Basic R functions

Jonas Björnerstedt

2021-10-19

Data management

The `dplyr` package is one of the packages that is loaded with `library(tidyverse)`

Documentation is available at: https://dplyr.tidyverse.org/. (Check the *Get started* and *Reference* menus.)

To modify a dataset `df`

Function	Description
`filter(df, conds)`	Select observations (rows) based on conditions `conds`
`select(df, vars)`	List variables (ie columns) `vars` to keep. To drop `var` write `select(df, -var)`
`summarise()`	Provide summary statistics for one or more variables
`group_by()`	List variable(s) that `summarise()` should calculate stats by
`mutate()`	Create new or modify existing variables
`rename()`	Change name of variable

Data analysis

Plotting The `ggplot2` package is one of the packages that is loaded with `library(tidyverse)`. It includes the `ggplot()` function.

Documentation is available at: https://ggplot2.tidyverse.org/. Use the **'ggplot2' builder** in the Addins menu. A plot contains at least three elements

ggplot function	Description
`ggplot(dataset)`	Create a plot object from dataset
`aes()`	Add mappings of variables, *aesthetics*
`geom_something()`	Specify a geometry, like `geom_point()`, `geom_line()`, `geom_smooth()` etc.

- To add a regression line to plot, add: `+ geom_smooth(method = "lm")`

Create summary tables as dataframes

Package	Function	Description
`library(statar)`	`sum_up()`	For summary statistics
`library(corrr)`	`correlate()`	For correlation table
`library(knitr)`	`kable()`	To present table in rmarkdown

Create and present regressions

Package	Function	Description
`library(estimatr)`	`lm_robust()`	*Modern* lm() function, robust SE
`library(huxtable)`	`huxreg()`	For regression tables