Web Attacks Data

Gaberial Campese

Individual Project 2

DATS 6103

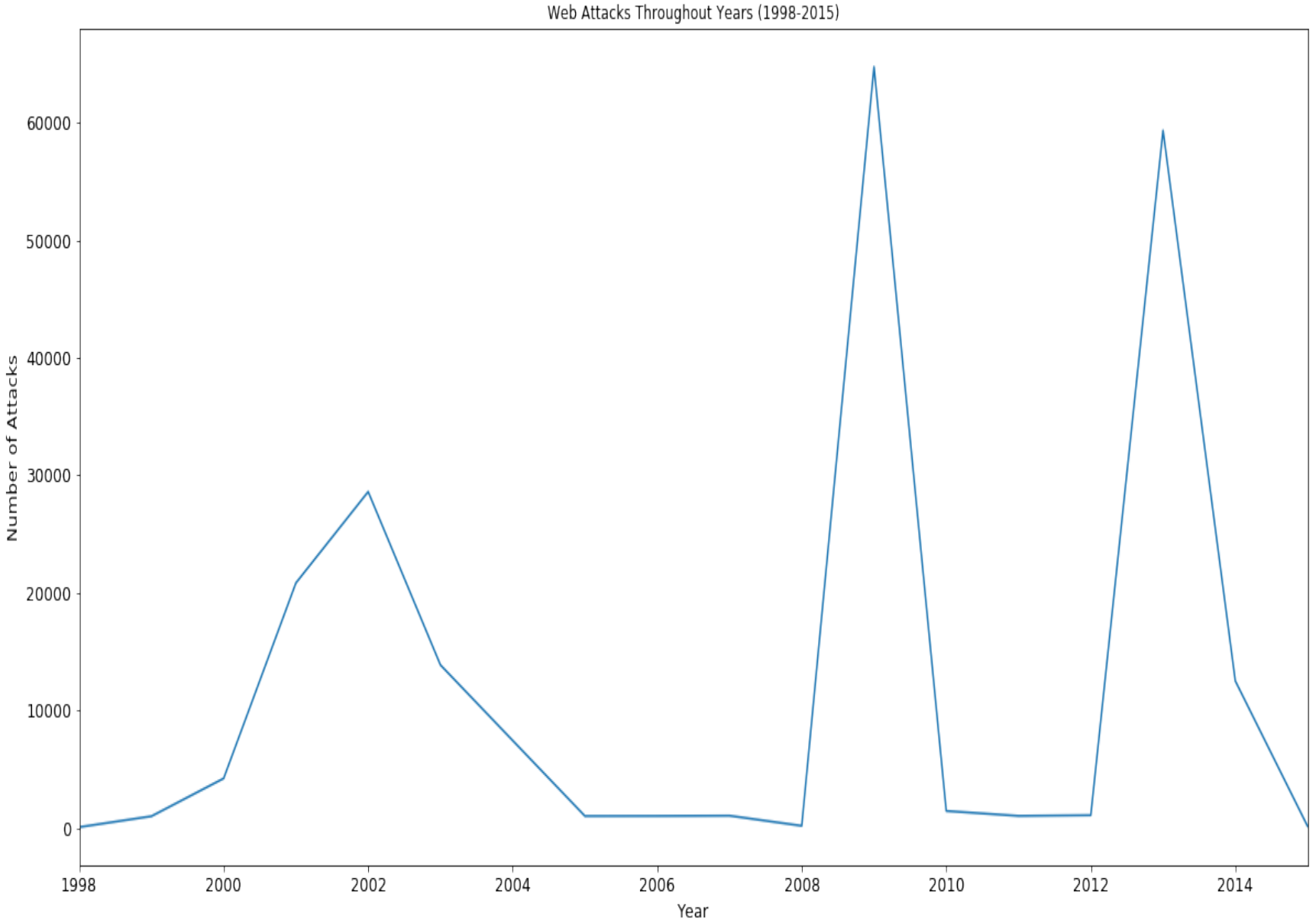
Introduction

- Web-Hacking Dataset
- Historical Overview of Web Attacks (1998-2015)
- Countries of Occurrence Analysis
- Top Victim OS' and Web Servers Overall
- Top Perpetrators based on number of attacks
- Most Targeted OS' and Web Servers among Top Perpetrators
- Most Targeted Countries among Top Perpetrators
- OS, Web Servers, and Countries targeted frequency by each individual group
- Code Review
- Conclusion

The Data

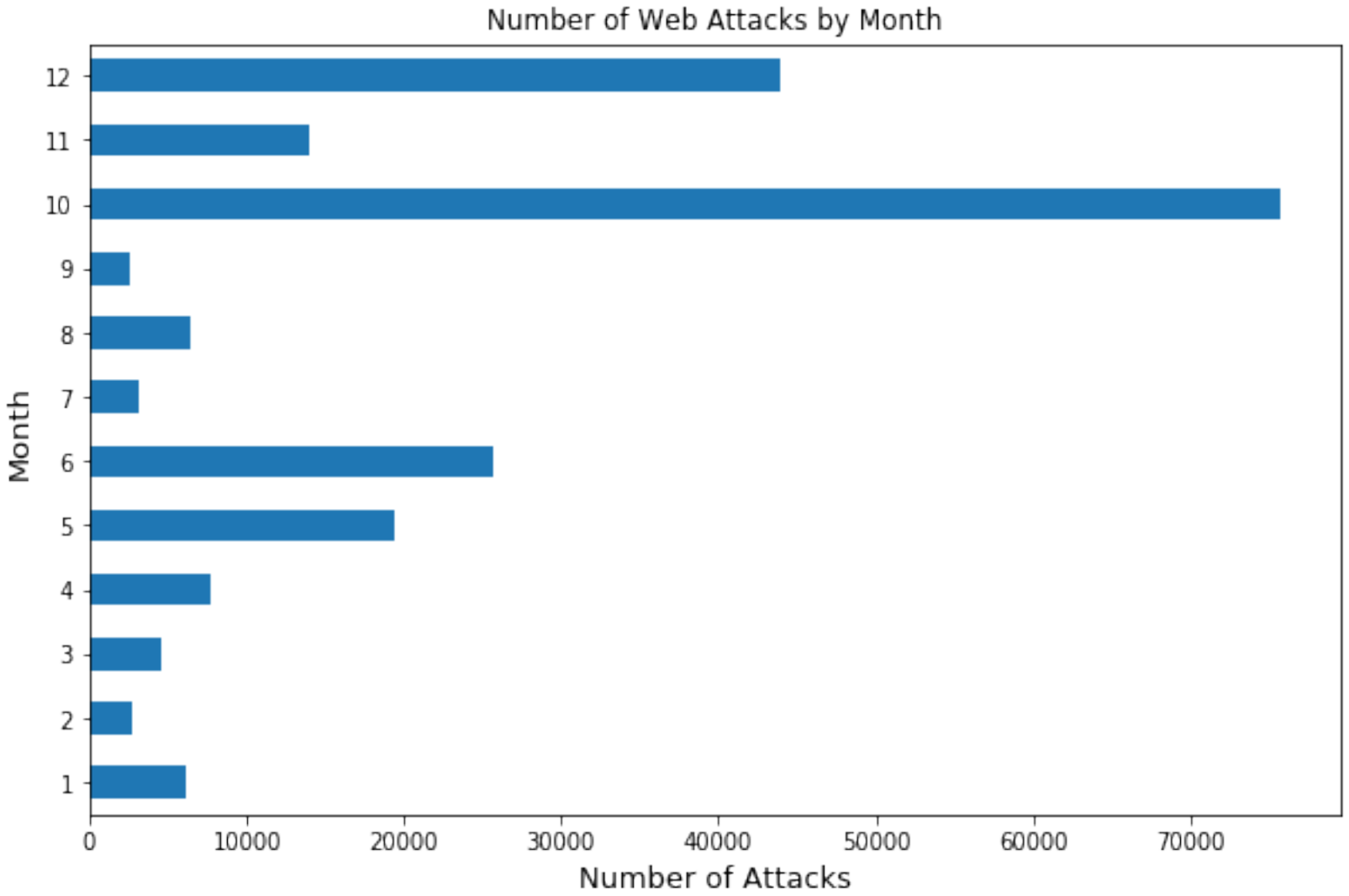
- Web-Hacking Excel Datasheet from Hacking and Countermeasure Research Lab
- 212,093 rows and 9 columns
- Years: 1998-2015
- Rows represent instances of attacks
- Columns: **Date**, **Notify**, URL, IP, **Country**, **OS**, **WebServer**, Encoding, Lang
- Used Python to conduct Data Mining & Visualization

Historical Overview of Web Hacking



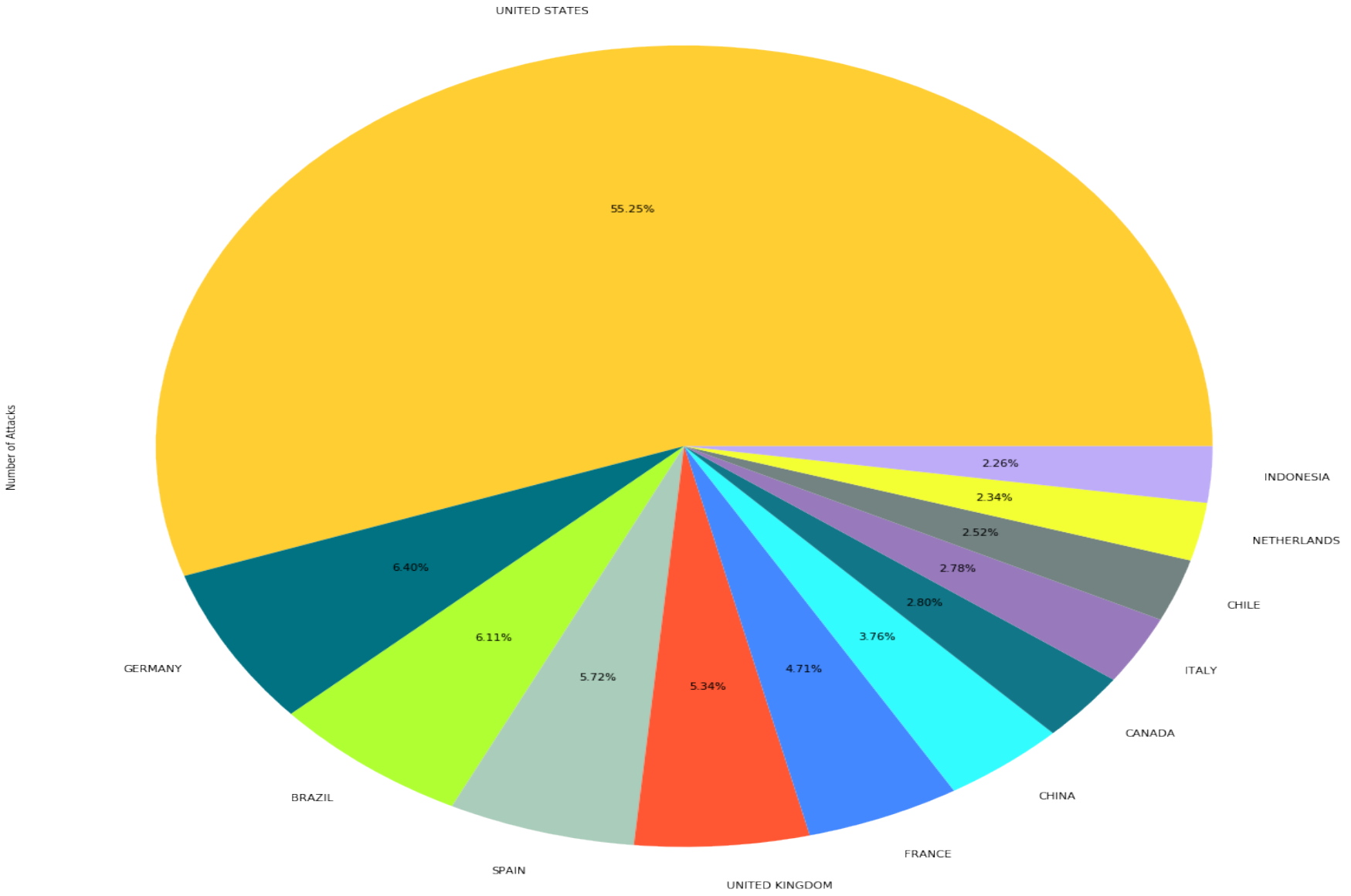
- Total number of web attacks
- Summed for each year
- Peak in 2009 followed by 2013
- Dataset ends in Sept. 2015

Top Months of Occurrence



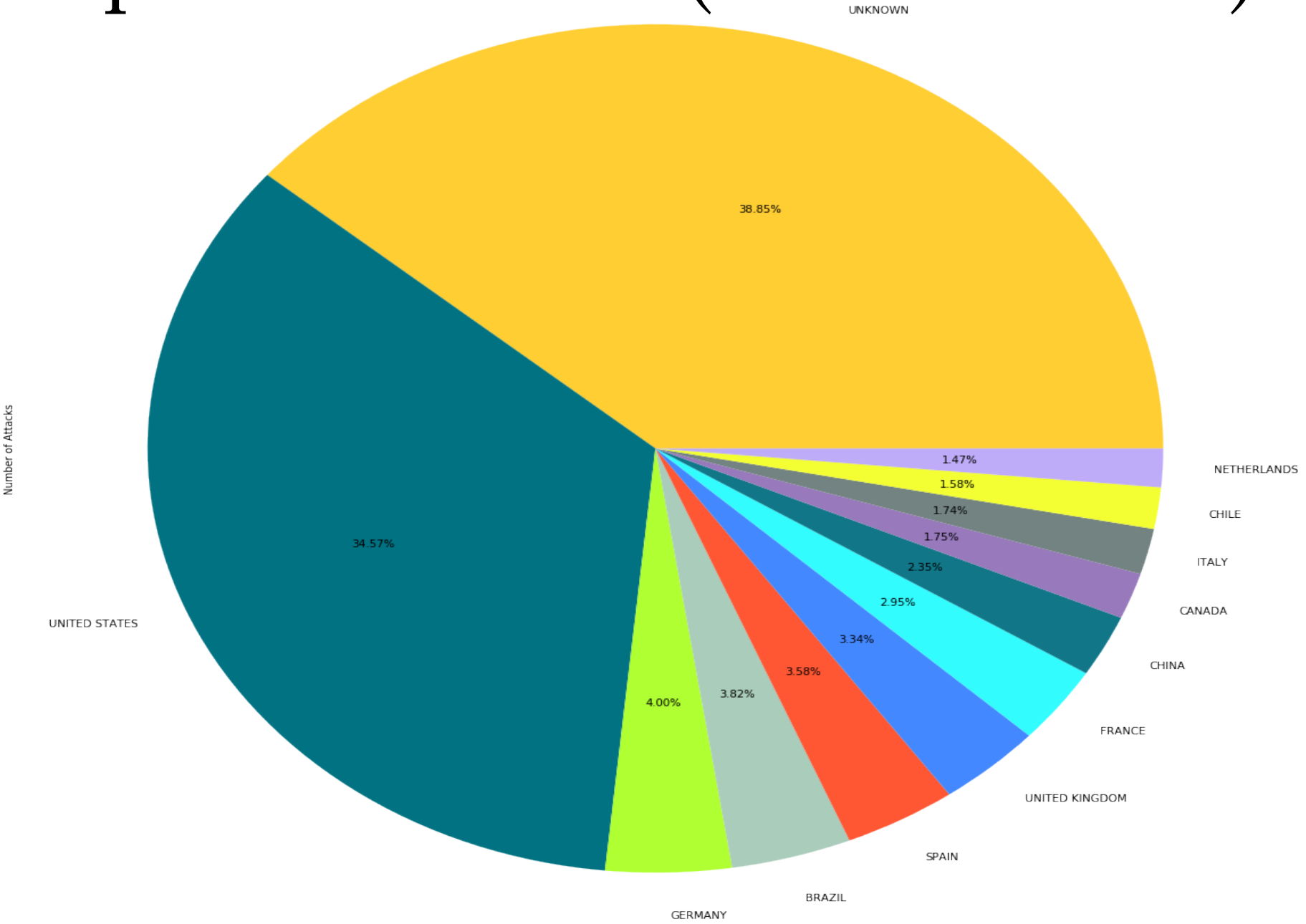
- Total number of web attacks
- Summed by each month
- October by significant margin
- December follows

Top 12 Countries



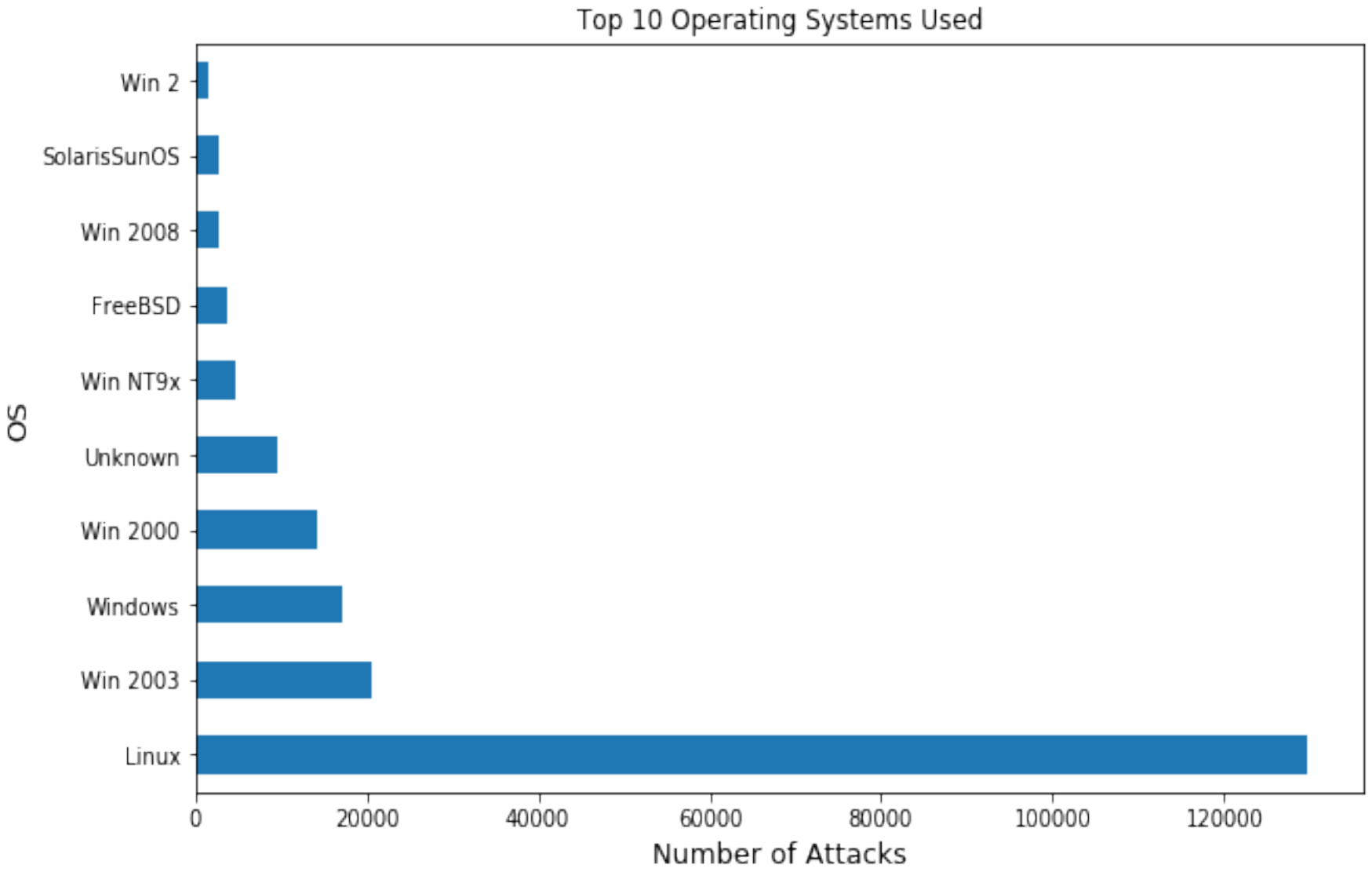
- Top 12 countries of occurrence
- United States accounts for 55.25% of top 12
- Germany follows at 6.4%

Top 12 Countries (w/ Unknowns)



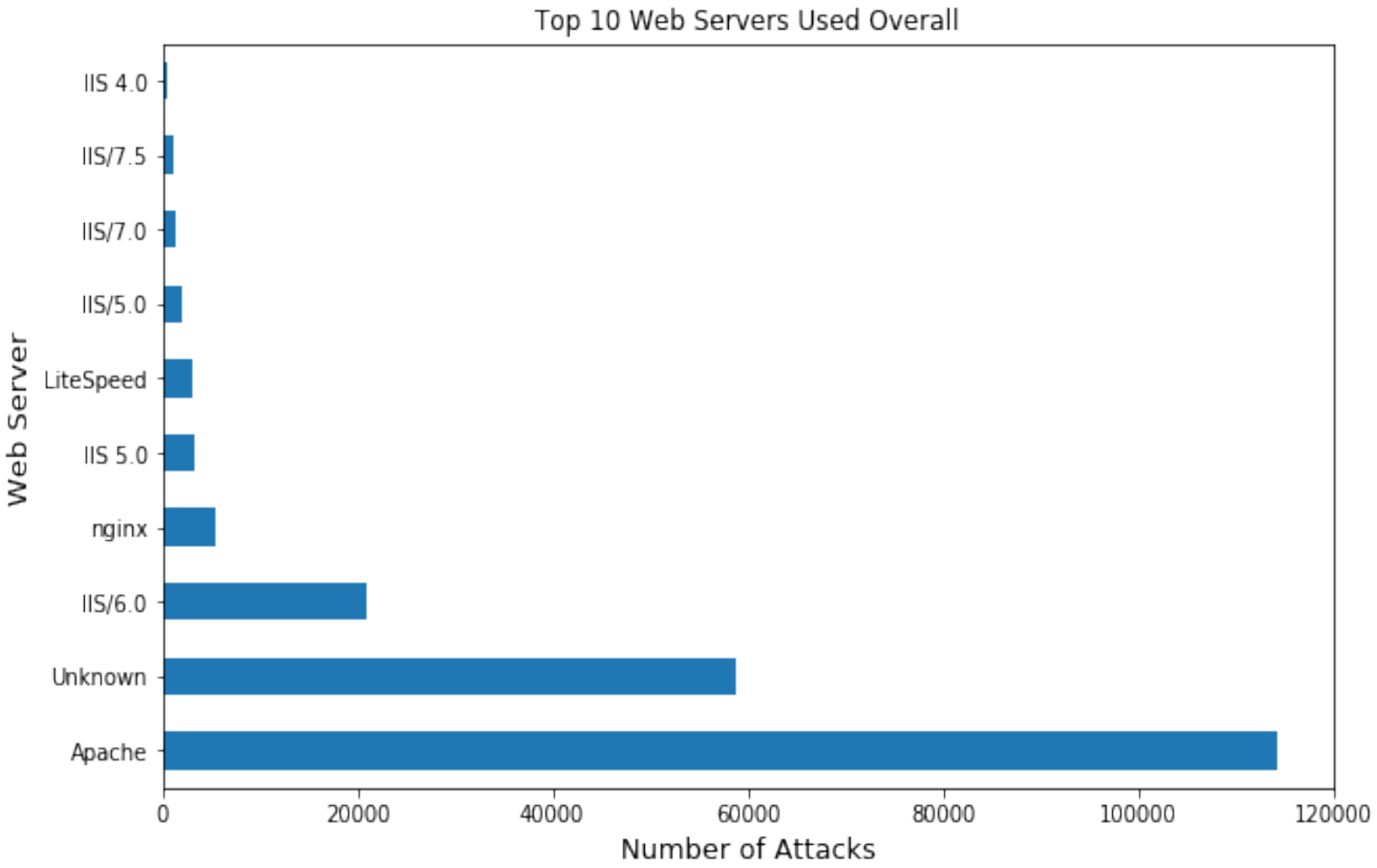
- United States still accounts for 34.57%
- Unknowns at 38.85%
- Indonesia drops

Top 10 Targeted OS' Overall



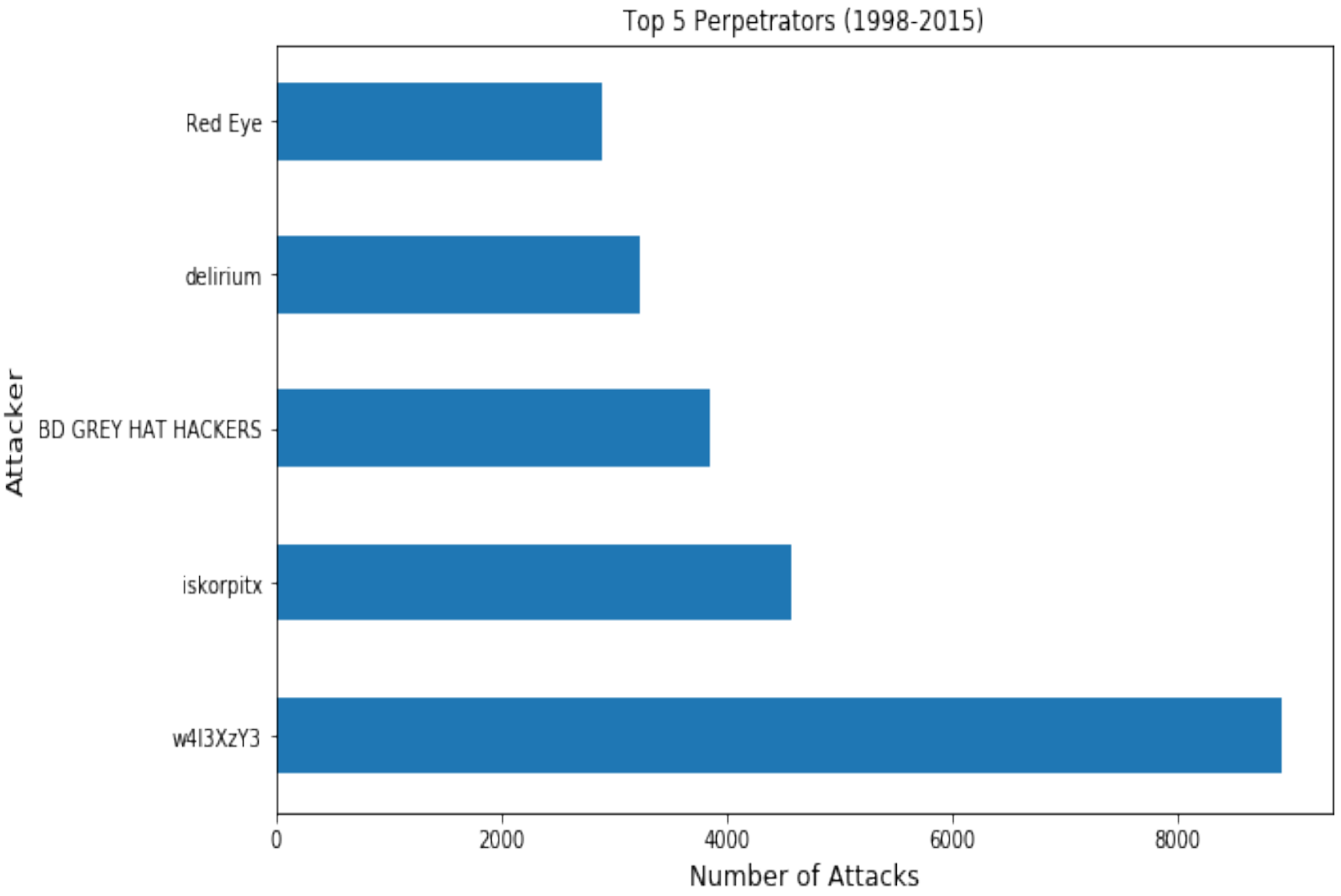
- Linux OS of choice by far
- ~61.2% Linux
- Notice Windows OS versions

Top 10 Targeted Web Servers Overall



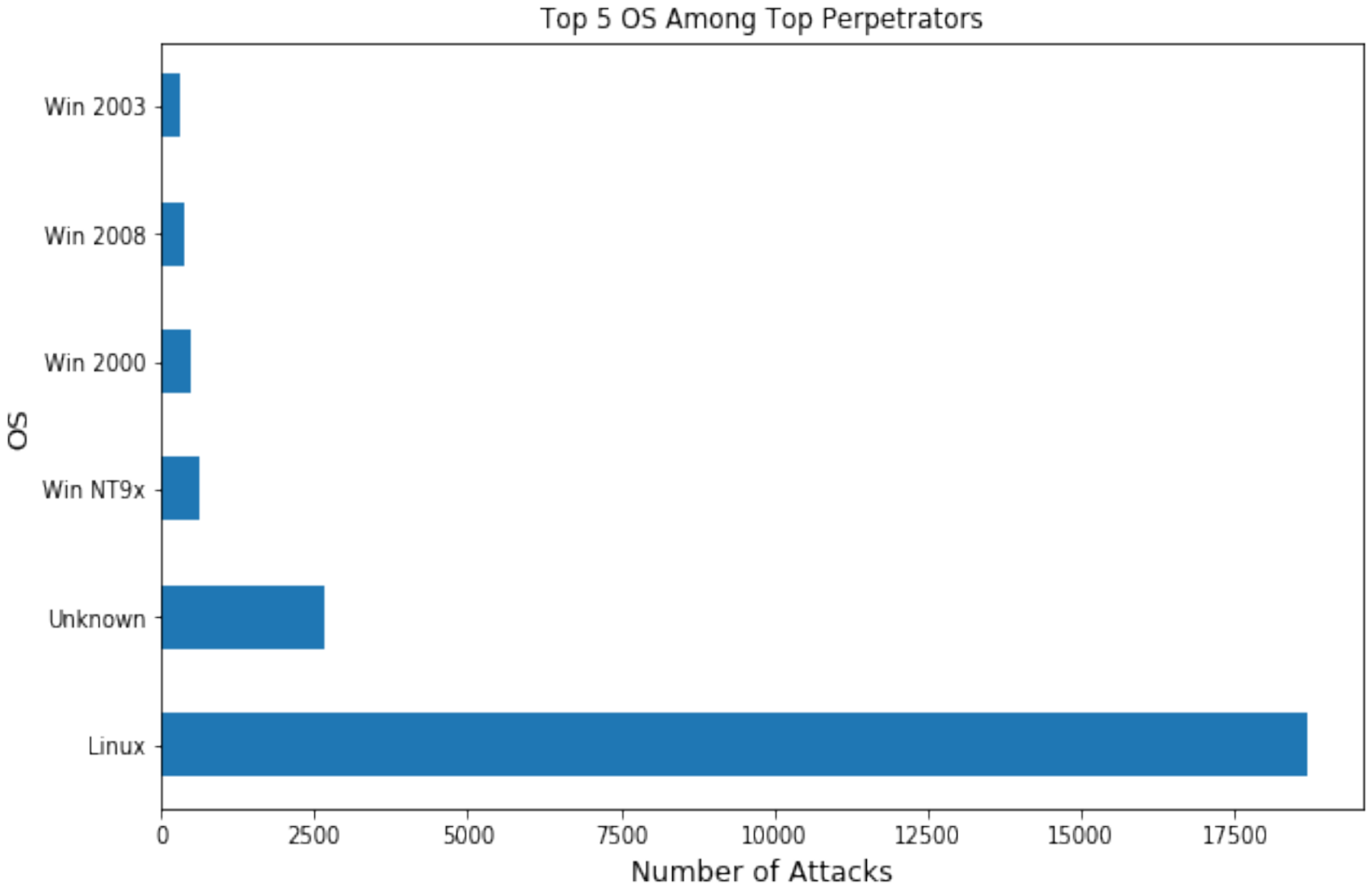
- Apache most popular Web Server
- Apache targeted ~53.9% of time
- Pay detail to IIS and Unknowns
- Still not close

Key Perpetrators



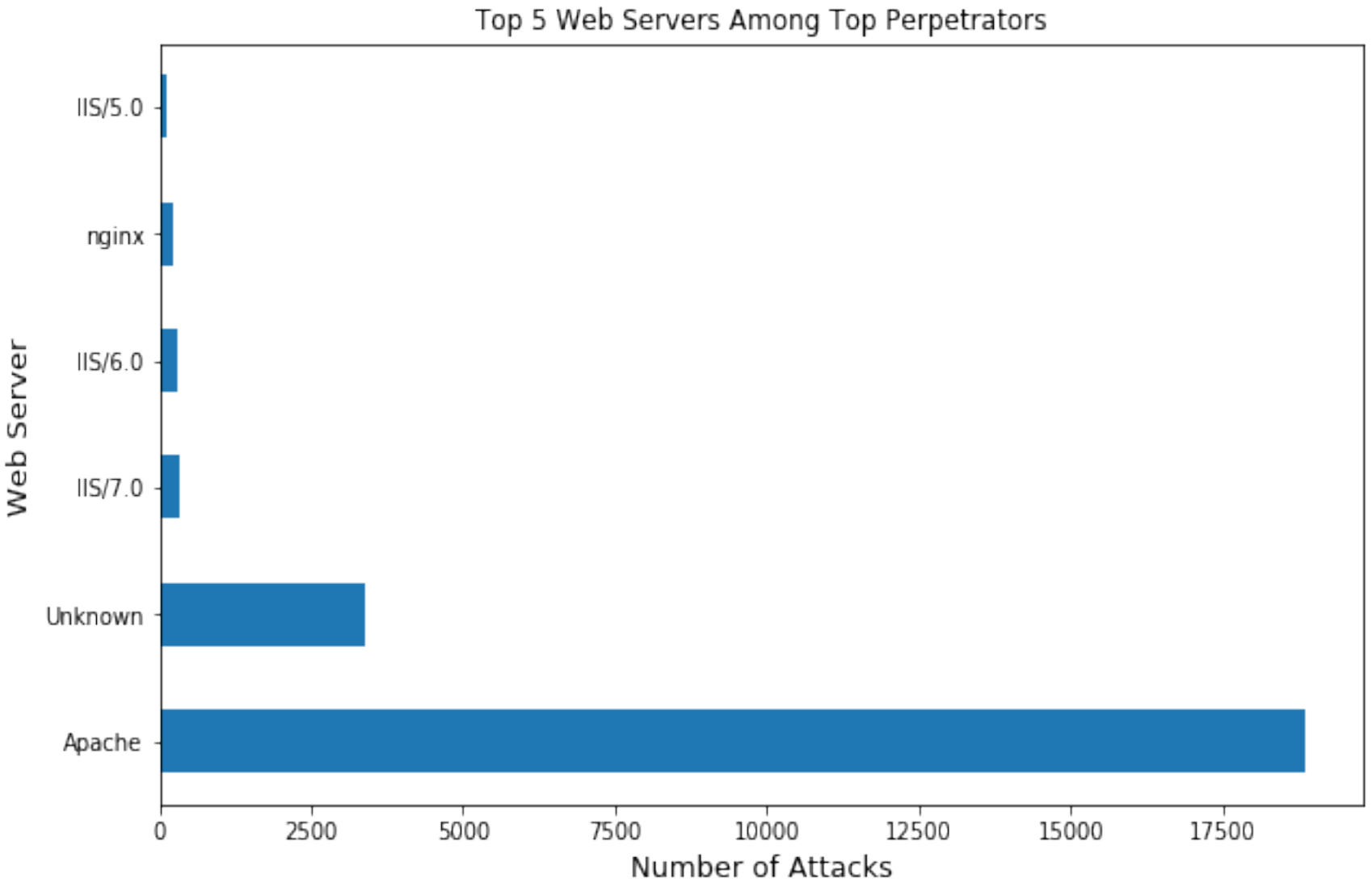
- w4l3XzY3 most active group with 8,928 attacks
- w4l3XzY3 ~4.2%
- Top 5 accounts for ~11.1% of total

Top Operating Systems Among Top Perpetrators



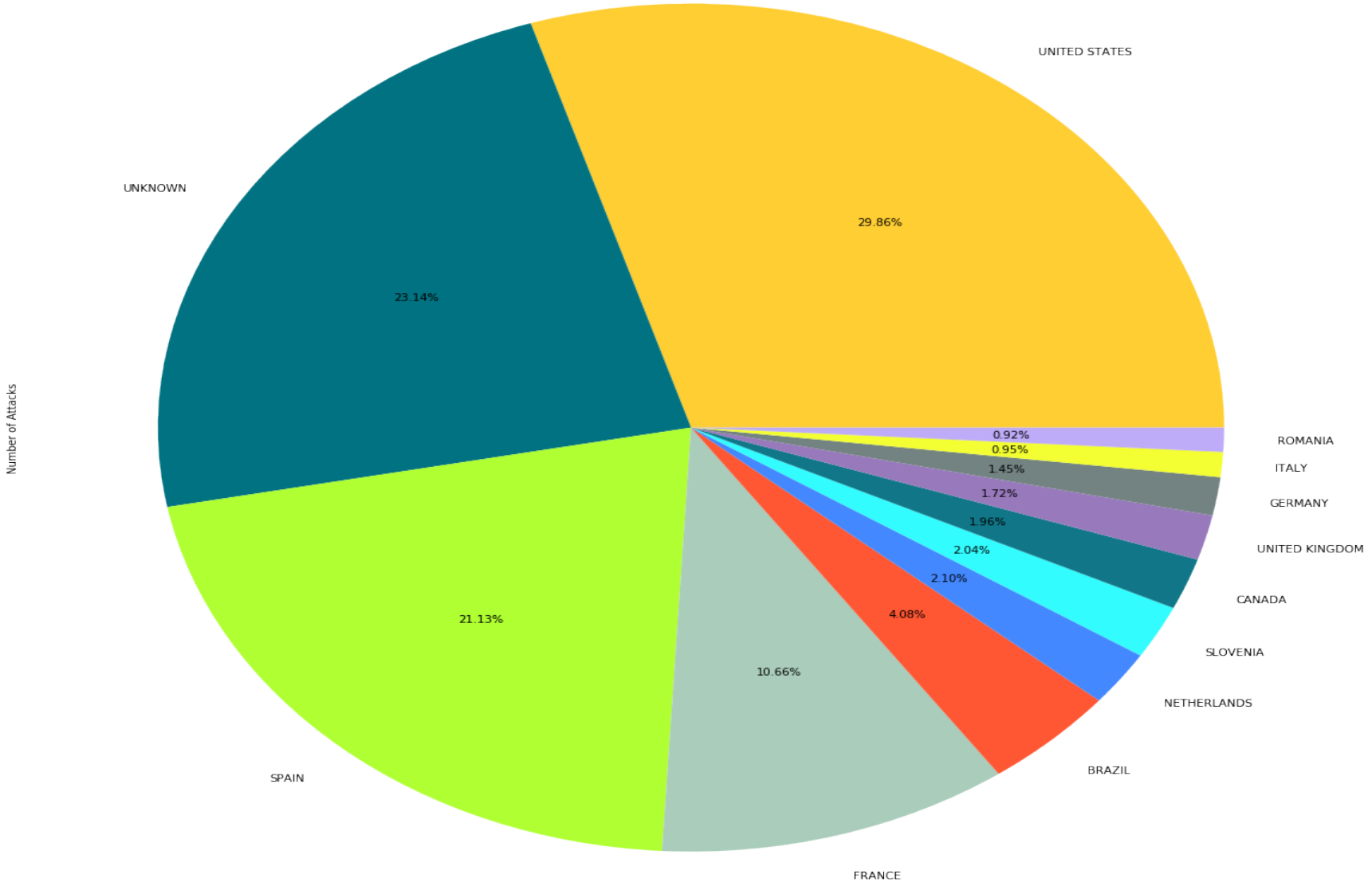
- Compare Linux frequency with Overall
- ~79.5% Top 5 attacks feature Linux
- 23,156/23,483
- Notice Windows versions

Top Web Servers Among Top Perpetrators



- Apache is most targeted Web Server among Top 5
- ~80.3% Apache
- Internet Information Services

Countries of Occurrence Among Top Perpetrators



- USA at 29.86%
- Unknowns at 23.14%
- Spain at 21.13%
- Recall Overall distribution

Individual Scoping of Top Perpetrators

- Developed 3 functions in Python
  - OS Function
  - Web Server Function
  - Countries of Occurrence
- Can be used for any Attacker
- Profile Example

“BD GREY HAT HACKERS” Profile

Country	Number of Attacks
UNITED STATES	3419
CANADA	76
UNITED KINGDOM	70
RUSSIAN FEDERATION	42
GERMANY	34

OS	Number of Attacks
Linux	3504
Unknown	195
Win 2008	55
FreeBSD	36
Citrix embedded	23

WebServer	Number of Attacks
Apache	3506
nginx	163
IIS/7.5	61
LiteSpeed	55
Oversee Turing v1.0.0	23

- Caution: Every group is different

Code Review

- Navigate to Jupyter Notebook
- Brief Code Review
- Python libraries used to analyze data
- Visualization

Key Findings/Analysis

- Over **30.5%** of web-hacking instances occurred in 2009, nearly **30%** in 2013
- Over **55%** of attacks occurred in United States among known countries (over **34%** w/ unknowns) and nearly **30%** among Top 5
- Pattern with overwhelming Linux usage both overall (**61.2%**) and within Top 5 (**79.5%**)
- Apache is most targeted Web Server by large margin in web-hacking at **53.9%** overall, **80.3%** among biggest perpetrators
- w4l3XzY3: Spain, **Linux**, **Apache**
- iskorpitx: **Unknown, Linux**, **Apache**
- BD GREY HAT HACKERS: **United States**, **Linux**, **Apache**
- delirium: **United States**, Unknown (Linux), Unknown (Apache)
- Red Eye: Unknown (USA), Linux, Unknown (Apache)

Conclusion/Predictions

- United States is prime country of web-hacking occurrences
- Data breach reporting
- High focus on Linux OS & Apache Web Server
- Overall vs. Top 5 Pattern Comparisons
- Significance of cyber criminal profiling
- Difficulties in recognizing patterns among perpetrators
  - Good to see patterns, yet hard to narrow down
- Identifying who, when, where, and how web-hacking occurs
- Volatile, but ever so prominent
- The Future
- My learning processes (in-class and experience)