



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Bundesamt für Statistik BFS
Office fédéral de la statistique OFS
Ufficio federale di statistica UST
Uffizi federal da statistica UST

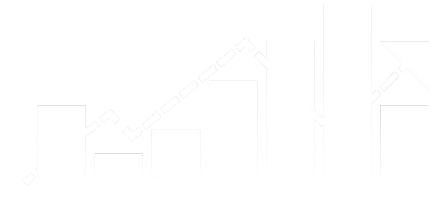


**8 December 2022,
at the FSFO, Neuchâtel**



Programm

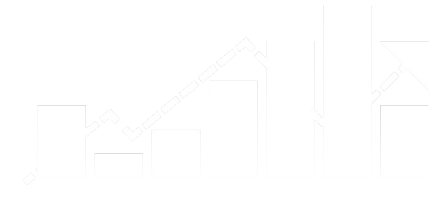
- Welcome (Renger van Nieuwkoop, SFSO)
- R at the SFSO (Renger van Nieuwkoop, SFSO)
- R/Python-API to Access Public SFSO Data (Fabian Santi, SFSO)
- Shiny Applications at VBZ, Yana Neuenschwander (VBZ)
- Après-Meetup: Drinks and Food





Welcome and Information

- All the presentations will be recorded and provided afterwards
- Each presentation will take about 15min with some additional time for questions
- Information on the building





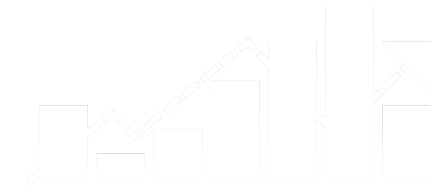
R at the Swiss Federal Statistical Office

Infrastructure and Examples

Autumn *meetup*

8.12.2022

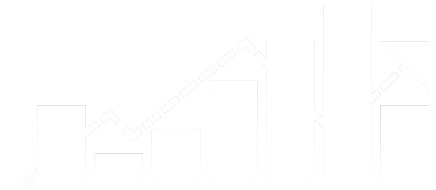
Renger van Nieuwkoop





R at the SFSO

Tools and Infrastructure





Tools at the SFSO

- Main tool until now is SAS, Excel, and Oracle
- Use of R, Python and GIT is steadily growing.
- In January 2018, R and Python were classified as equivalent to SAS, by decision of the Executive Board
- Most of the IT is centrally organised (Federal Office of Information, Technology, Systems, and Telecommunication, commonly referred to as «BIT»)
- The IT-Section at the FSO



SAS at the FSO

- More than 500 users, including 280 regular users
- ca. 40'000 SAS programs in the production
- Server installation from BIT (CPU 2700 MHz, 24 cores)
- SAS Support (250% full-time equivalent), the SAS Competence Center.





R

- > 100 active users, number of scripts in production > 1'000
- Starting 2022: each employee can install R (4.0.4) and R Studio on their own work computer (16GB RAM) as well as install packages from CRAN.
- Run R scripts via the Linux server (version 3.6.3, 2020, because of SAS)
- Renku (up to 11CPUs, 104G Memory and 512G Disk space). Renku includes Jupyter notebooks with R and Python kernels are available as well as Rstudio and Gitlab for version control.



Goals

Establish these tools and find a good infrastructure solution





Data-Analytics Tools Community

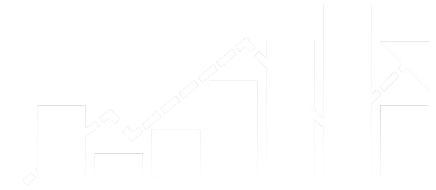


- The Data Analytics Tools Community is a group of internal experts that provide solutions related to data analytics tools consisting of representatives from almost all divisions of the FSO
- **Goal:**
 - Facilitation of the optimal use of data analytic tools at FSO through general knowledge transfer and support.
 - Support automation in data processing and promote re-use of existing processes.
- **Activities:**
 - weekly meetings
 - workshops/presentations
 - courses
 - documentation (Confluence site)



R at the SFSO

Examples





Input-Output Tables

Before

- ca. 1'400 files (1.4 Gb), ca. 600 Excel files with almost 5'000 worksheets, no version control
- Excel/Matlab/Gams/R and Mac vs Windows.
- Documentation partially outdated/lacking
- No automation
- Only for one year,
- Months of work

Now

- ca. 30 R-(Markdown)-Scripts, 60 Data files (350 Mb)
- R, Oracle, GIT, GAMS
- Rmarkdown for Documentation
- 80 - 90% automated
- Multiple years³
- ca. 1 – 2 weeks of work



30 Excel files to Eurostat (Section SNA)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ESA2010 Questionnaire 0301- Output													
2	TABLE_IDENTIFIER:	T0301	FREQ:		A									
3	REF_AREA:	CH	TIME:		2004		< >							
4	REF_PERIOD_DETAIL:	C	TIME_FORMAT:		P1Y									
5	REF_SECTOR:	S1	PRICES:		V									
6	COUNTERPART_AREA:	W2	REF_YEAR_PRICE:											
7	COUNTERPART_SECTOR:	S1	UNIT_MEASURE:		XDC									
8	STO:	Row 19	UNIT_MULT:		6									
9	INSTR_ASSET:	_Z	DECIMALS:		0									
10	ACTIVITY:	Column 1	ADJUSTMENT:		N									
11	EXPENDITURE:	_Z	TRANSFORMATION:		N									
12	ACCOUNTING_ENTRY:	Row 20	EMBARGO_DATE											
13		Fabrice Jacolet	LAST_UPDATE		2021-08-24T08:12:00									
14	Sender e-mail:	fabrice.jacolet@bfs.admin.ch	OBS_STATUS:		A									
15	Version:	Excel Template V2018-05	CONF_STATUS:		F									
17		Output		Intermediate consumption		Gross value added		Consumption of fixed capital		Net operational mix				
18														
19	STO ▶	P1	S T A T U S	S T A T U S	P2	S T A T U S	S T A T U S	B1G	S T A T U S	S T A T U S	P51C	S T A T U S	S T A T U S	B2A3M
20	ACCOUNTING_ENTRY ▶	C			D			B			D			B
21	ACTIVITY ▼	1			2			3=1-2			4			5
22	T	1'012'212	A	F	527'071	A	F	485'141	A	F	NaN	J	F	
23	A	12'481	A	F	7'785	A	F	4'696	A	F	NaN	J	F	
24	A01	11'815	A	N	7'383	A	N	4'433	A	N	NaN	J	F	
25	A02	630	A	N	392	A	N	238	A	N	NaN	J	F	
26	A03	35	A	N	10	A	N	25	A	N	NaN	J	F	
27	BTE	303'922	A	F	199'337	A	F	104'585	A	F	NaN	J	F	
28	B	1'806	A	F	1'093	A	F	713	A	F	NaN	J	F	
29	B05	0	A	N	0	A	N	0	A	N	NaN	J	F	
30	B06	0	A	N	0	A	N	0	A	N	NaN	J	F	
31	B07	0	A	N	0	A	N	0	A	N	NaN	J	F	
32	B08	1'806	A	N	1'093	A	N	713	A	N	NaN	J	F	

Before:

- sometimes with up to 75 worksheets
- copy-paste often the only option
- effort enorm and error-prone

After:

- Script: ca 300 lines of readable code
- one-time effort: 1 day programming
- Annual effort: < 5 minutes
- Main work: Adjust file names 😊
- Run script (3 minutes)



Other Examples

- Creation of official BFS documents
- Reading tables from PDF documents (instead of copy-paste)
- Some innovative projects using Machine Learning in R and Python. E.g:
 - NOGAuto: prediction system for NOGA codes from German, French and Italian activity descriptions
 - ADELE: Arealstatistik DEep LEarning
 - ML-SoSi: Machine Learning Soziale Sicherheit
- Imputation and optimization methods for FSO production systems
- etc.

Fehleranalyse erfolgen, um sicherzustellen, dass die Ergebnisse in allen Sektoren den Qualitätsstandards des BFS entsprechen.

2 Ökologische Faktoren

Für viele Menschen ist die ökologische Dimension der Nachhaltigkeit die erste, die mit «Nachhaltigkeit» assoziiert wird. Sie beschreibt, wie mit verschiedenen Ressourcen umgegangen wird. Die Ressourcen werden oft in Land, Wasser und Luft unterteilt.

Die Nutzung der Landflächen wird in diesem Bericht nicht behandelt, weil eine geographische Analyse nicht mit dem TSA kompatibel scheint (siehe Abschnitt 6). Eine eigenständige Analyse der Entwicklung der Landnutzung in Tourismusgemeinden scheint jedoch möglich. Diese wäre jedoch nicht Teil eines Indikatorensystems.

Ein anderer Aspekt der Landnutzung ist für diese Notiz analysiert worden – die Energienutzung. Das BFS hat Daten über die Energiequellen verschiedener Wirtschaftszweige. Es werden Indikatoren präsentiert, die die Energiequellen nach Tourismusprodukt als auch die Verteilung der Tourismusprodukte pro Energiequelle dargestellt. Wie auch für ökonomische Faktoren, könnte hier ein Niveau des Anteils einer einzelnen Energiequelle festgesetzt werden, ab welchem ein Tourismusprodukt (kurzfristig) abhängig von einer Energiequelle ist.

Abbildung 1: Verteilung der Energiequellen nach Tourismusprodukt

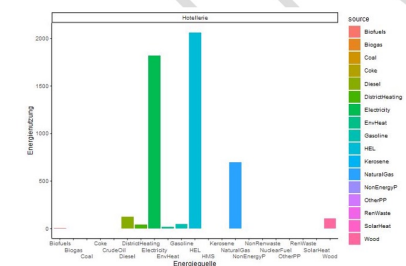
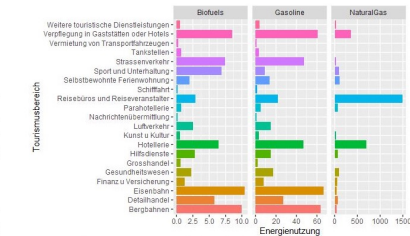
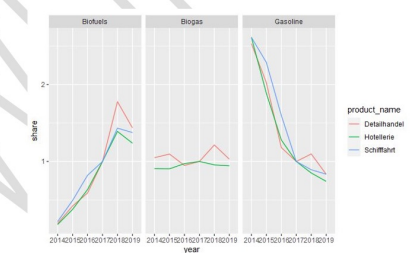


Abbildung 2: Verteilung der Tourismusprodukte nach Energiequelle



Es ist auffallend, dass mit Ausnahme der Luftfahrt, Diesel und Strom fast in allen Tourismusprodukten sehr ähnliche Anteile an den Energiequellen haben.

Abbildung 3: Zeitliche Entwicklung der Energiequellen für unterschiedliche Tourismusprodukte



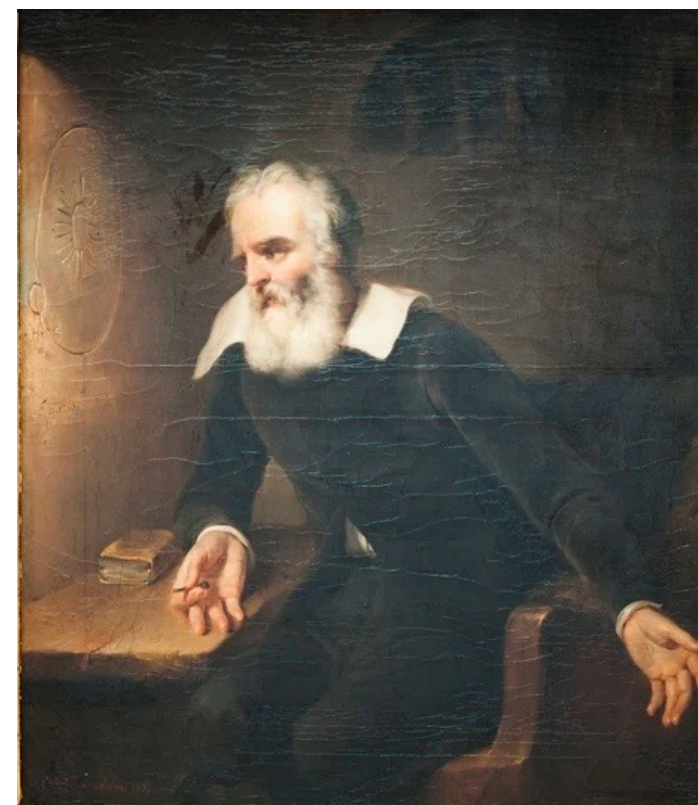
Die Nutzung von Biokraftstoffen hat stark im dargestellten Zeitraum zugenommen während Benzin stetig im gleichen Zeitraum abgenommen hat. Ein starker Anstieg der Nutzung in 2018 in fast allen Quellen fällt auf. Dies ist auf die TSA Ergebnisse und nicht die Energienutzung in diesem Jahr zurückzuführen. Die Nutzung von Biogas und normalem Gas ist volatil jedoch im Durchschnitt stabil.

Des Weiteren können Energiequellen gruppiert werden. Energiequellen können als erneuerbare Energie oder fossile Brenn-



Conclusions

- R is being used more and more at the SFSO and might become, together with Python one of the main tools in production
- Eppur si muove! We are in the process of setting up:
 - the proper infrastructure, and
 - the support
- We are eager to learn from your experience at your workplace!

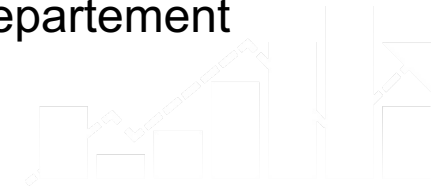




Questions/Discussions



We are particularly
interested in the solutions
for R at your departement



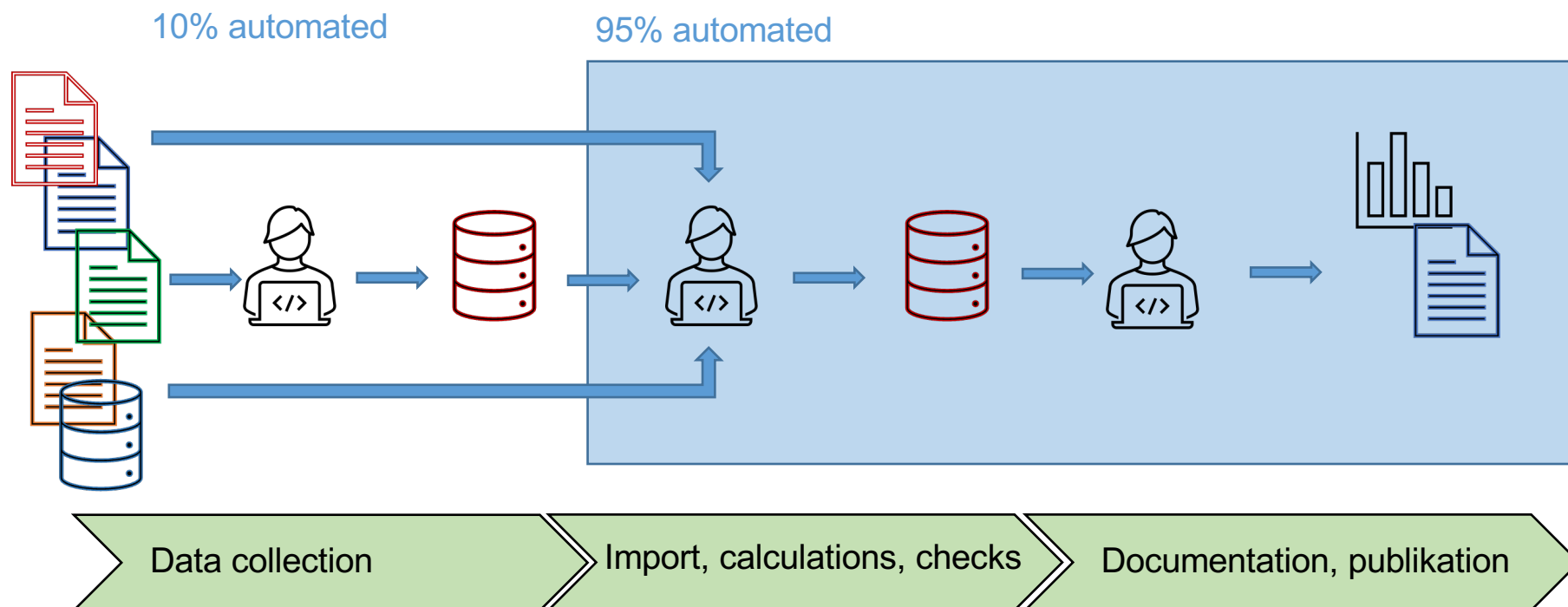


Reserve slides





Example: Input-Output Tables





Confluence

Confluence

Bereiche ▼ Personen Erstellen ...

Community - Data Analytics Tools

SEITENHIERARCHIE

▼ Public

> SAS

▼ R



> R - TTT: Tipps, Tricks and Tools

> Python

> Git


• Workshop 29. Nov 2022

> Internal

Seiten / Community - Data Analytics Tools / Public  

R

Erstellt von Didier Staudenmann, zuletzt geändert von Renger Van Nieuwkoop vor Kurzem



R is a programming language for statistical computing and graphics supported by the R Core Team and the R Foundation for Statistical Computing. Users have created packages to augment the functions of the R language.

According to user surveys and studies of scholarly literature databases, R is one of the most commonly used programming language used

The official R software environment is an open-source free software environment within the GNU package, available under the GNU General Public License. R has a command line interface.[8] Multiple third-party graphical user interfaces are also available, such as RStudio.

[Wikipedia \(visited 5.4.2022\)](#)

At the FSO **R 4.0.4** is available (release date February, 2021) as well as the IDE is [RStudio](#) (1.4.1106). Installation via [APS Kundenportal BIT](#) freely updated and installed.

Useful links to "R sites"

Link	Description
Awesome official statistics software	A list of open source software (R packages, Python libraries) for official statistics. R packages for e.g. access to official statistics, imputation,