

Psychology 350: Special Topics

An introduction to R for psychological research

Solving Conflicts between packages

William Revelle
Northwestern University
Evanston, Illinois USA

<https://personality-project.org/courses/350>



NORTHWESTERN
UNIVERSITY

Spring, 2024



Outline



Specifying packages at startup

Create a .First function and then quit and save your workspace

R code

```
.First <-  
function() {library(psych)  
  library(psychTools)  
  now <- as.POSIXlt(Sys.time())$hour  
  if (now < 12) {  
    cat("Good morning Bill. \nAre you ready to have fun?")  
  } else { if (now < 18) {cat("Good afternoon Bill.\nAre you  
    ready to have fun?") } else {  
    cat("Good evening Bill. \nAre you ready to have fun?")}  
  }  
}}
```



Active packages using SessionInfo

R code

```
sessionInfo()
```

Good afternoon Bill.

Are you ready to have fun?

attached base packages:

```
[1] stats      graphics  grDevices  utils      datasets  methods   base
```

other attached packages:

```
[1] psychTools_2.4.3 psych_2.4.3
```

loaded via a namespace (and not attached):

```
[1] compiler_4.3.3  tools_4.3.3      parallel_4.3.3   foreign_0.8-86   R.methodsS3_1.8.2
[6] nlme_3.1-164    mnormt_2.1.1     grid_4.3.3       rtf_0.4-14.1     R.oo_1.25.0
[11] lattice_0.22-5
```



describe is a psych package function

R code

```
describe(sat.act)
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
gender	1	700	1.65	0.48	2	1.68	0.00	1	2	1	-0.61	-1.62	0.02
education	2	700	3.16	1.43	3	3.31	1.48	0	5	5	-0.68	-0.07	0.05
age	3	700	25.59	9.50	22	23.86	5.93	13	65	52	1.64	2.42	0.36
ACT	4	700	28.55	4.82	29	28.84	4.45	3	36	33	-0.66	0.53	0.18
SATV	5	700	612.23	112.90	620	619.45	118.61	200	800	600	-0.64	0.33	4.27
SATQ	6	687	610.22	115.64	620	617.25	118.61	200	800	600	-0.59	-0.02	4.41



Debugging conflicts

1. R has many $> 21,000$ packages. Some of these conflict with each other.
2. Some packages have the same name for functions (e.g., `describe` in *psych* and `describe` in *Hmisc*).
3. depending upon the order in which packages are loaded, the other package will be *masked*.

R code

```
library(psych)  
library(Hmisc)
```

Attaching package: 'Hmisc'

The following object is masked from 'package:psych':

`describe`

The following objects are masked from 'package:base':

`format.pval`, `units`



Hmisc conflicts with psych

R code

```
library(Hmisc)
```

```
library(Hmisc)
sessionInfo()
```

Attaching package: 'Hmisc'

The following object is masked from 'package:psych':

describe

The following objects are masked from 'package:base':

format.pval, units

other attached packages:

```
[1] Hmisc_5.1-0          psychTools_2.4.3 psych_2.4.3
```

loaded via a namespace (and not attached):

[1] gtable_0.3.3	dplyr_1.1.1	compiler_4.3.3	rpart_4.1.23	tidyselect_1.2.0
[6] htmlTable_2.4.1	stringr_1.5.0	parallel_4.3.3	gridExtra_2.3	cluster_2.1.6
[11] scales_1.2.1	fastmap_1.1.1	lattice_0.22-5	ggplot2_3.4.1	R6_2.5.1
[16] generics_0.1.3	Formula_1.2-5	knitr_1.45	htmlwidgets_1.6.2	backports_1.4.1
[21] checkmate_2.2.0	tibble_3.2.1	munsell_0.5.0	nnet_7.3-19	pillar_1.9.0
[26] rlang_1.1.0	utf8_1.2.3	stringi_1.7.12	xfun_0.41	cli_3.6.1
[31] magrittr_2.0.3	digest_0.6.31	grid_4.3.3	rstudioapi_0.14	base64enc_0.1-3
[36] lifecycle_1.0.3	nlme_3.1-164	R.methodsS3_1.8.2	R.oo_1.25.0	vctrs_0.6.1
[41] mnormt_2.1.1	data.table_1.14.8	evaluate_0.20	glue_1.6.2	rtf_0.4-14.1
[46] fansi_1.0.4	colorspace_2.1-0	rmarkdown_2.21	foreign_0.8-86	htmltools_0.5.5
[51] tools_4.3.3	pkgconfig_2.0.3			

Now, the describe is the Hmisc function

R code

```
describe(sat.act)
```

```
sat.act
```

```
6 Variables      700 Observations
```

```
gender
```

n	missing	distinct	Info	Mean	Gmd
700	0	2	0.685	1.647	0.4574

Value	1	2
Frequency	247	453
Proportion	0.353	0.647

```
education
```

n	missing	distinct	Info	Mean	Gmd
700	0	6	0.922	3.164	1.532

Value	0	1	2	3	4	5
Frequency	57	45	44	275	138	141
Proportion	0.081	0.064	0.063	0.393	0.197	0.201

For the frequency table, variable is rounded to the nearest 0.05

```
age
```

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50
700	0	48	0.994	25.59	9.56	17	18	19	22
.90	.95								
39	46								

```
lowest : 13 14 15 16 17, highest: 57 58 61 62 65
```

```
ACT
```

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25



Can directly reference the package and function

R code

```
psych::describe(sat.act)
```

```
psych::describe(sat.act)
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
gender	1	700	1.65	0.48	2	1.68	0.00	1	2	1	-0.61	-1.62	0.02
education	2	700	3.16	1.43	3	3.31	1.48	0	5	5	-0.68	-0.07	0.05
age	3	700	25.59	9.50	22	23.86	5.93	13	65	52	1.64	2.42	0.36
ACT	4	700	28.55	4.82	29	28.84	4.45	3	36	33	-0.66	0.53	0.18
SATV	5	700	612.23	112.90	620	619.45	118.61	200	800	600	-0.64	0.33	4.27
SATQ	6	687	610.22	115.64	620	617.25	118.61	200	800	600	-0.59	-0.02	4.41



Multiple ways to read in the data can lead to confusion

Various problems with data structures and *tidyverse*

1. Core-R `read.table` vs `read.csv` (need to specify `header=TRUE`)
2. *psych* `read.file` will read csv, text, etc. depending upon suffix
3. *readr* in *tidyverse* makes ~~into a tibble~~

R code

```
fn <- file.choose()    #go find the file if on your machine
#or specify it

fn <- "https://personality-project.org/r/datasets/simulation.txt"
test1 <- read.table(fn)
dim(test1)
test1 <- read.table(fn, header=TRUE)
dim(test1)
test <- read.file(fn)
dim(test)
test.tidy <- read_delim(fn)
dim(test.tidy)
```



```
fn <- file.choose()    #go find the file if on your machine
#or specify it
fn <- "https://personality-project.org/r/datasets/simulation.txt"
test1 <- read.table(fn)
dim(test1)
test1 <- read.table(fn, header=TRUE)
dim(test1)
test <- read.file(fn)
dim(test)
test.tidy <- read_delim(fn)
dim(test.tidy)
```

```
> fn <- "https://personality-project.org/r/datasets/simulation.txt"
> test1 <- read.table(fn)
> dim(test1)
[1] 73  7
> test1 <- read.table(fn, header=TRUE)
> dim(test1)
[1] 72  7
> test <- read.file(fn)
Data from the .txt file https://personality-project.org/r/datasets/simulation.txt has been loaded
> dim(test)
[1] 72  7
> test.tidy <- read_delim(fn)
Rows: 72 Columns: 7
Column specification
Delimiter: "\t"
dbl (7): Time, Anxiety, Impulsivity, sex, Arousal, Tension, Performance
```

Use `spec()` to retrieve the full column specification for this data.
Specify the column types or set `show\_col\_types = FALSE` to quiet this message.



Tibbles drop row names!

```
headTail(test)
```

	Time	Anxiety	Impulsivity	sex	Arousal	Tension	Performance
1	9	4	9	1	50	55	40
2	19	8	8	1	70	64	90
3	9	5	10	2	50	69	48
4	9	4	1	2	57	55	68
...	...	...	...	...	...	...	...
69	19	6	1	1	66	53	88
70	9	5	10	2	48	63	40
71	19	6	8	2	69	60	95
72	19	10	1	2	66	48	93

versus

```
headTail(test.tidy)
```

	Time	Anxiety	Impulsivity	sex	Arousal	Tension	Performance
1	9	4	9	1	50	55	40
2	19	8	8	1	70	64	90
3	9	5	10	2	50	69	48
4	9	4	1	2	57	55	68
5	...	...	...	...	...	...	...
6	19	6	1	1	66	53	88
7	9	5	10	2	48	63	40
8	19	6	8	2	69	60	95
9	19	10	1	2	66	48	93



Tibbles sometime break some psych functions

R code

```
violinBy(test[2:4], grp="sex", grp.name=TRUE, las=3)
violinBy(test.tidy[2:4], grp="sex", las=3) #this works but
violinBy(test.tidy[2:4], las=3)          #this does not
```

```
violinBy(test[2:4], grp="sex", grp.name=TRUE, las=3)
> violinBy(test.tidy[2:4], grp="sex", las=3) #this works but
> violinBy(test.tidy[2:4], las=3)
Error in density.default(x[, i], na.rm = TRUE, adjust = adjust, from = minx[i], :
  argument 'x' must be numeric
```



How to find if a data.frame is a tibble and how to fix it

R code

```
str(test.tidy)
class(test.tidy)
```

```
spec_tbl_ [72 x 7] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ Time      : num [1:72] 9 19 9 9 19 19 19 9 19 19 ...
 $ Anxiety   : num [1:72] 4 8 5 4 4 3 5 7 5 7 ...
 $ Impulsivity: num [1:72] 9 8 10 1 8 2 2 2 8 2 ...
 $ sex       : num [1:72] 1 1 2 2 2 2 1 1 1 1 ...
 $ Arousal   : num [1:72] 50 70 50 57 70 69 69 56 69 67 ...
 $ Tension   : num [1:72] 55 64 69 55 50 58 61 54 55 56 ...
 $ Performance: num [1:72] 40 90 48 68 95 88 78 70 78 98 ...
- attr(*, "spec")=
 .. cols(
 ..   Time = col_double(),
 ..   Anxiety = col_double(),
 ..   Impulsivity = col_double(),
 ..   sex = col_double(),
 ..   Arousal = col_double(),
 ..   Tension = col_double(),
 ..   Performance = col_double()
 .. )
- attr(*, "problems")=<externalptr>

class(test.tidy)
[1] "spec_tbl_df" "tbl_df"      "tbl"         "data.frame"
```



Fixing a tibble = make it pure data frame

R code

```
test.fixed <- as.data.frame(test.tidy)
```

```
> str(test.fixed)
'data.frame':      72 obs. of  7 variables:
 $ Time      : num  9 19 9 9 19 19 19 9 19 19 ...
 $ Anxiety   : num  4 8 5 4 4 3 5 7 5 7 ...
 $ Impulsivity: num  9 8 10 1 8 2 2 2 8 2 ...
 $ sex       : num  1 1 2 2 2 2 1 1 1 1 ...
 $ Arousal   : num  50 70 50 57 70 69 69 56 69 67 ...
 $ Tension   : num  55 64 69 55 50 58 61 54 55 56 ...
 $ Performance: num  40 90 48 68 95 88 78 70 78 98 ...

class(test.tidy)
[1] "spec_tbl_df" "tbl_df"      "tbl"         "data.frame"
> class(test)
[1] "data.frame"
> class(test.fixed)
[1] "data.frame"
```

