



# Introduction to R

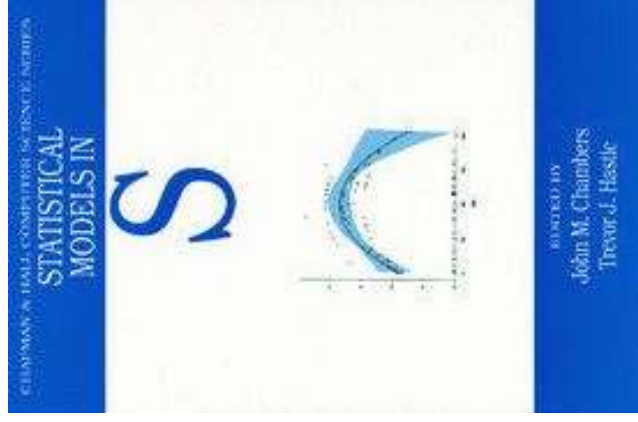


Diego Useche - [diego.useche@unisabana.edu.co](mailto:diego.useche@unisabana.edu.co)

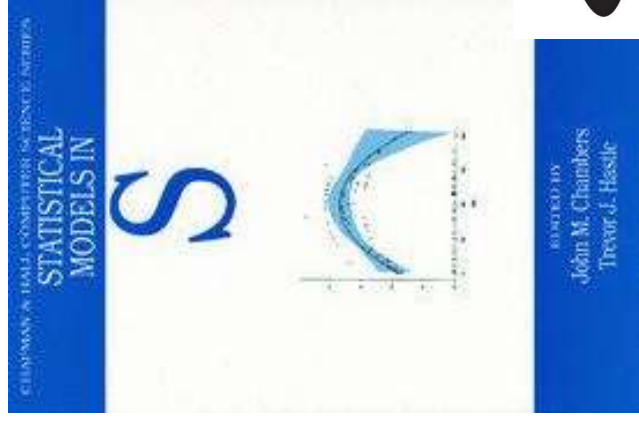


Probabilidad y estadística I  
Math, Physics and Statistics Department  
Universidad de la Sabana

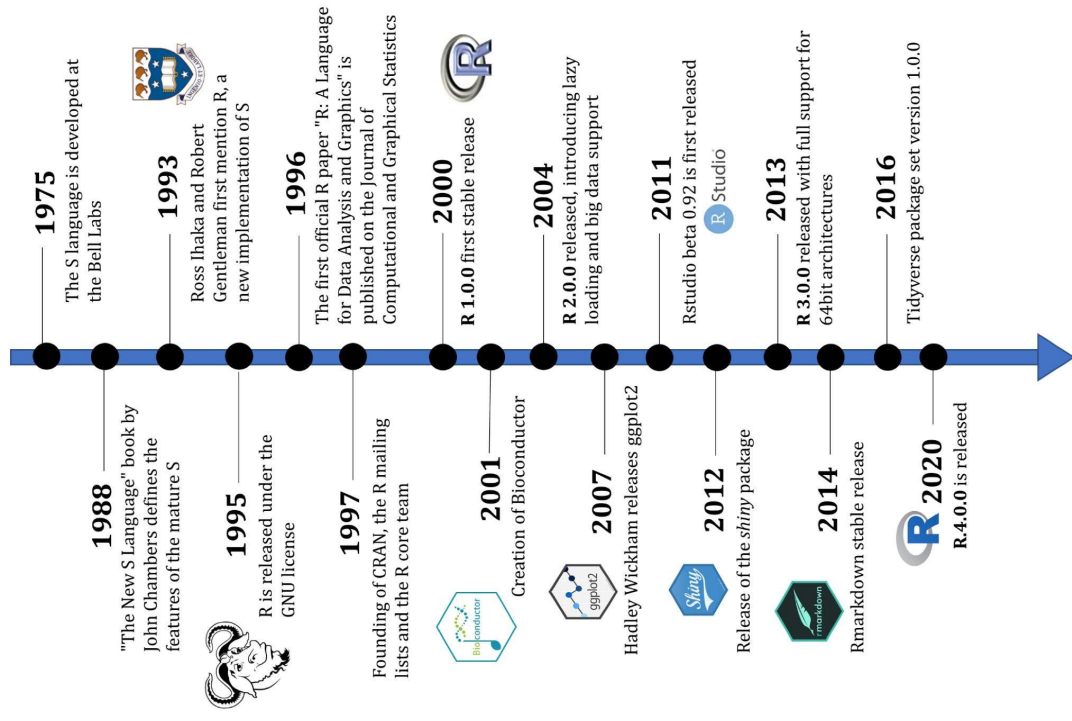
# Origins of R: Based on S language



# Origins of R: Based on S language



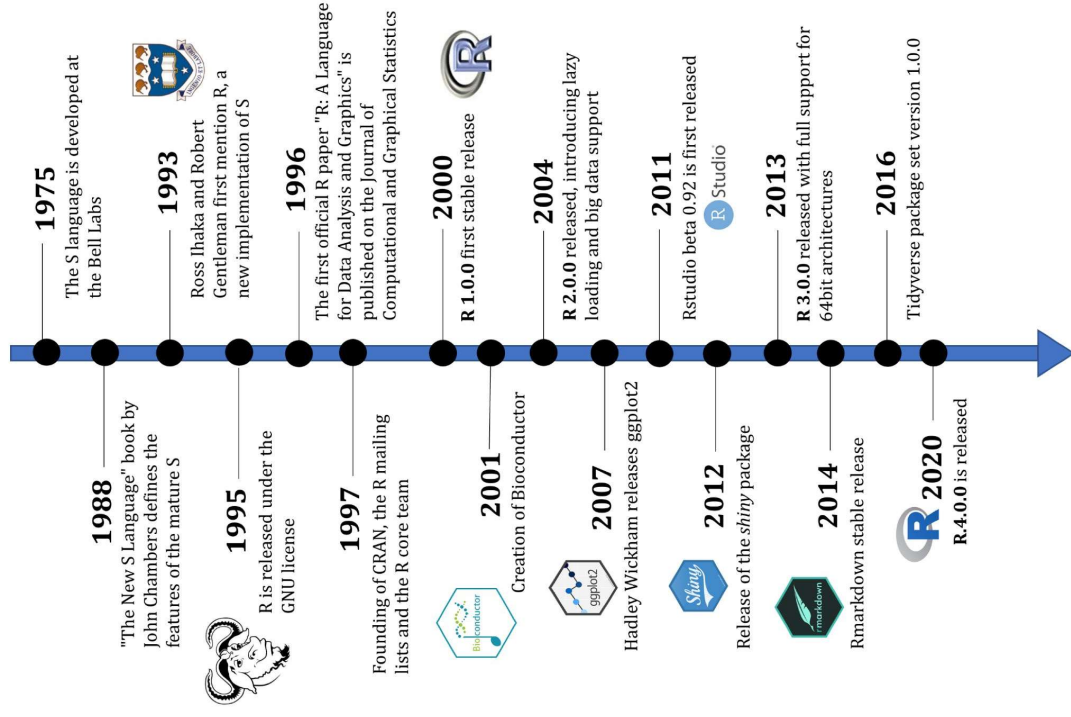
# Origins of R: Based on S language



# Origins of R: Based on S language



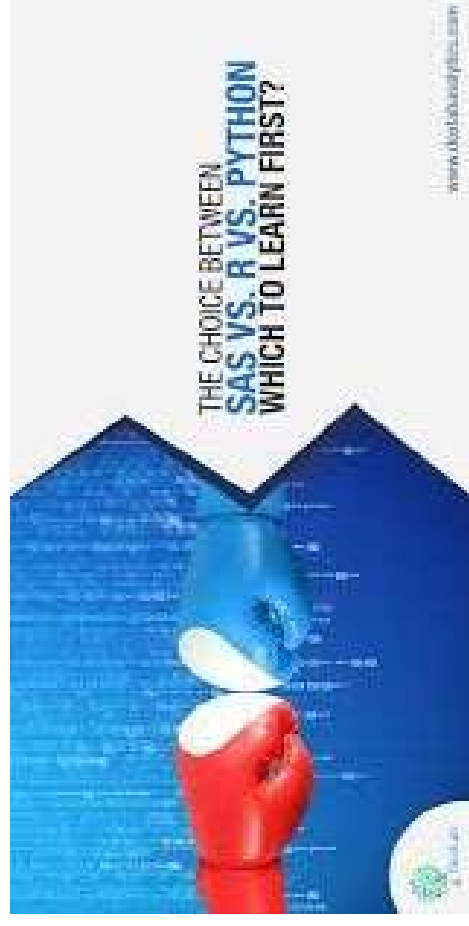
Ross Ihaka and Robert Gentleman



# R vs Python vs SAS



# R vs Python vs SAS



# R vs Python vs SAS

Parameter	SAS	R	Python
Availability/Cost	3	5	5
Ease of Learning	4.5	2.5	3.5
Data Handling Capabilities	4	4	4
Graphical Capabilities	3	4.5	4.5
Advancement in Tools	4	4.5	4.5
Job Scenario	4	4.5	4.5
Customer Service Support and Community	4	3.5	3.5
Deep Learning Support	2	3	4.5
Total	28.5	31.5	34



# R and R studio



Icon for R



Icon for RStudio

# R and R studio



Icon for R



Icon for RStudio

- R is like a car's engine
- RStudio is like a car's dashboard

**R: Engine**



**RStudio: Dashboard**



# General structure of R

The screenshot displays the RStudio interface with three main panes: Script, Environment, and Output.

**Script Pane:** Contains R code for loading ggplot2, plotting mpg data, and adding a color aesthetic for car class.

```
1 library(ggplot2)
2 ggplot(mpg, aes(x = displ, y = hwy)) +
3   geom_point(aes(color = class))
4 mp
5
```

**Environment Pane:** Shows the current environment with the message "Environment is empty".

**Output Pane:** Displays the result of the ggplot command, showing a scatter plot of highway mileage (hwy) versus engine displacement (displ) colored by car class. The legend indicates the following classes: 2seater, compact, midsize, minivan, pickup, subcompact, and suv.

**Console Pane:** Shows the execution of the R code, including the loading of ggplot2, the creation of the mpg object, and the execution of the ggplot command.

```
> library(ggplot2)
> ggplot(mpg, aes(x = displ, y = hwy)) +
+   geom_point(aes(color = class))
# A tibble: 234 x 11
#   manufacturer    model displ  year  cyl  trans      drv     cty   hwy fl    class
#   <chr>          <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>
1 audi          a4      1.8 1999    4    auto(l5) f      18    29 p      compact
2 audi          a4      1.8 1999    4    manual(m5) f      21    29 p      compact
3 audi          a4      2.0 2008    4    manual(m6) f      20    31 p      compact
4 audi          a4      2.0 2008    4    auto(av) f      21    30 p      compact
5 audi          a4      2.8 1999    6    auto(l5) f      16    26 p      compact
6 audi          a4      2.8 1999    6    manual(m5) f      18    26 p      compact
7 audi          a4      3.1 2008    6    auto(av) f      18    27 p      compact
8 audi          a4 quattro 1.8 1999    4    manual(m5) f      18    26 p      compact
9 audi          a4 quattro 1.8 1999    4    auto(l5) f      16    25 p      compact
10 audi          a4 quattro 2.0 2008    4    manual(m6) f      20    28 p      compact
# ... with 224 more rows
```

# Initial program

---

- Hello world program

R	OUTPUT
1 print("Hello World")	[1] "Hello World"

# ?print() to ask for help of a function

- How to ask for help in a function

R		OUTPUT
1	?print()	<div><div>Print Values</div><div>Description</div><div>print prints its argument and returns it <i>invisibly</i> (via <a href="#">invisible(x)</a>). It is a generic function which means that new printing methods can be easily added for new <a href="#">classes</a>.</div><div>Usage</div></div>

```
print(x, ...)
```

```
## S3 method for class 'factor'
print(x, quote = FALSE, max.levels = NULL,
      width = getOption("width"), ...)
```

# Basic concepts of R

---

- Functions
- Comments
- Variables
- Data types
- Vectors
- Pipes
- Data Frames

# Comment code

- comment #

R	OUTPUT
<pre>1 # This variable creates an int 2</pre>	No output.

# Comment code

- comment #

R	OUTPUT
<pre>1 # This variable creates an int 2</pre>	No output.



# Variable in R

---

## A value in R

- It can be stored in the computer for later use

# Variable in R

## A value in R

- It can be stored in the computer for later use
- Syntax: `a <- 2.3`

R		OUTPUT
1	<code>a &lt;- 3</code>	
2	<code>a</code>	<code>[1] 3</code>

# General structure of R



`function(variable)`

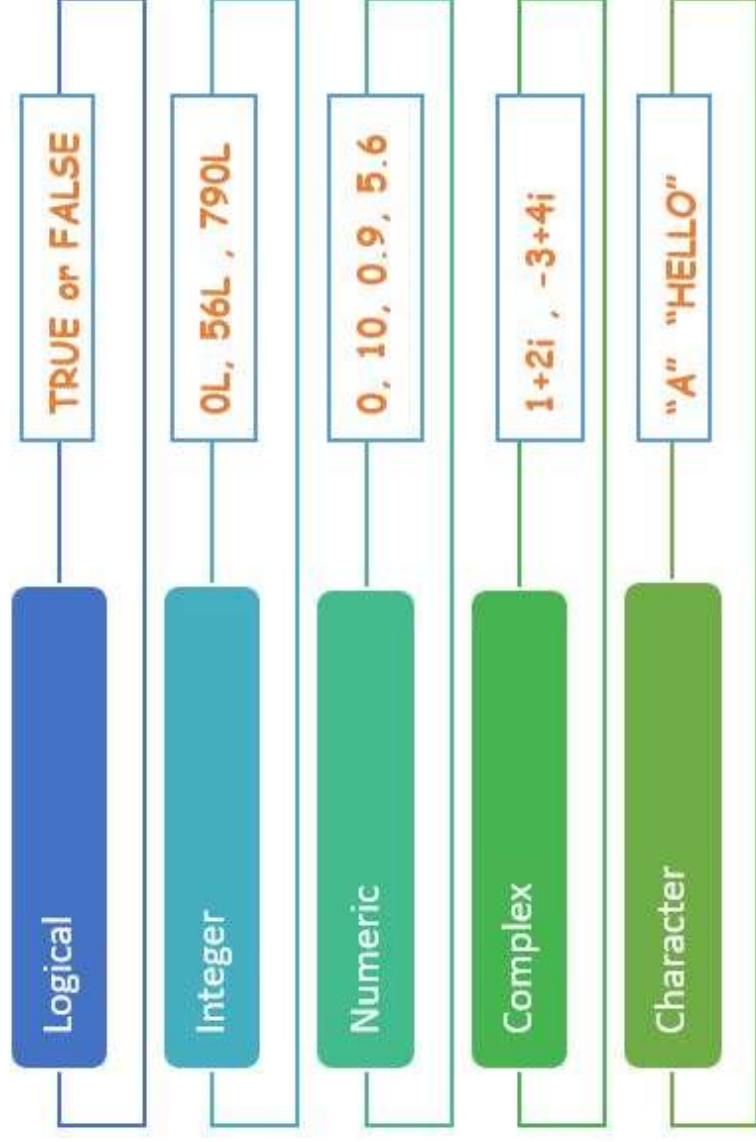
# General syntax structure of R

function(variable)

R	OUTPUT
<pre>1 print("Hello World")</pre>	<pre>[1] "Hello World"</pre>

# Data types

---



# Vector in R

---

A variable that stores a set of elements with the same data type

# Vector in R

---

Vector syntax

`c(1.2, 2.3, 5.6)`

# Vector in R

Vector syntax  
`c(1.2, 2.3, 5.6)`

R	OUTPUT
<pre>1   c(1.2, 2.3, 5.6)</pre>	<pre>[1] 1.2 2.3 5.6</pre>



## List in R

---

A variable that can store a set of elements  
with the different data type

# List in R

---

list syntax

```
list(1.2, 2L, "Happy coding")
```

# List in R

## list syntax

`list(1.2, 2L, "Happy coding")`

R		OUTPUT
1	<code>mylist &lt;- list(1.2, 2L, "Happy coding")</code>	<code>[[1]]</code>
2	<code>mylist</code>	<code>[1] 1.2</code>
3		
4	<code>mylist[[2]]</code>	<code>[[2]]</code>
5	<code>mylist[[1]]</code>	<code>[1] 2</code>
	<code>mylist[[3]]</code>	<code>[[3]]</code>
	<code>mylist[1]</code>	<code>[1] "Happy coding"</code>

# Data Frames in R

---

- To create data table
- Syntax: “data.frame()”

# Data Frames in R

---

- To create data table
- Syntax: “data.frame()”

# Data Frames in R

- To create data table
- Syntax: “data.frame()”

R

```
ages_students <- data.frame(names = c('Juan', 'Dario', 'Luciana'), ages = c(7L, 9L, 6L))  
ages_students
```

OUTPUT

```
  names ages  
1   Juan   7  
2  Dario   9  
3 Luciana   6
```

## Extract operator

---

- To extract a data sample indicate row and column
- Syntax: `data.frame()[row number, column number]`

## Extract operator

---

- To extract a data sample indicate row and column
- Syntax: `data.frame()[2, 1]`



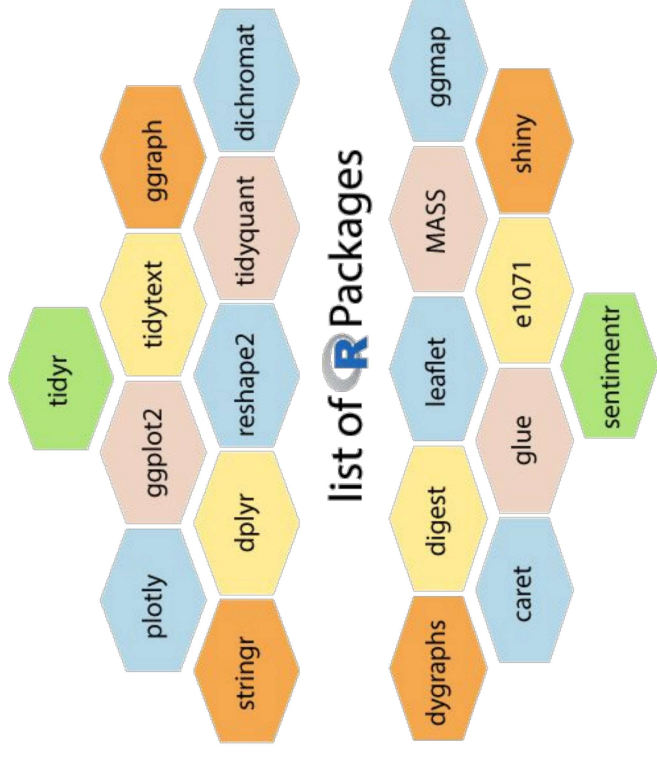
# Extract operator

- To extract a data sample indicate row and column
- Syntax: “data.frame()[2, 1]”

R	OUTPUT
<pre>ages_students &lt;- data.frame(names = c('Juan', 'Dario', 'Luciana'), ages = c(7L, 9L, 6L)) ages_students ages_students[3, 2] ages_students[2, 1]</pre>	<pre>names ages 1      Juan      7 2     Dario      9 3   Luciana      6 [1] 6 [1] "Dario"</pre>

# Packages in R language

- Prebuilt set of functions you can use.



## Install R Libraries in your computer

---

- Prebuilt functions you can use.
- Install only once.
- Syntax: “install.packages()”

# Install R Libraries in your computer

- Prebuilt functions you can use.
- Install only once.
- Syntax: “install.packages()”

R	OUTPUT
1 install.packages("readr")	

# Import R Libraries

---

- Prebuilt functions you can use.
- You need to import the library each time.
- Syntax: “library()”

# Import R Libraries

- Prebuilt functions you can use.
- You need to import the library each time.
- Syntax: “library()”

R	OUTPUT
<pre>1  import(readr)</pre>	

# References

---

[Hastie, T. J. \(Ed.\). \(2017\). Statistical models in S. Routledge. <https://www.coursera.org/learn/r-programming>](https://www.coursera.org/learn/r-programming)

<https://www.topuniversities.com/universities/university-auckland>,  
<https://www.youtube.com/watch?v=jQ-UL0IJTsw>

<https://m.dexlabanalytics.com/blog/the-choice-between-r-vs-python-which-to-learn-first>

<https://www.icertglobal.com/sas-vs-r-vs-python-which-is-best-for-data-analysis-in-2024-blog/detail>

<https://www.analyticsvidhya.com/blog/2017/09/sas-vs-vs-python-tool-learn/>

[https://jminnier-berd-r-courses.netlify.app/01-getting-started-v2/01\\_getting\\_started\\_slides.html#7](https://jminnier-berd-r-courses.netlify.app/01-getting-started-v2/01_getting_started_slides.html#7)

<https://datasciencedojo.com/blog/r-programming/>

<https://learnertutorials.com/r-programming/data-types>

<https://discuss.boardinfinity.com/t/list-of-some-r-packages/5444>