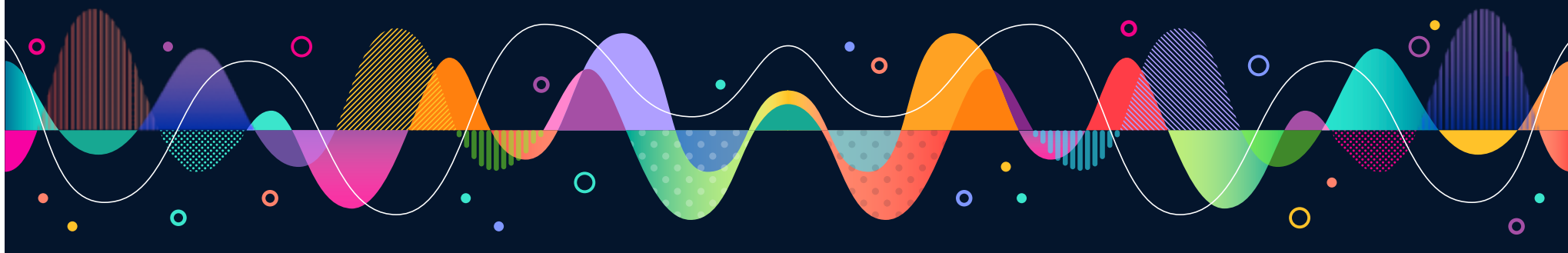


An Introduction to R and R Studio

Roy Roberts and Dr. Chris Bird



What is R?

```
royberts@LAPTOP-M413UJK4: ~
```

```
wahab-01:~> module load container_env mapdamage2
```

```
wahab-01:~> crun R
```

```
R version 4.0.1 (2020-06-06) -- "See Things Now"  
Copyright (C) 2020 The R Foundation for Statistical Computing  
Platform: x86_64-conda_cos6-linux-gnu (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.
```

```
Natural language support but running in an English locale
```

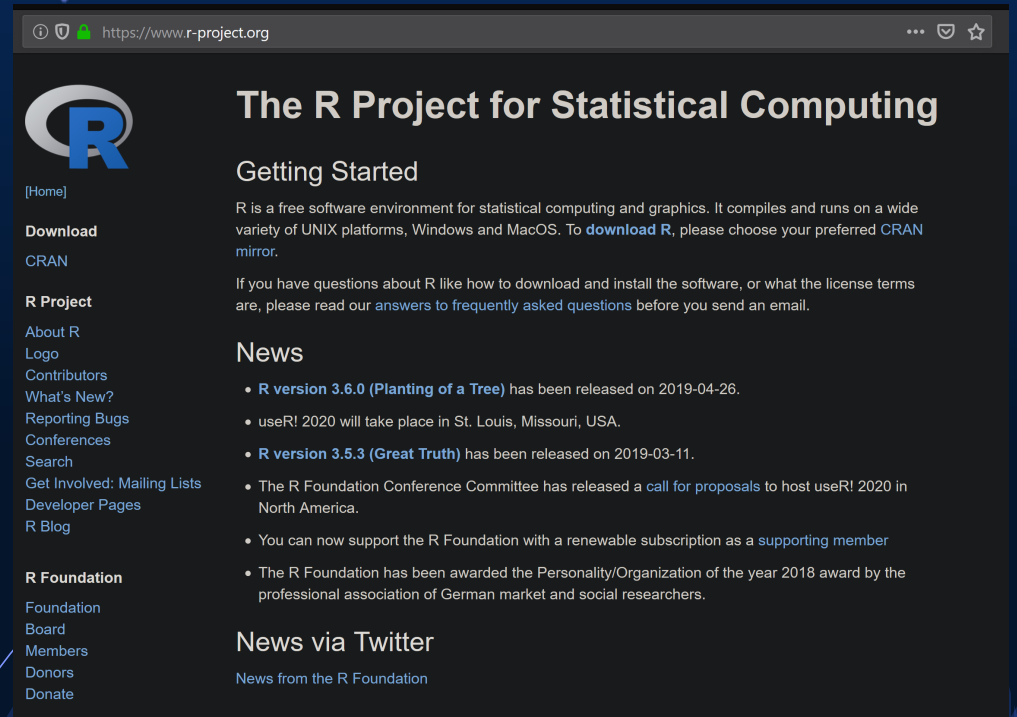
```
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.
```

```
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.
```

```
>  
>  
>  
>
```

Why use R ?

- ▷ Free
- ▷ Open source
- ▷ Popular among biologists
- ▷ Almost any analysis is available for free
- ▷ Easy to use
- ▷ Flexible
- ▷ Command line interface
- ▷ Linux, Win, Mac



R Studio

RStudio interface showing code, environment, and a plot.

```
1 rm(list = ls())
2
3
4 library(tidyverse)
5 library(lubridate)
6 library(vcfR)
7 library(adegenet)
8 library(adegraphics)
9 library(pegas)
10 library(seqinr)
11
12 setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
13
14 source("read_data_ssl.R")
15
16 metadata <- read_metadata_rrt("../metadata/")
17
18 #### VCFS ####
19
20 adu_mito_vcf_path <- "../summary_data_ssl/dDocentHPC_data/Adu_mito/vcf_filtering/adu.all.denovoSSL.Adu-
21 aur_mito_vcf_path <- "../summary_data_ssl/dDocentHPC_data/Aur_mito/vcf_filtering/aur.all.denovoSSL.Hte-
22 hte_mito_vcf_path <- "../summary_data_ssl/dDocentHPC_data/Hte_mito/vcf_filtering/hte.all.denovoSSL.Sob-
23 lle_mito_vcf_path <- "../summary_data_ssl/dDocentHPC_data/Lle_mito/vcf_filtering/lle.all.denovoSSL.Lle-
24 ssp_mito_vcf_path <- "../summary_data_ssl/dDocentHPC_data/Ssp_mito/vcf_filtering/ssp.all.denovoSSL.Sob-
25 tbi_mito_vcf_path <- "../summary_data_ssl/dDocentHPC_data/Tbi_mito/vcf_filtering/tbi.all.denovoSSL.Tbi-
26 consen_unfiltered_vcf_path <- "../summary_data_ssl/dDocentHPC_data/consensus_coi/TotalRawSNPs.denovoSSL
27
```

Environment:

Object	Value
dapc_consen_unf	Large dapc (19 elements, 1.2 MB)
grp_consen	List of 4
mat_consen	Large matrix (89320 elements, 796.9 kB)
metadata	143 obs. of 56 variables
vcf_consen	Large vcfR (2.7 MB)
vcfgenind_consen	Large genind (582.8 kB)

Values:

Object	Value
adu_mito_vcf_path	"../summary_data_ssl/dDocentHPC_data/Adu_mito/vcf_...
aur_mito_vcf_path	"../summary_data_ssl/dDocentHPC_data/Aur_mito/vcf_...
consen_unfiltered...	"../summary_data_ssl/dDocentHPC_data/consensus_co...
consen_vcf_path	"../summary_data_ssl/dDocentHPC_data/consensus_co...
grpval_pops_consen	Factor w/ 11 levels "Adu-A","Adu-C",...: 1 1 1 1 1...
hte_mito_vcf_path	"../summary_data_ssl/dDocentHPC_data/Hte_mito/vcf_...
lle_mito_vcf_path	"../summary_data_ssl/dDocentHPC_data/Lle_mito/vcf_...

Console:

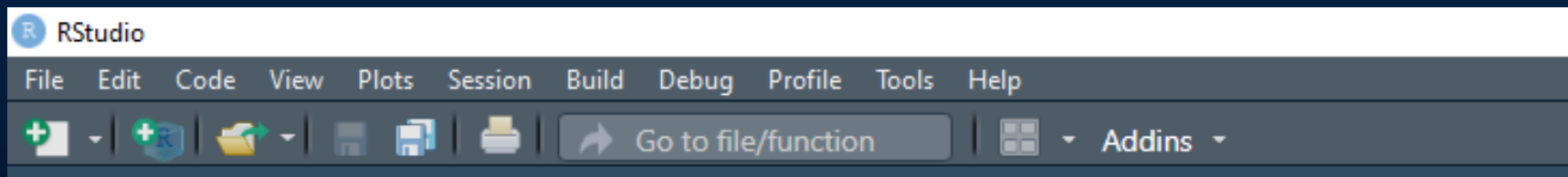
```
R 4.1.2 - C:/Users/Roy.R/Texas A&M University - Corpus Christi/Bird, Chris - Thesis_Roy_Roberts/repos/roberts_thesis/wrangling_data_ssl/
+      species
+      "site_era",
+      "repair",
+      "ng_dna"),
+
+      remove=FALSE,
+      sep = "-")%>%
+
+      select(~starts_with("temp")) %>%
+      mutate(site_era = str_replace(site_era,
+                                   "Apnb"
+                                   "APnb"))
+
+    }else{
+      view(.)
+    }
+  }
+ }
+ }
+ }
```

Plot:

membership probability

Libraries

R Studio

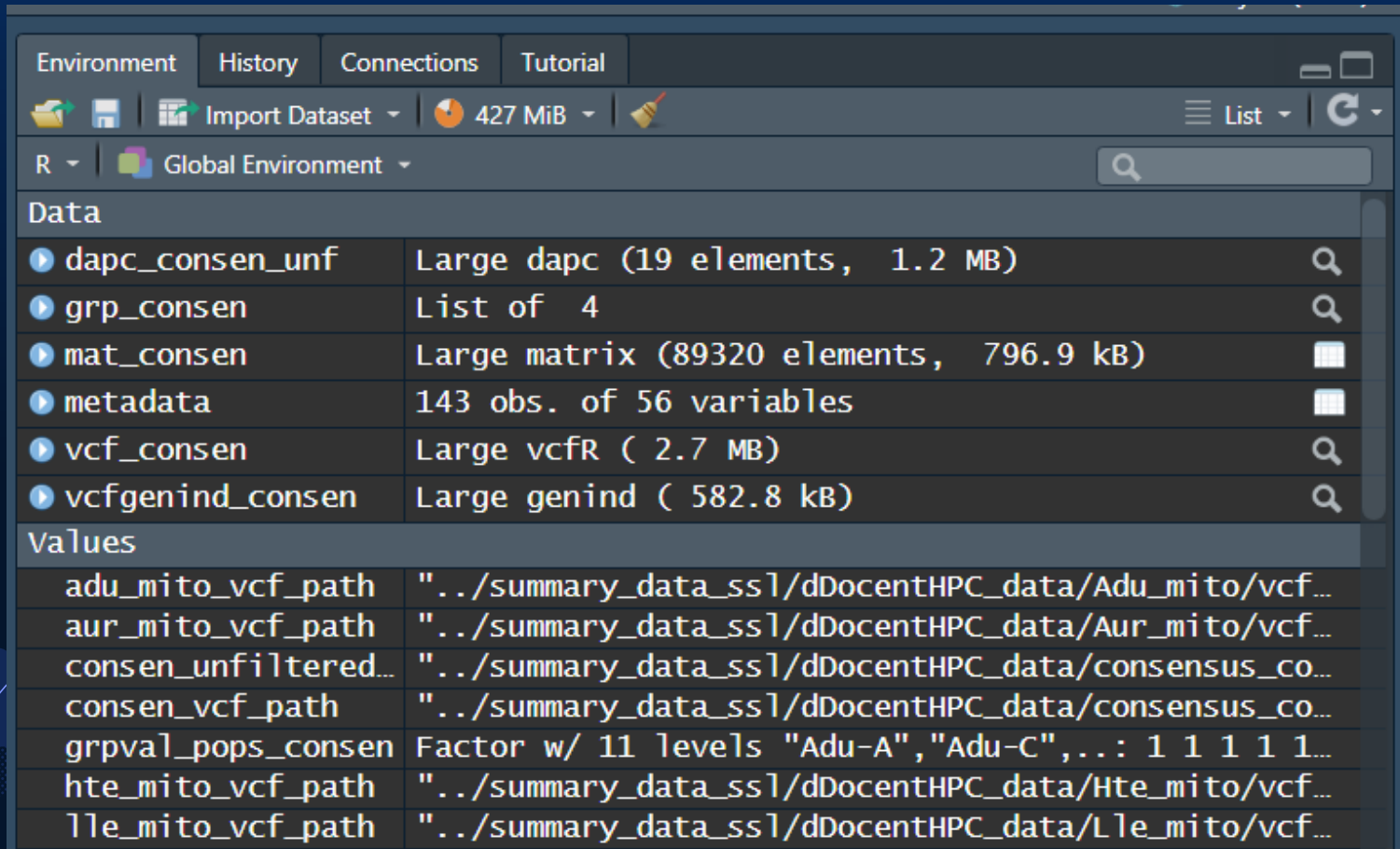


The screenshot shows the RStudio interface with the following elements:

- Top Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Icons for file operations and a "Go to file/function" search bar.
- Tab Bar:** thesis_plots.R, read_data_ssl.R, dapc_mito.R.
- Source Editor:** Contains the following R code:


```
1 rm(list = ls())
2
3
4 library(tidyverse)
5 library(lubridate)
6 library(vcfR)
7 library(adegenet)
8 library(adegraphics)
9 library(pegas)
10 library(seqinr)
11
12 setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
13
14 source("read_data_ssl.R")
15
16 metadata <- read_metadata_rrt("../metadata/")
17
18 ##### VCFS #####
19
20 adu_mito_vcf_path <- "../summary_data_ssl/dDocentHPC_data/Adu_mito/vcf_filtering/adu.all.denovoSSL.Adu-
21 aur_mito_vcf_path <- "../summary_data_ssl/dDocentHPC_data/Aur_mito/vcf_filtering/aur.all.denovoSSL.Hte-
22 hte_mito_vcf_path <- "../summary_data_ssl/dDocentHPC_data/Hte_mito/vcf_filtering/hte.all.denovoSSL.Sob-
23 lle_mito_vcf_path <- "../summary_data_ssl/dDocentHPC_data/Lle_mito/vcf_filtering/lle.all.denovoSSL.Lle-
24 ssp_mito_vcf_path <- "../summary_data_ssl/dDocentHPC_data/Ssp_mito/vcf_filtering/ssp.all.denovoSSL.Sob-
25 tbi_mito_vcf_path <- "../summary_data_ssl/dDocentHPC_data/Tbi_mito/vcf_filtering/tbi.all.denovoSSL.Tbi-
26 consen_unfiltered_vcf_path <- "../summary_data_ssl/dDocentHPC_data/consensus_coi/TotalRawSNPs.denovoSSL-
```
- Right Panel:** Run, Up, Down, and Source buttons.

R Studio



The screenshot shows the R Studio interface with the Environment pane active. The pane displays a list of objects in the Global Environment. The objects are categorized into 'Data' and 'Values'.

Data

Object	Description	Icon
dapc_consen_unf	Large dapc (19 elements, 1.2 MB)	Search icon
grp_consen	List of 4	Search icon
mat_consen	Large matrix (89320 elements, 796.9 kB)	Matrix icon
metadata	143 obs. of 56 variables	Table icon
vcf_consen	Large vcfR (2.7 MB)	Search icon
vcfgenind_consen	Large genind (582.8 kB)	Search icon

Values

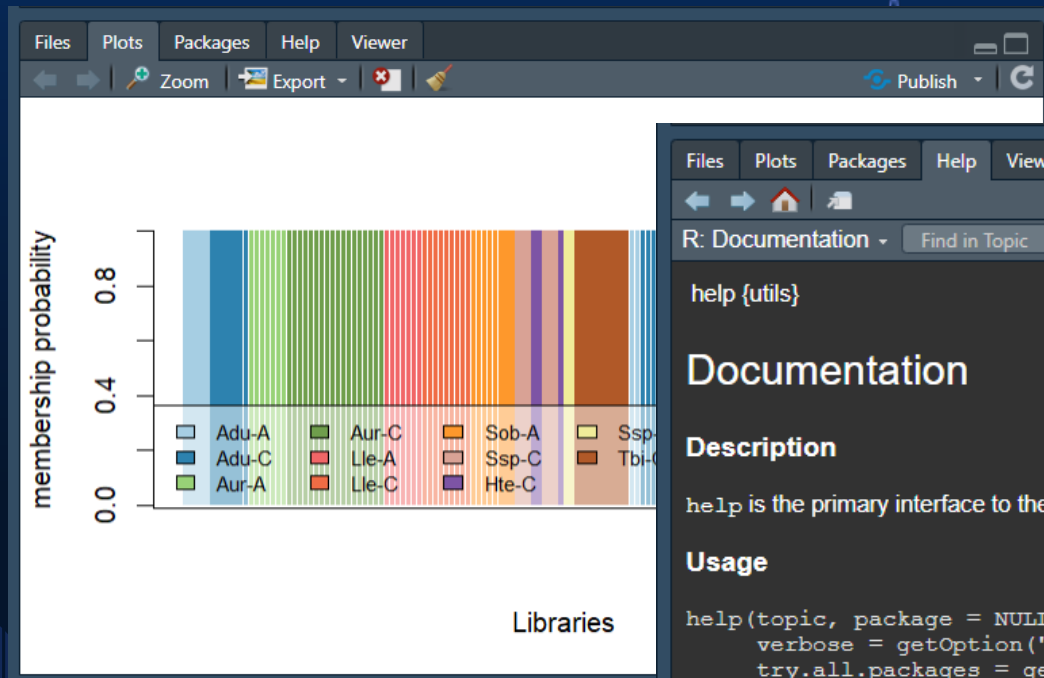
Object	Value
adu_mito_vcf_path	"../summary_data_ssl/dDocentHPC_data/Adu_mito/vcf..."
aur_mito_vcf_path	"../summary_data_ssl/dDocentHPC_data/Aur_mito/vcf..."
consen_unfiltered...	"../summary_data_ssl/dDocentHPC_data/consensus_co..."
consen_vcf_path	"../summary_data_ssl/dDocentHPC_data/consensus_co..."
grpval_pops_consen	Factor w/ 11 levels "Adu-A","Adu-C",...: 1 1 1 1 1...
hte_mito_vcf_path	"../summary_data_ssl/dDocentHPC_data/Hte_mito/vcf..."
lle_mito_vcf_path	"../summary_data_ssl/dDocentHPC_data/Lle_mito/vcf..."

R Studio



```
Console Terminal x Jobs x
R 4.1.2 · C:/Users/Roy_R/Texas A&M University - Corpus Christi/Bird, Chris - Thesis_Roy_Roberts/repos/rroberts_thesis/wrangling_data_ssl/
+       "species",
+       "site_era",
+       "repair",
+       "ng_dna"),
+       remove=FALSE,
+       sep = "-")%>%
+   select(-starts_with("temp")) %>%
+   mutate(site_era = str_replace(site_era,
+                                 "Apnb",
+                                 "APnb"))
+   }else{
+     view(.)
+   }
+ }
> |
```


R Studio



R: Documentation

help {utils}

Documentation

Description

help is the primary interface to the help systems.

Usage

```
help(topic, package = NULL, lib.loc = NULL,
      verbose = getOption("verbose"),
      try.all.packages = getOption("help.try.all.packages"),
      help_type = getOption("help_type"))
```

Arguments

topic	usually, a <u>name</u> or character string specifying the topic for which help is sought. A character string (enclosed in explicit single or double
-------	---

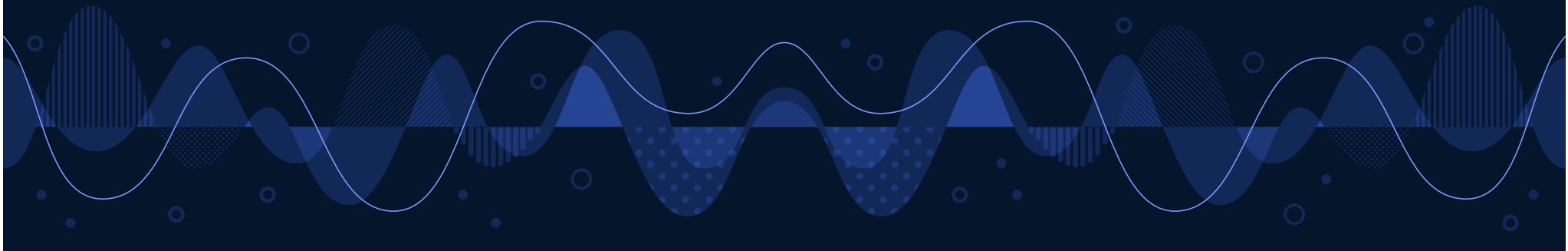
Yellow and Green



= Code that you don't change

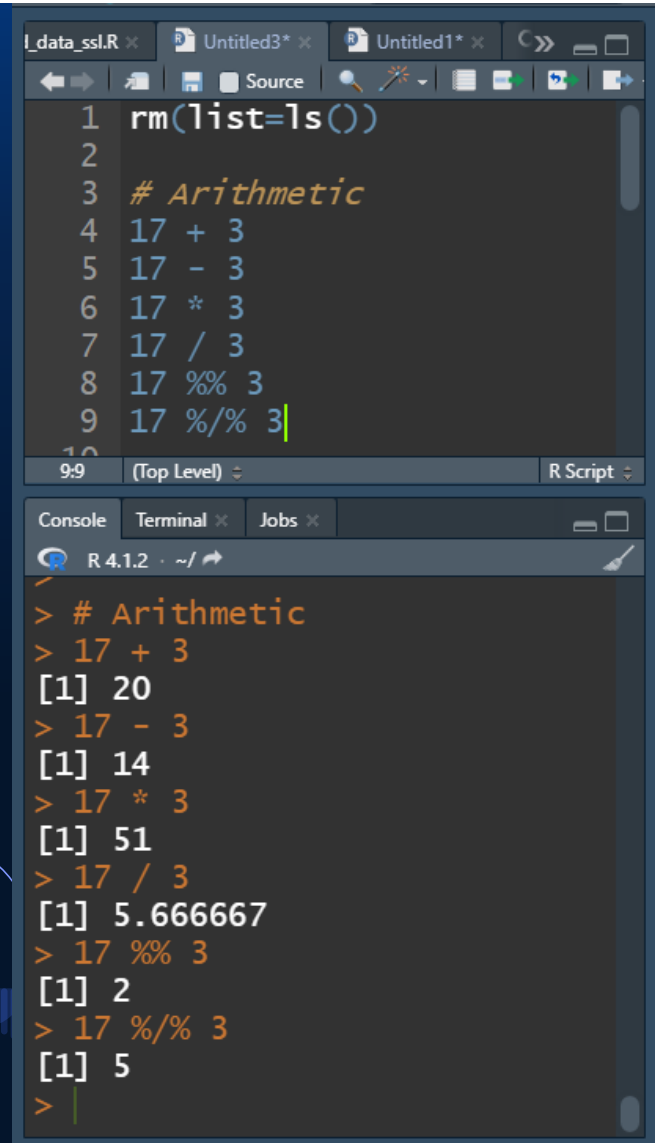


= Code that you customize



Arithmetic

- ▷ $+$ Addition
- ▷ $-$ Subtraction
- ▷ $*$ Multiplication
- ▷ $/$ Division
- ▷ $^$ Exponentiation
- ▷ $\%\%$ Modulo (remainder from division)
- ▷ $\%/\%$ Integer Division



```
1 rm(list=ls())
2
3 # Arithmetic
4 17 + 3
5 17 - 3
6 17 * 3
7 17 / 3
8 17 %% 3
9 17 %/% 3
```

```
> # Arithmetic
> 17 + 3
[1] 20
> 17 - 3
[1] 14
> 17 * 3
[1] 51
> 17 / 3
[1] 5.666667
> 17 %% 3
[1] 2
> 17 %/% 3
[1] 5
>
```

Logical Equalities

- ▷ **==** Equal to
- ▷ **!=** Not equal to
- ▷ **>** Greater than
- ▷ **<** Less than
- ▷ **>=** Greater than or equal
- ▷ **<=** Less than or equal
- ▷ **&** And
- ▷ **|** Or
- ▷ **!** Not

```
_data_ssl.R x  Untitled3* x  Untitled1* x  C>>  [Icons]
[Icons]  [Icons]  [Icons]  [Icons]  [Icons]  [Icons]  [Icons]  [Icons]
Source  [Icons]  [Icons]  [Icons]  [Icons]  [Icons]  [Icons]  [Icons]
11  # Equalities
12  17 == 3
13  17 != 3
14  17 >= 3
15  17 <= 3
16  17 > 3 | 17 < 3
17  17 > 3 & 17 < 3
18
19

18:1  (Top Level)  R Script

Console  Terminal x  Jobs x  [Icons]
R 4.1.2  ~ /  [Icons]
> # Equalities
> 17 == 3
[1] FALSE
> 17 != 3
[1] TRUE
> 17 >= 3
[1] TRUE
> 17 <= 3
[1] FALSE
> 17 > 3 | 17 < 3
[1] TRUE
> 17 > 3 & 17 < 3
[1] FALSE
>
```

Variables

- ▶ Variables hold data
 - ▶ Any alpha-numeric-logical-NA
- ▶ *Values* are numbers or strings
- ▶ Values can be assigned to variables using `<-`
- ▶ Clear all variables:
`rm(list=ls())`

The screenshot displays the RStudio interface. The Source editor on the left contains the following R code:

```
20  
21 # Variables  
22 x <- -1.2345  
23 y <- "Hi"  
24 z <- 6.34 * 10^5  
25 x  
26 y  
27 z  
28  
29  
30  
31  
32
```

The Environment pane on the right, titled 'Global Environment', shows the current values of the variables:

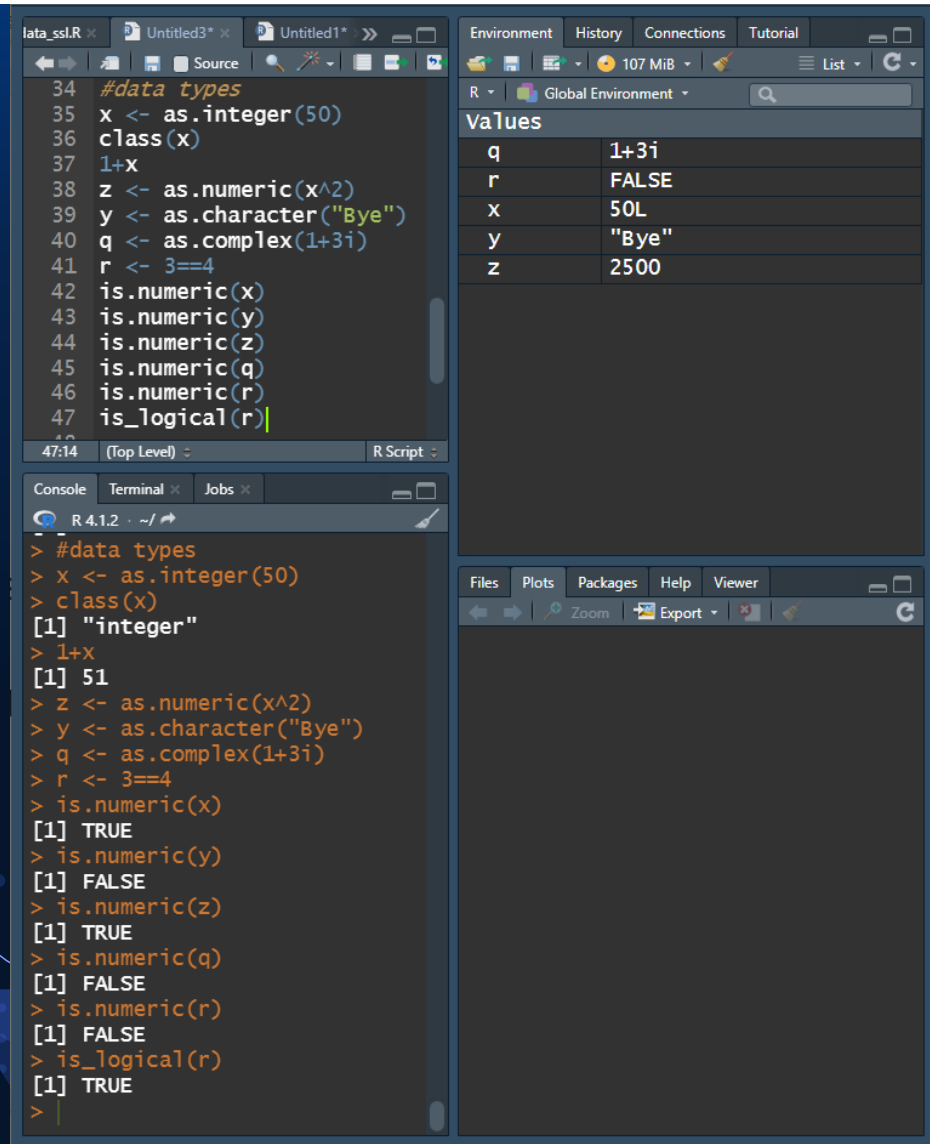
Values	
x	-1.2345
y	"Hi"
z	634000

The Console at the bottom shows the execution of the code, with the output of the variable inspection commands:

```
> # Variables  
> x <- -1.2345  
> y <- "Hi"  
> z <- 6.34 * 10^5  
> x  
[1] -1.2345  
> y  
[1] "Hi"  
> z  
[1] 634000  
>
```

Data Types

- **integer** Whole Numbers
 - **numeric** Real Numbers
 - **complex** Complex Numbers
 - **logical** TRUE or FALSE
 - **character** String
- Specify data type using **`as.type(variable)`**
 - Query data type use **`is.type(variable)`**



The screenshot displays the R Studio environment with three main panels. The top-left panel shows the R script editor with the following code:

```
34 #data types
35 x <- as.integer(50)
36 class(x)
37 1+x
38 z <- as.numeric(x^2)
39 y <- as.character("Bye")
40 q <- as.complex(1+3i)
41 r <- 3==4
42 is.numeric(x)
43 is.numeric(y)
44 is.numeric(z)
45 is.numeric(q)
46 is.numeric(r)
47 is_logical(r)
```

The bottom-left panel shows the console output for the same code:

```
> #data types
> x <- as.integer(50)
> class(x)
[1] "integer"
> 1+x
[1] 51
> z <- as.numeric(x^2)
> y <- as.character("Bye")
> q <- as.complex(1+3i)
> r <- 3==4
> is.numeric(x)
[1] TRUE
> is.numeric(y)
[1] FALSE
> is.numeric(z)
[1] TRUE
> is.numeric(q)
[1] FALSE
> is.numeric(r)
[1] FALSE
> is_logical(r)
[1] TRUE
>
```

The right panel shows the Environment window with the following variables and values:

Variable	Value
q	1+3i
r	FALSE
x	50L
y	"Bye"
z	2500

Mathematical Functions

<code>abs(x)</code>	absolute value
<code>sqrt(x)</code>	square root
<code>ceiling(x)</code>	round up
<code>floor(x)</code>	round down
<code>trunc(x)</code>	remove decimals
<code>round(x, digits = n)</code>	round x to n digits
<code>cos(x), sin(x), tan(x),</code>	etc. trig
<code>log(x)</code>	natural log
<code>exp(x)</code>	e^x
<code>log10(x)</code>	base 10 log

The screenshot displays the RStudio environment with three main panes: the Source editor, the Console, and the Environment pane.

Source Editor: Contains R code for mathematical functions. Lines 53-67 show the assignment of variables and the execution of various mathematical functions.

```
53  
54 #math functions  
55 rm(list=ls())  
56 x <- -1.234  
57 abs_x <- abs(x)  
58 sqrt_abs_x <- sqrt(abs_x)  
59 ceiling(x); ceiling(abs_x)  
60 floor(x); floor(abs_x)  
61 trunc(x); trunc(abs_x)  
62 (y <- round(x, 2))  
63 (cos_y <- cos(y))  
64 (z <- log(abs(y)))  
65 (exp_z <- exp(z))  
66  
67
```

Console: Shows the output of the R script. The output is a series of values printed to the console, corresponding to the operations performed in the source code.

```
> #math functions  
> rm(list=ls())  
> x <- -1.234  
> abs_x <- abs(x)  
> sqrt_abs_x <- sqrt(abs_x)  
> ceiling(x); ceiling(abs_x)  
[1] -1  
[1] 2  
> floor(x); floor(abs_x)  
[1] -2  
[1] 1  
> trunc(x); trunc(abs_x)  
[1] -1  
[1] 1  
> (y <- round(x, 2))  
[1] -1.23  
> (cos_y <- cos(y))  
[1] 0.3342377  
> (z <- log(abs(y)))  
[1] 0.2070142  
> (exp_z <- exp(z))  
[1] 1.23  
>
```

Environment Pane: Displays the current environment, showing the values of the variables created in the script. The variables and their values are listed in a table.

Values	
abs_x	1.234
cos_y	0.334237727124503
exp_z	1.23
sqrt_abs...	1.11085552615991
x	-1.234
y	-1.23
z	0.207014169384326

An Overview of Data Structures in R

- ▷ **Element** a single value or string
- ▷ **Vector** several elements of same data type
- ▷ **Matrix** 2-dimensional table of elements of same data type
- ▷ **Array** 3-dimensional table of elements of same data type
- ▷ **Tibble** 2-dimensional table of elements of different data types
- ▷ **Data frame** 2-dimensional table of elements of different data types
- ▷ **List** several elements or data structures of any type

Vectors

- ▶ A *vector* is a collection of *elements*
 - ▶ 1, 2, 3, 4, 5, 6
- ▶ Vectors can be specified using `c()`, or `seq()` or `rep()`
- ▶ Particular values in the vector can be specified by their index
`[index_num]`

The screenshot displays the RStudio interface with the following components:

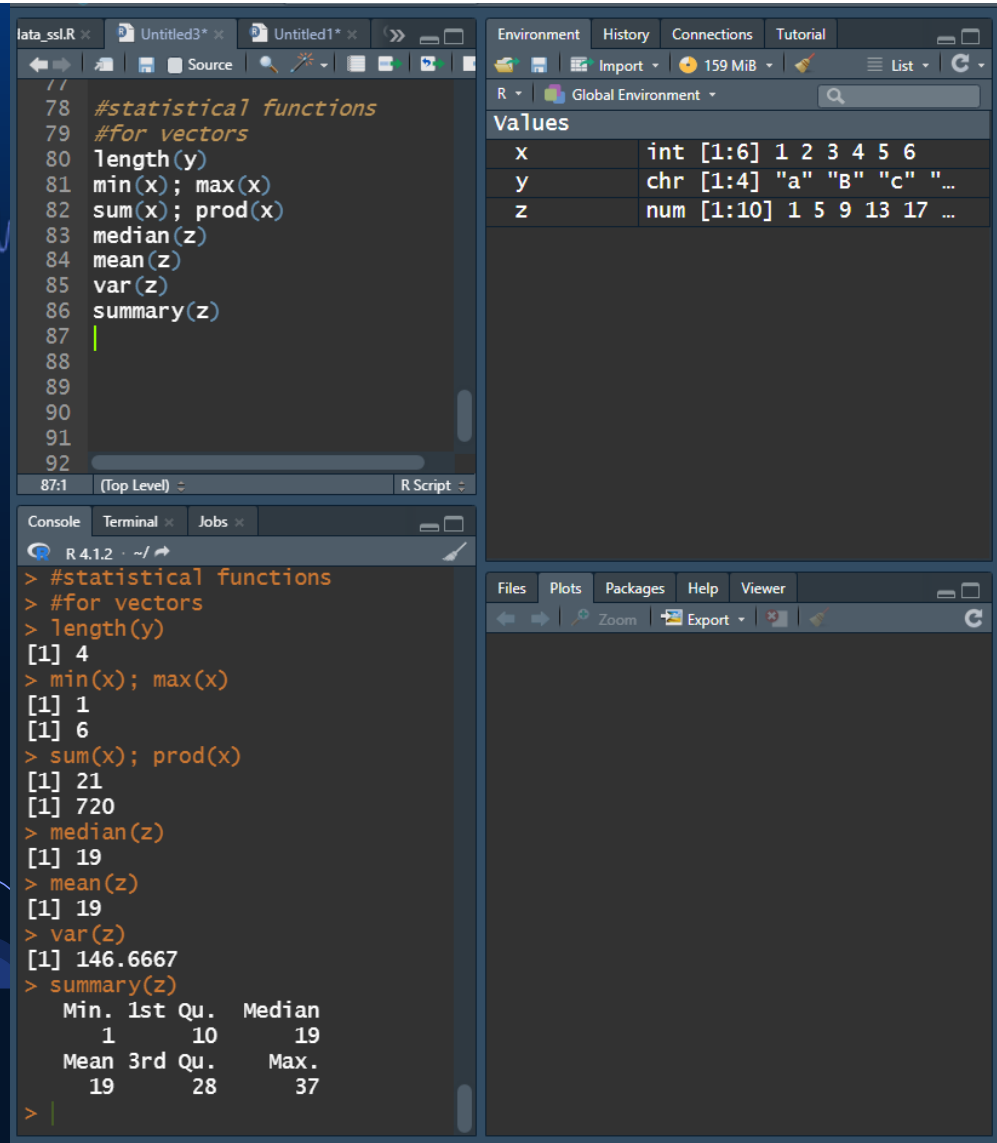
- Source Editor:** Contains R code for creating and manipulating vectors.

```
67 #vectors
68 rm(list=ls())
69 1:6
70 x <- c(1:6)
71 y <- c("a", "B", "c", "D")
72 y[3]
73 z <- seq(1, 40, 4)
74 z
75 z[3:6]
76 z[c(2,8,10)]
77
```
- Environment Pane:** Shows the current workspace with three objects: `x` (integer vector of length 6), `y` (character vector of length 4), and `z` (numeric vector of length 10).
- Console:** Shows the output of the commands entered in the source editor.

```
> rm(list=ls())
> 1:6
[1] 1 2 3 4 5 6
> x <- c(1:6)
> y <- c("a", "B", "c", "D")
> y[3]
[1] "c"
> z <- seq(1, 40, 4)
> z
[1] 1 5 9 13 17 21 25 29 33 37
> z[3:6]
[1] 9 13 17 21
> z[c(2,8,10)]
[1] 5 29 37
>
```

Statistical Functions

- ▷ **length(x)** # elements
- ▷ **min(x)** minimum
- ▷ **max(x)** maximum
- ▷ **sum(x)** sum, +
- ▷ **prod(x)** product, *
- ▷ **median(x)** central value
- ▷ **mean(x)** average
- ▷ **sd(x)** standard deviation
- ▷ **summary(x)**



The screenshot displays the RStudio environment. The top-left pane shows a script with the following code:

```
//  
78 #statistical functions  
79 #for vectors  
80 length(y)  
81 min(x); max(x)  
82 sum(x); prod(x)  
83 median(z)  
84 mean(z)  
85 var(z)  
86 summary(z)  
87  
88  
89  
90  
91  
92
```

The bottom-left pane (Console) shows the output of these functions:

```
> #statistical functions  
> #for vectors  
> length(y)  
[1] 4  
> min(x); max(x)  
[1] 1  
[1] 6  
> sum(x); prod(x)  
[1] 21  
[1] 720  
> median(z)  
[1] 19  
> mean(z)  
[1] 19  
> var(z)  
[1] 146.6667  
> summary(z)  
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
    19      28      37
```

The top-right pane (Environment) shows the current state of the workspace:

Variable	Class	Attributes	Values
x	int	[1:6]	1 2 3 4 5 6
y	chr	[1:4]	"a" "B" "c" "..."
z	num	[1:10]	1 5 9 13 17 ...

The bottom-right pane (Plots) is currently empty.



“

▸ Base R Mind Expander

<https://forms.gle/nE2CnAJj6rz8QVYo9>