

Basic R Commands

Viewing and Exploring Files in R Let's refresh our knowledge in R language. As you know, in R, we use **getwd()** and **setwd()** functions to show and set the current working directory respectively. In the same way, we use **list.dirs()** and **list.files()** functions, to list directories and files, sequentially.

```
# Get the current working directory. getwd()
```

```
## [1] "C:/Users/Administrator/Documents"
```

```
# List all folders and sub-folders in the current working directory. list.dirs()
```

```
## [1] "."
## [2] "./shiny2020"
## [3] "./shiny2020/Task-1"
## [4] "./shiny2020/Task-1/Shiny_app_Example"
## [5] "./shiny2020/Task-1/Shiny_app_Example/inc"
## [6] "./shiny2020/Task-1/Shiny_app_Example/www"
## [7] "./shiny2020/Task-1/Shiny_gadget_Example"
## [8] "./shiny2020/Task-1/Shiny_gadget_Example/inc"
## [9] "./shiny2020/Task-2"
## [10] "./shiny2020/Task-2/Multiple_File_App"
## [11] "./shiny2020/Task-2/single"
## [12] "./shiny2020/Task-2/Single_File_App"
## [13] "./shiny2020/Task-3"
## [14] "./shiny2020/Task-3/inputs"
## [15] "./shiny2020/Task-3/inputs/inc"
## [16] "./shiny2020/Task-4"
## [17] "./shiny2020/Task-4/outputs"
## [18] "./shiny2020/Task-4/outputs/inc"
## [19] "./shiny2020/Task-5"
## [20] "./shiny2020/Task-5/reactive"
## [21] "./shiny2020/Task-5/reactive/inc"
## [22] "./shiny2020/Task-5/reactive/www"
## [23] "./shiny2020/Task-6"
## [24] "./shiny2020/Task-6/themes"
## [25] "./shiny2020/Task-6/themes/inc"
## [26] "./shiny2020/Task-6/themes/www"
## [27] "./shiny2020/Task-7"
## [28] "./shiny2020/Task-7/dashboard"
## [29] "./shiny2020/Task-7/dashboard/inc"
## [30] "./shiny2020/Task-7/dashboard/www"
## [31] "./shiny2020/Task-8"
## [32] "./shiny2020/Task-8/gadgets"
## [33] "./shiny2020/Task-8/gadgets/inc"
```

```
# Change current working directory to be dashboard folder, # inside Task-7
folder inside the 'shiny2020' folder.
# We use the 'paste' function to append the path to the current working directory.
setwd(paste(getwd(), '/shiny2020/Task-7/dashboard', sep = ''))
```

```
# Get the current working directory to make sure it is changed. getwd()
```

```
## [1] "C:/Users/Administrator/Documents/shiny2020/Task-7/dashboard"
```

```
# List all contents in the current working directory. list.files()
```

```
## [1] "app.R" "inc"          "www"
```

As you can see in the previous code, we have one file **'app.R'** and two folders **'inc'** and **'www'**. Let's see what inside those two folders.

```
# List 'inc' folder contents. list.files('./inc')
```

```
## [1] "g1.csv"          "g2.csv"          "g3.csv"
## [4] "genes_names.csv" "help.html"       "information.html"
## [7] "introduction.html" "references.html"
```

Inside **'inc'** folder, we have four **CSV** files, **g1**, **g2**, **g3**, and **genes_names**. Also, we have four **HTML** files, **'help'**, **'information'**, **'introduction'**, and **'references'**.

```
# List 'www' folder contents. list.files('./www')
```

```
## [1] "bootstrap.css"          "bootstrap.min.css" "dropdown_one.png"
## [4] "dropdown_three.png" "dropdown_two.png"   "mytheme.css"
```

Inside **'www'** folder, we have a punch of **CSS** files with names **'bootstrap'**, **'bootstrap.min'**, and **'mytheme'**, and three **PNG** files with names **'dropdown_one'**, **'dropdown_three'**. **'dropdown_two'**, and Let's see some of those files.

```
# Read the g1.csv file.
```

```
g1 <- read.csv("./inc/g1.csv")
```

```
# Show the first part of the g1 data frame. head(g1)
```

```
##      Gene_ID                                     Gene_full_name
## 1 Anapc2      Anaphase promoting complex subunit 2 ## 2      Asb6
Ankyrin repeat and SOCS box-containing protein 6
## 3      Fbxw2      F-box and WD-40 domain protein 2 ## 4 Gpr107      G protein-
coupled receptor 107
## 5          Il1a                                     Interleukin 1 a
## 6          Il1rn      Interleukin 1 receptor anatagonist
##          Resistant_or_Susceptible_strains Relative_expression_levels      SD
```

```
## 1    Resistant      0.551 0.022 ## 2 Resistant      0.573 0.482 ## 3 Resistant
0.930 0.384 ## 4 Resistant      0.866 0.481 ## 5 Resistant      0.739 0.045 ## 6
Resistant      0.851 0.384
```

Change long names to shorter ones.

```
names(g1)[names(g1) == "Resistant_or_Susceptible_strains"] <- 'R_or_S_Strain' names(g1)[names(g1) ==
"Relative_expression_levels"] <- 'REL'
```

Show a summary of the g1 data frame. summary(g1)

```
##      Gene_ID                                     Gene_full_name
## Anapc2 : 2    Anaphase promoting complex subunit 2      : 2 ## Asb6      : 2
Ankyrin repeat and SOCS box-containing protein 6: 2
## Fbxw2 : 2    F-box and WD-40 domain protein 2 : 2 ## Gpr107 : 2 G protein-coupled
receptor 107      : 2
## Il1a      : 2    Interleukin 1 a                      : 2
## Il1rn : 2      Interleukin 1 receptor antagonist        : 2
## (Other):18      (Other)                                :18
##      R_or_S_Strain      REL      SD
## Resistant :15 Min. : 0.1620 Min. :0.0110 ## Susceptible:15 1st Qu.:
0.7185 1st Qu.:0.1520
##      Median : 0.9215 Median :0.2985 ##      Mean      : 4.7260
Mean      :0.3912 ##      3rd Qu.: 1.3802 3rd Qu.:0.5353 ##
Max.      :55.9090Max.      :1.8010
##
```

Use knitr to create a better table. library(knitr)

```
kable(g1[1:5, ], caption = 'The first 5 rows of g1 data frame')
```

Table 1: The first 5 rows of g1 data frame

Gene_ID	Gene_full_name	R_or_S_Strain	REL	SD
Anapc2	Anaphase promoting complex subunit 2	Resistant	0.551	0.022
Asb6	Ankyrin repeat and SOCS box-containing protein 6	Resistant	0.573	0.482
Fbxw2	F-box and WD-40 domain protein 2	Resistant	0.930	0.384
Gpr107	G protein-coupled receptor 107	Resistant	0.866	0.481
Il1a	Interleukin 1 a	Resistant	0.739	0.045

```
kable(g1[25:30, ], caption = 'The last 5 rows of g1 data frame')
```

Table 2: The last 5 rows of g1 data frame

	Gene_ID	Gene_full_name	R_or_S_Strain	REL	SD
25	Ptges	Prostaglandin E synthase	Susceptible	5.046	0.679
26	Ptges2	Prostaglandin E synthase 2	Susceptible	1.456	0.122
27	Rab14	RAB14, member RAS oncogene family	Susceptible	1.411	0.711
28	Sh2d3c	SH2 domain containing 3C	Susceptible	1.009	0.389
29	Phyhd1	Phytanoyl-CoA dioxygenase domain containing 1	Susceptible	1.288	0.288
30	Urm1	Ubiquitin related modifier 1 homolog (S. cerevisiae)	Susceptible	1.277	0.309

```
# Create a new data frame by merging the first 15 rows of 'g1'
# with the last 15 rows of 'g1' too. Include only columns 1, 3, 4, and 5.
# (we do not need to repeat the 'Gene_full_name' twice)
gA <- merge(g1[1:15,], g1[16:30,c(1,3,4,5)], by='Gene_ID')
# Create a table for the new data frame gA, # without showing column no 2
(Gene_full_name). kable(gA[1:6,-2 ], caption = 'The first 6 rows of gA data
frame')
```

Table 3: The first 6 rows of gA data frame

Gene_ID	R_or_S_Strain.x	REL.x	SD.x	R_or_S_Strain.y	REL.y	SD.y
Anapc2	Resistant	0.551	0.022	Susceptible	1.475	0.691
Asb6	Resistant	0.573	0.482	Susceptible	2.300	1.801
Fbxw2	Resistant	0.930	0.384	Susceptible	1.159	0.268
Gpr107	Resistant	0.866	0.481	Susceptible	2.196	0.758
Il1a	Resistant	0.739	0.045	Susceptible	54.687	0.011
Il1rn	Resistant	0.851	0.384	Susceptible	55.909	0.274

```
# Use library magic to show an image. library("magick") # Load the image
img <- magick::image_read('./www/dropdown_one.PNG')
# In Rstudio the image will be shown in the Viewer.
plot(img)
```

Choose Group to plot:

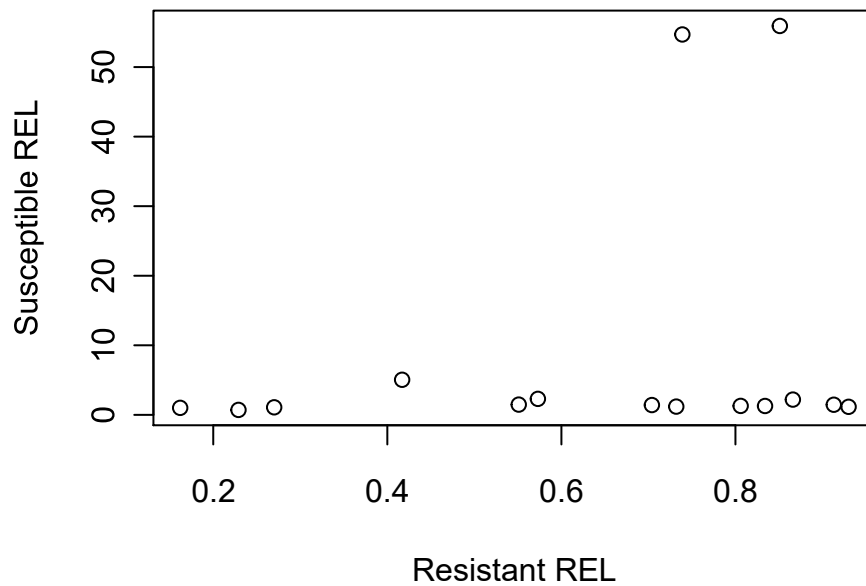
1- Genes down regulated in resistant while up regulated in susceptible

1- Genes down regulated in resistant while up regulated in susceptible

2- Genes down regulated in both resistant and susceptible

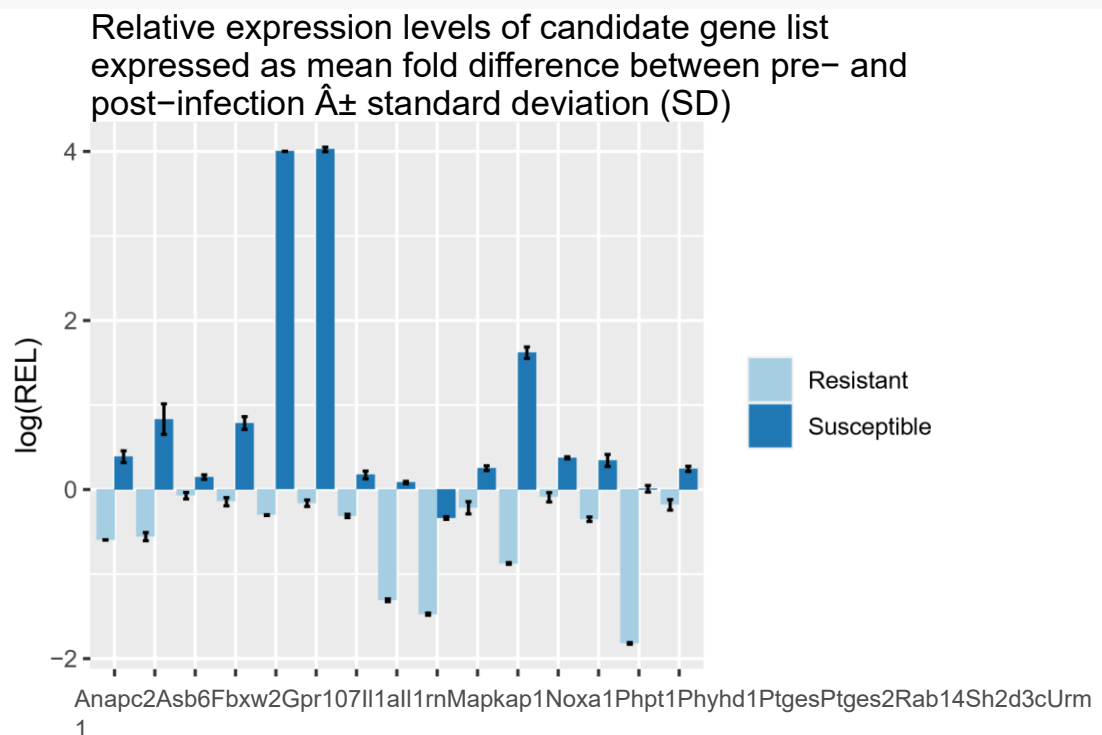
3- Genes up regulated in both resistant and susceptible

```
# Load ggplot2 package so we can create some graphs. library(ggplot2)
# Create a simple plot.
plot(g1$REL[1:15], g1$REL[16:30], xlab="Resistant REL",
      ylab="Susceptible REL")
```



Construct a plot with more details.

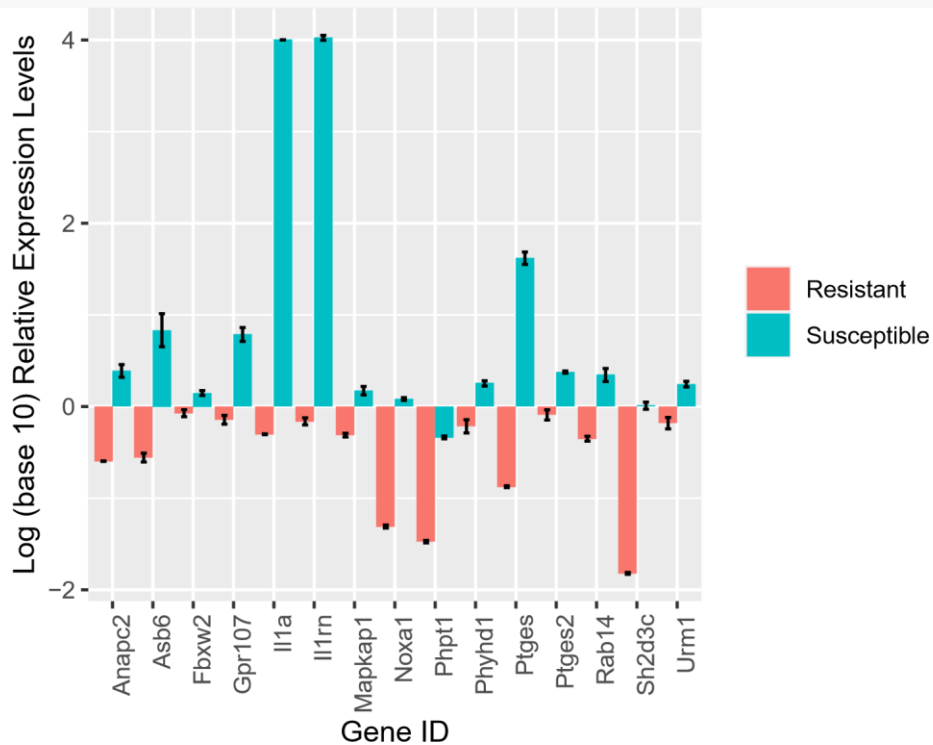
```
p <- ggplot(g1, aes(x=Gene_ID, y=log(REL), fill=R_or_S_Strain)) +  
  geom_bar(stat="identity", position=position_dodge()) + geom_errorbar(aes(ymin=log(REL)-  
    (SD/10), ymax=log(REL)+(SD/10)),width=.3, position=position_dodge(.9))  
p + scale_fill_brewer(palette="Paired") + ggtitle(paste("Relative expression levels of candidate  
  gene list", "\n",  
    "expressed as mean fold difference between pre- and",  
    "\n", "post-infection  $\hat{\pm}$  standard deviation (SD) ")) +  
  guides(fill=guide_legend(title=NULL))
```



Gene_ID

```
# We can add more sophisticated components like themes, # to change color and x axis labels
..etc. p$theme <- theme(axis.text.x = element_text(angle = 90, hjust = 1)) p$labels$x <- "Gene
ID"

p$labels$y <- "Log (base 10) Relative Expression Levels" p$labels$fill <-
NULL return(p)
```



Using R, we can even print different file contents.

```
# Show the g1.csv file contents. cat(readLines('./inc/g1.csv'), sep = '\n')
```

```
Gene_ID,Gene_full_name,Resistant_or_Susceptible_strains,Relative_expression_levels,SD
Anapc2,Anaphase promoting complex subunit 2,Resistant,0.551,0.022
Asb6,Ankyrin repeat and SOCS box-containing protein 6,Resistant,0.573,0.482
Fbxw2,F-box and WD-40 domain protein 2,Resistant,0.93,0.384 Gpr107,G protein-
coupled receptor 107,Resistant,0.866,0.481
Il1a,Interleukin 1 a ,Resistant,0.739,0.045
Il1rn,Interleukin 1 receptor antagonist,Resistant,0.851,0.384
Mapkap1,Mitogen-activated protein kinase associated protein
1,Resistant,0.732,0.199
Noxa1,NADPH oxidase activator 1,Resistant,0.27,0.164
Phpt1,Phosphohistidine phosphatase 1,Resistant,0.229,0.133
Ptges,Prostaglandin E synthase,Resistant,0.417,0.101
Ptges2,Prostaglandin E synthase 2,Resistant,0.913,0.553
Rab14,"RAB14, member RAS oncogene family",Resistant,0.704,0.268
```

Sh2d3c,SH2 domain containing 3C,Resistant,0.162,0.102
Phyhd1,Phytanoyl-CoA dioxygenase domain containing 1,Resistant,0.806,0.727
Urm1,Ubiquitin related modifier 1 homolog (S. cerevisiae),Resistant,0.834,0.625
Anapc2,Anaphase promoting complex subunit 2,Susceptible,1.475,0.691
Asb6,Ankyrin repeat and SOCS box-containing protein 6,Susceptible,2.3,1.801
Fbxw2,F-box and WD-40 domain protein 2,Susceptible,1.159,0.268
Gpr107,G protein-coupled receptor 107,Susceptible,2.196,0.758
Il1a,Interleukin 1 a ,Susceptible,54.687,0.011
Il1rn,Interleukin 1 receptor antagonist,Susceptible,55.909,0.274
Mapkap1,Mitogen-activated protein kinase associated protein
1,Susceptible,1.189,0.46
Noxa1,NADPH oxidase activator 1,Susceptible,1.086,0.151
Phpt1,Phosphohistidine phosphatase 1,Susceptible,0.714,0.155
Ptges,Prostaglandin E synthase,Susceptible,5.046,0.679
Ptges2,Prostaglandin E synthase 2,Susceptible,1.456,0.122
Rab14,"RAB14, member RAS oncogene family",Susceptible,1.411,0.711
Sh2d3c,SH2 domain containing 3C,Susceptible,1.009,0.389
Phyhd1,Phytanoyl-CoA dioxygenase domain containing 1,Susceptible,1.288,0.288 Urm1,Ubiquitin
related modifier 1 homolog (S. cerevisiae),Susceptible,1.277,0.309

```
# Show the help.html file contents. cat(readLines('./inc/help.html'), sep = '\n')
```

<h2>Help</h2>

A- From the First dropdown box (as shown in the next figure) you can select one of the three options:

1- Genes down regulated in resistant while up regulated in susceptible.

2- Genes down regulated in both resistant and susceptible.

3- Genes up regulated in both resistant and susceptible.
 Each option show one plot related to what you select.

B- From the Second dropdown box (as shown in the next figure) you can select Gene ID and the full gene name will display under the plot.

C- From the third dropdown box (as shown in the next figure) you can select:

1- Introduction, 2- Information, 3- Help (current), 4- References to show more information.


```
# Show mytheme.css file contents. cat(readLines('./www/mytheme.css'), sep = '\n')
```

```
body { background-color:black;
font-size: 1.80em; color:
yellow;
```

```

}

h2, h3, h4, h5, h6, h7 { margin: 0.5em
  0 0.4em; padding: 0; letter-
  spacing: 1px;
}
a {
  color: #bb0011; text-decoration: none;
}

```

```

# Load 'htmltools' package to create some HTML tags. library(htmltools)
# Create a clickable link with text and link. tags$a(href='https://www.coursera.org/',
'Go To Coursera')

```

Go To Coursera

```

# Print the code that creates the previous link.
print(tags$a(href='https://www.coursera.org/', 'Go To Coursera'))

```

```
## <a href="https://www.coursera.org/">Go To Coursera</a>
```

Currently, the Comprehensive R Archive Network (CRAN) package repository features more than **15K** available packages. One of those useful packages in genetics is Gene Book. You can use it to retrieve more information about any gene using the gene ID.

```

# First we load the library to use it. library(GeneBook)
# GeneCard_Symbol_Details is a function to get the detailed information # of the gene
interested. Here we search for 'Gpr107' gene.
GeneCard_Symbol_Details('Gpr107')

```

```
## [1] gene          type          description    summary_entrez
## [5] summary_genecard summary_uniport summary_Tocris summary_CIViC ## <0 rows> (or
0-length row.names)
```

The function returned five columns but with zero row! This is often due to the fact that the function requires specific formats for the gene ID. The GeneBook library has a function for transferring a single gene or multiple genes to the consistent gene symbols.

```

# Convert the gene Id to the proper format.
GeneCard_ID_Convert('Gpr107')

```

```
## [1] "GPR107" "Trust"
```

The library required that the gene ID is all caps. We can use the **GeneCard_ID_Convert** function first then, the **GeneCard_Symbol_Details** will retrieve all information available for that gene.

```
GeneCard_Symbol_Details(GeneCard_ID_Convert('Gpr107'))
```



```
## gene type description
## 1 "GPR107" Protein Coding G Protein-Coupled Receptor 107 ##
summary_entrez
## 1 Additional gene information for GPR107 Gene HGNC(17830)Entrez
Gene(57720)Ensembl(ENSG00000148358)UniProtKB(Q5VW38) Search for GPR107 at
DataMedSearch for GPR107 at HumanCyc
## summary_genecard
## 1 GeneCards Summary for GPR107 Gene GPR107 (G Protein-Coupled Receptor 107) is a Protein
Coding gene.An important paralog of this gene is GPR108.
## summary_uniprot
## 1 UniProtKB/Swiss-Prot for GPR107 GeneGP107_HUMAN,Q5VW38Involved in Golgi-to-ER
retrograde transport. Functions as a host factor required for infection by Pseudomonas aeruginosa
exotoxin A and Campylobacter jejuni CDT toxins.
## summary_Tocris summary_CIViC ## 1
<NA> <NA>
```

We have a lot of data here. Better to save to a list and access its parts as needed.

```
# Assign the results to 'gpr107_data'. gpr107_data <-
GeneCard_Symbol_Details(GeneCard_ID_Convert('Gpr107'))
# Print only 'description' inside 'gpr107_data'. gpr107_data$description
```

```
## [1] "G Protein-Coupled Receptor 107"
```

```
# Print only summary_genecard inside 'gpr107_data'
gpr107_data$summary_genecard
```

```
## [1] "GeneCards Summary for GPR107 Gene GPR107 (G Protein-Coupled Receptor 107) is a Protein
Coding gene.An important paralog of this gene is GPR108. "
```