

Data X

About Me:

Data-X:
Install instructions for Mac OSX / Linux
(also works for Windows)

Alexander Fred-Ojala
afo@berkeley.edu
Data-X at Berkeley

Install Anaconda with Python 3.X

<https://www.anaconda.com/download/>

Anaconda 2019.10 for macOS Installer

Python 3.7 version

Download

64-Bit Graphical Installer (654 MB)

64-Bit Command Line Installer (424 MB)

Python 2.7 version

Download

64-Bit Graphical Installer (637 MB)

64-Bit Command Line Installer (409 MB)

Windows Instructions

For Windows, when you install Anaconda, choose to also install **Anaconda Prompt.**

Python 3.6 version *
Graphical Installer (442 MB) ⓘ



DOWNLOAD

Command-Line Installer (380 MB) ⓘ

Python 2.7 version *
Graphical Installer (438 MB) ⓘ



DOWNLOAD

Command-Line Installer (375 MB) ⓘ

Create Virtual Environment for Data-X

- Open Terminal

- Run the command:

```
conda create -n data-x python=3 anaconda
```

To activate Virtual environment:

```
source activate data-x
```

on Windows: `activate data-x`

To deactivate Virtual environment:

```
source deactivate
```

on Windows: `deactivate`

Before you install packages or run a notebook Always Activate the Virtual Environment first!

(This way you will never run into problem with crashing your root Python / Anaconda installation)

Run:

```
source activate data-x
```

(on Windows: activate data-x)

every time you open a new terminal window.



```
~ >>> source activate data-x  
(data-x) ~ >>>
```

The word within the parenthesis at the start of every line in the command prompt indicate what Virtual Environment you have activated



Download the class content from

https://github.com/afo/data-x_prague19

Download by **cloning the Github repository** (if you know Git). Otherwise we recommend going to the website and downloading the content as a zip file.

The screenshot shows the GitHub repository page for 'afo/data-x_prague19'. The repository is currently on the 'master' branch. It has 5 commits, 1 branch, 0 releases, 1 contributor, and is licensed under Apache-2.0. The repository description states: 'No description, website, or topics provided.' Below this, there are buttons for 'Add topics' and 'Edit'. The repository structure is listed as follows:

File/Folder	Commit
d1s1-intro	first_push
d1s2-project-setup	first_push
d1s3-AI-stack	first_push
d1s4-ML-in-python	first_push
d2s1-innovation-leadership-and-webscraping	first_push

At the bottom right, there is a 'Clone or download' button. A dropdown menu is open, showing options to 'Clone with HTTPS', 'Use SSH', and 'Download ZIP'. The 'Download ZIP' option is highlighted with a red circle.

How to Install packages into your Virtual Environment

Anaconda comes with many packages pre-installed, but if you want to install additional packages (or update existing ones) you can run:

Install a package by running:

```
conda install [package name]
```

Install packages by running:

```
conda install [pkg1] [pkg2] [pkg3]
```

```
(data-x) → ~ conda install tensorflow keras html5lib
```

Data X

Required packages

The packages you need can be installed by running the command below:

Install a package by running:

```
conda install html5lib py-xgboost tensorflow
```


Run your first notebook

Anaconda comes with Jupyter notebooks installed.

In order to run Jupyter notebook, open the terminal, source your Virtual Environment, `cd` into the specific working directory and then run the command:

```
jupyter notebook
```

A new browser window with your current directory will open and you can create a new notebook or open an existing one.

```
~ ▶  
~ ▶ source activate data-x  
(data-x) ~ ▶ cd data-x  
(data-x) ~/data-x ▶ jupyter notebook  
[I 13:16:46.601 NotebookApp] Serving notebooks from local directory: /Users/F0/data-x  
[I 13:16:46.601 NotebookApp] 0 active kernels  
[I 13:16:46.601 NotebookApp] The Jupyter Notebook is running at: http://localhost:8888/  
?token=eae7a2506a950b2d995199cd59297bd7ddb70f33aba5f67b  
[I 13:16:46.601 NotebookApp] Use Control-C to stop this server and shut down all kernel  
s (twice to skip confirmation).  
[C 13:16:46.602 NotebookApp]
```

Copy/paste this URL into your browser when you connect for the first time,
to login with a token:

`http://localhost:8888/?token=eae7a2506a950b2d995199cd59297bd7ddb70f33aba5f67b`

```
[I 13:16:47.083 NotebookApp] Accepting one-time-token-authenticated connection from ::1
```

Troubleshooting / In-depth explanations

Please refer to the material below and / or Google if you encounter any problems or would like a more in-depth explanation:

- <https://machinelearningmastery.com/setup-python-environment-machine-learning-deep-learning-anaconda/>
- <https://medium.com/k-folds/setting-up-a-data-science-environment-5e6fd1cbd572>
- <https://drivendata.github.io/pydata-setup/>

OPTIONAL Install **pyspark** for Big Data locally:

<http://mortada.net/3-easy-steps-to-set-up-pyspark.html>



Good Luck!

