

データサイエンス におけるPythonの存在感

辻 真吾 (@tsjshg)
みんなのPython勉強会#10

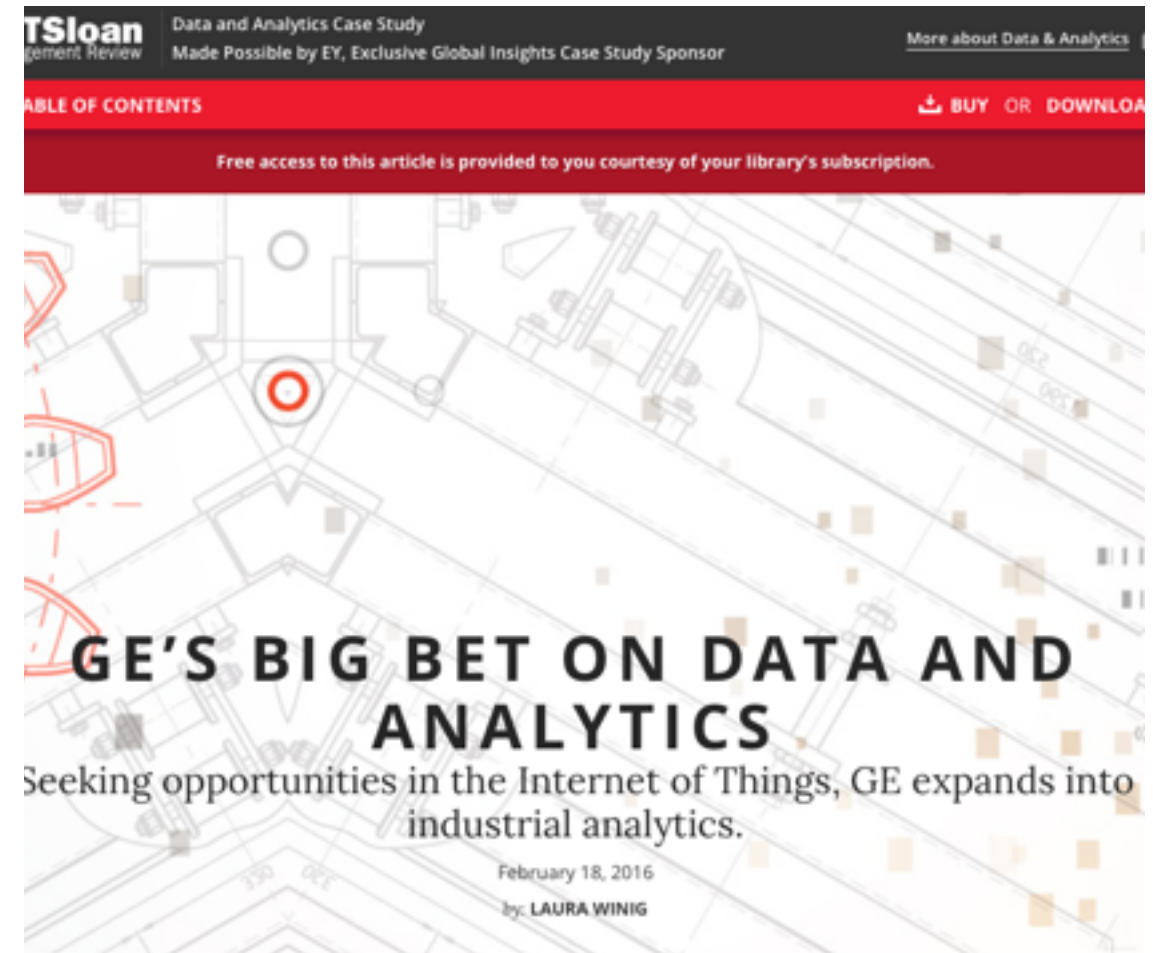
2016/3/8

自己紹介

- ・ 辻 真吾（つじ しんご）1975年生まれ（大厄）
- ・ 東京大学先端科学技術研究センター ゲノムサイエンス分野
 - ・ Bioinformaticsですが、3/31任期切れ予定
- ・ プログラミング経験
 - ・ MSX2でBASIC, Logo（小学生）
 - ・ C, C++（大学生）
 - ・ Java（IT系ベンチャー勤務時代）
 - ・ Python（10年ほど前からほとんどこれ）

ちょっと使い古された感がありますがBig Dataの時代

- ・ MBAで有名なMIT Sloan校のレポート
- ・ 大規模データの収集と解析に力を入れるGE
- ・ IoT（Internet of Things）の典型的な例



has bet big on the Industrial Internet — the convergence of industrial machines, data and the Internet. The company is putting sensors on gas turbines, jet engines, and other machines; connecting them to the cloud; and analyzing the resulting flow of data. The goal: identify ways to improve machine productivity and reliability. This MIT Sloan Management Review case study looks at how this traditional manufacturer is remaking itself into a modern digital business.

人材争奪戦

シリコンバレーでデータサイ
エンティストの仕事を探す

一流企業の求人がずらり

Find Jobs Find Resumes Employers / Post Job

indeed
one search. all jobs.

what: data scientist
job title, keywords or company

where: Silicon Valley, CA
city, state, or zip

data scientist jobs in Silicon Valley, CA Jobs 1 to 10

New! Join Indeed Prime - Get offers from great tech companies

Show: all jobs - 6 new jobs

Data Scientist - new
Intuit - ★★★★★ - 403 reviews - Mountain View, CA 94039
SAS, R, SQL, S-Plus, etc.). Quickly understand patterns within large quantity of data and to reference key characteristics using visualization techniques....
2 hours ago - save job - email - more...

Data Scientist, Analytics
Oculus VR - Menlo Park, CA
The Data Scientist Analytics role has work across the following four areas:. We're looking for data scientists to work on our core and business products at...
1 day ago - save job - email - more...

Data Scientist
Paypal - ★★★★★ - 279 reviews - San Jose, CA
Seeks Data Scientist in San Jose, CA:. Collaborate with architects, engineers, data scientists and PayPal Risk teams to architect and develop strategic and...
23 hours ago - save job - email - more...

Data Scientist, Analytics (Oculus)
Facebook - ★★★★★ - 78 reviews - Menlo Park, CA
Data Scientist, Analytics (Oculus) (Menlo Park, CA) Careers at Oculus A Facebook Company
We're looking for data scientists to work on our core and business...
1 day ago - save job - email - more...

Data Scientist
EBay - ★★★★★ - 627 reviews - San Jose, CA
The ideal candidate is an independent, solution-oriented thinker with a strong background processing huge data sets, applying analytical rigor and statistical...
1 day ago - save job - email - more...

Data Scientist, Consumer Applications
Apple - ★★★★★ - 2,055 reviews - Santa Clara Valley, CA
Solid knowledge in R, Perl, Python, Matlab, SAS, or other tools appropriate for large scale analysis of numerical and textual data....
8 days ago - save job - email - more...

Data Scientist, App Store Engineering R&D
Apple - ★★★★★ - 2,055 reviews - Santa Clara Valley, CA
Deep understanding of Machine Learning and Data Mining techniques. App Store team is looking for talented Data Scientist to work on high visibility project that...
11 hours ago - save job - email - more...

Data Scientist- Symantec
Symantec - ★★★★★ - 378 reviews - Mountain View, CA
Identify and develop high-impact machine learning data science use cases. Collect, wrangle and steward data assets from a variety of internal and external...
8 days ago - save job - email - more...

Data Scientist/Analyst
Cisco Systems, Inc. - ★★★★★ - 1,422 reviews - San Jose, CA

Sort by: relevance - date

Distance:
within 25 miles

Salary Estimate
\$0,000+ (1490)
\$0,000+ (1206)
\$5,000+ (925)
\$10,000+ (619)
\$25,000+ (340)

Job Type
full-time (1596)
internship (68)
contract (71)
temporary (21)
part-time (18)

Location
San Jose, CA (279)
Mountain View, CA (213)
Sunnyvale, CA (174)
San Jose, CA (145)
Redwood City, CA (137)
more »

Company
Microsoft (79)
Apple (72)
UberCoders (58)
Lead Sciences (54)
Google (50)
more »

どんなスキルが必要か？

- Python、Matlab、SAS、Rあたりが求められているらしい
- もちろん、Python推します！
- まず、MatlabやSASとの違いから

Qualifications/Requirements

- Master's Degree in a "STEM" subject (Science, Technology, Engineering, Mathematics) plus proven years analytics development for applications in a commercial/industrial setting OR Ph.D. in a "STEM" subject (Science, Technology, Engineering, Mathematics) plus some year analytics development for applications in a commercial/industrial setting
- Demonstrated skill in the use of analytic software tools or languages (e.g. Python, Matlab, SAS, R)
- SQL experience

Applications from job seekers who require sponsorship to work in the UK are welcome and will be considered alongside all other applications. However, non-EU/EEA candidates may not be appointed to a post if a suitably qualified, experienced and skilled EU/EEA candidate is available to take up the post, as the employing body is unlikely, in these circumstances, to satisfy the Resident Labour Market Test. For further information please visit the UK Border Agency website

<http://www.ukba.homeoffice.gov.uk/visas-immigration/working>

オープンではない

- ・ どちらも老舗
- ・ （ちょっとだけ使ったことありますが）超高性能
- ・ お金の問題？



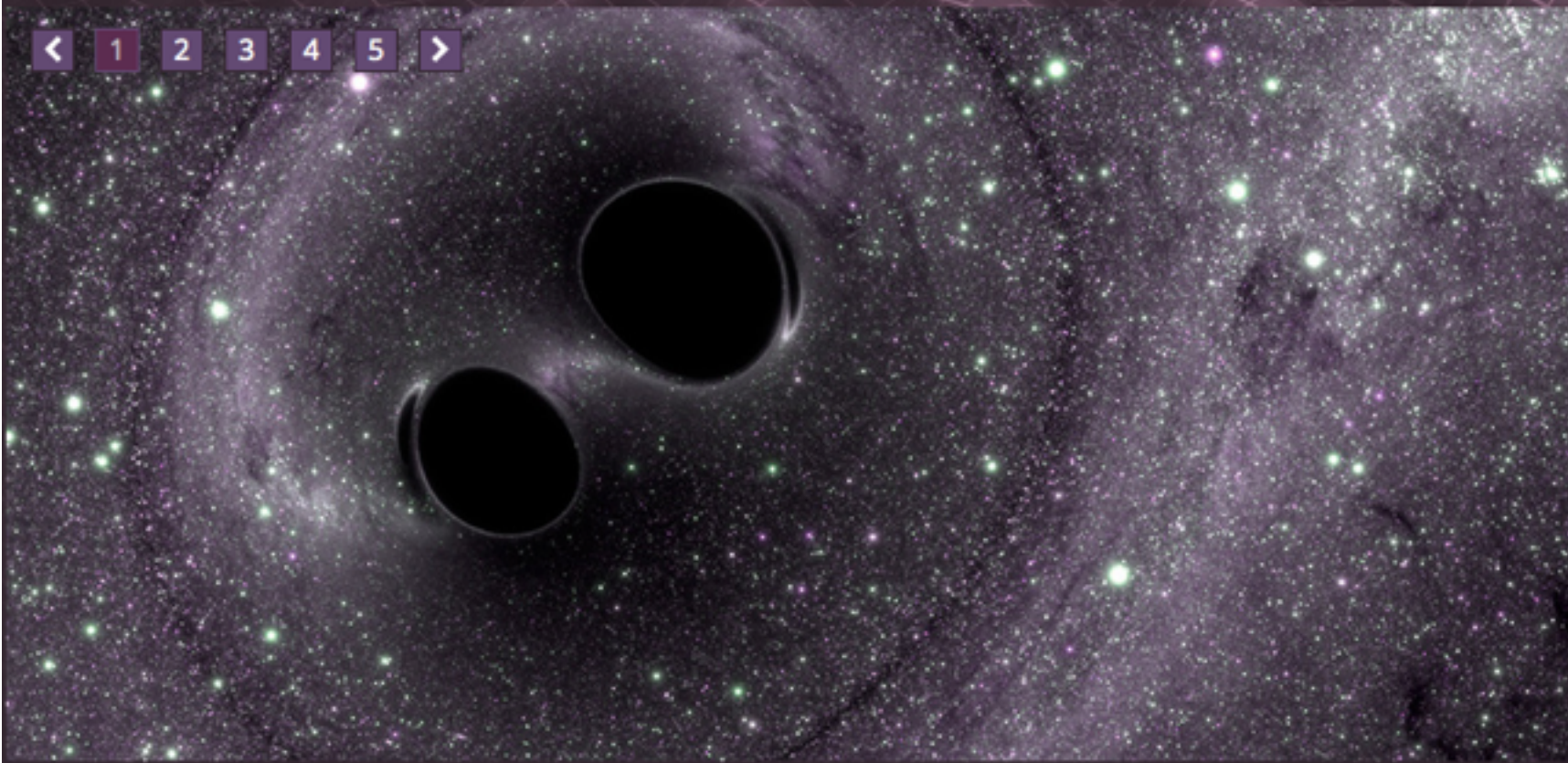


LIGO

Laser Interferometer
Gravitational-Wave Observatory
Supported by the National Science Foundation
Operated by Caltech and MIT

[About](#) [Learn More](#) [News](#) [Gallery](#) [Educational Resources](#) [For Scientists](#) [Study & Work](#)

[<](#) [1](#) [2](#) [3](#) [4](#) [5](#) [>](#)



LIGO Detects Gravitational Waves

Detection Portal • February 11, 2016

On September 14, 2015 at 5:51 a.m. Eastern Daylight Time (09:51 UTC), the twin Laser Interferometer Gravitational-wave Observatory (LIGO) detectors, located in Livingston, Louisiana, and Hanford, Washington, USA both measured ripples in the fabric of spacetime – gravitational waves – arriving at the Earth from a cataclysmic event in the distant universe. The new Advanced LIGO detectors had just been brought into operation for their first observing run when the very clear and strong signal was captured.

重力波を初めて観測

質量を持った物体が動くことで、時空のゆがみが波のように光速で伝わる現象をLIGO（ライゴ：レーザー干渉計重力波天文台）で初観測

オープンである 事の重要性

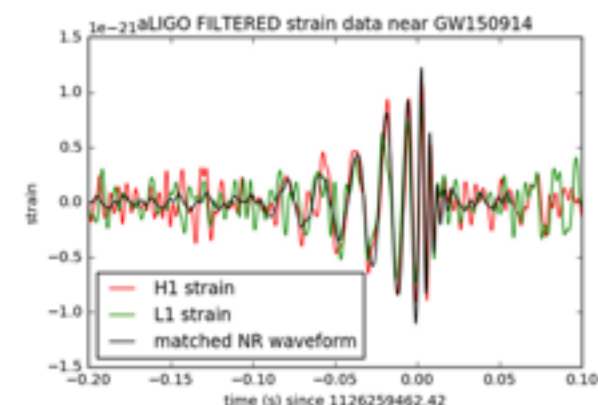
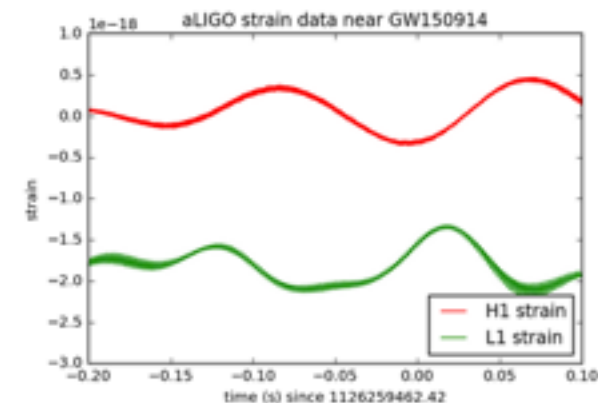
- ・ データはIPython notebookとともに公開されている
- ・ Pythonはフリー。誰でも解析を再実行できる！

```
In [14]: # filter the data:
strain_H1_filt = filter_data(strain_H1, coefs)
strain_L1_filt = filter_data(strain_L1, coefs)

# filter NR template as we do with the data:
NR_H1_filt = filter_data(NR_H1, coefs)

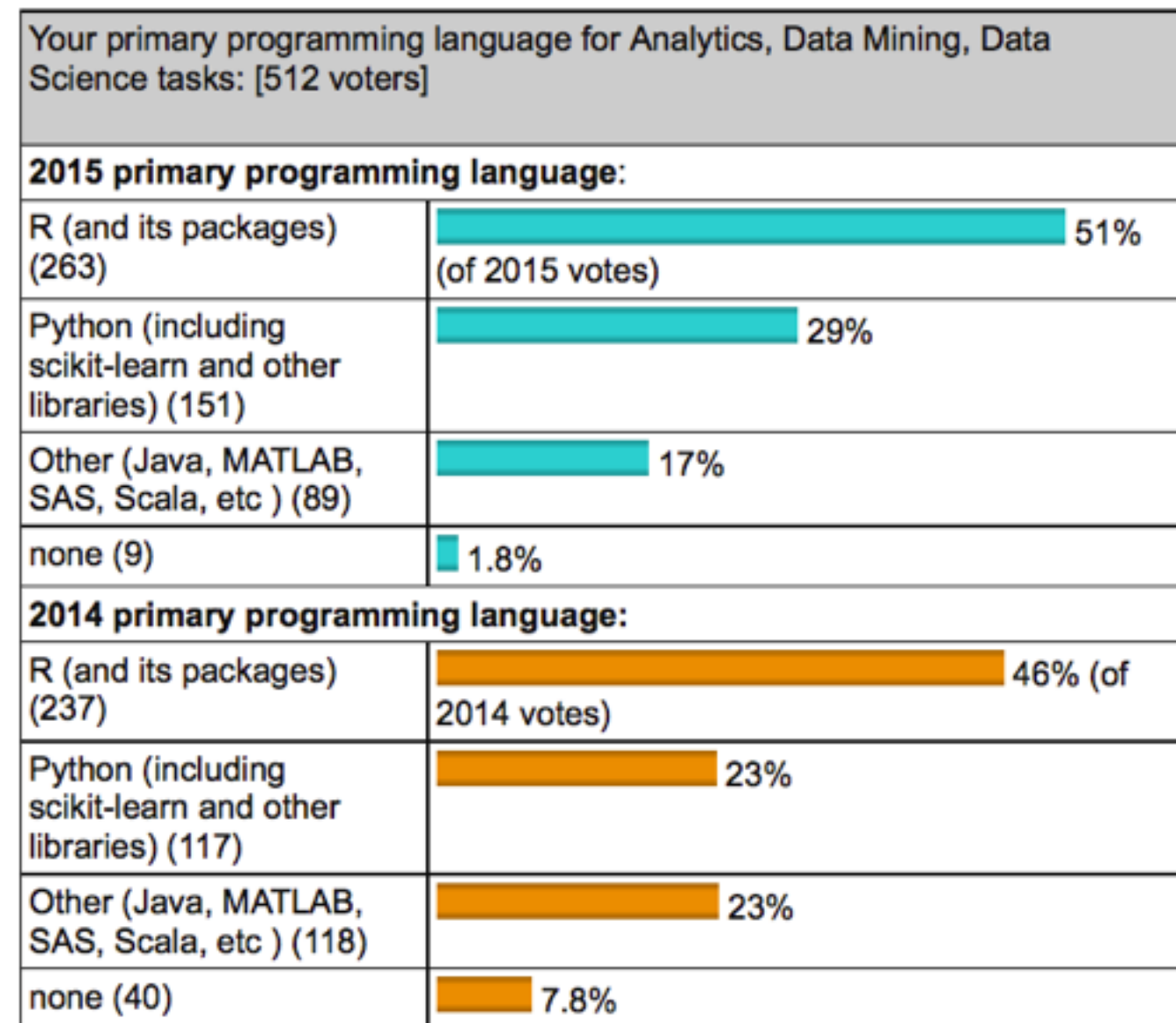
# plot the data prior to filtering:
plt.figure()
plt.plot(time-tevent,strain_H1,'r',label='H1 strain')
plt.plot(time-tevent,strain_L1,'g',label='L1 strain')
plt.xlim([-0.2,0.1])
plt.xlabel('time (s) since '+str(tevent))
plt.ylabel('strain')
plt.legend(loc='lower right')
plt.title('aLIGO strain data near GW150914')
plt.savefig('GW150914_H1_strain_unfiltered.png')

# plot the data after filtering:
# first, shift L1 by 7 ms, and invert. See the GW150914 detection paper for why!
strain_L1_filt = -np.roll(strain_L1_filt,int(0.007*fs))
# We also have to shift the NR template by 2 ms to get it to line up properly with the data
plt.figure()
plt.plot(time-tevent,strain_H1_filt,'r',label='H1 strain')
plt.plot(time-tevent,strain_L1_filt,'g',label='L1 strain')
plt.plot(NRtime+0.002,NR_H1_filt,'k',label='matched NR waveform')
plt.xlim([-0.2,0.1])
plt.ylim([-1.5e-21,1.5e-21])
plt.xlabel('time (s) since '+str(tevent))
plt.ylabel('strain')
plt.legend(loc='lower left')
plt.title('aLIGO FILTERED strain data near GW150914')
plt.savefig('GW150914_H1_strain_filtered.png')
```



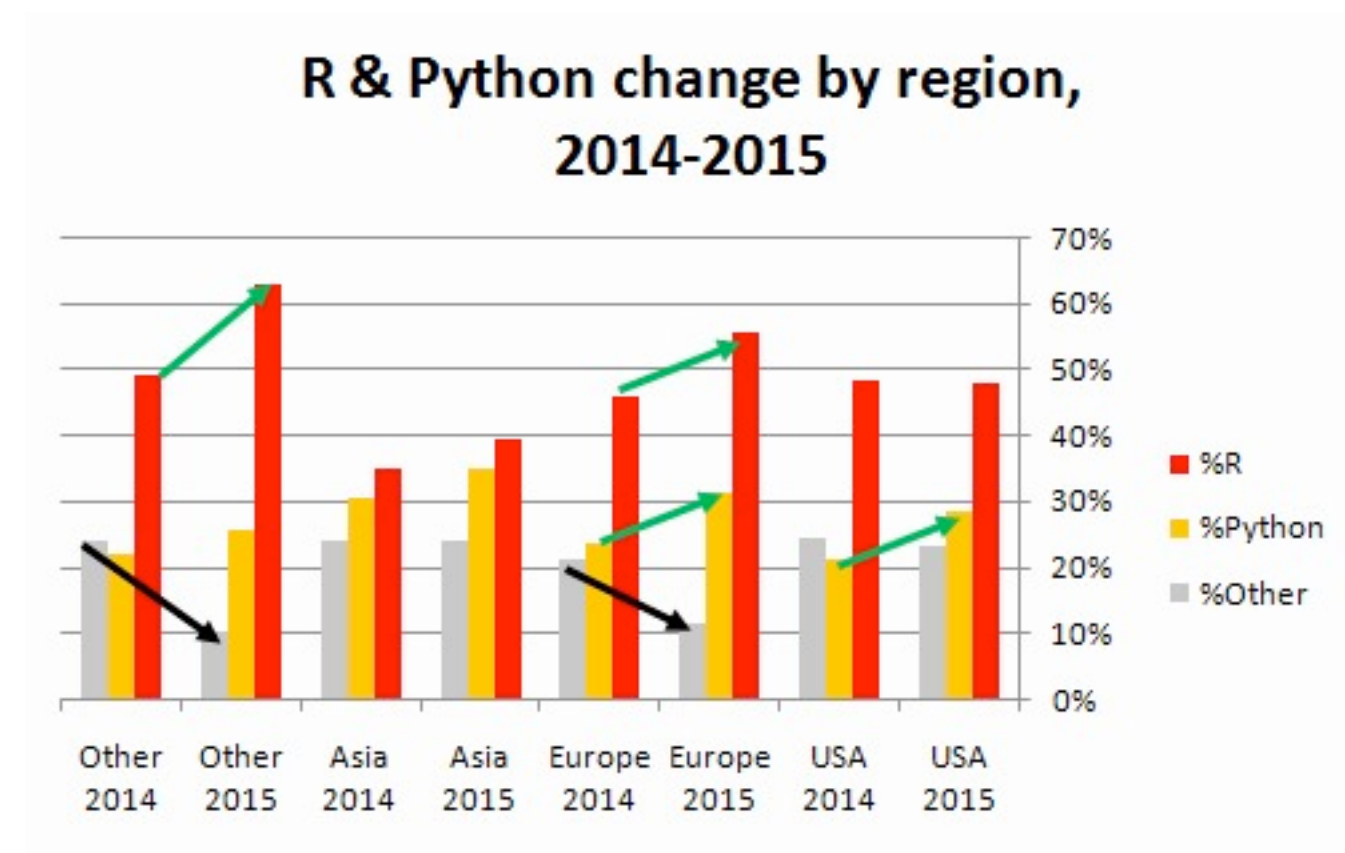
Rもオープン

- ・ KD nuggets調べ「データ解析で一番使っている言語は？」
- ・ Rが強い
- ・ でもPython伸びてる
- ・ さらに詳しく見てみると



地域別に見ると

- Pythonはどの地域でも伸びている
- データサイエンス先進国のUSAでRの伸びが鈍化！
- Pythonは汎用言語
 - 利用シーンはデータ解析だけじゃない



まとめ

	オープン	汎用
Python	○	○
Matlab	×	×
SAS	×	×
R	○	×

もうPythonでデータサイエンスやるしかない！
そんなあなたに・・・

urses 

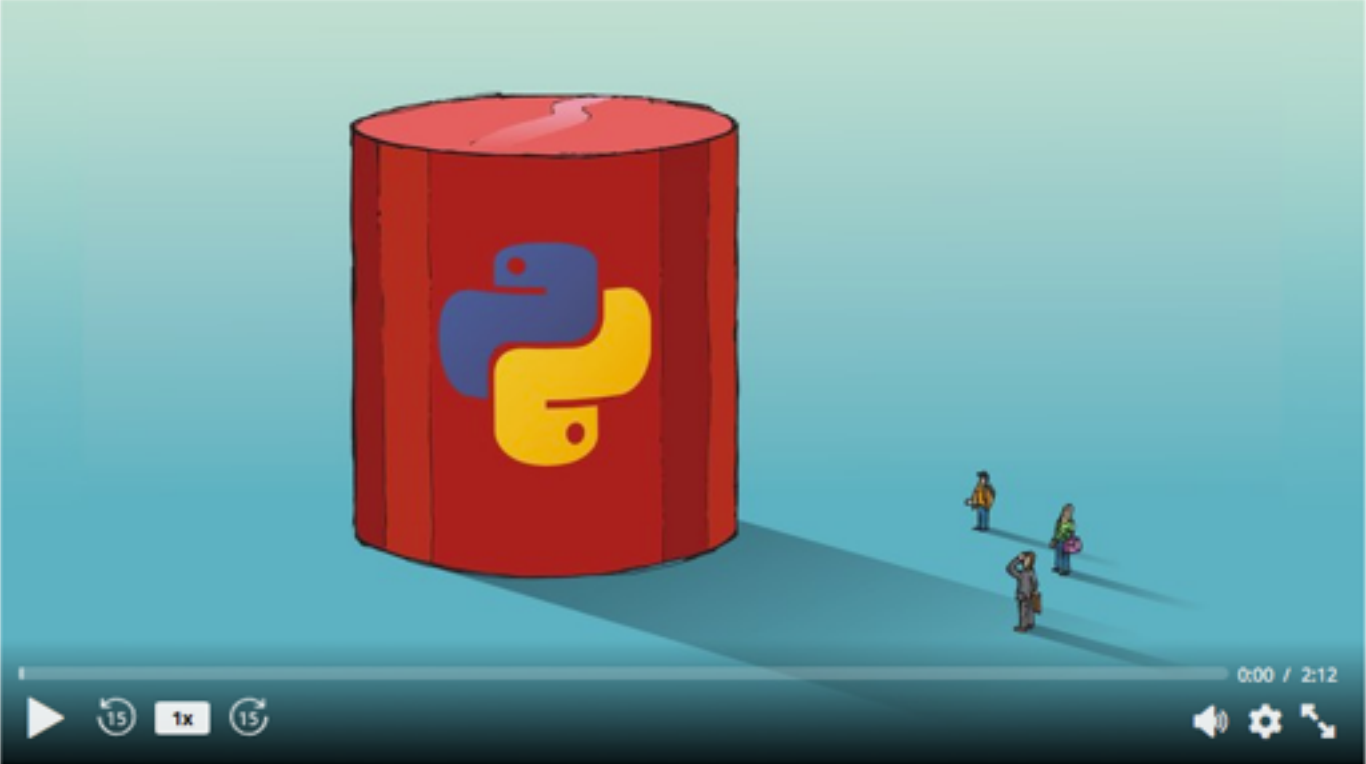
udemy

【世界で2万人が受講】 実践 Python データサイエンス

データ解析の基本、可視化、統計、機械学習などデータサイエンスに関するあらゆる実践的なスキルがPythonで身に付く！

★★★★★ 23 ratings, 323 students enrolled

Instructed by [Shingo Tsuji](#), [Jose Portilla](#) Development / Programming Languages



24,280 ¥

[Take This Course](#)

[Redeem a Coupon](#)

[Start Free Preview](#)

[More Options ▾](#)

Lectures	103
Video	17 Hours
Skill Level	All Levels
Languages	Japanese
Includes	Lifetime access 30 day money back guarantee! Available on iOS and Android Certificate of Completion

 [Wishlist](#)

いまだけ\$45

<http://tanopy.blog79.fc2.com/blog-entry-141.html>

お後がよろしいようで