

機器學習於材料資訊的應用

Machine Learning on Material Informatics

陳南佑(NAN-YOW CHEN)

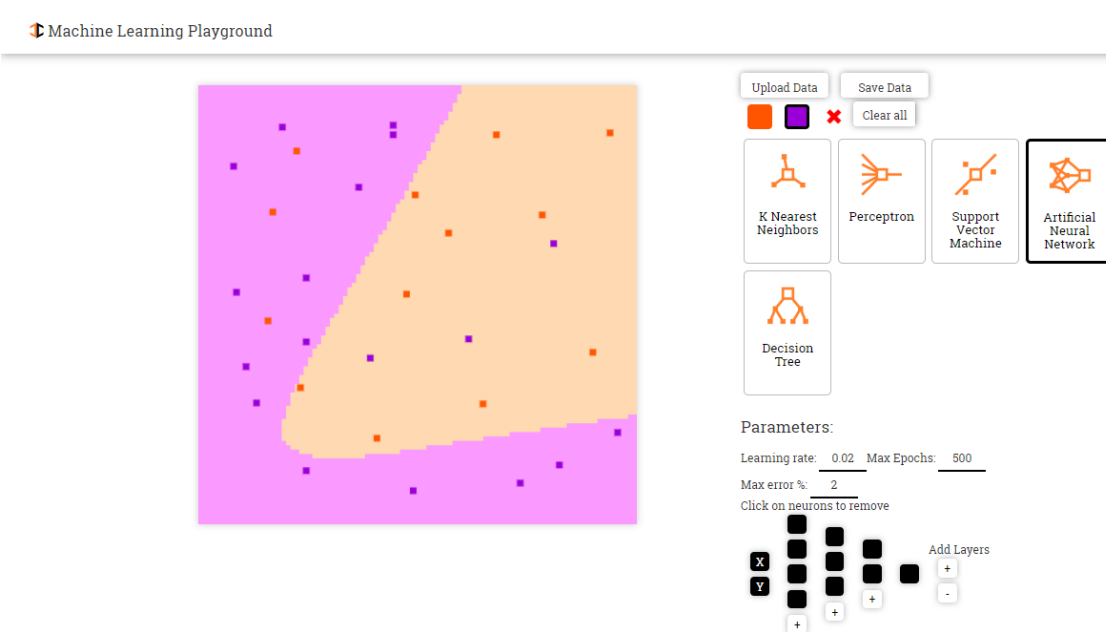
nanyow@narlabs.org.tw

楊安正(AN-CHENG YANG)

acyang@narlabs.org.tw

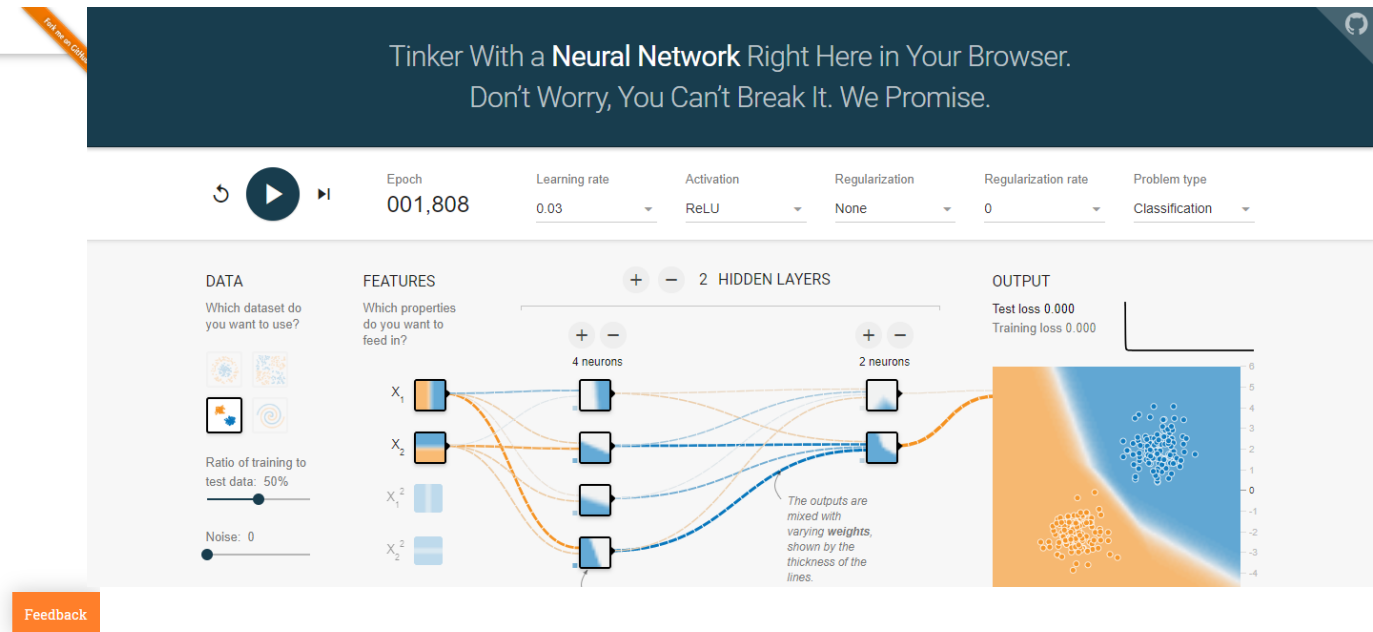
Two online Machine Learning playground

Machine Learning Playground



<http://ml-playground.com>

Play with neural networks!



<https://playground.tensorflow.org>

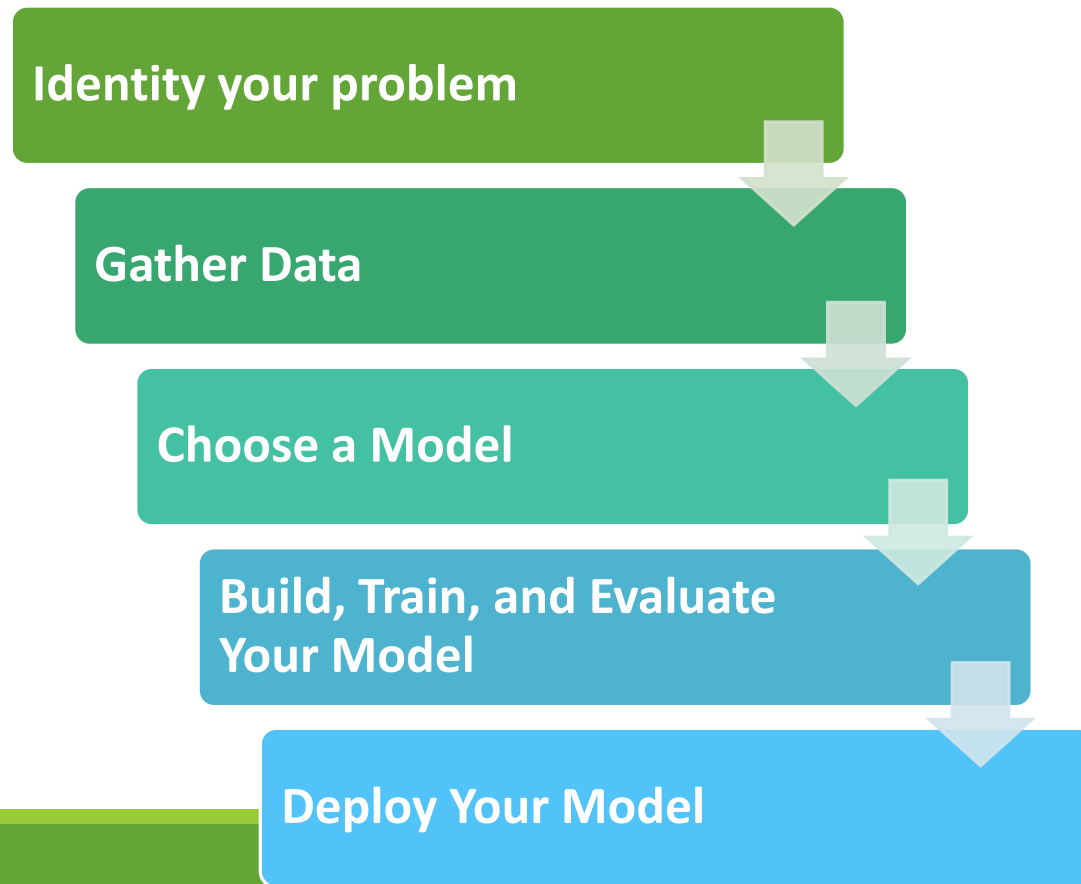
Development Environment For Machine Learning



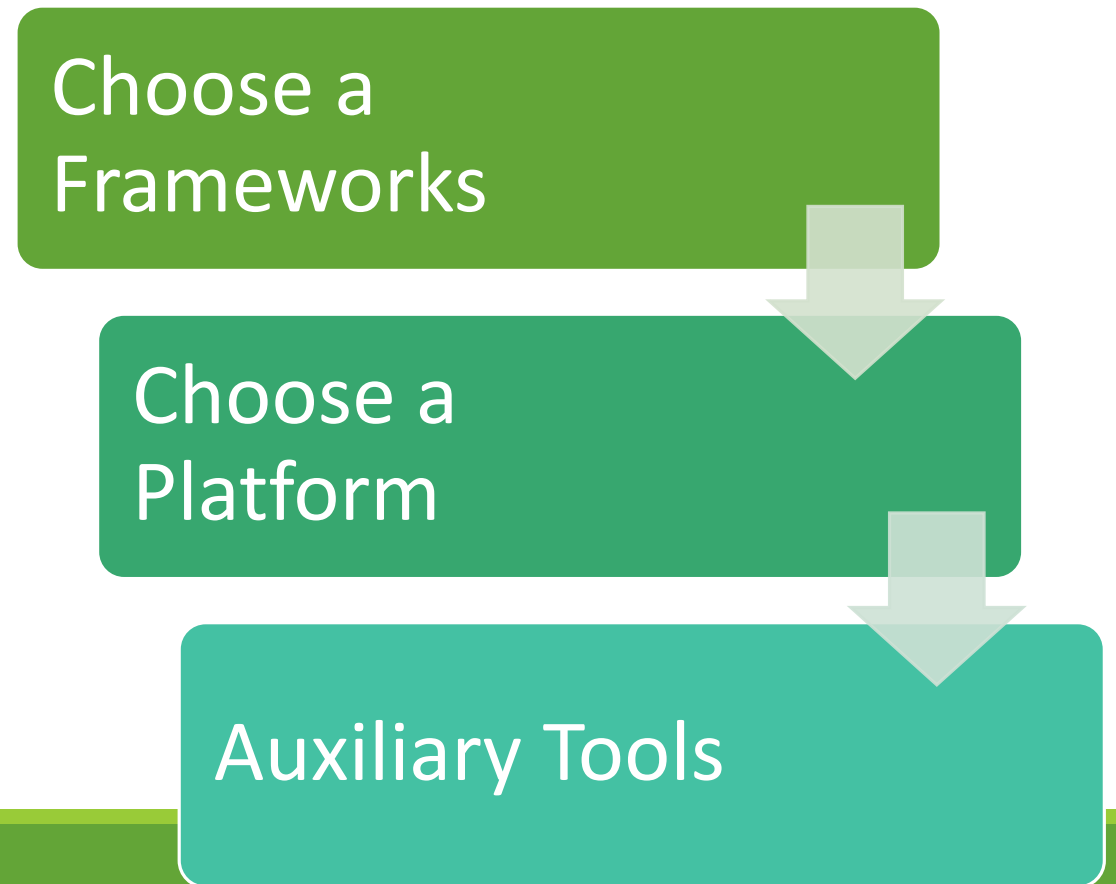
Getting started with machine learning

Two Perspectives

Analysis problem



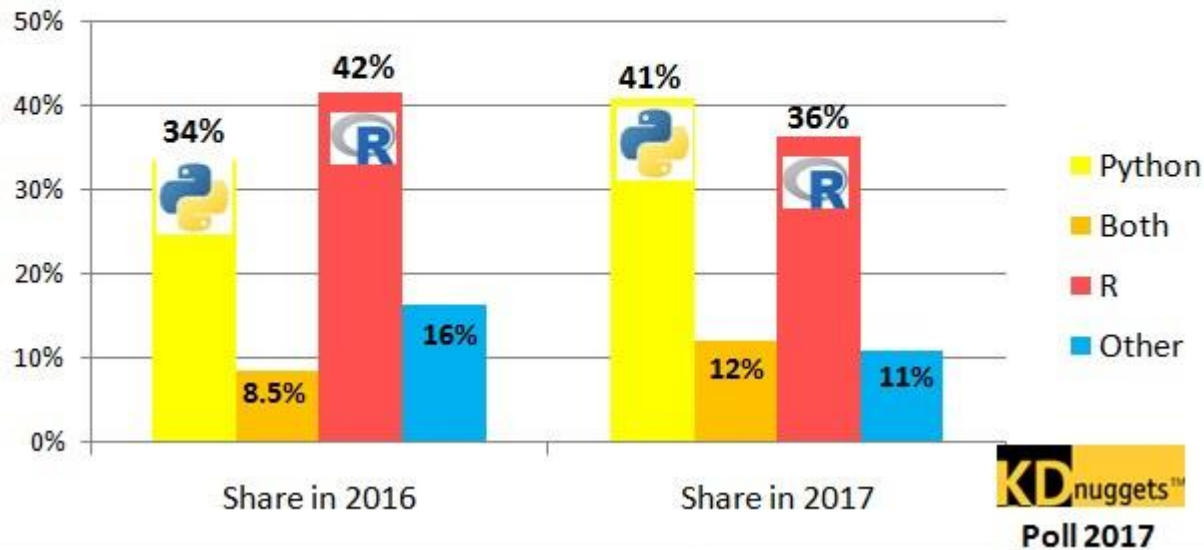
Implementation



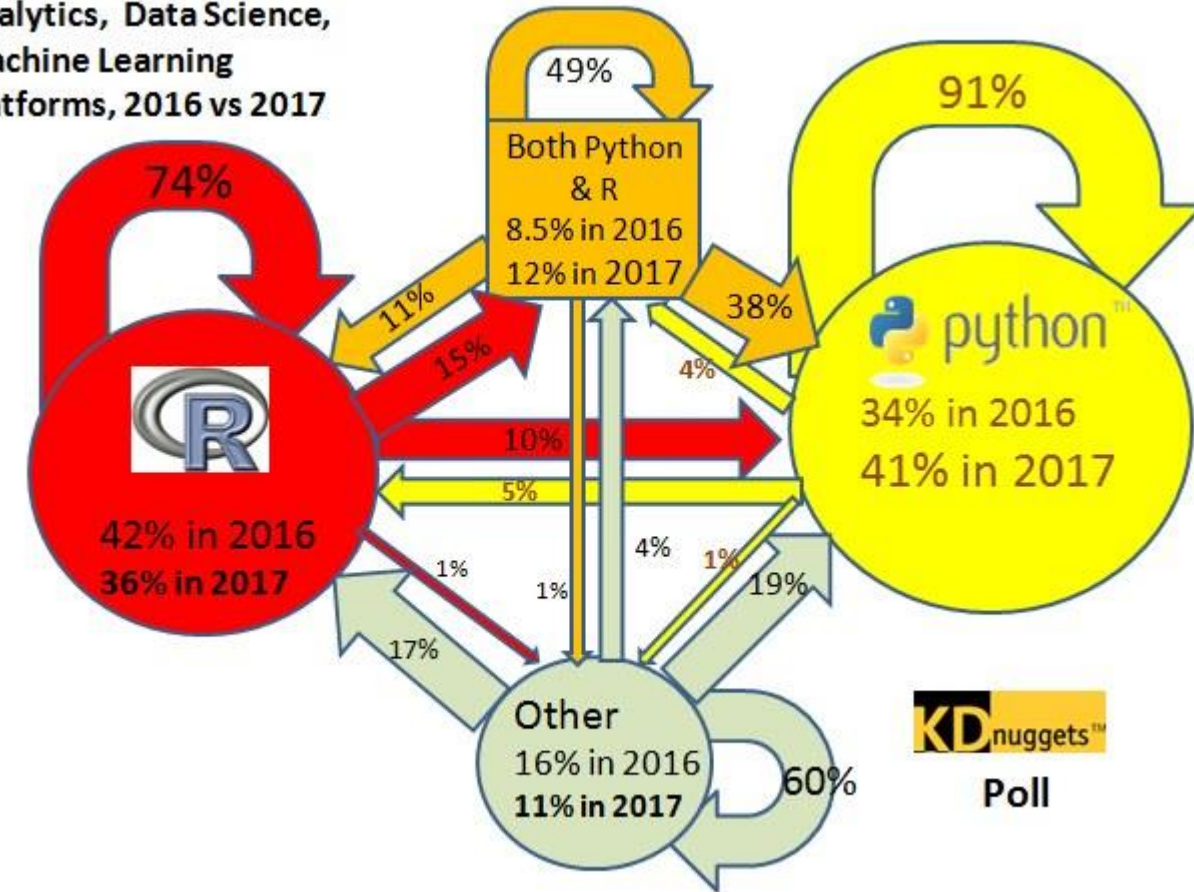
Choose a Frameworks

Popular languages in Data Science

Python, R, Both, or Other platforms for Analytics, Data Science, Machine Learning

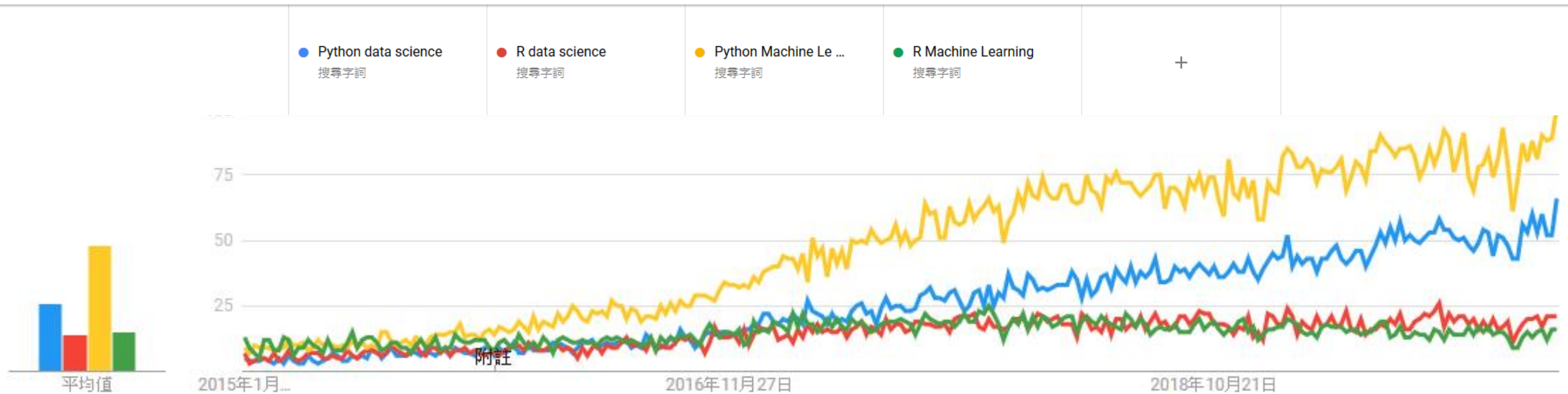


Analytics, Data Science, Machine Learning Platforms, 2016 vs 2017



Choose a Frameworks

Popular languages in Data Science & Machine Learning

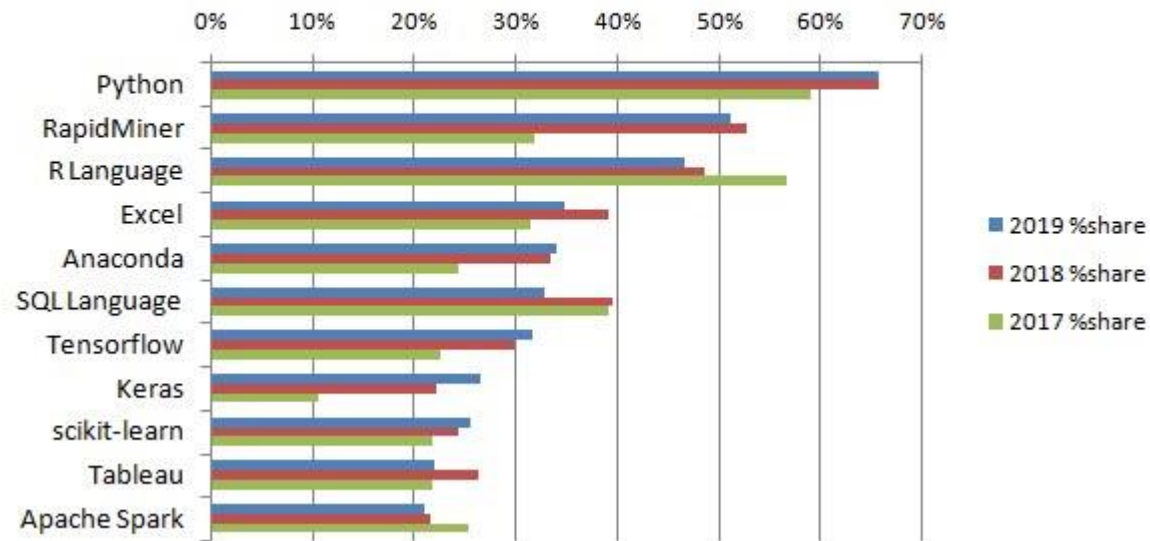


<https://trends.google.com/trends/explore?date=2015-01-01%202020-03-01&q=Python%20data%20science,R%20data%20science,Python%20Machine%20Learning,R%20Machine%20Learning>
<https://analyticsindiamag.com/top-programming-languages-for-machine-learning-based-on-aim-research/>

Choose a Frameworks

Popular languages in Data Science

Top Analytics, Data Science, Machine Learning Software 2017-2019, KDnuggets Poll



Advantages and Disadvantages of Python

Advantages

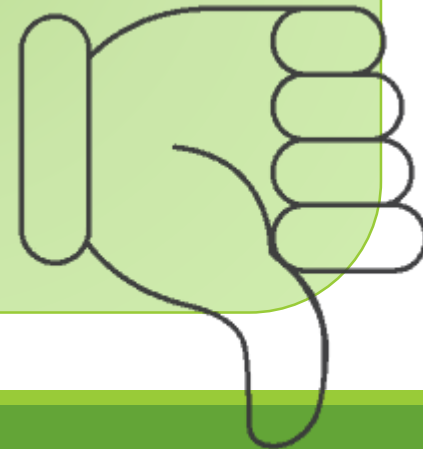
Extensive Support Libraries

Interactive
Object-Oriented, Procedure –
Oriented, Functional
programming
Extensible in C++ & C



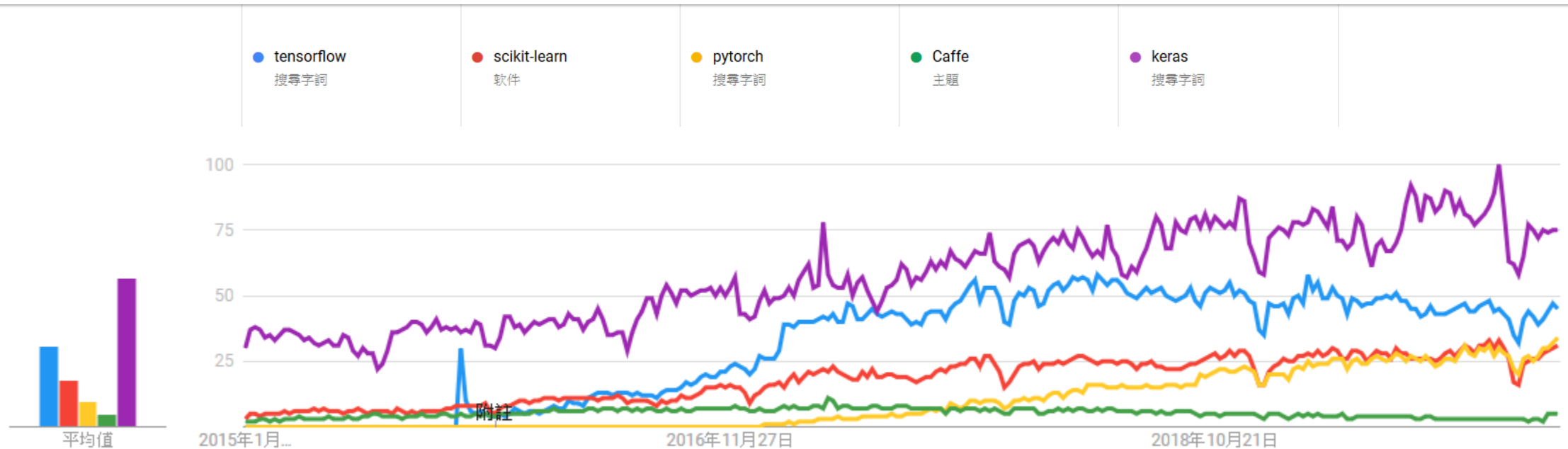
Disadvantages

Difficulty in Using Other
Languages
Gets Slow in Speed
Run-time Errors



Choose a Frameworks

Popular Machine Learning Frameworks (python)



<https://www.analyticsindiamag.com/top-10-programming-languages-data-scientists-learn-2018/>

Build your own Development Environment

Use python + pip + virtualenv

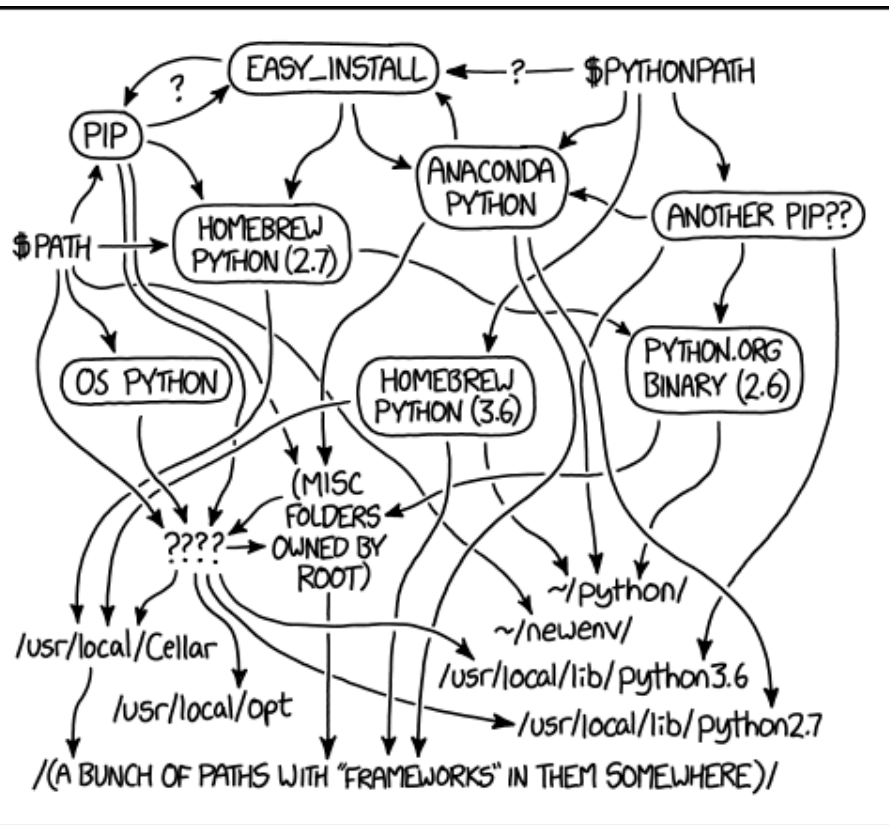
Windows

1. Download and install python.
2. Install virtualenv via pip
3. Activate virtualenv
4. Install package in virtualenv

Mac & Linux

Don't use system python
Don't use system python
Don't use system python

1. Download and compiler python.
2. Install virtualenv via pip
3. Activate virtualenv
4. Install package in virtualenv



MY PYTHON ENVIRONMENT HAS BECOME SO DEGRADED
THAT MY LAPTOP HAS BEEN DECLARED A SUPERFUND SITE.

Build your own Development Environment

Use Miniconda/Anaconda

Conda is a package management system.

Miniconda/Anaconda is a distribution for python.

Anaconda is owned by Continuum Analytics™.

Packages in Conda are released by binary not source code. This means that if conda decide not to release certain package you will ...



ANACONDA®

Anaconda Navigator
File Help

ANACONDA NAVIGATOR

Sign in to Anaconda Cloud

Home

Environments

Learning

Community

Documentation

Developer Blog



Applications on scikit-learn

Channels

Refresh



Jupyter

5.7.4

Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.

Launch



anypointools

1.1.1

Install



dioplas

0.4.1

Install



Glueviz

0.14.0

Multidimensional data visualization across files. Explore relationships within and among related datasets.

Install



hyperspyui

1.1.0



JupyterLab

0.35.3

An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.



Orange 3

3.20.1

Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows.



psyplo GUI

1.2.3

Small Practice

Create a new environment

1. Install numpy scipy matplotlib
2. Install scikit-learn
3. Install jupyter or spyder

Launch jupyter or spyder

Try to run some code

```
import numpy as np
```

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

```
# Compute the x and y coordinates for points on a sine curve
```

```
x = np.arange(0, 3 * np.pi, 0.1)
```

```
y = np.sin(x)
```

```
# Plot the points using matplotlib
```

```
plt.plot(x, y)
```

```
plt.show() # You must call plt.show() to make graphics appear.
```

Choose a Platform

Popular Machine Learning Frameworks

	 <p>Amazon Machine Learning</p>	
Google Colaboratory	Amazon ML	Microsoft Machine Learning Studio
https://colab.research.google.com/	https://aws.amazon.com/tw/machine-learning/	https://azure.microsoft.com/en-us/services/machine-learning-studio/

<https://analyticsindiamag.com/5-alternatives-to-google-colab-for-data-scientists/>

Auxiliary Tools

What is Version Control System & Why do you need it?

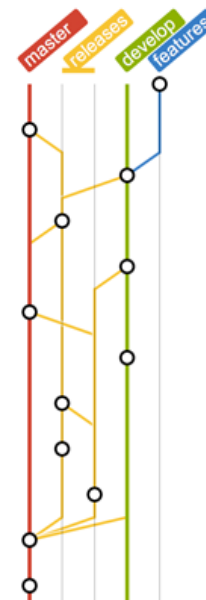
Folder / Filename style

VCS : git, svn

EVERY DESIGNER IN THIS WORLD



Git Flow Chart



FEATURE/JRA-1234	Working on JRA-1234	Remie Bolte	23:02:08 today	49a339b
MASTER	1.2.0 Merge branch 'release/1.2.0'	Remie Bolte	22:38:11 today	a38b6ec
DEVELOP	RELEASE/1.3.0 Merge branch 'release/1.2.0' into develop	Remie Bolte	22:38:11 today	07bf32f
	Merge branch 'master' into release/1.2.0	Remie Bolte	22:36:54 today	4b464c4
	Merge branch 'hotfix/1.1.3' into develop	Remie Bolte	22:36:41 today	8958747
1.1.3	Merge branch 'hotfix/1.1.3'	Remie Bolte	22:36:40 today	ae21f0
	Adding a feature specific file	Remie Bolte	22:34:46 today	3e8f9d1
	Merge branch 'hotfix/1.1.3' into release/1.2.0	Remie Bolte	22:33:55 today	55fb1b2
	Removing obsolete file	Remie Bolte	22:33:39 today	a93a486
	Adding release notes	Remie Bolte	22:33:08 today	b1b1d85
SIGNED_TAG	Copy C.zip as D.zip	Anna Buttfield	2011-02-03	0a943a2
	Add binary file C.zip	Anna Buttfield	2011-02-03	e0a3f6d

Auxiliary Tools

Git client

	msysGit+ TortoiseGit	SmartGit	SourceTree	Eclipse +egit
windows	⦿	⦿	⦿	⦿
mac		⦿	⦿	⦿
linux		⦿		⦿
GUI	⦿	⦿	⦿	⦿
portable	⦿	⦿		⦿
license	GNU GPL v2	Free for non-commercial purposes	Free for non-commercial purposes	EPL 1.0

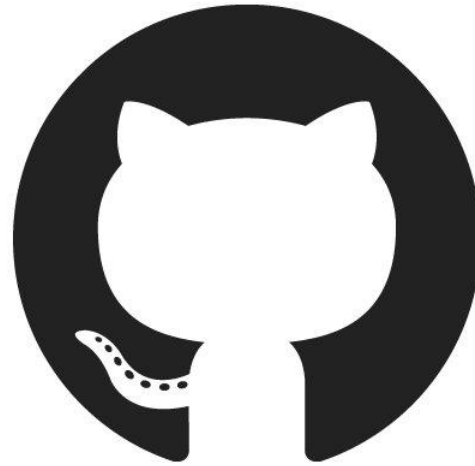
Auxiliary Tools

Popular Software repository



gitlab

<https://about.gitlab.com/>



github

<https://github.com/>



bitbucket

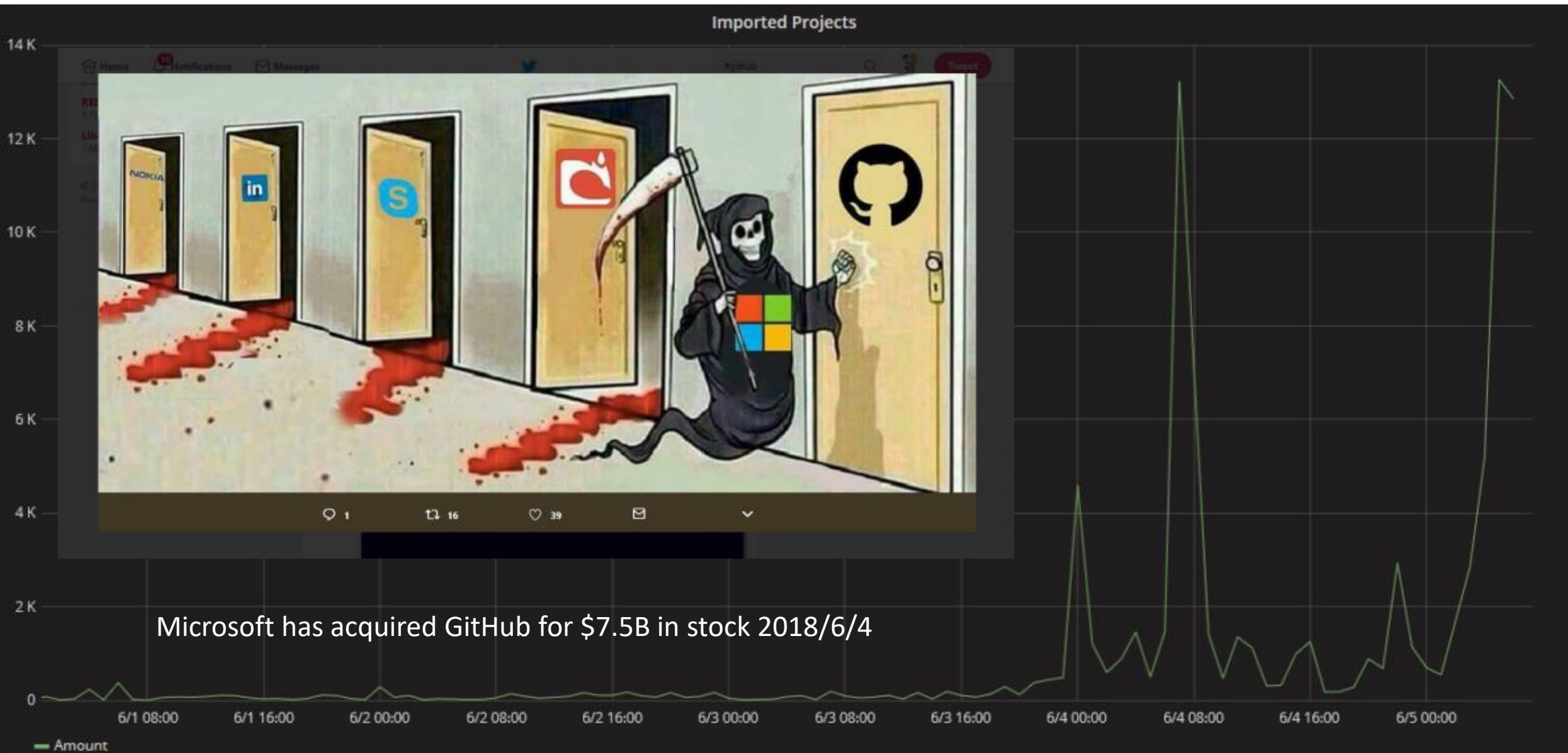
<https://bitbucket.org/>

Auxiliary Tools

Popular Software repository

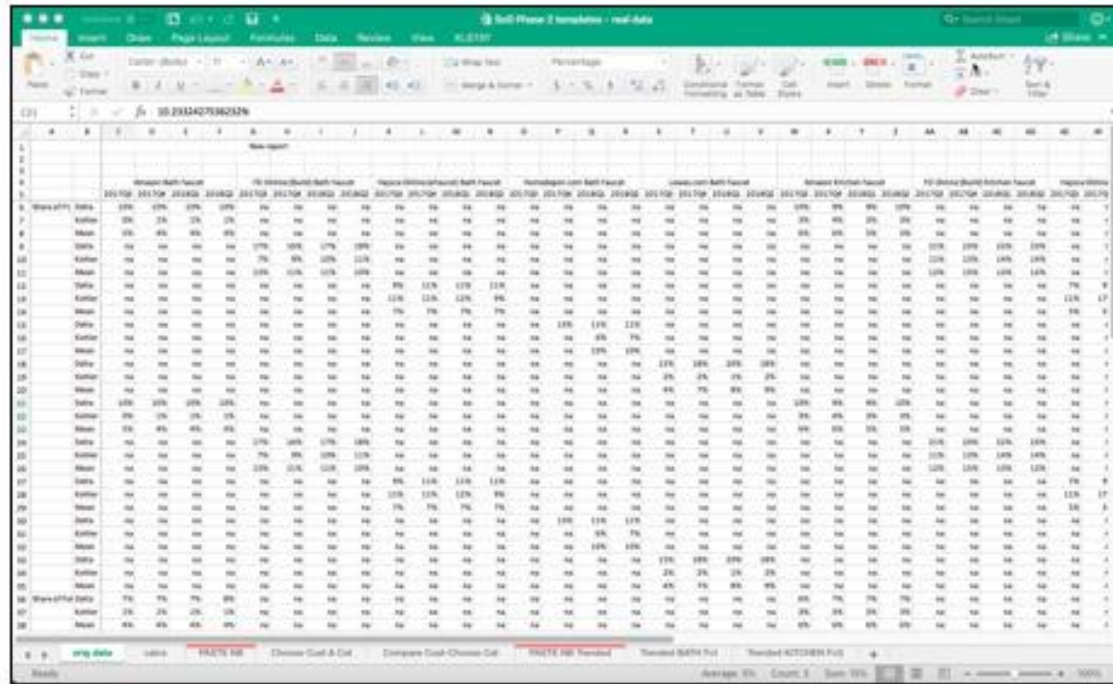
gitlab	github	bitbucket
GitLab 是一個利用 Ruby on Rails 開發的開源應用程式，實現一個自託管的 Git 項目倉庫，可通過 Web 介面進行訪問公開的或者私人項目。	GitHub 是第一個供「用Git進行版本控制系統的軟體開發項目」使用的基於Web的代碼託管服務，是目前全球最大的開源社交編程及代碼託管網站。	BitBucket 是 2008 年創建的原始碼託管網站，採用 Mercurial 和 Git 作為分布式版本控制系統，同時提供免費帳戶和商業計劃。2010 年被 Atlassian 收購。
Git	Git，SVN，HG，TFS	Git，HG，SVN
cloud-hosted plan 允許無限數量的用戶在無限數量的公共和私有項目上進行協作，並且每個存儲庫有 10GB 的空間限制	對磁碟使用沒有限制。但是，項目不能超過 1 GB和單個文件不能超過 100 MB。	Small teams plan 允許 5 個成員加入，公有/私有倉庫均免費。

Imported Projects of gitlab

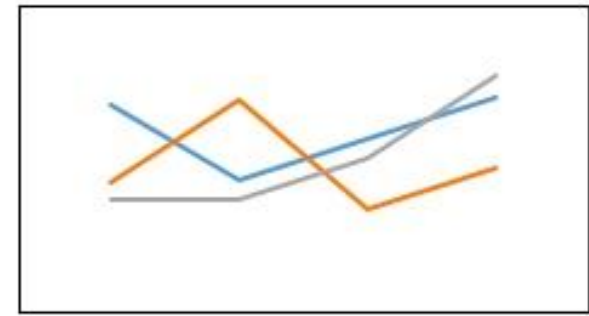
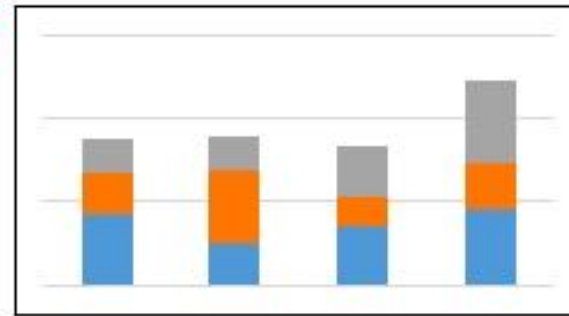
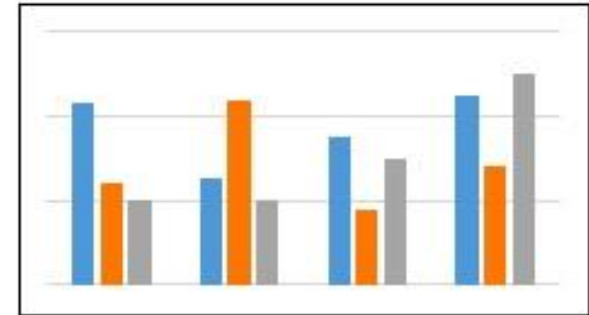
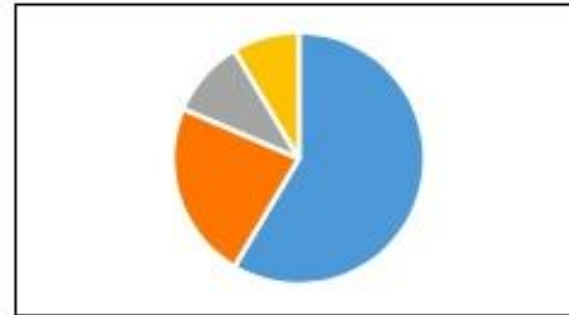


Auxiliary Tools

One picture worth ten thousand words --- Data visualization

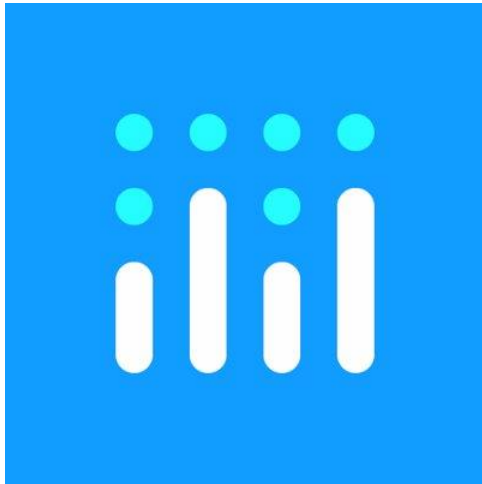


The screenshot shows a Microsoft Excel spreadsheet with a green title bar and a ribbon menu. The spreadsheet contains a large table of data with multiple columns and rows. The data appears to be organized into several sections, with headers and sub-headers. A red arrow points from the spreadsheet to the right, indicating the flow of data into the visualizations.



Auxiliary Tools

Data Analytic and visualization



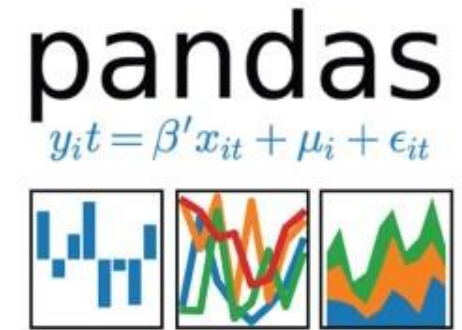
plotly

<https://plot.ly/>



Matplotlib

<https://matplotlib.org/>



Pandas

<https://pandas.pydata.org/>

20 Best Data Visualization Software Solutions of 2019

<https://financesonline.com/data-visualization/>

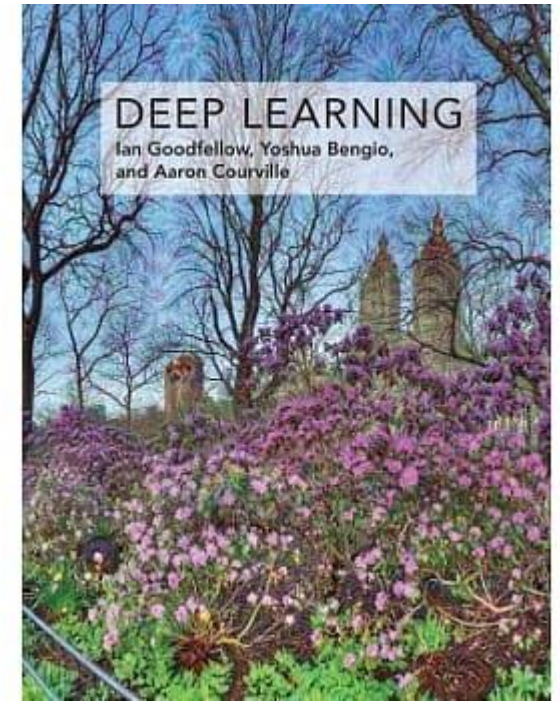
Learning Resources

Book

1. Deep Learning (ISBN : 9780262035613)
2. <https://d2l.ai/>, <https://github.com/d2l-ai>
3. 深度學習 (ISBN : 9789865021924)
4. Python程式設計：從入門到進階應用 (ISBN : 9789865030025)
5. 實戰TensorFlow：Google深度學習系統 (ISBN : 9789864764730)

Online Learning Resources

- Coursera : <https://zh-tw.coursera.org/learn/machine-learning>
- data-flair : <https://data-flair.training/blogs>
- Tensorflow : <https://www.tensorflow.org/tutorials>
- scikit-learn : <https://scikit-learn.org/stable/tutorial/index.html>
- Keras : <https://keras.io/>



Yann LeCun



YOSHUA BENGIO



GEOFFREY E HINTON