**Project Details**

1. Datasets and the related documents are in this folder

<https://georgetown.box.com/s/g2qm0wfqugkjljd1o0bq7dk2rusnlgyn>

1. A small description about the data and the problem which is the only information given for this data. You can use the information however you want to produce a great meaningful project output.

This is part of a cyber risk project for the NSA. Their task is to help the NSA develop new methods for prioritizing cyber risk among Defense Industrial Base companies. They are focusing on contract/supply chain risk as the key dimension of overall cyber risk - essentially hoping to build a model of how likely a defense company is to be targeted for cyber espionage based on their products and subcontractors.

The key database for analyzing & quantifying contract risk is the [USASpending](https://www.usaspending.gov/explorer/agency) website, which compiles comprehensive government award data from across federal agencies. We believe this data - properly understood - can meaningfully improve the security of America's industrial base.

Because we are trying to map out supply chains based on text data we think that NLP and network analysis would likely be the most relevant approaches - but we're happy to hear alternate suggestions!

For a bit more context, in essence we are trying to use the government's [list of critical technologies](https://www.whitehouse.gov/wp-content/uploads/2022/02/02-2022-Critical-and-Emerging-Technologies-List-Update.pdf) to identify important subcontractors within the defense industrial base who supply those products or inputs to them. From an adversarial perspective, the most important subcontractors to find are ones that are either pervasive (appearing in many critical supply chains) or provide a unique product (and are therefore a single point of failure).

**I've attached a spreadsheet which exemplifies the data we're trying to analyze.** This comes from [a contract for hypersonic missile development](https://www.usaspending.gov/award/CONT_AWD_HR001117C0025_9700_-NONE-_-NONE-) with a number of subcontractors (listed in the attachment). We are especially interested in the columns to the far right which describe the subcontractors' company info, geographical presence, and description of services rendered. There are hundreds, if not thousands, of similar contracts for hypersonics and other critical technologies. We are trying to build a proof of concept that can identify key supply chain participants.

Potentially you could analyze the contracts associated with one specific technology (Example: hypersonics) identified in the attached White House doc, or could look at a different technology as well.

USASpending (and this project in general) is totally open and unclassified, so please feell free to take an initial look at the website's data - it can be downloaded either manually as a CSV or through an API.

Time Series point: we are definitely interested in the shift of supply chain relationships over time. It will be key to understanding which firms are becoming more critical actors in the defense industrial base. But before we tackle that aspect of the data we want to have a robust static analysis of supply chain networks, to make sure that we can correctly extrapolate changes over time.

**I have also attached our original problem statement from NSA for reference.**