Project 1 Group 3 Proposal

Group Members: Daniel Hall, Rob Maysent, Ryan Johnson

Data: Our data topic was over anime

Inspiration: We decided to go with this topic due to us all having a common interest and passion when it comes to anime.

Here's a link as well to an Exploratory Data Analysis on Anime that I was very intrigued by as well [Exploratory Data Analysis on Anime Data Vinayak Nayak (elisonsherton.github.io)]

Research Questions:

- (1) What anime studio has the most top-rated anime?
- (2) What are the most popular genres in anime?
- (3) What anime was the most popular by year, and are there any trends in the top shows?

Datasets: We will be using two datasets from Kaggle to do our exploratory analysis.

https://www.kaggle.com/datasets/angadchau/anime-dataset

https://www.kaggle.com/datasets/brunobacelardc/myanimelist-top-1000-anime

Linear Regression: For our linear regression we might possibly find a correlation between the number of reviews left on a show and the rating of that show being higher. However, since the rating value is constrained within the bounds of a 1-10 rating scale this might cause

issues, so alternatively I may do a logistic regression. There is also a possible correlation between episode count and number of reviews left but that doesn't provide much useful information.

Types of Visualizations: We will be using various of charts based on story that we are trying to convey on each research question, so we will use bar charts, violin charts, scatter plots, pie charts, and etc..

Color Palette: 

Roles:

Daniel Hall – Data Cleaning, PowerPoint, Research Question 1

Robert Maysent -Data Cleaning, Research Question 3, Linear Regression

Ryan Johnson - Research Question 2, Written Report, PowerPoint touchups