

Data Science 1 - Final Project Proposal

Adam Haertter & Brennan Mulligan

Proposal

Your proposal is to be submitted in print. Spend time deciding on a topic that interests you. Think about compelling questions that can be answered with data. Once you have an idea of the data out there, try to picture your end product. What type of visualizations, tables, and/or statistical methods would be helpful. Don't think about coding, a particular dataset, or what you know how to do at first. This will increase the likelihood that you will come up with something ambitious and original, and you will be more motivated to learn new things as you work to accomplish your goal. The topic is completely open to your choice, but keep in mind the rules listed above.

Your proposal should contain the following content:

- **Title:** A Statistical Analysis on How Genre Affects the Video Game Sales Market
- **Purpose:** We would like to look into this list that shows the amount of sales by video games and video game series. We want to see how each genre affects the amount of sales. Further analysis can be done by seeing how certain developers produce more for a certain genre and comparing that to what genre of theirs makes the most sales. Publishers may find this information interesting when considering ideas for their next game to see what may be profitable. Specific publishers that we cover may also find it interesting to see if the genre they make the most games for is also the one that makes the most money by game/franchise.
- **Data:** We will be taking data from

<https://www.vgchartz.com/games/>

and scraping it using R, making it tidy, and putting it inside of a data frame so that we can access, analyze, and filter the data for our purposes. We want as much information as possible from the database from the results that the query provides us. It contains columns such as game/franchise names, publishers, amount of units sold by region/overall, and genre.

- **Variables:** Game Name: Strings of the name of the game or franchise

Publisher/Developer: Strings of the publisher or developer of the game

Amount of Units Sold: The amount of units in millions that said game has shipped. This includes amounts from specific regions and the overall amount.

Genre: Strings of the genre of the game

Release Date: The date of the release of the game

- **End Product:** We are producing a graph of the amount of units shipped per genre to compare what is most popular. This will be for a general amount of developers to understand what the market seems to favor. We also want to have a graph that compares amount of units shipped per genre by specific publishers. Then, we want to compare that to the amount of games that the publisher has created by genre. We will provide advice to specific publishers based on what seems to be the most profitable for them.