

Code_Lib_Import

Ruhika Chatterjee

2025-01-26

Getting the Data

Defining Workspace or directory

```
x <- getwd() # find working directory
x
```

```
## [1] "C:/Users/Owner/OneDrive/Documents/CompSci/CourseraDS/CourseraDataScience"
```

```
dir.create("testdir") # create a directory if doesn't exist, args: dir name, for nested recursive = true
```

```
## Warning in dir.create("testdir"): 'testdir' already exists
```

```
setwd("testdir") # set working dir
file.create("mytest.R") # create file in wd
```

```
## [1] TRUE
```

```
file.exists("mytest.R") # check if file or directory exists in wd
```

```
## [1] TRUE
```

```
file.info("mytest.R") # file metadata, use $ operator to grab specific items
```

```
##           size isdir mode                mtime                ctime
## mytest.R    0 FALSE  666 2025-01-27 19:44:00 2025-01-27 19:44:00
##           atime exe
## mytest.R 2025-01-27 19:44:00 no
```

```
file.rename("mytest.R", "mytest2.R") # rename
```

```
## [1] TRUE
```

```
file.copy("mytest2.R","mytest3.R") # copy file
```

```
## [1] FALSE
```

```
file.remove("mytest2.R") # remove file
```

```
## [1] TRUE
```

```
file.path("mytest3.R") # relative path
```

```
## [1] "mytest3.R"
```

```
setwd(x) # could use relative setwd("../") to move up one, setwd("./data"), or absolute path directly u
```

```
dir() # output files in directory. Also list.files()
```

```
## [1] "API_auth_ex.R"           "Code_lib_Export.html"
## [3] "Code_lib_Export.Rmd"      "Code_Lib_Import.pdf"
## [5] "Code_Lib_Import.Rmd"      "Code_Lib_Manipulate.pdf"
## [7] "Code_Lib_Manipulate.Rmd"   "coded.R"
## [9] "complete.R"               "corr.R"
## [11] "Course_Notes.pdf"         "Course_Notes.Rmd"
## [13] "CourseraDataScience.Rproj" "example.h5"
## [15] "getdata_data_ss06pid.csv" "getdata_wksst8110.for"
## [17] "hw1_data.csv"             "pollutantmean.R"
## [19] "Programming 2.3"          "quiz_data_3.1.2.xlsx"
## [21] "quiz_data_3.1.csv"        "Rprof.out"
## [23] "rprog_data_ProgAssignment3" "specdata"
## [25] "specdata.zip"             "testdir"
```

```
files_full <- list.files("specdata", full.names=TRUE) # pull all file names from a directory
ls() # prints the objects in work space
```

```
## [1] "files_full" "x"
```

```
rm(list=ls()) # clear workspace
rm(list=setdiff(ls(), "x")) # clear workspace except x
version #R info version
```

```
##
## platform      _
## arch          x86_64-w64-mingw32
## os            x86_64
## crt            mingw32
## crt           ucrt
## system        x86_64, mingw32
## status
## major         4
## minor         4.1
## year          2024
```

```
## month          06
## day            14
## svn rev        86737
## language       R
## version.string R version 4.4.1 (2024-06-14 ucrt)
## nickname       Race for Your Life
```

```
sessionInfo() #R info version, packages
```

```
## R version 4.4.1 (2024-06-14 ucrt)
## Platform: x86_64-w64-mingw32/x64
## Running under: Windows 11 x64 (build 22631)
##
## Matrix products: default
##
##
## locale:
## [1] LC_COLLATE=English_United States.utf8
## [2] LC_CTYPE=English_United States.utf8
## [3] LC_MONETARY=English_United States.utf8
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.utf8
##
## time zone: America/New_York
## tzcode source: internal
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## loaded via a namespace (and not attached):
## [1] compiler_4.4.1    fastmap_1.2.0     cli_3.6.3        tools_4.4.1
## [5] htmltools_0.5.8.1 rstudioapi_0.17.1 yaml_2.3.10      rmarkdown_2.29
## [9] knitr_1.49        xfun_0.49         digest_0.6.37    rlang_1.1.4
## [13] evaluate_1.0.3
```

```
source("coded.R") # load code into console
```

```
## [1] "Hello World"
```

```
args(ls()) # get arguments for a function
```

```
## NULL
```

```
help(ls) # access documentation on ls() function
```

```
## starting httpd help server ... done
```

```
?ls # same. for operator use ?`:`
```

```
# Interface to outside world
```

```
file(description = "hw1_data.csv") # open connection to standard, uncompressed file. Helps for partial
```

```
## A connection with
## description "hw1_data.csv"
## class      "file"
## mode       "r"
## text       "text"
## opened     "closed"
## can read   "yes"
## can write  "yes"
```

```
# gzfile() # connection to file w compression gzip
# bzfile() # connection to file w compression bzip2
jh <- url("http://www.jhsph.edu", "r") # connection to webpage
close(jh) # to end connection

# Download data from website
fileUrl <- "https://hub.arcgis.com/api/v3/datasets/42f8856d647a41b89561e10fb60bc98a_0/downloads/data?fo
if(!dir.exists("./testdir")) {
  dir.create("./testdir")
}
download.file(fileUrl, destfile = "./testdir/restdata.csv", method = "curl") # improve reproducibility.
dateDownloaded <- date() # Keep track of date downloaded for file
restdata <- read.csv("./testdir/restdata.csv")
```

CSV and XLSX

```
# Read table or csv data into R
x <- read.table("hw1_data.csv", header = TRUE, sep = ",") # reading tabular data from text files, return
head(x)
```

```
##      Ozone Solar.R Wind Temp Month Day
## 1      41      190  7.4   67     5   1
## 2      36      118  8.0   72     5   2
## 3      12      149 12.6   74     5   3
## 4      18      313 11.5   62     5   4
## 5      NA       NA 14.3   56     5   5
## 6      28       NA 14.9   66     5   6
```

```
x <- read.csv("hw1_data.csv") # Same but default separator is ", " and header = TRUE
# write.table(x)
```

```
# help read.table with colClasses with smaller sample
initial <- read.table("hw1_data.csv", header = TRUE, sep = ",", nrows = 100)
classes <- sapply(initial, class)
tabAll <- read.table("hw1_data.csv", header = TRUE, sep = ",", colClasses = classes)
```

```
# Read Excel file into R
x <- "https://stg-arcgisazurecdataproduct1.azurearcgis.com/exportfiles-8298-626/Hfai_6672659153760215301.xls
if(!file.exists("./testdir/foodavail.xlsx")) {
  file.create("./testdir/foodavail.xlsx") }
download.file(x, destfile = "./testdir/foodavail.xlsx", method = "curl") # improve
dateXl <- date()
library(readxl) # xlsx, XLConnect other options
```

```
## Warning: package 'readxl' was built under R version 4.4.2
```

```
cameraData <- read_excel("./testdir/foodavail.xlsx", sheet = 1, col_types = "text", n_max = 100)
head(cameraData)
```

```
## # A tibble: 6 x 7
##   OBJECTID 'Community Statistical Area (CSA)' CSA2010 '2012' '2015' Shape__Area
##   <chr>    <chr>                                <chr>    <chr> <chr>    <chr>
## 1 1      Allendale/Irvington/S. Hilton          Allenda~ 7.75   7.526~ 63770461.7~
## 2 2      Beechfield/Ten Hills/West Hills        Beechfi~ 15.58~ 10.4   47882527.7~
## 3 3      Belair-Edison                          Belair~~ 10.32~ 8.227~ 44950030.4~
## 4 4      Brooklyn/Curtis Bay/Hawkins Point      Brookly~ 8.588~ 8.793~ 176077742.~
## 5 5      Canton                                Canton   16.9   15.2   15408538.2~
## 6 6      Cedonia/Frankford                      Cedonia~ 12.33~ 8.956~ 71541340.3~
## # i 1 more variable: Shape__Length <chr>
```

```
# write also works
```

Read text file and R objects

```
lines <- readLines("coded.R") # reading lines of text file, return character vector
```

```
## Warning in readLines("coded.R"): incomplete final line found on 'coded.R'
```

```
writeLines("coded.R")
```

```
## coded.R
```

```
# editable textual format retains metadata, helpful for version control, corruption fixable, memory cos
dget("coded.R") # reading R objects deparsed into text files
```

```
## [1] "Hello World"
```

```
dput("coded.R") # takes R object, create R code to reconstruct object saving attributes, names
```

```
## "coded.R"
```

```
source("coded.R") # reading in R code files
```

```
## [1] "Hello World"
```

```
#dump() # multiple R objects
```

```
# load() # read in saved workspace read binary objects into R
# save()
# unserialize() # read single R objects in binary form
# serialize()
```

Just XML stuff

- XML (extensible markup language)
- <http://en.wikipedia.org/wiki/XML>
- store structured data in internet applications. From web scraping, API
- Components: markup (labels for structure) and content (text)
- Tags (labels): start
end
empty
- Elements: example of tag Hello, world
- Attributes: component of label
- XPath language: <http://www.stat.berkeley.edu/~statcur/Workshop2/Presentations/XML.pdf>
- /node (top level node), //node (node any level), node[@attr-name] node w attr name, node[@attr-name='bob']
- XML Short Intro: <https://www.omegahat.net/RXML/shortIntro.pdf>
- XML Long Intro: <https://www.omegahat.net/RXML/Tour.pdf>

```
# Reading file
library(XML)
```

```
## Warning: package 'XML' was built under R version 4.4.2
```

```
library(httr)
```

```
## Warning: package 'httr' was built under R version 4.4.2
```

```
fileUrl <- "https://www.w3schools.com/xml/simple.xml"
doc <- xmlTreeParse(rawToChar(GET(fileUrl)$content),useInternalNodes = TRUE) # read file into R to pars
rootNode <- xmlRoot(doc) # wrapper for structured element

# Access Data
xmlName(rootNode) # name of xml
```

```
## [1] "breakfast_menu"
```

```
names(rootNode) # names of root node
```

```
## food food food food food  
## "food" "food" "food" "food" "food"
```

```
rootNode[[1]] # first food element, get first subelement of element using [[1]][[1]]
```

```
## <food>  
## <name>Belgian Waffles</name>  
## <price>$5.95</price>  
## <description>Two of our famous Belgian Waffles with plenty of real maple syrup</description>  
## <calories>650</calories>  
## </food>
```

```
xmlSApply(rootNode,xmlValue) # give xml value of each root node
```

```
##  
## "Belgian Waffles$5.95Two of our famous Belgian Waffles with plenty of 1  
##  
## "Strawberry Belgian Waffles$7.95Light Belgian waffles covered with strawberries and  
##  
## "Berry-Berry Belgian Waffles$8.95Light Belgian waffles covered with an assortment of fresh berries and  
##  
## "French Toast$4.50Thick slices made from our homemade  
##  
## "Homestyle Breakfast$6.95Two eggs, bacon or sausage, toast, and our ever-popular
```

```
xpathSApply(rootNode,"//name",xmlValue) # get node xmlValue for each names
```

```
## [1] "Belgian Waffles" "Strawberry Belgian Waffles"  
## [3] "Berry-Berry Belgian Waffles" "French Toast"  
## [5] "Homestyle Breakfast"
```

```
xpathSApply(rootNode,"//price",xmlValue)
```

```
## [1] "$5.95" "$7.95" "$8.95" "$4.50" "$6.95"
```

```
# Pokemon  
library(rvest)
```

```
## Warning: package 'rvest' was built under R version 4.4.2
```

```
fileUrlBase <- "https://pokemondb.net/pokedex/stats/gen"  
generations <- 1:8  
# Scrape Data  
pokemonList <- lapply(generations,function(x){  
  html <- read_html(paste0(fileUrlBase,x)) # create URL  
  tableNode <- html_node(html,xpath="//*[(@id = 'pokedex')]") # read HTML, extract node using xpath sel  
  data <- html_table(html,header=TRUE)[[1]] # convert to data frame
```

```

colnames(data)[1] <- "ID"
data$Generation <- x # add generation ID
data # return data frame to list
})
thePokemon <- do.call(rbind,pokemonList) # one data frame
# Cleaning
types <- strsplit(gsub('[a-z]\\K(?=[A-Z])', ' ', thePokemon$Type, perl=T), " ") # split Type column
thePokemon$Type1 <- unlist(lapply(types,function(x) x[1])) # Type1 column
thePokemon$Type2 <- unlist(lapply(types,function(x) ifelse(is.na(x[2])," ",x[2]))) # Type2 column
pokemonNames <- strsplit(gsub('[a-z]\\K(?=[A-Z])', '--', thePokemon$Name, perl=T), "--") # split Name
thePokemon$Name <- unlist(lapply(pokemonNames,function(x) x[1]))
thePokemon$Form <- unlist(lapply(pokemonNames,function(x) ifelse(is.na(x[2])," ",x[2])))
thePokemon <- thePokemon[,c(1,2,14,12,13,4,5,6,7,8,9,10,11)] # reorder and remove unnecessary
thePokemon[29,"Name"] <- "Nidoran" # cleaning gender
thePokemon[29,"Form"] <- "Female"
thePokemon[32,"Name"] <- "Nidoran"
thePokemon[32,"Form"] <- "Male"

```

Just JSON Stuff

- Javascript Object Notation- light weight data storage, common format from APIs, similar to XML but different syntax/format.
- Data stored as: number(double), strings(double quoted), boolean, array(ordered, comma sep, [] enclosed), objects(unordered, comma sep, {} enclosed, key:value pairs).
- <http://en.wikipedia.org/wiki/JSON>
- <http://www.json.org/>
- jsonlite vignette
- <http://www.r-bloggers.com/new-package-jsonlite-a-smarter-json-encoderdecoder/>

```
library(jsonlite)
```

```
## Warning: package 'jsonlite' was built under R version 4.4.2
```

```
jsonData <- fromJSON("https://api.github.com/users/jtleek/repos") # can pass any format
names(jsonData) # names of data frame
```

```
## [1] "id" "node_id"
## [3] "name" "full_name"
## [5] "private" "owner"
## [7] "html_url" "description"
## [9] "fork" "url"
## [11] "forks_url" "keys_url"
## [13] "collaborators_url" "teams_url"
## [15] "hooks_url" "issue_events_url"
## [17] "events_url" "assignees_url"
## [19] "branches_url" "tags_url"
```



```
## [21] "blobs_url"           "git_tags_url"
## [23] "git_refs_url"        "trees_url"
## [25] "statuses_url"        "languages_url"
## [27] "stargazers_url"      "contributors_url"
## [29] "subscribers_url"     "subscription_url"
## [31] "commits_url"         "git_commits_url"
## [33] "comments_url"        "issue_comment_url"
## [35] "contents_url"        "compare_url"
## [37] "merges_url"          "archive_url"
## [39] "downloads_url"       "issues_url"
## [41] "pulls_url"           "milestones_url"
## [43] "notifications_url"   "labels_url"
## [45] "releases_url"        "deployments_url"
## [47] "created_at"          "updated_at"
## [49] "pushed_at"           "git_url"
## [51] "ssh_url"             "clone_url"
## [53] "svn_url"             "homepage"
## [55] "size"                "stargazers_count"
## [57] "watchers_count"      "language"
## [59] "has_issues"          "has_projects"
## [61] "has_downloads"       "has_wiki"
## [63] "has_pages"           "has_discussions"
## [65] "forks_count"         "mirror_url"
## [67] "archived"            "disabled"
## [69] "open_issues_count"   "license"
## [71] "allow_forking"        "is_template"
## [73] "web_commit_signoff_required" "topics"
## [75] "visibility"          "forks"
## [77] "open_issues"         "watchers"
## [79] "default_branch"
```

```
names(jsonData$owner) # names of column
```

```
## [1] "login"           "id"              "node_id"
## [4] "avatar_url"      "gravatar_id"     "url"
## [7] "html_url"        "followers_url"   "following_url"
## [10] "gists_url"       "starred_url"     "subscriptions_url"
## [13] "organizations_url" "repos_url"       "events_url"
## [16] "received_events_url" "type"            "user_view_type"
## [19] "site_admin"
```

```
jsonData$owner$login # who logged in
```

```
## [1] "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek"
## [9] "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek"
## [17] "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek"
## [25] "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek"
```

```
# write data frame to JSON
myjson <- toJSON(iris, pretty = TRUE)
cat(myjson)
```

```

## [
##   {
##     "Sepal.Length": 5.1,
##     "Sepal.Width": 3.5,
##     "Petal.Length": 1.4,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
##   },
##   {
##     "Sepal.Length": 4.9,
##     "Sepal.Width": 3,
##     "Petal.Length": 1.4,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
##   },
##   {
##     "Sepal.Length": 4.7,
##     "Sepal.Width": 3.2,
##     "Petal.Length": 1.3,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
##   },
##   {
##     "Sepal.Length": 4.6,
##     "Sepal.Width": 3.1,
##     "Petal.Length": 1.5,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
##   },
##   {
##     "Sepal.Length": 5,
##     "Sepal.Width": 3.6,
##     "Petal.Length": 1.4,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
##   },
##   {
##     "Sepal.Length": 5.4,
##     "Sepal.Width": 3.9,
##     "Petal.Length": 1.7,
##     "Petal.Width": 0.4,
##     "Species": "setosa"
##   },
##   {
##     "Sepal.Length": 4.6,
##     "Sepal.Width": 3.4,
##     "Petal.Length": 1.4,
##     "Petal.Width": 0.3,
##     "Species": "setosa"
##   },
##   {
##     "Sepal.Length": 5,
##     "Sepal.Width": 3.4,
##     "Petal.Length": 1.5,

```

```

##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 4.4,
##      "Sepal.Width": 2.9,
##      "Petal.Length": 1.4,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 4.9,
##      "Sepal.Width": 3.1,
##      "Petal.Length": 1.5,
##      "Petal.Width": 0.1,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5.4,
##      "Sepal.Width": 3.7,
##      "Petal.Length": 1.5,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 4.8,
##      "Sepal.Width": 3.4,
##      "Petal.Length": 1.6,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 4.8,
##      "Sepal.Width": 3,
##      "Petal.Length": 1.4,
##      "Petal.Width": 0.1,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 4.3,
##      "Sepal.Width": 3,
##      "Petal.Length": 1.1,
##      "Petal.Width": 0.1,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5.8,
##      "Sepal.Width": 4,
##      "Petal.Length": 1.2,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5.7,

```

```

##      "Sepal.Width": 4.4,
##      "Petal.Length": 1.5,
##      "Petal.Width": 0.4,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5.4,
##      "Sepal.Width": 3.9,
##      "Petal.Length": 1.3,
##      "Petal.Width": 0.4,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5.1,
##      "Sepal.Width": 3.5,
##      "Petal.Length": 1.4,
##      "Petal.Width": 0.3,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5.7,
##      "Sepal.Width": 3.8,
##      "Petal.Length": 1.7,
##      "Petal.Width": 0.3,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5.1,
##      "Sepal.Width": 3.8,
##      "Petal.Length": 1.5,
##      "Petal.Width": 0.3,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5.4,
##      "Sepal.Width": 3.4,
##      "Petal.Length": 1.7,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5.1,
##      "Sepal.Width": 3.7,
##      "Petal.Length": 1.5,
##      "Petal.Width": 0.4,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 4.6,
##      "Sepal.Width": 3.6,
##      "Petal.Length": 1,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##  },

```

```

## {
##   "Sepal.Length": 5.1,
##   "Sepal.Width": 3.3,
##   "Petal.Length": 1.7,
##   "Petal.Width": 0.5,
##   "Species": "setosa"
## },
## {
##   "Sepal.Length": 4.8,
##   "Sepal.Width": 3.4,
##   "Petal.Length": 1.9,
##   "Petal.Width": 0.2,
##   "Species": "setosa"
## },
## {
##   "Sepal.Length": 5,
##   "Sepal.Width": 3,
##   "Petal.Length": 1.6,
##   "Petal.Width": 0.2,
##   "Species": "setosa"
## },
## {
##   "Sepal.Length": 5,
##   "Sepal.Width": 3.4,
##   "Petal.Length": 1.6,
##   "Petal.Width": 0.4,
##   "Species": "setosa"
## },
## {
##   "Sepal.Length": 5.2,
##   "Sepal.Width": 3.5,
##   "Petal.Length": 1.5,
##   "Petal.Width": 0.2,
##   "Species": "setosa"
## },
## {
##   "Sepal.Length": 5.2,
##   "Sepal.Width": 3.4,
##   "Petal.Length": 1.4,
##   "Petal.Width": 0.2,
##   "Species": "setosa"
## },
## {
##   "Sepal.Length": 4.7,
##   "Sepal.Width": 3.2,
##   "Petal.Length": 1.6,
##   "Petal.Width": 0.2,
##   "Species": "setosa"
## },
## {
##   "Sepal.Length": 4.8,
##   "Sepal.Width": 3.1,
##   "Petal.Length": 1.6,
##   "Petal.Width": 0.2,

```

```

##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5.4,
##      "Sepal.Width": 3.4,
##      "Petal.Length": 1.5,
##      "Petal.Width": 0.4,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5.2,
##      "Sepal.Width": 4.1,
##      "Petal.Length": 1.5,
##      "Petal.Width": 0.1,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5.5,
##      "Sepal.Width": 4.2,
##      "Petal.Length": 1.4,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 4.9,
##      "Sepal.Width": 3.1,
##      "Petal.Length": 1.5,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5,
##      "Sepal.Width": 3.2,
##      "Petal.Length": 1.2,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5.5,
##      "Sepal.Width": 3.5,
##      "Petal.Length": 1.3,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 4.9,
##      "Sepal.Width": 3.6,
##      "Petal.Length": 1.4,
##      "Petal.Width": 0.1,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 4.4,
##      "Sepal.Width": 3,

```

```

##     "Petal.Length": 1.3,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 5.1,
##     "Sepal.Width": 3.4,
##     "Petal.Length": 1.5,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 5,
##     "Sepal.Width": 3.5,
##     "Petal.Length": 1.3,
##     "Petal.Width": 0.3,
##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 4.5,
##     "Sepal.Width": 2.3,
##     "Petal.Length": 1.3,
##     "Petal.Width": 0.3,
##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 4.4,
##     "Sepal.Width": 3.2,
##     "Petal.Length": 1.3,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 5,
##     "Sepal.Width": 3.5,
##     "Petal.Length": 1.6,
##     "Petal.Width": 0.6,
##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 5.1,
##     "Sepal.Width": 3.8,
##     "Petal.Length": 1.9,
##     "Petal.Width": 0.4,
##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 4.8,
##     "Sepal.Width": 3,
##     "Petal.Length": 1.4,
##     "Petal.Width": 0.3,
##     "Species": "setosa"
## },
## {

```

```

##      "Sepal.Length": 5.1,
##      "Sepal.Width": 3.8,
##      "Petal.Length": 1.6,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 4.6,
##      "Sepal.Width": 3.2,
##      "Petal.Length": 1.4,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5.3,
##      "Sepal.Width": 3.7,
##      "Petal.Length": 1.5,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5,
##      "Sepal.Width": 3.3,
##      "Petal.Length": 1.4,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 7,
##      "Sepal.Width": 3.2,
##      "Petal.Length": 4.7,
##      "Petal.Width": 1.4,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6.4,
##      "Sepal.Width": 3.2,
##      "Petal.Length": 4.5,
##      "Petal.Width": 1.5,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6.9,
##      "Sepal.Width": 3.1,
##      "Petal.Length": 4.9,
##      "Petal.Width": 1.5,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.5,
##      "Sepal.Width": 2.3,
##      "Petal.Length": 4,
##      "Petal.Width": 1.3,
##      "Species": "versicolor"

```



```

## },
## {
##     "Sepal.Length": 6.5,
##     "Sepal.Width": 2.8,
##     "Petal.Length": 4.6,
##     "Petal.Width": 1.5,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 5.7,
##     "Sepal.Width": 2.8,
##     "Petal.Length": 4.5,
##     "Petal.Width": 1.3,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 6.3,
##     "Sepal.Width": 3.3,
##     "Petal.Length": 4.7,
##     "Petal.Width": 1.6,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 4.9,
##     "Sepal.Width": 2.4,
##     "Petal.Length": 3.3,
##     "Petal.Width": 1,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 6.6,
##     "Sepal.Width": 2.9,
##     "Petal.Length": 4.6,
##     "Petal.Width": 1.3,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 5.2,
##     "Sepal.Width": 2.7,
##     "Petal.Length": 3.9,
##     "Petal.Width": 1.4,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 5,
##     "Sepal.Width": 2,
##     "Petal.Length": 3.5,
##     "Petal.Width": 1,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 5.9,
##     "Sepal.Width": 3,
##     "Petal.Length": 4.2,

```

```

##      "Petal.Width": 1.5,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6,
##      "Sepal.Width": 2.2,
##      "Petal.Length": 4,
##      "Petal.Width": 1,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6.1,
##      "Sepal.Width": 2.9,
##      "Petal.Length": 4.7,
##      "Petal.Width": 1.4,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.6,
##      "Sepal.Width": 2.9,
##      "Petal.Length": 3.6,
##      "Petal.Width": 1.3,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6.7,
##      "Sepal.Width": 3.1,
##      "Petal.Length": 4.4,
##      "Petal.Width": 1.4,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.6,
##      "Sepal.Width": 3,
##      "Petal.Length": 4.5,
##      "Petal.Width": 1.5,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.8,
##      "Sepal.Width": 2.7,
##      "Petal.Length": 4.1,
##      "Petal.Width": 1,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6.2,
##      "Sepal.Width": 2.2,
##      "Petal.Length": 4.5,
##      "Petal.Width": 1.5,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.6,

```

```

##     "Sepal.Width": 2.5,
##     "Petal.Length": 3.9,
##     "Petal.Width": 1.1,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 5.9,
##     "Sepal.Width": 3.2,
##     "Petal.Length": 4.8,
##     "Petal.Width": 1.8,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 6.1,
##     "Sepal.Width": 2.8,
##     "Petal.Length": 4,
##     "Petal.Width": 1.3,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 6.3,
##     "Sepal.Width": 2.5,
##     "Petal.Length": 4.9,
##     "Petal.Width": 1.5,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 6.1,
##     "Sepal.Width": 2.8,
##     "Petal.Length": 4.7,
##     "Petal.Width": 1.2,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 6.4,
##     "Sepal.Width": 2.9,
##     "Petal.Length": 4.3,
##     "Petal.Width": 1.3,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 6.6,
##     "Sepal.Width": 3,
##     "Petal.Length": 4.4,
##     "Petal.Width": 1.4,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 6.8,
##     "Sepal.Width": 2.8,
##     "Petal.Length": 4.8,
##     "Petal.Width": 1.4,
##     "Species": "versicolor"
## },

```

```

## {
##   "Sepal.Length": 6.7,
##   "Sepal.Width": 3,
##   "Petal.Length": 5,
##   "Petal.Width": 1.7,
##   "Species": "versicolor"
## },
## {
##   "Sepal.Length": 6,
##   "Sepal.Width": 2.9,
##   "Petal.Length": 4.5,
##   "Petal.Width": 1.5,
##   "Species": "versicolor"
## },
## {
##   "Sepal.Length": 5.7,
##   "Sepal.Width": 2.6,
##   "Petal.Length": 3.5,
##   "Petal.Width": 1,
##   "Species": "versicolor"
## },
## {
##   "Sepal.Length": 5.5,
##   "Sepal.Width": 2.4,
##   "Petal.Length": 3.8,
##   "Petal.Width": 1.1,
##   "Species": "versicolor"
## },
## {
##   "Sepal.Length": 5.5,
##   "Sepal.Width": 2.4,
##   "Petal.Length": 3.7,
##   "Petal.Width": 1,
##   "Species": "versicolor"
## },
## {
##   "Sepal.Length": 5.8,
##   "Sepal.Width": 2.7,
##   "Petal.Length": 3.9,
##   "Petal.Width": 1.2,
##   "Species": "versicolor"
## },
## {
##   "Sepal.Length": 6,
##   "Sepal.Width": 2.7,
##   "Petal.Length": 5.1,
##   "Petal.Width": 1.6,
##   "Species": "versicolor"
## },
## {
##   "Sepal.Length": 5.4,
##   "Sepal.Width": 3,
##   "Petal.Length": 4.5,
##   "Petal.Width": 1.5,

```

```

##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6,
##      "Sepal.Width": 3.4,
##      "Petal.Length": 4.5,
##      "Petal.Width": 1.6,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6.7,
##      "Sepal.Width": 3.1,
##      "Petal.Length": 4.7,
##      "Petal.Width": 1.5,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6.3,
##      "Sepal.Width": 2.3,
##      "Petal.Length": 4.4,
##      "Petal.Width": 1.3,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.6,
##      "Sepal.Width": 3,
##      "Petal.Length": 4.1,
##      "Petal.Width": 1.3,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.5,
##      "Sepal.Width": 2.5,
##      "Petal.Length": 4,
##      "Petal.Width": 1.3,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.5,
##      "Sepal.Width": 2.6,
##      "Petal.Length": 4.4,
##      "Petal.Width": 1.2,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6.1,
##      "Sepal.Width": 3,
##      "Petal.Length": 4.6,
##      "Petal.Width": 1.4,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.8,
##      "Sepal.Width": 2.6,

```

```

##      "Petal.Length": 4,
##      "Petal.Width": 1.2,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5,
##      "Sepal.Width": 2.3,
##      "Petal.Length": 3.3,
##      "Petal.Width": 1,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.6,
##      "Sepal.Width": 2.7,
##      "Petal.Length": 4.2,
##      "Petal.Width": 1.3,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.7,
##      "Sepal.Width": 3,
##      "Petal.Length": 4.2,
##      "Petal.Width": 1.2,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.7,
##      "Sepal.Width": 2.9,
##      "Petal.Length": 4.2,
##      "Petal.Width": 1.3,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6.2,
##      "Sepal.Width": 2.9,
##      "Petal.Length": 4.3,
##      "Petal.Width": 1.3,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.1,
##      "Sepal.Width": 2.5,
##      "Petal.Length": 3,
##      "Petal.Width": 1.1,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.7,
##      "Sepal.Width": 2.8,
##      "Petal.Length": 4.1,
##      "Petal.Width": 1.3,
##      "Species": "versicolor"
##    },
##    {

```

```

##      "Sepal.Length": 6.3,
##      "Sepal.Width": 3.3,
##      "Petal.Length": 6,
##      "Petal.Width": 2.5,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 5.8,
##      "Sepal.Width": 2.7,
##      "Petal.Length": 5.1,
##      "Petal.Width": 1.9,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 7.1,
##      "Sepal.Width": 3,
##      "Petal.Length": 5.9,
##      "Petal.Width": 2.1,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 6.3,
##      "Sepal.Width": 2.9,
##      "Petal.Length": 5.6,
##      "Petal.Width": 1.8,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 6.5,
##      "Sepal.Width": 3,
##      "Petal.Length": 5.8,
##      "Petal.Width": 2.2,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 7.6,
##      "Sepal.Width": 3,
##      "Petal.Length": 6.6,
##      "Petal.Width": 2.1,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 4.9,
##      "Sepal.Width": 2.5,
##      "Petal.Length": 4.5,
##      "Petal.Width": 1.7,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 7.3,
##      "Sepal.Width": 2.9,
##      "Petal.Length": 6.3,
##      "Petal.Width": 1.8,
##      "Species": "virginica"

```

```

## },
## {
##     "Sepal.Length": 6.7,
##     "Sepal.Width": 2.5,
##     "Petal.Length": 5.8,
##     "Petal.Width": 1.8,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 7.2,
##     "Sepal.Width": 3.6,
##     "Petal.Length": 6.1,
##     "Petal.Width": 2.5,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.5,
##     "Sepal.Width": 3.2,
##     "Petal.Length": 5.1,
##     "Petal.Width": 2,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.4,
##     "Sepal.Width": 2.7,
##     "Petal.Length": 5.3,
##     "Petal.Width": 1.9,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.8,
##     "Sepal.Width": 3,
##     "Petal.Length": 5.5,
##     "Petal.Width": 2.1,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 5.7,
##     "Sepal.Width": 2.5,
##     "Petal.Length": 5,
##     "Petal.Width": 2,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 5.8,
##     "Sepal.Width": 2.8,
##     "Petal.Length": 5.1,
##     "Petal.Width": 2.4,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.4,
##     "Sepal.Width": 3.2,
##     "Petal.Length": 5.3,

```



```

##      "Petal.Width": 2.3,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 6.5,
##      "Sepal.Width": 3,
##      "Petal.Length": 5.5,
##      "Petal.Width": 1.8,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 7.7,
##      "Sepal.Width": 3.8,
##      "Petal.Length": 6.7,
##      "Petal.Width": 2.2,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 7.7,
##      "Sepal.Width": 2.6,
##      "Petal.Length": 6.9,
##      "Petal.Width": 2.3,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 6,
##      "Sepal.Width": 2.2,
##      "Petal.Length": 5,
##      "Petal.Width": 1.5,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 6.9,
##      "Sepal.Width": 3.2,
##      "Petal.Length": 5.7,
##      "Petal.Width": 2.3,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 5.6,
##      "Sepal.Width": 2.8,
##      "Petal.Length": 4.9,
##      "Petal.Width": 2,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 7.7,
##      "Sepal.Width": 2.8,
##      "Petal.Length": 6.7,
##      "Petal.Width": 2,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 6.3,

```

```

##      "Sepal.Width": 2.7,
##      "Petal.Length": 4.9,
##      "Petal.Width": 1.8,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 6.7,
##      "Sepal.Width": 3.3,
##      "Petal.Length": 5.7,
##      "Petal.Width": 2.1,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 7.2,
##      "Sepal.Width": 3.2,
##      "Petal.Length": 6,
##      "Petal.Width": 1.8,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 6.2,
##      "Sepal.Width": 2.8,
##      "Petal.Length": 4.8,
##      "Petal.Width": 1.8,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 6.1,
##      "Sepal.Width": 3,
##      "Petal.Length": 4.9,
##      "Petal.Width": 1.8,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 6.4,
##      "Sepal.Width": 2.8,
##      "Petal.Length": 5.6,
##      "Petal.Width": 2.1,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 7.2,
##      "Sepal.Width": 3,
##      "Petal.Length": 5.8,
##      "Petal.Width": 1.6,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 7.4,
##      "Sepal.Width": 2.8,
##      "Petal.Length": 6.1,
##      "Petal.Width": 1.9,
##      "Species": "virginica"
##    },
##  },

```

```

## {
##   "Sepal.Length": 7.9,
##   "Sepal.Width": 3.8,
##   "Petal.Length": 6.4,
##   "Petal.Width": 2,
##   "Species": "virginica"
## },
## {
##   "Sepal.Length": 6.4,
##   "Sepal.Width": 2.8,
##   "Petal.Length": 5.6,
##   "Petal.Width": 2.2,
##   "Species": "virginica"
## },
## {
##   "Sepal.Length": 6.3,
##   "Sepal.Width": 2.8,
##   "Petal.Length": 5.1,
##   "Petal.Width": 1.5,
##   "Species": "virginica"
## },
## {
##   "Sepal.Length": 6.1,
##   "Sepal.Width": 2.6,
##   "Petal.Length": 5.6,
##   "Petal.Width": 1.4,
##   "Species": "virginica"
## },
## {
##   "Sepal.Length": 7.7,
##   "Sepal.Width": 3,
##   "Petal.Length": 6.1,
##   "Petal.Width": 2.3,
##   "Species": "virginica"
## },
## {
##   "Sepal.Length": 6.3,
##   "Sepal.Width": 3.4,
##   "Petal.Length": 5.6,
##   "Petal.Width": 2.4,
##   "Species": "virginica"
## },
## {
##   "Sepal.Length": 6.4,
##   "Sepal.Width": 3.1,
##   "Petal.Length": 5.5,
##   "Petal.Width": 1.8,
##   "Species": "virginica"
## },
## {
##   "Sepal.Length": 6,
##   "Sepal.Width": 3,
##   "Petal.Length": 4.8,
##   "Petal.Width": 1.8,

```

```

##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.9,
##     "Sepal.Width": 3.1,
##     "Petal.Length": 5.4,
##     "Petal.Width": 2.1,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.7,
##     "Sepal.Width": 3.1,
##     "Petal.Length": 5.6,
##     "Petal.Width": 2.4,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.9,
##     "Sepal.Width": 3.1,
##     "Petal.Length": 5.1,
##     "Petal.Width": 2.3,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 5.8,
##     "Sepal.Width": 2.7,
##     "Petal.Length": 5.1,
##     "Petal.Width": 1.9,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.8,
##     "Sepal.Width": 3.2,
##     "Petal.Length": 5.9,
##     "Petal.Width": 2.3,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.7,
##     "Sepal.Width": 3.3,
##     "Petal.Length": 5.7,
##     "Petal.Width": 2.5,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.7,
##     "Sepal.Width": 3,
##     "Petal.Length": 5.2,
##     "Petal.Width": 2.3,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.3,
##     "Sepal.Width": 2.5,

```

```
##      "Petal.Length": 5,
##      "Petal.Width": 1.9,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 6.5,
##      "Sepal.Width": 3,
##      "Petal.Length": 5.2,
##      "Petal.Width": 2,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 6.2,
##      "Sepal.Width": 3.4,
##      "Petal.Length": 5.4,
##      "Petal.Width": 2.3,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 5.9,
##      "Sepal.Width": 3,
##      "Petal.Length": 5.1,
##      "Petal.Width": 1.8,
##      "Species": "virginica"
##    }
##  ]
```

```
# upload again
iris2 <- fromJSON(myjson)
head(iris2)
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1           5.1         3.5         1.4         0.2  setosa
## 2           4.9         3.0         1.4         0.2  setosa
## 3           4.7         3.2         1.3         0.2  setosa
## 4           4.6         3.1         1.5         0.2  setosa
## 5           5.0         3.6         1.4         0.2  setosa
## 6           5.4         3.9         1.7         0.4  setosa
```

mySQL Stuff

- Free, widely-used open source database, esp in internet-based applications.
- Data structured in databases, inside is tables(dataset), inside is fields(column), row is record. IDs link databases.
- <http://en.wikipedia.org/wiki/MySQL> and <http://www.mysql.com/>
- Install MySQL: <https://dev.mysql.com/doc/refman/5.7/en/installing.html>
- <http://cran.r-project.org/web/packages/RMySQL/RMySQL.pdf>

- Commands: <http://www.pantz.org/software/mysql/mysqlcommands.html> and <http://www.r-bloggers.com/mysql-and-r>
- Be careful with SQL commands, *DO NOT PUSH UNLESS REQUIRED!*

```
library(RMySQL)
```

```
## Warning: package 'RMySQL' was built under R version 4.4.2
```

```
## Loading required package: DBI
```

```
## Warning: package 'DBI' was built under R version 4.4.2
```

```
ucscDb <- dbConnect(MySQL(), user="genome",host="genome-mysql.cse.ucsc.edu") # open a connection to the
result <- dbGetQuery(ucscDb, "show databases;"); dbDisconnect(ucscDb); # run the mySQL command "show da
```

```
## [1] TRUE
```

```
result # print available databases
```

```
##          Database
## 1          acaChl1
## 2          ailMel1
## 3          allMis1
## 4          allSin1
## 5          amaVit1
## 6          anaPla1
## 7          ancCey1
## 8          angJap1
## 9          anoCar1
## 10         anoCar2
## 11         anoGam1
## 12         anoGam3
## 13         apaSpi1
## 14         apaVit1
## 15         apiMel1
## 16         apiMel2
## 17  apiMel2_justpushed
## 18         aplCal1
## 19         aptFor1
## 20         aptMan1
## 21         aquChr2
## 22         araMac1
## 23         ascSuu1
## 24         balAcu1
## 25         balPav1
## 26         bisBis1
## 27         bosTau2
## 28         bosTau3
## 29         bosTau4
## 30         bosTau5
```

## 31	bosTau6
## 32	bosTau7
## 33	bosTau8
## 34	bosTau9
## 35	bosTauMd3
## 36	braFlo1
## 37	bruMal2
## 38	bucRhi1
## 39	burXyl1
## 40	caeAng2
## 41	caeJap1
## 42	caeJap4
## 43	caePb1
## 44	caePb2
## 45	caePb3
## 46	caeRem2
## 47	caeRem3
## 48	caeRem4
## 49	caeSp111
## 50	caeSp51
## 51	calAnn1
## 52	calJac1
## 53	calJac3
## 54	calJac4
## 55	calMil1
## 56	canFam1
## 57	canFam2
## 58	canFam3
## 59	canFam4
## 60	canFam5
## 61	canFam6
## 62	capCar1
## 63	carCri1
## 64	cavPor3
## 65	cb1
## 66	cb3
## 67	cb4
## 68	ce10
## 69	ce11
## 70	ce2
## 71	ce4
## 72	ce6
## 73	cerSim1
## 74	chaVoc2
## 75	cheMyd1
## 76	chlSab2
## 77	chlUnd1
## 78	choHof1
## 79	chrPic1
## 80	chrPic2
## 81	ci1
## 82	ci2
## 83	ci3
## 84	colLiv1

## 85	colStr1
## 86	corBra1
## 87	corCor1
## 88	cotJap2
## 89	criGri1
## 90	criGriChoV1
## 91	criGriChoV2
## 92	cucCan1
## 93	danRer1
## 94	danRer10
## 95	danRer11
## 96	danRer2
## 97	danRer3
## 98	danRer4
## 99	danRer5
## 100	danRer6
## 101	danRer7
## 102	dasNov3
## 103	dipOrd1
## 104	dirImm1
## 105	dm1
## 106	dm2
## 107	dm3
## 108	dm6
## 109	dp2
## 110	dp3
## 111	droAna1
## 112	droAna2
## 113	droEre1
## 114	droGri1
## 115	droMoj1
## 116	droMoj2
## 117	droPer1
## 118	droSec1
## 119	droSim1
## 120	droSim2
## 121	droVir1
## 122	droVir2
## 123	droYak1
## 124	droYak2
## 125	eboVir3
## 126	echTel1
## 127	echTel2
## 128	egrGar1
## 129	enhLutNer1
## 130	equCab1
## 131	equCab2
## 132	equCab3
## 133	eriEur1
## 134	eriEur2
## 135	eurHel1
## 136	falChe1
## 137	falPer1
## 138	felCat3


```

## 139         felCat4
## 140         felCat5
## 141         felCat8
## 142         felCat9
## 143         ficAlb2
## 144             fr1
## 145             fr2
## 146             fr3
## 147         fulGla1
## 148         gadMor1
## 149         galGal2
## 150         galGal3
## 151         galGal4
## 152         galGal5
## 153         galGal6
## 154         galVar1
## 155         gasAcu1
## 156         gavSte1
## 157         gbMeta
## 158         geoFor1
## 159             go
## 160         go080130
## 161         go140213
## 162         go150121
## 163         go180426
## 164         gorGor3
## 165         gorGor4
## 166         gorGor5
## 167         gorGor6
## 168         haeCon2
## 169         halAlb1
## 170         halLeu1
## 171         hetBac1
## 172         hetGla1
## 173         hetGla2
## 174             hg16
## 175             hg17
## 176             hg18
## 177             hg19
## 178         hg19Patch10
## 179         hg19Patch13
## 180     hg19_justpushed
## 181             hg38
## 182         hg38Patch11
## 183     hg38_justpushed
## 184             hgFixed
## 185         hgcentral
## 186             hs1
## 187     information_schema
## 188     knownGeneKentHg19
## 189         knownGeneV39
## 190         knownGeneV43
## 191         knownGeneV44
## 192         knownGeneV45

```

## 193	knownGeneV46
## 194	knownGeneVM27
## 195	knownGeneVM30
## 196	knownGeneVM32
## 197	latCha1
## 198	lepDis1
## 199	letCam1
## 200	loaLoa1
## 201	loxAfr3
## 202	macEug1
## 203	macEug2
## 204	macFas5
## 205	manPen1
## 206	melGal1
## 207	melGal5
## 208	melHap1
## 209	melInc2
## 210	melUnd1
## 211	merNub1
## 212	mesUni1
## 213	micMur1
## 214	micMur2
## 215	mm10
## 216	mm39
## 217	mm39_justpushed
## 218	mm5
## 219	mm6
## 220	mm7
## 221	mm8
## 222	mm9
## 223	monDom1
## 224	monDom4
## 225	monDom5
## 226	mpxvRivers
## 227	musFur1
## 228	myoLuc2
## 229	nanPar1
## 230	nasLar1
## 231	necAme1
## 232	neoSch1
## 233	nipNip1
## 234	nomLeu1
## 235	nomLeu2
## 236	nomLeu3
## 237	ochPri2
## 238	ochPri3
## 239	oncVol1
## 240	opiHoa1
## 241	oreNil1
## 242	oreNil2
## 243	oreNil3
## 244	ornAna1
## 245	ornAna2
## 246	oryCun2

```

## 247         oryLat2
## 248         otoGar3
## 249         oviAri1
## 250         oviAri3
## 251         oviAri4
## 252         panPan1
## 253         panPan2
## 254         panPan3
## 255         panRed1
## 256         panTro1
## 257         panTro2
## 258         panTro3
## 259         panTro4
## 260         panTro5
## 261         panTro6
## 262         papAnu2
## 263         papAnu4
## 264         papHam1
## 265         pelCri1
## 266         pelSin1
## 267 performance_schema
## 268         petMar1
## 269         petMar2
## 270         petMar3
## 271         phaCar1
## 272         phaLep1
## 273         phoRub1
## 274         picPub1
## 275         ponAbe2
## 276         ponAbe3
## 277         priExs1
## 278         priPac1
## 279         priPac3
## 280         proCap1
## 281         proteins120806
## 282         proteins121210
## 283         proteins140122
## 284         proteins150225
## 285         proteins160229
## 286         proteins180404
## 287         proteome
## 288         pteGut1
## 289         pteVam1
## 290         pygAde1
## 291         pytBiv1
## 292         rheMac1
## 293         rheMac10
## 294         rheMac2
## 295         rheMac3
## 296         rheMac8
## 297         rhiRox1
## 298         rn3
## 299         rn4
## 300         rn5

```

## 301	rn6
## 302	rn7
## 303	sacCer1
## 304	sacCer2
## 305	sacCer3
## 306	saiBol1
## 307	sarHar1
## 308	serCan1
## 309	sorAra1
## 310	sorAra2
## 311	sp120323
## 312	sp121210
## 313	sp140122
## 314	sp150225
## 315	sp160229
## 316	sp180404
## 317	speTri2
## 318	strCam1
## 319	strPur1
## 320	strPur2
## 321	strRat2
## 322	susScr11
## 323	susScr2
## 324	susScr3
## 325	taeGut1
## 326	taeGut2
## 327	tarSyr1
## 328	tarSyr2
## 329	tauEry1
## 330	tetNig1
## 331	tetNig2
## 332	thaSir1
## 333	tinGut2
## 334	triMan1
## 335	triSpi1
## 336	triSui1
## 337	tupBel1
## 338	turTru2
## 339	tytAlb1
## 340	uniProt
## 341	vicPac1
## 342	vicPac2
## 343	visiGene
## 344	wuhCor1
## 345	xenLae2
## 346	xenTro1
## 347	xenTro10
## 348	xenTro2
## 349	xenTro3
## 350	xenTro7
## 351	xenTro9
## 352	zonAlb1

```
hg19 <- dbConnect(MySQL(), user="genome", db = "hg19", host="genome-mysql.cse.ucsc.edu") # connection to
allTables <- dbListTables(hg19) # extract names of all tables
length(allTables)
```

```
## [1] 12690
```

```
allTables[1:5]
```

```
## [1] "HInv" "HInvGeneMrna" "acembly" "acemblyClass" "acemblyPep"
```

```
dbListFields(hg19, "affyU133Plus2") # fields/columns of table affyU133Plus2 (is a microarray)
```

```
## [1] "bin" "matches" "misMatches" "repMatches" "nCount"
## [6] "qNumInsert" "qBaseInsert" "tNumInsert" "tBaseInsert" "strand"
## [11] "qName" "qSize" "qStart" "qEnd" "tName"
## [16] "tSize" "tStart" "tEnd" "blockCount" "blockSizes"
## [21] "qStarts" "tStarts"
```

```
dbGetQuery(hg19, "select count(*) from affyU133Plus2") # how many rows/record. Pass mySQL command to connect
```

```
## count(*)
## 1 58463
```

```
library(knitr) # to suppress warnings
```

```
## Warning: package 'knitr' was built under R version 4.4.2
```

```
suppressWarnings({ affyData <- dbReadTable(hg19, "affyU133Plus2") }) # import data from table in mySQL
head(affyData)
```

```
## bin matches misMatches repMatches nCount qNumInsert qBaseInsert tNumInsert
## 1 585 530 4 0 23 3 41 3
## 2 585 3355 17 0 109 9 67 9
## 3 585 4156 14 0 83 16 18 2
## 4 585 4667 9 0 68 21 42 3
## 5 585 5180 14 0 167 10 38 1
## 6 585 468 5 0 14 0 0 0
## tBaseInsert strand qName qSize qStart qEnd tName tSize tStart
## 1 898 - 225995_x_at 637 5 603 chr1 249250621 14361
## 2 11621 - 225035_x_at 3635 0 3548 chr1 249250621 14381
## 3 93 - 226340_x_at 4318 3 4274 chr1 249250621 14399
## 4 5743 - 1557034_s_at 4834 48 4834 chr1 249250621 14406
## 5 29 - 231811_at 5399 0 5399 chr1 249250621 19688
## 6 0 - 236841_at 487 0 487 chr1 249250621 27542
## tEnd blockCount
## 1 15816 5
## 2 29483 17
## 3 18745 18
## 4 24893 23
```

```
## 5 25078      11
## 6 28029      1
##
##                                blockSize
## 1                                93,144,229,70,21,
## 2                73,375,71,165,303,360,198,661,201,1,260,250,74,73,98,155,163,
## 3                690,10,32,33,376,4,5,15,5,11,7,41,277,859,141,51,443,1253,
## 4 99,352,286,24,49,14,6,5,8,149,14,44,98,12,10,355,837,59,8,1500,133,624,58,
## 5                                131,26,1300,6,4,11,4,7,358,3359,155,
## 6                                487,
##
##                                qSt
## 1                                34,132,278,541,0
## 2                87,165,540,647,818,1123,1484,1682,2343,2545,2546,2808,3058,3133,3206,3317,3
## 3                44,735,746,779,813,1190,1195,1201,1217,1223,1235,1243,1285,1564,2423,2565,2617,3
## 4 0,99,452,739,764,814,829,836,842,851,1001,1016,1061,1160,1173,1184,1540,2381,2441,2450,3951,4103,4
## 5                                0,132,159,1460,1467,1472,1484,1489,1497,1856,5
## 6
##
## 1
## 2                14381,14454,14969,15075,15240,15543,15903,16104,16853,17054,17
## 3                14399,15089,15099,15131,15164,15540,15544,15549,15564,15569,15580,15
## 4 14406,20227,20579,20865,20889,20938,20952,20958,20963,20971,21120,21134,21178,21276,21288,21298,21
## 5                                19688,19819,19845,21145,21
## 6
```

```
suppressWarnings({ query <- dbSendQuery(hg19, "select * from affyU133Plus2 where misMatches between 1 and 1000000")
affyMis <- fetch(query); quantile(affyMis$misMatches) # fetch subset from database and run quantile on it
```

```
##    0%   25%   50%   75%  100%
##    1     1     2     2     3
```

```
affyMisSmall <- fetch(query,n=10); dbClearResult(query);# fetch top 10 records
```

```
## [1] TRUE
```

```
dim(affyMisSmall)
```

```
## [1] 10 22
```

```
dbDisconnect(hg19) # CLOSE CONNECTION
```

```
## [1] TRUE
```

Just HDF5 Stuff

- Large data sets, store range of data types, Hierarchical Data Format.
- Groups of 0+ data sets and their metadata: group header (with name and list of attributes) and group symbol table (with list of objects in groups).

- Datasets are multi-dimensional array of data elements with their metadata: header (with name, datatype, dataspace, and storage layout) and data array (containing data).
- <http://www.hdfgroup.org/>
- Install: if (!require("BiocManager", quietly = TRUE)) install.packages("BiocManager"); BiocManager::install(version = "3.20"); BiocManager::install(pkgs=c("rhdf5"));
- <https://www.bioconductor.org/packages/release/bioc/html/rhdf5.html>

```
library(rhdf5)
```

```
## Warning: package 'rhdf5' was built under R version 4.4.2
```

```
file.remove("example.h5")
```

```
## [1] TRUE
```

```
created = h5createFile("example.h5") # create hdf5 file
created
```

```
## [1] TRUE
```

```
created = h5createGroup("example.h5", "foo") # create group with given name inside h5
created = h5createGroup("example.h5", "baa")
created = h5createGroup("example.h5", "foo/foobaa") # create subgroup
h5ls("example.h5") # list components of h5 file
```

```
##   group   name      otype dclass dim
## 0      /    baa H5I_GROUP
## 1      /    foo H5I_GROUP
## 2 /foo foobaa H5I_GROUP
```

```
A <- matrix(1:10, nrow=5, ncol=2)
h5write(A, "example.h5", "foo/A") # write object to group
B <- array(seq(0.1, 2.0, by=0.1), dim=c(5, 2, 2))
attr(B, "scale") <- "liter" # can include metadata
h5write(B, "example.h5", "foo/foobaa/B")
h5ls("example.h5")
```

```
##      group   name      otype dclass      dim
## 0      /    baa H5I_GROUP
## 1      /    foo H5I_GROUP
## 2    /foo      A H5I_DATASET INTEGER    5 x 2
## 3    /foo foobaa H5I_GROUP
## 4 /foo/foobaa  B H5I_DATASET  FLOAT 5 x 2 x 2
```

```
df <- data.frame(1L:5L, seq(0, 1, length.out=5), c("ab", "cde", "fghi", "a", "s"), stringsAsFactors = FALSE)
h5write(df, "example.h5", "df") # write data set to top level group
h5ls("example.h5")
```

```
##      group  name      otype  dclass      dim
## 0      /    baa    H5I_GROUP
## 1      /    df    H5I_DATASET COMPOUND      5
## 2      /    foo    H5I_GROUP
## 3    /foo      A    H5I_DATASET  INTEGER    5 x 2
## 4    /foo foobaa    H5I_GROUP
## 5 /foo/foobaa      B    H5I_DATASET    FLOAT 5 x 2 x 2
```

```
h5read("example.h5", "foo/A") # read data from hf file
```

```
##      [,1] [,2]
## [1,]    1    6
## [2,]    2    7
## [3,]    3    8
## [4,]    4    9
## [5,]    5   10
```

```
h5write(c(12,13,14), "example.h5", "foo/A", index=list(1:3,1)) # can read and write (in this case edit) in
h5read("example.h5", "foo/A")
```

```
##      [,1] [,2]
## [1,]   12    6
## [2,]   13    7
## [3,]   14    8
## [4,]    4    9
## [5,]    5   10
```

HTML and Webscraping

- Webscraping: programatically extracting data from HTML code of websites. Be careful of Terms of Service and speed of scraping...
- http://en.wikipedia.org/wiki/Web_scraping
- <http://www.r-bloggers.com/?s=Web+Scraping>
- <http://cran.r-project.org/web/packages/httr/httr.pdf>

```
con = url("https://scholar.google.com/citations?user=HI-I6C0AAAAJ&hl=en") # open connection to website
htmlCode = readLines(con) # extract file lines into a char vector
```

```
## Warning in readLines(con): incomplete final line found on
## 'https://scholar.google.com/citations?user=HI-I6C0AAAAJ&hl=en'
```

```
close(con) # REMEMBER
htmlCode[1] # Hard to read
```

```
## [1] "<!doctype html><html><head><title>Jeff Leek - Google Scholar</title><meta http-equiv=\"Content-"
```



```
# alternate method: XML
library(XML)
url <- "http://web.archive.org/web/20130207021632/http://scholar.google.com:80/citations?user=HI-I6C0AA"
html <- htmlTreeParse(url,useInternalNodes = TRUE) # parsing file using internal nodes
xpathSApply(html, "//title", xmlValue) # use sApply for clean vector. Accessing title
```

```
## [1] "Jeff Leek - Google Scholar Citations"
```

```
xpathSApply(html, "//td[@id='col-citedby']", xmlValue) # accessing elements of table
```

```
## [1] "Cited by" "339"      "173"      "140"      "133"      "107"
## [7] "95"       "79"       "79"       "54"       "17"       "10"
## [13] "9"        "8"        "8"        "8"        "8"        "6"
## [19] "6"        "5"        "3"
```

```
# alternate method: httr and GET
library(httr)
html2 <- GET("https://scholar.google.com/citations?user=HI-I6C0AAAAJ&hl=en")
content2 <- content(html2, as = "text") # extract content from HTML page as one text string
parsedHtml <- htmlParse(content2, asText = TRUE) # parse the text, same as xml package result
xpathSApply(parsedHtml, "//title", xmlValue)
```

```
## [1] "Jeff Leek - Google Scholar"
```

```
# passwords
url <- "http://httpbin.org/basic-auth/user/passwd"
pg1 <- GET(url)
pg1 # not authenticated, passes status 401
```

```
## Response [http://httpbin.org/basic-auth/user/passwd]
##   Date: 2025-01-28 00:44
##   Status: 401
##   Content-Type: <unknown>
## <EMPTY BODY>
```

```
pg2 <- GET(url, authenticate("user","passwd")) # sends authentication
pg2
```

```
## Response [http://httpbin.org/basic-auth/user/passwd]
##   Date: 2025-01-28 00:44
##   Status: 200
##   Content-Type: application/json
##   Size: 47 B
## {
##   "authenticated": true,
##   "user": "user"
## }
```

```
names(pg2)
```

```
## [1] "url"          "status_code" "headers"      "all_headers" "cookies"
## [6] "content"      "date"         "times"        "request"     "handle"
```

```
# use handles
```

```
google <- handle("http://google.com") # to save authentication across multiple accesses to website
pg1 <- GET(handle = google, path = "/")
pg2 <- GET(handle = google, path = "search")
```

APIs

- Application programming interfaces.
- Developed platform will have GET request URLs (and parameters), which you use in httr to get data. Need a API/dev account, need to submit a request (for each project!). Receive: consumer key, consumer secret, request token URL, and Authorization URL.
- Able to GET, POST, PUT, DELETE with httr if authorized.
- Need oauth in most cases, sometimes username/password allowed.
- Sites: Facebook, Google, Twitter, GitHub...

```
# in personal vault :)
```

```
rm(list=ls())
```

Other Sources

Packages for Data Storage Mechanisms - Search Google: “data storage mechanism R package”

- ?connections to get info on creating connection (CLOSE)
- foreign package: loads data from Minitab, S, SAS, SPSS, Stata, Systat. read.lang. <http://cran.r-project.org/web/packages/foreign/foreign.pdf>
- RPostgreSQL: <https://code.google.com/p/rpostgresql/>, <http://cran.r-project.org/web/packages/RPostgreSQL/RPostgreSQL.pdf>
- RODBC (PostgreSQL, MySQL, Microsoft Access, SQLite): <http://cran.r-project.org/web/packages/RODBC/vignettes/RODBC.pdf>, <http://cran.r-project.org/web/packages/RODBC/RODBC.pdf>
- RMongo/rmongodb: <http://cran.r-project.org/web/packages/RMongo/RMongo.pdf>, <http://www.r-bloggers.com/r-and-mongodb>

Image Data - jpeg: <http://cran.r-project.org/web/packages/jpeg/index.html>

- readbitmap: <http://cran.r-project.org/web/packages/readbitmap/index.html>

- png: <http://cran.r-project.org/web/packages/png/index.html>

- EBImage (Bioconductor): <http://www.bioconductor.org/packages/2.13/bioc/html/EBImage.html>

GIS Data - raster: <http://cran.r-project.org/web/packages/raster/index.html>

- ‘sf’ and ‘terra’

Musical Data - tuneR: <http://cran.r-project.org/web/packages/tuneR/>

- seewave: <http://rug.mnhn.fr/seewave>