Scientific Computing for Information Security

Forging the missing link

Ryan Elkins

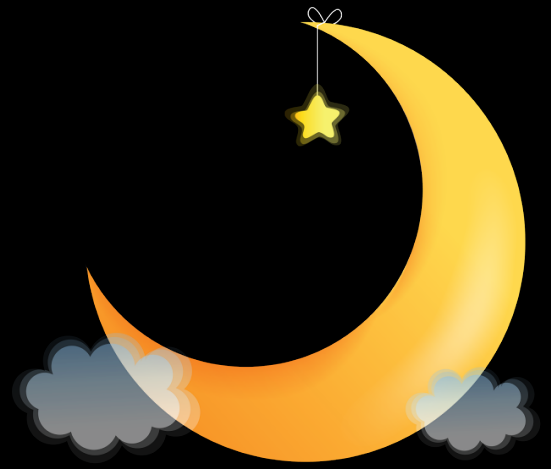
@ryanelkins

ryan-elkins@outlook.com

Information Security Architect

Ryan Elkins

~~@the_ryan_elkins~~
@ryanelkins



Security Researcher

My views are my own.

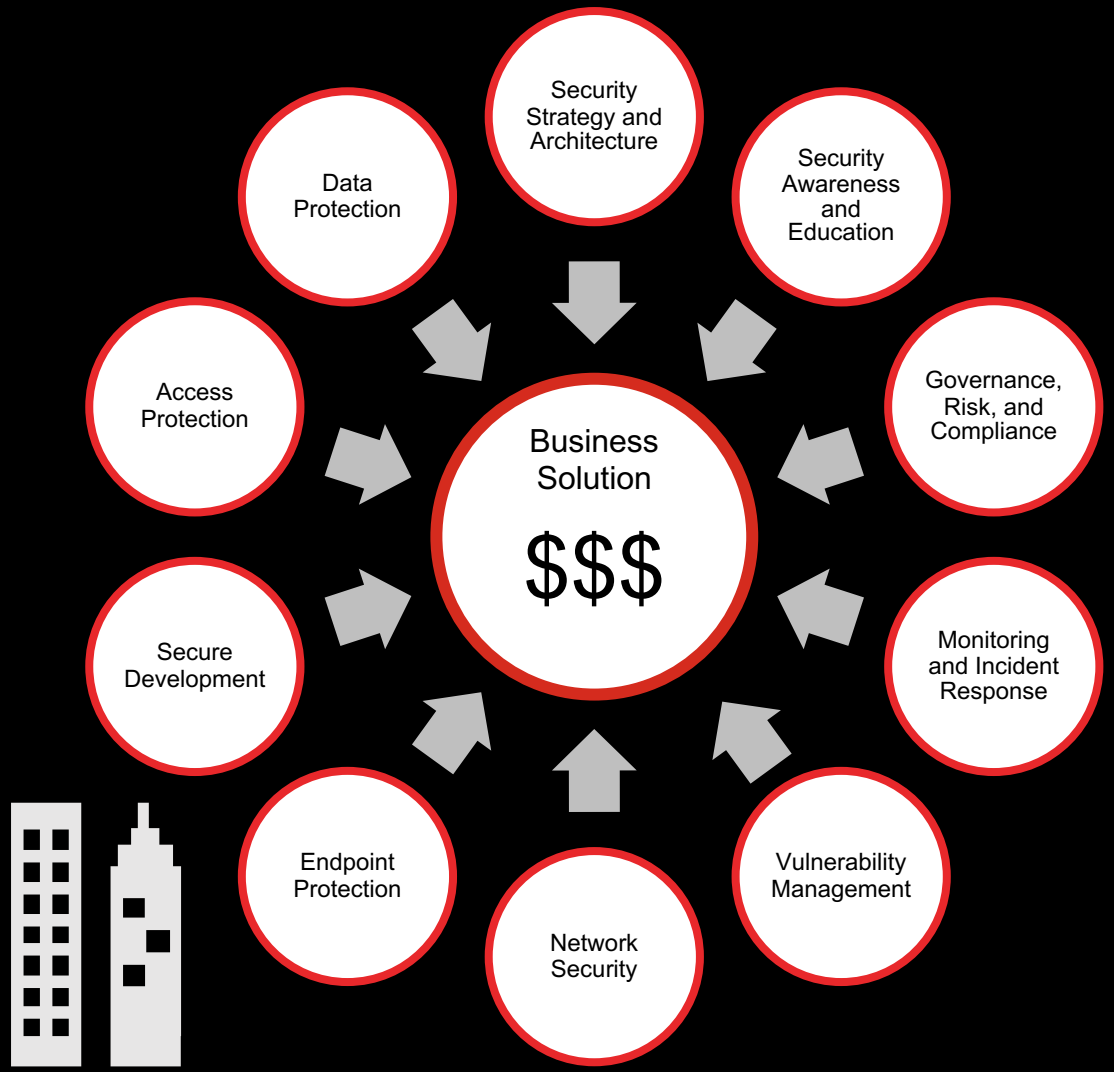
Thank you!

- Jeremiah Grossman (Influence/Reference)
- HD Moore (Project Sonar)
- Evan Perotti (AWS API)
- Dan Kaminsky (Influence/Reference)
- Justin Bollinger (crt.sh automation)
- Toby Hendricks (Jupyter and data science)

David Rowan – Wired – Atech 2017

Security by Design

Modular software, unopinionated deployment.

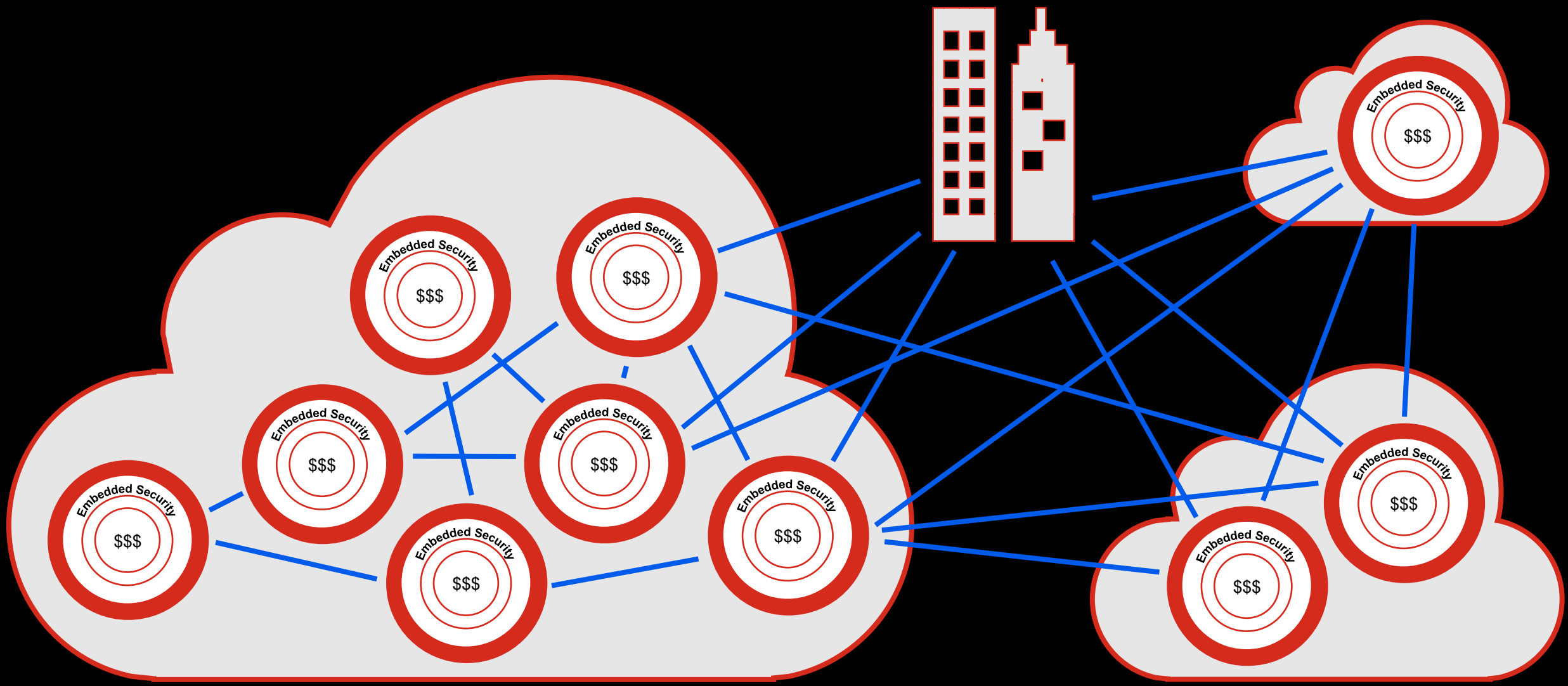


Bolt-on Security

Embedded Security

Hybrid Multi-Cloud

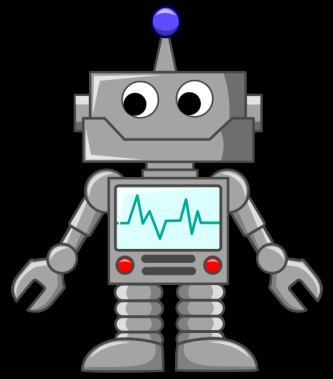
An interconnected, decentralized ecosystem.



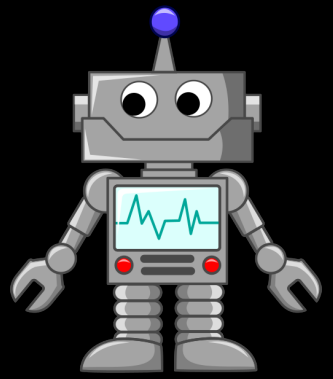
Robotic Process Automation

Eliminate manual processes.

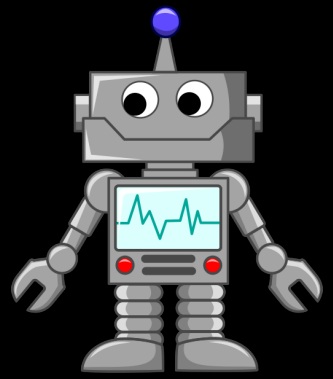
Systems Administration



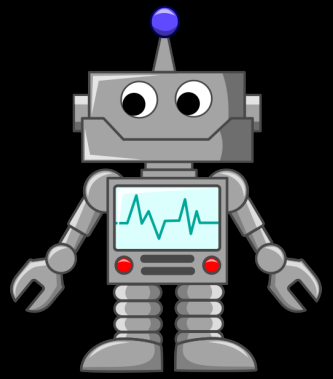
DevOps



Engineering

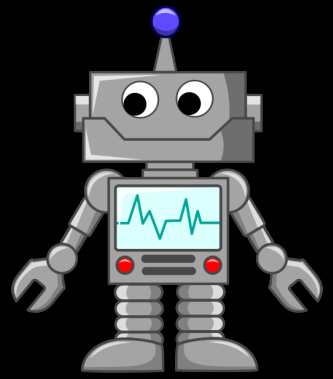


Quality Assurance

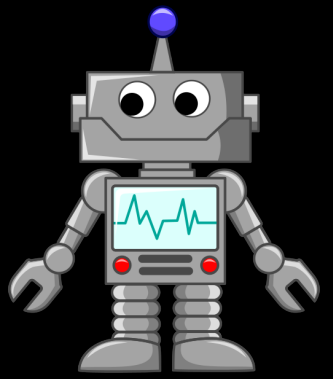


Information Security

Production Support



Lifecycle Management



Jeremiah Grossman @jeremiahg · Aug 7

Through actuarial data, there may come a day when cyber-insurance carriers conclude entire classes of security controls dont matter/work at all to reduce financial loss from breaches. And specific named products.

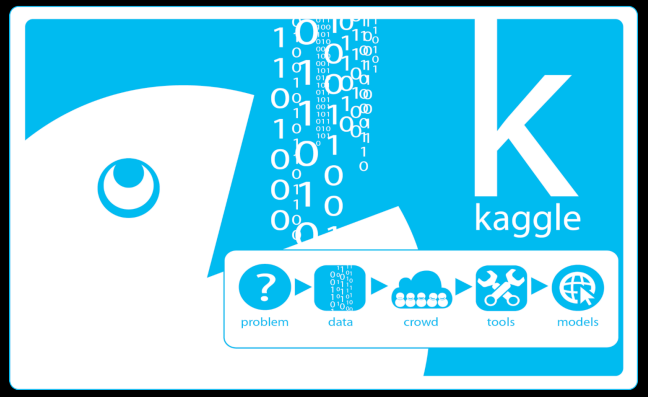
Resulting in the killing of entire product segments and companies.

11 32 77

Three major shifts driving mass adoption.

Democratized Accessibility

Increasing number of public data sets.



Google Dataset Search Beta

Search for Datasets

RAPID7 Open Data

Open Data

Offering researchers and community members open access to data from Project Sonar, which conducts internet-wide surveys to gain insights into global exposure to common vulnerabilities.

DATASETS: 13 FILES: 14,263 TOTAL SIZE: 27.4 TB

AWS Public Dataset Program

The AWS Public Dataset Program covers the cost of storage for publicly available high-value cloud-optimized datasets.

GitHub Repos

Github

5 months 3 TB 6.5 BigQuery

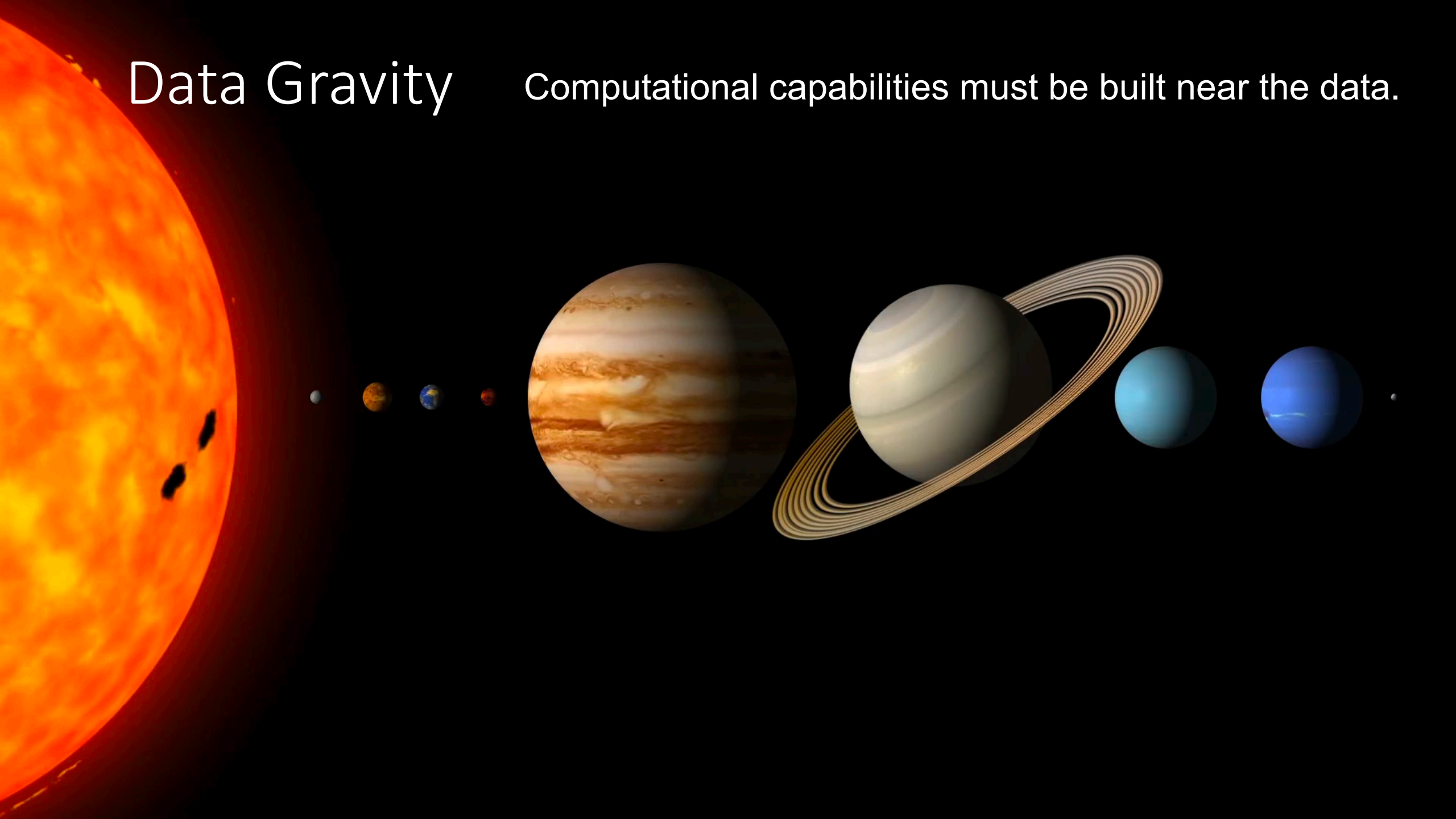
Common Crawl

encyclopedic internet machine learning natural language processing

A corpus of web crawl data composed of over 25 billion web pages.

Data Gravity

Computational capabilities must be built near the data.



Financially Viable

No longer cost prohibitive.

Google colab

Welcome to Colaboratory!

Colaboratory is a free Jupyter notebook environment that requires no setup and runs entirely in the cloud.

With Colaboratory you can write and execute code, save and share your analyses, and access powerful computing resources, all for free from your browser.

amazon web services

AWS Free Tier

Gain free, hands-on experience with the AWS platform, products, and services

Anaconda Distribution

The open-source Anaconda Distribution is the easiest way to perform Python/R data science and machine learning on Linux, Windows, and Mac OS X. With over 15 million users worldwide, it is the industry standard for developing, testing, and training on a single machine, enabling *individual data scientists* to:

Let’s get started!

Jupyter Notebook?

?

?

Jupyter Notebook?

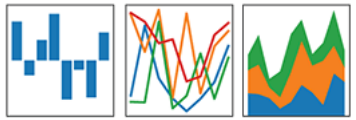
Jupyter – Julia, Python, and R

- Open source web application for interactive computing.
- Supports over 40 programming languages.
- Notebooks contain live code, equations, visualizations, and narrative text.

Seaborn

pandas

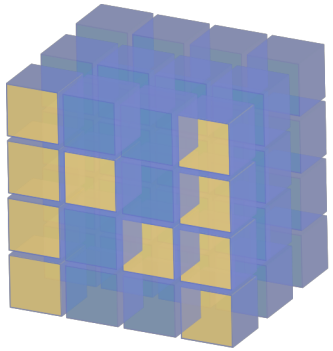
$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



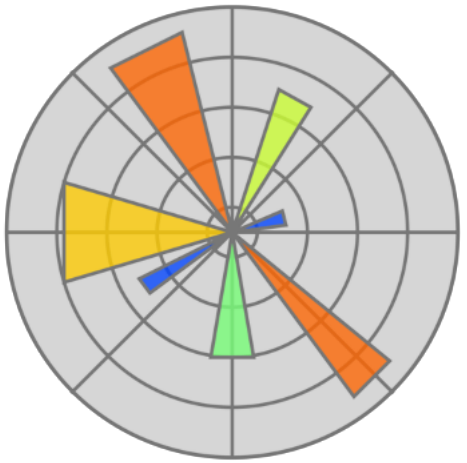
jupyter

IP[y]:

IPython



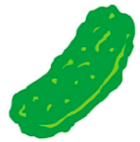
NumPy



bokeh

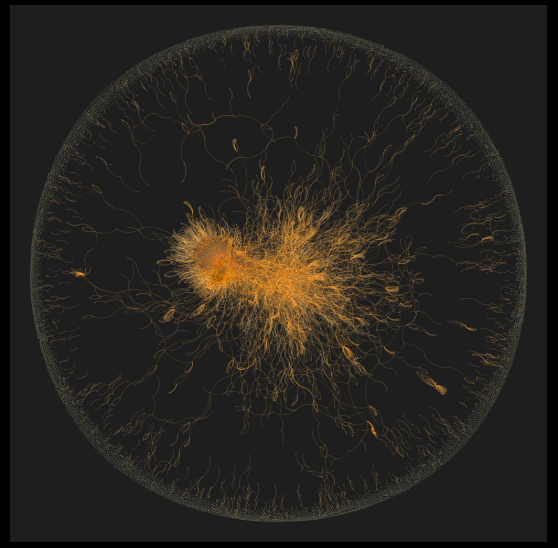
DASK

matplotlib

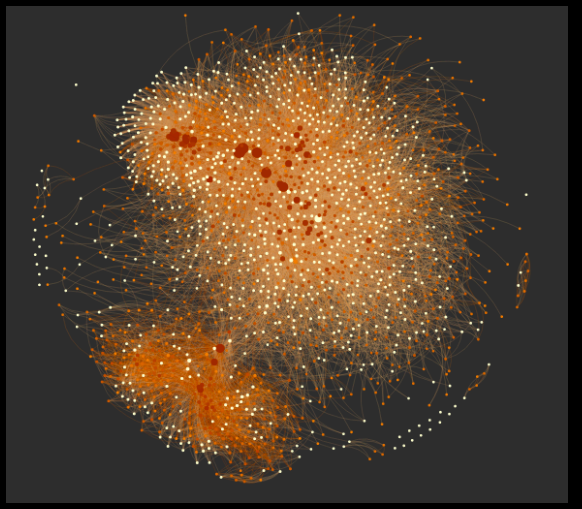


pickleDB

Security experts love to tell stories.



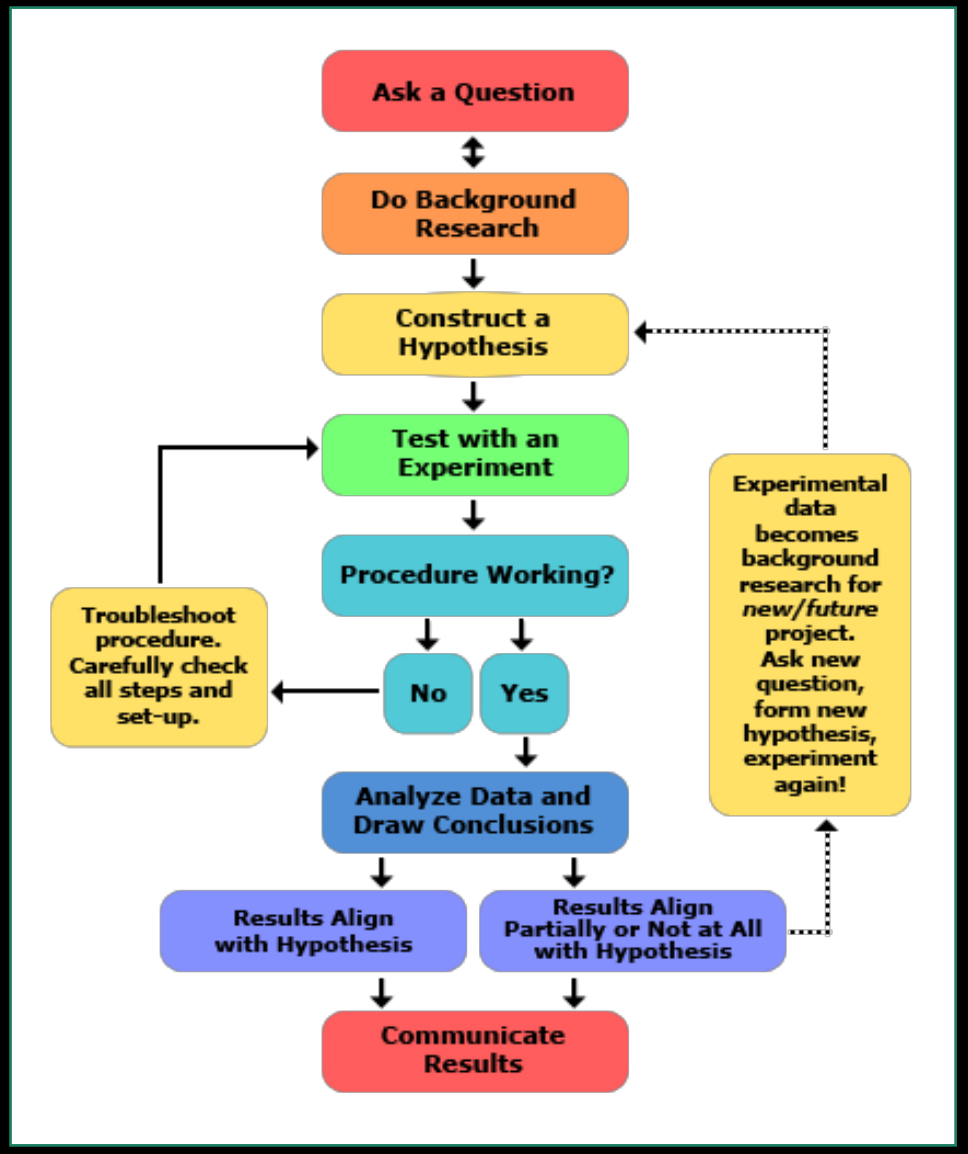
Data visualizations help us tell these stories.



5th Grade Science

Scientific Method

*The **scientific method** is an **empirical** method of acquiring knowledge that has characterized the development of **science** since at least the 17th century.*



Asking the right questions.

Sharing the right insights.

Asking the right questions.

- Adversary: How can I perform passive reconnaissance without touching the target?

Sharing the right insights.

Asking the right questions.

- Adversary: How can I perform passive reconnaissance without touching the target?
- CISO: What does our external attack surface look like?

Sharing the right insights.

Asking the right questions.

- Adversary: How can I perform passive reconnaissance without touching the target?
- CISO: What does our external attack surface look like?
- CEO: What does our company look like to our shareholders, customers, and competitors?

Sharing the right insights.

Passive Recon Demo

RAPID7 Open Data

Open Data

Offering researchers and community members open access to data from Project Sonar, which conducts internet-wide surveys to gain insights into global exposure to common vulnerabilities.

DATASETS: 13 FILES: 14,263 TOTAL SIZE: 27.4 TB

http://securityriskadvisors.com/blog/creating-a-project-sonar-fdns-api-with-aws/ - Evan Perotti

17766	2019-07-27 16:42:43.000
17767	2019-07-27 16:45:10.000
17768	2019-07-27 16:36:42.000
17769	2019-07-27 16:38:06.000
17770	2019-07-27 16:36:53.000
17771	2019-07-27 04:33:12.000

17772 rows × 4 columns

	latitude	longitude	count
0	-37.807	144.952	1
1	-36.851	174.768	1
2	-33.859	151.200	2
3	-33.810	151.131	4
4	-33.494	143.210	7
5	-22.831	-43.219	8
6	1.293	103.855	229
7	1.367	103.800	6
8	13.084	80.281	1
9	17.378	78.471	13
10	18.533	73.863	2
11	18.972	72.825	1
12	20.000	77.000	23
13	22.291	114.150	145
14	29.425	-98.493	335

Name	date_201907
Description	
Database	rapid7fqdns
Classification	parquet
Location	s3://rapid7-opendata/fdns/any/v1/date=201907/
Connection	
Deprecated	No
Last updated	Sat Aug 31 23:48:48 GMT-400 2019
Input format	org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat
Output format	org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat
Serde serialization lib	org.apache.hadoop.hive.ql.io.parquet.serde.ParquetHiveSerDe
Serde parameters	serialization.format 1
Table properties	sizeKey 45714202983 objectCount 153 UPDATED_BY_
	recordCount 2945369954 averageRecordSize 48 Craw

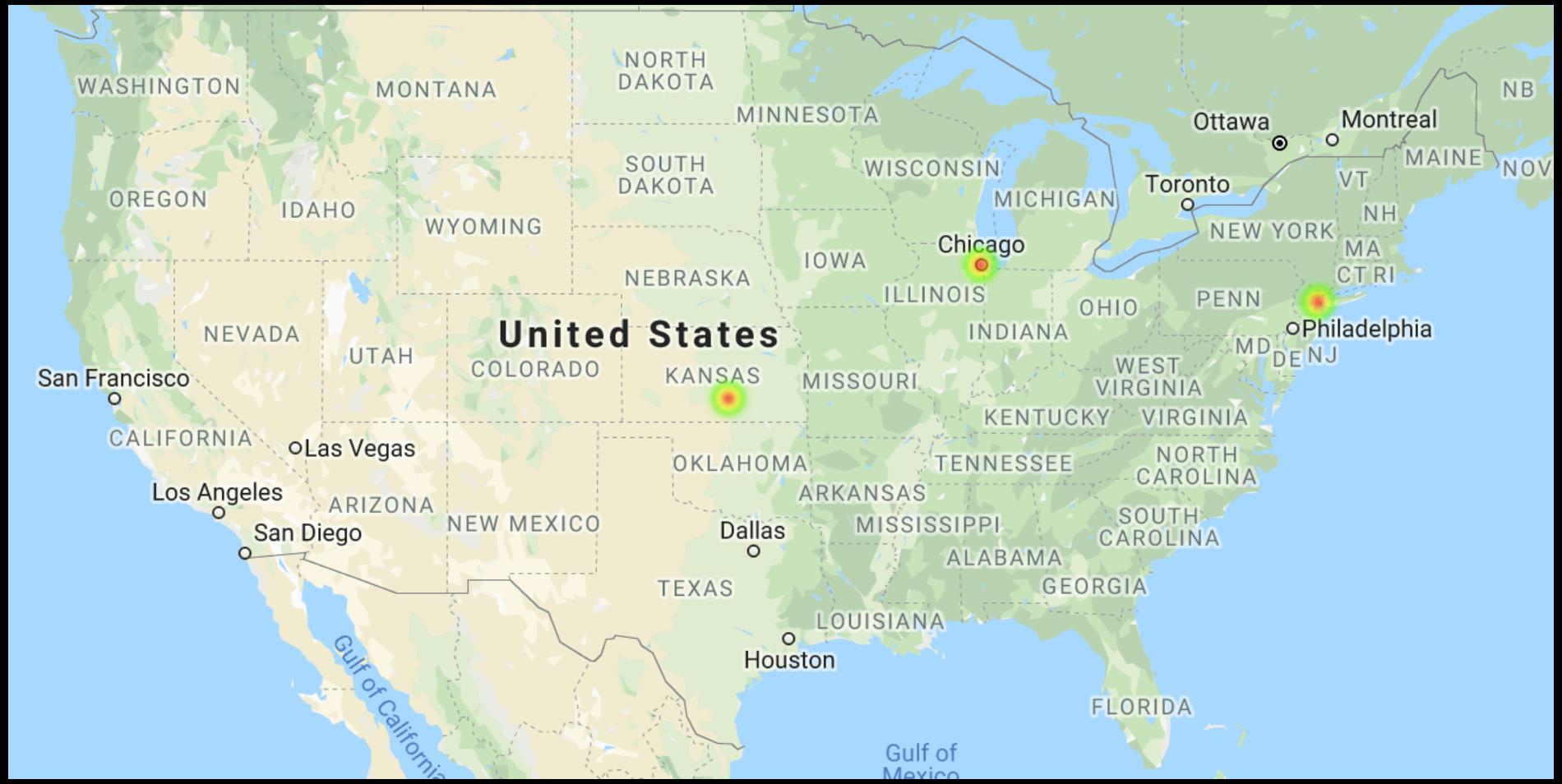

Russia
Norway
United Kingdom
Canada
Poland
Germany
Ukraine
Kazakhstan
Mongolia
France
Italy
Spain
Turkey
China
Japan
South Korea
United States
North Atlantic Ocean
Afghanistan
Iraq
Iran
Pakistan
Algeria
Libya
Egypt
Saudi Arabia
India
Thailand
Mexico
Mali
Niger
Sudan
Chad
Nigeria
Ethiopia
Venezuela
Colombia
Kenya
DRC
Indonesia
Papua New Guinea
Tanzania
Brazil
Peru
Angola
Bolivia
Namibia
Madagascar
Botswana
Indian Ocean
Chile
South Atlantic Ocean
Australia
South Africa
Argentina


Russia
Norway
United Kingdom
Canada
Poland
Germany
Ukraine
Kazakhstan
Mongolia
France
Italy
Spain
Turkey
China
Japan
South Korea
United States
Afghanistan
Iraq
Iran
Pakistan
North Atlantic Ocean
Algeria
Libya
Egypt
Saudi Arabia
India
Thailand
Mexico
Mali
Niger
Sudan
Chad
Nigeria
Ethiopia
Venezuela
Colombia
Kenya
DRC
Indonesia
Papua New Guinea
Tanzania
Brazil
Peru
Angola
Bolivia
Namibia
Madagascar
Indian Ocean
Botswana
Chile
South Atlantic Ocean
Australia
South Africa
Argentina


Iceland
Sweden
Norway
Russia
United Kingdom
Canada
Poland
Germany
Ukraine
Kazakhstan
Mongolia
France
Italy
Spain
Turkey
China
Japan
South Korea
United States
North Atlantic Ocean
Iraq
Iran
Afghanistan
Pakistan
Algeria
Libya
Egypt
Saudi Arabia
India
Thailand
Mexico
Mali
Niger
Sudan
Chad
Nigeria
Ethiopia
Venezuela
Colombia
Kenya
DRC
Indonesia
Papua New Guinea
Tanzania
Brazil
Peru
Angola
Bolivia
Namibia
Madagascar
Botswana
Indian Ocean
South Pacific Ocean
South Atlantic Ocean
Chile
Australia
South Africa
Argentina
New Zealand


Iceland
Sweden
Norway
Russia
United Kingdom
Canada
Poland
Germany
Ukraine
Kazakhstan
Mongolia
France
Italy
Spain
Turkey
China
South Korea
Japan
United States
North Atlantic Ocean
Afghanistan
Iraq
Iran
Pakistan
Algeria
Libya
Egypt
Saudi Arabia
India
Thailand
Mexico
Mali
Niger
Sudan
Chad
Nigeria
Ethiopia
Venezuela
Colombia
Kenya
DRC
Indonesia
Papua New Guinea
Tanzania
Brazil
Peru
Angola
Bolivia
Namibia
Madagascar
Indian Ocean
Botswana
Chile
South Atlantic Ocean
Australia
South Africa
Argentina
New Zealand


Russia
Norway
United Kingdom
Canada
Poland
Germany
Ukraine
Kazakhstan
Mongolia
France
Italy
Spain
Turkey
China
Japan
South Korea
United States
North Atlantic Ocean
Afghanistan
Iraq
Iran
Pakistan
India
Thailand
Algeria
Libya
Egypt
Saudi Arabia
Mexico
Mali
Niger
Sudan
Chad
Nigeria
Ethiopia
Venezuela
Colombia
Kenya
DRC
Indonesia
Tanzania
Brazil
Peru
Angola
Bolivia
Namibia
Madagascar
Indian


WASHINGTON
MONTANA
NORTH DAKOTA
MINNESOTA
NB
Ottawa
Montreal
SOUTH DAKOTA
WISCONSIN
MAINE
NOV
OREGON
IDAHO
WYOMING
MICHIGAN
Toronto
VT
NH
NEW YORK
Chicago
IOWA
MA
CT RI
NEBRASKA
ILLINOIS
PENN
OHIO
United States
NEVADA
UTAH
INDIANA
Philadelphia
MD
DE
NJ
COLORADO
KANSAS
MISSOURI
WEST VIRGINIA
San Francisco
KENTUCKY
VIRGINIA
CALIFORNIA
Las Vegas
OKLAHOMA
TENNESSEE
NORTH CAROLINA
Los Angeles
ARKANSAS
ARIZONA
NEW MEXICO
MISSISSIPPI
SOUTH CAROLINA
San Diego
Dallas
ALABAMA
TEXAS
GEORGIA
LOUISIANA
Houston
FLORIDA
Gulf of California
Gulf of Mexico

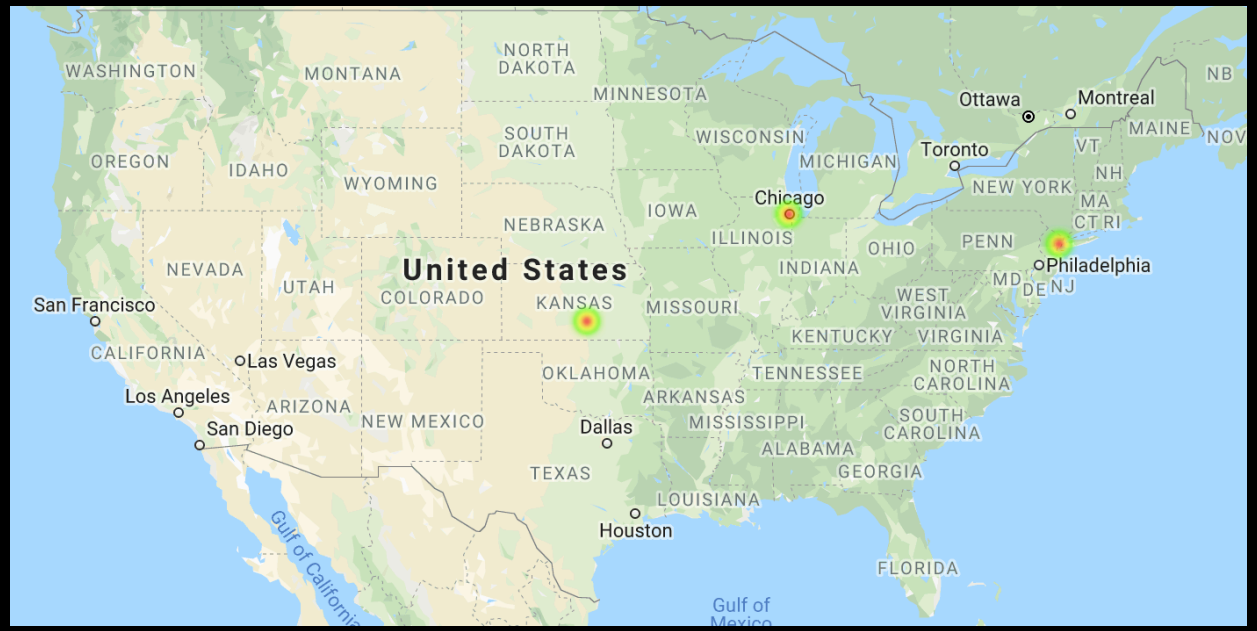
Derbycon.com

- Passive recon

```
In [7]: # Load an external notebook with normalized functions
        %run ./centralprocessing.ipynb
        df_min = get_location(dfhosts, 'value')
        df_min.head(10)
```

Out[7]:

	timestamp	name	value	type	latitude	longitude	country	state
0	2019-07-27 17:02:06.000	ns2.derbycon.com	198.241.11.53	a	37.751	-97.822	United States	None
1	2019-07-27 17:02:06.000	ns2.derbycon.com	2620:111:8001::53	aaaa	37.751	-97.822	United States	None
2	2019-07-27 17:03:11.000	ns3.derbycon.com	50.31.242.53	a	41.878	-87.638	United States	Chicago
3	2019-07-27 17:03:11.000	ns3.derbycon.com	2620:111:8002::53	aaaa	37.751	-97.822	United States	None
4	2019-07-27 17:16:15.000	ns4.derbycon.com	50.31.243.53	a	41.878	-87.638	United States	Chicago
5	2019-07-27 17:16:15.000	ns4.derbycon.com	2620:111:8003::53	aaaa	37.751	-97.822	United States	None
6	2019-07-27 19:14:36.000	scoreboard.ctf.derbycon.com	45.79.178.217	a	40.739	-74.170	United States	Newark
7	2019-07-27 16:59:09.000	ns1.derbycon.com	198.241.10.53	a	37.751	-97.822	United States	None
8	2019-07-27 16:59:09.000	ns1.derbycon.com	2620:111:8000::53	aaaa	37.751	-97.822	United States	None
9	2019-07-27 22:03:51.000	www.derbycon.com	45.79.164.10	a	40.739	-74.170	United States	Newark

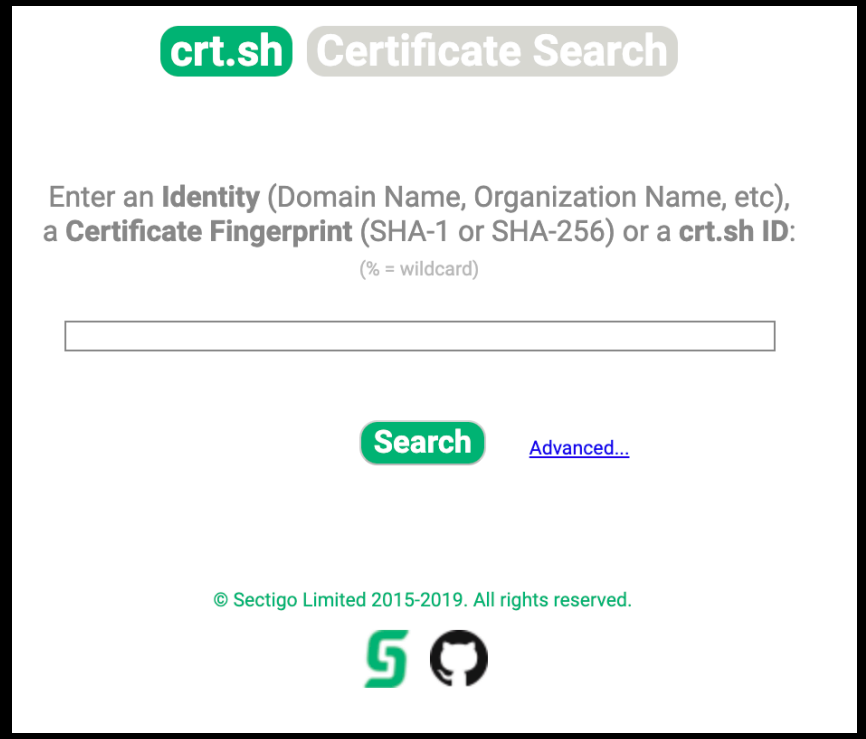



```
In [8]: # Load an external notebook with normalized functions
        %run ./centralprocessing.ipynb
        df_plot = prepare_location(df_min)
        df_plot.head(50)
```

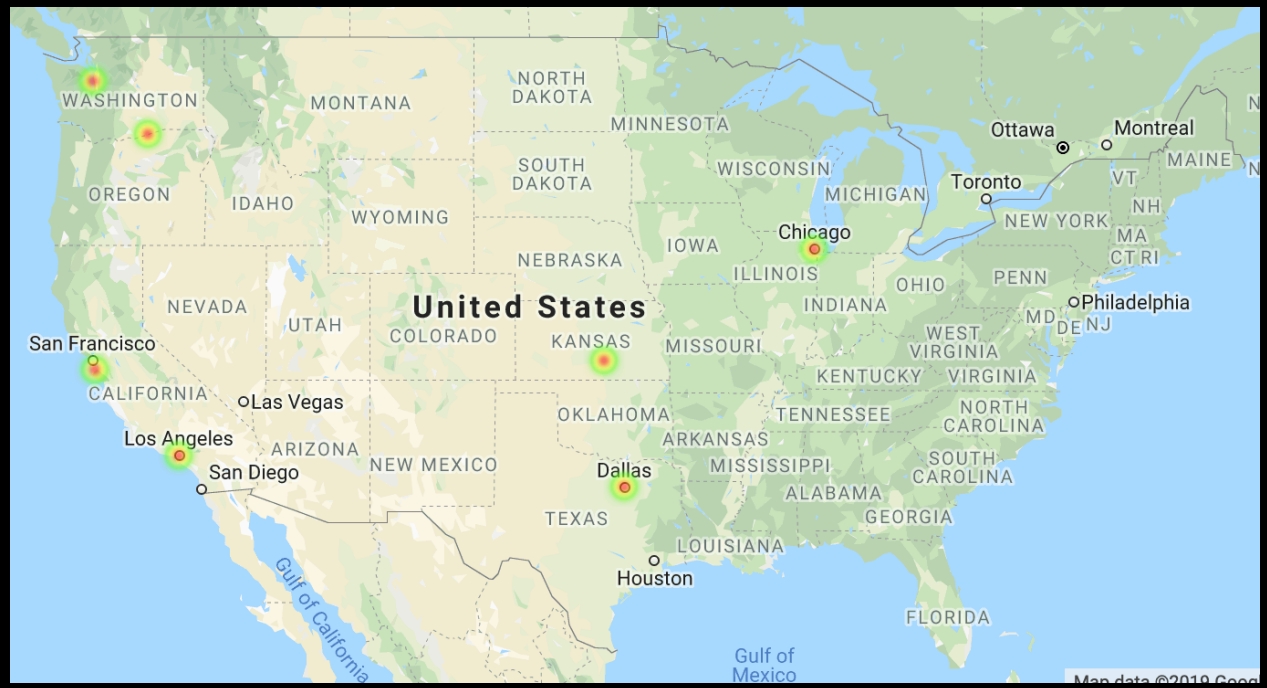
Out[8]:

	latitude	longitude	count
0	37.751	-97.822	6
1	40.739	-74.170	2
2	41.878	-87.638	2

Certificate Transparency Reporting



DomainName	addr	latitude	longitude	country	state
e25589.a.akamaiedge.net	205.185.204.29	32.789	-96.802	United States	Dallas
ab664b09fb79211e9acfe06505df8d1b-725225292.us-...	52.40.155.19	45.849	-119.714	United States	Boardman
tesla-2.sso.global.akadns.net	199.66.9.46	37.751	-97.822	United States	None
e16359.dscx.akamaiedge.net	23.198.246.244	37.751	-97.822	United States	None
e9056.b.akamaiedge.net	23.197.161.144	37.751	-97.822	United States	None
toolbox.tb.tesla.services	13.226.15.34	47.635	-122.345	United States	Seattle
mfamobile-dev.tesla.com	205.234.27.209	37.425	-122.296	United States	Redwood City
desk.cs.zohohost.com	8.39.54.74	37.751	-97.822	United States	None
e1792.dscx.akamaiedge.net	23.198.228.52	37.751	-97.822	United States	None
click.emails.tesla.com	13.111.48.179	37.751	-97.822	United States	None

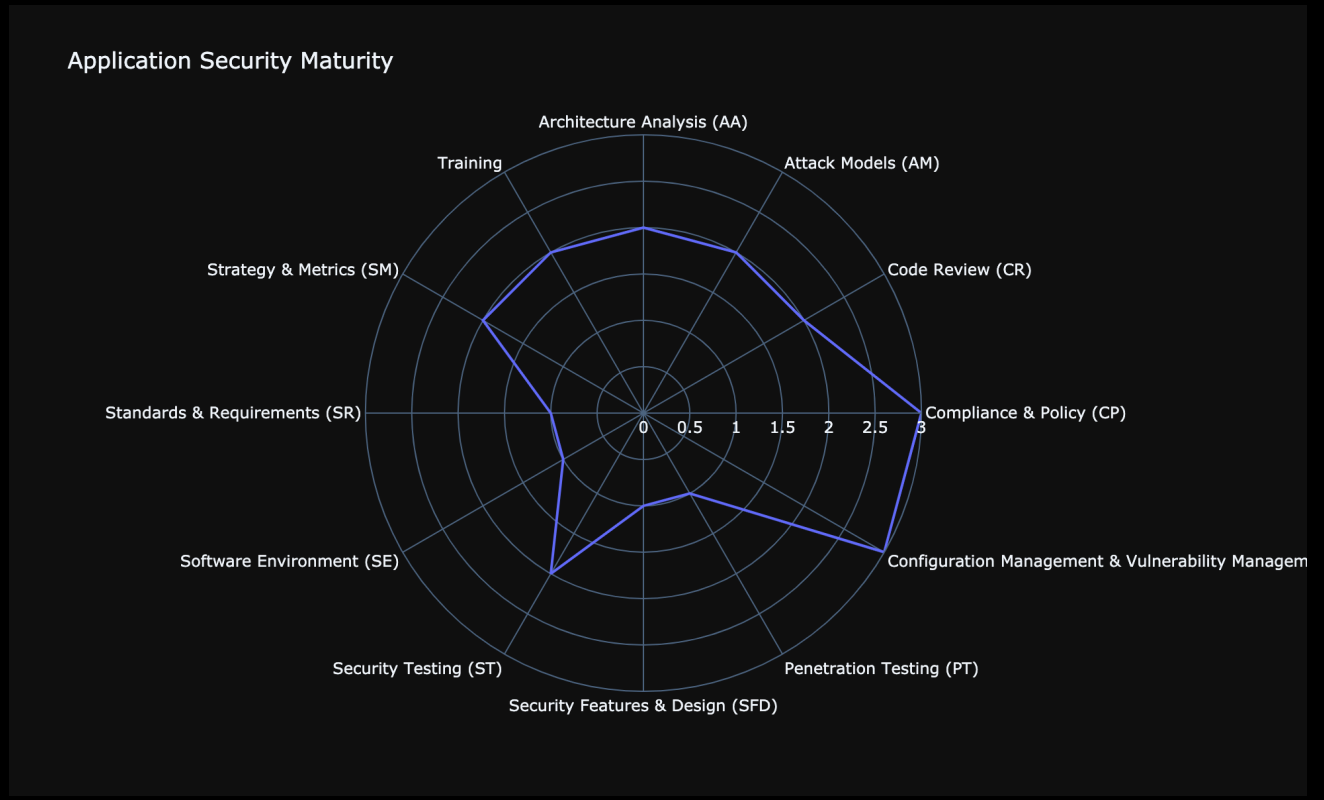


Special thanks to Justin Bollinger @Bandrel for developing the script to enumerate and parse crt.sh.

Program Value/Measurement

"Security technology management keeps the system fine tuned. But the secret sauce? That's in data enrichment. That's where the magic happens."

- Robert Herjavec



Category	Section	Activity Description	Activity	Level	Participant Percentage		Current	Target
SSDL Touchpoints	Security Testing (ST)	Ensure QA supports edge/boundary value cond	ST1.1	1	83.3%		TRUE	TRUE
SSDL Touchpoints	Security Testing (ST)	Drive tests with security requirements and secu	ST1.3	1	73.3%		TRUE	TRUE
SSDL Touchpoints	Security Testing (ST)	Integrate black-box security tools into the QA p	ST2.1	2	25.0%		FALSE	TRUE
SSDL Touchpoints	Security Testing (ST)	Share security results with QA.	ST2.4	2	11.7%		FALSE	TRUE
SSDL Touchpoints	Security Testing (ST)	Include security tests in QA automation.	ST2.5	2	10.0%		FALSE	FALSE
SSDL Touchpoints	Security Testing (ST)	Perform fuzz testing customized to application	ST2.6	2	10.8%		TRUE	TRUE
SSDL Touchpoints	Security Testing (ST)	Drive tests with risk analysis results.	ST3.3	3	3.3%		FALSE	FALSE
SSDL Touchpoints	Security Testing (ST)	Leverage coverage analysis.	ST3.4	3	2.5%		FALSE	FALSE
SSDL Touchpoints	Security Testing (ST)	Begin to build and apply adversarial security te	ST3.5	3	2.5%		FALSE	FALSE
Deployment	Penetration Testing (PT)	Use external penetration testers to nd problem	PT1.1	1	87.5%		TRUE	TRUE
Deployment	Penetration Testing (PT)	Feed results to the defect management and mit	PT1.2	1	74.2%		TRUE	TRUE

Asking the right questions.

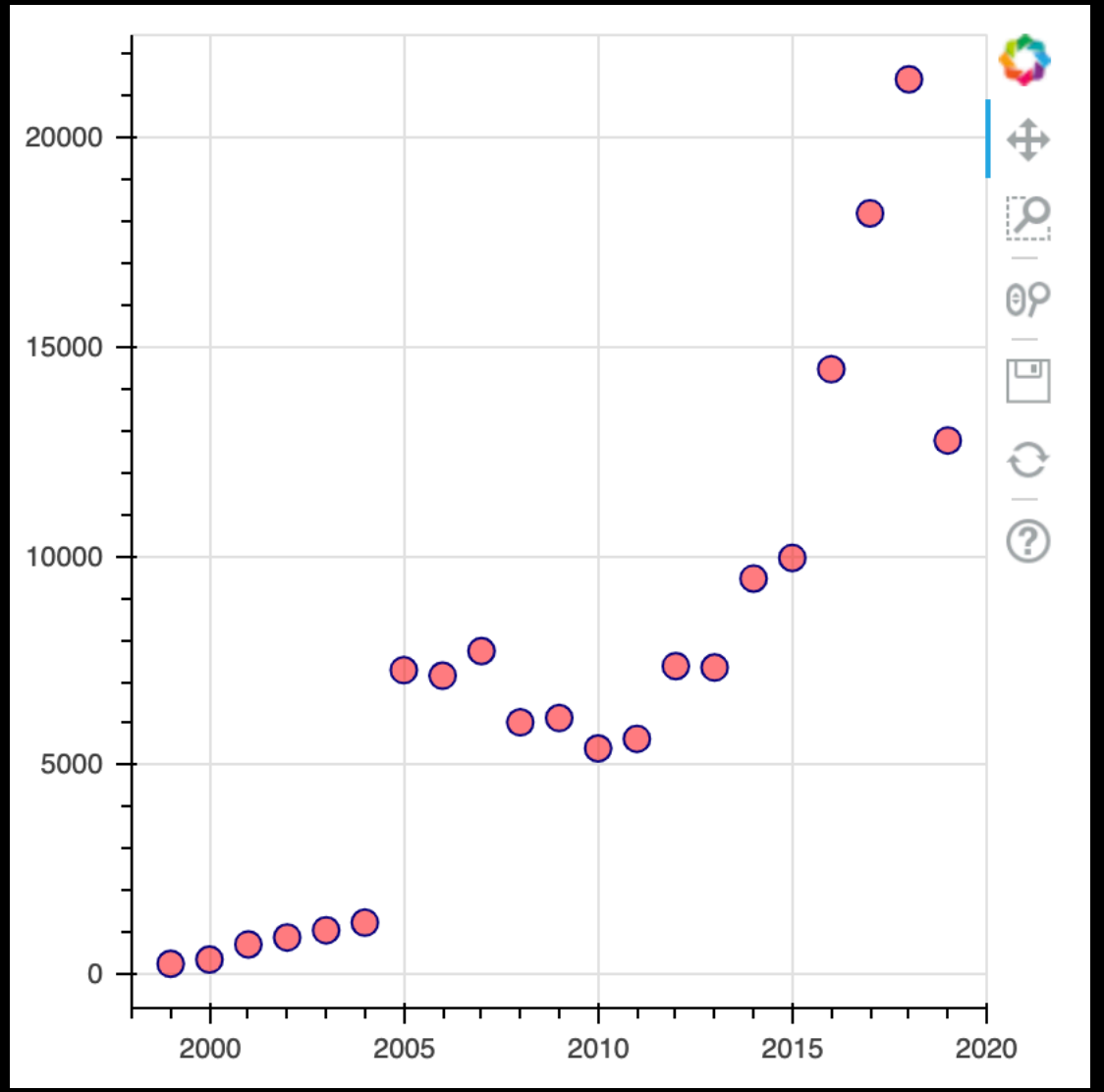
- Patch Management Team: Why are we so busy over the holidays?
- Security Manager: At what times do we need to increase staffing?
- CISO: Why are our vulnerability counts increasing?

Sharing the right insights.

Program Value/Measurement

Common Vulnerabilities & Exposures (CVEs)

- Growth over the past 20 years.



Month	Count
1	16600
2	11716
3	12446
4	10229
5	10760
6	12331
7	11862
8	12071
9	11282
10	10181
11	12374
12	18952

Year	Count
1999	249
2000	351
2001	713
2002	881
2003	1053
2004	1234
2005	7274
2006	7142
2007	7731
2008	6026
2009	6129
2010	5402
2011	5632
2012	7371
2013	7337
2014	9466
2015	9960
2016	14473
2017	18202
2018	21409
2019	12769

Challenges to Solve

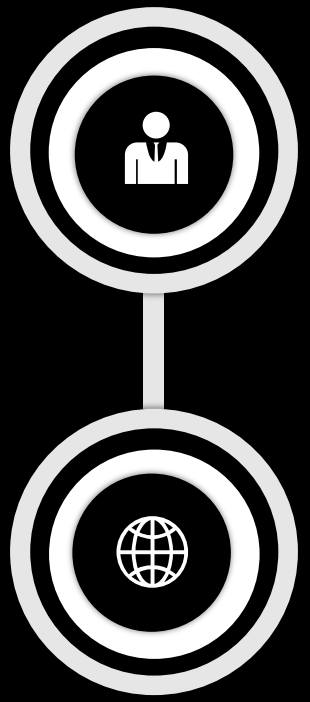
Expertise must keep up with an evolving attack surface and a dynamic nature of threats.

Competition for talent

“the average CISO tenure is only about 24 to 48 months”

- 38% leave for higher compensation
- 36% leave for lack of security culture
- 34% leave for not being active members with executive management or board of directors
- 31% leave from lack of budget

Challenges to Solve



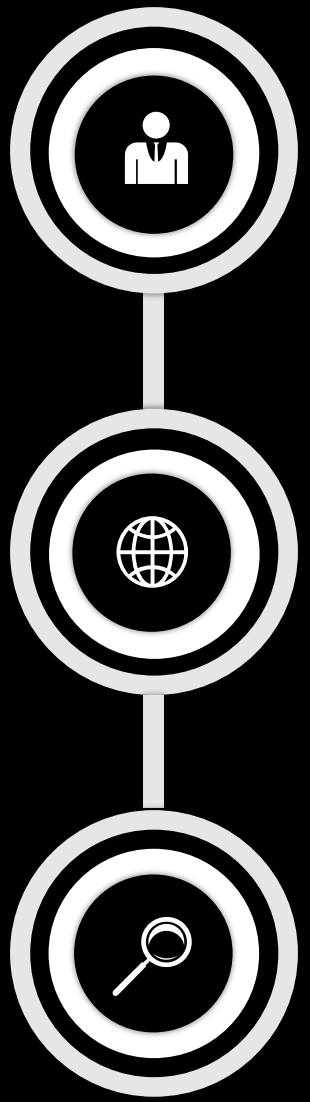
Expertise must keep up with an evolving attack surface and a dynamic nature of threats.

Competition for talent

Programs are reactive and rarely maintain a multi-year roadmap.

Program continuity

Challenges to Solve



Expertise must keep up with an evolving attack surface and a dynamic nature of threats.

Competition for talent

Programs are reactive and rarely maintain a multi-year roadmap.

Program continuity

Visibility and speed of execution must improve with automation.

Continuous risk measurement

Thank you!

- Name: Ryan Elkins
- Twitter: @ryanelkins
- Email: ryan-elkins@outlook.com