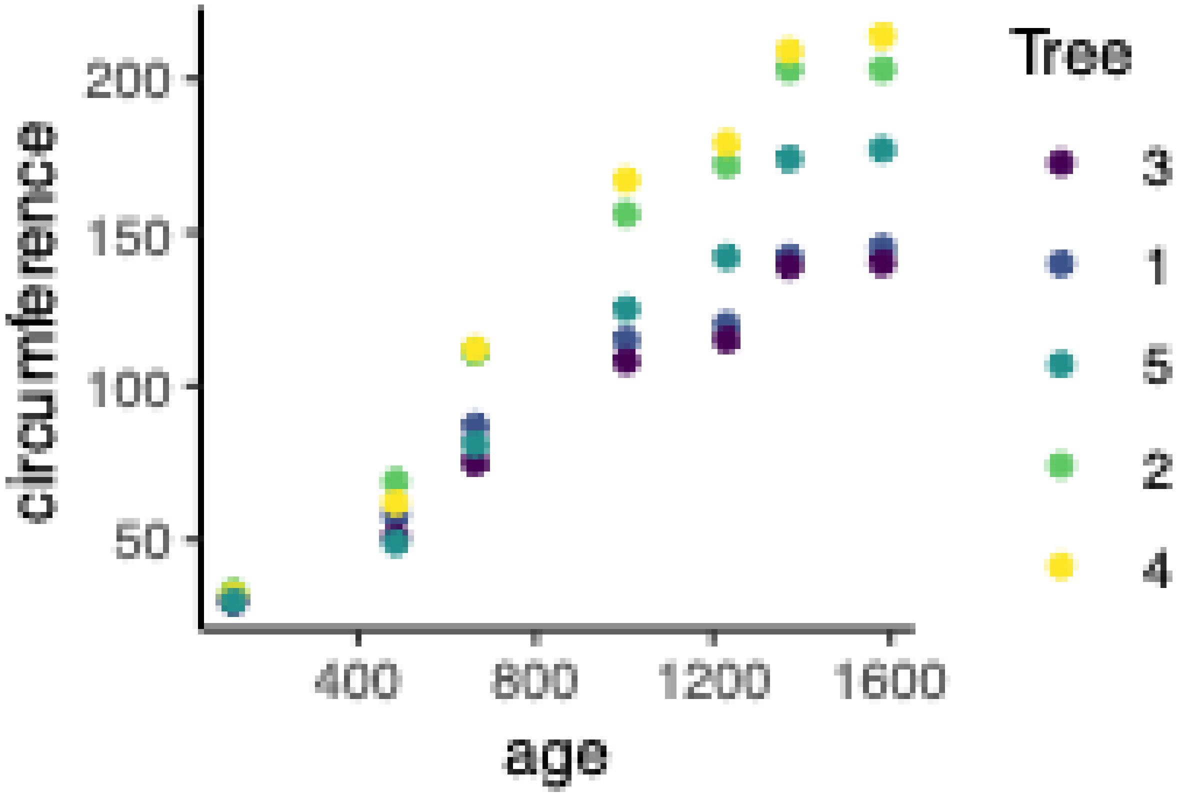
Fixed versus random effects

- `Orange` dataset – growth of orange trees

- **‘Repeated measures’** design
- Analysis needs to deal with both a fixed effect (age) and a random effect (Tree) so is a **‘mixed effects’** model



- Regression: numeric continuous response (circumference) with numeric continuous explanatory variable (age)
- Ok **BUT** only 5 trees – different measurements of the same tree are **pseudoreplicates** so need to include tree
- *Could* treat as ANCOVA with tree as a numeric categorical variable

```
lm(circumference ~ age * Tree, data = Orange)
```

- Fits lots of parameters, e.g. for exactly how Tree 4 is different from Tree 3

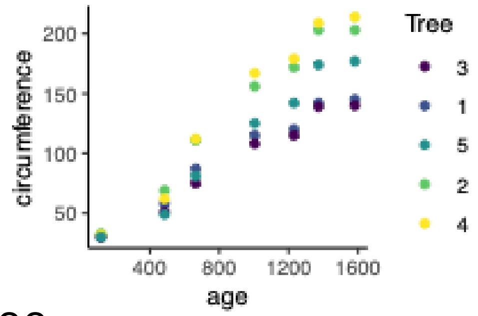
Do we care?

Not very parsimonious

What if there were 100 trees?

- Alternative: treat Tree as a ‘**random effect**’ and just estimate how much variation there is among trees
  - Reserve individual intercept/slope estimates for the ‘**fixed effect**’ (age)

Fixed versus random effects



Age | Tree

Fixed effect	Random effect
Interesting in itself	Often a nuisance
Response Mean of interest	Response Variability of interest
Measured without error	Random sample of possible values
Limited number of levels	Potentially infinite population
2 or more levels	At least 5 levels

Practically for fitting a model

- **Don't fit a random effect with <5 levels** (e.g. if your random effect is 'replicate')
- Not always obvious whether to treat an effect as fixed or random
  - See "Should I treat factor xxx as fixed or random?" at https://is.gd/glmmFAQ
- Can fit models with *only* random explanatory variables, or where the scientific question is about the random effects
  - known as '**Variance components** analysis'