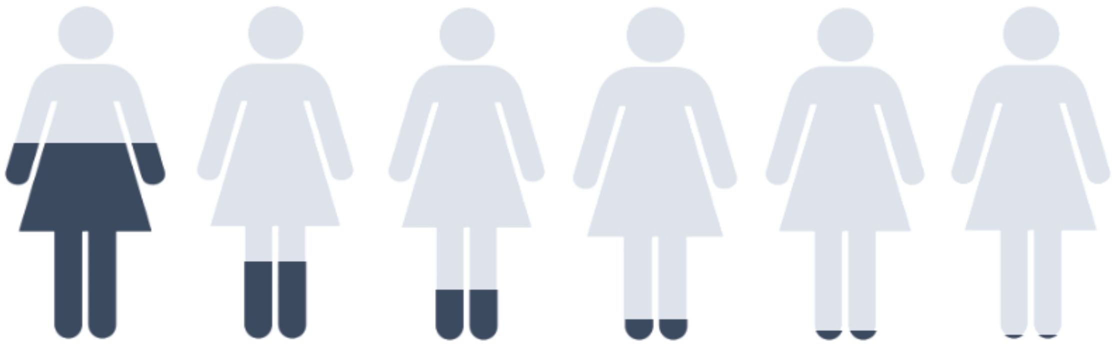
Presentation of homework no. 4

Missing Data:
PIMA INDIANS DIABETES DATSET

Tom Rupnik Medjedovič
Neža Kržan

Ljubljana, 10.1.2025

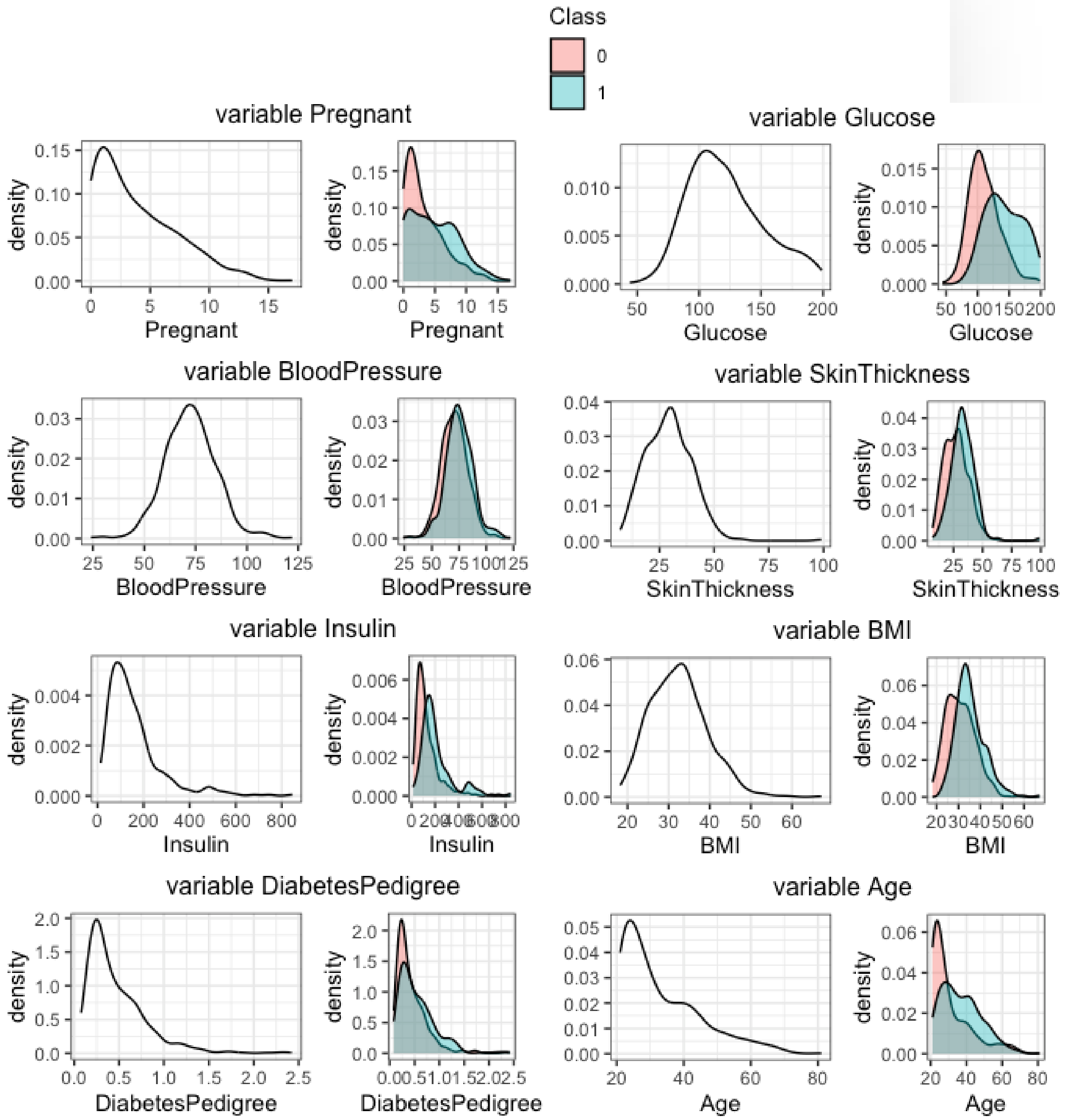
Health Data of Women from the Pima Indian Tribe



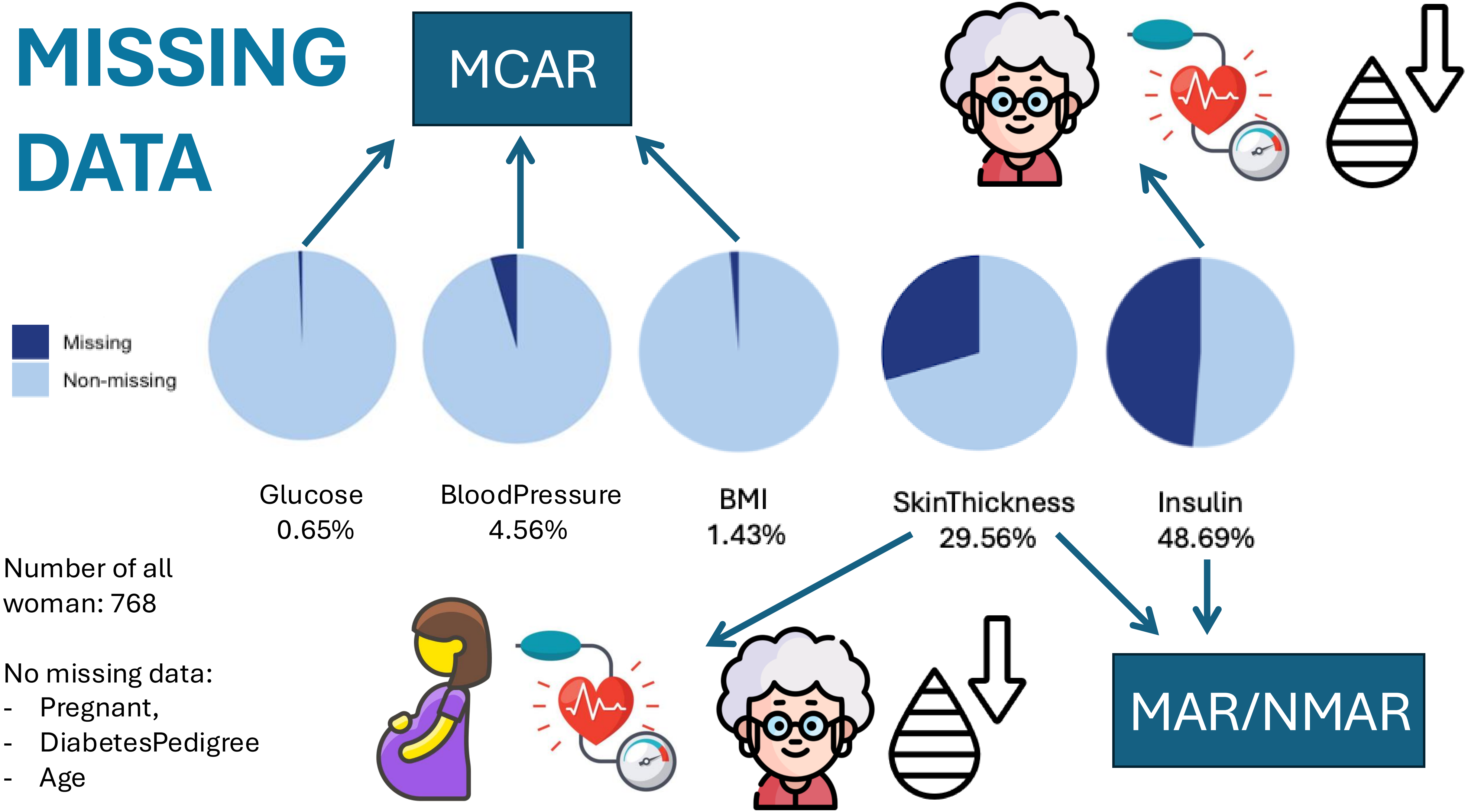
Age:	21-30	31-40	41-50	51-60	61-70	71-81
	54%	20%	15%	7%	3%	1%

Pregnant	*number of pregnancies*
Glucose	*plasma glucose concentration*
BloodPressure	*diastolic blood pressure(mm Hg)*
SkinThickness	*triceps skin fold thickness*
Insulin	*2-hour serum insulin*
BMI	*Body Mass Index*
DiabetesPedigree	*genetic predisposition*
Age	*age*
DiabetesClass	*Diabetes diagnosis (1 = Positive, 0 = Negative)*

- dataset primarily used for diabetes research,
- explore factors contributing to diabetes development



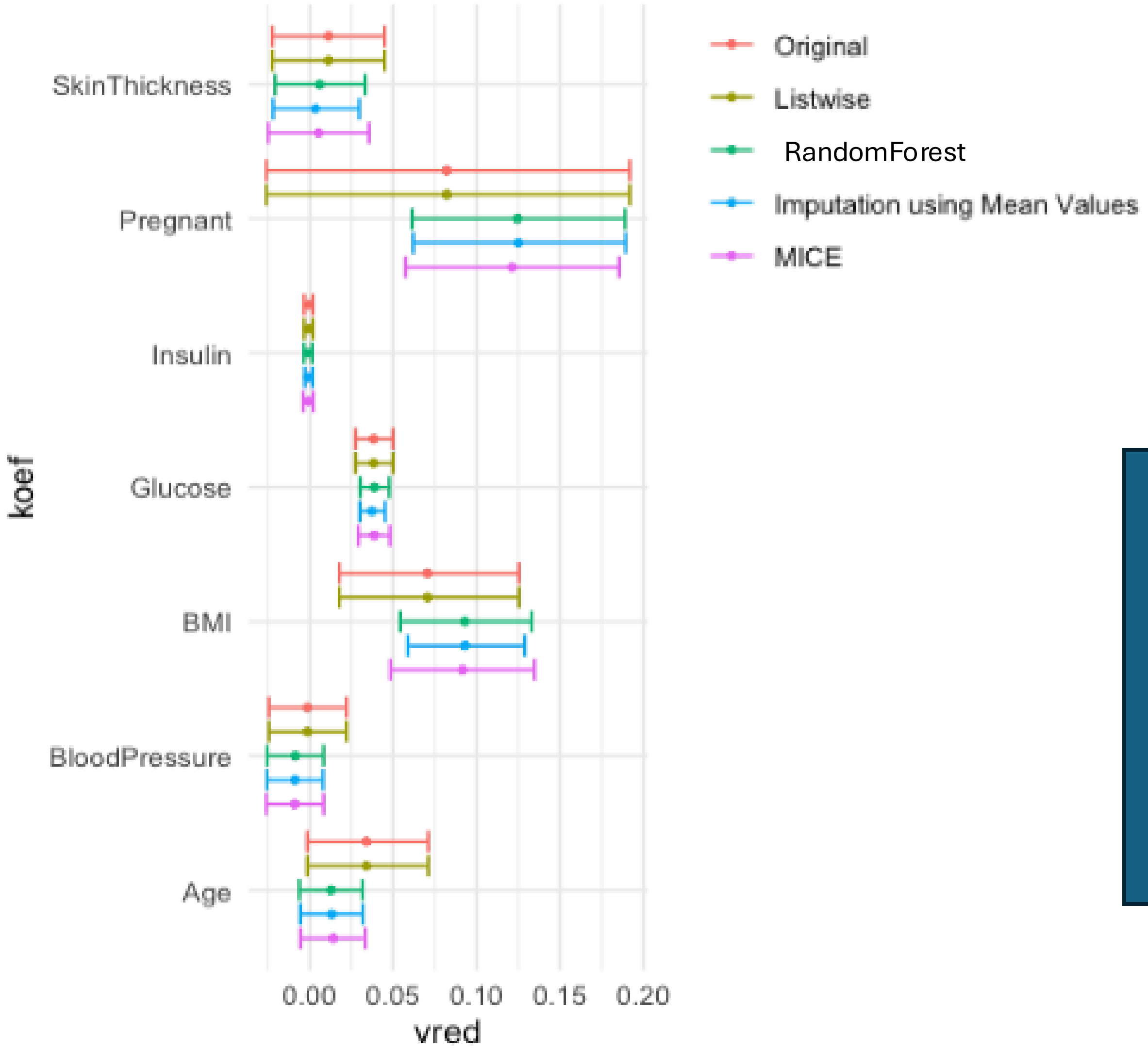
MISSING DATA



Number of all woman: 768

No missing data:
- Pregnant,
- DiabetesPedigree
- Age

MISSING VALUE IMPUTATION



Based on the results of our analysis, we can divide the methods into two groups.

listwise or pairwise deletion method

RandomForest,
Imputation using Mean Values,
MICE