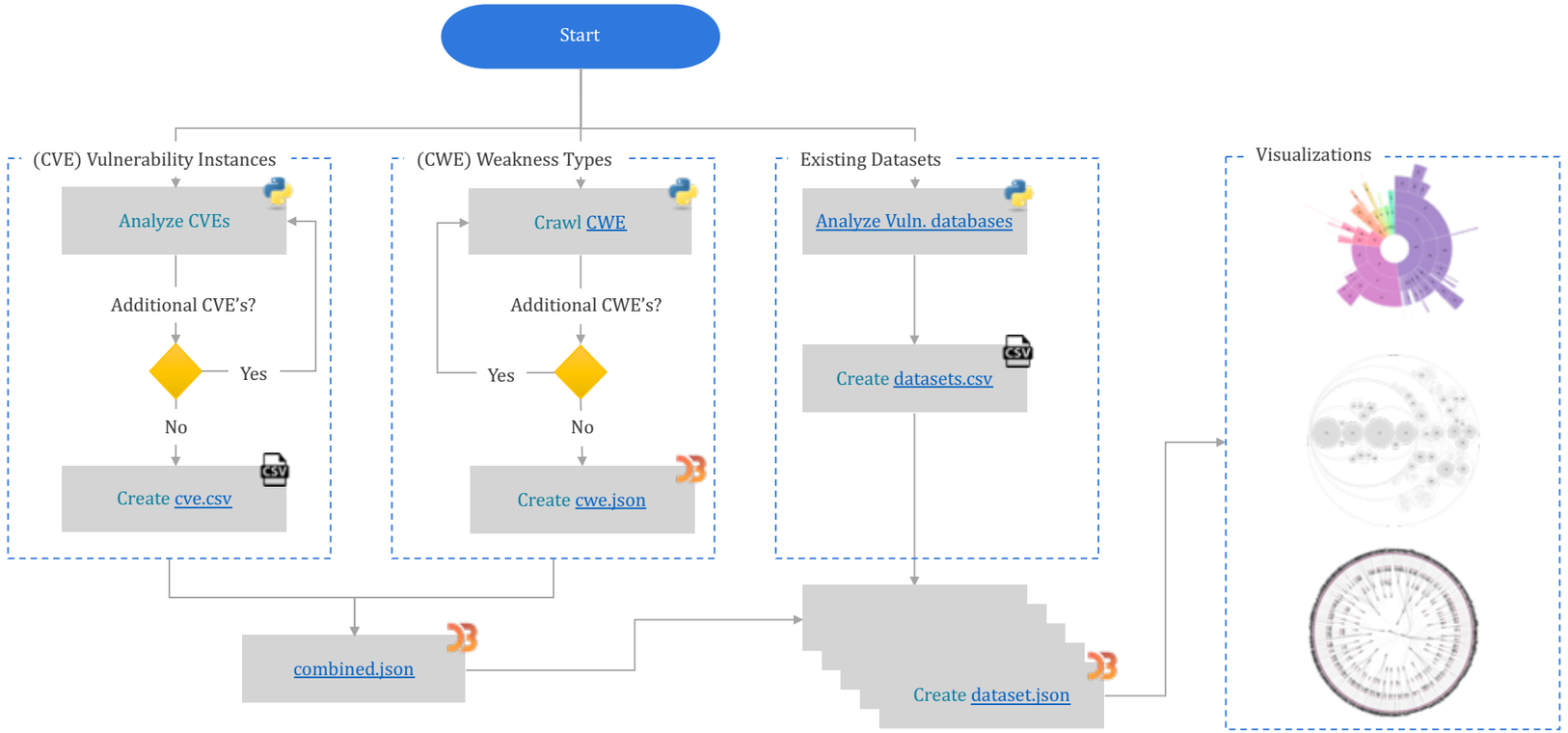
REPRESENTATIVENESS

in the Benchmark for Vulnerability Analysis Tools (**B-VAT**)

Kayla Afanador (Keen)
Naval Postgraduate School

Cynthia Irvine
Naval Postgraduate School

Preliminary Work Paper
Length: Short



Too many **vulnerabilities** to rely on manual analysis alone.

VATs compliment the analysis process, but there are **a lot** of tools…

No standard method (benchmark) to compare the tools.

Vulnerability types **disproportionately** represented

The Problem: No benchmark to compare VATs

Relevant

problems **representative** of reality

Repeatable

results should be consistently reproduced when the benchmark is run with the same tool

Verifiable

confidence that benchmark results are accurate

Usable

able to be used in multiple operating environments, and run with a variety of tools

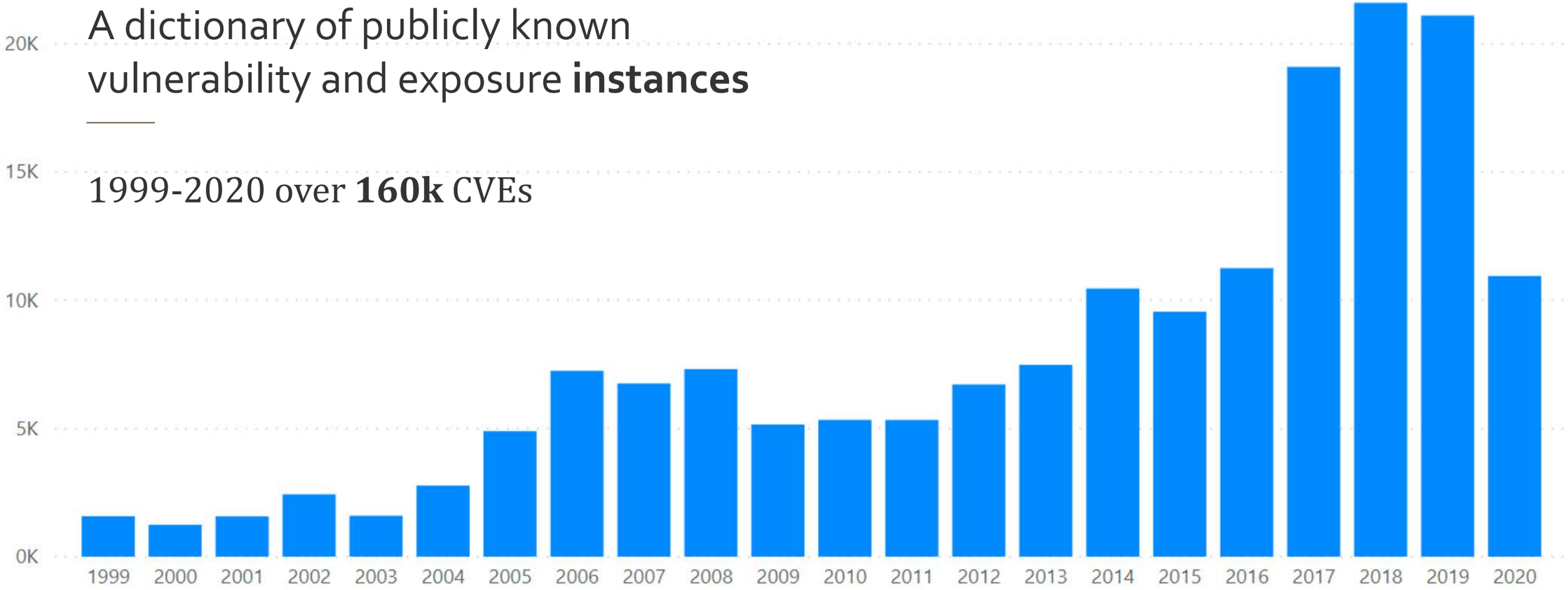
Fair

not be partial to any particular tool

The Solution: B-VAT

A dictionary of publicly known vulnerability and exposure **instances**

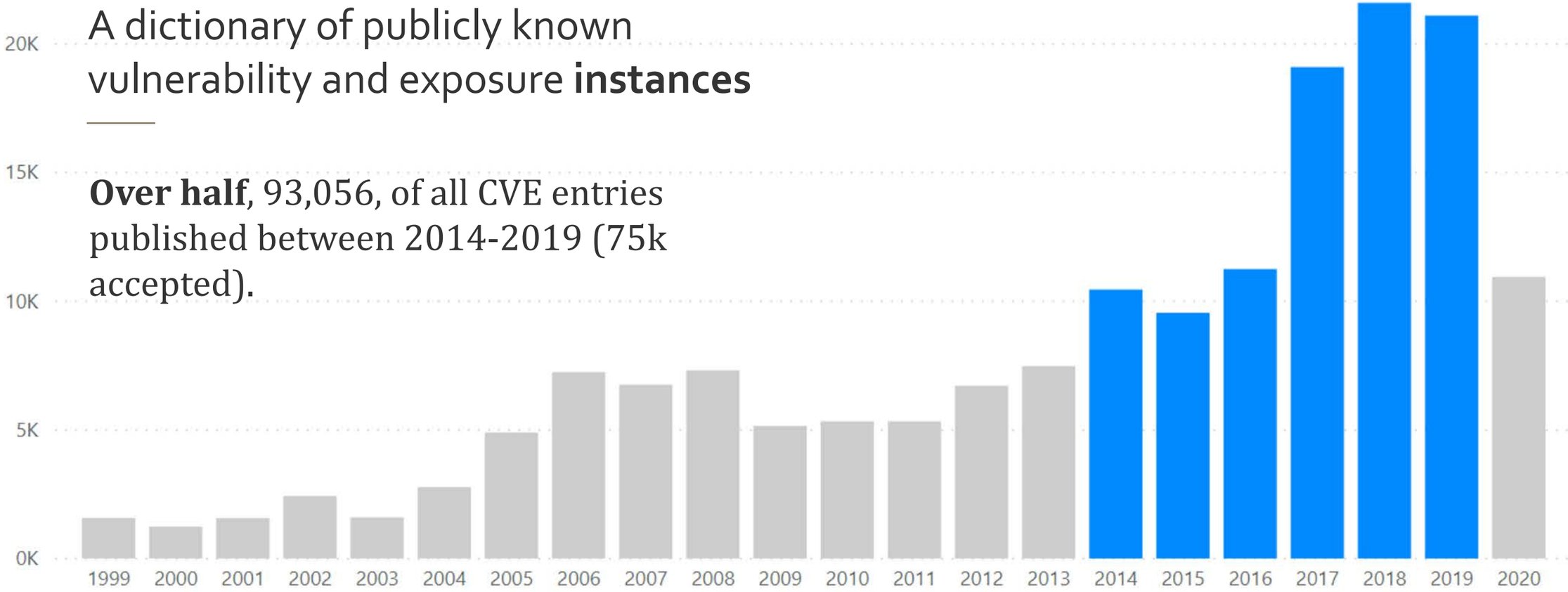
1999-2020 over **160k** CVEs



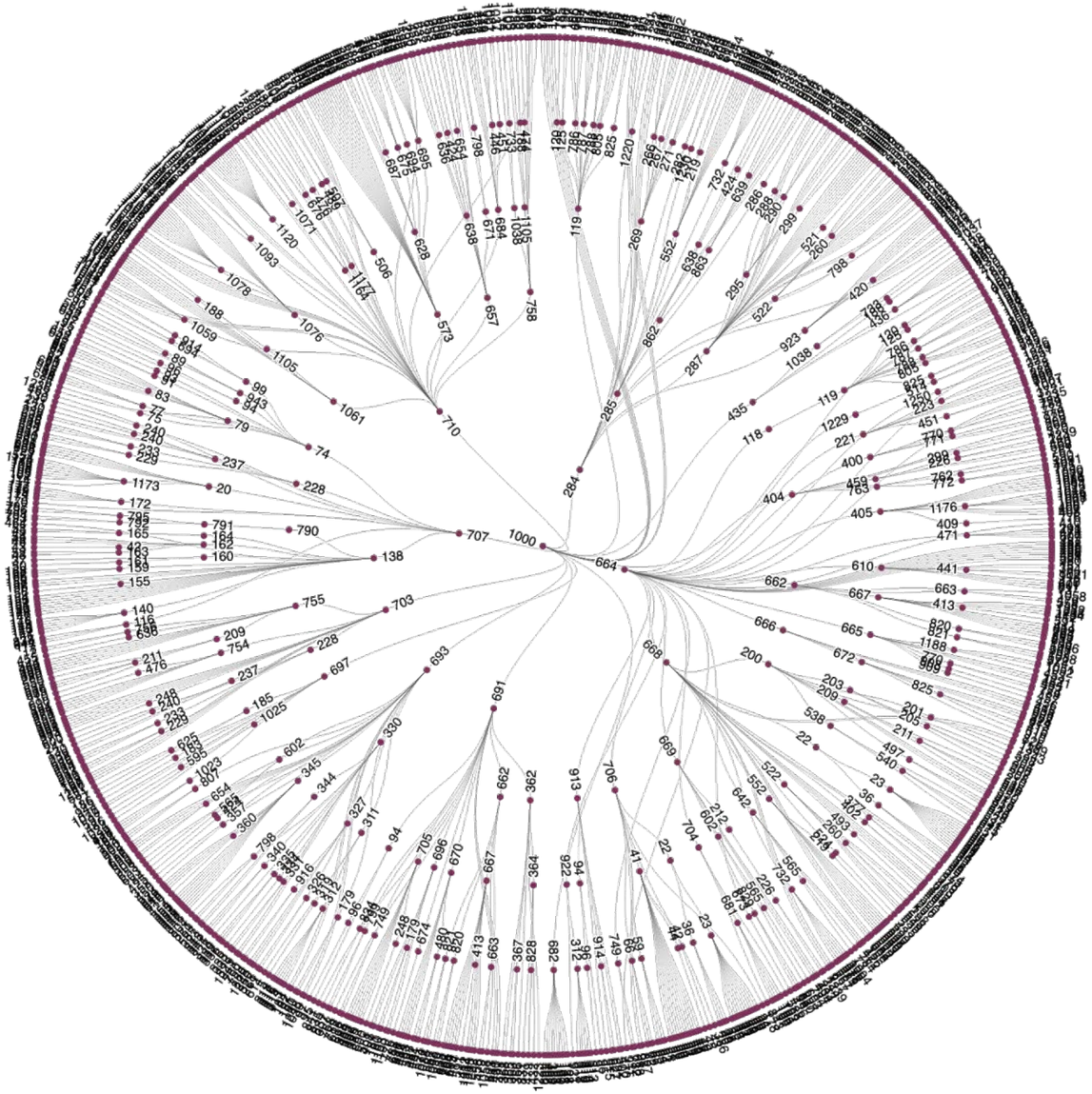
CVE's as vulnerability instances

A dictionary of publicly known vulnerability and exposure **instances**

Over half, 93,056, of all CVE entries published between 2014-2019 (75k accepted).



CVE's as vulnerability instances

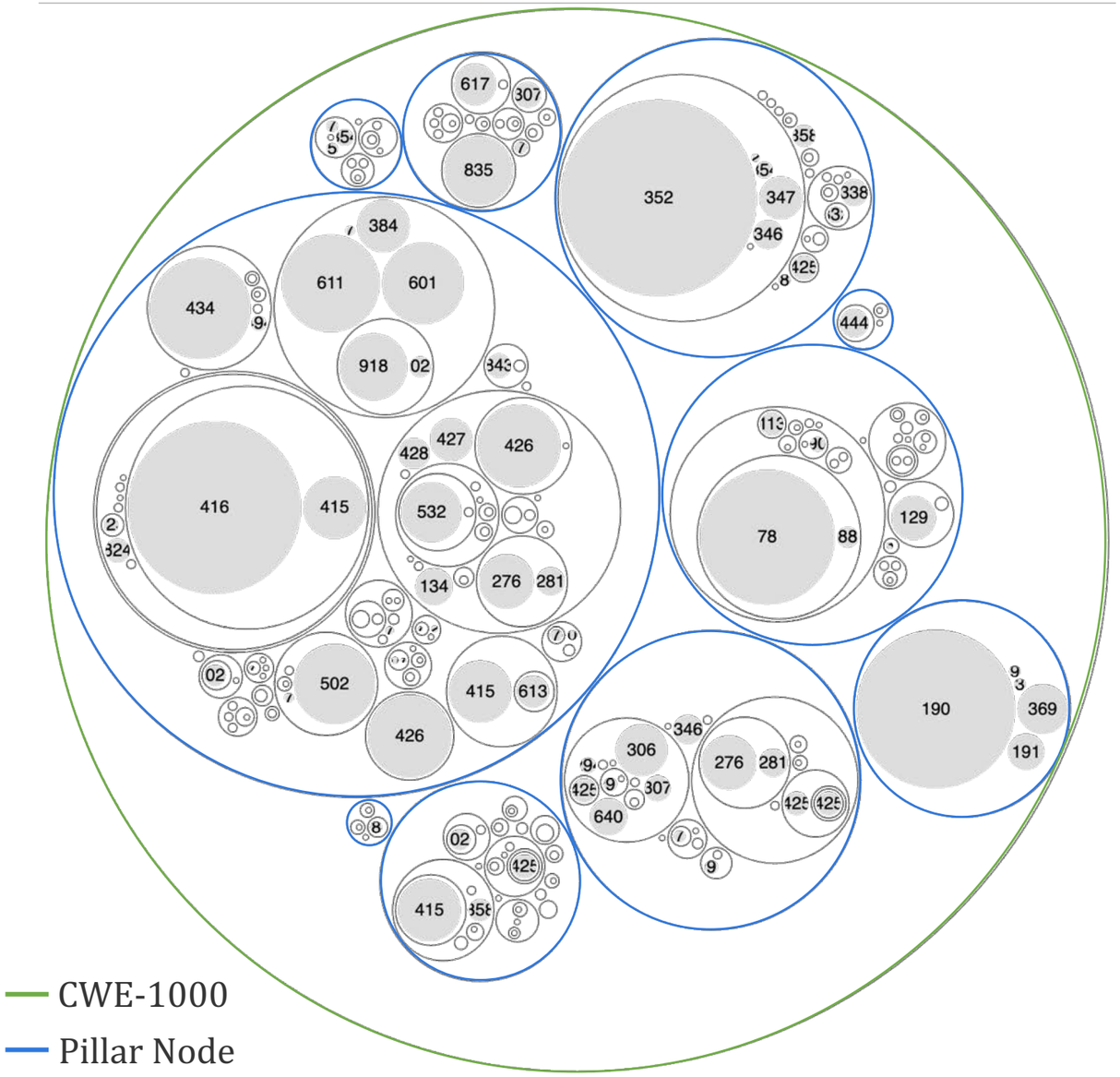


Community developed list of **weaknesses** with security ramifications

Crawled over **1k** CWE pages to create **tree data structures** for each of the ten CWE Pillars.

Use root node (1000) to create **single rooted tree**

CWE's as Weakness Types

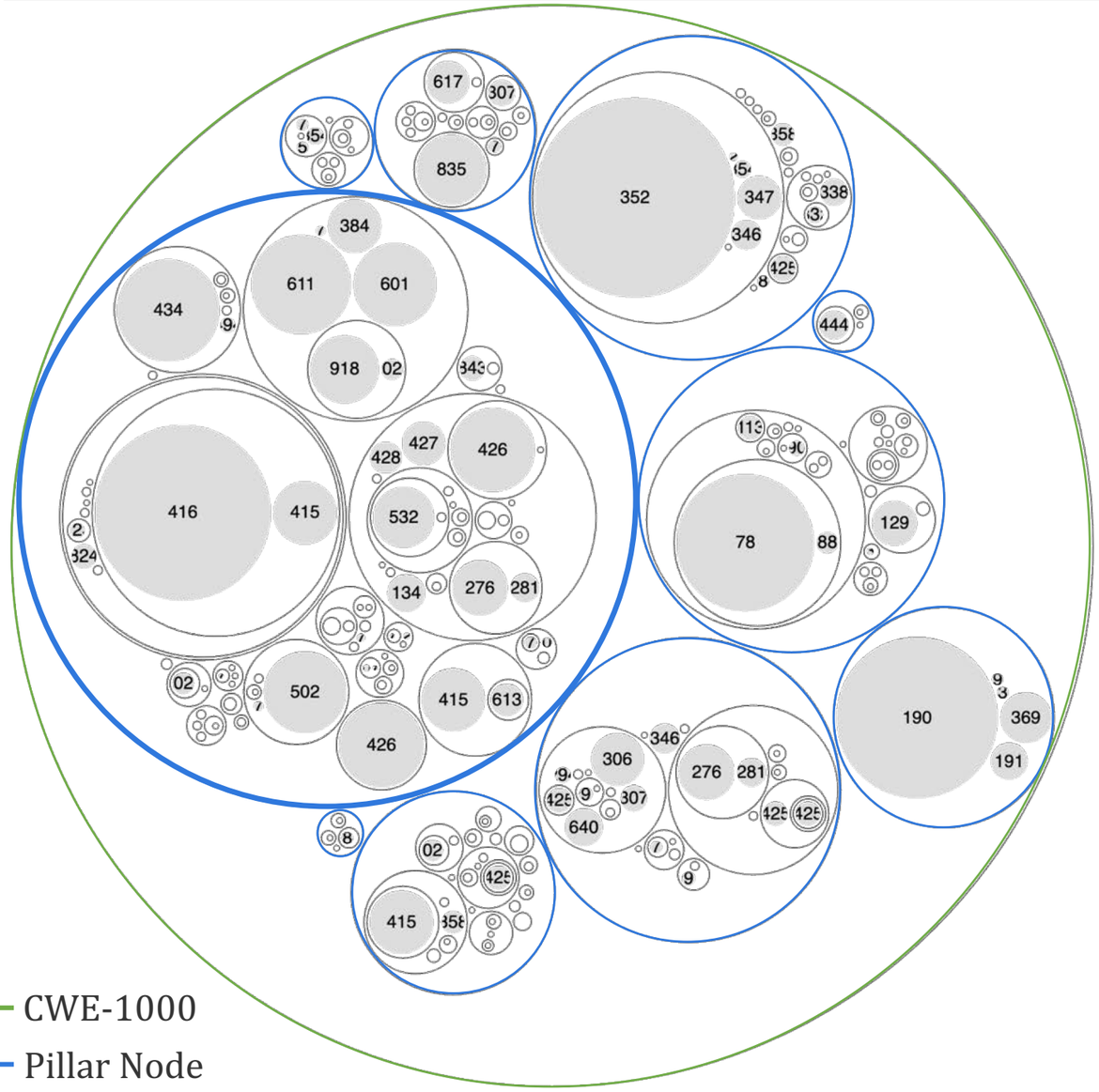


Use existing CVE/CWE correlation to classify **vulnerability instances** by associated **weakness type**

55,128 CVEs with associated CWE ID

Trace each CVE to 1 of **10 CWE pillars** (the most abstract weakness types)

CVE's & CWE's to create a representative set

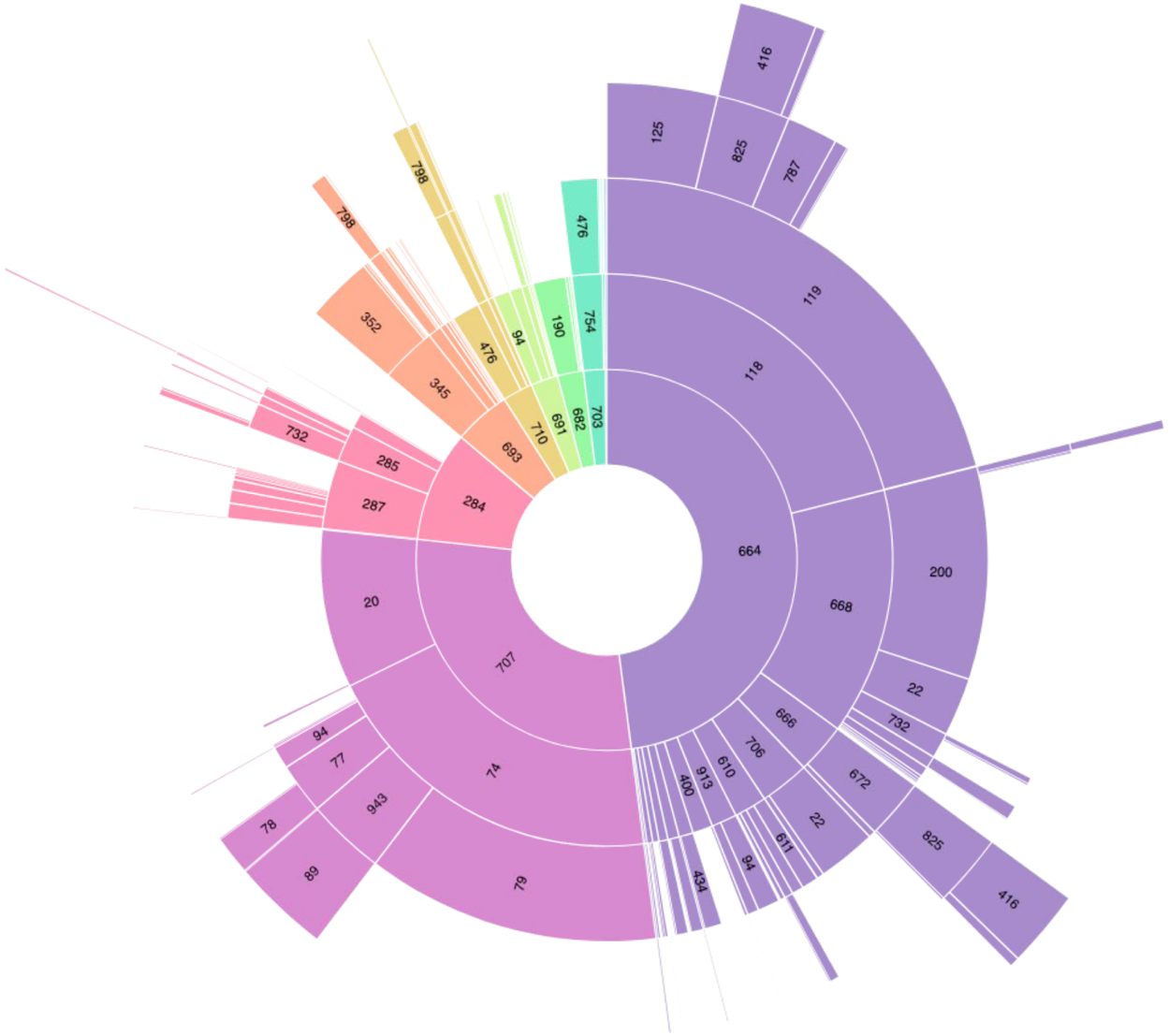


Representative Set:

a subset of test cases that adequately represents the larger set of known vulnerability **instances** and **types**

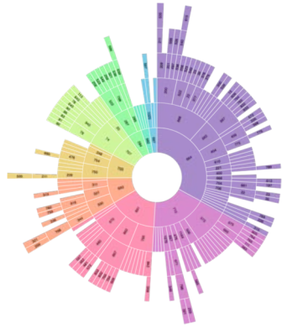
Pillar node **CWE-664** represent **45% of CVE's** from 2014-2019

CVE's & CWE's to create a representative set

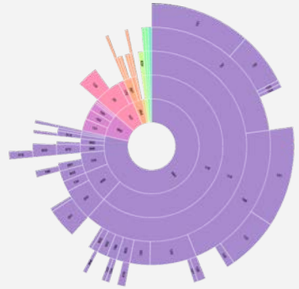


The representative set

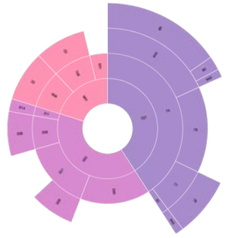
Juliet C/C++



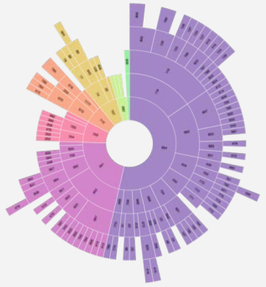
Juliet Java



CGC Corpus



OWASP Benchmark



Stonesoup

Coming Soon

B-VAT

Existing datasets may not be representative

Random sampling results in the **misrepresentation** of vulnerability instances and weakness types

Stratified Sample:

Allows sub-groups or "**strata**" to be proportionately represented

Provides a **representative** sample of a larger population

Preserves the relative proportions of each pillar

Pillar	CVE's	Stratified Sample
CWE-284	5,847	245
CWE-435	40	2
CWE-664	24,957	1,042
CWE-682	1,397	58
CWE-691	1,419	59
CWE-693	2,571	107
CWE-697	15	1
CWE-703	168	7
CWE-707	17,657	737
CWE-710	1,030	43
	55,128	**2,301**

Identifying a representative subset for B-VAT

Relevant

problems **representative** of reality

Repeatable

results should be consistently reproduced when the benchmark is run with the same tool

Usable

able to be used in multiple operating environments, and run with a variety of tools

Fair

not be partial to any particular tool

Verifiable

confidence that benchmark results are accurate

Pillar	Required Test Cases	Available Test Cases
CWE-284	245	3,309
CWE-435	2	42
CWE-664	1,042	92,733
CWE-682	58	28,876
CWE-691	59	1,511
CWE-693	107	3,321
CWE-697	1	76
CWE-703	7	1,117
CWE-707	737	34,417
CWE-710	43	7,236

Recap & Next Steps

THANK YOU

Special thanks to Dr. Lyn Whitaker for the valuable discussions

CONTACT US:

Kayla Afanador (Keen)
knkeen@nps.edu

Cynthia Irvine
irvine@nps.edu