Dear Smith School Students,

We are pleased to announce an exciting opportunity for you to flex your data analytics skills using real-world data from field research that aims to help prevent the further spread of HIV and AIDS, and possibly win the CHIDS Summer Data Clustering Challenge (plus an internship).

The uptake of HIV prevention products among adolescents and young women in Sub-Saharan Africa has been problematic. Suboptimal product use in recent prevention trials highlights the challenges for conferring protection to targeted users. Current methods of assessing acceptability have not predicted actual use by this population. New approaches are needed to better understand the sociocultural context and mindset of young women in Sub-Saharan Africa.

Towards this purpose, we administered a detailed survey to 1,500 low income (LSM 4-7) South African girls and women between the ages of 14 and 25 living in urban and small-urban locations using a face-to-face computer-aided personal interview. Anchored on two personal care product journeys, the survey covered information channels, media usage, shopping behaviors, key influencers, and attitudes, beliefs, lifestyles and demographics.

The challenge is the following… using sound data analysis methodologies for clustering and judgment, develop an optimal set of clusters that helps segment users into a typology consisting of distinct classes. For example, one may hypothesize that there are four distinct clusters, and that one of the clusters is that of mothers who consume social media, buy products based on similar dimensions, along with commonalities in their attitudes (the analyst should define these commonalities and other similarities of those in the same cluster). You will need to make decisions about what specific variables you want to include in the clustering, given the large set of individual factors that have been collected.

**Deliverable**
Participating students shall submit a PDF of a PowerPoint deck that includes at a minimum:

1. Methods - what one did and how
2. Results
3. Suggested Typology and Justification/Rationale
4. And a slide for each segment in the final suggested typology that defines the key characteristics of that segment

It is recommended to use R / Python / Tableau to conduct the analysis.
About the Data
The data contains 2,015 variables related to the information noted above. Variables include binary responses, Likert scales, and categorical data. The survey downstream flow depends on upstream answers, so not every participant will answer all the exact same questions, but the vast majority of questions will be the same across respondents.

Resources
- Background information deck
- Survey format
- Spreadsheet containing the Data in Label Format, Data in Numeric Format, Code Book, Variables List
- Professor Gao and Kenyon Crowley will host a short webinar on August 2nd at noon to provide additional context and answer questions. The webinar will be recorded and available for viewing. Join WebEx meeting at https://go.umd.edu/clusterbrief

The data for the challenge can be accessed using your Smith School email address at the following Box account link: https://go.umd.edu/clusterchallengedata

Deadline
The slide deck should be emailed to chids@rhsmith.umd.edu by Friday, August 10th at midnight, with the subject line YOURLASTNAME_Clustering Challenge. The code used should also be attached to the email. The top 5 individuals/teams will then be asked to present, which can be done via Skype.

First Prize
The winning student/team will be crowned the CHIDS Clustering Challenge winner, and may be offered an internship with CHIDS for the fall semester. (Team entries are allowable, but there is a single internship available), or $150
Second Prize: First runner-up award and $75
Third Prize: Second runner-up award and $50

About CHIDS
The Center for Health Information and Decision Systems (CHIDS) is an academic research center based in the Decision, Operations & Information Technologies (DO&IT) department in the Robert H. Smith School of Business, which collaborates closely with industry, government, and other key health system stakeholders. The research at CHIDS seeks to understand how digital technologies can be more effectively deployed to address outcomes such as quality, efficiency in healthcare delivery, patient safety, and a reduction in health disparities. CHIDS offers the benefit of a world-class research staff and renowned scholars in healthcare analytics and modeling, and health information technology design, adoption, and evaluation. CHIDS is a pioneer in the study of digitally enabled health system transformation, widely known for its thought leadership and research collaborations.