

Classy classifications in R

SUSIE JENTOFT

02.12.2020



Statistisk sentralbyrå
Statistics Norway

Coding of sex within Statistics Norway

1 = Male
2 = Female

'M' = Male
'F' = Female

'Male' = Male
'Female' = Female

1 = Male
2 = Female
3 = Unknown

2 = Male
1 = Female

'M' = Male
'W' = Female

'Menn' = Male
'Kvinner' = Female

1 = Male
2 = Female
9 = Unknown

+ ...

0 = Male
1 = Female

'M' = Male
'K' = Female

'Men' = Male
'Women' = Female

1 = Male
2 = Female
3 = Refusal
9 = Unknown

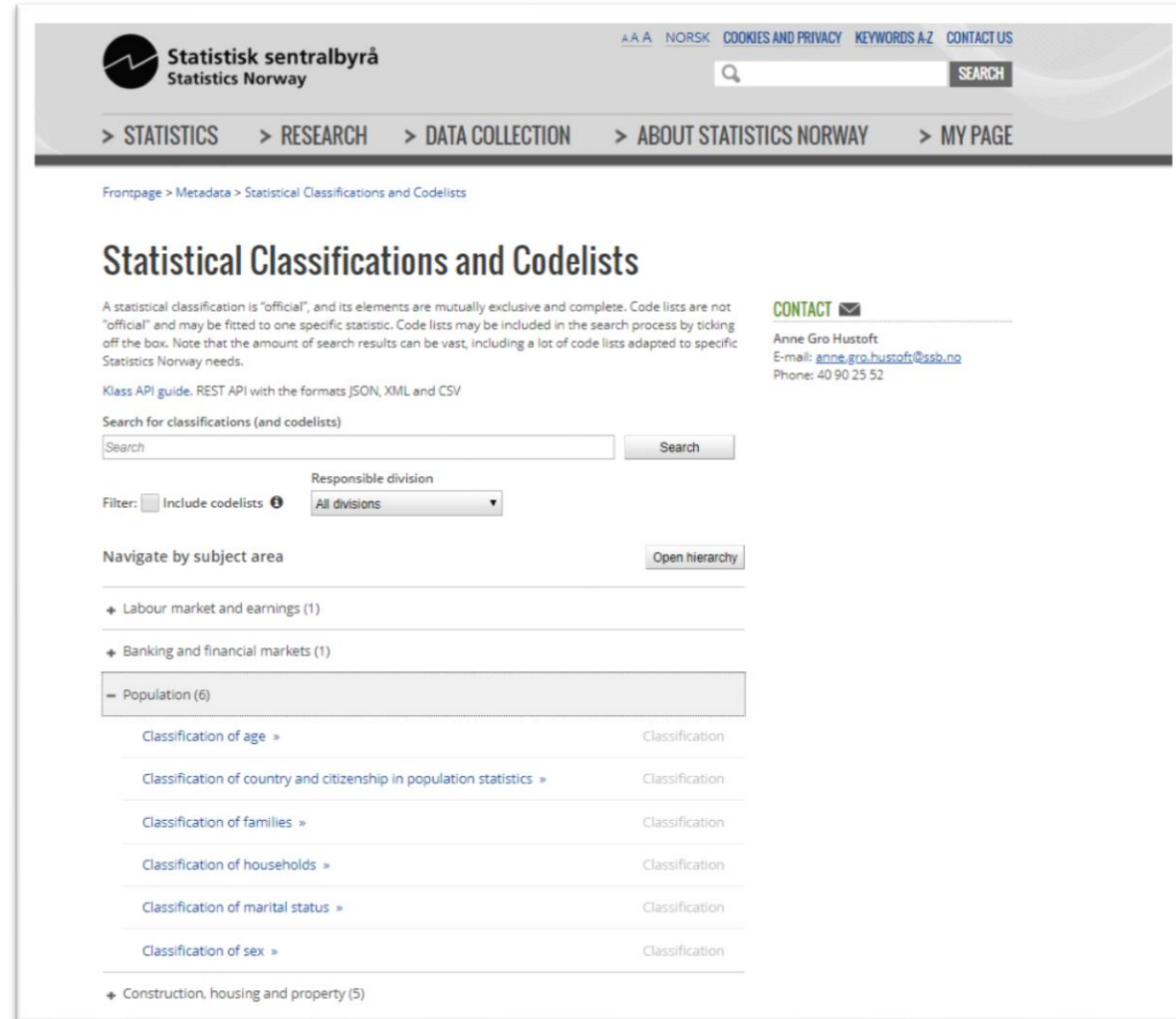
'm' = Male
'f' = Female

'Mann' = Male
'Dame' = Female



KLASS system

- Database for documenting and managing Statistical Classifications and Code Lists
- API for machine to machine communication
- Web page (www.ssb.no/klass) for searching and downloading Classifications and Code Lists
- Based on the GSIM – Statistical Classification model



The screenshot shows the website for Statistisk sentralbyrå (Statistics Norway). The header includes the logo, the name 'Statistisk sentralbyrå Statistics Norway', and navigation links for 'STATISTICS', 'RESEARCH', 'DATA COLLECTION', 'ABOUT STATISTICS NORWAY', and 'MY PAGE'. A search bar is also present.

The main content area is titled 'Statistical Classifications and Codelists'. It contains a search bar with the text 'Search for classifications (and codelists)', a filter for 'Responsible division' set to 'All divisions', and a 'Filter: Include codelists' option. Below this is a 'Navigate by subject area' section with a list of categories: 'Labour market and earnings (1)', 'Banking and financial markets (1)', 'Population (6)', and 'Construction, housing and property (5)'. The 'Population (6)' category is expanded, showing a list of classifications such as 'Classification of age', 'Classification of country and citizenship in population statistics', 'Classification of families', 'Classification of households', 'Classification of marital status', and 'Classification of sex'. Each item has a 'Classification' link next to it.

On the right side, there is a 'CONTACT' section with an email icon, the name 'Anne Gro Hustoft', and contact information: 'E-mail: anne.gro.hustoft@ssb.no' and 'Phone: 40 90 25 52'.



How to use with R?

- Previously: send SQL query, read json
- Hackathon project
- Developed R functions into a package: search, get and apply
- Marketing



Main functions in `klassR`

`SearchKlass()` Search for classifications

`GetKlass()` Fetch a classification

`ApplyKlass()` Use a classification on a variable

SearchKlass()

```
SearchKlass("municipality")
```

	klass_name	klass_nr
	Classification of county municipality	127
	Classification of municipality groups	129
	Classification of municipalities	131
	Classification of population density	102
	Classification of urban districts	103
	Classification of economic regions	108
	Classification of centrality	128
	Classification of industrial link	107
Classification of tree species and assortment used by commercial roundwood removals		24
Classification of statistical tract and basic statistical unit		1
Classification of regional breakdown of house prices - House price index		30
	Classification of Organization form	35



GetKlass()

Get current occupation codes

```
GetKlass(klass = 7)
```

Get current occupation codes
with English names

```
GetKlass(klass = 7, language = "en")
```

Get occupation codes valid at
specific date

```
GetKlass(klass = 7, date = "2007-01-01")
```

Get occupation codes valid for
an interval

```
GetKlass(klass = 7, date = c("2015-01-01", "2019-01-01"))
```

```
GetKlass(klass = 7, correspond = TRUE, date = c("2015-01-01", "2019-01-01"))
```

Get occupation code changes



Correspondence tables

- Between different classifications and time points/versions

Classification of occupations

The standard is based on the International Standard Classification of Occupations - ISCO-08, prepared by the International Labour Organization (ILO). The standard classification of occupations is designed for official statistics of Norway, and occupations are classified by 4 digits.

Current version: (Valid from January 2011)

Classification of Occupations 08 (STYRK-08)

Codes About the version Changes All versions **Correspondences** Variants

[<< Back to correspondence overview](#)

2016-12 - Classification of Occupations 08 (STYRK-08)
Responsible: Bakke, Stine, division 312 - Seksjon for arbeidsmarked og lønnsstatistikk
Published in: Norwegian (bokmål)

Filter by code or name

2016-12	Classification of Occupations 08 (STYRK-08)
0111101 -	0310 - Armed forces occupations, other ranks
0111102 -	0310 - Armed forces occupations, other ranks
0111103 -	0310 - Armed forces occupations, other ranks
0111104 -	0310 - Armed forces occupations, other ranks
0112101 -	0210 - Non-commissioned armed forces officers
0112102 -	0110 - Commissioned armed forces officers
0112103 -	0110 - Commissioned armed forces officers
0112104 -	0110 - Commissioned armed forces officers



ApplyKlass()

- Grupping/recoding
- Names
- Converting

ID	NACE3D
1	551
2	853
3	869
4	721
5	852
6	61
7	910
8	280
9	62
10	471

ID	NACE3D	Hovednaering
1	551	C
2	853	E
3	869	E
4	721	D
5	852	E
6	61	A
7	910	F
8	280	B
9	62	A
10	471	C

ID	NACE3D	Hovednaering
1	551	Industri
2	853	Vannforsyning
3	869	Vannforsyning
4	721	Gass
5	852	Vannforsyning
6	61	Jordbruk
7	910	Bygge
8	280	Bergverksdrift
9	62	Jordbruk
10	471	Industri

ID	NACE3D	NACE_EU
1	551	1110
2	853	2015
3	869	2015
4	721	1500
5	852	2015
6	61	7015
7	910	3110
8	280	1115
9	62	7015
10	471	1110



ApplyKlass()

```
> occupation <- c("1213", "0110")
>
> ApplyKlass(occupation, 7)
[1] "Strategi- og planleggingsjefer"      "Offiserer fra fenrik og høyere grad"
>
> ApplyKlass(occupation, 7, language = "en")
[1] "Policy and planning managers"      "Commissioned armed forces officers"
```



Marketing

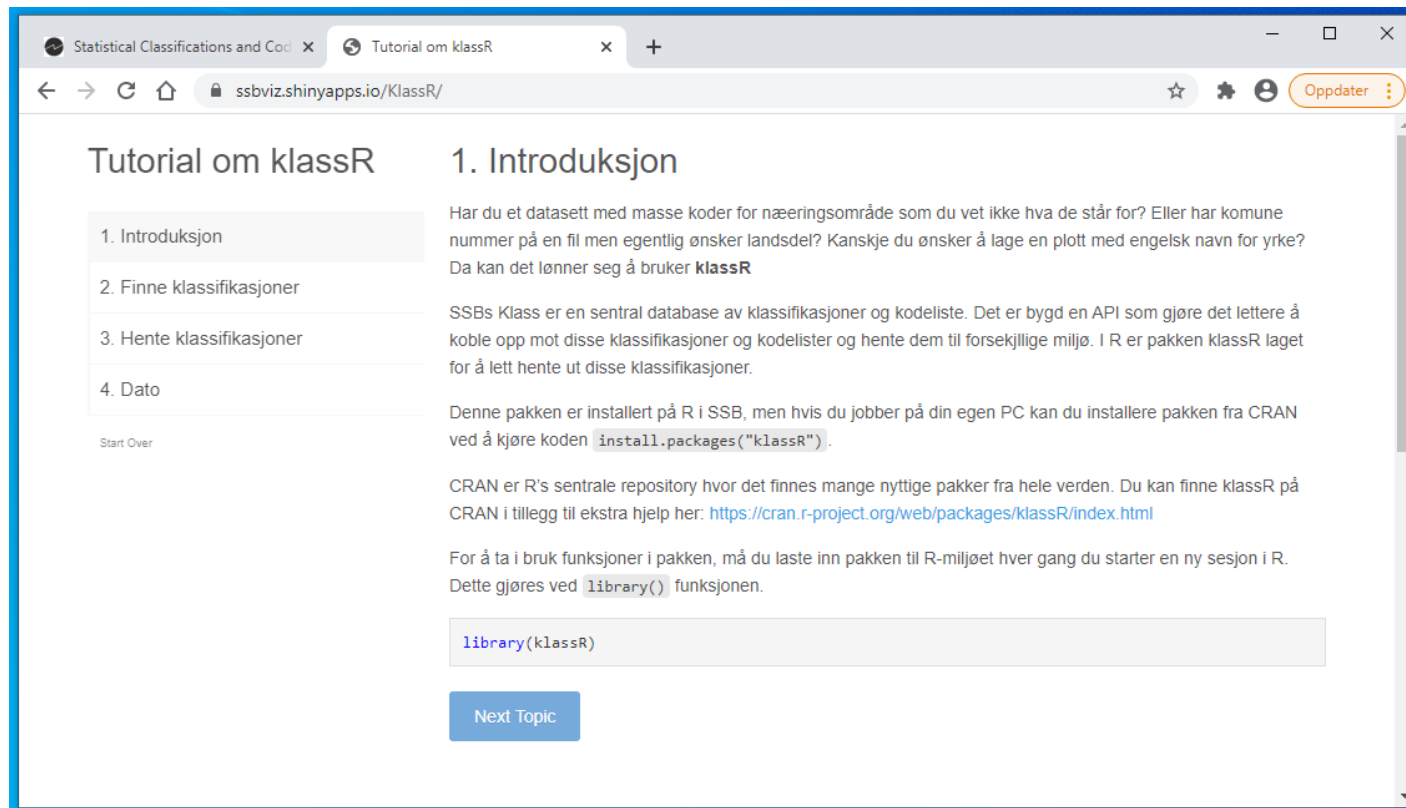
- CRAN/installed internally
- R-forum
- Coding club
- R introduction course
- Data cleaning course
- Internal groups/Yammer/Wiki

Formal decisions:

- Classifications must be collected from KLASS



Marketing



The screenshot shows a web browser window with two tabs: 'Statistical Classifications and Cod...' and 'Tutorial om klassR'. The address bar shows the URL 'ssbviz.shinyapps.io/KlassR/'. The page content includes a navigation menu on the left with items: '1. Introduksjon', '2. Finne klassifikasjoner', '3. Hente klassifikasjoner', and '4. Dato'. The main content area is titled '1. Introduksjon' and contains the following text:

Har du et datasett med masse koder for næringsområde som du vet ikke hva de står for? Eller har komune nummer på en fil men egentlig ønsker landsdel? Kanskje du ønsker å lage en plott med engelsk navn for yrke? Da kan det lønner seg å bruke **klassR**

SSBs Klass er en sentral database av klassifikasjoner og kodeliste. Det er bygd en API som gjøre det lettere å koble opp mot disse klassifikasjoner og kodelister og hente dem til forskjellige miljø. I R er pakken klassR laget for å lett hente ut disse klassifikasjoner.

Denne pakken er installert på R i SSB, men hvis du jobber på din egen PC kan du installere pakken fra CRAN ved å kjøre koden `install.packages("klassR")`.

CRAN er R's sentrale repository hvor det finnes mange nyttige pakker fra hele verden. Du kan finne klassR på CRAN i tillegg til ekstra hjelp her: <https://cran.r-project.org/web/packages/klassR/index.html>

For å ta i bruk funksjoner i pakken, må du laste inn pakken til R-miljøet hver gang du starter en ny sesjon i R. Dette gjøres ved `library()` funksjonen.

```
library(klassR)
```

Next Topic

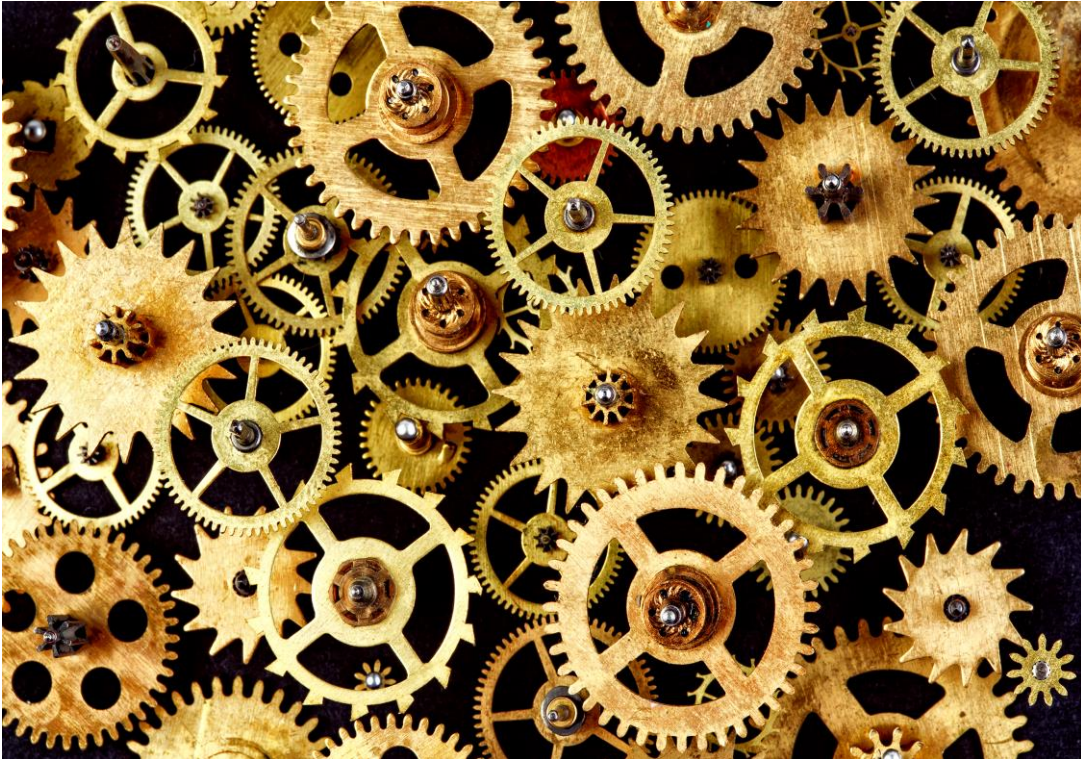
learnr and shiny

<https://ssbviz.shinyapps.io/KlassR/>



Statistisk sentralbyrå
Statistics Norway

Summary



- Standards important for data integration
- Package has increased access to standards
- Requires repeated encouragement for use