R for Data Analysis

WOMEN IN DATA
ACADEMY

TECH TALENT
ACADEMY

Session Content

- Understanding Data
- Installing R
- Installing Rstudio
- Exploring Rstudio

Understanding Data

Data is increasingly being collected by devices to analyse via ‘the cloud’

The ‘Internet of things’ has risen recently due to the ability to easily report vast amounts of data- often referred to as big data.

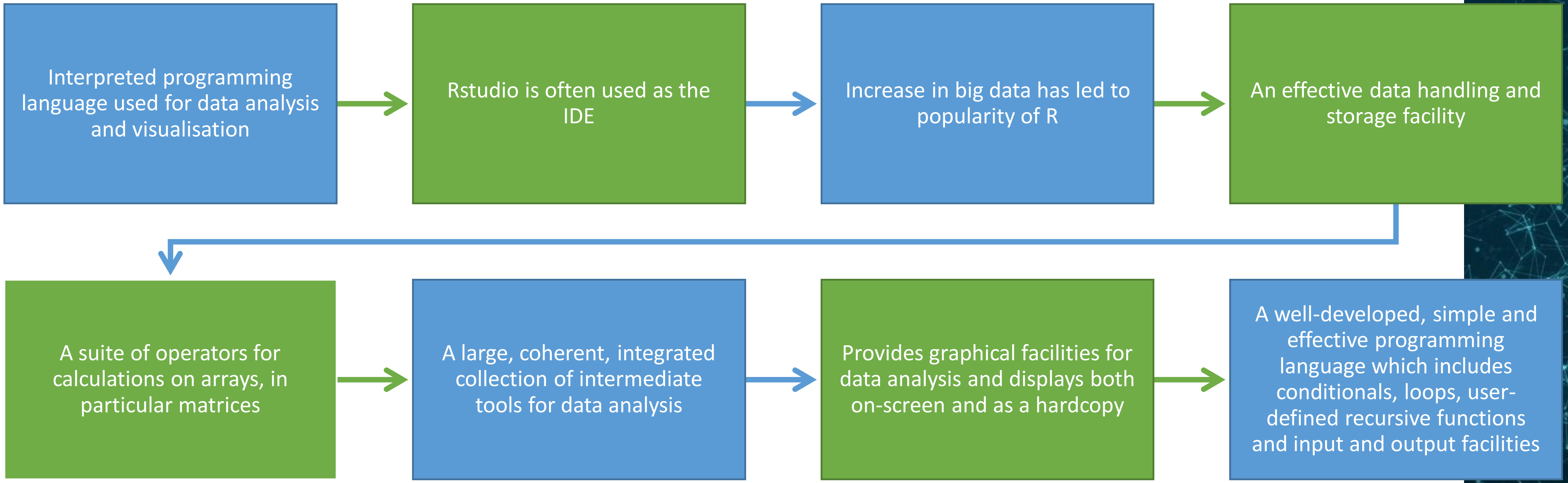
Data analysis is the practice of collecting data, converting it to useful information

Two procedures it must undergo ‘ data processing’ and ‘data cleaning’.

WOMEN IN DATA
ACADEMY

TECH TALENT
ACADEMY

What is R?



Interpreted programming language used for data analysis and visualisation

Rstudio is often used as the IDE

Increase in big data has led to popularity of R

An effective data handling and storage facility

A suite of operators for calculations on arrays, in particular matrices

A large, coherent, integrated collection of intermediate tools for data analysis

Provides graphical facilities for data analysis and displays both on-screen and as a hardcopy

A well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities

WOMEN IN DATA
ACADEMY

Step 1: Install R

- https://cran.r-project.org/
- Choose for Windows/mac
- Choose install for first time
- Choose DOWNLOAD R
- When it is completed – run the install wizard
- Choose to add core files, and 64 bit only in the select components
- Choose no (accept defaults)
- Choose a name E.g. R
- Choose additional – create desktop icon
- Installation should now begin

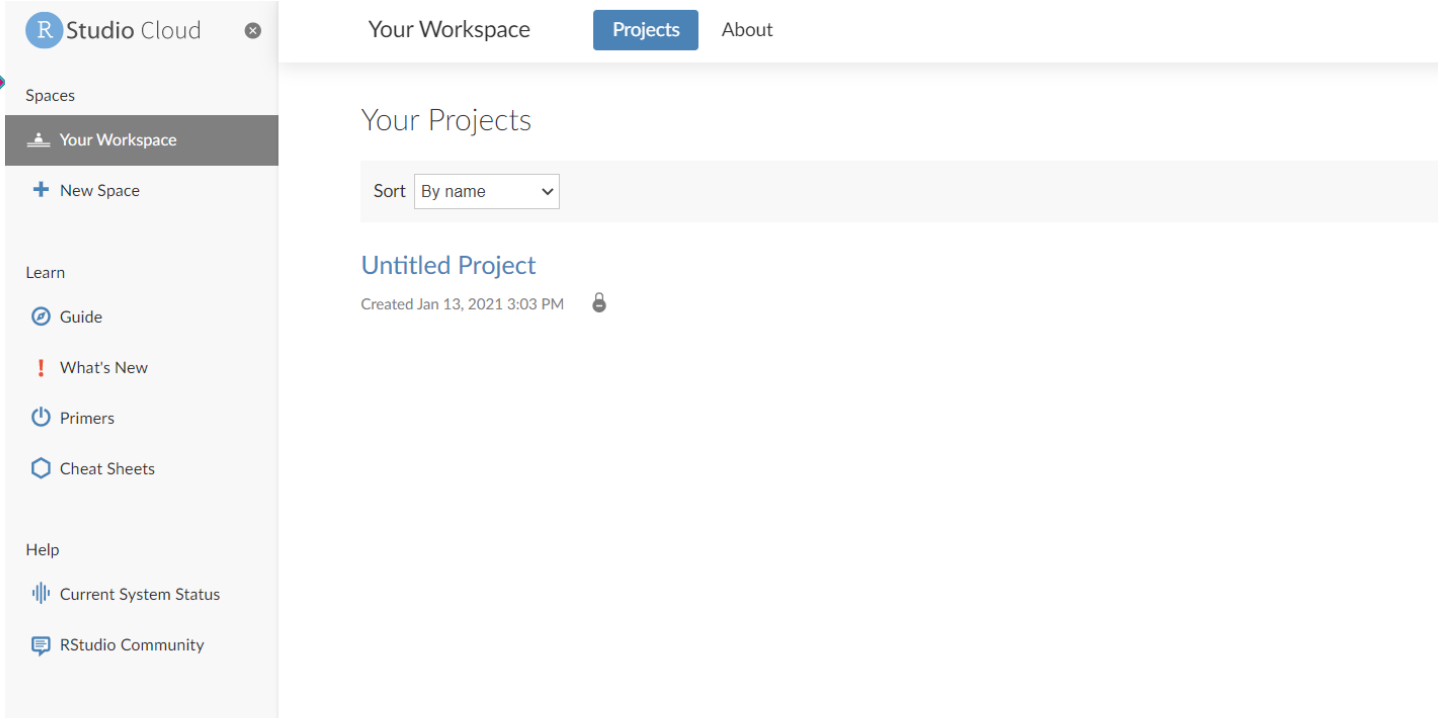
https://www.datacamp.com/community/tutorials/installing-R-windows-mac-ubuntu

Step 2: Install RStudio

- https://rstudio.com/products/rstudio/download/
- Scroll down the page – choose the free installer
- Install the one for your device (Windows)
- Follow the install process
- Accept start name of rstudio

RStudio Cloud Version

- https://rstudio.cloud/
- This saves you installing any further software and saves PC space
- Set-up your free account



Your Projects

Sort: By name

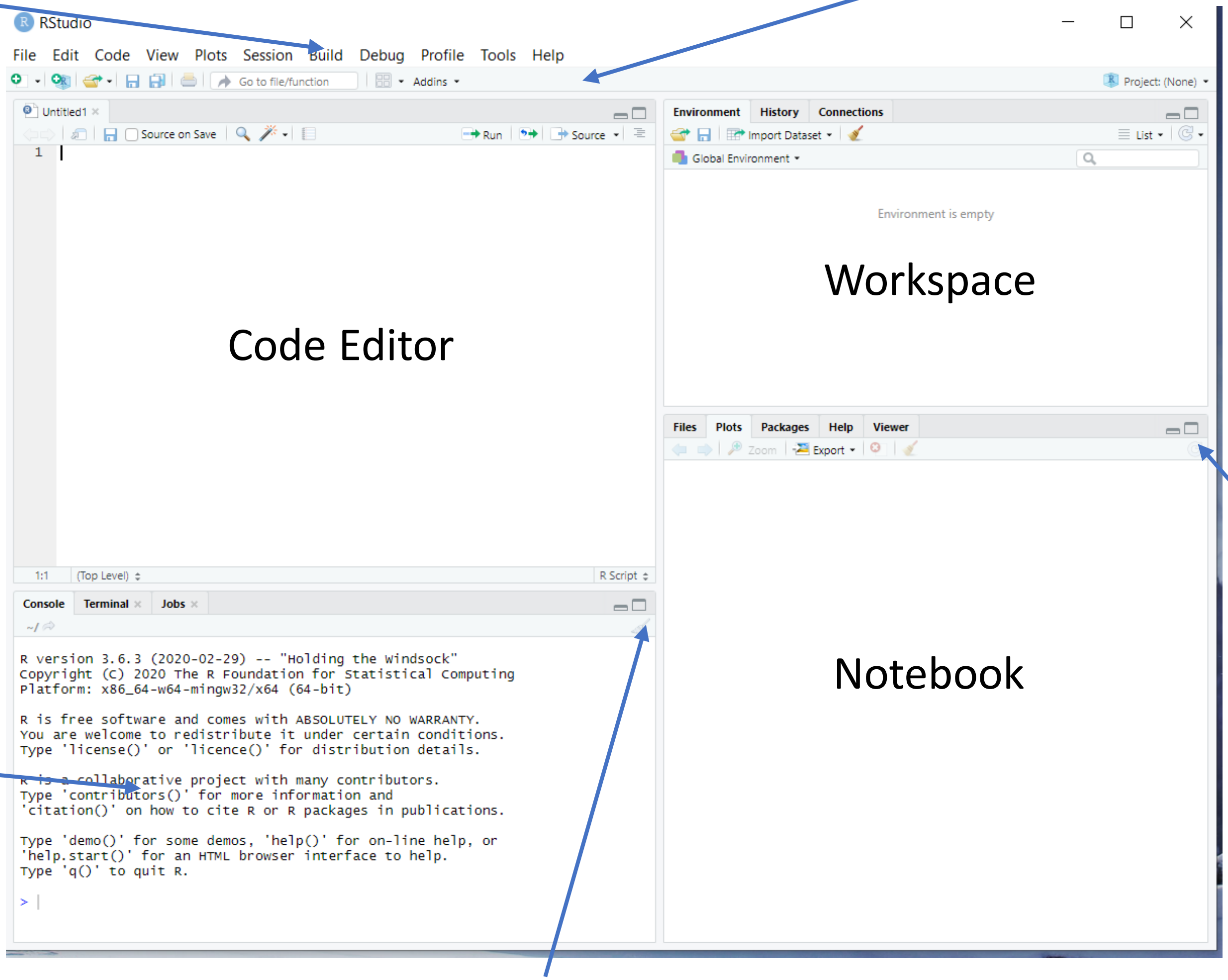
Untitled Project
Created Jan 13, 2021 3:03 PM

TECH TALENT
ACADEMY

Exploring RStudio

Menu Bar

Toolbar



```
R version 3.6.3 (2020-02-29) -- "Holding the Windsock"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> 
```

Code Editor

Workspace

Notebook

Maximize/minimize

Console

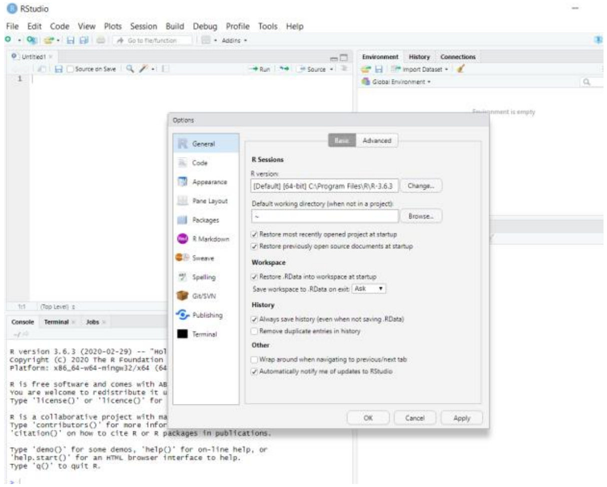
Drag Handle

WOMEN IN DATA
ACADEMY

TECH TALENT
ACADEMY

Settings

- Tools > Global options
- Change the default working directory to one of your choice



R Sessions

R version: [Default] [64-bit] C:\Program Files\R\R-3.6.3

Default working directory (when not in a project): ~

☑ Restore most recently opened project at startup

☑ Restore previously open source documents at startup

Workspace

☑ Restore .RData into workspace at startup

Save workspace to .RData on exit: Ask

History

☑ Always save history (even when not saving .RData)

□ Remove duplicate entries in history

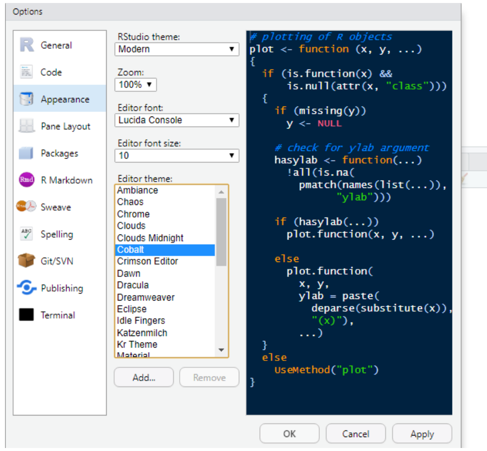
Other

□ Wrap around when navigating to previous/next tab

☑ Automatically notify me of updates to RStudio

```
R version 3.6.3 (2020-02-29) -- "Hol
Copyright (C) 2020 The R Foundation
Platform: x86_64-w64-mingw32/x64 (64

R is free software and comes with AB
You are welcome to redistribute it u
Type 'license()' or 'licence()' for

R is a collaborative project with ma
Type 'contributors()' for more infor
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>
```

- Change the appearance settings,
- cobalt is the black screen.



RStudio theme: Modern

Zoom: 100%

Editor font: Lucida Console

Editor font size: 10

Editor theme: Ambiance, Chaos, Chrome, Clouds, Clouds Midnight, Cobalt, Crimson Editor, Dawn, Dracula, Dreamweaver, Eclipse, Idle Fingers, Katzenmilch, Kr Theme, Material

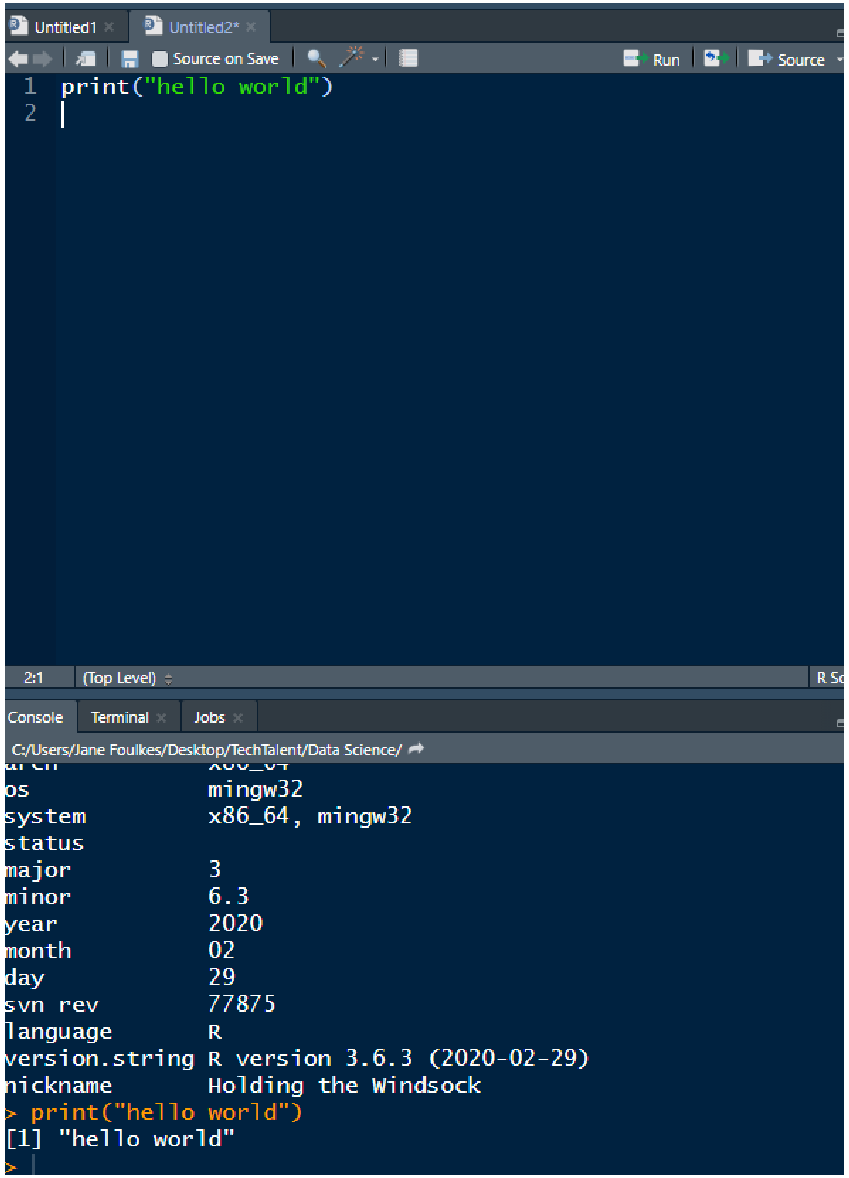
```
# plotting of R objects
plot <- function (x, y, ...)
{
  if (is.function(x) &&
      is.null(attr(x, "class")))
  {
    if (missing(y))
      y <- NULL

    # check for ylab argument
    hasylab <- function(...)
      !all(is.na(
        pmatch(names(list(...)),
               "ylab")))

    if (hasylab(...))
      plot.function(x, y, ...)

    else
      plot.function(
        x, y,
        ylab = paste(
          deparse(substitute(x)),
          "(x)"),
        ...)
  }
  else
    UseMethod("plot")
}
```

- Use ctrl + L or 🧹 on the console pane to clear
- Try typing in version you should now see your version in the console pane

WOMEN IN DATA
ACADEMY

Create First Script

- Choose File, new file
- Type in “print” in the code editor
- Add the (“hello world”)
- Click run or Ctrl+Enter
- Should see the code in the console
- Ctrl+S will save the file



```
print("hello world")

```

```
arch           x86_64
os             mingw32
system         x86_64, mingw32
status
major          3
minor          6.3
year           2020
month          02
day            29
svn rev        77875
language       R
version.string R version 3.6.3 (2020-02-29)
nickname       Holding the Windsock
> print("hello world")
[1] "hello world"
>
```

Session Content

Storing Values

Adding comments

Data types

Plotting stored values

Controlling objects

Storing Values (Variables)

The way to assign a value in R is using <-

shortcut = ALT -

```
name <- “username”
name <- “Satinder Sohal”
print (name)
```

Adding Comments

# is used to add comments

Readline and Paste Commands

Readline is a built-in function that accepts a string argument within its parentheses to output as a prompt. It waits for user to type in the input to assign to that variable.

```
name <- readline(“Enter your name: ”)
```

The **paste** function concatenates a string and an assigned variable to be output.

```
hello <- paste (“Welcome”, name, “!”)
```

Why do we now need to print? Notice the global variables??

Data Types

Data Type	Description	Example
Character	A Text character or string	“R” “R string”
Double	A Decimal Place	3.14
Integer	A Whole number	5
Boolean	A logical value	TRUE

Data Types

- Unlike other programming languages R does not require you to explicitly specify the data type when you create a variable.
- R automatically generates the data type based on the value that is typed in.
- To find out the data type you can use the typeof() function.

```
Title <- “Women in Data”
Type <- paste(“Type of title:”, typeof(title))
Print (type)
```

TRY THIS :

```
pi<- 3.1415
dozen <- 12L
print(paste(“Type of pi:”, typeof(pi)))
print(paste(“Type of dozen:”, typeof(dozen)))
flag <-T
print(paste( “is flag logical:”, is.logical(flag)))
```

WOMEN IN DATA
ACADEMY

Vectors: storing multiple values

- Variables are vectors in R, similar to an **array**.
- They have elements E.g. day = c(“Mon”, “Tues”, “Weds”).
- Not zero indexed. E.g. day[1] would retrieve Mon. They can be numbers too, but they must be the same data type.
- You use the built-in function c() to setup the vector, which is_assigned to a variable name
- To replace a value you would do day[3] = “Thurs” so Weds would become Thurs
- Vectors are flexible so expand easily, e.g. day [4] = “Fri” would automatically expand the vector
- length (day) would reveal all 4 elements

Store Mixed Data Types

- This is done using a **list** structure, so you can include **multiple** data types. *“You cannot do this in a vector! “*

- Lists are indexed to start at 1 like all other lists we have come across. Different values of data types are separated using commas.

For example, you can do this by

info <- list(21, 61.2, “Lisa”, TRUE)

If you want to add names to the elements

info <- list(age=21, weight=61.2, name=“Lisa”, female=TRUE)

Plotting Stored Value

Power of R allows you to provide graphic depictions of data stored within R script structures.

- Create two vectors
- *x<- c(1, 2, 3, 4, 5)*
- *Y <- c(21, 5, 7, 9, 3)*
- Add instruction to depict the vector values and then select all three lines. plot(x, y, type=“o”)
- Click Ctrl + Shift + S to run the script

Controlling Objects

When you create any R code to store data in any structure such as a variable or a list, a data structure object is created in the RStudio environment. You can use the built-in ls() function to call all objects/variables within the current console environment.

1. Create a list and two variables

```
airport.codesUK <-list ("Birmingham" = "BHX", "London Heathrow" = "LHR", "Bristol " = "BRS", “Manchester" = "MAN")

airport.dubai <-"DBX"

airport.sanfran <-"SFO"
```

1. Press CTRL + A to select the code then press CTRL + Enter to run the code and create the objects
2. Open the environment in list view and expand the list
3. Switch to grid view
4. Click the spyglass icon and expand the list
5. Return to the code editor and enter an instruction **ls()**
6. Add an instruction to remove some items **rm(airport.sanfran)**
7. Repeat the ls()
8. Select the three instructions then CTRL + Enter

Tasks

- ✓ Install R
- Install RStudio
- Set your settings
- Explore RStudio
- Add and print variables – add name variable (slide 12)
- Change to ask the user to input their name and print it - readline and paste commands (slide 14)
- Practical task (slide 16)
- Have a go at storing mixed data types, plotting stored values and controlling objects (slides 18-20)

Home Learning Tasks

1. Write a R program to take input from the user (name and age) and display the values.
2. Write a R program to get the details of the objects in memory. Hint: how do you list all the local variables?
3. Write a R program to create a sequence of numbers from 20 to 50 and find the mean of numbers from 20 to 60 and sum of numbers from 51 to 91.
4. Write a R program to create a vector which contains 10 random integer values between -50 and +50

WOMEN IN DATA ACADEMY

TECH TALENT
ACADEMY