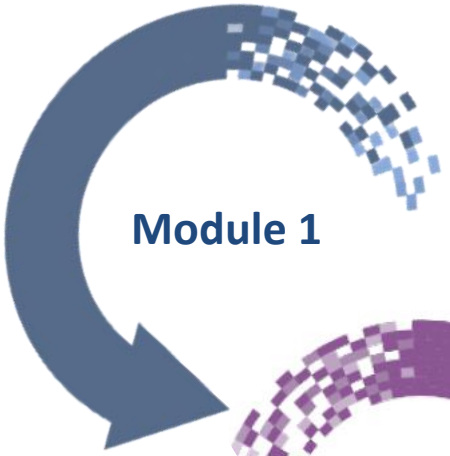
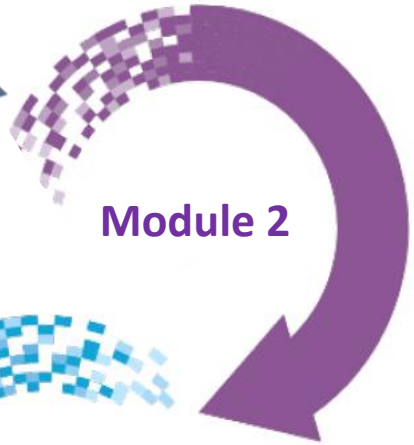
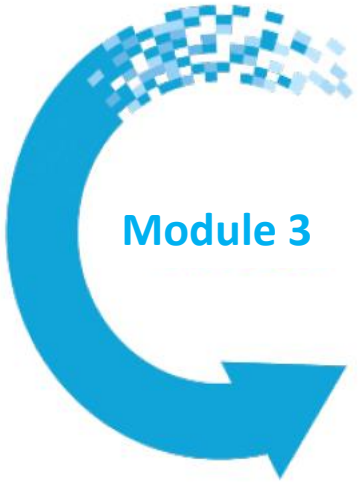
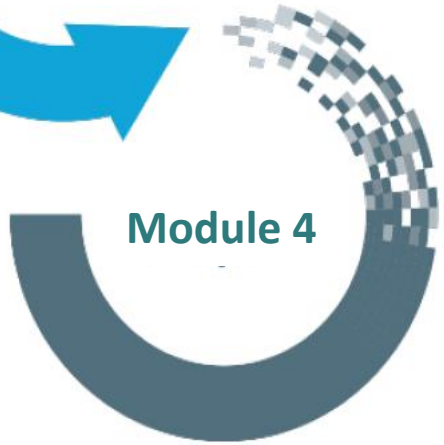








Exploratory Data Analysis

- Importing data
- Pre-processing phase
- Filtering phase

Network Analysis

- Building the correlation network
- Checking network integrity
- Finding communities in the network

Switch Mining

- Building the heat cartography map
- Identification of switch genes

Resilience Analysis

- Evaluation of network resilience

Module 0: Architecture

SWIMmeR is the R implementation of SWIM, a tool for the identification of a small pool of genes, called **switch genes**, which are likely to be critically associated with drastic changes in many biological settings.

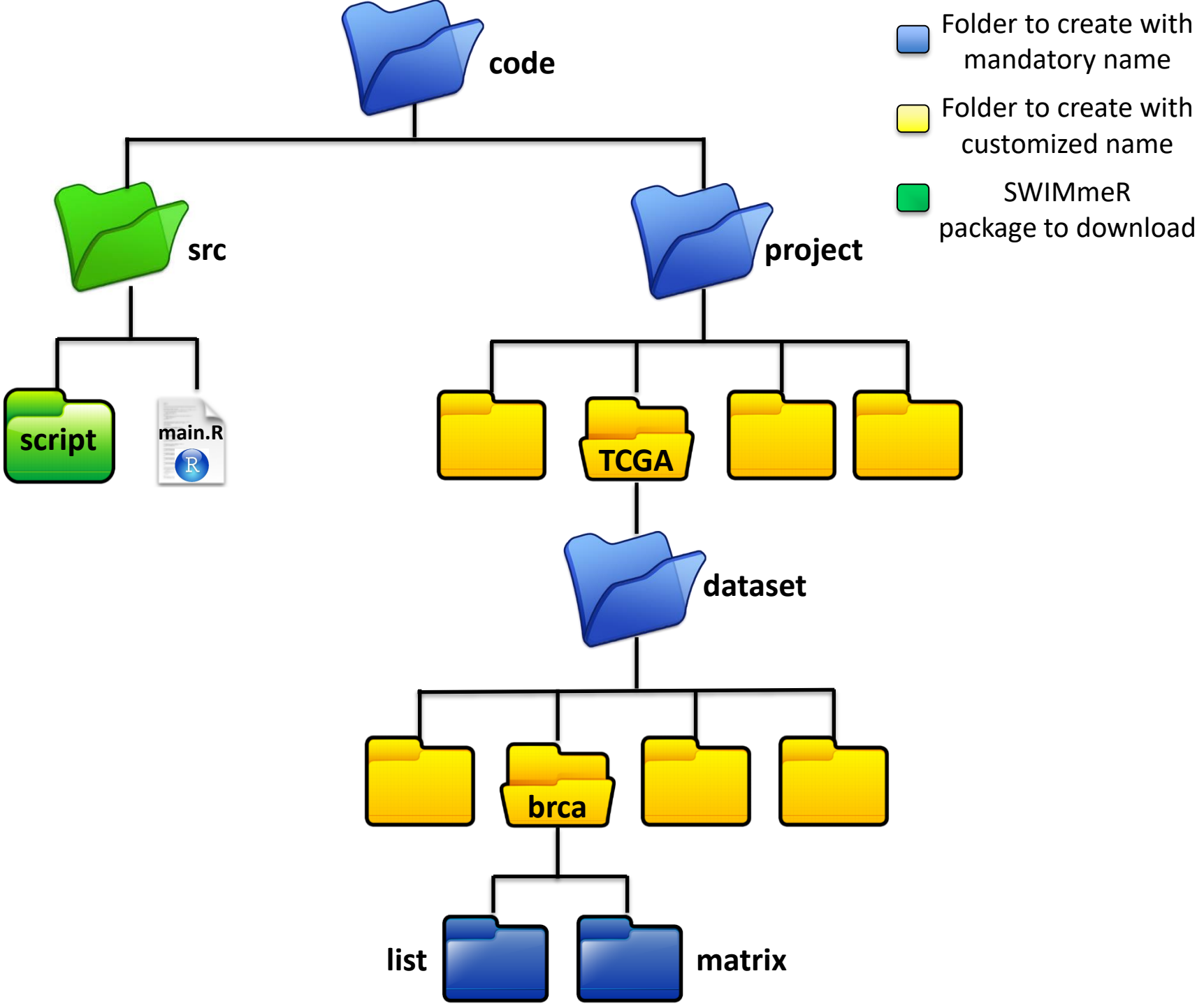
Scientific Reports 2017, **7**:44797

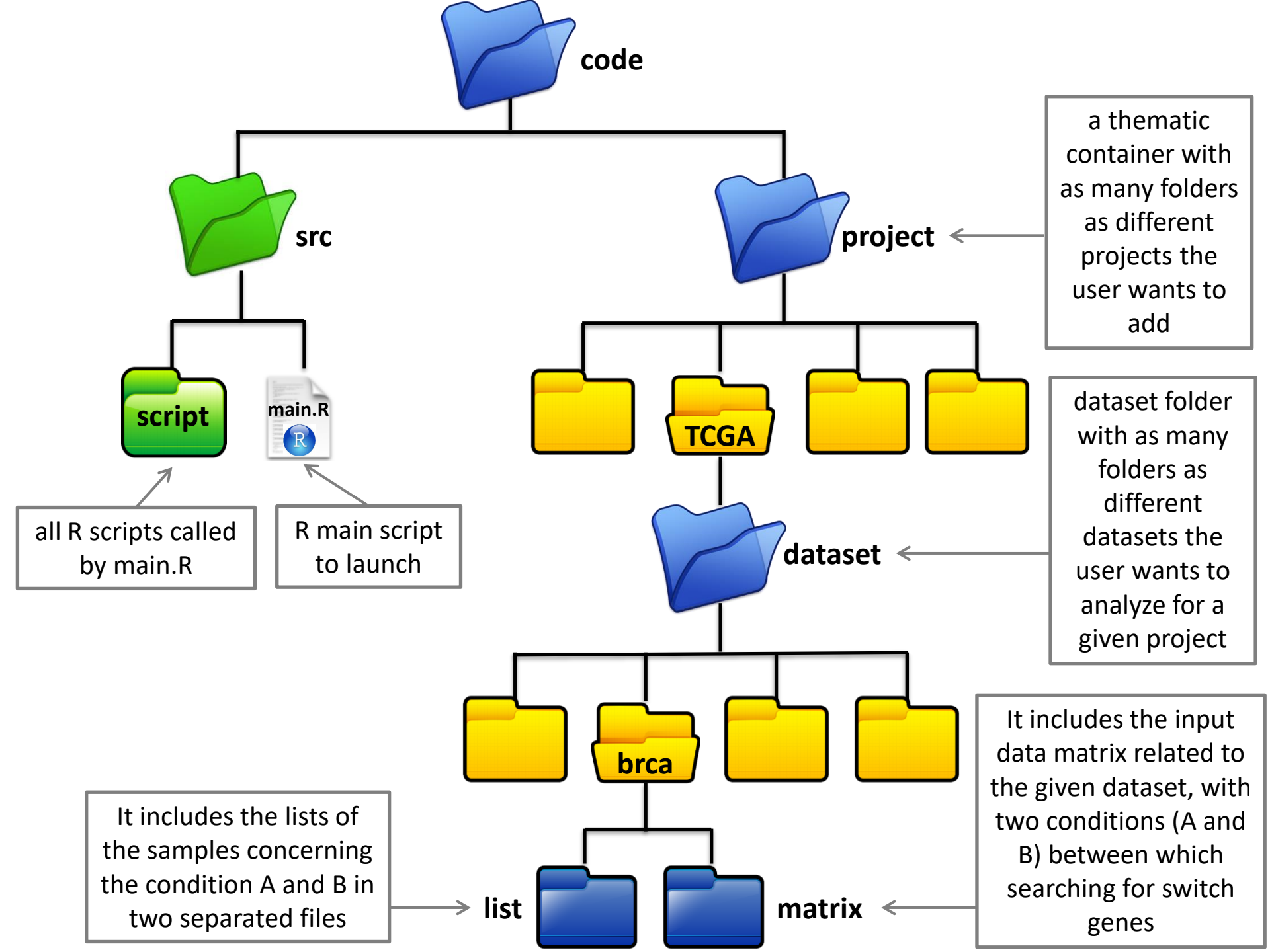
Software requirement

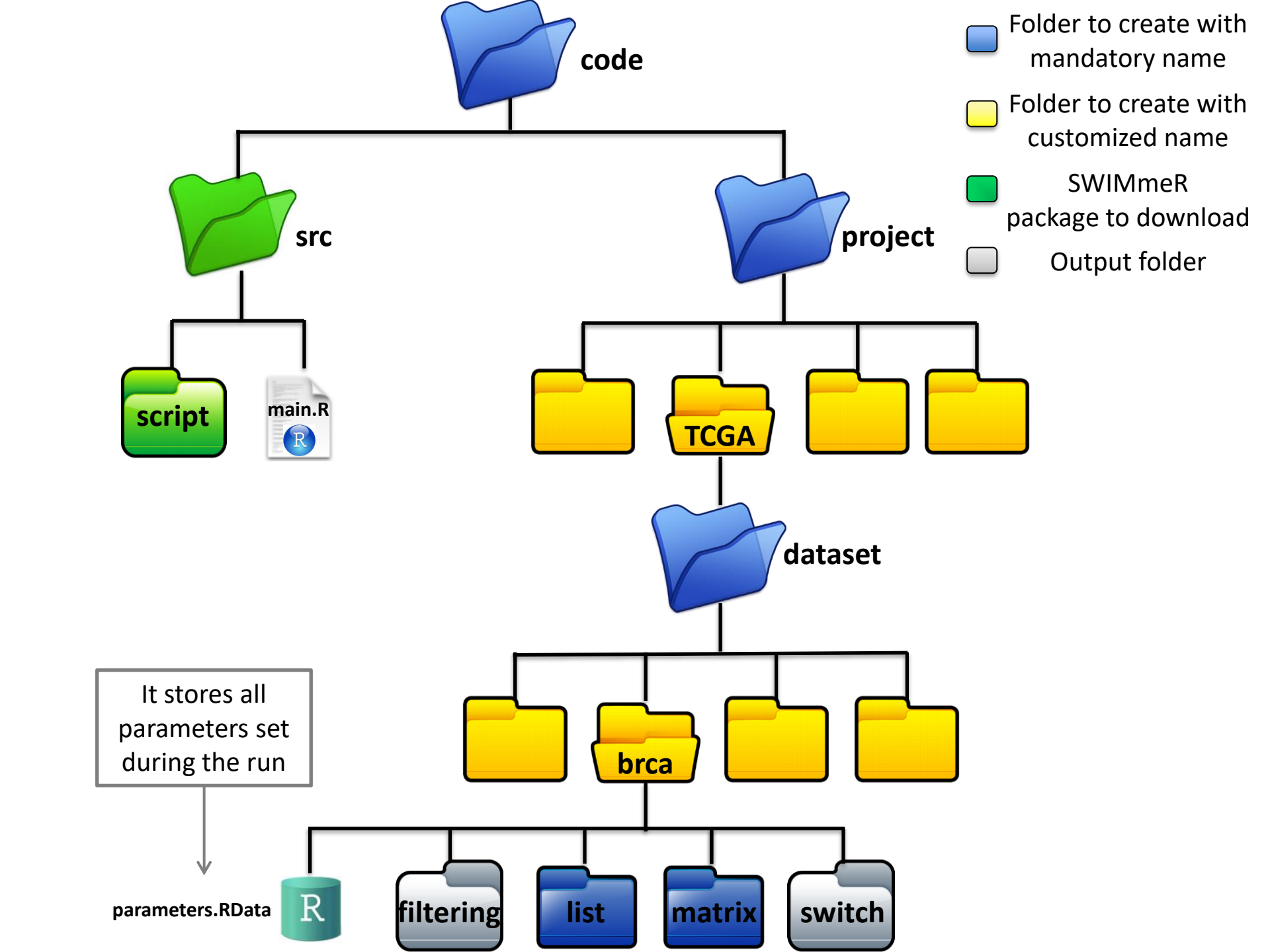
- SWIMmeR has been developed in R (version 3.6.1) and tested on the following operative systems:
  - macOS High Sierra 10.13.6
  - Windows 10 Pro

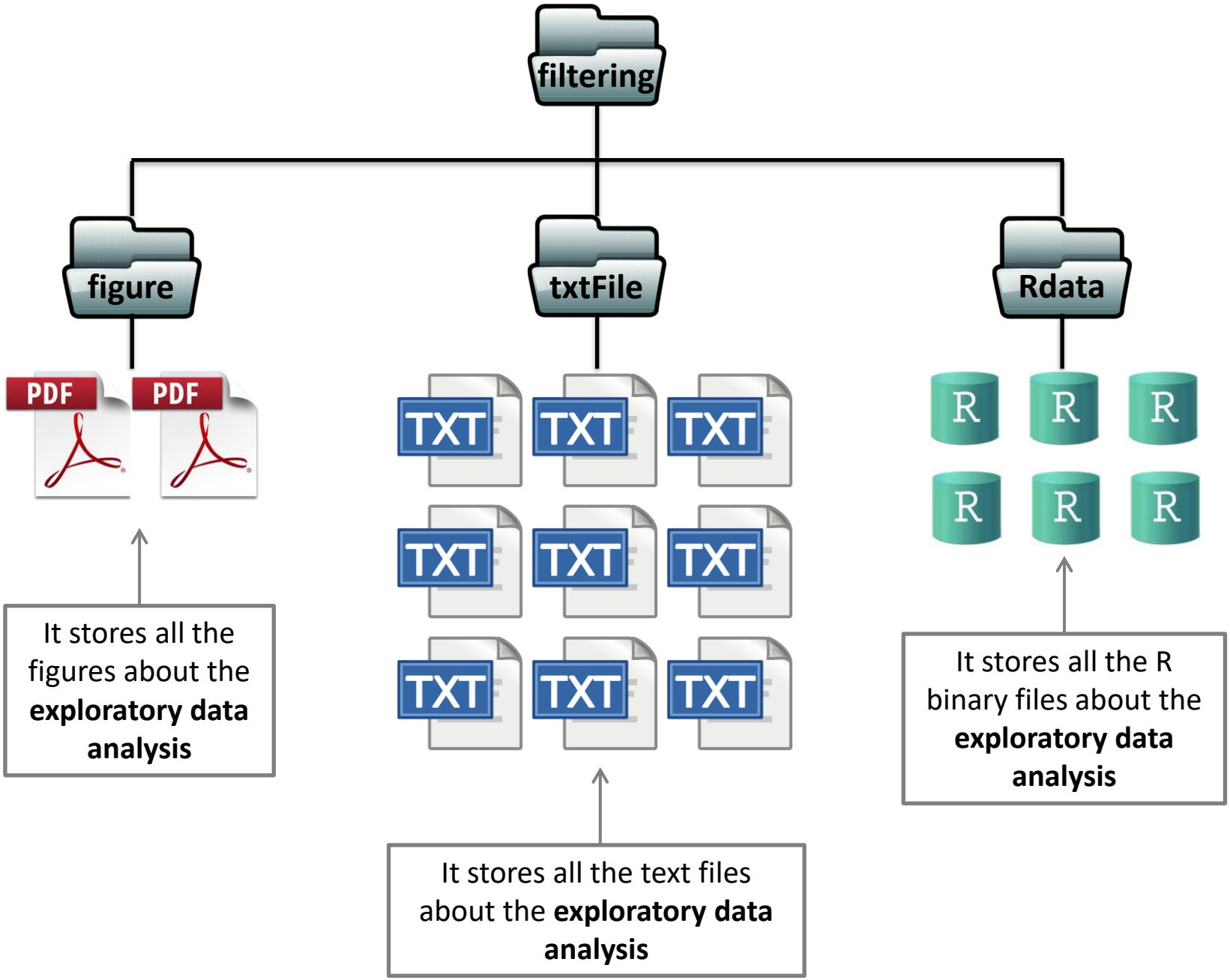
Setting up

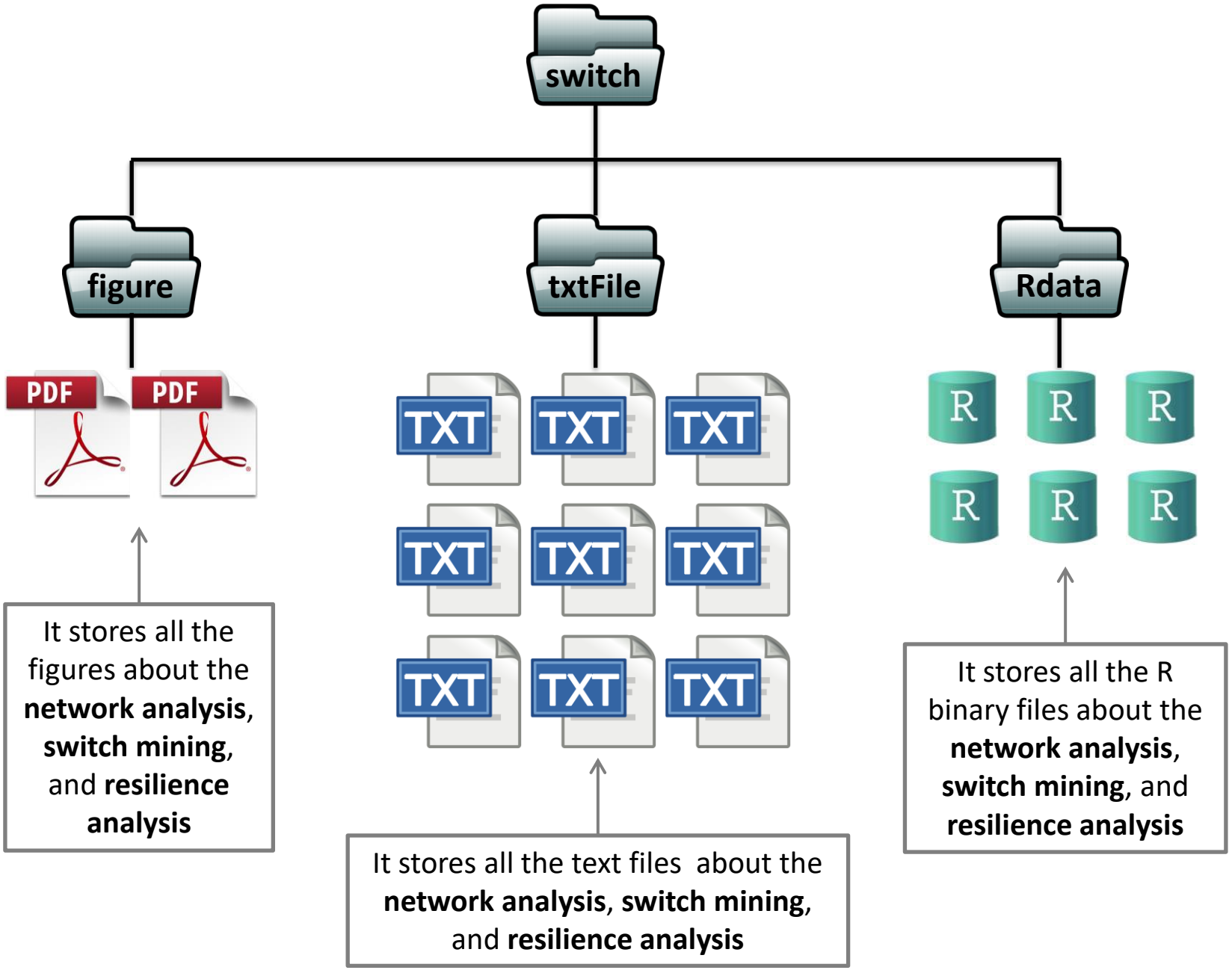
- Install R and R studio
- Download and unzip the SWIMmeR software package (“*src*” folder)


Folder to create with mandatory name
Folder to create with customized name
SWIMmeR package to download
code
src
project
script
main.R
TCGA
dataset
brca
list
matrix

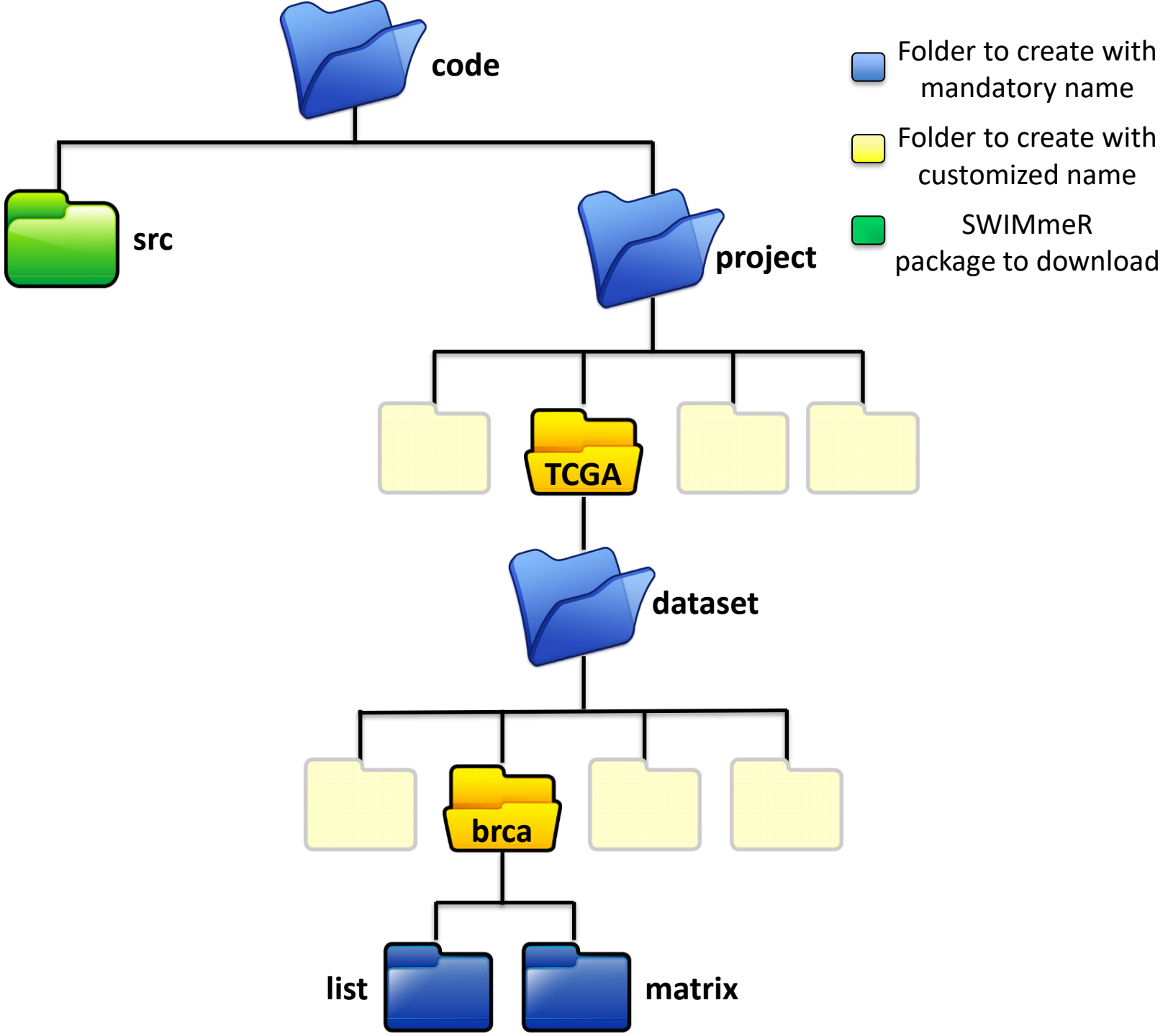

code
src
project
a thematic container with as many folders as different projects the user wants to add
script
main.R
TCGA
all R scripts called by main.R
R main script to launch
dataset
dataset folder with as many folders as different datasets the user wants to analyze for a given project
brca
It includes the input data matrix related to the given dataset, with two conditions (A and B) between which searching for switch genes
It includes the lists of the samples concerning the condition A and B in two separated files
list
matrix

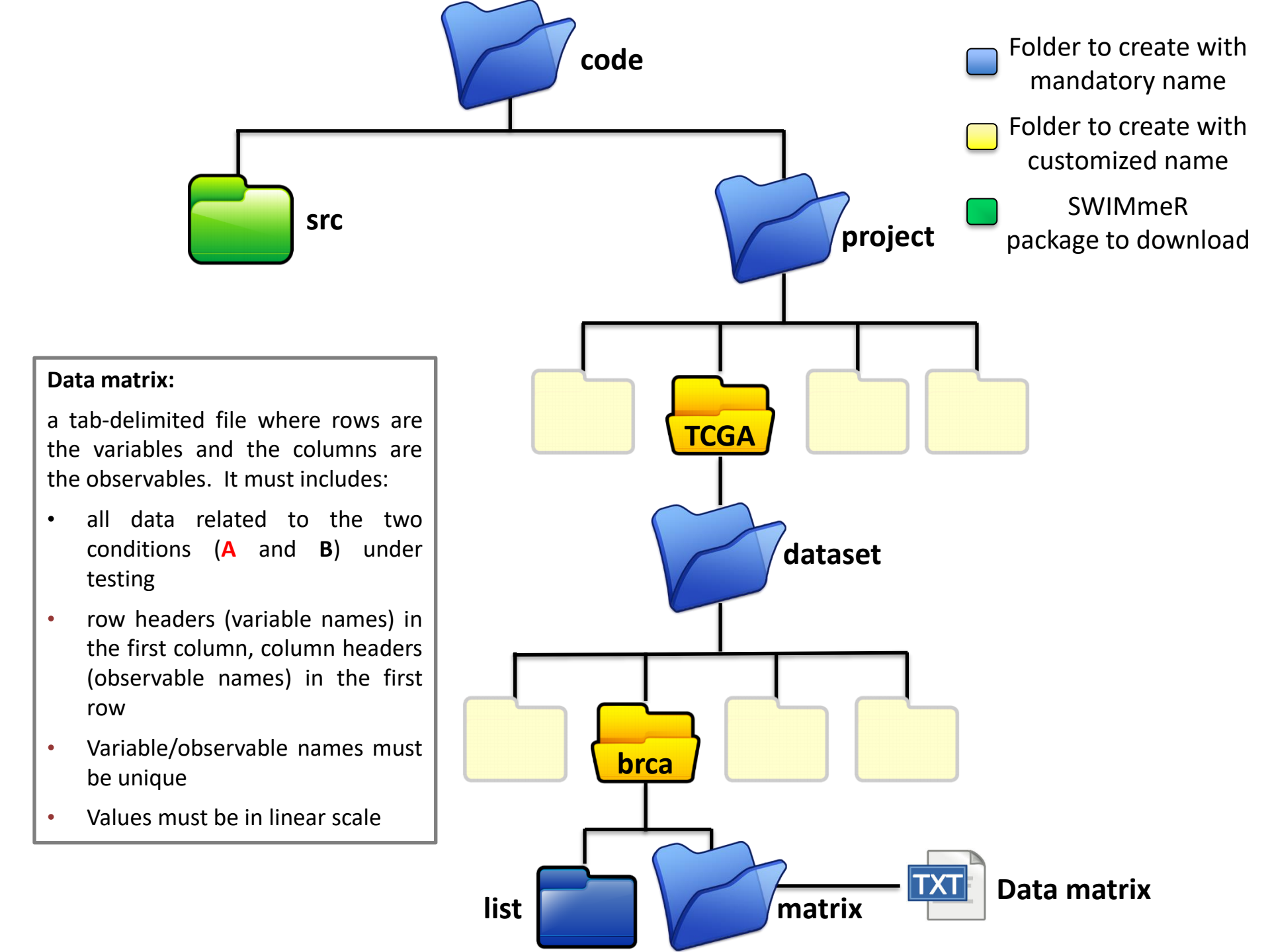

code
src
project
script
main.R
TCGA
dataset
brca
It stores all parameters set during the run
parameters.RData
filtering
list
matrix
switch
Folder to create with mandatory name
Folder to create with customized name
SWIMmeR package to download
Output folder


filtering
figure
txtFile
Rdata
PDF
PDF
TXT
R
It stores all the figures about the **exploratory data analysis**
It stores all the text files about the **exploratory data analysis**
It stores all the R binary files about the **exploratory data analysis**


switch
figure
txtFile
Rdata
PDF
PDF
TXT
R
It stores all the figures about the **network analysis**, **switch mining**, and **resilience analysis**
It stores all the text files about the **network analysis**, **switch mining**, and **resilience analysis**
It stores all the R binary files about the **network analysis**, **switch mining**, and **resilience analysis**

Let’s get started


code
src
project
TCGA
dataset
brca
list
matrix
Folder to create with mandatory name
Folder to create with customized name
SWIMmeR package to download



Data matrix:

a tab-delimited file where rows are the variables and the columns are the observables. It must includes:

- all data related to the two conditions (**A** and **B**) under testing
- row headers (variable names) in the first column, column headers (observable names) in the first row
- Variable/observable names must be unique
- Values must be in linear scale

Data matrix - example

Data matrix
N x M

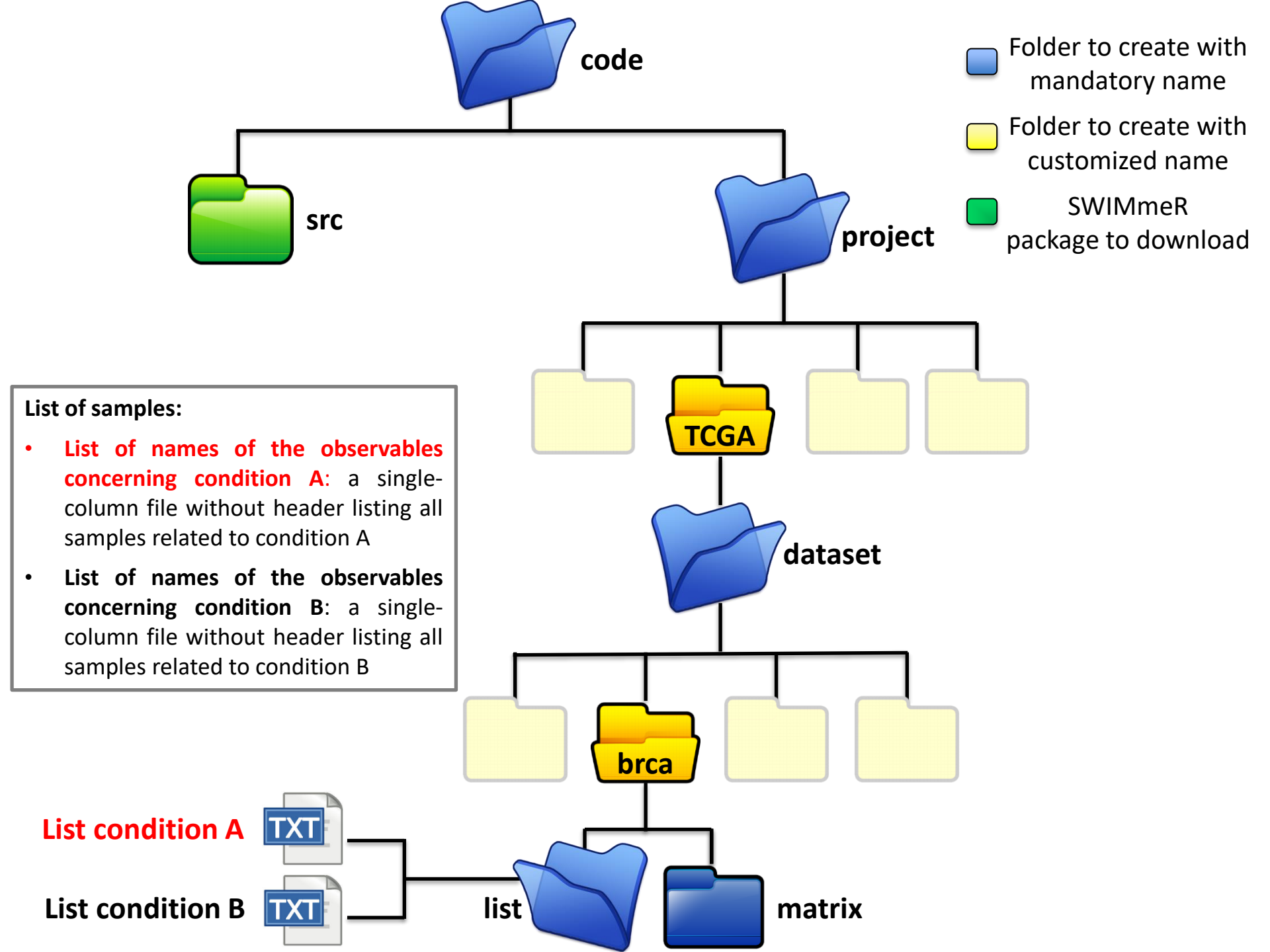
- N rows → Transcripts
- M columns → Samples
  - **Samples of Condition A – Tumor tissues**
  - **Samples of Condition B - Normal tissues**

Samples →

Transcripts ↓

	TCGA-A7-A0CE-11A-21R-A089-07	TCGA-A7-A0D9-11A-53R-A089-07	TCGA-A7-A0D9-01A-31R-A056-07	TCGA-A7-A0DB-01A-11R-A00Z-07	TCGA-A7-A13E-11A-61R-A12P-07	TCGA-A7-A13E-01A-11R-A12P-07	TCGA-A7-A13G-11A-51R-A13Q-07	TCGA-AC-A23H-11A-12R-A157-07	TCGA-AC-A2FB-01A-11R-A17B-07	TCGA-AC-A2FM-11B-32R-A19W-07	...	TCGA-A7-A0CE-01A-11R-A00Z-07	TCGA-A7-A0D9-11A-53R-A089-07	TCGA-A7-A0DB-11A-33R-A089-07	TCGA-A7-A0DC-01A-11R-A00Z-07	TCGA-A7-A13G-11A-51R-A13Q-07	TCGA-A7-A13F-01A-11R-A12P-07	TCGA-A7-A13G-01A-11R-A13Q-07	TCGA-AC-A23H-01A-11R-A157-07	TCGA-AC-A2FB-11A-13R-A17B-07	TCGA-AC-A2FM-01A-11R-A19W-07
AB1G	4.3	3.1	3.3	5.6	2.1	1.6	6.3	5.9	11.2	0.4	...	11.1	26.8	0.9	12.8	4.3	0.0	13.1	5.0	10.3	8.3
ADAM10	3.9	0.0	0.8	13.5	3.3	4.7	4.7	8.5	3.9	1.2	....	3.5	13.5	5.2	11.6	4.3	1.2	6.2	4.0	7.5	8.8
...											...										

⚠ **Caveat:** Transcript and samples names **must be** unique.
Matrix values **must be** linear. Missing values are treated as 0.


code
src
project
TCGA
dataset
brca
list
matrix
List condition A
List condition B
TXT
TXT
Folder to create with mandatory name
Folder to create with customized name
SWIMmeR package to download
List of samples:
• List of names of the observables concerning condition A: a single-column file without header listing all samples related to condition A
• List of names of the observables concerning condition B: a single-column file without header listing all samples related to condition B

Lists of samples - example

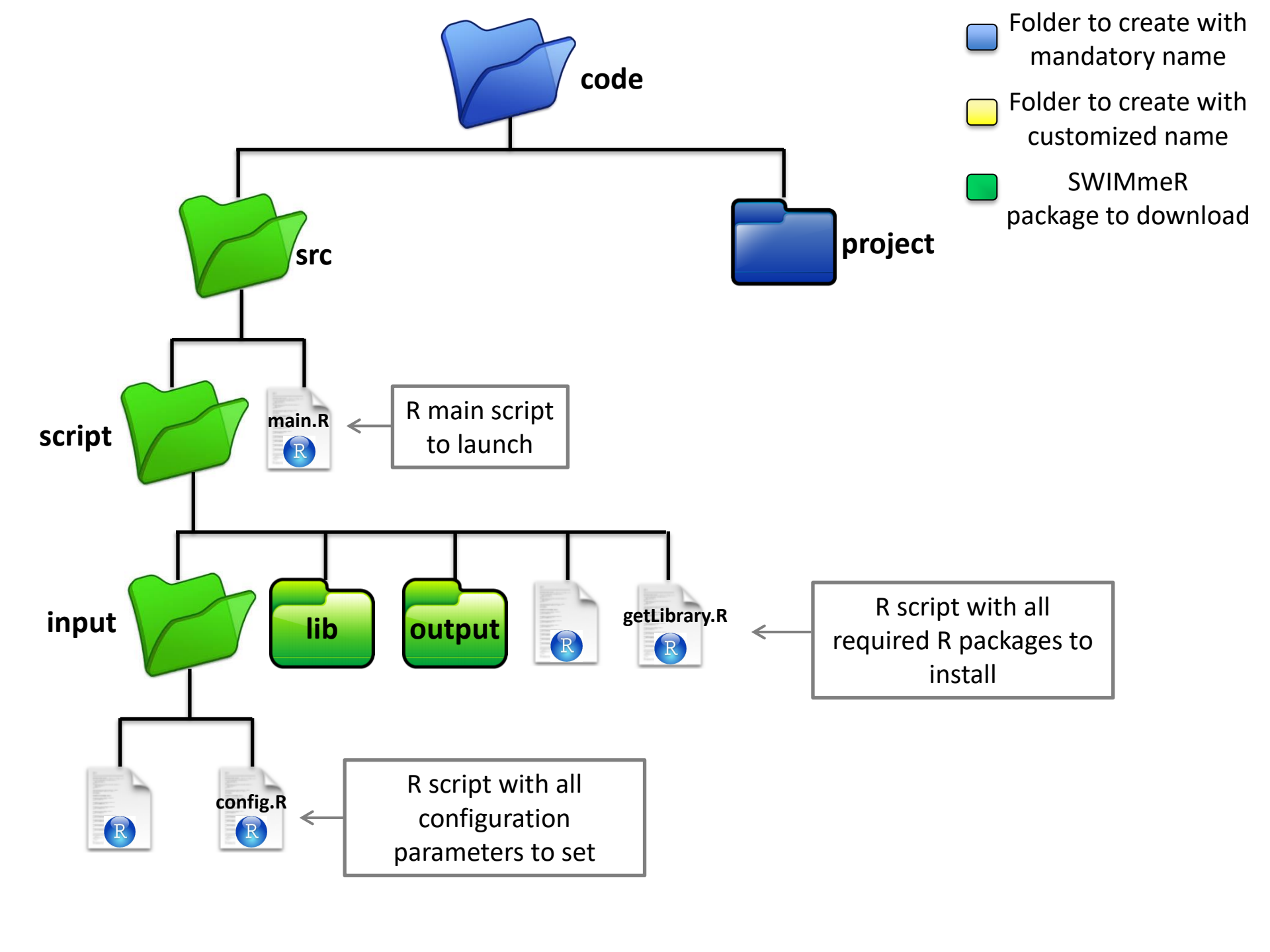
List of samples of condition A - case

TCGA-A7-A0CE-01A-11R-A00Z-07
TCGA-A7-A0D9-01A-31R-A056-07
TCGA-A7-A0DB-01A-11R-A00Z-07
TCGA-A7-A0DC-01A-11R-A00Z-07
TCGA-A7-A13E-01A-11R-A12P-07
TCGA-A7-A13F-01A-11R-A12P-07
TCGA-A7-A13G-01A-11R-A13Q-07
TCGA-AC-A23H-01A-11R-A157-07
.....
TCGA-AC-A2FB-01A-11R-A17B-07
TCGA-AC-A2FM-01A-11R-A19W-07

List of samples of condition B - control

TCGA-A7-A0CE-11A-21R-A089-07
TCGA-A7-A0D9-11A-53R-A089-07
TCGA-A7-A0DB-11A-33R-A089-07
TCGA-A7-A0DC-11A-41R-A089-07
TCGA-A7-A13E-11A-61R-A12P-07
TCGA-A7-A13F-11A-42R-A12P-07
TCGA-A7-A13G-11A-51R-A13Q-07
TCGA-AC-A23H-11A-12R-A157-07
.....
TCGA-AC-A2FB-11A-13R-A17B-07
TCGA-AC-A2FM-11B-32R-A19W-07

Ready to launch!


Folder to create with mandatory name
Folder to create with customized name
SWIMmeR package to download
code
src
project
script
main.R
R main script to launch
input
lib
output
getLibrary.R
R script with all required R packages to install
config.R
R script with all configuration parameters to set

Main file

Initial settings

main.R

```
rm(list=ls())

options(stringsAsFactors = F)

setwd("~/SWIMmeR/code")
####################################
source("src/script/getLibrary.R")
source("src/script/getSource.R")
####################################
getLibrary()
getSource()
input_parameter <- config()
input_file <- inputFiles()
output_file <- outputFiles()
####################################

data <- ExploratoryDataAnalysis()                                          Module 1

network <- NetworkAnalysis(data,checkNetIntegrity = T, screePlot = T)      Module 2

switch <- SwitchMining()                                                   Module 3

saveParameters()

if(input_parameter$removal_node == "yes") resilience <- ResilienceAnalysis()  Module 4

```

Environment is empty

Files

Home > SWIMmeR > code > src

Name	Size	Modified
..		
main.R	616 B	Oct 19, 2020, 5:58 PM
script		

Console

```
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.

>
```

Main file

Initial settings

main.R

```
rm(list=ls())

options(stringsAsFactors = F)

setwd("~/SWIMmeR/code")
####################################
source("src/script/getLibrary.R")
source("src/script/getSource.R")
####################################
getLibrary()
getSource()
input_parameter <- config()
input_file <- inputFiles()
output_file <- outputFiles()
####################################

data <- ExploratoryDataAnalysis()                                          Module 1

network <- NetworkAnalysis(data,checkNetIntegrity = T, screePlot = T)      Module 2

switch <- SwitchMining()                                                   Module 3

saveParameters()

if(input_parameter$removal_node == "yes") resilience <- ResilienceAnalysis()  Module 4
```

Environment is empty

Files

Home > SWIMmeR > code > src

Name	Size	Modified
..		
main.R	616 B	Oct 19, 2020, 5:58 PM
script		

Console

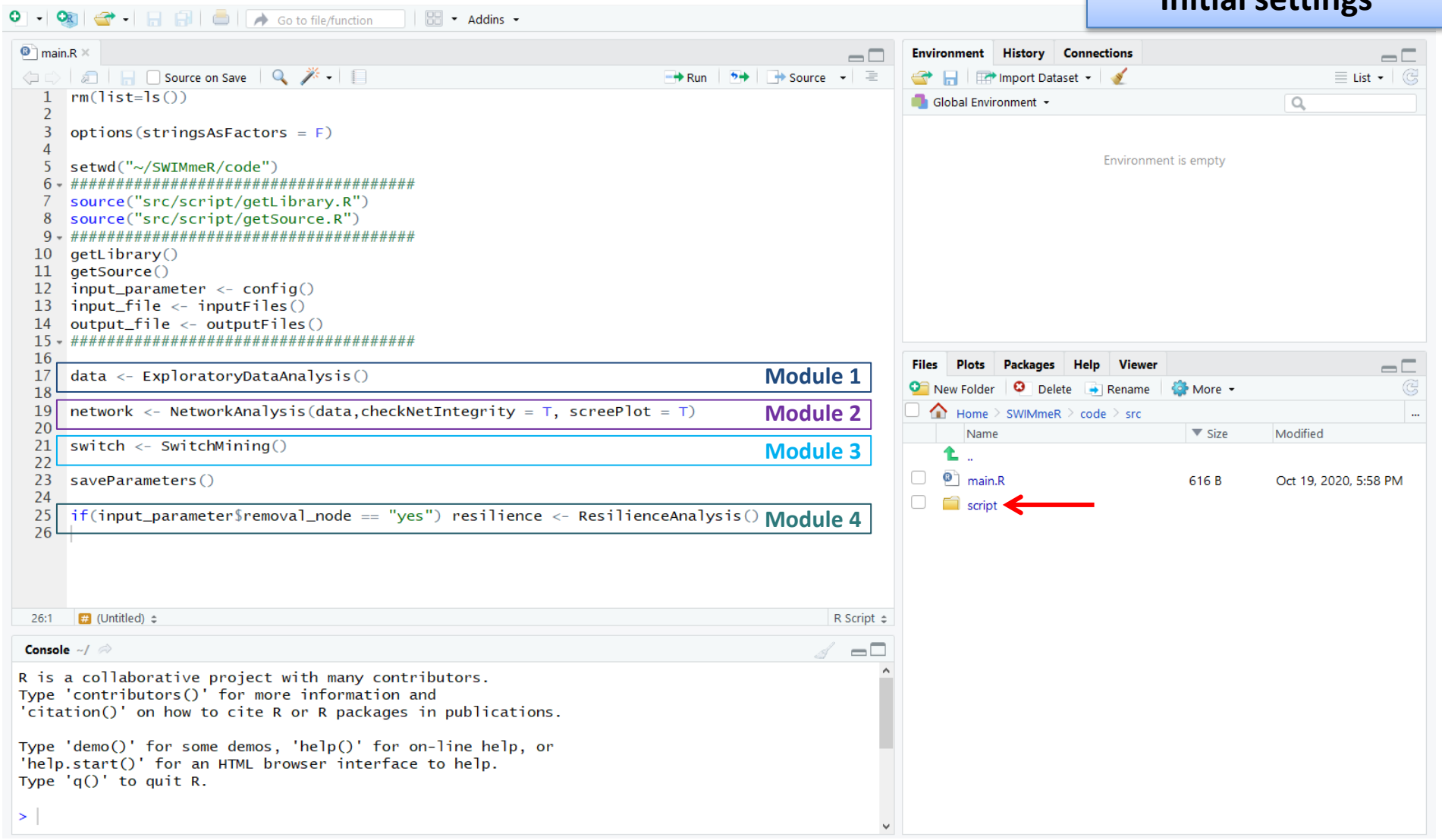
```
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

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.

>
```

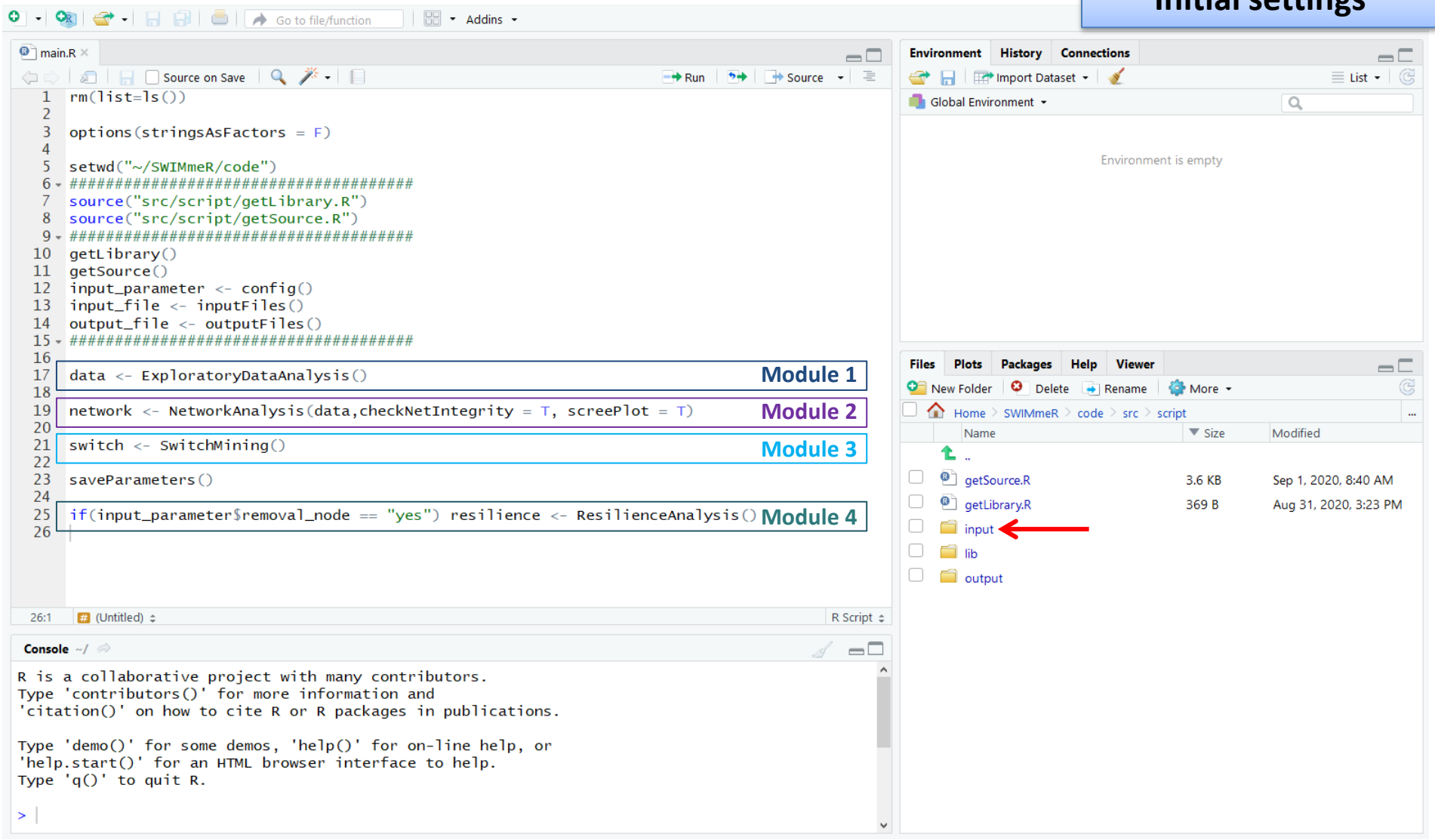
Main file

Initial settings



Main file

Initial settings



Main file

Initial settings

main.R

```
1  rm(list=ls())
2
3  options(stringsAsFactors = F)
4
5  setwd("~/SWIMmeR/code")
6  ###################################
7  source("src/script/getLibrary.R")
8  source("src/script/getSource.R")
9  ###################################
10 getLibrary()
11 getSource()
12 input_parameter <- config()
13 input_file <- inputFiles()
14 output_file <- outputFiles()
15 ###################################
16
17 data <- ExploratoryDataAnalysis()                                    Module 1
18
19 network <- NetworkAnalysis(data,checkNetIntegrity = T, screePlot = T)  Module 2
20
21 switch <- SwitchMining()                                             Module 3
22
23 saveParameters()
24
25 if(input_parameter$removal_node == "yes") resilience <- ResilienceAnalysis()  Module 4
26
```

26:1 (Untitled) R Script

Console ~/

```
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.

>
```

Environment | History | Connections

Import Dataset | List

Global Environment

Environment is empty

Files | Plots | Packages | Help | Viewer

New Folder | Delete | Rename | More

Home > SWIMmeR > code > src > script > input

Name	Size	Modified
..		
config.R	3.7 KB	Sep 29, 2020, 6:51 PM
inputFiles.R	2 KB	Sep 21, 2020, 4:08 PM

Configuration file

Initial settings

main.R | config.R

```
config <- function(){

  # for executing SWIMmer launch:
  # source('~/Documents/SWIMmeR/code/src/main.R')

  ################################
  project <- "TCGA"
  dataset <- "brca"
  miRNA <- "yes"

  path <- paste0("project/",project,"/dataset/",dataset)
  ################################
  # input files

  filename_data <- paste0(path,"/matrix/matrice__brca_RNASeq.txt")
  filename_CTRL <- paste0(path,"/list/Lista__RNASeq_Normal__brca__4wayData.txt")
  filename_CASE <- paste0(path,"/list/Lista__RNASeq_Tumor__brca__4wayData.txt")

  if(miRNA == "yes"){
    filename_data_miRNA <- paste0(path,"/matrix/matrice__brca_miRNASeq.txt")
    filename_CTRL_miRNA <- paste0(path,"/list/Lista__miRNASeq_Normal__brca__4wayData
.txt")
    filename_CASE_miRNA <- paste0(path,"/list/Lista__miRNASeq_Tumor__brca__4wayData.txt"
)
  }
  ################################
  # input parameters

  paired_ttest <- T            # a logical value indicating whether you want a paired t
-test
```

```
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.

>
```

Environment is empty

Name	Size	Modified
..		
config.R	3.7 KB	Sep 29, 2020, 6:51 PM
inputFiles.R	2 KB	Sep 21, 2020, 4:08 PM

Configuration file

Initial settings

main.R | config.R

```
config <- function(){

  # for executing SWIMmer launch:
  # source('~/Documents/SWIMmeR/code/src/main.R')

  ################################
  project <- "TCGA"
  dataset <- "brca"
  miRNA <- "yes"

  path <- paste0("project/",project,"/datase
  ################################
  # input files

  filename_data <- paste0(path,"/matrix/matr
  filename_CTRL <- paste0(path,"/list/Lista_
  filename_CASE <- paste0(path,"/list/Lista_

  if(miRNA == "yes"){
    filename_data_miRNA <- paste0(path,"/matrix/matrice__brca_miRNASeq.txt")
    filename_CTRL_miRNA <- paste0(path,"/list/Lista__miRNASeq_Normal__brca__4wayData
.txt")
    filename_CASE_miRNA <- paste0(path,"/list/Lista__miRNASeq_Tumor__brca__4wayData.txt"
)
  }
  ################################
  # input parameters

  paired_ttest <- T            # a logical value indicating whether you want a paired t
-test
```

Insert *project* name

Insert *dataset* name

Type “yes” if miRNA data are also available, “no” otherwise

Environment | History | Connections

Import Dataset | List

Global Environment

Environment is empty

Files | Plots | Packages | Help | Viewer

New Folder | Delete | Rename | More

Home > SWIMmeR > code > src > script > input

Name	Size	Modified
..		
config.R	3.7 KB	Sep 29, 2020, 6:51 PM
inputFiles.R	2 KB	Sep 21, 2020, 4:08 PM

Console

```
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.

>
```

Configuration file

Input files

```
config <- function(){

  # for executing SWIMmer launch:
  # source('~/Documents/SWIMmeR/code/src/main.R')

  ################################
  project <- "TCGA"
  dataset <- "brca"
  miRNA <- "yes"

  path <- paste0("project/",project,"/dataset/",dataset)
  ################################
  # input files

  filename_data <- paste0(path,"/matrix/matrice__brca_RNASeq.txt")
  filename_CTRL <- paste0(path,"/list/Lista__RNASeq_Normal__brca__4wayData.txt")
  filename_CASE <- paste0(path,"/list/Lista__RNASeq_Tumor__brca__4wayData.txt")

  if(miRNA == "yes"){
    filename_data_miRNA <- paste0(path,"/matrix/matrice__brca_miRNASeq.txt")
    filename_CTRL_miRNA <- paste0(path,"/list/Lista__miRNASeq_Normal__brca__4wayData
.txt")
    filename_CASE_miRNA <- paste0(path,"/list/Lista__miRNASeq_Tumor__brca__4wayData.txt"
)
  }
  ################################
  # input parameters

  paired_ttest <- T            # a logical value indicating whether you want a paired t
-test
```

Insert filename for *data matrix*

Insert filename for *list of samples* of **control** condition

Insert filename for *list of samples* of **case** condition

```
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.

>
```

Configuration file

Input files

main.R | config.R

```
config <- function(){

  # for executing SWIMmer launch:
  # source('~/Documents/SWIMmeR/code/src/main.R')

  ################################
  project <- "TCGA"
  dataset <- "brca"
  miRNA <- "yes"

  path <- paste0("project/",project,"/dataset/",dataset)
  ################################
  # input files

  filename_data <- paste0(path,"/matrix/matrice__brca_RNASeq.txt")
  filename_CTRL <- paste0(path,"/list/Lista__RNASeq_Normal__brca__4wayData.txt")
  filename_CASE <- paste0(path,"/list/Lista__RNASeq_Tumor__brca__4wayData.txt")

  if(miRNA == "yes"){
    filename_data_miRNA <- paste0(path,"/matrix/matrice__brca_miRNASeq.txt")
    filename_CTRL_miRNA <- paste0(path,"/list/Lista__miRNASeq_Normal__brca__4wayData
.txt")
    filename_CASE_miRNA <- paste0(path,"/list/Lista__miRNASeq_Tumor__brca__4wayData.txt"
)
  }
  ################################
  # input parameters

  paired_ttest <- T            # a logical value indicating whether you want a paired t
-test
```

Environment | History | Connections

Import Dataset | List

Global Environment

Environment is empty

If **miRNA** variable is set to "yes":

Insert filename for **miRNA** *data matrix*

Insert filename for **miRNA** *list of samples* of **control** condition

Insert filename for **miRNA** *list of samples* of **case** condition

Console

```
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.

>
```

Configuration file

Input parameters: Module 1

Type **T** if you want a **paired** t-test, **F** otherwise

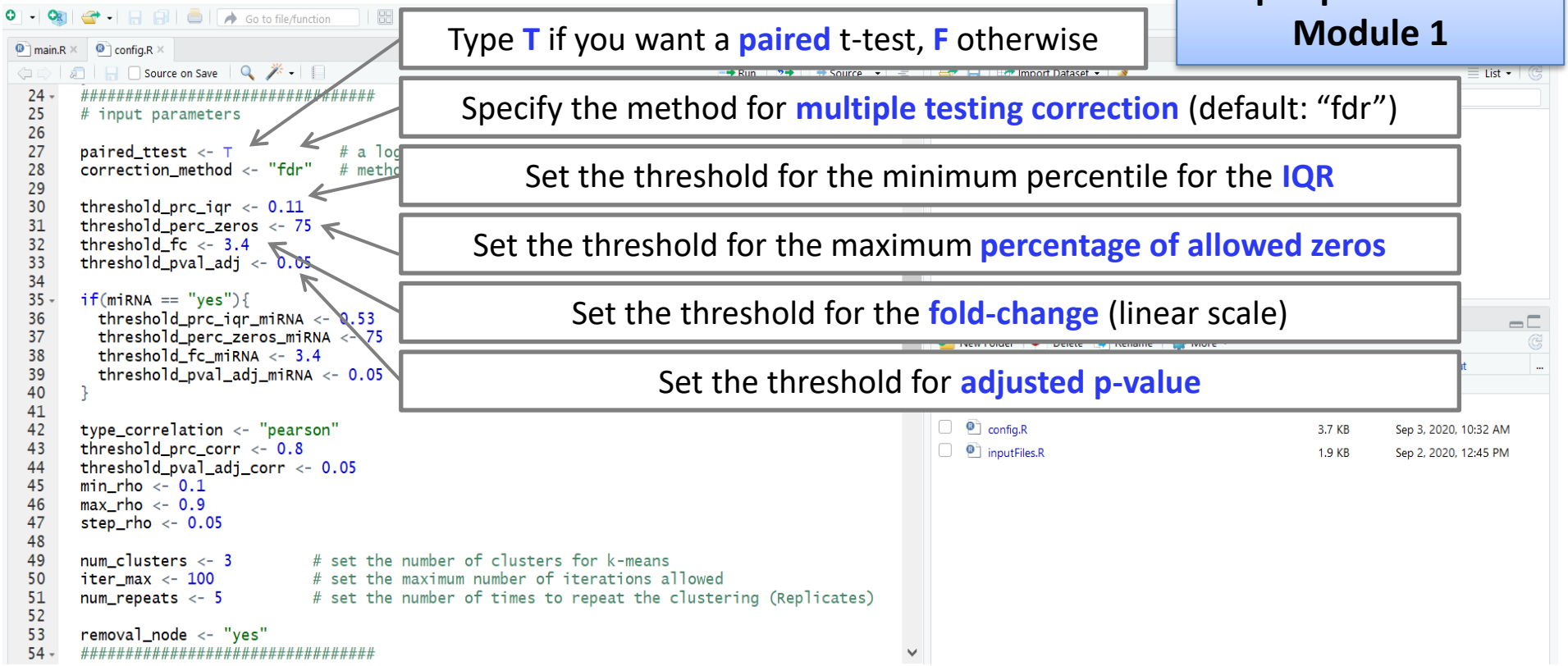
Specify the method for **multiple testing correction** (default: “fdr”)

Set the threshold for the minimum percentile for the **IQR**

Set the threshold for the maximum **percentage of allowed zeros**

Set the threshold for the **fold-change** (linear scale)

Set the threshold for **adjusted p-value**



Configuration file

Input parameters: Module 1

main.R | config.R

```
24  ##################################
25  # input parameters
26
27  paired_ttest <- T            # a logical value indicating whether y
28  correction_method <- "fdr"   # method for multiple testing correcti
29
30  threshold_prc_iqr <- 0.11
31  threshold_perc_zeros <- 75
32  threshold_fc <- 3.4
33  threshold_pval_adj <- 0.05
34
35  if(miRNA == "yes"){
36    threshold_prc_iqr_miRNA <- 0.53
37    threshold_perc_zeros_miRNA <- 75
38    threshold_fc_miRNA <- 3.4
39    threshold_pval_adj_miRNA <- 0.05
40  }
41
42  type_correlation <- "pearson"
43  threshold_prc_corr <- 0.8
44  threshold_pval_adj_corr <- 0.05
45  min_rho <- 0.1
46  max_rho <- 0.9
47  step_rho <- 0.05
48
49  num_clusters <- 3          # set the number of clusters for k-means
50  iter_max <- 100            # set the maximum number of iterations allowed
51  num_repeats <- 5           # set the number of times to repeat the clustering (Replicates)
52
53  removal_node <- "yes"
54  ##################################
```

If **miRNA** variable is set to “yes”:

Set the threshold for the minimum percentile for the **IQR** for **miRNA**

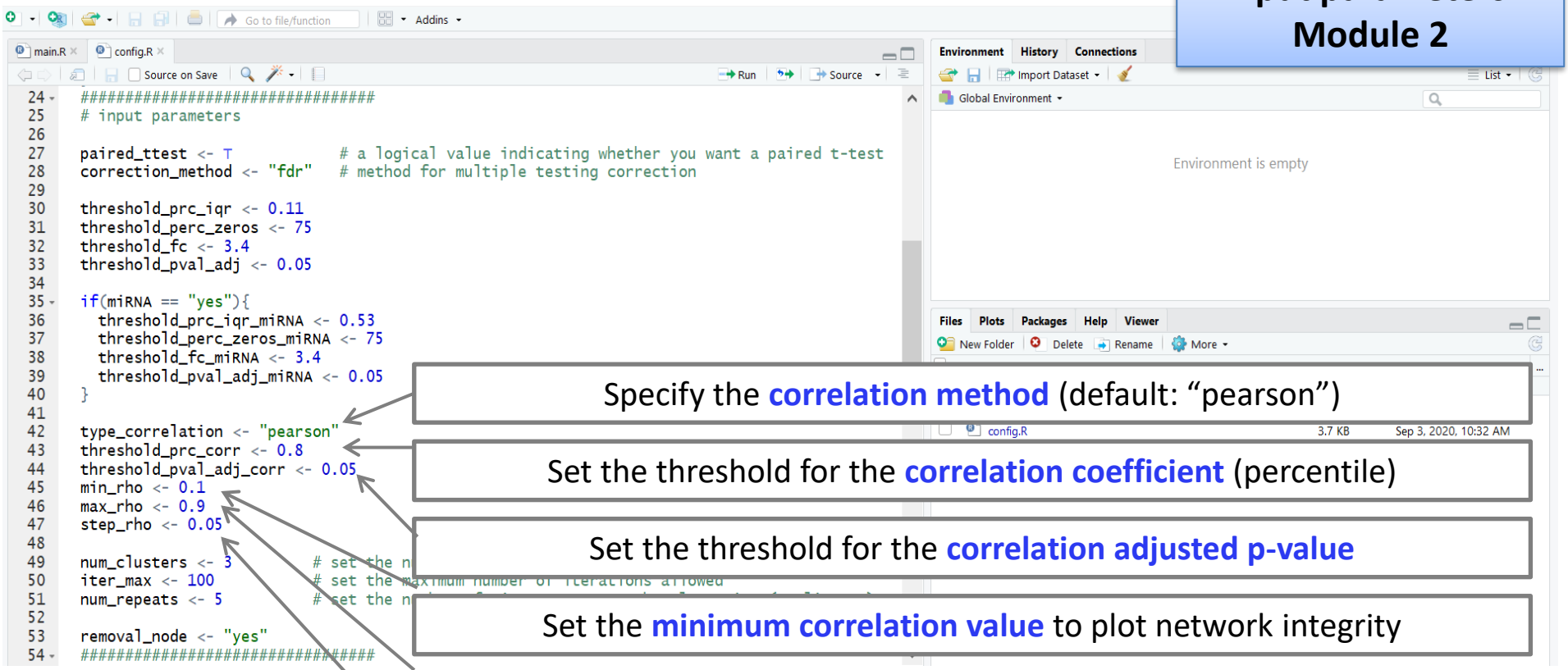
Set the threshold for the maximum **percentage of allowed zeros** for **miRNAs**

Set the threshold for the **fold-change** (linear scale) for **miRNAs**

Set the threshold for **adjusted p-value** for **miRNAs**

Configuration file

Input parameters: Module 2



Specify the **correlation method** (default: “pearson”)

Set the threshold for the **correlation coefficient** (percentile)

Set the threshold for the **correlation adjusted p-value**

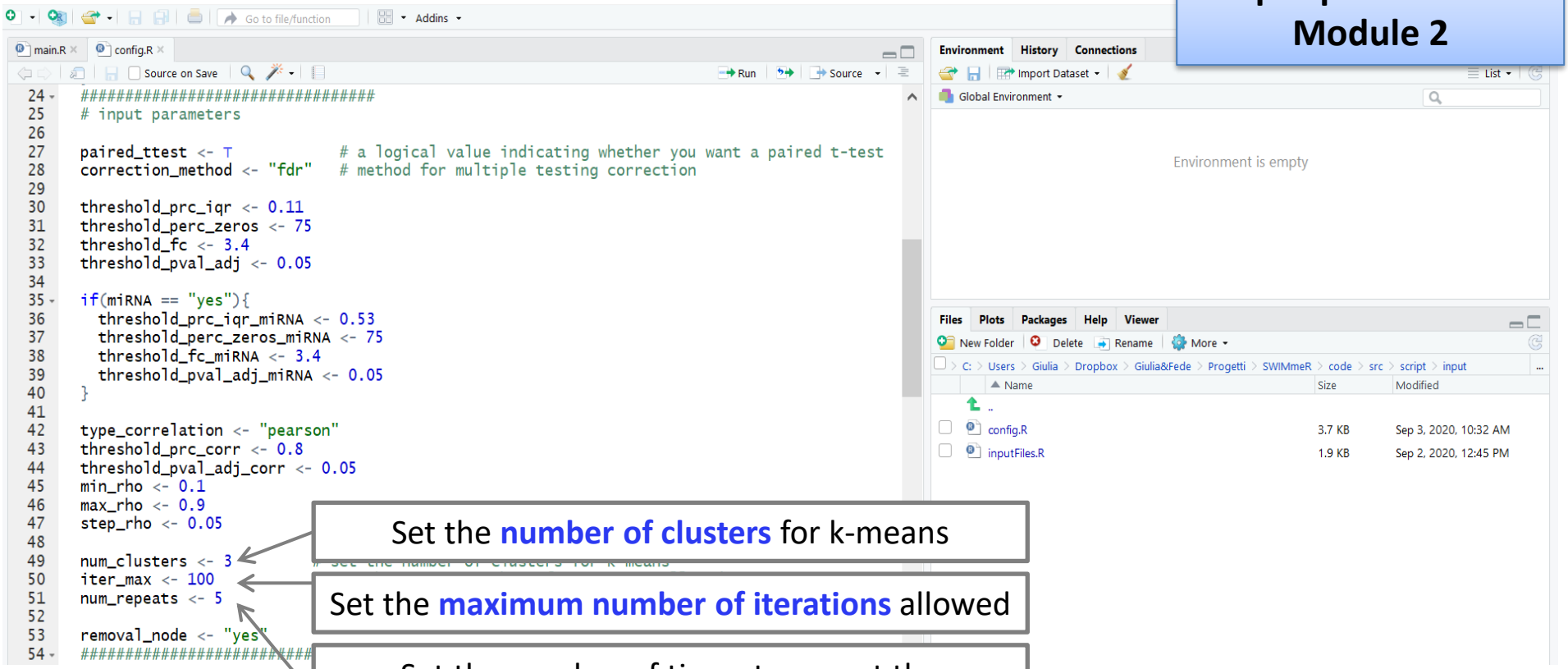
Set the **minimum correlation value** to plot network integrity

Set the **maximum correlation value** to plot network integrity

Set the **step** between min_rho and max_rho to plot network integrity

Configuration file

Input parameters: Module 2




```
################################
# input parameters

paired_ttest <- T            # a logical value indicating whether you want a paired t-test
correction_method <- "fdr"   # method for multiple testing correction

threshold_prc_iqr <- 0.11
threshold_perc_zeros <- 75
threshold_fc <- 3.4
threshold_pval_adj <- 0.05

if(miRNA == "yes"){
  threshold_prc_iqr_miRNA <- 0.53
  threshold_perc_zeros_miRNA <- 75
  threshold_fc_miRNA <- 3.4
  threshold_pval_adj_miRNA <- 0.05
}

type_correlation <- "pearson"
threshold_prc_corr <- 0.8
threshold_pval_adj_corr <- 0.05
min_rho <- 0.1
max_rho <- 0.9
step_rho <- 0.05

num_clusters <- 3
iter_max <- 100
num_repeats <- 5

removal_node <- "yes"
################################
```

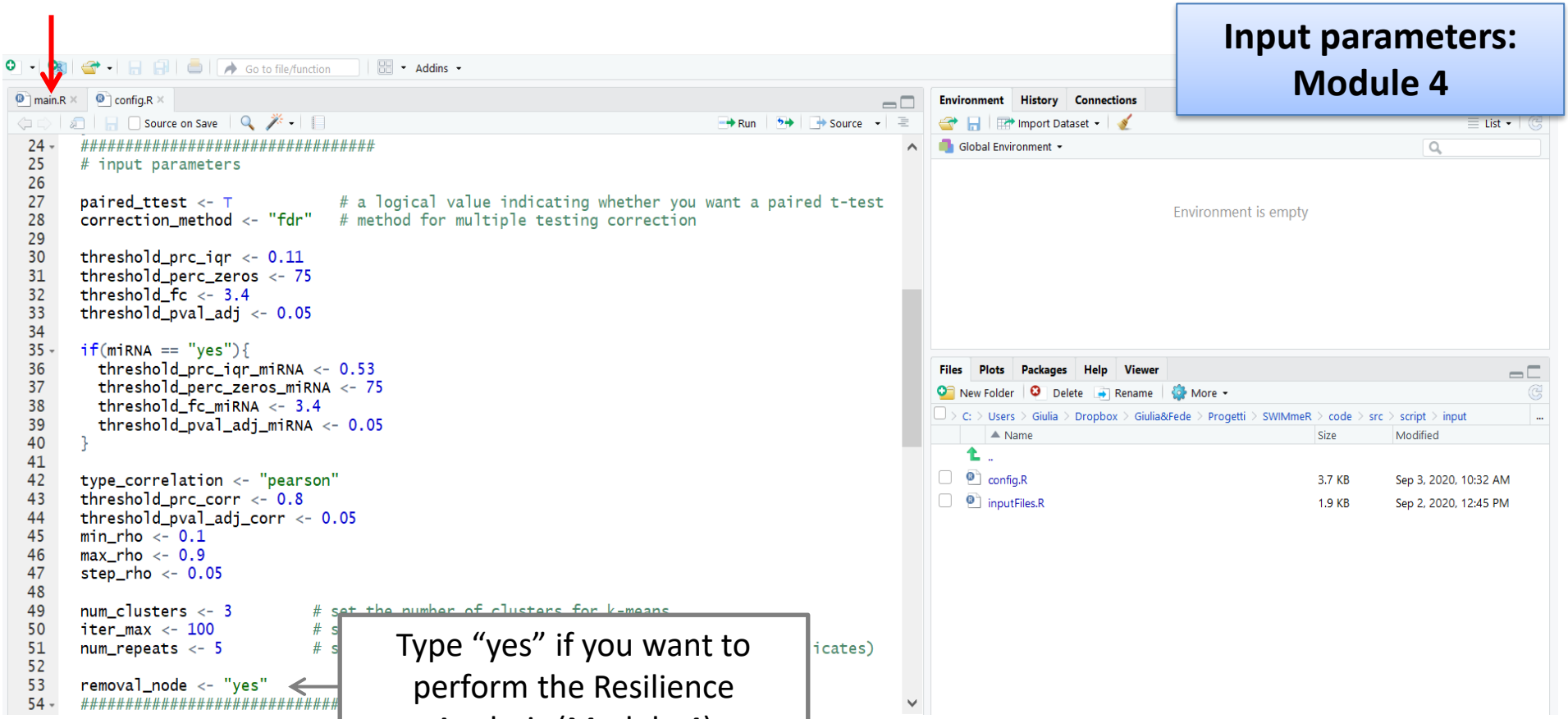
Set the **number of clusters** for k-means

Set the **maximum number of iterations** allowed

Set the number of times to repeat the clustering (**replicates**)

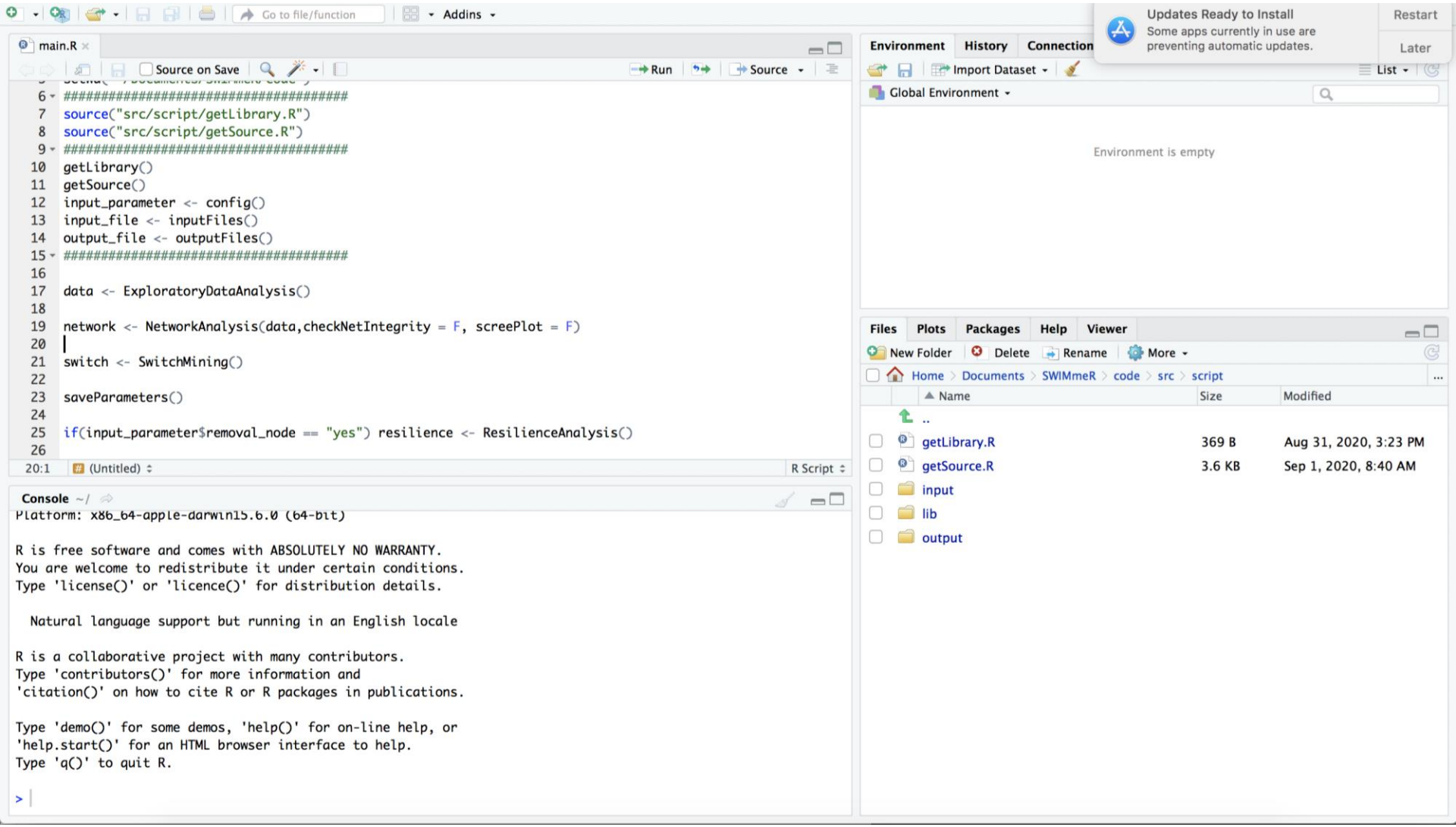
Configuration file

Input parameters: Module 4



Type “yes” if you want to perform the Resilience Analysis (Module 4)

SWIMmeR launch!



SWIMmeR launch!