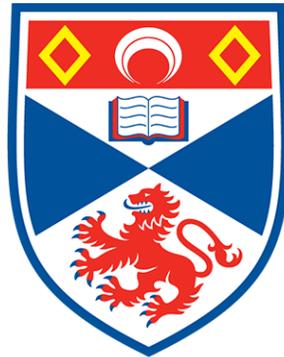


SURVEY ON DATA FRAGMENTATION ISSUES FOR USERS

Pooja Basavaraj Balekundargi

**A Thesis Submitted for the Degree of MSc
at the
University of St Andrews**



University of
St Andrews

2016

**Full metadata for this item is available in
St Andrews Research Repository
at:**

<http://research-repository.st-andrews.ac.uk/>

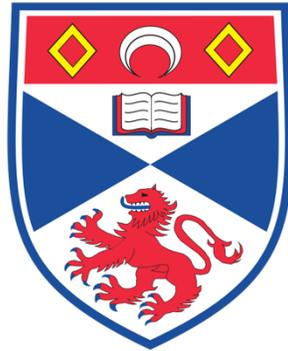
Please use this identifier to cite or link to this item:

<http://hdl.handle.net/10023/10004>

This item is protected by original copyright

UNIVERSITY OF ST ANDREWS

SCHOOL OF COMPUTER SCIENCE



IS 5199

Survey on Data Fragmentation Issues for Users

Pooja Basavaraj Balekundargi
150023241

Submitted in partial fulfilment of the requirements for the degree
of Master of Science in Information Technology
Date of submission - 19 August 2016

ACKNOWLEDGEMENTS

I would like to express my appreciation to those that helped me throughout the project.

I would like to thank my supervisors Dr. Graham Kirby and Dr. Miguel Nacenta from the School of Computer Science. They have provided immense guidance, support, and patiently helped me to achieve my project outcome, from the initial brainstorming ideas to the analysis.

I would like to thank Tom Kelsey for providing great suggestions and encouragement.

I would like to thank all the students and faculty for taking time out of their busy schedule to take part in this study or providing great suggestions and encouragement

Lastly, I would like to thank the teachers and staff members that taught me during my master's this year. I feel really honored to have studied in this great university.

ABSTRACT

Information is just data that is processed and given meaning. We live in a world where we use information to carry out the simplest tasks. We normally have different types of information, some very important and some less. Some of it is used for entertainment, some for the purpose of decision making, or just as a knowledge improvement tool. With new technologies being invented everyday, we tend to use different methods to store this information. We store information across various locations and across different services depending on personal preferences and needs. This causes information to get dispersed, most commonly known as data fragmentation. It is important to have efficient access to information at the time of need. However, although storing information in different places might aid in accessibility beyond geographical boundaries, it also hinders the process of finding and remembering the location of the right information. The following research study aims to gain insight into the methods used by individuals while storing information that is valuable to their daily activities. We also look at the mind sets of the users while they make decisions regarding storage methods. The empirical research carried out over the course of this dissertation provides insight on the causes and consequences of data fragmentation with regards to personal information. The findings have been analyzed and reported in readable format.

TABLE OF CONTENTS

1. INTRODUCTION	10
1.1 BACKGROUND	10
1.2 DEFINITIONS AND IMPORTANT TERMS	11
WHAT IS INFORMATION?	11
WHAT IS DATA FRAGMENTATION?	11
WHAT IS PERSONAL DATA?	12
WHAT IS PERSONAL INFORMATION?	12
WHAT IS PERSONAL INFORMATION MANAGEMENT?	12
2. PURPOSE OF THE PROJECT	13
2.1 OBJECTIVES	13
2.2 SCOPE	13
3. LITERATURE SURVEY	14
3.1 PAST TO PRESENT	14
3.2 BASED ON INFORMATION TYPE	15
3.2.1 EPHEMERAL INFORMATION	15
3.2.2 WORKING INFORMATION	15
3.2.3 ARCHIVED INFORMATION	16
3.3 BASED ON DIFFERENT FORMATS	16
3.4 BASED ON FOLDER ORGANIZATION	16
3.5 BASED COGNITIVE FACTORS	17
3.6 BASED ON BACK-UP.	18
4. CURRENT POPULAR TECHNOLOGIES AND SOLUTIONS:	18
4.1 CLOUD STORAGE SYSTEMS	18
A) THE GAP BETWEEN ONLINE AND OFFLINE	20
B) FILE STORAGE AND SYNCHRONIZATION	20
C) INTERFACE DESIGN	20
4.2 EXTERNAL STORAGE DEVICES AND BACK UP SOFTWARES	20
5. RESEARCH DESIGN AND METHODOLOGY	21
5.1 RESEARCH QUESTIONS	21
5.2 RATIONALE	21
5.3 DATA COLLECTION	21
A) QUESTIONNAIRE DESIGN	21
B) INTERVIEW DESIGN	22
5.4 PARTICIPANTS	22
5.5 ETHICS	23
5.6 FRAMEWORK FOR DATA ANALYSIS	23
A) INTERVIEW + QUESTIONNAIRE	23
B) SURVEY QUESTIONNAIRE	24
5.7 HYPOTHESIS TESTING	24

5.8 PROBLEMS FACED AND LIMITATIONS.	25
5.9 PROJECT PLAN	26
6. ANALYSIS: SURVEY QUESTIONNAIRE	27
RELATED SAMPLES	27
6.1 ACCESSIBILITY	27
6.1.1 FREQUENTLY USED PRIVATE INFORMATION AND FREQUENTLY USED PUBLIC INFORMATION.	28
6.2 RETRIEVAL	31
COMPARISONS ACROSS DIFFERENT TYPES OF DATA	32
6.2.1 FREQUENTLY USED PRIVATE INFORMATION AND FREQUENTLY USED PUBLIC INFORMATION.	32
6.3 SAFETY AND SECURITY	34
COMPARISONS ACROSS DIFFERENT TYPES OF DATA	35
6.3.1 NON-FREQUENTLY USED PRIVATE INFORMATION AND FREQUENTLY USED PRIVATE INFORMATION.	35
6.3.2 FREQUENTLY USED PRIVATE INFORMATION AND FREQUENTLY USED PUBLIC INFORMATION.	36
ONE SAMPLE TESTS	38
6.4 ACCESSIBILITY	38
6.4.1 I FIND THAT DOCUMENTS STORED IN DIFFERENT FORMATS ARE INACCESSIBLE BY DIFFERENT DEVICES	39
6.5 RETRIEVAL	41
6.5.1 I SOMETIMES HAVE MULTIPLE COPIES OF THE SAME INFORMATION IN DIFFERENT STORAGE LOCATIONS	42
6.6 SAFETY AND SECURITY:	43
6.6.1 IT WOULD BE TROUBLING AND/OR CAUSE ME DISTRESS IF MY DATA SUDDENLY BECOMES AVAILABLE	44
ADDITIONAL FACTORS:	45
6.6.2 STORING DATA IS EXPENSIVE	45
6.6.3 MY DATA IS FRAGMENTED/SPREAD OVER TOO MANY DIFFERENT SERVICES AND/OR DEVICES	46
6.6.4. USAGE OF MULTIPLE METHODS	47
7. ANALYSIS: INTERVIEWS	48
7.1 HOW MANY INFORMATION ITEMS DID THEY FIND?	48
7.2 WHERE DO PEOPLE STORE DIFFERENT TYPES OF DATA?	49
7.2.1 PUBLIC INFORMATION	49
7.2.2 PRIVATE INFORMATION	50
7.3 WHAT ARE THE FACTORS THAT INFLUENCE INDIVIDUALS WHILE MAKING DECISIONS ABOUT STORING INFORMATION?	52
7.3.1 ENVIRONMENTAL FACTORS	52
7.3.2 EMOTIONAL FACTORS	53
7.3.3 COGNITIVE FACTORS	54
8. RESULTS AND DISCUSSION	56
9. PERSONAL REFLECTION	57
10. RECOMMENDATIONS AND FUTURE WORK	57
11. CONCLUSION	58
12. REFERENCES	58
13 APPENDIX	61

- 13.1 FULL ETHICS FORM**
- 13.2 ETHICS APPROVAL FORM**
- 13.3 QUESTIONNAIRE**
- 13.4 PARTICIPANT INFORMATION SHEET**
- 13.5 PARTICIPANT CONSENT FORM**
- 13.6 DEBRIEFING FORM**
- 13.7 INTERVIEW QUESTIONS**
- 13.8 RELATED SAMPLE TESTS: ACCESSIBILITY**
- 13.9 RELATED SAMPLE TESTS: RETRIEVAL**
- 13.10 RELATED SAMPLE TESTS: SAFETY AND SECURITY**
- 13.11 ONE SAMPLE TESTS: ALL QUESTIONS**
- 13.12 FREQUENCY DISTRIBUTIONS OF ALL DATA**
- 13.13 GENDER DEMOGRAPHICS OF QUESTIONNAIRE DATA**
- 13.14 AGE DEMOGRAPHICS OF QUESTIONNAIRE DATA**
- 13.15 INCOME DEMOGRAPHICS OF QUESTIONNAIRE DATA**

Table of figures

FIGURE 1: EXAMPLE FOLDER ORGANIZATION BY SUBJECT FRAGMENTATION BY A STUDENT[10]	17
FIGURE 2: BROWSER COMPATIBILITY COMPARISONS OF GOOGLE DRIVE	19
FIGURE 3 :PICTURE SHOWING 4 CATEGORIES OF PERSONAL INFORMATION	22
FIGURE 4: HORIZONTAL TREE DIAGRAM OF DATA COLLECTION METHODS.	23
FIGURE 5: GANTT CHART SHOWING A BRIEF PLAN OF ACTION	26
FIGURE 6: BAR CHART SHOWING PARTICIPANTS RESPONSES ON ACCESSIBILITY OF DIFFERENT TYPES OF INFORMATION.	28
FIGURE 7: CHART SHOWING ACCESSIBILITY OF NON-FREQUENTLY USED PRIVATE INFORMATION WITH FREQUENTLY USED PRIVATE INFORMATION	29
FIGURE 8: ACCESSIBILITY OF NON-FREQUENTLY USED PUBLIC INFORMATION AND FREQUENTLY USED PUBLIC INFORMATION	30
FIGURE 9: BAR CHART SHOWING AN OVERVIEW OF THE PARTICIPANTS VIEWS ON THE QUICK RETRIEVAL OF DIFFERENT TYPES OF INFORMATION WITH THEIR CURRENT STORAGE SETUP.	32
FIGURE 10:CHART SHOWING RETRIEVAL OF FREQUENTLY USED PRIVATE INFORMATION AND FREQUENTLY USED PUBLIC DOCUMENTS	33
FIGURE 11: GROUPED BAR CHART SHOWING AN OVERVIEW OF THE PARTICIPANTS VIEWS ON SAFETY AND SECURITY OF DIFFERENT TYPES OF INFORMATION WITH THEIR CURRENT STORAGE SETUP.	35
FIGURE 12: SAFETY AND SECURITY OF NON-FREQUENTLY USED PRIVATE INFORMATION AND FREQUENTLY USED PRIVATE INFORMATION.	36
FIGURE 13: SAFETY AND SECURITY OF FREQUENTLY USED PRIVATE INFORMATION AND FREQUENTLY USED PUBLIC INFORMATION	37
FIGURE 14: HORIZONTAL BAR CHART SHOWING PARTICIPANTS VIEWS ON FILE FORMAT CAUSING INFORMATION TO BE INACCESSIBLE.	40
FIGURE 15: CHART SHOWING THE FREQUENCIES AND PERCENTAGE FREQUENCIES OF RESPONSES ABOUT MULTIPLE COPIES OF DATA	42
FIGURE 16:BAR CHART SHOWING PARTICIPANTS VIEWS ON DISTRESS CAUSED BY LOSS OR LEAK OF DATA.	44
FIGURE 17:BAR CHART SHOWING THE PARTICIPANTS VIEWS ON EXPENSES OF STORING DATA.	45
FIGURE 18: BAR CHART FOR PARTICIPANTS VIEWS ON DATA FRAGMENTATION	46
FIGURE 19: STALKED BAR SHOWING THE USAGE OF DIFFERENT STORAGE DEVICES AND SERVICES	47

FIGURE 20: FACTORS CAUSING DATA FRAGMENTATION. 52

FIGURE 21: INTERVIEW PARTICIPANTS VIEWS ON ACCESSIBILITY 55

Table of tables

TABLE 1: SYSTEMS INTEROPERABILITY OF CLOUD STORAGE APPLICATIONS. [39] 19

TABLE 2: CROSS TABLE CONTAINING 6 GROUPS OF INFORMATION COMPARISONS. 31

TABLE 3 : CROSS TABLE CONTAINING 6 GROUPS OF INFORMATION COMPARISONS 34

TABLE 4 : CROSS TABLE CONTAINING 6 GROUPS OF INFORMATION COMPARISONS 38

TABLE 5: TABLE SHOWING THE FREQUENCIES AND PERCENTAGE FREQUENCIES OF INFORMATION FOR ACCESSIBILITY. 39

TABLE 6: TABLE SHOWING FREQUENCIES AND PERCENTAGE FREQUENCIES OF PARTICIPANTS RESPONSES ON RETRIEVAL 41

TABLE 7: TABLE SHOWING FREQUENCIES AND PERCENTAGE FREQUENCIES OF PARTICIPANTS RESPONSES ON SAFETY AND SECURITY 43

TABLE 8: NUMBER OF ITEMS RETRIEVED BY THE PARTICIPANTS IN 15 MINUTES 48

TABLE 9: TABLE SHOWING MOST COMMONLY USED STORAGE METHODS FOR PUBLIC INFORMATION. 50

TABLE 10: TABLE SHOWING MOST COMMONLY USED STORAGE METHODS FOR PRIVATE INFORMATION. 50

1. INTRODUCTION

With the advent of technology, there is colossal amounts of data being created and stored on a daily basis. The following data is then stored as information files and used by organizations and individuals for efficient decision making. However, these information files are stored in different forms i.e. documents, text files, images, video files, appointment reminders, bookmarks etc. and stored on different types of devices and services i.e. public/private cloud storage services, mobile phones, paper based storage, computer devices, storage servers, external hard drives etc. This results in information being dispersed across various locations thus causing data fragmentation. Dispersal of information across several devices and services makes it accessible from different locations. However, inefficient management of an individual's storage setup can have an adverse effect in retrieving this information in the time of need.

In reality, making decisions on choices of storage methods is harder due to the variety of options available. It is important to consider factors like safety and security, back up and storage capability of the device in question while making these decisions. Information is used for all purposes by individuals of all generations. There are several forms in which we use our data. For example, one usually begins the day with reading the news either in a paper based format like a newspaper or in a digital text format from a mobile phone, tablet, desktop computer, television. This is followed by using data at work, or at school or while resting at home. All these activities, require us to come in contact with information that we access through information devices. Disruption of these activities can inevitably cause disturbance to one's entire likelihood. Loss of data or difficulties in retrieving them is a concern of many sectors.

In recent years, there have been major technological improvements in personal information storage devices and services but not enough improvement in management and organization of information stored across these devices and services to help increase the speed of retrieval and accessibility. We live in a world that is data driven. Whether, it is the education sector, businesses or even individuals carrying out their daily activities at their work environment or home. It is important to realize that information not accessible in the time of need can have a significant impact on the efficiency and productivity of daily activities. For this reason, this research project attempts to gain insight into issues caused by data fragmentation, with a specific focus on individuals.

1.1 Background

There have been several studies on fragmentation of data across organizations. A study evaluated the impact of data fragmentation on medical information of patients across several health care institutions[1]. Another study addressed the usability problems of accessing open data across private and public organizations caused by fragmentation[2]. Hermjakob's study on proteomics data addresses the problems of fragmentation across academic publications[3]. There is not enough research on data fragmentation across several applications and devices with regards to personal information belonging to individuals. With multiple options available to store

data, individuals generally choose devices based on their knowledge of and confidence in the storage device and/or service.

Personal information is very easy to create, store and even easier to destroy. Often this kind of information is created in a hurry which results in unclear and inconclusive naming of information, ultimately causing problems in retrieving them for later usage. This suggests that we need improvement in organization methods and systems which will ease the process of finding and retrieving information. In this study, we focus on information management problems caused due to data fragmentation of personal information and look at various devices and services that are currently being used by individuals.

1.2 Definitions and important terms

What is information?

Information is processed data that is used to carry out day today activities and as source for further learning.

“Information is a source of learning. But unless it is organized, processed, and available to the right people in a format for decision making, it is a burden, not a benefit.”- William Pollard [4].

What is data fragmentation?

Individuals generally store data/information across various devices and applications in order to make room for storage. However, this causes data to get dispersed across different locations. This is known as data fragmentation. The term data fragmentation takes different meaning depending on the field of research. It is most commonly used in the field of database management systems where data fragmentation is the breaking up of a single database into multiple fragments. These fragments are then stored individually and accessed by the database management system(DBMS) based of requirement[5].

Fragmentation of data can take place within a device spread across various applications i.e. mobile phone that contains email, address book, instant messages, applications like google drive, drop box as well as across various devices like USB sticks, CDS, mobile phones, laptops. There are many problems that are caused by storing data on this many sources which can take place on an organization level as well as individual level[6]. In the following report, we only look at fragmentation of personal information faced by individual's when their information is fragmented across email messages, paper documents, pictures etc.

What is personal data?

According to the data protection act by the ICO –

“Personal data means data which relate to a living individual who can be identified –

(a) from those data, or

(b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller,

and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual[7].”

What is Personal information?

Personal information is information that is valuable to an individual. It can be further divided into information that is private, the loss or leak of which could have the potential of causing danger or embarrassment to an individual and public information, the loss or leak of which will not have the potential of causing danger or embarrassment.

Personal information is generally used to help the development of an individuals learning technique or as a reminder of the past and the future. This includes emails, notes, bookmarks, references, files, pictures, receipts, home videos, text messages, meeting reminders, to do lists etc.[8]. It is different types of information that are of value to an individual and are worthy of archiving[9].

What is Personal information management?

Data fragmentation causes disorganization of personal information and has adverse effects on the process of personal information management (PIM). Therefore, it is important to define it. PIM is the management and organization of personal information of a single person[10] [11].It involves activities like storing, finding, retrieving, organization and maintenance of an individuals information[12], [13]. Over the years, research related to personal information management has also been termed as personal knowledge management[14].Information systems that are used for the purpose of PIM are known as personal information systems[15].

2. PURPOSE OF THE PROJECT

2.1 Objectives

Primary

- Identify and measure of severity of data fragmentation issues inducing inefficient personal information management.
- Outline the behavioral implications that lead to data fragmentation problems and providing suggestions to reduce the impact.
- A written analysis of the causes and consequences of data fragmentation on different types of users.
- Discuss the impact of physical distribution of storage devices hindering easy locating of data.

Secondary

- Provide a taxonomy of personal information storage systems while outlining the potential advantages and disadvantages.
- Identifying the security implications of data fragmentation and the need for back up.
- Provide recommendations regarding technological advancements that will help decrease the severity of the issues.
- Critically analyze the existing research in the following area and summarize the findings.

2.2 Scope

The following study will shed light on the the severity of data fragmentation and the stakeholders include:

1. Individuals:

Individuals who have incomplete knowledge of personal information storage services and are looking for clarity in integration of these services with their personal information storage setup. Users who are experiencing inconvenience regarding finding, retrieving and accessing information with their current information storage setup.

2. Businesses:

Businesses that are looking to create or develop information management systems that cater to the needs of their customers.

3. Researchers:

Researchers in the field of human-computer interaction and related fields.

3. LITERATURE SURVEY

3.1 Past to present

There have been several studies addressing issues related to the improper management of personal information. Evidence of data fragmentation and disorganization was noticed quite early on. In the past, individuals faced problems with filing and piling of paper documents on their desks. Large volumes of paper based documents used in their daily life activities and work environment were being stored and retrieved on a regular basis. With the digital era, paper based documents still do form a major part of information documents but have subsequently decreased. In 1983, Malone conducted a study on organizing patterns of paper document on desks in offices. He found that people didn't want to put things away to organize them because visually accessible documents are easily remembered. He also found differences in cognitive behaviors of people while making decisions about what is to be organized and how. He found that there were similar patterns in the behaviors of storing and retrieving paper documents and electronic documents. He then went on to propose that filing systems should integrate file classification functions based on similar patterns[16]. Over the years, studies of similar fashion have been conducted to help decrease data fragmentation problems in systems for improved management and organization with regards to paper-based documents[17], bookmarks[18], emails[19][20], memos, appointment reminders[21]. However, very few studies focus on cross tool organizational strategies and cross device organizational strategies. There have also been studies that address such issues by providing technological solutions. In 1999, Huhn's suggested building a personal ontology manager which would act as a personal search engine for all online documents[22]. In 2001, Gordon bell proposed a project called cyber all which would help organize, index, store and retrieve personal information for personal and work-related usage. The information items included were CDS, letters, memos emails, photos, papers[23]. Following Bells work, an SQL based platform for organizing and accessing information was proposed in 2006. However, the project required the users to manually assign file labels and classification for all pictures, documents, emails which should be done automatically[24]. In 2010, a study suggested that most storing decisions are either user-driven or system-driven and problems with user-driven decisions can be transferred to systems that make makes suggestions for appropriate locations[25].

Boardman addressed the cross tool perspective with the combination of emails, web bookmarks and documents [26]. In the last few decades, managing information has become a difficult task due to the availability of multiple storage devices and services. A study addressed the management of household data like photos and drawings made by children by proposing the "family archive device" that would organize data created by all the members of the family while carrying out household activities[27].

3.2 Based on Information Type

The differentiation of information management styles for professionals like librarians, employees and individuals who manage personal information has been done by Bergman[15]. There is a difference in the way an individual deals with information that related to private life as opposed to information that one needs as part of their work routine.

In 1982, Cole categorized personal information into three differentiable types for filing and organizing office documents as action information, personal work files and archived information[28]. Similar categories were adopted by a studies in 1995 to categorize information types for filing, organizing and finding documents in a desktop setting at the work place which were ephemeral, working and archived information[29][30].

3.2.1 Ephemeral information

Ephemeral data is most widely known as temporary information that is regularly created by the user and is most frequently received from an outside source like online articles, new letters. This might also include notes, to do lists, reminders etc. The results from Andersons study showed that the main problem for users while organizing ephemeral data was the short life span, therefore the users didn't have a set procedure to them. Users generally pile such information on the desktop screen as it is more visually accessible and acts as a reminder[29]. In 1997, a study on customers service executives on storing ephemeral data showed that they would rather write information like phone numbers and to do lists down on paper documents rather than computers to save space on their information systems and out of habit[31]. This causes information to be dispersed everywhere. Boardman and sasse found that some users prefer to file their ephemeral data like emails while the rest just leave it unfiled[32].

Barreau found that the most problematic ephemeral information to be emails. However, the study was conducted on an organizational level and most managers received more that 300 emails a day which made it hard for them to find and retrieve the important ones. Although some managers views emails as ephemeral data, these emails generally contained important working information that can be used as a reference tool later on[33]. However, the distinction of the data to be categorized into ephemeral data is quite relative and mainly depends on the user which also suggests that storage systems require flexible organizing procedures that cater to all users.

3.2.2 Working information

Studies found that user's interactions with working information is quite well structured as compared to ephemeral and achieved data, as this information is used quite regularly and needed in the completion of present projects. Documents like memos, research papers, project plans are most widely known as working information. The frequent nature of use of these documents aids in forced categorization and organization[29]. However as the project comes close to completion the location of these documents become less important and sometimes is easily ignored[30]. Therefore, it is hard to distinguish information based on type and mostly depends on human behavior and situation. On an organizational level, the volume of working

information is massive but managed quite efficiently. However, working information eventually turns dormant and if not deleted makes use of unnecessary resources. For example, in Bureau's study, a manager said that she sends information needed while preparing a speech to her own email and copy pastes it to her messages. such information is not deleted after use and are completely forgotten and useless[33].

3.2.3 Archived information

The main purpose of archived information is knowledge seeking from data that will be used for a longer duration of time, say 6 months to one year or more. It is generally non-frequently used and stored for specific purposes only. Examples for archived information are health insurance documents, old receipts, project reports, thesis. These documents are of great value and damage or loss would have undesired consequences. They are the final product from the information gained by using working data and ephemeral data. Bureau's study showed that users spent a lot of time deciding a structure for these documents because of how important they are[30]. However, once the shelf life period of such documents runs out, users generally forget about the placement of these documents.

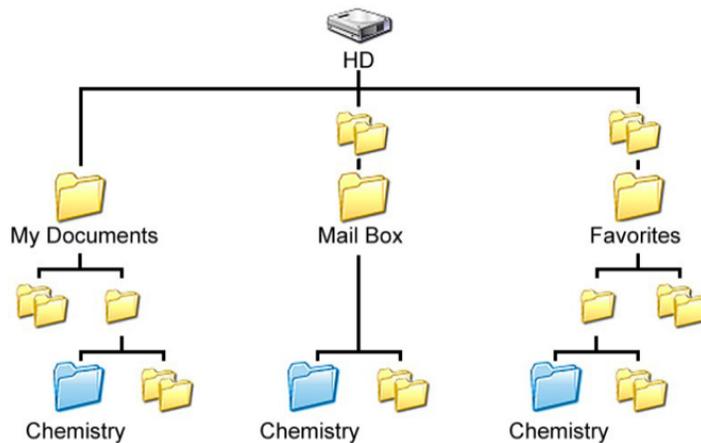
3.3 Based on different formats

There also a difference in analyzing if people store information together based on projects that are similar in context or depending on the formats[34]. Studies have shown that users generally ignore the format of the documents and store them according to the context of the projects that they are currently working on. When asked if users organize files based on formats or projects, users choose projects and mentioned that they would organize by formats only if the interface of the storage setup does not allow them to store different formats together[10]. However, this causes inconvenience if different formats in one folder cannot all be read by a device or application due to incompatibility issues.

3.4 Based on folder organization

Filing systems generally have folders which contain files with different titles belonging to a main directory, the hierarchy then generally contains sub-categories in the main file. Although this provides a structure to the storage system, it also becomes a dilemma if the directories are not well mapped[35]. Individuals generally tend to have sub-folders within main folders as a result of having excess of information, which results in a "divide and conquer" problem[36]. The main idea is to organize information in ways that aid in later retrieval. However due to the volume of information that we generally possess, the depth of the hierarchies increases extensively. This results in dispersal of different types of information in different folders.

Figure 1: Example folder organization by subject fragmentation by a student[10]



Folders are generally named by using leading characters like (“A”, “01”) based on the usage pattern. These folder names are sometimes to remember. With the progression of technology, there many extended hierarchies of folders in different applications like webpage bookmarks, emails, messages for different types of information like private information public information, work related information and different types of formats[36][10].Therefore, if not properly managed, retrieval can become a time-consuming task.

3.5 Based cognitive factors

A study conducted at the Tel Aviv University addressed the issues of fragmentation in PIM by suggesting information should be categorized based on the subjective classification principle which implies that all information related to the same subject or topic should be classified in the same folder[15] while other studies focus on search functions based on importance and value of the information[32]. Users generally use different strategies while choosing the PIM methodology. It is a well-known fact that users will use systems that are readily available to them for storing information instead of carefully employing a system that will help them with efficiency in the long run. The cognitive energy [37] and cognitive state[33] has a major influence on the saving , filing , classifying , acquiring, organizing , accessing and retrieving information. It drives an individual to choose the personal information system of their choice. Therefore PIM systems should be alterable and flexible to the requirements of the individuals using them.

Kwasnik suggested that retrieval of documents does not only depend upon the attributes i.e. type of document but also the context of the document i.e. the situation at which the document was retrieved. Although it is not possible to judge an entire population on certain results many similarities in behavior patterns were noticed. It is the decision making ability that influences one while choosing a location to store documents[38].

3.6 Based on back-up.

Most studies pointed out that there are several approaches to preserving valuable information. Some user's relied on self back-ups and rarely assessed the safety of their information while others have a set location and procedure to back-up strategies.

It was pointed out during a study that the users did not trust the technologies used for backup. They also had different strategies for backing up private information like pictures, videos and work-related information. While making decisions on what information to back up, users generally categorize them by the level of information, the financial cost of storing this information, value it brings to the customers and organization. However, this study was conducted on an organizational level and not an individual level. [12]

4. CURRENT POPULAR TECHNOLOGIES AND SOLUTIONS:

Various technologies do currently address the problems related to data fragmentation issues for individuals. The following chapter will be a brief analysis of services that are currently available for users.

4.1 Cloud storage systems

There are currently several cloud storage systems available for users to choose from. For example, Microsoft OneDrive, Google Drive, Dropbox, Box that allow online storage information on the cloud as well as file sharing, online documents editing. The advantages of these storage systems is that as long as one is connected to the internet it is fairly easy to access data from anywhere around the world. These systems automatically backup all online edits and overwritten changes. For example, edits made in desktop applications of drop box and google drive are automatically updated to the newer version. This eliminates the process of making edits offline and accessing the application, delete the older version and uploading of the newer version. However, google drive documents edits uses google tools which does not allow sharing formats. Therefore, in order to edit the document in a different program one will have to export it prior to sharing. Also with regards to storage capabilities, it is an obvious fact that most users use Gmail. Google drive shares storage space with Gmail, which implies that if there is an ample amount of emails, the storage capacity of the drive will be limited[39].

The main problem with these applications as observed in table 1 is the file accessibility across multiple devices. Certain Online storage services are not accessible by operating systems like Linux, windows for computers, windows phone for mobile phones and android. This leads to data being irretrievable if the physical location of the computer compatible with the application is not accessible.

Table 1: Systems interoperability of cloud storage applications. [39]

	Google Drive	One drive	ICloud	Dropbox	Box	Amazon cloud drive
Windows	Yes	yes	yes	yes	Yes	Yes
Windows phone	No	yes	no	no	Yes	No
Mac	Yes	yes	yes	yes	Yes	Yes
IOS	yes	yes	yes	yes	Yes	Yes
Android	yes	yes	no	yes	Yes	Yes
Linux	No	No	No	Yes	No	No

In addition, there are so certain file synchronization problems. For example, if one uses a safari browser on a Mac OS, they will not be able to upload complete folders. A work around for such a problem, is to download the google chrome browser or download the desktop application for google drive. The disadvantage of such a situation is wastage of time spent on uploading a single file as well as time and energy spend in figuring out workarounds to get the task of uploading done faster.

Figure 2: Browser compatibility comparisons of Google Drive

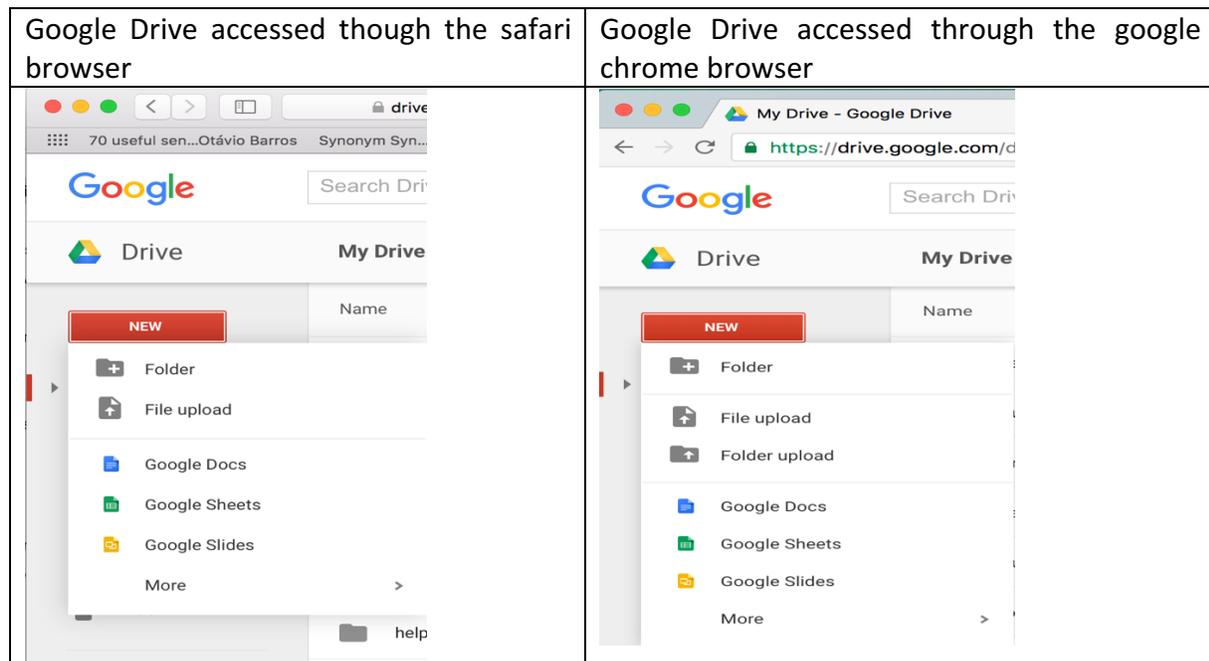


Image captured by the researcher

a) The gap between online and offline

The download of desktop applications like Mendeley desktop, Google Drive, OneDrive has many benefits. These files can be accessed by a computer or a phone while the device is not connected to the internet. The problem with backing up everything online is its inaccessibility when offline which has been improved upon in the last few years.

b) File storage and synchronization

With the advantage of having several file synchronization and file saving applications that cater to the needs of different users, the main disadvantage if this is that information stored in a certain application might not be compatible with another application.

c) Interface design

For files to be easily accessible, the visual design interface should allow users to choose the view based on preference. For example, while google drive allows users to choose between grid view and list view, drop box only allows list view. The disadvantages of this are difficulties in search and scroll functions in case of large amounts of data.

4.2 External storage devices and back up softwares

The most popularly used external storage devices like compact disks, digital versatile disks, blue-ray disks, solid state drives, flash memory cards, USBs, high capacity external drives are the traditional methods generally used to back up information. With the advancements in the cloud, the usage levels of these devices has decreased over the years. However, they are still used by many due to the simplicity and ease of use and mistrust of the cloud. These are most commonly known as direct attached storage (DAS).

The main disadvantage of DAS devices is limited scalability, because of which users are required to invest in additional storage devices and save data across these multiple devices. A possible solution here is using a network attached storage for individual usage. They are reliable, scalable and affordable. NAS also enables redundant array of inexpensive disks. This feature allows the use of multiple drives. NAS can be accessed from anywhere and it is storage locally for immediate retrieval[40].

The most important advantage is the scalability which does not apply to DAS devices. Although, the initial capital investment in NAS devices is higher, it has major benefits in the long run. Most commonly used NAS devices used by home users for personal information are QNAP TS-251, Synology DS214, WD My Cloud EX2. Options for backs are inbuilt back-up software's like time-machine for mac only and CrashPlan for Mac, Windows and Linux which allow online and offline back.

5. RESEARCH DESIGN AND METHODOLOGY

The following research study aims to gain insight on data fragmentation problems for individual users. The empirical research intends to answer the research questions while finding evidence of the severity data fragmentation issues. This chapter explains the design of research methodology which was carried out during the course of the project while addressing the data collection methods, methods of analysis and limitations. The research methodology includes a mixture of quantitative and qualitative analysis. Quantitative analysis consists of hypothesis testing and the qualitative analysis consists of interviews. Further description on research methodologies is stated in the framework for data analysis section.

5.1 Research questions

The research questions have been generated by knowledge in the existing literature and missing information. There are several studies that focus on cross-tool fragmentation and organization of personal information but not enough research related to cross-devices and cross-tool issues.

1. How do people treat different types of information?
2. Is data fragmentation causing inconvenience to users?
3. What are the factors that are influencing individuals to disperse their data across various devices, applications and services?
4. Are there any behavior patterns that are causing these problems? Can change of behavior decrease the degree of the problems associated with data fragmentation?

5.2 Rationale

The preliminary research involved getting a clear understanding of data fragmentation followed by brainstorming of ideas and making a preliminary literature survey which would help in producing the main research questions. After the main research questions were constructed the next step was designing the study and making decisions about the type of data collection. The results acquired from empirical analysis would allow us to analyze the awareness of the content, structure and location of the information that is dispersed across various devices. A mixture of questionnaire data and interview data will provide a comprehensive out-look of individuals experiences towards data fragmentation issues.

5.3 Data Collection

Primary data was collected by 2 methods. i.e. The deployment of a questionnaire survey on an online survey platform and a task-based semi structured interview.

a) Questionnaire design

The design of the questions on the questionnaire was heavily influenced by the preliminary research as well as primary and secondary objectives of the study. Since the focus of the

dissertation is on the behavior patterns of storing information across various devices and applications, the types of information have been categorized into 4 distinct types based on the ideas acquired by brainstorming with the supervisors.

Figure 3 :Picture showing 4 categories of personal information

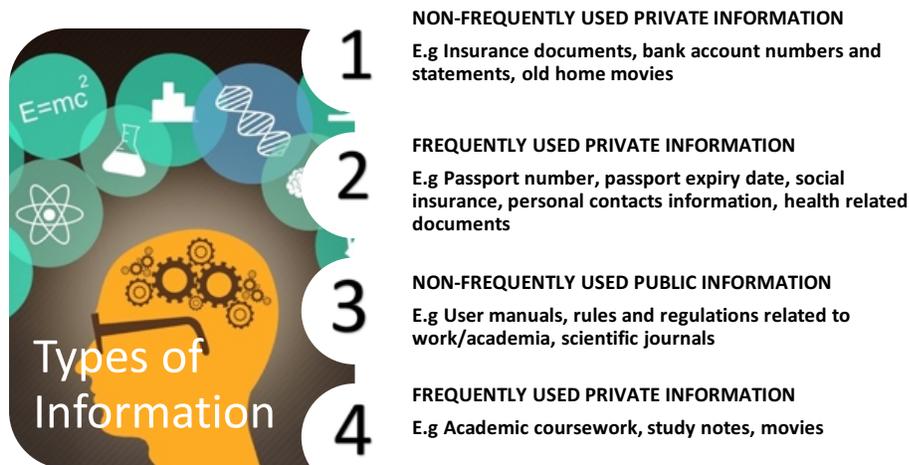


Image courtesy by Freerangestock.com and edited by researcher

The questions on the questionnaire were formed to answer the hypothesis. A five-point Likert scale was used as ordinal data .i.e. (i) Strongly agree (ii) Agree (iii) Neutral (iv) Disagree (v) Strongly disagree[41]. The questionnaire contains a set of 10 questions with open ended responses.

b) Interview design

The Interview was task based and semi-structured. The participants were asked to fill out the same questionnaire before commencing the task-based session. At the end of the the questionnaire, participants were provided with a list of projects related to their personal information storage activities. They were then asked to retrieve and access them. The task was given a timeline of 15 minutes and the number of items that the participants were able to retrieve were noted. After the completion of the task followed an open-ended interview to understand the overall experience of the participant while finding these documents and files. Only data related to user experience was recorded on an audio recording device. Participants were allowed to use any device that was available to them which was situational to the location of the study. As the study focuses with certain aspects of behavioral science, the location of the study depended on the daily activities of respective participants being students or employees.

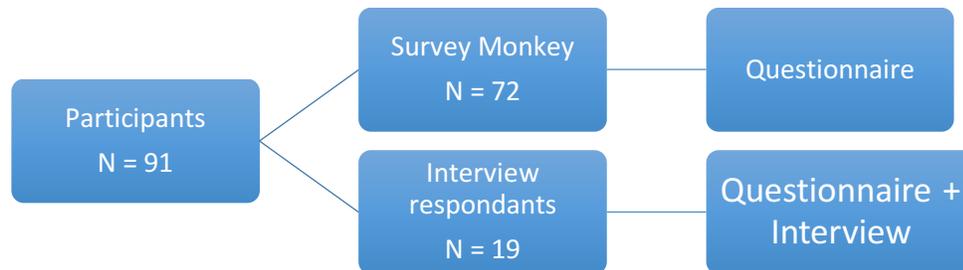
5.4 Participants

Convenience sampling was used for recruiting participants from the online survey. The audience were chosen from survey website and the participants were financially compensated. The age ratio, sex ratio and income ratio of the participants with complete responses have been attached

in the appendix. The data from this part of the study was collected by respondents that were recruited on Survey Monkey[42].

Convenience sampling was also used for the participants required for the interviews and email responses from the university. The second part of the study was a task based interview. The participants that took part in the interview were also asked to fill the questionnaire making the total number of respondents to a total of 91.

Figure 4: Horizontal tree diagram of data collection methods.



5.5 Ethics

The participants were chosen on a voluntary basis. The respondents of the interview as well as the questionnaire were provided with a participation information sheet and a consent form and a debriefing form. They were informed about the procedure of the study and risks involved before commencing the study. The data collected from the interviews was stored in an encrypted format using VeraCrypt and the access was provided to the researcher only. The ethics reference number for this research project is CS12314.

5.6 Framework for data analysis

There are 2 approaches to the analysis of the collected data. Both approaches have been analyzed separately and the results come together in the discussion.

a) Interview + questionnaire

The analysis of the interview data is descriptive. The interviews were audio taped and are being analyzed in an exploratory research format. The concepts have been categorized based on relevance and repetition. The data has been coded based on conceptualization of themes. As the participants also answered the questionnaire, the analysis of the interview will include the codes of the interview results as well as a bar chart with questionnaire answers.

b) Survey Questionnaire

The analysis of the survey data is quantitative where we try to answer our research questions by hypothesis testing. As the data from the questionnaire is ordinal and not normally distributed the Wilcoxon signed rank test has been selected as the primary non-parametric test.

5.7 Hypothesis testing

We used two variations of hypothesis in the following study. The hypothesis testing has been carried out by using SPSS which is a standardized statistics software package.

5.7.1 Related samples:

Overview: The following tests answer to a two-tailed hypothesis to check the median difference between two related samples from data collected from a 5-point Likert scale. If the null hypothesis is rejected, then the alternate hypothesis will be accepted which will mean that the two samples are significantly different.

H_o - The median of differences between (Related Sample A) and (Related sample B) equals 0.

H_a - The median of differences between (Related Sample A) and (Related sample B) does not equal 0.

Testing criteria:

- i. **Statistical Test:** For the following hypothesis, the non-parametric Wilcoxon signed ranked test has been chosen as it pertains to within subjects i.e. same participants are used under different conditions and the data are ordinal in nature.
- ii. **Significance level:** Let $\alpha = .05$. Alpha level has been set to the standard level.
- iii. **Reported statistics:** Median, P-value, Z score and effect size.
- iv. **Rejection criterion:** The hypothesis is two-tailed as the direction of the difference is not assumed beforehand. Therefore, if the P – value is found to be less than or equal to the Alpha value ($\alpha=.05$), the null hypothesis will be rejected thus accepting the alternate hypothesis.

5.7.2 One sample tests:

Overview: The following tests answer to a two-tailed hypothesis to check if the median is 3 which on a 5-point Likert scale falls on neutral. If the null hypothesis is rejected, then the alternate hypothesis will be accepted and which will mean that the the median falls on either end of the

extremes. We will then be reporting the median to check the direction of the data.

H_0 - The median of (one sample A) equals 3.00

H_a - The median of (one sample B) does not equal 3.00

Testing criteria:

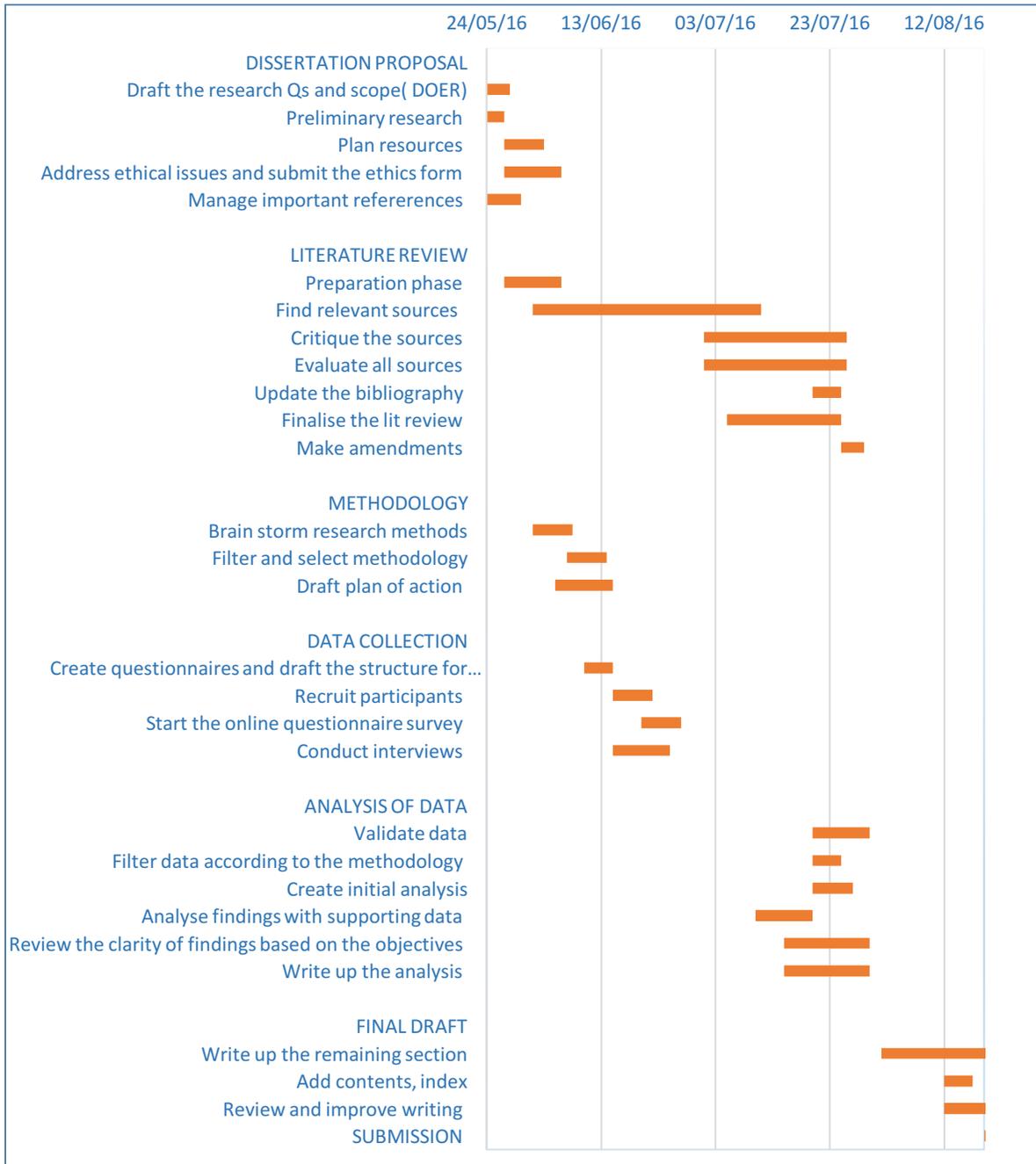
- i. **Statistical Test:** For the following hypothesis, the non-parametric Wilcoxon signed ranked test has been chosen as it pertains to a within subjects i.e. same participants are used under different conditions and the variables are ordinal in nature.
- ii. **Significance level:** Let $\alpha = .05$. Alpha level has been set to the standard level.
- iii. **Reported statistics:** Median, P-value, Z score and effect size.
- iv. **Rejection criterion:** The hypothesis is two-tailed as the direction of the difference is not assumed beforehand. Therefore, if the P – value is found to be less than or equal to the Alpha value ($\alpha=.05$), the null hypothesis will be rejected thus accepting the alternate hypothesis.

5.8 Problems faced and limitations.

1. **Schedule constraints:** Due to the high volume of data being collected, there were several time constraints to the completion of the analysis.
2. **Recruiting constraints:** The initial sampling size for the interviews was 20. All participants from the interview were asked to fill out the questionnaire before commencing the interview. However, due to technical difficulties a participant did not answer the questionnaire. Therefore, the sample size for the interviews is currently 19.
3. **Financial constraints:** A larger sample size for the online participant was not chosen due to financial constraints. Also, random random sampling of participants from the questionnaire was not chosen for the following reason.
4. **Sampling bias:** We fail to use probability sampling techniques due to time constraints and financial constraints which limits our results from applying to broader generalization.

5.9 Project Plan

Figure 5: Gantt Chart showing a brief plan of action



6. ANALYSIS: SURVEY QUESTIONNAIRE

The following chapter will contain a detailed analysis of the data collected from the participants of the questionnaire survey. The participants were asked to rate on a 5-point Likert scale on their experiences in finding different types of information and the data is represented in a bar chart format. The main purpose of the analysis is to answer the research questions. The analysis tries to answer the research questions by analyzing the data collected by each question in the survey. In the first part of the survey, we address the research questions

1. How do people treat different types of information?
2. Is data fragmentation causing inconvenience to users?
3. What are the the factors that are influencing individuals to disperse their data across various devices, applications and services?
4. Are there any behavior patterns that are causing these problems? Can change of behavior decrease the degree of the problems associated with data fragmentation?

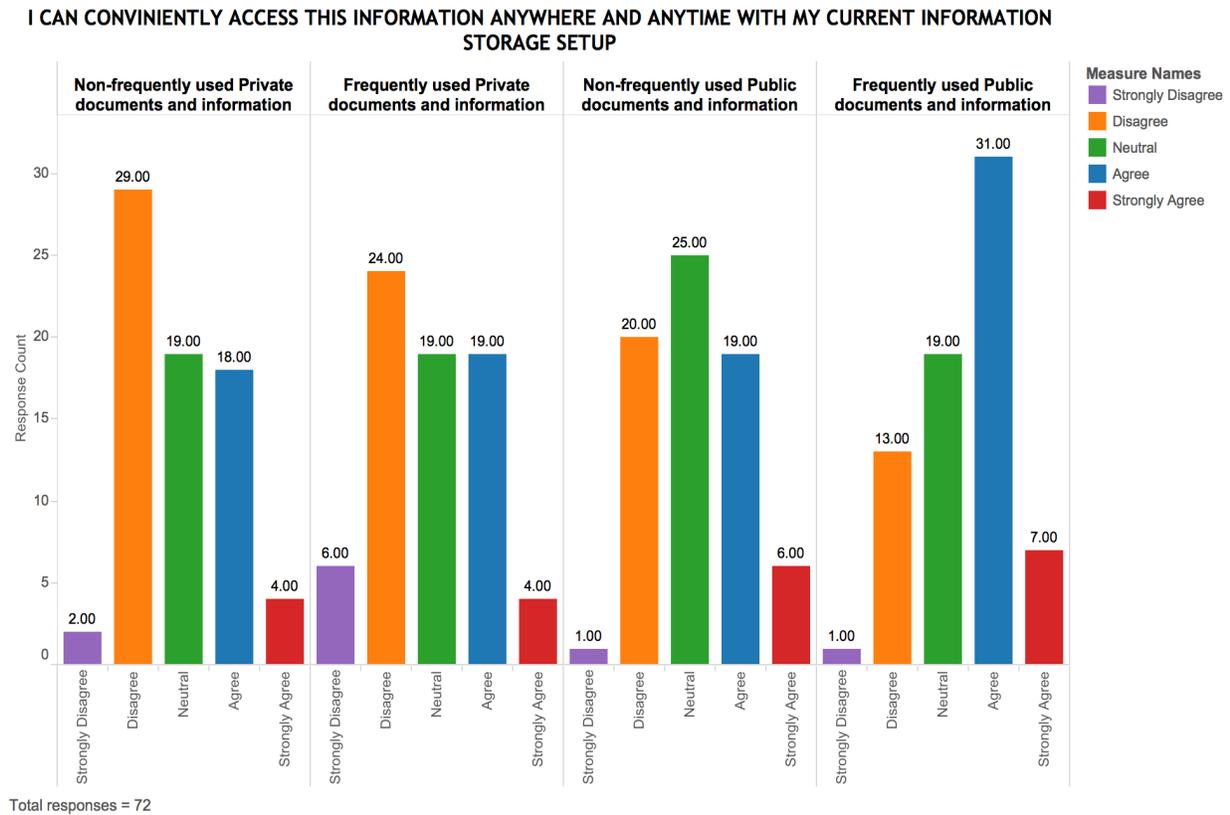
The process of information storage and retrieval generally involves organizing, storing, maintaining and retrieving information[43]. We look at users' experiences on accessing and retrieving of frequently used public and private information as well as non-frequently used public and private information. Technologies generally aid in finding information based on the name of the file or format. Along with the capabilities of information systems, it is also important to see how users interact with their data. Cole's study compared the interaction style of users with ephemeral , action and working data to improve user interaction capabilities of systems[28]

RELATED SAMPLES

6.1 ACCESSIBILITY

The following question was a measure of how conveniently an individual can access different types of information. With the advancements of technology in the last few decade, we now have the opportunity to access information from anywhere at the time of need, provided the information is stored online or on portable devices that have facilitated access. However, there are certain disadvantages to dispersed data if mismanaged. For Example, not remembering the location of storage or scrolling through folders timelessly, badly tagged folders[44] etc. which cause inconvenience in accessing them. The participants were asked to choose from a 5-point Likert scale ordered from strongly disagree to strongly agree. They were asked to rate the ease of accessibility of 4 types of information namely non-frequently used private information, frequently used private information, non-frequently used public information and frequently used public information.

Figure 6: Bar chart showing participants responses on accessibility of different types of information.



Description: The grouped bar chart above illustrates the frequency of responses on the convenience of accessing different types of public and private information. From a total of 72 participants, responses of each participant across 4 categories of information type have been graphically represented above.

Comparisons across different types of data

We now compare the accessibility of different types of information to see if people treat them differently by hypothesis testing.

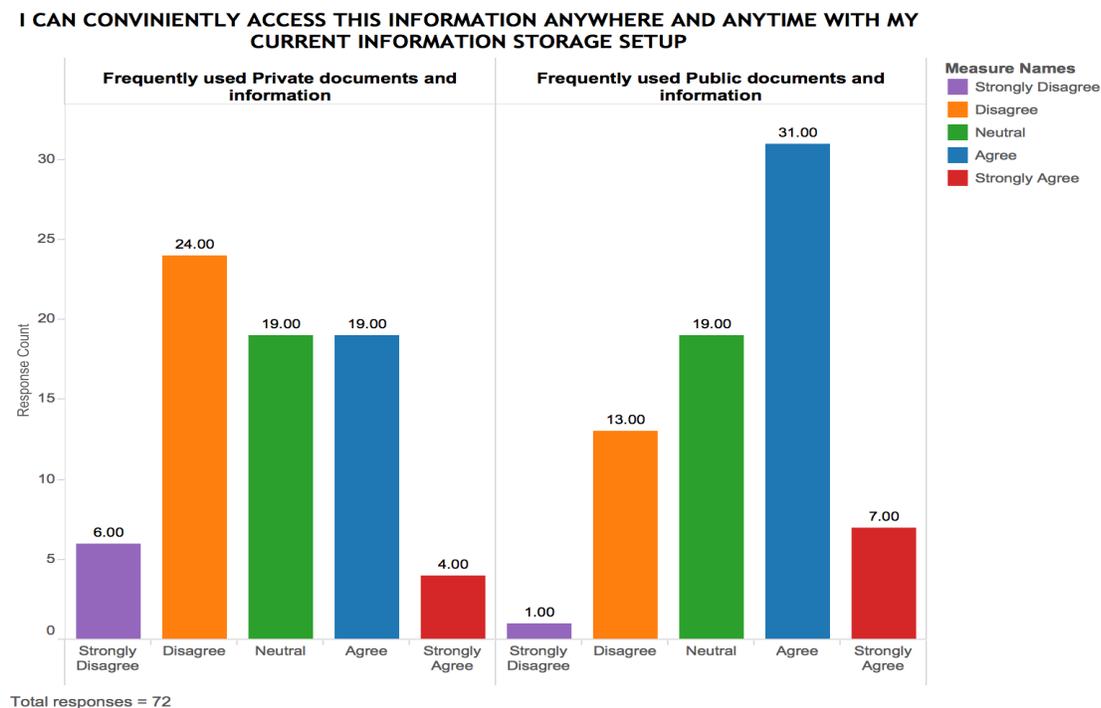
6.1.1 Frequently used private information and frequently used public information.

- i. Decision: **Reject H_0**
- ii. Wilcoxon signed-ranked test indicates that the median of differences between frequently

used private information and frequently used public information is not equal to zero. This suggests that individuals access frequently used private information differently from frequently used public information ($|Z| = 3.510$, $p < .000$, $r = 0.29$). The descriptive statistics show a median of 3 for frequently used private information and a median of 4 for frequently used public information. From a total of $N = 72$, positive differences are $N = 27$ and negative differences are $N = 10$ with a total of 35 ties. Therefore, we have significant evidence to reject the null hypothesis thus accepting the alternative hypothesis.

iii. Reporting:

Figure 7: Chart showing accessibility of Non-frequently used private information with frequently used private information

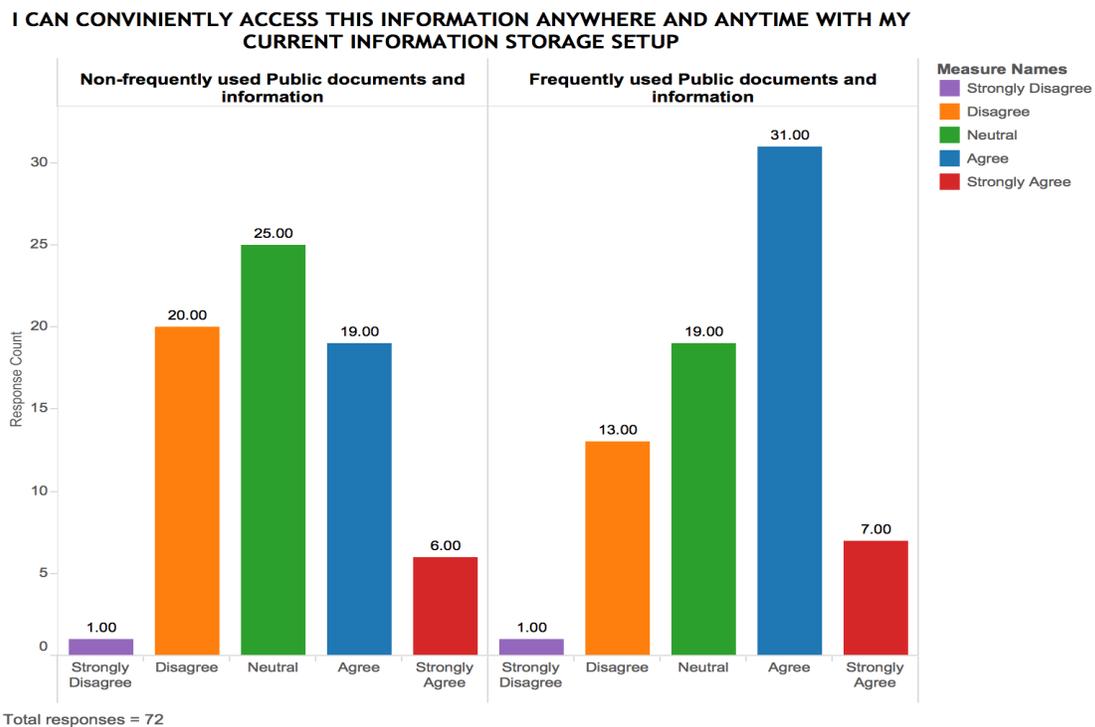


Description: The Chart above shows the frequency distribution of the responses for comparisons between frequently used private information and frequently used public information. The frequencies show that, while 24 participants disagree and 6 participants strongly disagree that they can conveniently access frequently used private information, only 13 participants disagree and 1 participant strongly disagrees that one can conveniently access frequently used public information and both sides have 19 neutral responses. This suggests that most people feel that they can access frequently used public information easily as compared to frequently used private information. Furthermore, while 19 participants agree and 4 participants strongly agree for the former, 31 participants agree and 7 strongly agree for the latter further confirms the judgment stated above.

6.1.2 Non-frequently used public information and frequently used public information.

- i. Decision: **Reject H_0**
- ii. Wilcoxon signed-ranked test indicates that the median of differences between non-frequently used public information and frequently used public information is not equal to zero. This suggests that individuals access non-frequently used public information differently from frequently used public information ($|Z| = 2.777$, $p < .005$, $r = 0.23$). The descriptive statistics show a median of 3 for non-frequently used public information and 4 for frequently used public information. From a total of $N = 72$, the positive differences are $N = 20$ and negative differences are $N = 9$ with a total of 43 ties. Therefore, we have significant evidence to reject the null hypothesis thus accepting the alternative hypothesis.
- iii. Reporting:

Figure 8: Accessibility of non-frequently used public information and frequently used public information



Description: The Chart above shows the frequency distribution of the responses for comparisons between non-frequently used public information and frequently used public information. The frequencies illustrate that, while 20 participants disagree and 1 participant strongly disagrees

that one can conveniently access non-frequently used public information, only 13 participants disagree and 1 participant strongly disagrees that one can conveniently access frequently used public information conveniently. This suggests that most people feel that they can access frequently used public information easily as compared to non-frequently used public information. Furthermore, while 19 participants agree and 6 participants strongly agree for the former, 31 participants agree and 7 strongly agree for the latter which further confirms the judgment stated above. The neutral responses for non-frequently used public information and frequently used public information are 25 and 19 respectively.

Table 2: Cross table containing 6 groups of information comparisons.

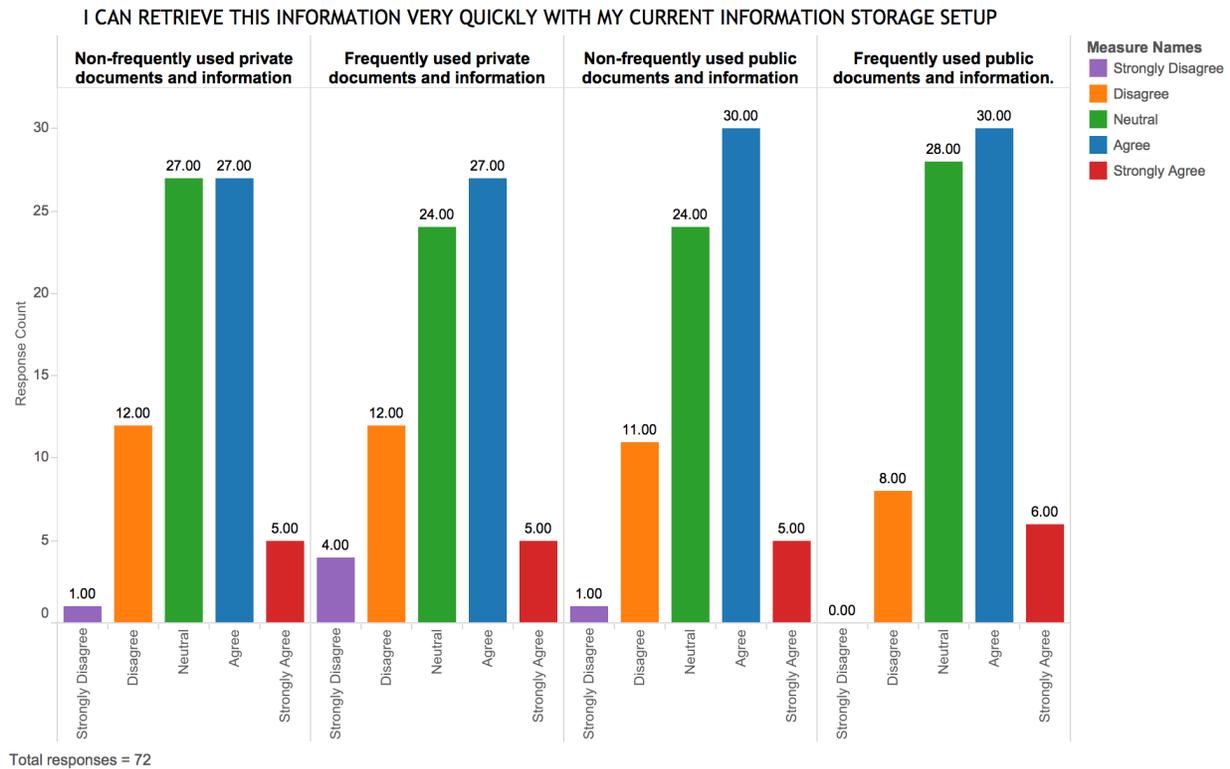
ACCESSIBILITY	Non frequently used private information	Frequently used private information.	Frequently used public information
Non frequently used public information	Retain H_0 P = .102	Retain H_0 P = .058	Reject H_0 P = .005
Frequently used public information	Reject H_0 P = .001	Reject H_0 P = .000	
Frequently used private information	Retain H_0 P = .900		

Description: The table above shows the results of hypothesis tests on different types of information tested against each other.

6.2 RETRIEVAL

The following question is an analysis of the participants' experiences with regards to the quick retrieval of information from their current information storage setup. As our data is stored across various devices and platforms, retrieving them in the time of need can be a compelling task. Retrieving information across several devices requires complete knowledge of the location of the device, name of the service or application used, location of the data file, name of the data file. For the cognitive system to remember all of or at-least one of these details in times of distress or emergency is a compelling task. Therefore, users generally categorize information across various services and devices based on type of information, formats, context etc. If one can retrieve a type of information faster than the other, this proves that they either store them by using different methods or that they are generally used differently.

Figure 9: Bar chart showing an overview of the participants views on the quick retrieval of different types of information with their current storage setup.



Description: The grouped bar chart above illustrates the frequency of responses on the quickness of retrieving different types of public and private information. From a total of 72 participants, responses of each participant across 4 categories of information type have been graphically represented above.

Comparisons across different types of data

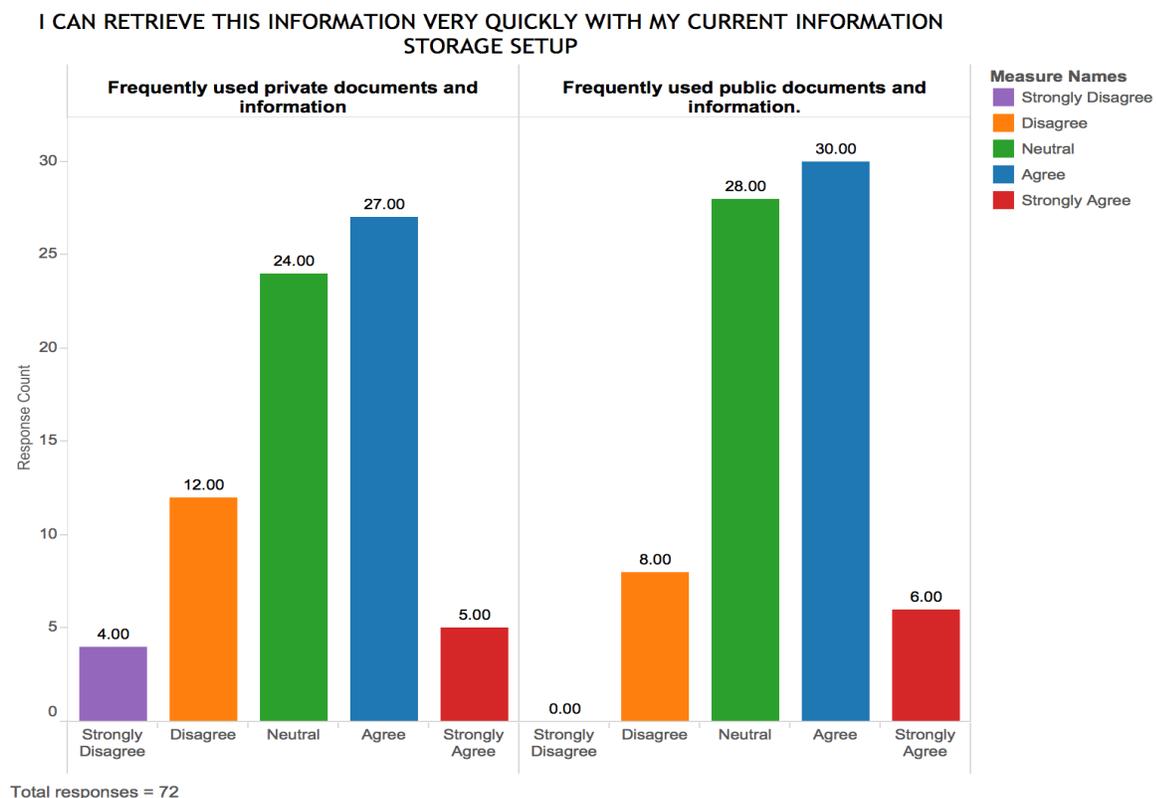
6.2.1 Frequently used private information and frequently used public information.

- i. Decision: **Reject H_0**
- ii. Wilcoxon signed-ranked test indicates that the median of differences between non-frequently used private information and frequently used public information is not equal to zero. This suggests that individuals can retrieve frequently used private information differently from frequently used public information ($|Z| = 2.002$, $p < .045$, $r = 0.16$). The

descriptive statistics show a median of 3 for frequently used private information and 3.5 for frequently used public information. From a total of N = 72, the positive differences are N = 13 and negative differences are N = 17 with a total of 42 ties. Therefore, we have significant evidence to reject the null hypothesis thus accepting the alternative hypothesis. However, as seen from the p-value of .045, although we have significant, the effect size is very small and the p-value is very close to the alpha value.

iii. Reporting:

Figure 10: Chart showing retrieval of frequently used private information and frequently used public documents



Description: The Chart above shows the frequency distribution of the responses for comparisons between frequently used private information and frequently used public information with regards to the quickness of retrieval. The frequencies illustrate that, while 12 participants disagree and 4 participants strongly disagree that they can quickly retrieve frequently used private information, only 8 participants disagree that one can quickly retrieve frequently used public information. However, while 27 participants agree and 5 participants strongly agree for the former, 30 participants agree and 6 strongly agree for the later which suggests that most people feel that they can retrieve frequently used public information much faster than frequently used private information. The neutral responses for frequently used private information and frequently used public information are 24 and 28 respectively.

Table 3 : Cross table containing 6 groups of information comparisons

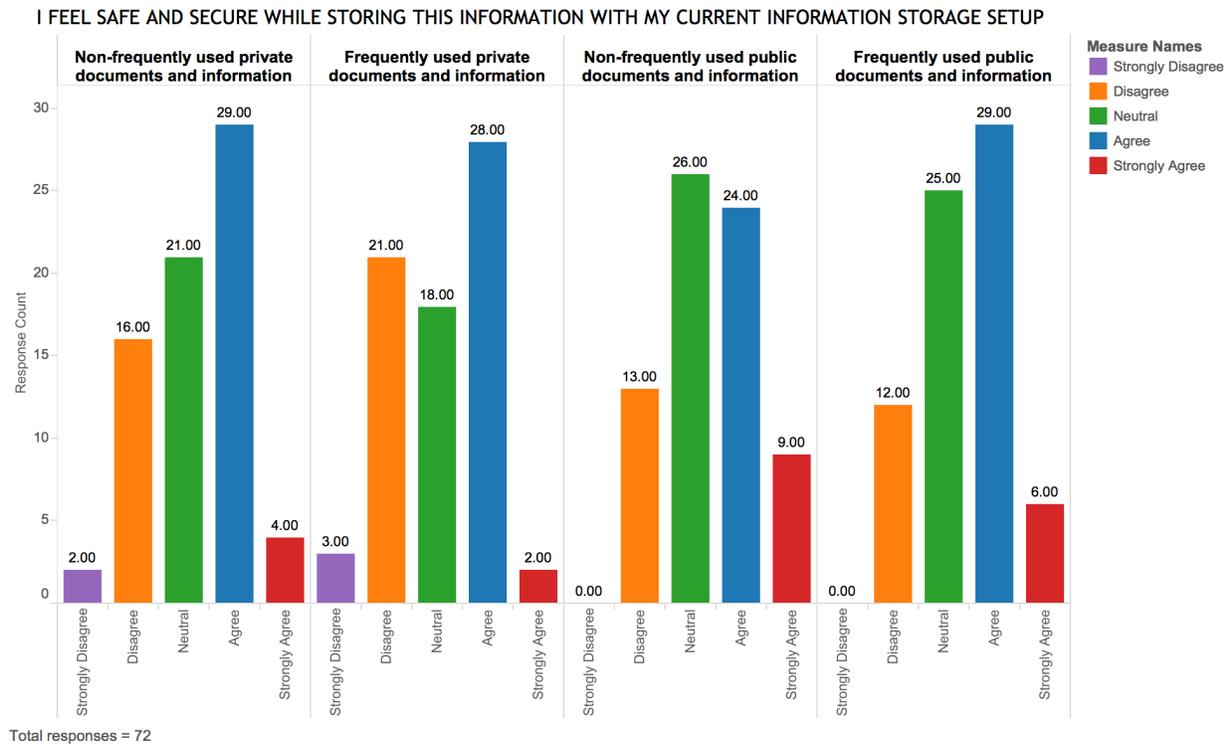
RETRIEVAL	Non used information	Frequently private information	Frequently used private information.	Frequently used public information
Non used information	Retain H_0 P = .524		Retain H_0 P = .250	Retain H_0 P = .205
Frequently used public information	Retain H_0 P = .119		Reject H_0 P = .045	
Frequently used private information	Retain H_0 .470			

Description: The table above shows the results of hypothesis tests on different types of information tested against each other.

6.3 SAFETY AND SECURITY

The following question an analysis of the participants' experiences with regards to the safety and security of information with their current information storage setup. Here, safety of information refers to loss of data and security of information refers to leak of information to a third party, both could lead to potential embarrassment, humiliation or identity theft. An individual generally makes decisions regarding storage, based on safety and security measures of certain applications or devices.

Figure 11: Grouped bar chart showing an overview of the participants views on safety and security of different types of information with their current storage setup.



Description: The grouped bar chart above illustrates the frequency of responses on each participant’s feelings on safety and security while storing different types of public and private information. From a total of 72 participants, responses of each participant across 4 categories of information type have been graphically represented above.

Comparisons across different types of data

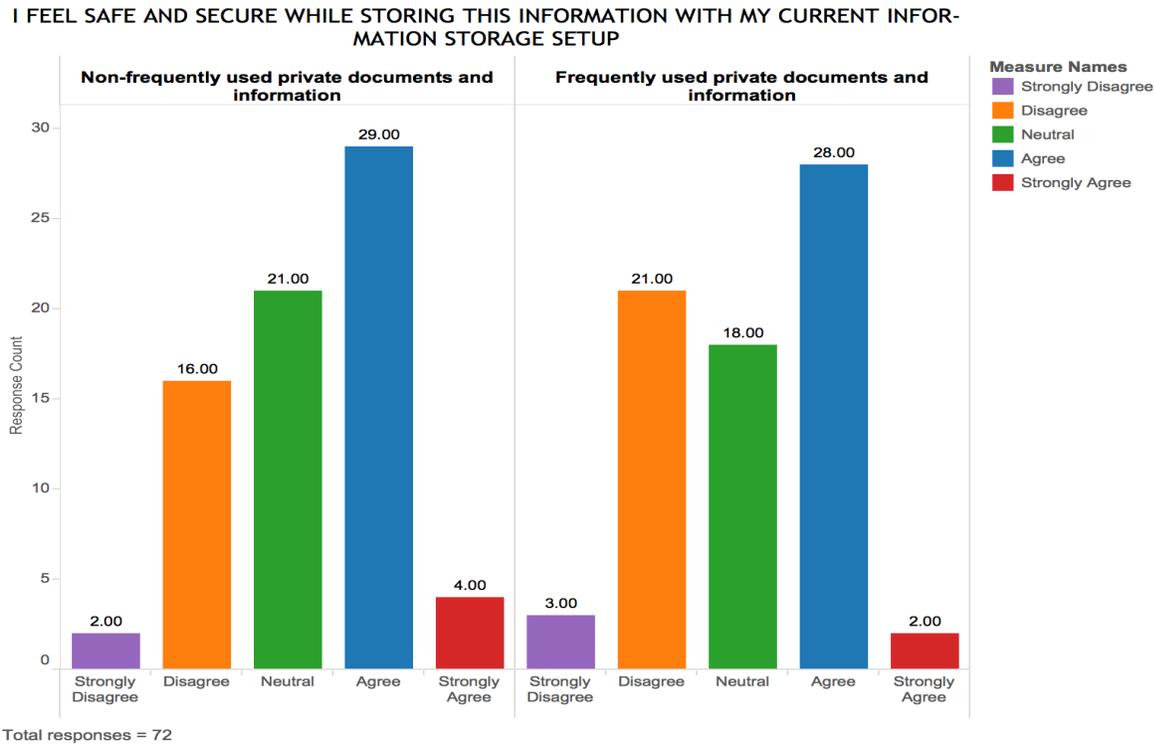
6.3.1 Non-frequently used private information and frequently used private information.

- i. Decision: **Reject H_0**
- ii. Wilcoxon signed-ranked test indicates that the median of differences between non-frequently used private information and frequently used private information is not equal to zero. This suggests that individuals do not feel the same level of safety and security with regards to non frequently used private information and frequently used private information ($|Z| = 2.180$, $p < .029$, $r = 0.18$). The descriptive statistics show that the median of frequently used private information and frequently used public information are the same (median = 3). From a total of $N = 72$, the positive differences are $N = 3$ and

negative differences are $N = 13$ with a total of 56 ties. Therefore, we have significant evidence to reject the null hypothesis thus accepting the alternative hypothesis.

iii. Reporting:

Figure 12: safety and security of non-frequently used private information and frequently used private information.

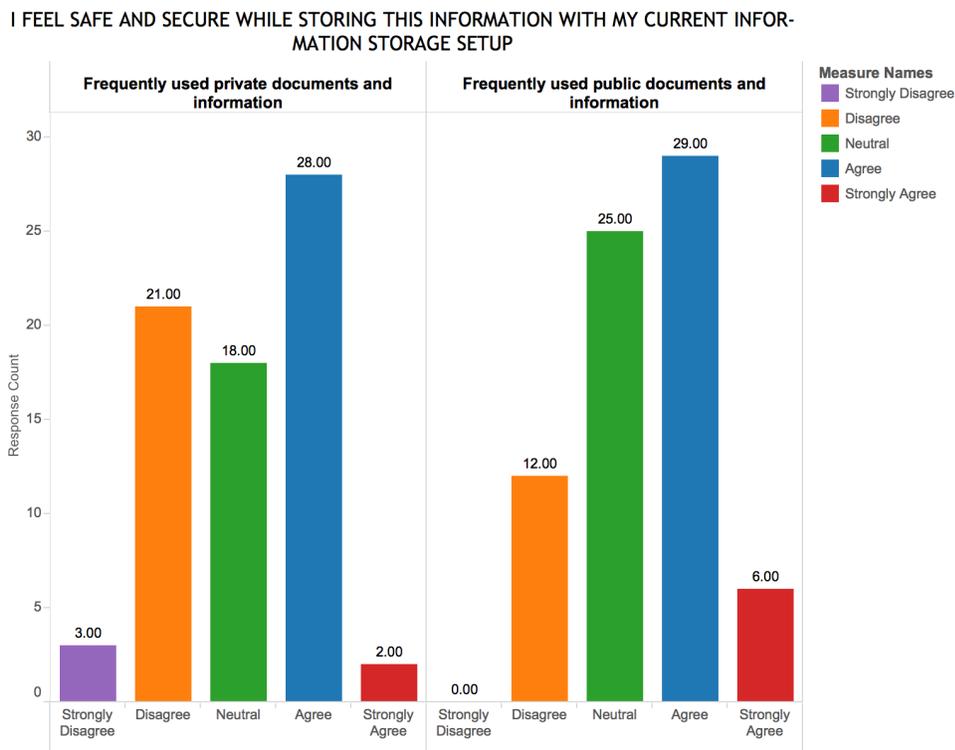


Description: The Chart above shows the frequency distribution of the responses with regards to safety and security while storing non-frequently used private information and frequently used private information. The frequencies illustrate that, while 16 participants disagree and 2 participants strongly disagree that they that they feel safe about non frequently used private information, 21 participants disagree and 3 participants strongly disagree for frequently used private information. This suggests that most people feel that their non frequently used private information is much safer than their frequently used private information. Furthermore, while 29 participants agree and 4 participants strongly agree for the former, 28 participants agree and 2 strongly agree for the later which further confirms the judgment stated above. The neutral responses for non frequently used private information and frequently used private information are 21 and 18 respectively.

6.3.2 Frequently used private information and frequently used public information.

- i. Decision: **Reject H_0**
- ii. Wilcoxon signed-ranked test indicates that the median of differences between frequently used private information and frequently used public information is not equal to zero. This suggests that individuals do not feel the same level of safety and security with regards to frequently used private information and frequently used public information ($|Z| = 3.190$, $p < .001$, $r = 0.26$). The descriptive statistics show a median of 3 for frequently used private information and 3.5 for frequently used public information. From a total of $N = 72$, the positive differences for $N = 19$ and negative differences for $N = 5$ with a total of 48 ties. Therefore, we have significant evidence to reject the null hypothesis thus accepting the alternative hypothesis.
- iii. Reporting:

Figure 13: safety and security of frequently used private information and frequently used public information



Total responses = 72

Description: The Chart above shows the frequency distribution of the responses with regards to safety and security while storing frequently used private information and frequently used public information. The frequencies illustrate that, while 21 participants disagree and 3 participants strongly disagree that they feel safe storing frequently used private information, 12 participants disagree and no participants strongly disagree for frequently used public information. This suggests that most people feel that their frequently used public information is much safer than

their frequently used private information. Furthermore, while 28 participants agree and 2 participants strongly agree for the former, 29 participants agree and 9 strongly agree for the latter which further confirms the judgment stated above. The neutral responses for frequently used private information and frequently used public information are 18 and 25 respectively

Table 4 : Cross table containing 6 groups of information comparisons

SAFETY AND SECURITY	Non frequently used private information	Frequently used private information.	Frequently used public information
Non frequently used public information	Retain H_0 P = .094	Reject H_0 P = .002	Retain H_0 1.000
Frequently used public information	Reject H_0 P = .046	Reject H_0 P = .001	
Frequently used private information	Reject H_0 P = .029		

Description: The table above shows the results of hypothesis tests on different types of information tested against each other.

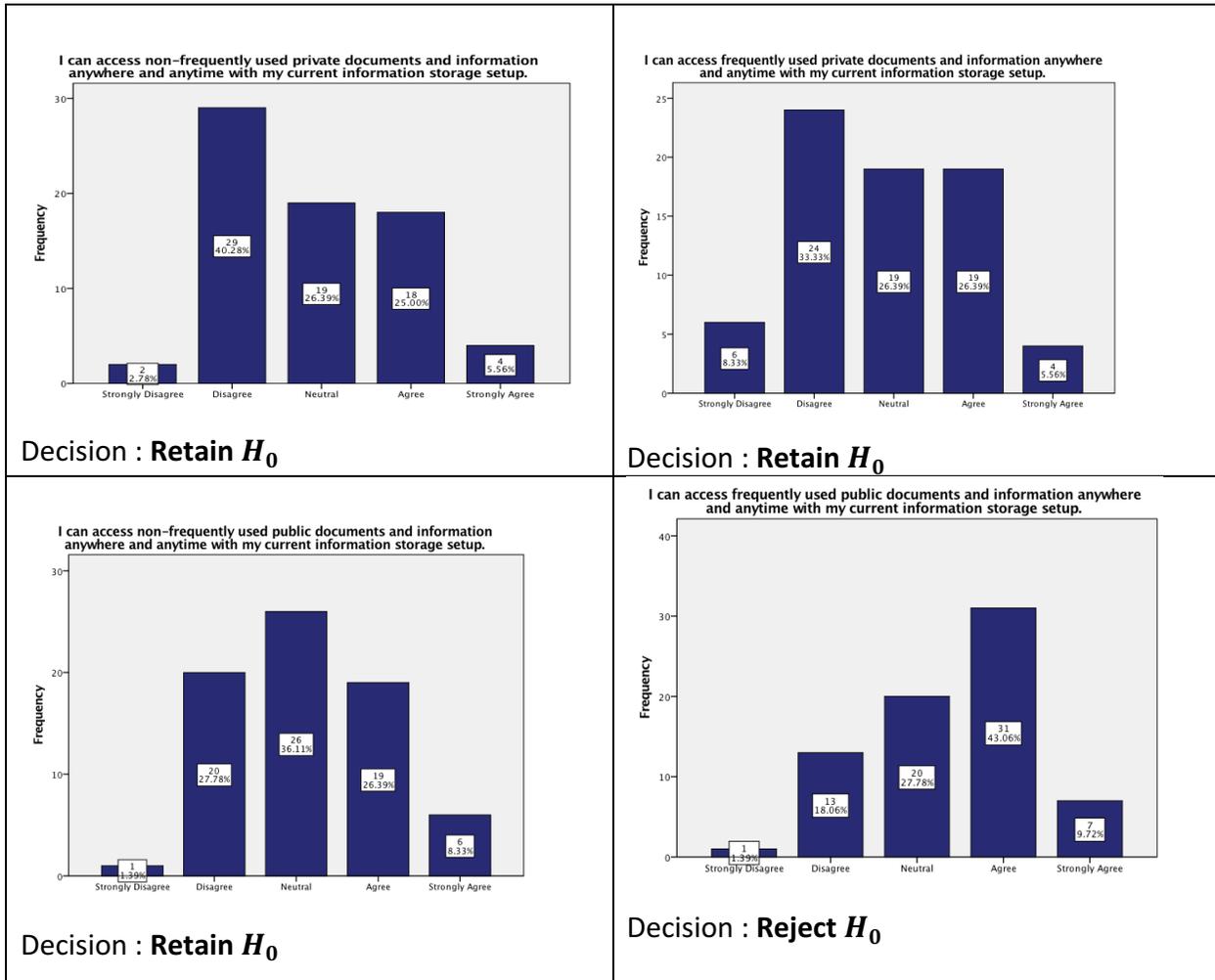
ONE SAMPLE TESTS

The following tests answer to a two tailed hypothesis to check if the median is 3 (neutral responses). If the null hypothesis is rejected, then the alternate hypothesis will be accepted and which will mean that the the median falls on either end of the Likert scale.

6.4 ACCESSIBILITY

- d) The Wilcoxon signed-ranked test indicates that the median of **frequently used public data** does not equal to 3 (neutral.) This proves that the results are directed towards either right side of the Likert scale. The observed median is 4 which as opposed to the hypothesized median of 3 ($|Z| = 3.433$, $p < .001$, $r = 0.28$). Therefore, we have significant evidence to reject the null hypothesis thus accepting the alternative hypothesis.

Table 5: Table showing the frequencies and percentage frequencies of information for accessibility.

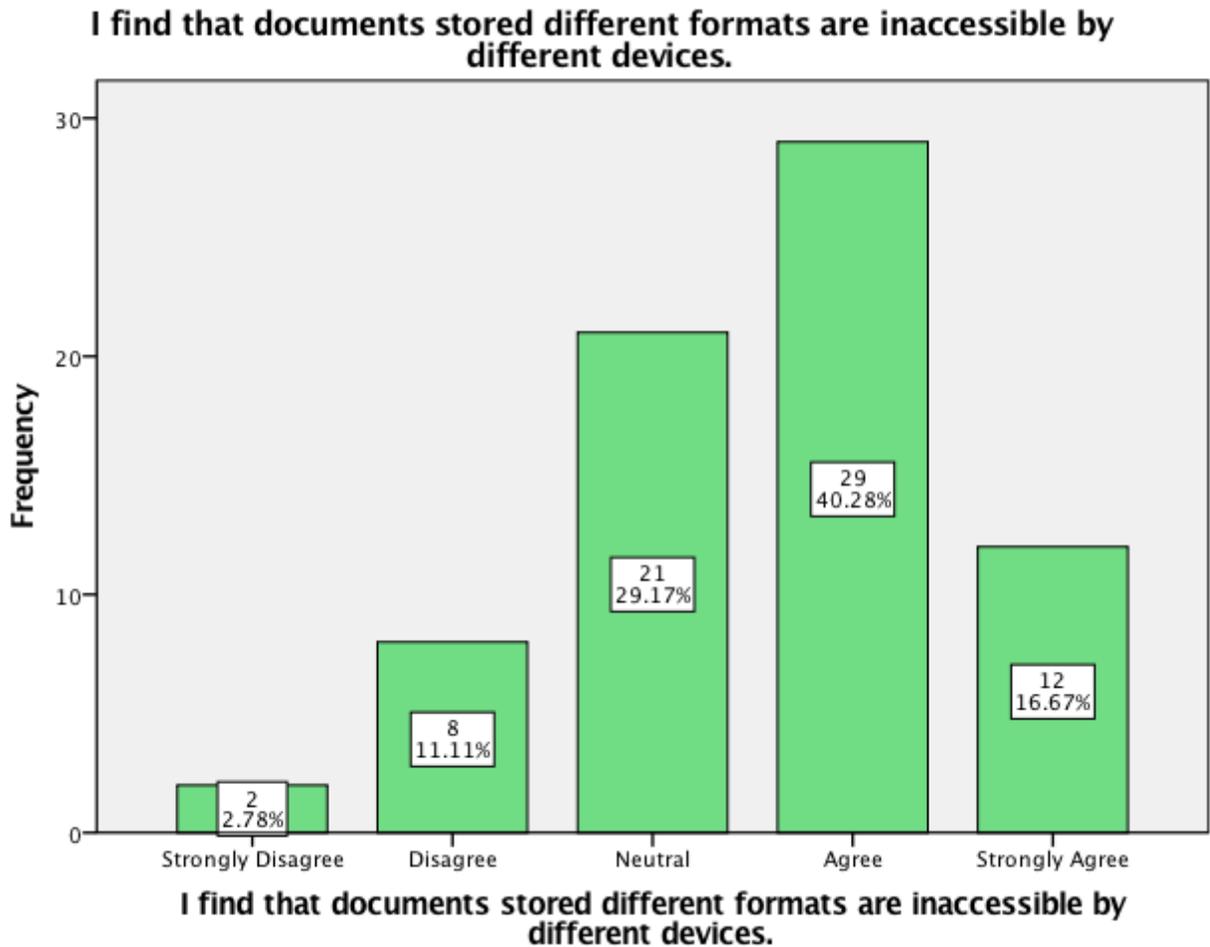


Justification: The chart above illustrates the frequencies and percent frequencies of responses for quick retrieval of different types of information. We only have significant evidence to reject the null hypothesis for accessibility of frequently used public documents.

6.4.1 I find that documents stored in different formats are inaccessible by different devices

- i. **Decision: Reject H_0**
- ii. The Wilcoxon signed-ranked test indicates that the median does not equal to 3 (neutral). This proves that the results are directed towards either ends of the Likert scale. The observed median is 4 (Agree) and hypothesized median is also 3 (Neutral) ($Z = 4.164$, $p < .000$, $r = 0.347$). There we have significant evidence to reject the null hypothesis thus accepting the alternative hypothesis.

Figure 14: Horizontal Bar chart showing participants views on file format causing information to be inaccessible.



Description: The chart above illustrates that 40.28 % of participants agree and 16.67 % strongly agree that that documents stored in different formats are inaccessible by different devices while only 11.11 % of the participants disagree and 2.78 % strongly disagree. There were 29.17 % of neutral responses.

6.5 RETRIEVAL

- a) The Wilcoxon signed-ranked test indicates that the median of **non-frequently used private data**, median of **non-frequently used public data** and median of **frequently used public data** do not equal to 3 (neutral). This proves that the results are directed towards either ends of the Likert scale.

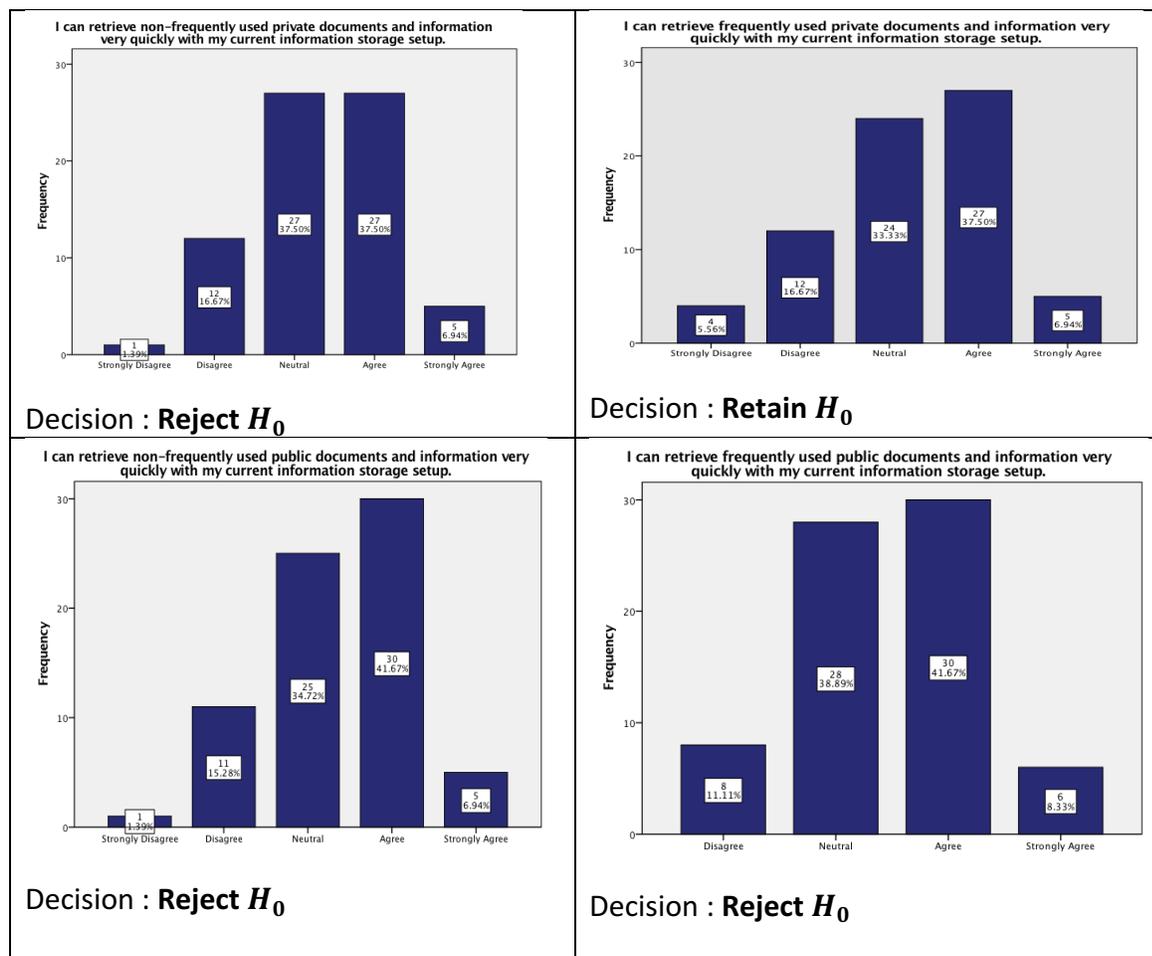
For non-frequently used private data, the observed median is 3 and hypothesized median is also 3 ($|Z| = 2.891$, $p < .004$, $r = 0.24$).

For non-frequently used public data, the observed median is 3 and hypothesized median is also 3 ($|Z| = 2.333$, $p < .001$, $r = 0.19$).

For frequently used public data, the observed median is 3.50 and hypothesized median is also 3 ($|Z| = 4.307$, $p < .000$, $r = 0.35$).

Therefore, we have significant evidence to reject the null hypothesis thus accepting the alternative hypothesis.

Table 6: table showing frequencies and percentage frequencies of participants responses on retrieval



Justification: The chart above illustrates the frequency and percent frequencies of responses for

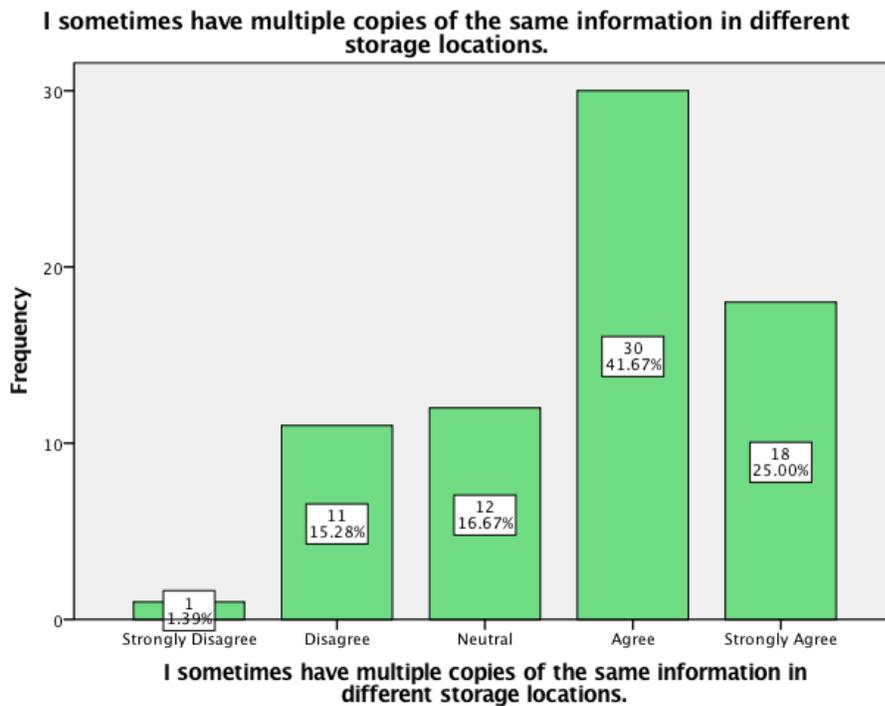
quick retrieval of different types of information. We only have significant evidence to reject the null hypothesis for retrieval of non-frequently used private documents, frequently used public documents and non-frequently used public documents.

6.5.1 I sometimes have multiple copies of the same information in different storage locations

i. Decision: Reject H_0

- ii. The Wilcoxon signed-ranked test indicates that the median of the study result is not 3 which stands for neutral responses. This proves that the results are directed towards either ends of the Likert scale. The observed median is 4 (Agree) and hypothesized median is 3 (Neutral) ($Z = 4.873$, $p < .000$, $r = .40$). Therefore, we have significant evidence to reject the null hypothesis thus accepting the alternative hypothesis.

Figure 15: chart showing the frequencies and percentage frequencies of responses about multiple copies of data



Description: The chart above illustrates that 40.67 % of participants agree and 25.00 % strongly agree that they have multiple copies of the same information in different storage locations while only 15.28 % of the participants disagree and 1.39 % strongly disagree. There were 16.67 % of neutral responses.

6.6 SAFETY AND SECURITY:

- a) The Wilcoxon signed-ranked test indicates that the median of **non-frequently used private data**, median of **non-frequently used public data** and median of **frequently used public data** do not equal to 3 (neutral). This proves that the results are directed towards either ends of the Likert scale.

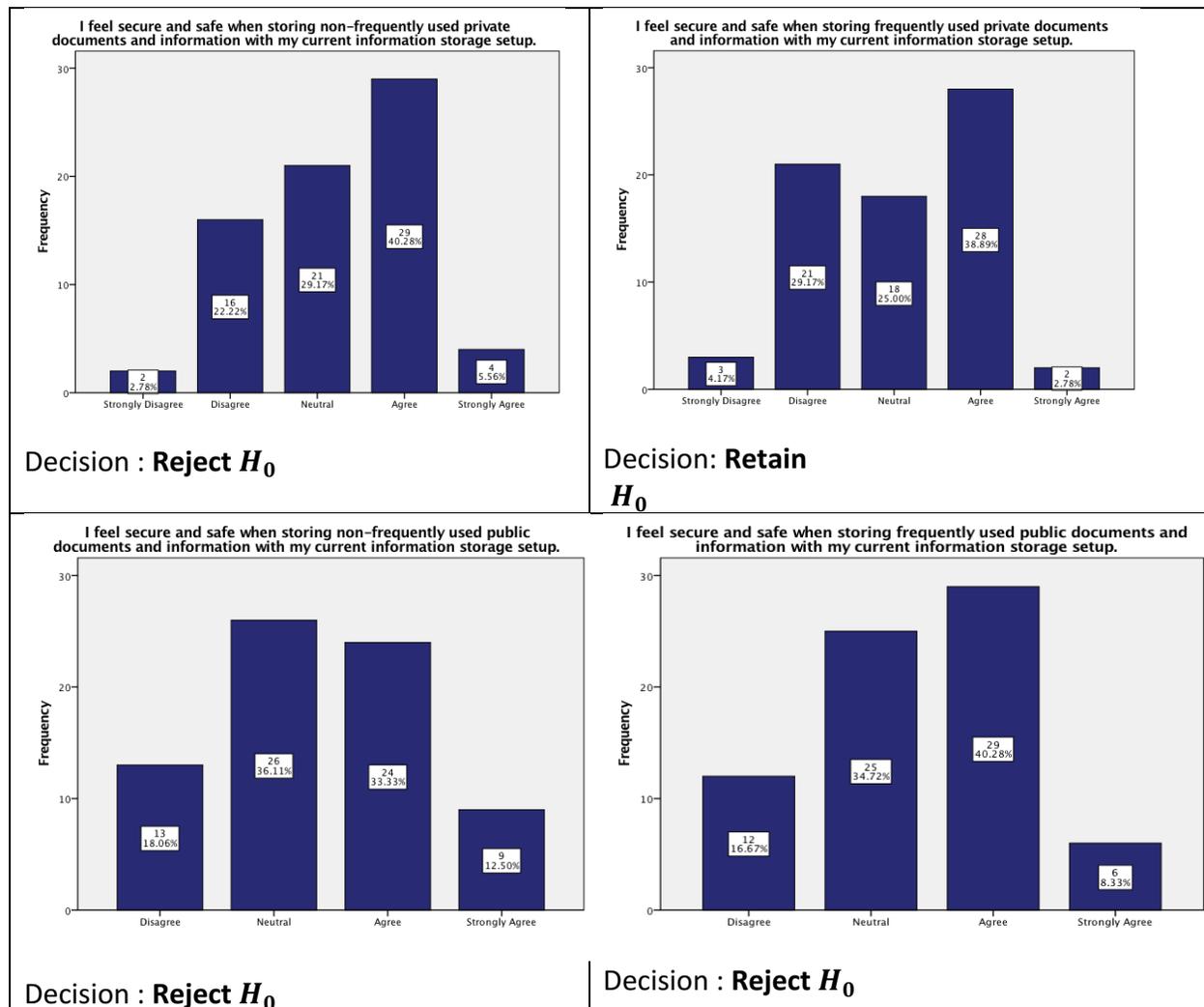
For non-frequently used private data, the observed median is 3 and hypothesized median is also 3 ($Z = 2.034$, $p < .042$, $r = 0.16$)

For non-frequently used public data, the observed median is 3.50 and hypothesized median is also 3 ($Z = 3.433$, $p < .001$, $r = 0.28$).

For frequently used public data, the observed median is 3.50 and hypothesized median is also 3 ($Z = 3.605$, $p < .000$, $r = 0.30$).

Therefore, we have significant evidence to reject the null hypothesis thus accepting the alternative hypothesis.

Table 7: table showing frequencies and percentage frequencies of participants responses on safety and security

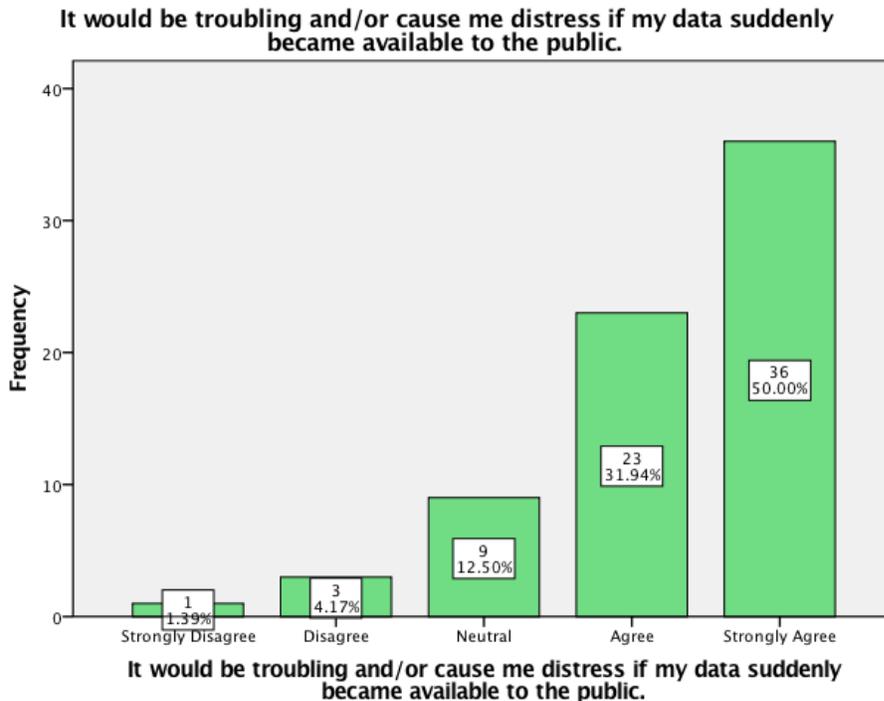


Justification: The chart above illustrates the frequency and percent frequencies of responses for safety and security of different types of information. We have significant evidence to reject the null hypothesis for safety and security of all documents.

6.6.1 It would be troubling and/or cause me distress if my data suddenly becomes available

- i. **Decision: Reject H_0**
- ii. The Wilcoxon signed-ranked test indicates that the median does not equal to 3 (neutral) This proves that the results are directed towards either ends of the Likert scale. The observed median is 4.50 (Agree – Strongly agree) and hypothesized median is 3 ($Z = 6.537$, $p < .000$, $r = 0.54$). Therefore, we have significant evidence to reject the null hypothesis thus accepting the alternative hypothesis.
- iii. Reporting:

Figure 16: Bar chart showing participants views on distress caused by loss or leak of data.



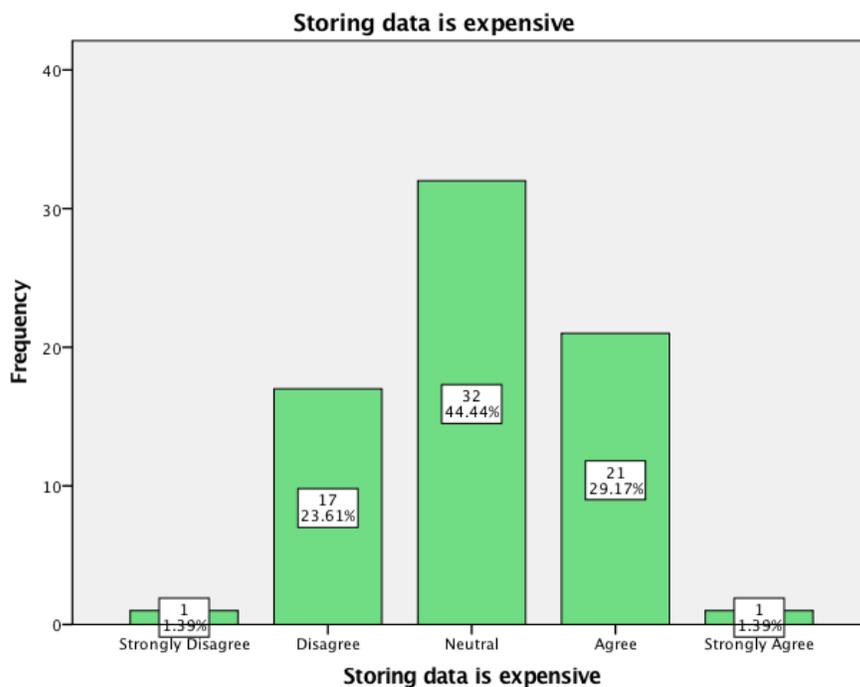
Description: The chart above illustrates that 31.94 % of participants agree and 50 % strongly agree that It would be troubling and cause distress if data suddenly becomes available while only 4.17 % of the participants disagree and 1.39 % strongly disagree. There were 12.50 % of neutral responses.

Additional factors:

6.6.2 Storing data is expensive

- i. **Decision: Retain H_0**
- ii. The Wilcoxon signed-ranked test indicates that the median of the study result is 3 which stands for neutral. This proves that the results are not directed towards either ends of the Likert scale. The observed median is 3 and hypothesized median is also 3 ($Z = .588$, $p < .556$, $r = 0.04$). Therefore, we have insignificant evidence to reject the null hypothesis.

Figure 17: Bar chart showing the participants views on expenses of storing data.

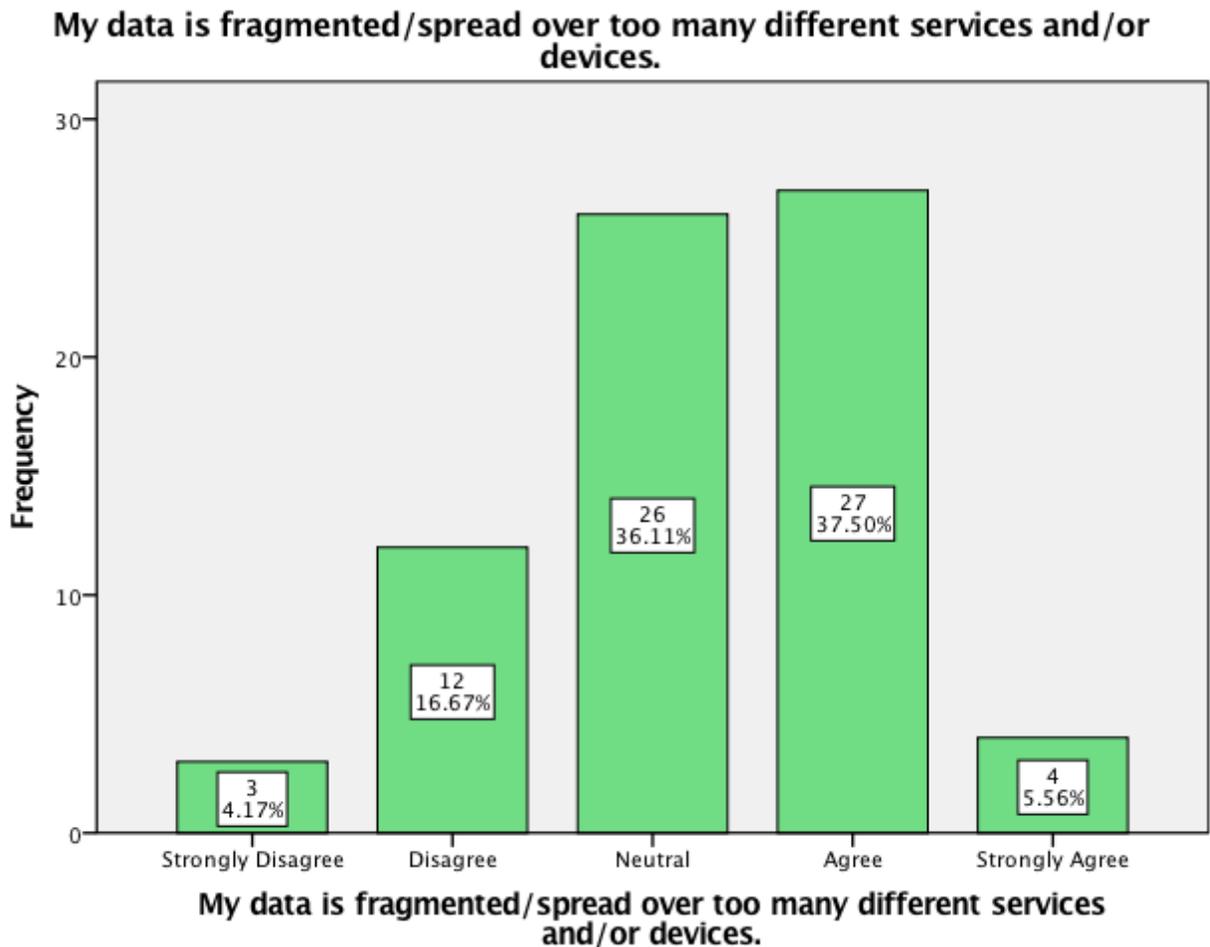


Description: The chart above illustrates that 29.17 % of participants agree and 1 % strongly agree that Storing data is expensive while 23.61 % of the participants disagree and 1.39 % strongly disagree. There were 44.44 % of neutral responses. This means that users have maximum neutral responses.

6.6.3 My data is fragmented/spread over too many different services and/or devices

- i. **Decision: Reject H_0**
- ii. The Wilcoxon signed-ranked test indicates that the median of the study result is 3 which stands for neutral. This proves that the results are directed towards either ends of the Likert scale. The observed median is 4.50 (Agree – Strongly agree) and hypothesized median is 3 ($Z = 2.030$, $p < .042$, $r = .16$). Therefore, we have significant evidence to reject the null hypothesis thus accepting the alternative hypothesis.

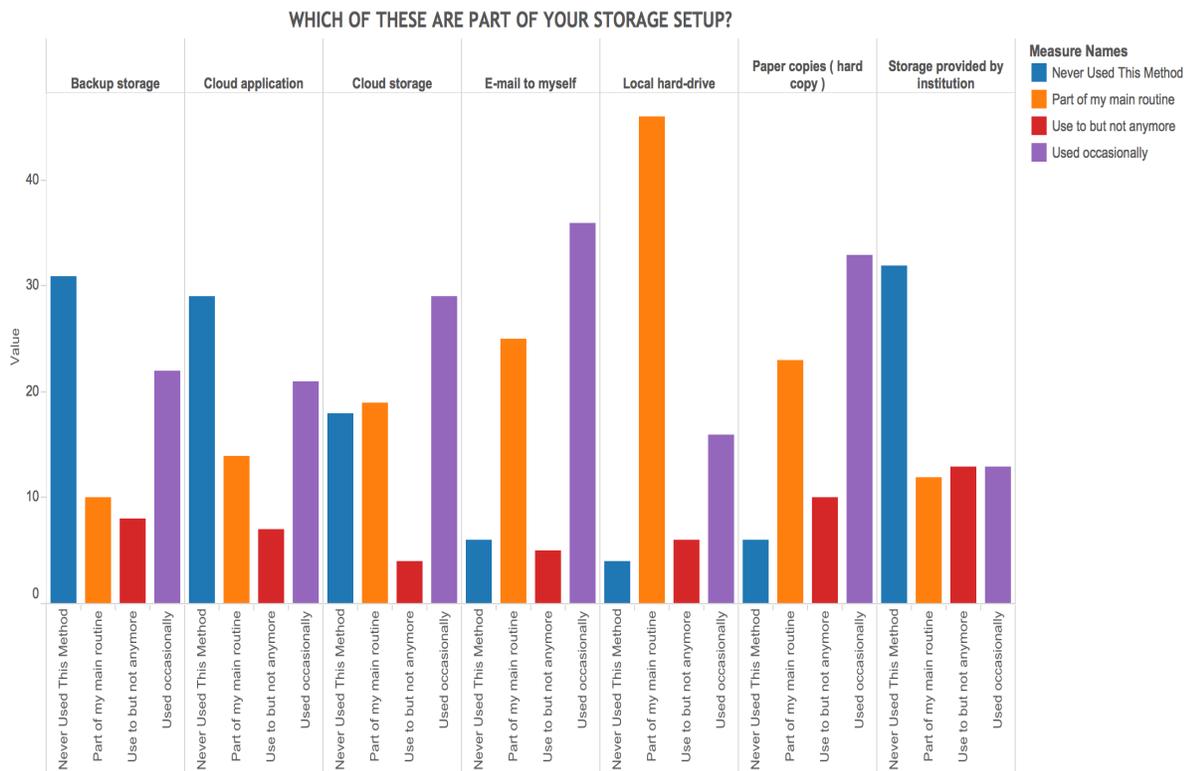
Figure 18: bar chart for participants views on data fragmentation



Description: The chart above illustrates that 37.50 % of participants agree and 5.56 % strongly agree that storing data is expensive while 16.67 % of the participants disagree and 4.17 % strongly disagree. There were 36.11 % of neutral responses.

6.6.4. Usage of multiple methods

Figure 19: stacked bar showing the usage of different storage devices and services



Description: The chart above illustrates the multiple methods of storage used by participants and the level of usage. Each participant chooses one category for each type of storage method.

The most commonly used storage service that is part of all user's routine is storing information on a local drive followed by email, paper copies, cloud storage, cloud applications, storage provided by the institution and lastly back up storage.

With regards to back up storage, cloud applications and storage provided by the institution most participants said that they never used this method.

With regards to cloud storage, emails and paper documents, most participants said that they occasionally used them.

With regards to local hard-drives, most participants said they they use this method as part of their routine.

7. ANALYSIS: INTERVIEWS

OVERVIEW

This will be a descriptive analysis of data collected from the task-based semi-structured interviews. The participants were asked to retrieve 20 different types of information documents. The number of items retrieved in 15 minutes were recorded. The type of information that they were asked to retrieve ranged from frequently used and non-frequently used private and public documents and information. During the task, participants were required them to use any device or application that were available to them to find different types of information. As the research is relating to everyday behavioral sciences of the participants, the location of the study was decided by the participant based on their convenience.

7.1 How many information items did they find?

Table 8: Number of items retrieved by the participants in 15 minutes

ID	CURRENT OCCUPATION	FIELD OF STUDY/WORK	AGE	DEVICES USED	NUMBER OF DOCUMENTS RETRIEVED
1	Senior Lecturer	Computer science	50 to 59	Work computer Phone	13
2	Secretary	Data Administration	50 to 59	work computer	10
3	Senior Academic Liaison	Library & Information Science	50 to 59	Work computer	11
4	Secretary	Data Administration	40 to 49	Work computer	12
5	Senior Manager	Librarianship	40 to 49	Work computer Phone	11
6	Repository Support Officer	Library & Information Science	40 to 49	work computer	13
7	Lecturer	Software Engineering	40 to 49	Work computer	15
8	Lecturer	Software Engineering	30 to 39	Work computer	12

9	Msc student	School of English	18 to 29	Personal computer	15
10	Msc student	Psychology	18 to 29	Work computer Mobile Phone	9
11	MSc student	Computing and IT	18 to 29	Lab computer Mobile phone	5
12	MSc student	Management and IT	18 to 29	Personal computer phone	17
13	MSc Student	Computer Science	18 to 29	Personal computer phone	15
14	MSc student	IT	18 to 29	Personal computer Mobile phone	12
15	MSc student	Software Engineering	18 to 29	Personal computer	13
16	MSc student	Advanced Computer Science.	18 to 29	Personal computer Phone	14
17	MSc student	Management & IT	18 to 29	Personal computer Phone	14
18	MSc student	Marketing	18 to 29	Personal computer phone	18
19	MSc Student	Law, Computing and IT	18 to 29	Personal computer Phone	17

7.2 Where do people store different types of data?

7.2.1 Public information

Participants were easily able to retrieve frequently and non-frequently used public information and documents like rules and regulations related to work/academia, online movies and bookmarks with ease because this information is available online and participants were connected to the Wi-Fi. A participant was not able to retrieve bookmarks as the student was using the computer provided at the university laboratory and the bookmarks were stored in a chrome browser of the personal computer at home for which the password was forgotten.

Table 9: Table showing most commonly used storage methods for public information.

Public information	Methods of storage used
Recently movie you watched	Netflix YouTube Amazon prime
Contact details of the head of teaching/employer	Online school website (school directory) Email search bar
Coursework/study notes /reference notes	Notes Evernote Online storage platforms (google drive, one drive, drop box) Online school website Written notes Organized folders on desktop
Rules and regulations related to work/academia	Online school website
Bookmarks saved for Work/academia	Chrome bookmark manager, safari bookmark manager, Pocket (app that saves bookmarks)
Birthdays of your closest friends/family.	Facebook calendar

7.2.2 Private Information

When asked to find frequently and non-frequently used private information like the passport number and expiry date, birth certificate, CV (resume), most participants either had only hard copies which were in a different location or had previously scanned copies which took a while to retrieval.

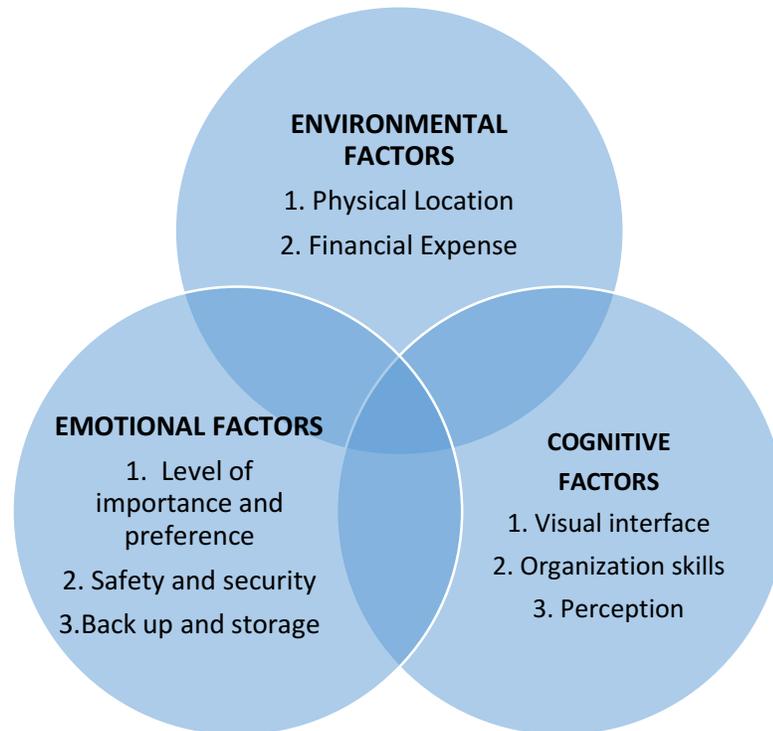
Table 10: Table showing most commonly used storage methods for private information.

Private information	Methods of storage used
Latest CV	Email Folder on a personal/work computer Online storage platforms(google drive , one drive, drop box)
Birth certificate	Hardcopy Email Folder on Desktop

log in to a bank account	Passwords saved in one password (app) Passwords saved in hard copy Passwords saved in notes application of a computer.
Receipt of a recent electronic purchase	Hard copy Amazon.com Email Budget manager
videos /home movies	Phone pictures iCloud Dropbox WhatsApp
Passport number and expiry date.	Hard copy Email Folder on a personal computer Evernote
log in to Gmail, OneDrive, Yahoo, LinkedIn	passwords saved and automatic login manual log ins
Professional pictures	Likened , folder on a personal computer, work computer , school webpage directory,
references for your dissertation/essay/grant paper	Mendeley Desktop Microsoft word PDF
Accommodation receipts or papers	Hard copy Online website Email
Educational/academic certificates	Hard disk (external drive), Hard copy Google drive One drive Folder on the desktop of a personal computer Email
Contact information of a classmate	Address book of a mobile phone Address book of outlook Search by typing in a Email ID

7.3 What are the factors that influence individuals while making decisions about storing information?

Figure 20: Factors causing data fragmentation.



Venn diagram created by the researcher.

7.3.1 Environmental factors

- **Physical Location**

The physical location of the data as well as the current location of the user has a major impact in the process of accessing and retrieving information. The participants were interviewed at the location of their preference depending on their daily activities. In a pool of 19 participants, 9 were interviewed at study laboratories, 8 at the work office and 2 at their home. The participants at the work office used work computers, personal iPads and/or phones. Most of them had most of the personal information stored on email and Dropbox but documents like passports and birth certificates were stored on personal computers at home or as hardcopies elsewhere. This made it impossible to access them. Some participants were unsure of where their information was stored. A participant who was interviewed at home stated "I know that the information exists, but I don't remember where". On asked about the overall experience of completing the task, a participant stated "I have my data all over the place, I usually use one drive for Microsoft office documents, drop box for personal data and google drive for university data, but it depends". Most participants found frequently used private information without a hassle

but non-frequently used private information was either hard to find or completely inaccessible.

- **Financial Expenses**

While individuals use different services for different types of data, the financial charges associated with storage services have a major impact on storage decisions.

There was contrast in the financial reasoning of the participants when asked about storage costs and decisions. Most participants who were employed stated that they were paying for storage services like Drop box and Google Drive for personal information while their employers provided them with free storage for work-related information. The participants who were unemployed said that they would rather delete information or store it in on a different device or service rather than paying for storage.

When asked about the overall experience of retrieving information a participant stated “Hard! It was hard because I use different services to store my data. When my data is full, I do not want to pay for storage services as I am unemployed. Therefore, I will just store it somewhere else. Although, I tend to forget where I put things.” Another participant stated “I have run out of storage before, but I just make a new account on a storage service website rather than paying for storage.

7.3.2 Emotional factors

- **Level of importance**

We generally organize documents based on the degree of importance of information. Most participants stored their current projects like references required for the completion of a thesis, the contact information of a colleague readily available on the desktop, where it is visually accessible. These documents were retrieved faster and accessed with great convenience. Participants were not able to retrieve resumes or educational documents because they had been working already and hence didn't feel the need to organize them or remember where they were stored. However, students could easily access this information as they were using these documents at this stage. The employed participants were not able to find passports for the same reason but and worried that if there was an immediate requirement, they would have to run home to access the passport number.

- **Safety and security**

Safety and security has become a major influence in the way we base our daily decisions. However, safety and security of personal information is differently interpreted by different individuals. Data is used by individuals as well as by organizations. The leak of sensitive information can be fateful for someone while the others might not have the same concern. In recent years, popular social media and cloud storage applications like Facebook, Twitter, Apple, Evernote and Microsoft have been subject to hacks which brings concern to many individuals[45]. A participant also stated that he does not trust

apps like Evernote after reading about the hack. One participant stated “I once used a storage service that provided 50 Giga bytes of free storage, but it was a new service, hence not as well known and dependable as google. New services generally don’t have money therefore they tend to shut and I might loose my data.” Users generally make unconscious storage decisions based on the sensitivity of certain information. Most participants stated that they are afraid of having so much information on their computer and are scared of someone hacking their computer and their whole life.

- **Back up and storage**

Most participants readily admitted to backing up only when the storage of the current device was full. Two employed participants stated they they had no idea how their information gets backed-up and assume that the university does it for them. Documents like passports are generally not backed-up and a participant stated that they are afraid of loosing their passport as they do not have a scanned copy. The participants that backed-up used a variety of devices like university free storage, one drive, google drive, external hard disks. A participant stated “I never look in the backup and need information on the desktop to remember it”. A participant stated “I have problems organizing my music because I download it from different places and its stored everywhere” while another stated “apps like Spotify have helped manage music as you do not have to download anything.

7.3.3 Cognitive factors

- **Visual interface**

The Visual interface of a device generally aids the cognitive process of remembering where information is stored. We use the desktop as a memory to help us remember where our information that is stored. Visual cues are very important in the process of finding and retrieving. Most participants admitted to having problems with remembering the names of folders. A participant admitted that while storing files in Google Drive he forgets to name them thus they are saved with the default name which makes it difficult to sort and retrieve in the time of need. Most participants feared that all of their information is online and if not connected to the internet it would be very difficult to retrieve them. Also, when documents are stored online they are away from the visual cognition.

- **Organization skills**

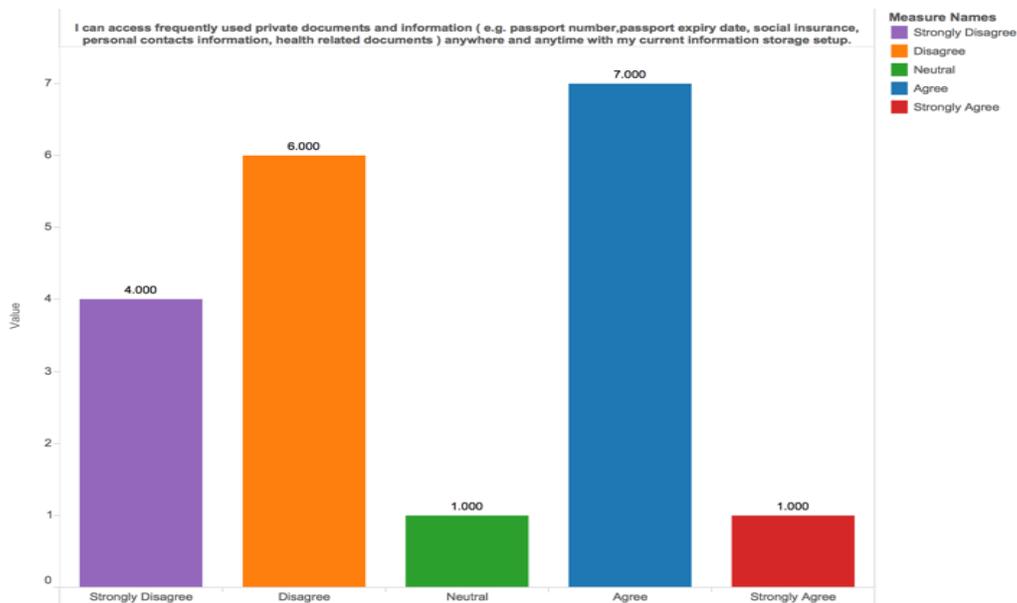
During the study, it was observed that the individuals who very well organized found the documents faster. The Most common problem faced by people was having different versions of documents stored across different storage services. For example, people had edited versions of CVs stored in folders and emails. Finding the right one was quite a task. This can also be applied to folders in computers and storage services. A participant stated “I am not good at naming folders but better than some others. I do spend time searching for the write version using modification dates.” Another interesting observation was that

employed individuals were more organized as compared to students. A participant stated “I am good at organizing because over the years I have improved”. It was seen that the younger generation used a higher number of storage services and spend more time looking for information.

- **Perception**

The most popular response when asked about the overall experience of retrieving documents was that they have now realized that they have their information all over the place. For example, when asked to retrieve frequently used private information, the participants who that agreed that they could access this information conveniently in the questionnaire were not able to do so in the task. They were either confused as to where the documents were stored or had no hard copies.

Figure 21: interview participants views on accessibility



Description: The examples provided for frequently used private information were Passport number and passport expiry date, social insurance, personal contacts information and health related documents. From a pool of 19 participants that were interviewed, 6 participants disagreed and 4 strongly disagreed that they can access their information conveniently, there was also 1 neutral response. While 7 participants agreed that they can access information and 1 strongly agreed on further analysis.

8. RESULTS AND DISCUSSION

While the main objective of this study was to identify and measure the severity of data fragmentation issues causing inefficient PIM practices, we first and foremost need to answer the question: Is personal information being subjected to data fragmentation? As outlined in chapter 3, there have been several studies in the past addressing the data fragmentation issues of personal information[10][6]. However, the term data fragmentation is relatively new in the context of personal information management as it is most commonly addressed in the field of database management systems. The literature review suggests that there is fragmentation of personal information across various tools like email, file manager, storage services. Very few studies addressed the issue of data fragmentation across different devices like desktop computers, laptops, external hard drives, mobile phones etc. Although, Kljun, Mariani and Dix's study on back up strategies of users on cross-device back-ups is close to our research field[12]. The results of the analysis of the interview data and the online survey data suggest that there is a fragmentation of information across devices and services. The causes were the differences in the way users of personal information treat different types of information. We found significant evidence that individuals treat frequently used private information and frequently used public information differently with respect to convenient accessibility, quickness in retrieval as well as safety and security differently (refer to index : 6.1.1, 6.2.1, 6.3.2). We also found significant evidence that people treat the accessibility of non-frequently used public information and frequently used public information differently and treat the safety and security of non-frequently used private information and frequently used private information differently (refer to index : 6.1.2, 6.3.1). Therefore, users tend to store them at different locations based on specific needs. Although we did not find significant evidence to prove the same for other comparisons, it was still clear that users feel differently about different types of documents from the literature[29], [30], [33]. The analysis of the interview transcripts also suggest the same. The most repeated answer when asked about the overall experience of find documents was "My information is all over the place"(refer to index : 7.3.1). It was found that there several factors that influenced the behaviors of users to fragment their data across several services and devices i.e. environmental factors, emotional factors and cognitive factors (refer to index : 7.3). For instance, individuals who had been organizing information for many years and were in the age range of 29 to 49 said that they have learned to organize information over years of experience and use specific set of tools and services for storing information. Consequently, the individuals belonging to the age range of 18 to 29 said that they found it hard because they use multiple free storage services due to financial shortcomings (refer to index : 7.3). Another important factor that has a major influence on the behavior patterns of individuals is the safety and security of their personal information. We found significant evidence that it would troubling and/or cause distress if data suddenly becomes available to the public, with a hypothesized median of 3 (neutral) and an observed median of 4.5 (agree – strongly agree). From the analysis of the transcripts of the interview, a participant said that he prefers to store data on services like google drive because more popular services are dependable.

Participants used various services like drop box, google drive, one drive, internal hard drive of computer etc. (refer to index: 7.2). From the quantitative analysis of the survey questionnaire, we found significant evidence that individuals have multiple copies of the same information in different storage locations where the hypothesized median was 3 (Neutral response) while the observed median was found to be 4 (Agree) This could be one of the issues that causes by data fragmentation problems as if it difficult to remember the name of the required file from the right location[33]. Participants from the interview had

multiple copies of their CV stored in different places which made it difficult to retrieve the latest version which further proves the case (refer to index: 7.3). We also found significant evidence saying that users data is fragmented/spread over too many different services and/or devices where the hypothesized median was 3 (Neutral response) and the observed median was found to be 4.5 (Agree- Strongly agree) which shows the severity of the data fragmentation. Vile (2013) addresses the data fragmentation across various applications and devices but in a business context[6]. The analysis of the transcripts from the task found that participants did not have access to documents like passports, receipts etc. because the documents were either in a paper-based format store somewhere else, or in different computer which was physically inaccessible due to the computer being stored at a different location (refer to index: 7.3). This shows that physical distribution of storage devices hinders the process of locating information. To sum up, Individual behaviors have major impacts on the severity of data fragmentation as stated in previous studies of this nature as well as this research project. Further research in the following field can have major benefits in providing suggestions for designing PIM systems that better cater to the needs of the users and help reduce data fragmentation.

9. PERSONAL REFLECTION

The overall experience of completing the following dissertation has had a good influence on the confidence levels of the researcher. The collection of primary data was the most interesting part of the study as it involved communicating with my peers and immediate superiors. During the course of collecting interview data, there were several instances that were surprising to the researcher. The responses from employees working within the university were very warm and welcoming. The most challenging task of the following project was time management. There were several re-adjustments to the timeline of the plan of actions due to improper management of time for collecting data, as there were two different forms of data collection i.e. interview, questionnaire survey. Analyzing the results for the quantitative methods took maximum effort due to unknowingness, however it aided in the learning experience. The supervisors of the study provided major support and guidance from the early stages of the project, brainstorming for ideas to the selection of non-parametric tests.

10. RECOMMENDATIONS AND FUTURE WORK

The results from this study indicate that personal behaviors have major influences on the causes of data fragmentation. The future of personal information will have increased volumes of data being created in different forms and saved across different upcoming digital storage devices. Personal information management systems need to be interactive, visually appealing as well as flexible to the needs of all users in order to reduce the consequences of data fragmentation across tools and devices.

The information management systems need to have the following capabilities:

- Improved organization functions for information that is user-centered.
- A personal information manager that keeps a record of the locations of personal information across tools as well as devices and aid in organizing, retrieving and accessing information based on context subjective approach.
- Further improvements on user-driven and system-driven storage strategies in already available personal informatics systems[25].

11. CONCLUSION

In conclusion, data fragmentation issues are only going to increase with the advancements in technology. There are several personal information storage systems available that allow users to store information based on their needs and usage capabilities but not a lot of research in the field of information retrieval aiding systems. This research project outlines the major causes and consequences of data fragmentation on personal information used for different purposes. When storing information, individuals consider factors like frequency of use, context, safety, security, back up, systems capabilities, physical access to the device etc. and not all storage systems take consideration of all the factors which forces users to use multiple services and devices. With further research in the following field, the results will benefit all stakeholders.

12. REFERENCES

- [1] W. Q. Wei, C. L. Leibson, J. E. Ransom, A. N. Kho, P. J. Caraballo, H. S. Chai, B. P. Yawn, J. A. Pacheco, and C. G. Chute, "Impact of data fragmentation across healthcare centers on the accuracy of a high-throughput clinical phenotyping algorithm for specifying subjects with type 2 diabetes mellitus," *J Am Med Inf. Assoc*, vol. 19, no. 2, pp. 219–224, 2012.
- [2] A. Zuiderwijk, M. Janssen, S. Choenni, R. Meijer, and R. S. Alibaks, "Socio-technical Impediments of Open Data," *Electron. J. e-Government*, vol. 10, no. 2, pp. 156–172, 2012.
- [3] H. Hermjakob, "The HUPO proteomics standards initiative - Overcoming the fragmentation of proteomics data," *Proteomics*, vol. 1, no. 1–2 SUPPL., pp. 34–38, 2006.
- [4] C. W. Pollard, *The Soul of the Firm*. Zondervan, 1996.
- [5] M. T. Özsu and P. Valduriez, "Distributed Database Design," *Princ. Distrib. Database Syst.*, no. Chapter 5, pp. 71–129, 2011.
- [6] D. Vile, "Storage Anywhere and Everywhere," *Free. Dyn. Ltd*, no. June, pp. 1–17, 2013.
- [7] "Key definitions of the Data Protection Act," 2016. [Online]. Available: <https://ico.org.uk/for-organisations/guide-to-data-protection/key-definitions/>. [Accessed: 11-Aug-2016].
- [8] W. Jones and W. Jones, "Personal Information Management," no. 1996, pp. 1–67, 2005.
- [9] L. Zhou, A. S. Mohammed, and D. Zhang, "Mobile personal information management agent: Supporting natural language interface and application integration," *Inf. Process*.

- Manag.*, vol. 48, no. 1, pp. 23–31, 2012.
- [10] O. Bergman, T. Aviv, R. Beyth-marom, and R. Nachmias, "The Project Fragmentation Problem in Personal Information Management," *CHI 2006 Proc. • Pers. Inf. Manag.*, no. September 2015, pp. 271–274, 2006.
- [11] M. W. Lansdale, "The psychology of personal information management," *Appl. Ergon.*, vol. 19, no. 1, pp. 55–66, 1988.
- [12] R. McKerlich, C. Ives, and R. McGreal, "Toward Understanding Short-Term Personal Information Preservation: A Study of Backup Strategies of End Users Matjaž," *Int. Rev. Res. Open Distance Learn.*, vol. 14, no. 4, pp. 90–103, 2013.
- [13] J. Teevan and W. Jones, "Personal Information Management (PIM) 2008," vol. 42, no. 2, pp. 99–105, 2008.
- [14] W. Jones, *Keeping found things found: The study and practice of personal information management*. Morgan Kaufmann, 2010.
- [15] O. Bergman, R. Beyth-Marom, and R. Nachmias, "The user-subjective approach to personal information management systems," *J. Am. Soc. Inf. Sci. Technol.*, vol. 54, no. 9, pp. 872–878, 2003.
- [16] T. W. Malone, "How Do People Organize Their Desks? Implications for the Design of Office Information Systems," vol. 1, no. 1, pp. 99–112, 1983.
- [17] S. Whittaker and J. Hirschberg, "The character, value and management of paper archives," vol. 8, no. 2, pp. 150–170, 2001.
- [18] D. Abrams, R. Baecker, and M. Chignell, "Information archiving with bookmarks: Personal Web Space Construction and Organization," *Proc. SIGCHI Conf. Hum. factors Comput. Syst. - CHI '98*, pp. 41–48, 1998.
- [19] W. E. Mackay, "More Than Just a Communication System : Diversity in the Us & of Electronic Mail," *CSCW-88 Proc. Conf. Comput. Co-op. Work*, pp. 344–353, 1988.
- [20] J. Gwizdka, "Email task management styles: the cleaners and the keepers," *Ext. Abstr. 2004 Conf. Hum. factors Comput. Syst. - CHI '04*, no. August, p. 1235, 2004.
- [21] E. Horvitz, S. Dumais, and P. Koch, "Models of Memorability : Learning , Experiments , and Applications ONLINE CALENDARS AND PROPERTIES," pp. 1–8.
- [22] M. N. Huhns and L. M. Stephens, "A Personal Ontology," *IEEE Internet Comput.*, no. October, pp. 85–87, 1999.
- [23] G. Bell, "A personal digital store," *Commun. ACM*, vol. 44, no. 1, pp. 86–91, 2001.
- [24] J. Gemmell, G. Bell, and R. Lueder, "MyLifeBits: a personal database for everything," *Commun. ACM*, vol. 49, no. 1, pp. 88–95, 2006.
- [25] I. Li, A. Dey, and J. Forlizzi, "A stage-based model of personal," pp. 557–566, 2010.
- [26] R. Boardman, "Too many hierarchies? The daily struggle for control of the workspace," *Proc. HCI Int. 2003*, vol. 1, no. March 2003, pp. 616–620, 2003.
- [27] D. S. Kirk, S. Izadi, A. Sellen, S. Taylor, R. Banks, and O. Hilliges, "Opening up the family archive," *Proc. 2010 ACM Conf. Comput. Support. Coop. Work (CSCW '10)*, pp. 261–270, 2010.
- [28] I. Cole, "Human aspects of office filing: Implications for the electronic office," *Proc. Hum. Factors Ergon. Soc. Annu. Meet. SAGE Publ.*, vol. 26, no. 1, pp. 59–63, 1982.
- [29] B. Nardi, K. Anderson, and T. Erickson, "Filing and finding computer files," *Proc. East-West HCI*, 1995.

- [30] D. Barreau and B. a Nardi, "Finding and Reminding," *SIGCHI Bull.*, vol. 27, no. 3, pp. 39–43, 1995.
- [31] D. K. Barreau, "Information systems for organizations and the problem of ephemeral information," *Univ. Maryl. Coll. Park.*, 1997.
- [32] R. Boardman and M. A. Sasse, "'Stuff goes into the computer and doesn't come out,'" *2004 Conf.*, vol. 6, no. 1, pp. 583–590, 2004.
- [33] D. Barreau, "The Persistence of Behavior and Form in the Organization of Personal Information," *J. Am. Soc. Inf. Sci. Technol.*, vol. 59, no. 2, pp. 307–317, 2008.
- [34] O. Bergman, R. Beyth-Marom, and R. Nachmias, "The User-Subjective Approach to Personal Information Management Systems Design: Evidence and Implementations Ofer," *J. Am. Soc. Inf. Sci. Technol.*, vol. 59, no. 2, pp. 235–246, 2008.
- [35] P. Dourish, W. K. Edwards, A. LaMarca, and M. Salisbury, "Presto: an experimental architecture for fluid interactive document spaces," *ACM Trans. Comput. Interact.*, vol. 6, no. 2, pp. 133–161, 1999.
- [36] W. Jones, W. Jones, A. J. Phuwanartnurak, A. J. Phuwanartnurak, R. Gill, R. Gill, H. Bruce, and H. Bruce, "Don't Take My Folders Away! Organizing Personal Information to Get Things Done," *CHI'05 Ext. Abstr. Hum. factors Comput. Syst.*, pp. 1505–1508, 2004.
- [37] D. K. Barreau, "Context as a factor in personal information management systems," *J. Am. Soc. Inf. Sci. Technol.*, vol. 46, no. 5, pp. 327–339, 1995.
- [38] B. H. Kwaznik, "The Importance of Factors That Are Not Document Attributes in the Organisation of Personal Documents," *J. Doc.*, vol. 47, no. 4, pp. 389–398, 1991.
- [39] "OneDrive, Dropbox, Google Drive and Box: Which cloud storage service is right for you? - CNET." [Online]. Available: <http://www.cnet.com/uk/how-to/onedrive-dropbox-google-drive-and-box-which-cloud-storage-service-is-right-for-you/>.
- [40] H. E. Growing, M. For, N. Storage, I. S. a Result, O. F. The, E. Demand, F. O. R. Storage, C. In, O. U. R. Increasingly, D. World, I. T. S. Tight, L. Market, A. Networks, N. A. Storage, a R. E. Two, P. Approaches, T. O. Networking, W. H. a S. One, W. Doesn, N. Practice, I. T. I. S. Often, and T. N. Hardware, "Network attached storage architecture," *Commun. ACM*, vol. 43, no. 11, pp. 37–45, 2000.
- [41] D. Bertram, "Likert Scales... are the meaning of life :," *Univ. Calagary, Dep. Comput. Sci.*, p. pages.cpsc.ucalgary.ca/~saul/wiki/uploads/CPSC681/, 2007.
- [42] "Audience: Market Research and Analysis Tool | SurveyMonkey Audience." [Online]. Available: <https://www.surveymonkey.com/mp/audience/>.
- [43] D. Soergel, *Organizing information: principles of data base and retrieval systems*. Elsevier, 1985.
- [44] A. Civan, W. Jones, P. Klasnja, and H. Bruce, "Better to organize personal information by folders or by tags?: The devil is in the details," *Proc. Am. Soc. Inf. Sci. Technol.*, vol. 45, no. 1, pp. 1–13, 2008.
- [45] "Hacker group that hit Twitter, Facebook, Apple and Microsoft intensifies attacks | PCWorld." [Online]. Available: <http://www.pcworld.com/article/2945632/hacker-group-that-hit-twitter-facebook-apple-and-microsoft-intensifies-attacks.html>.



University
of
St Andrews

INFORMATION AND CONSENT FORM

What is the study about?

The study investigates the storage of personal information across several platforms and the problems faced by individual users in retrieving this information i.e. timeliness, ease etc. The study is being conducted as part of the researcher's masters project for the School of Computer Science, University of St Andrews. Filling this survey takes about 30 minutes. Participation is voluntary and if you do decide to take part, you will be free to withdraw at any time without providing a reason.

What would I be required to do?

You will be required to fill out the online questionnaire regarding the methods of storing personal information as well as work related information across various devices and services. This will include the type of information (public as well as private data), the name of the storage device etc.

Do I have to take part?

This information page has been written to help you decide if you would like to take part. It is up to you and you alone whether or not to take part. If you do decide to take part, you will be free to withdraw at any time without providing a reason.

Will my participation be Anonymous and Confidential?

Only the researcher and supervisors will have access to the data which will be kept confidential and the data will be stored in an encrypted format.

Storage and Destruction of Data Collected

The data we collect will be accessible by the researchers and supervisor involved in this study only. It will be stored in an encrypted format. The collected data will be stored for 5 years and can be a valuable resource for future studies; therefore, we ask your consent to maintain it. Again, we

stress that all information is kept confidential, and the analyses of data will be done without specifying of individual identities.

What will happen to the results of the research study?

The results will be finalized by September 2016 and written up as part of the researcher's master dissertation.

Are there any potential risks to taking part?

No

Consent and Approval.

This research proposal has been scrutinized and been granted Ethical Approval through the University ethical approval process.

What should I do if I have concerns about this study?

A full outline of the procedures governed by the University Teaching and Research Ethical Committee is available at <http://www.st-andrews.ac.uk/utrec/guidelinespolicies/complaints/>

Questions

If you have any questions about this study, the results or your involvement, please feel free to email the researchers listed below.

Contact Details

Researcher:

Pooja Basavaraj Balekundargi Contact Details: email: pbbvb@st-andrews.ac.uk

Supervisors:

Dr. Graham Kirby Email: graham.kirby@st-andrews.ac.uk Phone: +44 (0)1334 46 3240

Dr. Miguel Nacenta Email: mans@st-andrews.ac.uk Phone: +44 (0)1334 463265

* By clicking on the “continue” button below you accept that:

You have read and understood the information sheet.

You have been given the opportunity to ask questions about the study.

You understand that you can withdraw from the study at any time without having to give an explanation.

You understand that your data will be confidential and that it will not contain any identifiable personal information.

You agree to take part in the study

CONTINUE

A survey on data fragmentation issues for users.

Please select your age range and specify your educational background and current occupation in the box below.

- 18 to 29
- 30 to 39
- 40 to 49
- 50 to 59
- 60 to 79
- 80 and above

For Example, say you have BSc in Economics and/or are currently working as a sales associate at an iron-ore manufacturing plant or waiting tables at a restaurant.

Select your agreement with the following sentences.

Strongly Disagree Disagree Neutral Agree Strongly Agree

My data is fragmented/spread over too many different services and/or devices.

I am good at storing and retrieving my own information.

I find that documents stored different formats are inaccessible by different devices.

It would be troubling and/or cause me distress if my data suddenly became available to the public.

I sometimes have multiple copies of the same information in different storage locations.

Storing data is expensive.

How conveniently can you **access** information anywhere and anytime with your current information storage setup? Select your agreement with the following sentences.

Strongly Disagree Disagree Neutral Agree Strongly Agree

I can access non-frequently used private documents and information (e.g. insurance documents, bank account numbers and statements, old home movies) anywhere and anytime with my current information storage setup.

Strongly Disagree

Disagree

Neutral

Agree

Strongly Agree

I can access frequently used private documents and information (e.g. passport number,passport expiry date, social insurance, personal contacts information, health related documents) anywhere and anytime with my current information storage setup.

I can access non-frequently used public documents and information (e.g. user manuals, rules and regulations related to work/academia, scientific journals) anywhere and anytime with my current information storage setup.

I can access my frequently used public documents and information (e.g. academic coursework, study notes, movies) anywhere and anytime with my current information storage setup.

How is your experience with regards to the **safety and security** of your information? Select your agreement with the following sentences.

Strongly Disagree Disagree Neutral Agree Strongly Agree

I feel secure and safe when storing non-frequently used private documents and information (e.g. insurance documents, bank account numbers and statements, old home movies) with my current information storage setup.

I feel secure and safe when storing frequently used private documents and information (e.g. passport number, passport expiry date, social insurance number, personal contacts information, health related documents) with my current information setup.

I feel secure and safe when storing non-frequently used public documents and information (e.g. user manuals, rules and regulations related to work/academia, Scientific journals) with my current information storage setup.

I feel secure and safe when storing frequently used public documents and information (e.g., academic coursework, study notes, movies) with my current information storage setup.

How long does it take to **retrieve** information from your current information storage setup? Select your agreement with the following sentences.

Strongly Disagree

Disagree

Neutral

Agree

Strongly Agree

I can retrieve non-frequently used private documents and information (e.g. insurance documents, bank account numbers and statements, old home movies) very quickly with my current information storage setup.

I can retrieve frequently used private documents and information (e.g. passport number, passport expiry date, social insurance number, personal contacts information, health related documents) very quickly with my current information storage setup.

I can retrieve non-frequently used public documents and information (e.g. user manuals, rules and regulations related to work/academia. Scientific journals) very quickly with my current information storage setup.

I can retrieve frequently used public documents and information (e.g. academic coursework, study notes, movies) very quickly with my current information storage setup.

Indicate storage options that you have used, use regularly or are part of your main methods to store information. Tick all that apply.

	Never Used This Method	Use to but not anymore	Used occasionally	Part of my main routine
Cloud storage (e.g. Google Drive, Microsoft OneDrive)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cloud application (e.g. Dropbox, Box)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
E-mail to myself	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Local hard-drive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Storage provided by institution (e.g. School home directory, SharePoint)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Backup storage (e.g. IDrive, CrashPlan)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Paper copies (hard copy)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Other (please specify)

Are you well informed about the storage location of your information?

Tick all that apply

- I always know where the required information is stored.
- I browse through folders and menus until I find the required information.
- I generally use the search/find option to find the required information.

Other (please specify)

What kind of device or storage service do you use to store the following kinds of information? (Name several if you use more than one).

Work-related documents and information:

Personal Contact Information:

Financial Information:

Academic Information:

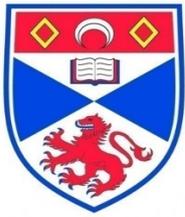
Health-related information:

Personal identification information:

What fraction of your information is backed up? Tick the box applies.

- All of my information is backed up.
- Some of my information is backed up.
- Only very important information is backed up.
- None of my information is backed up.

How often if at all, do you back up your information? What devices or services do you use to back up this information?



University
of
St Andrews

PARTICIPANT INFORMATION SHEET

Project Title: Survey on Data Fragmentation Issues for Users.

What is the study about?

The study investigates the storage of personal information across several platforms and the problems faced by individual users in retrieving this information i.e. timeliness, ease etc. The overall process will last approximately for 60 minutes. It will be done in two parts namely the questionnaire survey followed by the interview. The study is being conducted as part of the researcher's masters project for the School of Computer Science, University of St Andrews. Participation is voluntary and if you do decide to take part, you will be free to withdraw at any time without providing a reason.

What would I be required to do?

You will be required to fill out a questionnaire before commencing the face to face interview. At the end of the questionnaire you will be asked to list out several projects from their daily lives which you are comfortable with. For example, methods of storing personal information as well as work related information across various devices. This will include the type of information (purpose), name of the storage device and the approximate date of storage. One of these projects will be selected for the interview and you will then be asked open ended questions based on relevance. You will then be asked to give a guided tour of how you would retrieve this information from the location of storage. The interview will be recorded on a digital audio device which will be used for further analysis.

Will my participation be Anonymous and Confidential?

Only the researcher and supervisors will have access to the data which will be kept confidential. Your identity will not be revealed to anyone whatsoever, and all reference to you will be of a generic nature. Your permission may be sought in the Participant Consent form for the data you provide, which will be confidential, to be used for future scholarly purposes.

Storage and Destruction of Data Collected

The data we collect will be accessible by the researcher and supervisors involved in this study only. It will be stored in an encrypted format. The collected data will be stored for 5 years and can be a valuable resource for future studies; therefore, we ask your consent to maintain it. Again, we stress that all information is kept confidential, and the analyses of data will be done without specifying of individual identities.

What will happen to the results of the research study?

The results will be finalized by September 2016 and written up as part of the researcher's master dissertation.

Are there any potential risks to taking part?

No

Consent and Approval.

This research proposal has been scrutinized and been granted Ethical Approval through the University ethical approval process.

What should I do if I have concerns about this study?

A full outline of the procedures governed by the University Teaching and Research Ethical Committee is available at (<http://www.st-andrews.ac.uk/utrec/guidelinespolicies/complaints/>)

Questions.

If you have any questions about this study, the results or your involvement, please feel free to email the researcher.

Contact Details

Researcher: Pooja Basavaraj Balekundargi .*Contact Details:* email: pbbvb@st-andrews.ac.uk

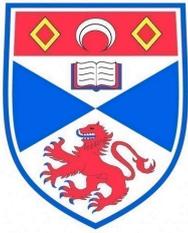
Phone: +44 7867213215

Supervisors: Dr. Graham Kirby Email: graham.kirby@st-andrews.ac.uk

Phone: +44 (0)1334 46 3240

Dr. Miguel Nacenta Email: mans@st-andrews.ac.uk

Phone: +44 (0)1334 463265



University
of
St Andrews

Participant Consent Form

Coded Data

Project Title

Survey on Data Fragmentation Issues for Users

Researcher(s) Name(s)

Pooja Basavaraj Balekundargi

Supervisors Names

Dr. Graham Kirby, Dr. Miguel Nacenta

The University of St Andrews attaches high priority to the ethical conduct of research. We therefore ask you to consider the following points before signing this form. Your signature confirms that you are happy to participate in the study.

What is Coded Data?

The term 'Coded Data' refers to when data collected by the researcher is identifiable as belonging to a particular participant but is kept with personal identifiers removed. The researcher retains a 'key' to the coded data which allows individual participants to be re-connected with their data at a later date. The un-coded data is kept confidential to the researcher and Supervisors. If consent is given to archive data, the participant may be contacted in the future by the original researcher.

Consent

The purpose of this form is to ensure that you are willing to take part in this study and to let you understand what it entails. Signing this form does not commit you to anything you do not wish to do and you are free to withdraw at any stage.

The raw collected data will be stored in a secure encrypted format and access will be provided to the researcher and the supervisors only. The processed data in the form of anonymised quotes, summaries and statistics will be published as part of my thesis and possibly academic publications.

By ticking yes in the boxes below with each statement concerning the collection and use of the research data you agree that:

- | | | |
|---|------------------------------|-----------------------------|
| You have read and understood the information sheet. | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| You have been given the opportunity to ask questions about the study. | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| You have had your questions answered satisfactorily. | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| You understand that you can withdraw from the study at any time without having to give an explanation. | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| You understand that your data will be confidential and that it will contain identifiable personal data but that will be stored with personal identifiers removed by the researcher and that only the researcher/supervisor will be able to decode this information as and when necessary. | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| You agree to take part in the study | <input type="checkbox"/> Yes | <input type="checkbox"/> No |

Part of the research involves taking audio recordings of the interviews that will be used for analysis and summary of findings. These recordings will be kept secure and stored with no identifying factors.

You agree to being audio recorded.

Yes No

You agree for tape recorded material to be published as part of this research

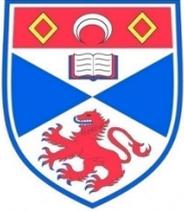
Yes No

Participation in this research is completely voluntary and your consent is required before you can participate in this research. If you decide at a later date that data should be destroyed, we will honour your request in writing.

Name in Block Capitals _____

Signature _____

Date _____



University
of
St Andrews

PARTICIPANT DEBRIEFING FORM

Project Title

Survey on Data Fragmentation Issues for Users

Researchers Name

Pooja Basavaraj Balekundargi

Nature of Project

This postgraduate research project was conducted to make an analysis of the difficulties faced by users and the severity of the issues caused due to data fragmentation. With the improvements in technology, there is colossal amounts of data being created and stored on a daily basis. The following data is then used by organizations and individuals for efficient decision making. However, this data is stored using different types platforms. I.e. public/private cloud storage services, mobile devices, paper based storage, computer devices/servers, hard disks etc. This results in information being dispersed at various locations hence causing data fragmentation problems.

The intention of this research project is to investigate on the problems faced by an individual user in the storage, maintenance and finding/retrieval of information in an easy and and timely manner. We expect to identify and measure the severity of the issues caused by data fragmentation and try to find the behavioral implications that are causing the issues.

Storage of Data

As outlined in the Participant Information Sheet your data will now be retained for a period of 5 years before being destroyed. Your data will remain accessible to only the researchers and supervisors. If you no longer wish for your data to be used in this manner you are free to withdraw your consent by contacting any of the researchers and or Supervisor.

What should I do if I have concerns about this study?

A full outline of the procedures governed by the University Teaching and Research Ethical Committee is available at (<http://www.st-andrews.ac.uk/utrec/guidelinespolicies/complaints/>)

Contact Details

Researcher: Pooja Basavaraj Balekundargi, Email: pbbvb@st-andrews.ac.uk

Phone: +44 7867213215

Supervisors: Dr. Graham Kirby Email: graham.kirby@st-andrews.ac.uk

Phone: +44 (0)1334 46 3240

Dr. Miguel Nacenta Email: mans@st-andrews.ac.uk

Phone: +44 (0)1334 463265

13.7 Interview Questions

Work/academia related documents	Type of document
1. Question: Can you retrieve the details of 2 meetings/appointment reminders in the first week of July. Details: the location , time and the names of the colleagues/peers that were involved in the meeting.	PR-F
2. Question: Can you retrieve the contact information of 2 of your classmates/co - worker Details: Email id and phone number.	PR-F AND PU-F
3. Bookmarks saved for the purpose of Work/academia. Question: Can you retrieve 5 bookmarks that you saved in the month of July. Details: The URL of the saved links.	PU-NF
4. Educational/academic certificates Question: can you retrieve your educational/academic certificates from your undergraduate degree.	PR-NF
5. School teacher's names: Can you retrieve the name and contact details of the head of teaching of our department/school?	PU-NF
6. Question: Do you live? can you retrieve the University accommodation receipts?	PU-NF
7. Coursework/study notes Retrieve study notes that you have recently used for your course work / exams / grant paper/ reference notes.	PR-NF
8. Rules and regulations Question: Can you retrieve this rule that applies to all St Andrews student. It is the Academic Appeals and Student Discipline or information and Communication Technologies of St Andrews	PU-NF
9. Can you 5 retrieve references for your dissertation/essay/grant paper?	PR-F
10. Can you retrieve Professional pictures that you have used for CV or anything similar to a job profile?	PR-NF and PU-NF

private documents	Type of document
1. Can you retrieve the passwords of your Gmail, OneDrive, yahoo, linkdn , online bank account .	PR-F and PR-NF
2. Can you retrieve a family videos / home movies of yourself.	PR-NF
3. What was the recent movie you watched that was not on the TV or the movie theater? Can you retrieve it for me?	PU-NF and PU-F
4. Can you retrieve the Contact details of your mother?	PR-NF
5. Can your retrieve your Passport number and expiry date.	PR-NF
6. Can you retrieve 5 birthdays of your closest friends/family.	PR-NF
7. Can you find me a copy of your birth certificate?	PR-NF
8. Can you find me your latest CV ?	
9. What was your latest/most recent electronic purchase ? can you find me the receipt for it?	PR-NF
10. How many bank accounts do u currently have? Can you log in to one of them?	PR-F and PR-NF

TASK: 15 minutes

INTERVIEW 5 to 10 minutes

The open ended questions:

1. how the overall experience of finding all this information?
2. What are you views on the safety and security of these documents?
3. Are you satisfied with the storage capabilities of the services and devices that you are currently using?

13.8 Related samples test : Accessibility

```

GET
  FILE='/Users/balekundargib/Desktop/First submission/NEW_ANALYSIS.sav'.
DATASET NAME DataSet1 WINDOW=FRONT.
NPAR TESTS
  /WILCOXON=NFPUIAccess NFPUIAccess NFPUIAccess WITH FPUiAccess FPIAccess N
FPIAccess (PAIRED)
  /MISSING ANALYSIS.

```

NPar Tests

		Notes
Output Created		13-AUG-2016 01:06:...
Comments		
Input	Data	/Users/balekundargib/Desktop/First submission/NEW_ANALYSIS.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	72
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics for each test are based on all cases with valid data for the variable(s) used in that test.
Syntax		NPAR TESTS /WILCOXON=NFPUIAccess NFPUIAccess NFPUIAccess WITH FPUiAccess FPIAccess NFPUIAccess (PAIRED) /MISSING ANALYSIS.
Resources	Processor Time	00:00:00.00
	Elapsed Time	00:00:00.00
	Number of Cases Allowed^a	349525

a. Based on availability of workspace memory.

```
[DataSet1] /Users/balekundargib/Desktop/First submission/NEW_ANALYSIS.sav
```

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
I can access frequently used public documents and information anywhere and anytime with my current information storage setup. - I can access non-frequently used public documents and information anywhere and anytime with my current information storage setup.	Negative Ranks	9 ^a	10.50	94.50
	Positive Ranks	20 ^b	17.03	340.50
	Ties	43 ^c		
	Total	72		
I can access frequently used private documents and information anywhere and anytime with my current information storage setup. - I can access non-frequently used public documents and information anywhere and anytime with my current information storage setup.	Negative Ranks	23 ^d	17.59	404.50
	Positive Ranks	11 ^e	17.32	190.50
	Ties	38 ^f		
	Total	72		
I can access non-frequently used private documents and information anywhere and anytime with my current information storage setup. - I can access non-frequently used public documents and information anywhere and anytime with my current information storage setup.	Negative Ranks	21 ^g	18.57	390.00
	Positive Ranks	13 ^h	15.77	205.00
	Ties	38 ⁱ		
	Total	72		

- a. I can access frequently used public documents and information anywhere and anytime with my current information storage setup. < I can access non-frequently used public documents and information anywhere and anytime with my current information storage setup.
- b. I can access frequently used public documents and information anywhere and anytime with my current information storage setup. > I can access non-frequently used public documents and information anywhere and anytime with my current information storage setup.
- c. I can access frequently used public documents and information anywhere and anytime with my current information storage setup. = I can access non-frequently used public documents and information anywhere and anytime with my current information storage setup.
- d. I can access frequently used private documents and information anywhere and anytime with my current information storage setup. < I can access non-frequently used public documents and information anywhere and anytime with my current information storage setup.
- e. I can access frequently used private documents and information anywhere and anytime with my current information storage setup. > I can access non-frequently used public documents and information anywhere and anytime with my current information storage setup.
- f. I can access frequently used private documents and information anywhere and anytime with my current information storage setup. = I can access non-frequently used public documents and information anywhere and anytime with my current information storage setup.
- g. I can access non-frequently used private documents and information anywhere and anytime with my current information storage setup. < I can access non-frequently used public documents and information anywhere and anytime with my current information storage setup.
- h. I can access non-frequently used private documents and information anywhere and anytime with my current information storage setup. > I can access non-frequently used public documents and information anywhere and anytime with my current information storage setup.
- i. I can access non-frequently used private documents and information anywhere and anytime with my current information storage setup. = I can access non-frequently used public documents and information anywhere and anytime with my current information storage setup.

Test Statistics^a

	I can access frequently used public documents and information anywhere and anytime with my current information storage setup. - I can access non-frequently used public documents and information anywhere and anytime with my current information storage setup.	I can access frequently used private documents and information anywhere and anytime with my current information storage setup. - I can access non-frequently used public documents and information anywhere and anytime with my current information storage setup.	I can access non-frequently used private documents and information anywhere and anytime with my current information storage setup. - I can access non-frequently used public documents and information anywhere and anytime with my current information storage setup.
Z	-2.777 ^b	-1.898 ^c	-1.633 ^c
Asymp. Sig. (2-tailed)	.005	.058	.102

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

c. Based on positive ranks.

```

NPAR TESTS
  /WILCOXON=FPUIAccess FPUIAccess FPIAccess WITH NFPIAccess FPIAccess NFPIA
ccess (PAIRED)
  /MISSING ANALYSIS.

```

NPar Tests

Notes

Output Created	13-AUG-2016 01:43:...	
Comments		
Input	Data	/Users/balekundargib/ Desktop/First submission/NEW_ANAL YSIS.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	72
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics for each test are based on all cases with valid data for the variable(s) used in that test.
Syntax	NPAR TESTS /WILCOXON=FPUIAcces s FPUIAccess FPIAccess WITH NFPIAccess FPIAccess NFPIAccess (PAIRED) /MISSING ANALYSIS.	
Resources	Processor Time	00:00:00.00
	Elapsed Time	00:00:00.00
	Number of Cases Allowed^a	393216

a. Based on availability of workspace memory.

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
I can access non-frequently used private documents and information anywhere and anytime with my current information storage setup. - I can access frequently used public documents and information anywhere and anytime with my current information storage setup.	Negative Ranks	30^a	22.53	676.00
	Positive Ranks	11^b	16.82	185.00
	Ties	31^c		
	Total	72		
I can access frequently used private documents and information anywhere and anytime with my current information storage setup. - I can access frequently used public documents and information anywhere and anytime with my current information storage setup.	Negative Ranks	27^d	21.46	579.50
	Positive Ranks	10^e	12.35	123.50
	Ties	35^f		
	Total	72		
I can access non-frequently used private documents and information anywhere and anytime with my current information storage setup. - I can access frequently used private documents and information anywhere and anytime with my current information storage setup.	Negative Ranks	10^g	13.40	134.00
	Positive Ranks	13^h	10.92	142.00
	Ties	49ⁱ		
	Total	72		

- a. I can access non-frequently used private documents and information anywhere and anytime with my current information storage setup. < I can access frequently used public documents and information anywhere and anytime with my current information storage setup.
- b. I can access non-frequently used private documents and information anywhere and anytime with my current information storage setup. > I can access frequently used public documents and information anywhere and anytime with my current information storage setup.
- c. I can access non-frequently used private documents and information anywhere and anytime with my current information storage setup. = I can access frequently used public documents and information anywhere and anytime with my current information storage setup.
- d. I can access frequently used private documents and information anywhere and anytime with my current information storage setup. < I can access frequently used public documents and information anywhere and anytime with my current information storage setup.
- e. I can access frequently used private documents and information anywhere and anytime with my current information storage setup. > I can access frequently used public documents and information anywhere and anytime with my current information storage setup.
- f. I can access frequently used private documents and information anywhere and anytime with my current information storage setup. = I can access frequently used public documents and information anywhere and anytime with my current information storage setup.
- g. I can access non-frequently used private documents and information anywhere and anytime with my current information storage setup. < I can access frequently used private documents and information anywhere and anytime with my current information storage setup.
- h. I can access non-frequently used private documents and information anywhere and anytime with my current information storage setup. > I can access frequently used private documents and information anywhere and anytime with my current information storage setup.
- i. I can access non-frequently used private documents and information anywhere and anytime with my current information storage setup. = I can access frequently used private documents and information anywhere and anytime with my current information storage setup.

Test Statistics^a

	<p>I can access non-frequently used private documents and information anywhere and anytime with my current information storage setup. - I can access frequently used public documents and information anywhere and anytime with my current information storage setup.</p>	<p>I can access frequently used private documents and information anywhere and anytime with my current information storage setup. - I can access frequently used public documents and information anywhere and anytime with my current information storage setup.</p>	<p>I can access non-frequently used private documents and information anywhere and anytime with my current information storage setup. - I can access frequently used private documents and information anywhere and anytime with my current information storage setup.</p>
Z	-3.256 ^b	-3.510 ^b	-.125 ^c
Asymp. Sig. (2-tailed)	.001	.000	.900

- a. Wilcoxon Signed Ranks Test
- b. Based on positive ranks.
- c. Based on negative ranks.

*Nonparametric Tests: One Sample.
NPTESTS

/ONESAMPLE TEST (Fragmentation Formats Storing Troubling Expensive) CHISQ
UARE(EXPECTED=EQUAL) WILCOXON(TESTVALUE=3)
/MISSING SCOPE=ANALYSIS USERMISSING=EXCLUDE
/CRITERIA ALPHA=0.05 CILEVEL=95.

Nonparametric Tests

Notes

Output Created	13-AUG-2016 02:09:...	
Comments		
Input	Data	/Users/balekundargib/ Desktop/First submission/NEW_ANAL YSIS.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	72
Syntax	NPTESTS /ONESAMPLE TEST (Fragmentation Formats Storing Troubling Expensive) CHISQUARE (EXPECTED=EQUAL) WILCOXON (TESTVALUE=3) /MISSING SCOPE=ANALYSIS USERMISSING=EXCLUDE /CRITERIA ALPHA=0. 05 CILEVEL=95.	
Resources	Processor Time	00:00:00.09
	Elapsed Time	00:00:01.00

13.9 Related samples test : Retrieval

NPAR TESTS

/WILCOXON=NFPUIRetrieve NFPUIRetrieve NFPUIRetrieve FPIUIRetrieve FPIUIRe
trieve FPIRetrieve WITH

NFPIRetrieve FPIRetrieve FPIUIRetrieve NFPIRetrieve FPIRetrieve NFPIRet
rieve (PAIRED)

/MISSING ANALYSIS.

NPar Tests

Notes

Output Created		13-AUG-2016 03:11:...
Comments		
Input	Data	/Users/balekundargib/ Desktop/First submission/NEW_ANAL YSIS.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	72
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics for each test are based on all cases with valid data for the variable(s) used in that test.
Syntax		NPAR TESTS /WILCOXON=NFPUIRetri eve NFPUIRetrieve NFPUIRetrieve FPIUIRetrieve FPIUIRetrieve FPIRetrieve WITH NFPIRetrieve FPIRetrieve FPIUIRetrieve NFPIRetrieve FPIRetrieve NFPIRetrieve (PAIRED) /MISSING ANALYSIS.
Resources	Processor Time	00:00:00.00
	Elapsed Time	00:00:00.00
	Number of Cases Allowed^a	349525

a. Based on availability of workspace memory.

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
I can retrieve non-frequently used private documents and information very quickly with my current information storage setup. - I can retrieve non-frequently used public documents and information very quickly with my current information storage setup.	Negative Ranks	11 ^a	11.05	121.50
	Positive Ranks	9 ^b	9.83	88.50
	Ties	52 ^c		
	Total	72		
I can retrieve frequently used private documents and information very quickly with my current information storage setup. - I can retrieve non-frequently used public documents and information very quickly with my current information storage setup.	Negative Ranks	14 ^d	14.57	204.00
	Positive Ranks	11 ^e	11.00	121.00
	Ties	47 ^f		
	Total	72		
I can retrieve frequently used public documents and information very quickly with my current information storage setup. - I can retrieve non-frequently used public documents and information very quickly with my current information storage setup.	Negative Ranks	6 ^g	6.50	39.00
	Positive Ranks	9 ^h	9.00	81.00
	Ties	57 ⁱ		
	Total	72		

Ranks

		N	Mean Rank	Sum of Ranks
I can retrieve non-frequently used private documents and information very quickly with my current information storage setup. - I can retrieve frequently used public documents and information very quickly with my current information storage setup.	Negative Ranks	15 ^j	17.93	269.00
	Positive Ranks	13 ^k	10.54	137.00
	Ties	44 ^l		
	Total	72		
I can retrieve frequently used private documents and information very quickly with my current information storage setup. - I can retrieve frequently used public documents and information very quickly with my current information storage setup.	Negative Ranks	17 ^m	19.26	327.50
	Positive Ranks	13 ⁿ	10.58	137.50
	Ties	42 ^o		
	Total	72		
I can retrieve non-frequently used private documents and information very quickly with my current information storage setup. - I can retrieve frequently used private documents and information very quickly with my current information storage setup.	Negative Ranks	7 ^p	9.93	69.50
	Positive Ranks	11 ^q	9.23	101.50
	Ties	54 ^r		
	Total	72		

a. I can retrieve non-frequently used private documents and information very quickly with my current information storage setup. < I can retrieve non-frequently used public documents and information very quickly with my current information storage setup.

b. I can retrieve non-frequently used private documents and information very quickly with my

Test Statistics^a

	I can retrieve non-frequently used private documents and information very quickly with my current information storage setup. - I can retrieve non-frequently used public documents and information very quickly with my current information storage setup.	I can retrieve frequently used private documents and information very quickly with my current information storage setup. - I can retrieve non-frequently used public documents and information very quickly with my current information storage setup.	I can retrieve frequently used public documents and information very quickly with my current information storage setup. - I can retrieve non-frequently used public documents and information very quickly with my current information storage setup.	I can retrieve non-frequently used private documents and information very quickly with my current information storage setup. - I can retrieve frequently used public documents and information very quickly with my current information storage setup.	I can retrieve frequently used private documents and information very quickly with my current information storage setup. - I can retrieve frequently used public documents and information very quickly with my current information storage setup.	I can retrieve non-frequently used private documents and information very quickly with my current information storage setup. - I can retrieve frequently used private documents and information very quickly with my current information storage setup.
Z	-.637 ^b	-1.149 ^b	-1.268 ^c	-1.559 ^b	-2.002 ^b	-.723 ^c
Asymp. Sig. (2-tailed)	.524	.250	.205	.119	.045	.470

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

c. Based on negative ranks.

13.10 Related samples test :: Safety and security

NPAR TESTS

```

/WILCOXON=NFPUISafety NFPUISafety NFPUISafety FPIUISafety FPIUISafety FPI
Safety WITH NFPISafety
      FPISafety FPIUISafety NFPISafety FPISafety NFPISafety (PAIRED)
/MISSING ANALYSIS.

```

NPar Tests

Notes

Output Created	13-AUG-2016 03:27:...	
Comments		
Input	Data	/Users/balekundargib/Desktop/First submission/NEW_ANALYSIS.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	72
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics for each test are based on all cases with valid data for the variable(s) used in that test.
Syntax	NPAR TESTS /WILCOXON=NFPUISafety NFPUISafety NFPUISafety FPIUISafety FPIUISafety FPIUISafety FPISafety WITH NFPISafety FPISafety FPIUISafety NFPISafety FPISafety NFPISafety (PAIRED) /MISSING ANALYSIS.	
Resources	Processor Time	00:00:00.01
	Elapsed Time	00:00:00.00
	Number of Cases Allowed^a	349525

a. Based on availability of workspace memory.

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
I feel secure and safe when storing non-frequently used private documents and information with my current information storage setup. - I feel secure and safe when storing non-frequently used public documents and information with my current information storage setup.	Negative Ranks	13 ^a	11.38	148.00
	Positive Ranks	7 ^b	8.86	62.00
	Ties	52 ^c		
	Total	72		
I feel secure and safe when storing frequently used private documents and information with my current information storage setup. - I feel secure and safe when storing non-frequently used public documents and information with my current information storage setup.	Negative Ranks	16 ^d	11.50	184.00
	Positive Ranks	4 ^e	6.50	26.00
	Ties	52 ^f		
	Total	72		
I feel secure and safe when storing frequently used public documents and information with my current information storage setup. - I feel secure and safe when storing non-frequently used public documents and information with my current information storage setup.	Negative Ranks	5 ^g	6.60	33.00
	Positive Ranks	6 ^h	5.50	33.00
	Ties	61 ⁱ		
	Total	72		

Ranks

		N	Mean Rank	Sum of Ranks
I feel secure and safe when storing non-frequently used private documents and information with my current information storage setup. - I feel secure and safe when storing frequently used public documents and information with my current information storage setup.	Negative Ranks	14 ^j	13.21	185.00
	Positive Ranks	8 ^k	8.50	68.00
	Ties	50 ^l		
	Total	72		
I feel secure and safe when storing frequently used private documents and information with my current information storage setup. - I feel secure and safe when storing frequently used public documents and information with my current information storage setup.	Negative Ranks	19 ^m	13.55	257.50
	Positive Ranks	5 ⁿ	8.50	42.50
	Ties	48 ^o		
	Total	72		
I feel secure and safe when storing non-frequently used private documents and information with my current information storage setup. - I feel secure and safe when storing frequently used private documents and information with my current information storage setup.	Negative Ranks	3 ^p	9.50	28.50
	Positive Ranks	13 ^q	8.27	107.50
	Ties	56 ^r		
	Total	72		

a. I feel secure and safe when storing non-frequently used private documents and information with my current information storage setup. < I feel secure and safe when storing non-frequently used public documents and information with my current information storage setup.

b. I feel secure and safe when storing non-frequently used private documents and information with

Test Statistics^a

	I feel secure and safe when storing non-frequently used private documents and information with my current information storage setup. - I feel secure and safe when storing non-frequently used public documents and information with my current information storage setup.	I feel secure and safe when storing frequently used private documents and information with my current information storage setup. - I feel secure and safe when storing non-frequently used public documents and information with my current information storage setup.	I feel secure and safe when storing frequently used public documents and information with my current information storage setup. - I feel secure and safe when storing non-frequently used private documents and information with my current information storage setup.	I feel secure and safe when storing non-frequently used private documents and information with my current information storage setup. - I feel secure and safe when storing frequently used public documents and information with my current information storage setup.	I feel secure and safe when storing frequently used private documents and information with my current information storage setup. - I feel secure and safe when storing non-frequently used public documents and information with my current information storage setup.	I feel secure and safe when storing non-frequently used private documents and information with my current information storage setup. - I feel secure and safe when storing frequently used public documents and information with my current information storage setup.
Z	-1.676 ^b	-3.031 ^b	.000 ^c	-1.996 ^b	-3.190 ^b	-2.180 ^d
Asymp. Sig. (2-tailed)	.094	.002	1.000	.046	.001	.029

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

c. The sum of negative ranks equals the sum of positive ranks.

d. Based on negative ranks.

*Nonparametric Tests: One Sample.
NPTESTS

```

/ONESAMPLE TEST (NFPIAccess FPIAccess NFPUIAccess FPUIAccess Fragmentatio
n Storing Formats Troubling Multiple Expensive NFPIRetrieve FPIRetrieve NFP
UIRetrieve FPIUIRetrieve NFPISafety FPISafety NFPUISafety FPIUISafety) WILC
OXON(TESTVALUE=3)
/MISSING SCOPE=ANALYSIS USERMISSING=EXCLUDE
/CRITERIA ALPHA=0.05 CILEVEL=95.

```

Nonparametric Tests

Notes

Output Created	18-AUG-2016 22:12:...	
Comments		
Input	Data	/Users/balekundargib/Desktop/First submission/NEW_ANALYSIS.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	72
Syntax	NPTESTS /ONESAMPLE TEST (NFPIAccess FPIAccess NFPUIAccess FPUIAccess Fragmentation Storing Formats Troubling Multiple Expensive NFPIRetrieve FPIRetrieve NFPUIRetrieve FPIUIRetrieve NFPISafety FPISafety NFPUISafety FPIUISafety) WILCOXON (TESTVALUE=3) /MISSING SCOPE=ANALYSIS USERMISSING=EXCLUDE /CRITERIA ALPHA=0. 05 CILEVEL=95.	
Resources	Processor Time	00:00:00.42
	Elapsed Time	00:00:01.00

13.11 : All one sample tests

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of I can access non-frequently used private documents and information anywhere and anytime with my current information storage setup. equals 3.0.	One-Sample Wilcoxon Signed Rank Test	.428	Retain the null hypothesis.
2	The median of I can access frequently used private documents and information anywhere and anytime with my current information storage setup. equals 3.00.	One-Sample Wilcoxon Signed Rank Test	.326	Retain the null hypothesis.
3	The median of I can access non-frequently used public documents and information anywhere and anytime with my current information storage setup. equals 3.00.	One-Sample Wilcoxon Signed Rank Test	.248	Retain the null hypothesis.
4	The median of I can access frequently used public documents and information anywhere and anytime with my current information storage setup. equals 3.00.	One-Sample Wilcoxon Signed Rank Test	.001	Reject the null hypothesis.
5	The median of My data is fragmented/spread over too many different services and/or devices. equals 3.00.	One-Sample Wilcoxon Signed Rank Test	.042	Reject the null hypothesis.
6	The median of I am good at storing and retrieving my own information. equals 3.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
7	The median of I find that documents stored different formats are inaccessible by different devices. equals 3.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
8	The median of It would be troubling and/or cause me distress if my data suddenly became available to the public. equals 3.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
9	The median of I sometimes have multiple copies of the same information in different storage locations. equals 3.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
10	The median of Storing data is expensive equals 3.00.	One-Sample Wilcoxon Signed Rank Test	.556	Retain the null hypothesis.
11	The median of I can retrieve non-frequently used private documents and information very quickly with my current information storage setup. equals 3.00.	One-Sample Wilcoxon Signed Rank Test	.004	Reject the null hypothesis.
12	The median of I can retrieve frequently used private documents and information very quickly with my current information storage setup. equals 3.00.	One-Sample Wilcoxon Signed Rank Test	.058	Retain the null hypothesis.
13	The median of I can retrieve non-frequently used public documents and information very quickly with my current information storage setup. equals 3.00.	One-Sample Wilcoxon Signed Rank Test	.001	Reject the null hypothesis.
14	The median of I can retrieve frequently used public documents and information very quickly with my current information storage setup. equals 3.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
15	The median of I feel secure and safe when storing non-frequently used private documents and information with my current information storage setup. equals 3.00.	One-Sample Wilcoxon Signed Rank Test	.042	Reject the null hypothesis.
16	The median of I feel secure and safe when storing frequently used private documents and information with my current information storage setup. equals 3.00.	One-Sample Wilcoxon Signed Rank Test	.558	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

(continued)

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
17	The median of I feel secure and safe when storing non-frequently used public documents and information with my current information storage setup. equals 3.00.	One-Sample Wilcoxon Signed Rank Test	.001	Reject the null hypothesis.
18	The median of I feel secure and safe when storing frequently used public documents and information with my current information storage setup. equals 3.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

```

FREQUENCIES VARIABLES=NFPUIAccess FPIAccess NFPUIAccess FPUIAccess NFPIRetri
eve FPIRetrieve
      NFPUIRetrieve FPIUIRetrieve NFPISafety FPISafety NFPUISafety FPIUISafet
Y
  /NTILES=4
  /STATISTICS=MEDIAN
  /BARCHART FREQ
  /ORDER=ANALYSIS.

```

Frequencies

13.12 Frequency distribution of all questions

Notes

Output Created	18-AUG-2016 22:13:...	
Comments		
Input	Data	/Users/balekundargib/Desktop/First submission/NEW_ANALYSIS.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	72
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics are based on all cases with valid data.
Syntax	FREQUENCIES VARIABLES=NFPUIAccess FPIAccess NFPUIAccess FPIAccess NFPIRetrieve FPIRetrieve NFPUIRetrieve FPIUIRetrieve NFPISafety FPISafety NFPUISafety FPIUISafety /NTILES=4 /STATISTICS=MEDIAN /BARCHART FREQ /ORDER=ANALYSIS.	
Resources	Processor Time	00:00:02.04
	Elapsed Time	00:00:02.00

Statistics

		I can access non-frequently used private documents and information anywhere and anytime with my current information storage setup.	I can access frequently used private documents and information anywhere and anytime with my current information storage setup.	I can access non-frequently used public documents and information anywhere and anytime with my current information storage setup.	I can access frequently used public documents and information anywhere and anytime with my current information storage setup.	I can retrieve non-frequently used private documents and information very quickly with my current information storage setup.	I can retrieve frequently used private documents and information very quickly with my current information storage setup.	I can retrieve non-frequently used public documents and information very quickly with my current information storage setup.	I can retrieve frequently used public documents and information very quickly with my current information storage setup.	I feel secure and safe when storing non-frequently used private documents and information with my current information storage setup.	I feel secure and safe when storing frequently used private documents and information with my current information storage setup.	I feel secure and safe when storing non-frequently used public documents and information with my current information storage setup.	I feel secure and safe when storing frequently used public documents and information with my current information storage setup.
N	Valid	72	72	72	72	72	72	72	72	72	72	72	72
	Missing	0	0	0	0	0	0	0	0	0	0	0	0
Median		3.0000	3.0000	3.0000	4.0000	3.0000	3.0000	3.0000	3.5000	3.0000	3.0000	3.0000	3.0000
Percentiles	25	2.0000	2.0000	2.0000	3.0000	3.0000	3.0000	3.0000	3.0000	2.2500	2.0000	3.0000	3.0000
	50	3.0000	3.0000	3.0000	4.0000	3.0000	3.0000	3.0000	3.5000	3.0000	3.0000	3.0000	3.0000
	75	4.0000	4.0000	4.0000	4.0000	4.0000	4.0000	4.0000	4.0000	4.0000	4.0000	4.0000	4.0000

Frequency Table

I can access non-frequently used private documents and information anywhere and anytime with my current information storage setup.

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Strongly Disagree	2	2.8	2.8	2.8
Disagree	29	40.3	40.3	43.1
Neutral	19	26.4	26.4	69.4
Agree	18	25.0	25.0	94.4
Strongly Agree	4	5.6	5.6	100.0
Total	72	100.0	100.0	

I can access frequently used private documents and information anywhere and anytime with my current information storage setup.

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Strongly Disagree	6	8.3	8.3	8.3
Disagree	24	33.3	33.3	41.7
Neutral	19	26.4	26.4	68.1
Agree	19	26.4	26.4	94.4
Strongly Agree	4	5.6	5.6	100.0
Total	72	100.0	100.0	

I can access non-frequently used public documents and information anywhere and anytime with my current information storage setup.

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Strongly Disagree	1	1.4	1.4	1.4
Disagree	20	27.8	27.8	29.2
Neutral	26	36.1	36.1	65.3
Agree	19	26.4	26.4	91.7
Strongly Agree	6	8.3	8.3	100.0
Total	72	100.0	100.0	

I can access frequently used public documents and information anywhere and anytime with my current information storage setup.

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Strongly Disagree	1	1.4	1.4	1.4
Disagree	13	18.1	18.1	19.4
Neutral	20	27.8	27.8	47.2
Agree	31	43.1	43.1	90.3
Strongly Agree	7	9.7	9.7	100.0
Total	72	100.0	100.0	

I can retrieve non-frequently used private documents and information very quickly with my current information storage setup.

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Strongly Disagree	1	1.4	1.4	1.4
Disagree	12	16.7	16.7	18.1
Neutral	27	37.5	37.5	55.6
Agree	27	37.5	37.5	93.1
Strongly Agree	5	6.9	6.9	100.0
Total	72	100.0	100.0	

I can retrieve frequently used private documents and information very quickly with my current information storage setup.

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Strongly Disagree	4	5.6	5.6	5.6
Disagree	12	16.7	16.7	22.2
Neutral	24	33.3	33.3	55.6
Agree	27	37.5	37.5	93.1
Strongly Agree	5	6.9	6.9	100.0
Total	72	100.0	100.0	

I can retrieve non-frequently used public documents and information very quickly with my current information storage setup.

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Strongly Disagree	1	1.4	1.4	1.4
Disagree	11	15.3	15.3	16.7
Neutral	25	34.7	34.7	51.4
Agree	30	41.7	41.7	93.1
Strongly Agree	5	6.9	6.9	100.0
Total	72	100.0	100.0	

I can retrieve frequently used public documents and information very quickly with my current information storage setup.

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Disagree	8	11.1	11.1	11.1
Neutral	28	38.9	38.9	50.0
Agree	30	41.7	41.7	91.7
Strongly Agree	6	8.3	8.3	100.0
Total	72	100.0	100.0	

I feel secure and safe when storing non-frequently used private documents and information with my current information storage setup.

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Strongly Disagree	2	2.8	2.8	2.8
Disagree	16	22.2	22.2	25.0
Neutral	21	29.2	29.2	54.2
Agree	29	40.3	40.3	94.4
Strongly Agree	4	5.6	5.6	100.0
Total	72	100.0	100.0	

I feel secure and safe when storing frequently used private documents and information with my current information storage setup.

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Strongly Disagree	3	4.2	4.2	4.2
Disagree	21	29.2	29.2	33.3
Neutral	18	25.0	25.0	58.3
Agree	28	38.9	38.9	97.2
Strongly Agree	2	2.8	2.8	100.0
Total	72	100.0	100.0	

I feel secure and safe when storing non-frequently used public documents and information with my current information storage setup.

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Disagree	13	18.1	18.1	18.1
Neutral	26	36.1	36.1	54.2
Agree	24	33.3	33.3	87.5
Strongly Agree	9	12.5	12.5	100.0
Total	72	100.0	100.0	

I feel secure and safe when storing frequently used public documents and information with my current information storage setup.

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Disagree	12	16.7	16.7	16.7
Neutral	25	34.7	34.7	51.4
Agree	29	40.3	40.3	91.7
Strongly Agree	6	8.3	8.3	100.0
Total	72	100.0	100.0	

Bar Chart

Notes

Output Created	19-AUG-2016 01:28:...	
Comments		
Input	Data	/Users/balekundargib/Desktop/First submission/NEW_ANALYSIS.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	72
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics are based on all cases with valid data.
Syntax	FREQUENCIES VARIABLES=Fragmentation Storing Formats Multiple Troubling Expensive /NTILES=4 /STATISTICS=MEDIAN /BARCHART FREQ /ORDER=ANALYSIS.	
Resources	Processor Time	00:00:00.69
	Elapsed Time	00:00:01.00

Statistics

		My data is fragmented/spread over too many different services and/or devices.	I am good at storing and retrieving my own information.	I find that documents stored in different formats are inaccessible by different devices.	I sometimes have multiple copies of the same information in different storage locations.	It would be troubling and/or cause me distress if my data suddenly became available to the public.	Storing data is expensive
N	Valid	72	72	72	72	72	72
	Missing	0	0	0	0	0	0
Median		3.0000	4.0000	4.0000	4.0000	4.5000	3.0000
Percentiles	25	3.0000	3.0000	3.0000	3.0000	4.0000	2.2500
	50	3.0000	4.0000	4.0000	4.0000	4.5000	3.0000
	75	4.0000	4.0000	4.0000	4.7500	5.0000	4.0000

Frequency Table

My data is fragmented/spread over too many different services and/or devices.

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Strongly Disagree	3	4.2	4.2	4.2
Disagree	12	16.7	16.7	20.8
Neutral	26	36.1	36.1	56.9
Agree	27	37.5	37.5	94.4
Strongly Agree	4	5.6	5.6	100.0
Total	72	100.0	100.0	

I am good at storing and retrieving my own information.

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Disagree	11	15.3	15.3	15.3
Neutral	19	26.4	26.4	41.7
Agree	30	41.7	41.7	83.3
Strongly Agree	12	16.7	16.7	100.0
Total	72	100.0	100.0	

I find that documents stored different formats are inaccessible by different devices.

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Strongly Disagree	2	2.8	2.8	2.8
Disagree	8	11.1	11.1	13.9
Neutral	21	29.2	29.2	43.1
Agree	29	40.3	40.3	83.3
Strongly Agree	12	16.7	16.7	100.0
Total	72	100.0	100.0	

I sometimes have multiple copies of the same information in different storage locations.

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Strongly Disagree	1	1.4	1.4	1.4
Disagree	11	15.3	15.3	16.7
Neutral	12	16.7	16.7	33.3
Agree	30	41.7	41.7	75.0
Strongly Agree	18	25.0	25.0	100.0
Total	72	100.0	100.0	

It would be troubling and/or cause me distress if my data suddenly became available to the public.

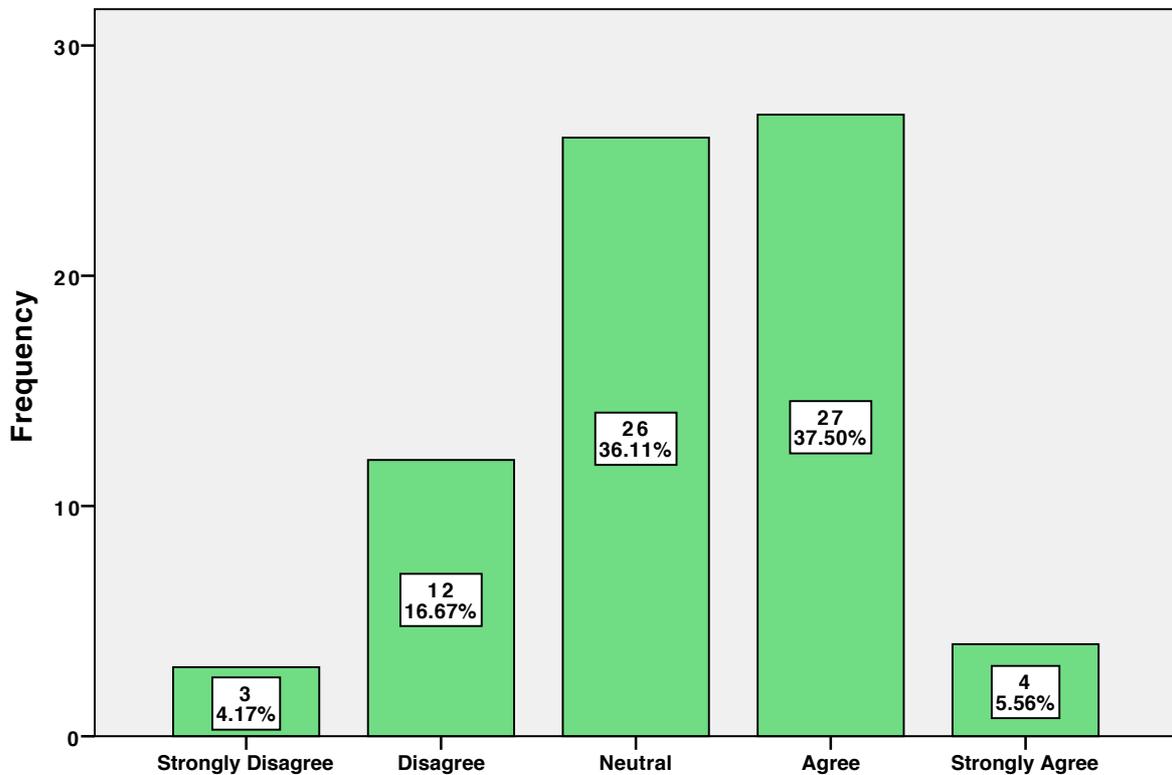
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Strongly Disagree	1	1.4	1.4	1.4
Disagree	3	4.2	4.2	5.6
Neutral	9	12.5	12.5	18.1
Agree	23	31.9	31.9	50.0
Strongly Agree	36	50.0	50.0	100.0
Total	72	100.0	100.0	

Storing data is expensive

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Strongly Disagree	1	1.4	1.4	1.4
Disagree	17	23.6	23.6	25.0
Neutral	32	44.4	44.4	69.4
Agree	21	29.2	29.2	98.6
Strongly Agree	1	1.4	1.4	100.0
Total	72	100.0	100.0	

Bar Chart

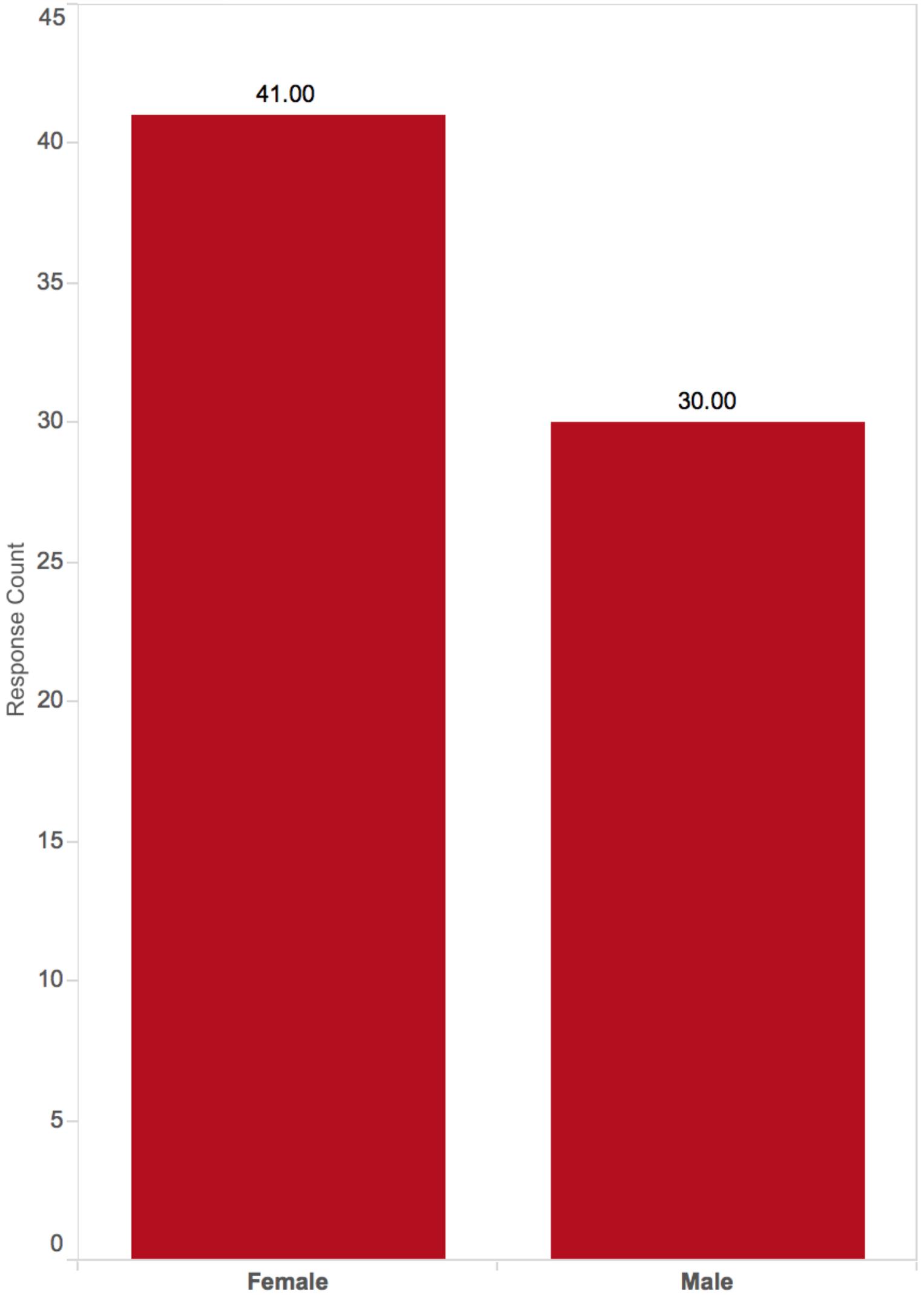
My data is fragmented/spread over too many different services and/or devices.



My data is fragmented/spread over too many different services and/or devices.

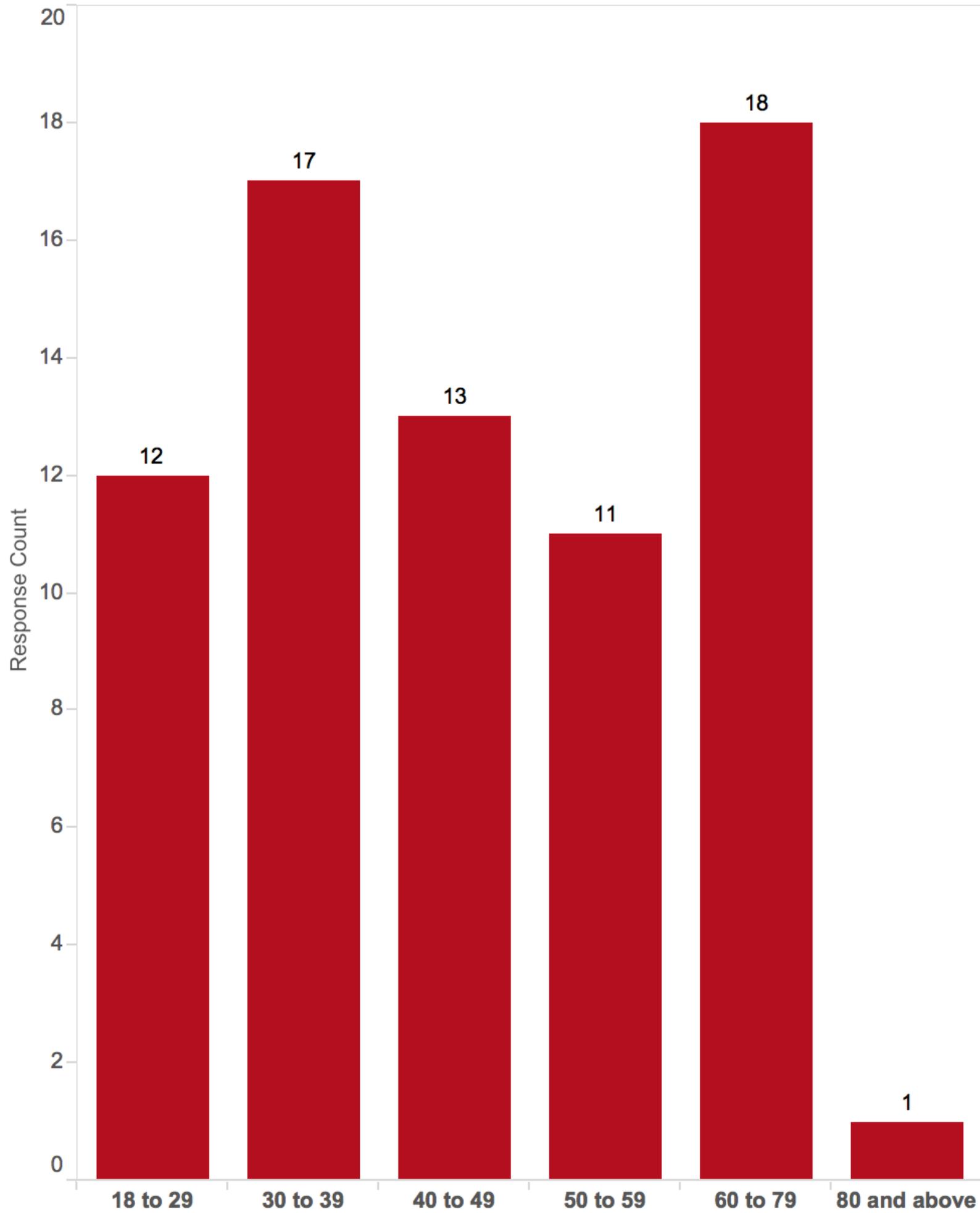
GENDER DEMOGRAPHY

13.13



13.14

AGE DEMOGRAPHY



INCOME DEMOGRAPHY

