The Second Shift Phenomenon

CSC 424 Team Members: Brandon Markwalder, Fiona Baenziger, Omar Ahram, and Steven Younan

Project Objective & Goal –

Exploring the Second Shift Phenomenon

Dataset Overview
American Time Use Survey, 2003-2015

- Comprehensive time survey conducted between 2003 – 2015
- Purpose of the study was to develop a national view of how Americans spend their time
- 170,842 respondents
- Response rate: 54.2%

- 17 Time modules
- 431 continuous variables
- 25 Categorical – Demographic data

https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/36268

Research Question

- How do women who work all day, only to come home and continue working the “Second Shift” caring for their home and families, spend their time?

Methodologies

- Documentation review
- General data exploration
- Data cleaning
- Principal Component Analysis
- Common Factor Analysis
- Correspondence Analysis
- Linear Regression

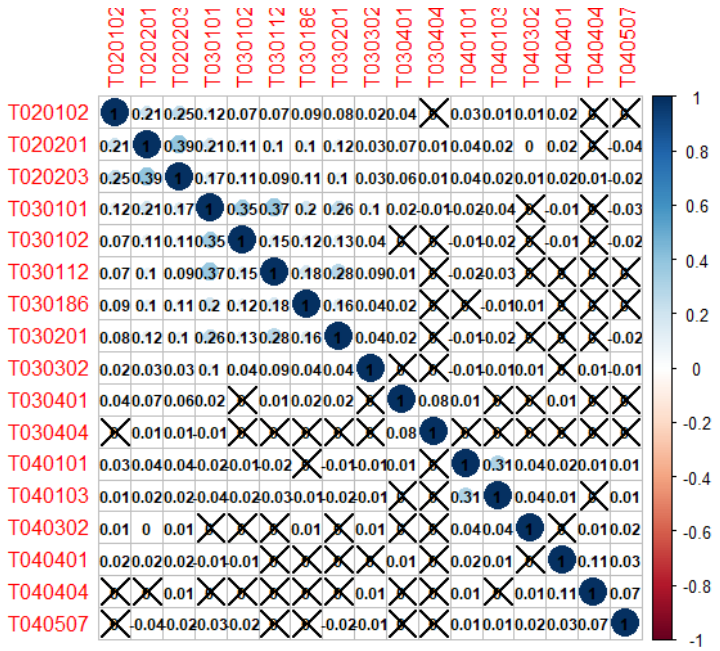
PCA – Why do it

- 431 continuous variables between 17 time modules
- Time modules contain as few as 5 to as many as 77 variables
- Modules include time spent caring for self and family, doing chores, advancing education, exercise, leisure, and work
- All could be important to the research question

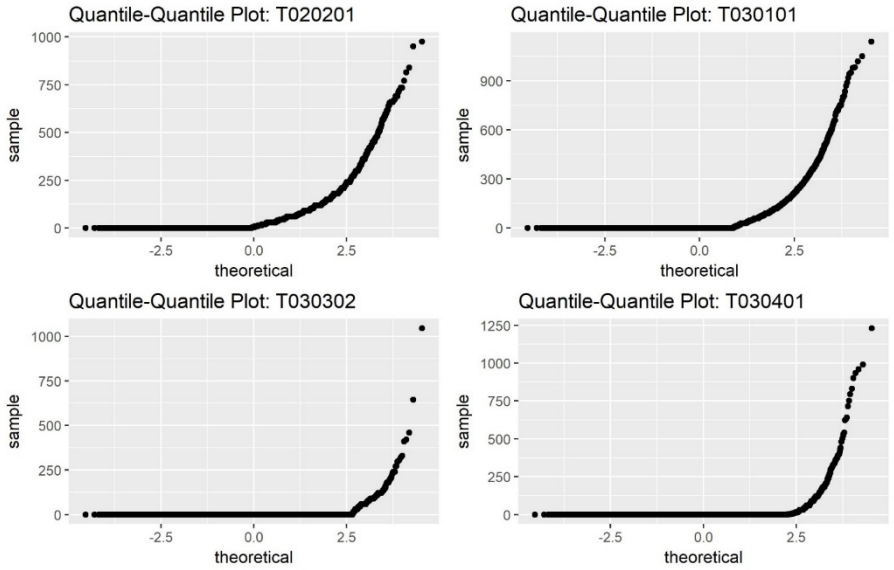
- Method: Principal Component Analysis
- Reduce variables down to easy to consume components for use in future analysis
- Modules of focus: Household Activities, Caring for Household Members, and Caring for Non-household Members – 102 variables
- Reduced to 3 components capable of explaining 77% of the data, but…

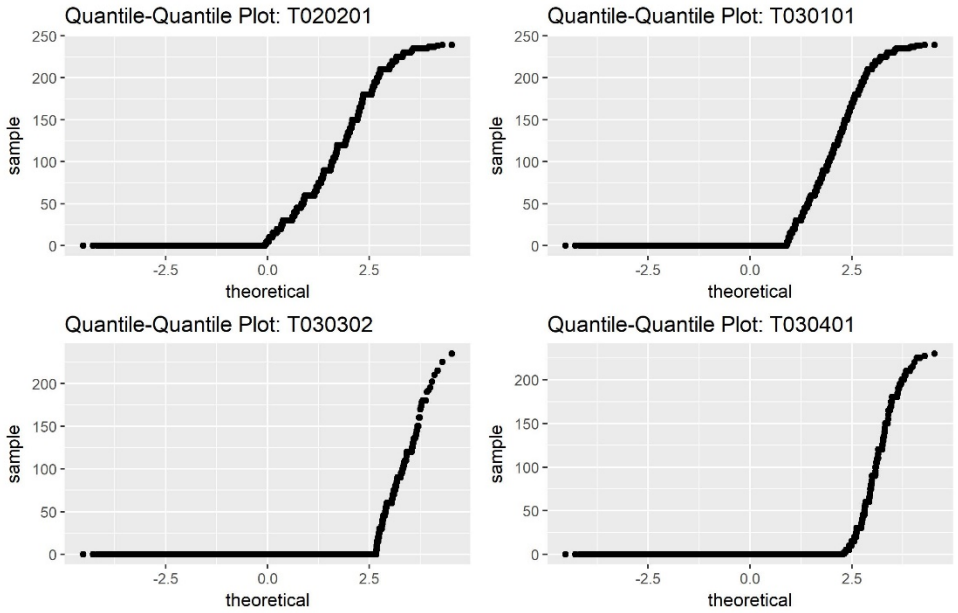
PCA – Assumptions

- What about the assumptions?
- Correlated data – The data is sparse.
- Normality – Time measured in minutes. The variance in the data is important to our research question, but the range is huge.
- Outliers – Laundry for 13 hours!



PCA – Wide Ranging Non-normal Data – A Tale of Two Models





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

PCA – Final Components

- PC1: Child Care (0.646 * T030101) + (0.507 * T030102) + (0.621 * T030112) + (0.489 * T030201)
- PC2: Chores (0.507 * T020102) + (0.712 * T020201) + (0.731 * T020203)
- PC3: Child Friendship Care (0.732 * T040101) + (0.682 * T040103)

	3 Component Solution		
Household Activities	Child Care	Chroes	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			

Common Factor Analysis

- Objective: To minimize the number of variables in our exploration of the Second Shift Phenomenon and determine correlated activities

- Each variable has a numeric value representing amount of time spent on a certain activity

- Modules Included:

  M01: Personal Care Activities

  M02: Household Activities

  M03: Caring For & Helping Household (HH) Members

  M04: Caring For & Helping Non-Household (NonHH) Members

  M05: Work & Work Related Activities

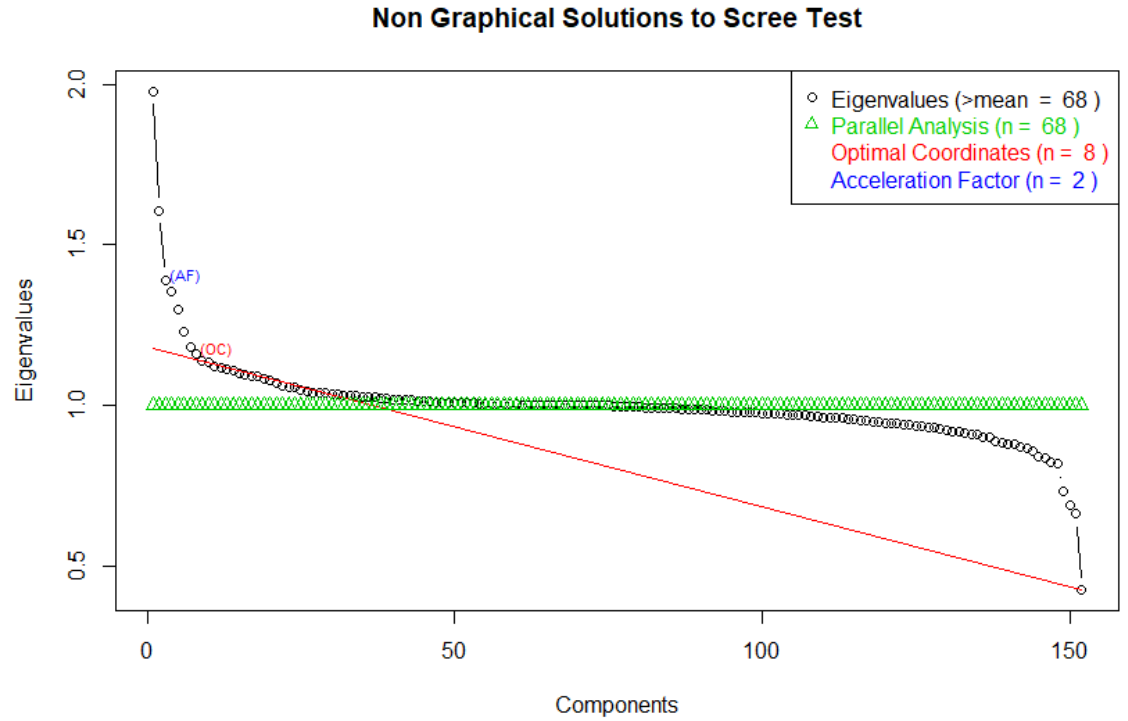
  M06: Education

CFA – The Process

- Initial Issues:

  Standard deviation of 0 = singular matrix → narrowed down the modules

- Computed eigenvalues to create a scree plot to determine ideal number of factors



CFA – Results

Factor 1: Sleep v Work

Factor 2: Household Chores

Factor 3: Children

Factor 4: Education

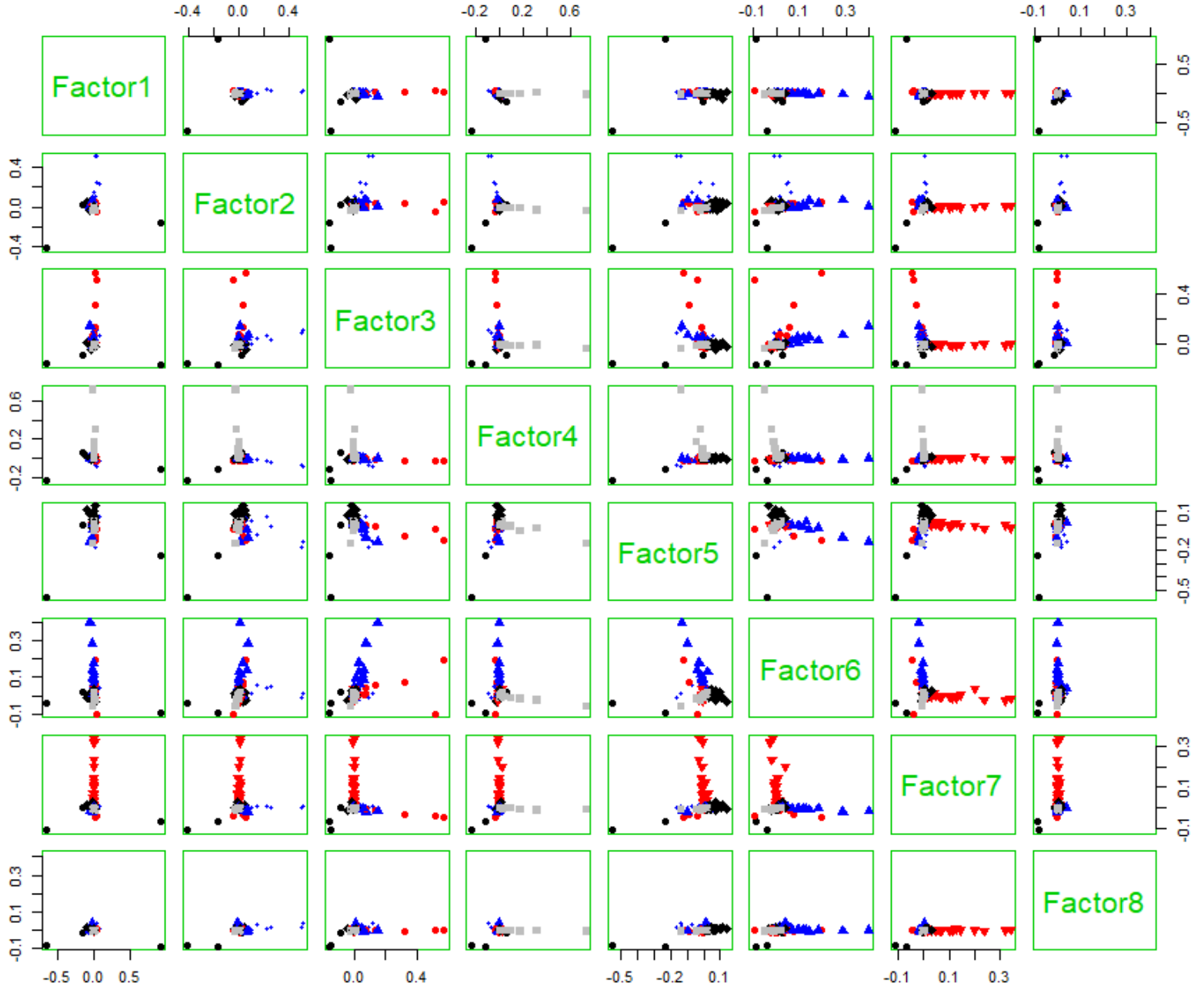
Factor 5: Working guardians, furthering their education

Factor 6: Good, attentive parental units

Factor 7:Teachers

Factor 8: Good family values

Future: Narrow down modules to focus solely on the relationships between more relevant variables.



CCA

- Method: Canonical Correlation Analysis
- Objective: Exploratory tool to test relation between two datasets. Looking to be able to describe time spent during second shift.
- Variables
  - Household Activities
    - Ex. Interior Cleaning, Laundry, and Food Prep
  - Caring for & Helping Household members
    - Ex. Time spent with children, attending children's events, and playing with children
  - Education
    - Ex. Time spent in class, on homework, and on research

CCA

Results:

- Household Activities
  - Busy with a lot of work hours => low time spent on cleaning, laundry, and food preparation
  - Educated with young children => more time spent on personal email and messaging, and on food/drink preparation
- Caring for & Helping Household members
  - Has young children and a busy spouse => most time spent caring for children
  - Both spouses working => focuses on playing with children and attending events
- Education
  - Older and Uneducated => less time spent on education and homework
  - Educated w/o children => more time spent on research and homework

Multiple Regression

- Second Shift Phenomenon and Minutes Slept
- A second shift would cause women to sleep less
- Variables Selected
  - Gender
  - Minutes Slept
  - Household Activities
  - Caring for Household Members
  - Caring for non-Household Members

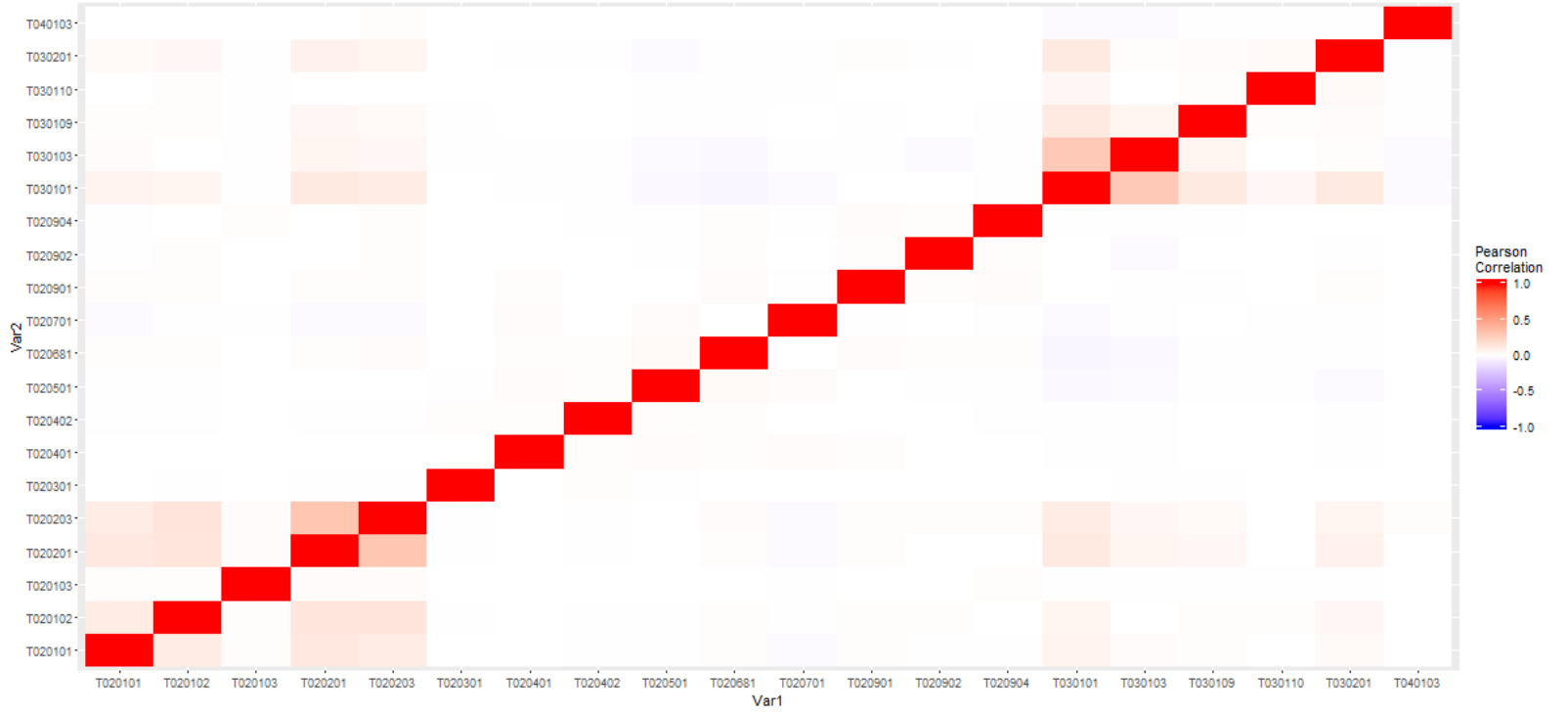
- Total of 102 surveyed questions
- Reduced to 20 variables
  - Issues of multicollinearity
  - Low impact (many zero values)
- Data is classified as Male or Female

Gender	Average Minutes Slept
Female	528
Male	520
Grand Total	525

Multiple Regression

- Correlation Matrix Heatmap
  - No Multicolinearity
- Backward Selection Regression
- 16 of 20 measurements included
- Very low R-Squared (.02)
- No way to increase by changing model buildout
- Multiple Regression not a good model for this study

- Correlation Heatmap



Multiple Regression

- All predictors are statistically significant
- F-test rejects null hypothesis
- All assumptions met
- Not enough variance is explained
- Predicted values were added for analysis

```
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
```

Multiple Regression

- Fitted model predicts 8.75 hours (525 minutes) slept a day
- Reasons Model did not work
  - Men/Women Sleep about the same time each night
  - Women have more productive non sleeping minutes, but also sleep more on average
  - Sleep is not a good response variable

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

Conclusion

- How do women who work all day, only to come home and continue working the “Second Shift” caring for their home and families, spend their time?
- We did not get quite that far.
- As a team we developed a deep understanding of the data.
- We carefully applied techniques that sum to robust data exploration.
- The results of which will lead to more manageable approaches to answering our question.
- PCA and CFA = independent variables.
- CCA = independent and dependent variables.
- Linear Regression = Confirmation that predictive models are feasible, if we are careful!
- Our group’s work adds up to simpler representations of the data to use in future predictive models – That may answer the Second Shift Phenomenon.