UNIVERSITÀ DI TORINO

Analisi e Visualizzazione di Reti Complesse

DV05 - (Statistical) Plotting in Python

Prof. Rossano Schifanella

Matplotlib

- **Tutorial on MatplotLib**
  - https://github.com/rschifan/avrc-2324/tree/main/dataviz/tutorials/01-matplotlib
- Other resources:
  - The Python Graph Gallery
    - https://github.com/holtzy/The-Python-Graph-Gallery
    - **400 Jupyter notebooks**
  - Scientific Visualization Book
    - **[github] [PDF]**
    - **[code examples]**

Seaborn

- Tutorial on Seaborn

Exercises (basic)

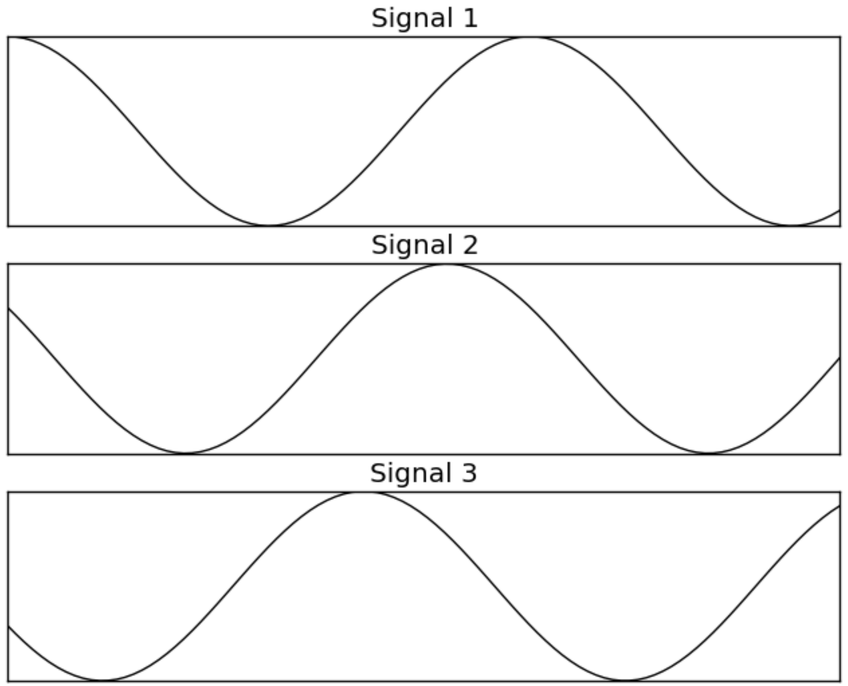
1. **Subplots with Shared Axis:** Create a 2x1 grid of subplots with shared x-axis. In the top subplot, plot y = sin(x), and in the bottom subplot, plot y = cos(x).
2. **Twin Axes:** Create a plot with twin axes (two y-axes that share the same x-axis). On the first y-axis, plot y1 = x, and on the second y-axis, plot y2 = e^x.
3. **Inset Plot:** Create a plot of the function y = sin(x) / x on the interval [-10, 10]. Add an inset plot that zooms in on the function near x = 0.
4. **Streamplot:** Generate a grid of points in the x-y plane and compute a vector (u, v) at each point that is given by (u, v) = (-y, x). Create a streamplot of this vector field.
5. **Image Display:** Load an image file and display it using imshow. Apply a colormap of your choice and display a colorbar.

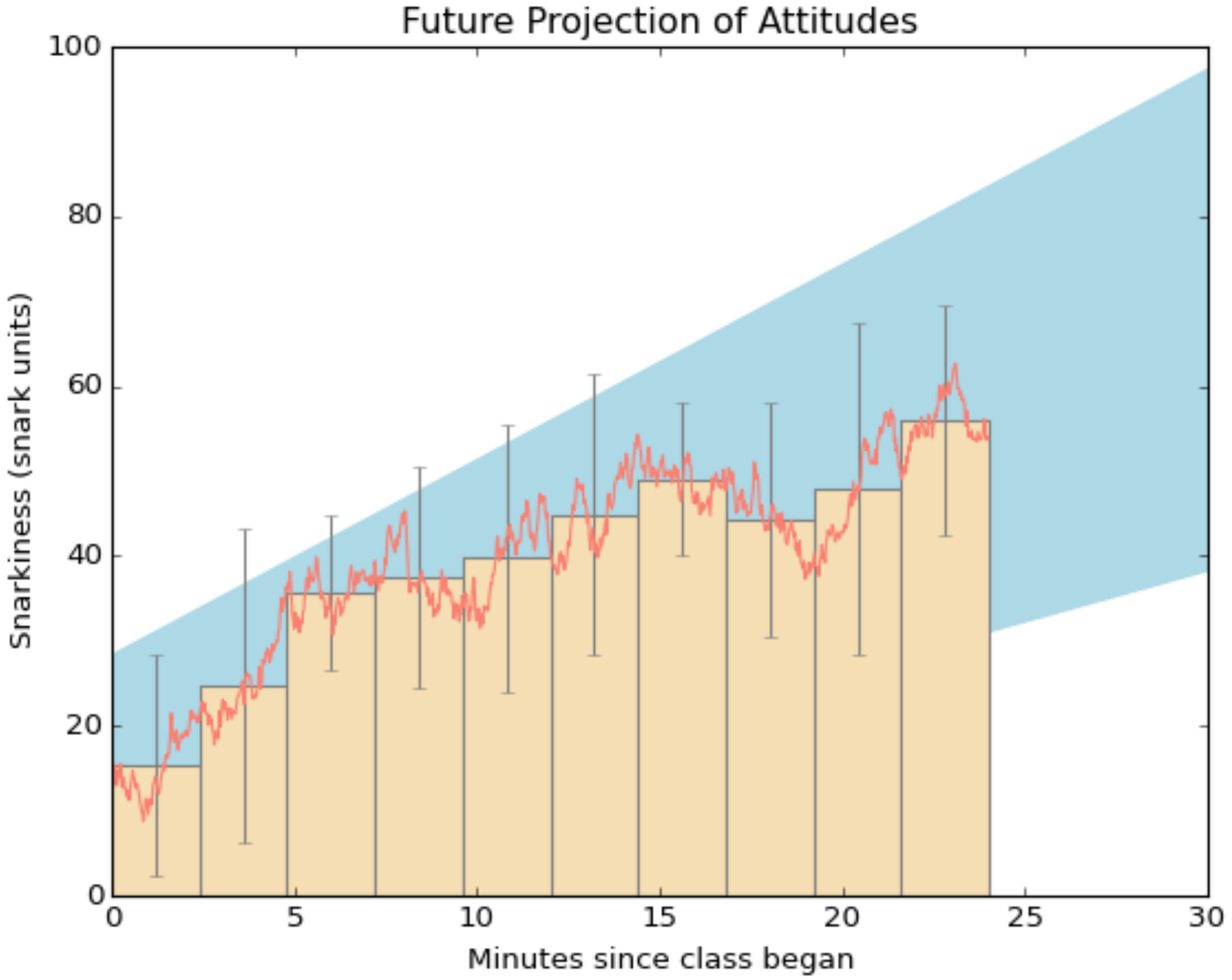
Exercises (basic)

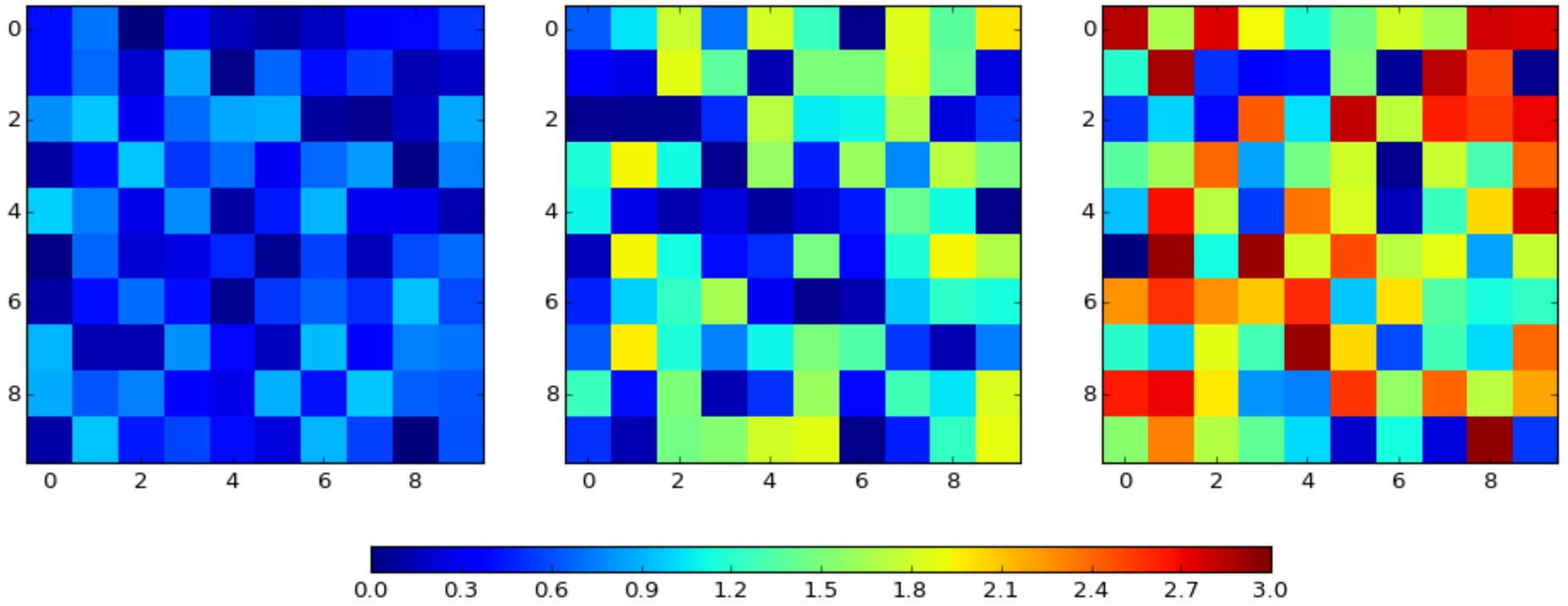
6. **Custom Legend:** Create a plot with several lines each representing a different dataset. Create a custom legend that accurately represents each line.
7. **Error Bars:** Create a line plot with error bars. Use the numpy.random.randn function to generate random error values.
8. **Filled Plots:** Create a plot of the function y = sin(x) on the interval [0, 2*pi]. Fill the area between the line and the x-axis.
9. **Histogram with Fit:** Generate 1000 random numbers from a normal distribution using numpy.random.randn and create a histogram. Fit a Gaussian function to the data and plot it on top of the histogram.
10. **Stacked Area Plot:** Create a stacked area plot using the following data: {'Category1': [1, 2, 3], 'Category2': [2, 2, 3], 'Category3': [3, 2, 3]}.

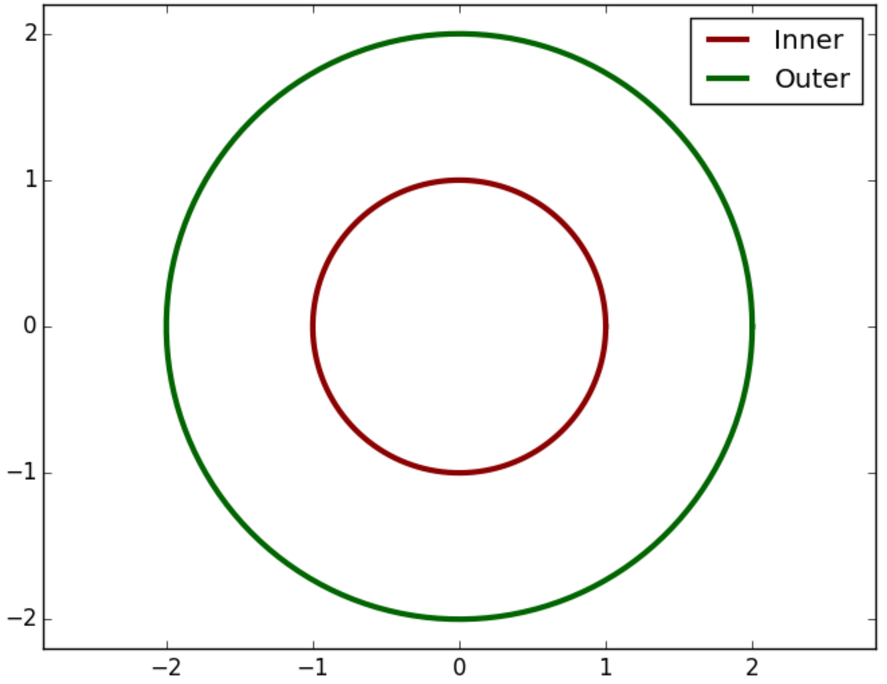
Exercise (basic)

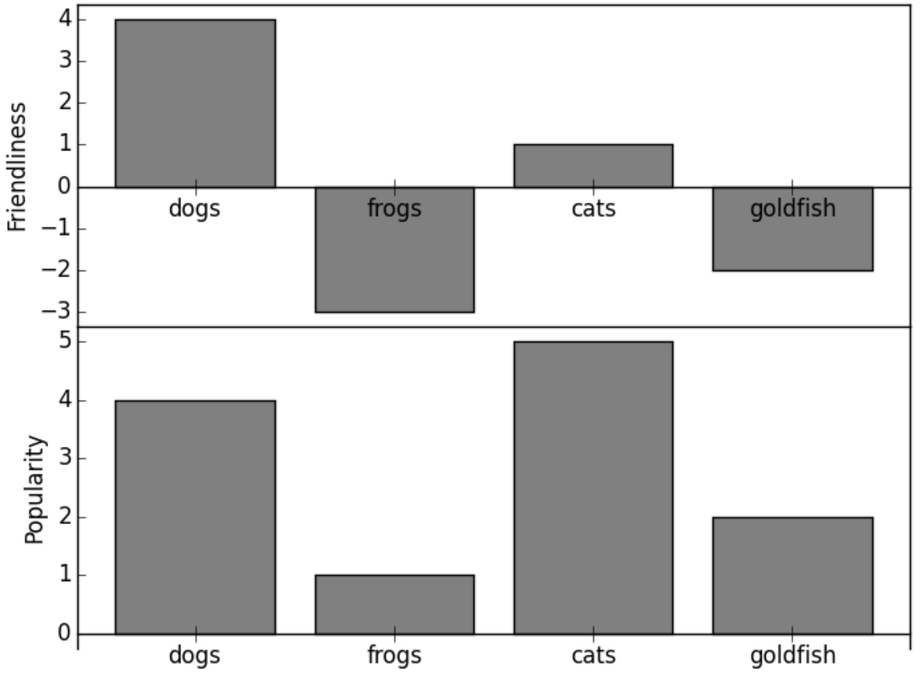
- Try to reproduce the following plots




Future Projection of Attitudes
Snarkiness (snark units)
Minutes since class began
0
20
40
60
80
100
0
5
10
15
20
25
30

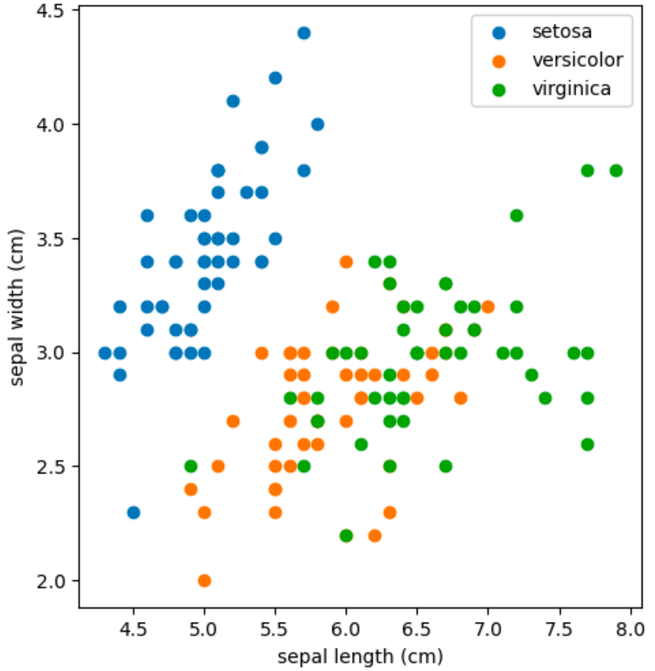

0
2
4
6
8
0.0
0.3
0.6
0.9
1.2
1.5
1.8
2.1
2.4
2.7
3.0

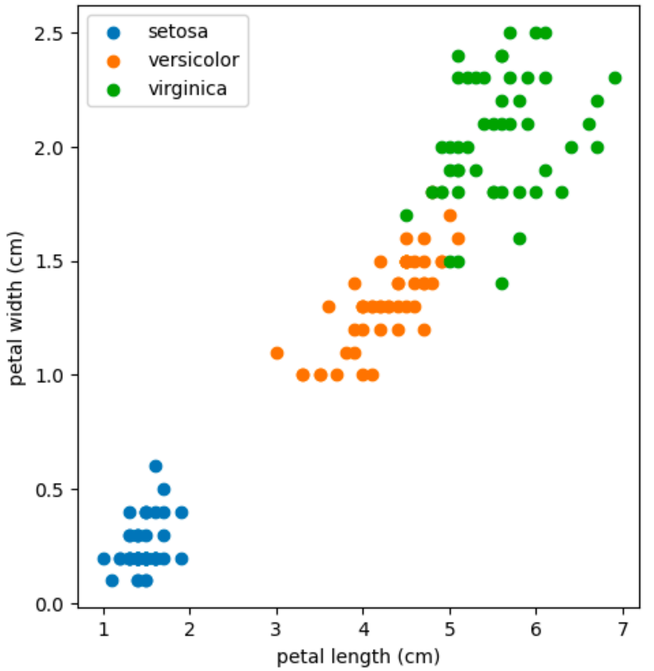

Inner
Outer
2
1
0
−1
−2
−2
−1
0
1
2

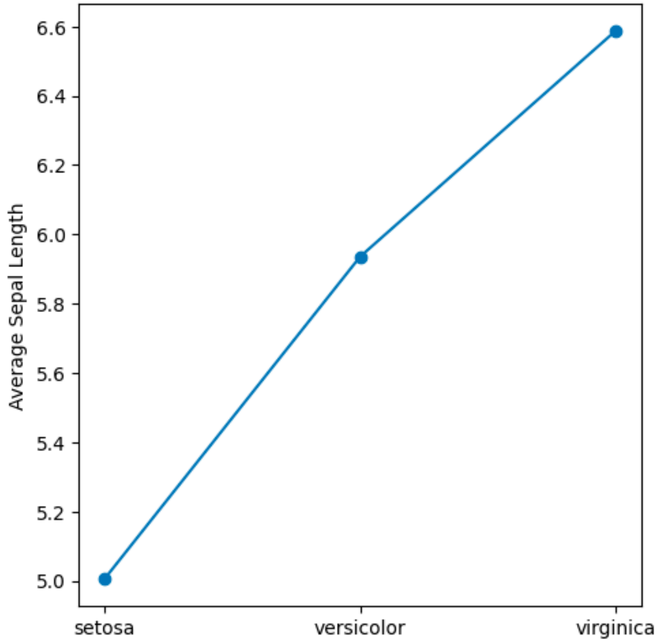

Friendliness
4
3
2
1
0
−1
−2
−3
dogs
frogs
cats
goldfish
Popularity
5
4
3
2
1
0
dogs
frogs
cats
goldfish

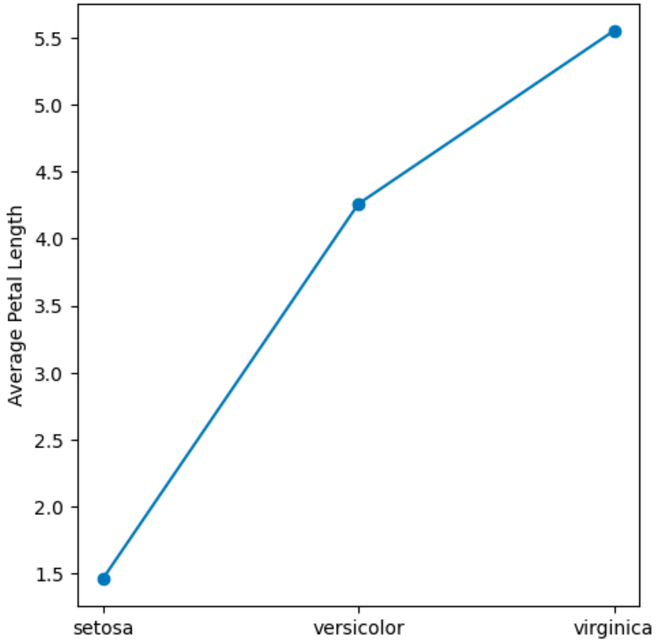
Exercise (intermediate)

- Create a 2x2 grid of subplots. In each subplot, plot the Iris dataset in a different way:
  - In the first subplot, create a scatter plot comparing sepal length and sepal width. Use different colors to represent different species.
  - In the second subplot, create a scatter plot comparing petal length and petal width. Again, use different colors for different species.
  - In the third subplot, create a line plot showing the average sepal length for each species.
  - In the fourth subplot, create a line plot showing the average petal length for each species.
- Remember to add appropriate labels and titles to each subplot.


setosa
versicolor
virginica
sepal width (cm)
sepal length (cm)
4.5
4.0
3.5
3.0
2.5
2.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0

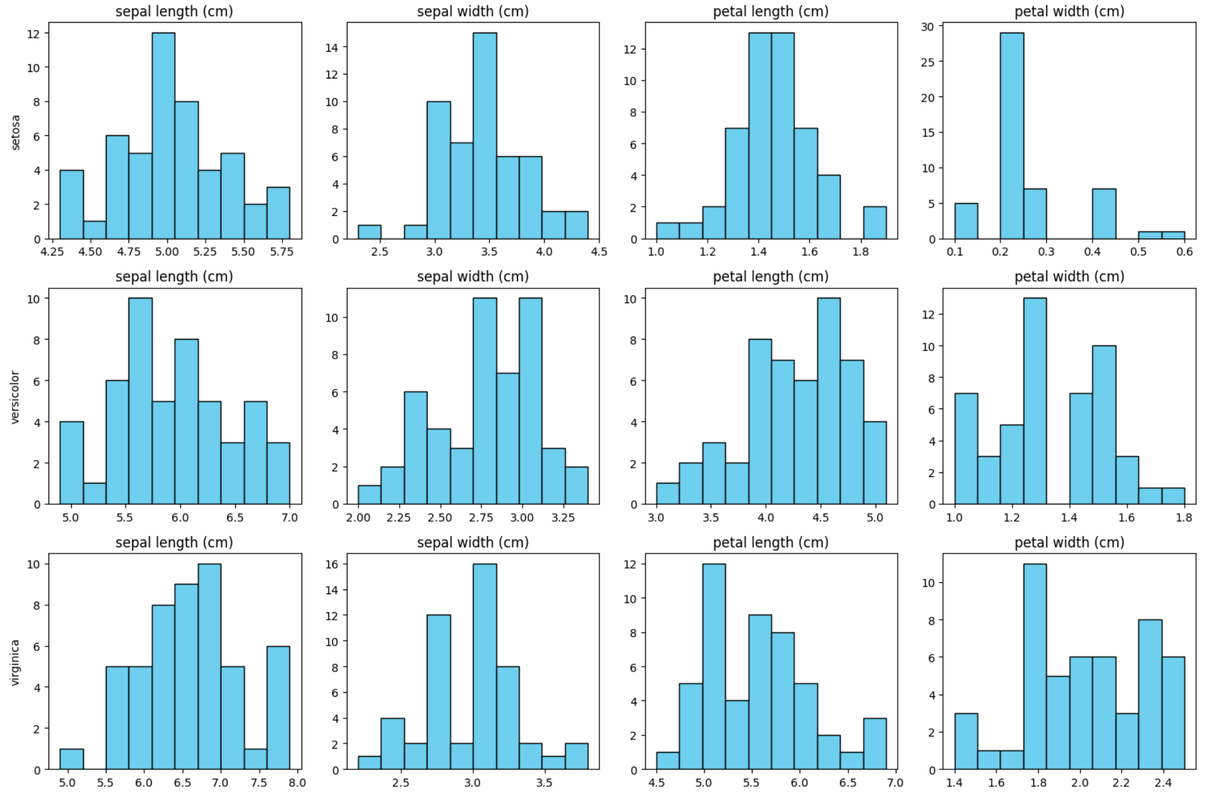

setosa
versicolor
virginica
petal width (cm)
petal length (cm)
2.5
2.0
1.5
1.0
0.5
0.0
1
2
3
4
5
6
7


Average Sepal Length
6.6
6.4
6.2
6.0
5.8
5.6
5.4
5.2
5.0
setosa
versicolor
virginica


Average Petal Length
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
setosa
versicolor
virginica

Exercise (intermediate)

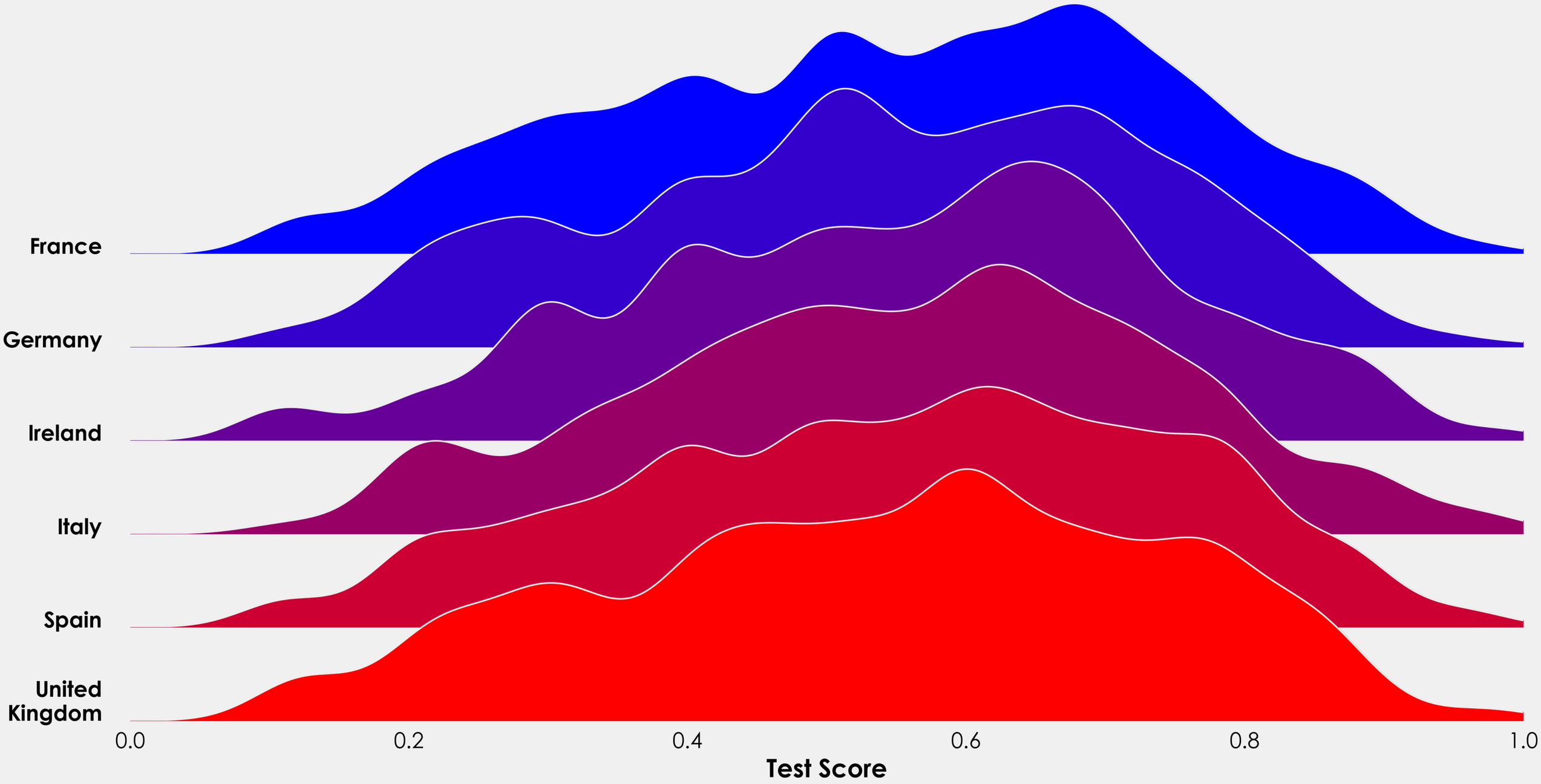
- Create a 3x4 grid of subplots, where each row represents a different species from the Iris dataset, and each column represents a different feature (sepal length, sepal width, petal length, petal width).
- In each subplot, create a histogram of the feature values for the corresponding species.



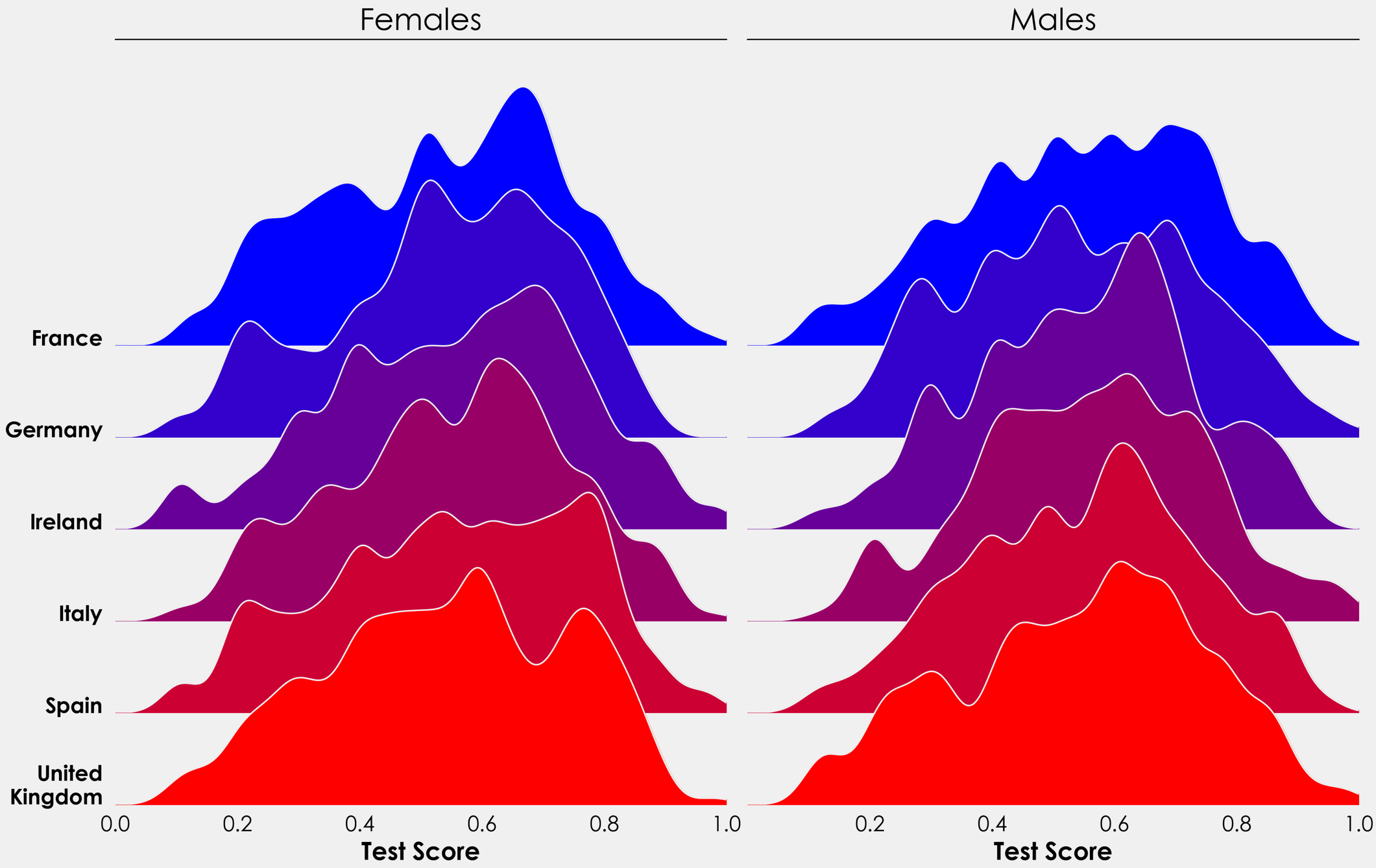
Exercise (advanced)

- This exercise will outline how we can leverage **gridspec** to create **ridgeplots** in Matplotlib.
- data source:
  - https://github.com/petermckeever/mock-data/blob/master/datasets/mock-european-test-results.csv
- To create the kde plots there are several alternatives:
  - **plot.kde** from **pandas**
  - **KernelDensity** module from **sklearn.neighbors** to compute the distribution and then **plot** and **fill_between** from **matplotlib**
- Replicate the exercise using **seaborn**

Distribution of Aptitude Test Results from 18 – 24 year-olds



Distribution of Aptitude Test Results from 18 – 24 year-olds

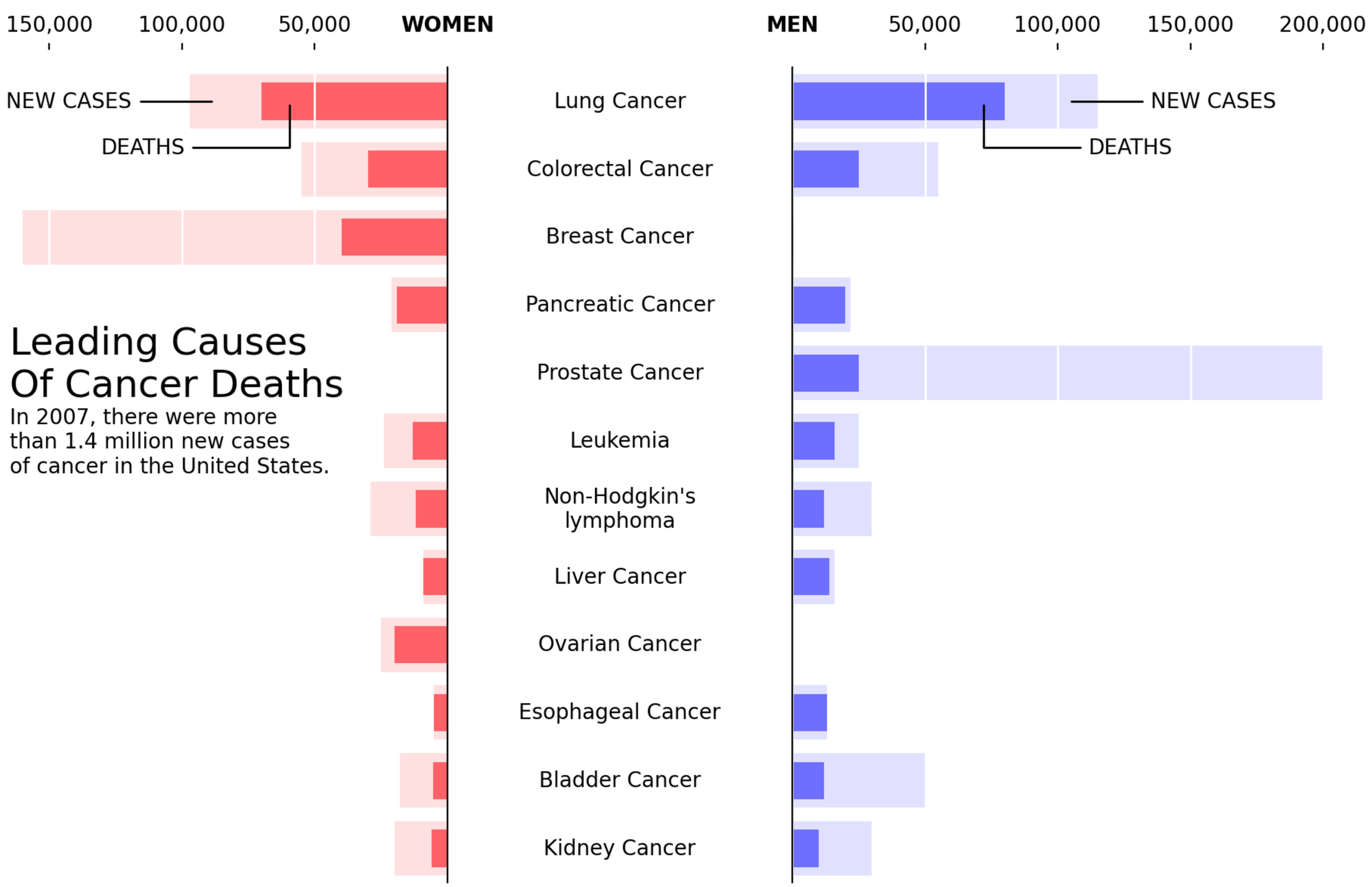


Exercise (advanced)

- Earth's temperatures are rising and nothing shows this in a simpler, more approachable graphic than the “**Warming Stripes**”. Introduced by Prof. Ed Hawkins they show the temperatures either for the global average or for your region as colored bars from blue to red for the last 170 years, available at #ShowYourStripes.
- The stripes have since become the logo of the **Scientists for Future**.
- HadCRUT4 dataset, published by the Met Office
- Replicate this using the **PatchCollection** and **Rectangle** objects

Exercise (advanced)

- This is a remake of a figure that was originally published in the New York Times (NYT) in 2007.
- The data is made of four series (men deaths/cases, women deaths/cases) of cancer
  - You can randomly assign the values to the classes
- This plot if entertaining for a general audience of a journal, it shows a serious drawback, which one?
  - Can you solve it?


150,000
100,000
50,000
WOMEN
MEN
50,000
100,000
150,000
200,000
NEW CASES
DEATHS
Lung Cancer
Colorectal Cancer
Breast Cancer
Pancreatic Cancer
Prostate Cancer
Leukemia
Non-Hodgkin's lymphoma
Liver Cancer
Ovarian Cancer
Esophageal Cancer
Bladder Cancer
Kidney Cancer
NEW CASES
DEATHS
Leading Causes Of Cancer Deaths
In 2007, there were more than 1.4 million new cases of cancer in the United States.

Material

Matplotlib

- Official Matplotlib User guide
- Official Matplotlib Tutorials
- Official Matplotlib Examples
- Matplotlib Cheatsheet (very useful!)

Seaborn

- Official Seaborn Tutorial
- Official Seaborn Gallery

Questions?


@rschifan


rossano.schifanella@unito.it


http://www.di.unito.it/~schifane