Project Report

Dean D'souza

2016-08-05

## Data Collection

The data collection process was done over a period of roughly 44 days (from 06/22/2016 - 08/04/2016) 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).  
 The data can hence be described as follows:

Variables Description

|  |  |
| --- | --- |
| Variable Name | Description |
| Day | The day of the week |
| Date | The date |
| Groceries&Supplies | Amount spent on groceries and other supplies for the house |
| EatingOut | Amount spent on ordering from restaurants or on other items such as coffee |
| Amazon | Amount spent on Amazon |
| Misc | Amount spent on other unrelated items that would normally be a one time purchase |
| Uber | Amount spent on Uber rides |
| OtherMOT | Amount spent on other modes of transpot(mostly buses and NYC Subway) |
| Total | Total Amount spent on that day |
| DistTrav | Distance traveled by modes of transport |
| DistWalk | Distance walked (leisurely or to stores and back) |
| Area | Area I ended up at the end of the day |
| HrsSleep | Hours of sleep I got (From the previous night to the current day) |
| HrsStudy | Time spent studying (including coding and attending class, in hours) |
| HrsRead | Time spent reading (News, Papers,Reddit, etc., in hours) |
| HrsYoutube | Time spent watching videos on YouTube (in hours) |
| HrsNetflix | Time spent watching series or movies on Netflix (in hours) |
| HrsESO | Time spent on Elder Scrolls Online (in hours) |
| Mood | My overall mood on that day |

Some of the variables above may need further explanation 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 Start off by first loading the dataset and keeping it ready for analysis as follows:

We Now move on to answering 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? We can do this with some simple operations as follows:

## [1] 22

## [1] 22

As we can see from the above counts, the number of times I have been successful and unsuccessful in maintating my limit is equal over the past 44 days. But this is not completely true due to the one time payments that needed to be made.Hence, essentialy, we put forward a new question, 'Have I kept my daily requirements expenditure less than $60?'. We calculate these as follows:

## [1] 29

Based on the above conditions (if we exclude expenditure due to other reasons), I was able to keep within my 'daily requirements' budget for 29 days out of 44 days.

Next, we try to answer the question of 'How well have I distributed my time acroos 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,] 0.2247191 0.2247191 0.2640449 0.1629213 0.123595506 1  
## [7,] 0.3050847 0.2372881 0.3050847 0.1016949 0.050847458 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 -0.24308629 0.20234009 0.04987505 0.09853612  
## HrsStudy -0.24308629 1.00000000 0.24957663 0.34070358 -0.09006723  
## HrsRead 0.20234009 0.24957663 1.00000000 0.43933405 0.51192495  
## HrsYoutube 0.04987505 0.34070358 0.43933405 1.00000000 0.11885145  
## HrsNetflix 0.09853612 -0.09006723 0.51192495 0.11885145 1.00000000  
## HrsESO 0.26054979 0.15538008 0.13119915 0.56471017 0.13266144  
## Mcode -0.06356676 -0.09605967 -0.01342152 0.03834176 0.23363750  
## HrsESO Mcode  
## HrsSleep 0.260549794 -0.063566764  
## HrsStudy 0.155380080 -0.096059671  
## HrsRead 0.131199154 -0.013421519  
## HrsYoutube 0.564710172 0.038341760  
## HrsNetflix 0.132661443 0.233637495  
## HrsESO 1.000000000 -0.001476042  
## Mcode -0.001476042 1.000000000

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 1.0000000 0.0709646  
## HrsSleep 0.0709646 1.0000000

From teh 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:

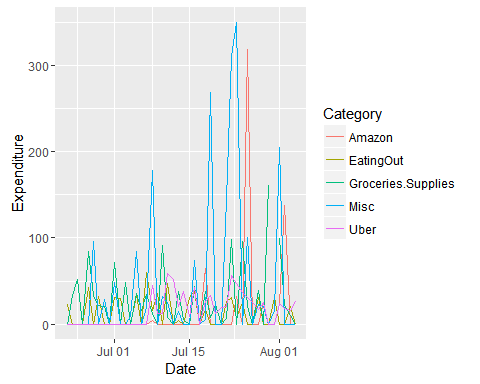
## DistTrav DistWalk HrsSleep  
## DistTrav 1.00000000 -0.3406476 -0.07743456  
## DistWalk -0.34064764 1.0000000 0.07096460  
## HrsSleep -0.07743456 0.0709646 1.00000000

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

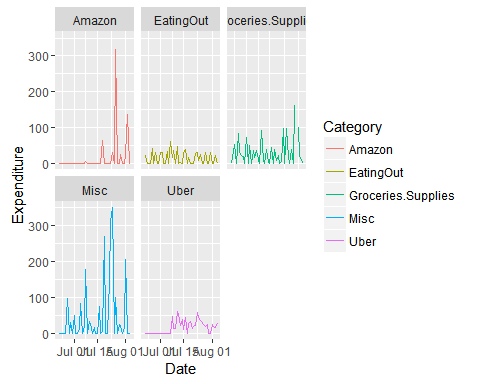
## Visualization methods

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

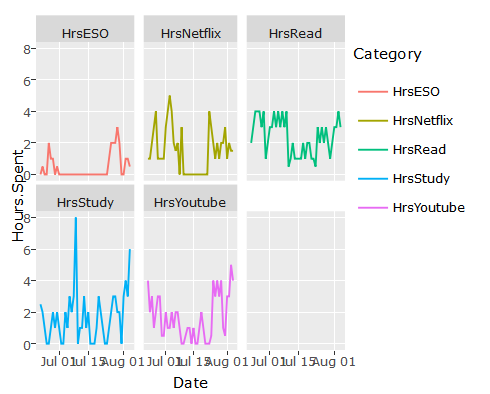
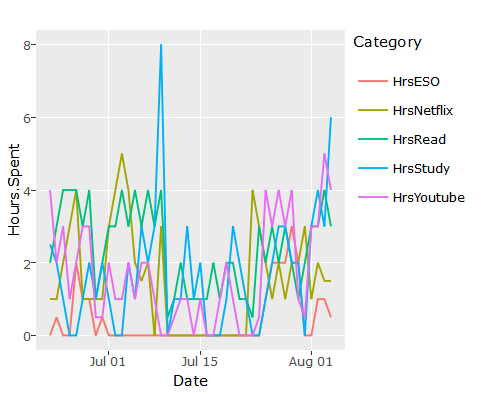
1. The first question can be visualized better with the help of a line plot distinguished by the amount spent in each category (in our data frame they are separate variables)



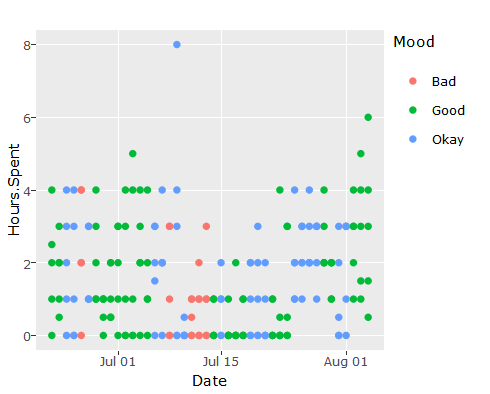
We could take it a step further by facetting the plot for clarity:



1. The second question can be visualized better with the help of a line plot as follows:



1. The third question can be better visualized with the help of a facet wrapped scatterplot as follows:



1. 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.

