



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN



repreport

# Facilitating reproducible research in Stata

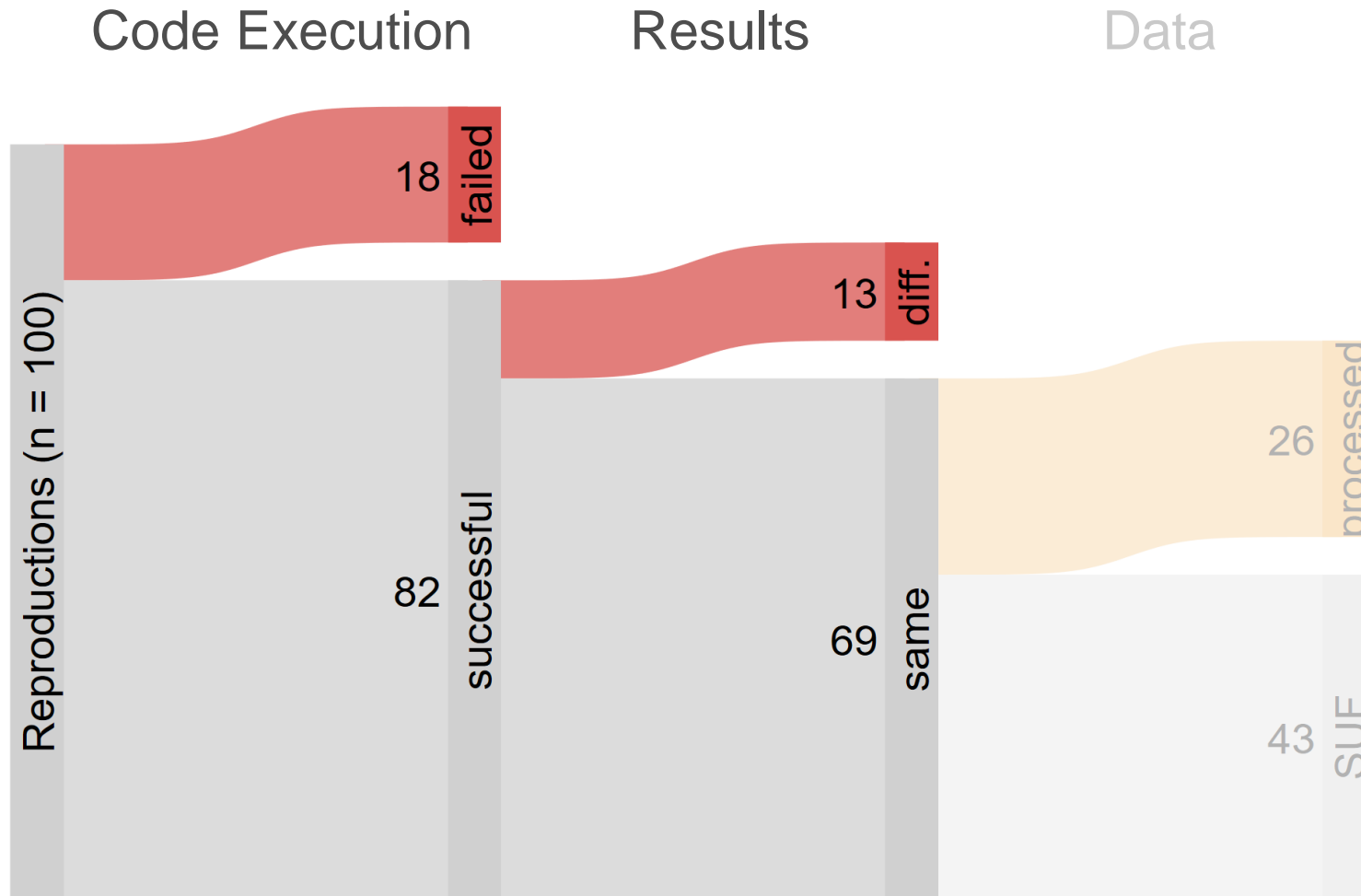
Daniel Krähmer

Department of Sociology | LMU Munich



Funding by the DFG through the Priority Program META-REP (Project 464507200) is gratefully acknowledged.

# “Houston, we have a problem“



**Alarming**  
(for science)



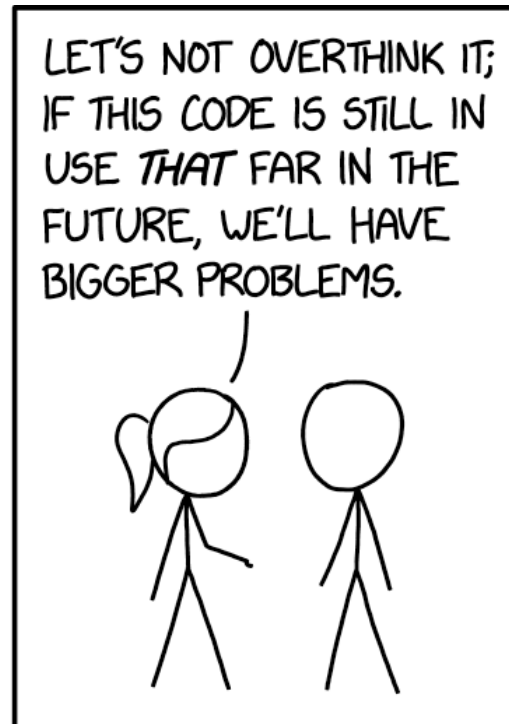
**Baffling**  
(for programmers)

Own reproducibility audit; *sankey* (A. Naqvi)

# “Houston, we have a problem“

## Code Execution

- Ados
- File paths
- File management (!)



## Results

- Randomness
- Runtime
- Unknown (?)

XKCD (modified)

# Are authors deliberately obstructing reproducibility?

**NO!**

“I would like to help, *but...*“



“I can't find the code“



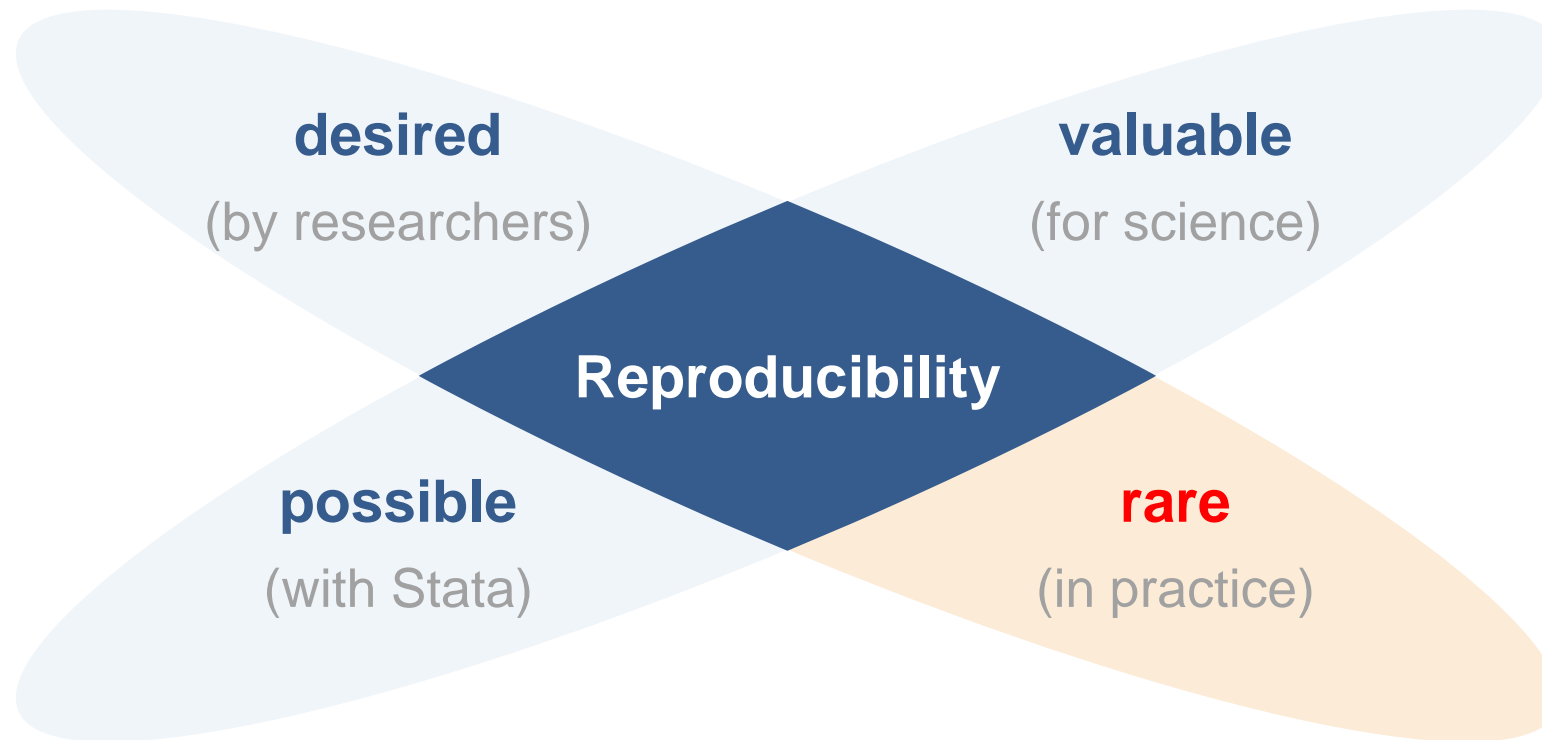
“I don't understand my code“



“I don't have the time“

→ It's about **skills**, not *willingness*

# The Status Quo



→ **Minimize costs** to work reproducibly in Stata

# Precursors to reprepare

## Log-files and user-written commands:

[callsado](#)

D. Klein

[dependencies](#)

D. Goldemberg

[iedorep](#)

B. Daniels

[ietoolkit](#)

DIME Analytics

[installpkg](#)

K. Du

[project](#)

R. Picard

[repkit](#)

DIME Analytics

[require](#)

S. Correia, M. Seay

[setroot](#)

S. Correia, M. Seay

[stata\\_linter](#)

DIME Analytics

[stata\\_reproducible](#)

B. Quistorff

[statapackagesearch](#)

L. Reiner

[texdoc](#)

B. Jann

[usepackage](#)

E. Booth

[zippkg](#)

D. Bela

...



# Using rereport: Frontend

\* 1. Initiate REPREPORT

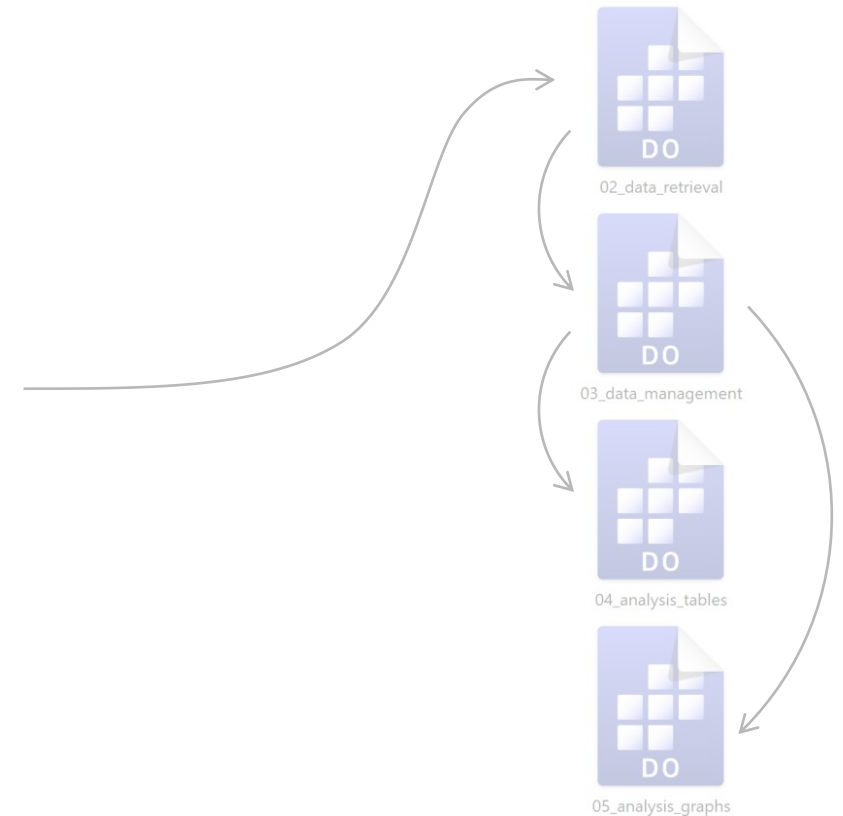
```
rereport_open
```

\* 2. Perform analysis (do-file based or interactively)

```
include 01_master.do
```

\* 3. Close REPREPORT

```
rereport_close using ../..
```



# Understanding rereport: Backend

log entire analysis

```

state
-----
name: demolog
log: C:\Users\m38fe\LRB\Sync\Share\PhD\13_Projekte\ado_repreport\example_ana
> lysia\do\..\REPREPORT.amcl
log type: smcl
opened on: 16 May 2024, 15:45:07

1. include 01_master.do
2. ***** Demo do-file to test the program REPREPORT *****
3.
4. /* Note: All do-files in this directory only serve the purpose of beta testing
   > the program REPREPORT. They carry no substantial meaning and are, in some cases,
   > deliberately quirky to test the program's boundaries.*/
5.
6. * 01 Master
7. * - Includes the setup and invokes all other do-files
8.
9. * Header
10. cap log close
11. macro drop _all
12. version 17
13. set more off
14. set varabbrev off
15.
16. * Subroutines
17. include 02_data_retrieval.do
18. ***** Demo do-file to test the program REPREPORT *****

```

import as data  
(smcl)

```

. **# Dataset 1: Auto
. use ../data/clean/auto_modified.dta, clear
(1978 automobile data)

. describe
Contains data from ../data/clean/auto_modified.dta
Observations:      50      1978 automobile data
Variables:         6      16 May 2024 15:53
                    (_dta has notes)
-----+-----
Variable      Storage   Display    Value
name          type      format    label      Variable label
-----+-----
make          str18    %-18s     Make and model
price         int       %8.0gc   Price
rep78         int       %8.0gc   Repair record 1978
weight        int       %8.0gc   Weight (lbs.)
length        int       %8.0gc   Length (in.)

```

reconstruct "clean"  
command history

```

esttab
tab rep78
local mycommand fre
`mycommand' rep78

```

substitute  
macros

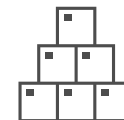
```

esttab
tab rep78
local mycommand fre
fre rep78

```

**Elicit reproducibility features**  
(ados, datasets, randomness, etc.)

generate output





# Output I: Reproducibility Report

## “Technical ReadMe“:

- General Information
- System Details
- Setup
- User-written commands  
(+ versions)

```

*****
Reproduction Report (created using ado 'REPREPORT')
*****

*** General Information
Date:           6 Jun 2024
Start:          10:35:46 (6 Jun 2024)
Stop:           10:35:52 (6 Jun 2024)
Runtime:        00:00:06

*** System Information
Stata Version:  17
Version (set):  17
Update level:   13 Feb 2024
Edition:        MP
OS:             Windows
Machine:        PC (64-bit x86-64)

*** Setup Information
Var. Abbrev.:   off
Niceness:       5

*** User-written commands
-----
Command          Version
-----
colorpalette     1.2.3
esttab           2.1.1
fre              1.2.5
  
```



# Output I: Reproducibility Report

## “Technical ReadMe“:

- General Information
- System Details
- Setup
- User-written commands  
(+ versions)
- Input files  
(+ location and signatures, if applicable)
- Subroutines  
(+ location)
- Randomness

```

*** User-written commands
-----
Command          Version
-----
colorpalette     1.2.3
esttab           2.1.1
fre              1.2.5

*** Input Files (i.e. datasets)
-----
File              Path              Data Signature (if applicable)
-----
C64284D_ALL_LATEST.csv  C:/Users/ado_repreport/data/raw
mydata.dta          C:/Users/ado_repreport/data/raw  69:12(71728):2063612797:516689108

*** Subroutines (i.e. other do-files)
-----
File              Original path
-----
01_master.do      C:/Users/ado_repreport/do
02_data_retrieval.do  C:/Users/ado_repreport/do
03_data_management.do  C:/Users/ado_repreport/do
04_analysis_tables.do  C:/Users/ado_repreport/do
05_analysis_graphs.do  C:/Users/ado_repreport/do

*** Randomness
Randomness:      Yes, randomness detected (change in rngstate)
RNG:             default
Seed set?        Yes
Seed value:      1234

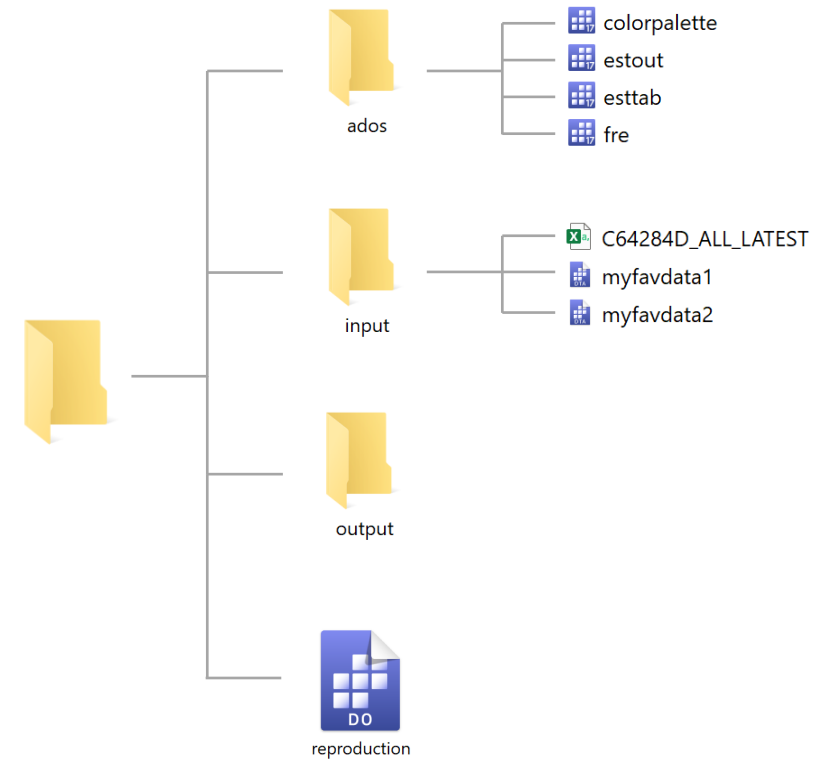
```



## Output II: Reproduction Materials

### File collection:

- 3 Directories:
  - Ados  
(all ados in user's current version)
  - Input  
(all input files, e.g. datasets, csv, etc.)
  - Output  
(storage for output)
- 1 Do-file



→ **Standardization increases chances for „push-button“ reproducibility**



## Beta-Testing (GitHub)

<https://tinyurl.com/repreport-beta>



**Please reach out to report bugs or suggest improvements!**

# Conclusion

repreport facilitates reproducibility at no extra cost.

## Who wins?

computational expertise

social scientists  
(in our sample)

most of  
you

some of you

...



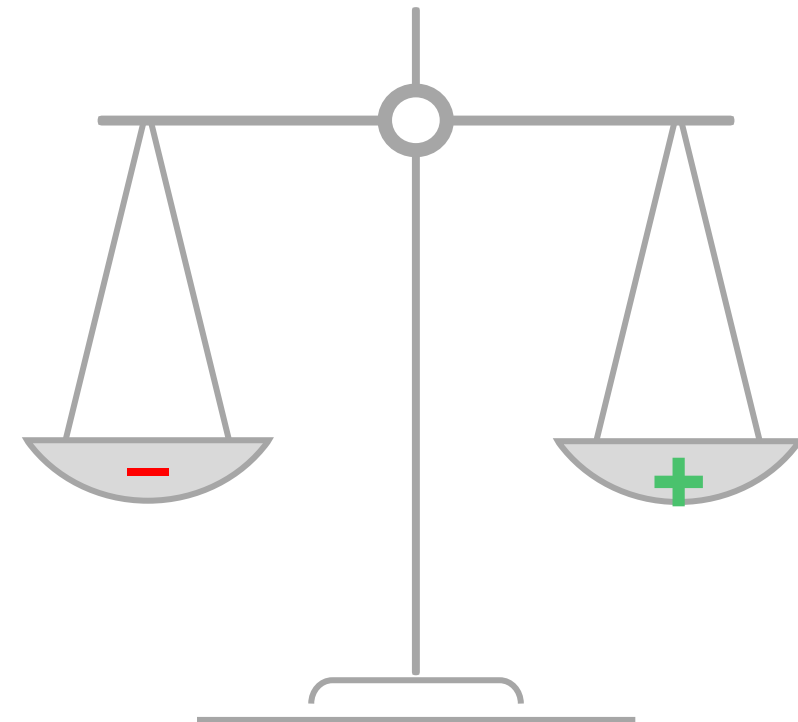
value of repreport

## Technical

- Ado dependencies
- Reliable detection of file paths (?)

## Substantive

- Caution with original data
- + “Brute force“ do-file is comprehensive  
(although it is not elegant)
- + Zero costs for users  
(unlike more elaborate solutions)



# THANKS!

✉ Daniel.Kraehmer@soziologie.uni-muenchen.de

🌐 www.danielkraehmer.com



<https://tinyurl.com/repreport-beta>

