












10th December 2018

# Joint **Christmas** **special**

Frank Kelly, Rob Charwood and Miquel Perello-Nieto

---

# Christmas Special

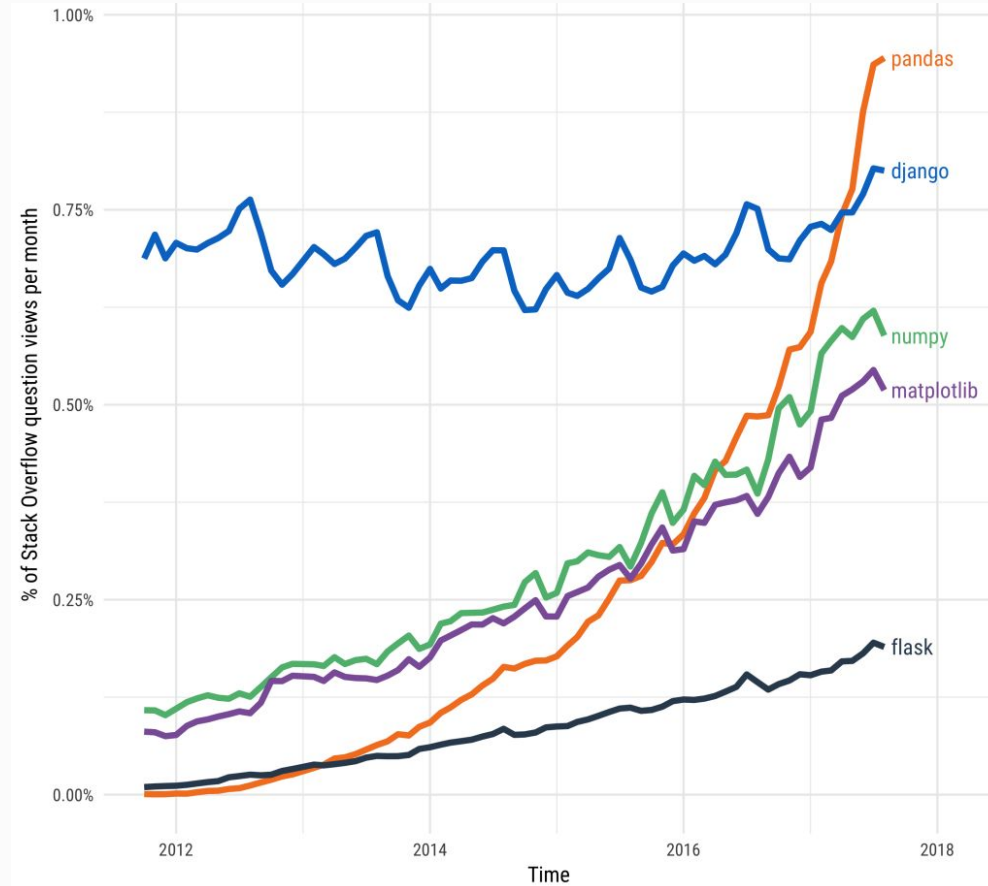
6:30	 Intro to the event	Meetup Organizers
6:40	 Santas Routing Problem - Observations on a Kaggle Competition	Bharat Kunwar
6:50	 1 day with Docker - What I learned in my first 8 hours	Rob Charlwood
7:00	  Break	
7:30	 Community announce	Attendants
7:40	 Trading using Reinforcement Learning	Matt Hamilton
8:00	  Networking and drinks	All

# Data Science vs Web Development



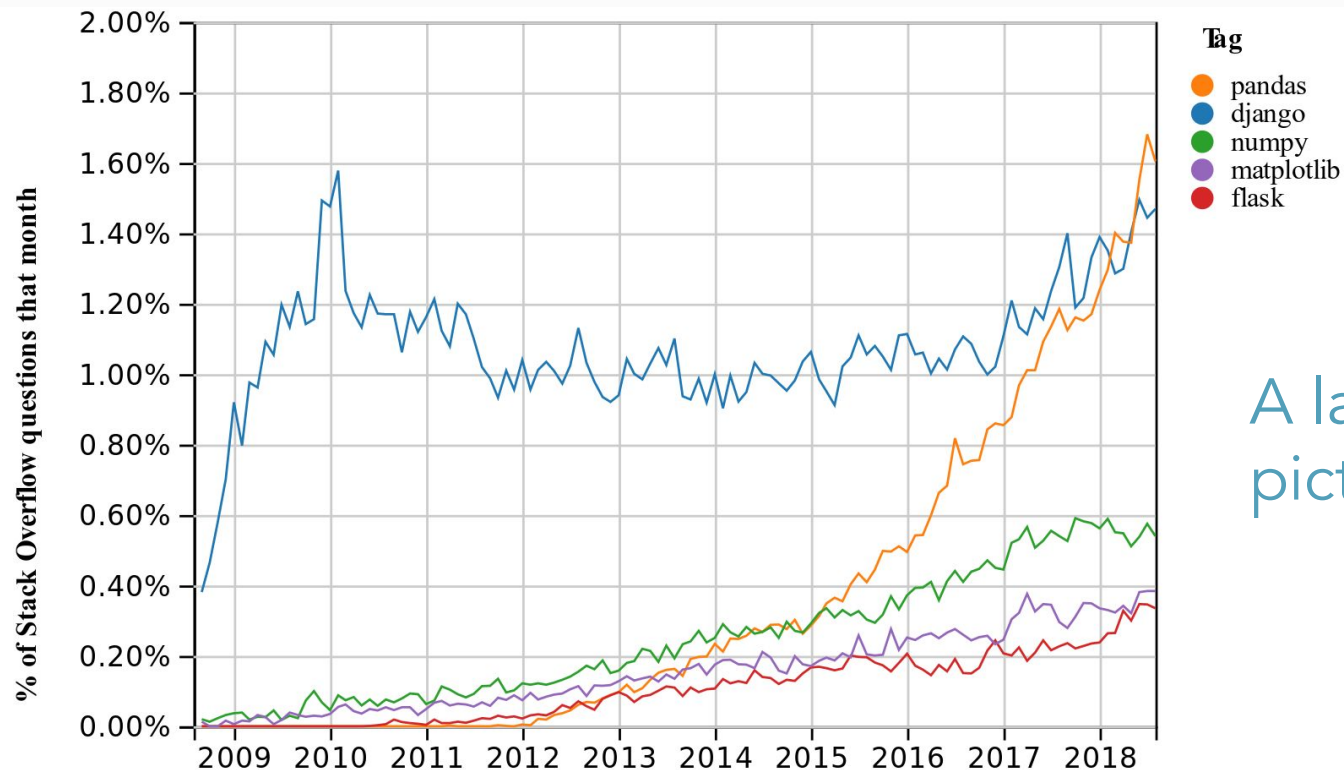
# Stack Overflow Traffic to Questions About Selected Python Packages

Based on visits to Stack Overflow questions from World Bank high-income countries



# Stack Overflow Traffic to Questions About Selected Python Packages

Based on visits to Stack Overflow questions from World Bank high-income countries



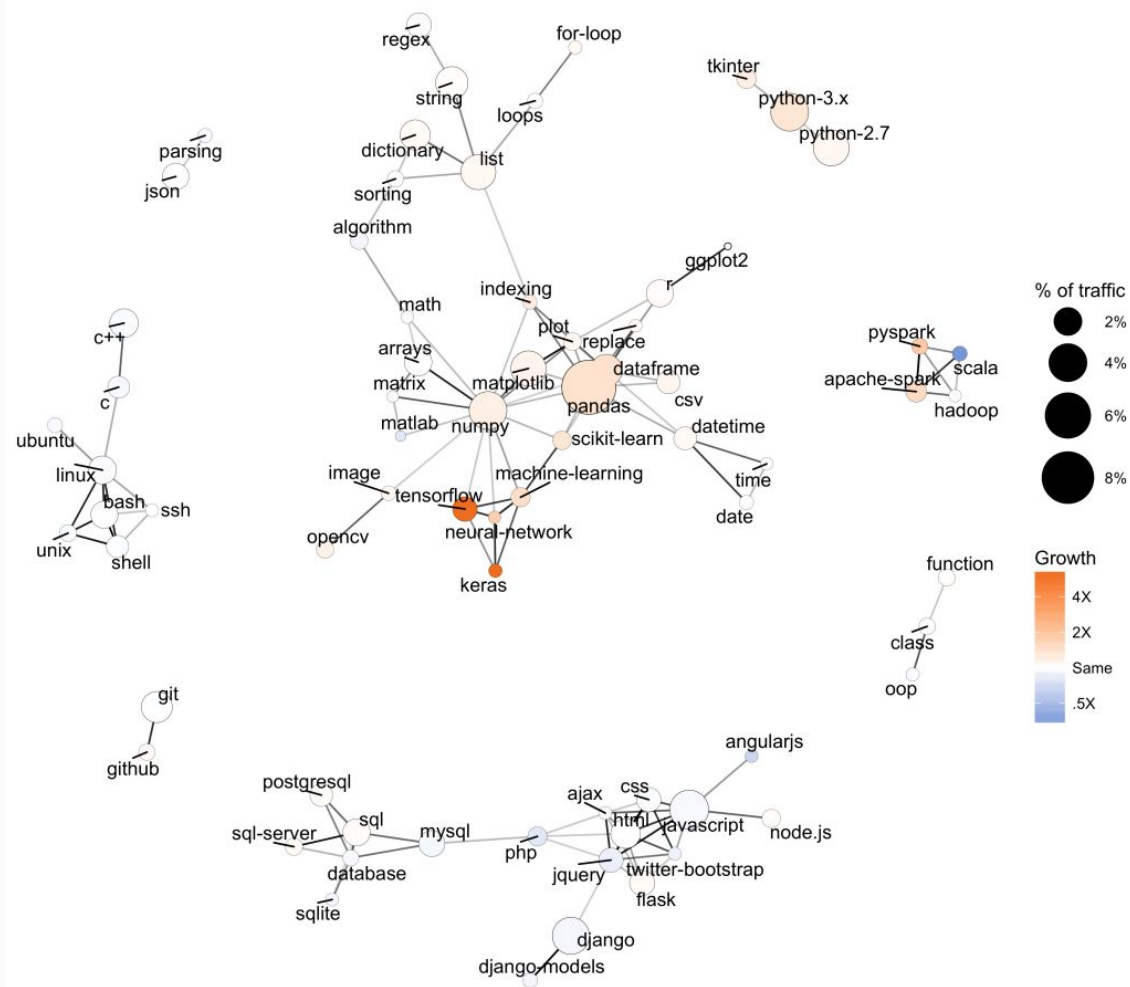
A larger picture

“Python’s popularity in data science and machine learning is probably the main reason for its fast growth”

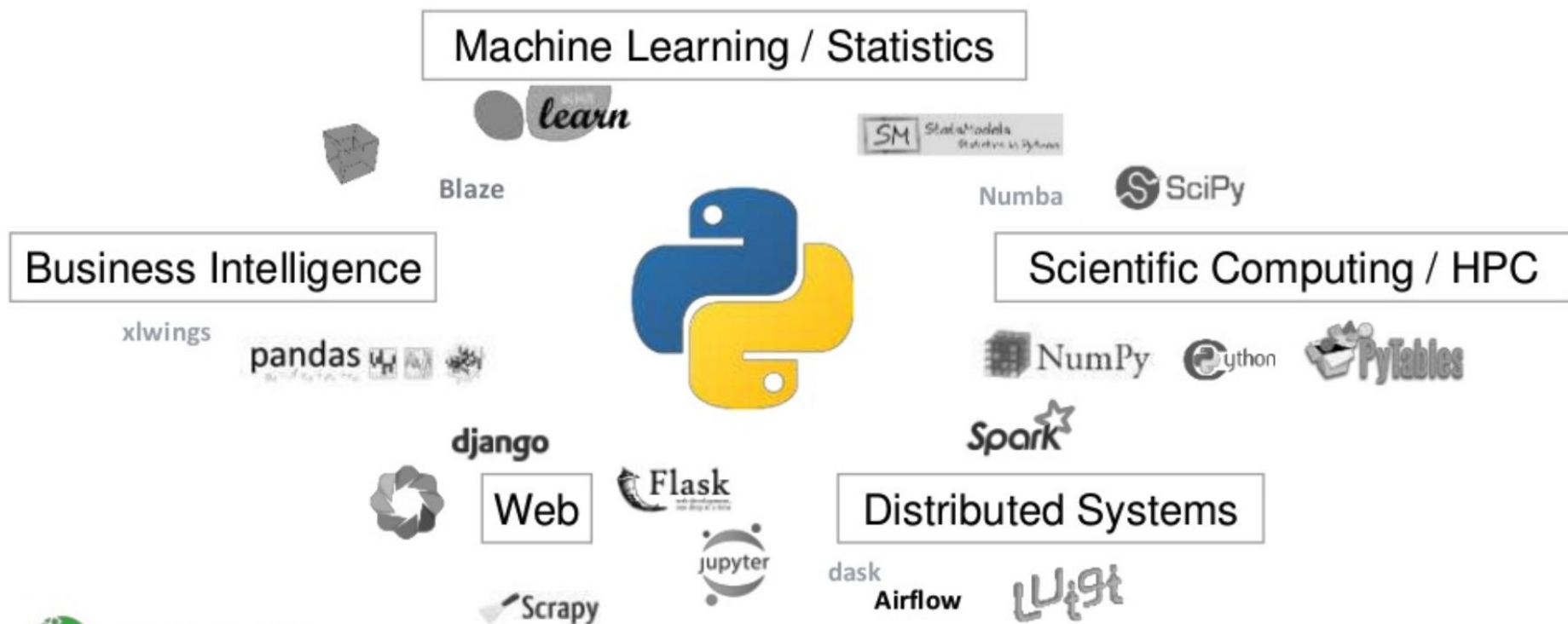
<https://stackoverflow.blog/2017/09/14/python-growing-quickly/>

### Network of Correlated Tags Visited by Python Visitors

Color represents 2017/2016 growth rate of Stack Overflow question views among Python visitors: whether the tag is becoming more (orange) or less (blue) frequently visited alongside Python.



# Python is the common language



# Upcoming events

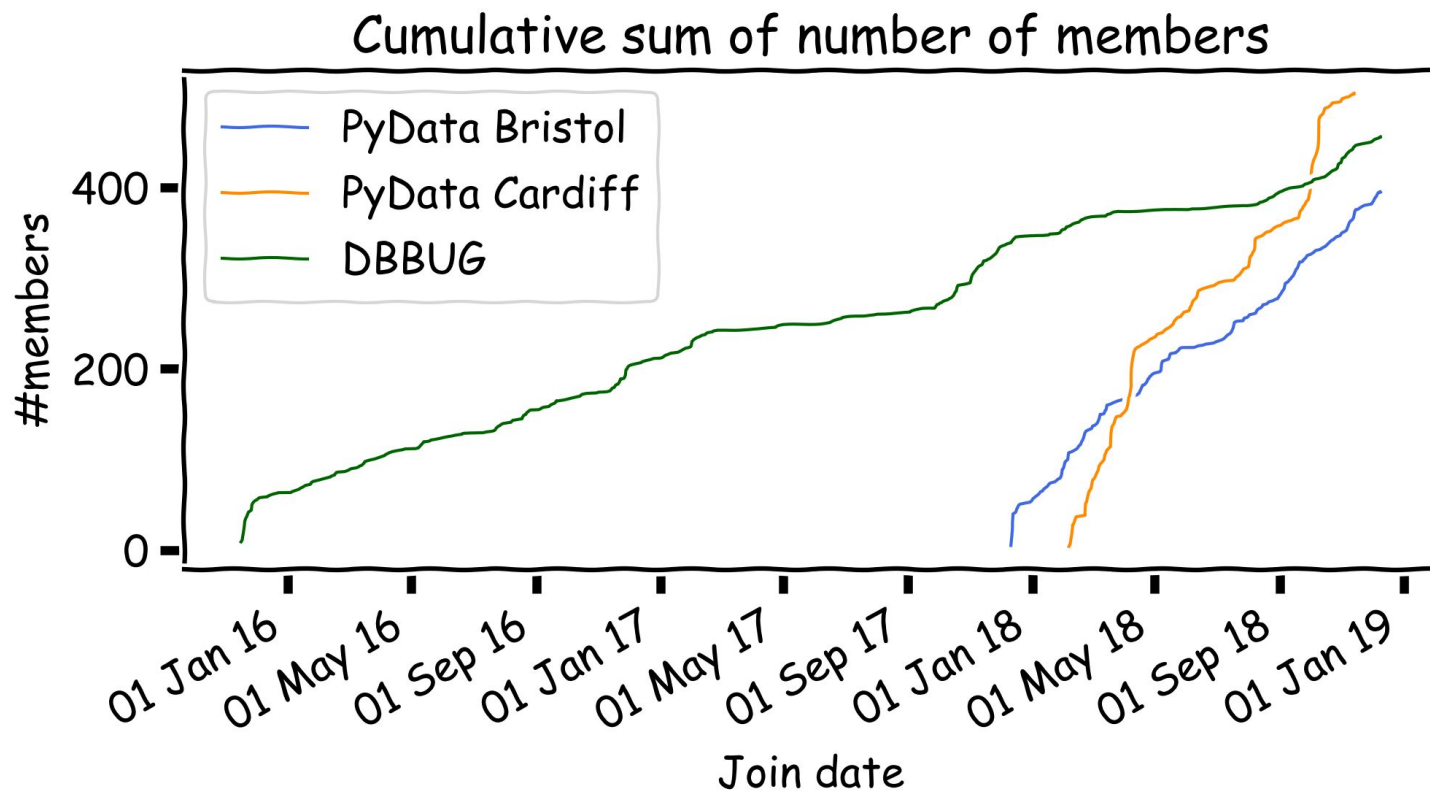
#PyData Cardiff *Meetup* (11 December, Tomorrow!)

#South West Data Social *Meetup* (11 December in Bristol, Tomorrow!)

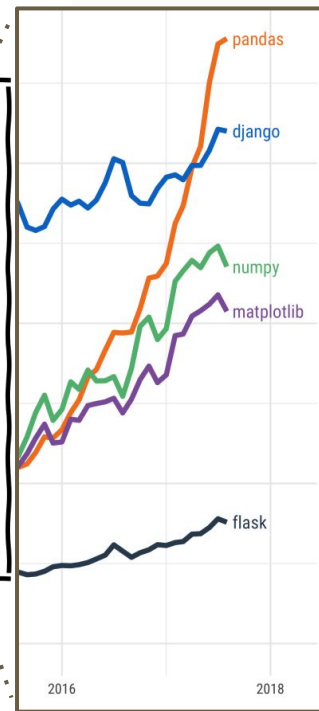
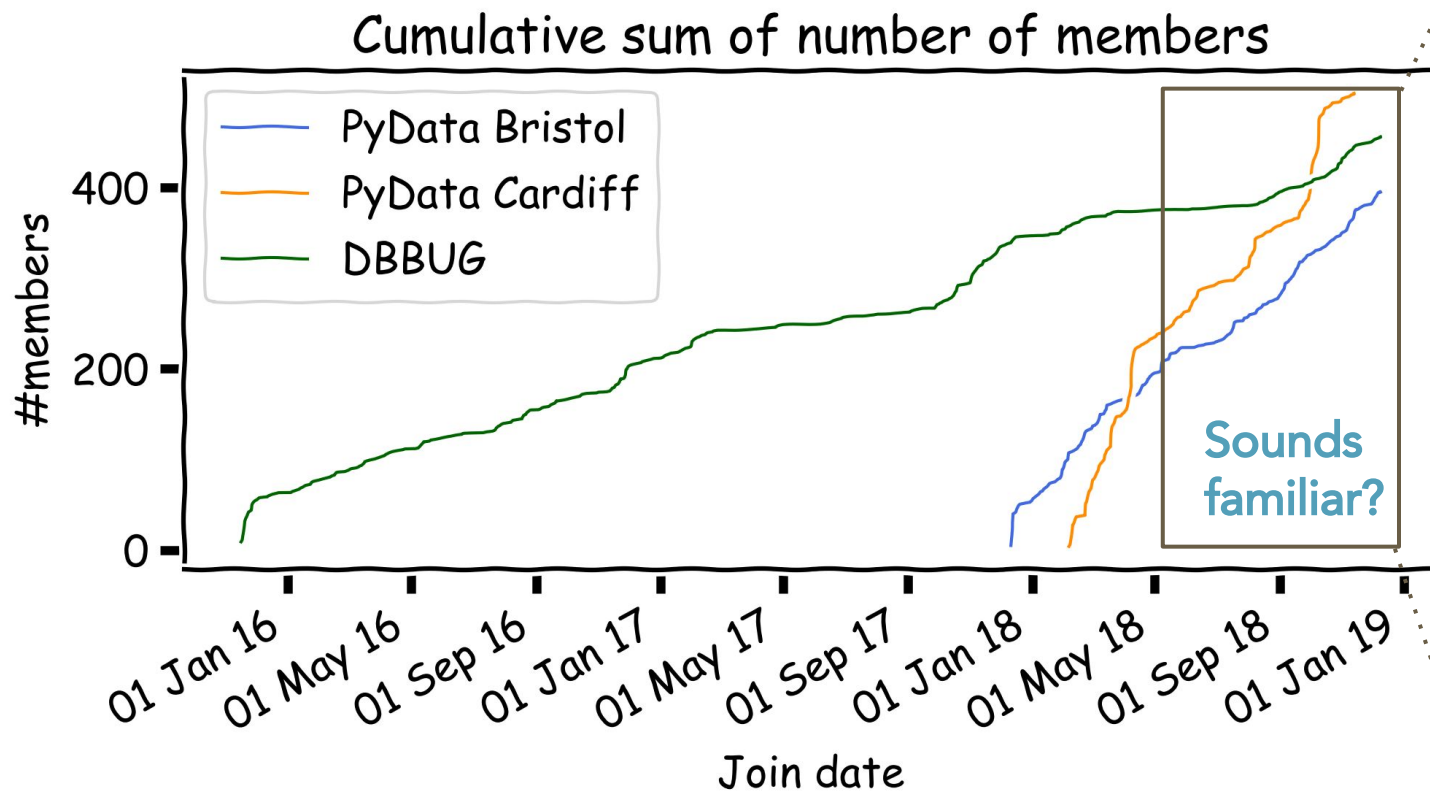
#PyData Bristol *Meetup* (17 January 2019)



# PyData Bristol/Cardiff, DBBUG members



# PyData Bristol/Cardiff, DBBUG members





# The Xmas package: pyjanitor

Tools for cleaning Pandas DataFrames

Implements method chaining

```
import pandas as pd
```

```
df = pd.DataFrame(...) # create a pandas DataFrame
                           somehow.
del df['column1'] # delete a column from the dataframe.
df = df.dropna(subset=['column2', 'column3']) # drop
rows that have empty values in column 2 and 3.
df = df.rename({'column2': 'unicorns', 'column3':
'dragons'}) # rename column2 and column3
df['new_column'] = ['iterable', 'of', 'items'] # add a
new column.
df.reset_index(inplace=True, drop=True) # reset index to
account for the missing row we removed above
```

```
pip install pyjanitor
```



## The pyjanitor approach

```
import pandas as pd
import pandas_flavor as pf

df = (
    pd.DataFrame(...)
    .remove_columns(['column1'])
    .dropna(subset=['column2', 'column3'])
    .rename_column('column2', 'unicorns')
    .rename_column('column3', 'dragons')
    .add_column('new_column', ['iterable', 'of',
'items'])
    .reset_index(inplace(drop=True))
)
```

<https://github.com/ericmjl/pyjanitor>

# Thanks to our sponsors



Elucidata

A series of vertical bars of varying heights and widths, some solid black and some white, arranged horizontally below the word "Elucidata".



- Margriet Groenendijk



NUMFOCUS

OPEN CODE = BETTER SCIENCE

The logo for NUMFOCUS, with "NUM" in red and "FOCUS" in blue. The letter "O" in "FOCUS" is replaced by a blue circle containing a red bracket "[ ]". Below the main text is the tagline "OPEN CODE = BETTER SCIENCE" in blue.

Do you want to be a sponsor?  
Come and talk to us.