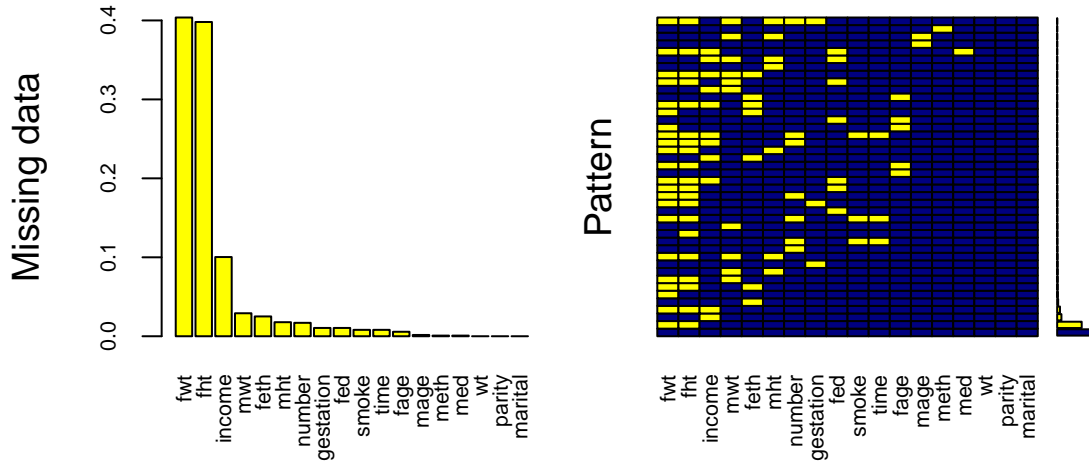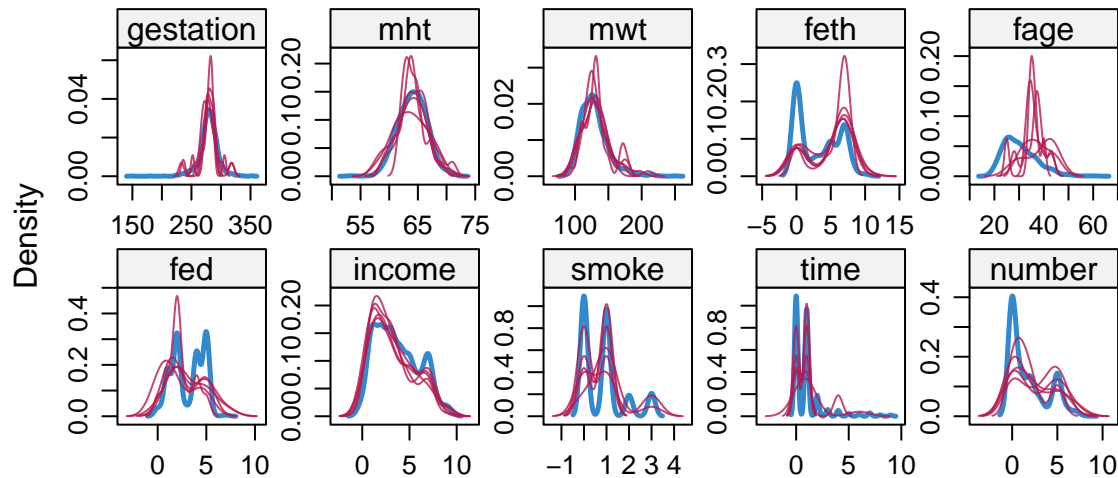# Data Imputation for Income

The functions in this file are dependent on the libraries VIM and mice

Visualizing the missing data

Looking the the above data frame, we might want to remove or impute values into the following features of the data, fwt : Fathers Weight,fht : Fathers Height, income : Fathers Income Since imputting may not work for fathers height and weight, we will remove those features from our data. Logically, it makes sense for the baby's weight to not depend on these features.

Using the mice library, we perform imputation for the income covariate

Blue - observed values Magenta - imputed values

From the density plot, we can say that imputed values are indeed plausible for the income covariate.

All the imputed data sets

Since the imputted values for income for all the imputed data sets follow a similar distribution, we choose imputed dataset 1, since it has the smallest difference in the means when compared the observed data

```
## [1] 0.01060839 0.23552749 0.07856956 0.08018768 0.02598056
```

We choose to include the imputed income data from imputed dataset 1 into our original dataset and then carry our model diagnostics and selection from there.