Option B:

Take any combination of the datasets listed on http://csr.lanl.gov/data/cyber1/ and tell us something interesting about the data.

For example, you could build a network graph out of the systems on the network, or build a predictive model to prevent red team attacks. For this option we are looking for you to tell us something valuable out of the datasets. You may use any languages, tools, and frameworks that you like and be prepared to give a small presentation on your findings.

**Dataset:** auth.txt.gz

This data represents authentication events collected from individual Windows-based desktop computers, servers, and Active Directory servers. Each event is on a separate line in the form of "time, source user@domain, destination user@domain,source computer,destination computer, authentication type, logon type, authentication orientation, success/failure" and represents an authentication event at the given time. The values are comma delimited and any fields that do not have a valid value are represented as a question mark ('?').

Authentication is the process of determining whether someone or something is, in fact, who or what it is declared to be. Logically, authentication precedes [authorization](http://searchsoftwarequality.techtarget.com/definition/authorization) (although they may often seem to be combined). The two terms are often used synonymously but they are two different processes.

Steps:

Data Collection

Data Analysis

Raw Data

Data Cleaning

Tidy Data

Panda/

Tableau

Auth.txt

CSV

MySqL

Pandas

csr.lanl.gov

Critical Log Review

https://zeltser.com/security-incident-log-review-checklist/