Clustering Pattern of Physical Activity in Pregnant Women

May 4, 2012

The data source was gathered from participants in the Blossom Project, an IRB approved project studying physical activity in pregnant women. Participants have been selected based on several criteria: age (18-45), pregnant with a single child, enroll prior to 19 weeks of gestation, not a smoker, no history of chronic diseases, and able to communicate without language or mental-status barriers. All the participants were provided with a SenseWear® Mini physical activity armband, StepWatch, and ActivPal activity monitor. These devices record multiple variables that characterize the intensity of a person's trend of activity throughout the day. More specifically, there are 20 subjects. For each subject, 15 variables (Step, METs, Sleep, Lying down, Energy expenditure, and so on) were recorded every minute. Unfortunately, there were some circumstances that led to missing values. For example, subjects would take off the devices to shower, and the devices would temporarily shut down sometimes. Therefore, the number of days for each enrolled person were not balanced.

We clustered individuals by their daily activity trends. The aim was to answer the following questions:

1. Are there clusters consisting of similar physical activity patterns in this sample of pregnant women?
2. Do different clustering algorithms produce similar or different results?

METHODS

1. Combine the three data files generated from three devices to one data file for each subject.
2. Decide which variables to use for clustering.
3. Generate means for each subject for each day, do this for all the variables that were used.
4. Define ID denoted as the combination of subject number and day.
5. Standardized all variables to zero mean and unit variance, then calculated Euclidean distance between each pair of ID's

Figure 1: Results for all the variables across three groups

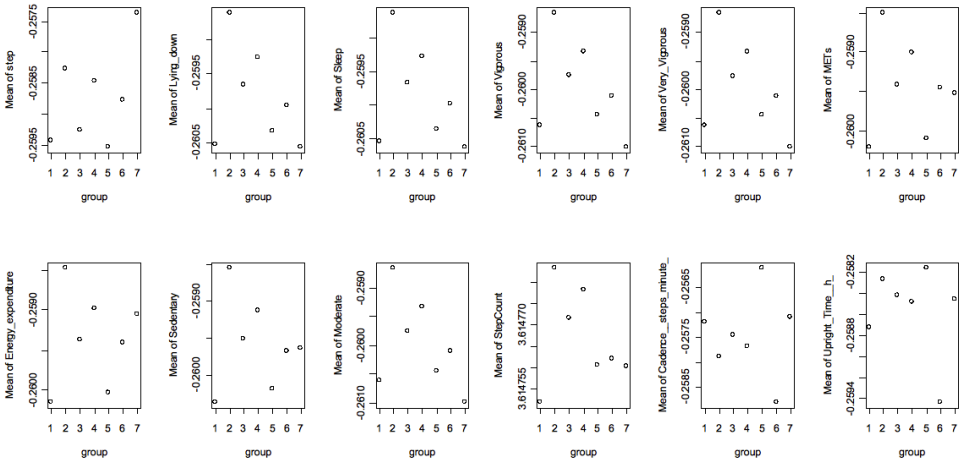
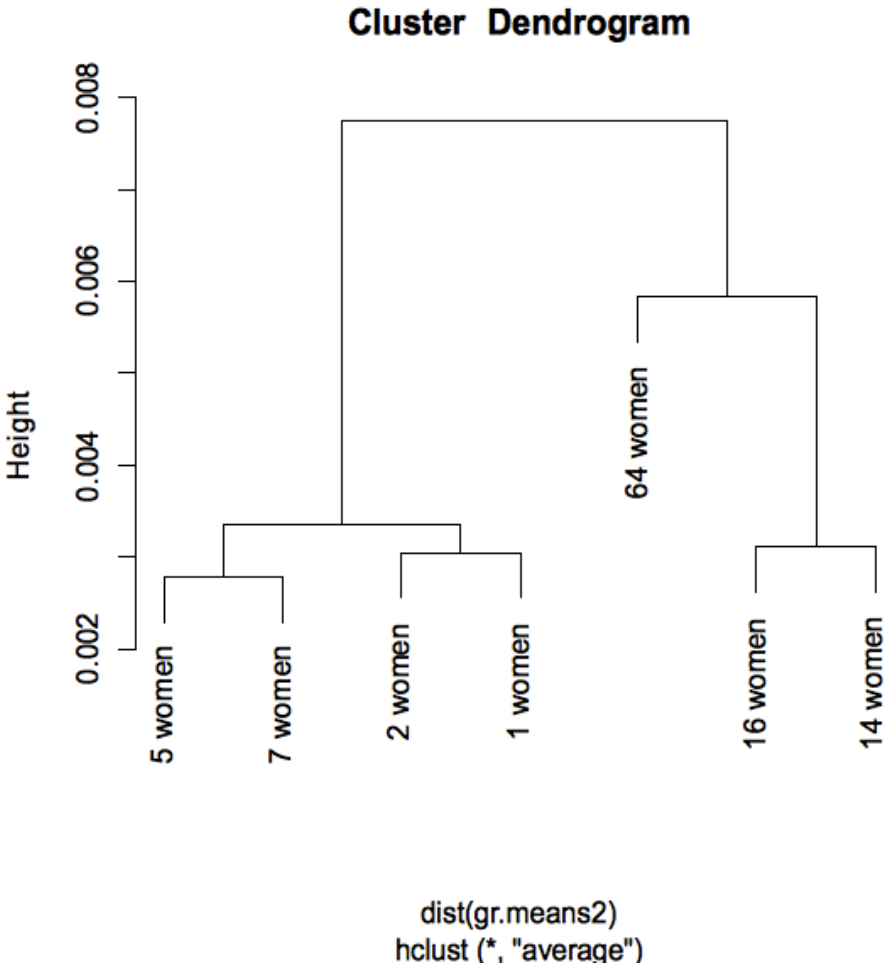


Figure 2: Clustering dendrogram for seven groups



6. Used hierarchical clustering with average linkage to identify groups of ID's with similar activity patterns. We chose average linkage because it had best correlation between the tree distance and the actual distance.

RESULTS AND INTERPRETATION

- Distribution of subjects into the three clusters
  - There are two groups have less than 3 observations. Among the 7 groups, one of than has much more observations than the others (64 observations, see the Figure 2).
- In order to explore the differences between groups, we calculate the mean of each variables across the three groups
  - We can see that variables Lying_down, Sleep, Moderate, Vigorous, Very_Vigorous, METs and StepCount share similar pattern.
  - Variables Sedentary_to_Upright_Movements and Upright_to_Sedentary_Movements share similar pattern