

R in meinem Arbeitsalltag

Jan-Philipp Kolb

16 November 2017

- VWL Studium in Trier (Diplom 2008)
- 2004 Erasmus Jahr an der Université Jean Moulin in Lyon
- Wissenschaftlicher Mitarbeiter am Lehrstuhl für Wirtschafts- und Sozialstatistik
- 2012 Promotion (Thema: Die Erzeugung von synthetischen Grundgesamtheiten)
- Seit 2012 am Gesis Leibniz Institut für Sozialwissenschaften - zunächst Team Statistik
- Seit 2017 Survey Statistik im Team Gesis Panel

Fünf wissenschaftliche Abteilungen

- Datenarchiv für Sozialwissenschaften (DAS)
- Dauerbeobachtung der Gesellschaft (DBG)
- Computational Social Sciences (CSS)
- Survey Design and Methodology (SDM)
- Wissenstechnologien für Sozialwissenschaften (WTS)

Gesis ist:

- Infrastruktureinrichtung für die Sozialwissenschaften
- mit über 250 MitarbeiterInnen an zwei Standorten (Köln und Mannheim)

GESIS bietet:

- Beratung zu Forschungsprojekten in allen Phasen
- Forschungsbasierte wissenschaftliche Dienstleistungen

- Probabilistisches mixed-mode Access Panel
- Deutsche Allgemeinbevölkerung
- Deutschsprachig - ab 18 Jahren
- Basierend auf Einwohnermeldeamtsstichprobe
- Mehrstufiger Rekrutierungsprozess, sequentielles mixed-mode Design
- Initiale Rekrutierung 2013
- 2016 Auffrischungsstichprobe

```
## [1] "tab_spatial"
```

	Kürzel	Studientitel
7	ag	Environmental Spatial Strategies
14	an	Leisure travel and subjective well-being
17	aq	Pro-environmental Behavior in High Cost Situations
48	bw	Space-sets: the scope and characteristics of national and in

- Panelbereinigung (bei Abmeldung oder Nonresponse)
- Anonymisierung und Kategorisierung
- Filterführung muss sich in den Missings widerspiegeln

- `foreign`, `readstata13` und `xlsx` zum Import von Daten
- `dplyr` Paket zur Datenaufbereitung
- `doParallel` zur Bearbeitung vieler Jobs
- `Rmarkdown` bei der Datendokumentation (Codebook, Wave Report)
- `caret` für maschinelles Lernen
- `Rstudio` git Interface zur Versionskontrolle

- Cheatsheet zum Umgang mit Strings

```
560 <div id="questiontable"><div class="qt311"><div id="qnameq23076">
561 <table cellpadding="0" cellspacing="0" width="100%" border="0" bordercolor="green">
562 <tbody><tr>
563 <td class="questiontext"><a name="2"> </a>Wie ähnlich ist Ihnen diese Person?
564 </td>
565 </tr>
566 </tbody></table><br>
567 <table border="0" class="answertable" bordercolor="purple" cellspacing="0" cellpadding="0">
568 <tbody><tr><td class="scaletitle"></td>
569 <td class="answerscale">ist mir überhaupt nicht ähnlich<br>1</td>
570 <td class="answerscale">ist mir nicht ähnlich<br>2</td>
571 <td class="answerscale">ist mir nur ein wenig ähnlich<br>3</td>
572 <td class="answerscale">ist mir einigermaßen ähnlich<br>4</td>
573 <td class="answerscale">ist mir ähnlich<br>5</td>
574 <td class="answerscale">ist mir sehr ähnlich<br>6</td>
575 <td width="100%" class="scaletitle"></td>
576 </tr>
```

Figure 1

Welche Programm werden genutzt

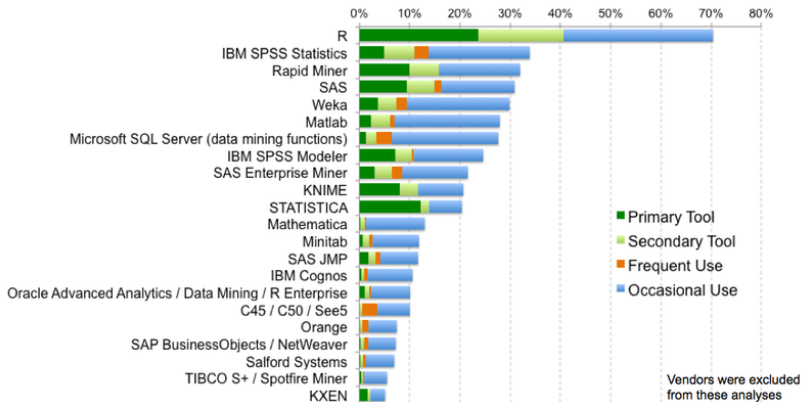


Figure 2

Open Science


Das Prinzip „Open Science“ hat das Ziel, wissenschaftliche Abläufe offen zugänglich, nachvollziehbar und nutzbar zu machen. Dazu werden verschiedene Ansätze verfolgt, beispielsweise Open Access, **Open Source**, Citizen Science und Open Educational Resources. Wie verschiedene Stellungnahmen der Europäischen Union und der G7 zeigen, gewinnt Open Science auch auf europäischer und internationaler Ebene an wissenschaftspolitischer Bedeutung. Die Leibniz-Gemeinschaft und ihre Mitgliedseinrichtungen unterstützen diese Entwicklung und gestalten sie mit. So setzen sie sich beispielsweise seit vielen Jahren mit zahlreichen Aktivitäten für **Open Access** , den freien Zugang zu wissenschaftlichen Publikationen, ein.

Figure 3



Eine neue, strategisch wichtige Herausforderung für GESIS ist die Forschung zu kollaborativen und partizipativen Modellen und Infrastrukturen, die Open Science-Prozesse in den Sozialwissenschaften unterstützen.

Figure 4

Job trends

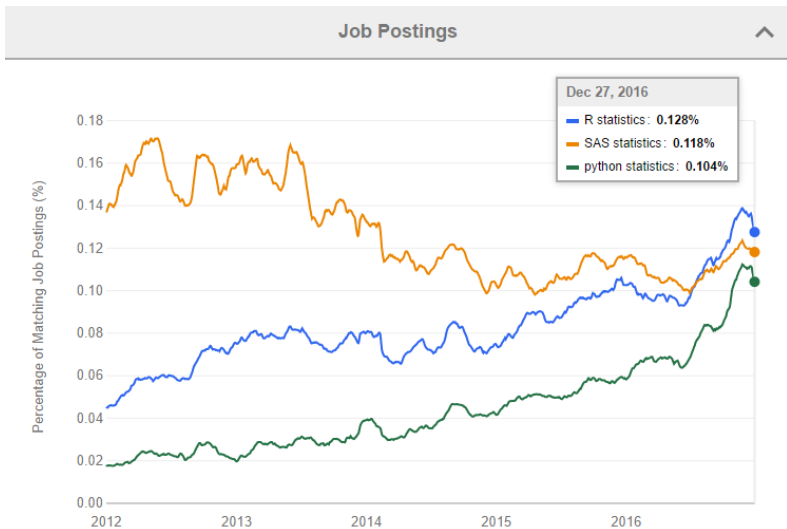


Figure 5

- Bedeutung von R scheint zuzunehmen
- Open Science Entwicklung
- Stata ist nach wie vor wichtig (Pfadabhängigkeit)
- Bedeutung von SPSS nimmt ab