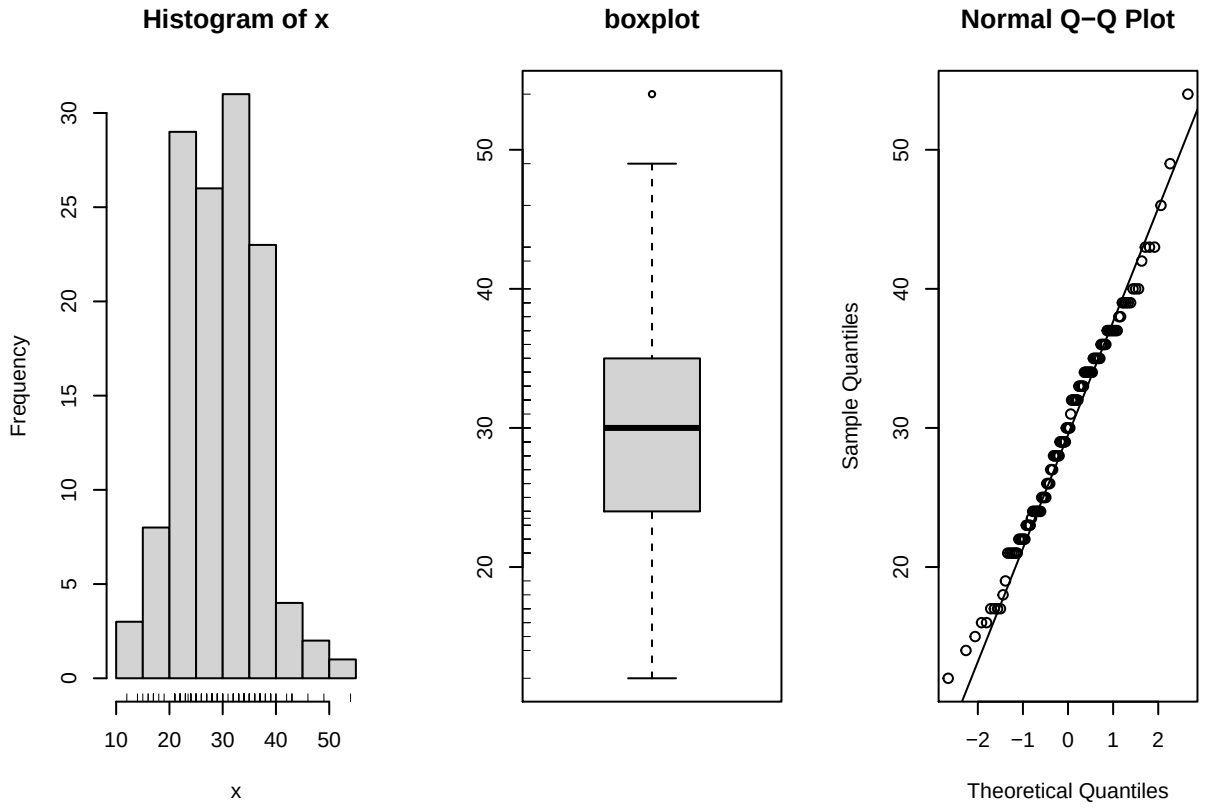
Hematocrit

```
all_dat %>%
  ggplot(., aes(x = hematocrit_percent)) +
  geom_histogram(color = "black", fill="steelblue", bins=50) +
  theme_classic() +
  xlab("Hematocrit (%)") +
  ylab("Count")
```

```
## Warning: Removed 12 rows containing non-finite values (stat_bin).
```




```
simple.eda(all_dat$hematocrit_percent)
```



```
shapiro.test(all_dat$hematocrit_percent)
```

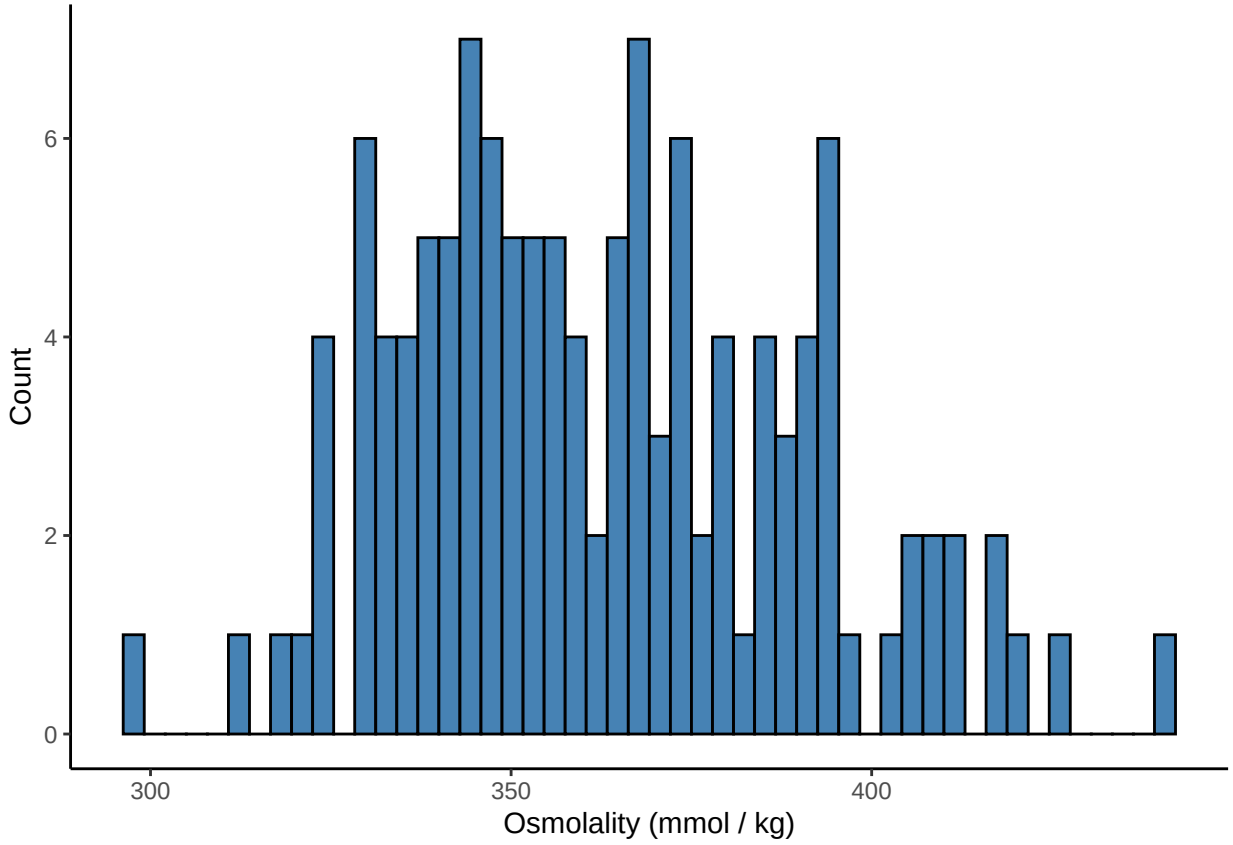
```
##
##  Shapiro-Wilk normality test
##
## data:  all_dat$hematocrit_percent
## W = 0.98872, p-value = 0.385
```

Visually, looks slightly skewed to the right, but statistically, the distribution of hematocrit is normal.

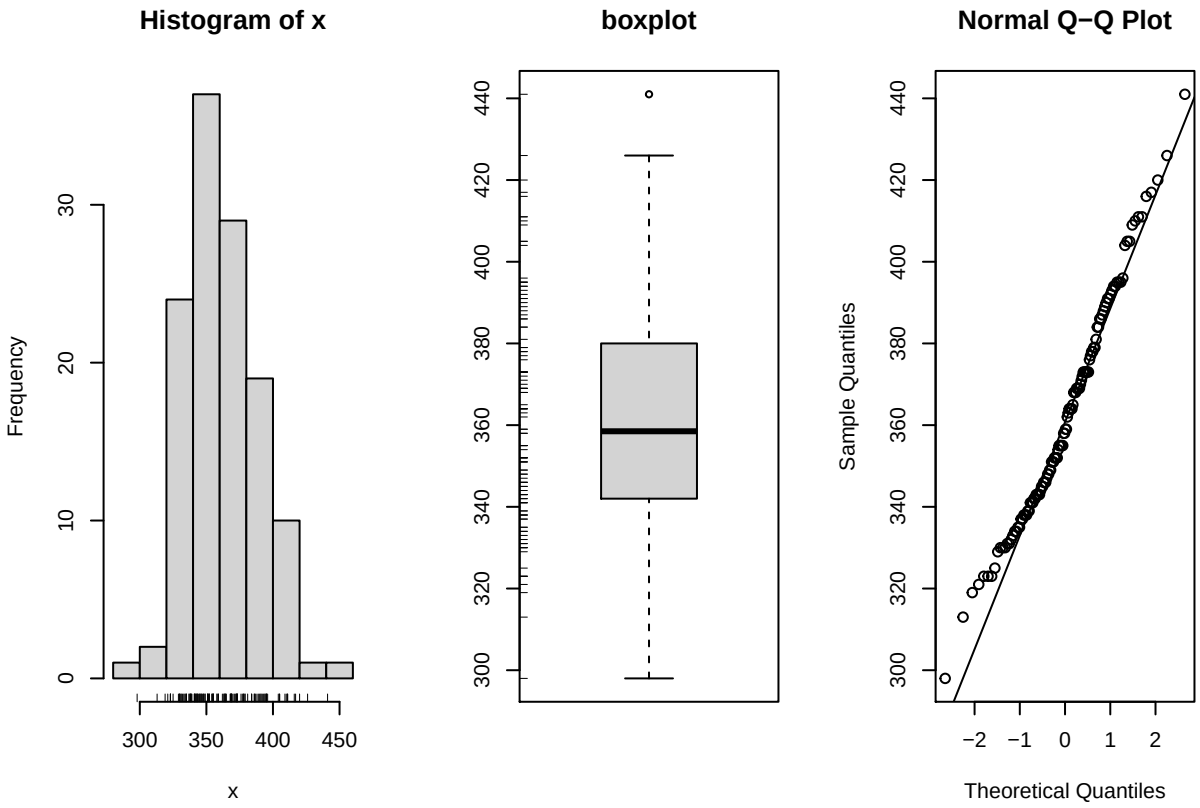
Osmolality

```
all_dat %>%
  ggplot(., aes(x = osmolality_mmol_kg)) +
  geom_histogram(color = "black", fill="steelblue", bins=50) +
  theme_classic() +
  xlab("Osmolality (mmol / kg)") +
  ylab("Count")
```

```
## Warning: Removed 15 rows containing non-finite values (stat_bin).
```




```
simple.eda(all_dat$osmolality_mmol_kg)
```




```
shapiro.test(all_dat$osmolality_mmol_kg)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  all_dat$osmolality_mmol_kg
## W = 0.9801, p-value = 0.06427
```

Visually, looks slightly skewed to the right, but statistically, the distribution of osmolality is normal.

Figures

Means to Overlay

```
all_dat_mean_SMI <- all_dat %>%
  group_by(humidity_tmt_percent, day) %>%
  summarise(SMI_mean = mean(SMI))
```

```
## `summarise()` regrouping output by 'humidity_tmt_percent' (override with `.groups` argument)
```

```
all_dat_mean_hct <- all_dat %>%
  dplyr::filter(complete.cases(hematocrit_percent)) %>%
  group_by(humidity_tmt_percent, day) %>%
  summarise(hct_mean = mean(hematocrit_percent))
```

```
## `summarise()` regrouping output by 'humidity_tmt_percent' (override with `.groups` argument)
```

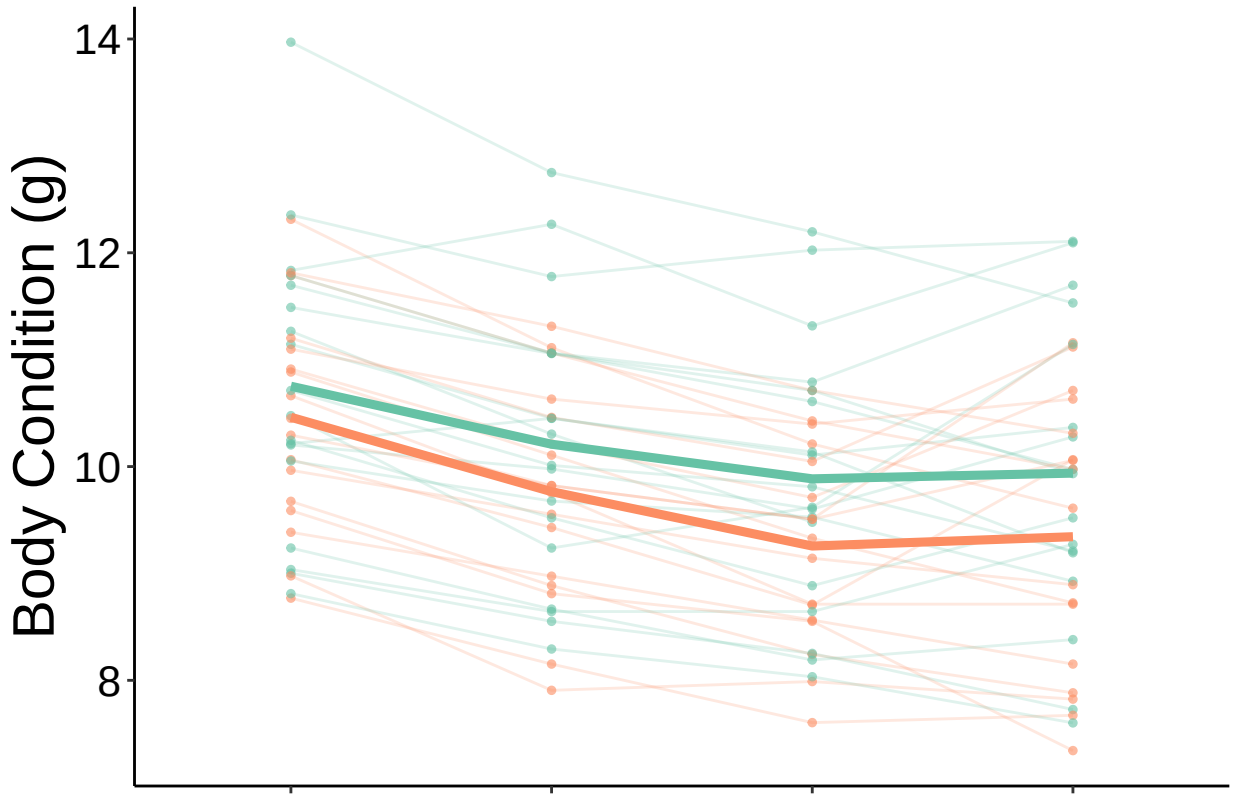
```
all_dat_mean_osml <- all_dat %>%
  dplyr::filter(complete.cases(osmolality_mmol_kg)) %>%
```

```
  group_by(humidity_tmt_percent, day) %>%
  summarise(osml_mean = mean(osmolality_mmol_kg))
```

```
## `summarise()` regrouping output by 'humidity_tmt_percent' (override with `.groups` argument)
```

SMI ~ Time

```
ggplot() +
  geom_point(data = all_dat,
             aes(x = day,
                 y = SMI,
                 color = humidity_tmt_percent
                 ),
             size = 1,
             alpha = 0.6) +
  geom_line(data = all_dat,
            aes(x = day,
                y = SMI,
                group = individual_ID,
                color = humidity_tmt_percent),
            alpha = 0.2) +
  geom_line(data = all_dat_mean_SMI,
            aes(x = day,
                y = SMI_mean,
                group = humidity_tmt_percent,
                color = humidity_tmt_percent),
            size = 1.6,
            alpha = 1) +
  theme_classic() +
  scale_color_brewer(palette = "Set2",
                     name = "") +
  xlab("") +
  ylab("Body Condition (g)") +
  theme(text = element_text(color = "black",
                            family = "sans",
                            size = 22),
        axis.text = element_text(color = "black",
                                 family = "sans",
                                 size = 16),
        axis.text.x = element_blank(),
        legend.text = element_text(color = "black",
                                 family = "sans",
                                 size = 24),
        legend.text.align = 0,
        legend.position = "none",
        plot.margin = unit(c(0.2,0,0,0.4), "cm")
) -> tmt_effects_SMI
tmt_effects_SMI
```




```
# export figure
#ggsave(filename = "tmt_effects_SMI.jpeg",
 #      plot = tmt_effects_SMI,
  #     path = "./final_figures",
   #    device = "jpeg",
    #   dpi = 1200,
     #  width = 5, height = 4)
```

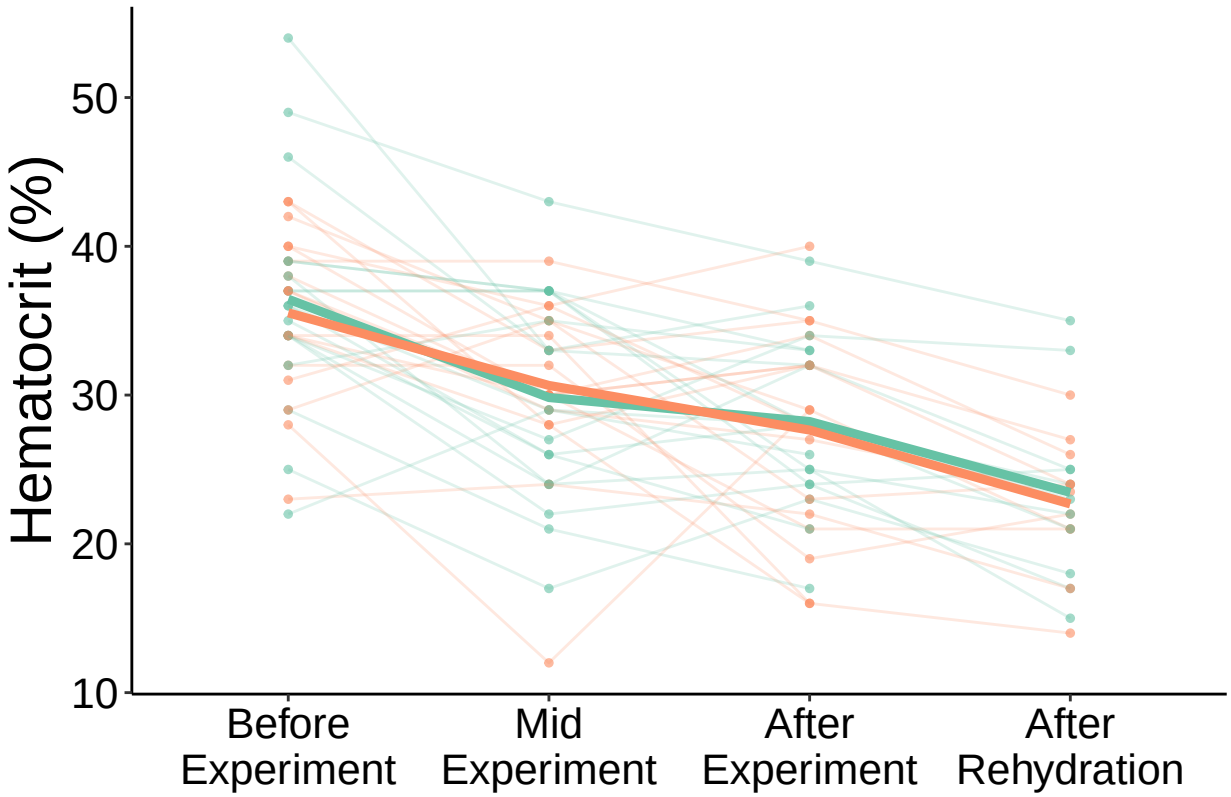
Hct ~ Time

```
ggplot() +
  geom_point(data = all_dat,
             aes(x = day,
                 y = hematocrit_percent,
                 color = humidity_tmt_percent
                 ),
             size = 1,
             alpha = 0.6) +
  geom_line(data = all_dat,
            aes(x = day,
                y = hematocrit_percent,
                group = individual_ID,
                color = humidity_tmt_percent),
            alpha = 0.2) +
  geom_line(data = all_dat_mean_hct,
            aes(x = day,
                y = hct_mean,
                group = humidity_tmt_percent,
```

```
                color = humidity_tmt_percent),
            size = 1.6,
            alpha = 1) +
  theme_classic() +
  scale_color_brewer(palette = "Set2",
                     name = "") +
  scale_x_discrete(labels = c("Before\nExperiment",
                              "Mid\nExperiment",
                              "After\nExperiment",
                              "After\nRehydration")) +
  xlab("") +
  ylab("Hematocrit (%)") +
  theme(text = element_text(color = "black",
                            family = "sans",
                            size = 22),
        axis.text = element_text(color = "black",
                                 family = "sans",
                                 size = 16),
        legend.text = element_text(color = "black",
                                 family = "sans",
                                 size = 24),
        legend.text.align = 0,
        legend.position = "none",
        plot.margin = unit(c(0.2,0,0,0.4), "cm")
        ) -> tmt_effects_hct
tmt_effects_hct
```

```
## Warning: Removed 12 rows containing missing values (geom_point).
```

```
## Warning: Removed 12 row(s) containing missing values (geom_path).
```




```
# export figure
#ggsave(filename = "tmt_effects_hct.jpeg",
 #       plot = tmt_effects_hct,
  #      path = "./final_figures",
   #     device = "jpeg",
    #    dpi = 1200,
     #   width = 5, height = 4)
```

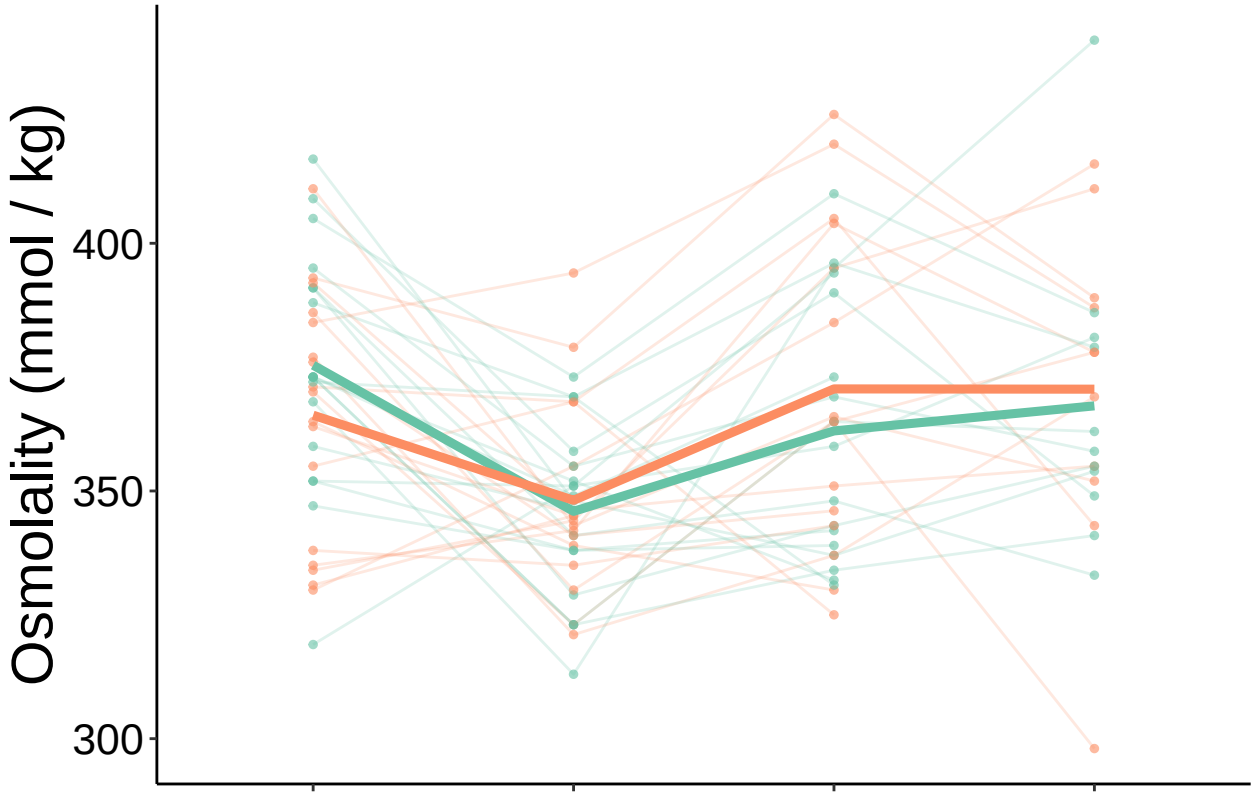
Osml ~ Time

```
ggplot() +
  geom_point(data = all_dat,
             aes(x = day,
                 y = osmolality_mmol_kg,
                 color = humidity_tmt_percent
                 ),
             size = 1,
             alpha = 0.6) +
  geom_line(data = all_dat,
            aes(x = day,
                y = osmolality_mmol_kg,
                group = individual_ID,
                color = humidity_tmt_percent),
            alpha = 0.2) +
  geom_line(data = all_dat_mean_osml,
            aes(x = day,
                y = osml_mean,
                group = humidity_tmt_percent,
```

```
                color = humidity_tmt_percent),
            size = 1.6,
            alpha = 1) +
  theme_classic() +
  scale_color_brewer(palette = "Set2",
                     name = "") +
  xlab("") +
  ylab("Osmolality (mmol / kg)") +
  theme(text = element_text(color = "black",
                            family = "sans",
                            size = 22),
        axis.text = element_text(color = "black",
                                 family = "sans",
                                 size = 16),
        axis.text.x = element_blank(),
        legend.text = element_text(color = "black",
                                 family = "sans",
                                 size = 24),
        legend.text.align = 0,
        legend.position = "none",
        plot.margin = unit(c(0.2,0,0,0.1), "cm")
        ) -> tmt_effects_osml
tmt_effects_osml
```

```
## Warning: Removed 15 rows containing missing values (geom_point).
```

```
## Warning: Removed 15 row(s) containing missing values (geom_path).
```



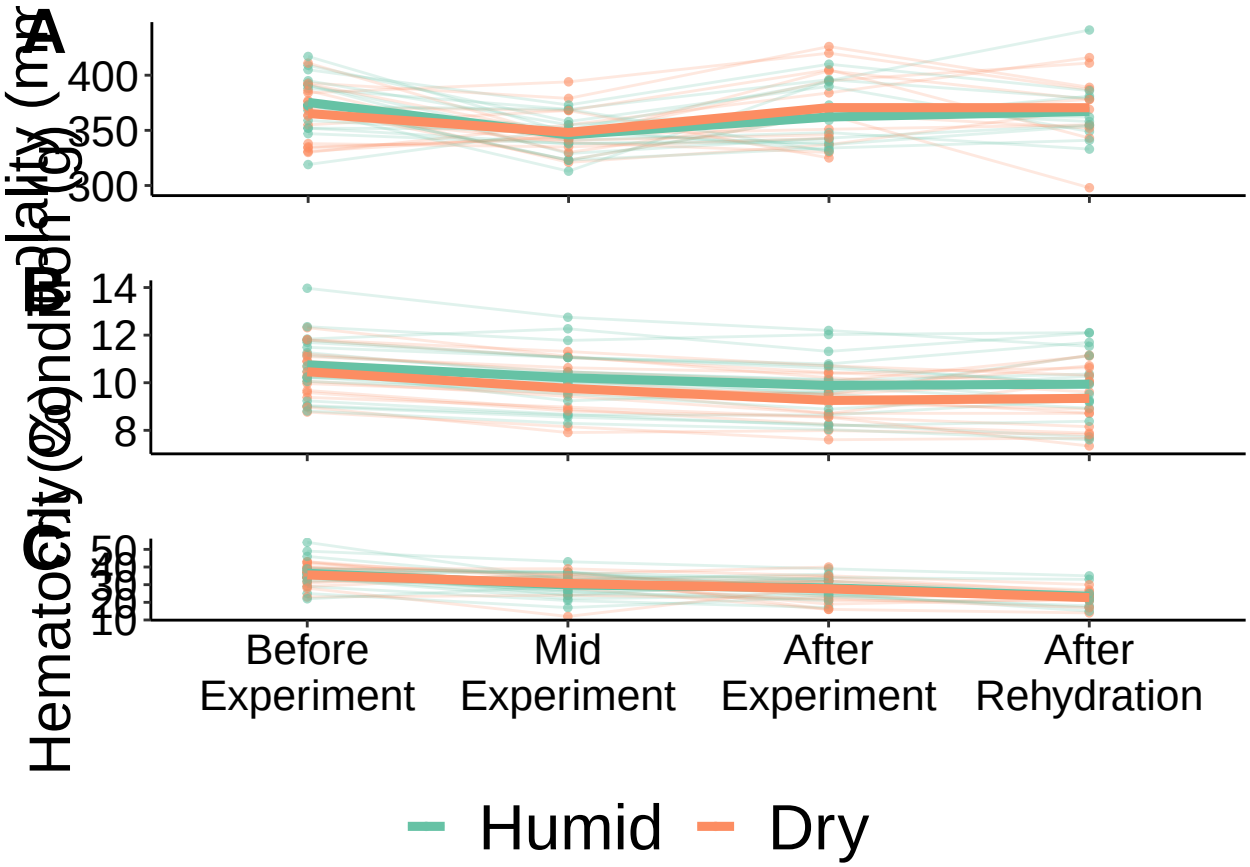
```
# export figure
#ggsave(filename = "tmt_effects_osml.jpeg",
 #      plot = tmt_effects_osml,
  #     path = "./final_figures",
   #    device = "jpeg",
    #   dpi = 1200,
     #  width = 5, height = 4)
```

Multi-Figure

```
ggarrange(tmt_effects_osml, tmt_effects_SMI, tmt_effects_hct,
          ncol = 1, nrow = 3,
          labels = c("A", "B", "C"),
          font.label = list(size = 24, face = "bold", color ="black"),
          vjust = c(1, 1, 1),
          common.legend = TRUE,
          legend = "bottom"
          ) -> tmt_multi_fig
```

```
## Warning: Removed 15 rows containing missing values (geom_point).
## Warning: Removed 15 row(s) containing missing values (geom_path).
## Warning: Removed 15 rows containing missing values (geom_point).
## Warning: Removed 15 row(s) containing missing values (geom_path).
## Warning: Removed 12 rows containing missing values (geom_point).
## Warning: Removed 12 row(s) containing missing values (geom_path).
```

```
tmt_multi_fig
```

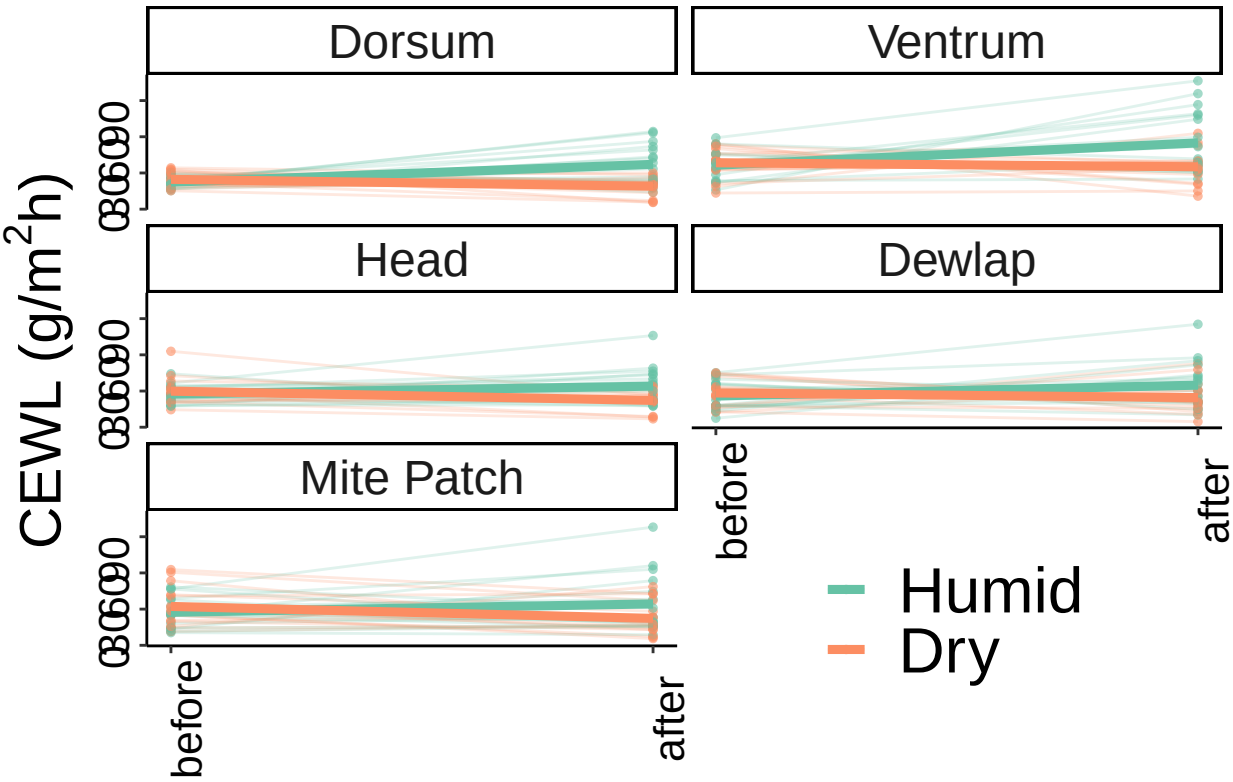



```
# export figure
ggsave(filename = "tmt_multi_fig.jpeg",
       plot = tmt_multi_fig,
       path = "./final_figures",
       device = "jpeg",
       dpi = 1200,
       width = 6, height = 12)
```

CEWL

```
CEWL %>%
  ggplot(data = .) +
  geom_point(aes(x = n_day,
                 y = TEWL_g_m2h,
                 color = humidity_tmt_percent
                 ),
             size = 1,
             alpha = 0.6) +
  geom_line(aes(x = n_day,
                y = TEWL_g_m2h,
                group = individual_ID,
                color = humidity_tmt_percent),
            alpha = 0.2) +
  stat_smooth(aes(x = n_day,
                  y = TEWL_g_m2h,
                  color = humidity_tmt_percent
```

```
                  ),
              formula = y ~ x,
              method = "lm",
              se = F,
              size = 1.6,
              alpha = 1) +
  theme_classic() +
  scale_color_brewer(palette = "Set2",
                     name = ""
                     #name = "Humidity\nTreatment"
                     ) +
  facet_wrap(~region, ncol = 2) +
  scale_x_continuous(breaks = c(0, 1),
                     labels = c("0" = "before", "1" = "after")
                     ) +
  xlab("") +
  ylab(bquote('CEWL (g/'*m^2*'h)')) +
  theme(text = element_text(color = "black",
                            family = "sans",
                            size = 22),
        axis.text = element_text(color = "black",
                                 family = "sans",
                                 size = 16,
                                 angle = 90),
        legend.text = element_text(color = "black",
                                 family = "sans",
                                 size = 24),
        legend.text.align = 0,
        legend.position = c(0.75,0.12),
        #legend.justification = c(1, 1)
) -> CEWL_tmt_fig
CEWL_tmt_fig
```




```
# export figure
ggsave(filename = "tmt_effects_CEWL.jpeg",
       plot = CEWL_tmt_fig,
       path = "./final_figures",
       device = "jpeg",
       dpi = 1200,
       width = 6, height = 9)
```

I saved the legend separately to make the figure layout better.

Models

SMI

Check whether means started out different:

```
SMI_diff_lm <- all_dat_no_rehab %>%
  dplyr::filter(day == "Before Experiment") %>%
  lm(data = ., SMI ~ humidity_tmt_percent)
summary(SMI_diff_lm)
```

```
##
## Call:
## lm(formula = SMI ~ humidity_tmt_percent, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9405 -0.7429 -0.0401  0.7385  3.2183
##
## Coefficients:
```

```
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               10.7517     0.2811   38.25   <2e-16 ***
## humidity_tmt_percentDry  -0.2904     0.4033   -0.72    0.476
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.192 on 33 degrees of freedom
## Multiple R-squared:  0.01547,    Adjusted R-squared:  -0.01436
## F-statistic: 0.5187 on 1 and 33 DF,  p-value: 0.4765
```

NOT significantly different, which is good.

Check whether means ended differently:

```
SMI_diff_lm_end <- all_dat_no_rehab %>%
  dplyr::filter(day == "After Experiment") %>%
  lm(data = ., SMI ~ humidity_tmt_percent)
summary(SMI_diff_lm_end)
```

```
##
## Call:
## lm(formula = SMI ~ humidity_tmt_percent, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.85144 -0.69873 -0.07453  0.80895  2.31088
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               9.8852     0.2585  38.246   <2e-16 ***
## humidity_tmt_percentDry  -0.6287     0.3709  -1.695   0.0994 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.097 on 33 degrees of freedom
## Multiple R-squared:  0.08012,    Adjusted R-squared:  0.05224
## F-statistic: 2.874 on 1 and 33 DF,  p-value: 0.09943
```

Build Model

```
SMI_mod <- lme4::lmer(data = all_dat_no_rehab,
               SMI ~ day*humidity_tmt_percent +
               (1|trial_number))
```

```
## boundary (singular) fit: see ?isSingular
```

```
summary(SMI_mod)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: SMI ~ day * humidity_tmt_percent + (1 | trial_number)
##    Data: all_dat_no_rehab
##
## REML criterion at convergence: 325.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -1.68654 -0.68344 -0.01058  0.71830  2.79715
##
## Random effects:
##  Groups       Name        Variance  Std.Dev.
##  trial_number (Intercept) 2.368e-20 1.539e-10
##  Residual                 1.324e+00 1.151e+00
## Number of obs: 105, groups:  trial_number, 4
##
## Fixed effects:
##                                             Estimate Std. Error t value
## (Intercept)                                  10.7517     0.2712  39.646
## dayMid Experiment                            -0.5428     0.3835  -1.415
## dayAfter Experiment                          -0.8665     0.3835  -2.259
## humidity_tmt_percentDry                      -0.2904     0.3891  -0.746
## dayMid Experiment:humidity_tmt_percentDry    -0.1534     0.5503  -0.279
## dayAfter Experiment:humidity_tmt_percentDry  -0.3383     0.5503  -0.615
##
## Correlation of Fixed Effects:
##             (Intr) dyMdEx dyAftE hmd__D dME:__
## dyMdExprmnt -0.707
## dyAftrExprm -0.707  0.500
## hmdty_tmt_D -0.697  0.493  0.493
## dyMExpr:__D  0.493 -0.697 -0.348 -0.707
## dyAExpr:__D  0.493 -0.348 -0.697 -0.707  0.500
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
drop1(SMI_mod)
```

```
## boundary (singular) fit: see ?isSingular
```

```
## Single term deletions
##
## Model:
## SMI ~ day * humidity_tmt_percent + (1 | trial_number)
##                           npar    AIC
## <none>                         337.25
## day:humidity_tmt_percent    2 333.66
```

```
# drop interaction term
SMI_mod2 <- lme4::lmer(data = all_dat_no_rehab,
               SMI ~ day + humidity_tmt_percent +
               (1|trial_number))
```

```
## boundary (singular) fit: see ?isSingular
```

```
summary(SMI_mod2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: SMI ~ day + humidity_tmt_percent + (1 | trial_number)
##    Data: all_dat_no_rehab
##
## REML criterion at convergence: 327.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.76998 -0.68801 -0.00593  0.73878  2.75012
```

```
##
## Random effects:
##  Groups       Name        Variance Std.Dev.
##  trial_number (Intercept) 0.000    0.000
##  Residual                 1.303    1.141
## Number of obs: 105, groups:  trial_number, 4
##
## Fixed effects:
##                          Estimate Std. Error t value
## (Intercept)               10.8313     0.2212  48.964
## dayMid Experiment         -0.6173     0.2728  -2.263
## dayAfter Experiment       -1.0308     0.2728  -3.778
## humidity_tmt_percentDry   -0.4543     0.2229  -2.039
##
## Correlation of Fixed Effects:
##             (Intr) dyMdEx dyAftE
## dyMdExprmnt -0.617
## dyAftrExprm -0.617  0.500
## hmdty_tmt_D -0.489  0.000  0.000
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
drop1(SMI_mod2)
```

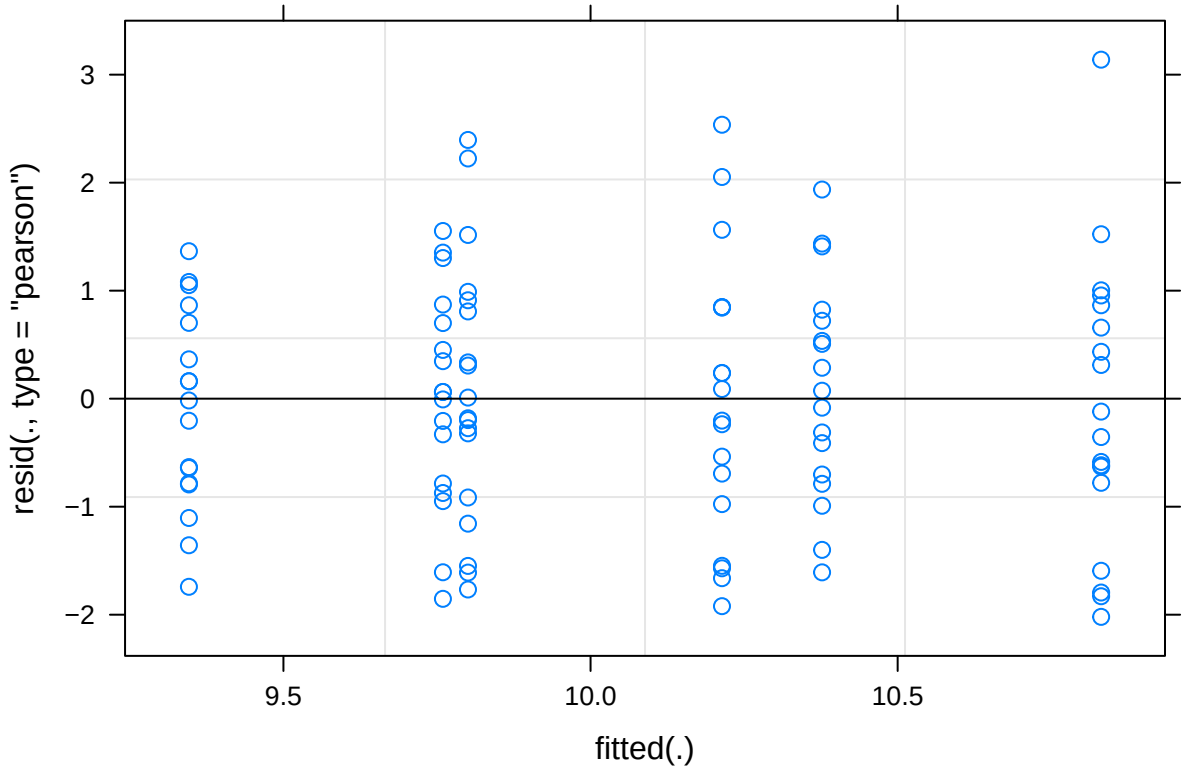
```
## boundary (singular) fit: see ?isSingular
## boundary (singular) fit: see ?isSingular
```

```
## Single term deletions
##
## Model:
## SMI ~ day + humidity_tmt_percent + (1 | trial_number)
##                       npar    AIC
## <none>                     333.66
## day                   2 343.71
## humidity_tmt_percent    1 335.89
```

Check Conditions

Is the function **linear**? Is there **equal** variance of the residuals? The residuals should be homoskedactic relative to y_hat (or x). Plotting residuals shows us whether the data meets linearity and equal variance assumptions:

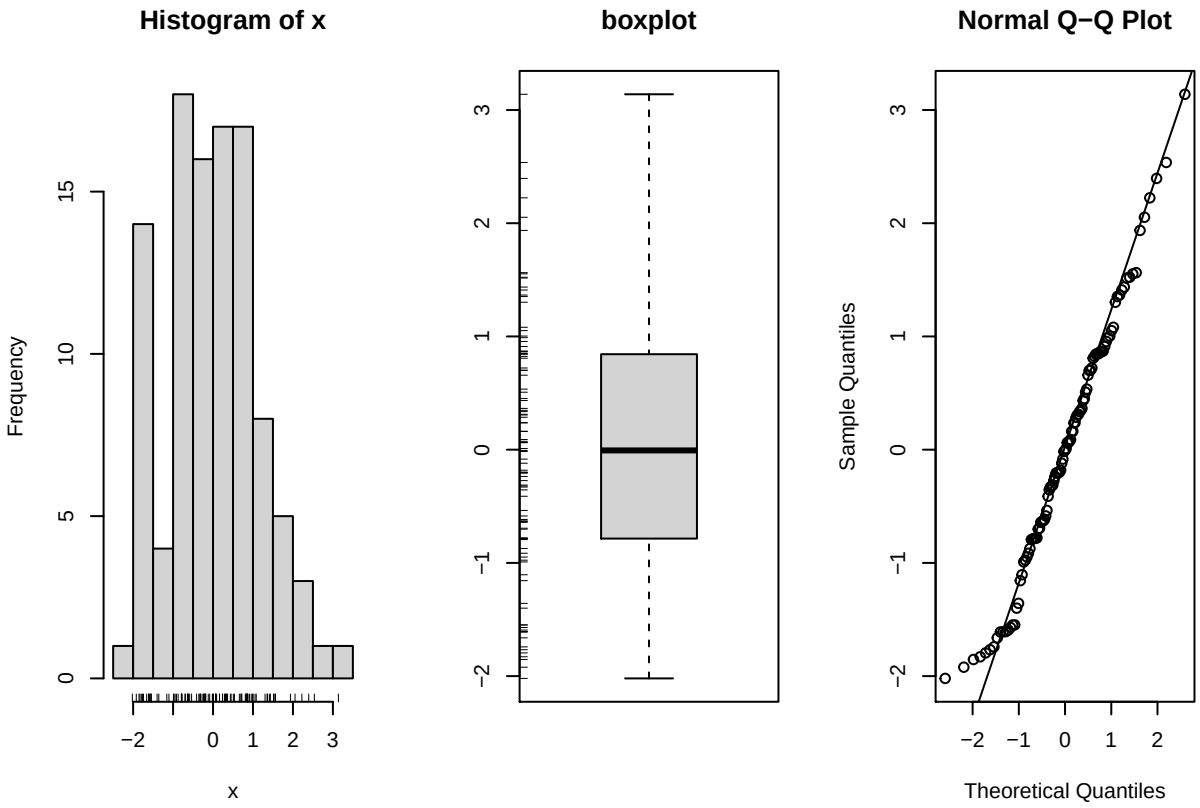
```
plot(SMI_mod2)
```



L & E both look good.

Is the distribution of residuals **normal**? use Shapiro-Wilk normality test: H0: data is NOT significantly different from normal distribution HA: data IS significantly different from normal distribution

```
simple.eda(residuals(SMI_mod2))
```




```
shapiro.test(residuals(SMI_mod2))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(SMI_mod2)
## W = 0.98165, p-value = 0.1551
```

Normality is fine.

Export Best

SMI is best predicted by day and treatment, but not including their interaction.

```
SMI_mod_final <- lmerTest::lmer(data = all_dat_no_rehab,
                                SMI ~ day + humidity_tmt_percent +
                                  (1|trial_number))
```

```
## boundary (singular) fit: see ?isSingular
```

```
write.csv(broom.mixed::tidy(SMI_mod_final),
          "./best models/exp_effects_SMI.csv")
```

Hematocrit

Build Model

```
hct_mod <- all_dat_no_rehab %>%
  dplyr::filter(complete.cases(hematocrit_percent)) %>%
  lme4::lmer(data = .,
```

```
              hematocrit_percent ~ day * humidity_tmt_percent +
              (1|trial_number))
summary(hct_mod)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: hematocrit_percent ~ day * humidity_tmt_percent + (1 | trial_number)
##    Data: .
##
## REML criterion at convergence: 665.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1677 -0.5501  0.0365  0.5956  2.6381
##
## Random effects:
##  Groups       Name        Variance Std.Dev.
##  trial_number (Intercept)  7.185   2.681
##  Residual                 38.959   6.242
## Number of obs: 105, groups:  trial_number, 4
##
## Fixed effects:
##                                             Estimate Std. Error t value
## (Intercept)                                  36.0220     1.9995  18.016
## dayMid Experiment                            -6.6111     2.0806  -3.178
## dayAfter Experiment                          -8.2222     2.0806  -3.952
## humidity_tmt_percentDry                      -0.8797     2.1115  -0.417
## dayMid Experiment:humidity_tmt_percentDry     1.7288     2.9853   0.579
## dayAfter Experiment:humidity_tmt_percentDry   0.3399     2.9853   0.114
##
## Correlation of Fixed Effects:
##             (Intr) dyMdEx dyAftE hmd__D dME:__
## dyMdExprmnt -0.520
## dyAftrExprm -0.520  0.500
## hmdty_tmt_D -0.514  0.493  0.493
## dyMExpr:__D  0.363 -0.697 -0.348 -0.707
## dyAExpr:__D  0.363 -0.348 -0.697 -0.707  0.500
```

```
drop1(hct_mod)
```

```
## Single term deletions
##
## Model:
## hematocrit_percent ~ day * humidity_tmt_percent + (1 | trial_number)
##                           npar    AIC
## <none>                         699.17
## day:humidity_tmt_percent    2 695.56
```

```
# drop day*humidity interaction
hct_mod2 <- all_dat_no_rehab %>%
  dplyr::filter(complete.cases(hematocrit_percent)) %>%
  lme4::lmer(data = .,
             hematocrit_percent ~ day + humidity_tmt_percent +
             (1|trial_number))
summary(hct_mod2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: hematocrit_percent ~ day + humidity_tmt_percent + (1 | trial_number)
##    Data: .
##
## REML criterion at convergence: 674
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.10869 -0.60033  0.02219  0.62265  2.71370
##
## Random effects:
##  Groups       Name        Variance Std.Dev.
##  trial_number (Intercept)  7.224   2.688
##  Residual                 38.312   6.190
## Number of obs: 105, groups:  trial_number, 4
##
## Fixed effects:
##                           Estimate Std. Error t value
## (Intercept)                 35.685      1.812  19.698
## dayMid Experiment           -5.771      1.480  -3.901
## dayAfter Experiment         -8.057      1.480  -5.445
## humidity_tmt_percentDry     -0.190      1.210  -0.157
##
## Correlation of Fixed Effects:
##             (Intr) dyMdEx dyAftE
## dyMdExprmnt -0.408
## dyAftrExprm -0.408  0.500
## hmdty_tmt_D -0.326  0.000  0.000
```

```
drop1(hct_mod2)
```

```
## Single term deletions
##
## Model:
## hematocrit_percent ~ day + humidity_tmt_percent + (1 | trial_number)
##                       npar    AIC
## <none>                     695.56
## day                      2 719.67
## humidity_tmt_percent     1 693.59
```

```
# drop humidity tmt
hct_mod3 <- all_dat_no_rehab %>%
  dplyr::filter(complete.cases(hematocrit_percent)) %>%
  lme4::lmer(data = .,
             hematocrit_percent ~ day +
             (1|trial_number))
summary(hct_mod3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: hematocrit_percent ~ day + (1 | trial_number)
##    Data: .
##
## REML criterion at convergence: 676.2
##
## Scaled residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -3.14009 -0.58850  0.00871  0.64072  2.74219
##
## Random effects:
##  Groups       Name        Variance Std.Dev.
##  trial_number (Intercept)  7.253   2.693
##  Residual                 37.932   6.159
## Number of obs: 105, groups:  trial_number, 4
##
## Fixed effects:
##                      Estimate Std. Error t value
## (Intercept)            35.592      1.712  20.794
## dayMid Experiment      -5.771      1.472  -3.920
## dayAfter Experiment    -8.057      1.472  -5.473
##
## Correlation of Fixed Effects:
##             (Intr) dyMdEx
## dyMdExprmnt -0.430
## dyAftrExprm -0.430  0.500
```

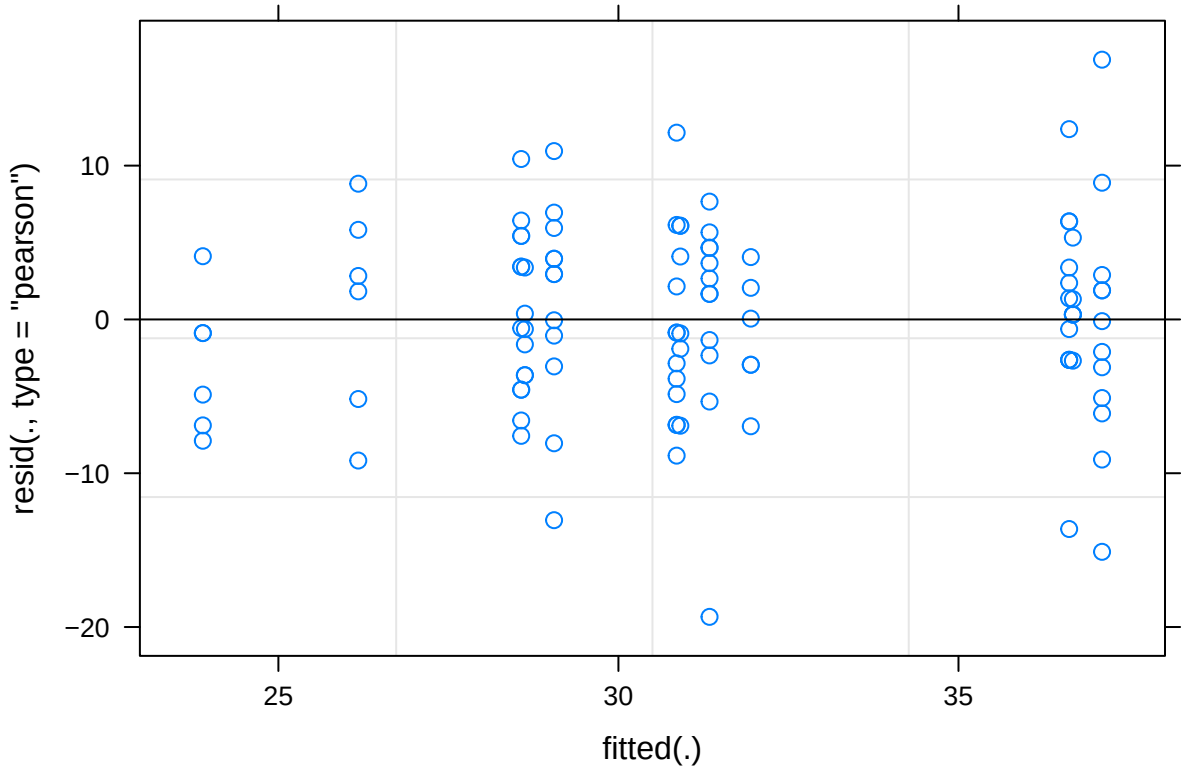
```
drop1(hct_mod3)
```

```
## Single term deletions
##
## Model:
## hematocrit_percent ~ day + (1 | trial_number)
##        npar    AIC
## <none>      693.59
## day       2 717.69
```

Check Conditions

Is the function **linear**? Is there **equal** variance of the residuals? The residuals should be homoskedactic relative to y_hat (or x). Plotting residuals shows us whether the data meets linearity and equal variance assumptions:

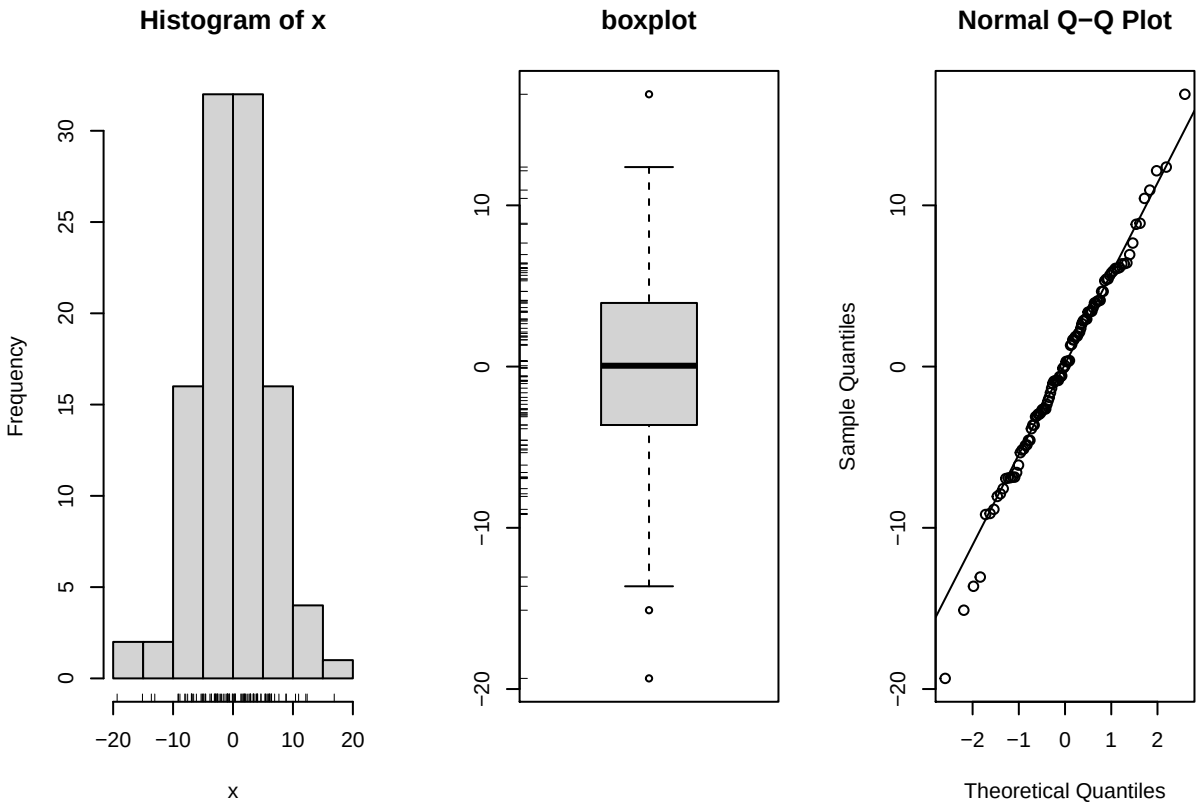
```
plot(hct_mod3)
```



L & E both seem fine. Slight fanning to the right, but otherwise no big pattern, so L is definitely okay. E could use help.

Is the distribution of residuals **normal**? use Shapiro-Wilk normality test: H0: data is NOT significantly different from normal distribution HA: data IS significantly different from normal distribution

```
simple.eda(residuals(hct_mod3))
```




```
shapiro.test(residuals(hct_mod3))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(hct_mod3)
## W = 0.98869, p-value = 0.5248
```

residuals are normally distributed!

Export Best

```
hct_mod_final <- all_dat_no_rehab %>%
  dplyr::filter(complete.cases(hematocrit_percent)) %>%
  lmerTest::lmer(data = .,
                 hematocrit_percent ~ day +
                 (1|trial_number))
write.csv(broom::tidy(hct_mod_final),
          "./best models/exp_effects_hct.csv")
```

Osmolality

Build Model

```
osml_mod <- all_dat_no_rehab %>%
  dplyr::filter(complete.cases(osmolality_mmol_kg)) %>%
  lme4::lmer(data = .,
                 osmolality_mmol_kg ~ day * humidity_tmt_percent +
```

```
              (1|trial_number))
summary(osml_mod)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: osmolality_mmol_kg ~ day * humidity_tmt_percent + (1 | trial_number)
##    Data: .
##
## REML criterion at convergence: 872.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.66849 -0.79364  0.00361  0.51469  2.41228
##
## Random effects:
##  Groups       Name        Variance Std.Dev.
##  trial_number (Intercept) 313.5    17.71
##  Residual                 393.8    19.84
## Number of obs: 102, groups:  trial_number, 4
##
## Fixed effects:
##                                             Estimate Std. Error t value
## (Intercept)                                  378.509     10.035  37.719
## dayMid Experiment                            -29.500      6.615  -4.460
## dayAfter Experiment                          -13.885      6.713  -2.068
## humidity_tmt_percentDry                      -10.215      6.713  -1.522
## dayMid Experiment:humidity_tmt_percentDry     12.324      9.491   1.298
## dayAfter Experiment:humidity_tmt_percentDry   17.783      9.721   1.829
##
## Correlation of Fixed Effects:
##             (Intr) dyMdEx dyAftE hmd__D dME:__
## dyMdExprmnt -0.330
## dyAftrExprm -0.326  0.493
## hmdty_tmt_D -0.326  0.493  0.485
## dyMExpr:__D  0.230 -0.697 -0.343 -0.707
## dyAExpr:__D  0.224 -0.340 -0.690 -0.690  0.488
```

```
drop1(osml_mod)
```

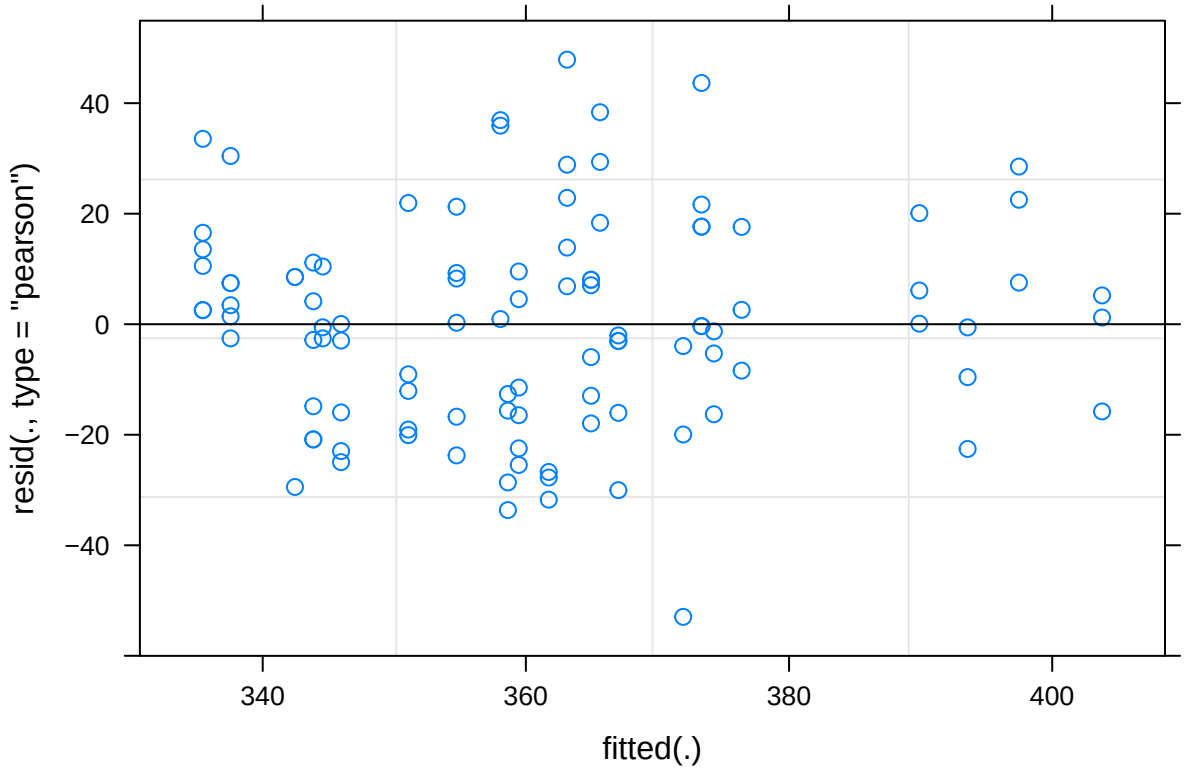
```
## Single term deletions
##
## Model:
## osmolality_mmol_kg ~ day * humidity_tmt_percent + (1 | trial_number)
##                           npar    AIC
## <none>                         920.78
## day:humidity_tmt_percent    2 920.47
```

This is the best model.

Check Conditions

Is the function **linear**? Is there **equal** variance of the residuals? The residuals should be homoskedactic relative to y_hat (or x). Plotting residuals shows us whether the data meets linearity and equal variance assumptions:

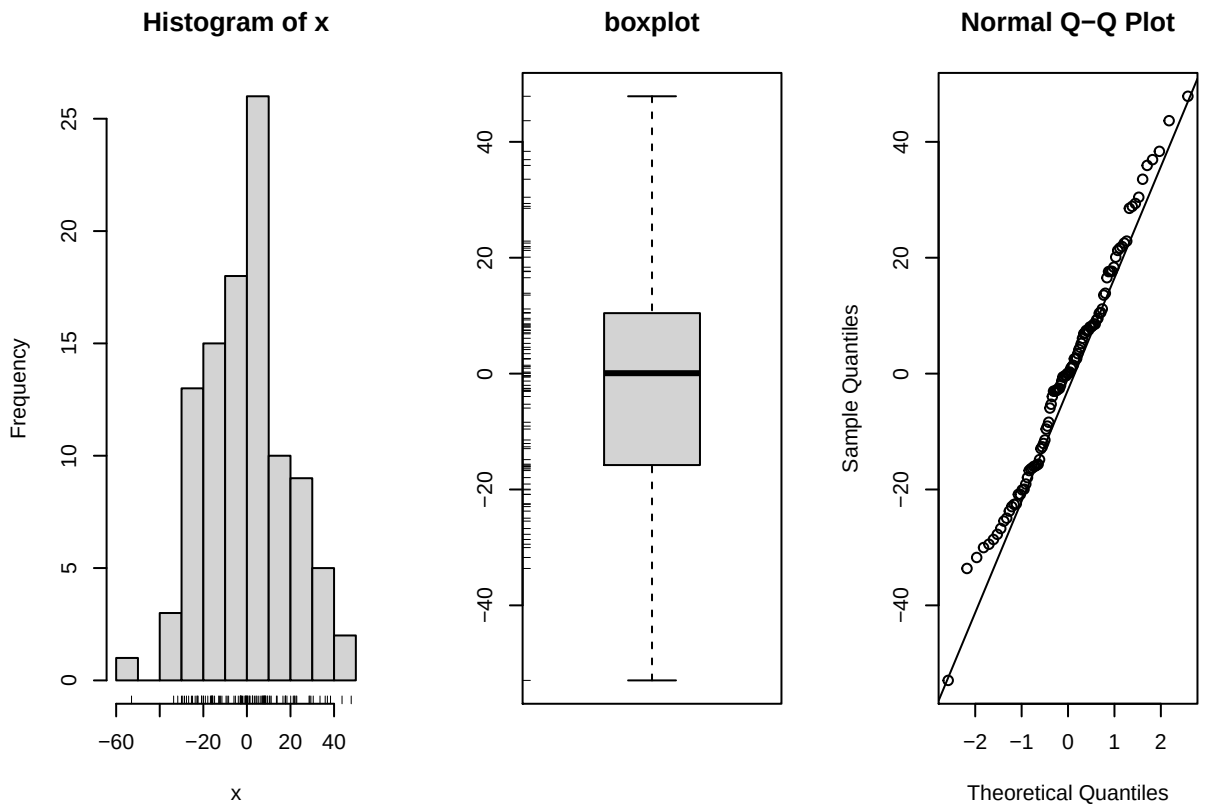
```
plot(osml_mod)
```



L & E both seem fine.

Is the distribution of residuals **normal**? use Shapiro-Wilk normality test: H0: data is NOT significantly different from normal distribution HA: data IS significantly different from normal distribution

```
simple.eda(residuals(osml_mod))
```




```
shapiro.test(residuals(osml_mod))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(osml_mod)
## W = 0.98957, p-value = 0.616
```

Normality is fine.

Export Best

The model seems good as-is.

```
osml_mod_final <- all_dat_no_rehab %>%
  dplyr::filter(complete.cases(osmolality_mmol_kg)) %>%
  lmerTest::lmer(data = .,
                 osmolality_mmol_kg ~ day * humidity_tmt_percent +
                 (1|trial_number))
write.csv(broom::tidy(osml_mod_final),
          "./best models/exp_effects_osml.csv")
```

CEWL

Build Model

```
CEWL_mod <- CEWL %>%
  lme4::lmer(data = .,
                 TEWL_g_m2h ~ day * humidity_tmt_percent * region +
```

```
              cloacal_temp_C +
              (1|trial_number/individual_ID))
summary(CEWL_mod)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: TEWL_g_m2h ~ day * humidity_tmt_percent * region + cloacal_temp_C +
##     (1 | trial_number/individual_ID)
##    Data: .
##
## REML criterion at convergence: 2441.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.4361 -0.5707 -0.0822  0.4555  4.1852
##
## Random effects:
##  Groups                      Name        Variance Std.Dev.
##  individual_ID:trial_number (Intercept)  38.212   6.182
##  trial_number               (Intercept)   8.141   2.853
##  Residual                                123.274  11.103
## Number of obs: 326, groups:  individual_ID:trial_number, 33; trial_number, 4
##
## Fixed effects:
##                                                  Estimate Std. Error t value
## (Intercept)                                     -69.91895   11.98454  -5.834
## dayAfter                                         20.55665    3.96057   5.190
## humidity_tmt_percentDry                           0.09390    4.44521   0.021
## regionVentrum                                    14.06471    3.80826   3.693
## regionHead                                        5.20176    3.80826   1.366
## regionDewlap                                      2.76945    3.87231   0.715
## regionMite Patch                                  5.02118    3.80826   1.318
## cloacal_temp_C                                    3.90353    0.48902   7.982
## dayAfter:humidity_tmt_percentDry                -22.17103    5.52280  -4.014
## dayAfter:regionVentrum                            4.82016    5.43126   0.887
## dayAfter:regionHead                              -7.17102    5.43126  -1.320
## dayAfter:regionDewlap                            -2.93470    5.46212  -0.537
## dayAfter:regionMite Patch                        -6.04716    5.47797  -1.104
## humidity_tmt_percentDry:regionVentrum            -0.34596    5.46920  -0.063
## humidity_tmt_percentDry:regionHead               -0.03395    5.46920  -0.006
## humidity_tmt_percentDry:regionDewlap              0.90930    5.51399   0.165
## humidity_tmt_percentDry:regionMite Patch          2.90477    5.52066   0.526
## dayAfter:humidity_tmt_percentDry:regionVentrum   -2.37516    7.76641  -0.306
## dayAfter:humidity_tmt_percentDry:regionHead        5.28070    7.76641   0.680
## dayAfter:humidity_tmt_percentDry:regionDewlap      5.66684    7.82338   0.724
## dayAfter:humidity_tmt_percentDry:regionMite Patch  1.47809    7.83657   0.189
##
## Correlation matrix not shown by default, as p = 21 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it
```

```
drop1(CEWL_mod)
```

```
## Single term deletions
##
```

```
## Model:
## TEWL_g_m2h ~ day * humidity_tmt_percent * region + cloacal_temp_C +
##     (1 | trial_number/individual_ID)
##                                 npar    AIC
## <none>                              2570.1
## cloacal_temp_C                     1 2628.2
## day:humidity_tmt_percent:region    4 2563.8
```

Drop triple interaction.

```
CEWL_mod2 <- CEWL %>%
  lme4::lmer(data = .,
             TEWL_g_m2h ~
             humidity_tmt_percent * (day + region) +
             cloacal_temp_C +
             (1|trial_number/individual_ID))
summary(CEWL_mod2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: TEWL_g_m2h ~ humidity_tmt_percent * (day + region) + cloacal_temp_C +
##     (1 | trial_number/individual_ID)
##    Data: .
##
## REML criterion at convergence: 2490.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.4790 -0.6180 -0.0854  0.4480  4.0055
##
## Random effects:
##  Groups                      Name        Variance Std.Dev.
##  individual_ID:trial_number (Intercept)  38.017   6.166
##  trial_number               (Intercept)   8.677   2.946
##  Residual                                123.628  11.119
## Number of obs: 326, groups:  individual_ID:trial_number, 33; trial_number, 4
##
## Fixed effects:
##                                          Estimate Std. Error t value
## (Intercept)                              -68.6125    11.9022  -5.765
## humidity_tmt_percentDry                   -0.8482     3.7242  -0.228
## dayAfter                                  18.2857     1.9294   9.478
## regionVentrum                             16.5143     2.7189   6.074
## regionHead                                 1.6557     2.7189   0.609
## regionDewlap                               1.3212     2.7199   0.486
## regionMite Patch                           2.0988     2.7414   0.766
## cloacal_temp_C                             3.8933     0.4894   7.956
## humidity_tmt_percentDry:dayAfter         -20.1573     2.4933  -8.084
## humidity_tmt_percentDry:regionVentrum     -1.5730     3.8883  -0.405
## humidity_tmt_percentDry:regionHead         2.5669     3.8883   0.660
## humidity_tmt_percentDry:regionDewlap       3.6718     3.9063   0.940
## humidity_tmt_percentDry:regionMite Patch   3.4656     3.9211   0.884
##
## Correlation matrix not shown by default, as p = 13 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it
```

```
drop1(CEWL_mod2)
```

```
## Single term deletions
##
## Model:
## TEWL_g_m2h ~ humidity_tmt_percent * (day + region) + cloacal_temp_C +
##     (1 | trial_number/individual_ID)
##                              npar    AIC
## <none>                            2563.5
## cloacal_temp_C                  1 2619.8
## humidity_tmt_percent:day        1 2622.6
## humidity_tmt_percent:region     4 2558.4
```

We can drop the humidity:region interaction.

```
CEWL_mod3 <- CEWL %>%
  dplyr::filter(complete.cases(.)) %>%
  lme4::lmer(data = .,
             TEWL_g_m2h ~
             day*humidity_tmt_percent + region + cloacal_temp_C +
             (1|trial_number/individual_ID))
summary(CEWL_mod3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: TEWL_g_m2h ~ day * humidity_tmt_percent + region + cloacal_temp_C +
##     (1 | trial_number/individual_ID)
##    Data: .
##
## REML criterion at convergence: 2510.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3478 -0.6058 -0.1117  0.4446  3.9319
##
## Random effects:
##  Groups                      Name        Variance Std.Dev.
##  individual_ID:trial_number (Intercept)  38.023   6.166
##  trial_number               (Intercept)   8.872   2.979
##  Residual                                123.111  11.096
## Number of obs: 326, groups:  individual_ID:trial_number, 33; trial_number, 4
##
## Fixed effects:
##                                   Estimate Std. Error t value
## (Intercept)                       -69.3758    11.8197  -5.869
## dayAfter                           18.2632     1.9251   9.487
## humidity_tmt_percentDry             0.7605     2.7927   0.272
## regionVentrum                      15.7651     1.9396   8.128
## regionHead                          2.9138     1.9396   1.502
## regionDewlap                        3.0977     1.9483   1.590
## regionMite Patch                    3.7912     1.9565   1.938
## cloacal_temp_C                      3.8920     0.4886   7.966
## dayAfter:humidity_tmt_percentDry -20.1375     2.4881  -8.094
##
## Correlation of Fixed Effects:
##             (Intr) dyAftr hmd__D rgnVnt regnHd rgnDwl rgnMtP clc__C
```

```
## dayAfter    -0.507
## hmdty_tmt_D  0.005  0.217
## regionVntrm -0.086 -0.004  0.000
## regionHead  -0.086 -0.004  0.000  0.504
## regionDewlp -0.098 -0.010 -0.006  0.502  0.502
## reginMtPtch -0.108  0.013  0.002  0.500  0.500  0.498
## clocl_tmp_C -0.972  0.456 -0.122  0.004  0.004  0.017  0.026
## dyAftr:h__D  0.191 -0.680 -0.418  0.004  0.004  0.017 -0.010 -0.146
```

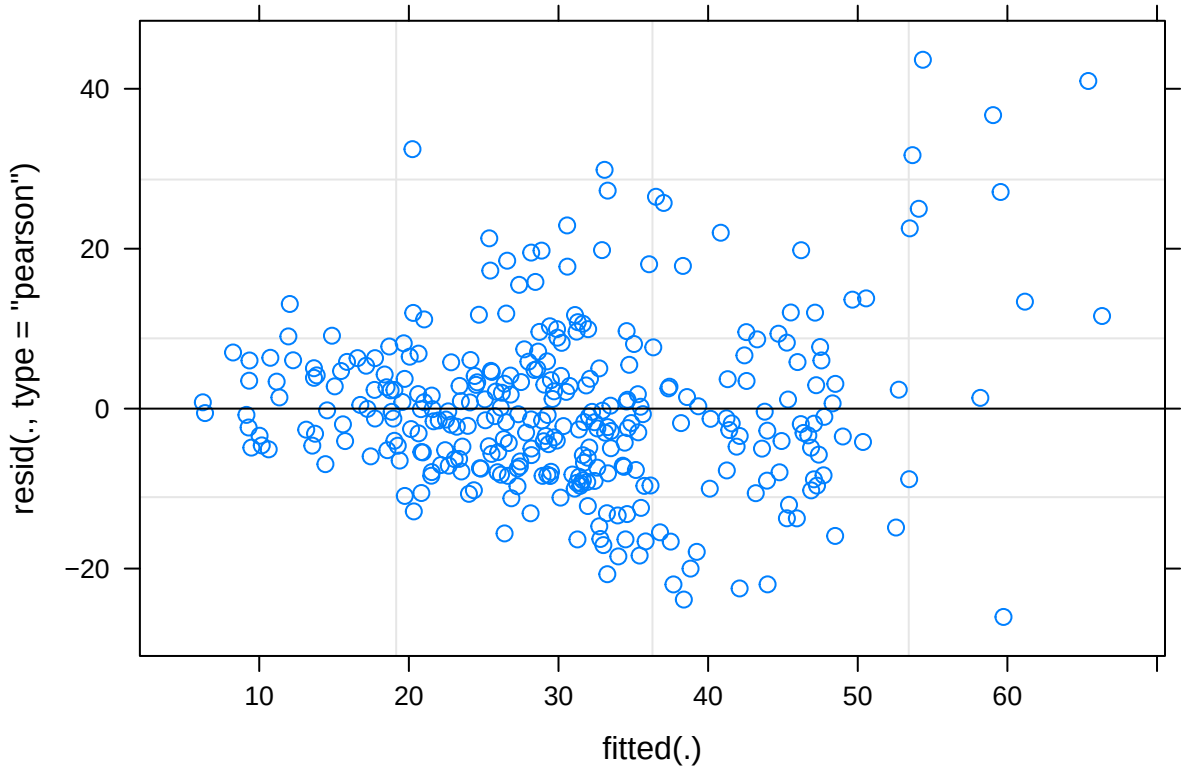
```
drop1(CEWL_mod3)
```

```
## Single term deletions
##
## Model:
## TEWL_g_m2h ~ day * humidity_tmt_percent + region + cloacal_temp_C +
##     (1 | trial_number/individual_ID)
##                            npar    AIC
## <none>                          2558.4
## region                        4 2622.8
## cloacal_temp_C                1 2614.1
## day:humidity_tmt_percent      1 2616.8
```

Check Conditions

Is the function **linear**? Is there **equal** variance of the residuals? The residuals should be homoskedactic relative to y_hat (or x). Plotting residuals shows us whether the data meets linearity and equal variance assumptions:

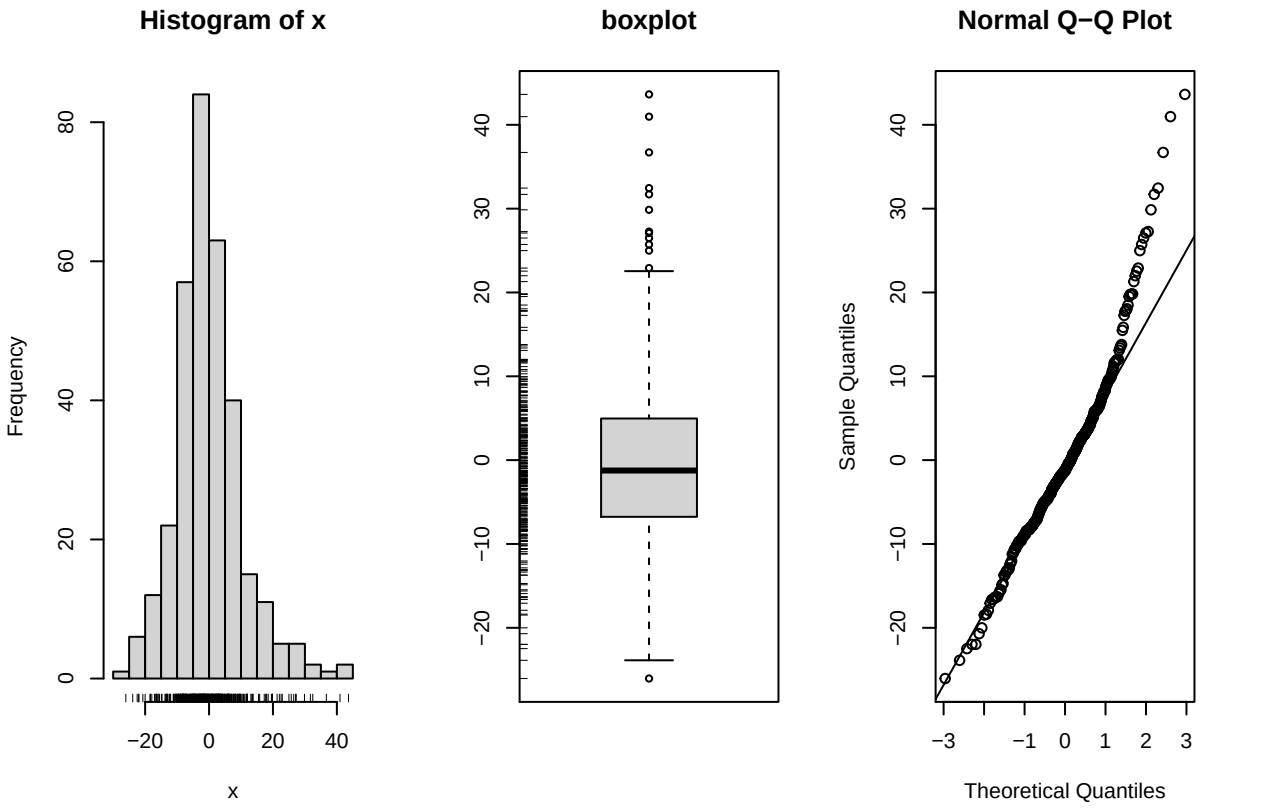
```
plot(CEWL_mod3)
```



clear fanning shape... L & E are not satisfied.

Is the distribution of residuals **normal**? use Shapiro-Wilk normality test: H0: data is NOT significantly different from normal distribution HA: data IS significantly different from normal distribution

```
simple.eda(residuals(CEWL_mod3))
```




```
shapiro.test(residuals(CEWL_mod3))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(CEWL_mod3)
## W = 0.95465, p-value = 1.701e-08
```

distribution not normal

Transform CEWL

```
CEWL_mod3_t <- CEWL %>%
  dplyr::filter(complete.cases(.)) %>%
  lme4::lmer(data = .,
             log(TEWL_g_m2h) ~
             day*humidity_tmt_percent + region + cloacal_temp_C +
             (1|trial_number/individual_ID))
summary(CEWL_mod3_t)
```

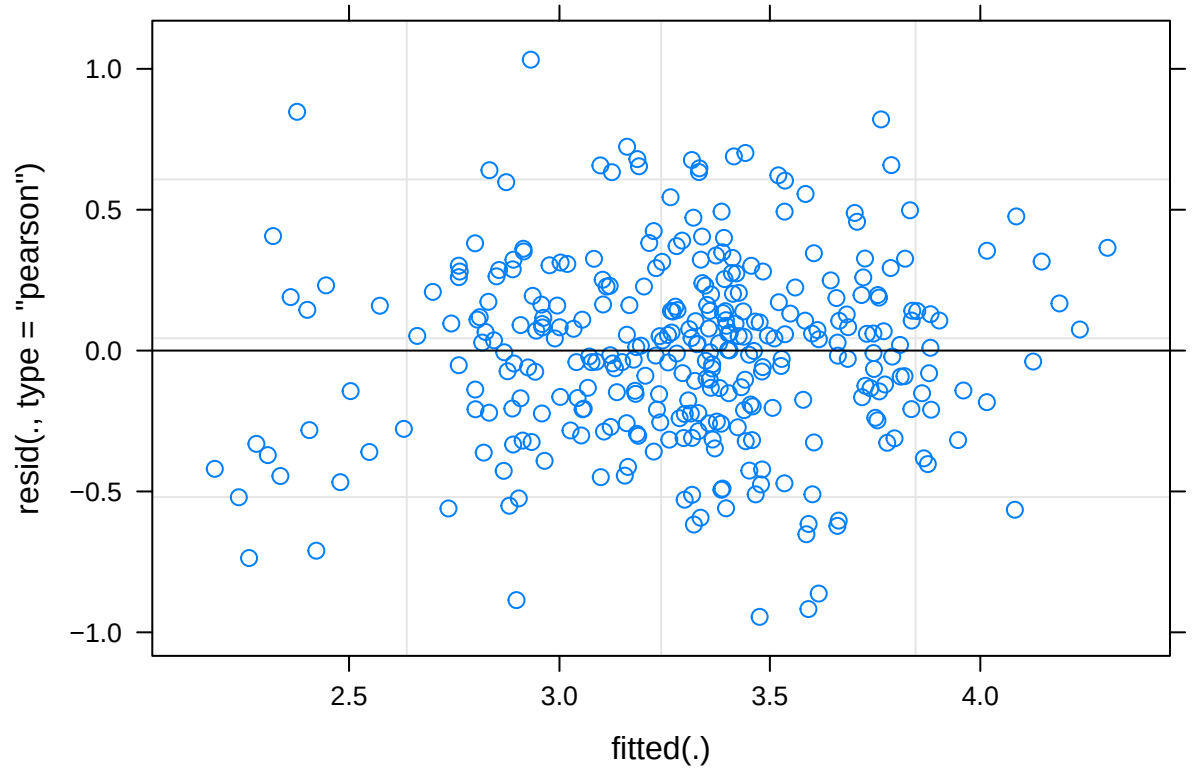
```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## log(TEWL_g_m2h) ~ day * humidity_tmt_percent + region + cloacal_temp_C +
##     (1 | trial_number/individual_ID)
##    Data: .
```

```
##
## REML criterion at convergence: 302.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.76069 -0.61262  0.05481  0.57357  3.01606
##
## Random effects:
##  Groups                      Name        Variance Std.Dev.
##  individual_ID:trial_number (Intercept) 0.02787  0.1669
##  trial_number               (Intercept) 0.03971  0.1993
##  Residual                               0.11720  0.3423
## Number of obs: 326, groups:  individual_ID:trial_number, 33; trial_number, 4
##
## Fixed effects:
##                                  Estimate Std. Error t value
## (Intercept)                      -0.16197    0.37173  -0.436
## dayAfter                          0.51526    0.05931   8.688
## humidity_tmt_percentDry           0.03209    0.08029   0.400
## regionVentrum                     0.48102    0.05984   8.038
## regionHead                        0.12591    0.05984   2.104
## regionDewlap                      0.08143    0.06011   1.355
## regionMite Patch                  0.05704    0.06036   0.945
## cloacal_temp_C                    0.13641    0.01496   9.117
## dayAfter:humidity_tmt_percentDry -0.65376    0.07676  -8.517
##
## Correlation of Fixed Effects:
##             (Intr) dyAftr hmd__D rgnVnt regnHd rgnDwl rgnMtP clc__C
## dayAfter    -0.491
## hmdty_tmt_D  0.028  0.231
## regionVntrm -0.085 -0.004  0.000
## regionHead  -0.085 -0.004  0.000  0.504
## regionDewlp -0.096 -0.010 -0.006  0.502  0.502
## reginMtPtch -0.106  0.013  0.002  0.500  0.500  0.498
## clocl_tmp_C -0.946  0.453 -0.139  0.004  0.004  0.017  0.027
## dyAftr:h__D  0.186 -0.680 -0.448  0.004  0.004  0.017 -0.010 -0.145
```

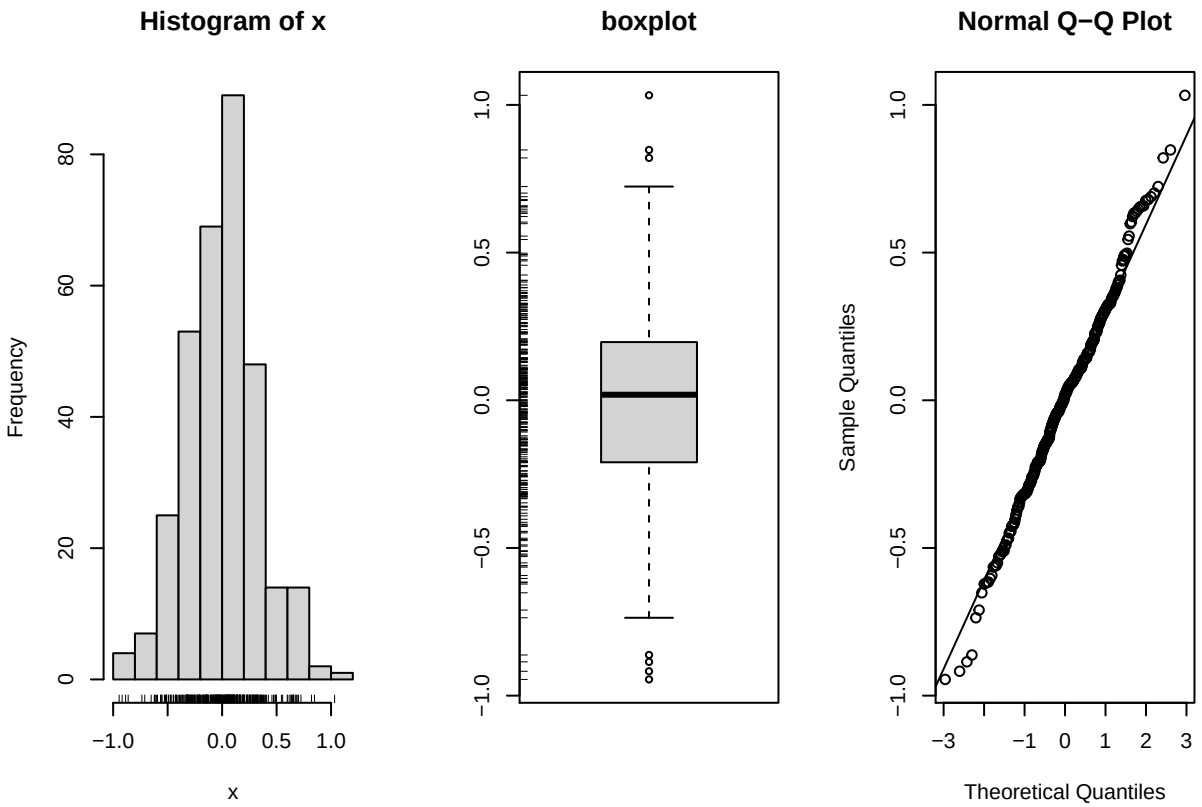
```
drop1(CEWL_mod3_t)
```

```
## Single term deletions
##
## Model:
## log(TEWL_g_m2h) ~ day * humidity_tmt_percent + region + cloacal_temp_C +
##     (1 | trial_number/individual_ID)
##                           npar    AIC
## <none>                         289.48
## region                     4 355.11
## cloacal_temp_C             1 362.73
## day:humidity_tmt_percent   1 353.42
```

```
plot(CEWL_mod3_t)
```




```
simple.eda(residuals(CEWL_mod3_t))
```



```
shapiro.test(residuals(CEWL_mod3_t))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(CEWL_mod3_t)
## W = 0.99524, p-value = 0.4174
```

L, E, and N are all satisfied now. :)

Export Best

The model is best with all the parameters currently included in model 3.

```
CEWL_mod_final <- CEWL %>%
  dplyr::filter(complete.cases(TEWL_g_m2h)) %>%
  lmerTest::lmer(data = .,
                 log(TEWL_g_m2h) ~
                 day*humidity_tmt_percent + region + cloacal_temp_C +
                 (1|trial_number/individual_ID))
write.csv(broom::tidy(CEWL_mod_final),
          "./best models/exp_effects_CEWL.csv")
```