

Data^X

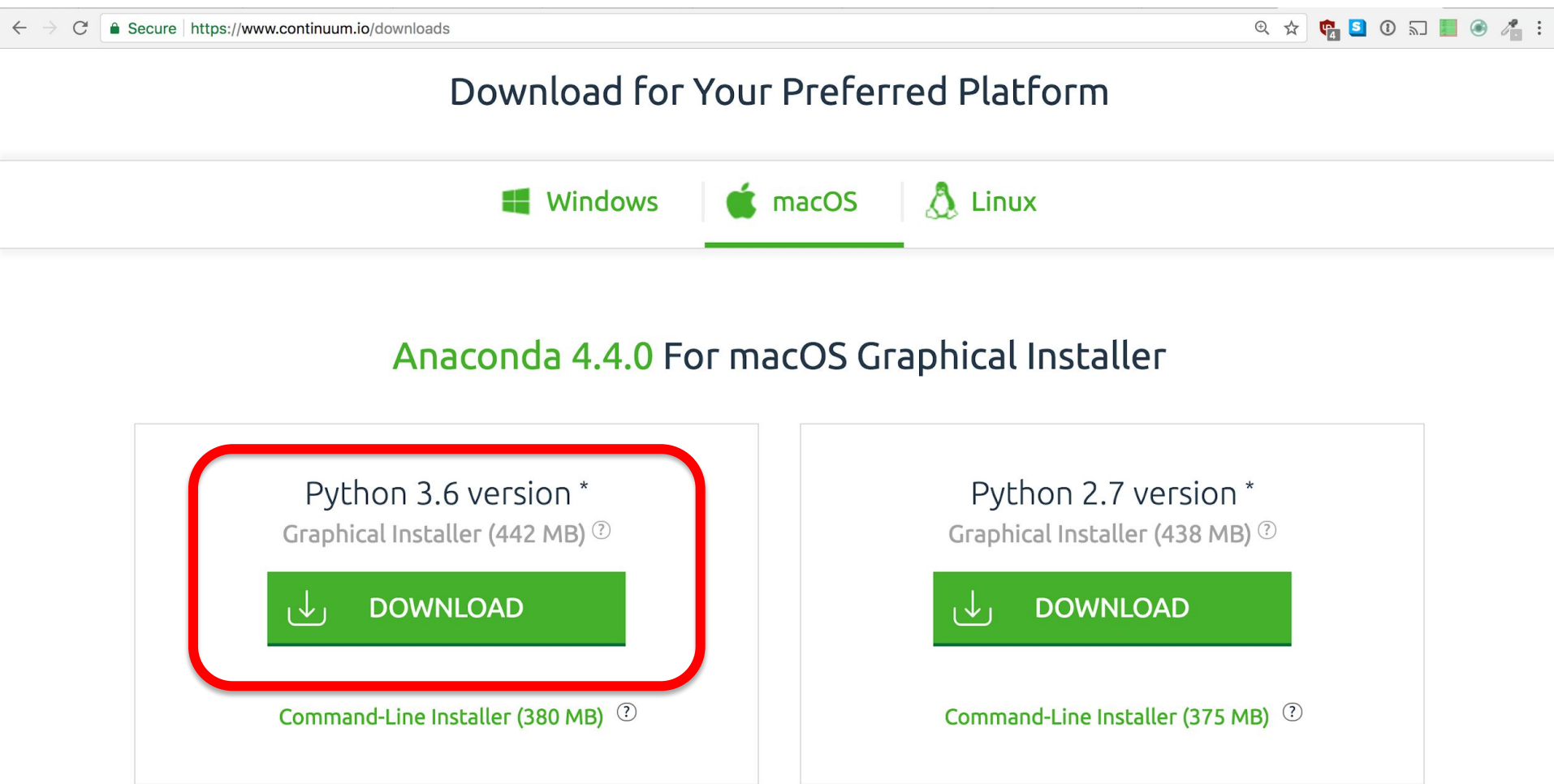
About Me:

Data-X at Berkeley Masterclass: Install instructions for MacOSX (also works for Windows)

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

Install Anaconda with Python 3.6

www.continuum.io/downloads



The screenshot shows a web browser window with the URL <https://www.continuum.io/downloads>. The page title is "Download for Your Preferred Platform". Below this, there are three tabs: "Windows", "macOS", and "Linux". The "macOS" tab is selected and highlighted with a green underline. Under the "macOS" tab, the text "Anaconda 4.4.0 For macOS Graphical Installer" is displayed. Below this text, there are two download options for macOS. The first option is for "Python 3.6 version *", which includes a "Graphical Installer (442 MB)" and a "Command-Line Installer (380 MB)". The "Graphical Installer" option is highlighted with a red rounded rectangle and has a green "DOWNLOAD" button with a download icon. The second option is for "Python 2.7 version *", which includes a "Graphical Installer (438 MB)" and a "Command-Line Installer (375 MB)". The "Graphical Installer" option also has a green "DOWNLOAD" button with a download icon.

Download for Your Preferred Platform

Windows macOS Linux

Anaconda 4.4.0 For macOS Graphical Installer

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) ⓘ



Extra Windows Instructions

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

This will make everything easier.

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

OPTIONAL: Create Virtual Environment (e.g. for Python 2.7)

We have chosen to work with Python 3.6 in this class, however it is easy to also install a Python 2.7 Virtual Environment(if you'd ever need it)

- **Open Terminal**

- **Run the command:**

```
conda create -n py2 python=2 anaconda
```

To activate the Python 2.7 Virtual environment:

```
source activate py2
```

on Windows: activate py2

To deactivate (any) Virtual environment:

```
source deactivate
```

on Windows: deactivate

Please note, many functions, modules and libraries differ between Python 2.x and Python 3.x (Python 3 is not backwards compatible). However, many scripts / notebooks can be compatible with both Python 3 and Python 2 by running the code below first in your script / notebook:

```
from __future__ import absolute_import, division, print_function
```

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

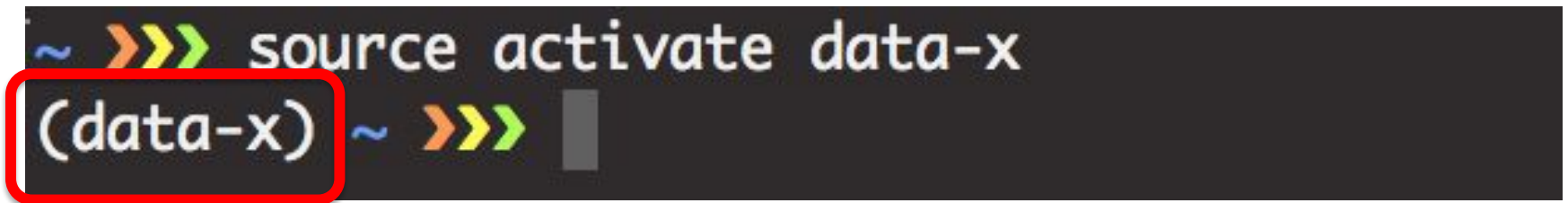
(This way you will never run into problem with crashing your root Python / Anaconda installation)

Run:

```
source activate data-x
```

(on Windows: activate data-x)

every time you open a new terminal window.

A terminal window with a dark background. The first line shows a prompt '~' followed by three green chevrons '»»»' and the command 'source activate data-x'. The second line shows the prompt '(data-x)' followed by '~' and three green chevrons '»»»'. A red rectangle highlights the '(data-x)' prompt. A grey cursor block is visible to the right of the second prompt.

```
~ »»» 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 Masterclass content from

<https://github.com/afo/dataXhkbu>

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 interface for the repository 'afo/dataXhkbu'. At the top, there are tabs for Code, Issues (0), Pull requests (0), Projects (0), Wiki, Insights, and Settings. Below the tabs, it says 'No description, website, or topics provided.' with an 'Edit' button and a link to 'Add topics'. A summary bar shows 5 commits, 1 branch, 0 releases, 1 contributor, and Apache-2.0 license. Below this, there are buttons for 'Branch: master', 'New pull request', 'Create new file', 'Upload files', 'Find file', and 'Clone or download'. The 'Clone or download' button is highlighted with a red circle, and its dropdown menu is open, showing 'Clone with HTTPS' (selected), 'Use SSH', and a text input field containing the repository URL. Below the URL, the 'Download ZIP' button is also highlighted with a red circle. The repository files list includes 'd1s1-intro', 'd1s2-project-setup', 'd1s3-AI-stack', 'd1s4-ML-in-python', and 'd2s1-innovation-leadership-and-webscraping', all marked as 'first_push'. The repository was updated 2 months ago.

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights Settings

No description, website, or topics provided. Edit

Add topics

5 commits 1 branch 0 releases 1 contributor Apache-2.0

Branch: master New pull request Create new file Upload files Find file Clone or download

afo Update README.md

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

2 months ago

Clone with HTTPS Use SSH

Use Git or checkout with SVN using the web URL.

<https://github.com/afo/dataXprague.gi>

Download ZIP

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 tensorflow keras html5lib py-xgboost
```

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

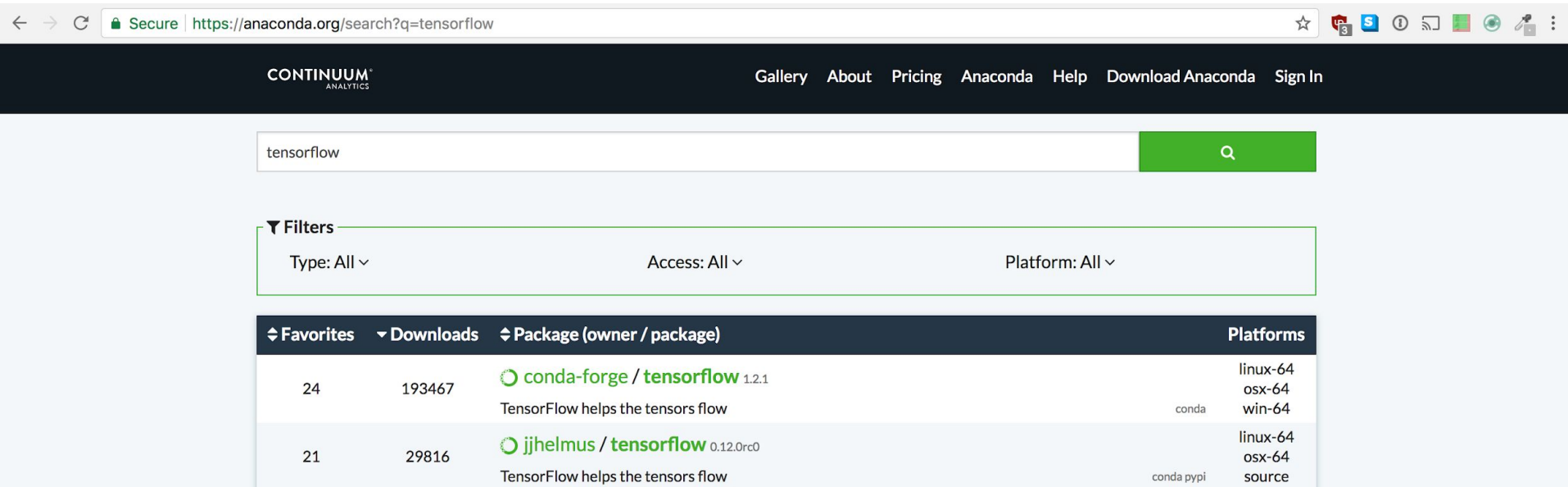
Data X

Installing packages not available via conda



Some packages are not available via conda, instead you can visit <https://anaconda.org/> (Anaconda Cloud, a package management service) and search for the package you want to install. Here you can usually find any Python package for your specific machine settings.

Install a package by (for example) running:

```
conda install -c conda-forge tensorflow
```



The screenshot shows a web browser window with the URL <https://anaconda.org/search?q=tensorflow>. The page header includes the Continuum Analytics logo and navigation links: Gallery, About, Pricing, Anaconda, Help, Download Anaconda, and Sign In. A search bar at the top contains the text 'tensorflow' with a green search button. Below the search bar, there are filter options: 'Type: All', 'Access: All', and 'Platform: All'. The main content area displays a table of search results for 'tensorflow'.

⬆ Favorites	⬇ Downloads	⬆ Package (owner / package)	Platforms
24	193467	 conda-forge / tensorflow 1.2.1 TensorFlow helps the tensors flow	linux-64 osx-64 win-64 conda
21	29816	 jjhelmus / tensorflow 0.12.0rc0 TensorFlow helps the tensors flow	linux-64 osx-64 source conda pypi

Run your first notebook

Anaconda comes with Jupyter notebooks which we will work with a lot. In order to run your first 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 either 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 kernels (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>



Preparation material

If you want to prepare for the bootcamp please feel free to dive into the following material:

- **PYTHON BOOTCAMP:**
<https://bids.berkeley.edu/news/python-boot-camp-fall-2016-training-videos-available-online>
- **NEURAL NETWORKS:**
<https://www.youtube.com/watch?v=aircAruvnKk>
- **Data-X resources:**
<https://data-x.blog/resources/>
- **Install geth for Ethereum development**
(and sync the Ethereum Blockchain -- requires ~75Gb of disk space)
<https://www.ethereum.org/cli>



See you soon!

