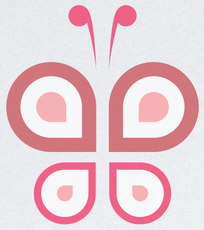


Soteria

security made **usable**

i290T Open Data

Ashley Evie Shreyas Victor

OBJECTIVE

Soteria

To provide **usable analysis** of **security data** to the

open community

DATA

Soteria

- Source: VERIS Community DB
- 3,084 breach incidents
- 321 breach features (cause &effect)
  - **attack.vectors** (270+ booleans, hacking, malware, ...)
  - **attack.effects** (victim.count, victim.revenueloss, ...)

Soteria

DATA CLEANING

- **Handle Missing Values** : NaN, Unknown
- **Recode Variables**
  - Eg: **timeline.discovery -> discovery.daycount**
- **Group Variables** (Using **CENSUS** API)
  - Eg: **Victim.IndustryName -> Victim.Industry.Category**

DATA EXPLORATION

Soteria

(aggregate statistics)

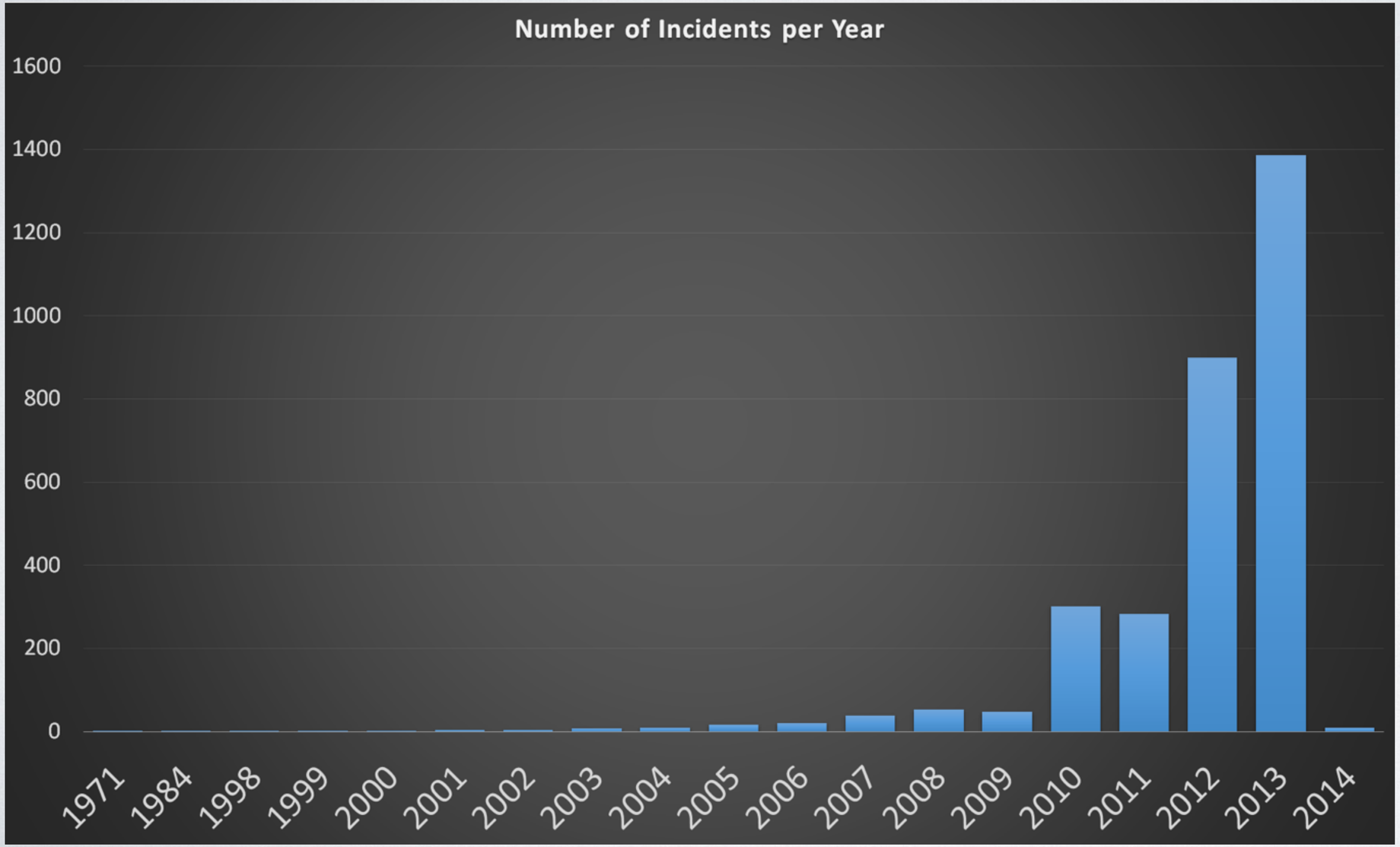
(one-variable-at-a-time)

(2-variables-at-a-time)

(all-key-variables-at-a-time)

AGGREGATE STATISTICS

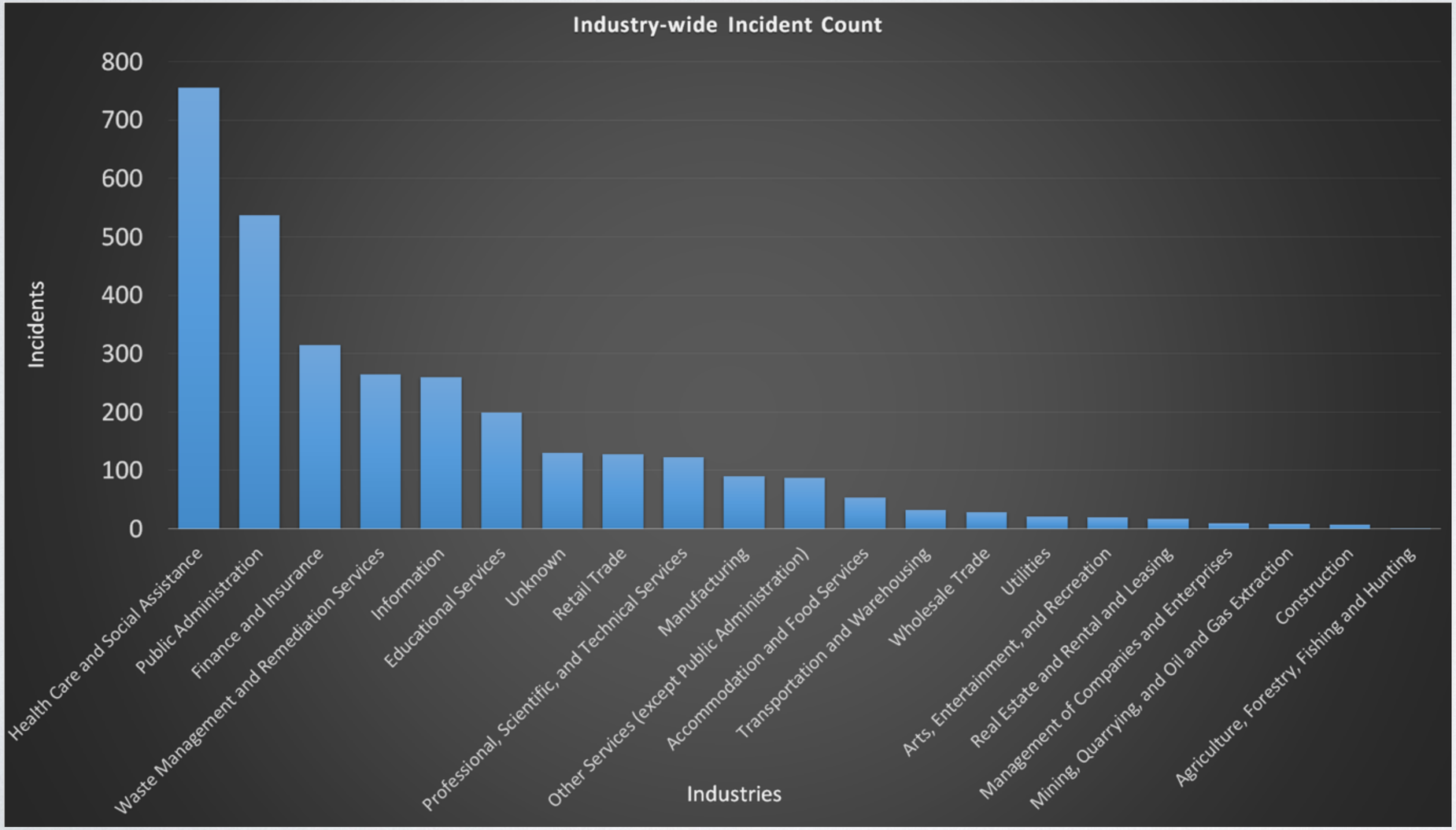
Soteria



incidents per **year**

AGGREGATE STATISTICS

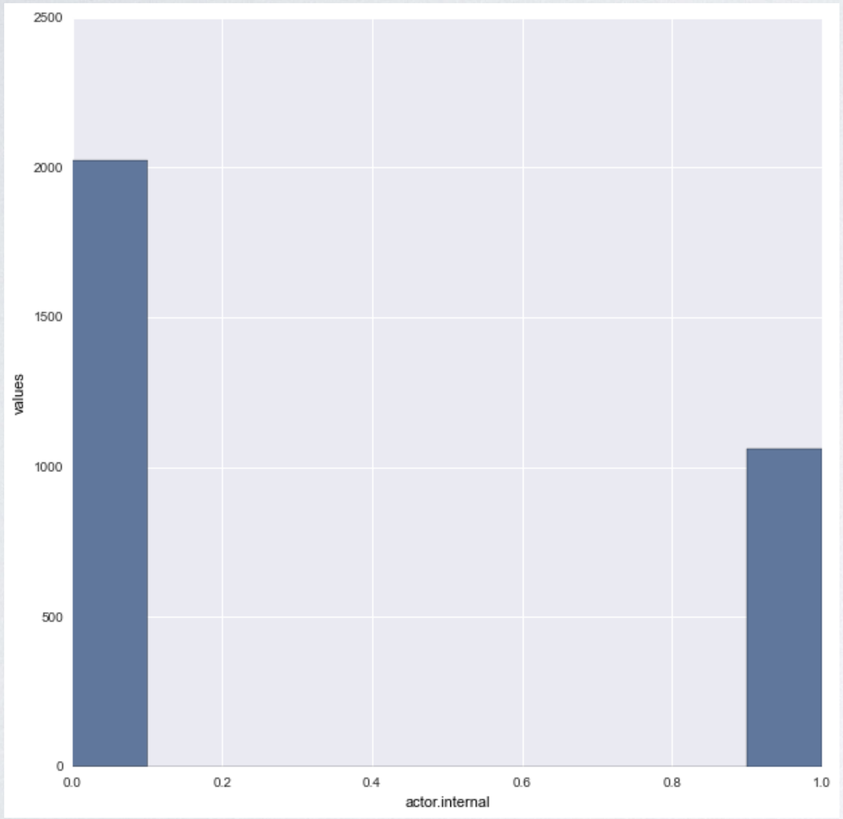
Soteria



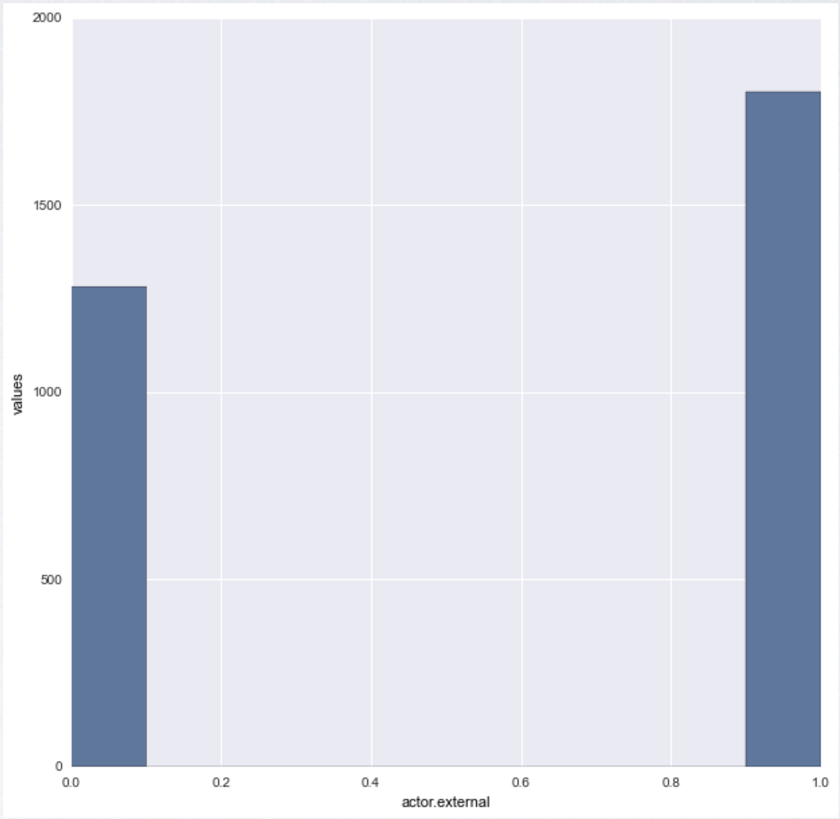
incidents per **industry**

ONE-VARIABLE

Soteria



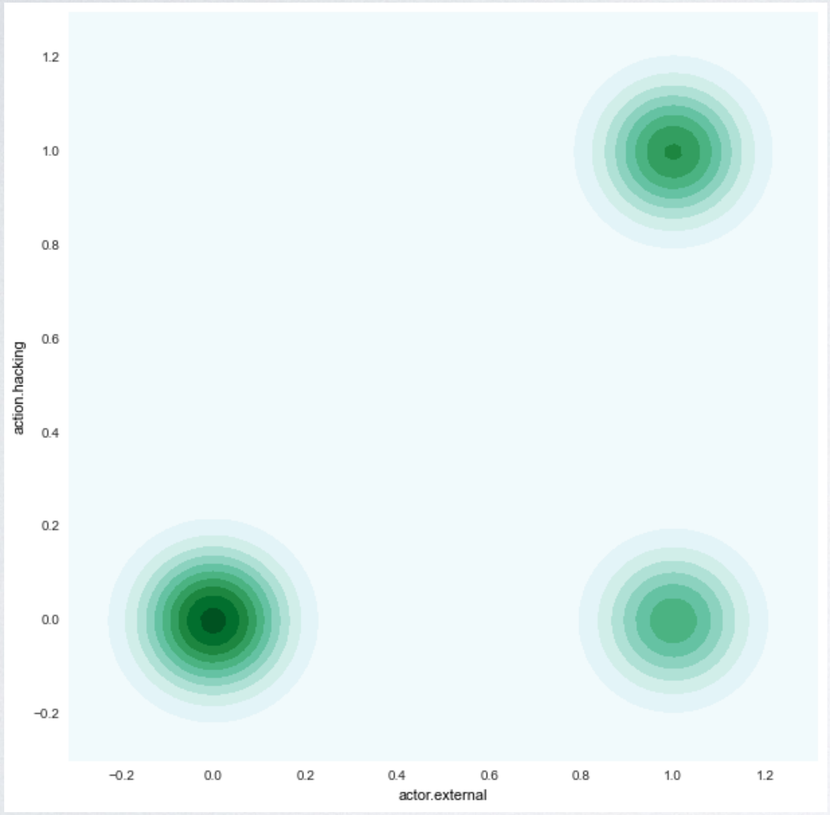
actor.internal



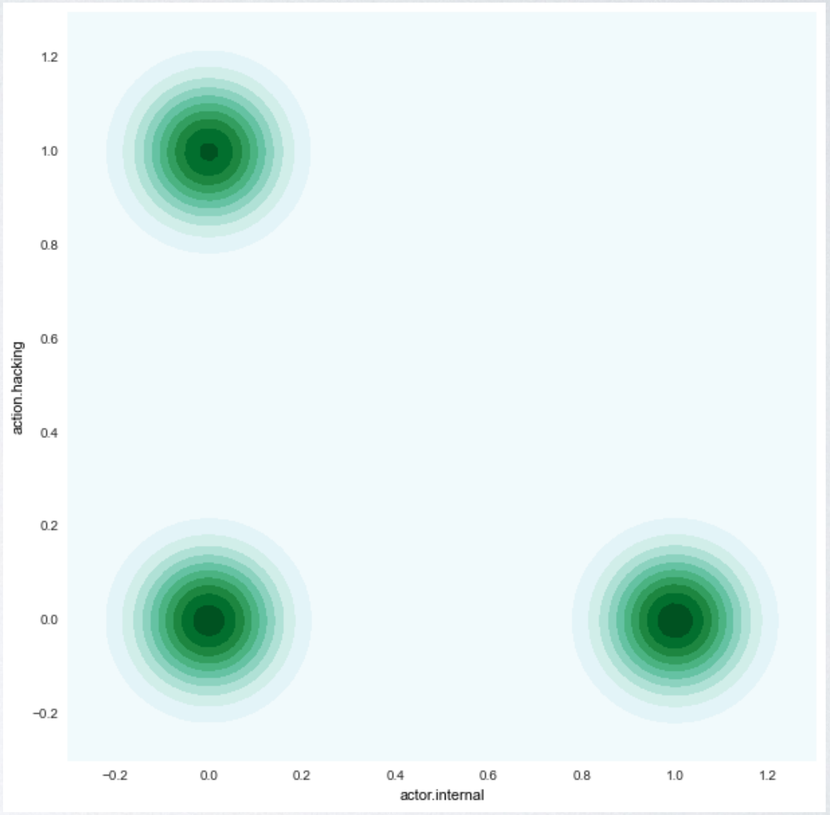
actor.external

TWO-VARIABLES

Soteria



actor.hacking



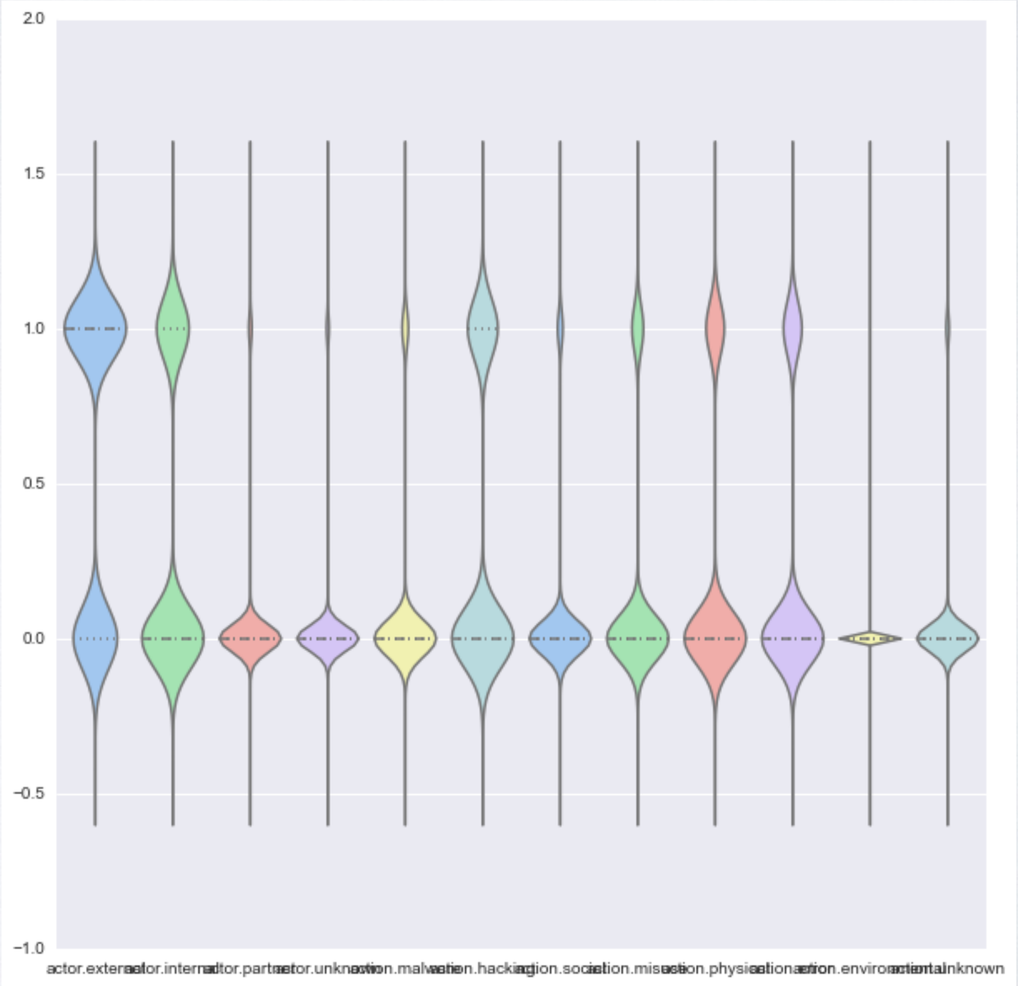
actor.internal

actor.external

KEY-VARIABLES

Soteria

ViolinPlot



KEY-VARIABLES

Soteria

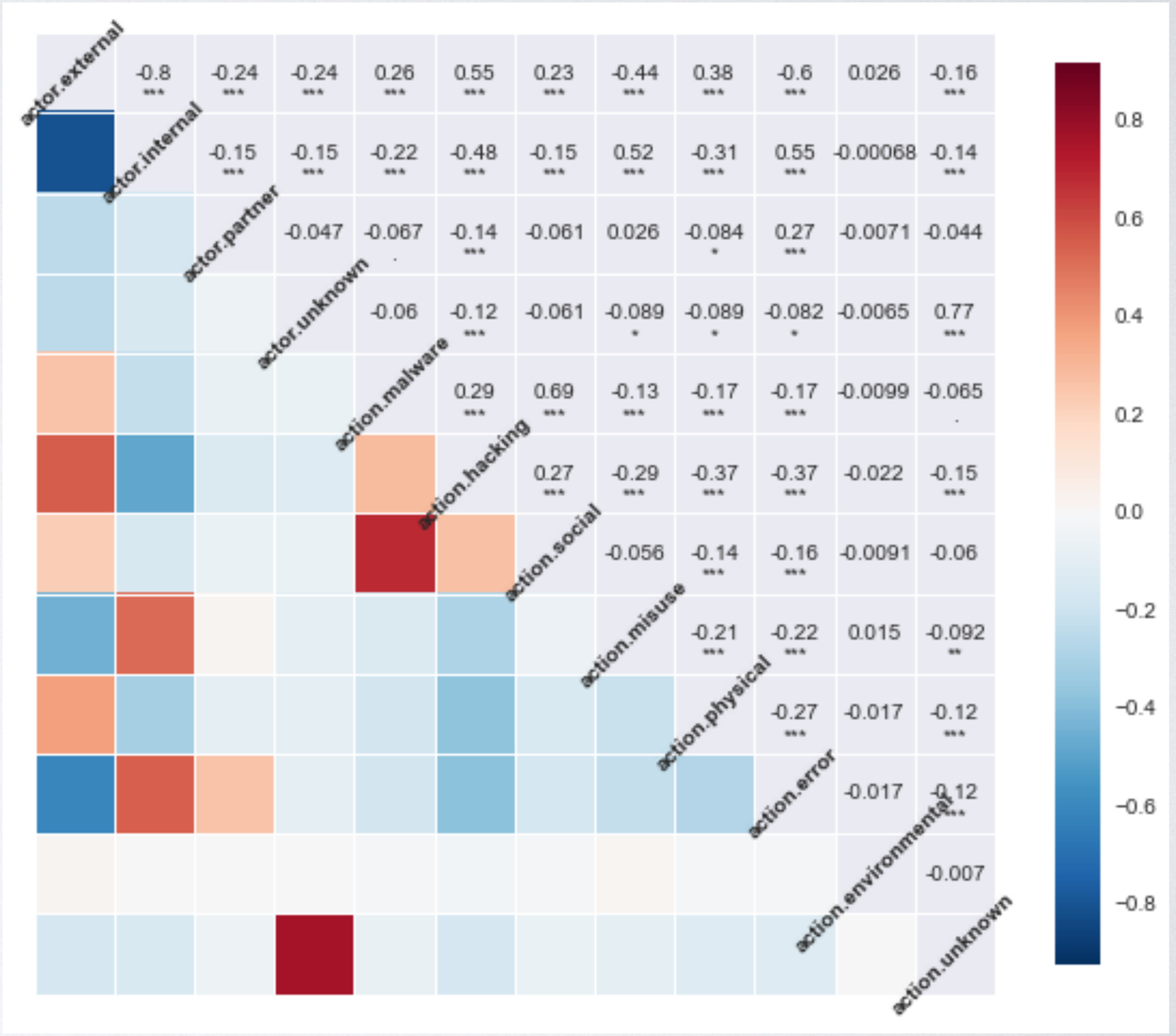
Correlations

(not-so-relevant)

actor.external actor.internal

actor.unknown action.unknown

actor.external action.hacking



KEY-VARIABLES

Soteria

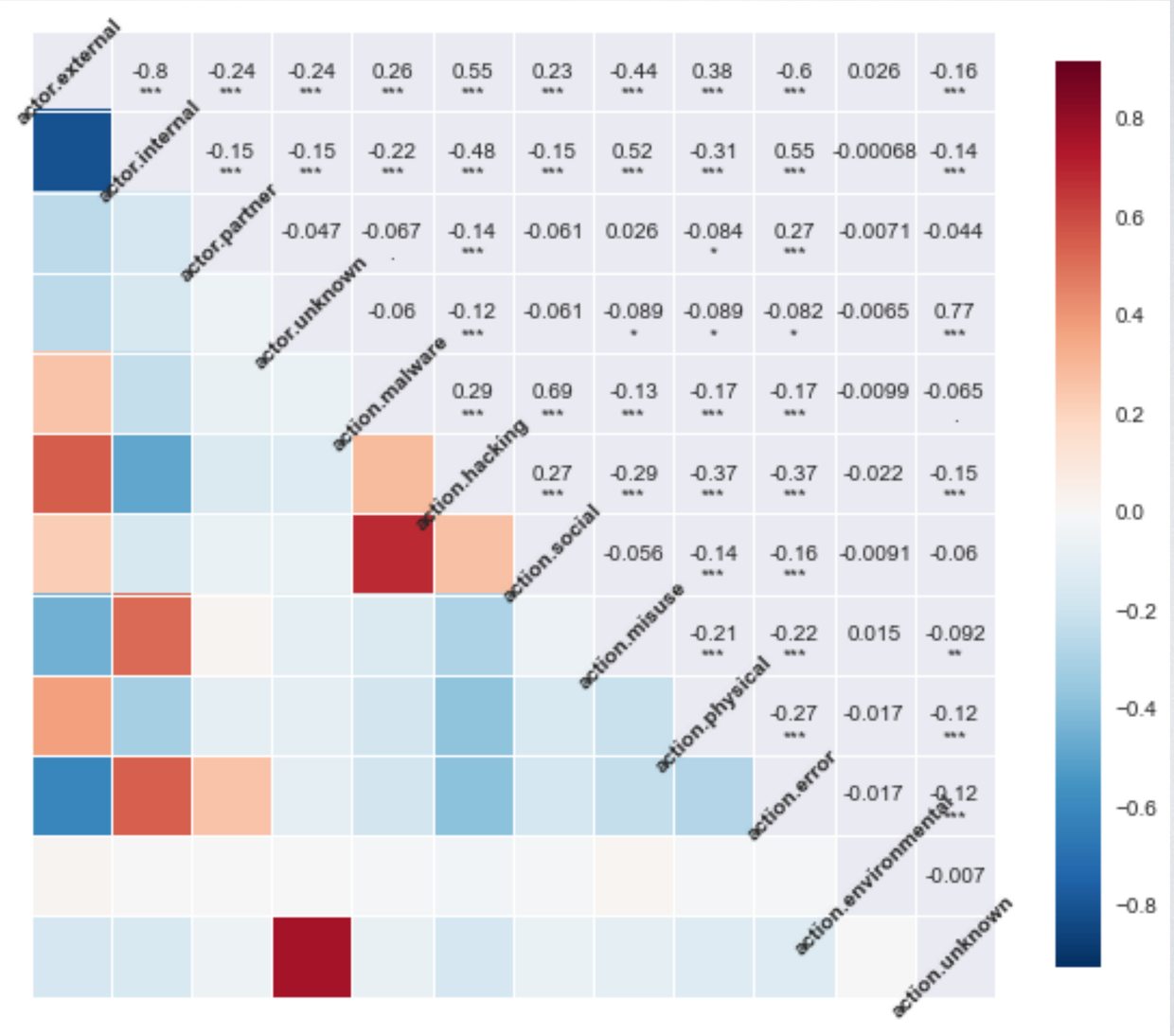
Correlations

(relevant)

action.malware action.social action.hacking

actor.internal action.misuse

actor.internal action.error



DATA SCIENCE

Soteria

- **cause** (booleans) & **effect** (victim) variables
- usable data exploration through filtering
- find pairs of similar incidents

USABLE FILTERING

Soteria

front-end demo

SIMILAR INCIDENTS

Soteria

attack_similarity.py

(MapReduce)

CONCLUSIONS (PRELIMINARY)

Soteria

- a **lot** of similarity
- Our inferences:
  - attacks are **reused**
  - breach **vocabulary** is still not **deep enough**

QUESTIONS ...