CGI
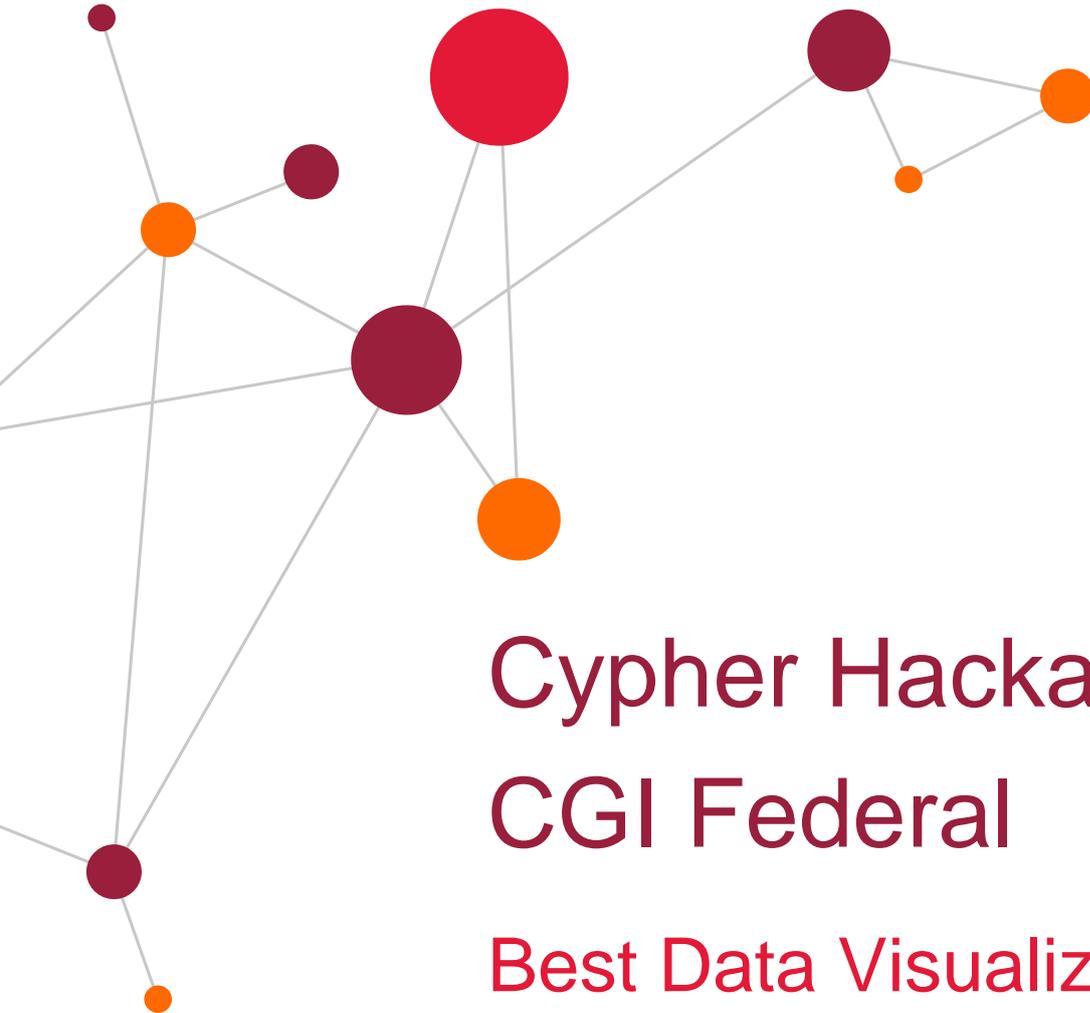
Experience the commitment®
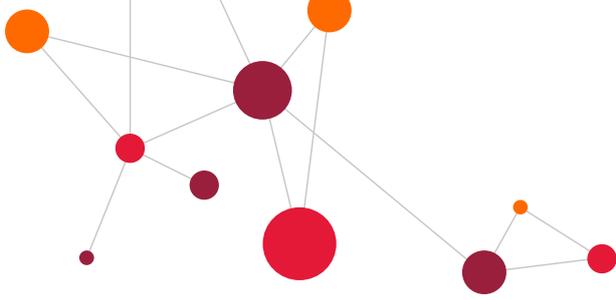
Cypher Hackathon & CGI Federal

Best Data Visualization (Exploratory Data Analysis)

For questions please contact:
hannah.slevin@cgifederal.com

April 9-11, 2021

Challenge Description

Produce a set of unique visualizations that is both parsimonious and sweeping in scope. Your visualization must capture interesting features and relationships within the dataset. You can use any tools to clean and present the data, including but not limited to R, Python, and Tableau. Your visualization will be judged based on the criteria below. You can present your visualizations in an interactive dashboard or slide deck (pdf). If you use python or R, be prepared to present your code.

Choose a Dataset

You can choose from one of the three datasets provided to create your visualization or you can bring your own dataset.

If you choose to bring your own, please provide us with the raw and clean dataset, and a written explanation that includes the goal of your code and the reasons why you produced the visualizations that you did. As a part of this option, you are welcome to submit classwork or other external projects that fit the criteria of the challenge.

Below are the links to 3 datasets that you can download from Kaggle:

Netflix Movies and TV Shows | Kaggle
Top 50 Spotify Tracks- 2020 | Kaggle
Chopped: 10+ Years of Episode Data | Kaggle

Please be aware that both the Netflix and Chopped datasets will require an extensive amount of text processing. The Spotify dataset contains primarily numeric data. Please choose your dataset accordingly.

Criteria

Each submission will be graded based on the following criteria from 1-10. The submission(s) with the highest aggregate score will win.

- Readability (documentation and inline comments)
- Captures core attributes and relationships
- Aesthetics
- Relevance
- Completion

Bonus points will be given for the ingenuity in analysis. For example, using algorithms from libraries such as NLTK, Sci-Kit Learn, Keras, PyTorch, etc, to *successfully* show relationships in the dataset is one way to earn extra points.