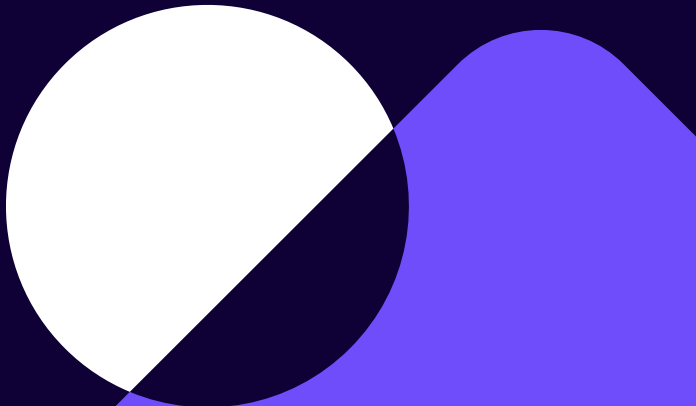


microdata.no

microdata.no provides instant, online access to large amounts of detailed and linkable microdata without any form of application.

microdata.no is open for employees and students at universities and colleges, approved research institutions, ministries and directorates.



Ørnulf Risnes
Product manager microdata.no
Sikt - Norwegian Agency for Shared Services in
Education and Research
<ornulf.risnes@sikt.no>

What is microdata.no?

Developed and operated by Sikt and Statistics Norway (SSB)

Live since spring 2018

A service for linkage and analysis of register data and other forms of microdata

Complete, non-anonymized data about the complete population

Access from day 1 (no paperwork) to all available variables (currently 405)

Large flexibility in linkage, data management and analyses

Browser based and location independent

Built-in privacy controls to keep data/results anonymous on the end-user side



Microdata.no won the Norwegian Data Protection Agency's (Datatilsynet) annual award for «Practical built-in privacy».

Article from Datatilsynet's website

<https://www.datatilsynet.no/aktuelt/aktuelle-nyheter-20192/microdata.no-vant-pris/>

Available data in microdata.no

Full population since 1964 (approx. 10 million)

Data in full detail, not anonymized

Variables about:

- Demography (sex, marital status, area of residence)
- Income
- Education
- Employment (jobs, etc)
- Welfare benefits
- Social background
- Sickleave
- Election participation
- Immigration

**ABSTRACT**

Early work experience is found to be an influential factor in young people's transitions from school to work. Still, we know little about whether early work experience can protect vulnerable young people from subsequent exclusion from labour and education in early adulthood. Our objective is therefore to examine how early work experience in adolescence influences the risk of being NEET (Not in Education, Employment or Training), and whether this relationship is stronger for early school leavers and young disabled people. We utilise Norwegian register data covering the entire 1985-birth cohort, followed from age 16 to 29 (n ~ 50 000). Linear probability models are used to estimate the NEET risk at age 25 and age 29. **The findings reveal that early work experience is related to a lower NEET risk for everyone, but more strongly for the young people with disabilities or early school leaving.** The findings support early work experience as a potentially important protective factor against subsequent NEET status, particularly among vulnerable young people.

- 25 (2): 323–334. doi:10.1007/s10926-014-9541-6.
- Widding-Havnerås, Tarjei. 2016. "Unge Voksne Som Verken Er I Arbeid Eller Utdanning: En Registerbasert Studie, 1993–2009. [Young People Not in Work or Education: A Registry Data Investigation, 1993–2009]." *Søkelys på arbeidslivet* 33 (4): 360–378. doi:10.18261/1.1504-7989-2016-04-05.
- Yates, Scott, and Malcolm Payne. 2006. "Not so NEET? A Critique of the Use of 'NEET' in Setting Targets for Interventions with Young People." *Journal of Youth Studies* 9 (3): 329–344. doi:10.1080/13676260600805671.

518  J. G. BALLO ET AL.

Appendix

Script used for analysis in www.microdata.no

```
// Can adolescent work experience protect vulnerable youth?  
// A population wide longitudinal study of young adults not in education, employment or training (NEET)
```

```
// Kobler til databank  
require no.ssb.fdb:2 as ds
```

```
create-dataset NEET  
import ds/BEFOLKNING_FOEDSELS_AAR_MND as alder15  
replace alder15 = 2015 - (int(alder15/100))  
keep if alder15 == 30  
import ds/BEFOLKNING_REGSTAT 2014-01-01 as status14  
import ds/BEFOLKNING_REGSTAT 2013-01-01 as status13  
import ds/BEFOLKNING_REGSTAT 2012-01-01 as status12
```



Reproducibility

Script from
paper's
appendix

The screenshot displays a Jupyter Notebook interface with a dark theme. The left pane shows Stata code for creating a dataset named 'NEET' from various data sources. The right pane shows the execution output, including messages about dataset creation, variable imports, and filtering. A third pane on the right shows a list of variables in the dataset.

```
6 require no.ssb.fdb:2 as ds
7
8 create-dataset NEET
9 import ds/BEFOLKNING_FOESELS_AAR_MND as alder15
10
11 replace alder15 = 2015 - int(alder15/100)
12 keep if alder15 == 30
13
14 import ds/BEFOLKNING_REGSTAT 2014-01-01 as status14
15 import ds/BEFOLKNING_REGSTAT 2013-01-01 as status13
16 import ds/BEFOLKNING_REGSTAT 2012-01-01 as status12
17 import ds/BEFOLKNING_REGSTAT 2011-01-01 as status11
18 import ds/BEFOLKNING_REGSTAT 2010-01-01 as status10
19 import ds/BEFOLKNING_REGSTAT 2009-01-01 as status09
20 import ds/BEFOLKNING_REGSTAT 2008-01-01 as status08
21 import ds/BEFOLKNING_REGSTAT 2007-01-01 as status07
22 import ds/BEFOLKNING_REGSTAT 2006-01-01 as status06
23 import ds/BEFOLKNING_REGSTAT 2005-01-01 as status05
24 import ds/BEFOLKNING_REGSTAT 2004-01-01 as status04
25 import ds/BEFOLKNING_REGSTAT 2003-01-01 as status03
26 import ds/BEFOLKNING_REGSTAT 2002-01-01 as status02
27 import ds/BEFOLKNING_REGSTAT 2001-01-01 as status01
28
29 keep if status14 == '1'
30 keep if status13 == '1'
31 keep if status12 == '1'
32 keep if status11 == '1'
33 keep if status10 == '1'
34 keep if status09 == '1'
35 keep if status08 == '1'
```

Execution output:

```
» require no.ssb.fdb:2 as ds
   Opprettet en kobling fra no.ssb.Fdb:2 til ds

» create-dataset NEET
   Et tomt dataset, NEET ble opprettet og valgt

NEET» import ds/BEFOLKNING_FOESELS_AAR_MND as alder15
   Importerte alder15 til NEET med 9 903 454 verdier

NEET» replace alder15 = 2015 - int(alder15/100)
   Byttet ut verdier i alder15
   Antall enheter: 9 903 454

NEET» keep if alder15 == 30
   9 788 810 enheter ble fjernet fra datasettet.

NEET» import ds/BEFOLKNING_REGSTAT 2014-01-01 as status14
   Importerte status14 til NEET med 114 644 verdier, hvorav 36 974 missingverdier

NEET» import ds/BEFOLKNING_REGSTAT 2013-01-01 as status13
   Importerte status13 til NEET med 114 644 verdier, hvorav 39 857 missingverdier

NEET» import ds/BEFOLKNING_REGSTAT 2012-01-01 as status12
   Importerte status12 til NEET med 114 644 verdier, hvorav 42 878 missingverdier

NEET» import ds/BEFOLKNING_REGSTAT 2011-01-01 as status11
   Importerte status11 til NEET med 114 644 verdier, hvorav 46 019 missingverdier

NEET» import ds/BEFOLKNING_REGSTAT 2010-01-01 as status10
   Importerte status10 til NEET med 114 644 verdier, hvorav 49 011 missingverdier

NEET» import ds/BEFOLKNING_REGSTAT 2009-01-01 as status09
   Importerte status09 til NEET med 114 644 verdier, hvorav 51 227 missingverdier
```

Datasett

✓ NEET
30 variabler, 50 506 enheter

- PERSONID_1
- alder15
- lønn16
- lønn17
- lønn18
- lønn25
- lønn29
- student25
- student29
- neet25
- neet29
- gs16
- hs16
- barn25
- barn29
- gift25
- gift29
- uhelse16
- sosbak
- tidlarb
- utd10
- dropout
- innvandr
- invbakgrunn
- landbakgrunn
- ikkevest
- kvinne
- ufør29juli
- ufør25juli
- bosted16

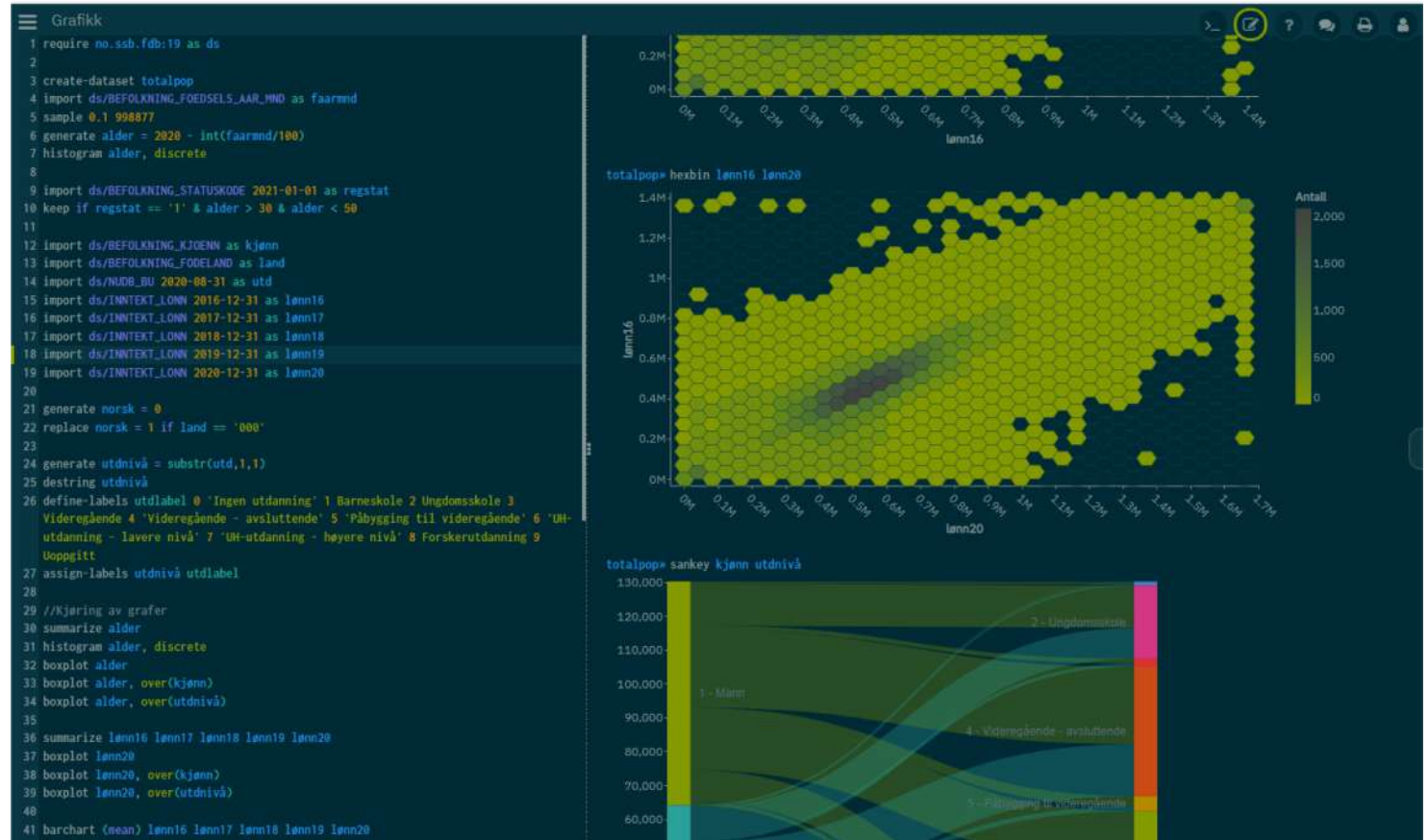
Send til kommandolinjen Kjør



Live demo

Income distribution by sex
at 2010 and 2021

In 3 minutes



Framework for built-in privacy –

«The 5 Safes» (Felix Ritchie -https://en.wikipedia.org/wiki/Five_safes)

Safe purposes	Is this use of the data appropriate?
Safe people	Can the users be trusted to use it in an appropriate manner? Do we know who they are? Can we monitor their behaviour in the system?
Safe settings	Does the access facility limit unauthorised use?
Safe data	Is there a disclosure risk in the data itself? (Yes)
Safe outputs	Are the statistical results non-disclosive?

Microdata.no and relative ranking of the 5 Safes

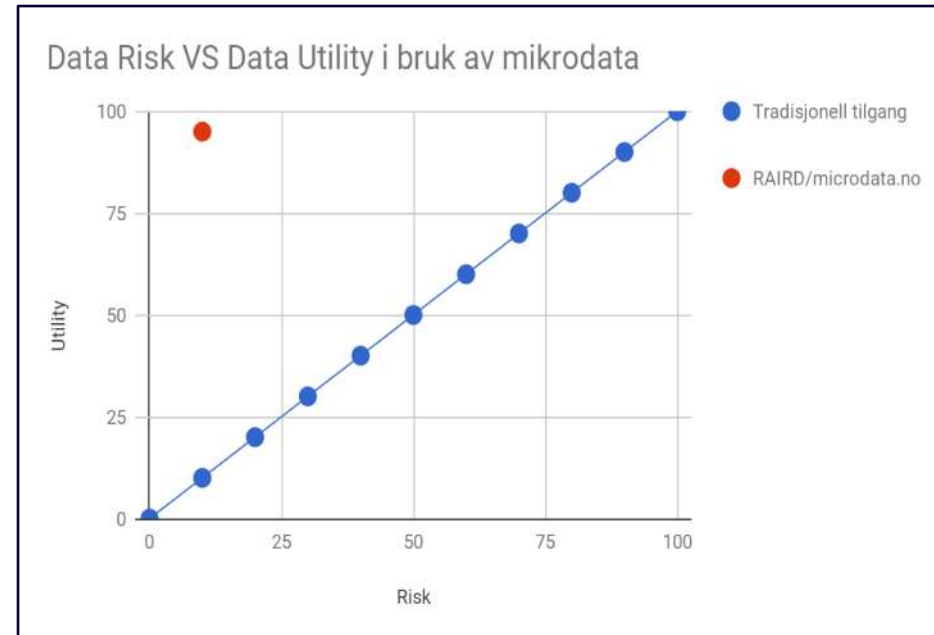
Safe settings

Safe outputs

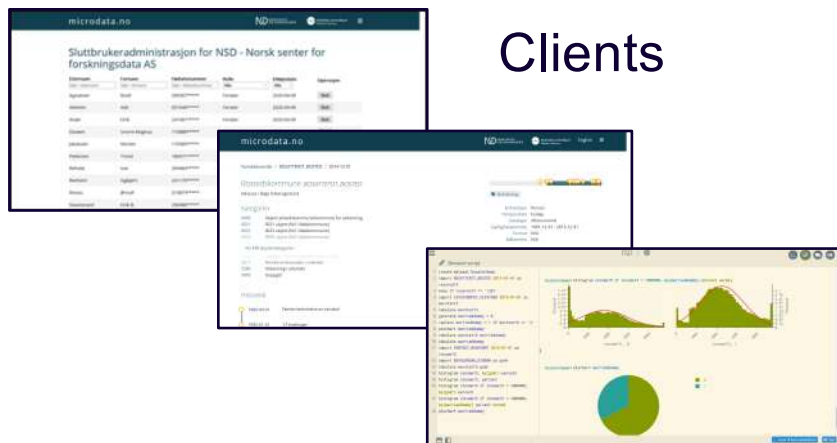
Safe people

Safe purposes

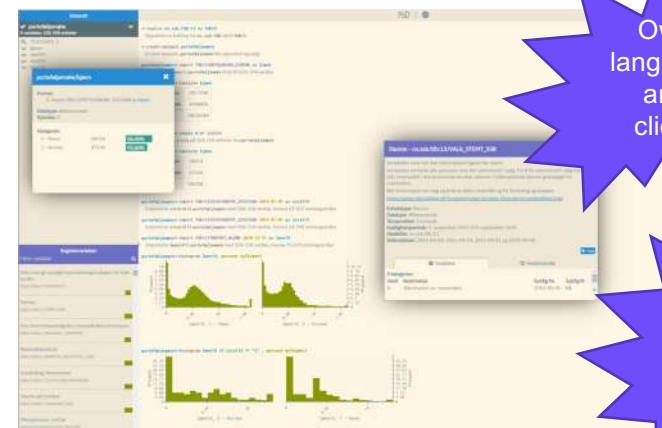
Safe data



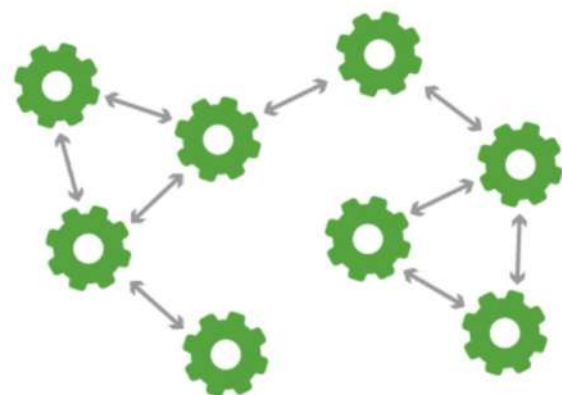
Safe settings – controlled execution



Clients



APIs in front of secured data settings

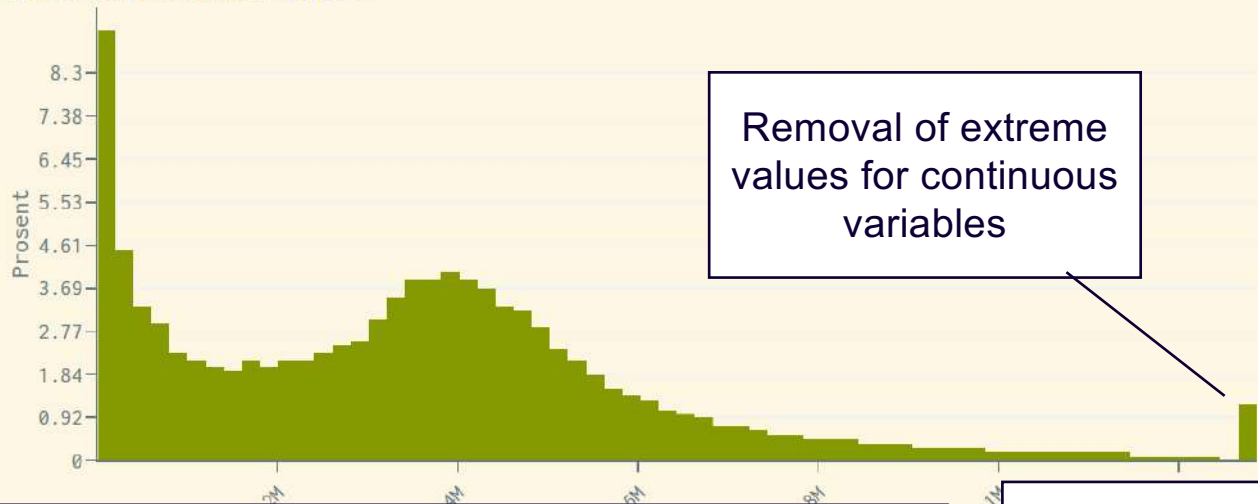


Services, data and metadata



Safe outputs – selected means

innt10» histogram lønn10, percent



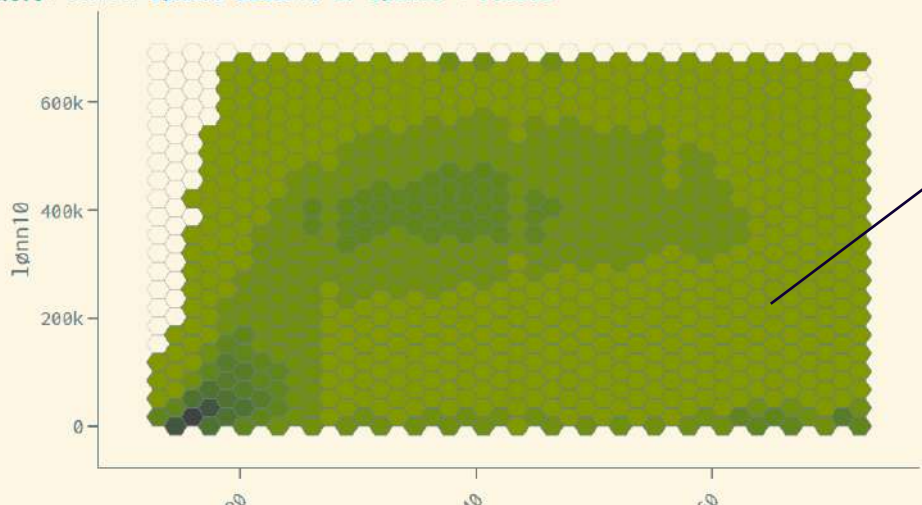
Removal of extreme values for continuous variables

te sivs10 kjonn

	kjonn		Total
	1 - Mann	2 - Kvinne	
ent	83	67	158
ft	1307038	1136146	
t	876008	856116	
kemann	49204	200874	
lt	160012	203571	363579
	32181	33120	65299
	1468	1099	2556
7 - Separert partner	181	145	316
8 - Skilt partner	436	383	814
9 - Gjenlevende partner	66	16	80
Total	2426658	2431519	4858185

Noise added to all frequencies

innt10» hexbin lønn10 alder10 if lønn10 < 700000



Coarse-grained scatterplots (hexbin)