**Семинар 6**

**Работа с библиотекой Matplotlib**

**Теоретическая часть**

**Терминология**

Figure представляет собой законченное окно или страницу, на которой построен график.

Axes – это область, на которой отображаются данные. Это может быть ось X, ось Y и т.д.

Spines – это линии, которые соединяют точки Axes.

**Линейный график**

Мы начнем с очень простого примера построения графика. Мы просто будем использовать два списка Python в качестве источника данных для точек графика. Напишем для этого фрагмент кода:

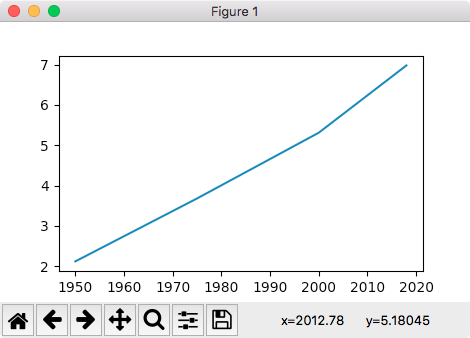
import matplotlib.pyplot as plt

year = [1950, 1975, 2000, 2018]

population = [2.12, 3.681, 5.312, 6.981]

plt.plot(year, population)

plt.show()



...

plt.xlabel('Year')

plt.ylabel('Population')

plt.title('World Population')

**Диаграмма рассеяния**

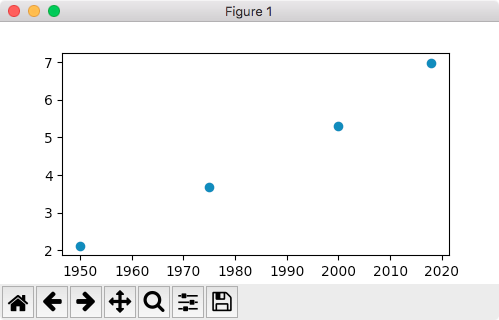
Вышеупомянутый график показывал точки, которые фактически не были переданы в массиве, поскольку он показывает линию. Что, если мы хотим видеть только фактические точки на графике? Диаграмма разброса достигает этого:

import matplotlib.pyplot as plt

year = [1950, 1975, 2000, 2018]

population = [2.12, 3.681, 5.312, 6.981]

plt.scatter(year, population) plt.show()



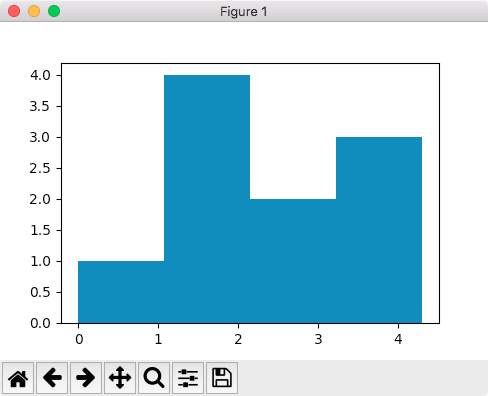
Гистограммы

В то время как графики информируют нас о том, как меняются наши данные, гистограмма описывает, как наши данные распределяются. Чем больше значений в диапазоне, тем выше полоса диапазона. Мы используем функцию hist() для построения гистограммы. У него есть 2 важных параметра: список значений для построения; количество диапазонов для распределения этих точек. Продемонстрируем это с помощью фрагмента кода:

import matplotlib.pyplot as plt

values = [0, 1.2, 1.3, 1.9, 4.3, 2.5, 2.7, 4.3, 1.3, 3.9]

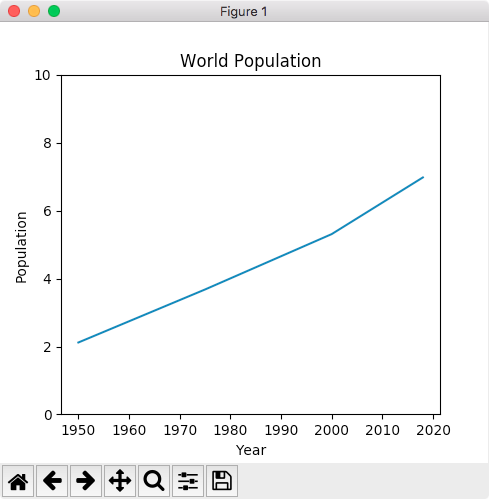
plt.hist(values, bins = 4) plt.show()



**Настройка Plo**t

На первом графике Line видно, что ось Y не начинается с 0. Можно изменить это:

plt.yticks([0, 2, 4, 6, 8, 10])



**Рисование нескольких кривых**

Распространено рисование нескольких кривых на одном графике для сравнения:

import numpy as np

import matplotlib.pyplot as plt

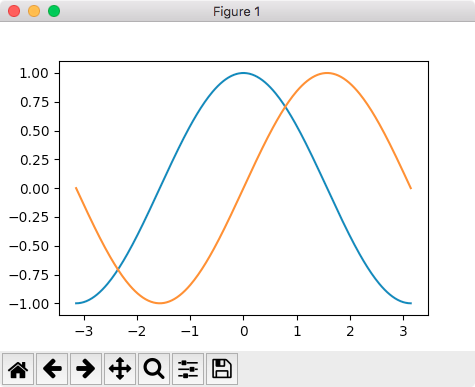
X = np.linspace(-np.pi, np.pi, 256, endpoint=True)

cos, sin = np.cos(X), np.sin(X)

plt.plot(X, cos)

plt.plot(X, sin)

plt.show()



**Изменение цвета и добавление надписи на графике**

Возможно изменить цвет кривых и подписать их:

import numpy as np

import matplotlib.pyplot as plt

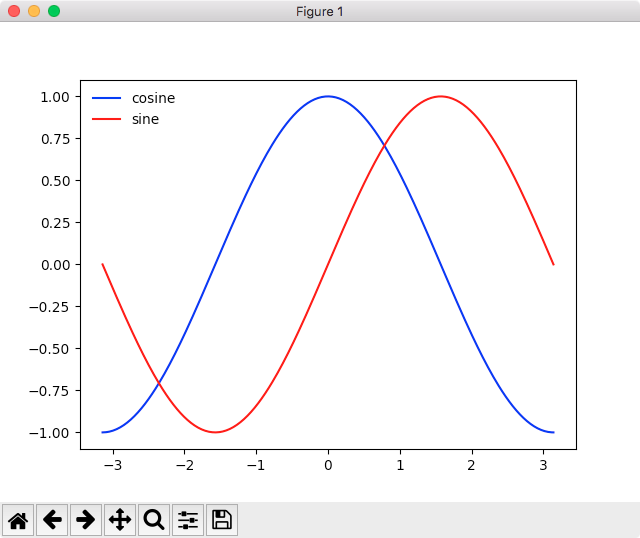
X = np.linspace(-np.pi, np.pi, 256, endpoint=True)

cos, sin = np.cos(X), np.sin(X)

plt.plot(X, cos, color='blue', label="cosine")

plt.plot(X, sin, color='red', label="sine")

plt.legend(loc='upper left', frameon=False) plt.show()



**Создание тепловых карт**

import numpy as np

import numpy.random

import matplotlib.pyplot as plt

# Create data

temperature = np.random.randn(4096)

anger = np.random.randn(4096)

# Create heatmap

heatmap, xedges, yedges = np.histogram2d(temperature, anger, bins=(64,64))

extent = [xedges[0], xedges[-1], yedges[0], yedges[-1]]

# Plot heatmap

plt.clf()

plt.ylabel('Anger')

plt.xlabel('Temp')

plt.imshow(heatmap, extent=extent)

plt.show()

**Задание.**

1. Напишите программу, которая будет создавать читать данные из файла CSV (по вариантам) и выводить их на экран в виде линейного графика. Параметры графика определяются в соответствии с вариантом. Вывести «легенду» к графикам. Отобразить на графике сетку.

*Четные варианты (по журналу)*

Название файла – passengers.csv

Построить графики линейные по столбцам (одновременно на одном рисунке): passengers\_avia и passengers\_jd

Постройте график распределения пассажиров по каждому году (гистограмма)

Цвета графиков: красный, зеленый.

Тип графиков: сплошная линия, точки.

Подписать оси: количество пассажиров и года.

*Нечетные варианты (по журналу)*

Название файла – oscar\_age\_female.csv, oscar\_age\_male.csv

Построить графики линейные по столбцам (одновременно на одном рисунке): Age (из обоих файлов)

Постройте график распределения возраста по каждому году (гистограмма)

Цвета графиков: красный, зеленый.

Тип графиков: сплошная линия, точки.

Подписать оси: Возраст, года.

2. Анализ размера цветков ириса (все варианты). Построить графики по данным файла iris.csv.

- Соотношение размеров чашелистиков и лепестков (диаграмма рассеяния)

- Соотношение размеров между чашелистиками и лепестками разных видов ириса (классифицированная диаграмма рассеяния)

- Распределение размеров чашелистиков и лепестков разных видов цветов ириса (гистограмма или прямоугольная диаграмма)

3. Анализ подсказок ресторана (все варианты). Построить графики по данным файла tips.csv.

- Связь между чаевыми и общим потреблением (диаграмма разброса)

- В будни и выходные дни, когда клиенты дают более щедрые чаевые (гистограммы)

- Влияние сочетания пола и курения на чаевые (гистограмма)

4. Для файла devices.csv постройте тепловую карту по столбцам. Оси подписать.