

# Streamlining Web Paradata Retrieval and Analysis: A Semi-Automated Approach for Enhanced Efficiency and Consistency

Gustaf Andersson, Methodologist, Department of Social Statistics and Analysis

Emma Stavås, Methodologist, Department of Data Management

Statistics Sweden

# Contents

- Introduction and background
- The report
- User cases and final reflections

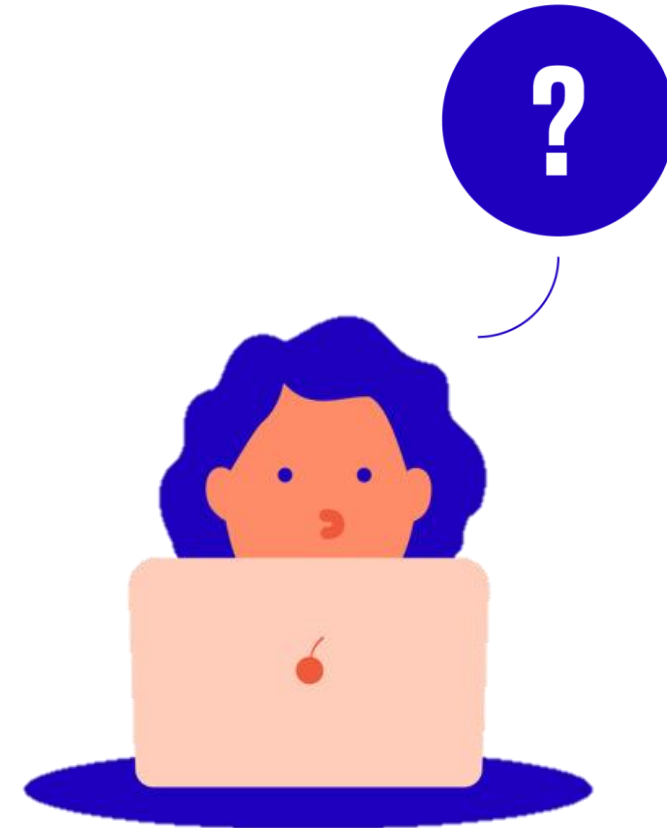


## What is Web Paradata?

- Paradata: Information about the process of collecting survey data
- For Web Surveys: The respondent's survey-taking behaviour

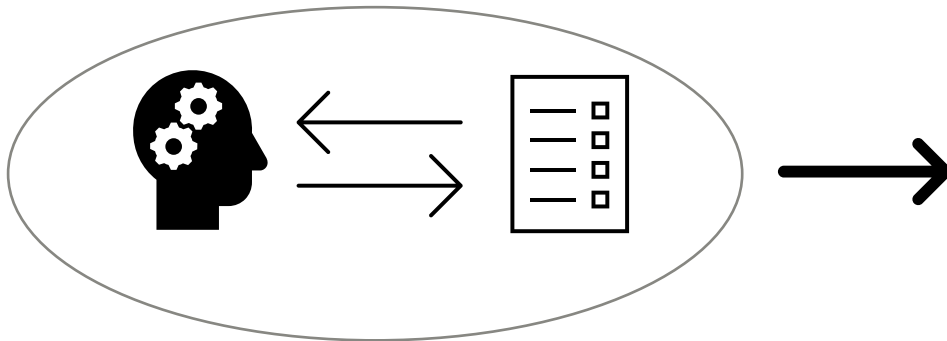
# Motivations Behind the Project

- Demand for paradata
  - Shift in primary data collection method
  - Research area under development
- Demand for a structured way of working
  - Retrieving and processing paradata perceived as difficult
  - Ensure consistent and uniformly defined variables



# How are input data generated and stored?

- Navigation data
- Device type
- ...



The same set-up across  
all web surveys!

Answer Log

Question ID	Answer	Unique ID
B4	1, 4	FU-14-FJVCI

Event Log

Respondent ID	Event Code	Unique ID
A	"Answer"	FU-14-FJVCI
B	"Control"	GH-29-VIGHE

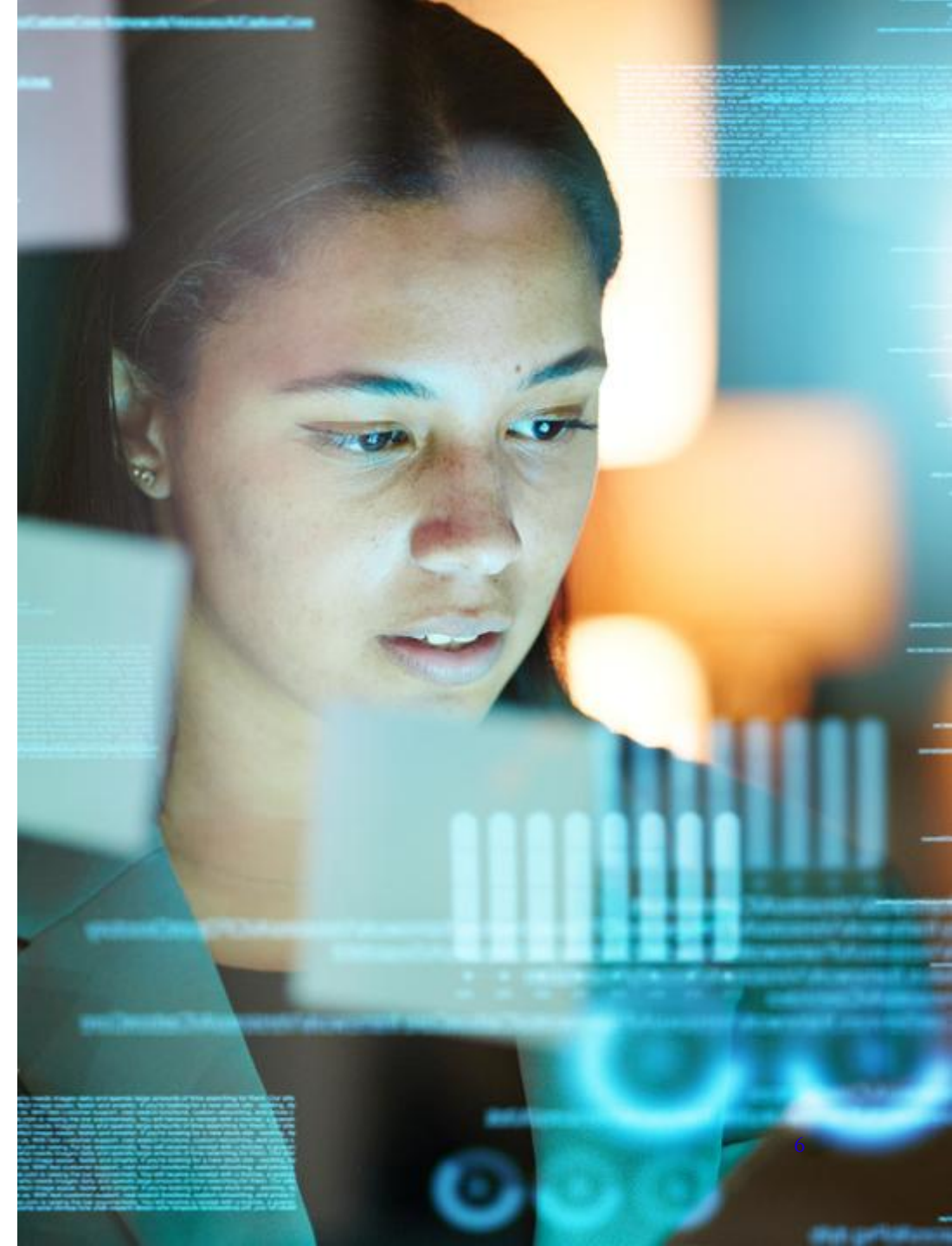
Control Log

Control Code	Control Message	Unique ID
B4contr	"Multiple selections not allowed"	GH-29-VIGHE

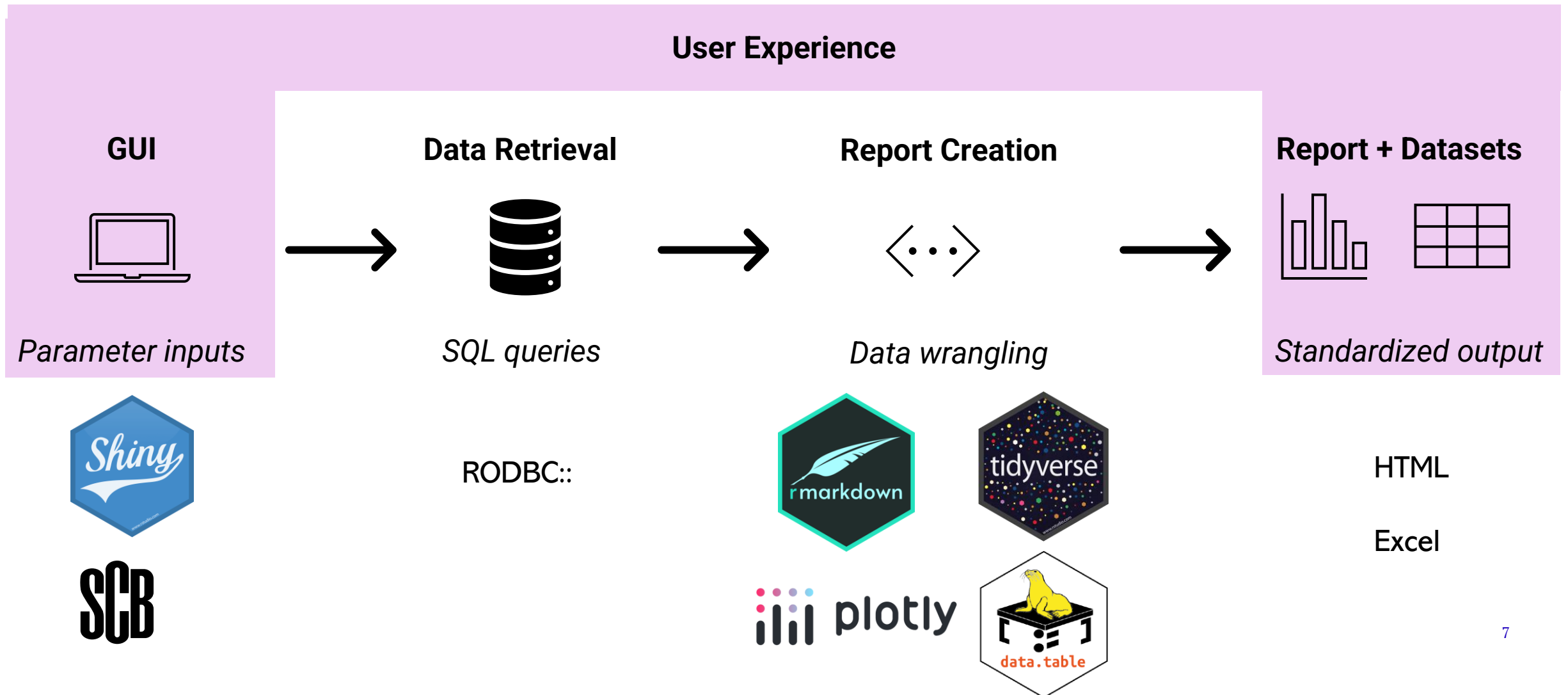


# Development philosophy

- **Keep it simple for users**
  - Point-and-click GUI
  - No coding required
  - Support documentation
- **Flexible output**
  - Dynamic plots/tables
  - Processed datasets



# How is a standard report generated?



# ... And in practice?

Copies of Input\_File.R and Background\_File.Rmd

Input\_File.R → “Run App” + parameters

Parameters → Background\_File.Rmd

Background\_File.Rmd → report and datasets

Datasets → possibly add other non-paradata  
background variables for grouped analyses



# Let's have a look

Choose from surveys you are authorised to access

Choose survey round

Date, start

Date, end

Select from options using checkboxes

Generate a report



## Standardiserad paradatarapport från SIV

Fyll i nedanstående uppgifter för att skapa din rapport.

Om du saknar din undersökning i rullistan nedan behöver du söka behörighet via Beställarportalen. Under "Ansök om behörighet till grupp eller mapp" fylls fältet "Grupp eller mapp" i med följande:

där {schemanamn} ersätts med undersökningens kortnamn.

Undersökning:

ULF\_SILC

Undersökningsperiod:

ULFSILC2021 test 02

Tidigaste datum som ska beaktas:

2021-04-22

Senaste datum som ska beaktas:

2021-10-12

☐ Standardrapporten ska redovisa grupperade resultat

☐ Dataset ska genereras tillsammans med standardrapporten

Skapa rapport

1. Processdataberättelse SIV

2. Teknisk utrustning

3. Genomströmning

3.1 Inloggningar och inskick

3.2 Tidpunkter för aktivitet

3.3 Aktiv svarstid

3.4 Rörelse genom blanketten

4. Användning av stödfunktioner

5. Uppgiftslämnargranskning



Light/Dark Mode

Dynamic chart: use  
cursor for more  
information, zoom in  
and out

"Clickable" Table of  
Contents

Händelser i blanketten registrerades mellan 2024-04-02 och 2024-04-28, ett intervall på 26 dagar. Efter 2 dagar hade 25 procent av händelserna som registrerats under tidsperioden skett. När 7 dagar passerat hade 50 procent av händelserna skett, och efter 15 dagar hade 75 procent av händelserna registrerats i händelseloggen. Flest händelser under ett enskilt dygn registrerades 2024-04-04 (16.17 procent av det totala antalet händelser).

Av 566 unika uppgiftslämnare var det flest unika uppgiftslämnare som använde blanketten 2024-04-04 (13.78 procent), följt av 2024-04-03 (12.54 procent) och 2024-04-08 (10.07 procent).

I diagrammen och tabellerna nedan presenteras fördelningarna av antalet händelser och aktiva unika uppgiftslämnare per datum. Ett exakt datum går att se genom att hålla muspekaren över en stapel i diagrammet.

Diagram 3.2.1.1. Händelser per datum

Tabell 3.2.1.1. Händelser per datum

Diagram 3.2.1.2. Unika uppgiftslämnare per datum

Tabell 3.2.1.2. Unika uppgiftslämnare per datum

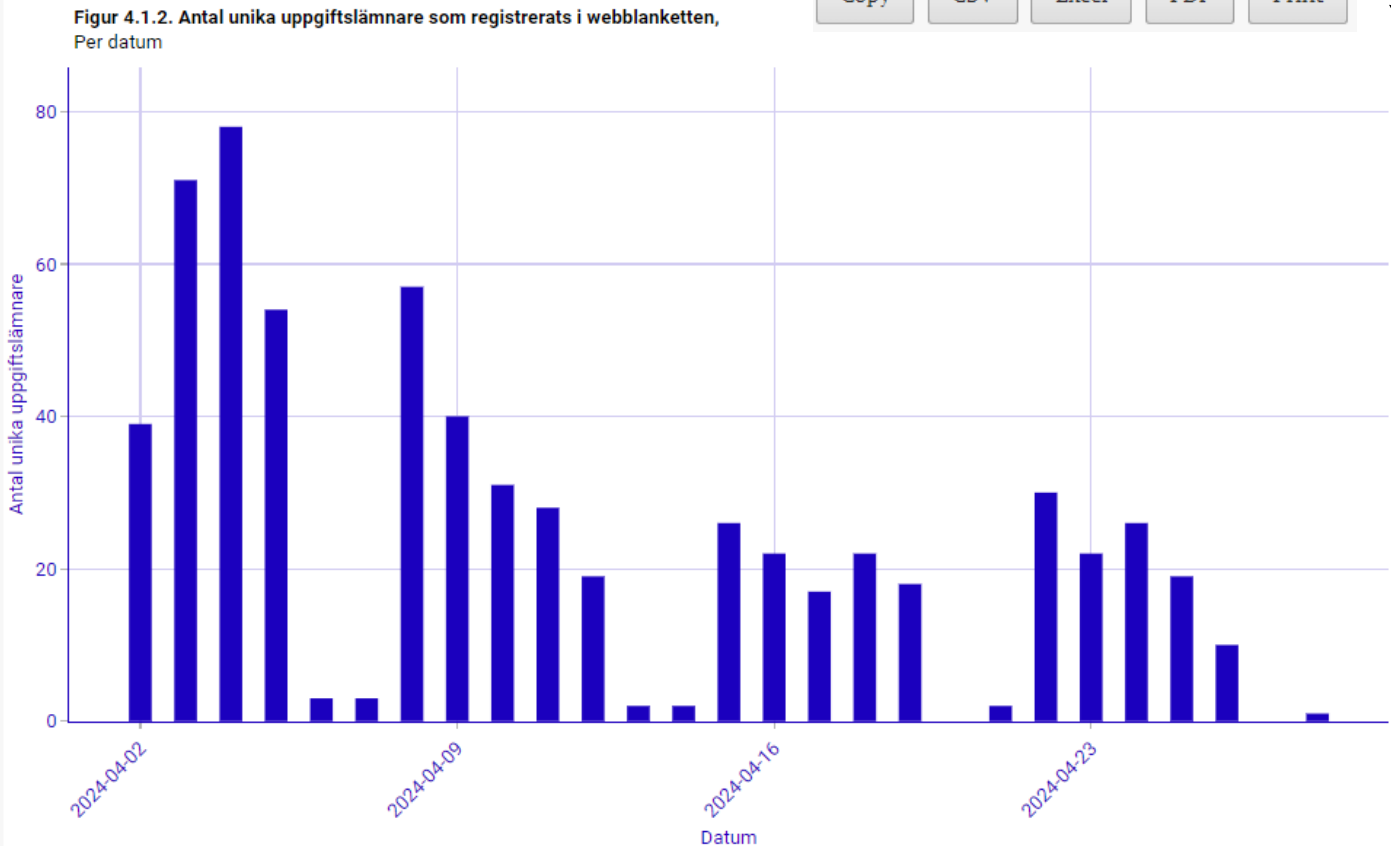
Copy

CSV

Excel

PDF

Print



Text based on input-  
data

Choose which chart  
or table to display

Copy, or export  
charts and tables as  
CSV, Excel, PDF or  
image

# User cases

Main use: follow up of data collection for a specific survey

Paradata is a new data source for many – exploratory analyses

R is new to some – still possible to retrieve standard reports

Case: SE-SILC - time spent on certain pages and use of help info – revise help info

Case: Citizen Survey - response times as Quality Assurance indicator – grouped analyses based on income and nationality

Measurement follow up: answer changes and quality control activations for certain questions

# Some final reflections

## Success Factors:

- Report integrated into Statistics Sweden's general production process
- Quick support
- Flexible coding with R
- Work based on demand from colleagues → interest in output

## Current Work:

- Paradata is new and a "secondary source" → building capacity and knowledge
- Business surveys
- Feature to add grouped analyses

# Thank you for your attention!

**Gustaf Andersson**

Statistics Sweden

[gustaf.andersson@scb.se](mailto:gustaf.andersson@scb.se)

**Emma Stavås**

Statistics Sweden

[emma.stavas@scb.se](mailto:emma.stavas@scb.se)