Pixelating Geo-Diversity

David Rodriguez, Jingchuan Chen, Dhia Mahjoub

Before we start:

https://github.com/DavidRdgz/hitbsec-notebook

Why us:

- We are maintaining a domain risk-score API used by our customers
- Last year we’ve been deploying tensorflow models into production
- GLMs
- Convolutional Neural Networks
- 500 Jobs a day in Complex Workflows (Hadoop Based)

Section 1

Introduce Requester Geo-Popularity Data

Section 2

Introduce Exponential Moving Averages (EMAs)

Section 3

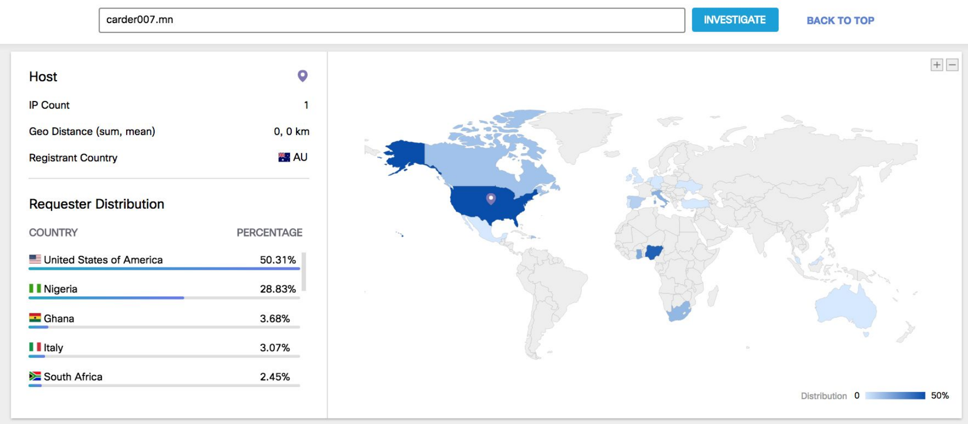
Convolutional Neural Networks

Section 1

Introduce Requester Geo-Popularity Data

+ Intuitions for modeling requests
+ Modeling requests at scale

Countries
+
Requests



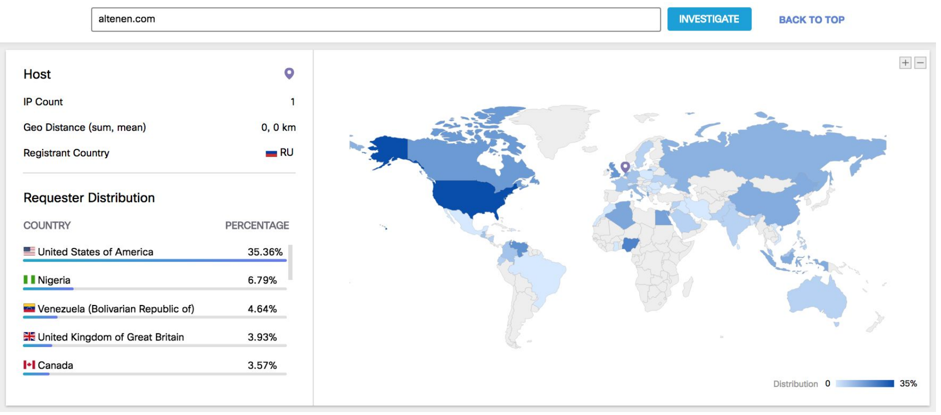
carder007.mn

Host

IP Count	1
Geo Distance (sum, mean)	0, 0 km
Registrant Country	AU

Requester Distribution

COUNTRY	PERCENTAGE
United States of America	50.31%
Nigeria	28.83%
Ghana	3.68%
Italy	3.07%
South Africa	2.45%



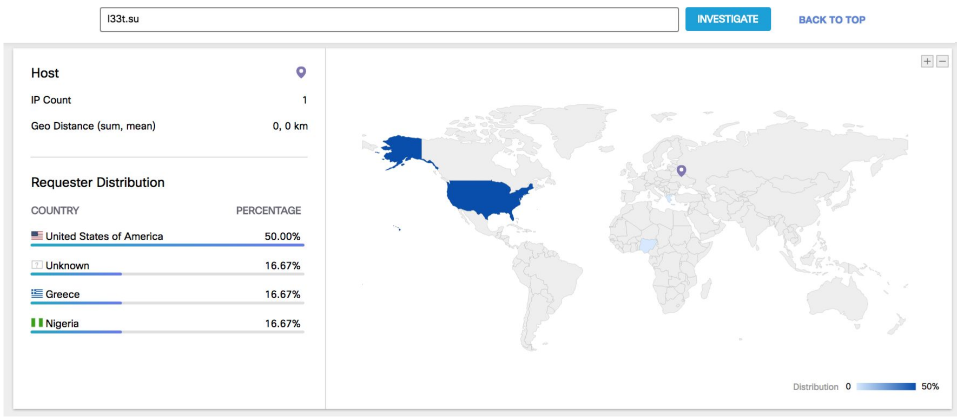
altenen.com

Host

IP Count	1
Geo Distance (sum, mean)	0, 0 km
Registrant Country	RU

Requester Distribution

COUNTRY	PERCENTAGE
United States of America	35.36%
Nigeria	6.79%
Venezuela (Bolivarian Republic of)	4.64%
United Kingdom of Great Britain	3.93%
Canada	3.57%



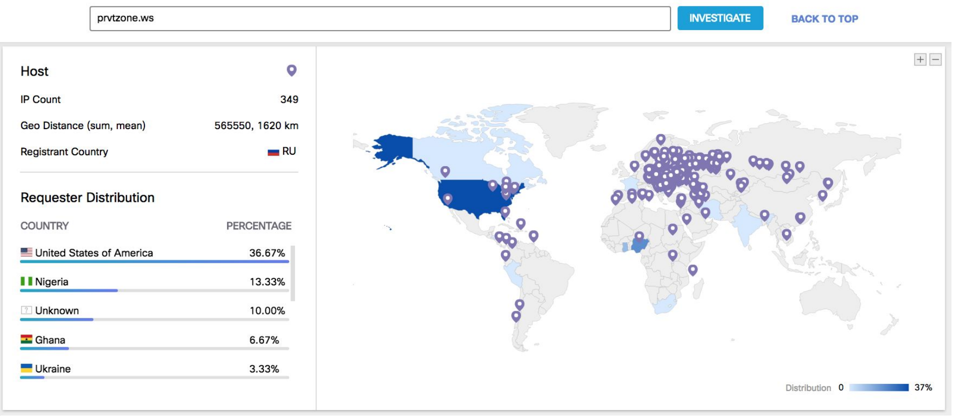
l33t.su

Host

IP Count	1
Geo Distance (sum, mean)	0, 0 km

Requester Distribution

COUNTRY	PERCENTAGE
United States of America	50.00%
Unknown	16.67%
Greece	16.67%
Nigeria	16.67%



prvtzone.ws

Host

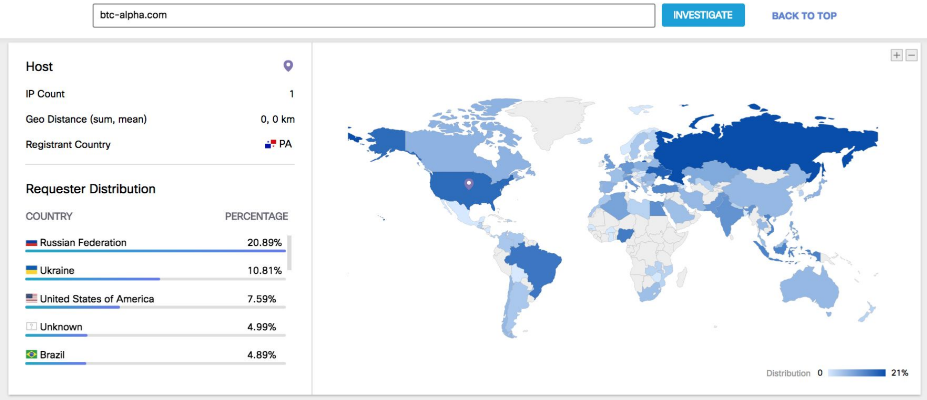
IP Count	349
Geo Distance (sum, mean)	565550, 1620 km
Registrant Country	RU

Requester Distribution

COUNTRY	PERCENTAGE
United States of America	36.67%
Nigeria	13.33%
Unknown	10.00%
Ghana	6.67%
Ukraine	3.33%

Carding sites

US
NG
DE
UA

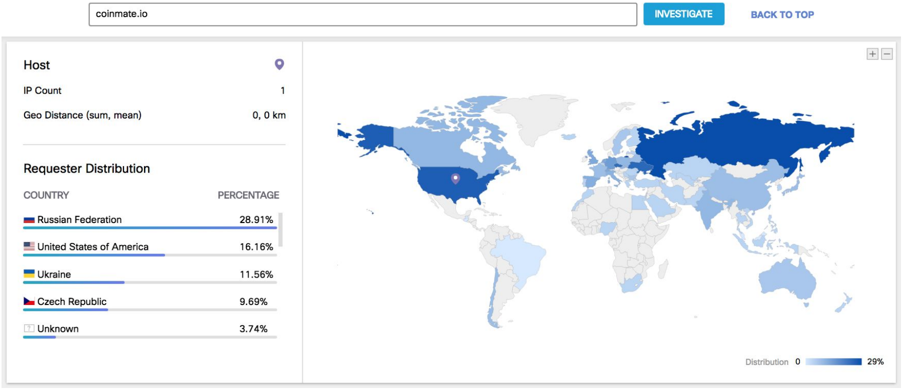


Host

IP Count	1
Geo Distance (sum, mean)	0, 0 km
Registrant Country	PA

Requester Distribution

COUNTRY	PERCENTAGE
Russian Federation	20.89%
Ukraine	10.81%
United States of America	7.59%
Unknown	4.99%
Brazil	4.89%

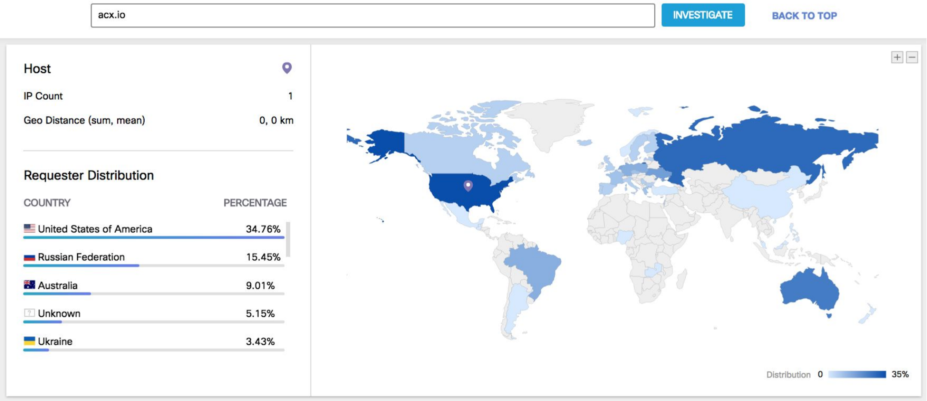


Host

IP Count	1
Geo Distance (sum, mean)	0, 0 km

Requester Distribution

COUNTRY	PERCENTAGE
Russian Federation	28.91%
United States of America	16.16%
Ukraine	11.56%
Czech Republic	9.69%
Unknown	3.74%

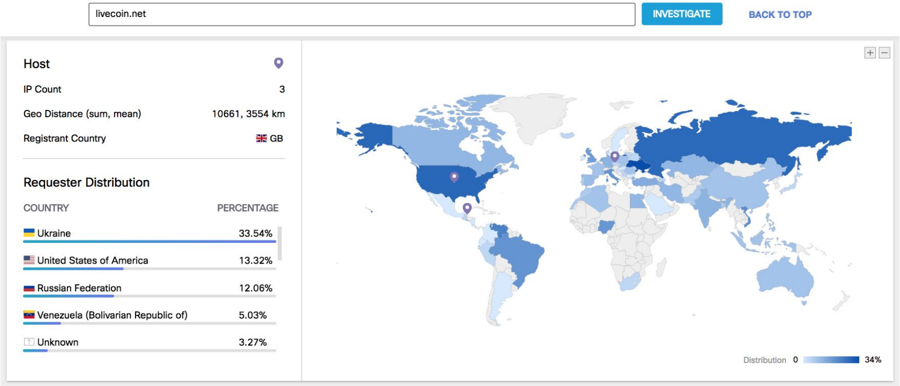


Host

IP Count	1
Geo Distance (sum, mean)	0, 0 km

Requester Distribution

COUNTRY	PERCENTAGE
United States of America	34.76%
Russian Federation	15.45%
Australia	9.01%
Unknown	5.15%
Ukraine	3.43%



Host

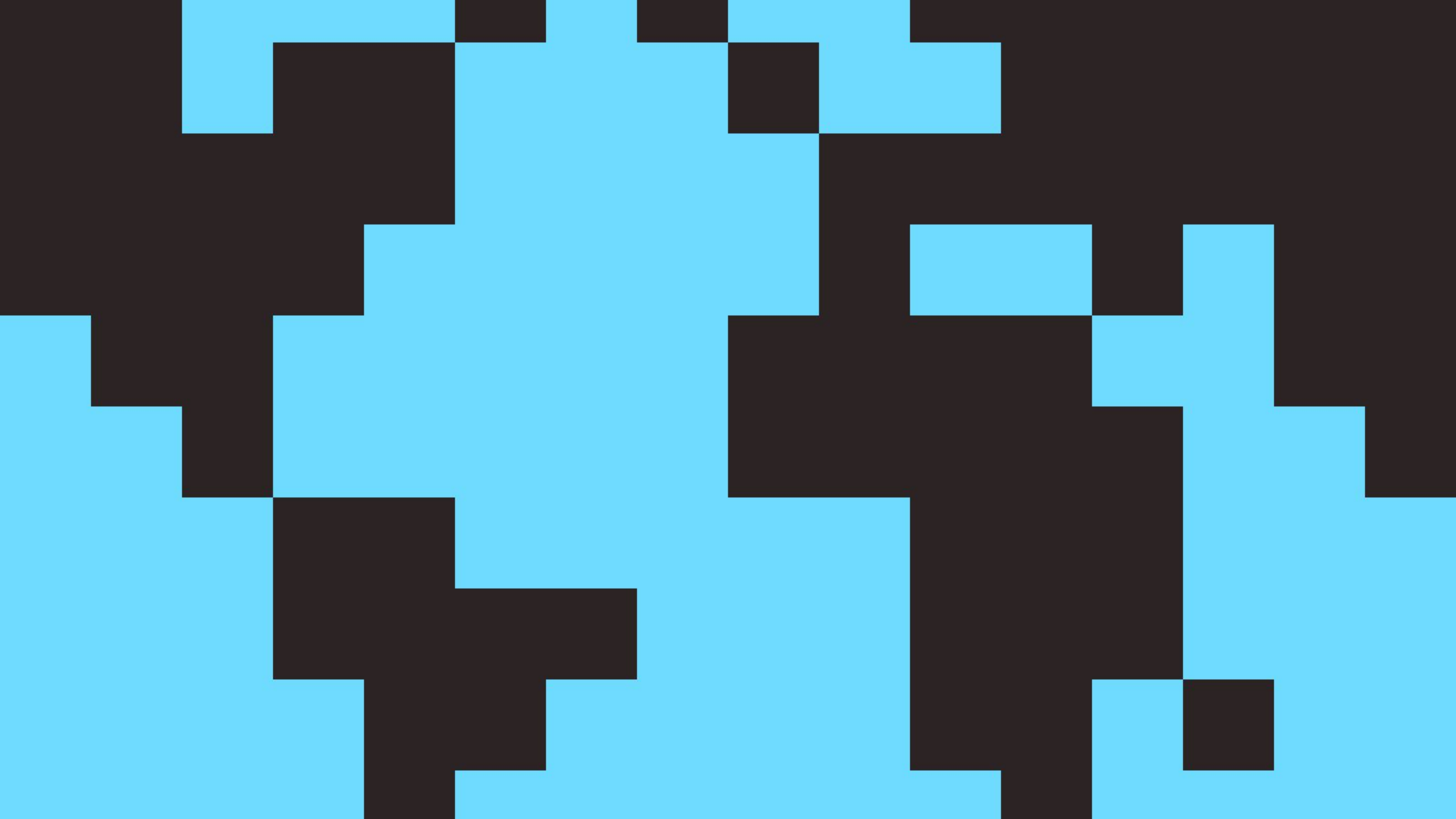
IP Count	3
Geo Distance (sum, mean)	10661, 3554 km
Registrant Country	GB

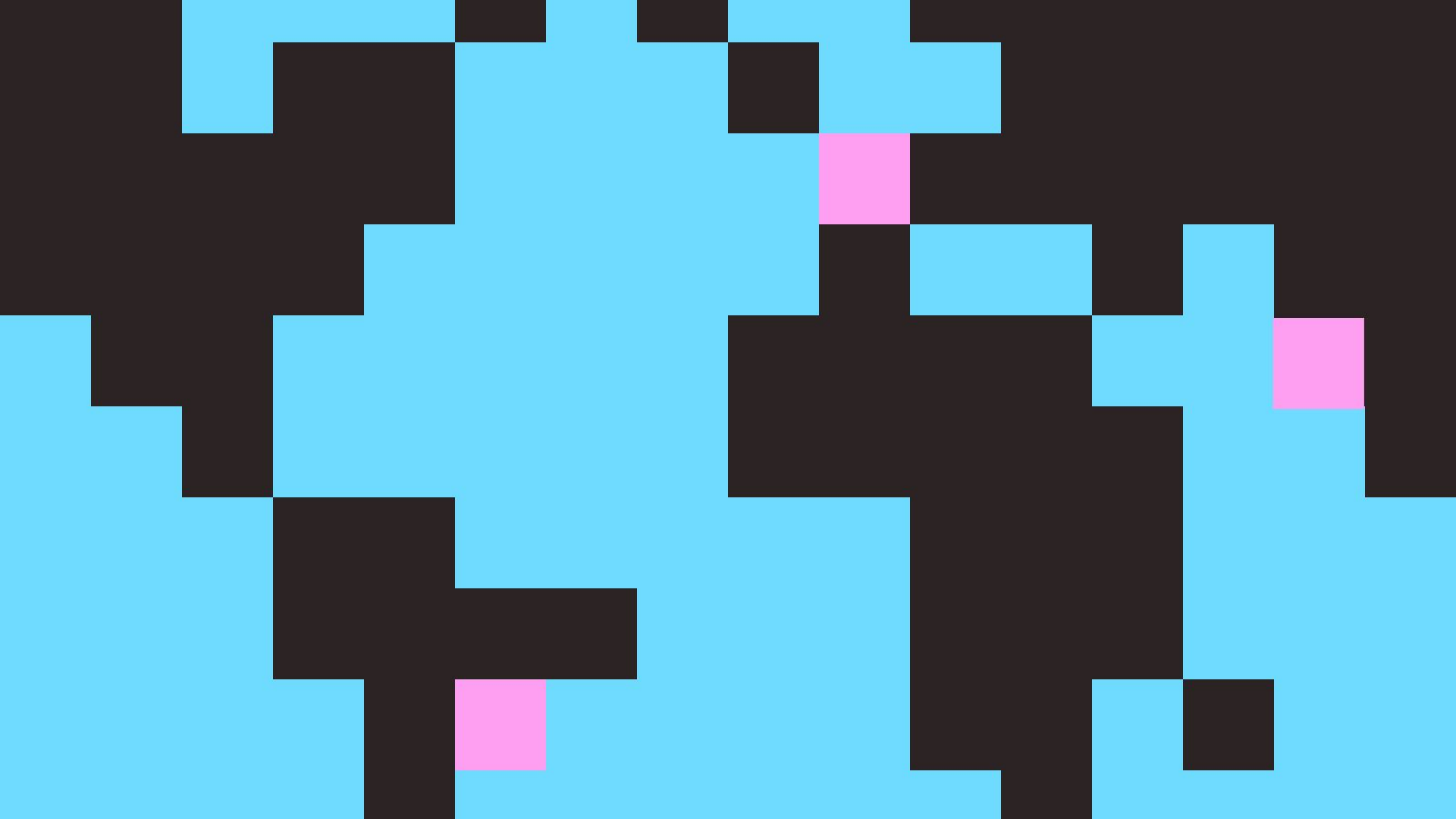
Requester Distribution

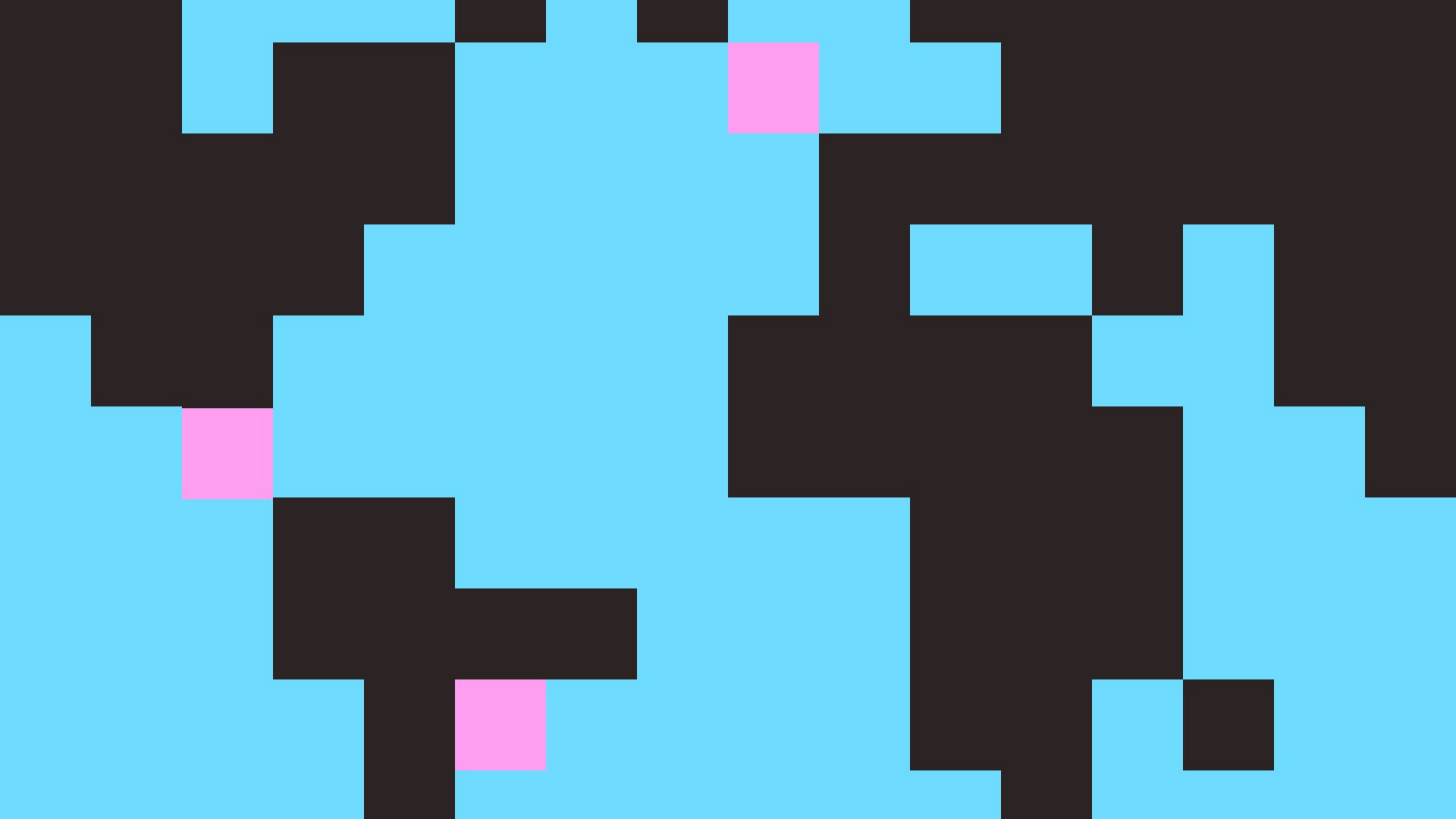
COUNTRY	PERCENTAGE
Ukraine	33.54%
United States of America	13.32%
Russian Federation	12.06%
Venezuela (Bolivarian Republic of)	5.03%
Unknown	3.27%

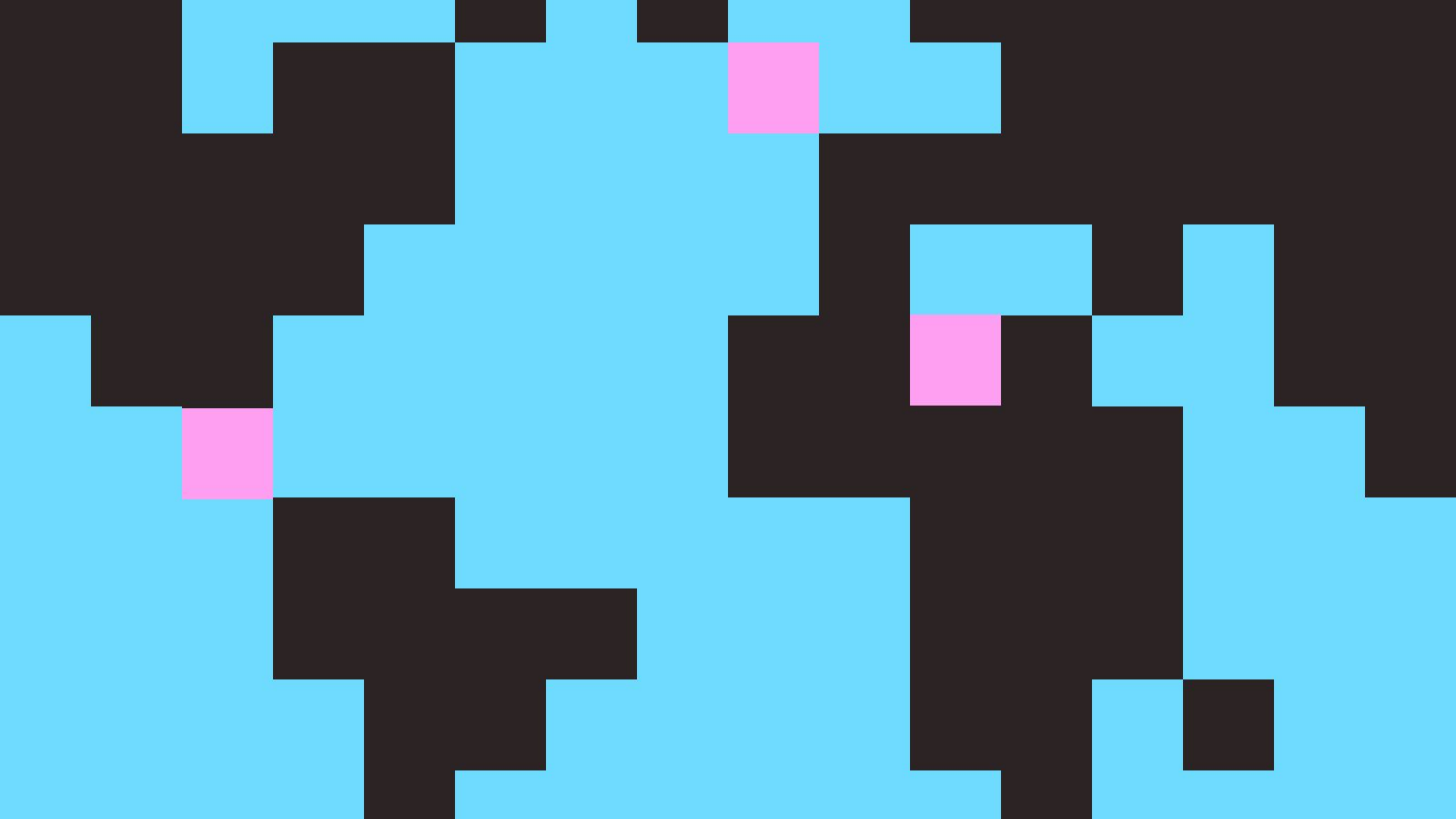
Altcoins sites

US, IT, GB, CA, UA, NG, RU





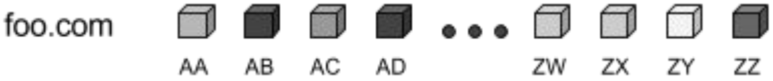


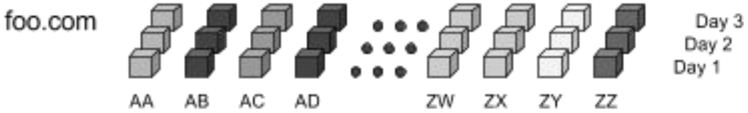


Countries

+

Counts


foo.com
AA
AB
AC
AD
ZW
ZX
ZY
ZZ


foo.com
Day 3
Day 2
Day 1
AA
AB
AC
AD
ZW
ZX
ZY
ZZ

Counts

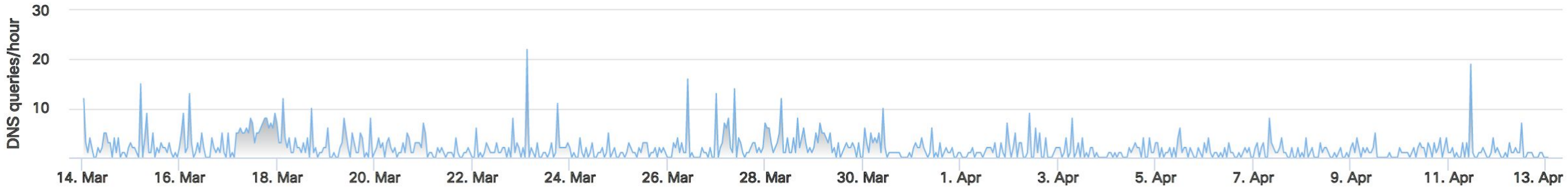
+

Time

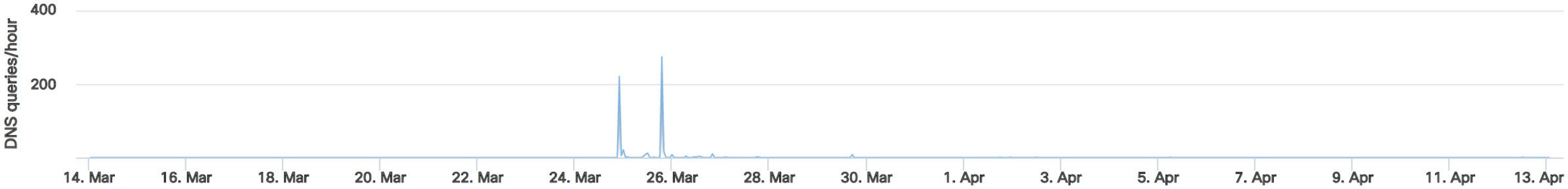
Section 2

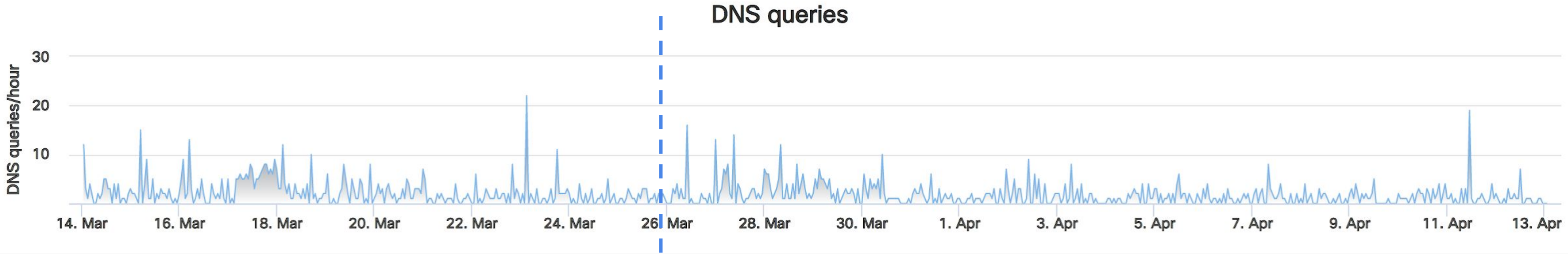
Introduce Exponential Moving Averages (EMAs)

DNS queries

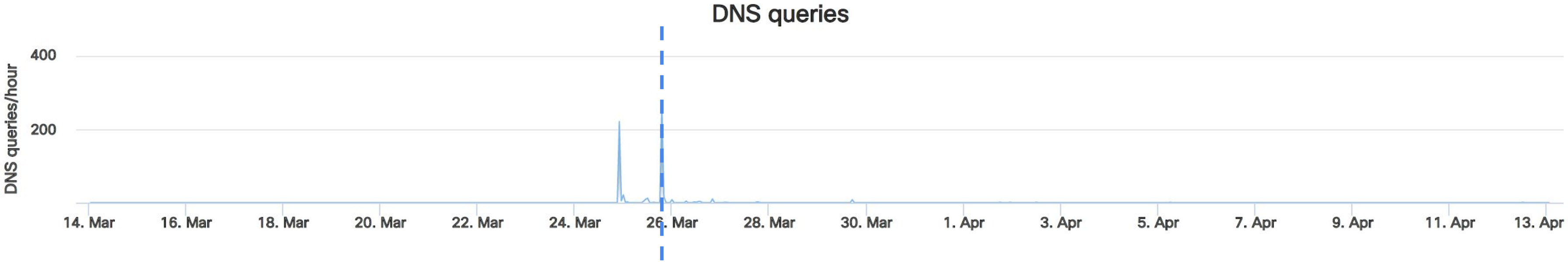


DNS queries

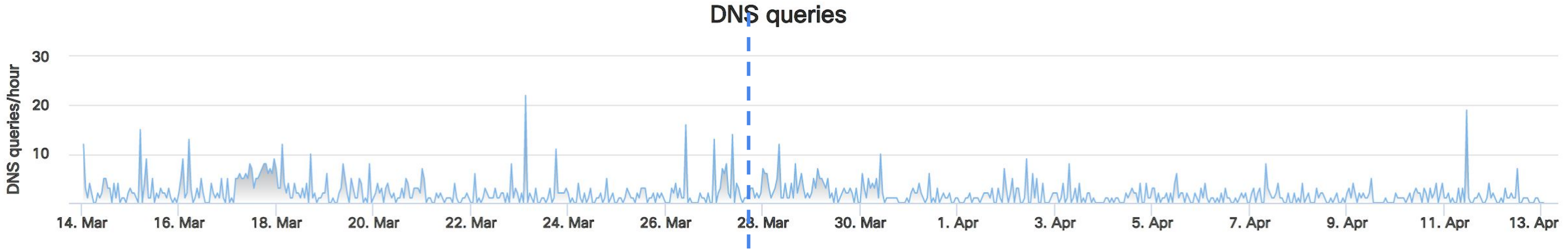


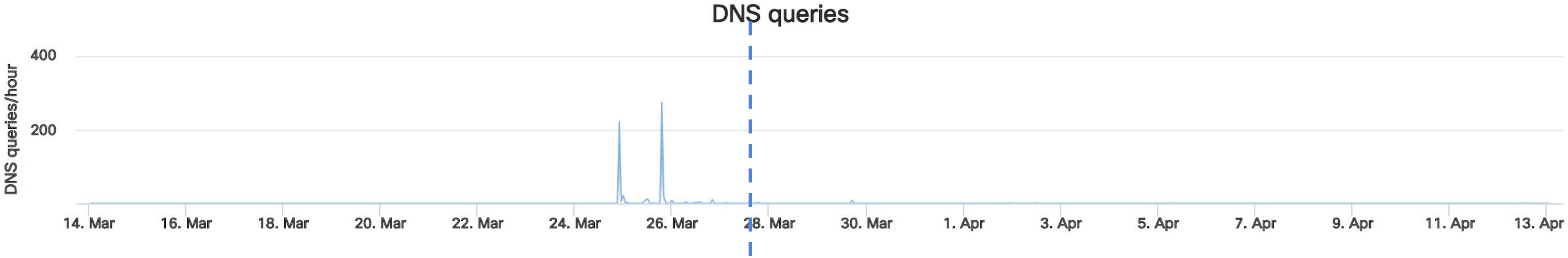


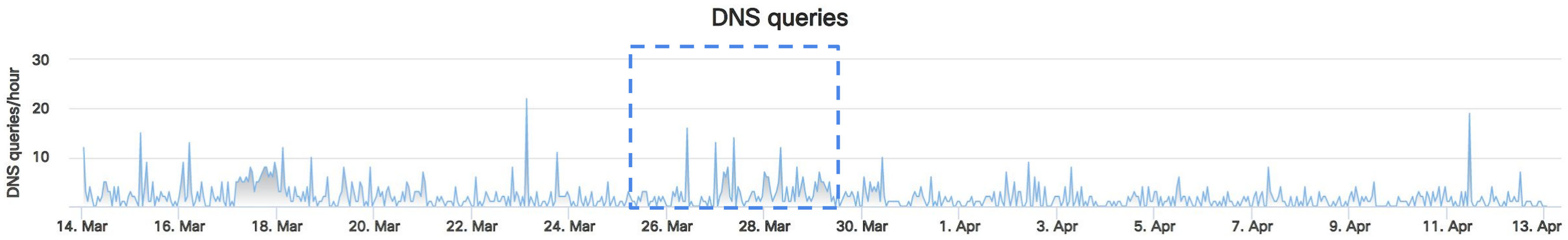
DNS queries
DNS queries/hour
30
20
10
14. Mar
16. Mar
18. Mar
20. Mar
22. Mar
24. Mar
26. Mar
28. Mar
30. Mar
1. Apr
3. Apr
5. Apr
7. Apr
9. Apr
11. Apr
13. Apr

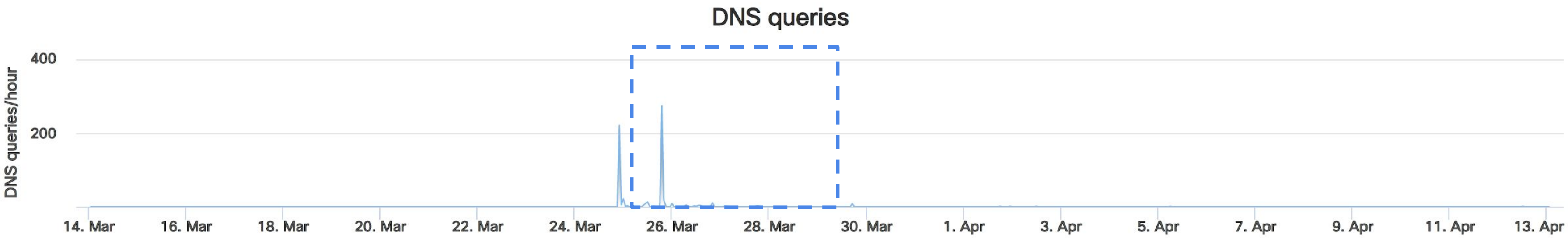


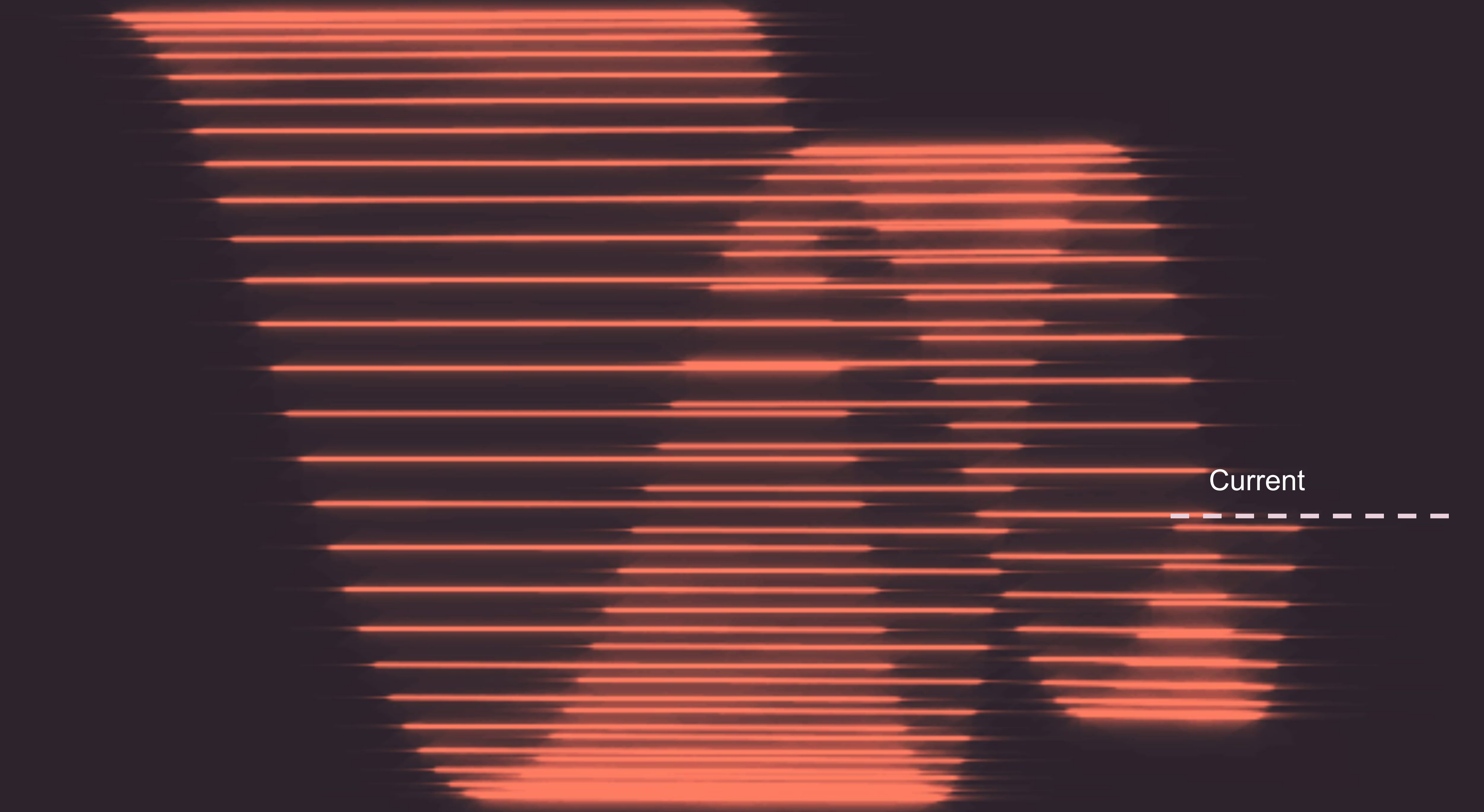
DNS queries
DNS queries/hour
400
200
14. Mar
16. Mar
18. Mar
20. Mar
22. Mar
24. Mar
26. Mar
28. Mar
30. Mar
1. Apr
3. Apr
5. Apr
7. Apr
9. Apr
11. Apr
13. Apr

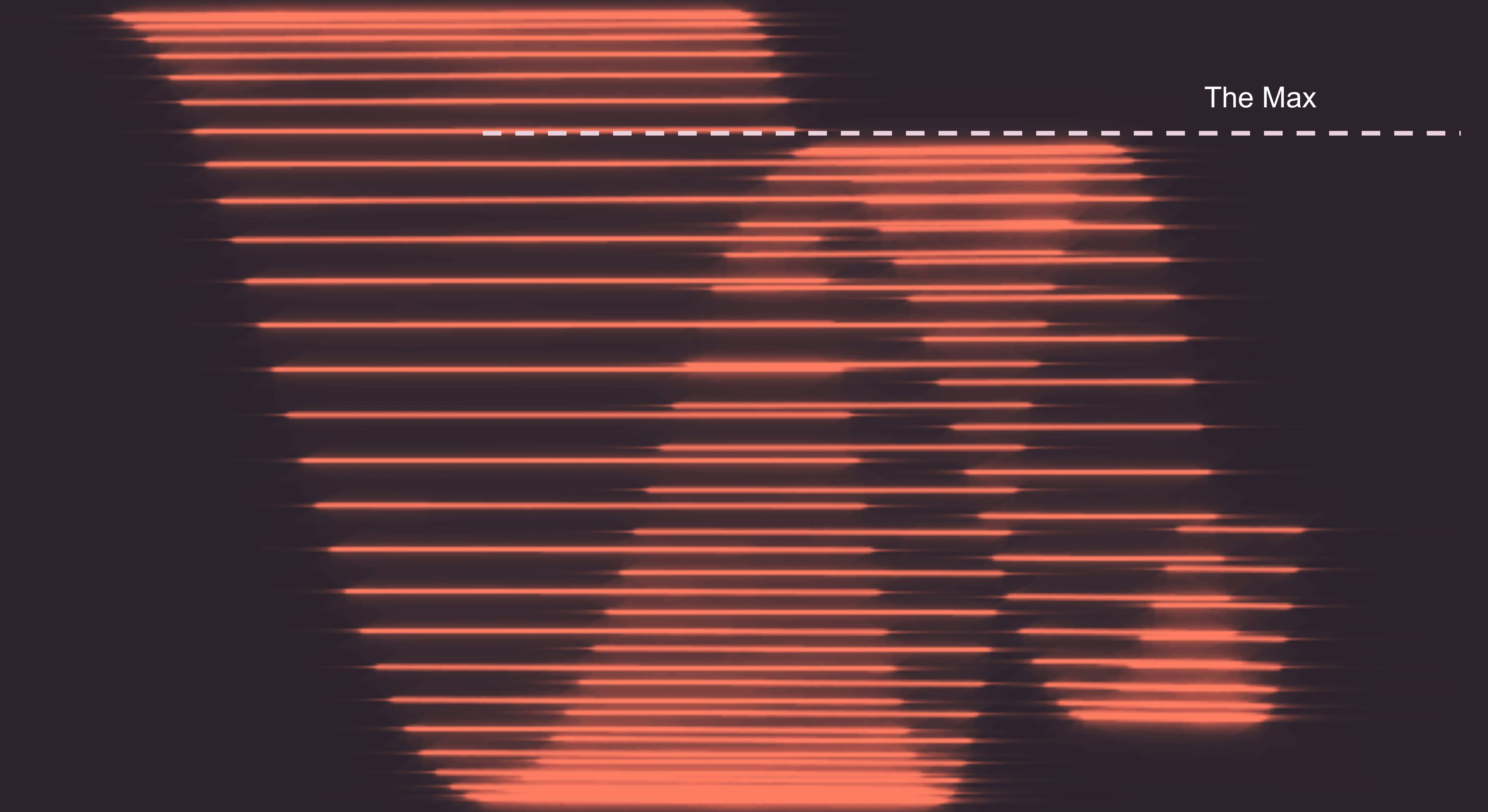

DNS queries
DNS queries/hour
30
20
10
14. Mar
16. Mar
18. Mar
20. Mar
22. Mar
24. Mar
26. Mar
28. Mar
30. Mar
1. Apr
3. Apr
5. Apr
7. Apr
9. Apr
11. Apr
13. Apr

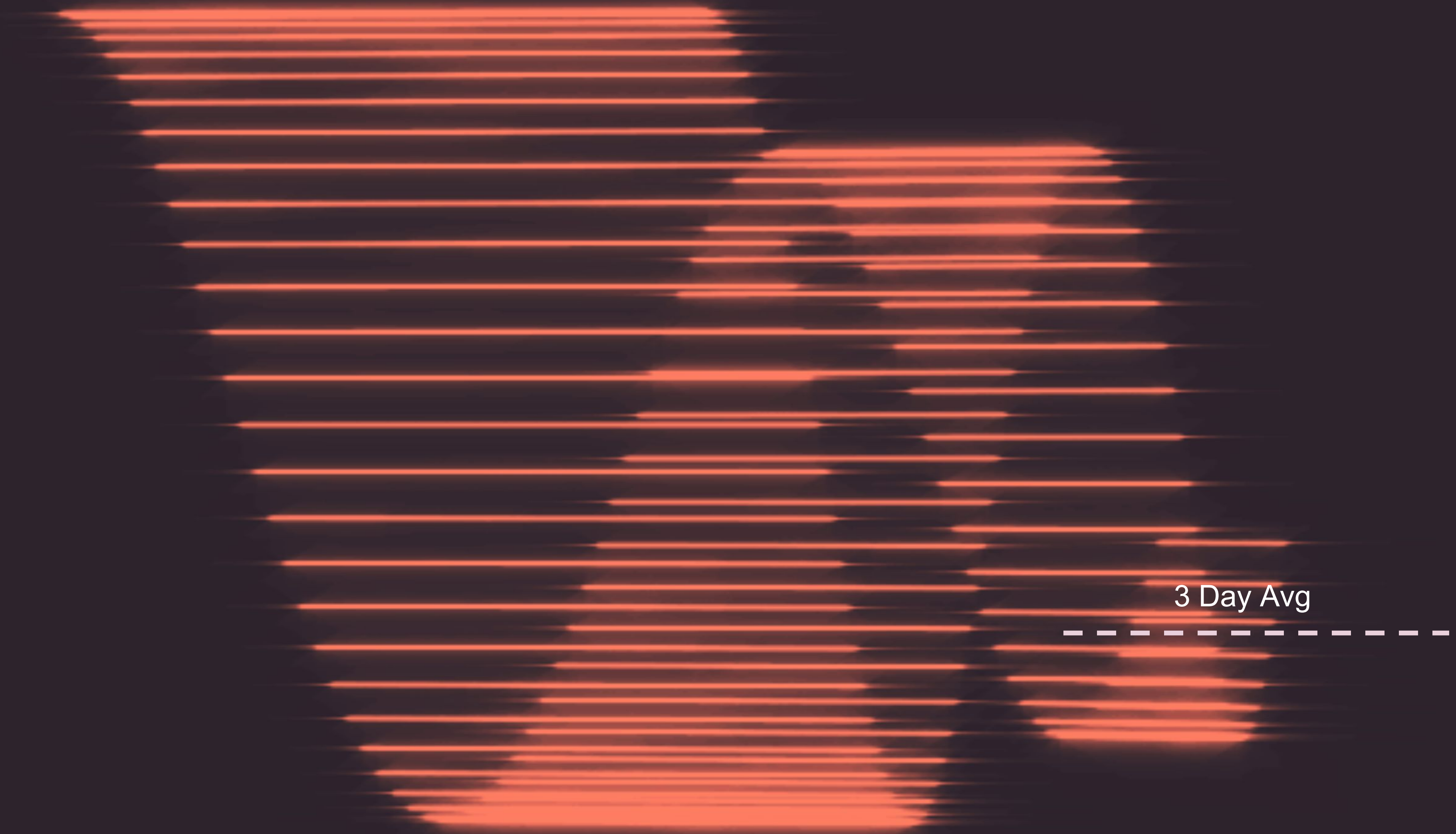

DNS queries
DNS queries/hour
400
200
14. Mar
16. Mar
18. Mar
20. Mar
22. Mar
24. Mar
26. Mar
28. Mar
30. Mar
1. Apr
3. Apr
5. Apr
7. Apr
9. Apr
11. Apr
13. Apr

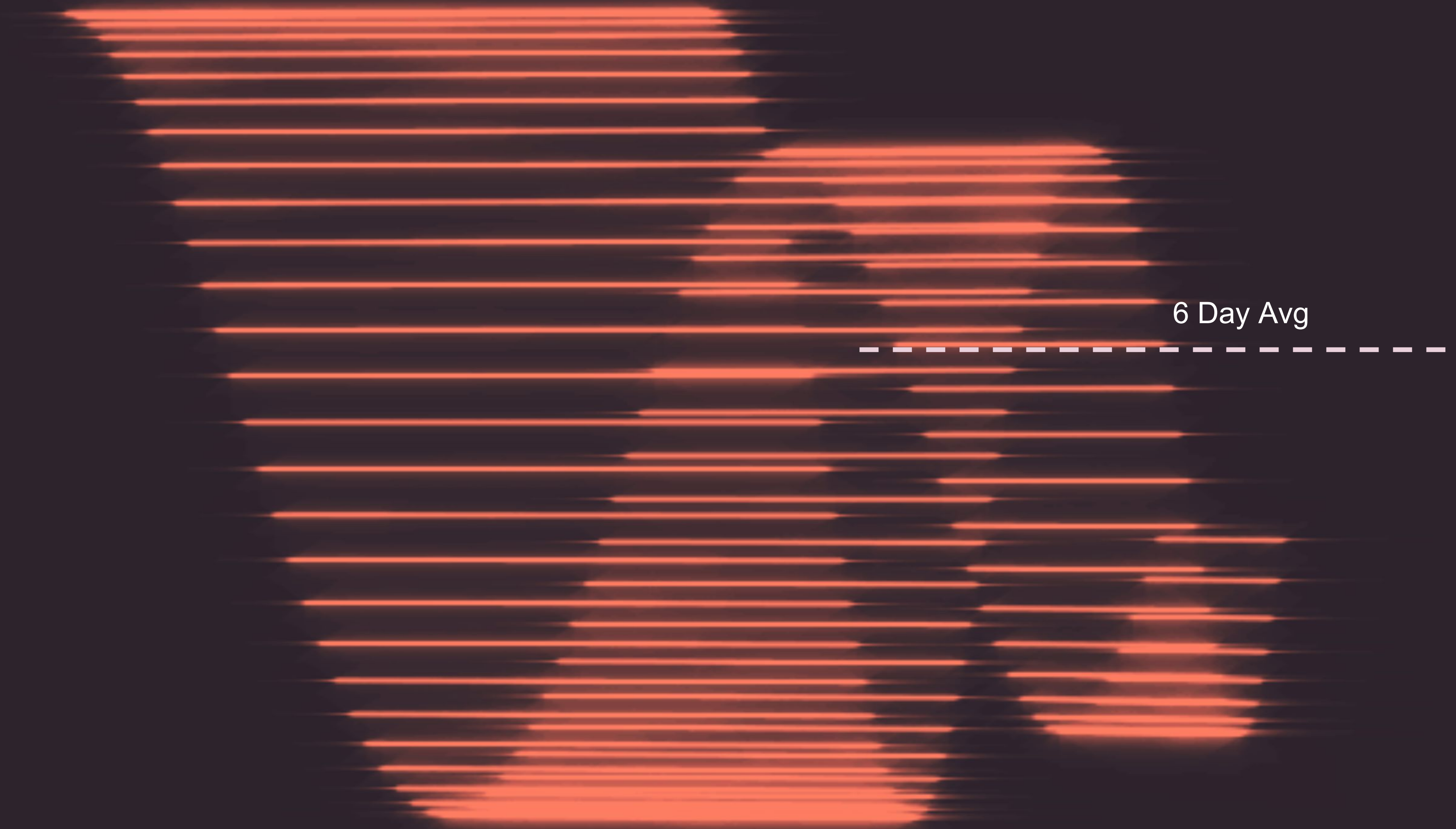

DNS queries
DNS queries/hour
30
20
10
14. Mar
16. Mar
18. Mar
20. Mar
22. Mar
24. Mar
26. Mar
28. Mar
30. Mar
1. Apr
3. Apr
5. Apr
7. Apr
9. Apr
11. Apr
13. Apr

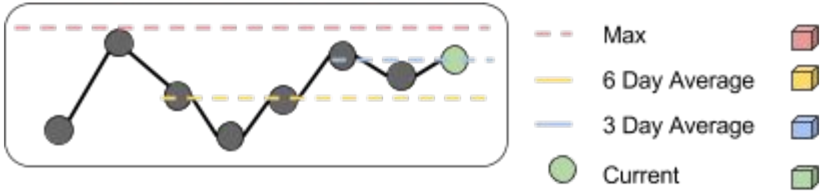

DNS queries
DNS queries/hour
400
200
14. Mar
16. Mar
18. Mar
20. Mar
22. Mar
24. Mar
26. Mar
28. Mar
30. Mar
1. Apr
3. Apr
5. Apr
7. Apr
9. Apr
11. Apr
13. Apr

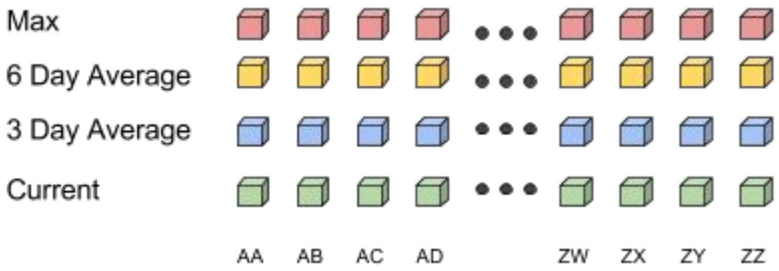

Current


The Max


3 Day Avg


6 Day Avg


Max
6 Day Average
3 Day Average
Current



Max

6 Day Average

3 Day Average

Current

AA AB AC AD ZW ZX ZY ZZ

Exponential Moving Average

$$S_t = \alpha Y_t + (1 - \alpha) S_{t-1}$$

Optional

$$S_t = \alpha[Y_{t-1} + (1 - \alpha) Y_{t-2} + (1 - \alpha)^2 Y_{t-3} + \cdots]$$

Convolutional Networks

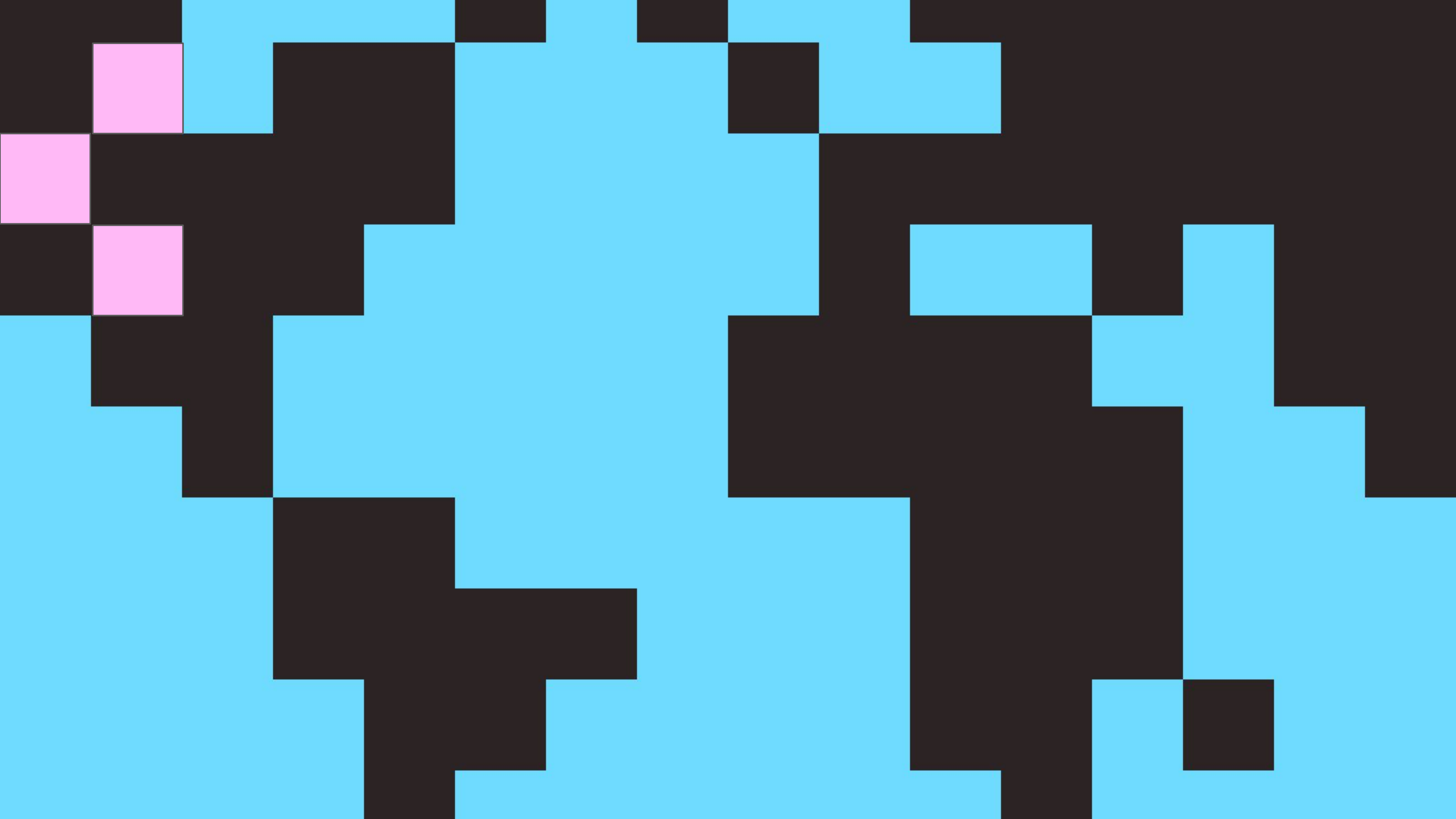
Section 2

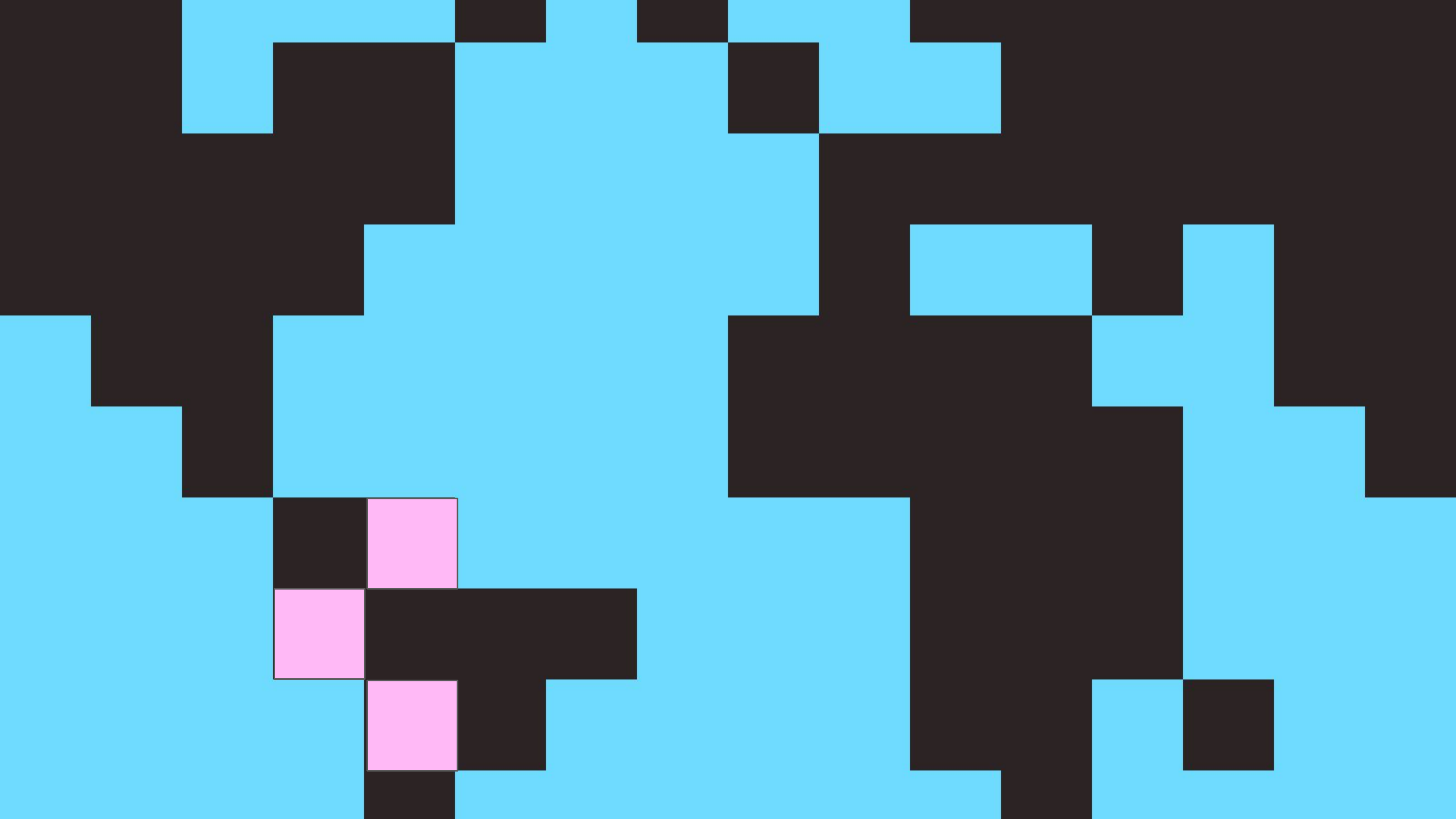
Convolutional Neural Networks

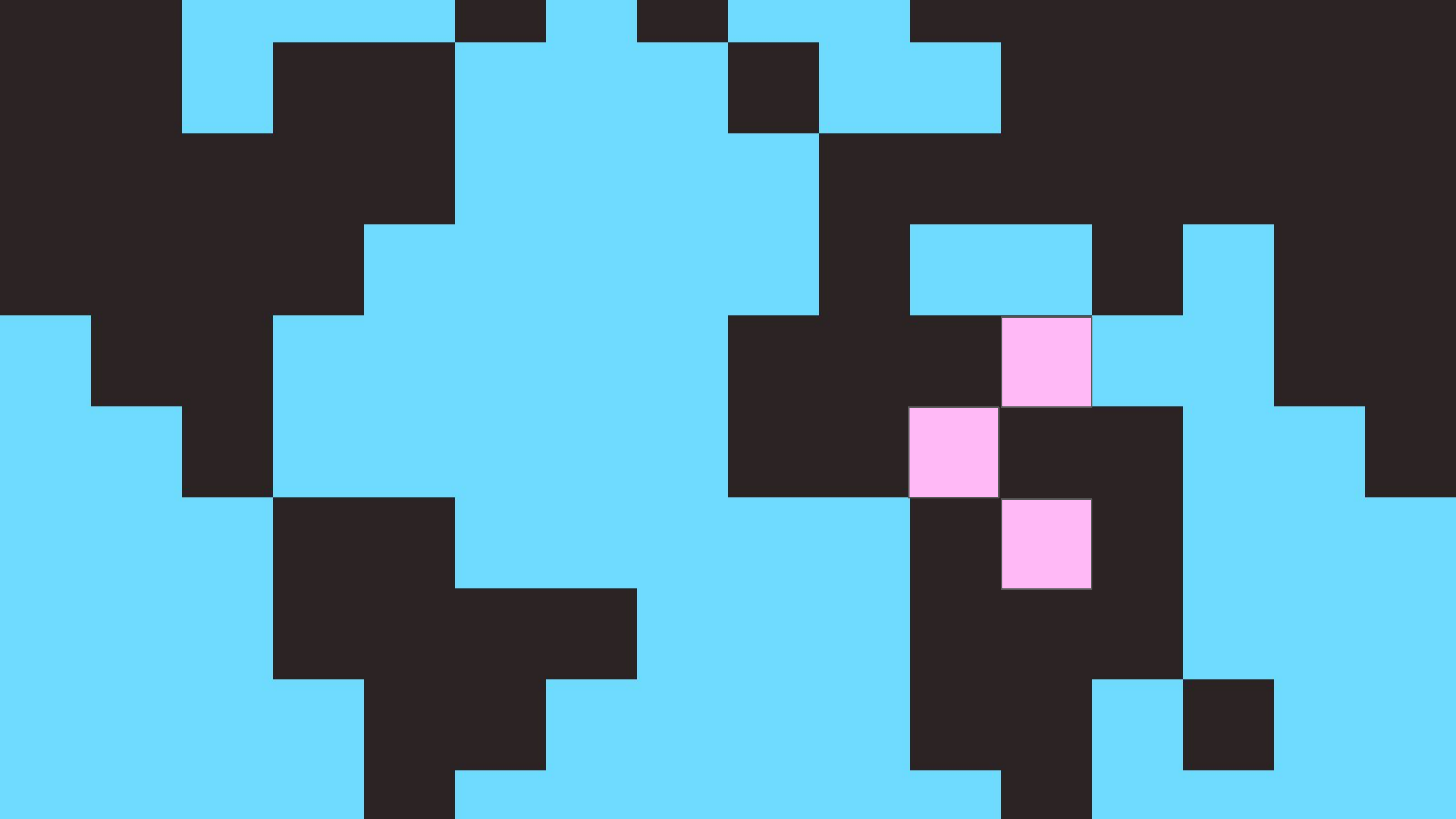
- Why these models?
- Inputs and Layers

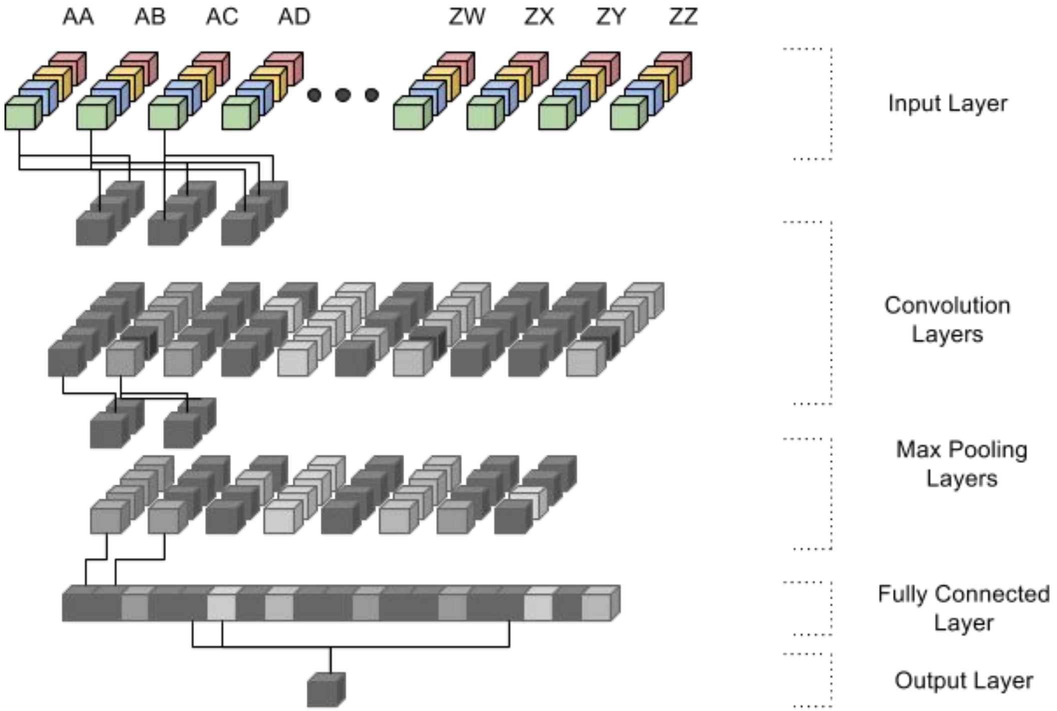
Zhang, Wei (1988). "Shift-invariant pattern recognition neural network and its optical architecture". Proceedings of annual conference of the Japan Society of Applied Physics.

Zhang, Wei (1990). "Parallel distributed processing model with local space-invariant interconnections and its optical architecture". Applied Optics. 29 (32): 4790–7. Bibcode:1990ApOpt..29.4790Z. doi:10.1364/AO.29.004790. PMID 20577468.

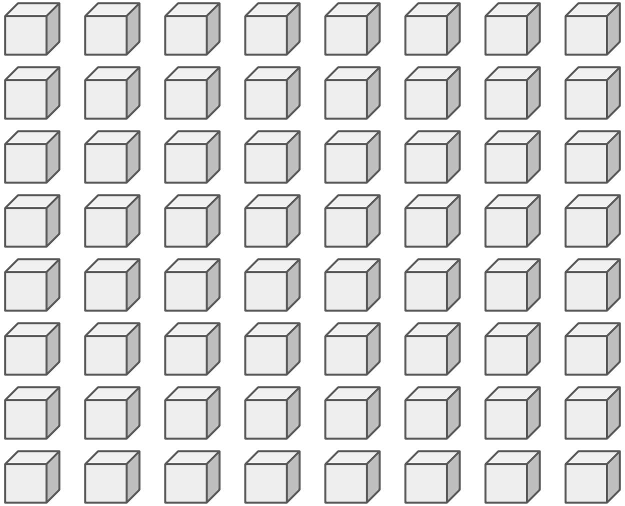




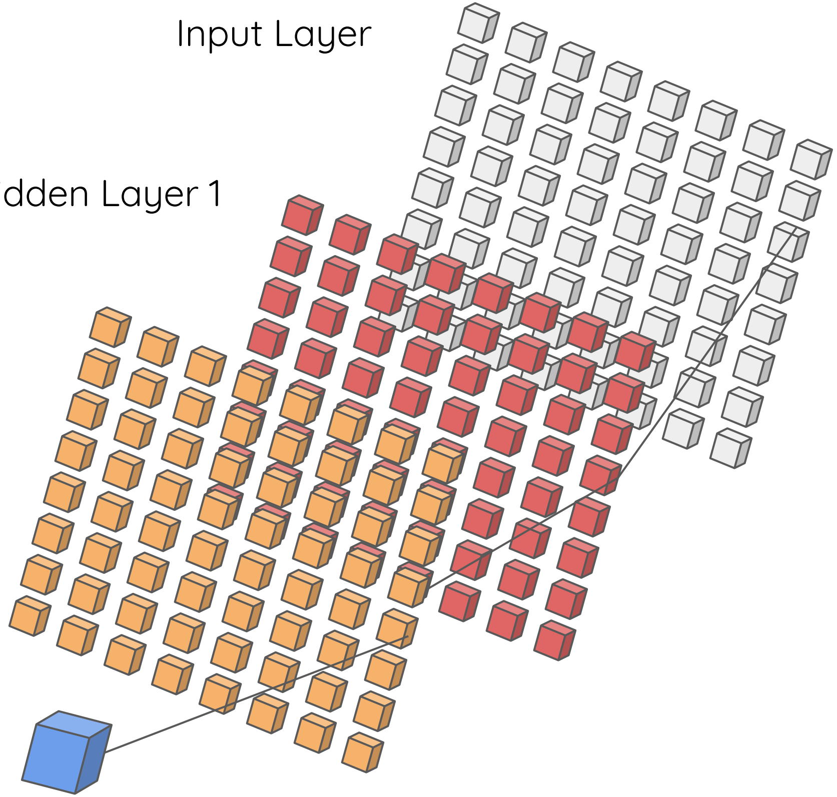



AA
AB
AC
AD
ZW
ZX
ZY
ZZ
Input Layer
Convolution Layers
Max Pooling Layers
Fully Connected Layer
Output Layer

Input Layer



Example of non-convolution net

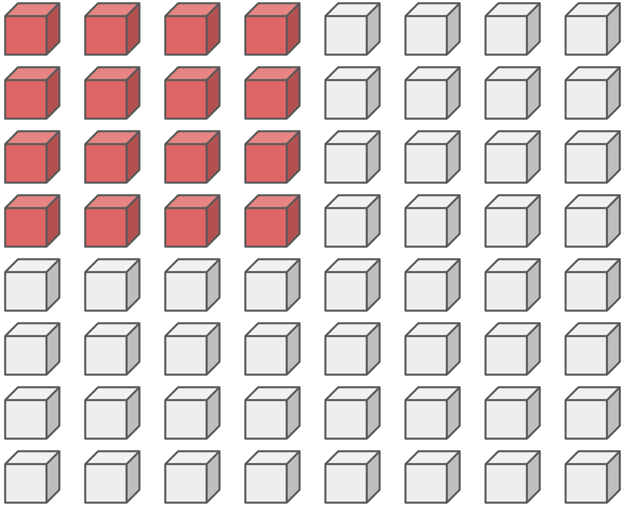


Input Layer

Hidden Layer 1

Hidden Layer 2

4 by 4

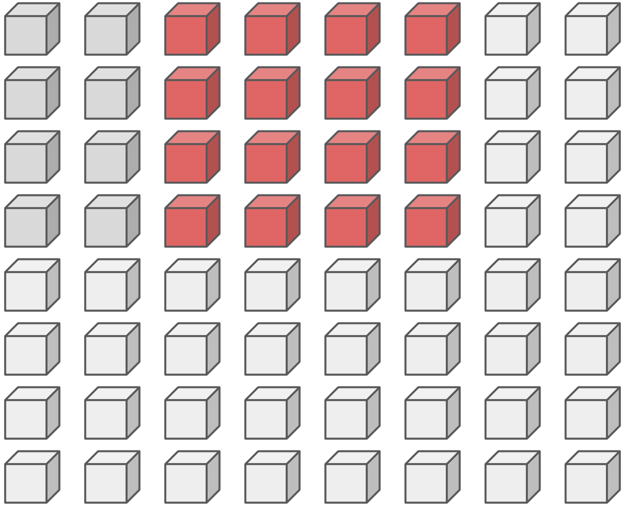


Convolution

Input Layer

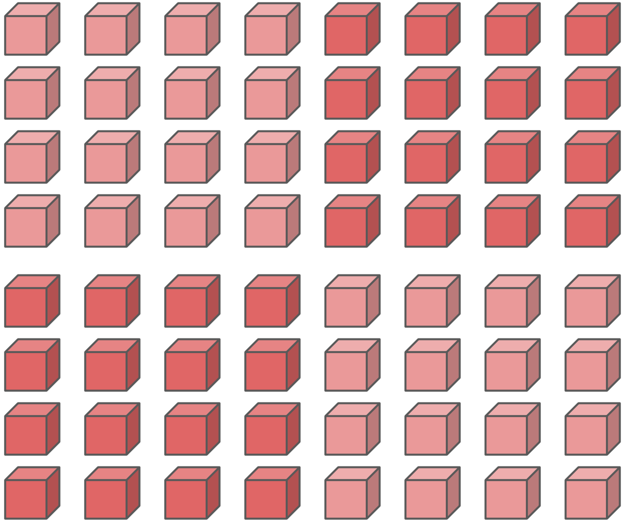
Stride by 2

Convolution



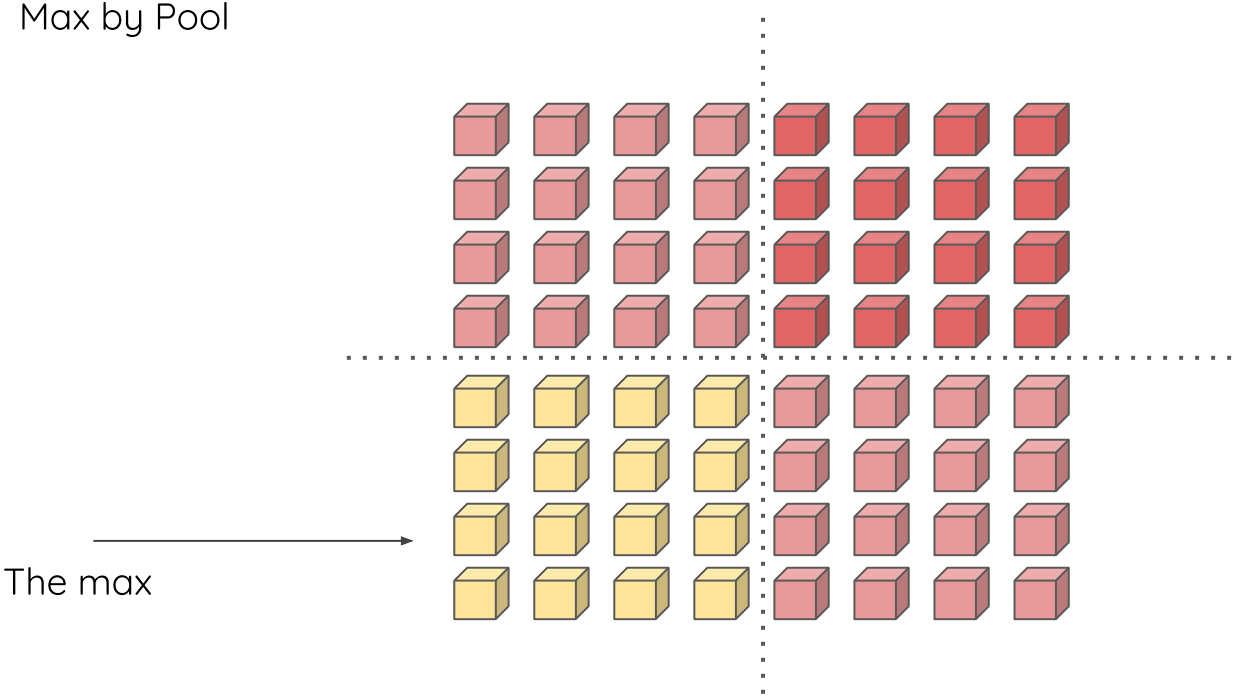
Input Layer

Convolution Output

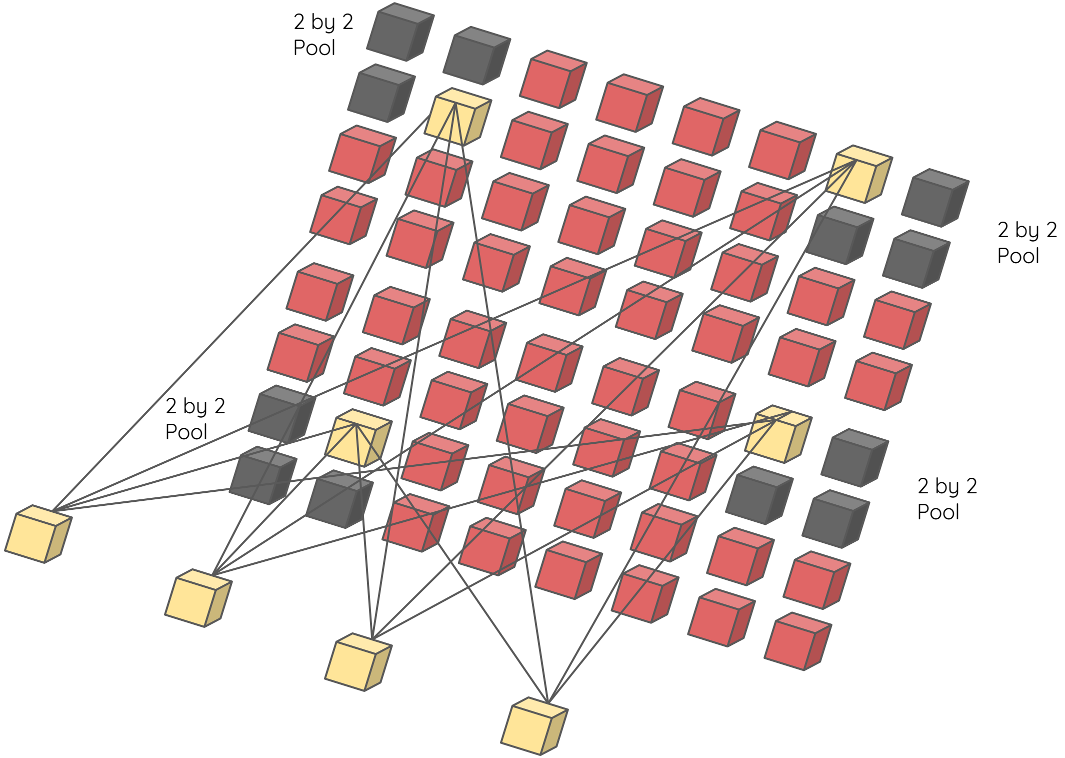


Sort of 4 outputs

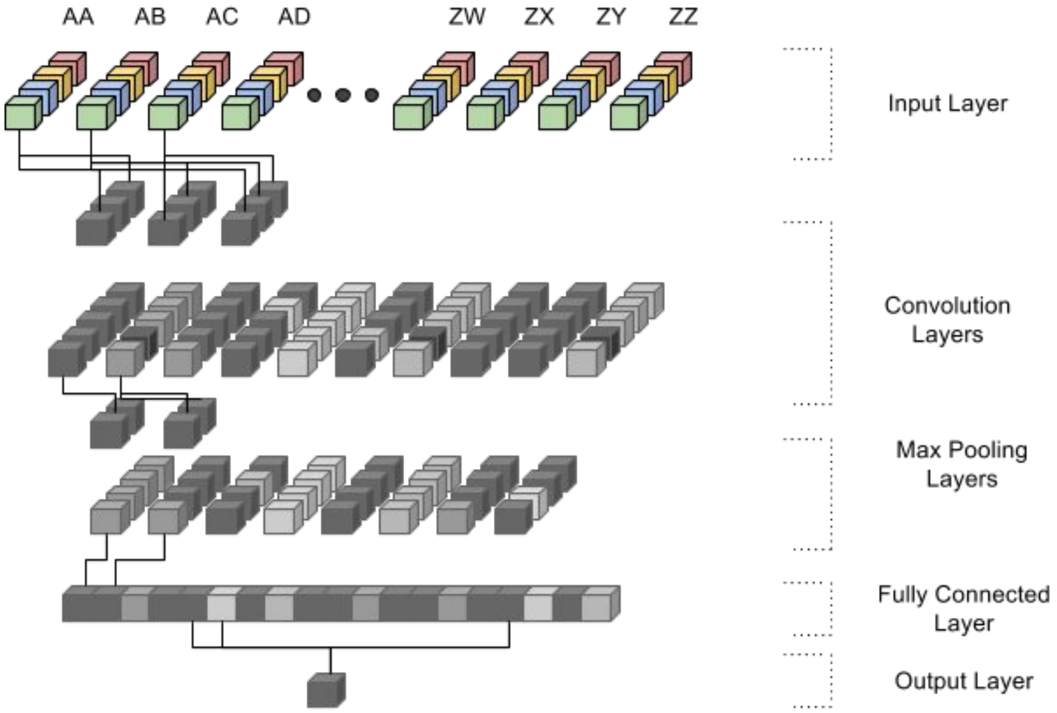
Max by Pool



2 by 2
Pool



Fully Connected Layer


AA
AB
AC
AD
ZW
ZX
ZY
ZZ
Input Layer
Convolution Layers
Max Pooling Layers
Fully Connected Layer
Output Layer

Conclusion

- Introduce Requester Geo-Popularity Data
- Introduce Exponential Moving Averages (EMAs)
- Convolutional Neural Networks on Geo-Popularity

Questions?