

statistical-modeling-r-examples.R

julian

Tue Jun 2 16:11:08 2015

```
#!/usr/bin/env Rscript

# implement statistical modeling examples in R

target.dir <- '~/GitHub/reproducible-research/Day-3/datasets'
target.file <- 'statistical-modeling-r-examples.txt'
sink(file = file.path(target.dir, target.file))

# published dataset -----

data.dir <- '~/GitHub/reproducible-research/Day-2/datasets'
data.file <- 'published-data-complete.csv'

# use base R to import
child.study.data <- read.csv(file.path(data.dir, data.file), header = TRUE)

# remove model fit values already included in the file
child.study.data$Fitted <- NULL
child.study.data$Resid <- NULL
child.study.data$ScaledResid <- NULL
child.study.data$Predicted <- NULL

head(child.study.data)
```

```
##      Subject Site Exclusions      DOS      DOB Age Age_Calc AgeGroup Gender
## 1 3002-102 UCSF          no 11/4/11 1/24/01  10  10.784   child  male
## 2 3002-102 UCSF          no 11/4/11 1/24/01  10  10.784   child  male
## 3 3002-102 UCSF          no 11/4/11 1/24/01  10  10.784   child  male
## 4 3002-102 UCSF          no 11/4/11 1/24/01  10  10.784   child  male
## 5 3002-102 UCSF          no 11/4/11 1/24/01  10  10.784   child  male
## 6 3002-102 UCSF          no 11/4/11 1/24/01  10  10.784   child  male
##   Handedness      Dx Chromosome      Case      Copies      ASD NVIQ VIQ
## 1      right proband      16p duplication duplication FALSE   77  86
## 2      right proband      16p duplication duplication FALSE   77  86
## 3      right proband      16p duplication duplication FALSE   77  86
## 4      right proband      16p duplication duplication FALSE   77  86
## 5      right proband      16p duplication duplication FALSE   77  86
## 6      right proband      16p duplication duplication FALSE   77  86
##   CELF.4 SRS_parent SRS_adult CTOPP      ICV2      cmICV Hem      Cond M50Lat
## 1      75      145      NA      6 1422314 1422.314 1-LH 1-200      120
## 2      75      145      NA      6 1422314 1422.314 1-LH 2-300      NA
## 3      75      145      NA      6 1422314 1422.314 1-LH 3-500      NA
## 4      75      145      NA      6 1422314 1422.314 1-LH 4-1000      NA
## 5      75      145      NA      6 1422314 1422.314 2-RH 1-200      NA
## 6      75      145      NA      6 1422314 1422.314 2-RH 2-300      NA
##   M100Lat M50LatCorr M100LatCorr dB.SL M100compCase M50compCase      cutAge
```

```
## 1      NA      92      NA      25      0      1 under-12
## 2      NA      NA      NA      25      0      0 under-12
## 3      NA      NA      NA      25      0      0 under-12
## 4      NA      NA      NA      25      0      0 under-12
## 5     162      NA     134      25      1      0 under-12
## 6     148      NA     120      25      1      0 under-12
## breakAge
## 1   (9,10]
## 2   (9,10]
## 3   (9,10]
## 4   (9,10]
## 5   (9,10]
## 6   (9,10]
```

```
str(child.study.data)
```

```
## 'data.frame': 792 obs. of 34 variables:
## $ Subject : Factor w/ 99 levels "3002-102","3003-101",...: 1 1 1 1 1 1 1 1 2 2 ...
## $ Site : Factor w/ 2 levels "CHOP","UCSF": 2 2 2 2 2 2 2 2 2 2 ...
## $ Exclusions : Factor w/ 1 level "no": 1 1 1 1 1 1 1 1 1 1 ...
## $ DOS : Factor w/ 87 levels "1/10/12","1/17/13",...: 20 20 20 20 20 20 20 20 17 17 ...
## $ DOB : Factor w/ 95 levels "1/15/01","1/16/04",...: 6 6 6 6 6 6 6 6 51 51 ...
## $ Age : int 10 10 10 10 10 10 10 10 12 12 ...
## $ Age_Calc : num 10.8 10.8 10.8 10.8 10.8 ...
## $ AgeGroup : Factor w/ 1 level "child": 1 1 1 1 1 1 1 1 1 1 ...
## $ Gender : Factor w/ 2 levels "female","male": 2 2 2 2 2 2 2 2 2 2 ...
## $ Handedness : Factor w/ 3 levels "ambidextrous",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ Dx : Factor w/ 2 levels "control","proband": 2 2 2 2 2 2 2 2 2 2 ...
## $ Chromosome : Factor w/ 2 levels "16p","control": 1 1 1 1 1 1 1 1 1 1 ...
## $ Case : Factor w/ 3 levels "control","deletion",...: 3 3 3 3 3 3 3 3 2 2 ...
## $ Copies : Factor w/ 3 levels "control","deletion",...: 3 3 3 3 3 3 3 3 2 2 ...
## $ ASD : Factor w/ 3 levels "control","FALSE",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ NVIQ : int 77 77 77 77 77 77 77 77 102 102 ...
## $ VIQ : int 86 86 86 86 86 86 86 86 106 106 ...
## $ CELF.4 : int 75 75 75 75 75 75 75 75 88 88 ...
## $ SRS_parent : int 145 145 145 145 145 145 145 145 6 6 ...
## $ SRS_adult : logi NA NA NA NA NA NA ...
## $ CTOPP : int 6 6 6 6 6 6 6 6 3 3 ...
## $ ICV2 : num 1422314 1422314 1422314 1422314 1422314 ...
## $ cmICV : num 1422 1422 1422 1422 1422 ...
## $ Hem : Factor w/ 2 levels "1-LH","2-RH": 1 1 1 1 2 2 2 2 1 1 ...
## $ Cond : Factor w/ 4 levels "1-200","2-300",...: 1 2 3 4 1 2 3 4 1 2 ...
## $ M50Lat : int 120 NA NA NA NA NA NA 116 NA 86 ...
## $ M100Lat : int NA NA NA NA 162 148 146 166 180 176 ...
## $ M50LatCorr : int 92 NA NA NA NA NA NA 88 NA 58 ...
## $ M100LatCorr : int NA NA NA NA 134 120 118 138 152 148 ...
## $ dB.SL : num 25 25 25 25 25 25 25 25 24 24 ...
## $ M100compCase : int 0 0 0 0 1 1 1 1 1 1 ...
## $ M50compCase : int 1 0 0 0 0 0 0 1 0 1 ...
## $ cutAge : Factor w/ 2 levels "12-and-over",...: 2 2 2 2 2 2 2 2 1 1 ...
## $ breakAge : Factor w/ 10 levels "(10,11]","(11,12]",...: 9 9 9 9 9 9 9 9 2 2 ...
```

```
summary(child.study.data)
```

```
##      Subject      Site Exclusions      DOS      DOB
## 3002-102: 8 CHOP:376 no:792 2/23/12 : 24 3/23/02: 16
## 3003-101: 8 UCSF:416 10/18/13: 16 4/26/98: 16
## 3003-102: 8 11/18/11: 16 6/5/06 : 16
## 3005-101: 8 11/19/12: 16 8/22/03: 16
## 3011-101: 8 11/8/13 : 16 1/15/01: 8
## 3014-101: 8 2/18/13 : 16 1/16/04: 8
## (Other) :744 (Other) :688 (Other):712
##      Age      Age_Calc      AgeGroup      Gender
## Min. : 7.00 Min. : 7.307 child:792 female:320
## 1st Qu.:10.00 1st Qu.: 9.953 male :472
## Median :11.00 Median :11.677
## Mean :11.49 Mean :11.958
## 3rd Qu.:14.00 3rd Qu.:14.216
## Max. :17.00 Max. :17.151
##
##      Handedness      Dx      Chromosome      Case
## ambidextrous:120 control:384 16p :408 control :384
## left :112 proband:408 control:384 deletion :280
## right :560 duplication:128
##
##
##
##
##      Copies      ASD      NVIQ      VIQ
## control :384 control:384 Min. : 53.00 Min. : 54.00
## deletion :280 FALSE :304 1st Qu.: 87.00 1st Qu.: 87.00
## duplication:128 TRUE :104 Median : 96.50 Median : 98.00
## Mean : 95.89 Mean : 96.38
## 3rd Qu.:105.00 3rd Qu.:107.00
## Max. :136.00 Max. :147.00
## NA's :40 NA's :40
##      CELF.4      SRS_parent      SRS_adult      CTOPP
## Min. : 40.00 Min. : 0.0 Mode:logical Min. : 1.000
## 1st Qu.: 79.00 1st Qu.: 12.0 NA's:792 1st Qu.: 6.000
## Median : 97.00 Median : 31.5 Median : 8.000
## Mean : 91.91 Mean : 46.9 Mean : 7.629
## 3rd Qu.:108.00 3rd Qu.: 77.0 3rd Qu.: 9.000
## Max. :132.00 Max. :145.0 Max. :13.000
## NA's :96 NA's :40 NA's :80
##      ICV2      cmICV      Hem      Cond      M50Lat
## Min. :1148733 Min. :1149 1-LH:396 1-200 :198 Min. : 52.00
## 1st Qu.:1384190 1st Qu.:1384 2-RH:396 2-300 :198 1st Qu.: 82.00
## Median :1528991 Median :1529 3-500 :198 Median : 91.00
## Mean :1522435 Mean :1522 4-1000:198 Mean : 94.88
## 3rd Qu.:1651057 3rd Qu.:1651 3rd Qu.:106.00
## Max. :1934972 Max. :1935 Max. :188.00
## NA's :128 NA's :128 NA's :370
##      M100Lat      M50LatCorr      M100LatCorr      dB.SL
## Min. :104.0 Min. : 31.00 Min. : 76.0 Min. : 0.00
## 1st Qu.:132.0 1st Qu.: 57.00 1st Qu.:107.0 1st Qu.:12.00
```

```
## Median :150.0 Median : 66.00 Median :125.5 Median :20.00
## Mean :157.6 Mean : 70.68 Mean :133.0 Mean :20.26
## 3rd Qu.:184.0 3rd Qu.: 82.00 3rd Qu.:159.0 3rd Qu.:28.00
## Max. :220.0 Max. :167.00 Max. :197.0 Max. :59.90
## NA's :258 NA's :370 NA's :258 NA's :16
## M100compCase M50compCase cutAge breakAge
## Min. :0.0000 Min. :0.0000 12-and-over:368 (9,10] :128
## 1st Qu.:0.0000 1st Qu.:0.0000 under-12 :424 [7,8] :120
## Median :1.0000 Median :1.0000 (11,12]:112
## Mean :0.6742 Mean :0.5328 (10,11]:104
## 3rd Qu.:1.0000 3rd Qu.:1.0000 (14,15]: 72
## Max. :1.0000 Max. :1.0000 (8,9] : 72
## (Other):184
```

```
dplyr::glimpse(child.study.data)
```

```
## Observations: 792
## Variables:
## $ Subject (fctr) 3002-102, 3002-102, 3002-102, 3002-102, 3002-102...
## $ Site (fctr) UCSF, UCSF, UCSF, UCSF, UCSF, UCSF, UCSF, UCSF, ...
## $ Exclusions (fctr) no, no, no, no, no, no, no, no, no, no, no, no, ...
## $ DOS (fctr) 11/4/11, 11/4/11, 11/4/11, 11/4/11, 11/4/11, 11/4/11, 11/4/11, 11/4/11, ...
## $ DOB (fctr) 1/24/01, 1/24/01, 1/24/01, 1/24/01, 1/24/01, 1/24/01, 1/24/01, 1/24/01, ...
## $ Age (int) 10, 10, 10, 10, 10, 10, 10, 10, 10, 12, 12, 12, 12, 1...
## $ Age_Calc (dbl) 10.784, 10.784, 10.784, 10.784, 10.784, 10.784, 10.784, 10.784, 1...
## $ AgeGroup (fctr) child, child, child, child, child, child, child, child, ...
## $ Gender (fctr) male, male, male, male, male, male, male, male, ...
## $ Handedness (fctr) right, right, right, right, right, right, right, right, ...
## $ Dx (fctr) proband, proband, proband, proband, proband, proband, proband, proband, ...
## $ Chromosome (fctr) 16p, 16p, 16p, 16p, 16p, 16p, 16p, 16p, 16p, 16p, 16p, 16p, ...
## $ Case (fctr) duplication, duplication, duplication, duplication, duplication, duplication, duplication, duplication, ...
## $ Copies (fctr) duplication, duplication, duplication, duplication, duplication, duplication, duplication, duplication, ...
## $ ASD (fctr) FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, ...
## $ NVIQ (int) 77, 77, 77, 77, 77, 77, 77, 77, 77, 102, 102, 102, 102, 10...
## $ VIQ (int) 86, 86, 86, 86, 86, 86, 86, 86, 86, 106, 106, 106, 106, 10...
## $ CELF.4 (int) 75, 75, 75, 75, 75, 75, 75, 75, 75, 88, 88, 88, 88, 8...
## $ SRS_parent (int) 145, 145, 145, 145, 145, 145, 145, 145, 145, 6, 6, 6, 6, 6...
## $ SRS_adult (lgl) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ CTOPP (int) 6, 6, 6, 6, 6, 6, 6, 6, 6, 3, 3, 3, 3, 3, 3, 3, 3, 3, 6...
## $ ICV2 (dbl) 1422314, 1422314, 1422314, 1422314, 1422314, 1422314, 1422314, 1422314, ...
## $ cmICV (dbl) 1422.314, 1422.314, 1422.314, 1422.314, 1422.314, 1422.314, 1422.314, 1422.314, ...
## $ Hem (fctr) 1-LH, 1-LH, 1-LH, 1-LH, 1-LH, 2-RH, 2-RH, 2-RH, 2-RH, ...
## $ Cond (fctr) 1-200, 2-300, 3-500, 4-1000, 1-200, 2-300, 3-500, 4-1000, ...
## $ M50Lat (int) 120, NA, NA, NA, NA, NA, NA, NA, NA, 116, NA, 86, NA, 86, ...
## $ M100Lat (int) NA, NA, NA, NA, NA, 162, 148, 146, 166, 180, 176, NA, NA, NA, ...
## $ M50LatCorr (int) 92, NA, NA, NA, NA, NA, NA, NA, NA, 88, NA, 58, NA, 58, 5...
## $ M100LatCorr (int) NA, NA, NA, NA, NA, 134, 120, 118, 138, 152, 148, NA, NA, NA, ...
## $ dB.SL (dbl) 25, 25, 25, 25, 25, 25, 25, 25, 25, 24, 24, 24, 24, 2...
## $ M100compCase (int) 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1...
## $ M50compCase (int) 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1...
## $ cutAge (fctr) under-12, under-12, under-12, under-12, under-12, under-12, under-12, under-12, ...
## $ breakAge (fctr) (9,10], (9,10], (9,10], (9,10], (9,10], (9,10], (9,10], (9,10], ...
```

```

# density estimate for each Case

library(ggplot2)

# color-blind friendly palette
cbPalette <-
  c("#999999", "#E69F00", "#56B4E9",
    "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7")

ggplot(child.study.data, aes(x = M100LatCorr)) +
  geom_density(aes(colour = Case, fill = Case, linetype = Case),
    size = 1.3, alpha = 0.3) +
  theme_bw() +
  labs(x = 'Latency (ms)', y = 'Density',
    title = 'Latency distribution for each Case') +
  scale_x_continuous(breaks = seq(90, 170, 20))+
  scale_y_continuous(breaks = seq(0, 0.1, 0.02))+
  theme(legend.position = 'bottom',
    legend.text = element_text(face = 'bold', size = 16),
    legend.title = element_blank(),
    axis.text.x = element_text(size = 15, face = 'bold'),
    axis.text.y = element_text(size = 15, face = 'bold'),
    axis.ticks = element_blank(),
    axis.title.x = element_text(size = 20, face = 'bold'),
    axis.title.y = element_text(size = 20, face = 'bold'),
    panel.grid.minor.x = element_blank(),
    panel.grid.major.x = element_blank(),
    panel.grid.minor.y = element_blank(),
    panel.grid.major.y = element_blank(),
    plot.title = element_text(size = 20, face = 'bold')) +
  scale_fill_manual(values = cbPalette) +
  scale_colour_manual(values = cbPalette) +
  scale_linetype_manual(values = c('solid', 'dashed', 'dotted'))

```

```

## Warning in loop_apply(n, do.ply): Removed 118 rows containing non-finite
## values (stat_density).

```

```

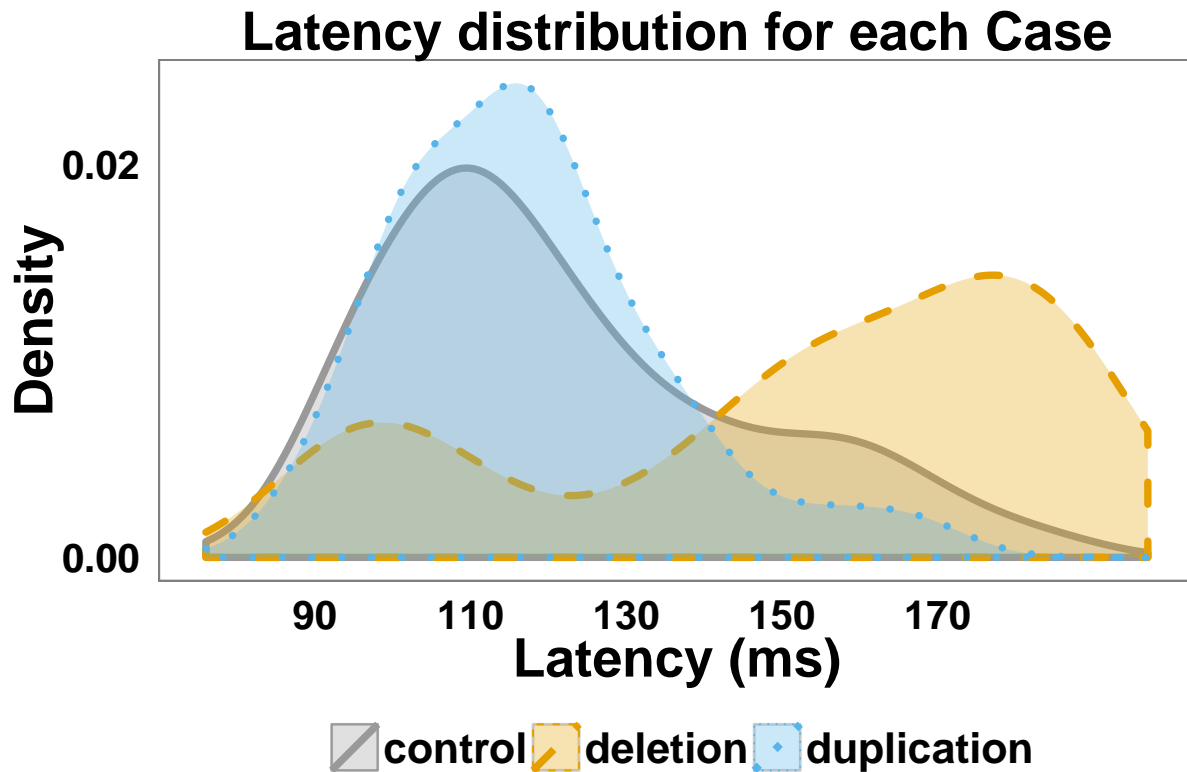
## Warning in loop_apply(n, do.ply): Removed 77 rows containing non-finite
## values (stat_density).

```

```

## Warning in loop_apply(n, do.ply): Removed 63 rows containing non-finite
## values (stat_density).

```



```
# plot latency against age for each case

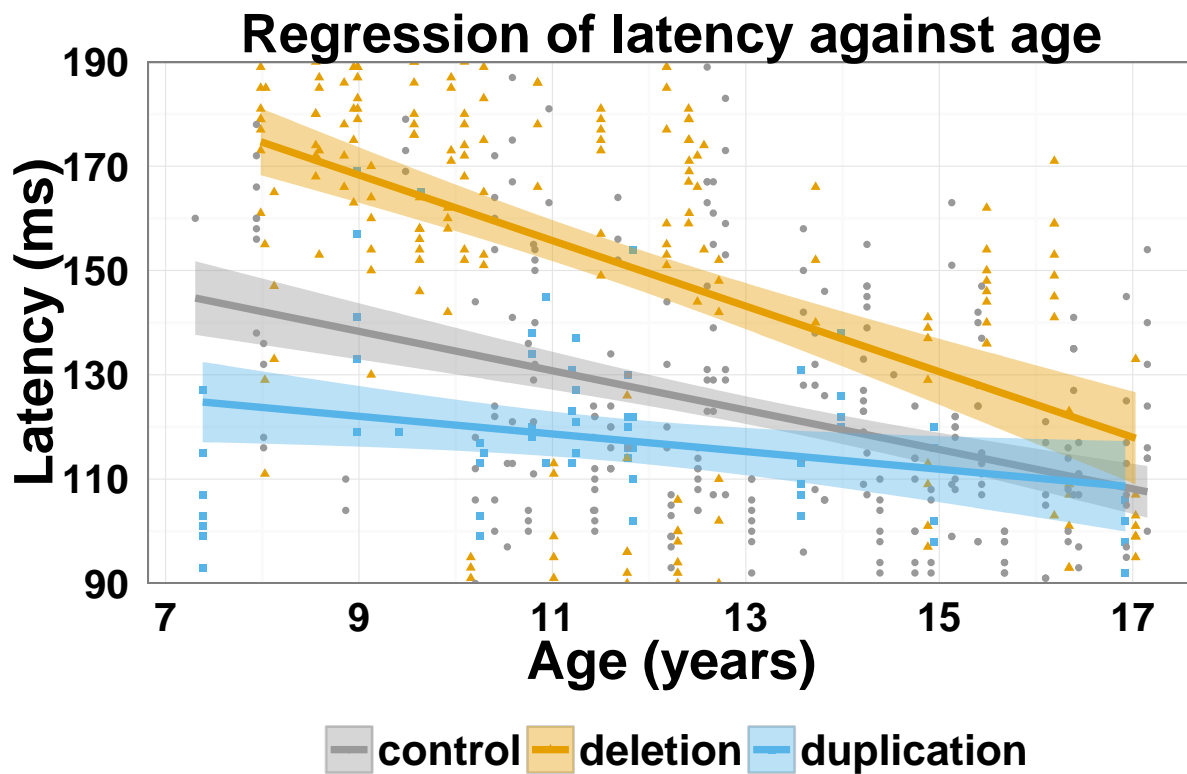
ggplot(child.study.data, aes(x = Age_Calc, y = M100LatCorr)) +
  geom_point(aes(colour = Case, shape = Case), size = 1.3) +
  geom_smooth(aes(colour = Case, fill = Case), method = 'lm', size = 1.3) +
  theme_bw() +
  scale_x_continuous(breaks = seq(5, 17, 2))+
  scale_y_continuous(breaks = seq(90, 215, 20))+
  coord_cartesian(ylim = c(90, 190)) +
  labs(x = 'Age (years)', y = 'Latency (ms)',
       title = 'Regression of latency against age') +
  theme(legend.position = 'bottom',
        legend.text = element_text(face = 'bold', size = 16),
        legend.title = element_blank(),
        axis.text.x = element_text(size = 15, face = 'bold'),
        axis.text.y = element_text(size = 15, face = 'bold'),
        axis.ticks = element_blank(),
        axis.title.x = element_text(size = 20, face = 'bold'),
        axis.title.y = element_text(size = 20, face = 'bold'),
        plot.title = element_text(size = 20, face = 'bold')) +
  scale_fill_manual(values = cbPalette) +
  scale_colour_manual(values = cbPalette) +
  scale_linetype_manual(values = c('solid', 'dashed', 'dotted'))
```

```
## Warning in loop_apply(n, do.ply): Removed 118 rows containing missing
## values (stat_smooth).
```

```
## Warning in loop_apply(n, do.ply): Removed 77 rows containing missing values
## (stat_smooth).
```

```
## Warning in loop_apply(n, do.ply): Removed 63 rows containing missing values
## (stat_smooth).
```

```
## Warning in loop_apply(n, do.ply): Removed 258 rows containing missing
## values (geom_point).
```



```
# regression of latency against age
```

```
lat.age.model <- lm(M100LatCorr ~ Age_Calc, data = child.study.data)
```

```
summary(lat.age.model)
```

```
##
## Call:
## lm(formula = M100LatCorr ~ Age_Calc, data = child.study.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -67.534 -20.699  -1.804   21.909   61.812
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   200.677     5.503   36.47  <2e-16 ***
## Age_Calc      -5.436     0.432  -12.58  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.08 on 532 degrees of freedom
```

```
## (258 observations deleted due to missingness)
## Multiple R-squared: 0.2294, Adjusted R-squared: 0.2279
## F-statistic: 158.4 on 1 and 532 DF, p-value: < 2.2e-16
```

```
# ANOVA-table style summary
summary.aov(lat.age.model)
```

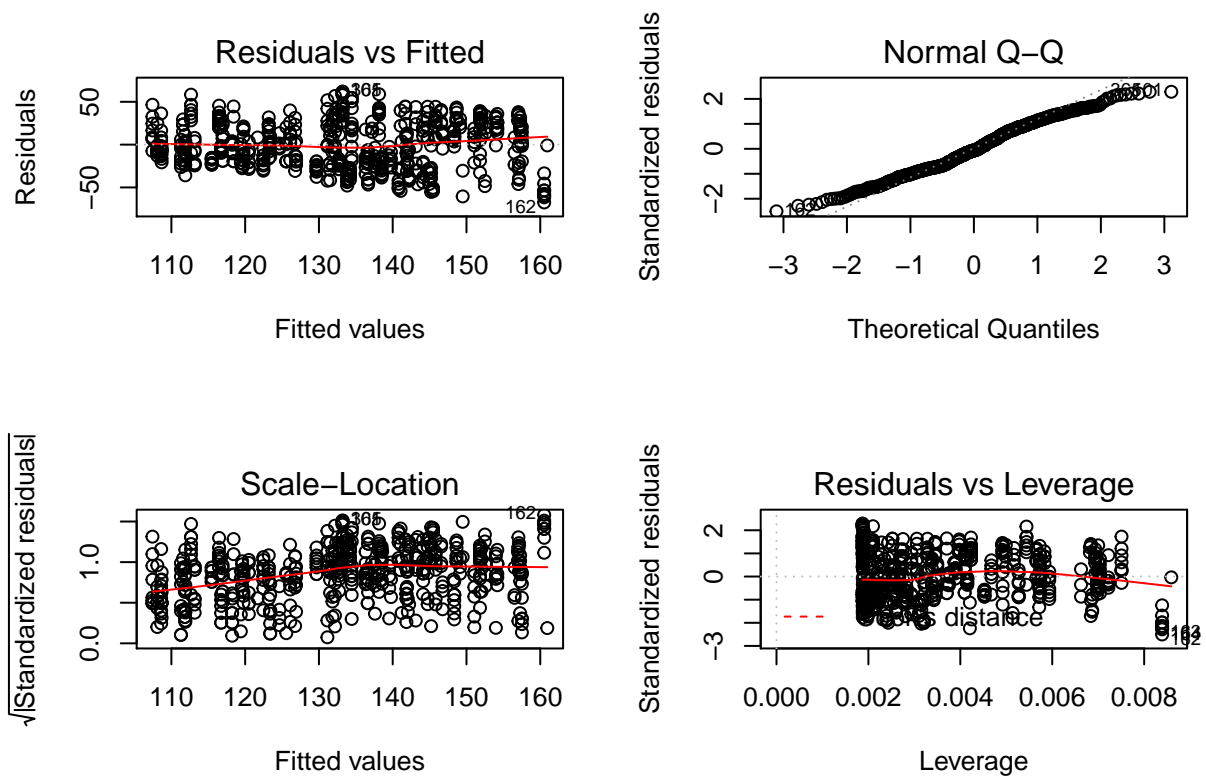
```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Age_Calc    1 116145  116145   158.4 <2e-16 ***
## Residuals  532 390177    733
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 258 observations deleted due to missingness
```

```
# treatment contrast summary
summary.lm(lat.age.model)
```

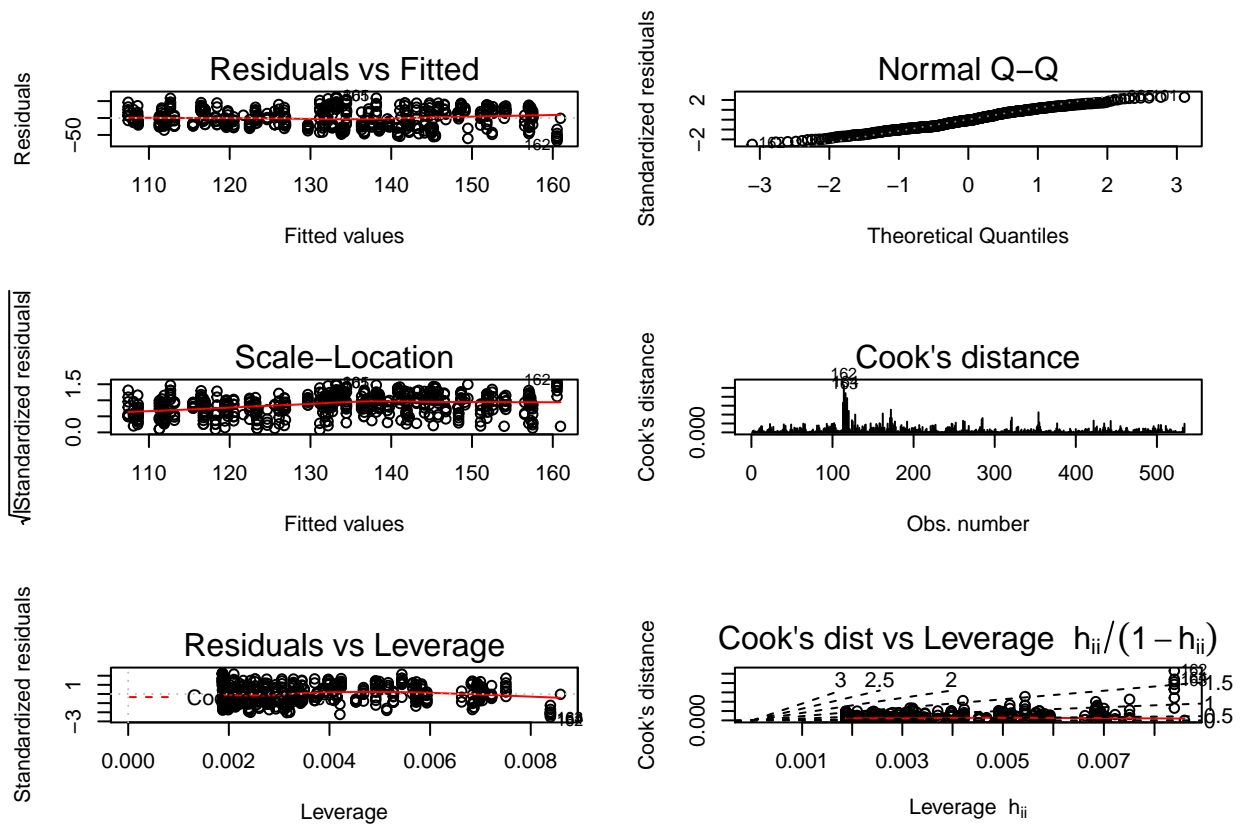
```
##
## Call:
## lm(formula = M100LatCorr ~ Age_Calc, data = child.study.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -67.534 -20.699  -1.804   21.909   61.812
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   200.677      5.503   36.47  <2e-16 ***
## Age_Calc       -5.436      0.432  -12.58  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.08 on 532 degrees of freedom
## (258 observations deleted due to missingness)
## Multiple R-squared: 0.2294, Adjusted R-squared: 0.2279
## F-statistic: 158.4 on 1 and 532 DF, p-value: < 2.2e-16
```

```
# plot model
```

```
# default plots
par(mfrow = c(2, 2))
plot(lat.age.model)
```

```
# all six diagnostic plots
par(mfrow = c(3, 2))
plot(lat.age.model, which = c(1:6))
```



```
# added functionality with ggplot2: append model fit values to the data.frame
```

```
# fortify model with data
```

```
lat.age.model.fortify <- fortify(lat.age.model)
```

```
# fortify parent data frame with model fit parameters
```

```
# refit without dropping NA; rows must match with fortifying
```

```
# data with model fit parameters
```

```
lat.age.model.2 <-
```

```
  lm(M100LatCorr ~ Age_Calc, data = child.study.data, na.action = na.exclude)
```

```
lat.age.model.data.fortify <- fortify(lat.age.model.2, child.study.data)
```

```
# for class lm(), glm() and anything that inherits from them
```

```
# you get yhat, sigma, cook's distance, fitted, residuals
```

```
# and standardized residuals
```

```
head(lat.age.model.fortify)
```

```
##      M100LatCorr Age_Calc      .hat      .sigma      .cooksd      .fitted
## 5           134   10.784 0.002574877 27.10488 1.143346e-04 142.0497
## 6           120   10.784 0.002574877 27.09019 8.578712e-04 142.0497
## 7           118   10.784 0.002574877 27.08698 1.020554e-03 142.0497
## 8           138   10.784 0.002574877 27.10656 2.893787e-05 142.0497
## 9           152   12.721 0.001892018 27.09253 5.431016e-04 131.5193
## 10          148   12.721 0.001892018 27.09768 3.516763e-04 131.5193
```

```
##      .resid .stdresid
## 5   -8.049723 -0.2976225
## 6  -22.049723 -0.8152447
## 7  -24.049723 -0.8891907
## 8   -4.049723 -0.1497305
## 9   20.480733  0.7569753
## 10  16.480733  0.6091339
```

```
summary(lat.age.model.fortify)
```

```
##      M100LatCorr      Age_Calc      .hat      .sigma
## Min.   : 76.0      Min.   : 7.307      Min.   :0.001873      Min.   :26.95
## 1st Qu.:107.0      1st Qu.:10.290      1st Qu.:0.002131      1st Qu.:27.07
## Median :125.5      Median :12.296      Median :0.003206      Median :27.09
## Mean   :133.0      Mean   :12.445      Mean   :0.003745      Mean   :27.08
## 3rd Qu.:159.0      3rd Qu.:14.882      3rd Qu.:0.005138      3rd Qu.:27.10
## Max.   :197.0      Max.   :17.151      Max.   :0.008591      Max.   :27.11
##      .cooksd      .fitted      .resid
## Min.   :2.900e-08      Min.   :107.4      Min.   : -67.534
## 1st Qu.:2.837e-04      1st Qu.:119.8      1st Qu.: -20.699
## Median :9.952e-04      Median :133.8      Median :  -1.804
## Mean   :1.798e-03      Mean   :133.0      Mean   :  0.000
## 3rd Qu.:2.250e-03      3rd Qu.:144.7      3rd Qu.: 21.909
## Max.   :2.653e-02      Max.   :161.0      Max.   : 61.812
##      .stdresid
## Min.   : -2.5042380
## 1st Qu.: -0.7658455
## Median : -0.0667208
## Mean   :  0.0000263
## 3rd Qu.:  0.8099385
## Max.   :  2.2845622
```

```
str(lat.age.model.fortify)
```

```
## 'data.frame': 534 obs. of 8 variables:
## $ M100LatCorr: int 134 120 118 138 152 148 142 110 102 90 ...
## $ Age_Calc : num 10.8 10.8 10.8 10.8 12.7 ...
## $ .hat : num 0.00257 0.00257 0.00257 0.00257 0.00189 ...
## $ .sigma : num 27.1 27.1 27.1 27.1 27.1 ...
## $ .cooksd : num 1.14e-04 8.58e-04 1.02e-03 2.89e-05 5.43e-04 ...
## $ .fitted : num 142 142 142 142 132 ...
## $ .resid : num -8.05 -22.05 -24.05 -4.05 20.48 ...
## $ .stdresid : num -0.298 -0.815 -0.889 -0.15 0.757 ...
## - attr(*, "terms")=Classes 'terms', 'formula' length 3 M100LatCorr ~ Age_Calc
## .. ..- attr(*, "variables")= language list(M100LatCorr, Age_Calc)
## .. ..- attr(*, "factors")= int [1:2, 1] 0 1
## .. ..- attr(*, "dimnames")=List of 2
## .. .. ..$ : chr [1:2] "M100LatCorr" "Age_Calc"
## .. .. ..$ : chr "Age_Calc"
## .. ..- attr(*, "term.labels")= chr "Age_Calc"
## .. ..- attr(*, "order")= int 1
## .. ..- attr(*, "intercept")= int 1
## .. ..- attr(*, "response")= int 1
```

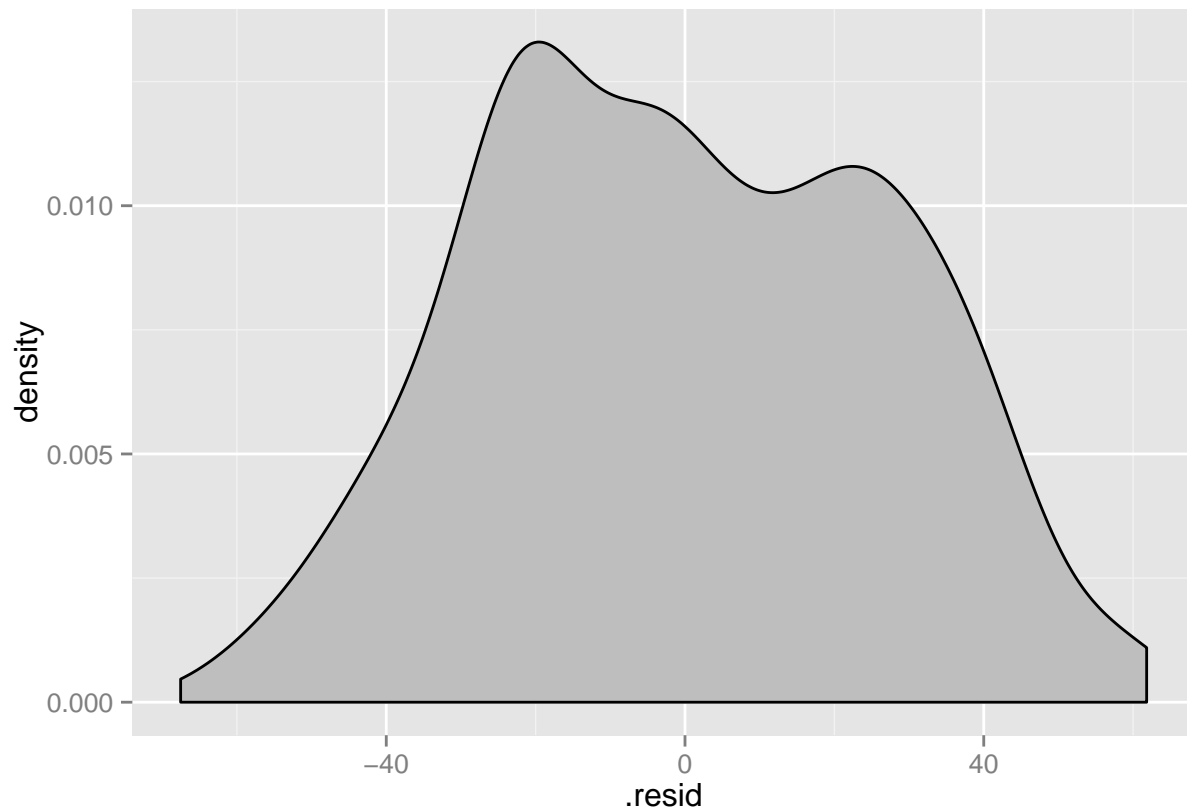
```
## ..- attr(*, ".Environment")=<environment: R_GlobalEnv>
## ..- attr(*, "predvars")= language list(M100LatCorr, Age_Calc)
## ..- attr(*, "dataClasses")= Named chr [1:2] "numeric" "numeric"
## ..- attr(*, "names")= chr [1:2] "M100LatCorr" "Age_Calc"
## - attr(*, "na.action")=Class 'omit' Named int [1:258] 1 2 3 4 11 15 21 22 23 24 ...
## ..- attr(*, "names")= chr [1:258] "1" "2" "3" "4" ...
```

```
dplyr::glimpse(lat.age.model.fortify)
```

```
## Observations: 534
## Variables:
## $ M100LatCorr (int) 134, 120, 118, 138, 152, 148, 142, 110, 102, 90, 1...
## $ Age_Calc     (dbl) 10.784, 10.784, 10.784, 10.784, 12.721, 12.721, 12...
## $ .hat         (dbl) 0.002574877, 0.002574877, 0.002574877, 0.002574877...
## $ .sigma       (dbl) 27.10488, 27.09019, 27.08698, 27.10656, 27.09253, ...
## $ .cooksds     (dbl) 1.143346e-04, 8.578712e-04, 1.020554e-03, 2.893787...
## $ .fitted      (dbl) 142.0497, 142.0497, 142.0497, 142.0497, 131.5193, ...
## $ .resid       (dbl) -8.049723, -22.049723, -24.049723, -4.049723, 20.4...
## $ .stdresid    (dbl) -0.2976225, -0.8152447, -0.8891907, -0.1497305, 0....
```

```
# plot distribution of residuals
```

```
ggplot(lat.age.model.fortify, aes(.resid)) + geom_density(fill = 'gray')
```



```
# additional diagnostics and plots with car package
library(car)
```

```
durbinWatsonTest(lat.age.model$fortify$resid)
```

```
## [1] 0.7242333
```

```
outlierTest(lat.age.model)
```

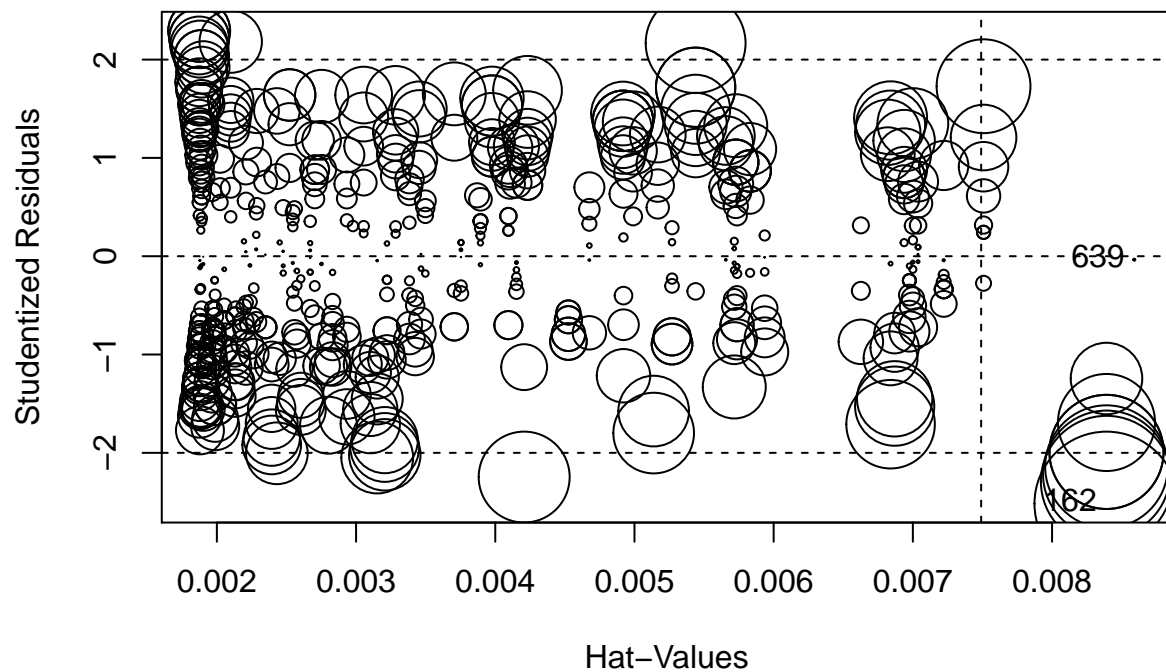
```
##  
## No Studentized residuals with Bonferonni p < 0.05  
## Largest |rstudent|:  
##      rstudent unadjusted p-value Bonferonni p  
## 162 -2.516761      0.012137      NA
```

```
sigmaHat(lat.age.model)
```

```
## [1] 27.08164
```

```
par(mfrow = c(1, 1))
```

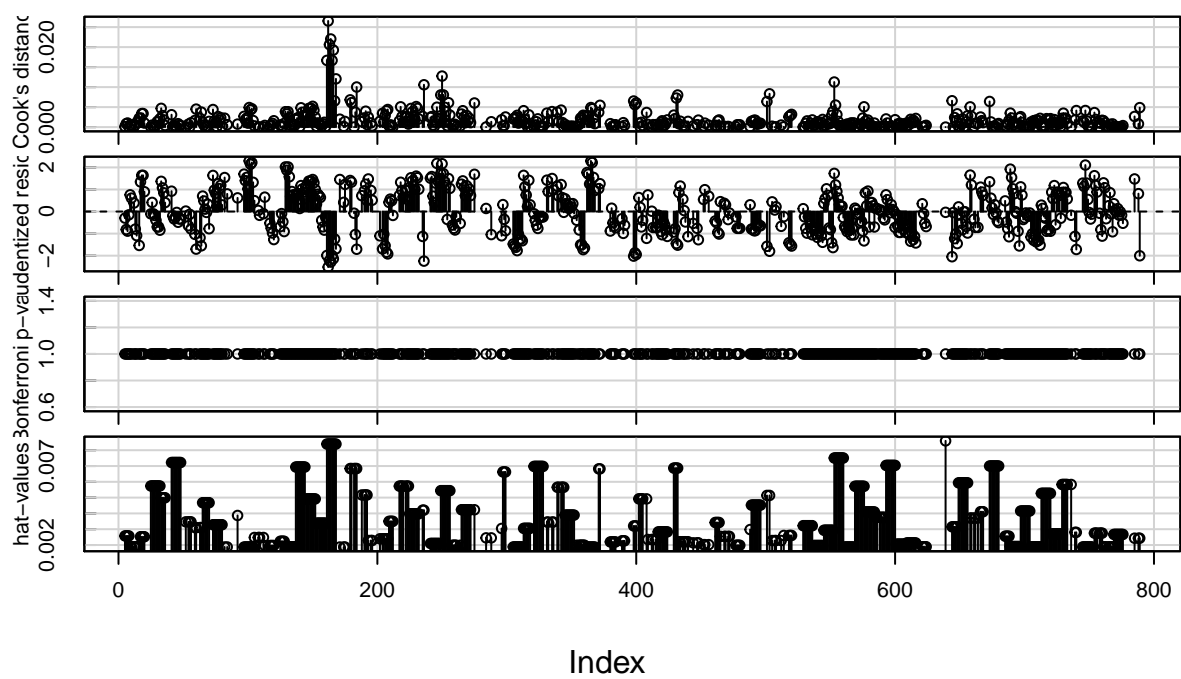
```
influencePlot(lat.age.model)
```



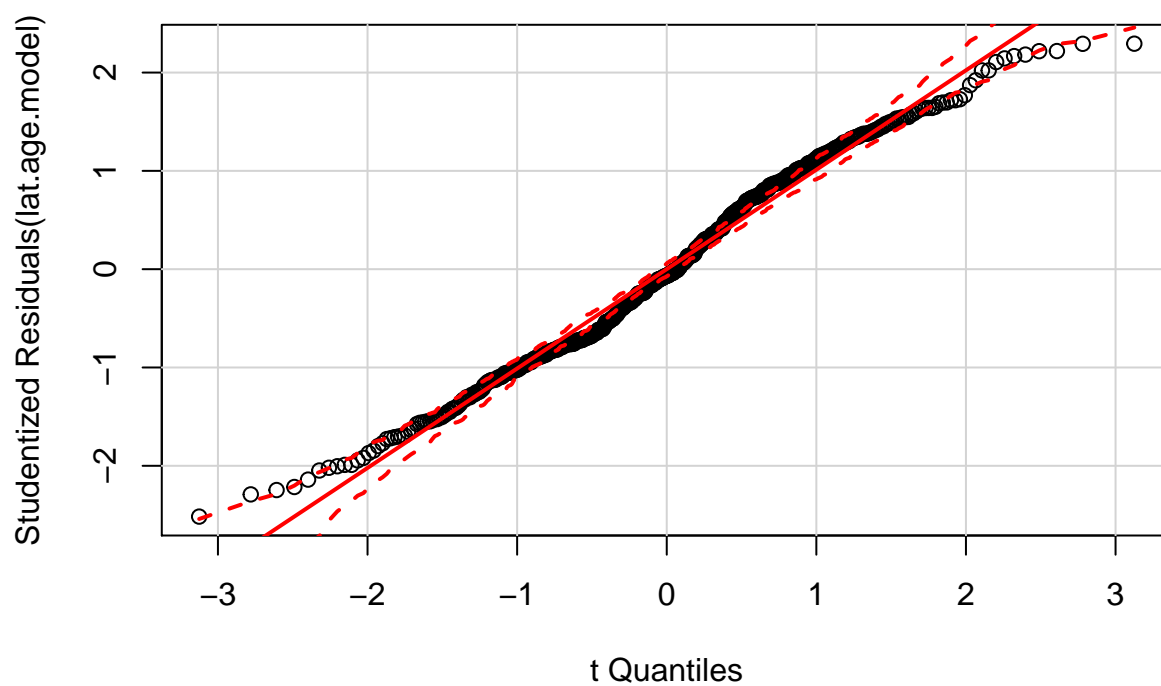
```
##      StudRes      Hat      CookD  
## 162 -2.51676103 0.008391043 0.162891471  
## 639 -0.03528496 0.008590890 0.002324745
```

```
infIndexPlot(lat.age.model)
```

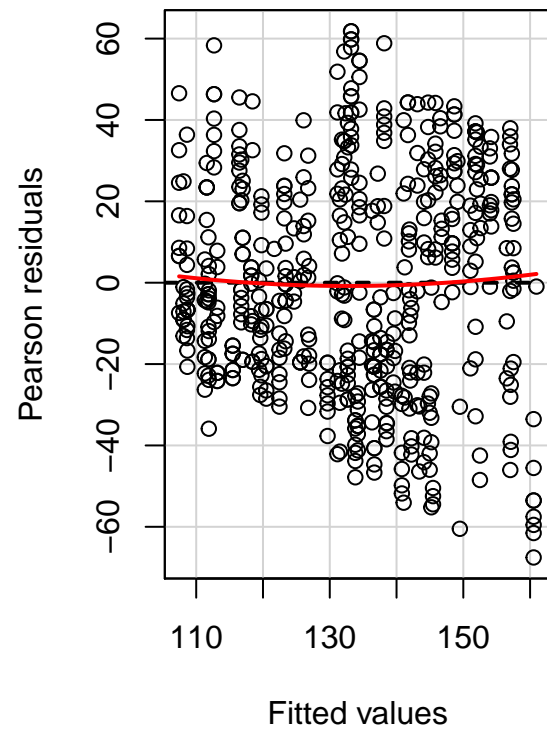
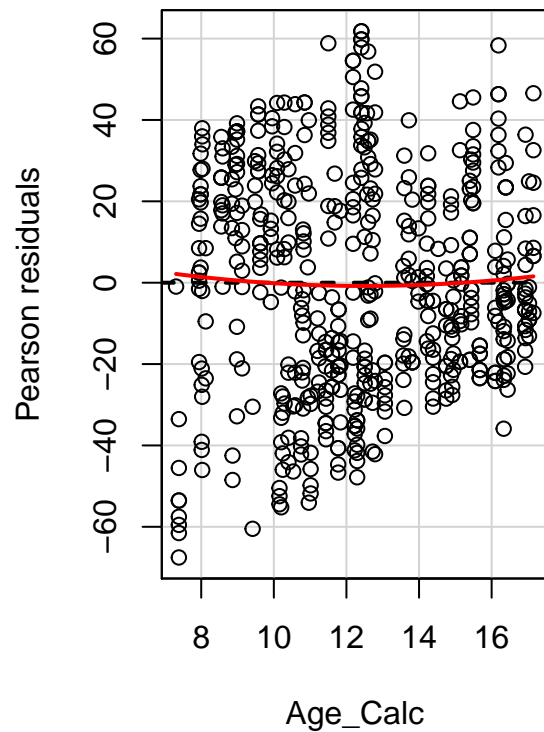
Diagnostic Plots



```
qqPlot(lat.age.model)
```



```
residualPlots(lat.age.model)
```



```
##          Test stat Pr(>|t|)
## Age_Calc      0.667   0.505
## Tukey test     0.667   0.504
```

```
# linear mixed effects models -----
```

```
library(lme4)
```

```
## Loading required package: Matrix
##
## Attaching package: 'Matrix'
##
## The following objects are masked from 'package:base':
##
##   crossprod, tcrossprod
##
## Loading required package: Rcpp
```

```
library(multcomp)
```

```
## Loading required package: mvtnorm
## Loading required package: survival
## Loading required package: TH.data
```

```
# internal numeric representation of the categorical predictors
```

```
unique(as.numeric(child.study.data$Hem))
```

```
## [1] 1 2
```

```
unique(as.numeric(child.study.data$Cond))
```

```
## [1] 1 2 3 4
```

```
unique(as.numeric(child.study.data$Case))
```

```
## [1] 3 2 1
```

```
unique(as.numeric(child.study.data$Site))
```

```
## [1] 2 1
```

```
levels(child.study.data$Hem)
```

```
## [1] "1-LH" "2-RH"
```

```
levels(child.study.data$Cond)
```

```
## [1] "1-200" "2-300" "3-500" "4-1000"
```

```
levels(child.study.data$Case)
```

```
## [1] "control" "deletion" "duplication"
```

```
levels(child.study.data$Site)
```

```
## [1] "CHOP" "UCSF"
```

```
# since missing values are removed by default, no need to drop  
# but we can include them if we want to impute missing values  
# or fortify the data.frame with model fit values
```

```
m100.lmm <-  
  lmer(M100LatCorr ~ Hem + Cond + Case + Site + Age_Calc +  
        (Hem + Cond | Subject), data = child.study.data,  
        REML = FALSE)
```

```
# include rows with missing values in the dependent variable
```

```
m100.lmm.2 <-  
  lmer(M100LatCorr ~ Hem + Cond + Case + Site + Age_Calc +  
        (Hem + Cond | Subject), data = child.study.data,  
        na.action = na.exclude, REML = FALSE)
```

```
# note: can only use summary() directly if NA values have been dropped
```

```
summary(m100.lmm)
```



```

## Linear mixed model fit by maximum likelihood ['lmerMod']
## Formula:
## M100LatCorr ~ Hem + Cond + Case + Site + Age_Calc + (Hem + Cond |
##   Subject)
##   Data: child.study.data
##
##      AIC      BIC   logLik deviance df.resid
##  4253.0   4360.0 -2101.5   4203.0     509
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.6691 -0.4596 -0.0101  0.4325  2.6082
##
## Random effects:
##   Groups   Name      Variance Std.Dev. Corr
##   Subject (Intercept) 479.23   21.891
##           Hem2-RH     364.62   19.095  -0.51
##           Cond2-300    13.28    3.644   0.32 -0.13
##           Cond3-500    41.98    6.479  -0.09  0.22  0.66
##           Cond4-1000   75.08    8.665  -0.31  0.00  0.61  0.89
##   Residual              49.67    7.048
## Number of obs: 534, groups: Subject, 96
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)   192.3710    10.4107  18.478
## Hem2-RH       -14.5908     2.2934  -6.362
## Cond2-300      -4.4720     1.0271  -4.354
## Cond3-500      -8.7888     1.1908  -7.381
## Cond4-1000    -16.5421     1.3258 -12.477
## Casedeletion   20.9438     4.3431   4.822
## Caseduplication -5.4485     5.5698  -0.978
## SiteUCSF       -2.4539     3.8051  -0.645
## Age_Calc       -3.9318     0.7395  -5.317
##
## Correlation of Fixed Effects:
##              (Intr) Hm2-RH C2-300 C3-500 C4-100 Csdltcn Csdplcn StUCSF
## Hem2-RH      -0.138
## Cond2-300    -0.057 -0.035
## Cond3-500    -0.092  0.132  0.546
## Cond4-1000  -0.103  0.014  0.516  0.697
## Casedeletin -0.459  0.064  0.002  0.005 -0.003
## Caseduplctn -0.332 -0.001 -0.029 -0.029 -0.017  0.374
## SiteUCSF     -0.228 -0.012  0.001 -0.004  0.009  0.157  0.148
## Age_Calc     -0.924  0.001  0.045  0.042  0.015  0.276  0.188  0.002

```

Anova(m100.lmm)

```

## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: M100LatCorr
##              Chisq Df Pr(>Chisq)
## Hem          40.4769  1  1.990e-10 ***
## Cond        162.9301  3  < 2.2e-16 ***

```

```
## Case      32.2650  2  9.857e-08 ***
## Site      0.4159  1    0.519
## Age_Calc  28.2661  1  1.057e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(glht(m100.lmm, linfct = mcp(Hem = 'Tukey', covariate_average = TRUE)))
```

```
##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmer(formula = M100LatCorr ~ Hem + Cond + Case + Site + Age_Calc +
## (Hem + Cond | Subject), data = child.study.data, REML = FALSE)
##
## Linear Hypotheses:
##              Estimate Std. Error z value Pr(>|z|)
## 2-RH - 1-LH == 0  -14.591      2.293  -6.362 1.99e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```
summary(glht(m100.lmm, linfct = mcp(Case = 'Tukey', covariate_average = TRUE)))
```

```
##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmer(formula = M100LatCorr ~ Hem + Cond + Case + Site + Age_Calc +
## (Hem + Cond | Subject), data = child.study.data, REML = FALSE)
##
## Linear Hypotheses:
##              Estimate Std. Error z value Pr(>|z|)
## deletion - control == 0    20.944      4.343   4.822 <1e-04 ***
## duplication - control == 0   -5.448      5.570  -0.978  0.588
## duplication - deletion == 0 -26.392      5.637  -4.682 <1e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```
summary(glht(m100.lmm, linfct = mcp(Cond = 'Tukey', covariate_average = TRUE)))
```

```
##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
```

```
## Fit: lmer(formula = M100LatCorr ~ Hem + Cond + Case + Site + Age_Calc +
##       (Hem + Cond | Subject), data = child.study.data, REML = FALSE)
##
## Linear Hypotheses:
##               Estimate Std. Error z value Pr(>|z|)
## 2-300 - 1-200 == 0   -4.4720    1.0271  -4.354  <1e-04 ***
## 3-500 - 1-200 == 0   -8.7888    1.1908  -7.381  <1e-04 ***
## 4-1000 - 1-200 == 0 -16.5421    1.3258 -12.477  <1e-04 ***
## 3-500 - 2-300 == 0   -4.3167    1.0668  -4.046   3e-04 ***
## 4-1000 - 2-300 == 0 -12.0701    1.1859 -10.178  <1e-04 ***
## 4-1000 - 3-500 == 0   -7.7533    0.9878  -7.849  <1e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

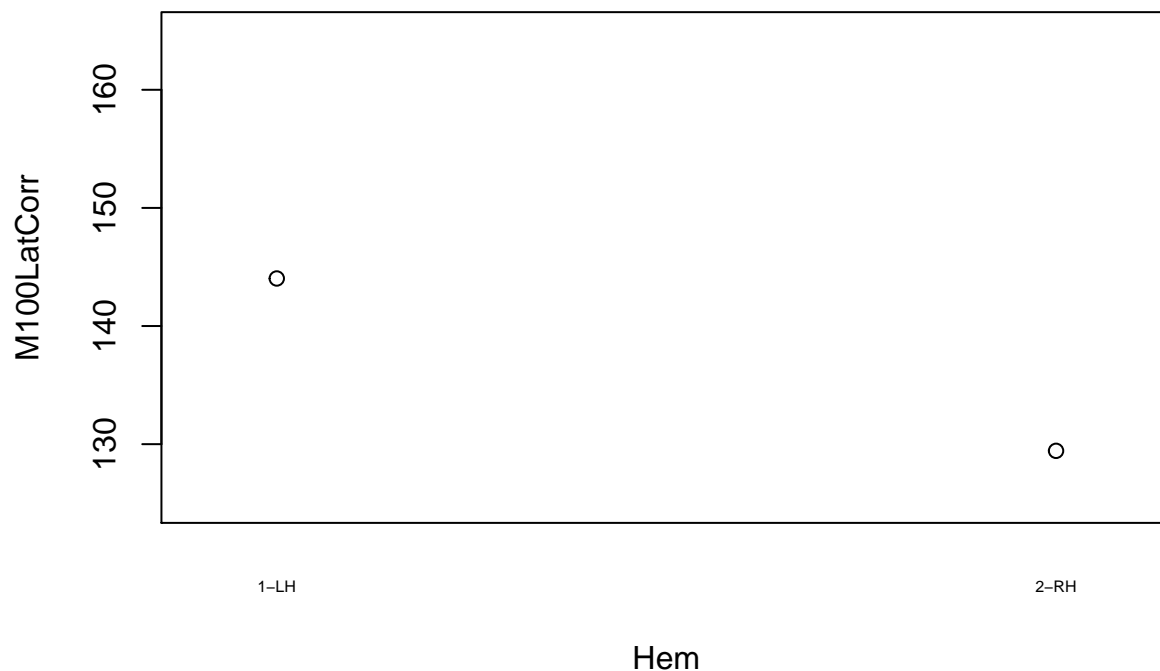
```
# plot with languageR
```

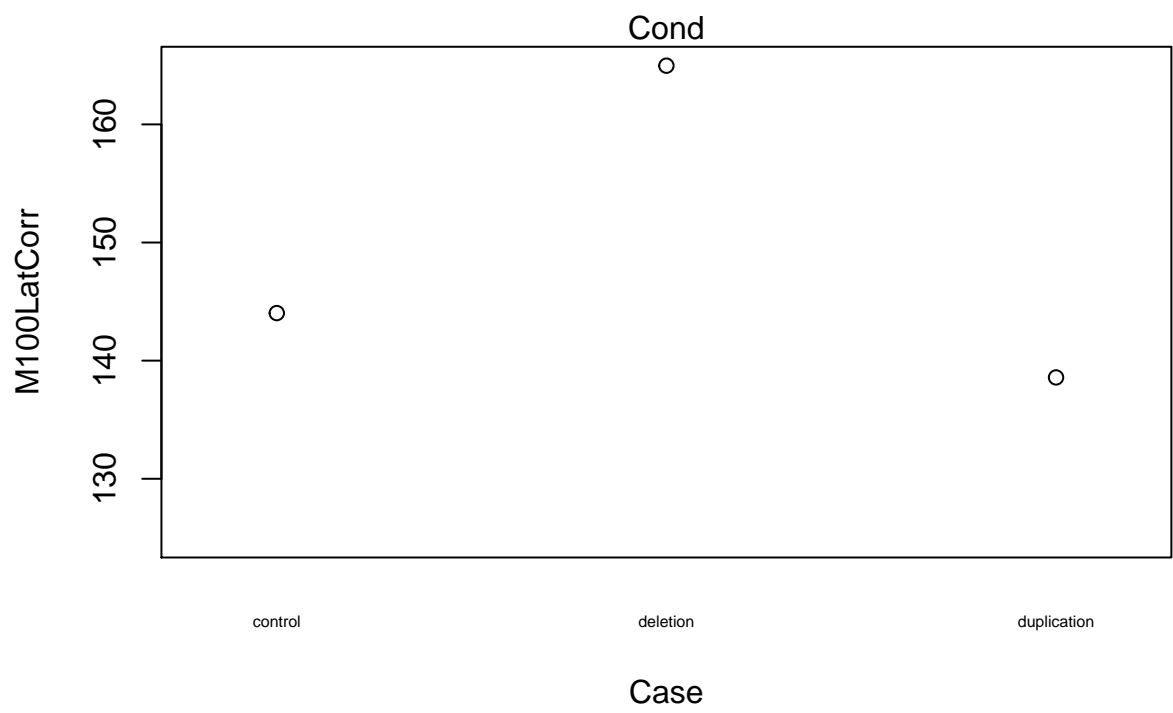
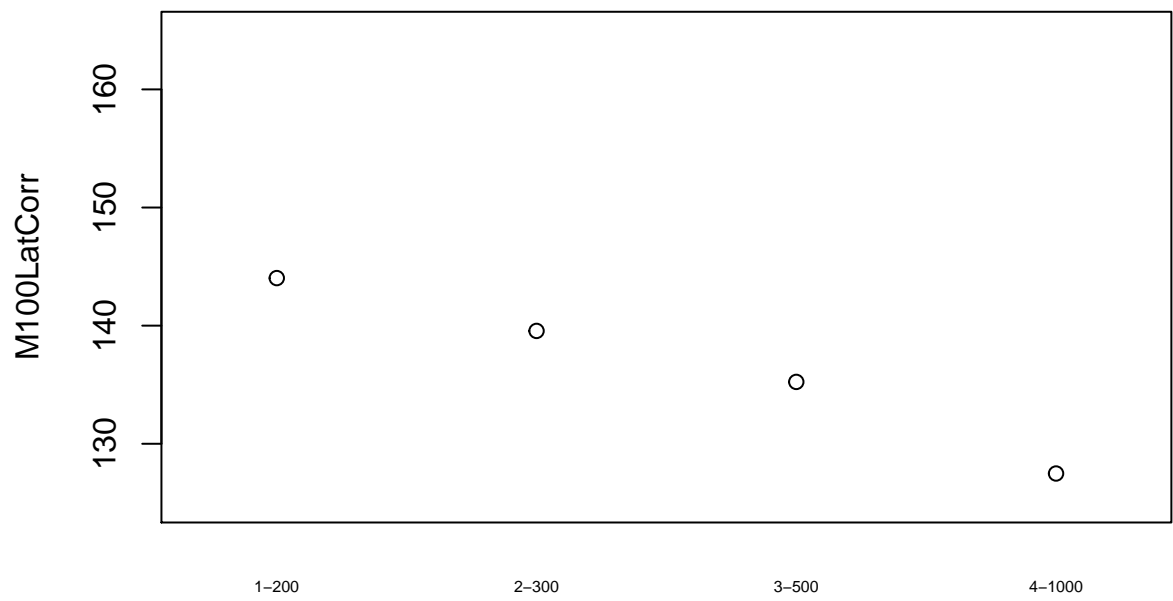
```
library(languageR)
```

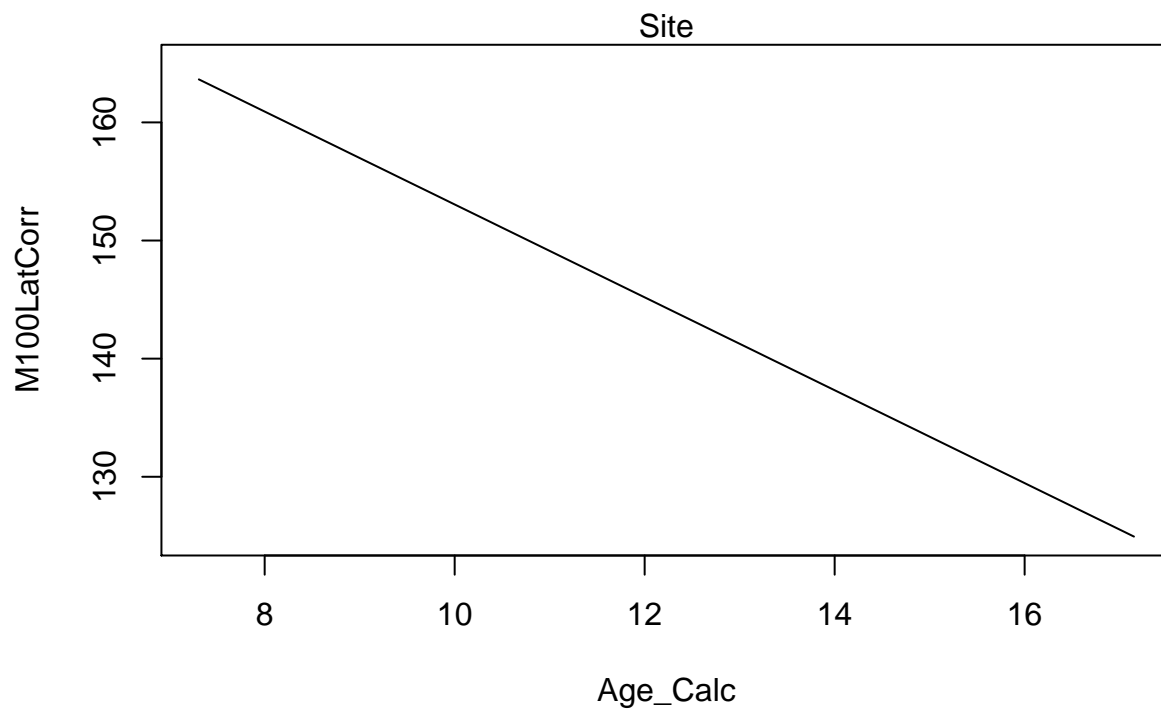
```
par(mfrow = c(1, 1))
```

```
plotLMER.fnc(m100.lmm, withList = TRUE)
```

```
## effect size (range) for Hem is 14.59082
## effect size (range) for Cond is 16.5421
## effect size (range) for Case is 26.39229
## effect size (range) for Site is 2.453915
## effect size (range) for Age_Calc is 38.70435
```







```
## $Hem
##   X      Y Predictor Type Interaction Levels
## 1 1 144.0260      Hem TRUE      FALSE 1-LH
## 2 2 129.4351      Hem TRUE      FALSE 2-RH
##
## $Cond
##   X      Y Predictor Type Interaction Levels
## 1 1 144.0260      Cond TRUE      FALSE 1-200
## 2 2 139.5539      Cond TRUE      FALSE 2-300
## 3 3 135.2372      Cond TRUE      FALSE 3-500
## 4 4 127.4839      Cond TRUE      FALSE 4-1000
```

```

##
## $Case
##   X          Y Predictor Type Interaction      Levels
## 1 1 144.0260      Case TRUE          FALSE    control
## 2 2 164.9698      Case TRUE          FALSE    deletion
## 3 3 138.5775      Case TRUE          FALSE    duplication
##
## $Site
##   X          Y Predictor Type Interaction Levels
## 1 1 144.026      Site TRUE          FALSE    CHOP
## 2 2 141.572      Site TRUE          FALSE    UCSF
##
## $Age_Calc
##           X          Y Predictor Type Interaction
## 1    7.307000 163.6416 Age_Calc FALSE          NA
## 2    7.406434 163.2506 Age_Calc FALSE          NA
## 3    7.505869 162.8597 Age_Calc FALSE          NA
## 4    7.605303 162.4687 Age_Calc FALSE          NA
## 5    7.704737 162.0778 Age_Calc FALSE          NA
## 6    7.804172 161.6868 Age_Calc FALSE          NA
## 7    7.903606 161.2958 Age_Calc FALSE          NA
## 8    8.003040 160.9049 Age_Calc FALSE          NA
## 9    8.102475 160.5139 Age_Calc FALSE          NA
## 10   8.201909 160.1230 Age_Calc FALSE          NA
## 11   8.301343 159.7320 Age_Calc FALSE          NA
## 12   8.400778 159.3411 Age_Calc FALSE          NA
## 13   8.500212 158.9501 Age_Calc FALSE          NA
## 14   8.599646 158.5592 Age_Calc FALSE          NA
## 15   8.699081 158.1682 Age_Calc FALSE          NA
## 16   8.798515 157.7773 Age_Calc FALSE          NA
## 17   8.897949 157.3863 Age_Calc FALSE          NA
## 18   8.997384 156.9954 Age_Calc FALSE          NA
## 19   9.096818 156.6044 Age_Calc FALSE          NA
## 20   9.196253 156.2135 Age_Calc FALSE          NA
## 21   9.295687 155.8225 Age_Calc FALSE          NA
## 22   9.395121 155.4316 Age_Calc FALSE          NA
## 23   9.494556 155.0406 Age_Calc FALSE          NA
## 24   9.593990 154.6496 Age_Calc FALSE          NA
## 25   9.693424 154.2587 Age_Calc FALSE          NA
## 26   9.792859 153.8677 Age_Calc FALSE          NA
## 27   9.892293 153.4768 Age_Calc FALSE          NA
## 28   9.991727 153.0858 Age_Calc FALSE          NA
## 29  10.091162 152.6949 Age_Calc FALSE          NA
## 30  10.190596 152.3039 Age_Calc FALSE          NA
## 31  10.290030 151.9130 Age_Calc FALSE          NA
## 32  10.389465 151.5220 Age_Calc FALSE          NA
## 33  10.488899 151.1311 Age_Calc FALSE          NA
## 34  10.588333 150.7401 Age_Calc FALSE          NA
## 35  10.687768 150.3492 Age_Calc FALSE          NA
## 36  10.787202 149.9582 Age_Calc FALSE          NA
## 37  10.886636 149.5673 Age_Calc FALSE          NA
## 38  10.986071 149.1763 Age_Calc FALSE          NA
## 39  11.085505 148.7853 Age_Calc FALSE          NA
## 40  11.184939 148.3944 Age_Calc FALSE          NA

```

## 41	11.284374	148.0034	Age_Calc	FALSE	NA
## 42	11.383808	147.6125	Age_Calc	FALSE	NA
## 43	11.483242	147.2215	Age_Calc	FALSE	NA
## 44	11.582677	146.8306	Age_Calc	FALSE	NA
## 45	11.682111	146.4396	Age_Calc	FALSE	NA
## 46	11.781545	146.0487	Age_Calc	FALSE	NA
## 47	11.880980	145.6577	Age_Calc	FALSE	NA
## 48	11.980414	145.2668	Age_Calc	FALSE	NA
## 49	12.079848	144.8758	Age_Calc	FALSE	NA
## 50	12.179283	144.4849	Age_Calc	FALSE	NA
## 51	12.278717	144.0939	Age_Calc	FALSE	NA
## 52	12.378152	143.7030	Age_Calc	FALSE	NA
## 53	12.477586	143.3120	Age_Calc	FALSE	NA
## 54	12.577020	142.9211	Age_Calc	FALSE	NA
## 55	12.676455	142.5301	Age_Calc	FALSE	NA
## 56	12.775889	142.1391	Age_Calc	FALSE	NA
## 57	12.875323	141.7482	Age_Calc	FALSE	NA
## 58	12.974758	141.3572	Age_Calc	FALSE	NA
## 59	13.074192	140.9663	Age_Calc	FALSE	NA
## 60	13.173626	140.5753	Age_Calc	FALSE	NA
## 61	13.273061	140.1844	Age_Calc	FALSE	NA
## 62	13.372495	139.7934	Age_Calc	FALSE	NA
## 63	13.471929	139.4025	Age_Calc	FALSE	NA
## 64	13.571364	139.0115	Age_Calc	FALSE	NA
## 65	13.670798	138.6206	Age_Calc	FALSE	NA
## 66	13.770232	138.2296	Age_Calc	FALSE	NA
## 67	13.869667	137.8387	Age_Calc	FALSE	NA
## 68	13.969101	137.4477	Age_Calc	FALSE	NA
## 69	14.068535	137.0568	Age_Calc	FALSE	NA
## 70	14.167970	136.6658	Age_Calc	FALSE	NA
## 71	14.267404	136.2749	Age_Calc	FALSE	NA
## 72	14.366838	135.8839	Age_Calc	FALSE	NA
## 73	14.466273	135.4929	Age_Calc	FALSE	NA
## 74	14.565707	135.1020	Age_Calc	FALSE	NA
## 75	14.665141	134.7110	Age_Calc	FALSE	NA
## 76	14.764576	134.3201	Age_Calc	FALSE	NA
## 77	14.864010	133.9291	Age_Calc	FALSE	NA
## 78	14.963444	133.5382	Age_Calc	FALSE	NA
## 79	15.062879	133.1472	Age_Calc	FALSE	NA
## 80	15.162313	132.7563	Age_Calc	FALSE	NA
## 81	15.261747	132.3653	Age_Calc	FALSE	NA
## 82	15.361182	131.9744	Age_Calc	FALSE	NA
## 83	15.460616	131.5834	Age_Calc	FALSE	NA
## 84	15.560051	131.1925	Age_Calc	FALSE	NA
## 85	15.659485	130.8015	Age_Calc	FALSE	NA
## 86	15.758919	130.4106	Age_Calc	FALSE	NA
## 87	15.858354	130.0196	Age_Calc	FALSE	NA
## 88	15.957788	129.6286	Age_Calc	FALSE	NA
## 89	16.057222	129.2377	Age_Calc	FALSE	NA
## 90	16.156657	128.8467	Age_Calc	FALSE	NA
## 91	16.256091	128.4558	Age_Calc	FALSE	NA
## 92	16.355525	128.0648	Age_Calc	FALSE	NA
## 93	16.454960	127.6739	Age_Calc	FALSE	NA
## 94	16.554394	127.2829	Age_Calc	FALSE	NA

```
## 95 16.653828 126.8920 Age_Calc FALSE NA
## 96 16.753263 126.5010 Age_Calc FALSE NA
## 97 16.852697 126.1101 Age_Calc FALSE NA
## 98 16.952131 125.7191 Age_Calc FALSE NA
## 99 17.051566 125.3282 Age_Calc FALSE NA
## 100 17.151000 124.9372 Age_Calc FALSE NA
```

```
# fortify with model fit parameters
# using model with NA rows included since rows must match

m100.lmm.2.fortify <- fortify(m100.lmm.2, child.study.data)

head(m100.lmm.2.fortify)
```

```
##      Subject Site Exclusions      DOS      DOB Age Age_Calc AgeGroup Gender
## 1 3002-102 UCSF          no 11/4/11 1/24/01 10 10.784 child male
## 2 3002-102 UCSF          no 11/4/11 1/24/01 10 10.784 child male
## 3 3002-102 UCSF          no 11/4/11 1/24/01 10 10.784 child male
## 4 3002-102 UCSF          no 11/4/11 1/24/01 10 10.784 child male
## 5 3002-102 UCSF          no 11/4/11 1/24/01 10 10.784 child male
## 6 3002-102 UCSF          no 11/4/11 1/24/01 10 10.784 child male
##      Handedness      Dx Chromosome      Case      Copies      ASD NVIQ VIQ
## 1      right proband      16p duplication duplication FALSE 77 86
## 2      right proband      16p duplication duplication FALSE 77 86
## 3      right proband      16p duplication duplication FALSE 77 86
## 4      right proband      16p duplication duplication FALSE 77 86
## 5      right proband      16p duplication duplication FALSE 77 86
## 6      right proband      16p duplication duplication FALSE 77 86
##      CELF.4 SRS_parent SRS_adult CTOPP      ICV2      cmICV Hem      Cond M50Lat
## 1      75      145      NA      6 1422314 1422.314 1-LH 1-200 120
## 2      75      145      NA      6 1422314 1422.314 1-LH 2-300 NA
## 3      75      145      NA      6 1422314 1422.314 1-LH 3-500 NA
## 4      75      145      NA      6 1422314 1422.314 1-LH 4-1000 NA
## 5      75      145      NA      6 1422314 1422.314 2-RH 1-200 NA
## 6      75      145      NA      6 1422314 1422.314 2-RH 2-300 NA
##      M100Lat M50LatCorr M100LatCorr dB.SL M100compCase M50compCase      cutAge
## 1      NA      92      NA      25      0      1 under-12
## 2      NA      NA      NA      25      0      0 under-12
## 3      NA      NA      NA      25      0      0 under-12
## 4      NA      NA      NA      25      0      0 under-12
## 5      162      NA      134      25      1      0 under-12
## 6      148      NA      120      25      1      0 under-12
##      breakAge .fitted .resid .sresid
## 1 (9,10]      NA      NA      NA
## 2 (9,10]      NA      NA      NA
## 3 (9,10]      NA      NA      NA
## 4 (9,10]      NA      NA      NA
## 5 (9,10] 129.7170 4.283048 0.6077238
## 6 (9,10] 127.7057 -7.705655 -1.0933591
```

```
# predict new values and allowing for levels/rows not present in fit

m100.lmm.2.fortify$Predicted <-
```



```
predict(m100.lmm.2, newdata = m100.lmm.2.fortify,
        na.action = na.pass, allow.new.levels = TRUE)

head(m100.lmm.2.fortify)
```

```
##      Subject Site Exclusions      DOS      DOB Age Age_Calc AgeGroup Gender
## 1 3002-102 UCSF          no 11/4/11 1/24/01  10  10.784   child   male
## 2 3002-102 UCSF          no 11/4/11 1/24/01  10  10.784   child   male
## 3 3002-102 UCSF          no 11/4/11 1/24/01  10  10.784   child   male
## 4 3002-102 UCSF          no 11/4/11 1/24/01  10  10.784   child   male
## 5 3002-102 UCSF          no 11/4/11 1/24/01  10  10.784   child   male
## 6 3002-102 UCSF          no 11/4/11 1/24/01  10  10.784   child   male
##      Handedness      Dx Chromosome      Case      Copies      ASD NVIQ VIQ
## 1      right proband      16p duplication duplication FALSE    77 86
## 2      right proband      16p duplication duplication FALSE    77 86
## 3      right proband      16p duplication duplication FALSE    77 86
## 4      right proband      16p duplication duplication FALSE    77 86
## 5      right proband      16p duplication duplication FALSE    77 86
## 6      right proband      16p duplication duplication FALSE    77 86
##      CELF.4 SRS_parent SRS_adult CTOPP      ICV2      cmICV      Hem      Cond M50Lat
## 1      75      145      NA      6 1422314 1422.314 1-LH 1-200    120
## 2      75      145      NA      6 1422314 1422.314 1-LH 2-300     NA
## 3      75      145      NA      6 1422314 1422.314 1-LH 3-500     NA
## 4      75      145      NA      6 1422314 1422.314 1-LH 4-1000    NA
## 5      75      145      NA      6 1422314 1422.314 2-RH 1-200     NA
## 6      75      145      NA      6 1422314 1422.314 2-RH 2-300     NA
##      M100Lat M50LatCorr M100LatCorr dB.SL M100compCase M50compCase      cutAge
## 1      NA      92      NA      25      0      1 under-12
## 2      NA      NA      NA      25      0      0 under-12
## 3      NA      NA      NA      25      0      0 under-12
## 4      NA      NA      NA      25      0      0 under-12
## 5     162      NA     134      25      1      0 under-12
## 6     148      NA     120      25      1      0 under-12
##      breakAge .fitted .resid .sresid Predicted
## 1 (9,10]      NA      NA      NA 138.4956
## 2 (9,10]      NA      NA      NA 136.4843
## 3 (9,10]      NA      NA      NA 136.6216
## 4 (9,10]      NA      NA      NA 130.1613
## 5 (9,10] 129.7170 4.283048 0.6077238 129.7170
## 6 (9,10] 127.7057 -7.705655 -1.0933591 127.7057
```

```
# refit latency-age relationship and plot the resulting model
```

```
library(dplyr)

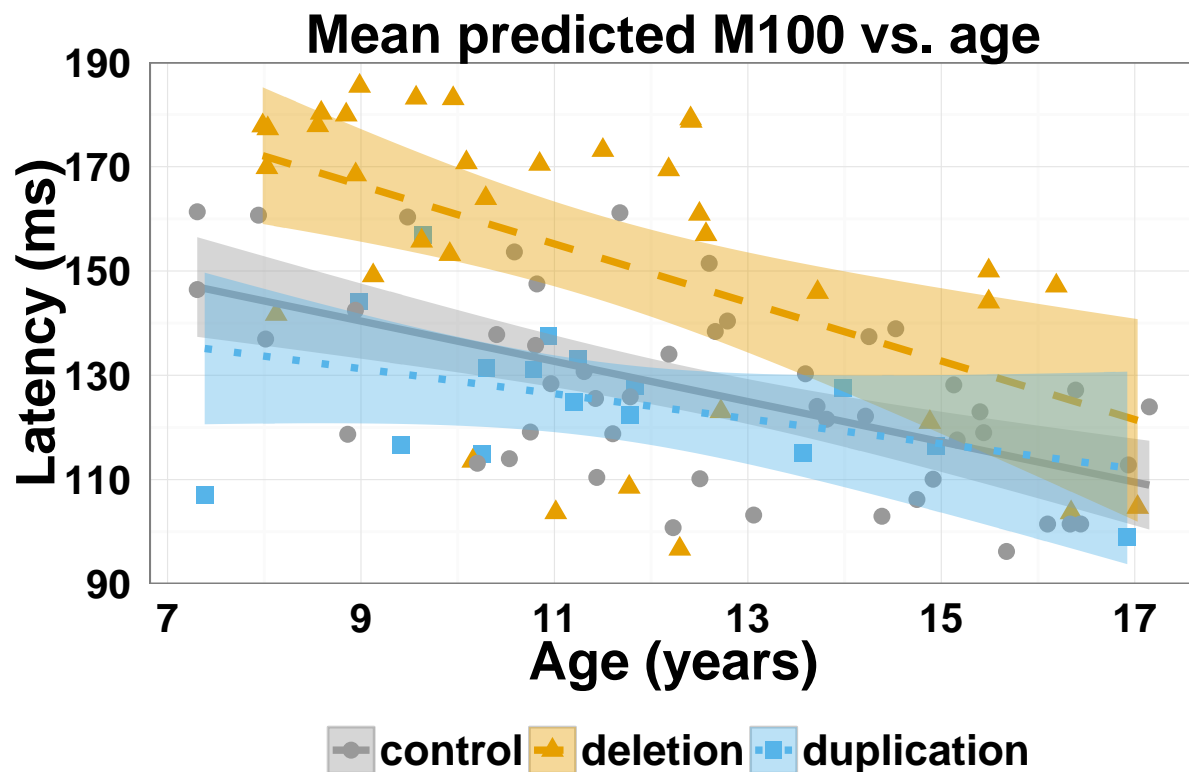
m100.lmm.2.fortify.summarize <-
  m100.lmm.2.fortify %>%
  group_by(Subject, Case, Age_Calc) %>%
  summarize(meanFitted = mean(.fitted), meanPredicted = mean(Predicted))

ggplot(m100.lmm.2.fortify.summarize,
       aes(y = meanPredicted, x = Age_Calc, shape = Case))+
  geom_point(size = 3, aes(colour = Case)) + theme_bw() +
```

```

labs(y = 'Latency (ms)', x = 'Age (years)',
     title = 'Mean predicted M100 vs. age',
     shape = 'Case', linetype = 'Case')+
geom_smooth(aes(linetype = Case, colour = Case, fill = Case),
            size = 1.3, method = 'lm', se = TRUE)+
scale_x_continuous(breaks = seq(5, 17, 2))+
scale_y_continuous(breaks = seq(90, 215, 20))+
coord_cartesian(ylim = c(90, 190))+
theme(legend.position = 'bottom',
      legend.text = element_text(face = 'bold', size = 16),
      legend.title = element_blank(),
      axis.text.x = element_text(size = 15, face = 'bold'),
      axis.text.y = element_text(size = 15, face = 'bold'),
      axis.ticks = element_blank(),
      axis.title.x = element_text(size = 20, face = 'bold'),
      axis.title.y = element_text(size = 20, face = 'bold'),
      plot.title = element_text(size = 20, face = 'bold')) +
scale_fill_manual(values = cbPalette) +
scale_colour_manual(values = cbPalette) +
scale_linetype_manual(values = c('solid', 'dashed', 'dotted'))

```



```

# overall model between latency and age
m100.age.dist.case.aov <-
  aov(meanPredicted ~ Age_Calc * Case, data = m100.lmm.2.fortify.summarize)

summary.lm(m100.age.dist.case.aov)

```

```
##
```

```
## Call:
## aov(formula = meanPredicted ~ Age_Calc * Case, data = m100.lmm.2.fortify.summarize)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -51.426 -11.795   2.298  13.478  31.924
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    175.1262    12.7816   13.701 < 2e-16 ***
## Age_Calc       -3.8589     0.9934   -3.885 0.000192 ***
## Casedeletion    41.7952    18.7085    2.234 0.027878 *
## Caseduplication -22.2545    26.0828   -0.853 0.395726
## Age_Calc:Casedeletion -1.7530    1.5395   -1.139 0.257754
## Age_Calc:Caseduplication  1.4583    2.1848    0.667 0.506120
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.15 on 93 degrees of freedom
## Multiple R-squared:  0.4921, Adjusted R-squared:  0.4647
## F-statistic: 18.02 on 5 and 93 DF,  p-value: 1.851e-12
```

```
summary.aov(m100.age.dist.case.aov)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Age_Calc      1  16832   16832  51.111 1.94e-10 ***
## Case          2  12049    6024  18.293 2.00e-07 ***
## Age_Calc:Case  2     789     394   1.197   0.307
## Residuals     93  30628     329
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

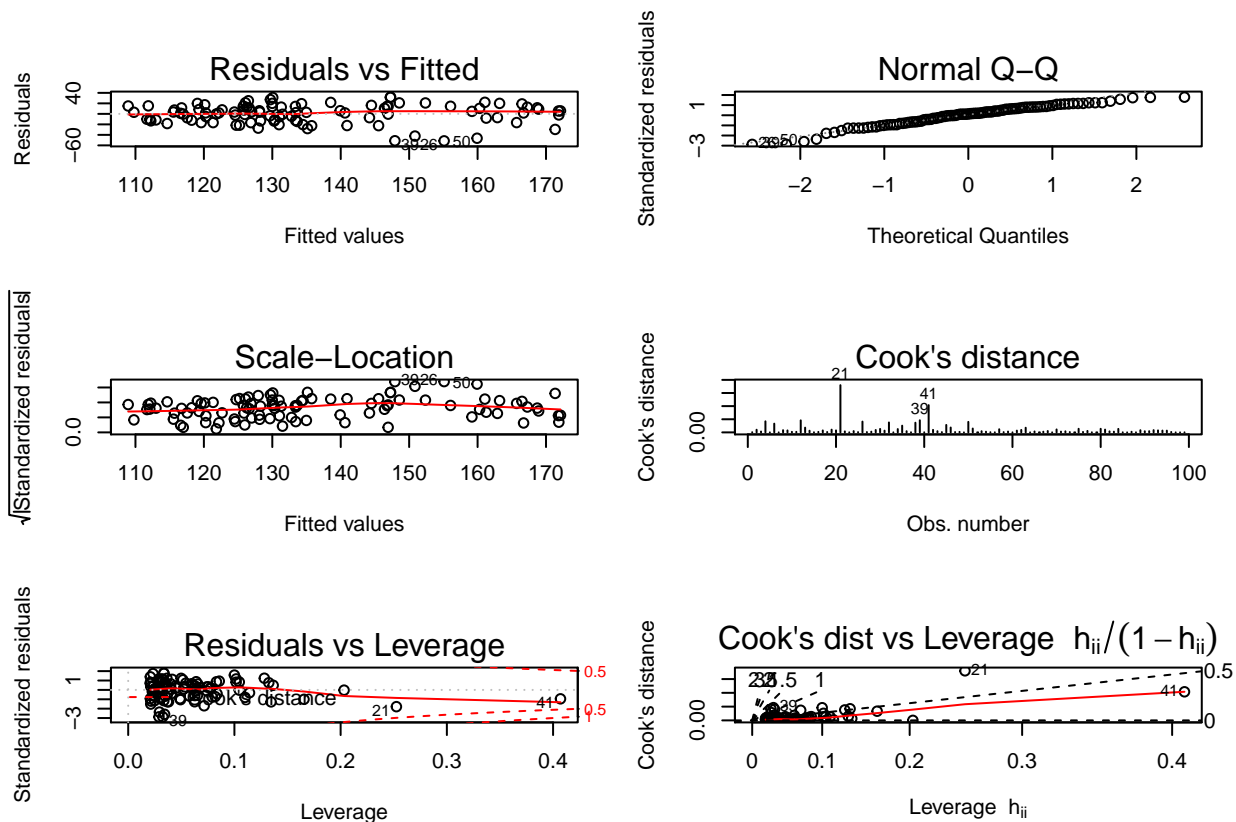
```
# slope for each case level
by(m100.lmm.2.fortify.summarize,
   m100.lmm.2.fortify.summarize[, 'Case'],
   function(x) summary(lm(meanPredicted ~ Age_Calc, data = x)))
```

```
## Case: control
##
## Call:
## lm(formula = meanPredicted ~ Age_Calc, data = x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -27.16 -11.53   1.41  12.64  31.10
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    175.1262    10.3359   16.943 < 2e-16 ***
## Age_Calc       -3.8589     0.8033   -4.804 1.69e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 14.68 on 46 degrees of freedom
## Multiple R-squared:  0.3341, Adjusted R-squared:  0.3196
## F-statistic: 23.08 on 1 and 46 DF,  p-value: 1.692e-05
##
## -----
## Case: deletion
##
## Call:
## lm(formula = meanPredicted ~ Age_Calc, data = x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -51.426 -14.502   5.997  16.700  31.924
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   216.921     17.667   12.28 7.5e-14 ***
## Age_Calc       -5.612       1.521   -3.69 0.000804 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.47 on 33 degrees of freedom
## Multiple R-squared:  0.2921, Adjusted R-squared:  0.2706
## F-statistic: 13.62 on 1 and 33 DF,  p-value: 0.0008036
##
## -----
## Case: duplication
##
## Call:
## lm(formula = meanPredicted ~ Age_Calc, data = x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28.047  -7.226   1.427   7.554  27.121
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   152.872     16.899   9.046 3.19e-07 ***
## Age_Calc       -2.401       1.446  -1.660   0.119
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.49 on 14 degrees of freedom
## Multiple R-squared:  0.1644, Adjusted R-squared:  0.1047
## F-statistic: 2.755 on 1 and 14 DF,  p-value: 0.1192
```

```
par(mfrow = c(3, 2))
```

```
plot(m100.age.dist.case.aov, which = c(1:6))
```



analysis of deviance -----

```
m100.complete <- glm(M100compCase ~ Case, binomial, data = child.study.data)
```

```
summary(m100.complete)
```

```
##
## Call:
## glm(formula = M100compCase ~ Case, family = binomial, data = child.study.data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6068  -1.1907   0.8020   0.8569   1.1642
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.8128    0.1106   7.349   2e-13 ***
## Casedeleletion  0.1566    0.1736   0.902  0.367132
## Caseduplication -0.7816    0.2085  -3.748 0.000178 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 999.72  on 791  degrees of freedom
## Residual deviance: 980.58  on 789  degrees of freedom
## AIC: 986.58
```

```
##
## Number of Fisher Scoring iterations: 4

# summary and observed proportions

m100.complete.summary <-
  child.study.data %>%
  group_by(Case) %>%
  summarize(counts = sum(M100compCase),
            total = sum(complete.cases(M100compCase)), pct = counts / total)

m100.complete.summary
```

```
## Source: local data frame [3 x 4]
##
##      Case counts total      pct
## 1   control      266    384 0.6927083
## 2  deletion      203    280 0.7250000
## 3 duplication      65    128 0.5078125
```

```
# psychophysical discrimination -----

source.dir <- '~/GitHub/reproducible-research/Day-2/datasets'
exact.path = 'psycho-data-april-2015.csv'

vowel.data <- read.csv(file.path(source.dir, exact.path), header = TRUE)

head(vowel.data)
```

```
##   SubjectAssgn  GenID ExpAssgn Language Training TrainType  Sex
## 1 a-i-subject-1 Subject1      a-i  english      yes    formal female
## 2 a-i-subject-1 Subject1      a-i  english      yes    formal female
## 3 a-i-subject-1 Subject1      a-i  english      yes    formal female
## 4 a-i-subject-1 Subject1      a-i  english      yes    formal female
## 5 a-i-subject-1 Subject1      a-i  english      yes    formal female
## 6 a-i-subject-1 Subject1      a-i  english      yes    formal female
##   Instrument InstYears InstPlay EarTrain EarTrainYear Vowel VowelHeight
## 1      piano         12   current      no           n/a      a          low
## 2      piano         12   current      no           n/a      a          low
## 3      piano         12   current      no           n/a      a          low
## 4      piano         12   current      no           n/a      a          low
## 5      piano         12   current      no           n/a      a          low
## 6      piano         12   current      no           n/a      a          low
##   VowelPosition  F1   F2   F3 F1.F2 F1.F3 F2.F3 BarkF2.F1 BarkF3.F1
## 1      back 768 1333 2522 0.576 0.305 0.529    3.306    7.539
## 2      back 768 1333 2522 0.576 0.305 0.529    3.306    7.539
## 3      back 768 1333 2522 0.576 0.305 0.529    3.306    7.539
## 4      back 768 1333 2522 0.576 0.305 0.529    3.306    7.539
## 5      back 768 1333 2522 0.576 0.305 0.529    3.306    7.539
## 6      back 768 1333 2522 0.576 0.305 0.529    3.306    7.539
##   BarkF3.F2 Item ItemType SCGDiff logSCGDiff      RT RespNum RespNum2
## 1    4.233 a-2-2      same        0          0 684.71        1    100
```

```
## 2      4.233 a-2-2      same      0      0 707.26      1      100
## 3      4.233 a-2-2      same      0      0 443.42      1      100
## 4      4.233 a-2-2      same      0      0 561.19      1      100
## 5      4.233 a-2-2      same      0      0 534.67      1      100
## 6      4.233 a-2-2      same      0      0 514.24      1      100
##      absRT logAbsRT ItemPair
## 1 684.71 2.835507      2-2
## 2 707.26 2.849579      2-2
## 3 443.42 2.646815      2-2
## 4 561.19 2.749110      2-2
## 5 534.67 2.728086      2-2
## 6 514.24 2.711166      2-2
```

see help(family) for more options

```
vowel.glm <- glm(RespNum ~ Vowel, binomial, data = vowel.data)
vowel.probit.glm <- glm(RespNum ~ Vowel, binomial(probit), data = vowel.data)
vowel.cauchit.glm <- glm(RespNum ~ Vowel, binomial(cauchit), data = vowel.data)

summary(vowel.glm)
```

```
##
## Call:
## glm(formula = RespNum ~ Vowel, family = binomial, data = vowel.data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8011  -1.2060   0.6634   1.1491   1.1491
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.06690    0.02501   2.675  0.00748 **
## Voweli       1.33509    0.05096  26.197 < 2e-16 ***
## Vowelu       0.84983    0.04645  18.297 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 16763  on 12799  degrees of freedom
## Residual deviance: 15874  on 12797  degrees of freedom
## AIC: 15880
##
## Number of Fisher Scoring iterations: 4
```

```
summary(vowel.probit.glm)
```

```
##
## Call:
## glm(formula = RespNum ~ Vowel, family = binomial(probit), data = vowel.data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8011  -1.2060   0.6634   1.1491   1.1491
```

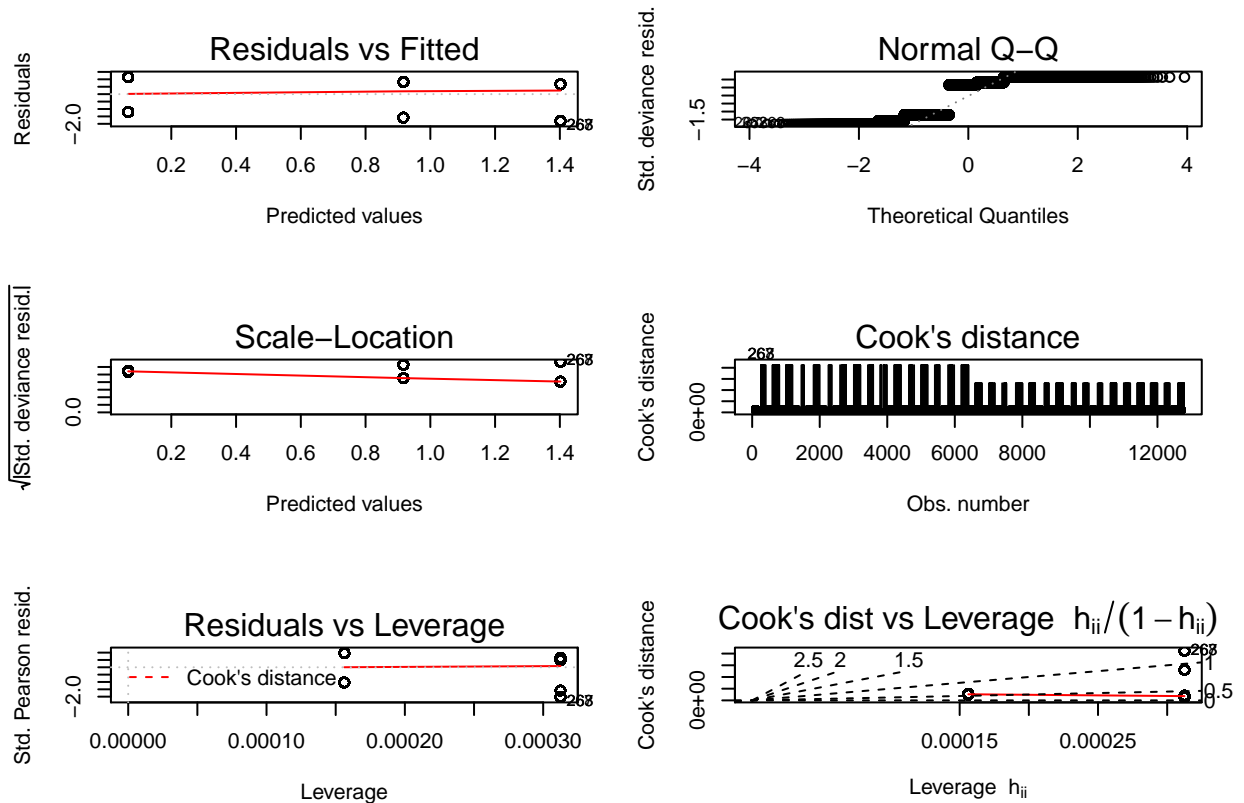
```
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.04192    0.01567   2.675  0.00747 **
## Voweli       0.80866    0.02979  27.150 < 2e-16 ***
## Vowelu       0.52429    0.02824  18.564 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 16763  on 12799  degrees of freedom
## Residual deviance: 15874  on 12797  degrees of freedom
## AIC: 15880
##
## Number of Fisher Scoring iterations: 4
```

```
summary(vowel.cauchit.glm)
```

```
##
## Call:
## glm(formula = RespNum ~ Vowel, family = binomial(cauchit), data = vowel.data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8011  -1.2060   0.6634   1.1491   1.1491
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.05257    0.01968   2.672  0.00755 **
## Voweli       1.34679    0.06830  19.718 < 2e-16 ***
## Vowelu       0.74536    0.04553  16.371 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 16763  on 12799  degrees of freedom
## Residual deviance: 15874  on 12797  degrees of freedom
## AIC: 15880
##
## Number of Fisher Scoring iterations: 5
```

```
# plot GLM; notice what is different between GLM and least-squares
par(mfrow = c(3, 2))

plot(vowel.glm, which = c(1:6))
```

```
# significance
Anova(vowel.glm)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: RespNum
##      LR Chisq Df Pr(>Chisq)
## Vowel   889.05  2 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(glht(vowel.glm, linfct = mcp(Vowel = 'Tukey')))
```

```
##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
## Fit: glm(formula = RespNum ~ Vowel, family = binomial, data = vowel.data)
##
## Linear Hypotheses:
##      Estimate Std. Error z value Pr(>|z|)
## i - a == 0  1.33509    0.05096  26.197 <1e-10 ***
## u - a == 0  0.84983    0.04645  18.297 <1e-10 ***
## u - i == 0 -0.48526    0.05919  -8.199 <1e-10 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```
# repeat with GLMM
```

```
vowel.glmm <-
  glmer(RespNum ~ Vowel + (Vowel | SubjectAssgn), binomial, data = vowel.data)
```

```
# notice that what works in the GLM case may not work in the GLMM case;
# random effects terms may have to be adjusted (hint: try (1 | SubjectAssgn))
```

```
vowel.probit.glmm <-
  glmer(RespNum ~ Vowel + (Vowel | SubjectAssgn),
        binomial(probit), data = vowel.data)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge: degenerate Hessian with 1 negative
## eigenvalues
```

```
vowel.cauchit.glmm <-
  glmer(RespNum ~ Vowel + (Vowel | SubjectAssgn),
        binomial(cauchit), data = vowel.data)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge: degenerate Hessian with 1 negative
## eigenvalues
```

```
summary(vowel.glmm)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: RespNum ~ Vowel + (Vowel | SubjectAssgn)
## Data: vowel.data
##
##      AIC      BIC   logLik deviance df.resid
## 15334.3 15401.5 -7658.2 15316.3    12791
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.7664 -0.9146  0.4424  0.7132  1.2721
##
## Random effects:
##  Groups      Name      Variance Std.Dev. Corr
## SubjectAssgn (Intercept) 0.1884  0.4341
##              Vowel1     0.1579  0.3973  -0.35
##              Vowel2     0.2064  0.4543   0.29 -0.53
## Number of obs: 12800, groups: SubjectAssgn, 32
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.07378    0.08088   0.912    0.362
## Vowel1      1.39440    0.11148  12.508 < 2e-16 ***
```

```

## Vowelu      0.93826    0.12576    7.461 8.62e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr) Voweli
## Voweli -0.279
## Vowelu  0.114 -0.032

# reaction time analysis

# remove timed-out responses
vowel.data.late.removed <- subset(vowel.data, absRT < 2000)
vowel.data.late.removed <- droplevels(vowel.data.late.removed)

# implement least-squares model for log RT
# three ways to do this
vowel.lm <- lm(logAbsRT ~ Vowel, data = vowel.data.late.removed)

vowel.lm.2 <-
  lm(logAbsRT ~ Vowel, data = vowel.data, subset = absRT < 2000)

vowel.lm.3 <-
  lm(log10(absRT) ~ Vowel, data = vowel.data, subset = absRT < 2000)

# linear mixed effects model for log RT

vowel.lmm <-
  lmer(logAbsRT ~ Vowel + (Vowel | SubjectAssgn),
    data = vowel.data.late.removed, REML = FALSE)

vowel.lmm.2 <-
  lmer(logAbsRT ~ Vowel + (Vowel | SubjectAssgn),
    data = vowel.data, subset = absRT < 2000, REML = FALSE)

vowel.lmm.3 <-
  lmer(log10(absRT) ~ Vowel + (Vowel | SubjectAssgn),
    data = vowel.data, subset = absRT < 2000, REML = FALSE)

# making more complicated models:

# no random slopes to avoid convergence issues
vowel.item.lmm <-
  lmer(log10(absRT) ~ Vowel * ItemPair + (1 | SubjectAssgn),
    data = vowel.data, subset = absRT < 2000, REML = FALSE)

vowel.item.glmm <-
  glmer(RespNum ~ ItemPair + (1 | SubjectAssgn),
    binomial, data = vowel.data, subset = absRT < 2000)

# significance
Anova(vowel.item.lmm)

## Analysis of Deviance Table (Type II Wald chisquare tests)

```

```
##
## Response: log10(absRT)
##           Chisq Df Pr(>Chisq)
## Vowel      4.6006 2    0.1002
## ItemPair    53.6165 9  2.233e-08 ***
## Vowel:ItemPair 136.0560 18 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova(vowel.item.glmm)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: RespNum
##           Chisq Df Pr(>Chisq)
## ItemPair 2592.5  9  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(glht(vowel.item.lmm,
             linfct = mcp(Vowel = 'Tukey', interaction_average = TRUE)))
```

```
##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmer(formula = log10(absRT) ~ Vowel * ItemPair + (1 | SubjectAssgn),
## data = vowel.data, REML = FALSE, subset = absRT < 2000)
##
## Linear Hypotheses:
##           Estimate Std. Error z value Pr(>|z|)
## i - a == 0 -0.0077655  0.0036520 -2.126  0.0794 .
## u - a == 0 -0.0004005  0.0036271 -0.110  0.9929
## u - i == 0  0.0073651  0.0051331  1.435  0.3092
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```
summary(glht(vowel.item.lmm,
             linfct = mcp(ItemPair = 'Tukey', interaction_average = TRUE)))
```

```
## Warning in RET$pfunction("adjusted", ...): Completion with error > abseps
## Warning in RET$pfunction("adjusted", ...): Completion with error > abseps
## Warning in RET$pfunction("adjusted", ...): Completion with error > abseps
## Warning in RET$pfunction("adjusted", ...): Completion with error > abseps
```

```

## Warning in RET$pffunction("adjusted", ...): Completion with error > abseps

## Warning in RET$pffunction("adjusted", ...): Completion with error > abseps

## Warning in RET$pffunction("adjusted", ...): Completion with error > abseps

## Warning in RET$pffunction("adjusted", ...): Completion with error > abseps

## Warning in RET$pffunction("adjusted", ...): Completion with error > abseps

## Warning in RET$pffunction("adjusted", ...): Completion with error > abseps

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmer(formula = log10(absRT) ~ Vowel * ItemPair + (1 | SubjectAssgn),
## data = vowel.data, REML = FALSE, subset = absRT < 2000)
##
## Linear Hypotheses:
##
## Estimate Std. Error z value Pr(>|z|)
## 2-square - 2-2 == 0 0.0184206 0.0060774 3.031 0.0738 .
## 23-2 - 2-2 == 0 0.0133986 0.0060680 2.208 0.4511
## 23-23 - 2-2 == 0 0.0150183 0.0060557 2.480 0.2798
## 23-4 - 2-2 == 0 0.0323296 0.0060978 5.302 <0.01 ***
## 23-square - 2-2 == 0 0.0182083 0.0060599 3.005 0.0789 .
## 4-2 - 2-2 == 0 0.0210645 0.0060779 3.466 0.0192 *
## 4-4 - 2-2 == 0 0.0123851 0.0060703 2.040 0.5703
## 4-square - 2-2 == 0 0.0209543 0.0060723 3.451 0.0201 *
## square-square - 2-2 == 0 0.0062432 0.0060616 1.030 0.9905
## 23-2 - 2-square == 0 -0.0050220 0.0060837 -0.825 0.9982
## 23-23 - 2-square == 0 -0.0034023 0.0060714 -0.560 0.9999
## 23-4 - 2-square == 0 0.0139090 0.0061136 2.275 0.4040
## 23-square - 2-square == 0 -0.0002122 0.0060761 -0.035 1.0000
## 4-2 - 2-square == 0 0.0026440 0.0060937 0.434 1.0000
## 4-4 - 2-square == 0 -0.0060355 0.0060861 -0.992 0.9928
## 4-square - 2-square == 0 0.0025337 0.0060879 0.416 1.0000
## square-square - 2-square == 0 -0.0121774 0.0060774 -2.004 0.5961
## 23-23 - 23-2 == 0 0.0016198 0.0060622 0.267 1.0000
## 23-4 - 23-2 == 0 0.0189310 0.0061043 3.101 0.0598 .
## 23-square - 23-2 == 0 0.0048098 0.0060664 0.793 0.9987
## 4-2 - 23-2 == 0 0.0076660 0.0060843 1.260 0.9622
## 4-4 - 23-2 == 0 -0.0010134 0.0060768 -0.167 1.0000
## 4-square - 23-2 == 0 0.0075558 0.0060786 1.243 0.9654
## square-square - 23-2 == 0 -0.0071554 0.0060679 -1.179 0.9756
## 23-4 - 23-23 == 0 0.0173113 0.0060919 2.842 0.1235
## 23-square - 23-23 == 0 0.0031900 0.0060543 0.527 1.0000
## 4-2 - 23-23 == 0 0.0060462 0.0060722 0.996 0.9926
## 4-4 - 23-23 == 0 -0.0026332 0.0060645 -0.434 1.0000
## 4-square - 23-23 == 0 0.0059360 0.0060663 0.979 0.9935

```

```
## square-square - 23-23 == 0      -0.0087751  0.0060558  -1.449   0.9116
## 23-square - 23-4 == 0           -0.0141212  0.0060963  -2.316   0.3781
## 4-2 - 23-4 == 0                 -0.0112650  0.0061141  -1.842   0.7079
## 4-4 - 23-4 == 0                 -0.0199445  0.0061066  -3.266   0.0363 *
## 4-square - 23-4 == 0            -0.0113752  0.0061084  -1.862   0.6944
## square-square - 23-4 == 0       -0.0260864  0.0060978  -4.278   <0.01 ***
## 4-2 - 23-square == 0            0.0028562  0.0060764   0.470   1.0000
## 4-4 - 23-square == 0            -0.0058232  0.0060688  -0.960   0.9944
## 4-square - 23-square == 0       0.0027460  0.0060707   0.452   1.0000
## square-square - 23-square == 0 -0.0119652  0.0060600  -1.974   0.6170
## 4-4 - 4-2 == 0                 -0.0086794  0.0060867  -1.426   0.9194
## 4-square - 4-2 == 0            -0.0001102  0.0060887  -0.018   1.0000
## square-square - 4-2 == 0       -0.0148214  0.0060779  -2.439   0.3021
## 4-square - 4-4 == 0            0.0085692  0.0060810   1.409   0.9248
## square-square - 4-4 == 0       -0.0061419  0.0060704  -1.012   0.9917
## square-square - 4-square == 0  -0.0147111  0.0060723  -2.423   0.3121
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```
summary(glht(vowel.item.glmm,
             linfct = mcp(ItemPair = 'Tukey', interaction_average = TRUE)))
```

```
## Warning in RET$pfunction("adjusted", ...): Completion with error > abseps
```

```
## Warning in RET$pfunction("adjusted", ...): Completion with error > abseps
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: glmer(formula = RespNum ~ ItemPair + (1 | SubjectAssgn), data = vowel.data,
##   family = binomial, subset = absRT < 2000)
##
## Linear Hypotheses:
##
##               Estimate Std. Error z value Pr(>|z|)
## 2-square - 2-2 == 0    -5.61454   0.17048 -32.934 <0.001 ***
## 23-2 - 2-2 == 0       -2.62096   0.15096 -17.362 <0.001 ***
## 23-23 - 2-2 == 0      -0.31014   0.18125  -1.711  0.7702
## 23-4 - 2-2 == 0       -3.85556   0.15187 -25.386 <0.001 ***
## 23-square - 2-2 == 0  -2.52559   0.15109 -16.715 <0.001 ***
## 4-2 - 2-2 == 0       -2.90344   0.15054 -19.287 <0.001 ***
## 4-4 - 2-2 == 0       -0.34032   0.18033  -1.887  0.6540
## 4-square - 2-2 == 0   -3.00976   0.15038 -20.015 <0.001 ***
## square-square - 2-2 == 0  0.19674   0.20115   0.978  0.9924
## 23-2 - 2-square == 0   2.99358   0.11733  25.514 <0.001 ***
## 23-23 - 2-square == 0  5.30440   0.15730  33.722 <0.001 ***
## 23-4 - 2-square == 0   1.75898   0.11645  15.105 <0.001 ***
## 23-square - 2-square == 0 3.08895   0.11778  26.226 <0.001 ***
## 4-2 - 2-square == 0   2.71110   0.11628  23.314 <0.001 ***
## 4-4 - 2-square == 0   5.27422   0.15630  33.745 <0.001 ***
```

```

## 4-square - 2-square == 0      2.60478    0.11590    22.474    <0.001 ***
## square-square - 2-square == 0  5.81128    0.18012    32.263    <0.001 ***
## 23-23 - 23-2 == 0            2.31083    0.13603    16.988    <0.001 ***
## 23-4 - 23-2 == 0            -1.23460    0.08898   -13.876    <0.001 ***
## 23-square - 23-2 == 0         0.09537    0.08907     1.071    0.9855
## 4-2 - 23-2 == 0              -0.28248    0.08756    -3.226    0.0365 *
## 4-4 - 23-2 == 0              2.28064    0.13487    16.910    <0.001 ***
## 4-square - 23-2 == 0         -0.38880    0.08719    -4.459    <0.001 ***
## square-square - 23-2 == 0     2.81770    0.16179    17.416    <0.001 ***
## 23-4 - 23-23 == 0           -3.54543    0.13697   -25.885    <0.001 ***
## 23-square - 23-23 == 0       -2.21545    0.13621   -16.265    <0.001 ***
## 4-2 - 23-23 == 0            -2.59330    0.13553   -19.135    <0.001 ***
## 4-4 - 23-23 == 0            -0.03018    0.16823    -0.179    1.0000
## 4-square - 23-23 == 0       -2.69962    0.13535   -19.946    <0.001 ***
## square-square - 23-23 == 0    0.50688    0.19031     2.663    0.1724
## 23-square - 23-4 == 0        1.32997    0.08949    14.861    <0.001 ***
## 4-2 - 23-4 == 0              0.95212    0.08775    10.851    <0.001 ***
## 4-4 - 23-4 == 0              3.51524    0.13582    25.881    <0.001 ***
## 4-square - 23-4 == 0         0.84580    0.08731     9.687    <0.001 ***
## square-square - 23-4 == 0     4.05230    0.16265    24.914    <0.001 ***
## 4-2 - 23-square == 0         -0.37785    0.08803    -4.293    <0.001 ***
## 4-4 - 23-square == 0         2.18527    0.13505    16.181    <0.001 ***
## 4-square - 23-square == 0     -0.48417    0.08766    -5.523    <0.001 ***
## square-square - 23-square == 0 2.72233    0.16193    16.812    <0.001 ***
## 4-4 - 4-2 == 0               2.56312    0.13436    19.076    <0.001 ***
## 4-square - 4-2 == 0          -0.10632    0.08605    -1.236    0.9621
## square-square - 4-2 == 0      3.10018    0.16138    19.211    <0.001 ***
## 4-square - 4-4 == 0          -2.66944    0.13418   -19.894    <0.001 ***
## square-square - 4-4 == 0      0.53706    0.18948     2.834    0.1130
## square-square - 4-square == 0 3.20650    0.16124    19.886    <0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)

```

```
# output from languageR
```

```

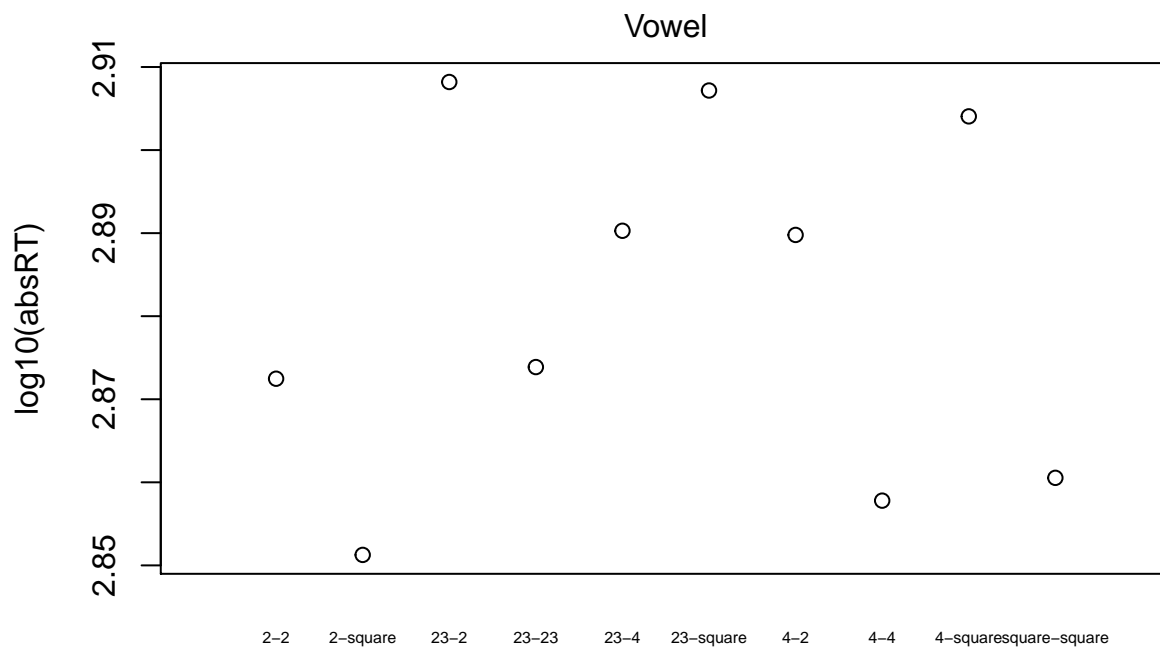
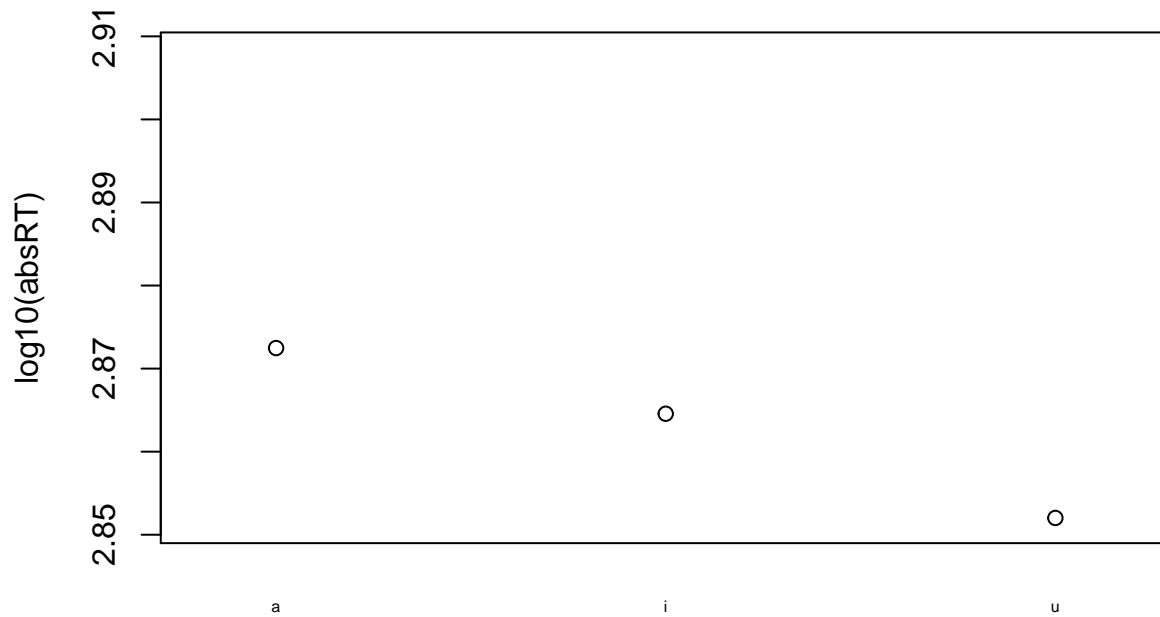
par(mfrow = c(1, 1))
plotLMER.fnc(vowel.item.lmm, withList = TRUE)

```

```

## effect size (range) for Vowel is 0.02045459
## effect size (range) for ItemPair is 0.0569331

```



ItemPair

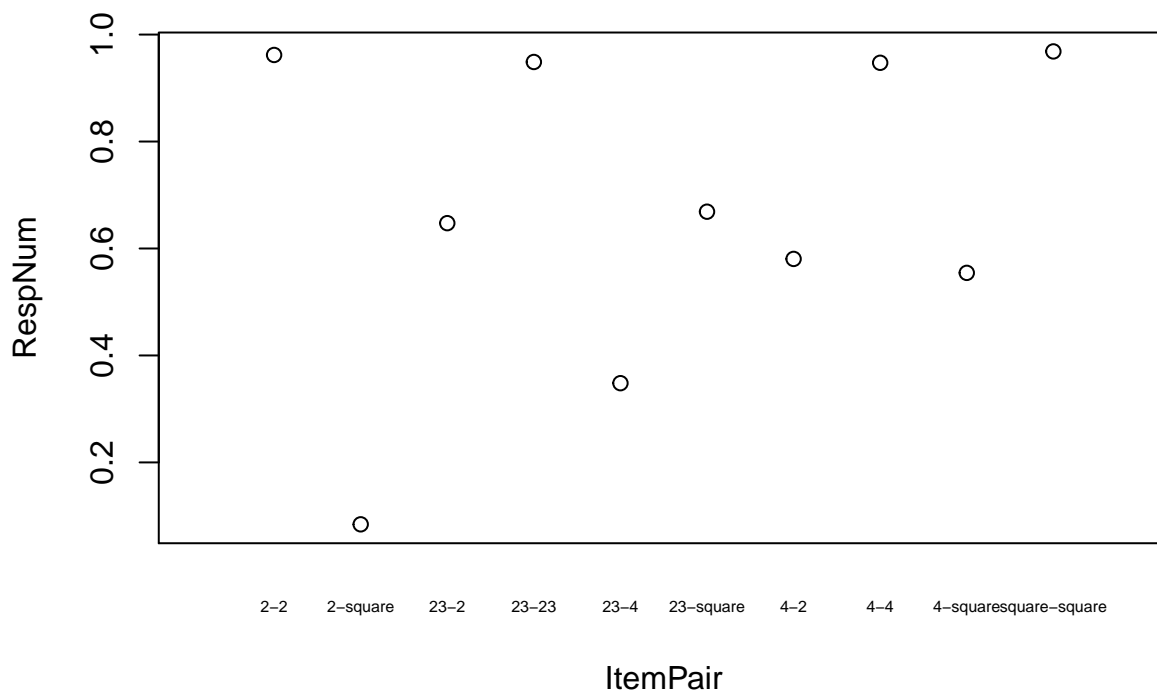
```
## $Vowel
##   X      Y Predictor Type Interaction Levels
## 1 1 2.872473   Vowel TRUE      FALSE    a
## 2 2 2.864566   Vowel TRUE      FALSE    i
## 3 3 2.852018   Vowel TRUE      FALSE    u
##
## $ItemPair
##   X      Y Predictor Type Interaction      Levels
## 1  1 2.872473  ItemPair TRUE      FALSE      2-2
## 2  2 2.851262  ItemPair TRUE      FALSE      2-square
```



```
## 3 3 2.908195 ItemPair TRUE FALSE 23-2
## 4 4 2.873876 ItemPair TRUE FALSE 23-23
## 5 5 2.890286 ItemPair TRUE FALSE 23-4
## 6 6 2.907172 ItemPair TRUE FALSE 23-square
## 7 7 2.889783 ItemPair TRUE FALSE 4-2
## 8 8 2.857793 ItemPair TRUE FALSE 4-4
## 9 9 2.904056 ItemPair TRUE FALSE 4-square
## 10 10 2.860537 ItemPair TRUE FALSE square-square
```

```
# notice that plogis() is set automatically
plotLMER.fnc(vowel.item.glmm, withList = TRUE)
```

```
## log odds are back-transformed to probabilities
## effect size (range) for ItemPair is 0.8842454
```



```
## $ItemPair
##      X      Y Predictor Type Interaction      Levels
## 1 1 0.96188913 ItemPair TRUE      FALSE      2-2
## 2 2 0.08423573 ItemPair TRUE      FALSE    2-square
## 3 3 0.64735624 ItemPair TRUE      FALSE      23-2
## 4 4 0.94874192 ItemPair TRUE      FALSE      23-23
## 5 5 0.34815394 ItemPair TRUE      FALSE      23-4
## 6 6 0.66881036 ItemPair TRUE      FALSE    23-square
## 7 7 0.58053260 ItemPair TRUE      FALSE      4-2
## 8 8 0.94725401 ItemPair TRUE      FALSE      4-4
## 9 9 0.55444313 ItemPair TRUE      FALSE    4-square
## 10 10 0.96848108 ItemPair TRUE      FALSE square-square
```

```
# construction of design matrices -----
```

```
subj.case.data <-
  m100.lmm.2.fortify %>%
  group_by(Subject, Case, Age_Calc) %>%
  summarize(meanPredicted = mean(Predicted))

design.mtx <- model.matrix(~ Case * Age_Calc, data = subj.case.data)

# notice the similarities to the Python example
design.mtx
```

##	(Intercept)	Casedeletion	Caseduplication	Age_Calc	Casedeletion:Age_Calc
## 1	1	0	1	10.784	0.000
## 2	1	1	0	12.721	12.721
## 3	1	1	0	10.847	10.847
## 4	1	1	0	16.342	16.342
## 5	1	1	0	8.945	8.945
## 6	1	1	0	17.030	17.030
## 7	1	0	1	14.948	0.000
## 8	1	0	1	10.247	0.000
## 9	1	1	0	9.126	9.126
## 10	1	1	0	10.090	10.090
## 11	1	1	0	12.567	12.567
## 12	1	0	1	9.644	0.000
## 13	1	1	0	12.414	12.414
## 14	1	0	1	13.986	0.000
## 15	1	0	1	11.836	0.000
## 16	1	0	1	11.247	0.000
## 17	1	1	0	12.181	12.181
## 18	1	1	0	7.984	7.984
## 19	1	1	0	8.986	8.986
## 20	1	1	0	14.882	14.882
## 21	1	0	1	7.384	0.000
## 22	1	1	0	12.499	12.499
## 23	1	1	0	8.027	8.027
## 24	1	1	0	8.847	8.847
## 25	1	1	0	13.721	13.721
## 26	1	1	0	11.014	11.014
## 27	1	1	0	9.918	9.918
## 28	1	1	0	8.553	8.553
## 29	1	1	0	9.570	9.570
## 30	1	0	1	9.416	0.000
## 31	1	1	0	11.501	11.501
## 32	1	1	0	16.189	16.189
## 33	1	0	1	11.784	0.000
## 34	1	1	0	15.490	15.490
## 35	1	1	0	15.490	15.490
## 36	1	0	1	10.937	0.000
## 37	1	0	1	10.296	0.000
## 38	1	1	0	8.123	8.123
## 39	1	1	0	12.296	12.296
## 40	1	1	0	10.290	10.290

## 41	1	0	1	16.923	0.000
## 42	1	1	0	9.953	9.953
## 43	1	1	0	8.589	8.589
## 44	1	1	0	9.627	9.627
## 45	1	1	0	11.773	11.773
## 46	1	1	0	12.408	12.408
## 47	1	1	0	8.036	8.036
## 48	1	0	1	13.570	0.000
## 49	1	0	1	11.203	0.000
## 50	1	1	0	10.156	10.156
## 51	1	0	1	8.984	0.000
## 52	1	0	0	13.814	0.000
## 53	1	0	0	14.386	0.000
## 54	1	0	0	8.014	0.000
## 55	1	0	0	13.595	0.000
## 56	1	0	0	11.425	0.000
## 57	1	0	0	11.677	0.000
## 58	1	0	0	14.915	0.000
## 59	1	0	0	10.819	0.000
## 60	1	0	0	11.784	0.000
## 61	1	0	0	14.529	0.000
## 62	1	0	0	15.674	0.000
## 63	1	0	0	8.863	0.000
## 64	1	0	0	13.715	0.000
## 65	1	0	0	10.751	0.000
## 66	1	0	0	11.307	0.000
## 67	1	0	0	14.748	0.000
## 68	1	0	0	13.060	0.000
## 69	1	0	0	10.403	0.000
## 70	1	0	0	17.151	0.000
## 71	1	0	0	12.504	0.000
## 72	1	0	0	16.332	0.000
## 73	1	0	0	15.400	0.000
## 74	1	0	0	15.164	0.000
## 75	1	0	0	7.940	0.000
## 76	1	0	0	11.603	0.000
## 77	1	0	0	11.438	0.000
## 78	1	0	0	12.184	0.000
## 79	1	0	0	7.307	0.000
## 80	1	0	0	7.307	0.000
## 81	1	0	0	10.203	0.000
## 82	1	0	0	16.441	0.000
## 83	1	0	0	15.129	0.000
## 84	1	0	0	9.482	0.000
## 85	1	0	0	16.934	0.000
## 86	1	0	0	10.805	0.000
## 87	1	0	0	12.789	0.000
## 88	1	0	0	15.438	0.000
## 89	1	0	0	12.227	0.000
## 90	1	0	0	16.099	0.000
## 91	1	0	0	14.252	0.000
## 92	1	0	0	16.389	0.000
## 93	1	0	0	10.534	0.000
## 94	1	0	0	12.600	0.000

## 95	1	0	0	10.586	0.000
## 96	1	0	0	12.663	0.000
## 97	1	0	0	14.216	0.000
## 98	1	0	0	8.942	0.000
## 99	1	0	0	10.964	0.000
##	Caseduplication:Age_Calc				
## 1				10.784	
## 2				0.000	
## 3				0.000	
## 4				0.000	
## 5				0.000	
## 6				0.000	
## 7				14.948	
## 8				10.247	
## 9				0.000	
## 10				0.000	
## 11				0.000	
## 12				9.644	
## 13				0.000	
## 14				13.986	
## 15				11.836	
## 16				11.247	
## 17				0.000	
## 18				0.000	
## 19				0.000	
## 20				0.000	
## 21				7.384	
## 22				0.000	
## 23				0.000	
## 24				0.000	
## 25				0.000	
## 26				0.000	
## 27				0.000	
## 28				0.000	
## 29				0.000	
## 30				9.416	
## 31				0.000	
## 32				0.000	
## 33				11.784	
## 34				0.000	
## 35				0.000	
## 36				10.937	
## 37				10.296	
## 38				0.000	
## 39				0.000	
## 40				0.000	
## 41				16.923	
## 42				0.000	
## 43				0.000	
## 44				0.000	
## 45				0.000	
## 46				0.000	
## 47				0.000	
## 48				13.570	

```

## 49          11.203
## 50          0.000
## 51          8.984
## 52          0.000
## 53          0.000
## 54          0.000
## 55          0.000
## 56          0.000
## 57          0.000
## 58          0.000
## 59          0.000
## 60          0.000
## 61          0.000
## 62          0.000
## 63          0.000
## 64          0.000
## 65          0.000
## 66          0.000
## 67          0.000
## 68          0.000
## 69          0.000
## 70          0.000
## 71          0.000
## 72          0.000
## 73          0.000
## 74          0.000
## 75          0.000
## 76          0.000
## 77          0.000
## 78          0.000
## 79          0.000
## 80          0.000
## 81          0.000
## 82          0.000
## 83          0.000
## 84          0.000
## 85          0.000
## 86          0.000
## 87          0.000
## 88          0.000
## 89          0.000
## 90          0.000
## 91          0.000
## 92          0.000
## 93          0.000
## 94          0.000
## 95          0.000
## 96          0.000
## 97          0.000
## 98          0.000
## 99          0.000
## attr("assign")
## [1] 0 1 1 2 3 3
## attr("contrasts")

```

```
## attr("contrasts")$Case  
## [1] "contr.treatment"
```

```
sink()
```