Intro. to OOP and S3 System in R

Ravichandran S.

ABCS, BIDS,
FNLCR

https://github.com/ravichas/OOP-S3-in-R

Scope

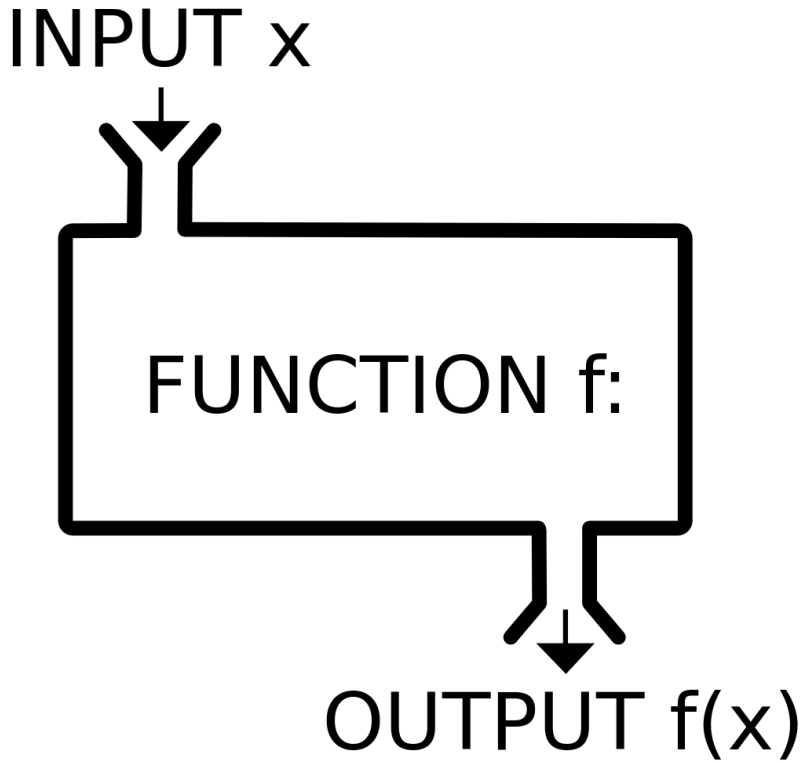
- OOP
  - Concepts might be similar to other languages, but …
- Specific to R
- Examples

Specific goals

- Note I am !here to teach OOP
- Reinforce concepts that you already know and associate them with OOP. In that process, I will remind/provide some definitions/examples of OOP
- Specific to R ; One-liners easy for other programmers

Functional programming

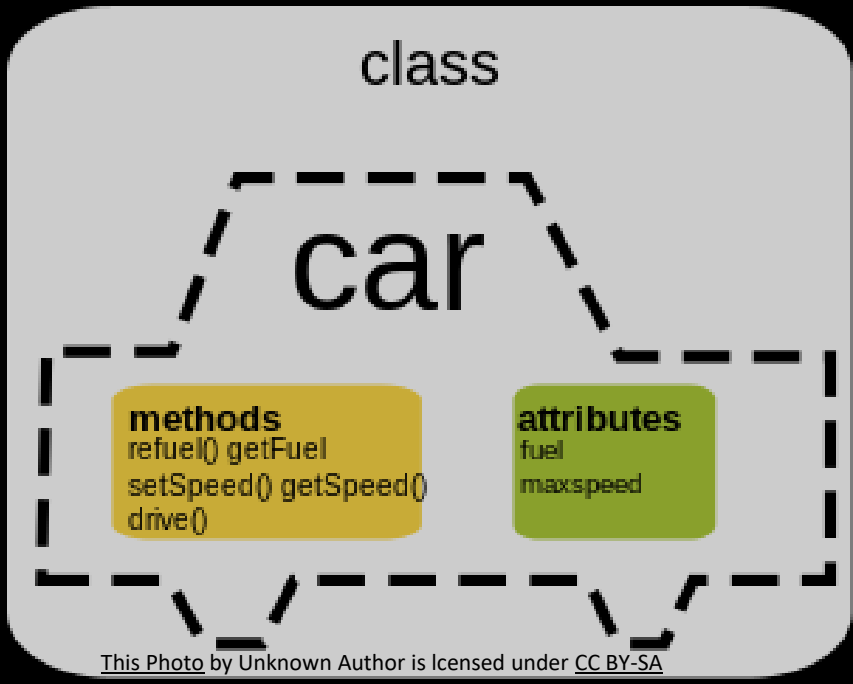
- Commonly used
- Focus is on functions
- Chain functions together to accomplish things
- Good for?
  - Data analysis, modeling etc.




```
my_add <- function(x,y)
{
   # do some task
    return(x + y)
}
```

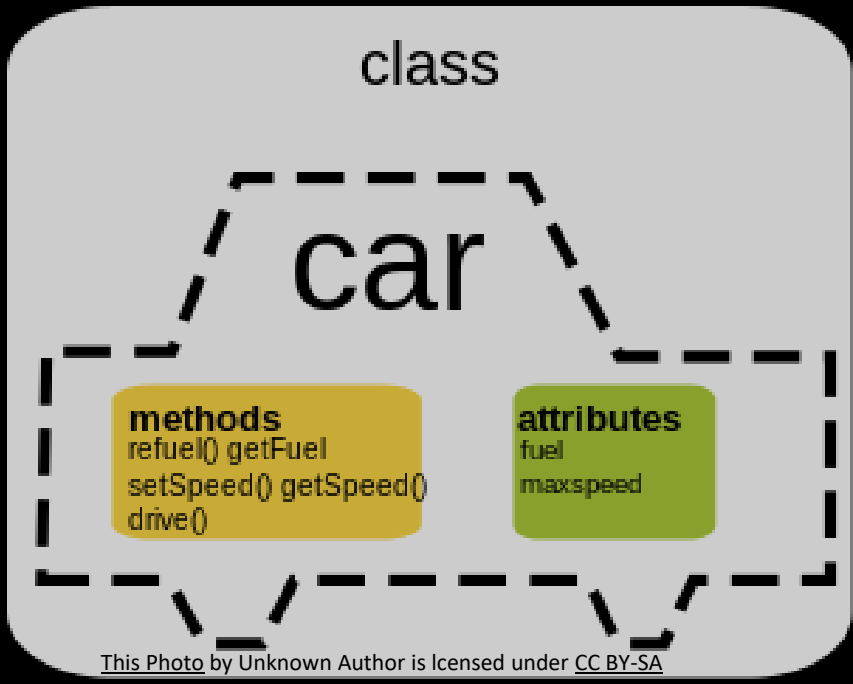
OOP

- Focus on objects
- Steps
  - Define Object
  - Describe its attributes (size, seats etc.)
  - Define methods to describe what object can do
- Note in OOP, functions are called METHODS



OOP

- What is OOP good for?
  - Developing tools, GUIs
  - Complex limited # of objects
  - Specifically when you know you can define the objects clearly
  - Developing GUIs (limited # of options)
    - Interface that can handle limited # of inputs
    - Bioconductor objects (complex but can be reused)

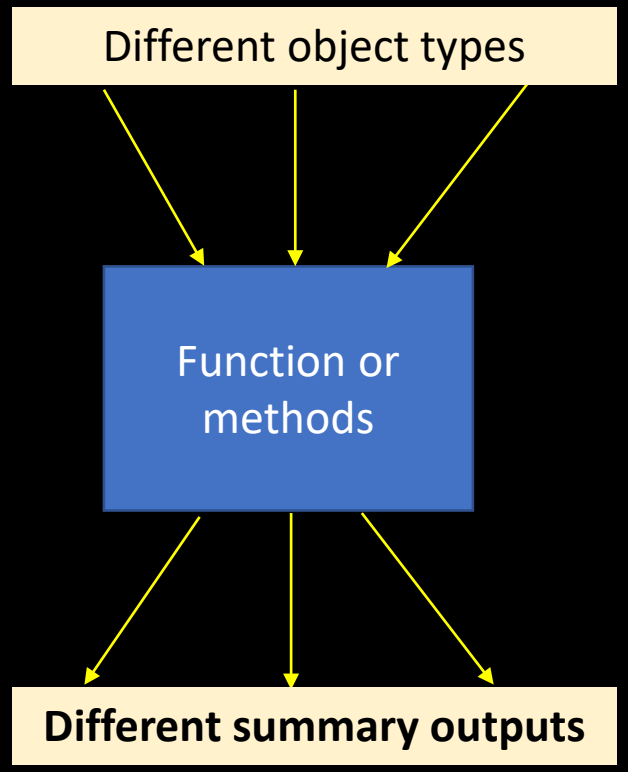


Object types in R

- ~ 20 types
- Integer, logical, numeric, data.frame, List, matrix, array, factor, formula, etc
- Most important type (create complex objects are:
  - List
    - Contain other types
- These constitute the building blocks that are needed for analysis

OOP concept: Functions behave differently for different object types

Polymorphism;
Function Overloading



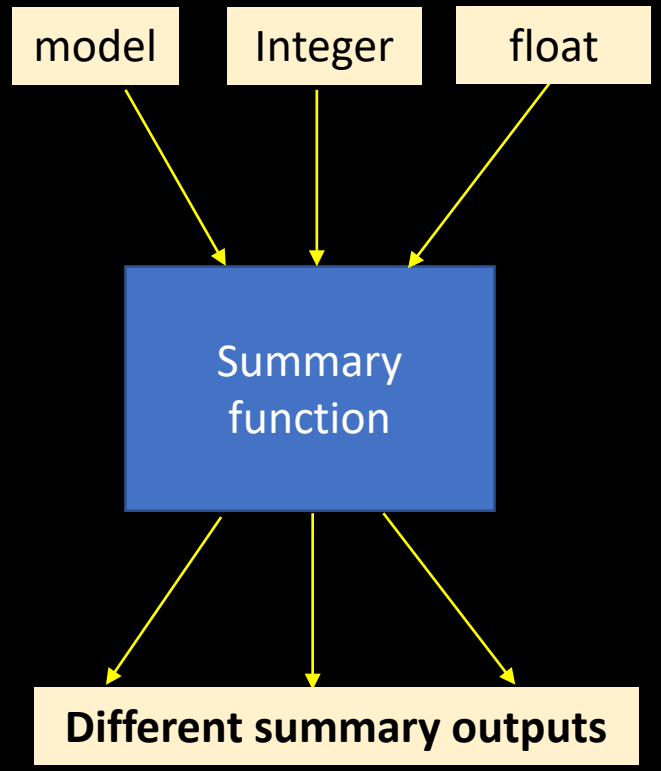
- Let us look at an example

- For OOP to work, R has to identify the class of the variables
  - How does R accomplishes this?

Hands-on 1

Object Types
Class of objects
A simple function to show that it behaves differently for different input types

OOP concept: Functions behave differently for different object types



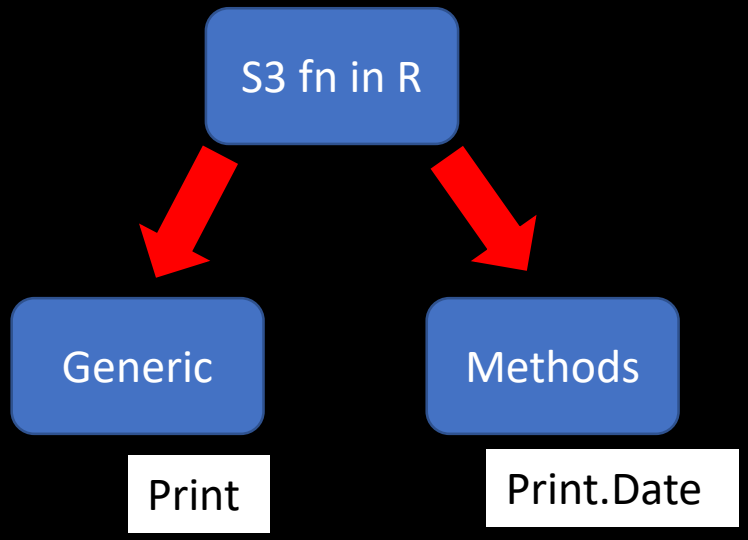
- Class (command: *class*)
  - Doesn't tell the whole story

- Typeof (command: *typeof* ; c-code)
  - Supplements class command

S3 System

- R accomplishes the OOP tasks using many Paradigms
- Why learn S3?
  - Most commonly used type
  - Simple; lacks formal definition
    - Freedom to be creative (comes with cost!)
  - Create custom class of objects and use S3 to accomplish complex tasks

S3 Object System in R



- Central players
  - Class & Method
- CLASS
  - defines type of object, its properties, how it works with other objects
- METHOD
  - Function associated with a particular object type
- OOP style in R is different than C++ or Java etc
- A generic function will decide what appropriate method to call

Generics and Methods

```
> head(mtcars)
                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

summary(mtcars)

Summary generic

```
> summary
function (object, ...)
UseMethod("summary")
```

class = data.frame

summary.data.frame

```
> summary.data.frame
function (object, maxsum = 7L, digits = max(3L, getOption("digits") -
    3L), ...)
{
    ncw <- function(x) {
        z <- nchar(x, type = "w")
        if (any(na <- is.na(z))) {
            z[na] <- nchar(encodeString(z[na]), "b")
        }
        z
    }
```

Output

```
      mpg             cyl             disp             hp             drat      
 Min.   :10.40   4:11   Min.   : 71.1   Min.   : 52.0   Min.   :2.760  
 1st Qu.:15.43   6: 7   1st Qu.:120.8   1st Qu.: 96.5   1st Qu.:3.080  
 Median :19.20   8:14   Median :196.3   Median :123.0   Median :3.695  
 Mean   :20.09          Mean   :230.7   Mean   :146.7   Mean   :3.597  
 3rd Qu.:22.80          3rd Qu.:326.0   3rd Qu.:180.0   3rd Qu.:3.920  
 Max.   :33.90          Max.   :472.0   Max.   :335.0   Max.   :4.930  
       wt             qsec             vs             am          gear      
 Min.   :1.513   Min.   :14.50   0:18   0:19   3:15   Min.   :1.000  
 1st Qu.:2.581   1st Qu.:16.89   1:14   1:13   4:12   1st Qu.:2.000  
 Median :3.325   Median :17.71                 5: 5   Median :2.000  
 Mean   :3.217   Mean   :17.85                        Mean   :2.812  
 3rd Qu.:3.610   3rd Qu.:18.90                        3rd Qu.:4.000  
 Max.   :5.424   Max.   :22.90                        Max.   :8.000  
```

How to name a Method?

- Standard notation for S3
- print.Date
  - generic.class
- Arguments should be same for both generic and UseMethod
- To avoid from being mistaken, don't name your variable/function with "dot"
  - DON'T: my.print
  - Maybe: my_print_function

```
> print
function (x, ...)
UseMethod("print")
<bytecode: 0x000000001e689540>
<environment: namespace:base>
```

generic	*UseMethod*
print	generic.class
	print.data.frame
	print.data.table*
	print.Date
	print.default
	print.dendrogram

generic	*UseMethod*
summary	generic.class
	summary.data.frame
	summary.data.table
	summary.factor
	summary.default

generic	*UseMethod*
print	generic.class
	print.data.frame
	print.data.table
	print.Date
	print.default
	print.dendrogram

Hands-on 2

- What functions are S3?
- How can I find out whether a function is S3?
- What methods are available for a S3 function?
- What functions are available for a class?

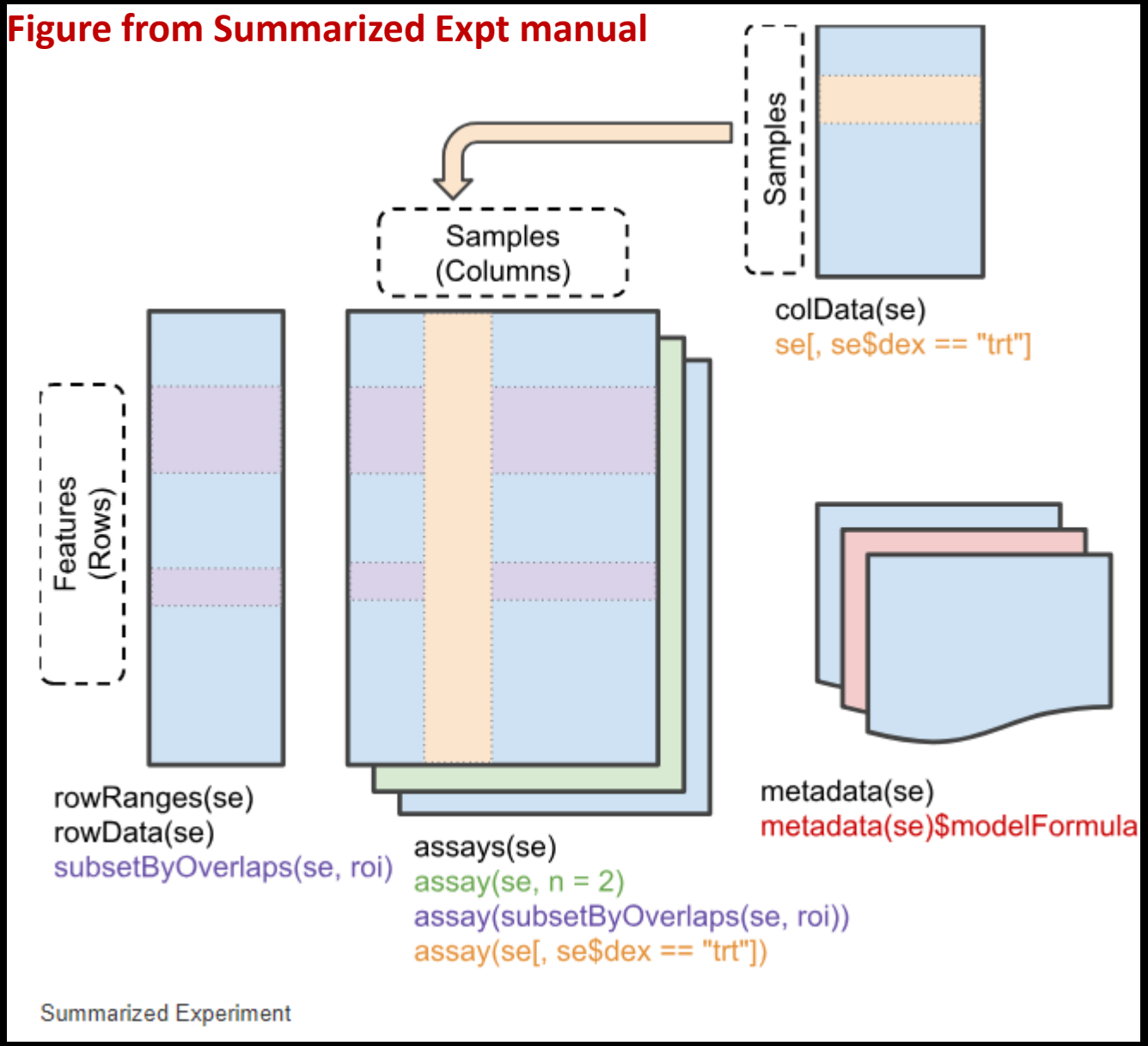
Other OOP Systems (frameworks) in R

- Important systems
- S3 (Introduced in 3rd version of S Language)
- S4 (4th version of S)
  - Bioconductor
- R6 (introduced in 6 version of S; more matured)
- ReferenceClasses (RC)

You can think of the systems as different packages for implementing OOP

S4

- Very useful to create new class
  - SummarizedExperiment
- Complex objects
  - Genomic objects
  - Elements of class are called slots
  - SetMethod to define methods for a class
- Reused in many contexts



```
> se <- airway
> se
class: RangedSummarizedExperiment
dim: 64102 8
metadata(1): ''
assays(1): counts
rownames(64102): ENSG00000000003 ENSG00000000005 ... LRG_98 LRG_99
rowData names(0):
colnames(8): SRR1039508 SRR1039509 ... SRR1039520 SRR1039521
colData names(9): SampleName cell ... Sample BioSample
> 
```

metadata accessor

```
> metadata(se)
[[1]]
Experiment data
  Experimenter name: Himes BE
  Laboratory: NA
  Contact information:
  Title: RNA-Seq transcriptome profiling identifies CRISPLD2 as a glucocorticoid responsive gene that modulates cytokine function in airway smooth muscle cells.
  URL: http://www.ncbi.nlm.nih.gov/pubmed/24926665
  PMIDs: 24926665

  Abstract: A 226 word abstract is available. Use 'abstract' method.
```

Acknowledgements

- Programmers' Corner team
- Presentations/Lectures/Tutorials from the following people good starting point for beginners
- Hadley Wickham
  - Adv. R tutorial
- Richie Cotton
  - Youtube
- Kelly Black, Univ Georgia
  - cyclismo.org

THANK YOU

https://github.com/ravichas/OOP-S3-in-R