

```

1 import matplotlib.pyplot as plt
2 import pandas as pd
3 import numpy as np
4 import seaborn as sns
5
6 titanic = pd.read_csv('/content/sample_data/train.csv')

```

Exercise 1 - Plotting two quantitative variables (value attributes) as a scatterplot in Python

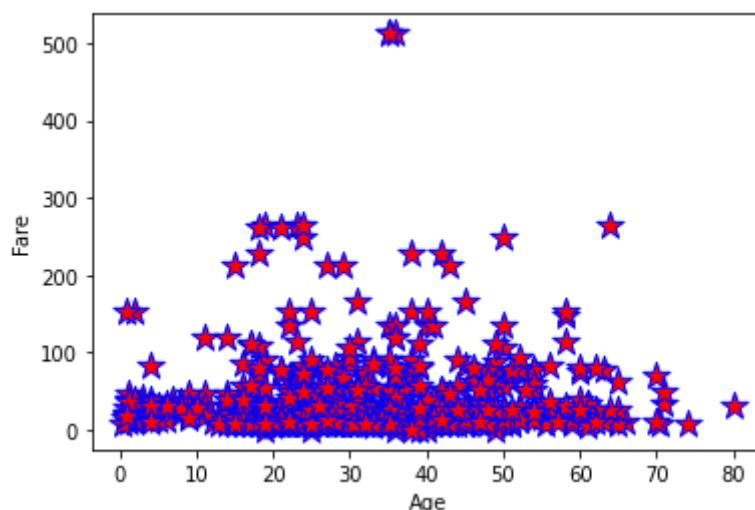
1  2  2 

- Plot the two quantitative variables (attribute values) from the titanic dataset (age and fare) and create a scatterplot to visualize the relationship between the two variables.
- Focus on making effective plots, keeping in mind what you have learned about in lecture.
- Reflect and discuss how the data is represented visually and why or why not you think it is effective. Explicitly state and comment on the marks and channels used in your visual encoding, the tasks that are well supported by it, and the scale of the data in terms of number of observations, and the number of levels of each categorical or ordered attribute (if categorical or ordered attributes are present).

```

1 x1 = titanic['Age']
2 y1 = titanic['Fare']
3 plt.scatter(x1, y1, c = "red",
4             linewidths = 1,
5             marker = "*",
6             edgecolor = "blue",
7             s = 200)
8 plt.title("Scatter Plot")
9 plt.xlabel("Age")
10 plt.ylabel("Fare")
11 plt.show()

```



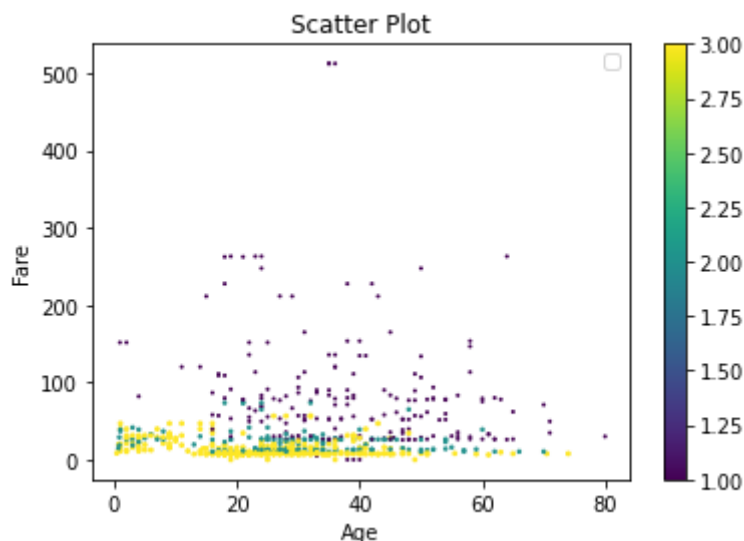
Exercise 2 (optional) - Plotting two quantitative variables (value attributes) and one categorical variable (key attribute) as a scatterplot in Python

1

- Plot the two quantitative variables (value attributes) from the titanic dataset (age and fare) and create a scatterplot where the points are coloured based upon a categorical variable (key attribute) you are interested in.
- Focus on making effective plots, keeping in mind what you have learned about in lecture.
- Reflect and discuss how the data is represented visually and why or why not you think it is effective. Explicitly state and comment on the marks and channels used in your visual encoding, the tasks that are well supported by it, and the scale of the data in terms of number of observations, and the number of levels of each categorical or ordered attribute.

```
1 plt.scatter(titanic['Age'], titanic['Fare'], c=titanic['Pclass'],s=titanic['Pclass'])
2 plt.title("Scatter Plot")
3 plt.xlabel('Age')
4 plt.ylabel('Fare')
5 plt.colorbar()
6 plt.legend()
7 plt.show()
```

WARNING:matplotlib.legend:No handles with labels found to put in legend.



Exercise 3 - Plotting two quantitative variables (value attributes) as a line-chart in Python

1 2 2

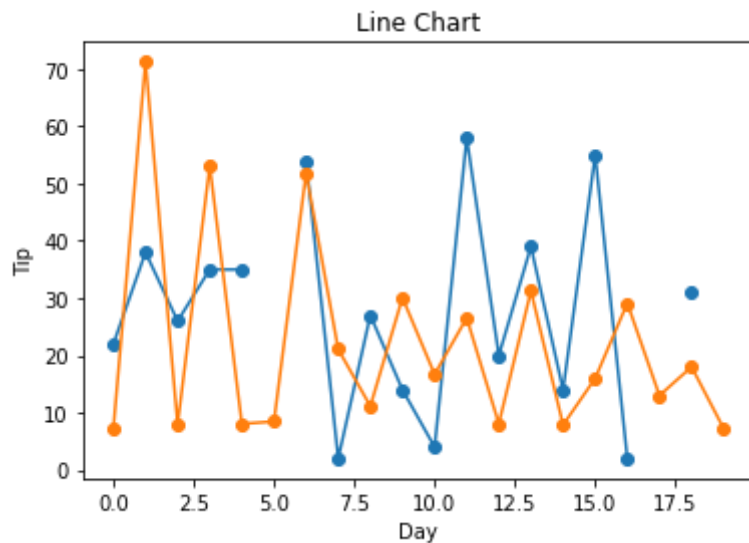
- Choose **one** country of interest from one of the these 10 Gapminder datasets and plot population versus year as a line-chart.
- Focus on making effective plots, keeping in mind what you have learned about in lecture.
- Reflect and discuss how the data is represented visually and why or why not you think it is effective. Explicitly state and comment on the marks and channels used in your visual encoding, the tasks that are well supported by it, and the scale of the data in terms of number of observations, and the number of levels of each categorical or ordered attribute.

```
1 titan=titanic.head(20)
2 plt.plot(titan['Age'],marker='o')
3 plt.plot(titan['Fare'],marker='o')
```

```

4 plt.title("Line Chart")
5 plt.xlabel('Day')
6 plt.ylabel('Tip')
7 plt.show()

```



Exercise 4 - Plotting a single variable in Python as a histogram

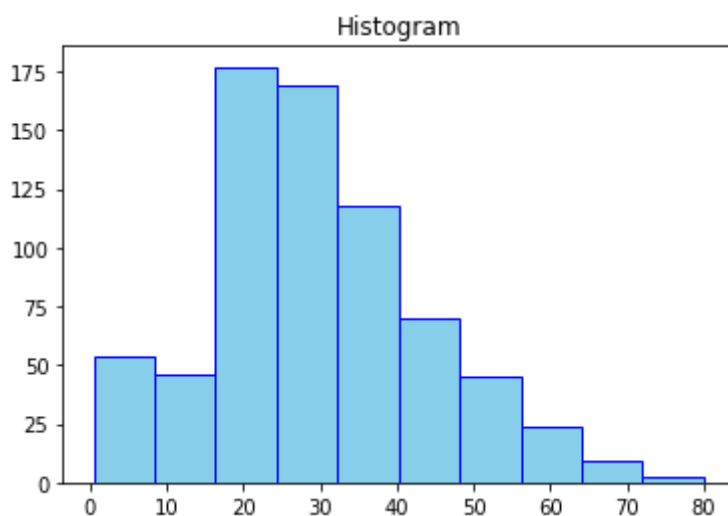
1 2 2

- Choose one variable from the titanic dataset you are interested in and plot the data as a histogram.
- Focus on making effective plots, keeping in mind what you have learned about in lecture.
- Reflect and discuss how the data is represented visually and why or why not you think it is effective. Explicitly state and comment on the marks and channels used in your visual encoding, the tasks that are well supported by it, and the scale of the data in terms of number of observations, and the number of levels of each categorical or ordered attribute.




```

1 plt.hist(titanic['Age'], color = "skyblue", ec="blue")
2 plt.title("Histogram")
3 plt.show()

```

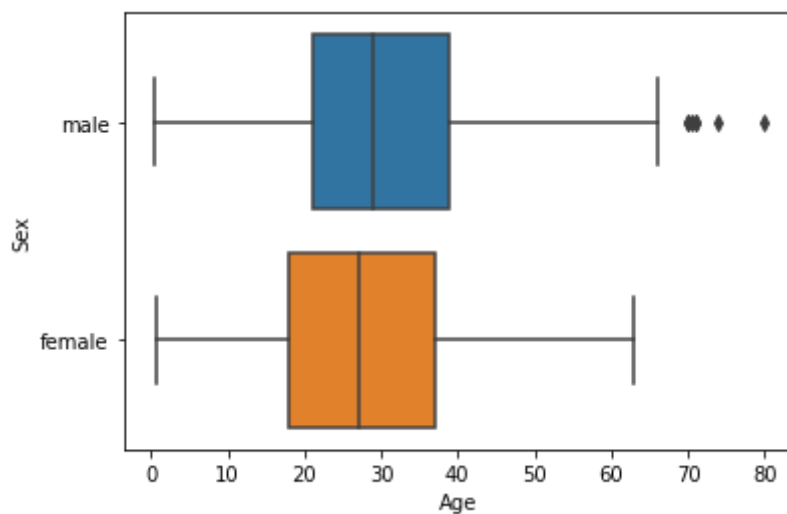


Exercise 5 - Plotting a single one quantitative variable (value) stratified by one categorical variable (key) in Python

1  2  2 

- Choose a quantitative variable (value attributes) from the titanic dataset (either fare or age) and one categorical variable (key attribute) you are interested in. Create boxplots to visualize the relationship between the two variables.
- Focus on making effective plots, keeping in mind what you have learned about in lecture.
- Reflect and discuss how the data is represented visually and why or why not you think it is effective. Explicitly state and comment on the marks and channels used in your visual encoding, the tasks that are well supported by it, and the scale of the data in terms of number of observations, and the number of levels of each categorical or ordered attribute.

```
1 sns.boxplot(x='Age', y="Sex",data=titanic );  
2 plt.show()
```



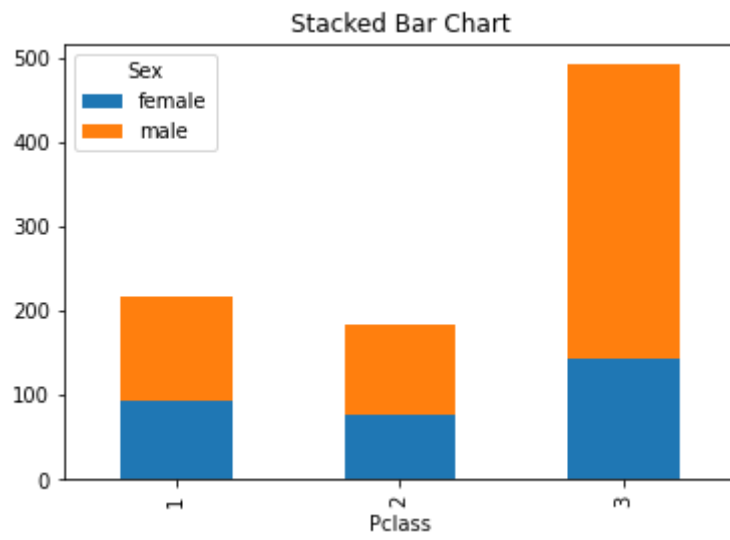
Exercise 6 - Plotting two categorical variables (keys) in Python

1  2  2 

- Pick two categorical variables/attributes from the titanic dataset and create a stacked bar-chart to visualize the relationship between the two variables/attributes.
- Focus on making effective plots, keeping in mind what you have learned about in lecture.
- Reflect and discuss how the data is represented visually and why or why not you think it is effective. Explicitly state and comment on the marks and channels used in your visual encoding, the tasks that are well supported by it, and the scale of the data in terms of number of observations, and the number of levels of each categorical or ordered attribute.

```
1 pd.crosstab(titanic['Pclass'],titanic['Sex']).plot(kind='bar', stacked='True')  
2 plt.title('Stacked Bar Chart')
```

Text(0.5, 1.0, 'Stacked Bar Chart')



[Colab paid products](#) - [Cancel contracts here](#)

✓ 0s completed at 5:57 PM

