Soteria

security made **usable**

i290T Open Data

Ashley Evie Shreyas Victor

OBJECTIVE

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

Soteria

DATA

- 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**

Soteria

DATA EXPLORATION

(aggregate statistics)

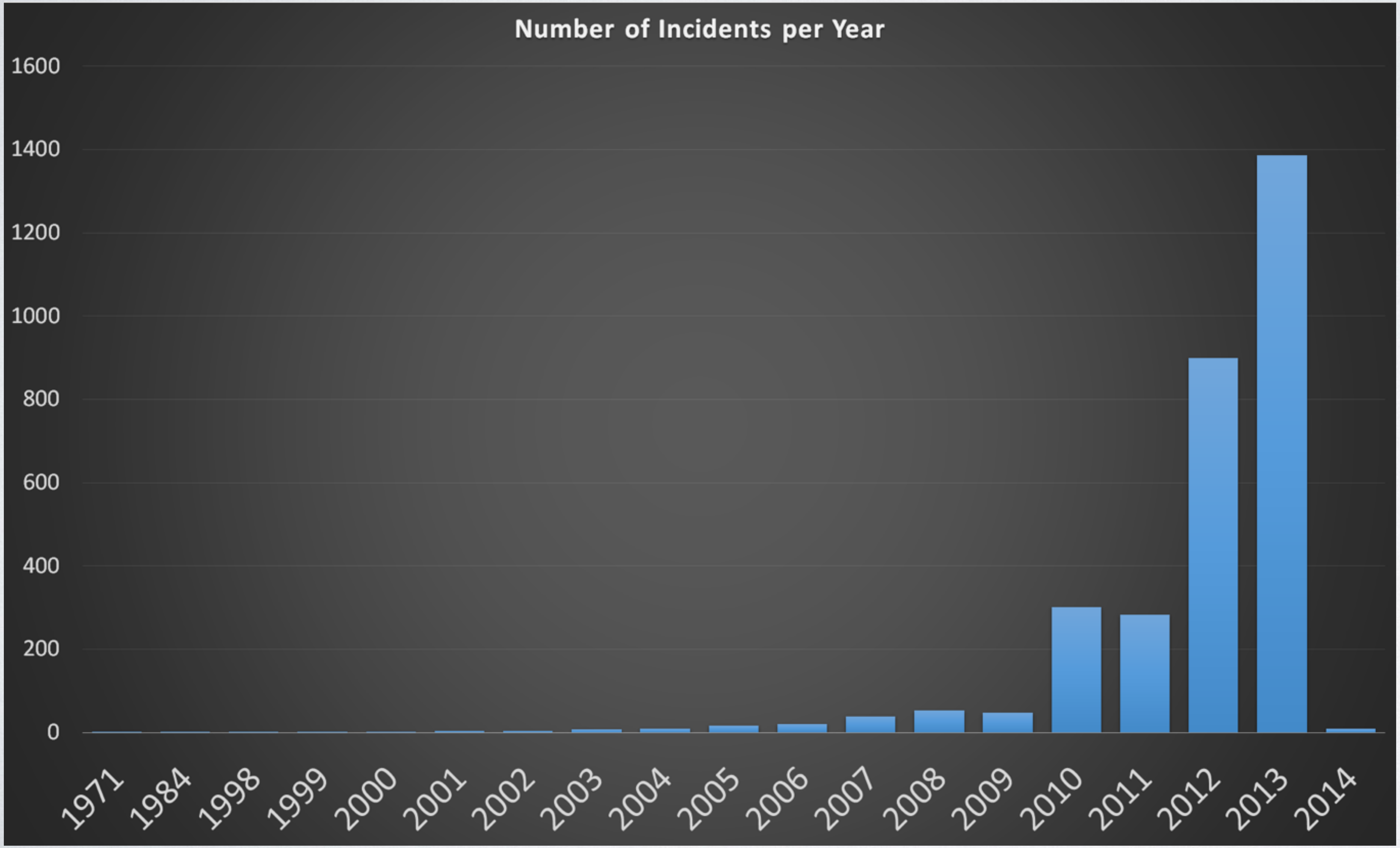
(one-variable-at-a-time)

(2-variables-at-a-time)

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

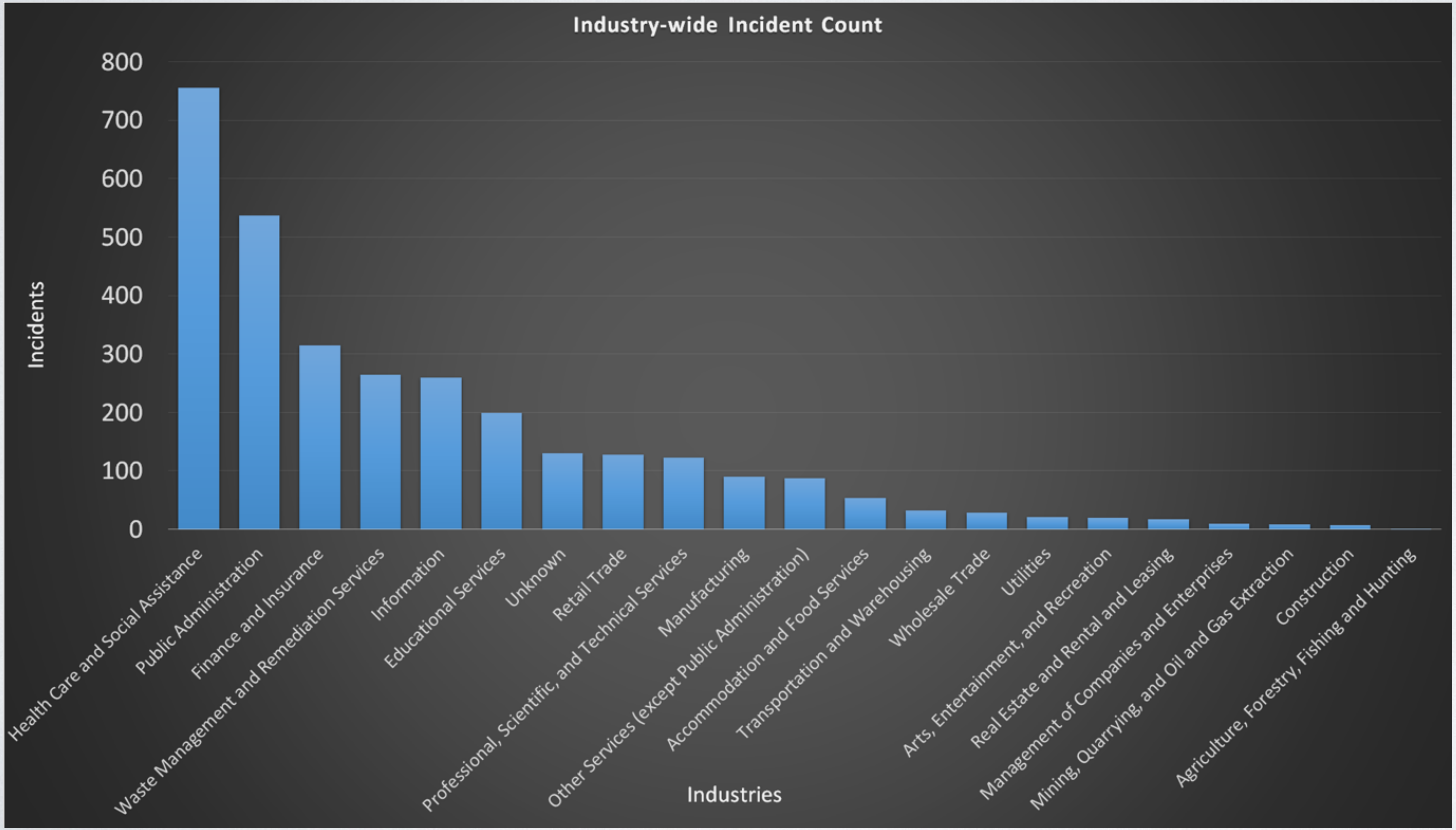
Soteria

AGGREGATE STATISTICS



incidents
per
year

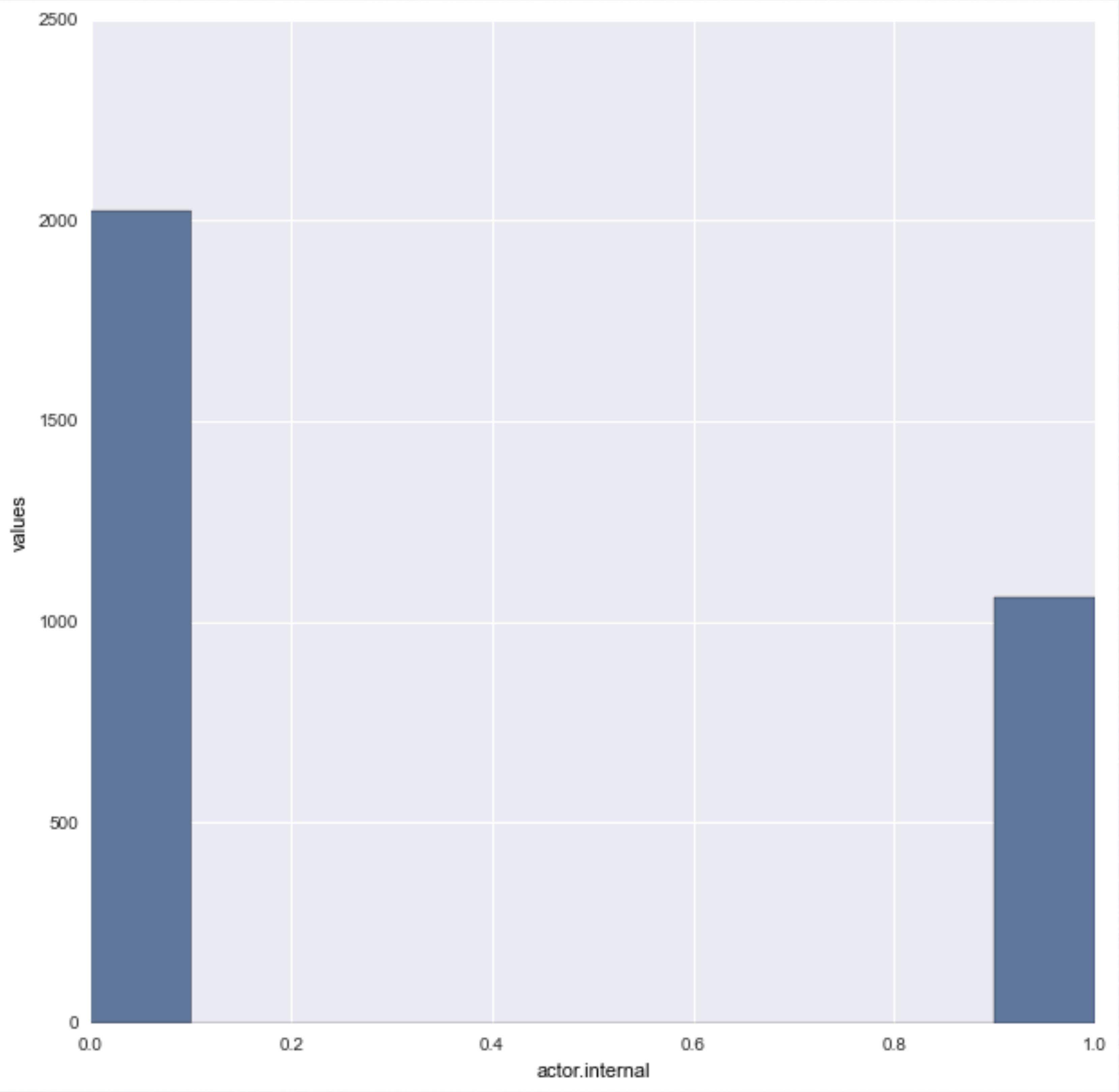
AGGREGATE STATISTICS

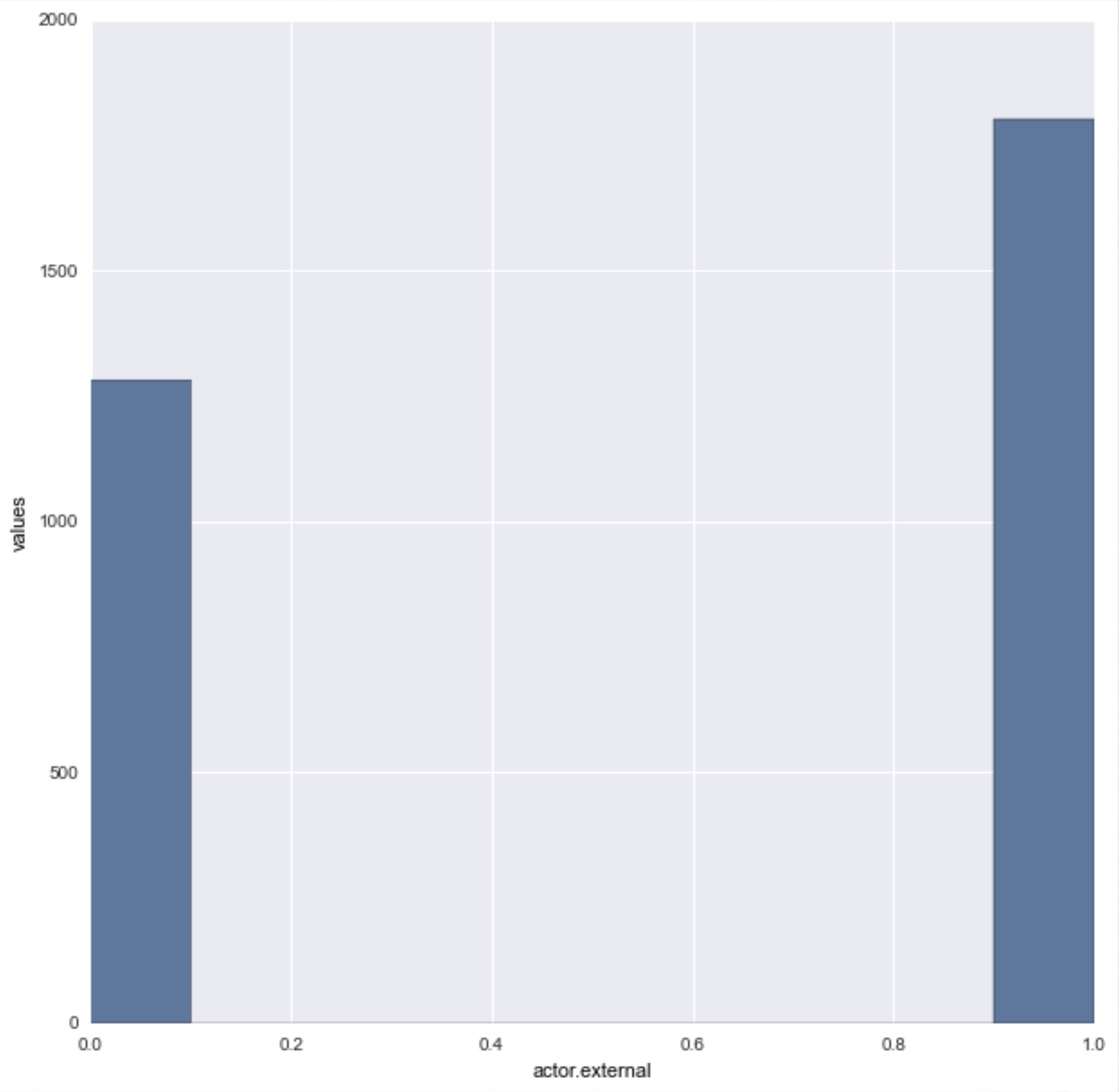


incidents
per
industry

Soteria

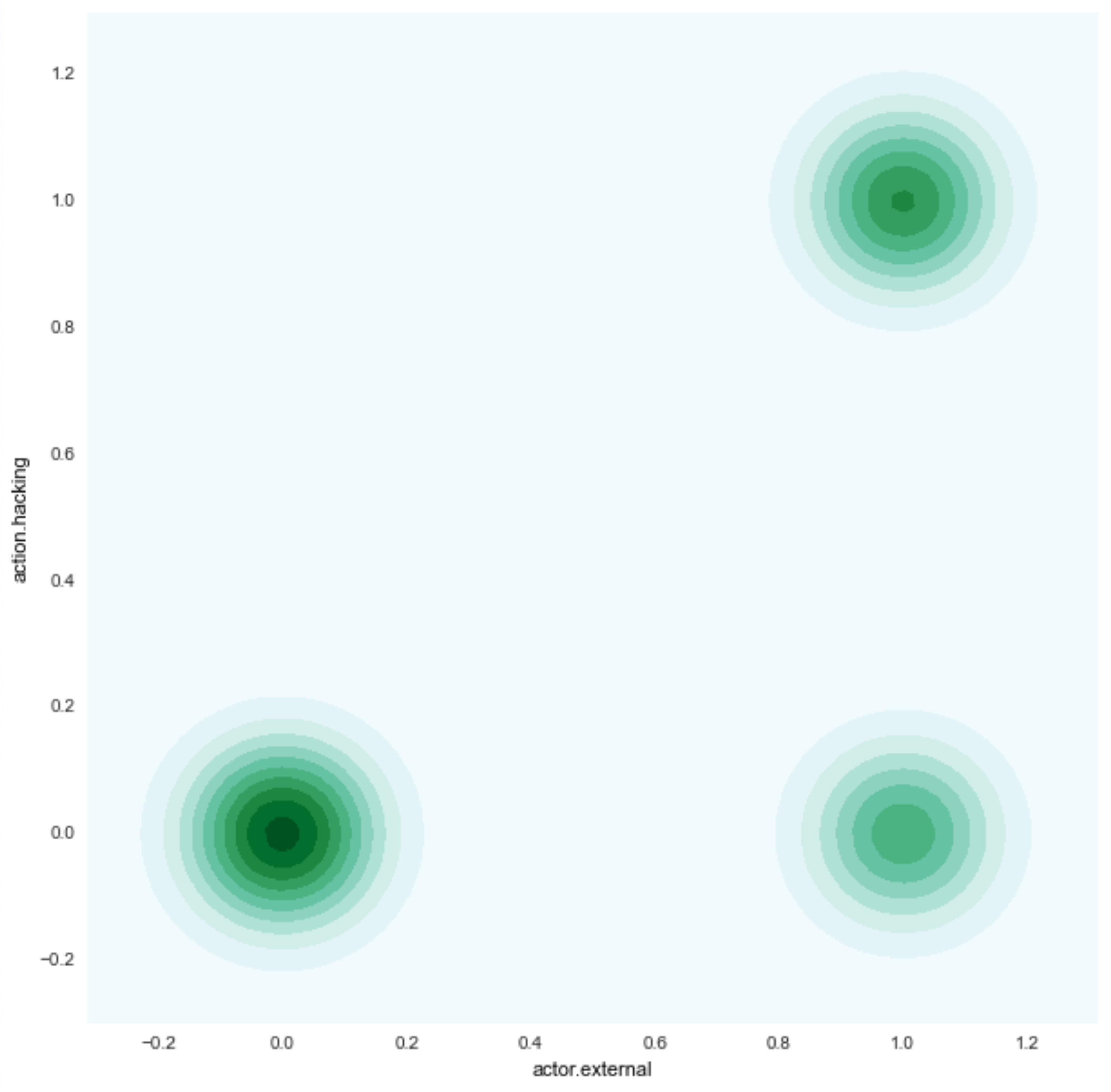
ONE-VARIABLE

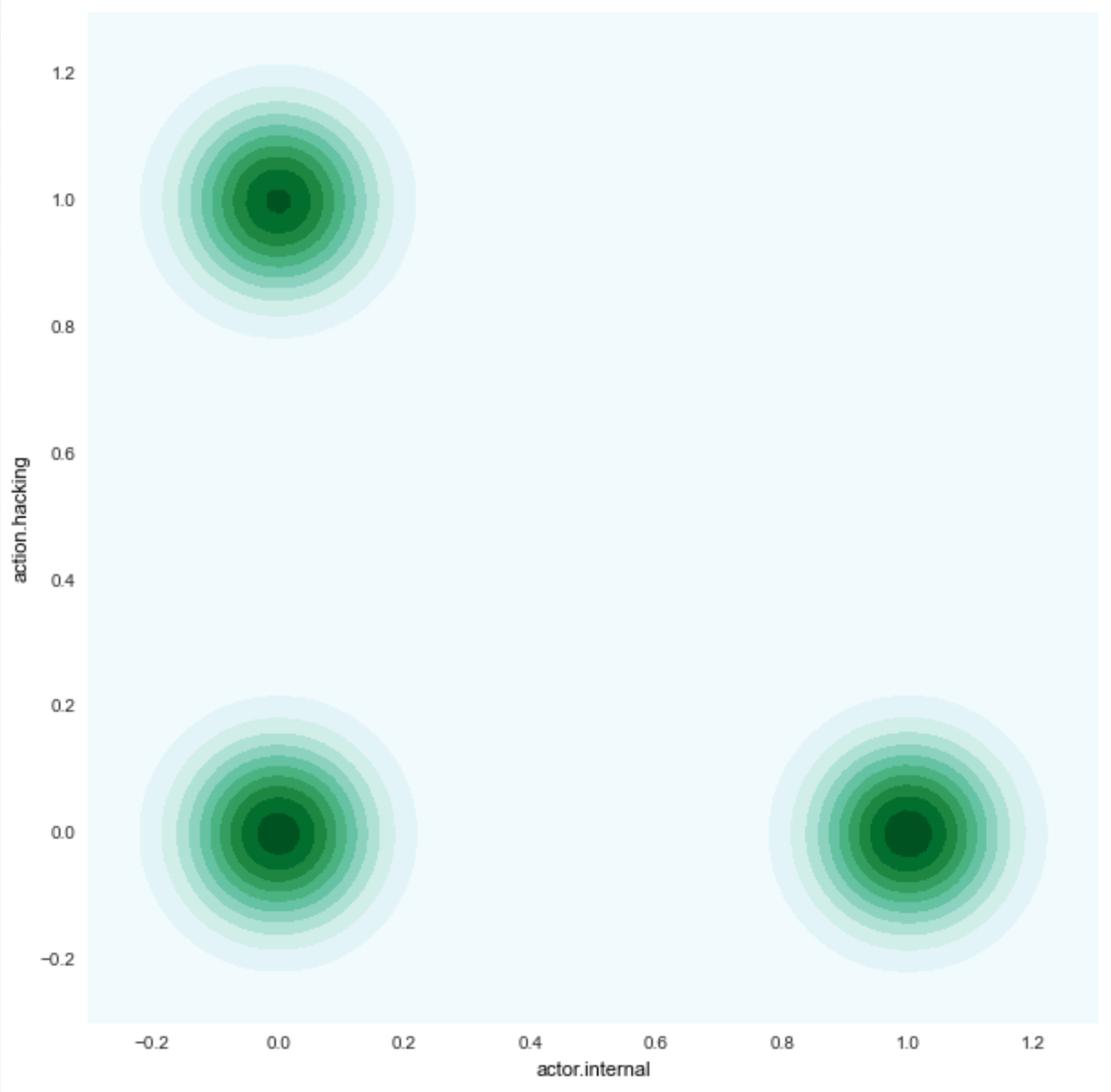




Soteria

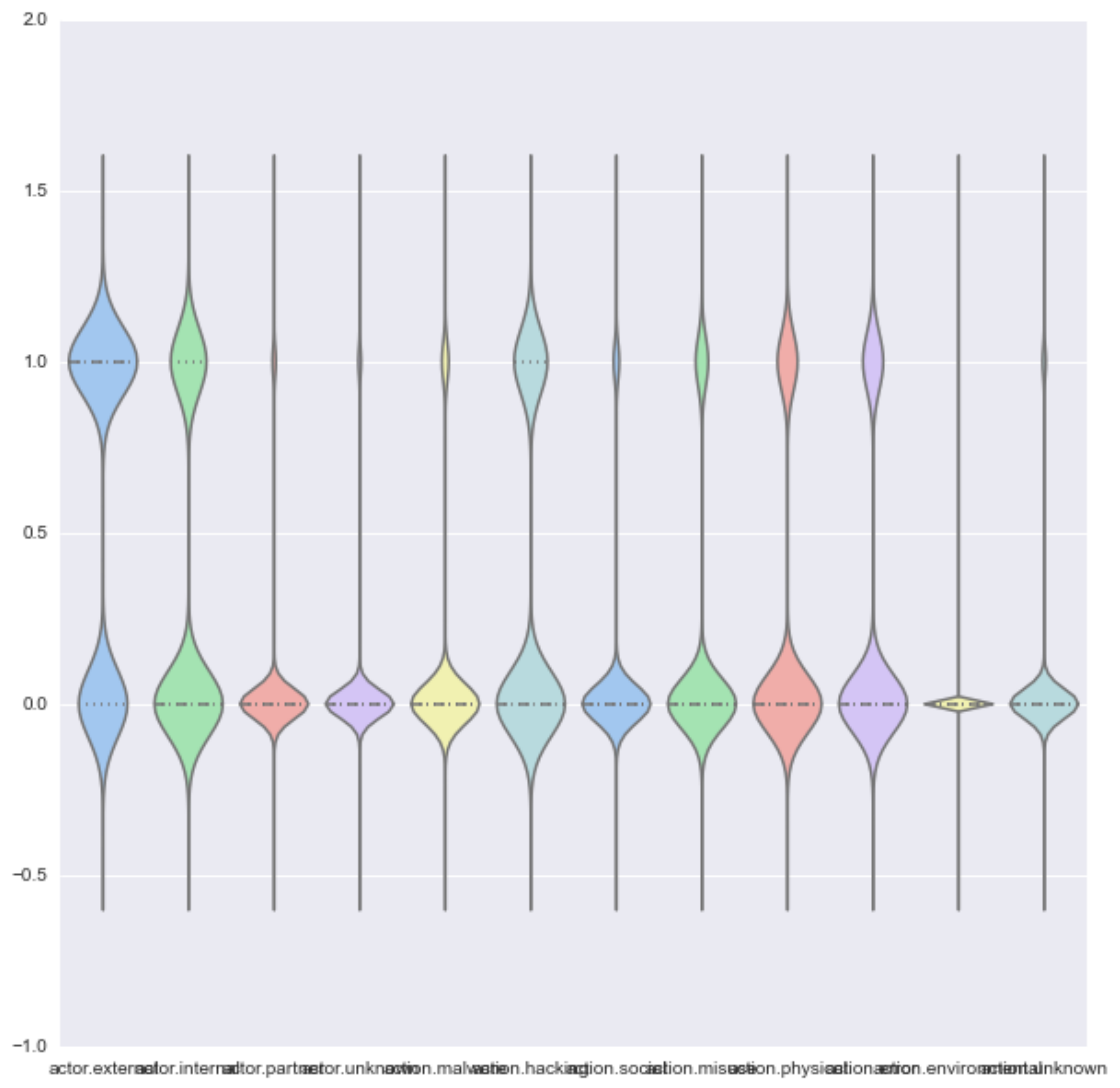
TWO-VARIABLES





KEY-VARIABLES

ViolinPlot



DATA SCIENCE

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

USABLE FILTERING

front-end demo

Soteria

SIMILAR INCIDENTS

attack_similarity.py

(MapReduce)

Soteria

CONCLUSIONS (PRELIMINARY)

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

QUESTIONS ...

Soteria