Open Source Vulnerability Metrics

Dr. Tate

Seth Goodwin, Michael Follari, Jaron Dunham,
Gabe Wilmoth, Rohit Gade

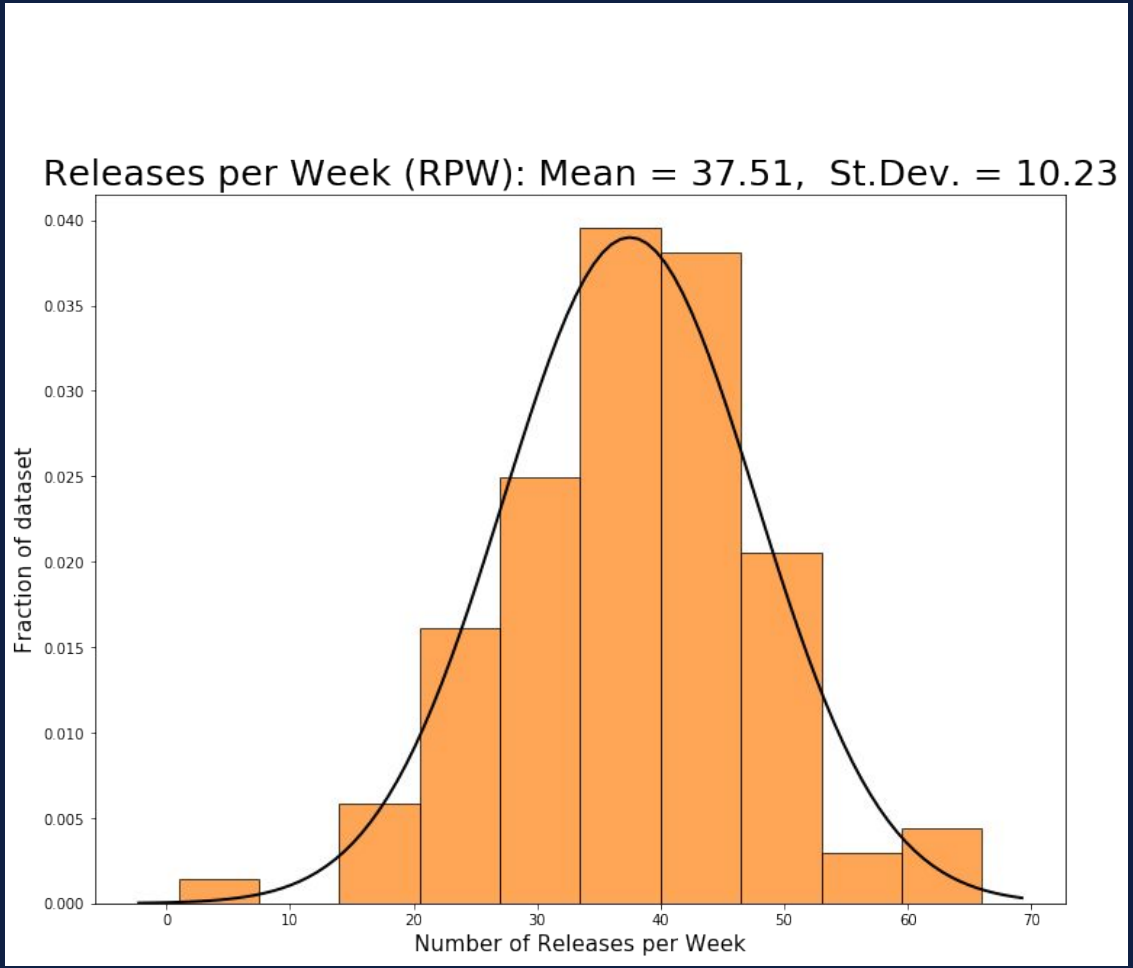
UNC GREENSBORO

Vulnerabilities vs. Software Releases

- Task
  - Statistics of Releases
- Hypothesis Test
  - H0: Weeks after a period of high or low amount of vulnerability discoveries will have the same number of releases as any other week.

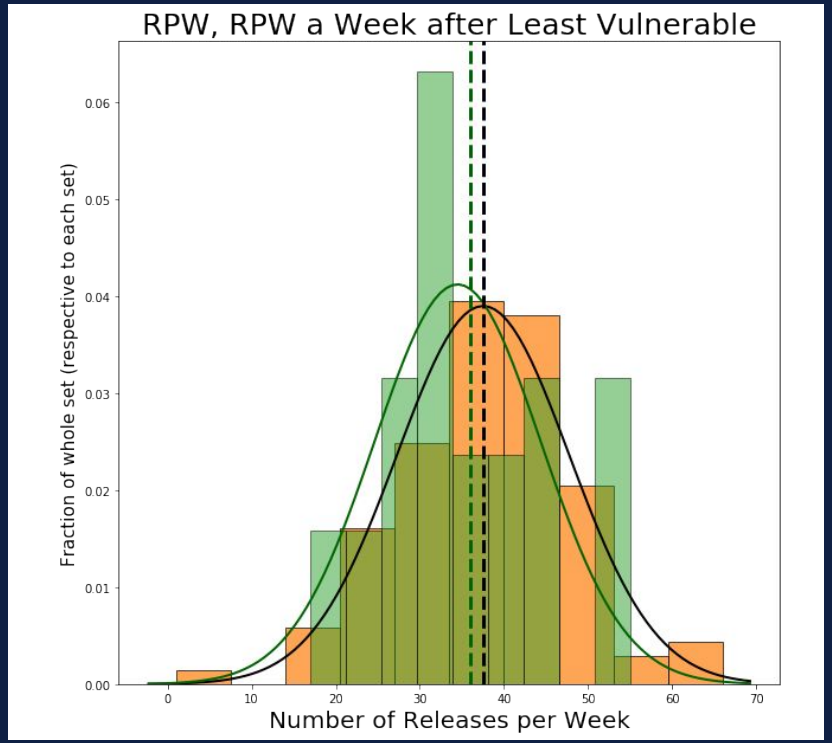
Statistics of Releases

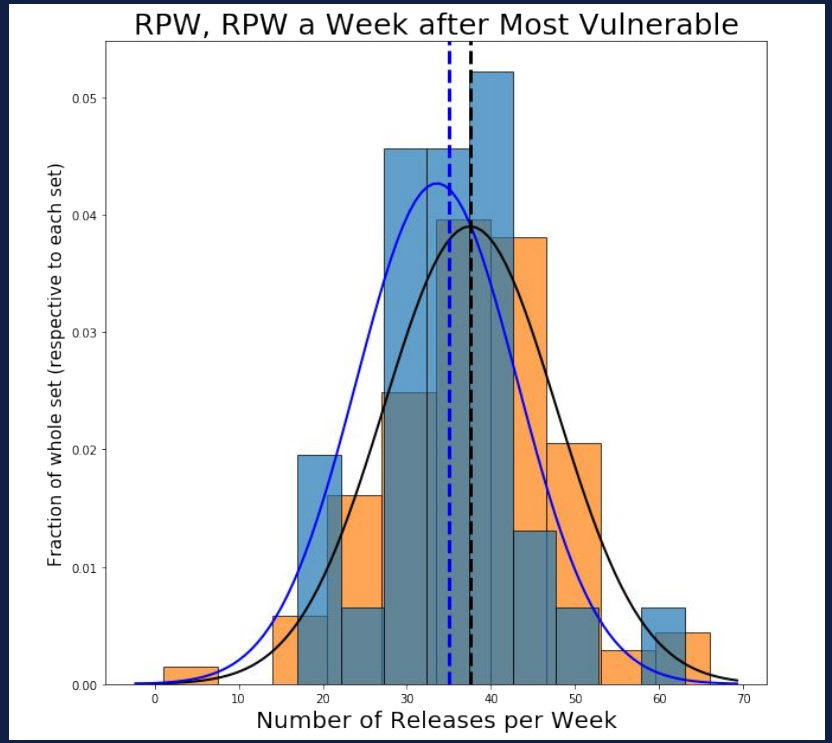
- Normal Distribution
- Mean: 37.51
- Median: 37.0
- Std. Dev: 10.23
- Variance: 105.7



UNC GREENSBORO

RPW One Week After Vulnerable Weeks





UNC GREENSBORO

Hypothesis

- Two Sample T-Test
  - P-Value for most vulnerable: 0.244
  - P-Value for least vulnerable: 0.46

Fail to reject the null hypothesis!

- Vulnerability and Release Correlation
  - Vulnerabilities and Releases: -0.005
  - V&R most: 0.025
  - V&R least: -0.023

Not much correlation going on here, huh?

UNC GREENSBORO

Frequency of Commits Vs. Severity

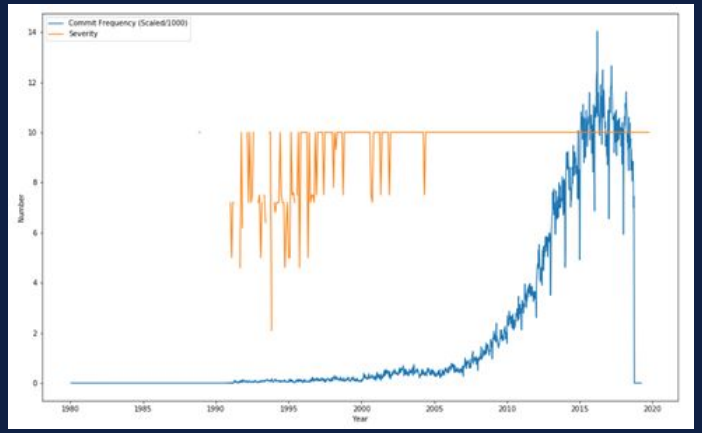
Comparing the Base Score (severity) of Vulnerabilities from NVD to the frequency of commits, there doesn't seem to be much of a link between the two parameters.

The top graph looks at the 1990s to 2020, although the discrepancy would be due to commits being more common as years go by. Also on this graph it seems like most base scores are in, but as we see the mean is around 6.

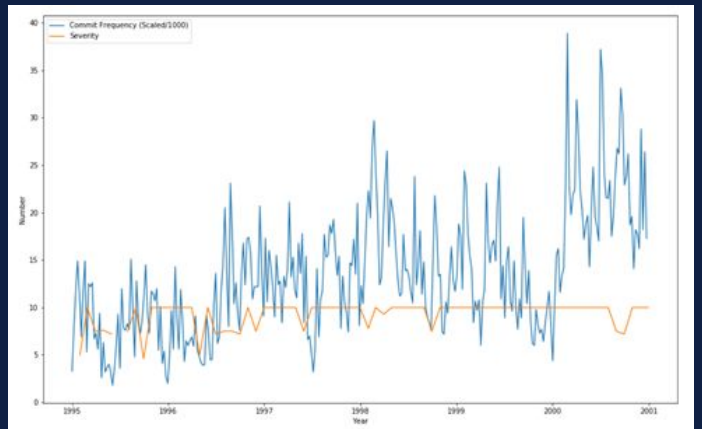
```
Overall mean of Base Score

6.1621878715815415
```

Looking at 1995 to 2000, an area with a lot more noticable differences in severity, the link between the two still doesn't seem feasible.



Commit Frequency (Scaled) vs. Severity



Severity(Weekly) vs. Frequency of Commits

UNC GREENSBORO

NVD JSON to CSV

More efficient algorithm to convert NVD Json to CSV

- (Simple) Converts with nested arrays
- (Detailed) Creates new rows that include the information from the nested arrays

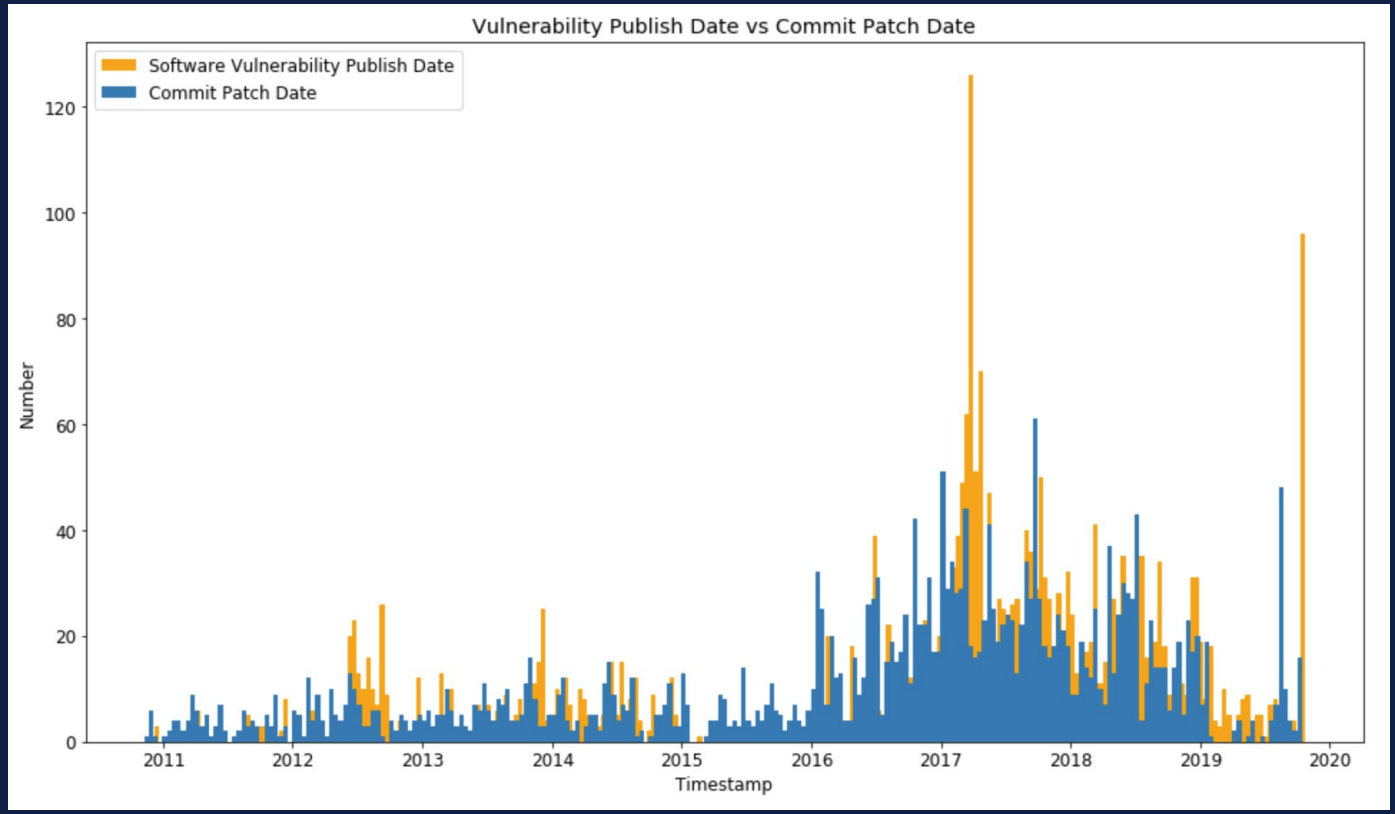
Vulnerability Publish Date vs. Commit Patch Date

Tasks

- More data wrangling
  - Extracting relevant data for statistical analysis
- Statistics of revisions and software vulnerabilities
- Distribution modeling
- Hypothesis testing

Vulnerability Publish Date vs. Commit Patch Date

Software vulnerabilities patched by GitHub commit



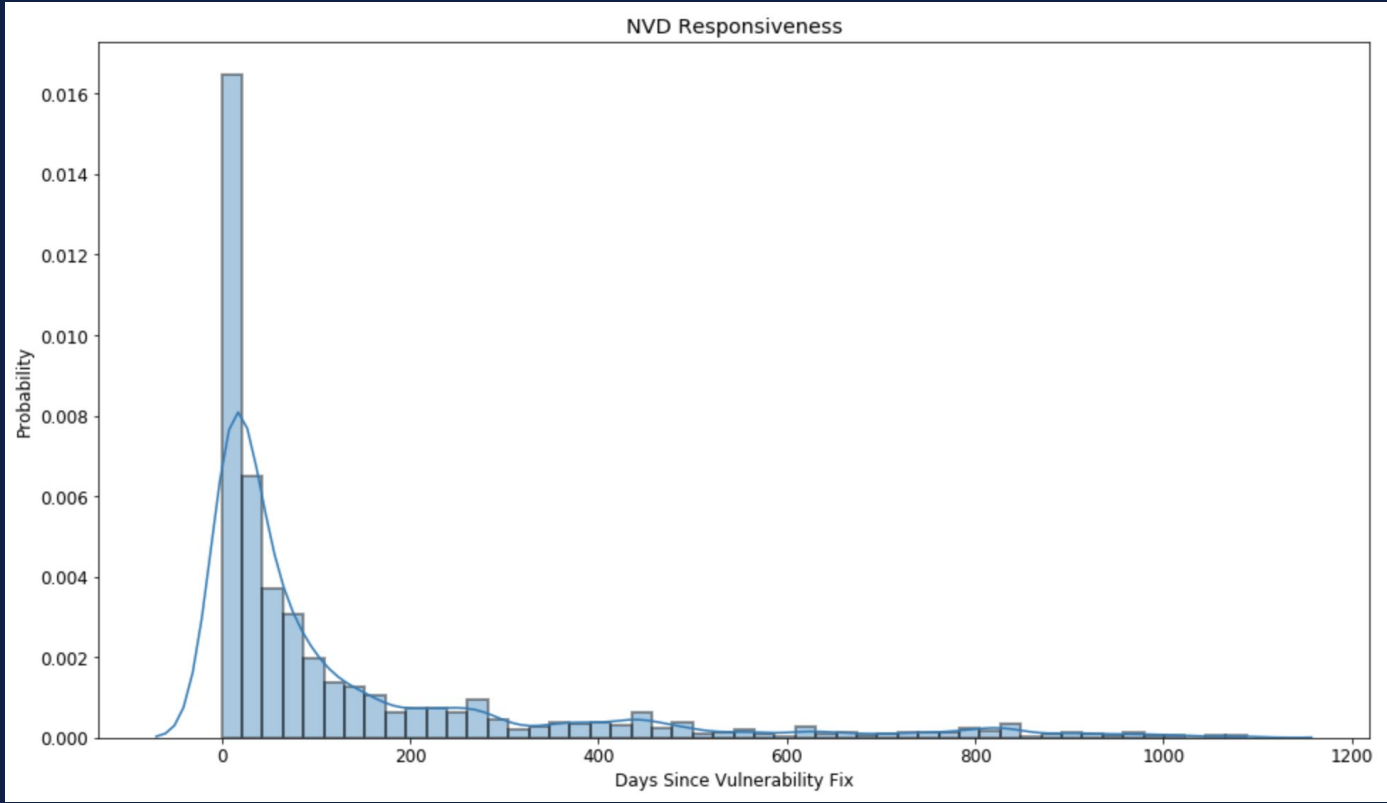
UNC GREENSBORO

Vulnerability Publish Date vs. Commit Patch Date

Publish date and commit patch date correlation:

- Daily: 0.226
- Weekly: 0.540
- Monthly: 0.701
- Yearly: 0.821

Vulnerability Publish Date vs. Commit Patch Date



UNC GREENSBORO

Vulnerability Publish Date vs. Commit Patch Date

H0: NVD will publish new software vulnerability info <= 4 months after there is a GitHub commit that fixes the vulnerability.

H1: NVD will publish new software vulnerability info > 4 months after there is a GitHub commit that fixes the vulnerability.

P-value = 7.17e-05 < 0.05 (significance level)

⇒ Reject H0

⇒ Avg time to submit new vulnerability info > 4 months…

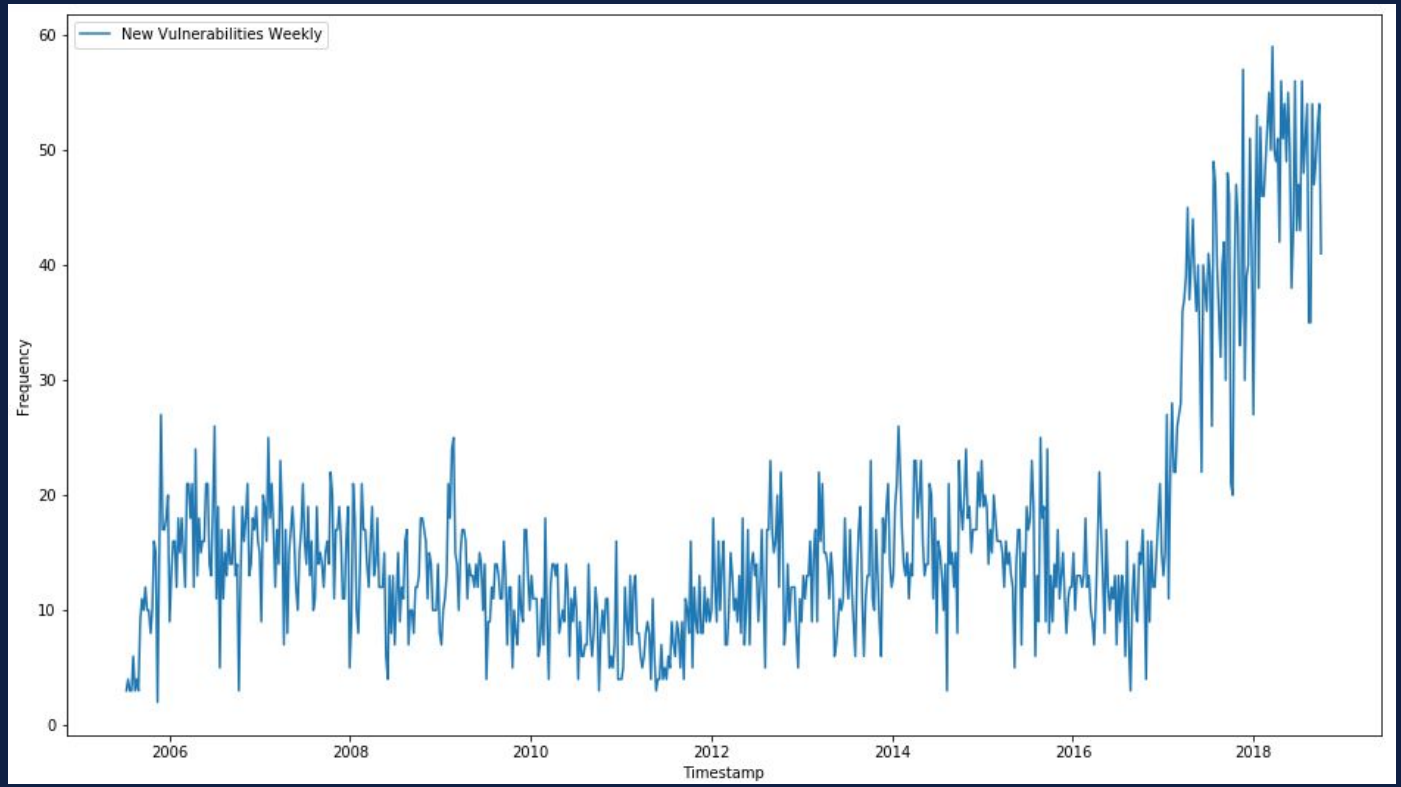
Releases vs. New Vulnerabilities

Task:

- Look into New Vulnerabilities Statistics
- Specifically Vulnerabilities per week

New Vulnerabilities (Per Week)

Large Increase in Vulnerabilities in recent years



UNC GREENSBORO

New Vulnerabilities (Per Week)

Obviously data is skewed to the right

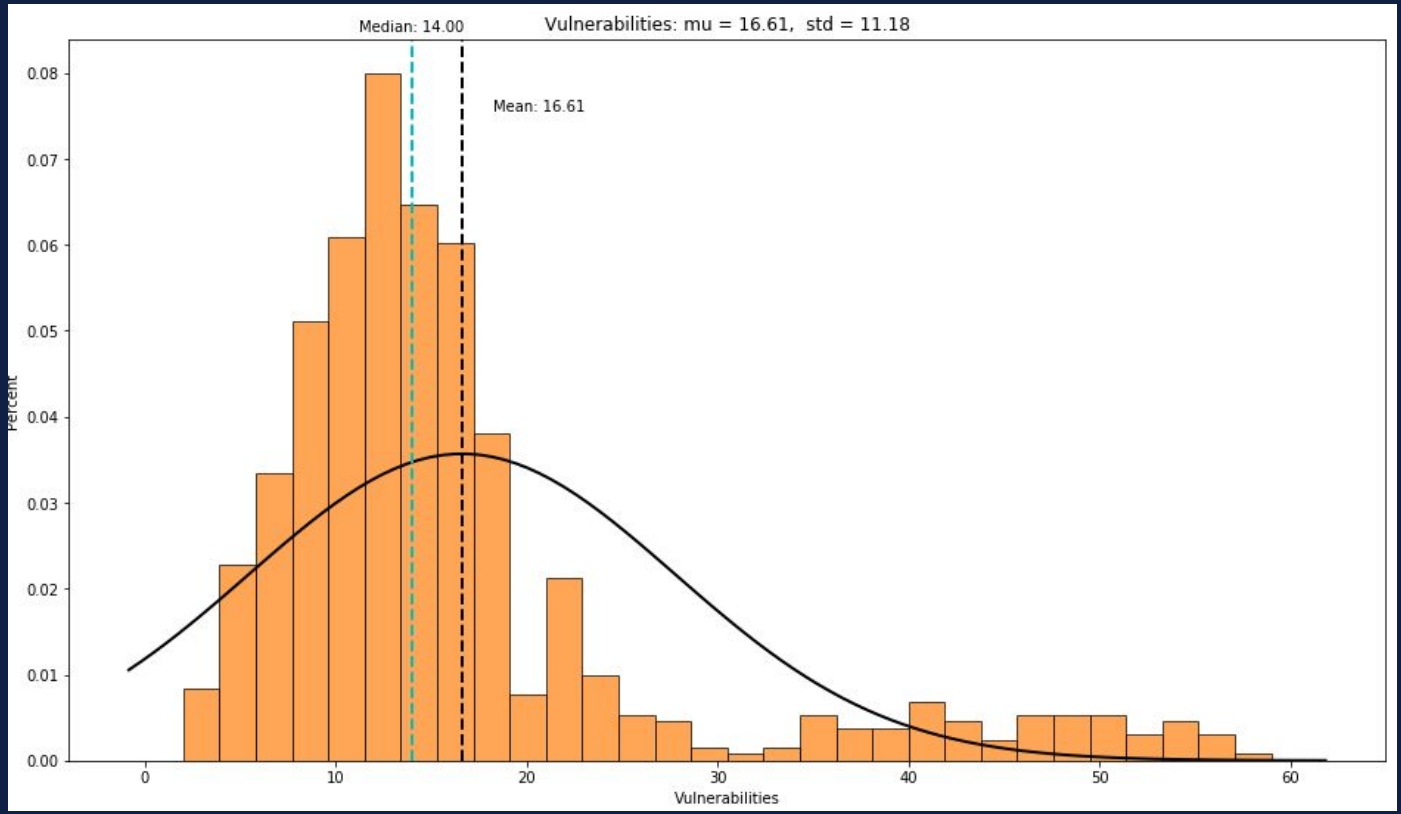
Mean: 16.61

Median: 14.00

Std: 11.18

CV: 0.67

Variance: 125.13



UNC GREENSBORO

New Vulnerabilities (Per Week)

Square Root of previous histogram results in a more normal distribution.

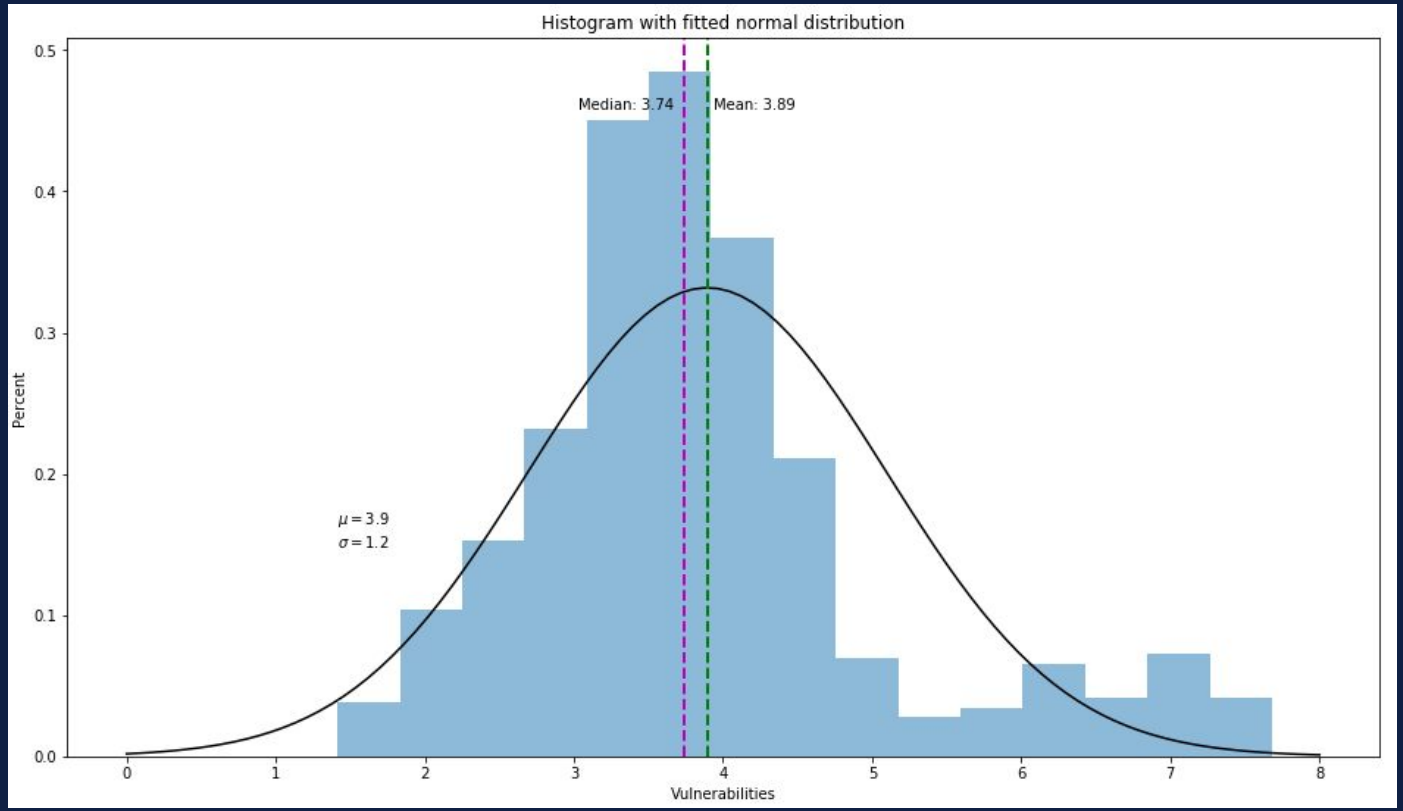
Mean: 3.89

Median: 3.74

Std: 1.2

CV: 0.31

Variance: 1.45

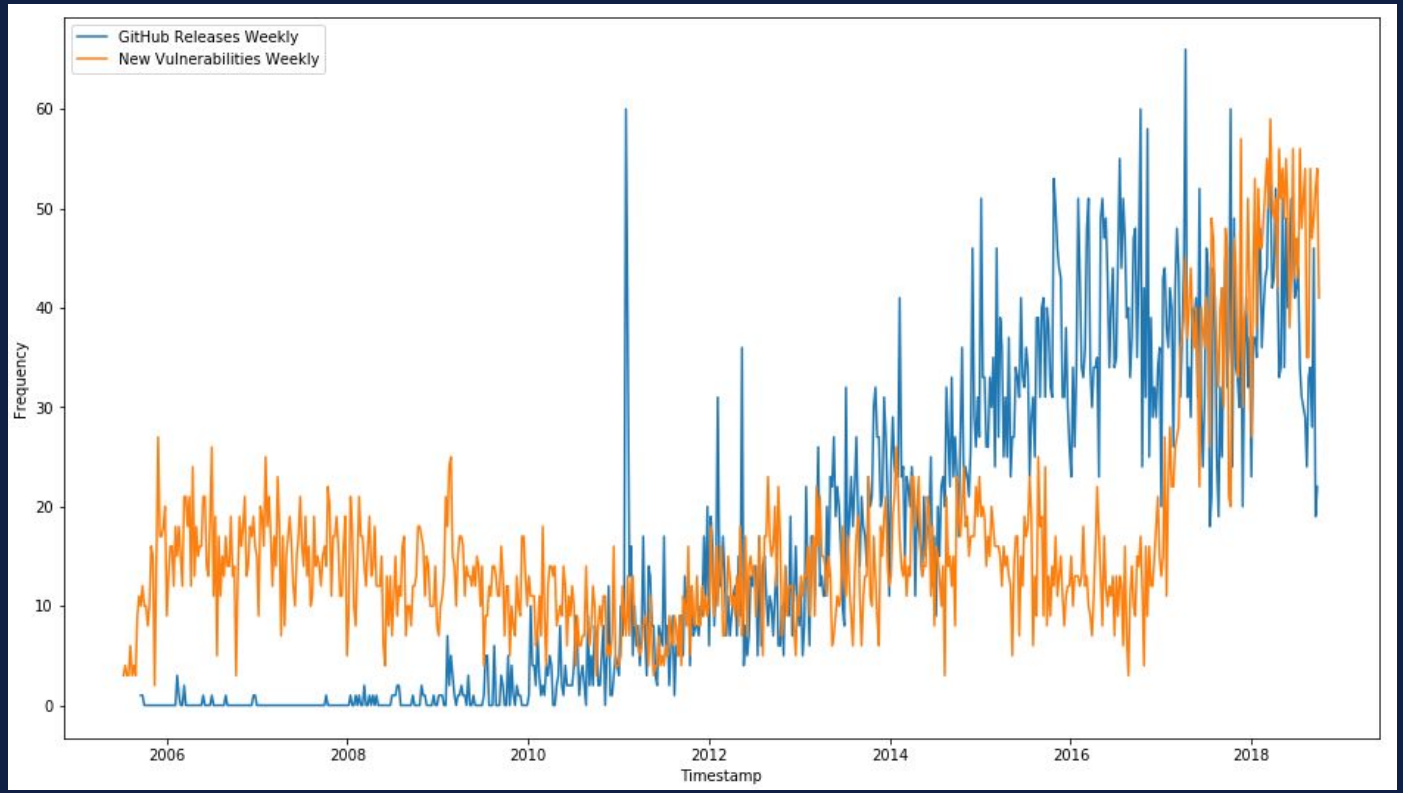


UNC GREENSBORO

Releases vs. New Vulnerabilities

Overlay of Vulnerabilities and Releases Per Week

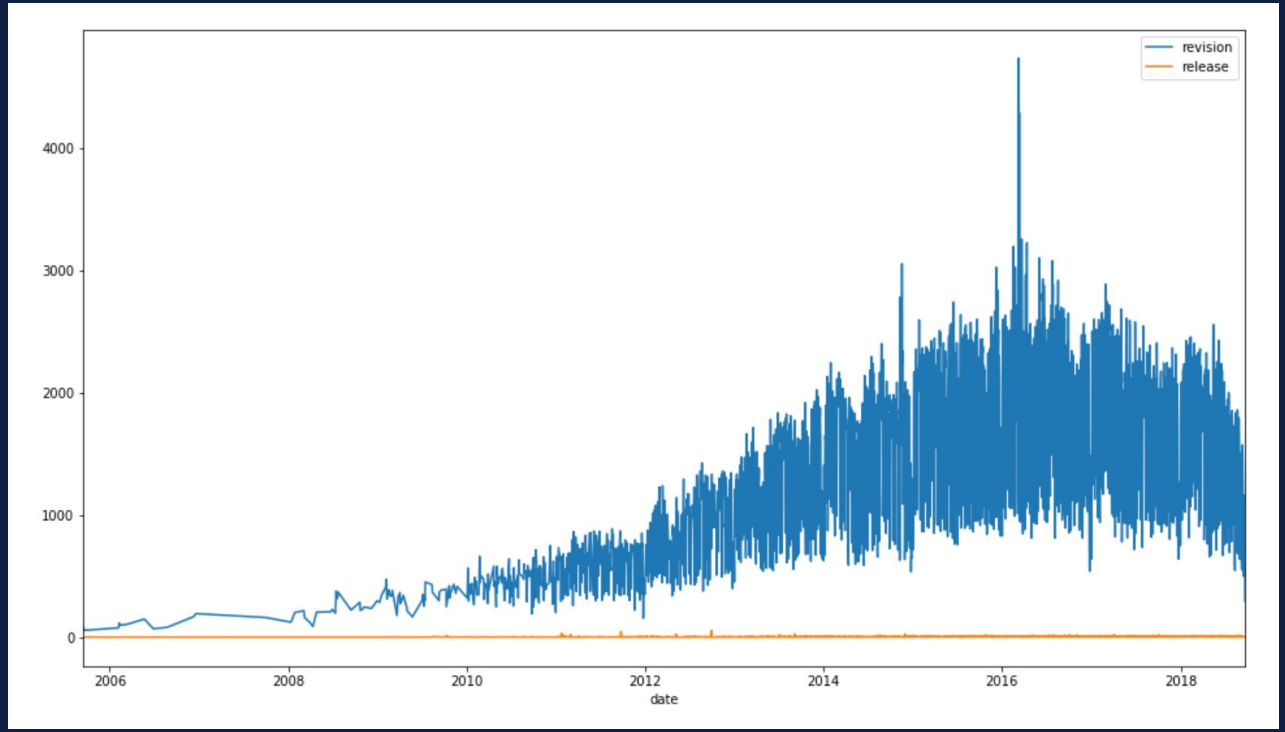
Correlation: 0.46



UNC GREENSBORO

Analysis between revisions (commits)and releases in general

A plot between Commits per-day And releases per Day over the years.



UNC GREENSBORO

Analysis Between Commits (Vul. Fixes) and Releases dates.

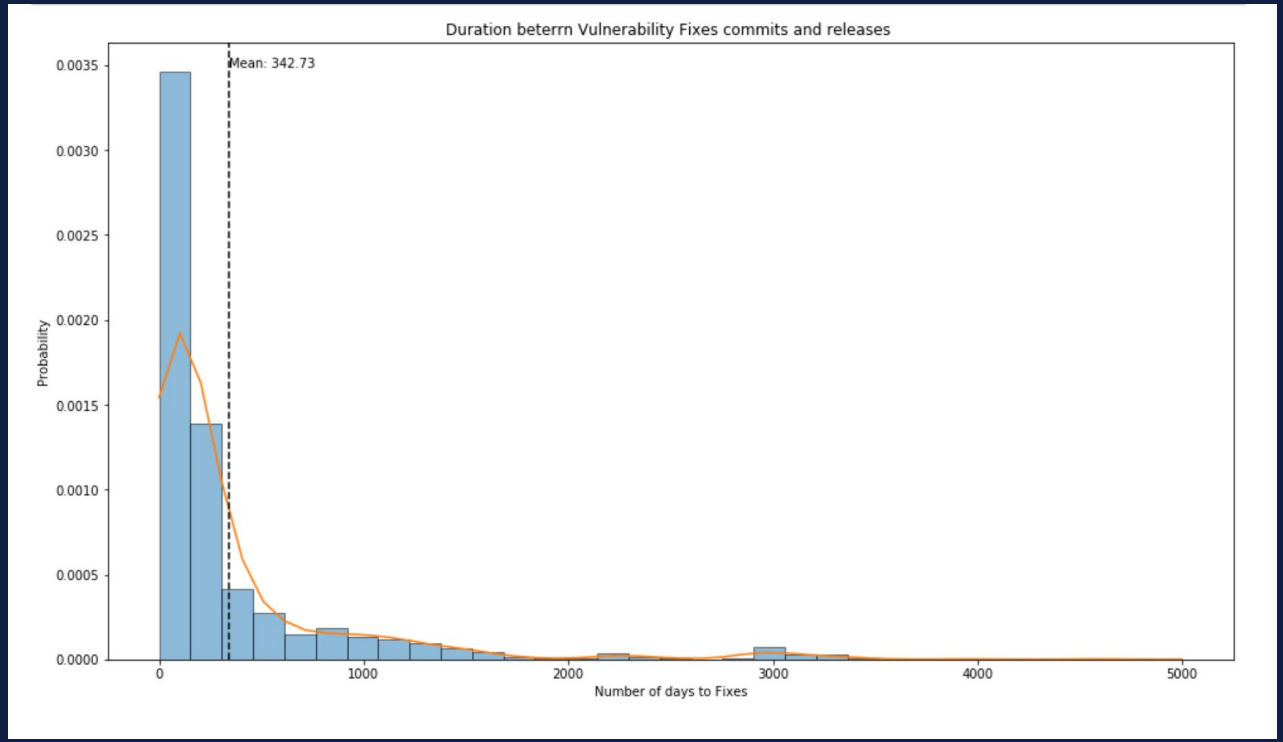
Applied Normal

Normal Dist.

Observed

Mean: 342.73

SD: 623.8



UNC GREENSBORO

Questions?