model.rmd

Yanran Li, Yijin Wang, Shubo Zhang

2024-03-20

```
library(mice)
```

```
##
## Attaching package: 'mice'

## The following object is masked from 'package:stats':
##
##     filter

## The following objects are masked from 'package:base':
##
##     cbind, rbind
```

```
library(tidyr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(visdat)
library(naniar)
library(ggplot2)
library(lme4)
```

```
## Loading required package: Matrix

## Warning: package 'Matrix' was built under R version 4.2.3

##
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```r
library(broom.mixed)
library(kableExtra)
```

```
## Warning: package 'kableExtra' was built under R version 4.2.3
```

```
##
## Attaching package: 'kableExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```r
data <- read.csv("data.csv") %>%
  janitor::clean_names() %>%
  mutate(day = as.factor(day))
```

```r
# pre-processing: expand design matrix and fill in NAs
full_data <- data %>%
  complete(day,
           nesting(subject_id, treatment_group)) %>%
  relocate(day, .after = subject_id) %>%
  relocate(treatment_group, .after = gender) %>%
  group_by(subject_id) %>%
  fill(age, gender, .direction = "downup") %>%
  arrange(subject_id, day)
```
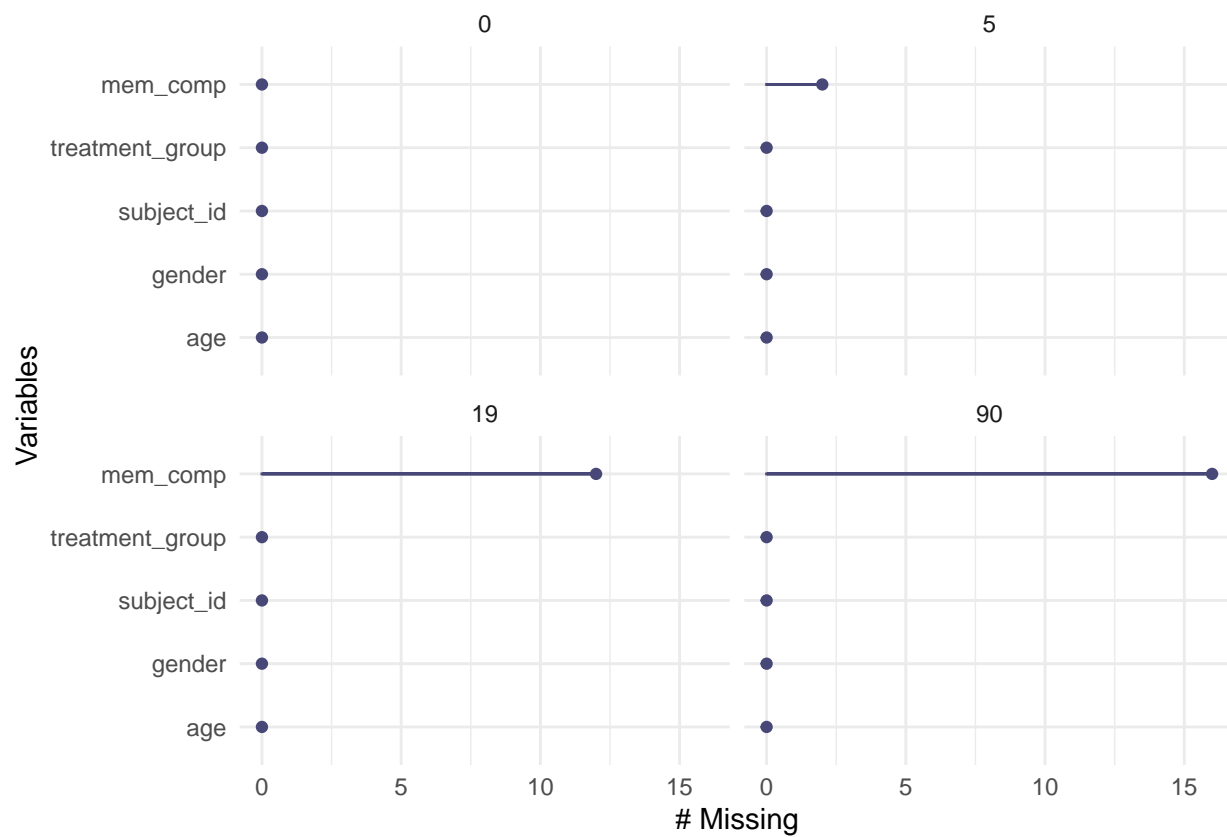
```r
vis_miss(full_data)
```
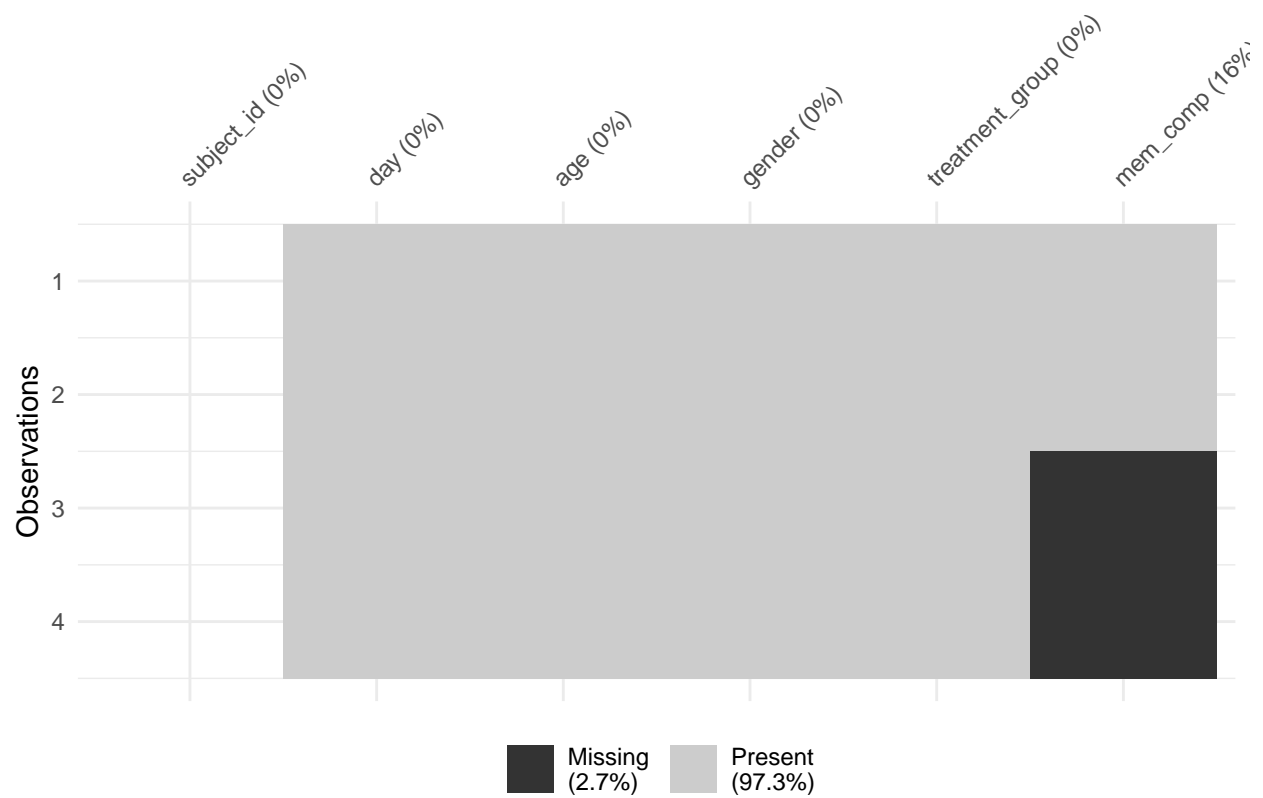
```
vis_miss(full_data,facet = day)
```
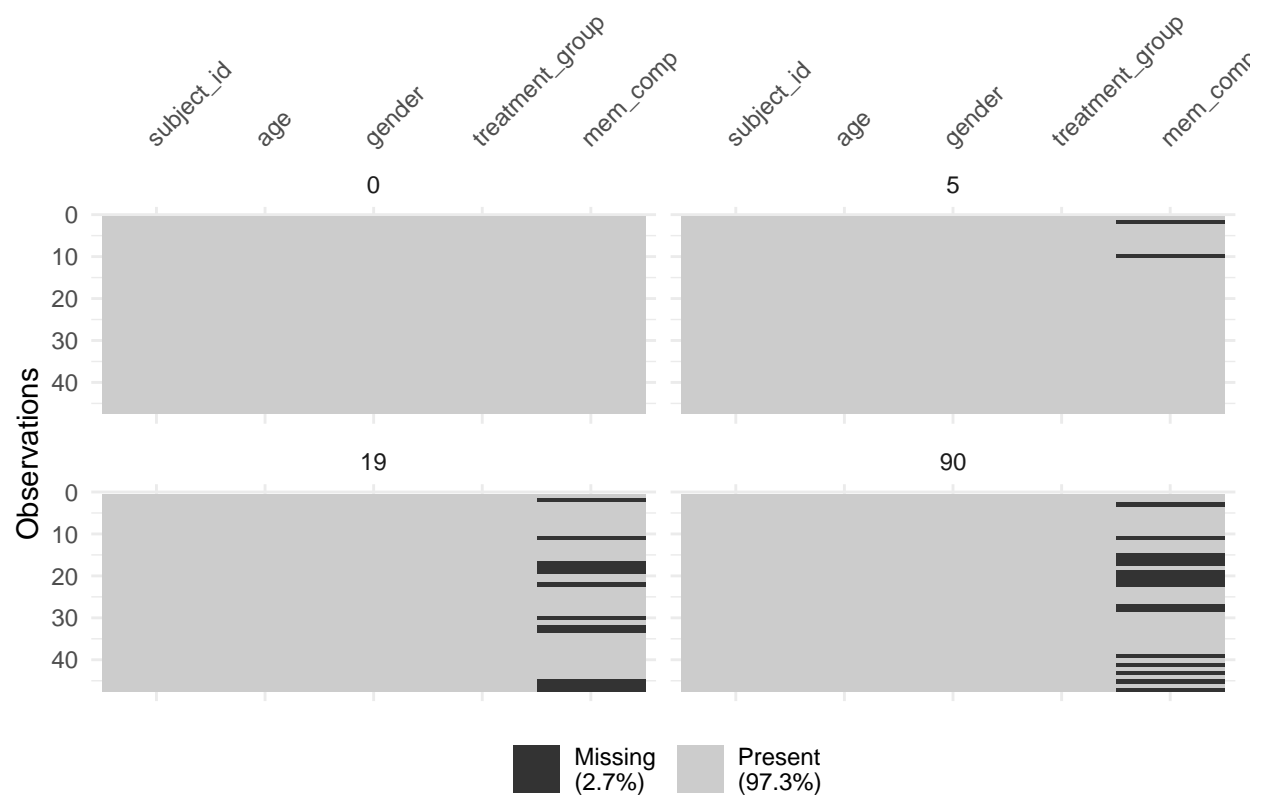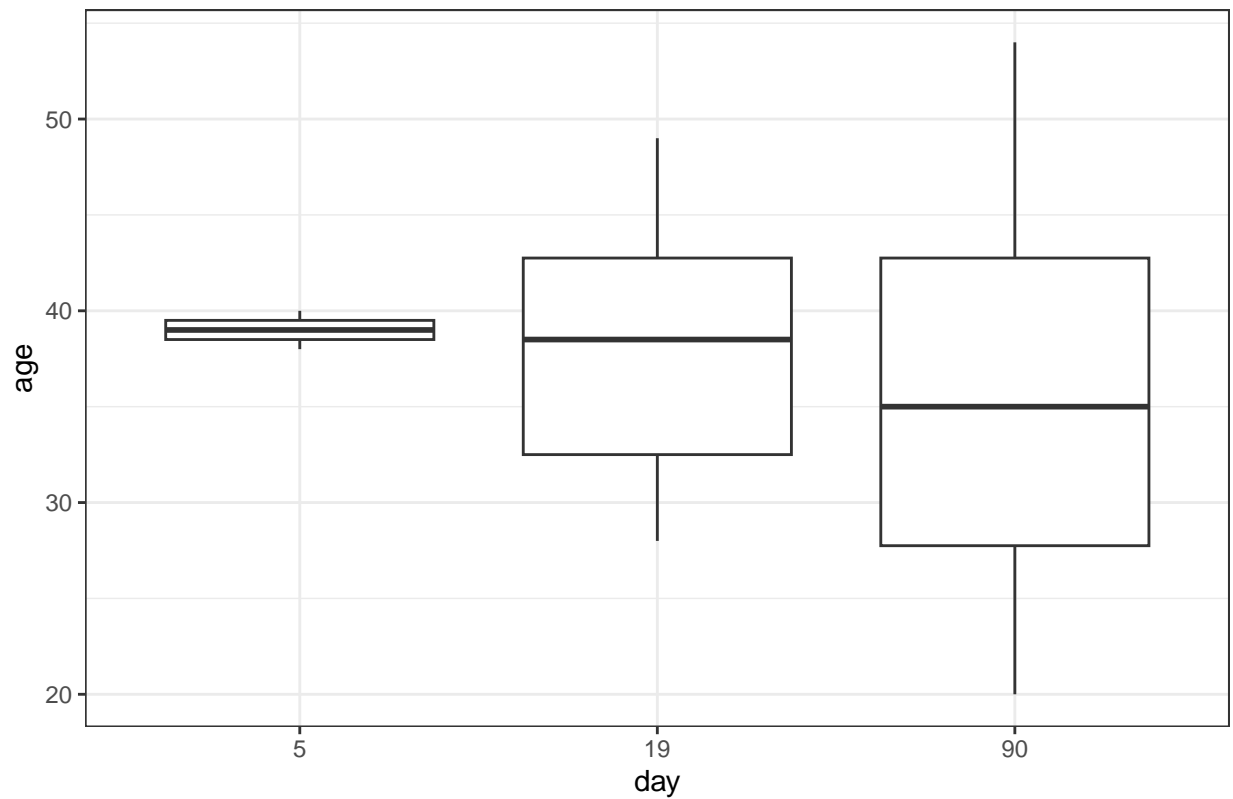
```
gg_miss_var(full_data,
            facet = day)
```

```
full_data %>% filter(is.na(mem_comp)) %>%
  ggplot(aes(x=as.factor(day)))+
  geom_boxplot(aes(y=age))+
  theme_bw()+
  xlab("day")+
  ggtitle("Relationship of age and missing values")
```
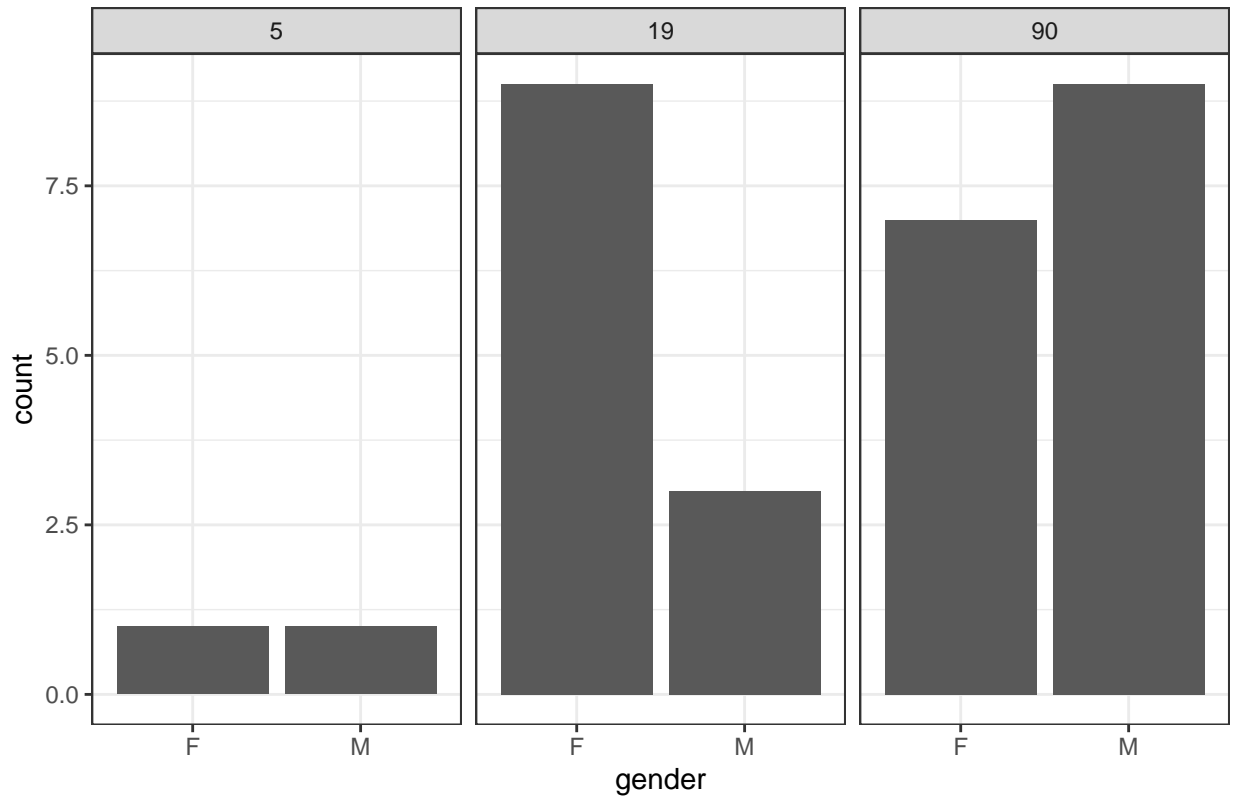
Relationship of age and missing values


```
full_data %>% filter(is.na(mem_comp)) %>%
  ggplot()+
  geom_bar(aes(x=gender, stat = "count"))+
  facet_wrap(day~.)+
  theme_bw()+
  xlab("gender")+
  ggtitle("Relationship of gender and missing values")
```

```
## Warning in geom_bar(aes(x = gender, stat = "count")): Ignoring unknown
## aesthetics: stat
```

Relationship of gender and missing values


```r
# assume MAR
# multiple imputations
imputed_data <- mice::mice(data = full_data, m = 6, seed =2024, print = FALSE)
```

```
## Warning: Number of logged events: 2
```

```r
imputed_data
```

```
## Class: mids
## Number of multiple imputations:  6
## Imputation methods:
##      subject_id           day           age         gender treatment_group
##             ""            ""            ""            ""             ""
##       mem_comp
##         "pmm"
## PredictorMatrix:
##                 subject_id day age gender treatment_group mem_comp
## subject_id               0   1   1      0               0        1
## day                      1   0   1      0               0        1
## age                      1   1   0      0               0        1
## gender                   1   1   1      0               0        1
## treatment_group          1   1   1      0               0        1
## mem_comp                 1   1   1      0               0        0
## Number of logged events:  2
##   it im dep     meth          out
```
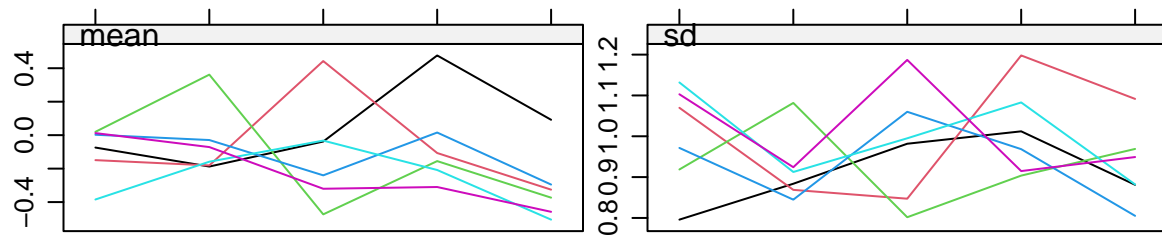
```
## 1  0  0      constant              gender
## 2  0  0      constant  treatment_group
```

```
# used pmm method
```
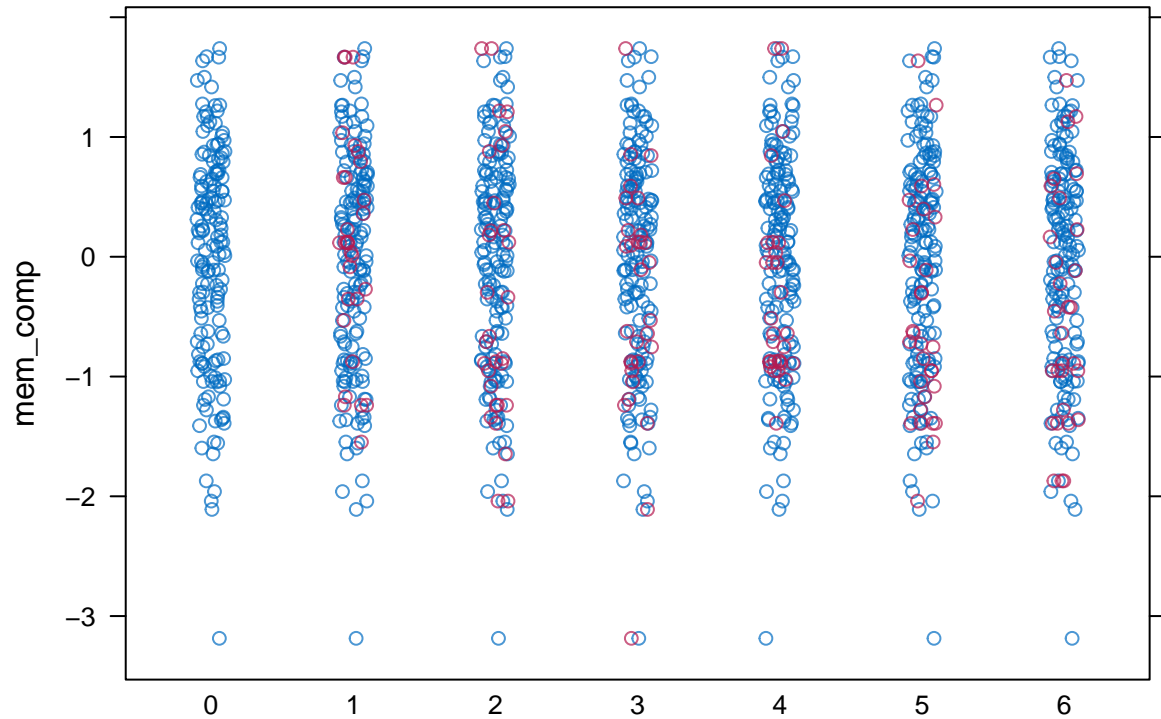
```
# check convergence
plot(imputed_data)
```


Iteration

```
# check if imputed data makes sense (ie. in a reasonable bound)
# red is imputed, blue is observed
mice::stripplot(imputed_data, mem_comp)
```

```r
# fit model to each imputed datasets

model <- with(imputed_data, lmer(mem_comp ~ age + gender + treatment_group+day+treatment_group:day + (1

pool.fit <- pool(model)

summary(pool.fit)
```

```
##                    term     estimate   std.error   statistic         df
## 1           (Intercept)  0.61259148  0.58498731   1.0471876  142.40918
## 2                   age -0.01435861  0.01378527  -1.0415912  125.55072
## 3               genderM  0.07118804  0.25591736   0.2781681   88.04616
## 4        treatment_groupB -0.07762312  0.34501125  -0.2249872  169.48896
## 5        treatment_groupC -0.18796858  0.33052182  -0.5687025  169.82702
## 6                  day5 -0.09577277  0.20913497  -0.4579472  170.01722
## 7                 day19 -0.29378401  0.21785002  -1.3485609  131.38638
## 8                 day90 -0.32331261  0.32115763  -1.0067100   12.46684
## 9   treatment_groupB:day5 -0.06276973  0.32782242  -0.1914748  117.93381
## 10  treatment_groupC:day5  0.34998520  0.30725190   1.1390823  152.67981
## 11 treatment_groupB:day19  0.09954652  0.36771610   0.2707157   41.12777
## 12 treatment_groupC:day19  0.57107906  0.31251365   1.8273732  132.74876
## 13 treatment_groupB:day90  0.12771299  0.40146403   0.3181181   24.01220
## 14 treatment_groupC:day90  0.31860668  0.34760696   0.9165716   48.25563
##       p.value
## 1  0.29678712
```

9

```
## 2   0.29960283
## 3   0.78153553
## 4   0.82226025
## 5   0.57030935
## 6   0.64757513
## 7   0.17979804
## 8   0.33321463
## 9   0.84848291
## 10 0.25645285
## 11 0.78796372
## 12 0.06988959
## 13 0.75314553
## 14 0.36392529
```

```r
# check the pooled model covariance
# get the full matrix ubar (instead of only the diagonal)
m <- pool.fit$m
ubar <- Reduce("+", lapply(model$analyses, vcov)) / (m)
b <- pool.fit$pooled$b # this one is still provided by mice

# # or by hand as well
# qbar <- getqbar(fil.pooled)  # pooled estimates
# b <- 1 / (m-1) * rowSums((sapply(fit.mi$analyses, coef) - qbar)^2)

t <- ubar + (1 + 1 / (m)) * b  # this is t as it used to be

# check versus the diagonal of t that is still provided
all.equal(as.numeric(diag(t)), pool.fit$pooled$t) # check
```

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

```r
calc_pooled_covar<- function(pool.fit,model){
  # check the pooled model covariance
  # get the full matrix ubar (instead of only the diagonal)
  m <- pool.fit$m
  ubar <- Reduce("+", lapply(model$analyses, vcov)) / (m)
  b <- pool.fit$pooled$b # this one is still provided by mice

  # # or by hand as well
  # qbar <- getqbar(fil.pooled)  # pooled estimates
  # b <- 1 / (m-1) * rowSums((sapply(fit.mi$analyses, coef) - qbar)^2)

  t <- ubar + (1 + 1 / (m)) * b  # this is t as it used to be

  # check versus the diagonal of t that is still provided
  check = all.equal(as.numeric(diag(t)), pool.fit$pooled$t) # check
  if (check) {
    return(t)
  } else{
    stop("error!!!!!!!")
  }
}
```

```r
smy<-summary(pool.fit)
# Hypothesis test to see whether treatment effects at day 5, 19, and 90 are significant

output <- data.frame(
  Estimate = round(c(smy$estimate[c(6:8, 9, 11, 13, 10, 12, 14)], smy$estimate[6]+smy$estimate[9], smy$e
  `Standard error` = round(c(smy$std.error[c(6:8, 9, 11, 13, 10, 12, 14)], sqrt((t[6,6]) + (t[9,9]) + 2*
  `95% CI` = rep(NA,15)
)
colnames(output)[c(2,3)] <- c("Standard error", "95% CI")
output$`95% CI` <- paste0( "(", as.character(round(output$Estimate - 1.96*output$`Standard error`, 3)),
output$Significance <- ifelse((output$Estimate - 1.96*output$`Standard error`) * (output$Estimate + 1.96
rownames(output) <- c("$\\beta_5$", "$\\beta_6$", "$\\beta_7$", "$\\beta_8$", "$\\beta_9$", "$\\beta_{10

# Print the table using kable
knitr::kable(output,caption = "Summary table")
```

Table 1: Summary table

	Estimate	Standard error	95% CI	Significance
β_5	-0.096	0.209	(-0.506,0.314)	FALSE
β_6	-0.294	0.218	(-0.721,0.133)	FALSE
β_7	-0.323	0.321	(-0.952,0.306)	FALSE
β_8	-0.063	0.328	(-0.706,0.58)	FALSE
β_9	0.100	0.368	(-0.621,0.821)	FALSE
β_{10}	0.128	0.401	(-0.658,0.914)	FALSE
β_{11}	0.350	0.307	(-0.252,0.952)	FALSE
β_{12}	0.571	0.313	(-0.042,1.184)	FALSE
β_{13}	0.319	0.348	(-0.363,1.001)	FALSE
$\beta_5 + \beta_8$	-0.159	0.252	(-0.653,0.335)	FALSE
$\beta_6 + \beta_9$	-0.194	0.320	(-0.821,0.433)	FALSE
$\beta_7 + \beta_{10}$	-0.196	0.544	(-1.262,0.87)	FALSE
$\beta_5 + \beta_{11}$	0.254	0.225	(-0.187,0.695)	FALSE
$\beta_6 + \beta_{12}$	0.277	0.255	(-0.223,0.777)	FALSE
$\beta_7 + \beta_{13}$	-0.005	0.505	(-0.995,0.985)	FALSE

```r
# pool coef estimates and standard errors to create 95% CIs
pool_coef_ci_table <- function(pool.fit, t, caption){
  smy<-summary(pool.fit)
# Hypothesis test to see whether treatment effects at day 5, 19, and 90 are significant

output <- data.frame(
  Estimate = round(c(smy$estimate[c(6:8, 9, 11, 13, 10, 12, 14)], smy$estimate[6]+smy$estimate[9], smy$e
  `Standard error` = round(c(smy$std.error[c(6:8, 9, 11, 13, 10, 12, 14)], sqrt((t[6,6]) + (t[9,9]) + 2*
  `95% CI` = rep(NA,15)
)
colnames(output)[c(2,3)] <- c("Standard error", "95% CI")
output$`95% CI` <- paste0( "(", as.character(round(output$Estimate - 1.96*output$`Standard error`, 3)),
output$Significance <- ifelse((output$Estimate - 1.96*output$`Standard error`) * (output$Estimate + 1.96
rownames(output) <- c("$\\beta_5$", "$\\beta_6$", "$\\beta_7$", "$\\beta_8$", "$\\beta_9$", "$\\beta_{10

# Print the table using kable
```

```
return (knitr::kable(output,caption = caption,escape = TRUE))
}


# sensitivity analysis for each data missing assumptions
# mcar - fit data with completers data
completers_data <- full_data %>%
  group_by(subject_id) %>%
  mutate(missing = any(is.na(mem_comp))) %>%
  ungroup()
completers_fit <- lmer(mem_comp ~ age + gender + treatment_group+day +treatment_group:day + (1|subject_
summary(completers_fit)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## mem_comp ~ age + gender + treatment_group + day + treatment_group:day +
##     (1 | subject_id)
##    Data: completers_data %>% filter(missing == FALSE)
##
## REML criterion at convergence: 173.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.7688 -0.5756  0.1020  0.6102  1.7294
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  subject_id (Intercept) 0.6790   0.8240
##  Residual               0.1597   0.3997
## Number of obs: 96, groups:  subject_id, 24
##
## Fixed effects:
##                          Estimate Std. Error t value
## (Intercept)            -0.2777878  0.9983716  -0.278
## age                     0.0004539  0.0227681   0.020
## genderM                 0.2159847  0.3696430   0.584
## treatment_groupB        0.0579773  0.4876764   0.119
## treatment_groupC       -0.0553875  0.4461694  -0.124
## day5                   -0.2121626  0.1998282  -1.062
## day19                  -0.3476050  0.1998282  -1.740
## day90                  -0.2289847  0.1998282  -1.146
## treatment_groupB:day5   0.0978170  0.2925186   0.334
## treatment_groupC:day5   0.3899459  0.2746377   1.420
## treatment_groupB:day19  0.3457121  0.2925186   1.182
## treatment_groupC:day19  0.6461374  0.2746377   2.353
## treatment_groupB:day90  0.3018569  0.2925186   1.032
## treatment_groupC:day90  0.2610949  0.2746377   0.951
##
##
## Correlation matrix not shown by default, as p = 14 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it
```

```r
t <-vcov(completers_fit)
```

```r
completers_fit_se <- summary(completers_fit)[["coefficients"]][,2]
# calculate CIs
output <- data.frame(
  Estimate = round(c(completers_fit@beta[c(6:8, 9, 11, 13, 10, 12, 14)], completers_fit@beta[6]+complete
  `Standard error` = round(c(completers_fit_se[c(6:8, 9, 11, 13, 10, 12, 14)], sqrt((t[6,6]) + (t[9,9])
  `95% CI` = rep(NA,15)
)
colnames(output)[c(2,3)] <- c("Standard error", "95% CI")
output$`95% CI` <- paste0( "(", as.character(round(output$Estimate - 1.96*output$`Standard error`, 3)),
output$Significance <- ifelse((output$Estimate - 1.96*output$`Standard error`) * (output$Estimate + 1.9
rownames(output) <- c("$\\beta_5$", "$\\beta_6$", "$\\beta_7$", "$\\beta_8$", "$\\beta_9$", "$\\beta_{1(

# Print the table using kable
knitr::kable(output,caption = "Completers' Model Summary table")
```

Table 2: Completers' Model Summary table

	Estimate	Standard error	95% CI	Significance
β_5	-0.212	0.200	(-0.604,0.18)	FALSE
β_6	-0.348	0.200	(-0.74,0.044)	FALSE
β_7	-0.229	0.200	(-0.621,0.163)	FALSE
β_8	0.098	0.293	(-0.476,0.672)	FALSE
β_9	0.346	0.293	(-0.228,0.92)	FALSE
β_{10}	0.302	0.293	(-0.272,0.876)	FALSE
β_{11}	0.390	0.275	(-0.149,0.929)	FALSE
β_{12}	0.646	0.275	(0.107,1.185)	TRUE
β_{13}	0.261	0.275	(-0.278,0.8)	FALSE
$\beta_5 + \beta_8$	-0.114	0.214	(-0.533,0.305)	FALSE
$\beta_6 + \beta_9$	-0.002	0.214	(-0.421,0.417)	FALSE
$\beta_7 + \beta_{10}$	0.073	0.214	(-0.346,0.492)	FALSE
$\beta_5 + \beta_{11}$	0.178	0.188	(-0.19,0.546)	FALSE
$\beta_6 + \beta_{12}$	0.299	0.188	(-0.069,0.667)	FALSE
$\beta_7 + \beta_{13}$	0.032	0.188	(-0.336,0.4)	FALSE

```r
# mnar - fill na's with different delta --> fit model --> CI's

# Create a delta vector that represent the following adjustment values for mmHg: 0 for MAR, and -0.7, -(
delta <- c(0, -0.7, -0.2, 0.2, 0.7 )

imp.all <- vector("list", length(delta))
post <- imputed_data$post
for (i in 1:length(delta)){
  d <- delta[i]
  cmd <- paste("imp[[j]][,i] <- imp[[j]][,i] +", d)
  post["mem_comp"] <- cmd
  imp <- mice::mice(full_data, post = post,m = 6,seed = i, print = FALSE)
  imp.all[[i]] <- imp
}
```
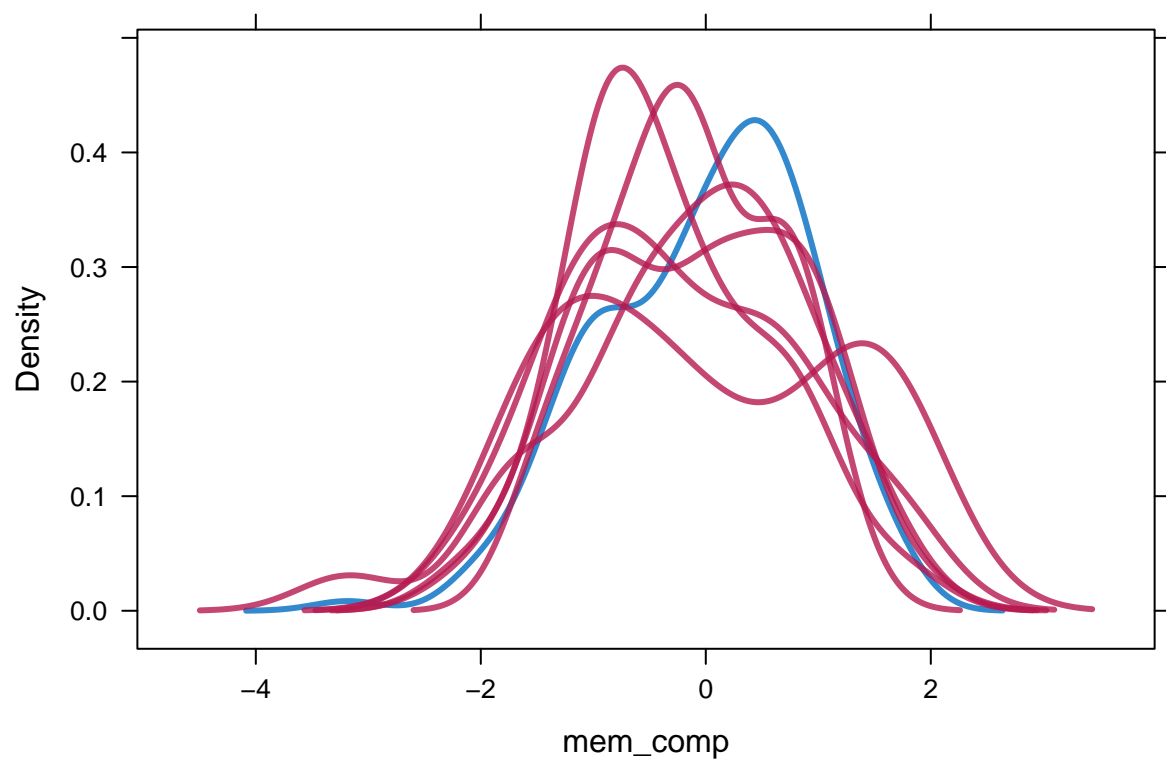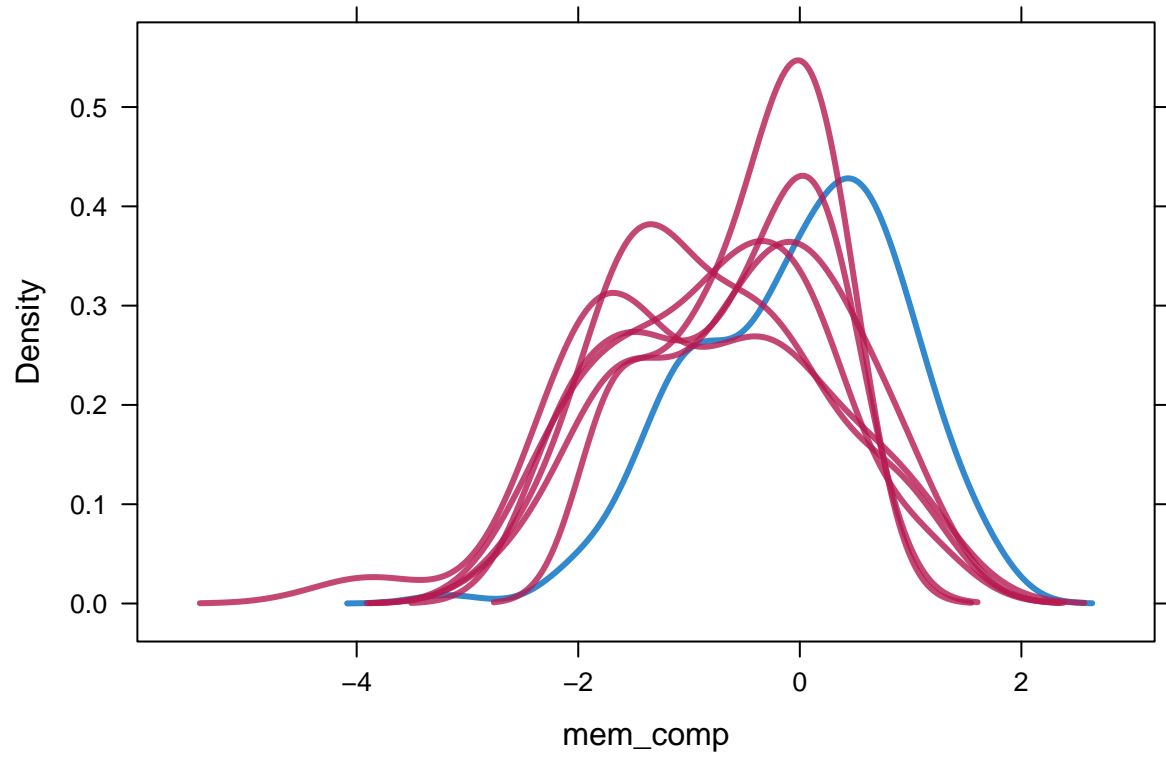
```
## Warning: Number of logged events: 2
```

```
## Warning: Number of logged events: 2

## Warning: Number of logged events: 2

## Warning: Number of logged events: 2

## Warning: Number of logged events: 2
```
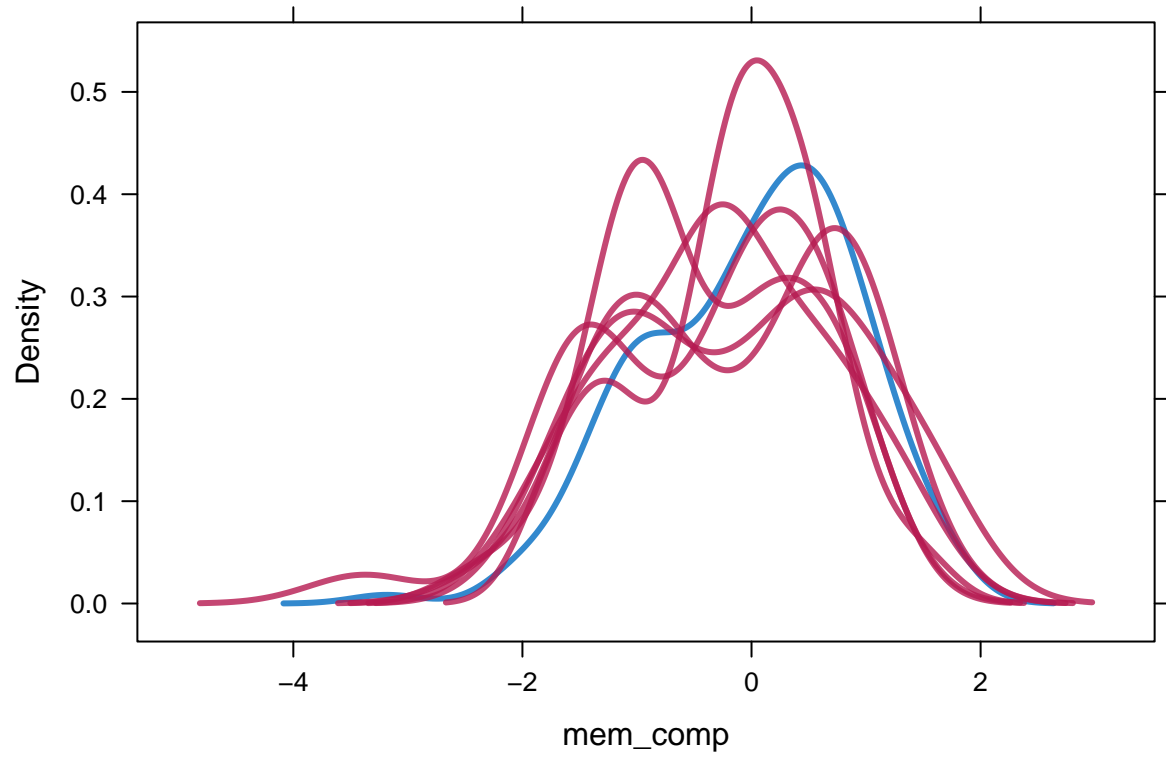
```
# imputation with no adjustment (delta = 0)
densityplot(imp.all[[1]], lwd = 3)
```



```
# imputation with adjustment (delta = -0.7)
densityplot(imp.all[[2]], lwd = 3)
```
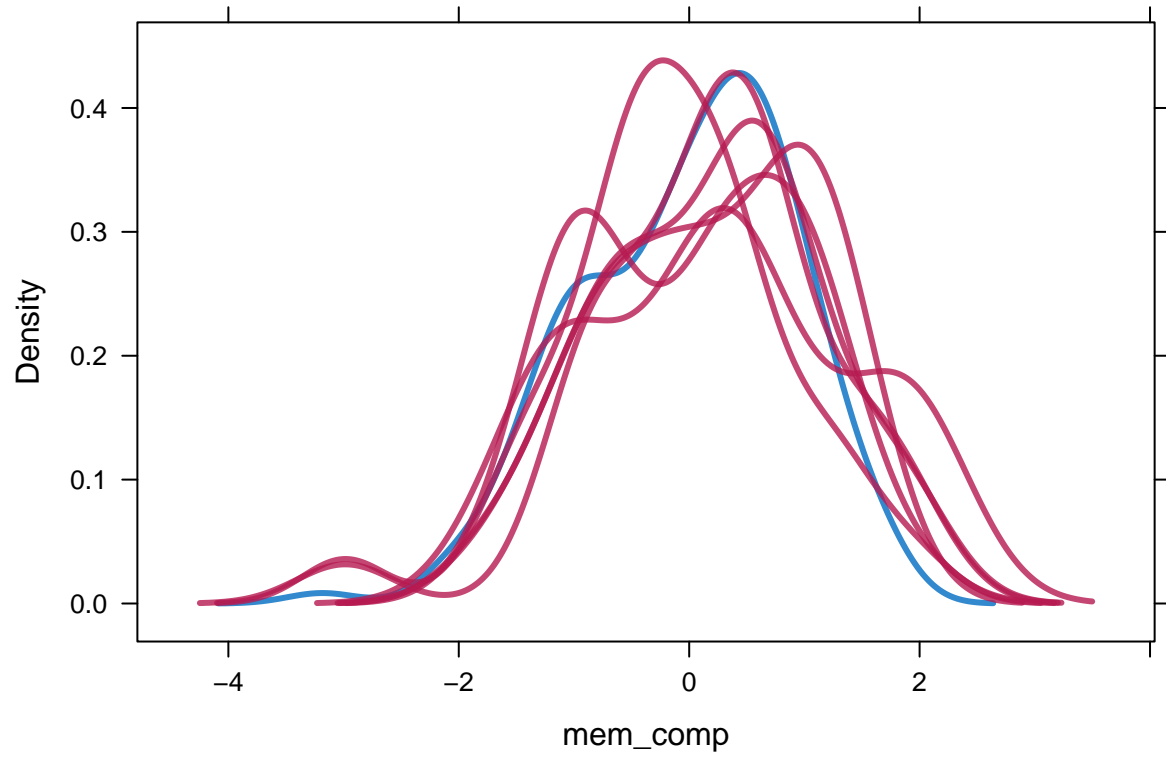
```
# imputation with adjustment (delta = -0.2)
densityplot(imp.all[[3]], lwd = 3)
```
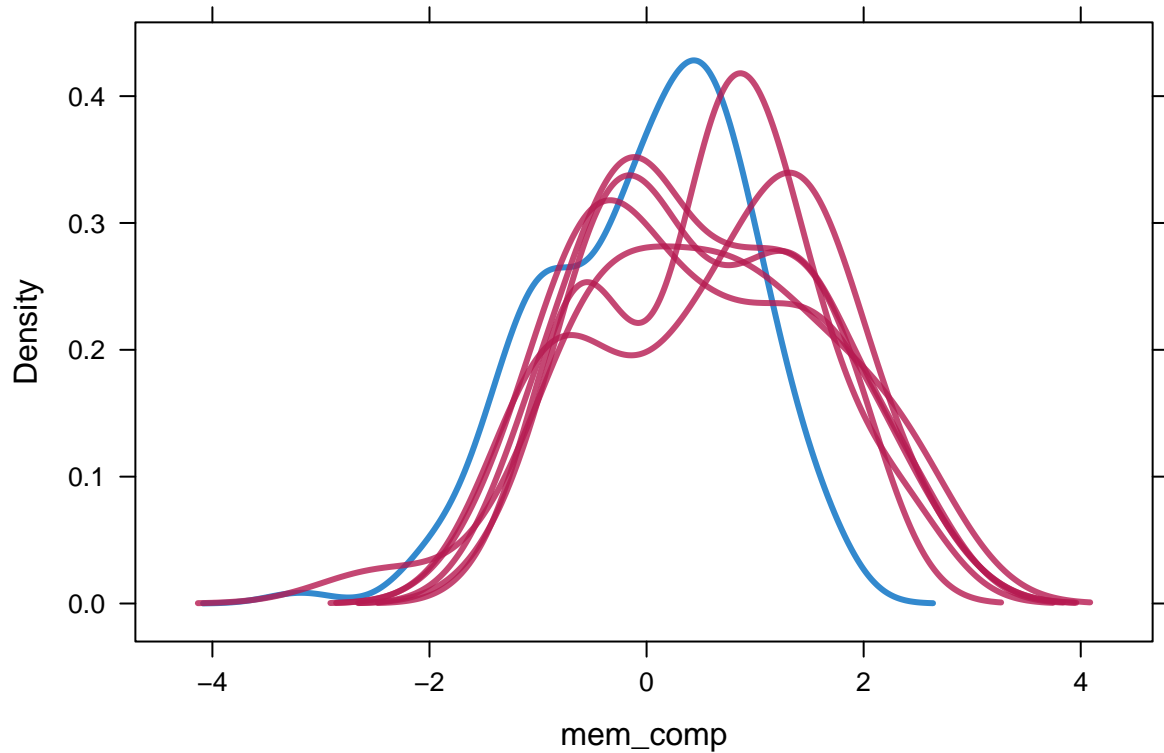
```
# imputation with adjustment (delta = 0.2)
densityplot(imp.all[[4]], lwd = 3)
```

```
# imputation with adjustment (delta = 0.7)
densityplot(imp.all[[5]], lwd = 3)
```

```r
# calculate results for imputation+adjustment datasets
for (i in 2:length(delta)){
  curr_data <- imp.all[[i]]

  # fit model
  curr_model <- with(curr_data, lmer(mem_comp ~ age + gender + treatment_group+day+treatment_group:day +

  curr.pool.fit <- pool(curr_model)

  # pool covar matrix
  curr_t<-calc_pooled_covar(curr.pool.fit,curr_model)

  # generate ci table
  curr_table <- pool_coef_ci_table(curr.pool.fit, curr_t, caption = paste0("delta=", delta[i]," Summary



  print(curr_table)
}
```

Table 3: delta=-0.7 Summary Table

	Estimate	Standard error	95% CI	Significance
β_5	-0.096	0.231	(-0.549,0.357)	FALSE
β_6	-0.361	0.259	(-0.869,0.147)	FALSE
β_7	-0.446	0.254	(-0.944,0.052)	FALSE
β_8	-0.040	0.352	(-0.73,0.65)	FALSE
β_9	0.039	0.409	(-0.763,0.841)	FALSE
β_{10}	0.107	0.365	(-0.608,0.822)	FALSE
β_{11}	0.303	0.336	(-0.356,0.962)	FALSE
β_{12}	0.622	0.379	(-0.121,1.365)	FALSE
β_{13}	0.200	0.377	(-0.539,0.939)	FALSE
$\beta_5 + \beta_8$	-0.135	0.265	(-0.654,0.384)	FALSE
$\beta_6 + \beta_9$	-0.322	0.394	(-1.094,0.45)	FALSE
$\beta_7 + \beta_{10}$	-0.339	0.337	(-1,0.322)	FALSE
$\beta_5 + \beta_{11}$	0.207	0.245	(-0.273,0.687)	FALSE
$\beta_6 + \beta_{12}$	0.261	0.364	(-0.452,0.974)	FALSE
$\beta_7 + \beta_{13}$	-0.245	0.350	(-0.931,0.441)	FALSE

Table 4: delta=-0.2 Summary Table

	Estimate	Standard error	95% CI	Significance
β_5	-0.096	0.214	(-0.515,0.323)	FALSE
β_6	-0.275	0.236	(-0.738,0.188)	FALSE
β_7	-0.326	0.255	(-0.826,0.174)	FALSE
β_8	-0.022	0.326	(-0.661,0.617)	FALSE
β_9	0.206	0.355	(-0.49,0.902)	FALSE
β_{10}	0.143	0.374	(-0.59,0.876)	FALSE
β_{11}	0.295	0.313	(-0.318,0.908)	FALSE
β_{12}	0.584	0.332	(-0.067,1.235)	FALSE
β_{13}	0.289	0.388	(-0.471,1.049)	FALSE
$\beta_5 + \beta_8$	-0.118	0.246	(-0.6,0.364)	FALSE
$\beta_6 + \beta_9$	-0.069	0.333	(-0.722,0.584)	FALSE
$\beta_7 + \beta_{10}$	-0.183	0.390	(-0.947,0.581)	FALSE
$\beta_5 + \beta_{11}$	0.199	0.228	(-0.248,0.646)	FALSE
$\beta_6 + \beta_{12}$	0.309	0.307	(-0.293,0.911)	FALSE
$\beta_7 + \beta_{13}$	-0.037	0.404	(-0.829,0.755)	FALSE

Table 5: delta=0.2 Summary Table

	Estimate	Standard error	95% CI	Significance
β_5	-0.096	0.201	(-0.49,0.298)	FALSE
β_6	-0.260	0.231	(-0.713,0.193)	FALSE
β_7	-0.116	0.234	(-0.575,0.343)	FALSE
β_8	-0.018	0.326	(-0.657,0.621)	FALSE
β_9	0.199	0.366	(-0.518,0.916)	FALSE
β_{10}	0.080	0.352	(-0.61,0.77)	FALSE
β_{11}	0.384	0.293	(-0.19,0.958)	FALSE
β_{12}	0.598	0.349	(-0.086,1.282)	FALSE
β_{13}	0.271	0.330	(-0.376,0.918)	FALSE

	Estimate	Standard error	95% CI	Significance
$\beta_5 + \beta_8$	-0.114	0.256	(-0.616,0.388)	FALSE
$\beta_6 + \beta_9$	-0.061	0.363	(-0.772,0.65)	FALSE
$\beta_7 + \beta_{10}$	-0.036	0.354	(-0.73,0.658)	FALSE
$\beta_5 + \beta_{11}$	0.288	0.213	(-0.129,0.705)	FALSE
$\beta_6 + \beta_{12}$	0.339	0.346	(-0.339,1.017)	FALSE
$\beta_7 + \beta_{13}$	0.156	0.332	(-0.495,0.807)	FALSE

Table 6: delta=0.7 Summary Table

	Estimate	Standard error	95% CI	Significance
β_5	-0.096	0.220	(-0.527,0.335)	FALSE
β_6	-0.072	0.226	(-0.515,0.371)	FALSE
β_7	-0.079	0.263	(-0.594,0.436)	FALSE
β_8	0.051	0.336	(-0.608,0.71)	FALSE
β_9	0.216	0.405	(-0.578,1.01)	FALSE
β_{10}	0.205	0.419	(-0.616,1.026)	FALSE
β_{11}	0.365	0.325	(-0.272,1.002)	FALSE
β_{12}	0.630	0.362	(-0.08,1.34)	FALSE
β_{13}	0.178	0.424	(-0.653,1.009)	FALSE
$\beta_5 + \beta_8$	-0.045	0.254	(-0.543,0.453)	FALSE
$\beta_6 + \beta_9$	0.144	0.354	(-0.55,0.838)	FALSE
$\beta_7 + \beta_{10}$	0.126	0.437	(-0.731,0.983)	FALSE
$\beta_5 + \beta_{11}$	0.269	0.239	(-0.199,0.737)	FALSE
$\beta_6 + \beta_{12}$	0.558	0.303	(-0.036,1.152)	FALSE
$\beta_7 + \beta_{13}$	0.099	0.441	(-0.765,0.963)	FALSE