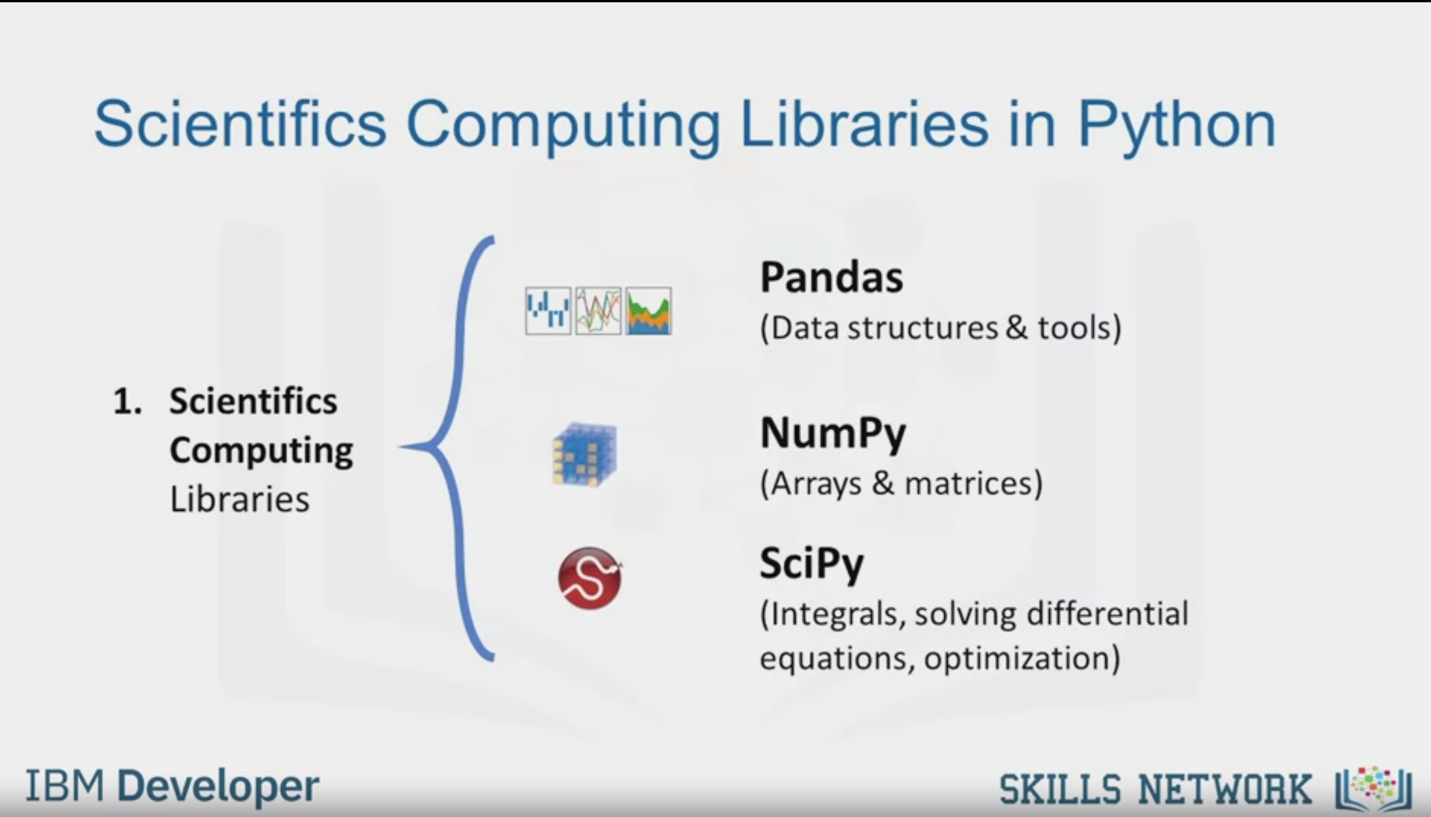


Python Packages for

Data Science

IBM Developer

SKILLS NETWORK



Scientific Computing Libraries in Python

1. Scientifics

Computing

Libraries

Pandas

(Data structures & tools)

NumPy

(Arrays & matrices)

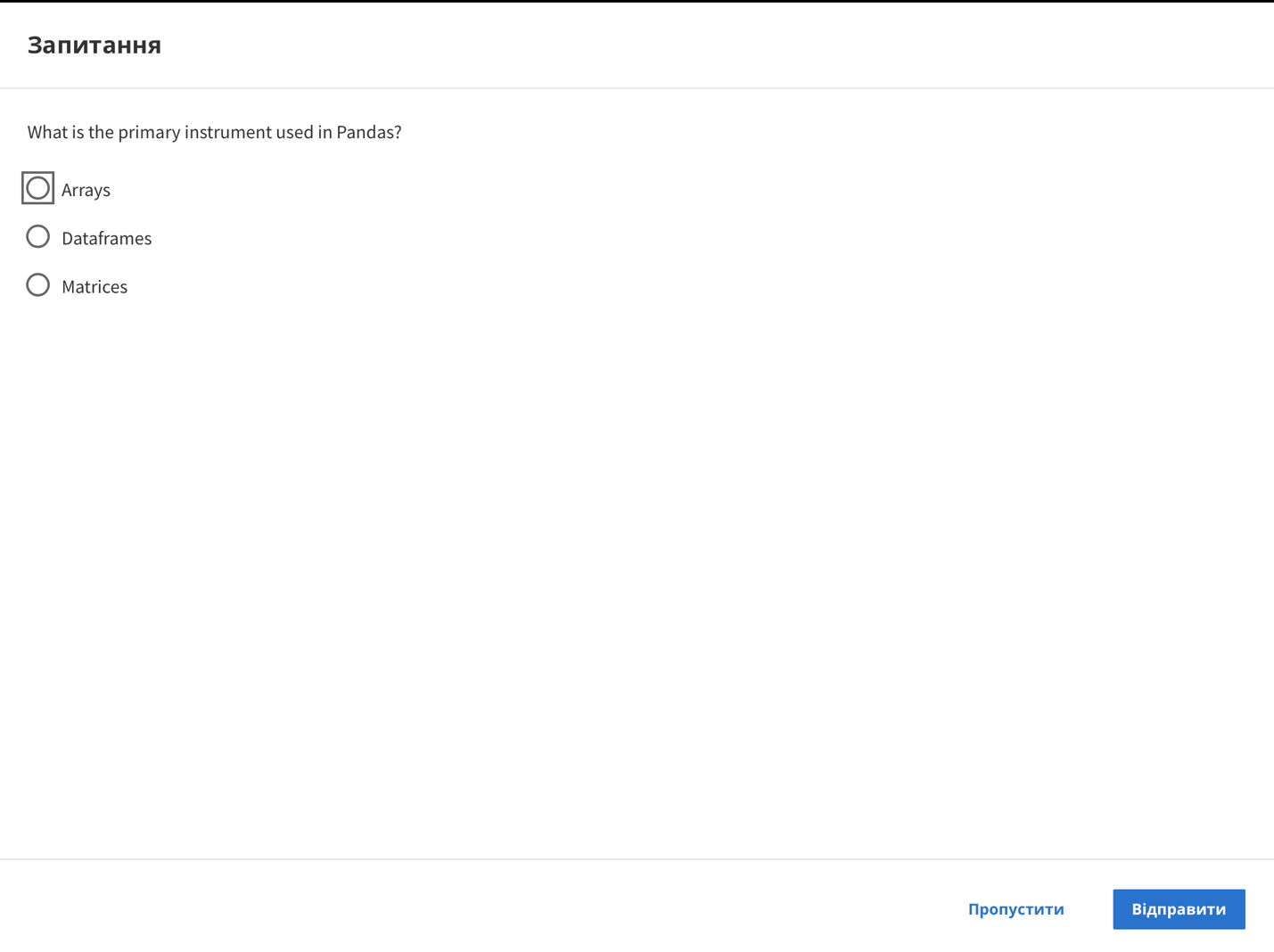
SciPy

(Integrals, solving differential

equations, optimization)

IBM Developer

SKILLS NETWORK



Question

What is the primary instrument used in Pandas?

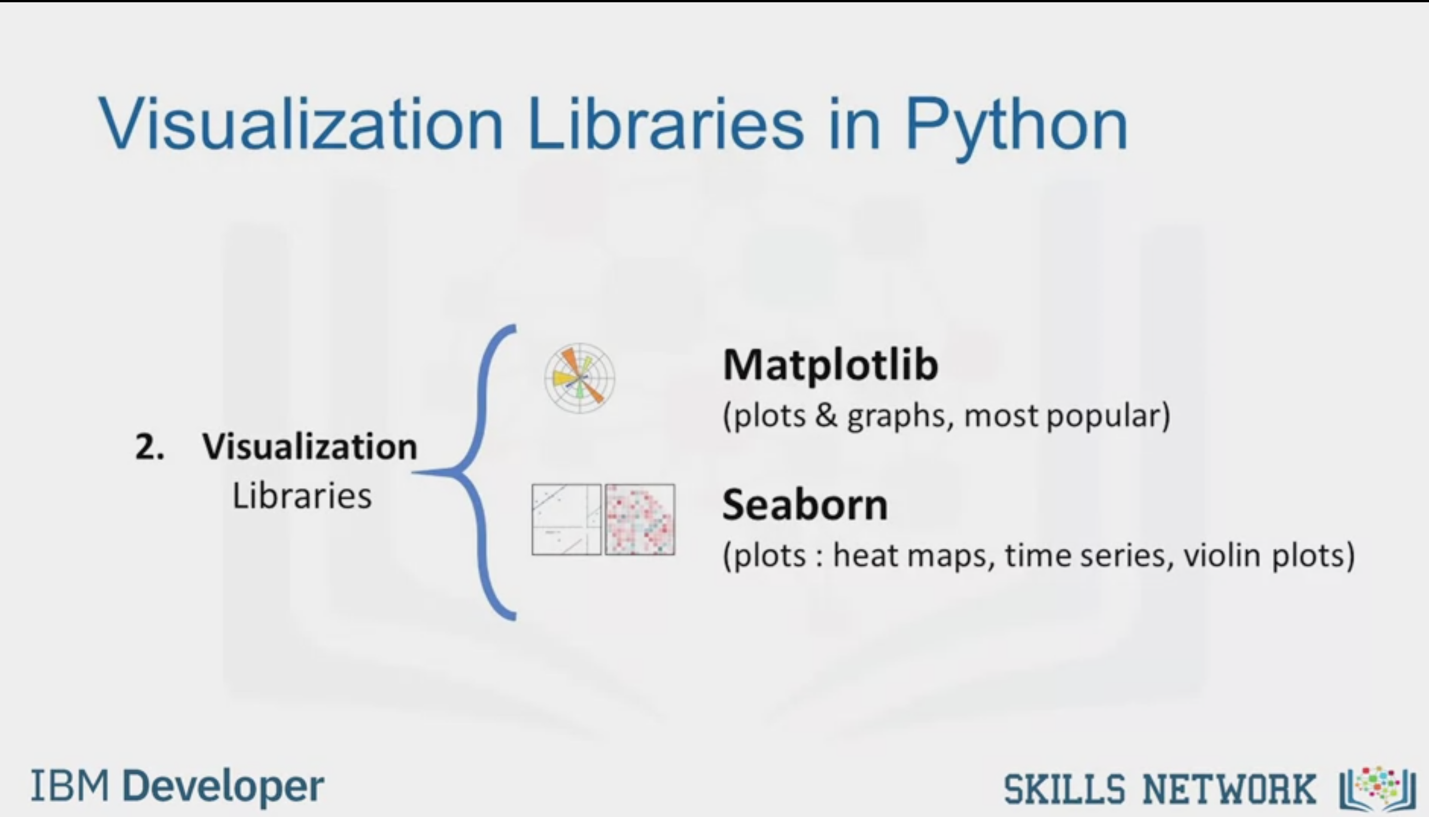
Arrays

Dataframes

Matrices

Skip

Send



Visualization Libraries in Python

2. Visualization

Libraries

Matplotlib

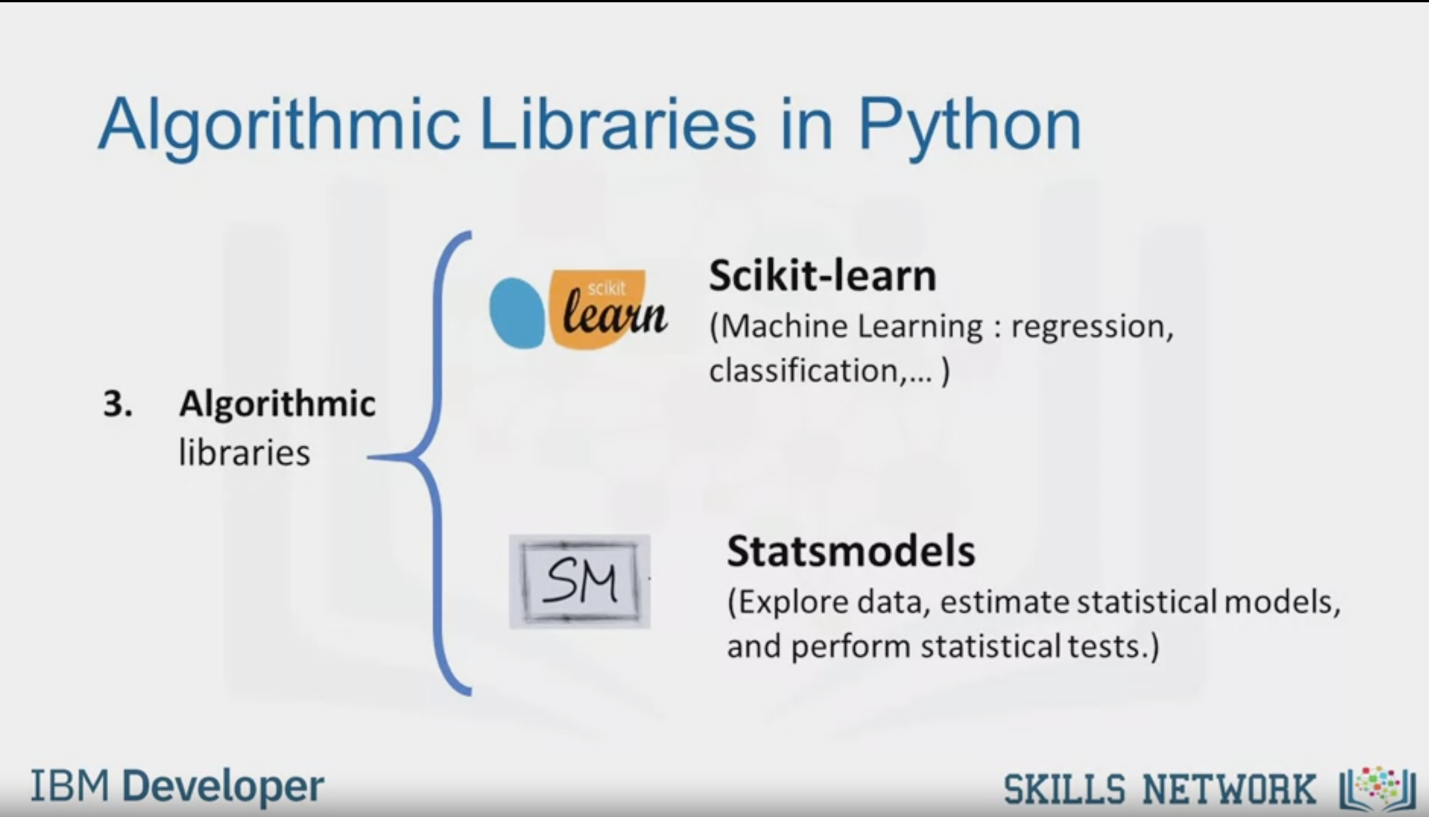
(plots & graphs, most popular)

Seaborn

(plots : heat maps, time series, violin plots)

IBM Developer

SKILLS NETWORK



Algorithmic Libraries in Python

3. Algorithmic

libraries

Scikit-learn

scikit

learn

(Machine Learning : regression,

classification,...)

Statsmodels

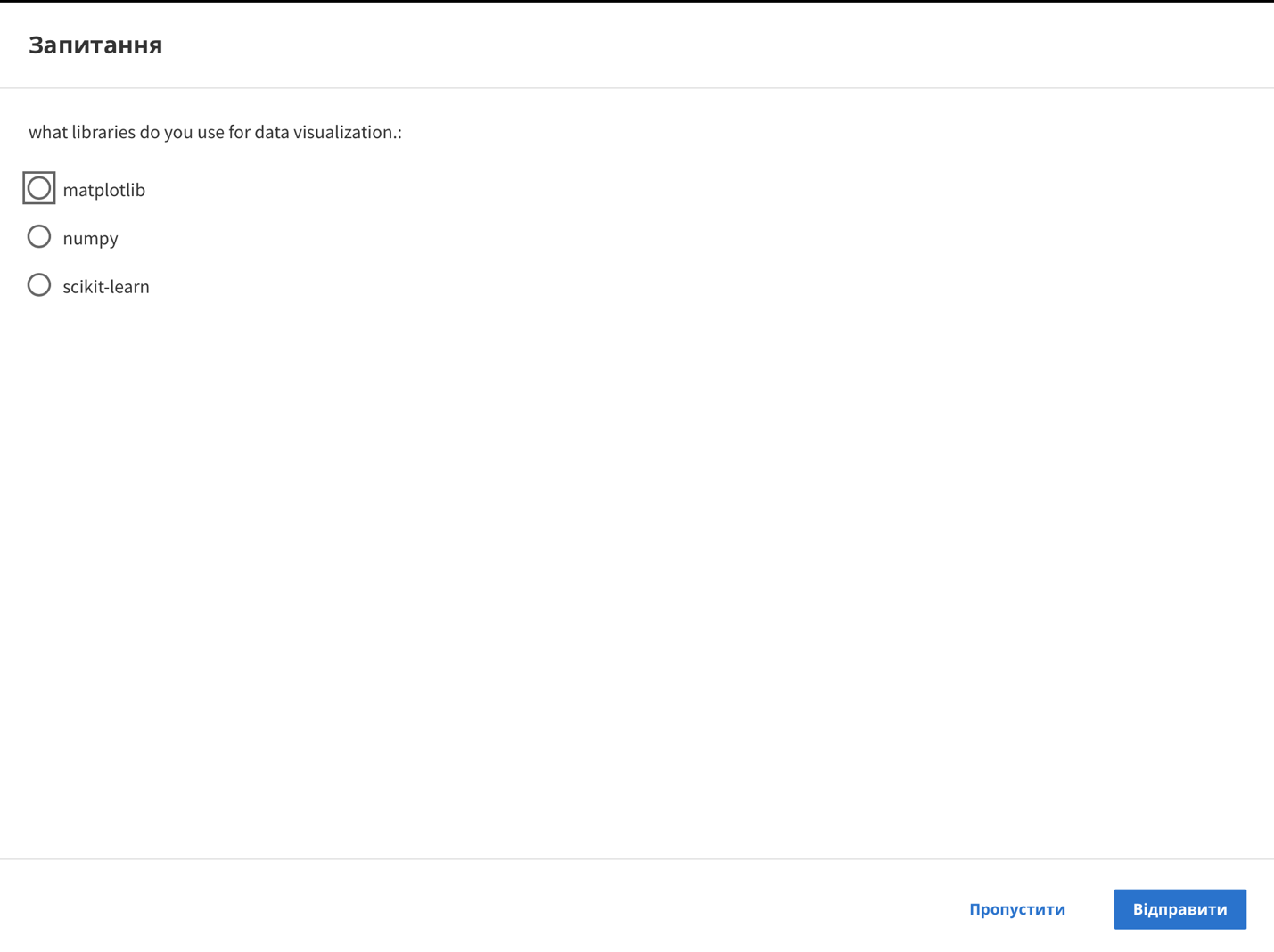
SM

(Explore data, estimate statistical models,

and perform statistical tests.)

IBM Developer

SKILLS NETWORK



Question

what libraries do you use for data visualization.:

matplotlib

numpy

scikit-learn

Skip

Send



SKILLS NETWORK

IBM Developer

In order to do data analysis in Python, we should first tell you a little bit about

the main packages relevant to analysis in Python.

A Python library is a collection of functions and

methods that allow you to perform lots of actions without writing any code.

The libraries usually contain built in modules providing different

functionalities which you can use directly.

And there are extensive libraries offering a broad range of facilities.

We have divided the Python data analysis libraries into three groups.

The first group is called scientific computing libraries.

Pandas offers data structure and tools for effective data manipulation and analysis.

It provides facts, access to structured data.

The primary instrument of Pandas is the two dimensional table consisting of column

and row labels, which are called a data frame.

It is designed to provid easy indexing functionality.

The NumPy library uses arrays for its inputs and outputs.

It can be extended to objects for matrices and

with minor coding changes, developers can perform fast array processing.

SciPy includes functions for

some advanced math problems as listed on this slide, as well as data visualization.

Using data visualization methods is the best way to communicate with others,

showing them meaningful results of analysis.

These libraries enable you to create graphs, charts and maps.

The Matplotlib package is the most well known library for data visualization.

It is great for making graphs and plots.

The graphs are also highly customizable.

Another high level visualization library is Seaborn.

It is based on Matplotlib.

It's very easy to generate various plots such as heat maps,

time series and violin plots.

With machine learning algorithms,

we're able to develop a model using our data set and obtain predictions.

The algorithmic libraries tackles the machine learning tasks from basic

to complex.

Here we introduce two packages,

the Scikit-learn library contains tools statistical modeling,

including regression, classification, clustering, and so on.

This library is built on NumPy, SciPy and Matplotib.

Statsmodels is also a Python module that allows users to explore data,

estimate statistical models and perform statistical tests.