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

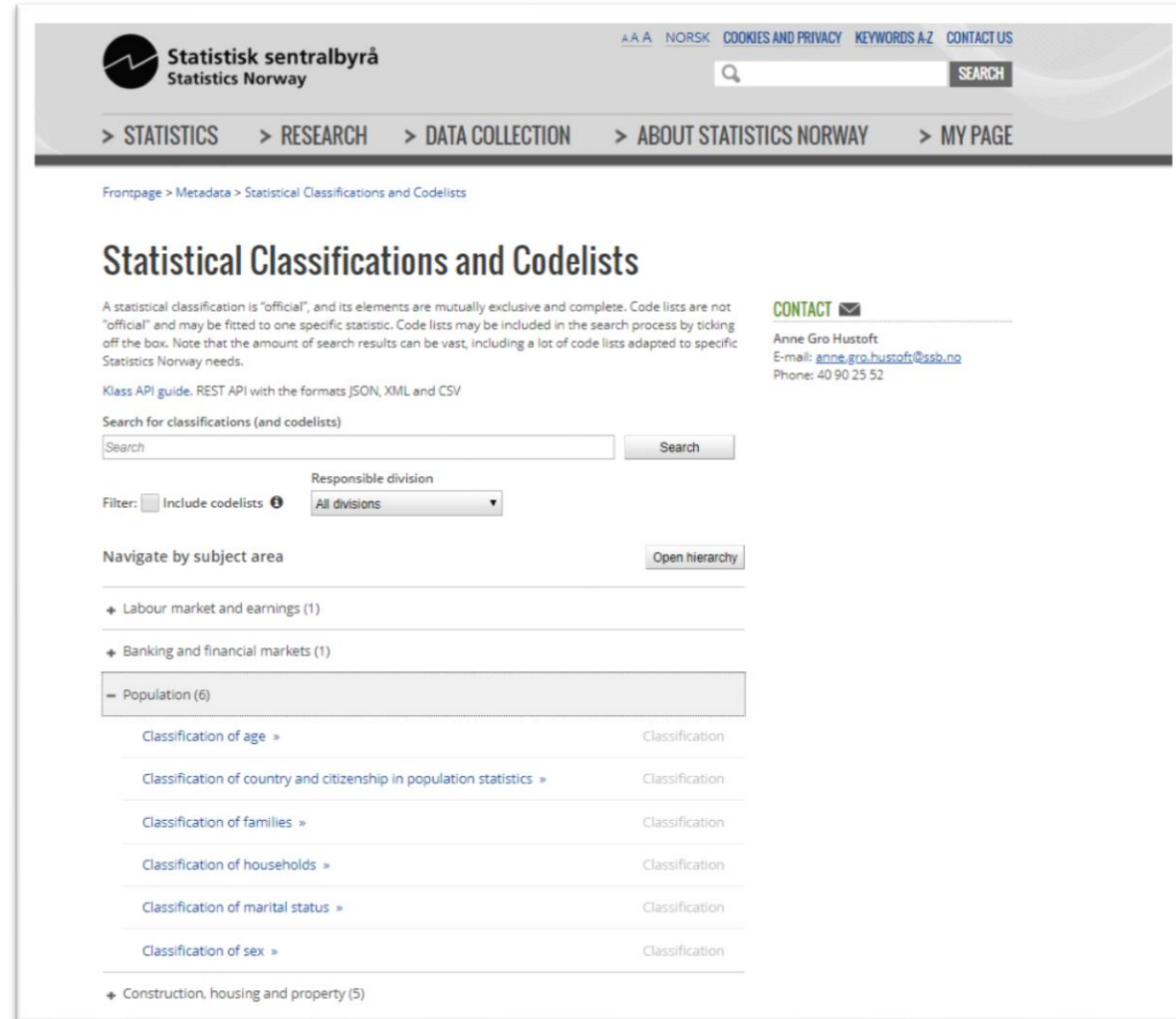'Men' = Male
'Women' = Female

'm' = Male
'f' = Female

'Mann' = Male
'Dame' = Female

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

Statistisk sentralbyrå
Statistics Norway

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

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

Statistisk sentralbyrå
Statistics Norway

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

Statistisk sentralbyrå
Statistics Norway

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

Statistisk sentralbyrå
Statistics Norway

Correspondence tables

- Between different classifications
  and time points/versions

ApplyKlass()

- Grupping/recoding

- Names

- Converting

ApplyKlass()

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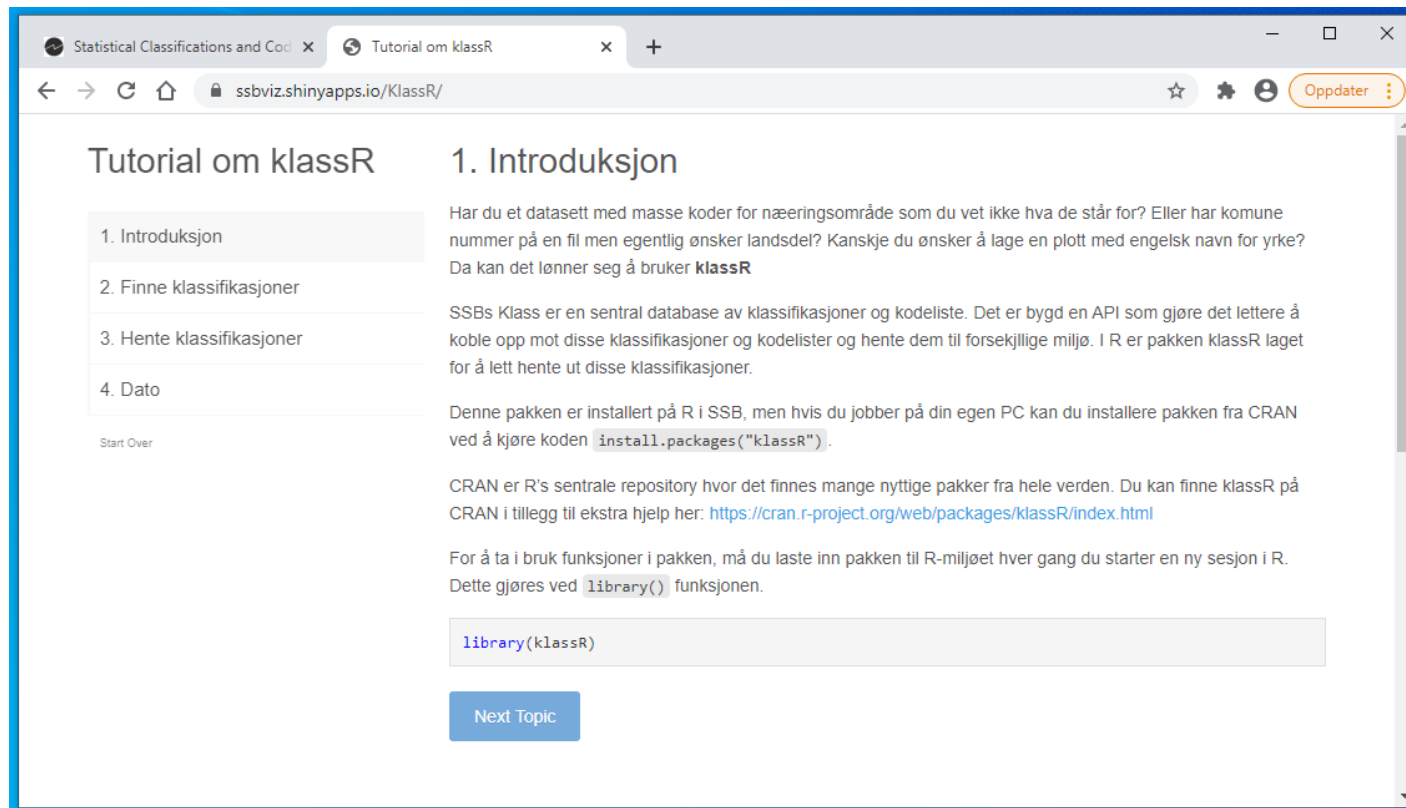
Marketing

- 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

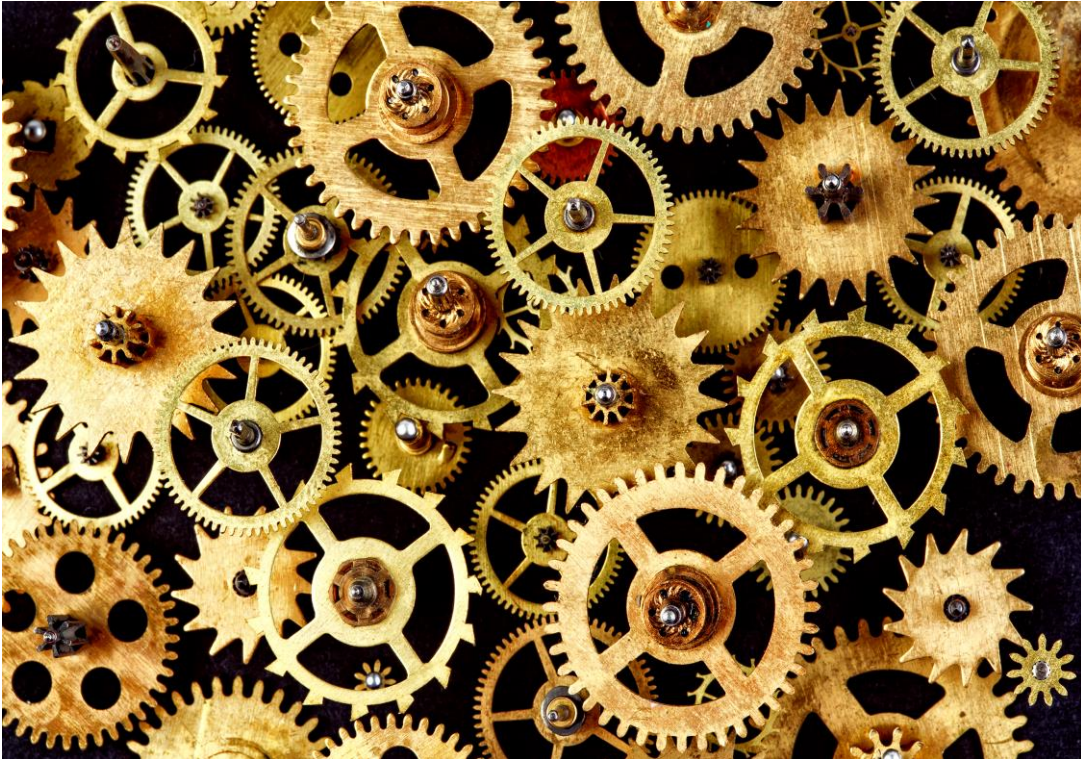# Marketing

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

Statistisk sentralbyrå
Statistics Norway