Project Report

Dean D'souza

2016-08-11

## Data Collection

The data collection process was done over a period of roughly 51 days (from 06/22/2016 - 08/11/2016) and was driven by the following main questions:

1. How well have I managed my expenditure in various categories over the span of the month?(Keeping a $60 a day as acceptable and ideal)
2. On average how well have I distributed my time across activities I perform on a near regular basis?
3. Does spending time on any particular activity have an influence of my overall mood on that day?
4. Does the amount I walk around in a day help or in any way affect the amount I sleep?

The data I gathered to answer these questions were limited due to various constraints and were mainly manually input into a microsoft excel file which was converted to .csv format for ease of use with R (mainly due to familiarity on working with the format).  
 Due to space constraints, the data table is not described here but is present under the **About** page tab in the dashboard file.However, some of the variables may need further explanation and are as follows:

1. Area - As this data set had to be built during a time where I needed to travel a lot and shift to a new apartment, I had to visit and stay at a few places before finally settling down in a new apartment. The values are namely 'Jc'-Jersey City, 'BK'-Brooklyn, 'NR'-New Rochelle, 'HAR'-Harrisburg, 'KOP'-King of Prussia and 'DVN'-Devon.   
   
 2. Hrs.. - This set of variables are mostly an approximate as they try to take into account distractions experienced in between. Further,'HrsYoutube' also includes the time I would play a music Playlist and leave it on in the background.   
   
 3. Mood - This categorical variable doesn't properly describe the changes in mood over the day but is an estimate of my overall mood at the end of the day.

## Analysis

We try to answer the first question of 'How well have I managed my expenditure?'. Put in other words, have I been able to keep within my limit of $60 per day? However, most columns contatin one time payments that should not be considered, so we put forward a new question, 'Have I kept my daily requirements expenditure less than $60?'. We calculate these as follows:

## [1] 34

Based on the above conditions (if we exclude expenditure due to other reasons), I was able to keep within my 'daily requirements' budget for most days. But when it comes to visualizing the data, it would be better to place **CutOff** values for each important category.

Next, we try to answer the question of 'How well have I distributed my time across activities?'. In order to do this we try to compare the averages over each week.

## HrsStudy HrsRead HrsYoutube HrsNetflix HrsESO Total  
## [1,] 0.1328671 0.3496503 0.2587413 0.1958042 0.062937063 1  
## [2,] 0.1538462 0.3538462 0.1538462 0.3307692 0.007692308 1  
## [3,] 0.3925234 0.3457944 0.1401869 0.1214953 0.000000000 1  
## [4,] 0.3703704 0.4074074 0.2222222 0.0000000 0.000000000 1  
## [5,] 0.2142857 0.2589286 0.2232143 0.2142857 0.089285714 1  
## [6,] 0.2639594 0.2131980 0.2741117 0.1319797 0.116751269 1  
## [7,] 0.3615819 0.1920904 0.2485876 0.1073446 0.090395480 1

From the above matrix of ratios we can see that for the most part, time allocated is distributed unevenly per week which indicates that better management of time would be required.(personal opinion).

Next we try to answer the question on 'Whether spending time on any particular activity have an effect on my overall mood?'. We do this by seeing the correlation among variables as follows:

## HrsSleep HrsStudy HrsRead HrsYoutube HrsNetflix  
## HrsSleep 1.00000000 -0.3951737 0.09310362 -0.07606003 0.17297856  
## HrsStudy -0.39517370 1.0000000 0.19004251 0.38735901 -0.12223558  
## HrsRead 0.09310362 0.1900425 1.00000000 0.48458925 0.47366266  
## HrsYoutube -0.07606003 0.3873590 0.48458925 1.00000000 0.07056844  
## HrsNetflix 0.17297856 -0.1222356 0.47366266 0.07056844 1.00000000  
## HrsESO 0.11131413 0.2240996 0.22643317 0.66740646 0.09490167  
## Mcode -0.02786148 -0.1339527 -0.05828840 -0.06302671 0.27876724  
## HrsESO Mcode  
## HrsSleep 0.11131413 -0.02786148  
## HrsStudy 0.22409963 -0.13395272  
## HrsRead 0.22643317 -0.05828840  
## HrsYoutube 0.66740646 -0.06302671  
## HrsNetflix 0.09490167 0.27876724  
## HrsESO 1.00000000 -0.04960916  
## Mcode -0.04960916 1.00000000

From the above correlation matrix we can see that when it comes to 'Mood', 'HrsNetflix' seems to have the most positive effect. The remaining covaraiates have a mix of negative and positive effect but not as much.

Next we try to answer the question of 'does the amount I walk have an effect on my sleep?'. We do this as follows:

## DistWalk HrsSleep  
## DistWalk 1.00000000 -0.03705786  
## HrsSleep -0.03705786 1.00000000

From the correlation matrix we see that it does have a decent effect, but let's take into account distance traveled as well. We do it as follows:

## DistTrav DistWalk HrsSleep  
## DistTrav 1.0000000 -0.24030848 -0.16524737  
## DistWalk -0.2403085 1.00000000 -0.03705786  
## HrsSleep -0.1652474 -0.03705786 1.00000000

From the above correlation matrix we can see that distance traveled has a negative correlation with the amount of sleep I get.

## Visualization methods

In order to get a better idea through visualization means, we select the following visualization methods:

1. The first question can be visualized better with the help of a line plot distinguished by the amount spent in each category (in our data frame they are separate variables). We further improve on it by setting **CutOff** values for the three important categories of **Groceries.Supplies**, **EatingOut** and **Uber**, and further facetting the plot for better clarity.
2. The second question can be visualized better with the help of a heatmap plotted usign the geom\_tile() function in ggplot2.
3. The third question can be better visualized with the help of a polar coordnate plot that is made to function more as a pie chart such that based on the mood we can see which acticities are performed and how often.
4. The final question can be better visualized with the help of a scatterplot keeping the distance travelled as a factor for the size of points as well as colour of points for better clarity.