INFS4203/7203 Data Mining
The University of Queensland, Australia
Semester 2, 2018

Tutorial Week 5:
Introduction to R for Data Mining

Chandra Prasetyo Utomo
c.utomo@uq.edu.au

Objectives

1. To gain familiarity with basic R syntax and RStudio IDE.
2. To learn some best practices of data mining project.
3. To be able to load external data into R data frame.
4. To tour some methods of exploring a data frame.

Outline

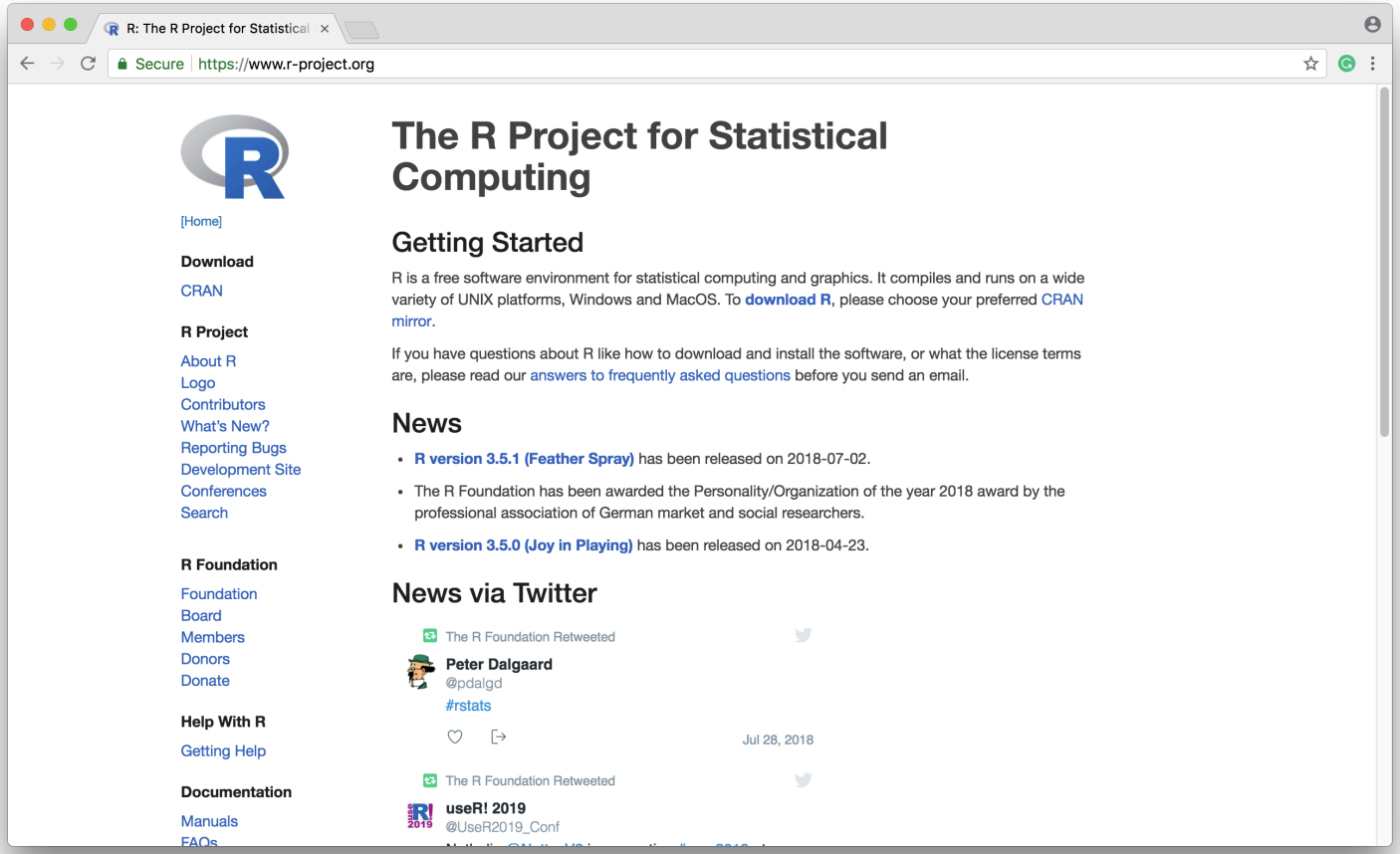
1. Installation and Introduction to RStudio (*10 minutes*)
2. Best Practice Data Mining Project (*10 minutes*)
3. Data Extraction (10 minutes)
4. Data Exploration (*20 minutes*)

Part 1

Installation & Introduction to RStudio

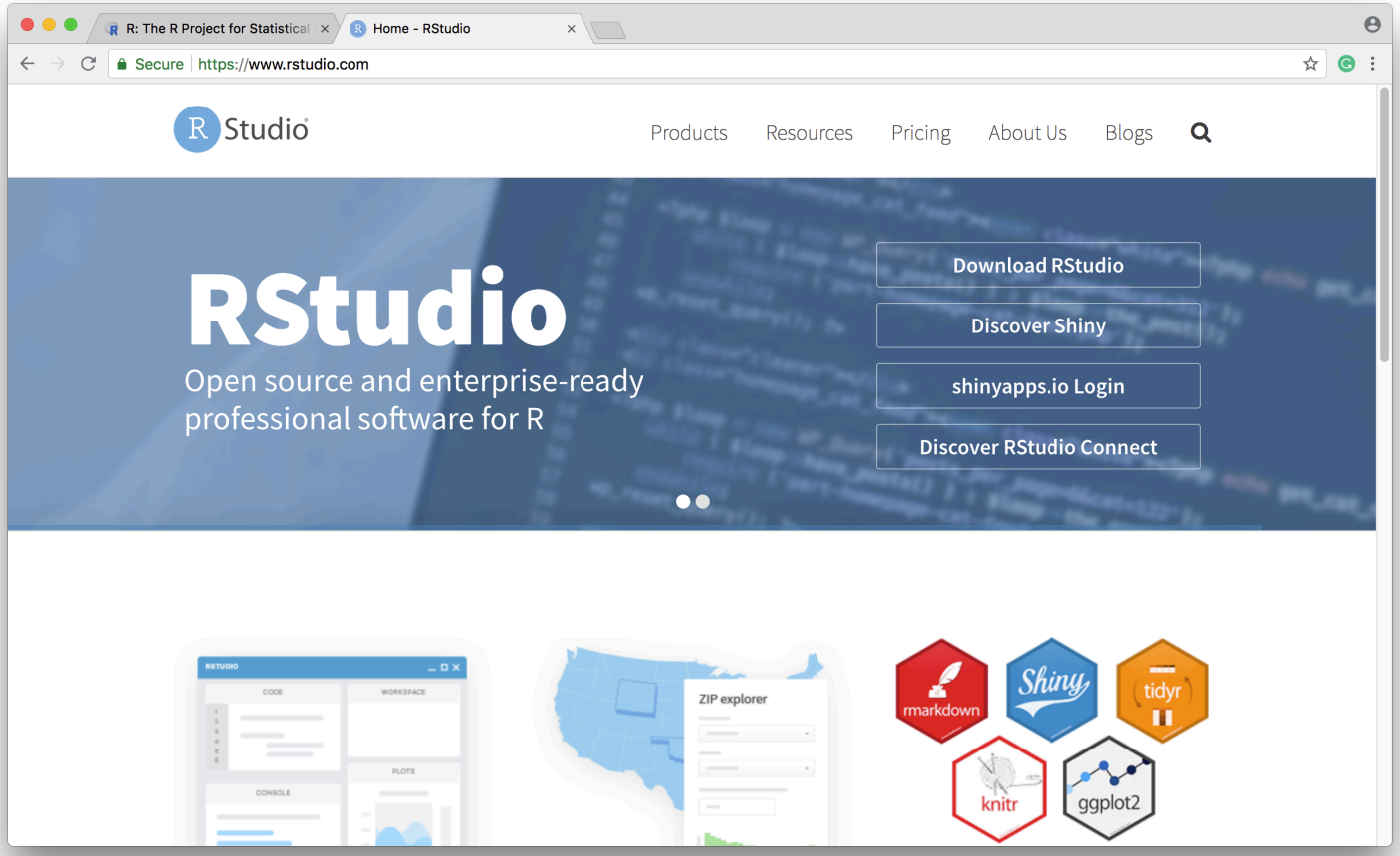
Install R

- https://www.r-project.org/

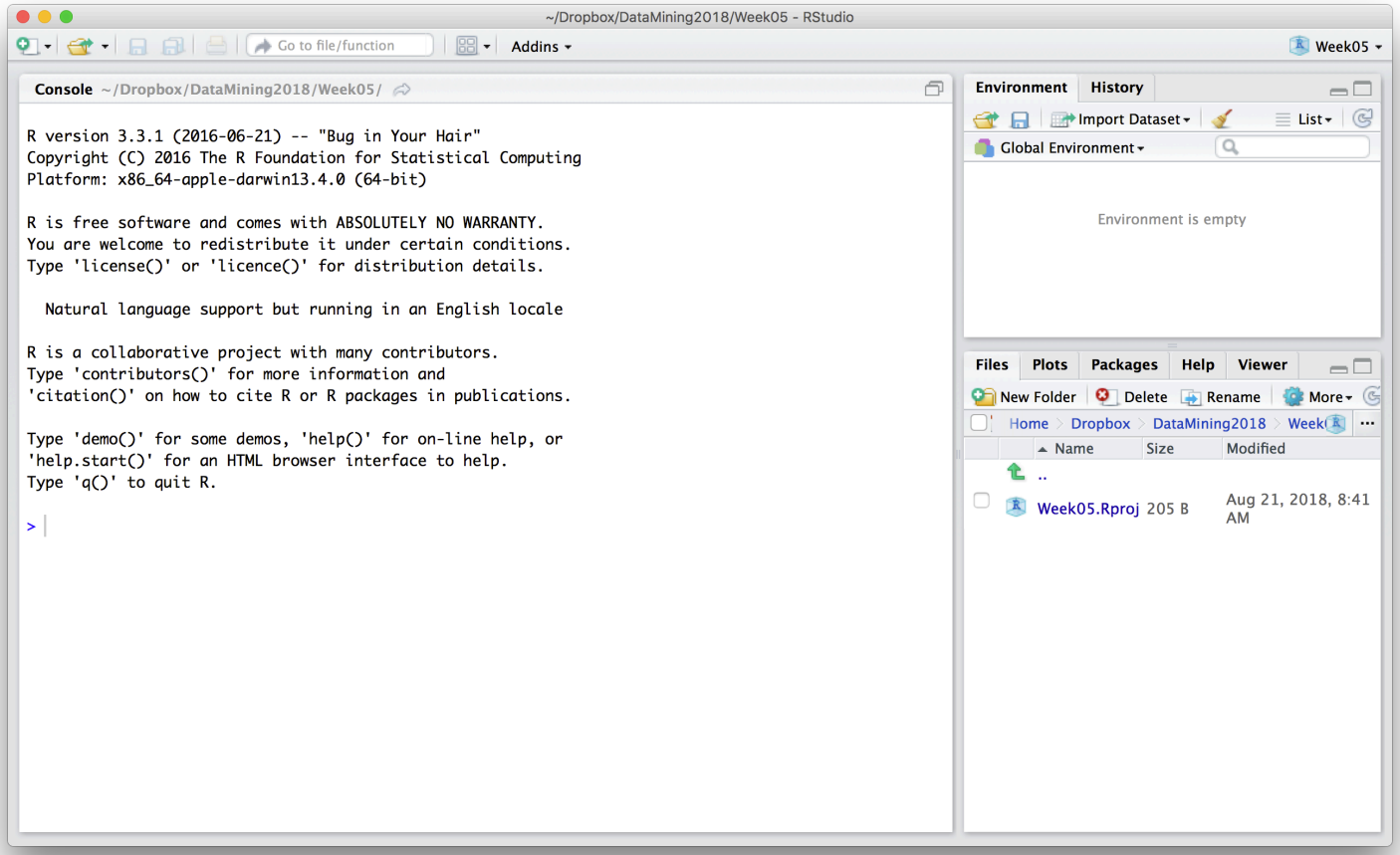


Install RStudio

- https://www.rstudio.com/



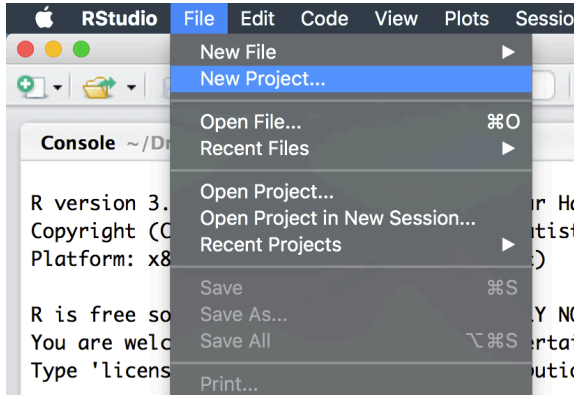
RStudio

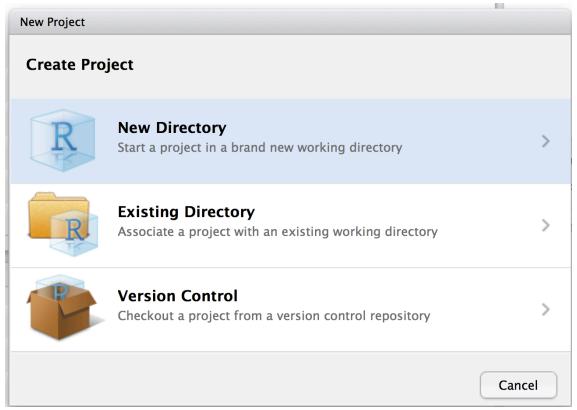

~/Dropbox/DataMining2018/Week05 - RStudio
Go to file/function
Addins
Week05
Console ~/Dropbox/DataMining2018/Week05/
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin13.4.0 (64-bit)
R is free software and comes 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.
>
Environment
History
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Home
Dropbox
DataMining2018
Week
Name
Size
Modified
..
Week05.Rproj
205 B
Aug 21, 2018, 8:41 AM

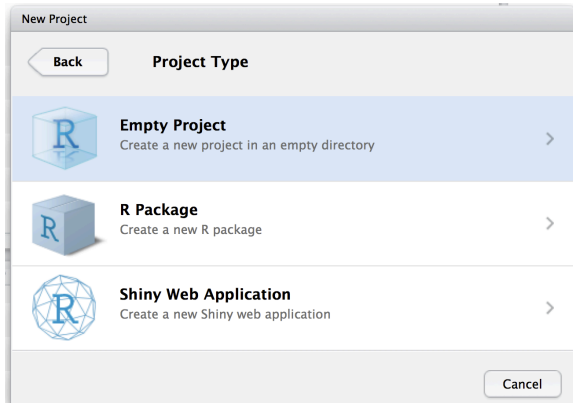
Part 2

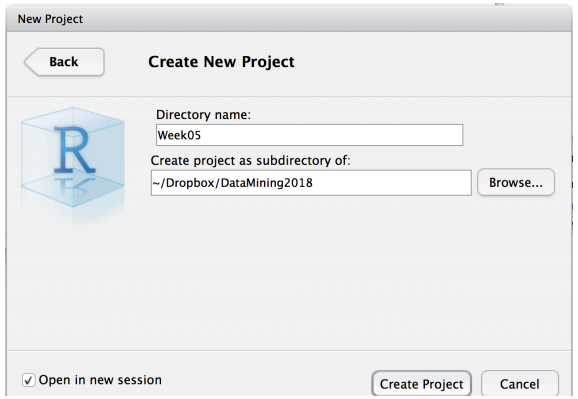
Best Practice Data Mining Project

New Project

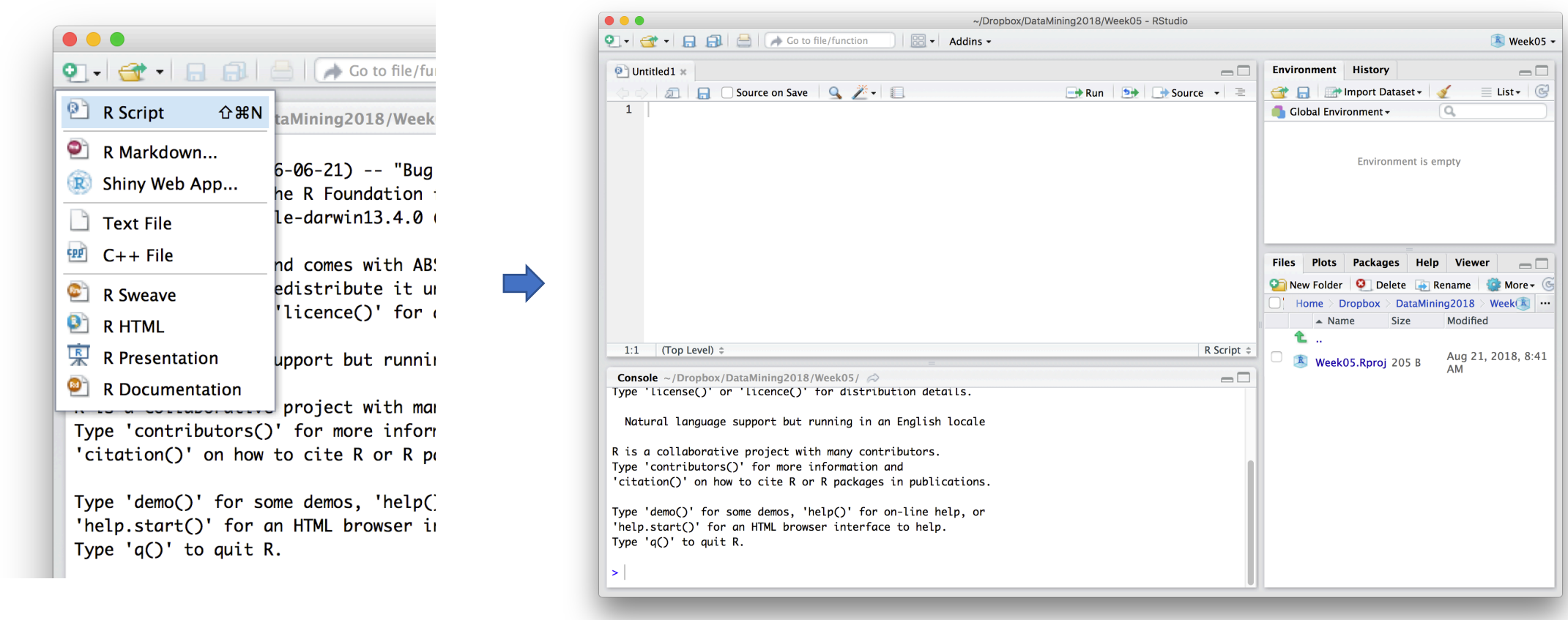

RStudio
File
Edit
Code
View
Plots
Session
New File
New Project...
Open File... ⌘O
Recent Files
Open Project...
Open Project in New Session...
Recent Projects
Save ⌘S
Save As...
Save All ⌥⌘S
Print...
Console


New Project
Create Project
New Directory
Start a project in a brand new working directory
Existing Directory
Associate a project with an existing working directory
Version Control
Checkout a project from a version control repository
Cancel

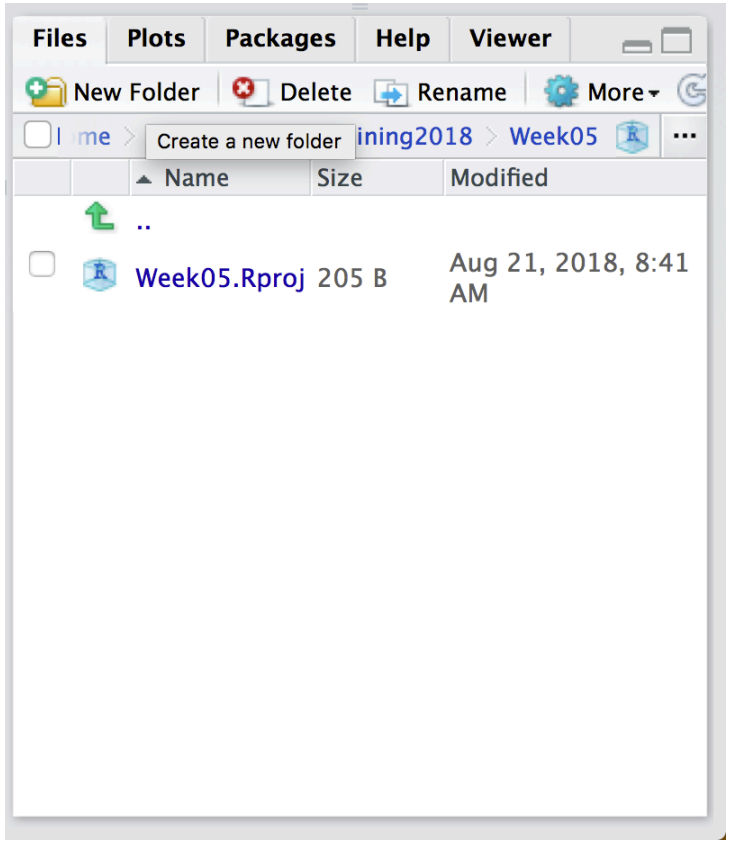

New Project
Back
Project Type
Empty Project
Create a new project in an empty directory
R Package
Create a new R package
Shiny Web Application
Create a new Shiny web application
Cancel

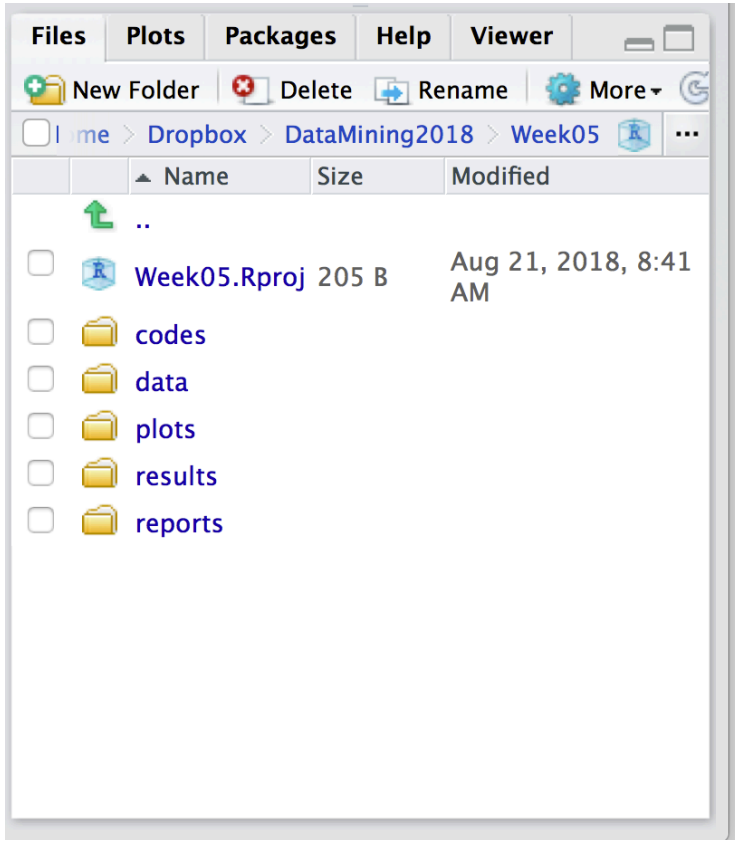

New Project
Back
Create New Project
Directory name:
Week05
Create project as subdirectory of:
~/Dropbox/DataMining2018
Browse...
Open in new session
Create Project
Cancel

R Script


R Script ⇧⌘N
R Markdown...
Shiny Web App...
Text File
C++ File
R Sweave
R HTML
R Presentation
R Documentation
Type 'contributors()' for more infor
'citation()' on how to cite R or R p
Type 'demo()' for some demos, 'help(
'help.start()' for an HTML browser i
Type 'q()' to quit R.
~/Dropbox/DataMining2018/Week05 - RStudio
Untitled1
Source on Save
Run
Source
Environment
History
Import Dataset
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Home
Dropbox
DataMining2018
Name
Size
Modified
Week05.Rproj 205 B
Aug 21, 2018, 8:41 AM
Console ~/Dropbox/DataMining2018/Week05/
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.

Folders

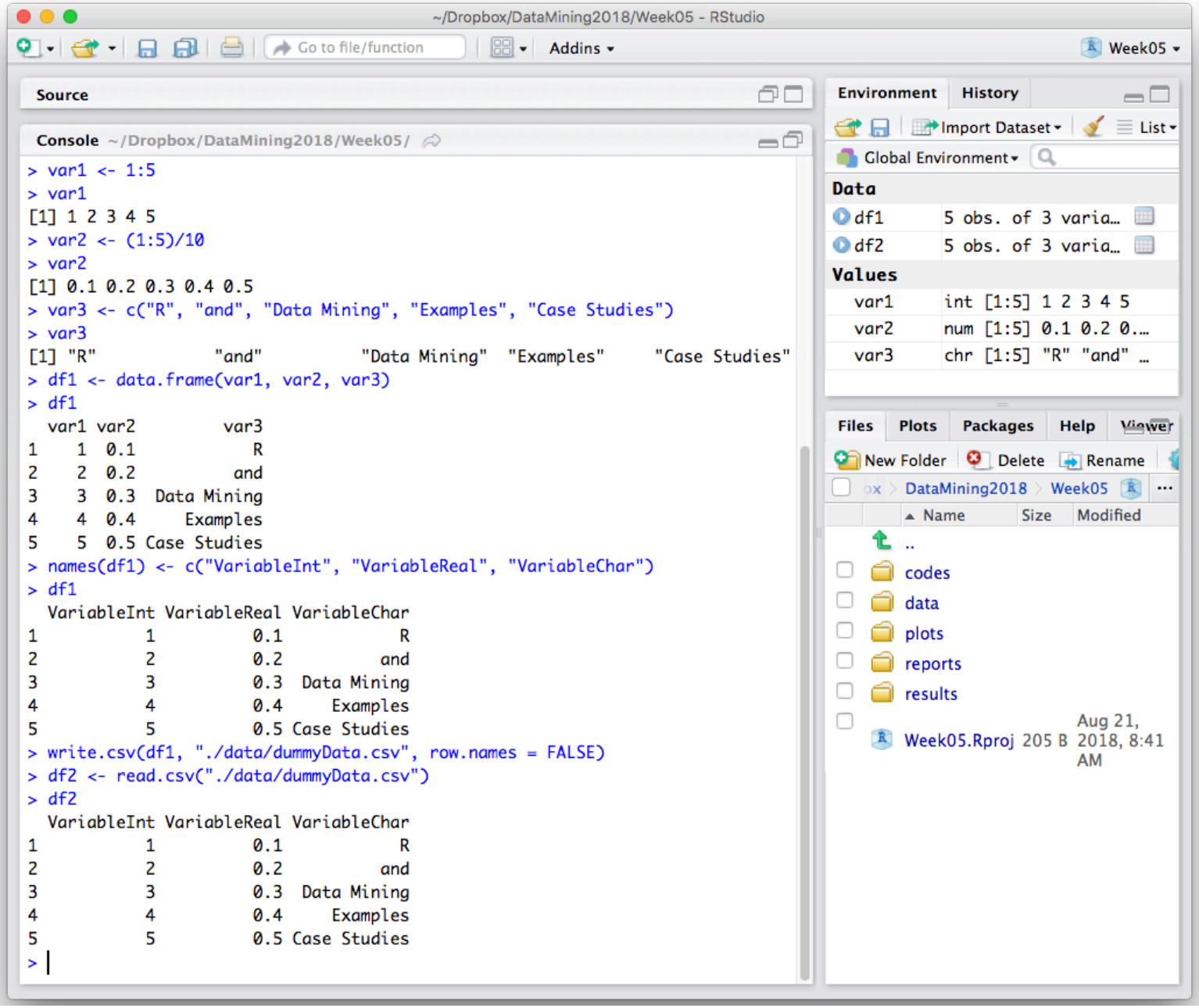

Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Create a new folder
ining2018
Week05
Name
Size
Modified
..
Week05.Rproj
205 B
Aug 21, 2018, 8:41 AM


Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Dropbox
DataMining2018
Week05
Name
Size
Modified
..
Week05.Rproj
205 B
Aug 21, 2018, 8:41 AM
codes
data
plots
results
reports

Part 3

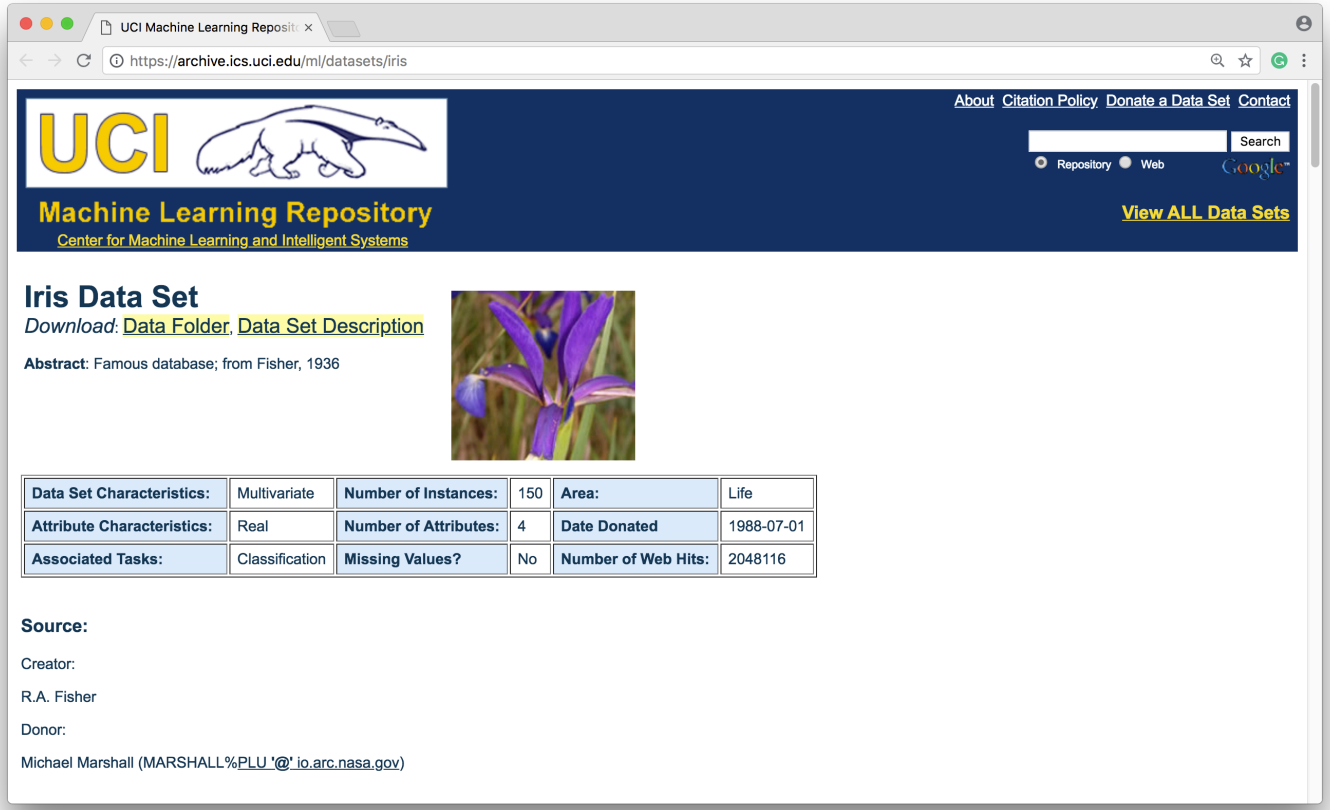
Data Extraction

Write and Read Data

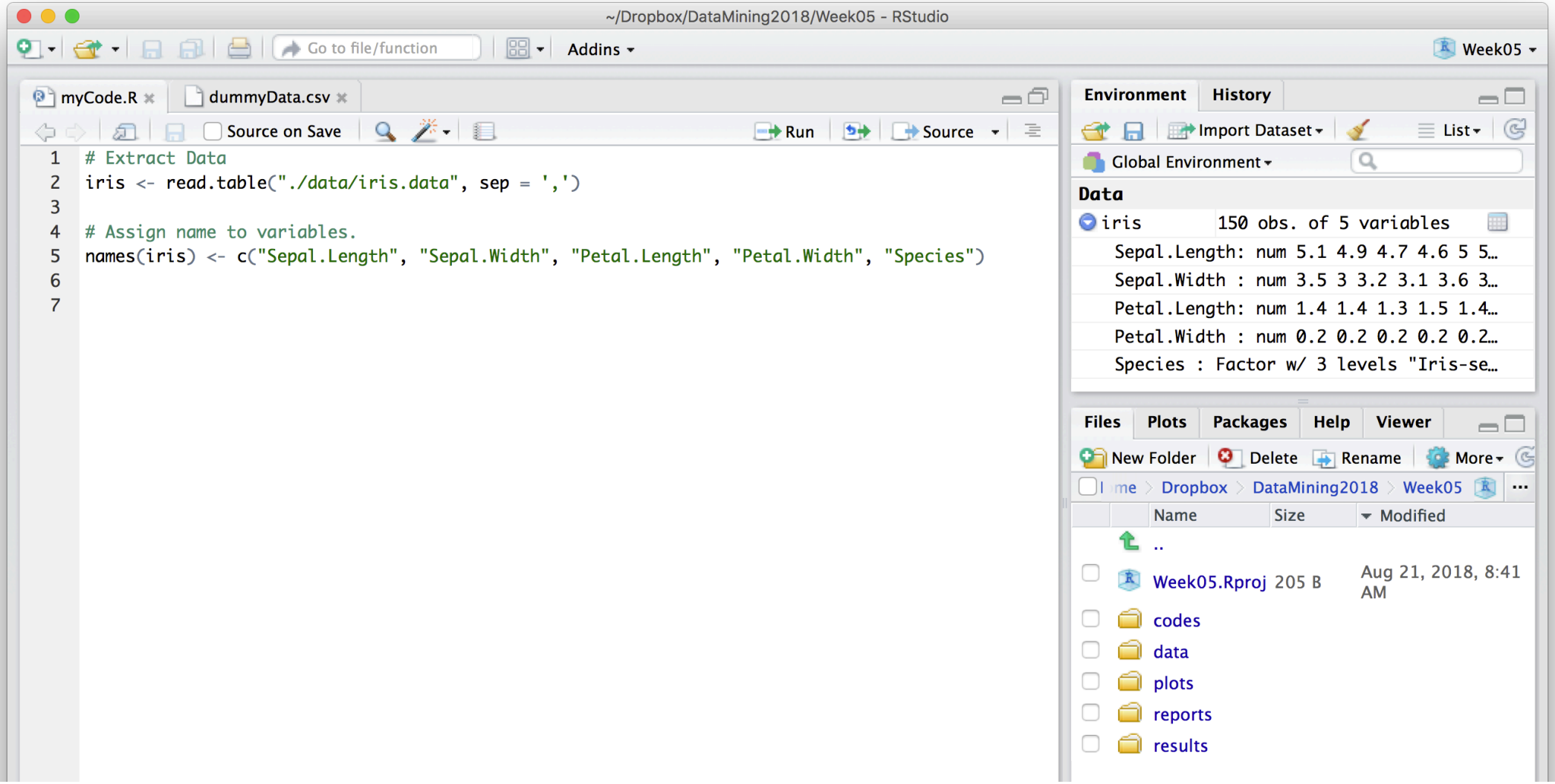


Download Dataset

- Iris Dataset: https://archive.ics.uci.edu/ml/datasets/iris
- File “iris.data”: https://archive.ics.uci.edu/ml/machine-learning-databases/iris/
- Save to data folder



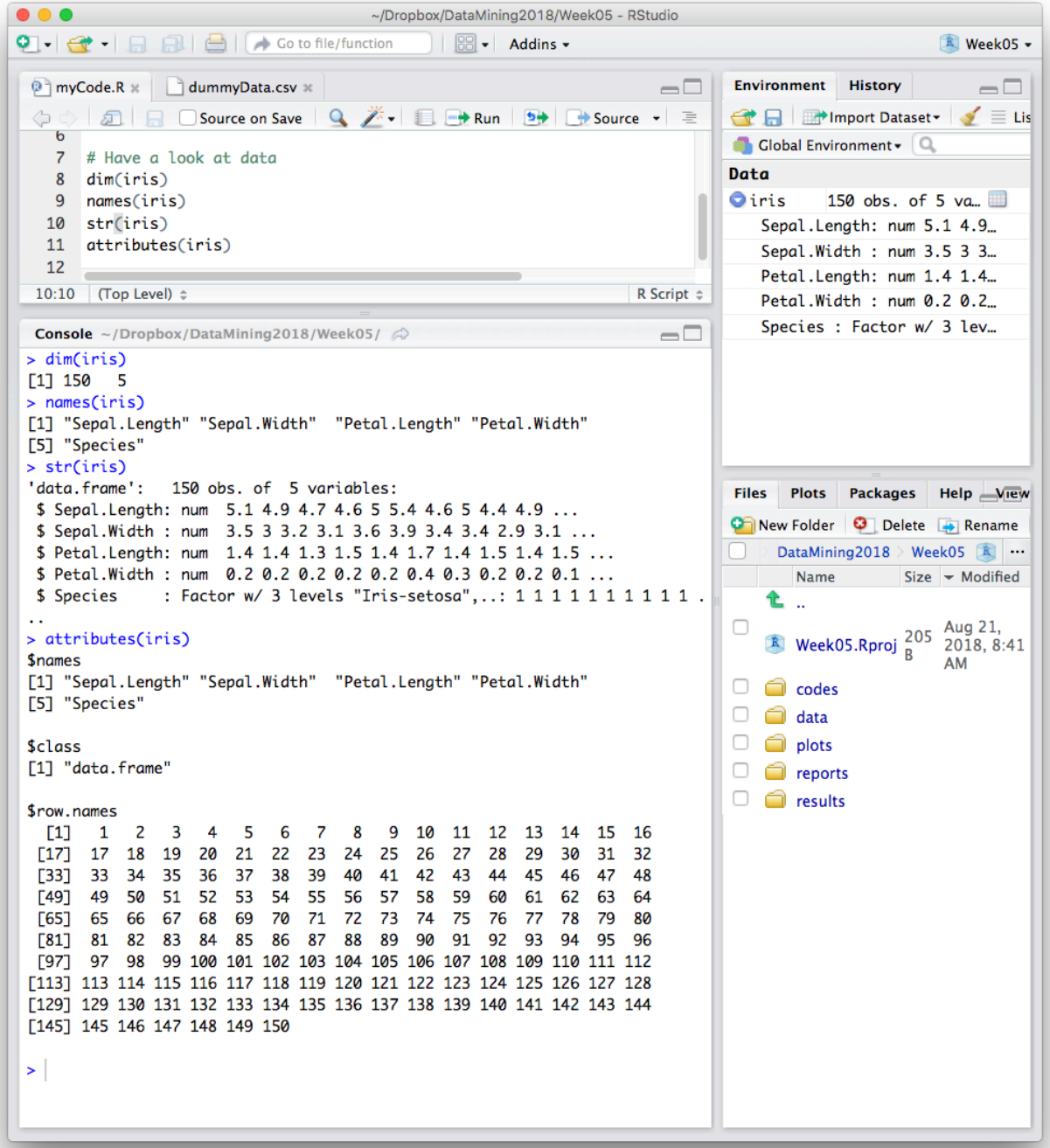
Extract Data



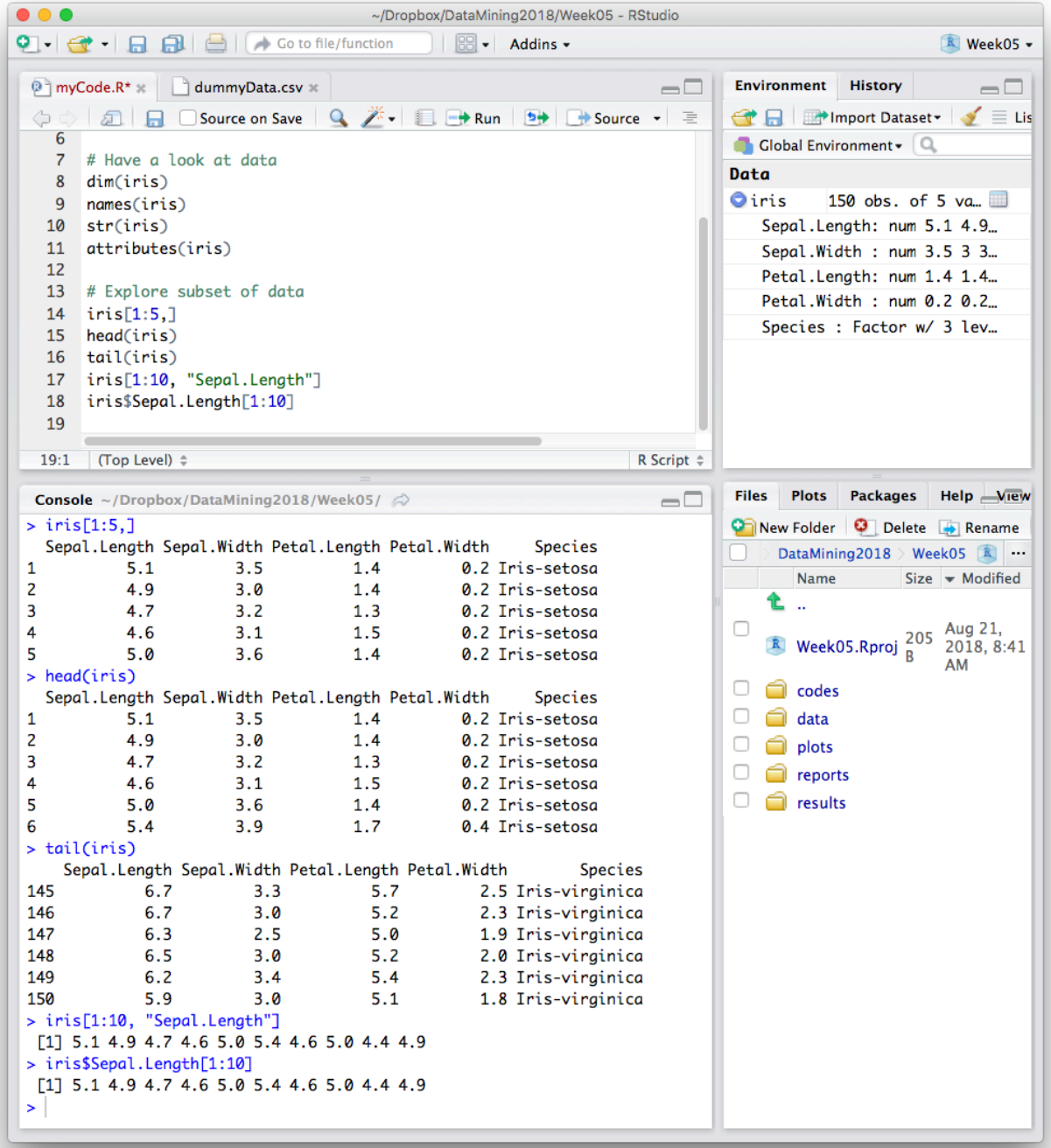
Part 4

Data Exploration

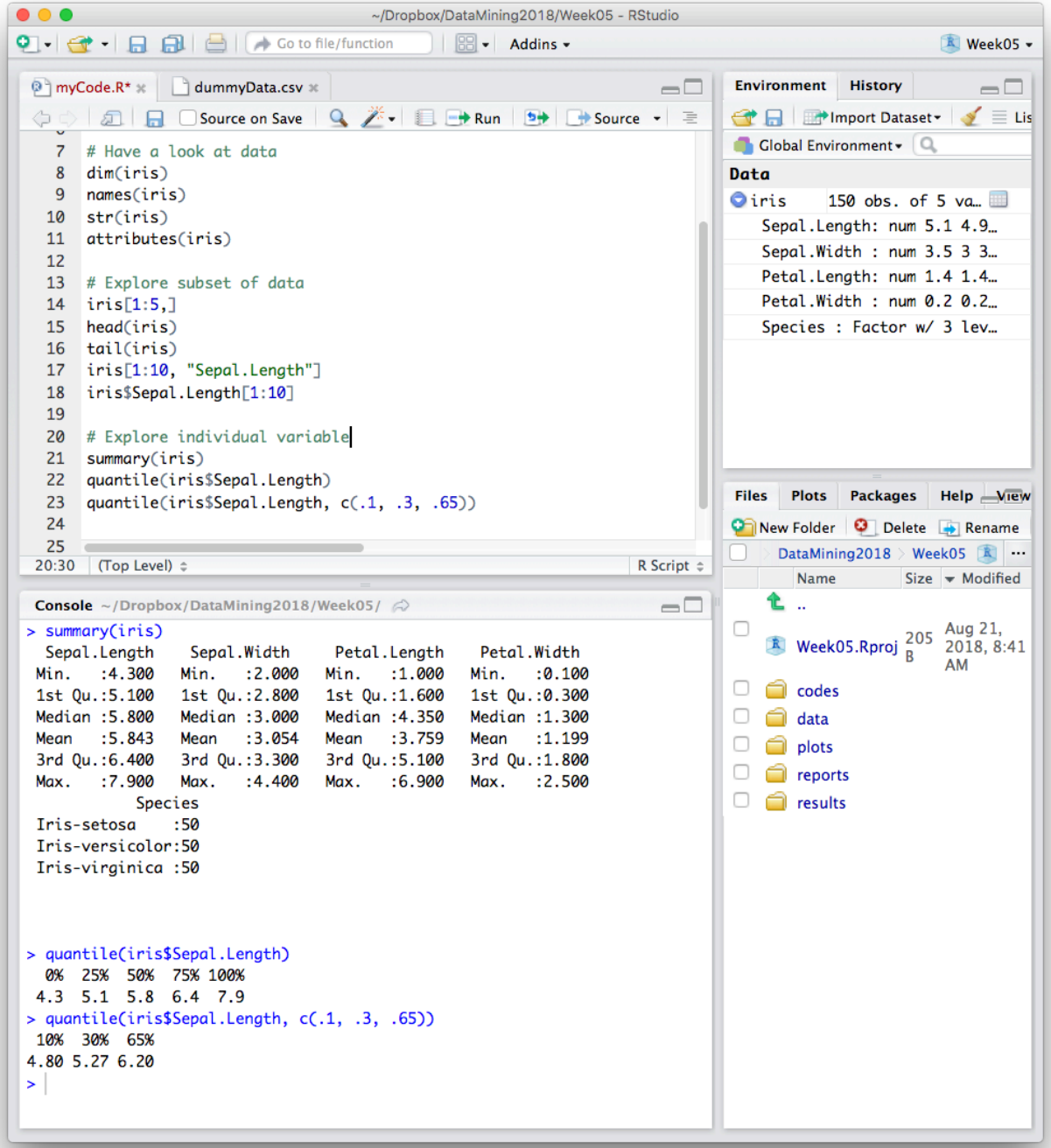
Explore Dataset



Explore Subset of Dataset



Explore Individual Variable



Visualise the frequency

~/Dropbox/DataMining2018/Week05 - RStudio

myCode.R* | dummyData.csv

```
12
13  # Explore subset of data
14  iris[1:5,]
15  head(iris)
16  tail(iris)
17  iris[1:10, "Sepal.Length"]
18  iris$Sepal.Length[1:10]
19
20  # Explore individual variable
21  summary(iris)
22  quantile(iris$Sepal.Length)
23  quantile(iris$Sepal.Length, c(.1, .3, .65))
24
25  # Visualise the frequency
26  hist(iris$Sepal.Length)
27  table(iris$Species)
28  pie(table(iris$Species))
29  barplot(table(iris$Species))
30
```

30:1 (Top Level) R Script

Environment | History

Data

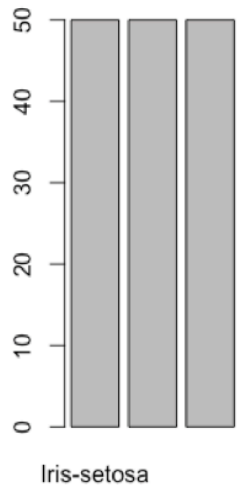
iris 150 obs. of 5 va…
Sepal.Length: num 5.1 4.9…
Sepal.Width : num 3.5 3 3…
Petal.Length: num 1.4 1.4…
Petal.Width : num 0.2 0.2…
Species : Factor w/ 3 lev…

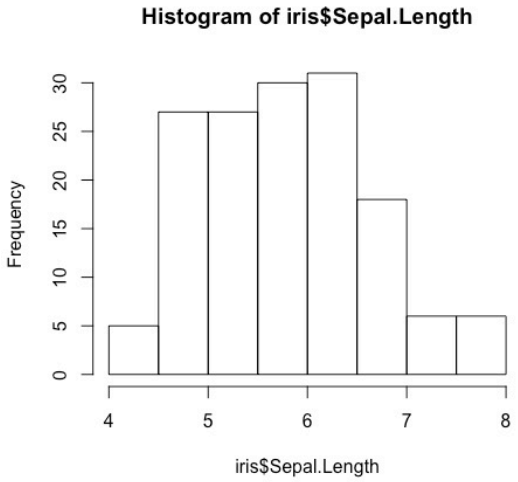
Console ~/Dropbox/DataMining2018/Week05/

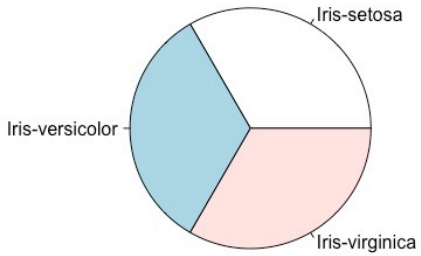
```
> hist(iris$Sepal.Length)
> table(iris$Species)

    Iris-setosa Iris-versicolor  Iris-virginica
             50              50              50
> pie(table(iris$Species))
> barplot(table(iris$Species))
>
```

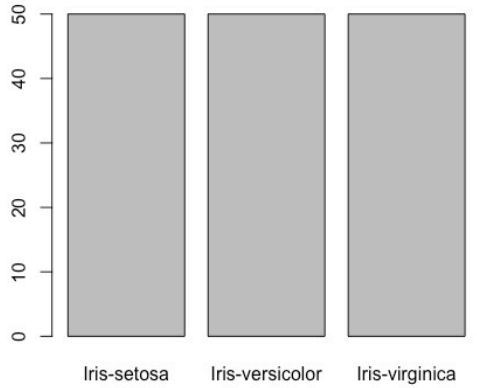
Files | Plots | Packages | Help | Viewer

Zoom | Export









Reference:

- R and Data Mining. Yangchan Zhao. Academic Press 2012. https://www.sciencedirect.com/book/9780123969637/r-and-data-mining