ST606 - The growth(height) curve analysis for the 6–18-year-old children

The growth curve analysis by using the SuperImposition by Translation And Rotation (SITAR) model

Yang Wang (19250003)

Contents

1 Abstract

[TODO]

2 Introduction

[TODO]

3 Background

[TODO]

A growth curve is an empircal model of the evolution of a quantity over time. The growth curves are widely used in biology for quantities such as population size or biomass (in population ecology and demography, for population growth analysis), individual body height or biomass (in physiology, for growth analysis of individuals).

Growth is influenced by many factors such as heredity, genetic or congenital factors, illness and medications, nutrition, hormones, and psychosocial environment. A child growth curve is an important standard to help parents to find out whether or not their children follow a normal growth pattern. In fact, hundreds of medical issues can affect children's growth. An abnormal growth pattern is often the earliest indication of an underlying medical condition. Following a normal growth pattern is a good indicator of a child's overall good health.

The growth curve analysis is a statistical issue in life course. Height in puberty involves a growth spurt, the timing and intensity of which varies between individuals.

4 Methods

I used the SuperImposition by Translation And Rotation (SITAR) model, a shape invariant model with a single fitted curve. Curves for individuals are matched to the mean curve by shifting their curve up–down (representing differences in mean size) and left–right (for differences in growth tempo), and the age scale is also shrunk or stretched to indicate how fast time passes in the individual (i.e. velocity). These three parameters per individual are estimated as random effects while fitting the curve. The outcome is a mean curve plus triplets of parameters per individual (size, tempo and velocity) that summarize the individual growth patterns.

4.1 The SITAR model

The method used to summarize the individual growth curves. This is a shape invariant model that involves fitting the following random effects model to the set of height growth curves.

$$y_{it} = \alpha_i + h\left(\frac{t - \beta_i}{\exp(-\gamma_i)}\right)$$

where yit is height for subject i at age t, h(t) is a natural cubic spline curve of height vs age, and αi, βi and γi are subject-specific random effects. The aim is to choose the values of αi, βi and γi to make the individual growth curves as similar as possible.

1. αi is a random height intercept that adjusts for differences in mean height—here it is termed size. Geometrically it can be thought of as a subject-specific shift up or down or translation in the spline curve, with αi smaller for shorter children.

2. βi is a random age intercept to adjust for differences in the timing of the pubertal growth spurt in individuals, and it is here called tempo. Geometrically it corresponds to a subject-specific left–right shift or translation in the spline curve, with βi negative for early puberty and positive for late.

3. γi is a random age scaling that adjusts for the duration of the growth spurt in individuals. Its parameterization as exp(–γi) ensures that both positive and negative values are permissible,

with zero corresponding to average velocity, and the minus sign means that it measures velocity rather than its inverse.

Geometrically it corresponds to a shrinking or stretching of the age scale. Thus for a child with a relatively short spurt and a steep growth curve, γi is positive to stretch the age scale and reduce the slope (i.e. peak velocity). Conversely, if γi is negative the age scale is shrunk and the curve slope increased. For this reason the parameter is termed velocity.

5 Clean Data

1. Read the data from excel file.
2. Delete the data that the height is NA.
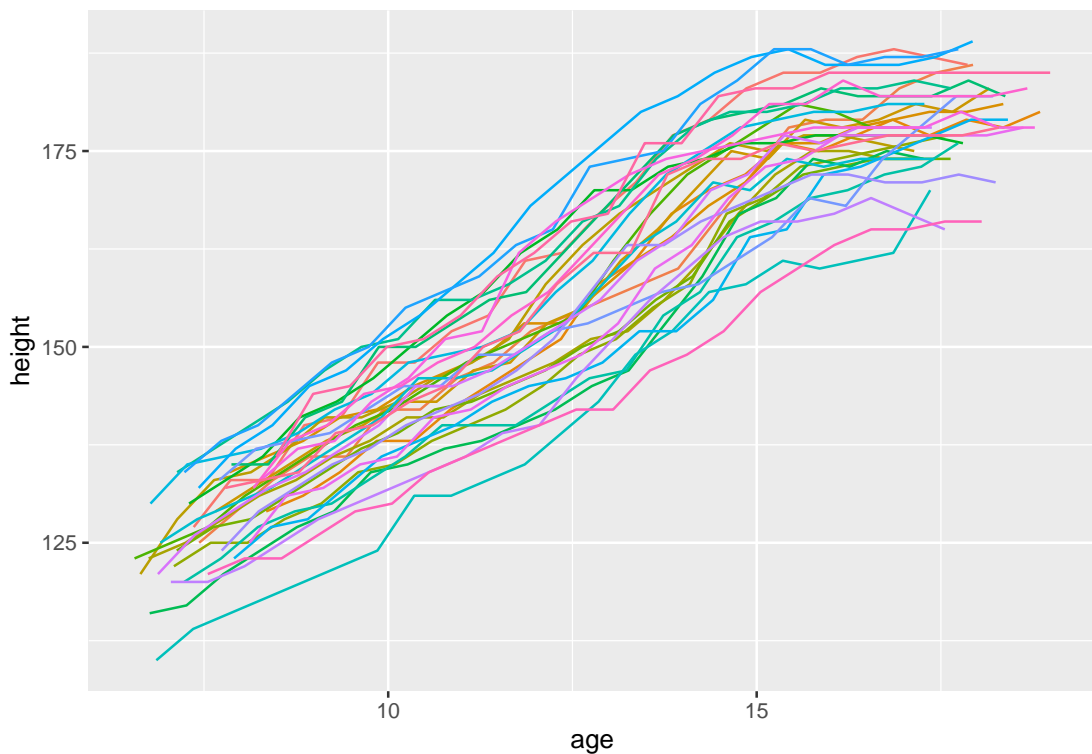3. Rename the column name.

```
##       All Object Deleted
## 1 102642  94888     7754
```

6 The DataSets

The datasets are heights for boy(31) and girl(38) at the stage between 2007 and 2018 (6-18 year-old), and have the complete set of 22 observations between the ages of 6 and 18 years.

7 The first dataset (Boys)

The 31 boys who have the complete set of 22 observations.

7.1 Fit the data to find the best model
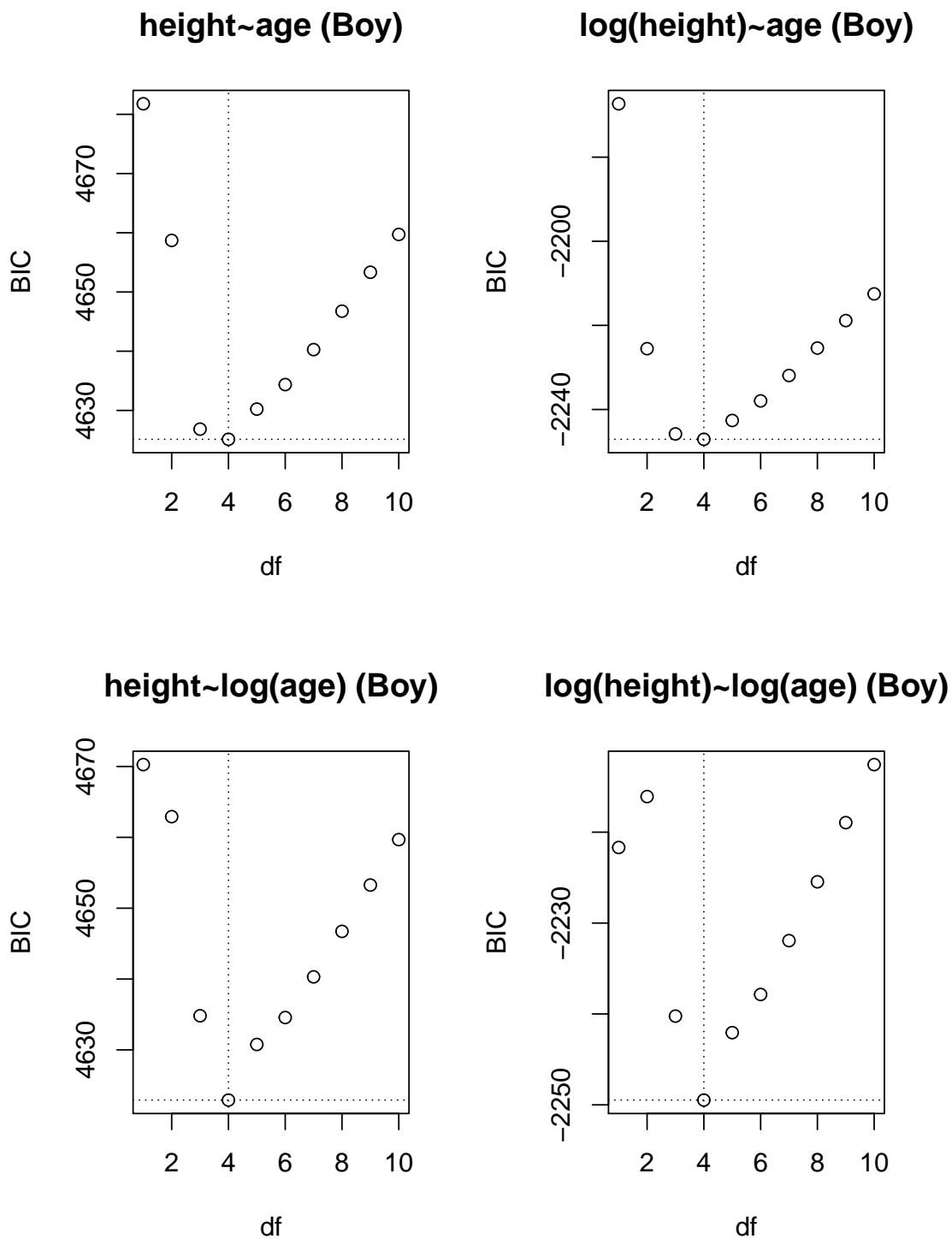
model 1: height~age

model 2: log(height)~age

model 3: height~log(age)

model 4: log(height)~log(age)

7.2 Find the the degrees of freedom (By BIC)

The degrees of freedom (DF) indicate the number of independent values that can vary in an analysis without breaking any constraints.

When Df = 4, the BICs are the lowest, choose the df = 4 for the four models.

7.3 Fit the models by using sitar model

fit the four models and compare the variance, residual and AIC to find the best model.

```
## Warning in nlme.formula(y ~ fitnlme(x, s1, s2, s3, s4, a, b, c), fixed = s1 + :
## Iteration 3, LME step: nlminb() did not converge (code = 1). PORT message: false
## convergence (8)
```

```
##                  formula variance    residual        AIC Converge
## 1            height~age    96.56   1.2995116    2718.42     True
## 2       log(height)~age    96.47   0.00854324  -4139.344    false
## 3       height~log(age)   96.45   1.31863340   2735.403     True
## 4 log(height)~log(age)   96.41 0.008600255   -4131.781     True
```
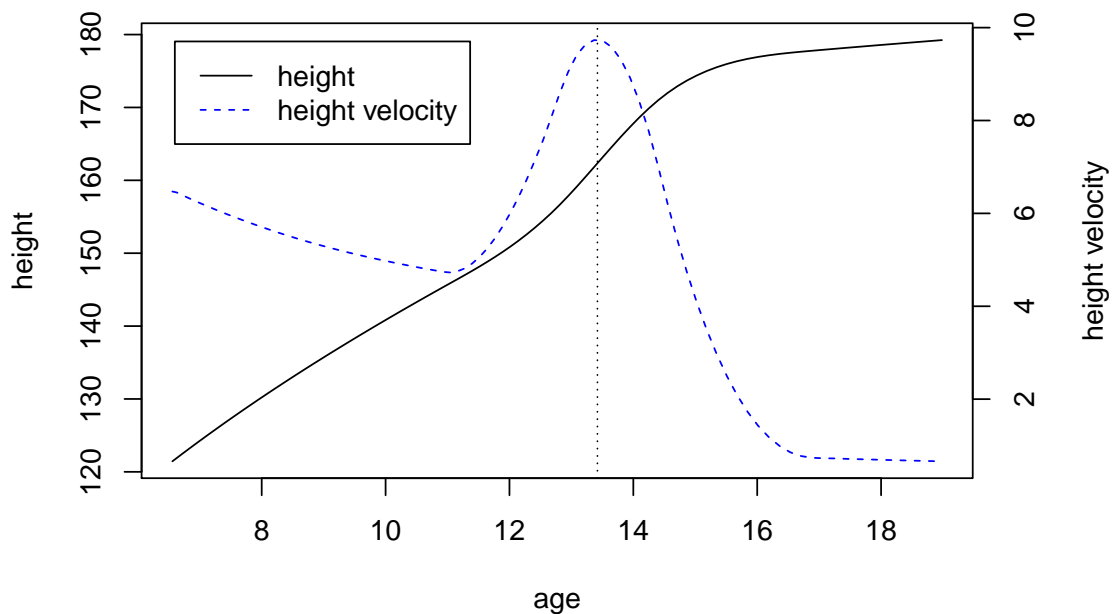
Although, the log(height)~age model is the best, however, it has convergence error. So I chose the log(height)~log(age) model.

7.4 The distance and velocity curves

Distance curve(left y axis)(Black solid line): the amount of height achieved at a given age.

Velocity curve(right y axis)(Blue dash line): the rate of growth at a given age.


```
##    apv     pv
## 13.420  9.739
```

1. The growth rate:

   The boys grow fast between 6~16 year-old, grow slowly after 16 year-old.

2. The growth velocity:

   a) It reached the peak of 9.685 at the age of 13.42.
   b) It declined slowly between 6~11 year-old.
   c) It increased rapidly between 11~13.42 year-old.
   d) It declined rapidly between 13.42~17 year-old.
   e) It is relatively constant after 17 year-old.

7.5 Check the significant of fixed and random effect factors

```
##      numDF denDF   F-value p-value
## s1      1   645     977.3  <.0001
## s2      1   645    1660.6  <.0001
## s3      1   645   24774.2  <.0001
## s4      1   645  289070.1  <.0001
## a       1   645  466446.2  <.0001
## b       1   645       3.8  0.0532
## c       1   645      58.0  <.0001
```

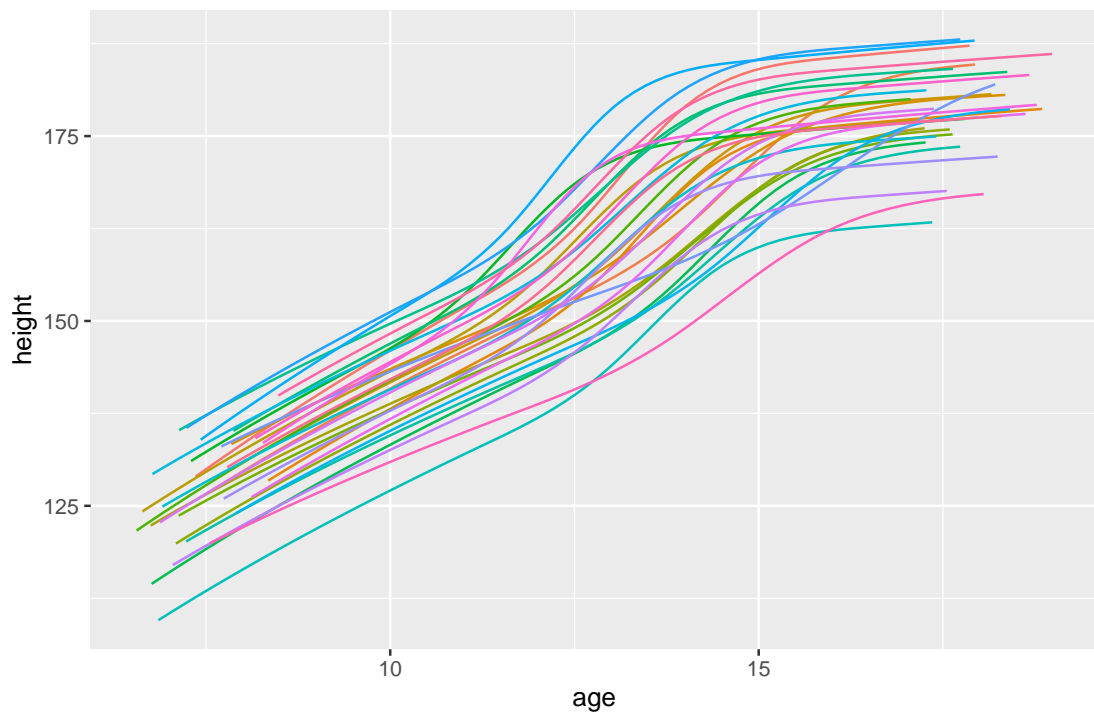The fixed effect s1, s2, s3, s4 are singificant.(p-value < 0.05).

The random effect of α(a) and γ(c) are singificant.(p-value < 0.05).

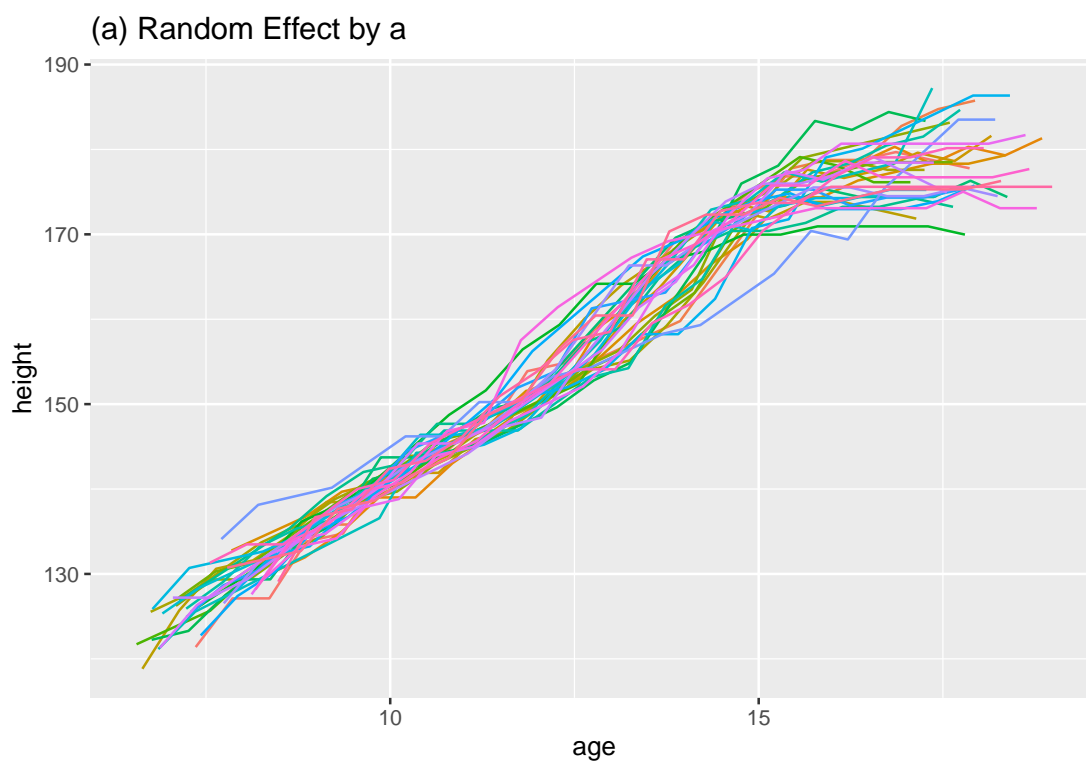The random effect of β(b) is not significant(p-value > 0.05).

7.6 Compare the random effect $(\alpha(a) + \beta(b) + \gamma(c))$

Increase each random effect SITAR parameter in model, figure out the parameters how to effect the model
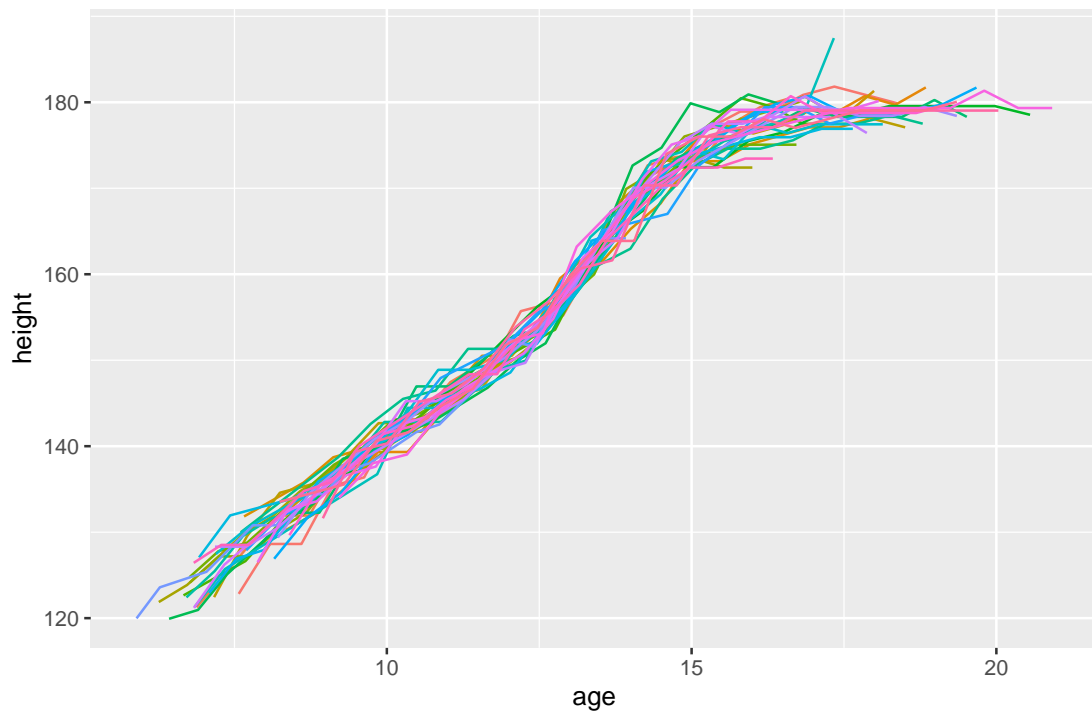

The growth curve for each boy

(a) Random Effect by a


```
## SITAR nonlinear mixed-effects model fit by maximum likelihood
##   Call: sitar(x = log(age), y = log(height), id = ID, data = data_f_boys,
##     df = bestDf3[1], random = "a")
##        AIC        BIC    logLik
##   -3424.986 -3393.311 1719.493
##
## Random effects:
##  Formula: a ~ 1 | id
##                 a   Residual
## StdDev: 0.04177228 0.01741566
##
## Fixed effects: s1 + s2 + s3 + s4 + a ~ 1
##       Value    Std.Error  DF  t-value p-value
## s1 0.242042 0.004643054 647  52.1299       0
## s2 0.371876 0.004327533 647  85.9324       0
## s3 0.509318 0.011623187 647  43.8192       0
## s4 0.312884 0.005252838 647  59.5648       0
## a  4.786553 0.009022328 647 530.5231       0
##  Correlation:
##    s1     s2     s3     s4
## s2  0.479
## s3  0.789  0.682
## s4  0.327 -0.372  0.175
## a  -0.468 -0.419 -0.532 -0.040
##
## Standardized Within-Group Residuals:
##          Min            Q1           Med            Q3           Max
## -2.9596930115 -0.6373855678 -0.0001611906  0.6417735461  2.8967537443
##
## Number of Observations: 682
## Number of Groups: 31
##
##                     Type variance   residual        AIC        BIC
## mba Random Effect by a     85.27 0.01741566 -3424.986 -3393.311
```
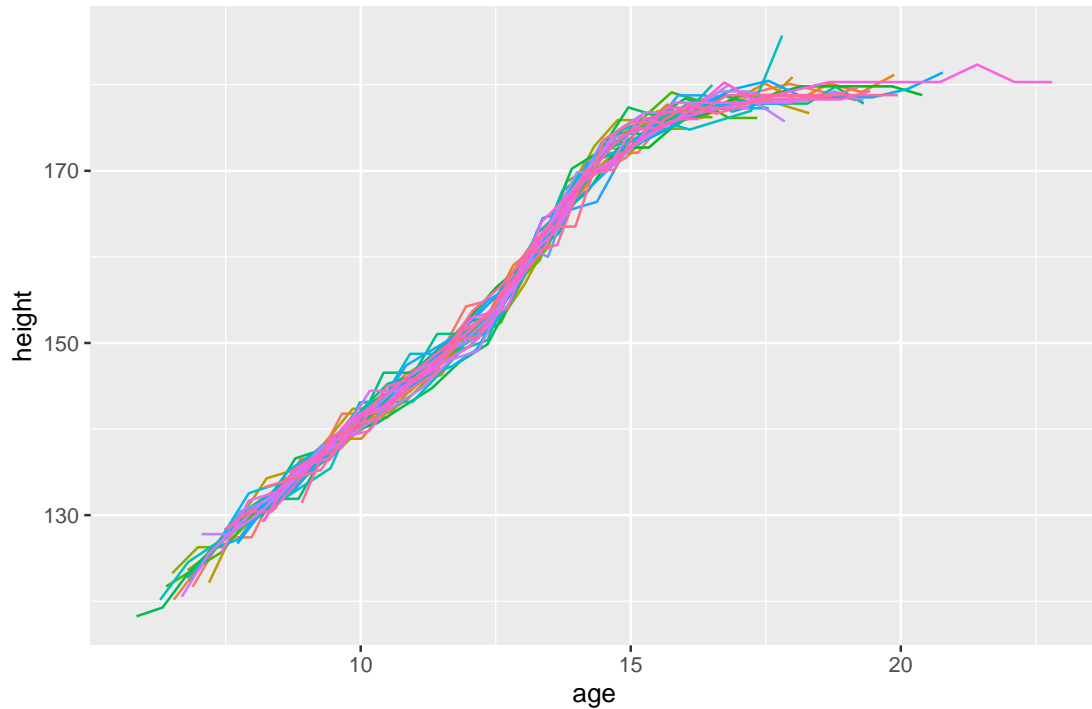
(b) Random Effect by a + b


```
## SITAR nonlinear mixed-effects model fit by maximum likelihood
##   Call: sitar(x = log(age), y = log(height), id = ID, data = data_f_boys,
##     df = bestDf3[1], random = "a+b")
##        AIC       BIC   logLik
##   -3845.512 -3800.262 1932.756
##
## Random effects:
##  Formula: list(a ~ 1, b ~ 1)
##  Level: id
##  Structure: General positive-definite, Log-Cholesky parametrization
##         StdDev     Corr
## a       0.03674339 a
## b       0.07995842 0.216
## Residual 0.01158098
##
## Fixed effects: s1 + s2 + s3 + s4 + a + b ~ 1
##        Value    Std.Error DF  t-value  p-value
## s1  0.224951 0.003294340 646  68.2840   0.000
## s2  0.376430 0.002661578 646 141.4313   0.000
## s3  0.522630 0.009288782 646  56.2647   0.000
## s4  0.332925 0.004197426 646  79.3166   0.000
## a   4.781255 0.008049173 646 594.0058   0.000
## b  -0.024822 0.015729438 646  -1.5781   0.115
##  Correlation:
##    s1     s2     s3     s4     a
## s2 0.564
## s3 0.393  0.474
## s4 0.014 -0.194  0.593
## a -0.179 -0.220 -0.549 -0.375
## b  0.139  0.085 -0.230 -0.301  0.324
##
## Standardized Within-Group Residuals:
##          Min          Q1         Med          Q3         Max
```

```
## -3.607552439 -0.589380617  0.009570318  0.603159768  3.875430032
##
## Number of Observations: 682
## Number of Groups: 31

##                         Type variance   residual      AIC       BIC
## mbab Random Effect by a + b    93.48 0.01158098 -3845.512 -3800.262
```

(c) Random Effect by a + b + c


```
## SITAR nonlinear mixed-effects model fit by maximum likelihood
##   Call: sitar(x = log(age), y = log(height), id = ID, data = data_f_boys,
##     df = bestDf3[1])
##        AIC       BIC    logLik
##   -4131.781 -4068.43 2079.89
##
## Random effects:
##  Formula: list(a ~ 1, b ~ 1, c ~ 1)
##  Level: id
##  Structure: General positive-definite, Log-Cholesky parametrization
##          StdDev      Corr
## a        0.032928723 a       b
## b        0.069095211  0.006
## c        0.095246807 -0.133 -0.178
## Residual 0.008600255
##
## Fixed effects: s1 + s2 + s3 + s4 + a + b + c ~ 1
##        Value Std.Error  DF  t-value p-value
## s1 0.151213 0.00673157 645  22.4632  0.0000
## s2 0.263661 0.01101298 645  23.9409  0.0000
## s3 0.369560 0.01736324 645  21.2841  0.0000
## s4 0.253757 0.00938032 645  27.0521  0.0000
## a  4.880337 0.01271720 645 383.7586  0.0000
## b  0.022506 0.01483057 645   1.5176  0.1296
## c  0.471799 0.06196562 645   7.6139  0.0000
##  Correlation:
```

```
##     s1     s2     s3     s4      a      b
## s2  0.947
## s3  0.929  0.991
## s4  0.937  0.990  0.990
## a  -0.813 -0.871 -0.880 -0.877
## b  -0.406 -0.477 -0.495 -0.495  0.453
## c  -0.906 -0.954 -0.955 -0.955  0.826  0.429
##
## Standardized Within-Group Residuals:
##          Min           Q1          Med           Q3          Max
## -3.305702664 -0.551354419  0.002992165  0.570881310  4.643728253
##
## Number of Observations: 682
## Number of Groups: 31

##                                   Type variance    residual       AIC       BIC
## fitHB3 Random Effect by a + b + c    96.41 0.008600255 -4131.781 -4068.43

##                          Type variance    residual       AIC       BIC
## 1         Random Effect by a    85.27  0.01741566 -3424.986 -3393.311
## 2     Random Effect by a + b    93.48  0.01158098 -3845.512 -3800.262
## 3 Random Effect by a + b + c    96.41 0.008600255 -4131.781  -4068.43
```
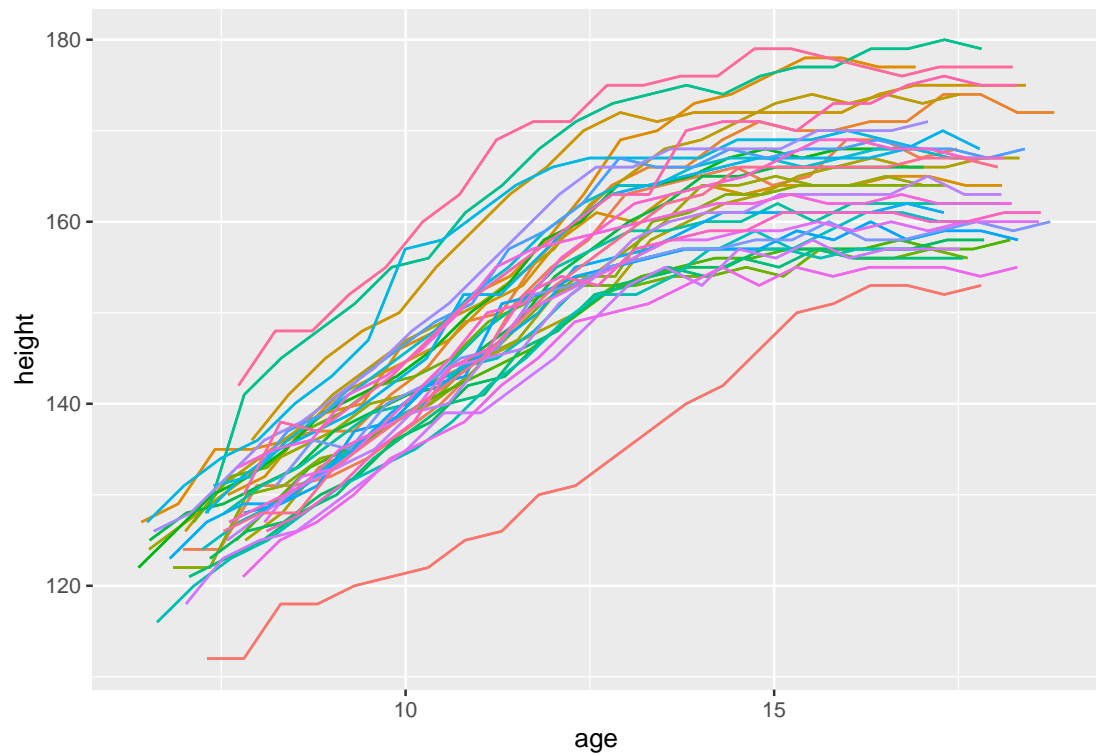
The model with all random effects is best.

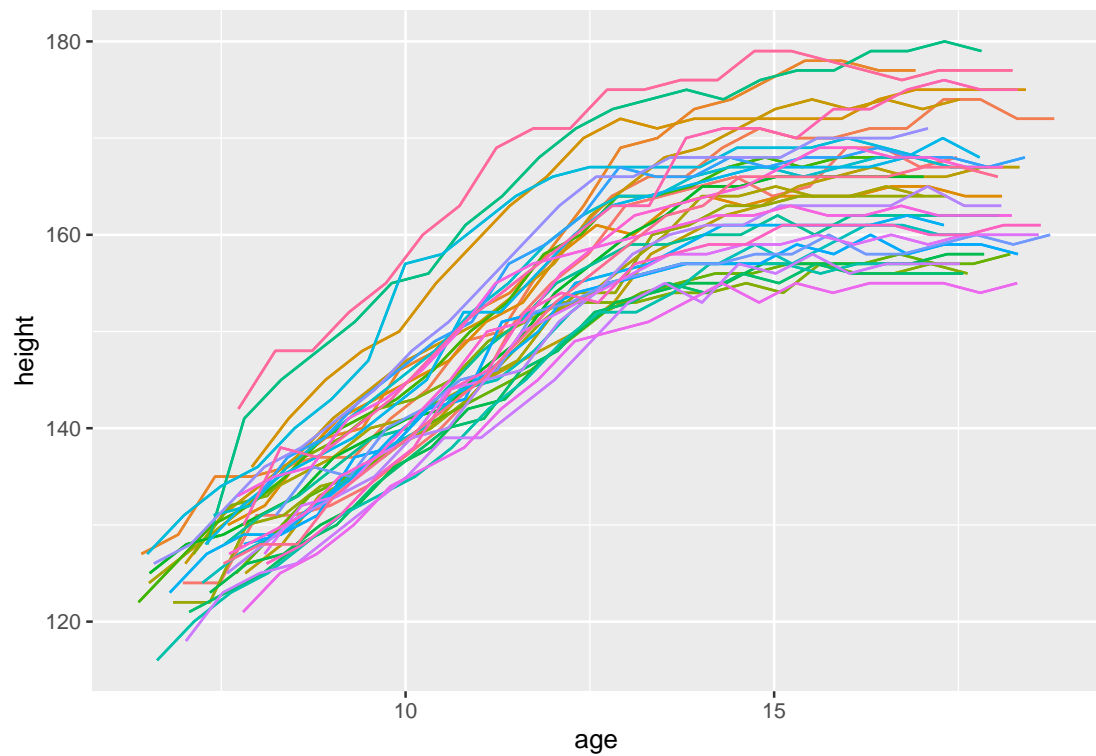### (d) Growth Curve with Random Effect by a + b + b
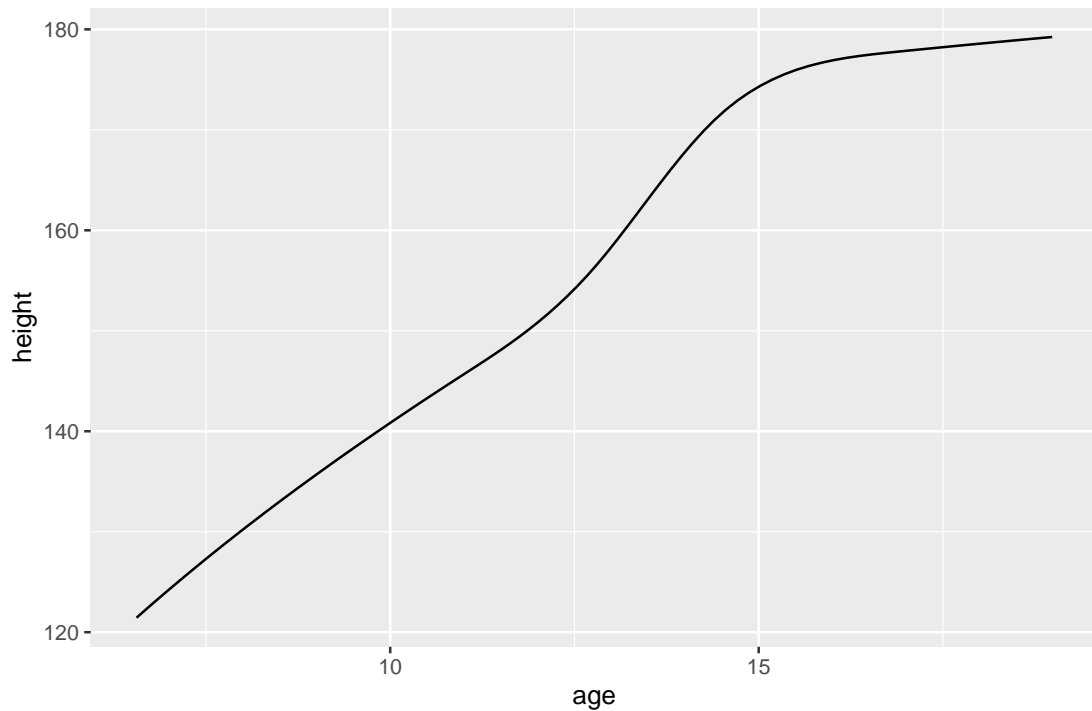
8 The second dataset (Girls)

The 38 girls who have the complete set of 22 observations.

The record(ID = 140) is an outlier.

## 8.1 Fit the data to find the best model

model 1: height~age

model 2: log(height)~age

model 3: height~log(age)

model 4: log(height)~log(age)

8.2 Find the the degrees of freedom (By BIC)

The degrees of freedom (DF) indicate the number of independent values that can vary in an analysis without breaking any constraints.

height~age (girl)

log(height)~age (girl)

height~log(age) (girl)

log(height)~log(age) (girl)
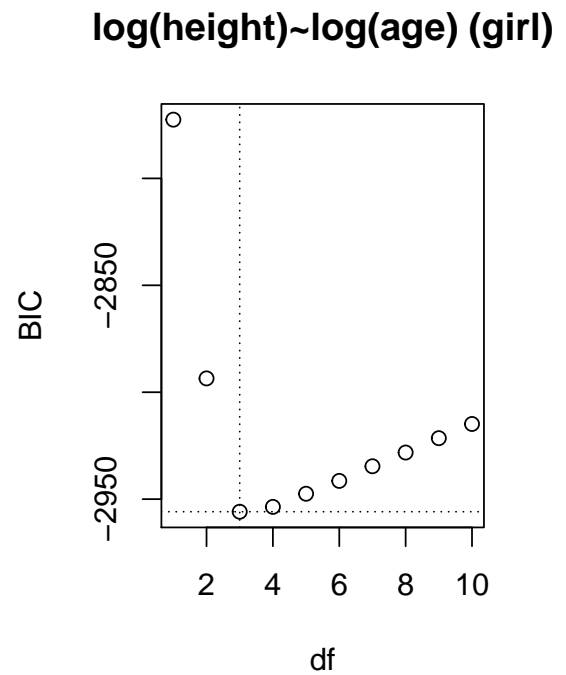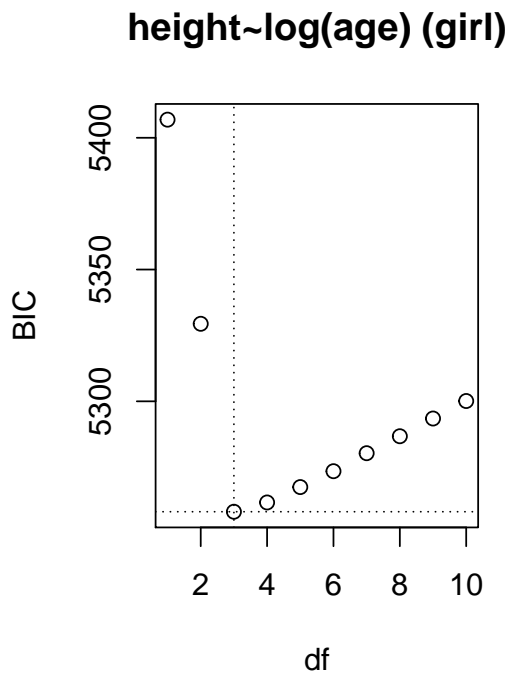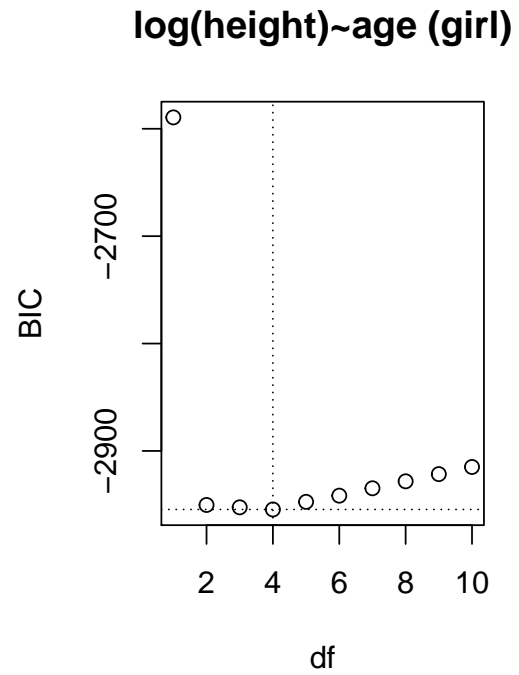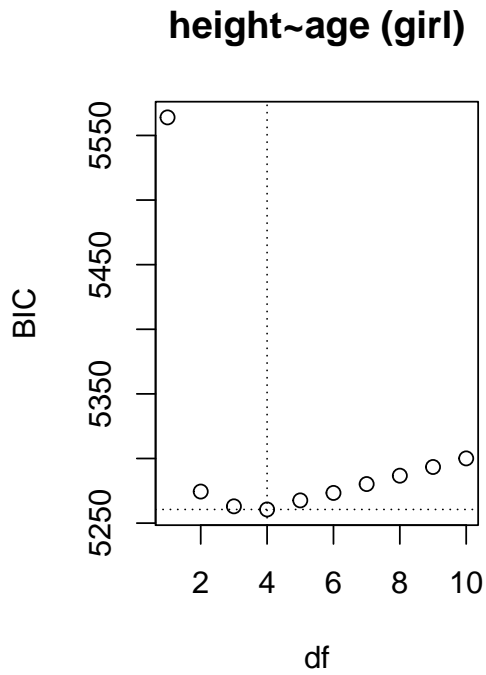
When Df = 3, the BIC is the lowest, choose the df = 3 for fitting model (height~log(age), log(height)~log(age)). When Df = 4, the BIC is the lowest, choose the df = 4 for fitting model (height~age, log(height)~age).

8.3 Fit the models by using sitar model

Fit the four models and compare the variance, residual and AIC to find the best model.

```
## Warning in nlme.formula(y ~ fitnlme(x, s1, s2, s3, s4, a, b, c), fixed = s1 + :
## Iteration 1, LME step: nlminb() did not converge (code = 1). PORT message: false
## convergence (8)
```

```
##                   formula variance    residual        AIC Converge
## 1            height~age      96.68   1.0929107   2930.909     True
## 2       log(height)~age      96.63 0.007419857  -5201.022    False
## 3       height~log(age)      96.30  1.10268133   2942.887     True
## 4 log(height)~log(age)      96.29 0.007452629   -5194.83     True
```
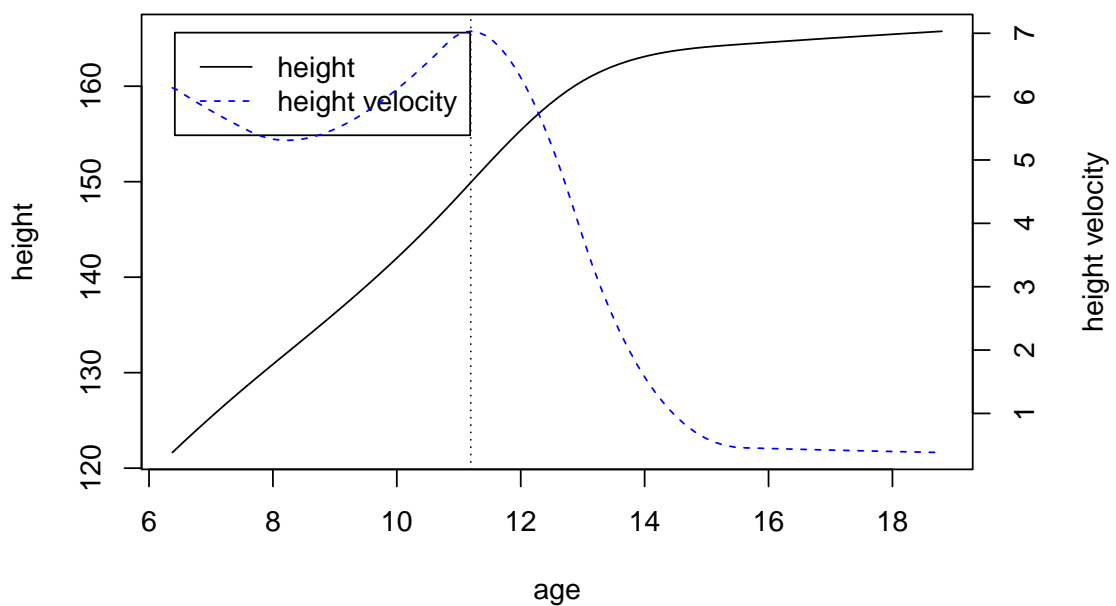
Although, the log(height)~age model is the best, however, it has convergence error. So I chose the log(height)~log(age) model.

8.4 The distance and velocity curves

Distance curve(left y axis)(Black solid line): the amount of height achieved at a given age.

Velocity curve(right y axis)(Blue dash line): the rate of growth at a given age.


```
##    apv     pv
## 11.190  7.031
```

1. The growth rate:

   The girls grow fast between 6~14 year-old, grow slowly after 14 year-old.

2. The growth velocity:

   a) It reached the peak of 7.031 at the age of 11.19.
   b) It declined between 6~8 year-old.

c) It increased between 8~11.19 year-old.
d) It declined rapidly between 11.19~15 year-old.
e) It is relatively constant after 15 year-old.

8.5 Check the significant of fixed and random effect factors

```
##      numDF denDF  F-value p-value
## s1      1   772   8177.9  <.0001
## s2      1   772     13.3   3e-04
## s3      1   772   4233.0  <.0001
## a       1   772 805085.5  <.0001
## b       1   772     75.5  <.0001
## c       1   772     86.7  <.0001
```
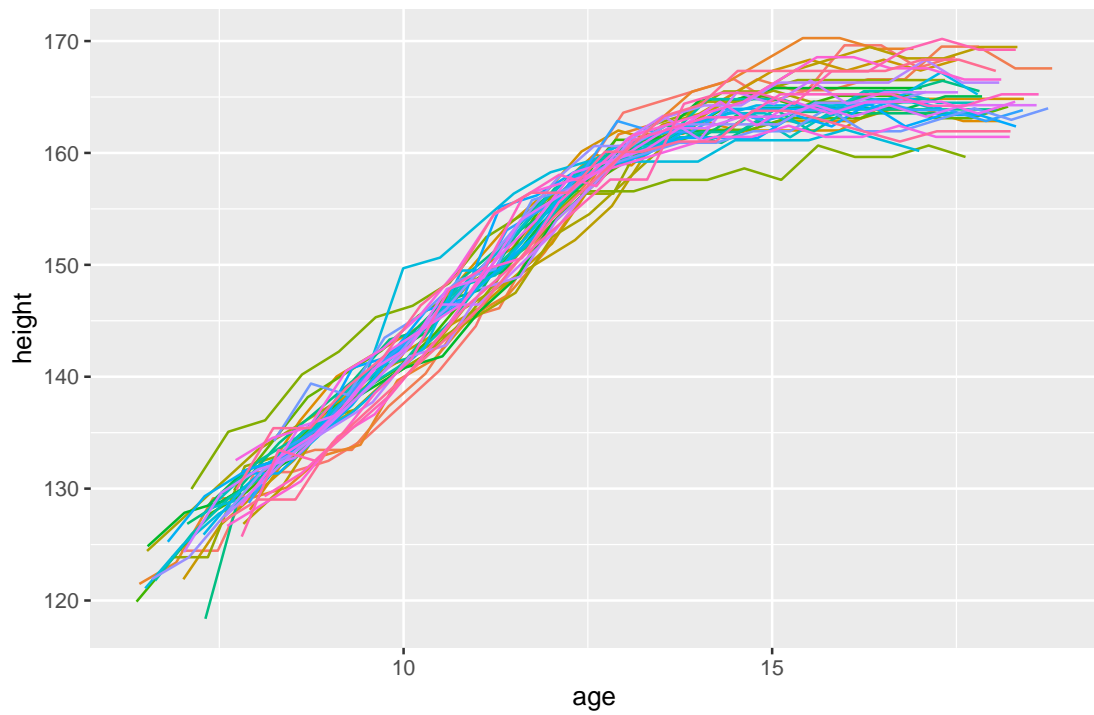
The fixed effect s1, s2, s3, s4 and the random effect of α(a), β(b), γ(c) are singificant.(p-value < 0.05).

8.6 Compare the random effect $(\alpha(\mathbf{a}) + \beta(\mathbf{b}) + \gamma(\mathbf{c}))$

Increase each random effect SITAR parameter in model, figure out the parameters how to effect the model.

The growth curve for each girl
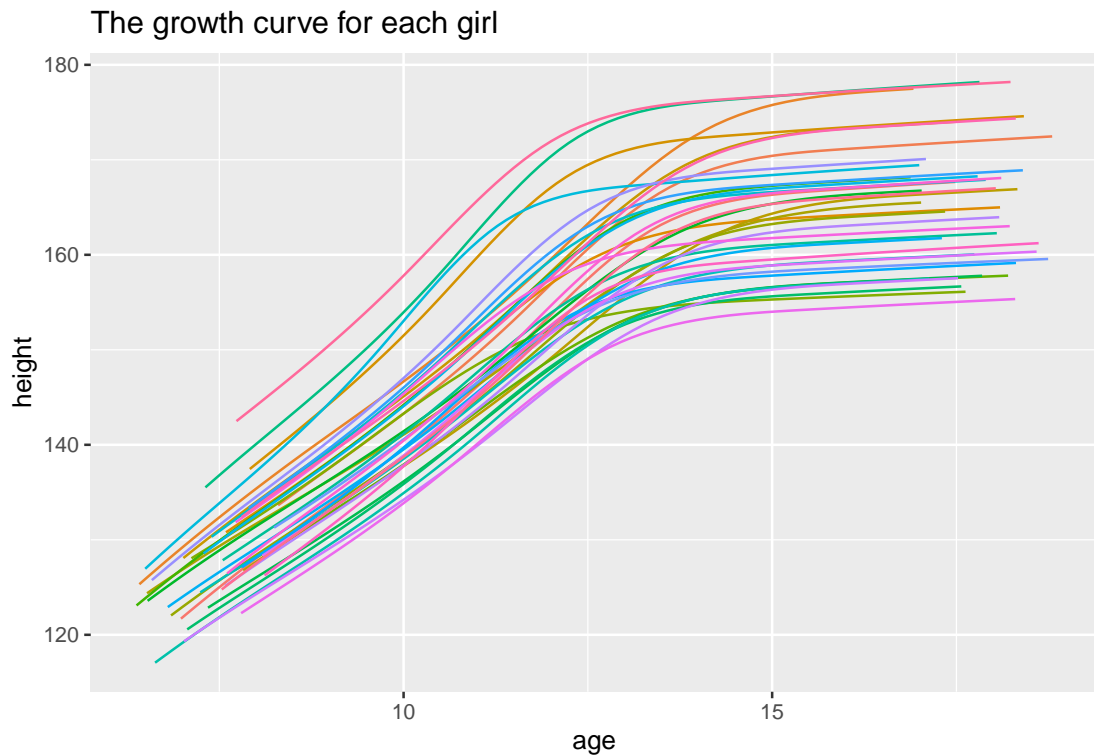
(a) Random Effect by a
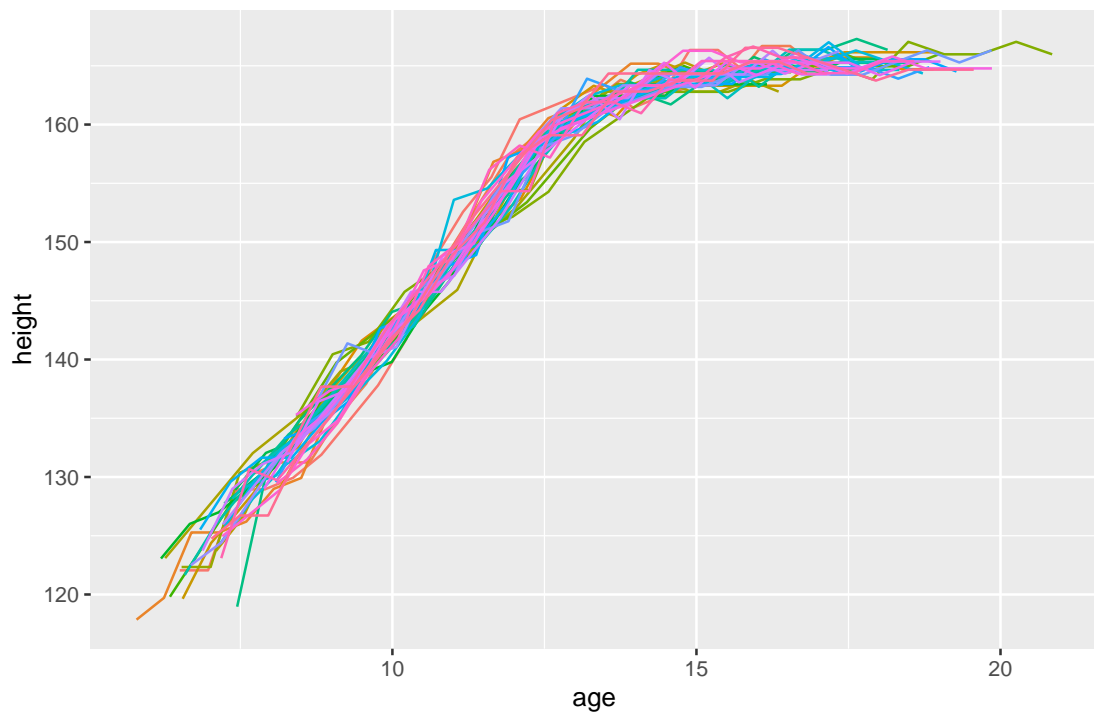
```
##                       Type variance   residual       AIC       BIC
## mga Random Effect by a      86.45 0.01423831 -4416.925 -4388.713
```
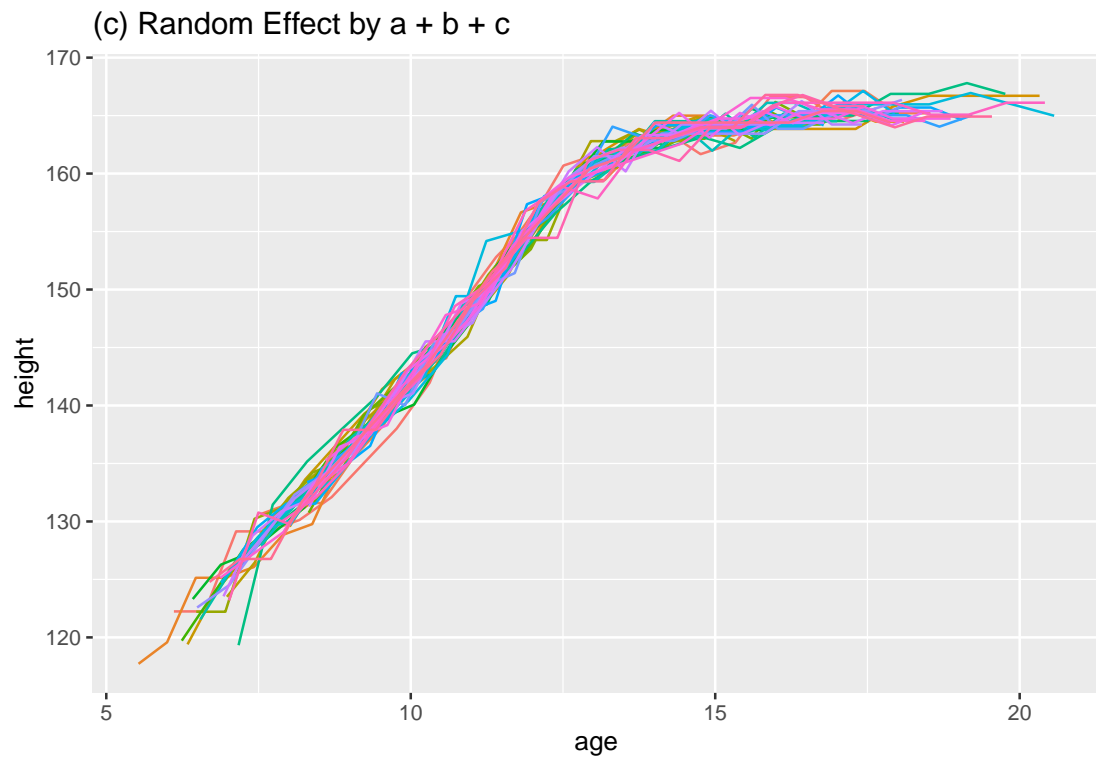

(b) Random Effect by a + b

```
##                           Type variance  residual      AIC      BIC
## mgab Random Effect by a + b      94.27 1.3751748 3195.192 3242.212
```

(c) Random Effect by a + b + c


```
##                          Type variance    residual       AIC       BIC
## fitHG4 Random Effect by a + b + c    96.29 0.007452629 -5194.83 -5133.704
```

```
##                          Type variance    residual       AIC       BIC
## 1        Random Effect by a    86.45  0.01423831 -4416.925 -4388.713
## 2    Random Effect by a + b    94.27   1.3751748  3195.192  3242.212
## 3 Random Effect by a + b + c    96.29 0.007452629  -5194.83 -5133.704
```

The model with all random effects is best.

(d) Growth Curve with Random Effect by a + b + c
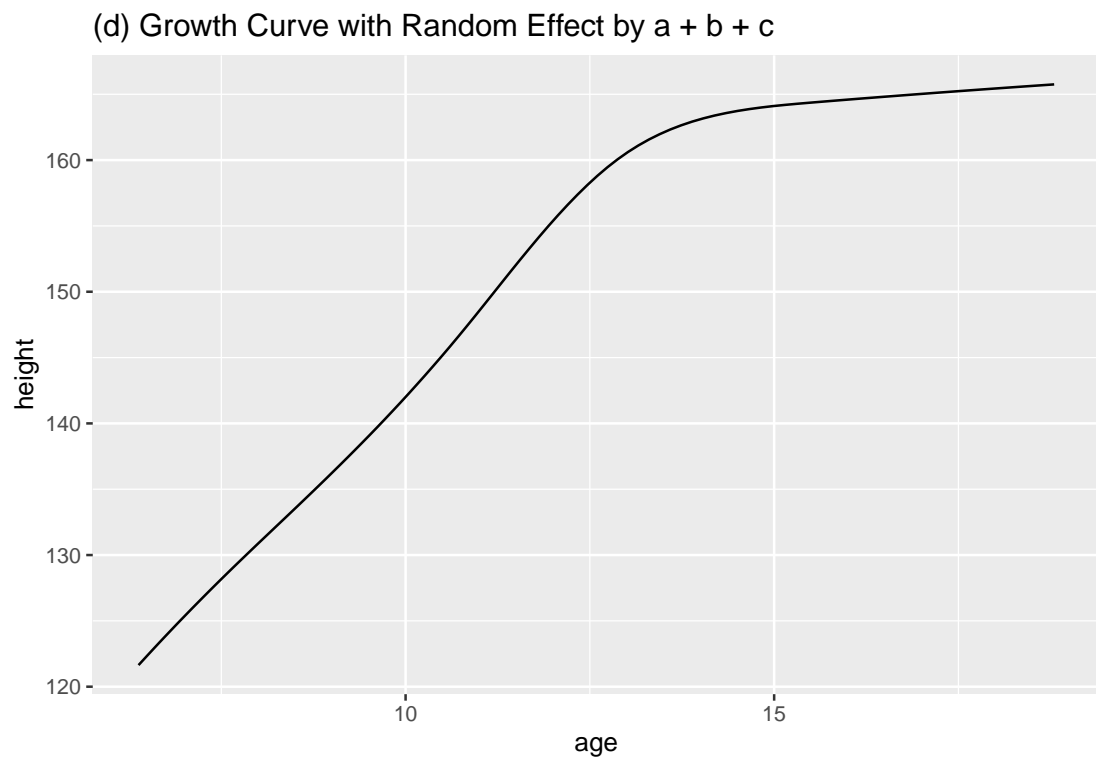
9 Compare the growth rate and velocity of height by gender

The Growth Curve (Boy:blue, Girl:red)
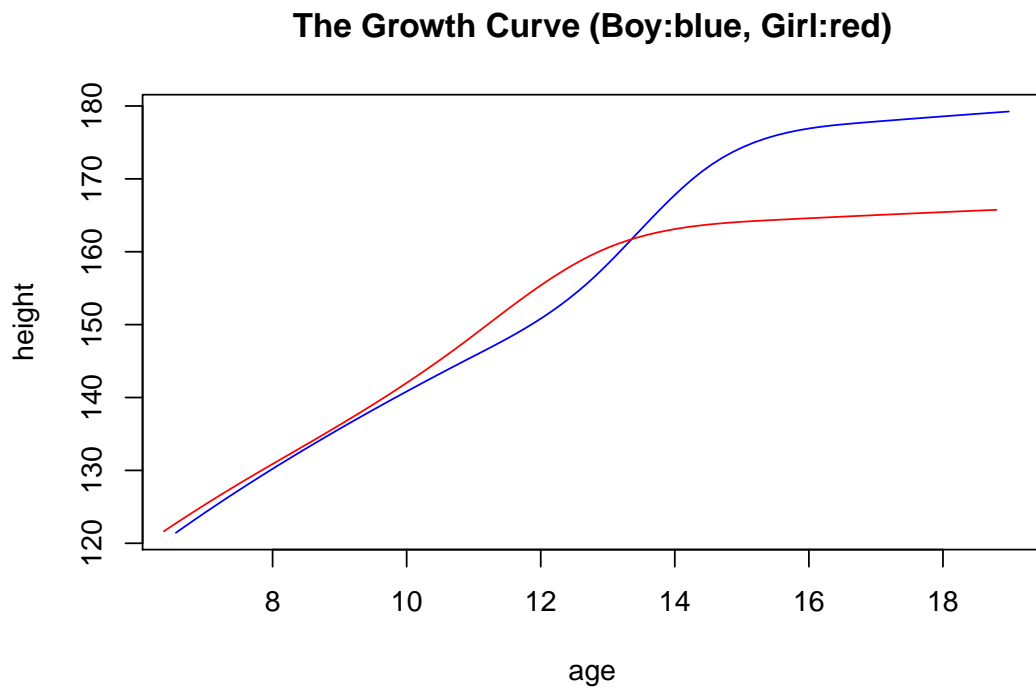
1. 6-10 year-old: Boys and Girls almost have the same rate.

2. 10~13.5 year-old: Girls are a little higer than boys.

3. 13.6~18 year-old: Boys are significant higher than Girls.

The Growth Curve : velocity (Boy:blue, Girl:red)
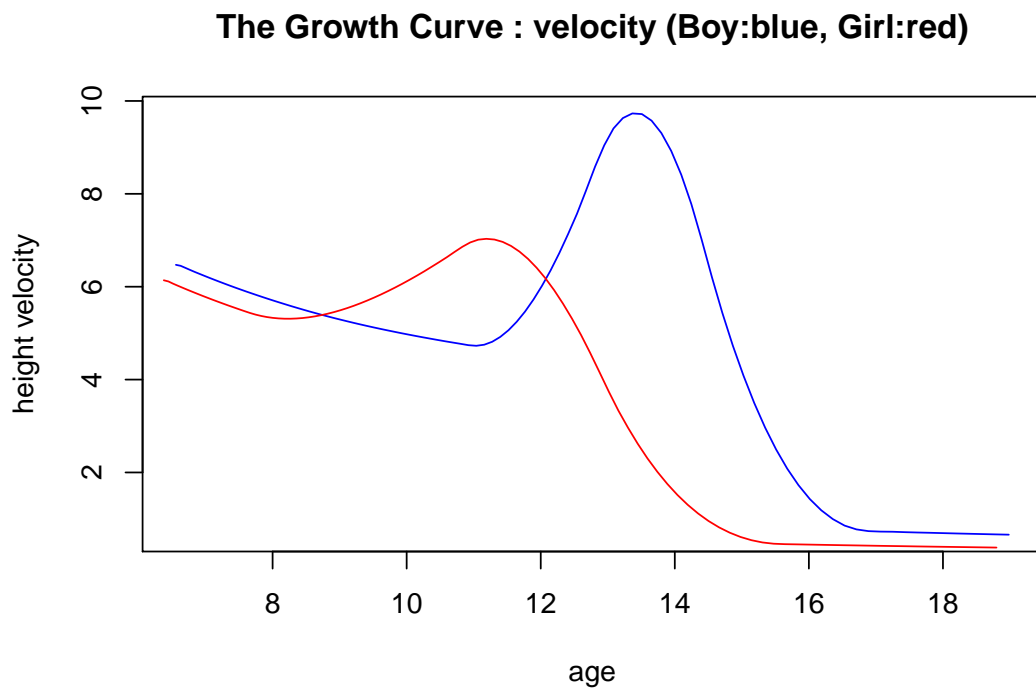
1. 6~9 year-old: Boys and Girls are near to each other and deline.

2. 9~11 year-old: Boys and girls have different trend, Boys continue delining, Girls increase and reach the top at 11 year-old.

3. 11~15.5 year-old: Girls keep declining fast, Boys increase fast and reach the top at 13.5, then declined fast til 16.5 year-old.

4. The girls keep constant from 15.5 year-old, the boys keep constant from 17 year-old and near to each other.

10   Conclusion

The fact that the model fits well both in boys and girls. Although, the log height scale versus age is the best model, however, the convergence error happened when fit the model, so chose the log height scale versus log age scale as the model. Compared the growth curve, we can see that the difference of growth rate and velocity by gender at a certain age.