HELMi metaData overview

Goals and objectives:

This documents aims to provide an overview on the HELMi contextual dataset extracted from the families questionnaires.

The following aspects are considered:

- Types and sparcity of the contextual dataset
- Family compliance

Dataset description:

infants questionnaires:

- A_background
- F1_NutritionMotorDev_year1_version1 or F1_NutritionMotorDev_year1_version2
- F2_NutritionMotorDev_year2_version1 or F2_NutritionMotorDev_year2_version2

infants frequency questionnaires:

- E_Previous3Months
- C_4to6months
- D_7to12months
- B_0to3months

parents questionnaires:

- G_ParentF_FoodFrequency
- G_ParentM_FoodFrequency
- G_MotherDepression
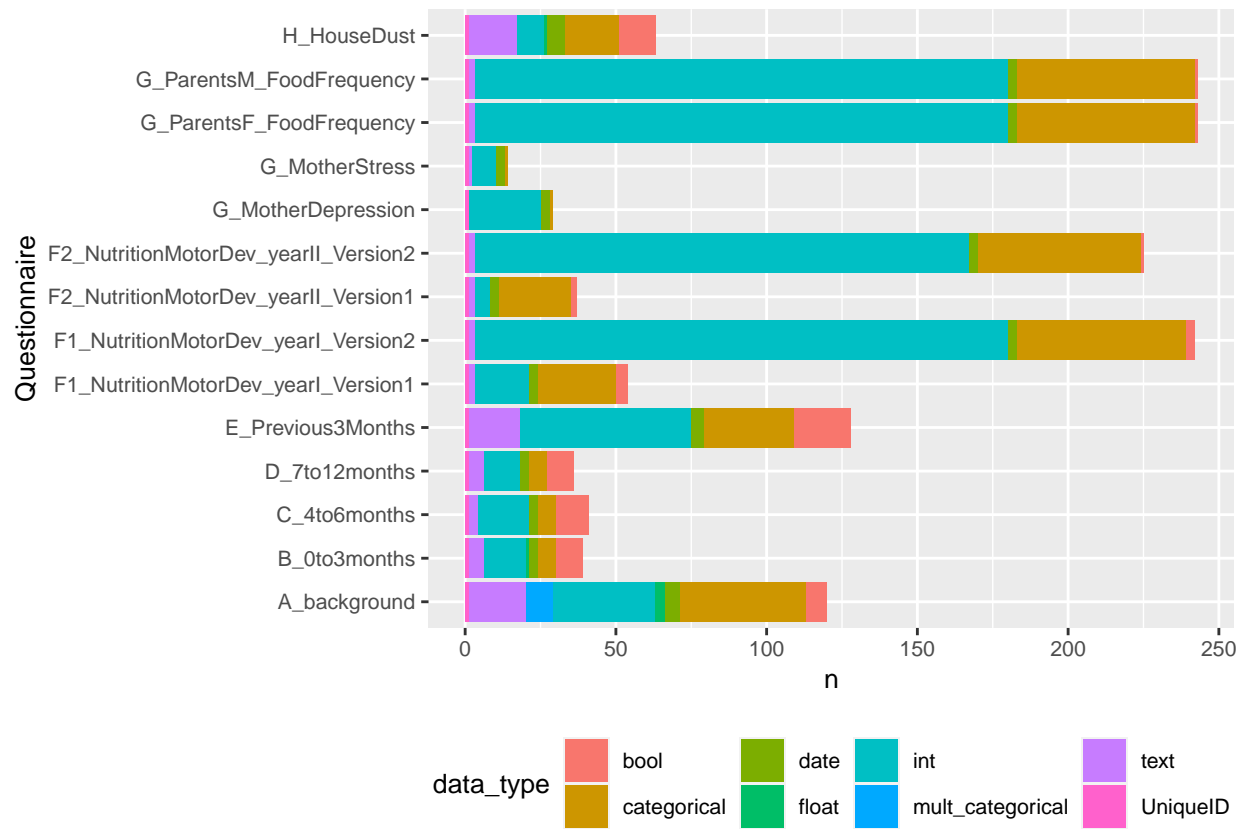- G_MotherStress (in progress)

Other questionnaires:

- H_HouseCollection

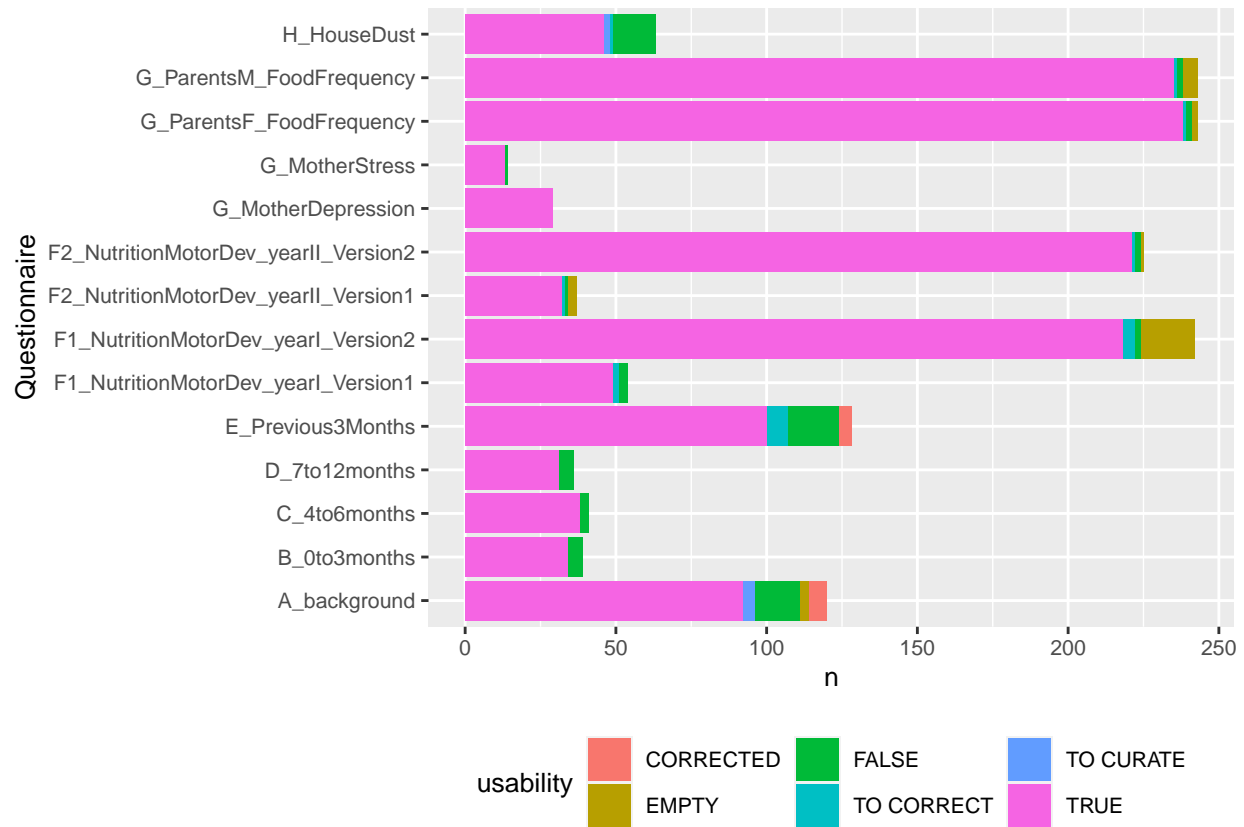| Field_category	Field_Subcategory	Background	Birth	0-3 months	4-6 months	7-12 months	12-24 months
Environment_Lifestyle	Animals_Pets	x	x	x	x	x	x
	Cleaning habits	x	x	x	x	x	x
	Home_Household	x	x	x	x	x	x
Familly	Father health	x					
	Mother health	x	x				
	Mother lifestyle	x					
	Mother mental health						
	Mother nutrition					x	x
	Partner lifestyle	x					
	Partner nutrition					x	x
	Sibling	x					
Infant general	Infant general	x	x	x	x	x	x
Infant health	Allergy_Skin			x	x	x	x
	General health			x	x	x	x
	GI			x	x	x	x
	Medications_Vaccines			x	x	x	x
	motor					x	
	Respiratory			x	x	x	x
Infant nutrition	Breastfeeding		x	x	x	x	x
	Other			x	x	x	x
Pregnancy_Birth	Birth general		x				
	Health after birth		x				
	Health during pregnancy	x	x				
	Lifestyle before pregnancy	x					
	Lifestyle during pregnancy	x	x				
	Pregnancy general	x	x				

General overview

Types of Metadata

Overview of the type of metadata in each questionnaire. Note that "mult_categorical" corresponds to multiple choice categories.
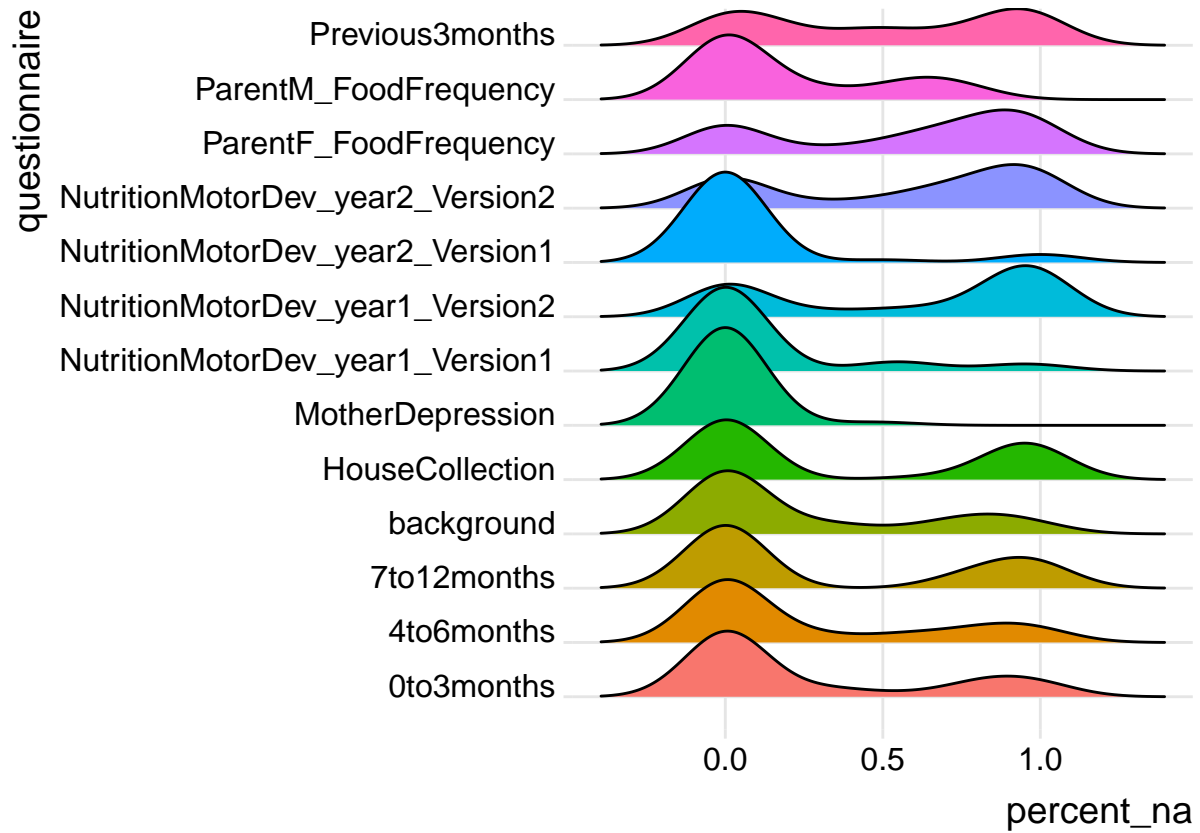
Usability of Metadata

Overview of the direct usability of the contextual dataset. "TO CURATE" indicates fields where a manual extraction of the data would be vaulable. "TO CORRECT" indicates fields where a conversion or type correction would be valuable. "CORRECTED" indicates a field where the data was converted or corrected in a major way (this do not include the simple category translation).

Questionnaire sparcity

Looking at the number of NA per column in each questionnaire
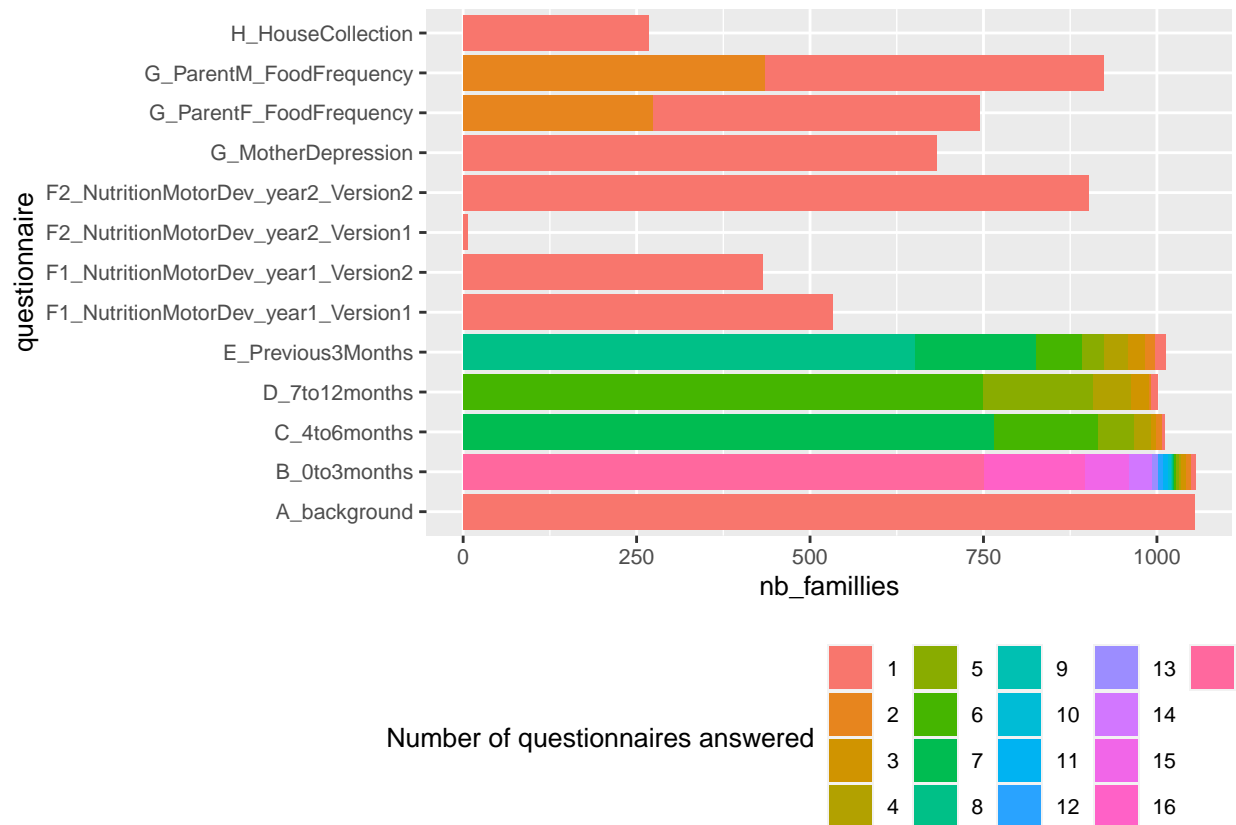
```
## Picking joint bandwidth of 0.132
```

Familly compliance

Compliance overview

Table 1: number of famillies answering the questionnaire at least once

questionnaire	nb of famillies
A__background	1055
B__0to3months	1056
C__4to6months	1011
D__7to12months	1002
E__Previous3Months	1012
F1__NutritionMotorDev__year1__Version1	532
F1__NutritionMotorDev__year1__Version2	431
F2__NutritionMotorDev__year2__Version1	6
F2__NutritionMotorDev__year2__Version2	901
G__MotherDepression	683
G__ParentF__FoodFrequency	745
G__ParentM__FoodFrequency	923
H__HouseCollection	268

Family compliance in questionnaires

overview per familly: Famillies answering all the regular questionnaires

measures	nb families
all infants	890
all infants frequency	426
all parents	199
all major questionnaires	121
all major infants and frequency	422
all major infants and most frequency	766