

# Installing R and RStudio

Augustin Luna  
17 January, 2016

Research Fellow  
Department of Biostatistics and Computational Biology  
Dana-Farber Cancer Institute

# Topics to be Covered

- R: Language Basics, Plotting, Getting Help
- Using the RStudio Editor

# What is R?

- Free, open source
- Started in 1993
- Geared towards scientific computing
  - Created by Ross Ihaka and Robert Gentleman (statisticians)
- Interpreted; similar to Python and MATLAB

# Why is R Popular?

- Free, open source
- Interactive data analysis
  - Script-driven rather than menu-driven helps reproducibility
- Flexible and powerful plotting support
- Excellent package management system

# R Package Management System

- Large and growing collection of statistical analysis methods
- Simple package installation; dependency management
- R scripts usually portable to other platforms
- Package repositories ensure functionality, documentation, and interoperability
- Vignettes (tutorials) provided as runnable analyses
- Large diversity of packages for data analysts
  - This presentation was produced with R packages

# Extending R and Package Repositories

- Comprehensive R Archive Network (CRAN)
  - 5,800 R packages (as of June 2014)
  - Many packages call C, C++, Fortran, or Java code for speedups
- Bioconductor
  - 800+ R packages focused on bioinformatics
  - 50+ packages dedicated to pathway analysis
- Devtools
  - R package that allows package installation from code repositories

# RStudio

- <https://www.rstudio.com/>
- Available for Windows, OSX, and Linux
- Simplifies common tasks: plotting, package installation, accessing files, viewing variables, etc.
- Provides code-completion so users do not have to remember whole lists functions and tons parameters

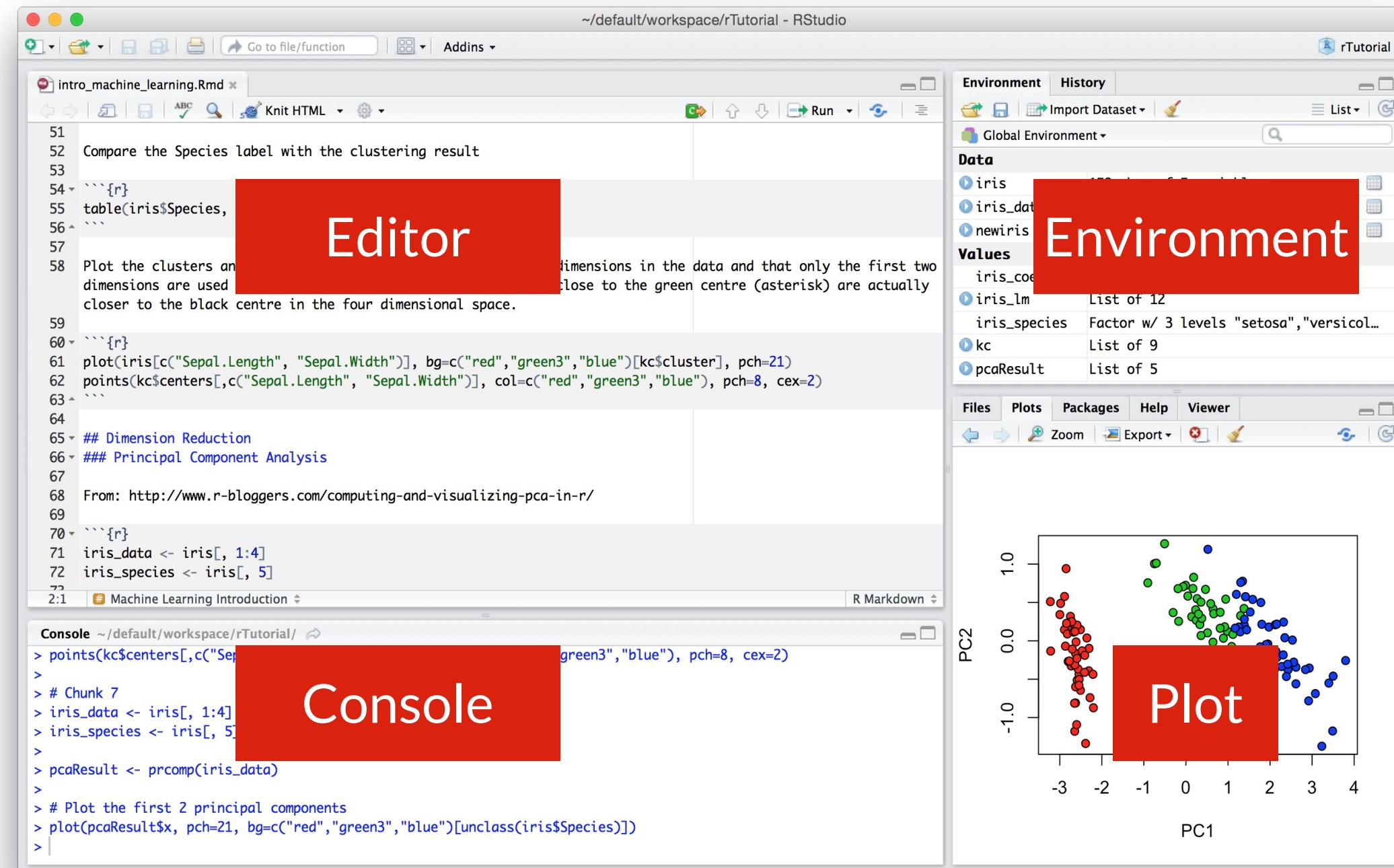
# Installing R and RStudio

- Install R
  - <https://cran.rstudio.com/>
- Install RStudio
  - <https://www.rstudio.com/products/rstudio/download/>
- RStudio does not come with R and R must be installed first

# YouTube Video Guides to Install R and RStudio

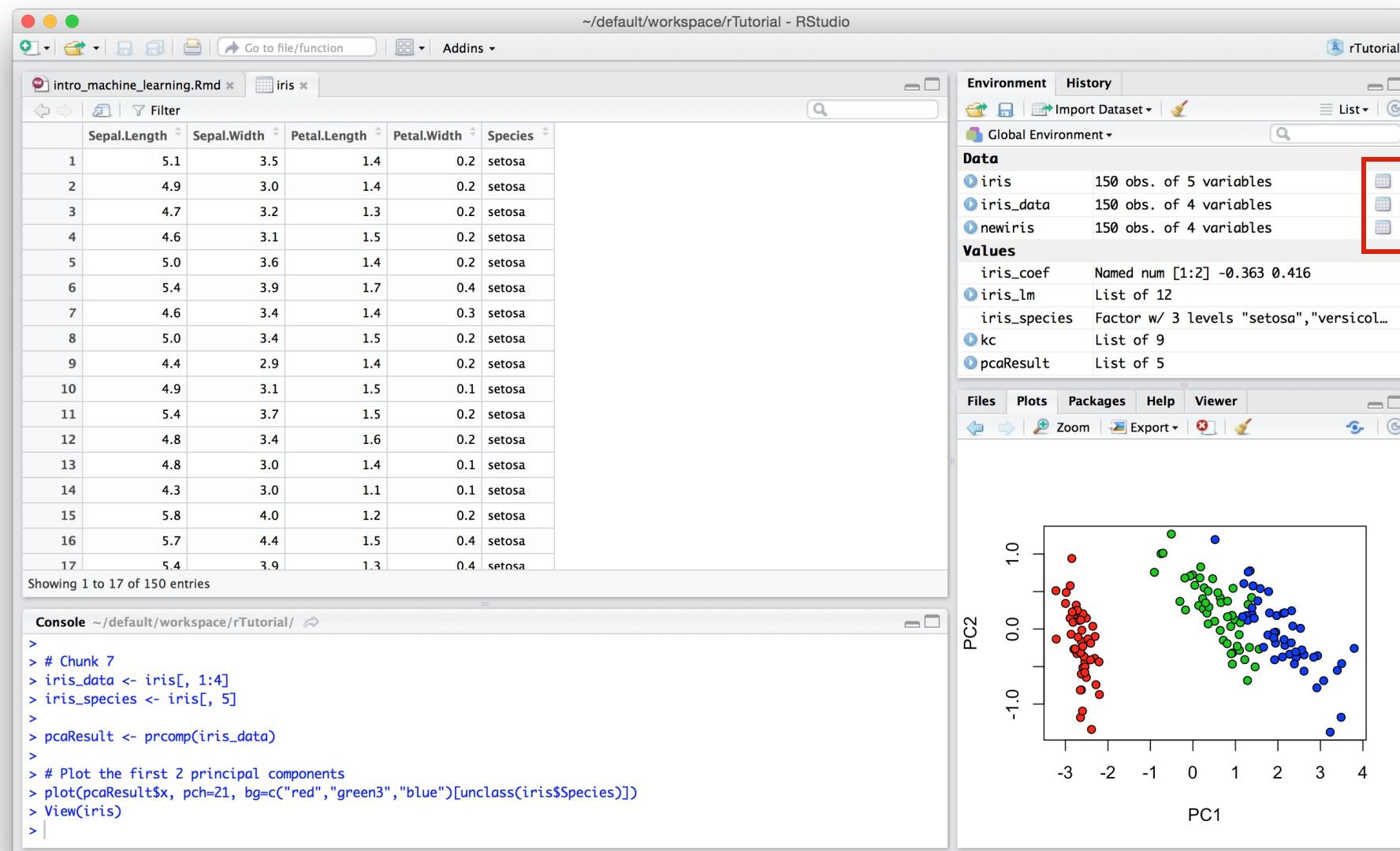
- The following videos show how to install R and RStudio from scratch
  - Include instructions for many common dependencies needed for pathway analysis package paxtoolsr
  - OS X: <https://youtu.be/lUwP6KncMOo>
  - Windows: <https://youtu.be/LcnCngOlbJc>
  - Linux: <https://youtu.be/JICy9lwZrOk>

# RStudio Overview



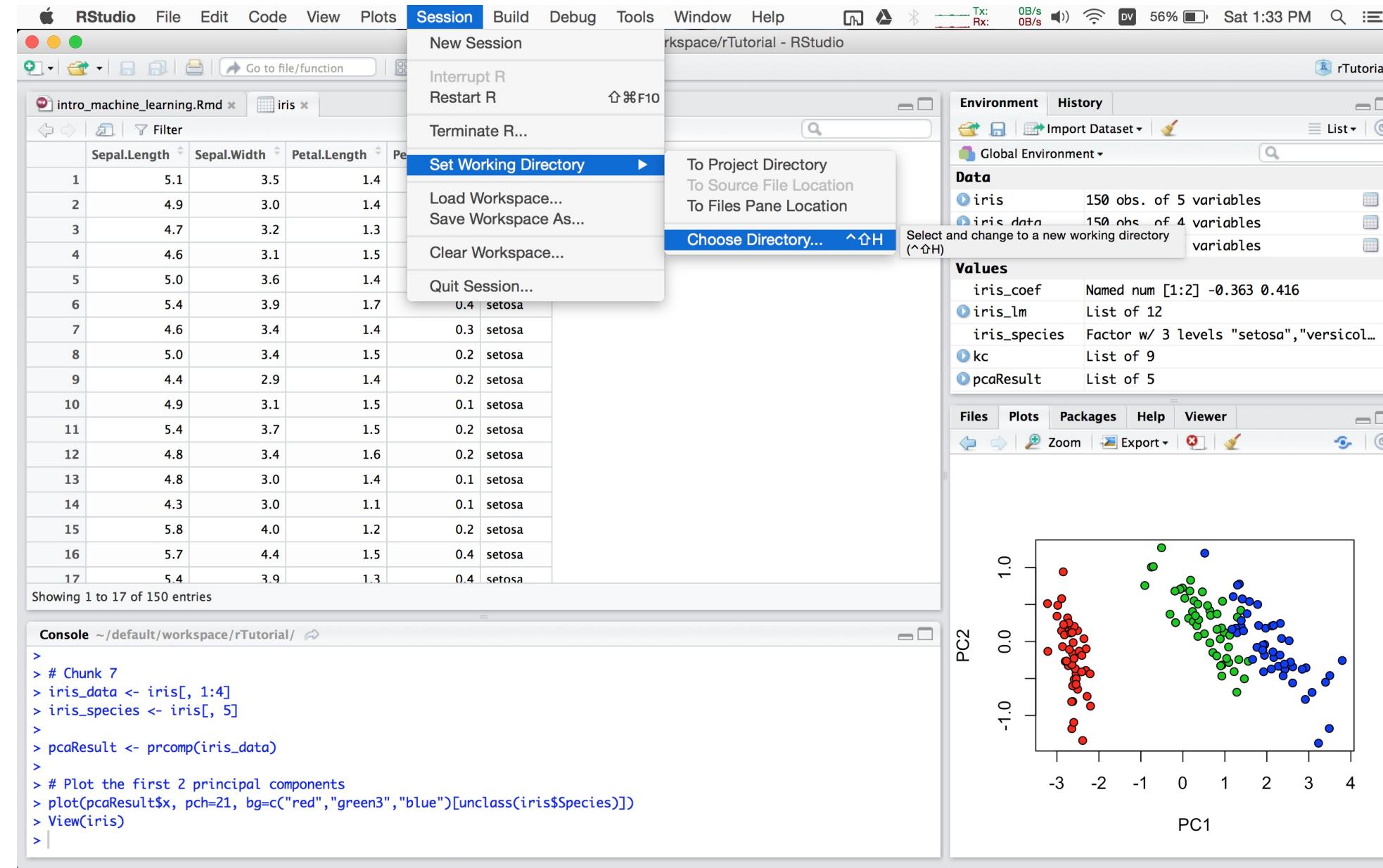
# Table View of Variables

- Highlighted boxes open a table view of variable contents



# Change Current Directory

- Highlighted boxes open a table view of variable contents



# Making a New R Script

The screenshot shows the RStudio interface with the following components:

- File Menu:** The "File" menu is open, showing options like "New File" (selected), "New Project...", "Open File...", "Save", "Print...", "Close", "Close Project", and "Quit Session...".
- Code Editor:** A data frame titled "Sepal.Length" is displayed, showing columns for Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, and Species (setosa).
- Console:** The console shows R code for loading the iris dataset, creating data frames for species and data, performing PCA, and plotting the results.
- Data View:** The "Data" pane shows variables: iris (150 obs. of 5 variables), iris\_data (150 obs. of 4 variables), newiris (150 obs. of 4 variables), iris\_coef (Named num [1:2] -0.363 0.416), iris\_lm (List of 12), iris\_species (Factor w/ 3 levels "setosa", "versicol..."), kc (List of 9), and pcaResult (List of 5).
- Plots:** A scatter plot titled "PC1" vs "PC2" shows the first two principal components for the iris dataset, with points colored by species (red for setosa, green for versicolor, blue for virginica).

# First Script: Hello World!

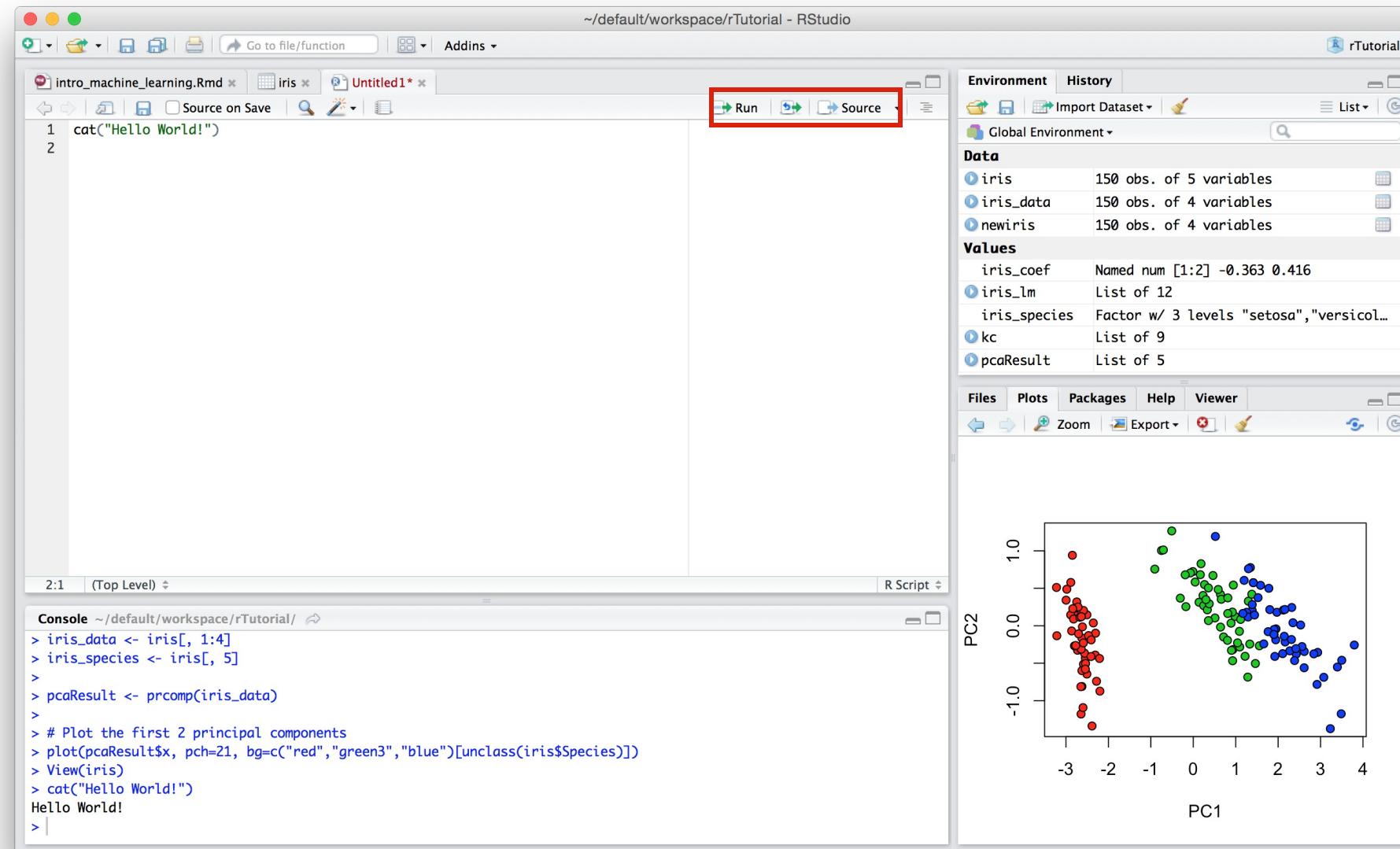
- `cat()` prints a simple message in the console

```
cat("Hello World!")
```

```
Hello World!
```

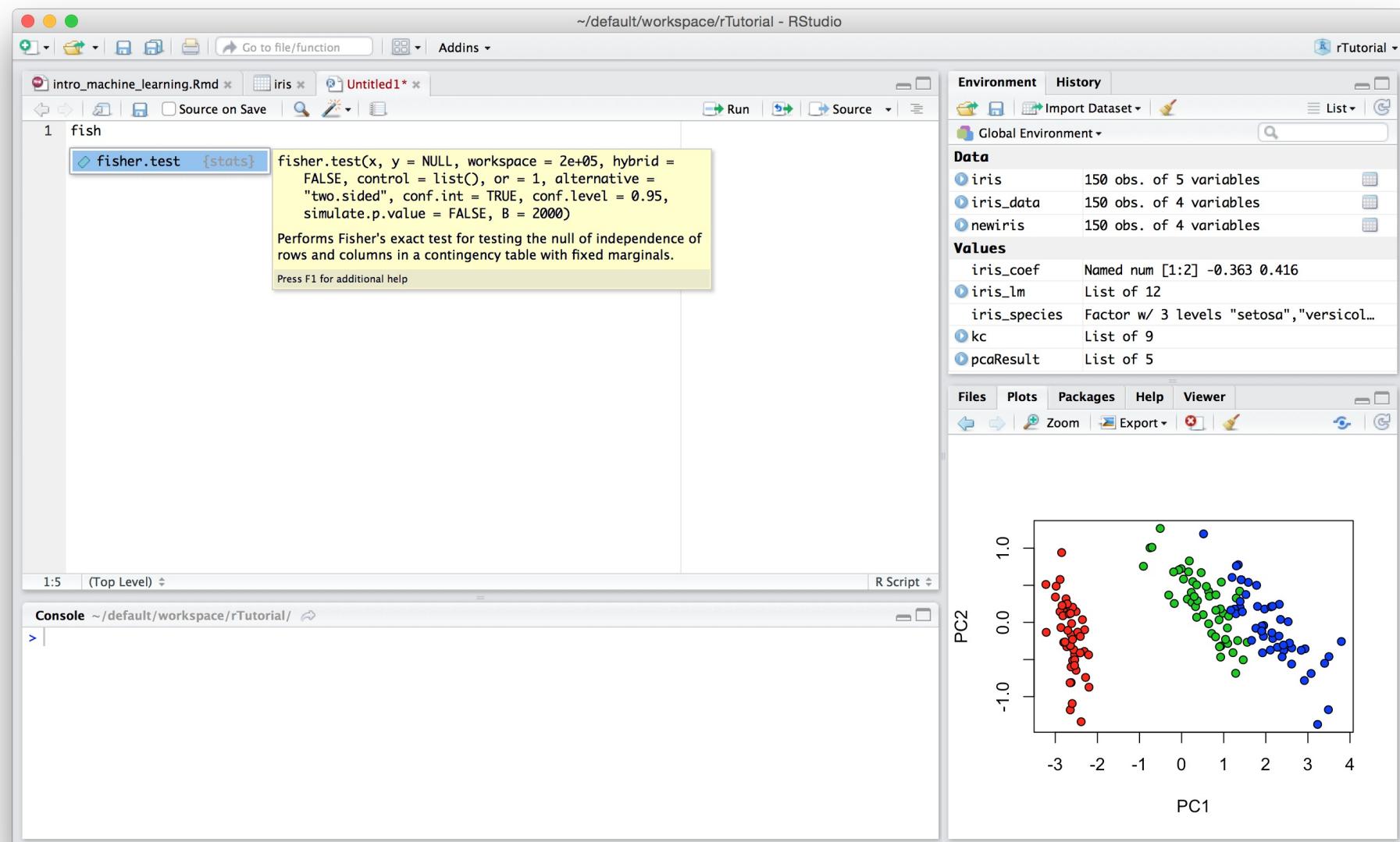
# Running Hello World Script

- “Run” button runs current line or selected lines
- “Source” button runs all lines in file



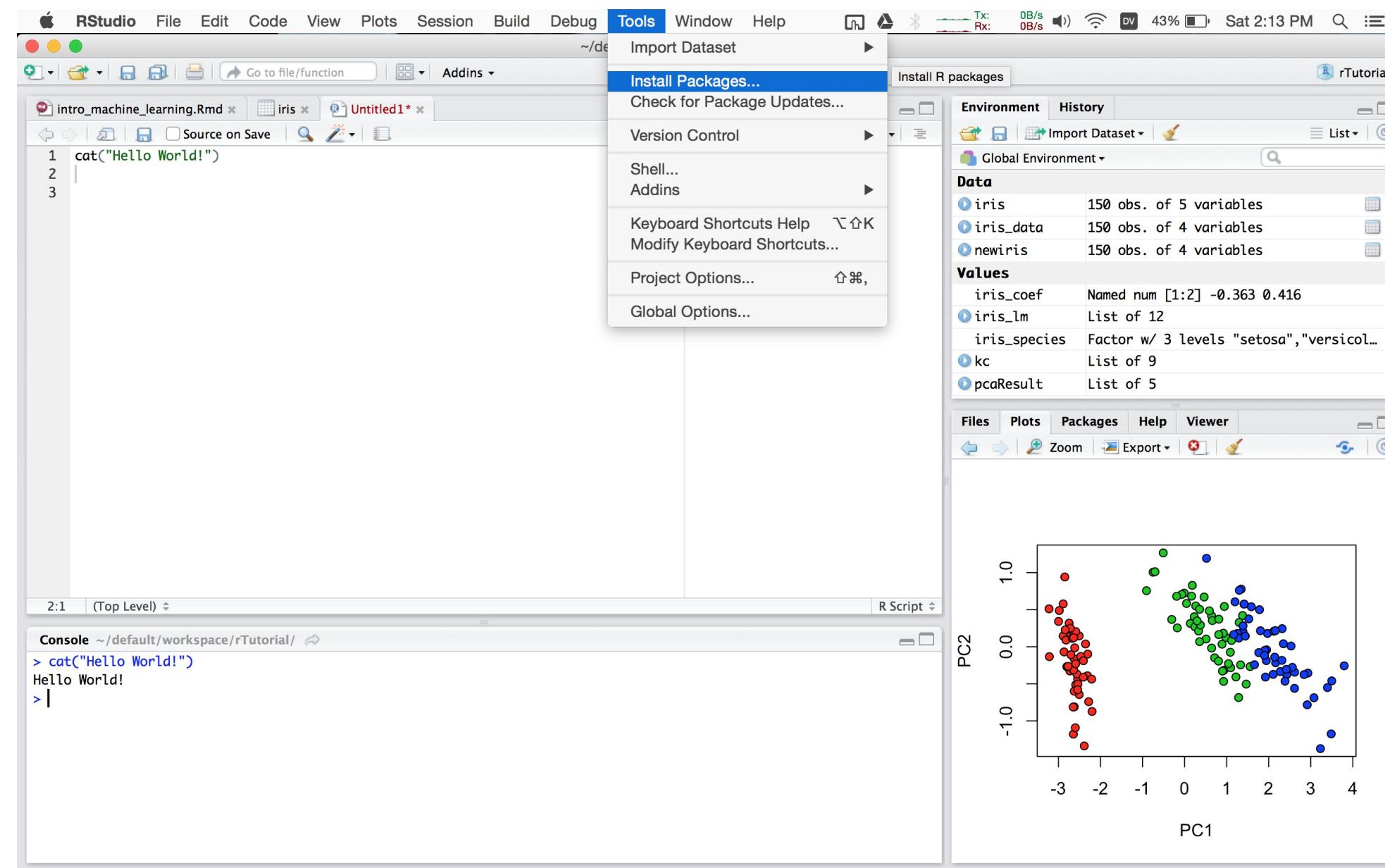
# Code Completion

- Pressing “Enter” will complete the name of the function



# Installing Packages

- CRAN packages can be installed using RStudio or `install.packages()`



# Creating Reports with R

- R and RStudio simplifies creating PDF and HTML (webpage) reports that can include:
  - Code
  - R code results
  - Formatted text
  - Tables
  - Figures
- R Markdown files have the extension .Rmd instead of .R
- Tutorials (vignettes) are commonly written in this format
- Markdown cheatsheet:  
<https://www.rstudio.com/resources/cheatsheets/>

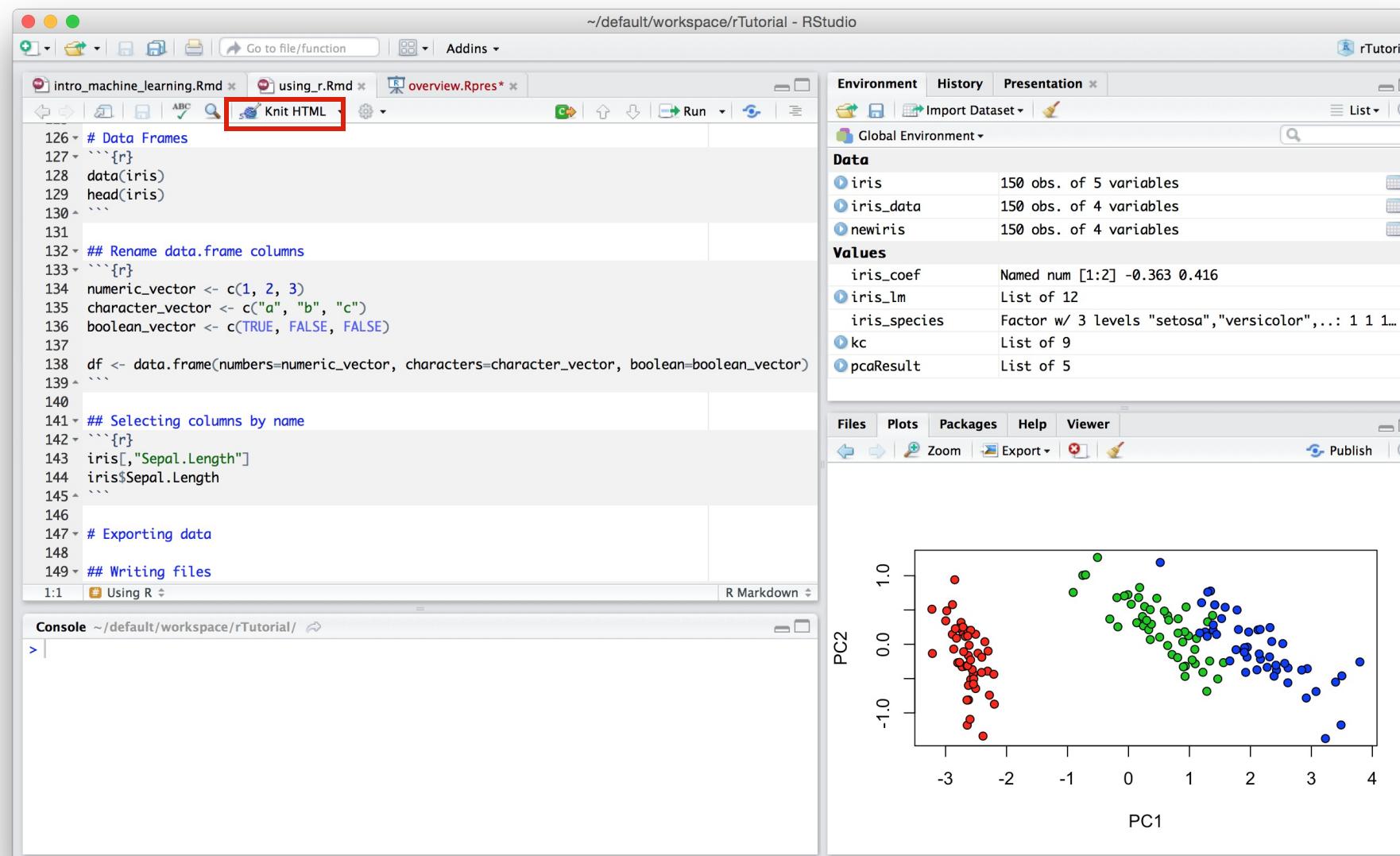
# Example RMarkdown (Rmd) Content

- NOTE: Remove apostrophes before each line when trying the Rmd file

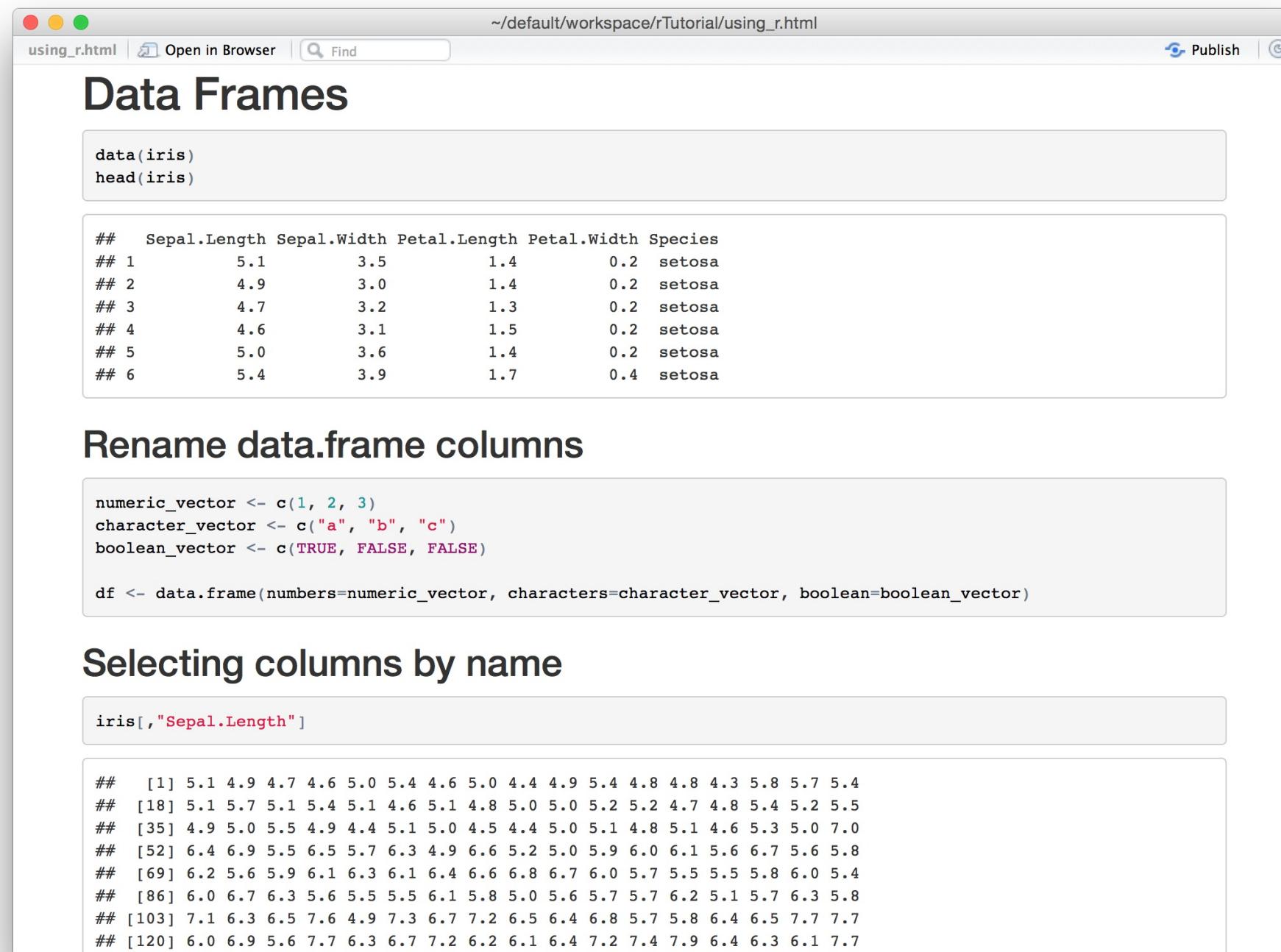
```
' # Heading
' ## Sub-heading
'
' Text
'
' ``{r}
' #R code
' cat("Hello")
'
```

# Converting Rmd to HTML

- Rendering .Rmd to HTML is done with the Knit HTML button in RStudio



# Example Rendered .Rmd



The screenshot shows a web browser window displaying an R Markdown document. The title bar reads "using\_r.html" and "Open in Browser". The main content area is divided into sections by h2 tags:

- Data Frames**

```
data(iris)
head(iris)
```

	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
- Rename data.frame columns**

```
numeric_vector <- c(1, 2, 3)
character_vector <- c("a", "b", "c")
boolean_vector <- c(TRUE, FALSE, FALSE)

df <- data.frame(numbers=numeric_vector, characters=character_vector, boolean=boolean_vector)
```
- Selecting columns by name**

```
iris[, "Sepal.Length"]
```

## [1] 5.1 4.9 4.7 4.6 5.0 5.4 4.6 5.0 4.4 4.9 5.4 4.8 4.8 4.3 5.8 5.7 5.4
## [18] 5.1 5.7 5.1 5.4 5.1 4.6 5.1 4.8 5.0 5.0 5.2 5.2 4.7 4.8 5.4 5.2 5.5
## [35] 4.9 5.0 5.5 4.9 4.4 5.1 5.0 4.5 4.4 5.0 5.1 4.8 5.1 4.6 5.3 5.0 7.0
## [52] 6.4 6.9 5.5 6.5 5.7 6.3 4.9 6.6 5.2 5.0 5.9 6.0 6.1 5.6 6.7 5.6 5.8
## [69] 6.2 5.6 5.9 6.1 6.3 6.1 6.4 6.6 6.8 6.7 6.0 5.7 5.5 5.5 5.8 6.0 5.4
## [86] 6.0 6.7 6.3 5.6 5.5 5.5 6.1 5.8 5.0 5.6 5.7 5.7 6.2 5.1 5.7 6.3 5.8
## [103] 7.1 6.3 6.5 7.6 4.9 7.3 6.7 7.2 6.5 6.4 6.8 5.7 5.8 6.4 6.5 7.7 7.7
## [120] 6.0 6.9 5.6 7.7 6.3 6.7 7.2 6.2 6.1 6.4 7.2 7.4 7.9 6.4 6.3 6.1 7.7

# Getting Help

- Solutions to many R problems have been posted online
  - Search the web with the error messages
- Question/Answer Sites
  - Stack Overflow: <http://stackoverflow.com/>
  - Biostars (Bioinformatics): <https://www.biostars.org/>
- Cheatsheets
  - Summarize available functionality
  - R: <https://cran.r-project.org/doc/contrib/Short-refcard.pdf>
  - RStudio:  
<https://www.rstudio.com/resources/cheatsheets/>