

Data Wrangling and Visualization 101 in R - Class 2

Florian Jaeger

Fall 2020

Contents

1	Applying these methods to your own data	1
1.1	Importing your data	1
1.1.1	DeAngelis & Haefner group	1
1.1.2	Jaeger group	2
1.2	Plotting your data	2
1.2.1	DeAngelis & Haefner group	2
1.2.2	Jaeger group	3
2	Session info	5

1 Applying these methods to your own data

1.1 Importing your data

1.1.1 DeAngelis & Haefner group

```
## experiment subject trialNumber flowCondition apertureSize
## 1:365 1:243 Min. : 1.00 Control :188 0:164
## 2:312 2:215 1st Qu.: 18.00 Full :166 1:148
## 3:219 Median : 38.00 Global : 86 2:187
## 4: 0 Mean : 44.64 Local : 95 3:178
## 3rd Qu.: 69.00 Opposite: 71
## Max. :131.00 Same : 71
##
## probeEccentricity probeAngle sceneIndex relativeTilt absoluteTilt
## 0:338 -15:213 3 :188 Min. : -26.729 Min. : -22.587
## 1:339 0 :243 6 : 95 1st Qu.: 0.000 1st Qu.: 0.000
## 15 :221 4 : 88 Median : 4.901 Median : 10.252
## 5 : 86 Mean : 7.177 Mean : 7.354
## 7 : 78 3rd Qu.: 15.000 3rd Qu.: 16.600
## 8 : 71 Max. : 44.264 Max. : 42.584
## (Other): 71
## reactionTime stimulusTime probeVelX probeVelY
## Min. : 0.0680 Min. :2 Min. : -0.0517638 Min. : -0.2000
## 1st Qu.: 0.9714 1st Qu.:2 1st Qu.: -0.0517638 1st Qu.: -0.2000
## Median : 1.2694 Median :2 Median : 0.0000000 Median : -0.1932
## Mean : 2.5943 Mean :2 Mean : -0.0006117 Mean : -0.1956
## 3rd Qu.: 2.1624 3rd Qu.:2 3rd Qu.: 0.0517638 3rd Qu.: -0.1932
```

```
## Max. :361.6860 Max. :2 Max. : 0.0517638 Max. : -0.1932
##
## probeStartLocationX probeStartLocationY probeEndLocationX probeEndLocationY
## Min. :1.000 Min. :0 Min. :0.9482 Min. : -0.20000
## 1st Qu.:1.000 1st Qu.:0 1st Qu.:1.0000 1st Qu.: -0.20000
## Median :1.000 Median :0 Median :1.0000 Median : -0.19319
## Mean :1.151 Mean :0 Mean :1.1351 Mean : -0.06728
## 3rd Qu.:1.500 3rd Qu.:0 3rd Qu.:1.4482 3rd Qu.: 0.20000
## Max. :1.500 Max. :0 Max. :1.5000 Max. : 0.20000
##
```

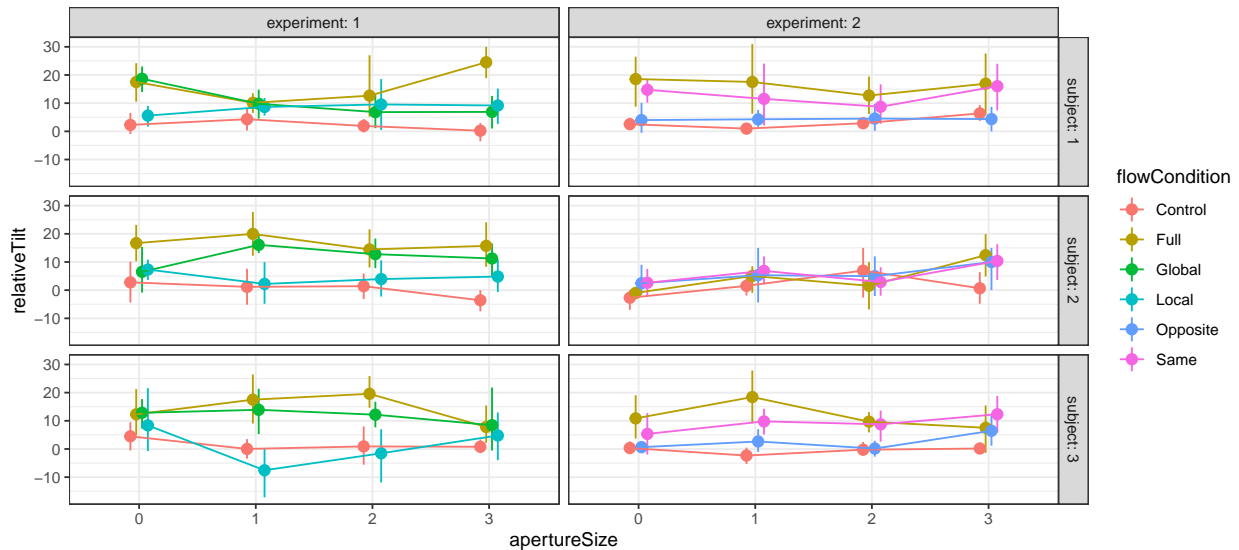
1.1.2 Jaeger group

1.2 Plotting your data

1.2.1 DeAngelis & Haefner group

1.2.1.1 Trial exclusions Are there any criteria that would make you think that a trial should be excluded from further analysis?

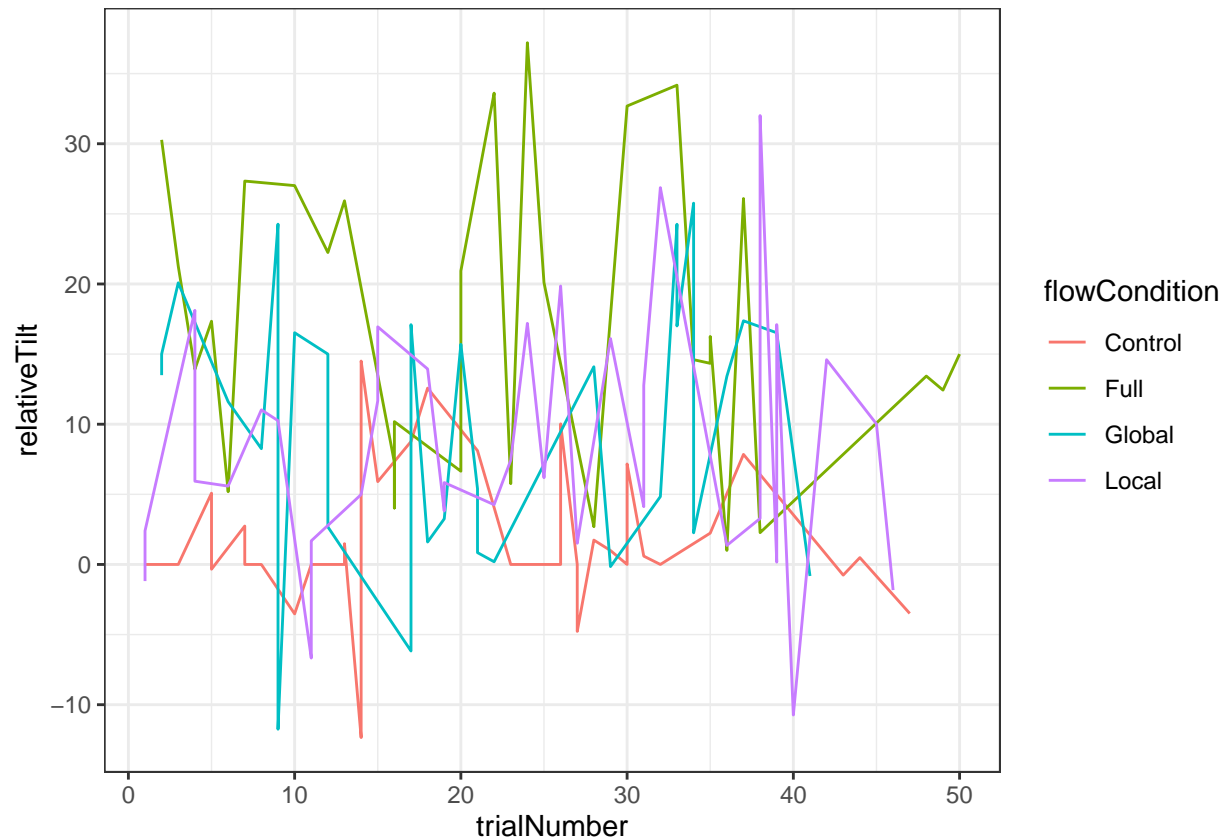
1.2.1.2 Relative tilt Here is a plot for all subjects that builds on what Ji-ze posted on Slack. **How would you go about changing the titles for the two axes and the condition legend?** Hint: use scales! The R primer I mentioned in the tutorial for the first class talks about scales, too, if you prefer an introduction to reading help files.



Now let's zoom in on Subject 2 in Experiment 1. Plot the relative tilt as a function of the aperture size and condition, as in the above plot (but only for that subject). **Now think about how you can further show the data separated by probe's angle (probeAngle, 3 values) and eccentricity (probeEccentricity, 2 values).** Hint: you might use shape, transparency (alpha), or faceting to show the data split by additional variables.

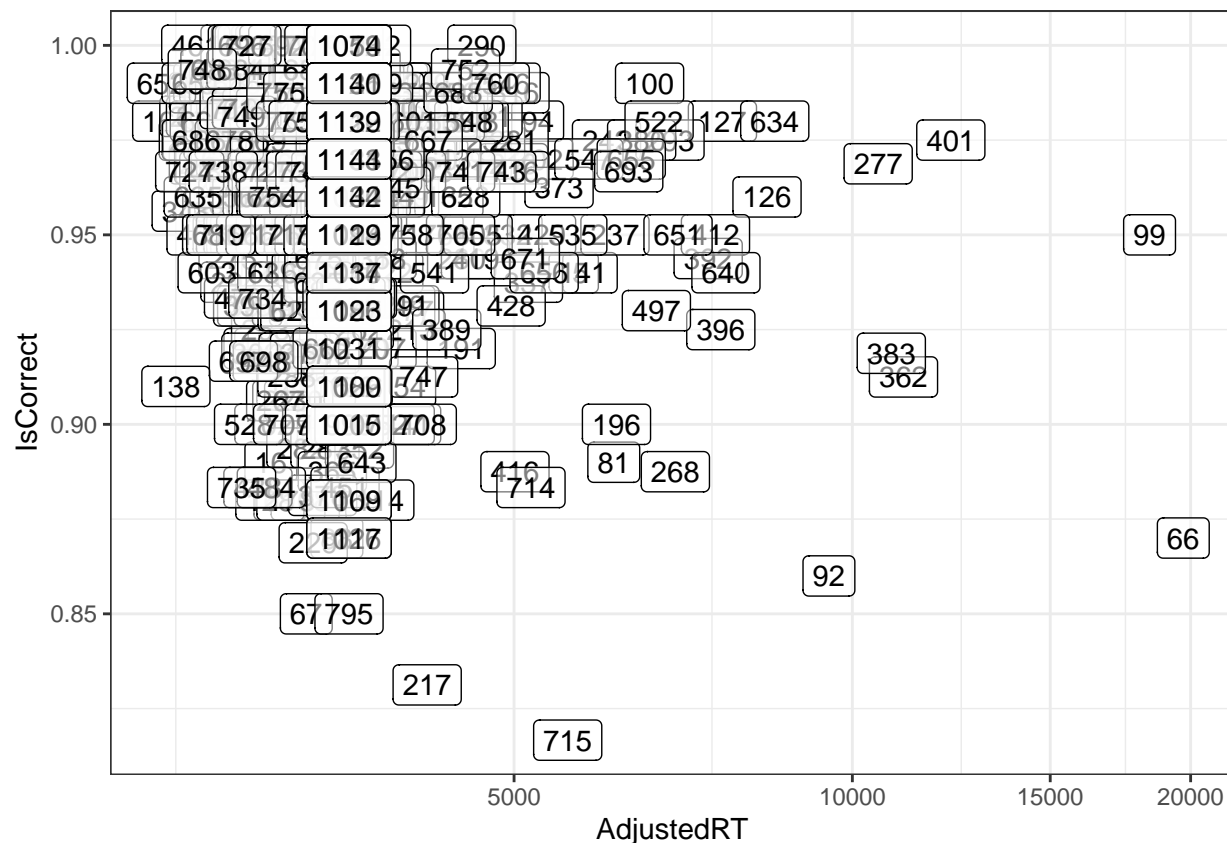
1.2.1.3 Plotting changes across trials Some of you also plotted changes across trials. Arya, for example, looked at changes in RTs across trials. Try to plot changes in the relative tilt effect across trials.

Here's a plot for Subject 1 in Experiment 1. **How would you combine the data from subjects 1-3 from Experiment 1 and then plot an average for those subjects?** Hint: you don't need to do any manual averaging. Look into `geom_smooth`, which let's you plot trend lines.



1.2.2 Jaeger group

1.2.2.1 Plotting all subjects' performance during exposure We are getting the average reaction time (AdjustedRT) and accuracy (IsCorrect) of each subject and visualize the distribution of subjects with regard to these two variables. Note that I'm using a log-transformed coordinate system on the x-axis since some RTs can be very high. **What would you conclude from this plot? Should you exclude subjects if they are very slow or fast? Should you look into whether *some* of their trials are very very slow or fast? What would be a good exclusion criterion (if any) based on RTs? Do you think all subjects performed with sufficiently high accuracy to be included in the analysis?**

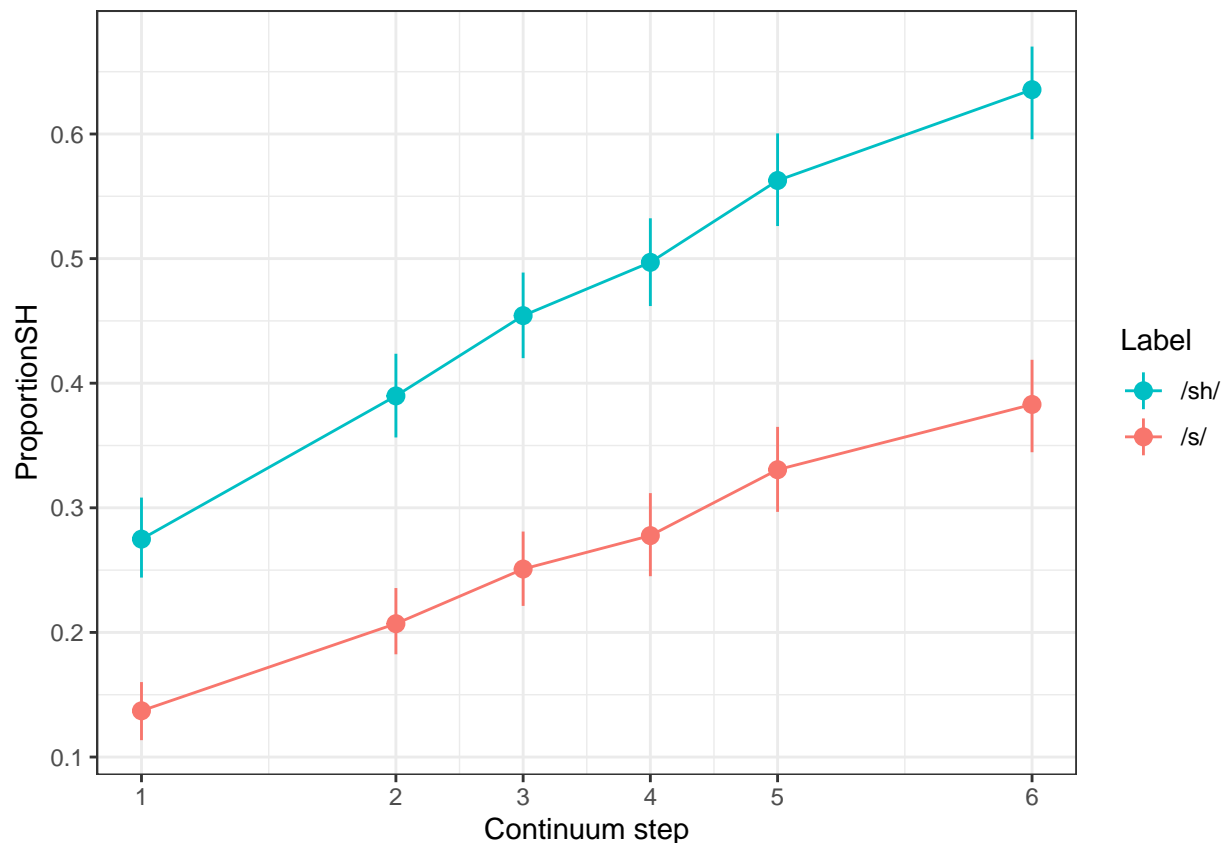


Only after you've settled on an exclusion criterion, try to modify this graph: Specifically, how would you modify this plot to a) only show Experiment 1 and b) color subjects based on the label condition and the pen condition? Hint: labels have both fill (aesthetic fill) and border color (aesthetic color). You can use these two visual means to express the label and pen condition.

How might you additionally indicate in this plot which subjects your exclusion criteria leave for analysis? Hint: you could use a `geom_rect` to draw a semi-transparent rectangle or you could use `geom_segment`, `geom_hline`, and/or `geom_vline` to draw borders in the RT-by-accuracy space that to indicate your exclusion criteria. Look up those geoms.

1.2.2.2 Plotting all subjects' performance during test Here's the basic plot for the test data. Note that we first summarized the data down to the subject level—i.e., one data point per subject, label condition, and subject. We then get the 95% bootstrapped CIs over those *by-subject means*. This is avoiding overly confident (small) CIs that would result from us failing to acknowledge that repeated measures taken from the same subject are *not* independent of each other. This is different from the motion group, since they are plotting their data separately by subject.

``summarise()`` has grouped output by 'Subject', 'Label'. You can override using the ``.groups`` argument



How would you change the y-axis label? How would you make sure that the y-axis actually goes from 0 to 0? Hint: both things can be changed through a scale component. And, how would you add information about the pen-in-the-hand vs. pen-in-the-mouth condition to this plot? Hint: shape and linetype provide you with additional visual means, as does faceting.

1.2.2.3 Relating subjects' performance during exposure and test If you figured out the above, go ahead and try to plot the boundary shift during test against the proportion of shifted words that were rated as words. This will require *combining* the exposure and test data. You will first have to aggregate (summarise) each of the two data down to the by-subject level, and then you can *join* the two data frames (look up `?left_join`).

2 Session info

```
## - Session info -----
## hash: merperson, raising hands: medium skin tone, right arrow curving up
##
## setting value
## version R version 4.1.0 (2021-05-18)
## os      macOS High Sierra 10.13.6
## system  x86_64, darwin17.0
## ui      X11
## language (EN)
## collate en_US.UTF-8
## ctype   en_US.UTF-8
```

```
## tz      America/New_York
## date    2021-11-06
## pandoc  2.14.2 @ /Applications/RStudio.app/Contents/MacOS/pandoc/ (via rmarkdown)
##
## - Packages -----
## package      * version date (UTC) lib source
## assertthat   0.2.1   2019-03-21 [1] CRAN (R 4.1.0)
## backports    1.3.0   2021-10-27 [1] CRAN (R 4.1.0)
## base64enc    0.1-3   2015-07-28 [1] CRAN (R 4.1.0)
## bit          4.0.4   2020-08-04 [1] CRAN (R 4.1.0)
## bit64        4.0.5   2020-08-30 [1] CRAN (R 4.1.0)
## broom        0.7.10  2021-10-31 [1] CRAN (R 4.1.0)
## cachem       1.0.6   2021-08-19 [1] CRAN (R 4.1.0)
## callr        3.7.0   2021-04-20 [1] CRAN (R 4.1.0)
## cellranger    1.1.0   2016-07-27 [1] CRAN (R 4.1.0)
## checkmate    2.0.0   2020-02-06 [1] CRAN (R 4.1.0)
## cli          3.1.0   2021-10-27 [1] CRAN (R 4.1.0)
## cluster      2.1.2   2021-04-17 [1] CRAN (R 4.1.0)
## colorspace   2.0-2   2021-06-24 [1] CRAN (R 4.1.0)
## cowplot      * 1.1.1   2020-12-30 [1] CRAN (R 4.1.0)
## crayon       1.4.2   2021-10-29 [1] CRAN (R 4.1.0)
## data.table   1.14.2  2021-09-27 [1] CRAN (R 4.1.0)
## DBI          1.1.1   2021-01-15 [1] CRAN (R 4.1.0)
## dbplyr       2.1.1   2021-04-06 [1] CRAN (R 4.1.0)
## desc        1.4.0   2021-09-28 [1] CRAN (R 4.1.0)
## devtools     2.4.2   2021-06-07 [1] CRAN (R 4.1.0)
## digest       0.6.28  2021-09-23 [1] CRAN (R 4.1.0)
## dplyr        * 1.0.7   2021-06-18 [1] CRAN (R 4.1.0)
## ellipsis     0.3.2   2021-04-29 [1] CRAN (R 4.1.0)
## evaluate     0.14    2019-05-28 [1] CRAN (R 4.1.0)
## fansi        0.5.0   2021-05-25 [1] CRAN (R 4.1.0)
## farver       2.1.0   2021-02-28 [1] CRAN (R 4.1.0)
## fastmap      1.1.0   2021-01-25 [1] CRAN (R 4.1.0)
## forcats      * 0.5.1   2021-01-27 [1] CRAN (R 4.1.0)
## foreign      0.8-81  2020-12-22 [1] CRAN (R 4.1.0)
## Formula      1.2-4   2020-10-16 [1] CRAN (R 4.1.0)
## fs           1.5.0   2020-07-31 [1] CRAN (R 4.1.0)
## generics     0.1.1   2021-10-25 [1] CRAN (R 4.1.0)
## ggplot2      * 3.3.5   2021-06-25 [1] CRAN (R 4.1.0)
## glue         1.4.2   2020-08-27 [1] CRAN (R 4.1.0)
## gridExtra    2.3     2017-09-09 [1] CRAN (R 4.1.0)
## gtable       0.3.0   2019-03-25 [1] CRAN (R 4.1.0)
## haven        2.4.3   2021-08-04 [1] CRAN (R 4.1.0)
## Hmisc        4.6-0   2021-10-07 [1] CRAN (R 4.1.0)
## hms          1.1.1   2021-09-26 [1] CRAN (R 4.1.0)
## htmlTable    2.3.0   2021-10-12 [1] CRAN (R 4.1.0)
## htmltools    0.5.2   2021-08-25 [1] CRAN (R 4.1.0)
## htmlwidgets  1.5.4   2021-09-08 [1] CRAN (R 4.1.0)
## httr         1.4.2   2020-07-20 [1] CRAN (R 4.1.0)
## jpeg         0.1-9   2021-07-24 [1] CRAN (R 4.1.0)
## jsonlite     1.7.2   2020-12-09 [1] CRAN (R 4.1.0)
## knitr        1.36    2021-09-29 [1] CRAN (R 4.1.0)
## labeling     0.4.2   2020-10-20 [1] CRAN (R 4.1.0)
## lattice      0.20-45 2021-09-22 [1] CRAN (R 4.1.0)
```

##	latticeExtra	0.6-29	2019-12-19	[1]	CRAN	(R 4.1.0)
##	lazyeval	0.2.2	2019-03-15	[1]	CRAN	(R 4.1.0)
##	lifecycle	1.0.1	2021-09-24	[1]	CRAN	(R 4.1.0)
##	lubridate	1.8.0	2021-10-07	[1]	CRAN	(R 4.1.0)
##	magrittr	* 2.0.1	2020-11-17	[1]	CRAN	(R 4.1.0)
##	Matrix	1.3-4	2021-06-01	[1]	CRAN	(R 4.1.0)
##	memoise	2.0.0	2021-01-26	[1]	CRAN	(R 4.1.0)
##	modelr	0.1.8	2020-05-19	[1]	CRAN	(R 4.1.0)
##	munsell	0.5.0	2018-06-12	[1]	CRAN	(R 4.1.0)
##	nnet	7.3-16	2021-05-03	[1]	CRAN	(R 4.1.0)
##	pillar	1.6.4	2021-10-18	[1]	CRAN	(R 4.1.0)
##	pkgbuild	1.2.0	2020-12-15	[1]	CRAN	(R 4.1.0)
##	pkgconfig	2.0.3	2019-09-22	[1]	CRAN	(R 4.1.0)
##	pkgload	1.2.3	2021-10-13	[1]	CRAN	(R 4.1.0)
##	plotly	* 4.10.0	2021-10-09	[1]	CRAN	(R 4.1.0)
##	png	0.1-7	2013-12-03	[1]	CRAN	(R 4.1.0)
##	prettyunits	1.1.1	2020-01-24	[1]	CRAN	(R 4.1.0)
##	processx	3.5.2	2021-04-30	[1]	CRAN	(R 4.1.0)
##	ps	1.6.0	2021-02-28	[1]	CRAN	(R 4.1.0)
##	purrr	* 0.3.4	2020-04-17	[1]	CRAN	(R 4.1.0)
##	R.matlab	* 3.6.2	2018-09-27	[1]	CRAN	(R 4.1.0)
##	R.methodsS3	1.8.1	2020-08-26	[1]	CRAN	(R 4.1.0)
##	R.oo	1.24.0	2020-08-26	[1]	CRAN	(R 4.1.0)
##	R.utils	2.11.0	2021-09-26	[1]	CRAN	(R 4.1.0)
##	R6	2.5.1	2021-08-19	[1]	CRAN	(R 4.1.0)
##	RColorBrewer	1.1-2	2014-12-07	[1]	CRAN	(R 4.1.0)
##	Rcpp	1.0.7	2021-07-07	[1]	CRAN	(R 4.1.0)
##	readr	* 2.0.2	2021-09-27	[1]	CRAN	(R 4.1.0)
##	readxl	* 1.3.1	2019-03-13	[1]	CRAN	(R 4.1.0)
##	remotes	2.4.1	2021-09-29	[1]	CRAN	(R 4.1.0)
##	reprex	2.0.1	2021-08-05	[1]	CRAN	(R 4.1.0)
##	rlang	0.4.12	2021-10-18	[1]	CRAN	(R 4.1.0)
##	rmarkdown	2.11	2021-09-14	[1]	CRAN	(R 4.1.0)
##	rpart	4.1-15	2019-04-12	[1]	CRAN	(R 4.1.0)
##	rprojroot	2.0.2	2020-11-15	[1]	CRAN	(R 4.1.0)
##	rstudioapi	0.13	2020-11-12	[1]	CRAN	(R 4.1.0)
##	rvest	1.0.2	2021-10-16	[1]	CRAN	(R 4.1.0)
##	scales	1.1.1	2020-05-11	[1]	CRAN	(R 4.1.0)
##	sessioninfo	1.2.1	2021-11-02	[1]	CRAN	(R 4.1.0)
##	stringi	1.7.5	2021-10-04	[1]	CRAN	(R 4.1.0)
##	stringr	* 1.4.0	2019-02-10	[1]	CRAN	(R 4.1.0)
##	survival	3.2-13	2021-08-24	[1]	CRAN	(R 4.1.0)
##	testthat	3.1.0	2021-10-04	[1]	CRAN	(R 4.1.0)
##	tibble	* 3.1.5	2021-09-30	[1]	CRAN	(R 4.1.0)
##	tidyr	* 1.1.4	2021-09-27	[1]	CRAN	(R 4.1.0)
##	tidyselect	1.1.1	2021-04-30	[1]	CRAN	(R 4.1.0)
##	tidyverse	* 1.3.1	2021-04-15	[1]	CRAN	(R 4.1.0)
##	tzdb	0.2.0	2021-10-27	[1]	CRAN	(R 4.1.0)
##	usethis	2.1.3	2021-10-27	[1]	CRAN	(R 4.1.0)
##	utf8	1.2.2	2021-07-24	[1]	CRAN	(R 4.1.0)
##	vctrs	0.3.8	2021-04-29	[1]	CRAN	(R 4.1.0)
##	viridisLite	0.4.0	2021-04-13	[1]	CRAN	(R 4.1.0)
##	vroom	1.5.5	2021-09-14	[1]	CRAN	(R 4.1.0)
##	withr	2.4.2	2021-04-18	[1]	CRAN	(R 4.1.0)

```
## xfun          0.27    2021-10-18 [1] CRAN (R 4.1.0)
## xml2          1.3.2    2020-04-23 [1] CRAN (R 4.1.0)
## yaml          2.2.1    2020-02-01 [1] CRAN (R 4.1.0)
##
## [1] /Library/Frameworks/R.framework/Versions/4.1/Resources/library
##
## -----
```