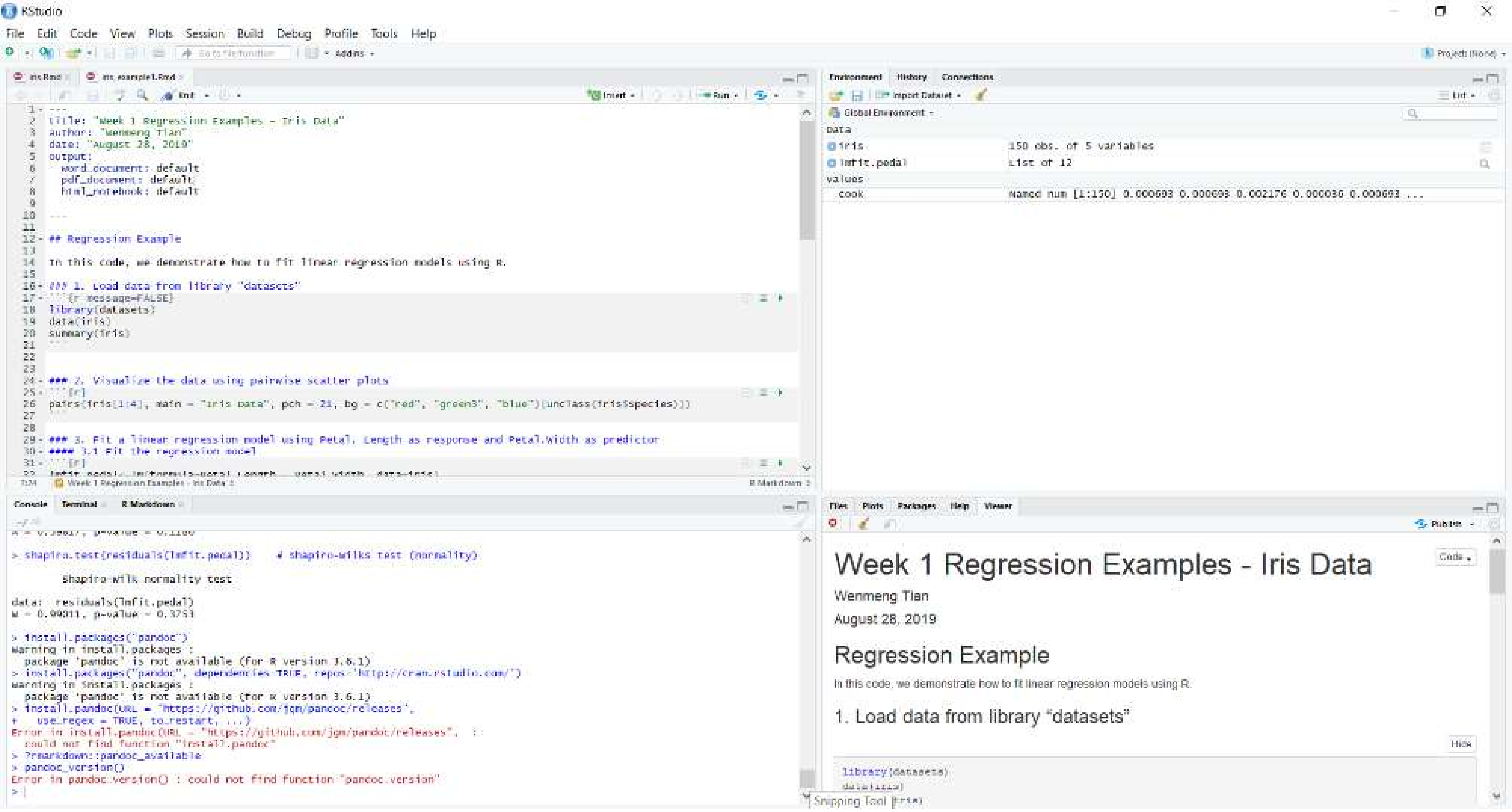
IE8990: Adv. Data Analytics for Complex Systems

- Lab 1
  - Getting started with R and Rstudio
  - Use R to fit regression model and perform diagnostics

R and Rstudio

- R Download
  - https://cloud.r-project.org/
- Rstudio Download
  - https://www.rstudio.com/products/rstudio/download/

Rstudio interface


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
iris.Rmd
iris_example1.Rmd
Knit
Insert
Run
Week 1 Regression Examples - Iris Data
R Markdown
Environment History Connections
Import Dataset
List
Global Environment
Console Terminal R Markdown
Files Plots Packages Help Viewer
Publish
Code
Hide
Snipping Tool

~~~
---
title: "Week 1 Regression Examples - Iris Data"
author: "Wenmeng Tian"
date: "August 28, 2019"
output:
  word_document: default
  pdf_document: default
  html_notebook: default

---

## Regression Example

In this code, we demonstrate how to fit linear regression models using R.

### 1. Load data from library "datasets"
```{r message=FALSE}
library(datasets)
data(iris)
summary(iris)
~~~

2. Visualize the data using pairwise scatter plots
```{r}
pairs(iris[1:4], main = "Iris Data", pch = 21, bg = c("red", "green3", "blue")[unclass(iris$Species)])
```

3. Fit a linear regression model using Petal. Length as response and Petal.Width as predictor
3.1 Fit the regression model
```{r}
lmfit.pedal <- lm(formula=Petal.Length ~ Petal.Width, data=iris)
```

```
Data
iris                 150 obs. of 5 variables
lmfit.pedal          List of 12
Values
  cook               Named num [1:150] 0.000693 0.000693 0.002176 0.000036 0.000693 ...
```

```
> shapiro.test(residuals(lmfit.pedal))    # shapiro-wilks test (normality)

        Shapiro-Wilk normality test

data:  residuals(lmfit.pedal)
W = 0.99011, p-value = 0.3751

> install.packages("pandoc")
Warning in install.packages :
  package 'pandoc' is not available (for R version 3.6.1)
> install.packages("pandoc", dependencies=TRUE, repos='http://cran.rstudio.com/')
Warning in install.packages :
  package 'pandoc' is not available (for R version 3.6.1)
> install.pandoc(URL = "https://github.com/jgm/pandoc/releases",
+   use_regex = TRUE, to_restart, ...)
Error in install.pandoc(URL = "https://github.com/jgm/pandoc/releases",  :
  could not find function "install.pandoc"
> ?rmarkdown::pandoc_available
> pandoc_version()
Error in pandoc_version() : could not find function "pandoc_version"
> 
```

Week 1 Regression Examples - Iris Data

Wenmeng Tian

August 28, 2019

Regression Example

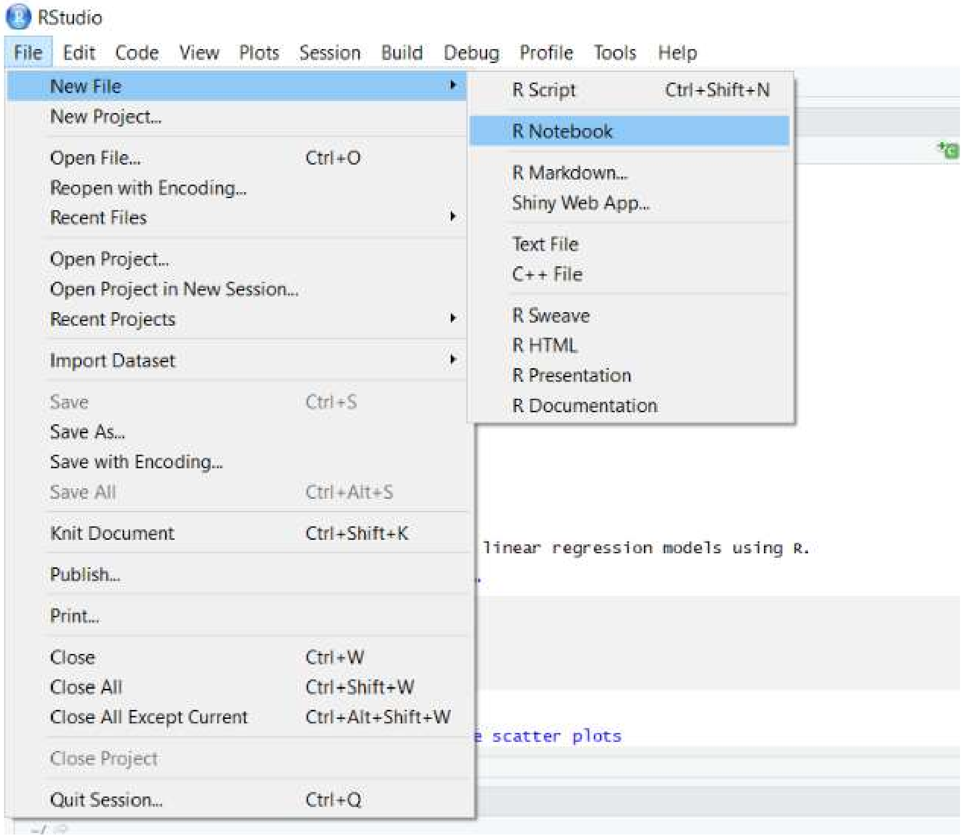
In this code, we demonstrate how to fit linear regression models using R.

1. Load data from library "datasets"

```
library(datasets)
data(iris)
```

R notebook and markdown

- The structure and functions of R notebook and R markdown are almost the same.
- Coding is exactly the same for these two.
- One major difference is that R notebook is able to provide a preview of the generate report.



```
linear regression models using R.
```

```
e scatter plots
```

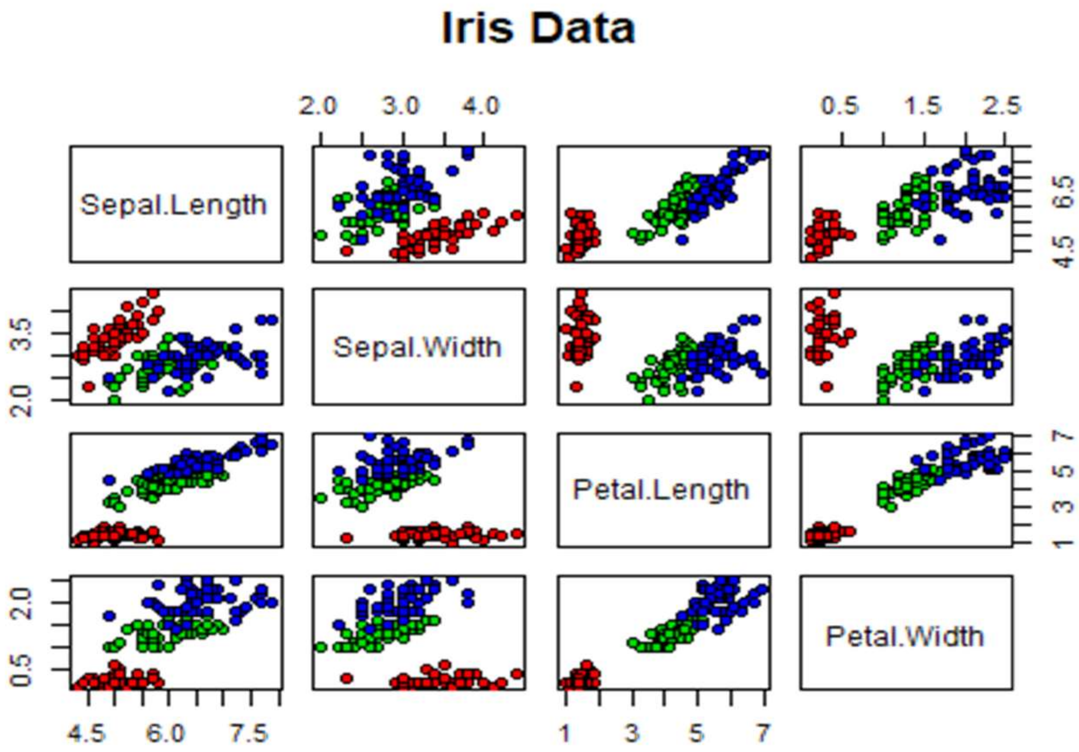
Example 1: Iris data

- Load the data
- Data visualization
- Fit a linear regression model using Petal. Length as response and Petal.Width as predictor
- Evaluate the performance of this model
- Find confidence interval of the parameters
- Diagnostics
  - Residual plots
  - Statistical tests

How do we interpret the results?

Data visualization

- Response variable
  - Sepal.Length



Homework 1: Question 1 – Iris data cont.

1. Fit Model 0 using Sepal.Length as response and Sepal.Width as the only one predictor
   a. Find all the parameters, write down their interpretations
   b. Evaluate Model 0 using adj R^2 as the criteria
   c. Is Model 0 good for prediction? If no, how do you want to improve the model?
2. Based on your answer in 1c, fit a new model (Model 1) using Sepal.Length as response, Sepal.Width and Species as predictors
   a. Find all the parameters, write down their interpretations
   b. Is Model 1 a good model for prediction?
3. Can you come up with another model that have better adj R^2 than Model 1?

We will discuss what models you have fitted next Thursday.