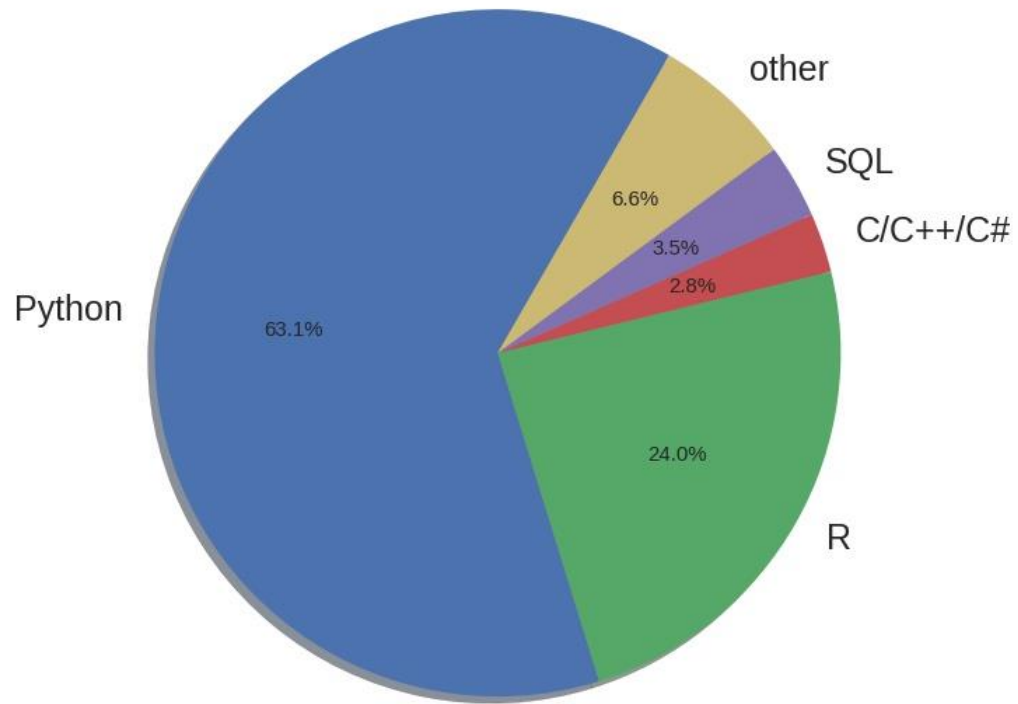


# Technical details

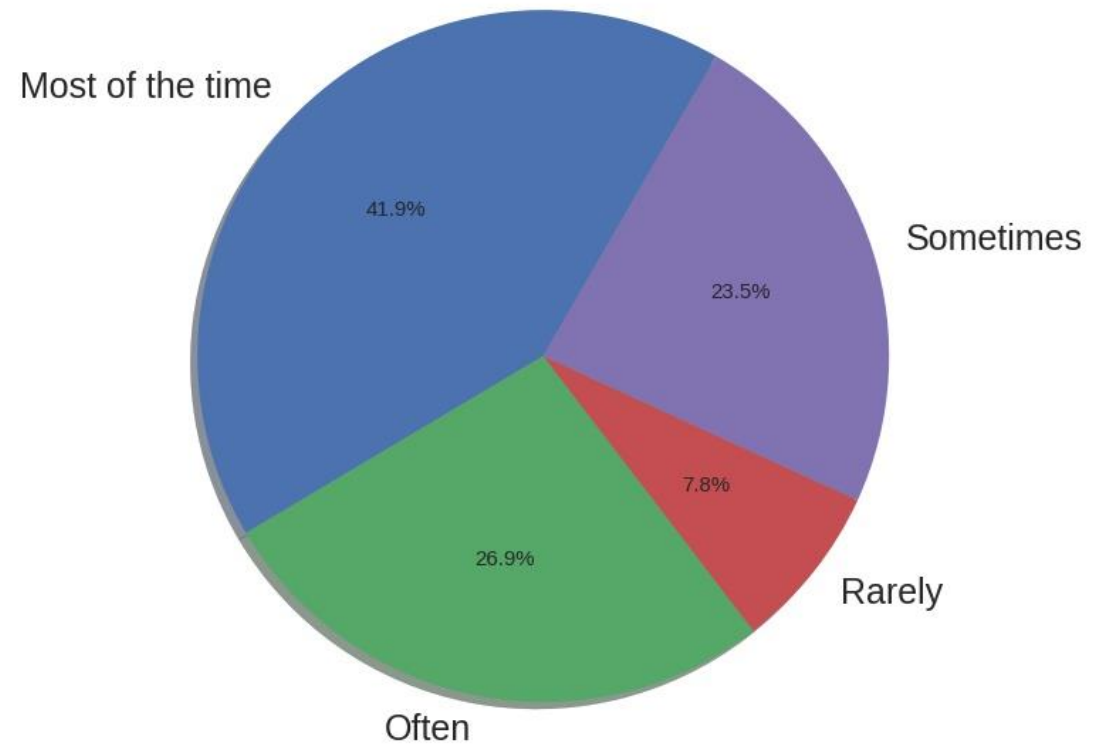
Working tools, python, numpy, jupyter notebooks, github, Kaggle, ssh

Examples at the course site: [qati.me/dl-class.html](http://qati.me/dl-class.html)  
(Bálint Ármin Pataki)

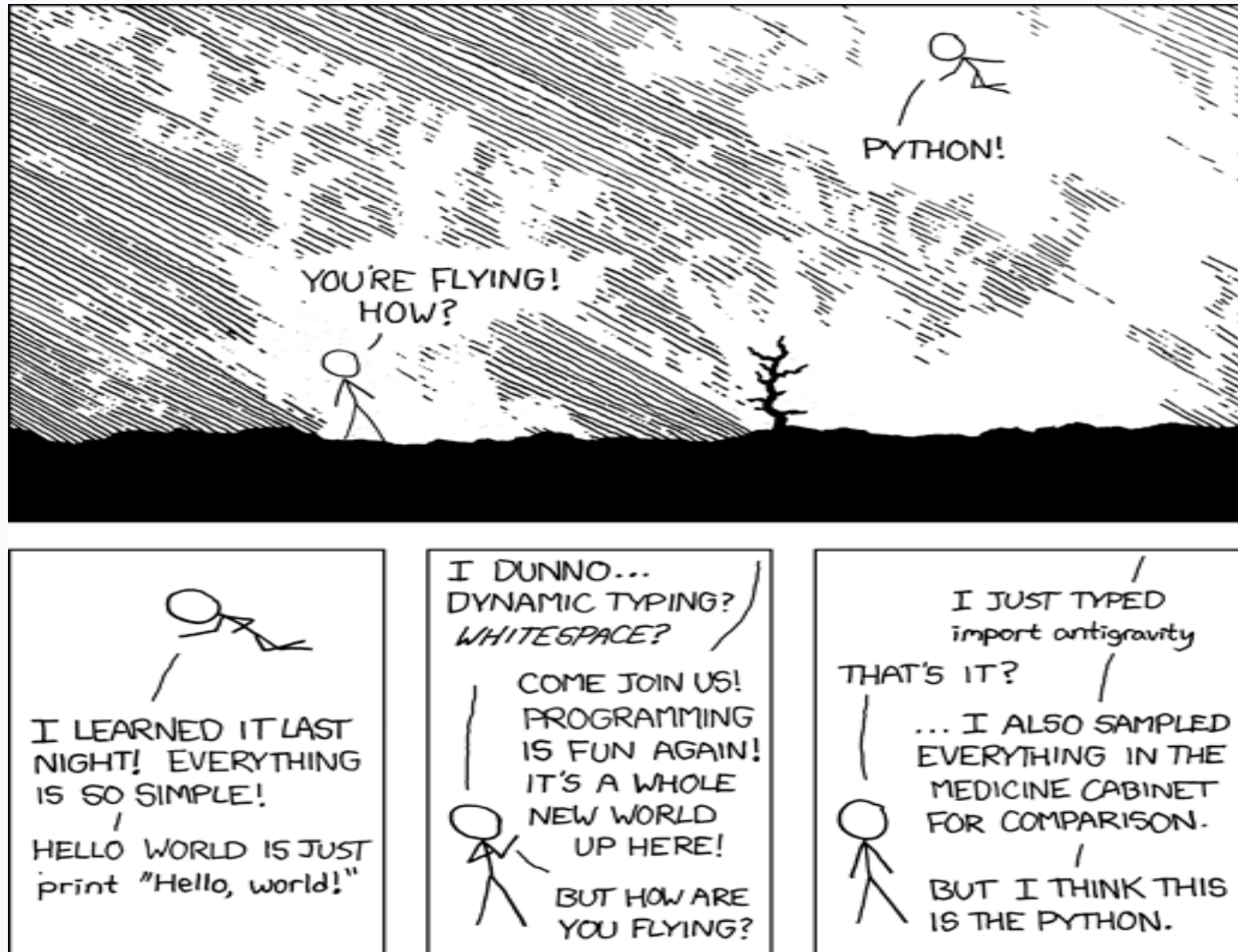
LanguageRecommendationSelect



WorkToolsFrequencyJupyter



[Kaggle ML and DataScience Survey, 2017]



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

Kooplex Reports Projects Gitlab Owncloud Help

Hello bpataki! Log off

My projects Public projects New project

Start SNPHeight Using image:basic

Start CAUSAL ( owned by icsabai ) Using image:basic

Start Data Exploration and Visualisation ( owned by jegesm ) Using image:basic

Open DeepPhotoZ ( owned by icsabai ) Using image:basic

Using image:basic

Using image:basic

Using image:basic

Using image:basic

localhost:5555/notebooks/DeepLearningCourse/local/class1.ipynb

jupyter class1 Last Checkpoint: 7 minutes ago (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Python [default]

```
ax1.axis('equal')
plt.savefig('pichart.jpg')
plt.show()
```

LanguageRecommendationSelect

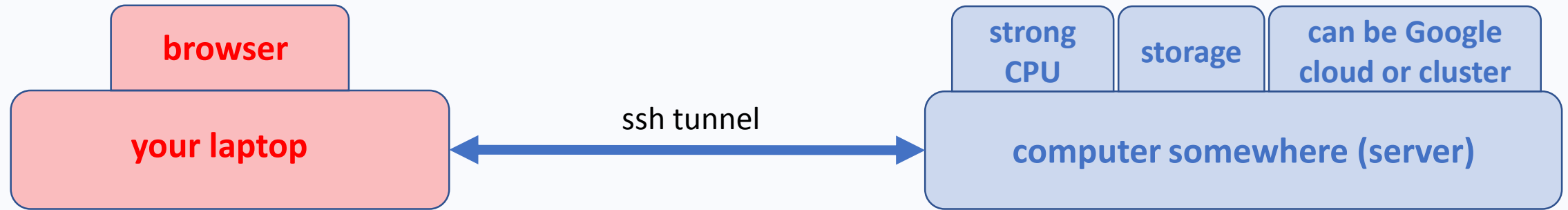
Language	Percentage
Python	63.1%
R	24.0%
other	6.6%
SQL	3.5%
C/C++/C#	2.8%

```
In [10]: w = tmp.WorkToolsFrequencyJupyter.dropna().tolist()
wFreq = dict(Counter(w))
wFreq
```

Out[10]: {'Most of the time': 1327, 'Often': 851, 'Rarely': 246, 'Sometimes': 745}

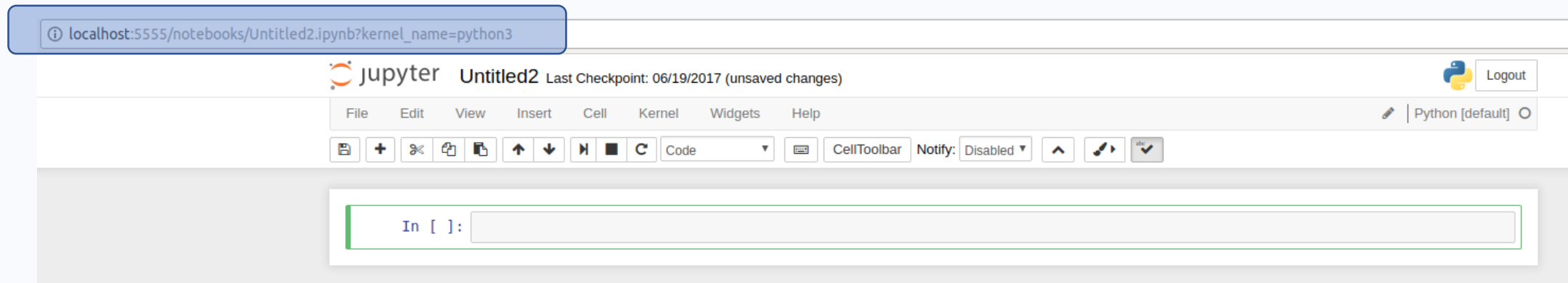


# ssh port forwarding & jupyter. General scheme.



`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 examples: [\[oroszl.web.elte.hu/fiznum1/\]](http://oroszl.web.elte.hu/fiznum1/)  
+ at our course github repository

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

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

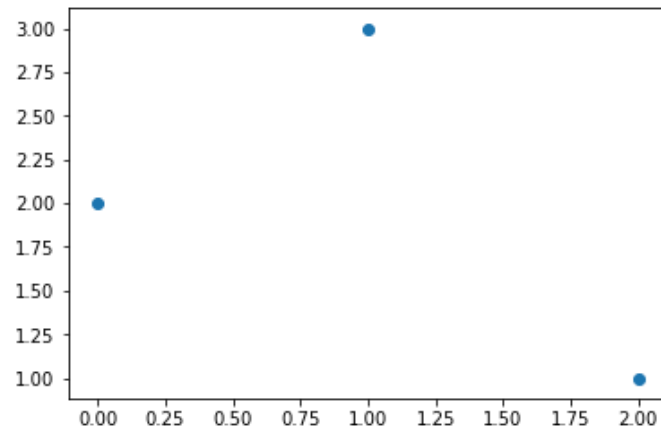
```
Out[2]: 2
```

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/](http://oroszl.web.elte.hu/fiznum1/)  
+ at our course github repository

```
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: [\[oroszl.web.elte.hu/fiznum1/\]](http://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!

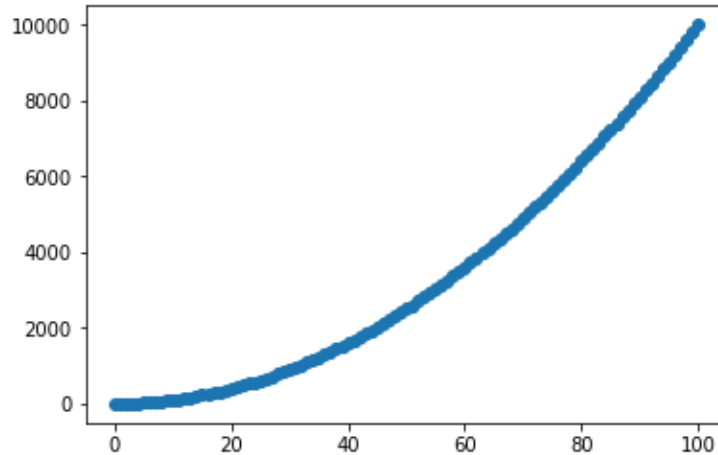
Plot the  $x^2$  function between 0 and 100.

## 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()
```



```
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]])
```

```
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.])
```

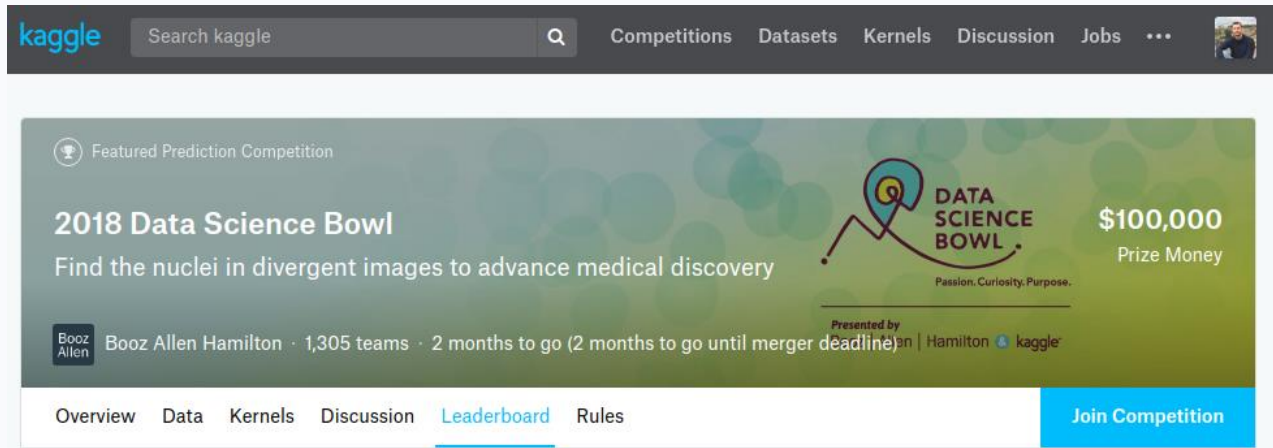
```
In [8]: a.dot(c)
```

```
-----
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



**2018 Data Science Bowl**  
Find the nuclei in divergent images to advance medical discovery

**\$100,000**  
Prize Money

Booz Allen Hamilton · 1,305 teams · 2 months to go (2 months to go until merger deadline)

Overview Data Kernels Discussion **Leaderboard** Rules

[Join Competition](#)

Public Leaderboard Private Leaderboard

This leaderboard is calculated with all of the test data. [Raw Data](#) [Refresh](#)

In the money Gold Silver Bronze

#	Δ1w	Team Name	Kernel	Team Members	Score	Entries	Last
1	—	Allen Goodman (not prize elig...			0.634	14	3d
2	▲36	Mickey			0.526	46	3h
3	new	Tariq S.			0.521	10	14h
4	▼2	Malong Tech.			0.508	24	13h
5	▼2	Tran Dang Dinh Ang			0.497	42	1d
6	▲6	Geoffrey French			0.482	40	5h
7	▲100	Technitos			0.472	28	3d

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



**DREAM CHALLENGES**

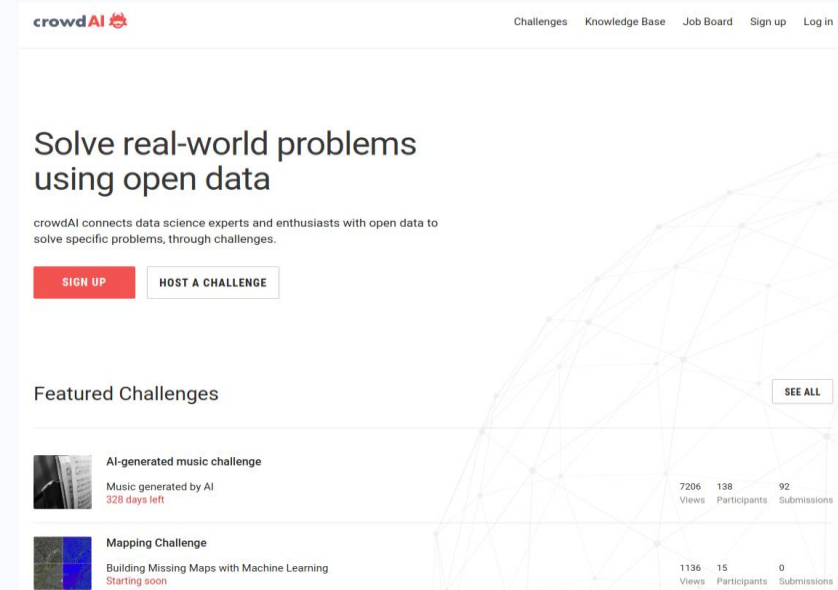
Community insight  
Community impact

**DREAM CHALLENGES**

Multi-targeting Drug DREAM Challenge

Open October 5, 2017 - February 26, 2018

[<https://dreamchallenges.com>]



**crowdAI**

Solve real-world problems using open data

crowdAI connects data science experts and enthusiasts with open data to solve specific problems, through challenges.

[SIGN UP](#) [HOST A CHALLENGE](#)

Featured Challenges

AI-generated music challenge  
Music generated by AI  
528 days left  
7206 Views 138 Participants 92 Submissions

Mapping Challenge  
Building Missing Maps with Machine Learning  
Starting soon  
1136 Views 15 Participants 0 Submissions

[<https://crowdai.org>]

# Happiness detector: one of the your projects



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

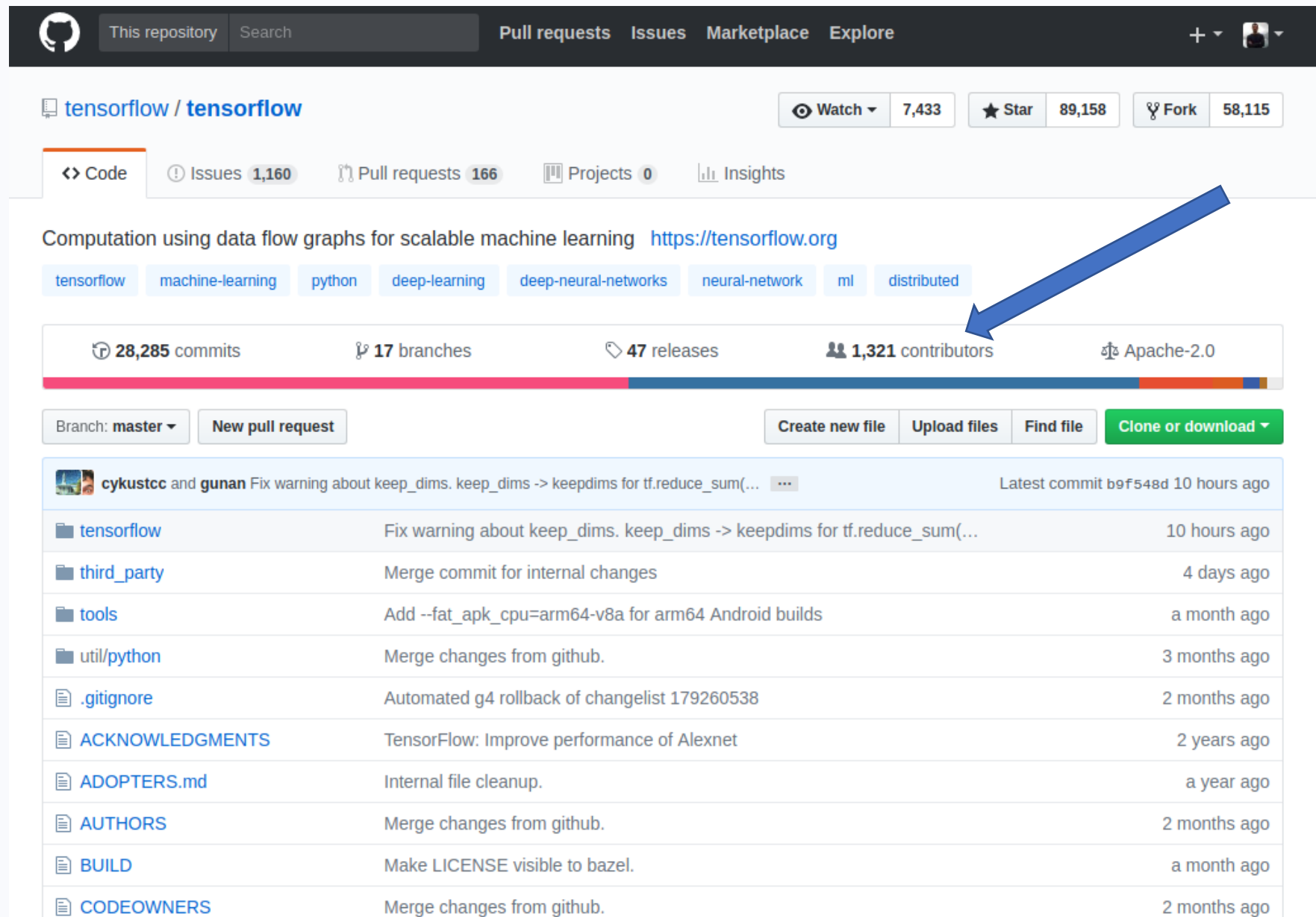
## In theory

```
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



GitHub repository page for **tensorflow / tensorflow**. The page shows repository statistics and a list of recent commits.

Repository statistics:

- Watch: 7,433
- Star: 89,158
- Fork: 58,115

Navigation tabs: Code, Issues (1,160), Pull requests (166), Projects (0), Insights.

Repository description: Computation using data flow graphs for scalable machine learning <https://tensorflow.org>

Tags: tensorflow, machine-learning, python, deep-learning, deep-neural-networks, neural-network, ml, distributed.

Repository statistics:

- 28,285 commits
- 17 branches
- 47 releases
- 1,321 contributors
- Apache-2.0 license

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

Recent commits:

Commit	Message	Time
<a href="#">cykustcc and gunan</a>	Fix warning about keep_dims. keep_dims -> keepdims for tf.reduce_sum(...)	10 hours ago
<a href="#">tensorflow</a>	Fix warning about keep_dims. keep_dims -> keepdims for tf.reduce_sum(...)	10 hours ago
<a href="#">third_party</a>	Merge commit for internal changes	4 days ago
<a href="#">tools</a>	Add --fat_apk_cpu=arm64-v8a for arm64 Android builds	a month ago
<a href="#">util/python</a>	Merge changes from github.	3 months ago
<a href="#">.gitignore</a>	Automated g4 rollback of changelist 179260538	2 months ago
<a href="#">ACKNOWLEDGMENTS</a>	TensorFlow: Improve performance of Alexnet	2 years ago
<a href="#">ADOPTERS.md</a>	Internal file cleanup.	a year ago
<a href="#">AUTHORS</a>	Merge changes from github.	2 months ago
<a href="#">BUILD</a>	Make LICENSE visible to bazel.	a month ago
<a href="#">CODEOWNERS</a>	Merge changes from github.	2 months ago



# Git + github: git blame

tensorflow / tensorflow

Watch 7,433 Star 89,158 Fork 58,115

Code Issues 1,160 Pull requests 166 Projects 0 Insights

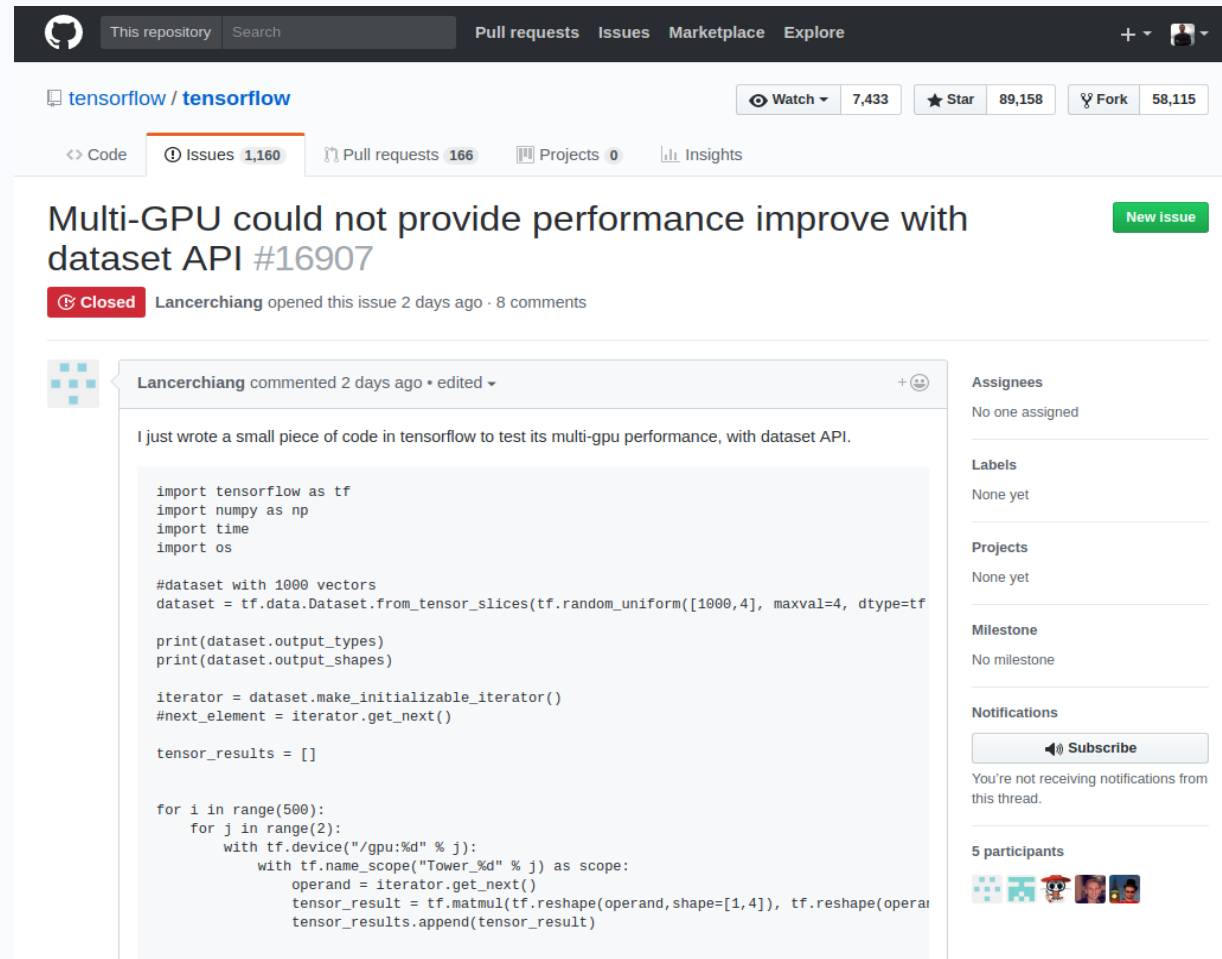
tensorflow / tensorflow / c / c\_api.cc

100644 | 2730 lines (2406 sloc) | 97.1 KB

Raw Normal view History

Commit	Author	Time	Line	Code
Update copyright for 3p/tf/core.		2 years ago	1	/* Copyright 2015 The TensorFlow Authors. All Rights Reserved.
TensorFlow: Improve performance of Alexnet		2 years ago	2	
			3	Licensed under the Apache License, Version 2.0 (the "License");
			4	you may not use this file except in compliance with the License.
			5	You may obtain a copy of the License at
			6	
			7	http://www.apache.org/licenses/LICENSE-2.0
			8	
			9	Unless required by applicable law or agreed to in writing, software
			10	distributed under the License is distributed on an "AS IS" BASIS,
			11	WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
			12	See the License for the specific language governing permissions and
			13	limitations under the License.
			14	===== */
			15	
Move C API files (and related files used by SWIG wrappers)		2 years ago	16	#include "tensorflow/c/c_api.h"
TensorFlow: Initial commit of TensorFlow library.		2 years ago	17	
C-API: Add TF_ColocateWith to enable colocation.		a year ago	18	#include <algorithm>
C-API: Stick to a convention around use of size_t and int.		a year ago	19	#include <limits>
TensorFlow: Initial commit of TensorFlow library.		2 years ago	20	#include <memory>
Move #include <vector> out of port.h to users of std::vector...		2 years ago	21	#include <vector>
TensorFlow: Initial commit of TensorFlow library.		2 years ago	22	
Java/C API: Make them Android friendly.		a year ago	23	#ifndef __ANDROID__

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



The screenshot shows the GitHub interface for the tensorflow/tensorflow repository. The issue title is "Multi-GPU could not provide performance improve with dataset API #16907", marked as "Closed". A comment by Lancerchiang, posted 2 days ago, includes a code snippet for testing multi-gpu performance. The code imports tensorflow, numpy, time, and os, then creates a dataset with 1000 vectors and iterates over it for 500 iterations across 2 GPUs. The right sidebar shows no assignees, labels, projects, or milestones, and a "Subscribe" button.

tensorflow / tensorflow

Watch 7,433 Star 89,158 Fork 58,115

Code Issues 1,160 Pull requests 166 Projects 0 Insights

### Multi-GPU could not provide performance improve with dataset API #16907

**Closed** Lancerchiang opened this issue 2 days ago · 8 comments

Lancerchiang commented 2 days ago • edited

I just wrote a small piece of code in tensorflow to test its multi-gpu performance, with dataset API.

```
import tensorflow as tf
import numpy as np
import time
import os

#dataset with 1000 vectors
dataset = tf.data.Dataset.from_tensor_slices(tf.random_uniform([1000,4], maxval=4, dtype=tf.float32))

print(dataset.output_types)
print(dataset.output_shapes)

iterator = dataset.make_initializable_iterator()
#next_element = iterator.get_next()

tensor_results = []

for i in range(500):
    for j in range(2):
        with tf.device("/gpu:%d" % j):
            with tf.name_scope("Tower_%d" % j) as scope:
                operand = iterator.get_next()
                tensor_result = tf.matmul(tf.reshape(operand,shape=[1,4]), tf.reshape(operand,shape=[1,4]))
                tensor_results.append(tensor_result)
```

**Assignees**  
No one assigned

**Labels**  
None yet

**Projects**  
None yet

**Milestone**  
No milestone

**Notifications**  
Subscribe  
You're not receiving notifications from this thread.

**5 participants**

GitHub, Inc. [US] | <https://github.com/qati/DeepLearningCourse>

This repository Search Pull requests Issues Marketplace Explore

qati / DeepLearningCourse Watch 2 Star 0 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights

The official repository of Deep learning and machine learning in science (deeplea17em) course @ ELTE.

1 commit 1 branch 0 releases 1 contributor

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

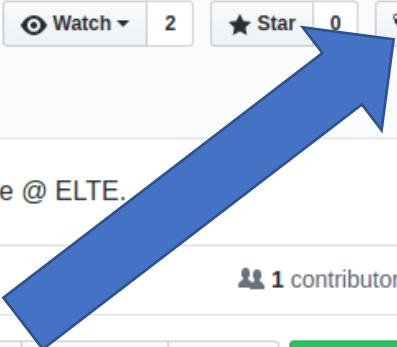
qati Initial commit Latest commit d6e7510 4 days ago

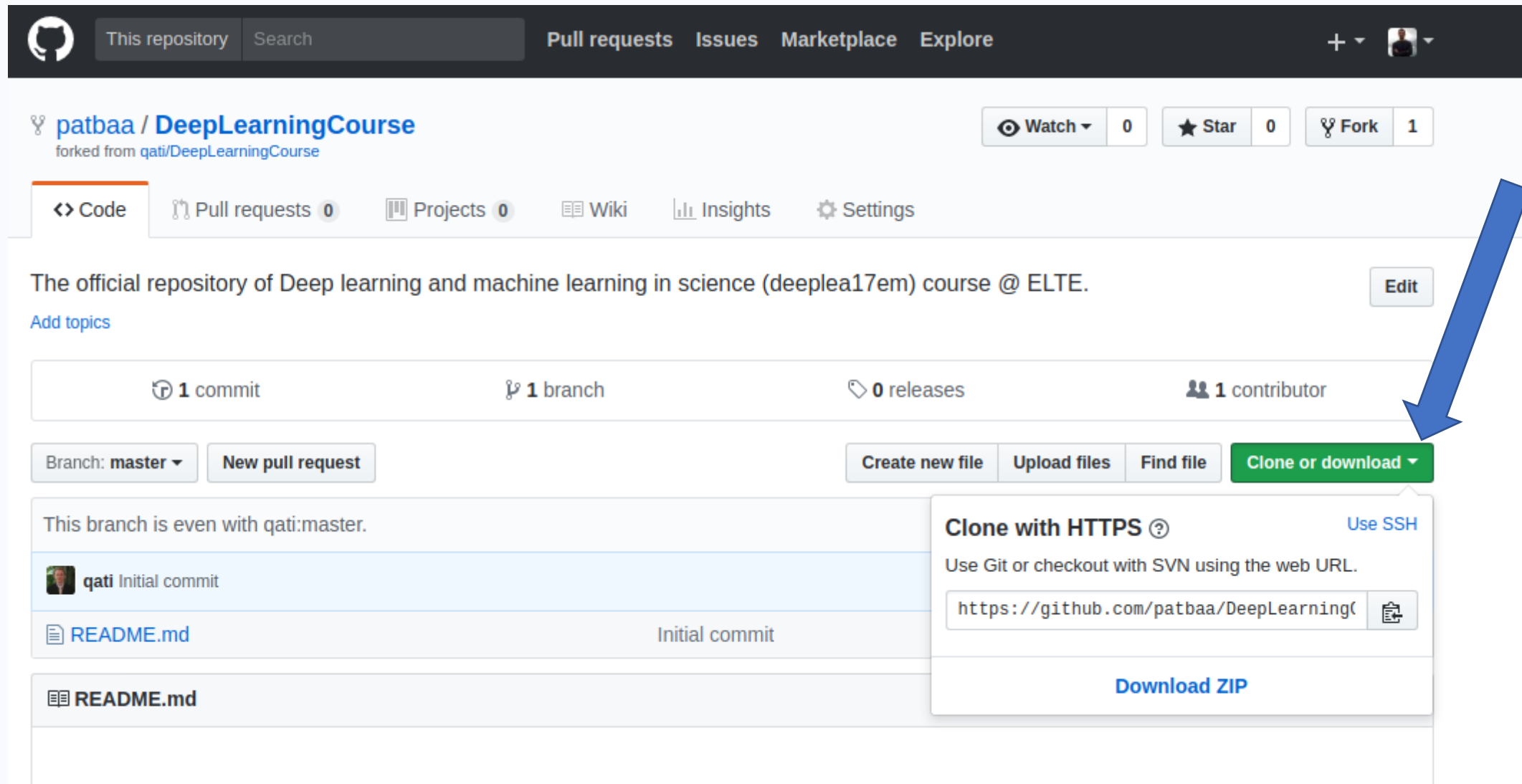
README.md Initial commit 4 days ago

README.md

## DeepLearningCourse

The official repository of Deep learning and machine learning in science (deeplea17em) course @ ELTE.





The screenshot shows the GitHub interface for the repository 'patbaa / DeepLearningCourse', which is a fork of 'qati/DeepLearningCourse'. The repository description is 'The official repository of Deep learning and machine learning in science (deeplea17em) course @ ELTE.' The repository statistics show 1 commit, 1 branch, 0 releases, and 1 contributor. A blue arrow points to the 'Clone or download' button. A dropdown menu is open, showing options to 'Clone with HTTPS' (with a question mark icon), 'Use SSH', and 'Download ZIP'. The HTTPS URL is 'https://github.com/patbaa/DeepLearningCourse/'. The repository also has a 'README.md' file listed.

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.

qati Initial commit

README.md Initial commit

README.md

Clone with HTTPS ? Use SSH

Use Git or checkout with SVN using the web URL.

https://github.com/patbaa/DeepLearningCourse/

Download ZIP

```
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$
```



The screenshot shows the GitHub interface for the repository 'patbaa / DeepLearningCourse'. The repository is a fork of 'qati/DeepLearningCourse'. The repository description is 'The official repository of Deep learning and machine learning in science (deeplea17em) course @ ELTE.' The repository has 2 commits, 1 branch, 0 releases, and 1 contributor. The current branch is 'master', which is 1 commit ahead of 'qati:master'. The commit history shows two commits: 'patbaa just playing around' (latest commit 1aba520 a minute ago) and 'Initial commit' (5 days ago). The file list shows 'README.md' and 'testFile'. A blue arrow points to the 'testFile' entry.

GitHub repository page for **patbaa / DeepLearningCourse**, forked from [qati/DeepLearningCourse](#).

Repository description: The official repository of Deep learning and machine learning in science (deeplea17em) course @ ELTE.

Repository statistics: 2 commits, 1 branch, 0 releases, 1 contributor.

Branch: **master** (1 commit ahead of qati:master). Buttons: New pull request, Create new file, Upload files, Find file, Clone or download.

Commit history:

- patbaa just playing around (Latest commit 1aba520 a minute ago)
- Initial commit (5 days ago)

File list:

- README.md
- testFile (highlighted by a blue arrow)

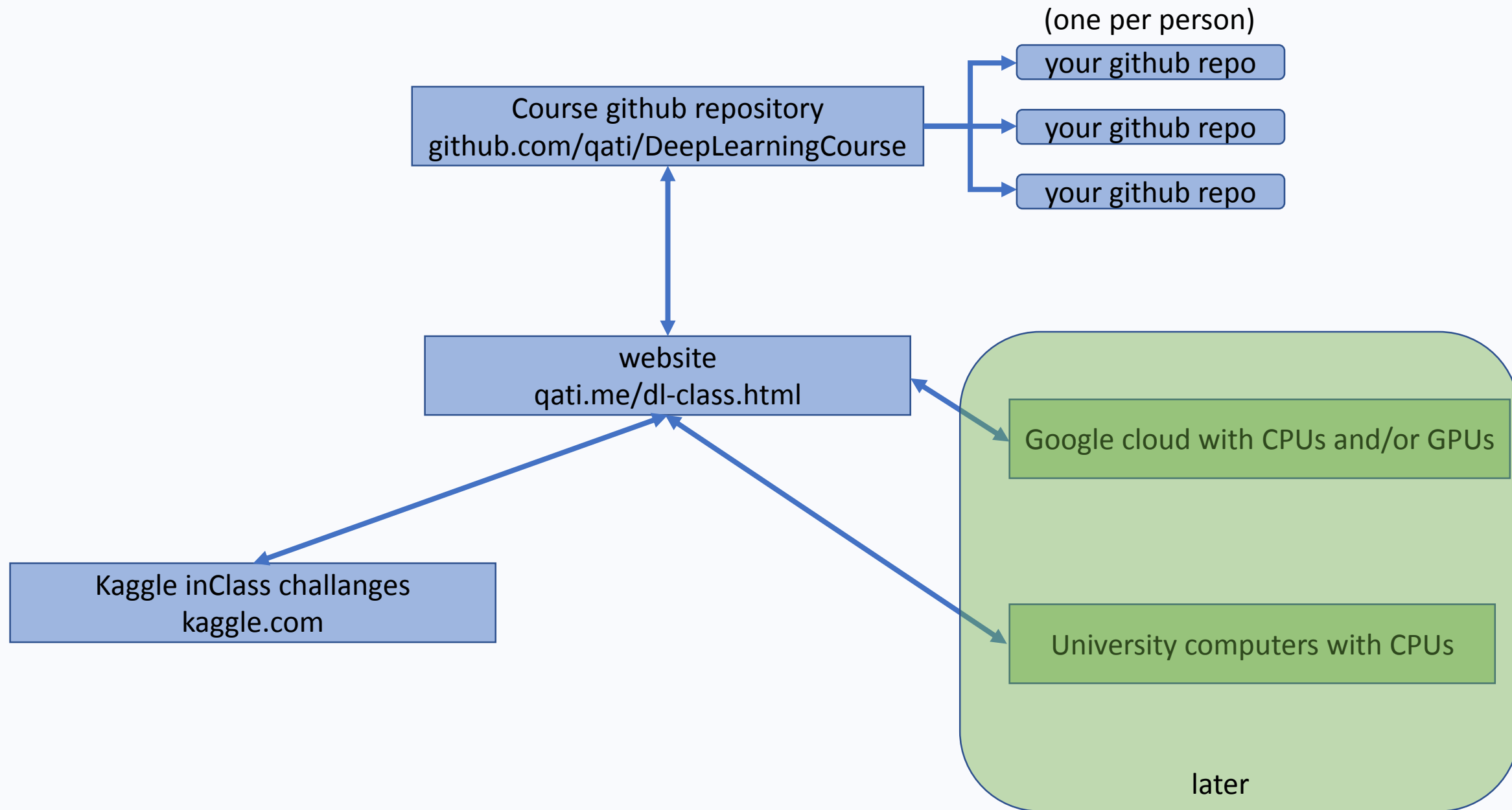
TIP: after registration add your *caesar* email to unlock some premium features

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/>