Methods Library

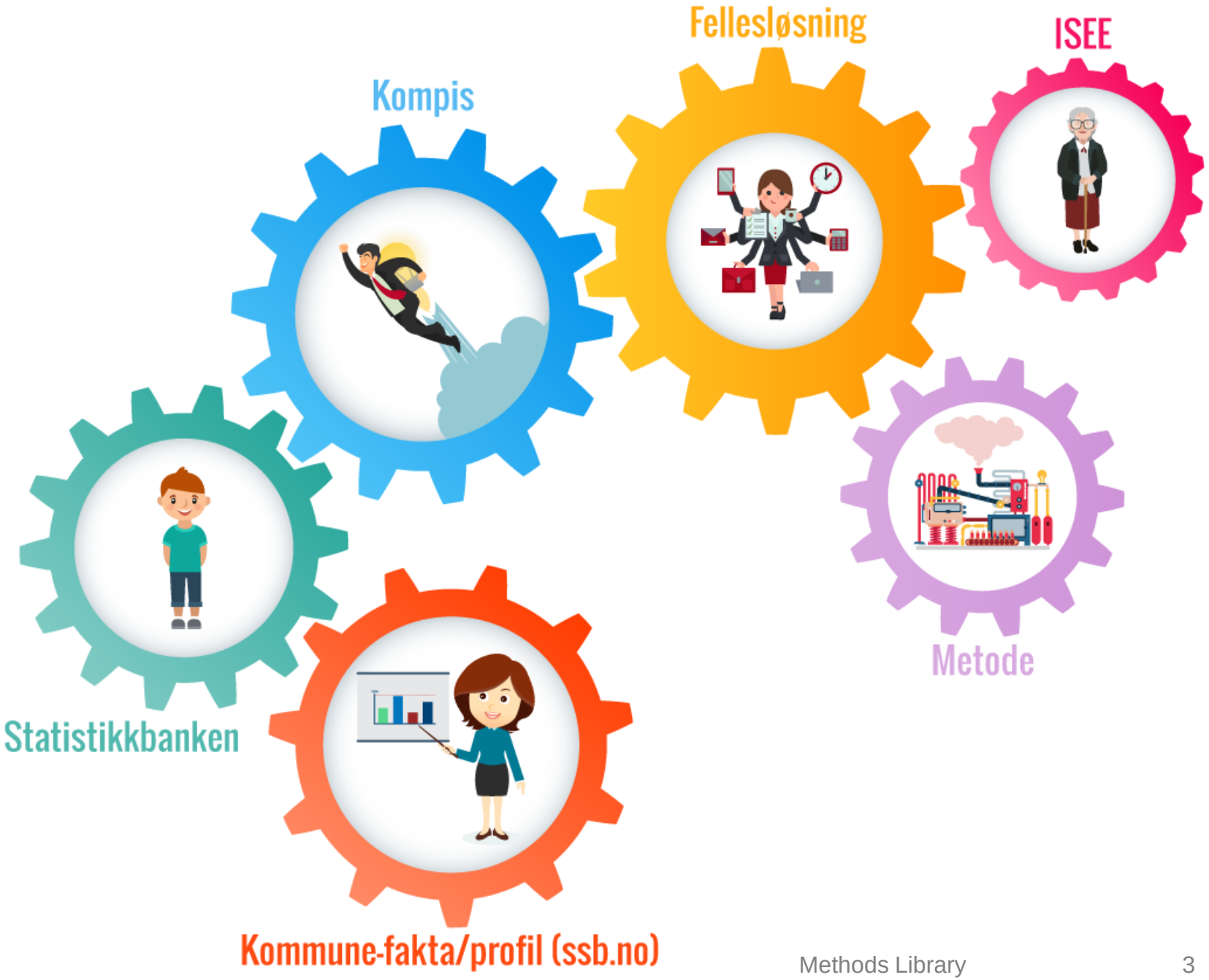
as part of the modernization of the statistical production in Norway

Statistisk sentralbyrå
Statistics Norway

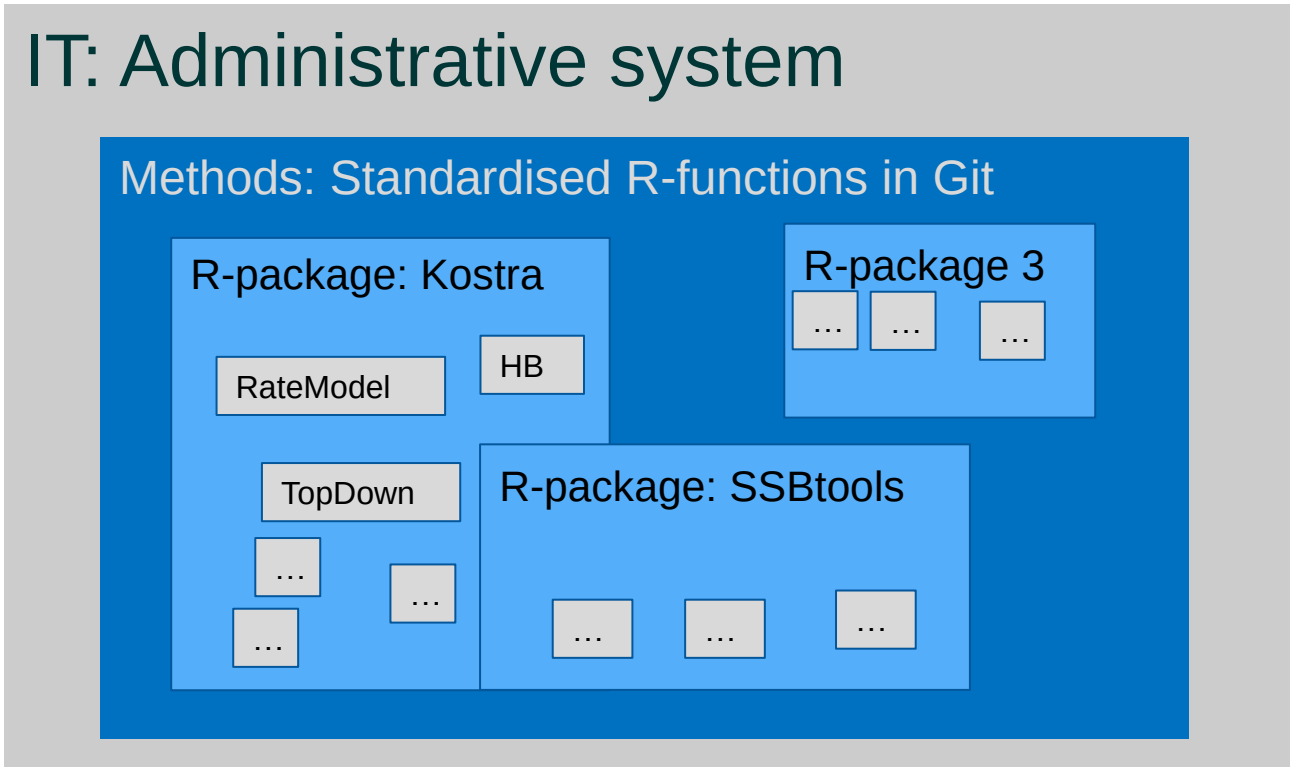
UNECE - Work Session on Statistical Data Editing
24-26 April 2017

Susie Jentoft on behalf of Aslaug, Øyvind & Ane

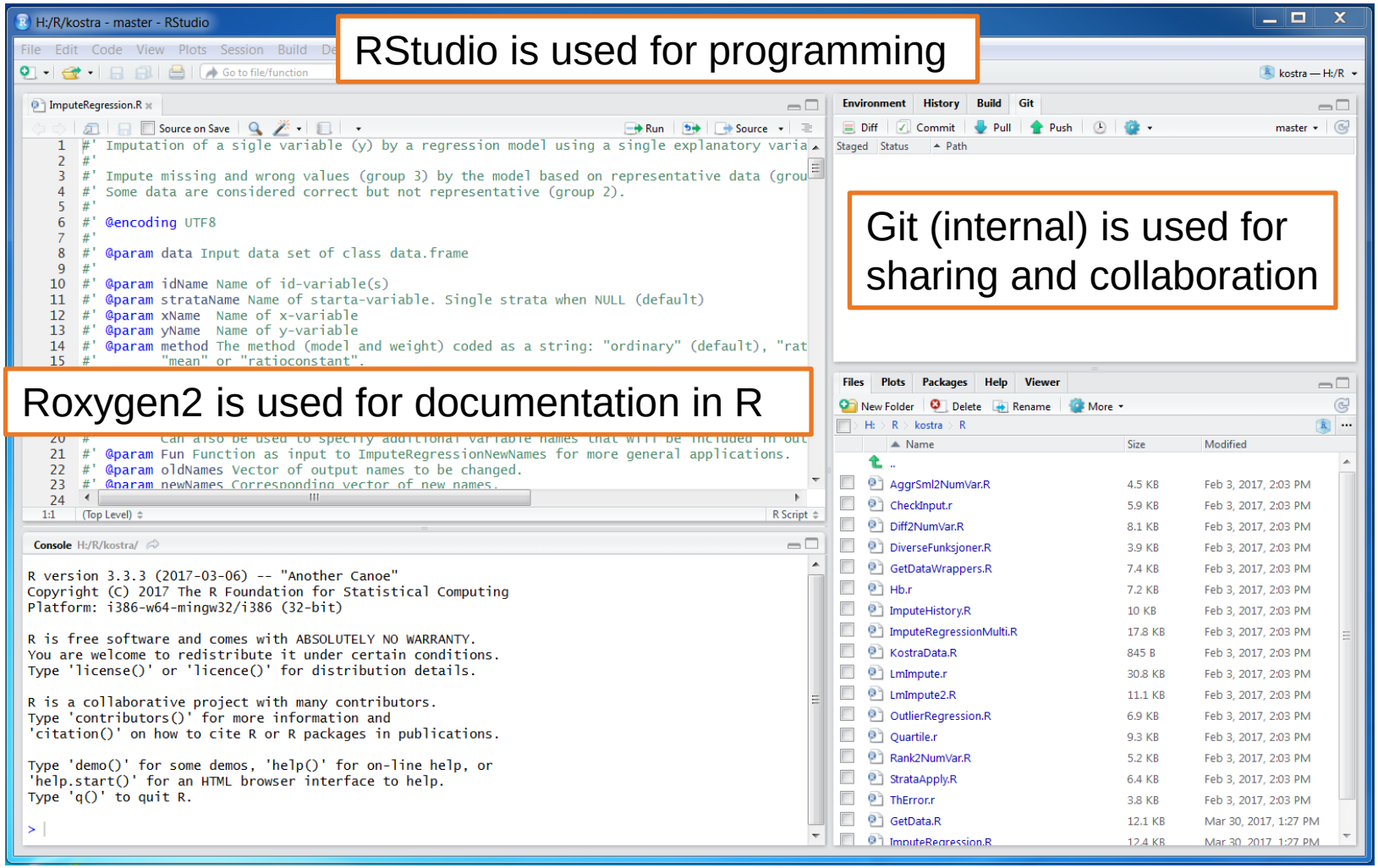
NORGE


Fellesløsning
ISEE
Kompis
Statistikkbanken
Metode
Kommune-fakta/profil (ssb.no)

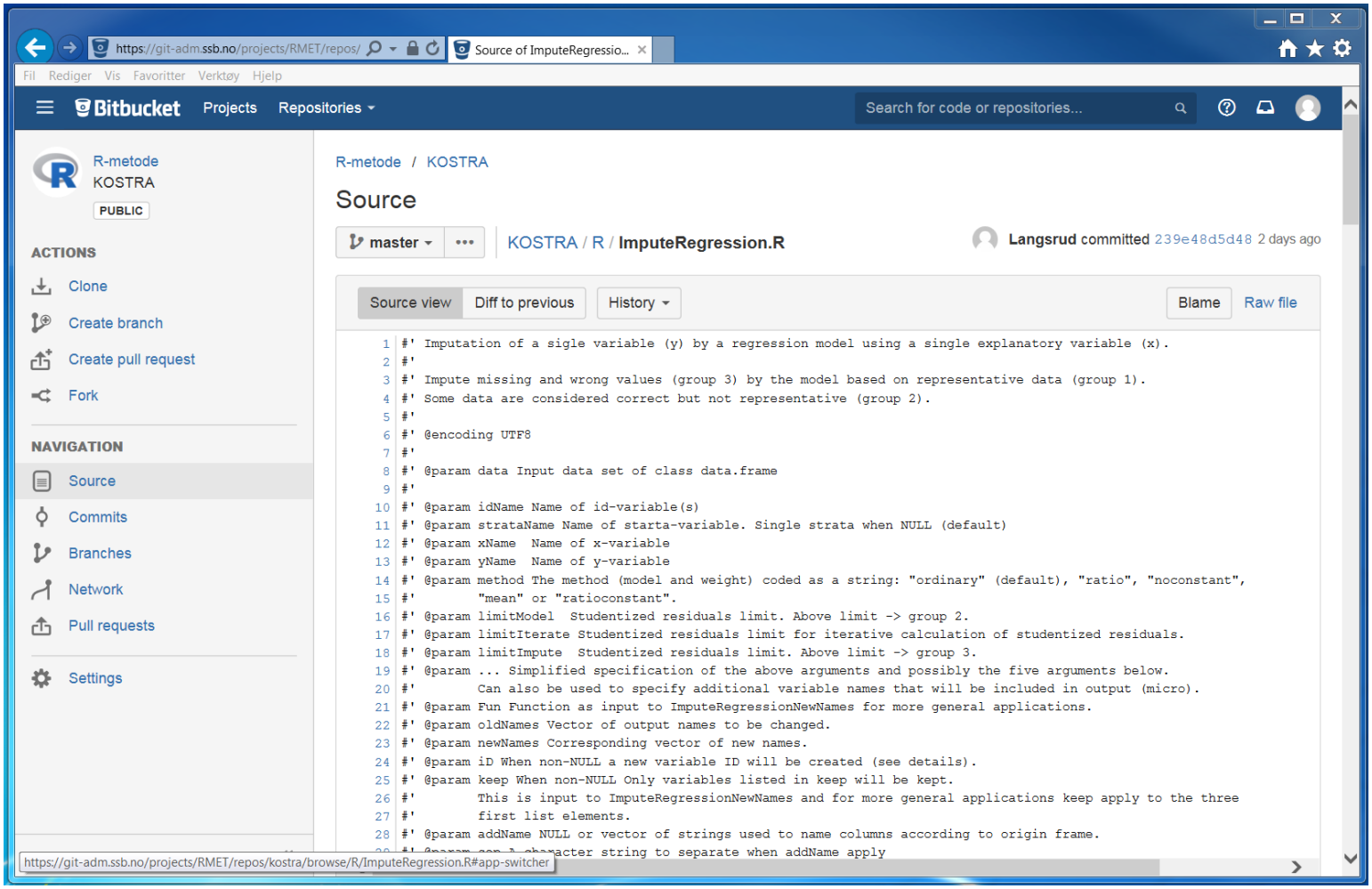
Methods Library



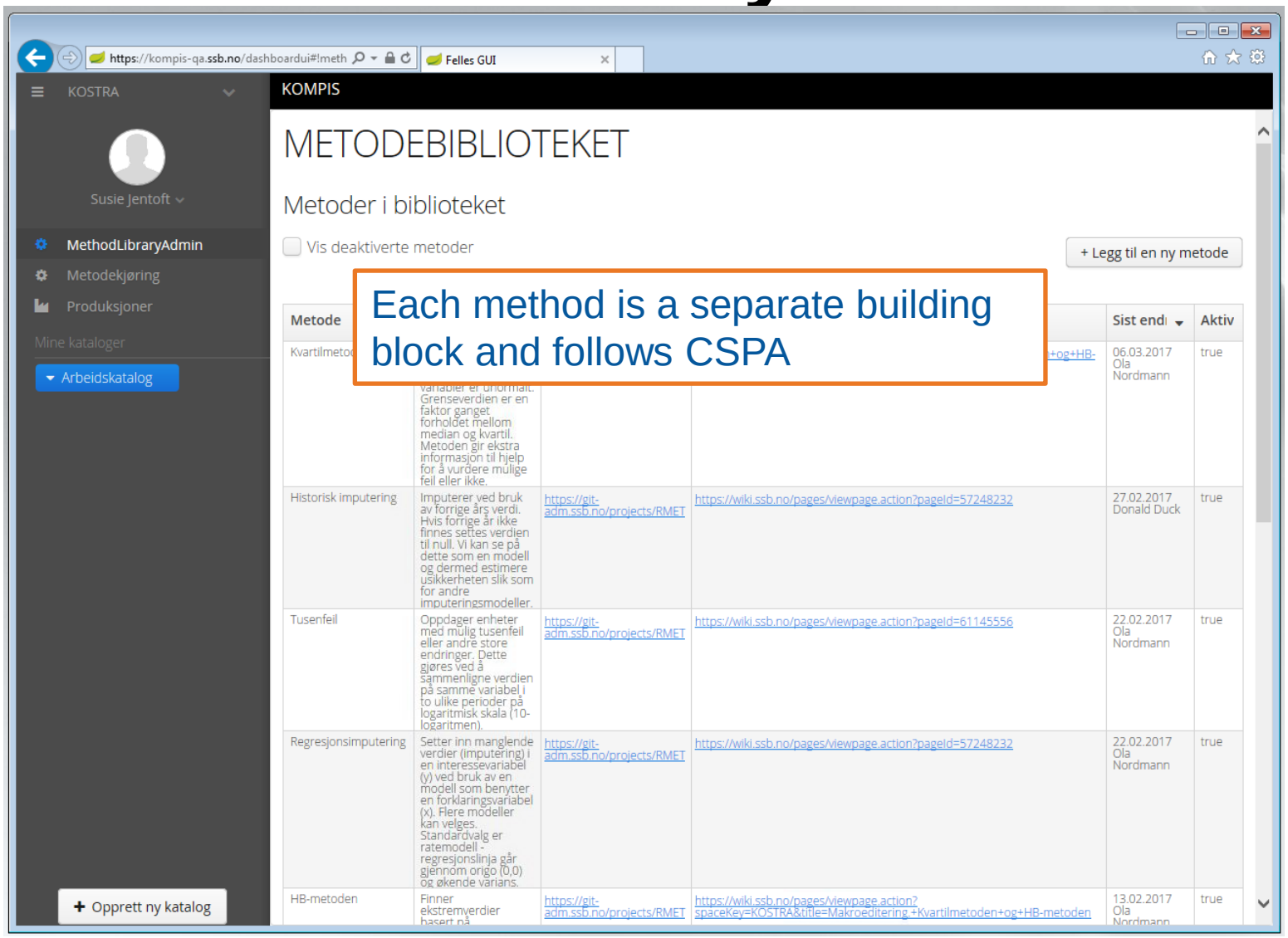
R functions and packages



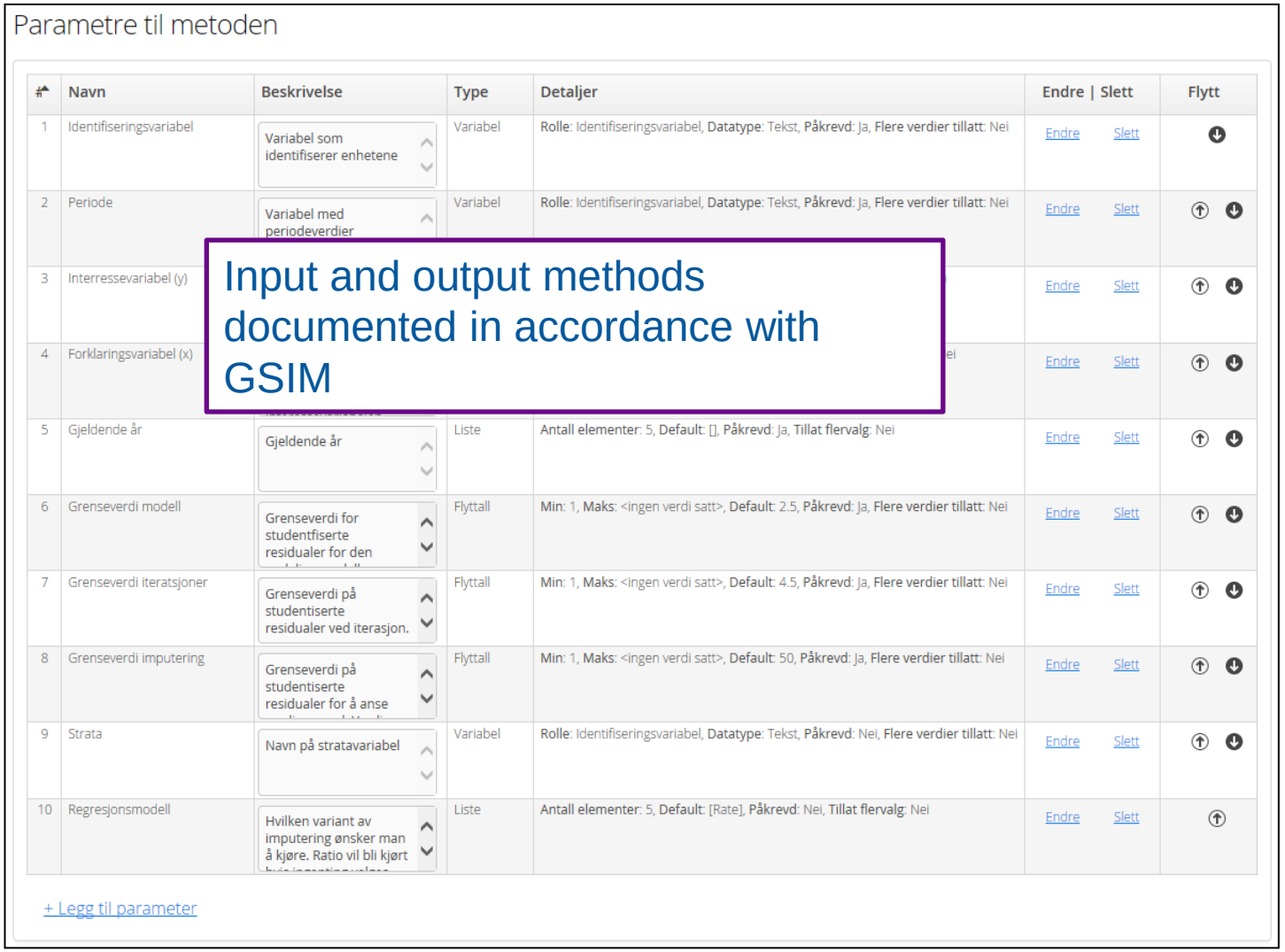
Git


https://git-adm.ssb.no/projects/RMET/repos/
Source of ImputeRegressio...
Bitbucket
Projects
Repositories
Search for code or repositories...
R-metode
KOSTRA
PUBLIC
ACTIONS
Clone
Create branch
Create pull request
Fork
NAVIGATION
Source
Commits
Branches
Network
Pull requests
Settings
R-metode / KOSTRA
Source
master
KOSTRA / R / ImputeRegression.R
Langsrud committed 239e48d5d48 2 days ago
Source view
Diff to previous
History
Blame
Raw file
#' Imputation of a sigle variable (y) by a regression model using a single explanatory variable (x).
#'
#' Impute missing and wrong values (group 3) by the model based on representative data (group 1).
#' Some data are considered correct but not representative (group 2).
#'
#' @encoding UTF8
#'
#' @param data Input data set of class data.frame
#'
#' @param idName Name of id-variable(s)
#' @param strataName Name of starta-variable. Single strata when NULL (default)
#' @param xName Name of x-variable
#' @param yName Name of y-variable
#' @param method The method (model and weight) coded as a string: "ordinary" (default), "ratio", "noconstant",
#'   "mean" or "ratioconstant".
#' @param limitModel Studentized residuals limit. Above limit -> group 2.
#' @param limitIterate Studentized residuals limit for iterative calculation of studentized residuals.
#' @param limitImpute Studentized residuals limit. Above limit -> group 3.
#' @param ... Simplified specification of the above arguments and possibly the five arguments below.
#'   Can also be used to specify additional variable names that will be included in output (micro).
#' @param Fun Function as input to ImputeRegressionNewNames for more general applications.
#' @param oldNames Vector of output names to be changed.
#' @param newNames Corresponding vector of new names.
#' @param iD When non-NULL a new variable ID will be created (see details).
#' @param keep When non-NULL Only variables listed in keep will be kept.
#'   This is input to ImputeRegressionNewNames and for more general applications keep apply to the three
#'   first list elements.
#' @param addName NULL or vector of strings used to name columns according to origin frame.
https://git-adm.ssb.no/projects/RMET/repos/kostra/browse/R/ImputeRegression.R#app-switcher

Administrative system



Administrative system



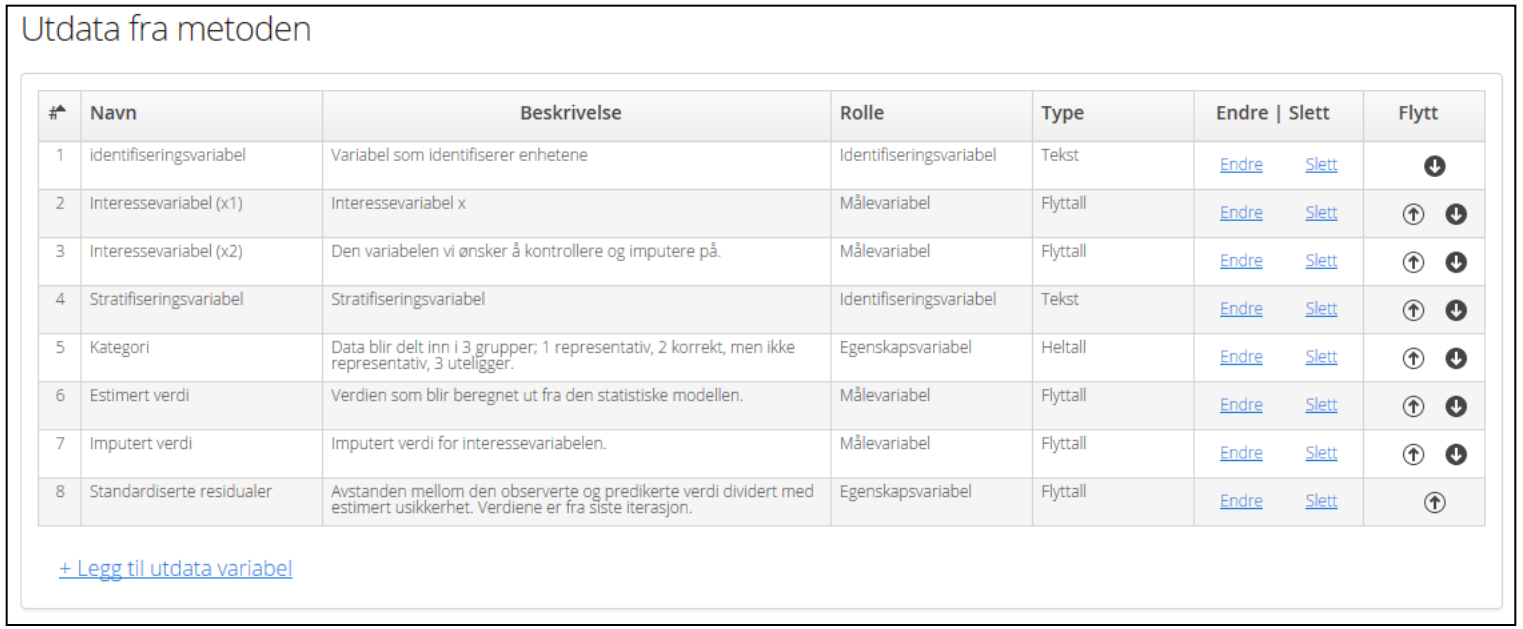
Parametre til metoden

#	Navn	Beskrivelse	Type	Detaljer	Endre | Slett	Flytt
1	Identifiseringsvariabel	Variabel som identifiserer enhetene	Variabel	**Rolle**: Identifiseringsvariabel, **Datatype**: Tekst, **Påkrevd**: Ja, **Flere verdier tillatt**: Nei	Endre Slett	
2	Periode	Variabel med periodeverdier	Variabel	**Rolle**: Identifiseringsvariabel, **Datatype**: Tekst, **Påkrevd**: Ja, **Flere verdier tillatt**: Nei	Endre Slett	
3	Interressevariabel (y)				Endre Slett	
4	Forklaringsvariabel (x)				Endre Slett	
5	Gjeldende år	Gjeldende år	Liste	**Antall elementer**: 5, **Default**: [], **Påkrevd**: Ja, **Tillat flervalg**: Nei	Endre Slett	
6	Grenseverdi modell	Grenseverdi for studentfiserte residualer for den	Flyttall	**Min**: 1, **Maks**: <ingen verdi satt>, **Default**: 2.5, **Påkrevd**: Ja, **Flere verdier tillatt**: Nei	Endre Slett	
7	Grenseverdi iteratsjoner	Grenseverdi på studentiserte residualer ved iterasjon.	Flyttall	**Min**: 1, **Maks**: <ingen verdi satt>, **Default**: 4.5, **Påkrevd**: Ja, **Flere verdier tillatt**: Nei	Endre Slett	
8	Grenseverdi imputering	Grenseverdi på studentiserte residualer for å anse	Flyttall	**Min**: 1, **Maks**: <ingen verdi satt>, **Default**: 50, **Påkrevd**: Ja, **Flere verdier tillatt**: Nei	Endre Slett	
9	Strata	Navn på stratavariabel	Variabel	**Rolle**: Identifiseringsvariabel, **Datatype**: Tekst, **Påkrevd**: Nei, **Flere verdier tillatt**: Nei	Endre Slett	
10	Regresjonsmodell	Hvilken variant av imputering ønsker man å kjøre. Ratio vil bli kjørt	Liste	**Antall elementer**: 5, **Default**: [Rate], **Påkrevd**: Nei, **Tillat flervalg**: Nei	Endre Slett	

+ Legg til parameter

Input and output methods documented in accordance with GSIM

Administrative system



Utdata fra metoden

#	Navn	Beskrivelse	Rolle	Type	Endre | Slett	Flytt
1	identifiseringsvariabel	Variabel som identifiserer enhetene	Identifiseringsvariabel	Tekst	Endre Slett	
2	Interessevariabel (x1)	Interessevariabel x	Målevariabel	Flyttall	Endre Slett	
3	Interessevariabel (x2)	Den variabelen vi ønsker å kontrollere og imputere på.	Målevariabel	Flyttall	Endre Slett	
4	Stratifiseringsvariabel	Stratifiseringsvariabel	Identifiseringsvariabel	Tekst	Endre Slett	
5	Kategori	Data blir delt inn i 3 grupper; 1 representativ, 2 korrekt, men ikke representativ, 3 utelligger.	Egenskapsvariabel	Heltall	Endre Slett	
6	Estimert verdi	Verdien som blir beregnet ut fra den statistiske modellen.	Målevariabel	Flyttall	Endre Slett	
7	Imputert verdi	Imputert verdi for interessevariabelen.	Målevariabel	Flyttall	Endre Slett	
8	Standardiserte residualer	Avstanden mellom den observerte og predikerte verdi dividert med estimert usikkerhet. Verdiene er fra siste iterasjon.	Egenskapsvariabel	Flyttall	Endre Slett	

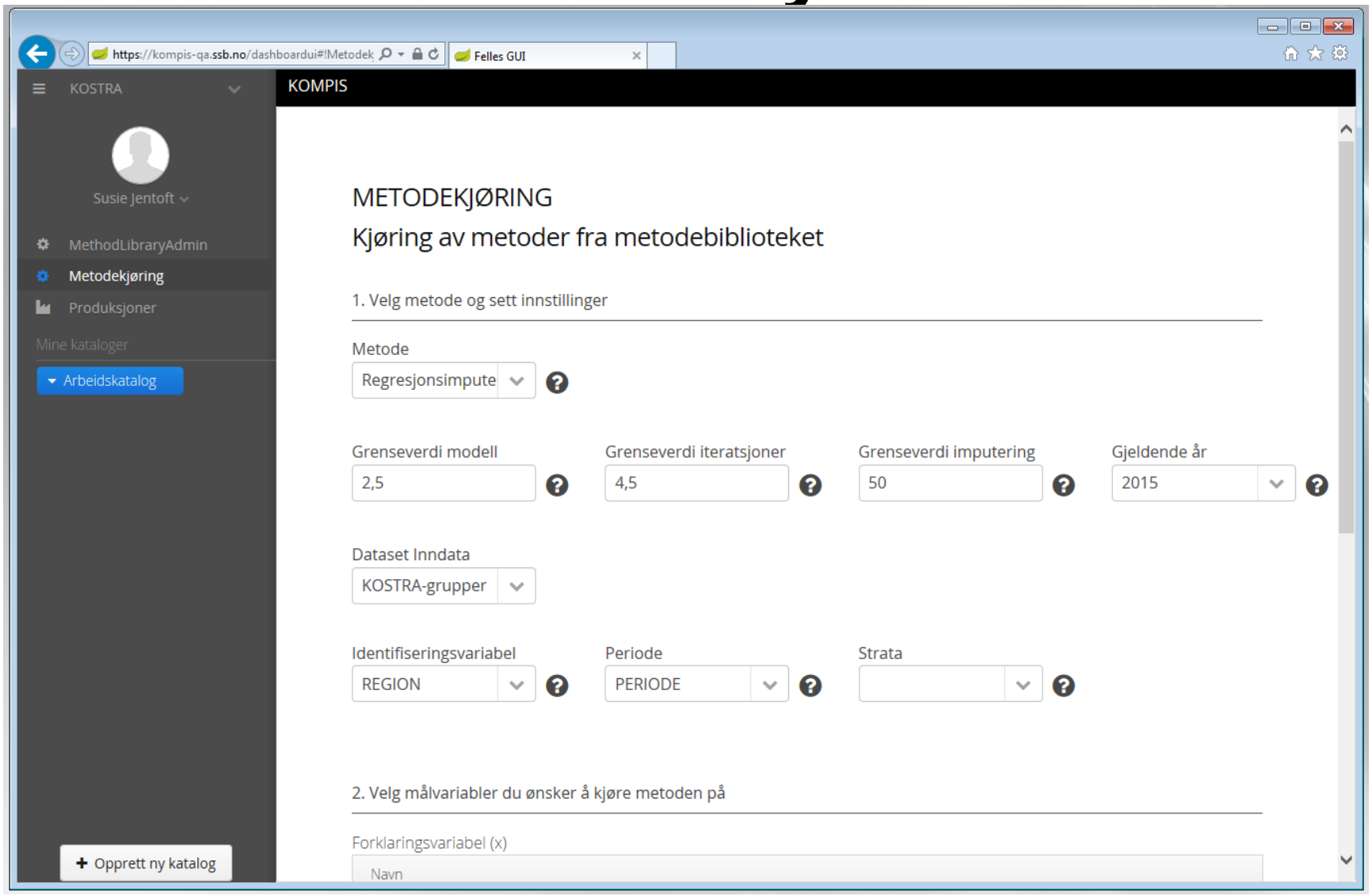
+ Legg til utdata variabel

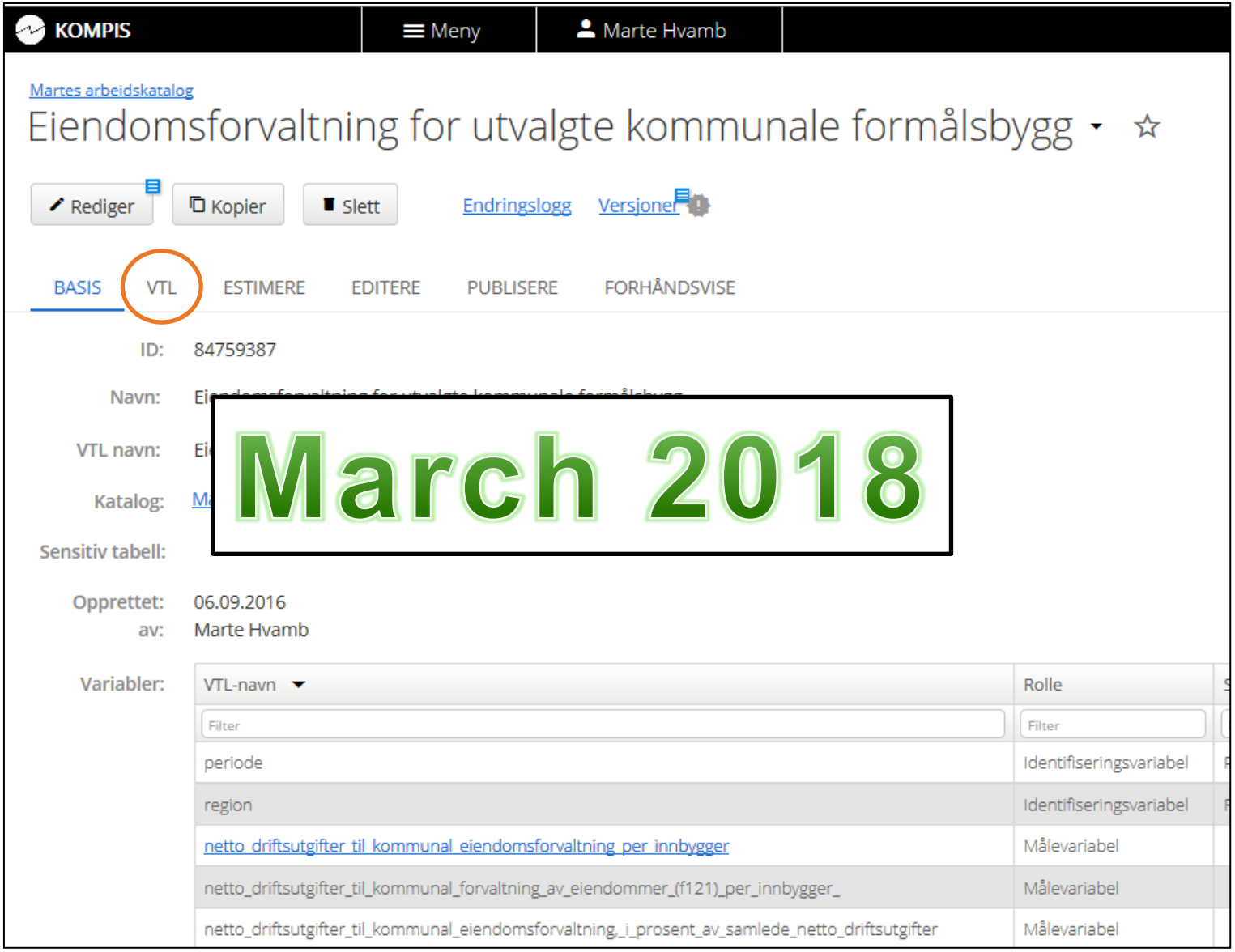
Grafkonfigurasjoner til metoden

Graftype	Endre | Slett
Scatterplot	Endre Slett

+ Legg til grafkonfigurasjon

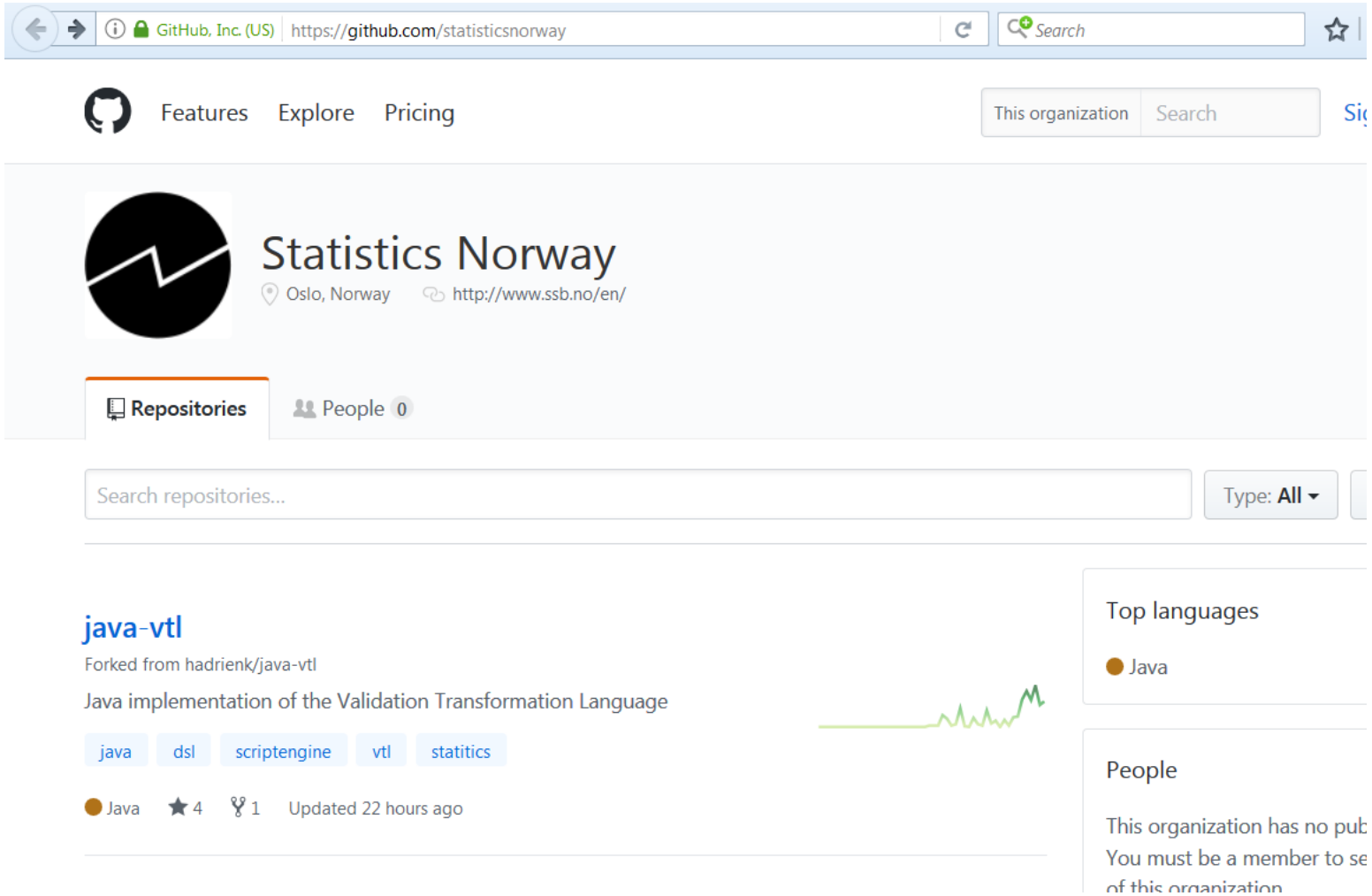
Administrative system




KOMPIS
Meny
Marte Hvamb
Martes arbeidskatalog
Eiendomsforvaltning for utvalgte kommunale formålsbygg
Rediger
Kopier
Slett
Endringslogg
Versjoner
BASIS
VTL
ESTIMERE
EDITERE
PUBLISERE
FORHÅNDSVISE
ID: 84759387
Navn:
VTL navn:
Katalog:
Sensitiv tabell:
March 2018
Opprettet: 06.09.2016
av: Marte Hvamb
Variabler:
VTL-navn
Rolle
Filter
Filter
periode
Identifiseringsvariabel
region
Identifiseringsvariabel
netto_driftsutgifter_til_kommunal_eiendomsforvaltning_per_innbygger
Målevariabel
netto_driftsutgifter_til_kommunal_forvaltning_av_eiendommer_(f121)_per_innbygger_
Målevariabel
netto_driftsutgifter_til_kommunal_eiendomsforvaltning,_i_prosent_av_samlede_netto_driftsutgifter
Målevariabel

https://vtl-demo.ssb.no/

https://github.com/statisticsnorway



Thank you

susie.jentoft@ssb.no