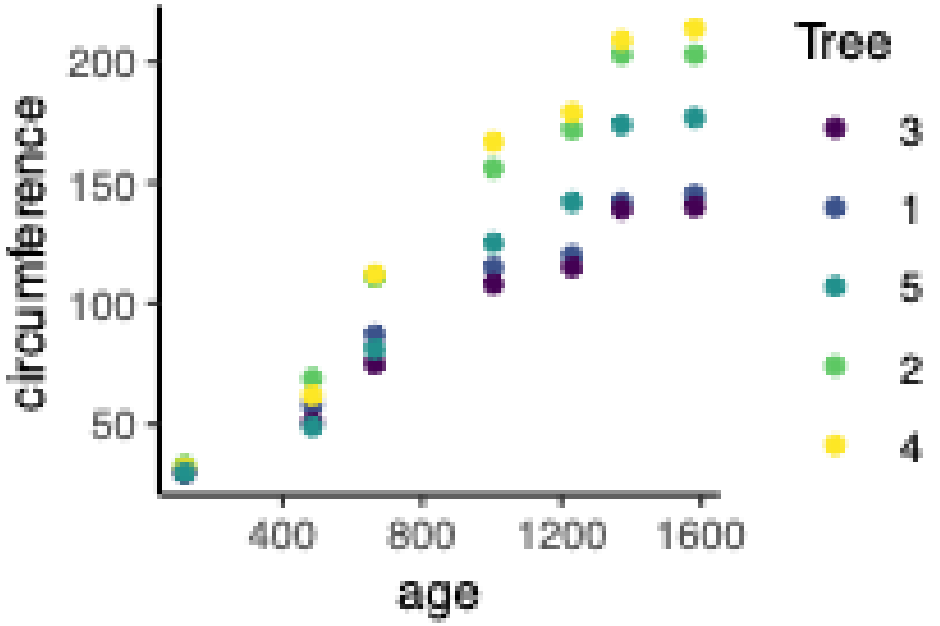
Mixed effects Modelling 1 – getting started

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

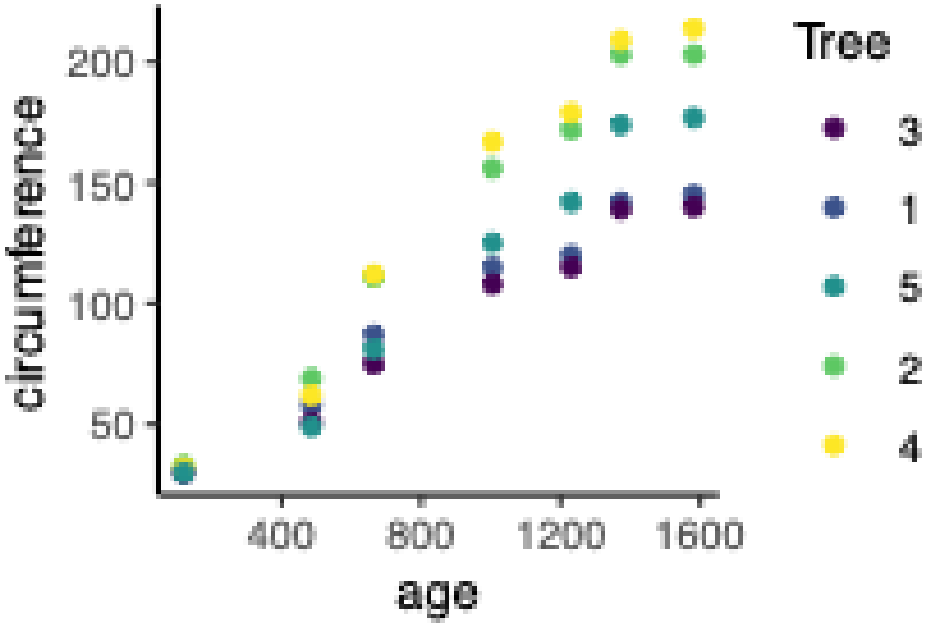
- Growth of orange trees
- Three variables:
  - Circumference (numeric, continuous)
  - Age (numeric, continuous, fixed)
  - Tree (numeric, categorical, random)



- 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 – questions

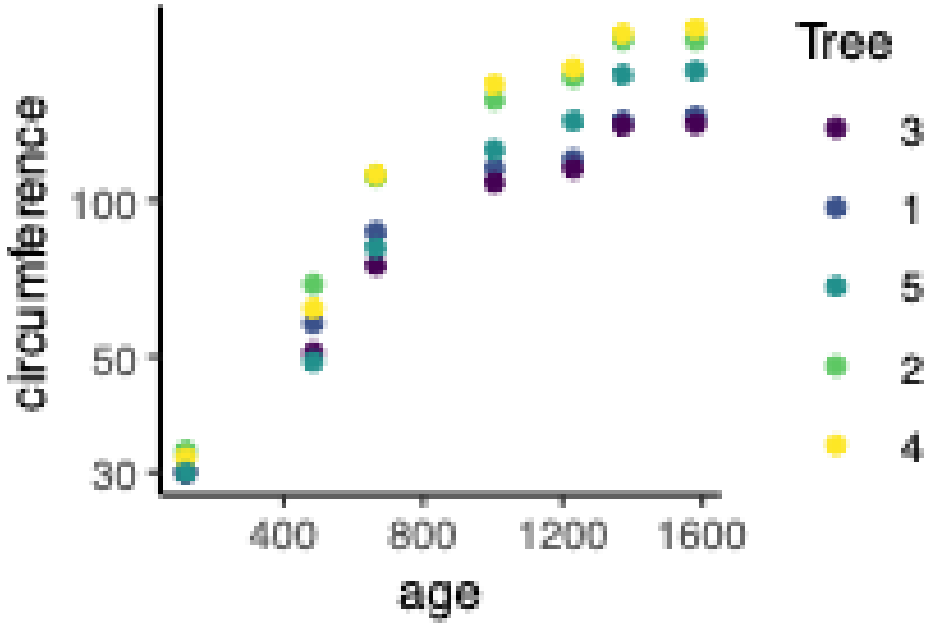
- Growth of orange trees
- Three variables:
  - Circumference (numeric, continuous)
  - Age (numeric, continuous, fixed)
  - Tree (numeric, categorical, random)
- Questions
  - How do orange trees grow?
- A bit too vague



- Scientific questions about the relationship between response and explanatory variables
- Clarity about what is your response variable and what are explanatory variables – which are fixed and which random

Example Experiment – data

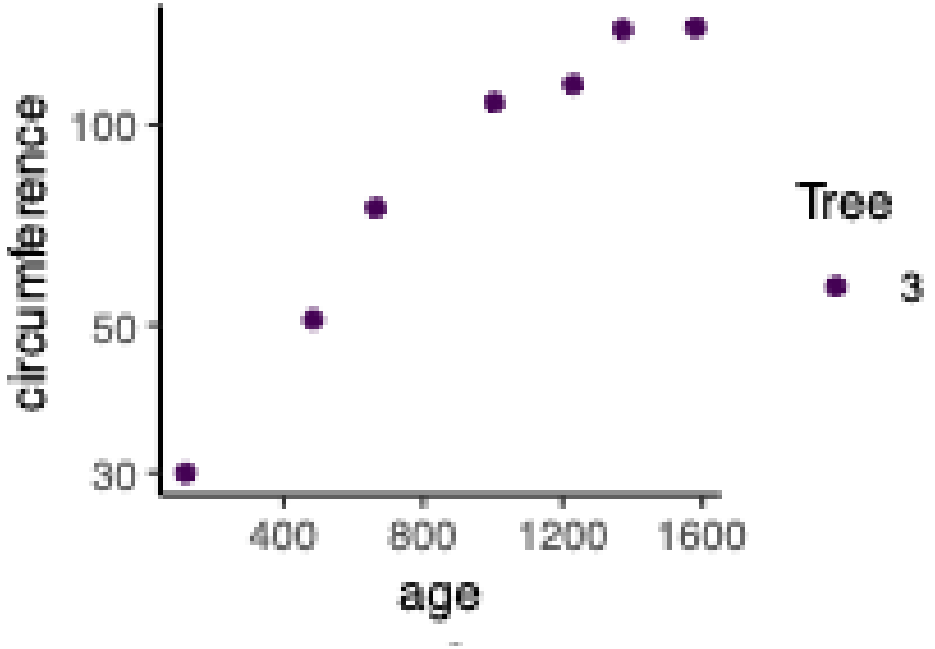
- Growth of orange trees
- Three variables:
  - Circumference (numeric, continuous)
  - Age (numeric, continuous, fixed)
  - Tree (numeric, categorical, random)
- Questions
  - How do orange trees grow?
- A bit too vague
- Data looks a bit problematic
  - Doesn't seem very linear
  - Really bunched at low levels and spread out at high ones
- Try looking at the data on a logarithmic scale
- Can be appropriate for growth
  - If grows a certain % per time, rather than a certain distance
  - Body size is never negative
  - Can sort out non-normality as well as heteroscedasticity
  - Often try to be more precise when measuring small things
- Scientific questions about the relationship between response and explanatory variables
- Clarity about what is your response variable and what are explanatory variables – which are fixed and which random



- Don't start reading up on non-linear modelling just yet
- '**Heteroscedasticity**' – a problem

- A log scale is a whole lot better
  (though heteroscedasticity not completely sorted)

Example Experiment – subset



- Three variables:
  - Circumference (numeric, continuous)
  - Age (numeric, continuous, fixed)
  - Tree (numeric, categorical, random)
- Questions
  - How do orange trees grow?
- A bit too vague
- Data still looks curvy
  - Try fitting a quadratic curve – ask does logarithmic growth slow with age?
- Start with one level of the random effect

```
> model <- lm(log(circumference) ~ age + I(age^2), data = Orange3)

> summary(model)
```

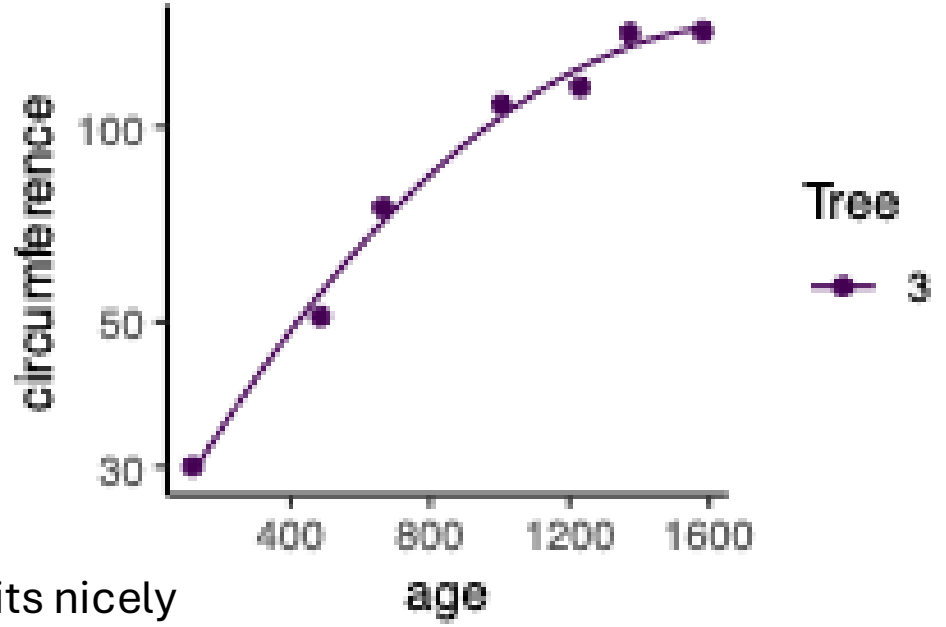
- Scientific questions about the relationship between response and explanatory variables
- Clarity about what is your response variable and what are explanatory variables – which are fixed and which random

Example Experiment – subset

- Growth of orange trees

```
> model <- lm(log(circumference) ~ age + I(age^2), data = Orange3)
> summary(model)

Call:
lm(formula = log(circumference) ~ age + I(age^2), data = Orange3)

Residuals:
       1        2        3        4        5        6        7
 0.02282 -0.08209  0.05004  0.04198 -0.06675  0.04741 -0.01341

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.139e+00  8.372e-02  37.494 3.02e-06 ***
age          2.098e-03  2.243e-04   9.355 0.000727 ***
I(age^2)    -6.006e-07  1.271e-07  -4.726 0.009132 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06784 on 4 degrees of freedom
Multiple R-squared:  0.9908,    Adjusted R-squared:  0.9862
F-statistic:   215 on 2 and 4 DF,  p-value: 8.491e-05
```

- Bad if want to extrapolate



- Fits nicely
- Good if just want parameters or hypothesis tests
  Does logarithmic growth slow with age?
- Yes, logarithmic growth slows with age $P = 0.009$