

Medizinische Statistik, Informatik und Telemedizin

1. Praktikum

- Messniveaus
- Daten importieren aus .xls(x) und .csv Dateien

optional

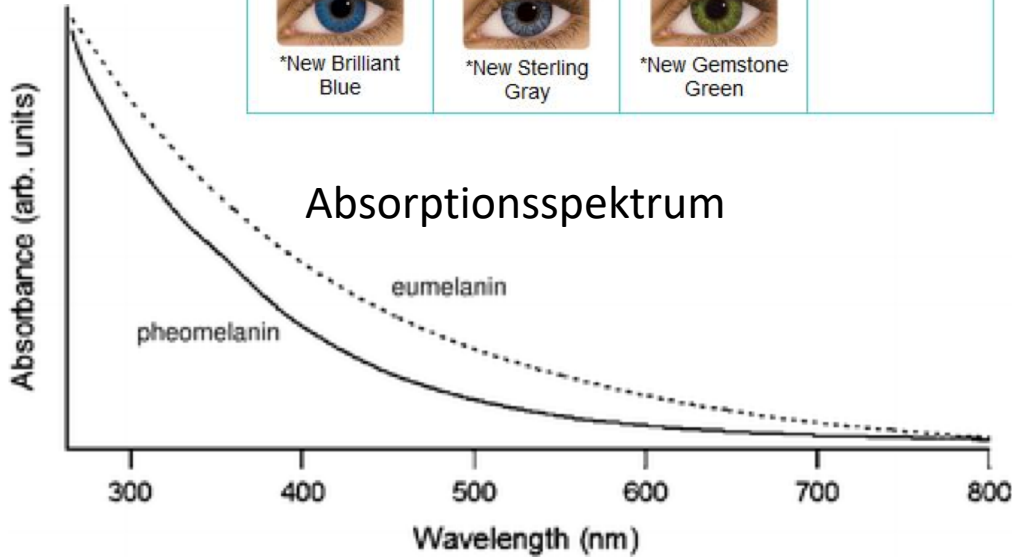
Statistische Variablen

Alles was wir messen oder beobachten können, z.B.:

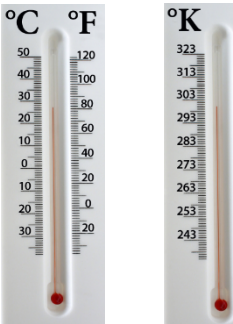
Münzversuch



Augenfarbe



Temperatur



Zahl der Zähne



Würfelwerfen



Kalenderdatum



Druck



Platz bei einem Wettlauf

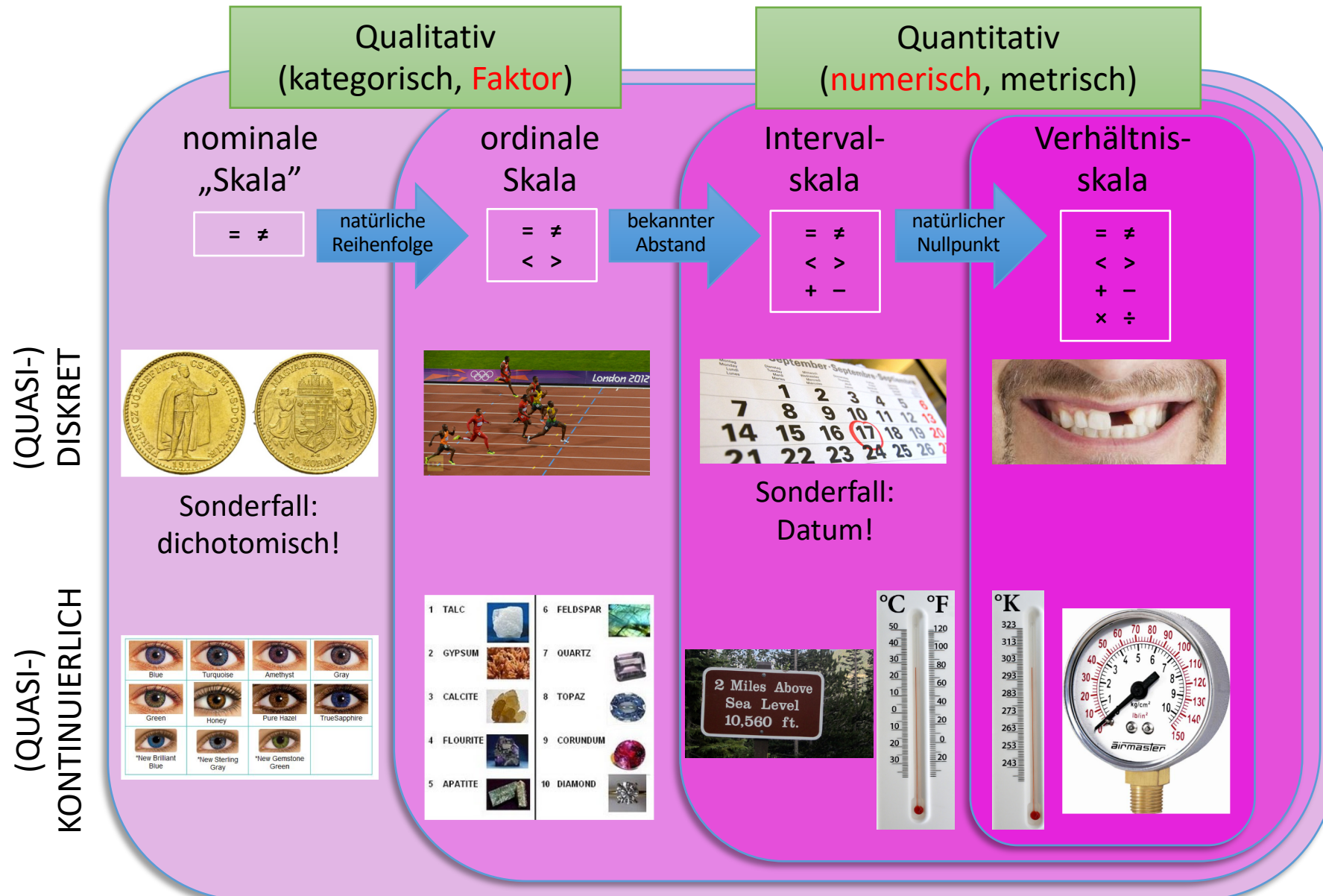
Härtewert nach Mohs

1 TALC		6 FELDSPAR	
2 GYPSUM		7 QUARTZ	
3 CALCITE		8 TOPAZ	
4 FLOURITE		9 CORUNDUM	
5 APATITE		10 DIAMOND	

Höhe über dem Meeresspiegel



Typ der Variablen, Messskalen, Operationen



Messniveaus nach
Stanley Smith Stevens
(1906–1973),
Amerikanischer Psychologe

Eine Datenmatrix aus Excel-Arbeitsblatt importieren

Verwenden Sie die **frmgham2_period3_onlyrisk_NA_factor.xls** Datei.

Öffnen Sie die Datei zunächst in Excel und prüfen Sie:

- Enthält die erste Zeile die Namen der Variablen?
- Was wird verwendet, um fehlende Daten anzuzeigen?

Geschlecht
(nominal als Zahl
eingegeben)

Gesamt-
serumcholesterin,
mg/dl (numerisch)

aktueller
Raucherstatus
(nominal)

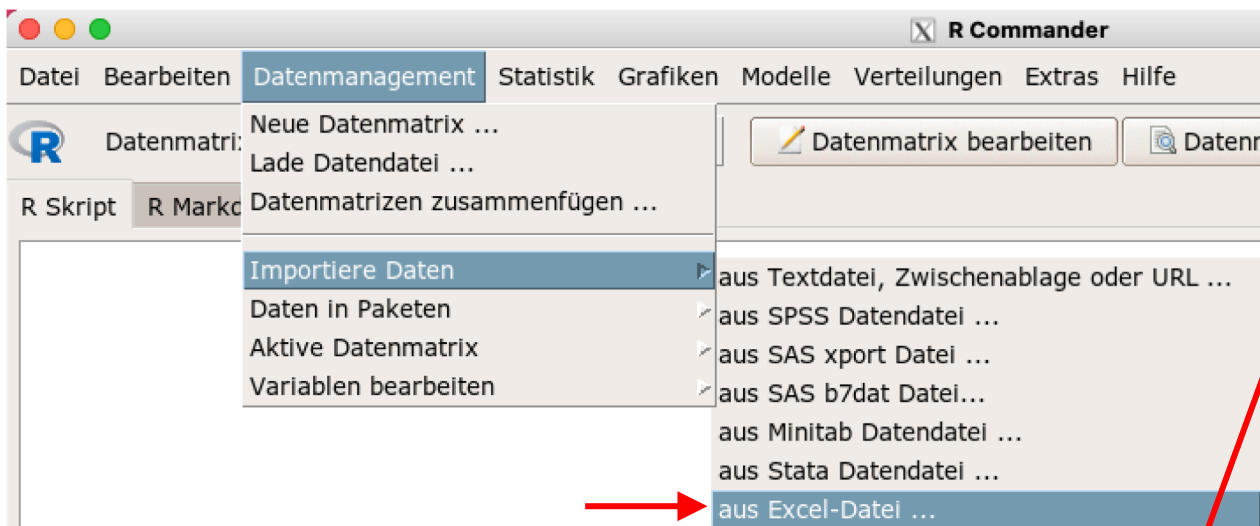
Body-Mass-
Index
(numerisch)

fehlende Daten: NA
(alle Großbuchstaben!)

Bildungsniveau
(ordinal als Zahl
eingegeben)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	RANDID	SEX	TOTCHOL	AGE	SYSBP	DIABP	CURSMOKE	CIGPDAY	BMI	DIABETES	BPMEDS	HEARTRTE	GLUCOSE	EDUC
2	2448	1	209	52	121	66	n	0	NA	0	0	69	92	4
3	6238	2	237	58	108	66	n	0	28.5	0	0	80	71	2
4	11252	2	NA	58	155	90	y	30	24.61	0	0	74	NA	3
5	11263	2	220	55	180	106	n	0	31.17	1	1	86	81	2
6	12806	2	320	57	110	46	y	30	22.02	0	0	75	87	2
7	14367	1	280	64	168	100	n	0	25.72	0	0	92	82	1
8	16365	1	211	55	173	123	n	0	29.11	0	1	75	85	1
9	16799	2	291	62	120	70	n	0	21.98	0	NA	62	83	1
10	22727	2	150	52	121	78	n	0	26.62	0	0	68	125	2

Eine Datenmatrix aus Excel-Arbeitsblatt importieren

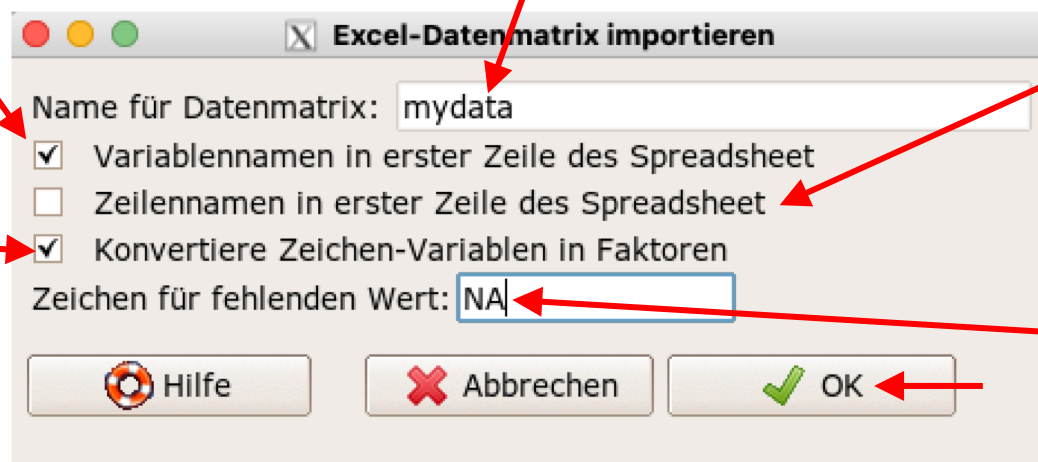


Geben Sie einen Namen für Ihre Datenmatrix ein (behalten Sie nicht den Standardnamen "Dataset" bei)

Text (Engl. string) Eingaben in R:

- verwenden Sie nur das lateinische (englische) Alphabet
- verwenden Sie vorzugsweise englischen Text
- Sie können auch Ziffern oder Unterstriche (_) einfügen
- Verwenden Sie **keine** Leerzeichen oder andere Sonderzeichen (z.B. .,;?!%+=-*/^\\@~&\$<>'()[]{}" #) da diese bei der Codierung stören können

Wie wir bereits gesehen haben, enthält die erste Zeile des Excel-Arbeitsblatts die Namen der Variablen

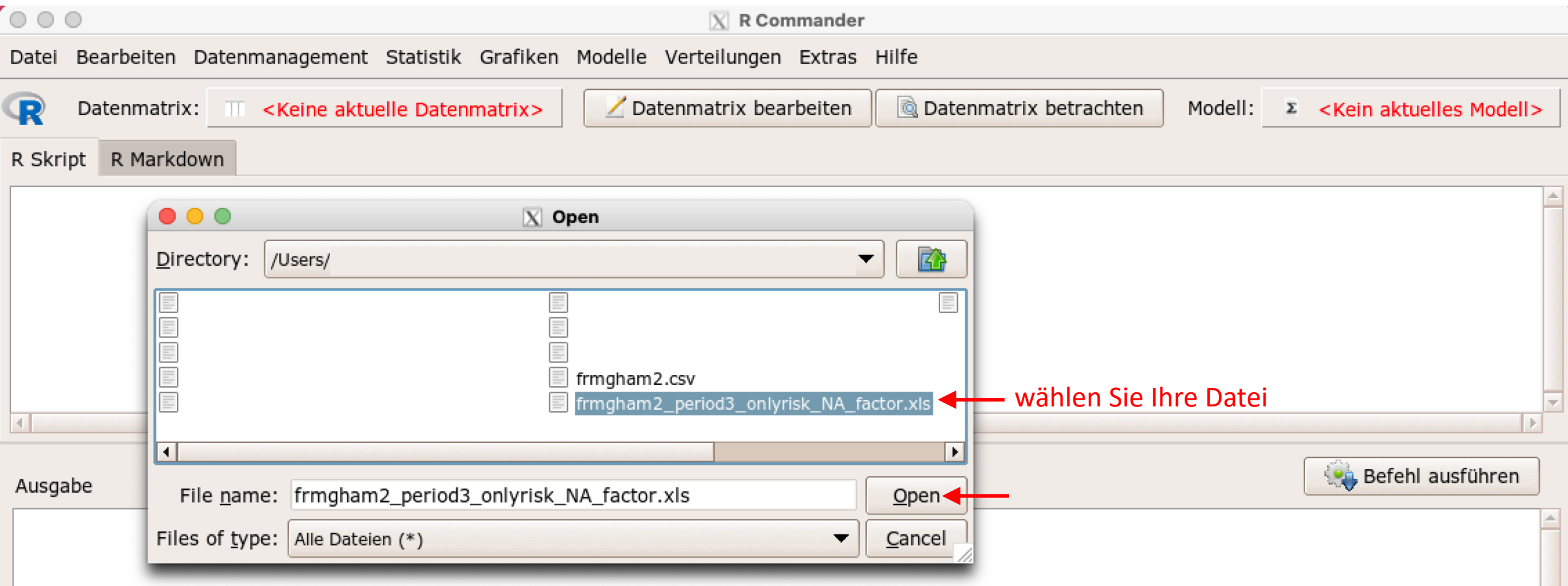


wir könnten die RANDID-Variablenwerte verwenden, um Zeilen zu benennen, aber wir brauchen sie nicht, also lassen Sie diese Option unmarkiert

im Falle einer Faktorvariable wird jeder Stufe der kategorischen Variable eine Zahl zugewiesen; wir konvertieren kategorische (Zeichen-) Variablen immer in Faktoren

fehlende Daten: NA (alle Großbuchstaben, kein Schrägstrich!)

Eine Datenmatrix aus Excel-Arbeitsblatt importieren



Eine Datenmatrix aus Excel-Arbeitsblatt importieren

The screenshot shows the R Commander window with the following components and annotations:

- Menu Bar:** Datei, Bearbeiten, Datenmanagement, Statistik, Grafiken, Modelle, Verteilungen, Extras, Hilfe.
- Buttons:** **Datenmatrix:** mydata (highlighted with a red arrow and text: "die soeben importierte Datenmatrix wird zu unserem aktiven Datenmatrix"). **Datenmatrix bearbeiten** (highlighted with a red arrow and text: "klicken Sie hier, um die aktive Datenmatrix anzuzeigen"). **Datenmatrix betrachten**. **Modell:** <Kein aktuelles Modell>.
- Script Editor:** Contains the R code:

```
mydata <-  
  readXL("/Users/  
  rownames=FALSE, header=TRUE, na="NA", sheet="frmgham2", stringsAsFactors=TRUE)
```

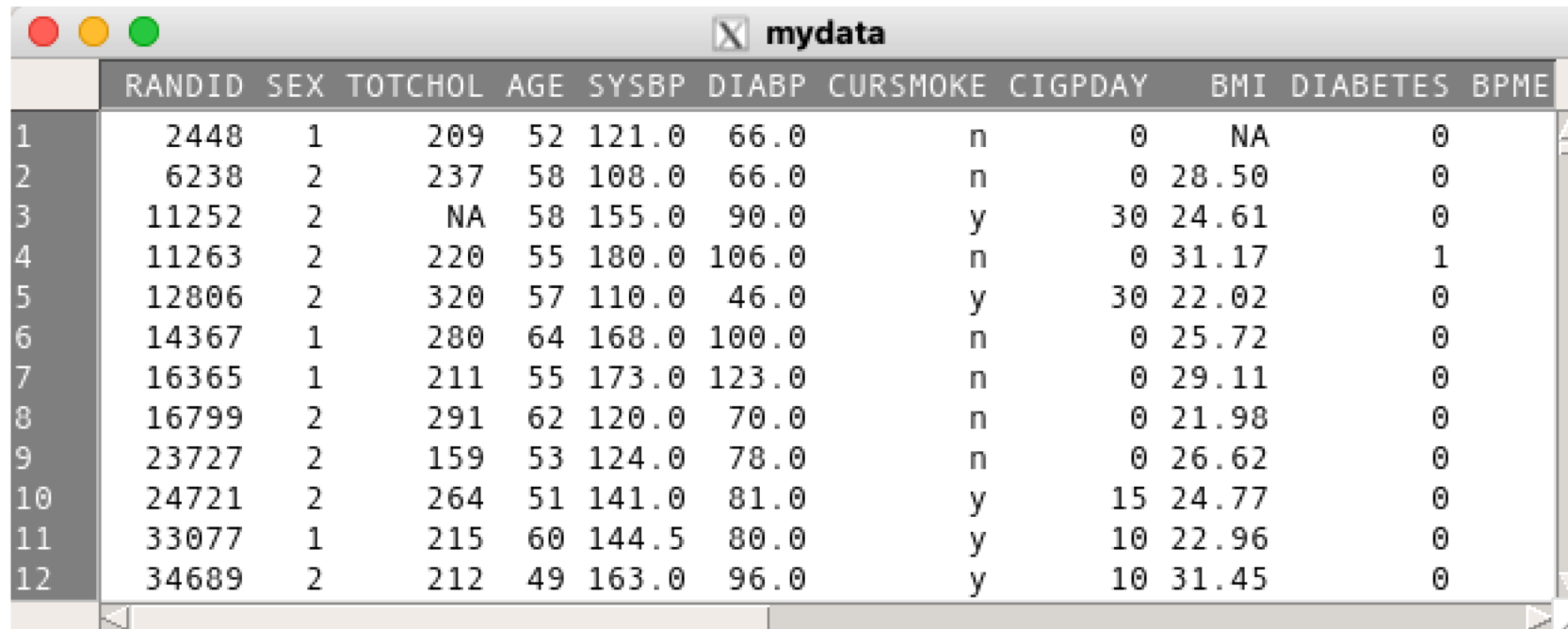
 (highlighted with a red arrow and text: "der eigentliche R-Code, der ausgeführt wird").
- Output Console:** Shows the executed command:

```
> mydata <-  
+   readXL("/Users/  
+   rownames=FALSE, header=TRUE, na="NA", sheet="frmgham2", stringsAsFactors=TRUE)
```
- Messages Console:** Shows the message:

```
[4] HINWEIS: Die Datenmatrix 'mydata' hat 3263 Zeilen und 23 Spalten.
```

 (highlighted with a red arrow and text: "Größe der importierten Datenmatrix").

Eine Datenmatrix aus Excel-Arbeitsblatt importieren



	RANDID	SEX	TOTCHOL	AGE	SYSBP	DIABP	CURSMOKE	CIGPDAY	BMI	DIABETES	BPME
1	2448	1	209	52	121.0	66.0	n	0	NA	0	
2	6238	2	237	58	108.0	66.0	n	0	28.50	0	
3	11252	2	NA	58	155.0	90.0	y	30	24.61	0	
4	11263	2	220	55	180.0	106.0	n	0	31.17	1	
5	12806	2	320	57	110.0	46.0	y	30	22.02	0	
6	14367	1	280	64	168.0	100.0	n	0	25.72	0	
7	16365	1	211	55	173.0	123.0	n	0	29.11	0	
8	16799	2	291	62	120.0	70.0	n	0	21.98	0	
9	23727	2	159	53	124.0	78.0	n	0	26.62	0	
10	24721	2	264	51	141.0	81.0	y	15	24.77	0	
11	33077	1	215	60	144.5	80.0	y	10	22.96	0	
12	34689	2	212	49	163.0	96.0	y	10	31.45	0	

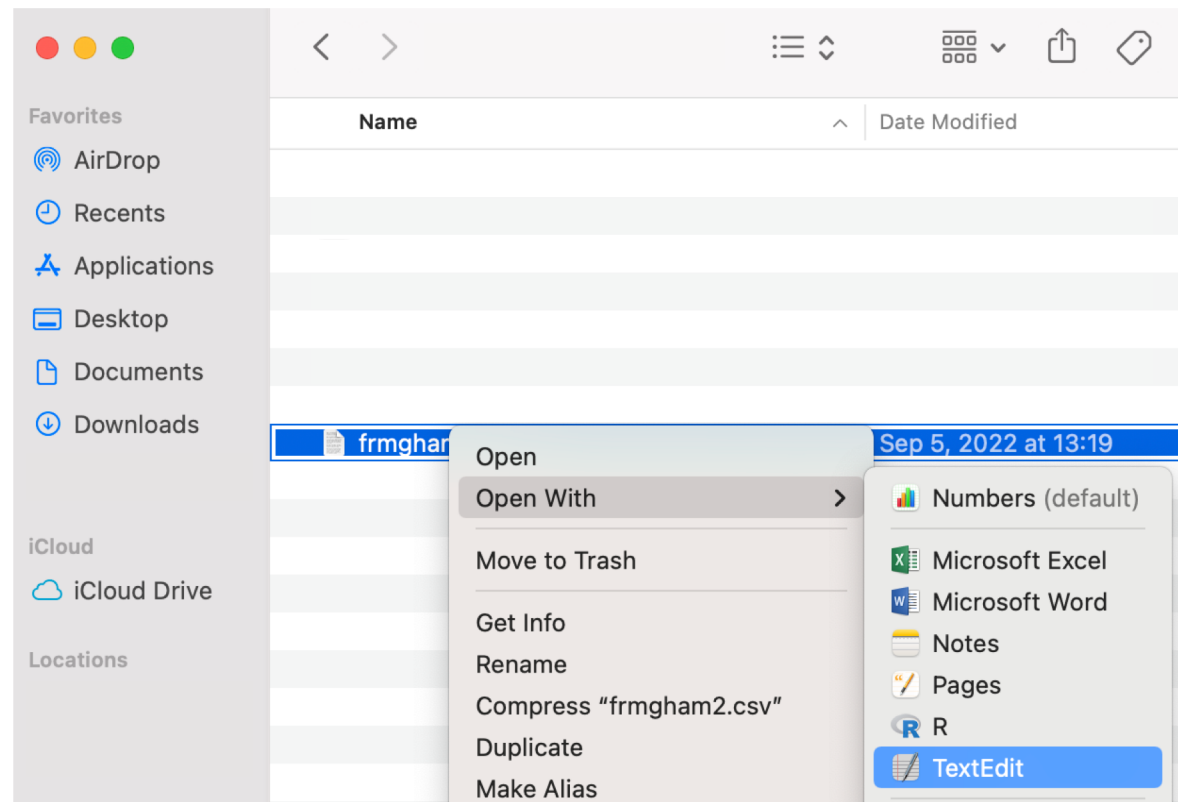
jetzt können wir sehen, dass unsere
Datenmatrix erfolgreich importiert wurde

Eine Datenmatrix aus .csv-Datei importieren

Verwenden Sie die **frmgham2.csv** Datei. Die .csv Dateinamenserweiterung steht für “comma separated values” (Komma-getrennte Werte), und es ist eine einfache Textdatei.

Öffnen Sie ihn zunächst in einem einfachen Textverarbeitungsprogramm und prüfen Sie:

- Enthält die erste Zeile die Namen der Variablen?
- Was wird verwendet, um fehlende Daten anzuzeigen?
- Was ist das Feldtrennzeichen (Symbol zur Trennung von Werten)?
- Was ist das Dezimaltrennzeichen?



erste Zeile:
Variablen-
namen

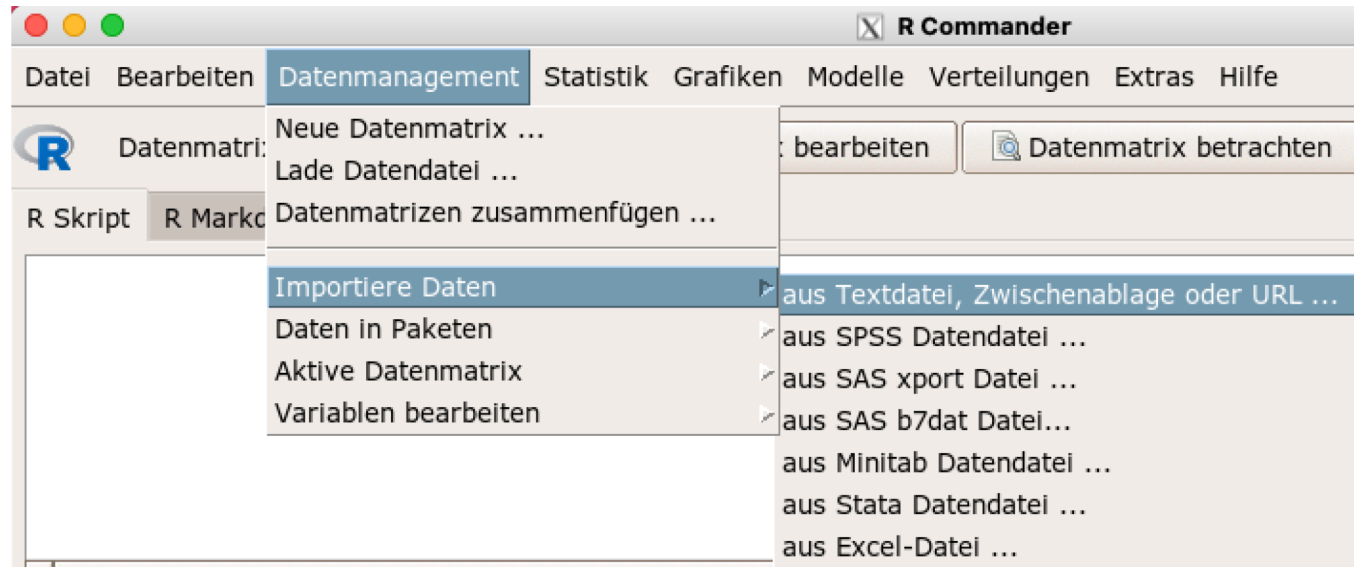
Dezimal-
trennzeichen:
Punkt

Feldtrenn-
zeichen:
Komma

fehlende
Daten:
(leer)

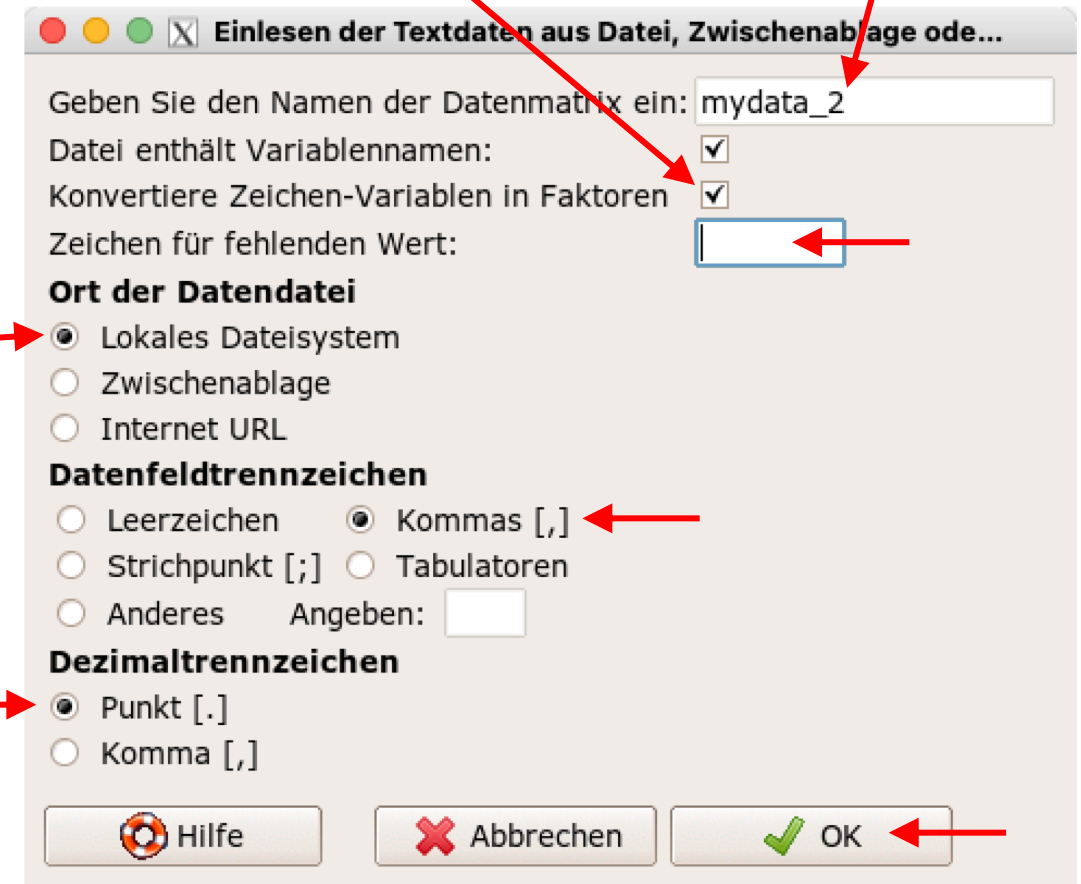
```
frmgham2.csv
RANDID,SEX,TOTCHOL,AGE,SYSBP,DIABP,CURSMOKE,CIGPDAY,BMI,DIABETES,BPMEDS,HEARTRTE,GLUCOSE,EDUC,PREVCHD,PREVAP,PREVMI,PREVSTRK,PREVHYP,TIME,
PERIOD,HDLC,LDLC,DEATH,ANGINA,HOSPMI,MI_FCHD,ANYCHD,STROKE,CVD,HYPERTEN,TIMEAP,TIMEMI,TIMEMIFC,TIMECHD,TIMESTRK,TIMECVD,TIMEDTH,TIMEHYP
2448,1,195,39,106,70,0,0,26.97,0,0,80,77,4,0,0,0,0,0,1,,0,0,1,1,1,0,1,0,8766,6438,6438,6438,8766,6438,8766,8766
2448,1,209,52,121,66,0,0,,0,0,69,92,4,0,0,0,0,0,4628,3,31,178,0,0,1,1,1,0,1,0,8766,6438,6438,6438,8766,6438,8766,8766
6238,2,250,46,121,81,0,0,28.73,0,0,95,76,2,0,0,0,0,0,0,1,,0,0,0,0,0,0,0,8766,8766,8766,8766,8766,8766,8766,8766
6238,2,260,52,105,69.5,0,0,29.43,0,0,80,86,2,0,0,0,0,0,2156,2,,,0,0,0,0,0,0,0,8766,8766,8766,8766,8766,8766,8766,8766
6238,2,237,58,108,66,0,0,28.5,0,0,80,71,2,0,0,0,0,0,4344,3,54,141,0,0,0,0,0,0,0,8766,8766,8766,8766,8766,8766,8766,8766
9428,1,245,48,127.5,80,1,20,25.34,0,0,75,70,1,0,0,0,0,0,0,1,,0,0,0,0,0,0,0,8766,8766,8766,8766,8766,8766,8766,8766
```

Eine Datenmatrix aus .csv-Datei importieren



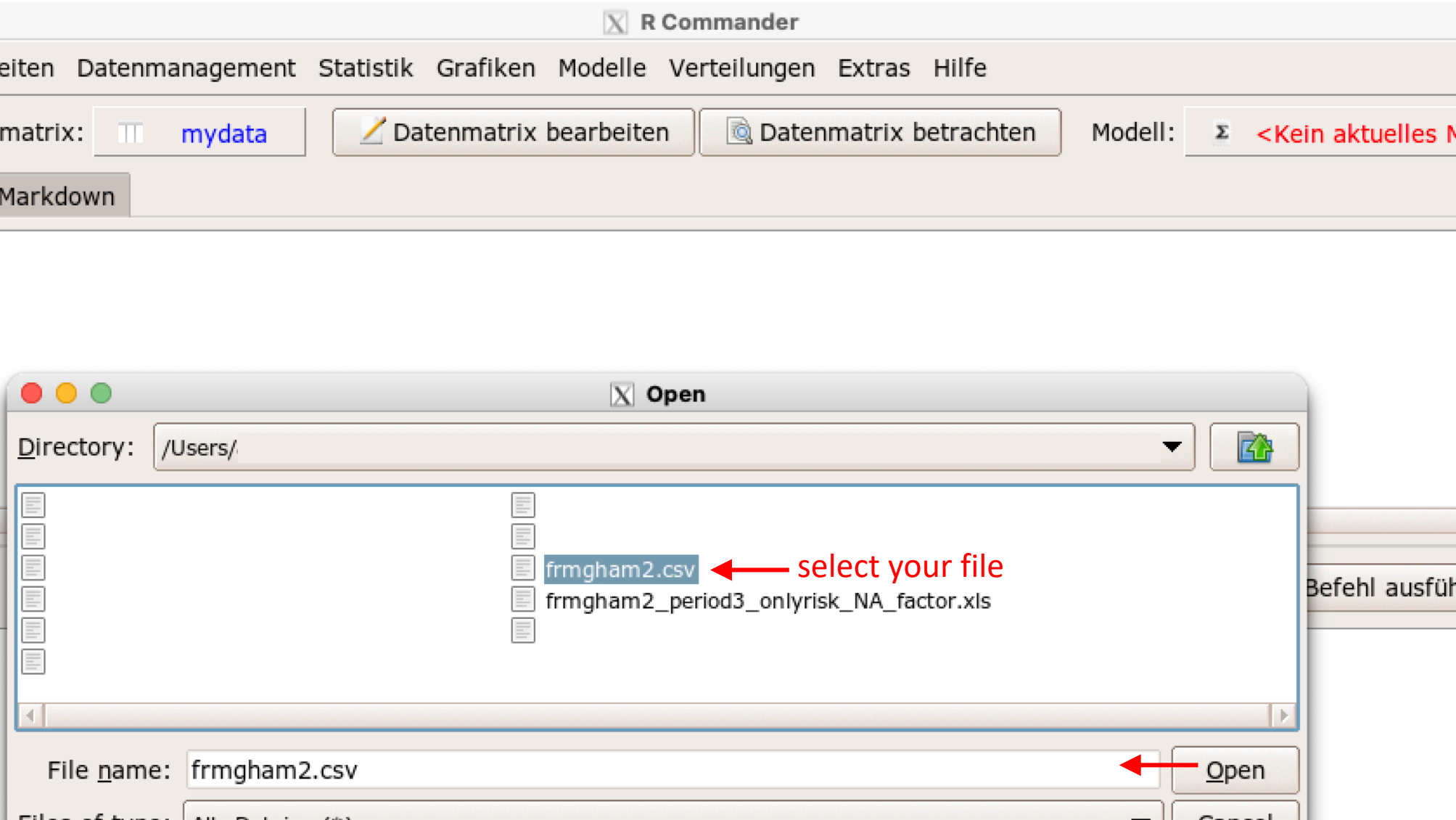
wie bereits erwähnt,
werden wir immer
Zeichen-Variablen in
Faktoren umwandeln
wollen

verwenden wir einen
neuen Namen für
diese Datenmatrix

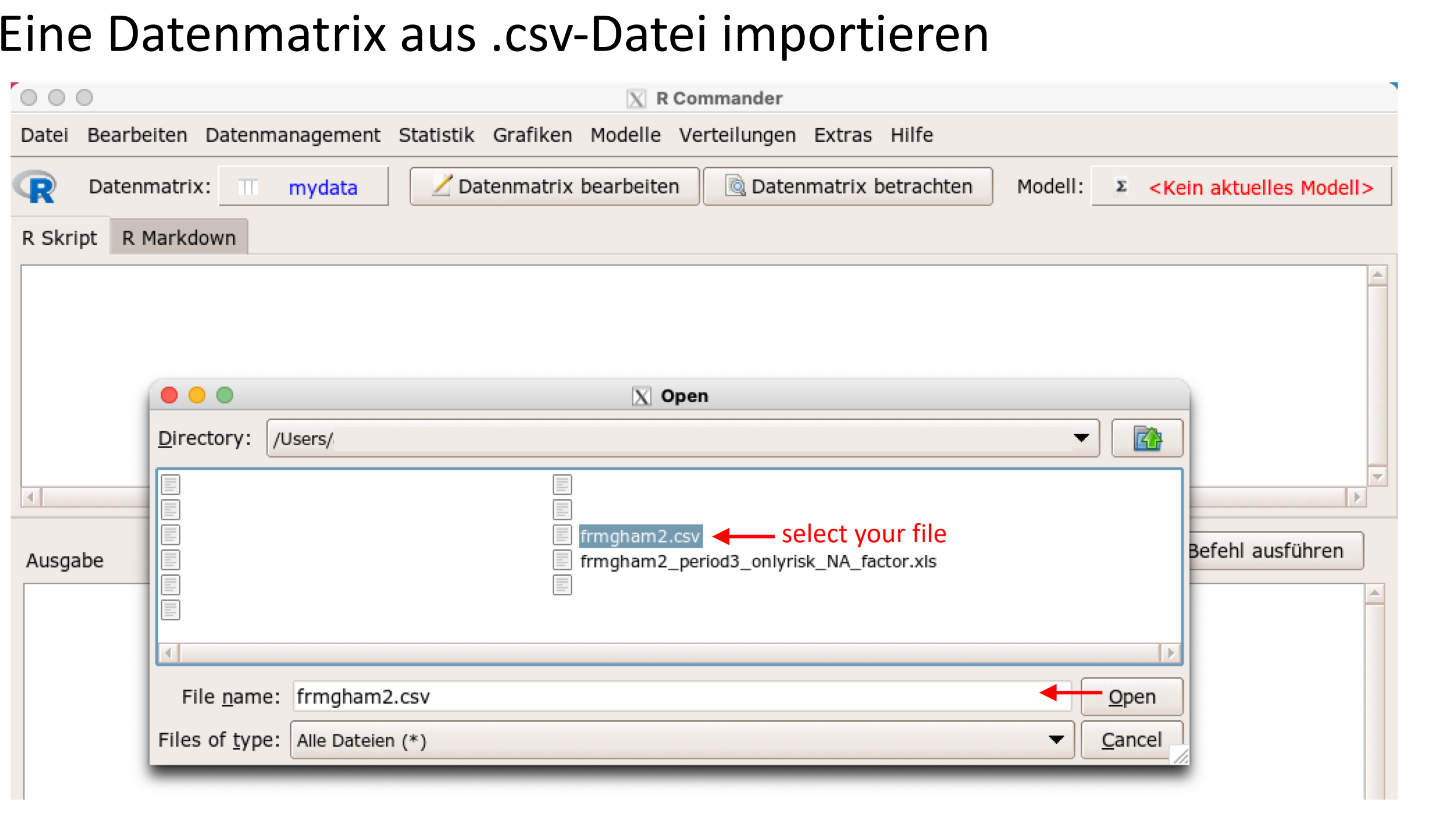


wir werden das Dokument auf
unserem Computer finden

Eine Datenmatrix aus .csv-Datei importieren



The screenshot shows the R Commander interface with the 'Open' dialog box open. The dialog box displays the directory '/Users/' and lists two files: 'frmgham2.csv' and 'frmgham2_period3_onlyrisk_NA_factor.xls'. The file 'frmgham2.csv' is selected. A red arrow points to the selected file with the text 'select your file'. Below the file list, the 'File name' field contains 'frmgham2.csv'. The 'Files of type' dropdown is set to 'Alle Dateien (*)'. The 'Open' button is highlighted with a red arrow.



Eine Datenmatrix aus .csv-Datei importieren

The screenshot shows the R Commander window with the following components:

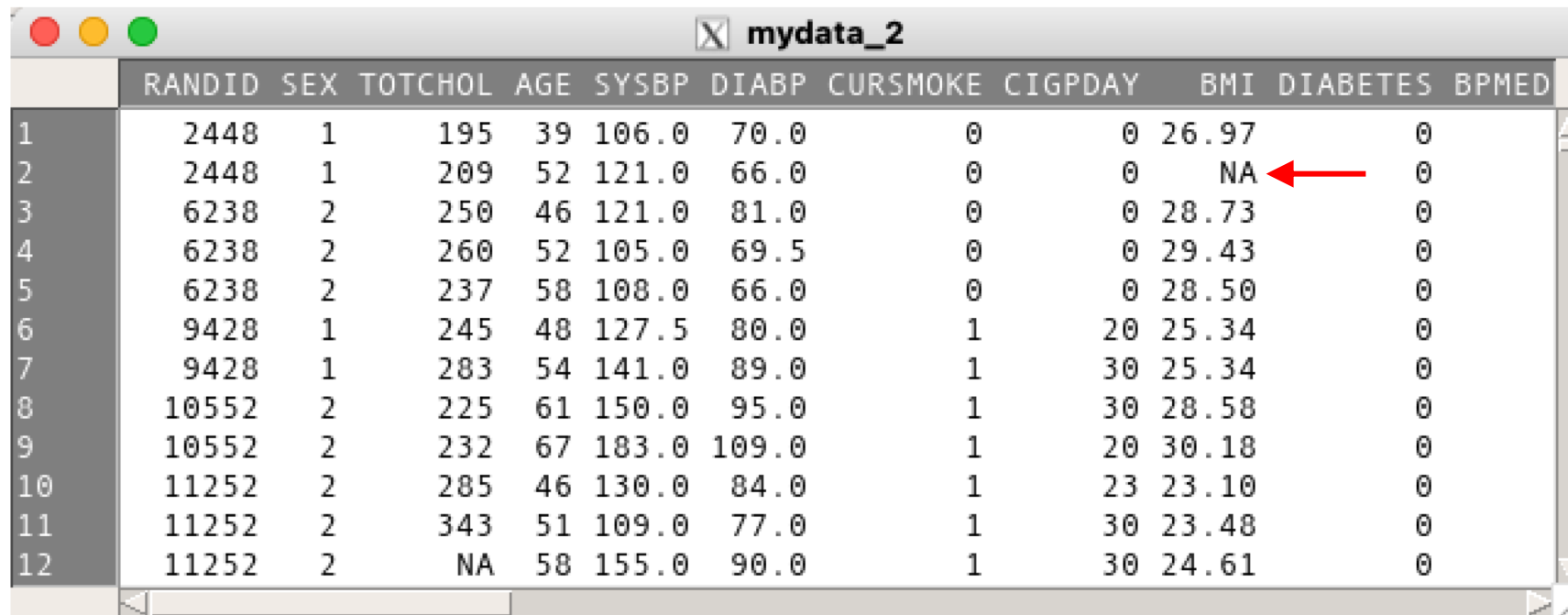
- Menu Bar:** Datei, Bearbeiten, Datenmanagement, Statistik, Grafiken, Modelle, Verteilungen, Extras, Hilfe.
- Datenmatrix:** A list of datasets including `mydata_2`, which is highlighted with a red arrow and the text "the dataset we have just imported becomes our new active dataset".
- Buttons:** "Datenmatrix bearbeiten" and "Datenmatrix betrachten" (highlighted with a red arrow and the text "click here to view the active dataset").
- Modell:** A dropdown menu showing "<Kein aktuelles Modell>".
- R Skript / R Markdown:** A tab for editing scripts or markdown.
- Code Editor:** Contains the R command:

```
mydata_2 <- read.table("/Users/  
header=TRUE, stringsAsFactors=TRUE, sep=",", na.strings="", dec=".", strip.white=TRUE)
```
- Ausgabe:** A tab for viewing the output of the command, showing the same command being executed.
- Meldungen:** A console window at the bottom showing messages:

```
[4] HINWEIS: Die Datenmatrix 'mydata' hat 3263 Zeilen und 23 Spalten.  
[5] HINWEIS: Die Datenmatrix 'mydata_2' hat 11627 Zeilen und 39 Spalten.
```

A red arrow points from the text "extent of the imported dataset" to the second message.

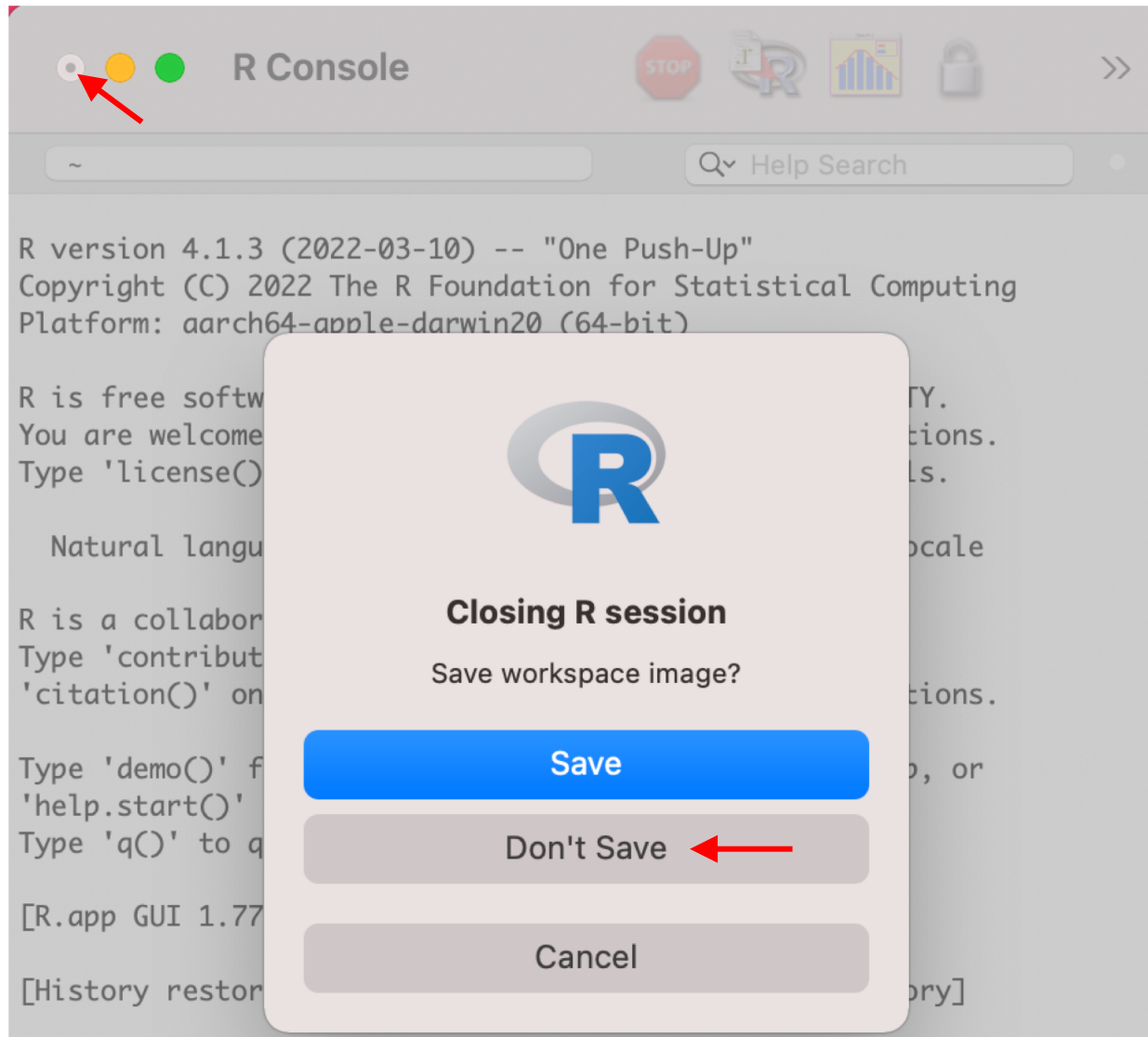
Eine Datenmatrix aus .csv-Datei importieren



	RANDID	SEX	TOTCHOL	AGE	SYSBP	DIABP	CURSMOKE	CIGPDAY	BMI	DIABETES	BPMED
1	2448	1	195	39	106.0	70.0	0	0	26.97	0	
2	2448	1	209	52	121.0	66.0	0	0	NA	0	
3	6238	2	250	46	121.0	81.0	0	0	28.73	0	
4	6238	2	260	52	105.0	69.5	0	0	29.43	0	
5	6238	2	237	58	108.0	66.0	0	0	28.50	0	
6	9428	1	245	48	127.5	80.0	1	20	25.34	0	
7	9428	1	283	54	141.0	89.0	1	30	25.34	0	
8	10552	2	225	61	150.0	95.0	1	30	28.58	0	
9	10552	2	232	67	183.0	109.0	1	20	30.18	0	
10	11252	2	285	46	130.0	84.0	1	23	23.10	0	
11	11252	2	343	51	109.0	77.0	1	30	23.48	0	
12	11252	2	NA	58	155.0	90.0	1	30	24.61	0	

wir können sehen, dass unser zweiter Datensatz
ebenfalls erfolgreich importiert wurde

Das R Programm beenden



- gehen wir zur R Console (Base R, das normalerweise nur im Hintergrund läuft, während wir R Commander verwenden)
- schließen wir sie und wählen wir bei der Frage "Save workspace image?" die Option "Don't Save".
- dadurch werden R Commander und R Console ebenfalls geschlossen.