



北京大学
PEKING UNIVERSITY



DATA VISUALIZATION AND ANALYSIS METHODS

Final Recap and Final Project: Fall 2025-2026 Semester

Zhaorui(Elijah) JIANG

December 30, 2025

Outline

Recap

Positron

rpy2

Final Project

Recap

What you have learned from this course?

- Environment configuration tools: conda; pip; uv
- Computer Systems: Windows(wsl); macOS; Linux;
- Basic usage for Python
- Basic usage for R
- Essential skills for data visualization and analysis
- Advanced analysis methods: machine/deep learning

If you want to learn more useful and graceful computer science knowledge and skills in life science, welcome to select next semester's course: Environmental Bioinformatics Methods

Positron

The Data Science IDE: Positron

Positron unifies exploration and production work in one free, AI-assisted environment, empowering the full spectrum of data science in Python and R.

GitHub: <https://github.com/posit-dev/positron>

Official Website: <https://positron.posit.co>

It supports both windows and macOs computer systems.

Positron

Positron, a next-generation data science IDE

The screenshot shows the Positron IDE interface. At the top, there's a menu bar with File, Edit, View, Insert, Tools, Help, and Viewer. Below the menu is a toolbar with icons for New, Open, Save, Run, Stop, and others. The main area is divided into several panes:

- Code Editor:** Displays two Python scripts: `pandas-ex.py` and `seaborn-plot.py`. The `seaborn-plot.py` script contains code for generating plots from the diamonds dataset.
- Interpreter Session Selector:** A dropdown menu titled "Select Interpreter Session" lists "R 4.5.0" and "Python 3.12.0 (Uv_venv) Currently Selected". It also includes options for "New Interpreter Session..." and "Filter".
- Data View:** Shows the diamonds DataFrame with columns: carat, cut, color, clarity, and depth.
- Plots View:** Displays four plots:
 - A histogram of diamond prices.
 - A violin plot comparing price distribution by cut quality (Ideal, Premium, Good, Fair).
 - A scatter plot of carat vs price by clarity.
 - A violin plot comparing price distribution by color (D, E, F, G, H, I, J).
- Console:** Shows the command history for the current session.
- Terminal:** Shows the command history for the current session.
- Ports:** Shows the command history for the current session.
- Output:** Shows the command history for the current session.
- Problems:** Shows the command history for the current session.
- Debug Console:** Shows the command history for the current session.

rpy2

What is rpy2 Python package?

rpy2 is a Python package that provides interface to use R from Python

Read the Docs: <https://rpy2.readthedocs.io/en/latest>

```
1 pip install rpy2
2 pip install 'rpy2[all]'
3 # or pip install uv first and then
4 uv pip install rpy2
5 uv pip install 'rpy2[all]'
6 # Automatically match a compatible R version
7 conda install -c conda-forge rpy2 r-base
```

rpy2

Listing 1: rpy2 Basic Usage

```
1 import rpy2.robj as robjects
2 # Execute R code
3 robjects.r('''
4     x <- rnorm(100)
5     mean_x <- mean(x)
6     sd_x <- sd(x)
7 ''')
8 # Access R variables from Python
9 mean_x = robjects.r['mean_x']
10 sd_x = robjects.r['sd_x']
11 print(f"Mean: {mean_x[0]}, Standard Deviation: {sd_x[0]}")
```

Listing 2: rpy2 Dataframe Conversion

```
1 import pandas as pd
2 import rpy2.robj as ro
3 from rpy2.robj import pandas2ri
4 from rpy2.robj.packages import importr
5
6 # Activate conversion between pandas and R
7 pandas2ri.activate()
8
9 # Create pandas DataFrame
10 pdf = pd.DataFrame({
11     'a': [1, 2, 3, 4, 5],
12     'b': ['a', 'b', 'c', 'd', 'e'],
13     'c': [True, False, True, False, True]
14 })
```

rpy2

```
15  
16 # Convert to R DataFrame  
17 rdf = pandas2ri.py2rpy(pdf)  
18  
19 # Use R functions on the dataframe  
20 stats = importr('stats')  
21 result = stats.aggregate(rdf.rx2('a'), rdf.rx2('c'), ro.r('mean'))  
22  
23 # Convert result back to Python  
24 py_result = pandas2ri.rpy2py(result)
```

Dataframe the key data format during data visualization and analysis no what you are using Python or R.

rpy2

Listing 3: rpy2 ggplot Graphics

```
1 import rpy2.robj as ro
2 from rpy2.robj.packages import importr
3 from rpy2.robj import pandas2ri
4 import pandas as pd
5
6 # Activate pandas to R conversion
7 pandas2ri.activate()
8
9 # Import ggplot2 and other packages
10 ggplot2 = importr('ggplot2')
11 datasets = importr('datasets')
12
13 # Get mtcars dataset
14 mtcars = ro.r('mtcars')
```

rpy2

```
15 # For Python object
16 mtcars_pd = pandas2ri.rpy2py(mtcars)
17
18 # Create a ggplot
19 pp = ggplot2.ggplot(mtcars) +
20     ggplot2.aes_string(x='wt', y='mpg', color=
21         factor(cyl)) +
22     ggplot2.geom_point() +
23     ggplot2.ggttitle('MPG vs Weight by Cylinders')
24
25 # Display the plot
26 ro.r.X11()
27 pp.plot()
```

Final Project

DVAM Official Website: <https://github.com/PKU-EMBL/Data-Visualization-and-Analysis-Methods>

Requirements are as follows:

- The topic is up to your group's major
- Each group should be up most to 4 people
- Visualize and analysis a meaningful datasets for research
- Write detailed README file and technical report (mandatory)
- Provide a user-friendly website (optional)

Those requirements can also be seen in GitHub repository

Final Project

Please fork the original GitHub repository and pull your team's project external link with revised version README as submission.

 README  MIT license



PKU EMBL | Data Visualization and Analysis Methods

News

- [2025/12/29]:🌟 We release Course Data-Visualization-and-Analysis-Methods Community.

Students Projects

Group ID: Projects Topic: Related GitHub link with detailed README files — Team numbers

G1: Cryo-ET Analysis and Visualization Platform: [GitHub Link](#) — James, Christy, Vivek

G2:

G3:

G4: