

# **Epoch Fail:**

## **Forecasting Vulnerabilities Amid Temporal Discontinuity**

**Dr. Ben Edwards, Cyentia Institute**

**Sander Vinberg, F5 Labs**

# Agenda

- A Brief History of the CVE
- Modeling Challenges and Solutions
- Analyzing the CVE Data Landscape
- Identifying structural and not so structural changes.
- Conclusion: Implications for Vulnerability Forecasting

# A Brief History of the CVE Framework

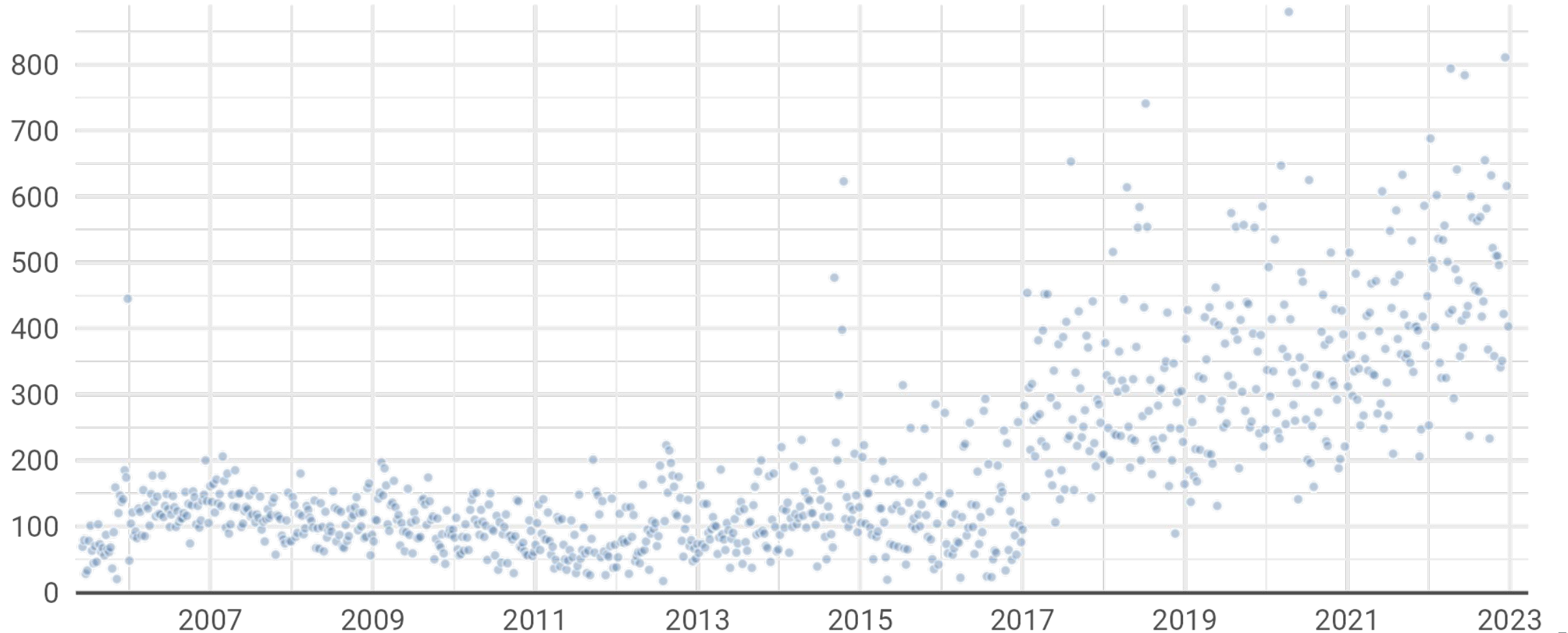
- **1999:** Dave Mann and Steve Christey present framework at a workshop. Later that year an initial list of 321 vulns is published
- **2000:** Two vuln lists emerge, now MITRE CVE and NIST ICAT
- **2001:** NIST defunds ICAT, SANS keeps it running w/ grad students
- **2004:** DHS funds ICAT, renames to NVD
- **2007:** CVSSv2 adopted, CWEs introduced
- **2008:** Common Product Enumeration (CPE) revised
- **2016:** CWEs revised
- **2016:** CVE Number Authorities (CNAs) introduced

# Source Data Considerations

- All CVE data analyzed here come from the National Vulnerability Database (NVD)
- 191,000 CVEs analyzed
  - Reserved/Rejected CVEs not included
- Timing based on **publication date**
- Data up through December 31, 2022

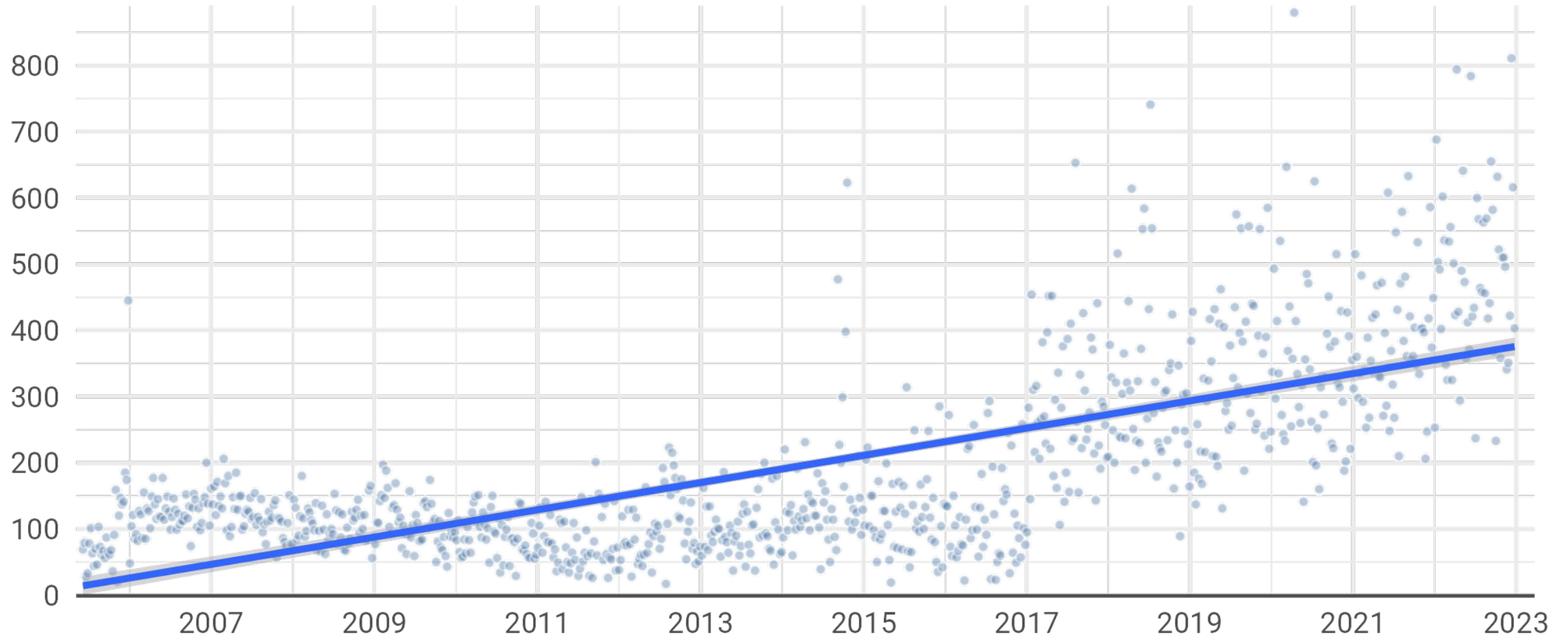
# Let's forecast some vulnerability volume

Number of CVEs published in NVD each week

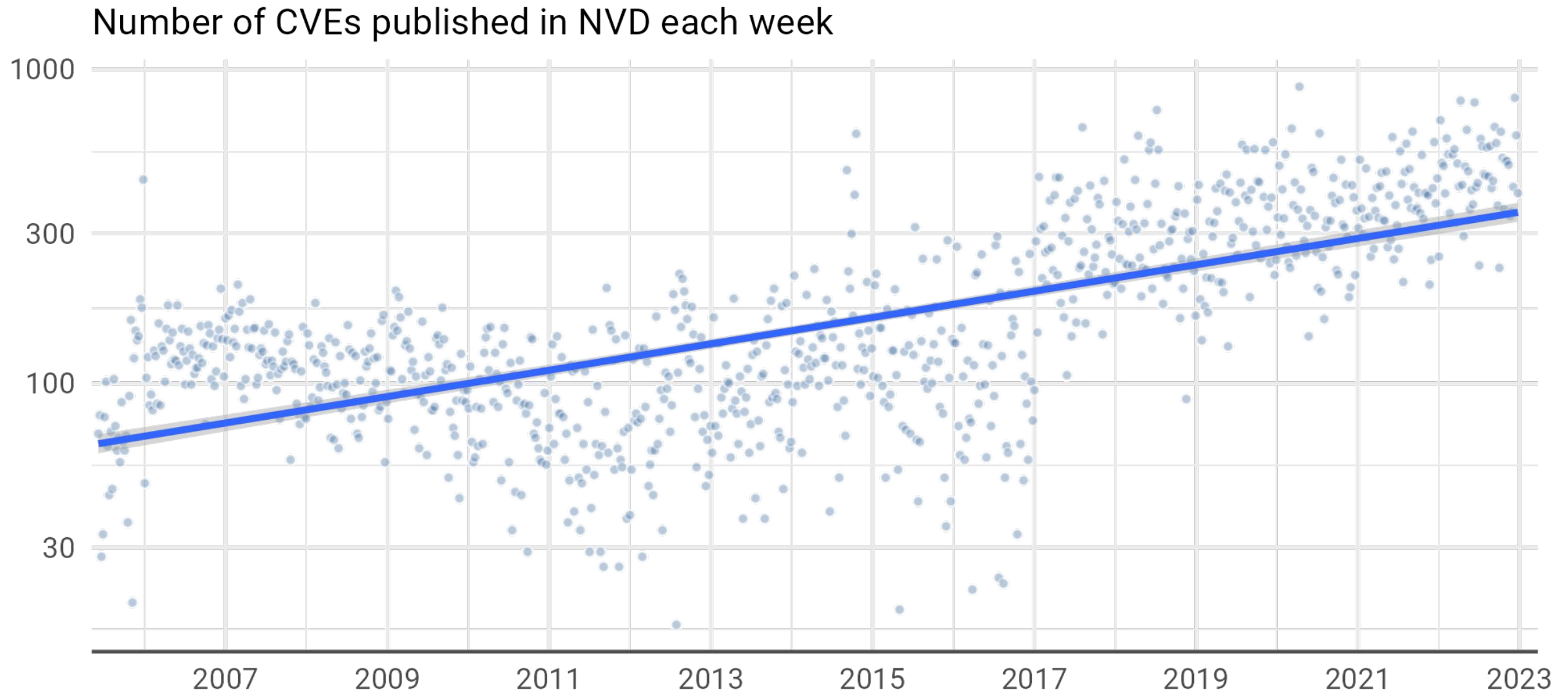


# Linear models aren't going to be enough

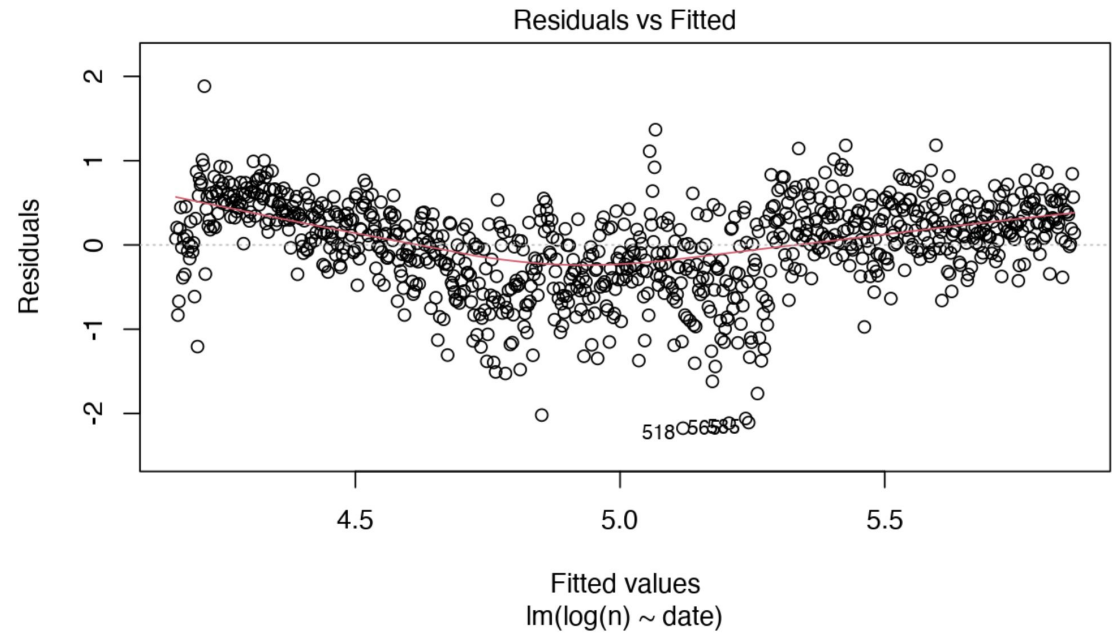
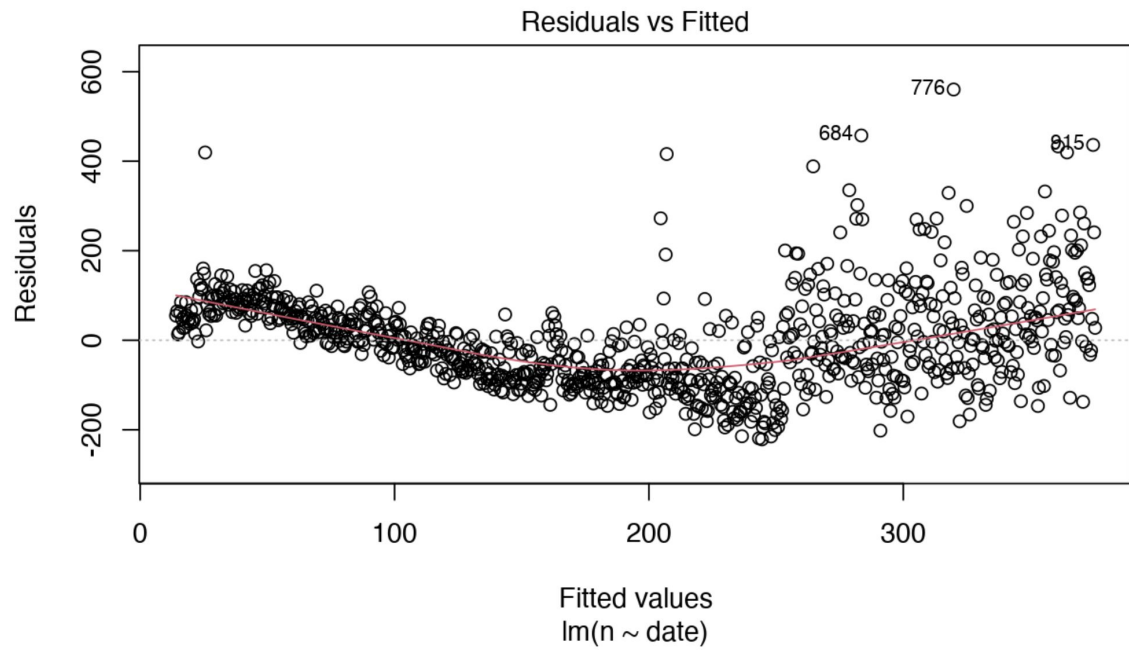
Number of CVEs published in NVD each week



# Log/Linear Models aren't going to be enough



# The fits, they are had



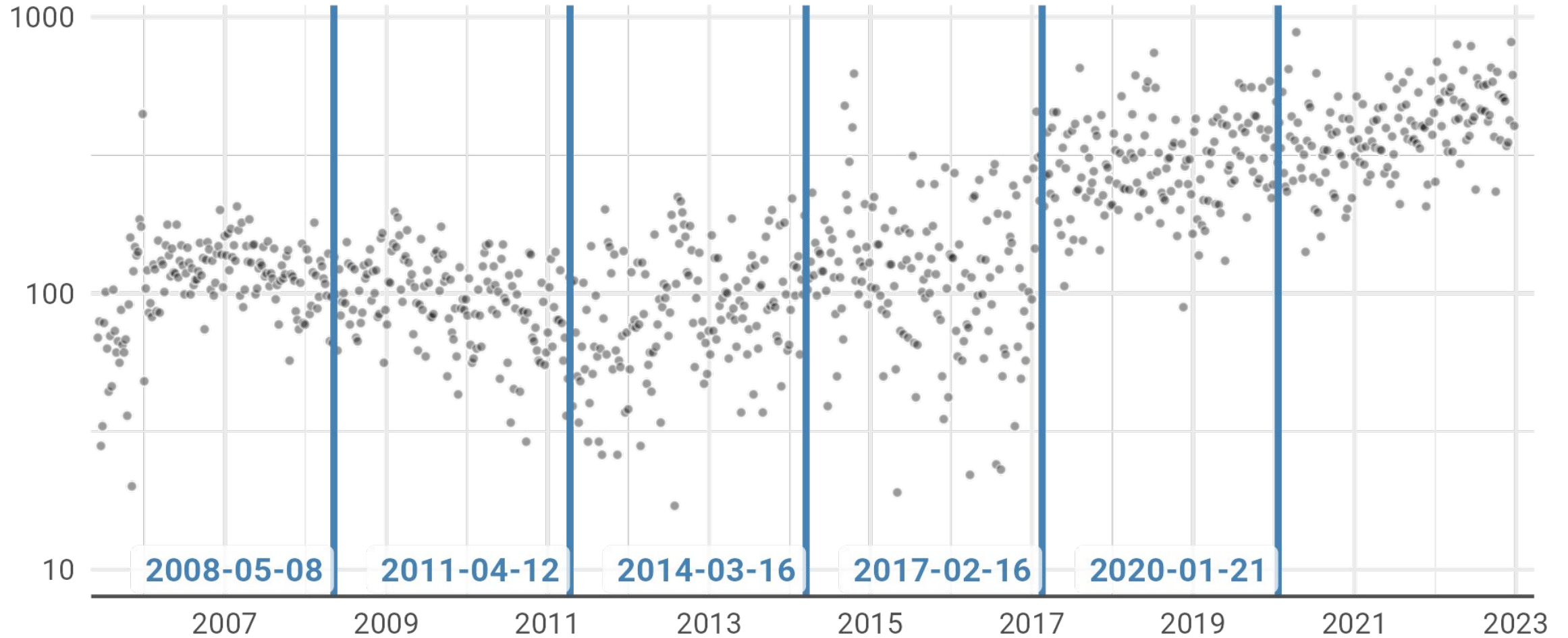


# Segmented Models

1. Break up the time series into pre-selected number of segments
2. Fit linear (or log linear) models within each segment
3. Constrain linear models to be continuous (but not differentiable)
4. Adjust segment break locations in such a way to maximize model log likelihood
5. Select best number of segments using model selection criteria
6. Interpret

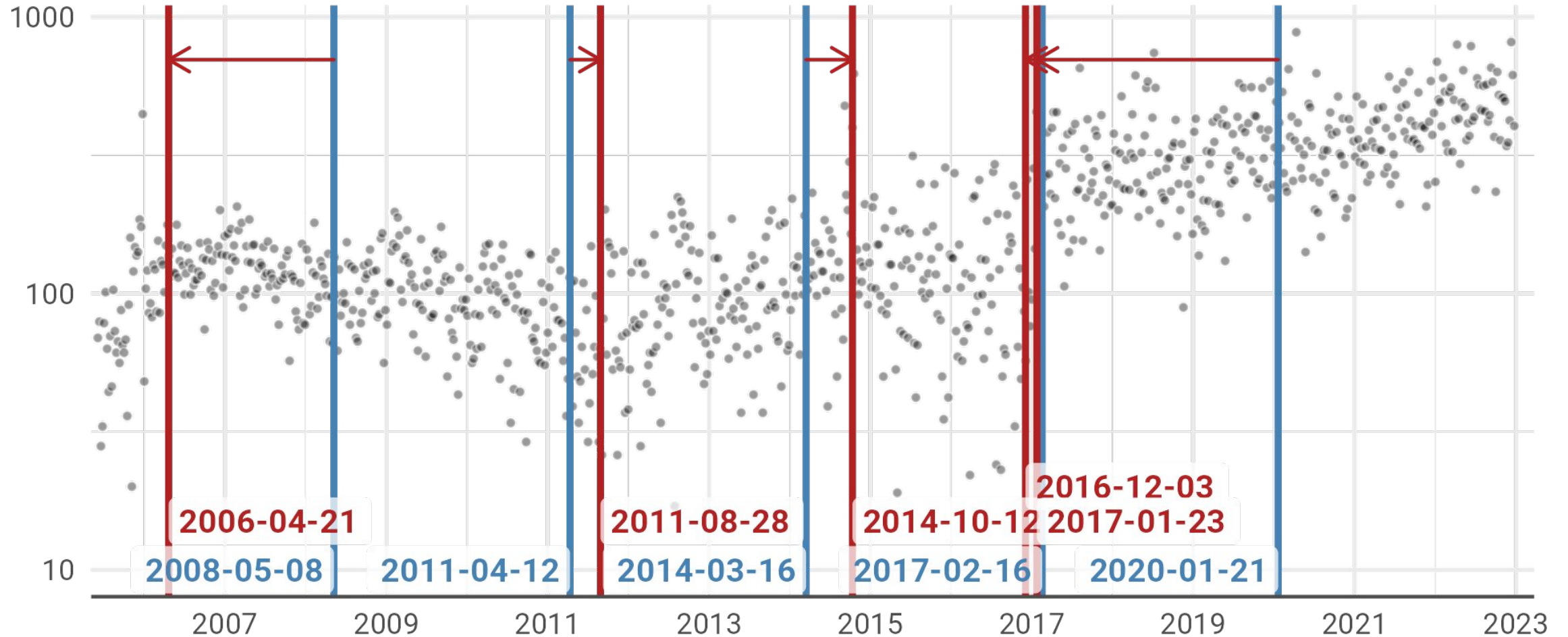
# Segmented Models: An example

Number of CVEs published in NVD each week



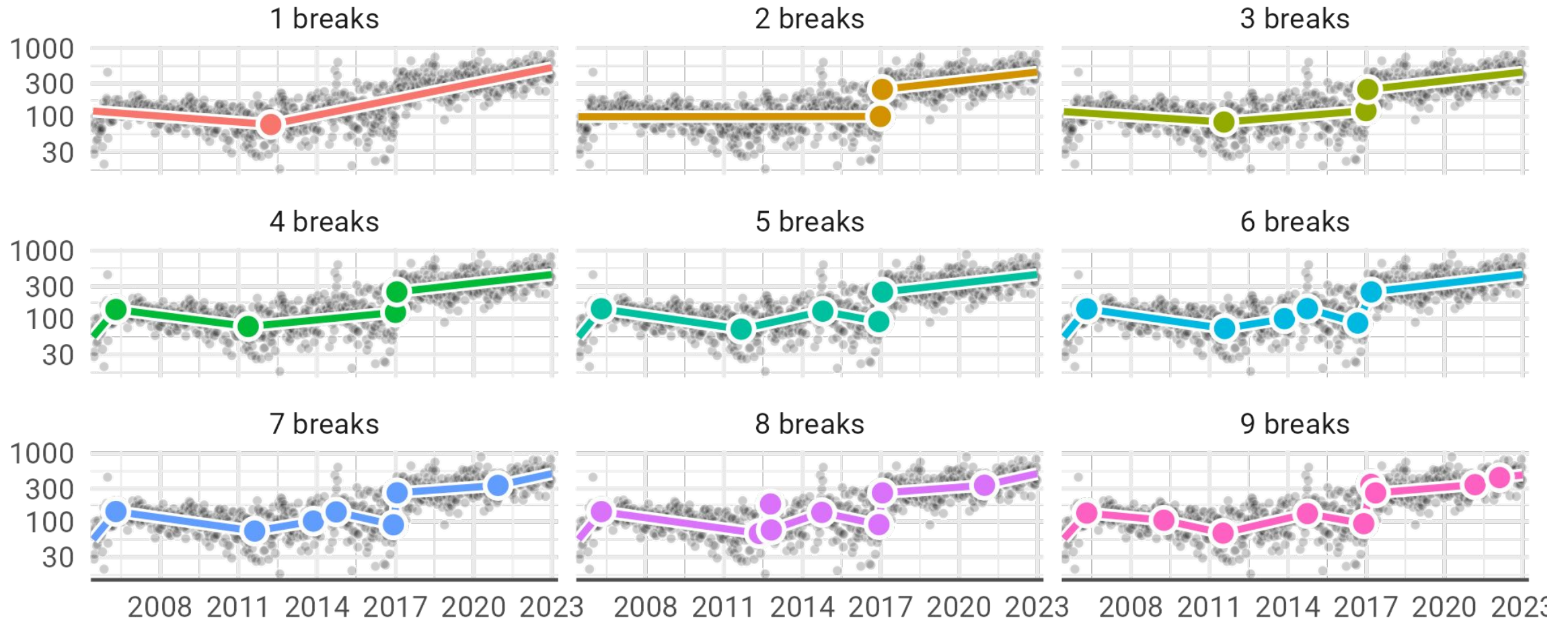
# Segmented Models: An example

Number of CVEs published in NVD each week

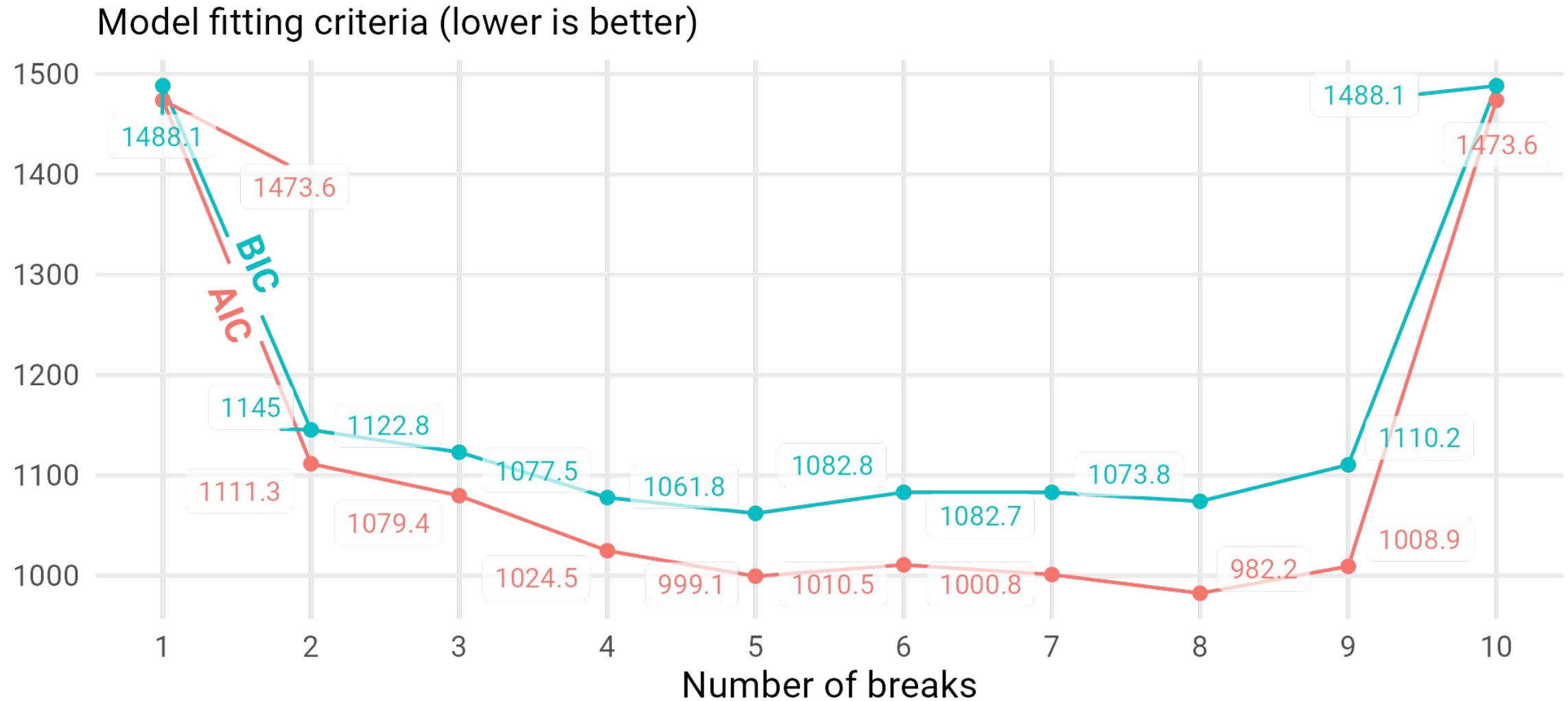


# Segmented Models: An example

Number of CVEs published in NVD each week

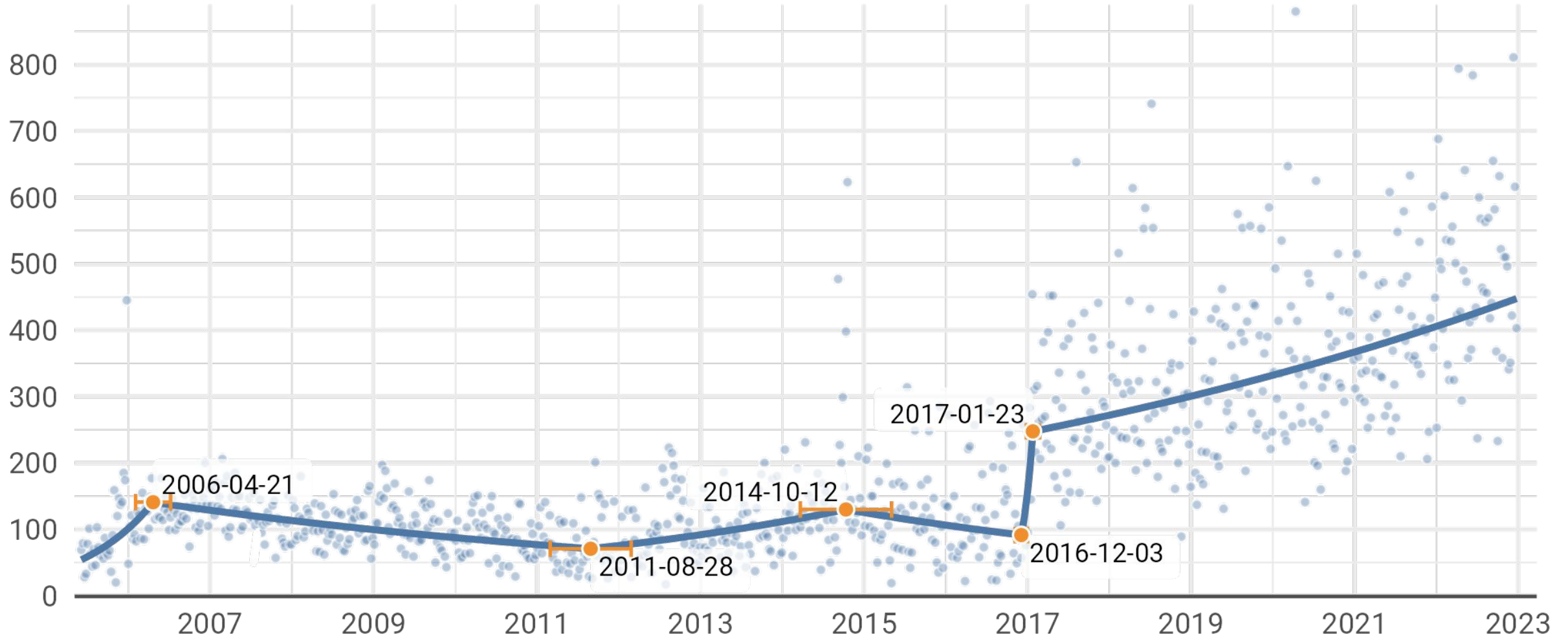


# Segmented Models: An example

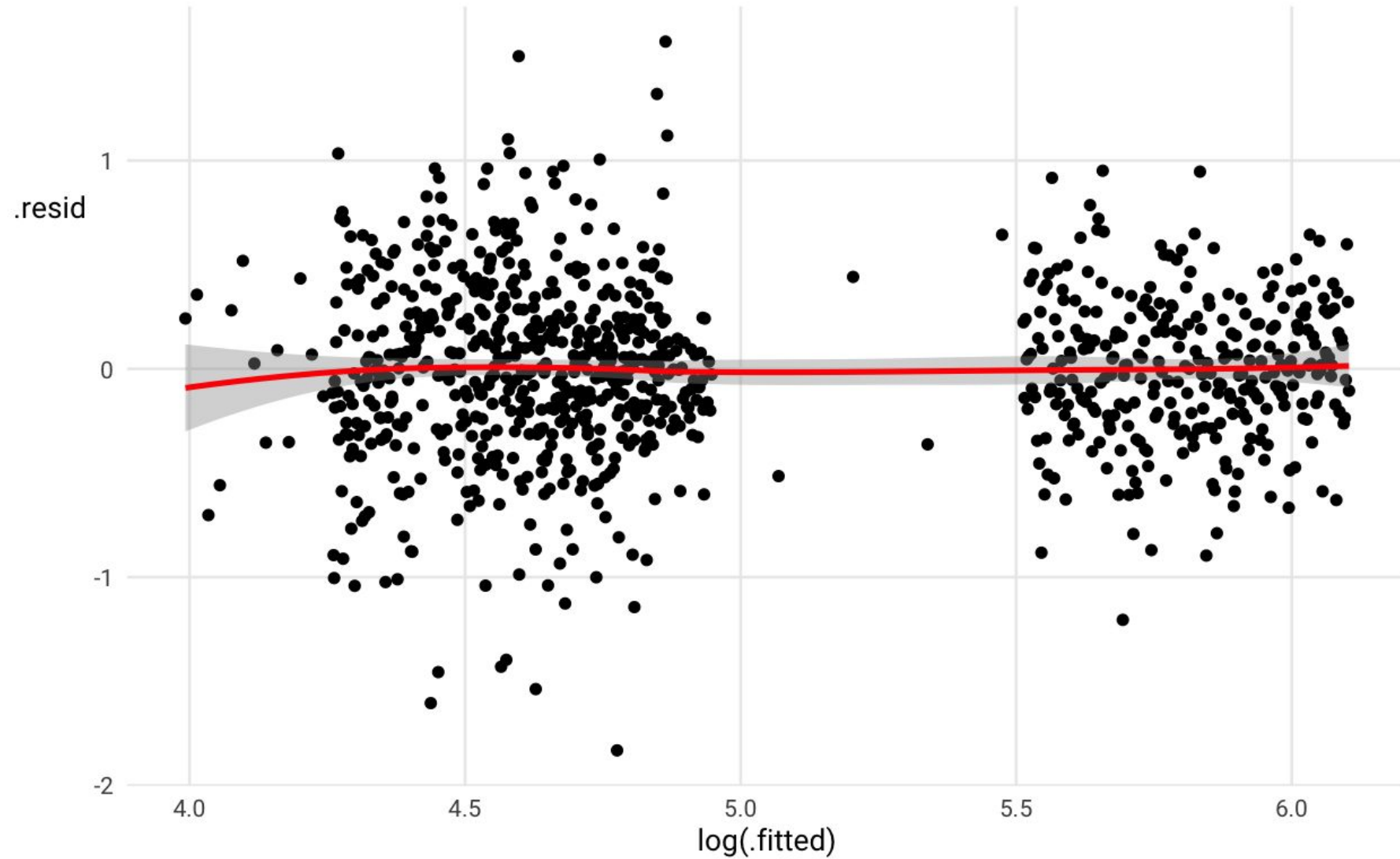


# Segmented models: An example

Number of CVEs published in NVD each week



# Segmented models: An example



# Segmented Models Conclusions

## Implications for Vulnerability data

Vulnerability volume is non-linear, prediction and inference should be as well.

The cause of non-linearity is *deliberate* changes in the CVE process.

Models need to account for development of CNAs or only use data post 2016.

## Segmented Model Limitations

Forced continuity (other models can address this)

Model selection ambiguity

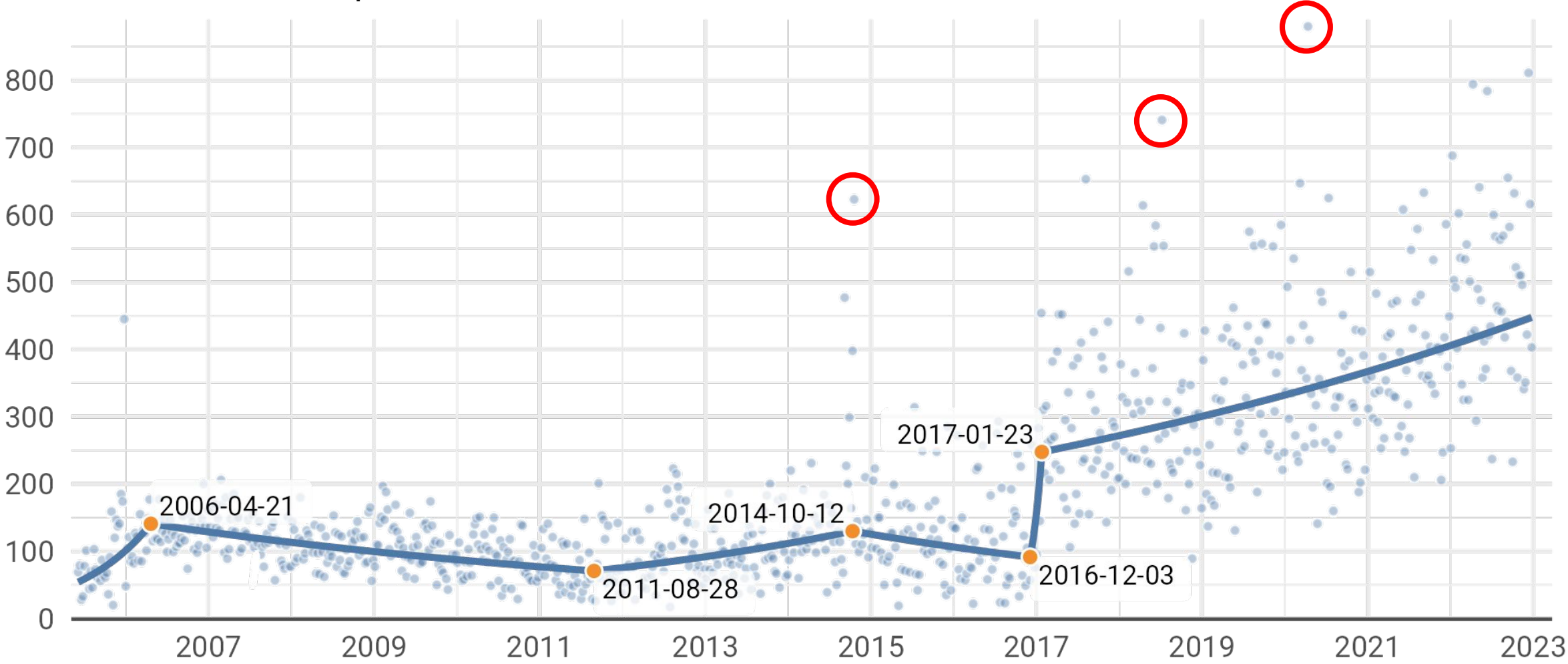
Limited model flexibility (theory not as well developed as other methods, e.g. GAMs)

Very limited predictive capability (for looking at the past)



# Noise Alert: Days of Many Vulnerabilities

Number of CVEs published in NVD each week



# Generalized Additive Models

Fit splines that balance between error reduction and smoothness

Spline knots are predetermined and uniformly distributed

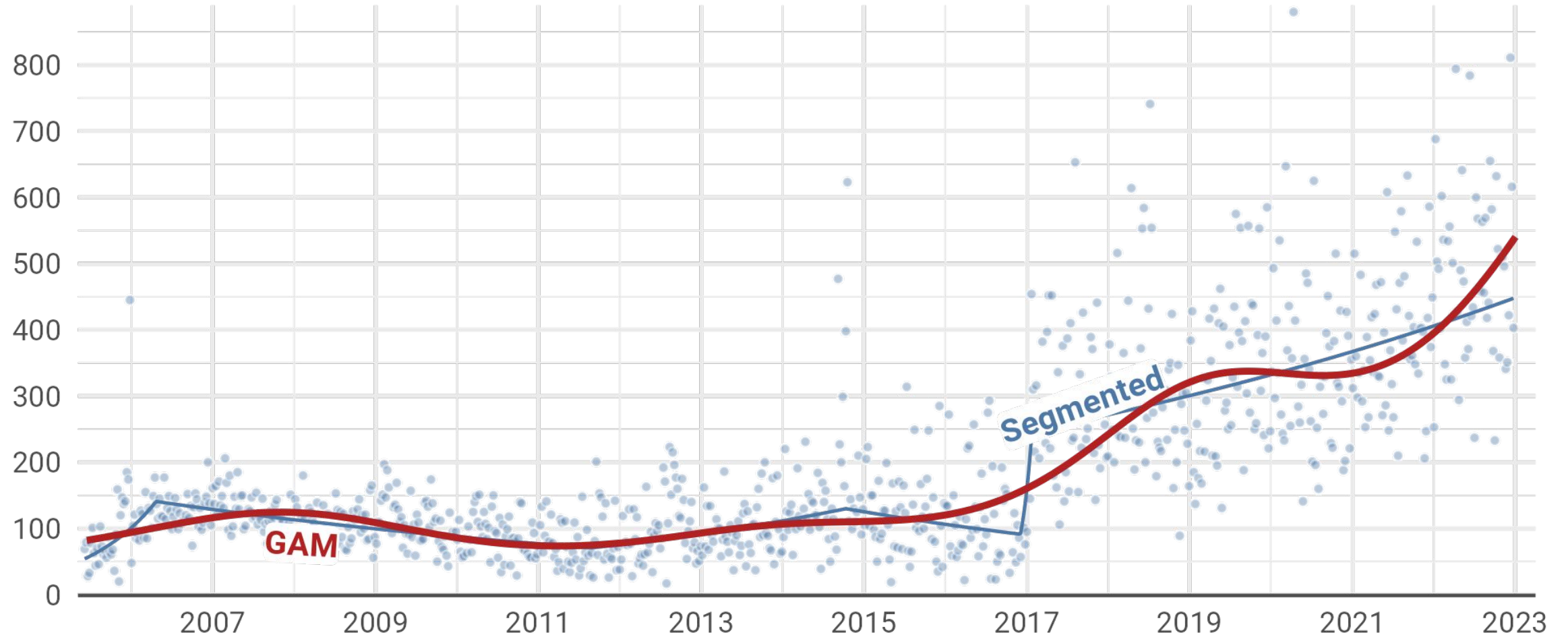
Multiple methods for splines and smooth regularization exist

Extensions to multiple dimensions “thin plates”

Generally visual interpretation of models

# GAMs on our volume data

Number of CVEs published in NVD each week



# Generalized Additive Models

## Pros

- Can fit non-linearities without assumptions on form
- Can handle interaction terms
- Robust and well developed software
- Relative simplicity makes them suitable for “medium” data
- No discrete break points (smoother likelihood function)

## Cons

- Future prediction quickly has wide CIs
- No distinct “break points” as in segmented models
- Resulting model can be difficult to interpret

# Generalized Additive Models Daily with covariates

$$\log(n(t)) \propto s(t) + \text{month}(t) * \text{day\_of\_week}(t) + \text{is\_holiday}(t) * \text{day\_of\_week}(t) + \text{holiday}(t)$$

*t*: days since start of data

*s*: Smoothed regression function

*month(t)*: month of the specific date

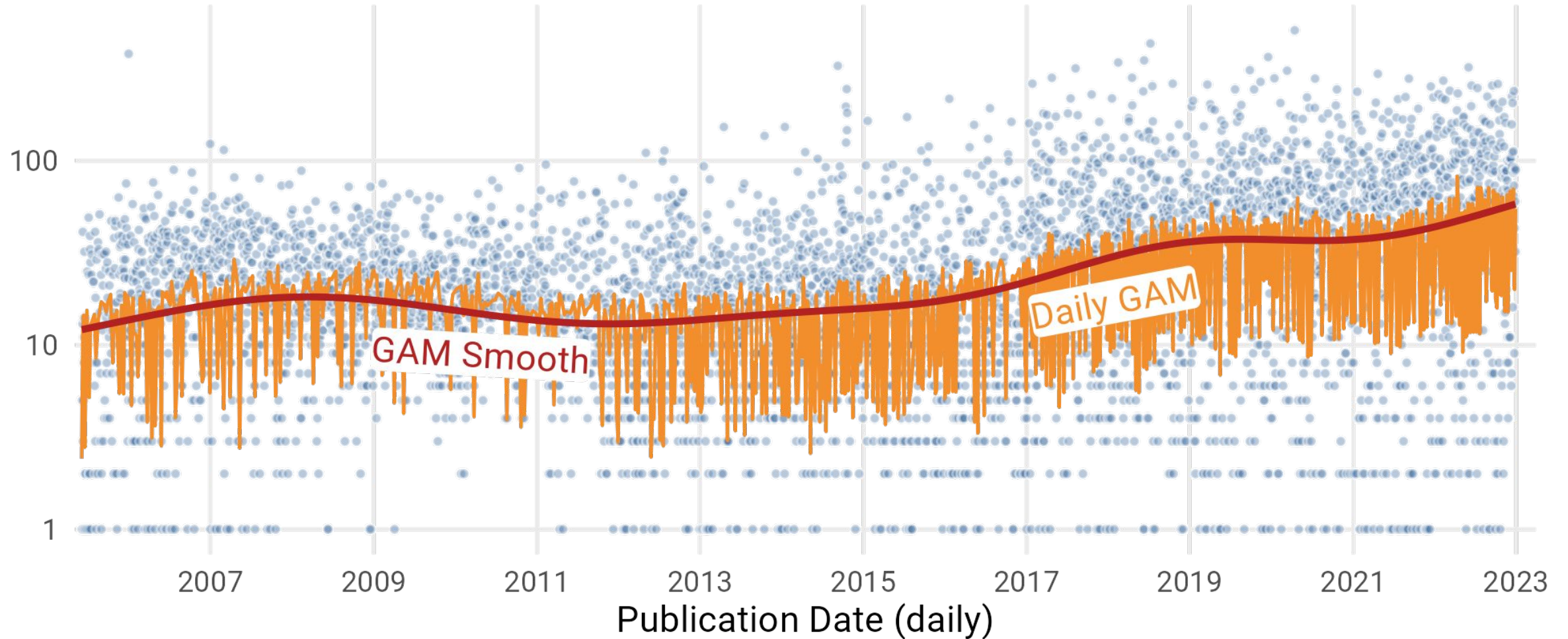
*day\_of\_week(t)*: day of the week for specific date

*is\_holiday(t)*: Binary indicating whether the day is a holiday

*holiday(t)*: Factor indicating which holiday the day is, with a base level of “no holiday”

# Generalized Additive Models Daily with covariates

Number of CVEs published each day



# Generalized Additive Models

## Pros

- Can fit non-linearities without assumptions on form
- Can handle interaction terms
- Robust and well developed software
- Relative simplicity makes them suitable for “medium” data
- No discrete break points (smoother likelihood function)

## Cons

- Future prediction quickly has wide CIs
- No distinct “break points” as in segmented models
- Resulting model can be difficult to interpret***

# Average Marginal Effects

*General approach to understanding non-linear statistical models*

Use model to predict CVE volume for all observations

Predict all observations with a variable of interest changed

Average the change in prediction

Compare across subsets of data

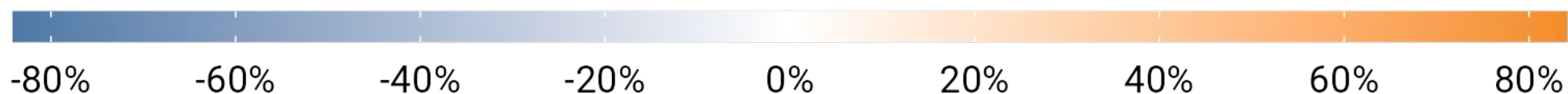
*Knowledge of models statistical form can be used to calculate confidence intervals*



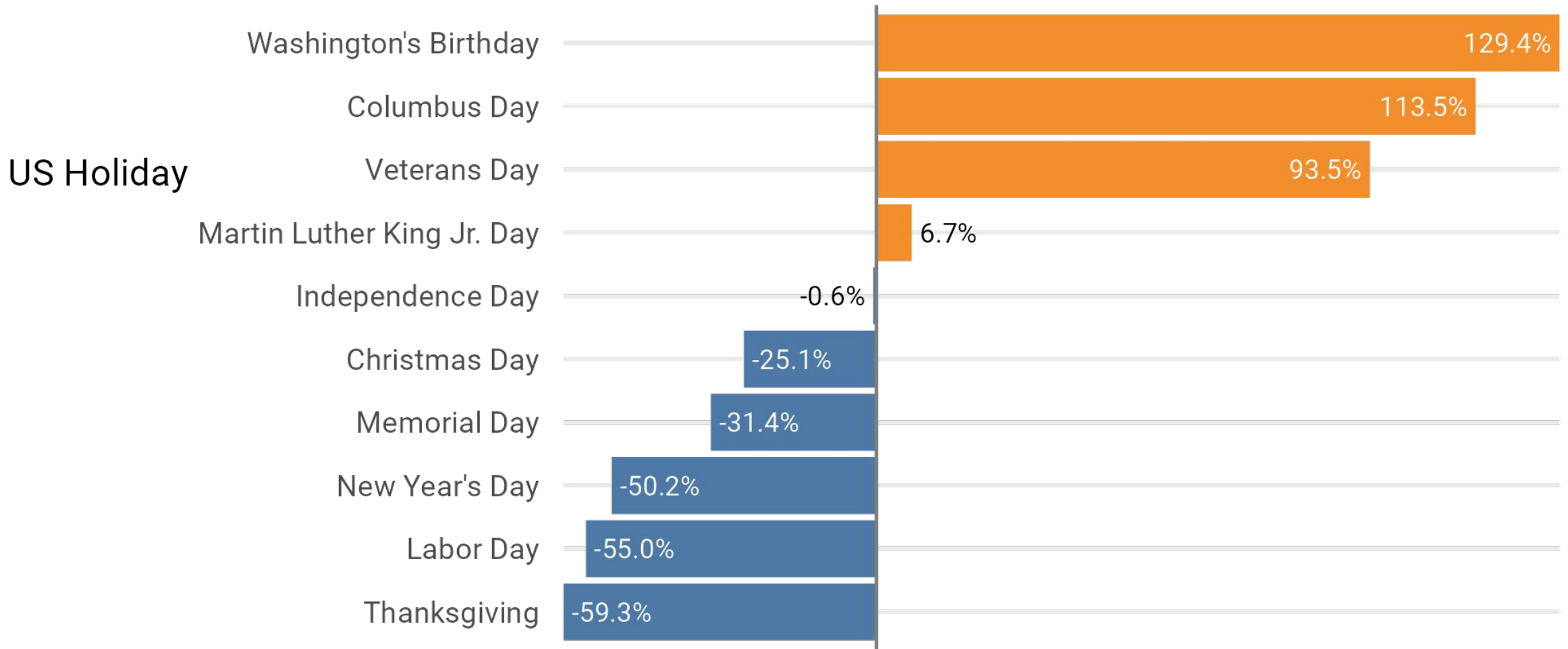
# Publication Timing

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Jan	-31.7%	2.0%	34.5%	38.4%	26.4%	-68.7%	-72.0%
Feb	-7.8%	3.4%	38.1%	7.2%	7.1%	-57.5%	-73.3%
Mar	-9.6%	24.1%	35.7%	29.3%	13.7%	-74.2%	-68.1%
Apr	5.4%	29.5%	76.0%	30.7%	12.0%	-78.8%	-67.1%
May	13.9%	-8.4%	17.7%	23.5%	19.1%	-75.5%	-83.9%
Jun	4.6%	25.0%	28.6%	34.1%	8.8%	-71.0%	-79.6%
Jul	12.5%	43.0%	52.4%	28.8%	-7.0%	-79.3%	-68.1%
Aug	15.9%	21.2%	33.2%	8.4%	4.2%	-72.8%	-70.2%
Sep	43.8%	22.4%	43.0%	36.9%	23.2%	-67.4%	-65.9%
Oct	1.7%	32.9%	59.7%	18.9%	14.8%	-75.2%	-64.5%
Nov	-7.7%	20.0%	21.4%	-3.2%	-10.7%	-70.7%	-74.4%
Dec	0.0%	-3.5%	51.0%	25.6%	-4.8%	-65.1%	-65.8%

Change in daily published CVEs



# Publication Timing



Effect on Daily CVE Publication Count

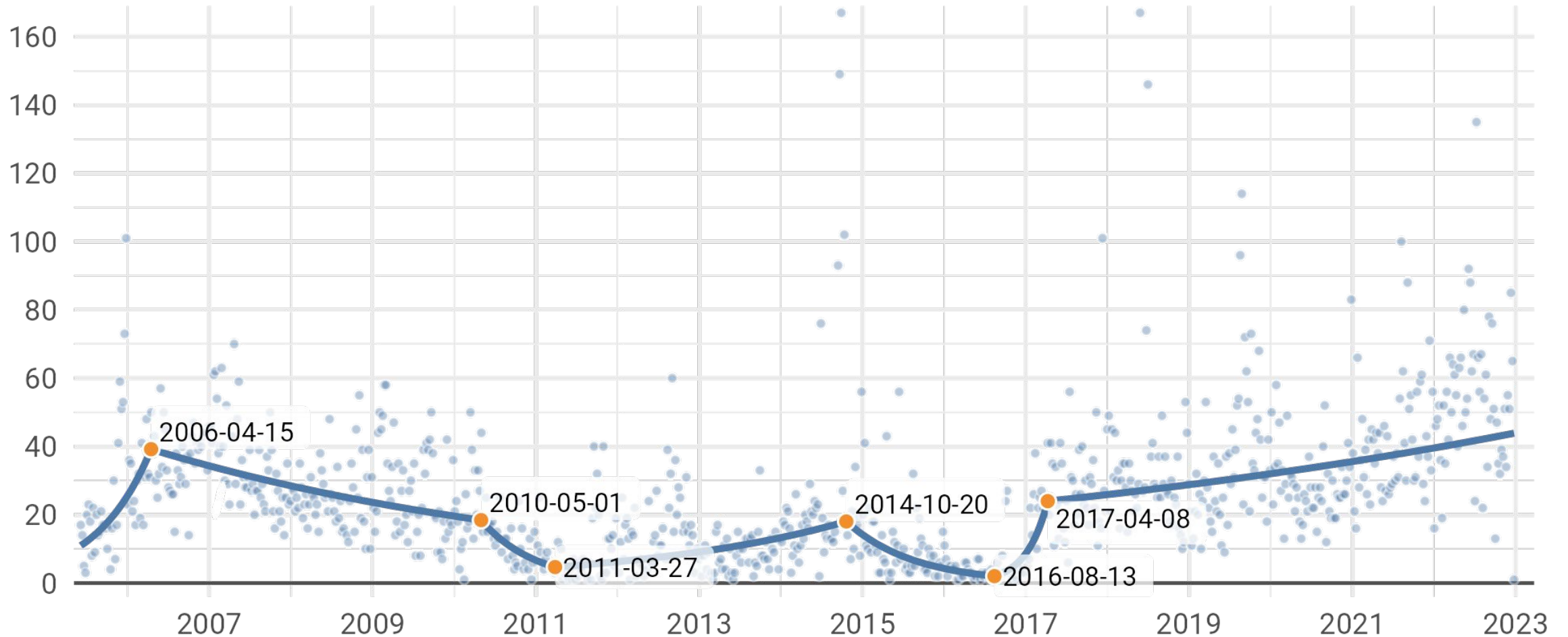
# Publication Timing

Monthly Percent of CVEs in a year (by publication date)

	'05	'06	'07	'08	'09	'10	'11	'12	'13	'14	'15	'16	'17	'18	'19	'20	'21	'22
Jan	7%	6%	9%	9%	8%	7%	10%	6%	8%	7%	11%	10%	7%	8%	7%	9%	8%	8%
Feb	2%	7%	8%	9%	12%	7%	9%	7%	8%	5%	7%	6%	7%	8%	5%	8%	7%	8%
Mar	3%	8%	11%	9%	10%	11%	9%	8%	8%	7%	7%	5%	9%	8%	7%	10%	7%	8%
Apr	2%	9%	10%	8%	10%	11%	8%	4%	8%	8%	8%	10%	11%	10%	9%	12%	9%	8%
May	25%	9%	9%	7%	6%	9%	7%	8%	7%	7%	6%	9%	7%	7%	8%	6%	7%	8%
Jun	5%	9%	9%	8%	8%	11%	7%	8%	7%	6%	8%	8%	7%	11%	7%	10%	9%	9%
Jul	6%	8%	9%	9%	8%	7%	7%	10%	9%	8%	10%	11%	9%	13%	9%	8%	8%	8%
Aug	7%	8%	8%	7%	9%	8%	7%	13%	7%	5%	9%	6%	10%	6%	12%	6%	10%	9%
Sep	6%	8%	7%	8%	10%	6%	9%	12%	9%	14%	8%	9%	8%	7%	9%	9%	9%	9%
Oct	6%	8%	8%	9%	6%	9%	12%	10%	11%	18%	11%	11%	10%	9%	9%	9%	8%	7%
Nov	10%	8%	7%	8%	5%	6%	7%	8%	8%	6%	6%	6%	7%	6%	10%	7%	7%	8%
Dec	20%	12%	7%	9%	8%	8%	8%	5%	9%	8%	9%	8%	7%	7%	9%	8%	10%	10%

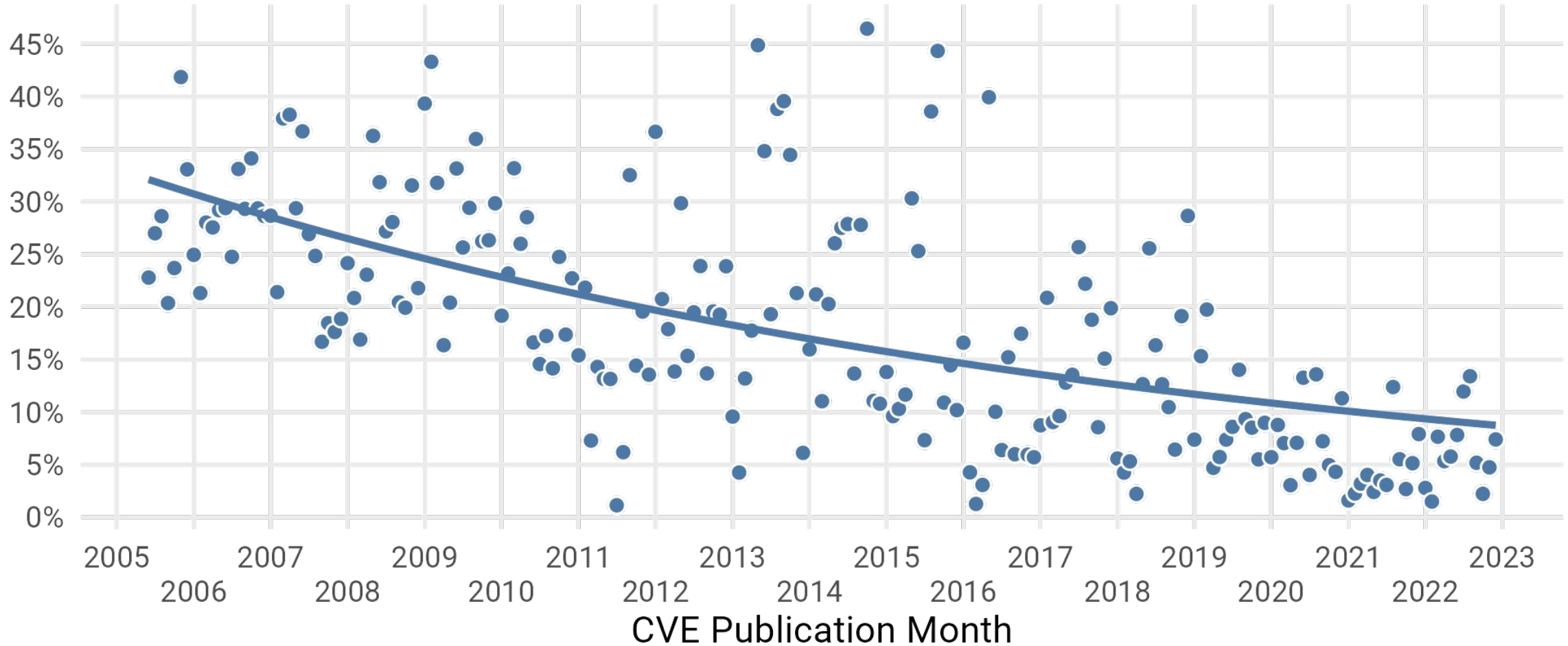
# Vendors & Their Vulns

Total number of vendors experiencing their first CVE



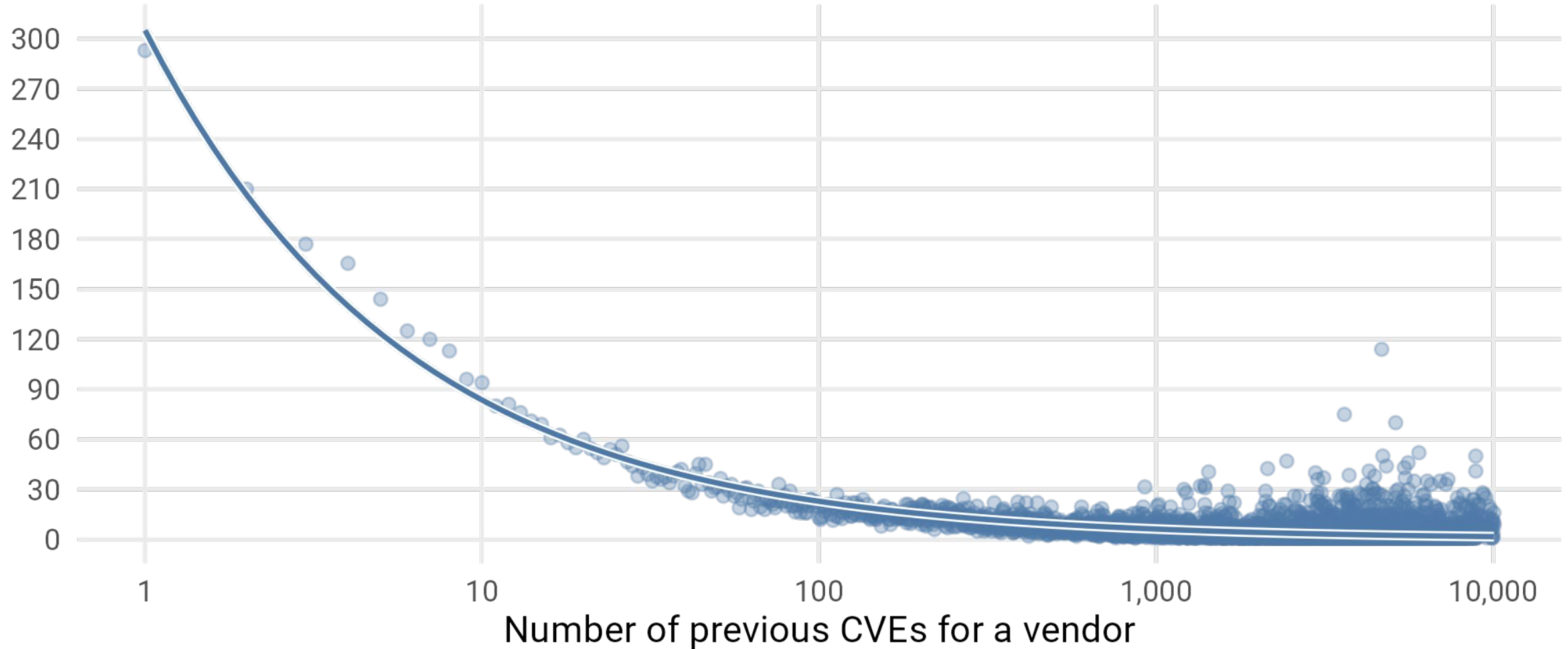
# Vendors & Their Vulns

Percentage of CVEs that are a vendor's first



# Vendors & Their Vulns

Median days between CVEs

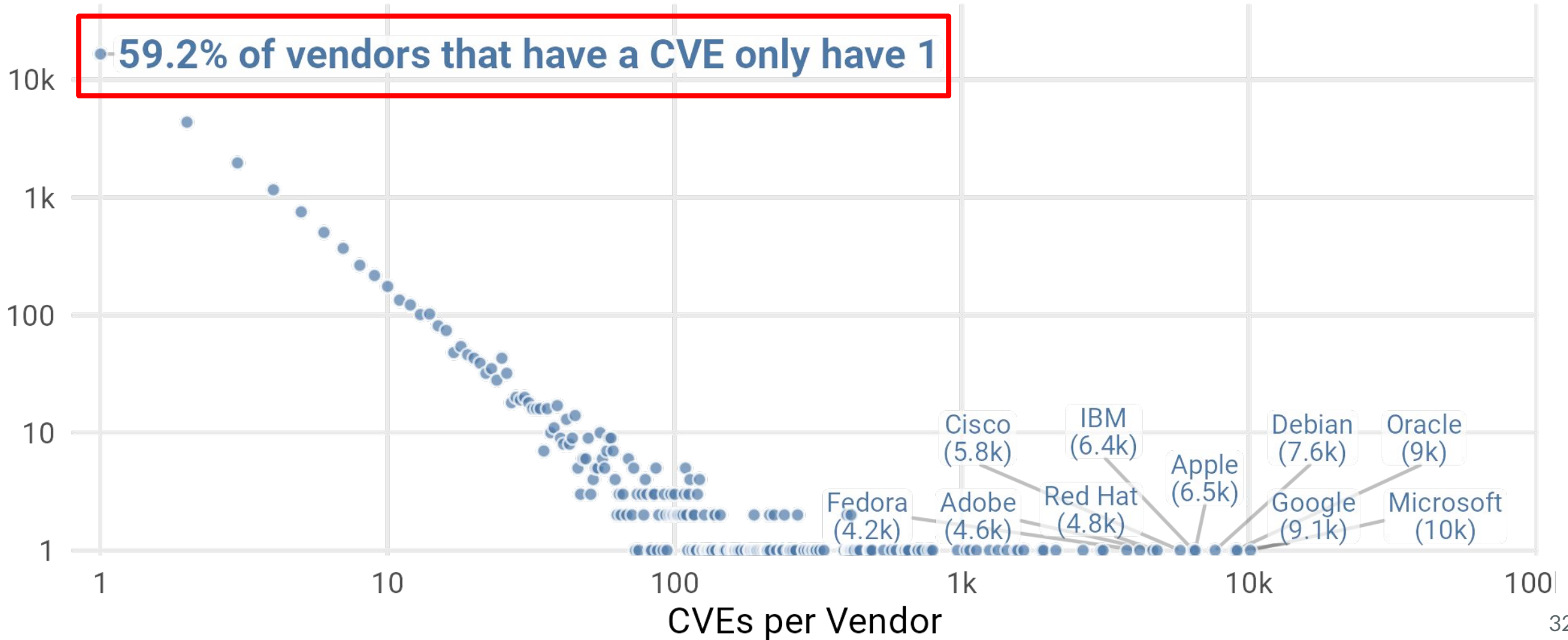


# CVE One Hit Wonders

- Among ~28,000 vendors, 382 (1.4%) published only one vuln at least 10 years ago.
- Notable vendors include:
  - ReactOS (CVE-2006-7136)
  - Casio (CVE-2006-3893)
  - Deutsche Telekom (CVE-2008-1252)
  - IronMountain (CVE-2011-2397)

# Vendors & Their Vulns

Number of vendors with a particular number of CVEs



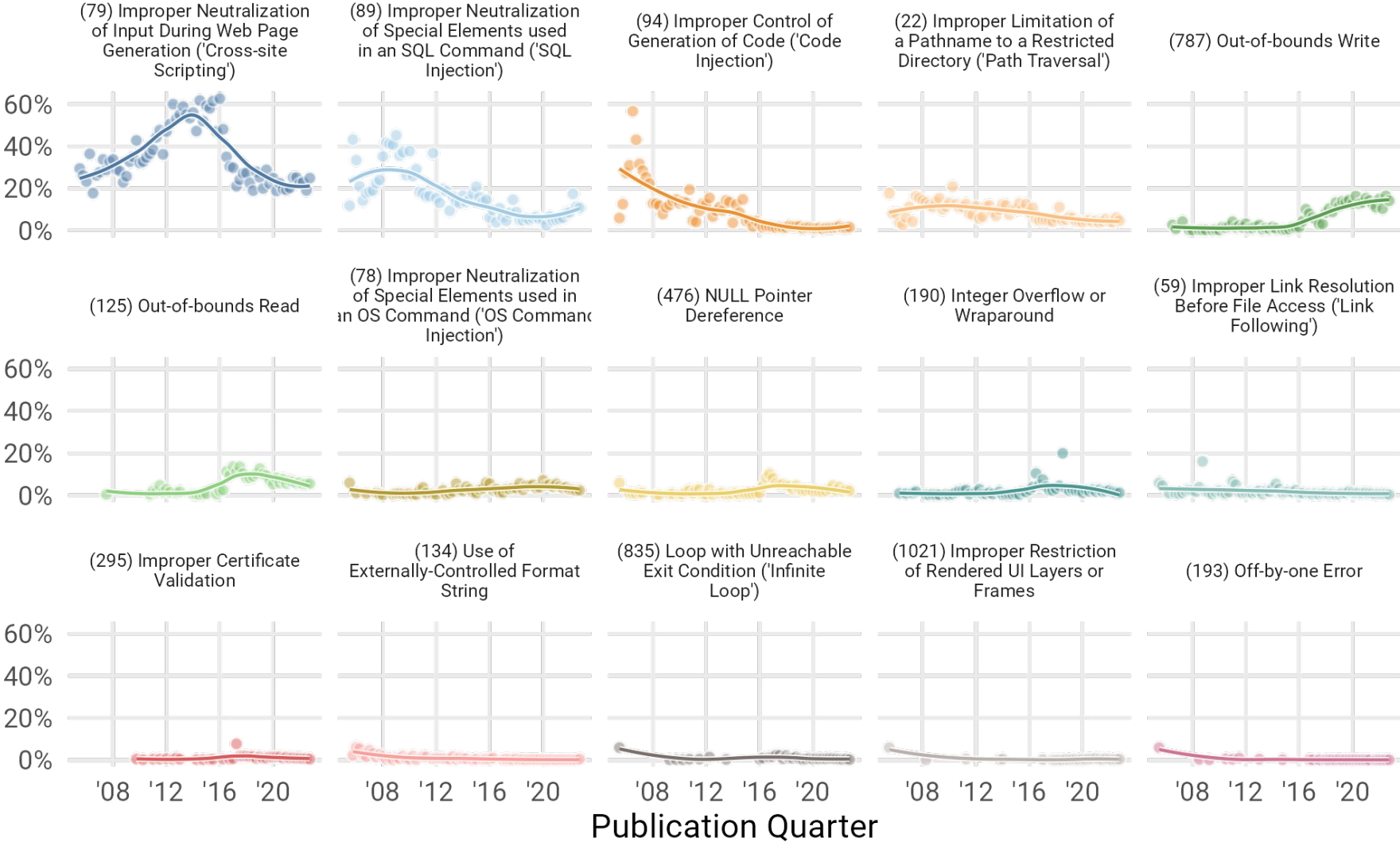


# Noise Alert: Wide-Ranging CVEs (by CPE)

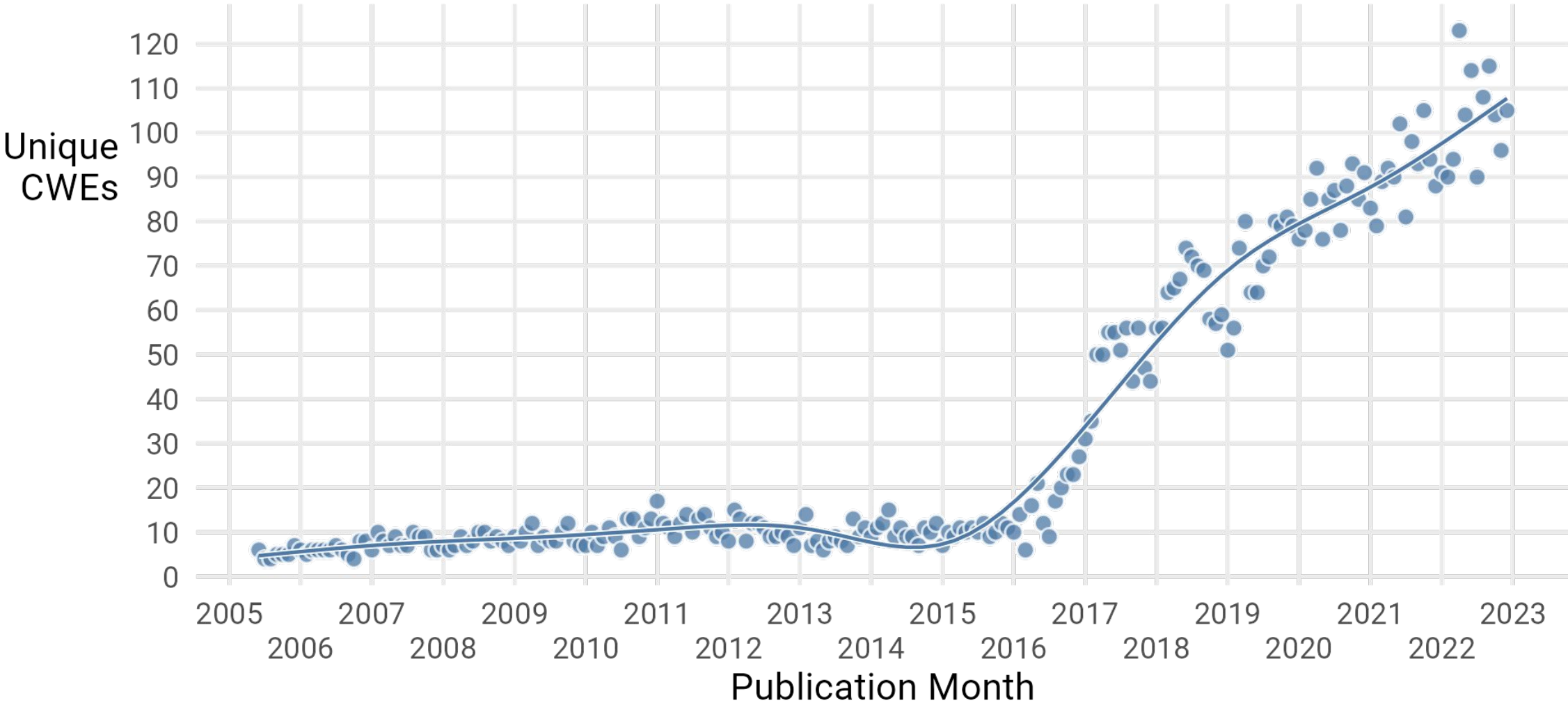
- Common Platform Enumeration maps CVEs to software vendors, products, and versions.
- Most CVEs are narrowly focused
  - 90% affect one vendor
  - 74% affect one product
  - 49% affect one version of one product
- The tail, however, is long...
- CVE-2017-15361: **35 different Chromebook manufacturers** using faulty TPM chip.
- CVE-2015-12207: page table invalidation flaw for VMs running on Intel chips. **1,532 distinct products!**
- CVE-2016-1409: vuln in Cisco implementation of NDP for IPv6. **4,891 versions affected** (mostly due to granular versioning)

# Describing CVEs with CWEs

Percent of CVEs labeled with CWE



# CWE Use Over Time



# CWE Collision


- CVEs are supposed to have a single CWE.
- 5% of CVEs do not.
- May (or may not) be why we can't have nice things.

## References to Advisories, Solutions, and Tools

By selecting these links, you will be leaving NIST webspace. We have provided these links to other web sites because they may have information that would be of interest to you. No inferences should be drawn on account of other sites being referenced, or not, from this page. There may be other web sites that are more appropriate for your purpose. NIST does not necessarily endorse the views expressed, or concur with the facts presented on these sites. Further, NIST does not endorse any commercial products that may be mentioned on these sites. Please address comments about this page to [nvd@nist.gov](mailto:nvd@nist.gov).

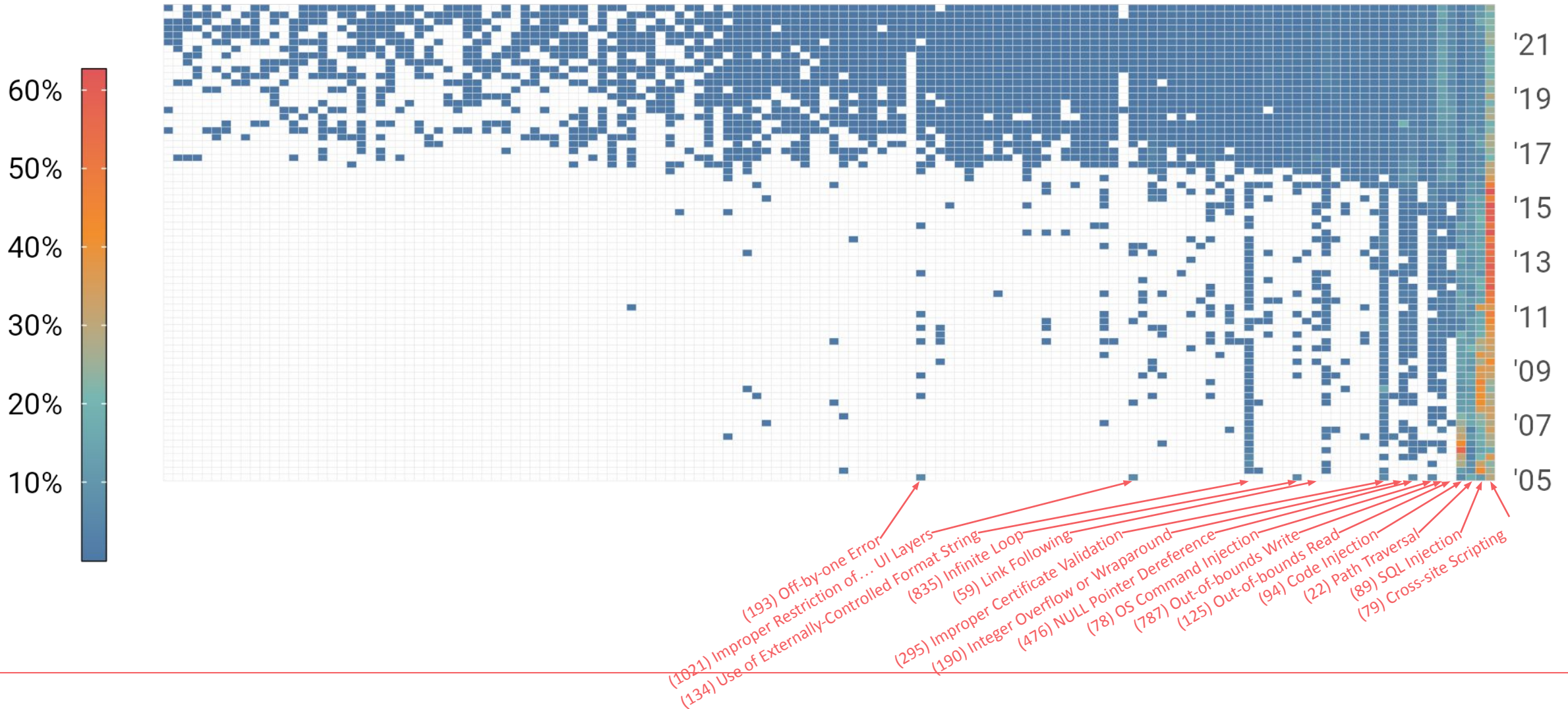
Hyperlink	Resource
<a href="https://support.citrix.com/article/CTX463706/citrix-gateway-and-citrix-adc-security-bulletin-for-cve202227510-cve202227513-and-cve202227516">https://support.citrix.com/article/CTX463706/citrix-gateway-and-citrix-adc-security-bulletin-for-cve202227510-cve202227513-and-cve202227516</a>	Vendor Advisory

## Weakness Enumeration

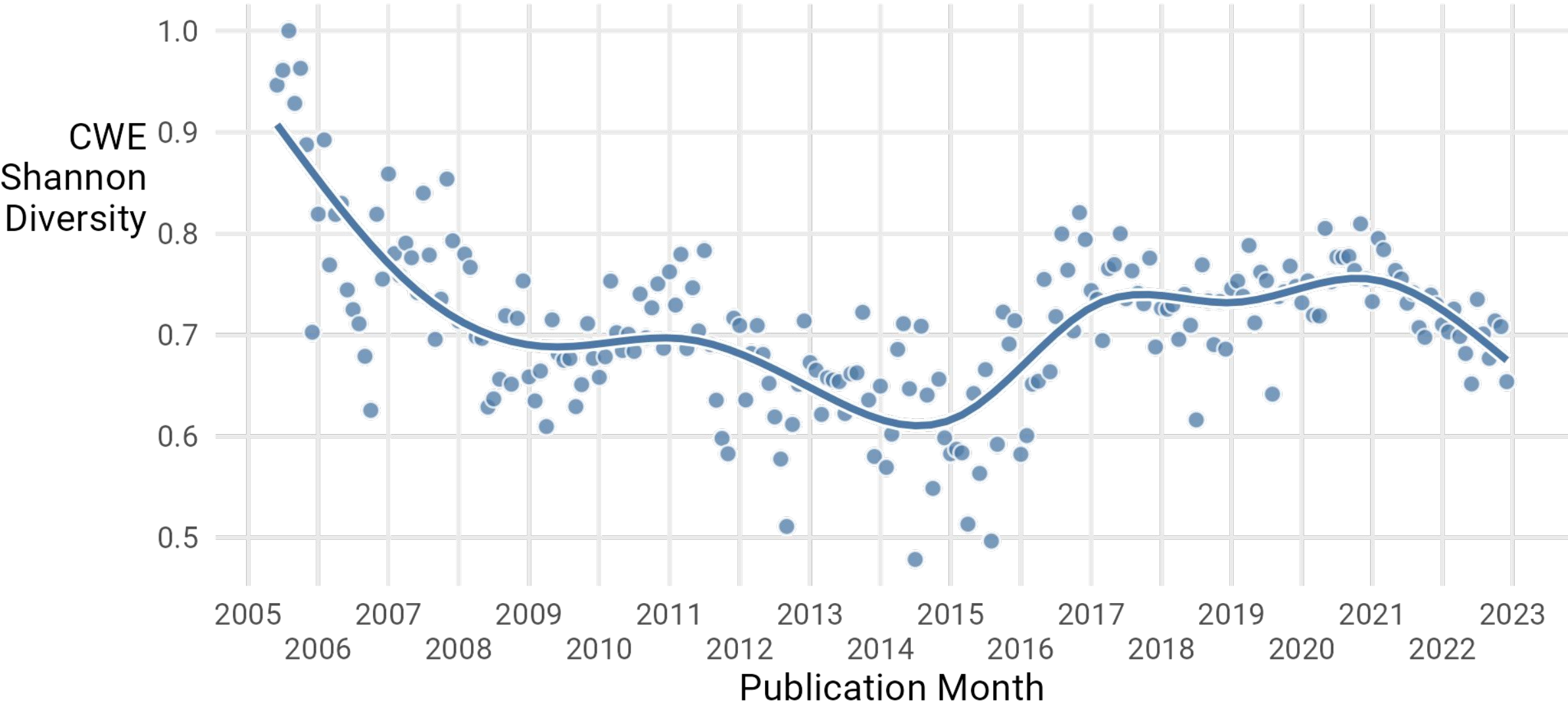
CWE-ID	CWE Name	Source
CWE-287	Improper Authentication	 NIST
CWE-288	Authentication Bypass Using an Alternate Path or Channel	 Citrix Systems, Inc.

# CWE: CVE Diversity

Percentage of CVEs with a particular CWE by quarter

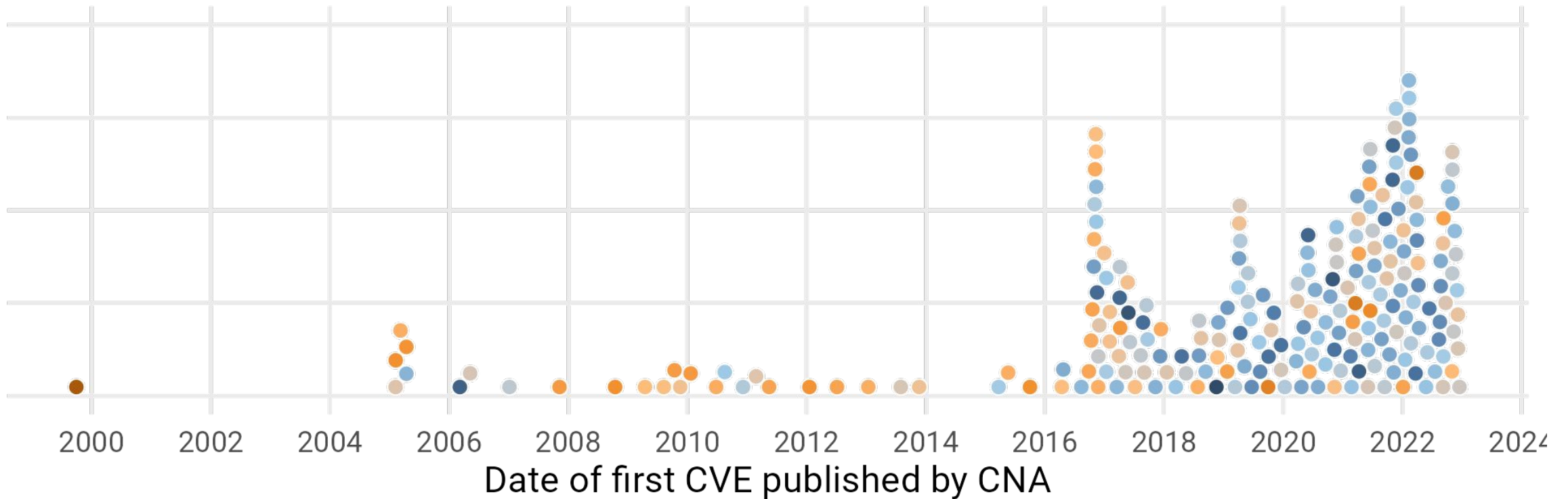


# CWE: CVE Diversity

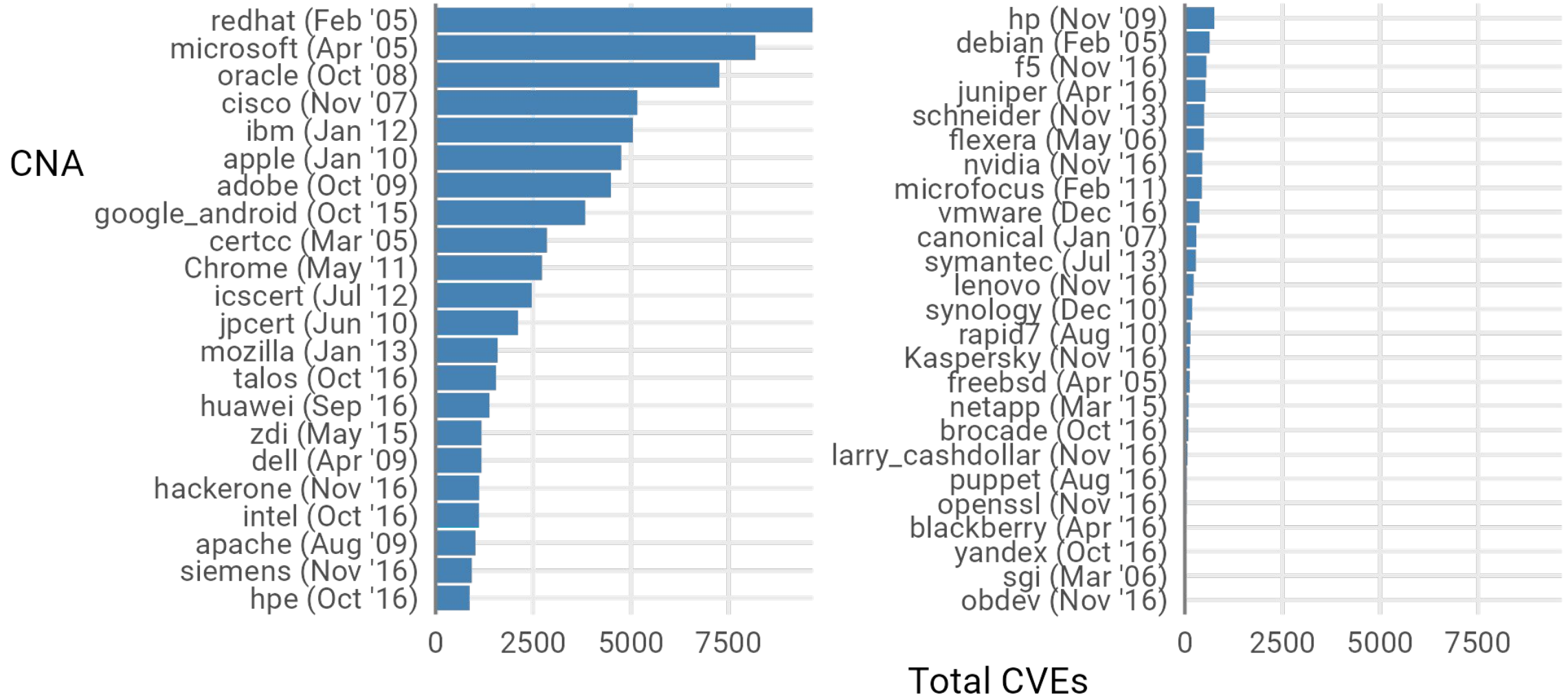


# CNAs and their rates. When the data is not the data.

Weekly CVE publication rate over life of CNA



# 48 CNAs publish their first CVE before the CNA process.



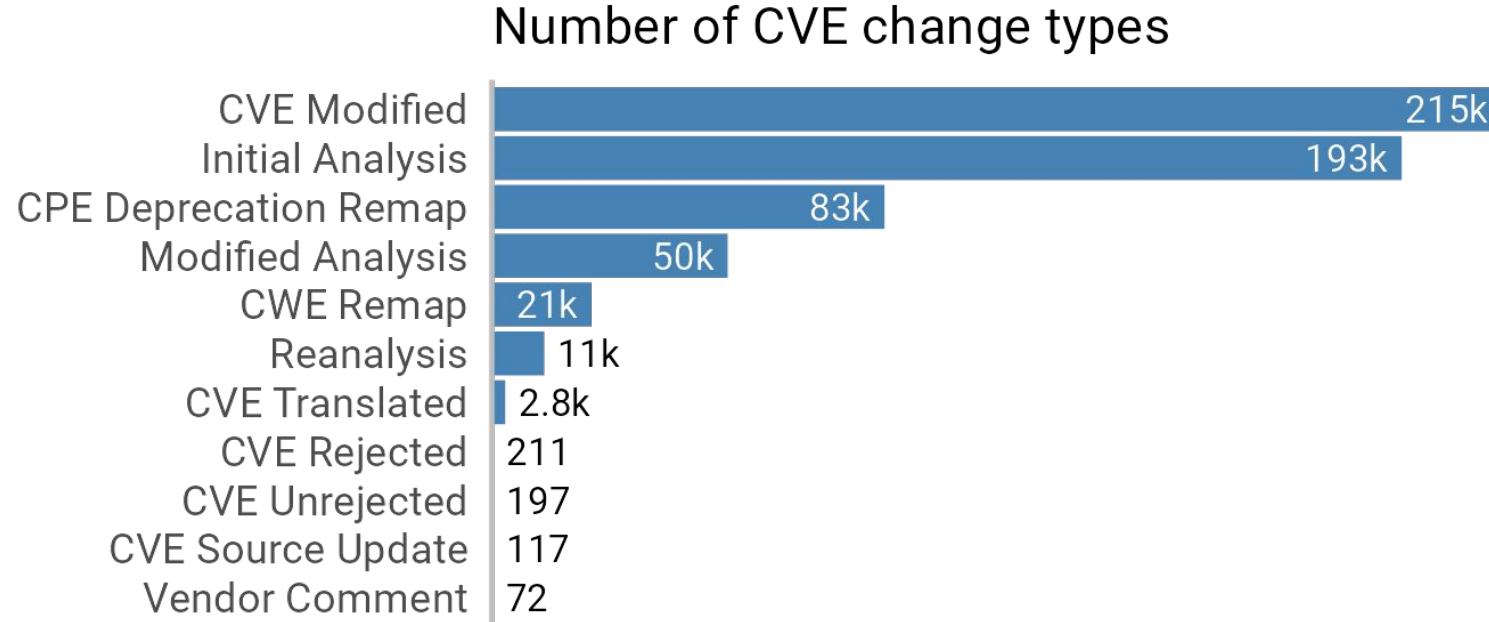


# What happened? CVE Changes

381,020 total changes made (more than 1 per CVE on average)

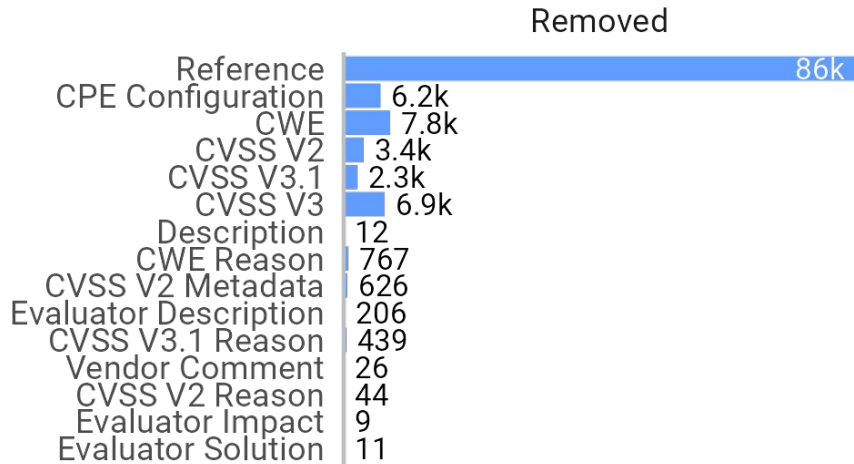
123,660 have been modified after initial analysis

751,794 Individual CVE fields have been changed

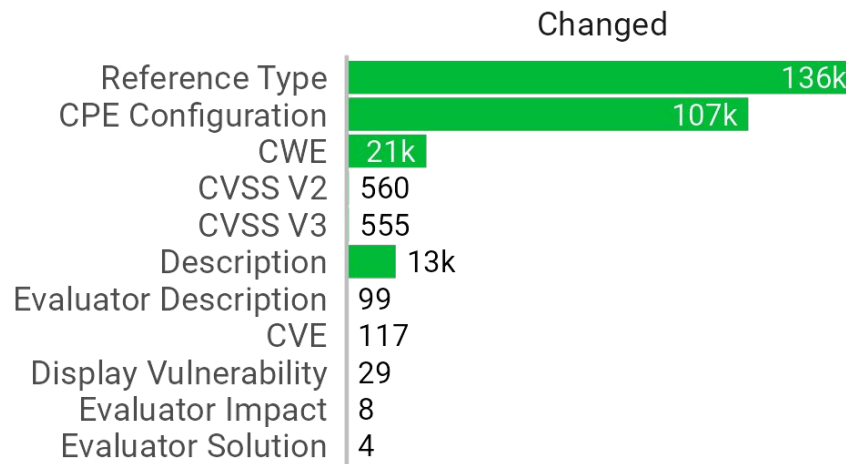
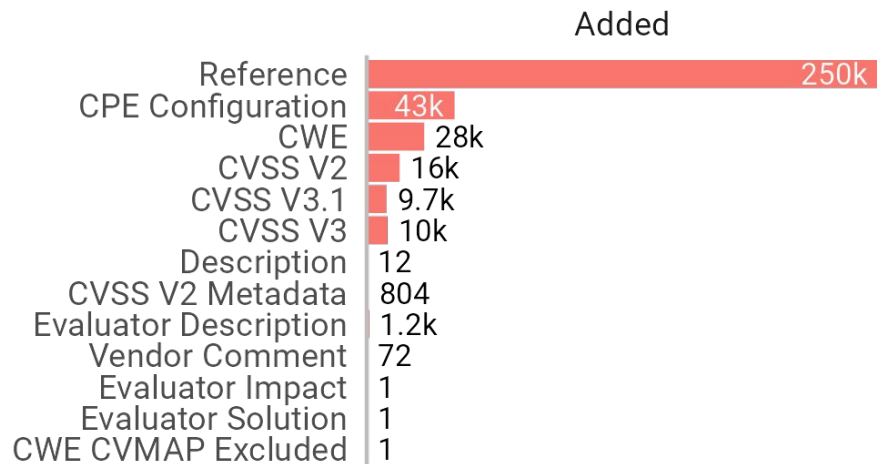


# Changes are where we might expect

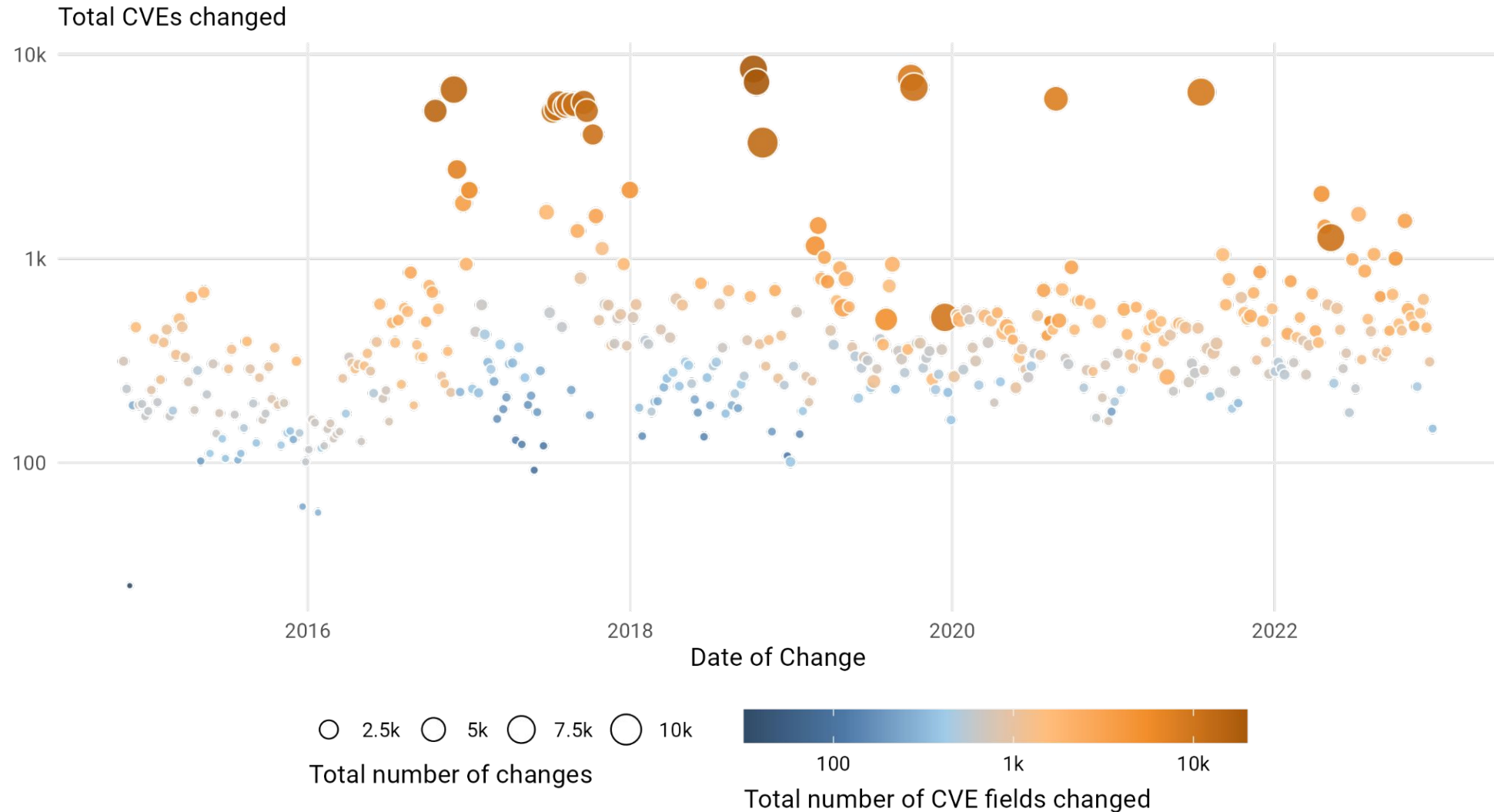
Number of CVE change types



Reference, CPE, CWE, and Description are the most common



# Changes are frequent and sporadic



# So what happened with those CNAs?

NVD does not maintain CNA information.

MITRE does not have easily accessible historical data.

Purdue “maintains” an index of MITRE changes

In March 2019 a huge number of CVEs were changed

Including 99.97% of the 123,433 pre CNA cves (all but 43)

Entries labeled “Changes in comments/references”

Plausible these are miscraped changes in “Assigning CNA”

Data history may be fluid and/or unknown

Name	Last modified	Size	Description
Parent Directory		-	
<a href="#">CVE.2001.11.html</a>	2016-03-03 14:37	23K	
<a href="#">CVE.2001.12.html</a>	2016-03-03 14:37	26K	
<a href="#">CVE.2002.01.html</a>	2016-03-03 14:37	20K	
<a href="#">CVE.2002.02.html</a>	2016-03-03 14:37	150K	
<a href="#">CVE.2002.03.html</a>	2016-03-03 14:37	191K	
<a href="#">CVE.2002.04.html</a>	2016-03-03 14:37	67K	
<a href="#">CVE.2002.05.html</a>	2016-03-03 14:37	156K	
<a href="#">CVE.2002.06.html</a>	2016-03-03 14:37	149K	
<a href="#">CVE.2002.07.html</a>	2016-03-03 14:37	122K	

## References

**Note:** [References](#) are provided for the convenience of the user.

- MS:MS05-018
- [URL:https://docs.microsoft.com/en-u](https://docs.microsoft.com/en-us/...)
- OVAL:oval:org.mitre.oval:def:1271
- [URL:https://oval.cisecurity.org/repos](https://oval.cisecurity.org/repos)
- OVAL:oval:org.mitre.oval:def:2043
- [URL:https://oval.cisecurity.org/repos](https://oval.cisecurity.org/repos)
- OVAL:oval:org.mitre.oval:def:4397
- [URL:https://oval.cisecurity.org/repos](https://oval.cisecurity.org/repos)
- OVAL:oval:org.mitre.oval:def:4832
- [URL:https://oval.cisecurity.org/repos](https://oval.cisecurity.org/repos)

## Assigning CNA

Microsoft Corporation

## Date Record Created

**20050226**

Disclaimer: The vulnerability wa

## Phase (Legacy)

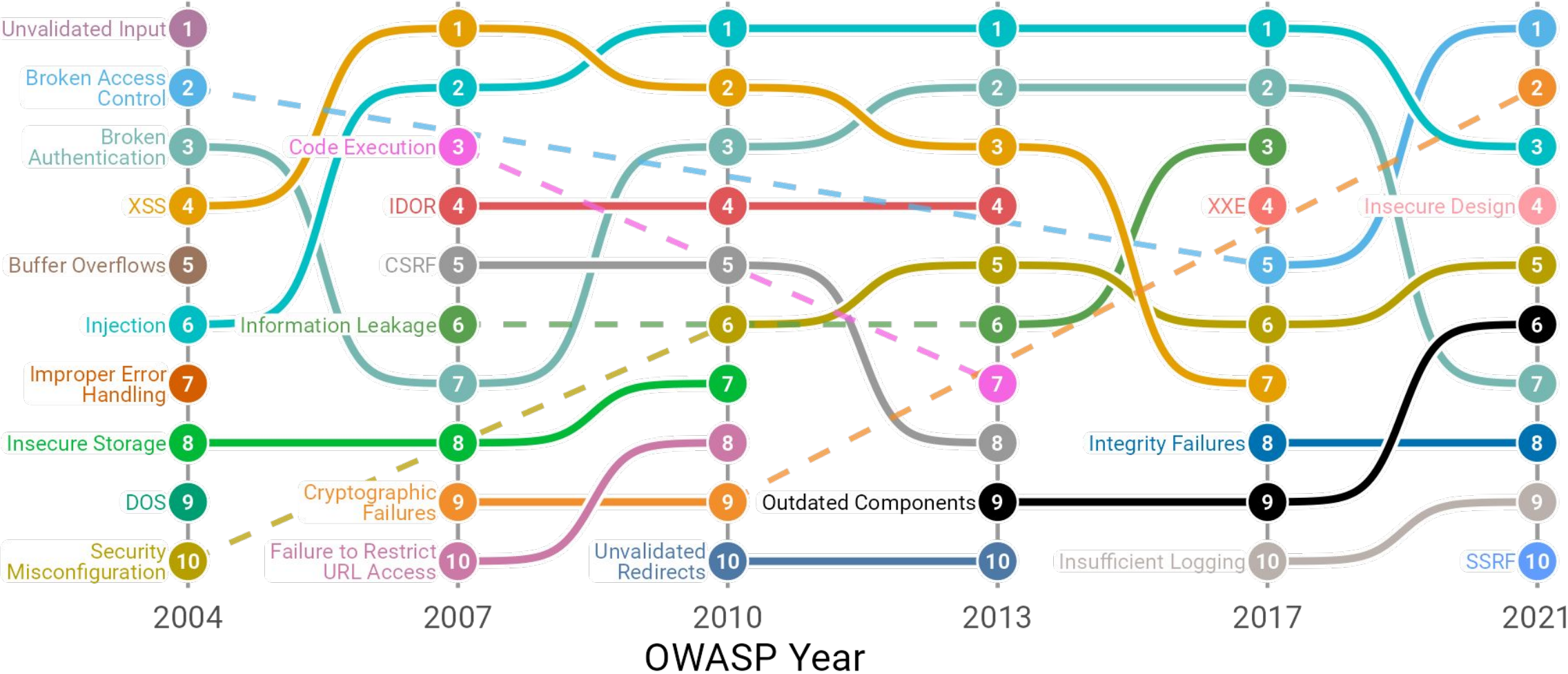
Assigned (20050226)

## Votes (Legacy)

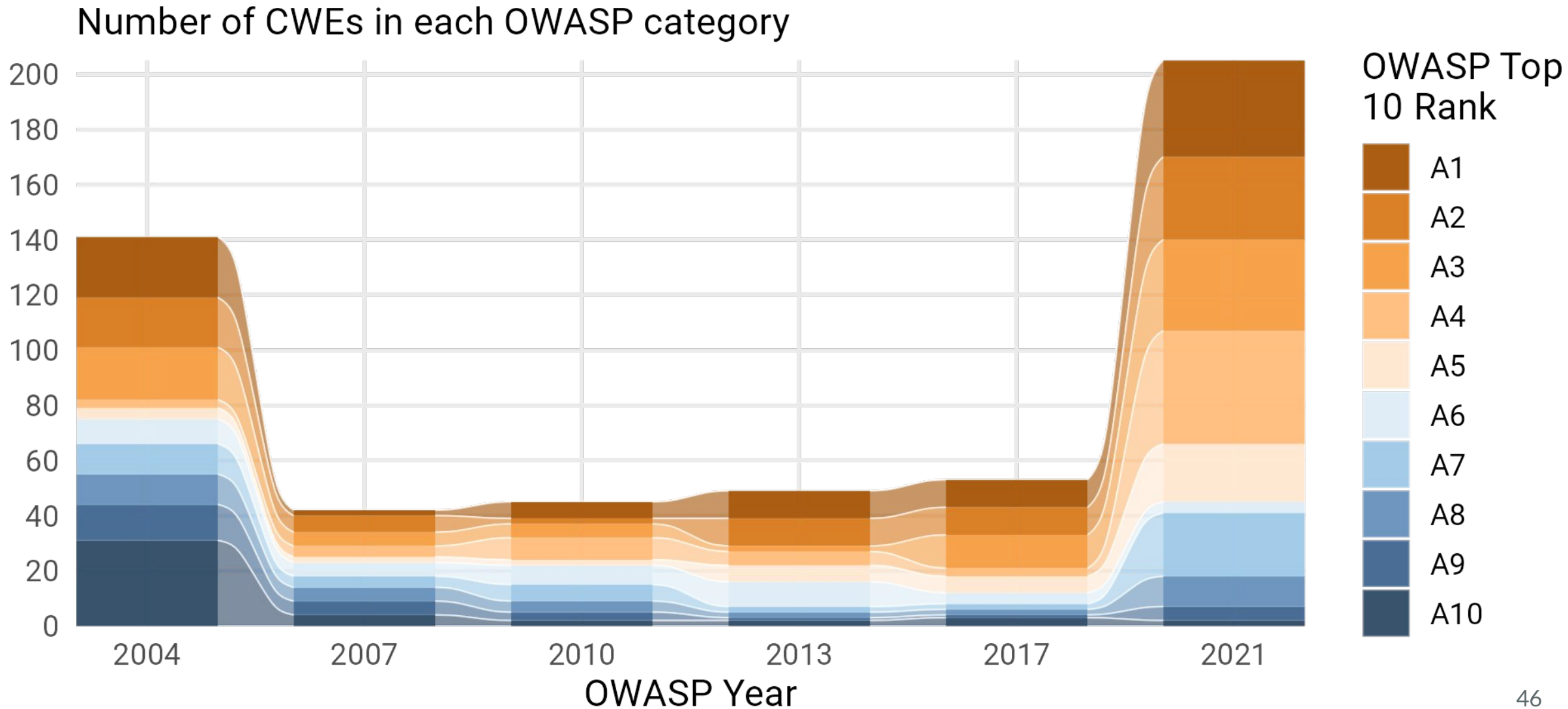
## Comments (Legacy)

# Noise Alert: OWASP All Over the Place (I)

OWASP Rank for Weaknesses by Year

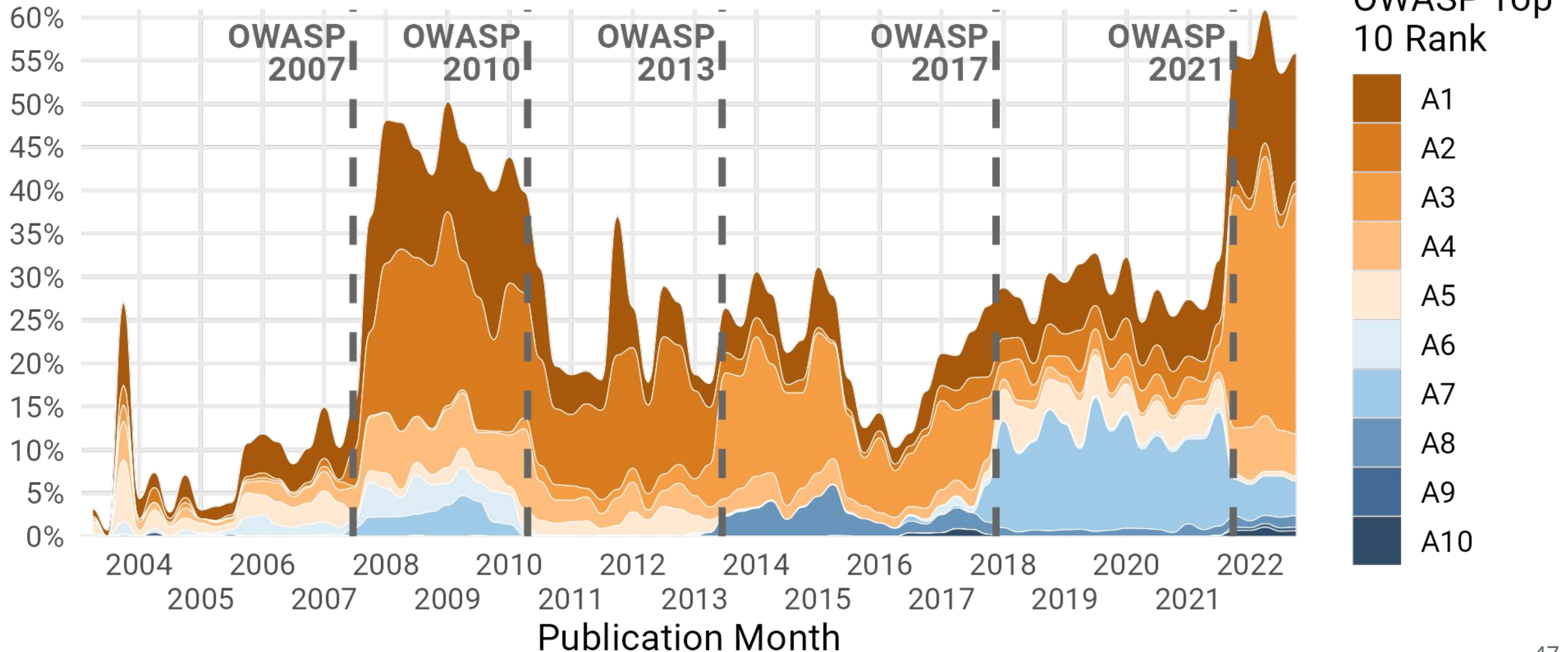


# OWASP:CWE Mapping Over Time



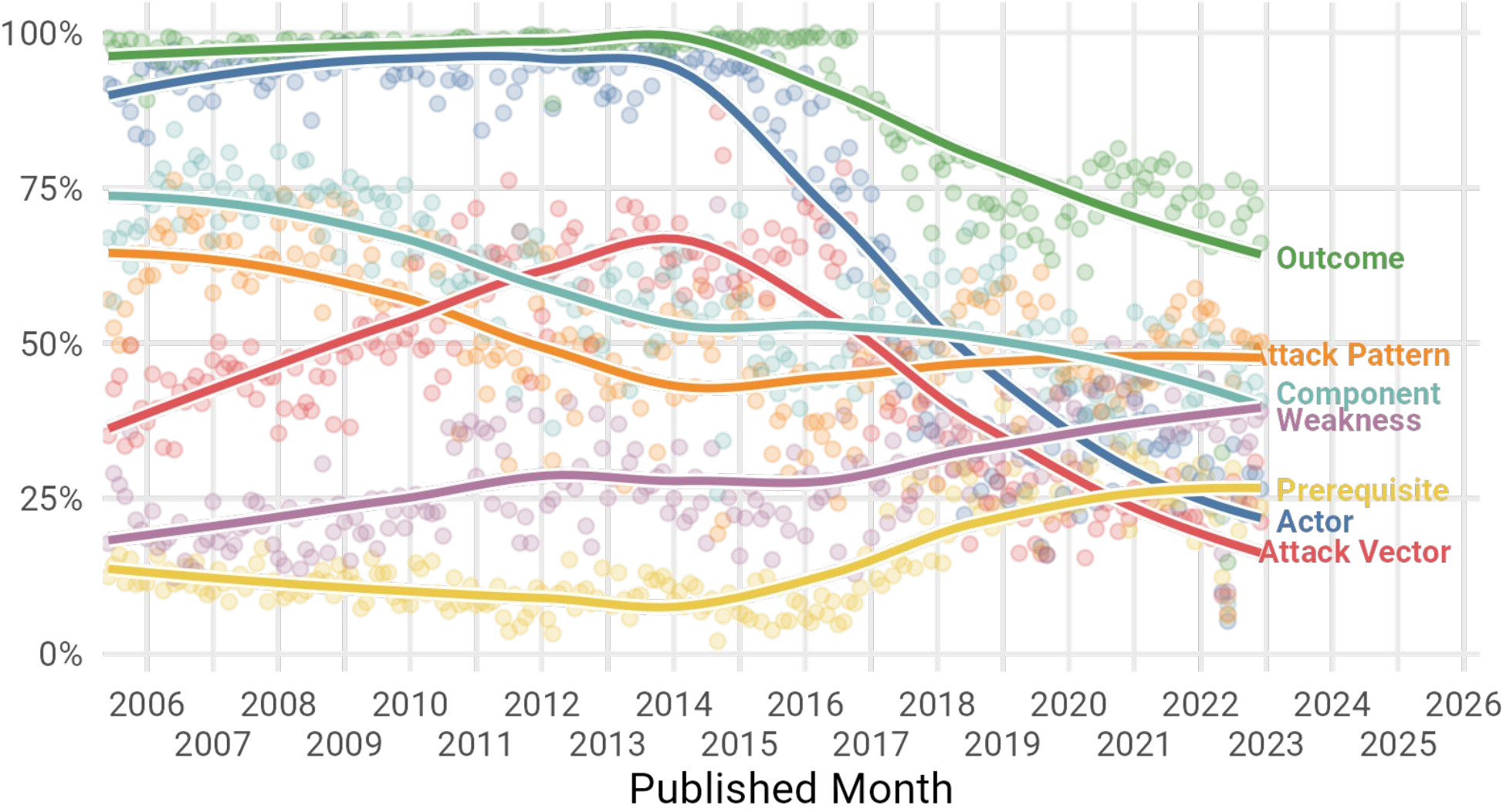
# CVEs Containing OWASP CWEs

Percent of CVEs with a CWE in the most recent OWASP top 10



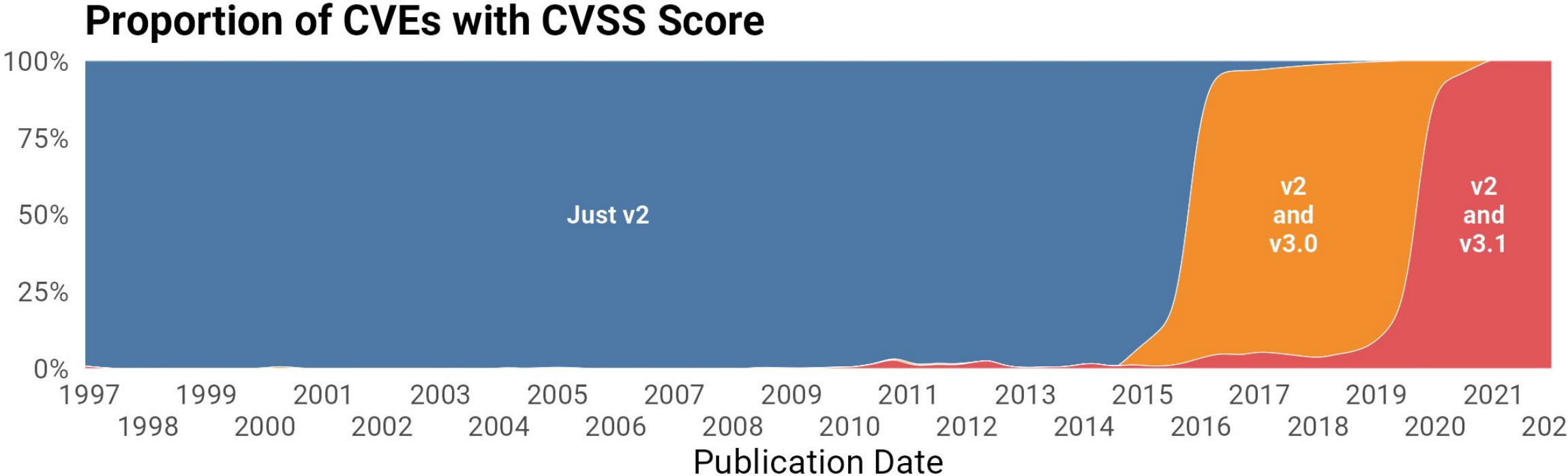
# Named Entity Recognition Analysis of CVEs

Percent of CVEs With Entity in Description



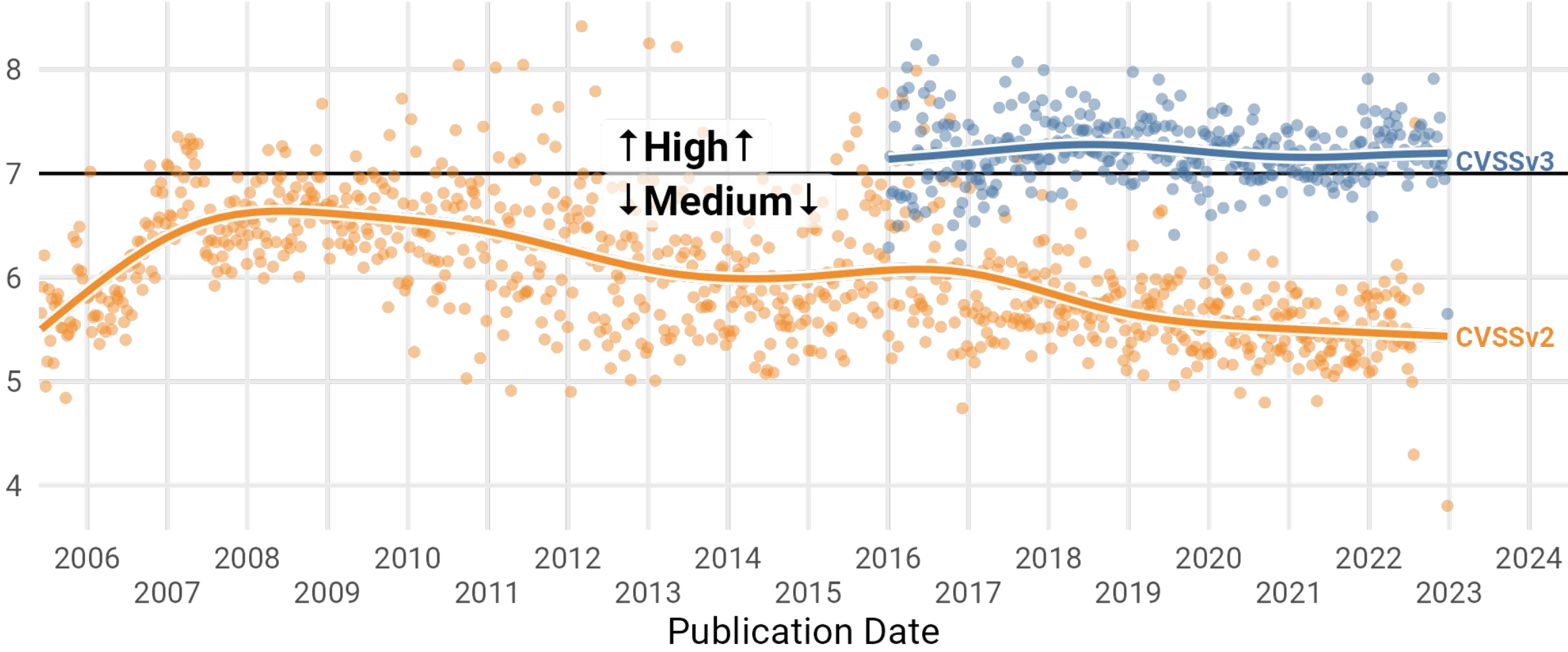


# CVSS Version Distributions Over Time



# CVE Severity Over Time

Mean Weekly CVSS base score



# Noise Alert: Severity Disagreement

## NVD CVSSv3 Score

Critical

High

Medium

Low

CNA  
CVSSv3  
Score

Critical

3.1%  
(n=432)

4.2%  
(n=586)

0.79%  
(n=110)

High

7.2%  
(n=1k)

17%  
(n=2.4k)

9.2%  
(n=1.3k)

0.050%  
(n=7)

Medium

3.9%  
(n=543)

18%  
(n=2.6k)

24%  
(n=3.3k)

0.82%  
(n=115)

Low

0.26%  
(n=36)

1.9%  
(n=263)

7.5%  
(n=1.1k)

1.9%  
(n=270)

None

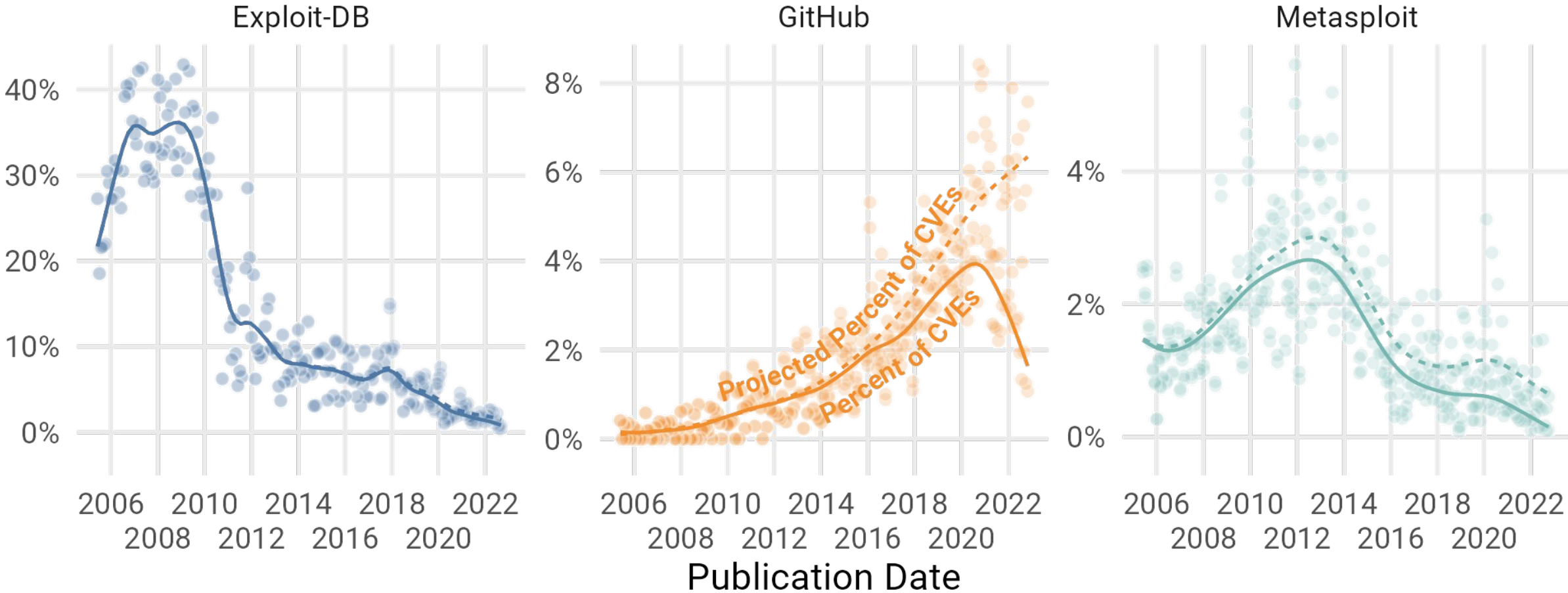
0.029%  
(n=4)

0.043%  
(n=6)

0.029%  
(n=4)

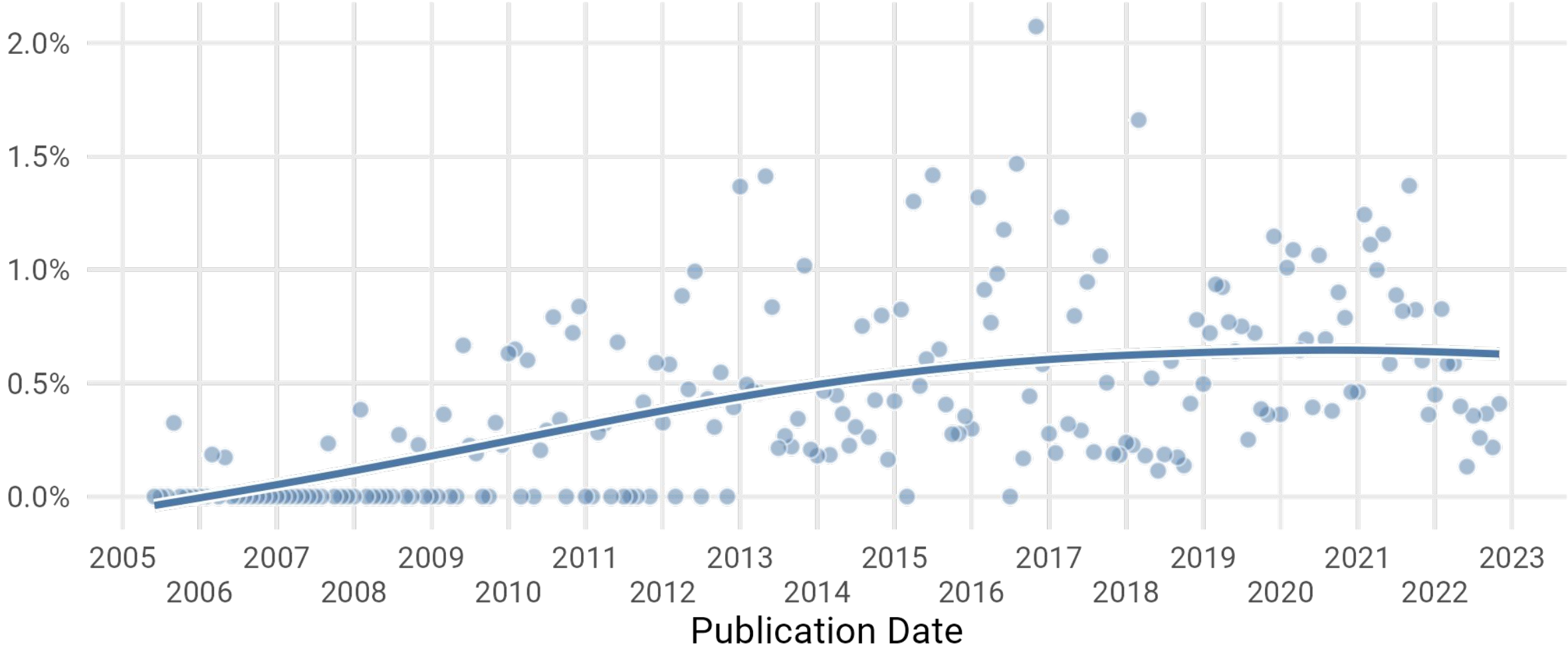
# Exploit Sharing by Repository

Percent of CVEs with Exploits in Various Repositories



# Known Exploitation Activity

Percent of CVEs appearing in the KEV



# Summary & Implications

CVE data processes have changed multiple times

Forecasting requires methods that can handle these changes

Do not use vulnerability data from pre-CNA period for forecasting volume, or incorporate the CNA process into your models

Frameworks are subject to change, CVE, CVSS and CWE (OWASP) have all changed.

The data is not the data - historical data is subject to modification, in ways that may or may not be visible or historically consistent.





<https://www.cyentia.com/>



[f5.com/labs](https://f5.com/labs)