Mixed effects Modelling 2 – fitting

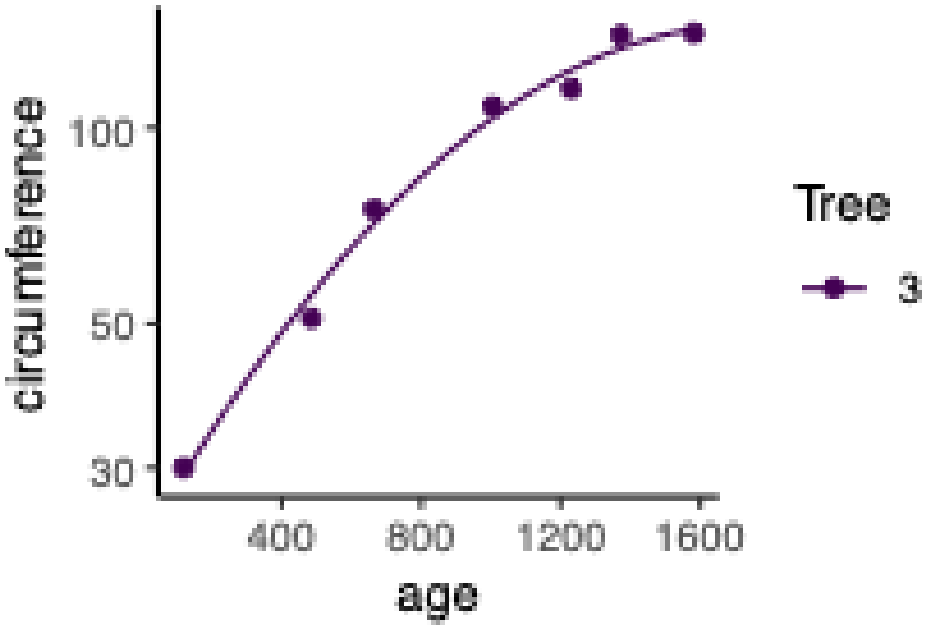
Starting point:

- Clarity about what is your response variable and what are explanatory variables – which are fixed and which random
- Scientific questions about the relationship between response and explanatory variables
- Data in a long-form dataframe or tibble

Example Experiment – variables

- Growth of orange trees

- Three variables:
  - Circumference (numeric, continuous)
  - Age (numeric, continuous, fixed)
  - Tree (numeric, categorical, random)



- Roles in experiment
  - **Response variable**
    - What we care about
    - What we measured
    - Want to know if it responds to explanatory variables
    - Not set as part of experiment
    - Biological and technical '**noise**' in the measurement
  - **Explanatory variable – fixed**
    - Something we can manipulate
    - What we manipulated
    - Want to know if it affects the response variable
    - Set precisely as part of the experiment
    - Little or no '**noise**' in the value
  - **Explanatory variable – random**
    - Something we have done or picked or sampled
    - What we sampled from
    - Want to know *how* it affects the response variable
    - Set as part of the experiment
    - No '**noise**' in the value

- Clarity about what is your response variable and what are explanatory variables – which are fixed and which random

Example Experiment – variables

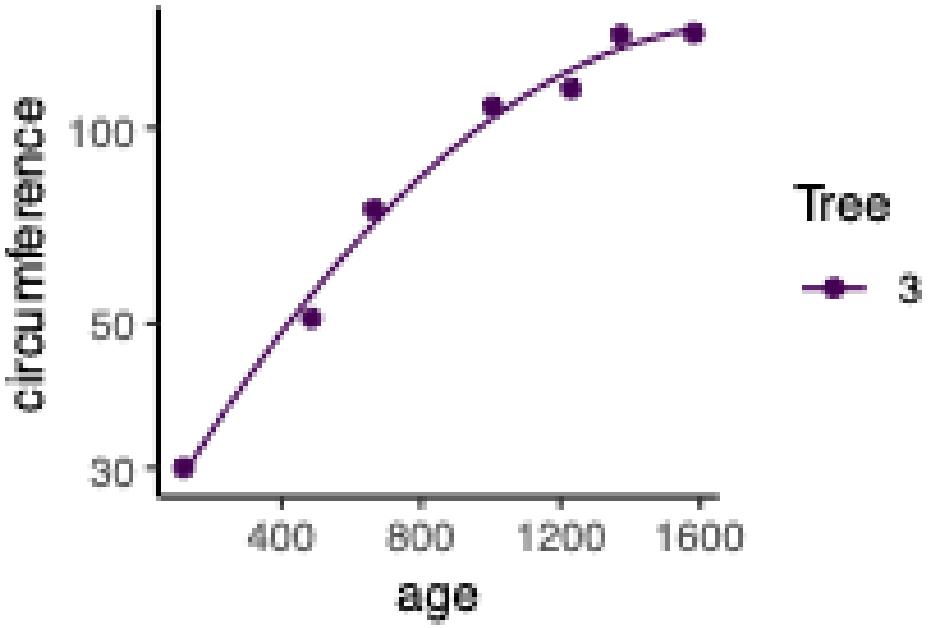
- Growth of orange trees

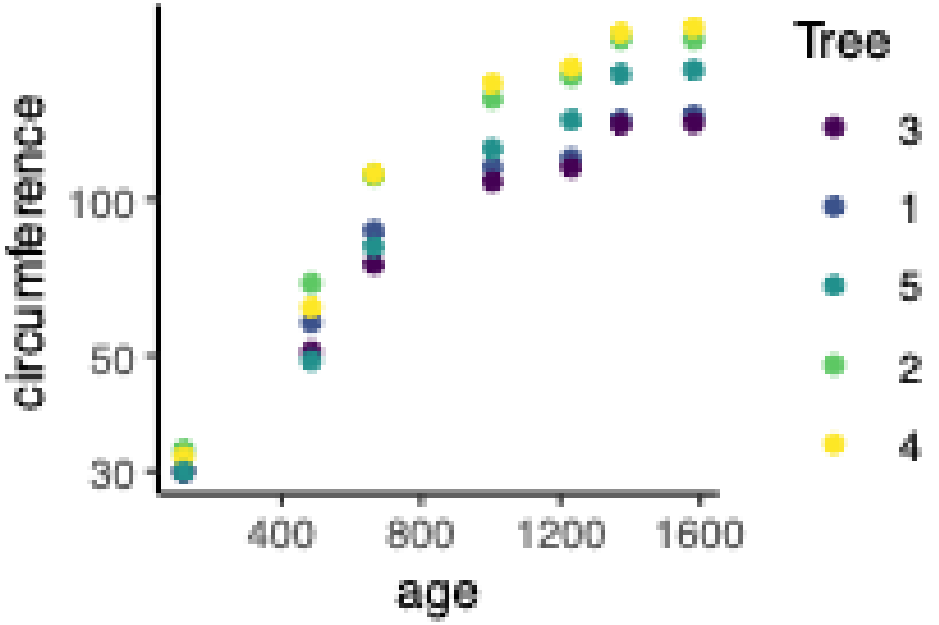
- Three variables:
  - Circumference (numeric, continuous)
  - Age (numeric, continuous, fixed)
  - Tree (numeric, categorical, random)

- Have a nice fit for an individual tree
- Want to fit a random effect at the same time to capture the variation among individual trees

```
lm(log(circumference) ~ age + I(age^2))

library(lme4)

lmer(log(circumference) ~ age + I(age^2) + (1|Tree))
```

- Various packages fit mixed effects models, espcially `nlme` and `brms`





Example Experiment – variables

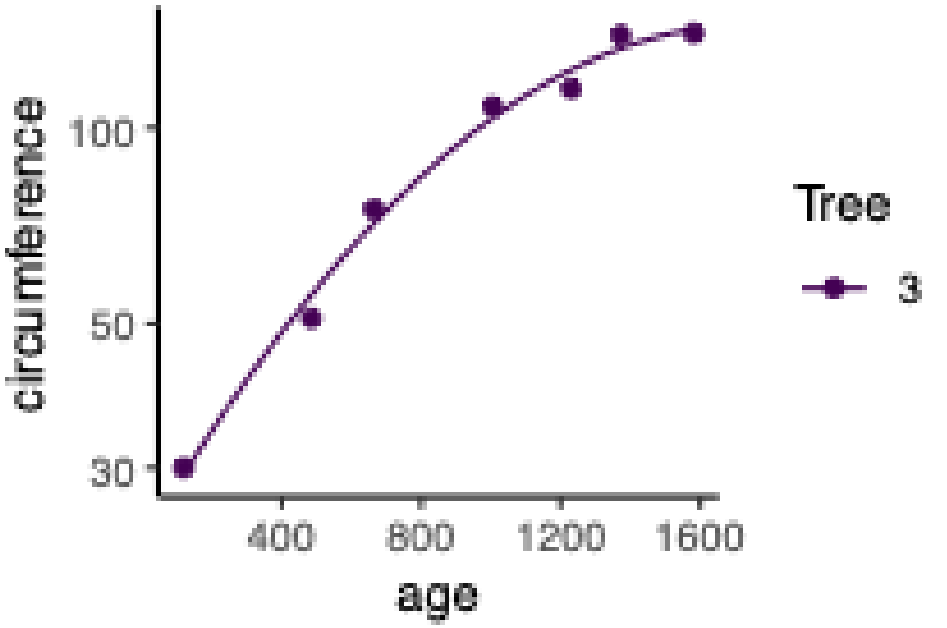
- Growth of orange trees

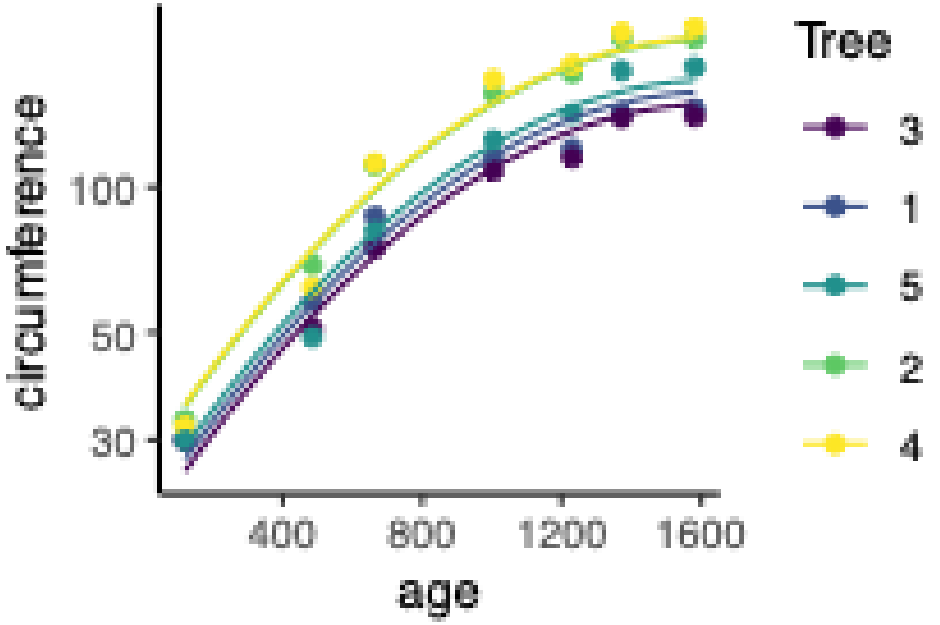
- Three variables:
  - Circumference (numeric, continuous)
  - Age (numeric, continuous, fixed)
  - Tree (numeric, categorical, random)

- Have a nice fit for an individual tree
- Want to fit a random effect at the same time to capture the variation among individual trees

```
lm(log(circumference) ~ age + I(age^2))

library(lme4)

lmer(log(circumference) ~ age + I(age^2) + (1|Tree))
```

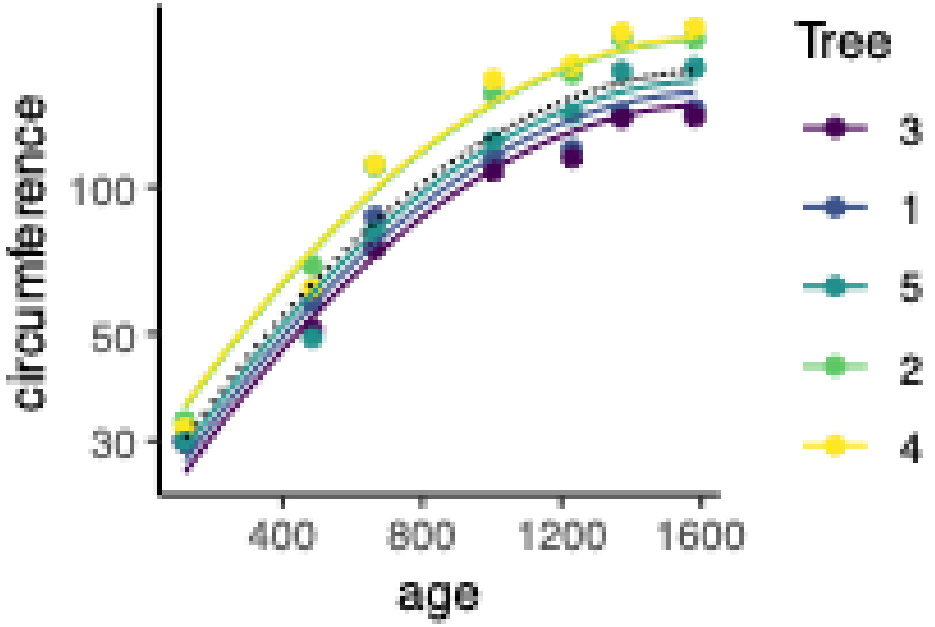
- Various packages fit mixed effects models, espcially `nlme` and `brms`





Example Experiment – variables

- Growth of orange trees
- Three variables:
  - Circumference (numeric, continuous)
  - Age (numeric, continuous, fixed)
  - Tree (numeric, categorical, random)



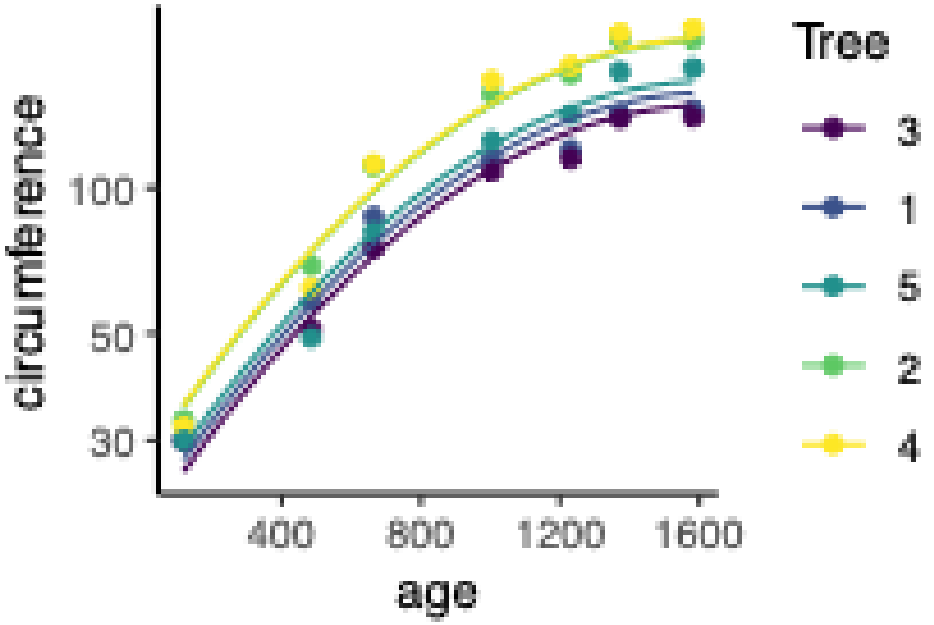

```
> summary(model)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: log(circumference) ~ age + I(age^2) + (1 | Tree)
   Data: Orange

REML criterion at convergence: 1.1

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.15991 -0.57319  0.01656  0.61931  1.54516

Random effects:
 Groups   Name        Variance Std.Dev.
 Tree     (Intercept) 0.022835 0.15111
 Residual             0.009594 0.09795
Number of obs: 35, groups:  Tree, 5

Fixed effects:
             Estimate Std. Error t value
(Intercept)  3.130e+00  8.654e-02  36.166
age          2.470e-03  1.448e-04  17.056
I(age^2)    -7.507e-07  8.206e-08  -9.148

Correlation of Fixed Effects:
        (Intr) age
age     -0.543
I(age^2)  0.470 -0.972
fit warnings:
Some predictor variables are on very different scales: consider rescaling
```

Example Experiment – questions

- Growth of orange trees

- Three variables:
  - Circumference (numeric, continuous)
  - Age (numeric, continuous, fixed)
  - Tree (numeric, categorical, random)




```
> summary(model)
```

```
Random effects:
 Groups   Name        Variance Std.Dev.
 Tree     (Intercept) 0.022835 0.15111
 Residual             0.009594 0.09795
Number of obs: 35, groups:  Tree, 5
```

- How much does circumference vary among trees?
- Pick either the Variance or SD to report (one is the square of the other)
  - Often SD, because it's on the same scale as the response variable
- Here the response variable is on a log scale, so 0.15 not so helpful
- Variances can be better because they add up
  - Proportion of variance that's among trees = Variance among trees/(Variance among trees + residual variance) = 0.023/(0.023 + 0.096) = 0.70
- i.e. 70% of the variation we see from the growth curve is variation among trees