# Mobile Phone Survey

*Ben Arancibia*

*October 21, 2014*

## Looking at Mobile Data (Survey on Mobile Phone Usage)

This is going to look at the mobile survey data from survey data collection done by Significance Labs. There are two surveys done, one for users that Smart Phones and users that Basic Phones. The idea of this report is to look at the usage of phones and demographics associated with different phones. This quick overview will go over the Survey Findings document provided and will also include code in R to show results.

***Survey Findings Document***

**1. Smartphone ownership is really high among low-income populations in New York City.**

This statment is not really true. If you look at the data presented and the statement in the text: "Of those who owned a phone, 84% of respondents owned a smartphone compared to 16% of respondents who owned a basic phone." 84% and 16% are calculated by total smart phone + total basic phone and dividing responses by total. In order for this statement to be true what needs to occur is finding what is defined as low-income and finding percentages.

```
require(plyr)
```

```
## Loading required package: plyr
```

```
smart_phone <- read.csv(file="/users/bcarancibia/bca_mobilephone_BRF/data/Smart_Table.csv",head=TRUE,se
count(smart_phone, "income")
```

```
##   income freq
## 1      0  276
## 2      1  280
## 3      2  428
## 4      3  307
## 5      4  127
## 6     NA   21
```

As you can see from the table you there are about 150 responses that are not below the poverty line. Income with value 3 (20,000-29,000) is hard to tell because the poverty line is around 23,000.

So smartphone ownership among low-income populations surveyed are: $276 + 280 + 428 + 307/1439$

1291/1439 equals about 90%

Looking at this number, I think there is some sampling bias because of the survey location, so I think whoever is presentating this data needs to be cautious in how it is phrased.

I would create a graphic that is total smartphone vs total basic phone and then show the above calculation.

**3. Low-income New Yorkers use Android and iPhone operating systems.**

I think this statment is wrong: "48% reported using an Android operating system and 33% reported using iOS. Only 2% reported using other systems like Windows." The reason is because 48+2+33 does not equal 100, it only equals 83%. The graphic next to it shows both smartphones and basic phones. You can't show a graphic including basic phones and smart-phones but only mention smart-phones in text.

```
count(smart_phone, "os")
```

```
##           os freq
## 1               13
## 2          0  564
## 3          1  801
## 4          2   43
## 5          3   10
## 6 Blackberry    7
## 7    Windows    1
```

Doing a quick count shows you have inconsistencies in the data. On the data enter form you only have have 0-3 recording, but you have blackberry or windows listed in the data. Should blackberry be in 2 and Windows in 3? It's an insignificant amount of data points, but something I just came across.

**4. People use texting as their primary mode to communicate on their smartphone.** "When asked to rank the mode of communication used most frequently on their smartphone, 40% of respondents reported texting. This is followed by calling and Facebook, which 29% and 25% of respondents reported them as most used."

Looking at the data behind this I see Facebook. Was Facebook answered as Facebook First? What is the second? *Look into this*

**5. Low-income populations are willing to pay for apps.**

```
count(smart_phone, "paidapp")
```

```
##   paidapp freq
## 1       0  459
## 2       1  961
## 3       2    3
## 4      NA   16
```

If you look at the table, you have 4 answers but on the survey you only have 2 possible options to answer. Need to take a look at the data. Not significant amount of data, but something to look.

*Coorelation Plot* will try to do this later time need to reformat data

```
#require(ggplot2)
#require(reshape2)
#qplot(x=Var1, y=Var2, data=melt(cor(smart_phone)), fill=value, geom="title") Need to make everything n
```

### Conclusions

Overall, I think it looks good. Couple suggestions, there needs to be some data clean up because there are a lot of data inconsistencies. See above. I tried looking at some coorelations and its hard to find anything. Some fields are data ranges and you have a lot of bias in the survey based on the locations. It would be cool to figure out a way to do this survey over the entire city.

For the next survey I would be careful about the ranges you present. Technical definition of poverty-line is 23k, but your range is 20k-29k, hard to figure out what is below the poverty line. I am not sure the exact definition of low income in terms of numbers, but maybe you can let me know.

If you are looking for a really awesome organization that knows alot about surveys, I can put you in touch with a nonprofit called Keystone Accountability. They are well known for conducting surveys and getting a lot of feedback/closing feedback loops.

Bill had mentioned something about d3 visualization, but he was having trouble. This might be a useful site, basically you can take a .csv, which are attached in the email, put it in into this website: http://raw.densitydesign.org/ and it spits out a d3 visualization. It is completely free, you can download the visualization it makes, or you can embed it into a website. More than happy to help if needed.