Lecture 5

Random Effects Models

Colin Rundel

02/01/2017

Random Effects Models

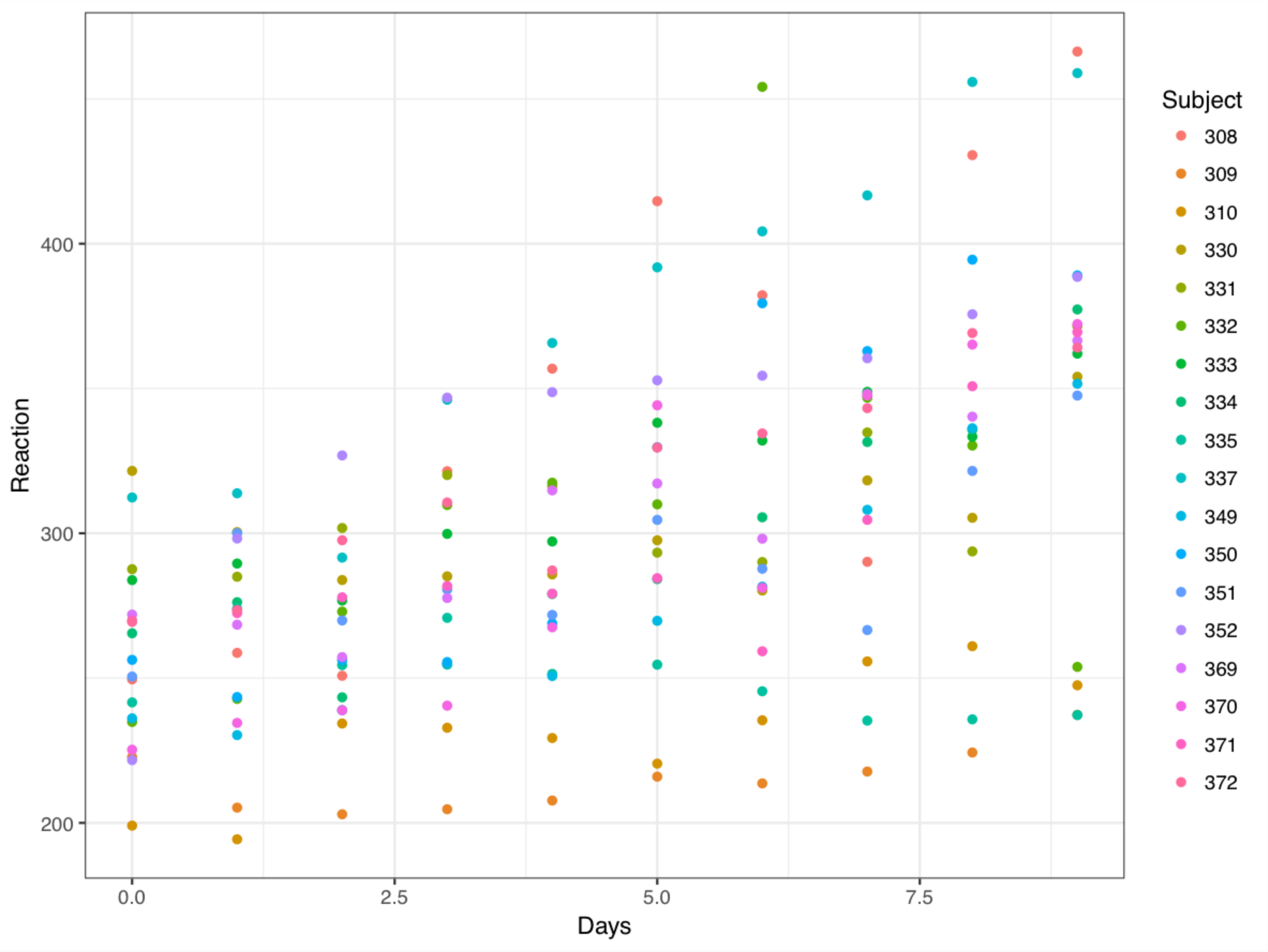
Sleep Study Data

The average reaction time per day for subjects in a sleep deprivation study. On day 0 the subjects had their normal amount of sleep. Starting that night they were restricted to 3 hours of sleep per night. The observations represent the average reaction time on a series of tests given each day to each subject.

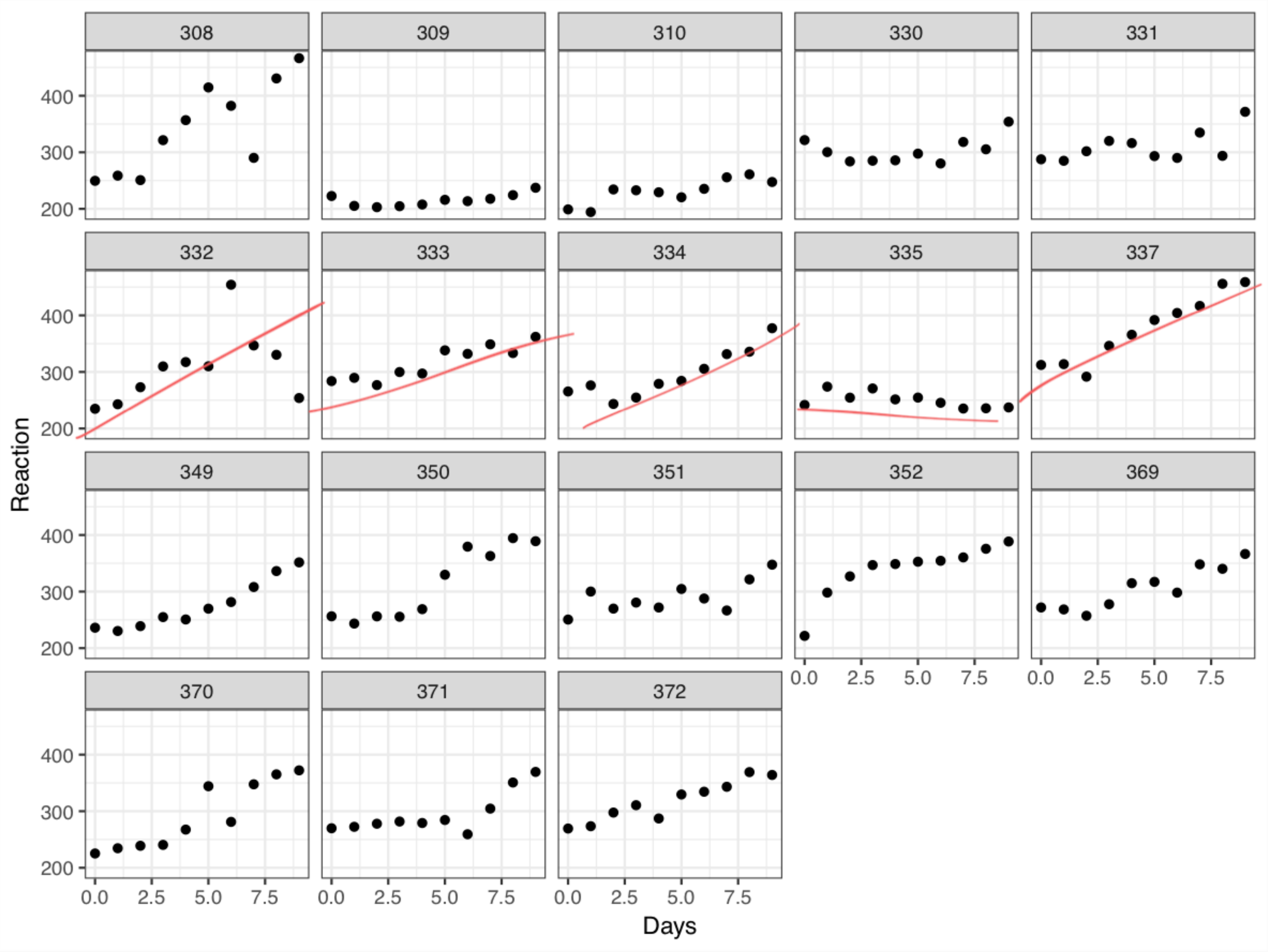
```
library(lme4)

sleepstudy %>% tbl_df()
## # A tibble: 180 × 3
##    Reaction  Days Subject
##       <dbl> <dbl>  <fctr>
## 1  249.5600     0     308
## 2  258.7047     1     308
## 3  250.8006     2     308
## 4  321.4398     3     308
## 5  356.8519     4     308
## 6  414.6901     5     308
## 7  382.2038     6     308
## 8  290.1486     7     308
## 9  430.5853     8     308
## 10 466.3535     9     308
## # ... with 170 more rows
```

EDA


Reaction
400
300
200
0.0
2.5
5.0
7.5
Days
Subject
308
309
310
330
331
332
333
334
335
337
349
350
351
352
369
370
371
372

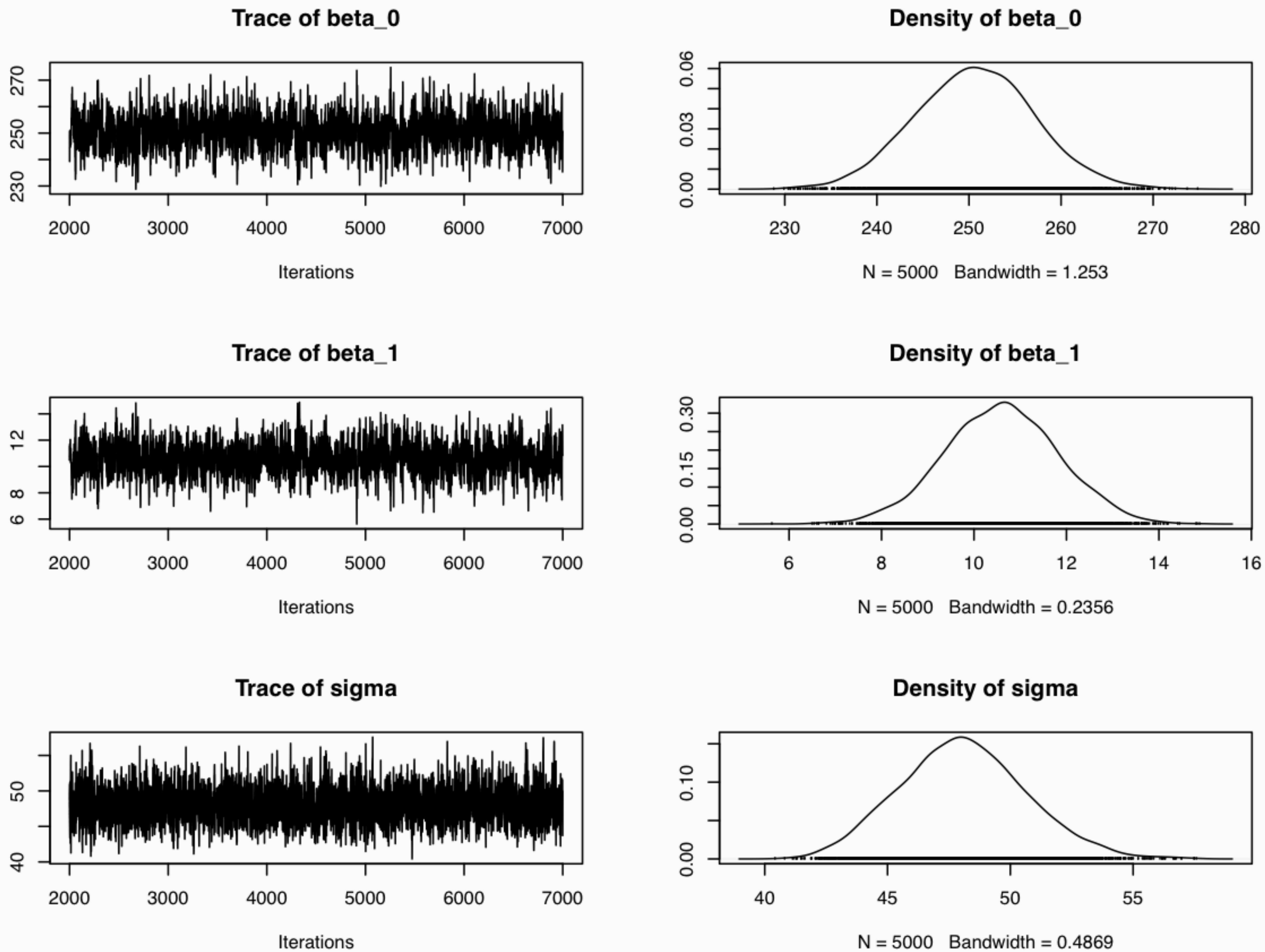
EDA (small multiples)



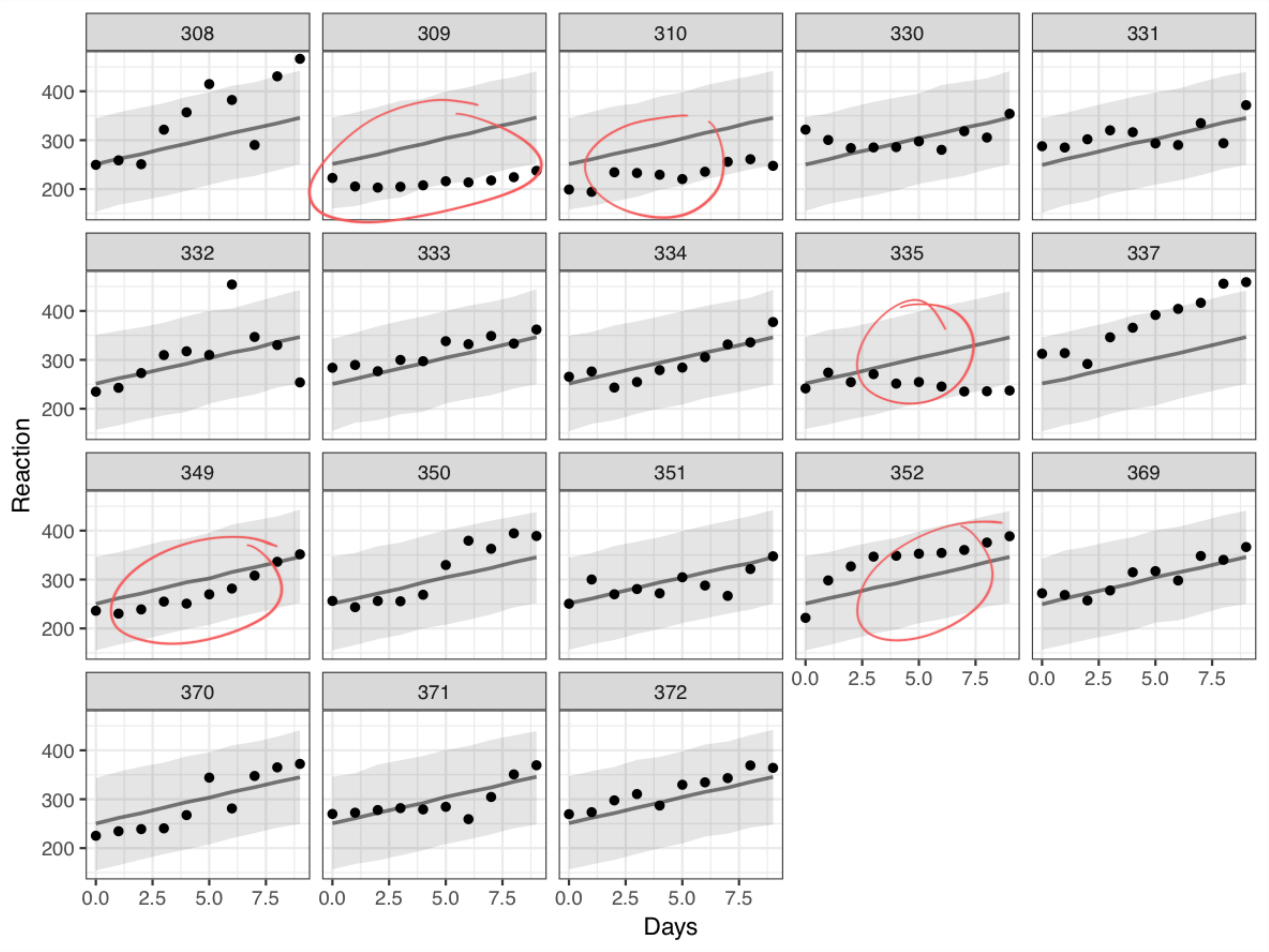
Bayesian Linear Model

```
## model{
##   # Likelihood
##   for(i in 1:length(Reaction)){
##     Reaction[i] ~ dnorm(mu[i],tau2)
##     mu[i] <- beta_0 + beta_1*Days[i]
##
##     Y_hat[i] ~ dnorm(mu[i],tau2)
##   }
##
##   # Prior for beta
##   beta_0 ~ dnorm(0,1/10000)
##   beta_1 ~ dnorm(0,1/10000)
##
##   # Prior for sigma / tau2
##   sigma ~ dunif(0, 100)
##   tau2 <- 1/(sigma*sigma)
## }
```

MCMC Diagnostics

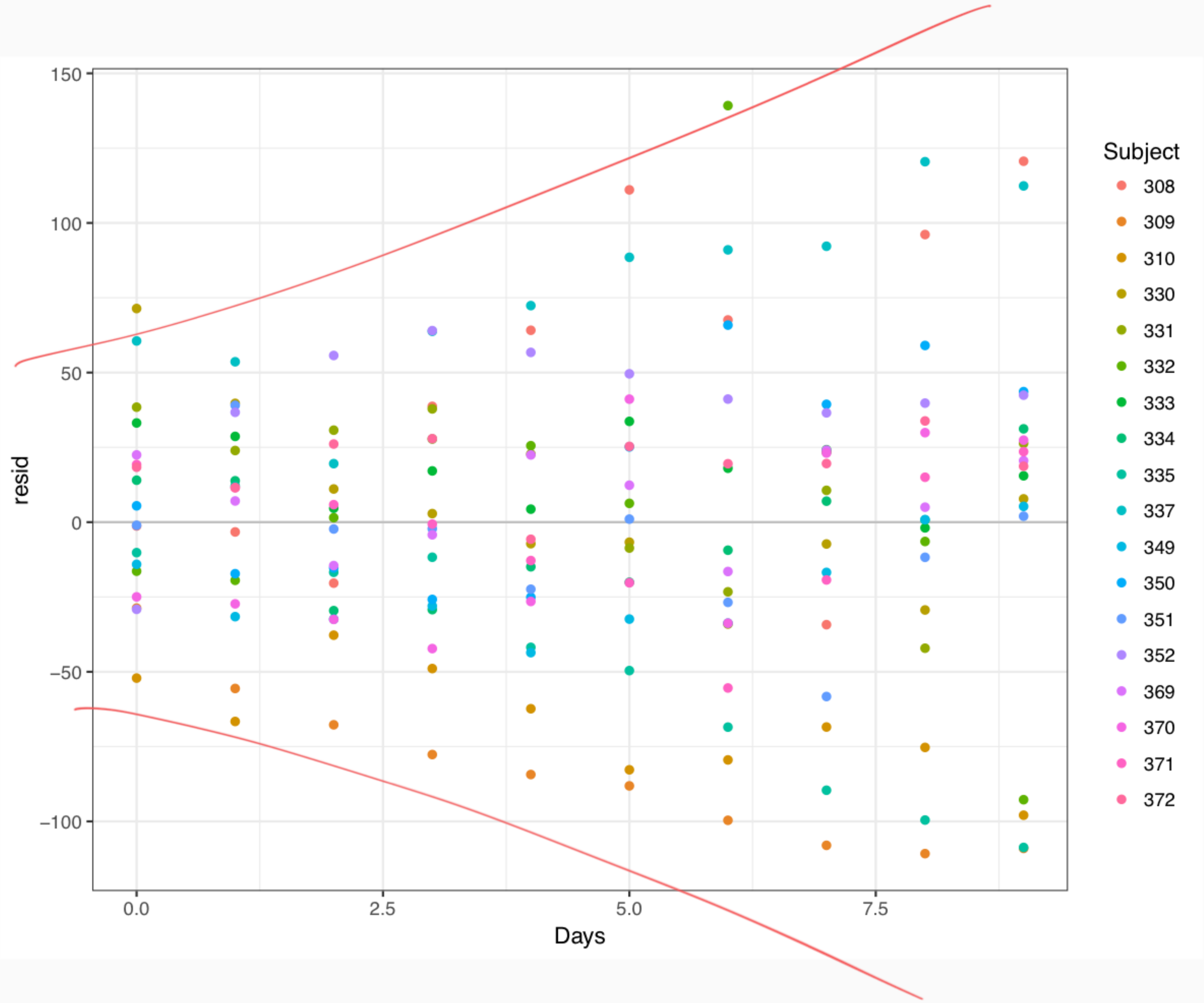


Model fit

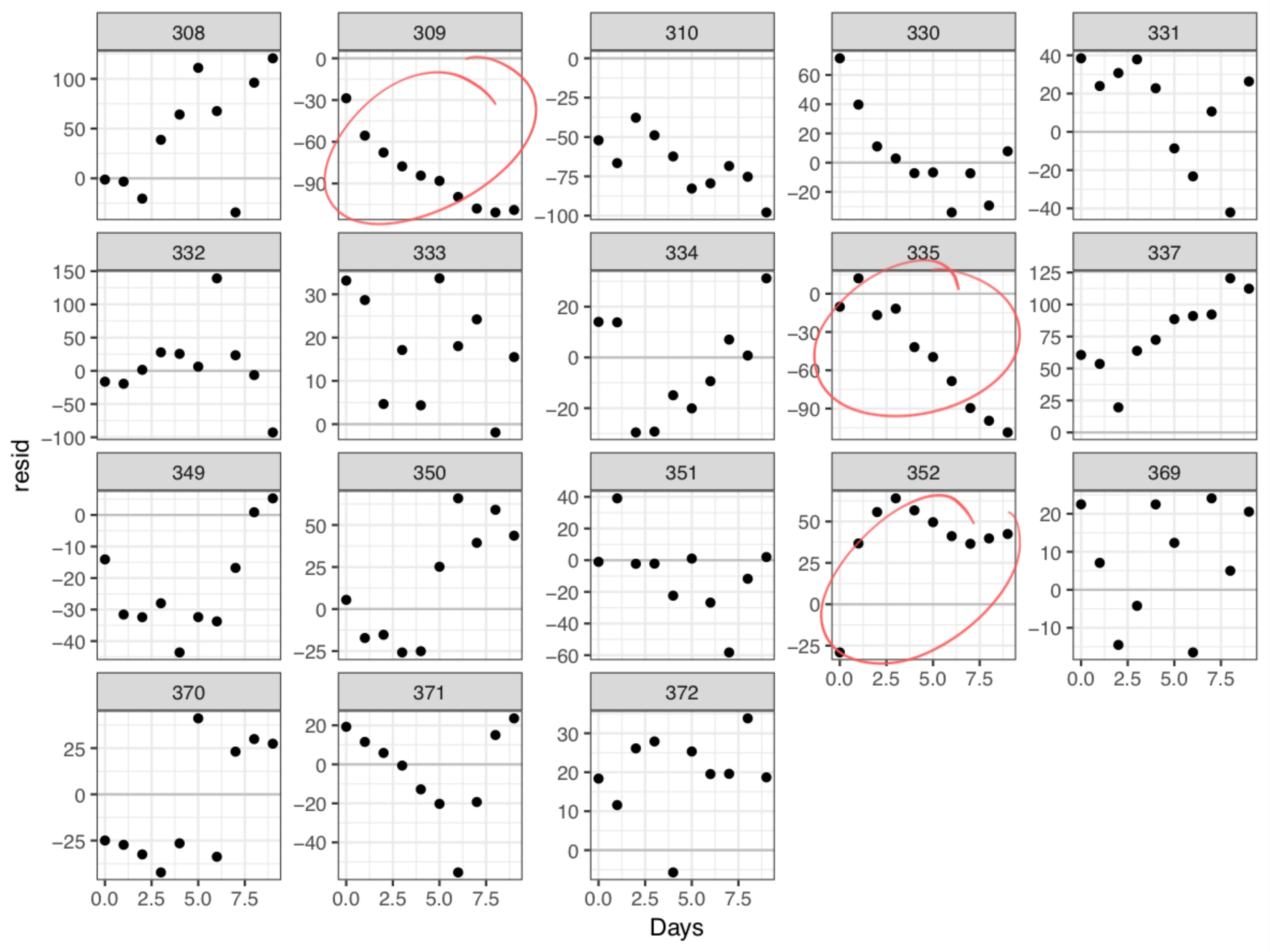


308
309
310
330
331
332
333
334
335
337
349
350
351
352
369
370
371
372
Reaction
Days
200
300
400
0.0
2.5
5.0
7.5

Residuals


150
100
50
0
−50
−100
resid
0.0
2.5
5.0
7.5
Days
Subject
308
309
310
330
331
332
333
334
335
337
349
350
351
352
369
370
371
372

Residuals by subject



Random Intercept Model

Data coding

```
sleepstudy = sleepstudy %>%
  mutate(Subject_index = as.integer(Subject))

sleepstudy[c(1:2,11:12,21:22,31:32),]
##    Reaction Days Subject Subject_index
## 1  249.5600    0     308             1
## 2  258.7047    1     308             1
## 11 222.7339    0     309             2
## 12 205.2658    1     309             2
## 21 199.0539    0     310             3
## 22 194.3322    1     310             3
## 31 321.5426    0     330             4
## 32 300.4002    1     330             4
```

Random Intercept Model

Let i represent each observation and $j(i)$ be subject in oberservation i then

$$Y_i = \alpha_{j(i)} + \beta_1 \times \text{Days} + \epsilon_i$$

$$\alpha_j \sim \mathcal{N}(\beta_0,\ \sigma_\alpha^2)$$
$$\epsilon_i \sim \mathcal{N}(0,\ \sigma^2)$$

$$\beta_0, \beta_1 \sim \mathcal{N}(0, 10000)$$
$$\sigma, \sigma_\alpha \sim \text{Unif}(0, 100)$$

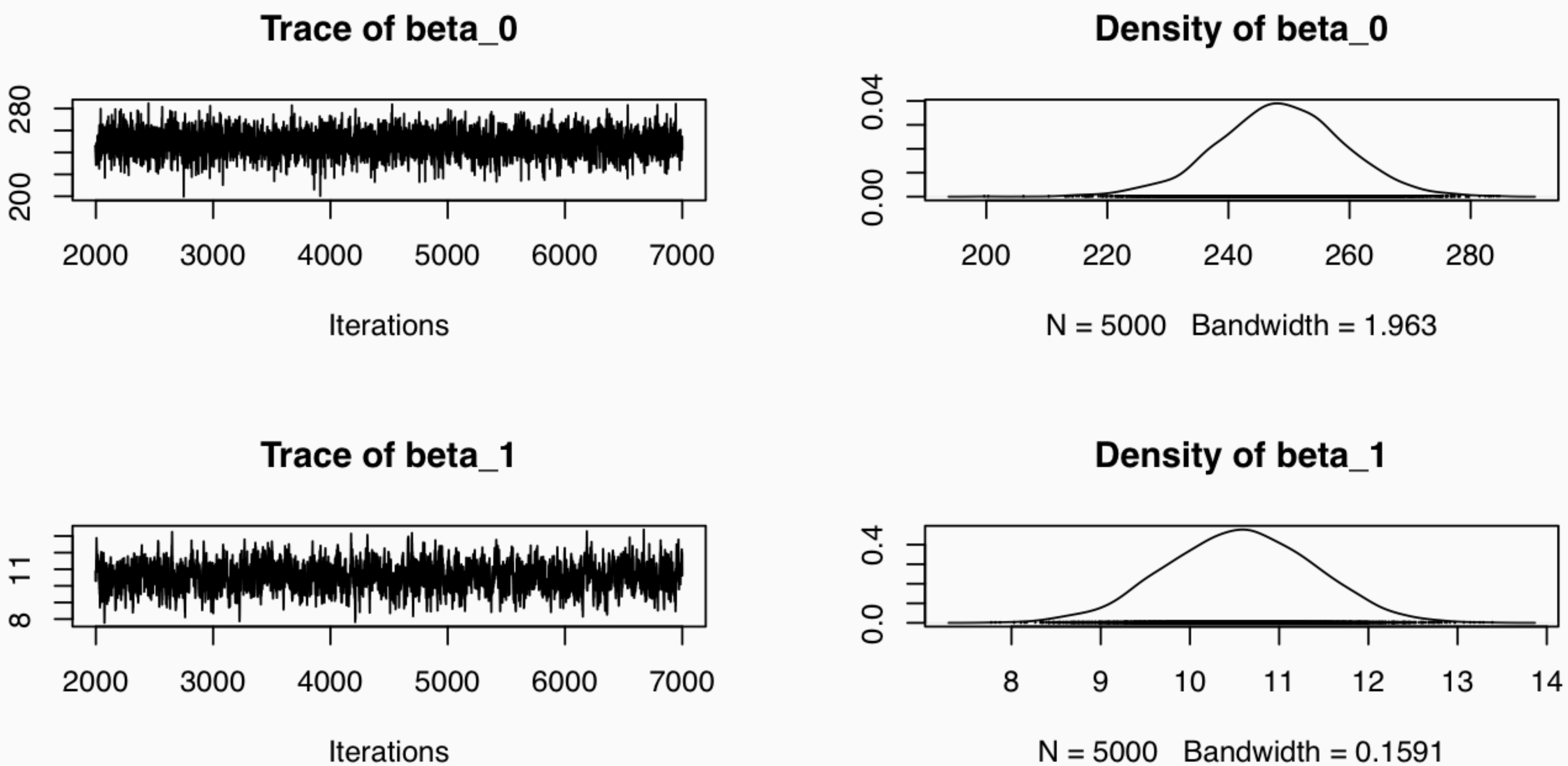
$$Y_i = \beta_0 + \beta_1 \, \text{days}_i + \alpha_{j(i)} + \epsilon_i$$

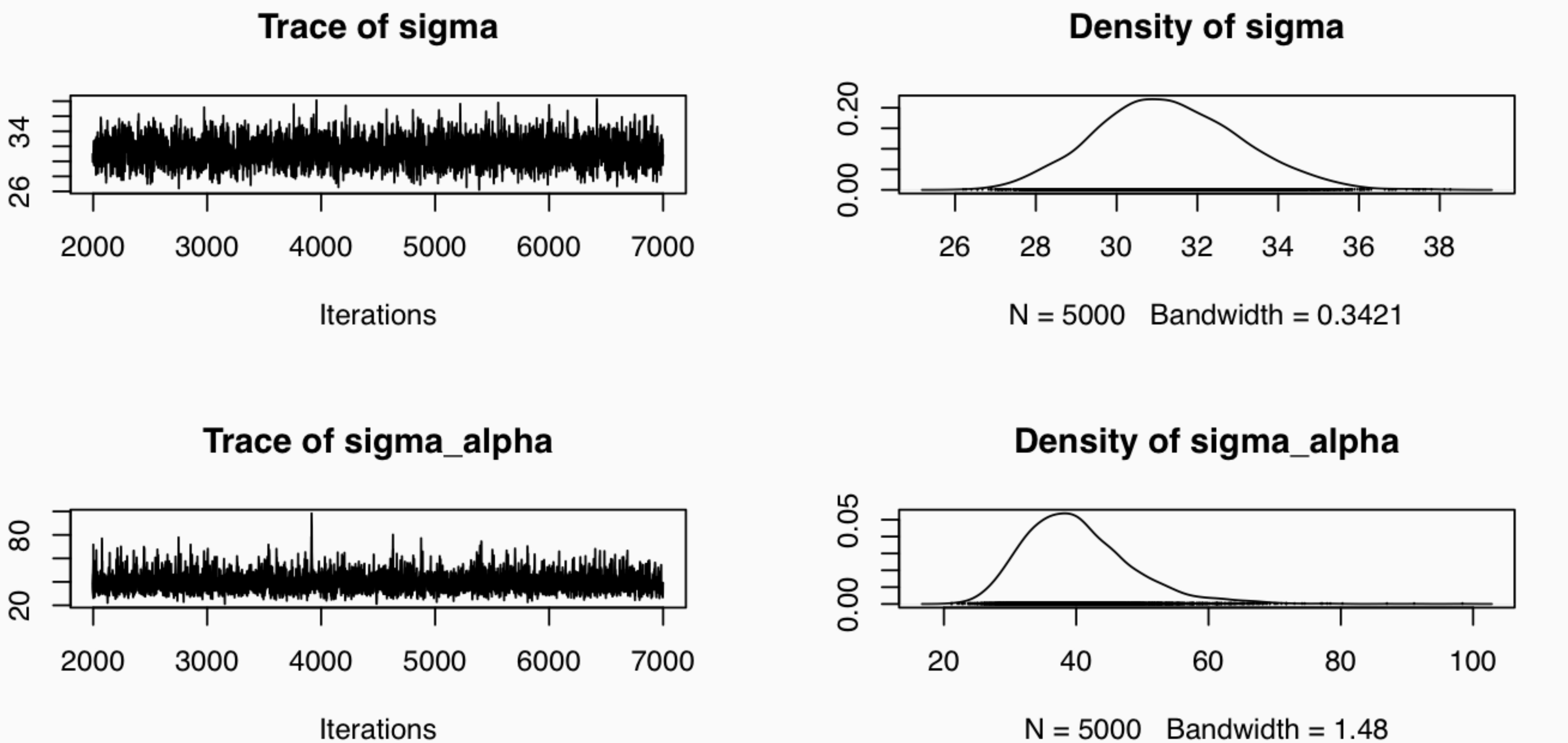
$$\alpha_{j(i)} \sim N(0, \sigma_\alpha^2)$$

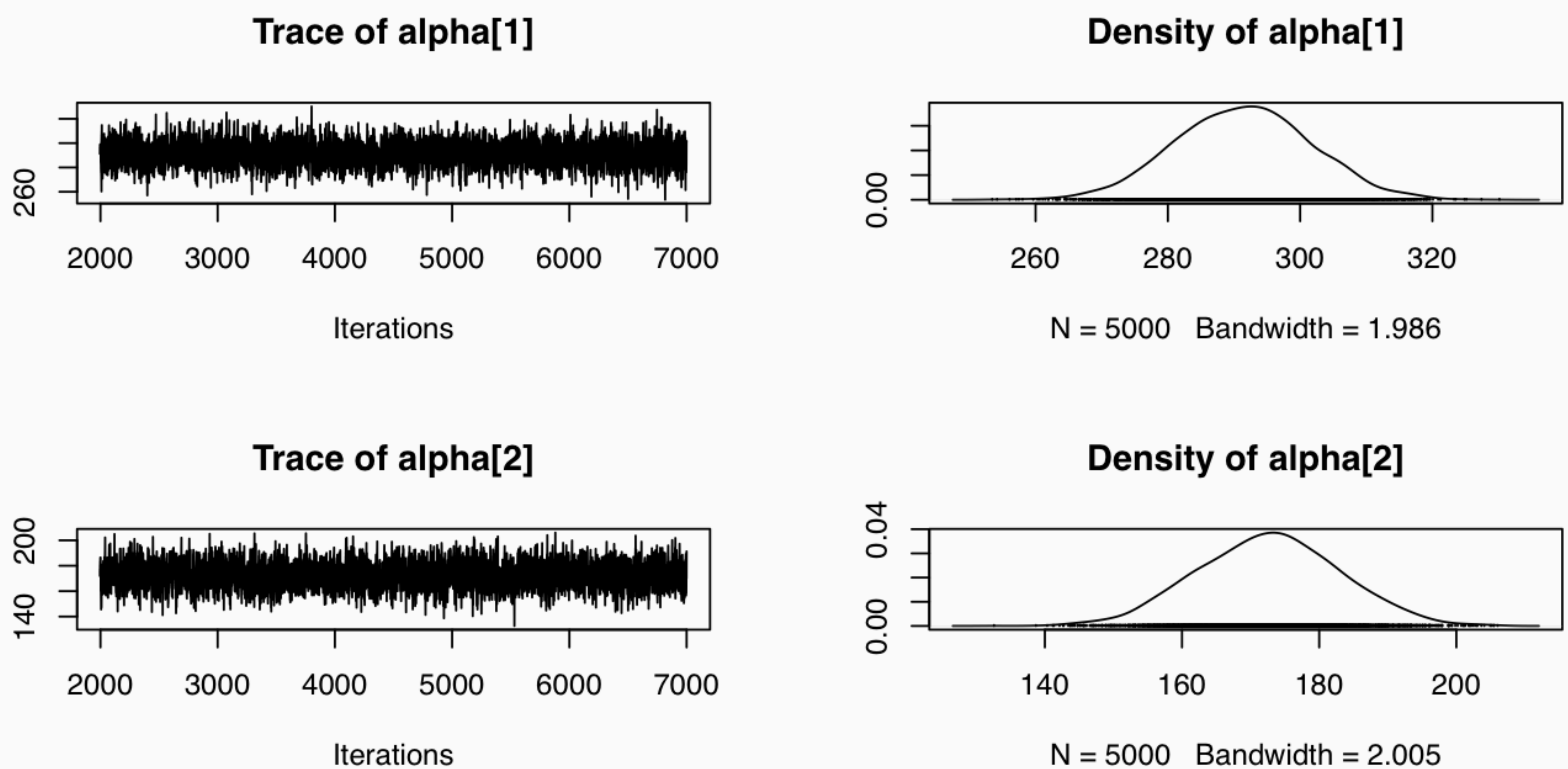
Random Intercept Model - JAGS

```
## model{
##   for(i in 1:length(Reaction)) {
##     Reaction[i] ~ dnorm(mu[i],tau2)
##     mu[i] <- alpha[Subject_index[i]] + beta_1*Days[i]
##
##     Y_hat[i] ~ dnorm(mu[i],tau2)
##   }
##
##   for(j in 1:18) {
##     alpha[j] ~ dnorm(beta_0, tau2_alpha)
##   }
##
##   sigma_alpha ~ dunif(0, 100)
##   tau2_alpha <- 1/(sigma_alpha*sigma_alpha)
##
##   beta_0 ~ dnorm(0,1/10000)
##   beta_1 ~ dnorm(0,1/10000)
##
##   sigma ~ dunif(0, 100)
##   tau2 <- 1/(sigma*sigma)
## }
```

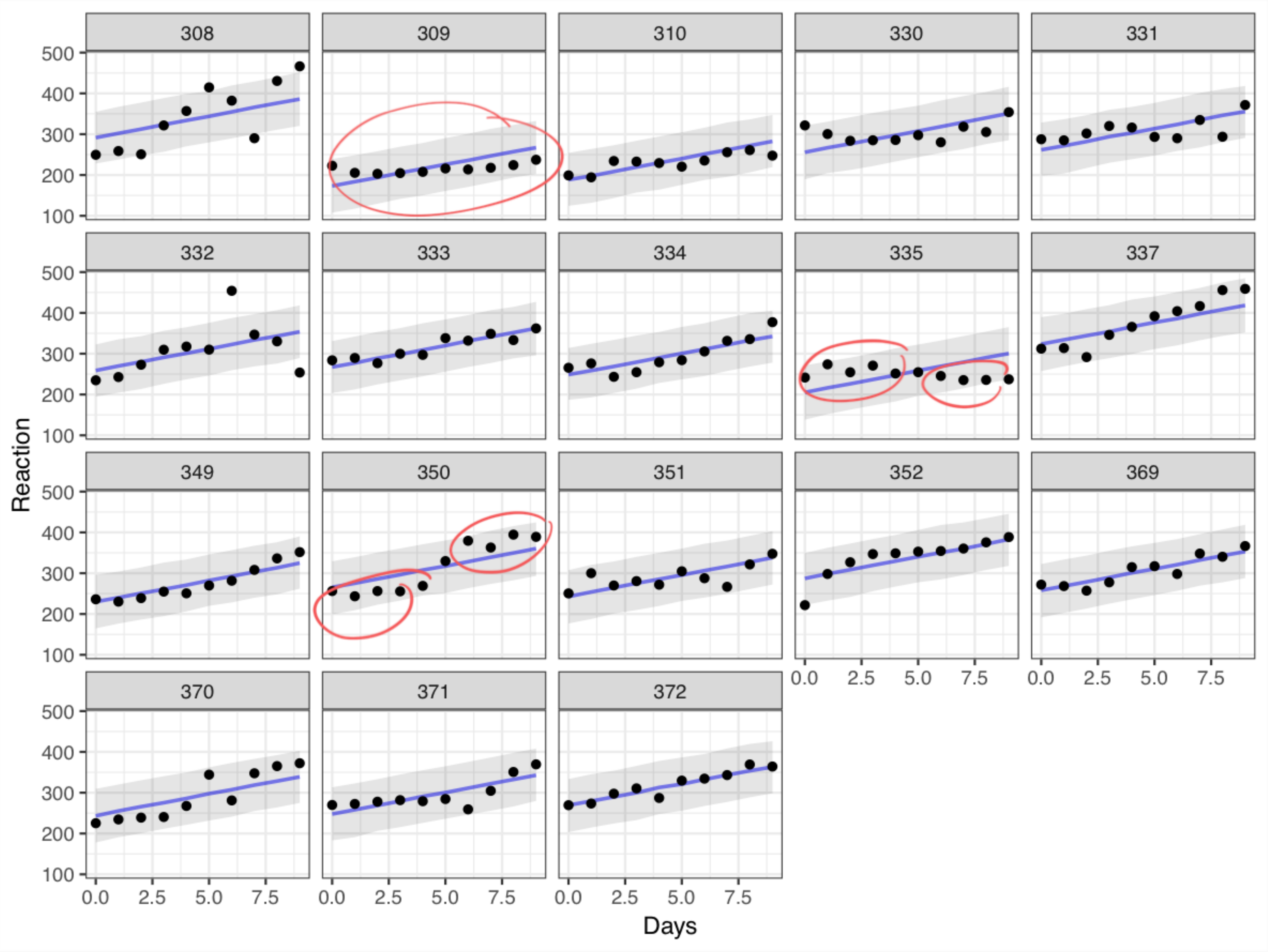
MCMC Diagnostics



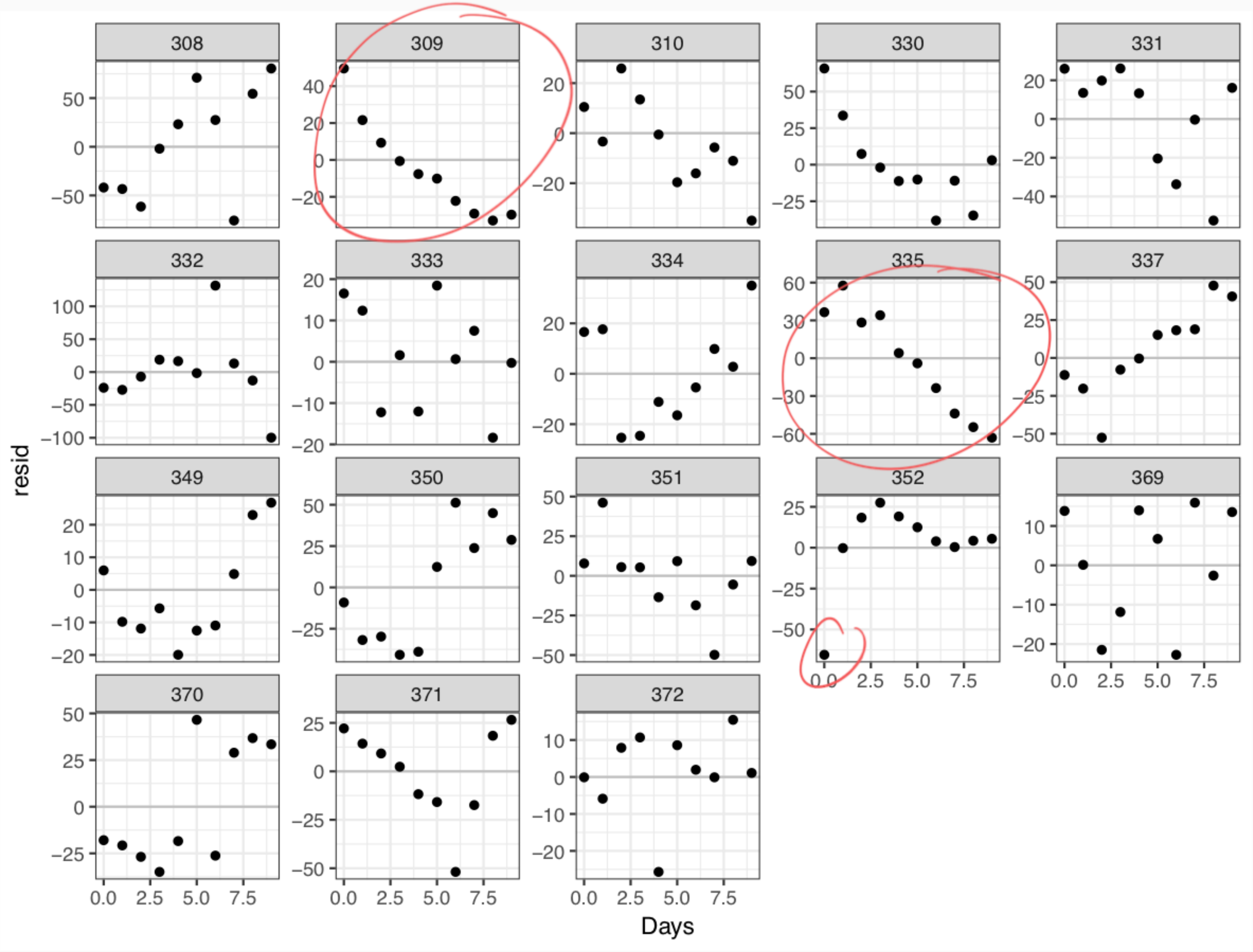

Trace of sigma
26
34
2000
3000
4000
5000
6000
7000
Iterations
Density of sigma
0.00
0.20
26
28
30
32
34
36
38
N = 5000 Bandwidth = 0.3421
Trace of sigma_alpha
20
80
2000
3000
4000
5000
6000
7000
Iterations
Density of sigma_alpha
0.00
0.05
20
40
60
80
100
N = 5000 Bandwidth = 1.48


Trace of alpha[1]
260
2000
3000
4000
5000
6000
7000
Iterations
Density of alpha[1]
0.00
260
280
300
320
N = 5000 Bandwidth = 1.986
Trace of alpha[2]
200
140
2000
3000
4000
5000
6000
7000
Iterations
Density of alpha[2]
0.04
0.00
140
160
180
200
N = 5000 Bandwidth = 2.005

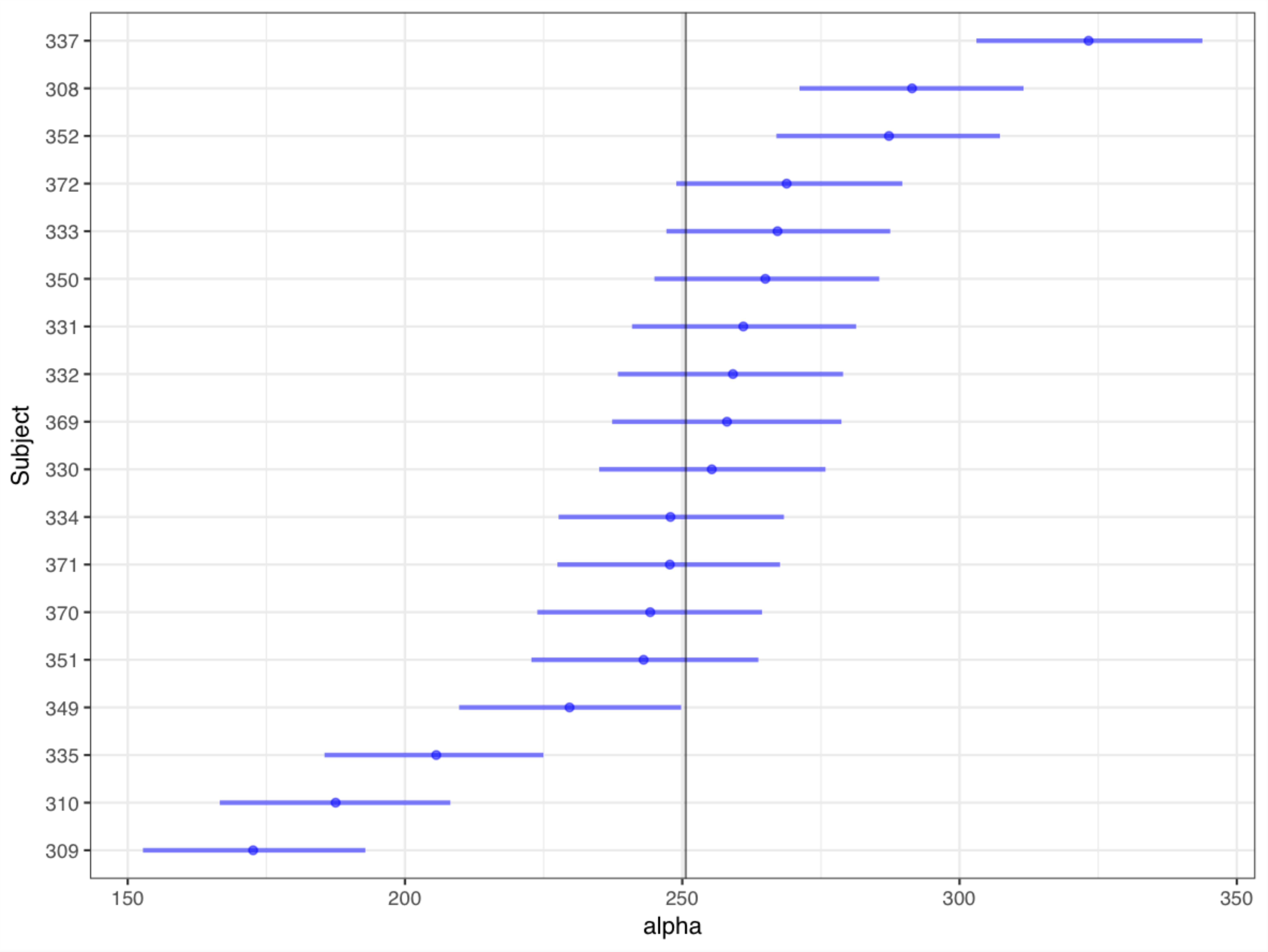
Model fit


308
309
310
330
331
332
333
334
335
337
349
350
351
352
369
370
371
372
Reaction
Days
500
400
300
200
100
0.0
2.5
5.0
7.5

Residuals by subject



Random effects

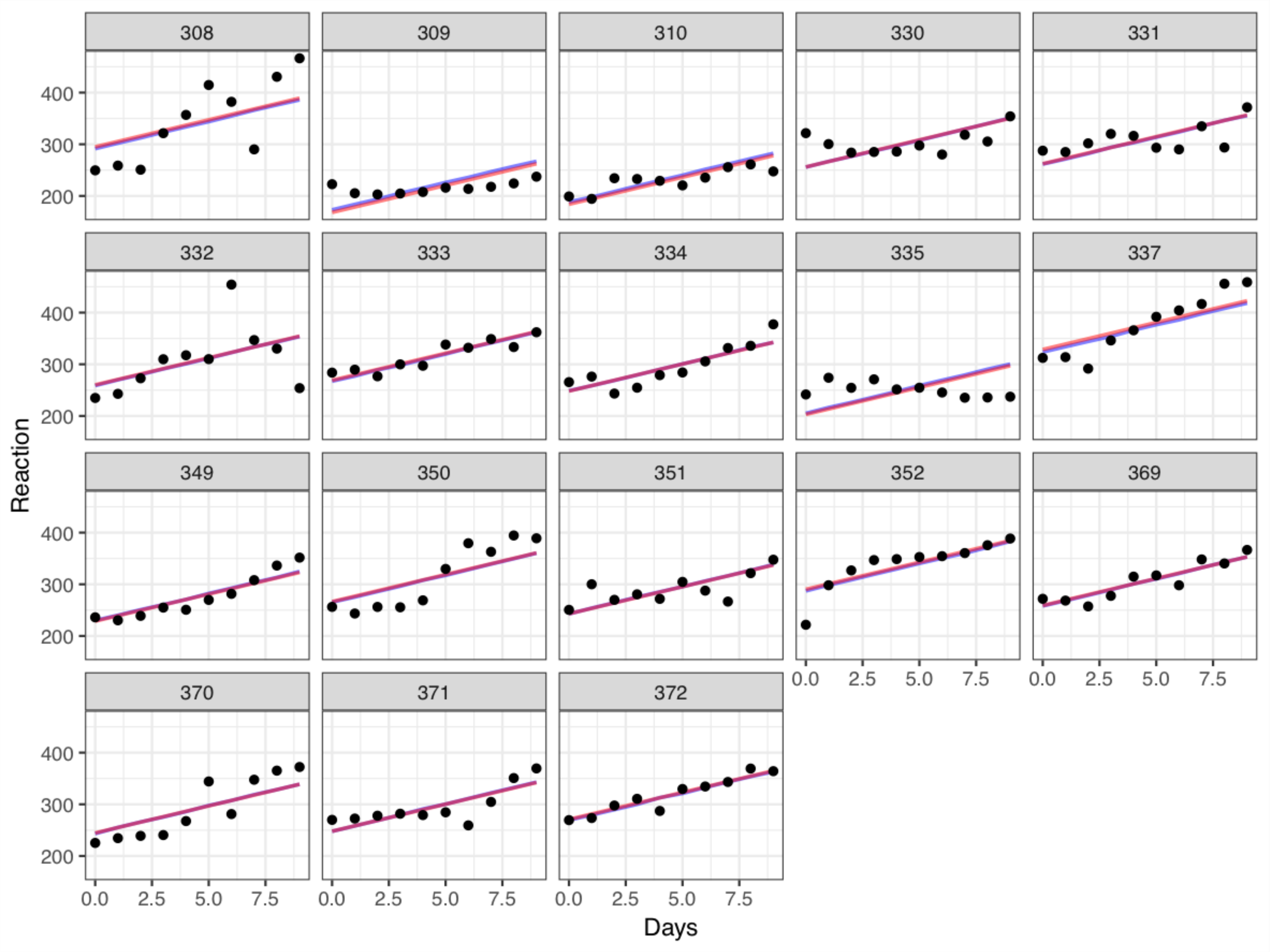


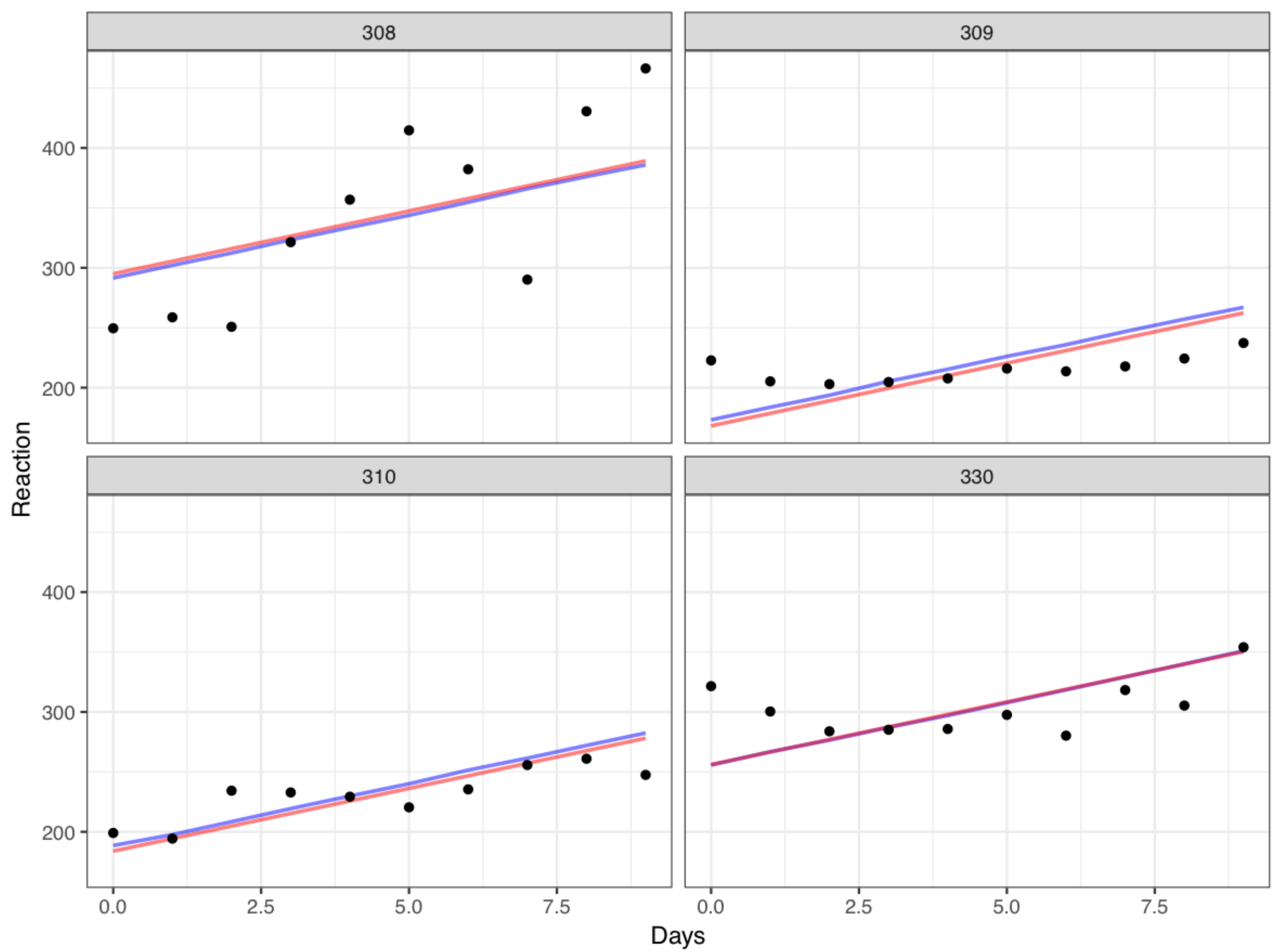
Why not a fixed effect for Subject?

Not going to bother with the Bayesian model here because of all the dummy coding and betas … . . .

```
l = lm(Reaction ~ Days + Subject - 1, data=sleepstudy)
summary(l)
##
## Call:
## lm(formula = Reaction ~ Days + Subject - 1, data = sleepstudy)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -100.540  -16.389   -0.341   15.215  131.159
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## Days         10.4673     0.8042   13.02   <2e-16 ***
## Subject308 295.0310    10.4471   28.24   <2e-16 ***
## Subject309 168.1302    10.4471   16.09   <2e-16 ***
## Subject310 183.8985    10.4471   17.60   <2e-16 ***
## Subject330 256.1186    10.4471   24.52   <2e-16 ***
## Subject331 262.3333    10.4471   25.11   <2e-16 ***
## Subject332 260.1993    10.4471   24.91   <2e-16 ***
## Subject333 269.0555    10.4471   25.75   <2e-16 ***
## Subject334 248.1993    10.4471   23.76   <2e-16 ***
```

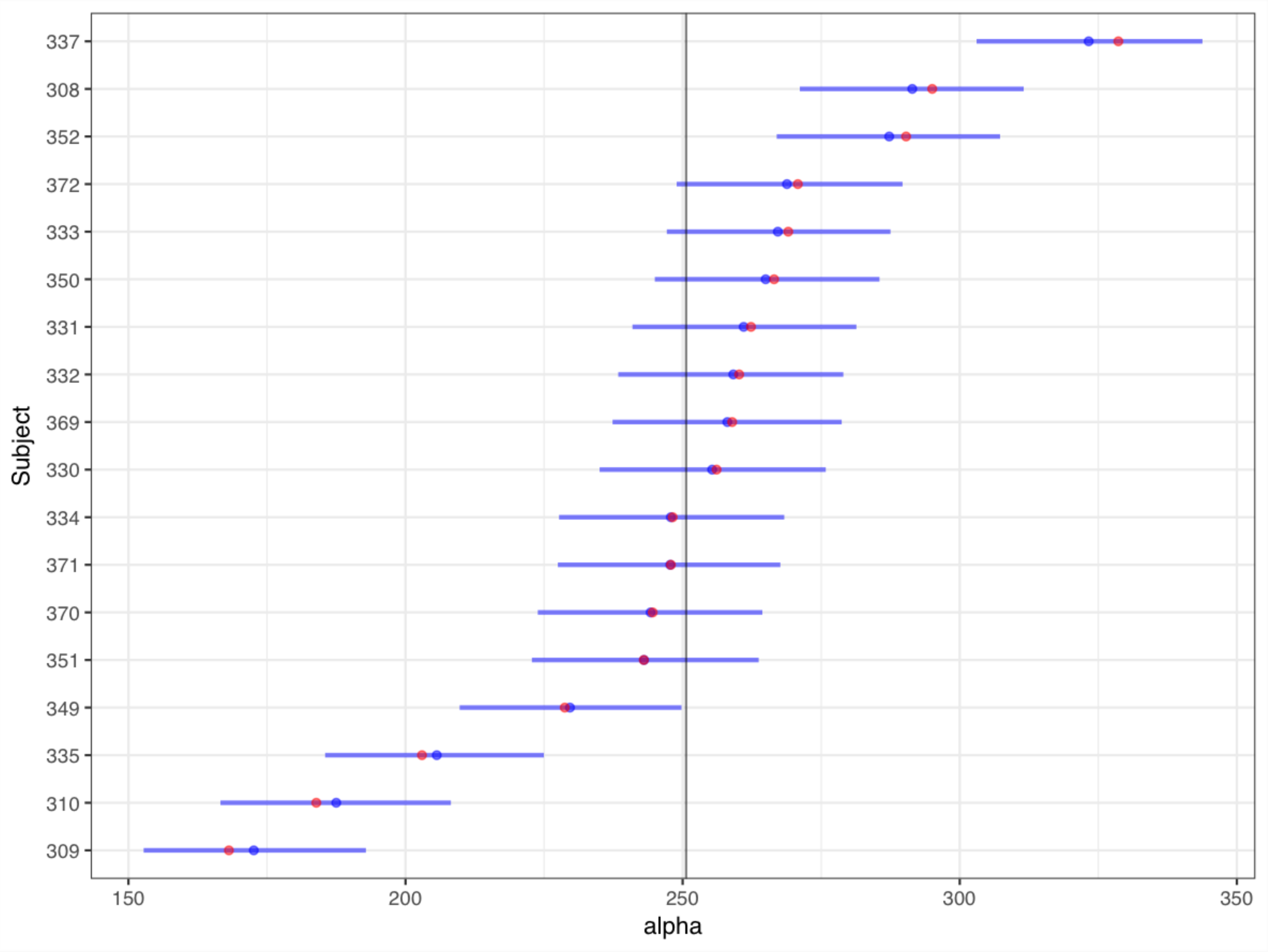
Comparing Model fit





308
309
310
330
Reaction
Days
400
300
200
0.0
2.5
5.0
7.5

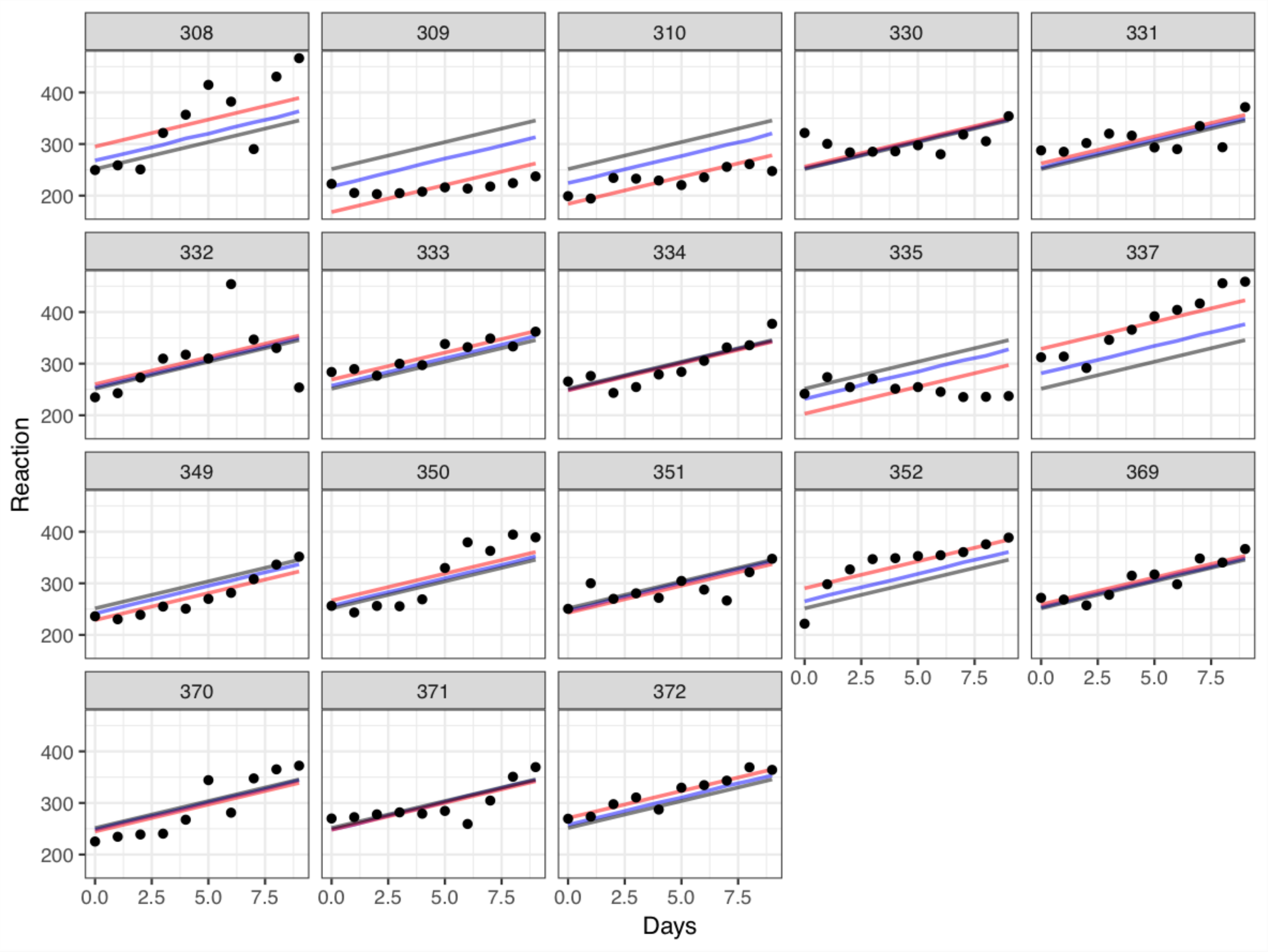
Random effects vs fixed effects

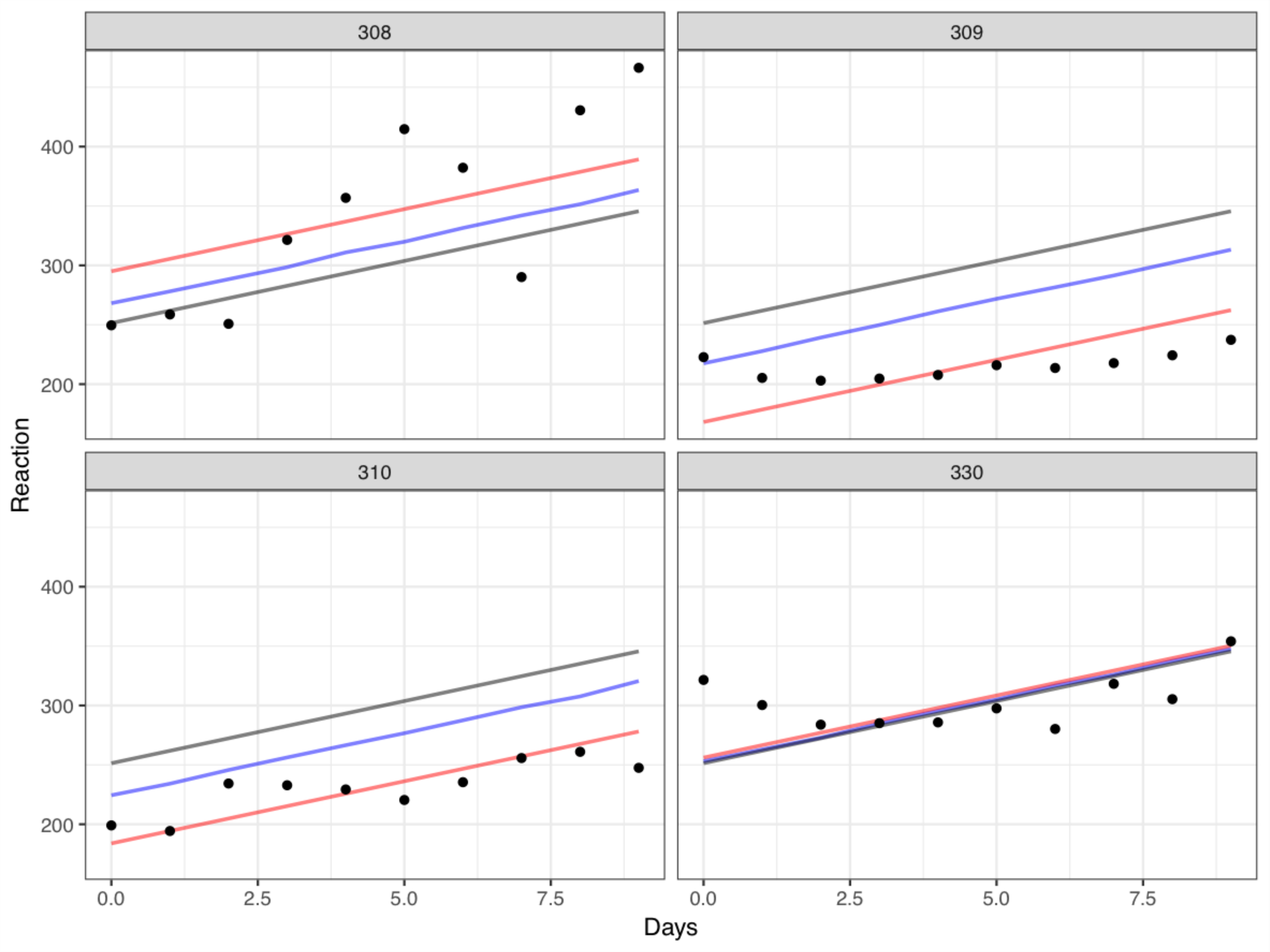


Random Intercept Model (Informative prior for σ_α)

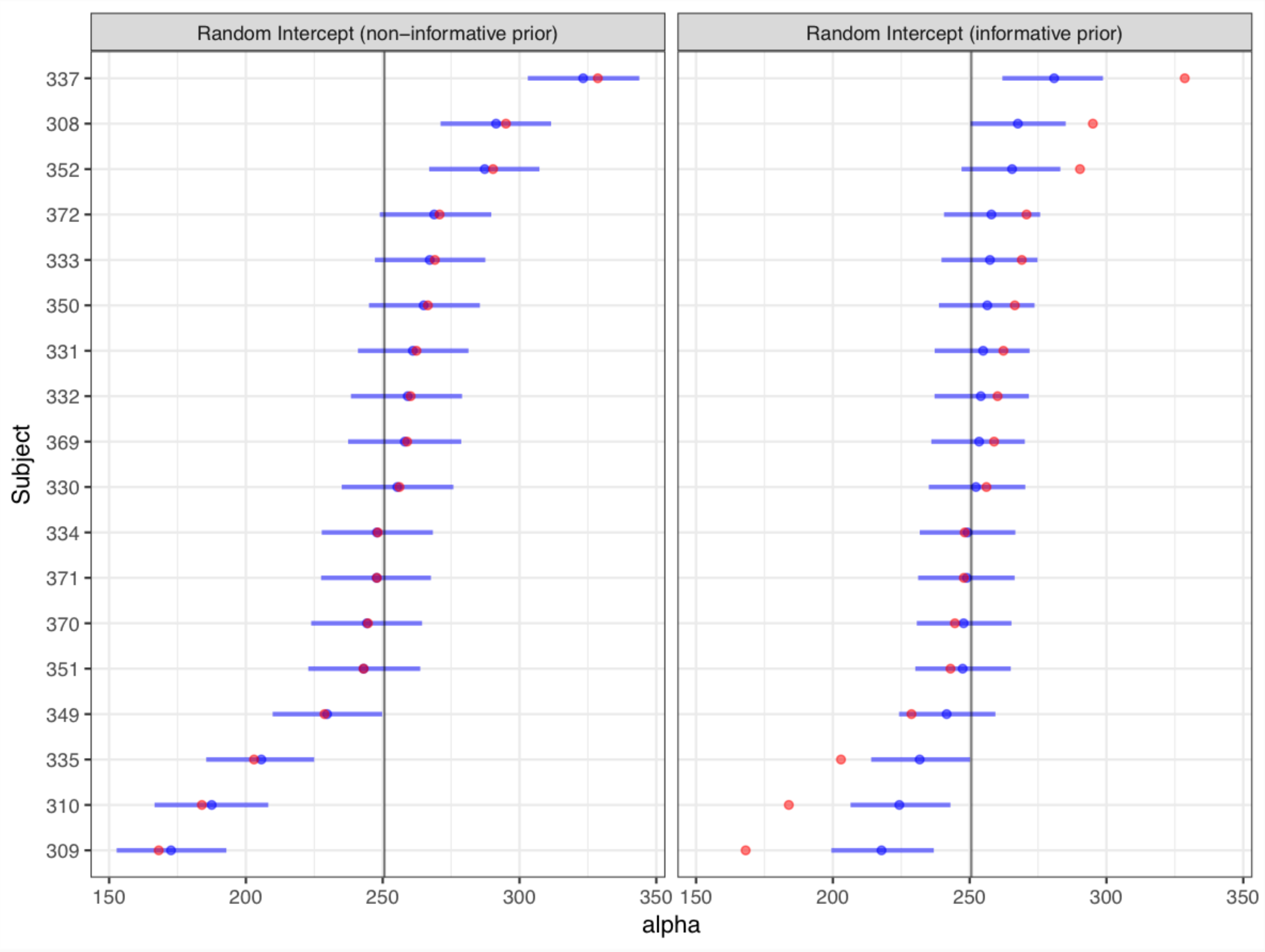
```
## model{
##   for(i in 1:length(Reaction)) {
##     Reaction[i] ~ dnorm(mu[i],tau2)
##     mu[i] <- alpha[Subject_index[i]] + beta_1*Days[i]
##
##     Y_hat[i] ~ dnorm(mu[i],tau2)
##   }
##
##   for(j in 1:18) {
##     alpha[j] ~ dnorm(beta_0, tau2_alpha)
##   }
##
##   sigma_alpha ~ dunif(0, 10)
##   tau2_alpha <- 1/(sigma_alpha*sigma_alpha)
##
##   beta_0 ~ dnorm(0,1/10000)
##   beta_1 ~ dnorm(0,1/10000)
##
##   sigma ~ dunif(0, 100)
##   tau2 <- 1/(sigma*sigma)
## }
```

Comparing Model fit (Constrainged α)




308
309
310
330
Reaction
Days
400
300
200
0.0
2.5
5.0
7.5

Prior Effect on α



Some Distribution Theory (about $Y \mid \beta_0, \beta_1, \sigma, \sigma_\alpha$)

$$Y_i = \beta_0 + \beta_1 \, day_i + \alpha_{j(i)} + \varepsilon_i$$

$$\alpha_{j(i)} \sim N(0, \sigma_\alpha^2) \qquad \varepsilon_i \sim N(0, \sigma^2)$$

$$\underline{Y} \mid \cdot \;\sim\; ? \quad MVN(\underline{\mu}, \underline{\Sigma})$$

$$E(Y_i \mid \cdot) = E(\beta_0 + \beta_1 \, day_i + \alpha_{j(i)} + \varepsilon_i)$$

$$= E(\beta_0 + \beta_1 \, day_i) + \cancel{E(\alpha_i)}^{\,0} + \cancel{E(\varepsilon_i)}^{\,0}$$

$$= \beta_0 + \beta_1 \, day_i$$

$$\mu = E(\underline{Y} \mid \cdot) = \beta_0 + \beta_1 \, \underline{day}$$

$$\{\Sigma\}_{i,k} = Cov(y_i|\cdot, y_k|\cdot)$$

$$= Cov(\beta_0 + \beta_1 day_i + \alpha_{j(i)} + \varepsilon_i, \beta_0 + \beta_1 day_k + \alpha_{j(k)} + \varepsilon_k)$$

$$= Cov(\alpha_{j(i)} + \varepsilon_i, \alpha_{j(k)} + \varepsilon_k)$$

$$= Cov(\alpha_{j(i)}, \alpha_{j(k)}) + \cancelto{0}{Cov(\alpha_{j(i)}, \varepsilon_k)}$$

$$+ \cancelto{0}{Cov(\varepsilon_i, \alpha_{j(k)})} + Cov(\varepsilon_i, \varepsilon_k)$$

$$= \mathrm{Cov}\left(\alpha_{j(i)}, \alpha_{j(k)}\right) + \mathrm{Cov}\left(\cancel{\alpha}\,\varepsilon_i, \varepsilon_k\right)$$

$$= \begin{cases} \sigma_\alpha^2 + \sigma^2 & \text{if } i = k \quad \text{(same obs)} \\ \sigma_\alpha^2 & \text{else if } j(i) = j(k) \quad \text{(same subject diff obs)} \\ 0 & \text{otherwise} \end{cases}$$

$$
\Sigma = \begin{array}{c c}
 & \begin{array}{ccccc} k: & 1 & 2 & 3 & \cdots\ 10 \end{array} \\
\begin{array}{c} i \\ 1 \\ 2 \\ 3 \\ \vdots \\ 10 \\ \\ \\ \\ \end{array} &
\left[\begin{array}{c|c}
\begin{array}{cccc}
\sigma_\alpha^2+\sigma^2 & \sigma_\alpha^2 & & \sigma_\alpha^2 \\
\sigma_\alpha^2 & \sigma_\alpha^2+\sigma^2 & & \\
\sigma_\alpha^2 & \sigma_\alpha^2 & \ddots & \\
 & & & \sigma_\alpha^2+\sigma^2
\end{array} & 0 \\
\hline
0 & \begin{array}{cccc}
\sigma_\alpha^2+\sigma^2 & & & \sigma_\alpha^2 \\
 & \ddots & & \\
\sigma_\alpha^2 & & \ddots & \\
 & & & \ddots
\end{array}
\end{array}\right]
\end{array}
$$

Random intercept and slope model

Model

Let i represent each observation and $j(i)$ be the subject in oberservation i then

$$Y_i = \alpha_{j(i)} + \beta_{j(i)} \times \text{Days} + \epsilon_i$$

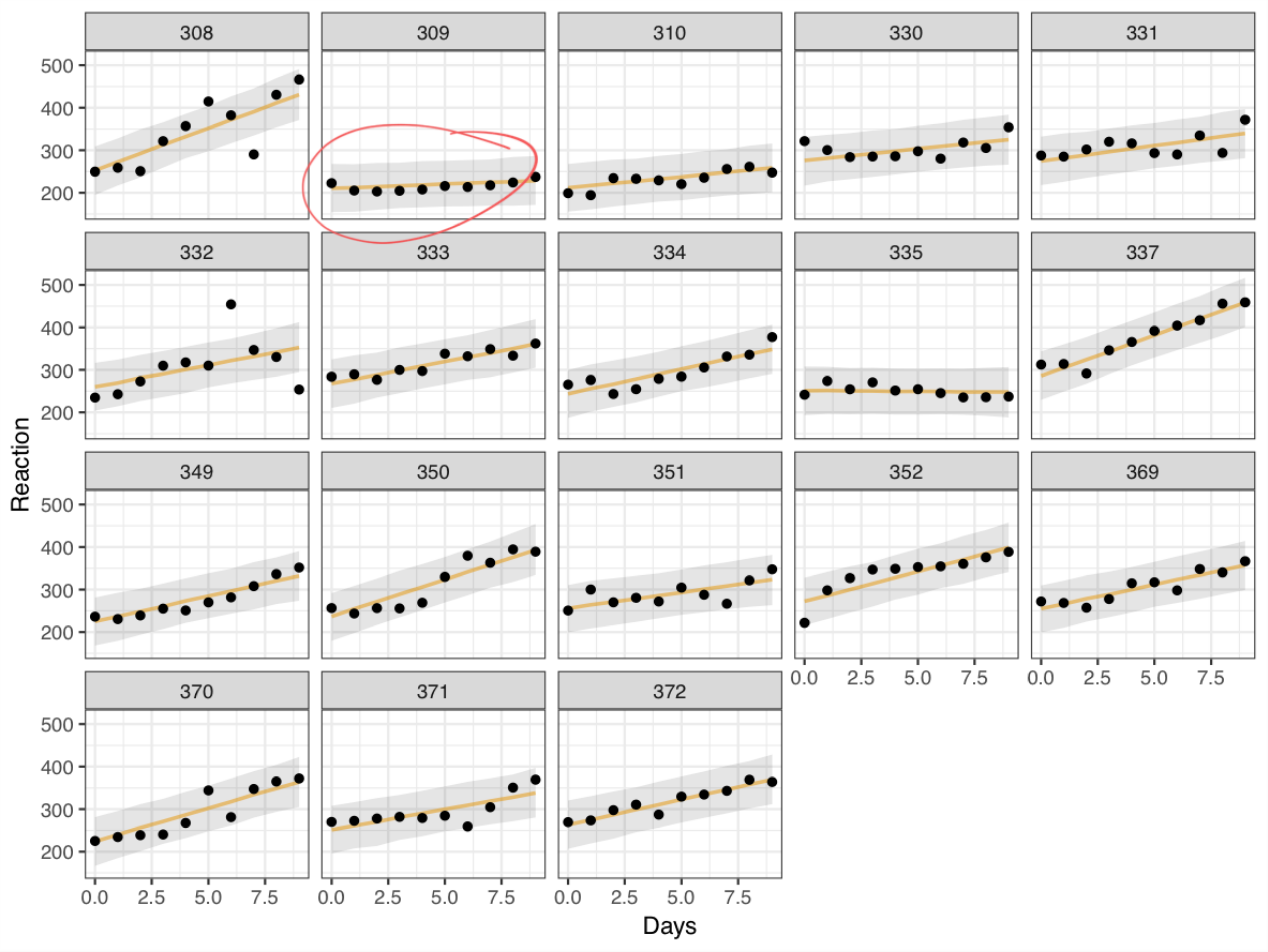
$$\begin{aligned}
\alpha_j &\sim \mathcal{N}(\beta_0,\ \sigma^2_\alpha) \\
\beta_j &\sim \mathcal{N}(\beta_1,\ \sigma^2_\beta) \\
\epsilon_i &\sim \mathcal{N}(0,\ \sigma^2)
\end{aligned}$$

$$\begin{aligned}
\beta_0, \beta_1 &\sim \mathcal{N}(0, 10000) \\
\sigma, \sigma_\alpha, \sigma_\beta &\sim \text{Unif}(0, 100)
\end{aligned}$$

Model - JAGS

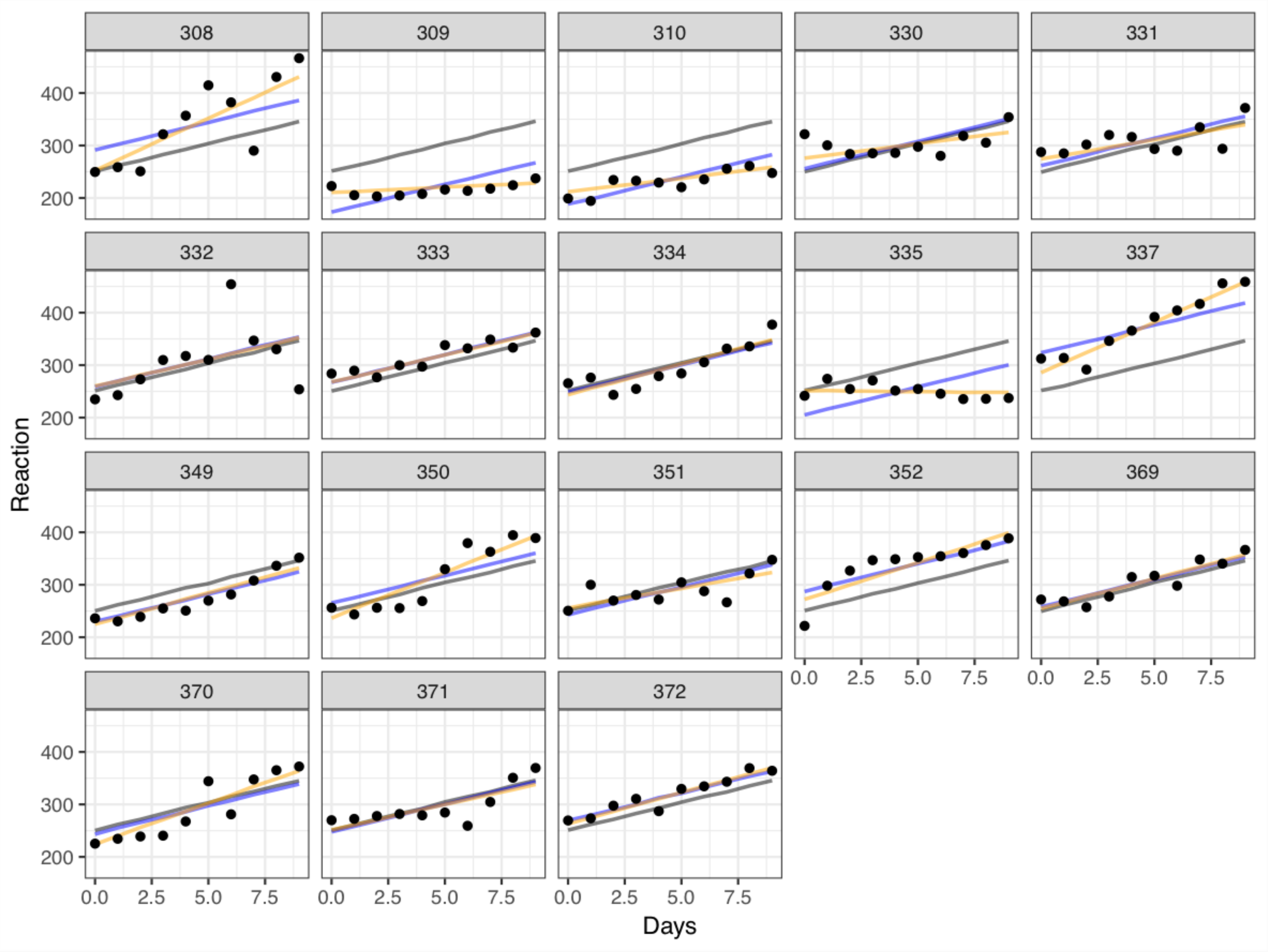
```
## model{
##   for(i in 1:length(Reaction)) {
##     Reaction[i] ~ dnorm(mu[i],tau2)
##     mu[i] <- alpha[Subject_index[i]] + beta[Subject_index[i]]*Days[i]
##     Y_hat[i] ~ dnorm(mu[i],tau2)
##   }
##
##   sigma ~ dunif(0, 100)
##   tau2 <- 1/(sigma*sigma)
##
##   for(j in 1:18) {
##     alpha[j] ~ dnorm(beta_0, tau2_alpha)
##     beta[j]  ~ dnorm(beta_1, tau2_beta)
##   }
##
##   beta_0 ~ dnorm(0,1/10000)
##   beta_1 ~ dnorm(0,1/10000)
##
##   sigma_alpha ~ dunif(0, 100)
##   tau2_alpha <- 1/(sigma_alpha*sigma_alpha)
##
##   sigma_beta ~ dunif(0, 100)
##   tau2_beta <- 1/(sigma_beta*sigma_beta)
## }
```

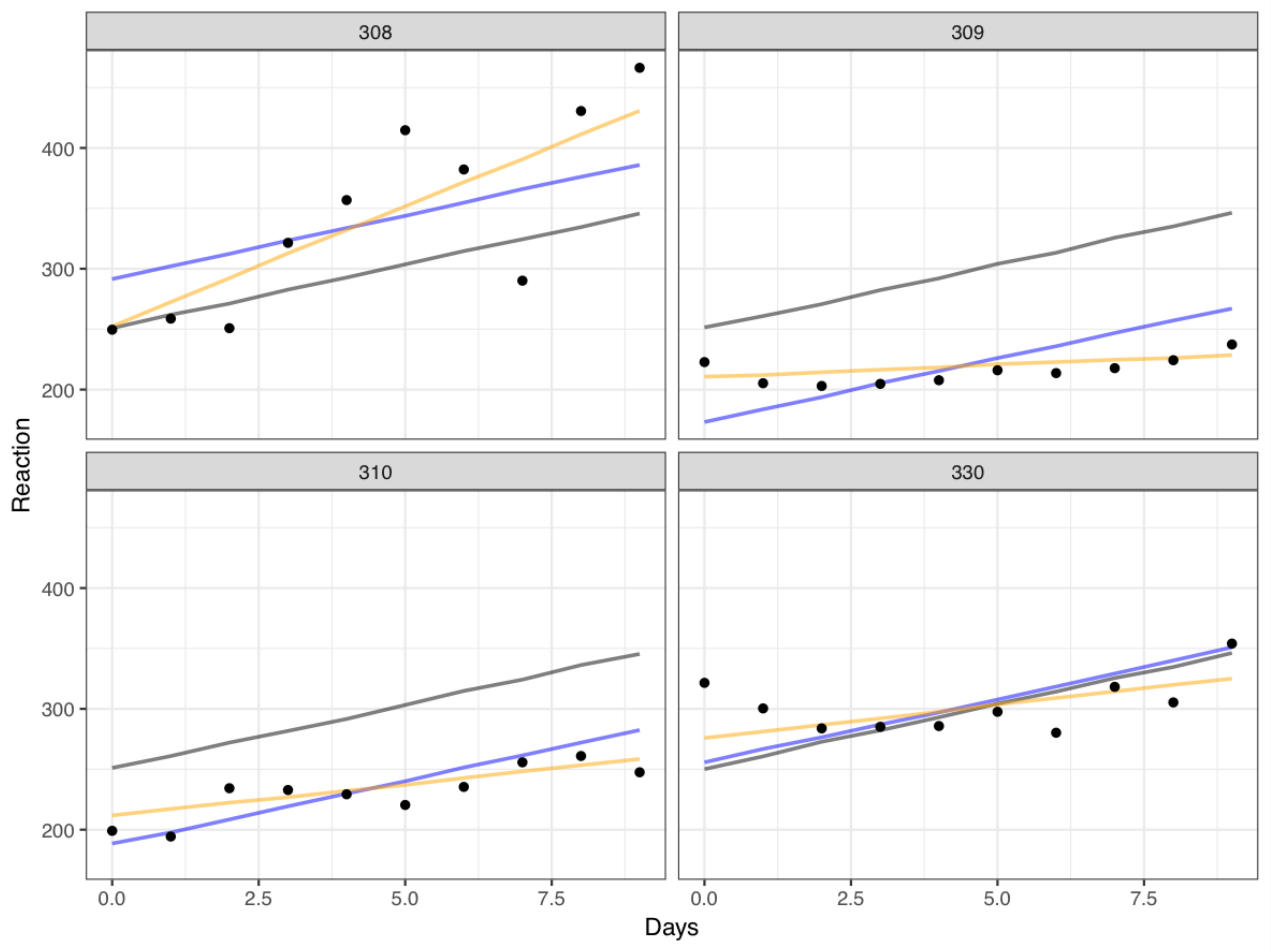
Model fit



308
309
310
330
331
332
333
334
335
337
349
350
351
352
369
370
371
372
Reaction
500
400
300
200
0.0
2.5
5.0
7.5
Days

Comparison


308
309
310
330
331
332
333
334
335
337
349
350
351
352
369
370
371
372
Reaction
400
300
200
0.0
2.5
5.0
7.5
Days


308
309
310
330
Reaction
Days
400
300
200
0.0
2.5
5.0
7.5

Residuals by subject