Data import and export

Lukáš Lehotský and Petr Ocelík

Summary of the last lesson

- Logical conditions or index positions allow for object filtering, when inserted into square brackets
- Square brackets must include all object dimensions
- When filtering data frames, filtering condition usually applies to rows

```
condition <- df[,3] > 1000000
df[ condition , ]
    cars type    price    consumption
1 BMW   3    1200000   6.2
2 Audi   A4    1164000   5.9
```

```
df.sub <- df[ condition , ]</pre>
```

Working directory

- Folder, where all imports and exports are taking place
- Makes data import and export easier
- Functions setwd() and getwd()

```
getwd()
[1] "C:/Users/Lukas/...
```

setwd("C:\\Users\\Lukas\\Documents\\R intro")

```
setwd("C:/Users/Lukas/Documents/R intro")
```

Working directory

- There are few issues/limitations
- R does **not accept single backslash** \ in Win path
 - Typically, this is path copied from the Win explorer
 - It's necessary to replace backslash \ with forwardslash / or double backslash \\
- Sometimes, there's an issue with non-standard letters in the path
 - Typically, in user-name paths in Win
 - Solution is to move R and working directory out

8		RStudio				_ 🗆	x
<u>File Edit Code View Plots</u>	Session Build Debug Profile Tools Help	0					
9 - 🞯 - 🔒 🚔	New Session					🗴 Project: (None) 🝷
🕑 Untitled1 🗴	Interrupt R			Environmen	t History		
🔷 🖒 🔎 🗐 🔚 🗌 Source (Terminate R	📑 Run 🛛 📑 Source	• =	🞯 🔒 🛙	🜁 Import Dataset 👻 🛛 🎻	📃 List 🔹	. C
1	Restart R Ctrl+Shift+F10			💧 Global Er	nvironment •		
	Set <u>W</u> orking Directory	To Source File Location To Files Pane Location					
	Load Workspace				Environment is empty		
	Save Workspace As	Choose Directory Ctrl+Shift+H			Environmenters empty		
	<u>C</u> lear Workspace						
	Quit Session Ctrl+Q						
				Files Plot	s Packages Help Viewer		-0
				OL Install	Update		G
				Name	Description	Version	
				User Library	y		^
1:1 (Top Level) \$			R Script \$	abind	Combine Multidimensional Arrays	1.4-5	8
Console ~/ 🖈	=		-0	acepac	 ACE and AVAS for Selecting Multiple Regression Transformations 	1.4.1	8
>				assertt	hat Easy Pre and Post Assertions	0.2.0	8
				audio	Audio Interface for R	0.1-5	8
				backpo	orts Reimplementations of Functions Introduced Since R-3.0.0	1.1.0	8
				base64	enc Tools for base64 encoding	0.1-3	8
				beepr	Easily Play Notification Sounds on any Platform	1.2	8
				BH	Boost C++ Header Files	1.62.0-1	8
				bindr	Parametrized Active Bindings	0.1	8
				bindro		0.2	8
				bitops		1.0-6	8
				Cairo	R graphics device using cairo graphics library for creating high-quality bitmap (PNG, JPEG TIFF), vector (PDF, SVG, PostScript) and display (X11 and Win32) output		8

Data import

- R Studio visual import
- Functions
 - Structured data (tables)
 - Unstructured data
 - R native files

8	RStudio		_ □	×
File Edit Code View Plots Session Build Debug Profile Tools Help				
💽 🔹 🥶 🖌 🔒 🍰 K Go to file/function 🛛 🔡 🔹 Addins 👻			🖄 Project:	: (None) 💌
Ontitled1 ×	-0	Environment History		
♦ Online of a set	Run 🕒 🕞 Source 🔹 🚍	🕣 🕞 📑 Import Dataset 🔹 🎸	≡ List	- C
		Global Environment	 [Q_	<u> </u>
		Environr	ment is empty	
			=	
			ewer	
		(= =) 🏠 🖨 🔊	Q,	6
		Home Find in Topic		_
				^
1:1 (Top Level) \$	R Script 💲	R Resources	🗷 RStudio	
Console ~/ 🛱	-0	Learning R Online	RStudio IDE Support	
>		CRAN Task Views	RStudio Cheat Sheets	
		R on StackOverflow	RStudio Tip of the Day	
		Getting Help with R	RStudio Packages	
			RStudio Products	
		Manuala		
		Manuals		
		An Introduction to R	The R Language Definition	
		Writing R Extensions	R Installation and	
		R Data Import/Export	Administration R Internals	
			n internais	
		Reference		
		Packages	Search Engine & Keywords	
				~
l		Missellen en Meter	: = 1	*

- Most common format of tabular data
 - CSV comma-separated value
 - TSV tab-separated value
 - TXT plain text file
 - Other formats (XLSX, ...)
- CSV/TSV are most desired
- Essentially, CSV and TSV files are text files

PSPad - [C:\Users\Lukas\Downloads\BALIBOMBING2002_2002.csv] -	×
File Projects Edit Search View Format Tools HTML Settings Window Help(x)	- 8 ×
🗗 🐔 🛃 - 🐔 🚥 🕵 🗔 - 🗲 - 🔚 - 🕞 🍃 🔎 🥳 号 🕀 💷 👰 😓 🖉 🏷 🗶 🗈 🐔 🗙 🗊 🤊 🔄 🔂 🛪	EU, ▼ - >
1BALIBOMBING2002 2002.csv	
0 10 20 30 40 50 60 70 80 90 100 110 12	20
"",151,153,155,160,162,163,175,177,180,183,189,650,1579,176,1549,1564,1576,524,588,1532,1597,590,1503,1512,1513,1529,159	4 ^
151,0,2,2,2,2,2,2,2,2,2,2,2,2,2,0,0,0,0,0,	
153,2,0,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,0,0,0,0,0,0,0,0,0,0,0 155,2,2,0,2,3,3,2,2,2,2,2,0,2,2,2,2,2,2,2,2	
160,2,2,2,0,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2	
162,2,2,3,2,0,3,2,2,2,2,0,0,2,2,2,2,0,0,2,2,0,0,0,0	
163,2,2,3,2,3,0,2,2,2,2,0,0,2,2,2,2,2,2,2,	
175,2,2,2,2,2,0,2,2,2,0,0,2,2,1,2,1,0,0,2,2,0,0,0,0	
177,2,2,2,2,2,2,2,0,2,2,2,0,0,0,0,0,0,0,0	
180,2,2,2,2,2,2,2,2,0,2,2,0,2,0,0,0,0,0,0	
183,2,2,2,2,2,2,2,2,2,0,0,0,2,2,2,2,2,0	
189,2,2,2,2,0,0,0,2,2,0,0,0,2,0,0,0,0,0,0,	
1579,2,2,2,2,2,2,2,2,2,2,2,2,2,0,0,2,0,2,0,	
176,0,2,2,2,2,2,2,0,0,2,0,0,2,0,0,2,2,0,0,0,0,0,0,0	
1549,0,2,0,2,2,2,1,0,0,2,0,0,0,2,0,2,2,0,0,0,2,0,0,0,0	
1564,0,2,2,2,2,2,2,0,0,2,0,0,2,2,2,0,0,0,2,2,0,0,0,0,0,0,0	
1576,0,2,2,2,2,1,0,0,2,0,0,2,2,2,2,0,0,0,0,2,0,0,0,0	
524,0,0,2,2,0,2,0,0,0,0,0,0,0,0,0,0,0,0,2,0,2,0,2,2,2,2,2,2	
588,0,0,2,2,0,2,0,0,0,2,0,0,0,0,0,0,2,0,0,2,0,0,0,0,0,0	
1532,0,0,2,2,2,2,2,0,0,2,0,0,0,2,0,2,0,0,0,0,2,0	
1597,0,0,2,2,2,2,2,0,0,2,0,0,2,2,2,2,2,2,0,0,2,2,2,0,0,2	
590,0,0,0,2,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,	
1503,0,0,0,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,0,0,2,2,2,2,2 1512,0,0,0,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0	
1513,0,0,0,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,	
1529,0,0,0,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0	
1594,0,0,0,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,0,2,2,2,2,0	
1:47/29 [1727]	-

	151	153	155	160	162
151	0	2	2	2	2
153	2	0	2	2	2
155	2	2	0	2	3
160	2	2	2	0	2

- BALIBOMBING2002_2002.csv
 - Matrix
 - Stored as a comma-separated value
 - First row column names (note first cell is empty)
 - First column row names
 - No missing values/texts

- Most common function read.table()
- Import any structured (tabular) data
- Highly customizable many arguments
- Always imports data as data frame

```
df <- read.table( file = "BALIBOMBING2002_2002.csv",
    header = TRUE,
    sep = ",",
    row.names = 1,
    na.strings = "NA",
    stringsAsFactors = FALSE,
    fileEncoding = "UTF-8")
```

ß	RStudio																	_ [×				
ile	Edit Code View Plots Session Build Debug Profile Tools Help ▼ Image: Session Image: Session <t< th=""><th></th><th></th><th></th><th></th><th></th><th></th></t<>																						
- C	OR	4	• •			🍌	Go to f	file/funct	tion	-	Addins 👻										🔋 Project	: (None) 👻	
	lf ×																	Envi	ronment Histor	y Connections			
$\langle \neg$		<i>a</i>	🝸 Filt	er									Q,					a	📊 🛛 🖙 Import	📃 List	- I C		
-	X151	\$	X153	÷ X1	55	X160	÷ X1	62 $^{\circ}$	X163 ÷	X175 [÷]	X177 [÷]	X180 [÷]	X183	÷ 🗙	189	X65	50	Global Environment -					
151		0		2	2	2	2	2	2	2	2	2		2	2	2 /	^	Dat	ta				
153		2		0	2	2	2	2	2	2	2	2		2	2	2		D d	f	27 obs. of 27 variables			
155		2		2	C)	2	3	3	2	2	2		2	2	2							
160		2		2	2	2	0	2	2	2		2		2	2	2							
162		2		2	3		2	0	3	2		2		2	(,							
163		2		2			2	3	0					2									
175																							
177																							
180																							
183																Files	Plots Packa	ges Help Viewer					
189		2		2	2		2	0	0	0		2		0))			Install 💿 Upda				
650		2		2	-		0	0	0	0		0		0		, ,			Name	Description	Version		
1579		2		2	2		2	2	2	2				2				User	Library			^	
<		-		-		-	-	-	-	-	-	-		-		>	~		abind	Combine Multidimensional Arrays	1.4-5	8	
_	ing 1 t	o 15 o	of 27 ent	ries												-			acepack	ACE and AVAS for Selecting Multiple Regression Transformations	1.4.1	8	
Cons	ole	Termi	inal ×													_			antiword	Extract Text from Microsoft Word Documents	1.1	8	
C:/l	Jsers/L	.ukas/	Downlo	oads/ /															assertthat	Easy Pre and Post Assertions	0.2.0	8	
> 0	; 1F <	(- r	read	.tab	le(fil	e =	"BALI	вомві	IG2002	2002.	csv"						audio	Audio Interface for R	0.1-5	8	
+					(hea	der	= TR						,					backports	Reimplementations of Functions Introduced Since F 3.0.0	- 1.1.2	\otimes	
+							= "												base64enc	Tools for base64 encoding	0.1-3	8	
+							.nam												beepr	Easily Play Notification Sounds on any Platform	1.2	8	
+									= "NA										BH	Boost C++ Header Files	1.66.0-1	8	
+							_		ctors										bindr	Parametrized Active Bindings	0.1.1	8	
+ .	+ fileEncoding = "UTF-8")														<u> </u>	bindrcpp	An 'Rcpp' Interface to Active Bindings	0.2.2	8				
>	>															bit	A Class for Vectors of 1-Bit Booleans	1.1-13	8				
															bit64 bitops	A S3 Class for Vectors of 64bit Integers Bitwise Operations	0.9-7	8					
															bitops blob	A Simple S3 Class for Representing Vectors of Bina		8					
															000	Data ('BLOBS')	y 1.1.1						
																			brew	Templating Framework for Report Generation	1.0-6	8	
																			са	Simple, Multiple and Joint Correspondence Analysi	s 0.70	⊗ ∽	

- Most common to use function readLines()
- Can import any unstructured data as a character vector
- Data conversions/mining must follow

Data import: RData files

- R has native data format
- Allows to save whole environment (all objects) or particular object for later use
- Easiest way to access/share previous R work
- Function load()

```
load(file = "balibombing.rdata")
```

R												RStu	idio				_	×
<u>F</u> ile	<u>E</u> dit	<u>C</u> ode	<u>V</u> iew	<u>P</u> lots	Session	<u>B</u> uild <u>D</u> e	ebug <u>P</u> ro	ofile <u>T</u> oo	ols <u>H</u> elp									
•	•	1	•		📥 🍺	Go to file/f	function		• Addins •								🔋 Project:	(None) 👻
	df ×														Environment History	Connections		
		a	🝸 Filte	r							Q				🕣 📊 🖙 Import Data	set 🗸 🔏	📃 List	1.00
-	X151	۵.	(153	X155	÷ X160	÷ X162	° X163	° X175	÷ X177	X180	X183	° X189	° X65	0	🜗 Global Environment 👻		Q,	
15		0		2	2	2	2	2	2	2 2	2	2	2 ^	•	Data			
153		2	(D	2	2	2	2	2	2 2	2	2	2		⊘df	27 obs. of 27 variables		
15		2		2	0	2	3	3	2	2 2	2	2	2					
16		2		2	2	0	2	2			2	2	2					
162		2		2	3	2	0	3			2	2	0					
16		2		2	3	2	3	0			2	2	0					
17		2		2	2	2	2	2			2	2	0					
17		2		2	2	2	2	2			2	2	2					
180		2		2	2	2	2	2			0	2	2					
18		2		2	2	2	2	2			2	0	0		Files Plots Packages	Help Viewer		
							0	0							Install O Update	Q		
189		2		2	2	2		-			2	0	0		Name	Description	Version	
650		2		2	0	0	0	0			D -	0	0		User Library			^
1579	2	2		2	2	2	2	2	2	2 2	2	2	2 ~	•	abind	Combine Multidimensional Arrays	1.4-5	\otimes
<													>		acepack	ACE and AVAS for Selecting Multiple Regression	1.4.1	8
Sho	ving 1 t	o 15 of	27 entri	es												Transformations		
Cor	sole	Termir	nal ×												antiword	Extract Text from Microsoft Word Documents	1.1	8
C:/	Users/L	.ukas/[Downloa	ads/ 🔿											assertthat	Easy Pre and Post Assertions	0.2.0	8
>	str(df)											^		audio	Audio Interface for R Reimplementations of Functions Introduced Since R-	0.1-5	
	ata.		me':	27	7 obs.	of 2	7 var:	iables	:						backports	3.0.0	1.1.2	8
\$	X15	51 :	int	0 2	2 2 2 2	222	222								base64enc	Tools for base64 encoding	0.1-3	8
\$	X15	33:	int	26	9222	222	222	•••							beepr	Easily Play Notification Sounds on any Platform	1.2	8
\$	X15	55 :	int	2 2	2023	332	222								BH	Boost C++ Header Files	1.66.0-1	\otimes
\$	X16	60:	int	2 2	2202	222	222							1	bindr	Parametrized Active Bindings	0.1.1	8
			int		2320		222	• • •							bindrcpp	An 'Rcpp' Interface to Active Bindings	0.2.2	8
			int		2323		222								bit	A Class for Vectors of 1-Bit Booleans	1.1-13	8
			int	2 2			2 2 2								bit64 bitops	A S3 Class for Vectors of 64bit Integers Bitwise Operations	0.9-7	8
			int	2 2			022	•••							blob	A Simple S3 Class for Representing Vectors of Binary		8
	X18		int				202	•••								Data ('BLOBS')		w
			int		2222			•••							brew	Templating Framework for Report Generation	1.0-6	8
Þ	×18		int		2220		2 2 0	•••					~	•	Са	Simple, Multiple and Joint Correspondence Analysis	0.70	8 ¥

	RStudio — <u>File Edit Code View Plots Session Build Debug Profile Tools H</u> elp																					
_		N.C.	DI 1	C	D. 11.1	D 1	D (1)	T 1				Not	uulo									
v • •	8 4				🔶 Go to fi	ile/functior	1	∷ • A	ddins 👻												🐴 Proje	ect: (None) 👻
🗍 df ×													_		Envi	ironment Hi	istory	Connect	ions			
		C elle			_										1	🕞 🛛 🖙 İmp	port Data	set 🕶	1		Ξι	ist 🕶 🛛 🕲
1 X	151 🔅 🕽	X153 🔅	X155	÷ X160	0 [÷] X16	52 [÷] X1	63 🗘	X175 🔅	X177 $^{\circ}$	X180	[©] X183	÷ X189	÷ 🗙	650	4	Global Environr	ment 🝷				Q,	
151	U	2		2	2	2	2	۷	2		2	2	2		Dat	ta						
153	2	0		2	2	2	2	2	2		2	2	2		Od	f		27	obs.	of 27 variables		
155	2	2		0	2	3	3	2	2		2	2	2									
160	2	2		2	0	2	2	2	2		2	2	2									
162	2	2		3	2	0	3	2	2		2	2	0									
163	2	2		3	2	3	0	2	2		2	2	0									
	2						-															
175	-	2		2	2	2	2	0	2		2	2	0									
177	2	2		2	2	2	2	2	0		2	2	2									
180	2	2		2	2	2	2	2	2		0	2	2									_
183	2	2		2	2	2	2	2	2		2	0	0		Files		ackages	Help	Viewer			
189	2	2		2	2	0	0	0	2		2	0	0		O		Jpdate	Deseri	at a		Q,	
650	2	2		0	0	0	0	0	0		0	0	0			Name		Descri	ption		Version	^
1579	2	2		2	2	2	2	2	2		2	2	2	~	User	Library						
<													>		<u> </u>	abind				dimensional Arrays	1.4-5	0
Showing	1 to 15 of	27 entrie	s													acepack			nd AVAS ormation	for Selecting Multiple Regression s	1.4.1	\otimes
Console	Termi	nal ×												_		antiword		Extrac	t Text fro	m Microsoft Word Documents	1.1	8
	rs/Lukas/I		ls/ 🔊						_					/		assertthat				ost Assertions	0.2.0	8
> st	r(df)													^		audio			Interface		0.1-5	8
	a.fra	me':	27	obs	. of	27 va	ariab	les:								backports		Reimp 3.0.0	lementat	ions of Functions Introduced Sin	ce R- 1.1.2	\otimes
	191 :								•							base64enc				4 encoding	0.1-3	8
	153 :				222			-								beepr				fication Sounds on any Platform	1.2	8
\$ X					332			•								BH				ader Files	1.66.0-1	_
	160 :		22				2	•						-		bindr				ctive Bindings	0.1.1	8
\$ X			2 2			222	2	•								bindrcpp bit				face to Active Bindings tors of 1-Bit Booleans	0.2.2	8
\$ X			2 2			222	2	•								bit bit64				Vectors of 64bit Integers	0.9-7	8
\$ X				22		922		•								bitops			e Operati	-	1.0-6	8
\$ X		int	22 22	2 2	2 2 2	202 220	2 2	•								blob		A Sim		ass for Representing Vectors of Bi		0
\$ X			22			222		•								brew				mework for Report Generation	1.0-6	8
\$ X	189 :	int	2 2	22	000	922	0	•						~		са		Simple	e, Multipl	e and Joint Correspondence Ana	ysis 0.70	⊗ ∽
															_							

Data conversions

- Data frame may be converted to other object types
 - Function as.matrix() turns DF to matrix
 - Function as.list() will turn DF into list of vectors
- Other object types may be converted to a data frame as well
 - Function as.data.frame()







Data conversions: DF to matrix

- Data are loaded as data frames by default for networks, conversion is necessary
- Class of data needs to be considered beforehand
 - If text included, the whole matrix will be coerced to character
 - Conversion of data classes after the DF conversion is painful – needs to be done column by column

```
mat <- as.matrix(df)</pre>
```

R	RStudio e <u>E</u> dit <u>C</u> ode <u>V</u> iew <u>P</u> lots <u>S</u> ession <u>B</u> uild <u>D</u> ebug <u>P</u> rofile <u>T</u> ools <u>H</u> elp																_	×
<u>F</u> ile	<u>E</u> dit	<u>C</u> ode	View	<u>P</u> lots	Session E	<u>B</u> uild <u>D</u> e	bug <u>P</u> rofil	e <u>T</u> ools	<u>H</u> elp									
• •	0	1	-		-	Go to file/fu	unction	-	Addins 👻								🔋 Project:	(None) 👻
	df ×												$-\Box$	Er	vironment History	Connections		-
		a l	𝖓 Filter								Q,			6	🕈 📊 🖙 Import Data	set 👻 💰	List	- C
-	X151	÷]	X153 [‡]	X155	[‡] X160	[÷] X162	° X163 °	X175 [÷]	X177 [‡]	X180 [‡]	X183	° X189	÷ X650		Global Environment 👻		Q,	
151		0	2		2	2	2 2	2	2	2		2	2 ^	Da	ata			
153		2	0		2	2	2 2	2	2	2		2	2	0	df	27 obs. of 27 variables		
155		2	2		0	2	3 3	2	2	2		2	2					
160		2	2		2	0	2 2	2	2	2		2	2					
162		2	2		3	2	0 3	2	2	2		2	0					
163		2	2				з с			2		2	0					
175		2	2				2 2			2		2	0					
177		2	2				2 2			2		2	2					
180		2	2				2 2			0		2	2					
183		2	2				2 2			2		0	0	Fi	les Plots Packages	Help Viewer		-
189		2	2				0 0			2		0	0		Install 💿 Update	٩		
650		2	2				0 0			0		0	0		Name	Description	Version	
		2	2				2 2			2		2	2	Us	er Library			^
1579		2	2		2	2	2 2	2	2	2		2	2 🗸		-	Combine Multidimensional Arrays	1.4-5	8
<													>			ACE and AVAS for Selecting Multiple Regression	1.4.1	8
Show	ving 1 t	:o 15 of	27 entries	S												Transformations		
Con	sole	Termi	nal ×												antiword	Extract Text from Microsoft Word Documents	1.1	8
C:/	Users/l	Lukas/I	Download	ls/ 🔿									1		assertthat audio	Easy Pre and Post Assertions Audio Interface for R	0.2.0	8
>	str((df)											^	H	backports	Reimplementations of Functions Introduced Since R-	1.1.2	8
'd	ata.	fra	me':	27	obs.	of 2	7 varia	bles:							j buckports	3.0.0	1.1.2	
	X15		int	02			222.) base64enc	Tools for base64 encoding	0.1-3	8
\$	X15	53	int	20			222.									Easily Play Notification Sounds on any Platform	1.2	8
	X15		int	22			222.									Boost C++ Header Files	1.66.0-1	8
	X16		int				222.								bindr	Parametrized Active Bindings	0.1.1	8
	X16		int				222.									An 'Rcpp' Interface to Active Bindings	0.2.2	8
	X16	-	int	2 2			222.	••						F	bit64	A Class for Vectors of 1-Bit Booleans A S3 Class for Vectors of 64bit Integers	0.9-7	8
	X17		int	22			222.	•••						F	bitops	Bitwise Operations	1.0-6	8
	X17 X18		int	22			922.	••							blob	A Simple S3 Class for Representing Vectors of Binary	1.1.1	8
	X18		int int	22 22			202. 220.	••								Data ('BLOBS')		
	X18		int	22			220. 220.								brew	Templating Framework for Report Generation	1.0-6	8
ې •			• •		2 2 0			•••					~) ca	Simple, Multiple and Joint Correspondence Analysis	0.70	8 ¥

Data conversions: matrix to DF

- Inconvenient process
- Argument stringsAsFactors is important otherwise any text becomes factor variable
- Data class of matrix will be preserved
 - No automatic recognition built in on conversion
- If matrix is unnamed, column names are **assigned automatically**

mat2df <- as.data.frame(mat)</pre>

mat2df <- as.data.frame(mat, stringsAsFactors = FALSE)</pre>

Names of rows and columns

- Column names are an issue in data frames
- Data frame columns must have proper column (variable) names
 - Names starting with number are not allowed (default transformation behavior to "X#")
 - Missing column names lead to **automatically generated** names (V1, V2, ..., Vn)
 - Conversion (e.g. from list) might result into crazy column names

Names of rows and columns

- If issue with names is present, renaming needs to follow
- Functions rownames () and colnames () allow to access existing names
- Renaming is counterintuitive we need to assign vector with names to the function result
- It must be of a **same length** as number of columns

colnames(df) <- c("var1" , "var2" , ... , "varN")</pre>

Names of rows and columns

- Names may be generated automatically using paste() or paste0() functions – these functions collapse text bits together
 - Function paste() adds spaces between texts
 - Function paste0 () collapses texts without spaces

colnames(df) <- paste0("actor" , 1:27)</pre>

 In case of a matrix, each dimension needs to be named separately

colnames(mat) <- paste0("actor" , 1:27)</pre>

colnames(mat) <- rownames(mat)</pre>

Data export

- Export of entire workspace
- Export of particular objects
 - R native files
 - Structured data (tables)
 - Unstructured data

Data export: saving entire workspace

- Function save.image()
 - Rdata format
 - No additional arguments required

save.image(file = "workspace_2018_07_04.rdata")

B	RStudio Edit <u>C</u> ode <u>V</u> iew <u>P</u> lots <u>S</u> ession <u>B</u> uild <u>D</u> ebug <u>P</u> rofile <u>T</u> ools <u>H</u> elp																_ □	×			
<u>F</u> ile <u>E</u> d	it <u>C</u> ode	<u>V</u> iew	<u>P</u> lots	<u>S</u> essio	n <u>B</u> u	uild <u>D</u> e	bug j	<u>P</u> rofile	e <u>T</u> ool	s <u>H</u> elp											
0 - 0	8 🗠	-		- [🔶 G	o to file/fu	inction		-	Addins	•									🚯 Project	: (None) 🝷
df 3															-		Envi	ronment History			
	1	🝸 Filte	r										Q,				e	📊 🛛 🖙 Import Data	📃 List		
^ x	151 🔅 🛛	X153	×155	÷ X10	50 ÷	X162	÷ X16	з÷	X175	÷ X177	÷ X180	÷ x	83 0	X189	÷ X6	50	-	Global Environment 👻		Q	
151	0		2	2	2		2	2		2	2	2	2		2		Dat				
153	2		0	2	2		2	2		2	2	2	2		2		Od	f	27 obs. of 27 variables		
155	2		2	0	2		3	3		2	2	2			2						
160	2		2	2	0		2	2		2	2	2	2		2						
162	2		2	3	2		0	3		2	2	2	2		0						
163	2		2	3	2		3	0		2	2	2	2		0						
175	2		2	2	2		2	2		0	2	2	2		0						
177	2		2	2	2		2	2		2	0	2	2		2						
180	2		2	2	2		2	2		2	2	0	2		2						
183	2		2	2	2		2	2		2	2	2	0		0		Files	Plots Packages	Help Viewer		
189	2		2	2	2		0	0		0	2	2	0		0		0	install 💿 Update	Q		16
650	2		2	0	0		0	0		0	0	0	0		0			Name	Description	Version	
1579	2		2	2	2		2	2		2	2	2	2		2		User	Library			^
<														-	>	*		abind	Combine Multidimensional Arrays	1.4-5	8
Showing	1 to 15 of	27 entr	ies															acepack	ACE and AVAS for Selecting Multiple Regression Transformations	1.4.1	8
Consol	Termi	nal ×																antiword	Extract Text from Microsoft Word Documents	1.1	8
C:/Use	rs/Lukas/I	Downlo	ads/ 🔿													/	<u> </u>	assertthat	Easy Pre and Post Assertions	0.2.0	8
> Vi	ew(df)																audio backports	Audio Interface for R Reimplementations of Functions Introduced Since R	0.1-5 - 1.1.2	8
>																		backports	3.0.0	- 1.1.2	~
																		base64enc	Tools for base64 encoding	0.1-3	8
																		beepr	Easily Play Notification Sounds on any Platform	1.2	\otimes
																		BH	Boost C++ Header Files	1.66.0-1	\otimes
																		bindr	Parametrized Active Bindings	0.1.1	\otimes
																		bindrcpp	An 'Rcpp' Interface to Active Bindings	0.2.2	\otimes
																		bit	A Class for Vectors of 1-Bit Booleans	1.1-13	\otimes
																		bit64	A S3 Class for Vectors of 64bit Integers	0.9-7	8
																		bitops	Bitwise Operations	1.0-6	
																		blob	A Simple S3 Class for Representing Vectors of Binar Data ('BLOBS')	y 1.1.1	\otimes
																		brew	Templating Framework for Report Generation	1.0-6	8
																		са	Simple, Multiple and Joint Correspondence Analysis	0.70	⊗ ∨

Data export: saving particular object as Rdata

- Function save()
- Rdata is R native format
 - Preserves the object as is, with its object name
 - Allows to avoid export and import issues

```
save(df,
    file = "BALIBOMBING_export.rdata")
```

Data export: saving tabular data

- Function write.table()
- Object which is exported has to be specified
- Otherwise similar to read.table()

```
write.table(df,
    "BALIBOMBING_export.csv",
    sep = ",",
    row.names = TRUE,
    col.names = TRUE,
    fileEncoding = "UTF-8")
```

Data export: saving unstructured data

- Function writeLines()
- Similar to readLines ()
- Basically no arguments

Other input and output options

- Built-in CSV-specific functions
 - read.csv() and write.csv() functions
 - read.csv() requires less arguments than read.table() but has more limited functionality
- Libraries "xlsx" or "openxlsx" for XLSX input
 - Both packages contain functions read.xlsx() and write.xlsx() - similar to read.table()
- Package "readtext" designed to read unstructured text data (PDFs, Word documents, etc.)
- Other packages for specific formats (XML, HTML, Json, SQL, ...)

Practice 1

- Download the package "MEB433_434_03A_practice.zip" into your computer
 - Unpack it into any folder
 - Set the folder as a working directory
- Import the data "nrg_122a.csv" into R
 - Explore and adjust the data set in a text editor, if necessary
 - Achieve the following
 - Dataset has proper column names
 - Missing values are treated correctly (find function arguments)
 - All columns are numeric (there's a catch here)

Practice 2

- Get rid of the summary rows in the data frame
- Get rid of the missing values for the last year
- Extract only the last 3 years into a separate object
- Export the subset in rdata format
- Export the subset in CSV format