STAT 33B Lecture – April 29

- Review:
  - Creating Closures
  - Creating S3 Objects
- Formulas
  - Built-in plot functions
- dplyr & magrittr
- Non-standard Evaluation
- What Now?

To the notebook!

Formulas

Formulas

Many of R's plotting and modeling functions accept a **formula**:

`x ~ y` means "the relationship between x and y"

A formula can also indicate conditioning:

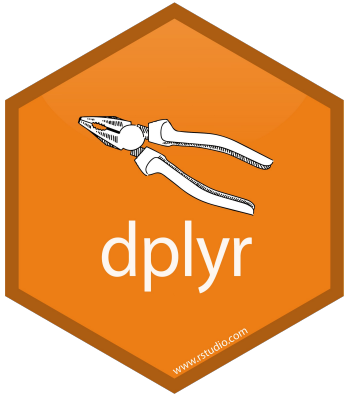
`x ~ y | z` means "the relationship between x and y, given z"

dplyr & magrittr

dplyr

Alternative interface for subsetting data frames.

- Part of the Tidyverse
- There's online documentation
- There's a cheat sheet
- Popular



How dplyr Compares

Task	Base R	dplyr
Select rows based on a condition	`x[condition, ]`	`filter(x, condition)`
Select rows by name or position	`x[index, ]`	`slice(x, index)`
Select columns by name or position	`x[, index]`	`select(x, index)`
Sort rows by a column	`x[order(x$col), ]`	`arrange(x, col)`
Insert a new column	`x$col = f(a, b)`	`x = mutate(x, col = f(a, b))`
Group elements or rows by a factor	`split(x$col, groups)`	`group_by(x, groups)`
Compute statistics on groups	`sapply(by_group, mean)`	`summarise(by_group, mean(col))`

magrittr

Alternative interface for calling functions — the **pipe operator**:

`x %>% f` instead of `f(x)`

- Part of the Tidyverse
- There's online documentation
- Popular, but controversial
- Loaded automatically by dplyr

Non-standard Evaluation

Non-standard Evaluation

Some functions evaluate arguments in a non-standard way:

- `library()`
- `subset()`
- dplyr functions
- `plot()`
- `curve()`

These sometimes don’t require quotes around strings

All of these circumvent R’s standard evaluation rules.

What Now?

What Now?

- **DATA 8**: For everyone. Basic statistics skills. A chance to learn Python.
- **DATA 100**: If you want a data science career. Round out your skill set: databases, visualizations, modern statistical methods.
- **STAT 133**: If you want to know more about (how statisticians use) R.