

4 February 2023

Review the sample projects in ISLR and suggest data project examples using your own experience and data (e.g. cancer gene identification, spam classification, income etc.)

In 2019, several local governments across the United States experienced a cyber-attack using some variation of Ransomware. While the attacks appeared largely at random, a question remains if certain organizational factors of local government targets influenced the selection by the Threat Actors.

a. Present a proposed hypothesis:

Overall Question: Does budget or personnel allocation influence a Threat Actor's cyber-attack target selection decision?

Overall Hypothesis: Budget and personnel allocation do play a role in influencing a Threat Actor's cyber-attack target selection decision.

b. How data can be collected?

Cyber-attack data: Seculore.com actively collects data on all publicly reported cyber-attacks in the United States. There store is information in an easily accessible database on their website. The categorize the information into different target designations, one of which is "government." A list of all local governments nation-wide that experienced a Ransomware attack in 2019 can be collected.

Local Govt Data: Many local governments post their budget and personnel data, online their organization website. A complete annual budget will have total government budget, individual department budgets, and personnel assignments and allocation for each department.

c. What methods could be considered?

Data Collection: SecurLore Solutions data may be able to be data mined, but most government budget data is included in PDF forms and will require downloads and actual data collection by hand.

Data Consolidation: The collected data can then be consolidated to a single form.

Data Analysis: The data table can then be imported into R-suite or STATA for analysis.

d. How to start the data project?

By the formulation of the question and hypothesis and developing a plan such as this, is the start of the data project. Next, would involve the initial collection of the rough data to be sorted through and then inputted into a single, data table for eventual importing and analysis.