Multivariate Time Series Clustering for Mobile Apps Data

Thevaa Chandereng
Department of Biostatistics
Columbia University
New York, NY 10032
thevaasiinen@gmail.com

Ying Kuen Cheung
Department of Biostatistics
Columbia University
New York, NY 10032
yc632@columbia.edu

Abstract

Clustering can generate intuitive and interpretable groups when it is tailored to the unique characteristics of the data. Mobile app usage data are typically highly sparse. When identifying features of app usage patterns, it is important to characterize the timing of usage as well as amount of usage. As such, direct application of traditional clustering methods, such as K-means clustering and hierarchical clustering may fail to incorporate important temporal features. We propose a multivariate quantile coarsening algorithm (mQCA) to analyze time series data from multiple apps per user. Briefly, a univariate QCA performs quantile transformation to time series data so that the time series is represented by quantiles of app activity time. An mQCA applies univariate QCA separately to each individual app, and links the multiple apps by anchoring the multivariate time series at the same time origin. While fitting separate univariate QCA reduces dimensionality and sparsity of the data, defining appropriate contrasts of quantiles further untangles the anchoring of the multivariate time series thus allowing for interpretation of individual app patterns. We can then apply traditional clustering methods such as K-means after applying mQCA to the raw data. We demonstrate the proposed method to identify app usage patterns of a suite of mental health apps in 14527 users. Our method compares favorably to direct application of K-means to the app usage amount, it produce better association with user characteristics.

1 Introduction

The advancement in technology, including the exponential growth in number of smartphones have allowed the use of mobile apps for health and well-being promotion. Mobile phones are pervasive as 73% of teens and 85% of young adults own a smartphone [12]. As a result, a large number of apps are developed for mobile users with different functionality. For example, as of the end of April 2021, there are about 1.96 and 2.87 million apps available at the Apple App and Google Play stores respectively, and these numbers are still growing dramatically. However, retaining user engagement on these apps are challenging. The 30-day retention rate for using health and fitness apps is only 47% with a mean usage of 2.7 times a week [8]. Mobile apps provide a rich source of data for understanding both user preference and behavior.

Time series data are collected extensively in every discipline to study the complexity and dynamic of a group of entities. Temporal features are important across every domain (for example, finance, sales, and biology) because it reveals vital information. For instance, tracking the app usage data over time helps identify user patterns. Similarity in user patterns can correspond to similarity in both preference and behavior. Thus, this can aid the enhancement of the app, distribution of ads in the app, or even personalization of the app. Clustering algorithms group data with similar characteristics together. The notion of "similar" is subjective and domain-specific. In time series datasets, an observation

Preprint. Under review.

is highly dependent on the past observations, thus the similarity needs to account for the temporal structure. Specifically, for management of app usage, it is critical to differentiate a user who engages with the apps consistently from another user who use apps intensively early on with rapid drop in usage, even both users may have the same aggregate usage over a period of time.

Many time series clustering algorithms have been introduced to group similar entities together. K-means clustering has been widely because it is relatively simple to implement, scales to large data sets, guarantees convergence, and generalizes to clusters of different shapes and sizes [11]. The other clustering approaches are usually either hierarchical or model-based. In a model-based clustering method, the data is assumed to be generated by a model and the method tries to recover the original model from the data [3]. On the other hand, the hierarchical method builds a hierarchy of clusters.

The EM algorithm is a commonly used algorithm for model-based clustering. This is an iterative method to find the maximum likelihood or maximum a posteriori estimates of parameters. There are many clustering methods developed using the EM algorithm with different assumptions [6, 19, 14]. Dahl proposes a clustering algorithm using Markov Chain Monte Carlo where entities with similar Dirichlet process mixture components are grouped together and the model is fit [7]. McDowell et al. introduce a Dirichlet process Gaussian process mixture model as the posterior distribution and a Gaussian process as the prior to cluster time series data [18].

Hierarchical clustering methods, such as clustering with correlation or transformed Euclidean distance for similarity, were a common choice before the proliferation of time series-specific algorithms [9] and continue to be widely used for temporal data [18, 2]. It is still widely used due to the simplicity of the method. The Lag-Penalized Weighted Correlation (LPWC) approach to fully take into account the lagged temporal profiles between entities by assigning a Gaussian kernel penalty score to reduce the chance of higher weighted correlations for such entities [3]. Short time series (STS) method computes the similarity matrix by comparing the rate of change in temporal profiles between neighboring timepoints [21]. Vilar et al. use forecasting density adopted from auto-regressive (AR) models to compute the similarity between time series [25]. Dynamic Time Warping (DTW) computes the optimal (least cumulative distance) alignment between points of two time series [10, 1]. DTW has been applied to temporal sequences of video and audio data.

Despite the abundance of time series clustering algorithms, clustering time series data for mobile apps is not straightforward. The data is usually high-dimensional and highly sparse. Traditional algorithms, such as K-means and hierarchical clustering and most model-based clustering algorithms are not suitable due to the nature of the data. The most effective way to cluster both sparse and high-dimensional is to perform data reduction while preserving key information before applying a clustering algorithm. There are a small number of methods proposed for reducing the volume of time series data [15, 5, 22]. For instance, Jang et al. proposed a method by constructing multiple sets of binned data with varying size and starting position, and then merging the clustering results from the binned data [15]. The quantile coarsening algorithm (QCA) performs quantile transformation to a time series [5]. However, these above mentioned methods are applicable for univariate time series data.

In this paper, we propose a multivariate quantile coarsening algorithm (mQCA) to analyze time series data from multiple apps per user. In section 2, we describe the mQCA in detail and the selection of input features in the mQCA for data reduction. Also in section 2, we will demonstrate how to pick the optimal number of clusters (k) and evaluate the clustering results and comparing them with the other methods. In the next section, we will describe the IntelliCare data. In section 4, we will apply the proposed method to the IntelliCare data and discuss the results. This article is ended with a discussion in Section 5. Additional details (including supplementary figures and tables) are included as separate supplementary material.

2 Methods

2.1 mQCA

Let $Y_{ij}(t)$ be the non-negative count at location t of user i for app j on the continuous time domain $t \in [0, T]$. The cumulative count up to location t of subject i for app j , $S_{ij}(t)$ is defined as following: $S_{ij}(t) = \int_0^t Y_{ij}(s)ds$ and $A_{ij} = S_{ij}(T)$ indicates the total count of subject i for app j. Since $Y_{ij}(t) \geq 0$, the function $S_{ij}(t)/A_i$ can be seen as a distribution function of a random count location.

Let

$$T_{ij}(p) = \inf\{t : S_{ij}(t)/A_{ij} \geq p\} = \inf\{t : \int_0^t Y_{ij}(r)dr \geq pA_{ij}\}$$

denote the location where $100p$ percent of the total count has been achieved. This can also be referred as the pth quantile of count location for this individual. Note that we assume that the domain of $T_{ij}(p)$ is contained in the finite interval [0, T].

We propose here to represent the entire function $Y_{ij}(t)$ by quantiles $T_{ij}(p)$ for a pre-specified set of percents (p), along with the total count A_{ij}. This is related to the concept of coarse data extensively discussed in [13]. The general idea of quantile coarsening is to represent a time series $Y_{ij}(t)$ using multiple quantiles $T_{ij}(p_j)$ for a prespecified set of $0 = p_0 < p_1 < p_2 < ... < p_k < p_{k+1} = 1$, together with the total counts $S_{ij}(T)$. In this manner, the volume of data can be reduced without losing important about the important features of the function. The Kth order, quantile coarsening function (QCF) of $Y_{ij}(t)$ is defined as

$$C_K Y_{ij}(t) = A_{ij} * \frac{p_{k+1} - p_k}{T_{ij}(p_{k+1}) - T_{ij}(p_k)},$$

for $T_{ij}(p_k|x) \leq t < T_{ij}(p_{k+1}|x)$. In a special scenario, where p_k's are evenly spaced, the formula reduced down to

$$C_K Y_{ij}(t) = \frac{A_{ij}}{(K+1) * (T_{ij}(p_{k+1}) - T_{ij}(p_k))},$$

for $T_{ij}(p_k|x) \leq t < T_{ij}(p_{k+1}|x)$, where $p_k = k/(K+1)$ for $k = 0, .., K+1$. Note that $T_{ij}(0) = 0$ and $T_{ij}(1) = T$ and $T(p)$ is monotone increasing, thus the QCF is always well-defined.

It can be shown that the QCF $C_K Y_{ij}(t)$ is a step function with the following properties:

$$\int_0^T C_K Y_{ij}(t) = S_{ij}(T) = A_{ij},$$

and

$$\inf\{t : \int_0^T C_K Y_{ij}(r)dr \geq p_k A_{ij}\} = T_{ij}(p_k).$$

That is, applying quantile transformation to $C_K Y_{ij}(t)$ on the grid p_k will yield identical results to quantile transformation of $Y_{ij}(t)$. In other words, inference about p_kth quantile is invariant under the Kth order quantile-coarsening mechanism. In practice, instead of a continuous function, we observe a discrete-time signal (or time series) Y_{ijt} at $t = 1, \ldots, T$. In these scenarios, the integrals above shall be replaced with summations.

2.2 Cluster analysis

We performed cluster analysis using K-means clustering with $C_K Y(t)$ as the input parameter for each individual app (mQCA kmeans) [11]. The time series data for app j, $Y_j(t)$, was represented by 5 different features, namely $T_j(0.05)$, $T_j(0.5)$, $T_j(0.75) - T_j(0.25)$, $T_j(0.90) - T_j(0.10)$, and $S_j(T)$. $T_j(0.05)$ provides a close proximation to when the mobile app j was downloaded. $T_j(0.90) - T_j(0.10)$ and $T_j(0.75) - T_j(0.25)$ yield the time differences between the 90% and 10% of the total usage and the 75% and 25% of the total usage for app j respectively. Both $T_j(0.90) - T_j(0.10)$ and $T_j(0.75) - T_j(0.25)$ provide proxies of user retention and how consistently a user engages with the app. $S_j(T)$ represents the total number of times the app j was used. We compare mQCA kmeans with Euclidean distance with hierarchical clustering (heuc) and K-means clustering (kmeans) applied to the app usage data without data reduction. These algorithms include some of the most widely used general clustering approaches.

2.3 Cluster evaluation

One of the biggest challenges for clustering data is choosing the number of clusters, which can be addressed with the silhouette method [24]. The silhouette value assesses how similar an object is to its own cluster compared to other clusters. We select the number of clusters that maximizes the average silhouette width.

In real dataset true clusters are usually unknown, thus evaluation of clustering methods is difficult. The Rand index is used as a metric to compare two clustering results [23]. The adjusted Rand index (ARI) is a corrected-for-chance version of the Rand index which is a more appropriate metric to compare clustering results [23]. The ARI is 1 for a perfect clustering that matches the true cluster labels. On the other hand, a score close to 0 indicates a poor clustering.

One way to evaluate time series clustering algorithms without ground truth labels is by assessing how important the temporal information is to the clustering results. We obtain clusters using the original data and then permute the data by randomly reordering the time (Day) (the app usage observations do not change). The permutations destroy the true temporal dependencies in the data. If a clustering algorithm does not use the temporal information, the ARI score when comparing its clusters on the original and permuted data will be close to 1, which is undesirable. In the IntelliCare data, we repeat the timepoint permutation 100 times for each clustering algorithm and assess the distribution of ARI scores. The results are discussed in the Supplement.

3 IntelliCare Data

The IntelliCare platform was designed by multidisciplinary team using the Behavioral Intervention Technology model [20] to improve symptoms of depression and anxiety. The platform contains multiple apps (13) rather than single app with the goal of being simple and brief. As part of IntelliCare, the Hub is a central app that navigates a user's experience with the other IntelliCare apps [20, 17]. The Hub makes weekly recommendation for new apps to be installed and explored, it also manages messages and notifications from the other apps. Besides the Hub, the IntelliCare platform consists of 12 other apps (Aspire, Day to Day, Daily Feats, Worry Knot, Social Force, My Mantra, Thought Challenger, iCope, Purple Chill, MoveMe, Slumber Time, and Boost Me) [4]. The 12 apps are divided into five different functionalities (thinking, calming, Checklists, activity, and other). The five functionalities can be grouped as follows: "Thinking" - Thought Challenger, MyMantra, Day to Day, and iCope; "Calming" - Purple Chill and Slumber Time; "Checklists" - Aspire and Daily Feats; "Activity" - Boost Me and MoveMe; and "Other" - Me Locate, Social Force, and Worry Knot. Previous studies have shown that the 12 app usages can be clustered into the five functionality [16].

The IntelliCare apps were made freely available to the public and were placed in stages on the Google Play Store starting September 22, 2014 [4], and improved upon based on observed usage patterns and user feedback. The users were presented with a user acknowledgment agreement that notifies users that the app usage information would be stored and analyzed for quality assurance purposes upon the first app was downloaded. This study included all users who downloaded their first IntelliCare app(s) during the period between April 1, 2015 and April 30, 2017, with the exception that Hub users who did not download Hub as their first app were excluded. The analyses included all app usage data up to July 31, 2017, so that we had at least 12-week of data for each user. The rationale behind this exclusion is due to users who had downloaded an app other than the Hub and then decided to download the Hub were arguably motivated and inclined to continue exploring the IntelliCare platform which introduces selection bias [4]. Second, by including the time and usage occurring before the Hub download would artificially inflate the engagement duration of these Hub users.

A detailed description of the participants and study details has been published elsewhere [20]. We identified records of a total of 14738 users. Out of the 14738 users, we excluded 211 users who downloaded a non-Hub app as their first app. All the app usage for 14,527 users were tracked for at least 84 days.

4 Results

In our analysis, we combined the app usage data into 6 distinct groups (the Hub app and 5 different functionality in Section 3). Each group contains 14,527 users and number of app usage for 84 days. Before clustering each group, we removed non-group users and labeled them as cluster 0. For example, if a user never downloaded both the Purple Chill and Slumber Time, we labeled them as cluster 0 for the calming group. For each group, we used mQCA to transform the data and cluster them using K-means (Table 1 and Figure 1). The cluster sizes were selected based on the maximum average silhouette width (Figure S3). Table 1 shows the cluster sizes for 6 different app categories for all 3 algorithms.

In all six groups, the app usage decrease over time (Figure 1). For the Hub app, the clusters had a similar temporal shape but they were divided by number of usages. In all the clusters for the Hub app, the number of usages decline as time progresses. In the thinking group, clusters 2, 3, and 4 have similar patterns, but cluster 3 is separated from clusters 2 and 4 by the amount of usage. In the calming, checklists, activity, and other groups, there are 3 distinct clusters with similar temporal shapes across the groups. In the 3 distinct clusters, one of the cluster has a low app usage between day 1 and day 20 before a spike in app usage and the app usage decreases after day 40. The other two clusters begin with a spike in usage and then drops after the first few days, with the distinction being no usage or very little usage after the first few days. The spike at day 84 in both calming and other groups is discussed in the Supplementary document.

Methods	mQCA kmeans					kmeans			heuc			non-user
Clusters	1	2	3	4	5	1	2	3	1	2	3	0
Hub	63	240	3755	1294	-	125	4540	687	5338	10	4	9175
Thinking	97	5253	328	1212	-	131	495	6264	6706	153	31	7637
Calming	161	2252	570	-	-	2325	542	116	2972	3	8	11544
Checklists	550	2471	69	154	-	134	3109	1	3242	1	1	11283
Activity	50	1	305	1790	165	1	139	2171	2308	2	1	12216
Other	3069	613	163	-	-	3699	1	145	3842	1	2	10682

Table 1: Cluster sizes for 6 different app categories for all three algorithms. Note: Cluster 0 represent the users who did not download all the apps in the specific category. Based on the average silhouette width, the optimal number of clusters for both kmeans and heuc are 3 for all app categories. The optimal number of clusters for mQCA kmeans is 3 (Calming and Other), 4 (Hub, Thinking, and Checklists), and 5(Activity).

We also clustered the app usage data using both kmeans and heuc (Figures S1 and S2). Hierarchical clustering with Eucledian distance is not suitable for mobile apps data due to the large number of singleton and small cluster sizes (less than 10 users) (Table 1). In all groups, hierarchical clustering fails to capture all the different temporal shapes (Figure S2). K-means clustering on the app data provides a similar clustering pattern to the mQCA kmeans for the Hub app (Figure S1). For the other 5 groups, kmeans fails to divide groups that have a high spike in usage in the first few days and no usage after the first few days and a high spike in usage in the first few days and low usage after the first few days. Compared to heuc, kmeans performs better on the app usage data. Using kmeans, there are only 3 singleton and no small cluster sizes out of the 18 clusters.

Since, kmeans clustering performs well on the app usage compared. We used patient characteristics to compare both kmeans and mQCA kmeans. For the hub app, both kmeans and mQCA kmeans performs well as illustrated in Figures S1 and 1. Tables S2 and S3 show similar pattern. The active users (high usage) group in the Hub app are dominantly older, White, female, non-Hispanic, and highly educated. Using mQCA kmeans to cluster, in the thinking, calming, and checklists app groups, the active users tend to be older, White, female, and highly educated (Tables S4, S6, and S8). The same results were found using kmeans, except we found no difference in age among all the clustering groups and no difference in education level for the checklists group (Tables S5, S7 and S9). For the activity app group, using mQCA kmeans, the active users tend to be older, White, and highly educated (Table S10). However, using kmeans, there was no difference in age among all the clusters but the active users tend to be White and highly educated (Table S11). Even though, we reported the patient characteristics for the other group using both kmeans and mQCA kmeans, it is had to compare the patient characteristics due to the different functionalities of the apps (Table S12 and S13). Based on both the patient characteristics and cluster figures (Figures 1 and S1), mQCA kmeans outperforms kmeans for the IntelliCare data.

5 Discussion

mQCA applies univariate QCA separately to each individual group of apps, and links the multiple apps by anchoring the multivariate time series at the same time origin. mQCA is designed to reduce dimensionality and sparsity without compromising vital information about the data. The feature selection for mQCA is subjective. Future work needs to address the optimal feature selection for

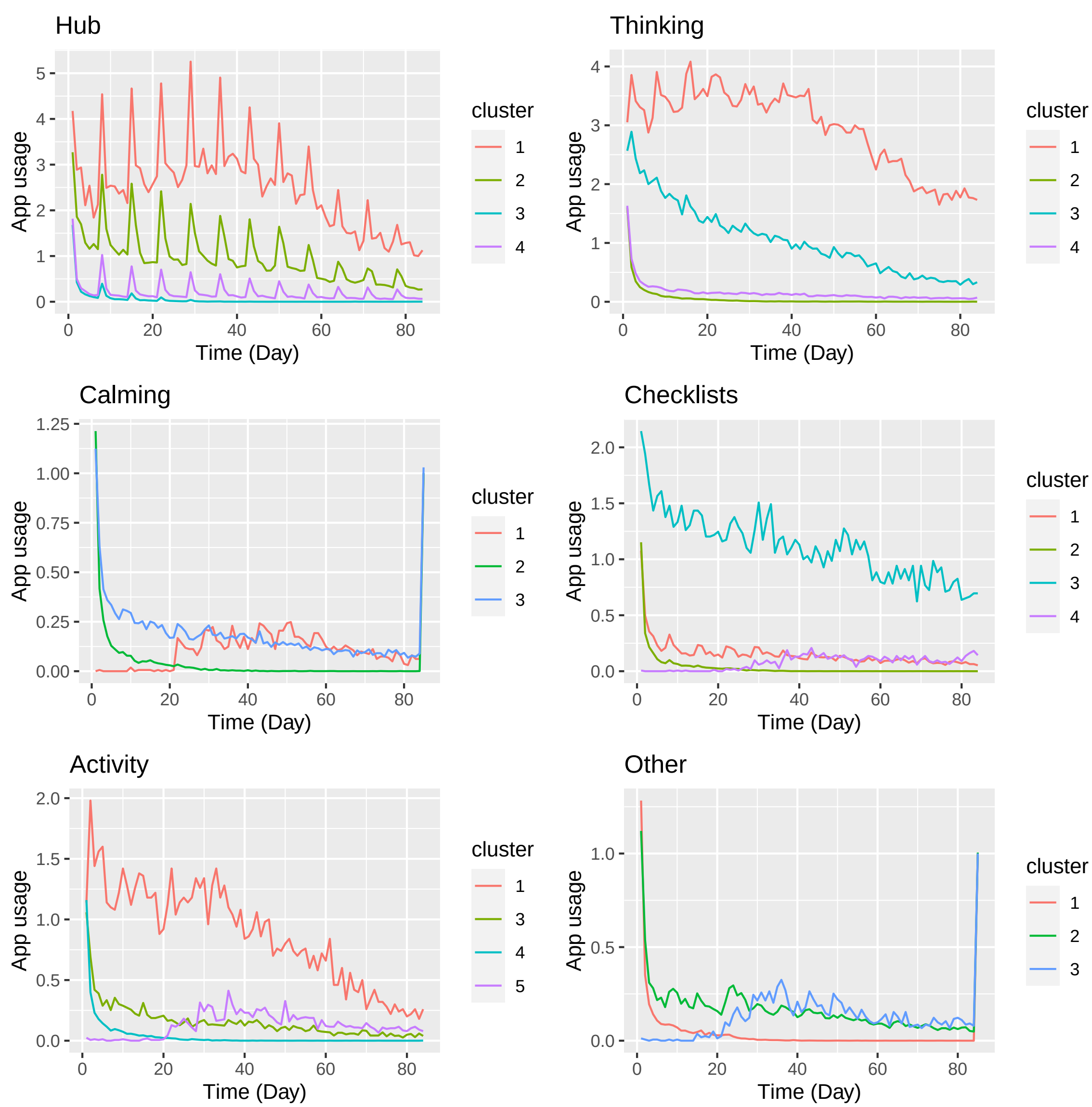

Figure 1: Mean app usage over time (Day) of all clusters for all the 6 different app categories (hub, thinking, calming, checklists, and other) using QCA kmeans clustering. Cluster size of less than 10 is removed.

dimensional reduction. Compared to the traditional clustering methods, mQCA performs better in the IntelliCare data. It produce better association with user characteristics.

Acknowledgments and Disclosure of Funding

This project is supported by NIH grant R01 MH109496.

References

[1] John Aach and George M Church. Aligning gene expression time series with time warping algorithms. *Bioinformatics*, 17(6):495–508, 2001.

[2] Afnan A Al-Subaihin, Federica Sarro, Sue Black, Licia Capra, Mark Harman, Yue Jia, and Yuanyuan Zhang. Clustering mobile apps based on mined textual features. In *Proceedings of the 10th ACM/IEEE international symposium on empirical software engineering and measurement*, pages 1–10, 2016.

[3] Thevaa Chandereng and Anthony Gitter. Lag penalized weighted correlation for time series clustering. *BMC bioinformatics*, 21(1):1–15, 2020.

[4] Ken Cheung, Wodan Ling, Chris J Karr, Kenneth Weingardt, Stephen M Schueller, and David C Mohr. Evaluation of a recommender app for apps for the treatment of depression and anxiety: an analysis of longitudinal user engagement. *Journal of the American Medical Informatics Association*, 25(8):955–962, 2018.

[5] Ying Kuen Cheung, Pei-Yun Sabrina Hsueh, Ipek Ensari, Joshua Z Willey, and Keith M Diaz. Quantile coarsening analysis of high-volume wearable activity data in a longitudinal observational study. *Sensors*, 18(9):3056, 2018.

[6] Emma J Cooke, Richard S Savage, Paul DW Kirk, Robert Darkins, and David L Wild. Bayesian hierarchical clustering for microarray time series data with replicates and outlier measurements. *BMC Bioinformatics*, 12(1):399, 2011.

[7] David B. Dahl. Model-Based Clustering for Expression Data via a Dirichlet Process Mixture Model. In Kim-Anh Do, Marina Vannucci, and Peter Müller, editors, *Bayesian Inference for Gene Expression and Proteomics*, pages 201–218. Cambridge University Press, 2006.

[8] Peter Farago. App engagement: The matrix reloaded. *Flurry Analytics Blog*, 2012.

[9] Francis D. Gibbons and Frederick P. Roth. Judging the quality of gene expression-based clustering methods using gene annotation. *Genome Research*, 12(10):1574–1581, October 2002.

[10] Toni Giorgino. Computing and visualizing dynamic time warping alignments in r: the dtw package. *Journal of Statistical Software*, 31(7):1–24, 2009.

[11] John A Hartigan and Manchek A Wong. Ak-means clustering algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 28(1):100–108, 1979.

[12] Sarra L Hedden. *Behavioral health trends in the United States: Results from the 2014 national survey on drug use and health*. Substance Abuse and Mental Health Services Administration, Department of ?, 2015.

[13] Daniel F Heitjan and Donald B Rubin. Ignorability and coarse data. *The annals of statistics*, pages 2244–2253, 1991.

[14] Christian Hennig, Marina Meila, Fionn Murtagh, and Roberto Rocci. *Handbook of cluster analysis*. CRC Press, 2015.

[15] Ja-Yoon Jang, Hee-Seok Oh, Yaeji Lim, and Ying Kuen Cheung. Ensemble clustering for step data via binning. *Biometrics*, 77(1):293–304, 2021.

[16] Mary J Kwasny, Stephen M Schueller, Emily Lattie, Elizabeth L Gray, and David C Mohr. Exploring the use of multiple mental health apps within a platform: secondary analysis of the intellicare field trial. *JMIR mental health*, 6(3):e11572, 2019.

[17] Emily G Lattie, Stephen M Schueller, Elizabeth Sargent, Colleen Stiles-Shields, Kathryn Noth Tomasino, Marya E Corden, Mark Begale, Chris J Karr, and David C Mohr. Uptake and usage of intellicare: a publicly available suite of mental health and well-being apps. *Internet Interventions*, 4:152–158, 2016.

[18] Ian C McDowell, Dinesh Manandhar, Christopher M Vockley, Amy K Schmid, Timothy E Reddy, and Barbara E Engelhardt. Clustering gene expression time series data using an infinite gaussian process mixture model. *PLoS Computational Biology*, 14(1):e1005896, 2018.

[19] Mario Medvedovic and Siva Sivaganesan. Bayesian infinite mixture model based clustering of gene expression profiles. *Bioinformatics*, 18(9):1194–1206, 2002.

[20] David C Mohr, Kathryn Noth Tomasino, Emily G Lattie, Hannah L Palac, Mary J Kwasny, Kenneth Weingardt, Chris J Karr, Susan M Kaiser, Rebecca C Rossom, Leland R Bardsley, et al. Intellicare: an eclectic, skills-based app suite for the treatment of depression and anxiety. *Journal of medical Internet research*, 19(1):e10, 2017.

[21] Carla S. Möller-Levet, Frank Klawonn, Kwang-Hyun Cho, and Olaf Wolkenhauer. Fuzzy Clustering of Short Time-Series and Unevenly Distributed Sampling Points. In *Advances in Intelligent Data Analysis V*, Lecture Notes in Computer Science, pages 330–340. Springer, Berlin, Heidelberg, August 2003.

[22] Jin-Hong Park, TN Sriram, and Xiangrong Yin. Dimension reduction in time series. *Statistica Sinica*, pages 747–770, 2010.

[23] William M. Rand. Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association*, 66(336):846–850, 1971.

[24] Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(Supplement C):53–65, November 1987.

[25] José Antonio Vilar, Andrés M Alonso, and Juan Manuel Vilar. Non-linear time series clustering based on non-parametric forecast densities. *Computational Statistics & Data Analysis*, 54(11):2850–2865, 2010.

Multivariate Time Series Clustering for Mobile Apps Data: Supplementary Information

1 Supplementary Figures

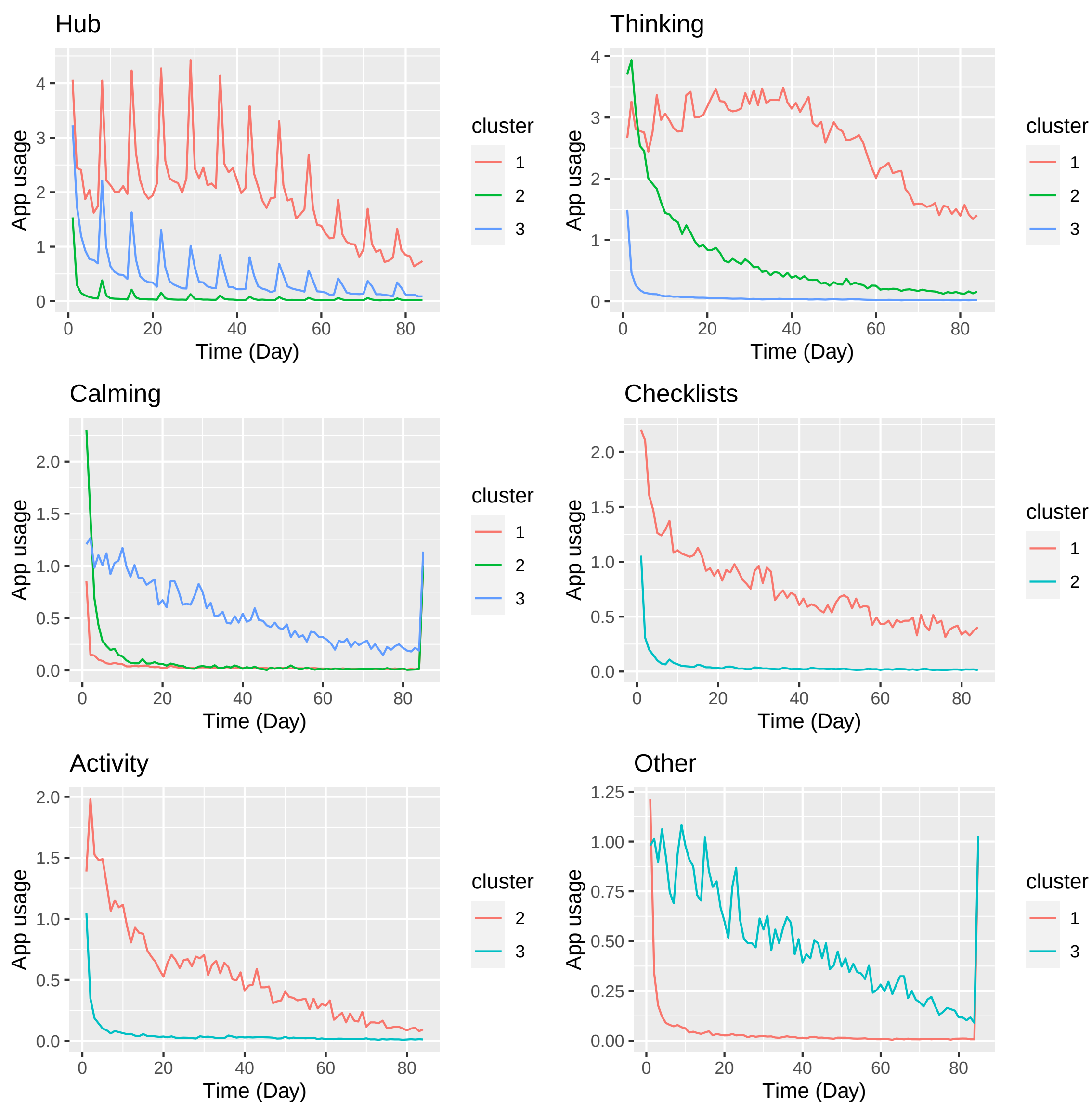

Fig. S1: Mean app usage over time (Day) of all clusters for all the 6 different app categories (hub, thinking, calming, checklists, and other) using kmeans clustering.

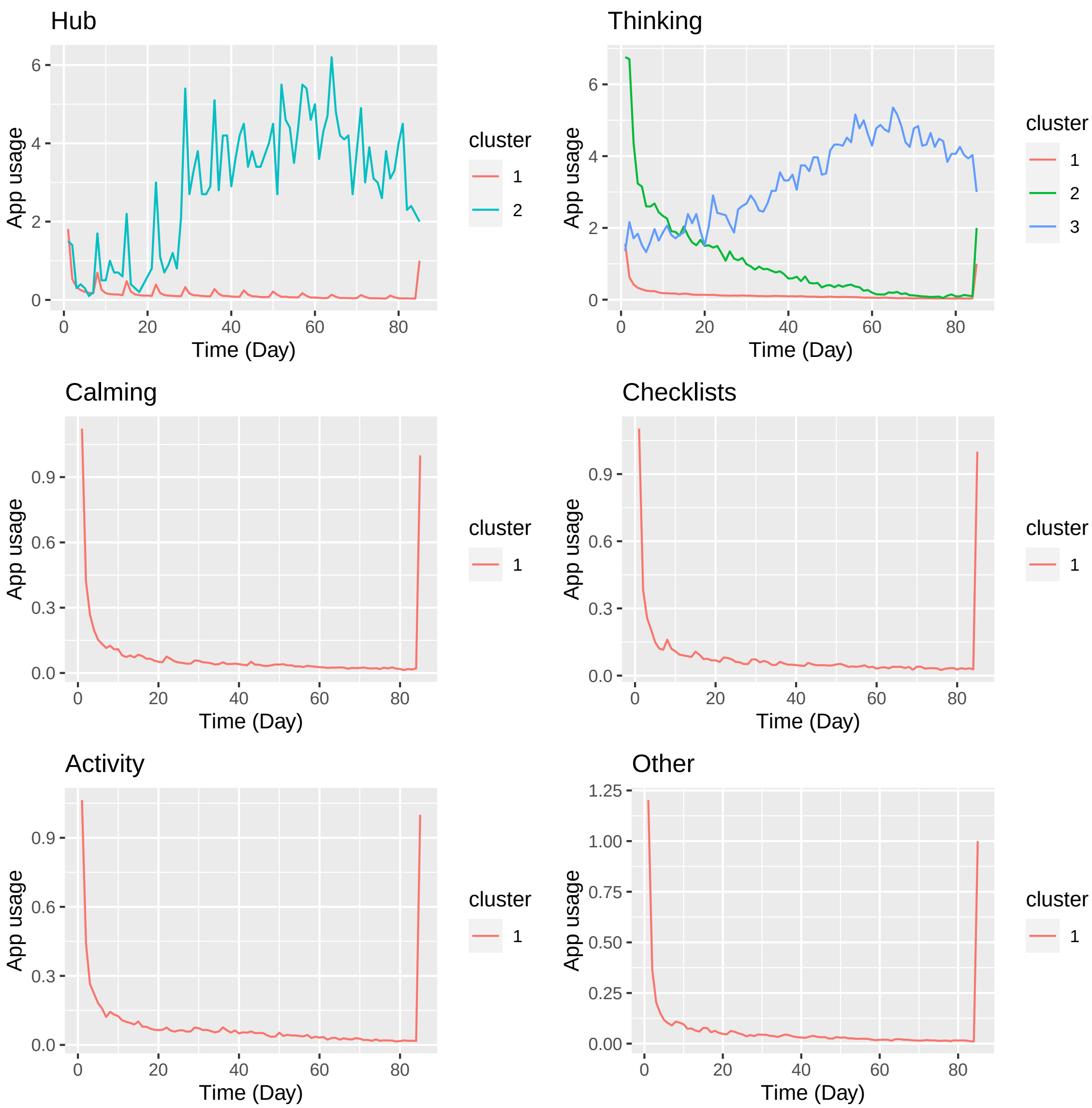

Fig. S2: Mean app usage over time (Day) of all clusters for all the 6 different app categories (hub, thinking, calming, checklists, and other) using hierarchical clustering.

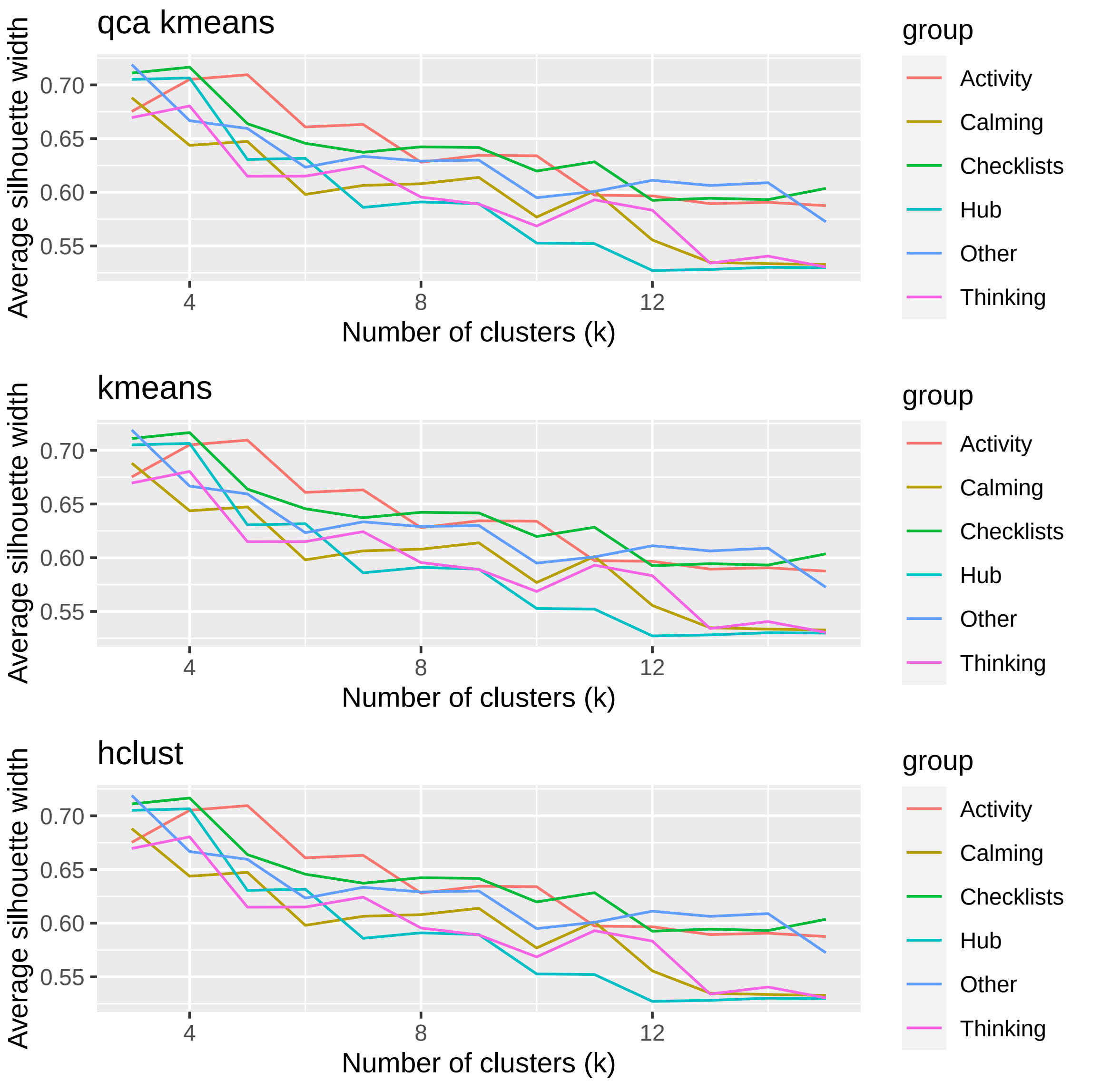

Fig. S3: Average silhouette width for all the 6 different app categories (hub, thinking, calming, checklists, and other) and 3 different clustering methods.

2 Supplementary Tables

Group of Apps	qca kmeans	kmeans	hclust
Hub	0.38 (0.10)	1.00 (0.00)	1.00 (0.00)
Thinking	0.25 (0.06)	1.00 (0.00)	1.00 (0.00)
Calming	0.23 (0.12)	1.00 (0.00)	1.00 (0.00)
Checklists	0.24 (0.08)	1.00 (0.00)	1.00 (0.00)
Activity	0.17 (0.15)	1.00 (0.00)	1.00 (0.00)
Other	0.45 (0.19)	1.00 (0.00)	1.00 (0.00)

Table S1: Mean and (standard deviation) of ARI scores for different clustering methods for permuted app data. The ARI scores were obtained by comparing the clusters from the permuted data to the clusters from the original data. The optimal number of cluster of the permuted app data is selected based on the optimal number of cluster for the original app data.

Characteristic	Cluster 1, N = 50	Cluster 2, N = 199	Cluster 3, N = 2,502	Cluster 4, N = 934	Cluster 0, N = 5,147	p-value	q-value
Total Usage	261 (101)	105 (50)	5 (8)	31 (37)	0 (6)	< 0.001	< 0.001
Age	39 (12)	39 (13)	36 (13)	40 (14)	35 (13)	< 0.001	< 0.001
Gender						< 0.001	< 0.001
Female	42 (84%)	139 (71%)	1,331 (53%)	515 (55%)	3,341 (65%)		
Male	8 (16%)	57 (29%)	1,147 (46%)	411 (44%)	1,702 (33%)		
Other	0 (0%)	1 (0.5%)	17 (0.7%)	5 (0.5%)	60 (1.2%)		
Race						< 0.001	< 0.001
Asian	2 (4.0%)	15 (7.5%)	465 (19%)	88 (9.4%)	571 (11%)		
Black or African American	1 (2.0%)	7 (3.5%)	110 (4.4%)	60 (6.4%)	306 (5.9%)		
Other	3 (6.0%)	12 (6.0%)	248 (9.9%)	83 (8.9%)	647 (13%)		
White	44 (88%)	165 (83%)	1,679 (67%)	703 (75%)	3,623 (70%)		
Ethnicity						< 0.001	< 0.001
Hispanic	2 (4.3%)	9 (4.7%)	206 (10.0%)	93 (11%)	552 (13%)		
Non-Hispanic	45 (96%)	184 (95%)	1,864 (90%)	770 (89%)	3,835 (87%)		
Education						< 0.001	< 0.001
Did not complete high school	2 (4.0%)	2 (1.0%)	91 (3.6%)	21 (2.2%)	279 (5.4%)		
Completed high school	2 (4.0%)	9 (4.5%)	309 (12%)	62 (6.6%)	861 (17%)		
Some college	11 (22%)	62 (31%)	650 (26%)	200 (21%)	1,570 (31%)		
Bachelor's degree	19 (38%)	74 (37%)	824 (33%)	319 (34%)	1,412 (27%)		
Graduate degree (Master's, Ph.D., J.D., M.D., etc.)	16 (32%)	52 (26%)	628 (25%)	332 (36%)	1,025 (20%)		

Table S2: Characteristics of app users (hub) by clusters for mQCA kmeans. The mean (standard deviation) are reported for all continuous variables and the n (%) are reported for all categorical variables. The p-values and q-values (adjusted p-values for false discovery rate) are reported for both one-way anova (continuous variables) and Fisher's exact test (categorical variables).

Characteristic	Cluster 1, N = 101	Cluster 2, N = 3,010	Cluster 3, N = 574	Cluster 0, N = 5,147	p-value	q-value
Total Usage	202 (101)	9 (17)	52 (46)	0 (6)	< 0.001	< 0.01
Age	37 (11)	37 (13)	39 (13)	35 (13)	< 0.001	< 0.01
Gender					< 0.001	< 0.01
Female	79 (78%)	1,588 (53%)	360 (63%)	3,341 (65%)		
Male	22 (22%)	1,393 (46%)	208 (36%)	1,702 (33%)		
Other	0 (0%)	21 (0.7%)	2 (0.4%)	60 (1.2%)		
Race					< 0.001	< 0.01
Asian	7 (6.9%)	512 (17%)	51 (8.9%)	571 (11%)		
Black or African American	2 (2.0%)	157 (5.2%)	19 (3.3%)	306 (5.9%)		
Other	7 (6.9%)	298 (9.9%)	41 (7.1%)	647 (13%)		
White	85 (84%)	2,043 (68%)	463 (81%)	3,623 (70%)		
Ethnicity					< 0.001	< 0.01
Hispanic	5 (5.1%)	260 (10%)	45 (8.4%)	552 (13%)		
Non-Hispanic	93 (95%)	2,277 (90%)	493 (92%)	3,835 (87%)		
Education					< 0.001	< 0.01
Did not complete high school	3 (3.0%)	104 (3.5%)	9 (1.6%)	279 (5.4%)		
Completed high school	3 (3.0%)	343 (11%)	36 (6.3%)	861 (17%)		
Some college	30 (30%)	739 (25%)	154 (27%)	1,570 (31%)		
Bachelor's degree	39 (39%)	1,004 (33%)	193 (34%)	1,412 (27%)		
Graduate degree (Master's, Ph.D., J.D., M.D., etc.)	26 (26%)	820 (27%)	182 (32%)	1,025 (20%)		

Table S3: Characteristics of app users (hub) by clusters for kmeans. The mean (standard deviation) are reported for all continuous variables and the n (%) are reported for all categorical variables. The p-values and q-values (adjusted p-values for false discovery rate) are reported for both one-way anova (continuous variables) and Fisher's exact test (categorical variables).

Characteristic	Cluster 1, N = 75	Cluster 2, N = 3,438	Cluster 3, N = 245	Cluster 4, N = 816	Cluster 0, N = 4,258	p-value	q-value
Total Usage	325 (160)	6 (12)	109 (60)	20 (27)	0 (4)	< 0.001	< 0.001
Age	39 (12)	36 (12)	37 (12)	38 (13)	35 (13)	< 0.001	< 0.001
Gender						< 0.001	< 0.001
Female	54 (72%)	2,172 (64%)	179 (73%)	536 (67%)	2,427 (57%)		
Male	21 (28%)	1,206 (35%)	63 (26%)	265 (33%)	1,770 (42%)		
Other	0 (0%)	31 (0.9%)	2 (0.8%)	4 (0.5%)	46 (1.1%)		
Race						< 0.001	< 0.001
Asian	5 (6.7%)	359 (10%)	20 (8.2%)	85 (10%)	672 (16%)		
Black or African American	1 (1.3%)	161 (4.7%)	8 (3.3%)	57 (7.0%)	257 (6.0%)		
Other	6 (8.0%)	403 (12%)	21 (8.6%)	100 (12%)	463 (11%)		
White	63 (84%)	2,515 (73%)	196 (80%)	574 (70%)	2,866 (67%)		
Ethnicity						0.034	0.034
Hispanic	3 (4.3%)	338 (11%)	18 (7.7%)	75 (10%)	428 (12%)		
Non-Hispanic	67 (96%)	2,683 (89%)	217 (92%)	653 (90%)	3,078 (88%)		
Education						< 0.001	< 0.001
Did not complete high school	1 (1.3%)	178 (5.2%)	4 (1.6%)	23 (2.8%)	189 (4.4%)		
Completed high school	8 (11%)	465 (14%)	15 (6.1%)	76 (9.3%)	679 (16%)		
Some college	13 (17%)	1,055 (31%)	73 (30%)	214 (26%)	1,138 (27%)		
Bachelor's degree	27 (36%)	994 (29%)	92 (38%)	255 (31%)	1,280 (30%)		
Graduate degree (Master's, Ph.D., J.D., M.D., etc.)	26 (35%)	746 (22%)	61 (25%)	248 (30%)	972 (23%)		

Table S4: Characteristics of app users (thinking) by clusters for mQCA kmeans. The mean (standard deviation) are reported for all continuous variables and the n (%) are reported for all categorical variables. The p-values and q-values (adjusted p-values for false discovery rate) are reported for both one-way anova (continuous variables) and Fisher's exact test (categorical variables).

Characteristic	Cluster 1, N = 100	Cluster 2, N = 400	Cluster 3, N = 4,074	Cluster 0, N = 4,258	p-value	q-value
Total Usage	288 (156)	66 (49)	8 (19)	0 (4)	< 0.001	< 0.01
Age	38 (12)	37 (12)	36 (13)	35 (13)	0.14	0.17
Gender					< 0.001	< 0.01
Female	70 (70%)	288 (72%)	2,583 (64%)	2,427 (57%)		
Male	30 (30%)	107 (27%)	1,418 (35%)	1,770 (42%)		
Other	0 (0%)	3 (0.8%)	34 (0.8%)	46 (1.1%)		
Race					< 0.001	< 0.01
Asian	6 (6.0%)	35 (8.8%)	428 (11%)	672 (16%)		
Black or African American	1 (1.0%)	13 (3.2%)	213 (5.2%)	257 (6.0%)		
Other	8 (8.0%)	45 (11%)	477 (12%)	463 (11%)		
White	85 (85%)	307 (77%)	2,956 (73%)	2,866 (67%)		
Ethnicity					0.086	0.086
Hispanic	6 (6.4%)	35 (9.4%)	393 (11%)	428 (12%)		
Non-Hispanic	88 (94%)	337 (91%)	3,195 (89%)	3,078 (88%)		
Education					< 0.001	< 0.01
Did not complete high school	1 (1.0%)	9 (2.2%)	196 (4.8%)	189 (4.4%)		
Completed high school	9 (9.0%)	28 (7.0%)	527 (13%)	679 (16%)		
Some college	22 (22%)	122 (30%)	1,211 (30%)	1,138 (27%)		
Bachelor's degree	35 (35%)	136 (34%)	1,197 (29%)	1,280 (30%)		
Graduate degree (Master's, Ph.D., J.D., M.D., etc.)	33 (33%)	105 (26%)	943 (23%)	972 (23%)		

Table S5: Characteristics of app users (thinking) by clusters for kmeans. The mean (standard deviation) are reported for all continuous variables and the n (%) are reported for all categorical variables. The p-values and q-values (adjusted p-values for false discovery rate) are reported for both one-way anova (continuous variables) and Fisher's exact test (categorical variables).

Characteristic	Cluster 1, N = 118	Cluster 2, N = 1,524	Cluster 3, N = 401	Cluster 0, N = 6,789	p-value	q-value
Total Usage	11.58 (13.09)	4.24 (8.10)	23.16 (33.97)	0.08 (2.08)	< 0.001	< 0.001
Age	39 (13)	36 (13)	39 (13)	36 (13)	< 0.001	< 0.001
Gender					< 0.001	< 0.001
Female	85 (73%)	973 (64%)	271 (68%)	4,039 (60%)		
Male	32 (27%)	535 (35%)	124 (31%)	2,634 (39%)		
Other	0 (0%)	7 (0.5%)	5 (1.2%)	71 (1.1%)		
Race					< 0.001	0.001
Asian	6 (5.1%)	167 (11%)	44 (11%)	924 (14%)		
Black or African American	6 (5.1%)	67 (4.4%)	14 (3.5%)	397 (5.8%)		
Other	11 (9.3%)	171 (11%)	35 (8.7%)	776 (11%)		
White	95 (81%)	1,119 (73%)	308 (77%)	4,692 (69%)		
Ethnicity					0.004	0.004
Hispanic	9 (7.8%)	130 (9.7%)	29 (7.8%)	694 (12%)		
Non-Hispanic	106 (92%)	1,212 (90%)	343 (92%)	5,037 (88%)		
Education					< 0.001	< 0.001
Did not complete high school	1 (0.8%)	60 (3.9%)	17 (4.2%)	317 (4.7%)		
Completed high school	4 (3.4%)	194 (13%)	31 (7.7%)	1,014 (15%)		
Some college	33 (28%)	437 (29%)	90 (22%)	1,933 (28%)		
Bachelor's degree	42 (36%)	479 (31%)	153 (38%)	1,974 (29%)		
Graduate degree (Master's, Ph.D., J.D., M.D., etc.)	38 (32%)	354 (23%)	110 (27%)	1,551 (23%)		

Table S6: Characteristics of app users (calming) by clusters for mQCA kmeans. The mean (standard deviation) are reported for all continuous variables and the n (%) are reported for all categorical variables. The p-values and q-values (adjusted p-values for false discovery rate) are reported for both one-way anova (continuous variables) and Fisher's exact test (categorical variables).

Characteristic	Cluster 1, N = 1,542	Cluster 2, N = 411	Cluster 3, N = 90	Cluster 0, N = 6,789	p-value	q-value
Total_Usage	4.76 (9.94)	9.76 (10.98)	63.97 (46.90)	0.08 (2.08)	< 0.001	< 0.01
Age	36 (13)	36 (12)	39 (14)	36 (13)	0.11	0.11
Gender					< 0.001	0.001
Female	1,000 (65%)	269 (66%)	60 (67%)	4,039 (60%)		
Male	525 (34%)	137 (34%)	29 (32%)	2,634 (39%)		
Other	10 (0.7%)	1 (0.2%)	1 (1.1%)	71 (1.1%)		
Race					< 0.001	0.001
Asian	159 (10%)	48 (12%)	10 (11%)	924 (14%)		
Black or African American	72 (4.7%)	12 (2.9%)	3 (3.3%)	397 (5.8%)		
Other	165 (11%)	49 (12%)	3 (3.3%)	776 (11%)		
White	1,146 (74%)	302 (73%)	74 (82%)	4,692 (69%)		
Ethnicity					< 0.001	0.001
Hispanic	124 (9.1%)	41 (11%)	3 (3.5%)	694 (12%)		
Non-Hispanic	1,242 (91%)	337 (89%)	82 (96%)	5,037 (88%)		
Education					< 0.001	0.001
Did not complete high school	60 (3.9%)	13 (3.2%)	5 (5.6%)	317 (4.7%)		
Completed high school	176 (11%)	44 (11%)	9 (10%)	1,014 (15%)		
Some college	424 (27%)	118 (29%)	18 (20%)	1,933 (28%)		
Bachelor's degree	505 (33%)	134 (33%)	35 (39%)	1,974 (29%)		
Graduate degree (Master's, Ph.D., J.D., M.D., etc.)	377 (24%)	102 (25%)	23 (26%)	1,551 (23%)		

Table S7: Characteristics of app users (calming) by clusters for kmeans. The mean (standard deviation) are reported for all continuous variables and the n (%) are reported for all categorical variables. The p-values and q-values (adjusted p-values for false discovery rate) are reported for both one-way anova (continuous variables) and Fisher's exact test (categorical variables).

Characteristic	Cluster 1, N = 391	Cluster 2, N = 1,696	Cluster 3, N = 52	Cluster 4, N = 113	Cluster 0, N = 6,580	p-value	q-value
Total Usage	20.17 (27.74)	4.00 (15.42)	137.08 (113.69)	19.61 (50.67)	0.07 (2.78)	< 0.001	< 0.001
Age	36 (13)	34 (12)	37 (13)	39 (12)	36 (13)	< 0.001	< 0.001
Gender						< 0.001	< 0.001
Female	285 (74%)	1,128 (67%)	38 (75%)	71 (63%)	3,846 (59%)		
Male	92 (24%)	540 (32%)	12 (24%)	40 (35%)	2,641 (40%)		
Other	10 (2.6%)	20 (1.2%)	1 (2.0%)	2 (1.8%)	50 (0.8%)		
Race						< 0.001	< 0.001
Asian	26 (6.6%)	163 (9.6%)	2 (3.8%)	10 (8.8%)	940 (14%)		
Black or African American	12 (3.1%)	80 (4.7%)	2 (3.8%)	7 (6.2%)	383 (5.8%)		
Other	32 (8.2%)	180 (11%)	2 (3.8%)	8 (7.1%)	771 (12%)		
White	321 (82%)	1,273 (75%)	46 (88%)	88 (78%)	4,486 (68%)		
Ethnicity						0.12	0.12
Hispanic	40 (11%)	154 (10%)	2 (4.1%)	8 (7.3%)	658 (12%)		
Non-Hispanic	333 (89%)	1,337 (90%)	47 (96%)	102 (93%)	4,879 (88%)		
Education						< 0.001	< 0.001
Did not complete high school	10 (2.6%)	83 (4.9%)	2 (3.8%)	0 (0%)	300 (4.6%)		
Completed high school	38 (9.7%)	263 (16%)	7 (13%)	1 (0.9%)	934 (14%)		
Some college	125 (32%)	520 (31%)	10 (19%)	35 (31%)	1,803 (27%)		
Bachelor's degree	129 (33%)	466 (27%)	15 (29%)	42 (37%)	1,996 (30%)		
Graduate degree (Master's, Ph.D., J.D., M.D., etc.)	89 (23%)	364 (21%)	18 (35%)	35 (31%)	1,547 (24%)		

Table S8: Characteristics of app users (checklists) by clusters for mQCA kmeans. The mean (standard deviation) are reported for all continuous variables and the n (%) are reported for all categorical variables. The p-values and q-values (adjusted p-values for false discovery rate) are reported for both one-way anova (continuous variables) and Fisher's exact test (categorical variables).

Characteristic	1, N = 102	2, N = 2,149	0, N = 6,580	**p-value**	**q-value**
Total Usage	87.52 (54.94)	6.62 (21.03)	0.07 (2.78)	< 0.001	< 0.01
Age	37 (13)	35 (13)	36 (13)	0.08	0.10
Gender				< 0.001	< 0.01
Female	77 (76%)	1,445 (68%)	3,846 (59%)		
Male	20 (20%)	663 (31%)	2,641 (40%)		
Other	4 (4.0%)	29 (1.4%)	50 (0.8%)		
Race				< 0.001	< 0.01
Asian	8 (7.8%)	193 (9.0%)	940 (14%)		
Black or African American	3 (2.9%)	98 (4.6%)	383 (5.8%)		
Other	5 (4.9%)	217 (10%)	771 (12%)		
White	86 (84%)	1,641 (76%)	4,486 (68%)		
Ethnicity				0.2	0.2
Hispanic	8 (8.2%)	196 (10%)	658 (12%)		
Non-Hispanic	89 (92%)	1,729 (90%)	4,879 (88%)		
Education				0.2	0.2
Did not complete high school	3 (2.9%)	92 (4.3%)	300 (4.6%)		
Completed high school	15 (15%)	294 (14%)	934 (14%)		
Some college	24 (24%)	666 (31%)	1,803 (27%)		
Bachelor's degree	32 (31%)	619 (29%)	1,996 (30%)		
Graduate degree (Master's, Ph.D., J.D., M.D., etc.)	28 (27%)	478 (22%)	1,547 (24%)		

Table S9: Characteristics of app users (checklists) by clusters for kmeans. The mean (standard deviation) are reported for all continuous variables and the n (%) are reported for all categorical variables. The p-values and q-values (adjusted p-values for false discovery rate) are reported for both one-way anova (continuous variables) and Fisher's exact test (categorical variables). Cluster 3 was removed because it had less than 10 users.

Characteristic	Cluster 1, N = 33	Cluster 3, N = 217	Cluster 4, N = 1,096	Cluster 5, N = 125	Cluster 0, N = 7,361	p-value	q-value
Total Usage	86.88 (46.65)	15.76 (15.42)	3.76 (5.02)	15.61 (19.60)	0.05 (1.01)	< 0.001	< 0.001
Age	38 (12)	40 (14)	35 (12)	38 (12)	36 (13)	< 0.001	< 0.001
Gender						0.6	0.6
Female	23 (72%)	142 (66%)	663 (61%)	85 (68%)	4,455 (61%)		
Male	9 (28%)	72 (33%)	419 (38%)	39 (31%)	2,786 (38%)		
Other	0 (0%)	1 (0.5%)	10 (0.9%)	1 (0.8%)	71 (1.0%)		
Race						0.002	0.004
Asian	0 (0%)	21 (9.7%)	146 (13%)	10 (8.0%)	964 (13%)		
Black or African American	4 (12%)	13 (6.0%)	43 (3.9%)	4 (3.2%)	420 (5.7%)		
Other	2 (6.1%)	16 (7.4%)	110 (10%)	8 (6.4%)	857 (12%)		
White	27 (82%)	167 (77%)	797 (73%)	103 (82%)	5,120 (70%)		
Ethnicity						0.11	0.13
Hispanic	2 (6.1%)	14 (6.8%)	100 (10%)	12 (9.8%)	734 (12%)		
Non-Hispanic	31 (94%)	193 (93%)	870 (90%)	110 (90%)	5,494 (88%)		
Education						< 0.001	< 0.001
Did not complete high school	3 (9.1%)	5 (2.3%)	35 (3.2%)	2 (1.6%)	350 (4.8%)		
Completed high school	4 (12%)	16 (7.4%)	126 (11%)	3 (2.4%)	1,094 (15%)		
Some college	9 (27%)	63 (29%)	295 (27%)	36 (29%)	2,090 (28%)		
Bachelor's degree	11 (33%)	75 (35%)	343 (31%)	42 (34%)	2,177 (30%)		
Graduate degree (Master's, Ph.D., J.D., M.D., etc.)	6 (18%)	58 (27%)	297 (27%)	42 (34%)	1,650 (22%)		

Table S10: Characteristics of app users (activity) by clusters for mQCA kmeans. The mean (standard deviation) are reported for all continuous variables and the n (%) are reported for all categorical variables. The p-values and q-values (adjusted p-values for false discovery rate) are reported for both one-way anova (continuous variables) and Fisher's exact test (categorical variables). Cluster 2 was removed because it had less than 10 users.

Characteristic	Cluster 2, N = 98	Cluster 3, N = 1373	Cluster 0, N = 7,361	p-value	q-value
Total Usage	49.03 (39.18)	5.50 (9.32)	0.05 (1.01)	< 0.001	< 0.01
Age	38 (13)	36 (13)	36 (13)	0.041	0.049
Gender				0.13	0.13
Female	71 (73%)	842 (62%)	4,455 (61%)		
Male	25 (26%)	514 (38%)	2,786 (38%)		
Other	1 (1.0%)	11 (0.8%)	71 (1.0%)		
Race				< 0.001	< 0.01
Asian	4 (4.1%)	173 (13%)	964 (13%)		
Black or African American	6 (6.1%)	58 (4.2%)	420 (5.7%)		
Other	4 (4.1%)	132 (9.6%)	857 (12%)		
White	84 (86%)	1,010 (74%)	5,120 (70%)		
Ethnicity				0.029	0.044
Hispanic	5 (5.3%)	123 (9.9%)	734 (12%)		
Non-Hispanic	90 (95%)	1,114 (90%)	5,494 (88%)		
Education				< 0.001	< 0.01
Did not complete high school	3 (3.1%)	42 (3.1%)	350 (4.8%)		
Completed high school	9 (9.2%)	140 (10%)	1,094 (15%)		
Some college	26 (27%)	377 (27%)	2,090 (28%)		
Bachelor's degree	36 (37%)	435 (32%)	2,177 (30%)		
Graduate degree (Master's, Ph.D., J.D., M.D., etc.)	24 (24%)	379 (28%)	1,650 (22%)		

Table S11: Characteristics of app users (activity) by clusters for kmeans. The mean (standard deviation) are reported for all continuous variables and the n (%) are reported for all categorical variables. The p-values and q-values (adjusted p-values for false discovery rate) are reported for both one-way anova (continuous variables) and Fisher's exact test (categorical variables). Cluster 1 was removed because it had less than 10 users.

Characteristic	Cluster 1, N = 2,183	Cluster 2, N = 442	Cluster 3, N = 108	Cluster 0, N = 6,099	p-value	q-value
Total Usage	3.55 (4.65)	17.00 (25.49)	14.20 (25.18)	0.06 (1.47)	< 0.001	< 0.001
Age	36 (12)	39 (14)	38 (13)	36 (13)	< 0.001	< 0.001
Gender					< 0.001	< 0.001
Female	1,362 (63%)	285 (65%)	79 (73%)	3,642 (60%)		
Male	797 (37%)	151 (35%)	28 (26%)	2,349 (39%)		
Other	11 (0.5%)	1 (0.2%)	1 (0.9%)	70 (1.2%)		
Race					< 0.001	< 0.001
Asian	243 (11%)	44 (10.0%)	9 (8.3%)	845 (14%)		
Black or African American	101 (4.6%)	26 (5.9%)	4 (3.7%)	353 (5.8%)		
Other	185 (8.5%)	40 (9.0%)	6 (5.6%)	762 (12%)		
White	1,654 (76%)	332 (75%)	89 (82%)	4,139 (68%)		
Ethnicity					0.5	0.5
Hispanic	210 (11%)	41 (10%)	9 (8.6%)	602 (12%)		
Non-Hispanic	1,733 (89%)	363 (90%)	96 (91%)	4,506 (88%)		
Education					< 0.001	< 0.001
Did not complete high school	83 (3.8%)	12 (2.7%)	0 (0%)	300 (4.9%)		
Completed high school	316 (14%)	31 (7.0%)	5 (4.6%)	891 (15%)		
Some college	627 (29%)	98 (22%)	31 (29%)	1,737 (28%)		
Bachelor's degree	648 (30%)	147 (33%)	43 (40%)	1,810 (30%)		
Graduate degree (Master's, Ph.D., J.D., M.D., etc.)	509 (23%)	154 (35%)	29 (27%)	1,361 (22%)		

Table S12: Characteristics of app users (other) by clusters for mQCA kmeans. The mean (standard deviation) are reported for all continuous variables and the n (%) are reported for all categorical variables. The p-values and q-values (adjusted p-values for false discovery rate) are reported for both one-way anova (continuous variables) and Fisher's exact test (categorical variables).

Characteristic	Cluster 1, N = 2,625	Cluster 3, N = 108	Cluster 0, N = 6,099	p-value	q-value
Total_Usage	4.41 (7.17)	48.35 (36.21)	0.06 (1.47)	< 0.001	< 0.01
Age	36 (13)	41 (13)	36 (13)	< 0.001	< 0.01
Gender				< 0.001	0.001
Female	1,647 (63%)	79 (73%)	3,642 (60%)		
Male	948 (36%)	28 (26%)	2,349 (39%)		
Other	12 (0.5%)	1 (0.9%)	70 (1.2%)		
Race				< 0.001	< 0.01
Asian	290 (11%)	6 (5.6%)	845 (14%)		
Black or African American	128 (4.9%)	3 (2.8%)	353 (5.8%)		
Other	225 (8.6%)	6 (5.6%)	762 (12%)		
White	1,982 (76%)	93 (86%)	4,139 (68%)		
Ethnicity				0.038	0.038
Hispanic	255 (11%)	5 (4.7%)	602 (12%)		
Non-Hispanic	2,091 (89%)	101 (95%)	4,506 (88%)		
Education				< 0.001	< 0.01
Did not complete high school	94 (3.6%)	1 (0.9%)	300 (4.9%)		
Completed high school	345 (13%)	7 (6.5%)	891 (15%)		
Some college	737 (28%)	19 (18%)	1,737 (28%)		
Bachelor's degree	795 (30%)	43 (40%)	1,810 (30%)		
Graduate degree (Master's, Ph.D., J.D., M.D., etc.)	654 (25%)	38 (35%)	1,361 (22%)		

Table S13: Characteristics of app users (other) by clusters for kmeans. The mean (standard deviation) are reported for all continuous variables and the n (%) are reported for all categorical variables. The p-values and q-values (adjusted p-values for false discovery rate) are reported for both one-way anova (continuous variables) and Fisher's exact test (categorical variables). Cluster 2 was removed because it had less than 10 users.