The variety of R syntaxes give you many ways to “say” the same thing.

![Diagram showing various R syntaxes for different tasks such as summarizing, plotting, and wrangling data.]](image-url)
R Syntax Comparison :: CHEAT SHEET

Syntax is the set of rules that govern what code works and doesn’t work in a programming language. Most programming languages offer one standardized syntax, but R allows package developers to specify their own syntax. As a result, there is a large variety of (equally valid) R syntaxes.

The three most prevalent R syntaxes are:

1. **The dollar sign syntax**, sometimes called **base R syntax**, expected by most base R functions. It is characterized by the use of `dataset$variablename`, and is also associated with square bracket subsetting, as in `dataset[1,2]`. Almost all R functions will accept things passed to them in dollar sign syntax.

2. **The formula syntax**, used by modeling functions like `lm()`, lattice graphics, and mosaic summary statistics. It uses the tilde (~~) to connect a response variable and one (or many) predictors. Many base R functions will accept formula syntax.

3. **The tidyverse syntax** used by `dplyr`, `tidyr`, and more. These functions expect data to be the first argument, which allows them to work with the “pipe” (~~~) from the `magrittr` package. Typically, `ggplot2` is thought of as part of the tidyverse, although it has its own flavor of the syntax using plus signs (~~~) to string pieces together. `ggplot2` author Hadley Wickham has said the package would have had different syntax if he had written it after learning about the pipe.

Educators often try to teach within one unified syntax, but most R programmers use some combination of all the syntaxes.

### Even more ways to say the same thing

Even within one syntax, there are often variations that are equally valid. As a case study, let’s look at the `ggplot2` syntax. `ggplot2` is the plotting package that lives within the tidyverse. If you read down this column, all the code here produces the same graphic.

**quickplot**

`qplot()` stands for quickplot, and allows you to make quick plots. It doesn’t have the full power of `ggplot2`, and it uses a slightly different syntax than the rest of the package.

```r
qplot(x=disp, y=mpg, data=mtcars, geom="point")
```

```r
ggplot2::qplot(x=disp, y=mpg, data=mtcars)
```

```r
ggplot2::qplot(disp, mpg, data=mtcars)
```

**ggplot**

To unlock the power of `ggplot2`, you need to use the `ggplot()` function (which sets up a plotting region) and add geoms to the plot.

```r
ggplot2::ggplot(mtcars)+ geom_point(aes(x=disp, y=mpg))
```

```r
ggplot2::ggplot(mtcars, aes(x=disp, y=mpg))+ geom_point()
```

```r
ggplot2::ggplot(mtcars, aes(x=disp))+ geom_point(aes(y=mpg))
```

```r
ggplot2::ggplot()+ geom_point(mtcars, aes(x=disp, y=mpg))
```

**ggformula**

The “third and a half way” to use the formula syntax, but get `ggplot2`-style graphics

```r
ggformula::gf_point(mpg~cyl, data= mtcars)
```

Internet research tip:
If you are searching on google, StackOverflow, or another favorite online source and see code in a syntax you don’t recognize:
- Check to see if the code is using one of the three common syntaxes listed on this cheatsheet
- Try your search again, using a keyword from the syntax name (“tidyverse”) or a relevant package (“mosaic”)

Sometimes particular syntaxes work, but are considered dangerous to use, because they are so easy to get wrong. For example, passing variable names without assigning them to a named argument.