

Prediction of toddler vocabulary from infant speech and non-speech processing

Alejandrina Cristia

Main version 2020-07-30

History

- 2020-07-30: First version
- 2020-07-31: Final version of analyses
- 2020-08-04: Making sure that everything is in the same order as in the manuscript
- 2020-12-16: Integrating changes by YYW, adding table 6, checking for reproducibility

Read and preprocess data

```
## [1] "Subject.."                 "Gender"
## [3] "X18mCDIAge"                "X18mCom"
## [5] "X18mComAll"                 "X18mComGenderSpec"
## [7] "X18mSay"                   "X18mSayAll"
## [9] "X18mSayGenderSpec"          "X24mCDIAge"
## [11] "X24mCom"                   "X24msay"
## [13] "X24mSayALL"                "X24mSayGenderSpec"
## [15] "Mullen.VR"                 "MullenExL"
## [17] "stress...VRM.age"           "total.tro.pref.quotient"
## [19] "VRM.trials.completed"       "VRM.trials.coded"
## [21] "VRM.age"                   "Novelty.VRM"
## [23] "faces.VRM"                 "Shapes.VRM"
## [25] "vowels.Age"                "Vowel.Alt.pref.quotient"
## [27] "A.not.B.Age"               "A.not.B.score"
## [29] "notes1"                     "notes2"

## [1] 97 30

##   Subject..   Gender   X18mCDIAge      X18mCom      X18mComAll
##   Min.    :3425   F:43   Min.    :17.40   Min.    : 15.0   Min.    : 5.00
##   1st Qu.:3587   M:54   1st Qu.:17.69   1st Qu.:174.0  1st Qu.:20.45
##   Median  :3716                    Median :17.96   Median :232.0   Median :35.42
##   Mean    :3676                    Mean   :18.09   Mean   :231.1   Mean   :44.49
##   3rd Qu.:3771                    3rd Qu.:18.32   3rd Qu.:304.0  3rd Qu.:69.62
##   Max.    :3874                    Max.   :19.28   Max.   :395.0   Max.   :99.00
##   NA's    :32                      NA's   :32     NA's   :32     NA's   :32
##   X18mComGenderSpec   X18mSay      X18mSayAll   X18mSayGenderSpec
##   Min.    : 5.00   Min.    : 2.00   Min.    : 5.556   Min.    : 6.25
##   1st Qu.:21.11  1st Qu.: 21.00  1st Qu.:10.625  1st Qu.:11.75
##   Median  :33.06  Median  : 52.00  Median :18.571  Median :21.88
##   Mean    :44.49  Mean    : 70.17  Mean   :34.826  Mean   :35.43
```

```

## 3rd Qu.:71.67      3rd Qu.:105.00     3rd Qu.:60.000    3rd Qu.:56.25
## Max.   :99.00      Max.   :339.00     Max.   :97.894    Max.   :99.00
## NA's    :32         NA's    :32        NA's    :32       NA's    :32
## X24mCDIAge          X24mCom       X24msay      X24mSayALL
## Min.   :22.93      Min.   :50.0       Min.   : 4.0      Min.   : 5.00
## 1st Qu.:23.75      1st Qu.:350.5     1st Qu.:115.0    1st Qu.:17.96
## Median  :24.05      Median  :490.0      Median  :260.0    Median  :45.34
## Mean    :24.18      Mean    :468.9      Mean    :294.0    Mean    :47.54
## 3rd Qu.:24.42      3rd Qu.:586.8     3rd Qu.:473.5    3rd Qu.:78.50
## Max.   :29.24      Max.   :684.0      Max.   :653.0    Max.   :98.70
## NA's    :33         NA's    :33        NA's    :34       NA's    :34
## X24mSayGenderSpec   Mullen.VR       MullenExL     stress...VRM.age
## Min.   : 5.00       Min.   :32.00      Min.   :20.00    Min.   :5.000
## 1st Qu.:17.04       1st Qu.:46.00      1st Qu.:44.00    1st Qu.:5.330
## Median  :46.48       Median  :49.00      Median  :51.00    Median  :5.560
## Mean    :47.84       Mean    :53.38      Mean    :48.67    Mean    :5.675
## 3rd Qu.:82.05       3rd Qu.:60.50      3rd Qu.:57.00    3rd Qu.:6.020
## Max.   :99.00       Max.   :80.00      Max.   :70.00    Max.   :6.910
## NA's    :34         NA's    :71        NA's    :76       NA's    :2
## total.tro.pref.quotient VRM.trials.completed VRM.trials.coded   VRM.age
## Min.   :0.2338       Min.   :2.000      Min.   :2.000    Min.   :5.000
## 1st Qu.:0.4070       1st Qu.:7.000      1st Qu.:6.000    1st Qu.:5.315
## Median  :0.4885       Median  :9.000      Median  :9.000    Median  :5.560
## Mean    :0.5074       Mean    :7.404      Mean    :7.277    Mean    :5.678
## 3rd Qu.:0.6219       3rd Qu.:9.000      3rd Qu.:9.000    3rd Qu.:6.035
## Max.   :0.8202       Max.   :9.000      Max.   :9.000    Max.   :6.910
## NA's    :3           NA's    :3         NA's    :3       NA's    :2
## Novelty.VRM          faces.VRM       Shapes.VRM     vowels.Age
## Min.   :0.3486       Min.   :0.2802     Min.   :0.3886    Min.   :6.350
## 1st Qu.:0.5550       1st Qu.:0.5271     1st Qu.:0.5584    1st Qu.:6.680
## Median  :0.5914       Median  :0.5782     Median  :0.6126    Median  :6.810
## Mean    :0.5934       Mean    :0.5756     Mean    :0.6194    Mean    :6.918
## 3rd Qu.:0.6296       3rd Qu.:0.6286     3rd Qu.:0.6670    3rd Qu.:7.110
## Max.   :0.8349       Max.   :0.8883     Max.   :0.8938    Max.   :8.390
## NA's    :2           NA's    :3         NA's    :3       NA's    :2
## Vowel.Alt.pref.quotient A.not.B.Age   A.not.B.score
## Min.   :0.2576       Min.   :6.350      Min.   :1.000
## 1st Qu.:0.4585       1st Qu.:6.680      1st Qu.:1.000
## Median  :0.5083       Median  :6.810      Median  :2.000
## Mean    :0.5170       Mean    :6.938      Mean    :1.862
## 3rd Qu.:0.5786       3rd Qu.:7.110      3rd Qu.:2.000
## Max.   :0.7306       Max.   :8.060      Max.   :3.000
## NA's    :2           NA's    :3         NA's    :3
## notes1            notes2
## :90              :94
## incomplete infant data: 7  error=2: 1
## fail=1 : 1
## pass=3 : 1
##
## Subject..  Gender  Stress       VRM       VRM.age
## 3425   : 3   F:129  Min.   :0.2338  Min.   :0.3486  Min.   :5.000

```

```

##  3427   : 3   M:162   1st Qu.:0.4070   1st Qu.:0.5537   1st Qu.:5.300
##  3435   : 3           Median :0.4885   Median :0.5914   Median :5.560
##  3436   : 3           Mean   :0.5074   Mean   :0.5934   Mean   :5.678
##  3437   : 3           3rd Qu.:0.6219   3rd Qu.:0.6301   3rd Qu.:6.050
##  3438   : 3           Max.   :0.8202   Max.   :0.8349   Max.   :6.910
##  (Other):273          NA's    :6       NA's    :6
##      Vowel            anotB            vowels.Age     A.not.B.Age
##  Min.   :0.2576   Length:291        Min.   :6.350   Min.   :6.350
##  1st Qu.:0.4577   Class :character  1st Qu.:6.680   1st Qu.:6.680
##  Median :0.5083   Mode  :character  Median :6.810   Median :6.810
##  Mean   :0.5170           NA's    :6       Mean   :6.918   Mean   :6.938
##  3rd Qu.:0.5842           NA's    :6       3rd Qu.:7.110   3rd Qu.:7.110
##  Max.   :0.7306           NA's    :6       Max.   :8.390   Max.   :8.060
##  NA's    :6             NA's    :9
##      values            ind
##  Min.   :-2.2646  com18z:97
##  1st Qu.:-0.7612  say18z:97
##  Median : -0.1084 say24z:97
##  Mean   : 0.0000
##  3rd Qu.: 0.7315
##  Max.   : 4.0585
##  NA's    :99

```

Beginning of results section

Analyses were implemented in the R environment (R Core Team, 2014), and are available from supplementary materials. Table 1 presents the summary statistics for each measure.

Table 1

```

doline<-function(myage,myvec,prettyname,dod=F){
  #myage=mydat$vowels.Age
  #myvec=mydat$Vowel.Alt.pref.quotient
  #prettyname="Vowel"
  #dod=T
  cbind(prettyname,sum(!is.na(myvec)),
    paste0(round(mean(myage,na.rm=T),2)," (",round(sd(myage,na.rm=T),2),")"), #age(sd)
    paste0(round(mean(myvec,na.rm=T),2)," (",round(sd(myvec,na.rm=T),2),")"), #score(sd)
    ifelse(dod,round(cohensd(myvec),3), ""))
  )
}

tab1=rbind(
  doline(mydat$VRM.age,mydat$Stress,"Stress",T),
  doline(mydat$VRM.age,mydat$VRM,"VRM",T),
  doline(mydat$vowels.Age,mydat$Vowel,"Vowel",T),
  doline(mydat$A.not.B.Age,as.numeric(mydat$A.not.B.score),"A-not-B",F), !!!! it doesn't make sense to
  doline(mydat$X18mCDIAge,mydat$X18mCom,"18m receptive",F),
  doline(mydat$X18mCDIAge,mydat$X18mSay,"18m expressive",F),
  doline(mydat$X24mCDIAge,mydat$X24msay,"24m expressive",F)
)

```

Distribution checks

Prior to any analysis, we inspected the distributions of each measure using histograms and QQ plots. The distribution of the 3 continuous infant measures, namely, Stress, Vowel, and VRM, were approximately normal. However, the 3 vocabulary measures were not normally distributed. Therefore, we standardized the three outcome measures by applying z-scoring and then trimmed values more than 3 SD ($N = 1$) to improve the residual distribution, which yielded an approximately normal distribution of the outcome measures.

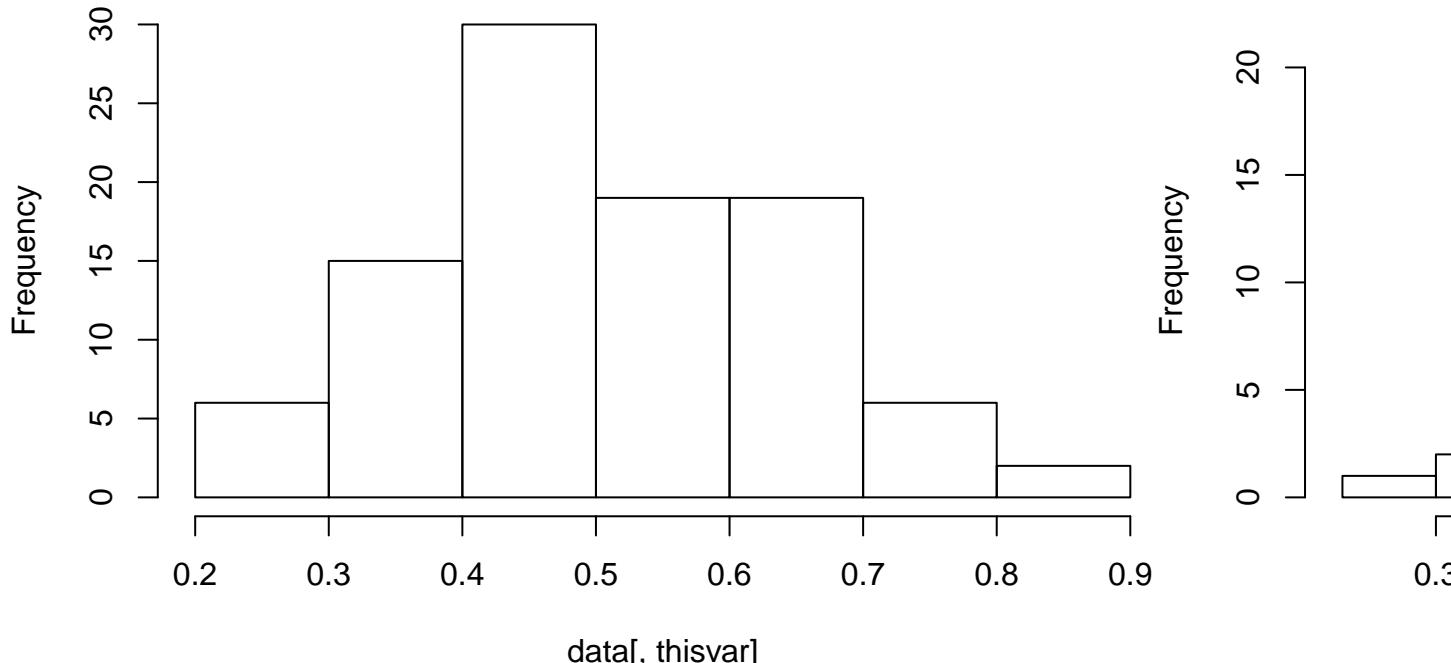
```
#outcomes:
outcomes = c("X18mCom", "X18mComGenderSpec", "X18mSay", "X18mSayGenderSpec", "X24mCom", "X24msay", "X24mSayGenderSpec")
#predictions:
predictors = c("total.tro.pref.quotient", "Vowel.Alt.pref.quotient", "Novelty.VRM")
# "A.not.B.score" is categorical

table(data$A.not.B.score)

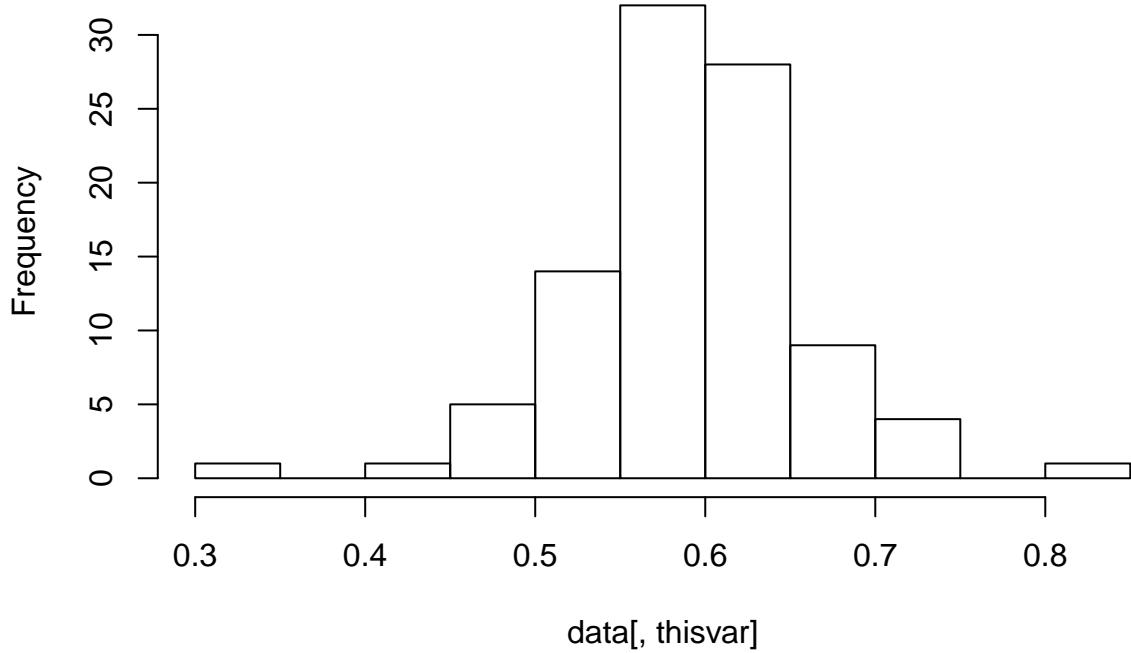
##
##   1   2   3
## 36 35 23

#outcomes
for(thisvar in predictors) hist(data[,thisvar],main=thisvar)
```

total.tro.pref.quotient

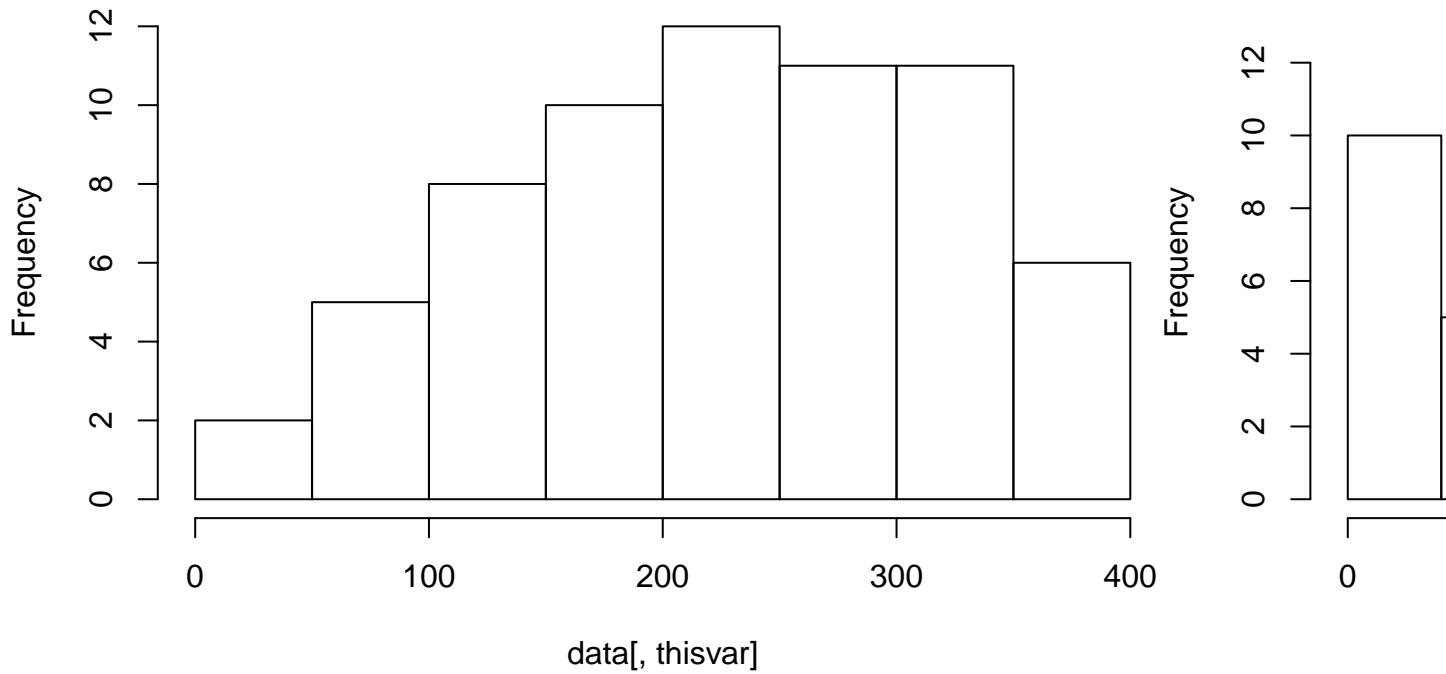


Novelty.VRM

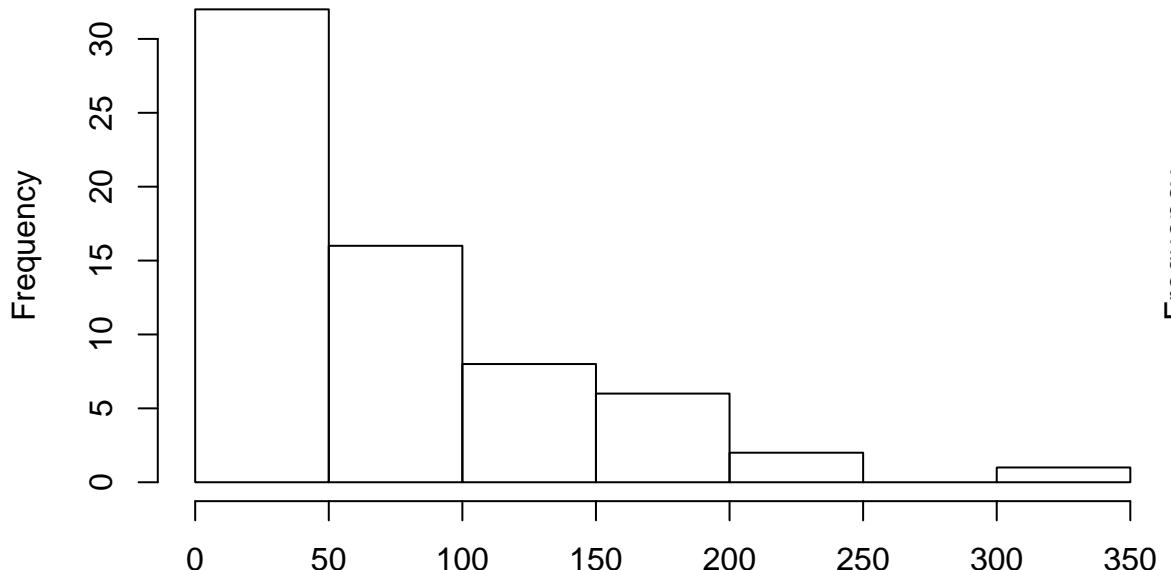


```
#outcomes - separated  
for(thisvar in outcomes) hist(data[,thisvar],main=thisvar)
```

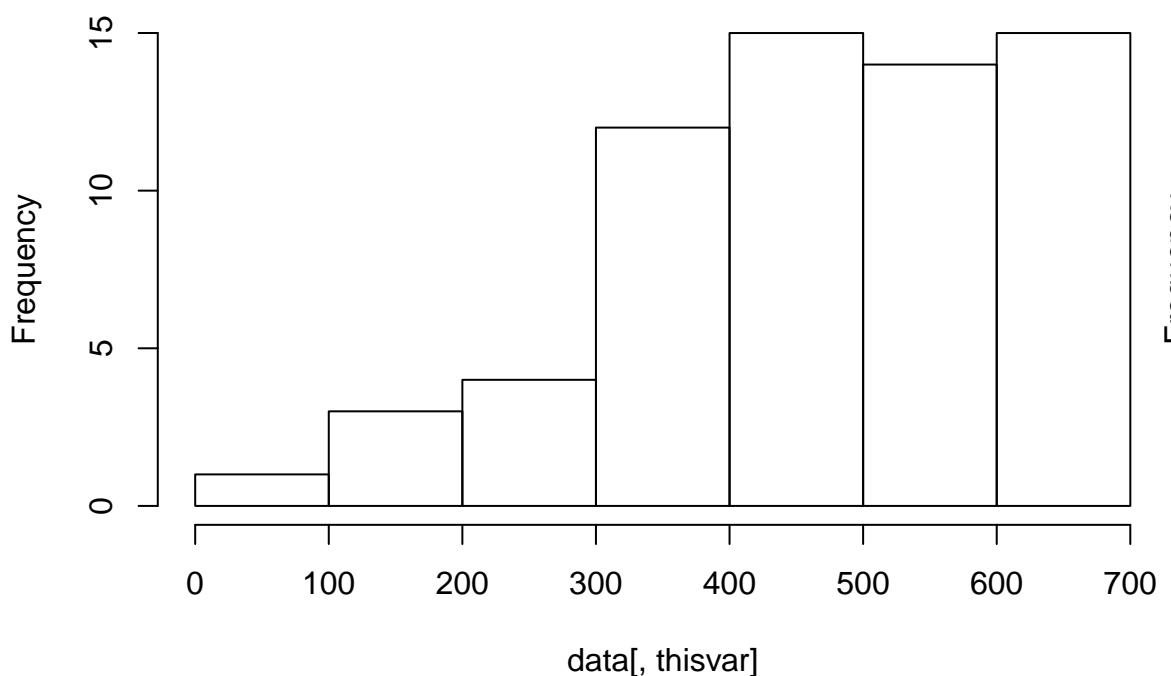
X18mCom



X18mSay

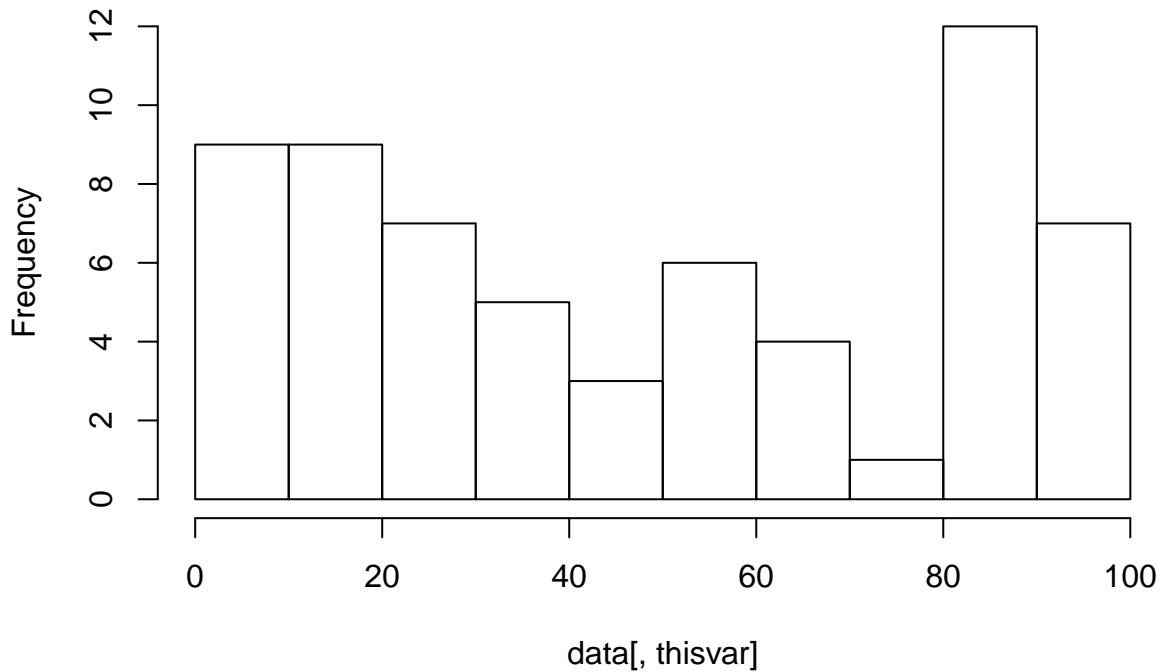


data[, thisvar]
X24mCom



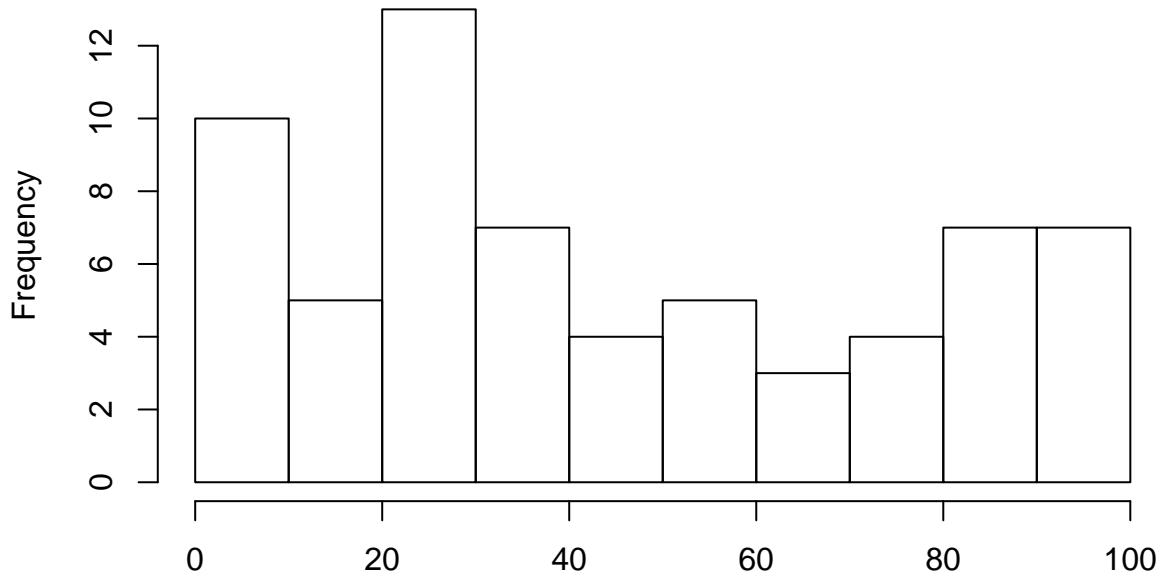
data[, thisvar]

X24mSayGenderSpec



```
stdat=cbind( data[,c("Subject..","total.tro.pref.quotient", "Vowel.Alt.pref.quotient",
  "Novelty.VRM", "A.not.B.score")], stack(data[,outcomes]))  
  
## Warning in data.frame(..., check.names = FALSE): row names were found from a  
## short variable and have been discarded  
  
#outcomes - combined - percentiles  
hist(stdat$values[stdat$ind %in% c('X18mComGenderSpec')], main='X18mComGenderSpec')
```

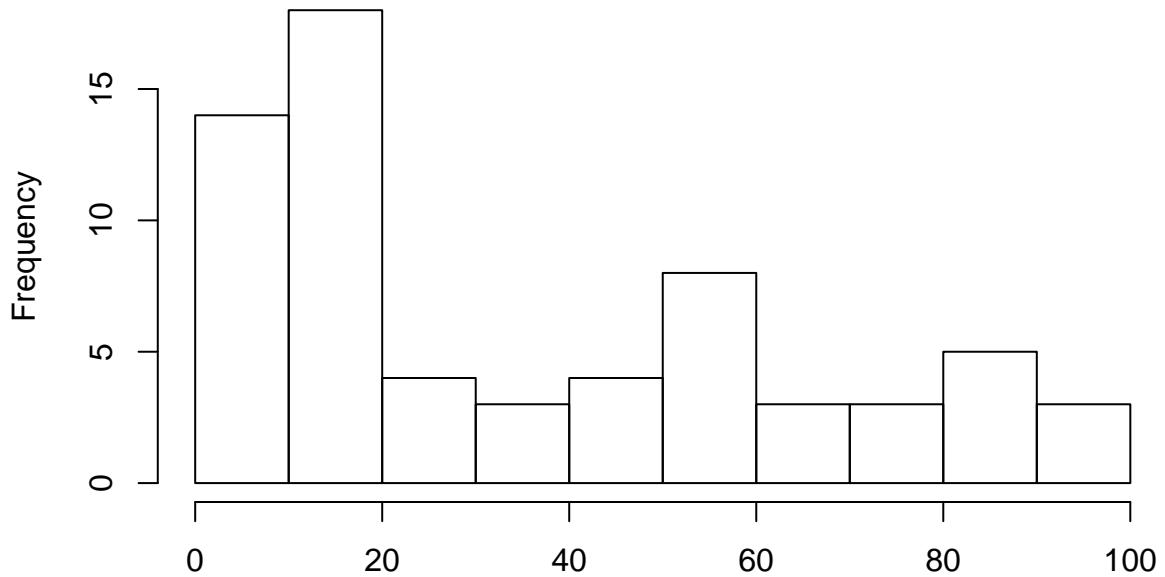
X18mComGenderSpec



```
stdat$values[stdat$ind %in% c("X18mComGenderSpec")]
```

```
hist(stdat$values[stdat$ind %in% c('X18mSayGenderSpec')], main='X18mSayGenderSpec')
```

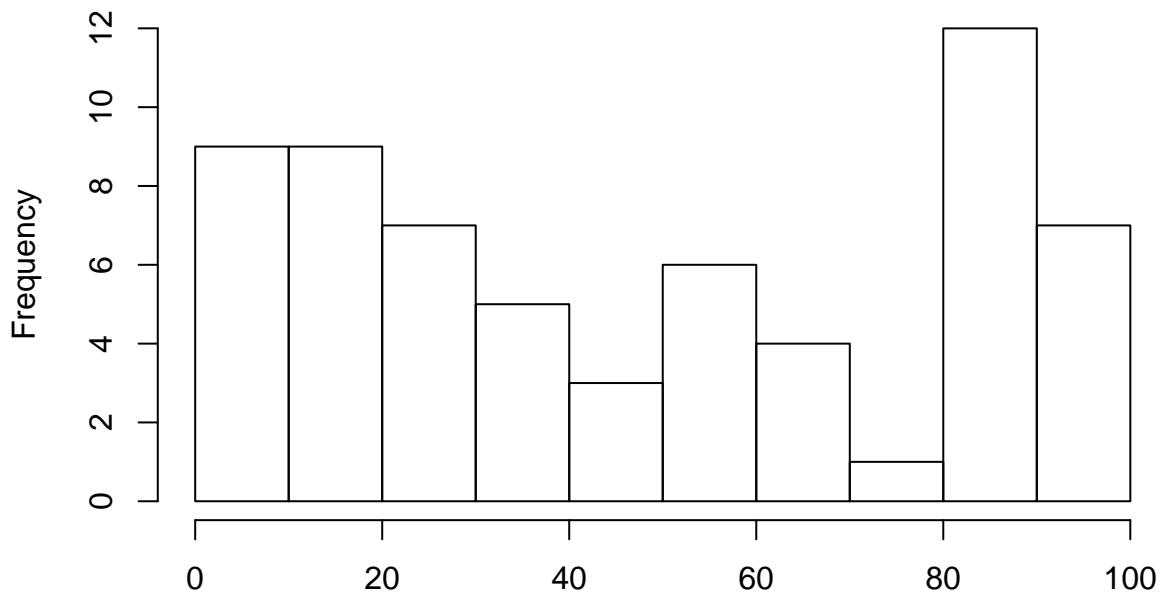
X18mSayGenderSpec



```
stdat$values[stdat$ind %in% c("X18mSayGenderSpec")]
```

```
hist(stdat$values[stdat$ind %in% c('X24mSayGenderSpec')], main='X24mSayGenderSpec')
```

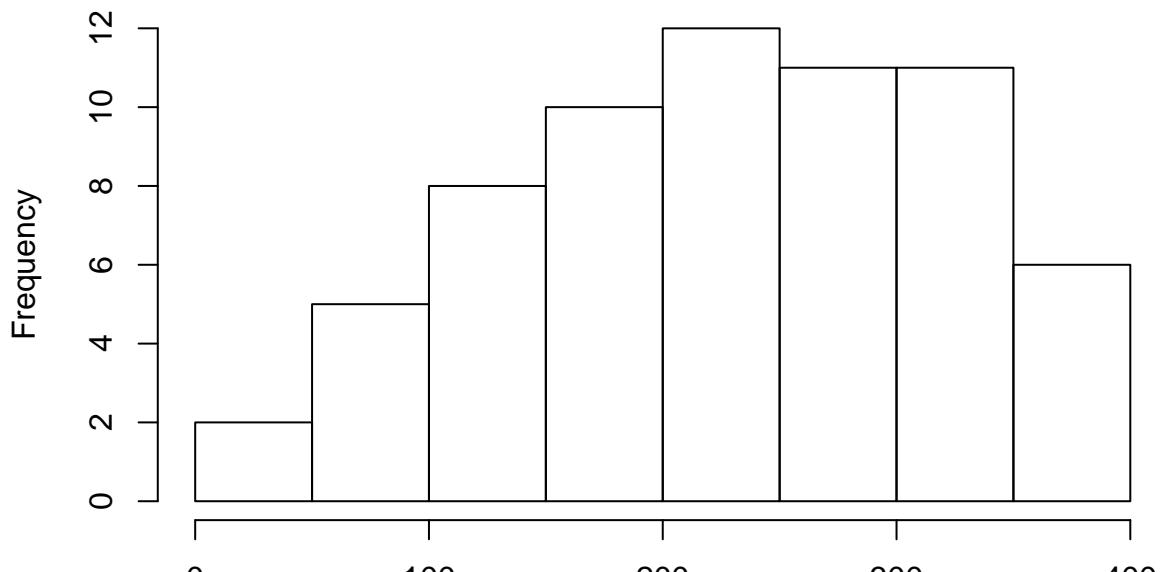
X24mSayGenderSpec



```
stdat$values[stdat$ind %in% c("X24mSayGenderSpec")]
```

```
#outcomes - combined - raw  
hist(stdat$values[stdat$ind %in% c('X18mCom')], main='X18mCom')
```

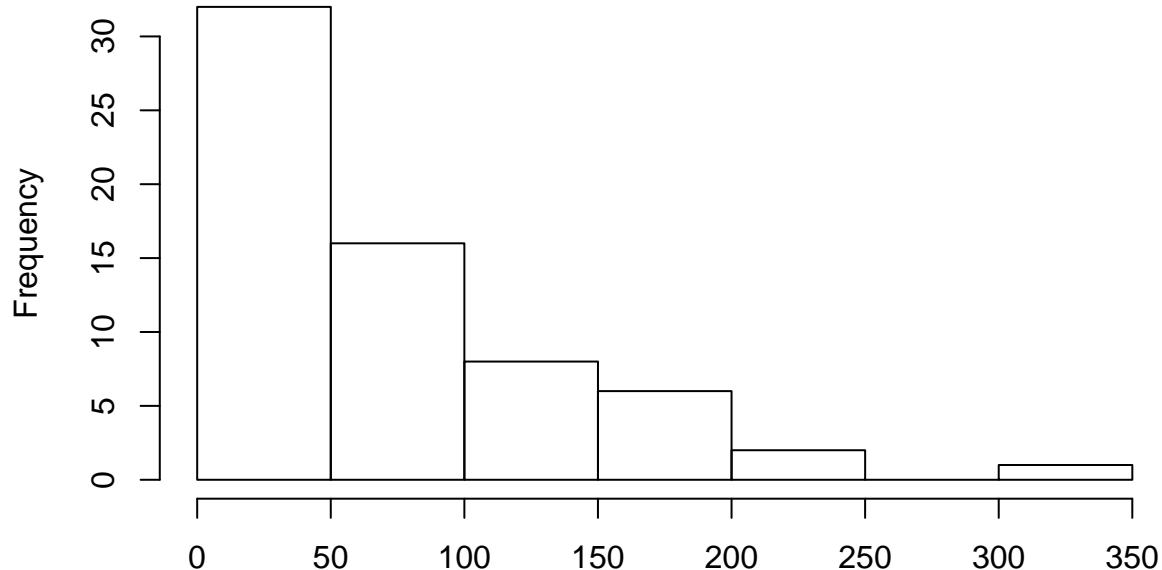
X18mCom



```
stdat$values[stdat$ind %in% c("X18mCom")]
```

```
hist(stdat$values[stdat$ind %in% c('X18mSay')], main='X18mSay')
```

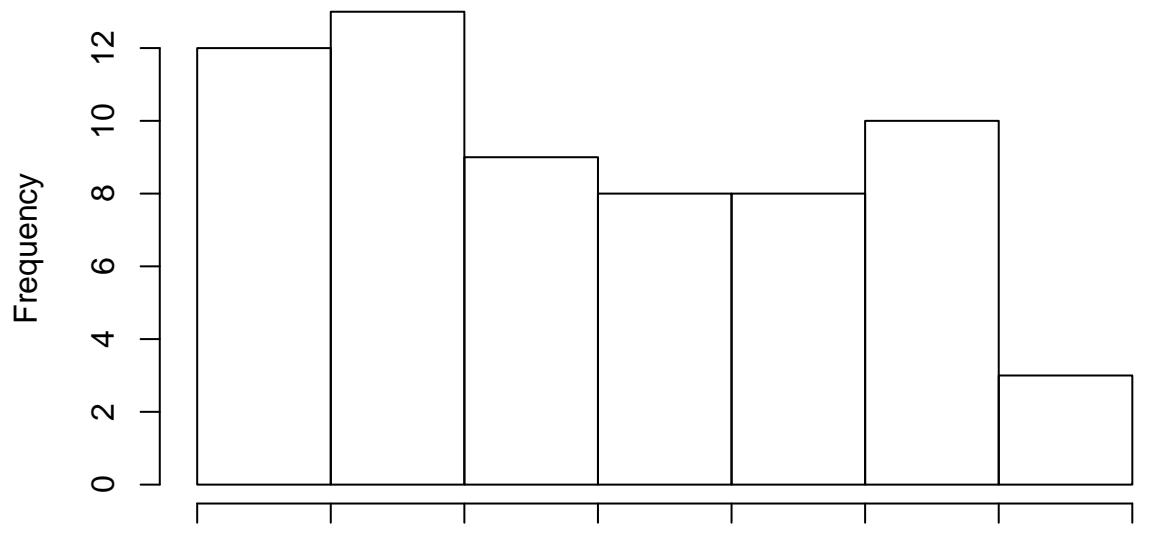
X18mSay



stdat\$values[stdat\$ind %in% c("X18mSay")]

```
hist(stdat$values[stdat$ind %in% c('X24msay')], main='X24msay')
```

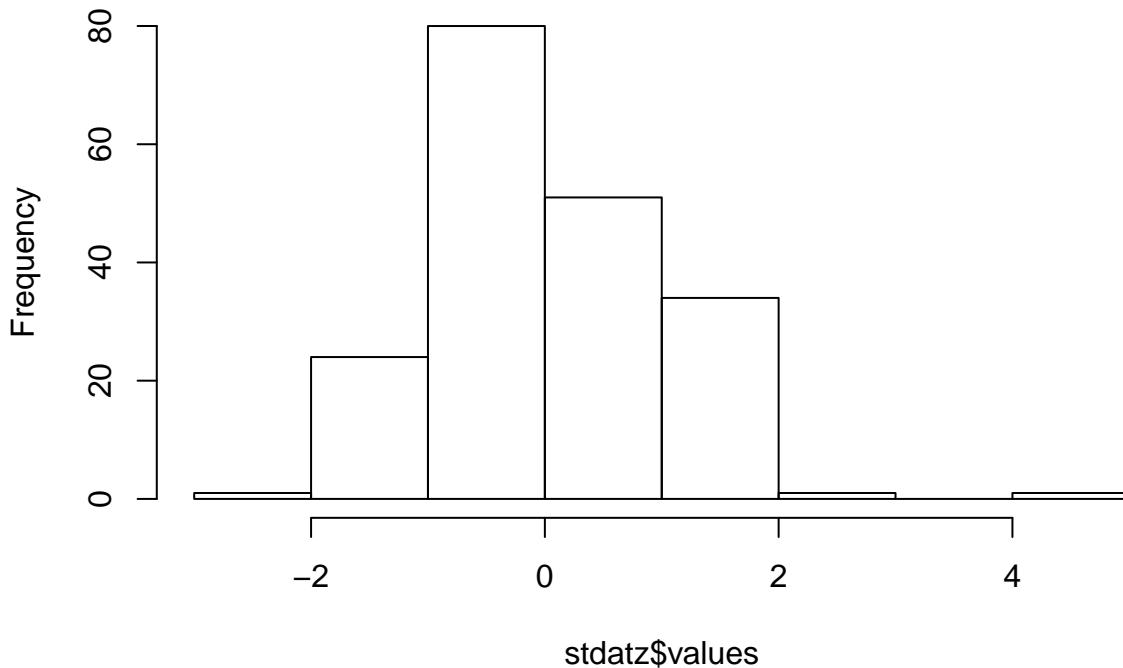
X24msay



stdat\$values[stdat\$ind %in% c("X24msay")]

```
#outcomes - combined - raw+Z
hist(stdatz$values,main="Combined-raw+Z")
```

Combined-raw+Z



```
# find outliers for each variable
for(thisvar in c(outcomes,predictors)) {
  #create a z-score for the observations
  print(thisvar)
  z = (data[,thisvar] - mean(data[,thisvar], na.rm=T) )/ sd(data[,thisvar], na.rm=T)
  print(paste("there are",sum(abs(z)> 3, na.rm=T), "observations that are outliers
             at 3 SD in", thisvar))
}

## [1] "X18mCom"
## [1] "there are 0 observations that are outliers \n
## [1] "X18mComGenderSpec"
## [1] "there are 0 observations that are outliers \n
## [1] "X18mSay"
## [1] "there are 1 observations that are outliers \n
## [1] "X18mSayGenderSpec"
## [1] "there are 0 observations that are outliers \n
## [1] "X24mCom"
## [1] "there are 0 observations that are outliers \n
## [1] "X24msay"
## [1] "there are 0 observations that are outliers \n
## [1] "X24mSayGenderSpec"
## [1] "there are 0 observations that are outliers \n
## [1] "total.tro.pref.quotient"
## [1] "there are 0 observations that are outliers \n
## [1] "Vowel.Alt.pref.quotient"
## [1] "there are 0 observations that are outliers \n
```

at 3 SD in X18mCom"

at 3 SD in X18mComGenderSpec"

at 3 SD in X18mSay"

at 3 SD in X18mSayGenderSpec"

at 3 SD in X24mCom"

at 3 SD in X24msay"

at 3 SD in X24mSayGenderSpec"

at 3 SD in total.tro.pref.quotient"

at 3 SD in Vowel.Alt.pref.quotient"

```

## [1] "Novelty.VRM"
## [1] "there are 2 observations that are outliers \n                                at 3 SD in Novelty.VRM"
table(abs(stdatz$values)>3)

##
## FALSE  TRUE
##   191     1

```

Bivariate relationship among predictors

Correlation matrix

Table 2 presents the bivariate correlations between infant measures that were continuous (Stress, Vowel, VRM), and language outcome measures (18m receptive, 18m expressive, and 24m expressive vocabulary).

```

numvars=as.matrix(mydat[,c("Stress", "VRM","Vowel",
                           "com18z","say18z","say24z")])
colnames(numvars)<-c("Stress","VRM","Vowel","18m receptive","18m expressive","24m expressive")
cor.mat=rcorr(numvars)

# r values
cor.mat$r

##                   Stress          VRM          Vowel 18m receptive
## Stress      1.00000000000  0.0002924873  0.23098790  0.01890751
## VRM        0.0002924873  1.00000000000  0.03903981 -0.02086173
## Vowel      0.2309879016  0.0390398121  1.000000000  0.16547139
## 18m receptive 0.0189075116 -0.0208617293  0.16547139  1.000000000
## 18m expressive 0.0320740429 -0.0293004167  0.21423357  0.59658044
## 24m expressive 0.1371903275  0.0918073053  0.23559252  0.58363815
##                   18m expressive 24m expressive
## Stress      0.03207404  0.13719033
## VRM        -0.02930042  0.09180731
## Vowel      0.21423357  0.23559252
## 18m receptive 0.59658044  0.58363815
## 18m expressive 1.00000000  0.70403815
## 24m expressive 0.70403815  1.00000000

# p values
cor.mat$p

##                   Stress          VRM          Vowel 18m receptive 18m expressive
## Stress      NA 0.9977555 0.02431365  8.811648e-01  7.997760e-01
## VRM        0.99775549 NA 0.71023825  8.700285e-01  8.182217e-01
## Vowel      0.02431365 0.7102382  NA 1.913114e-01  8.915203e-02
## 18m receptive 0.88116481 0.8700285 0.19131139  NA 1.568061e-07
## 18m expressive 0.79977603 0.8182217 0.08915203  1.568061e-07  NA
## 24m expressive 0.28764991 0.4816245 0.06997017  2.344685e-06  1.418632e-09
##                   24m expressive
## Stress      2.876499e-01
## VRM        4.816245e-01
## Vowel      6.997017e-02
## 18m receptive 2.344685e-06
## 18m expressive 1.418632e-09
## 24m expressive NA

```

```

# Ns
cor.mat$n

##          Stress VRM Vowel 18m receptive 18m expressive 24m expressive
## Stress      97  95   95       65       65       62
## VRM        95  95   93       64       64       61
## Vowel      95  93   95       64       64       60
## 18m receptive 65  64   64       65       65       56
## 18m expressive 65  64   64       65       65       56
## 24m expressive 62  61   60       56       56       62

# create star

cor.mat$star=ifelse(cor.mat$P<.001,"**",
                     ifelse(cor.mat$P<.05,"*",
                           ifelse(cor.mat$P<.1,"†","")))

#compose the table

#round
cor.mat$r=format(round(cor.mat$r,2),nsmall=2)

#add dfs
my.cor.mat=matrix(Map("paste", cor.mat$r, (cor.mat$n-2), sep=" ("), nrow=dim(cor.mat$r)[1])
my.cor.mat=matrix(Map("paste", my.cor.mat, ")"), sep=""), nrow=dim(cor.mat$r)[1])
#add *
my.cor.mat=matrix(Map("paste", my.cor.mat, cor.mat$star, sep=""), nrow=dim(cor.mat$r)[1])

colnames(my.cor.mat)=colnames(cor.mat$r)
rownames(my.cor.mat)=colnames(cor.mat$r)

#remove the values at or below the diagonal
for(i in 1:dim(my.cor.mat)[1]) for(j in 1:dim(my.cor.mat)[1]) if(i>=j) my.cor.mat[i,j]<- ""

write.table(my.cor.mat,file="tab2.txt",row.names=T,quote=T,sep="\t")

```

Regressions for AnotB

Given that the A-not-B measure was categorical, we performed 6 separate one-way analyses of variance (ANOVAs) to examine whether there were differences on the other tasks with continuous data as a function of A-not-B outcome.

```

dothiscell<-function(thistype,thisvar){
  paste0(round(mean(mydat$anotB==thistype,thisvar),na.rm=T),2),
  " (",
  round(sd(mydat$anotB==thistype,thisvar),na.rm=T),2),"; ",
  sum(!is.na(mydat$anotB==thistype,thisvar))),
  ")")
}

regtab=NULL

for(thisvar in c("Stress","VRM","Vowel",
                "com18z","say18z","say24z")){
  myaov=summary(aov(mydat[,thisvar]~anotB,data=mydat))
}

```

```

regtab=rbind(regtab,
             cbind(thisvar,
                    dothiscell("Hit",thisvar),dothiscell("Random",thisvar),dothiscell("Error",thisvar)
                    paste0("F(", unlist(myaov)[ "Df2"], ")=", round(unlist(myaov)[ "F value1"],2),
                           " p=", round(unlist(myaov)[ "Pr(>F)1"],3)))
            ))}

colnames(regtab)<-c(
  "Task",
  "Hit",
  "Error",
  "Random",
  "F(df), p"
)
regtab[4:6,'Task']<-c("18m receptive","18m expressive","24m expressive")
regtab

##      Task          Hit         Error        Random
## [1,] "Stress"    "0.47 (0.12; 23)" "0.52 (0.16; 35)" "0.52 (0.13; 36)"
## [2,] "VRM"       "0.61 (0.05; 23)" "0.59 (0.06; 34)" "0.59 (0.08; 35)"
## [3,] "Vowel"     "0.5 (0.09; 22)"  "0.52 (0.1; 35)" "0.53 (0.1; 35)"
## [4,] "18m receptive"  "-0.01 (1.05; 19)" "-0.12 (0.88; 21)" "0.11 (1.08; 25)"
## [5,] "18m expressive" "0.09 (0.91; 19)"  "-0.25 (0.69; 21)" "0.15 (1.25; 25)"
## [6,] "24m expressive" "0.4 (1.05; 18)"   "-0.2 (0.89; 21)" "-0.13 (1; 23)"
##      F(df), p
## [1,] "F(91)=1.29 p=0.28"
## [2,] "F(89)=0.93 p=0.398"
## [3,] "F(89)=0.63 p=0.537"
## [4,] "F(62)=0.3 p=0.745"
## [5,] "F(62)=1.03 p=0.363"
## [6,] "F(59)=2.15 p=0.125"

write.table(regtab,file="tab3.txt",row.names=F,quote=T,sep="\t")

```

Setting up models

If RECALC is TRUE then the next chunk will not be done...

```

readRDS("vowel_main_model.rds")->vowel
readRDS("anotb_main_model.rds")->anb
readRDS("stress_main_model.rds")->strs
readRDS("vrm_main_model.rds")->vrm
readRDS("main_model.rds")->main
main_summary=readRDS("main_model_summary.rds")
main_BF=readRDS("main_model_bf.rds")

```

... but this one will (and vice versa):

```

niter=4000
nwarmup=500

```

```

#values for scaled variables -- this cannot be passed as a variable to stan, but it's noted here for cl
nu=3
s=1

```

```

our_priors <- c(prior("student_t(3,0,1)", class = b),
                 prior("student_t(3,0,1)", class = Intercept)
               )

```

Prediction of vocabulary from each infant measure

Univariate Bayesian models

Stress

```

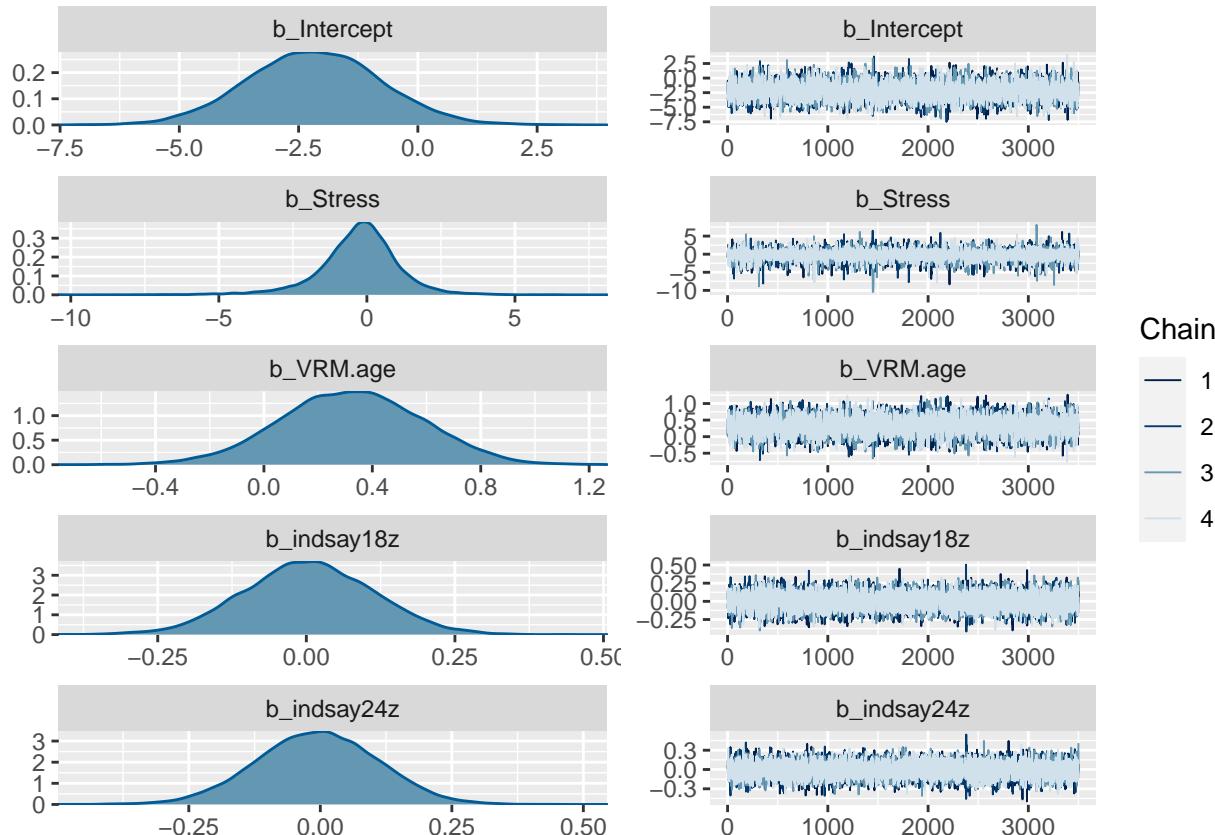
strs = brm(values ~
  Stress*VRM.age +
  + ind + (1 | Subject..), data=stdatz,
  prior = our_priors,
  iter=niter, warmup=nwarmup, chains=4, cores=2,
  seed=12,
  save_all_pars = T,
  sample_prior = T
)

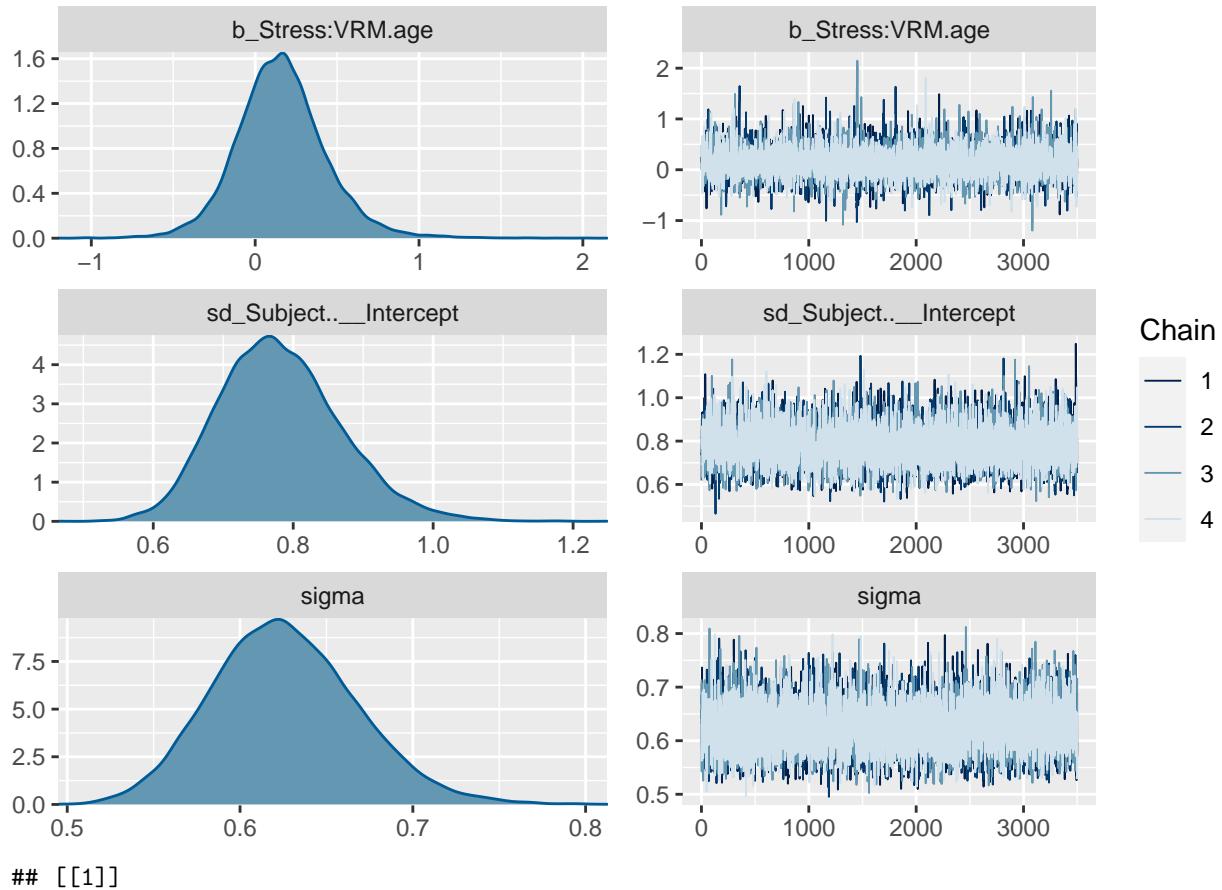
saveRDS(strs, file = paste("stress", "main_model.rds", sep="_"))

fit_uni_print(strs,"stress")

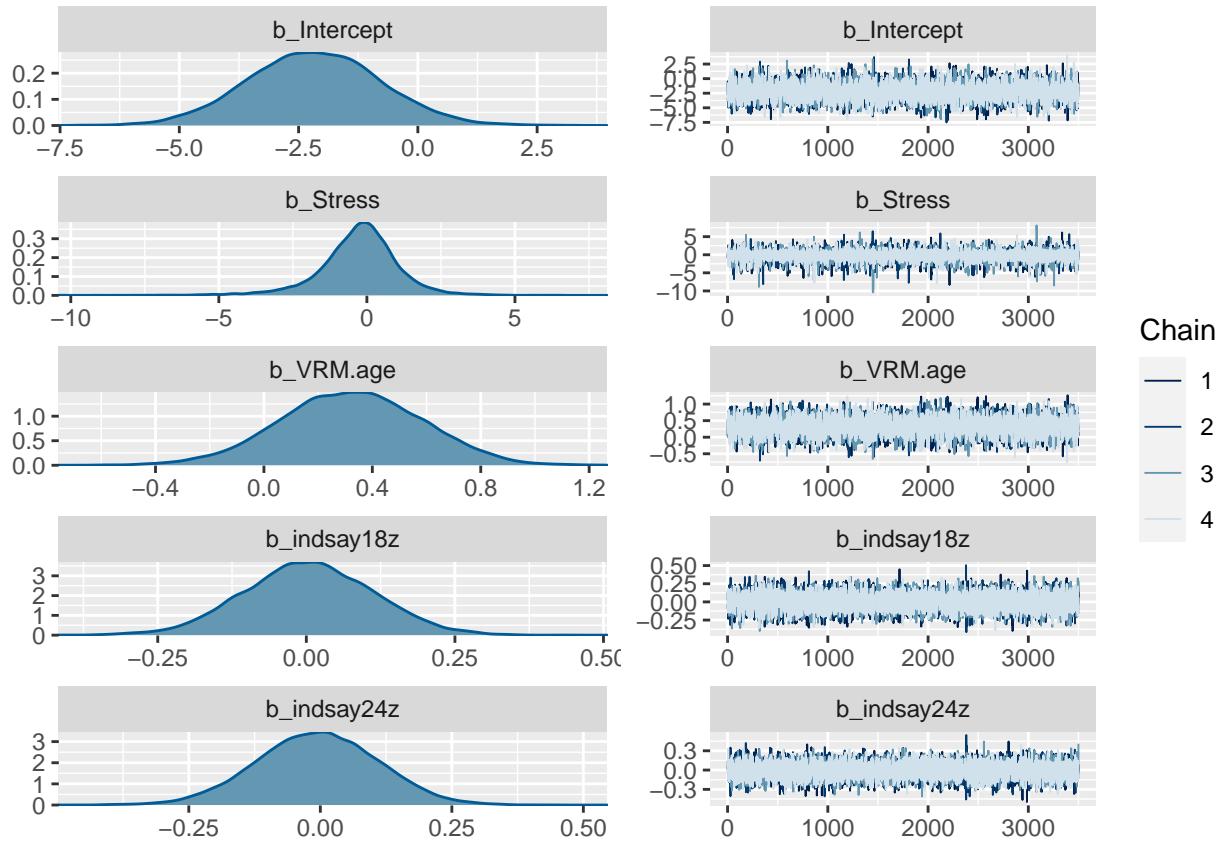
```

[1] "stress"





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



```

## 
## [[2]]
## 
## Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: values ~ Stress * VRM.age + +ind + (1 | Subject..)
##   Data: stdatz (Number of observations: 189)
## Samples: 4 chains, each with iter = 4000; warmup = 500; thin = 1;
##          total post-warmup samples = 14000
## 
## Group-Level Effects:
## ~Subject.. (Number of levels: 70)
##             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)    0.78      0.09     0.63     0.97 1.00    4212     7235
## 
## Population-Level Effects:
##             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept      -2.17      1.40    -4.89     0.59 1.00    5236     7774
## Stress        -0.27      1.37    -3.34     2.29 1.00   12368     5896
## VRM.age        0.32      0.26    -0.19     0.83 1.00    4926     7385
## indsay18z       0.00      0.11    -0.21     0.22 1.00   15867    11041
## indsay24z      -0.01      0.12    -0.23     0.21 1.00   14584     9907
## Stress:VRM.age   0.16      0.28    -0.36     0.74 1.00    7690     6142
## 
## Family Specific Parameters:
##             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma       0.63      0.04     0.55     0.71 1.00    8395   10117

```

```

##  

## Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS  

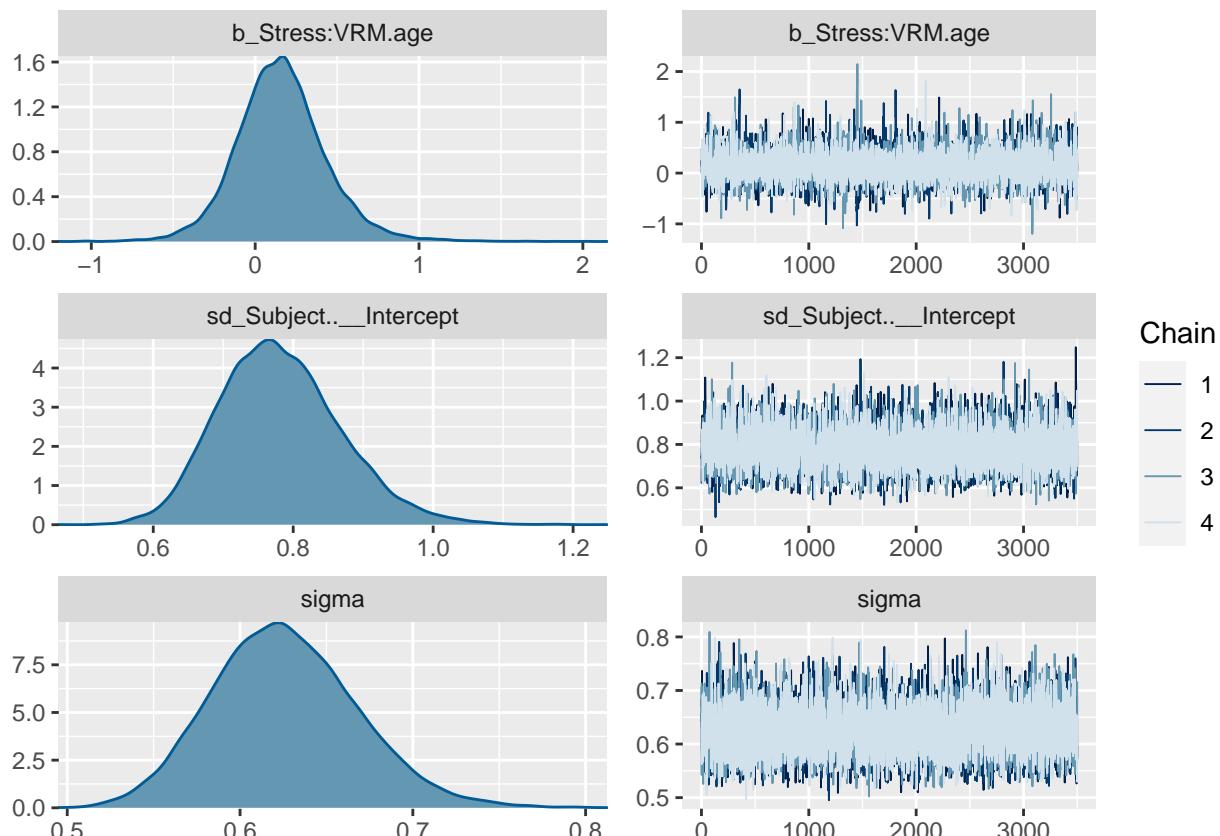
## and Tail_ESS are effective sample size measures, and Rhat is the potential  

## scale reduction factor on split chains (at convergence, Rhat = 1).  

## Sampling priors, please wait...  

## Loading required namespace: logspline

```



```

## # Bayes Factor (Savage-Dickey density ratio)
##  

## Parameter | BF
## -----  

## Intercept | 0.617
## Stress | 0.929
## VRM.age | 0.550
## indsay18z | 0.098
## indsay24z | 0.102
## Stress.VRM.age | 0.260
##  

## * Evidence Against The Null: [0]

```

VRM

```

vrm = brm(values ~  

  VRM*VRM.age +  

  + ind + (1 | Subject..), data=stdatz,  

  prior = our_priors,

```

```

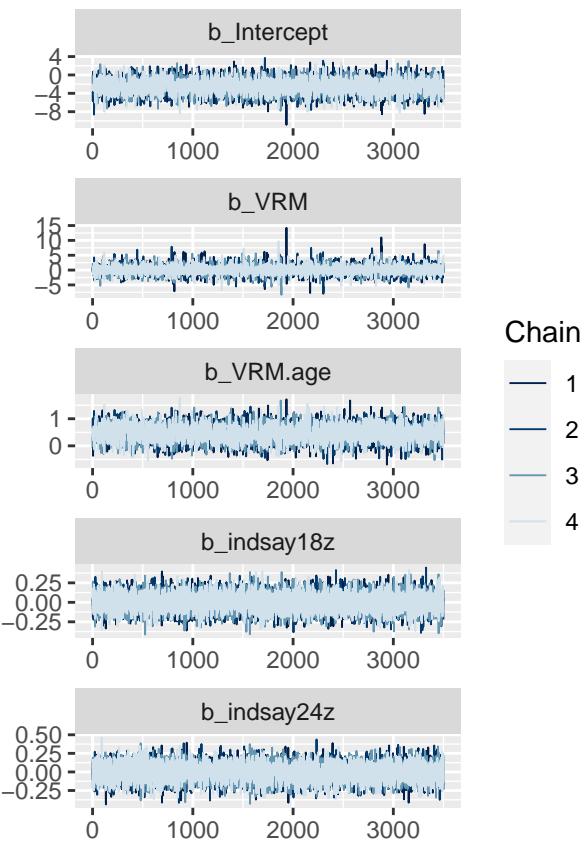
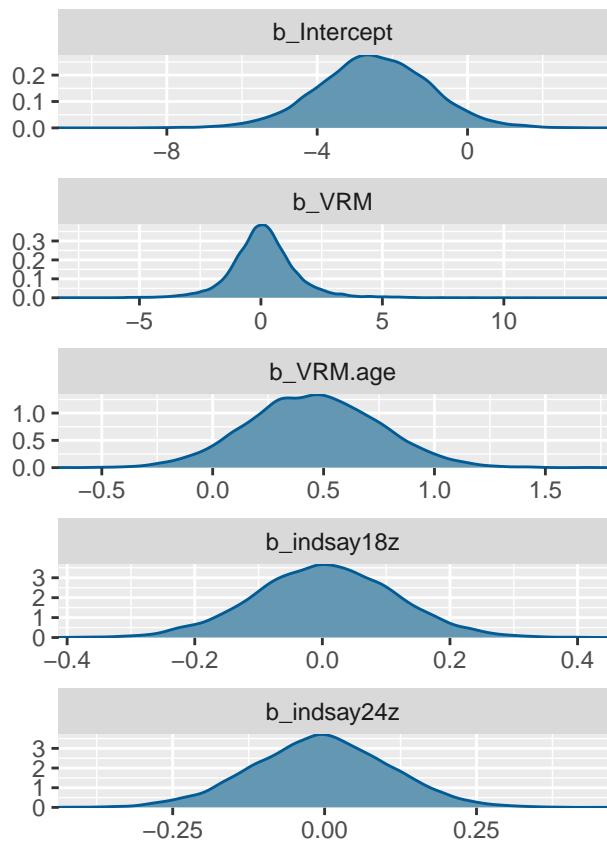
    iter=niter, warmup=nwarmup, chains=4,cores=2,
    seed=12,
    save_all_pars = T,
    sample_prior = T
  )

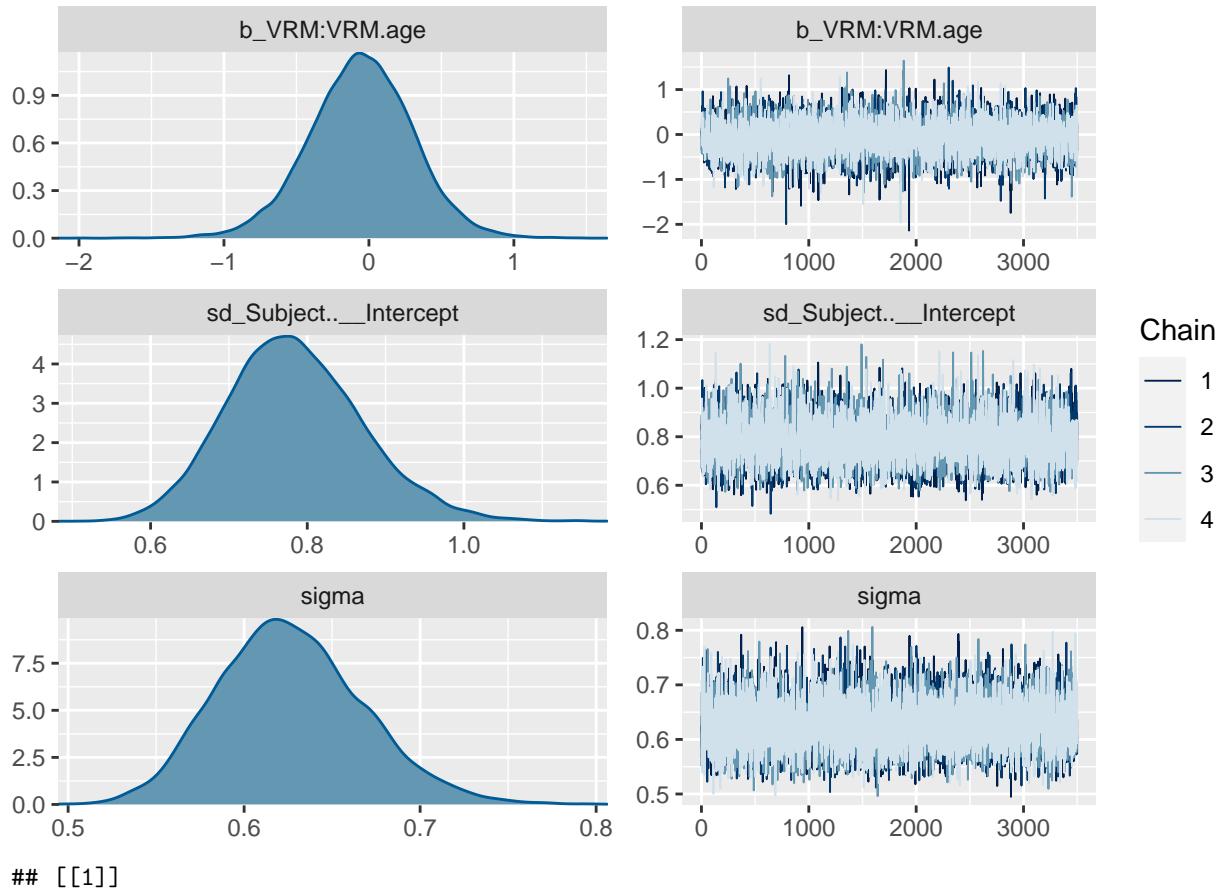
saveRDS(vrm, file = paste("vrm", "main_model.rds",sep="_"))

fit_uni_print(vrm,"vrm")

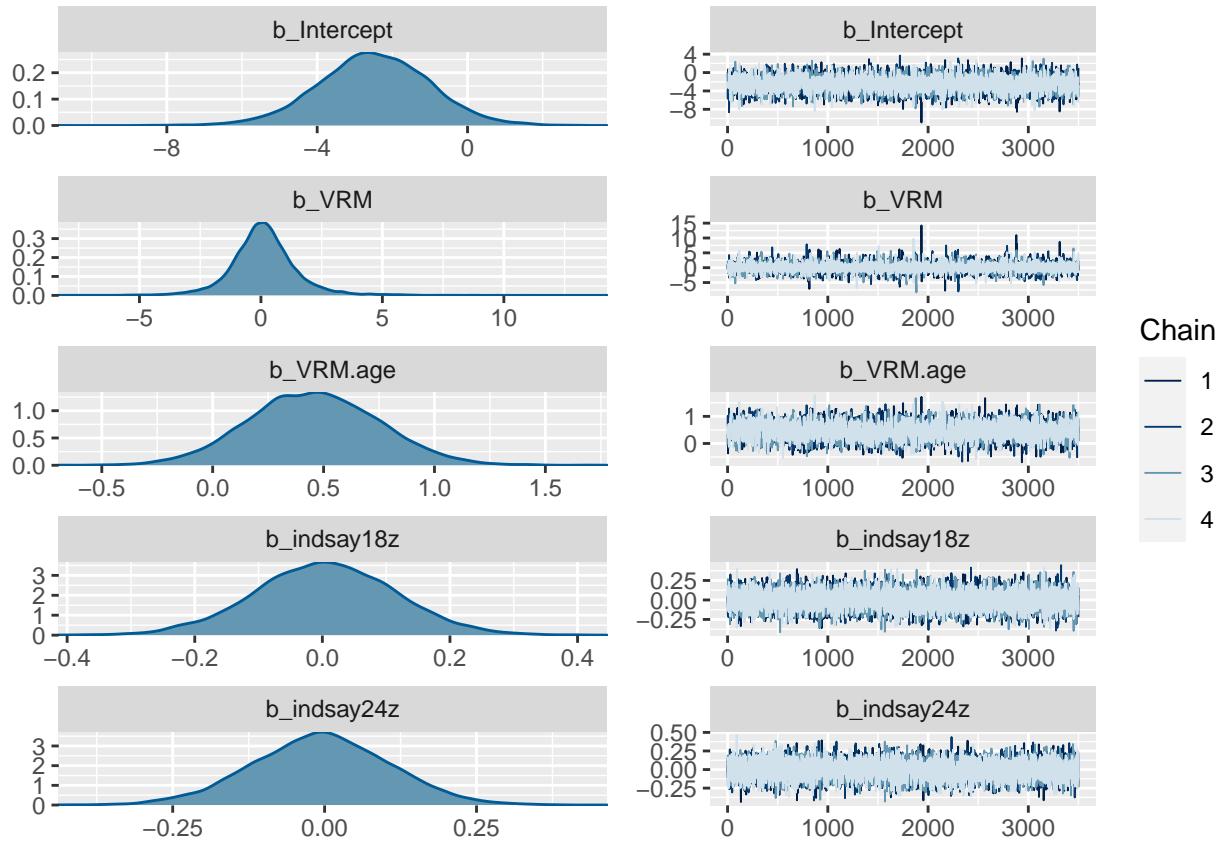
```

[1] "vrm"





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



```
##  
## [[2]]  
  
##  
## Family: gaussian  
## Links: mu = identity; sigma = identity  
## Formula: values ~ VRM * VRM.age + +ind + (1 | Subject..)  
## Data: stdatz (Number of observations: 189)  
## Samples: 4 chains, each with iter = 4000; warmup = 500; thin = 1;  
## total post-warmup samples = 14000  
##  
## Group-Level Effects:  
## ~Subject.. (Number of levels: 70)  
## Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS  
## sd(Intercept) 0.78 0.09 0.63 0.97 1.00 4262 7975  
##  
## Population-Level Effects:  
## Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS  
## Intercept -2.49 1.48 -5.45 0.40 1.00 4817 7481  
## VRM 0.11 1.38 -2.57 3.09 1.00 9895 5813  
## VRM.age 0.46 0.29 -0.11 1.04 1.00 4099 6404  
## indsay18z 0.00 0.11 -0.22 0.22 1.00 13633 10245  
## indsay24z -0.01 0.11 -0.23 0.21 1.00 12337 10628  
## VRM:VRM.age -0.06 0.36 -0.77 0.64 1.00 5011 6789  
##  
## Family Specific Parameters:  
## Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS  
## sigma 0.63 0.04 0.55 0.71 1.00 7826 10576
```

```

##  

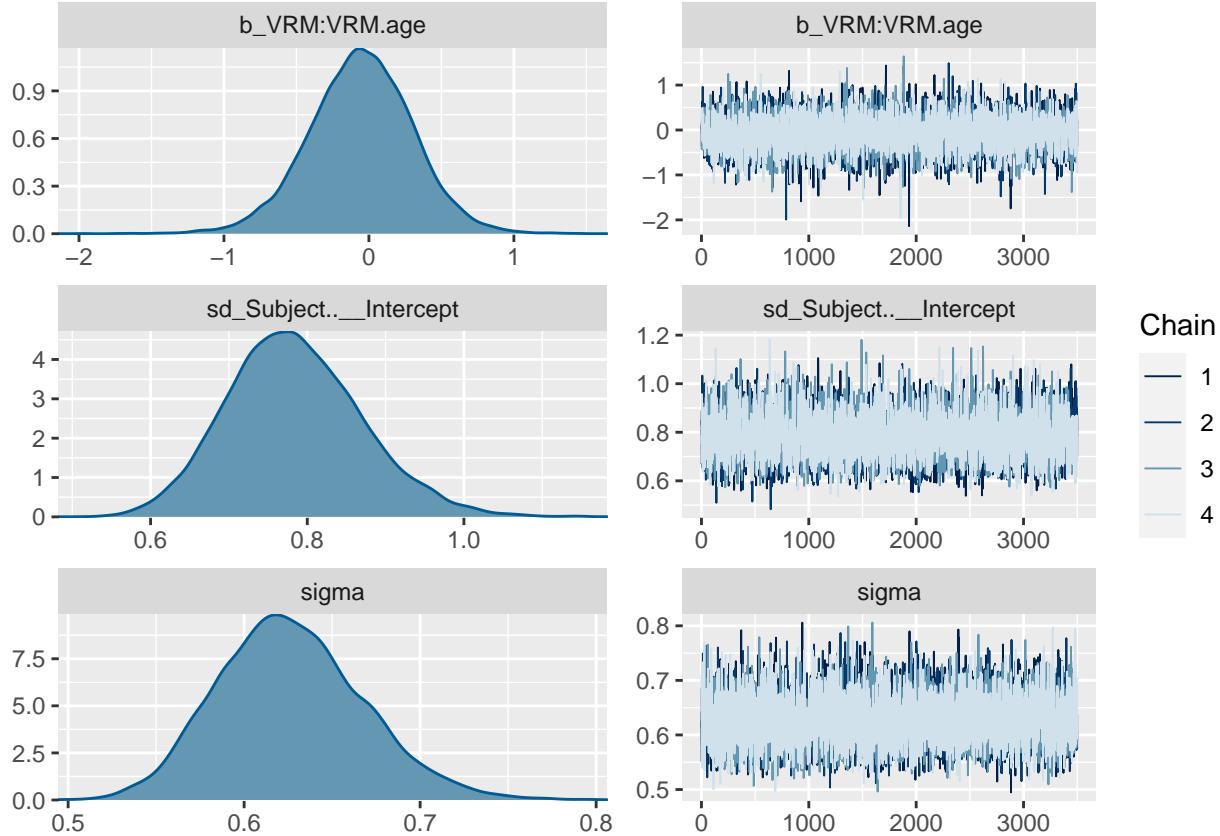
## Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS  

## and Tail_ESS are effective sample size measures, and Rhat is the potential  

## scale reduction factor on split chains (at convergence, Rhat = 1).  

## Sampling priors, please wait...

```



```

## # Bayes Factor (Savage-Dickey density ratio)
##  

## Parameter | BF
## -----
## Intercept | 0.806
## VRM       | 0.907
## VRM.age   | 0.937
## indsay18z | 0.105
## indsay24z | 0.098
## VRM.VRM.age | 0.323
##  

## * Evidence Against The Null: [0]

```

Vowel

```

vowel = brm(values ~  

  Vowel*vowels.Age +  

  + ind + (1 | Subject..), data=stdatz,  

  prior = our_priors,  

  iter=niter, warmup=nwarmup, chains=4, cores=2,  

  seed=12,

```

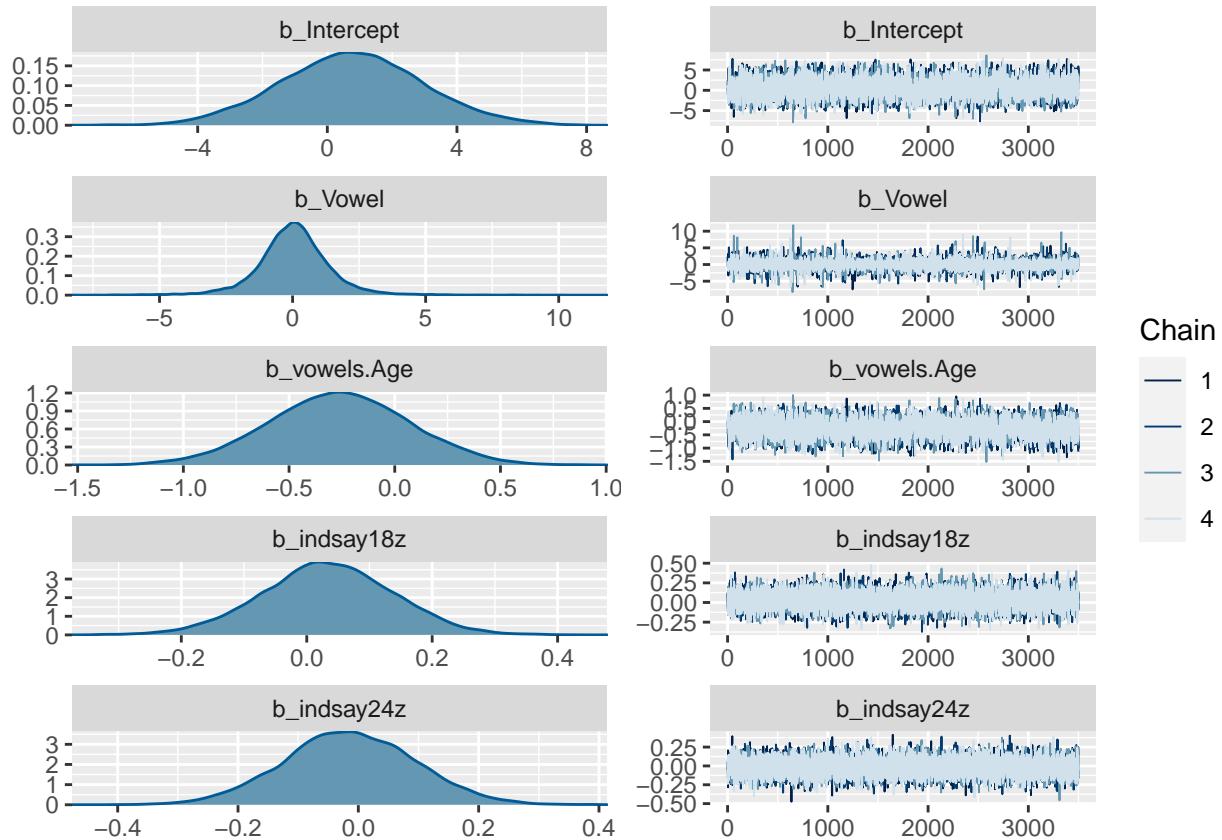
```

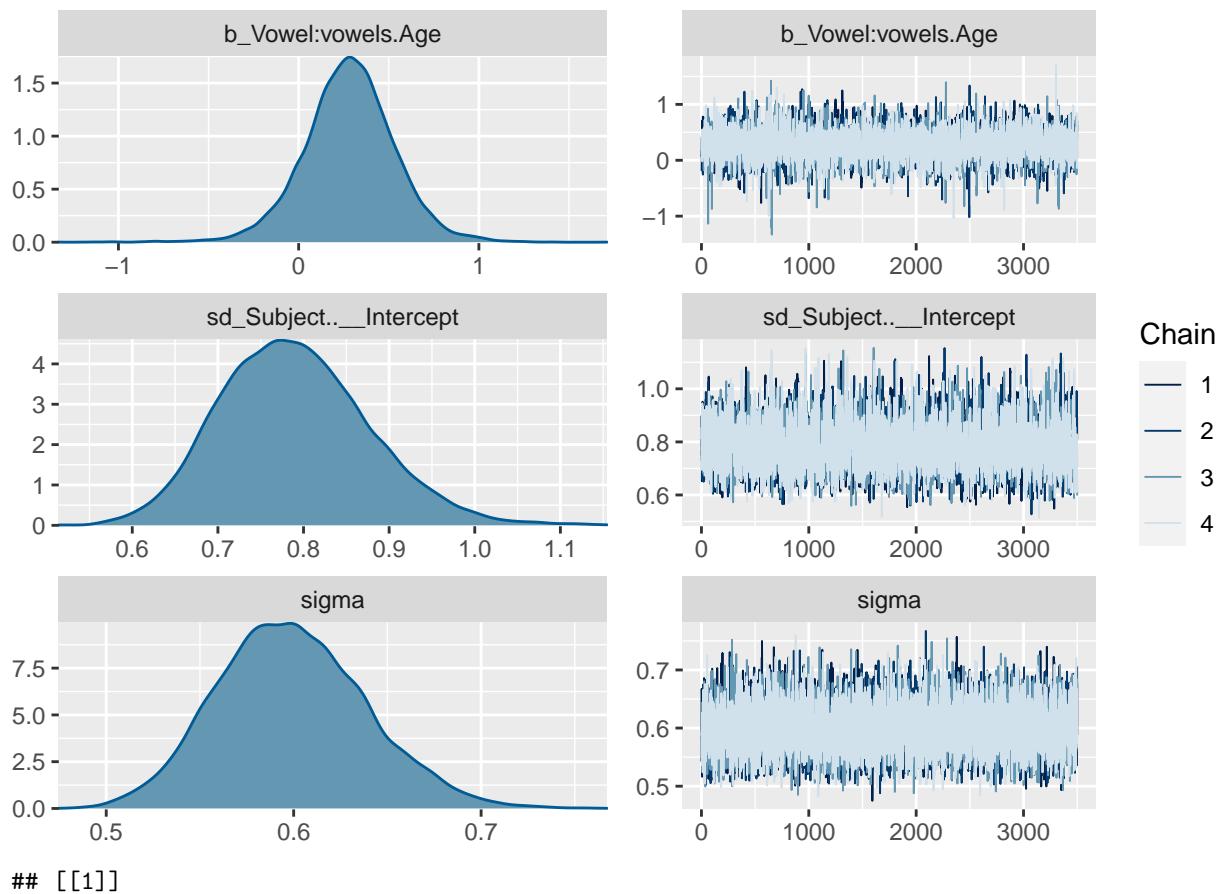
        save_all_pars = T,
        sample_prior = T
    )
saveRDS(vowel, file = paste("vowel", "main_model.rds", sep="_"))

fit_uni_print(vowel,"vowel")

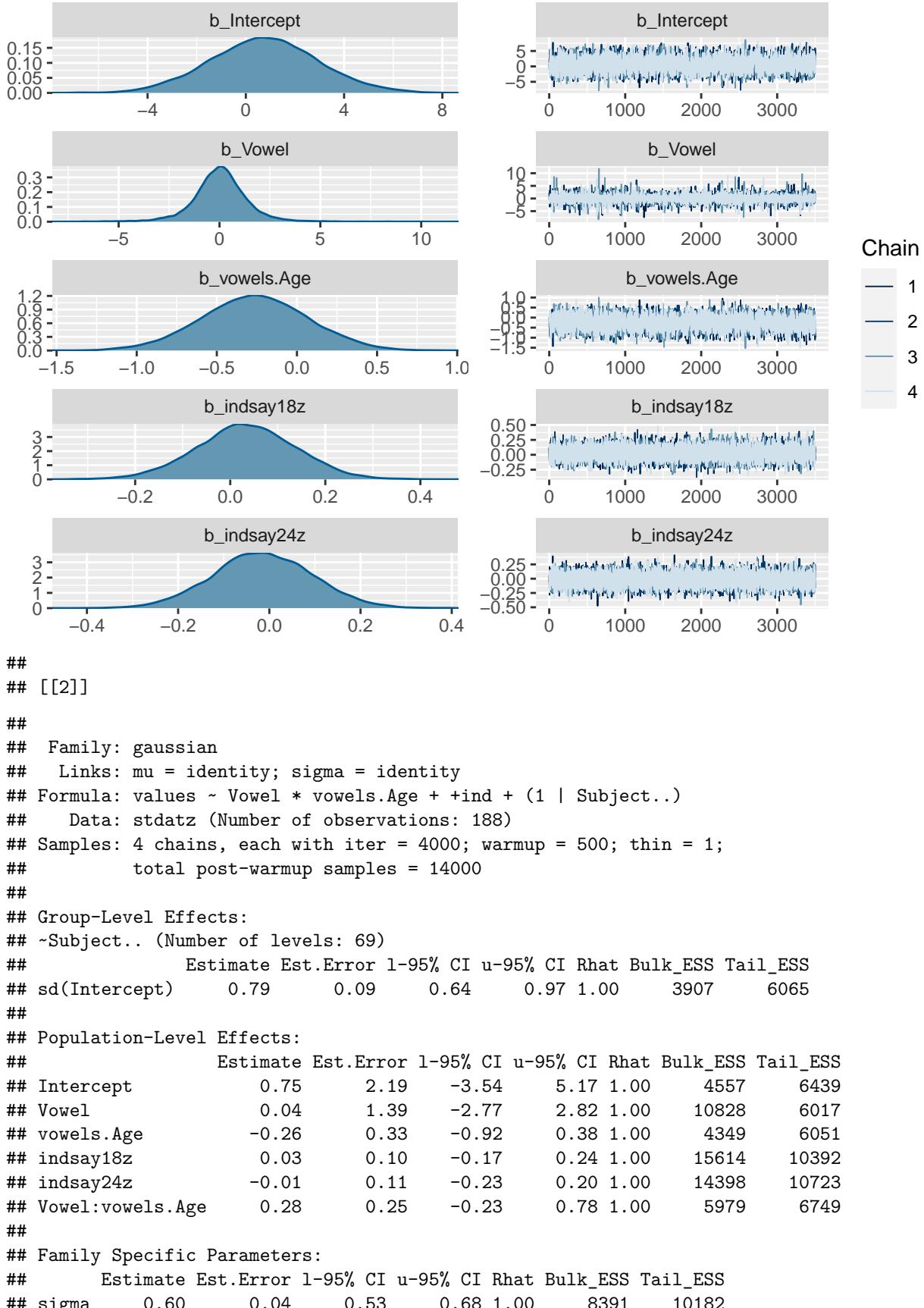
## [1] "vowel"

```





```
## [1]
```



```

##  

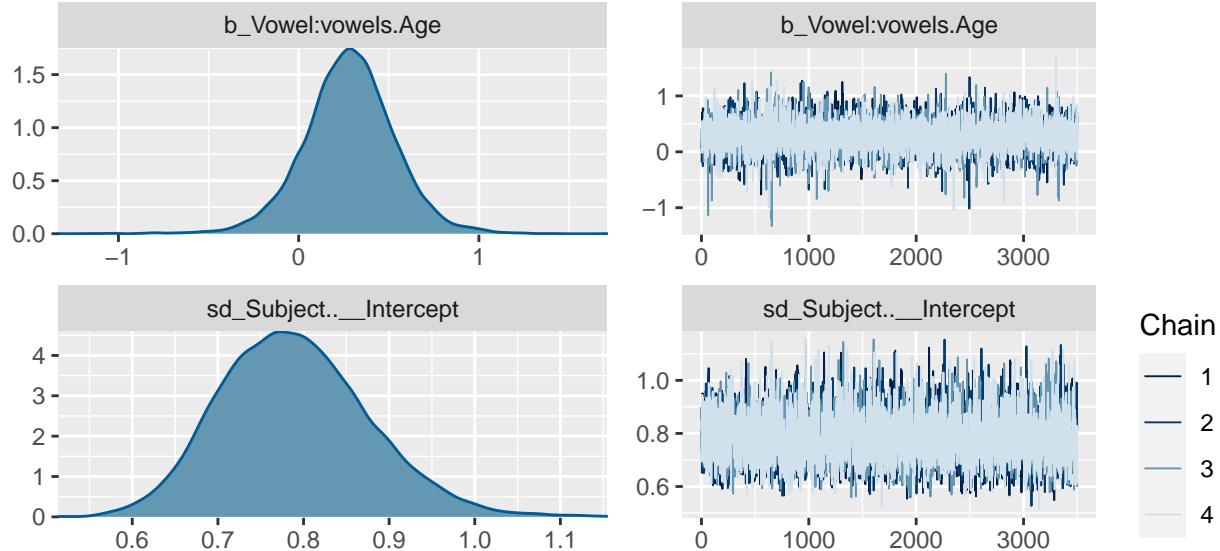
## Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS  

## and Tail_ESS are effective sample size measures, and Rhat is the potential  

## scale reduction factor on split chains (at convergence, Rhat = 1).  

## Sampling priors, please wait...

```



```

## # Bayes Factor (Savage-Dickey density ratio)
##  

## Parameter | BF
## -----
## Intercept | 0.242
## Vowel     | 0.956
## vowels.Age | 0.415
## indsay18z | 0.101
## indsay24z | 0.098
## Vowel.vowels.Age | 0.521
##  

## * Evidence Against The Null: [0]

```

A-not-B

```

anb = brm(values ~  

  anotB*A.not.B.Age +  

  + ind + (1 | Subject..), data=stdatz,  

  prior = our_priors,  

  iter=niter, warmup=nwarmup, chains=4, cores=2,  

  seed=12,

```

```

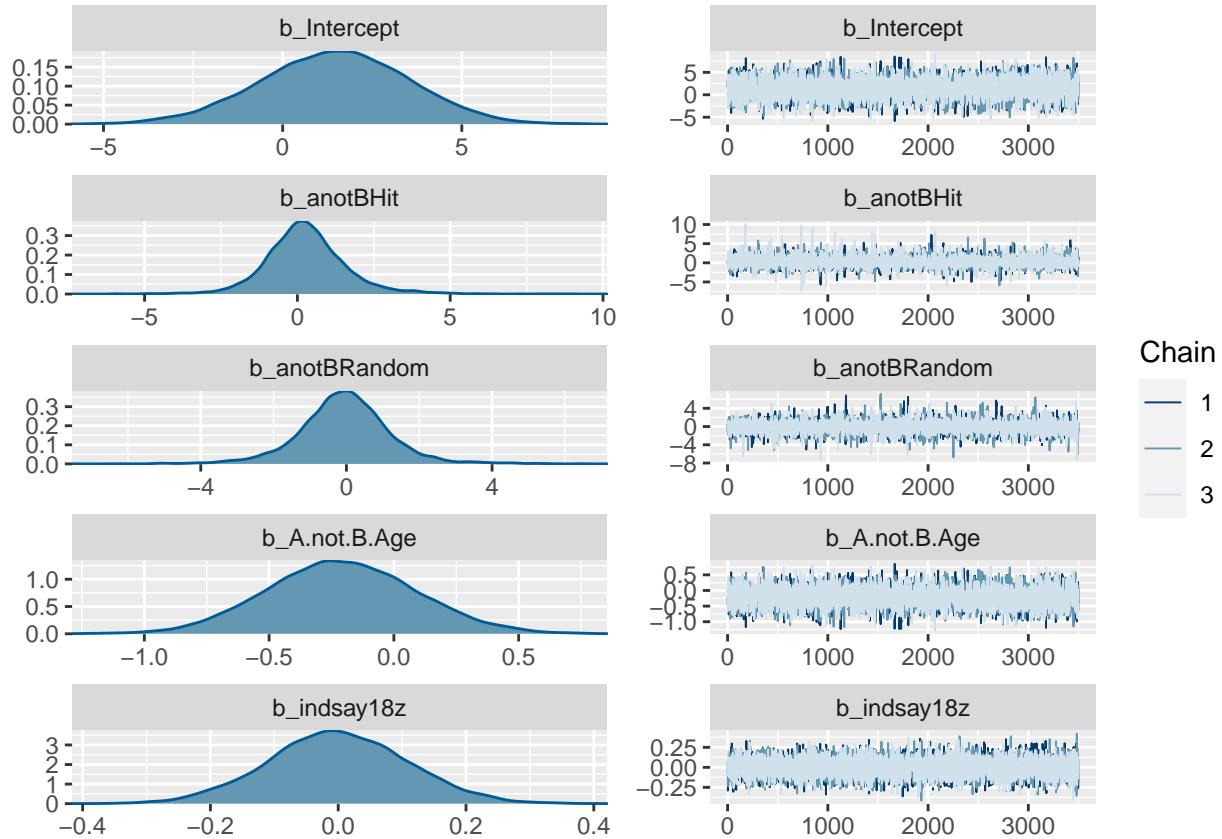
    save_all_pars = T,
    sample_prior = T
)

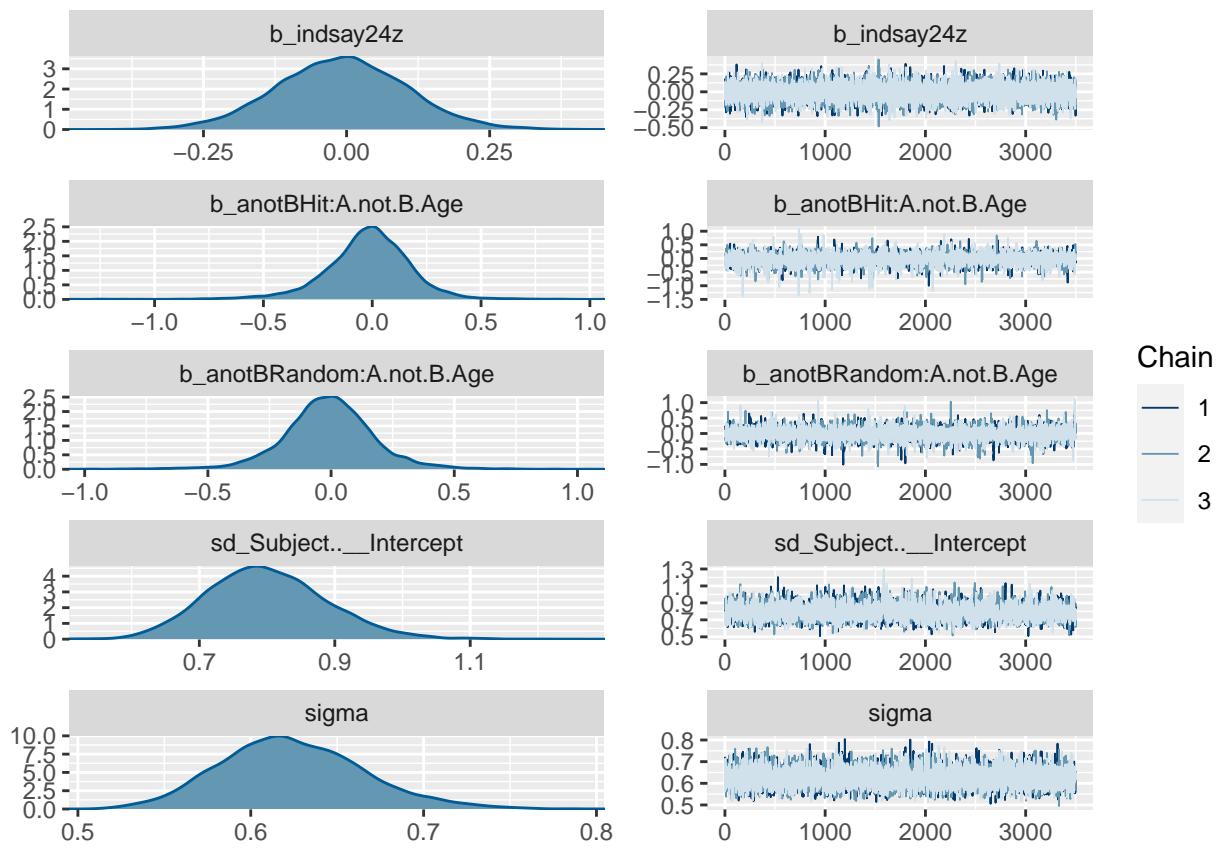
saveRDS(anb, file = paste("anotb", "main_model.rds", sep="_"))

fit_uni_print(anb,"anotb")

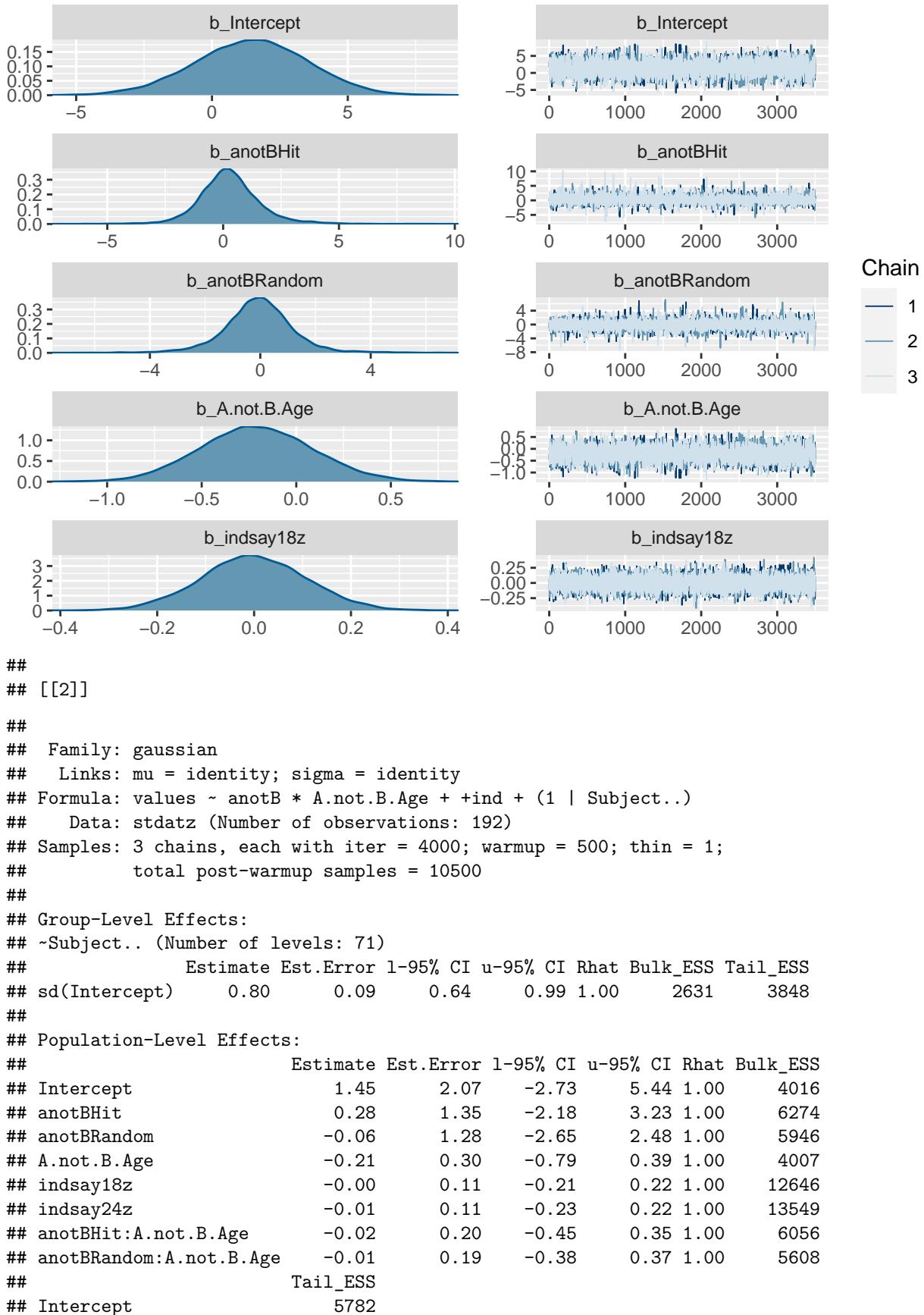
## [1] "anotb"

```





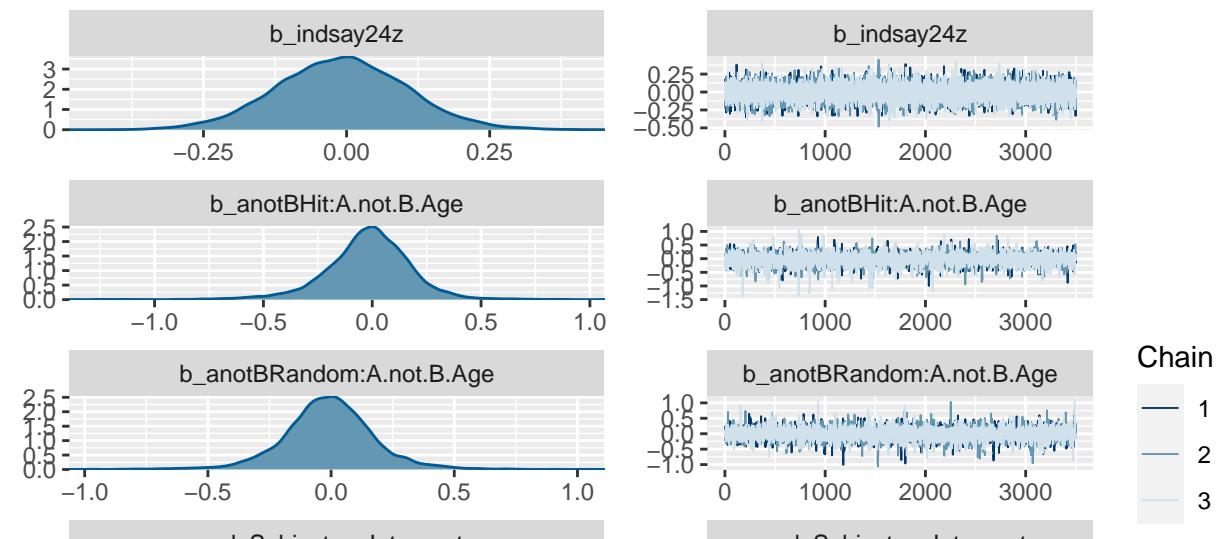
```
## [[1]]
```



```

## anotBHit          4377
## anotBRandom       4777
## A.not.B.Age       5427
## indsay18z          7787
## indsay24z          7527
## anotBHit:A.not.B.Age 4421
## anotBRandom:A.not.B.Age 4403
##
## Family Specific Parameters:
##           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma      0.62      0.04     0.55     0.71 1.00      6171      7177
##
## Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
##
## Sampling priors, please wait...

```



```

## # Bayes Factor (Savage-Dickey density ratio)
##
## Parameter | BF
## -----
## Intercept | 0.283
## anotBHit  | 0.986
## anotBRandom | 0.996
## A.not.B.Age | 0.340
## indsay18z  | 0.091
## indsay24z  | 0.096

```

```

## anotBHit.A.not.B.Age | 0.154
## anotBRandom.A.not.B.Age | 0.148
##
## * Evidence Against The Null: [0]

```

Add estimates to table 1

```

myvec=c(paste0(round(fixef(strs)[["Stress","Estimate"]],2)," (",round(fixef(strs)[["Stress","Est.Error"]],2),
  paste0(round(fixef(vrm)[["VRM","Estimate"]],2)," (",round(fixef(vrm)[["VRM","Est.Error"]],2),")"),
  paste0(round(fixef(vowel)[["Vowel","Estimate"]],2)," (",round(fixef(vowel)[["Vowel","Est.Error"]],2),
  paste0("<",round(max(fixef(anb)[c("anotBHit","anotBRandom"),"Estimate"])),2)," >",round(min(fixef(anb)[c("anotBHit","anotBRandom"),"Estimate"])),2),
  "___",
  "___",
  "___"
)
cbind(tab1,myvec)->tab1
colnames(tab1)<-c("Measure",      "No.",    "Age (SD)", "Score (SD)",   "Cohen's d", "Estimate (SE)")
write.table(tab1,file="tab1.txt",row.names=F,quote=F,sep="\t")

```

Table 4

Estimate (standard error), 95 percent credibility interval, and Bayes Factor for each measure as predictor of vocabulary (in a Bayesian mixed model declaring all vocabulary measures together), as well as the same metrics for this predictor's interaction with age in that model. Null>Alt indicates that the Bayes Factor shows more support for the null than the alternative equal to the ratio provided. No predictor or interaction showed more support for the alternative than the null.

```

# Predictor Interaction with age
#Measure, Estimate (SE), 95% CI, BF

dolinetab4<-function(mymod,varname){
  #print(varname)
  #mymod=strs ; varname="Stress"
  if(varname %in% c("anotBHit","anotBRandom")) readRDS(file = paste("anotb", "main_model_bf.rds",sep="_"))

  age_interaction_line=grep(varname,rownames(fixef(mymod)))[2]

  cbind(varname, paste0(round(fixef(mymod)[varname,"Estimate"],2)," (",round(fixef(mymod)[varname,"Est.Error"],2),
    paste0("[",round(fixef(mymod)[varname,"Q2.5"],2),",",",round(fixef(mymod)[varname,"Q97.5"],2),"]"),
    round(1/temp[grep(varname,temp$Parameter),"BF"])[1],2), #bf #this is ugly but basically there are
    #repeat for interaction with age
    paste0(round(fixef(mymod)[age_interaction_line,"Estimate"],2)," (",round(fixef(mymod)[age_interaction_line,"Est.Error"],2),
    paste0("[",round(fixef(mymod)[age_interaction_line,"Q2.5"],2),",",",round(fixef(mymod)[age_interaction_line,"Q97.5"],2),"]"),
    round(1/temp[grep(varname,temp$Parameter),"BF"])[2],2) #bf
  )
}

tab4=rbind(
  dolinetab4(strs,"Stress"),
  dolinetab4(vrm,"VRM"),
  dolinetab4(vowel,"Vowel"),
  dolinetab4(anb,"anotBHit"),
  dolinetab4(anb,"anotBRandom")
)

```

```

)
colnames(tab4)<-c("Task", "Estimate (SE)", "95% CI", "Null>Alt BF", "Estimate (SE)", "95% CI"
tab4

##      Task      Estimate (SE)    95% CI Null>Alt BF Estimate (SE)    95% CI
## [1,] "Stress"   "-0.27 (1.37)" "[ -3.34, 2.29]" "1.08"     "0.16 (0.28)"
## [2,] "VRM"      "0.11 (1.38)"  "[ -2.57, 3.09]" "1.1"      "0.46 (0.29)"
## [3,] "Vowel"    "0.04 (1.39)"  "[ -2.77, 2.82]" "1.05"     "0.28 (0.25)"
## [4,] "anotBHit" "0.28 (1.35)"  "[ -2.18, 3.23]" "1.01"     "-0.02 (0.2)"
## [5,] "anotBRandom" "-0.06 (1.28)" "[ -2.65, 2.48]" "1"        "-0.01 (0.19)"
##      95% CI Null>Alt BF
## [1,] "[ -0.36, 0.74]" "3.85"
## [2,] "[ -0.11, 1.04]" "1.07"
## [3,] "[ -0.23, 0.78]" "1.92"
## [4,] "[ -0.45, 0.35]" "6.47"
## [5,] "[ -0.38, 0.37]" "6.77"
write.table(tab4,file="tab4.txt",row.names=F,quote=F,sep="\t")

```

Evaluating the relative importance of different predictors

Frequentist analysis

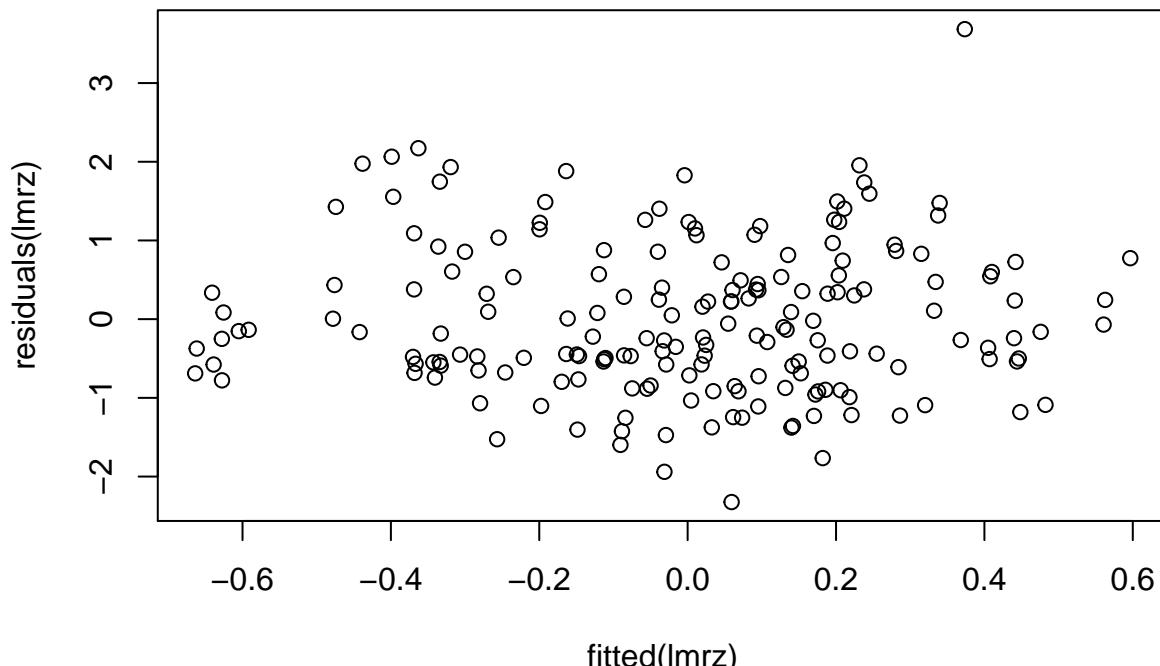
```

lm(values ~ Stress + Vowel +
  VRM + anotB + ind + Gender + (1/Subject..), data= stdatz) -> lmrz

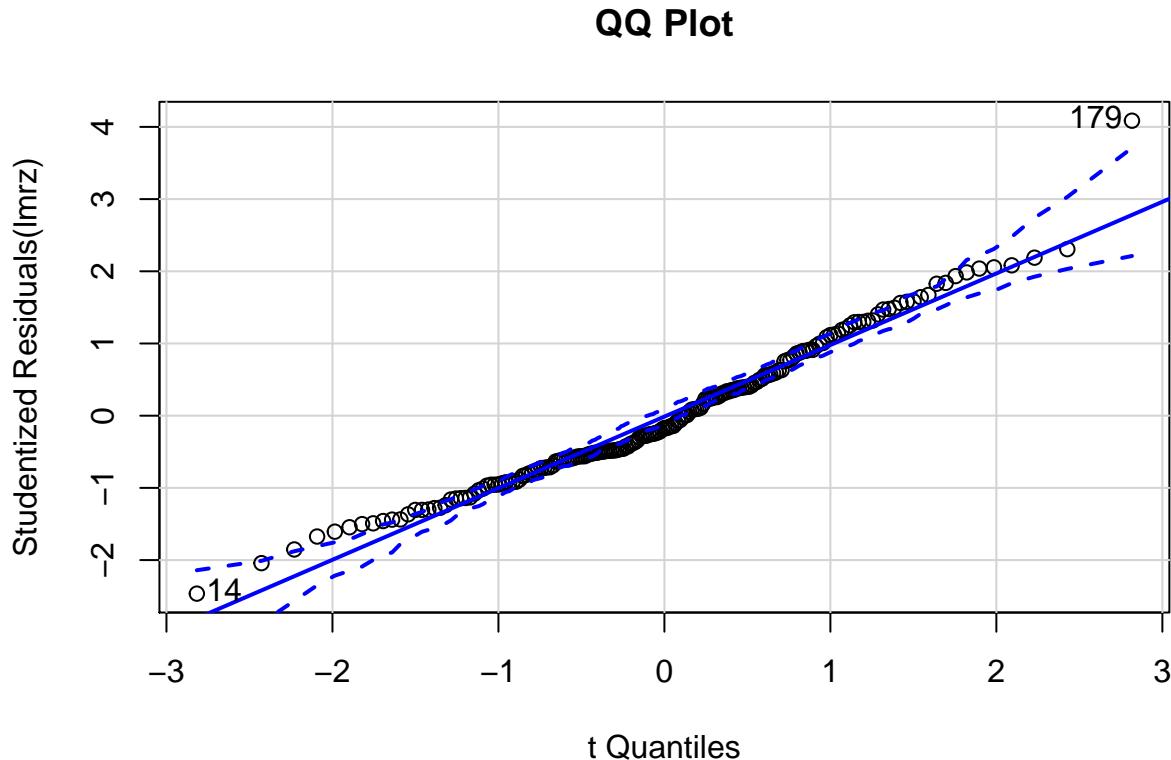
outlierTest(lmrz)

##      rstudent unadjusted p-value Bonferroni p
## 179 4.085668       6.6826e-05     0.012363
plot(residuals(lmrz) ~ fitted(lmrz))

```



```
qqPlot(lmrz, main="QQ Plot")
```



```
## [1] 14 179
```

```
gvlma(lmrz)
```

```
##  
## Call:  
## lm(formula = values ~ Stress + Vowel + VRM + anotB + ind + Gender +  
##       (1/Subject..), data = stdatz)  
##  
## Coefficients:  
## (Intercept)      Stress      Vowel        VRM      anotBHit  anotBRandom  
## -1.277461     0.373654    2.313114   -0.038771     0.163936   -0.228052  
## indsay18z    indsay24z    GenderM  
##  0.035954     0.002047   -0.185360  
##  
##  
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS  
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:  
## Level of Significance =  0.05  
##  
## Call:  
## gvlma(x = lmrz)  
##  
##          Value  p-value           Decision  
## Global Stat 11.0163 0.026381 Assumptions NOT satisfied!  
## Skewness     8.6680 0.003238 Assumptions NOT satisfied!  
## Kurtosis     0.8480 0.357105 Assumptions acceptable.  
## Link Function 1.3819 0.239784 Assumptions acceptable.
```

```
## Heteroscedasticity 0.1184 0.730805 Assumptions acceptable.
```

Conclusion: should do something about outlier; non-linear relationship between fitted and residuals?, qqplot looks OK, all assumptions are acceptable -> it's a go!!

Decision point: I'll trim the value of the outlier

```
stdatz.noout<-stdatz
stdatz.noout[stdatz.noout$values>2 & !is.na(stdatz.noout$values),] #yields 2 observations, the outlier

##      Subject.. Gender   Stress      VRM VRM.age   Vowel anotB vowels.Age
## 112       3560     F 0.4779087 0.6277496   5.66 0.5701395 Error      7.37
## 179       3796     M 0.6159650 0.7024363   5.23 0.6906250 Error      7.30
##      A.not.B.Age values    ind
## 112       7.37 2.186494 say18z
## 179       7.30 4.058508 say18z

stdatz.noout[stdatz.noout$values>4 & !is.na(stdatz.noout$values), "values"]<-2.186494 #this line assigns

lm(values ~ Stress + Vowel +
  VRM + anotB + Gender + ind + (1/Subject..), data= stdatz.noout) -> lmrzno
outlierTest(lmrzno)

## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##      rstudent unadjusted p-value Bonferroni p
## 14 -2.566679          0.011105           NA

gvlma(lmrzno) #all assumptions met

## 
## Call:
## lm(formula = values ~ Stress + Vowel + VRM + anotB + Gender +
##     ind + (1/Subject..), data = stdatz.noout)
## 
## Coefficients:
## (Intercept)      Stress        Vowel        VRM      anotBHit  anotBRandom
## -0.9325506     0.3052995    2.1241920   -0.3995276    0.1919743   -0.1955495
## GenderM      indsay18z  indsay24z
## -0.2099522     0.0062400    0.0009692
## 
## 
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance = 0.05
## 
## Call:
## gvlma(x = lmrzno)
## 
##             Value p-value      Decision
## Global Stat      6.2274 0.18280 Assumptions acceptable.
## Skewness         3.2467 0.07157 Assumptions acceptable.
## Kurtosis         1.6129 0.20409 Assumptions acceptable.
## Link Function    1.0285 0.31051 Assumptions acceptable.
## Heteroscedasticity 0.3394 0.56017 Assumptions acceptable.
```

Table 5

```

summary(lmrzno)

##
## Call:
## lm(formula = values ~ Stress + Vowel + VRM + anotB + Gender +
##     ind + (1/Subject..), data = stdatz.noout)
##
## Residuals:
##    Min      1Q  Median      3Q     Max 
## -2.3398 -0.6066 -0.1712  0.5739  2.2032 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -0.9325506  0.8778952 -1.062  0.28957  
## Stress       0.3052995  0.5971116  0.511  0.60979  
## Vowel        2.1241920  0.7527664  2.822  0.00532 ** 
## VRM         -0.3995276  1.1547304 -0.346  0.72976  
## anotBHit    0.1919743  0.1779022  1.079  0.28202  
## anotBRandom -0.1955495  0.1688243 -1.158  0.24831  
## GenderM     -0.2099522  0.1443785 -1.454  0.14768  
## indsay18z   0.0062400  0.1685745  0.037  0.97051  
## indsay24z   0.0009692  0.1715231  0.006  0.99550  
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9461 on 176 degrees of freedom
##   (106 observations deleted due to missingness)
## Multiple R-squared:  0.07634,   Adjusted R-squared:  0.03436 
## F-statistic: 1.818 on 8 and 176 DF,  p-value: 0.07639

```

```

anova(lmrzno)

## Analysis of Variance Table
##
## Response: values
##             Df  Sum Sq Mean Sq F value    Pr(>F)    
## Stress       1  0.234  0.2343  0.2618 0.609530  
## Vowel        1  6.160  6.1600  6.8816 0.009473 ** 
## VRM          1  0.003  0.0032  0.0036 0.952446  
## anotB        2  4.729  2.3647  2.6416 0.074068 .  
## Gender       1  1.894  1.8935  2.1153 0.147613  
## ind          2  0.001  0.0007  0.0008 0.999212  
## Residuals 176 157.546  0.8951  
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
apa.reg.table(lmrzno, filename = "tab5.rtf", table.number = 5)

```

```

##
## Table 5
##
## Regression results using values as the criterion
##
##
```

```

## Predictor      b      b_95%_CI sr2   sr2_95%_CI          Fit
## (Intercept) -0.93 [-2.67, 0.80]
##      Stress   0.31 [-0.87, 1.48] .00 [-.01, .01]
##      Vowel  2.12** [0.64, 3.61] .04 [-.01, .10]
##      VRM    -0.40 [-2.68, 1.88] .00 [-.01, .01]
##      anotBHit  0.19 [-0.16, 0.54] .01 [-.02, .03]
##  anotBRandom -0.20 [-0.53, 0.14] .01 [-.02, .03]
##      GenderM -0.21 [-0.49, 0.07] .01 [-.02, .04]
##  indsday18z   0.01 [-0.33, 0.34] .00 [-.00, .00]
##  indsday24z   0.00 [-0.34, 0.34] .00 [-.00, .00]
##                                         R2 = .076
##                                         95% CI [.00,.12]
##
##
## Note. A significant b-weight indicates the semi-partial correlation is also significant.
## b represents unstandardized regression weights.
## sr2 represents the semi-partial correlation squared.
## Square brackets are used to enclose the lower and upper limits of a confidence interval.
## * indicates p < .05. ** indicates p < .01.
##

```

Figure 3

```

png("fig3.png")
mycol=c("black","darkgray","lightgray")
mypch=c(4,1,20)
names(mycol)=names(mypch)<-levels(factor(stdatz.noout$ind))
plot(values~Vowel,data=stdatz.noout,xlab="Preference quotient in the Vowel task",ylab="Vocabulary (z-score)",cex=1.5)
abline(lm(values~Vowel,data=stdatz.noout))
dev.off()

## pdf
## 2
png("fig4.png")
mycol=c("black","darkgray","lightgray")
mypch=c(4,1,20)
names(mycol)=names(mypch)<-levels(factor(stdatz.noout$ind))
plot(values~Stress,data=stdatz.noout,xlab="Preference quotient in the Stress task",ylab="Vocabulary (z-score)",cex=1.5)
abline(lm(values~Stress,data=stdatz.noout))
dev.off()

## pdf
## 2

```

Multivariate Bayesian model

```

main = brm(values ~
  total.tro.pref.quotient + Vowel.Alt.pref.quotient + Novelty.VRM + A.not.B.score +
  Gender + ind + (1 | Subject..), data = stdatz,
  prior = our_priors,
  iter=niter, warmup=nwarmup, chains=4, cores=2,
  seed=12,
  save_all_pars = T,

```

```

        sample_prior = T
    )
saveRDS(main,file="main_model.rds")

main_summary=summary(main)

saveRDS(main_summary,file="main_model_summary.rds")

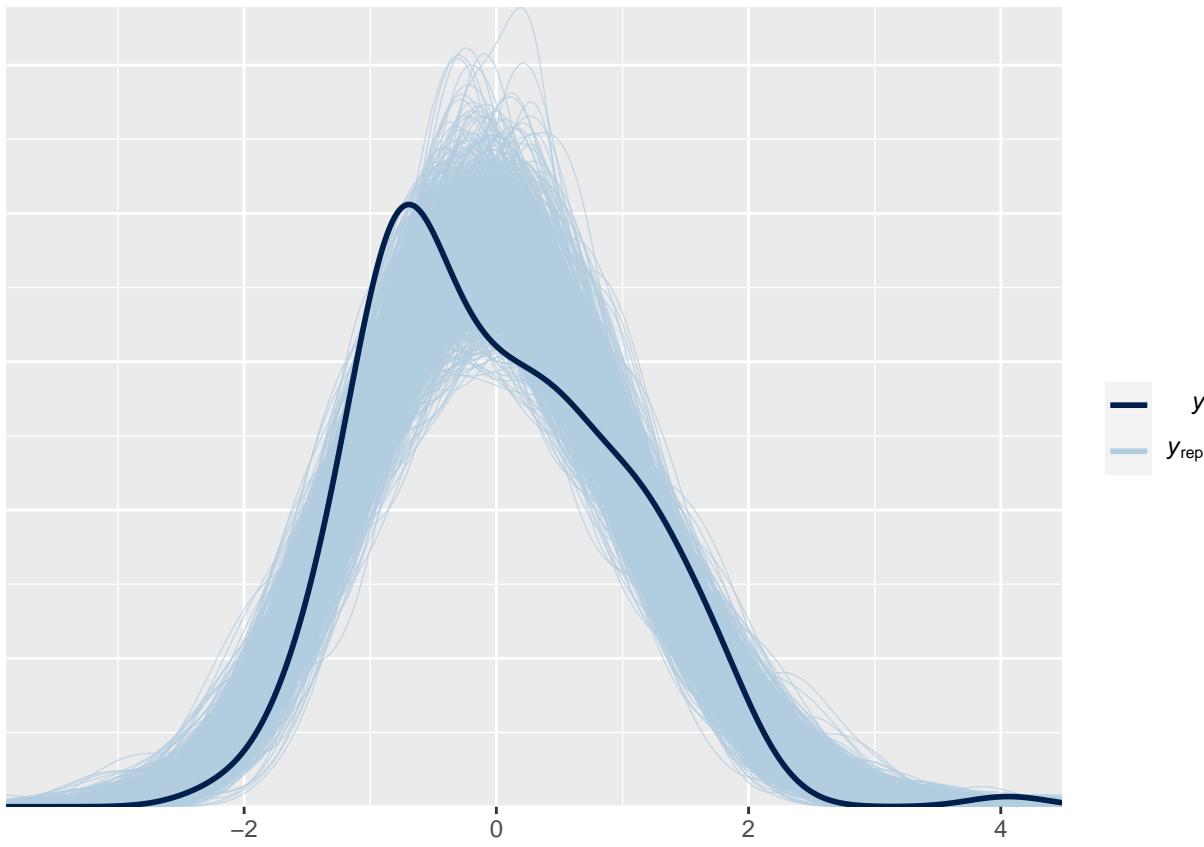
main_BF = bayesfactor_parameters(main, null = 0)

saveRDS(main_BF,file="main_model_bf.rds")

#plot(main)

pp_check(main, nsamples = 1000)

```



```
main
```

```

##  Family: gaussian
##  Links: mu = identity; sigma = identity
## Formula: values ~ total.tro.pref.quotient + Vowel.Alt.pref.quotient + Novelty.VRM + A.not.B.score +
##            Data: stdatz (Number of observations: 185)
## Samples: 4 chains, each with iter = 4000; warmup = 500; thin = 1;
##           total post-warmup samples = 14000
##
## Group-Level Effects:
##   ~Subject.. (Number of levels: 68)
##             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS

```

```

## sd(Intercept)      0.80      0.09      0.64      0.99 1.00      4483      8262
##
## Population-Level Effects:
##                               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## Intercept                  -0.55      0.87     -2.28      1.14 1.00      7559
## total.tro.pref.quotient    0.38      0.70     -0.94      1.82 1.00      7414
## Vowel.Alt.pref.quotient   1.06      0.94     -0.60      3.09 1.00      5956
## Novelty.VRM                -0.18      0.99     -2.20      1.79 1.00      9721
## A.not.B.score2             -0.12      0.25     -0.61      0.38 1.00      4807
## A.not.B.score3              0.16      0.26     -0.34      0.67 1.00      4791
## GenderM                     -0.19      0.21     -0.61      0.22 1.00      4968
## indsay18z                   0.04      0.11     -0.17      0.25 1.00      19170
## indsay24z                   -0.02      0.11     -0.24      0.20 1.00      18908
##                               Tail_ESS
## Intercept                  9703
## total.tro.pref.quotient   8596
## Vowel.Alt.pref.quotient  7178
## Novelty.VRM                8943
## A.not.B.score2             7238
## A.not.B.score3              7569
## GenderM                     8018
## indsay18z                  10329
## indsay24z                  10727
##
## Family Specific Parameters:
##                               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma          0.60      0.04      0.53      0.69 1.00      10228      10728
##
## Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
main_BF

## # Bayes Factor (Savage-Dickey density ratio)
##
## Parameter | BF
## -----
## Intercept | 0.460
## total.tro.pref.quotient | 0.646
## Vowel.Alt.pref.quotient | 1.488
## Novelty.VRM | 0.827
## A.not.B.score2 | 0.253
## A.not.B.score3 | 0.288
## GenderM | 0.294
## indsay18z | 0.104
## indsay24z | 0.101
##
## * Evidence Against The Null: [0]
plot(main_BF)

```

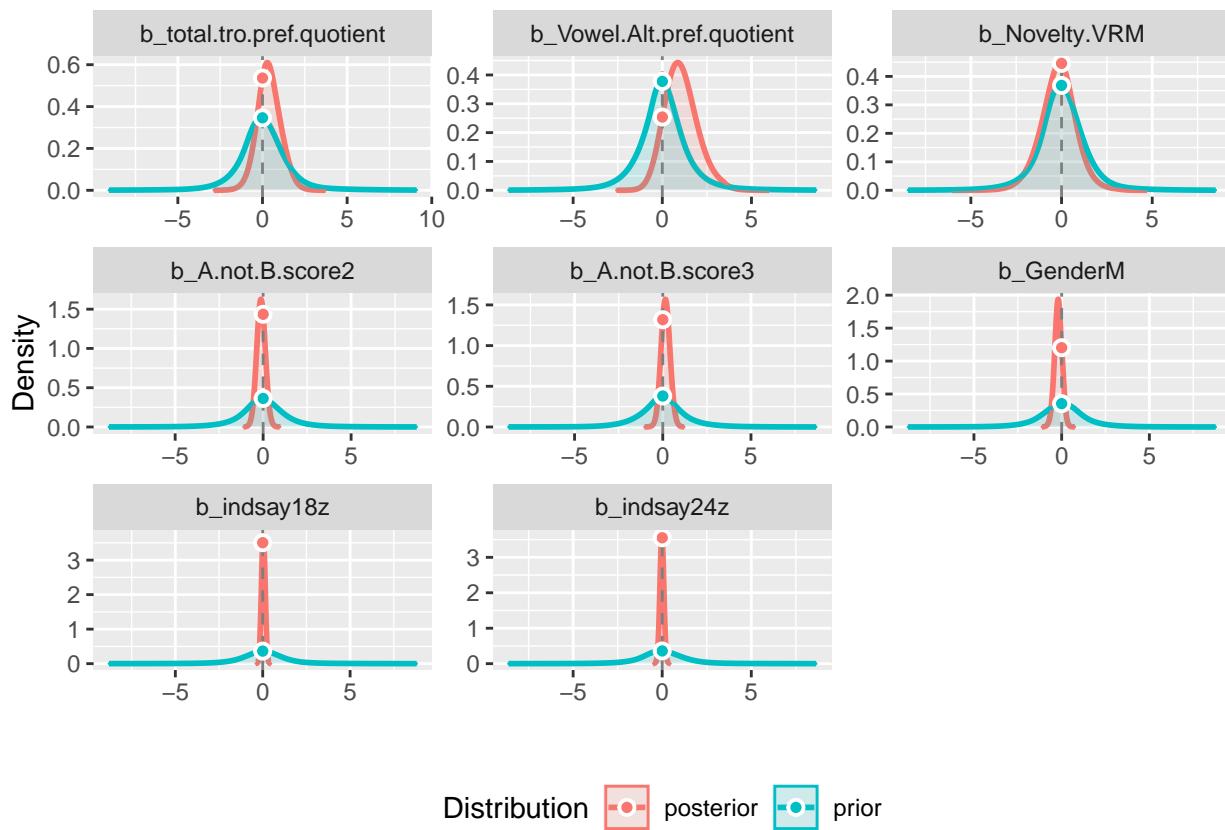


Table 6

```

mytab=cbind(main_summary$fixed[, "Estimate"],main_summary$fixed[, "1-95% CI"],main_summary$fixed[, "u-95% CI"])
mytab=round(mytab,3)
colnames(mytab)<-c("B", "LCI", "HCI", "BF10")
write.table(mytab, "tab6.txt", sep="\t")

```