Group 1: Christopher Hainzl, Christopher Barbieri, Christopher Hakkenberg, Ryan Podzielny, Peter Bitanga

1. **The task / problem we want to address -**

   We would like to analyze sales of games, user score, and the relationship to sales of game consoles. More specifically, we are trying to see if there are trends between video game sales and game console sales. We also would like to go further and see if there is an increase of certain companies in the stock market, if possible. Using the stock market with the video game dataset is somewhat complicated and might not be appropriate in all instances, but with proper analysis we should be able to combine the two. Meaning do sales of video games, and game consoles have an influence on the stock of the companies that sell these products. Platform sales by year may be interesting as well, seeing consoles rise and become obsolete - though not our main focus.

2. **The data science skills / tools we will use in the study -**

   The main tools we will be using are R Markdown and RStudio Cloud. Ranges from tidying data to using models. We will of course use our toolkit for reproducibility to ensure what we are doing can be reproduced. As for the data itself, as mentioned, we will do some data tidying and wrangling, as well as visualization with graphs such as histograms and scatter plots. In essence we will be "crunching" the data and using EDA to find certain patterns to use specific models like linear or logistic regression. Cluster could also be used if we want to divide our data into groups, but we need to do further EDA to see if this is appropriate.

3. **The source / background of our data -**

   Stock Market Data - https://www.marketwatch.com/ and https://research.investors.com/

   Source: Market Watch and Investors Business Daily

   Game Data - Video Games Sales Dataset | Kaggle

   Source: Kaggle Datasets

Background: The Video Games Sales dataset contains names of video games, genres, platforms, as well as sales. It also includes ratings and scores that were collected off of Metacritic, along with the names of the companies who developed each game found in the dataset. Furthermore, it also includes the region of sales and some more information where the sale was. As for the stock market, our idea is to use numbers from the data set then compare the stock price on the NYSE (New York Stock Exchange) and the TYO (Tokyo Stock Exchange). So, the companies we will be looking into are as follows: Microsoft, Sony, Nintendo, Activision, Electronic Arts, Ubisoft, just to name a few.

4. **How we will prepare the data for analysis**

   We will first tidy and wrangle the data. So basically checking for missing values, linearity, data dimensions, etc. Also with visualizations such as frequency and density plots we can get a better understanding of the analysis we want to use. If we encounter missing data, we can remove it when we bring everything into R after looking at the proportion of missing data found. But if we discover that there are too many missing data to remove or drop, we will handle accordingly, i.e. impute when necessary. After we do EDA we will move on accordingly. Currently we would like to use linear regression, but if the data is unsuitable for it, we will try other models that better represent the data.

5. **What analytical methods we will use to test hypotheses, how does it suit our data** - We will use scatter plots to analyze the relationship between sales and other predictors such as genre, publisher, and platform. We will also utilize a linear regression to see which of those predictors have the most notable relationship with sales. Of course this will come after our EDA and making sure that these methods are appropriate.