# kleine Einführung in R

Dies ist eine kleine Einführung in die Arbeit mit R aus einem Seminar. Es ist noch nicht final für die Veröffentlichung bearbeitet / korrigiert. Sie können es aber schon als Einstieg in R nutzen.

# 1. Was ist R?

R ist eine Open-Source-Software (also eine Software, deren Quellcode frei verfügbar/öffentlich ist) – und eine Programmiersprache, die hauptsächlich für die Statistik benutzt wird. R kann man sowohl für die statistische Datenanalyse benutzen als auch zu der Erstellung von Grafiken.

### 1.1 Kleiner Einblick in die Geschichte von R

Mitte der 70er Jahre – also vor ca. 50 Jahren wurde die Programmiersprache S von John Chambers et al. in den USA entwickelt, um <u>statistische Simulationen und Grafiken</u> zu erstellen. **Ende der 80er Jahre** wurde daraus S-PLUS für den kommerziellen Gebrauch entwickelt. Seither wird die Software zur <u>Datenanalyse</u> und zur <u>statistischen Modellierung</u> genutzt. R ist nichts Weiteres als eine Weiterentwicklung der Programmiersprache, die auf S aufbaut und wurde Anfang der 90er Jahre initiiert. Inzwischen ist R ein weit **verbreitetes Tool** und wird in vielen Bereich benutzt.

Die Informationen sind von folgender Quelle:

https://tu-dresden.de/gsw/phil/iso/mes/ressourcen/dateien/prof/lehre/freieS/Dateien/Einfuehrung\_R.pdf ?lang=de

### 1.2 R im Vergleich zu anderen Statistikprogrammen

#### R

Ein großes Argument für R ist, dass R und R Studio **kostenlos** sind. Egal ob man R als einzelne\*r Student\*in für die Uni benutzt, oder in einem nationalen Statistikunternehmen arbeitet – R kann kostenlos runtergeladen und genutzt werden. Es gibt auch **keine Upgrade Version**, die einen dazu "anstiftet", eine kostenpflichtige Version zu kaufen.

*Vorteilhaft* ist, dass du zu R sehr viele Anleitungen findest. Egal ob du Anfänger\*in bist, oder Fortgeschrittene\*r, meistens findest du eine Lösung zu deinem Problem. Zudem ist es egal, auf welcher Software du R runterladen willst: MacOS, Windows, Linux o.Ä.!

Wenn du keine vereinzelten Kenntnisse beim Coden mitbringst, wirst du anfangs wahrscheinlich etwas verwirrt und überfordert sein, allerdings kommt das mit der Zeit. R ist wie eine gesprochene Sprache: Du musst sie regelmäßig üben und wenn du das Programm eine Zeit lang nicht benutzt hast, kann es auch etwas dauern, wieder reinzukommen. Allerdings gewöhnst du dich recht schnell wieder daran.

*Unvorteilhaft:* Bei sehr umfangreichen Rechenoperationen oder Datensätzen kann das Auswerten etwas länger dauern, v.a. bei Windows und/oder, wenn das Programm nicht das Neueste Update hat. Aber v.a. für Basics an der Uni ist das Programm gut geeignet, um reinzukommen.

### Stata

Stata habe ich persönlich noch nicht benutzt, aber das ist auch ein gängiges Statistikprogramm.

*Vorteilhaft:* auch dieses Programm ist über diverse Betriebssysteme verfügbar, egal ob MacOS, Windows oder Unix-Betriebssysteme (Linux gehört z.B. dazu).

Unvorteilhaft: Stata ist für niemanden kostenfrei. Das Unternehmen bietet unterschiedliche "Packages" an, die von ca. 57€ für Studierende (für 6 Monate) zu ca. 1090€ für einen nicht akademischen Gebrauch (für 1 Jahr) reichen. [In den Beschreibungen steht immer "ab", also kann es auch noch teurer werden]. Die Lizenz muss also immer wieder erneuert werden. Aber: wer sich unsicher ist, ob Stata das passende Programm ist: es gibt eine 30-tägige Testversion, die kostenfrei ist.

Die Informationen sind von folgender Quelle: https://stata-germany.com/?gad\_source=1&gclid=Cj0KCQjw4cS-BhDGARIsABg4\_J2KjN-41gcEo7-jxPrt8 gBGTcaXLUytI Mmwnm6-KwZeZcgbFipwu8aAvrIEALw wcB#!/Preise

#### SPSS

Auch SPSS habe ich persönlich noch nicht benutzt, dennoch kenne ich einige, die damit im universitären Kontext schon gearbeitet haben.

*Vorteilhaft:* Vor allem Menschen, die sich nicht gerne mit dem Schreiben von Codes beschäftigen, können von diesem Programm profitieren. SPSS bietet eine Drag- and-Drop- Funktionalität und kann die **Daten analysieren, ohne Codes schreiben zu müssen**, was z.B. bei R auf jeden Fall erfordert wird. Der Output kann nach persönlichen Bedürfnissen optimiert werden und geht meistens auch recht schnell. Die Syntax kann z.B. mit R oder Python ergänzt werden.

Unvorteilhaft: SPSS ist auch für jeden kostenpflichtig. Das **Basisabonnement beläuft sich auf 1188€** pro Jahr, und kann noch mit 3 Add-Ons erweitert werden, die jeweils auch ca. 950€ jährlich kosten. Studierende hingegen können SPSS bei unterschiedlichen Anbietern erwerben, und so unterschiedliche Lizenzen erwerben – so kann man für **29€ eine sechsmonatige Lizenz erwerben und für 143€ eine dreijährige Lizenz**. (Es gibt auch teurere Versionen für Studierende, das hängt davon ab, welche zusätzlichen Tools man braucht [] Hier kann man das nachschauen: https://studentdiscounts.com/.

Die Informationen sind von folgender Quelle: https://www.ibm.com/de-de/products/spss-statistics

**Persönliche Begründung unserer Wahl:** In unserem Seminar haben wir uns alle geeinigt, R zu nutzen. Einige kannten schon den Umgang damit und hatten das Programm bereits auf dem Laptop installiert. Zudem brauchte niemand das Programm zu kaufen. Im Unikontext gibt es oft die Möglichkeit, SPSS zu nutzen, das muss dann über die PC's in der Universität laufen – mit Laptop ist man aber flexibler und wir konnten einen Raum nutzen, der nicht unbedingt andere PC's erfordert hat.

### 2. Grundlagen von R

### 2.1 Installation von R

Damit du R nutzen kannst, musst du das Programm natürlich erstmal runterladen. Am besten lädst du zusätzlich RStudio runter. R funktioniert zwar auch ohne RStudio, allerdings vereinfacht RStudio die Arbeit mit R erheblich. Aber was ist der Unterschied? R ist die Programmiersprache für statistische Analysen (wie vorhin schon beschrieben). Wenn ihr nachher arbeitet, arbeitet ihr also mit R. RStudio hingegen ist eine Entwicklungsumgebung für R.

Der erste Schritt ist natürlich der Download von R:

Hierfür musst du auf folgende Seite: https://cran.r-project.org/ Auf der Seite werden euch Links angezeigt, auf die ihr zum Runterladen greifen könnt, je nachdem mit welchem Betriebssystem euer Gerät funktioniert. Bei mir ist das MacOS.



Wenn du da drauf klickst, werden dir unterschiedliche Informationen zu verschiedenen Versionen gegeben. Je nachdem wie neu oder alt dein Macbook ist, hat es nur Zugriff auf bestimmte MacOs Versionen und demnach musst du dich orientieren. Hier kannst du zwischen zwei Versionen unterscheiden.



[Exkurs: Dafür muss dir bewusst sein, welche Version du auf dem Laptop hast. Das findest du beim Macbook in den Systemeinstellungen raus, wenn du auf "Allgemein" drückst.]

Last update: 2025/04/22 lv-wikis-oeffentlich:boot2024:2.\_arbeiten\_mit\_r https://institut.soziologie.uni-freiburg.de/dokuwiki/doku.php?id=lv-wikis-oeffentlich:boot2024:2.\_arbeiten\_mit\_r 2025/04/22



Dann wählst du die passende Version und das Paket wird geladen – in der Regel geht das sehr schnell und braucht nur ein paar Sekunden. Wenn es runtergeladen ist, klickst du drauf und folgendes Fenster öffnet sich:



Dem folgst du nacheinander, die Schritte sind selbsterklärend. Wichtig: hier lädst du nur R runter!

Wenn du zusätzlich RStudio runterladen willst, dann kannst du die folgende Anleitung beachten: Zuerst öffnest du diese Webseite dafür: https://posit.co/download/rstudio-desktop/ und dann klickst du auf "Install R-Studio":



Auch das ist normalerweise recht schnell runtergeladen. Wenn du drauf klickst, öffnet sich folgendes Fenster. Zur Benutzung von R öffnest du einfach die App "RStudio".



Wahrscheinlich wirst du gefragt, ob du die App wirklich öffnen willst, dann klickst du einfach auf "öffnen". Dieses Nachfragen kann nervig erscheinen, das dient aber nur zu deinem Schutz und ist i.d.R. normal bei Apple.

#### 2.2 Der Aufbau von R

Wenn du R nun öffnest, wirst du zu Beginn vielleicht von all den Informationen und Eindrücken erschlagen. Deswegen arbeiten wir das jetzt Schritt für Schritt durch, damit du ein tiefgründiges Verständnis für das Programm aufbaust. Was du nach dem Öffnen zuerst siehst, ist das GUI: das Graphical User Interface und dieses besteht aus vier Bereichen:

I + No. all + Month American Control Address				10	
# (bestel - # Datelaneilled : Billder Detaker : Dellater	- 14 - I - I - I - I - I - I - I - I - I -	Revisement Matters Con	analises Tutarial		-
CONTRACTOR OF A	rates we have a brance of the	P CI IP reportingent 1	A state of the		-
		· · · · · · · · · · · · · · · · · · ·		10	
		The second second	The second secon		
		Contract of the	AND BELLET 7 VERSION		
		and a second sec	And a second sec		
		0 101, 101 pm	Large Humerts Limbe elements, 108.1 Mg		
			Rend nor (2-1) 4.30 4.94 4.08		
		NO.	Read one (213) 4.4587 4.8582 4.4383		
		ton par	Ranad Aut & BOD		
		Film, MARTYL	Renal rule & GEB		
		Plan, Unigers	Environment, History		
Editor		Peer, tracely	Read nue -4.8001		
			X68 (3.2) 4.47 5.25		
		1.7	Manad And (3-2) & MIGH K-10846		
		and come	NU.		
		spreating line	4.750608208382		
		Contract	HAR (3.1986) & 255 K-648 K-623 1.5067 (1.484)		
		1	196.		
		8800.007*	-8. LORDER THEOR		
		man.com/	-8.18H0H7H081		
		46/0.09/9	4.808844.80587		
		1,040	198		
		num. incoming many	200		
		Particular'	-0.6%40/L018000		
		person.com/d	-4.4%4ACD28008C		
		M.A.	Name and (\$110) 8-562 8-681 1-685		
11. Option 1	A brief i	11.1	1 ADACOUTES		_
anale Territal - Antigeneticity -	-0	the firs belage in	No Front Franklahe		-
- 8442		Clinetal @ Later			
		No.	Decopios	Territoria.	
zersion 6.4.3 (2001-02.20) = "Trophy East"		waters almost	Diote Darksone		
aprophi (c) was the it hander on rar Material case targeting		2 has	The billion Parlian	14.7	
store accession where we not		1 hours	Restorms Resulting Objects in Inspire Cash for B	1.1-18	-
In the settings and case with principle of the settings.		1 citra	Austions for Eleveloption	24.20	
IN THE MEXICAL TO ANTIDETIMATE IT, WHAT SAFARA CONTENTS.		- Ander	"Nucling Decape in Entit" Clusters Analysis Estenden/Heuseneuperst al.	2.1.4	
of "Transc" or "topoll" for distribution dealth.		(Martinette	Ends Analysis Tasis for 8	0.4.20	
		1 complete	The E-Complex Relates	44.3	
Nound, tangage aspent but numbry to an traitab focable		2 depends	The Information Participat	11.8.2	
to a callaboration project with many carterilations. Con	wole	- Renage	Rest Onto Streed by Workey, 75, 7457, 7407, 7607	1.8-18	- 0
er 'antribland)' for men informition and		2 graphics	The KG sphin Perkage	4.8.3	
charlow()" on how to other K or K postupes in publications.		d photos	The Ericiphus Rodom-analysigned for Estimate and Party	1.4.1	
en hierofi de seu deux. Dela V de sed las lada se		1 914	The Drafforgaliko Package	4.8.3	
and another the same among "shipping in bola.		turnin anti-	Renations for Kernel Tonacohing Repairting World & Jones (2005)	2.28-24	
as 'all' to mak #.		Lation .	Tratta Dogetica for K	1.11+	- 0
		I MARK	Repport Putertises, and Spherica for Hendlike and Hybry's NAVA	2.8.68	
pringers Looked from/ Mersi		- Materia	Ignores and Denne Matrix-Communand Helikola	1.9-0	
		2 million	Normal Nethods and theses	1.4.3	
PanderowsJaneta		- main	Rivad DRV Computation Rehabiliered Automatic Secondarias Saturation	1.1-1	
		i ulma	Simular and Northinger Mircell ("Reins Medals	14.187	
			Read Research Research and Read and State of the Read		

Fangen wir mit dem "Editor" an: hiermit wirst du wohl am meisten arbeiten. Du kannst ihn dir wie ein Bereich vorstellen, indem du dein Skript erstellst. Du schreibst hier also deinen Code. Hier kannst du von unterschiedlichen Funktionen Gebrauch machen. Wenn du auf das weiße "Blatt" mit Plus klickst, erstellst du eine neue "Seite", in der du ein neues Skript beginnen kannst.

<u>**</u> *							RS
0 - 00 @-	🖶 🔐 🚔 🗇 Go ta	flie/function	🖂 + Addins +				
Untitled1* ×	DaterSammi.Rmd ×	15511_de ×	Trst_clean ×	Trut50_clean ×	Trat =		
0 0 0	Source on Save   Q	p• 🗆 👘			- Run	50 0 0 IB	Source • 🔍

Mit dem Symbol daneben (R mit einem Plus) kannst du ein neues Projekt starten und zwischen unterschiedlichen Möglichkeiten wählen, je nachdem was du brauchst.

•	<b>A</b> .		file/function	🔄 + Addins +				R
0	Unreled1" ×	DaterSammLRmd ×	E5511_de ×	Trst_clean ×	Trat50_clean ×	Trat ×		-0
	100	Source on Save   Q	<b>∦•</b>   ⊟			-+ Run	₩ 0-0.	_in Source + ≥

Auf das Symbol mit dem Ordner kannst du klicken, wenn du vorhandene Dateien auf deinem Laptop in R laden magst.

•••							R
o - 9 🖝	🖯 🕄 📥 🛛 🖈 🚥	file/function	🔝 + Addins +				
• Unstied1* ×	DaterSammLRmd ×	E5511_de ×	Trat_clean ×	Trat50_clean =	Trat ×		
1	Source on Save   Q	ו18			-+ Run	🍉 🕆 👌 🕞 Soi	ince 🔹 🗦

Die folgenden drei Symbole sind dir wahrscheinlich von anderen Systemen wie "Word" bekannt. Die ersten beiden sind zum Speichern, mit dem anderen Symbol kannst du dein Skript drucken.

•••								RS
0.00 @.		A Coto	file/function	🔄 + Addins +				
• Untitled 1" ×	Caterdamm	LRmd ×	E5511_de ×	Trat_clean ×	Trat50_clean =	Trat ×		
0000	Source on Si	ive 🔍	ו18			-+ Run	🍉 🔿 👌 🕞 Source	∗ ≥
1								

Gleich darunter werden dir die Dateien angezeigt, die du gerade offen hast. Bei mir ist das gerade ein Skript (Das Dokument "DatenSamml.rmd", die anderen Dateien sind Tabellen (z.B. zur Übersicht, die ich über das Environment geöffnet habe). Ich kann auf jede Datei zugreifen, ohne dass der Code von einer anderen Datei verloren geht. Im Rahmen dieser Anleitung werde ich vorerst mit dem Dokument "Untitled1" arbeiten.



Diese Datei kann natürlich auch umbenannt werden, indem du die Datei speicherst. Achte darauf, wo du deine Datei auf deinem Laptop abspeicherst, damit du sie auch immer findest, wenn du auf sie zugreifen willst.



In der "Console" wird dir der ausgeführte Code zusammen mit den Ergebnissen angezeigt. Dort führt R also deine Befehle aus. Auch Fehlermeldungen werden dir dort angezeigt, die dir dann den Hinweis liefern, wieso und wo genau eine Fehlermeldung eingetreten ist. Mit dem "Environment" erhälst du einen Überblick über die selbst erstellten Objekte. Es werden dir Daten, Variablen, Werte, usw. angezeigt. Im unteren rechten Bereich werden dir unterschiedliche Dateien, Plots, Hilfsangebote und eine Übersicht über die verschiedenen Pakete angezeigt, die du dir installieren kannst. Zu diesen drei Bereichen werden im Laufe der Anleitung ausführlichere Übersichten und Erklärungen kommen.

### 3. Erste Arbeitsschritte mit R

Jetzt geht's ans Arbeiten mit R. Wie bereits erwähnt, findet der größte Teil der Arbeit im Editor statt. Fangen wir mal mit grundlegenden Rechnungen an. Jetzt willst du natürlich die Ergebnisse der jeweiligen Rechnungen haben. Hierfür kannst du unterschiedliche Ausführungsmöglichkeiten nutzen:

1. Du kannst oben auf "Run" drücken (dann wird aber nur eine Zeile ausgerechnet, und zwar die, die du gerade ausgewählt hast, in diesem Falle würde nur die vierte Zeile ausgerechnet werden, also 30\*80.).

2. Wenn du die Ergebnisse schneller und einfacher haben willst, kannst du auch folgenden Shortcut benutzen (dieser berechnet dann alle Zeilen): cmd+shift+enter beim Mac, strg+shift+enter bei Windows/Linux.

#### 2025/06/08 06:49

i a 🕸 🐲 i la 🔐 🗇 la Constatuente i 🗇 e MARCO.	
F Addung L - F Condensitied -	-0
COLO DI Chamanna (S. 2 - 1)	and the phone 2
1. 28-0	
1 1 1 1	
4 19710	
an marane	Escap 1
Second Torontal - Background Jake -	-0
R-BAS-OF	
sare, consequence and consequer, second)	
8-4	
1.04	
21-1	
0.26	
56/18	
10.9	
10-10	
() 2400	

Wie du siehst, erscheinen die Ergebnisse unten in der "Console".

Für komplexere Geschichten in R, müssen z.B. Variablen vorerst definiert werden, damit du anschließend damit weiterrechnen kannst. Das geht, indem du einer Variable (z.B. x) einen Wert zuweist, dafür brauchst du einen Pfeil ( $\leftarrow$ ). Der Pfeil steht fürs Definieren (z.B. x definiert 10, y definiert 30. Also x+y=40). Das kannst du theoretisch auch mit beliebigen anderen Wörtern/Namen machen:

•••	
R = 🕸 🐲 🖓 🕼 () 🕼 () to tradictories () 📋 + ABH +	
E manual - E monomotive -	-0
COLDER Devolution 19, 21 - 11 - 446	6.144.340404-131
1 20-4	
8 8 10 10 10 10 10 10 10 10 10 10 10 10 10	
a sy b W Data e W	
12. Huar - 18. 12. Orts - Huge	
12 SF Physical I	These I
Consels Tarmind - Redupment Jobs -	-7
R44371	
> 58/10 CC3 F	
- 3740	
0.0 1444	
210.03	
>p < M	
(1) M	
s Date at 20	
have as 15	
S Dits + Not (3) IB	

Im Moment, indem du die Berechnung durchläufst, ploppt im "Environment" etwas auf. Hier kannst du dann deine Variablen einsehen – bei mir sind jetzt recht viele, die von einem Seminar sind. Wenn du nach einer Zeit den Überblick verlierst, so wie bei mir, dann kannst du diese Liste auch "säubern", indem du oben auf den Besen klickst. Dann wird die Liste geleert. Du kannst zudem auch Funktionen oder Daten einsehen, die du im Laufe der Zeit definiert hast. Wenn du etwas genauer deine Schritte verfolgen magst, kannst du deine letzten Schritte unter "History" aufrufen.

new .		A Date of the local of
Endrannam History S	anumiera Naturial	
🐨 🔛 🛛 🗁 input Based	- 1 (* 201 Mill + 1 🗶	i the s life s
R - 🚯 Cobel Stationard	-	9
6	1008.	
NUM, OFT	-0. MOMENTATION IN	
man_corrit	-0.230043044754985	
wah_carth	8.40W0644383087	
4.3001	1899	
man Jacob Admap	1899	
process, corr	-0.85088108M080	
pearset_sorr3	-0.474684334564682	
HLX.	Named Nam (21: 5) 4. 500 4. 601. 1. 809	
mLy .	C.4MINCREASED	
spall in	144 (2) 20 7 8 4 7 4 8 6 7 8 7	
10007501.0077	-0.3278 PEUR0EHG	
spanner, servit	8.4271633C7163068	
COMPLEMENTS.	Let D: MR 3435555566	
ward.	141 (2) (2) 0 0 7 10 9 6 8 30 10 8	
1987	144 (5) (8) ( 2 2 3 5 4 3 5 5 1 2 3 C	
sart.	IN THE PLAN AND A DAMAGE	
1	10	
and the second sec	and the second	
	19	

Definieren wir mal eine Variable und rechnen den Mittelwert aus. Zusätzlich wollen wir ein Histogramm erstellen. In der "Console" werden dir die Ergebnisse und deine eingegebenen Definitionen angezeigt. Im "Environment" entsteht dann deine Variable (grün umrandet), die du einsehen kannst. Wenn du beispielsweise Histogramme erstellst, werden dir die im "Plot"-Bereich Last update: 2025/04/22 22:35

#### angezeigt.

Elaboration Elaboration		Annual State of	maline famile	
And	THE R. LEWIS CO.		THE PARTY OF THE P	
1.2				
1.4.4		-	100000000000	
1.44		1.818		
			-	
		1000 C		
		and the second s	The second se	
10 Jan -				
10.00		20		
and the second se			a contraction and a second second	
		10000-000	the second se	
8, 1522-823			THE DESIGN OF A DESIGN AND A DE	
			the same size of the same size	
			The second se	
		-		
		1.48		
		-		
		and the set		
		second and second	Carrier (see astrong	
		And the second second	and the second second	-
test test testers		the second second	the state to an and the state of the state o	
1. 1.00			and the second se	
			The second se	
1.1				
		A		
1982 1 2				
11				
and the second sec				
- and the second second second second				
(Apple and				
Charlest and				

#### 3.1 Pakete

Wenn du R runterlädst, hast du das Basic Paket, mit dem du einfache statistische und mathematische Dinge machen/berechnen kannst. Je tiefgründiger du aber mit R arbeitest, desto sinnvoller und notwendiger werden zusätzliche Pakete (Packages). Keine Angst, die sind alle kostenlos und diese Pakete runterzuladen kostet meistens nur einige Sekunden. Es gibt unterschiedliche Methoden, diese runterzuladen. Am besten du entscheidest dich, je nachdem, welche und wie viele Pakete du runterladen willst.

1. Methode: hauptsächlich, wenn du nur vereinzelte Pakete runterladen willst: beispielsweise "Tidyverse". Dieses Paket beinhaltet automatisch mehrere Pakete, die hilfreich sein können:

Palet.	beckreitung
appietz	Erstellen von Detenvisuelisierungen mit einem "Grammar of Graphics"-Amatz.
datar -	Datemanipulation, wie Pillern, Orappiener und Transformieren zen Daten.
Ney:	Daterau/Benetiang daniti Umdraktorierang von Beiten (J.B. Postleren).
NeX	Importieren und Lesen von rechtschigen Datenalitzen (CSP; T3P).
part	Funktionale Programmierung mit Funktionen wie Imgelit our bereden.
shirle	Verbeiserte Delenframes mit bessene Onzekaungabe und Funktionen zur sicheren Handhalsung von Oston.
shingr	Söling-Manipulation und Textremerbeitung.
hereasts.	Paritilizenen aur Jaharil mit kallegarialen Variablen: Paktoreni.

(Die Tabelle habe ich aus Chatgpt, hier der Link für die Konversation: https://chatgpt.com/share/67e14a77-eaa4-8012-bdf5-8acc8bdbf9a7)

Du kannst das Paket "manuell" runterladen, indem du unten rechts in der Abteilung "Packages" auf "Install" klickst. Dann öffnet sich das Fenster und du kannst dein gewünschtes Paket aufrufen. Anschließend klickst du auf "Install".

and a second sec				_	_
second displaying		And and the state of	a late		
m - 1			A COMPANY OF A COMPANY		
**			COMPANY AND		
			-		
		and the second	1 A MARK OF A MARK		
			A Profession of the second secon		
		18.4	THE R. LEWIS CO., LANSING MICH.		
		10.1			
		100			
			and the second se		
			CONTRACTOR CONTRACTOR OF CONTRACTOR		
and the second			the local sectors and and a		
	ACCESS TO A				
			the state of the s		
	the second second				
	The second se				
	100 B 100 B 100				
			100 0100 0.000 00		
		and interaction			
- Les Longest		and the same same	The second		
		and the second second			
			100 B B		
-		with the second			
			An other strength		
the second se			And a street water street at the		
the second second second second second second			Trade free of the lines done free to see to		
		1.00	the inclusion of the life has been been been	-	
the second s			An operation of the second sec		
100 L 10			A 10 Million and a second second		
			Contraction and Contraction of Contr		
		and and a second se	Institute & doorsel instantion instantion into the second limit.		
The second second second second			ALTERNATION AND		
			Sale of the second second second	100	
			Book on from here there as being		
			The second		
What when the second		-	And the statement of the second statement of the		
the first of a part of a state of the second se					
			and frame have been at all and an and and and		

2. Methode: wenn du mehrere Packages runterladen musst, kannst du eine Variable mit den Packages erstellen und diese im anschließenden Schritt ausführen. Das sieht dann folgendermaßen aus:

2025/06/08 06:49

_	Annual Barristani Annual Carlos	
	and the summarian in the state of	
۰.	E Contra de la presión de la contra de la co	and the second of the
- 2	84	
- 1	9.9	
÷	17.00	
6.		
. 4	a M	
	3 - 10	
	n y	
	Annual March 199	
- 22	former - Bana	
12		
14	seriable = =3.4.1.8.8.4.4	
18.	marcarialat)	
14	Mat(variable)	
10		
18	instants - s(	
- 22	contrar a participant participant	
17	Contract ( Sector )	
12	a strange based	
	Chapter 1	18 borest
	m Tenting - Referentiate -	1 boar
ine.	m Terrinal - Ratigmentate -	1 boat ==2
(ana ()) -	Departed 1 als Terrelative Reduperand Mer Red. 1	-
Comu () -	The Constant I and a second state in a second state in a second state in a second state in a second state second state state second state	
Comu (R - 1.1)	- Transmitter - Kalaparant Alle - 18443 - 117 - 18443 - 117 - 1871 - Martinger (Martin) 1987 - High Andrea (Martin) - Genetic Antonio Negli Ser-antek (Martin Antonio Seria 1987 - Status (Martin)	
Carras (i) - 1.14 Style Style	Internal - Matyawa Jako - Ek J	
Came () -	The Tennish I Bellgement Adv Tel. J 4 <sup>-1</sup> Tel. J 4 <sup>-1</sup> Tel. Appl. Compar. Tel. Bell, Section 2014, Section 2014, ACA Systems In Light Telephonological and Telephonological Academics In Light Telephonological and Telephonological Academics Section 2014	1040.00°
(i) - 10	(Republic ) Belgewei als : EK3	42.4 Pop*
	(Figure 1) August data: REALT-α <sup>2</sup> REALT-α <sup>2</sup>	
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	The Section 4 - Intervention 4 - In	100 124 Mgr
1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 -	(mp. 10) Response data : RALS = r <sup>(1)</sup> (RM), Set (RM), RM = RALS = r <sup>(1)</sup> (RM), RM = RALS = r <sup>(1)</sup> (RM), RM = RALS =	all the second s
	(Percent)     (         Monton)     (	
	The stands - 1 <b>B A Solution - 1</b> <b>B A Solution - 1</b> <b>A Solution - 1</b>	
	In particular Sequence data : RALS - σ <sup>+2</sup> EMD, and Particular Sequence data : EMD, and Particular Sequence and Compared Register and Argument Mark Ard Markers In Care Particular Sequence data : Sequence 2000; Versil COSI 800 In Sequence 2000; Sequence 2000; Sequence 2000; Sequence 2000; Sequence 2000; Sequence 2000; Sequence 2000; Sequence 2000; Sequence 2000;	
Cares (R - 1	Topper 1     Bandwine 1     Ban	
Carras (iii - 1.5) (iii - 1.5) (iii) - 1.5)	The stands - 1 (Section - 1 (Se	i kon ut A Ray ' 'Bage'
- 10	Impaired 1 Market All and a set of the set	LEADUR"
1 - 10 - 10 - 10 - 10 - 10	The Section 1     The Section 2     The Sec	A box and "Blogs" "Blogs" "Higgs 2.4.8 —
1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 -	The share is a second set of the second set of t	i kon ant Stage Hages Hilgens 2.4.8 –
Care and a second secon	The second seco	Alber and SECRETUP "Enge" "Highere 2.4.8 —
1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 -	The second	4 km mi 42.4.9.14* 
	The Share I is a Share I is a second set of the Share I is a s	4 box 20.4.8 tgr 
Card and a second secon	Impaired 1 Sector - σ <sup>2</sup> Sector -	مرا س میر س میر س میر س میر س میر س میر س میر میر میر میر میر میر میر میر میر میر
	The second seco	i kon al al al al al al al al al al
	(marked)	
	In processi In processi Address - Comparison of the comparison o	ник с

#### 3.2 Good to know

In R gibt es die Funktion eine "Markdown-Datei" zu erstellen. Diese ermöglicht es, Codes, Grafiken usw. in einem Dokument miteinander zu vereinen. Man kann sie auch als "Cheat-Sheet" betrachten. Wenn du auf "R Markdown…" klickst, kannst du dem Dokument und dem Autoren/der Autorin einen Namen geben. Zudem wird das Erstelldatum gespeichert.

					New R Harldown		
					Document	Title	Untitled
4 x95x84	18 10 Date 194 B	di 160		00 97024 900	Presentation	Author	Asceym
•••	Harris		Rikelpe 1111	10	Shiny	Date	2025-01-18
* Almost	realition.		Quarte Documenti.			Date:	
1 40	Oper He	80	Quarte Presentation -	Ren March Laborer - 11	<ul> <li>From Template</li> </ul>	Use cu	ment date when rendering document
1 61	Open the in New Column Response the Encoding Response Theory	,	d Markdown. d faintalaidd			Default 0	utput Format:
1 11	Oper Project.		Bring Hele App., Phankar (IP).			Recommend	inded format for authoring iyou can switch to
2 -4	Recard Projects	>	CPIe			PDF or We	ord output anytime).
10 Coller -	Ingo/10earet	5	Con File Header/File			O PDF	
18.	2000	11.1	Markdoore Film			PDF outp	ut requires TeX (MikTeX on Windows, MacTeX
15 marcal	Dava mil.		HTM. No.			2013+ 0	I US X, TEX DVE 2013+ On Linux).
10. 10.00.00	Sam with December		COLUMN TWO			O Word	
18 participa	fam til		C018/W			Previewin	g Word documents requires an installation of
18 18140	Cample Report		Public Sciel			MS Word	(or Libre/Open Office on Linux).
10 1000	Pend		The Doug				
15 10.000	China		IKG. Buript				
in prints	Child-All		That Fig.				
it cannot	Close All Except Current						
15			ErDA.		Country Economy Documents		OK Canal
1111 -7444	Gair Season		R thorumentation -	Resource 1	Court ordeb reconsent		un cann

Eine neue Datei (.rmd) öffnet sich und darin arbeitest du am besten, v.a. wenn es um komplexere Datensätze geht. In der Übersicht wird dir die Markdown kurz erklärt und wozu sie genutzt wird. Das ist deine Übersicht in R:

statung h : @ sentral : @ interfurnitional :	-7
The second states of the second states of the	March 1997 August Manager
bart find	10.00
a sale manager	
a sea the second	
The second second second	
and the second sec	
7	
Rear In state included \$200	
3 holds under cherdlandfahr - 7007	
2	
2	
To M I Related	
13	
24. This is in 8 Northner storent, Montone is a pages foresting	provide for exchange (196), NO, and Hill Real
deciminate. For some details on onling it Workdown part datas china	etates, catalog, and
70	
20 Wen you click the "Month" bottom a document will be generated output of our opheniat \$ only change at \$115, 100 document. The	I the technic beh carant as will as the
22	
30. · · · D. 10/11	
I MARCYCOND	
58.; · · ·	
6	
E. B. Societa Pate	
E In Securing Plats	
<ul> <li>If (scholing Fiel)</li> <li>To on size and plots, for example:</li> </ul>	
1 In Second Plats 1 Tay on size and plats, for events:	

#### 3.3 Datensätze in R laden

Vor allem im universitären Kontext kann es vorkommen, dass du mit Datensätzen arbeitest, die bereits existieren. Somit musst du die Studien und Umfragen nicht unbedingt selbst durchführen, kannst aber dein Verständnis im statistischen Bereich trotzdem so sehr vertiefen wie du willst. Wir haben mit den ESS Datensätzen gearbeitet, deswegen werden diese im Fokus stehen. Wir haben mit der 11. Runde des ESS Datensatzes gearbeitet, das ist die rezenteste. Du kommst mit folgenden Schritten dahin. Zuerst rufst du die ESS Seite auf, dahin gelingst du mit folgendem Link: https://ess.sikt.no/en/datafile/242aaa39-3bbb-40f5-98bf-bfb1ce53d8ef

Mit welchem Datensatz du arbeitest, entscheidest du natürlich selbst (oder im Seminar), je nach Interesse. Dann wählst du unter den vier Möglichkeiten "ESS11- integrated file, edition 2.0 aus". (Wichtig!: Damit du die Daten auch aufrufen/runterladen kannst, musst du eingeloggt sein, das geht ganz einfach mit einem Google Account und es ist alles kostenlos.)

a - it is an investigated					
Barrison Province 4	Conception Research Conception of Conception Research     Encoded     Encoded Sciences (2002)     Sector Research Conception Research Conception     Transmission Research Conception Research     Transmission Research				
	Read Toronto and the contract of				
	Market 1: Millionering Sphered works				
	All control and the participants are been				
	In some till delse steller offense forse forsjo				
	the same and strapping the strapping state				
	the same and the second allocations are used.				
	Bur Property II				
	BE LET'T DR. MALE ALLE. AND				
	El contro del frequente Associatione				
	The second is the second second second provide the second se				

Dann hast du die Option, zwischen mehreren Programmen zu wählen, mit denen du den Datensatz runterlädst. Am einfachsten ist das SPSS (.sav) Format oder auch das Stata (.dta) Format. Aber auch das (.CSV) Format funktioniert, dafür musst du aber das Paket "readr"-Paket installiert haben. Ich habe das Format für SPSS gewählt. Das Runterladen geht recht schnell, und braucht i.d.R. nur einige Sekunden, maximal eine Minute.



Jetzt willst du natürlich den Datensatz in R laden, damit du auch damit arbeiten kannst. Dafür musst du beachten, wo du deinen geladenen Datensatz im Computer abspeicherst, am besten kennst du den "Pfad" zu deinem Datensatz. Ansonsten könnte folgende Fehlermeldung aufleuchten und das hindert dich an der Arbeit mit dem Datensatz, denn dein Computer findet den Datensatz einfach nicht:



**Tipp**: Falls du es nicht weißt: Um den "Pfad" zu finden, kannst du oben in der Leiste über die "Session" die "Working Directory" auswählen. Dann bekommst du Auskunft über den Pfad deiner Datei.

	new teacor		hudhs
2 - Borger G. Brief Let an annexes 2 - Manager - Romann - Romannesse 2 - Jan B. Dennis - Romannesse 2	Internet I Internet R, Instant II and Date Oxford Instant II and Date Oxford Instant II and III All Oxells Jot Restant Breach Land Westing Devology II Land Westing Devology II Land Westing Devology II Date Restant Re.	andi Balance Schementer Schementer	Radianad Ma Control (Control (Contro) (Control (Contro) (Contro) (Contro) (Contro)
Fri C. B. Mark, Includentials Friends, Settlementation - 1885 Set	Chill Portugios.		0 detail 0 1304 0 1304, de
12- "P Warted 13 ToherryMaterial 14 ToherryMaterial 13 ToherryMaterial 13 ToherryMaterial			0 farmery 0 tote 0 tote

In der "Console" wird dir dann der "Aufenthaltsort" der Datei beschrieben. Spätestens jetzt ist alles bereit, dass du den Datensatz in R laden kannst. Zuerst überlegst du dir einen Namen, der den Datensatz definiert, z.B. "ESSdatensatz". Das machst du mit dem Pfeil, wie du vorhin beim Definieren von verschiedenen Variablen gelernt hast. Dann soll die Datei "gelesen" werden, die du auf deinem Laptop abgespeichert hast und die du ja wie beschrieben mit dem Pfad findest.

California P. 2 Address also And 2 & Spectromited 1	= 2
COLD M Determs 12 9, where the	The Contractor in Marcol
Terrs Wood	1.0404
E talan 'B kadawa'' B adden 'Bened' B adden 'Bened' B adden 'Bened' B adden 'Bened' B adden 'Bened'	
B. T. In star, include/1423) B. India: optimation()etc. (1003) 381 (2014) 10.	
<ul> <li>The state bady</li> <li>Theory (applied)</li> <li>Theory (applied)</li> <li>Theory (applied)</li> <li>Theory (applied)</li> <li>Theory (applied)</li> <li>Theory (applied)</li> </ul>	

Wenn du auf den grünen Pfeil rechts klickst, durchläuft R immer den ganzen Chunk (das sind die einzelnen grauen "Bereiche"). Das kannst du aber auch ganz einfach mit einem Shortcut durchlaufen lassen, wie ich euch im Kapitel zu Shortcuts zeigen werde. Wenn der Chunk durchlaufen wurde, erscheint im "Environment" eine neue "Data".

nisener Moats G	enorien. Faschal	-0			
G Insection	2 (2 (2 ( # # - / #	11 Mart - 1 17 -			
· A tabe burning					
Elistemate	4000 site, of \$40 sectorizes				

Wenn du auf diese klickst, öffnet sich ein Tab mit den ganzen Angaben (also vom ganzen Datensatz, den du runtergeladen hast). Das kannst du aber auch einfach wieder zuklappen, diese Angaben dienen nur zur genaueren Information und du brauchst sie in der Regel nicht dauerhaft. Das sieht dann ungefähr so aus:

	Topol Bake +
Informat Miles Interim Total	-7
🕼 🔝 The Server Dataset + 🗈 (2016) + 🖌	10 Mar - 1 10 -
8 - Color Deleased -	9
Losse	
I second a set D 40 NO D D D D D D D D D D D	
Satistics 1 to Transfer Contractor and the	
A second start of the second real and the second start the second start and	
I then I are 12-46000 teller teller teller teller teller teller	
S also in the station of the rate of the	
S desight ( ) and [1.00050] S.005 B.04 S.0.00 B.555 B.755	
8 manufet 1 mar [2:48180] 8.361 8.381 6.031 6.041 1.361	
3 peright 1 rue [146:814.1014.1014.1014.1014.101.	
S asserged 1 was (2 MEN) & 12 MINO 3 SEP A REP A 2001	
is named non [2:00:00] the test to us do use of 120 20 120	
Entranti : ne DAGRO EED EED EED	
3 Aduste 1 No 2148201 39 59 9 98 98 98 13	
Significant 1 and [2.48.00] Sidd Sidd SiZ 4 P.3	
Emploit : not [24630] 0494 287 884	
Tartile I we DARRIELEA AND TAREA AND TAREA	
\$ pd 543 1 Hot 12 MENO Example concentration to be the target	
Greenper i nat (LABAC) EE EZ EZ EZ EZ E	
E-minutge - nor (1-4636) E-1 4 3 1 5 3 4 3 3	
Research I we thereight 4 h 4 h 4 h 4 h 5 h 1 h 1 h	
\$ organia : na (1.4650) 5 2 3 3 5 3 7 5 2 7	
6 majori - 1 mar (2.66640) 6 6 7 6 6 3 6 6 6 8 4	
3 (moligit) 1 mm 12 (MERC) 8 (0.1 (0.1 (0.1 (0.1 (0.1 (0.1 (0.1 (0.1	

#### 3.4 Mit Datensätzen arbeiten

Mit diesen Datensätzen kannst du natürlich viele unterschiedliche Dinge berechnen und analysieren, je nachdem was dein Fokus ist. Die Datensätze von ESS beinhalten Daten von vielen unterschiedlichen Ländern und je nachdem ist dein Fokus auf einem bestimmten Aspekt, oder einem bestimmten Land. Dann ist es sinnvoll, die Daten zu filtern, damit du ausschließlich diese analysieren kannst. Nehmen wir mal an, du willst die Daten aus Deutschland analysieren. Dafür kannst du einen Code benutzen, der aus den Datensätzen nur die Daten aus Deutschland anzeigt. Anschließend kannst du dann damit weiter analysieren.

1		-			-	-	The local data of the local data		- 1
12	1						Times dia citati		
		11				-			
	15								
-	10					-			
-	1		27		2	-			
1	1		47	1	2	1			
-	1	-	11	1	1	111			
1.111	- Provent		11	1	11114	1111			
1.1111	- prover	100		1	11114	1000	10 TO, TOM TO TOO TOOLO		
	a proven			1000	1000	1000	No No. You No Han Madage		
1.11111	a berner			P IIIIII	1000	· · · · · · · · · · · · · · · · · · ·	N N, NA N N N NA		-
-	- Province			10000	10000	COLUMN T	NI NI VAI NI NI NI NAAN	2.0	~
-	- provenue			100000	NAMES OF A	· · · · · · · · · · · · · · · · · · ·	N NL FOR N NL FORM	1.0	1
	- p			P IIIIIII	100000	PHONE:		-	

Im "Environment" kannst du dann die Variablen dazu einsehen und wenn du darauf klickst, öffnet sich ein neuer Tab, in dem du die ganze Tabelle im Überblick hast. So sieht das dann aus (das ist natürlich nur ein Ausschnitt, also 26 von 2.460):

£)	keininang R	<ul> <li>B Adet</li> </ul>	enni Opiers	itane. Rend -	- Kilder	enastelii i	· · Dated	(ammi Rimal -		
		West Lot	0.001	0.18				0,		
	name 1	exercised in the second	adding 1	problem	idea 1	many 1	daulight -	prepagate 1	panight 1	1
1	00111-02	11	1	RILEN	10004	08	0.00100.00	1.4094082	1.0000	
2	15411-02			2011-3634	50052	06	4-9958686	6.6296065	2.00007	
3	0011-02	11	1	RILINH	10011	08	0.00100.00	4.943400	1.0000	
	0411492	11		20112034	50080	06	6-9958667	1.094512	2.00007	
1	00110-02	11	1	RILEH	10070	08	0.0000000	4.143400	LHH	
16	0411492			20112034	50088	06	14047323	1.638879	2.00000	
2	0011-02	11	1	RILINH	10048	08	0.0000000	1,016,892	LHHH	
	0411492			20.11.2024	50000	- 16	4-9938686	1.000.005	2.00007	
	0011-02	11	1	10.11.HUH	10089	08	0.0000000	0.220804	LHH	
	0411492	11		10112434	50600	16	1.2238085	1.030-01	2.99897	
23	STREET, NO.	11	1	IOTLINH.	10LTP	08	0.0000000	0.007039	LHH	
12	0411492	11		10112034	50138	-	1.9258086	0.025304	2.998997	
28	0011-02	11	- 1	DOL:NH	10048	08	0.0000000	LANDON	1.0001	
94	01102	11		10112034	501.55	16	1.2258086	0.0707758	2.99897	
28	0011-02	11	- 1	10.11.HUH	10045	08	0.00030000	1.707488	1.0001	
16	011102	11		10112034	50058	16	1.9258086	1206360	2.99897	
27	0011-02	11	- 1	10.11.HUH	100 H	08	0.0000000	1.1099887	2.0001	
10	011102	11		10112004	50004	16	1.9958086	1202410	2.99897	
38	0011-02	11	- 1	10.11.HUH	10018	06	0.0000000	0.0270018	2.00017	
en.	011102	11		10112004	50857	96	1.11110000	1.2503.000	2.99897	
21	0011-02	11	- 1	DOL:NH	10048	06	0.0008687	1.826347	2.00007	
ee.	011102	11		10.11.0014	10075	98	1.11110000	1,00008	1.791197	
23	00110-02	11	- 1	10.11.30H	10080	06	0.0008080	0.0007718	2.00007	
64	0111472	11	1	1011-0014	10030	98	1.2238080	4,3455400	1.791197	
21	00110-02	11	- 1	10.11.HUH	10014	06	0.0008686	1.0148942	2.00007	
-	011592	11		10.11.0104	10012	98	1.11100.001	0.969(2.90	1.791197	
33	REIMO	11	1	10.11.1014	10444	14	A-841848	2 Indext at	1-44042	

Ein nerviges Problem, das beim Analysieren deine Werte manipulieren kann, (indem besonders abweichende und unrealistische Werte z.B. bei der Schiefe oder Kurtosis rauskommen) sind ungültige Werte. Diese entstehen wenn bei der Umfrage (z.B. Fragen mit Antwort, bei der man eine Skala bis 10 hat) nichts oder falsche Antworten angegeben wurden. In dem Fall werden die Antworten oft mit "77", "99" (meistens aber mit "88", zumindest bei mir) gekennzeichnet. Generell kann die angegebene Zahl aber nicht höher als 10 sein. Deswegen müssen diese fehlerhaften Werte ausgefiltert werden. Sagen wir, du interessierst dich für folgende Variablen: "trstprl", "trstlgl", "trstplc", "trstplt", "trstprt", "trstep", "trstun" und willst aus diesen Variablen die ungültigen Werte filtern. Dafür kannst du eine Variable erstellen, die alle 7 enthält und sie definieren, wie du willst. Da es um Variablen geht, in denen das Wort "trust" vorkommt, nennen wir die Variable jetzt einfach "Trst":

Or an and a second sec					S fraject (Kern
* Adving U Method Sciences And		-3	<b>Gentronment</b>	Bidary Connectors Tylestal	-1
OT OTHER TRANSPORT OF A CONTRACT OF	1 C 1		OF ALL PROPERTY.	per lister - 🕐 151 mil - 🖌	line of the
laurer Vesal		≥ Dutine	B r Galad	Engineering v	0.
3			Dete		
		C 7 4	ESSabitaneat	2 4856 dot. of 648 variable	
31+ *** bd			ESSaturaneat	201 2428 stor. of 648 variables	
23 Test or Elizatematality, all instants, firstally, firstally, firstally, firstally, firstally, firstary, firstary,	CONNECTION OF		Tyst.	2428 also, of 7 variables	
11					

Dann schreibst du R vor, alle Werte, die größer als 10 sind (>10) durch N.A. zu ersetzen. Du kannst auch einen anderen Code nehmen, z.B. "Trst[Trst == 77 | Trst == 88 | Trst==99]  $\leftarrow$  NA. Das ist dir selbst überlassen, im Endeffekt werden die ungültigen Werte durch NA ersetzt, was ja das Ziel ist.



Wenn du nun die Tabelle öffnest, kannst du sehen, dass einige N.A.-Werte im Datensatz vorhanden sind:

	· · ·							a restaurante
		-						
	**	ranje	nays.	TANK .	sum.	Table .	Table .	
				5	1		- N	
				5			- P	
							10	
1								
							- P	
							- P	
	-							
							-	
6.0								
						_	_	

Bei so einer großen Tabelle kann es natürlich vorkommen, dass du den Überblick verlierst. Damit du einen Überblick erhältst, ob NA Werte vorhanden sind, kannst du einen weiteren Code aufrufen:



Wahrscheinlich wirst du bei der Analyse Stichproben ziehen, mit der Grundgesamtheit arbeiten ist oft für Seminarkontexte zu viel und generell nicht besonders sinnvoll. Um zu vermeiden, dass du Stichproben ziehst, die NA Werte haben, kannst du eine Variable erstellen, die von solchen Werten befreit ist. Am besten gibst du ihr einen Namen, an dem du erkennen kannst, dass sich die "gesäuberten" Werte darin befinden. Deswegen nenne ich sie jetzt "Trst clean".

21 -	- · · · (r)	0 = +
22	<pre>Trst &lt;&gt; ESSdatersatz0E[, c("trstpl","trstpl","trstplt","trstplt","trstprt","trstep","trsten")]</pre>	
23	Trat[Trat > 10] <- NA	
24		
25	<pre>sapply(ESSdatersatz0E, function(x) sum(is.no(x)))</pre>	
26		
27	Trst_clean <> Trst[complete.cases(Trst), ]	
28		

Die Tabelle mit den sauberen Werten kannst du dann auch immer wieder im "Environment"-Bereich öffnen und einsehen.

						**************************************	second And second rates	
	-						A REPORT AND A REPORT OF	1.00
		 		1.000	1.000		the Brandward of Contract of C	
							hein-	
							Printeenter 1878 de si 18 - ministre	
							Witnessender 100 als of 100 metalline	
2		 -	-		2 2			_
		 -					Winkshop 100 do of Landshop	
-		 		-				
		-						
-								
-		-	-					
-		 -	-	-				
-					-			
-		 			-			
-		-						

#### 3.5 Stichproben ziehen

Oft arbeitest du mit gezogenen Stichproben und vergleichst diese mit anderen, denn die Grundgesamtheit zu untersuchen ist sehr aufwendig und ineffizient. Von der Stichprobe schließt du dann auf die Grundgesamtheit. Wenn du dann unterschiedliche Stichprobengrößen mit anderen vergleichst, kannst du Muster und Zusammenhänge erkennen und Hypothesen testen. Das geht auch in R ganz einfach mit einem Code, der dir ermöglicht deine erwünschte Stichprobengröße zu erzeugen. Du definierst deine Stichprobe, ich habe sie jetzt "Stichprobe100" genannt und du brauchst dafür natürlich deinen gesäuberten Datensatz, damit keine abweichenden Werte rauskommen. Deswegen habe ich den Datensatz "Trst\_clean" genommen und nicht "Trst". Dieser Code wählt dann zufällig 100 Zeilen-frames aus dem Datensatz "Trst clean" aus, erstellt also eine 100er Stichprobe.

29 30 - 31	<pre>```{r} ```{r} Stichprobe100 &lt;- Trst_clean [sample(nrow(Trst_clean), size = 100), ]</pre>	⊙ <b>±</b> •
32 -		
33		
34		
35		
34:1	(Top Level) 4	R Markdown 1

Wenn du den Code ausführst, kommt im "Environment" eine Variable dazu: "Stichprobe100". Auch hier kannst du drauf klicken, dann öffnet sich ein Tab, der die Tabelle mit euren 100 Stichproben anzeigt:



Wichtig: Wenn du den Code wie oben angezeigt ausführst, entstehen jedes Mal andere Stichproben! Das ist ungünstig, wenn du über längere Zeit damit arbeitest. Jedes Mal wenn du R öffnest, musst du den Code nämlich nochmal durchlaufen lassen. Es entstehen dann jedes Mal neue 100er Stichproben, was für deine Analysen manipulierend ist. Deswegen macht es Sinn, dass du den Code anpasst, das geht mit "set.seed(123)". Das ermöglicht, dass die Stichprobenziehung beim Durchlaufen immer

Last update: 2025/04/22 lv-wikis-oeffentlich:boot2024:2.\_arbeiten\_mit\_r https://institut.soziologie.uni-freiburg.de/dokuwiki/doku.php?id=lv-wikis-oeffentlich:boot2024:2.\_arbeiten\_mit\_r 22:35

identisch ist und sollte deswegen unbedingt anpegasst werden.

29 - ""(r) 30 unt.see(123) 31 Stidgrobe100 ~ Trst\_clean [sample(nrow(Trst\_clean), size = 180), ]] 32 - """

**Tipp**: Du kannst auch mehrere Stichproben gleichzeitig ziehen, du musst nicht alle Stichproben in einzelnen Schritten ziehen. Klickst du einmal auf den grünen Pfeil, entstehen gleich drei Stichproben.

34 -	····{r}						0 =	F
35	set.seed(123)							
36	Stichprobe100	Trst_clean	[sample(nrow(Trst_cleon),	size	- 100), ]			
37	Stichprobe250 <-	Trst_clean	[sample(nrow(Trst_clean),	size	- 250), ]			
38	Stichprobe500 <-	Trst_clean	[semple(nrow(Trst.cleon),	size	= <b>SNO</b> ), ]			
39 -	***							

### 4. R und seine unterschiedlichen Logiken

Wer sich mit R befasst und sich beispielsweise in einer Gruppe mit anderen darüber unterhält, wird bemerken, dass es in R nicht nur "den einen richtigen Weg" gibt, um zum "Ziel" zu kommen. Manchmal vergleichst du deine Codes mit anderen und ihr alle habt, wenn alles gut geht, die Ergebnisse, die ihr haben wolltet. Du hast aber eventuell einen anderen Weg/Code benutzt. Trotzdem funktionieren alle eure Codes. Es gibt eben unterschiedliche Logiken, die euch aber alle zum Ziel führen können.

In diesem Kapitel gehe ich auf unterschiedliche Logiken ein, die auftauchen können. In diesem Kapitel stelle ich dir drei gängige Verfahren vor und unterschiedliche Wege, wie du sie in R anwenden kannst. Es gibt in den meisten Fällen die Option, ein Verfahren durchzuführen, ohne unbedingt ein bestimmtes Paket installiert zu haben. Die Pakete vereinfachen das Coden aber in den meisten Fällen. Allerdings gilt: wenn du tiefgründigeres Verständnis für die Logik von R entwickeln willst, arbeitest du am besten mit so wenig Paketen wie möglich.

#### (1) Deskriptive Statistik

Die Berechnung dieser Verfahren verläuft recht einfach mit den Basisfunktionen von R. Für die Berechnung von Mittelwert, Median und Standardabweichung sind also keine Pakete notwendig. Dahingegen kannst du aber auch ein Paket anwenden, das dir diese Verfahren berechnet. So kannst du beispielsweise alternativ das "dplyr"-Paket anwenden:

Martin - Lik, Ki, K, K, K, Ki, Ki Martin - Lik, Ki, K, K, K, Ki, Ki Martin - Like - Like - Martin - Like -	Amendung ches Paleat	
() 0.000 () 0.00 () 0.000 () 0		
<ul> <li>The second second</li></ul>		Anwednung mit Paker "Apyr"
	Production of the	
	P.D.W	

Der Unterschied besteht nur daraus, dass dir bei der Variante ohne Paket mit dem Befehl "summary" weitere Daten angegeben werden, wie z.B. das 1. Und 3. Quartil, zudem das Minimum und das Maximum. Wenn du die rausfinden willst, musst du das bei der Variante mit Paket explizit angeben. In diesem Fall brauchst du jetzt nicht unbedingt ein zusätzliches Paket, aber hier geht es auch um einfache Berechnungen.

### (2) Lineare Regression

Auch eine lineare Regression kannst du mit unterschiedlichen Codes berechnen.

A CONTRACTOR OF A CONTRACT OF A CONTRACT	
Electrolistic Electrolistic Id. (April )	-1
	B
10.0 No.	1.000
I Compared a compared	
Minatory of Manager	
weight with the set of the s	Anwendung ohne Pakol
Particul Life of 1924 K. Police Rule	
E Internation	
<ul> <li>New - de tento d' New - de tento d' New - de tento de tento New - de tento</li></ul>	Anwentung mit Pakat "späyr"
Minimum         Minimum         Minimum         Minimum           Minimum         Minimum         Minimum         Minimum         Minimum           Minim         Minim <td></td>	

Die beiden Varianten unterscheiden sich in der Datenstruktur. Bei der Anwendung ohne Paket werden x und y als numerische Vektoren definiert und bei der Anwendung mit einem Paket wird ein Datenrahmen "Daten" erstellt. Das lineare Modell wird mit der Formel "y ~ x" und mit "Daten" spezifiziert. Diese Methode ist allerdings eher sinnvoller für Datensätze, die etwas komplexer sind. R greift bei der ersten Variante auf die Vektoren zu, die sich im "Environment" befinden. Bei der zweiten Variante greift R explizit auf "Daten" zu, was die Arbeit mit komplexen Datensätzen erleichtert.

### (3) T-Test

Einen T-Test kannst du auch mit verschiedener Syntax durchführen, du musst nicht mal unbedingt für die zweite Variante ein besonderes Paket installieren:

A REAL PROPERTY OF A REAL PROPERTY OF A REAL PROPERTY.		
history and the second se		
of the second of the second of the	5- C	-
8 (page) (10, 10, 10, 10, 10, 10)		
2 All 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		
kinetpagel, propil, an end - NO		
Here langter a research		
hits presed on presed		
tantish, malk basa san		
e prose contante transit.		
ALL REAL REAL REAL REAL REAL REAL REAL R		
And of a line of a		
This are a set that		
5 1 1 1 A 1 A 1 A 1 A 1 A 1 A 1 A 1 A 1		
Londers - logs, dog - moders		
NAME OF COMPACING AND ADDRESS OF COMPACT ADDRES		
Marcol and America a control		
And A line in the line of the		
10-100 V-000		
No. And A. Frage 1 - 4 (19), A - 1 (10), project - 4,00 Project and page 10 R array of Phase Learning	and in our and in the	
No. No. 6 (1998) 1 - 4 (1998) & -1 (100), project -4.00 distributions for the second states in the second states prove forget and property If a prove of the second states in the second states prove forget and property -1 (1998) and 1000	aged to not apply in th	
But, New & State (1998). A set of the set	agari 1 ani agari 1 di	

Die erste Variante führt einen T-Test durch, indem die zwei Gruppen als separate Vektoren definiert werden. Dabei trifft die Option "var.equal = True" die Annahmen gleicher Varianzen. Bei der zweiten Variante werden die Werte von beiden Gruppen im Datenframe "ttestdaten" zusammengefasst. Spalte "Gruppe" markiert die Zugehörigkeit zu "Gruppe1" und "Gruppe2". Der T-Test wird mit der Formel Werte ~ Gruppe durchgeführt, und R vergleicht die Mittelwerte der beiden Gruppen im Datenrahmen.

Der Unterschied der Syntax in den meisten Fällen ist also, dass der Code bei der ersten Variante mit getrennten Vektoren arbeitet, während die zweite Variante die Daten in einem Datenrahmen speichert. Wie bereits erwähnt, ist diese Variante v.a. vorteilhafter, wenn du mit größeren Datensätzen arbeitest und komplexere Analysen durchführen willst. Wenn du deine Syntax mit anderen vergleichst, kann es durchaus vorkommen, dass ihr eine unterschiedliche Syntax habt. Oft ist auch erlaubt, mit KI-Systemen zu arbeiten und es kann sein, dass Programme wie chatgpt dem einen die 1. Syntax vorschlägt und dem anderen die 2. Syntax. Im Endeffekt ist es wichtig zu verstehen, was Last update: 2025/04/22 lv-wikis-oeffentlich:boot2024:2.\_arbeiten\_mit\_r https://institut.soziologie.uni-freiburg.de/dokuwiki/doku.php?id=lv-wikis-oeffentlich:boot2024:2.\_arbeiten\_mit\_r 22:35

R genau mit den Codes macht und wie das Programm damit arbeitet.

## 5. Shortcuts und nützliche Tools in R, die ihr kennen solltet

Shortcuts vereinfachen generell das Arbeiten mit Computern, nicht nur in R und ermöglichen ein schnelleres Arbeiten. Grundsätzlich sind folgende Shortcuts praktisch und effizient, wenn du in R schneller vorankommen willst. Deswegen solltest du diese kennen.

#### Nicht nur in R, aber zur Erinnerung

- Command + S -> speichert das Dokument. (Macbook) / Strg + S (Windows, Linux)

- Command + Z -> den letzten Schritt rückgängig machen (Macbook) / Strg + Z (Windows, Linux)

- Command + c -> kopieren / Command + v -> das Kopierte einfügen (Macbook) / Strg + c / Strg + v (Windows, Linux)

**Spezifisch in R** - Zur Erstellung eines neuen Markdown Chunks: Option + Command + I (Macbook) / Strg + Alt + I (Windows, Linux)

- Zum Ausführen des aktuellen Chunks: Shift + Command + Enter (Macbook) / Strg + Shift + Enter (Windows, Linux)

- Der Zuweisungspfeil: Option + - (Macbook) / Alt + - (Windows, Linux)

**Tools um das Arbeiten mit R zu erleichtern #** hinter dem Code bedeutet, dass alles dahinter nicht miteinberechnet wird. Das ist besonders am Anfang eine praktische Hilfe, weil du hinter jedem Code beschreiben kannst, was da passiert, ohne dass R bei der Berechnung verwirrt wird und eine Fehlermeldung ausgibt.

21 + "(r) 22 Trst = 4556sternst. 23 Trst[Trst > 10] < \* # Siles non größer ols 10 ist wird durch WA ersetzt. Auch Redwungen werden micht berechnet, Z+6 44

Das **"\$"** bedeutet, dass du auf eine bestimmte Spalte zugreifen kannst, ohne die anderen Variablen im Code zu beachten.

```
41× ""(r)
42 Personelists <- list(Nome - "Hens", Alter = 24, Interessen = c("Lesen", "Sport"))
43 ListelName Face greaft the in der Liste nur auf dem Nomen zu
44× ""
```

Ein **Code**, den du auf **jeden Fall** kennen sollst, ist ein Code, der statistische Analysen anhand einer Schleife durchführt. Dafür erstellst du zuerst eine Liste (hier namens "spalte"), diese enthält fünf Dataframes (hier im Beispiel: Stichprobe100) und du gibst die ausgewählten Variablen ein. Die Schleife iteriert dann über die jede angegeben Spalte in der Liste, und das nur mit einem Mal ausführen. So sparst du bei einer hohen Anzahl von Spalten enorm viel Zeit, weil alles nacheinander durchgerechnet wird. Bei diesem Code wird dann mit jeder Spalte, jede einzelne angegebenes Verfahren durchlaufen, also z.B. der Mittelwert, der Median, usw. Das **"print"** gibt eine Textart aus, das die Ausgabe für jede Spalte kennzeichnet. So etwa sieht die Ausgabe dann aus:

#### 2025/06/08 06:49

#### Empfehlung für weiteres Lernen von und mit R:

Folgender Blog ist empfehlenswert: https://r4ds.hadley.nz/

Spezifischere Fragen/Probleme: https://blog.r-project.org/

Auch wenn nicht immer 100% Verlass ist auf die KI (immer überprüfen): https://chatgpt.com/

From:

 $https://institut.soziologie.uni-freiburg.de/dokuwiki/ \ - Institut \ für \ Soziologie \ - \ Lehrwiki$ 

Permanent link: https://institut.soziologie.uni-freiburg.de/dokuwiki/doku.php?id=lv-wikis-oeffentlich:boot2024:2.\_arbeiten\_mit\_

Last update: 2025/04/22 22:35

