Matplotlib

Introduction to Matplotlib and Setting Things Up

Democratizing Data Science Learning

DPhi

Learning Objectives

- Introduction to Matplotlib
- Advantages
- Installation and Importing Matplotlib
- Matplotlib Chart
- Importing Libraries
- About the Dataset

Introduction

- Matplotlib is the most popular Python plotting library.
- It is a comprehensive library for creating static, animated, and interactive visualizations in Python.
- It is useful for those working with Python and NumPy.
- It is a low-level library with a Matlab like interface which offers lots of freedom at the cost of having to write more code.

Advantages

- extremely powerful
- fast and efficient
- open source tool
- works well with many operating systems and graphic backends
- high-quality graphics and plots
- ability to print and view a range of graphs
- large community support and cross-platform support
- full control over graph or plot styles

Installation

- Before Matplotlib's plotting functions can be used, Matplotlib needs to be installed.

- The Anaconda distribution of Python as well as Google Colab come with Matplotlib pre-installed and no further installation steps are necessary.

- However, if you're not using any of those, you can install it by running a simple pip command in your terminal:

```
pip install matplotlib
```

Pyplot

The most used module of Matplotib is Pyplot which is matplotlib's plotting framework.

Each pyplot function makes some change to a figure: e.g., creates a figure, creates a plotting area in a figure, plots some lines in a plotting area, decorates the plot with labels, etc.

Importing matplotlib

There are a lot of modules in Matplotlib but for plotting purposes, we only import pyplot.

Just as we use the np shorthand for NumPy and the pd shorthand for Pandas, we will use the standard shorthand plt for the Matplotlib import:

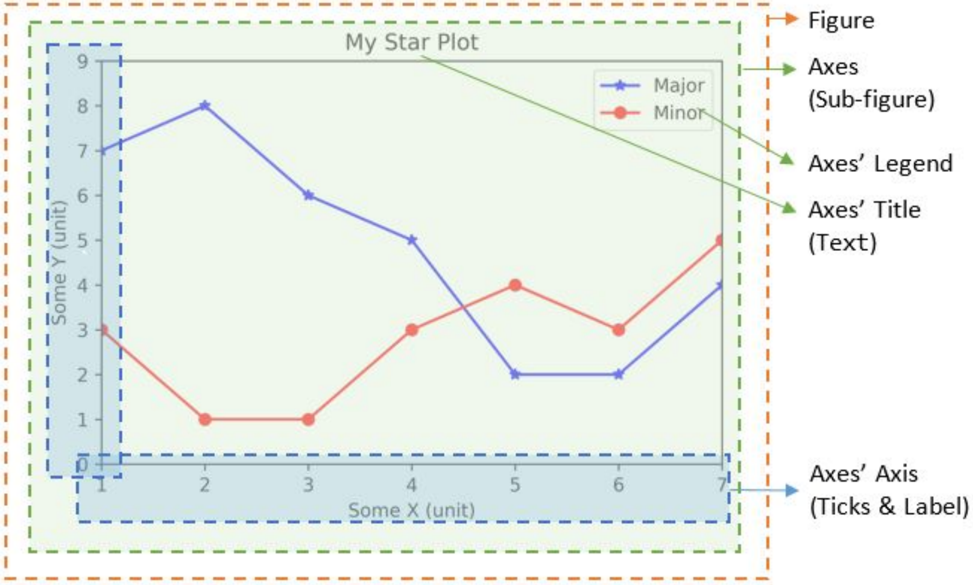
```
import matplotlib.pyplot as plt
```

Matplotlib Chart

Matplotlib’s charts are made up of two main components:

1. **The axes:** the lines that delimit the area of the chart
2. **The figure:** where we draw the axes, titles and elements that come out of the area of the axes.

PS. Unlike the plural of axis, axes in Matplotlib is actually the whole rectangular box containing the axis lines, ticks and labels.

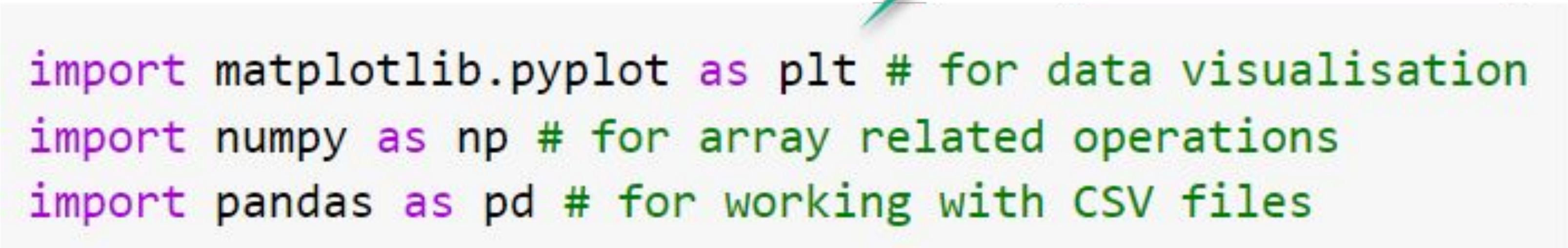


Importing libraries

All Python capabilities are not loaded to our working environment by default (even they are already installed in your system). So, we import each and every library that we want to use.

We'll import numpy, pandas and matplotlib with their respective aliases: np, pd and plt.

Importing matplotlib, numpy and pandas as plt, np and pd



```
import matplotlib.pyplot as plt # for data visualisation
import numpy as np # for array related operations
import pandas as pd # for working with CSV files
```

About the Dataset

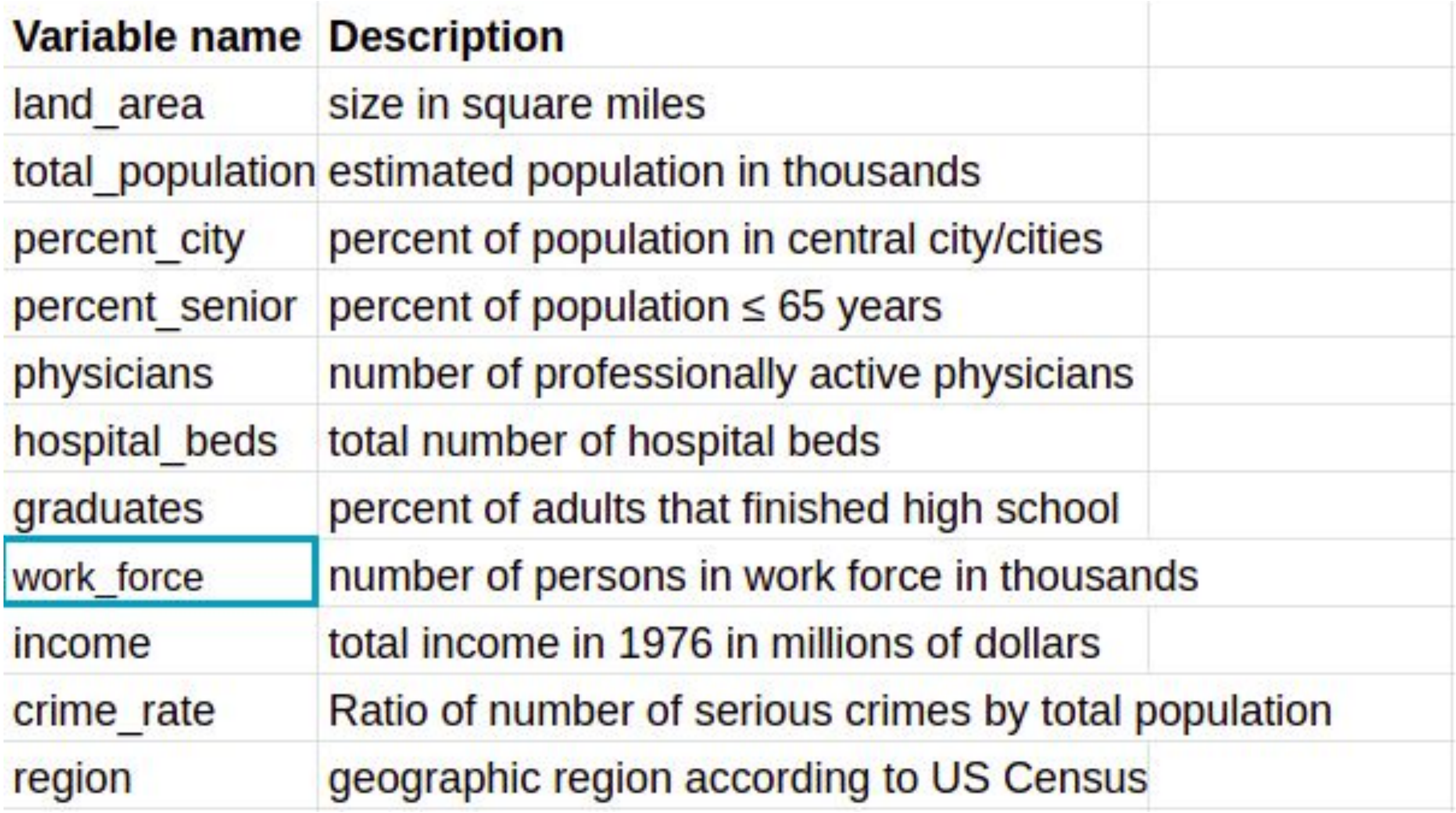
Dataset Name: Standard Metropolitan Dataset

Dataset Link:
https://raw.githubusercontent.com/dphi-official/Datasets/master/Standard_Metropolitan_Areas_Data-data.csv

Dataset Description:

- It contains data of 99 standard metropolitan areas in the US i.e 99 row entries.
- The dataset provides information on 11 variables for each area for the period 1976-1977.
- The areas have been divided into 4 geographic regions: 1=North-East, 2=North-Central, 3=South, 4=West.
- The variables provided are listed in the table in the next slide:

Dataset Description



Variable name	Description
land_area	size in square miles
total_population	estimated population in thousands
percent_city	percent of population in central city/cities
percent_senior	percent of population ≤ 65 years
physicians	number of professionally active physicians
hospital_beds	total number of hospital beds
graduates	percent of adults that finished high school
work_force	number of persons in work force in thousands
income	total income in 1976 in millions of dollars
crime_rate	Ratio of number of serious crimes by total population
region	geographic region according to US Census

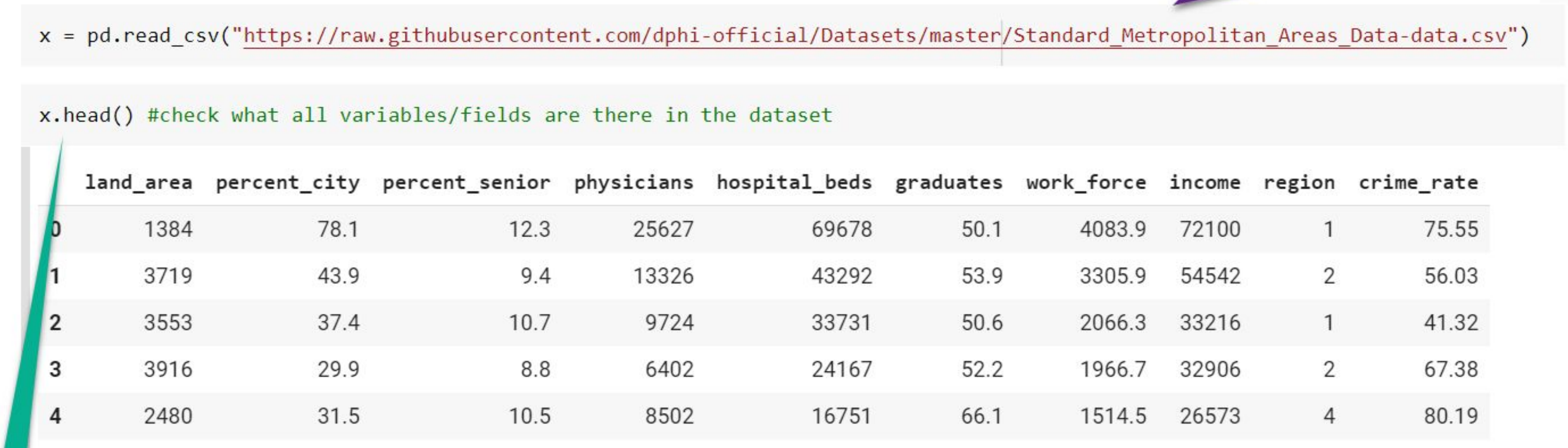
Objective

The problem objective is to find the crime rate of each area.

Loading the dataset

For loading the dataset, we'll use the Pandas read_csv method.

Link of the dataset that has been uploaded on GitHub



```
x = pd.read_csv("https://raw.githubusercontent.com/dphi-official/Datasets/master/Standard_Metropolitan_Areas_Data-data.csv")
```

```
x.head() #check what all variables/fields are there in the dataset
```

	land_area	percent_city	percent_senior	physicians	hospital_beds	graduates	work_force	income	region	crime_rate
0	1384	78.1	12.3	25627	69678	50.1	4083.9	72100	1	75.55
1	3719	43.9	9.4	13326	43292	53.9	3305.9	54542	2	56.03
2	3553	37.4	10.7	9724	33731	50.6	2066.3	33216	1	41.32
3	3916	29.9	8.8	6402	24167	52.2	1966.7	32906	2	67.38
4	2480	31.5	10.5	8502	16751	66.1	1514.5	26573	4	80.19

To display the first 5 rows of the dataset

Slide Download Link

You can download these slides from the below link:

https://docs.google.com/presentation/d/1z1WEMPuYtrBUOmfJG8MvK-kT45OSt14RLq4JtwnqPrE/edit?usp=sharing

That’s it for this unit. Thank you!

Feel free to post any queries on Discuss.

DPhi