



ПРЕЗИДЕНТСКАЯ  
АКАДЕМИЯ

2025

# «Чему учить студентов», или анализ использования СПО в исследованиях

Шалаев Н. Е., к. полит. н.,  
доцент ФМОПИ СЗИУ РАНХиГС

Переславль-  
Залесский 2025

РАНХиГС

# СПО в социальных науках

## “Индустриальный стандарт”

- С 90х в отрасли укоренилось проприетарное ПО
- Вопрос лицензий и сохранения доступа к ПО никого не волнует
- В самой насыщенной области — статистике — концентрация проприетарного ПО особенно высока
- Для перехода на СПО нет очевидных (объективных) аргументов
- Есть СПО в ролях “на подхвате”
- Много предложений в формате SaaS, что удручет



# Откуда берётся идея о “стандарте”?

## опыт окружающих + сила привычки

- Про проприетарное ПО написано и прочитано огромное количество книг и статей
- Переучиваться никто не хочет, а оглядываясь на окружающих видит, что никто и не мигрирует
- Проектов систематического продвижения СПО практически нет

 BROWN UNIVERSITY LIBRARY

[All Guides/ Social Sciences Data Resources](#)

**Social Sciences Data Resources**

Resources, tools and tops for finding and managing data across the social sciences.

**Overview**

**Demographic**

**US Census Data**

**Economic & Financial**

**Elections & Polls**

**Government Information**

**GIS & Spatial Data**

**Statistical Software**

Software at Brown

Brown OIT manages access to proprietary statistical software packages like SAS, SPSS, and STATA. For a complete overview of software available for all Brown community members visit the [Brown University Software Catalog](#).

The library can support research in Python, R, and SQL, and can provide basic support for STATA.

**Python**

Python is an open source, general purpose scripting and programming language with good support for statistical analysis and data visualization.

**Download and Installation Instructions**

- [Python Homepage](#)
- [Anacondas \(Python distro bundled with common modules like NumPY, Pandas, and Matplotlib\)](#)

**Resources for learning Python**

- [Python manuals and documentation](#)
- [Statistics in Python Tutorial \(with SciPy\)](#)
- [Statistical Analysis in Python Using Pandas](#)

**Practical Programming** by Paul Gries; Jennifer Campbell; Jason Montojo  
Publication Date: 2017-12-26

**Python for Data Analysis** by Wes McKinney

Search BruKnow

 Berkeley Initiative for Transparency in the Social Sciences

[Home](#) [About](#) [Events](#) [Evidence](#) [Education](#) [Ecosystem](#) [Resources](#)

**Subscribe to Blog via Email**

Enter your email address to subscribe to this blog and receive notifications of new posts by email.

Join 4,461 other subscribers

**Contribute to the Blog**

Interested in authoring a post on the BITSS blog? Review our [Guidance for Authors](#) and contact [Jo Weech](#).

**Open Source Software for Reproducible Social Science**

January 8, 2016  
Garret Christensen –BITSS Project Scientist

BITSS offers grad student workshops in reproducible research, where we give a hands-on introduction to software that can help make your work more reproducible. A lot of the software is listed on the [Software](#) section of our Resources page, but I wanted to create a quick narrative summary of software that people can use to make their research more reproducible.

**1: Dynamic Documents**

You can write your code and your paper in one place. This means you won't screw anything up copying and pasting, and you'll never have to wonder what code it was that produced which figure, and where on earth you saved it, or whether the paper has the updated version.

In R, this can be done with R Markdown, which is built into R Studio. So download and install [R](#) and [R Studio](#). When you open a new R Markdown file in R Studio, it starts with a really simple example, or you can learn more [here](#).

# Где взять аргументы против статуса quo?

## “Мировые практики”

- Рейтингов ПО нет
- Есть рейтинги языков программирования
- Рейтинги языков программирования показывают “не то, что надо”
- Популярность ≠ релевантность
- Спрос у работодателей в целом ≠ востребованность в специфической области
- Число строк кода ≠ важность
- Но тем не менее...

**TIOBE** the software quality company

Schedule a demo

Jan 2025 Jan 2024 Change Programming Language Ratings Change

Rank	Rank	Change	Icon	Language	Rating	Change
1	1		Python	23.28%	+9.32%	
2	3	▲	C++	10.29%	+0.33%	
3	4	▲	Java	10.15%	+2.28%	
4	2	▼	C	8.86%	-2.59%	
5	5		C#	4.45%	-2.71%	
6	6		JavaScript	4.20%	+1.43%	
7	11	▲	Go	2.61%	+1.24%	
8	9	▲	SQL	2.41%	+0.95%	
9	8	▼	Visual Basic	2.37%	+0.77%	
10	12	▲	Fortran	2.04%	+0.94%	
11	13	▲	Delphi/Object Pascal	1.79%	+0.70%	
12	10	▼	Scratch	1.55%	+0.11%	
13	7	▼	PHP	1.38%	-0.41%	
14	19	▲	Rust	1.16%	+0.37%	
15	14	▼	MATLAB	1.07%	+0.09%	
16	18	▲	Ruby	1.06%	+0.25%	
17	15	▼	Assembly language	1.01%	+0.10%	
18	23	▲	R	1.00%	+0.27%	
19	16	▼	Swift	0.99%	+0.10%	

# Что можно увидеть?

Worldwide, Feb 2025 :

Rank	Change	Language	Share	1-year trend
1		Python	29.85 %	+1.6 %
2		Java	15.15 %	-0.7 %
3		JavaScript	7.92 %	-0.8 %
4		C/C++	7.19 %	+0.5 %
5		C#	6.13 %	-0.5 %
6		R	4.55 %	-0.1 %
7		PHP	3.72 %	
8	↑↑	Rust	3.07 %	
9	↑↑	Objective-C	2.86 %	
10	↓↓	TypeScript	2.74 %	
11	↓↓	Swift	2.46 %	
12		Go	2.07 %	
13		Kotlin	1.87 %	
14		Matlab	1.7 %	
15	↑	Ada	1.25 %	
16	↑	Ruby	1.02 %	

2025

## “Кругом Питон (спасибо, что хоть не JS)”

- Python является лидером везде
- R занимает высокое положение только в PYPL
- Но и тот же Matlab тоже не в лидерах
- Для специализированных языков такой метод, видимо, не подходит.

		Python	R	Matlab	Fortran	Java	Perl
	TIOBE	1	18	13	8	3	27
	IEEE Spectrum	1	20	24	45	2	25
	PYPL (на рис.)	1	6	14	—	2	—
	GitHub	1	33	43	40	2	19
	StackOverflow	3	20	19	34	5	39
	Zdnet	1	17	—	—	3	—

# Где искать “нишевую популярность”?

“В репах!”

- Воспроизводимые исследования и открытые данные → репозитории данных
- Данные генерируются профессиональными исследователями
- Репозиториев куча (Zenodo, Figshare, Harvard Dataverse и проч.)
- Mendeley Data: репозиторий + агрегатор
- Тут и будем искать

 Mendeley Data



Sign In / Register

## Share your research data

Mendeley Data is a free and secure cloud-based communal repository where you can store your data, ensuring it is easy to share, access and cite, wherever you are.

[Create a Dataset](#)

Find out more about our institutional offering, [Digital Commons Data](#)

## Search the repository



[Advanced search](#)

Search results powered by Data Monitor

### Recently published

[Frameworks For Assessing Agroecological Status In agricultural Systems: A Systematic Review 2015-2024](#)

parra gonzález, cristian nacianceno  
Published 4 February 2025 | Mendeley Data  
Original data and after compiled and purified

[Dataset](#)

Export: [APA](#) | [BibTeX](#) | [DataCite](#) | [RIS](#)



### The Generalist Repository Ecosystem Initiative

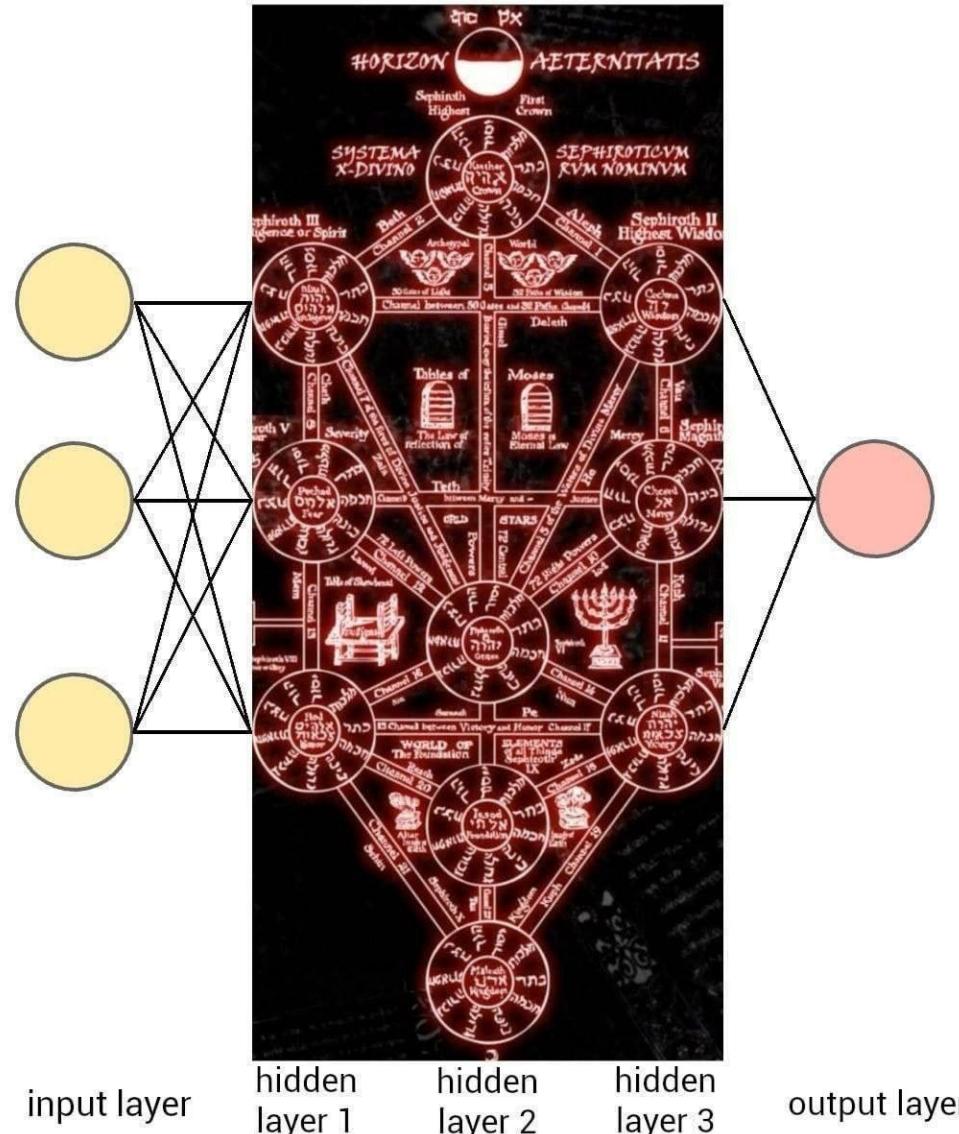
Elsevier's Mendeley Data repository is a participating member of the National Institutes of Health (NIH) Office of Data Science Strategy (ODSS) GREI project. The GREI includes seven established generalist repositories funded by the NIH to work together to establish consistent metadata, develop use cases for data sharing, train and educate researchers on FAIR data and the importance of data sharing, and more.

[Find out more](#)

# Дизайн эксперимента

“пароль стандартный — несколько звёздочек”

- Мы не гордые, парсим всё на Perl
  - Поиск специфических и неспецифических терминов
  - Неспецифические термины — “как есть”
  - Специфические: термин + уточняющие токены
    - Примеры: “file”, “data”, “software”
    - “program”, “script”, “language”, “source code”
  - Неспецифический поиск (что ожидаемо) нормально не работает
  - Специфический поиск, видимо, требует доработки
  - Истина где-то посередине. Во всяком случае, между.



# Группа ПО 1: статистика

Search	Non-specific	Specific	Figshare ARS	Zenodo	DRYAD	Mendeley Data	Harvard Dataverse
Excel	116827	24519	7442	5100	3659	1687	632
JMP	881	46	11	6	17	9	0
Julia	1321838	407	54	250	14	30	1
KNIME	416		167	172	12	21	15
PSPP	50		6	8	1	16	2
Python	62713	11290	2262	5205	758	445	101
R	5164122	23840	8468	4300	5192	718	944
RStudio	2672		535	867	596	182	65
jamovi	167		67	39	6	23	3
JASP	338	23	2	3		3	
SAS	79765	448	76	38	22	30	113
SPSS	20504	1917	285	70	34	139	688
Stata	18927	1746	71	62	13	70	1173

# Группа ПО 2: сетевой анализ (и gnuplot)

Search	Non-specific	Specific	Figshare ARS	Zenodo	DRYAD	Mendele y Data	Harvard Datavers e
<b>Cytoscape</b>	2088	98	69	12	4	2	0
<b>gephi</b>	641		309	130	11	69	19
<b>graphviz</b>	230		57	118	5	14	11
<b>Pajek</b>	199	3	2	1			
<b>SocNetV</b>	3		2			1	
<b>gnuplot</b>	385	128	34	47	4	15	0

## Примечания

- ORA (CASOS ORA, ORA-PRO, ORA-LITE) видимо, померла (*sic semper tyrannis*)
- Tulip (<https://tulip.labri.fr/site/>) толком не найти
- Gnuplot — почётный гость

# Группа ПО 3: ЯП числодробильные (и не только)

Search	Non-specific	Specific	figshare ARS	Zenodo	DRYAD	Mendeley Data	Harvard Dataverse
<b>Fortran</b>	4019	842	63	179	26	180	23
<b>Java</b>	23833	1136	201	643	32	115	4
<b>Perl</b>	5127	1044	403	145	257	29	7
<b>Maple</b>	4188	85	20	26	9	16	0
<b>Matlab</b>	33779	7457	1903	2015	572	518	175
<b>Maxima</b>	17634	6	2	2	0	2	0
<b>Octave</b>	2675	108	44	31	1	6	0
<b>PEBL</b>	34		18	2	1	2	

## Примечания

- Perl наше всё
- PEBL = Psychology Experiment Building Language

# Выводы

**“СПО в массы! Хотя оно, собственно, уже там.”**

- R успешно конкурирует с проприетарными тяжеловесами
- R — ядро концепции “воспроизводимых исследований”
- Python популярен не только среди разработчиков, но и исследователей — его тоже надо учить
- В некоторых областях СПО является де-факто стандартом (для QCA, см. илл: R + пакеты, никакой tosmana, kirq и проч.), для сетевого анализа отдельное ПО или опять-таки R с пакетами от Statnet.
- Даже весьма нишевое СПО (gnuplot, PEBL, Veusz) можно найти в репозиториях исследовательских данных
- Доминирование SPSS и компаний — живучий миф

Software	Environment	Methodology
<a href="#">acg</a>	Unix shell	csQCA, fsQCA
<a href="#">FSGoF</a>	DOS	Fit test for fuzzy-set analysis
<a href="#">fs/QCA</a>	Windows, macOS	csQCA, fsQCA
<a href="#">fuzzy</a>	Stata	csQCA, fsQCA
<a href="#">Kirq</a>	Windows, macOS, Linux	csQCA, fsQCA
<a href="#">MDSO/MSDO</a>	Web app	MDSO/MSDO
<a href="#">EvalC3</a>	Microsoft Excel (Windows)	QCA type configurational analysis
<a href="#">TOSMANA</a>	Windows (.NET/Mono)	csQCA, mvQCA, fsQCA
<a href="#">QCA Add-in for Excel</a>	MS Excel	csQCA, mvQCA, fsQCA
<a href="#">causalHyperGraph</a>	R	Draw causal hypergraphs
<a href="#">cna</a>	R	Coincidence Analysis
<a href="#">SetMethods</a>	R	Additional functions for QCA
<a href="#">LogicOpt</a>	R	Truth table minimization
<a href="#">NCA</a>	R	Necessary conditions analysis
<a href="#">QCA</a>	R	csQCA, mvQCA, fsQCA
<a href="#">QCA3</a>	R	csQCA, mvQCA, fsQCA
<a href="#">QCAfalsePositive</a>	R	Type I error test
<a href="#">QCApro</a>	R	csQCA, mvQCA, fsQCA
<a href="#">QCAtools</a>	R	Additional functions for QCA
<a href="#">Venn</a>	R	Draw Venn diagrams



ПРЕЗИДЕНТСКАЯ  
АКАДЕМИЯ

**СПАСИБО ЗА ВНИМАНИЕ!**

**[shalaev-ne@ranepa.ru](mailto:shalaev-ne@ranepa.ru)**

Переславль-  
Залесский 2025

2025

РАНХиГС