

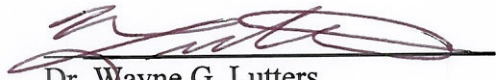


APPROVAL SHEET

Title of Thesis: Users' Privacy and Security Behaviors on Mobile Devices

Name of Candidate: Charles Lenward Blount

Thesis and Abstract Approved:



Dr. Wayne G. Lutters  
Associate Professor  
Department of Information Systems

Date Approved: April 30<sup>th</sup>, 2014

## Curriculum Vitae

Name: Charles Lenward Blount

Degree and date to be conferred: M.S., 2014.

Secondary education: LaGrange High School, Lagrange, GA., 2000.

Collegiate institutions attended:

University of Maryland Baltimore County

Major:

Bachelor of Science, Interdisciplinary Studies - Studies in Human and Computer Interaction, 2005.

Bachelor of Arts, Psychology, 2005.

Master of Science, Human-Centered Computing, 2014 (pending).

Professional positions held:

2006-2008 Engineer, General Dynamics Robotics Systems, Westminster, MD.

2008-2011 Senior Engineer, General Dynamics, Crystal City, VA.

2011-2013 Engineer, Mandiant, Alexandria, VA.

## ABSTRACT

Title of Document:                   USERS' PRIVACY AND SECURITY  
  BEHAVIORS ON MOBILE DEVICES

  Charles Lenward Blount, Master of Science,  
  2014

Directed By:                           Dr. Wayne G. Lutters , Associate Professor,  
  Information Systems

Preferences and behaviors for privacy management with mobile applications are difficult to capture. Previous measures are mostly based on self-report data, which often does not accurately predict actual user behavior. A deeper understanding was sought, gleaned from observing actual practices. This thesis analyzes 11,777 applications from the Google Play marketplace in order to determine the impact of privacy settings on purchase behavior. This was done by looking at the effect of the number of privacy concessions as well as the effect of individual concessions and category on number of downloads. It was found that users of paid applications do not have a preference for fewer privacy concessions. This study further reinforces the disconnect between the user's often stated preference for privacy and their actual behavior -- a discrepancy known as the "privacy paradox". Theoretical and practical implications are discussed.

USERS' PRIVACY AND SECURITY BEHAVIORS ON MOBILE DEVICES.

By

Charles Lenward Blount

Thesis submitted to the Faculty of the Graduate School of the  
University of Maryland, Baltimore County, in partial fulfillment  
of the requirements for the degree of  
Master of Science  
2014

UMI Number: 1571723

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 1571723

Published by ProQuest LLC (2014). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code



ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 - 1346

## ABSTRACT

Title of Document:

USERS' PRIVACY AND SECURITY  
BEHAVIORS ON MOBILE DEVICES

Charles Lenward Blount, Master of Science,  
2014

Directed By:

Dr. Wayne G. Lutters , Associate Professor,  
Information Systems

Preferences and behaviors for privacy management with mobile applications are difficult to capture. Previous measures are mostly based on self-report data, which often does not accurately predict actual user behavior. A deeper understanding was sought, gleaned from observing actual practices. This thesis analyzes 11,777 applications from the Google Play marketplace in order to determine the impact of privacy settings on purchase behavior. This was done by looking at the effect of the number of privacy concessions as well as the effect of individual concessions and category on number of downloads. It was found that users of paid applications do not have a preference for fewer privacy concessions. This study further reinforces the disconnect between the user's often stated preference for privacy and their actual behavior -- a discrepancy known as the "privacy paradox". Theoretical and practical implications are discussed.

USERS' PRIVACY AND SECURITY BEHAVIORS ON MOBILE DEVICES.

By

Charles Lenward Blount

Thesis submitted to the Faculty of the Graduate School of the  
University of Maryland, Baltimore County, in partial fulfillment  
of the requirements for the degree of  
Master of Science  
2014



© Copyright by  
Charles Lenward Blount  
2014



# Table of Contents

Table of Contents .....	ii
List of Tables .....	iv
List of Figures .....	v
Chapter 1: Introduction .....	1
Background .....	1
An Abstract Illustration .....	3
Chapter 2: Related work .....	6
Privacy .....	6
Defining privacy and security .....	6
The Construct of Trust .....	6
The Construct of Privacy .....	7
Irregularities .....	9
Validity .....	10
Challenges to the Internal Validity .....	10
Substantiating External Validity .....	11
Operationalization and Theoretical Foundations .....	12
<i>General Findings on Security and Privacy Management</i> .....	13
Measurement and Categorization .....	13
Consumer Privacy Index .....	13
Factor based Models .....	14
Prior Findings .....	17
Desired Privacy Depends on Subject .....	17
Privacy Concern is Impacted by Demographics .....	17
Privacy Behaviors in Mobile Devices .....	18
Chapter 3: Study Design .....	19
Objectives .....	22
Methodology .....	23
Analysis Techniques .....	24
ANOVA .....	25
Linear Regression .....	26
Dealing with Censoring of Variables .....	26
Chapter 4: Findings .....	27
Domain Summary .....	27
Data Summary .....	28
Quantitative Analysis .....	29
Privacy Concession Count to Number of Downloads .....	31
Individual Concessions .....	32
Application Category Effects .....	56
Qualitative Analysis .....	57
Interesting Case of Personalization .....	57
Chapter 5: Discussion .....	59
The Effect of Privacy Concession Count .....	59

The Effect of Individual Privacy Concessions.....	60
Categorical Effects.....	61
Chapter 6: Conclusion.....	63
Limitations .....	63
Practical Implications.....	65
Design Implications .....	65
Policy Implications .....	67
Governmental /Societal.....	67
Business .....	68
Personal.....	68
Theoretical Implications .....	69
Bolstering Related Findings.....	69
Current Models .....	74
Cohort Effects .....	74
Future Work .....	75
Appendices.....	79
Terms of Service.....	79
Graphs .....	80
Categorical Descriptions.....	84
Bibliography .....	85

## List of Tables

Table 1 Summary of Domain.....	27
Table 2 Histogram of Privacy Concessions .....	28
Table 4 Category Descriptions.....	84

## List of Figures

Figure 1 Download Distribution .....	29
Figure 2 Privacy Count Across Downloads.....	31
Figure 3 Privacy Count by Category .....	32
Figure 4 The Effect of Full Network Access on Downloads.....	34
Figure 5 The Effect of the Phone Status Concession on Downloads .....	35
Figure 6 The Effect of the USB Storage Concession on Downloads .....	36
Figure 7 The Effect of the View Network Connection Concession on Downloads ...	37
Figure 8 The Effect of the View Wi-Fi Connections Concession on Downloads .....	38
Figure 9 The Effect of the Test Protected Storage Concession on Downloads .....	39
Figure 10 The Effect of the Prevent Sleep Concession on Downloads .....	40
Figure 11 The Effect of the Wi-Fi Connection Status Concession on Downloads ....	41
Figure 12 The Effect of the Run at Startup Concession on Downloads .....	42
Figure 13 The Effect of the Running App Retrieval Concession on Downloads.....	43
Figure 14 The Effect of the Control Vibration Concession on Downloads.....	44
Figure 15 The Effect of the Modify System Settings Concession on Downloads.....	45
Figure 16 The Effect of the Receive Data Concession on Downloads.....	46
Figure 17 The Effect of the Find Accounts Concession on Downloads.....	47
Figure 18 The Effect of the Pair Bluetooth Concession on Downloads.....	48
Figure 19 The Effect of the Access Bluetooth Settings Concession on Downloads ..	49
Figure 20 The Effect of the Change Network Connectivity Concession on Downloads .....	50
Figure 21 The Effect of the Read Sync Settings Concession on Downloads .....	51
Figure 22 The Effect of the Close Other Apps, Set an Alarm Concession on Downloads .....	52
Figure 23 The Effect of the Change Screen Orientation Concession on Downloads. 53	53
Figure 24 The Effect of the Control Near Field Communication Concession on Downloads .....	54
Figure 25 The Effect of the Read Your Own Contact Card Concession on Downloads .....	55
Figure 26 The Effect of the Add Words to User Defined Dictionary Concession on Downloads .....	56
Figure 27 Concession Count Versus Download Rate by Category .....	58
Figure 28 Standard Errors of Downloads Across Concessions Counts .....	80
Figure 29: Difference in the Music category .....	81
Figure 30: Write To call log in Health Applications .....	82
Figure 31: Modify Contacts in Health Applications.....	83

# Chapter 1: Introduction

## Background

Mobile devices have become more common with the popularity of cell phone, with smart-phones making a large amount of information and capabilities available to their users. Increasingly, this information flow has been emerging as a bidirectional one. People often expect their data to be accessible and omnipresent. With the rise of cloud computing, the data and services are provided to the user whenever they might want them. While this is useful and convenient, these capabilities come with choices that the user must make which will allow others access to their device.

As Moore's law takes its effect on mobile devices, the cost of smart-phones decreases over time. This has had the effect of widening the smart-phone market, and increasing the popularity of smart-phones in general. Recently, smart-phones made up 30 percent of the mobile phone market (Nielson 2010).

High-end smartphones have many abilities, from their integrated sensors and abilities that increase the exposure of its users' personal life to the applications that run on it (Grudin, 2001). High end smartphones can have myriad features, such as front and rear cameras, microphones, gyroscopes, a GPS receiver, call history, the transcripts of voicemail messages, the text message history, a call history, contact lists, software usage statistics, web browsing history, and a list of nearby Bluetooth devices.

The addition of features and an increased pervasiveness, have led to smartphones becoming more integral to people's lives and have caused the security and privacy issues of one's smartphone and its applications to influence the users' personal security and privacy. For example, the average camera being used in the United States is in a phone. (Nielson 2010)

Given the pervasiveness and functionality of today's smartphones, these devices could provide a veritable cornucopia of possible information to others. This could be intentionally provided or taken without permission; an important question is how people deal with this, and how this can be predicted, changed, and measured.

Although some have argued that privacy and security concerns should be separated, from the perspective of the user, the actions of an approved applications can have the same negative effect as those of unsanctioned entities regardless of whether or not the unwanted behavior was somehow disclosed beforehand (e.g. the disclosure of information from a trojan to a third party is the same as the disclosure of the same information from one company to a third party). This is compounded by the fact that if the user does not read the user agreement the unwanted effects may be unexpected. Therefore, the analysis of the security and privacy decisions of the users are often analyzed and discussed together in this paper, for the purpose of defining a behavioral and establishing the user's mental model.



### An Abstract Illustration

Consider the following scenario in which John Doe would like a music player for his smartphone. He uses the built-in store on his phone in order to identify and download this music player. He browses some, and, upon finding one that interests him, he chooses to purchase it. Upon clicking the purchase button, he will be reminded of what kind of access this application requires in order to proceed. After proceeding he decides he likes it and continues to use it. However, one day an update is required which notifies him of a change in the permissions that the application requires. He reads this, understands what he reads, and deeming this acceptable, he hits accept, downloading the update.

This example presents a best-case scenario, in which the marketplace notifies the user clearly of how the application is to interact with his smartphone, in the form of permissions. In this case, our user understands this interaction, and makes an informed decision which takes all of this into account. Whether or not these assumptions are being met in the real world, situations like this are occurring on a more frequent basis, and what is at stake each time this decision is made is as important as the data on the phone itself.

When John, the sophisticated user from our example, had an advantage over many users when he was looking for the application, he had an idea what he was looking for. Therefore, at some point in his search, he will either decide that has found that which he seeks, or that he will not find it, as suggested by information foraging theory. Then he will finally decide whether to install an application, or not. It is this

decision step that is irreducible. Although one day there may be agents that could serve as a proxy for the user's role in this decision, this decision must still be made. Even if the rest of the security holes could be plugged, the question of how users value the various privacy aspects of applications remains. After all, the mobile platform differentiates itself from other computing platforms not because of its power, security or privacy, but because of its mobility and convenience.

A user's security and privacy concerns could influence any step of the illustrated search process. It has recently been found that some users take permissions into account when choosing which applications to use, install and keep. For example, 51% of teens who download apps have reported avoiding an app due to permissions (Madden, 2012). This is similar to the finding that 54% of the adult population have chosen not to install a mobile application based on concern over personal information disclosure, based on self-report (Boyles, 2012). If there is a privacy or security breach afterwards, it will not have affected this decision, as application consumers are not assumed to be gifted with prescience. While a breach may affect future decisions, a user's opinions on security and privacy as concepts may have to do with actual past practices and experiences, values, or even stories in the media that they have heard.

There are several additional unresolved question. First, is whether average users have the knowledge to make informed decisions regarding this issue. That is, do their mental models match the actual functioning of smartphones closely enough for their decisions to correspond to their intended end result. Another is how best to measure and predict privacy and security concerns. If one wanted to observe privacy and

security actions in order to infer the concerns of users, one could spy on user behavior using means such as a malicious application. However, this would ironically breach their privacy and security. Any volunteer study is subject to selection bias, with those less concerned with privacy and security opting to participate in the study more often. Furthermore, any study relying on self-report is subject to the well-established biases that come with it, such as the observer effect and issues with poor recall.

We live in a world where private information is very valuable, whether to the government, commercial institutions, or to other people. Given the increased integration of mobile devices into our lives, and the value that people say they place on privacy, it is important to figure out how people actually behave regarding the management of their privacy. Misunderstanding people's actual tendencies may have a consequence of us adopting policies, consuming products, and structuring our lives in ways that prove detrimental in the long run. For example, the adoption of new laws could either help or hinder the protection of people's data.

## Chapter 2: Related work

### Privacy

#### Defining privacy and security

As one of the large pieces of Human-Computer Interaction (HCI) theory, a part which is interdisciplinary and well established, privacy and trust are some of the few parts of HCI which have both legal implications, as well as the prospect of affecting the potential of firms (Corritore et al. 2003). Yet there are still many obstacles to overcome in order to transform these nascent theories, models, and concepts, into a more mature line of inquiry. This section endeavors to explore the state of these concepts and their validity as scientific concepts by exploring the issues with, and state of, both privacy and trust.

#### The Construct of Trust

A major issue with the application of the social science conceptions of trust is that there are many applications which apply different definitions of trust. One popular article, which provides an illuminating example, defines a subtype of trust (online-trust) as “an attitude of confident expectation in an online situation of risk that one’s vulnerabilities will not be exploited.” (Corritore et al., 2003) The authors go on to distinguish their definition of trust from faith, competence, credibility, reliance, and trustworthiness. Additionally, the authors highlight multiple types of trust (slow versus swift and cognitive versus emotional), degrees of trust, and stages of trust.

Degrees of trust is a concept described by Corritore et al. (2003) to be “the depth of trust that an individual has.” However, in the primary source from which this definition is cited as being derived says that these forms of trust (note that these were not stated as existing on a continuum of a singular modality) are “distinguished in terms of the particular contexts in which trust plays a role”(Brenkert, 1998). The gulf between these two definitions is enormous, and although the examples listed by Corritore fit as a result of meeting the explicit descriptions of the constraints as described by Brenkert. The problem is that Brenkert then lists the constraints on which each degree of trust is built. Each degree (basic, guarded, and extended) has the preceding conditions as prerequisites for its own (the basic degree has 2 conditions which it needs to occur, guarded requires each of the basic degree’s context’s and has its own, and extended requires each of the guarded degree’s contexts, and has its own). This is important because while the more modern definition of online trust takes a reductionist approach, the inclusion of a feature that only exists (and in fact derives from ) in the context of observation as a part of the definition conflicts with this approach.

#### The Construct of Privacy

In early studies on privacy a large amount of care was taken when defining and comparing privacy concerns. One early and thorough framing of privacy is laid forth by Altman (1977). Altman first presents framework in which to conceptualize privacy as the ability or act of regulating social interaction. It is described as a dialectic process (similar to the idea of the ‘unity of opposites’) wherein the ideal privacy is

revealed by its seemingly contradictory facets (qualities) which are exposed at various points in time and varying circumstances. Secondly, privacy is described as an optimization in which the optimal amount (which varies according to the aforementioned dialectic process), is not that of maximal or minimal value, but one which crowding and isolation are both to be avoided. Finally, privacy is described as being modulated by behaviors across modalities in response to stimuli that also cross modalities (e.g., verbal responses to physical space intrusion). (Altman 1977) This framing is then used to describe how privacy (as a process and as behavioral actions) differs between cultures.

Following up on this heavily influential paper, this framework was reinterpreted in an attempt to deal with technological improvements that have occurred in the preceding three decades, this is done by illuminating the various ways in which technology is part of the context of privacy and as a mode of mediating privacy. These various ways were enumerated by the listing of the boundaries negotiated in order to have privacy. The first boundary listed is the self vs others (ie is my action truly my own (or that of my larger group, for example the way a professor's research reflects the school as much as himself), and who is the observer of the action). Secondly, the temporal boundary is mentioned (eg how do past actions inform current ones, and which past actions are also present actions by virtue of the fact that things on the Internet often have no shelf life). Finally the boundary between the exposure and isolation is highlighted (Palen and Dourish 2003), (of course, this was seen as part of the very definition of privacy by Altman (Altman, 1977)).

Currently there are many different definitions of privacy. For example, when studying the effect of privacy concerns on user behavior, researchers found that of 16 studies used to lay the groundwork for the study, none shared the exact same definition. Once broken into 5 different criterion, however, they each shared the concept of privacy concern requiring “Internet customer’s concern for controlling the acquisition and subsequent use of the information that is generated on him or acquired on the Internet”(Castañeda and Montoro, 2007). This is to be taken with a grain of salt, as the research was marketing oriented and the construct could be argued to be different across the two disciplines. Furthermore, it was not an exhaustive search, and their sampling of the literature was not even described as being in any way random.

### *Irregularities*

There are several studies that completely eschew this foundational notion of privacy as a process or even as the amount of self-control. Their operationalization of privacy (which they use interchangeably with the term E-privacy) is closer to that of the degree to which the actions of a site or Internet application acts with adequate fiduciary responsibility.

One proponent of this version of the construct is Nickel and Shaumburg (2004), who put forth a model (termed Electronic Privacy) which trifurcates the concept of privacy (in the domain of website personal information disclosure) into the Openness (which is commonly known as transparency), Control (which seems analogous to the ability to alter or remove previously entered information, as well as confining the information disclosed to the website to which it was disclosed unless explicit

permission is given), and Security. This study does much to bolster the reliability of other studies which have shown that several earlier studies with their result that higher amount of self-reported trust led to a higher amount of disclosure. As trust is generally accepted to be a willingness to expose vulnerabilities to others based on their expected reaction (Riegelsberger, J. 2005) (as found by a researcher applying questionnaire made by Kammerer, M. (2000)) . Then based on this definition, the authors assumption that trust leads to disclosure succumbs to the fallacy of reification, as a construct cannot cause action stated to require the construct as part of the construct's very definition (for example that's why one cannot say that alcoholism causes the need to consume alcohol, indeed, if you don't need to consume alcohol, you do not suffer from alcoholism). In this study, the construct validity of privacy is low. The representation validity for the model itself is low as the three supposed factors to perceived privacy are both difficult to operationalize (as done by the authors informally operationalizing through the presence of an explicit statement), and difficult to make discriminant as they allegedly influence the same factor (perceived privacy), and are not stated to have an effect on anything else (hence these three constructs have low discriminant validity)( Cronbach and Meehl, 1955).

### Validity

#### Challenges to the Internal Validity

Due to the proliferation of user self-report as a method of determining the preferences and potential actions of users one might assume that the external validity of this



approach has been well established. This, is actually far from the truth, in fact, the opposite has been demonstrated. In one such study, the users' behavior contrasted with their reported level of privacy concern. The users were grouped according to an established method, and the subsequent experiment (which involved a shopping interface that the users, through a clever use of deception, thought they were evaluating, and in doing so revealed private information). The researchers, perhaps not missing the irony that would have resulted if they had neglected to do so, placed a caveat on their claims by relaying the fact that their sample was heavily biased towards college students which may challenge the external validity of their own findings as a result (Spiekermann et al., 2001).

#### Substantiating External Validity

Fortunately, there have been some studies that have established the External validity of Privacy and Security Measures. One shining example (due to its thoroughness and its completeness), is one done by Metzger (2006). This design used multiple data collection techniques: A pre-test questionnaire; a fictitious, yet verisimilar website; and a posttest user assessment. Then these measures were used to provide statistically significant evidence which supported each of their hypotheses. The most relevant to this discussion are that: 1. "Internet users' concern for privacy online negatively influences their past online information disclosure. " and 2. "Internet users' (a) trust of a company's Web site and (b) past online disclosure positively influence their current information disclosure to the company's Web site." This simultaneously bolsters the convergent and predictive validity of the questionnaire, and the convergent validity of the application of the surrogate website. As the authors

point out, the actual user statistics, which would corroborate this, and thereby bolster its ecological validity, are generally proprietary and therefore difficult obtain.

### Operationalization and Theoretical Foundations

One problem with those that form a theory composed of “intervening variables interpolated between a set of measurable antecedent conditions” is that the independent and dependent variables proposed are often complex ideas which couldn’t be quantified, and which an entire field of study could be devoted (Koch, 1992). One example highlighted by Koch (1992) is that of Tolman (1936). In this example, the components (independent and dependent variables) include Environmental Stimuli, physiological drive, heredity previous training and maturity. These are examples of variables which are would be of little use in operational analysis as outlined by Bridgman (Koch, 1992). However, these variables bear a stunning to the types of variables listed in the construction of the concepts, and models, of privacy and trust. For example, one researcher defines trust as consisting as being influenced by social indicators, personal experience, understanding, communality, social indicator, personal experience, understanding, communality (Tan and Thoen, 2000). These are not operationalized in the paper, but they are so broad as concepts, it is easy to doubt that they ever could be. If there is no path from the theory to measurable, objective data, the question is obvious. Are the current conceptions of models scientific in nature, and are the theories, in this line of research, scientific theories?

## General Findings on Security and Privacy Management

### Measurement and Categorization

#### Consumer Privacy Index

The study of how to measure and categorize people's propensity for privacy and trust has been varied, and been studied in, and applied to many disciplines. Some of the earliest research in this area evolved from an early work by Alan Westin, a professor of public law and government ethics, *Privacy and Freedom* (Westin, 1967). They spoke provided in only definition for privacy, and also examined some of the threats to privacy by a society and technology. Early participation in framing the debate around and the definition of privacy lent credence and influence to some of his early studies on privacy. Westin participated in the development of a survey of privacy, which resulted in creating categories of users' attitudes, and influencing models of users' attitudes (Kumaraguru and Cranor, 2005). Although this study was quite influential, it was funded by Equifax and kept private (unpublished), the said executive summary is the often cited manifestation of its results.

The following year the Westin conducted a follow-up survey (Kumaraguru and Cranor, 2005). This survey included a survey questions regarding attitudes towards divulging information which would normally be private. Westin's Consumer Privacy Concern Index used the responses to divide respondents into three categories, Low(0-1 concerns), Medium(2 concerns), and High(3-4 concerns).

This index evolved into what was called the privacy segmentation index in 1995. Its creation was first published for a national survey undertaken by Dr. Westin while he

worked at Louis Harris and associates. This index divides the public (living in America) into three groups, privacy fundamentalists, privacy and concerned, and privacy pragmatists. Privacy fundamentalists have a high amount of privacy concern, privacy pragmatists have the set of balanced attitudes, while the privacy unconcerned have little to no concern (Westin, 1990).

The groupings devised by Westin have been derived in other manners as well, for example, by using SAS's partitional clustering to identify groups of general online privacy attitudes (Ackerman, Cranor and Reagle, 1999). This study, however, divided general concern into the following three groups: privacy fundamentalists, the pragmatic majority, and the marginally concerned names which differed from those previously mentioned.

#### Factor based Models

##### Smith et al.'s Information Privacy Instrument

In response to the one-dimensional instruments used to test and define people's privacy concerns, Smith et al. endeavored to create a privacy instrument which would discern the nature of people's privacy concerns as well as demonstrating its validity (Smith et al., 1996). Smith et al. felt that a new study and an instrument was needed as the "dimensionality is neither absolute, nor static". After a thorough and methodical process of studying existing definitions and measures of privacy, an instrument was derived which measures an individual's privacy concern by how concerned they are in about their personally identifiable information is being collected, if it is being internally, or externally (relative to the collector) misused,

whether it can be improperly accessed, or whether or not there are deliberate or accidental errors in the personal data.

This index was validated empirically and popularized a name, which seems to have stuck: concern for information privacy (CFIP) (Stewart and Segars, 2002). The original study (Smith et al., 1996) validated both with experts, and by comparing with other privacy scales, while the study was additionally validated (found internally consistent while finding evidence of second order underlying factors) by doing a Confirmatory Factor Analysis on surveys distributed to consumers using the “mail intercept” method in DC, Georgia, California, and Texas.

### **Internet Users' Information Privacy Concerns**

Tying together on social contract theory and the Concern For Information Privacy scale, Malhotra, Kim, and Agarwal created and validated a model for information privacy concern which included second order relationships which influenced privacy behaviors. This construct, called Internet Users' Information Privacy Concern (IUIPC), formed a model in which IUIPC was a second order factor related to personal dispositions with three factors, collection, control, and awareness. These factors are derived from social contract theory. In addition, IUIPC was hypothesized to be have an effect on trusting beliefs and risk beliefs.

### **Privacy Theory**

It has been asserted that customers disclose private information if benefits of their disclosure exceed their risks. (Culnan and Armstrong, 1999) This contrasts with some older more esoteric concept privacy regulation theory (positing a dynamic dialectic

regulation process (Altman, 1977). This theory has not been well established, and has been shown to have its limits, as people regulating their privacy online have been shown to act on incomplete information, exhibit bounded rationality (limits on the complexity of mental model, and the amount of information used at once), and have psychological biases affect their decisions (Acquisti and Grosslags, J.). Privacy has also been measured as a cost has been measured on a self-reported scale (Nickel and Schaumburg, 2004).

### **Limitations**

Unfortunately the testing of privacy and privacy concern is fraught with several practical issues of concern. The first is that the research of privacy concern, especially the early studies, are proprietary, not open source academic research. Some privacy studies are industry studies or done by corporately sponsored research groups (e.g. Harris-Equifax), and often kept proprietary. This means that finding the publicly available summaries of these studies is often only possible through the Internet archive, and the original full report is often elusive. The conclusions, while interesting for providing context, may lack the rigor of an academic study.

### **Definition**

For the purposes of this study, the concept of privacy borrows heavily from that of Castañeda and Montoro (2007). We treat privacy as the extent to which a person can control the acquisition and subsequent use of the information that is generated about him or her. By using this definition, we are sticking to a consensus definition and one that can be tested. If one gives data to many different parties, their privacy is decreased. If they trust these parties, they are trusting that their desires in terms of

dissemination of their personal data will be followed, dampening, yet not eliminating, the decrease of privacy.

### Prior Findings

#### Desired Privacy Depends on Subject

Ahern et al. (2007) looked at the ratio of private to public photos and found that people were significantly more private with photos containing people than they were overall, additionally, they were more private with photos containing people than with either places, events, objects, and activities. Additionally, they found that locations where photographs were frequently taken were more likely to have their privacy setting set to private. It has also been shown that people will give away private information for less if the information paints them in a more socially desirable light or if they personally value privacy little (Huberman et al., 2005).

#### Privacy Concern is Impacted by Demographics

Education has been shown to be inversely related to privacy concern (Zukowski and Brown, 2007). The more educated the user less they scored on the IUIPC. In this same study, the Internet concern was shown to be directly proportional to age; older users were significantly more concerned than younger ones (Little et al, 2011). Concern For Information Privacy (CFIP) has been found to be related to the amount of privacy regulation in one's country (Bellman et al. 2004).

Surveys have also found males were more likely than females to clear their browsing history on mobile devices. (Boyles et al. 2012) The same survey also found that

people with no college education at all, were more likely to have never decided not to install a mobile application based on personal information concerns (45% of high school graduates or less vs. 57% for those attending some college and 60% for college graduates).

#### Privacy Behaviors in Mobile Devices

Other researchers have used the Android market as a target domain for research. For example, applications were analyzed in order to look for permissions associated with malicious software (Teufl et al., 2012). Although in the aforementioned study, the researchers used "various websites" and an unofficial Application Program Interface (API) package to gather their information, and they didn't disclose any details on the specific methods involved. Another instance is in the Federal Trade Communication (FTC) report examining Children's privacy on smartphones (FTC, 2012), which sampled and manually examined 400 applications.



## **Chapter 3: Study Design**

The shortcomings of not having data for actual user behaviors could be surmounted by ascertaining the choices of a large number of users via publicly available data. In order to study this, the platform's application store itself was selected to determine whether privacy and security data (what can be shared with others), was taken into consideration when purchasing applications.

To examine the privacy behaviors of users, their actions need to be observable in some way. The various application stores handle application permissions in different ways. Two popular stores are the Google Play store and the Apple iTunes App store. In the Apple App store, permissions are requested at run-time as opposed to the Google Play store, which requires permission acceptance before download or purchase is possible. Because the Google Play Store requires acceptance of the concessions at download, its downloads can be seen as acceptance of the terms, giving us visibility into the privacy decisions of the users.

### **Benefits of Choosing Google Play**

There are many advantages to picking the Google Play marketplace as a source of data. One of the largest advantages is the availability of the data itself; this type of data is often proprietary, or hard to obtain. Google provides data on applications to users of the Google Play marketplace in order to aid them in making an informed decision. But this data has many different dimensions which may prove interesting in looking at the relationship between people's behavior and the privacy requirements of a given application. Two big features which would indicate people's behavior is that

download count of an application and the rating of an application. This will tell us much about aggregate behavior that we could not get by looking at the micro scale.

Another advantage is that this collection suffers less than most from sampling bias by casting a wide net. When looking for applications to install from the Google Play marketplace for the first time, if users do not know the exact application that they want and are looking for an application, the domain of their search could include all of that can be browsed applications on the Google Play web site. This sample encompasses all of the applications that a user can see by browsing the Google Play web site by category. Because of the complete nature of the data set, we can make stronger statements on the specific domain. The fact that it is a census this rather than a sample of this domain means that we do not have to account for sample errors including response bias, coverage bias, and selection bias.

Yet another advantage to using the Google Play store for the sample is that, as an instrument, it is very accurate. The most commonly accepted definition of accuracy is the similarity between the measured value and its true underlying quantity, that that is, the closeness between measured reality and truth. As the sales statistics and other facets of applications in the Google play store are a direct representation of their true values, barring some unseen manipulation behind the scenes, the numbers examined in this study will perfectly represent truth. For example, a 4.5 star rating on an application is its real rating, not a sample or measurement of it.

These advantages would be irrelevant if the users were not representative of all smartphone users. The Android is the most common smartphone (Smith, 2013) and the Google Play store is the default means of app download. Furthermore, the demographics are largely similar between Android and iPhone users (who combined make up the large majority of the smartphone market) (Smith 2013). While there is some survey data to suggest that Android users may be more privacy aware, this potential difference which will be addressed in the Discussion Chapter (Benenson et al., 2013). Furthermore, there has not been data to suggest that the privacy-related behavior of Android users is any different. With these facts (and caveat) at hand, it is reasonable to claim that the users of this group are representative of smartphone users as a whole.

### **Tradeoffs**

Although there are many benefits to using the Google Play data, there are also some downsides. The first downside is that although it is a census, it is only a snapshot in time. A consequence of this is that it is difficult to say how downloads were affected by past actions. For example, the permissions requested could change in the form of an update, thus an application could have gained popularity with fewer permissions than it currently has.

Another downside is that the data is on the applications, not the users. It would be nice to see how users react to different programs or to see their individual purchasing behavior. Instead, we have to settle for analyzing how all users who have seen an

application react to it, in terms of whether they download it or not, and how they rate it.

### **Domain Research**

Other researchers have used the Android market as a target domain for research. For example, applications were analyzed in order to look for permissions associated with malicious software (Teufl et al. 2012). although in the aforementioned study, the researchers used "various websites" and an unofficial API package to gather their information , and they didn't disclose any details on the specific methods involved. Another instance is in the FTC report examining Children's privacy on smartphones (FTC, 2012), which sampled and manually examined 400 applications.

### **Demographics**

### ***Objectives***

There are several ways that one can ascertain the relationship between the privacy concessions requested by an application and the popularity of the application. One straightforward way is by analyzing the relationship between the number of privacy concessions and the number of downloads. Yet another approach would be to analyze applications with an extreme number of concessions to see what sets those applications apart, and to qualitatively analyze the differences that might contribute to its popularity. In order to control for the different needs of different applications, analysis of the effects of individual concessions are examined within categories. This is important because the necessary features may vary between categories and be confounded with privacy concessions. For example, one might expect contacts list modification to be necessary in a communications application, but not necessary in an

action or arcade game. In order to see if this confounding factor has an effect, categories will be analyzed separately and in aggregate.

### Methodology

For the purpose of evaluating the impact of privacy on preference, the purchase of an application, and subsequent downloads, was used as an indicator of preference (as people must spend money, and thus, on average think about a purchase, as opposed to downloading a free application which might require less thought). Paid applications, therefore, are the only applications analyzed. The downside to choosing Paid applications is that most apps are free (FTC, 2012). Teenagers have said that they prefer free applications since they don't have to get their parent's permission (Madden et al. 2012). Despite this, since people value things they are willing to pay more for, the salience of the differentiating attributes (ie privacy permissions) are likely increased in paid applications. What is intended is that after this study the effect of the privacy permissions required by an application on the likelihood that someone will purchase this application can be determined.

All of the data was recorded from the website of the Google Play Application store (URL: <https://play.google.com/store/apps>). On this site, users can find paid apps under each category. There is, at the time of data collection, a maximum number of applications that are visible to the user without searching specifically for an application as the website only displays the first 20 pages of applications in any given category. At the time of data collection, a browsing user can browse, at most, approximately 480 applications per category. These 20 pages of applications were

browsed, and the hyperlinks linking to the apps themselves were navigated to, via a simple Python Script. The pertinent data for each application at this address was collected, and stored on a local file for aggregation and later statistical analysis. The file was stored as a CSV format, which needed to be imported into R for analysis. R was chosen as it has the necessary analysis modules, is an established piece of analysis software, and is freely available.

### Analysis Techniques

It has been argued that hypothesis testing is often misused in order to lend results scientific credibility (Cortina and Landis). Keeping in line with the main research goals, this study will make use of standard descriptive techniques and charts to portray the distributions in a manner most fitting a study of this kind. Although there is no hypothesis testing performed, in order to compare the results of this exercise to that of other studies, a Bonferroni correction on a standard level of acceptance can be done for reference (e.g. if 100 concessions are found in multiple applications, the standard acceptance of  $\alpha=.05$ , would require an individual test threshold of 0.0005). Any good scientific observations have to be able to make predictions. As such, although the samples in this study represent the entire population of applications described earlier, for the purposes of quantifying errors and determining significance, the samples will be treated as part of a population of potential programs. The primary intent of the analysis is to understand the effects of the concessions on the download/purchase rates. Because of this, of particular interest is under which conditions the average number of downloads is likely to have changed. However, the

conditions in which the shape of the distributions may have changed are also of interest. With this in mind, the linear regression and F-Test (used in analysis of variance, ANOVA) use the following assumptions:

**Independence:**

That the results of one measurement do not affect another is self-evident from the study design. The study is done as a survey of data, with no experimental treatment; additionally, it was done rapidly, as to dampen temporal effects.

**Identical distribution (Homogeneity of variance, and normality)**

*Homogeneity of variance:* Each comparison has to have its variance checked.

However, as this is not a small sample but a census of a domain, the differences in variances can also be looked at as an interesting result, and indeed a possible difference between the groups, precisely what we are looking for.

*Normality:* Mitigated by a large sample size, where this fails, useful relationships can still be seen or differences due to the failure.

**ANOVA**

The ANOVA is used when one needs to analyze the differences in group means.

The appropriateness of ANOVA is determined by the assumptions. The intent is to use an ANOVA to determine if there is any relation between the presence of individual concessions and the number of downloads. Because the download count is interval censored, the categorical range will be used as a basis of comparison. This will understate the relationship, as the categories have more of a geometric

relationship (with the endpoints going 5, 10, 15, 20, 50, 100 etc.). This has the effect of being conservative, but is deemed appropriate due to the uncertainty of the actual values. This is further applied to the individual categories in order to determine whether or not the individual concessions have differing effects when applied to different types of application.

## **Linear Regression**

The linear regression is useful when describing the relationship between continuous (i.e. non-categorical) variables. The usefulness of a linear regression is bound by the existence of several assumptions: a linear relationship between the independent and dependent variables, independence of errors and constant variance of errors.

## **Dealing with Censoring of Variables**

Censored Data is continuous data whose value is partially hidden. In this case, the most important hidden data is that of the download count, which is hidden beneath an interval (range of possible values). There have been many studies examining how to deal with this unfortunate circumstance. The interval censored data will be analyzed using a parametric survival regression model. This method is implemented using the Survival Module (commonly used for survival models), as described in “Tutorial on methods for interval-censored data and their implementation in R” (Gomez et al., 2009).



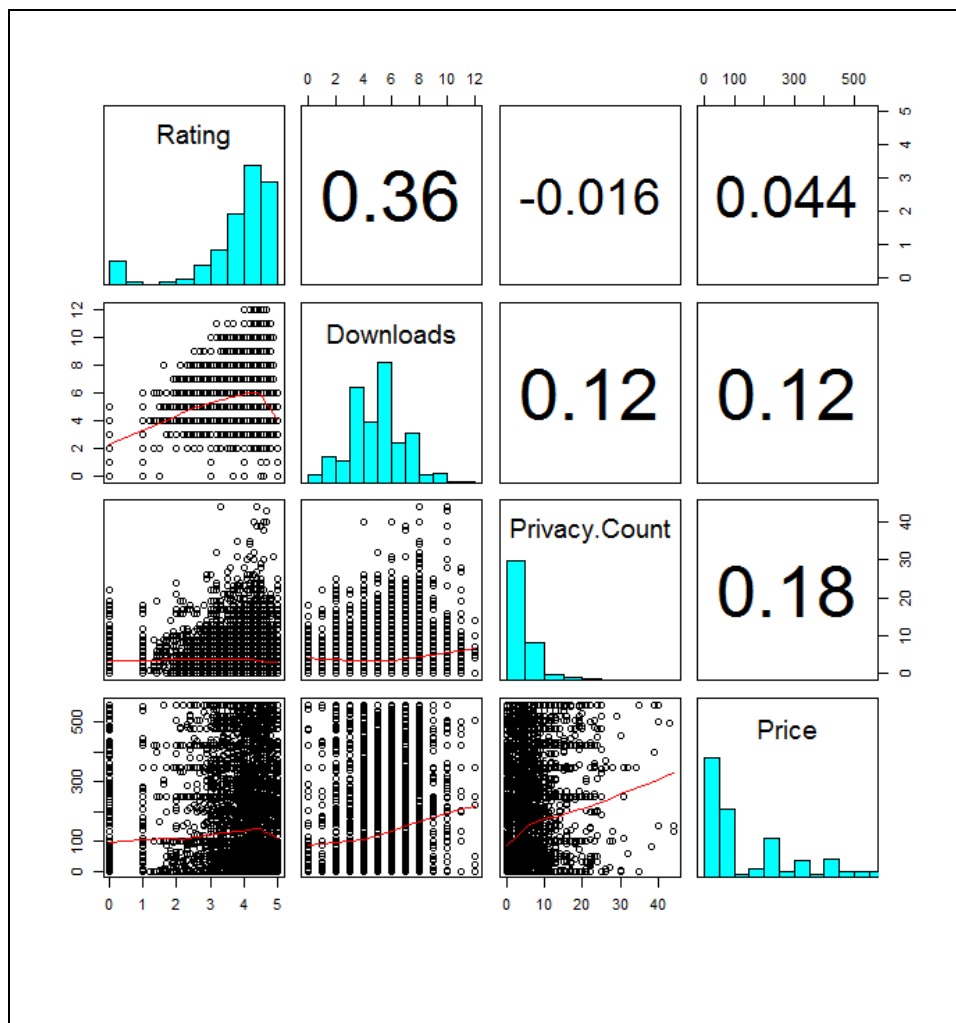
## Chapter 4: Findings

### Domain Summary

The domain that we are looking at is the Android applications that users of android can find by browsing and can buy for money. In this domain, there were 8 categories of games, and 26 categories of applications, which contained a total of 11777 applications.

### Key attributes

Table 1 Summary of Domain

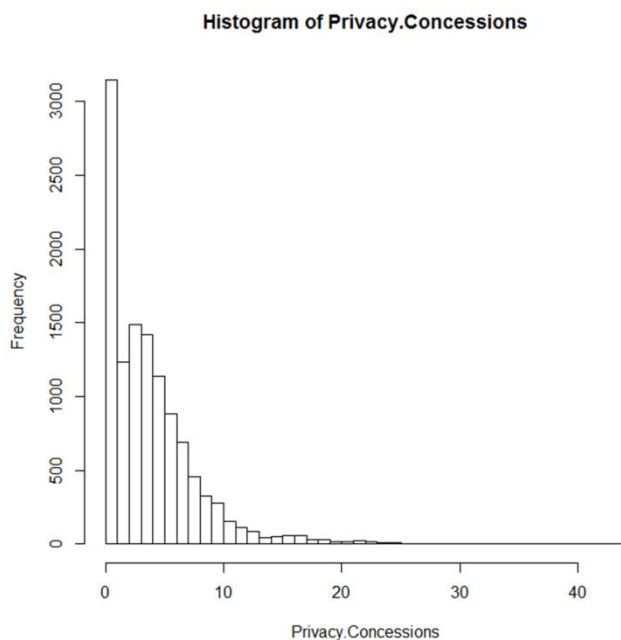


The number of downloads of each category is presented as a range. There are 13 such ranges(1-5, 5-10, 10-50, 50-100, 100-500, 500-1K, 1K-5K, 5K-10K, 10K-50K, 50K-100K, 100K-500K, 500K-1M, 1M-5M). Each application also has a list of privacy permissions that are required for the application to run. In order to determine the effect of the privacy permissions required by an application on the likelihood that someone might purchase this application, the number of privacy settings are compared to the download count.

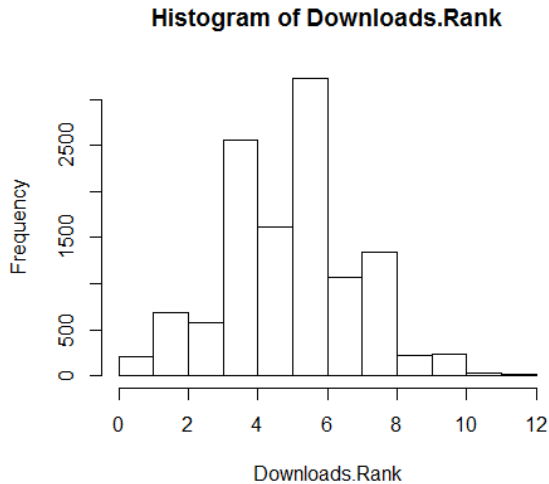
### Data Summary

The number of privacy concessions required by applications is not normally distributed, as seen in the figure below. When dealing with the entire domain, we can be reassured that the skew is not likely a result of selection bias. The mean number of privacy concessions is 4.26, while the median is 4. The standard deviation is 4.16.

**Table 2 Histogram of Privacy Concessions**



Download rank is fairly regular, however given the nonlinearity, the actual download rates are less so.



**Figure 1**Download Distribution

### Quantitative Analysis

In order to determine the relationship between the number of downloads and their rating, and the number of downloads and the number of privacy concessions, the survival package was used in order to come up with a continuous scale of download count from the interval data collected.

In order to do this, a distribution must be assumed. The distribution assumed in this case is log-logistic. The survival package estimates the significance of the model, which in this case is difficult to interpret ( $p=0$ ). In order to interpret whether  $p=0$  meant that the model was extremely significant or insignificant, the regression was recreated with three successive subsets of the data. The regression was done on the first 50 data points, and the model was found to be insignificant (with 2 degrees of

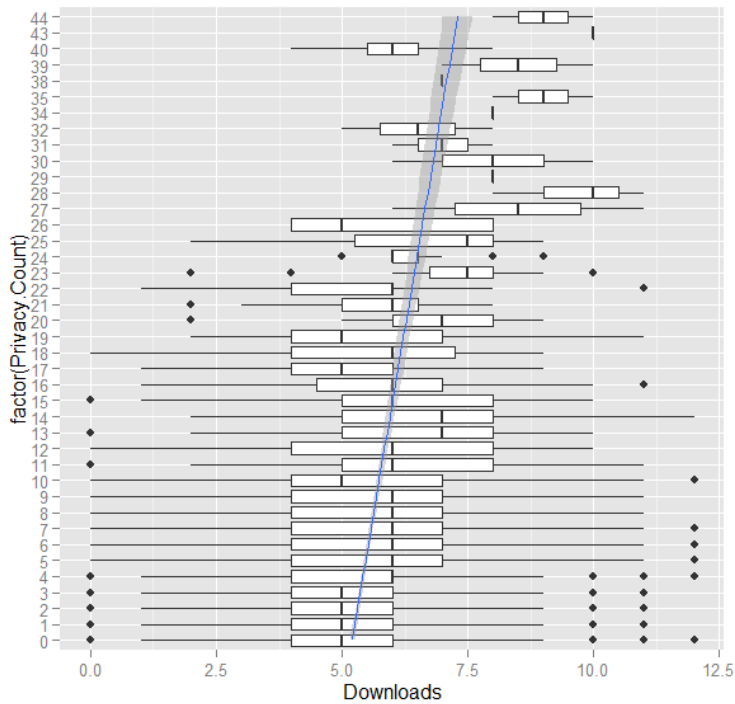
freedom,  $p= 0.58$ ). The regression was then done on the first 500 data points, and the model was found to be significant (with 2 degrees of freedom,  $p= 4e-13$ ). The regression was then done on the first 5,000 data points and the model was found to be  $p = 0$ . Given this trend, the model is interpreted to be extremely significant when applied the full sample.

This package produced a regression with some hard-to-interpret interval regression coefficients. Privacy concession count's relationship to the number of downloads was significant ( $p < 1.20 e-59$ ), while user rating's relationship to the number of downloads showed a p value of 0. In order to interpret this, the regression was recreated with three successive subsets of the data. The regression was done on the first 50 data points, where there was an insignificant relationship between rating and downloads ( $p = 8.19 e-01$ ). The regression was then done on the first 500 data points, and there was a significant relationship between rating and downloads ( $p = 4.44 e-15$ ). The regression was then done on the first 5,000 data points and there was a significant relationship between rating and downloads ( $p = 1.78 e-211$ ). Given the trend of these significance values, taken with the fact that the bigger the size of the sample, the smaller the p value, the p value of 0 is interpreted to be extremely significant.

Each of these relationships is an independent effect, so that we can see that there is a significant effect of privacy count when adjusting for rating and a significant effect of rating on download count when adjusting for privacy count. A relationship was not able to be established between rating and privacy count.

## Privacy Concession Count to Number of Downloads

The download count (the coded, 1-12, interval) is positively correlated to the number of privacy concessions. The correlation is weak (0.053530), but highly significant (two tailed  $p < 2e-16$ ), with a standard error of 12.59.



Coefficients:

Estimate Std.

**Figure 2 Privacy Count Across Downloads**

Error t value Pr(>|t|)

(Intercept) 5.197540 0.025309 205.36 <2e-16 \*\*\*

Privacy.Count 0.053530 0.004253 12.59 <2e-16 \*\*\*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.919 on 11775 degrees of freedom

Multiple R-squared: 0.01328, Adjusted R-squared: 0.01319

F-statistic: 158.4 on 1 and 11775 DF, p-value: < 2.2e-16

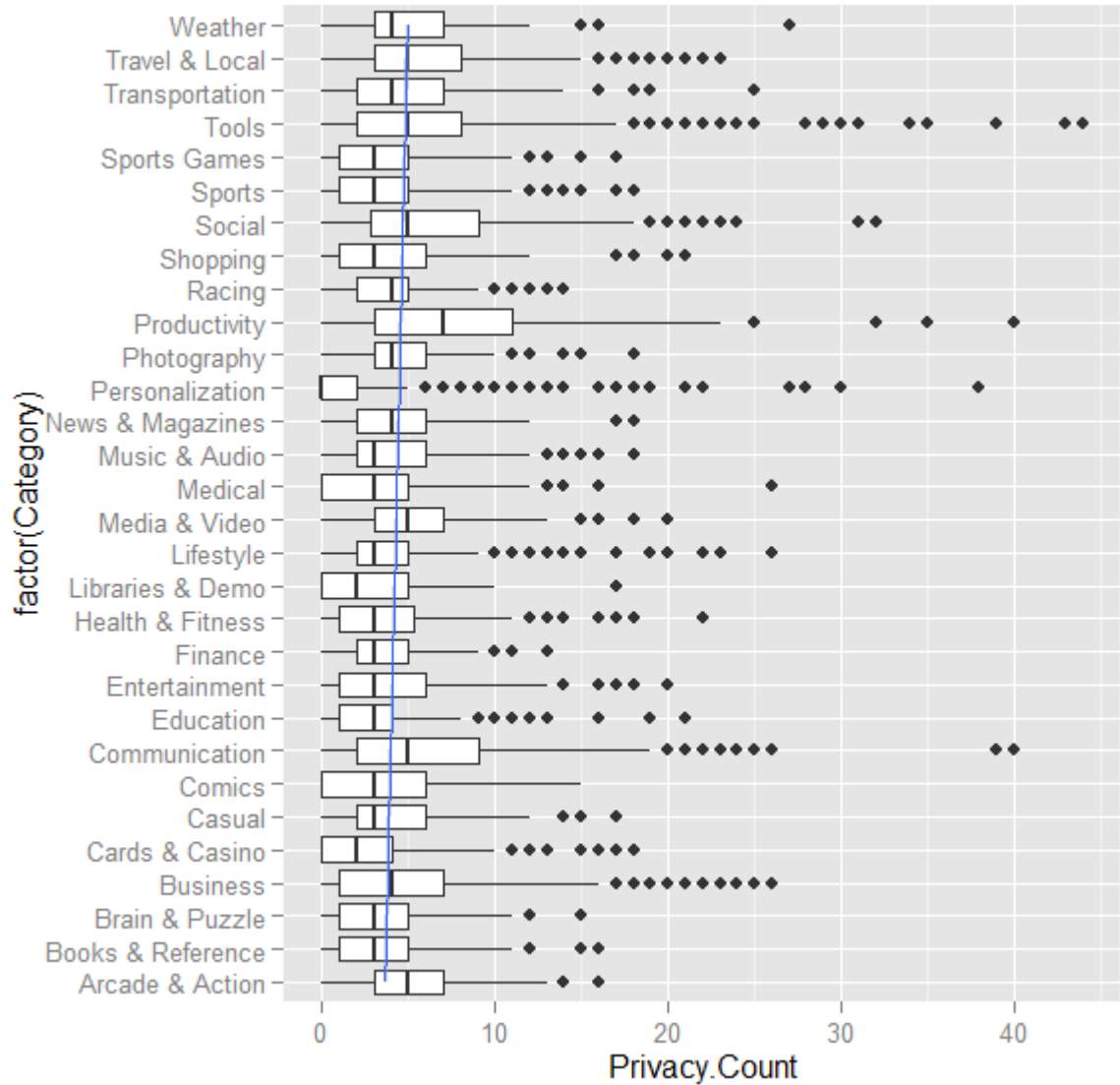


Figure 3 Privacy Count by Category

### Individual Concessions

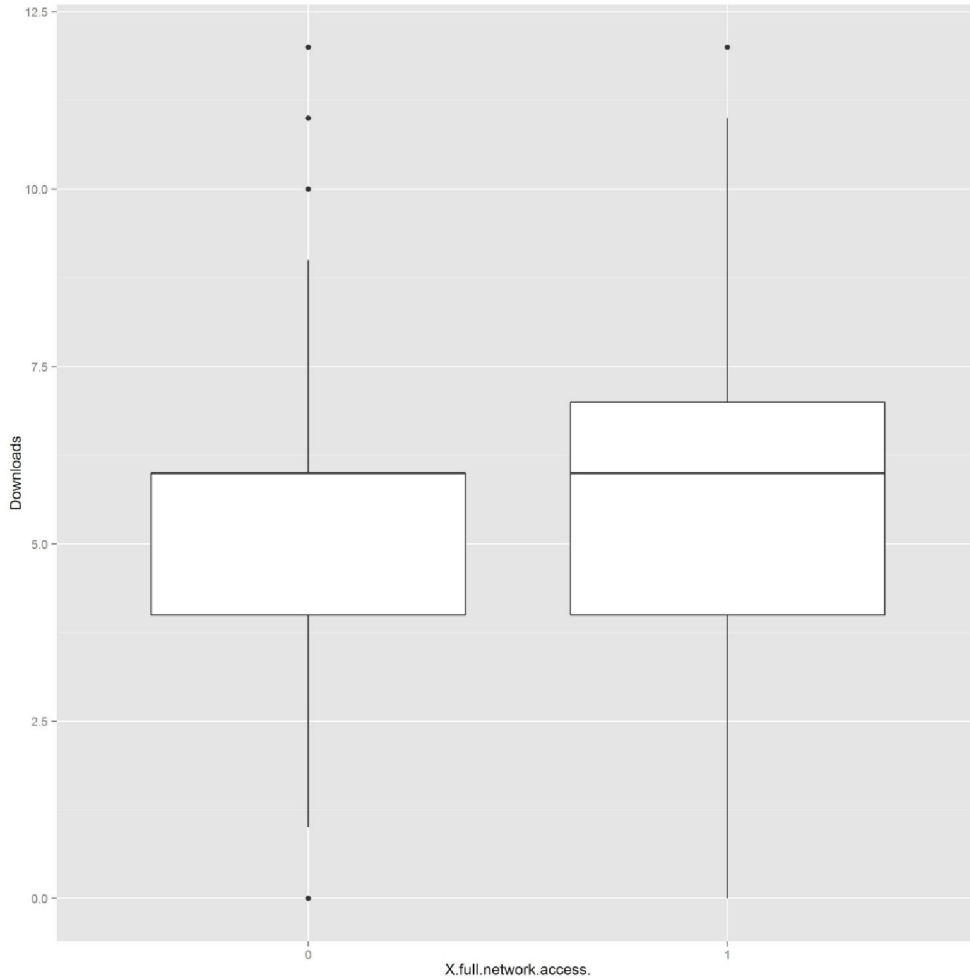
In order to determine whether the presence of any of the concessions had significant impact on the download rate, an ANOVA was first run with a Tukey-hsd post hoc

test. There were many highly significant impacts of concessions on the number of downloads, however they were all positive relationships. When analyzed as a whole, each individual concession was revealed to either have no effect on the number of downloads, or was found to increase the number of downloads. Specifically, the following relationships were found (with 137 concessions found in applications, a Bonferroni correction reveals that the standard acceptance of  $\alpha=.05$  would require an individual test threshold of .0003). The following had a significant positive relationship to download count: full network access ( $p<0.0001$ ), read phone status and identity ( $p<0.0001$ ), modify or delete the contents of your USB storage modify or delete the contents of your SD card ( $p<0.0001$ ), view network connections ( $p<0.0001$ ), view wi-fi connections ( $p<0.0001$ ), test access to protected storage test access to protected storage ( $p<0.0001$ ), prevent phone from sleeping ( $p<0.0001$ ), connect and disconnect from Wi Fi ( $p<0.0001$ ), run at startup ( $p<0.0001$ ), retrieve running apps ( $p<0.0001$ ), control vibration ( $p<0.0001$ ), modify system settings ( $p<0.0001$ ), receive data from Internet ( $p<0.0001$ ), find accounts on the device ( $p<0.0001$ ), pair with Bluetooth devices ( $p<0.0001$ ), access Bluetooth settings ( $p<0.0001$ ), change network connectivity ( $p<0.0001$ ), read sync settings ( $p<0.0001$ ), close other apps, set an alarm ( $p<0.0001$ ), change screen orientation ( $p<0.0001$ ), control near field communication ( $p<0.0001$ ), read your own contact card ( $p<0.0001$ ), add words to user defined dictionary ( $p<0.0001$ ). There were other significant positive relationships: change your audio settings ( $p<.0117$ ), read sync statistics ( $p<.0008$ ), allow wi-fi multicast reception ( $p<.0103$ ). Additional,

relationships can be found in the Appendix. None of the additional significant relationships were in the opposite direction.

**Full Network Access:**

With an F value of 16.10, applications with full network access also have a larger range than those without, although the median download category is the same with

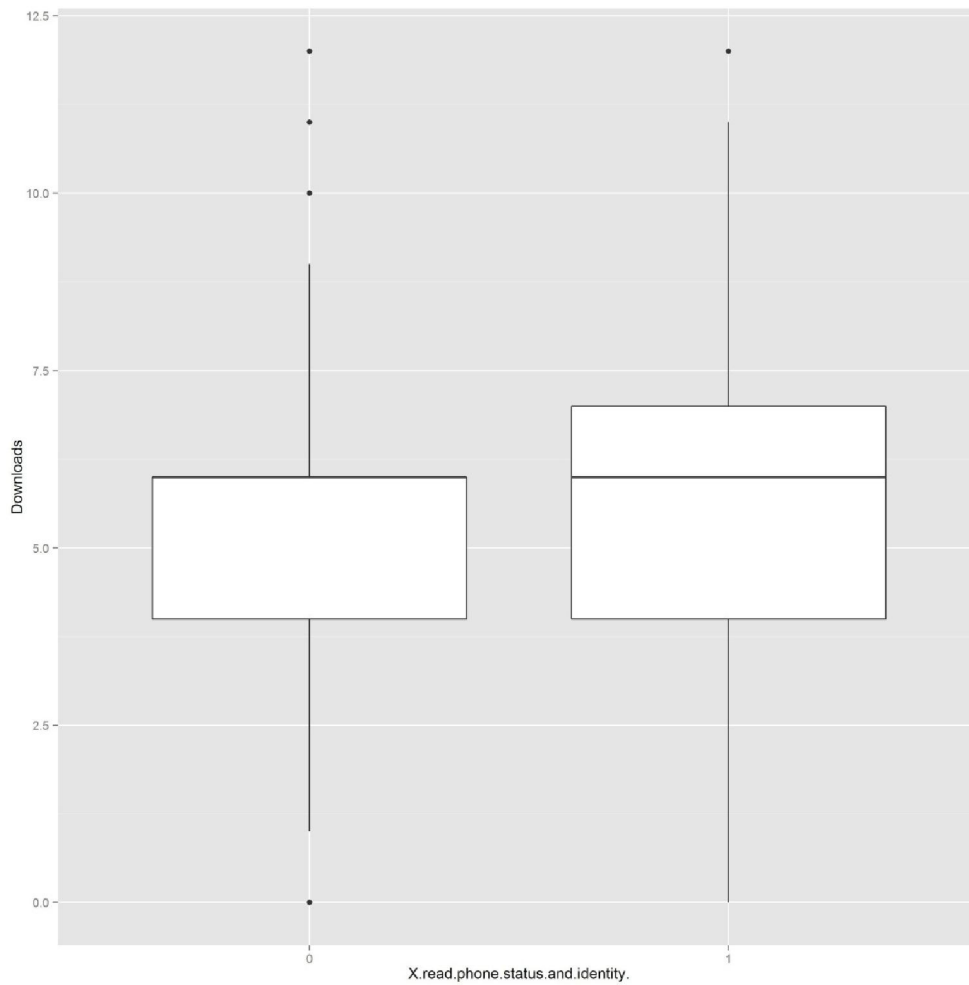


**Figure 4 The Effect of Full Network Access on Downloads and without this permission.**



### Read Phone Status and Identity:

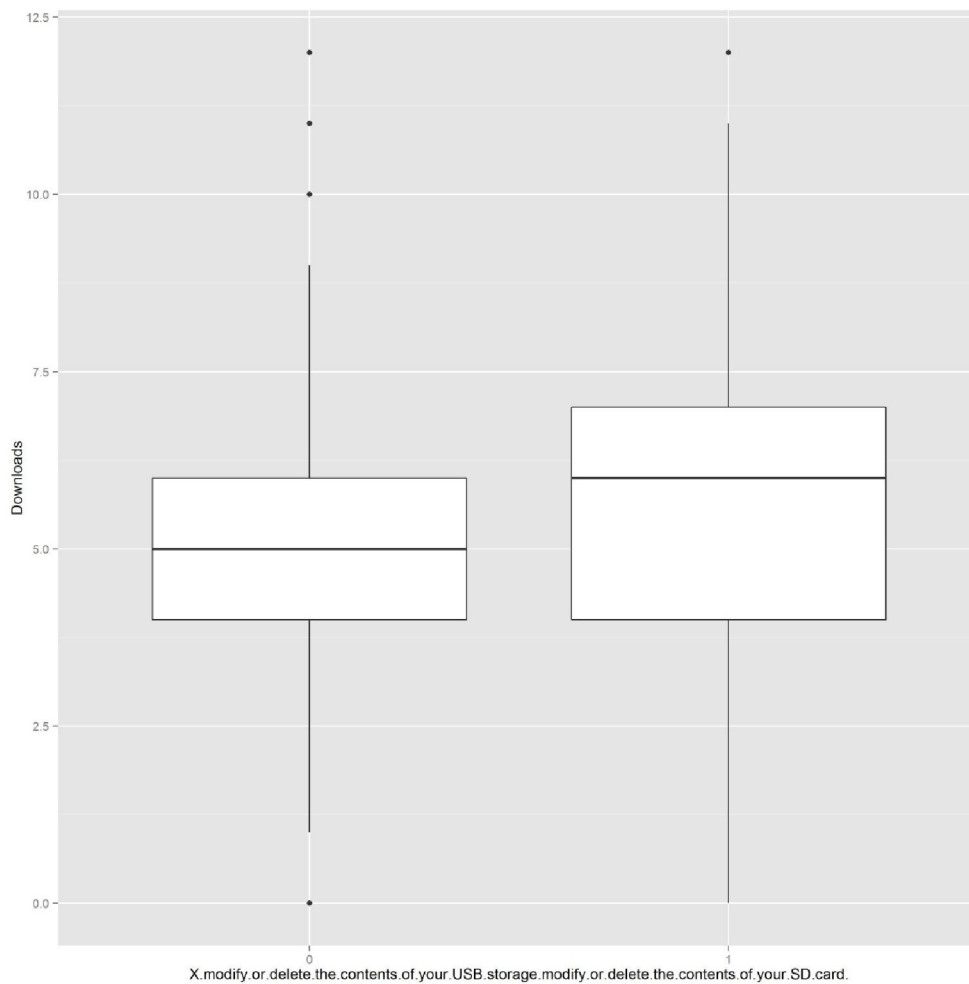
The F value of 29.50 seems to be an effect of applications with the read phone status and identity permissions having a less skewed distribution than those without the permission.



**Figure 5 The Effect of the Phone Status Concession on Downloads**

**Modify or delete the contents of your USB storage/modify or delete the contents of your SD card:**

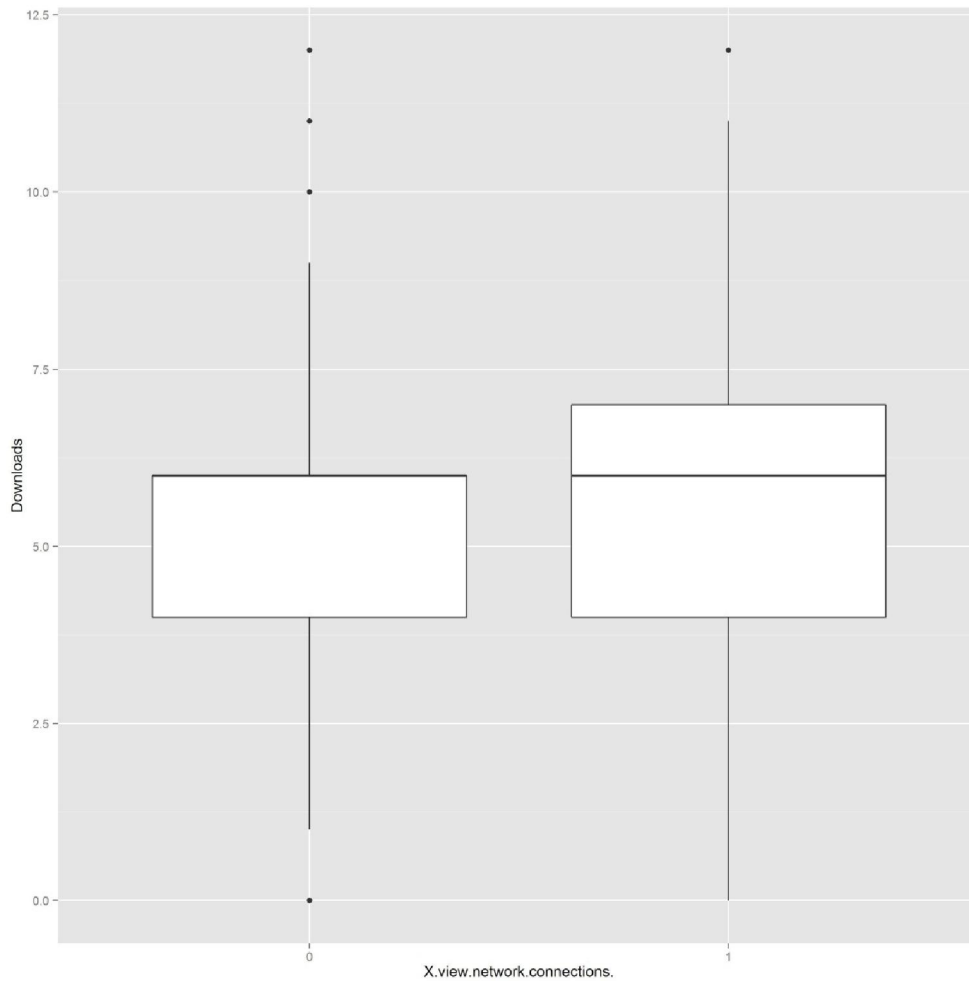
The F value of 22.95 seems to be an effect of applications with this permission having a more skewed distribution than those without the permission.



**Figure 6 The Effect of the USB Storage Concession on Downloads**

**View network connections:**

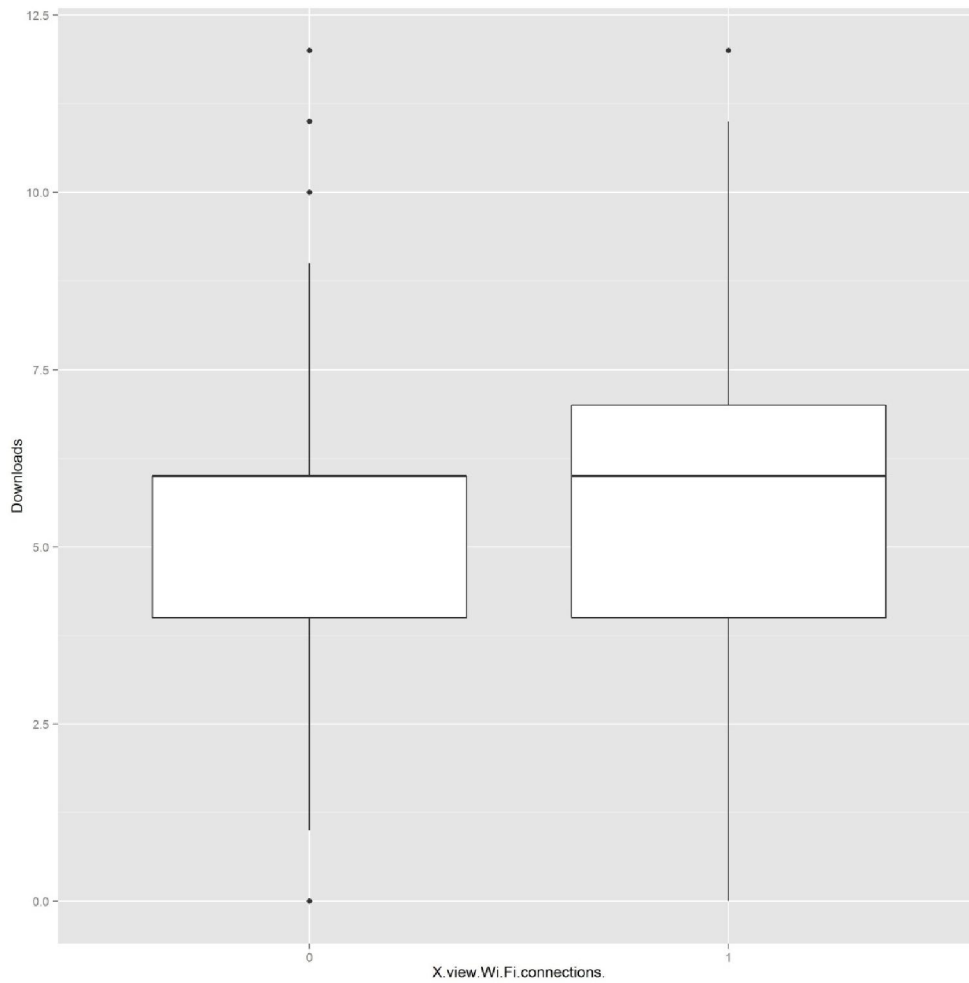
The F value of 23.39 seems to be an effect of applications with this permission having a less skewed distribution than those without the permission.



**Figure 7 The Effect of the View Network Connection Concession on Downloads**

### View Wi Fi connections

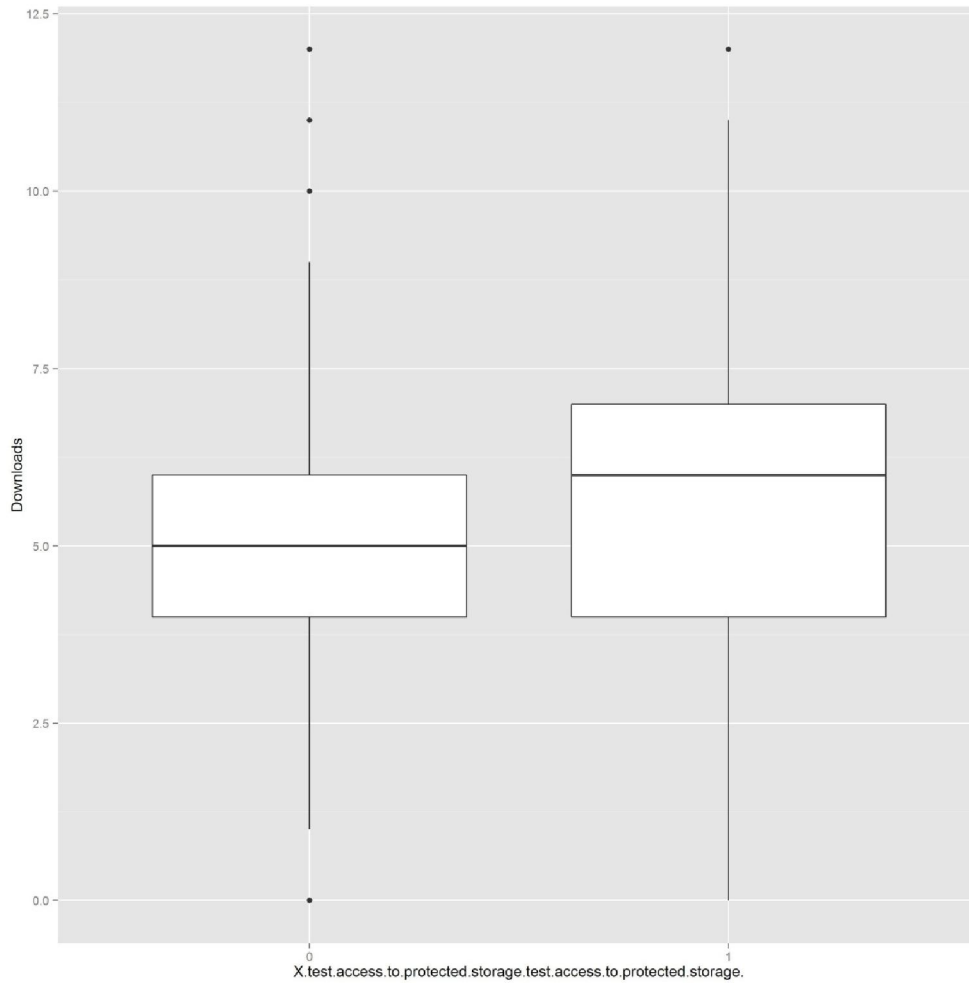
The F value of 71.01 seems to be an effect of applications with this permission having a more skewed distribution than those without the permission.



**Figure 8 The Effect of the View Wi-Fi Connections Concession on Downloads**

### Test access to protected storage test access to protected storage

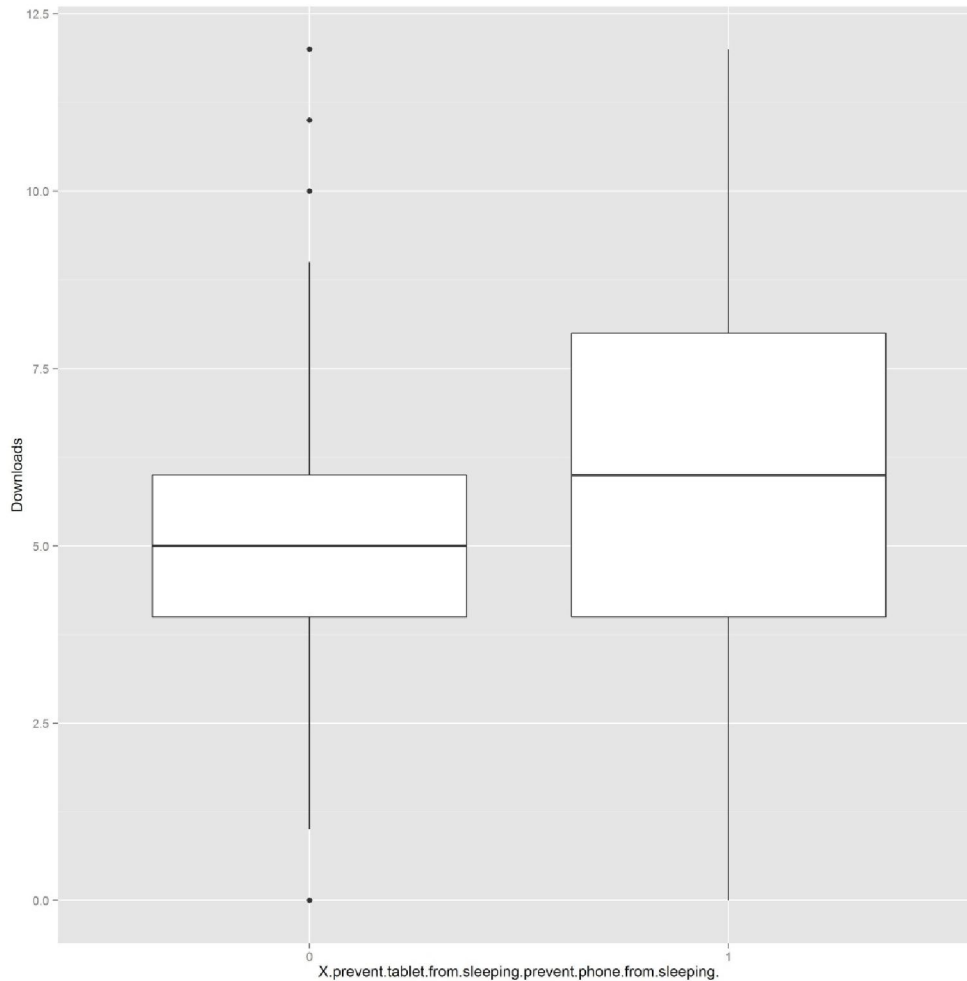
The F value of 24.47 seems to be an effect of applications with this permission having a less skewed distribution than those without the permission.



**Figure 9 The Effect of the Test Protected Storage Concession on Downloads**

### Prevent tablet from sleeping prevent phone from sleeping

The F value of 48.79 seems to be an effect of applications with this permission having a wider distribution, with a higher median than those without the permission.

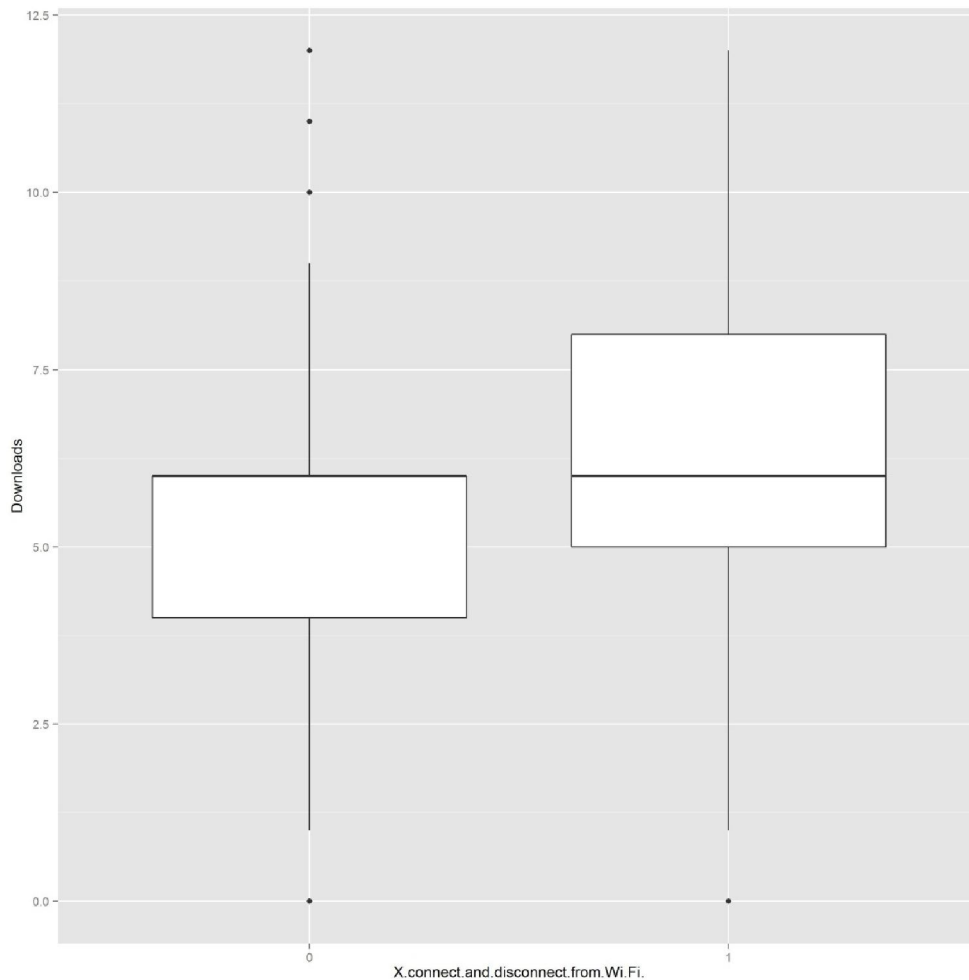


**Figure 10 The Effect of the Prevent Sleep Concession on Downloads**

### Connect and disconnect from Wi Fi:

The F value of 66.28 seems to be an effect of applications with this permission having a less skewed distribution than those without the

permission.

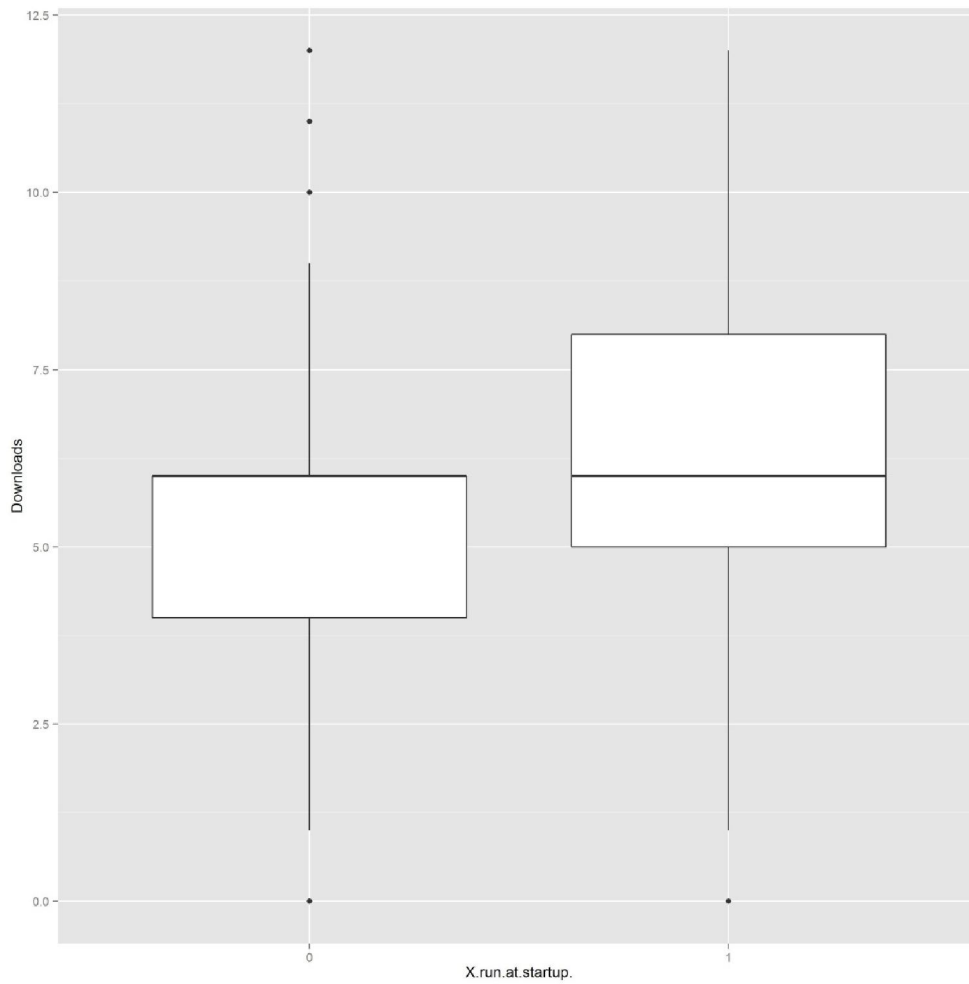


**Figure 11**The Effect of the Wi-Fi Connection Status Concession on Downloads

### Run at startup

The F value of 27.43 seems to be an effect of applications with this permission having a less skewed distribution, with a wider interquartile range than those without the

permission.

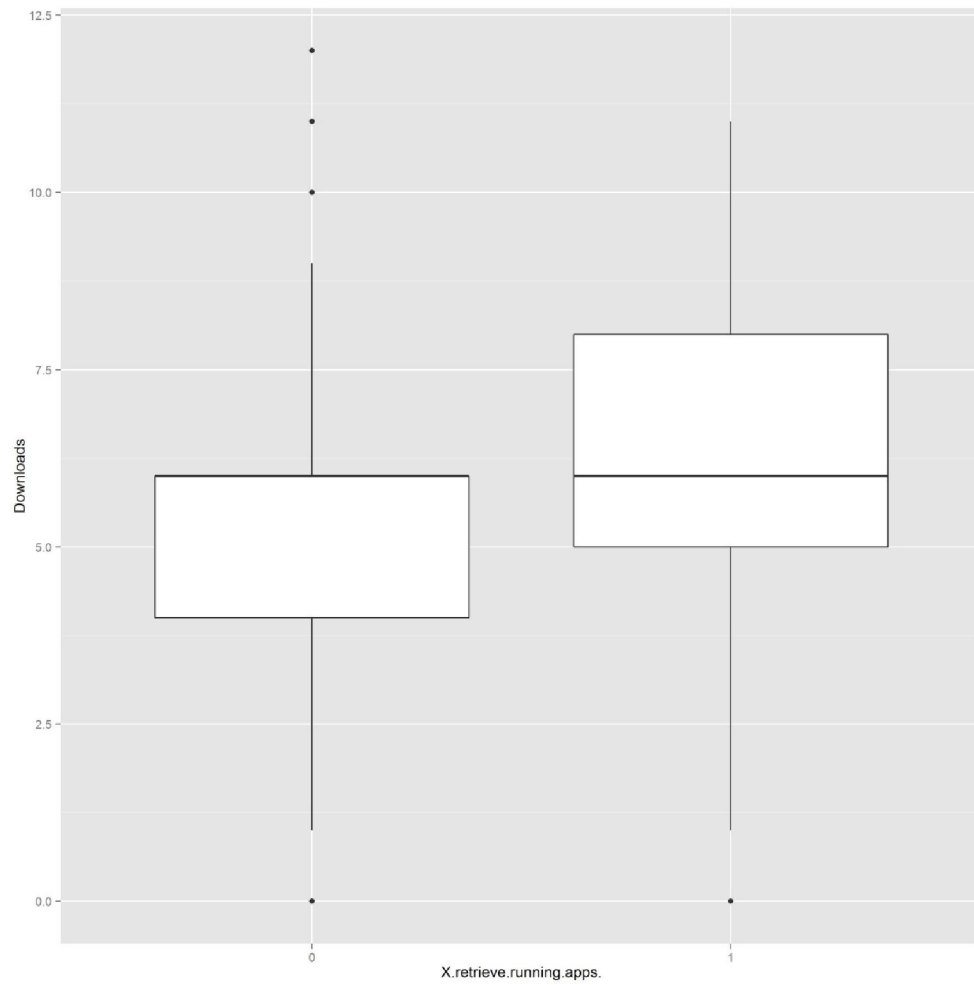


**Figure 12 The Effect of the Run at Startup Concession on Downloads**



## Retrieve Running Apps

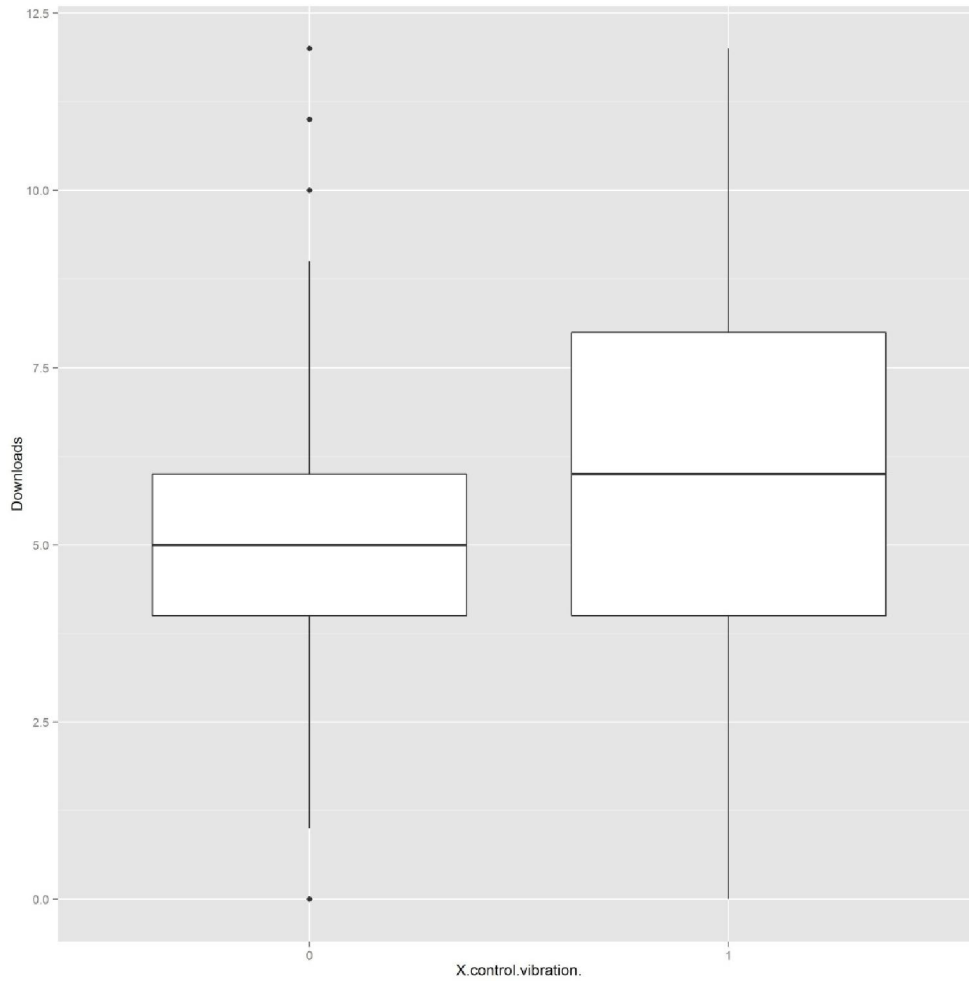
The F value of 14.55 seems to be an effect of applications with this permission having a less skewed distribution than those without the permission.



**Figure 13** The Effect of the Running App Retrieval Concession on Downloads

**Control vibration:**

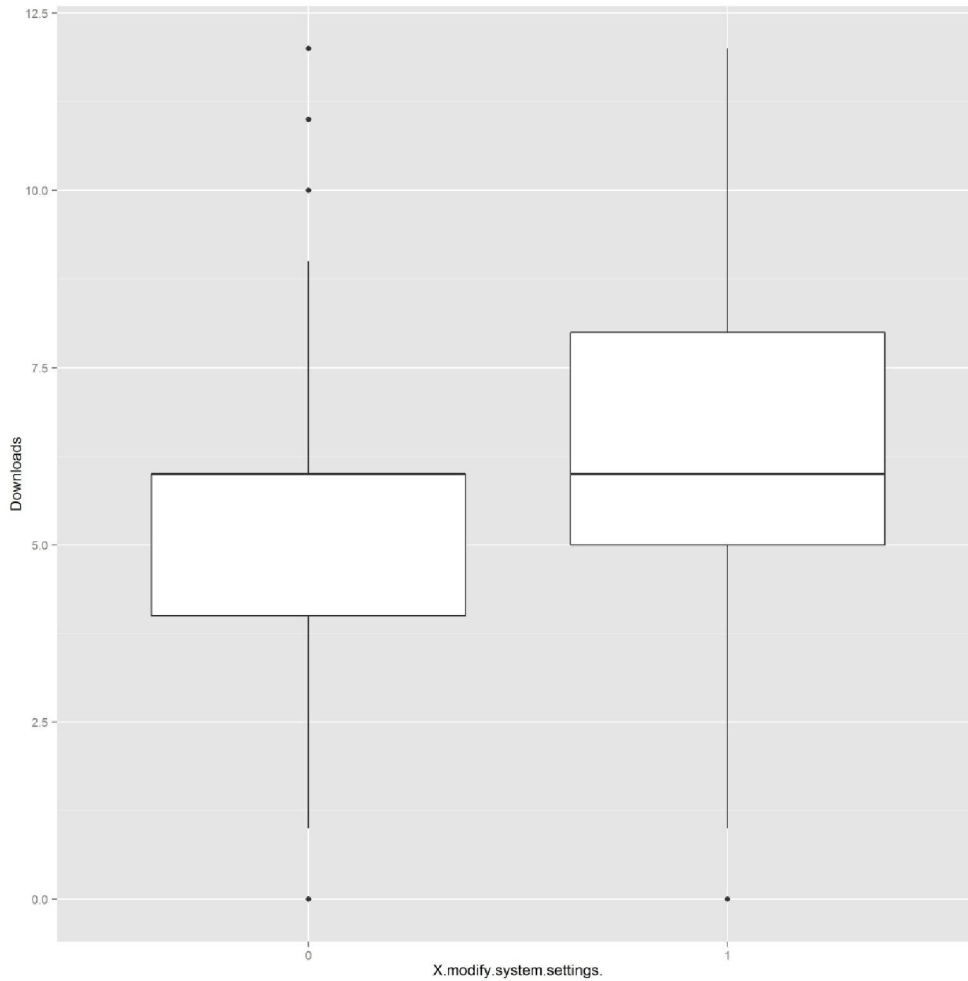
The F value of 32.48 seems to be an effect of applications with this permission having a wider distribution, with a higher median than those without the permission.



**Figure 14**The Effect of the Control Vibration Concession on Downloads

### Modify system settings

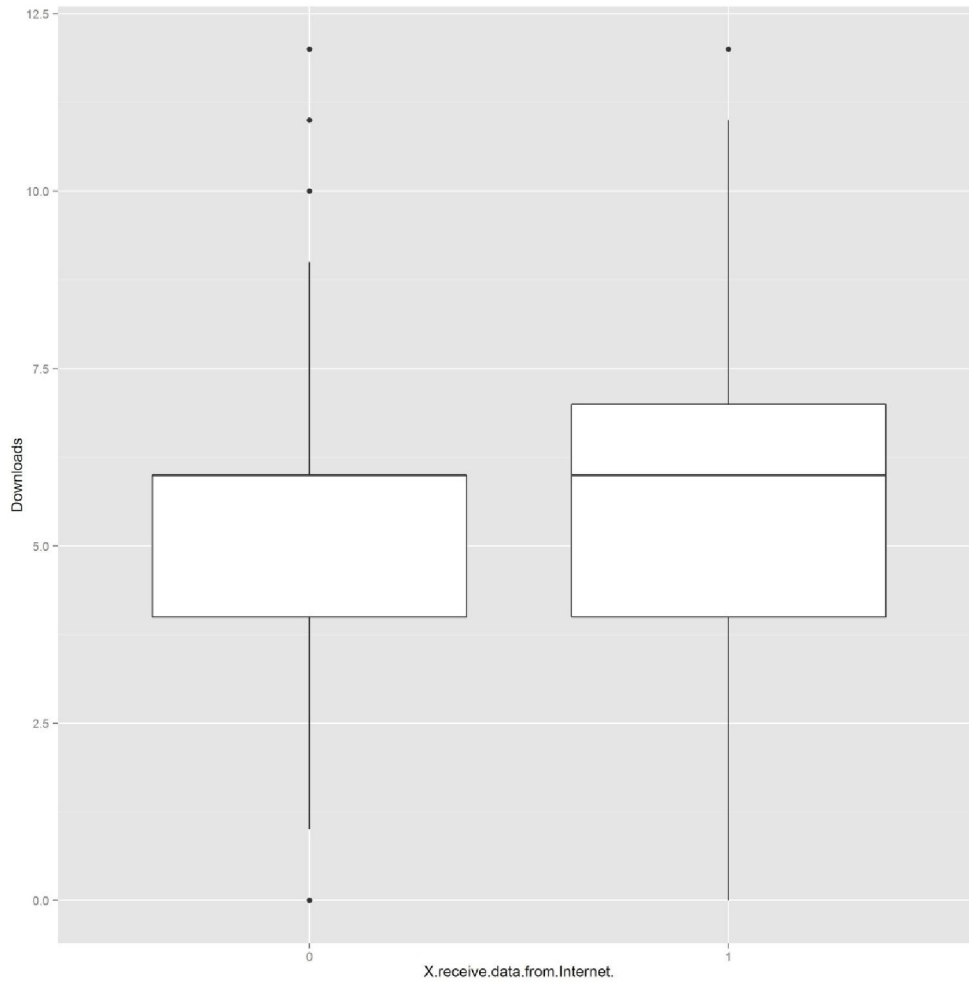
The F value of 32.48 seems to be an effect of applications with this permission having a less skewed and wider distribution than those without the permission.



**Figure 15 The Effect of the Modify System Settings Concession on Downloads**

### Receive Data from Internet

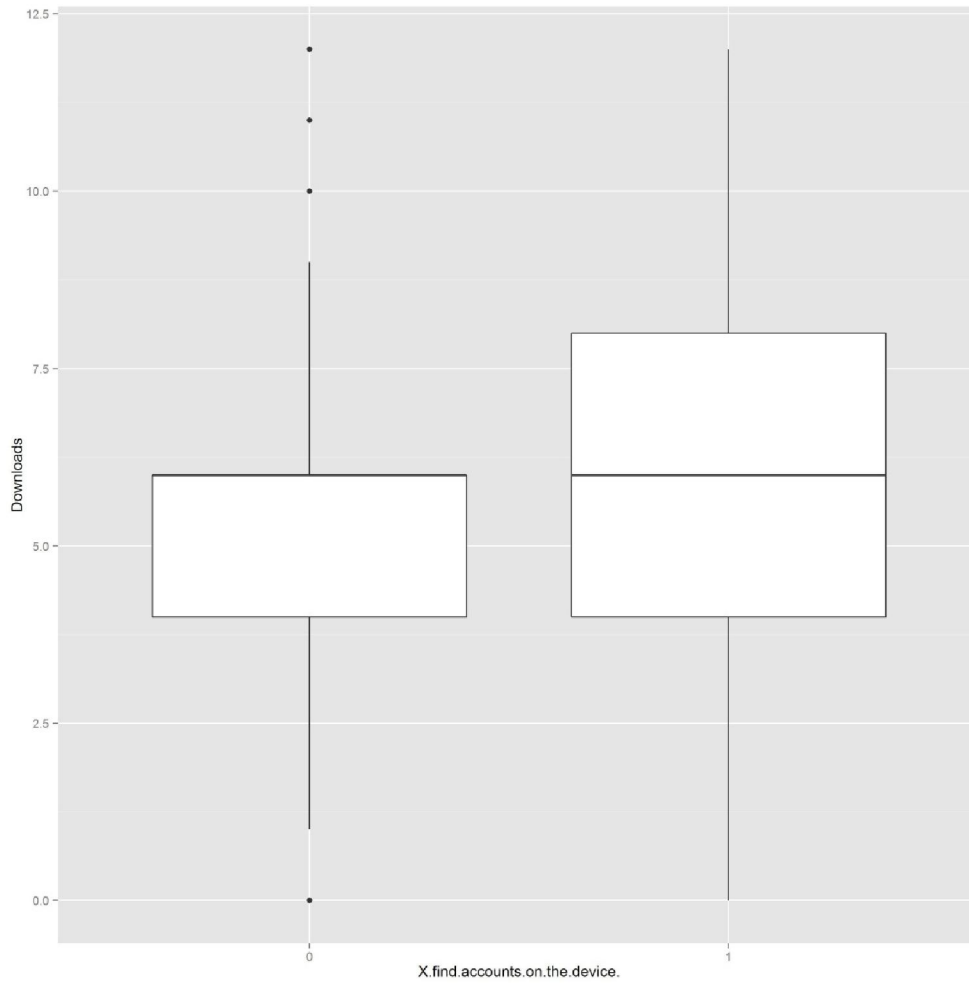
The F value of 27.42 seems to be an effect of applications with this permission having a less skewed and wider distribution, with an interquartile range translated upward, as compared to those without the permission.



**Figure 16 The Effect of the Receive Data Concession on Downloads**

### **Find Accounts on the Device**

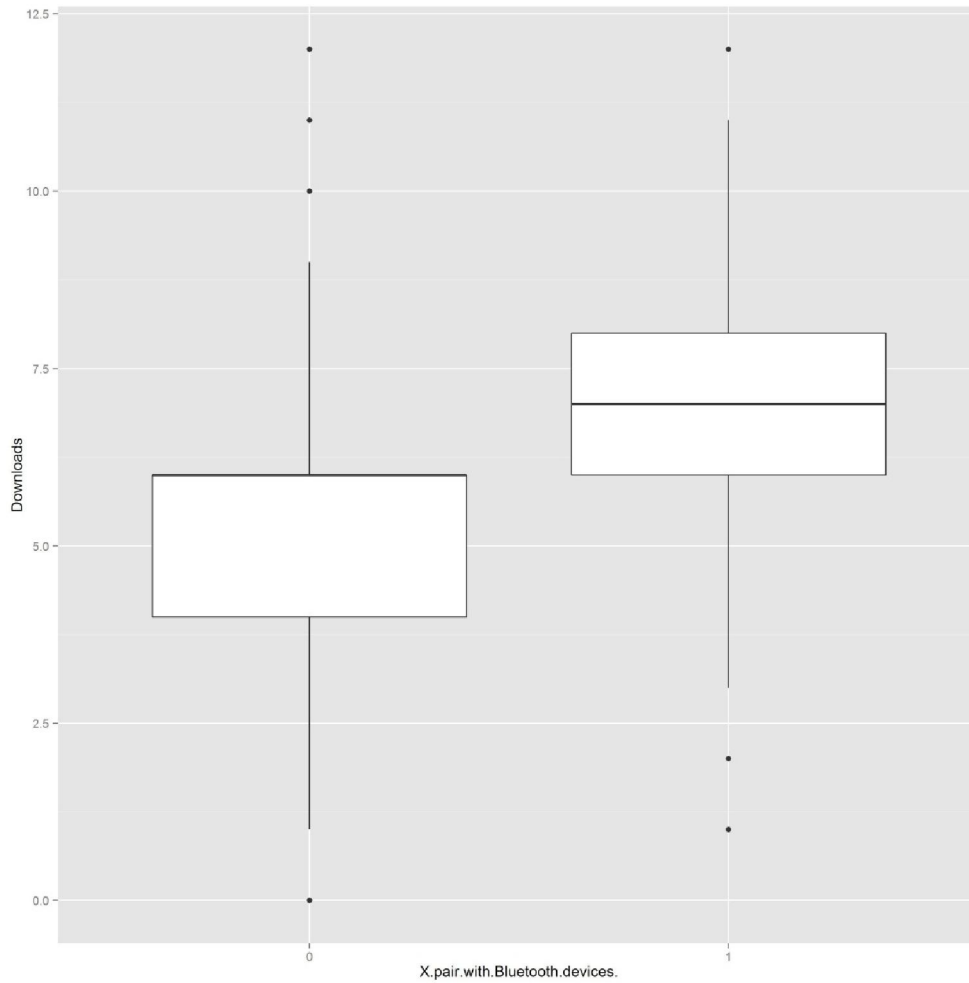
The F value of 22.67 seems to be an effect of applications with this permission having a less skewed and wider distribution.



**Figure 17 The Effect of the Find Accounts Concession on Downloads**

**Pair with Bluetooth devices**

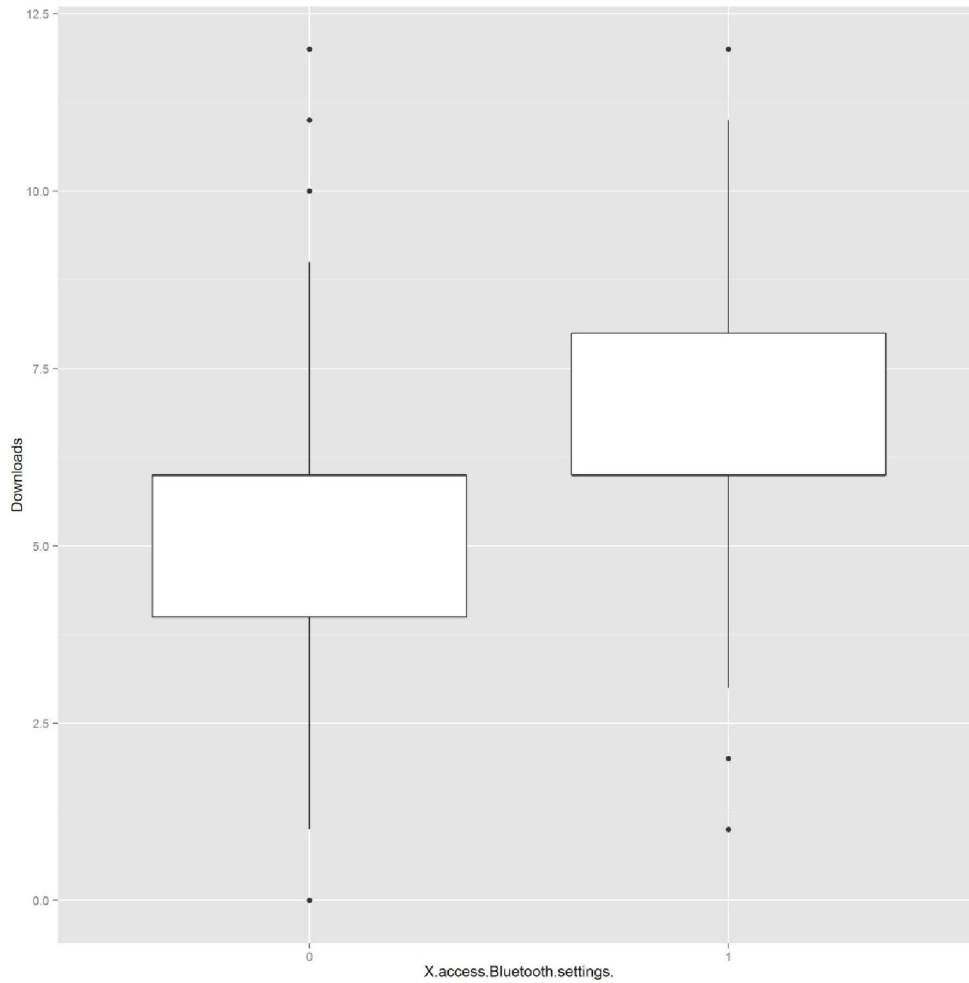
The F value of 34 seems to be an effect of applications with this permission having a less skewed and wider distribution than those without the permission.



**Figure 18 The Effect of the Pair Bluetooth Concession on Downloads**

### **Access Bluetooth settings**

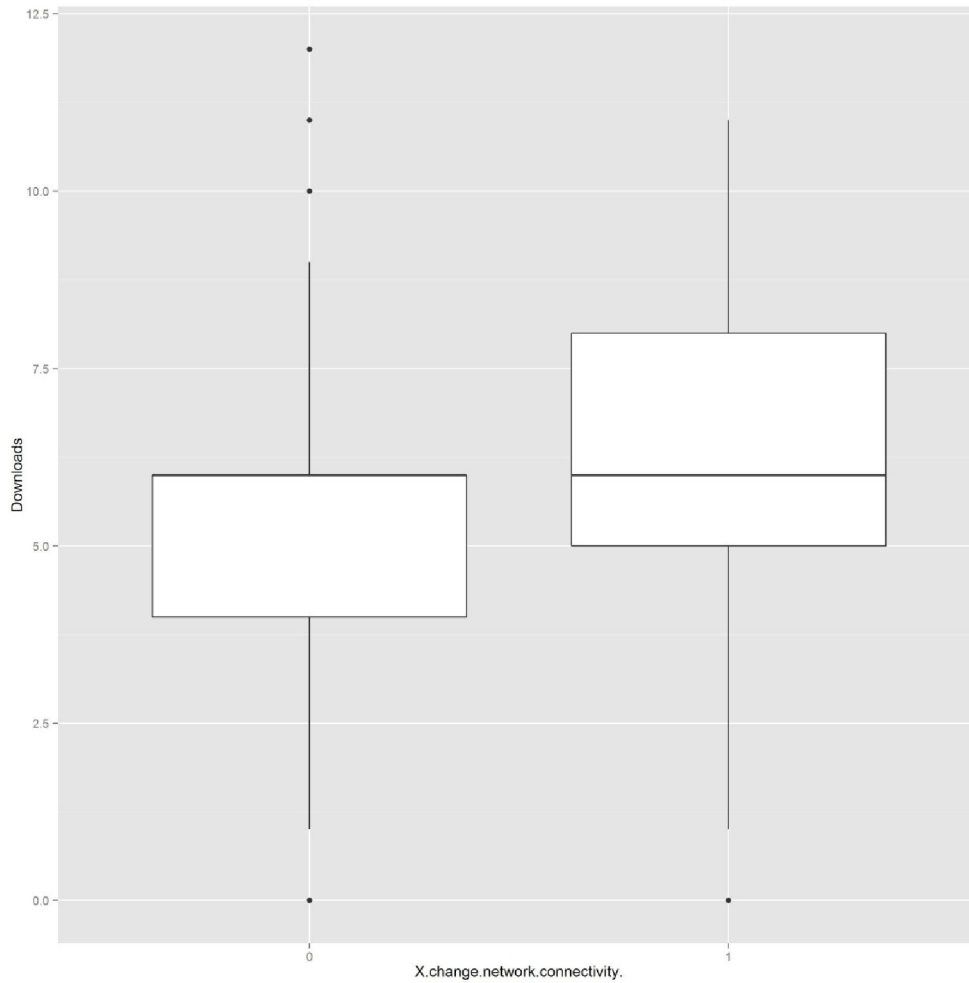
The F value of 23.28 seems to be an effect of applications with this permission having an opposite skewed to those without the permission.



**Figure 19 The Effect of the Access Bluetooth Settings Concession on Downloads**

### Change Network Connectivity

The F value of 50.79 seems to be an effect of applications with this permission having a less skewed and wider distribution than those without the permission.

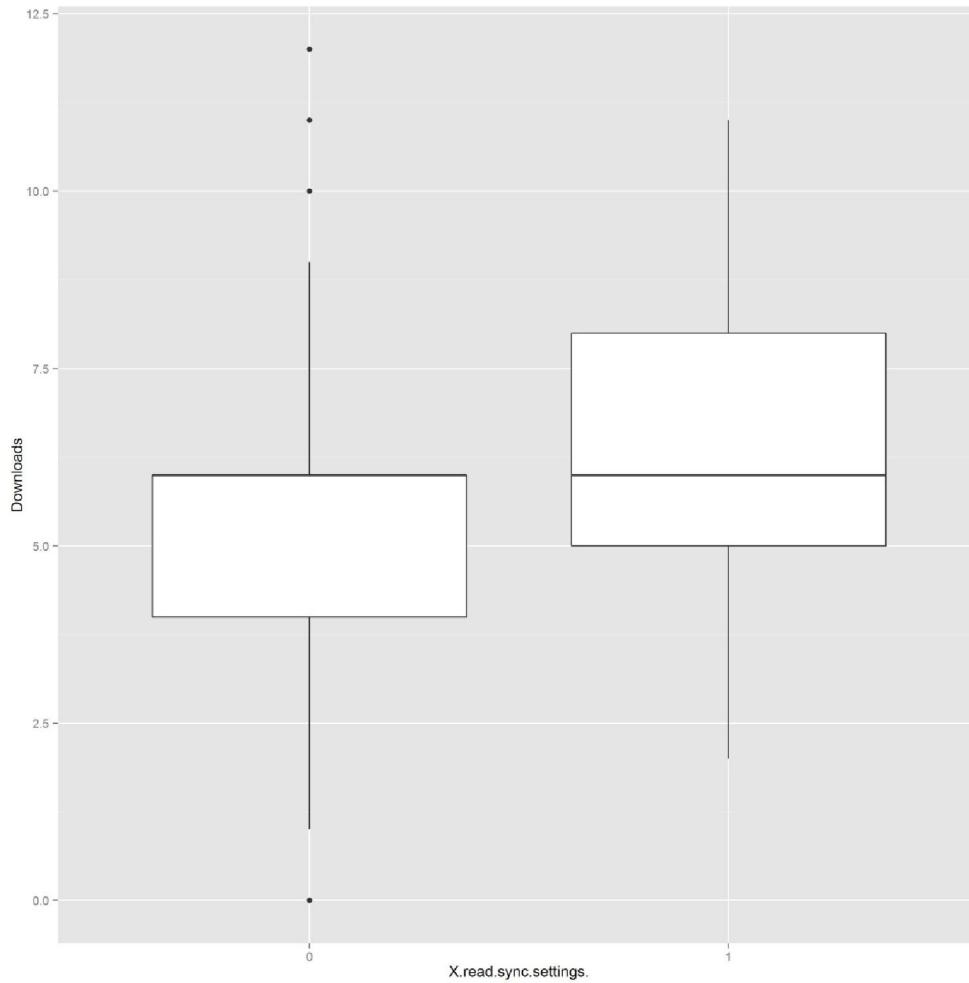


**Figure 20 The Effect of the Change Network Connectivity Concession on Downloads**

### Read Sync Settings

The F value of 18.9 seems to be an effect of applications with this permission having a less skewed and wider distribution than those without the permission.

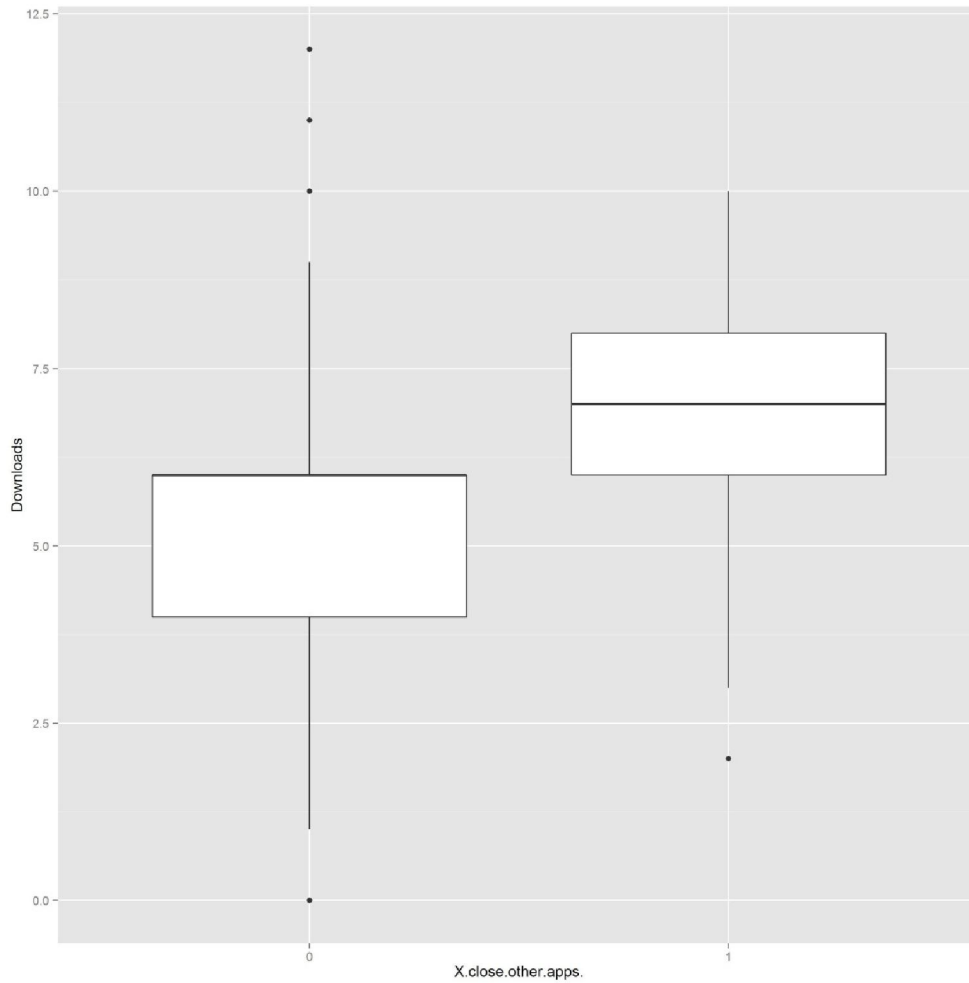




**Figure 21 The Effect of the Read Sync Settings Concession on Downloads**

**Close Other Apps, Set An Alarm**

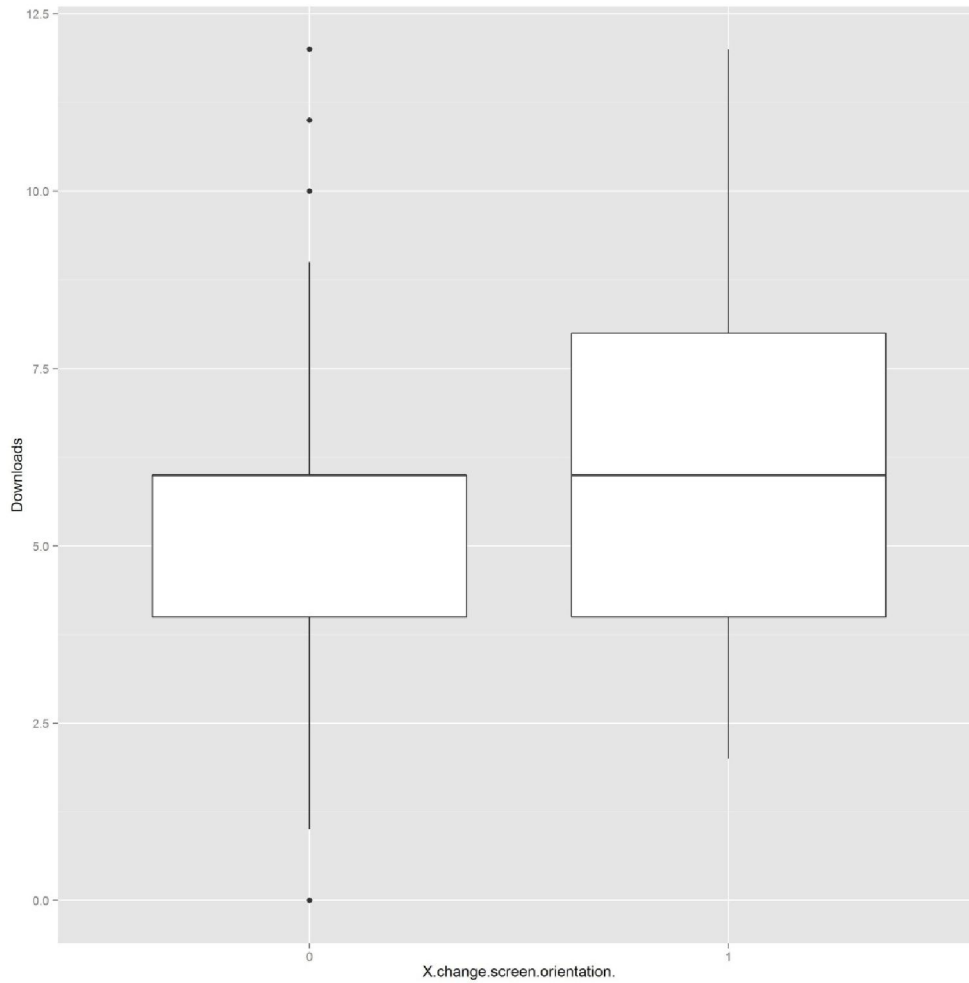
The F value of 17.94 seems to be an effect of applications with this permission having a less skewed distribution and an interquartile range translated upward as compared to those without the permission.



**Figure 22**The Effect of the Close Other Apps, Set an Alarm Concession on Downloads

### Change Screen Orientation

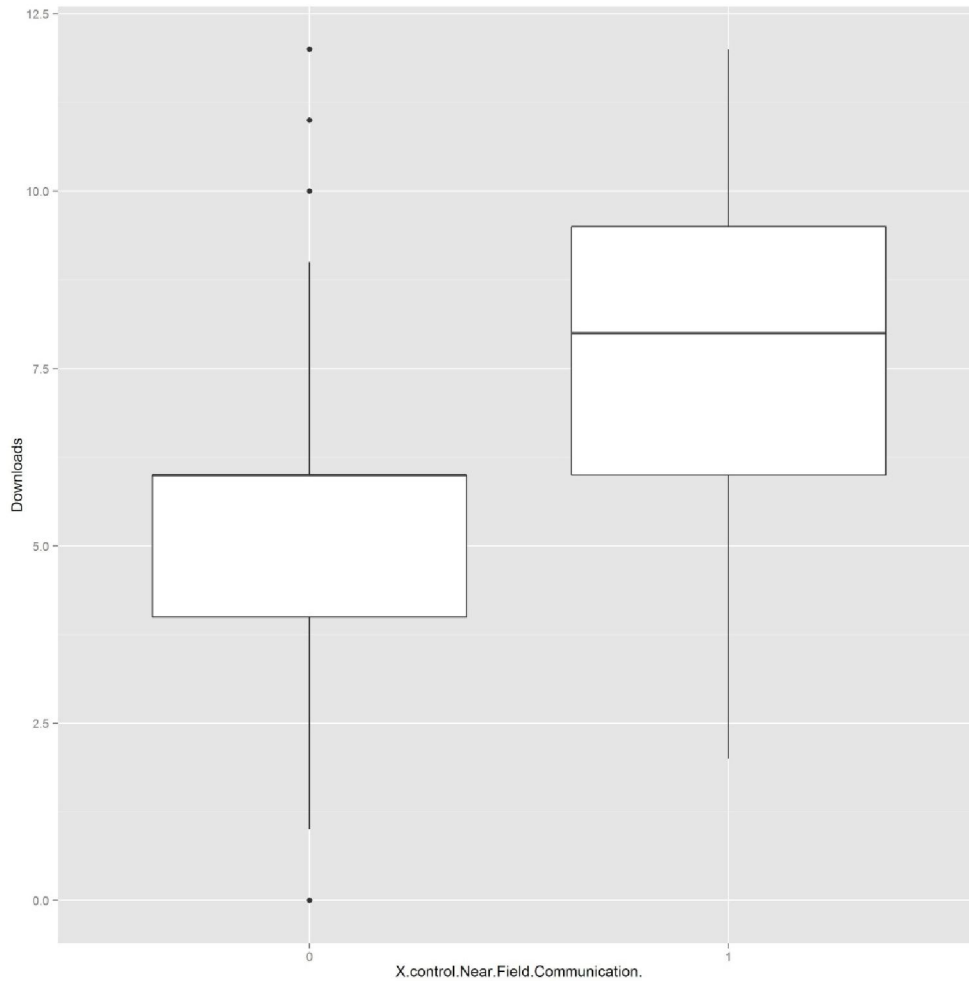
The F value of 14.1 seems to be an effect of applications with this permission having a less skewed and wider interquartile range than those without the permission.



**Figure 23 The Effect of the Change Screen Orientation Concession on Downloads**

### **Control Near Field Communication**

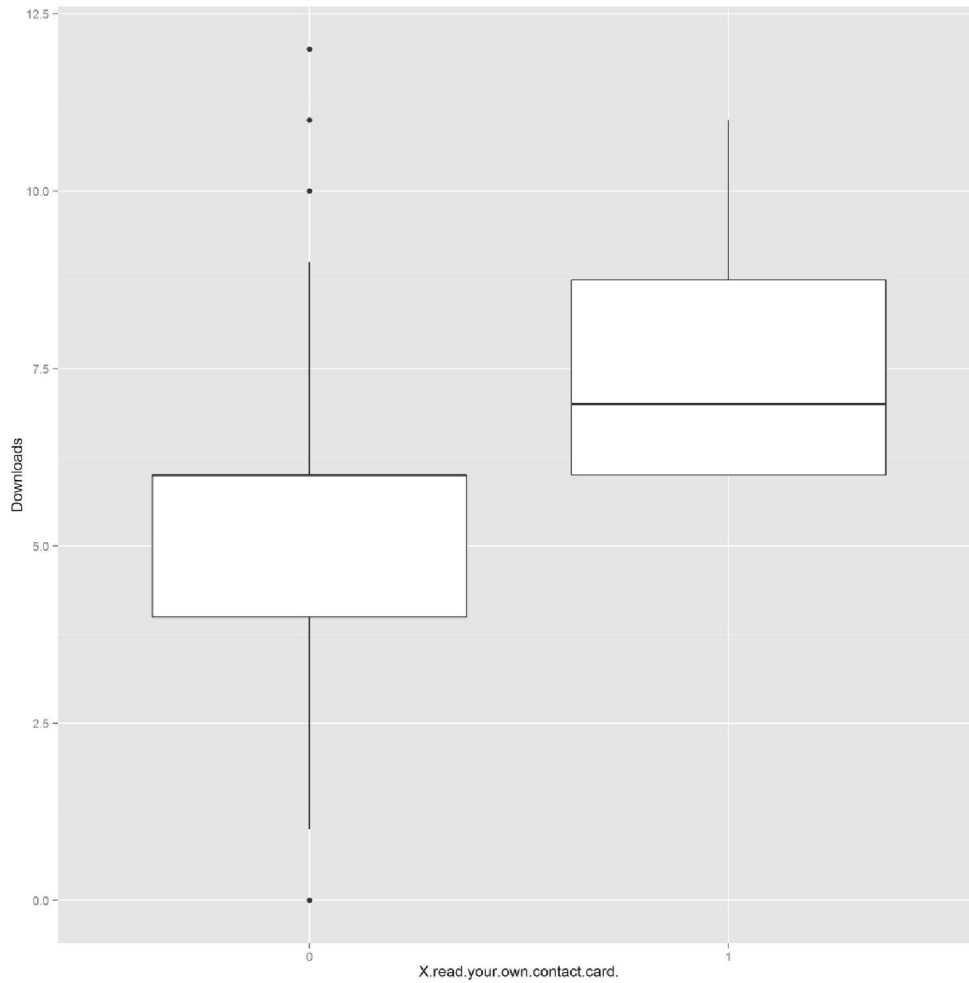
The F value of 17.94 seems to be an effect of applications with this permission having a less skewed, wider distribution and an interquartile range translated upward as compared to those without the permission.



**Figure 24 The Effect of the Control Near Field Communication Concession on Downloads**

### **Read Your Own Contact Card**

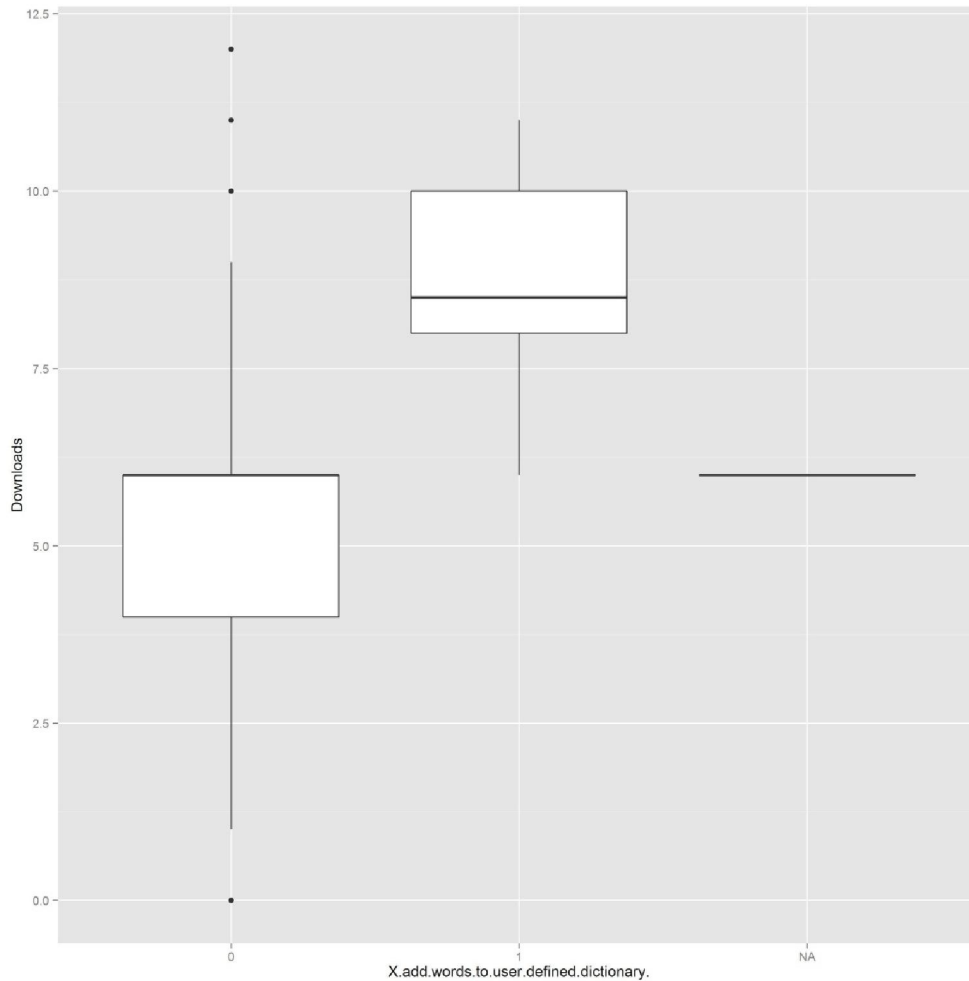
The F value of 18.79 seems to be an effect of applications with this permission having a wider distribution, with a higher median than those without the permission.



**Figure 25**The Effect of the Read Your Own Contact Card Concession on Downloads

### Add Words to User Defined Dictionary

The F value of 30.22 seems to be an effect of applications with this permission having a less skewed distribution and an interquartile range translated upward as compared to those without the permission.



**Figure 26 The Effect of the Add Words to User Defined Dictionary Concession on Downloads**

#### Application Category Effects

The question of whether separate categories of applications have this same relationship to privacy concessions is answered in a similar fashion. The same analysis of differences is run on the data from each category , using an ANOVA, to reveal whether or not the same relationship holds. The main relationship that was found was generally the same as was found in the categorically aggregated case. Specifically, any given privacy concession was associated with a net increase in the number of downloads or had no effect at all.

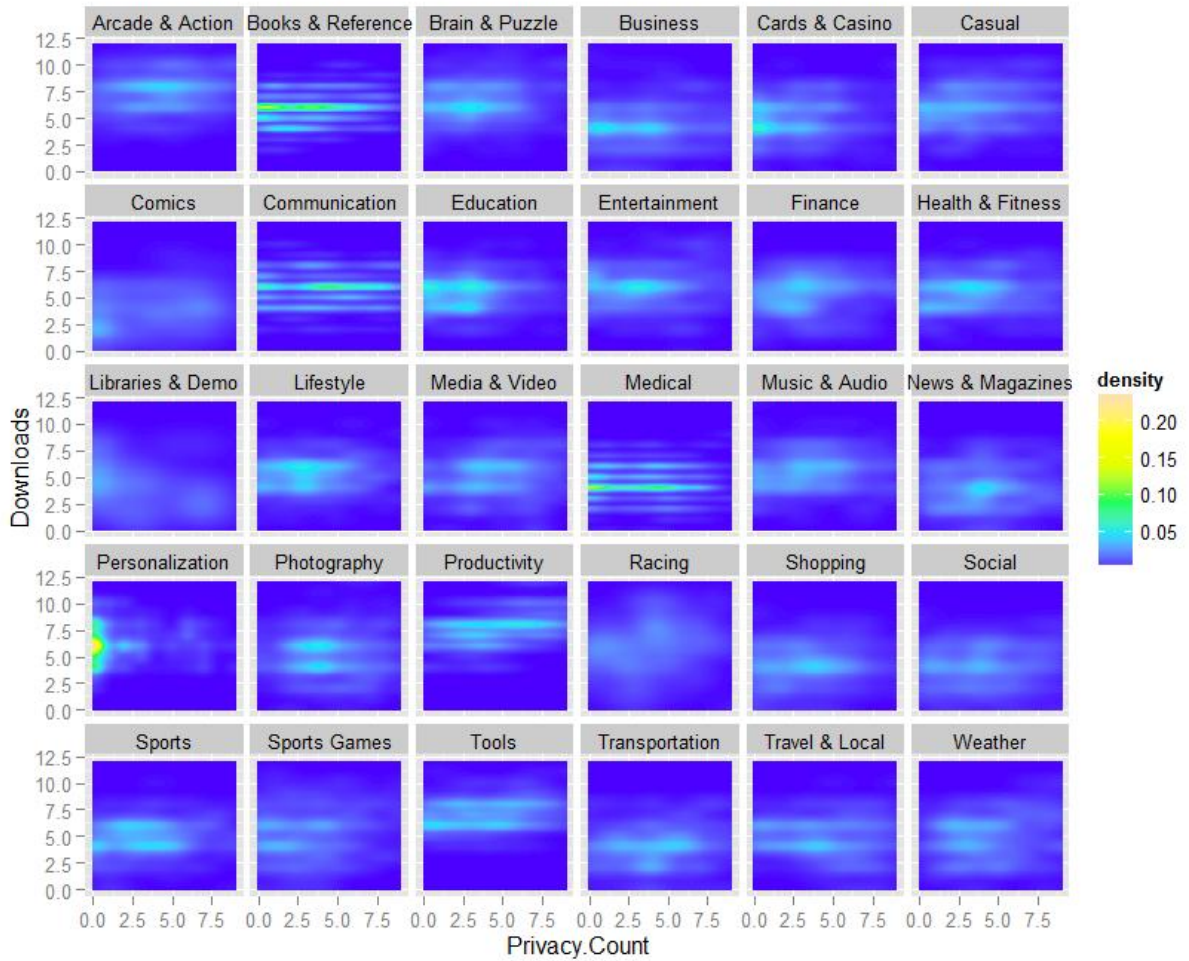
There were, however, several significant exceptions to this trend. Most significant amongst these is within the category of music applications. Analysis revealed that the ability to determine precise location (GPS and network-based) resulted in a significantly ( $p < .001$ ) lower rate of download. In the category health and fitness, the ability to modify your contacts led to a significantly lower rate of download ( $p < .0001$ ), as did writing to the call log ( $p < .0001$ ). Although there were few applications in the category Libraries and Demos, it is interesting to note that the strongest positive relationship was found at  $p < .094$  while most were negative and one was significantly so (modify or delete the contents of your USB storage modify or delete the contents of your SD card ( $p < .02$ )). See the Appendix for graphs of these differences.

### Qualitative Analysis

#### Interesting Case of Personalization

Visually inspecting the Categories VS Privacy Count reveals several categories that are significantly different in terms of their average privacy count. One that stands out the most is Personalization, as the interquartile range is below the median of most of the others. Indeed, examining a density map (**Error! Reference source not found.**) shows the Personalization category to stand out due to its concentration of cases in

the low count category (across all download ranges).



**Figure 27 Concession Count Versus Download Rate by Category**

An interesting follow-up question might be why might make Personalization different from the other categories. Personalization seems to be an interesting category as the objective of the applications contained in the category is to customize the platform itself. An exercise which implies that it would have no purpose other than the gratification via customization. Many hardware functions would scarcely see use in this category which would include changes to backgrounds, ringtones and lock screens as mentioned in **Error! Reference source not found.** Given this, there is not



likely to be a high amount of utility from additional functionality, which would necessitate permissions. This could have two explanatory effects. One is that users do not want applications with more permissions; two, is that the developers making said applications are not exercising said functionality since there is no need (both could of course be true).

## Chapter 5: Discussion

### *The Effect of Privacy Concession Count*

The privacy concessions do not appear to generally be perceived as a cost to the users of applications. If anything, the privacy concessions exist as benefit. As the number of downloads appear to increase with the number of privacy concessions, one can surmise that the users either want to cede more privacy, or that they are keen on purchasing something confounded with the privacy concessions. The main confound with the number of concessions is functionality; applications that are asking for more privacy concessions may indeed be doing more with that data. Thus, the privacy concessions required may, on average, be doing more for the user, causing the user to be more likely to download it.

This would conveniently explain the positive relationship between the number of privacy concessions and the number of downloads. The qualitative analysis on the applications that require an inordinately high number of privacy concessions indicates that applications which require many concessions are highly functional. Indeed, they are highly multifunctional.

All measures indicate the demand for an application as being inelastic to increasing the privacy cost across applications. The counter-examples to the supposition that a high number of privacy concessions convey a high cost which would inhibit downloading are found in those applications which require an extremely high amount of privacy concessions. They hint that, at least when the numbers of concessions are high, the functionality is high.

#### *The Effect of Individual Privacy Concessions*

Across categories, in aggregate, no privacy concession was associated with a decrease in the number of downloads. The number of highly significant positive relationships is quite stunning and lead to statements that can be counter-intuitive to those who would expect the privacy concession to have the negative influence of a cost. It is interesting to interpret this result in light of the other result of the number of concessions having a positive influence on the number of downloads. Specifically, the fact that the number of concessions is positively correlated with the number of purchases of an application and the fact that the individual concessions are demonstrated to increase the number of downloads lead to the conclusion that the presence of any given concession is associated with an increase in sales. Whether this increase is associated with an increase in functionality warrants further study, but is indicated by the qualitative analysis in this study.

### Categorical Effects

One important effort in establishing either a correlational or causal relationship is the elimination of confounding variables. With this in mind, it was useful to search for the effects of specific variables within categories. The result is that the trend of the aggregated variables (of individual concessions leading to an increase in the number of downloads) remains quite consistent with a few notable exceptions. The biggest exception of libraries and demos is the least significant exception, but perhaps the most telling. The overwhelming trend is completely bucked in this category, and the cause may be due to one of three factors. The factor that would be the most convenient is that the selection of the applications is so small that applications that would never have been good enough to make it into the ranks of browseable applications in other sections made it to the ranks here. This seems unlikely as there were other categories that had a small number of applications (e.g. there were only 95 Productivity applications compared with the 72 Library and Demo applications).

Another explanation is that the sample size is so small that the significant results are due to a sampling error. This also seems unlikely, as the trend is so thoroughly bucked that it would be necessary for there to be a sampling error that occurred repeatedly, which may go against the very concept of the sampling error, since we have the entire population of Library and Demo applications available at the point in time of the sampling. Another explanation, which seems the most plausible, is that the audience is different for the Libraries and Demo application, or that they at least have different expectations for these applications. This category consists of software

libraries (e.g., text to speech voices or variable speed playback for music) and plugins for existing applications (e.g. visual effects for a photo editor). The differing expectations could stem from the relationship between the library and the standalone application itself; the application itself would do the heavy lifting, if any is to be done. For example, the writing to storage would be handled by an application, not by a plugin (e.g., visual effects for a photo editor) which should just be expanding the already extant functionality. This would also explain the most significant positive relationship, as the ability to see what other applications are running could aid in integrating with the intended main application.

A final explanation, which dovetails with the aforementioned expectation explanation, is that those libraries and demos which do require additional functionality happen to be applications which fewer people want. In other words, that there is a confounding of quality and popularity, by definition. While this might be plausible, it is not a very testable explanation, as the confounding variable would be confounding by definition. Because this explanation does not offer any predictive power outside of saying that these applications do not meet the expectations of users, and as its testable results are identical to the aforementioned explanation, the two can be restated concurrently thusly: the functionality is not expected to require most permissions and as such, applications successfully adopted by those interested and savvy enough to customize their own applications are more likely to download library and demo applications without this extra permission requirement.

## Chapter 6: Conclusion

### Limitations

One of the biggest shortcomings of this work is the inability to tell if users are self-selecting themselves out of the marketplace entirely. Although those that do not participate in the market do not having an effect on what we can say about those that do participate, research on those that abstain could tell us more about the generalizability of these results to other domains, as well as what could be done in order to bring potential consumers into the ranks of paid application purchasers on their smartphone.

Additionally, it is not apparent what users do after they have downloaded the application. Do they actually use the application? Do they decline an upgrade with updated permissions? Do they uninstall it? Any of these behaviors, all of which are invisible to us in this study, could be indicative of the withdrawal of consent for the privacy concessions.

Another limitation is the censoring of data. With actual download numbers, rather than a range, conclusions might have been stronger, especially as they could be based on analysis techniques that have stricter assumptions and are more nuanced.

Obtaining this uncensored data and performing an analysis on it would be a

worthwhile undertaking if it could be accomplished. That being said, this data is likely to be difficult to obtain, perhaps due to this data's commercial value.

An additional limitation is that the categories are not strictly defined. They are chosen by developers; developers are given very little on which to base their characterization of the application. At this time, it consists of a few examples of application types, at most (e.g. Music and Audio has the examples "Music services, radios, music players"). Similarly, users are also not given the definition of categories. This is a pragmatic approach, and this characterization relies on fuzzy concepts anyway. However, for the purposes of examining categorical relationships, this methodology lacks the ideal amount of rigor. This may be mitigated by the sample sizes. Specifically, if there is on average a common understanding for the definition and delineation of categories, a representative sample should reflect these categories accurately.

Another limitation is that this study was done on the Google Play store only, omitting competing stores such as iTunes. There is some evidence that the users of Android are more likely to be young, less likely to have high income, and more likely to have lower income (Smith, 2013). Black people are also more likely to be Android users than iPhone users (Smith, 2013). Additionally, there is some evidence that Android users, and therefore the users in this study, are more privacy concerned than other smartphone users (Benenson et al., 2013). While more privacy concern should mean that our result of low privacy concern is one that would be stronger in the wider

population, further study on these differences is needed. The other differences deserve further study as well, since our population is not fully demographically representative. Because of this, when applying the finding to the population of smartphone users as a whole, there is selection bias, and more study may be needed to safely apply these findings more generally.

Another shortcoming is that the only applications analyzed were paid applications. It is common knowledge that the vast majority of applications on the mobile platform are free. This poses a problem in two ways. The users that purchase applications may be a different set of users from those that download them for free. Alternatively, users may treat the decision differently when they have to pay for an application. For example, they may give their selection more thought if they have to pay for the application, due to the increased cost. Alternatively, they may be more likely to trust the vendor of paid applications, perhaps thinking that if something is free, there might be a hidden cost.

### *Practical Implications*

#### Design Implications

##### **For Application Designers**

Privacy Concessions required for the addition of features should not have a negative influence when weighing whether are not to include the feature in the implementation and design of an application. Furthermore, one would be advised not to cut features in order to preserve user privacy. Of course, both of these implications are predicated on the existence of a market similar to the one analyzed in this study. Specifically, the

distribution platform must be trusted, and the market must be open to making privacy concessions in the first place.

Aside from considering the effect of the behavior potential customers on one's own application sales, one should also consider the potential effects on the ecosystem in which the application is sold. Specifically, the fact that the USB storage which stores one's own applications may be readily shared with other applications implies the need for potential safeguarding of potentially sensitive user and proprietary business data. Consider, for example, the potential backlash if a company with malicious intentions were to gain access to the passwords for your site, or some trade secrets upon which your business relies. The negative implications of such a breach could potentially be quite troublesome.

### **For Distribution Sites**

The distribution of applications has been analyzed from the perspective of how to increase trust or how to convey the trustworthiness of individual applications or sites. What has scarcely been addressed is the positive effect of having such a system. Within the domain examined in this study, Google has regularly purged malicious applications from the store. In the resulting domain of the reasonably popular applications, all of which are browsable without searching directly for the application, there is no discernible negative effect of privacy concessions on sales. Therefore, full disclosure would be the most financially responsible posture when listing applications or websites. This is because the disclosure had no measurable negative effect on sales, but in the highly trusted domain studied, price had no negative effect on sales



either. More specifically, although one cannot surmise that it will increase sales, one can state that more listed disclosures are associated with more sales, and at the very least it will not harm them. This, of course does not apply to free applications, which far outnumber paid applications (FTC, 2012), but generate no revenue on their own. Such a distribution service would have to be advertising supported, be subscription based, or rely on donations.

### Policy Implications

#### Governmental /Societal

People have reported that they value their privacy, in Westin Surveys, for example. However, when considering the aggregate effect of the privacy paradox, there are societal implications that may have implications for governmental policy. This circumstance is very similar to the situation with sustainable food, where the positive stated attitudes towards sustainable food do not match the limited purchase and popularity of such foods. (Vermeir and Verbeke, 2006). Government action to mitigate this has included the exploration into ways to capture externalities and increase public awareness. Similar approaches may be appropriate for the loss of privacy data to third parties and the cloud.

Additionally, if governments intend to ensure the ownership or control of people's own personal information, they could ensure the perpetual control or traceability through legislation, as controlling later dissemination on an individual level could prove difficult.

Alternatively, regulations might be useful in limiting the collection of data, especially from certain vulnerable groups such as children, who are incapable of giving legal consent to such arrangements. Many, if not all products are allowed to be marketed to, or, alternatively, sold to children. Personal data collection could be similarly restricted.

#### Business

Bring your own Devices is a new policy some companies are adopting. One thing that businesses have to consider is the impact of the user's application ecosystem (on their mobile device). This may result in restrictions placed on the user in terms of what they are allowed to do with the phone. Other ameliorating measures may include the deployment of apps to monitor the ecosystem for specific behaviors, or a rethinking of the extent to which Bring your own Device is an appropriate policy.

#### Personal

As an individual, it may be very important to know about the lack of effect of permissions effect on purchasing behavior (even within categories). The tendency towards a lax personal regulation of privacy means that the individual might benefit from some assistance in this regulation. Researchers have examined agents to make decisions for people, as well as standardized labels, and which have been implemented in several studies (e.g. Tsai et al., 2011), this study's results show that in order to have human like behavior any agent would have to take into account

functionality when evaluating the protection of privacy. Additionally the reach of such applications may be limited to privacy fundamentalists (as described by Westin, 1990).

### *Theoretical Implications*

#### Bolstering Related Findings

This study finds that in general, trading their privacy as part for an application's functionality does not affect people's installation behavior. This conclusion is based on observational data, but is backed up by poll data, which has found that persistent identifying information is acceptable to most (78% of) users if it used for a customer service. (Ackerman et al., 1999) Additionally, it is backed up by experimental data; this study's results bolster the results of a recent study which found that privacy was not a factor in peoples purchasing decisions (Beresford et al., 2012). The study allowed participants to buy from one of two stores; in the first experimental treatment, the cheaper store required more personal data, and the users opted to buy from the cheaper store. In the second treatment, the stores had the same prices, and one required more personal data. In the second treatment, participants showed no preference for one store over the other.

#### A Contemporaneous Finding

While this study was being completed, a similar examination to determine how privacy concessions impact purchasing decisions was being published (Egelman, Felt, and Wagner, 2013). Researchers examined this in two ways. They presented subjects

four mocked up applications to compare based on their attributes (rating description privacy concessions etc.). The number privacy concessions were inversely proportional to the price, and the researchers found that a quarter of participants would pay an extra \$1.50 on a \$0.49 the application fewest permissions, however over forty percent of users preferred the cheapest application which required the most permissions. In a second experiment, they performed a reverse Vickrey auction (participants would name the price they would be paid to beta test an app without knowing the other bids) on applications that required various permissions. The applications in this experiment had between one and four permissions, two of which were unrelated to the applications functionality. They also asked about the factors that influenced the bids as well as a question about targeted ads. They found that only the ability of an application to read contact list, which was unrelated to the any of the functionality that the application purported to have, had any significant impact on the bid placed. They also found that 77% of users would take targeted ads if it meant saving a dollar regardless of what private data was required in order to do so.

These findings of Egelman, Felt, and Wagner (2013) are corroborative with the findings of this study in several ways. In their sample, a minority of people were willing to pay more for an application with fewer permissions, however even amongst these privacy concerned participants, privacy was ranked fourth amongst the self-reported decision factors. This bolsters the finding that users do not highly value their Privacy, and in fact value the other attributes of an application much more. Additionally, the finding that when all things are not being held equal, e.g. price

being inversely related to the number of permissions required, the applications with the cheaper price will have a higher rate of acceptance seems to support the finding that the individual privacy concerns do not in general have a negative impact on the number of downloads.

The main divergent finding is that Egelman, Felt, and Wagner (2013), found that a correlation between the number of privacy concessions and the bid. Consequently, with all other things being equal, the number of permissions positively affected the amount of compensation required to evaluate it. In other words, the users valued their privacy and treated the number of privacy concessions as a cost. This divergent finding may be an effect of the negative impact of a completely unneeded permission (in this case, the Read Contacts concession), as there were only four permissions in the study. Another explanation is that the positive relationship found between downloads and privacy concession count may be due to a lack of control in our study for the functionality of each application. Specifically, the positive relationship found in this study could be due to the potentially confounding variable of functionality, while the inverse relationship found by Egelman, Felt, and Wagner (2013) could be due to the concession count itself. Furthermore, this may be related to our finding that certain categories had specific permissions that were associated with a lower acceptance rate. In each of these cases the permissions which lowered download rate were in no way related to the main function of category. This finding is further bolstered by the finding of Egelman, Felt, and Wagner (2013), that an unrelated

permission (the Read Contacts concession), lowered user acceptance, while the other permissions did not have any effect.

### **The Privacy Paradox**

The most obvious concept that is informed by this research is the concept of the privacy paradox. The privacy paradox is generally thought of as the reporting of a high amount of privacy concern expressed by people, while the same people are willing to divulge this personal information. This phenomenon has been demonstrated in empirically in a college population (Spiekermann et al., 2001), internet users (Ackerman et al., 2001), ethnographically and phenomenologically explored in supermarket shoppers with club cards (Sayre and Horne), and measured in an experimental setting (Norberg et al., 2007.). One study, in fact found that while risk had an effect on people's intention to disclose private information, it had no effect on their actual behavior. These studies imply that we do not currently fully understand the relationship that privacy considerations play in an individual's behavior, and neither do the individuals (Jensen et al., 2005). The findings of this study bolsters these findings in the following way: rather than demonstrating that there was a difference between stated privacy concern and privacy disclosure, it demonstrates that there is rarely any negative impact created by disclosure at all.

This paradox may be rooted in some of the unspoken assumptions of privacy research. One assumption is that when people make conscious predictions about the about the impact of present decisions on their future, they can do so accurately. This has been demonstrated to be false when it comes to affect; specifically, impact bias,

wherein people “overestimate the intensity and duration of future events” (Wilson and Gilbert, 2005).

One assumption is that people are able to articulate their privacy preferences. Some studies have found a significant correlation between privacy preferences and stated hypothetical future actions. For example, one study validated the users’ general privacy concern by comparing it to the users stated intention to provide sensitive data (Castañada and Montoro, 2007). Rather than refuting the privacy paradox this seems to highlight what makes it nuanced, that people’s stated preference can line up with their stated intentions; this does not contradict that they behave radically differently in everyday situations.

If it is assumed that the applications that are directly findable through browsing on the Google Play store are on average asking for the permissions necessary to perform their stated functionality, and not some clandestine purpose, then the number of permissions should be proportional to the breadth of functionality. Given the aforementioned low value of privacy in actions, one would expect the findings obtained in this study: that specific privacy concessions and the number of privacy concessions are associated with more user acceptance. This would follow, because as concessions are required, functionality would increase, and this would be unmitigated by privacy concerns, despite any verbal behavior to the contrary.

### Current Models

Traditional means of describing user population's relationship to their privacy, have often been based on their stated privacy attitude (e.g. the Westin groupings), however, with the privacy paradox in mind, further research is required on the individual behaviors of individual in the real world. Previous empirical studies (e.g. Norberg et al., 2007), have demonstrated the divorce of stated intention to attitude, but in light of this foundational analyses of privacy desires and attitudes may need to be re-evaluated

### Cohort Effects

Much research has been devoted to analyzing the differences in attitude and behavior in technology as generations pass. It has been found that second generation Digital Natives (people born after 1993, surveyed in 2012) have more positive attitudes towards the internet and use technology more than first generation Digital Natives (people born since 1980, surveyed in 2002). Although this finding must be caveated by the confounding variable of the different survey time (maybe the attitudes of all cohorts changed over this period), this finding of cohort effects has been seen in other countries; for example, in China differences in privacy attitudes has been demonstrated (Shengming and Xiaoping, 2006) as well.



### Future Work

In addition to the aforementioned findings, it is been shown that privacy concerns are related to technology usage and literacy (Bellman et al., 2004). Since one might expect technology usage to increase in future cohorts, as Internet experience, for example has been shown to be inversely proportional to age (Zukowski and Brown, 2007). These facts together mean that any established analysis of what people want and how they will behave in terms of privacy are potentially subject to change over time. On the other hand based on previous studies on what may be needed for a more robust analysis, is a more comprehensive model of privacy regulation in people in general.

It has also shown that when people (especially who attempt to maximize the value of their preferences) are presented with fewer choices, they can be happier with their decisions (e.g. Iyengar et al., 2006 and Schwartz et al. 2002). Might the primacy of the Google Play store (similar to other mobile form application marketplaces), lead to a perceived lack of choice, and therefore to a userbase more happy with the applications that they are presented with? This could be examined at length, with other study designs.

In order to gain a fuller understanding of the privacy regulation on mobile platforms further examination of user behavior in this domain could be examined. As this study only looked at this issue from the applications point of view, to gain a deeper understanding one could look at the user's behavior more from the user's perspective.

For example one could look at the user's purchasing behavior across applications across applications. Additionally, one could examine related usage outside of purchasing behavior. For example the user may uninstall shortly after purchasing which might help us figure out the extent to which desired functionality confounds aversion to privacy concessions. Alternatively to accomplish this, frequency of use could be examined.

With a similar methodology, other data sets could be examined. For example, data from iTunes could be examined. Rather than looking at the applications shown in the categorical listing, which was limited to a subset of applications, all applications contained in the store could be analyzed as well. Additionally, the behavior of users on paid applications could be compared to the download behavior of users on free applications. Similarly, rather than looking at the initially browsable applications only, all of the applications could be looked at. This would be especially useful for figuring out whether Personalization applications have fewer required concessions was a result of a difference in the design of applications, or whether this was due to users preferring to purchase personalization applications with fewer concessions. Further analysis in this vein could be used to examine the finding of a decrease in downloads associated with fewer permissions. As this may conflict with a study which controlled for functionality (Egelman, Felt, and Wagner, 2013), a methodologically similar study which sampled applications with similar functionality could compare download rates and look for the relationships with privacy in this narrowed sample. This would tease out the variable of functionality, which might be confounded with permissions.

Older individuals have been shown to pick up less on facial cues of untrustworthiness, an antecedent condition of which has been demonstrated to be a muted reaction in the anterior insular, as observed in an fMRI (Castle et al., 2012). In order to bolster the validity of privacy research, as well as connect it to physical research fMRI (e.g. Riedl et al., 2010), and other biometric studies could be undertaken in order to determine whether or not privacy behaviors can be associated with related cortical structures, and, furthermore whether the postulated privacy preference groupings have associated physical basis.

There is still much more research that needs to be done on the privacy and security of mobile devices. The mobile platform is a new one as is the concept of cloud computing. The study of how people manage their privacy is in its infancy as well. In contrast, more established fields such as physics and engineering predate modern science. Before evolution of the scientific method, these fields brought us many technologies, from the wheel and metallurgy, to the aqueducts and the construction of cities. The nuclear, Internet and industrial ages, however, would not have been possible without the scientific method, and the maturity of concepts and theories that came with rigorous forms of inquiry. When Bridgman wrote his reflections on operational concepts and methods, it was largely motivated because, “We must demand that the set of operations equivalent to any concept be a unique set, for otherwise there are possibilities of ambiguity in practical applications which we cannot admit.” (Bridgman, 1927) This statement was in the context of physics where

there is far less ambiguity than there currently are in the field of HCI (especially in Trust and Privacy research). With all of the advancements and cost reductions in the recent past, perhaps one limitation (and cause of stagnation) to the current state of HCI is that the precision and rigor of the hard sciences may be necessary in order to experience a renaissance.

## Appendices

### ***Terms of Service***

Google Play Terms of Service: [http://play.google.com/intl/en\\_us/about/play-terms.html](http://play.google.com/intl/en_us/about/play-terms.html)

Google Terms of Service: <https://www.google.com/policies/terms/>

# Graphs

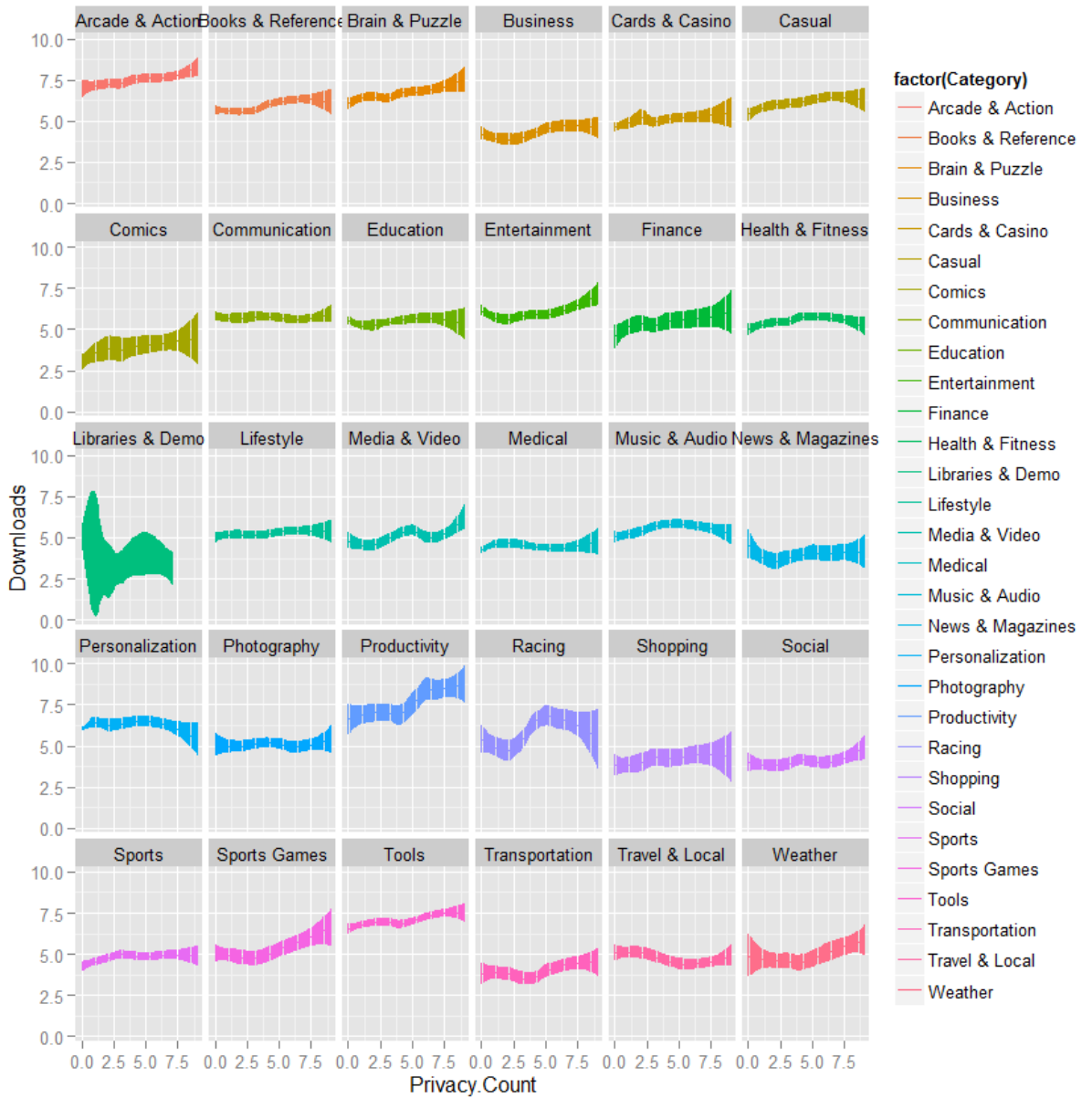
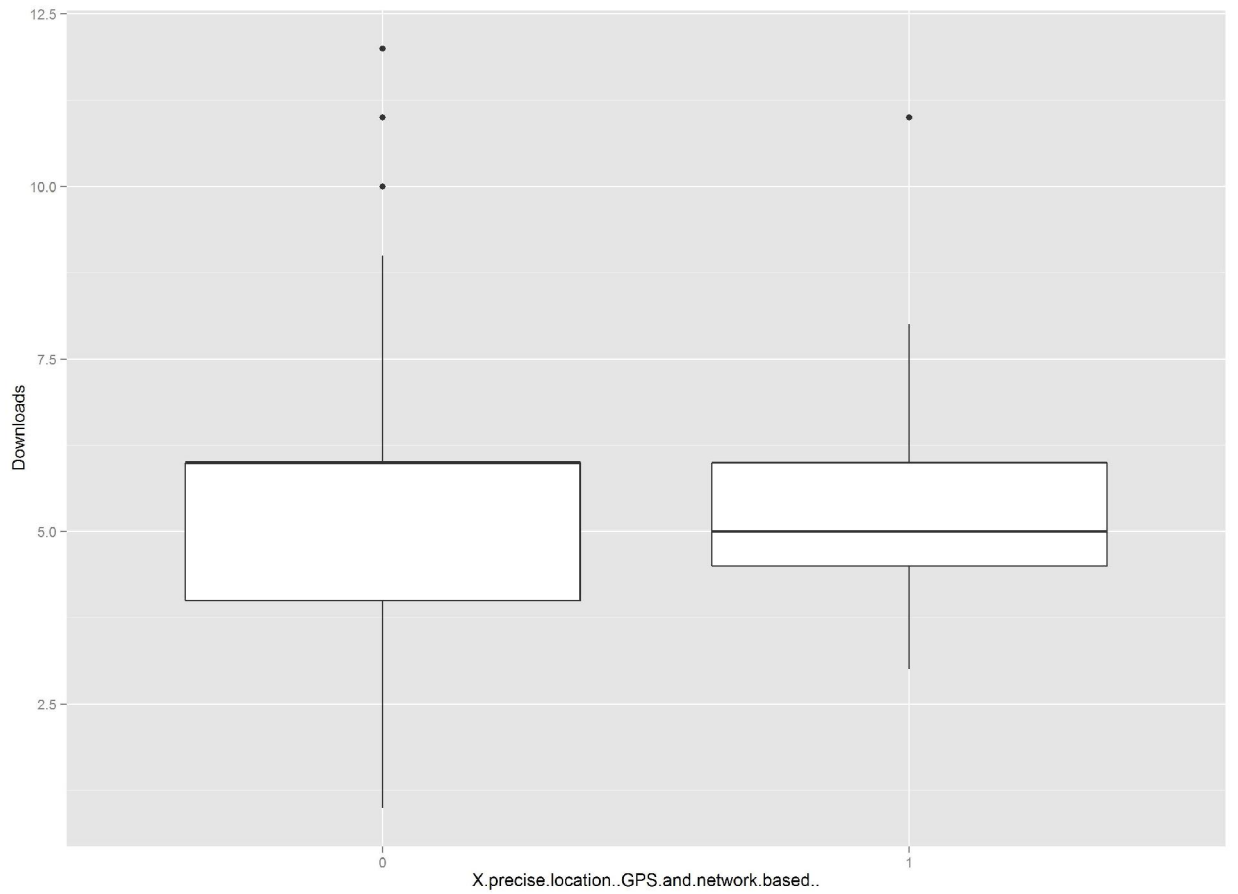
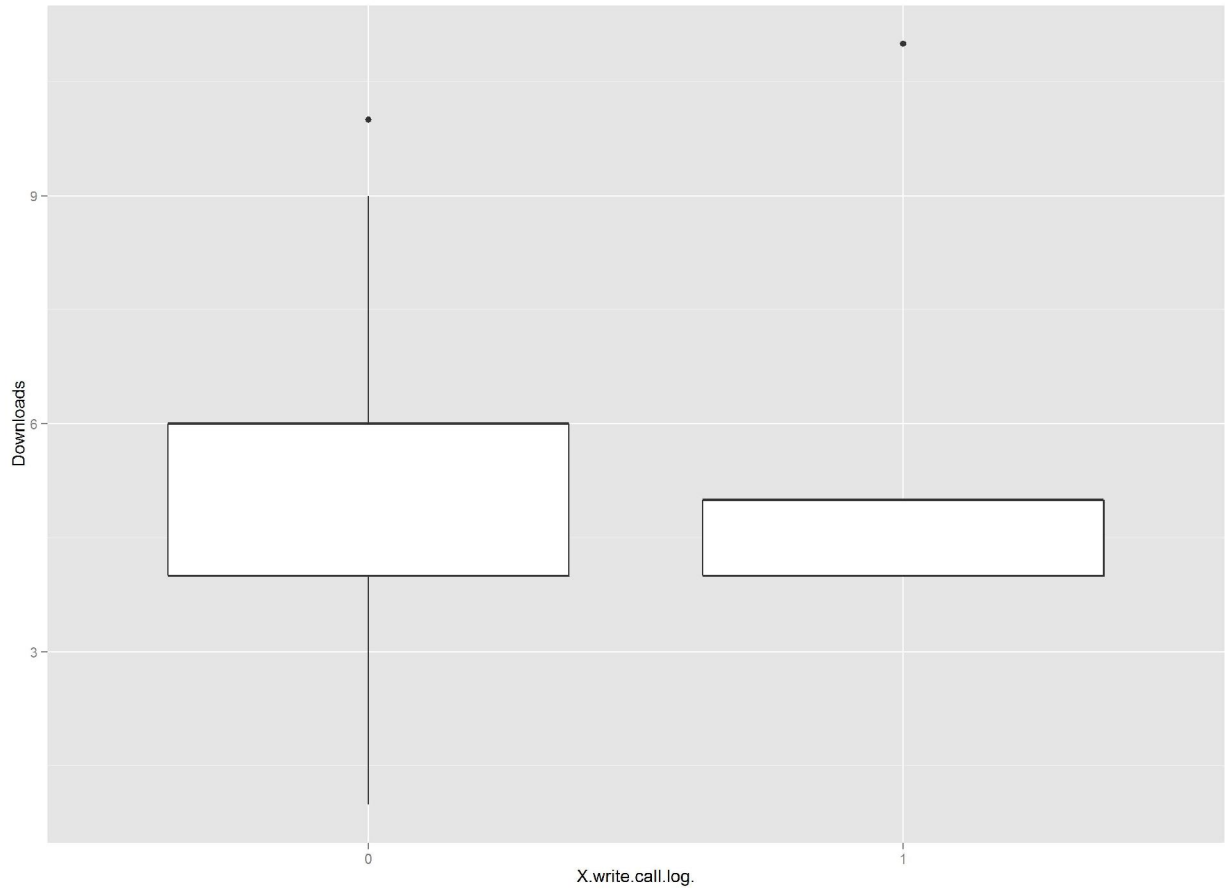


Figure 28 Standard Errors of Downloads Across Concessions Counts

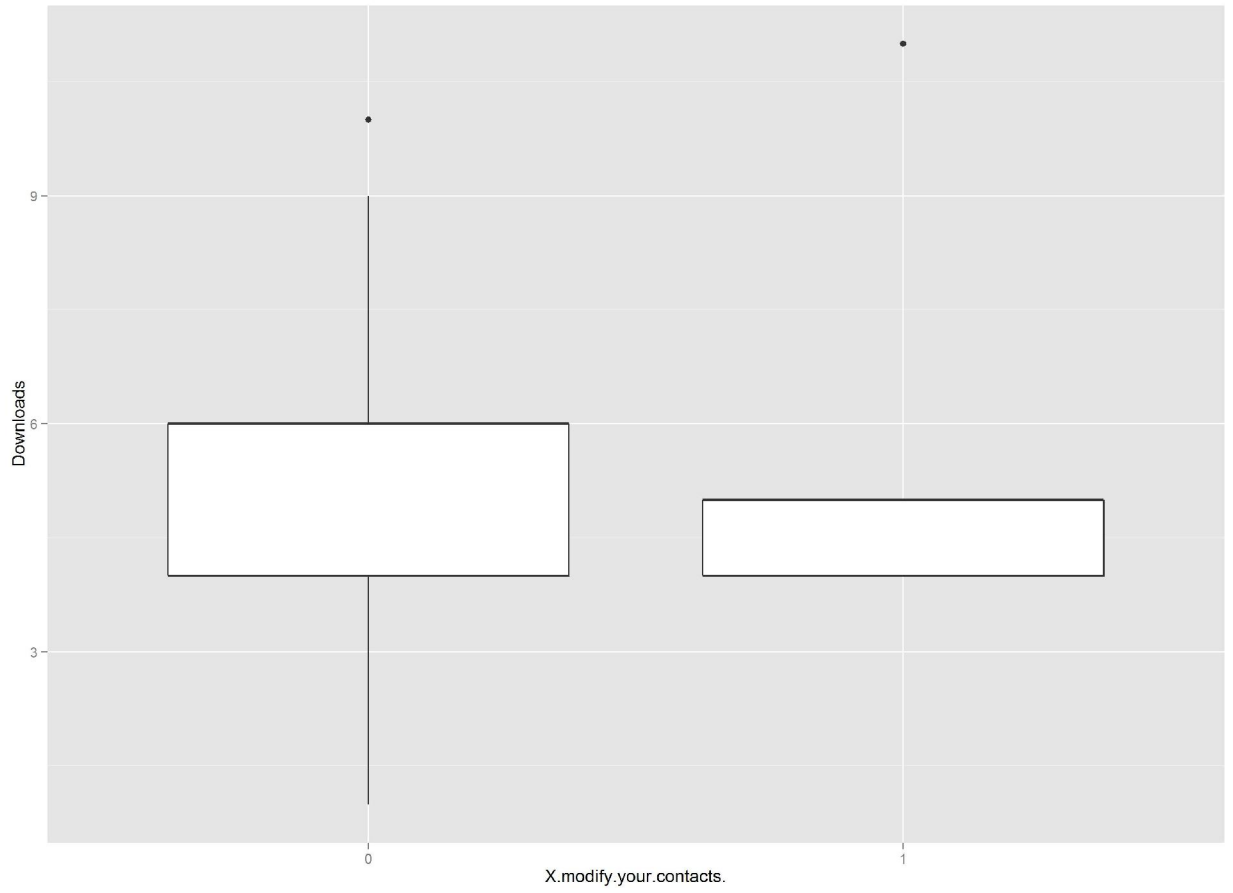


**Figure 29: Difference in the Music category**



**Figure 30: Write To call log in Health Applications**





**Figure 31: Modify Contacts in Health Applications**

## ***Categorical Descriptions***

Comics	Comic players, comic titles
Communications	Messaging, chat/IM, dialers, address books, browsers, call management, etc.
Finance	Banking, payment, ATM finders, financial news, insurance, taxes, portfolio/trading, tip calculators, etc.
Health & Fitness	Personal fitness, workout tracking, diet and nutritional tips, health & safety etc.
Medical	Drug & clinical references, calculators, handbooks for health-care providers, medical journals & news, etc.
Lifestyle	Recipes, style guides
Media & Video	Subscription movie services, remote controls, media/video players
Music & Audio	Music services, radios, music players
Photography	Cameras, photo editing tools, photo management and sharing
News & Magazines	Newspapers, news aggregators, magazines, blogging, etc.
Weather	Weather reports
Productivity	Notepad, to do list, keyboard, printing, calendar, backup, calculator, conversion, etc.
Business	Document editor/reader, package tracking, remote desktop, email management, job search, etc.
Books & Reference	Book readers, reference books, text books, dictionaries, thesaurus, wikis, etc.
Education	Exam preparations, study-aids, vocabulary, educational games, language learning, etc.
Shopping	Online shopping, auctions, coupons, price comparison, grocery lists, product reviews, etc.
Social	Social networking, check-in, blogging, etc.
Sports	Sports News & Commentary, score tracking, fantasy team management, game Coverage, etc.
Personalization	Wallpapers, live wallpapers, home screen, lock screen, ringtones
Tools	
Travel & Local	City guides, local business information, trip management tools
Libraries & Demo	Software Libraries

**Table 3 Category Descriptions**

## Bibliography

- Ackerman, M. S., Cranor, L. F., & Reagle, J. (1999). Privacy in e-commerce (pp. 1–8). ACM Press. doi:10.1145/336992.336995
- Acquisti, A., & Grossklags, J. (2005). Privacy and rationality in individual decision making. *IEEE Security and Privacy Magazine*, 3(1), 26–33. doi:10.1109/MSP.2005.22
- Ahern, S., Eckles, D., Good, N. S., King, S., Naaman, M., & Nair, R. (2007). Overexposed? (p. 357). ACM Press. doi:10.1145/1240624.1240683
- Alan Westin. (1990). PrivacyExchange.org. Retrieved November 3, 2013, from <http://web.archive.org/web/20040305145947/http://www.privacyexchange.org/iss/surveys/eqfx.execsum.1990.html>
- Altman, I. (1977). Privacy Regulation: Culturally Universal or Culturally Specific? *Journal of Social Issues*, 33(3), 66–84. doi:10.1111/j.1540-4560.1977.tb01883.x
- Bellman, S., Johnson, E. J., Kobrin, S. J., & Lohse, G. L. (2004). International Differences in Information Privacy Concerns: A Global Survey of Consumers. *The Information Society*, 20(5), 313–324. doi:10.1080/01972240490507956
- Benenson, Z., Gassmann, F., & Reinfelder, L. (2013). Android and iOS users' differences concerning security and privacy (p. 817). ACM Press. doi:10.1145/2468356.2468502
- Beresford, A. R., Kübler, D., & Preibusch, S. (2012). Unwillingness to pay for privacy: A field experiment. *Economics Letters*, 117(1), 25–27. doi:10.1016/j.econlet.2012.04.077

- Boyles, J. L., Smith, A., & Madden, M. (2012). *Privacy and data management on mobile devices*. Retrieved from <http://pewinternet.org/Reports/2012/Mobile-Privacy.aspx>
- Brenkert, G. G. (1998). TRUST, MORALITY AND INTERNATIONAL BUSINESS. *Business Ethics Quarterly*, 8(2), 293–317. doi:Article
- Bridgman, P. W. (1927). *The logic of modern physics*. Macmillan New York.
- Castañeda, J. A., & Montoro, F. J. (2007). The effect of Internet general privacy concern on customer behavior. *Electronic Commerce Research*, 7(2), 117–141.  
doi:10.1007/s10660-007-9000-y
- Castle, E., Eisenberger, N. I., Seeman, T. E., Moons, W. G., Boggero, I. A., Grinblatt, M. S., & Taylor, S. E. (2012). Neural and behavioral bases of age differences in perceptions of trust. *Proceedings of the National Academy of Sciences*, 109(51), 20848–20852. doi:10.1073/pnas.1218518109
- Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003). On-line trust: concepts, evolving themes, a model. *International Journal of Human-Computer Studies*, 58(6), 737–758.  
doi:10.1016/S1071-5819(03)00041-7
- Cortina, J. M., & Landis, R. S. (2010). The Earth Is Not Round (p = .00). *Organizational Research Methods*, 14(2), 332–349. doi:10.1177/1094428110391542
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. doi:10.1037/h0040957
- Culnan, M. J., & Armstrong, P. K. (1999). Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organization Science*, 10(1), 104–115.

- Davi, L., Dmitrienko, A., Sadeghi, A.-R., & Winandy, M. (2011). Privilege Escalation Attacks on Android. In M. Burmester, G. Tsudik, S. Magliveras, & I. Ilić (Eds.), *Information Security* (Vol. 6531, pp. 346–360). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from [http://www.springerlink.com/index/10.1007/978-3-642-18178-8\\_30](http://www.springerlink.com/index/10.1007/978-3-642-18178-8_30)
- Dolnicar, S., & Jordaan, Y. (2007). A Market-Oriented Approach to Responsibly Managing Information Privacy Concerns in Direct Marketing. *Journal of Advertising*, 36(2), 123–149. doi:10.2753/JOA0091-3367360209
- Dourish, P., & Anderson, K. (2006). Collective Information Practice: Exploring Privacy and Security as Social and Cultural Phenomena. *Human-Computer Interaction*, 21(3), 319–342. doi:10.1207/s15327051hci2103\_2
- Dourish, P., Grinter, R. E., Delgado de la Flor, J., & Joseph, M. (2004). Security in the wild: user strategies for managing security as an everyday, practical problem. *Personal and Ubiquitous Computing*, 8(6), 391–401. doi:10.1007/s00779-004-0308-5
- Egelman, S., Felt, A. P., & Wagner, D. (2013). Choice Architecture and Smartphone Privacy: There's a Price for That. In R. Böhme (Ed.), *The Economics of Information Security and Privacy* (pp. 211–236). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from [http://link.springer.com/10.1007/978-3-642-39498-0\\_10](http://link.springer.com/10.1007/978-3-642-39498-0_10)
- Featherman, M. S., & Wells, J. D. (2010). The intangibility of e-services. *ACM SIGMIS Database*, 41(2), 110. doi:10.1145/1795377.1795384
- Federal Trade Commission. (2012, February 16). FTC Report Raises Privacy Questions About Mobile Applications for Children. Retrieved November 1, 2013, from [http://www.ftc.gov/opa/2012/02/mobileapps\\_kids.shtml](http://www.ftc.gov/opa/2012/02/mobileapps_kids.shtml)

- Felt, A. P., Finifter, M., Chin, E., Hanna, S., & Wagner, D. (2011). A survey of mobile malware in the wild (p. 3). ACM Press. doi:10.1145/2046614.2046618
- Gomez, G., Calle, M. L., Oller, R., & Langohr, K. (2009). Tutorial on methods for interval-censored data and their implementation in R. *Statistical Modelling*, 9(4), 259–297. doi:10.1177/1471082X0900900402
- Grudin, J. (2001). Desituating Action: Digital Representation of Context. *Human-Computer Interaction*, 16(2), 269–286. doi:10.1207/S15327051HCI16234\_10
- Huberman, B. A., Adar, E., & Fine, L. R. (2005). Valuating privacy. *Security & Privacy, IEEE*, 3(5), 22–25.
- Iyengar, S. S., Wells, R. E., & Schwartz, B. (2006). Doing Better but Feeling Worse: Looking for the “Best” Job Undermines Satisfaction. *Psychological Science*, 17(2), 143–150. doi:10.1111/j.1467-9280.2006.01677.x
- Jensen, C., Potts, C., & Jensen, C. (2005). Privacy practices of Internet users: Self-reports versus observed behavior. *International Journal of Human-Computer Studies*, 63(1-2), 203–227. doi:10.1016/j.ijhcs.2005.04.019
- Kammerer, M. (2000). Die Bestimmung von Vertrauen in Internetangebote. Lizensiatsarbeit der Philosophischen Fakultät der Universität Zürich.  
[http://opac.nebis.ch/F/?local\\_base=NEBIS&CON\\_LNG=GER&func=find-b&find\\_code=SYS&request=005496779](http://opac.nebis.ch/F/?local_base=NEBIS&CON_LNG=GER&func=find-b&find_code=SYS&request=005496779)
- Koch, S. (1992). Psychology’s Bridgman vs Bridgman’s Bridgman: An Essay in Reconstruction. *Theory & Psychology*, 2(3), 261–290.  
doi:10.1177/0959354392023002

- Kumaraguru, P., & Cranor, L. F. (2005). Privacy indexes: A survey of westins studies. *Institute for Software Research International*.
- Little, L., Briggs, P., & Coventry, L. (2011). Who knows about me?: an analysis of age-related disclosure preferences. In *Proceedings of the 25th BCS Conference on Human-Computer Interaction* (pp. 84–87). Swinton, UK, UK: British Computer Society. Retrieved from <http://dl.acm.org.proxy-bc.researchport.umd.edu/citation.cfm?id=2305316.2305332>
- Madden, M., Lenhart, A., Cortesi, S., & Gasser, U. (2012). *Teens and Mobile Apps Privacy*. Retrieved from <http://www.pewinternet.org/Reports/2013/Teens-and-Mobile-Apps-Privacy.aspx>
- Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research*, 15(4), 336–355. doi:10.1287/isre.1040.0032
- Metzger, M. J. (2006). Privacy, Trust, and Disclosure: Exploring Barriers to Electronic Commerce. *Journal of Computer-Mediated Communication*, 9(4), 00–00. doi:10.1111/j.1083-6101.2004.tb00292.x
- Nickel, J., & Schaumburg, H. (2004). Electronic privacy, trust and self-disclosure in e-recruitment (p. 1231). ACM Press. doi:10.1145/985921.986031
- Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. *Journal of Consumer Affairs*, 41(1), 100–126. doi:10.1111/j.1745-6606.2006.00070.x
- Palen, L., & Dourish, P. (2003). Unpacking “privacy” for a networked world (p. 129). ACM Press. doi:10.1145/642611.642635

- Privacy Revelations for Web and Mobile Apps. (n.d.). Retrieved from [http://www.usenix.org/event/hotos11/tech/final\\_files/Wetherall.pdf](http://www.usenix.org/event/hotos11/tech/final_files/Wetherall.pdf)
- Riedl, R., Hubert, M., & Kenning, P. (2010). Are there neural gender differences in online trust? an fMRI study on the perceived trustworthiness of ebay offers. *MIS Q.*, *34*(2), 397–428.
- Riegelsberger, J. (2005). Trust in mediated interactions. Dissertation submitted to the University College, London, UK.
- Riegelsberger, Jens, Sasse, M. A., & McCarthy, J. D. (2003). The researcher's dilemma: evaluating trust in computer-mediated communication. *International Journal of Human-Computer Studies*, *58*(6), 759–781. doi:10.1016/S1071-5819(03)00042-9
- Riegelsberger, Jens, Vasalou, A., Bonhard, P., & Adams, A. (2006). Reinventing trust, collaboration and compliance in social systems (p. 1687). ACM Press. doi:10.1145/1125451.1125763
- Sayre, S., & Horne, D. (2000). Trading Secrets for Savings: How Concerned are Consumers About Club Cards as a Privacy Threat?. *Advances in Consumer Research*, *27*(1), 151 – 155.
- Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., & Lehman, D. R. (2002). Maximizing versus satisficing: Happiness is a matter of choice. *Journal of Personality and Social Psychology*, *83*(5), 1178–1197. doi:10.1037/0022-3514.83.5.1178
- Singh, S., & Morley, C. (2009). Young Australians' privacy, security and trust in internet banking (p. 121). ACM Press. doi:10.1145/1738826.1738846



- Smith, A. (2013). Smartphone ownership—2013 update. *Pew Research Center: Washington DC*. Retrieved from <http://www.pewinternet.org/2013/06/05/smartphone-ownership-2013/>
- Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information Privacy: Measuring Individuals' Concerns about Organizational Practices. *MIS Quarterly*, 20(2), 167. doi:10.2307/249477
- Spiekermann, S., Grossklags, J., & Berendt, B. (2001). E-privacy in 2nd generation E-commerce. In *Proceedings of the 3rd ACM conference on Electronic Commerce - EC '01* (pp. 38–47). Tampa, Florida, USA. doi:10.1145/501158.501163
- Stewart, K. A., & Segars, A. H. (2002). An Empirical Examination of the Concern for Information Privacy Instrument. *Information Systems Research*, 13(1), 36–49. doi:10.1287/isre.13.1.36.97
- Tan, Y.-H., & Thoen, W. (2000). Toward a Generic Model of Trust for Electronic Commerce. *International Journal of Electronic Commerce*, 5(2), 61–74.
- Teufl, P., Kraxberger, S., Orthacker, C., Lackner, G., Gissing, M., Marsalek, A., ... Prevenhieber, O. (2012). Android Market Analysis with Activation Patterns. In R. Prasad, K. Farkas, A. U. Schmidt, A. Liroy, G. Russello, & F. L. Luccio (Eds.), *Security and Privacy in Mobile Information and Communication Systems* (Vol. 94, pp. 1–12). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from [http://www.springerlink.com/index/10.1007/978-3-642-30244-2\\_1](http://www.springerlink.com/index/10.1007/978-3-642-30244-2_1)
- Tolman, E. C. (1936). Operational behaviorism and current trends in psychology. *Proceedings. 25th Anniv. Celebration of Inauguration of Graduate Studies, Univ. of Southern Califor*, 89–103.

- Tsai, J. Y., Egelman, S., Cranor, L., & Acquisti, A. (2011). The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study. *Information Systems Research*, 22(2), 254–268. doi:10.1287/isre.1090.0260
- Vermeir, I., & Verbeke, W. (2006). Sustainable Food Consumption: Exploring the Consumer “Attitude – Behavioral Intention” Gap. *Journal of Agricultural and Environmental Ethics*, 19(2), 169–194. doi:10.1007/s10806-005-5485-3
- Wilson, T. D., & Gilbert, D. T. (2005). Affective Forecasting. Knowing What to Want. *Current Directions in Psychological Science*, 14(3), 131–134. doi:10.1111/j.0963-7214.2005.00355.x
- Zukowski, T., & Brown, I. (2007). Examining the influence of demographic factors on internet users’ information privacy concerns (pp. 197–204). ACM Press.  
doi:10.1145/1292491.1292514

