# Games All Around the World

**Daniel Glazer - 318282860**

**Valentin Volovik -**

# Table of Contents

[Games All Around the World 1](#_Toc492211555)

[Table of Contents 2](#_Toc492211556)

[1. Abstract 2](#_Toc492211557)

[2. Introduction 3](#_Toc492211558)

[3. Background 3](#_Toc492211559)

[4. Methodology 3](#_Toc492211560)

[4.1. Data Collection 4](#_Toc492211561)

[4.2. Data Validation 4](#_Toc492211562)

[4.3. Data Analysis 4](#_Toc492211563)

[4.3.1. What? 4](#_Toc492211564)

[4.3.2. Why? 6](#_Toc492211565)

[4.3.3. How? 7](#_Toc492211566)

[5. Evaluation 9](#_Toc492211567)

[6. Conclusions 10](#_Toc492211568)

[7. References 10](#_Toc492211569)

[8. Appendix 10](#_Toc492211570)

## **Abstract**

## **Introduction**

Following the nested model learned in the course (3 question) that was inspired by Tamara Munzner

(Munzner, 2009) and (Meyer, Sedlmair, & Munzner, 2012).

## **Background**

The Valve Corporation a video game developer founded the Steam digital distribution service a decade and a half ago.

## **Methodology**

In this chapter, there is a full documentation on how we collected the data on users and games from the Steam system and the data about countries. We also explain how we validate and analyze the data using visualization as will be shown further in the Data Analysis section.

## **Data Collection**

Valve Corporation, the company that owns and operates Steam, provides a Steam Web API, for gathering information about users' profiles, friendships, game ownerships and playtimes, group memberships, and more. In the relatively new paper (O'Neill, Vaziripour, Wu, & Zappala, 2016) they use this very API to crawl 716 million games and more than 108 million Steam accounts, along with the information that is associated with each account.

Since their paper mostly focus on the user’s relations like friendships and group memberships, we decided to focus on some interesting aspects other than the social aspect like Economy, Games and Gamer distribution.

The dataset comprises of our queries result on the dataset collected by (O'Neill, Vaziripour, Wu, & Zappala, 2016) on all Steam accounts available at the time of collection that specified the country they live in. We also obtained information countries worldwide with a GeoJSON dataset from Natural Earth. Both datasets mentioned above where collected in the past – the Steam dataset crawler collected data in 2013 – 2014 and the GeoJSON data is relevant to the year 2011.

Hence by definition this study is an observational study, or to be more precise a retrospective study –this means that while we can observe the data and establish associations / correlations we cannot establish causation between the explanatory and the response variable.

## **Data Validation**

The data found in the Steam dataset was sampled manually to assure that accounts were associated with real users, both by randomly sampling hundreds of accounts and also by examining all accounts that exhibited extreme behaviors (the scrutiny includes examining their name, friends, and posts on their public profile). Needless to say, all the data collection was done by legal means – the data collected concerning user accounts is

publicly accessible from player profiles, through both the Steam website and client.

Regarding the GeoJSON sourced from Natural Earth it is designed to meet the needs of production cartographers using a variety of software applications so we believe that the data is reliable.

## **Data Analysis**

## **What?**

**Data and Dataset Types**

Identifying the type of data is always the first step in the data analysis process.

The dataset is a combination of 2 datasets – GeoJSON dataset sourced from Natural Earth (that can be produced here <https://geojson-maps.ash.ms/)> and dataset that contains the results from queries (specified in the documentation to the derived data) on the Steam library dataset. The combination of those two results in a dataset in which there is both spatial data and relational data (tables), hence the type of the dataset is both relational and spatial. The dataset availability is static.

In this section only the variables of the derived data are shown (you can read about the variables of the raw data in the Steam website <https://steam.internet.byu.edu/> or in (O'Neill, Vaziripour, Wu, & Zappala, 2016) paper.

For each **country**, we have:

**Numerical**:

**Discrete**:

gdp\_md\_est – an estimation of the country’s GDP

money\_spent – the amount of money spent by the country’s players on games in the Steam library (in US Dollars)

pop\_est – estimation of the population in the country

country\_owners - the number of country’s owners

country\_active - the number of country’s active users

avg\_play\_time - the country’s average playtime (minutes)

num\_casual\_users - the number of country’s casual users

num\_moderate\_users- the number of country’s moderate users

num\_excessive\_users - the number of country’s excessive users

for each X in range of 1 to 10 (for the 10 specific games selected)

gameXowners – the number of country’s owners of game X

gameXactive\_users - the number of country’s active users of game X

gameXavg\_play\_time - the country’s average playtime in game X(minutes)

gameXcasual\_users - the number of country’s casual users of game X

gameXmoderate\_users - the number of country’s moderate users of game X

gameXexcessive\_users - the number of country’s excessive users of game X

**Categorical**:

**Regular Categorical**:

continent – the continent’s country

**Ordinal**:

economy – the country’s economy group

income\_grp - the country’s income group

For each **game**, we have:

**Numerical**:

**Discrete**:

Appid – the game id in the Steam store

Is\_Multiplayer – 1 if the game is multiplayer game, 0 otherwise

price – price payed to purchase a game

Required\_Age – 0 if is suitable for all ages

Rating – the game rating (not for all games the ratings is specified)

**Categorical**:

**Regular Categorical**:

Genres – the game genres such as action, strategy etc.

**Note**: some of the properties are not specified but helped us to present the data to the user (such as country’s name and iso\_a2).

The next step in the data analysis process one would make is looking for relationships between variables.

A relationship between 2 variables could be either described as associated(dependent) or independent. Association can be Further described as either positive or negative.

### **Why?**

So why would we even need a visualization of this dataset?

In general, any subset of statistical terms comes to mind can be computed in seconds and give as basic understanding of the dataset, however, this is only a general feeling of the data and will never give as the “full picture” (Anscombe's quartet is the most vivid example to this fact).

Specifically, in the Steam dataset …. //TODO

**User tasks**

1. Present players distribution in various places
2. Identify \ Locate(?) places with high percentage of addicts for specific game
3. Compare games’ addictiveness
4. Compare game popularity
5. Present players distribution in various places
6. Identify \ Locate(?) places with high percentage of addicts for specific game
7. Identify addictive games (genres)
8. Compare amounts of addicts between different games
9. Compare game popularity
10. Explore similarities / disparities in the same genre
11. Identify correlations or similarities between game’s rating to the active players avg game playing time

### **How?**

## Evaluation

## Conclusions

## References

Meyer, M. D., Sedlmair, M., & Munzner, T. (2012). *The four-level nested model revisited: blocks and guidelines*. Retrieved 8 17, 2017, from http://dl.acm.org/citation.cfm?id=2442587

Munzner, T. (2009). A Nested Model for Visualization Design and Validation. *IEEE Transactions on Visualization and Computer Graphics, 15*(6), 921-928. Retrieved 8 18, 2017, from http://dl.acm.org/citation.cfm?id=1639181

O'Neill, M., Vaziripour, E., Wu, J., & Zappala, D. (2016). *Condensing Steam: Distilling the Diversity of Gamer Behavior*. Retrieved 8 22, 2017, from http://dblp.uni-trier.de/db/conf/imc/imc2016.html

## Appendix