Information Visualization – G13-A – Checkpoint V Report

| Francisco Campaniço  IST  Lisboa, Portugal  francisco.campanico@tecnico.ulisboa.pt | João Rafael  IST  Caldas da Rainha, Portugal  joaolucasrafael@tecnico.ulisboa.pt | Rodrigo Oliveira  IST  Lisboa, Portugal  rodrigo.domingues.oliveira@tecnico.ulisboa.pt |
| --- | --- | --- |

# INTRODUCTION

Our project is related to electronic sports, more commonly referred to as Esports (eSports is a common way of spelling this as well, but it’s an incorrect one). Esports differ from “normal” sports because instead of being a primarily physical activity; while any sport will have a component of tactics behind it, there’s also a much higher requirement for physical prowess. The “electronic” component refers then to the fact that this kind of sport is related to playing video games in competitive tournaments, which will, of course, change the necessary requirements for success (there’s a lesser reliance on more “broad” physical conditions, but a higher one in hand-eye coordination, reflexes and tactics (and, depending on the game being played, additional requirements may include a good aim, or being able to manage several characters both effective and quickly, for example).

Esports have become a prevalent medium of entertainment in the past decade, as they combine the current generation’s rising interest in videogames with the already established interest in sports (for example in Portugal, where football fandom is prevalent). They combine the appeal of videogames with the stories that can come out of sports, such as the up-and-comers defeating big established teams, the newest star who rose to the top in their first year, the dismaying loss of a highly praised team in only the first games of a tournament. Our primary motivation is that we both enjoy playing videogames and watching Esports tournaments, as they provide great entertainment.

While it’s easy to find quite a lot of information for Esports, this information is either not organized in a particularly appealing fashion for quick assessment (one cannot conclude many correlations by browsing several player or team pages in wikis) or the only thing that’s able to be assessed is the games which award the highest prize pools, or the highest earning players. It’s not possible to correlate the number of players in a country with its more general statistics (such as, for example, urban population, which would be interesting to do in this case due to the electronic nature of Esports, meaning it requires access to the Internet).

Our approach allows the assessment of more interesting statistics related to Esports besides simpler information such as “highest earning teams”. In Checkpoint I, we decided that our visualization should have the following tasks:

* Analyzing how esports have evolved through the years;
* Searching for the countries with the highest player earnings;
* Identifying at what ages players earn the most;
* Comparing different teams;
* Comparing different games;
* Comparing at which months most tournaments were held.

Our visualization allows all these tasks. Moreover, it specifies some of them further, and adds some new ones:

* It’s possible to not only identify at what ages players earn the most, but one can also filter this information per the desired country;
* It’s possible to compare different games and teams through both monetary statistics (prize pool sum for games, earnings for teams) but number of tournaments hosted/played;
* One can correlate a country’s number of players or their earnings with some of the country’s statistics.

# RELATED WORK

We didn’t find much related work in the field of Esports. It’s common to find infographics (images displaying several interesting stats) such as [this one](https://www.reddit.com/r/DotA2/comments/6jxwcg/infographic_for_the_ti7_open_qualifiers/) for a particularly popular tournament, or “top 100” rankings for earnings or prize pools, such as the ones found on [*esportsearnings.com*](https://www.esportsearnings.com/), but we didn’t take an inspiration in any “in-depth” visualization for the theme. As we stated in the introduction, there weren’t visualizations that went beyond very game-specific statistics (such as the infographic above) or comparing prize pools.

# THE DATA

Our data was obtained from two websites’ APIs: [*esportsearnings.com*](https://www.esportsearnings.com/)for esports-related dataand [*worldbank.org*](http://www.worldbank.org/) for more general country statistics. The statistics for *WorldBank* included urban population, unemployment, GDP and education expenditure. Sadly, the education expenditure statistics were too incomplete for us to use them. The *EsportsEarnings* API didn’t display all the information available on the website, and since the data relied on user submissions, we were unable to use data for additional tasks. For example, while something such as the location where tournaments are hosted would be quite interesting, and the data was available in the API, due to it being user submitted there was too much variation in location format, such as “California, USA” (state, country) and “Greater Warsaw, Warsaw, Poland” (metro area, city, country), and we couldn’t use it.

Another problem with the API is that each user is allowed only one query every two seconds, and the queries are very limited. This proved particularly annoying when we decided to add some interactivity where you can see what teams won (and how many tournaments they won) in a month of the user’s choosing. To get all the tournaments (which we already had to do), we could only query the “100 most recent” with an offset which allowed us to move back to the first tournament in the website’s data (which took place in 1996, for the curious). Since the website had around 32 thousand tournaments, this required 320 queries, and if one does the maths, 10 minutes to perform all necessary queries.

This isn’t much, but the thing is this query doesn’t provide information for who won the tournament. To obtain that kind of information, a different query is required, one for every tournament. This adds up to about 17(!) hours of leaving an automation script querying everything. We put it running on a Google server (thankfully IST provides free server uptime in these services) and we only remembered to check back a week after.

# VISUALIZATION

Leave me here to keep the formatting, please.

## Overall Description

Leave me here to keep the formatting, please.

## Rationale

Leave me here to keep the formatting, please.

## Potential

Leave me here to keep the formatting, please.

# IMPLEMENTATION DETAILS

Leave me here to keep the formatting, please.

# CONCLUSION & FUTURE WORK

Leave me here to keep the formatting, please.