About Me:

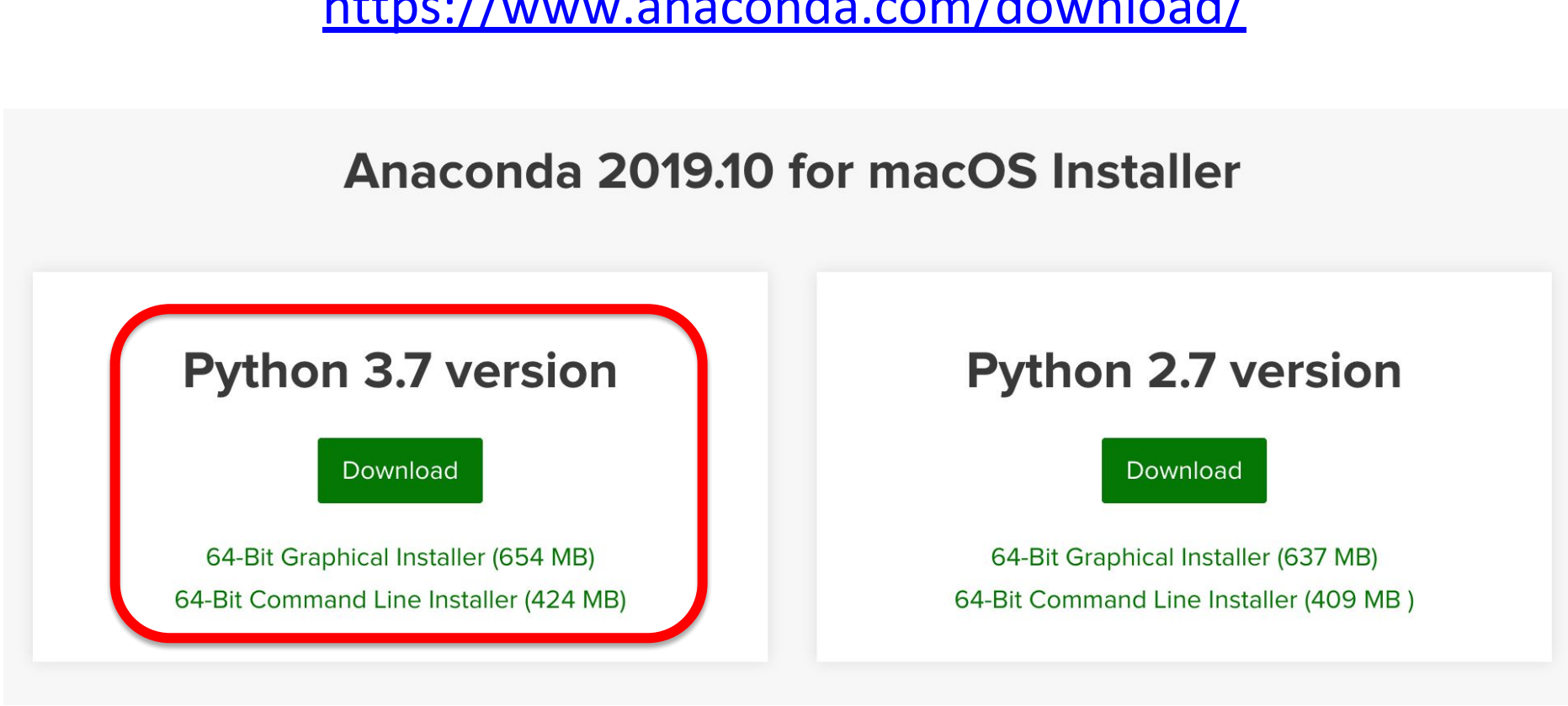
Data-X:

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/



Windows Instructions

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

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!

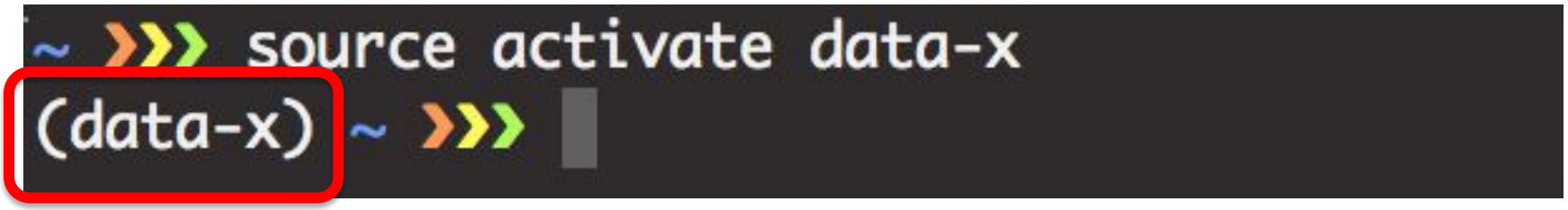
(This way you will never run into problem with crashing your root Python / Anaconda installation)

Run:

`source activate data-x`

(on Windows: `activate data-x`)

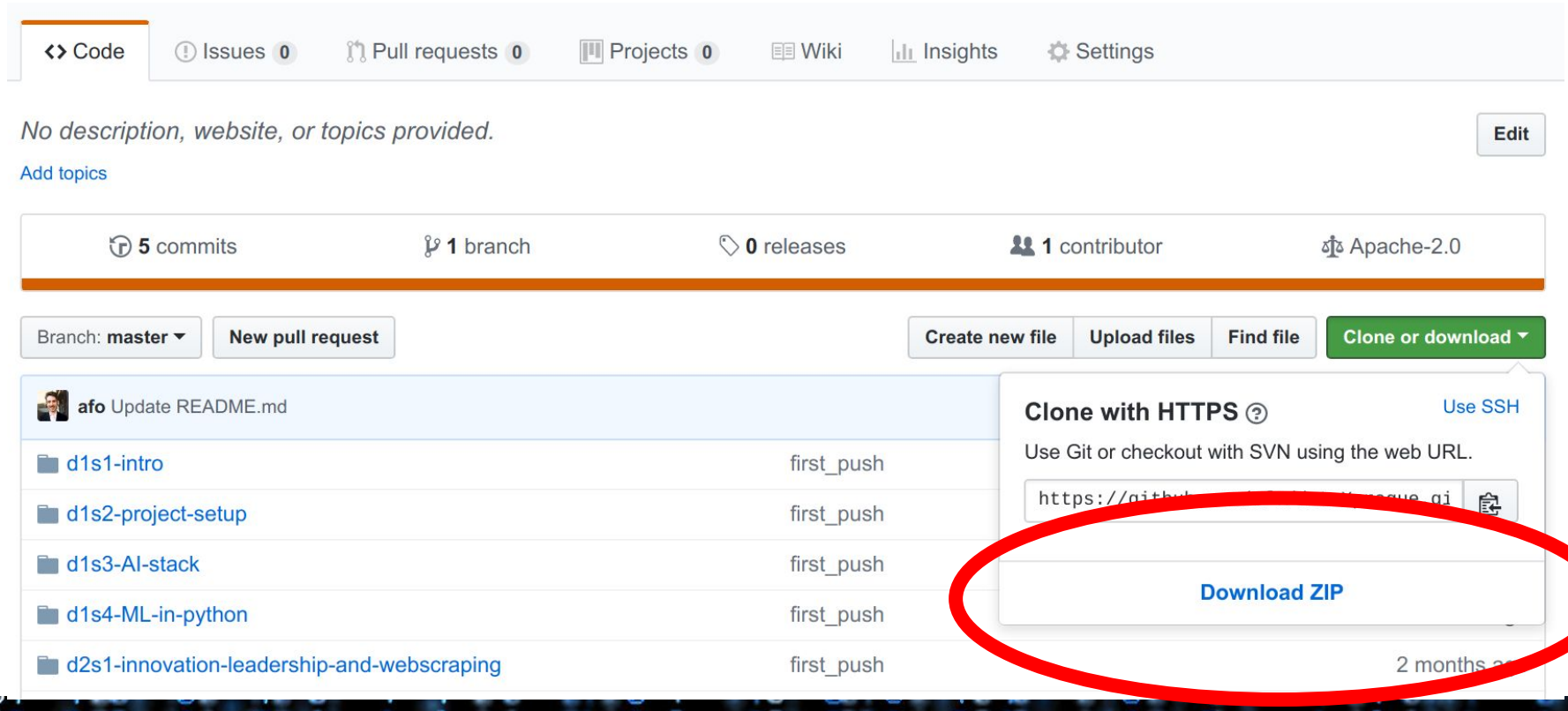
every time you open a new terminal window.



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.



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
```

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/FO/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!