

Project 3 Data Report

Data Description and Sources

A. 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. (10 pts)

We used original data from an online Qualtrics survey conducted as part of a project for INFO 4450 - Computer Mediated Communications in the Spring Semester of 2015 by one of our team members. The survey was open to the public and captured data on League of Legends gamers, their demographics, game habits and ranks in different game types, and their opinions about different in-game scenarios. The survey results were downloaded in a CSV format with survey answers downloaded as coded text values (ex: country name included instead of a number that represents a country id).

We cleaned the data using Excel and rearranged it into a new CSV file. This included the following steps:

- Deleted survey variables not relevant to our visualization (for example: we did not believe that showing the in-game chat methods that players used or their opinions on toxicity were relevant for this visualization so we didn't include them)
- Deleted responses by participants that had entered only obscenities or other inappropriate content in the text field
- Added an original ID to each entry
- Converted the text values for game ranks, years playing and hours playing to a number to allow for analysis (for example: do not play rank = 0, bronze rank = 1)

The table below describes the final variables in our data set (formatData.csv), their sources and what was done to the data to prepare the CSV:

Variable	Description	Manipulation Done
ID	Custom identifier added by the team.	Not Applicable
age	Age of gamer. Corresponding variable in original data set: "What is your age?"	A small portion of ages that were entered as text values (e.g., nineteen instead of 19) were manually converted to numeric values via Excel.

gender	Gender of gamer. Corresponding variable in original data set: "What is your gender?"	Change "Other:" value to "Other" for aesthetics purposes.
country	Country where the gamer resides. Answer to survey question: "In what country do you reside?" Corresponding variable in original data set: "List of Countries"	None.
years	Length that gamer has played LoL. Corresponding variable in original data set: "Approximately long have you been playing League of Legends?"	Converted text fields to numbers in Excel: "Less than 6 months" = 0 "6 months to 1 year" = 1 "1-2 years" = 2 "2-3 years" = 3 "3-4 years" = 4 "4-5 years" = 5 "5+ years" = 6
hours	Number of hours gamer plays per week. Corresponding variable in original data set: "How many hours a week do you play League of Legend?"	Converted text fields to numbers in Excel: "5 hours or less" = 0 "5 - 10 hours" = 1 "10 - 15 hours" = 2 "15 - 20 hours" = 3 "More than 20 hours" = 4
rank1	Corresponding variable in original data set: "What is the highest rank you achieved?-Ranked 3's"	Converted text fields to numbers in Excel: "Unranked" = 0 "Bronze" = 1 "Silver" = 2 "Gold" = 3 "Platinum" = 4 "Diamond" = 5 "Masters" = 6 "Challenger" = 7
rank2	Corresponding variable in original data set: "What is the highest rank you achieved?-Solo Queue"	
rank3	Corresponding variable in original data set: "What is the highest rank you achieved?-Ranked 5's"	

Mapping of Data to Visual Elements

B. 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. (10 pts)

Data Manipulation Conducted:

- A formula in our code parsed the ages variable into 5 blocks to use in the pie and bar graph. The blocks include:
 - ≤ 18 year olds, 19~21 year olds, 22~24 year olds, 25~27 year olds, and older than 27
- A formula in our code combined countries which had 4 or less gamers into a category titled "Other" to use in the pie graph.

Regression Graph:

In the regression graph, we drew rectangles along a grid. Each rectangle represents a group of people in a specific rank that play a specific length of time. We set the opacity of the rectangles to 0.04 per person, which means that each additional person in the same rank, that has the same playing habit will increase the opacity by 0.04 and make the rectangle darker. Rectangles with 25 people or more were assigned the darkest color - each additional person after 25 does not change the opacity of the rectangle because it is over 100%. We also calculated the linear regression, and the result shows that, gamers are more likely to improve in rank in the solo game if play very often or play a long time.

Bar Graph:

In the bar chart, each small rectangle means a person. The X axis means the rank of the person, and the color of the rectangle reflects the age of that person. We can clearly see how many people are in a specific rank with ages in specific blocks. We can also see the distribution of ranks for people in each age block (Orange - ≤ 18 year olds, Red - 19~21 year olds, Blue - 22~24 year olds, Purple - 25~27 year olds, Green - older than 27). As we can see, people are more likely to get higher rank in solo game, most players are under 18 years old, older people are not likely to get higher rank in games, etc.

Pie Chart:

For the pie chart, data was sorted in d3 into three different variables based on the user's selection: age, gender or country variables. An ordinal scale was used to assign three different shades to each pie slice. The age data was parsed using the formula described in the section above.

Code from the following sources was integrated or used as inspiration or examples to develop the chart:

// nesting: <http://bl.ocks.org/phoebebright/raw/3176159/>

// pie: <http://bl.ocks.org/mbostock/3887235>

// onclick + tooltip: <http://jsfiddle.net/rrvtechdiamond/t5S3F/1/>

Header image source:

<http://wallylol.deviantart.com/art/League-of-Legends-Division-Renders-399539315>

Data Story

C. The story. What does your visualization tell us? What was surprising about it? (5 pts)

Although the demographics data we saw in the visualization seems skewed for a standard population, it was not surprising for a LoL population. In 2013, IGN.com reported that 51% of LoL gamers were between the ages of 19-25 years of age, and that 92% were male. They also reported that LoL gamers “chat on social media in 16 different languages every day”, confirming the diversity of countries we observed.

As expected, we also saw a positive correlation between rank and length of time a gamer played LoL, both in years played and in weekly gaming, meaning that the longer time or more hours they spent playing, the better their rank. What was surprising is how much higher, on average, gamers were ranked in Solo queue compared to the other game types. It may be possible that 3 VS. 3 and 5 VS. 5 games are harder, or possibly more frustrating, making it more difficult to get a better rank. This hypothesis is supported by the fact that many gamers chose to play unranked in 3 VS. 3 and 5 VS. 5, while the majority playing Solo did. These observations indicate that further research and investigation is needed why it takes less time to get a better rank Solo, and why gamers choose to play it ranked but do not do so in the other games.

Two other things surprised us: 1. Older players are not likely to have a higher rank. 2. Some players do not play very often on a weekly basis, but have a high rank, while some guys that play very often but do not rank high. It may be possible that this type of a player only plays one kind of game, so although they play very often, they do not spend much time on the other types. However, they may play their game of choice more often within a time frame, and might take less time to rank higher than the players that divide their time between the games.

It is also interesting that we can see less people in Bronze rank than in its neighbor, unranked and silver. This may be because it is faster and easier to upgrade to Silver.

References:

<http://www.ign.com/articles/2013/10/18/league-of-legends-infographic-highlights-eye-popping-numbers>