Module 1

Factors and R packages

Factors in R

Categorical variables - a variable that can take on one of a limited, and usually fixed, number of possible values

- The roll of a six-sided die: possible outcomes are 1,2,3,4,5, or 6.
- Demographic information of a population: gender.
- The blood type of a person: A, B, AB or O.

Factor is a used to store categorical variables and store it on multiple levels.

```r
blood <- c("B","AB","O","A","O","O","A","B")
str(blood)
blood_factor <- factor(blood)
str(blood_factor)
```

```
> blood <- c("B","AB","O","A","O","O","A","B")
> str(blood)
 chr [1:8] "B" "AB" "O" "A" "O" "O" "A" "B"
> blood_factor <- factor(blood)
> str(blood_factor)
 Factor w/ 4 levels "A","AB","B","O": 3 2 4 1 4 4 1 3
```

Levels and changing the order

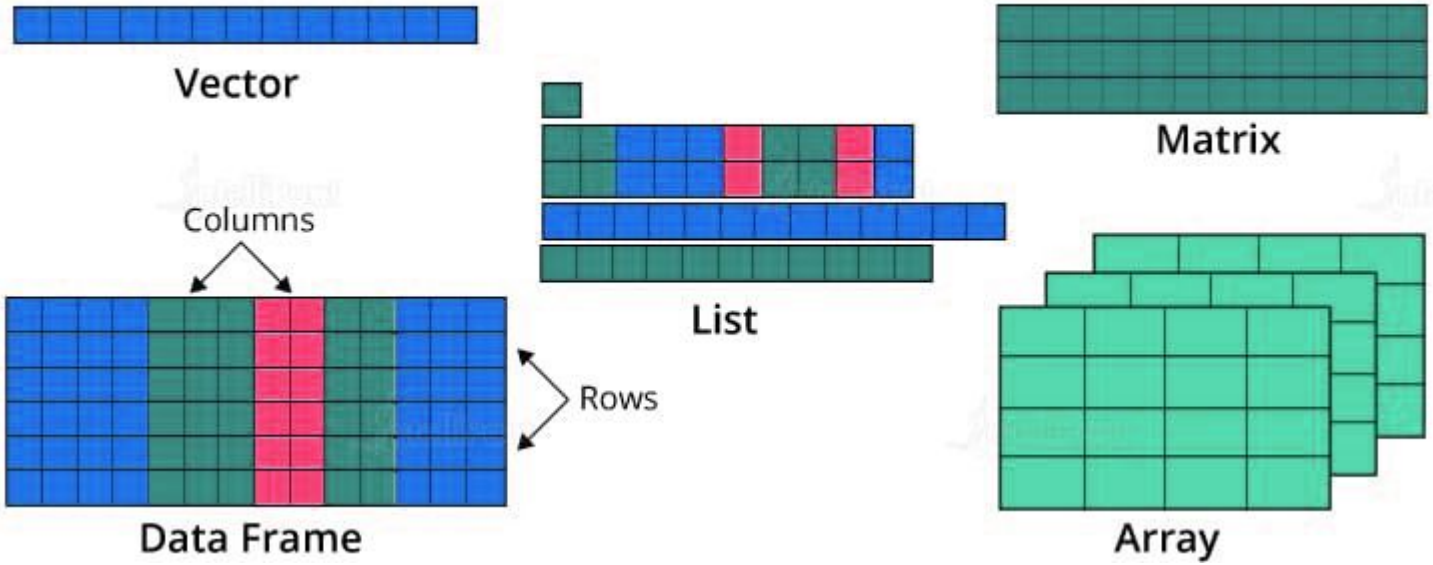
```
> blood <- c("B","AB","O","A","O","O","A","B")
> blood_factor <- factor(blood)
> str(blood_factor)
 Factor w/ 4 levels "A","AB","B","O": 3 2 4 1 4 4 1 3
> #To change the level name
> levels(blood_factor) <- c("BG_A","BG_AB","BG_B","BG_O")
> str(blood_factor)
 Factor w/ 4 levels "BG_A","BG_AB",..: 3 2 4 1 4 4 1 3
> #To change rhe order of levels
> blood_new_factor <- factor(blood,levels=c("O","A","B","AB"),
+                            labels=c("BT_O","BT_A","BT_B","BT_AB"))
> str(blood_new_factor)
 Factor w/ 4 levels "BT_O","BT_A",..: 3 4 1 2 1 1 2 3
> 
```

Ordinal Factors - when natural ordering exists

```
> tshirt <- c("L","S","M","M","S","L","L","S")
> tshirt_factors <- factor(tshirt,ordered = TRUE,levels= c("S","M","L"))
> tshirt_factors
[1] L S M M S L L S
Levels: S < M < L
```

We can compare ordered factors in R

```
> tshirt_factors[2]<tshirt_factors[1]
[1] TRUE
```


Vector
Matrix
Columns
List
Rows
Data Frame
Array

Packages in R

- R packages are collections of functions and data sets developed by the community
- Package will include
  - code (not only R code!),
  - documentation for the package and the functions inside,
  - some tests to check everything works as it should,
  - data sets.
- basic information about a package is provided in the DESCRIPTION file. For example to get the description of package graphics

```
packageDescription("graphics")
help(package = "graphics")
```

```
> packageDescription("graphics")
Package: graphics
Version: 4.3.1
Priority: base
Title: The R Graphics Package
Author: R Core Team and contributors worldwide
Maintainer: R Core Team <do-use-Contact-address@r-project.org>
Contact: R-help mailing list <r-help@r-project.org>
Description: R functions for base graphics.
Imports: grDevices
License: Part of R 4.3.1
NeedsCompilation: yes
Built: R 4.3.1; x86_64-w64-mingw32; 2023-06-16 07:33:30 UTC; windows

-- File: C:/Program Files/R/R-4.3.1/library/graphics/Meta/package.rds
```

Untitled1* | Untitled1* | Untitled2*

Source

```
1  packageDescription("graphics")
2  help(package = "graphics")
3
4
```

4:1 (Top Level) R Script

Console | Terminal | Background Jobs

R 4.3.1 · ~/

```
Version: 4.3.1
Priority: base
Title: The R Graphics Package
Author: R Core Team and contributors worldw
ide
Maintainer: R Core Team <do-use-Contact-add
ress@r-project.org>
```

Environment | History | Connections | Tutorial

Import Dataset | 122 MiB

R | Global Environment

Data	
A	num [1:2, 1:2] 1 2 3 4
Values	
a	3
age	30
b	4
colors	chr [1:4] "red" "green" "blue" "yellow"
f	14

Files | Plots | Packages | Help | Viewer | Presentation

R: The R Graphics Package | Find in Topic

The R Graphics Package

R Repositories

- a place where packages are located so you can install them from it.

  - CRAN: the official repository (The Comprehensive R Archive Network)
  - Bioconductor: this is a topic-specific repository intended for open-source software for bioinformatics.
  - Github

Install R packages(from CRAN)

```
install.packages("package_name")
```

```
> install.packages("vioplot")
WARNING: Rtools is required to build R packages but is not currently installed. Please
ownload and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/geoge/AppData/Local/R/win-library/4.3'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/vioplot_0.4.0.zip'
Content type 'application/zip' length 1390668 bytes (1.3 MB)
downloaded 1.3 MB

package 'vioplot' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\geoge\AppData\Local\Temp\RtmpolPWvl\downloaded_packages
```

```
install.packages(c("vioplot", "MASS"))
```

- To check what packages are installed on your computer, you can use:

```
> installed.packages()
             Package         LibPath
babynames    "babynames"     "C:/Users/geoge/AppData/Local/R/win-library/4.3"
cli          "cli"           "C:/Users/geoge/AppData/Local/R/win-library/4.3"
colorspace   "colorspace"    "C:/Users/geoge/AppData/Local/R/win-library/4.3"
fansi        "fansi"         "C:/Users/geoge/AppData/Local/R/win-library/4.3"
farver       "farver"        "C:/Users/geoge/AppData/Local/R/win-library/4.3"
ggplot2      "ggplot2"       "C:/Users/geoge/AppData/Local/R/win-library/4.3"
glue         "glue"          "C:/Users/geoge/AppData/Local/R/win-library/4.3"
gtable       "gtable"        "C:/Users/geoge/AppData/Local/R/win-library/4.3"
isoband      "isoband"       "C:/Users/geoge/AppData/Local/R/win-library/4.3"
labeling     "labeling"      "C:/Users/geoge/AppData/Local/R/win-library/4.3"
lifecycle    "lifecycle"     "C:/Users/geoge/AppData/Local/R/win-library/4.3"
magrittr     "magrittr"      "C:/Users/geoge/AppData/Local/R/win-library/4.3"
munsell      "munsell"       "C:/Users/geoge/AppData/Local/R/win-library/4.3"
pillar       "pillar"        "C:/Users/geoge/AppData/Local/R/win-library/4.3"
pkgconfig    "pkgconfig"     "C:/Users/geoge/AppData/Local/R/win-library/4.3"
R6           "R6"            "C:/Users/geoge/AppData/Local/R/win-library/4.3"
```

- Uninstalling a package is straightforward with the function **remove.packages()**, in your case:

```
> remove.packages("vioplot")
Removing package from 'C:/Users/geoge/AppData/Local/R/win-library/4.3'
(as 'lib' is unspecified)
```

- You can check what packages need an update with a call to the function:

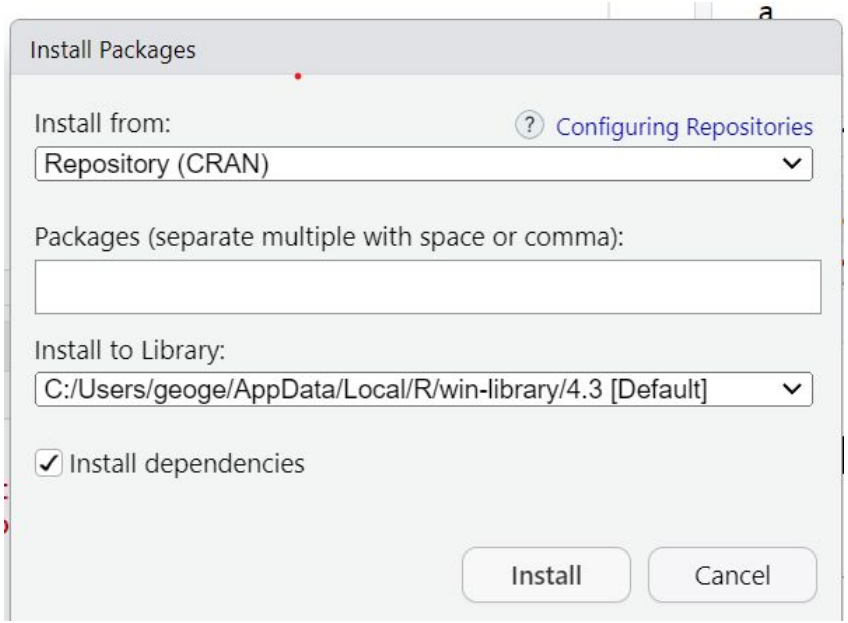
```
> old.packages()
           Package      LibPath                                  Installed Built
foreign    "foreign"    "C:/Program Files/R/R-4.3.1/library" "0.8-84"  "4.3.1"
KernSmooth "KernSmooth" "C:/Program Files/R/R-4.3.1/library" "2.23-21" "4.3.1"
Matrix     "Matrix"     "C:/Program Files/R/R-4.3.1/library" "1.5-4.1" "4.3.1"
mgcv       "mgcv"       "C:/Program Files/R/R-4.3.1/library" "1.8-42"  "4.3.1"
nlme       "nlme"       "C:/Program Files/R/R-4.3.1/library" "3.1-162" "4.3.1"
spatial    "spatial"    "C:/Program Files/R/R-4.3.1/library" "7.3-16"  "4.3.1"
survival   "survival"   "C:/Program Files/R/R-4.3.1/library" "3.5-5"   "4.3.1"
           ReposVer  Repository
foreign    "0.8-85"  "https://cran.rstudio.com/src/contrib"
KernSmooth "2.23-22" "https://cran.rstudio.com/src/contrib"
Matrix     "1.6-1.1" "https://cran.rstudio.com/src/contrib"
mgcv       "1.9-0"   "https://cran.rstudio.com/src/contrib"
nlme       "3.1-163" "https://cran.rstudio.com/src/contrib"
spatial    "7.3-17"  "https://cran.rstudio.com/src/contrib"
survival   "3.5-7"   "https://cran.rstudio.com/src/contrib"
```

- You can update all packages by using:

```
> update.packages()
foreign :
 Version 0.8-84 installed in C:/Program Files/R/R-4.3.1/library
 Version 0.8-85 available at https://cran.rstudio.com
```

User Interface to install

Tools -> Install Packages



Load Packages

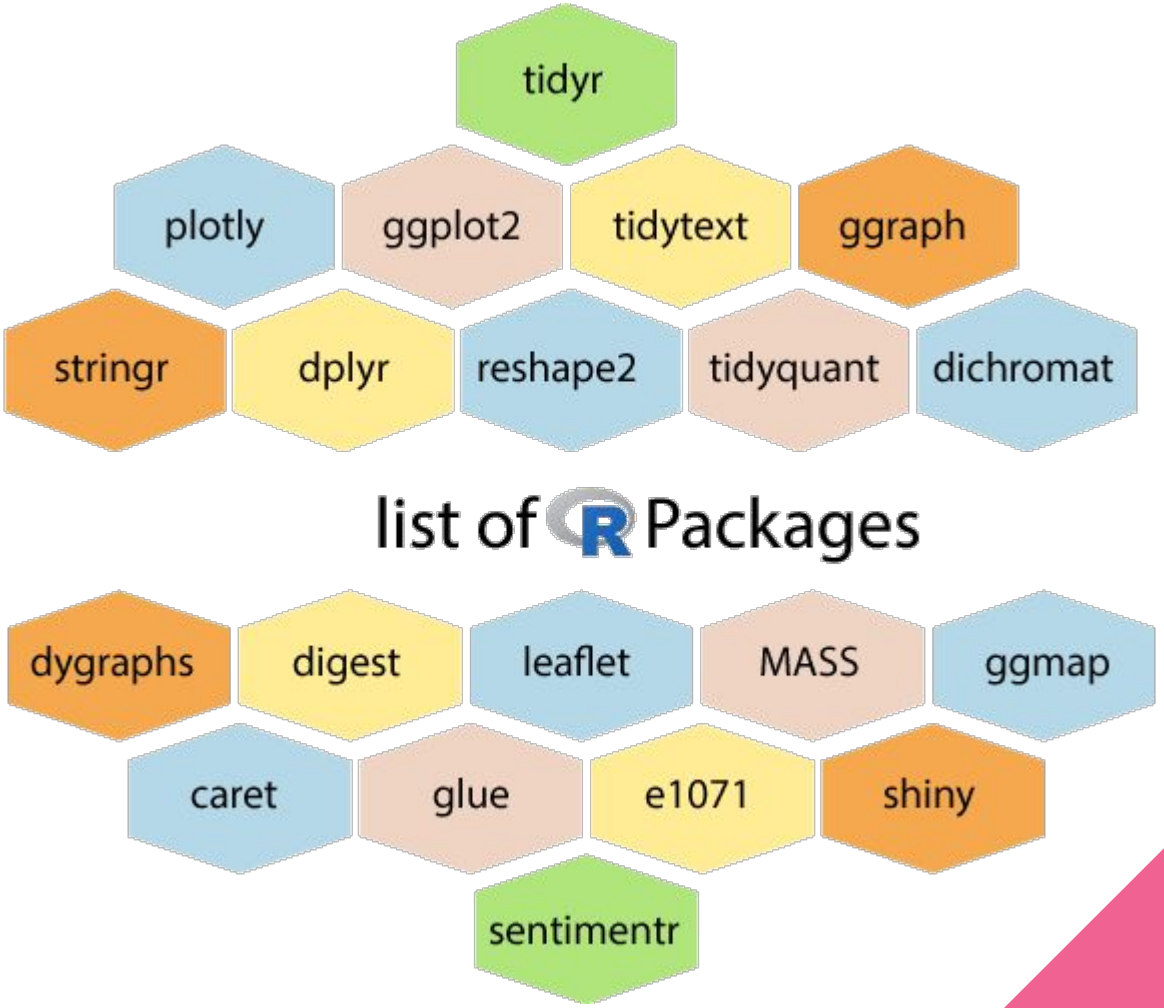
To load the package to memory

```
library(babynames)
births
```

packagename::functionname()

```
> babynames::births
# A tibble: 109 × 2
    year  births
   <int>   <int>
 1  1909 2718000
 2  1910 2777000
 3  1911 2809000
 4  1912 2840000
 5  1913 2869000
 6  1914 2966000
 7  1915 2965000
 8  1916 2964000
 9  1917 2944000
10  1918 2948000
# ℹ 99 more rows
# ℹ Use `print(n = ...)` to see more rows
```

Note: library() is the command used to load a package, and it refers to the place where the package is contained, usually a folder on your computer, while a package is the collection of functions bundled conveniently.


tidyr
plotly
ggplot2
tidytext
ggraph
stringr
dplyr
reshape2
tidyquant
dichromat
list of R Packages
dygraphs
digest
leaflet
MASS
ggmap
caret
glue
e1071
shiny
sentimentr