

---

# AGENS - Umfrage zu Auswertungssoftware im FDZ Gesundheit

---

## Ergebnisse

### Umfrage 288572

---

Anzahl der Datensätze in dieser Abfrage:	192
Gesamtzahl der Datensätze dieser Umfrage:	192
Anteil in Prozent:	100.00%

---

## Zusammenfassung für Q00

Welche der folgenden Softwaretools würden Sie nutzen, wenn Sie in den virtuellen Analyseräumen des FDZ Gesundheit Auswertungen durchführen würden?

---

Antwort	Anzahl	Prozent
MATLAB (SQ001)	1	0.52%
Python (SQ002)	51	26.56%
R (SQ003)	129	67.19%
SAS (SQ004)	75	39.06%
SPSS (SQ005)	53	27.60%
SQL (SQ006)	63	32.81%
STATA (SQ007)	45	23.44%
Sonstiges	9	4.69%

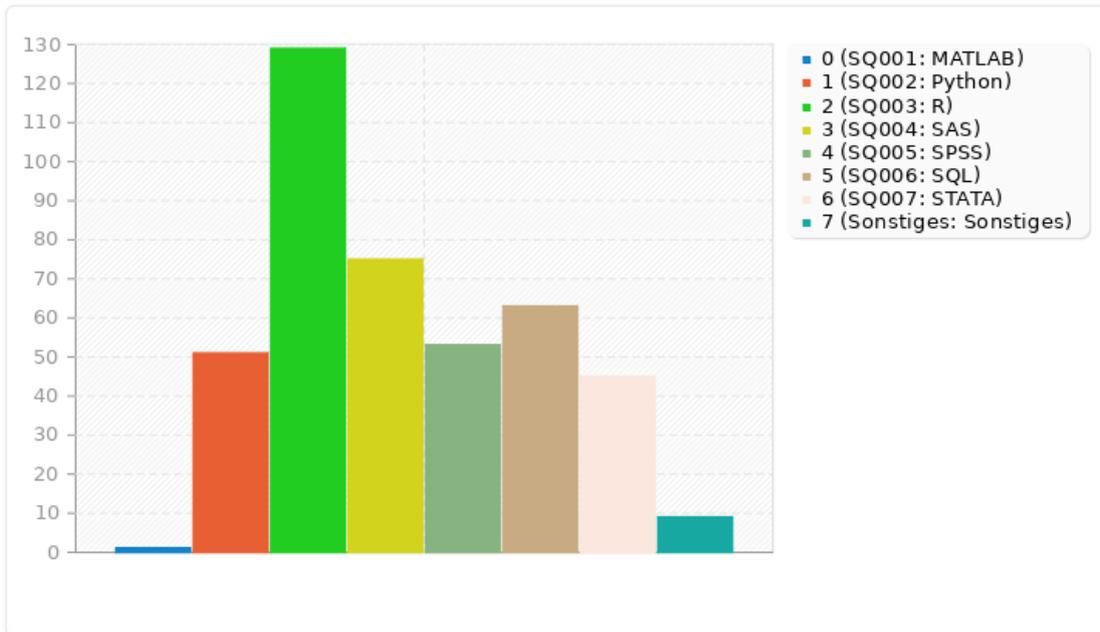
ID	Antwort
26	SPSS Modeler
39	Jupyter Notebook
41	KNIME
91	Epicure, Joinpoint Regression Program (NCI)
92	PowerBI, Toad
104	Excel
175	Excel
215	Weitere Ergebnisaufbereitung in Excel
229	Excel

---

## Zusammenfassung für Q00

Welche der folgenden Softwaretools würden Sie nutzen, wenn Sie in den virtuellen Analyseräumen des FDZ Gesundheit Auswertungen durchführen würden?

---



## Zusammenfassung für Q01

Falls Sie dazu Kommentare oder Fragen haben, formulieren Sie diese bitte im nachfolgenden Freitextfeld:

Antwort	Anzahl	Prozent
Antwort	21	10.94%
Keine Antwort	171	89.06%
Nicht gezeigt	0	0.00%

ID	Antwort
19	Gerade für R wäre es äußerst hilfreich, könnte man Pakete selbst installieren. Vielleicht zumindest die Pakete, die bei CRAN sind.
34	Seit Jahren Nutzer von SAS beim FDZ
39	Wichtig ist, dass die Softwaretools grundsätzlich aktuell gehalten werden. Startet eine Projekt sollten die bereitgestellten Softwaretools einen möglichst aktuellen Stand haben oder die Möglichkeit gegeben sein, den Stand aus einem anderen Projekt zu übernehmen. Darüber hinaus sollte es möglich sein, pro Projekt zu entscheiden, ob die im Projekt verwendeten Softwaretools aktualisiert werden. Es sollte klar geregelt sein, welchen Support das FDZ, dass FDZ unter diesen Umständen leisten kann, da ein Support verschiedenster Softwaretools, die dazu noch in verschiedenen Version genutzt werden dürfen, nur zu Teilen möglich ist.
55	R und Python bisher noch nicht genutzt
72	Ich finde die "Stewardship-Entscheidung" des FDZ, auf Open-Source-Programme zu setzen, SEHR gut, da sich dies auf die Ausbildung zukünftiger Versorgungsforscher auswirken wird. Die Zeiten kommerzieller Statistik-Software sind meiner Meinung nach zu Recht vorbei, und es gibt meiner Meinung nach - außer Bequemlichkeit - absolut keinen Grund, Support für unterlegene Produkte einzufordern. Wir als Forschende sollten mit dem begrenzten Budget des FDZ verantwortungsvoll umgehen, denn jeder Euro, der in Richtung der Dinosaurier fließt, fehlt woanders. Für alle Legacy-Nutzer gibt es genug kostenfreie Ressourcen, um R (oder Python, wobei die Lernkurve hier steiler ist) zu lernen.
77	Aus meiner Sicht tun wir gut daran, auf Open-Source-Lösungen zu setzen und das FDZ nicht durch kommerzielle Statistiksoftware zu bottlenecken und Geld zu verschwenden, das an anderer Stelle besser eingesetzt wäre.
83	die Open Source Sprachen R und Python bieten eine gute Grundlage um für alle Forscher gleiche Bedingungen zu schaffen, da jeder die Möglichkeit hat diese selbstständig zu erlernen ohne für große Summen Lizenzen zu kaufen.
110	STATA sollte auch nutzbar sein.
125	Schon vorab, nachfolgend wird nach benötigten R Paketen gefragt. Ich nutze unzählige verschiedene Pakete. Beim Versuch möglichst viele wichtige Pakete in dieser Befragung anzugeben, habe ich jedes mal die Sitzungszeit überschritten. Es ist sinnlos nach einzelnen Pakten zu fragen. Es muss die Möglichkeit geben, bei Bedarf R um jeweils benötigte Pakete zu erweitern. Z. B. anhand einer Anfrage an den Admin der R Umgebung o. Ä.
132	Es wäre wichtig, dass verschiedene Programme (z.B. Rstudio) auf Packages möglichst flexibel zugreifen können.
133	das sind die wesentlichen Softwarepakete - mit weniger wird es schwierig, weil die Analysemöglichkeiten dann eingeschränkt werden.
145	Die Frage ist zu unkonkret. Um sich einen Datensatz über mehrere Datenbankobjekte sinnvoll zusammenzustellen ist SQL absolut sinnvoll. Für die Analyse eines Datensatzes kommt es drauf an, welcher Schwerpunkt gesetzt wird. Ist es Statistik/Gesundheitsökonomie wähle ich R. Sind es u. a. Verfahren des maschinellen Lernens wähle ich Python.
173	Ich würde eine aktuelle R Version sowie die für die Analyse relevanten Packages nutzen. Auch würde ich versuchen, ein Versionskontrolle über git und eine "reproducible environment" über eine package-Verwaltung wie renv zu nutzen.
184	Bisher bin ich bei der Analyse so vorgegangen, dass ich einfache Abfragen mit SQL gemacht habe und logistische Regressionsanalysen etc. mit SPSS auf Syntaxisbasis.
207	Ich habe den Survey bereits einmal ausgefüllt. Hatte aber vermutlich die igrph bibliothek vergessen.
208	SQL via SAS
215	SQL ist auch innerhalb von SAS implementiert, ich nutze es dort und habe es insofern anderweitig nicht angegeben.
217	Wir arbeiten im gesamten GKV-Routinedaten-Team mit Stata

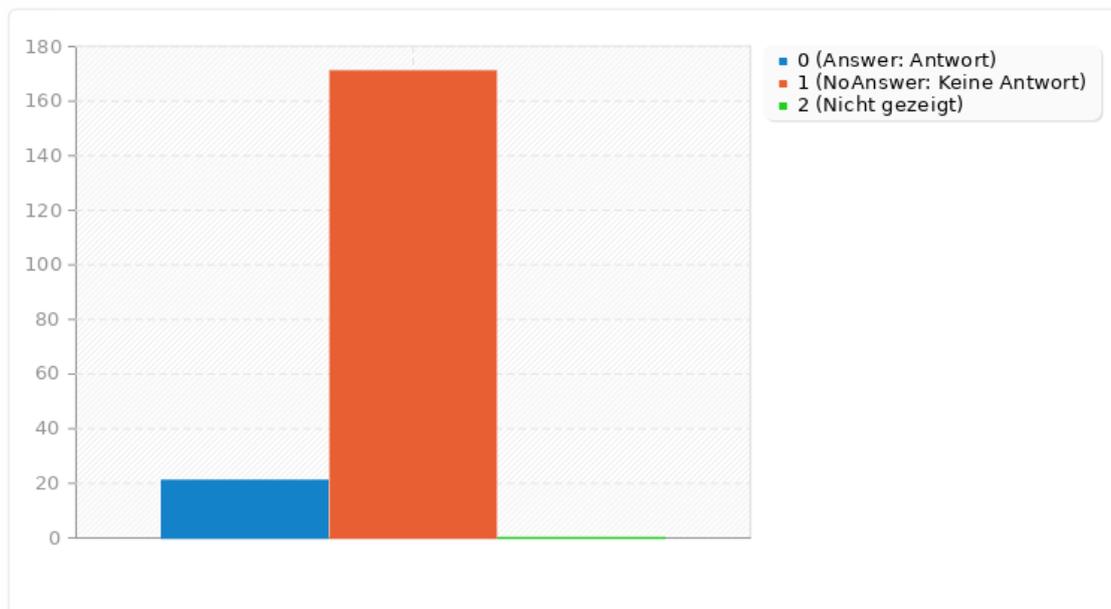
- 
- 232 Python ist zwar angekreuzt allerdings bei uns noch nicht im Einsatz
- 236 Wichtig grade für R wäre ein offener Port zur Package Installation, sowie die Möglichkeit das packages in der verwendeten version die für ein projekt zusammen mit der R Version im Package gespeichert werden. Ohne das diese von Server updates verändert werden.
- 262 Für die Verwendung von R ist die Bereitstellung von einigen Packages erforderlich, da wahrscheinlich keine Möglichkeit besteht in der sicheren Umgebung diese aus dem Web zu installieren.

---

### Zusammenfassung für Q01

Falls Sie dazu Kommentare oder Fragen haben, formulieren Sie diese bitte im nachfolgenden Freitextfeld:

---



---

## Zusammenfassung für G01Q01

Aktuell plant das FDZ Gesundheit, dass neben SQL (SAP/HANA) in den virtuellen Analyseräumen ausschließlich R und Python genutzt werden kann. Finden Sie diese Auswahl ausreichend?

Antwort	Anzahl	Prozent
Ja, diese Auswahl würde mir ausreichen (AO01)	90	46.88%
Nein, diese Auswahl würde mir nicht ausreichen (AO02)	102	53.12%
Keine Antwort	0	0.00%
Nicht gezeigt	0	0.00%

---

## Zusammenfassung für G01Q01

Aktuell plant das FDZ Gesundheit, dass neben SQL (SAP/HANA) in den virtuellen Analyseräumen ausschließlich R und Python genutzt werden kann. Finden Sie diese Auswahl ausreichend?



## Zusammenfassung für G01Q02

Falls Sie dazu Kommentare oder Fragen haben, formulieren Sie diese bitte im nachfolgenden Freitextfeld:

Antwort	Anzahl	Prozent
Antwort	39	20.31%
Keine Antwort	153	79.69%
Nicht gezeigt	0	0.00%

ID	Antwort
17	Weitere Tools wären wünschenswert, aber R und Python sich sicher weit verbreitet.
28	würde lieber mit STATA arbeiten, R wäre 2. Wahl
34	Mit nur noch R und Phyton grenzt das FDZ SAS-User aus, was nicht akzetabel ist.
40	bitte unbedingt SAS erhalten!
47	Kann mit R und Stata umgehen, bin aber wesentlich erfahrener in Stata, eine Stata-Option wäre daher schön.
50	Da ich in meiner Forschung ausschließlich SAS für meine Analysen verwende, ist die Nichtverwendung von SAS ein großes Hindernis bei der Durchführung meiner Analysen, und es erfordert mehr Zeit, eine neue Statistiksoftware wie SAS oder Python zu erlernen, was in meinem Fall schwierig ist, da ich meine Dissertation abschließen muss.
55	SAS wäre sinnvoll
58	SPSS sollte ebenfalls gegeben sein.
64	SPSS und Stata sollten ebenfalls möglich sein, es beherrschen nicht alle die bislang vorgesehenen Softwareprogramme.
66	SAS ist und bleibt m.E. einer der Standards in diesem Umfeld - sicherlich ist dies mit höheren Kosten verbunden, diese müssten dann über Nutzergebühren finanziert werden.
68	Mittelfristig würde vielleicht Python ausreichen.
72	Diese Auswahl ist nicht nur ausreichend, sondern absolut begrüßenswert.
77	Aus meiner Sicht tun wir gut daran, auf Open-Source-Lösungen zu setzen und das FDZ nicht durch kommerzielle Statistiksoftware zu bottlenecken und Geld zu verschwenden, das an anderer Stelle besser eingesetzt wäre.
84	SAS kann mit den anfallenden Datenmengen umgehen. Bei den anderen Lösungen bin ich mir nicht sicher.
92	Wir haben in der Vergangenheit vermehrt mit SAS gearbeitet und die Mitarbeitenden sind in der Anwendung dieser Software am fittesten bzw. haben zum Teil nie vorher mit R gearbeitet (in der Regel fehlt die Zeit der Einarbeitung in neue Programme, obwohl ich grundsätzlich die Anwendung von Open Source Software befürworte)
102	Das FDZ des Statistischen Bundesamtes bietet SAS und Stata, sowie R und SPSS am Gastwissenschaftsarbeitsplatz mit vorher eingeschränktem Datenmaterial. Nur SAS und Stata können mit diesen großen Datenmengen performant umgehen.
109	Völlig an den Nutzern vorbei. Beide Softwarepakete sind unüblich bei Routinedatenanalysen.
110	Es ist ein guter Start, anderes sollte in Ausnahmefällen möglich sein.
115	aber denn lernen wir eben r bzw. python ;-)
118	Für die Nutzung von R werden sicherlich zahlreiche Packages erforderlich, die je nach Fragestellung und Komplexität der Analyse von den einzelnen Nutzer:innen erfragt werden. Dies könnte das Arbeiten verlangsamen.
123	Vorteil bei z.B. Stata als weiterer Option ist, dass der Funktionsumfang meist ausreicht und man ohne Packages auskommt
125	SAS!!!
132	Auch wenn der Trend in der Community in die Richtung R/Python geht, gibt es dennoch viele Anwender, die damit eher wenig Erfahrung haben. Dazu zählen z. B. auch Studierende. Es sollte zumindest auch mind. ein "niedrigschwelliges" Analyseprogramm wie SPSS geben.
133	s. Kommentar zur Frage 1
136	Viele Programmierer*innen haben Erfahrung mit SAS, daher sollte mMn auch diese Software vorhanden sein.
144	Ich bezweifele, dass sich mit R un PYTHON Datenmenge dieser Größenordnung performant bearbeiten lassen
145	passt ;- ) siehe Kommentar zur 1ten Frage,
173	Ich würde mich langfristig explizit gegen proprietäre Software aussprechen. Evtl. braucht man aber eine Übergangszeit
182	Wir nutzen in unserer Forschungsgruppe weder R noch Python

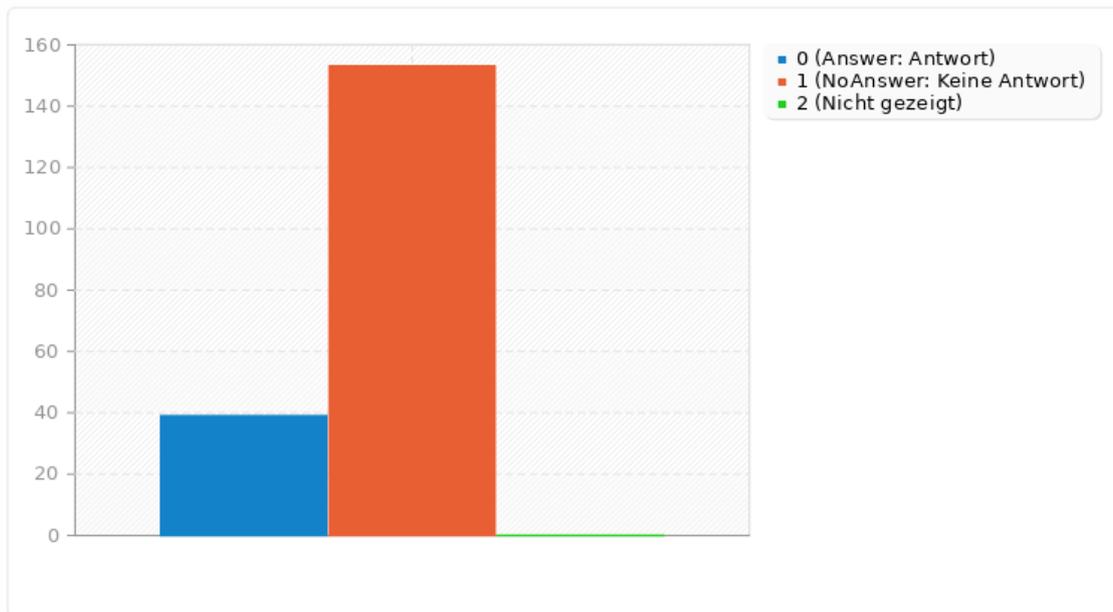
- 
- 184 Ich finde SPSS (oder Stata) in zweierlei Hinsicht vorteilhaft gegenüber den bisher geplanten Optionen: 1. Die Datenansicht gibt einen guten Überblick über vorhandene Variablen und ihre Eigenschaften. Das hat mir ungemein geholfen, die Daten besser kennenzulernen. Eine separate Datensatzbeschreibung ist super und notwendig, aber ein Blick in die Daten mit direkt verfügbaren Metadaten bietet aus meiner Sicht hierzu einen Mehrwert. 2. Forschende, die noch nicht so sicher mit Syntax bzw. Programmiersprache sind, können eine Eingabemaske nutzen und darüber ihre Syntax erstellen und so ihre Analysen dokumentieren. Auch finde ich - so viele Vorteile R bietet - dass es deutlich länger braucht, diese Sprache zu lernen im Vergleich zu SPSS oder Stata. Daher bietet aus meiner Sicht eine Ergänzung von R um SPSS (oder Stata) einen deutlichen Mehrwert für die Nutzerfreundlichkeit und Nutzbarkeit der FDZ-Daten. Auch die FDZ der statistischen Landesämter und GESIS als die größte europäische Infrastruktureinrichtung der Sozialwissenschaften ermöglichen die Anwendung von SPSS (und/oder Stata).
- 185 SAS kann Daten deutlich schneller verarbeiten, das ist vor allem für die Datenaufbereitung und -plausibilisierung interessant. Je nachdem, wie die Daten vorliegen werden, muss SAS aber nicht unbedingt genutzt werden. Teilweise gibt es auch in R sehr schnelle Verarbeitungsmöglichkeiten. Verschiedene Algorithmen sind in verschiedenen Softwares besser implementiert. Beispielsweise ist die aktuell beste Implementierung von Entropy Balancing in Stata zu finden.
- 199 Es ist nicht nur die Umstellung auf eine für mich ungewohnte Software. Software unterscheidet sich in den Paketen und Libraries, so dass für eine Fragestellung die eine Software und für eine andere Fragestellung eine andere Software geeignet sein mag.
- 208 Aus meiner Sicht wäre SAS vor allem für die Datenaufbereitung zusätzlich hilfreich.
- 209 SAS ist die standard auswertungssoftware für zulassungsanträge an die FDA und EMEA. SPSS ist vermutlich das langlebigste der grossen Statistikpakete, seit einigen Versionen ein IBM Produkt
- 215 Viele eigene Auswertungsskripte mit universellem Charakter sind in SAS vorhanden, die liessen sich ohne SAS nicht nutzen. Wie gut R mit extrem grossen Dateien klar kommt, könnte problematisch werden.
- 217 Was bedeutet das? Kann man die Daten dann in STATA transferieren? Und werden einem dann MitarbeiterInnen an die Seite gestellt?
- 229 R und Python werden in Gesundheitsämtern noch nicht standardmäßig genutzt, u.a. aus Datenschutzgründen. Eine grundsätzliche Unbedenklichkeitserklärung hinsichtlich des Datenschutzes bezogen auf R wäre sehr hilfreich, um lokale Datenschützer\*innen in den Gesundheitsämtern zu unterstützen. Und es braucht Schulungs-/Qualifizierungsangebote für beide Programme für die Mitarbeiter\*innen im ÖGD
- 236 Mir reicht das durchaus, sollte jedoch in Zukunft Zulassungsrelevante Studien laufen, kann es sein das SAS in diesem Kontext nötig werden wird. Daher würde ich dies im Aufbau mit denken und evtl. direkt mit implementieren, auch wenn es zunächst nicht genutzt wird. Ich kenne mich nicht gut aus, aber ein hinzufügen im Nachhinein ist immer komplizierter als es gleich mit zu implementieren.
- 266 Ich kann mir gut vorstellen, dass bei einem Fokus auf R und Python viele außen vor gelassen werden, welche nicht so technikaffin sind, aber trotzdem gerne in diesem Bereich etwas auswerten würden. SPSS fängt solche Personen auf, weil es einfach zu bedienen ist und sehr leicht ein passendes Skript aufbereitet werden kann.

---

## Zusammenfassung für G01Q02

Falls Sie dazu Kommentare oder Fragen haben, formulieren Sie diese bitte im nachfolgenden Freitextfeld:

---



## Zusammenfassung für G02Q01 [Welche R-Pakete würden Sie in Ihrem Analyseraum benötigen?]

Für alle Nutzerinnen und Nutzer von R oder Python:

Antwort	Anzahl	Prozent
Antwort	75	39.06%
Keine Antwort	117	60.94%
Nicht gezeigt	0	0.00%

ID	Antwort
9	MASS, boot, dplyr
13	Ich benutze jede Menge R-packages; auch selbst erstellte. Z. Zt. benutze ich lokal 561 unterschiedliche packages die 2.3GB Raum einnehmen. Das wird sehr schwierig eine Auswahl vorzunehmen, die vielen Usern gerecht wird. Aber ist es überhaupt nötig zentral verwaltete packages zu pflegen? Warum überläßt man das nicht dem User in seinem Verzeichnis die für ihn nötigen packages zu installieren? Ich benutze z. B. den high-performance-cluster meiner Uni und dort werden die aktuellen Versionen von R bereitgehalten. Aber die packages pflegt jeder user in seinem home-directory.
18	tidyverse, data.table, nlme, lme4, Hmisc, magrittr, ggplot2, mice, coin, rstatix, survival; sicher auch noch weitere
19	Weightit, MatchIt, Lubridate, Tidyverse, glue, janitor, datawizard, performance, cobalt, flextable, officer, glmmTMB, lme4, survery, sandwich, boot
24	Kann ich aktuell nicht abschätzen
27	- mediation - tidyverse - spgwr - tmap - spatialreg - foreign - ggplot2
31	R müsste ich erst mal lernen, ist aber ok, da eh schon lange auf der to do Liste. Daher noch keine Aussage möglich.
32	tidyr, dplyr, ggplot2, lubridate, epiR, nlme
33	möglichst viele für Regressionen und Zeitreihenbewertung
35	dplyr, tidyr, lubridate, TraMineR, stringr, ggplot2, ggseqplot, ggthemes, plaetteer, cowplot, svglite, cluster, psych, summarytools
39	Das hängt von der Analyse ab und der zum Zeitpunkt der Analyse verfügbaren und verwendbaren Paketen.
41	Über 50 Packages im Einsatz
43	tidyverse, ggplot2, tmap, sf, sp, terra, forecast, survival, nlme  (Frage: Wird es möglich sein, im laufenden Betrieb die Installation noch fehlender Pakete anzufordern?)
47	R-Studio
55	keine Erfahrung
59	Die lässt sich nicht allgemein beantworten, da es 21091 Pakete allein auf dem CRAN-Servern gibt, von denen man je nach Aufgabe mal das eine und mal das andere Paket nutzt. Falls die Frage auf einem KERN von Pakten bezieht, die man generell benutzt fallen mir spontan die Pakete des Core-tidyverse ein, also ggplot2, dplyr, tidyr, purrr, stringr, readr, tibble, forcats ein. Dazu noch data.table, DBI. Dazu dann jeweils die Pakete die man für die Analyse verwendet.
63	tidyverse, summarytools, gt, gtsummary, writeexl, openxlsx, survey, ggpattern, jtools, scales, broom, plyr, (todor)
65	dplyr, dbplyr, data.table, dtplyr, fst, ICD10gm, ggplot2, tidyr, stringr, lubridate, MatchIt, cobalt, purr, survival, kableExtra, markdown, survminer, fmsv, sandwich, lmttest
67	Langzeitanalysen
72	Wurden bereits an das FDZ übermittelt.

75	data.table, dplyr, etc.
77	Ausführliche Liste wurde bereits direkt mit dem FDZ geteilt.
80	tidyverse, psych, , ggpubr, survival, survminer, interactions, lm.beta
81	DBI, odbc, dbplyr, tidyverse, magrittr, skimr, lubridate, hms, haven, questionr, modelsummary, estimatr, PanelMatch, ebal, WeightIt, cobalt, knitr, openxlsx, xlsx, plm, did, DRDID, pstest, DIDmultiplegt, fixest, did2s, didimputation, bacondcomp, synth, tidysynth, ivreg, boot, sandwich, fwildclusterboot, meboot, ranger, SuperLearner, glmnet, xgboost, caret, ipred, e1071, survival, rms, NADA, coxrobust, Hmisc, mice, zoo, ggplot2, comorbidity, stringr, grf, gridExtra, betareg, hdm, AER, scales, moments, margins, prediction, stargazer, lmtree, erer, DoubleML
83	schon über die Umfrage des FDZ direkt mitgeteilt
85	tidyverse
87	library(readxl) library(ggplot2) library(writexl) library(sf) library(data.table) options(datatable.showProgress = T) library(tidyverse) library(ggpubr) library(scales) library(stringr) library(gridExtra) library(shiny) library(leaflet) library(leaflet.extras) library(dplyr) library(tidytext) library(psych) library(spdep)  library(ggspatial) library(rgeos) library(ggrepel) library(viridis) library(RColorBrewer) library(ggplot2) library(geodata) library(raster)
93	Das hängt von den konkreten Analysen ab, abgesehen von einigen Standard-Paketen. M.E. ist es notwendig, dass R-Pakete während des Betriebs/der Analysen aus der R-Software heraus heruntergeladen und installiert werden können. Ich habe keinen Überblick, wie viele (allgemein genutzte) R-Pakete es gibt, aber diese alle vorzuhalten scheint mir schwierig.
95	dplyr, survival, lme4, pROC und je nach Anwendungsfall weitere Pakete.
96	Das kommt leider ganz auf die Fragestellung der Datenauswertung an. Es sollten die bekannten und oft genutzten Pakete für Basis Datenmanagement, -auswertung, -modellierung und graphische Darstellung (z.B. dplyr, glm, ggplot2, tidyverse) vorhanden sein. Aber es sollte auch die Möglichkeit geben spezielle Pakete (z.B. comorbidity, pccc) installieren zu können.
98	lubridate, dplyr, comorbidity, weights, boot, ICD10gm, WeightIt, readxl, ggplot2
100	tidyverse, lubridate, summarytools, openxlsx, ICD10gm
101	data.table, tidyverse (komplett), glmnet, survival, formatR, foreach, rpart, parallel, splines, spatial, stats4, nlme, Matrix, MASS, lattice, grid, compiler, boot, cluster, class, codetools, XML, xml2, tydiselect, utf8, vctrs, stringi, rJava, rlang, reshape2, rematch, readr, Rcpp, pbkrtest, nnet, lubridate, knitr, devtools, styler, dplyr, dbplyr, purrr, CDMConnector, CohortSurvival, KMSurv, cowplot, ggplot2, corrplot, contTimeCausal, tmlr, autoCovariateSelection
105	data.table, tidyverse, glmnet, survival, formatR, foreach, rpart, parallel, splines, spatial, stats4, nlme, Matrix, MASS, lattice, grid, compiler, boot, cluster, class, codetools, XML, xml2, tydiselect, utf8, vctrs, stringi, rJava, rlang, reshape2, rematch, readr, Rcpp, pbkrtest, nnet, lubridate, knitr, devtools, styler, dplyr, dbplyr, purrr, CDMConnector, CohortSurvival, KMSurv, cowplot, ggplot2, corrplot, contTimeCausal, tmlr, autoCovariateSelection
110	Das sollte offen bleiben.
120	data.table, tidyverse (komplett), glmnet, survival, formatR, foreach, rpart, parallel, splines, spatial, stats4, nlme, Matrix, MASS, lattice, grid, compiler, boot, cluster, class, codetools,

	XML, xml2, tidyselect, utf8, vctrs, stringi, rJava, rlang, reshape2, rematch, readr, Rcpp, pbkrtest, nnet, lubridate, knitr, devtools, styler, dplyr, dbplyr, purrr, CDMConnector, CohortSurvival, KMSurv, cowplot, ggplot2, corrplot, contTimeCausal, tmlc, autoCovariateSelection
123	ggplot, ggplot2, ggmap, psych, factoextra, tidyverse, dplyr, caret, stringr Welche noch ist schwer zu sagen, analysebezogen sicherlich mehr
125	UNZÄHLIGE!!!
130	Projektabhängig z.B. nlme, rpart, survival, Epi, epitools,...
132	Alle! Bei R gilt doch das Prinzip "Viele Wege führen nach Rom" und man kann zu Beginn einer Analyse gar nicht so genau wissen, was man letztlich alles so braucht um irgendein Problem zu lösen. Dass man benötigte Pakete einfach über Befehl installieren kann und damit den Funktionsumfang von R ständig ausbaut, ist einer der zentralen Vorteile des Programms. Es wäre also sehr schön, wenn dies auch in der Terminal-Server Umgebung des FDZs möglich wäre. Eine manuelle Installation wird schnell zu Frustration führen, wenn irgendwelche dependencies fehlen. Minimal sind dabei aber auf jeden Fall die "Standardpakete" "dplyr", "tidyverse", "ggplot2" usw.. Für Datumsangaben braucht es aber auch Pakete wie "lubridate". Zum Einlesen/Speichern vermutlich "xlsx", "readxl" o. ä.. Für die Aufbereitung von Diagnosen "ICD10gm" und "comorbidity". Für Analysen braucht es die ganzen Pakete, um Regressionen rechnen zu können.
133	in unseren Projekten verwenden wir zahlreiche R-Pakete und programmieren auch selbst neue Pakete und Funktionalitäten. Nicht nur unsere Kompetenzen und Erfahrungen, sondern auch die Qualitätssicherung und Dokumentationsstandards erfordern ein sehr breites Spektrum von Analysemöglichkeiten in der sicheren Umgebung.
145	Dies ist abschließend nicht zu beantworten. Ein CRAN/GitHub Zugang wäre sinnvoll. Oder eine unkomplizierte individuelle Ergänzung der Pakete.
146	tidyverse bibliotheken wie tidyr und dplyr, stringr, janitor, Hmisc, ggplot2, igraph, DT
147	Dplyr, tidyverse, purrr, glue ...
149	v.a. tidyverse
162	plyr dplyr tidyr lubridate data.table gmodels janitor ggplot2 survival prodim gtsummary ggsci comorbidity stringr purrr glue readxl knitr kableExtra tidyr dbplyr tidyverse bit64
163	Ich arbeite regelmäßig mit einer (geschätzt) dreistelligen Anzahl von R-Paketen (inklusive abhängige Pakete). Bei mir lokal installiert ist eine vierstellige Anzahl von R-Paketen (inklusive abhängige R-Pakete). Bevorzugt - und um die Interessen aller Nutzerinnen und Nutzer zu befriedigen - sollten alle R-Pakete installiert sein und zwei mal im Jahr geupdated werden. Ansonsten Gewährleistung zeitnaher Nachinstallation.
172	tidyverse, dbplyr, DBI, quarto, lmer, plm und andere
173	Vorallem: "broom", "conflicted", "dbplyr", "dplyr", "dtplyr", "forcats", "ggplot2", "haven", "hms", "httr", "lubridate", "magrittr", "modelr", "pillar", "purrr", "rvest", "stringr", "tibble", "tidyr", "survival", "gtsummary", "gt", "glue", "DiagrammeR", "labelled", "cowplot", "performance", "msm"
175	MASS psych dplyr ggplot2

	lubridate
	epitools
	stringr
	tidyverse
176	MAGMA, MatchIt, survival, nett,
178	tidyverse; tidytext; skimr; janitor; fastDummies; anytime; psych; gtsummary; effsize;
183	am besten ein lokales Repo einrichten, dann muss man nicht jetzt schon entscheiden, was man in einigen Jahren braucht ;-)
185	Auswahl: AsioHeaders, askpass, assertthat, backports, base64enc, BH, BiocGenerics, BiocManager, BiocVersion, bit, bit64, bitops, blob, brew, brio, broom, bslib, cachem, callr, cellranger, checkmate, chromote, chron, classInt, cli, clipr, collapsibleTree, colorspace, colourpicker, commonmark, cowplot, cpp11, crayon, credentials, crosstalk, curl, data.table, data.tree, DBI, dbplyr, desc, devtools, DiagrammeR, DiagrammeRsvg, diffobj, digest, docxtractr, downloader, dplyr, dtplyr, e1071, ellipsis, evaluate, fansI, farver,

---

fastmap,  
filesstrings,  
flextable,  
forcats,  
Formula,  
fs,  
fuzzyjoin,  
gargle,  
gdtools,  
generics,  
geosphere,  
gert,  
ggplot2,  
ggplotify,  
ggridges,  
ggthemes,  
gh,  
gitcreds,  
glue,  
googledrive,  
googlesheets4,  
gridExtra,  
gridGraphics,  
gsubfn,  
gtable,  
haven,  
highr,  
Hmisc,  
hms,  
htmlTable,  
htmltools,  
htmlwidgets,  
httpuv,  
httr,  
ids,  
igraph,  
ini,  
installr,  
IRanges,  
isoband,  
jpeg,  
jquerylib,  
jsonlite,  
knitr,  
labeling,  
later,  
latticeExtra,  
lazyeval,  
lifecycle,  
listviewer,  
lme4,  
lubridate,  
magrittr,  
markdown,  
memoise,  
mime,  
miniUI,  
minqa,  
modelr,  
munsell,  
networkD3,  
nloptr,  
officer,  
openssl,  
osmdata,

---

pander,  
pdftools,  
pillar,  
pkgbuild,  
pkgconfig,  
pkgload,  
plogr,  
plotly,  
plyr,  
png,  
praise,  
prettyunits,  
processx,  
progress,  
promises,  
proto,  
proxy,  
ps,  
purrr,  
pwr,  
pythonic,  
qpdf,  
R6,  
rappdirs,  
rcmdcheck,  
RColorBrewer,  
Rcpp,  
RcppEigen,  
RCurl,  
readr,  
readxl,  
rematch,  
rematch2,  
remotes,  
reprex,  
rJava,  
rjson,  
RJSONIO,  
rlang,  
rmarkdown,  
rmdexamples,  
roxygen2,  
rprojroot,  
RSQLite,  
rstudioapi,  
rvcheck,  
rversions,  
rvest,  
S4Vectors,  
sass,  
scales,  
selectr,  
servr,  
sessioninfo,  
sf,  
shiny,  
shinyDND,  
shinyjs,  
slidex,  
sourcetools,  
sp,  
sqldf,  
strex,  
stringdist,  
stringi,

---

stringr,  
sys,  
systemfonts,  
testthat,  
tibble,  
tidyr,  
tidyselect,  
tidyverse,  
tinytex,  
translations,  
units,  
usethis,  
utf8,  
uuid,  
vctrs,  
viridis,  
viridisLite,  
visNetwork,  
waldo,  
webshot,  
webshot2,  
websocket,  
whisker,  
withr,  
xaringan,  
xfun,  
xlsx,  
xlsxjars,  
xml2,  
xopen,  
xtable,  
yaml,  
zip

- 188 tidyverse, gtsummary, arrow, lubridate, knitr [und sehr wichtig: Quarto sollte installiert sein]
- 189 tidyverse packages, data table, lubridate
- 190 Das ist in der Regel projektspezifisch und für mich nicht vorab zu bestimmen.
- 192 data.table, tableone, tidyverse, glmnet, survival, formatR, foreach, rpart, parallel, splines, spatial, stats4, nlme, lme4, mgcv, pammttools, randomForestSRC, lcmm, Hmisc, rms, Matrix, MASS, lattice, grid, compiler, boot, cluster, class, codetools, XML, xml2, tidyselect, utf8, vctrs, stringi, rJava, rlang, reshape2, rematch, readr, Rcpp, pbkrtest, nnet, lubridate, knitr, devtools, styler, dplyr, dbplyr, purrr, CDMConnector, CohortSurvival, KMSurv, cowplot, ggplot2, corrplot, contTimeCausal, tmle, autoCovariateSelection, FactoMineR
- 201 So allgemein schwer zu beantworten, da das sehr vom jeweiligen Projekt abhängt. Zur Datenaufbereitung auf jeden Fall das package data.table. Bestenfalls besteht die Möglichkeit für eigene Projekte packages zu installieren oder installieren zu lassen.
- 204 u. a. dplyr, ggplot2, tidyr, tidyverse, haven, car, data.table, readr
- 207 igraph
- 208 Dplyr  
Tidyverse  
tidymodels
- 211 Es sollte möglichst die gesamte Bandbreite von CRAN zur Verfügung stehen  
tidyverse, survival, survminer, table1, library(dbplyr)  
library(lubridate)  
library(scales)  
library(plotly)  
library(ggh4x)  
library(cowplot)  
library(skimr)  
library(table1)  
library(flextable)  
library(comorbidity)  
library(survival)  
library(ggsurvfit)  
library(survminer)

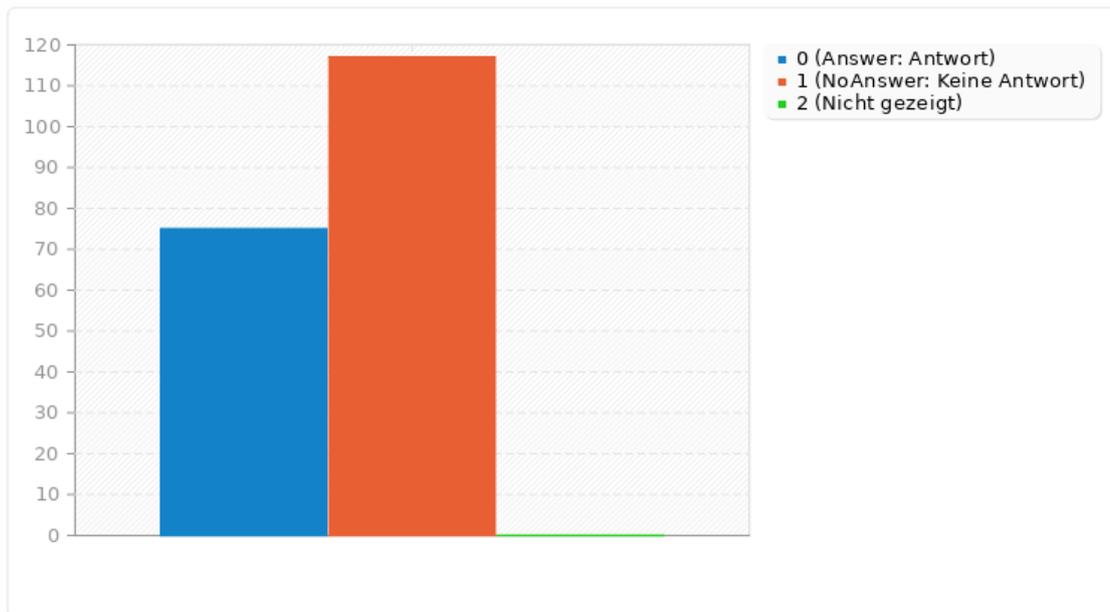
	library(UpSetR) library(fastDummies)
214	dbplyr, tidyverse, caret, UpSetR, table1, flextable, comorbidity, lubridate, survival, ggsurvfit, survminer, fastDummies, here, comorbidity, epiR
215	-
236	Es gibt sehr generelle Pakete und sehr Projekt bezogene Pakete. Das ist eine Liste meiner am meisten verwendeten Pakete: "askpass", "base64enc", "bookdown", "boot", "callr", "cli", "codebook", "colorspace", "cpp11", "crayon", "curl", "DiagrammeR", "digest", "dmetar", "dplyr", "dplyrAssist", "ellipsis", "esquisse", "evaluate", "fans", "farver", "flextable", "forcats", "foreign", "fs", "future", "ggplot2", "ggtext", "glue", "gt", "haven", "highr", "htmltools", "httr", "irr", "irr", "irr", "janitor", "jsonlite", "kableExtra", "KappaGUI", "knitr", "labeling", "labelled", "lifecycle", "lubridate", "magrittr", "markdown", "metadat", "metafor", "metafor", "mime", "munsell", "netmeta", "netmeta", "officedown", "officer", "openssl", "openxlsx", "pillar", "pkgconfig", "processx", "ps", "psy", "psych", "R6", "RColorBrewer", "Rcpp", "readxl", "rio", "rlang", "rlang", "rmarkdown", "rstudioapi", "rsvg", "rsvg", "rvest", "scales", "selectr", "snakecase", "stringi", "stringr", "svglite", "sys", "systemfonts", "tibble", "tidyverse", "tinytex", "utf8", "vcd", "vctrs", "viridisLite", "webshot", "xfun", "xml2", "yaml",
238	Survey Dplyr ggplot2 summarytools
239	tidyverse, mosaic, dplyr, stringr, lubridate, tidyr, ggplot
245	tidyverse, readxl, readr, summarytools, dplyr, markdown, shiny, ggplot2, tidyr
251	(1) tidyverse (alle darin enthaltenen Pakete, insbesondere dplyr, tidyr, purrr, stringr, forcats, ggplot2); (2) survey; (3) broom; (4) MASS; (5) summarytools; (6) gt; (7) gtsummary
252	Vor allem die Pakete aus dem Tidyverse. Extrem wichtig wäre aber, eine Möglichkeit zu schaffen, unkompliziert und kurzfristig Pakete ergänzen (lassen) zu können, die noch nicht zur Verfügung stehen. Bei der Arbeit mit R wäre es eine extreme Einschränkung, wenn nur eine feste Auswahl an Paketen zur Verfügung stünde. Nicht selten wird erst während der Feinkonzeption der Analysen deutlich, dass man weitere Pakete benötigt. Zudem ist die Entwicklung rasant: Auf neue Pakete, die die Arbeit erleichtern, die aber eben jetzt noch nicht bekannt sind, möchte man ggf. nicht verzichten. Eine Ergänzung der zur Verfügung stehenden Pakete muss mit kurzen Vorlaufzeiten möglich sein. Bitte auch unbedingt berücksichtigen, dass die zur Auswahl stehenden Pakete immer auf dem neuesten Stand gehalten werden (regelmäßige Updates).
260	base, stats, dplyr, stringr, utils, data.table, ggpubr, splines, survival, cmprsk, survminer, Rcpp, grDevices, ggplot, tidyr, coxph,
262	rms & predtools - Für die Entwicklung und Validierung von Prediction Modellen; tidyverse inkl. dplyr - Datenmanagement; lubridate - Sehr hilfreich zum arbeiten mit verschiedenen Datumsformaten
268	regression; Leaflet; tidyverse
270	tidyverse, dplyr, ggplot2, kable, psych, lavaan usw.

---

Zusammenfassung für G02Q01 [Welche R-Pakete würden Sie in Ihrem Analyseraum benötigen?]

Für alle Nutzerinnen und Nutzer von R oder Python:

---



## Zusammenfassung für G02Q01 [Welche Python-Pakete würden Sie in Ihrem Analyseraum benötigen?]

Für alle Nutzerinnen und Nutzer von R oder Python:

Antwort	Anzahl	Prozent
Antwort	25	13.02%
Keine Antwort	167	86.98%
Nicht gezeigt	0	0.00%

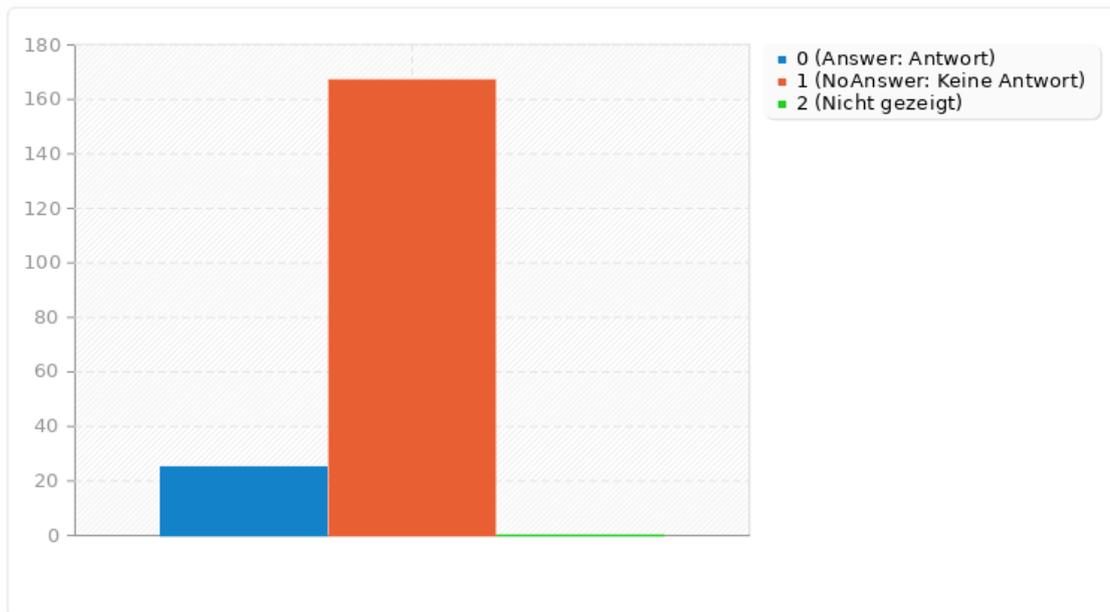
ID	Antwort
14	Pandas, Numpy, Seaborn, Matplotlib, SciPy, scikit-learn, TensorFlow, Pendulum
39	Das hängt von der Analyse ab und der zum Zeitpunkt der Analyse verfügbaren und verwendbaren Paketen.
41	Über 20 Packages im Einsatz
55	keine Erfahrung
72	Wurden bereits an das FDZ übermittelt.
77	Ausführliche Liste wurde bereits direkt mit dem FDZ geteilt.
101	pyreadstat, scikit-learn, scipy, numpy, pandas, tqdm, ipykernel, seaborn, matplotlib, openpyxl, pytest, xgboost, plotnine, rpy2, econml, vulture, ruff, black, shap, psmypy, sqlalchemy und ein database connector wie psycopg2 wenn PostgreSQL verfügbar ist, dowhy, causalmml, statsmodels
105	pyreadstat, scikit-learn, scipy, numpy, pandas, tqdm, ipykernel, seaborn, matplotlib, openpyxl, pytest, xgboost, plotnine, rpy2, econml, vulture, ruff, black, shap, psmypy, sqlalchemy, dowhy, causalmml, statsmodels
107	Pandas, Numpy, Matplotlib, Seaborn, SciPy, SciKit-Learn
108	pandas, numpy, matplotlib, seaborn, scipy.stats, statistics, sklearn, math, statsmodel
120	pyreadstat, scikit-learn, scipy, numpy, pandas, tqdm, ipykernel, seaborn, matplotlib, openpyxl, pytest, xgboost, plotnine, rpy2, econml, vulture, ruff, black, shap, psmypy, sqlalchemy und ein database connector wie psycopg2 wenn PostgreSQL verfügbar ist, dowhy, causalmml, statsmodels
132	Grundproblem gleich
133	in unseren Projekten verwenden wir zahlreiche Python-Programme und programmieren auch selbst neue Funktionalitäten. Nicht nur unsere Kompetenzen und Erfahrungen, sondern auch die Qualitätssicherung und Dokumentationsstandards erfordern ein sehr breites Spektrum von Analysemöglichkeiten in der sicheren Umgebung.
145	Dies ist abschließend nicht zu beantworten. Ein CRAN/GitHub Zugang wäre sinnvoll. Oder eine unkomplizierte individuelle Ergänzung der Pakete.
149	tensorflow, scikit-learn, numpy, pandas u.ä.
183	s.o.
188	pandas, siuba, scikit-learn, plotnine, jupyter
189	Pandas
192	pyreadstat, scikit-learn, scipy, numpy, pandas, tqdm, ipykernel, seaborn, matplotlib, openpyxl, pytest, xgboost, plotnine, rpy2, econml, vulture, ruff, black, shap, psmypy, dowhy, causalmml, statsmodels, catboost, umap.learn, hdbscan, pingouin
207	igraph
214	pandas, plotly, pm4py
215	-
257	pandas, numpy, matplotlib, sklearn, tensorflow, statistics
262	Aktuell arbeite ich nicht mit Python
270	-

---

Zusammenfassung für G02Q01 [Welche Python-Pakete würden Sie in Ihrem  
Analyseraum benötigen?]

Für alle Nutzerinnen und Nutzer von R oder Python:

---



---

## Zusammenfassung für G03Q01

Ich arbeite mit GKV-Routinedaten:

---

Antwort	Anzahl	Prozent
Bisher noch nie (AO01)	18	9.38%
Seit weniger als einem Jahr (AO02)	16	8.33%
Seit ein bis fünf Jahren (AO03)	71	36.98%
Länger als fünf Jahre (AO04)	87	45.31%
Keine Antwort	0	0.00%
Nicht gezeigt	0	0.00%

---

## Zusammenfassung für G03Q01

Ich arbeite mit GKV-Routinedaten:

---

