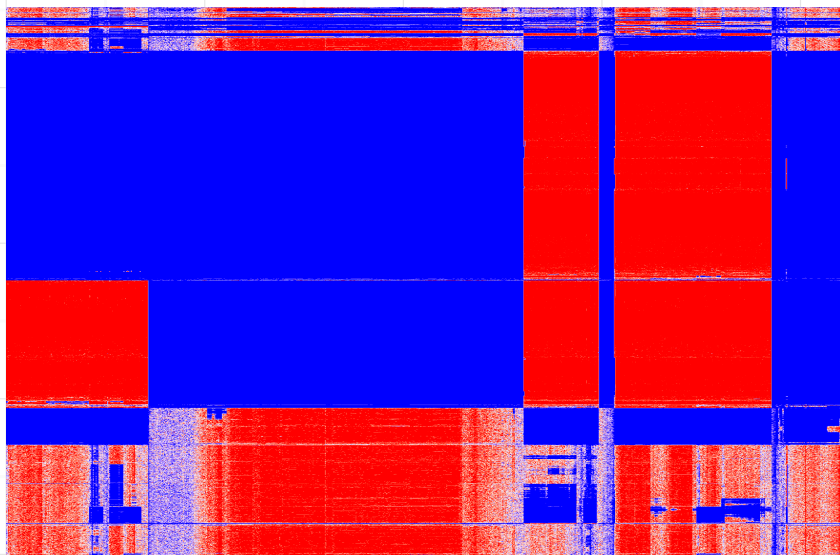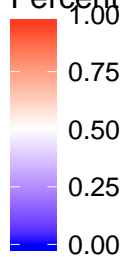Missing data pre-cleaning