Analysis of variance

Packages

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.2     v readr     2.1.4
v forcats   0.5.0     v stringr   1.5.0
v ggplot2   3.4.2     v tibble    3.2.1
v lubridate 1.9.2     v tidyr     1.3.0
v purrr     1.0.1
-- 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 beco
```

```
library(smmr)
library(PMCMRplus)
```

Jumping rats

- Link between exercise and healthy bones (many studies).
- Exercise stresses bones and causes them to get stronger.
- Study (Purdue): effect of jumping on bone density of growing rats.
- 30 rats, randomly assigned to 1 of 3 treatments:
  - No jumping (control)
  - Low-jump treatment (30 cm)
  - High-jump treatment (60 cm)
- 8 weeks, 10 jumps/day, 5 days/week.

- Bone density of rats (mg/cm^3) measured at end.
- See whether larger amount of exercise (jumping) went with higher bone density.
- Random assignment: rats in each group similar in all important ways.
- So entitled to draw conclusions about cause and effect.

Reading the data

Values separated by spaces:

```
my_url <- "http://ritsokiguess.site/datafiles/jumping.txt"
rats <- read_delim(my_url," ")
```

```
Rows: 30 Columns: 2
-- Column specification --------------------------------------------------------
Delimiter: " "
chr (1): group
dbl (1): density

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

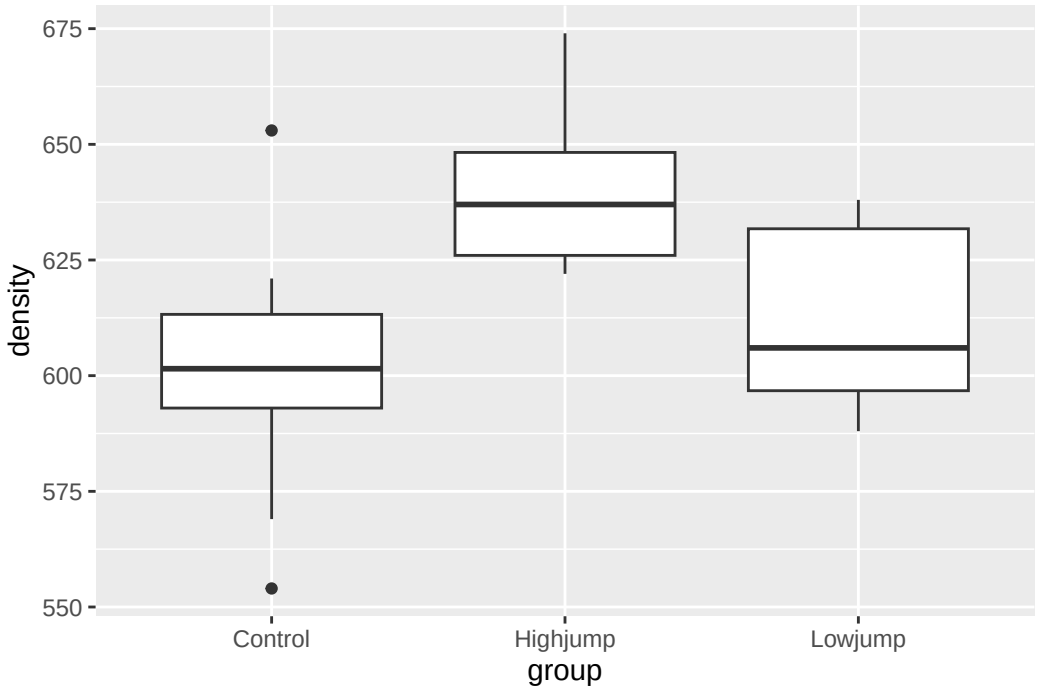
The data (some random rows)

```
rats %>% slice_sample(n=12)
```

```
# A tibble: 12 x 2
   group    density
   <chr>      <dbl>
 1 Control      569
 2 Control      614
 3 Highjump     622
 4 Control      554
 5 Highjump     631
 6 Control      593
 7 Highjump     622
 8 Highjump     650
 9 Control      621
10 Lowjump      605
11 Highjump     626
12 Highjump     650
```

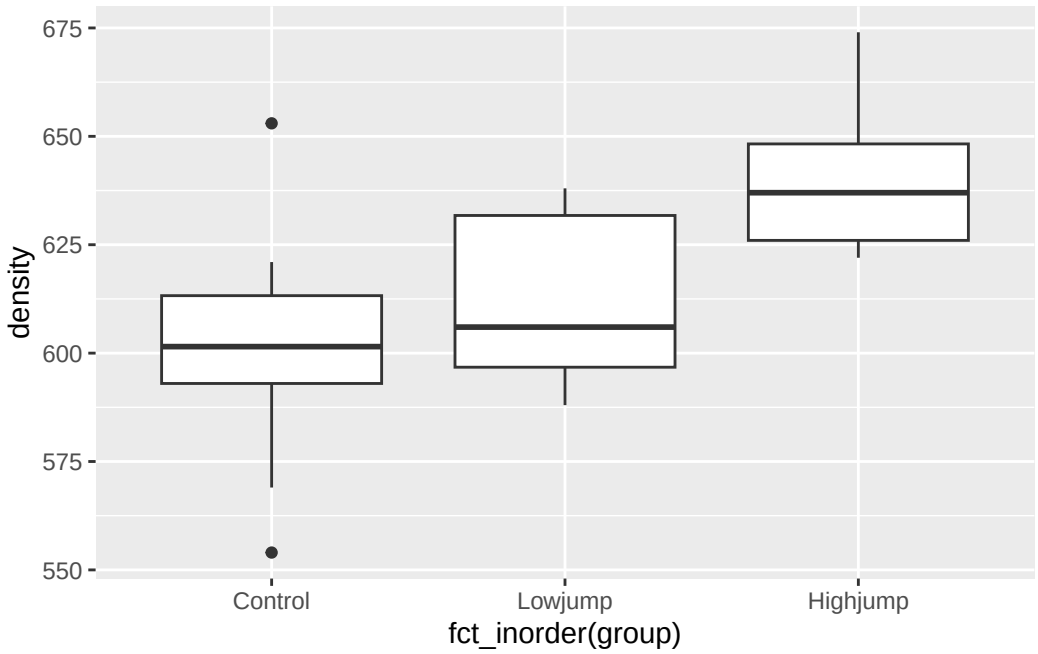
Boxplots

```
ggplot(rats, aes(y=density, x=group)) + geom_boxplot()
```



Or, arranging groups in data (logical) order

```
ggplot(rats, aes(y=density, x=fct_inorder(group))) +
  geom_boxplot()
```



Analysis of Variance

- Comparing > 2 groups of independent observations (each rat only does one amount of jumping).
- Standard procedure: analysis of variance (ANOVA).
- Null hypothesis: all groups have same mean.
- Alternative: "not all means the same", at least one is different from others.

Testing: ANOVA in R

```
rats.aov <- aov(density~group,data=rats)
summary(rats.aov)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
group        2   7434    3717   7.978 0.0019 **
Residuals   27  12579     466
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Usual ANOVA table, small P-value: significant result.
- Conclude that the mean bone densities are not all equal.
- Reject null, but not very useful finding.

Which groups are different from which?

- ANOVA really only answers half our questions: it says "there are differences", but doesn't tell us which groups different.
- One possibility (not the best): compare all possible pairs of groups, via two-sample t.
- First pick out each group:

```
rats %>% filter(group=="Control") -> controls
rats %>% filter(group=="Lowjump") -> lows
rats %>% filter(group=="Highjump") -> highs
```

Control vs. low

```
t.test(controls$density, lows$density)
```

```
    Welch Two Sample t-test

data:  controls$density and lows$density
t = -1.0761, df = 16.191, p-value = 0.2977
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -33.83725  11.03725
sample estimates:
mean of x mean of y
    601.1     612.5
```

No sig. difference here.

Control vs. high

```
t.test(controls$density, highs$density)
```

```
    Welch Two Sample t-test

data:  controls$density and highs$density
t = -3.7155, df = 14.831, p-value = 0.002109
```