L8: More Random Models…
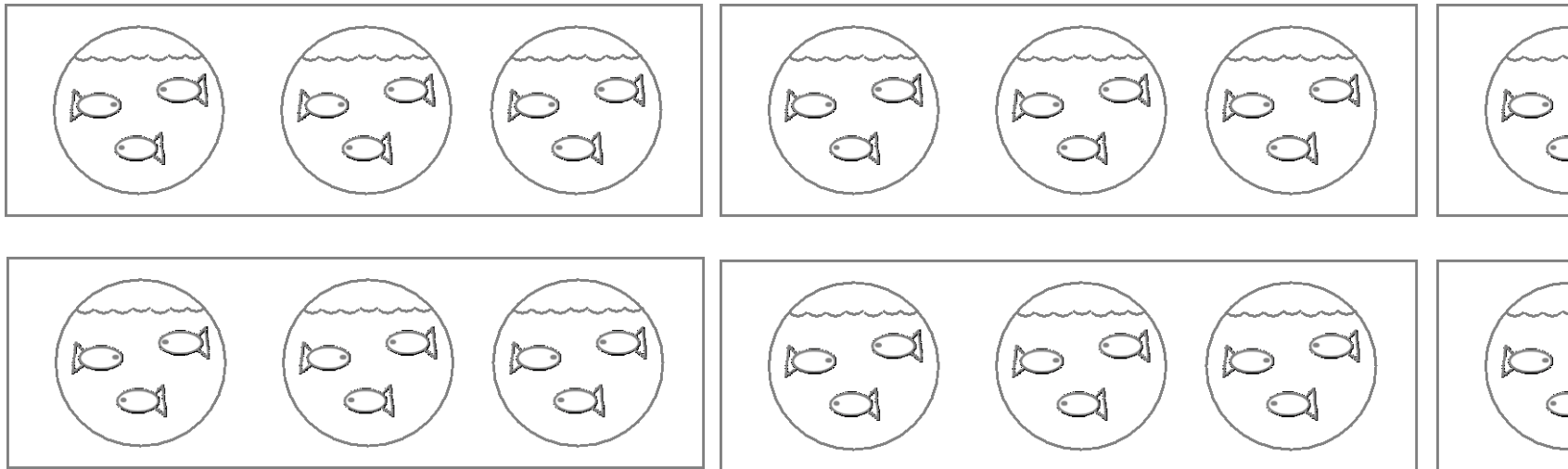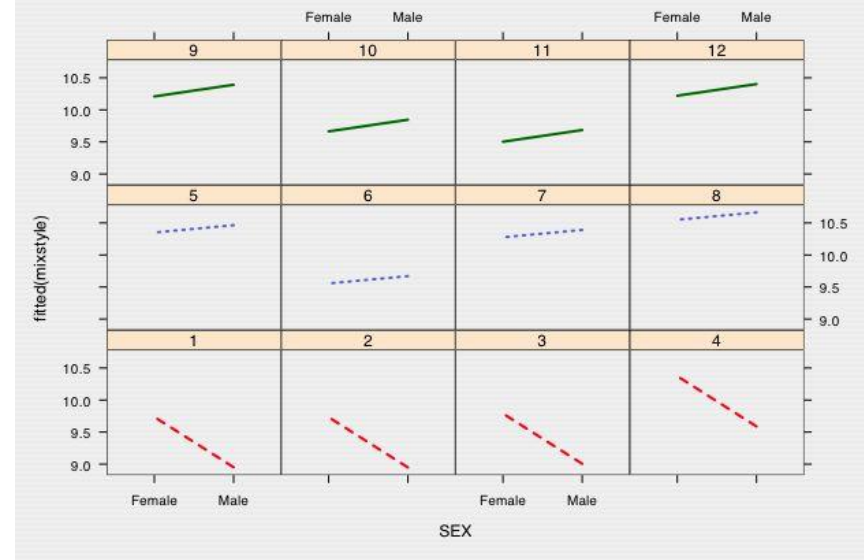
This morning…

- Intercepts were allowed to vary among nest boxes… under the influence of a treatment factor

- Model: $\texttt{weight = a.sex}_i \texttt{ + b.temp}_j \texttt{ + box}_j \texttt{ + error}_i$

- Could the effect of sex vary among nest boxes as well?

Data = Mean + Treatment + Error + Group Treatment + Group error + Group Treatment error

Example 2: Children's growth rates
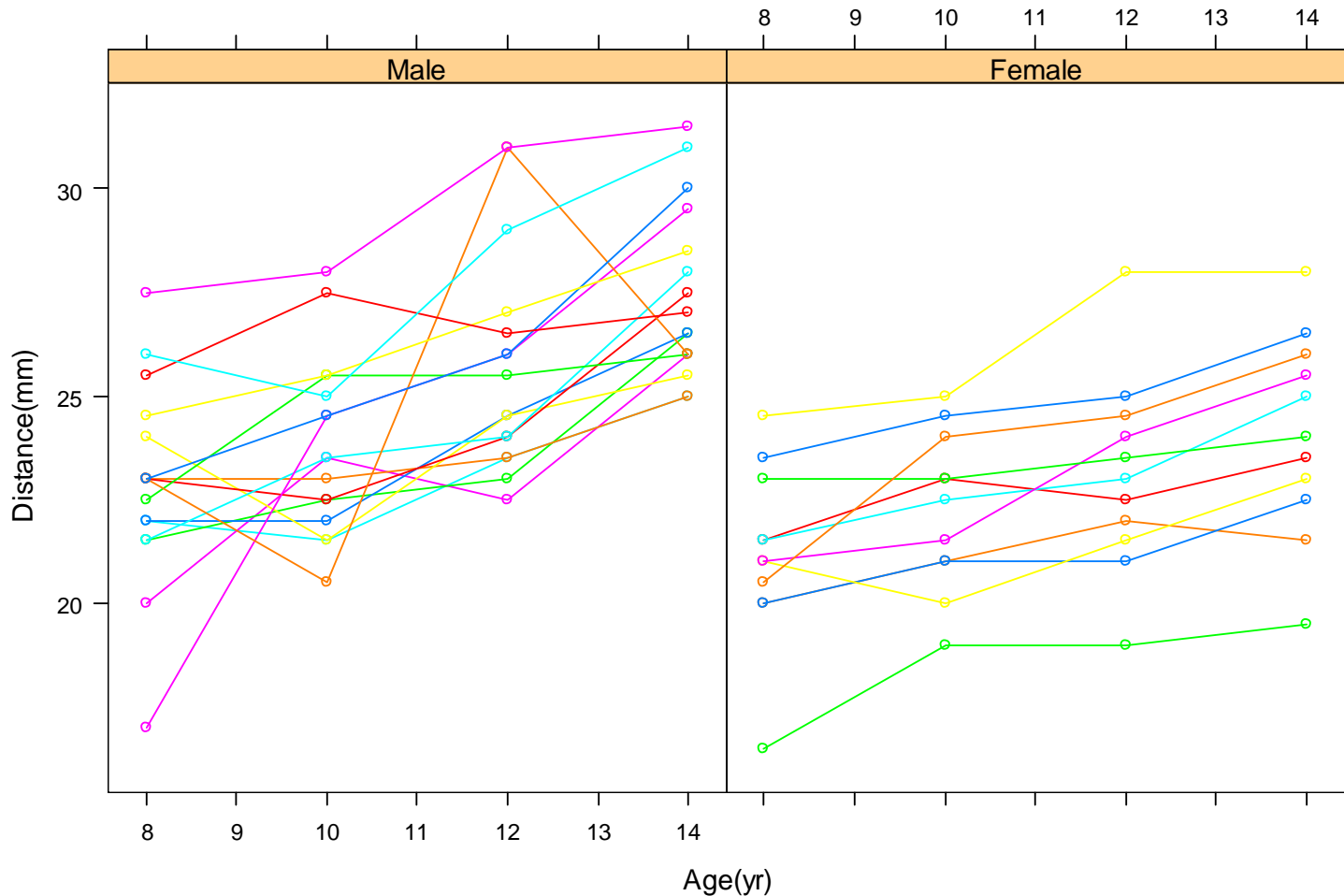
A longitudinal dataset:

Data of skull measurements from children taken every two years from 8 until 14 from 27 children (16 males and 11 females).

• Orthodont dataset provided with library(nlme)

• Do the growth rates differ among the boxes?
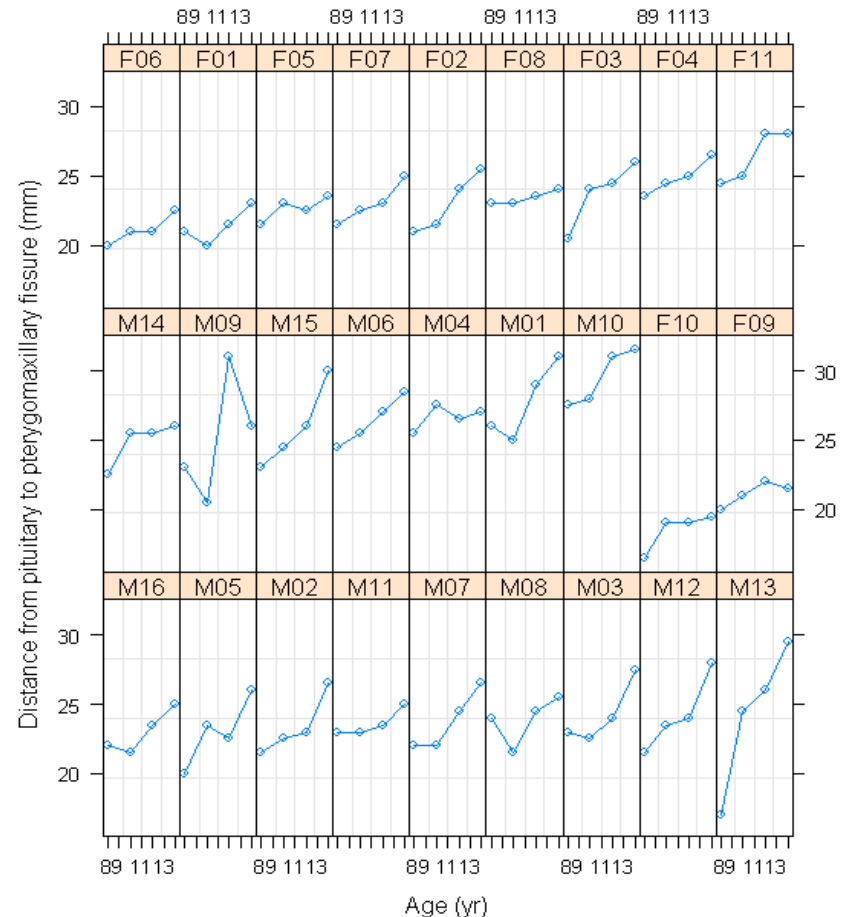
`plot(Orthodont)`

Are **slopes** the same among boxes, within each sex?

Allow varying intercepts only:

```
> lme1 <- lme(distance~Sex*age, data=Orthodont, random= ~1|Subject)

Linear mixed-effects model fit by REML
 Data: Orthodont
       AIC       BIC       logLik
  445.7572 461.6236 -216.8786

Random effects:
 Formula: ~1 | Subject
        (Intercept) Residual
StdDev:    1.816214 1.386382

Fixed effects: distance ~ Sex * age
                   Value Std.Error DF   t-value p-value
(Intercept)     16.340625 0.9813122 79 16.651810  0.0000
SexFemale        1.032102 1.5374208 25  0.671321  0.5082
age              0.784375 0.0775011 79 10.120824  0.0000
SexFemale:age   -0.304830 0.1214209 79 -2.510520  0.0141
:
:
Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-3.59804402 -0.45461690  0.01578364  0.50244657  3.68620794

Number of Observations: 108
Number of Groups: 27
```

Predictions with varying intercepts

`plot( augPred(lme1,level=c(0:1)) )`

The level= argument plots the predictions of

- the fixed effects for each sex (blue line) and
- the fixed effect plus random effect for each individual (purple line)

- *Note that the random effect has allowed the **intercept** of the slope to differ for each individual from the predicted fixed effect, but **slope** for fixed and random effect is the same for each subject*
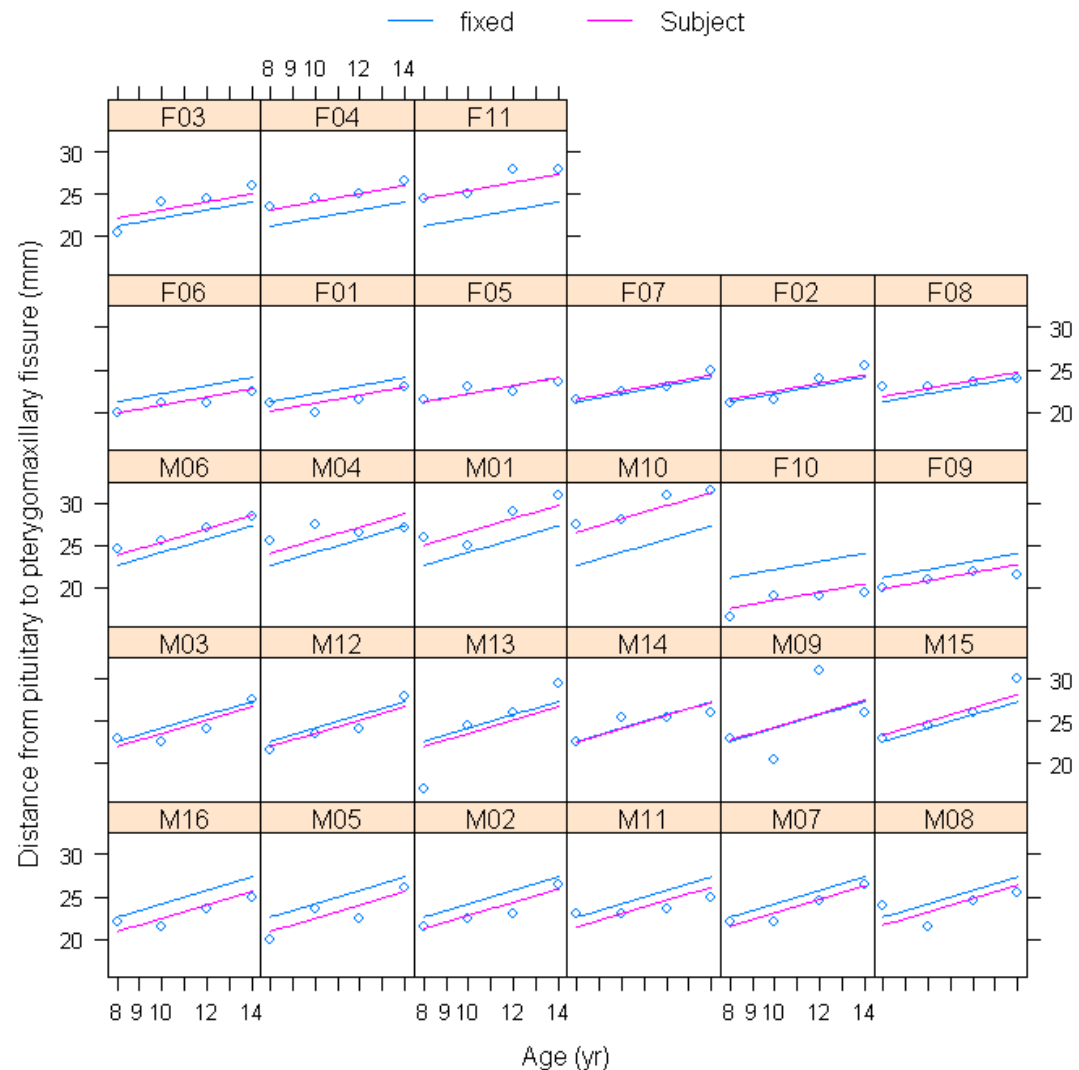
Allowing varying intercepts and slopes

```
> lme2 <- lme(distance ~ Sex*age, data=Orthodont, random= ~age|Subject)
> summary(lme2)
Linear mixed-effects model fit by REML
 Data: Orthodont
       AIC      BIC    logLik
  448.5817 469.7368 -216.2908

Random effects:
 Formula: ~age | Subject
 Structure: General positive-definite, Log-Cholesky
parametrization
            StdDev    Corr
(Intercept) 2.4059143 (Intr)
age         0.1804303 -0.668
Residual    1.3100046

Fixed effects: distance ~ Sex * age
                Value Std.Error DF   t-value p-value
(Intercept)   16.340625 1.0185753 79 16.042629  0.0000
SexFemale      1.032102 1.5958008 25  0.646761  0.5237
age            0.784375 0.0860090 79  9.119688  0.0000
SexFemale:age -0.304830 0.1347502 79 -2.262183  0.0264
:
:
Standardized Within-Group Residuals:
       Min          Q1         Med          Q3          Max
-3.167729145 -0.385826727  0.007027997  0.445020334  3.849531112

Number of Observations: 108
Number of Groups: 27
```
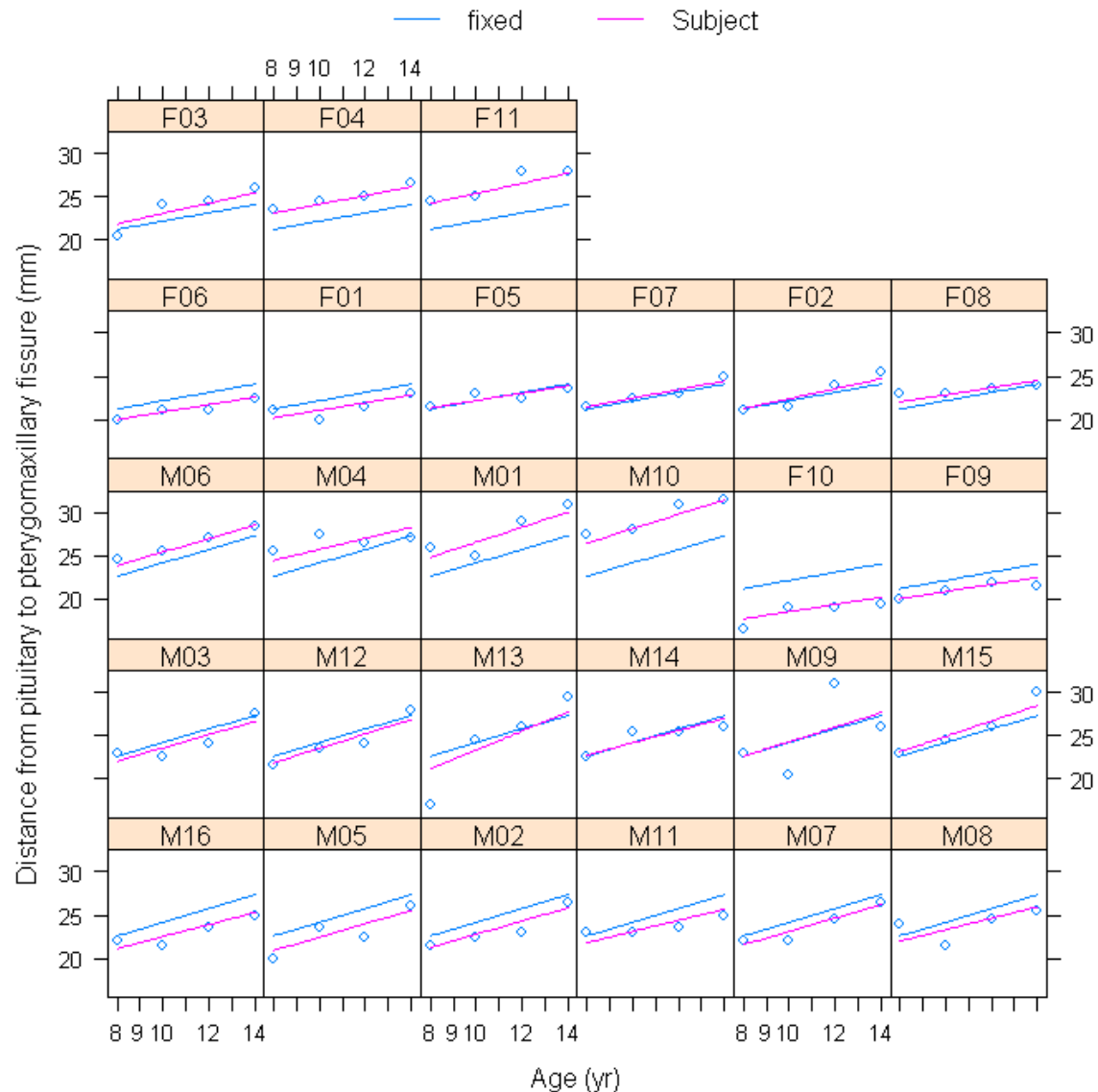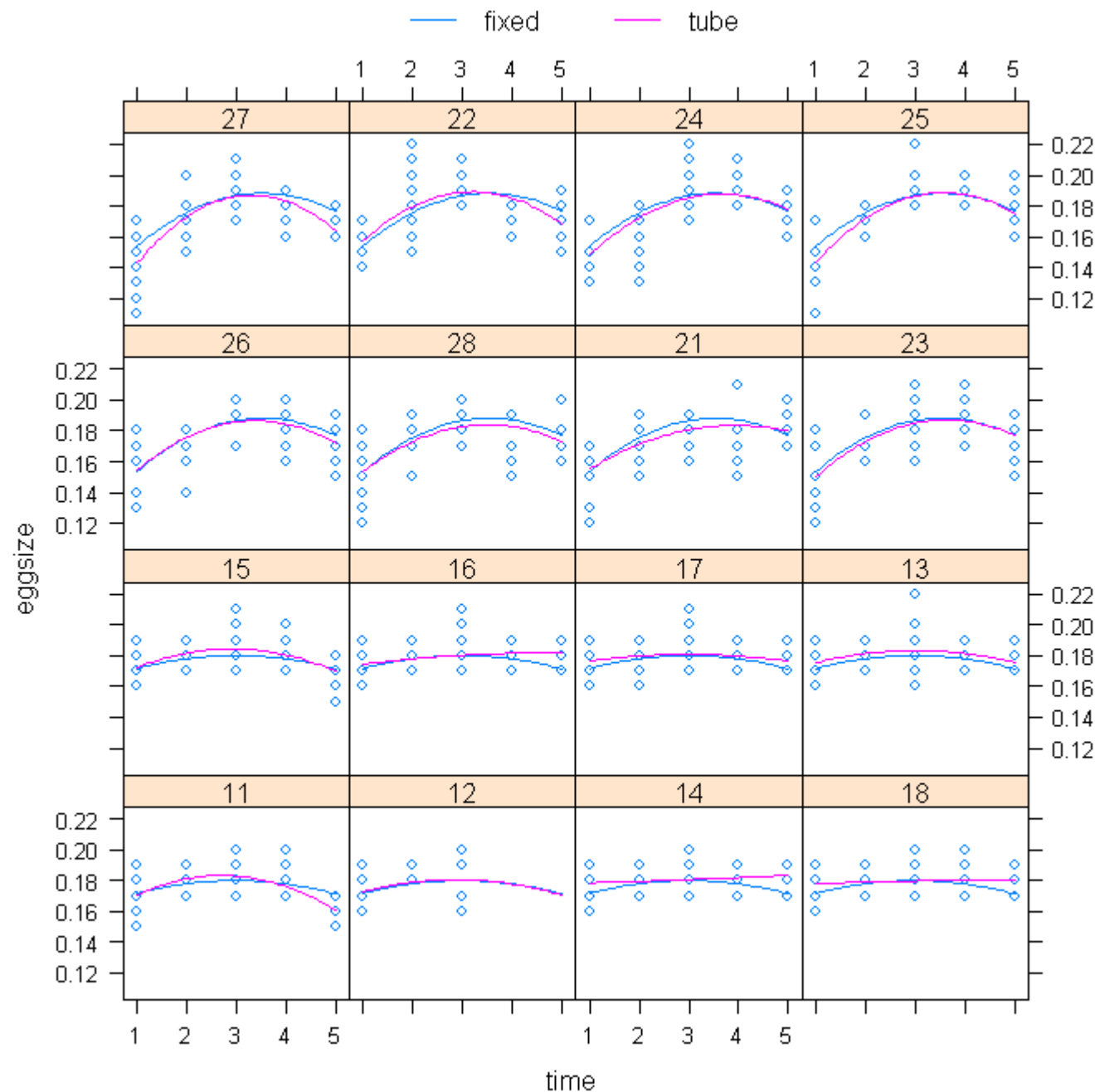
Predictions with varying intercepts and slopes

• Predictions for each subject now vary in both intercept and slope (purple line; compare blue line)

Paternal effects in mites

lme(
eggsize ~
poly(time,2) + food,

random =
~ maleID + poly(time,2)
 | tube)

Hypothesis testing

- Do groups need both slopes and intercepts to vary?

```
lme1 <- lme( distance~Sex*age, data=Orthodont, random= ~age|Subject )
lme2 <- lme( distance~Sex*age, data=Orthodont, random= ~1|Subject )

anova(lme1,lme2)
     Model df      AIC      BIC   logLik   Test  L.Ratio p-value
lme1     1  8 448.5817 469.7368 -216.2908
lme2     2  6 445.7572 461.6236 -216.8786 1 vs 2 1.175587  0.5556
```

→ No, a model with random intercepts appears sufficient.

- Is there a significant effect of sex on growth (i.e. the slope of the distance~age effect)?

```
anova(lme2)
            numDF denDF   F-value p-value
(Intercept)     1    79  4123.156  <.0001
Sex             1    25     9.292  0.0054
age             1    79   122.450  <.0001
Sex:age         1    79     6.303  0.0141
```

→ Yes, growth is significantly affected by sex.

Model structures
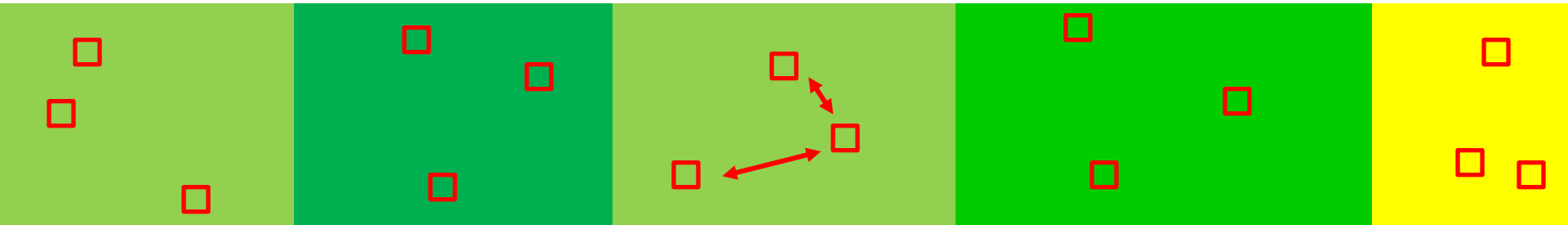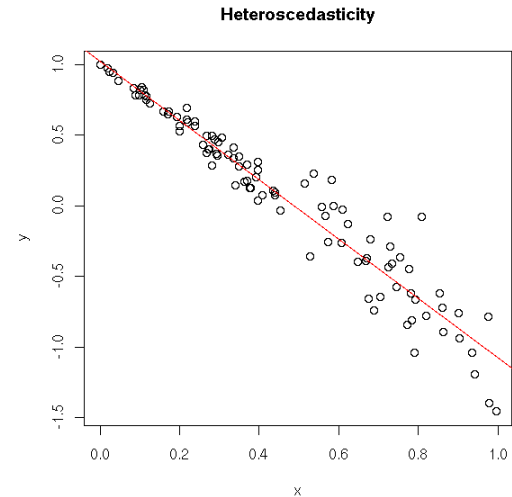# for mean and variance changing
in time and space

Complex model structures with library "nlme"

A "random factor" is a grouping variable.

– You may need to fit different **variances** for different groups.

"Random effects" may also include:

– **autocorrelation** if nearby replicates are more similar.

Heteroscedasticity

How do we model variance structures within a "regression" framework to incorporate changes in variance and autocorrelation?

(1) Variance changing with covariates

- Transform or fit appropriate model

- GLM with **`quasi(link=ident,var="mu")`**

- GLS with variance family
  - **`gls()`** if you don't have groups
  - Allows errors to be correlated and/or with different variances

Heteroscedasticity
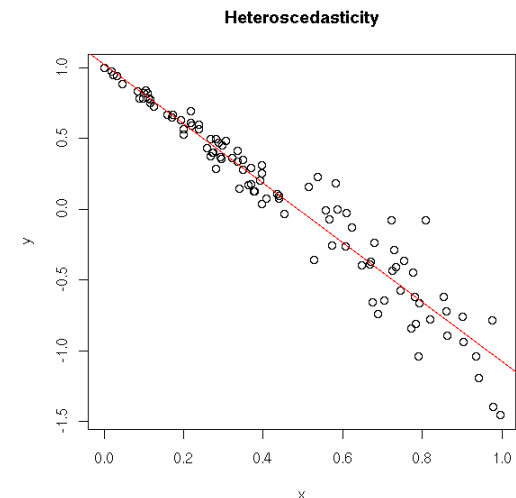
`varClasses`

- **`varIdent`**: constant variance(s), generally used to allow different variances according to the levels of a classification factor.
- **`varExp`**: exponential of a variance covariate.
- **`varPower`**: power of a variance covariate.
- **`varConstPower`**: constant plus power of a variance covariate.
- **`varFixed`**: fixed weights, determined by a variance covariate.
- **`varComb`**: combination of variance functions.
- Eg:

```
var <- varPower(form= ~size)
Mod1 <- gls(weight ~size, data=data, weights= var)
```
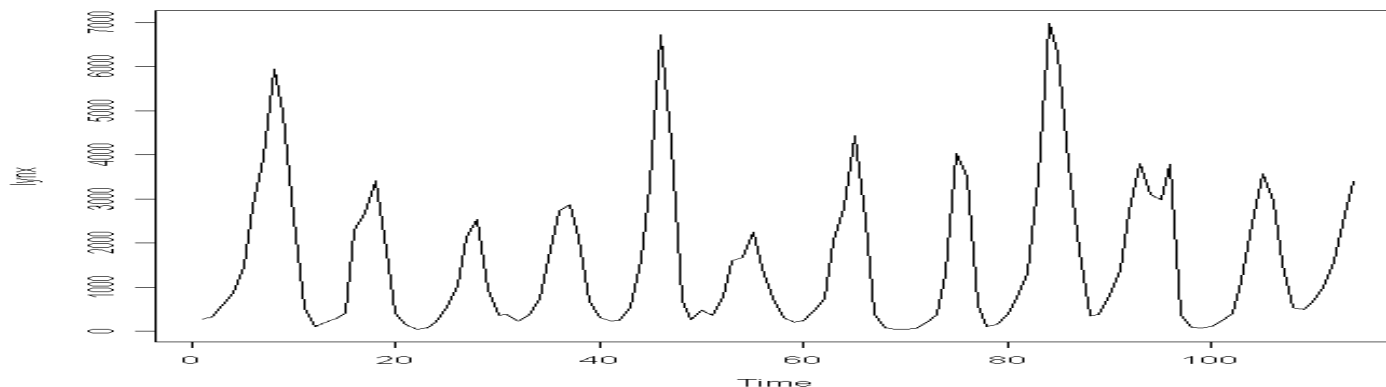
(2) Hierarchical grouping – spatial sampling designs

- E.g. repeated samples from different sites so you are looking at space as discrete

```
lme(plant_height ~ pesticide, random= ~1|farm/field)
```

Random effects structures:
(3) Correlation in time


- Temporal autocorrelation (TSA for another time)

```
> lme1 <- lme(distance ~ Sex*age, data=Orthodont, random=~1|Subject)
> lme3 <- update(lme1, correlation= corAR1())
> anova(lme3,lme1)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
lme3	1	7	447.7081	466.2188	-216.8541			
lme1	2	6	445.7572	461.6236	-216.8786	1 vs 2	0.04913688	0.8246

This tests for a correlation in the resids between consecutive points (i.e. autocorrelation of lag 1)

(4) Correlation in space

```
wheat <- lme( yield~Treatment,
   random= ~1|variety, data=Wheat3,
   correlation= corSpher(c(28,0.2),
        form= ~ latitude + longitude,
        nugget= TRUE) )
```
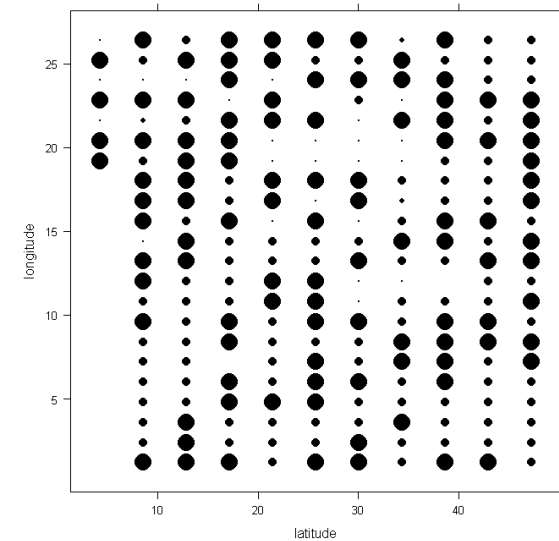

corSpher(c(28,0.2):

- 28 refers to <u>range</u> of correlations and
- 0.2 to the size of the *nugget* effect
  - correlation due to measurement error – how correlated are plants when close together? They should be about the same size, so plants measured at close points in space should be correlated at about zero, but it may be a small amount
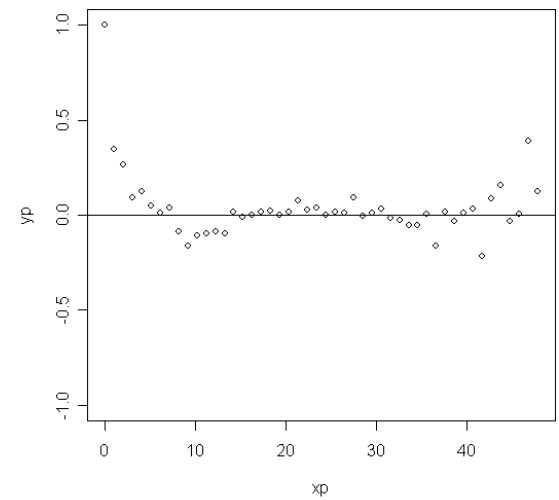
corClasses

- **corAR1**: autoregressive process of order 1.

- **corARMA**: autoregressive moving average process, with arbitrary orders for the autoregressive and moving average components.

- **corCAR1**: continuous autoregressive process (AR(1) process for a continuous time covariate).

- **corCompSymm**: compound symmetry structure corresponding to a constant correlation.

- **corExp**: exponential spatial correlation.

- **corGaus**: Gaussian spatial correlation.

- **corLin**: linear spatial correlation.

- **corRatio**: Rational quadratics spatial correlation.

- **corSpher**: spherical spatial correlation.

- **corSymm**: general correlation matrix, with no additional structure.

Non-Normal Errors

Modelling non-normal responses

- GLMs are a generalisation of LMs where error structure is non normal

- What do you do with LMEs where errors are not normal?

  - GLMMs

    ```
    glmm.pql()  library(MASS) – penalised QuasiL
    lmer()      library(lme4)
    ```

  - Transform and use lme()?

- Active research field… To avoid believing silly answers from GLMM, do transformed LME first as guide

Table I. Techniques for GLMM parameter estimation, their advantages and disadvantages and the software packages that implement them

Technique	Advantages	Disadvantages	Software
Penalized quasilikelihood	Flexible, widely implemented	Likelihood inference inappropriate; biased for large variance or small means	PROC GLIMMIX (SAS), GLMM (Genstat), glmmPQL (R), glmer (R)
Laplace approximation	More accurate than PQL	Slower and less flexible than PQL	PROC GLIMMIX [56], glmer (R), glmm.admb (R), AD Model Builder, HLM
Gauss-Hermite quadrature	More accurate than Laplace	Slower than Laplace; limited to 2–3 random effects	PROC GLIMMIX [56], PROC NLMIXED (SAS), glmer (R), glmmML (R)
Markov chain Monte Carlo	Highly flexible, arbitrary number of random effects; accurate	Very slow, technically challenging, Bayesian framework	WinBUGS, JAGS, MCMCpack, (R), AD Model Builder

Generalized linear mixed models: a practical guide for ecology and evolution

Benjamin M. Bolker[1], Mollie E. Brooks[1], Connie J. Clark[1], Shane W. Geange[2], John R. Poulsen[1], M. Henry H. Stevens[3] and Jada-Simone S. White[1]

[1] Department of Botany and Zoology, University of Florida, PO Box 118525, Gainesville, FL 32611-8525, USA
[2] School of Biological Sciences, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand
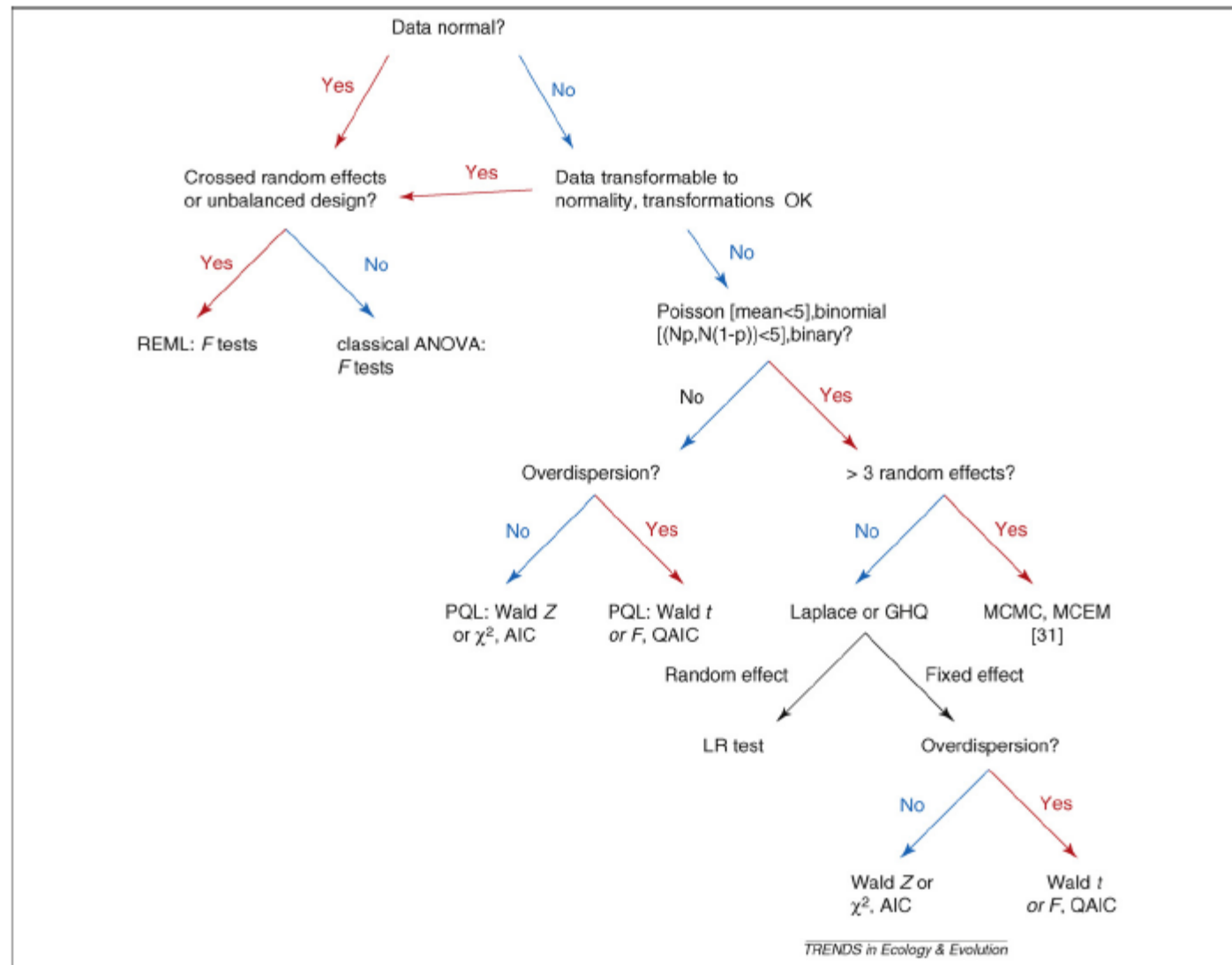[3] Department of Botany, Miami University, Oxford, OH 45056, USA

Figure 1. Decision tree for GLMM fitting and inference. Conditions on the Poisson and binomial distributions along the right branch refer to penalized quasilikelihood (PQL) rules of thumb [30]: to use PQL, Poisson distributions should have mean > 5 and binomial distributions should have the minimum of the number of successes and failures > 5. MCEM = Monte Carlo expectation-maximization [40].

Modelling non-normal responses

Various options in R (still a research topic)

```
> library(lme4)

> lmer1 <- lmer(y ~ x +(1|z), family=
  binomial)
```

- where z is the random factor

```
> library(MASS)

> glmm1 <- glmmPQL(distance ~ Sex*age,
  data=Orthodont, random= ~1|Subject,
  family="gaussian")
```

```
Linear mixed-effects model fit by REML
 Data: Orthodont
       AIC       BIC    logLik
  445.7572  461.6236  -216.8786


Random effects:
 Formula: ~1 | Subject
        (Intercept) Residual
StdDev:    1.816214 1.386382


Fixed effects: distance ~ Sex * age
                Value Std.Error DF   t-value p-value
(Intercept)  16.340625 0.9813122 79 16.651810  0.0000
SexFemale     1.032102 1.5374208 25  0.671321  0.5082
age           0.784375 0.0775011 79 10.120823  0.0000
SexFemale:age -0.304830 0.1214209 79 -2.510520  0.0141
```

**> lmer1 <- lmer(distance~sex*age
  +(1|Subject),
  family="gaussian",
  data=Orthodont)**
> summary(lmer1)

```
Linear mixed-effects model fit by maximum likelihood
 Data: Orthodont
  AIC BIC logLik
   NA  NA     NA


Random effects:
 Formula: ~1 | Subject
        (Intercept) Residual
StdDev:    1.740851 1.369159


Variance function:
 Structure: fixed weights
 Formula: ~invwt
Fixed effects: distance ~ Sex * age
                Value Std.Error DF   t-value p-value
(Intercept)  16.340625 0.9814310 79 16.649795  0.0000
SexFemale     1.032102 1.5376069 25  0.671239  0.5082
age           0.784375 0.0779963 79 10.056564  0.0000
SexFemale:age -0.304830 0.1221968 79 -2.494580  0.0147
```

**> glmm1 <- glmmPQL(
  distance~Sex*age,data=
  Orthodont,
  random=~1|Subject,
  family="gaussian")**
> summary(glmm1)

Ce**ll**
P R E S S

Generalized linear mixed models: a practical guide for ecology and evolution

Benjamin M. Bolker[1], Mollie E. Brooks[1], Connie J. Clark[1], Shane W. Geange[2], John R. Poulsen[1], M. Henry H. Stevens[3] and Jada-Simone S. White[1]

[1] Department of Botany and Zoology, University of Florida, PO Box 118525, Gainesville, FL 32611-8525, USA
[2] School of Biological Sciences, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand
[3] Department of Botany, Miami University, Oxford, OH 45056, USA