

# DASC 3240, L05

## R and Rstudio – II: R basics

---

2024/25 SPRING SEMESTER

MASAYUKI USHIO, ASSISTANT PROFESSOR AT OCES

[USHIO@UST.HK](mailto:USHIO@UST.HK) (RM CYT-2013)

# Outline (L05)

6. R basics

7. Package repositories

8. Hands-on

- Creating objects in R
- Data types
- Missing data

# R basics

---

SUMMARY OF KEY POINTS COMPARED WITH PYTHON

# Difference btwn R & Python 1 | Creating an object

- You can use “=” or “<-”. Recommendation is “<-”

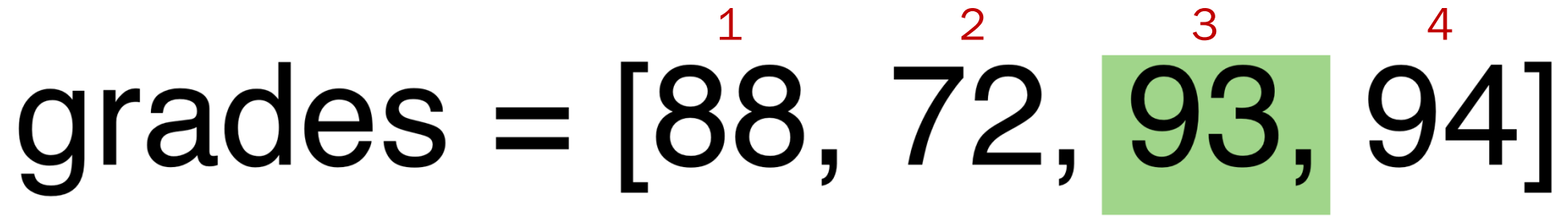
```
a = 5    # ok  
a <- 5    # ok  
b <- 3 + 5 # ok
```

# Difference btwn R & Python 2 | Indexing

- Indexing start with “1”, not “0”

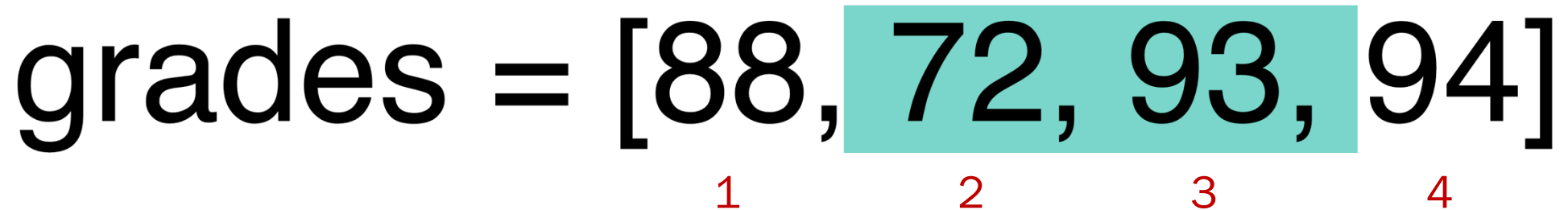
indexing: getting a specific element

grades = [88, 72, 93, 94]



slicing: selecting a set of elements

grades = [88, 72, 93, 94]



> grades[2:3]

# Difference btwn R & Python 3 | No indent

- You can use “{ }” for “for” loop and “if” statement.

```
# Create a for loop
```

```
for (i in 1:10) {  
    print(i)  
}
```

```
for (i in 1:10) { print(i) }
```

```
for (i in 1:10) print(i)
```

# Difference btwn R & Python 3 | No indent

- You can use “{}” for “for” loop and “if” statement.

# Create an if loop

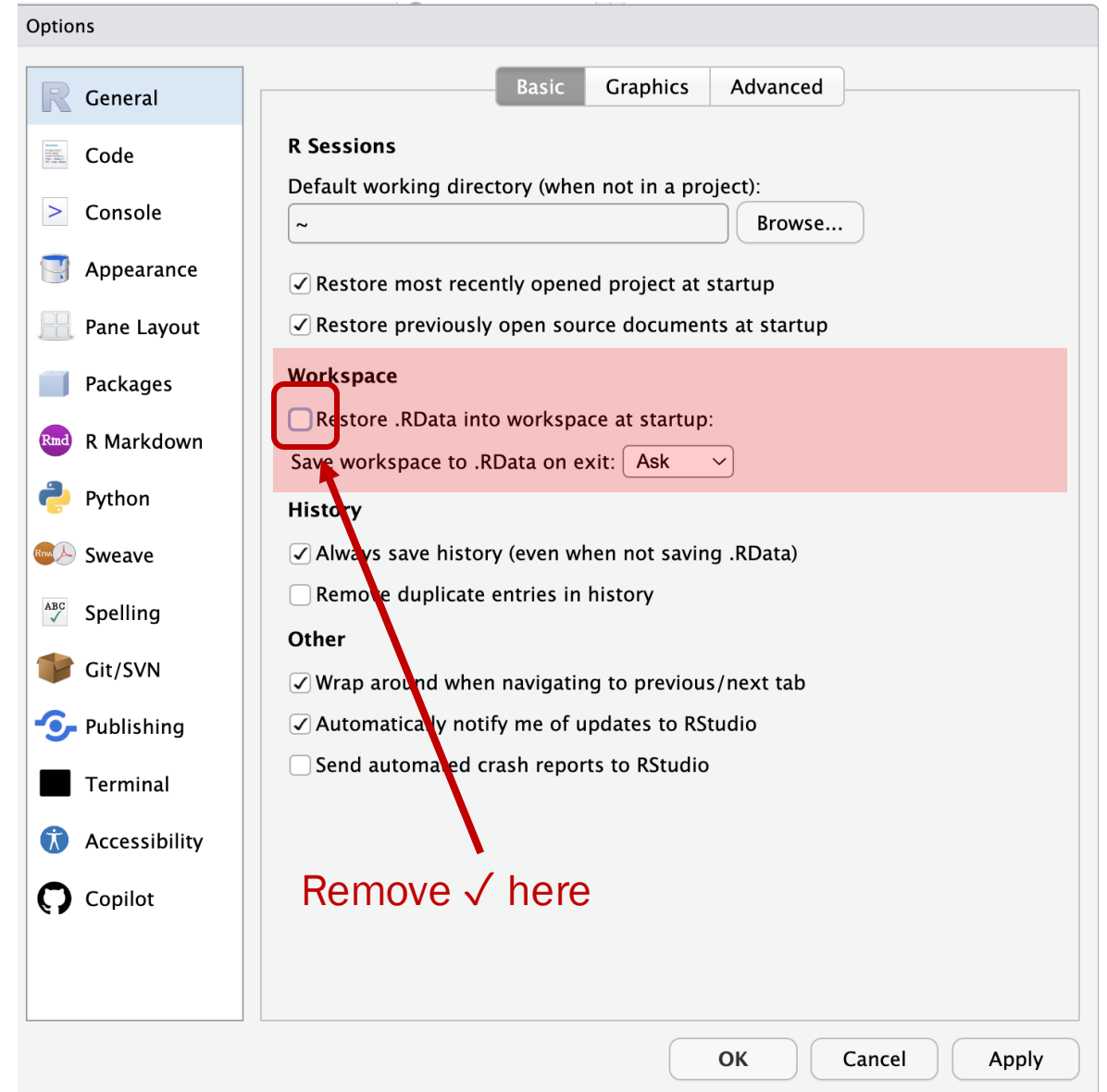
```
if (i == 10) {  
    print("i is 10")  
}
```

```
if (i == 10) { print("i is 10") }
```

```
if (i == 10) print("i is 10")
```

# Some R tips

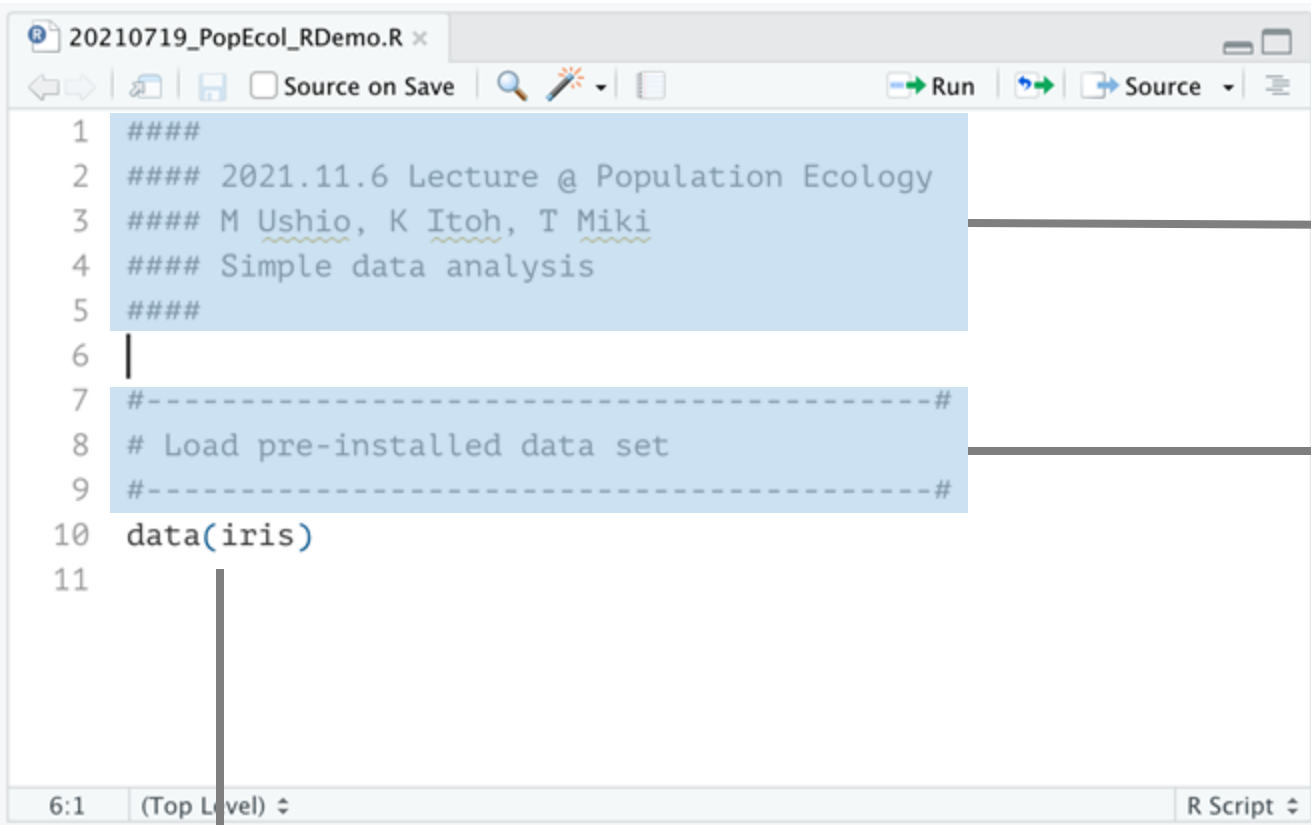
Go to “Tool” menu => “Global Options”





# Some R tips

You have to make your code readable...



```
1 #####
2 ##### 2021.11.6 Lecture @ Population Ecology
3 ##### M Ushio, K Itoh, T Miki
4 ##### Simple data analysis
5 #####
6 |
7 #-----#
8 # Load pre-installed data set
9 #-----#
10 data(iris)
11
```

Describe code information

If you type “#”, then the following characters will be recognized as a comment (not executed).  
(I put four “#”s, but this is a personal preference...)

Describe what you did in this section

**Caveats:** Always make detailed notes and comments, even if it's a bit tedious, as I am sure that **you will forget why you wrote that code, what you did there, and what you wanted to do**

“iris” is a famous dataset included in R by default. After loading the “iris” dataset by typing “**data(iris)**”, press “**Shift + Enter**” on Windows or “**Command + Enter**” on macOS to execute the command.

# Some R tips

The screenshot shows the RStudio interface. On the left, a data preview table for the 'iris' dataset is displayed, showing the first 13 rows. The columns are Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, and Species. The 'Species' column shows 'setosa' for all rows. On the right, the 'Environment' pane shows the 'Global Environment' with the 'iris' object listed under the 'Data' section. The 'iris' object is highlighted with a red box. An arrow points from the red box to the text '“iris” is now loaded in your environment'. Another arrow points from the text 'You can check the content by double-clicking this' to the 'iris' object in the Environment pane.

	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
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa
11	5.4	3.7	1.5	0.2	setosa
12	4.8	3.4	1.6	0.2	setosa

Showing 1 to 13 of 150 entries, 5 total columns

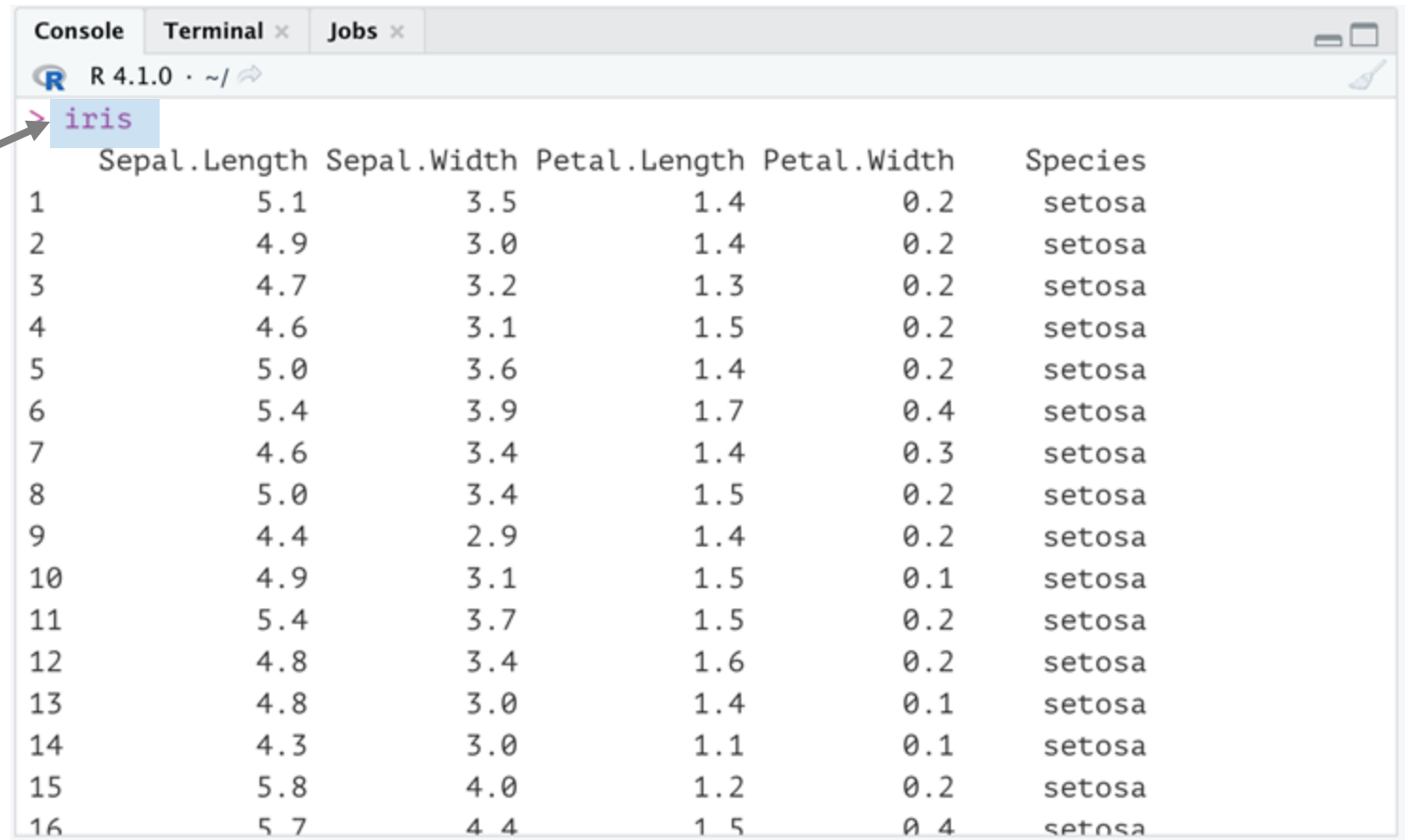
Environment History Connections Tutorial  
R Global Environment  
Data  
iris 150 obs. of 5 variables

“iris” is now loaded in your environment

You can check the content by double-clicking this

# Some R tips

Or, you can just type “iris” in the console.



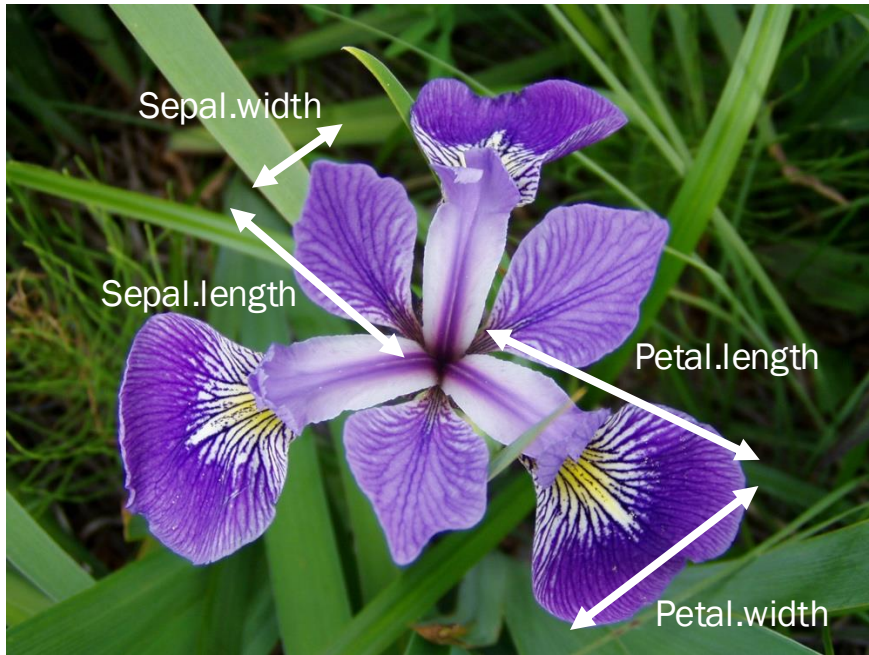
Console Terminal x Jobs x

R 4.1.0 · ~/

```
> 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
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa
11	5.4	3.7	1.5	0.2	setosa
12	4.8	3.4	1.6	0.2	setosa
13	4.8	3.0	1.4	0.1	setosa
14	4.3	3.0	1.1	0.1	setosa
15	5.8	4.0	1.2	0.2	setosa
16	5.7	4.4	1.5	0.4	setosa

# Some R tips



[https://en.wikipedia.org/wiki/Iris\\_flower\\_data\\_set#/media/File:Iris\\_versicolor\\_3.jpg](https://en.wikipedia.org/wiki/Iris_flower_data_set#/media/File:Iris_versicolor_3.jpg)

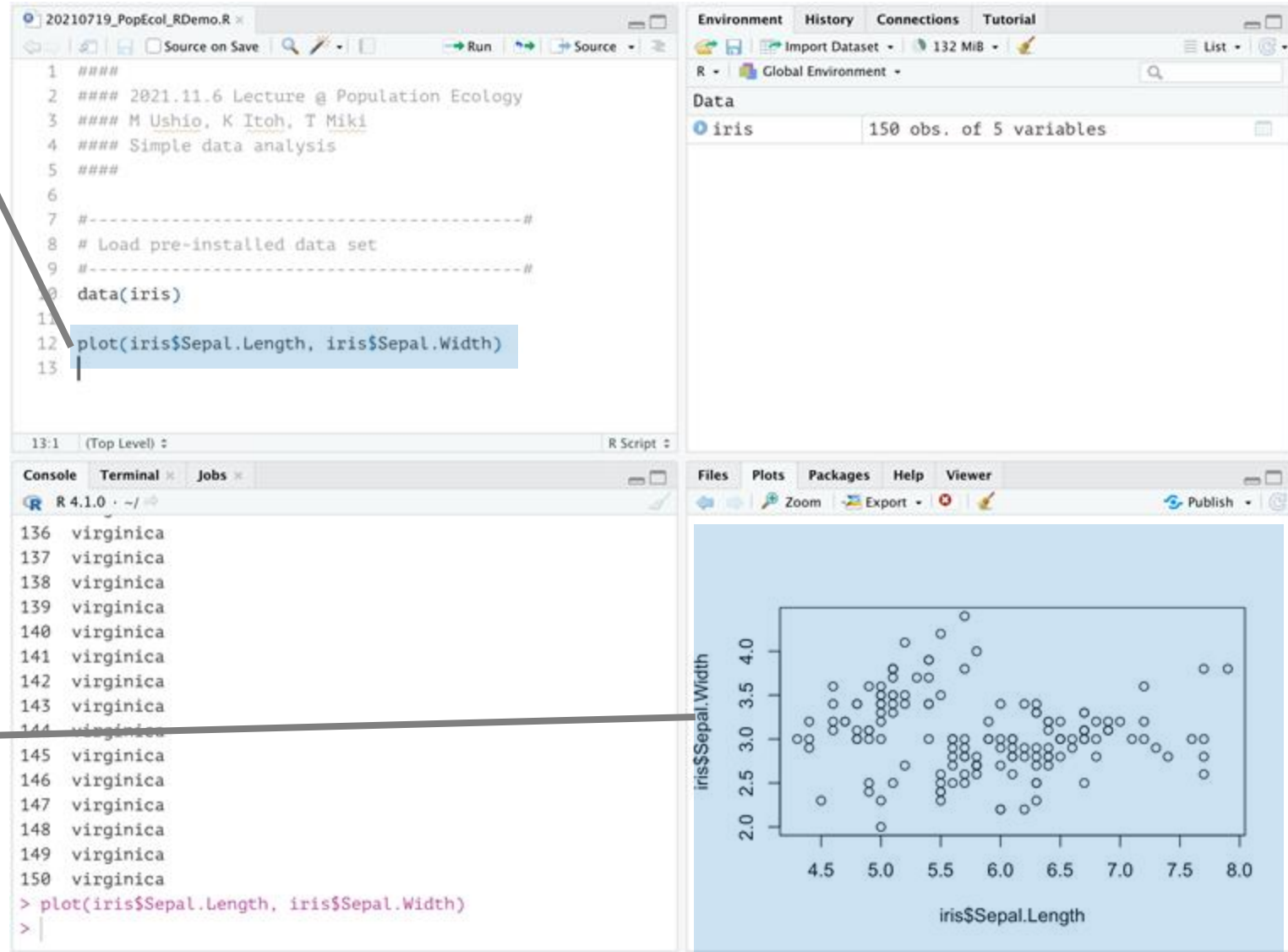
	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa
11	5.4	3.7	1.5	0.2	setosa
12	4.8	3.4	1.6	0.2	setosa
13	4.8	3.0	1.4	0.1	setosa
14	4.3	3.0	1.1	0.1	setosa
15	5.8	4.0	1.2	0.2	setosa
16	5.7	4.4	1.5	0.4	setosa

# Some R tips

Visualize the relationship between variables using plot() function.

- "\$" after "iris" is used to access a variable in the data frame.
- Visualize the relationship between "iris\$Sepal.Length" and "iris\$Sepal.Width"

Select the command, press "Shift+Enter" (Win) or "Command+Enter" (Mac). Then, the command will be executed and a plot will be shown at right-bottom.



# Package repositories

---

# Package repositories

## Python

- PyPi (<https://pypi.org/>)
- Anaconda (<https://anaconda.org/>)



 ANACONDA.ORG

 Search Anaconda.org

**anaconda / packages**

## R

- CRAN (<https://cran.r-project.org/>)
- Bioconductor (<https://www.bioconductor.org/>)

The Comprehensive R Archive Network





# Package repositories

R packages are registered in **The Comprehensive R Archive Network (CRAN)**  
<https://cran.r-project.org/>

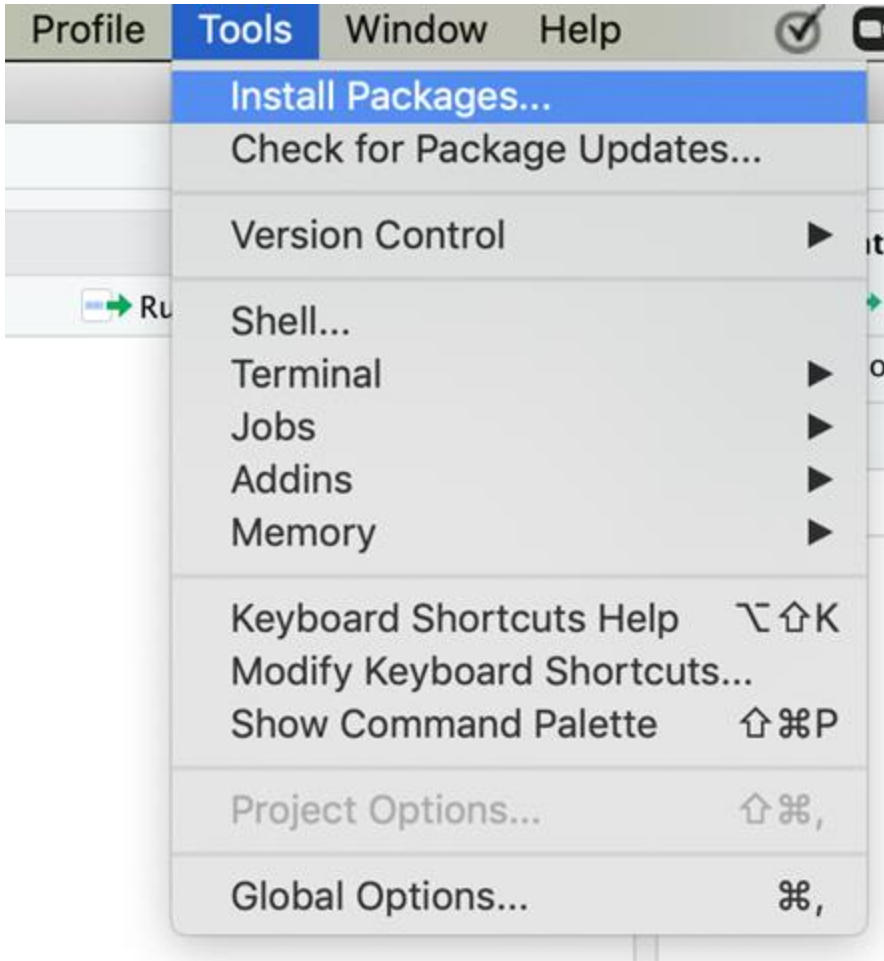
## Available CRAN Packages By Name

[A](#)[B](#)[C](#)[D](#)[E](#)[F](#)[G](#)[H](#)[I](#)[J](#)[K](#)[L](#)[M](#)[N](#)[O](#)[P](#)[Q](#)[R](#)[S](#)[T](#)[U](#)[V](#)[W](#)[X](#)[Y](#)[Z](#)

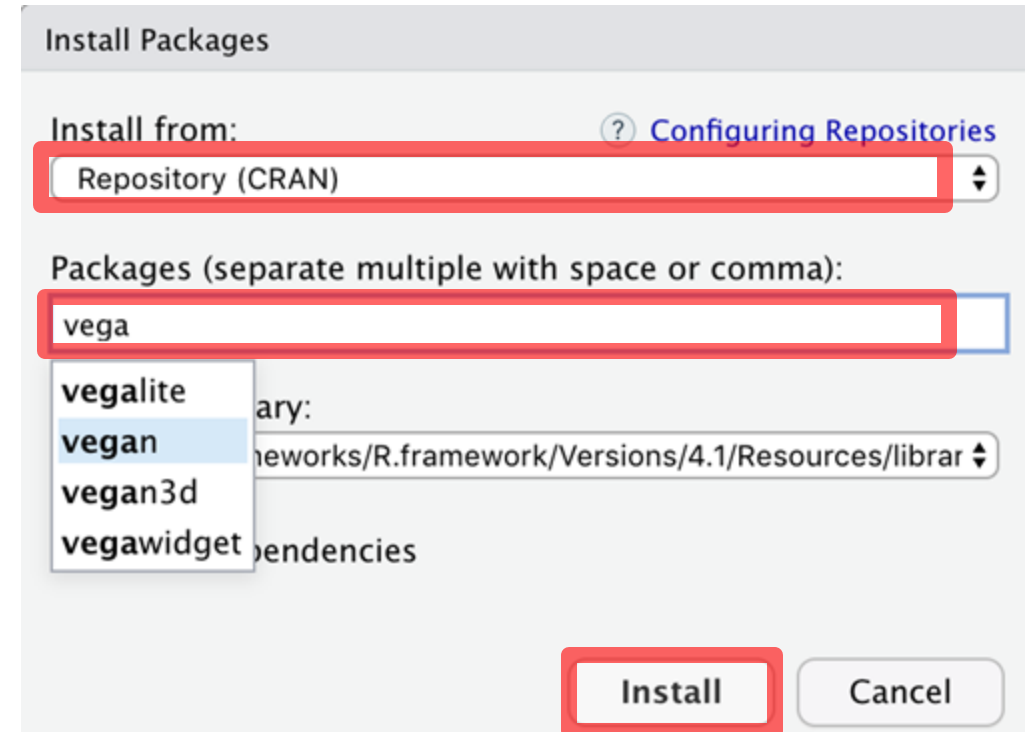
<a href="#">A3</a>	Accurate, Adaptable, and Accessible Error Metrics for Predictive Models
<a href="#">aaSEA</a>	Amino Acid Substitution Effect Analyser
<a href="#">AATtools</a>	Reliability and Scoring Routines for the Approach-Avoidance Task
<a href="#">ABACUS</a>	Apps Based Activities for Communicating and Understanding Statistics
<a href="#">abbyyR</a>	Access to Abbyy Optical Character Recognition (OCR) API
<a href="#">abc</a>	Tools for Approximate Bayesian Computation (ABC)
<a href="#">abc.data</a>	Data Only: Tools for Approximate Bayesian Computation (ABC)
<a href="#">ABC.RAP</a>	Array Based CpG Region Analysis Pipeline
<a href="#">abcADM</a>	Fit Accumulated Damage Models and Estimate Reliability using ABC
<a href="#">ABCanalysis</a>	Computed ABC Analysis
<a href="#">abcdeFBA</a>	ABCDE_FBA: A-Biologist-Can-Do-Everything of Flux Balance Analysis with this package
<a href="#">ABCOptim</a>	Implementation of Artificial Bee Colony (ABC) Optimization
<a href="#">ABCp2</a>	Approximate Bayesian Computational Model for Estimating P2



# Package repositories



- Select “Tools” in the RStudio menu bar → Select “Install Packages”
- Select “CRAN” in the “Install from” tab and input a package name you want to install
- Click “Install”



# Package repositories

## The Comprehensive R Archive Network

```
install.packages(XXXXX)
```



```
If (!require("BiocManager", quietly = TRUE))  
  install.packages("BiocManager")  
BiocManager::install("XXXXX", version = "XXX")
```



```
install.packages("remotes")  
remotes::install_github("XXXXX")
```

```
install.packages("devtools")  
devtools::install_github("XXXXX")
```

# Package repositories

:: (specify namespace)

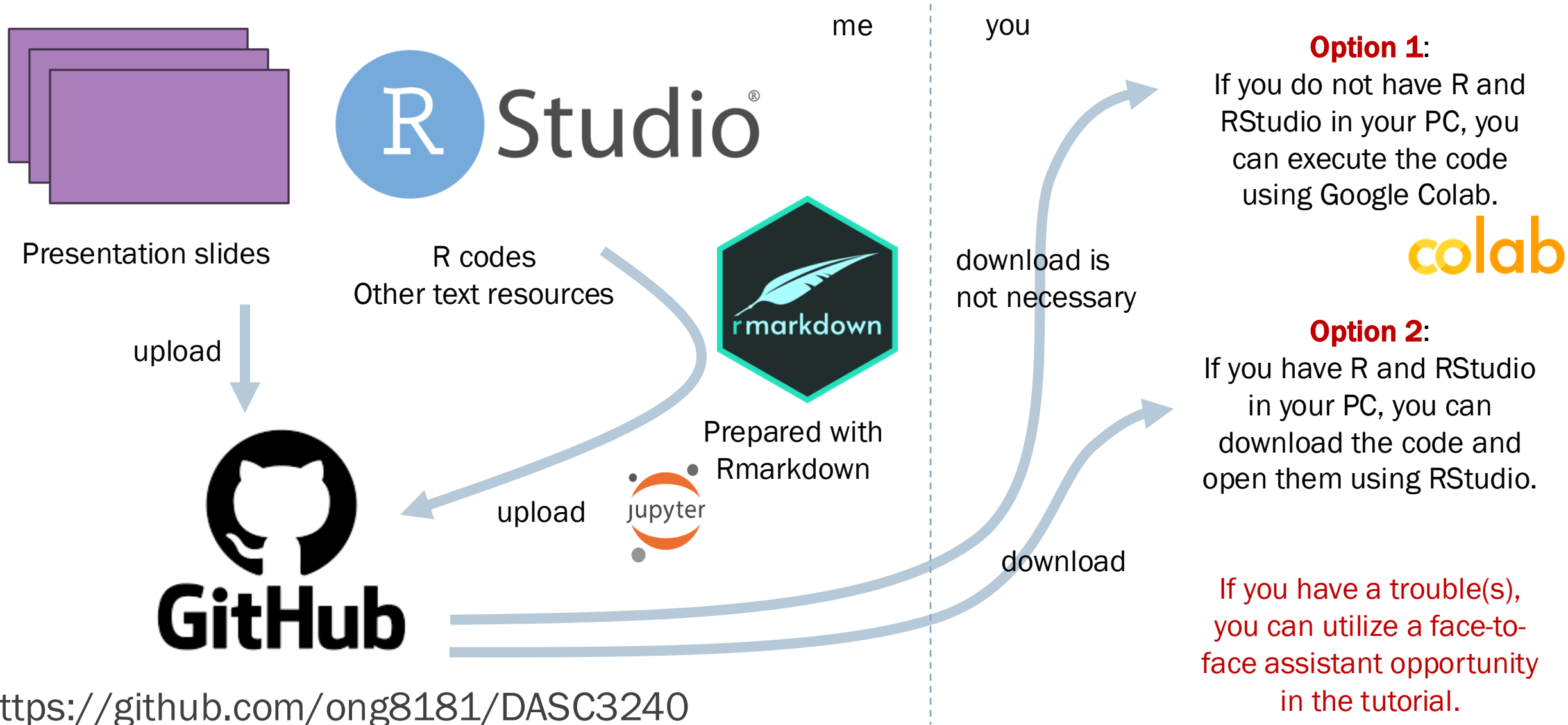
- In R, “::” is used to specify “namespace”. In each namespace, many functions are defined. If you specify namespace, you can avoid conflicts (e.g., different packages define functions with the same name).
- “xxxx::yyyy” means that you will use a function “yyyy” defined by a package “xxxx”




```
install.packages("remotes")  
remotes::install_github("XXXXXX")
```


```
install.packages("devtools")  
devtools::install_github("XXXXXX")
```

# Course materials













# Jupyter notebook




 jupyter L18\_RBasics01 Last Checkpoint: 1 minute ago



File Edit View Run Kernel Settings Help

Trusted

          Markdown ▾

JupyterLab   

Github repository is here: <https://github.com/ong8181/DASC3240>

## R Basics

The contents basically follow this website (<https://datacarpentry.org/R-ecology-lesson/01-intro-to-r.html>), but some contents are omitted/modified so that they fit DASC 3240.

### References

- Data Analysis and Visualization in R for Ecologists: <https://datacarpentry.org/R-ecology-lesson>
- R basics: <https://datacarpentry.org/R-ecology-lesson/01-intro-to-r.html>
- Zenodo archive: <http://doi.org/10.5281/zenodo.3264888>
- License: <https://github.com/datacarpentry/R-ecology-lesson?tab=License-1-ov-file>

## Contents

1. Creating objects in R
2. Vectors and data types
3. Missing data

### Select Kernel

Select kernel for: "Untitled.ipynb"

R

☐ Always start the preferred kernel

No Kernel

Select

# Jupyter notebook

- If you cannot change the kernel (language), please try the following in R console.
  - Launch R or Rstudio
  - In R console, type

```
install.packages(c('repr', 'IRdisplay', 'evaluate', 'crayon', 'pbdZMQ', 'devtools', 'uuid', 'digest'))  
install.packages('IRkernel')  
IRkernel::installspec()
```

The contents of this lecture generally follow the tutorial below.

## Data Analysis and Visualization in R for Ecologists



**DATA CARPENTRY**

BUILDING COMMUNITIES TEACHING UNIVERSAL DATA LITERACY

<https://datacarpentry.org/>  
<https://datacarpentry.github.io/R-ecology-lesson/>

Licensed under CC-BY 4.0

📖 README

📄 GPL-3.0 license




## DASC3240 Data Visualization in Science

This repository contains lecture PDFs and R code for 2024-25 Spring Semester to learn data visualization using R and RStudio.

### DASC3240\_code

This folder contains R code files.

- L05 R Basics:  [Open in Colab](#) or [Download L05\\_Basics.Rmd](#)
- L06 Data manipulation and ggplot2
- L07 Plot Basics