>library(mice)

>library(lattice)

>library(dplyr)

>library(VIM)

>set.seed(271)

# Choose the sample size of 200.

>n<-200

# Simulate a random data frame

> data <- data.frame(sex = sample (c("F","M")), age = rnorm(n,15:80), bmi = rnorm(n,18:50), sbp = rnorm(n, 40:180), dbp = rnorm(n, 50:200), insulin = rnorm(n,1:50), smoke = rep(c(1, 2), 200))

# Print the first six rows of the dataset

>head(data)

# Manually add some missing values

> missing.data <- data %>%mutate(age = "is.na<-"(age, age <25 | age >75), bmi = "is.na<-"(bmi, bmi >44 | bmi <18), sbp = "is.na<-"(sbp, sbp >45 | sbp <20), dbp = "is.na<-"(dbp, dbp >180 | dbp <65))

> head(missing.data)

# How many patterns are there where the "bmi" variable is missing.

>mpattern <- md.pattern(missing.data)

> sum(mpattern[, "bmi"] == 0)

# Draw the aggr plot graph.

> aggr\_plot <- aggr(missing.data, col=c('red','yellow'), numbers=TRUE, sortVars=TRUE, labels=names(data), cex.axis=.7, gap=3, ylab=c("Histogram of missing data","Pattern"))

# Draw a box plot graph.

> marginplot(missing.data[c(6,3)])

# Impute missing values using the mice package

> imputation <- mice(missing.data, method = "mean", m = 4, maxit = 1)

> head(complete(imputation))

# Pooling the results and fitting a linear model

>ModelFit <- with(imputation, lm(insulin ~ bmi+age+sbp))

# Combine the results of the 4 models produced

> pool(ModelFit)

>summary(pool(ModelFit))

> densityplot(missing.data$bmi)

# Plot the imputed density graph.

>densityplot(imputation)

>imp <- mice(missing.data, seed = 271, print = FALSE)

Density plot original and imputed dataset

>densityplot(imp)

# Find the distribution of insulin variable according to other variables

> stripplot(imp, insulin ~ bmi+age+sbp+dbp, pch = 3, cex = 0.5)

# Filter the dataset.

> Orig.Df <- missing.data %>% dplyr::select(age, bmi, sbp, dbp)

>imp1 <- mice(Orig.Df, seed = 271, print = FALSE)

# Plotting scatterplots of observed and imputed data.

> stripplot(imp1)

# Check the convergence of the algorithm used

>imp2 <- mice(missing.data)

# Draw the trace lines of the variables.

>plot(imp2, c("bmi", "sbp", "dbp"))