Ethics in Online Social Network Research – Initial Analysis

# Hypotheses

1. There are differences between the simulated experiments. (Health and Mobile)
2. There are differences in sharing behaviour for different types of information. (Status updates, Photos, Check-ins, etc.)
3. There are differences in sharing behaviour for items with different levels of privacy. (Friends, Everyone, etc.)

# Hypothesis #1

## Chi-Squared Test

A Chi-Squared test looking at participant responses against the experiment group they were in yielded the result **1.669e-06**.

## Kruskal-Wallis Test

A Kruskal-Wallis test on this data yielded the result **1.486e-06**.

## Box Plot

Figure 1

One of the conclusions that can be drawn from the above plot is that participants in each group behaved fairly similarly. In each case there were people who shared lots of data as well as people who shared very little.  
It is surprising to note that, from looking at the data, people were more willing to share when they were placed in group 2. This goes against what we originally anticipated – that those placed in group 1 (Social Networks & Health) would share more.

## Paired T-Test

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Check-In | Friend | Like / Interest | Photo | Photo Album | Profile Info |
| Friend | 1.00000 | - | - | - | - | - |
| Like / Interest | < 2e-16 | < 2e-16 | - | - | - | - |
| Photo | < 2e-16 | < 2e-16 | < 2e-16 | - | - | - |
| Photo Album | 1.3e-12 | 2.5e-16 | < 2e-16 | 5.8e-06 | - | - |
| Profile Info | 0.00016 | 1.0e-08 | 1.8e-13 | < 2e-16 | < 2e-16 | - |
| Status Update | 1.00000 | 0.57526 | < 2e-16 | < 2e-16 | 1.1e-15 | 0.00120 |

# Hypothesis #2

## Chi-Squared Test

A Chi-Squared test looking at participant responses against different types of information resulted in a p value of less than **2.2e-16**.

## Kruskal-Wallis Test

A Kruskal-Wallis test on this data resulted in a p value of less than **2.2e-16**.

## Box Plot

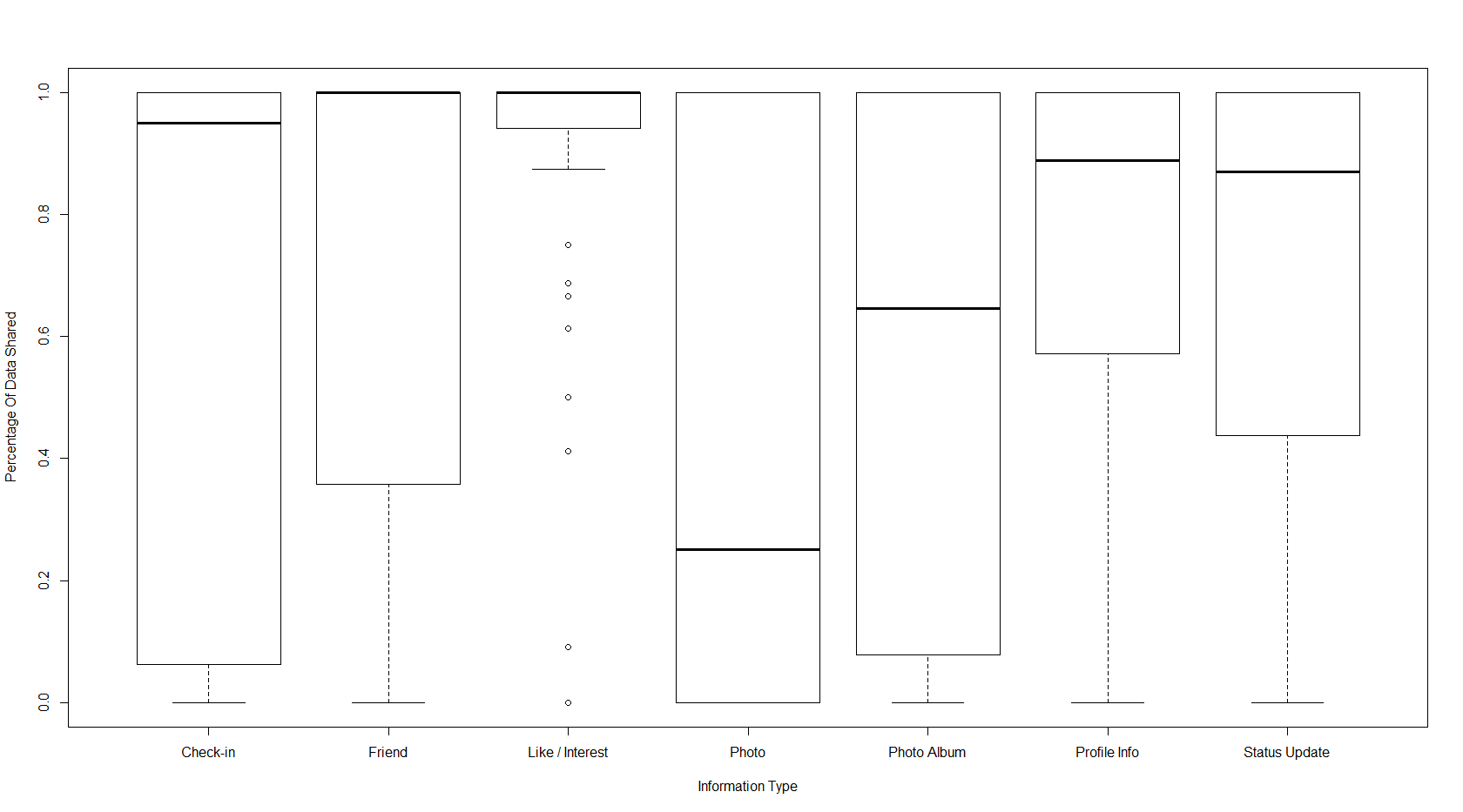


Figure 2

An observation that can be made from the above plot is that people’s sharing habits do indeed depend on the type of information they are asked to share. For example, we can see that likes and interests are shared by almost everyone, as opposed to photos which are shared very little. Similarly to Figure 1, we also see that sharing varies between participant, with some people being willing to share all types of information, and others being reluctant to share anything at all.

Some of the observations therefore mimic what one would expect. People were very willing to share their likes and interests. In the real world, we often see people wearing band t-shirts and so on. This suggests that likes and interests are a very public thing and people are happy to share them. Photos, on the other hand, are somewhat more personal and less likely to be shared.  
Informal feedback from a number of participants suggests that some people did not wish to share photos they were embarrassed about or felt they did not look “good” in.

# Hypothesis #3

## Chi-Squared Test

A Chi-Squared test looking at participant responses against the privacy of different types of information yielded the result **4.295e-13**.

## Kruskal-Wallis Test

A Kruskal-Wallis test for this data resulted in the p value **4.311e-13**.

## Box Plot

Ignoring the **N/A** plot (Where no privacy information actually exists), Figure 3 above allows a number of observations to be made. Firstly,

# Notes

All the data used for this analysis can be found in the **misc.** folder on Mercurial.

* **Results\_Privacy\_Reduced.txt** – Tab-separated file containing the results from the experiment. In this data set, the privacy attribute has been reduced to one of FRIENDS, FRIENDS-OF-FRIENDS, EVERYONE, CUSTOM or N/A.
* **Results\_Complete.txt** – Tab-separated file containing the results from the experiment in full. The privacy attribute has not been modified and so contains some potentially identifiable information. (Facebook includes full names for some reason. Eg: Privacy = FRIENDS; EXCEPT: JOE BLOGGS)
* **Results\_Analysis.rdata** – R workbench.
* **Results\_Analysis\_History** – R command history.