

## Project 2 Report

Tara Chari/tc455 Wanming Hu/wh298 John Lidstone/jl2463

*A description of the data. Report where you got the data. Describe the variables. If you had to reformat the data or filter it in any way, provide enough details that someone could repeat your results. If you combined multiple datasets, specify how you integrated them. Mention any additional data that you used, such as shape files for maps. **Editing is important! You are not required to use every part of the dataset. Selectively choosing a subset can improve usability. Describe any criteria you used for data selection.***

- Matching top 50 games of each country to their tags

All data was downloaded from SteamSpy. One csv (eg. usTags.csv) for each country with Steam data was downloaded, contained the top 50 games played in that country. There was also player data in these csv's (hours played, number of players, etc).

For the tags (essentially genres) used on Steam, to define games, one csv per tag was downloaded containing the top games in that tag (and other player data as well). In the attached dataScrape.js, each game in each country's csv was matched with its corresponding tags. These tags were added as a string to end of each game's row (under the variable name 'Tag').

All edited csv's saved in the newCountryData folder.

- Finding the top tags for each country

We created a python script that takes in all csv data files for the 10 countries (usTags.csv, etc.). For each country, it counted how many times a type of genre are played, added up the counts, and ranked the top 3 game genres being played for each country. The file "country\_top3\_tags.json" contains the cleaned data for this part.

*A description of the mapping from data to visual elements. Describe the scales you used, such as position, color, or shape. Mention any transformations you performed, such as log scales.*

- Top Games Genres Played by Selected Country on World Map
  - Created map interactive visualization with geoEquirectangular projection for all countries globally. Changed color scheme of map to more accurately represent the countries with the most Steam usage (USA, China, United Kingdom, etc.). Map is interactive, allowing users to hover on a nation to see that nation's top video game genres (to their own expectation or surprise!). Utilized iso numerical data for country codes and cross-referenced that with names (e.g. United Kindom = GBR = 826) for interactivity.
  - There are 4 span that you can hover on above the svg map. When you hover on "Strategic", for example, countries that have strategic games within their top 3 game genres will be highlighted. The same functionality applies for the other 3

spans similarly. This is accomplished by iterating through country\_top3\_tags data and see what are the countries that each genre is popular in. We cross-referenced the id of the span and the genre, and we colorred the specific country based on the id of the path (the id of the path is iso number of the country, which can be access by the iso\_a variable from the country abbreviation).

- Doughnut chart with global ownership data
  - Each of the tag csv's, mentioned previously, also had data on the number of owners for each game. The number of owners for the top game (ranked by Steam) in each tag were saved into a separate file, containing the game's name, its number of owners, and an image of the game (taken from Steam as well). These owner numbers were then mapped onto a circular arc, taking up their calculated percentage of that circle. d3.pie() was used to create the overall shape. An outer arc was also created to place annotations around the chart, corresponding to each type of game. Shadows were placed under paths and text to enhance the visualization of these items as clickable. When hovered over, the colors of the chart slices become more grey to indicate the hover, changing the grey percentage of the original slice's color.

*The story. What does your visualization tell us? What was surprising about it?*

We wanted to look at whether our preconceived notions about what video games people like to play are true. America, for example, when compared to other countries loves its guns - it is also home to Hollywood, where a plethora of highly successful violent action movies are made - does this mean that Americans love violent video games? We are also aware of the massive success that violent video game see, like Call Of Duty, so we thought that this example could be extrapolated to the rest of the industry.

We also asked fifteen of our friends (5 per group member) what they thought the most popular video game genre in America was - mostly everyone said shooter-type games, with a few outliers hinging on strategy.

Rather than speculate, we decided to compile data from steamspy (tool that scrapes data from Steam, the most popular online video game distributor), more specifically the top 50 games by ownership (number of people who own a copy of the game), and make a visualization. We were definitely surprised when everything started to fall together.

First, we were *sort-of* right about America loving shooting games - the shooter genre is number the 2nd most popular. Open-world, however, was the first - proving many wrong. Maybe Americans don't love their guns so much after all. More interestingly, Germany, with some of the strictest gun-laws in Europe, has the Shooter category in a very clear first place. We very

quickly learned that our team, and our friends, were definitely wrong in assuming that generalizations about culture would translate to video game ownership.

The overall global game ownership chart makes for more interesting observations. We all assumed that violent video games would make up for a majority of games purchased, however we were very wrong. The Diplomacy genre, which includes games like Sid Meier's Civilization, and Europa Universalis, is very obviously the biggest genre by ownership. Interestingly, if we assume the Gore, Zombies, Stealth, Shooter, Fighting, and Medieval categories contain violence and compare that to diplomacy, we can conclude that humans love resolving conflicts as much as they love creating them!

Overall, we definitely learned a lot and we think this visualization provides a pretty useful insight into the ownership rates of different genres, and that generalizations and expectations could definitely be broken.