The Second Shift Phenomenon

Omar Ahram, Fiona Baenziger, Brandon Markwalder, and Steve Younan

Abstract

We aim to achieve a deeper understanding of how the second shift phenomenon impacts the health and well-being of American women. The attributes we are interested in include their current education level, presence of a spouse in the household, potential spouse's employment status, and location of residence. We are also looking to discover the cascading impact that the second shift phenomenon might have on the entire family unit.

The Principle Component Analysis was used to reduce the number variables requiring exploration. The Common Factor Analysis was also used to minimize the number of variables used in our dataset. The Canonical Correlation Analysis was run to be used as an exploratory tool to test two sets of data to understand how they are related. Through these methods of exploration, we were able to produce relations between the data and multiple sets like Household Activities, Caring for Household Members, and Education. For example, education plays a role in the Second Shift Phenomenon and shows that those who work less hours tend to spend time on research and homework for their personal interest. Also, those older in age with no young children tend to spend little time on education and homework.

The Second Shift Phenomenon impacts the health and well-being of American women. The Second Shift is affected by household activities, education, and care needed for others in the household. The dataset helps provide a great foundation for understanding this phenomenon and there is potential in future analysis.

Introduction

The Second Shift is a relatively new phenomenon that describes the "second shift" that women partake in after paid work hours - such as household chores and activities. Resources that will allude to this shift are both time and energy. The data set we will be analyzing contains nearly 170,000 records and 12 years of survey data about how people spend their time, from personal care to education to work. Our objective is to better understand the relationship between women, their work, their education and how they spend their time at home. In this report, we analyze how this Second Shift Phenomenon manifests through potential causing factors. We use different methods of analysis, such as Principal Component Analysis, Common Factor Analysis, Canonical Correlation and Linear Regression. Throughout this report, we attempt to understand more of the connections between certain activities and apply those connections to how it relates to the Second Shift Phenomenon.

Literature Review

The Second Shift Phenomenon was first coined in 1989 when Arlie Hochschild wrote a book titled "The Second Shift" that addressed the double burden of work and household chores on modern women. Since then, it has been explored by authors and researchers alike from across different fields. Literature that directly references this topic and has added to the analysis of our data include the book that started it all "The Second Shift" by Arlie Hochschild, which addresses the topic of "Is it still relevant now?", and "Intersectionality" by Kimberle Crenshaw, which examines the differences in the effect between race/class.

Specifically pertaining to the Second Shift Phenomenon, studies have shown that "policies such as paternity leave increase male participation in their household, enhance women's participation in the workforce, and promote gender equity in both the home and work setting".[1] Noting that there can be correlations across genders can help increase our ability to analyze the data set. These are conclusions that we hope we can come to by the end of our analysis.

From their research and writing, there are important notes to consider while going through the process of sifting and analyzing this large data set. While there already exists theories and studies about how social norms, the glass ceiling and gender equity affect women in the workplace, we will be exploring daily activities that correlate with this phenomenon.

Methods

The methods chosen to analyze the data set are as follows: Principal Component Analysis, Common Factor Analysis, and Canonical Correlation Analysis. We focused primarily on variable reduction techniques to determine what activities related most to our exploration of the Second Shift Phenomenon.

Principal Component Analysis

A Principal Component Analysis (PCA) was run to reduce the number variables requiring exploration in our efforts to answer the Second Shift Phenomenon. Specifically, the PCA focused on the combination of the Household Activities (M02), the Caring For & Helping Household Members (M03), and the Caring For & Helping Non-Household Members (M04) modules. Collectively, M02, M03, and M04, contain 102 variables covering tasks related to keeping a household running. A sampling of these tasks includes time spent caring for children, cleaning, cooking, doing laundry, performing repairs, and yardwork. With PCA, we aim to reduce the 102 variables in the "Combined Household Activities" module to a subset of latent components capable of explaining how time was spent caring for a household. The data has both desirable and undesirable traits considering the standard assumptions required for PCA. The Kaiser-Meyer-Olkin test for sampling adequacy reveals a mediocre score of 0.63[2], but we consider this acceptable with respect to the size of this social science dataset. Unfortunately, the continuous variables studied during the PCA are not normally distributed. Many of the observations are reported as zero minutes spent doing an activity. There exists wide ranges for select variables which is also cause for concern. We believe the PCA results achieved in this study are still of value, however, they must be interpreted with caution due to the lack of normalcy in the data.

A Common Factor Analysis (CFA) was completed to minimize the number of variables used in our exploration of the Second Shift Phenomenon. With the data set being expansive, we aimed to determine what variables would be the most significant in our deeper analysis of the question. Each variables has a numeric value that represents the number of minutes spent. While there were no null values in the data set, there were a plethora of variables with values of zero. CFA was used to understand the relationship between the 456 variables in order to rule those that showed no significance and those that showed correlation with each other. The variables with correlation values larger than our test number represented related activities and was used as a preliminary to the CFA. The presence of the zeroes in

[1] Gonzales, Teresa Irene. "The Second Shift and Workplace Policies." Everyday Sociology Blog, 6 Feb. 2005

[2] See page 10 of Package 'REdaS' documentation: https://cran.r-project.org/web/packages/REdaS/REdaS.pdf

our data set would present issues in the process of CFA relating to the inability to determine exact values due to the singularity of the matrix. Despite this issue, we feel that CFA will provide good insight into which variables are significant in our exploration of the Second Shift Phenomenon.

A Canonical Correlation Analysis was run as an exploratory tool to test two sets of data for their relation. Looking to see if we can predict a set from another one. The CCA was ran for the variables within the sets of Household Activities (M02), the Caring For & Helping Household Members (M03), and Education (M06) modules. M02 looks at activities around the house like interior cleaning, laundry, and food preparation. M03 focuses on care provided for other members with variables like time spent playing with children, caring for children, and attending children events. M06 is about education and is relevant as time spent focused on education can play a big role as it is time consuming in many ways. It can be time spent in class, on homework, or on research. Using these sets of variables and running them against demographic information helps understand what can be used as a predictor variable.

A Multiple Regression Analysis was conducted on the survey data in order to predict minutes slept based on activities conducted after work. The regression model used minutes slept as a response variables and the predictor variables classified as the following (all in minutes spent): Household Activities (M02), the Caring For & Helping Household Members (M03), Caring for & Helping Nonhousehold (NonHH) Members (M04). A multiple regression model can use those variables to help calculate a prediction value for minutes slept. When analyzing the second shift for woman it's important to note that I am using a regression model because I already assume that a second shift exists. Instead of proving the existence of the second shift, I will create a regression model for minutes slept to help identify what factor affects men and women's total sleep. A regression model will find the most important predictors and will give coefficients that will either increase or decrease minutes slept. Thus, if a certain activity is conducted more by women, it can reduce the amount of minutes slept (analysis will prove that certain activities have a much higher average minutes spent by women).

Discussion and Results

Two separate Principal Component Analyses[3], PCA1[4] and PCA2[5], were conducted on two variations of the Combined Household Activities module[6]. PCA1 and PCA2 each have three components capable of explaining 77.4 and 79.1 percent of the variance in the data respectively. An initial correlation matrix was computed using the spearman method as the data is continuous and it is not normally distributed. Examination of the initial correlation matrix revealed that many of the 102 variables were either weakly correlated or not correlated at all. The correlation matrix was first refined to only include variables with correlation coefficients greater than 0.20 resulting in 12 of the original 102 variables. The threshold was lowered to 0.15 resulting in a more adequate 17 of the original 102 variables. The remaining variables were checked for significant correlations using a confidence interval of 0.95 percent and a significance level of $p < 0.05$. Examination of the correlation matrix shown in Figure 1 demonstrates that our 17 variables are correlated with each other such that we can proceed with the PCA with confidence that our results will be of value.

---

[3] See Appendix for complete PCA R code

[4] Comprised of Modules 02, 03, and 04, totally 102 variables

[5] Data used for PCA1 included all observations

[6] Data used for PCA2 eliminated observations reporting time spent > 240 minutes (4 hours) for any one activity

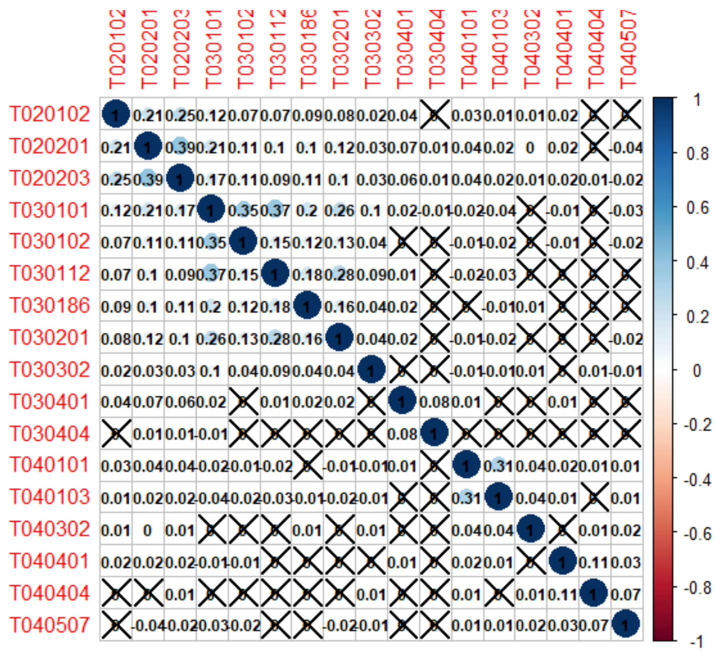

Figure 1

Before running a PCA on the data we first must address the non-normalcy of the data. Each of the 17 variables we have identified for PCA are continuous representing time spent in minutes performing a given household activity. Four of the variables in particular are of concern as they represent a wide range in time spent across the 170,000 observations. The variables in question are shown in Table 1. We have elected against scaling the data as the units of measurement are uniform across all variables, therefore the variance in the data could be of importance. An alternative exploration was conducted in which all observations for a given activity that exceeded 240 minutes were dropped. Time capping the observations both helps to normalize the data and mitigate the extreme cases. We make the assumption that in most cases, individuals do not regularly spend 13 hours doing laundry in a single 24 hour period. We strongly caution, that the PCA2 results which are based off the time capped data are for comparison purposes only, as they are not representative of the data in its entirety. The data was reduced from 170,842 to 167,442 observations after removing those that exceeded 240 minutes for a given activity. The side by side quantile-quantile plot shown in figure 2 demonstrates that all though we mitigated the extreme outliers, we are still working with data that is not normally distributed.

Code	Variable	Time in minutes	Time in hours
T020102	Laundry	810	13.5
T020201	Food and Drink Preparation	975	16.25
T030101	Physical Care for HH Children	1140	19
T030302	Obtaining Medical Care for HH Children	1045	17.42
T030401	Physical Care for HH Adults	1230	20.5

Table 1

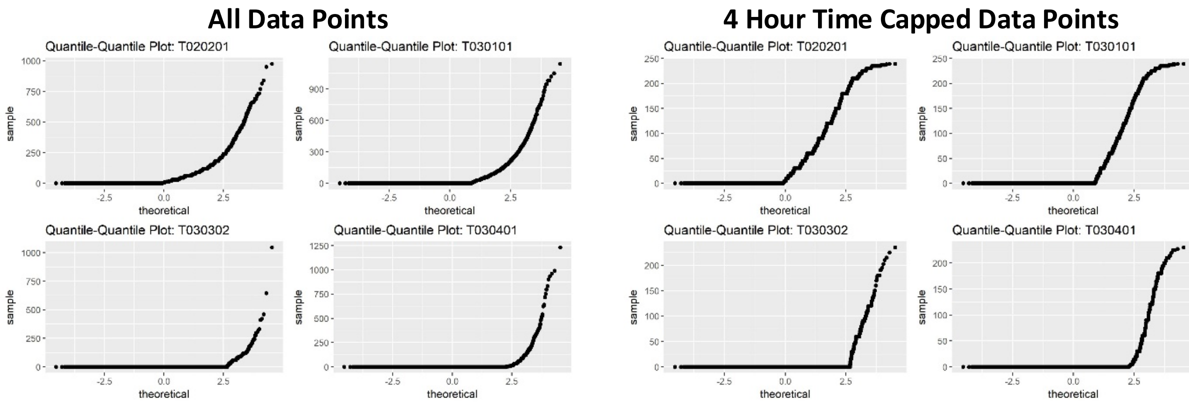

Figure 2

Initially, PCA1 and PCA2 were computed using the prcomp() R function from the Stats R library. PCA1 explained 83.2 percent of the variance in the data whereas PCA2 explained 86.1 percent of the variance in the data. The scree plots for both PCA1 and PCA2 indicated a clear knee at three components. Examination of the loadings for n=3 components and n=4 components revealed easier to interpret results from the use of three components. PCA1 and PCA2 were recomputed using the principal() function from the Psych R library, with varimax rotation set to True and n equal to three components. Technically, PCA2[7] out performed PCA1, however we have elected to use the results for PCA1[8] as they encompass all of the data, with a minimal reduction in performance. The final version of PCA1 is comprised of three components that explain 77.4 percent of the variance in the data. The root mean square of the residuals (RMSR) is 0.08. The formulas for the three components are shown in table 2.

Component	PCA1 Component Formulas
PC1 =	(0.646 * T030101) + (0.507 * T030102) + (0.621 * T030112) + (0.489 * T030201)
PC2 =	(0.507 * T020102) + (0.712 * T020201) + (0.731 * T020203)
PC3 =	(0.732 * T040101) + (0.682 * T040103)

Table 2

The PCA successfully reduced the 102 variables contained within the Combined Household Activities module to three easy to interpret components. We chose a cutoff threshold of 0.30, as it produced an easy to interpret rotation matrix as shown in Table 3. The four variables contained in Principal Component 1 (PC1) show a moderate positive correlation with the component. The variables of PC1 are related to caring for a child that lives in the household, therefore we have named PC1 *Child Care*. The three variables contained in PC2 show a moderate to strong positive correlation with the component. The variables of PC2 are related to household chores and food preparation, lending the name *Chores* to the component. The two variables contained in PC3 show a strong positive correlation with the component. The variables of PC3 capture time spent both caring for, and playing with, children that do not live in the household. Therefore we have name PC3, *Child Friend Care*. The three principal components capture a large amount of information surrounding how time is spent related to maintaining all aspects of a household. Parents and caregivers know that caring for a child and their friends, preparing food, and cleaning the kitchen afterwards, and doing the laundry often dominate how

[7] PCA2 explains 95 percent of the variance in the data with three components and possesses a RMSR = 0.07.
[8] Future studies will include the use of non-parametric PCA to account for the outliers in the data

time is spent at home. The components can be used in further study of the data to uncover how they relate to the Second Shift Phenomenon, allowing for simplified comparisons between household management and activities that are not directly associated with the Second Shift Phenomenon, including advancing education, leisure, and exercise.

	3 Component Solution		
Household Activities	Child Care	Chores	Child Friend Care
Physical Care For HH Children	**0.646**		
Reading to/with HH Children	**0.507**		
Picking up/dropping Off HH Children	**0.621**		
Laundry		**0.507**	
Food and Drink Preparation		**0.712**	
Kitchen and Food Clean-up		**0.731**	
Physical Care for NONHH Children			**0.732**
Playing with NONHH Children, Not Sports			**0.682**
Talking with/listening to HH Children			
Homework (HH Children)	**0.489**		
Obtaining Medical Care for HH Children			
Physical Care for HH Adults			
Obtaining Medical and Care Services for HH Adult			
Obtaining Medical Care for NONHH Children			
Physical Care for NONHH Adults			
Obtaining Medical and Care Services for NONHH Adult			
Picking up/dropping off NONHH Adult			

Table 3

Common Factor Analysis

The Common Factor Analysis (CFA) was completed in order to minimize the number of variables used in our exploration of the Second Shift Phenomenon. Therefore, we focused on the exploratory aspects of CFA versus confirming previous theories based on research. Two CFAs were conducted, the second only because the first one failed to achieve notable results.

In order to prepare the data for this method, we considered the assumptions of Common Factor Analysis. The 170,000 observations deemed it a large enough dataset and every variable was coded as a number of minutes spent on an activity so the scaling was correct. For the sake of our time, we did consider an outlier analysis for CFA but considered it a huge stretch for the amount of time we had. As Factor Analysis is highly reliant on the data being put into the model, outliers can severely affect the outcome of the analysis. For the objective of this project, we will continue on without detecting outliers.

In the first CFA, we used all the variables in the dataset except for demographic information to see if there were correlations amongst the different sections. There is a total of 456 variables, all of which we were attempting to compute factors with; however, there were issues creating an actual factor analysis model as there was a standard deviation of 0, also known as the singular matrix.

The first method to narrow down the variables was to analyze the correlation matrix and use only the pairs that had a correlation greater than 0.3 - as this is a social science data set known for low correlation rates. After using only these highly correlated variables, the factanal() was still reporting a singular matrix so we decided to use a different version of CFA: fa(). There were a total of 4 errors about the model and while it was created, it was not great and the factors did not make sense.

Therefore, we decided to conduct another CFA on a smaller portion of the data, as we determined that only certain sections pertained to our research question. The modules are listed out below:

M01: Personal Care Activities - 12 variables
M02: Household Activities - 31 variables
M03: Caring For & Helping Household (HH) Members - 33 variables
M04: Caring For & Helping Non-Household (NonHH) Members - 38 variables
M05: Work & Work Related Activities - 20 variables
M06: Education - 18 variables

By narrowing down the variables in our preprocessing, we were then able to conduct the factor analysis with 152 attributes as opposed to 456. After computing the eigenvalues and creating a scree plot, we determined that 8 factors was the ideal number, as shown in Figure 3. The 8 factors and the variables comprising them can be examined in Table 4.

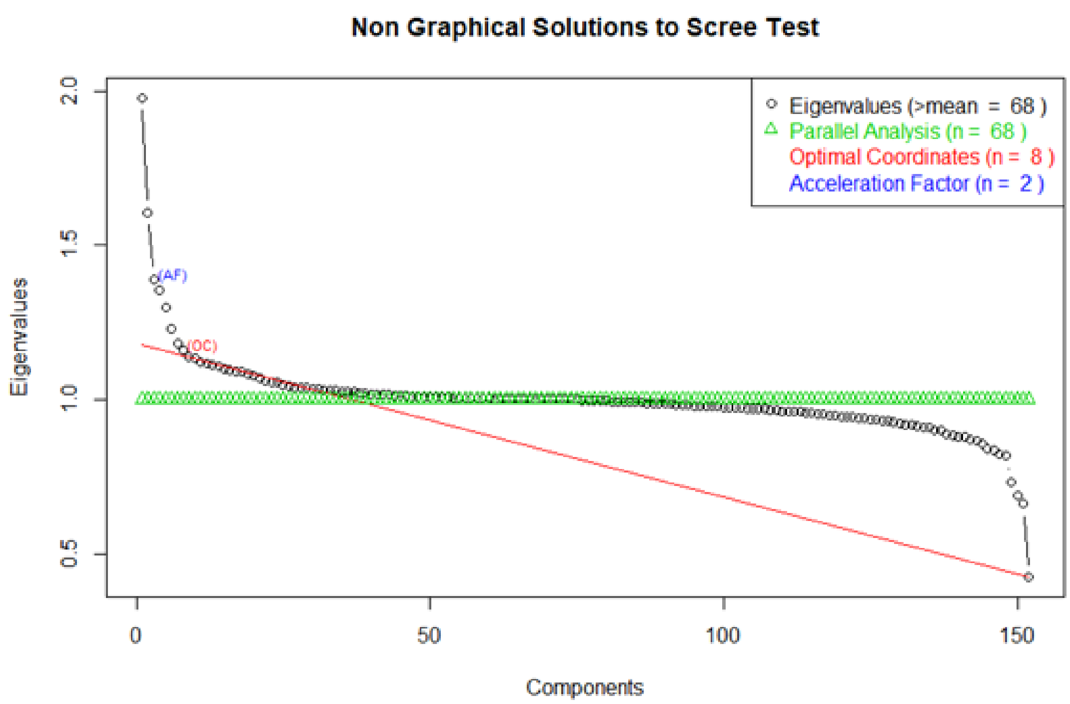

Figure 3

	Code	Variable Name	Value
Sleep v Work	T010101	Sleeping	0.925
	T010201	Sleeplessness	-0.151
	T050101	Work, main job	-0.652
Household Chores	**Code**	**Variable Name**	**Values**
	T010101	Sleeping	-0.161
	T020101	Interior Cleaning	0.228
	T020102	Laundry	0.252
	T020104	Storing interior hh items	0.102
	T020201	Food and drink prep.	0.509
	T020202	Food presentation	0.149
	T020203	Kitchen and food clean up	0.506
	T050101	Work, main job	-0.406
Children	**Code**	**Variable Name**	**Values**
	T010101	Sleeping	-0.156
	T020201	Interior Cleaning	0.119
	T030101	Physical care for hh children	0.569
	T030102	Reading to/with hh children	0.317
	T030103	Playing with hh children	0.514
	T030109	Looking after hh children	0.137
	T030112	Picking up, dropping off hh children	0.148
	T050101	Work, main job	-0.15
Education	**Code**	**Variable Name**	**Values**
	T010101	Sleeping	-0.112
	T050101	Work, main job	-0.238
	T060101	Taking class for degree	0.729
	T060103	Waiting associated with taking classes	0.173
	T060301	Research/homework for class for degree	0.314
Working guardians further their education	**Code**	**Variable Name**	**Values**
	T010101	Sleeping	-0.237
	T010102	Sleeplessness	0.116
	T010301	Health-related self care	0.103
	T020201	Food and drink preparation	-0.136
	T020203	Kitchen and food cleanup	-0.169
	T030101	Physical care for hh children	-0.122
	T030112	Picking up/dropping off hh children	-0.132
	T050101	Work, main job	-0.553
	T060101	Taking class for degree	-0.142
Good Parental Units	**Code**	**Variable Name**	**Values**
	T030101	Physical care for hh children	0.196
	T030108	Organization & planning for hh children	0.114

	T030110	Attending hh children's events	0.128
	T030111	Waiting for/with hh children	0.18
	T030112	Picking up/dropping off hh children	0.399
	T030186	Talking with/listening to hh children	0.142
	T030201	Homework (hh children)	0.286
Teachers	**Code**	**Variable Name**	**Values**
	T040101	Physical care for non hh children	0.339
	T040102	Reading to/with non hh children	0.236
	T040103	Playing with non-hh children	0.321
	T040105	Playing sports with non hh children	0.101
	T040109	Looking after non-hh children	0.126
	T040111	Waiting for/with non-hh childrean	0.145
	T040112	Dropping off/picking up non hh children	0.201
	T040186	Talking with/listening to non-hh children	0.111
	T040201	Homework (non-hh children)	0.104
	T050101	Work, main job	-0.11
Good Family Values	**Code**	**Variable Name**	**Values**
	T030401	Physical care for hh adults	0.409
	T030402	Looking after hh adult	0.224
	T030403	Providing medical care to hh adult	0.246
	T030404	Obtaining medical and care services for hh adults	0.148
	T030405	Waiting associated with caring for household adults	0.15

Table 4

The majority of the attributes stayed consistent with their individual models, as shown in Table 4. There were a couple factors that revealed some interesting information:

Sleep v Work - Factor 1: Sleep and work are inversely correlated. More work = less sleep.

Household Chores - Factor 2: Sleep and work are inversely correlated with the amount of time spent on chores

Children - Factor 3: Sleep and work are inversely correlated with the amount of time spent on children

Education - Factor 4: Sleep and work are inversely correlated to time spent on education.

Working guardians, furthering their education - Factor 5: This one is particularly interesting as it reveals a cross sections of parents who work and are pursuing a higher education. The more time spent on children, education and work, the more likely that more time will be spent being tired and more time focused on fixing health-related issues.

Good parental units - Factor 6: These are the parents that spend a lot of their time with their children. Work and sleep do not appear to be strongly correlated enough to make a difference in this factor.

Teachers - Factor 7: This factor represents those that spend a lot of their time helping non-household children. We are still not sure how to interpret this exactly, but those help non-household children are committed. Potential teachers, though work is inversely correlated.

Good family values - Factor 8: This factor contains those who spend a lot of their time helping either their elderly parents or those living with them. Though we previously did not consider taking care of children or adults as a part of that second shift, this data shows that it has more importance than previously thought.

The limitations of this method are the fact that the model is so sensitive to data that is not normalized. This is a reason why the model is not the best that it could be. With more time, we would like to narrow down the modules to focus solely on the relationships between more relevant variables. We would also like to conduct a more thorough outlier analysis to truly create a more realistic and accurate data set for exploring the variables.

Overall, the Common Factor Analysis revealed that sleep and work related activities contribute to every aspect of our lives which can play a key role in determining the presence and predictors of an individual who participates in the Second Shift Phenomenon. We also created profiles of different types of individuals and where their priorities lie, either with family, work, education or their children. This is good insight as to the dynamic between these aspects of their lives which allows us to make more connections between this and the "Second Shift".

Canonical Correlation Analysis

A Canonical Correlation Analysis (PCA) was run to test the relations between two datasets for predictive purposes. The CCA looked at description of variables and time spent to associate with second shift. We used CCA coefficients and .3 as a benchmark for scoring. We determined canonical variates with Wilks Score. The CCA was run with the following variables:

M02 Household Activities:

	WilksL	F	df1	df2	p
[1,]	0.74	72.03	736	3194933.1	0.00
[2,]	0.88	30.93	682	3069545.5	0.00
[3,]	0.96	12.63	580	2817685.7	0.00

Variates: [1] 0.410710065 0.223565524 0.111878343

Coefficients:

	[,1]	[,2]	[,3]
T020101	0.005	-0.002	0.001
T020102	0.007	-0.004	0.003
T020103	0.007	0.008	0.008
T020104	0.012	0.013	0.005
T020199	0.004	0.010	0.006
T020201	0.011	0.001	-0.007
T020202	0.022	-0.022	-0.045
T020203	0.018	0.006	-0.007
T020299	0.032	-0.026	0.042

T020301	0.000	0.004	-0.004
T020302	-0.002	0.008	-0.006
T020303	-0.005	0.012	-0.014
T020399	-0.005	0.016	-0.017
T020401	-0.003	0.010	-0.010
T020402	-0.002	0.006	-0.008
T020499	0.001	0.006	-0.012
T020501	-0.001	0.011	-0.006
T020502	-0.001	0.009	-0.012
T020599	-0.001	0.005	-0.012
T020681	0.001	0.010	0.014
T020699	0.004	0.002	-0.003
T020701	-0.006	0.004	-0.013
T020799	-0.003	-0.006	-0.039
T020801	-0.005	0.011	-0.007
T020899	-0.019	0.044	-0.027
T020901	0.002	0.017	0.010
T020902	0.003	0.006	0.007
T020903	0.008	0.033	0.028
T020904	0.002	0.013	0.026
T020905	0.002	0.029	0.001
T020999	0.002	0.002	0.010
T029999	-0.001	0.009	-0.005

Scores:

	[,1]	[,2]	[,3]
T020101	**-0.486009629**	0.106523691	**-0.421197483**
T020102	**-0.442050866**	0.133392636	**-0.306821429**
T020103	-0.175587737	-0.182506546	-0.249957240
T020104	-0.172363035	-0.130347573	-0.174147224
T020199	-0.011238133	-0.031132925	-0.065649506
T020201	**-0.695749200**	-0.044931675	**0.301773774**
T020202	-0.191712073	0.075917486	0.289945399
T020203	**-0.618236210**	-0.119389995	0.152203655
T020299	-0.002161927	0.003491589	0.029898179
T020301	0.029363315	-0.150217851	0.048168592
T020302	0.017445524	-0.084420176	0.041692247
T020303	0.064900136	-0.143959853	-0.073085670
T020399	0.014355853	-0.036800530	-0.010198427
T020401	0.054970966	-0.225581748	-0.044554086
T020402	0.080470456	-0.164984131	-0.059877586
T020499	-0.001454171	-0.025103638	-0.015740413
T020501	0.084421992	**-0.589254226**	-0.157086855
T020502	0.006286848	-0.072947697	0.047056573
T020599	0.004212930	-0.008143330	-0.017788192
T020681	-0.044547792	-0.274434038	-0.191122636
T020699	-0.004762530	-0.013358284	0.005032641
T020701	0.188338210	-0.113659914	-0.138594814

T020799	0.007758961	0.003609461	-0.011471394
T020801	0.090641944	-0.185256444	0.052095966
T020899	0.006987431	-0.020111108	-0.017561894
T020901	-0.051723813	**-0.309126896**	0.114367724
T020902	-0.114109171	-0.195851493	0.248273212
T020903	-0.118732312	**-0.405528965**	-0.072908339
T020904	-0.030955566	-0.198733416	**0.485197868**
T020905	-0.011383597	-0.103320022	-0.017561796
T020999	-0.003883587	-0.004188327	0.003314968
T029999	0.009814900	-0.062058922	-0.035490275

	[,1]	[,2]	[,3]
GEMETSTA	0.0048007823	-0.10105117	-0.19256387
GTMETSTA	-0.0073421688	0.01483896	0.05751726
PEEDUCA	-0.0003467662	**-0.30347422**	**0.46479134**
PEHSPNON	0.1178855747	**-0.30272473**	0.11222777
PTDTRACE	0.0007386017	0.20513127	0.28384238
TEAGE	-0.2537673107	**-0.79388426**	-0.20186869
TELFS	**-0.3689857411**	**-0.33685890**	0.13965190
TEMJOT	**0.3468519152**	**0.31048856**	-0.13684389
TESCHENR	0.0749060993	**0.63472731**	0.12561557
TESCHLVL	0.2244337624	**0.35149865**	0.22245309
TESEX	**-0.8364913855**	0.27295202	-0.14394719
TESPEMPNOT	-0.1436952588	-0.28147949	0.22565003
TRCHILDNUM	-0.1863573726	**0.42074477**	**0.30289885**
TRDPFTPT	0.2969512692	0.29964029	-0.09949980
TRERNWA	0.2814212945	0.06792043	0.03133568
TRHOLIDAY	-0.0740759172	-0.01911836	0.22524777
TRSPFTPT	-0.2236148455	-0.04452347	0.11854760
TRSPPRES	0.2057387659	0.24529152	-0.22371166
TRYHHCHILD	-0.0017590043	0.28253871	0.06403358
TUDIARYDAY	0.0439078247	-0.04932327	**-0.43133239**
TUFNWGTP	0.2022417878	0.11671880	0.20996609
TEHRUSLT	**0.3952920366**	0.27571784	-0.15430886
TUYEAR	-0.0147729247	0.10130701	0.15708981

Explanation of Variates:

Canonical Variate 1: Don’t spend time Interior Cleaning, laundry, food and drink prep, kitchen and food cleanup.
For: Higher hours worked, Gender not important, Had one job.

Canonical Variate 2: Don’t spend time on lawn, garden, and houseplant care. Don’t spend time on financial management, or on mail and messaging.
For: Non-Hispanic, non-educated, had one job, but enrolled in school and have young children.

Canonical Variate 3: Doesn’t spend time on interior cleaning and laundry. Spends time on food and drink preparation. Spend time on personal email and messaging.
For: Educated, with young children.

M03 Household_Members_Care:

	WilksL	F	df1	df2	p
[1,]	0.62	108.63	759	3227689.9	0.00
[2,]	0.90	26.95	704	3101429.5	0.00
[3,]	0.96	11.44	651	2974767.4	0.00

Variates: [1] 0.553592919 0.254909217 0.110631715

Coefficients:

	[,1]	[,2]	[,3]
T030101	-0.016	-0.012	0.013
T030102	-0.024	-0.004	-0.031
T030103	-0.007	-0.007	-0.017
T030104	-0.010	0.002	-0.006
T030105	-0.011	0.009	-0.045
T030108	-0.018	0.014	-0.014
T030109	-0.009	0.002	0.002
T030110	-0.008	0.017	-0.026
T030111	-0.014	0.036	0.011
T030112	-0.045	0.069	0.014
T030186	-0.013	0.045	0.010
T030199	-0.008	0.009	0.009
T030201	-0.011	0.027	0.007
T030202	-0.007	0.024	-0.004
T030203	-0.011	0.006	0.024
T030204	0.009	0.037	0.035
T030299	-0.008	0.033	-0.015
T030301	-0.004	0.001	0.003
T030302	-0.006	0.005	0.014
T030303	-0.004	0.006	0.010
T030399	-0.004	0.000	-0.003
T030401	0.001	-0.002	0.016
T030402	0.001	0.000	0.003
T030403	0.002	-0.005	0.020
T030404	0.001	0.000	0.015
T030405	0.000	0.000	0.007
T030499	0.002	-0.001	0.004
T030501	-0.002	-0.002	0.006
T030502	-0.004	-0.005	-0.001
T030503	-0.012	0.012	-0.086
T030504	-0.001	-0.002	-0.005
T030599	-0.001	-0.005	-0.005
T039999	-0.012	-0.006	0.010

Scores:

	[,1]	[,2]	[,3]
T030101	**0.818954263**	**0.3655626624**	-0.2921197879
T030102	**0.395363682**	0.0622424334	0.2223464463

T030103	**0.483161431**	**0.3726128895**	**0.5262376997**
T030104	0.078318342	0.0082704003	0.0267225206
T030105	0.116088368	-0.0364431111	0.2480231760
T030108	0.162959261	-0.0943553838	0.0546929482
T030109	0.278786555	-0.0126800853	-0.0430244386
T030110	0.202286017	**-0.3558902991**	**0.4813860246**
T030111	0.152564522	-0.2432249521	-0.0600691642
T030112	**0.442600372**	**-0.4058021002**	-0.1223513658
T030186	0.226723658	**-0.4934825106**	-0.1050536844
T030199	0.058701965	-0.0371133663	-0.0407273490
T030201	**0.300174477**	**-0.4240110184**	-0.1320717381
T030202	0.067745050	-0.1437921317	0.0077688622
T030203	0.107692647	-0.0529073569	-0.1712383948
T030204	0.010534724	-0.0369322279	-0.0070648579
T030299	0.034088293	-0.0785795969	0.0157250820
T030301	0.094319172	-0.0051683764	-0.0429548139
T030302	0.102883061	-0.0372121915	-0.1096900785
T030303	0.062951110	-0.0370767772	-0.0682100865
T030399	0.024842082	-0.0000622599	0.0103333367
T030401	-0.024655518	0.0274733608	-0.1969158567
T030402	-0.010226243	0.0029101148	-0.0343608923
T030403	-0.016153283	0.0275587672	-0.1159656371
T030404	-0.010446904	0.0004963342	-0.0873166200
T030405	-0.005247423	0.0018411369	-0.0530850726
T030499	-0.010924767	0.0071903806	-0.0158782520
T030501	0.012262065	0.0125258699	-0.0363550599
T030502	0.015734034	0.0129478163	0.0006633258
T030503	0.036469746	-0.0290571312	0.1551605423
T030504	0.008245836	0.0098100769	0.0139271450
T030599	0.005370511	0.0128683671	0.0160760259
T039999	0.022853712	0.0016866682	-0.0054220640

	[,1]	[,2]	[,3]
GEMETSTA	-0.007101999	0.017427151	-0.15350158
GTMETSTA	-0.015906533	-0.005566088	0.13047875
PEEDUCA	0.224436369	-0.022837727	**0.33608292**
PEHSPNON	-0.006371152	0.021129793	0.18359750
PTDTRACE	0.017927119	-0.019063666	-0.11863697
TEAGE	**-0.443977533**	-0.117058574	-0.14848931
TELFS	-0.052450019	0.121478589	**-0.36609206**
TEMJOT	0.055038650	-0.106574583	**0.34856230**
TESCHENR	**0.582375858**	-0.054335622	0.24616950
TESCHLVL	-0.040126741	0.031820913	-0.02079969
TESEX	0.255147455	-0.097672341	**-0.58636007**
TESPEMPNOT	0.297222479	0.031019168	**0.31047071**
TRCHILDNUM	**0.811452402**	**-0.324689445**	-0.00397124
TRDPFTPT	0.069150521	-0.117693558	0.27436203
TRERNWA	0.074999494	-0.040319982	0.51661268

TRHOLIDAY	-0.016459555	0.077422093	0.03677358
TRSPFTPT	**0.436354990**	-0.063015933	0.26637909
TRSPPRES	**-0.393801057**	0.010248938	**-0.32681456**
TRYHHCHILD	0.177429123	**-0.803995662**	0.16957014
TUDIARYDAY	-0.019298067	-0.063267900	0.26165140
TUFNWGTP	-0.162004342	-0.007410879	-0.30221900
TEHRUSLT	0.011960775	-0.070474553	**0.40563438**
TUYEAR	-0.022085048	-0.004370423	0.16304187

Explanation of Variates:

Canonical Variate 1: Spends time on: Physical care for children, reading to/with children, playing with children, picking up/dropping off children, homework
For: Younger Children, younger Age, in school, spouse employed but not present

Canonical Variate 2: Spends time on: Physical care for children, playing with children
Does not spend time on: Attending Children events, picking up/dropping off children, talking/listening to children, homework
For: No young children,

Canonical Variate 3: Spends time on: Playing with children, attends children events
For: Educated with a job, and spouse employed not present

M06 Education:

	WilksL	F	df1	df2	p
[1,]	0.68	161.30	414	2428282.6	0.00
[2,]	0.96	19.32	374	2305093.0	0.00
[3,]	1.00	2.53	336	2181644.6	0.00

Variates: [1] 0.539153636 0.191231159 0.044539277

Coefficients:

	[,1]	[,2]	[,3]
T060101	0.012	0.014	0.000
T060102	0.003	0.001	-0.059
T060103	0.024	-0.013	-0.001
T060104	0.059	0.493	1.227
T060199	0.012	0.006	-0.012
T060201	0.006	0.015	0.015
T060202	0.008	0.023	0.028
T060203	0.007	0.074	0.029
T060289	0.011	0.012	0.007
T060301	0.014	-0.019	0.002
T060302	0.003	-0.011	-0.069
T060303	-0.092	0.020	0.769
T060399	0.013	-0.007	-0.066
T060401	0.017	0.000	-0.032
T060402	0.006	-0.031	-0.154

T060403	0.079	0.016	-0.202
T060499	0.017	0.026	-0.024
T069999	0.013	-0.014	0.012

Scores:

	[,1]	[,2]	[,3]
T060101	**-0.80348843**	**-0.566260923**	-0.007471913
T060102	-0.04530732	-0.008888972	**0.762348305**
T060103	-0.14950689	-0.051294901	0.020277226
T060104	-0.02244402	-0.040015179	-0.050889372
T060199	-0.07479552	-0.020139255	0.090433950
T060201	-0.05731038	-0.066134455	-0.058935382
T060202	-0.11378228	-0.170045152	-0.159465127
T060203	-0.03441873	-0.056708245	-0.020750885
T060289	-0.05262038	-0.040464563	-0.018317002
T060301	**-0.74353748**	**0.645582657**	-0.086384587
T060302	-0.01865904	0.084499448	**0.564274441**
T060303	-0.02477027	0.017548338	-0.095460513
T060399	-0.07092723	0.018605668	0.266657300
T060401	-0.08043867	0.005191578	0.116668626
T060402	-0.00680722	0.006253971	0.072884514
T060403	-0.03812378	-0.008206214	0.069738705
T060499	-0.04559759	-0.051124803	0.053229403
T069999	-0.09712087	0.079207551	-0.063037012

	[,1]	[,2]	[,3]
GEMETSTA	-0.003295640	-0.048399170	0.136647877
GTMETSTA	0.020468923	0.009342824	-0.169635259
PEEDUCA	0.242996850	**0.522495439**	0.242165583
PEHSPNON	0.029009935	0.060086787	-0.265058400
PTDTRACE	-0.058831008	0.078532646	0.089863118
TEAGE	**0.482150053**	0.184769530	0.002471504
TELFS	-0.157702093	-0.143117830	0.268009106
TEMJOT	0.166735703	0.144976509	-0.298191073
TESCHENR	-0.083494539	-0.009176893	-0.027297171
TESCHLVL	**-0.942068539**	**0.297603215**	-0.084822872
TESEX	0.001596272	0.128918028	**0.299928351**
TESPEMPNOT	0.264734551	0.159774058	0.017098470
TRCHILDNUM	-0.168011764	-0.247840298	-0.141380814
TRDPFTPT	0.102675499	0.140014883	-0.206696514
TRERNWA	0.185345013	0.105523052	-0.259568324
TRHOLIDAY	0.029244652	0.055349746	-0.124847572
TRSPFTPT	0.168424293	0.149547624	-0.035841905
TRSPPRES	-0.268732499	-0.170001235	-0.016795966
TRYHHCHILD	**-0.316871855**	**-0.386641560**	-0.057333305
TUDIARYDAY	0.022466473	-0.114958086	-0.165148091
TUFNWGTP	-0.250204345	-0.301486961	0.138568284
TEHRUSLT	0.228209081	0.132435771	**-0.360699860**
TUYEAR	0.014229757	0.083873050	**-0.463487117**

Explanation of Variates:

Canonical Variate 1: Negative relationship with time spent on education, and on research/homework for class.
For: Older age, high school education or less, no young children.

Canonical Variate 2: Less Negative relationship with time spent on education, and but strong relationship with research/homework for class.
For: Educated, college or university, no young children.

Canonical Variate 3: Personal Interest for research and homework, and taking classes for personal interest.
For: Gender, year doesn't matter, less hours worked.

Multiple Linear Regression

Initially there were too many predictors to select from in creating this model, so the first step was to reduce them. Since this is survey data, the easiest way to select data that is accounted for enough. For example, certain variables had almost a zero average for minutes spent each day. Thus, I took the top 20 highest averages for the variables selected. Therefore now we can start checking model assumptions for these 20 variables. Tests for normality were conducted, and that assumption passed without requiring a transformation. The next model assumption was to check for multicollinearity amongst the variables.

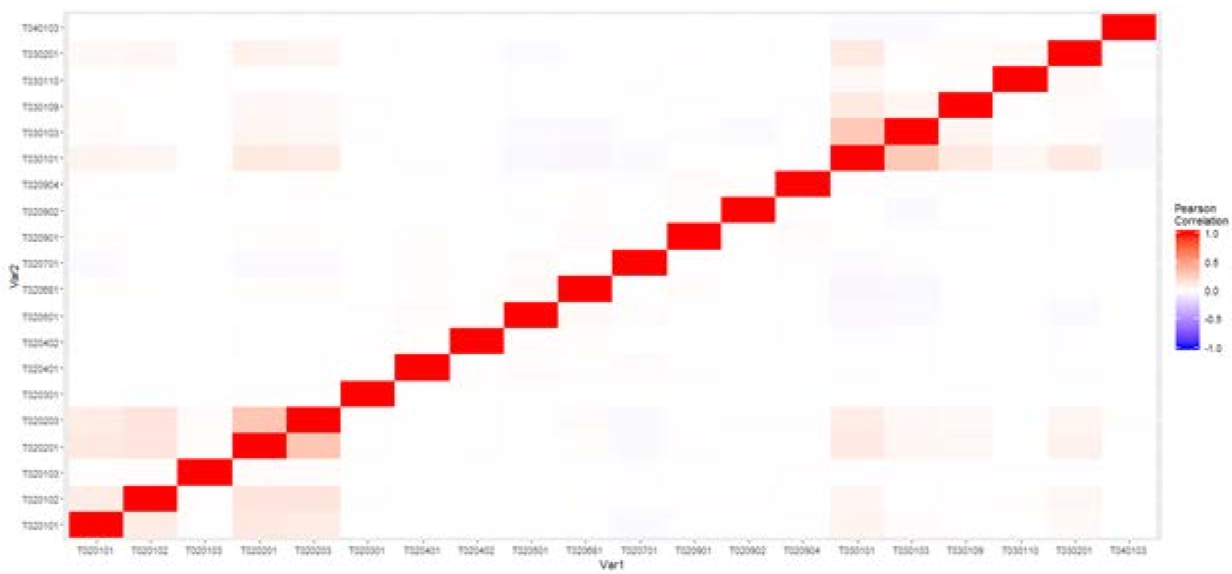

Figure 4

From the correlation matrix shown in Figure 4, we can see that none of the variables are strongly correlated to each other. Thus, multicollinearity will not be an issue once conducting a multiple linear regression to predict time slept. When running all 20 variables in a standard model buildout, I found that some were not statistically significant. Therefore I decided to use the backward selection model with an alpha of .05. The final model that passed all model assumptions had 16 of the 20 variables included. The model is shown in Figure 5.

```
Coefficients:
               Estimate Std. Error  t value Pr(>|t|)    
(Intercept) 532.629546   0.426364 1249.237  < 2e-16 ***
T020101       0.015830   0.005139    3.081 0.002066 ** 
T020102       0.031557   0.008736    3.612 0.000304 ***
T020103      -0.056735   0.014726   -3.853 0.000117 ***
T020203      -0.131445   0.016249   -8.090 6.03e-16 ***
T020301      -0.065746   0.009201   -7.145 9.01e-13 ***
T020401      -0.058828   0.016520   -3.561 0.000370 ***
T020402      -0.065895   0.013564   -4.858 1.19e-06 ***
T020501      -0.028923   0.006306   -4.586 4.51e-06 ***
T020681      -0.269951   0.014513  -18.600  < 2e-16 ***
T020901      -0.148809   0.019815   -7.510 5.94e-14 ***
T020902      -0.163160   0.010413  -15.668  < 2e-16 ***
T020904      -0.258373   0.024144  -10.702  < 2e-16 ***
T030101      -0.214707   0.009139  -23.494  < 2e-16 ***
T030103       0.021801   0.009307    2.342 0.019159 *  
T030110      -0.157945   0.017082   -9.246  < 2e-16 ***
T030201      -0.281113   0.024263  -11.586  < 2e-16 ***
T040103      -0.069533   0.018007   -3.862 0.000113 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 135.6 on 170824 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.01084,	Adjusted R-squared:  0.01074 
F-statistic: 110.1 on 17 and 170824 DF,  p-value: < 2.2e-16
```

Figure 5

The average model predicted that both men and women slept an average of 8.75 hours a night. In relation to the second shift phenomenon, it would be assumed that women are sleeping less hours than men. However, the data collected saw that men and women sleep nearly the exact amount of time each night (8.67 for men vs 8.80 for women). The reason for this is that women seem to record more minutes in the survey. For example, Table 5 shows the minutes breakdown on certain household chores and activities:

	Average Daily Minutes Spent		
Gender	Laundry	Interior Cleaning	Outdoor Cleaning
Female	18	36	9
Male	5	13	17
Grand Total	12	26	12

Table 5

The reason that this linear model isn't good is that we are comparing men and women separately without including that as non-continuous variable in the model. If this model was rerun, I would include a binary indicator for gender. That being said, since women are filling out more minutes than men, it ruins the concept of a linear model. Women will fill in more time for almost every category including minutes slept. This breaks the model comparison for gender since at the end of the day minutes slept is almost identical. The model itself yields a .02 R squared value, which is very low. Therefore, it might be best to include more non numeric variables in a multiple linear regression as a final model for time slept is not affected by gender.

Conclusion

The Second Shift Phenomenon is a concept that requires time and effort to break down factors, understand causation, and predict those who are demographically more inclined to partake in it. The study that we have conducted has helped us to understand this concept on an analytical level, which we then interpreted our results to group together the time activities that are the most relevant to those experiencing “Second Shifts”. Considering the complexity of this subject, we broke our analysis into two phases.

From phase one of our analysis, we successfully reduced the number of time activities using a multitude of techniques. Principal Component Analysis revealed time spent caring for children of the household, as well as children’s friends, and time spent performing chores in the household. Common Factor Analysis showed similar results to PCA, with the addition of sleeping, working, and helping the family’s wellbeing. We were able to use these reduction techniques to significantly reduce the number of time activities and focus of the ones that people spent the most consistent amount of time on. This contributes a greater look at what activities Americans are spending their time on during a potential “Second Shift” scenario. In addition, Canonical Correlation Analysis helped us to understand the relationships between how time is spent based on demographic status, and education levels. Lastly, we built a Linear Regression model to predict how someone is most likely to spend their time on a day to day basis based on a parameter of interest, such as sex, marriage or parental status and more.

These methods are only the start to answering the questions related to the Second Shift Phenomenon. Phase two of our study pertains to the ability to predict certain aspects of the second shift based on the variables hand-chosen in our reduction techniques. Answering these questions by creating predictive models will help us to better understand the corporate and societal cultures based on gender equity and social norms that contribute to the Second Shift Phenomenon, in order to analyze the lives and well beings of modern-day American women.

References

Gonzales, T. I. (2015, Feb 6). *The Second Shift and Workplace Policies.* Retrieved from Everyday Sociology Blog: http://www.everydaysociologyblog.com/2015/02/the-second-shift-and-workplace-policies.html

Hochschild, A. R. (1989). *The second shift : working parents and the revolution at home.* New York: Viking.

Appendix

time_survey_pca.R

Brandon

Wed Jun 06 20:04:38 2018

```
#Brandon Markwalder
#CSC 424 - Spring 2018
#Final project - American Time Survey
#Principal Compnent Analysis

library(car)
library(stats)
library(corrplot)
library(caret)
library(gridExtra)
library(psych)
library(tidyverse)
library(ggplot2)
library(GGally)
library(outliers)
library(nortest)

####
#Laod the data
####

setwd("C:/Users/Brandon/Documents/GitHub/csc424/project/raw_data_R")
load(file = "DS0001/36268-0001-Data.rda")
load(file = "DS0002/36268-0002-Data.rda")

#Create a dataframe for the weight functions for each variable
df2 <- (da36268.0002)
df2_head <- head(df2)

#Create traditional R dataframes with all variables and numeric only variable
s
```

```
df <- (da36268.0001)
df_head <- head(df)
df_numeric <- df[, c(26:456)]
df_head_numeric <- head(df_numeric)

#Create tidyverse tibble - the new fancy dataframe
tb <- as_tibble(da36268.0001)
tb_head <- head(tb)
tb_numeric <- tb[, c(26:456)]
tb_dv <- tb[, c(1:25)]
tb_dv_head <- head(tb_dv)

####
#Correlation Analysis
####

#Combine stuff
household_activites <- select(tb, starts_with("T02"), starts_with("T03"), sta
rts_with("T04"))
ha_head <- head(household_activites)
tb_pca <- household_activites
head <- head(tb_pca)
names(tb_pca)

#Check for na values
apply(tb_pca, 2, function(x) any(is.na(x)))

#The data is continuous and non-normal so we will use the spearman method
#for the correlation matrix
cor.df <- cor(tb_pca, method="spearman")

#Filter for correlation coefficients above 0.15, filtering for the 0.2 thresh
old
#exludes too much of the data
cors <- sort(findCorrelation(cor.df, cutoff = .15, verbose = FALSE, exact = T
RUE))
cors

#Create our new tibble from the reduced dataset computed above
tbr <- as_tibble(tb_pca[c(cors)])
names(tbr)

#Re-compute the correlation matrix and check for significance
cor.df <- cor(tbr, method="spearman")
sig <- cor.mtest(tbr, conf.level = .95, method="spearman")
corrplot(cor.df, p.mat = sig$p, sig.level = .05, addCoef.col = "black", numbe
r.cex=0.75)

#Compute significant correlation counts for each variable
options("scipen"=100, "digits"=5)
```

```
round(cor.df, 2)
MCorrTest = corr.test(tbr, adjust="none", method="spearman")
MCorrTest

M = MCorrTest$p
M

# Now, for each element, see if it is < .01 (or whatever significance) and se
t the entry to
# true = significant or false
MTest = ifelse(M < .05, T, F)
MTest

# Now lets see how many significant correlations there are for each variable.
We can do
# this by summing the columns of the matrix
colSums(MTest) - 1  # We have to subtract 1 for the diagonal elements (self-c
orrelation)

####
#Check Assumptions
####

stats <- as_tibble(describe(tbr))
ggplot(gather(tbr), aes(value)) +
  geom_histogram(bins = 10) +
  facet_wrap(~key, scales = 'free_x')

#The data is not normal and there are very large ranges present in some of th
e variables
#We do not want to scale because all of the observations have the same unit o
f measurement
#Instead let's create a filtered version of the data with a cutoff
#for any one activity at 4 hours in a single day
names(tbr)
tbr_f <- filter(tbr, T020102 < 240,
                T020201 < 240,
                T020203 < 240,
                T030101 < 240,
                T030102 < 240,
                T030112 < 240,
                T030186 < 240,
                T030201 < 240,
                T030302 < 240,
                T030401 < 240,
                T030404 < 240,
                T040101 < 240,
                T040103 < 240,
                T040302 < 240,
                T040401 < 240,
```

```
                T040404 < 240,
                T040507 < 240)

s_stats <- as_tibble(describe(tbr_f))
ggplot(gather(tbr_f), aes(value)) +
  geom_histogram(bins = 10) +
  facet_wrap(~key, scales = 'free_x')

#QQPlots for raw data
#tbr_p1  <- ggplot(tbr, aes(sample = T020102)) + stat_qq() + ggtitle("Quantil
e-Quantile Plot: T020102")
tbr_p2  <- ggplot(tbr, aes(sample = T020201)) + stat_qq() + ggtitle("Quantile
-Quantile Plot: T020201")
#tbr_p3  <- ggplot(tbr, aes(sample = T020203)) + stat_qq() + ggtitle("Quantil
e-Quantile Plot: T020203")
tbr_p4  <- ggplot(tbr, aes(sample = T030101)) + stat_qq() + ggtitle("Quantile
-Quantile Plot: T030101")
#tbr_p5  <- ggplot(tbr, aes(sample = T030102)) + stat_qq() + ggtitle("Quantil
e-Quantile Plot: T030102")
#tbr_p6  <- ggplot(tbr, aes(sample = T030112)) + stat_qq() + ggtitle("Quantil
e-Quantile Plot: T030112")
#tbr_p7  <- ggplot(tbr, aes(sample = T030186)) + stat_qq() + ggtitle("Quantil
e-Quantile Plot: T030186")
#tbr_p8  <- ggplot(tbr, aes(sample = T030201)) + stat_qq() + ggtitle("Quantil
e-Quantile Plot: T030201")
tbr_p9  <- ggplot(tbr, aes(sample = T030302)) + stat_qq() + ggtitle("Quantile
-Quantile Plot: T030302")
tbr_p10 <- ggplot(tbr, aes(sample = T030401)) + stat_qq() + ggtitle("Quantile
-Quantile Plot: T030401")
#tbr_p11 <- ggplot(tbr, aes(sample = T030404)) + stat_qq() + ggtitle("Quantil
e-Quantile Plot: T030404")
#tbr_p12 <- ggplot(tbr, aes(sample = T040101)) + stat_qq() + ggtitle("Quantil
e-Quantile Plot: T040101")
#tbr_p13 <- ggplot(tbr, aes(sample = T040103)) + stat_qq() + ggtitle("Quantil
e-Quantile Plot: T040103")
#tbr_p14 <- ggplot(tbr, aes(sample = T040302)) + stat_qq() + ggtitle("Quantil
e-Quantile Plot: T040302")
#tbr_p15 <- ggplot(tbr, aes(sample = T040401)) + stat_qq() + ggtitle("Quantil
e-Quantile Plot: T040401")
#tbr_p16 <- ggplot(tbr, aes(sample = T040404)) + stat_qq() + ggtitle("Quantil
e-Quantile Plot: T040404")
#tbr_p17 <- ggplot(tbr, aes(sample = T040507)) + stat_qq() + ggtitle("Quantil
e-Quantile Plot: T040507")

#Highest range QQ Plots
g1 <- grid.arrange(tbr_p2, tbr_p4, tbr_p9, tbr_p10, nrow = 2)
ggsave(file="tbr_qqplot.jpg", g2)

#The above plots are computationally expensive so here is a simple
#Printed addition to let me know when the g1 plot finally saved to disk
```

```r
print(11+22)

#QQPlots for filtered data
#tbr_f_p1  <- ggplot(tbr_f, aes(sample = T020102)) + stat_qq() + ggtitle("Quantile-Quantile Plot: T020102")
tbr_f_p2  <- ggplot(tbr_f, aes(sample = T020201)) + stat_qq() + ggtitle("Quantile-Quantile Plot: T020201")
#tbr_f_p3  <- ggplot(tbr_f, aes(sample = T020203)) + stat_qq() + ggtitle("Quantile-Quantile Plot: T020203")
tbr_f_p4  <- ggplot(tbr_f, aes(sample = T030101)) + stat_qq() + ggtitle("Quantile-Quantile Plot: T030101")
#tbr_f_p5  <- ggplot(tbr_f, aes(sample = T030102)) + stat_qq() + ggtitle("Quantile-Quantile Plot: T030102")
#tbr_f_p6  <- ggplot(tbr_f, aes(sample = T030112)) + stat_qq() + ggtitle("Quantile-Quantile Plot: T030112")
#tbr_f_p7  <- ggplot(tbr_f, aes(sample = T030186)) + stat_qq() + ggtitle("Quantile-Quantile Plot: T030186")
#tbr_f_p8  <- ggplot(tbr_f, aes(sample = T030201)) + stat_qq() + ggtitle("Quantile-Quantile Plot: T030201")
tbr_f_p9  <- ggplot(tbr_f, aes(sample = T030302)) + stat_qq() + ggtitle("Quantile-Quantile Plot: T030302")
tbr_f_p10 <- ggplot(tbr_f, aes(sample = T030401)) + stat_qq() + ggtitle("Quantile-Quantile Plot: T030401")
#tbr_f_p11 <- ggplot(tbr_f, aes(sample = T030404)) + stat_qq() + ggtitle("Quantile-Quantile Plot: T030404")
#tbr_f_p12 <- ggplot(tbr_f, aes(sample = T040101)) + stat_qq() + ggtitle("Quantile-Quantile Plot: T040101")
#tbr_f_p13 <- ggplot(tbr_f, aes(sample = T040103)) + stat_qq() + ggtitle("Quantile-Quantile Plot: T040103")
#tbr_f_p14 <- ggplot(tbr_f, aes(sample = T040302)) + stat_qq() + ggtitle("Quantile-Quantile Plot: T040302")
#tbr_f_p15 <- ggplot(tbr_f, aes(sample = T040401)) + stat_qq() + ggtitle("Quantile-Quantile Plot: T040401")
#tbr_f_p16 <- ggplot(tbr_f, aes(sample = T040404)) + stat_qq() + ggtitle("Quantile-Quantile Plot: T040404")
#tbr_f_p17 <- ggplot(tbr_f, aes(sample = T040507)) + stat_qq() + ggtitle("Quantile-Quantile Plot: T040507")

#Highest range QQ Plots
g2 <- grid.arrange(tbr_f_p2, tbr_f_p4, tbr_f_p9, tbr_f_p10, nrow = 2)
ggsave(file="tbr_f_qqplot.jpg", g2)

#The above plots are computationally expensive so here is a simple
#Printed addition to let me know when the g1 plot finally saved to disk
print(03+23)

#Assumption testing/checking playground
#chisq.out.test(tbr_s$T020102)
#rm.outlier(tbr_s)
```

```r
#ad.test(tbr$T020102)
#chisq.out.test(tbr$T020102)

#bartlett.test(dfr$T030102)

#KMO(tbr)
#cortest.bartlett(cor.df)

####
#PCA Analysis
####

#?prcomp
p <- prcomp(tbr)
print(p)
summary(p)
plot(p, xlab="components", main="Household Activities - No Time Cap")

#?principal #Search principal help to try the different rotations.
p2 = principal(tbr, rotate="varimax", nfactors=3, scores=TRUE)
p2
summary(p2)
print(p2$loadings, cutoff=.3, sort=T)
p2$loadings
p2$values
p2$communality
p2$scores
p2$rot.mat

KMO(tbr_f)


p_f <- prcomp(tbr_f)
print(p)
summary(p_f)
plot(p_f, xlab="components", main="Household Activities - 4 Hour Time Cap")

p2_f = principal(tbr_f, rotate="varimax", nfactors=3, scores=TRUE)
p2_f
print(p2_f$loadings, cutoff=.4, sort=T)
p2_f$loadings
p2_f$values
p2_f$scores
p2_F$rot.mat
```