

rempsyc: Convenience functions for psychology

- 2 Rémi Thériault 10 1
- 1 Department of Psychology, Université du Québec à Montréal, Québec, Canada

DOI: 10.xxxxx/draft

Software

- Review 🗗
- Repository 🗗
- Archive ♂

Editor: Open Journals ♂

Reviewers:

@openjournals

Submitted: 01 January 1970 **Published:** unpublished

License

Authors of papers retain copyrights and release the work under a Creative Commons Attribution 4.0. International License (CC BY 4.0)

Summary

{rempsyc} is an R package of convenience functions that make the analysis-to-publication workflow faster, easier, and less error-prone. It enables nice APA tables exportable to Word (via {flextable}) and easily customizable APA-compliant plots (via {ggplot2}). It makes it easy to run statistical tests, check assumptions, and automate various tasks common in psychology research.

Statement of need

There are many reasons to use R (R Core Team, 2022) for analyzing and reporting data from research studies, such as being compatible with the ideals of open science (Quintana, 2020). However, R has a major downside for novices: its steep learning curve due to its programmatic interface, in contrast to perhaps more user-friendly point-and-click software. Of course, this flexibility is also a strength, as the R community can and does come together to produce packages that make using R increasingly easier and more user-friendly (e.g., the *easystats* ecosystem Lüdecke et al., 2019/2023). The {rempsyc} package (Really Easy Methods for Psychology) contributes to this momentum by providing convenience functions that remove as much friction as possible between your script and your manuscript (in particular, if you are using Microsoft Word).

There are mainly three things that go into a manuscript: text, tables, and figures. {rempsyc} does not generate publication-ready text summarizing analyses; for this, see the {report} package (Makowski et al., 2021/2023). Instead, {rempsyc} focuses on the production of publication-ready tables and figures. Below, I go over a few quick examples of those.

Examples Features

Publication-Ready Tables

Many researchers using R still copy-paste the values from the R console to their manuscript, or retype them manually. Yet, this approach increases the risks of copy-paste and retyping errors so common in psychology. This problem is not trivial given that according to some estimates, up to 50% of articles in psychology have at least one statistical error (Nuijten et al., 2016). Ideally, one should be able to format the table directly in R, and to export it to Word directly.

Formatting your table properly in R is already a time-consuming task, but fortunately several packages take care of this step (e.g., the {broom} or {report} packages, Makowski et al., 2021/2023; Robinson et al., 2022, and there are several others). Exporting these formatted tables to Microsoft Word remains a challenge however. Some packages do export to Word (e.g., Stanley & Spence, 2018), but their formatting is often rigid especially when using analyzes or

table formats that are not supported by default.



- {rempsyc} solves this problem by allowing maximum flexibility: you manually create the data frame exactly the way you want, and then only use the magical function, nice_table(), on the resulting data frame. nice_table() works on any data frame, even non-statistical ones
- 41 like mtcars.
- One of its main benefit however is the automatic formatting of statistical symbols and its integration with other packages. We can for example create a {broom} table and then apply nice_table() on it. It suits particularly well the pipe workflow.
- 45 library(rempsyc)
 46

```
lm(mpg ~ cyl + wt * hp, mtcars) |>
broom::tidy(conf.int = TRUE) |>
nice table(broom = "lm")
```

Term	b	SE	t	p	95% CI
(Intercept)	49.49	3.66	13.51	<.001	[41.97, 57.01]
cyl	-0.37	0.51	-0.72	.479	[-1.41, 0.68]
wt	-7.63	1.52	-5.01	< .001	[-10.75, -4.51]
hp	-0.11	0.03	-3.64	.001	[-0.17, -0.05]
$wt \times hp$	0.03	0.01	3.23	.003	[0.01, 0.04]

```
_{51} We can do the same with a {report} table.
```

```
52 stats.table <- lm(mpg ~ cyl + wt * hp, mtcars) |>
53 report::report() |>
54 as.data.frame()
```

56 nice_table(stats.table)



57

my_table

Parameter	Fit	b	95% CI (b)	t	df	p	β	95% CI (β)
(Intercept)		49.49	[41.97, 57.01]	13.51	27	< .001	-0.18	[-0.36, -0.01]
cyl		-0.37	[-1.41, 0.68]	-0.72	27	.479	-0.11	[-0.42, 0.20]
wt		-7.63	[-10.75, -4.51]	-5.01	27	< .001	-0.62	[-0.85, -0.40]
hp		-0.11	[-0.17, -0.05]	-3.64	27	.001	-0.29	[-0.53, -0.04]
$\mathbf{wt} \times \mathbf{hp}$		0.03	[0.01, 0.04]	3.23	27	.003	0.29	[0.11, 0.47]
AIC	147.01							
AICc	150.37							
BIC	155.80							
R2	0.89							
R2 (adj.)	0.87							
Sigma	2.17							

The {report} package provides quite comprehensive tables, so one may request an abbreviated table with the 'short' argument. For convenience, it is also possible to highlight significant results for better visual discrimination, using the 'highlight' argument.[1] Once satisfied with the table, we can add a title and note.

my_table <- nice_table(
stats.table, short = TRUE, highlight = 0.001,
title = c("Table 1", "A Pretty Regression Model"),
note = c("The data was extracted from the 1974 Motor Trend US magazine.",

"Greyed rows represent statistically significant differences, p < .001."))



Table 1

A Pretty Regression Model

Parameter	b	t	df	p	β	95% CI (β)
(Intercept)	49.49	13.51	27	<.001	-0.18	[-0.36, -0.01]
cyl	-0.37	-0.72	27	.479	-0.11	[-0.42, 0.20]
wt	-7.63	-5.01	27	<.001	-0.62	[-0.85, -0.40]
hp	-0.11	-3.64	27	.001	-0.29	[-0.53, -0.04]
$wt \times hp$	0.03	3.23	27	.003	0.29	[0.11, 0.47]

Note. The data was extracted from the 1974 Motor Trend US magazine.

Greyed rows represent statistically significant differences, p < .001.

- One can then easily save the resulting table to Word with flextable::save_as_docx(), specifying the object name and desired path.
- 71 flextable::save_as_docx(my_table, path = "nice_tablehere.docx")
- Additionally, tables created with nice_table() are {flextable} objects (Gohel & Skintzos,
- ⁷³ 2022), and can be modified as such.[2]

Formattting Results of Analyses

75 {rempsyc} also provides its own set of functions to prepare statistical tables before they can be 76 fed to nice_table() and saved to Word.

t tests

•	Dependent Variable	t	df	p	d	95% CI
	mpg	-3.77	18.33	.001	-1.48	[-2.27, -0.67]
	disp	4.20	29.26	<.001	1.45	[0.64, 2.23]
	drat	-5.65	27.20	<.001	-2.00	[-2.86, -1.12]



Contrasts

```
nice_contrasts(data = mtcars,
response = c("mpg", "disp"),
response =
```

Dependent Variable	Comparison	df	t	p	d	95% CI
	4 - 8	28	3.66	.001	3.59	[2.76, 4.51]
mpg	6 - 8	28	1.29	.207	1.44	[0.85, 1.98]
	4 - 6	28	3.64	.001	2.15	[1.35, 3.04]
	4 - 8	28	-6.04	<.001	-4.80	[-5.84, -3.91]
disp	6 - 8	28	-4.86	<.001	-3.29	[-4.31, -2.27]
	4 - 6	28	-2.70	.012	-1.51	[-2.30, -0.87]

Regressions

```
model1 <- lm(mpg ~ cyl + wt * hp, mtcars)
model2 <- lm(qsec ~ disp + drat * carb, mtcars)

nice_lm(list(model1, model2)) |>
nice_table(highlight = TRUE)
```

Dependent Variable	Predictor	df	b	t	p	sr^2	95% CI
	cyl	27	-0.37	-0.72	.479	.00	[0.00, 0.01]
mna	wt	27	-7.63	-5.01	<.001	.11	[0.01, 0.20]
mpg	hp	27	-0.11	-3.64	.001	.06	[0.00, 0.12]
	wt × hp	27	0.03	3.23	.003	.04	[0.00, 0.10]
	disp	27	-0.01	-1.97	.059	.07	[0.00, 0.20]
2222	drat	27	0.23	0.20	.845	.00	[0.00, 0.01]
qsec	carb	27	1.15	0.72	.479	.01	[0.00, 0.06]
	$drat \times carb$	27	-0.48	-1.08	.289	.02	[0.00, 0.09]



Simple Slopes

model1 <- lm(mpg ~ gear * wt, mtcars)</pre>

```
model2 <- lm(disp ~ gear * wt, mtcars)
my.models <- list(model1, model2)

nice_lm_slopes(my.models, predictor = "gear", moderator = "wt") |>
nice table()
```

Dependent Variable	Predictor (+/-1 SD)	df	b	t	p	Sr^2	95% CI
	gear (LOW-wt)	28	7.54	2.01	.054	.03	[0.00, 0.09]
mpg	gear (MEAN-wt)	28	5.62	1.94	.062	.03	[0.00, 0.08]
	gear (HIGH-wt)	28	3.69	1.80	.083	.02	[0.00, 0.08]
	gear (LOW-wt)	28	50.51	0.67	.511	.00	[0.00, 0.02]
disp	gear (MEAN-wt)	28	35.80	0.61	.545	.00	[0.00, 0.02]
	gear (HIGH-wt)	28	21.08	0.51	.616	.00	[0.00, 0.02]

Correlation Matrices

105

106

107

108

109

111

It is also possible to export a colour-coded correlation matrix to Microsoft Excel. The cormatrix_excel() function has several benefits over conventional approaches. The base R cor() function for example does not use rounded values and the console is impractical for large matrices. One may manually round values and export it to a .csv file, which is an improvement but still unsatisfying.

The {apaTables} package (Stanley & Spence, 2018) allows exporting the correlation matrix to Word in an APA format, and in many cases this is very satisfying for APA requirements.

Hovever, the Word format is not suitable for large matrices, as it will often spread beyond the document's margin limits.

Another approach is to export to an image, like the {correlation} package does (Makowski et al., 2020).[3] For very small matrices, this works extremely well, and the colour is an immense help to quickly identify which correlations are strong or weak, positive or negative, and significant or non-significant. Again, however, this does not work so well for large matrices because labels might overlap or navigating the large figure becomes difficult.

When the goal is more exploratory, rather than reporting, and we have large matrices, it can be more useful to export it to Excel. In {rempsyc}, we combine the idea of using a coloured correlation matrix from the {correlation} package with the idea of exporting to Excel using {openxlsx2} (Barbone & Garbuszus, 2023).

We also provide some usability improvements, like freezing the first row and column so as to be able to easily see which variables correlates with which other variable, regardless of how far or deep those variables are located within the matrix.

The colour represents the strength of the correlation, whereas the stars represent how significant the *p* value is.[4] The exact *p* values are provided in a second tab for reference purposes, so all information is readily available in a convenient format.

```
cormatrix_excel(data = infert,
filename = "cormatrix1",
```



select = c("age", "parity", "induced", "case", "spontaneous",

"stratum", "pooled.stratum"))

	- 4					_	_			
	4	Α .	В	C	D	Е	F	G	H	. '
	1	Paramete		parity	induced	case	spontaneo		pooled.str	atum
	2	age	1.0	.08	10	.0	08	21 ***		
	3	parity	.08	1.0	.45 ***	.01	.31 ***	31 ***	.12	
	4	induced	10	.45 ***	1.0	.02	27 ***	10	.16 *	
	5	case	.0	.01	.02	1.0	.36 ***	.0	.0	
	6	spontaneo	08	.31 ***	27 ***	.36 ***	1.0	.06	.21 ***	
	7	stratum	21 ***	31 ***	10	.0	.06	1.0	.75 ***	
	8	pooled.sti	17 *	.12	.16 *	.0	.21 ***	.75 ***	1.0	
	9									
135			r_value	p_val	ues	\oplus				
	4	Α	В	С	D	E	F	G	н	- 1
	1	Paramete			induced	case	spontaneo	ctrotum	and the state of	
			age	parity	muuceu	case	spontaneo	Stratum	pooled.str	ratum
	2	age	.0	.194	.113	.956	.186	.001	.006	ratum
	3									ratum
		age	.0	.194	.113	.956	.186	.001	.006	ratum
	3	age parity	.0 .194	.194 .0	.113	.956 .889	.186	.001 .0	.006 .059	ratum
	3	age parity induced	.0 .194 .113 .956	.194 .0 .0	.113 .0 .0	.956 .889 .789	.186 .0 .0	.001 .0 .113	.006 .059 .010	ratum
	3 4 5	age parity induced case	.0 .194 .113 .956	.194 .0 .0 .889	.113 .0 .0 .789	.956 .889 .789	.186 .0 .0 .0	.001 .0 .113 .952	.006 .059 .010 .939	ratum
	3 4 5 6	age parity induced case spontaneo	.0 .194 .113 .956 .186	.194 .0 .0 .889	.113 .0 .0 .789	.956 .889 .789 .0	.186 .0 .0 .0	.001 .0 .113 .952 .341	.006 .059 .010 .939	atum
	3 4 5 6 7	age parity induced case spontaneo stratum	.0 .194 .113 .956 .186	.194 .0 .0 .889 .0	.113 .0 .0 .789 .0	.956 .889 .789 .0 .0	.186 .0 .0 .0 .0 .0	.001 .0 .113 .952 .341	.006 .059 .010 .939 .001	atum

Publication-Ready Figures

Preparing figures according to APA style, having them look good, and being able to save them in high-resolution with the proper ratios is often challenging. Working with {ggplot2} (Wickham, 2016) provides tremendous flexibility, but an unintended consequence is that doing even trivial operations can at times be daunting.

This is why {rempsyc} setups a few default plot types, ready to be saved to your preferred format (.pdf, .tiff, or .png).

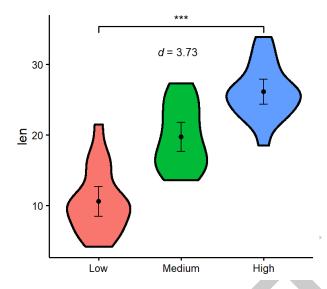
Violin Plots

137

139

140





For an example of such use in publication, see Thériault et al. (2021).

One can easily save the resulting figure with ggplot2::ggsave(), specifying the desired file name, extension, and resolution.

Recommended dimensions for saving {rempsyc} figures is 7 inches wide and 7 inches high at 300 dpi, which makes sure that the resolution is high enough even if saving to non-vector graphics formats like .png. That said, scalable vector graphics formats like .pdf or .eps are still recommended for high-resolution submissions to scientific journals.

Scatter Plots

153

154

155

156

157

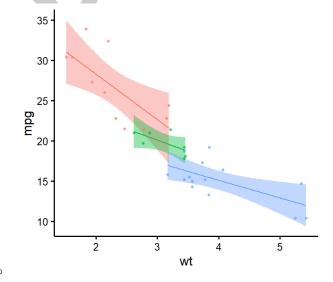
158

160

161

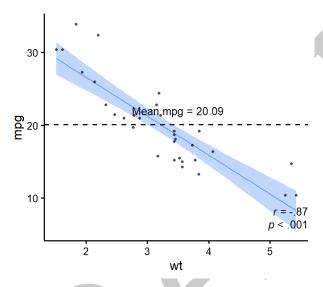
162

Figures are {ggplot2} objects (Wickham, 2016), and can be modified as such.





```
nice_scatter(data = mtcars,
171
                  predictor = "wt",
172
                  response = "mpg",
173
                  has.confband = TRUE,
174
                  has.r = TRUE,
175
                  has.p = TRUE) +
176
      ggplot2::geom_hline(yintercept = mean(mtcars$mpg), colour = "black",
177
                           linewidth = 1.4, linetype = "dashed") +
178
      ggplot2::annotate("text", x = 3.5, y = 22, size = 7,
179
                         label = paste("Mean mpg =", round(mean(mtcars$mpg), 2)))
180
```



For an example of such use in publication, see Krol et al. (2020).

Overlapping Circles

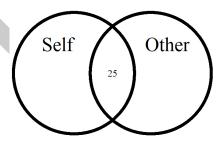
182

183

189

For psychologists using the Inclusion of Other in the the Self Scale (Aron et al., 1992), it can be useful to interpolate the original discrete scores (1 to 7) into a group average representation of the conceptual self-other overlap. For example, assuming the group mean is 3.5 on the 1 to 7 scale:

ss overlap_circle(3.5)



For an example of such use in publication, see Thériault et al. (2021).



Testing assumptions

When comes time to test assumptions of a linear model, the best option is the check_model()
function from easystats' {performance} package, which allows direct visual evaluation of
assumptions (Lüdecke, Ben-Shachar, et al., 2021). Indeed, visual assessment of diagnostic
plots is recommended over statistical tests since they are overpowered in large samples and
underpowered in small samples (Kozak & Piepho, 2018).

That said, if for whatever reason one wants to check objective asumption tests for a linear model, rempsyc makes this easy with the nice_assumptions() function, which provide *p* values for normality (Shapiro–Wilk), homoscedasticity (Breusch–Pagan) and autocorrelation of residuals (Durbin–Watson) in one call.

Categorical Predictors

197

198

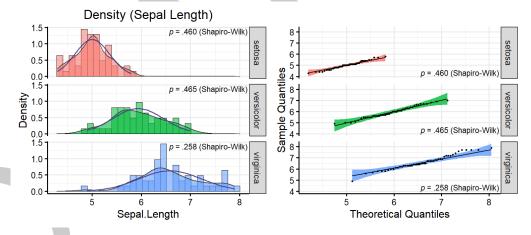
199

201

211

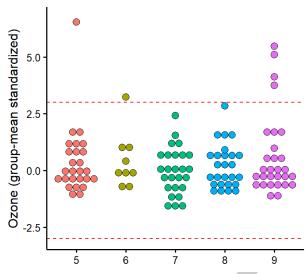
212

nice_normality() makes it easy to visually check normality in the case of categorical predictors
(i.e., when using groups), through a combination of quantile-quantile plots, density plots, and
histograms.



Similarly for univariate outliers using the median absolute deviation (MAD, Leys et al., 2013).

```
plot_outliers(airquality,
group = "Month",
response = "Ozone")
```



```
216
    Univariate outliers based on the MAD can also be simply requested with find_mad().[5]
217
    find_mad(airquality, names(airquality), criteria = 3)
218
   ## 8 outlier(s) based on 3 median absolute deviations for variable(s):
220
        Ozone, Solar.R, Wind, Temp, Month, Day
221
   ##
222
   ## Outliers per variable:
223
   ##
224
   ##
       $0zone
225
   ##
         Row Ozone_mad
226
          30
              3.218284
    ## 1
227
          62
               3.989131
228
          99
               3.488081
       3
229
               3.025573
       4 101
230
   ## 5 117
               5.261028
   ## 6 121 3.333911
232
   ##
233
   ## $Wind
         Row Wind_mad
235
           9 3.049871
236
          48 3.225825
237
    Homoscedasticity can also be checked numerically with nice_var() or visually with
    nice_varplot().
239
    nice_var(data = iris,
240
              variable = names(iris[1:4]),
241
              group = "Species") |>
242
      nice_table()
243
   nice_varplot(data = iris,
244
                  variable = "Sepal.Length",
245
                  group = "Species")
246
```

Utility functions

247

Finally, with the idea of making the analysis workflow easier in mind, $\{\text{rempsyc}\}$ also provides a few other utility functions. nice_na() allows reporting item-level missing values per scale,



- as well as participant's maximum number of missing items by scale, as per recommendations (Parent, 2013).
- extract_duplicates() creates a data frame of only observations with a duplicated ID or
- participant number, so they can be investigated more thoroughly. best_duplicate() allows
- to follow-up on this investigation and only keep the "best" duplicate, meaning those with the
- fewer number of missing values, and in case of ties, the first one.
- $_{256}$ nice_reverse() permits the automatic reverse-coding of scores so common for psychology
- questionnaires, provided the minimum and maximum score values are known.
- 258 There are other functions that the reader can explore at their leisure on the package official
- 259 website. However, hopefully, this overview has given the reader a gentle introduction to this
- 260 package.

61 Licensing and Availability

- The {rempsyc} package is licensed under the GNU General Public License (GPL v3.0). It is
- available on CRAN, and can be installed using install.packages("rempsyc"). The full tutorial
- website can be accessed at: https://rempsyc.remi-theriault.com/. All code is open-source and
- hosted on GitHub, and bugs can be reported at https://github.com/rempsyc/rempsyc/issues/.

266 Acknowledgements

- ²⁶⁷ I would like to thank Jay Olson, Hugues Leduc, Charles-Étienne Lavoie, and Björn Büdenbender
- for statistical or technical advice that helped inform some functions of this package and/or
- useful feedback on this manuscript. I would also like to acknowledge funding from the Social
- Sciences and Humanities Research Council of Canada.

References

- Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the
- 273 structure of interpersonal closeness. Journal of Personality and Social Psychology, 63(4), 596.
- https://doi.org/10.1037/0022-3514.63.4.596
- Barbone, J. M., & Garbuszus, J. M. (2023). openxlsx2: Read, write and edit 'xlsx' files.
- 276 https://github.com/JanMarvin/openxlsx2
- Gohel, D., & Skintzos, P. (2022). Flextable: Functions for tabular reporting. https://CRAN.
- 278 R-project.org/package=flextable
- yo Kozak, M., & Piepho, H.-P. (2018). What's normal anyway? Residual plots are more telling
- than significance tests when checking ANOVA assumptions. Journal of Agronomy and Crop
- ²⁸¹ Science, 204(1), 86-98. https://doi.org/10.1111/jac.12220
- ²⁸² Krol, S. A., Thériault, R., Olson, J. A., Raz, A., & Bartz, J. A. (2020). Self-concept clarity
- ²⁸³ and the bodily self: Malleability across modalities. Personality and Social Psychology Bulletin,
- 284 46(5), 808-820. https://doi.org/10.1177/0146167219879126
- Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use
- standard deviation around the mean, use absolute deviation around the median. Journal of
- 287 Experimental Social Psychology, 49(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013
- Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance:
- An R package for assessment, comparison and testing of statistical models. Journal of Open
- 290 Source Software, 6(60), 3139. https://doi.org/10.21105/joss.03139



- Lüdecke, D., Makowski, D., Ben-Shachar, M. S., Patil, I., Wiernik, B. M., Bacher, E., & Thériault, R. (2023). *easystats: Streamline model interpretation, visualization, and reporting*.
- 293 https://easystats.github.io/easystats/ (Original work published 2019)
- ²⁹⁴ Lüdecke, D., Patil, I., Ben-Shachar, M. S., Wiernik, B. M., Waggoner, P., & Makowski, D.
- ²⁹⁵ (2021). see: An R package for visualizing statistical models. *Journal of Open Source Software*,
- ²⁹⁶ 6(64), 3393. https://doi.org/10.21105/joss.03393
- Makowski, D., Ben-Shachar, M. S., Patil, I., & Lüdecke, D. (2020). Methods and algorithms
- for correlation analysis in R. Journal of Open Source Software, 5(51), 2306. https://doi.org/
- 299 10.21105/joss.02306
- Makowski, D., Lüdecke, D., Patil, I., Thériault, R., Ben-Shachar, M. S., & Wiernik, B. M.
- 301 (2023). report: Automated reporting of results and statistical models. https://easystats.github.
- 302 io/report/ (Original work published 2021)
- Nuijten, M. B., Hartgerink, C. H., Van Assen, M. A., Epskamp, S., & Wicherts, J. M. (2016).
- The prevalence of statistical reporting errors in psychology (1985–2013). Behavior Research
- 305 Methods, 48, 1205–1226. https://doi.org/doi.org/10.3758/s13428-015-0664-2
- Parent, M. C. (2013). Handling item-level missing data: Simpler is just as good. The
- ³⁰⁷ Counseling Psychologist, 41(4), 568–600. https://doi.org/10.1177%2F0011000012445176
- Quintana, D. S. (2020). Five things about open and reproducible science that every early
- career researcher should know. https://osf.io/2jt9u
- R Core Team. (2022). R: A language and environment for statistical computing. R Foundation
- for Statistical Computing. https://www.R-project.org/
- Robinson, D., Hayes, A., & Couch, S. (2022). broom: Convert statistical objects into tidy
- tibbles. https://CRAN.R-project.org/package=broom
- Stanley, D. J., & Spence, J. R. (2018). Reproducible tables in psychology using the apaTables
- 316 //doi.org/10.1177/2515245918773743
- Thériault, R., Olson, J. A., Krol, S. A., & Raz, A. (2021). Body swapping with a Black person
- boosts empathy: Using virtual reality to embody another. Quarterly Journal of Experimental
- ³¹⁹ Psychology, 74(12), 2057–2074. https://doi.org/10.1177/17470218211024826
- Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. Springer-Verlag New York.
- https://ggplot2.tidyverse.org
- [1] This argument can be used logically, as 'TRUE' or 'FALSE', but can also be provided with
- $_{23}$ a numeric value representing the cut-off threshold for the p value
- [2] A great resource for this is the {flextable} e-book: https://ardata-fr.github.io/
- 325 flextable-book/
- $_{226}$ [3] Exporting the correlation matrix to an image through the $\{$ correlation $\}$ package also requires
- the {see} package (Lüdecke, Patil, et al., 2021)
- $_{128}$ [4] For convenience, colours are only used when the corresponding p value is at least smaller
- 29 than .05
- 330 [5] Once one has identified outliers, it is also possible ot winsorize them with the
- winsorize_mad() function.