

Matrix 1.12.

Imputation

- get Matrix with missing values

- impute only missing values

→ Decision Tree / Random Forest

→ can't we write a general function which calculates
a Decision Tree or anything else?

Meeting Jeroen

Questions:

- What does count for grading?
- How would be a real prototype handled?
- Is it best practice to have different datasets?

Data Preparation

Possible Datasets:

- t0 without pcm
- t0
- t0 + Questionnaire
- t0 + CBS
- t0 + CBS + Questionnaire

Cleaning Data:

- calculate MQ-score
- calculate MQ-category
- convert numerical columns to numeric
- remove outliers
- remove nan
- remove columns with $\text{nan} \geq 20\%$

Fellende Columns:

- zip-code
- MQ
- MQ category

TO Data

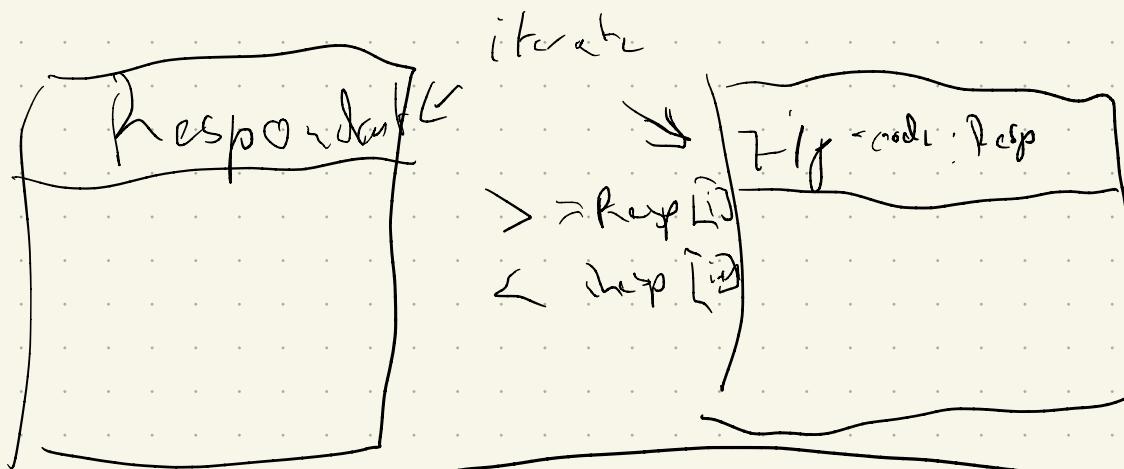
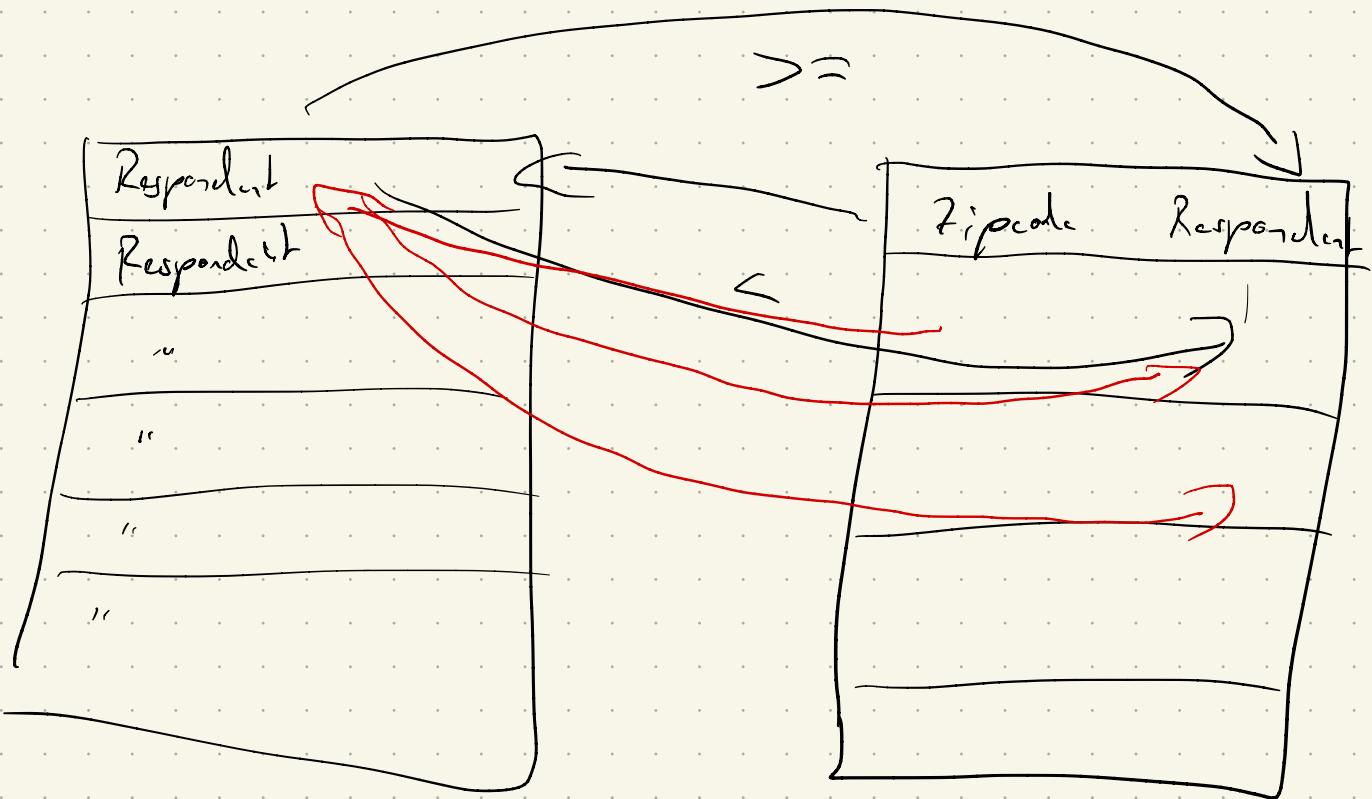
- Groningen
 - Total TO 221120
 - Copy of schools -> postcodes
- } merges

Problem

- Groningen TO Data column names are different
- convert columns konvertiert Geburtsdaten
→ dadurch werden alle gelöscht

Zip codes

- von Respondentnummer zwischen df[0] und df[1] dann min - df[0]
- ansonsten versch df[1] und df[2]
- speicher den Wert dann in der



Respondent[0] $\xrightarrow{>=}$ Zipcode[0] = Resp[0]

Respondent[1] $\xrightarrow{>=}$ Zipcode[1] = Resp[1]

Respondent[2] $\xrightarrow{>=}$ Zipcode[2] = Resp[2]