**Project Draft: Video Game Sales  
Zain Nasrullah**

**Dataset**: VGChartz Global Sales Game Database  
<http://www.vgchartz.com/gamedb/>

**Inspiration:**  With a fondness of video games instilled in me as I child, I thought it would be interesting to investigate this industry through an analytical framework. Upon doing a quick search, I found a Kaggle activity with a reference to a BeautifulSoup web-scraping script and some exploratory analysis for reference.  
<https://www.kaggle.com/gregorut/videogamesales>

**Hypothesis Test:** I plan to investigate different avenues of the data.

1. Is there a relationship between Release Date & Global Sales?
   1. Implication: Do older games have more global sales because they’ve been out longer?
   2. Method: Correlation Coefficient & Visual Inspection
   3. Null Hypothesis: There is not sufficient evidence to suggest a relationship exists between Release Date & Global Sales.
2. Is there a relationship between Genre & Mean Global Sales?
   1. Implication: Do games of a certain genre sell better than games of another genre?
   2. Method: Crosstab Analysis & Linear Regression Fitting
   3. Null Hypothesis: There is not sufficient evidence to suggest a relationship exists between Genre & Mean Global Sales
3. Is there a console developer that produces better-selling games, on average, than another?  
    [*If Time Permits*]
   1. Implication: Is one console developer much better than the rest?
   2. Method: Random Sample Simulation
   3. Null Hypothesis: There is not sufficient evidence to suggest a relationship exists between Platform Developer & Mean Global Sales.

A notable weakness in this dataset is that it only includes direct video game sales which, in today’s world, does not translate into profitability. Many modern games have opted for free-to-play models where microtransactions or paid expansion content act as the primary revenue generators. These exact dollar values or sales figures would not be captured on VGChartz. The data also does not account for second-hand game sales and renting which were popular trends in the early 2000s before the industry digitized. Additional sources of error also include target demographics and marketing budgets which are missing from the dataset but play a major role in sales. Thus, one should only utilize this dataset to get an idea about how a video game is received by the public.