SOCIABILITY OF MUSICAL INSTRUMENTS

DOES PLAYING A CERTAIN INSTRUMENT CORRELATE WITH HAVING FEW OR MANY FRIENDS?

JIM LAING - HUNG TRUONG - SAMEER HALAI

OVERVIEW

We used data from a facebook application that tracks people's favorite musical instruments. We combined this data with the list of actual friends the people have to test our hypothesis that there could be a correlation between the number of friends people have and the kind of musical instruments they play. We conducted an additional survey to determine which instruments were perceived to be more social than others by the general population.

THE DATASET

DESCRIPTION

The dataset was derived from a facebook application which lets users specify which musical instruments they usually pay. The dataset had 8603 rows each corresponding to a user in the dataset. Each row has the following fields:

- User IDs – unique identifiers for each user
- No. of Friends – the no. of friends
- Instrument – a free text field containing

PLOTS

Some of the plots of the dataset show what the data looks like. The boxplot of friends grouped by each category of instruments shows the overall variance and mean for each category. Most seem to be similar but some of them appear to deviate from the average. The plot showing the mean number of friends by instruments shows a normal distribution – the central limit theorem. The raw plot of the number of friends in the system shows a long-tail distribution. Very few people have more than 600 friends. Most people have less than 200 friends.

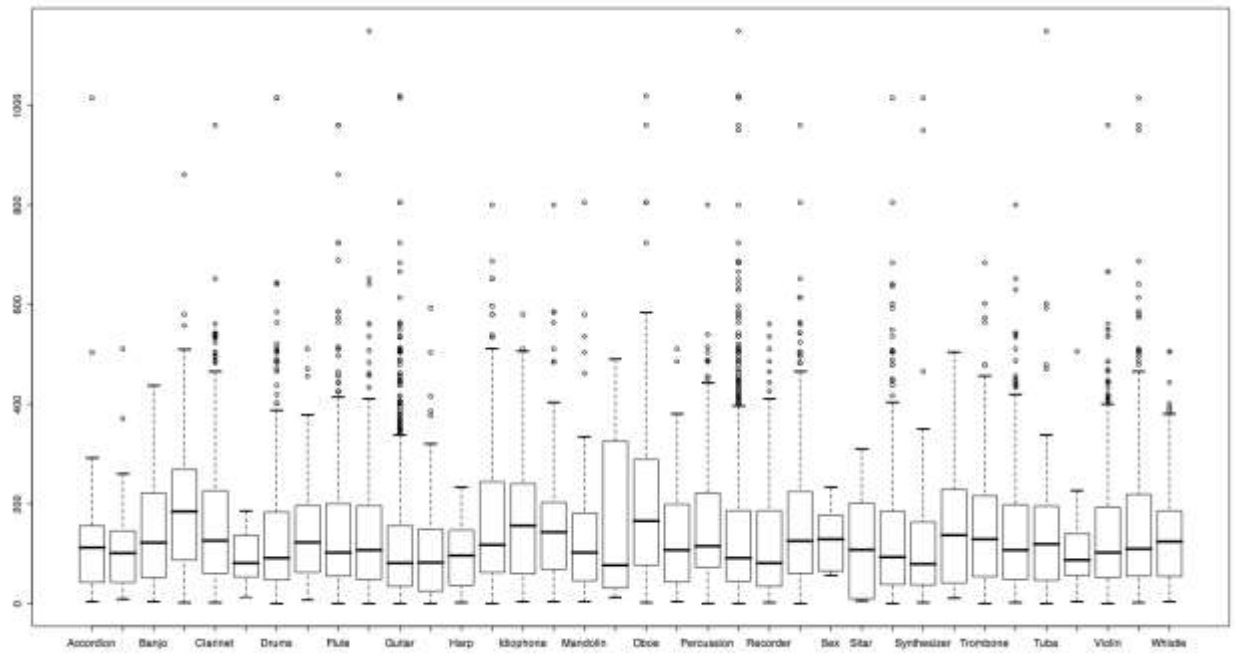

Figure 1: Box plot of number of friends in each category of instruments

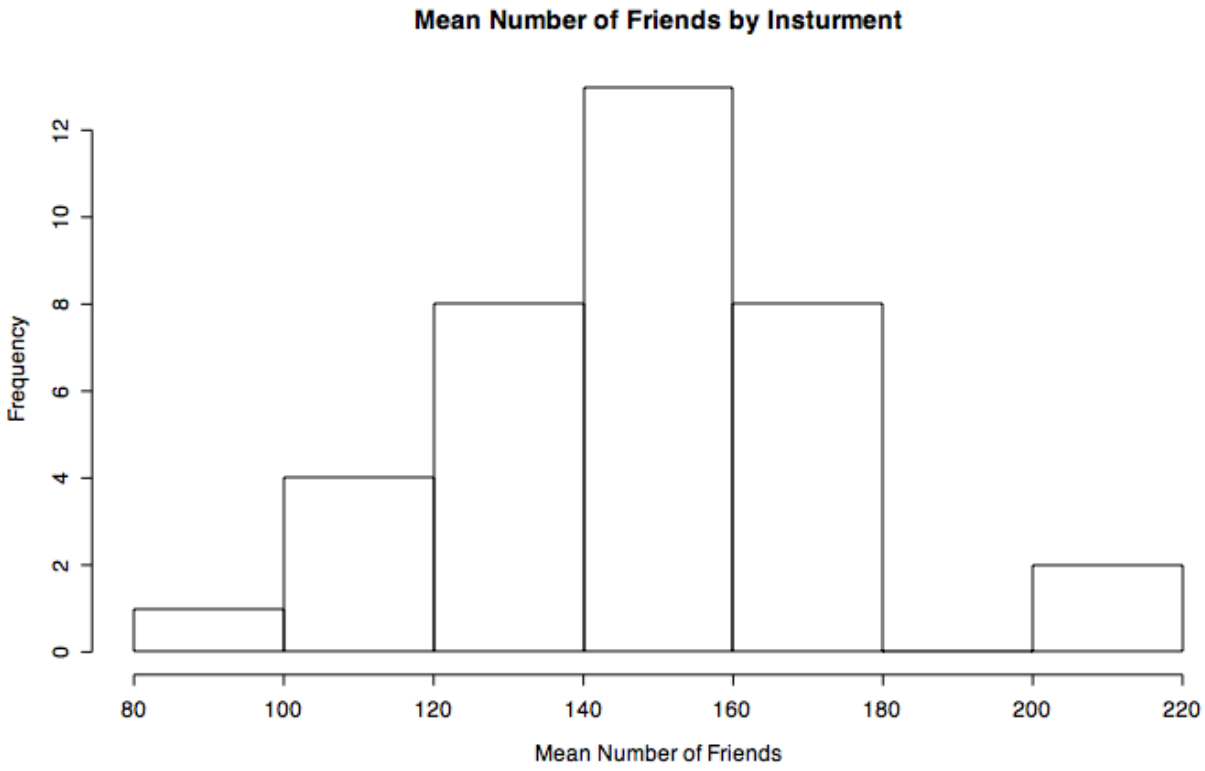

Figure 2: The mean number of friends in each instrument category has a normal distribution

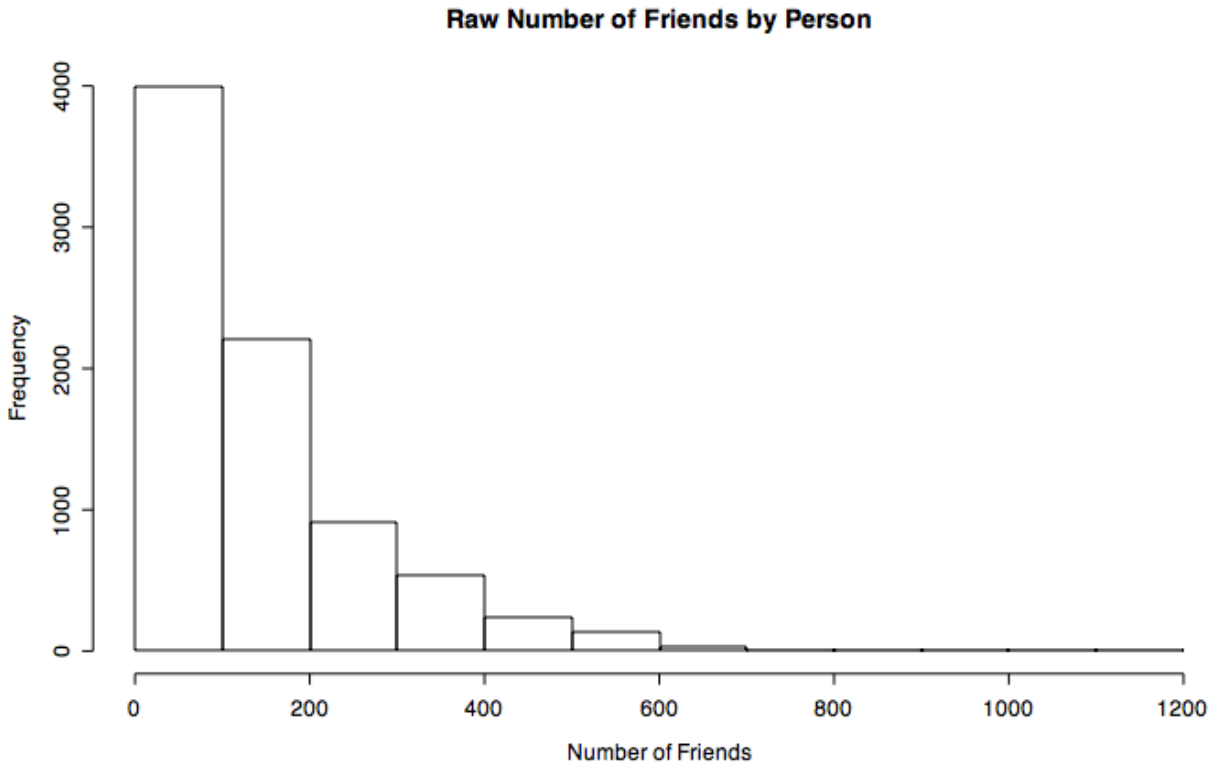

Figure 3: The number of friends per user has a long tail

SURVEY

We sent out a survey which asked people to rank the following instruments by their perceived sociability. We intentionally didn't explain the exact meaning of the word "sociability" because we wanted to our respondents to use their own perception of the word in deciding the rank. We, however, found that most of the users seem to converge on their ranking of perceived sociability.

HYPOTHESIS

We wanted to test if there was any correlation between the number of friends people have and the kind of instruments they play. Our null hypothesis was "People who play instruments which are considered to be more social are likely to have significantly more friends on facebook".

ANALYSIS

PROCESS

We did a lot of qualitative cleaning up of the dataset, specifically for the last column, Instruments, which was free text. We normalized the spelling variations in individual instrument names. We also collapsed instrument names into slightly broader categories retaining the contextual variations in terms of sociability and genre e.g. classical guitar was collapsed with acoustic guitar. We used Wikipedia music classifications to qualitatively assess those parameters. After this cleaning up and consolidation step, we had 8126 rows of very high quality data.

The mean no. of friends for the entire sample was 143.50 with a standard deviation of 137.7345. We did a pairwise t-test across all the simplified categories to find out significant differences in the mean number of friends. We decided to focus mainly on only the 16 instruments that we had survey data for but also retained other significant differences in each pair.

RESULT

Our sociability ranking tally is on the left. The survey results suggest that Vocals are the most social "instrument" closely followed by Guitar Hero (a game) and by Guitar. The Recorder, Flute, Trombone and Saxophone had the highest rank indicating they are the least social.

Instrument	Average Rank
Vocals	5.5
Guitar Hero	5.9
Guitar	6
Percussion	6.9
Drums	7
String Bass	7.3
Piano	7.6
Trumpet	10.2
Violin	10.5
Saxophone	10.7
Clarinet	11.1
Trombone	11.1
Flute	11.7
Recorder	12.4

The matrix below shows only those pairs of instruments which had significantly different no. of average friends in each pair. The bolded headings correspond to instruments that correspond to survey data.

As can be seen from the matrix, there are significant differences in the mean number of friends between people who play the Guitar as opposed to people who play the Clarinet, Flute, Saxophone and Vocals.

	Bassoon	Clarinet	Drums	Flute	Guitar	Harmonica	Oboe	Piano
Guitar	0.00278	1.10E-05	1	0.00569	-	-	-	-
Harmonica	0.03374	0.73672	1	1	1	-	-	-
Horn	1	1	0.07346	1	1.80E-05	0.1343	-	-
Idiophone	1	1	0.6273	1	0.004	0.3262	-	-
Mallet	1	1	1	1	0.00773	0.85867	-	-
Oboe	1	0.31308	3.00E-05	0.02031	2.70E-08	0.00012	-	-
Piano	0.1596	0.55602	1	1	0.67423	1	4.30E-05	-
Recorder	0.1252	1	1	1	1	1	0.00021	1
Saxophone	1	1	0.05529	1	6.70E-08	0.29976	0.54869	0.04352
StringBass	0.21148	1	1	1	1	1	0.00015	1
Trumpet	1	1	1	1	1	1	0.00558	1
Violin	0.50045	1	1	1	1	1	0.00062	1
Vocals	1	1	1	1	0.00069	1	0.07672	1

Figure 4. Pairwise significance results

The interesting finding from these results is that people thought there is a significant difference between the sociability of a guitar over a saxophone and that a guitar is more social than a saxophone. Our analysis of the facebook data reveals that though there is indeed a significant difference between the two categories, people who play the guitar have significantly less friends as compared to people who play the saxophone. People who play the saxphone had an average of 20 more than the mean no of friends of the sample, while people who play the guitar had an average of 20 less than the mean of the sample. We found similar findings for other instrument pairs e.g. people that play the clarinet have a statistically significant no. of average no. of friends more than people who play the guitar.
This conflicts with people's perception that a guitar is more social than a clarinet. In fact, guitar came out to be the least social compared to all the other instruments like Clarinet, Flute and Saxophone which were rated as less social than the guitar.

STATISTICAL SIGNIFICANCE

Since we had a large enough dataset, we were able to find a statistically significant result. Out of the total of 16 instruments that we could have tested our hypothesis we found at least 10 instruments to have statistically significant results. Because our dataset had a long tail, if we had to test statistical significance for other less frequent categories we would need a massive dataset.

POSSIBLE BIASES

Our main dataset has been derived from a facebook application. There might be a bias in the demographics of this population. However, the diversity of musical instruments across all types indicates a rich sample. The survey to test sociability was sent out to si.all.open. Though this list is pretty varied in demographics, we have no way of tracking who responded and there could’ve been some self-selection bias. We did hear qualitative responses which indicated that many were music.

CONCLUSION

We assimilated a dataset that combined data from a facebook application, a survey and qualitative normalization of the dataset to test our hypothesis that people who play instruments which are more sociable might have more friends on facebook. We found statistically significant responses which indicated that our hypothesis is not always true and the results, in fact, indicate a contrary hypothesis.