


Spring 6-14-2024

The Institutional Challenges of a Quantified Self Study: An Attempt to Ascertain How Data Collected From a Mobile Device Can Be an Indicator of Personal Mental Health Over Time

Julian Lazaras
Portland State University

Follow this and additional works at: <https://pdxscholar.library.pdx.edu/honorsthesis>

 Part of the [Applied Behavior Analysis Commons](#), [Databases and Information Systems Commons](#), [Data Science Commons](#), [Experimental Analysis of Behavior Commons](#), [Health Psychology Commons](#), [Other Computer Sciences Commons](#), [Quantitative Psychology Commons](#), [Software Engineering Commons](#), and the [Theory and Philosophy Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Lazaras, Julian, "The Institutional Challenges of a Quantified Self Study: An Attempt to Ascertain How Data Collected From a Mobile Device Can Be an Indicator of Personal Mental Health Over Time" (2024). *University Honors Theses*. Paper 1485.
<https://doi.org/10.15760/honors.1517>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in University Honors Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

The Institutional Challenges of a Quantified Self Study

An attempt to ascertain how data collected from a mobile device can be an indicator of personal mental health over time.

By: Julian Lazaras.

Abstract:

The adoption of an application of new technology always comes with a bias, this is never more true for the case of human behavioral analytics within higher education. While movements such as the quantified self movement make strides to reinterpret the realm of data analytics, psychology, and computer science, there are inevitably limitations to the adoption and application of such approaches within the standard realm of research. Herein is presented a case where an effort to evaluate the prospect of use of mobile phone data as secondary indicators of personal mental health through the lens of data analysis was put at odds with institutional knowledge especially that of the institutional review board and expectation both for the disciplines of psychology, computer science, data science and that of software engineering.

The question of the application of data analytics is not uncommon, rather it seems to be of great concern as the advent of information technology systems continues to grow. However an analysis of the bias that is used in the application of such data, the intent behind its analysis, the means of its collection, the ethics of its intention all must be questions that are thoroughly considered within this discourse. The irony of this fact being that to truly understand the application of a tool such as this, it is imperative to view the wielder of such tools, what their motivations are, what their intentions are, and what effects they have, especially the unintended ones. Inevitably this pursuit is vastly added by the use of such tools as big data, or large scale, and small scale data analytics can provide, when a purely explanative or procedural effort to do so is likely to fail. In essence the application of data analysis must be viewed through a lens of leveraging itself to examine the intentions of those who deploy it, a uniquely self-referential or self dependent tool. Put another way, when classical psychology, that of psychoanalysis, or even individual behavioral analysis fails to examine the effects, and motivations of the use of potentially risky technologies, there is a lot of value in using behavioral trend analysis, and

behavioral deviation indicators to examine how individuals, and communities interact and are motivated. This idea is integral to the notion of the quantified self, a philosophy that believes self exploration can be vastly added by mathematical, and scientific correlative data analysis.

Background:

The role, responsibility, and limitations of higher education

It's well known that there are difficulties transitioning between secondary education and employment or post-secondary education. This is often a time in which individuals begin to foster the relationships that are likely to last into greater adulthood, indeed while limited in scope the term intimacy versus isolation as coined by american psychologist Erik Erikson. This time is one in which adolescents formally transition their attachment from that of a primary caregiver to that of a romantic partner, friends, or themselves. As such this period is a pivotal time for the development of adults, and shaping how people will react within the greater socio-cultural context in which they exist. By its nature this is a complex topic that present many challenges to its examination, within the period of time between high school graduation, and the full development into adulthood, around age 25 there is a lot of variance in trajectory that any particular individual may face, as such this research aims develop a model especially applicable to those within higher educational institutions, that may be applicable beyond.

Higher educational institutions present a fascinating lens through which to examine not only the developmental component of the transition between adolescence and adulthood, but also a best case for resources available during this time, and a framing of our cultures best attempts at self reflection, and exploration. Definition: higher educational institutes are any post-secondary educational institutes that provide as a core component of their operation a service that facilitates the advancement of an individual's knowledge, or understanding of any

particular area of study. However, given the pivotal role these institutions adopt by the age range of individuals who utilize them, they have taken on a far greater significance in recent decades especially. This significance principally being, as outlined above, the transition from adolescence into adulthood. This is facilitated through resources such as career centers, first year housing experiences, and group social activities in addition to the core educational requirements that are set out by organizations such as ABET (Accreditation Board for Engineering and Technology).

Unfortunately however, there are many limitations to such institutions, especially in cases where the institution becomes seen merely as a means of providing an accredited degree, in an isolated discipline. In these cases especially it becomes substantially more challenging to track and maintain a high quality of life, including with respect to mental health, as individuals become increasingly more isolated from a lack of facilitation, or desired participation within extra curricular activities, or integrated academic social functions.

Motivations:

By providing a means of identifying individuals whose behaviors have become increasingly isolationist, and by measuring progress made towards breaking that isolation within a standardized mathematically verifiable manner allows not only for tailored individualized assistance towards social reintegration, but also provides a means of addressing potentially systemic institutional problems, and making progress towards adjusting or correcting for such issues. Take for example a case in which a student group has disbanded or an individual was asked to leave a particular social group, this group being a primary source of such integration this student may be less inclined towards finding alternative social occupations, have difficulty completing coursework, and in the worst cases have to leave the institution. By using a quantified self perspective, and the development of an application which can track either

primary, or secondary social connectedness indicators, any individual can be identified as needing additional assistance and much more easily build a case for the additional aid to be supplied. This study aims to provide a working model of an application that uses secondary metrics to provide a quantifiable measure of social isolation through the use of mobile phone data. In so doing providing a theoretical framework for the identification, and markers of progress towards social and academic reintegration.

This application of data collection, and verifiable behavioral change presents a potentially groundbreaking reapplication of large, and small scale data analytics. In so doing the goal of providing such utility also has the added benefit of introducing a more easily digestible application of data analytics to populations that many not otherwise have had the concept, ability, or desire to leverage such tools, and thus not have a vested interest in developing an understanding of the tools available to them.

Prior Research:

This study was inspired by the student life research project. A research project out of Dartmouth university the aim of the student life research project was to use secondary mobile phone indicators such as location, text message frequency, call frequency, and other metrics to identify students under stress, and correlate that to times of high stress throughout the course of a term. The biggest limitations of this study being 1) the lack of scalability, and 2) the lack of reproducibility by not providing an open source implementation. This study was conducted on only 60 participants over the course of a single term. (why does my study matter?, am I advocating a repeat? Or will my thesis simply be a re-evaluation of existing work?)

The issue of lack of reproducibility, and the issue of modern research practices:

It's well accepted in disciplines such as physics, chemistry, and philosophy the need for a study to be easily reproduced, However disciplines such as sociology, psychology, and computer science (discipline leveraged in this research) often lack some of the appreciation, and easy of replicability, within psychology, and sociology this is sometimes inevitable, take for example the case of case studies, however this is not always the case. Nevertheless the motivation seems to be within research communities, even those of chemistry, biology, and medicine to produce "original work" that needs to be "protected" thereby discouraging the easy reapplication of the study on a different sample population. This presents a distinct challenge when attempting to verify the authenticity of a particular study, especially given the stringent requirements placed on researchers that deal with human subjects on the part of uneducated or ill informed institutional review boards. By providing a tool, a framework, and sample data, this study aims to set a precedent of transparency, verifiability, and ease of reproduction.

Timeline:

Software Development Period:

Example simple interface for data collection.

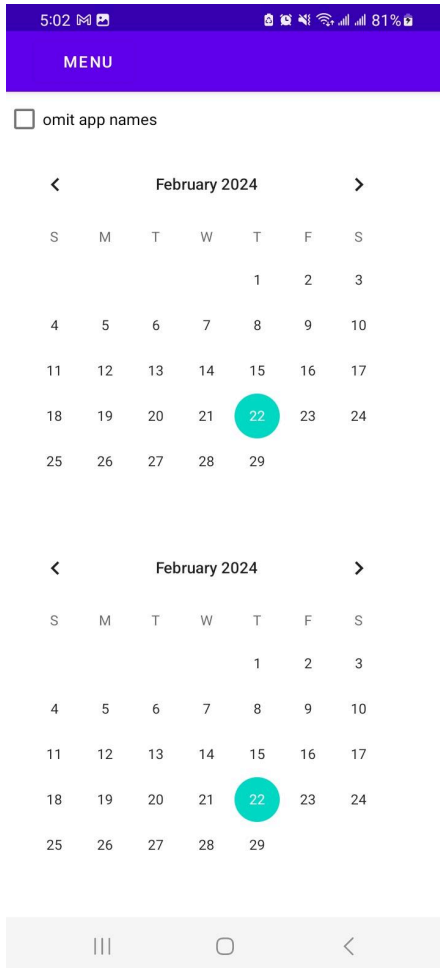


Figure 1: selecting date ranges for app usage events



Figure 2: selecting to export queryable metrics

Use of Permissions systems and user choice:

The intention behind this project being the ability of individuals to participate in their own data analysis the principal goal of the software development was not only to supply users with a means of exporting their data for their own analysis, but also selecting what data they want exported and shared with the servers used to collect data in this project. Before requesting permissions from users, the nature of the data being collected, and an explanation of how it's

being collected and analyzed should be prepared in order to provide users the ability to have fully informed consent to any research performed.

Backend:

The backend implemented in this project used rust + rocket though because it never reached a minimum viable product, it has not been published as part of this work. The database used was PostgreSQL as it is a lightweight, easy to use open source platform with relatively high availability.

What should be done:

The largest component of the process of completing this study, the development of the tool used to collect data from participants. This step must include determining the metrics used within the study to observe behavior of participants over time, to do so select from data available through apis on the target platform. In this case, this involved determining what data could relatively easily be collected from android devices, it was determined the easiest data to collect was, text message, calendar events, call logs, and app usage over time. While other more selective metrics such as activity tracking were available, the practicality of deploying these methods of data collection were much more challenging as documentation of the process, and code examples were few and far between.

The largest component of the process of completing this study, the development of the tool used to collect data from participants. This step must include determining the metrics used within the study to observe behavior of participants over time, to do so select from data available through apis on the target platform. In this case, this involved determining what data could relatively easily be collected from android devices, it was determined the easiest data to collect was, text message, calendar events, call logs, and app usage over time. While other more selective metrics such as activity tracking were available, the practicality of deploying these methods of data collection were much more challenging as documentation of the process, and code examples were few and far between.

Institutional review board application process, The problem with IRBs:

Likely the most bureaucratic, and potentially time consuming process of this type of research is the institutional review board (IRB) exemption process which required common rule training, and a demonstration of lack of harm to participants. This was also the most frustrating as the administrators of this process had difficulty understanding the nature of the data collected for this study. It is advised that this process of applying for approval for human subjects research be performed well in advance of any recruitment or even in advance of the release of any software as this process tends to take a long time. To mitigate these delays, it is important to consider and present clearly what the research focus is, and if the common rule applies, for example not only did this research not qualify as research under title 45 of code of federal regulations (45 CFR 46) as this research is not generalizable, but it also did not generate private identifiable information. To ensure this data scrubbing practices including removing any and all demographic information, and identifiers was used to ensure participant anonymity. By clearly presenting this to the IRB an exemption may be applicable, or the research may qualify for expedited review.

The common rule or 45 CFR 46 outlines the specifications for what is, and is not exempt from institutional review board approval, specifically in order for something to be classified as human subjects research it must meet both the qualifications of research, as well as meet the qualifications of human subjects who are identifiable. In order for something to be classified as research it must be “a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge.” (1)

Private sector approach:

Unlike research at a research institute, especially a public institute, the private sector, corporations have a lot more leeway when it comes to collecting information about people. While on the one hand that ease of information collection for analysis does make research a lot easier, it does present the distinct ethical issue of its use. Unlike the public sector, rarely is it the

case that companies perform analytics for the purposes of pure knowledge, rather it is performed for profit, which is very much in opposition to the quantified self movement, and the purpose behind this research.

Recruitment Period:

During the recruitment period, the most important step is the selection of incentives, why have people participate. In the first version of this research, with limited funds available the primary point of sale was that of academic curiosity. This approach allowed for a sense of personal investment on the part of the participants by involving them in the academic theory of the research. This also worked well with the motivation for performing the research, principally the expansion of the concept of data analytics for uses that are more heavily oriented towards participant involvement and cooperation within a traditionally highly stratified approach to market research.

Data Collection:

With all software development, and recruitment complete the process of data collection is fairly easy, and can be done with relatively little to no human interaction. For example the participant would be able to create an account within the app, answer a questionnaire to verify that they are within the target demographic, sign a release form, and then select what metrics they are willing to have submitted to the web backend.

The most challenging component of this phase is to address any bugs, or security concerns that weren't discovered during the initial development phase. In order to mitigate this, having a beta testing or trial data collection phase with a small number of participants who are more communicative or know the researcher would help to find potential bugs before release to unknown participants. In addition it's recommended to have a secondary means of answering questions, or submitting bug reports, such as could be provided through email, or github. Once initial trial data has been collected the experiment can be extended to a wide range of

participants, and ask more detailed questions related to mood, quality of life, sense of self esteem, and general metrics the research is evaluating.

Test condition:

Owing to the limited data availability the analysis will be set up such that it only compares across time to prior data collected from the individual, rather than comparing to a group. In addition this provides the benefit of detecting change in behavior, rather than attempting to assume there is an “expected” normal usage of text messages, calls, or apps between individuals, expanding this study to sleep and activity tracking does however present an opportunity to compare across individuals, however that is unfortunately out of scope for this research at present, but would likely be required for a reasonable deployment of such a tool as outlined here.

Application Event classifications:

The primary app usage events are app start, app stop, keyboard shown, keyboard hidden, screen interactive, screen non interactive, user interaction, foreground start, foreground stop, and app paused.

Text and Call Frequency:

For text messages sent and received, as well as incoming and outgoing calls, the primary point of analysis will be how often an event is registered in a weekly time frame, which will be a discrete time domain function in the same manner as application events.

Analysis:

Each event, whether it is application event, text message, or call, will be summed per weeks worth of data, these individual weeks will then be compared using median and quartiles to determine statistically significant outliers for each app, and for each of text message conversation, and call recipient if enough data is present to see significant differences between text conversations and call recipients respectively. Median and quartile fence methods are used

over that of mean and standard deviation as this framework aims to operate, and identify a statistically significant left skew from periods of inactivity.

Sample Equation:

mod:

The modulo operator takes the remainder after dividing the first number by the second.

Q1,3

first, and third quartiles representing points that are each half of the median of observed data.

$$Q1(X) = X\left[\frac{n + (n \bmod 4)}{4}\right]$$

$$Q3(x) = x\left[3\left(\frac{[n + (n \bmod 4)]}{4}\right)\right]$$

$$IQR = Q3 - Q1$$

$$v(x) < Q1 - 1.5IQR$$

Data Safety Practices:

For the purposes of user safety, Android OS defines a permissions system that the user must grant before any data can be collected from the participants (excluding IP and thus coarse location, which won't be used in this study). The app provided to collect data for this project uses as a primary feature, the ability to export data to json files, and save to external storage of the phone (accessible by apps such as google drive), as such data submission is completely optimal. In addition any extension to the underlying code is provided free of charge on condition that the usage of data collected from this application is disclosed to the user prior to any automated collection of said data.

Conclusion:

A replication and expansion of this study could substantially change how big data is recognized, and appreciated, providing a unique framing for the development of new therapeutic practices, and improved treatment of conditions such as anxiety, and depression. This study also provides

a model by which an increased investment in data analytics can be made more approachable, and more interesting to those who otherwise would have little to no need or desire to investigate the flow, and interpretation of data.

Works Cited:

Swan M. The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery. *Big Data*. 2013 June;1(2):85-99. doi: 10.1089/big.2012.0002. PMID: 27442063.

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., & Campbell, A. (2014). StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 3–14.
<https://doi.org/10.1145/2632048.2632054>

[Estrin 14a] [“small data, where n = me” Links to an external site.](#), Deborah Estrin, CACM, April 2014.

Title 45 of the code of federal regulations (CFR) section 46 “revised common rule”, Department of Health and Human Services (HHS) United State Federal Government. Office for Human Research Protections. June 18th 2018.
<https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html>