Introduction to R

November 5, 2025

Alliance Bioversity International & CIAT

What is R?

- It is an object-oriented programming language, focused on computation or statistical processing.
- It is free and open source.
- It integrates a wide variety of graphical and statistical techniques.
- It is a very understandable programming language.

Why to use R

- It is a tool that allows us to automate many processes
- It is easy to understand and use
- It has a large community of developers worldwide
- It is equipped with the most up-to-date statistical analysis techniques
- And many more …

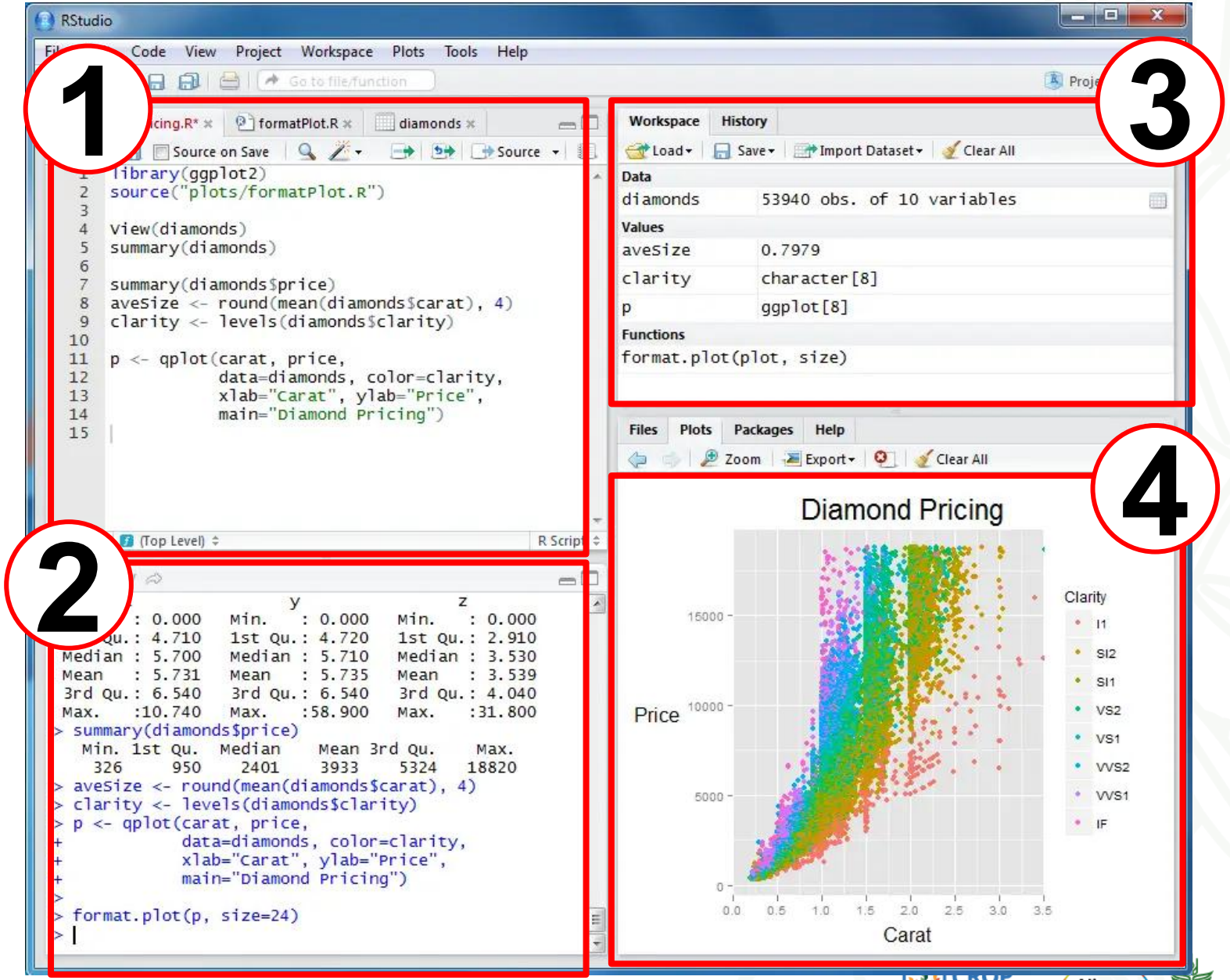
RStudio Desktop

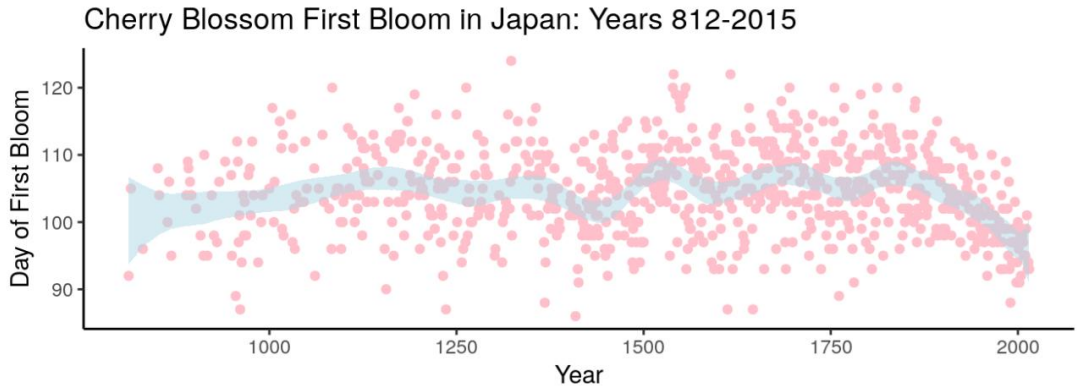
- It is an IDE or integrated development environment.
- It is a program that integrates several functionalities oriented to facilitate the programming of routines.
- It has a more developed graphical GUI interface.In most cases they are multiplatform.
- By integrating all the tools in one place it makes our life easier.
- We can download it from this link:

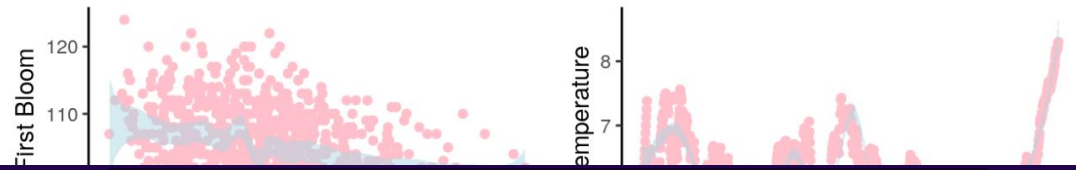
https://posit.co/downloads/

RStudio (GUI)

1. Script editor
2. Console/Terminal
3. Workspace
4. Graph viewer, library and help page



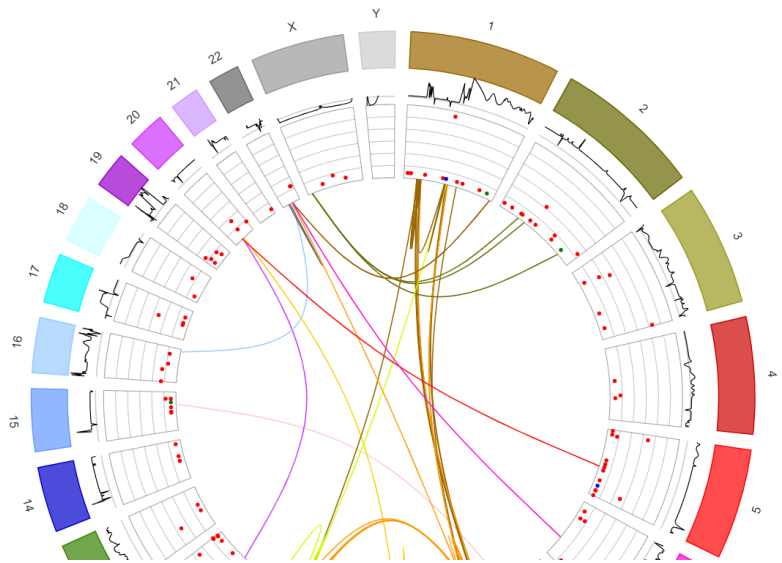




gallery.shinyapps.io/genome_browser/

ICGC PANCREATIC CANCER (DUCTAL ADENOCARCINOMA) - GENOME VIEWER

More Information



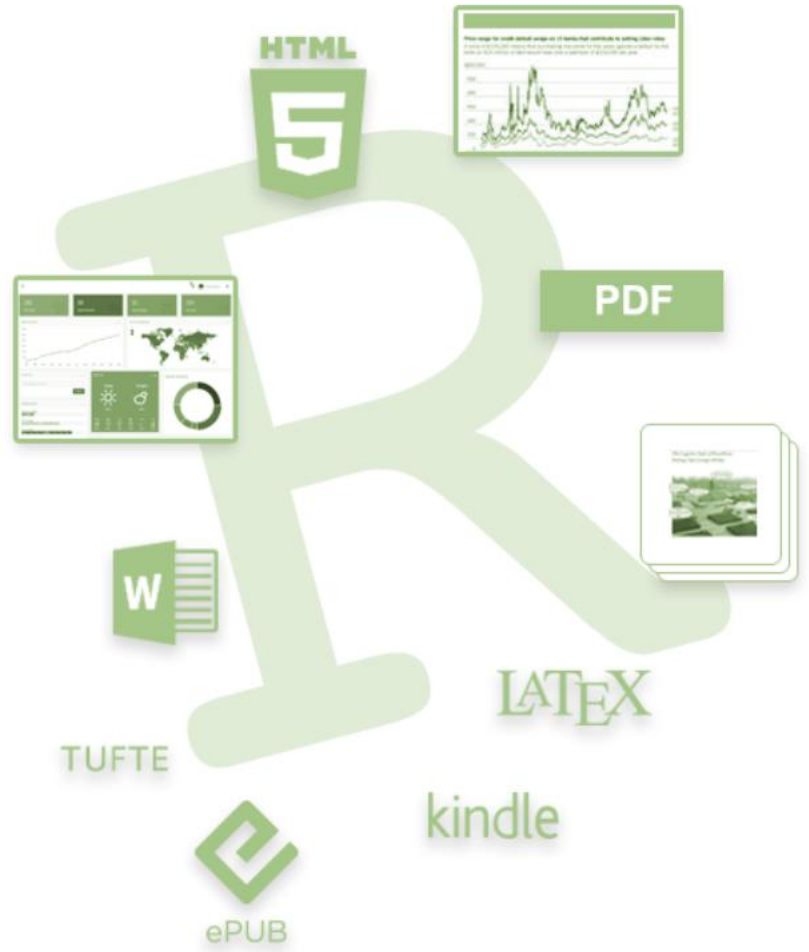
Cohort Top ClinVar Gene Summary:

HGNC	Chr	Start	From	To	Consequence	Count
SMARCA4	19	11144847	C	T	missense_variant	18
TP53	17	7578437	G	A	stop_gained	18
KRAS	12	25398284	C	T	missense_variant	12
TP53	17	7578437	G	A	exon_variant	10
SMARCA4	19	11144847	C	T	exon_variant	8
TP53	17	7577121	G	A	downstream_gene_variant	6
TP53	17	7577121	G	A	missense_variant	6
KRAS	12	25398285	C	G	missense_variant	4
SMARCA4	19	11144847	C	T	downstream_gene_variant	4
TP53	17	7578437	G	A	downstream_gene_variant	4

Please select a donor ID:

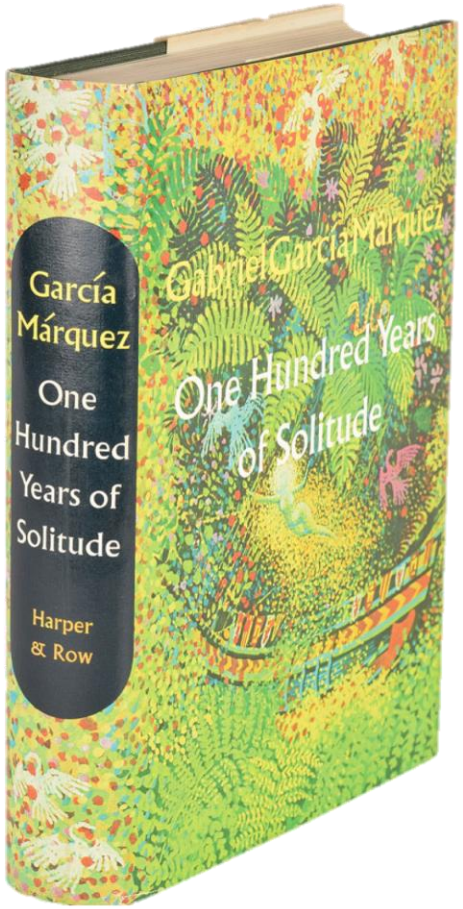
DO49184

SNP Consequences



CROP TRUST
Alianza Bioversity & CIAT
CGIAR

Book



Paragraph

MANY YEARS LATER, as he faced the firing squad, Colonel Aureliano Buendía was to remember that distant afternoon when his father took him to discover ice. At that time Macondo was a village of twenty adobe houses, built on the bank of a river of clear water that ran along a bed of polished stones, which were white and enormous, like prehistoric eggs. The world was so recent that many things lacked names, and in order to indicate them it was necessary to point. Every year during the month of March a family of ragged gypsies would set up their tents near the village, and with a great uproar of pipes and kettledrums they would display new inventions. First they brought the magnet. A heavy gypsy with an untamed beard and sparrow hands, who introduced himself as Melquíades, put on a bold public

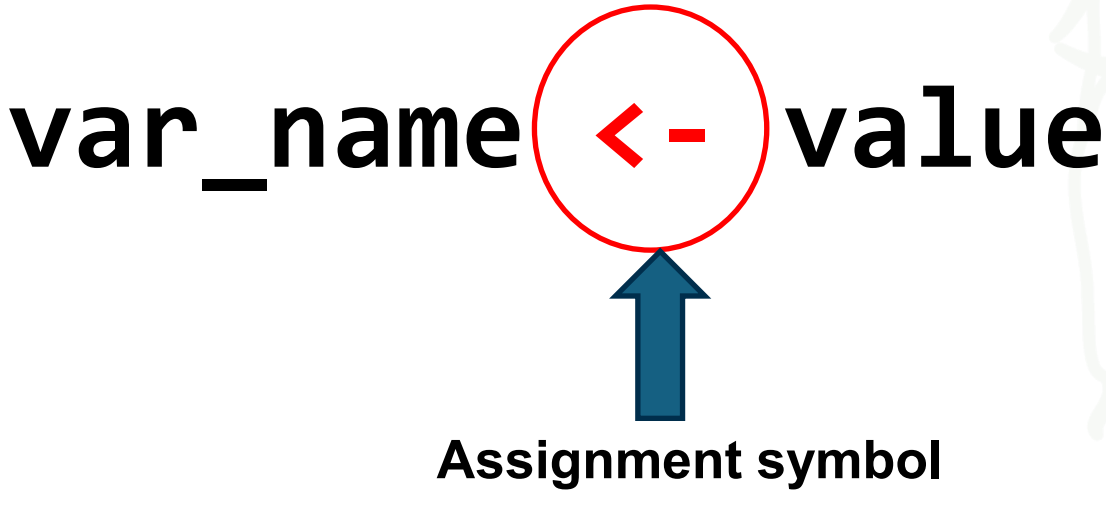
Words

Letters

Basic concepts

What is an object/variable?

- In object-oriented programming, an object is an empty container that the programmer can fill with whatever they decide.
- These containers are often referred to as "variables" in R due to the common use of the term in statistics.
- Variables are composed of values that have different types, such as text, numeric, logical, integer, and others.
- Variable names cannot contain spaces, special characters, or numbers at the beginning.



Types of values in R

- Logical: `TRUE, FALSE`
- Integers: `1, 2, 3, 4, ...`
- Real Numbers: `1.342, 3.45565, 0.0879985`
- Character/Text: `“string,” “text”`
- Missing Values: `NA`
- Non-calculable values: `NaN`
- Very large/small values: `Inf, -Inf`
- Null values: `NULL`

How can you identify the type of values in a variable?

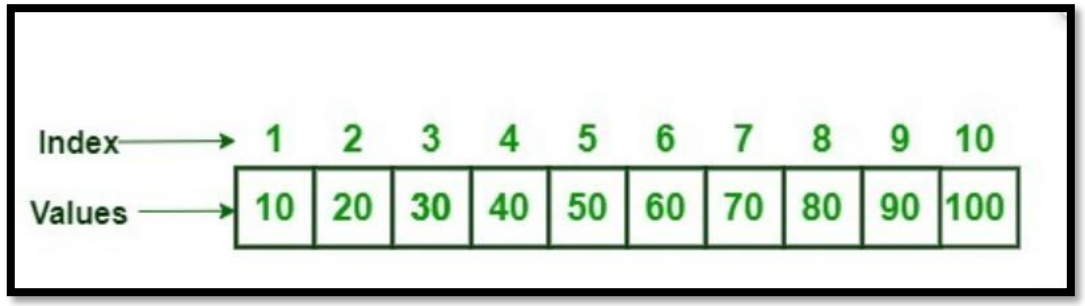
R provides tools for determining the types of values:

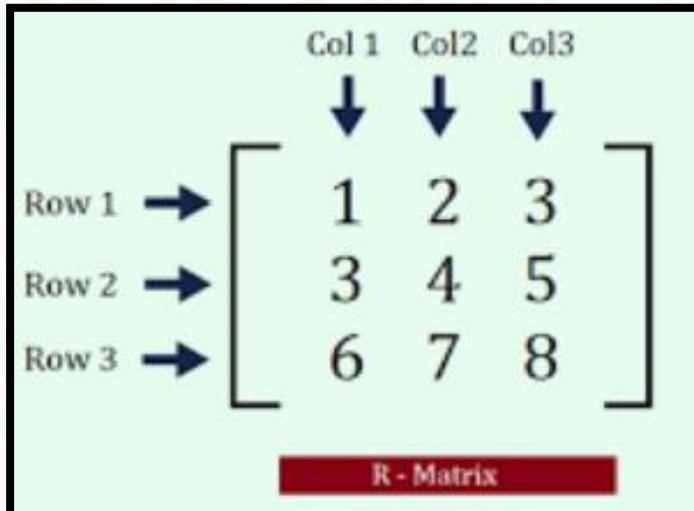
`is``(variable)` and **`class`**`(variable)`. These commands display the value type on the screen.

CROP TRUST
Alianza Bioversity & CIAT
CGIAR

Vectors and matrices – Basic data structures

- **Vectors:**
  - In programming, objects often consist of multiple ordered and indexed values.
  - These are referred to as “vectors" in R and are fundamental to the programming process.
  - Often, variables are composed of vectors.
- **Matrices:**
  - A collection of elements (numbers) arranged in rows and columns.
  - Every row and column is considered a vector.
  - Matrices must have a fixed number of rows and columns.





Indexing

- All objects/variables in R have a way to access their elements.

- **Vectors**: Elements are ordered sequentially, starting from position 1.
  - Access syntax: Variable[position].

- **Matrices**: Access elements by specifying the row and column indices.
  - Matrix[row number, column number] – extracts a single cell value.
  - Matrix[row number,] – extracts the entire row.
  - Matrix[, column number] – extracts the entire column.

CROP TRUST
Alianza Bioversity & CIAT
CGIAR

Exercise

1. Create a vector containing the **first ten letters** of the alphabet.
2. Assign it to the variable "`vec_exercise`".
3. Obtain the letters in **positions 3, 7, and 9**, then print them in the console.
4. Obtain the **first five letters**.
5. Obtain the **last five letters**.
6. Reverse the order of the "`vec_exercise`".

Data frames

- They are like matrices.
- They are organized into rows and columns.
- You can name both rows and columns.
- They allow all types of values and do not affect each other.
- This is the most used format.
- All of its elements are easily accessible.

	puntos.de.ventas	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
1	cali	6290408	2504943	2701923	7565694	6585867	7494628	4846782	2860766	3948181	5888232	3812555	9623078
2	bogota	5155575	7822896	5235454	8007954	235291	7310997	412739	2515766	2325974	2845886	4732304	5092811
3	medelllin	7693013	6310346	7233947	4048892	6945718	3026637	9908147	4941736	2774863	1673729	6142101	4452125
4	cartagena	1635552	848500	1298844	9388080	9651571	7266864	3863462	2196088	9467877	9404306	8298670	1411678
5	pasto	2058038	6393276	7917958	6435180	4803311	9422771	3453430	6301717	7770590	8671723	3483237	4524928

CROP TRUST
Alianza Bioversity & CIAT
CGIAR

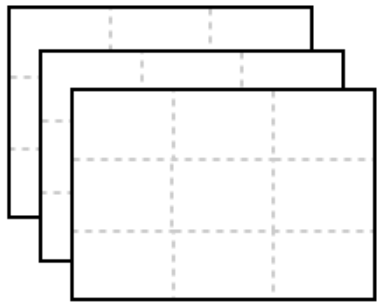
Filter by row

- Filters are always necessary when managing and processing databases.
- R provides several methods for filtering vectors, matrices, and data frames.

```
vector[ logical condition ]
data.frame[ logical condition , ]
```

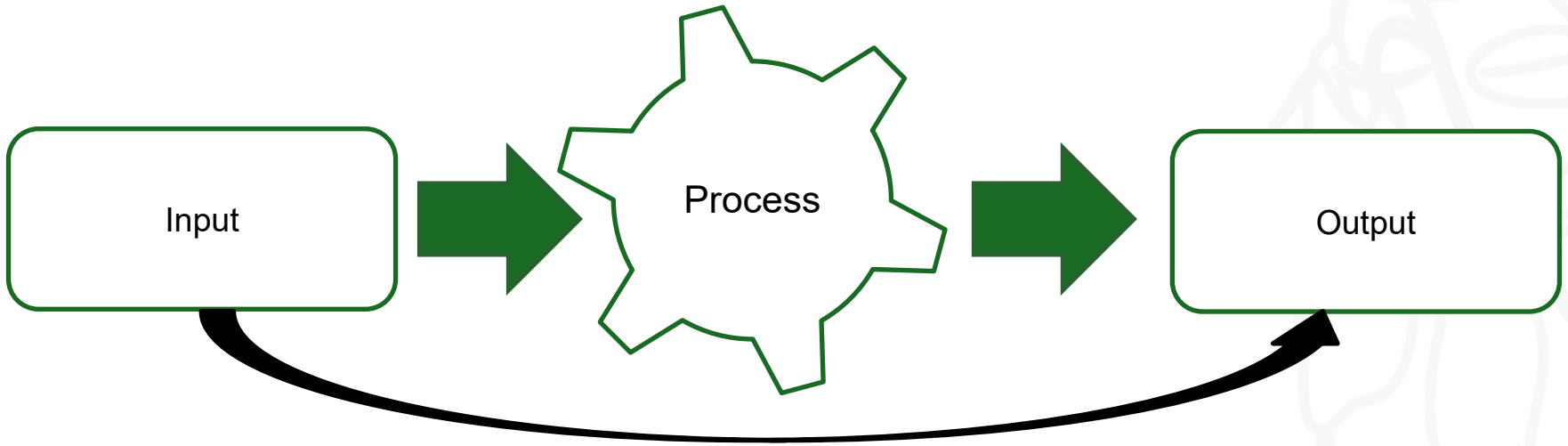
CROP TRUST
Alianza Bioversity & CIAT
CGIAR

	single type	multiple types
1D	Vector	List
2D	Matrix	Data frame
nD	Array	



Algorithmic thinking

- A set of systematic operations performed to solve a type of problem.
- Algorithmic thinking involves abstraction, modeling the problem, logical deduction, synthesizing a solution, and writing the correct algorithm.



Exercise

Scenario:

The first monthly sales totals for four stores in the listed cities for 2024 are provided. Calculate each store's total sales for the quarter.

1. Identify the inputs.
2. Define the desired outcome.
3. Establish steps to process inputs and convert them into the desired outcomes.

	Month		
Sales points	**January**	**February**	**March**
Cali	3643	4652	2145
Bogotá	2011	4345	4708
Medellín	2820	1995	3273
Cartagena	4487	3574	4832

Solution

	Month			
Sales points	**January**	**February**	**March**	**Total**
Cali	3643	4652	2145	**10440**
Bogotá	2011	4345	4708	**11064**
Medellín	2820	1995	3273	**8088**
Cartagena	4487	3574	4832	**12893**

- **Inputs**:
  - Table showing monthly sales totals.
  - The rows contain records of sales points in cities.
  - The columns contain monthly sales data.
- **Processing:**
  - Sum the rows and enter the total in the last column.
- **Outputs**:
  - Table showing the sum of monthly sales (quarter total) for each sales point.

What if I am interested in total sales per year?

Functions

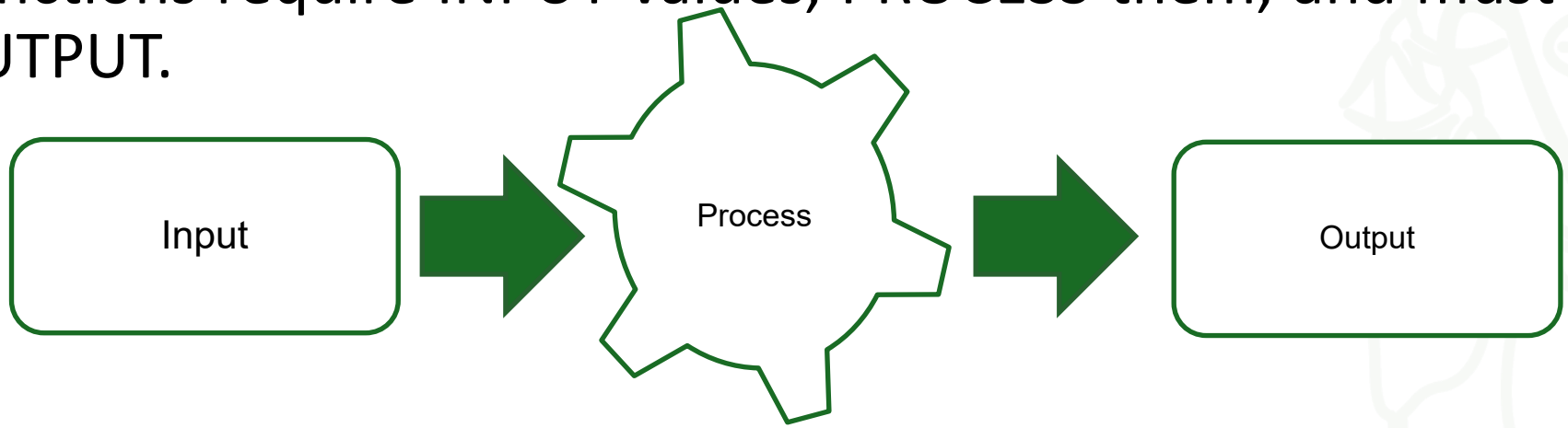
$$\sigma^2 = \frac{\sum_1^N (x_i - \bar{X})^2}{N}$$

$$g(x) = \frac{x+3}{2x+1}$$

$$Mean = \bar{\chi} = \frac{\sum_i^n \chi_i}{n}$$

What is a function?

- A function is a type of object.
- It is the foundation of all programming in R.
- A function is defined as a set of processes or algorithms that perform a specific task or calculation.
- Functions require INPUT values, PROCESS them, and must return an OUTPUT.



How are functions defined in R?

Functions have a specific structure:

1. A name must be assigned to the function.
2. Input values must be specified. These values can be of any type or structure that we have seen before, such as data frames, vectors, or matrices.
3. The function should perform an operation based on the input values.
4. Finally, it must return at least one object, which can be of any type.

CROP TRUST
Alianza Bioversity & CIAT
CGIAR

Function	Description
mean()	Calculates the arithmetic mean of a numeric vector.
sum()	It returns the sum of all the elements in a vector or array.
length()	It returns the number of elements in a vector or list.
head()	Displays the first few rows of a data frame or the initial elements of a vector.
tail()	Displays the last few rows of a data frame or the final elements of a vector.
seq()	Generates a sequence of numbers based on a specified start number, end number, and step size.
sort()	It sorts a vector in ascending or descending order.
paste()	Concatenates strings using a specified separator.
which()	It identifies the indices of the elements in a vector that meet a given condition.
table()	It creates a frequency table of the elements in a vector or factor.

R libraries

reticulate
applicable
tidyverse
gt
timetk
Shiny
tibble
purrr
R6
readr
rlang
dplyr
modeltime
spatialsample
sparklyr
tidymodels
ggplot2
vctrs
torch
Vetiver
scales
plumber
fs
CLOCK
R2D3
TUNE

What is a library?

- In R, packages or libraries are sets of functions designed to solve a variety of problems, primarily statistical ones.
- R comes with a series of installed packages called "base“, which contain basic statistical processing functions.
- R provides the ability to install new packages when our operations or processes require more specialized functions that are not included in the "base" package.
- Each package includes helpful documentation explaining the contents, functions, and other details.
- R offers a massive catalog of packages thanks to the large community dedicated to developing new ones.
- Essentially, there is a package designed for almost any task we can think of or need to perform. You just need to know how to find it.

How are packages installed in R?

`install.packages("package_name")`

- This command downloads the package from the CRAN repository.
- This command should only be called once.
- The package will be installed, but it has not been called yet.

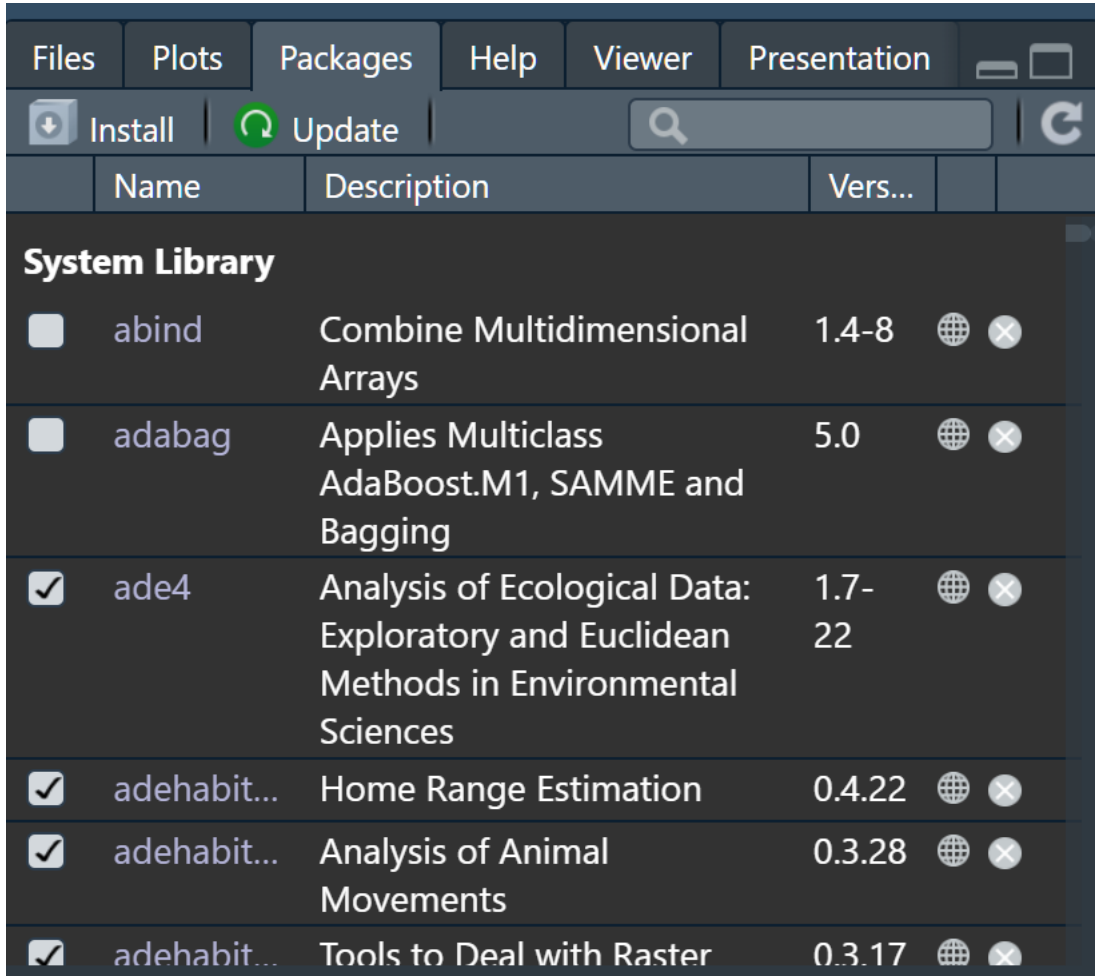
Unquoted

`library(package_name)`

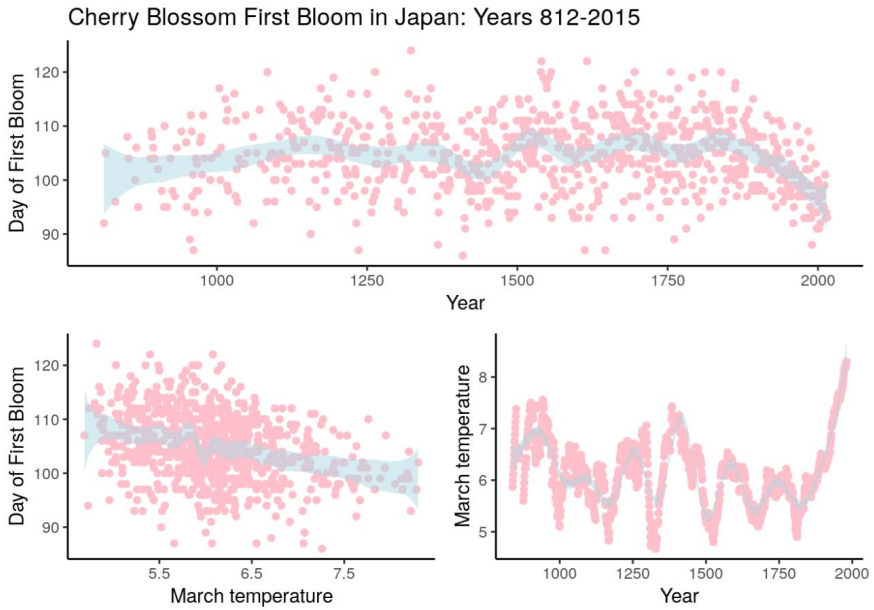
- This function loads the package functions into the R session.
- It's good practice to load the libraries at the beginning of the script only.

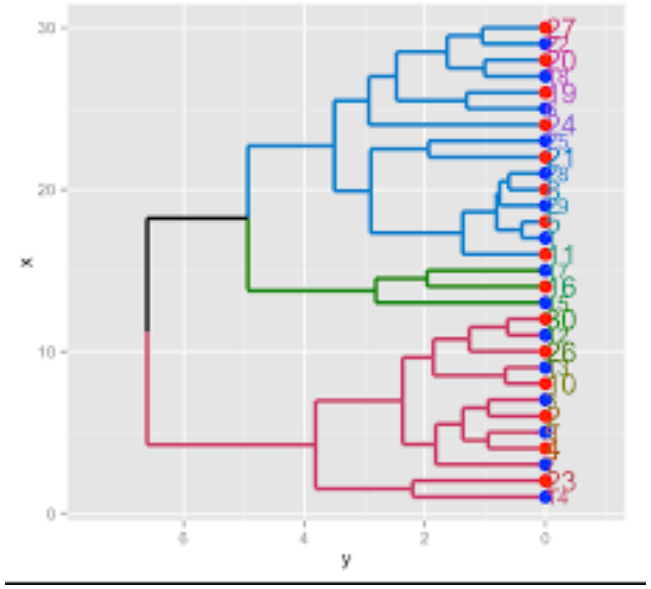
or simply:

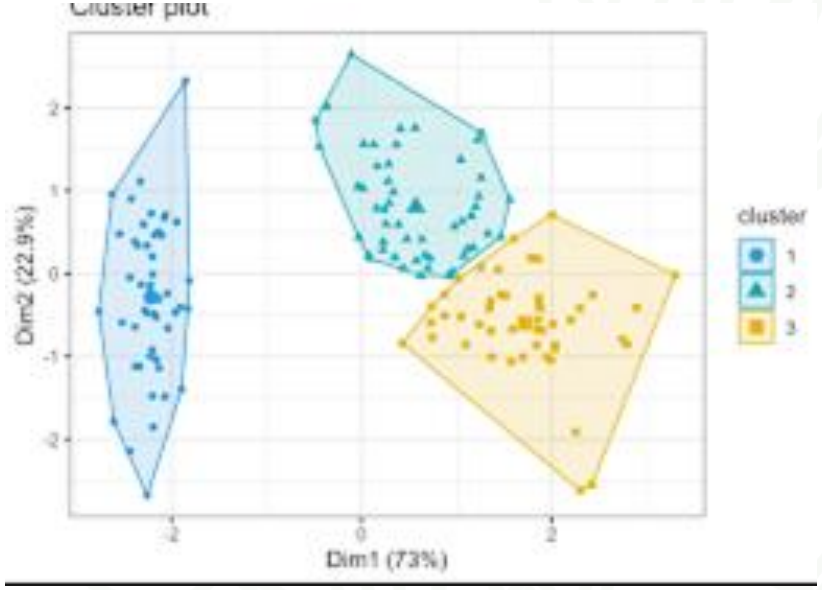
`library(pacman); pacman::p_load(package_name)`

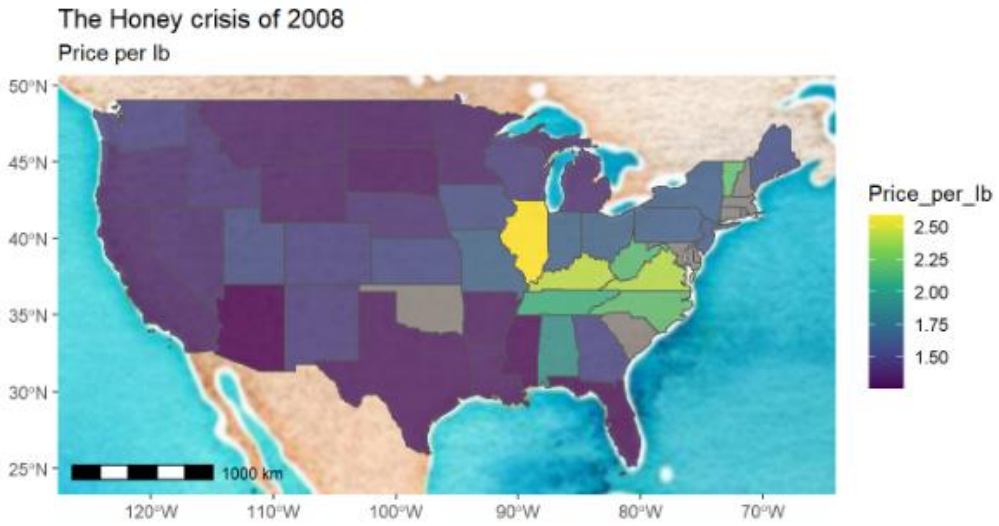


You can check which packages are installed and loaded into the R session.

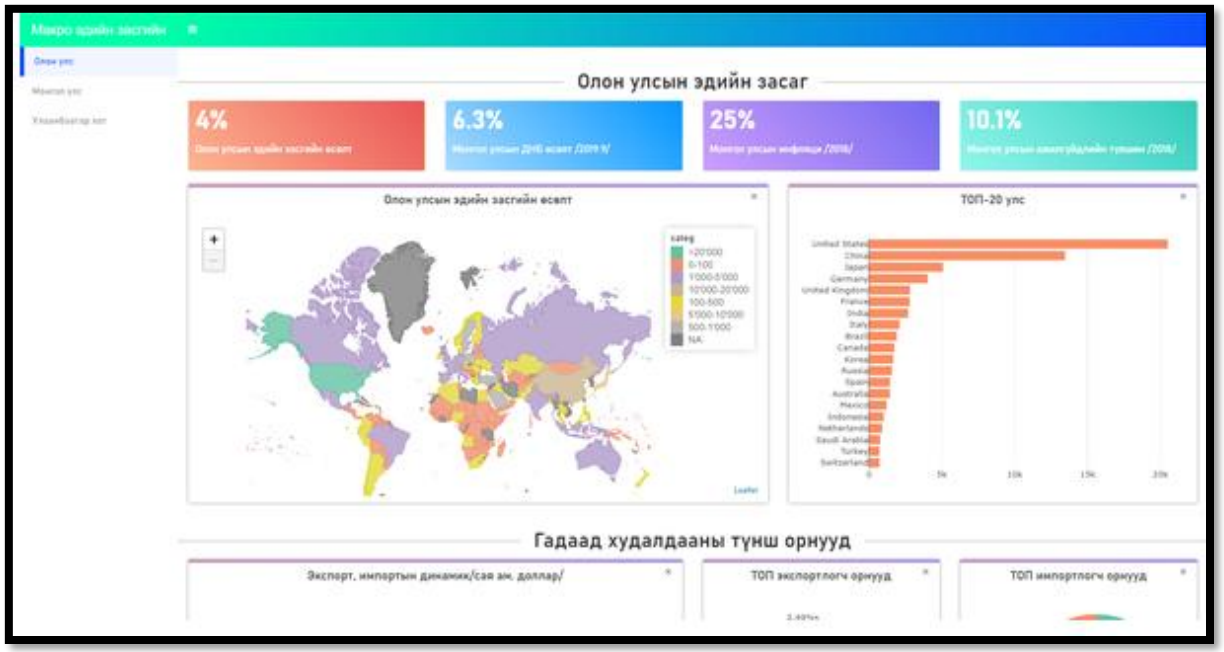

Cherry Blossom First Bloom in Japan: Years 812-2015
Day of First Bloom
Year
March temperature
Day of First Bloom
March temperature
Year

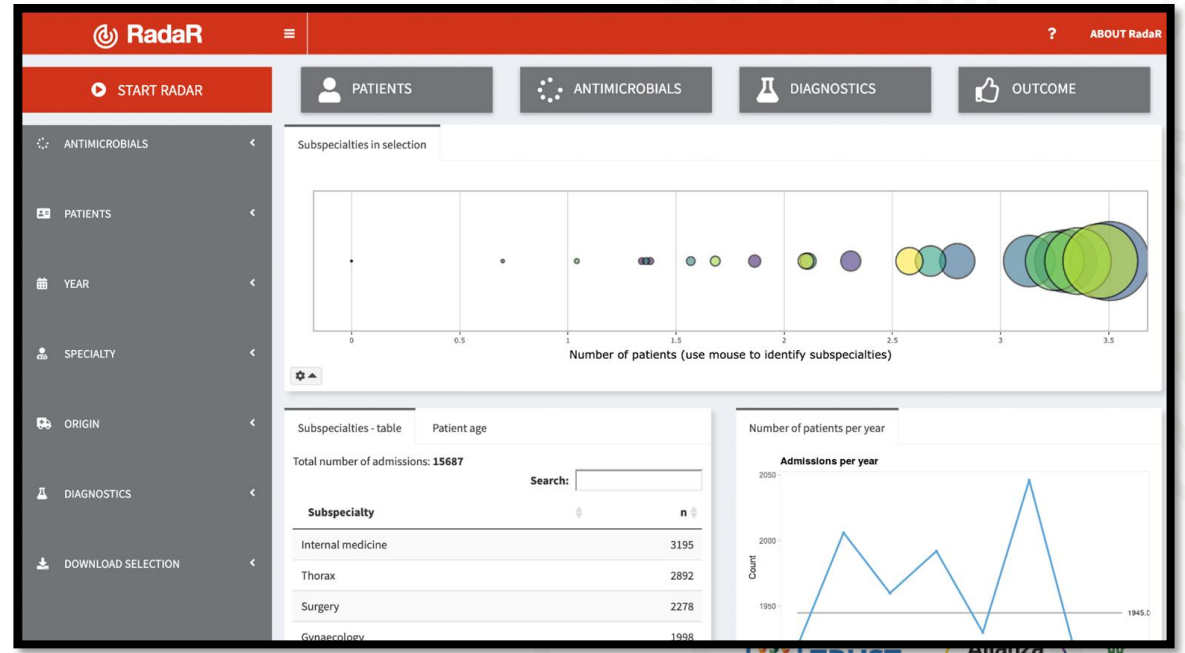

x
y

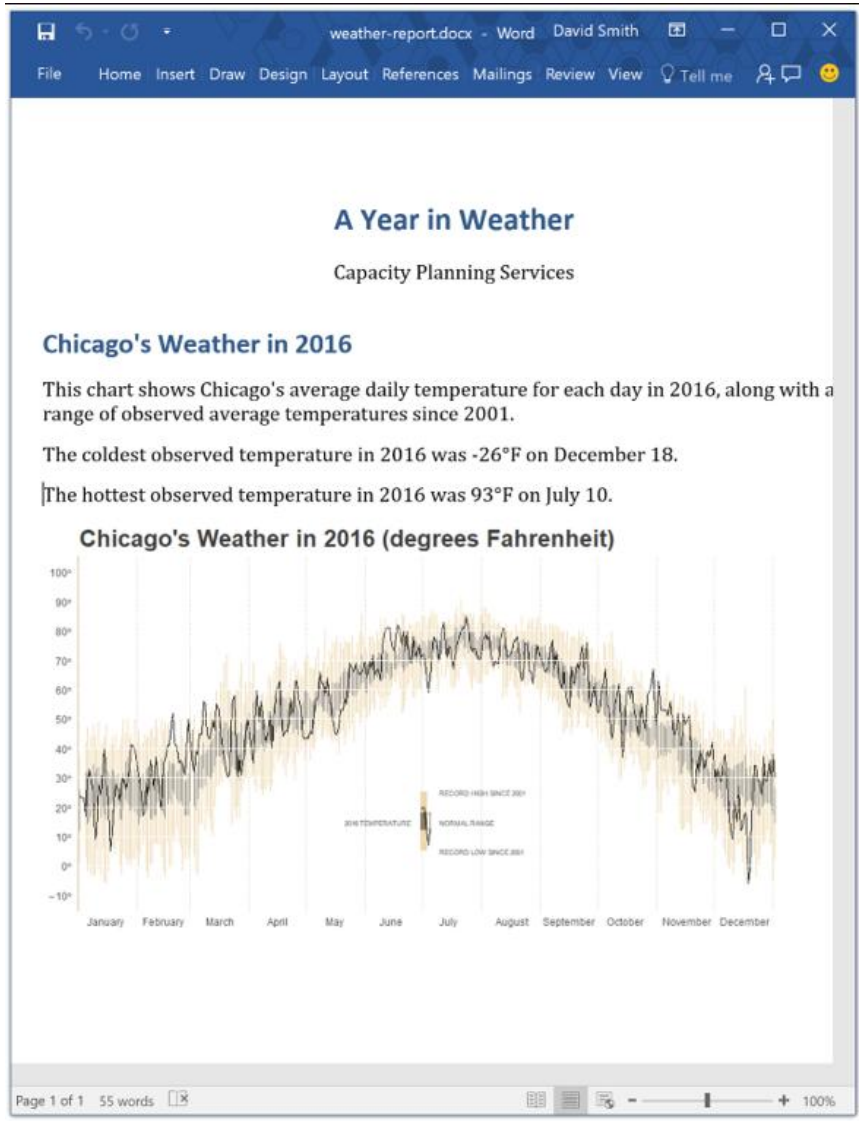

Cluster plot
Dim2 (22.9%)
Dim1 (73%)
cluster
1
2
3


The Honey crisis of 2008
Price per lb
Price_per_lb
2.50
2.25
2.00
1.75
1.50
1000 km

level
500
1000
level
1000
500


Олон улсын эдийн засаг
4%
6.3%
25%
10.1%
Олон улсын эдийн засгийн өсөлт
ТОП-20 улс
Гадаад худалдааны түнш орнууд
Экспорт, импортын динамик/сая ам. доллар/
ТОП экспортлогч орнууд
ТОП импортлогч орнууд


RadaR
ABOUT RadaR
START RADAR
PATIENTS
ANTIMICROBIALS
DIAGNOSTICS
OUTCOME
Subspecialties in selection
Number of patients (use mouse to identify subspecialties)
Subspecialties - table
Patient age
Total number of admissions: 15687
Search:
Subspecialty
n
Internal medicine 3195
Thorax 2892
Surgery 2278
Gynaecology 1998
Number of patients per year
Admissions per year
Count


weather-report.docx - Word
David Smith
File
Home
Insert
Draw
Design
Layout
References
Mailings
Review
View
Tell me
A Year in Weather
Capacity Planning Services
Chicago's Weather in 2016
This chart shows Chicago's average daily temperature for each day in 2016, along with a range of observed average temperatures since 2001.
The coldest observed temperature in 2016 was -26°F on December 18.
The hottest observed temperature in 2016 was 93°F on July 10.
Chicago's Weather in 2016 (degrees Fahrenheit)
100°
90°
80°
70°
60°
50°
40°
30°
20°
10°
0°
−10°
January
February
March
April
May
June
July
August
September
October
November
December
Page 1 of 1
55 words
100%

rmarkdown
www.rstudio.com

Questions?