北京大学
PEKING UNIVERSITY

PKU EMBL
Yu's Lab AI

# 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
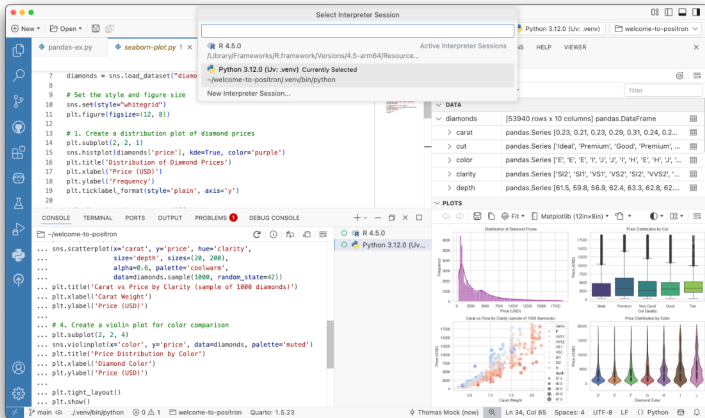
# rpy2

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

```python
import rpy2.robjects as robjects
# Execute R code
robjects.r('''
    x <- rnorm(100)
    mean_x <- mean(x)
    sd_x <- sd(x)
''')
# Access R variables from Python
mean_x = robjects.r['mean_x']
sd_x = robjects.r['sd_x']
print(f"Mean: {mean_x[0]}, Standard Deviation: {sd_x[0]}")
```

# rpy2

Listing 2: rpy2 Dataframe Conversion

```python
import pandas as pd
import rpy2.robjects as ro
from rpy2.robjects import pandas2ri
from rpy2.robjects.packages import importr

# Activate conversion between pandas and R
pandas2ri.activate()

# Create pandas DataFrame
pdf = pd.DataFrame({
    'a': [1, 2, 3, 4, 5],
    'b': ['a', 'b', 'c', 'd', 'e'],
    'c': [True, False, True, False, True]
})
```

# 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

```python
import rpy2.robjects as ro
from rpy2.robjects.packages import importr
from rpy2.robjects import pandas2ri
import pandas as pd

# Activate pandas to R conversion
pandas2ri.activate()

# Import ggplot2 and other packages
ggplot2 = importr('ggplot2')
datasets = importr('datasets')

# Get mtcars dataset
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='
            factor(cyl)') + \
21      ggplot2.geom_point() + \
22      ggplot2.ggtitle('MPG vs Weight by Cylinders
            ')
23
24  # Display the plot
25  ro.r.X11()
26  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 anlysis a meaningfule datasets for research
- Write detialed README file and technicla report (mandatory)
- Provide a user-friendly website (optional)

Those reuirements 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:

PKU EMBL
Yu's Lab