# Visualization Project – Report

**Daniel Glazer - 318282860**

**Valentin Volovik -**

# Table of Contents

[Table of Contents 2](#_Toc491540054)

[1. Abstract 2](#_Toc491540055)

[2. Introduction 2](#_Toc491540056)

[3. Background 2](#_Toc491540057)

[4. Methodology 3](#_Toc491540058)

[4.1. Data Collection 3](#_Toc491540059)

[4.2. Data Validation 3](#_Toc491540060)

[4.3. Data Analysis 3](#_Toc491540061)

[5. Evaluation 6](#_Toc491540062)

[6. Conclusions 7](#_Toc491540063)

[7. References 7](#_Toc491540064)

[8. Appendix 7](#_Toc491540065)

## 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.

The Steam digital distribution service was started

in 2003 and is owned and operated by the Valve Cor-

poration, a video game developer. In this section we

describe the Steam platform and related work that has

used Steam to measure gamers.

## **Methodology**

## **Data Collection**

## **Data Validation**

## **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