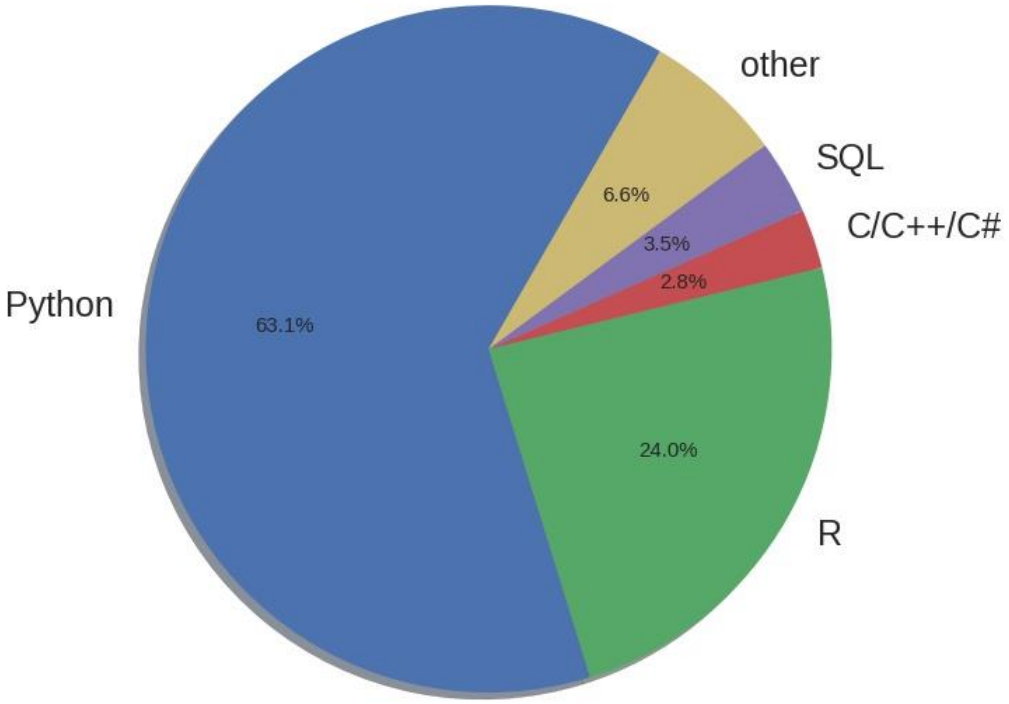
Technical details

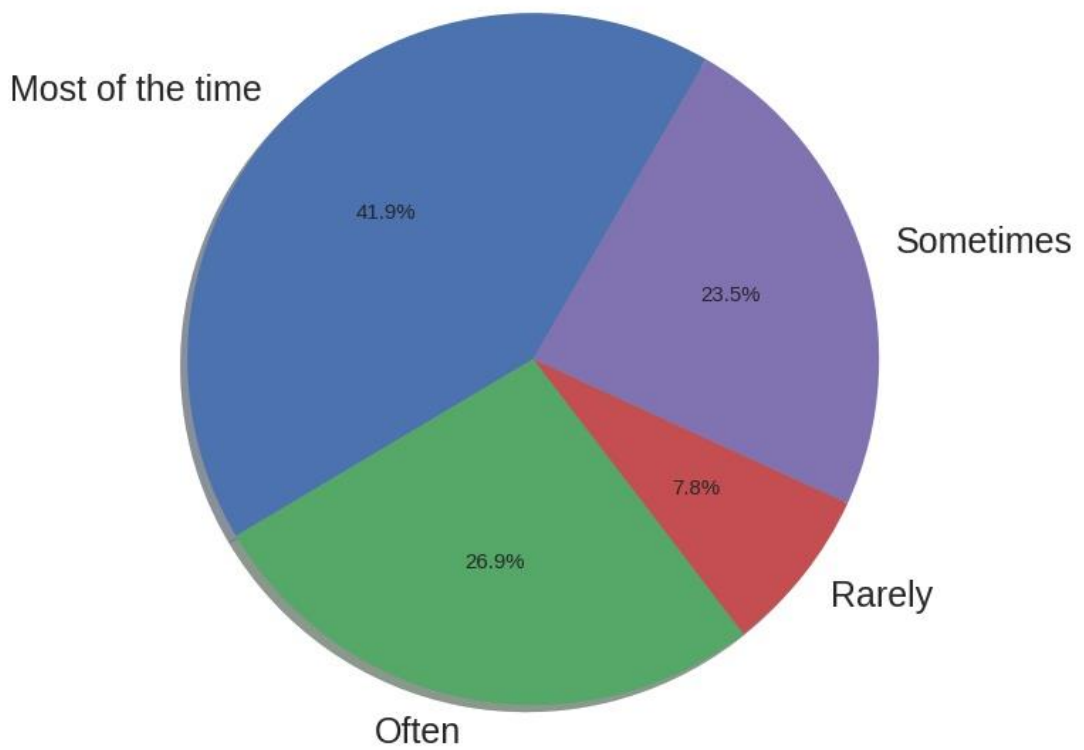
Working tools, python, numpy, jupyter notebooks, github, Kaggle, ssh

Examples at the course site: qati.me/dl-class.html

(Bálint Ármin Pataki)

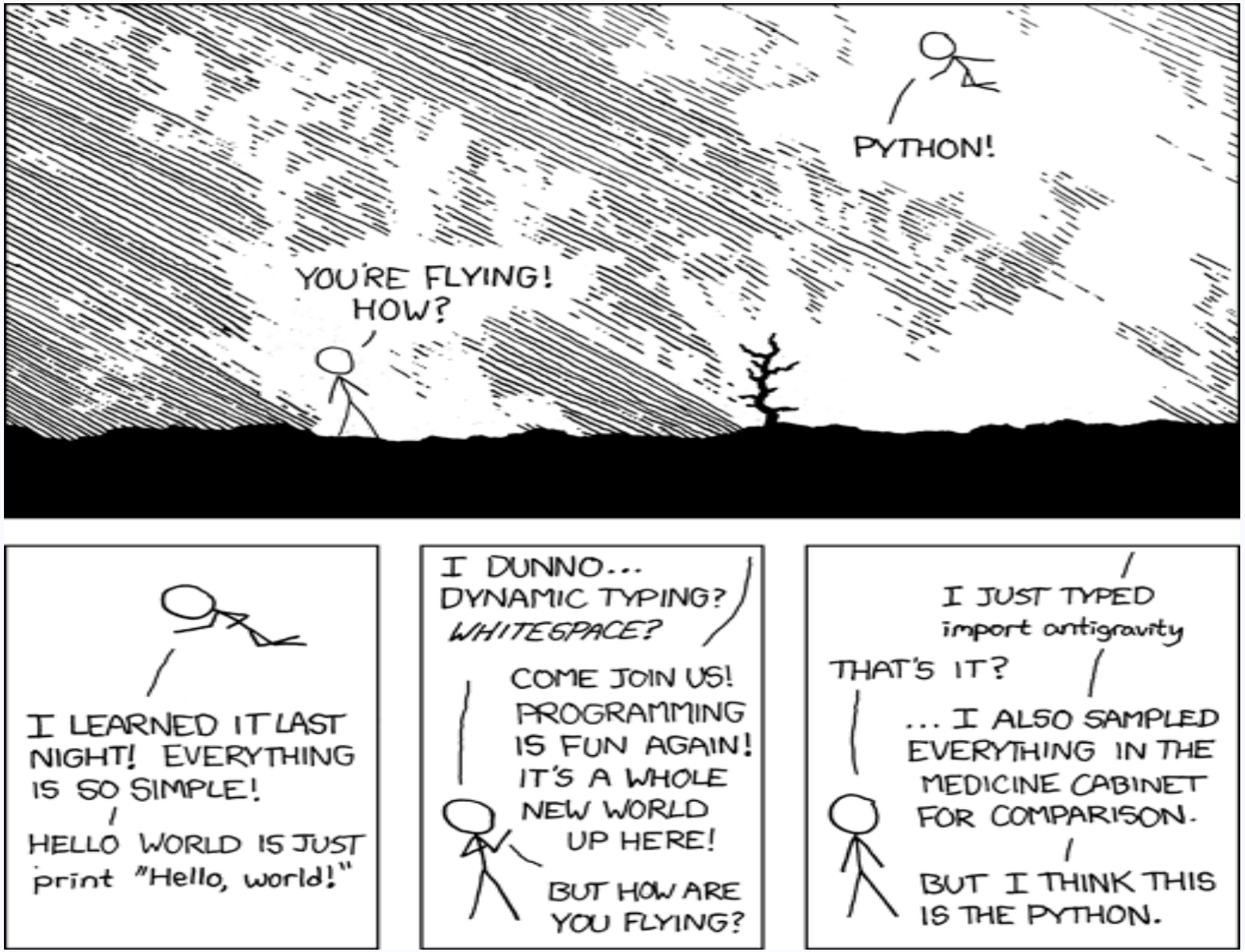
Working tools: language + IDE





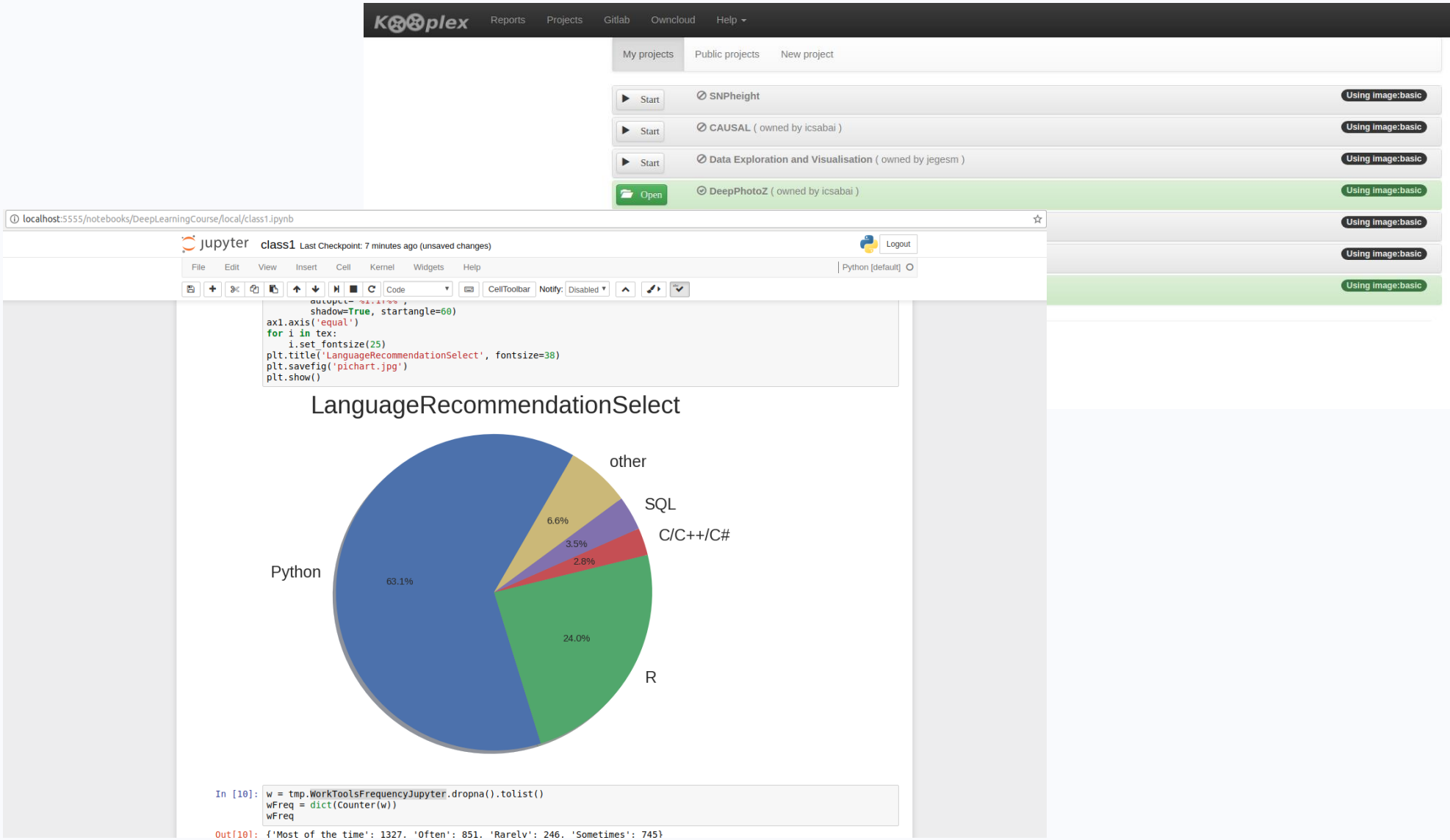
[Kaggle ML and DataScience Survey, 2017]

Working tools: language + IDE

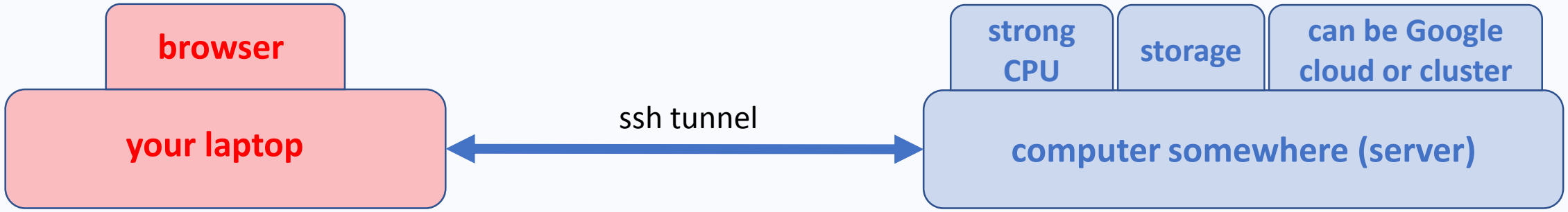


[xkcd, https://xkcd.com/353/]

Jupyter and Kooplex



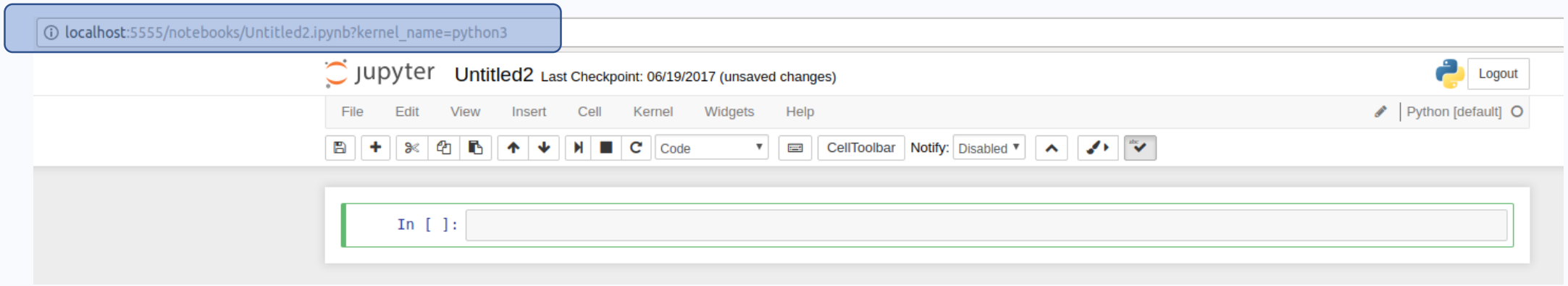
ssh port forwarding &jupyter. General scheme.



ssh –L 1234:localhost:5678 username@server

- launch a jupyter notebook at the server: *jupyter notebook --port=5678*
- open up a browser at your laptop and go to: localhost:1234
- If you launch the notebook via: *nohup jupyter notebook --port=5678&*
  then it won't stop at closing terminal (or you can use *screen/tmux* as well)

Jupyter notebook



Jupyter notebook examples: [oroszl.web.elte.hu/fiznum1/]
+ at our course github repository

Jupyter notebook

In [1]:
```
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:
```
1+1
```

Out[2]: 2

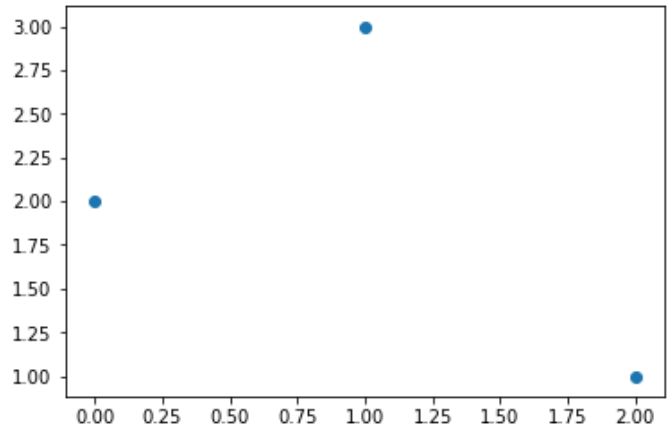
This is a $\LaTeX$ symbol with some math

$$\frac{x}{\sqrt{1+x^2}}$$

.

In [3]:
```
plt.scatter([0, 1, 2], [2, 3, 1])
```

Out[3]: <matplotlib.collections.PathCollection at 0x7f4473fea2b0>



Jupyter notebook examples: [oroszl.web.elte.hu/fiznum1/]
+ at our course github repository

Jupyter notebook

```
In [4]: !ls DeepLearningCourse/local/

class1.ipynb  multipleChoiceResponses.csv  pichart2.jpg  pichart.pdf
github        pataki_1.pptx                pichart.jpg   Untitled.ipynb

In [5]: lineCounts = !wc DeepLearningCourse/local/*
        lineCounts

Out[5]: ['     669     1487   146461 DeepLearningCourse/local/class1.ipynb',
         'wc: DeepLearningCourse/local/github: Is a directory',
         '       0        0        0 DeepLearningCourse/local/github',
         '   16716  1906084 24876561 DeepLearningCourse/local/multipleChoiceResponses.csv',
         '    8311    52263  3443821 DeepLearningCourse/local/pataki_1.pptx',
         '     370      906    47514 DeepLearningCourse/local/pichart2.jpg',
         '     440      911    50256 DeepLearningCourse/local/pichart.jpg',
         '     416     1646    20229 DeepLearningCourse/local/pichart.pdf',
         '     300      572     5290 DeepLearningCourse/local/Untitled.ipynb',
         '   27222  1963869 28590132 total']

In [6]: [[i.split()[-1], i.split()[0]] for i in lineCounts]

Out[6]: [['DeepLearningCourse/local/class1.ipynb', '669'],
         ['directory', 'wc:'],
         ['DeepLearningCourse/local/github', '0'],
         ['DeepLearningCourse/local/multipleChoiceResponses.csv', '16716'],
         ['DeepLearningCourse/local/pataki_1.pptx', '8311'],
         ['DeepLearningCourse/local/pichart2.jpg', '370'],
         ['DeepLearningCourse/local/pichart.jpg', '440'],
         ['DeepLearningCourse/local/pichart.pdf', '416'],
         ['DeepLearningCourse/local/Untitled.ipynb', '300'],
         ['total', '27222']]
```

Jupyter notebook examples: [oroszl.web.elte.hu/fiznum1/]
+ at our course github repository

Running locally:

- $pip install jupyter or pip3 install jupyter
- $jupyter notebook --no-browser --port=8888
- Open a browser and go to: localhost:8888

Windows:

- https://www.anaconda.com/download/

Running online:

- https://try.jupyter.org (can be full)
- https://ipy-nb.elte.hu (with caesar)

You will get access to institutional educational jupyter soon.

Ask if something doesn't work!

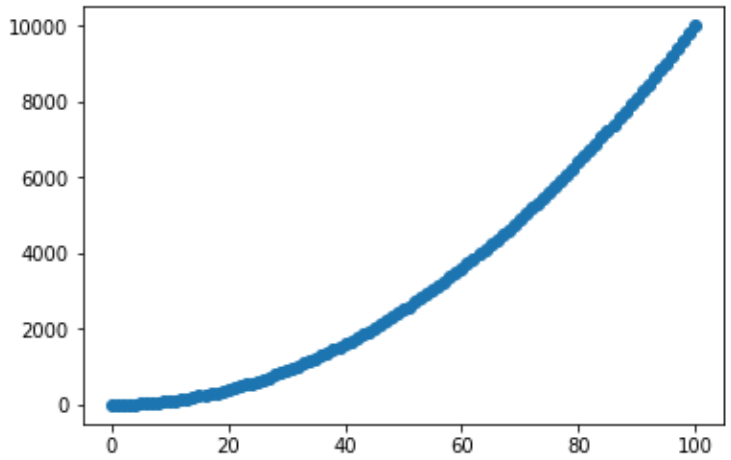
Plot the x^2 function between 0 and 100.

One solution:

In [1]:
```
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:
```
X = np.linspace(0, 100, 101)
Y = np.square(X)
```

In [3]:
```
plt.scatter(X, Y)
plt.show()
```



numpy

```
In [1]: import numpy as np
```

```
In [2]: mat1 = np.array([[0., 1., 2.],
                         [10., 20., 30.],
                         [100., 201., 300.]])
        print(mat1.shape)
        mat1
```

```
(3, 3)
```

```
Out[2]: array([[   0.,    1.,    2.],
               [  10.,   20.,   30.],
               [ 100.,  201.,  300.]])
```

```
In [3]: mat1.sum(axis = 0)
```

```
Out[3]: array([ 110.,  222.,  332.])
```

```
In [4]: mat1.sum(axis = 1)
```

```
Out[4]: array([   3.,   60.,  601.])
```

```
In [5]: mat1.sum()
```

```
Out[5]: 664.0
```

```
In [6]: mat1.T
```

```
Out[6]: array([[   0.,   10.,  100.],
               [   1.,   20.,  201.],
               [   2.,   30.,  300.]])
```

```
In [7]: inverse = np.linalg.inv(mat1)
        inverse
```

```
Out[7]: array([[-1.5,  5.1, -0.5],
               [ 0. , -10. ,  1. ],
               [ 0.5,  5. , -0.5]])
```

```
In [8]: inverse.dot(mat1)
```

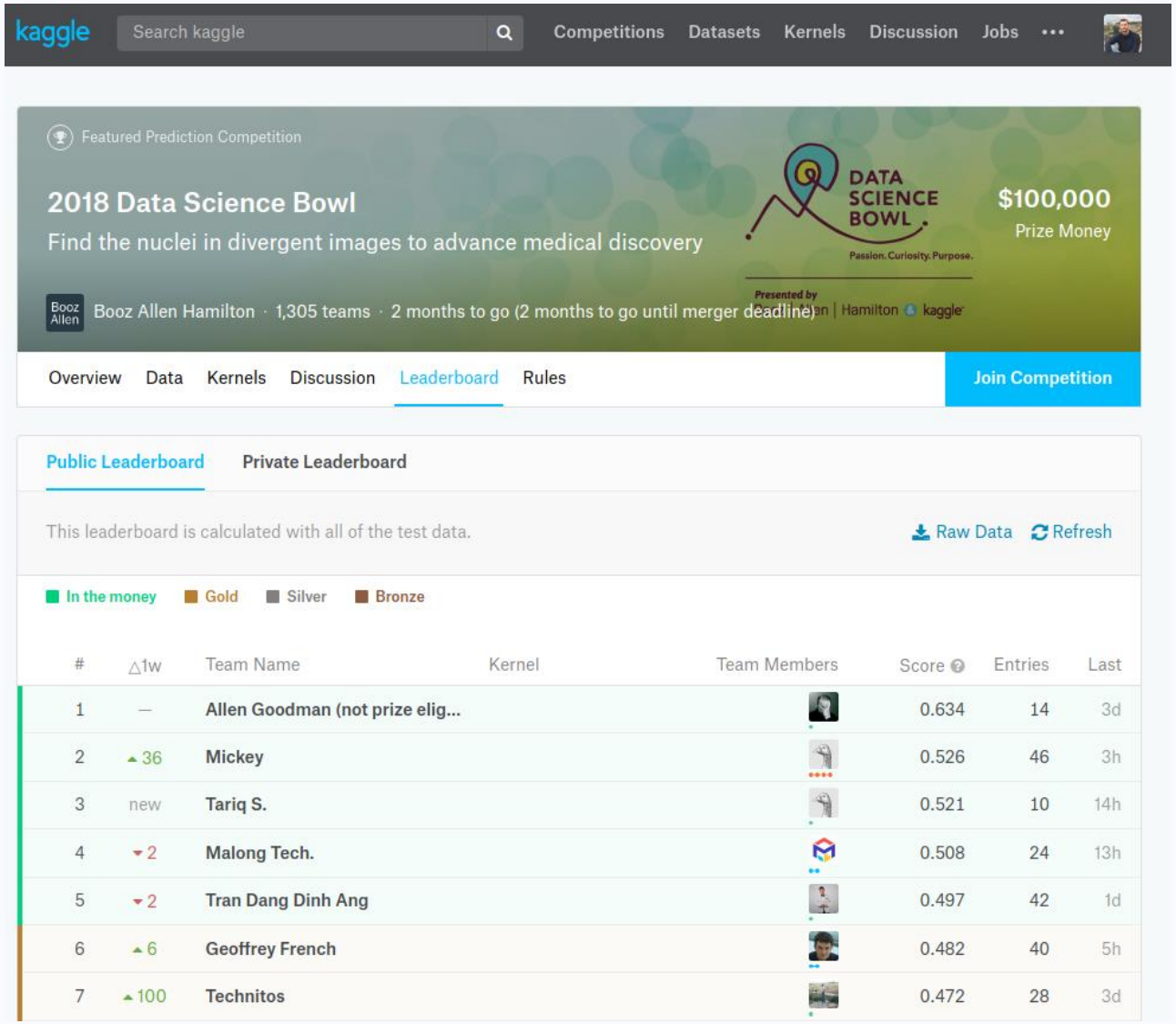
```
Out[8]: array([[  1.00000000e+00,  -2.84217094e-14,  -2.84217094e-14],
               [  0.00000000e+00,   1.00000000e+00,   0.00000000e+00],
               [  0.00000000e+00,   0.00000000e+00,   1.00000000e+00]])
```

numpy

```
In [1]: import numpy as np

In [2]: mat = np.random.randn(4, 4)
        mat

Out[2]: array([[ 0.24663924,  0.38151637, -0.45373846,  1.52077717],
               [ 1.25339265, -0.23388767,  1.14458239, -0.19646592],
               [ 0.11703608, -0.69976066, -1.34154909,  0.30857629],
               [-0.10569809,  0.35166965, -0.55060552,  2.15144935]])

In [3]: a = np.array([0., 1., 2., 5.]) #rank 1 array strange behaviour
        b = np.array([[0.], [1.], [2.], [5.]]) # column vector
        c = np.array([[0., 1., 2., 5.]]) # row vector

In [4]: print(a)
        print(a.reshape(4, 1)) # this is b actually

        [ 0.  1.  2.  5.]
        [[ 0.]
         [ 1.]
         [ 2.]
         [ 5.]]
```

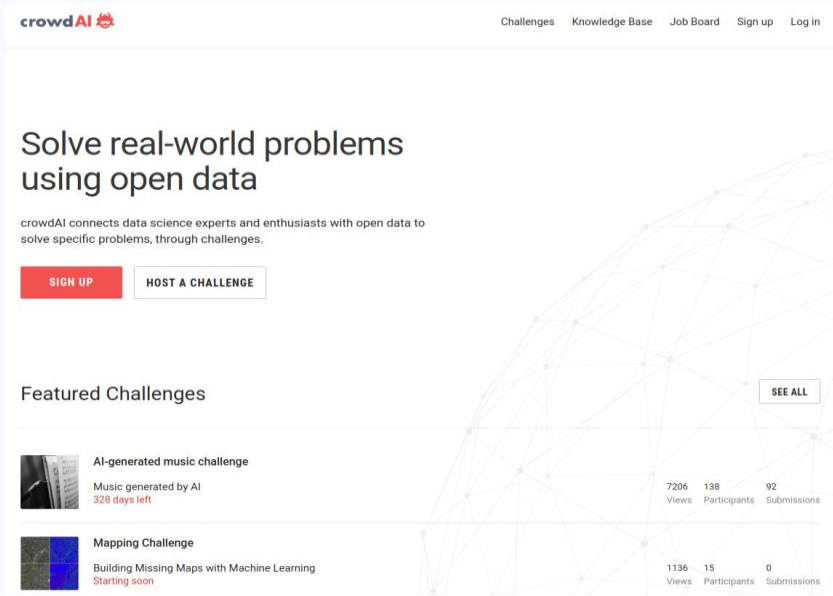
```
In [5]: print('a:', a.shape, a.T.shape) # better not to use rank 1 array
                                       # (or be sure what you are doing!)
        print('b:', b.shape, b.T.shape)
        print('c:', c.shape, c.T.shape)

        a: (4,) (4,)
        b: (4, 1) (1, 4)
        c: (1, 4) (4, 1)

In [6]: b.dot(c)

Out[6]: array([[  0.,   0.,   0.,   0.],
               [  0.,   1.,   2.,   5.],
               [  0.,   2.,   4.,  10.],
               [  0.,   5.,  10.,  25.]])

In [7]: a.dot(b)

Out[7]: array([ 30.])

In [8]: a.dot(c)

        ---------------------------------------------------------------------------
        ValueError                                Traceback (most recent call last)
        <ipython-input-8-5d91888f8066> in <module>()
        ----> 1 a.dot(c)

        ValueError: shapes (4,) and (1,4) not aligned: 4 (dim 0) != 1 (dim 0)
```

Competitive machine learning



[https://kaggle.com/c/data-science-bowl-2018]

[https://dreamchallenges.com]



[https://crowdai.org]

Happiness detector: one of the your projects

A Kaggle inClass challange hosted for you! Link will be sent out soon.

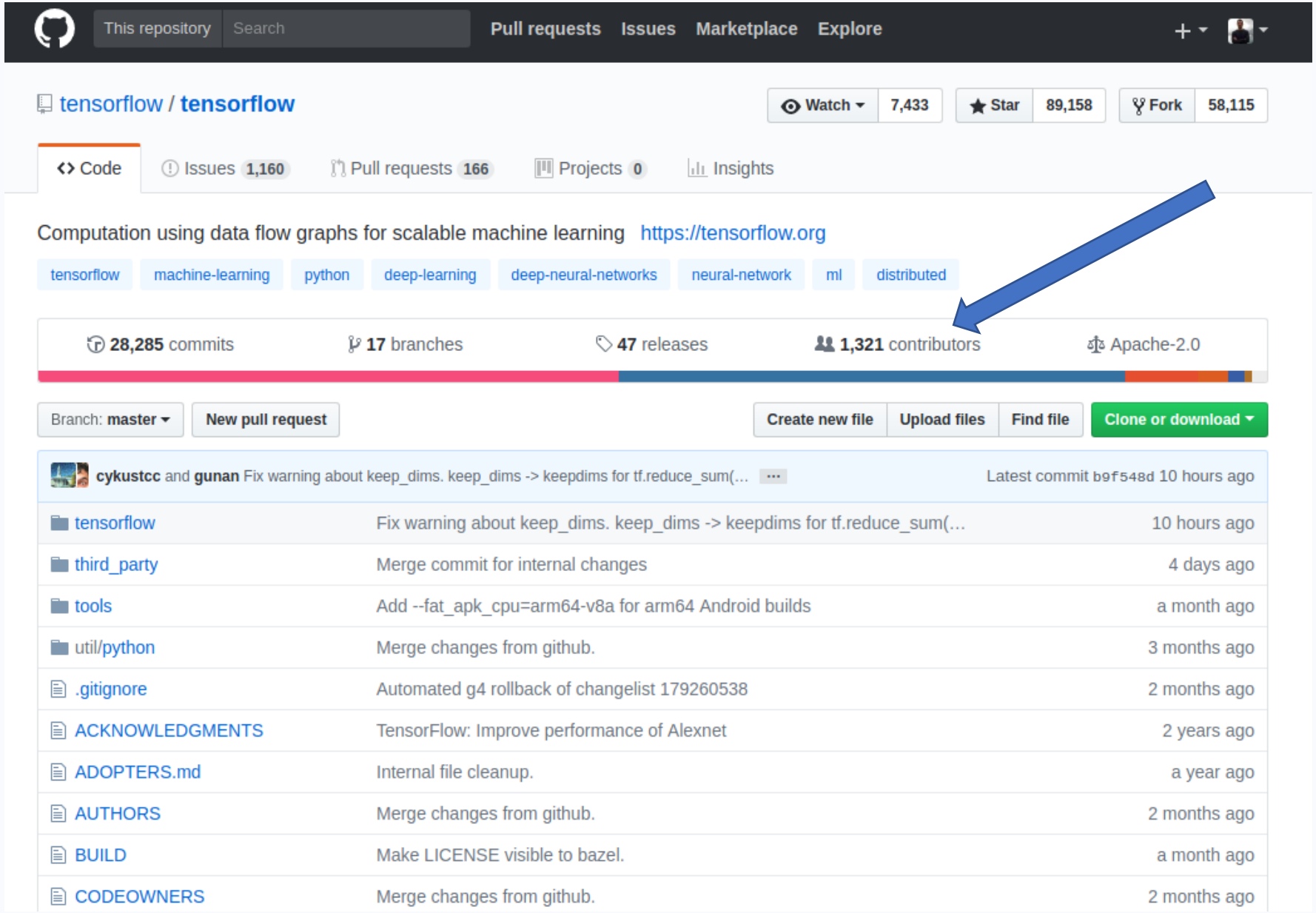
In theory

```
patbaa@eltesgep:~/TMP/thesis$ ls
dev   final
```

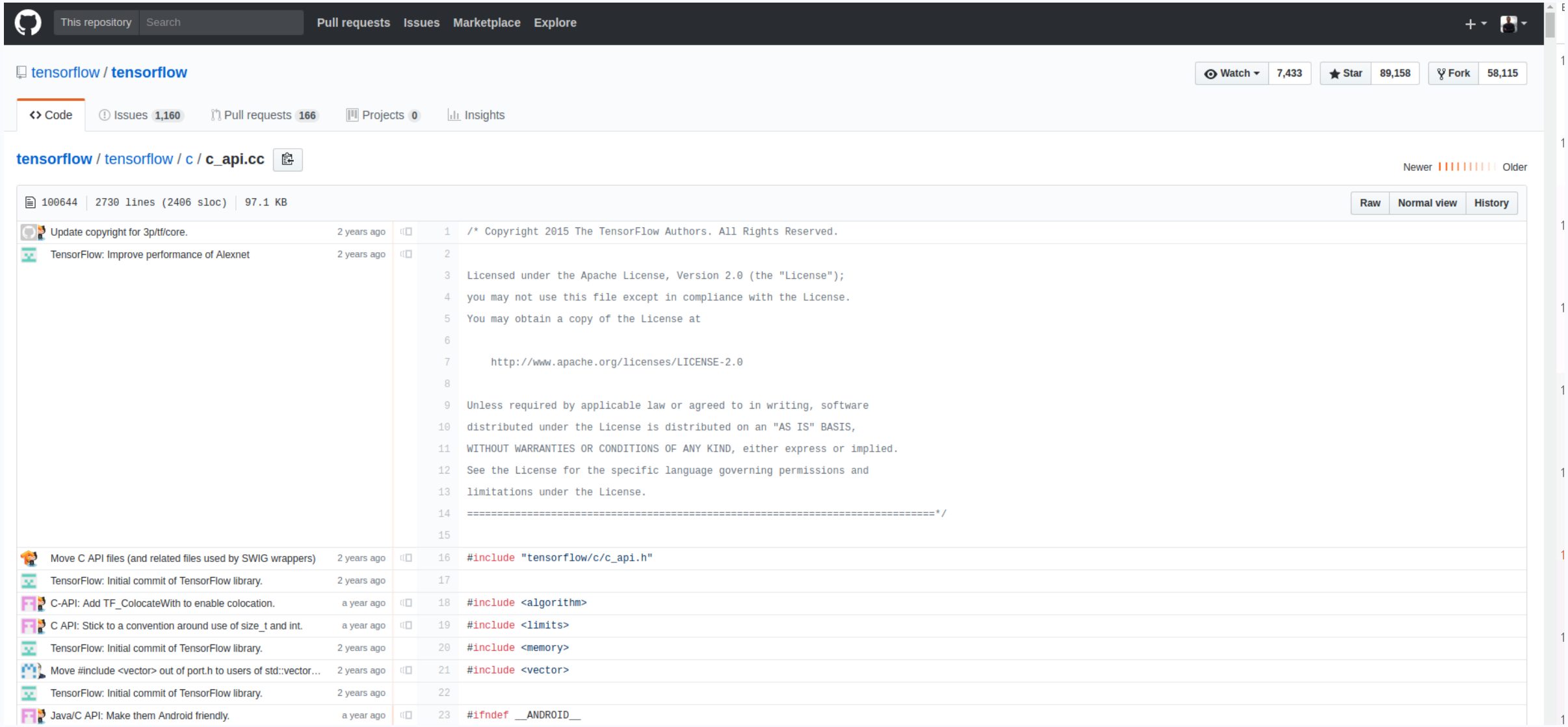
In reality

```
patbaa@eltesgep:~/TMP/without_git$ ls -lh
total 40K
drwxrwxr-x 2 patbaa patbaa 4,0K febr  11 14:28 final
drwxrwxr-x 2 patbaa patbaa 4,0K febr  11 14:28 FINAL
drwxrwxr-x 2 patbaa patbaa 4,0K febr  11 14:28 final2
drwxrwxr-x 2 patbaa patbaa 4,0K febr  11 14:28 final3
drwxrwxr-x 2 patbaa patbaa 4,0K febr  11 14:29 final4
drwxrwxr-x 2 patbaa patbaa 4,0K febr  11 14:28 finalfinal
drwxrwxr-x 2 patbaa patbaa 4,0K febr  11 14:28 finalfinal2
drwxrwxr-x 2 patbaa patbaa 4,0K febr  11 14:29 finalfinal3
drwxrwxr-x 2 patbaa patbaa 4,0K febr  11 14:28 notfinal
drwxrwxr-x 2 patbaa patbaa 4,0K febr  11 14:28 reallyfinal
patbaa@eltesgep:~/TMP/without_git$
```

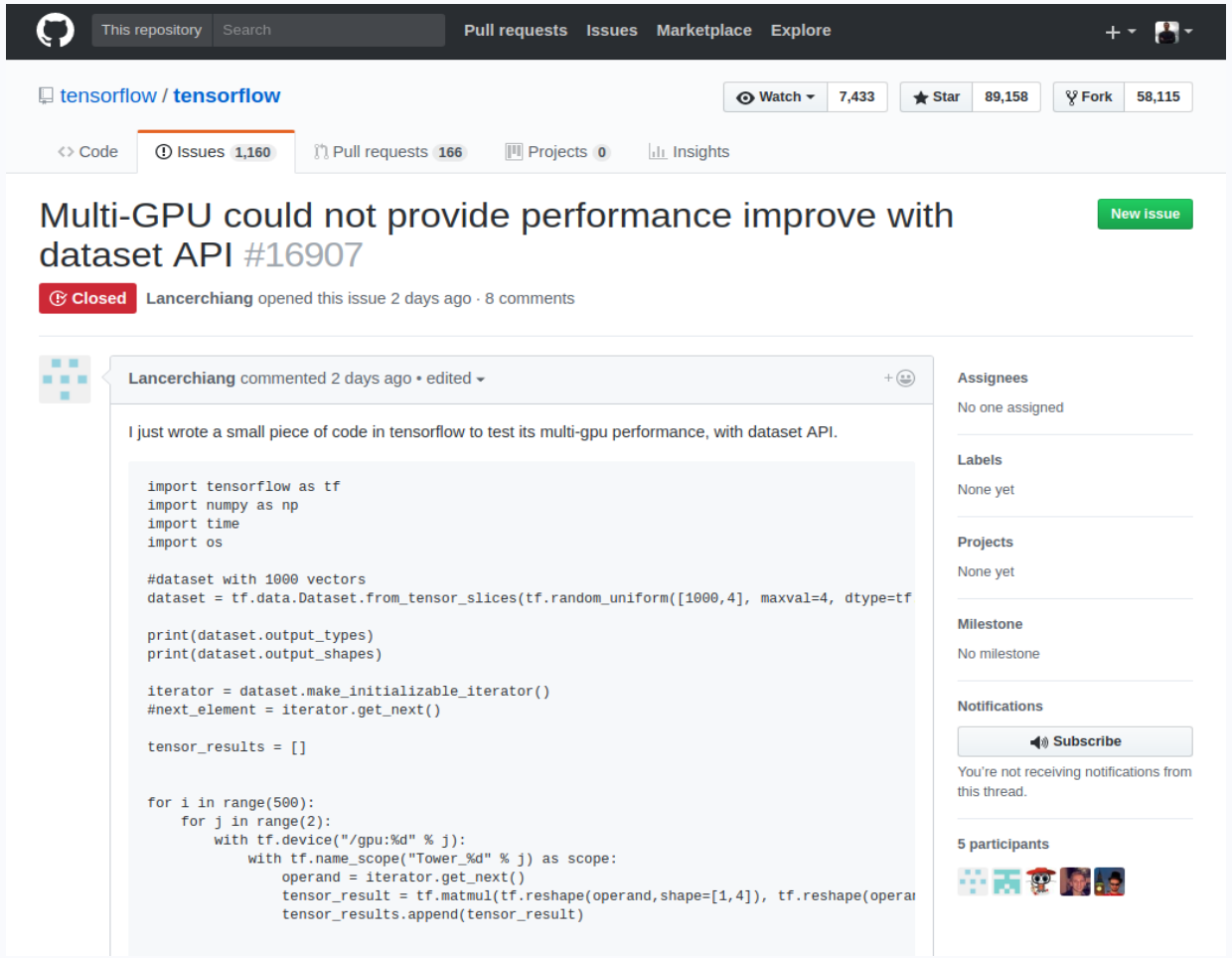
Git + github: collaboration &version control



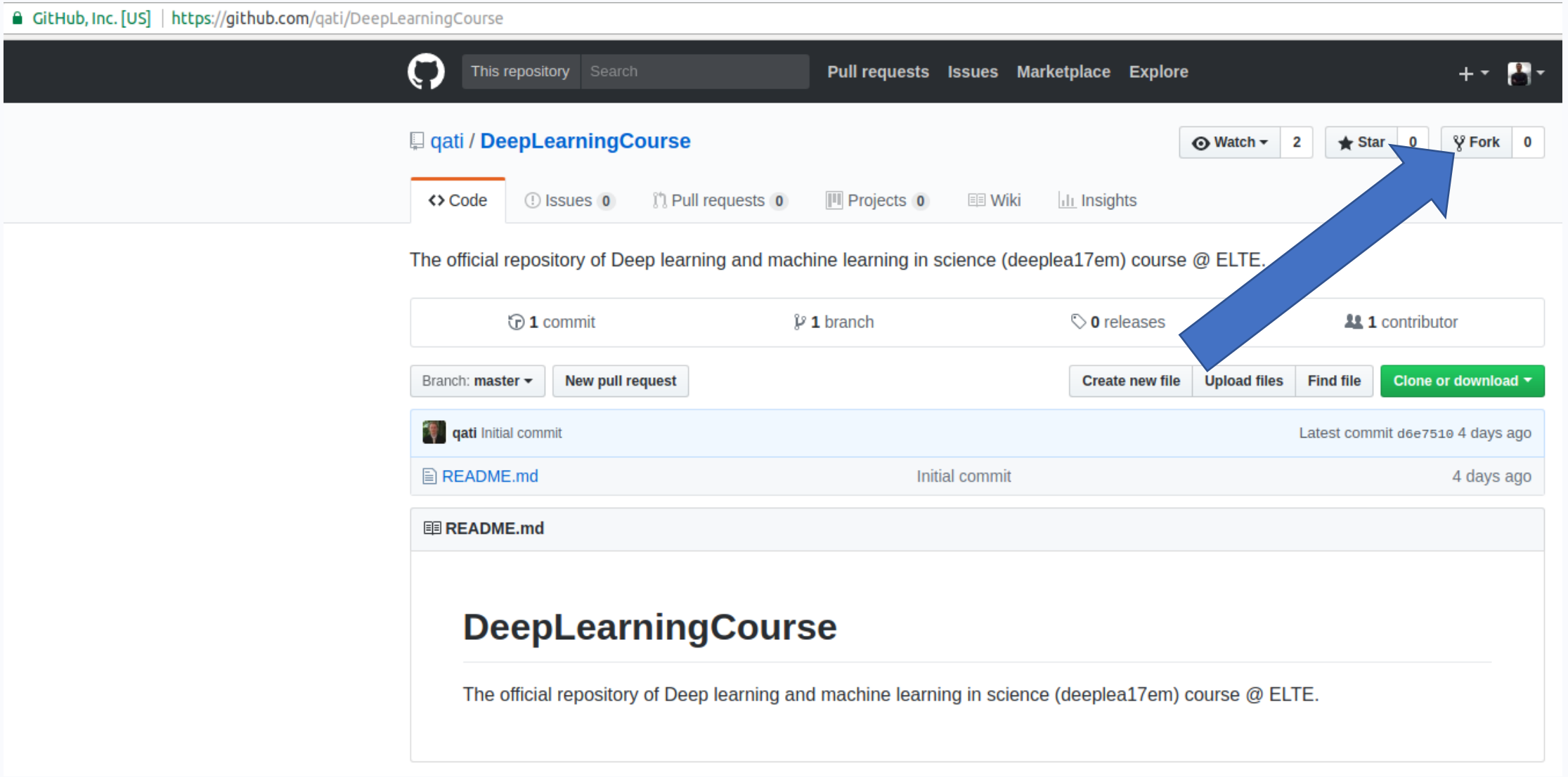
Git + github: git blame



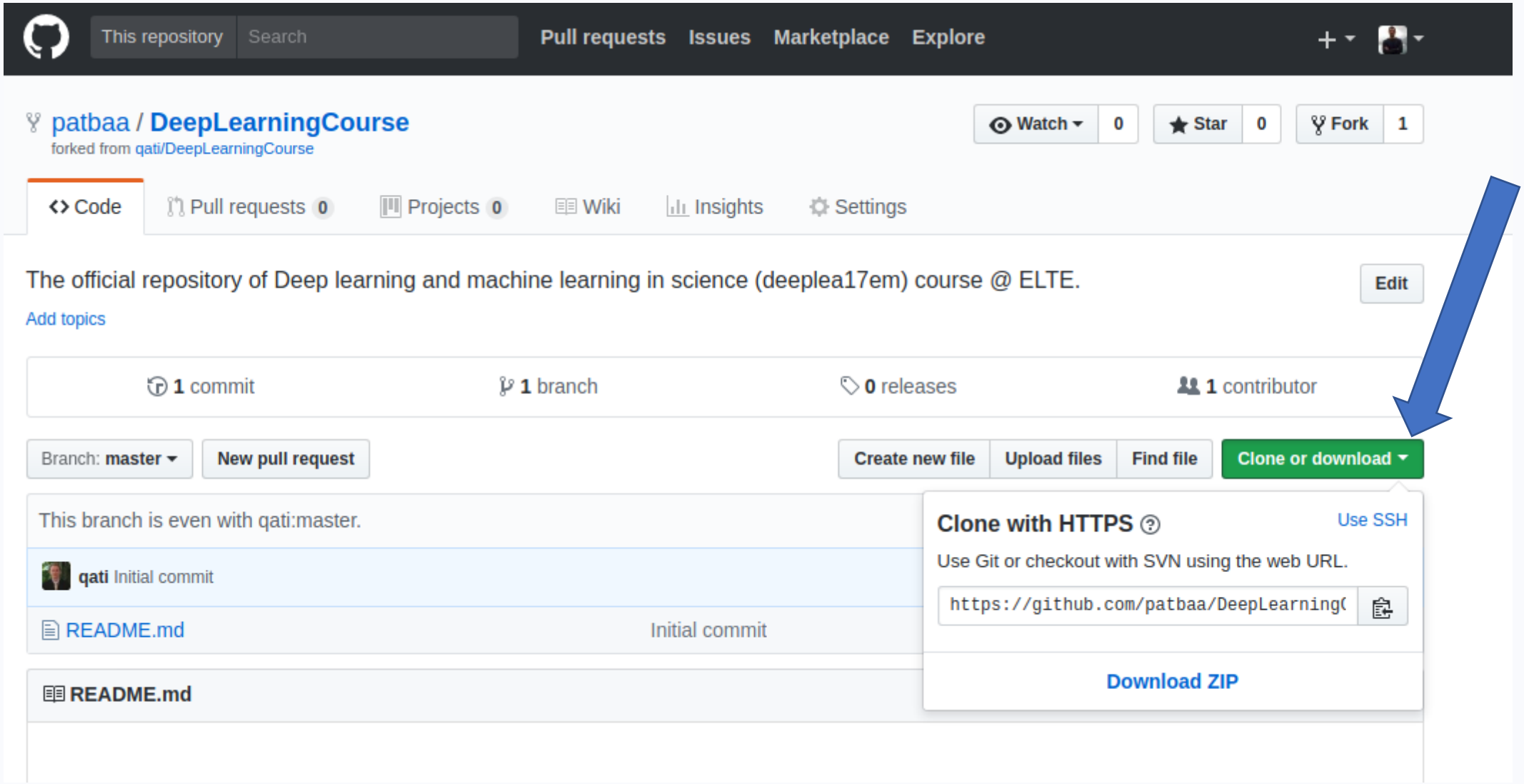
If you have question that can be useful for others too, open an issue!



Git + github: fork



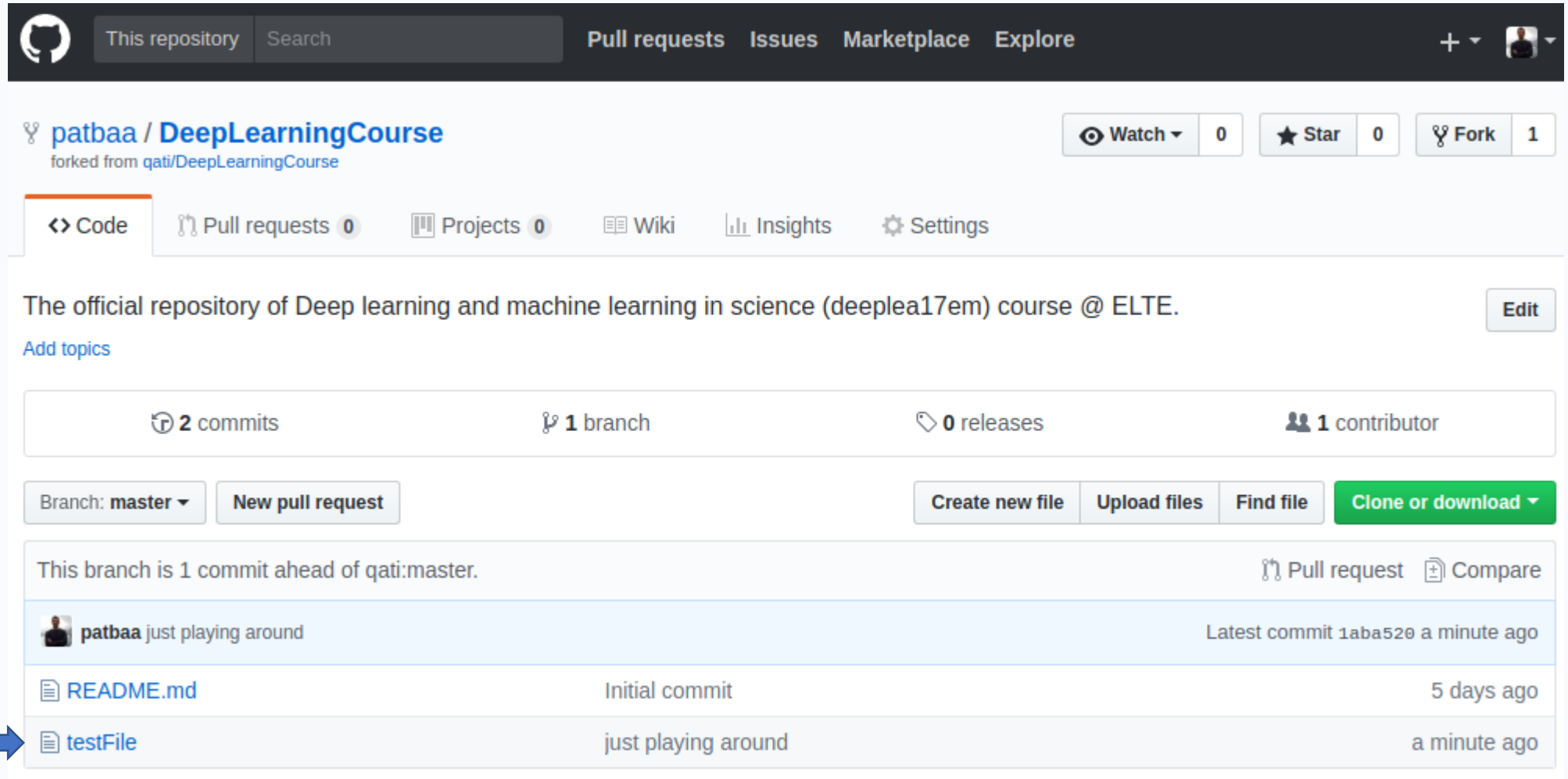
Git + github


This repository
Search
Pull requests
Issues
Marketplace
Explore
patbaa / DeepLearningCourse
forked from qati/DeepLearningCourse
Watch 0
Star 0
Fork 1
Code
Pull requests 0
Projects 0
Wiki
Insights
Settings
The official repository of Deep learning and machine learning in science (deeplea17em) course @ ELTE.
Edit
Add topics
1 commit
1 branch
0 releases
1 contributor
Branch: master
New pull request
Create new file
Upload files
Find file
Clone or download
This branch is even with qati:master.
Clone with HTTPS
Use SSH
Use Git or checkout with SVN using the web URL.
https://github.com/patbaa/DeepLearning(
qati Initial commit
README.md
Initial commit
Download ZIP
README.md

Git + github

```
patbaa@eltesgep:~/TMP$ git clone https://github.com/patbaa/DeepLearningCourse.git
Cloning into 'DeepLearningCourse'...
remote: Counting objects: 3, done.
remote: Compressing objects: 100% (2/2), done.
remote: Total 3 (delta 0), reused 0 (delta 0), pack-reused 0
Unpacking objects: 100% (3/3), done.
Checking connectivity... done.
patbaa@eltesgep:~/TMP$ cd DeepLearningCourse/
patbaa@eltesgep:~/TMP/DeepLearningCourse$ echo 'testtest' > testFile
patbaa@eltesgep:~/TMP/DeepLearningCourse$ git add testFile
patbaa@eltesgep:~/TMP/DeepLearningCourse$ git commit -m 'just playing around'
[master 1aba520] just playing around
 1 file changed, 1 insertion(+)
 create mode 100644 testFile
patbaa@eltesgep:~/TMP/DeepLearningCourse$ git push origin master
Username for 'https://github.com': patbaa
Password for 'https://patbaa@github.com':
Counting objects: 3, done.
Delta compression using up to 8 threads.
Compressing objects: 100% (2/2), done.
Writing objects: 100% (3/3), 296 bytes | 0 bytes/s, done.
Total 3 (delta 0), reused 0 (delta 0)
To https://github.com/patbaa/DeepLearningCourse.git
   d6e7510..1aba520  master -> master
patbaa@eltesgep:~/TMP/DeepLearningCourse$
```

Git + github



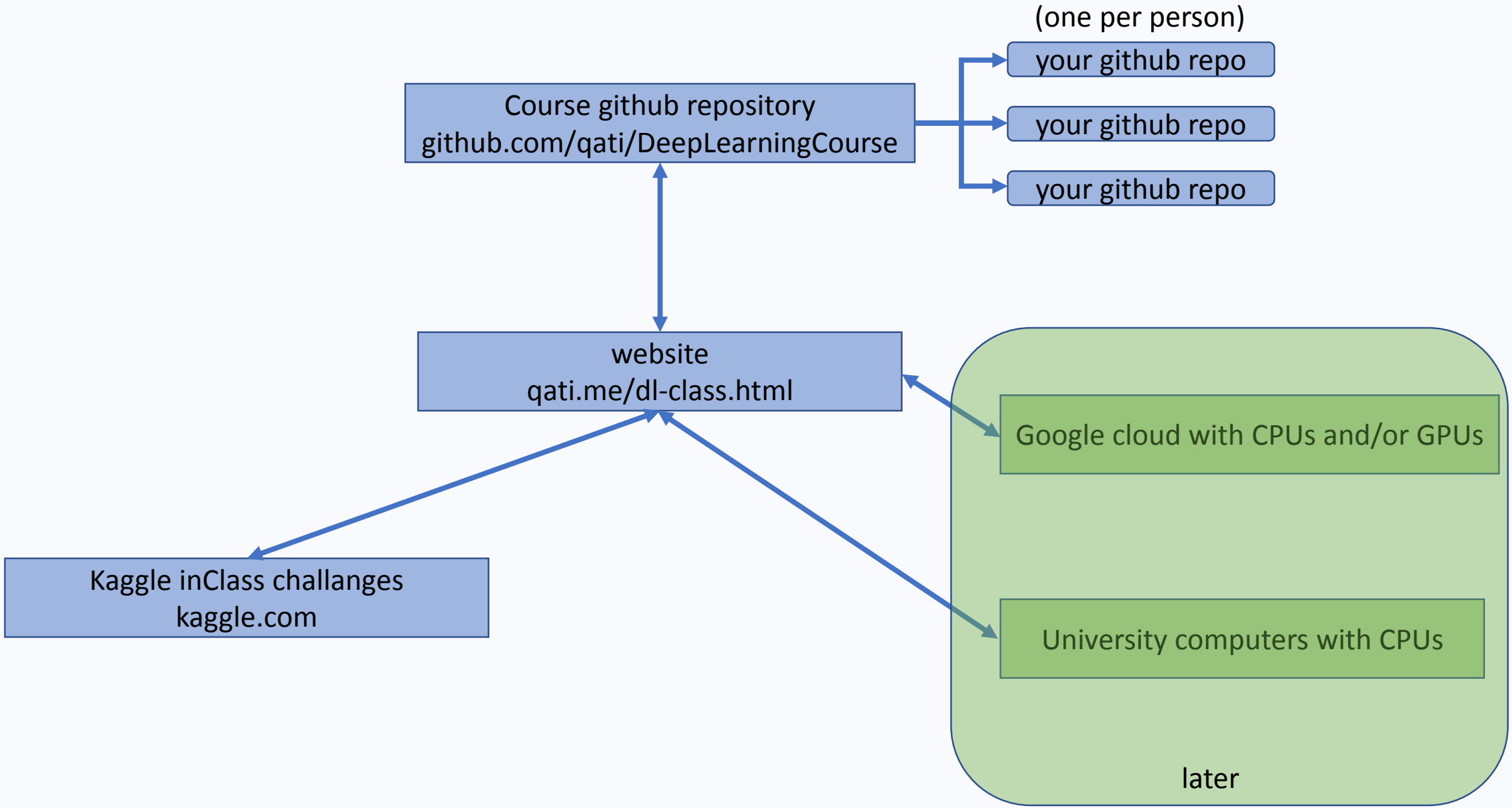
TIP: after registration add your *caesar* email to unlock some premium features

Git + github: steps to follow

1. Register* at https://github.com
2. Fork the qati/DeepLearningCourse repository
3. Download git (sudo apt-get install git) or https://git-scm.com/downloads (for windows)
4. git config --global user.name 'Balint Armin Pataki'
5. git config --global user.email 'balintpataki@example.com'
6. Clone the forked (your own) repository to your computer
7. Do some work.
8. Git add workfile.txt
9. git commit -m 'this is my commit message'
10. git push origin master

*TIP: after registration add your *caesar* email to unlock some premium features
Really useful step-by-step git intro: https://try.github.io

The structure



Register at the following sites:

- https://github.com
- https://kaggle.com

Install the following programs:

- git
- python3
- Jupyter notebook

Visit our github and course site for the course materials (be updated weekly):

- https://github.com/qati/DeepLearningCourse
- http://qati.me/dl-class.html (fill the form under technical infos)

Complete the first assignment(s) that you can find in our github repo folder:

https://github.com/qati/DeepLearningCourse/assignments/