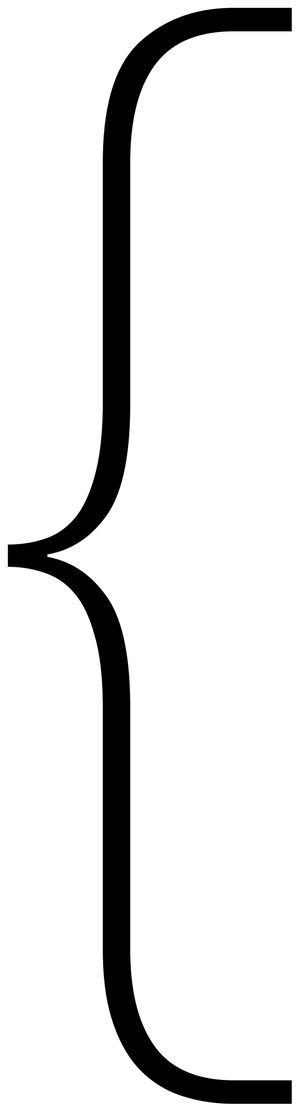
Learning about Developer Culture using

Hacker News

Gabriel Ruiz
GOOD DATA.

External Links



Hacker News new | comments | show | ask | jobs | submit login

1. Don't use Hadoop when your data isn't that big (2013) (chrisstucchio.com)
   217 points by tosh 3 hours ago | hide | 148 comments
2. The Blockstack Browser: A Gateway to a New, Decentralized Internet (blockstack.org)
   102 points by adunk 1 hour ago | hide | 28 comments
3. Chaos Computer Clubs Breaks Iris Recognition System of the Samsung Galaxy S8 (ccc.de)
   459 points by morsch 7 hours ago | hide | 117 comments
4. At The Dawn Of Recorded Sound, No One Cared (npr.org)
   23 points by collapse 52 minutes ago | hide | 4 comments
5. How jeans conquered the world (2012) (bbc.com)
   23 points by prmph 54 minutes ago | hide | 2 comments
6. I've become worse at programming (fascinatedbox.blogspot.com)
   189 points by kumaranvpl 5 hours ago | hide | 54 comments
7. New Surface Pro (microsoft.com)
   190 points by pierre-renaux 5 hours ago | hide | 315 comments
8. BuildZoom (YC W13) is hiring a VP of Sales (lever.co)
   16 minutes ago | hide
9. 22-Year-Old Lidar Whiz Claims Breakthrough (ieee.org)
   9 points by deepnotderp 30 minutes ago | hide | discuss
10. Show HN: 15-question programming quiz with answers (triplebyte.com)
    21 points by ammon 49 minutes ago | hide | 9 comments
11. Stack Overflow: Helping One Million Developers Exit Vim (stackoverflow.blog)
    9 points by var_explained 9 minutes ago | hide | discuss
12. How to build your own particle detector (2015) (symmetrymagazine.org)
    15 points by gus_massa 1 hour ago | hide | 1 comment
13. Choosing The Right Freelancing Niche [audio] (simplecast.com)
    44 points by chris_hawk 3 hours ago | hide | 5 comments
14. Sand and gravel mining "greatly exceeds natural renewal rates" (newyorker.com)
    8 points by sergeant3 1 hour ago | hide | 2 comments
15. Kickstarter set to launch in Japan, where hardware startups are finding it hard (techinasia.com)
    6 points by williswee 1 hour ago | hide | discuss
16. Why Did a Chinese Peroxide Company Pay $1B for a Talking Cat? (bloomberg.com)
    4 points by lumisota 37 minutes ago | hide | discuss
17. Jamboard is now available (blog.google)
    12 points by happy-go-lucky 1 hour ago | hide | 3 comments
18. A Man Who Made the Mistake of Trying to Help Wikileaks (vice.com)
    4 points by dsr12 45 minutes ago | hide | discuss
19. Chasing the Harvest: 'If You Want to Die, Stay at the Ranch' (longreads.com)
    5 points by DiabloD3 1 hour ago | hide | discuss
20. Zuckerberg-Backed Data Trove Exposes the Injustices of Criminal Justice (wired.com)
    74 points by sprucely 2 hours ago | hide | 17 comments
21. Arq 5.8.5 for Mac Fixes a Bad Bug (arqbackup.com)
    107 points by ivank 6 hours ago | hide | 30 comments

+

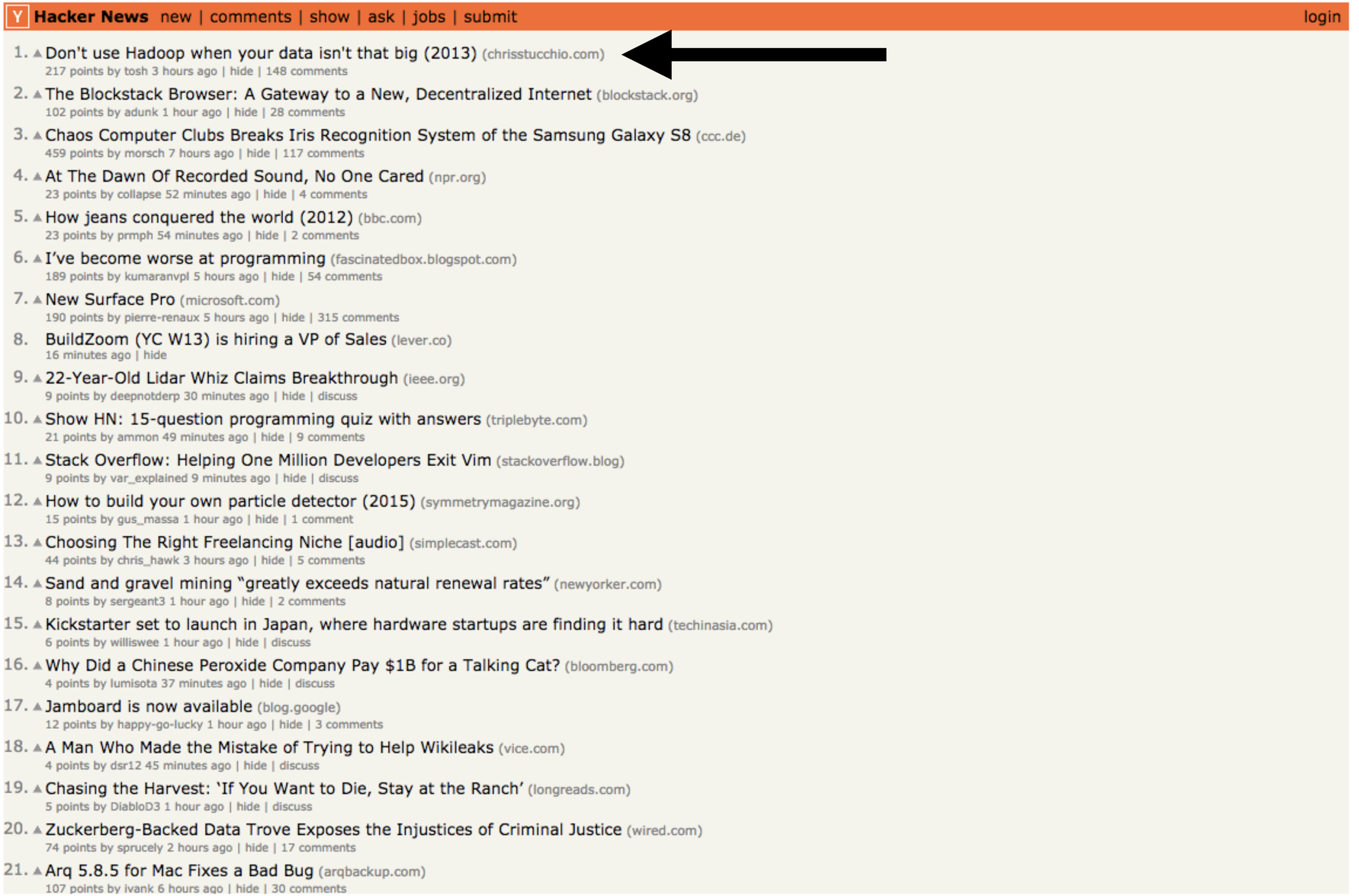
Entrepreneurship **Computer Science**

The Land of Opportunity

Capture Markets

Add to Developer Toolkits

What kind of topics are on Hacker News?


Hacker News new | comments | show | ask | jobs | submit
login
1. Don't use Hadoop when your data isn't that big (2013) (chrisstucchio.com)
217 points by tosh 3 hours ago | hide | 148 comments
2. The Blockstack Browser: A Gateway to a New, Decentralized Internet (blockstack.org)
102 points by adunk 1 hour ago | hide | 28 comments
3. Chaos Computer Clubs Breaks Iris Recognition System of the Samsung Galaxy S8 (ccc.de)
459 points by morsch 7 hours ago | hide | 117 comments
4. At The Dawn Of Recorded Sound, No One Cared (npr.org)
23 points by collapse 52 minutes ago | hide | 4 comments
5. How jeans conquered the world (2012) (bbc.com)
23 points by prmph 54 minutes ago | hide | 2 comments
6. I've become worse at programming (fascinatedbox.blogspot.com)
189 points by kumaranvpl 5 hours ago | hide | 54 comments
7. New Surface Pro (microsoft.com)
190 points by pierre-renaux 5 hours ago | hide | 315 comments
8. BuildZoom (YC W13) is hiring a VP of Sales (lever.co)
16 minutes ago | hide
9. 22-Year-Old Lidar Whiz Claims Breakthrough (ieee.org)
9 points by deepnotderp 30 minutes ago | hide | discuss
10. Show HN: 15-question programming quiz with answers (triplebyte.com)
21 points by ammon 49 minutes ago | hide | 9 comments
11. Stack Overflow: Helping One Million Developers Exit Vim (stackoverflow.blog)
9 points by var_explained 9 minutes ago | hide | discuss
12. How to build your own particle detector (2015) (symmetrymagazine.org)
15 points by gus_massa 1 hour ago | hide | 1 comment
13. Choosing The Right Freelancing Niche [audio] (simplecast.com)
44 points by chris_hawk 3 hours ago | hide | 5 comments
14. Sand and gravel mining "greatly exceeds natural renewal rates" (newyorker.com)
8 points by sergeant3 1 hour ago | hide | 2 comments
15. Kickstarter set to launch in Japan, where hardware startups are finding it hard (techinasia.com)
6 points by williswee 1 hour ago | hide | discuss
16. Why Did a Chinese Peroxide Company Pay $1B for a Talking Cat? (bloomberg.com)
4 points by lumisota 37 minutes ago | hide | discuss
17. Jamboard is now available (blog.google)
12 points by happy-go-lucky 1 hour ago | hide | 3 comments
18. A Man Who Made the Mistake of Trying to Help Wikileaks (vice.com)
4 points by dsr12 45 minutes ago | hide | discuss
19. Chasing the Harvest: 'If You Want to Die, Stay at the Ranch' (longreads.com)
5 points by DiabloD3 1 hour ago | hide | discuss
20. Zuckerberg-Backed Data Trove Exposes the Injustices of Criminal Justice (wired.com)
74 points by sprucely 2 hours ago | hide | 17 comments
21. Arq 5.8.5 for Mac Fixes a Bad Bug (arqbackup.com)
107 points by ivank 6 hours ago | hide | 30 comments

Hacker News new | comments | show | ask | jobs | submit login

▲ Don't use Hadoop when your data isn't that big (2013) (chrisstucchio.com)

221 points by tosh 3 hours ago | hide | past | web | 152 comments | favorite

add comment

▲ lettergram 1 hour ago [-]

I have seen so so so many projects get bogged down by the need to use a "big data" stack.

I think my favorite example, was a team that spent six months trying to build a system to take in files, parse them, and store them. Files came through a little less than one per second, which translated to about 100kb. This translated to about 2.5Gb a day of data. The data only needed to be stored for a year, and could easily be compressed.

They felt the need to setup a cluster, with 1Tb of RAM to handle processing the documents, they had a 25 Kafka instances, etc. It was just insane.

I just pulled out a python script, and combine that with Postgres and within an afternoon I had completed the project (albeit not production ready). This is so typical within companies it makes me gag. They were easily spending $100k a month just on infrastructure, my solution cost ~$400 ($1200 with replication).

The sad part, is that convincing management to use my solution was the hardest part. Basically, I had to explain how my system was more robust, faster, cheaper, etc. Even side-by-side comparisons didn't seem to convince them, they just felt the other solution was better some how... Eventually, I convinced them, after about a month of debates, and an endless stream of proof.

reply

▲ quantumhobbit 1 hour ago [-]

The problem you had was that you were trying to convince management to act in the best interest of the company rather than their own best interest. They would much rather be in charge of a $100k a month project instead of a $1k a month project. Also the big solution required a bunch of engineers working full time which puts them higher up the ladder compared to your 1 engineer working part time solution.

You were saving the company money but hurting their resumes.

reply

▲ nostrademons 49 minutes ago [-]

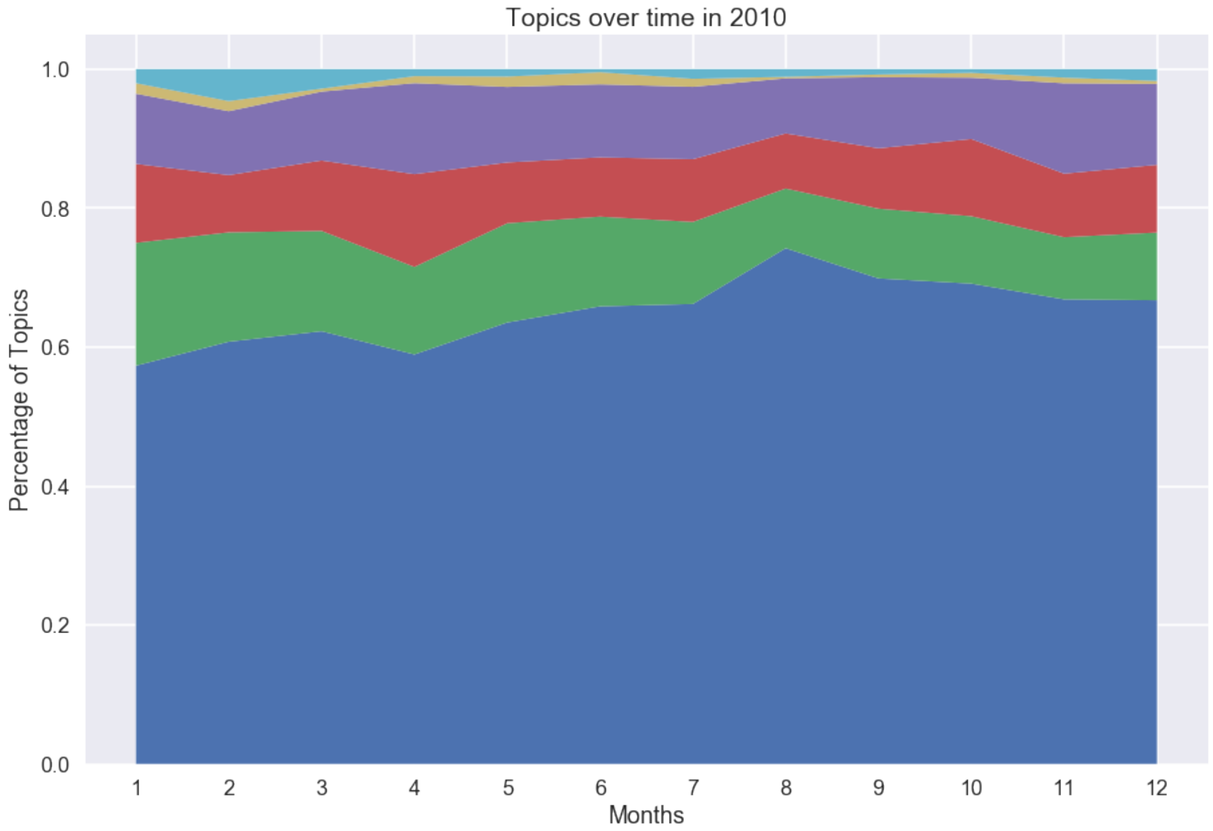
The other point that's commonly missed is that Hadoop is really only useful when both inputs *and* outputs are too large to fit on one machine. If you have a big-data input but a small-data output (which is very common in a lot of exploratory problems), you can get away with a simple work-queue setup that sends results into a shared RDBMS or filesystem.

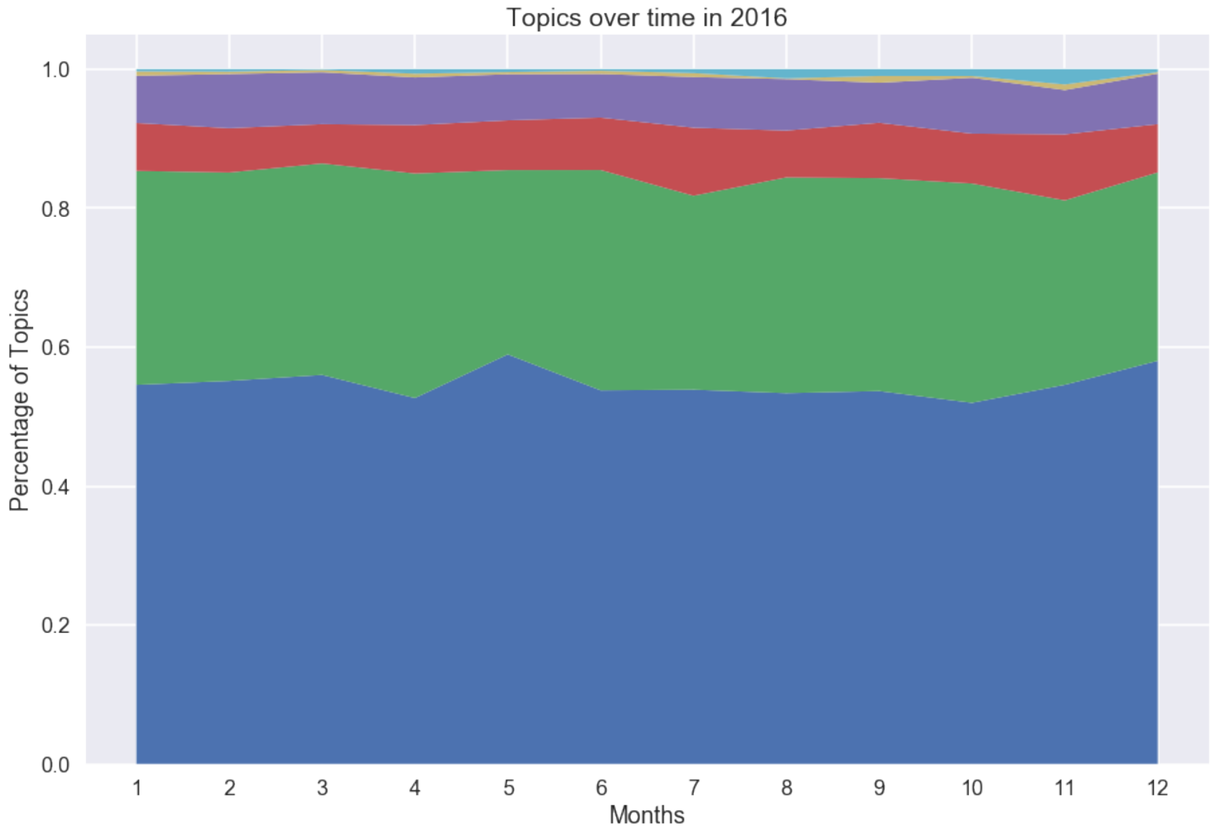
At the beginning of my current project, I had a job that involved 35T of input, but the vast majority of records would be ignored, and then for each successful one, only a few hundred bytes of output would be generated. Rather than Hadoop, I setup a simple system where a number of worker processes would query Postgres for the next available shard, mark it as in-progress, and then stream it from S3 and process it. When they finished, they'd write a CSV file back to S3. The reduce phase was just 'cat'.

The resulting system took a few hours to build (few days, including the actual algorithms run), and it was *much* more debuggable than Hadoop would be. You could inspect exactly where the job was, what shards had errored out, and which were currently running on machines, and download & view intermediate results before the whole computation finished. You could run the workers locally on a MBP if you needed to debug a shard, with no setup needed.


smelly marcus chisel phoneme esperanto masculine carving tw triviality slavic
entropy bloomberg certainty linkedin ec matt po siri receipt ic
listen music amp song copyright artist video voice audio sound
criminal claim police legal state law gun case crime information
alan async ebay joining amazon publisher auction idle asynchronous affiliate
flesh milk marking bb brown australia watt wheat australian foster
science computer class student university learn school job math college
device phone iphone game use app apps mobile android screen
secure key encrypted security encryption password dns attack certificate access
tee sierra proportionally vienna fab toy ubi austria counterfeit elevation
turner dane buffalo truffle informant imgur intimidate induction sewer bison
money founder team startup employee company product investor business market
gender white sex boy emission woman girl female dating male
ui chrome webkit chromium browser html web tab plugin jquery
apartment swift fish nvidia steam distance sugar cook eat taste
use page search site facebook google people user email link
twofold colorspace gimp preprocessing psychopath psychopathy gamma waived psychotic favicons
drug research body science food doctor study medical health effect
recommend reading gt author read learning list book writing history
neue uuid analyser coda outdoor snowboarding ner aria silk trinity
penis websocket comet divine cellular typewriter prophet naval morgan dreamed
alphabet framework django flask signature salt rail rack asp crystal
apartheid jockey rodent spurring quot envoy zork cacheable irrelevance oldschool
japanese truck oil power vote plane global electric voting japan
pay tax market cost money government economy bank income stock
vending cloning nickel allergic icmp signalling delaying spoof repairing organise
model node function case number data result example algorithm value
branch remote commit github working project work git repo repository
parking driving mile car lane driver ticket traffic road drive
trove substitution uris regexes uri mx homer barebones audiobooks rw
geometry confuses palantir src logo design grail designer ant colour
thing one think make people good even time get work
sphere dyson airbnb evernote def groupon digest hex jason usr
service database server data system use client cloud using network
colorado prisoner activation election ballot ga cure cycling spiral prison
doj delayed wilson bloated nt fred horizon headset retina vessel
poland oxygen inc combustion grammar queen bloom dice ajax spelling
syllabus maddening smuggling heist irompython freezer summarise logitech reagent compassionate
vega deterrent reservation openstreetmap galileo franchise hotel osm booking kindly
society kid city cancer brain population age human child intelligence
supersonic alcoholic watson booze optimizer hoover nacl boredom bro strapped
lilypond supplementing likelyhood polygon methionine aire fudge latvian dynos emotive
neglect rethinkdb jarvis invalid undertake refugee unfinished afaict putin meego
racism race marriage cheese rust nazi muslim floppy racist babel
boeing varnish conditioned soldier terrain plumber stl diy ai marine
eating beer shipping airline ca tesla shirt shoe free bag
simcity guncorn telomere addendum ox schizophrenia columbus seaside capitan jennifer
nand she additive polymer consolidated sabotage gh mythical kindness slicing
kernel command run shell terminal install window linux system unix
desktop device computer apple ipad hardware flash mac microsoft laptop
exporting iran gulf south refined north blender syria korea korean
price product free money customer company cost service business pay
package open developer io free software code license source project
tag document folder mapped registry file hierarchy directory dropbox filesystem
chinese china government country political internet state speech american english
need really time one thing work make get use much
disposition repayment barter tatari gmt flaking astounded mediated vox robinson lettuce
marijuana semantic turtle diagram temple minded sydney shameless spoiler
year energy could water technology space interesting see light article
library javascript programming language project use code java python write


smelly
marcus
chisel
esperanto
masculine
carving
slavic
entropy
bloomberg
linkedin
music
amp
video
audio
sound
criminal
police
legal
law
case
crime
amazon
publisher
science
computer
student
school
math
college
phone
game
app
android
secure
key
security
password
attack
sierra
toy
ubi
austria
buffalo
truffle
imgur
induction
money
founder
startup
employee
company
product
investor
business
market
gender
white
sex
woman
girl
female
male
chrome
browser
html
tab
plugin
jquery
webkit
chromium
apartment
swift
distance
sugar
eat
google
site
user
gimp
psychopath
gamma
psychotic
drug
research
food
study
health
effect
reading
gt
read
book
writing
uuid
outdoor
silk
penis
websocket
divine
cellular
typewriter
morgan
dreamed
alphabet
framework
django
flask
rail
rack
asp
quot
irrelevance
oldschool
japanese
power
oil
vote
japan
electric
global
pay
tax
money
bank
income
cloning
nickel
model
node
function
data
example
value
branch
remote
commit
github
project
git
repo
driving
car
driver
traffic
road
drive
trove
substitution
uris
regexes
uri
mx
design
designer
logo
thing
one
think
people
make
good
even
time
work
sphere
airbnb
evernote
def
groupon
digest
jason
service
database
server
data
cloud
client
colorado
prisoner
activation
election
ballot
cure
cycling
spiral
prison
delayed
bloated
horizon
headset
retina
grammar
ajax
spelling
syllabus
maddening
freezer
logitech
reagent
vega
reservation
franchise
hotel
booking
kindly
city
brain
human
child
alcoholic
watson
optimizer
boredom
bro
polygon
methionine
dynos
neglect
rethinkdb
invalid
refugee
putin
racism
race
rust
muslim
racist
babel
boeing
varnish
soldier
diy
marine
beer
shipping
tesla
shoe
free
bag
nand
polymer
mythical
kindness
kernel
command
shell
window
linux
system
unix
apple
hardware
microsoft
laptop
iran
south
north
korea
korean
price
product
customer
service
business
open
developer
software
code
source
document
folder
file
directory
dropbox
china
government
country
time
one
work
use
gmt
vox
repayment
semantic
diagram
minded
energy
could
space
water
light
article
library
programming
language
project
code
java
python





Startup / Business

OS / Linux

Middle East

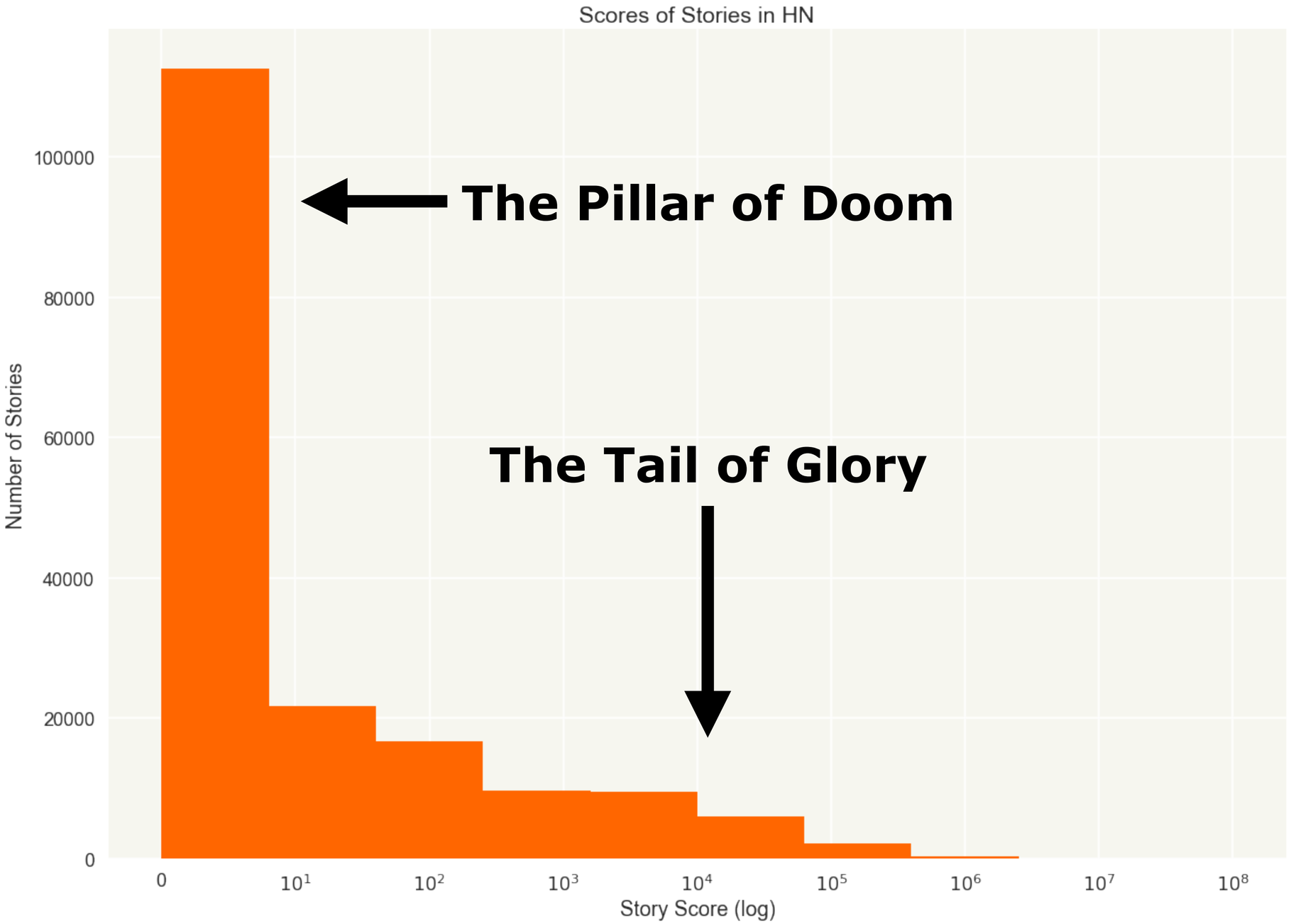
Power / Politics

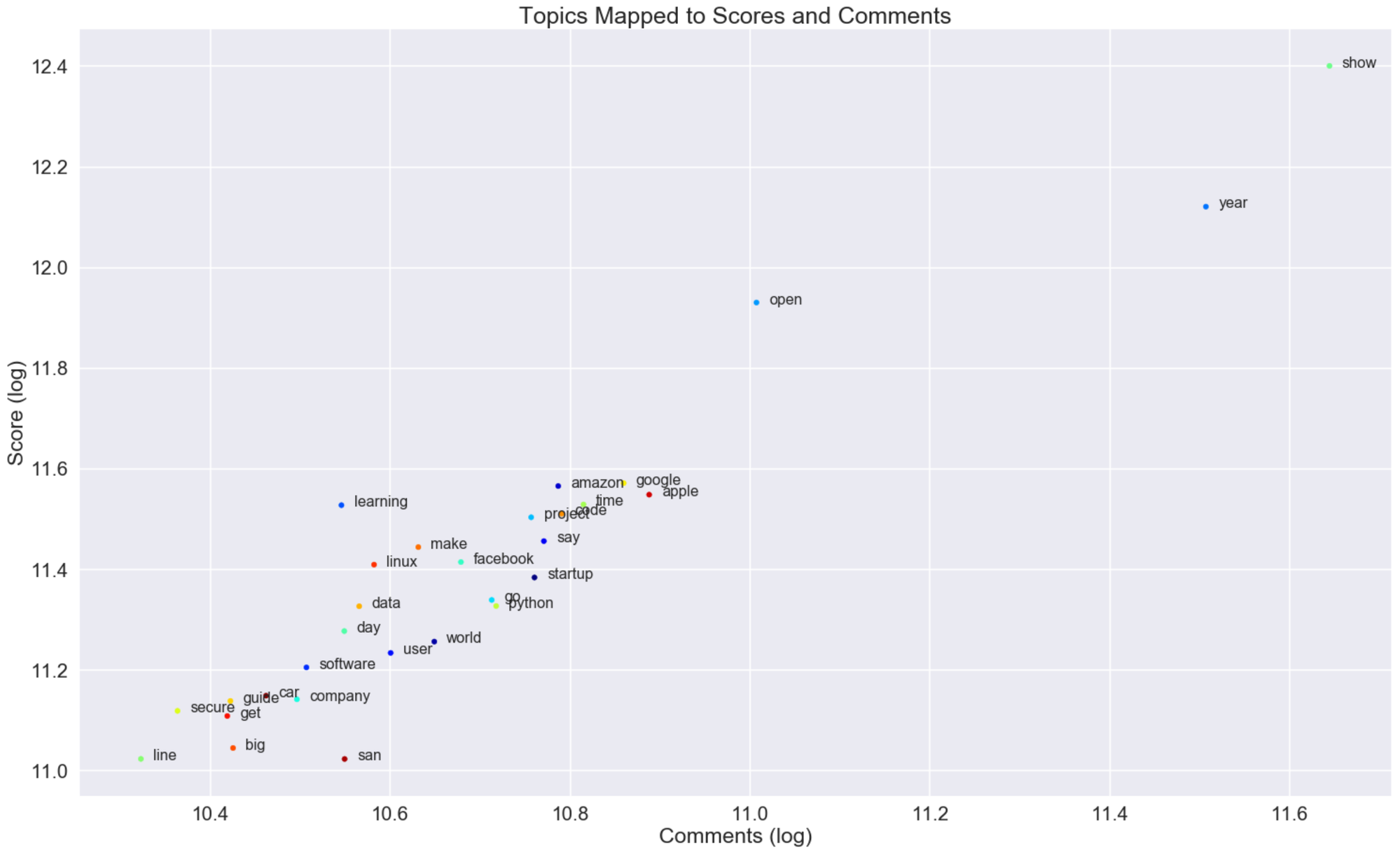
Web / Browsers

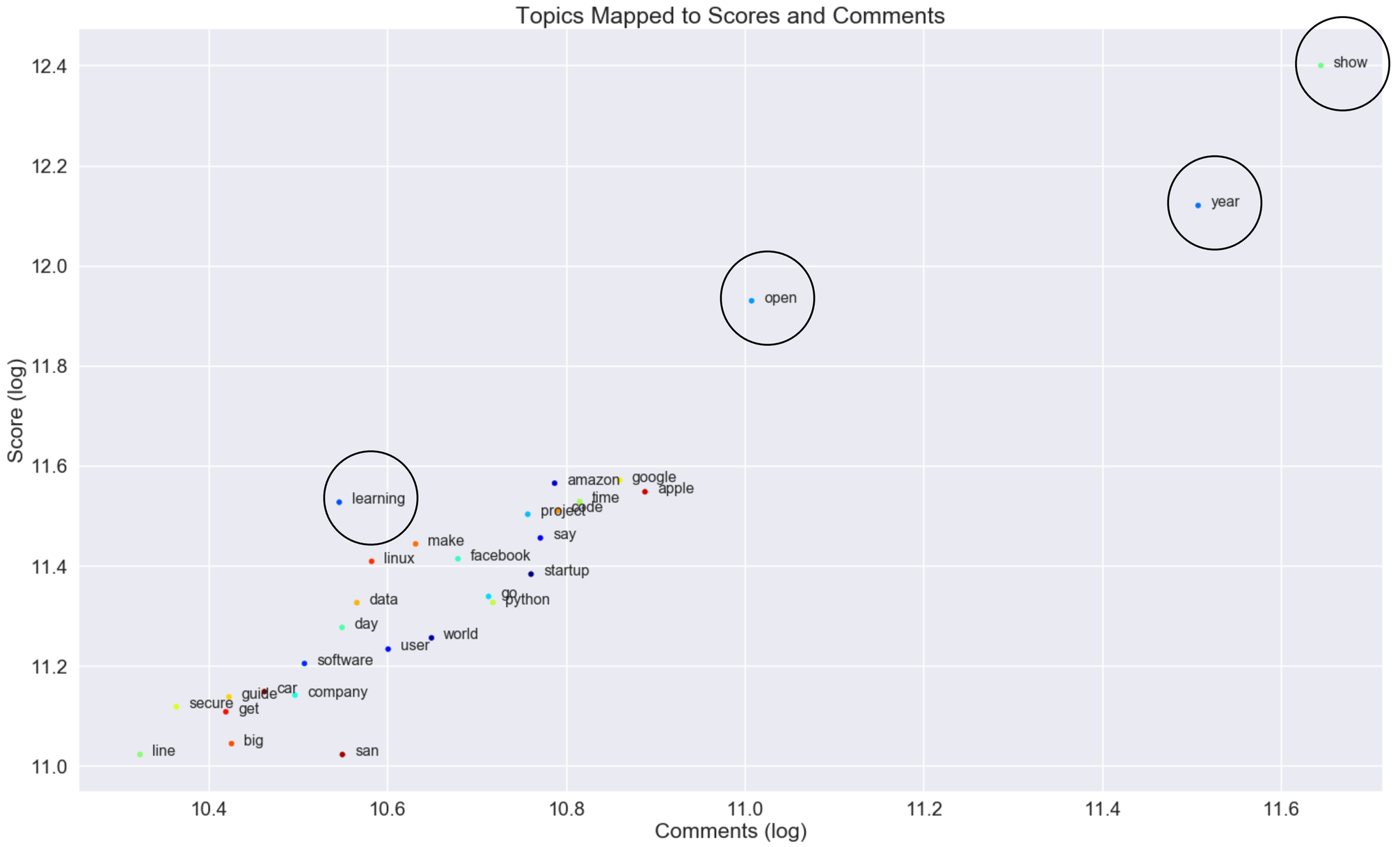
Election / Justice

Understanding Culture Through HN Score


Hacker News new | comments | show | ask | jobs | submit login
1. Don't use Hadoop when your data isn't that big (2013) (chrisstucchio.com)
217 points by tosh 3 hours ago | hide | 148 comments
2. The Blockstack Browser: A Gateway to a New, Decentralized Internet (blockstack.org)
102 points by adunk 1 hour ago | hide | 28 comments
Chaos Computer Clubs Breaks Iris Recognition System of the Samsung Galaxy S8 (ccc.de)
459 points by morsch 7 hours ago | hide | 117 comments
At The Dawn Of Recorded Sound, No One Cared (npr.org)
23 points by collapse 52 minutes ago | hide | 4 comments
How jeans conquered the world (2012) (bbc.com)
23 points by prmph 54 minutes ago | hide | 2 comments
I've become worse at programming (fascinatedbox.blogspot.com)
189 points by kumaranvpl 5 hours ago | hide | 54 comments
New Surface Pro (microsoft.com)
190 points by pierre-renaux 5 hours ago | hide | 315 comments
BuildZoom (YC W13) is hiring a VP of Sales (lever.co)
16 minutes ago | hide
22-Year-Old Lidar Whiz Claims Breakthrough (ieee.org)
9 points by deepnotderp 30 minutes ago | hide | discuss
Show HN: 15-question programming quiz with answers (triplebyte.com)
21 points by ammon 49 minutes ago | hide | 9 comments
Stack Overflow: Helping One Million Developers Exit Vim (stackoverflow.blog)
9 points by var_explained 9 minutes ago | hide | discuss
How to build your own particle detector (2015) (symmetrymagazine.org)
15 points by gus_massa 1 hour ago | hide | 1 comment
Choosing The Right Freelancing Niche [audio] (simplecast.com)
44 points by chris_hawk 3 hours ago | hide | 5 comments
Sand and gravel mining "greatly exceeds natural renewal rates" (newyorker.com)
8 points by sergeant3 1 hour ago | hide | 2 comments
Kickstarter set to launch in Japan, where hardware startups are finding it hard (techinasia.com)
6 points by williswee 1 hour ago | hide | discuss
Why Did a Chinese Peroxide Company Pay $1B for a Talking Cat? (bloomberg.com)
4 points by lumisota 37 minutes ago | hide | discuss
Jamboard is now available (blog.google)
12 points by happy-go-lucky 1 hour ago | hide | 3 comments
A Man Who Made the Mistake of Trying to Help Wikileaks (vice.com)
4 points by dsr12 45 minutes ago | hide | discuss
Chasing the Harvest: 'If You Want to Die, Stay at the Ranch' (longreads.com)
5 points by DiabloD3 1 hour ago | hide | discuss
Zuckerberg-Backed Data Trove Exposes the Injustices of Criminal Justice (wired.com)
74 points by sprucely 2 hours ago | hide | 17 comments
Arq 5.8.5 for Mac Fixes a Bad Bug (arqbackup.com)
107 points by ivank 6 hours ago | hide | 30 comments


Scores of Stories in HN
The Pillar of Doom
The Tail of Glory
Number of Stories
100000
80000
60000
40000
20000
0
0
10^1
10^2
10^3
10^4
10^5
10^6
10^7
10^8
Story Score (log)


Topics Mapped to Scores and Comments
Score (log)
Comments (log)
12.4
12.2
12.0
11.8
11.6
11.4
11.2
11.0
10.4
10.6
10.8
11.0
11.2
11.4
11.6
show
year
open
amazon
google
apple
learning
time
code
project
say
make
linux
facebook
startup
go
python
data
day
world
user
software
guide
car
company
secure
get
big
line
san


Topics Mapped to Scores and Comments
Score (log)
Comments (log)
12.4
12.2
12.0
11.8
11.6
11.4
11.2
11.0
10.4
10.6
10.8
11.0
11.2
11.4
11.6
show
year
open
learning
amazon
google
apple
time
code
project
say
make
linux
facebook
startup
go
python
data
day
world
user
software
guide
car
company
secure
get
big
line
san

Topic 1: startup computer tech problem application
Topic 2: world support wrong looking everything
Topic 3: amazon rust book real text
Topic 4: say design may month government
Topic 5: user phone model tesla ever
Topic 6: software case business git memory

Topic 7: learning deep building start neural

Topic 8: year programming one pdf developer
Topic 9: open source system could library

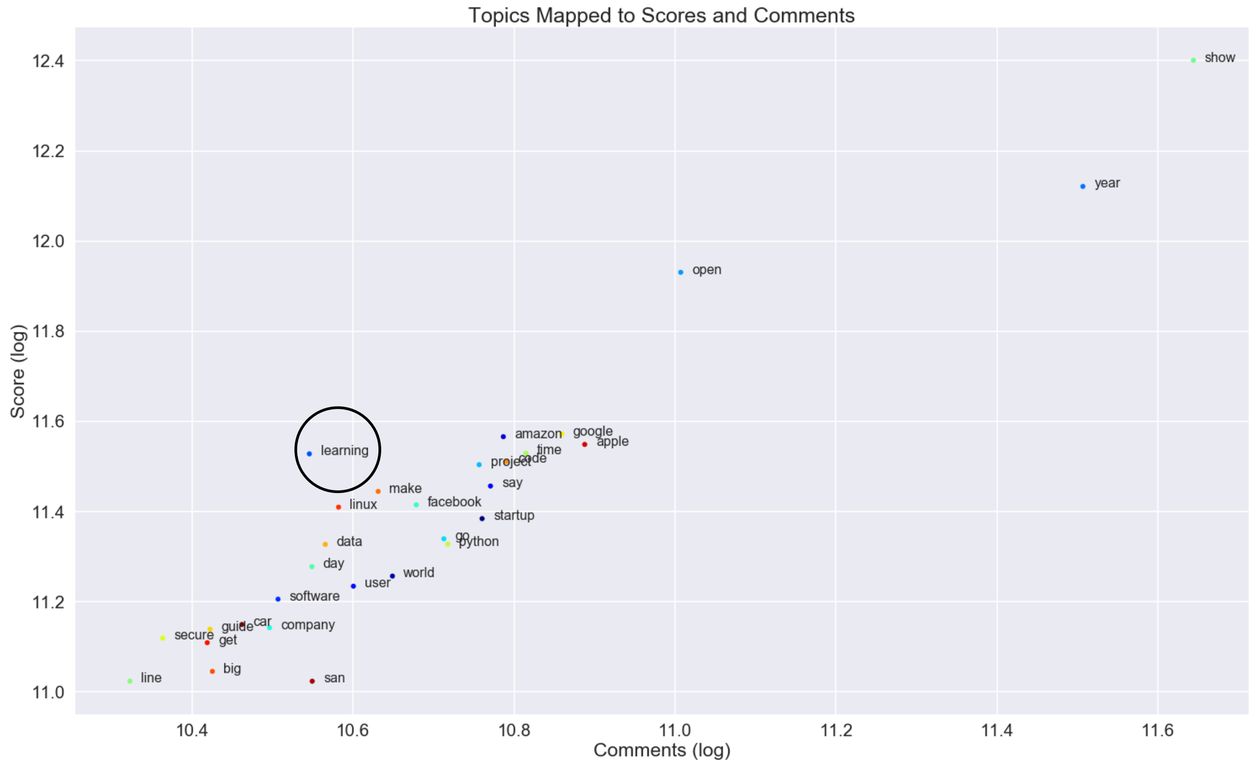
Topic 10: project microsoft window side built
Topic 11: go china change search result
Topic 12: company american police died operating
Topic 13: facebook fast employee news story
Topic 14: day tool right human platform

Topic 15: show new work web video

Topic 16: line mac study program basic
Topic 17: time way build learn still
Topic 18: python github people secret pro
Topic 19: secure nasa living ground giant
Topic 20: google research founder live bitcoin
Topic 21: guide online law record minute
Topic 22: data hacker many future bad
Topic 23: code machine uber part better
Topic 24: make released internet rule name
Topic 25: big engineer making ceo mobile
Topic 26: linux react engine power community
Topic 27: get self page stack keep
Topic 28: apple game network security life
Topic 29: san francisco lisp office macos
Topic 30: car test crash camera swift

Machine Learning

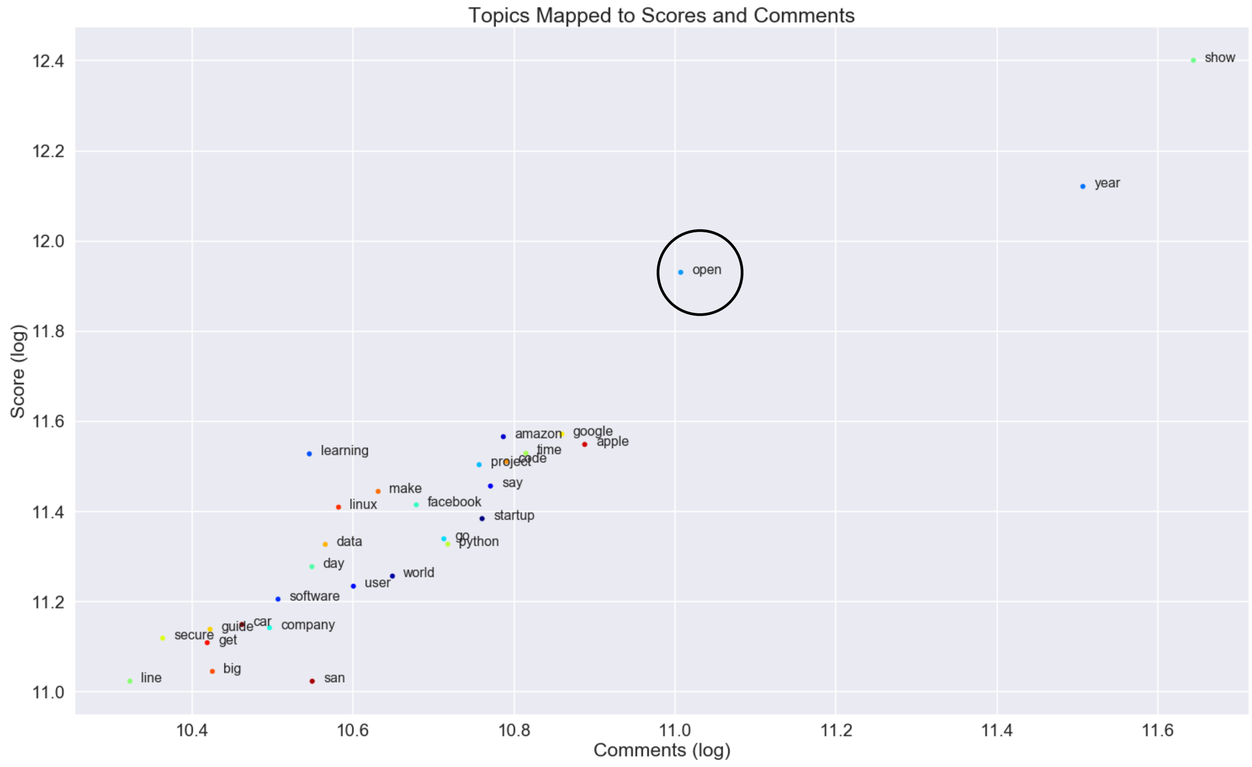
1. LSTM Neural Networks for Time Series Prediction
2. Symbolic Machine Learning
3. A Guide to Deep Learning
4. Learning Machine Learning: A beginner's journey
5. Advice on ***learning*** Python efficiently
6. An introduction to Machine Learning



Topic 1: startup computer tech problem application
Topic 2: world support wrong looking everything
Topic 3: amazon rust book real text
Topic 4: say design may month government
Topic 5: user phone model tesla ever
Topic 6: software case business git memory
Topic 7: learning deep building start neural
Topic 8: year programming one pdf developer
Topic 9: open source system could library
Topic 10: project microsoft window side built
Topic 11: go china change search result
Topic 12: company american police died operating
Topic 13: facebook fast employee news story
Topic 14: day tool right human platform
Topic 15: show new work web video
Topic 16: line mac study program basic
Topic 17: time way build learn still
Topic 18: python github people secret pro
Topic 19: secure nasa living ground giant
Topic 20: google research founder live bitcoin
Topic 21: guide online law record minute
Topic 22: data hacker many future bad
Topic 23: code machine uber part better
Topic 24: make released internet rule name
Topic 25: big engineer making ceo mobile
Topic 26: linux react engine power community
Topic 27: get self page stack keep
Topic 28: apple game network security life
Topic 29: san francisco lisp office macos
Topic 30: car test crash camera swift

Open Source

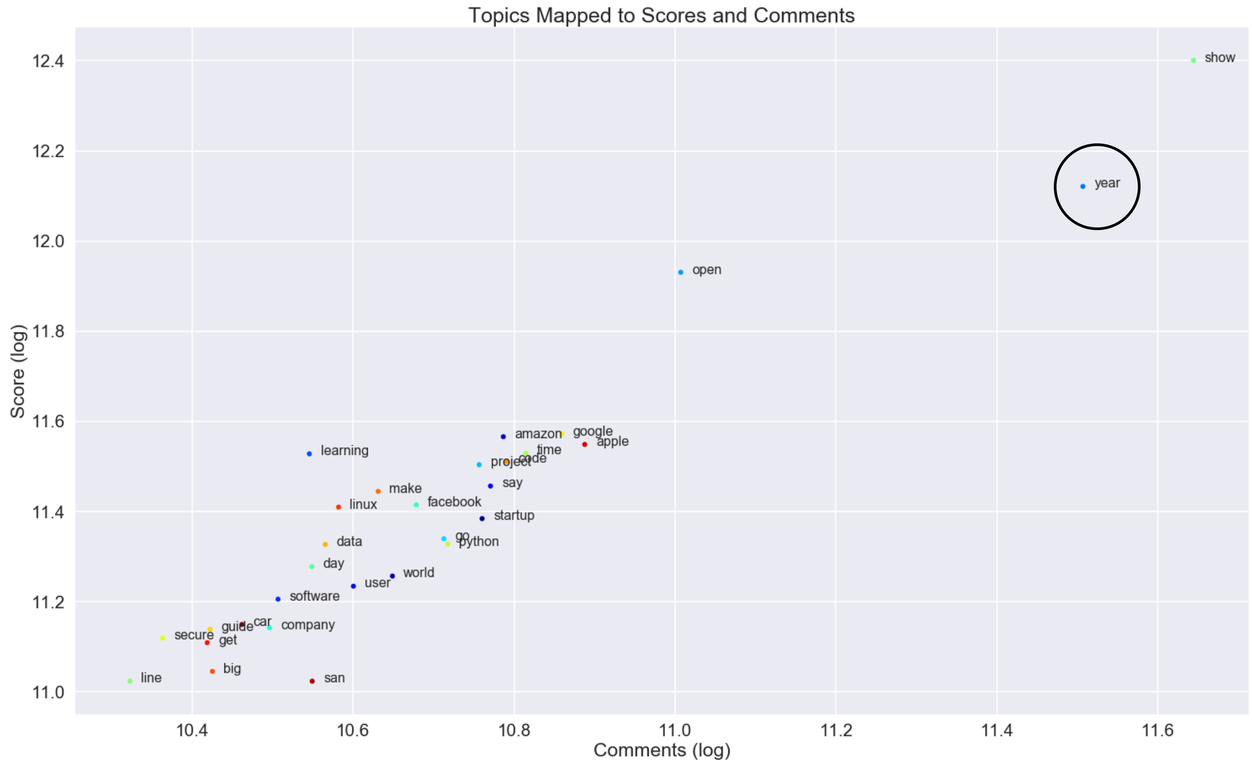
1. I Open-Sourced All My Business Ideas
2. OneOps – Open-source cloud ops platform from Walmart
3. List of high-quality open datasets in public domains
4. Apache Arrow: A new open source in-memory columnar data format



Topic 1: startup computer tech problem application
Topic 2: world support wrong looking everything
Topic 3: amazon rust book real text
Topic 4: say design may month government
Topic 5: user phone model tesla ever
Topic 6: software case business git memory
Topic 7: learning deep building start neural
Topic 8: year programming one pdf developer
Topic 9: open source system could library
Topic 10: project microsoft window side built
Topic 11: go china change search result
Topic 12: company american police died operating
Topic 13: facebook fast employee news story
Topic 14: day tool right human platform
Topic 15: show new work web video
Topic 16: line mac study program basic
Topic 17: time way build learn still
Topic 18: python github people secret pro
Topic 19: secure nasa living ground giant
Topic 20: google research founder live bitcoin
Topic 21: guide online law record minute
Topic 22: data hacker many future bad
Topic 23: code machine uber part better
Topic 24: make released internet rule name
Topic 25: big engineer making ceo mobile
Topic 26: linux react engine power community
Topic 27: get self page stack keep
Topic 28: apple game network security life
Topic 29: san francisco lisp office macos
Topic 30: car test crash camera swift

Programming / Reflection

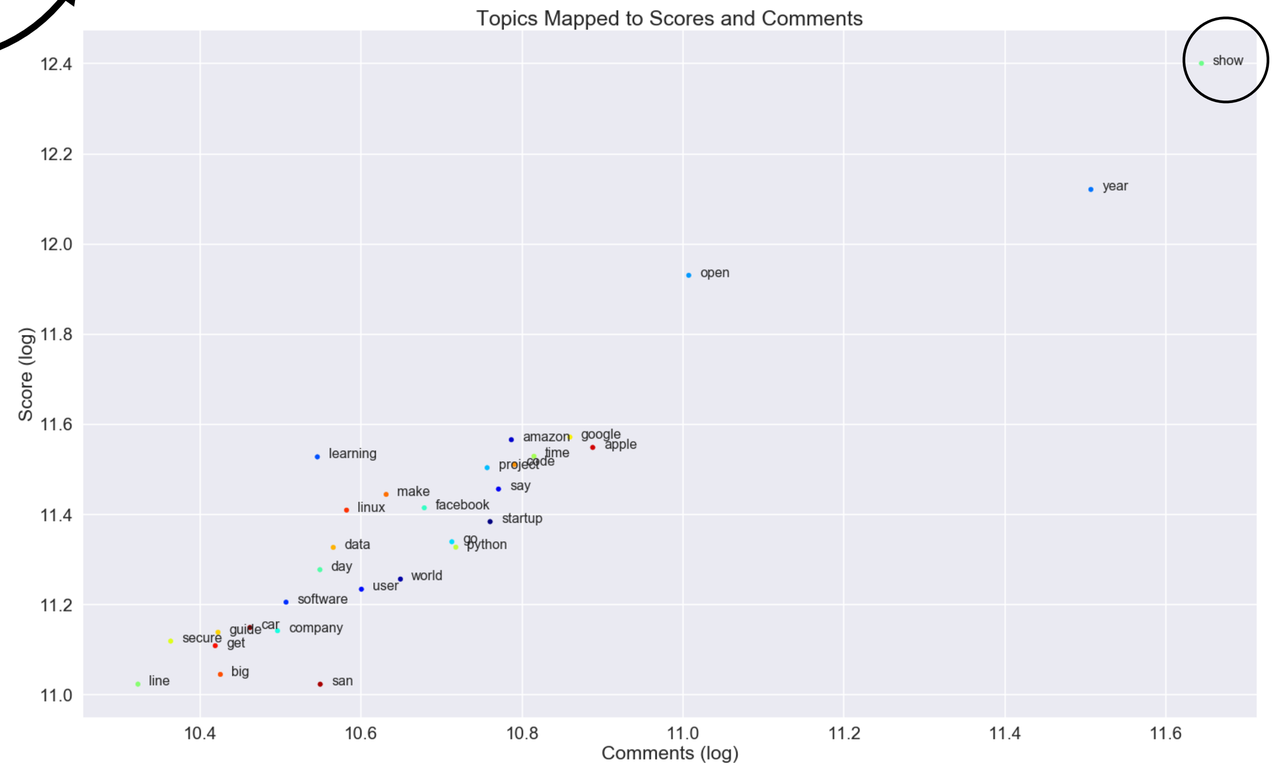
1. What's wrong with 2006 **programming**? (2010)
2. Happy new **year**
3. Choosing Functional **Programming** for Our Game
4. The Crystal **Programming** Language
5. Category Theory and Declarative **Programming**
6. Three **Years** as a One-Man Startup



Topic 1: startup computer tech problem application
Topic 2: world support wrong looking everything
Topic 3: amazon rust book real text
Topic 4: say design may month government
Topic 5: user phone model tesla ever
Topic 6: software case business git memory
Topic 7: learning deep building start neural
Topic 8: year programming one pdf developer
Topic 9: open source system could library
Topic 10: project microsoft window side built
Topic 11: go china change search result
Topic 12: company american police died operating
Topic 13: facebook fast employee news story
Topic 14: day tool right human platform
Topic 15: show new work web video
Topic 16: line mac study program basic
Topic 17: time way build learn still
Topic 18: python github people secret pro
Topic 19: secure nasa living ground giant
Topic 20: google research founder live bitcoin
Topic 21: guide online law record minute
Topic 22: data hacker many future bad
Topic 23: code machine uber part better
Topic 24: make released internet rule name
Topic 25: big engineer making ceo mobile
Topic 26: linux react engine power community
Topic 27: get self page stack keep
Topic 28: apple game network security life
Topic 29: san francisco lisp office macos
Topic 30: car test crash camera swift

Show HN

1. Show HN: Scalable reverse image search built on Kubernetes and Elasticsearch
2. Show HN: Using Google AMP to build a Medium-style Jekyll site that loads in 65ms
3. Show HN: Localkube – zero to Kubernetes 1.2 in one command
4. Show HN: Micro – a microservice toolkit



Developer Culture

Not just about the bits

Business

Politics

Tools

Programming Languages

Next Steps

- Scrape the text from the **content posted**
- Topic analysis on the **job postings**
- Topic analysis across **all of time**

Thank You.

Reference:

1. https://blog.kissmetrics.com/reddit-marketing-guide/
2. https://techcrunch.com/2013/05/18/the-evolution-of-hacker-news/
3. http://www.heavybit.com/library/podcasts/practical-product/ep-8-launch-day-trifecta-hackernews-producthunt-techcrunch/
4. https://wiredcraft.com/blog/how-to-post-on-hacker-news/
5. https://jacquesmattheij.com/how-to-make-the-hacker-news-homepa

https://news.ycombinator.com/item?id=10047481
http://blog.datadive.net/which-topics-get-the-upvote-on-hacker-news/


Scrapy


mongoDB.


scikit
learn