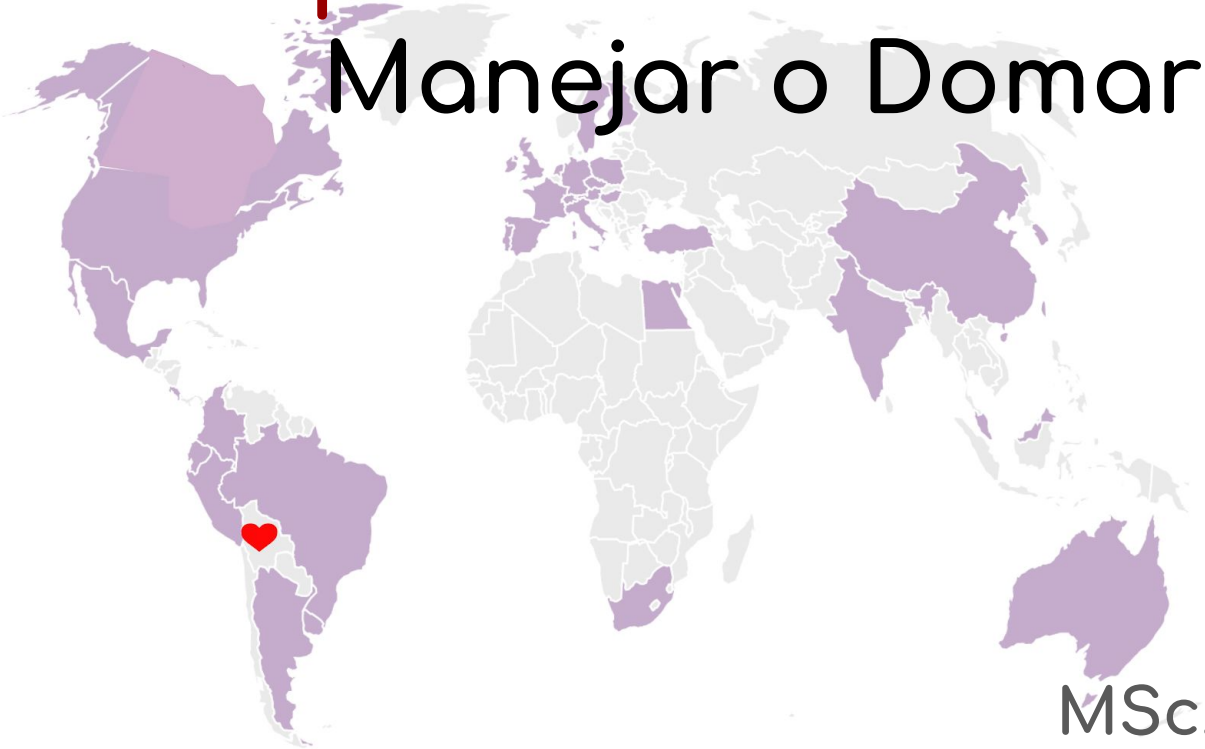
“R para Ciencia de Datos”

Manejar o Domar Datos

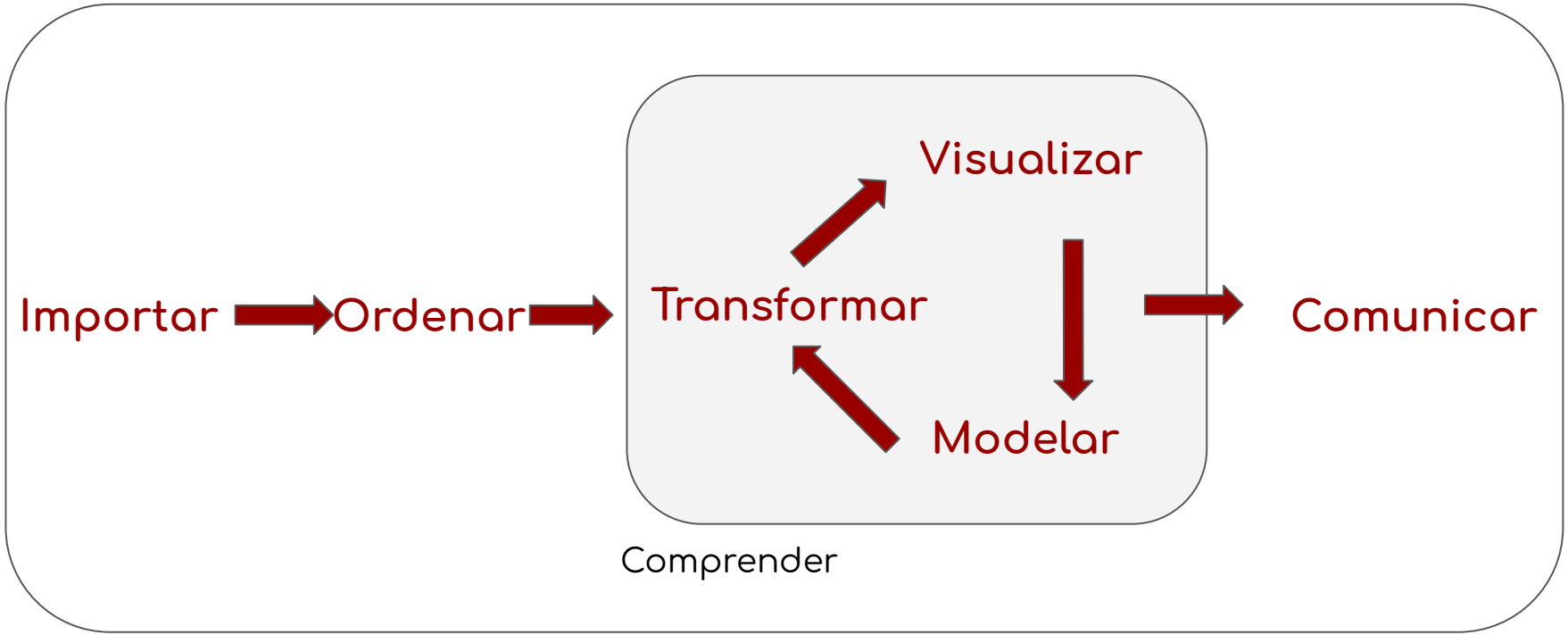


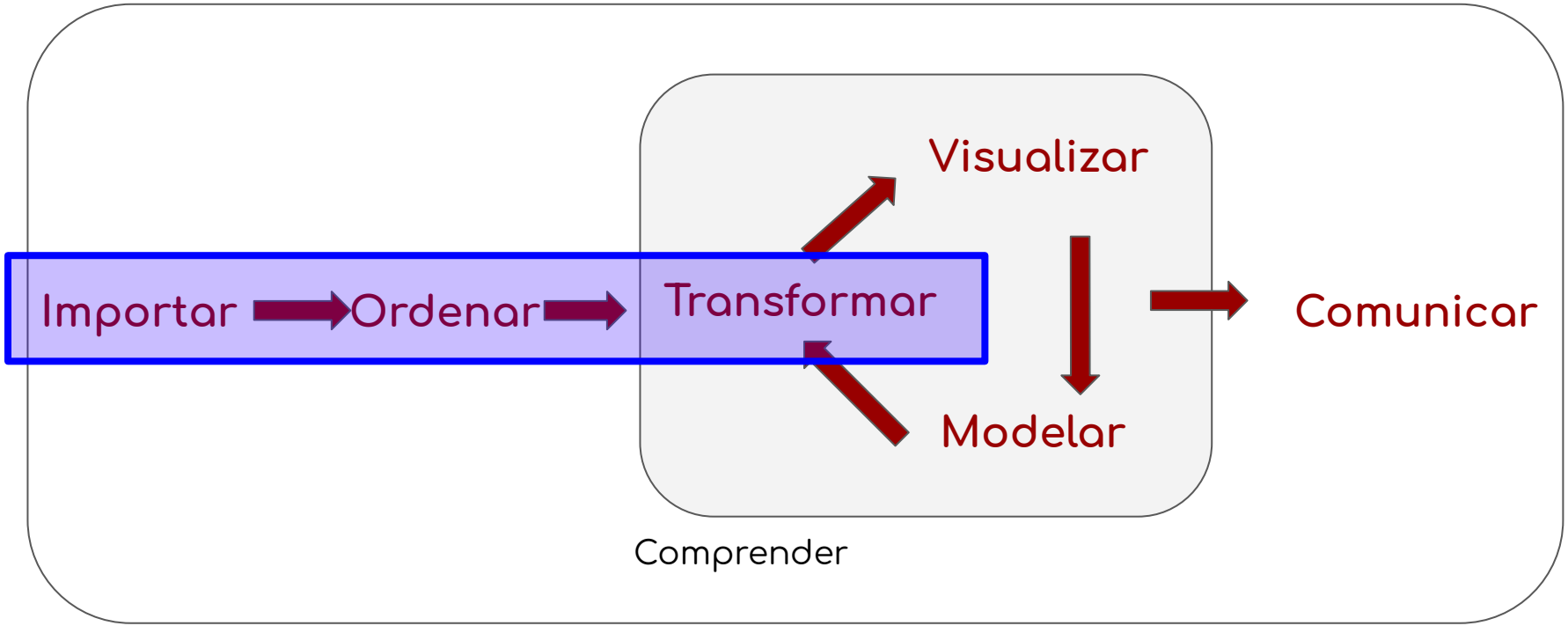
R-Ladies La Paz

MSc. Ing. Ruth Chirinos

Domar Datos (wrangling)

“Es el arte de tener tus datos en R de una forma conveniente para su visualización y modelado”


Importar
Ordenar
Visualizar
Transformar
Modelar
Comunicar
Comprender
Programar


Visualizar
Importar
Ordenar
Transformar
Comunicar
Modelar
Comprender
Programar

tibble 3.0.0

Tibble is a modern re-imagining of the data frame

```
tibble(
  x = 1:5,
  y = 1,
  z = x^2 + y
)
#> # A tibble: 5 x 3
#>       x     y     z
#>   <int> <dbl> <dbl>
#> 1     1     1     2
#> 2     2     1     5
#> 3     3     1    10
#> 4     4     1    17
#> 5     5     1    26
```

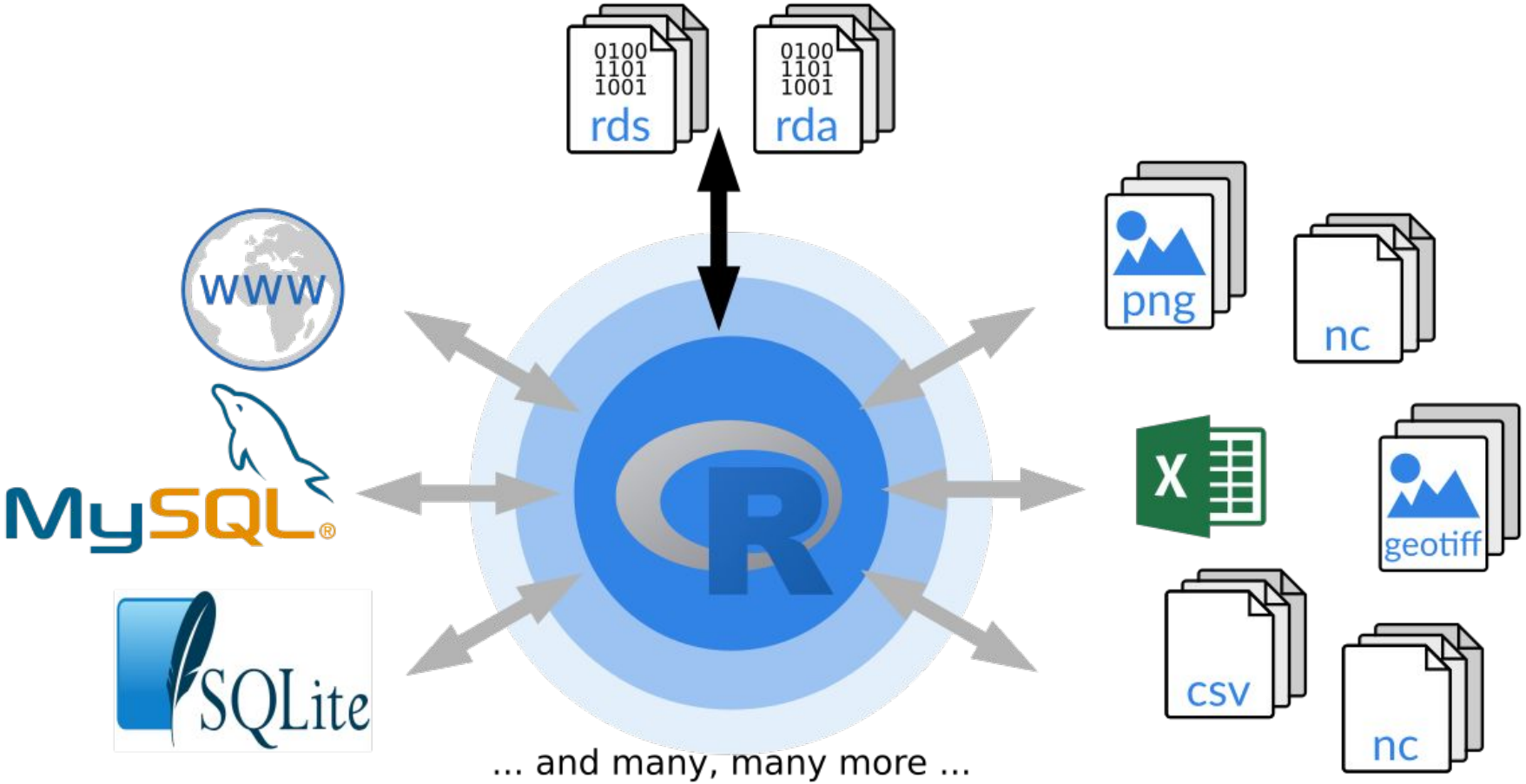
tibble vs data.frame

1. Impresion en consola
2. Selección de subconjuntos

tribble

```
tribble(
  ~x, ~y, ~z,
  #--|--|----
  "a", 2, 3.6,
  "b", 1, 8.5
)
#> # A tibble: 2 x 3
#>   x           y     z
#>   <chr> <dbl> <dbl>
#> 1 a           2   3.6
#> 2 b           1   8.5
```

Importar Datos


0100
1101
1001
rds
0100
1101
1001
rda
www
MySQL
SQLite
R
png
nc
X
geotiff
csv
nc
... and many, many more ...

Read Tabular Data - These functions share the common arguments:

read_*(file, col_names = TRUE, col_types = NULL, locale = default_locale(), na = c("", "NA"), quoted_na = TRUE, comment = "", trim_ws = TRUE, skip = 0, n_max = Inf, guess_max = min(1000, n_max), progress = interactive()**)**

A	B	C
1	2	3
4	5	NA

Comma Delimited Files

read_csv("file.csv"**)**

To make file.csv run:

write_file(x = "a,b,c\n1,2,3\n4,5,NA", path = "file.csv")

A	B	C
1	2	3
4	5	NA

Semi-colon Delimited Files

read_csv2("file2.csv"**)**

write_file(x = "a;b;c\n1;2;3\n4;5;NA", path = "file2.csv")

A	B	C
1	2	3
4	5	NA

Files with Any Delimiter

read_delim("file.txt", delim = "|"**)**

write_file(x = "a|b|c\n1|2|3\n4|5|NA", path = "file.txt")

A	B	C
1	2	3
4	5	NA

Fixed Width Files

read_fwf("file.fwf", col_positions = c(1, 3, 5)**)**

write_file(x = "a b c\n1 2 3\n4 5 NA", path = "file.fwf")

Tab Delimited Files

read_tsv("file.tsv"**)** Also **read_table().**

write_file(x = "a\tb\tc\n1\t2\t3\n4\t5\tNA", path = "file.tsv")

Save Data

Save **x**, an R object, to **path**, a file path, as:

Comma delimited file

```
write_csv(x, path, na = "NA", append = FALSE,
  col_names = !append)
```

File with arbitrary delimiter

```
write_delim(x, path, delim = " ", na = "NA",
  append = FALSE, col_names = !append)
```

CSV for excel

```
write_excel_csv(x, path, na = "NA", append =
  FALSE, col_names = !append)
```

String to file

```
write_file(x, path, append = FALSE)
```

String vector to file, one element per line

```
write_lines(x,path, na = "NA", append = FALSE)
```

Object to RDS file

```
write_rds(x, path, compress = c("none", "gz",
  "bz2", "xz"), ...)
```

Tab delimited files

```
write_tsv(x, path, na = "NA", append = FALSE,
  col_names = !append)
```

Gracias!