



PyCon MY 2017

PYTHON: IMPACTING THE BUSINESS WORLD

Keynote Presentation by Luis M. Sánchez

WHY PYTHON?

1. **PYTHON HAS A HEALTHY, ACTIVE AND SUPPORTIVE COMMUNITY:** HIGH VOLUME OF USERS IN STACKOVERFLOW, GITHUB, ETC.
2. **PYTHON HAS SOME GREAT CORPORATE SPONSORS:** GOOGLE ADOPTED IT HEAVILY IN 2006
3. **PYTHON HAS BIG DATA:** TRADITIONALLY DOMINATED BY R, THAT'S NO LONGER THE CASE
4. **PYTHON HAS AMAZING LIBRARIES:** NUMPY, PANDAS, ETC.
5. **PYTHON IS RELIABLE AND EFFICIENT:** DESKTOP, MOBILE APPS, HARDWARE ETC.
6. **PYTHON IS ACCESSIBLE:** SIMPLIFIED SYNTAX WITH EMPHASIS ON NATURAL LANGUAGE

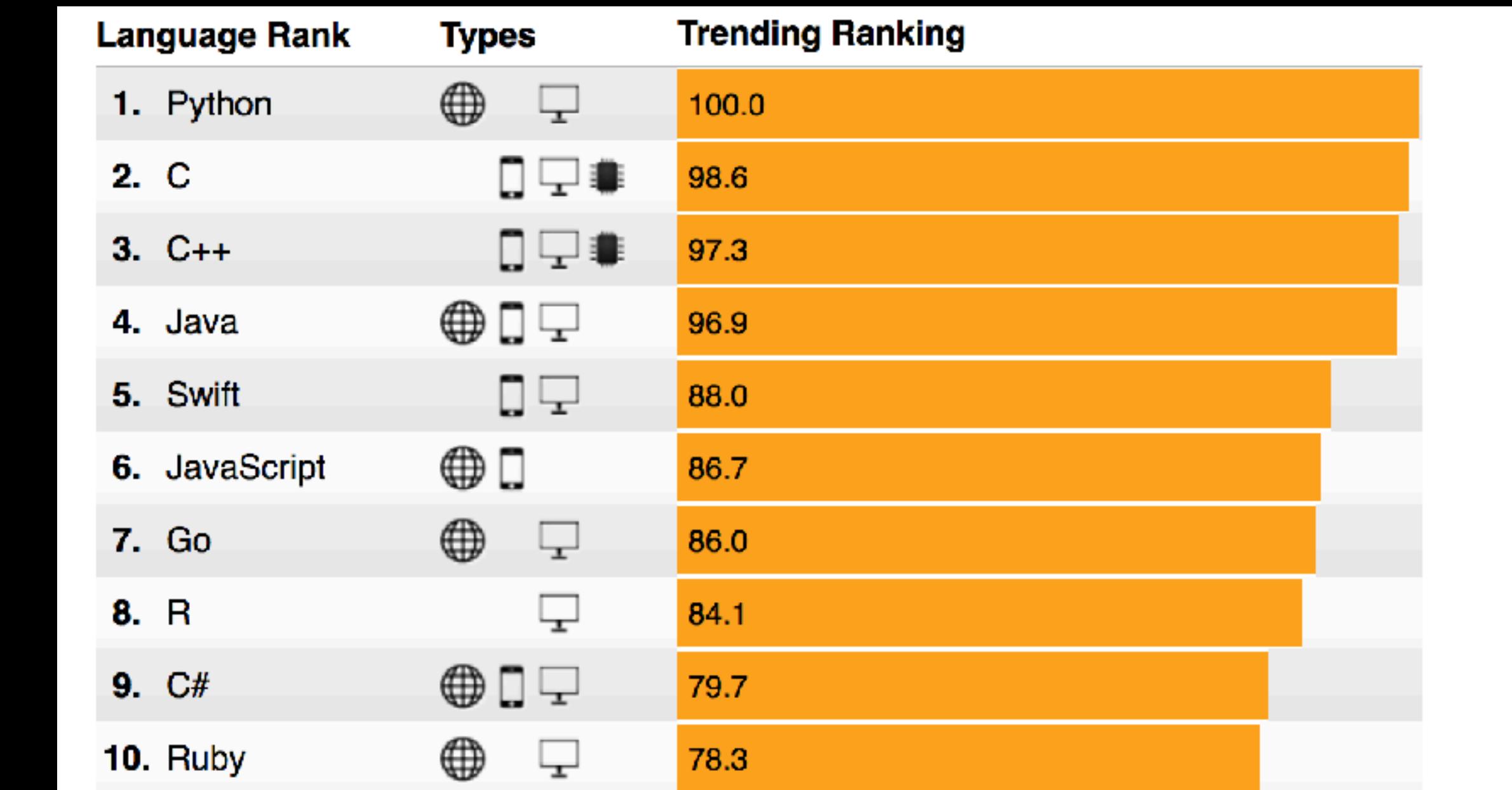
PHYTON IS “NUMERO 1”

According to the Institute of Electrical and Electronics Engineers (IEEE), a professional association with its corporate office in New York City and its operations center in Piscataway, New Jersey (It was formed in 1963 from the amalgamation of the American Institute of Electrical Engineers and the Institute of Radio Engineers.)



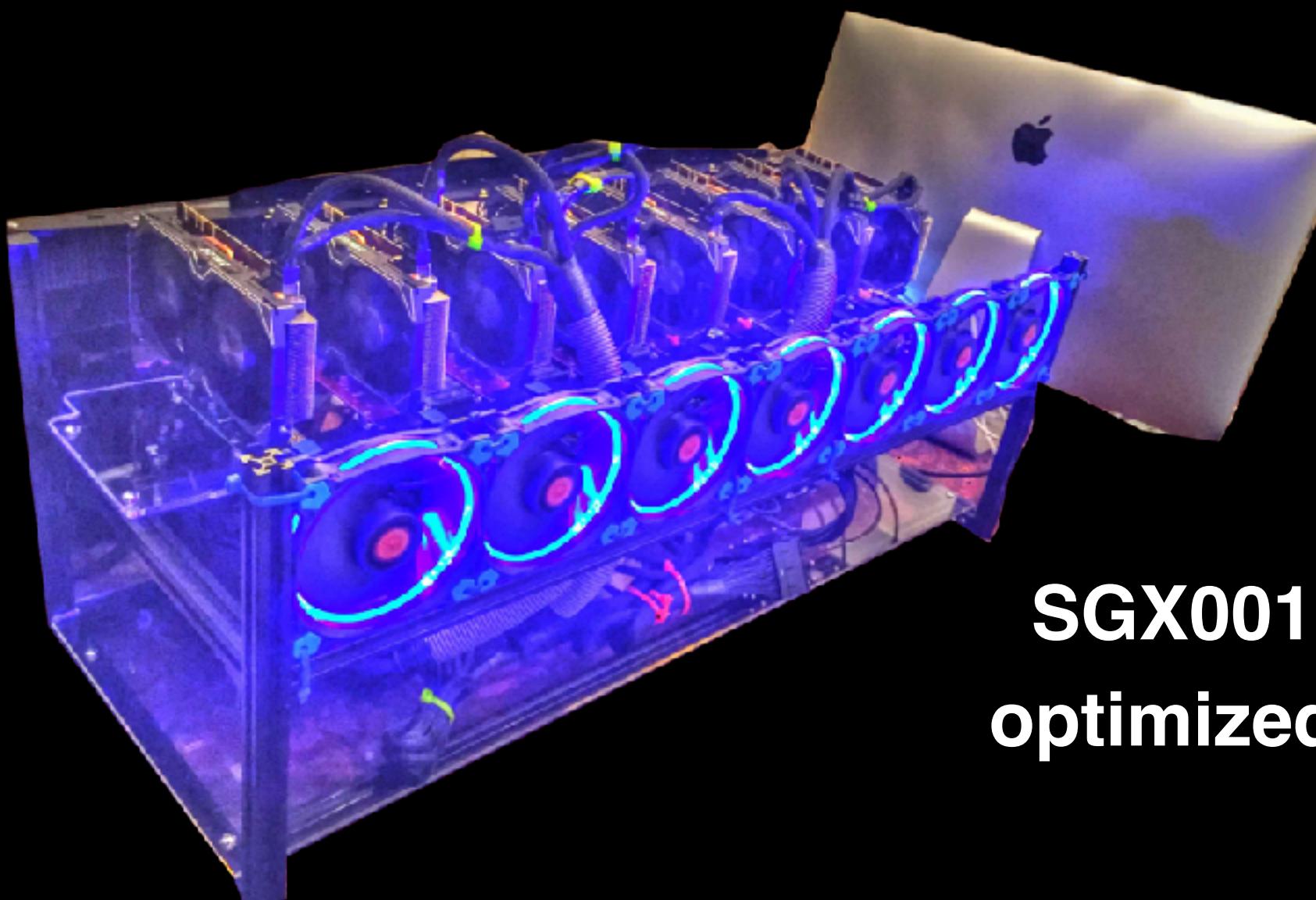
SUMMER 2017 STUDY

12 METRICS FROM
10 CAREFULLY
CHOSEN ONLINE
SOURCES TO RANK
48 LANGUAGES...



**WHAT SPECIFIC
TOOLS? TECHNIQUES?**

For dNN, choose your architecture



SGX001 DevBox, 8 Pascal GPUs,
optimized for financial applications



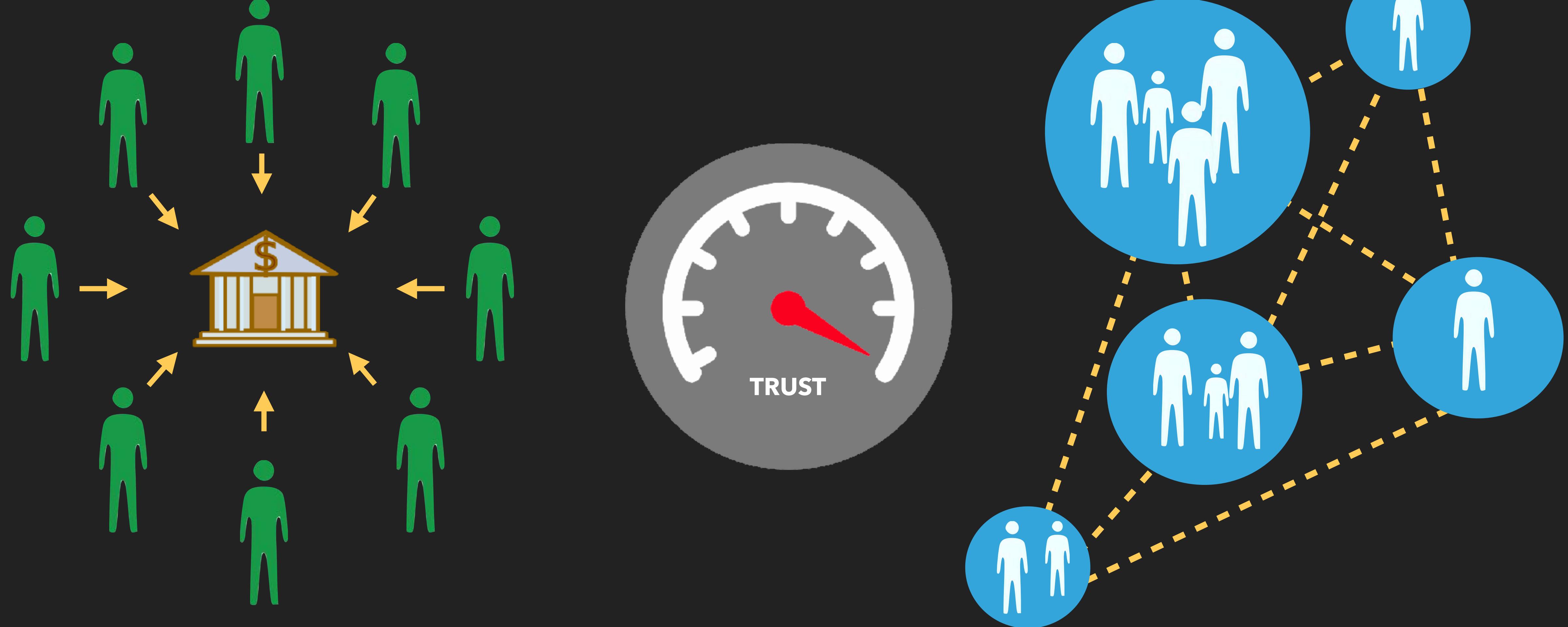
NVIDIA DevBox, 4 TitanX, general
dNN

“AS PEOPLE'S ACCESS TO THE INTERNET GROWS WE'RE SEEING THE SHARING ECONOMY BOOM – I THINK OUR OBSESSION WITH OWNERSHIP IS AT A TIPPING POINT AND THE SHARING ECONOMY IS PART OF THE ANTIDOTE FOR THAT.”

Sir Richard Branson



THE GREAT TRUST SHIFT: FROM INSTITUTIONS TO INDIVIDUALS:



SHARING ECONOMY

- Peer economy
- Crowdfunding
- Open source
- Maker spaces
- Collaborative consumption
- Crypto currency mining

SHARING ECONOMY

An economy built on distributed networks of connected individuals and communities as opposed to centralized institutions, transforming how we can produce, consume, learn, insure, and finance:



Production

Quirky, TechShop,
CrowdFlower



Consumption

AirBnB, Uber,
Lyft, Zipcar



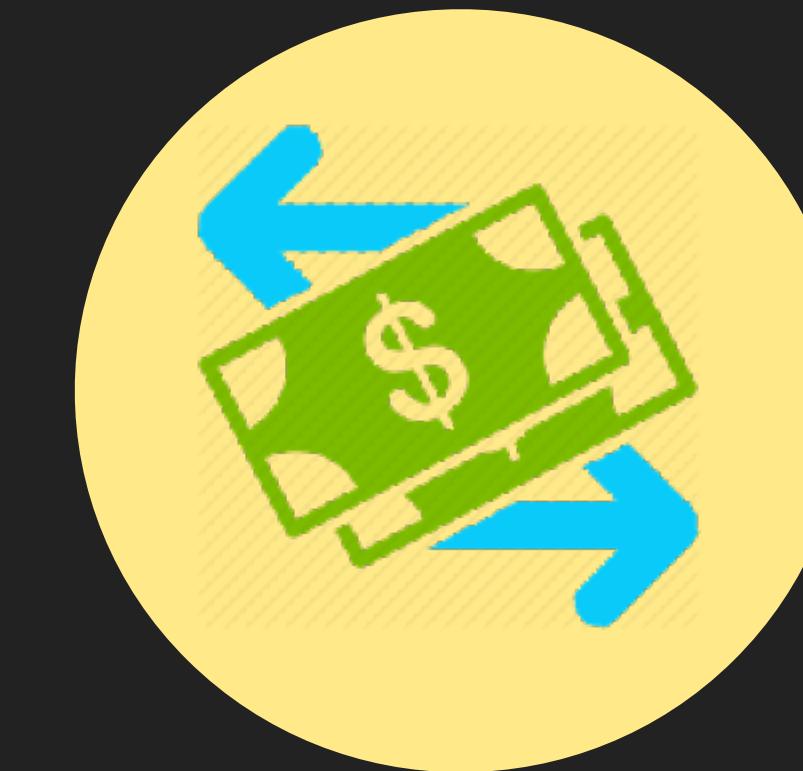
Education & Entertainment

Coursera, Udemy
EDx, GitHub



Insurance

CatRisk, Lemonade



Finance

KickStarter, Coinbase,
Crowdster, Indigogo,
Cryptocurrencies

SOME OF THE PYTHON BUSINESS APPLICATIONS WHERE MY FIRM HAS BEEN OR IS INVOLVED WITH

**Stock market
prediction**

**Design of financial
instruments to provide
protection against the
economic impact of
catastrophic natural events**

**Cryptocurrency mining
and trading**

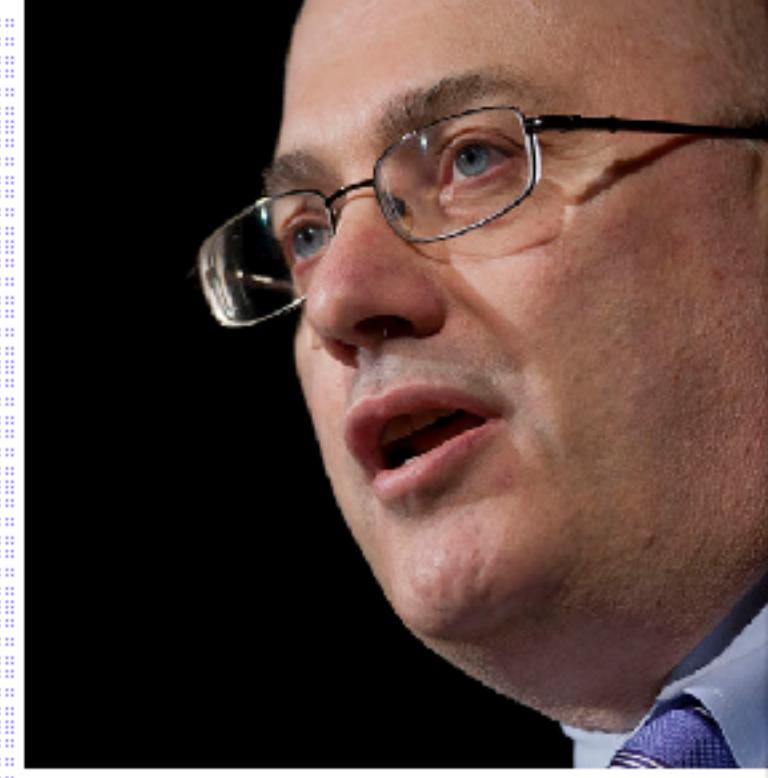
Music composition

STOCK MARKET PREDICTION

Early in 2014, I was approached by Cohen's Point72 to 'advise' on Data Investing and to learn about the code I had developed, an offshoot of my semantic search engine/data collection algorithms. They also inquired about the possibility of me selling the data I had collected for their own backtests. I did not sell my data and analytics. Later on...

BloombergBusiness News Markets Insights Video

Cohen's Point72 Hires 30 People for Big Data Investing



FINANCE More: Hedge Funds Stocks

Steve Cohen's fund is going quant

STEPHANIE YANG MAR. 10, 2015, 10:28 AM | 4,816 ▾ 6

[FACEBOOK](#) [LINKEDIN](#) [TWITTER](#) [EMAIL](#)

Steve Cohen's Point72 Asset Management is moving into quantitative investing.

According to Bloomberg, the firm has 30 new hires dedicated to building investing models that use computer analysis of public data. A spokesman for Point72 Asset Management said the new project was called Aperio.

He said Aperio would work with all seven of the firm's equity units to provide a technological and data-driven edge.

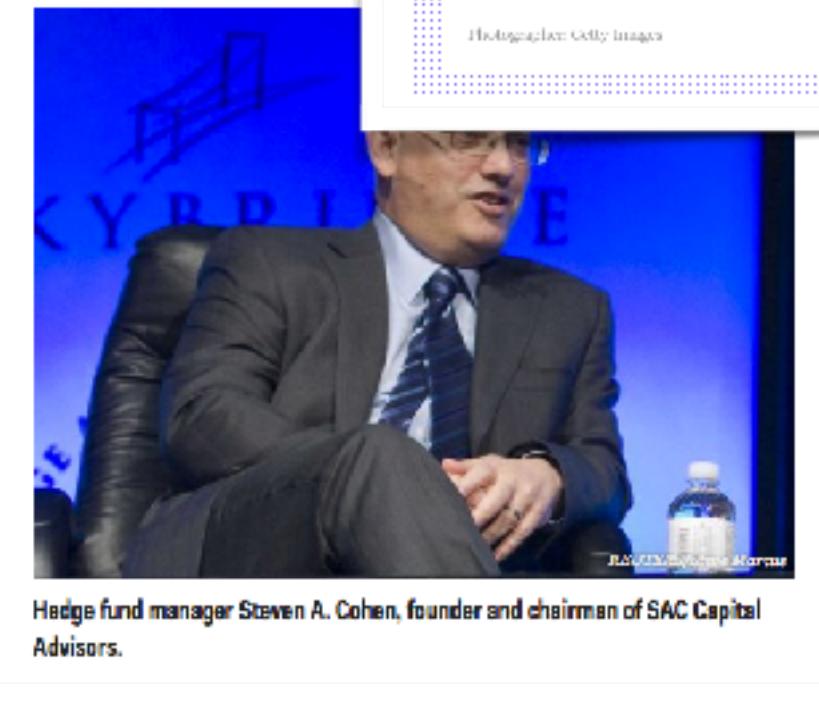
BloombergBusiness News Markets Insights Video

The Crucial Piece of Information That Big Traders Get Before Everyone Else

Everyone else is flying blind



Photo: Getty Images



What are earnings per share (EPS)? Why are they important?

- EPS serves as an indicator of a company's profitability
- It is the portion of a company's profit (or loss) allocated to each outstanding share of common stock in a given quarter.

Example: GRPN EPS Q2 2013

Top Line Revenue:	\$608,747	← top line revenue
Cost of Net Revenues:	(\$224,053)	
Gross Profit:	\$384,694	
Administrative and General Expenses:	(\$357,282)	
Income from Operations:	\$27,412	
Interest + other income/(expenses) + tax provision:	(\$34,986)	
Profit/Loss:	-\$7,574	
Common shares outstanding	662,361	
EPS	-\$0.01	← negative earnings

What are quarterly earnings announcements? Why are they important?

- Formal reports mandated by the SEC, called 10-Q reports
- Public companies in the US are subject to SEC Rule 10b-5: strict rules/penalties for erroneous misleading disclosure

When are the announcements made?

- Typically, 14-28 days after the fiscal end of Q1, Q2, and Q3
- Up to 90 days after end of Q4, since results need to be audited by a independent public accounting firm.

There is plenty of time for ‘information leakage’ by insiders that could be reflected in anomalies in buy/sell patterns of stocks and/or stock options, open interest spikes, etc. Some of these patterns could be detected by machine learning algorithms.

There is also a more ‘proactive’ approach using data collection and machine learning algorithms to ‘forecast’ top line revenues and other key components of a company’s balance sheet.

How good are Wall Street analysts estimating top line revenues / earnings as a group?

-Let's first define Earnings Surprises

**Positive
Earnings**

**Negative
Earnings**

**Negative
Surprise**

Consensus EPS = \$0.10
Actual EPS = \$0.08
Surprise = -\$0.02
“Miss” of 20%

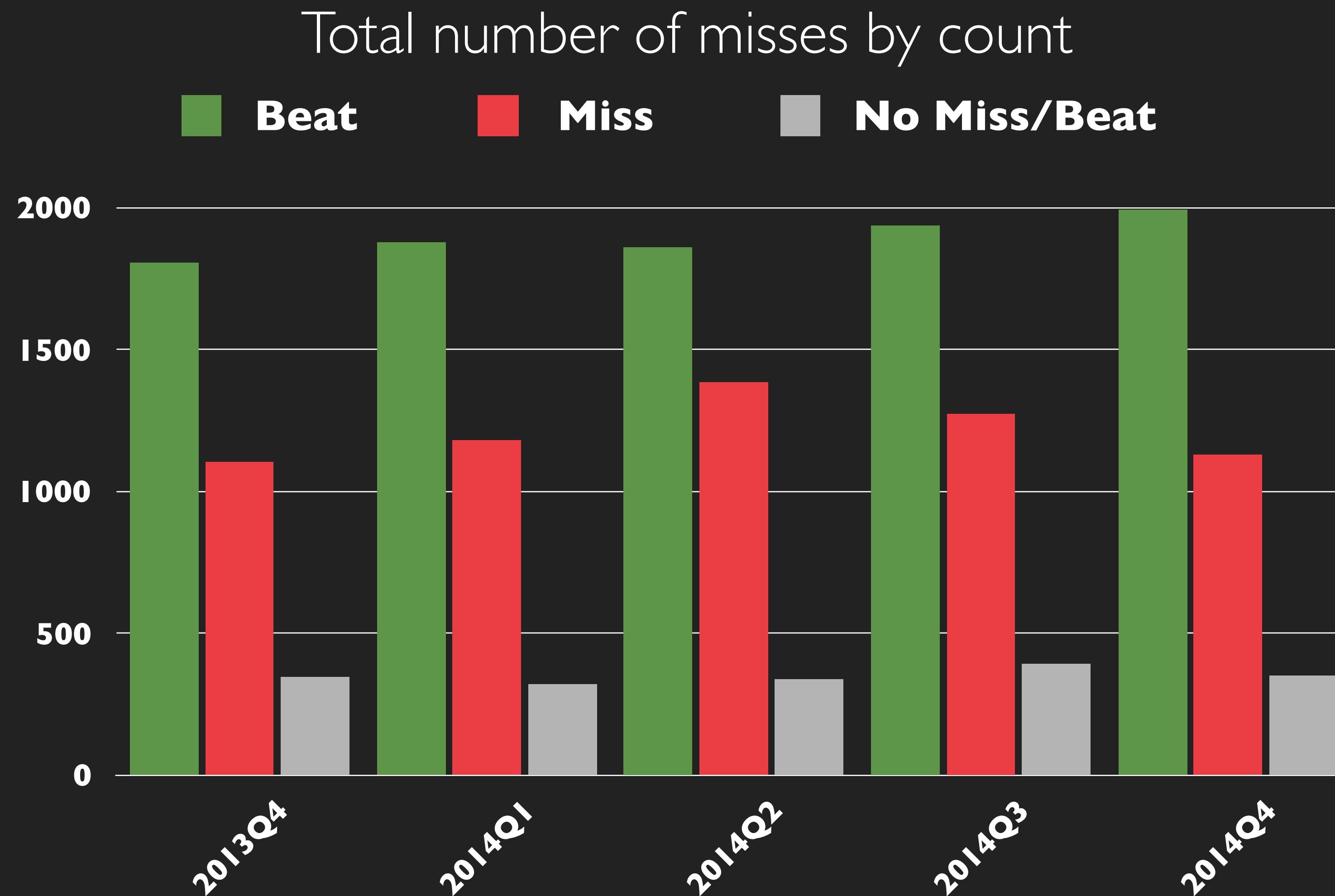
Consensus EPS = -\$0.03
Actual EPS = -\$0.06
Surprise = -\$0.03
“Miss” of 100%

**Positive
Surprise**

Consensus EPS = \$0.20
Actual EPS = \$0.25
Surprise = +\$0.05
“Beat” of 25%

Consensus EPS = -\$0.15
Actual EPS = -\$0.10
Surprise = +\$0.05
“Beat” of 33%

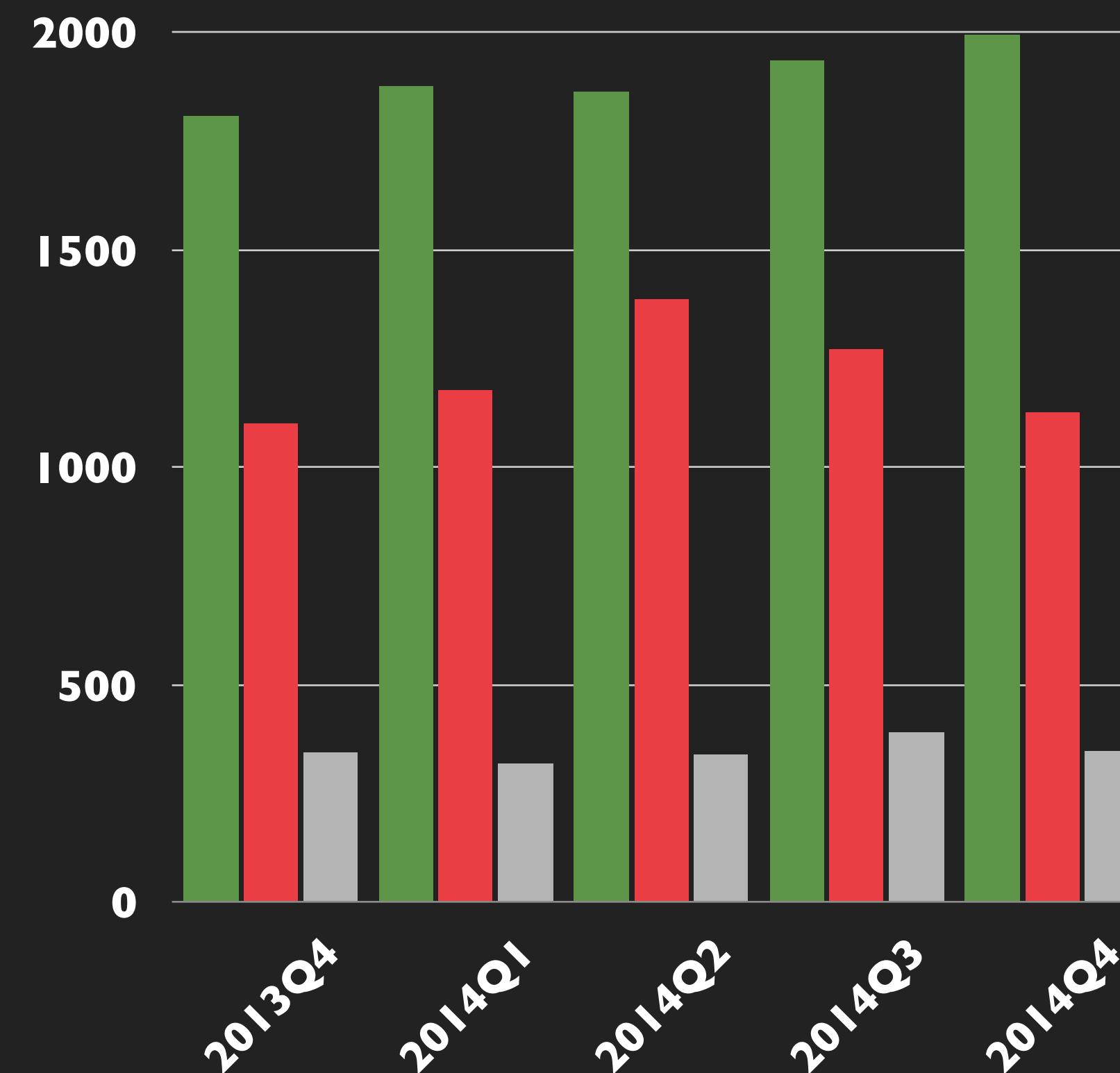
Wall Street tends to be more wrong than right when it comes to guessing earnings..



They are also large in dollar value.

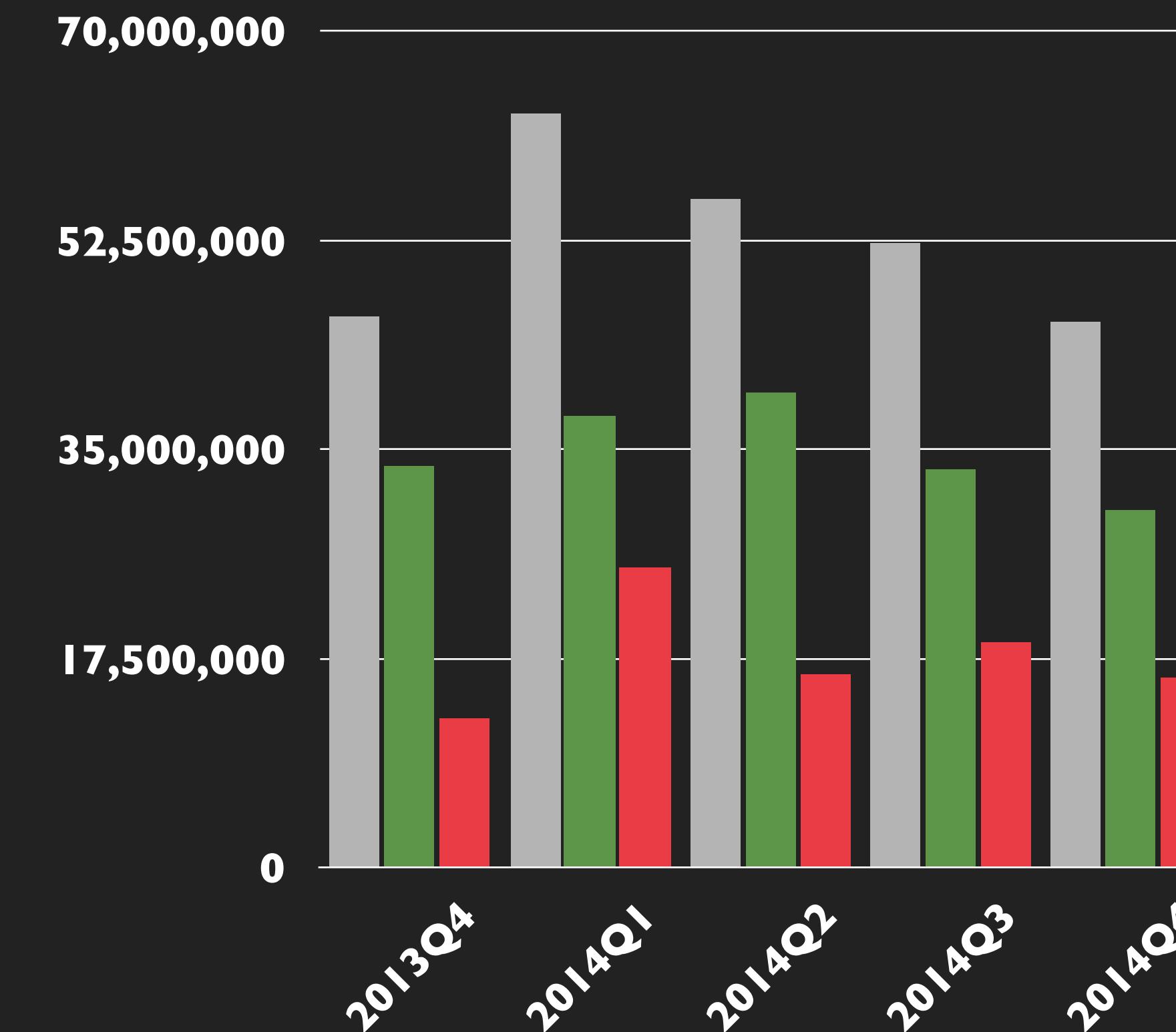
Total number of misses by count

Beat Miss No Miss/Beat



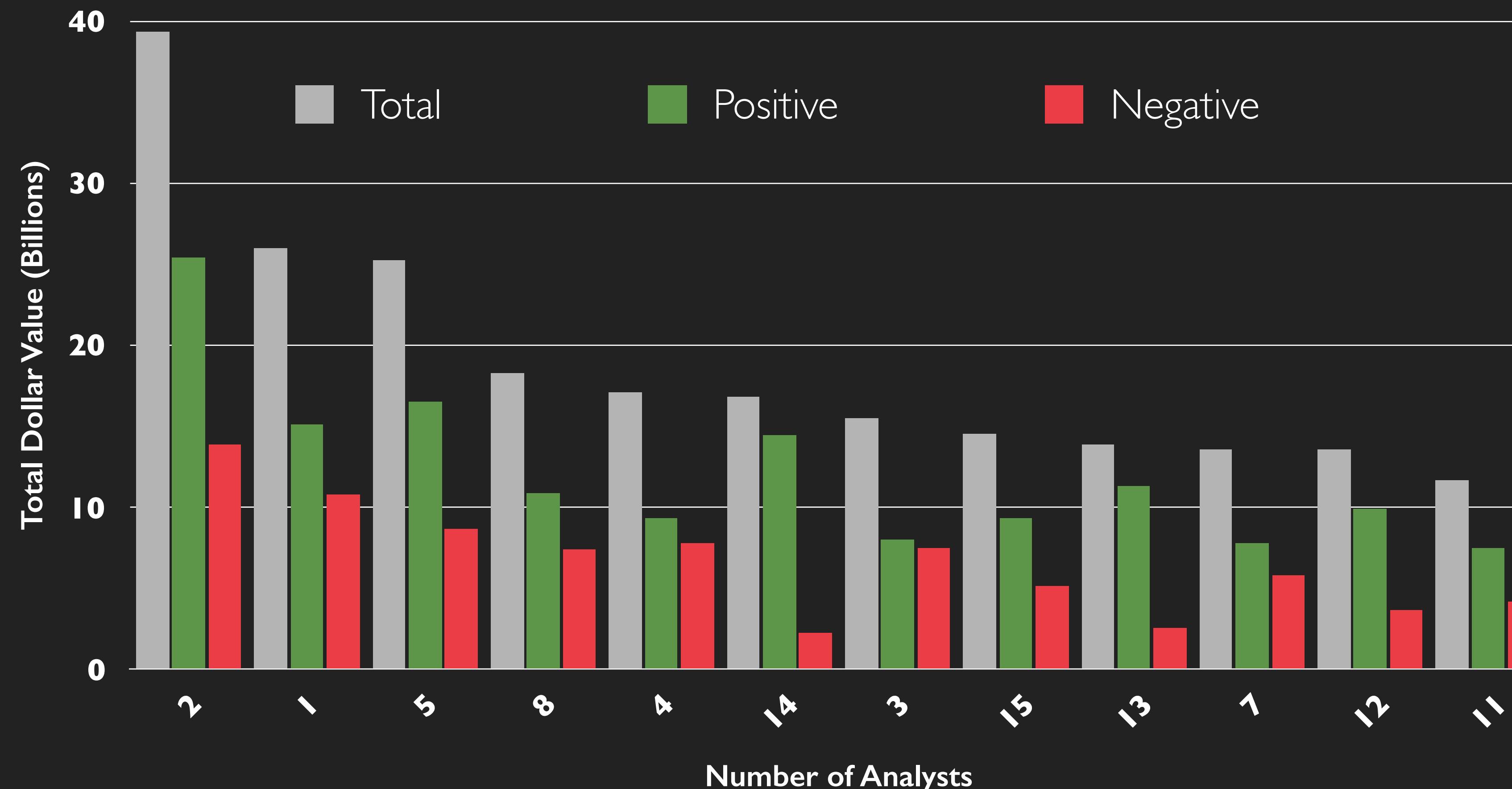
Total number of misses by USD value

Total Beat Miss



Misses/Beats by number of analysts

Aggregate value of earnings expectation misses/beats by number of analysts covering a company



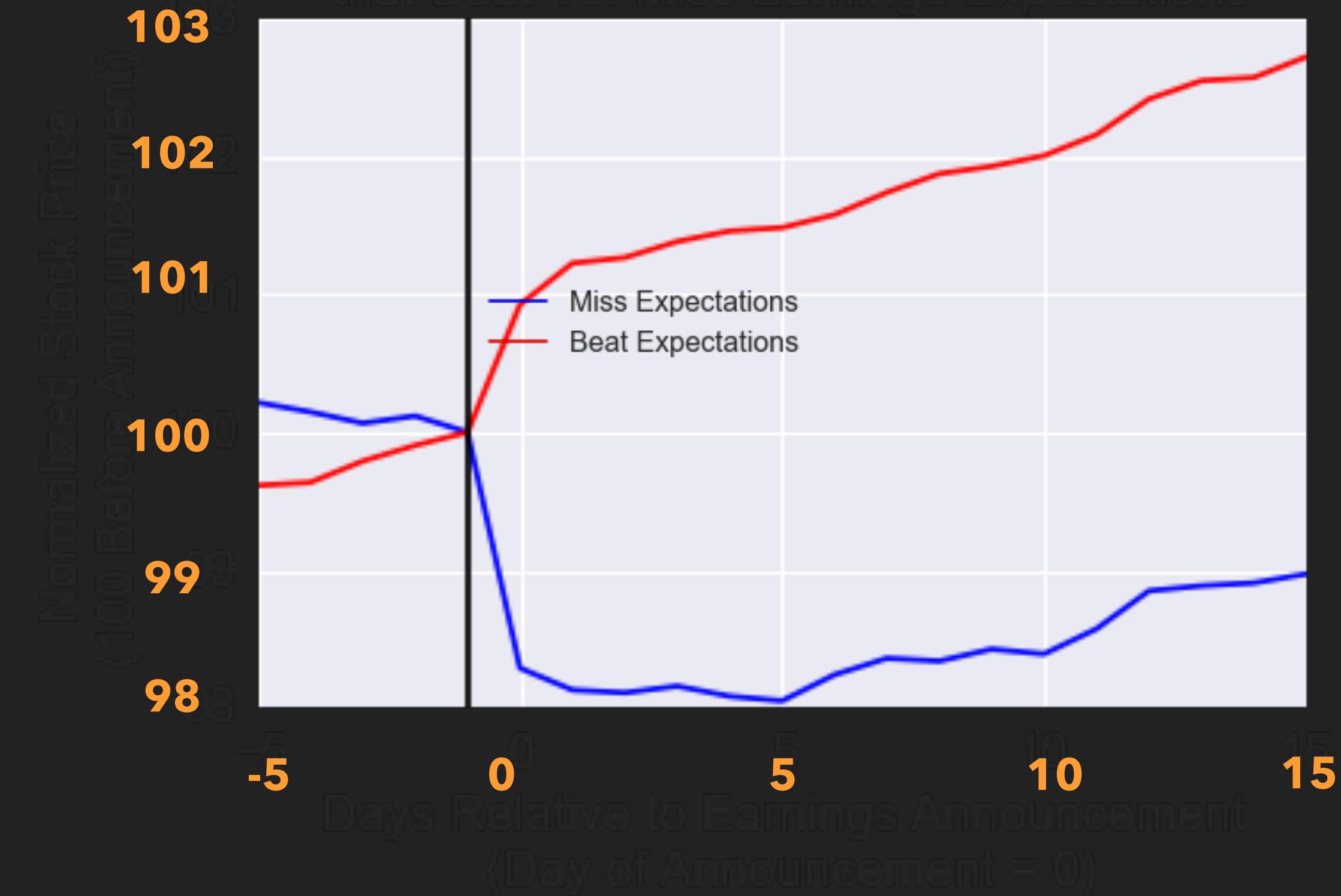
What sort of pattern to look for?: identify stocks with high probability of trading paths similar to historical benchmark we discovered plus calibration with story told by public data aggregation algorithms.

Short term
 ‘short’ trade -
max leverage with knock out
OTC put option or right vol adjusted option spread

**Normalized Stock Price
 (100 before
 announcement)**

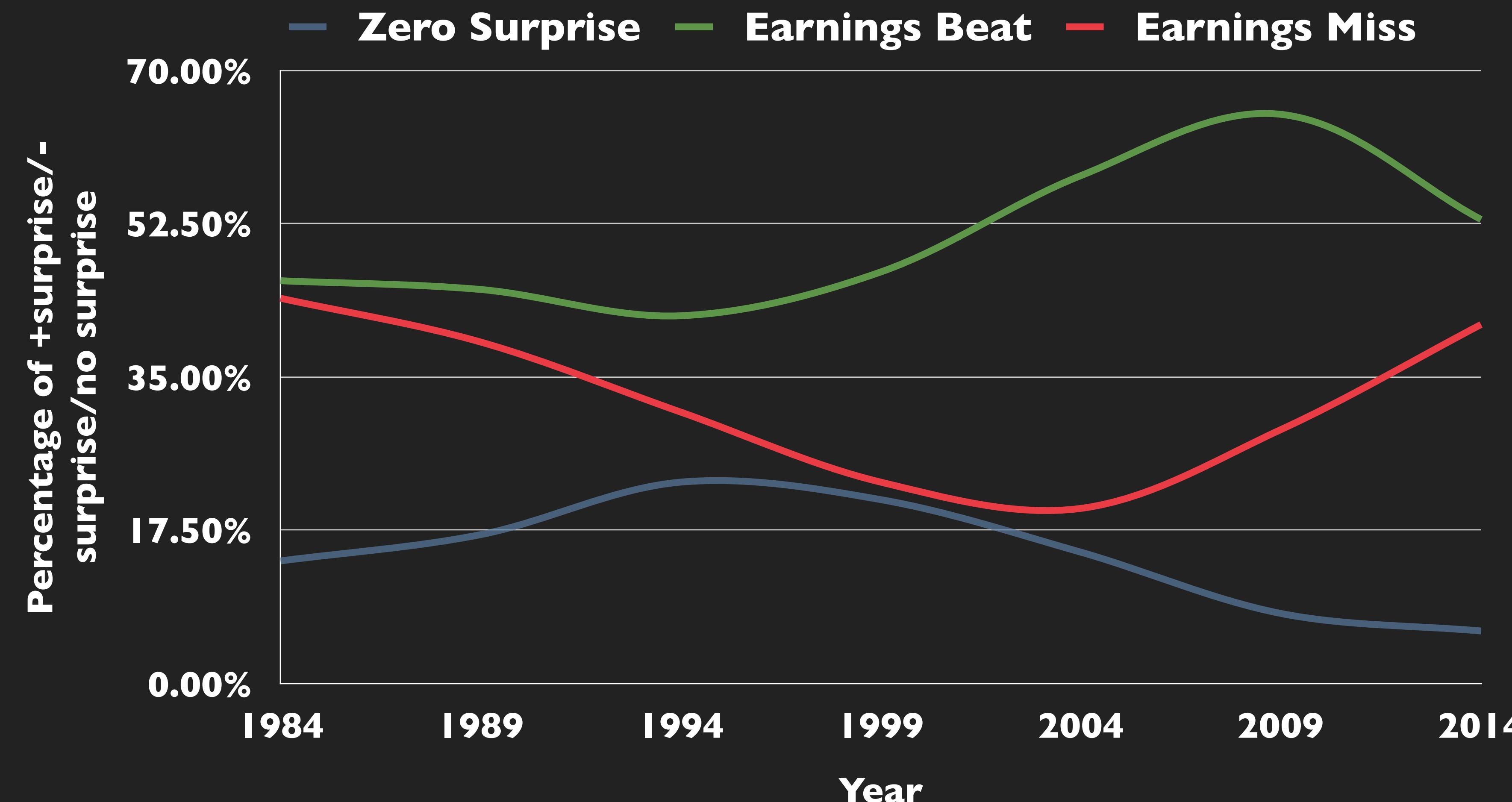
Medium term
 ‘long’ trade -
max leverage with knock out
OTC call option or right vol adjusted option spread

Median Price Response of Stocks that Miss vs Beat Earnings Expectations



With more information available, are analysts getting worse or better?

They are getting worse: with all the data available, zero surprises should be the norm, not the exception. Surprises mean opportunities for those with superior knowledge.

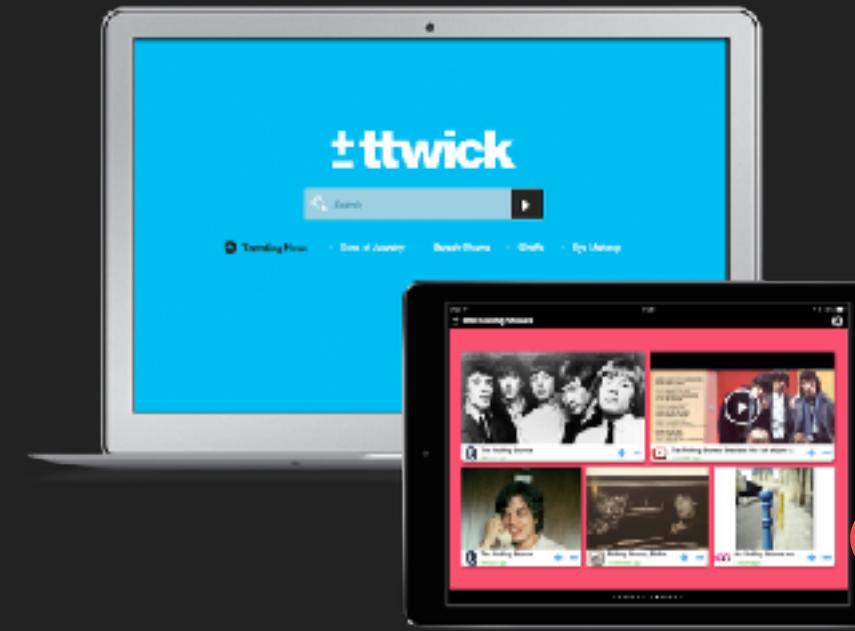


Can an ensemble of hybrid quant /data science models on top of a custom search application do a better job than Wall Street and Hedge Fund analysts?

Example: GRPN EPS Q2 2013	
Top Line Revenue:	\$608,747
Cost of Net Revenues:	(\$724,053)
Gross Profit:	\$384,694
Administrative and General Expenses:	(\$357,282)
Income from Operations:	\$27,412
Interest + other income/(expenses) + tax provision:	(\$34,986)
Profit/_Loss:	-\$7,574
Common shares outstanding	652,361
EPS	-\$0.01

← Top Line Revenue

Yes. The lucky discovery.....



ttwick
News & Social Search

- Web, iOS for iPad (February 2015)
- Facebook connect login
- Digital assets are used to build portfolios of content
- Social content, news summaries, blogs, etc.
- Computational linguistics technology, data collection and analytics with important, proven forecasting applications



ttwick deals

- Web, iOS, and Android
- Lists more deals than Groupon, Living Social, Tippr, etc. combined
- Covers deals in the US, Canada, UK, Singapore, Malaysia, Indonesia
- Valuable database of information with important financial trading implications
- Will be redesigned in 2015 to serve a more strategic market



blacksheep
SMARTER SEARCH

- AI based semantic search that provides specific answers to questions rather than forcing users to fish for information
- In alpha stage
- Initial database ingestion is Wikipedia and MusicBrainz but can be deployed in most types of unstructured or structured datasets

**SWITCH TO TABLEAU
DASHBOARD FOR DEMO**

**Non correlated datasets collected from news, social sites,
etc. plus data augmentation and selected info from
EDGAR's SEC database allowed me to validate the “Data
Investing” concept**

**Python libraries used: scrappy, pandas, numpy, scipy,
seaborn**

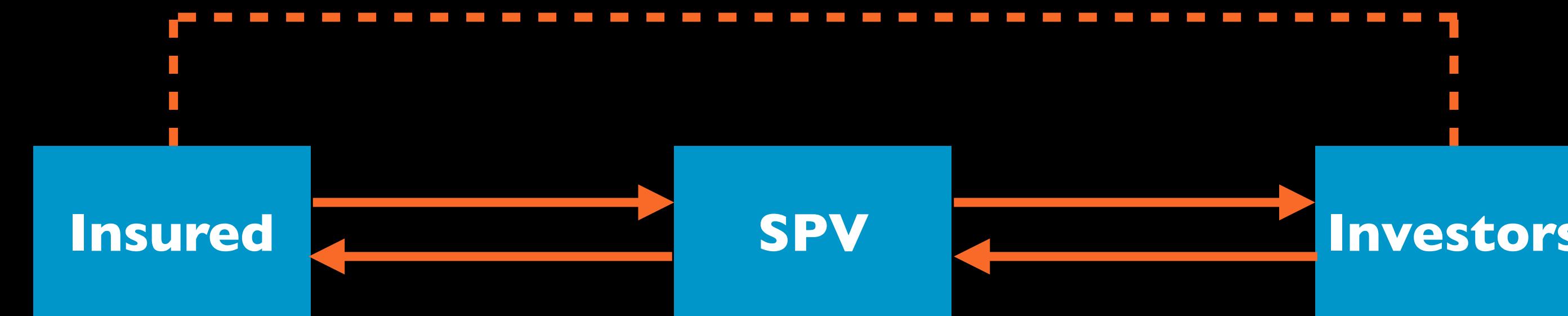
**DESIGN FINANCIAL INSTRUMENTS
TO PROTECT AGAINST FINANCIAL
IMPACT OF CATASTROPHIC NATURAL
EVENTS**

CAT Risk is a type of Insurance Linked Security (ILS)

- **ILS is a highly structured security that requires knowledge of financial quantitative analysis plus domain expertise in credit, market, and operational risk, plus engineering (mechanical, civil, etc.), time series analysis and data science.**
- **ILS returns do not have significant correlation with the market's returns.**
- **ILS coupon and principal payments depend on a pool or index linked to natural or man made catastrophic risk instead of a credit event:**
 - ✓ **Earthquakes**
 - ✓ **Hurricanes**
 - ✓ **Weather**
 - ✓ **Windstorms**
 - ✓ **Terrorist attacks**
 - ✓ **Oil spills**
 - ✓ **Satellite launches**
 - ✓ **Other**
- **Generally, deals have the principal amount plus coupon payments fully collateralized by risk free and/or highly rated securities, which reduces credit risk.**
- **Rating of the transaction is isolated from the sponsor's rating, and it is tied to the probability of occurrence of the catastrophe.**

Mechanics:

- An offshore special purpose vehicle (“SPV”) is established for the sole purpose of entering into a insurance contract with the insured.
- The SPV issues liabilities to the capital markets to collateralize its obligations under the insurance contract.
- The proceeds from the SPV liability issuance are invested in highly rated securities, i.e: US Treasuries or corporate “AAA” rated securities.
- Earnings on the SPV’s investments as well as the insurance premiums received from the insured are used to cover the coupons on the SPV’s liabilities as well as the SPV’s operating costs.



Insurance risk structures

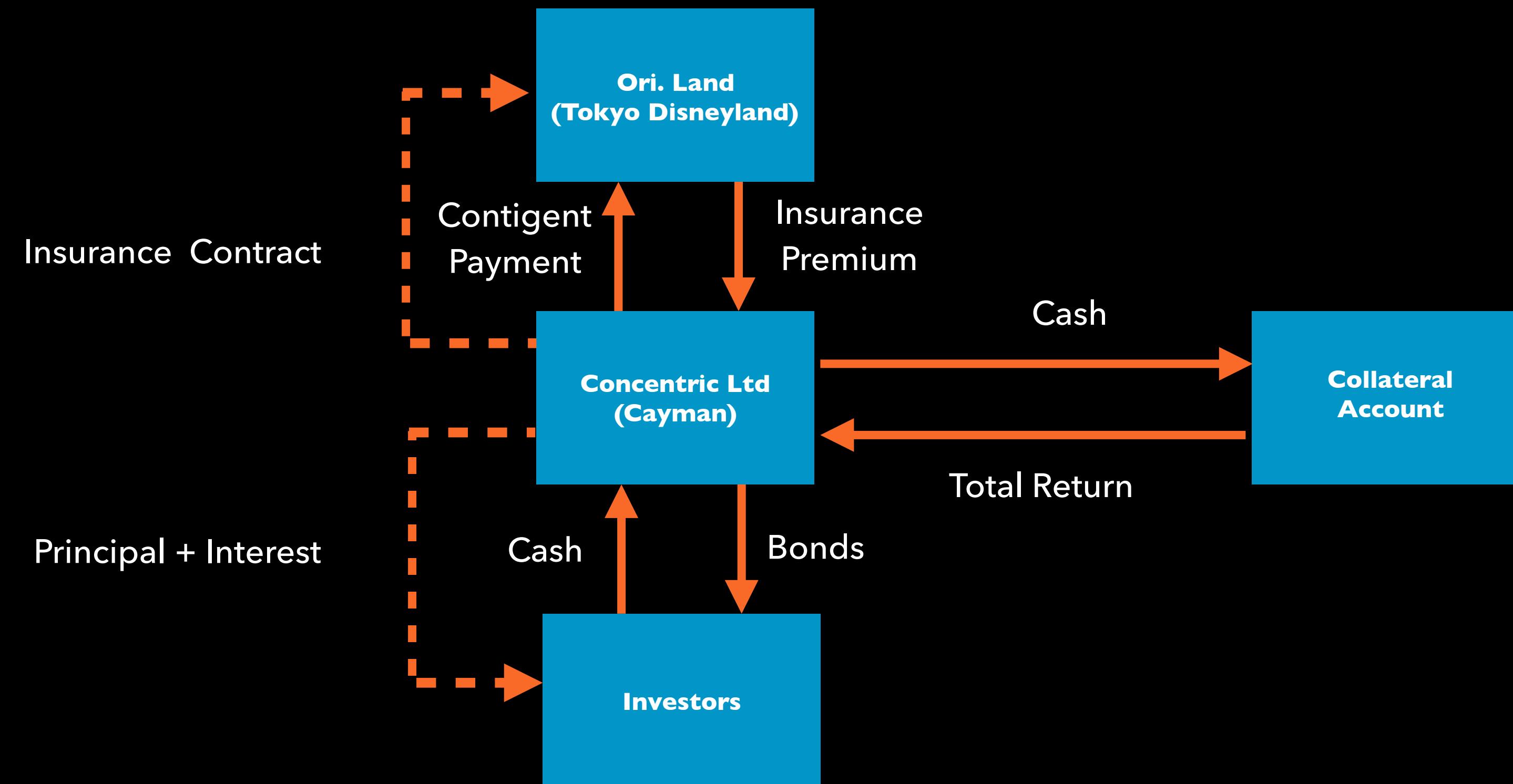
Insurance Risk Structure	Characteristics	Issuer's Concerns	Investors Concerns	Report of Losses
Parametric Structure	Bondholder losses are based on model that takes parameters from the event	Basis risk, if model does not accurately represent the losses of the sponsor	Potential financial gain for the sponsor	Losses based on parameters observed by independent reporting agencies

Parametric bonds: loss triggers

Risk	Parameter	Intermediate steps	Loss to investors occurs when
Earthquake	- Magnitude - Epicenter - Depth	- Creation of a time series earthquake risk - Selection of area - Probability of survival	- Magnitude is equal or higher than a defined trigger level - Location is within a defined area specified in the offering circular
Terrorist attacks	- Mortality risk - Landmark damages - Backstop limit	- Regulatory approval - Insurance facility	- Loss of life exceeds a pre-determined threshold - Designated landmarks are totally or partially destroyed - Monetary damage exceeds pre defined thresholds
Windstorm	- Wind speed	- Creation of a time series wind indices	- Location is within a defined area specified in the offering circular
Temperature	- Accumulated temperature in a defined region and period of time	- Creation of a CDD and HDD time series indices	- HDD or CDD value is greater or lower than a predetermined value ("Strike")

Real life example: Monetization of earthquake data via structured finance using proprietary analytical models

- First ever data/securitization product of this type: bypassed completely the insurance industry.
- Cayman Islands' SPV set by investment banking team with feedback from Front Office Quants (FOQ)¹

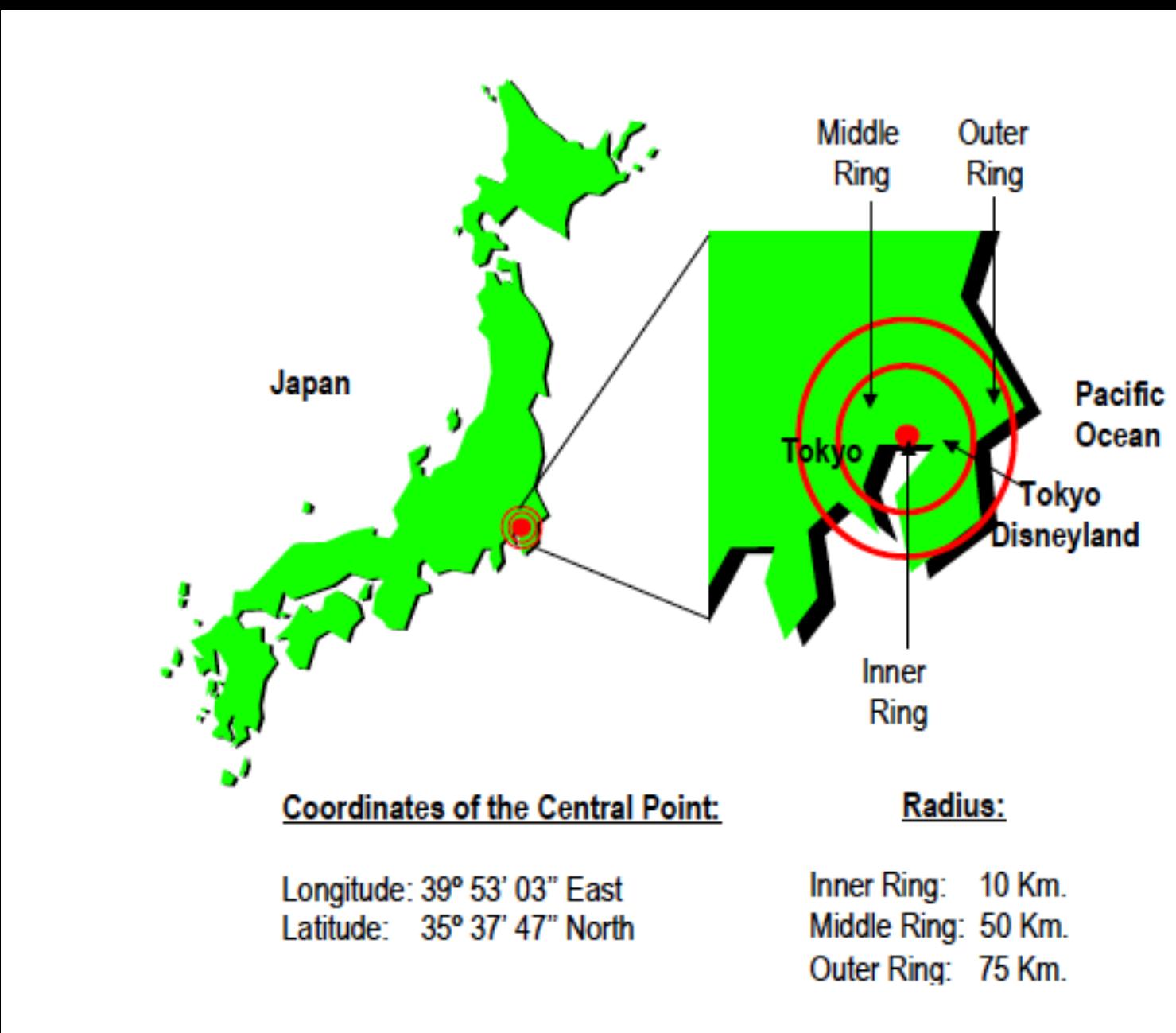


¹ http://en.wikipedia.org/wiki/Quantitative_analyst#Front_office_quantitative_analyst

Real life example: Monetization of earthquake data via structured finance using proprietary analytical models

- Total issuance amount was \$100 million and obtained a Ba1 rating from Moody's, and a BB+ rating from Standard and Poor's.
- Bonds paid 6m LIBOR + 310 bps.

Geographical Parameters



Bond Loss Parameters

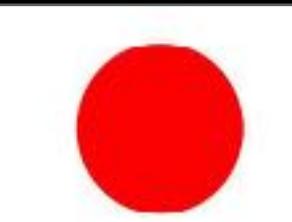
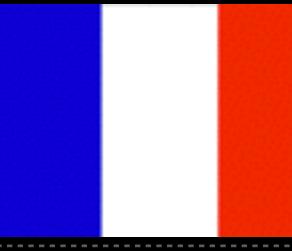
Richter Magnitude measured by the JMA	If the epicenter is in the Inner Ring, the loss to investors is (%):	If the epicenter is in the Middle Ring, the loss to investors is (%):	If the epicenter is in the Outer Ring, the loss to investors is (%):
6.5	25.0	0.0	0.0
6.6	32.5	0.0	0.0
6.7	40.0	0.0	0.0
6.8	47.5	0.0	0.0
6.9	55.0	0.0	0.0
7.0	62.5	0.0	0.0
7.1	70.0	25.0	0.0
7.2	77.5	37.5	0.0
7.3	85.0	50.0	0.0
7.4	92.5	62.5	0.0
7.5	100.0	75.0	0.0
7.6	100.0	87.5	25.0
7.7	100.0	100.0	50.0
7.8	100.0	100.0	75.0
7.9 or greater	100.0	100.0	100.0

SWITCH TO TABLEAU

DASHBOARD

<https://public.tableau.com/profile/catrisk/vizhome/CatRisk/OverviewoftheCatRiskMarket>

Some reporting agencies for trigger events

Country	Agency	Risk
	- Japan Meteorological Agency (JMA)	- Earthquakes, Storms, Typhoons
	- Météo-France	- Windstorm, Temperature
	- The Meteorological Office	- Windstorm, Temperature
	- Deutscher Wetterdienst	- Windstorm, Temperature
	- Royal Meteorological Institute of Belgium	- Windstorm, Temperature
	- Royal Netherlands Meteorological Institute	- Windstorm, Temperature
	- National Hurricane Center (NHC) - US Geological Survey (USGS)	- Hurricane, Windstorms - Earthquakes

Some Asian sovereigns with high cat risk exposure



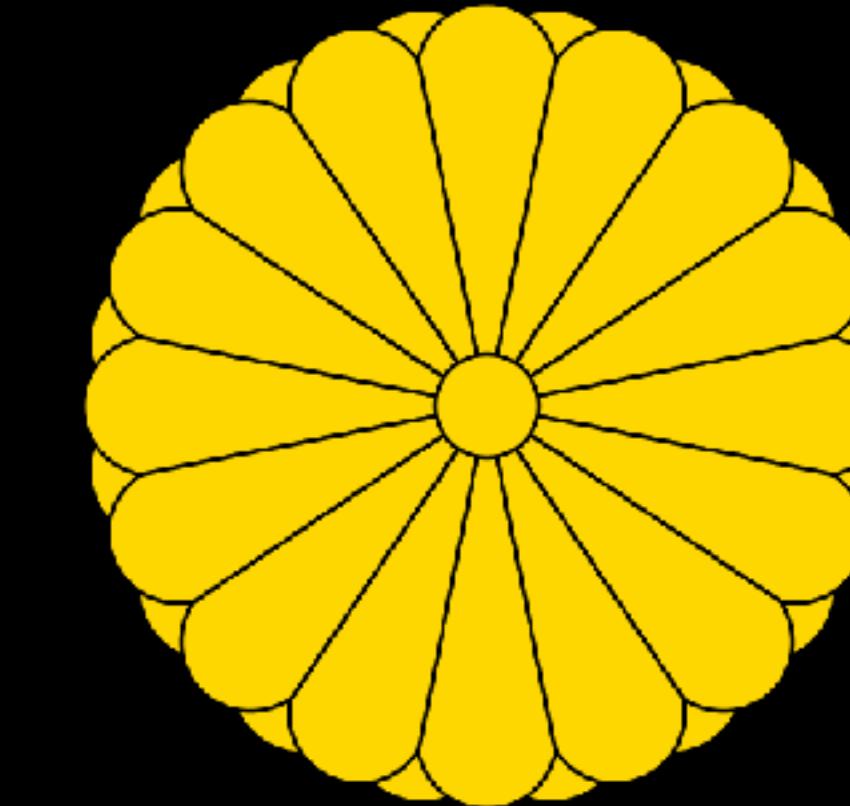
Indonesia



China



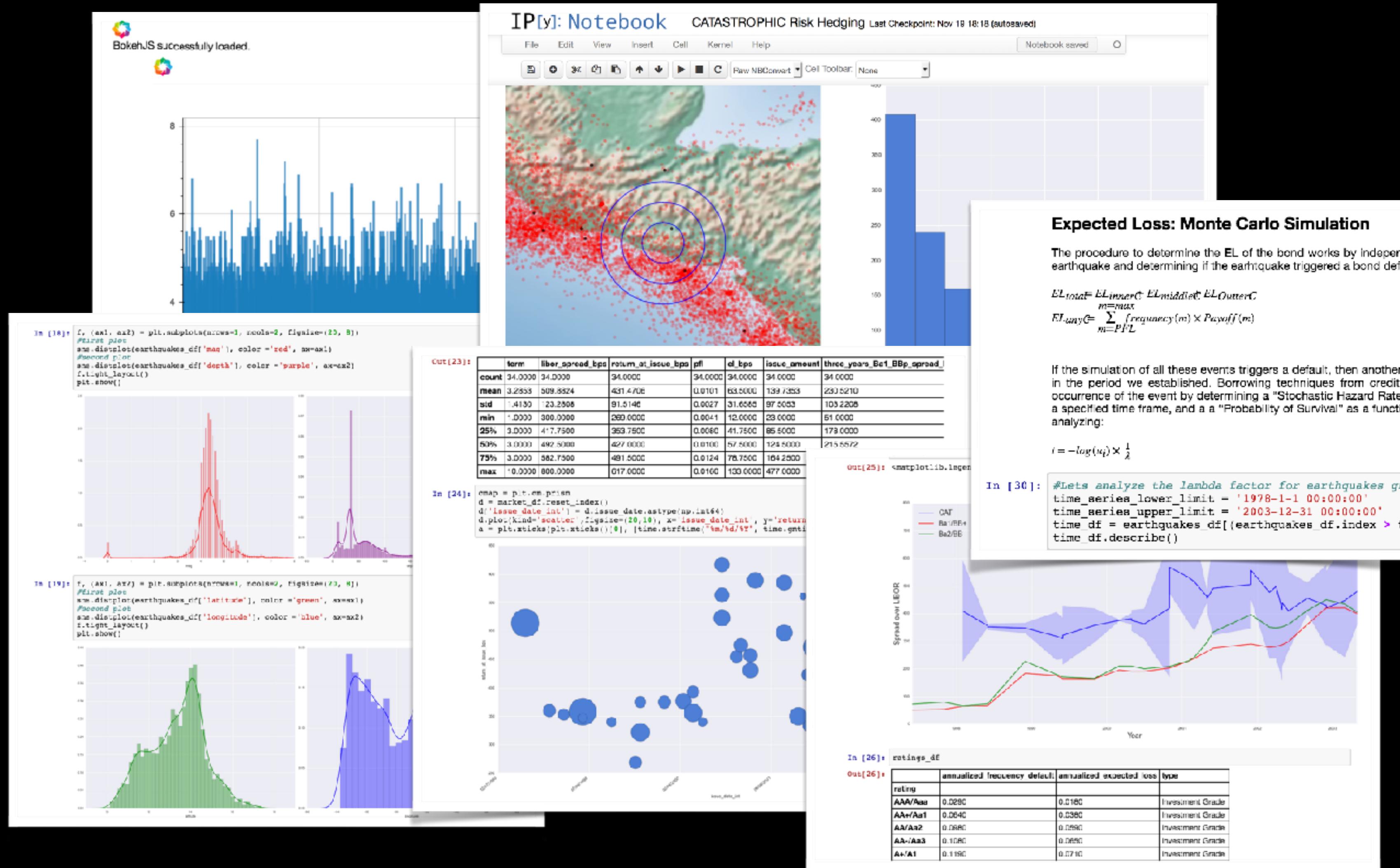
India



Japan

Some multilateral agencies with significant cat risk exposure





Non correlated natural events such as floods, temperature fluctuation, earthquakes, hurricanes, etc. present a very interesting asset class with a great deal of funds available for exposure in th

Python libraries used: beautiful soup, pandas, numpy, scipy, seaborn, bokeh, scikit learn

FOR MORE INFORMATION:

<http://catrisk.com/>

CRYPTO CURRENCY MINING AND TRADING



“BITCOIN IS BETTER THAN CURRENCY IN THAT YOU DON’T HAVE TO BE PHYSICALLY IN THE SAME PLACE AND, OF COURSE, FOR LARGE TRANSACTIONS, CURRENCY CAN GET PRETTY INCONVENIENT... IN THE FUTURE, FINANCIAL TRANSACTIONS WILL EVENTUALLY BE DIGITAL, UNIVERSAL AND ALMOST FREE.”

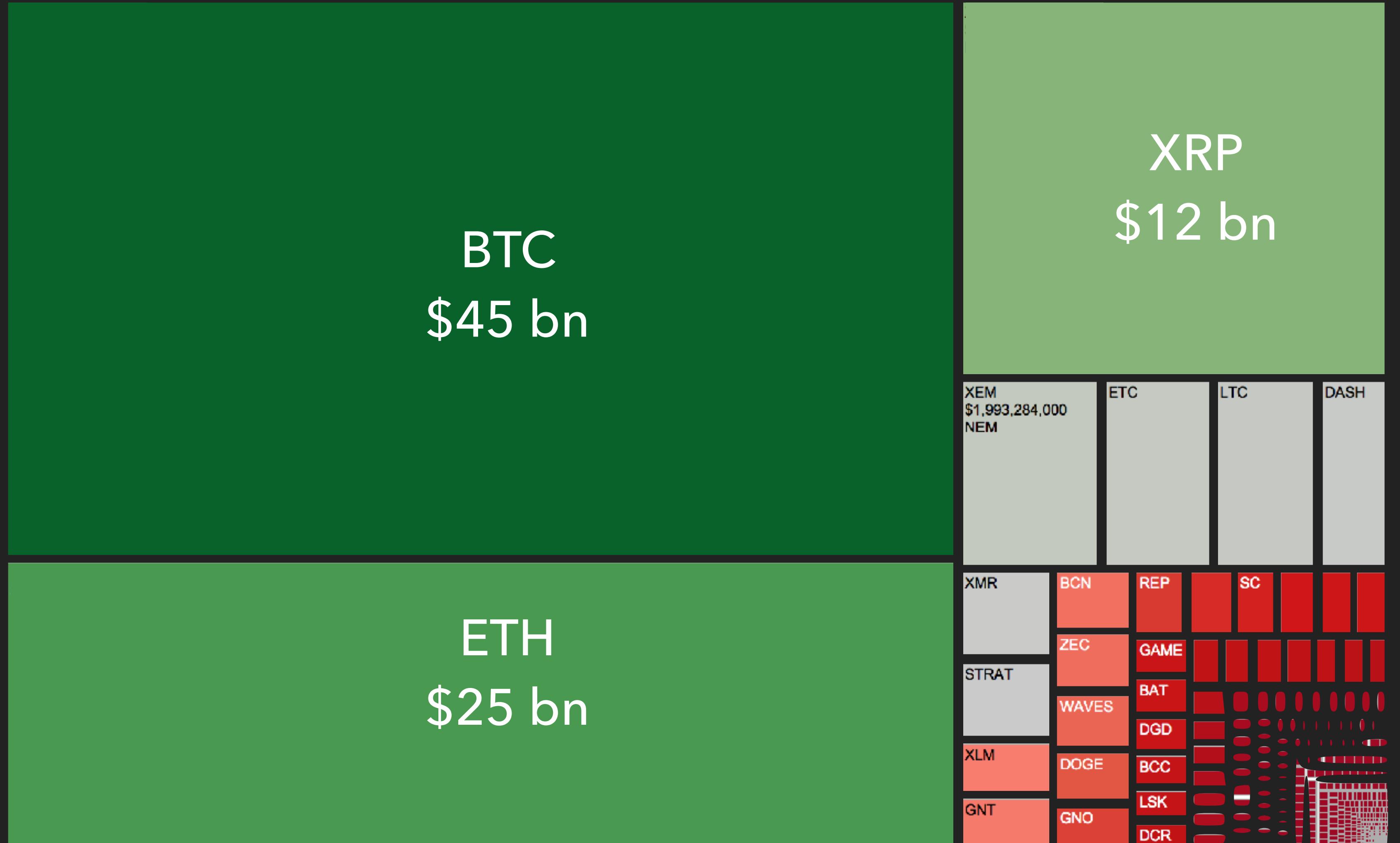
Bill Gates

CRYPTO CURRENCIES AS A TRADEABLE ASSET CLASS

- ▶ Bitcoin is not the only tradable crypto currency, but it was the first one
- ▶ Currently about 800 crypto currencies with unique features
- ▶ Asset class has shown tremendous growth with low correlation with other asset classes
- ▶ High volatility
- ▶ Asset class returns show correlation to electricity prices measured in kilowatts per hour

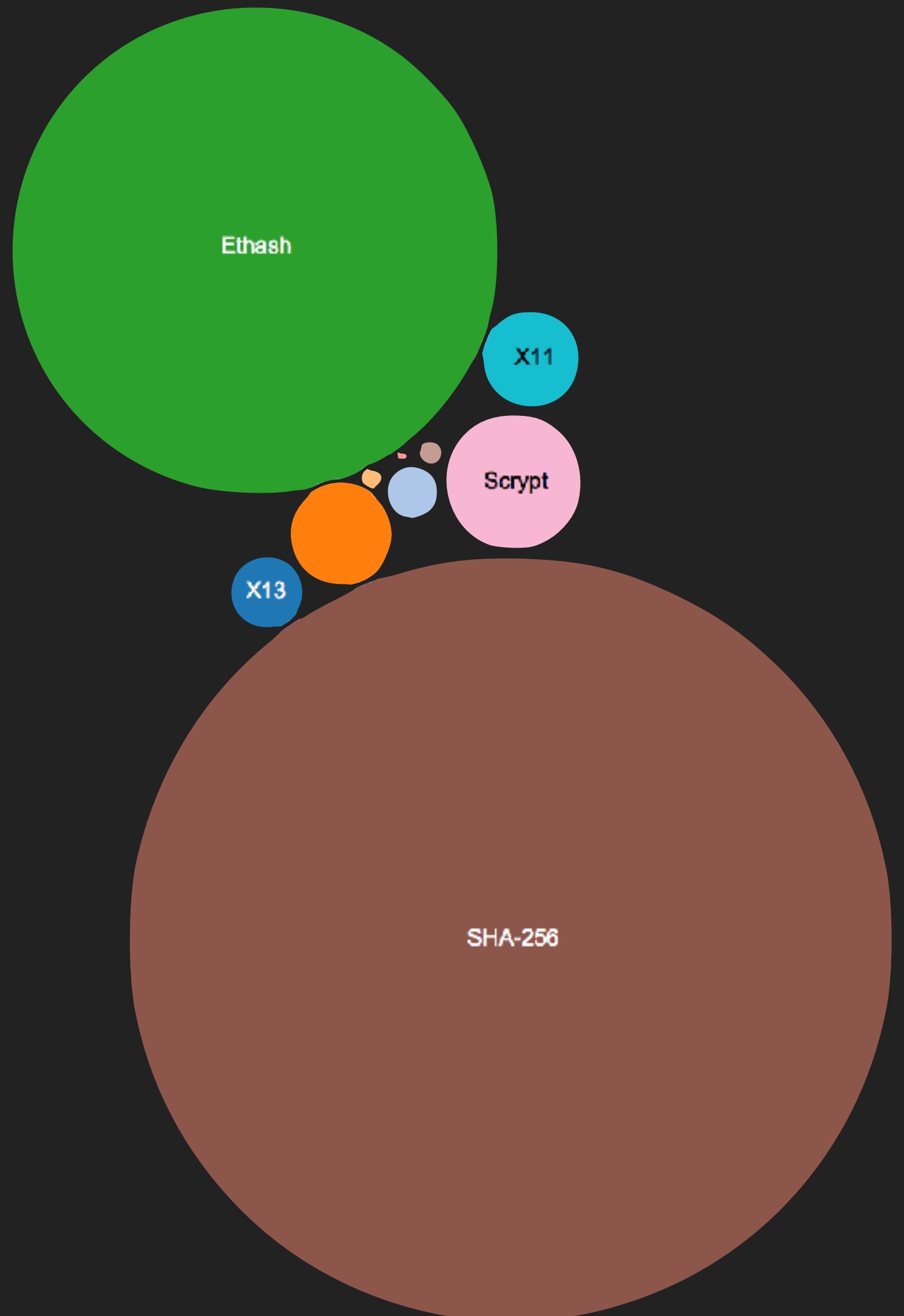


CRYPTO CURRENCIES AS A TRADABLE ASSET CLASS



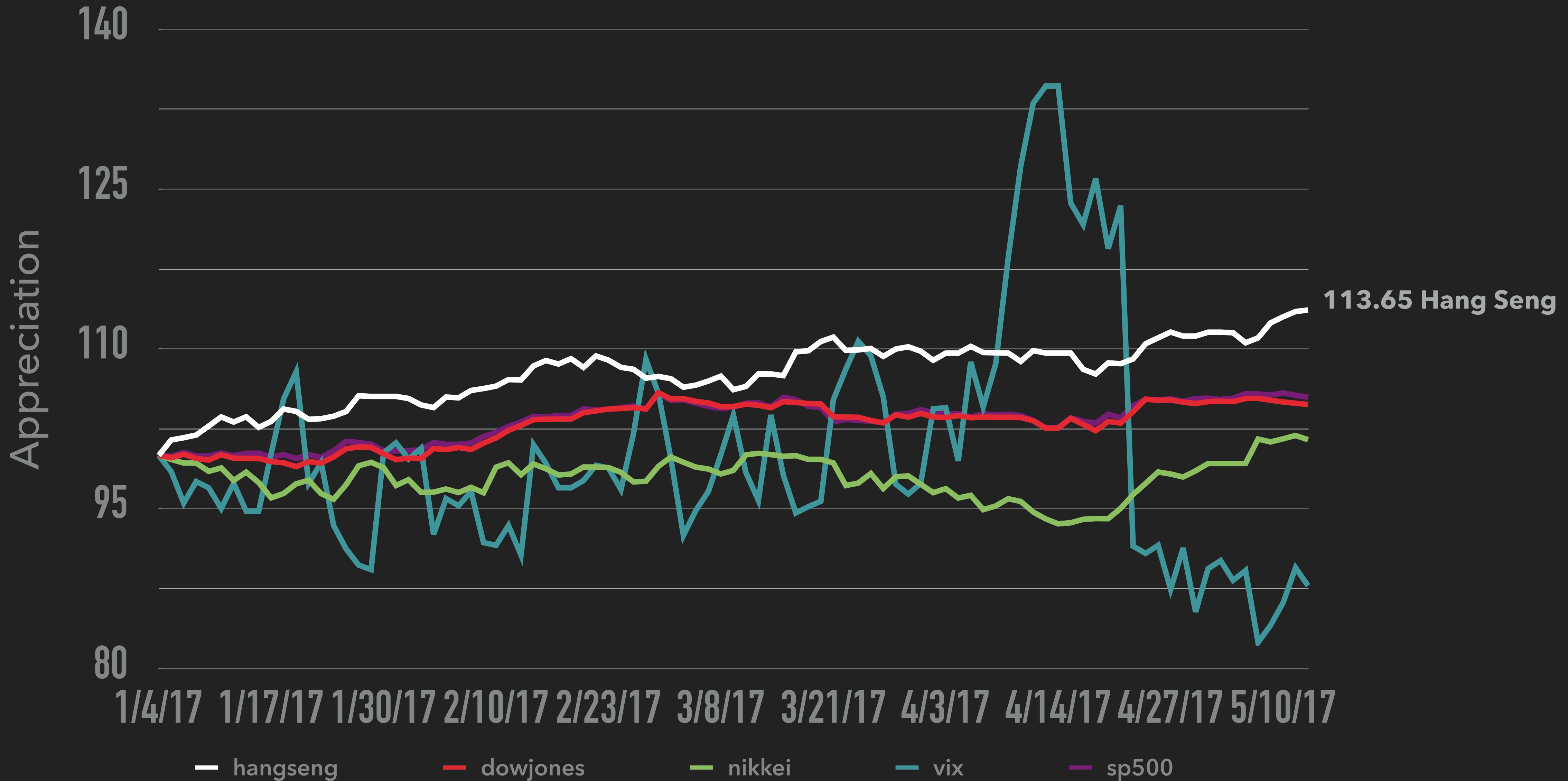
- ▶ ~\$100 bn USD total market cap as of June 12, 2017
- ▶ SGX defines tradable classes as those with market capitalizations larger than 100 MM USD as of June 12, 2017
- ▶ BTC, ETH, and XRP account for over 80% of the total market cap

HASHING ALGORITHMS

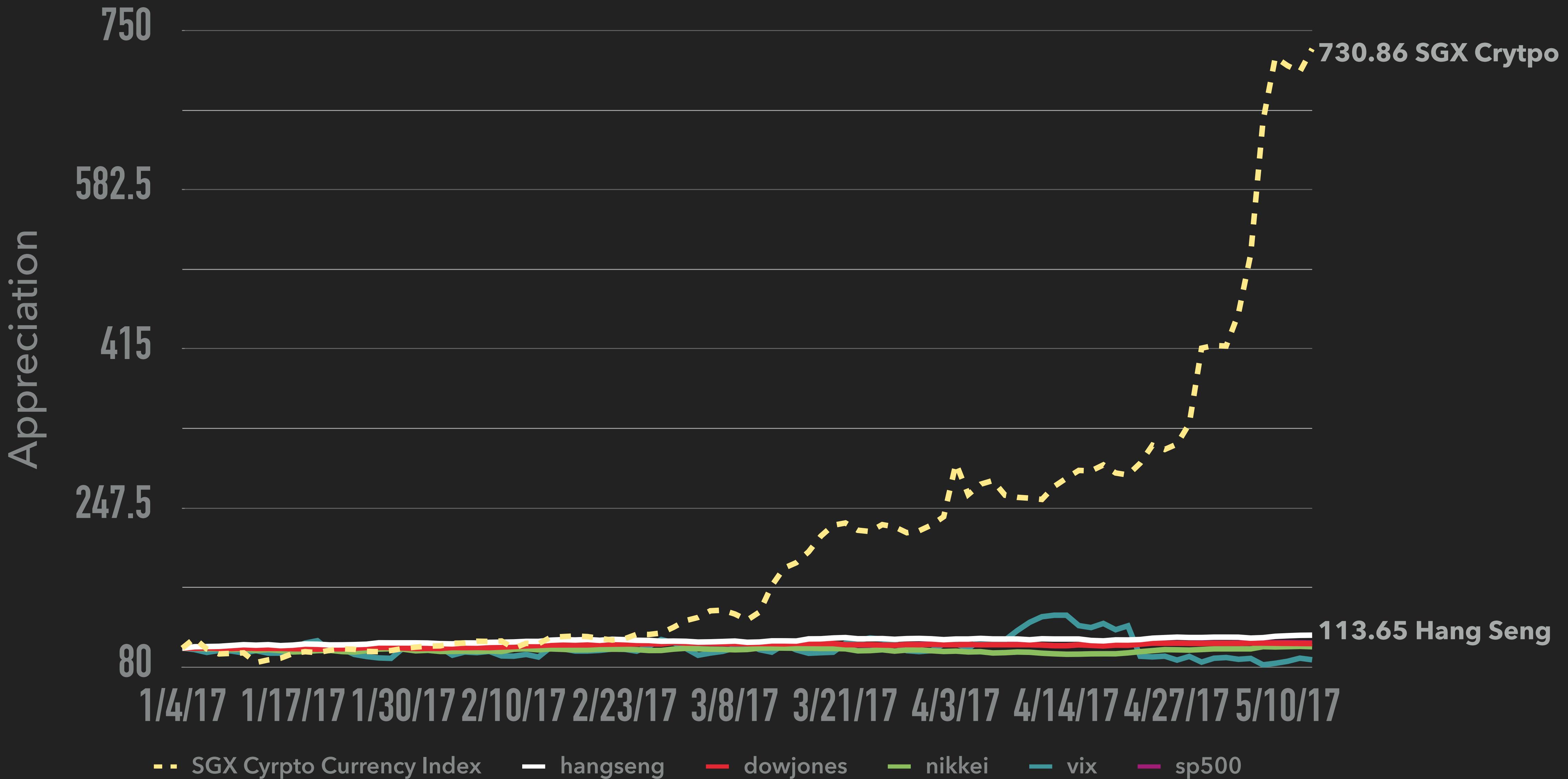


- ▶ SHA-256 and ETHash take more than 90% of the market cap
- ▶ ETHash seems to be gaining more momentum than any other algorithms

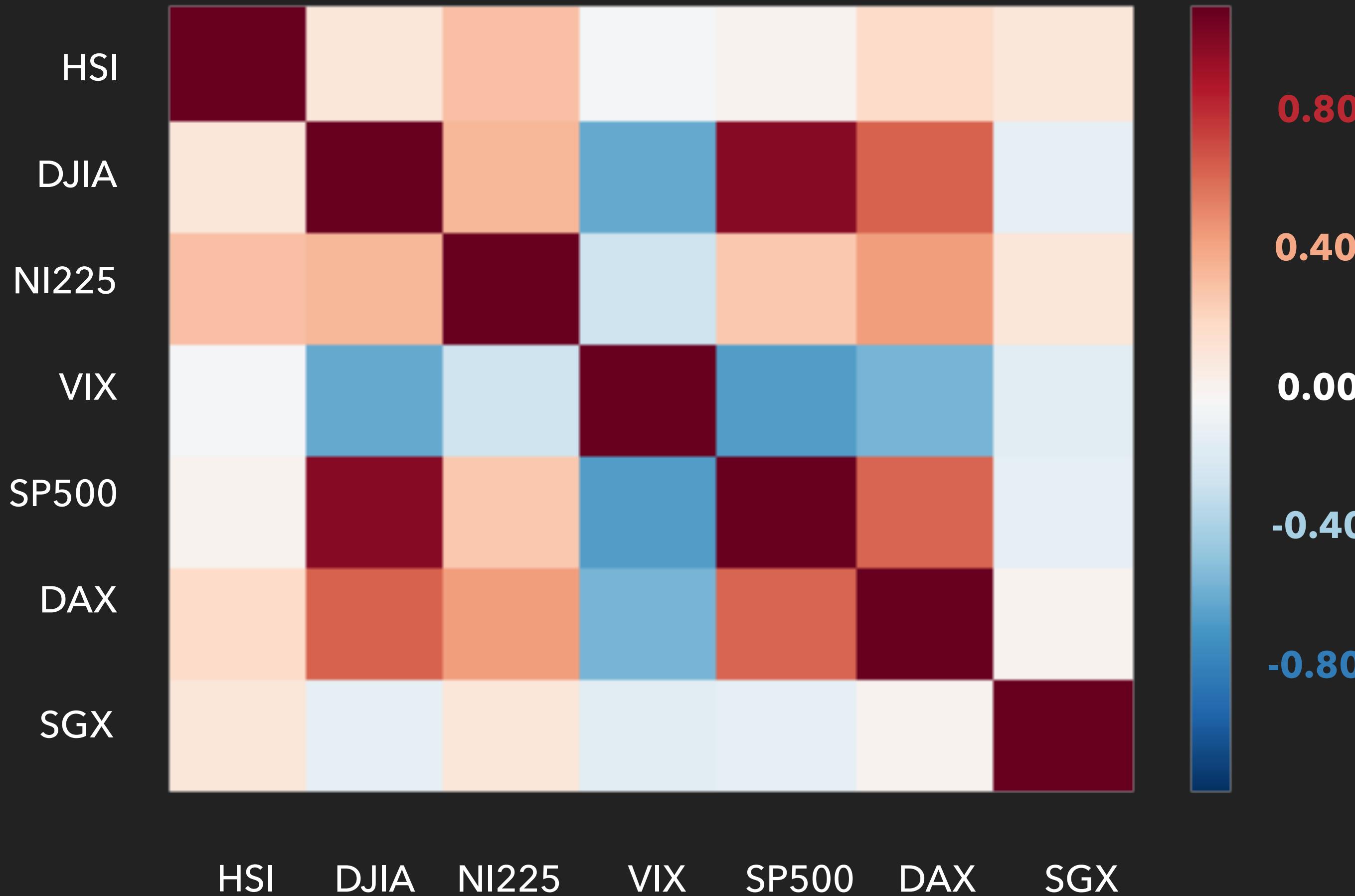
EQUITY INDICES VS SGX CRYPTO INDEX APPRECIATION YTD



EQUITY INDICES VS SGX CRYPTO INDEX APPRECIATION YTD

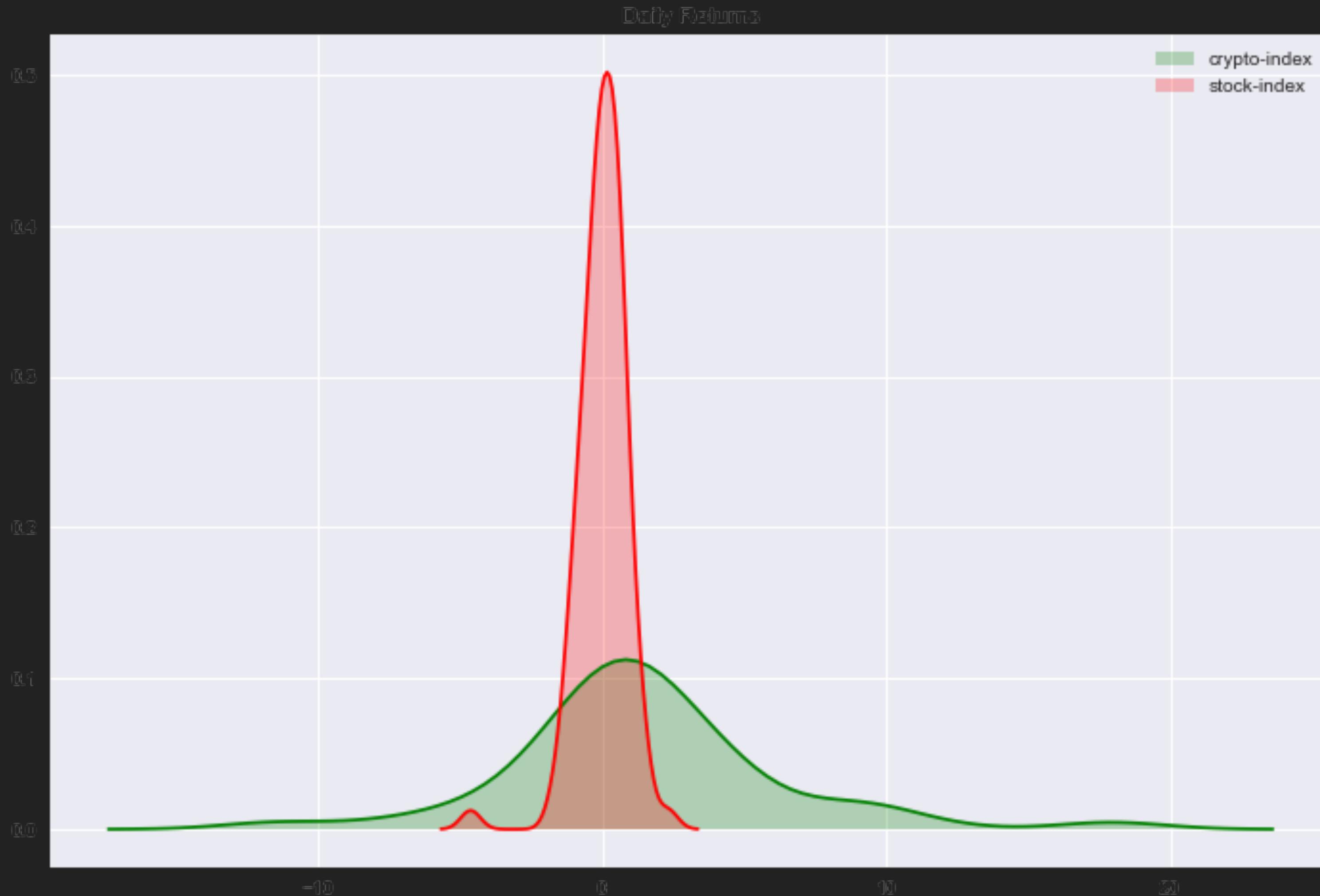


SGX CRYPTO INDEX



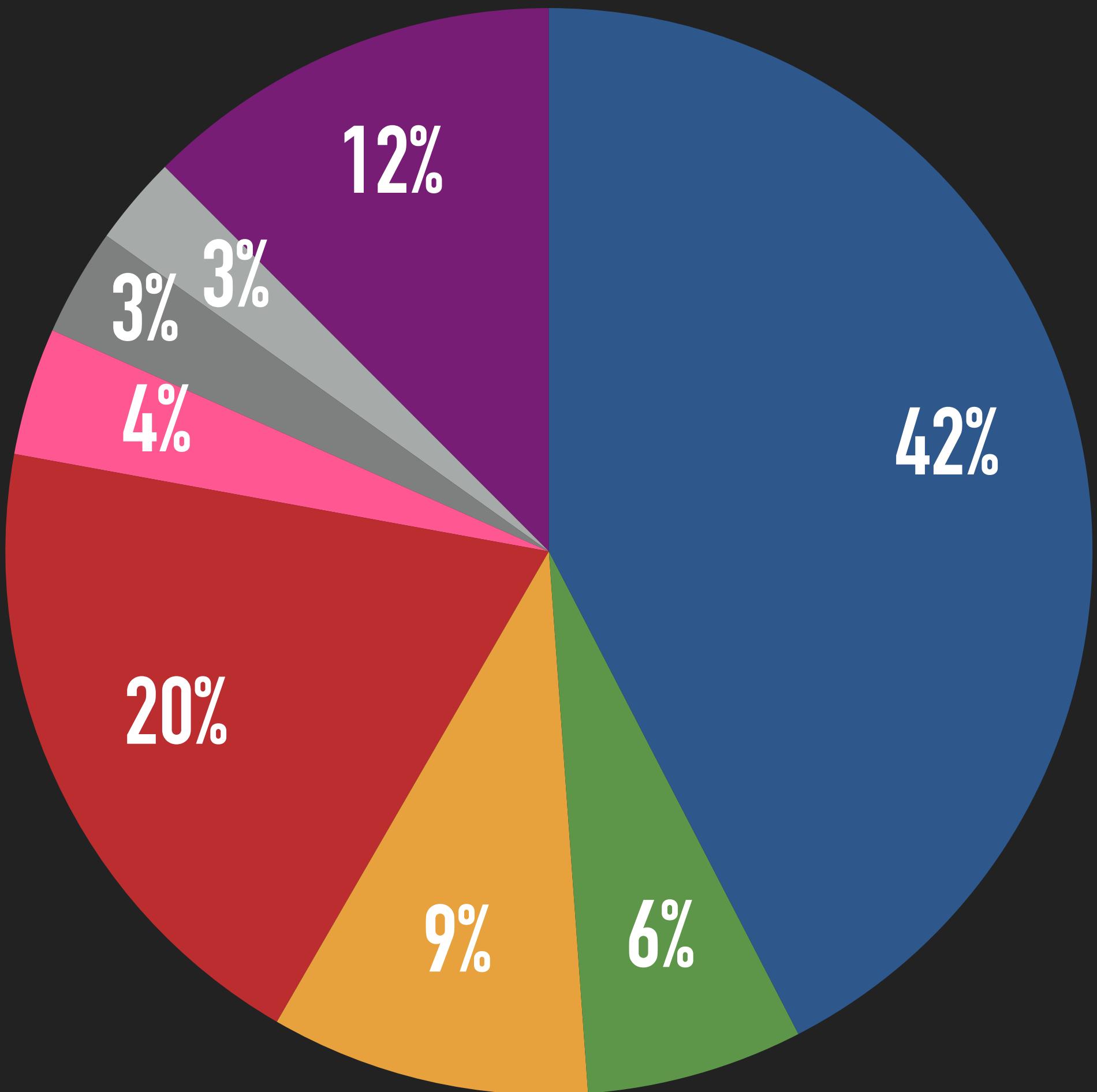
- ▶ Total daily trading volume of about ~4.0 bn USD
- ▶ Low or zero correlation with other asset classes

SGX CRYPTO INDEX



- ▶ High annualized volatility, around 112%
- ▶ Possible mean reversion
- ▶ Stock indices returns are p25, p50, p75, -0.44%, 0.04%, 0.54%
- ▶ SGX Crypto index p25, p50, p75 are -0.42%, 0.98%, 3.39%

FOREX MARKET VS SGX CRYPTO INDEX



- ▶ The FX OTC market is about 5.3 trillion USD per day in trading volume, dwarfing the equities and futures markets
- ▶ The USD is the most traded currency
- ▶ Only 1% allocation of the FX OTC market to crypto currencies over the next few years would represent over 12x change in volume, from the current 4.0bn/day to over 50bn/day.

● USD

● GBP

● JPY

● EUR

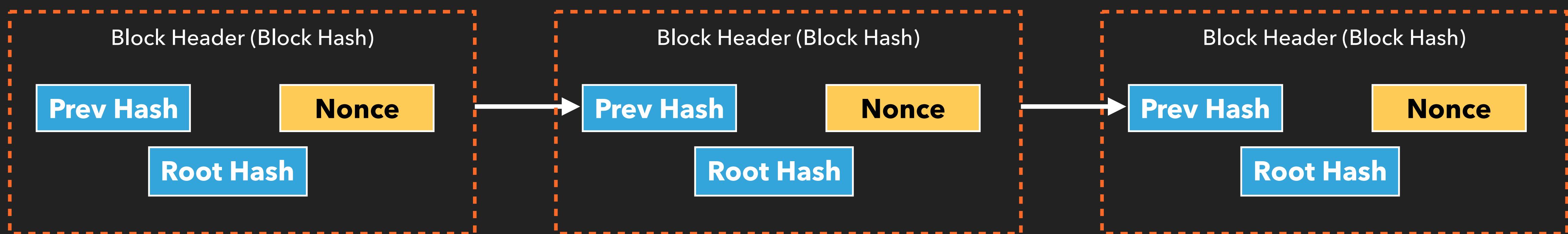
● AUD

● CHF

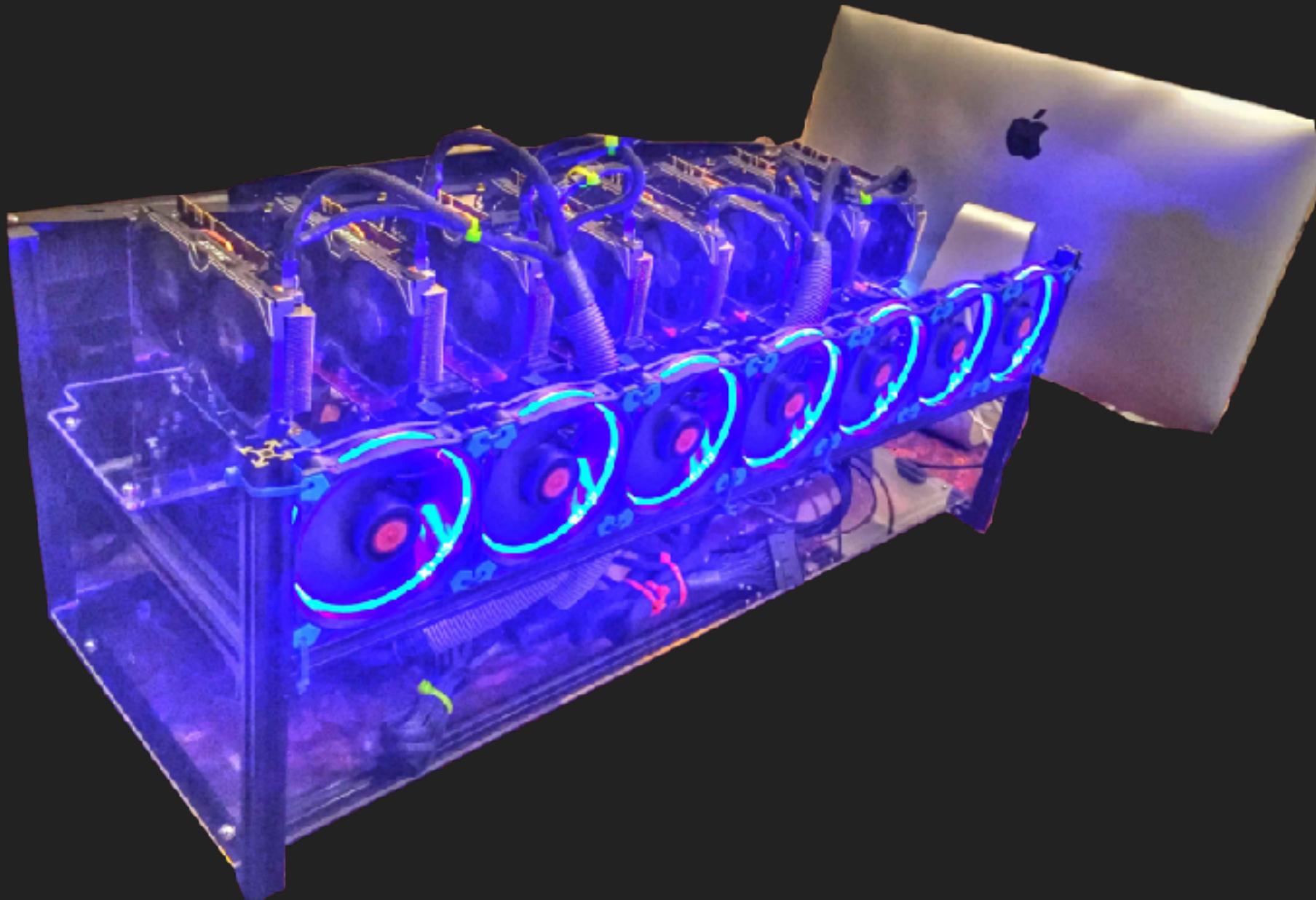
● CAD

● Others

HASH CHAINING



SGX PROPRIETARY MINING RIG - SGX1



- ▶ Custom hashing and mining code written by SGX Analytics
- ▶ Short payback period
- ▶ SGX001 mines multiple coins and proprietary code developed in Python trades coins at favorable times
- ▶ Uses probabilistic graphical models and times series analysis for best coins to hold over the long run

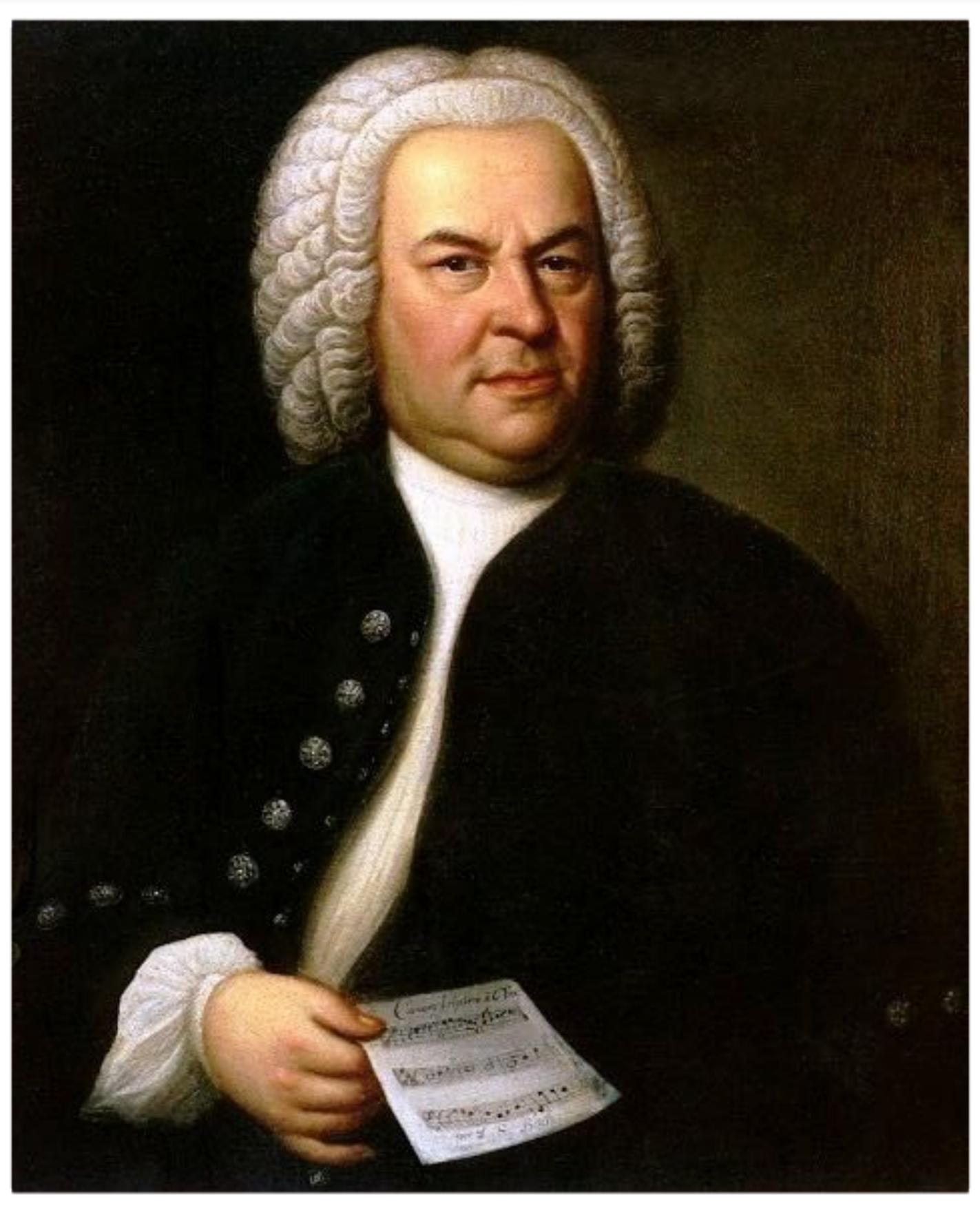
Miners		Active Devices		
Active	Total	GPU	FPGA	ASIC
2	2	12	0	0
Exchange Rate		Revenue		
Bitcoin		Day	Month	
\$4,609.50		\$29.81	\$903.25	
Coins				
Unspecified Groestl	1 active	282.72 MH/s (282.72 MH/s)	\$24.40	
Ethereum (ETH)	1 active	78.88 MH/s (78.88 MH/s)	\$5.41	
Algorithms				
Groestl	1 active	282.72 MH/s (282.72 MH/s)	\$24.40	
Ethereum	1 active	78.88 MH/s (78.88 MH/s)	\$5.41	

Crypto currencies is a new, emerging asset class with over 150 billion of USD in market capitalization and starting to gain attention of institutional investors. It is one of the best opportunities for Python coders with expertise and background in numerical analysis, cuda programming, hashing and Cython

Python libraries used: pandas, numpy, scipy, seaborn, bokeh, scikit learn, custom built NN, prophet, cython

MUSIC COMPOSITION

PARTICULAR INTEREST: JOHANN SEBASTIAN BACH'S GOLDBERG VARIATIONS



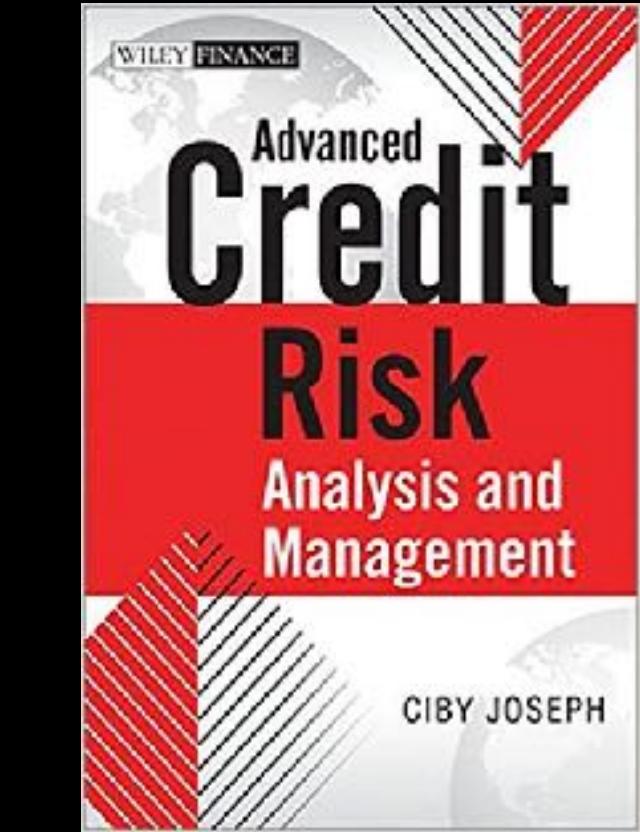
- Back in 1977 as a 12 year old kid, I fell in love with “The Goldberg Variations”
- The Goldberg Variations is a compendium written for harpsichord by Johann Sebastian Bach.
- The work is considered to be one of the most important examples of variation form in classical music.
- Rather unusually for Bach's works, the Goldberg Variations were published in his own lifetime, in 1741.

WHY IS IT RARE & VALUABLE?

- As an “Intellectual Property” asset, it is rare:
- It consists of an aria and a set of 30 variations, with 0 probability of Bach creating another one (he is dead).
- The Variations are named after Johann Gottlieb Goldberg, who may have been the first and best performer (also dead)
- The best performer in recent times was probably Glenn Gould one of the best-known and most celebrated classical pianists of the 20th century.
- Let's listen to Glenn Gould.



CREATIVE ARTIFICIAL INTELLIGENCE MUSIC COMPOSITION



David Cope (born May 17, 1941 in San Francisco, California) is an American author, composer, programmer, and professor of music at the University of California, Santa Cruz. His primary area of research involves artificial intelligence and creativity with focus in music; he codes in LISP and has developed algorithms that can analyze existing music and create new compositions in the style of the original input music. He taught a summer Workshop in Artificial Intelligence & Music that was only opened to enrolled music PhD students at UCSC, using LISP as the main tool.



An exception to the music PhD background was granted to me, who then used Python and iPython/Jupyter notebooks plus a lot of Data Science, Graph Theory, and algorithms from Credit Risk Analysis to bring algorithmic music composition to a whole new level by cross-pollinating ideas from different knowledge domains.

SONIFICATION: MY EARLY EXPERIMENTS CREATING SOUND

Sonification is the use of non-speech audio to convey information or perceptualize data. Auditory perception has advantages in temporal, spatial, amplitude, and frequency resolution that open possibilities as an alternative or complement to visualization techniques.

Steps:

Find dataset

Normalize dataset to piano keyboard

Play it

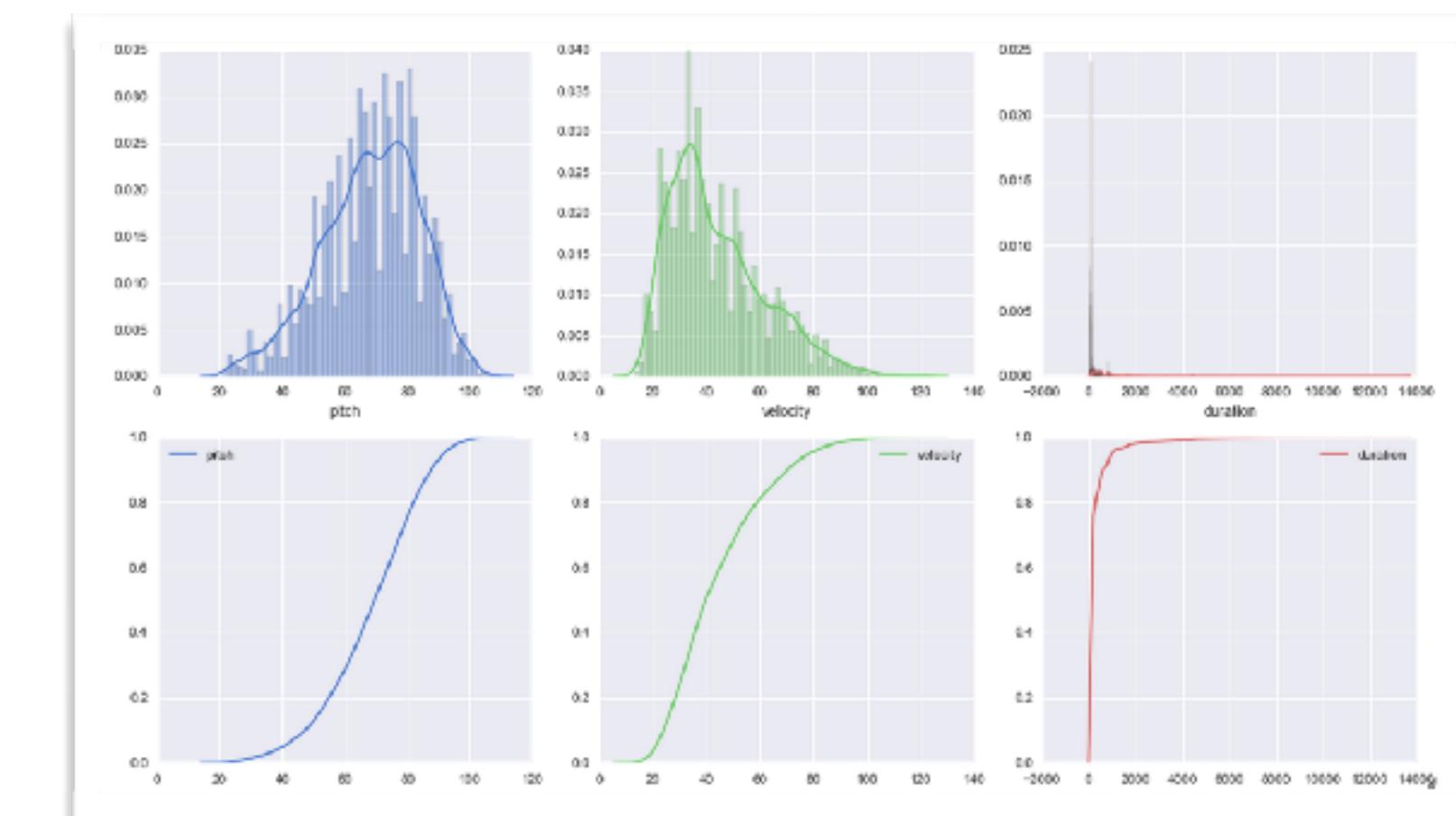
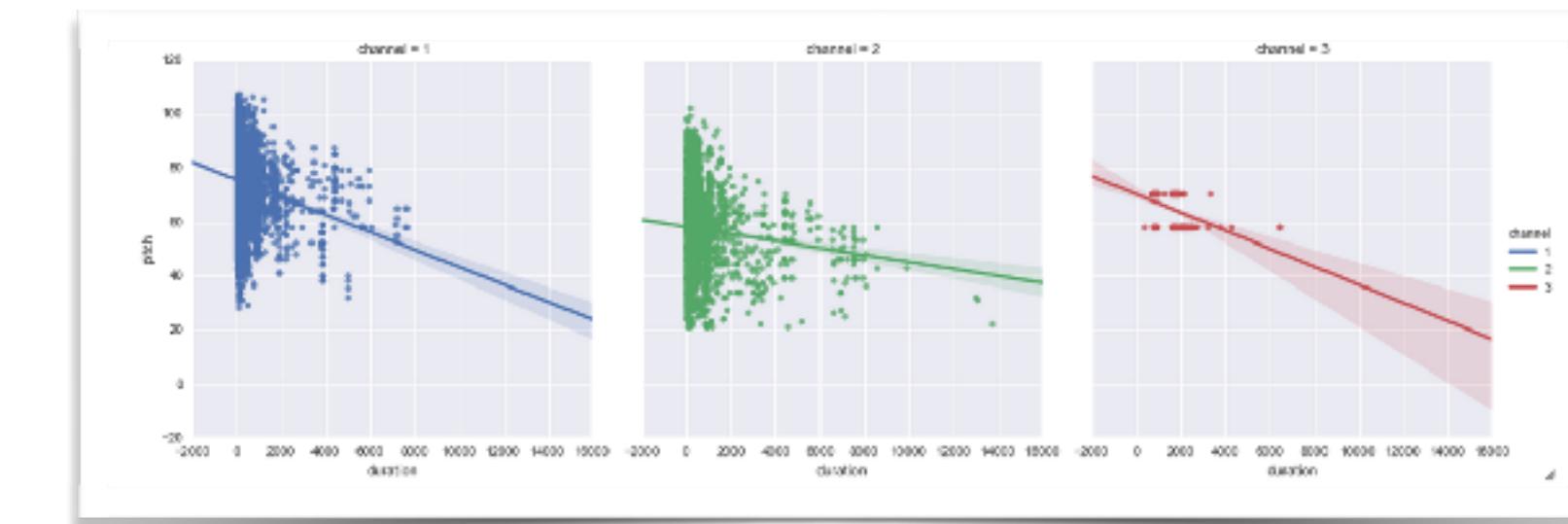
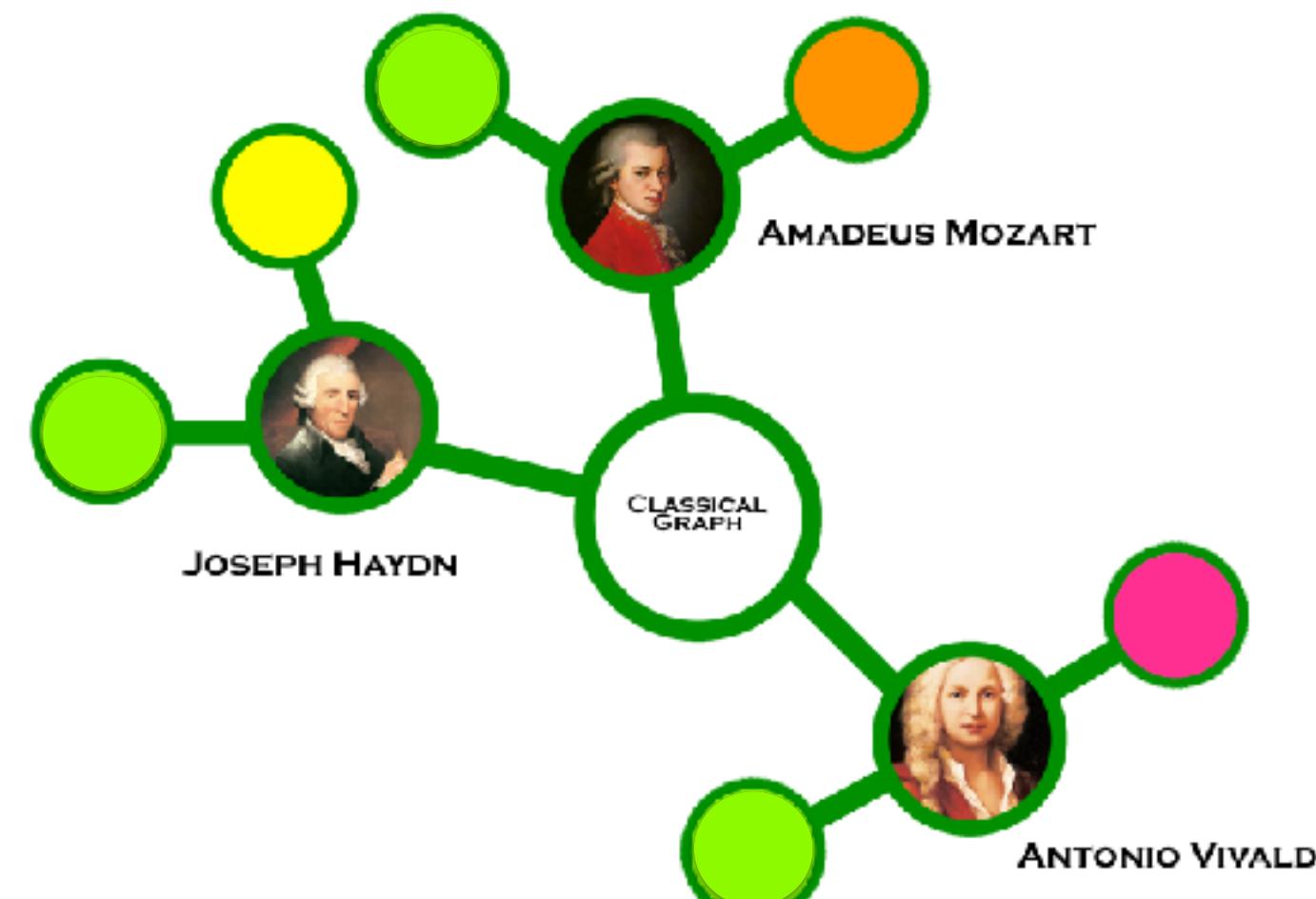
Editing: optional

Early Days - Composing a Jingle for a TV Show, mixing



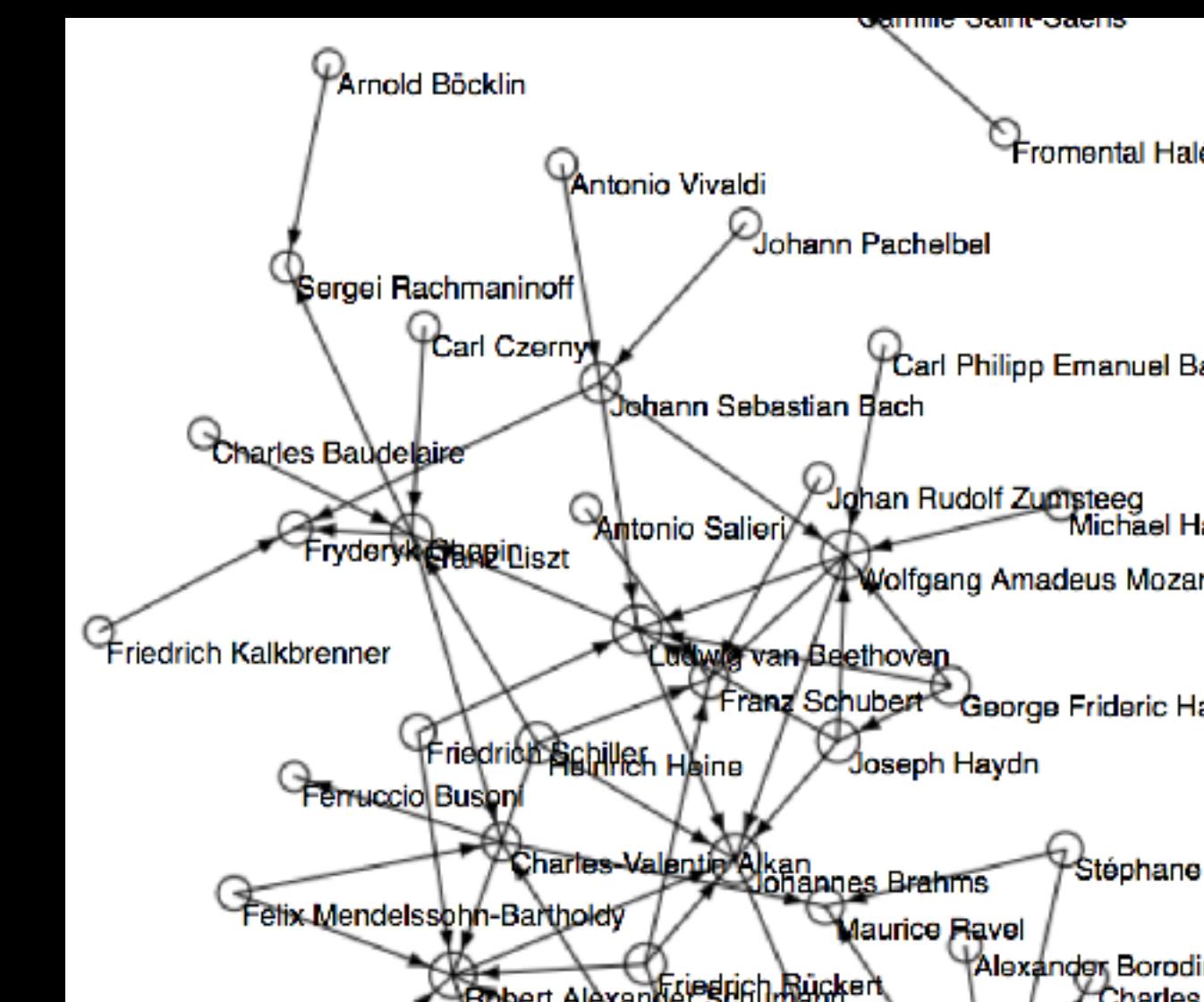
44.02 seconds

Clusters of “Musical Structure”



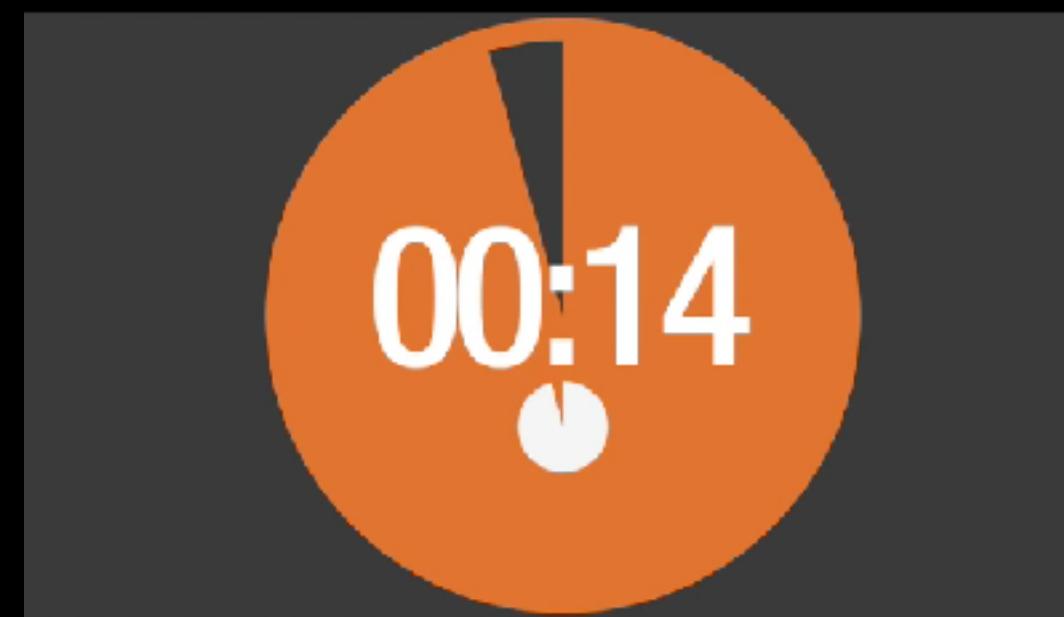
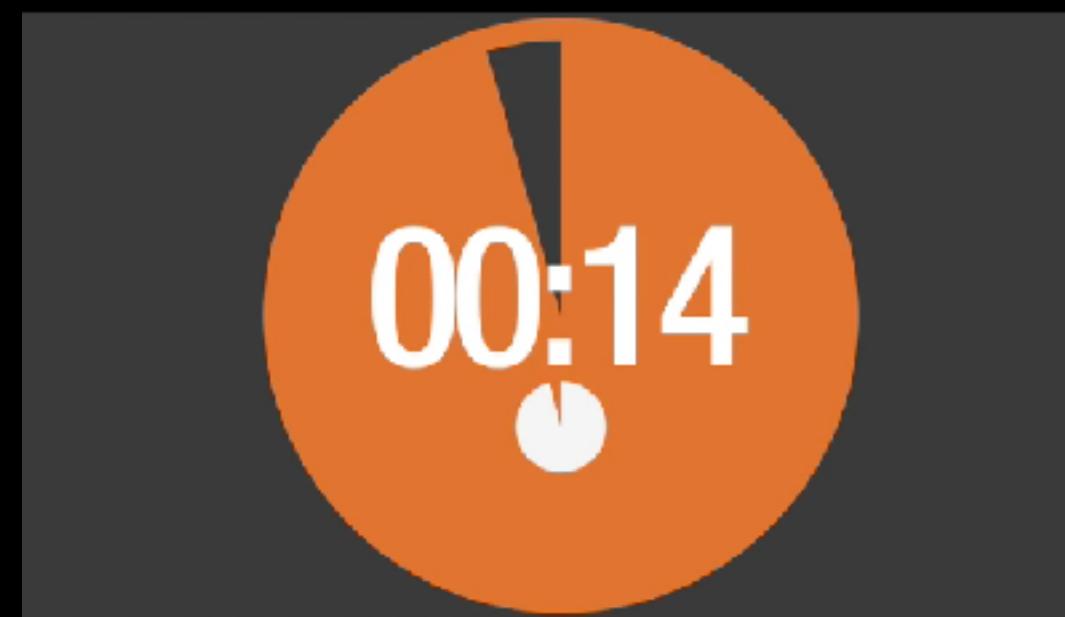
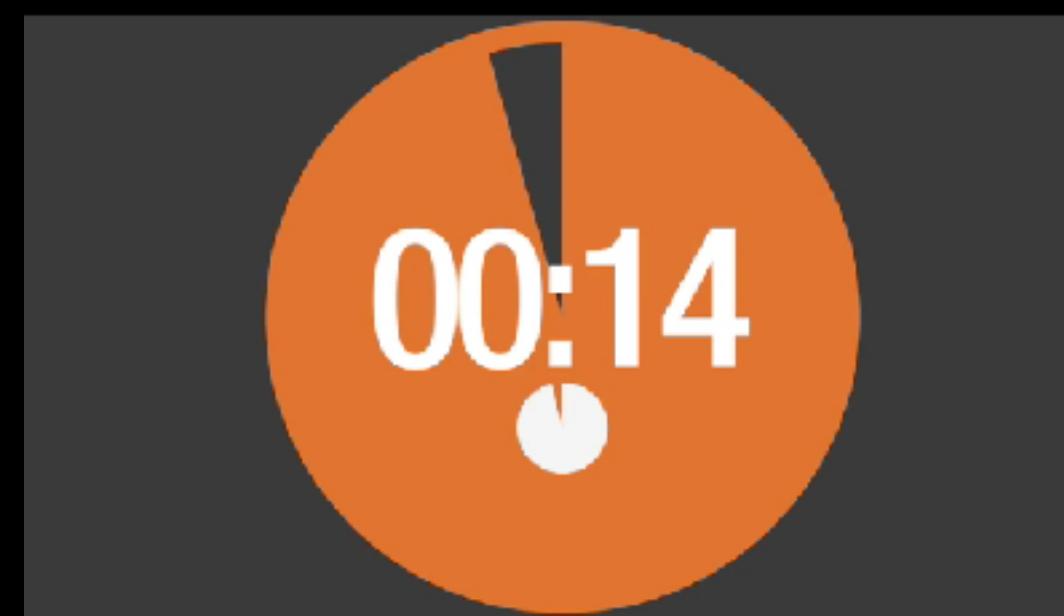
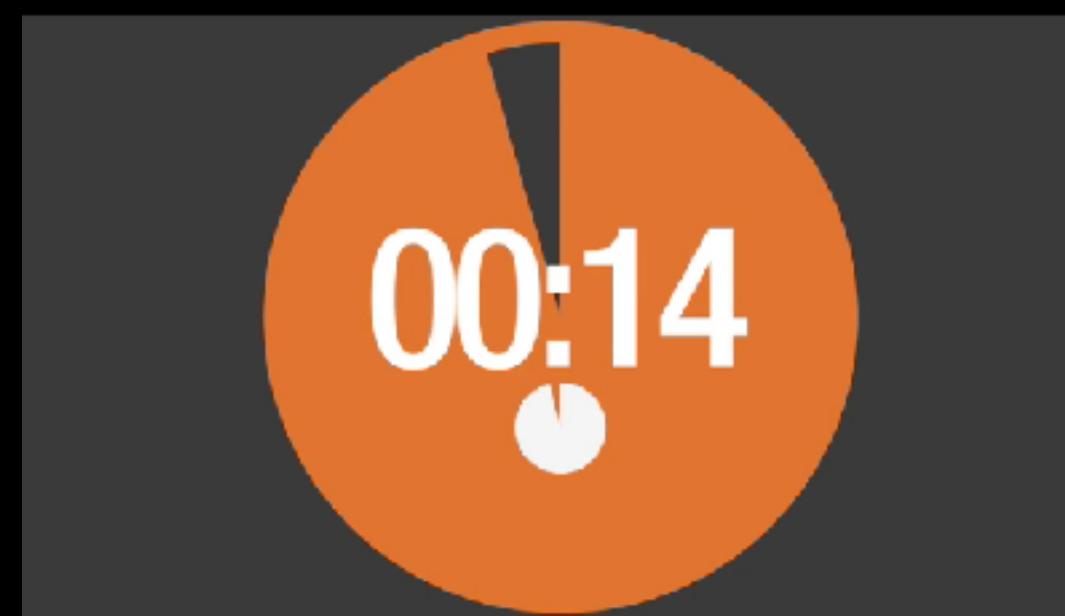
IN MUSIC, WHO INFLUENCED WHO? INTELLIGENCE DOES NOT EVOLVE IN ISOLATION

Switch to HTML graph



Early Days - Composing Original Baby Lullabys

(collaboration with Michael Hasak of Palo Alto, California)



More Advanced - Making Inferences

AI: Create something that Mozart would have composed, limited to 60" approximately



1' 3"

<https://soundcloud.com/luis-m-sanchez-1/>

AI: Bach only wrote 30 Goldberg Variations for Harpsichord. Extrapolate Bach's work and compose an original Goldberg Variation of about 2" in length and with a Baroque orchestra



2' 25"

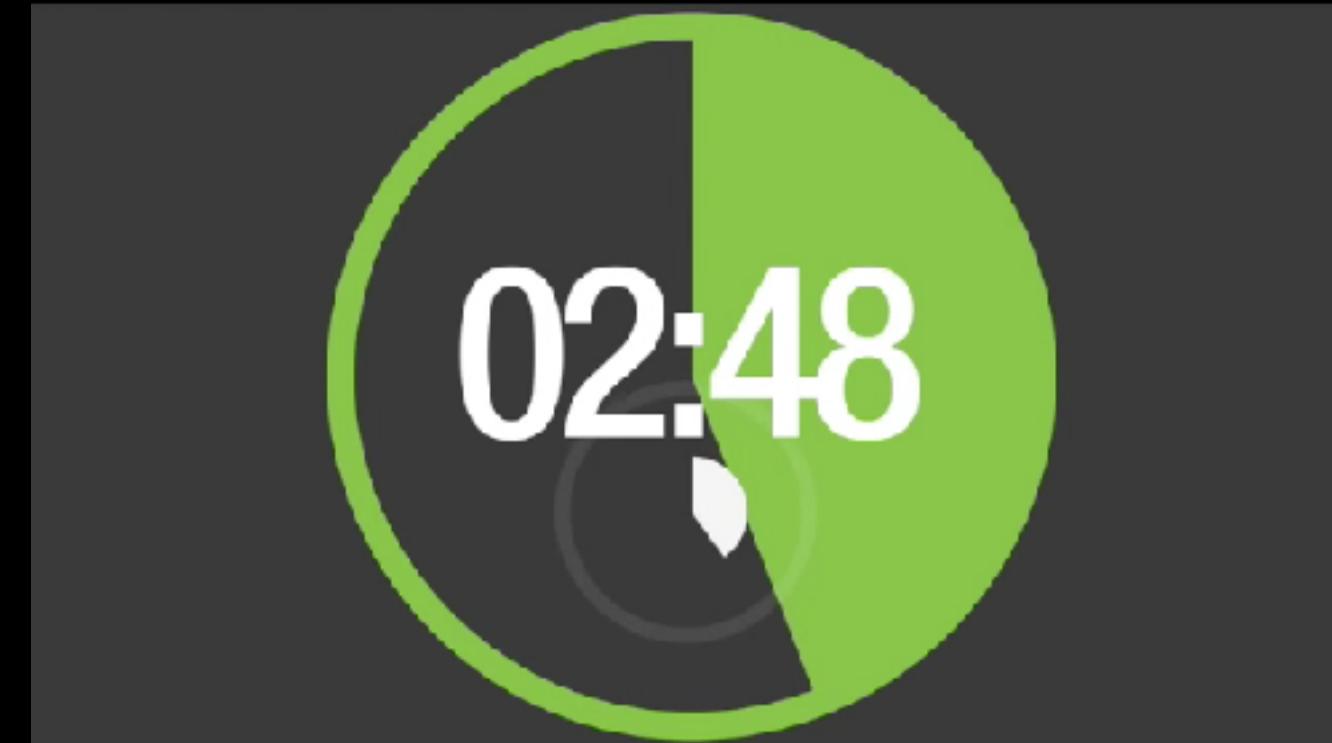
Half way - Imitating Styles

Create something that sounds like Bach's Cello Suite No.1 in G - prelude



Original

2' 39"



Artificial Intelligence variation

2' 48"

Even More Advanced - Making Inferences

**AI: Create a slow Jazz piece in piano with influence from
Glenn Miller, Miles Davis, and Duke Ellington**



<https://soundcloud.com/luis-m-sanchez-1/>

Evolved - Creating Original Pieces

AI: Create any song

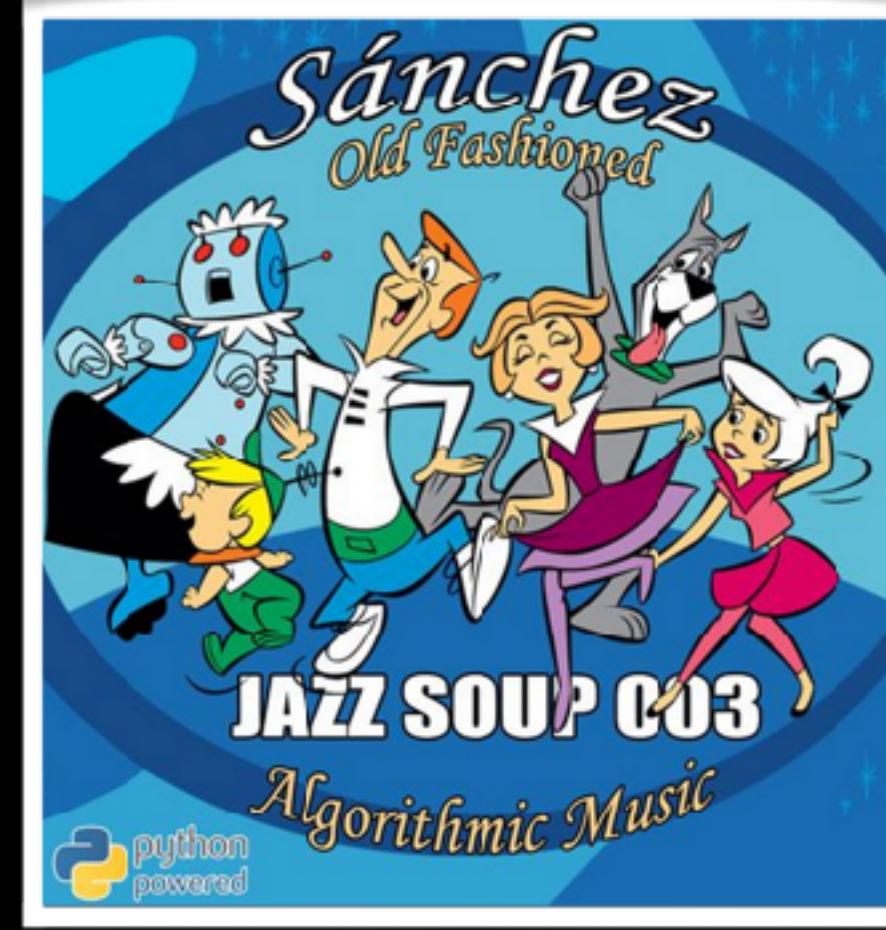
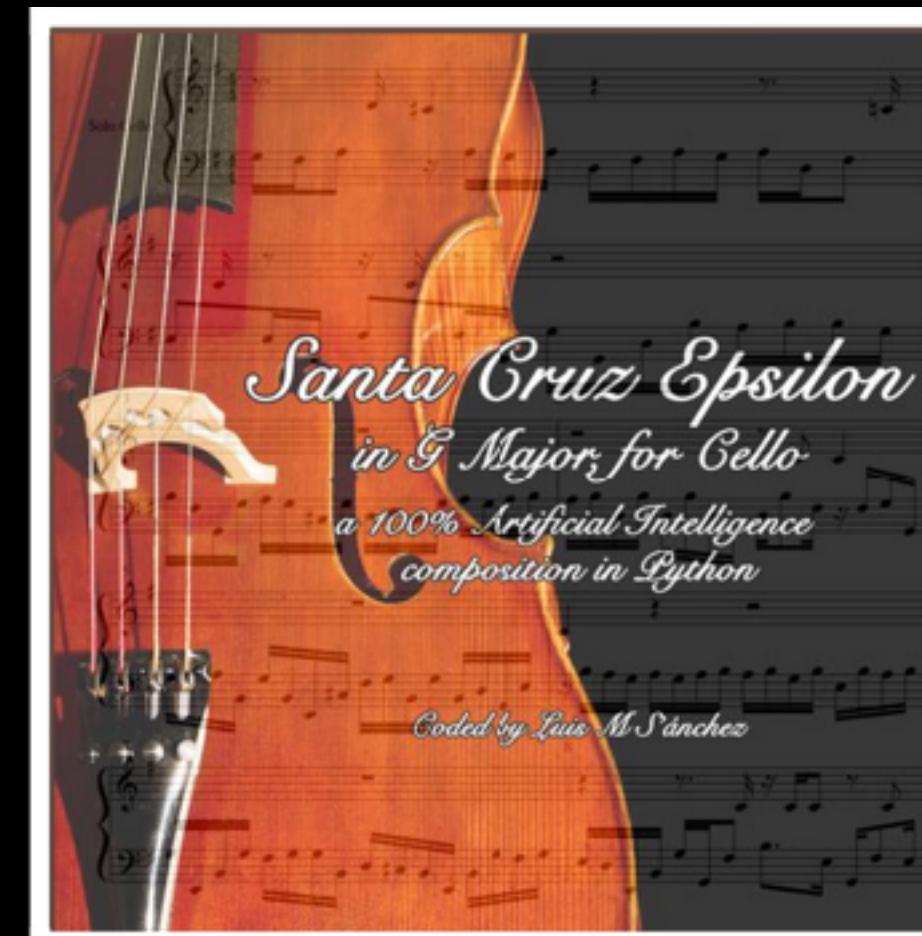
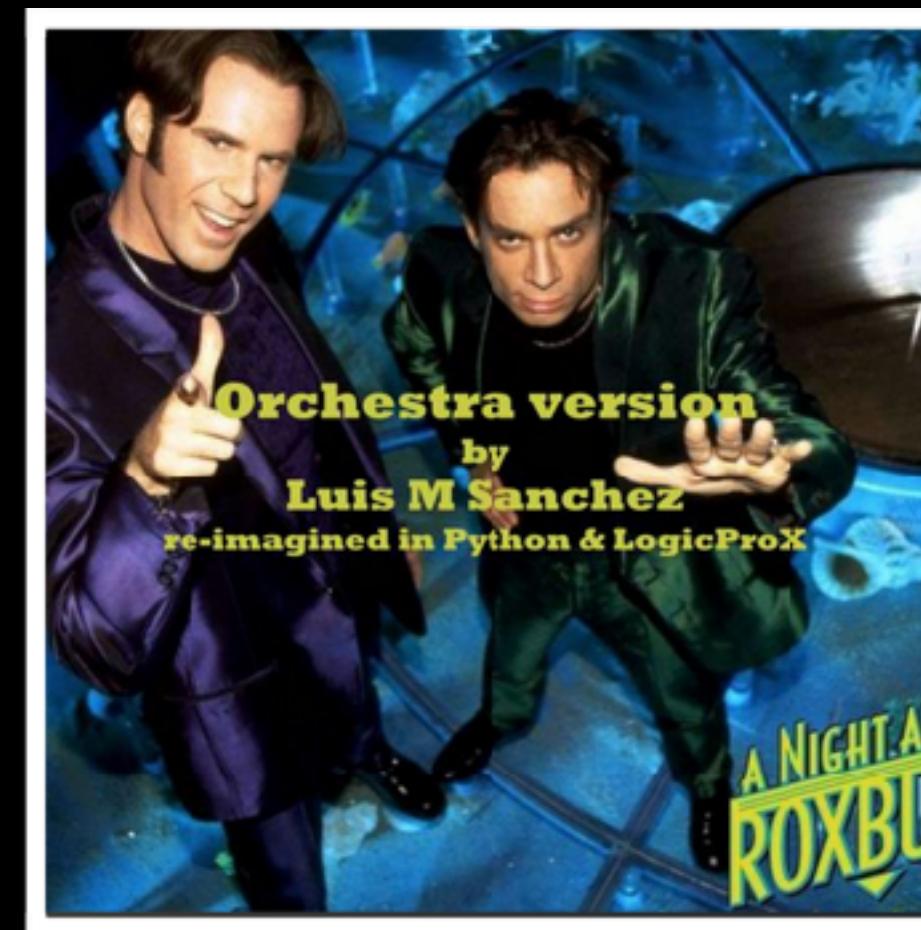


<https://soundcloud.com/luis-m-sanchez-1/>

Evolved - AI & Human Collaboration



<https://soundcloud.com/luis-m-sanchez-1/>



<https://soundcloud.com/luis-m-sanchez-1/>

Algorithmic Music Composition is the most fun project I have been involved using python. Although there is revenue already from royalty payments from songs sold and/or streamed in amazon, Apple Music, iTunes, Spotify, etc. a strategic partner in the music business would be required to make this a stand alone business.

Python libraries used: pandas, numpy, scipy, seaborn, bokeh, proprietary MusicQuant library, Music 21 (MIT), seaborn, librosa

FOR MORE INFORMATION:

<http://algorhythm.ai/>

**SWITCH TO JUPYTER
NOTEBOOK FOR DEMO**

**FOR MORE INFORMATION, JOIN
MY MEETUP GROUP**

[https://www.meetup.com/Algorithmic-Art-Visual-Musical-
Composition-Quadrivium/](https://www.meetup.com/Algorithmic-Art-Visual-Musical-Composition-Quadrivium/)

Conclusions

WORK: sgxanalytics.ai

STACKOVERFLOW: Luis Miguel

GITHUB: <https://github.com/lmsanch>

TWITTER: lmsanch

SOUNDCLOUD: <https://soundcloud.com/luis-m-sanchez-1>

<https://soundcloud.com/algorythm-ai>

APPLE MUSIC: <https://itunes.apple.com/us/artist/luis-miguel-sanchez/id950482695>

SPOTIFY: <https://open.spotify.com/album/28rza84rTb9O3jGolDa4m8>

PROJECTS: catrisk.com, Algorythm.ai

LINKEDIN: <https://www.linkedin.com/in/lmsanch/>

**THANK YOU FOR
YOUR ATTENTION!
FEEL FREE TO
CONTACT AND ENJOY
PYCON MY 2017!**