Best practices for parsing a formula in R programming

Ronald (Ryy) Glenn Thomas

2024-03-14

Table of contents

1 Introduction

R developers can make use of the `formula` object class to allow succinct (unquoted) input to functions. However, its a bit unclear what the best strategy is to parse the formula inside the function to extract and validate the instructions the formula provides.

In this post we'll present several alternative methods for consideration. As with most coding challenges there is no single best answer but some are more elegant than others.

2 Methods

2.1 References:

"the formula is used to specify the symbolic model as well as generating the intended design matrix" The R Formula Method: The Good Parts · R Views

" You're probably familiar with formulas from linear models (e.g. lm(mpg ~ displ, data = mtcars)) but formulas are more than just a tool for modelling: they are a general way of capturing an unevaluated expression"

"because a formula captures two things:

```
An unevaluated expression.
The context (environment) in which the expression was created.
```

~ is a single character that allows you to say: "I want to capture the meaning of this code, without evaluating it right away". For that reason, the formula can be thought of as a "quoting" operator."

Non-standard evaluation

Advanced Programming and Non-Standard Evaluation with dplyr | by Ryan Boyer | Shipt Tech