Activities We Give Up When We're Online

Table of Contents

Overview

The goal of this research is to measure the opportunity cost of online leisure activity. Inspired by Wallsten's research, we will be using a similar method to calculate the opportunity cost for the years 2011-2015. In other words, we will be measuring how online leisure activty crowds out other activities.

Wallsten uses the ATUS (American Time Survey) dataset for the years 2003-2011. He estimates 18 versions of equations (one for each major activity and one for an unknown category), and uses the coefficient (and t-statistic) on the computer leisure variable from each regression as a measurement of the crowd-out effect of computer leisure on the major category (the major category being the dependent variable in each regression).

The challenge in the ATUS dataset is that it does not explicitely include a variable indicating whether a respondent has access to the internet or not. To overcome this obstacle, Wallsten followed Goldfarb and Prince's methodology to estimate this variable, using two-stage least squares regression.

His method implied that only 17 percent of households had access to the internet in 2010, when the US Census (based on the Current Population Survey) estimated that more than 70 percent actually had access. For that reason, we will be taking a different approach for estimating the internet access variable on the ATUS dataset. The ATUS dataset is in fact a subsample of a larger dataset: CPS (Current Population Survey). Unlike the ATUS dataset, the CPS dataset includes a variable that indicates whether a subject has internet access or not. As such, we will be using similar type of variables as the one used in Wallsten's regression, and will look at the common variables found in both the ATUS and CPS datasets in order to construct a decision tree algorithm, classifiying our records into: subject has internet access or subject does not have internet access.

After we construct our algorithm and apply it to the ATUS dataset, we will compare the percentage of households estimated to have internet access with the actual percentage of the population using the internet.

Then, we select a subsample of the ATUS dataset based on the portion of households that are estimated to have Internet access, and we will run the regression analysis to measure the crowd out effect of computer leisure on the 17 major activities.

Finally, we compare the crowd effect measurements from year 2003-2011 (based on Wallsten's findings), and those from years 2011-2015 (based on our findings), and present our conclusions.

Literature Review

1. Carver, Jeffrey C. "Towards reporting guidelines for experimental replications: A proposal." *1st International Workshop on Replication in Empirical Software Engineering*. 2010.
   An explanation on how to replicate a paper.

2. Goldfarb, Avi, and Jeff Prince. "Internet adoption and usage patterns are different: Implications for the digital divide." *Information Economics and Policy* 20.1 (2008): 2-15.
   The data for this study come from a detailed survey of technology choices conducted by Forrester Research. The data set is a random sub-sample of the Forrester data and contains 18,439 American household respondents, collected in December 2001. They estimate usage and adoption using a Type-II Tobit regression.

3. Wallsten, Scott. *What are we not doing when we're online*. No. w19549. National Bureau of Economic Research, 2013.
   A detailed explanation of this paper is provided below. In summary, Wallsten found that 1 minute of online leisure activity translates into:
   - 0.29 fewer minutes spent on all other types of leisure:
     - 0.145 coming from time spent watching TV and video
     - 0.05 coming from on offline socializing
     - 0.04 coming from relaxing and thinking
     - 0.055 coming from attending parties, cultural events and listening to the radio
   - 0.27 fewer minutes spent working
   - 0.12 fewer minutes spent on sleeping
   - 0.10 fewer minutes spent in travel time
   - 0.07 fewer minutes in household activities
   - 0.06 fewer minutes in educational activities
   - 0.08 fewer minutes spent in sports, helping other people, eating and drinking and leisure activities

Dataset

American Time Use Survey (ATUS)

The American Time Use Survey interviews respondents about how they spent their time on the previous day, where they were, and whom they were with. The goal is to measure how people divide their time among life's activities. Individuals are randomly selected from a subset of households previously interviewed in the Current Population Survey (CPS).

- Link: **http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/36268?timePeriodFrom=2010&timePeriodTo=2017&sortBy=&searchSource=revise&q=time+use**

Documentation Used:

- American Time Use Survey technical documentation: add pdf.
- American Time Use Survey user guide: add pdf.

Current Population Survey: Computer and Internet Use Supplement

The Current Population Survey (CPS) interviews around 56,000 households monthly, scientifically selected on the basis of area of residence to represent the nation as a whole, individual states, and other specified areas. The main purpose of the survey is to collect information on the employment situation, as well as other information on demographic characteristics such as age, sex, race, marital status, educational attainment, family relationship, occupation and industry etc.

- Link: **http://thedataweb.rm.census.gov/ftp/cps_ftp.html#cpssupps**

Documentation Used:

- Current Population Use Survey July 2011 technical documentation.
- Current Population Use Survey July 2011 user guide.

- Current Population Use Survey July 2013 technical documentation.
- Current Population Use Survey July 2013 user guide.

- Current Population Use Survey July 2015 technical documentation.
- Current Population Use Survey July 2015 user guide.

Phase 1: Data Formatting & Data Cleaning

In Step 1, we will clean and format both our datasets: CPS and ATUS, in order to have a consistent data format in both datasets.

The ATUS dataset is already in R format. However, the CPS dataset has to be read as a delim file. Below, we will read our CPS datasets.

Creating CPS_prep dataframe

As shown, the dataset is read in a single column dataframe, and includes 1259, 1173 and 1174 characters for years 2011, 2013 and 2015 respectively. Each row represents the answers of one respondent to the survey, and each character a response to a question (In 2011 for example, 1259 questions were asked to each subject during the survey, as per the CPS codebook guide).

```
CPS_2011 <- read.delim("july2011cps.dat", header = FALSE, sep="\t")
CPS_2013 <- read.delim("july2013cps.dat", header = FALSE, sep="\t")
CPS_2015 <- read.delim("july2015cps.dat", header = FALSE, sep="\t")

CPS_2015 <- as.data.frame(sapply(CPS_2015[,1], as.character))
CPS_2013 <- as.data.frame(sapply(CPS_2013[,1], as.character))
CPS_2011 <- as.data.frame(sapply(CPS_2011[,1], as.character))

colnames(CPS_2011) <- "RAW_DATA"
colnames(CPS_2013) <- "RAW_DATA"
colnames(CPS_2015) <- "RAW_DATA"

CPS_2011[1,]

## [1] 138009000100198 72011 220100 1 1 1-1 115-1-1-1  38893409 1 2 1 2 2 0 2
89001 2  2-1-1-1-1 36301 115000154102000     2 1-1540 1 2 1 2   244 1-1 9
1   1 1 1 2 1 2 57 57 57 1 0 0 2 1 1-1-1-1-1-1-1-1-1-1-1-1-1-1-1 2-124-1-1
24 2-1 2-1 2-1 11640-1 40-1      2 5 4 3 2 3 5 2-1-1-1-1-1 -1-1-1-1-1-1-1-1-1
-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1 -1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1
-1-1-1-1 1-118 1 1 2 2-1-118 -1  -1 1 9-1-1-1-1-1-1-1 4 2      -1-1      -1-1
-1 1 2 4-1 1 6-132-1 2-1 1 8-1 1-1 1 1 1 1 0-1-1-1-1  -1  -1  -1  -10-1
-10          -1      -10  -1-1-1-1-1-1-1-1-1-1-1  38893409  52811218
  0  37458457  36305764 0 0  -1-1 0 0 0 1 0-1 050 0 0 0 0 1 0 0 0-1-1-1   0 0
-1 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-
1 0 0 0-1-1-1-1-1-1-1-1    -1-1-1-1-1-1-1-1-1-1-1-1    1 1-1-1-1-1-1 2 1 1 1
1 1 0  371544046870 850  -1  -111-1-1-1-1-1 0-1-1-1-1-1 1 1 1 1 1 2 2 2 2 2 2
 2 0 0 0 0 0 0 0                 -1-1-1-1-1-1-1-1-1-1-1 22010 1-1 2 2 1 2 2 2
 2-1-1-1-1-1-1 1 1 1 1 2150-1 1 2 2-1-1-1-1 1 1 2 1-1-1 2-1 1 1 2 1 2 2-1-1-1
-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1
0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

```
 0 0 0 0 0 0 0 0 0 0 0 0         0
## 152260 Levels: 000011141520290 72011 120100 1 1 1-1 116-1-1-1  28390832 1
2 1 3 2 0 289001 1  2-1-1-1-1 35451 000000003200000      2 1-1590 1 2 2 1   2
40 1-1 9   1   1 1 1 2 1 2 57 57 57 1 0 0 1 1 1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1
 2-156-1-1 56-1-1-1-1 2-1 2-156-1 56-1      2 5 5-1 2 3 5 2-1-1-1-1-1 -1-1-1-
1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1 -1-1-1-1-1-1-1-1-1-1-1-1-
1-1-1-1-1-1-1-1-1-1 1-121 1 1 1 6-1-1-1 -1  -1 1 2-1-1-1-1 1 2 1 4 1      -1-
1       4 3 3 1 2 4-1 1 6-119-121-1 1 4-1 9-1 6 1 1 1 0-1-1-1-1  -1  -1  -1
-10-1      -10          -1      -10  -1-1-1-1-1-1-1-1-1-1-1-1  28390832  3994
5191         0  28390832  34270334 0 0  -1-1 0 0 0 1 0-1 050 0 0 0 0 0 0 0 0-
1-1-1   0 0-1 1 1 0 1 0 1 1 0 1 1 1 0 1 0 1 1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1
-1-1-1-1-1-1 0 0 0-1-1-1-1-1-1-1-1    -1-1-1-1-1-1-1-1-1-1-114913 1-1-1 3-1
-1-1 1 1 0 1 1 1  2811415521908965  -1  -1 5-1 3-1-1-1 0-1-1-1-1-1 1 1 1 1 1
2 2 2 2 2 2 2 0 0 0 0 0 0 0                  1 1 1 1 1 1 2 2 2 2 1 32010 1-1
2 1 2 2 2 2 2-1-1-1-1-1-1 1 2 2 1 2 40-1 1 7 1-1-1-1-1 1 1 2 1-1-1 2-1 1 1 1
1 2 2 1 0 2 2 1 2 2 1 2 1 1 1 2 2 2 2 2 2 2 2 3-1-1-1-1-1-1-1 1 1 1 1 1 2 3 2
 1 2 2 2 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  44940687 ...
```

```r
first_row_2011 <- CPS_2011[1,1]
first_row_2011 <- as.character(first_row_2011)
nchar(first_row_2011)
```

```
## [1] 1259
```

```r
first_row_2013 <- CPS_2013[1,1]
first_row_2013 <- as.character(first_row_2013)
nchar(first_row_2013)
```

```
## [1] 1173
```

```r
first_row_2015 <- CPS_2015[1,1]
first_row_2015 <- as.character(first_row_2015)
nchar(first_row_2015)
```

```
## [1] 1174
```

Now we want to structure our 2011, 2013 and 2015 dataset into dataframes, selecting the variables required. The variable selections have been qualitatevely selected based on the similar variables found in the CPS and ATUS datasets. The location of each variable is indicated in the CPS codebooks.

```r
#2011 Dataset:
CPS_2011$PERSON_TYPE <- substr(CPS_2011$RAW_DATA, 161,162)
CPS_2011$LABOUR_FORCE_STATUS <- substr(CPS_2011$RAW_DATA, 180,181)
CPS_2011$AGE <- substr(CPS_2011$RAW_DATA, 122,123)
CPS_2011$HISPANIC <- substr(CPS_2011$RAW_DATA, 157,158)
```

```
CPS_2011$SEX <- substr(CPS_2011$RAW_DATA, 129,130)
CPS_2011$MORE_THAN_1_JOB <- substr(CPS_2011$RAW_DATA, 214,215)
CPS_2011$HOURS_PER_WEEK <- substr(CPS_2011$RAW_DATA, 224,226)
CPS_2011$FULL_TIME_PART_TIME <- substr(CPS_2011$RAW_DATA, 2,3)
CPS_2011$WEEKLY_EARNINGS <- substr(CPS_2011$RAW_DATA, 527,534)
CPS_2011$EDUCATION <- substr(CPS_2011$RAW_DATA, 137,138)
CPS_2011$SPOUSE <- substr(CPS_2011$RAW_DATA, 125,126)
CPS_2011$CHILDREN <- substr(CPS_2011$RAW_DATA, 635,636)
CPS_2011$AGE_YOUNGEST_CHILD <- substr(CPS_2011$RAW_DATA, 633,634)
CPS_2011$METROPOLITAN_STATUS <- substr(CPS_2011$RAW_DATA, 105,105)
CPS_2011$HOME_INTERNET_ACCESS <- substr(CPS_2011$RAW_DATA, 979,980)

#2013 Dataset:
CPS_2013$PERSON_TYPE <- substr(CPS_2013$RAW_DATA, 161,162)
CPS_2013$LABOUR_FORCE_STATUS <- substr(CPS_2013$RAW_DATA, 180,181)
CPS_2013$AGE <- substr(CPS_2013$RAW_DATA, 122,123)
CPS_2013$HISPANIC <- substr(CPS_2013$RAW_DATA, 157,158)
CPS_2013$SEX <- substr(CPS_2013$RAW_DATA, 129,130)
CPS_2013$MORE_THAN_1_JOB <- substr(CPS_2013$RAW_DATA, 214,215)
CPS_2013$HOURS_PER_WEEK <- substr(CPS_2013$RAW_DATA, 224,226)
CPS_2013$FULL_TIME_PART_TIME <- substr(CPS_2013$RAW_DATA, 2,3)
CPS_2013$WEEKLY_EARNINGS <- substr(CPS_2013$RAW_DATA, 527,534)
CPS_2013$EDUCATION <- substr(CPS_2013$RAW_DATA, 137,138)
CPS_2013$SPOUSE <- substr(CPS_2013$RAW_DATA, 125,126)
CPS_2013$CHILDREN <- substr(CPS_2013$RAW_DATA, 635,636)
CPS_2013$AGE_YOUNGEST_CHILD <- substr(CPS_2013$RAW_DATA, 633,634)
CPS_2013$METROPOLITAN_STATUS <- substr(CPS_2013$RAW_DATA, 105,105)
CPS_2013$HOME_INTERNET_ACCESS <- substr(CPS_2013$RAW_DATA, 977,978)

#2015 Dataset:
CPS_2015$PERSON_TYPE <- substr(CPS_2015$RAW_DATA, 161,162)
CPS_2015$LABOUR_FORCE_STATUS <- substr(CPS_2015$RAW_DATA, 180,181)
CPS_2015$AGE <- substr(CPS_2015$RAW_DATA, 122,123)
CPS_2015$HISPANIC <- substr(CPS_2015$RAW_DATA, 157,158)
CPS_2015$SEX <- substr(CPS_2015$RAW_DATA, 129,130)
CPS_2015$MORE_THAN_1_JOB <- substr(CPS_2015$RAW_DATA, 214,215)
CPS_2015$HOURS_PER_WEEK <- substr(CPS_2015$RAW_DATA, 224,226)
CPS_2015$FULL_TIME_PART_TIME <- substr(CPS_2015$RAW_DATA, 2,3)
CPS_2015$WEEKLY_EARNINGS <- substr(CPS_2015$RAW_DATA, 527,534)
CPS_2015$EDUCATION <- substr(CPS_2015$RAW_DATA, 137,138)
CPS_2015$SPOUSE <- substr(CPS_2015$RAW_DATA, 125,126)
CPS_2015$CHILDREN <- substr(CPS_2015$RAW_DATA, 635,636)
CPS_2015$AGE_YOUNGEST_CHILD <- substr(CPS_2015$RAW_DATA, 633,634)
```

```
CPS_2015$METROPOLITAN_STATUS <- substr(CPS_2015$RAW_DATA, 105,105)
CPS_2015$HOME_INTERNET_ACCESS <- substr(CPS_2015$RAW_DATA, 975,976)
```

Then, we will set the data type to numeric (for easier data manipulation), combine the datasets, and view our generated CPS_prep dataset:

```
CPS_2015 <- as.data.frame(sapply(CPS_2015[,1:16], as.numeric))
CPS_2013 <- as.data.frame(sapply(CPS_2013[,1:16], as.numeric))
CPS_2011 <- as.data.frame(sapply(CPS_2011[,1:16], as.numeric))
CPS_prep <- rbind(CPS_2011, CPS_2013, CPS_2015)

CPS_prep[1:3,]

##   RAW_DATA PERSON_TYPE LABOUR_FORCE_STATUS AGE HISPANIC SEX
## 1    31793           2                   1  54        2   1
## 2    31794           2                   1  55        2   2
## 3    65285          -1                  -1  -1       -1  -1
##   MORE_THAN_1_JOB HOURS_PER_WEEK FULL_TIME_PART_TIME WEEKLY_EARNINGS
## 1               2             24                  38              -1
## 2               2             24                  38              -1
## 3              -1             -1                   0              -1
##   EDUCATION SPOUSE CHILDREN AGE_YOUNGEST_CHILD METROPOLITAN_STATUS
## 1        44      1        0                  0                   1
## 2        45      1        0                  0                   1
## 3        -1     -1       -1                 -1                   1
##   HOME_INTERNET_ACCESS
## 1                    1
## 2                    1
## 3                   -1
```

To reduce our data, we will drop the first column RAW data from the CPS dataframe.

```
CPS_prep[,1] <- NULL
CPS_prep[1:3,]

##   PERSON_TYPE LABOUR_FORCE_STATUS AGE HISPANIC SEX MORE_THAN_1_JOB
## 1           2                   1  54        2   1               2
## 2           2                   1  55        2   2               2
## 3          -1                  -1  -1       -1  -1              -1
##   HOURS_PER_WEEK FULL_TIME_PART_TIME WEEKLY_EARNINGS EDUCATION SPOUSE
## 1             24                  38              -1        44      1
## 2             24                  38              -1        45      1
## 3             -1                   0              -1        -1     -1
##   CHILDREN AGE_YOUNGEST_CHILD METROPOLITAN_STATUS HOME_INTERNET_ACCESS
## 1        0                  0                   1                    1
```

```
## 2        0                  0                   1                    1
## 3       -1                 -1                   1                   -1
```

Creating ATUS_prep and ATUS_time dataframes

In the section below, we will load the ATUS dataset, select all records from years 2011-105, and then select the variables we require.

ATUS_prep contains the same variables found in CPS dataset. The model (decision tree derived from CPS) will be applied to the ATUS_prep dataset.

ATUS_prep includes all demographic and geographic variables of subjects. ATUS_time includes the time each subject spent on various activities.

```
load("/Users/daliashanshal/Desktop/Capstone_Project/ATUS R Format/DS0001/3626
8-0001-Data.rda")

ATUS_years_select <- da36268.0001[which(da36268.0001$TUYEAR>=2011),]

Variables <- c(9, 8, 6, 13, 10, 24, 16, 17,5,20,15,21, 4)

ATUS_prep <- ATUS_years_select[, Variables]

names(ATUS_prep)[names(ATUS_prep)=="TELFS"] <- "LABOUR_FORCE_STATUS"
names(ATUS_prep)[names(ATUS_prep)=="TEAGE"] <- "AGE"
names(ATUS_prep)[names(ATUS_prep)=="PEHSPNON"] <- "HISPANIC"
names(ATUS_prep)[names(ATUS_prep)=="TESEX"] <- "SEX"
names(ATUS_prep)[names(ATUS_prep)=="TEMJOT"] <- "MORE_THAN_1_JOB"
names(ATUS_prep)[names(ATUS_prep)=="TEHRUSLT"] <- "HOURS_PER_WEEK"
names(ATUS_prep)[names(ATUS_prep)=="TRDPFTPT"] <- "FULL_TIME_PART_TIME"
names(ATUS_prep)[names(ATUS_prep)=="TRERNWA"] <- "WEEKLY_EARNINGS"
names(ATUS_prep)[names(ATUS_prep)=="PEEDUCA"] <- "EDUCATION"
names(ATUS_prep)[names(ATUS_prep)=="TRSPPRES"] <- "SPOUSE"
names(ATUS_prep)[names(ATUS_prep)=="TRCHILDNUM"] <- "CHILDREN"
names(ATUS_prep)[names(ATUS_prep)=="TRYHHCHILD"] <- "AGE_YOUNGEST_CHILD"
names(ATUS_prep)[names(ATUS_prep)=="GTMETSTA"] <- "METROPOLITAN_STATUS"
```

ATUS_time contains the variables required for the last phase of this project. It includes the time spent on each of the major activities. For the first variable: Leisure (excluding computer), we want to substract the subcategory 'computer use for leisure', for other subcategories, we have to manually add the subcategories together.

```
#Leisure
which( colnames(ATUS_years_select)=="T120101")

## [1] 250
```

```
which( colnames(ATUS_years_select)=="T120308")

## [1] 262

Leisure_Excl_Computer <- (ATUS_years_select$T120101 - ATUS_years_select$T1203
08)

#Personal Care Including Sleep
Variable_Care_Sleep <- (c(which( colnames(ATUS_years_select)=="T010101"), whi
ch( colnames(ATUS_years_select)=="T010102"), which( colnames(ATUS_years_selec
t)=="T010199"), which( colnames(ATUS_years_select)=="T010201"), which( colnam
es(ATUS_years_select)=="T010299"), which( colnames(ATUS_years_select)=="T0103
01"), which( colnames(ATUS_years_select)=="T010399"), which( colnames(ATUS_ye
ars_select)=="T010401"), which( colnames(ATUS_years_select)=="T010499"), whic
h( colnames(ATUS_years_select)=="T010501"), which( colnames(ATUS_years_select
)=="T010599"), which( colnames(ATUS_years_select)=="T019999")))
Personal_Care_Sleep_Pre <- ATUS_years_select[,Variable_Care_Sleep]
Personal_Care_Sleep_Pre$Personal_Care_Sleep <- rowSums(Personal_Care_Sleep_Pr
e[1:12])
Personal_Care_Sleep <- data.frame(Personal_Care_Sleep_Pre$Personal_Care_Sleep
)

# Work Activities
Variable_Work <- (c(which( colnames(ATUS_years_select)=="T050101"), which( co
lnames(ATUS_years_select)=="T050102"), which( colnames(ATUS_years_select)=="T
050103"), which( colnames(ATUS_years_select)=="T050189"), which( colnames(ATU
S_years_select)=="T050201"), which( colnames(ATUS_years_select)=="T050202"),
which( colnames(ATUS_years_select)=="T050203"), which( colnames(ATUS_years_se
lect)=="T050204"), which( colnames(ATUS_years_select)=="T050289"), which( col
names(ATUS_years_select)=="T050301"), which( colnames(ATUS_years_select)=="T0
50302"), which( colnames(ATUS_years_select)=="T05030"), which( colnames(ATUS_
years_select)=="T050304"), which( colnames(ATUS_years_select)=="T050389"), wh
ich( colnames(ATUS_years_select)=="T050403"), which( colnames(ATUS_years_sele
ct)=="T050404"), which( colnames(ATUS_years_select)=="T050405"), which( colna
mes(ATUS_years_select)=="T050481"), which( colnames(ATUS_years_select)=="T050
499"), which( colnames(ATUS_years_select)=="T059999")))
Pre_Work <- ATUS_years_select[,Variable_Work]
ncol(Pre_Work)

## [1] 19

Pre_Work$Work <- rowSums(Pre_Work[1:19])
Work <- data.frame(Pre_Work$Work)

#Travel
```

```r
Variable_Travel <- (c(which( colnames(ATUS_years_select)=="T080101"), which(
colnames(ATUS_years_select)=="T089999")))
Variable_Travel

## [1] 189 214

Pre_Travel <- ATUS_years_select[,189:214]
Pre_Travel$Travel <- rowSums(Pre_Travel[ncol(Pre_Travel)])
Travel <- data.frame(Pre_Travel$Travel)

#Household Activities
Variable_HH_Activities <- (c(which( colnames(ATUS_years_select)=="T020101"),
which( colnames(ATUS_years_select)=="T029999")))
Variable_HH_Activities

## [1] 38 69

Pre_HH_Activities <- ATUS_years_select[,38:69]
ncol(Pre_HH_Activities)

## [1] 32

Pre_HH_Activities$HH_Activities <- rowSums(Pre_HH_Activities[1:32])
HH_Activities <- data.frame(Pre_HH_Activities$HH_Activities)

#Education
Variable_Educ <- (c(which( colnames(ATUS_years_select)=="T060101"), which( co
lnames(ATUS_years_select)=="T069999")))
Variable_Educ

## [1] 160 177

Pre_Educ <- ATUS_years_select[,160:177]
ncol(Pre_Educ)

## [1] 18

Pre_Educ$Education <- rowSums(Pre_Educ[1:18])
Education <- data.frame(Pre_Educ$Education)

#Sports
Variable_Sports <- (c(which( colnames(ATUS_years_select)=="T130101"), which(
colnames(ATUS_years_select)=="T139999")))
Variable_Sports

## [1] 281 357
```

```
Pre_Sports <- ATUS_years_select[,281:357]
ncol(Pre_Sports)

## [1] 77

Pre_Sports$Sports <- rowSums(Pre_Sports[1:77])
Sports <- data.frame(Pre_Sports$Sports)

#Helping household members
Variable_Helping_HH <- (c(which( colnames(ATUS_years_select)=="T030101"), whi
ch( colnames(ATUS_years_select)=="T039999")))
Variable_Helping_HH

## [1]  70 102

Pre_Helping_HH<- ATUS_years_select[,70:102]
ncol(Pre_Helping_HH)

## [1] 33

Pre_Helping_HH$Helping_HH <- rowSums(Pre_Helping_HH[1:33])
Helping_HH <- data.frame(Pre_Helping_HH$Helping_HH)

#Eating and drinking
Variable_Eat_Drink <- (c(which( colnames(ATUS_years_select)=="T110101"), whic
h( colnames(ATUS_years_select)=="T119999")))
Variable_Eat_Drink

## [1] 245 249

Pre_Eat_Drink <- ATUS_years_select[,245:249]
ncol(Pre_Eat_Drink)

## [1] 5

Pre_Eat_Drink$Eat_Drink <- rowSums(Pre_Eat_Drink[1:5])
Eat_Drink <- data.frame(Pre_Eat_Drink$Eat_Drink)

#Helping non-housheold members
Variable_Helping_NONHH <- (c(which( colnames(ATUS_years_select)=="T040101"),
which( colnames(ATUS_years_select)=="T049999")))
Variable_Helping_NONHH

## [1] 103 139

Pre_Helping_NONHH<- ATUS_years_select[,103:139]
ncol(Pre_Helping_NONHH)
```

```
## [1] 37

Pre_Helping_NONHH$Helping_NONHH <- rowSums(Pre_Helping_NONHH[1:37])
Helping_NONHH <- data.frame(Pre_Helping_NONHH$Helping_NONHH)

#Religion
Variable_Rel <- (c(which( colnames(ATUS_years_select)=="T140101"), which( col
names(ATUS_years_select)=="T149999")))
Variable_Rel

## [1] 358 363

Pre_Rel <- ATUS_years_select[,358:363]
ncol(Pre_Rel)

## [1] 6

Pre_Rel$Religion <- rowSums(Pre_Rel[1:6])
Religion <- data.frame(Pre_Rel$Religion)

#Volunteer
Variable_Vol <- (c(which( colnames(ATUS_years_select)=="T150101"), which( col
names(ATUS_years_select)=="T159989")))
Variable_Vol

## [1] 364 387

Pre_Vol <- ATUS_years_select[,364:387]
ncol(Pre_Vol)

## [1] 24

Pre_Vol$Volunteer <- rowSums(Pre_Vol[1:24])
Volunteer <- data.frame(Pre_Vol$Volunteer)

#Professional care and services
Variable_Prof_Care <- (c(which( colnames(ATUS_years_select)=="T080101"), whic
h( colnames(ATUS_years_select)=="T089999")))
Variable_Prof_Care

## [1] 189 214

Pre_Prof_Care <- ATUS_years_select[,189:214]
ncol(Pre_Prof_Care)

## [1] 26
```

```r
Pre_Prof_Care$Professional_Care <- rowSums(Pre_Prof_Care[1:26])
Professional_Care <- data.frame(Pre_Prof_Care$Professional_Care)

#Household services
Variable_HH_Services <- (c(which( colnames(ATUS_years_select)=="T090101"), wh
ich( colnames(ATUS_years_select)=="T099999")))
Variable_HH_Services

## [1] 215 232

Pre_HH_Serv <- ATUS_years_select[,215:232]
ncol(Pre_HH_Serv)

## [1] 18

Pre_HH_Serv$HH_Services <- rowSums(Pre_HH_Serv[1:18])
HH_Services <- data.frame(Pre_HH_Serv$HH_Services)

#Government and civic obligations
Variable_Gov_and_Civic_Obligations <- (c(which( colnames(ATUS_years_select)==
"T100101"), which( colnames(ATUS_years_select)=="T109999")))
Variable_Gov_and_Civic_Obligations

## [1] 233 244

Pre_Gov <- ATUS_years_select[,233:244]
ncol(Pre_Gov)

## [1] 12

Pre_Gov$Gov_and_Civic_Obligations <- rowSums(Pre_Gov[1:12])
Gov_and_Civic_Obligations <- data.frame(Pre_Gov$Gov_and_Civic_Obligations)

#Consumer Purchases
Variable_Cons_Purch <- (c(which( colnames(ATUS_years_select)=="T070101"), whi
ch( colnames(ATUS_years_select)=="T079999")))
Variable_Cons_Purch

## [1] 178 188

Pre_Cons_Purch <- ATUS_years_select[,178:188]
ncol(Pre_Cons_Purch)

## [1] 11

Pre_Cons_Purch$Consumer_Purchases <- rowSums(Pre_Cons_Purch[1:11])
Consumer_Purchases <- data.frame(Pre_Cons_Purch$Consumer_Purchases)
```

```r
#Phone calls
Variable_Phone <- (c(which( colnames(ATUS_years_select)=="T160101"), which( c
olnames(ATUS_years_select)=="T169989")))
Variable_Phone

## [1] 388 396

Pre_Phone <- ATUS_years_select[,388:396]
ncol(Pre_Phone)

## [1] 9

Pre_Phone$Phone_calls <- rowSums(Pre_Phone[1:9])
Phone_calls <- data.frame(Pre_Phone$Phone_calls)

#Computer Leisure (Excluding Games)
Computer_leisure <- data.frame(ATUS_years_select$T120308)

#ATUS_time
ATUS_time <- cbind(Phone_calls, Consumer_Purchases, Gov_and_Civic_Obligations
, HH_Services, Professional_Care, Volunteer, Religion, Helping_HH, Helping_NO
NHH, Sports, Education, HH_Activities, Travel, Personal_Care_Sleep, Work, Lei
sure_Excl_Computer, Eat_Drink ,Computer_leisure)

#Adjusting column names
colnames(ATUS_time) <- c("Phone_calls", "Consumer_Purchases", "Gov_and_Civic_
Obligations", "HH_Services", "Professional_Care", "Volunteer", "Religion", "H
elping_HH", "Helping_NONHH", "Sports", "Education", "HH_Activities", "Travel"
, "Personal_Care_Sleep", "Work", "Leisure_Excl_Computer", "Eat_Drink", "Compu
ter_leisure")

write.csv(ATUS_time, "Cap_ATUS_time.csv")
```

Formatting ATUS data

Using MySQL commands, we will format our data in order to have a consistent data format for both our CPS dataset and ATUS dataset.

First we add the library

```r
library(sqldf)

## Loading required package: gsubfn

## Loading required package: proto
```

```
## Loading required package: RSQLite
```

As we can see below, the categorical variables in the CPS dataset include only the category number. In the ATUS dataset, the whole response is included. We want to format the ATUS dataset in such a way that both our datasets contain the same data format.

```
CPS_prep[1:3,]

##   PERSON_TYPE LABOUR_FORCE_STATUS AGE HISPANIC SEX MORE_THAN_1_JOB
## 1           2                   1  54        2   1               2
## 2           2                   1  55        2   2               2
## 3          -1                  -1  -1       -1  -1              -1
##   HOURS_PER_WEEK FULL_TIME_PART_TIME WEEKLY_EARNINGS EDUCATION SPOUSE
## 1             24                  38              -1        44      1
## 2             24                  38              -1        45      1
## 3             -1                   0              -1        -1     -1
##   CHILDREN AGE_YOUNGEST_CHILD METROPOLITAN_STATUS HOME_INTERNET_ACCESS
## 1        0                  0                   1                    1
## 2        0                  0                   1                    1
## 3       -1                 -1                   1                   -1

ATUS_prep[1:3,]

##           LABOUR_FORCE_STATUS AGE         HISPANIC        SEX
## 112039 (5) Not in labor force  62 (2) Non-Hispanic (2) Female
## 112040 (1) Employed - at work  22 (2) Non-Hispanic (2) Female
## 112041 (1) Employed - at work  33 (2) Non-Hispanic   (1) Male
##        MORE_THAN_1_JOB HOURS_PER_WEEK FULL_TIME_PART_TIME WEEKLY_EARNINGS
## 112039            <NA>             NA                <NA>              NA
## 112040          (2) No             40       (1) Full time             150
## 112041          (2) No             42       (1) Full time             350
##                                                 EDUCATION
## 112039                                   (37) 11th grade
## 112040 (39) High school graduate - diploma or equivalent [GED]
## 112041                                   (36) 10th grade
##                                         SPOUSE CHILDREN
## 112039 (3) No spouse or unmarried partner present        1
## 112040 (3) No spouse or unmarried partner present        0
## 112041                       (1) Spouse present        1
##        AGE_YOUNGEST_CHILD  METROPOLITAN_STATUS
## 112039                  9      (1) Metropolitan
## 112040                 NA      (1) Metropolitan
## 112041                 15 (2) Non-metropolitan
```

```
#LABOUR_FORCE_STATUS
ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET LABOUR_FORCE_STATUS=1 where LABOUR
_FORCE_STATUS LIKE '%1%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET LABOUR_FORCE_STATUS=2 where LABOUR
_FORCE_STATUS LIKE '%2%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET LABOUR_FORCE_STATUS=3 where LABOUR
_FORCE_STATUS LIKE '%3%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET LABOUR_FORCE_STATUS=4 where LABOUR
_FORCE_STATUS LIKE '%4%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET LABOUR_FORCE_STATUS=5 where LABOUR
_FORCE_STATUS LIKE '%5%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

#HISPANIC
ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET HISPANIC=1 where HISPANIC LIKE '%(
1)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET HISPANIC=2 where HISPANIC LIKE '%(
2)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

#SEX
ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET SEX=1 where SEX LIKE '%(1)%'", "SE
LECT * from ATUS_prep"))
```

```
## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET SEX=2 where SEX LIKE '%(2)%'", "SE
LECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

#MORE_THAN_1_JOB
ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET MORE_THAN_1_JOB=1 where MORE_THAN_
1_JOB LIKE '%(1)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET MORE_THAN_1_JOB=2 where MORE_THAN_
1_JOB LIKE '%(2)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

#FULL_TIME_PART_TIME
ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET FULL_TIME_PART_TIME=1 where FULL_T
IME_PART_TIME LIKE '%(1)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET FULL_TIME_PART_TIME=2 where FULL_T
IME_PART_TIME LIKE '%(2)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

#EDUCATION
ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET EDUCATION=31 where EDUCATION LIKE
'%(31)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET EDUCATION=32 where EDUCATION LIKE
'%(32)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries
```

```
ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET EDUCATION=33 where EDUCATION LIKE
'%(33)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET EDUCATION=34 where EDUCATION LIKE
'%(34)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET EDUCATION=35 where EDUCATION LIKE
'%(35)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET EDUCATION=36 where EDUCATION LIKE
'%(36)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET EDUCATION=37 where EDUCATION LIKE
'%(37)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET EDUCATION=38 where EDUCATION LIKE
'%(38)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET EDUCATION=39 where EDUCATION LIKE
'%(39)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET EDUCATION=40 where EDUCATION LIKE
'%(40)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries
```

```r
ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET EDUCATION=41 where EDUCATION LIKE
'%(41)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET EDUCATION=42 where EDUCATION LIKE
'%(42)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET EDUCATION=43 where EDUCATION LIKE
'%(43)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET EDUCATION=44 where EDUCATION LIKE
'%(44)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET EDUCATION=45 where EDUCATION LIKE
'%(45)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET EDUCATION=46 where EDUCATION LIKE
'%(46)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

#SPOUSE
ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET SPOUSE=1 where SPOUSE LIKE '%(1)%'
", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET SPOUSE=2 where SPOUSE LIKE '%(2)%'
", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries
```

```r
ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET SPOUSE=3 where SPOUSE LIKE '%(3)%'
", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

#METROPOLITAN_STATUS
ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET METROPOLITAN_STATUS=1 where METROP
OLITAN_STATUS LIKE '%(1)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET METROPOLITAN_STATUS=2 where METROP
OLITAN_STATUS LIKE '%(2)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET METROPOLITAN_STATUS=3 where METROP
OLITAN_STATUS LIKE '%(3)%'", "SELECT * from ATUS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries
```

As we can seen below, both our datasets have the same data format now.

```r
#As seen below, both our datasets have the same data format.
CPS_prep[1:3,]

##   PERSON_TYPE LABOUR_FORCE_STATUS AGE HISPANIC SEX MORE_THAN_1_JOB
## 1           2                   1  54        2   1               2
## 2           2                   1  55        2   2               2
## 3          -1                  -1  -1       -1  -1              -1
##   HOURS_PER_WEEK FULL_TIME_PART_TIME WEEKLY_EARNINGS EDUCATION SPOUSE
## 1             24                  38              -1        44      1
## 2             24                  38              -1        45      1
## 3             -1                   0              -1        -1     -1
##   CHILDREN AGE_YOUNGEST_CHILD METROPOLITAN_STATUS HOME_INTERNET_ACCESS
## 1        0                  0                   1                    1
## 2        0                  0                   1                    1
## 3       -1                 -1                   1                   -1

ATUS_prep[1:3,]

##   LABOUR_FORCE_STATUS AGE HISPANIC SEX MORE_THAN_1_JOB HOURS_PER_WEEK
## 1                   5  62        2   2            <NA>             NA
```

```
## 2                    1  22         2   2               2             40
## 3                    1  33         2   1               2             42
##   FULL_TIME_PART_TIME WEEKLY_EARNINGS EDUCATION SPOUSE CHILDREN
## 1                <NA>              NA        37      3        1
## 2                   1             150        39      3        0
## 3                   1             350        36      1        1
##   AGE_YOUNGEST_CHILD METROPOLITAN_STATUS
## 1                  9                    1
## 2                 NA                    1
## 3                 15                    2
```

Dealing with variable consistency and data-type

The next step would be to look at each variable and manipulate our data to have consistent results in ATUS and CPS for each variables.

We want to make sure the ATUS_prep and CPS_prep data-type are set to numeric for easier data manipulation using the sqldf package.

```
ATUS_prep <- as.data.frame(sapply(ATUS_prep[,1:13], as.numeric))
CPS_prep <- as.data.frame(sapply(CPS_prep[,1:15], as.numeric))
```

PERSON_TYPE. This variable is present only in the CPS dataset. Because the ATUS dataset include only Adult Civilian Household Members (15+), we will select only those in category 2 (Adult Civilian Household Members 15+). The CPS dataframe will have 136017 fewer observations (from 451984 to 315967). Once all PERSON_TYPE=Adult Civilian Household Members (2) is selected, we can drop this column.

```
CPS_prep <- sqldf("SELECT * FROM CPS_prep WHERE PERSON_TYPE = 2")
CPS_prep[,1] <- NULL
```

LABOUR_FORCE_STATUS: Labor force Status. The ATUS dataset has 5 categories: 1-Employed-at-work, 2-Employed-absent, 3-Unemployed-on-layoff, 4-Unemployed-looking, 5-Not-in-laborforce. The CPS dataset has 7 categories. The first 4 are similar to the ATUS ones, however, the not-in-laborforce is expandeed into 3 other categories: 5-Not-in-laborforce-Retired, 6-Not-in-laborforce-Disables, 7-Not-in-laborforce-Other. We will merge these 3 categories into one category: not-in-laborforce, so that both the ATUS and CPS datasets have 5 factors. Then we set the correct data-type for this variable: factor.

```
CPS_prep <- sqldf(c("UPDATE CPS_prep SET LABOUR_FORCE_STATUS = 5 WHERE LABOUR
_FORCE_STATUS = 6 OR LABOUR_FORCE_STATUS = 7", "SELECT * FROM CPS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

CPS_prep$LABOUR_FORCE_STATUS <- as.factor(CPS_prep$LABOUR_FORCE_STATUS)
ATUS_prep$LABOUR_FORCE_STATUS <- as.factor(ATUS_prep$LABOUR_FORCE_STATUS)
```

Age: Age. We leave the data-type as numeric for this variable.

HISPANIC: Hispanic. ATUS and CPS have the same categories.

```
CPS_prep$HISPANIC <- as.factor(CPS_prep$HISPANIC)
ATUS_prep$HISPANIC <- as.factor(ATUS_prep$HISPANIC)
```

SEX: Sex. ATUS and CPS have the same categories. We will only adjust the data type to factor.

```
CPS_prep$SEX <- as.factor(CPS_prep$SEX)
ATUS_prep$SEX <- as.factor(ATUS_prep$SEX)
```

MORE_THAN_1_JOB: More than one job. ATUS and CPS have the same categories. We will only adjust the data type to factor.

```
CPS_prep$MORE_THAN_1_JOB <- as.factor(CPS_prep$MORE_THAN_1_JOB)
ATUS_prep$MORE_THAN_1_JOB <- as.factor(ATUS_prep$MORE_THAN_1_JOB)
```

HOURS_PER_WEEK: Total hours worked per week. We leave the data-type as numeric for this variable.

FULL-TIME OR PART-TIME. This variable is included in the ATUS dataset, but not in the CPS dataset. We will determine whether a person works full-time or part-time based on the following condition: if a person works 35 hours or more per week, he/she is considered to be working as full-time, otherwise, he/she is considered to be working as part-time. The total hours worked per week are found in variable HOURS_PER_WEEK.

```
CPS_prep$FULL_TIME_PART_TIME <- CPS_prep$HOURS_PER_WEEK
CPS_prep <- sqldf(c("UPDATE CPS_prep SET FULL_TIME_PART_TIME=-1 WHERE FULL_TI
ME_PART_TIME=-1", "SELECT * FROM CPS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

CPS_prep <- sqldf(c("UPDATE CPS_prep SET FULL_TIME_PART_TIME=1000 WHERE FULL_
TIME_PART_TIME > 35", "SELECT * FROM CPS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

CPS_prep <-sqldf(c("UPDATE CPS_prep SET FULL_TIME_PART_TIME=2000 WHERE FULL_T
IME_PART_TIME BETWEEN 0 AND 35", "SELECT * FROM CPS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

CPS_prep <- sqldf(c("UPDATE CPS_prep SET FULL_TIME_PART_TIME=1 WHERE FULL_TIM
E_PART_TIME = 1000", "select * from CPS_prep"))
```

```
## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries
```

```
CPS_prep <- sqldf(c("UPDATE CPS_prep SET FULL_TIME_PART_TIME=2 WHERE FULL_TIM
E_PART_TIME = 2000", "select * from CPS_prep"))
```

```
## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries
```

WEEKLY EARNINGS. We leave the data-type as numeric for this variable.

EDUCATION: Highest Level of Education. ATUS and CPS have the same categories. We will only adjust the data type to factor.

```
CPS_prep$EDUCATION <- as.factor(CPS_prep$EDUCATION)
ATUS_prep$EDUCATION <- as.factor(ATUS_prep$EDUCATION)
```

SPOUSE: Presence of spouse in the household. The ATUS dataset has 3 responses to this variable: 1-Spouse present, 2-Unmarried partner present, 3-No spouse or unmarried partner present. In the CPS dataset, there are 6 categories: 1-Married spouse present, 2-Married spouse absent, 3-Widowed, 4-Divorced, 5-Separated, 6-Never Married. For convenience, we want our categories to be: 1-Spouse present, 2-Spouse absent or no spouse. As such, we will merge ATUS categories 2 and 3 into one category: 2-Spouse absent or no spouse.; and category 1 will be: 1-Spouse present. As for the CPS, we will merge categories 2, 3, 4, 5, 6 into one category: 2-Spouse absent or no spouse, and category 1 will be: 1-Spouse present.

```
ATUS_prep$SPOUSE <- as.numeric(ATUS_prep$SPOUSE)
ATUS_prep <- sqldf(c("UPDATE ATUS_prep SET SPOUSE = 2 WHERE SPOUSE  = 3", "SE
LECT * FROM ATUS_prep"))
```

```
## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries
```

```
ATUS_prep$SPOUSE <- as.factor(ATUS_prep$SPOUSE)


CPS_prep <- sqldf(c("UPDATE CPS_prep SET SPOUSE = 2 WHERE SPOUSE!=1 ", "SELEC
T * FROM CPS_prep"))
```

```
## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries
```

```
CPS_prep$SPOUSE <- as.factor(CPS_prep$SPOUSE)
```

CHILDREN: Number of children in the household. We leave this data type as numeric.

```
ATUS_prep$CHILDREN <- as.numeric(ATUS_prep$CHILDREN)
CPS_prep$CHILDREN <- as.numeric(CPS_prep$CHILDREN)
```

AGE_YOUNGEST_CHILD: Age of youngest child (<18) in the household. The ATUS records for this variable is a numeric data ranging from 0 to 17. However, the similar variable in the CPS dataset include range group of age (for the CPS category details, the CPS codebook can be consulted). To overcome this challenge, we will tranform our data into the following categories: 1-[0-2], 2-[3-5], 3-[6-13], 4-[14-17].

```
CPS_prep <- sqldf(c("UPDATE CPS_prep SET AGE_YOUNGEST_CHILD = -100 WHERE AGE_
YOUNGEST_CHILD = -1 OR AGE_YOUNGEST_CHILD = 0", "SELECT * FROM CPS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

CPS_prep <- sqldf(c("UPDATE CPS_prep SET AGE_YOUNGEST_CHILD = 100 WHERE AGE_Y
OUNGEST_CHILD IN (1,5,6,7,11,12,13,15)", "SELECT * FROM CPS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

CPS_prep <- sqldf(c("UPDATE CPS_prep SET AGE_YOUNGEST_CHILD = 200 WHERE AGE_Y
OUNGEST_CHILD IN (14,2,8,9)", "SELECT * FROM CPS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

CPS_prep <- sqldf(c("UPDATE CPS_prep SET AGE_YOUNGEST_CHILD = 300 WHERE AGE_Y
OUNGEST_CHILD IN (3,10)", "SELECT * FROM CPS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

CPS_prep <- sqldf(c("UPDATE CPS_prep SET AGE_YOUNGEST_CHILD = 400 WHERE AGE_Y
OUNGEST_CHILD=4", "SELECT * FROM CPS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

CPS_prep <- sqldf(c("UPDATE CPS_prep SET AGE_YOUNGEST_CHILD = -1 WHERE AGE_YO
UNGEST_CHILD=-100", "SELECT * FROM CPS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

CPS_prep <- sqldf(c("UPDATE CPS_prep SET AGE_YOUNGEST_CHILD = 1 WHERE AGE_YOU
NGEST_CHILD= 100", "SELECT * FROM CPS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries
```

```r
CPS_prep <- sqldf(c("UPDATE CPS_prep SET AGE_YOUNGEST_CHILD = 2 WHERE AGE_YOU
NGEST_CHILD= 200", "SELECT * FROM CPS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

CPS_prep <- sqldf(c("UPDATE CPS_prep SET AGE_YOUNGEST_CHILD = 3 WHERE AGE_YOU
NGEST_CHILD= 300", "SELECT * FROM CPS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

CPS_prep <- sqldf(c("UPDATE CPS_prep SET AGE_YOUNGEST_CHILD = 4 WHERE AGE_YOU
NGEST_CHILD= 400", "SELECT * FROM CPS_prep"))

## Warning in rsqlite_fetch(res@ptr, n = n): Don't need to call dbFetch() for
## statements, only for queries

CPS_prep$AGE_YOUNGEST_CHILD <- as.factor(CPS_prep$AGE_YOUNGEST_CHILD)

ATUS_prep$AGE_YOUNGEST_CHILD <- cut(ATUS_prep$AGE_YOUNGEST_CHILD, c(0,2,5,13,
17))
levels(ATUS_prep$AGE_YOUNGEST_CHILD) <- c(1, 2, 3, 4)
```

METROPOLITAN_STATUS: Metropolitan Status. ATUS and CPS have the same categories. We will only adjust the data type to factor.

```r
CPS_prep$METROPOLITAN_STATUS <- as.factor(CPS_prep$METROPOLITAN_STATUS)
ATUS_prep$METROPOLITAN_STATUS <- as.factor(ATUS_prep$METROPOLITAN_STATUS)
```

HOME_INTERNET_ACCESS: Internet Access. This variable is present only in the CPS dataset. It will be used for constructing our model. We change the data type to factor.

```r
CPS_prep$HOME_INTERNET_ACCESS <- as.factor(CPS_prep$HOME_INTERNET_ACCESS)
```

The CPS dataset marks NAs as -1 whereas in the ATUS dataset, missing values are correctly labeled NA. We will replace all values of -1 in the CPS dataset with NAs, so that the CPS and ATUS datasets become consistent.

```r
sum(CPS_prep == -1)

## [1] 927963

CPS_prep[CPS_prep == -1] <- NA
sum(is.na(CPS_prep) == T)

## [1] 927963
```

Exploring our datasets

There are 315967 rows in the CPS data and 58804 in the ATUS data (ATUS_time and ATUS_prep). There is a total NA number of 927963 out of 4423538 observations in the CPS dataset. Thereis a total NA number of 134816 out of 764452 observations in the ATUS_prep dataset. There is no NA in the ATUS_time dataset, out of the 999668 observations. We intend to remove the NA's from the Datasets form ATUS_prep and CPS_prep.

```
nrow(CPS_prep)

## [1] 315967

nrow(ATUS_prep)

## [1] 58804

nrow(ATUS_time)

## [1] 58804

sum(is.na(CPS_prep))

## [1] 927963

nrow(CPS_prep) * ncol(CPS_prep)

## [1] 4423538

sum(is.na(ATUS_prep))

## [1] 134816

nrow(ATUS_prep) * ncol(ATUS_prep)

## [1] 764452

sum(is.na(ATUS_time))

## [1] 0

nrow(ATUS_time) * ncol(ATUS_time)

## [1] 1058472
```

Dealing with Missing Values

First, we see if there are any missing value in variable HOME_INTERNET_ACCESS of the CPS dataset, and we remove those rows.

```
sum(is.na(CPS_prep$HOME_INTERNET_ACCESS))
```

```
## [1] 26078

CPS_prep<- CPS_prep[-which(is.na(CPS_prep$HOME_INTERNET_ACCESS)), ]
```

Then we will explore the NAs proportion in each column, and remove the columns that have more than 0.33% missing values in both the CPS and ATUS dataset using the function included in sapply() below.

MORE_THAN_1_JOB has 41.4% NAs in CPS and 39.7% in ATUS. HOURS_PER_WEEK has 41.4% NAs in CPS and 43.3% in ATUS. FULL_TIME_PART_TIME has 41.4% in CPS and 39.7% in ATUS. WEEKLY_EARNINGS has 86% NAs in the CPS dataset and 46% in the ATUS dataset. AGE_YOUNGEST_CHILD has 74% NAs in the CPS dataset and 60% in the ATUS dataset.

We start with 14 column in CPS and reduce the columns to 9. We start with 13 column in CPS and reduce the columns to 8.

```
CPS_row_col <- c(nrow(CPS_prep), ncol(CPS_prep))
ATUS_row_col <- c(nrow(ATUS_prep), ncol(ATUS_prep))

CPS_row_col

## [1] 289889     14

ATUS_row_col

## [1] 58804    13

na_count <-sapply(CPS_prep, function(CPS_prep) sum(length(which(is.na(CPS_pre
p)))))/nrow(CPS_prep)
na_count

##  LABOUR_FORCE_STATUS                  AGE             HISPANIC
##            0.0000000            0.0000000            0.0000000
##                  SEX      MORE_THAN_1_JOB       HOURS_PER_WEEK
##            0.0000000            0.4139343            0.4139343
##  FULL_TIME_PART_TIME      WEEKLY_EARNINGS            EDUCATION
##            0.4139343            0.8577559            0.0000000
##               SPOUSE             CHILDREN   AGE_YOUNGEST_CHILD
##            0.0000000            0.0000000            0.7436812
##  METROPOLITAN_STATUS HOME_INTERNET_ACCESS
##            0.0000000            0.0000000

na_count2 <-sapply(ATUS_prep, function(ATUS_prep) sum(length(which(is.na(ATUS
_prep)))))/nrow(ATUS_prep)
na_count2
```

```
## LABOUR_FORCE_STATUS                  AGE           HISPANIC
##           0.0000000           0.0000000          0.0000000
##                 SEX      MORE_THAN_1_JOB      HOURS_PER_WEEK
##           0.0000000           0.3971669          0.4328107
## FULL_TIME_PART_TIME      WEEKLY_EARNINGS           EDUCATION
##           0.3971669           0.4635909          0.0000000
##              SPOUSE            CHILDREN  AGE_YOUNGEST_CHILD
##           0.0000000           0.0000000          0.6018978
## METROPOLITAN_STATUS
##           0.0000000
```

```
CPS_prep <- subset(CPS_prep, select = -c(MORE_THAN_1_JOB, HOURS_PER_WEEK, FUL
L_TIME_PART_TIME, WEEKLY_EARNINGS, AGE_YOUNGEST_CHILD))
ATUS_prep <- subset(ATUS_prep, select = -c(MORE_THAN_1_JOB, HOURS_PER_WEEK, F
ULL_TIME_PART_TIME, WEEKLY_EARNINGS, AGE_YOUNGEST_CHILD))

CPS_row_col <- c(nrow(CPS_prep), ncol(CPS_prep))
ATUS_row_col <- c(nrow(ATUS_prep), ncol(ATUS_prep))

CPS_row_col
```

```
## [1] 289889      9
```

```
ATUS_row_col
```

```
## [1] 58804     8
```

As we can see below, following the previous step, all NAs have been removed from both our datasets CPS and ATUS.

```
na_count3 <-sapply(CPS_prep, function(CPS_prep) sum(length(which(is.na(CPS_pr
ep)))))/nrow(CPS_prep)
na_count3
```

```
##   LABOUR_FORCE_STATUS                  AGE             HISPANIC
##                     0                    0                    0
##                   SEX            EDUCATION               SPOUSE
##                     0                    0                    0
##              CHILDREN  METROPOLITAN_STATUS HOME_INTERNET_ACCESS
##                     0                    0                    0
```

```
na_count4 <-sapply(ATUS_prep, function(ATUS_prep) sum(length(which(is.na(ATUS
_prep)))))/nrow(ATUS_prep)
na_count4
```

```
## LABOUR_FORCE_STATUS                 AGE            HISPANIC
##                   0                   0                   0
##                 SEX           EDUCATION              SPOUSE
##                   0                   0                   0
##            CHILDREN METROPOLITAN_STATUS
##                   0                   0
```

```
sum(is.na(CPS_prep))
```

```
## [1] 0
```

```
sum(is.na(ATUS_prep))
```

```
## [1] 0
```

Dealing with Outliers (Demographics Variables)

Now we want to remove all outliers from our datasets.

Because we will be using the ATUS_time dataset later in phase 3 (for the application of the decision tree model), we need to make sure that whenever we remove rows from the ATUS_prep during data cleaning, the same rows are removed from the ATUS_time dataframe. Otherwise, when joining the internet_predicted column with the ATUS_time dataframe, we will get an error due to unmatching number of rows. Therefore, below, we normalize our data:

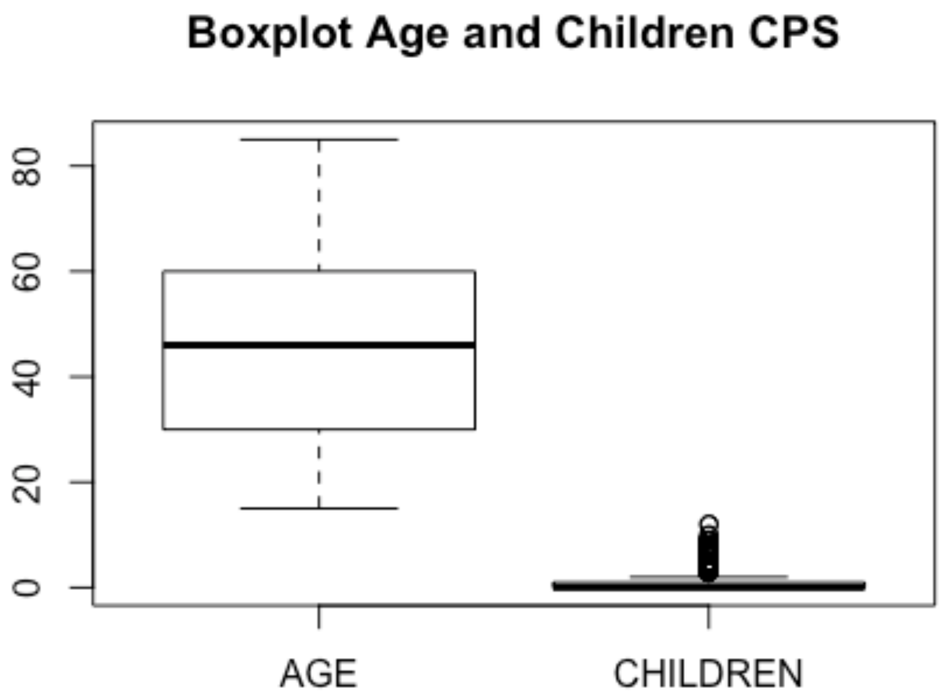
```
ATUS_norm <- cbind(ATUS_prep, ATUS_time)
c(nrow(ATUS_norm), ncol(ATUS_norm))
```

```
## [1] 58804    26
```

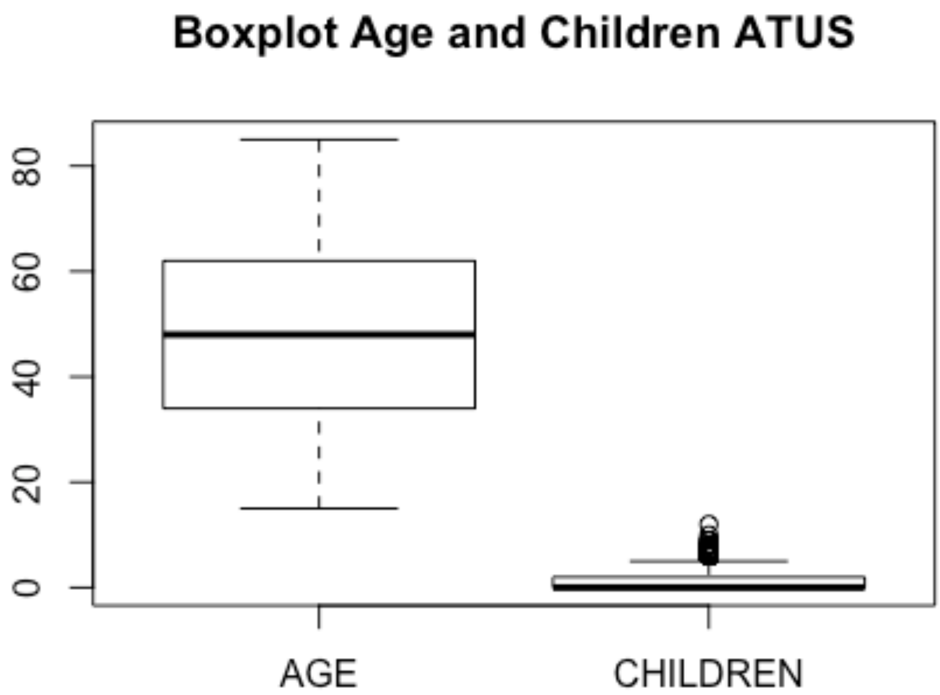
We start with the Age variable in both the CPS and ATUS dataset.

First we will visualize the outliers. This applies to the numerical variables: Age and Children. By looking at the boxplot, we can see that there are no outliers for Age, but there are some for children. However, if we look at the histogram for Children, it makes sense that only very few individuals have more than 6 children. As such, having up to 12 children doesn't seem to be an outlier, but a rare case in the population.

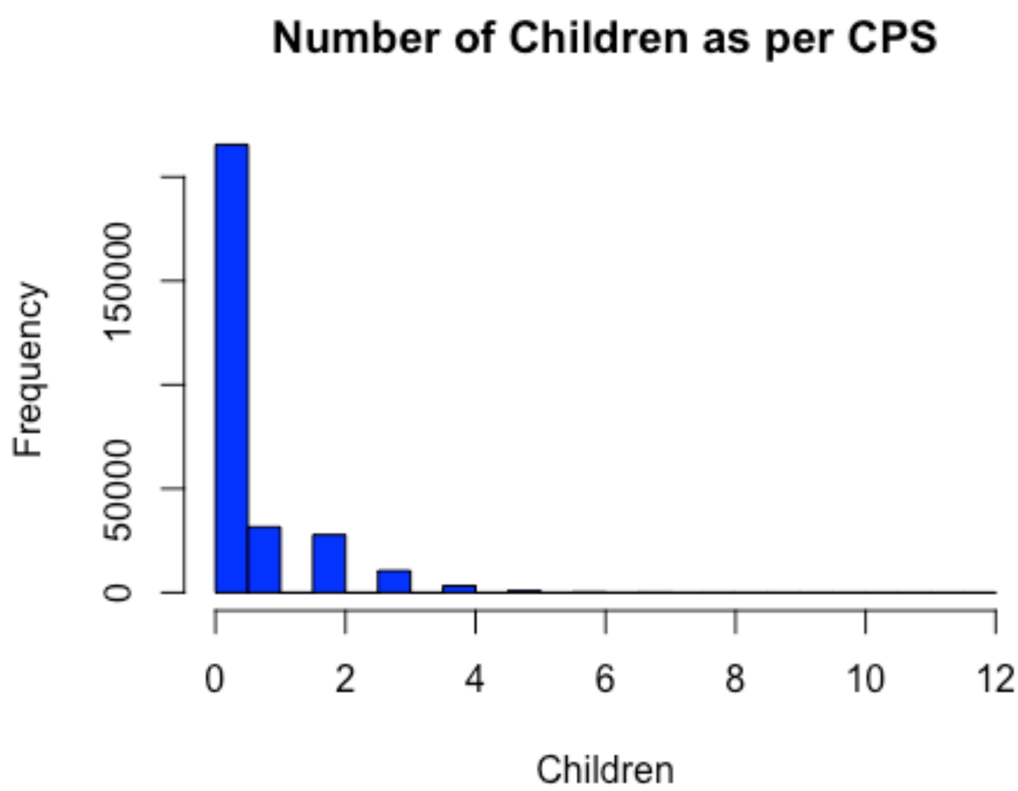
```
boxplot(CPS_prep[, c(2,7)], main="Boxplot Age and Children CPS")
```



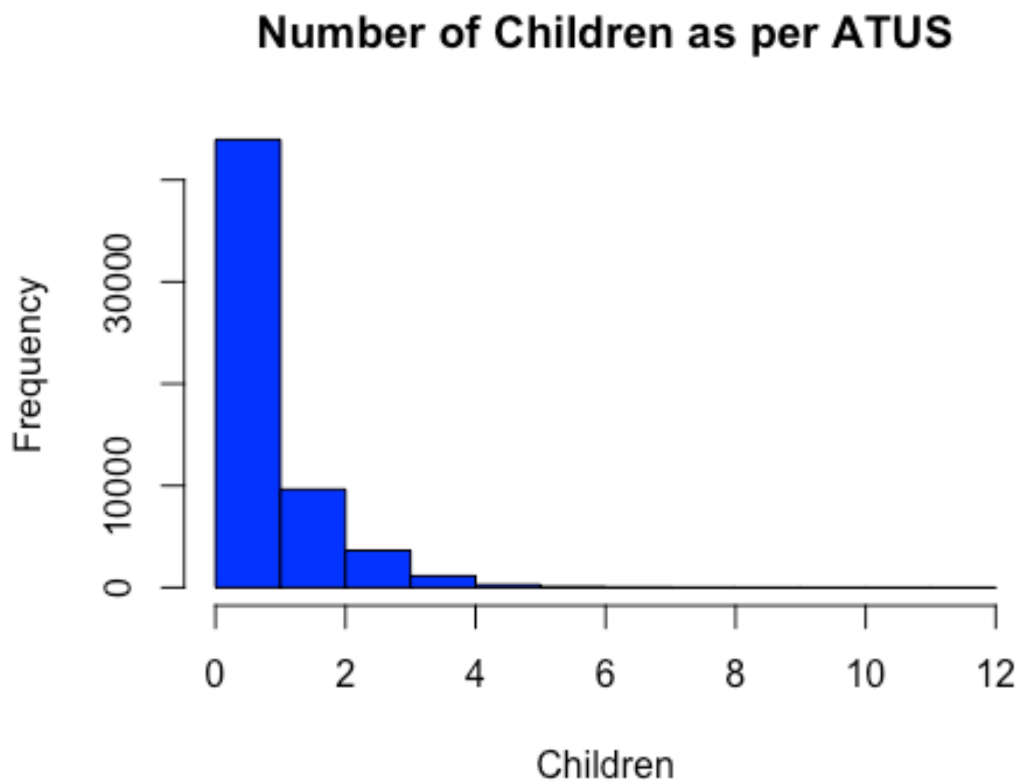

```
boxplot(ATUS_norm[, c(2,7)], main="Boxplot Age and Children ATUS")
```




```
hist(CPS_prep$CHILDREN, main="Number of Children as per CPS", xlab = "Children", col="blue")
```




```
hist(ATUS_norm$CHILDREN, main="Number of Children as per ATUS", xlab = "Child
ren", col="blue")
```

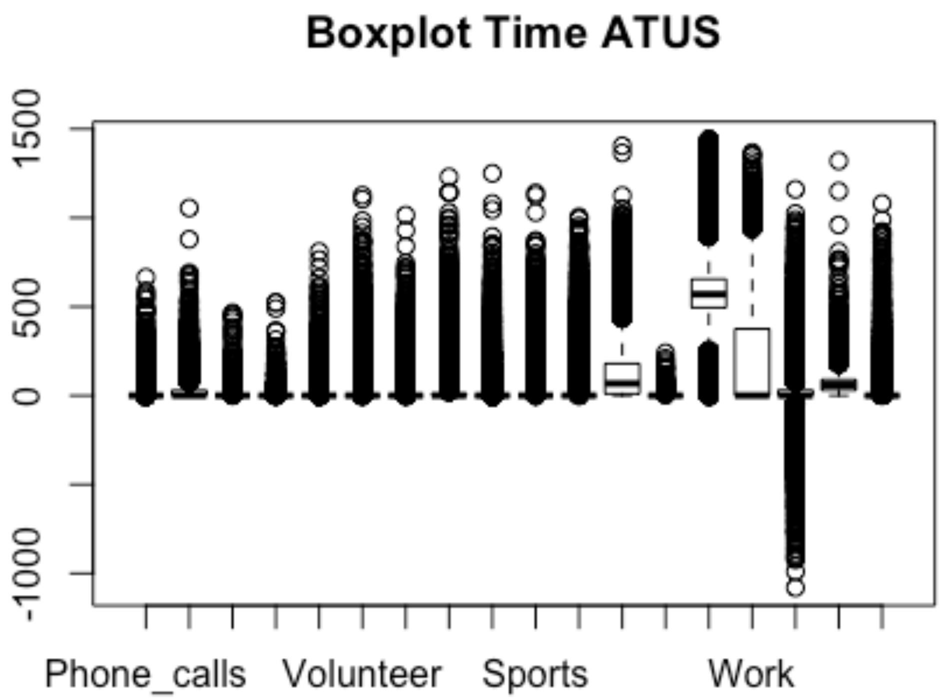


Dealing with Outliers (Time Variables)

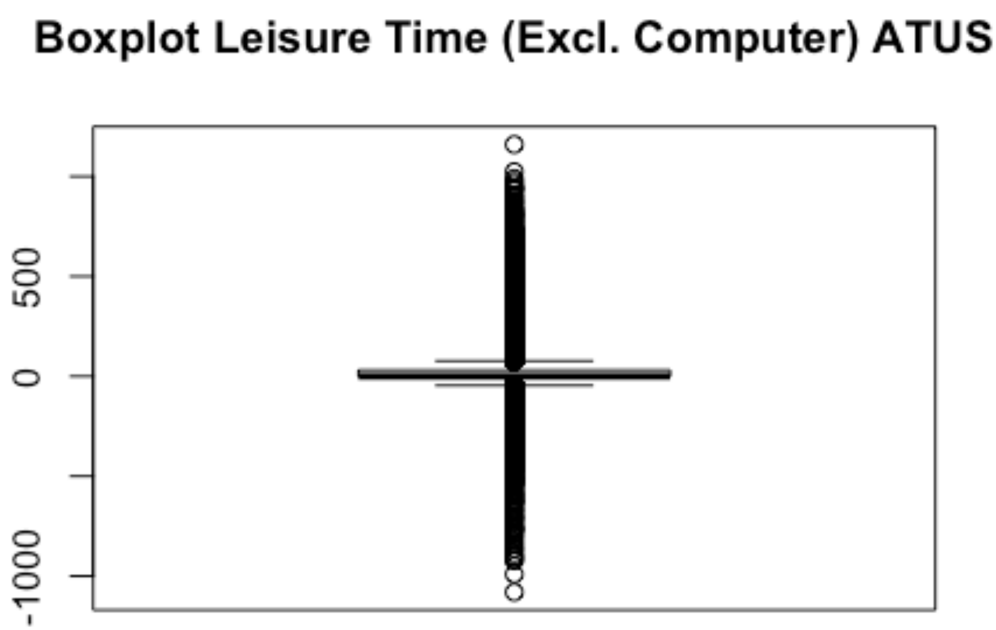
Now we would want to remove outliers in the ATUS (time) portion of the dataset.

First, we visualize the outliers. We notice an odd boxplot: Leisure Time (excl. Computer). This variable seems to have negative values. When looking at the summary, we see there is a value of -1080. It does not make sense to spend -1080 minutes on leisure time. Therefore, we want to delete all rows that have negative numbers.

```
boxplot(ATUS_norm[, 9:ncol(ATUS_norm)], main="Boxplot Time ATUS")
```




```
boxplot(ATUS_norm[, 24], main="Boxplot Leisure Time (Excl. Computer) ATUS")
```




```
summary(ATUS_norm$Leisure_Excl_Computer)

##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -1080.00     0.00     0.00    29.49    31.00  1160.00
```

There are 5769 records with a negative leisure time. Let's remove those rows.

```
neg <- ATUS_norm$Leisure_Excl_Computer < 0
sum(neg==T)

## [1] 5769
```

First we'll add a neg column to dataset, then we want to keep those rows that have negative = FALSE. This leaves us with 53035 records.

```
ATUS_norm <- cbind(ATUS_norm, as.data.frame(neg))
summary(ATUS_norm$neg)

##    Mode   FALSE    TRUE
## logical   53035    5769

ATUS_norm <- subset(ATUS_norm, neg == "FALSE")
summary(ATUS_norm$neg)
```

```
##     Mode   FALSE
## logical   53035

nrow(ATUS_norm)

## [1] 53035
```

Now we drop the column 'neg'.

```
ATUS_norm[,ncol(ATUS_norm)] <- NULL
```

Final Datasets

Finally, we now have CPS (289889 records and 9 variables) and ATUS (53035 records and 25 variables) as our final dataset to perform feature selection and the decision tree model in phase 2.

```
CPS <- CPS_prep
ATUS <- ATUS_norm

c(nrow(CPS), ncol(CPS))

## [1] 289889      9

c(nrow(ATUS), ncol(ATUS))

## [1] 53035    26

write.csv(CPS, "Final_CPS.csv")
write.csv(ATUS, "Final_ATUS.csv")
```

The final datasets looks like this:

```
CPS[1:3,]

##   LABOUR_FORCE_STATUS AGE HISPANIC SEX EDUCATION SPOUSE CHILDREN
## 1                   1  54        2   1        44      1        0
## 2                   1  55        2   2        45      1        0
## 3                   1  57        2   1        39      1        0
##   METROPOLITAN_STATUS HOME_INTERNET_ACCESS
## 1                   1                    1
## 2                   1                    1
## 3                   1                    1

ATUS[1:3,]

##   LABOUR_FORCE_STATUS AGE HISPANIC SEX EDUCATION SPOUSE CHILDREN
## 1                   5  62        2   2        37      2        1
## 2                   1  22        2   2        39      2        0
## 3                   1  33        2   1        36      1        1
```

```
##   METROPOLITAN_STATUS Phone_calls Consumer_Purchases
## 1                   1           0                 20
## 2                   1           0                  0
## 3                   2           0                  0
##   Gov_and_Civic_Obligations HH_Services Professional_Care Volunteer
## 1                         0           0                 0         0
## 2                         0           0                 0         0
## 3                         0           0                 0         0
##   Religion Helping_HH Helping_NONHH Sports Education HH_Activities Travel
## 1        0          0            15      0         0           155      0
## 2        0          0             0      0         0           120      0
## 3        0          0             0      0         0           300      0
##   Personal_Care_Sleep Work Leisure_Excl_Computer Eat_Drink
## 1                 540    0                     0        10
## 2                 600  600                     0        95
## 3                 400    0                     0        25
##   Computer_leisure
## 1                0
## 2                0
## 3                0
```

Additional Functions

Prior to removing the columns, I have created a function to remove outliers, as well as a function and loop to replace any remaining NAs to mean or mode. Below are the function and loop, which we will not be using as we removed all NAs.

This Function Returns the Mode. Mode <- function (x, na.rm) { xtab <- table(x) xmode <- names(which(xtab == max(xtab))) if (length(xmode) > 1) xmode <- ">1 mode" return(xmode) }

We use the Mode Function in the Following Loop for Mean and Mode value Imputation. for (var in 1:ncol(Input_Data4)) { if (class(Input_Data4[,var])%in% c("integer", "numeric")) { Input_Data4[is.na(Input_Data4[,var]),var] <- mean(Input_Data4[,var], na.rm = TRUE) } else if (class(Input_Data4[,var]) %in% c("character", "factor")) { Input_Data4[is.na(Input_Data4[,var]),var] <- Mode(Input_Data4[,var], na.rm = TRUE) } }

Because it won't make sense to remove outliers in this particular dataset, we have kept the data as is. However, if we had to remove outliers, the following function helps: remove_outliers <- function(x, na.rm = TRUE, ...) { qnt <- quantile(x, probs=c(.25, .75), na.rm = na.rm, ...) H <- 1.5 * IQR(x, na.rm = na.rm) y <- x y[x < (qnt[1] - H)] <- NA y[x > (qnt[2] + H)] <- NA y }

Phase 2: Correlation, Decision Tree Building and Internet Access Percentage Comparison

In this section, we want to create a model that will predict variable: HOME INTERNET ACCESS. Before proceeding, we want to look at the correlation between the variables.

Correlation

First, we will look at the correlation between our independent variables as well as the significance levels in order to understand our variables in an attempt to reduce our dataset.

We will use the libraries below as well the function below:

```
library("ggplot2")
library("lattice")
library("Formula")
library("survival")
library("Hmisc")
```

```
##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, round.POSIXt, trunc.POSIXt, units
```

```
flattenCorrMatrix <- function(cormat, pmat) {
  ut <- upper.tri(cormat)
  data.frame(
    row = rownames(cormat)[row(cormat)[ut]],
    column = rownames(cormat)[col(cormat)[ut]],
    cor  =(cormat)[ut],
    p = pmat[ut]
    )
}
```

Below is a table showing the correlation and the correlation and the p value.

This dataset does not have highly correlated variables, hence we do not remove any variable.

```
Correlations <- rcorr(as.matrix(CPS))
flattenCorrMatrix(Correlations$r, Correlations$P)
```

```
##                    row               column          cor            p
## 1  LABOUR_FORCE_STATUS                  AGE  0.257411629 0.0000000000
## 2  LABOUR_FORCE_STATUS             HISPANIC  0.006978087 0.0001718765
## 3                  AGE             HISPANIC  0.140601471 0.0000000000
## 4  LABOUR_FORCE_STATUS                  SEX  0.123676710 0.0000000000
## 5                  AGE                  SEX  0.033636000 0.0000000000
## 6             HISPANIC                  SEX  0.003711284 0.0456946389
## 7  LABOUR_FORCE_STATUS            EDUCATION -0.243129492 0.0000000000
## 8                  AGE            EDUCATION  0.090587601 0.0000000000
## 9             HISPANIC            EDUCATION  0.237461835 0.0000000000
## 10                 SEX            EDUCATION  0.021360371 0.0000000000
## 11 LABOUR_FORCE_STATUS               SPOUSE  0.086016573 0.0000000000
## 12                 AGE               SPOUSE -0.275922388 0.0000000000
## 13            HISPANIC               SPOUSE -0.039414484 0.0000000000
## 14                 SEX               SPOUSE  0.040946145 0.0000000000
## 15           EDUCATION               SPOUSE -0.191730753 0.0000000000
## 16 LABOUR_FORCE_STATUS             CHILDREN -0.163317204 0.0000000000
## 17                 AGE             CHILDREN -0.199073061 0.0000000000
## 18            HISPANIC             CHILDREN -0.095136084 0.0000000000
## 19                 SEX             CHILDREN  0.037577201 0.0000000000
## 20           EDUCATION             CHILDREN  0.074135557 0.0000000000
## 21              SPOUSE             CHILDREN -0.284729689 0.0000000000
## 22 LABOUR_FORCE_STATUS  METROPOLITAN_STATUS  0.025444033 0.0000000000
## 23                 AGE  METROPOLITAN_STATUS  0.053411063 0.0000000000
## 24            HISPANIC  METROPOLITAN_STATUS  0.104046255 0.0000000000
## 25                 SEX  METROPOLITAN_STATUS -0.005345389 0.0040015769
## 26           EDUCATION  METROPOLITAN_STATUS -0.075496979 0.0000000000
## 27              SPOUSE  METROPOLITAN_STATUS -0.037985239 0.0000000000
## 28            CHILDREN  METROPOLITAN_STATUS -0.006100686 0.0010208831
## 29 LABOUR_FORCE_STATUS HOME_INTERNET_ACCESS  0.192443743 0.0000000000
## 30                 AGE HOME_INTERNET_ACCESS  0.175807983 0.0000000000
## 31            HISPANIC HOME_INTERNET_ACCESS -0.093050882 0.0000000000
## 32                 SEX HOME_INTERNET_ACCESS  0.016564004 0.0000000000
## 33           EDUCATION HOME_INTERNET_ACCESS -0.260926008 0.0000000000
## 34              SPOUSE HOME_INTERNET_ACCESS  0.133594140 0.0000000000
## 35            CHILDREN HOME_INTERNET_ACCESS -0.072740041 0.0000000000
## 36 METROPOLITAN_STATUS HOME_INTERNET_ACCESS  0.073360600 0.0000000000
```

Conditional Inference Tree Building

Let's build our model.

```
library(grid)
library(mvtnorm)
library(modeltools)

## Loading required package: stats4

library(stats4)
library(strucchange)

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: sandwich

library(zoo)
library(party)
library(sandwich)
```

We divide our dataset into training and testing (70-30%).

```
train_index <- sample(1:nrow(CPS), 0.7 * nrow(CPS))
train.set <- CPS[train_index,]
test.set <- CPS[-train_index,]
```

Running the model on the training set.

```
internet_ctree_model <- ctree(HOME_INTERNET_ACCESS ~ LABOUR_FORCE_STATUS + AG
E + HISPANIC + SEX + EDUCATION + SPOUSE + CHILDREN  + METROPOLITAN_STATUS, da
ta=train.set)
```

Now let's make our prediction on the test set.

```
internet_ctree_prediction <- predict(internet_ctree_model, test.set)
head(internet_ctree_prediction)

## [1] 1 1 1 2 2 2
## Levels: -1 1 2

table(internet_ctree_prediction, test.set$HOME_INTERNET_ACCESS)

##
## internet_ctree_prediction    -1     1     2
##                        -1     0     0     0
```

```
##                           1       0 63886 14778
##                           2       0  3260  5043
```

Mesuring accuracy, precision, recall and F score.

```
accuracy = (64027+5229)/nrow(test.set)
accuracy

## [1] 0.796348

recall = (5229)/(3169+5229)
recall

## [1] 0.6226482

precision = (5229)/(14542+5229)
precision

## [1] 0.2644783

f1score = 2*((precision*recall)/(precision+recall))
f1score

## [1] 0.3712592
```

Precision is measure of accurateness,i.e., how much percentage of items have we classified correctly. Recall is measure of correctness,i.e., what percentage of items were classified as positive. F-Score is harmonic mean of Precision and Recall.So, ideally model should have F-Score of 1, i.e., both precision and recall should have same values. So a model with f-score closer to 1 is considered as better. Our f score is 0.37.

We will try building a traditional decision tree instead and compare the F score.

Traditional Decision Tree Building

We divide our dataset into training and testing again (70-30%).

```
train_index <- sample(1:nrow(CPS), 0.7 * nrow(CPS))
train.set2 <- CPS[train_index,]
test.set2 <- CPS[-train_index,]
```

Running the model on the training set.

```
library(rpart)
library(randomForest)

## randomForest 4.6-12

## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'

## The following object is masked from 'package:Hmisc':
##
##     combine

## The following object is masked from 'package:ggplot2':
##
##     margin

library(caTools)

internet_rpart_model <- ctree(HOME_INTERNET_ACCESS ~ LABOUR_FORCE_STATUS + AG
E + HISPANIC + SEX + EDUCATION + SPOUSE + CHILDREN  + METROPOLITAN_STATUS, da
ta=train.set2)
```

Now let's make our prediction on the test set.

```
internet_rpart_prediction <- predict(internet_rpart_model, test.set2)
table(internet_rpart_prediction, test.set2$HOME_INTERNET_ACCESS)

##
## internet_rpart_prediction    -1     1     2
##                        -1     0     0     0
##                        1      0 63908 14776
##                        2      0  3207  5076
```

Mesuring accuracy, precision, recall and F score.

```
accuracy2 = (63474+5483)/nrow(test.set2)
accuracy2

## [1] 0.79291

recall2 = (5483)/(3521+5483)
recall2

## [1] 0.6089516

precision2 = (5483)/(14489+5483)
precision2

## [1] 0.2745343

f1score2 = 2*((precision2*recall2)/(precision2+recall2))
f1score2
```

```
## [1] 0.3784511
```

Our f-score for the traditional decision tree model is very close to the conditional inference tree model, equaling 0.38. As such, although it is not a big diffrence, we will use the traditional tree rpart model.

Decision Tree Algorithm Application on the ATUS Dataset

In this section we will look into the precentage of population using the internet during the years 2011-2015 (as per the Current Population Survey US Census). Then we will apply the decision tree model (internet_rpart_model) to the ATUS dataset, see the percentage of people having internet access and see how it compares with the US Census.

Percentage of individuals having access to the internet years: 2011-2015. Over the years 2011-2015, 77.16% of the US population is estimated to have internet access.

```
CPS_Internet <- subset(CPS, HOME_INTERNET_ACCESS == 1)
nrow(CPS_Internet)/nrow(CPS)*100

## [1] 77.1616
```

Let's see what percentage of the ATUS sample will our model predict as having internet access.

```
ATUS_internet_pred <- predict(internet_rpart_model, ATUS)
summary(ATUS_internet_pred)

##    -1     1     2
##     0 46386  6649

sum(ATUS_internet_pred == 1)/length(ATUS_internet_pred)

## [1] 0.87463
```

Percentage Comparison

Our model esimates that 87% of the population. This is way beyond the 77.16% estimate. However, we only took years 2011, 2013 and 2015 (CPS data) into account to get the 77.16% estimate, whereas the ATUS include 2011 up to 2015. This may be one of the reasons why we're getting a higher percentage of people having access to the internet. Furthermore, as per another source, the Pew Research Center, the percentage of population using the internet increased from 83% to 84% from year 2011 to 2015 (http://www.pewinternet.org/fact-sheet/internet-broadband/). Therefore, according to the Pew Research Center, our findings seem more or less accurate.

Phase 3: Linear Regression and Coefficient Analysis

We select rows where ATUS_internet_pred == 1 (TRUE) and we run our 17 version of regression for analysis.

```
ATUS_internet_pred_df <- as.data.frame(ATUS_internet_pred)
ATUS_phase4_prep <- cbind(ATUS, ATUS_internet_pred_df)
ATUS_phase4 <- subset(ATUS_phase4_prep, ATUS_internet_pred ==1)
str(ATUS_phase4)
```

```
## 'data.frame':    46386 obs. of  27 variables:
##  $ LABOUR_FORCE_STATUS      : Factor w/ 5 levels "1","2","3","4",..: 1 1 1
 1 4 1 1 1 1 1 ...
##  $ AGE                      : num  22 33 45 24 29 29 31 35 33 61 ...
##  $ HISPANIC                 : Factor w/ 2 levels "1","2": 2 2 1 2 2 1 2 2
2 2 ...
##  $ SEX                      : Factor w/ 2 levels "1","2": 2 1 1 2 2 1 2 1
2 2 ...
##  $ EDUCATION                : Factor w/ 16 levels "31","32","33",..: 9 6 9
 9 9 10 9 10 10 10 ...
##  $ SPOUSE                   : Factor w/ 2 levels "1","2": 2 1 2 1 2 2 1 2
1 2 ...
##  $ CHILDREN                 : num  0 1 0 2 2 1 0 1 3 0 ...
##  $ METROPOLITAN_STATUS      : Factor w/ 3 levels "1","2","3": 1 2 1 1 2 1
2 2 1 2 ...
##  $ Phone_calls              : num  0 0 0 0 0 0 0 0 0 105 ...
##  $ Consumer_Purchases       : num  0 0 0 0 5 0 0 0 0 25 ...
##  $ Gov_and_Civic_Obligations: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ HH_Services              : num  0 0 0 0 0 0 0 0 15 0 ...
##  $ Professional_Care        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Volunteer                : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Religion                 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Helping_HH               : num  0 0 0 60 120 320 0 0 230 0 ...
##  $ Helping_NONHH            : num  0 0 0 0 0 0 20 0 0 0 ...
##  $ Sports                   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Education                : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ HH_Activities            : num  120 300 2 0 75 70 35 300 445 355 ...
##  $ Travel                   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Personal_Care_Sleep      : num  600 400 540 600 705 670 480 540 575 600
```

```
 ...
## $ Work : num 600 0 0 575 0 0 610 0 0 0 ...
## $ Leisure_Excl_Computer : num 0 0 0 0 0 315 0 270 0 15 ...
## $ Eat_Drink : num 95 25 30 95 25 30 40 60 90 0 ...
## $ Computer_leisure : num 0 0 0 0 0 0 0 0 0 0 ...
## $ ATUS_internet_pred : Factor w/ 3 levels "-1","1","2": 2 2 2 2 2 2
 2 2 2 2 ...
```

Running the Linear Regressions

Now we run the 17 linear regression versions.

1) Phone calls.

```
attach(ATUS_phase4)

## The following objects are masked _by_ .GlobalEnv:
##
##     ATUS_internet_pred, Computer_leisure, Consumer_Purchases,
##     Eat_Drink, Education, Gov_and_Civic_Obligations, Helping_HH,
##     Helping_NONHH, HH_Activities, HH_Services,
##     Leisure_Excl_Computer, Personal_Care_Sleep, Phone_calls,
##     Professional_Care, Religion, Sports, Travel, Volunteer, Work

phonecalls_regr <- lm(Phone_calls~Computer_leisure+LABOUR_FORCE_STATUS+AGE+HI
SPANIC+SEX+EDUCATION+SPOUSE+CHILDREN+METROPOLITAN_STATUS,data=ATUS_phase4)
summary(phonecalls_regr)

##
## Call:
## lm(formula = Phone_calls ~ Computer_leisure + LABOUR_FORCE_STATUS +
##     AGE + HISPANIC + SEX + EDUCATION + SPOUSE + CHILDREN + METROPOLITAN_ST
ATUS,
##     data = ATUS_phase4)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -28.42  -8.12  -4.85  -1.19 651.43
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -3.544586   4.568388  -0.776  0.43782
## Computer_leisure      0.050564   0.010148   4.983 6.29e-07 ***
## LABOUR_FORCE_STATUS2  0.692247   0.680748   1.017  0.30921
## LABOUR_FORCE_STATUS3  6.148049   1.614617   3.808  0.00014 ***
## LABOUR_FORCE_STATUS4  4.447254   0.555839   8.001 1.26e-15 ***
```

```
## LABOUR_FORCE_STATUS5  2.984066   0.277074  10.770  < 2e-16 ***
## AGE                   0.040972   0.008585   4.773 1.82e-06 ***
## HISPANIC2             1.340558   0.336338   3.986 6.74e-05 ***
## SEX2                  3.695078   0.229197  16.122  < 2e-16 ***
## EDUCATION32           0.226806   5.144774   0.044  0.96484
## EDUCATION33           2.565143   4.816400   0.533  0.59432
## EDUCATION34           3.510445   4.667749   0.752  0.45202
## EDUCATION35           3.692260   4.609933   0.801  0.42317
## EDUCATION36           3.287740   4.614130   0.713  0.47614
## EDUCATION37           2.991263   4.599887   0.650  0.51551
## EDUCATION38           3.392473   4.682679   0.724  0.46878
## EDUCATION39           0.913180   4.546414   0.201  0.84081
## EDUCATION40           1.326304   4.548849   0.292  0.77062
## EDUCATION41           0.646211   4.571073   0.141  0.88758
## EDUCATION42           1.422185   4.563701   0.312  0.75532
## EDUCATION43           1.495197   4.547298   0.329  0.74230
## EDUCATION44           3.045406   4.554555   0.669  0.50372
## EDUCATION45           3.849815   4.616555   0.834  0.40433
## EDUCATION46           3.999885   4.606944   0.868  0.38527
## SPOUSE2               4.539943   0.243277  18.662  < 2e-16 ***
## CHILDREN             -0.539891   0.113690  -4.749 2.05e-06 ***
## METROPOLITAN_STATUS2 -0.743305   0.314443  -2.364  0.01809 *
## METROPOLITAN_STATUS3  1.238950   1.264979   0.979  0.32738
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 24 on 46358 degrees of freedom
## Multiple R-squared:  0.02693,    Adjusted R-squared:  0.02636
## F-statistic: 47.51 on 27 and 46358 DF,  p-value: < 2.2e-16
```

2) Consumer Purchases.

```
purchases_regr <- lm(Consumer_Purchases~Computer_leisure+LABOUR_FORCE_STATUS+
AGE+HISPANIC+SEX+EDUCATION+SPOUSE+CHILDREN+METROPOLITAN_STATUS,data=ATUS_phas
e4)
summary(purchases_regr)

##
## Call:
## lm(formula = Consumer_Purchases ~ Computer_leisure + LABOUR_FORCE_STATUS +

##     AGE + HISPANIC + SEX + EDUCATION + SPOUSE + CHILDREN + METROPOLITAN_ST
ATUS,
##     data = ATUS_phase4)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -41.92  -26.47  -19.33    2.87 1022.50
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           26.813133   9.737284   2.754   0.0059 **
## Computer_leisure      -0.008901   0.021630  -0.412   0.6807
## LABOUR_FORCE_STATUS2   2.334525   1.450978   1.609   0.1076
## LABOUR_FORCE_STATUS3   3.603713   3.441473   1.047   0.2950
## LABOUR_FORCE_STATUS4   2.832810   1.184743   2.391   0.0168 *
## LABOUR_FORCE_STATUS5  -0.288421   0.590569  -0.488   0.6253
## AGE                    0.043898   0.018298   2.399   0.0164 *
## HISPANIC2             -5.240836   0.716887  -7.311 2.70e-13 ***
## SEX2                   9.460543   0.488523  19.366  < 2e-16 ***
## EDUCATION32           -8.992341  10.965821  -0.820   0.4122
## EDUCATION33           -3.294574  10.265909  -0.321   0.7483
## EDUCATION34           -6.671791   9.949067  -0.671   0.5025
## EDUCATION35           -8.668453   9.825835  -0.882   0.3777
## EDUCATION36           -8.115086   9.834780  -0.825   0.4093
## EDUCATION37           -5.333321   9.804422  -0.544   0.5865
## EDUCATION38           -4.853982   9.980890  -0.486   0.6267
## EDUCATION39           -4.540464   9.690447  -0.469   0.6394
## EDUCATION40           -3.886998   9.695637  -0.401   0.6885
## EDUCATION41           -2.856846   9.743006  -0.293   0.7694
## EDUCATION42           -3.573719   9.727293  -0.367   0.7133
## EDUCATION43           -0.413910   9.692332  -0.043   0.9659
## EDUCATION44           -0.380969   9.707799  -0.039   0.9687
## EDUCATION45           -5.084430   9.839948  -0.517   0.6054
## EDUCATION46           -2.730741   9.819464  -0.278   0.7809
## SPOUSE2               -2.502995   0.518532  -4.827 1.39e-06 ***
## CHILDREN               0.279628   0.242325   1.154   0.2485
## METROPOLITAN_STATUS2  -2.675927   0.670219  -3.993 6.54e-05 ***
## METROPOLITAN_STATUS3  -0.180268   2.696238  -0.067   0.9467
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 51.16 on 46358 degrees of freedom
## Multiple R-squared:  0.01269,    Adjusted R-squared:  0.01211
## F-statistic: 22.07 on 27 and 46358 DF,  p-value: < 2.2e-16
```

3) Government and civic obligations.

```
GovCivivObligations_regr <- lm(Gov_and_Civic_Obligations~Computer_leisure+LAB
OUR_FORCE_STATUS+AGE+HISPANIC+SEX+EDUCATION+SPOUSE+CHILDREN+METROPOLITAN_STAT
```

```
US,data=ATUS_phase4)
summary(GovCivivObligations_regr)

##
## Call:
## lm(formula = Gov_and_Civic_Obligations ~ Computer_leisure + LABOUR_FORCE_S
TATUS +
##     AGE + HISPANIC + SEX + EDUCATION + SPOUSE + CHILDREN + METROPOLITAN_ST
ATUS,
##     data = ATUS_phase4)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -2.47  -0.45  -0.31  -0.17 463.56
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -0.2229053  1.5755917  -0.141 0.887496
## Computer_leisure     -0.0030381  0.0034999  -0.868 0.385379
## LABOUR_FORCE_STATUS2  0.8221661  0.2347831   3.502 0.000463 ***
## LABOUR_FORCE_STATUS3 -0.2944874  0.5568654  -0.529 0.596926
## LABOUR_FORCE_STATUS4  0.8512956  0.1917035   4.441 8.99e-06 ***
## LABOUR_FORCE_STATUS5  0.0003659  0.0955601   0.004 0.996945
## AGE                   0.0015654  0.0029608   0.529 0.597006
## HISPANIC2            -0.0246463  0.1159996  -0.212 0.831742
## SEX2                  0.0324200  0.0790479   0.410 0.681711
## EDUCATION32           0.0471405  1.7743815   0.027 0.978805
## EDUCATION33           0.0754653  1.6611287   0.045 0.963765
## EDUCATION34           0.1476251  1.6098604   0.092 0.926936
## EDUCATION35           0.0859381  1.5899201   0.054 0.956894
## EDUCATION36          -0.0058406  1.5913676  -0.004 0.997072
## EDUCATION37           0.8692418  1.5864553   0.548 0.583753
## EDUCATION38           0.2949600  1.6150096   0.183 0.855084
## EDUCATION39           0.3933078  1.5680129   0.251 0.801945
## EDUCATION40           0.5022929  1.5688528   0.320 0.748844
## EDUCATION41           0.4564183  1.5765176   0.290 0.772192
## EDUCATION42           0.2459950  1.5739752   0.156 0.875806
## EDUCATION43           0.2890891  1.5683180   0.184 0.853755
## EDUCATION44           0.2898338  1.5708208   0.185 0.853613
## EDUCATION45           0.1023548  1.5922039   0.064 0.948744
## EDUCATION46           0.1415773  1.5888894   0.089 0.928999
## SPOUSE2               0.1997856  0.0839037   2.381 0.017264 *
## CHILDREN              0.0403297  0.0392107   1.029 0.303703
## METROPOLITAN_STATUS2 -0.1847033  0.1084483  -1.703 0.088547 .
```

```
## METROPOLITAN_STATUS3  0.6666349  0.4362788   1.528 0.126519
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.279 on 46358 degrees of freedom
## Multiple R-squared:  0.001351,   Adjusted R-squared:  0.0007697
## F-statistic: 2.323 on 27 and 46358 DF,  p-value: 0.0001153
```

4) Household services.

```
HHServices_regr <- lm(HH_Services ~Computer_leisure+LABOUR_FORCE_STATUS+AGE+H
ISPANIC+SEX+EDUCATION+SPOUSE+CHILDREN+METROPOLITAN_STATUS,data=ATUS_phase4)
summary(HHServices_regr)

##
## Call:
## lm(formula = HH_Services ~ Computer_leisure + LABOUR_FORCE_STATUS +
##     AGE + HISPANIC + SEX + EDUCATION + SPOUSE + CHILDREN + METROPOLITAN_ST
ATUS,
##     data = ATUS_phase4)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -2.02  -1.03  -0.79  -0.51 518.95
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -0.490864   1.756527  -0.279 0.779900
## Computer_leisure        -0.001102   0.003902  -0.283 0.777519
## LABOUR_FORCE_STATUS2 -0.185151   0.261745  -0.707 0.479339
## LABOUR_FORCE_STATUS3  0.380190   0.620814   0.612 0.540272
## LABOUR_FORCE_STATUS4  0.267429   0.213718   1.251 0.210825
## LABOUR_FORCE_STATUS5  0.177421   0.106534   1.665 0.095841 .
## AGE                      0.010933   0.003301   3.312 0.000926 ***
## HISPANIC2               -0.074758   0.129321  -0.578 0.563213
## SEX2                    -0.122575   0.088125  -1.391 0.164258
## EDUCATION32              1.538244   1.978145   0.778 0.436797
## EDUCATION33              0.076131   1.851886   0.041 0.967208
## EDUCATION34              1.232402   1.794731   0.687 0.492289
## EDUCATION35              0.190912   1.772500   0.108 0.914228
## EDUCATION36              0.151838   1.774114   0.086 0.931796
## EDUCATION37              0.886090   1.768638   0.501 0.616372
## EDUCATION38              0.171307   1.800471   0.095 0.924200
## EDUCATION39              0.553666   1.748077   0.317 0.751451
## EDUCATION40              0.621637   1.749014   0.355 0.722276
```

```
## EDUCATION41           0.675900   1.757559   0.385 0.700560
## EDUCATION42           0.762821   1.754724   0.435 0.663765
## EDUCATION43           1.110236   1.748418   0.635 0.525435
## EDUCATION44           1.189317   1.751208   0.679 0.497052
## EDUCATION45           1.046596   1.775046   0.590 0.555451
## EDUCATION46           1.038840   1.771351   0.586 0.557564
## SPOUSE2               0.160854   0.093539   1.720 0.085504 .
## CHILDREN             -0.003177   0.043714  -0.073 0.942058
## METROPOLITAN_STATUS2 -0.015911   0.120902  -0.132 0.895298
## METROPOLITAN_STATUS3  0.402160   0.486379   0.827 0.408330
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.229 on 46358 degrees of freedom
## Multiple R-squared:  0.001617,   Adjusted R-squared:  0.001035
## F-statistic:  2.78 on 27 and 46358 DF,  p-value: 2.091e-06
```

5) Professional Care.

```
ProfessionalCare_regr <- lm(Professional_Care ~Computer_leisure+LABOUR_FORCE_
STATUS+AGE+HISPANIC+SEX+EDUCATION+SPOUSE+CHILDREN+METROPOLITAN_STATUS,data=AT
US_phase4)
summary(ProfessionalCare_regr)
```

```
##
## Call:
## lm(formula = Professional_Care ~ Computer_leisure + LABOUR_FORCE_STATUS +
##     AGE + HISPANIC + SEX + EDUCATION + SPOUSE + CHILDREN + METROPOLITAN_ST
ATUS,
##     data = ATUS_phase4)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -11.13  -5.55  -3.99  -2.35 804.98
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -2.286031   4.750752  -0.481 0.630381
## Computer_leisure     -0.016877   0.010553  -1.599 0.109771
## LABOUR_FORCE_STATUS2  3.053135   0.707922   4.313 1.62e-05 ***
## LABOUR_FORCE_STATUS3  0.655450   1.679070   0.390 0.696268
## LABOUR_FORCE_STATUS4  2.156926   0.578028   3.732 0.000191 ***
## LABOUR_FORCE_STATUS5  1.963985   0.288135   6.816 9.46e-12 ***
## AGE                   0.051588   0.008928   5.778 7.59e-09 ***
## HISPANIC2            -0.557332   0.349764  -1.593 0.111066
```

```
## SEX2                   2.129710   0.238347   8.935  < 2e-16 ***
## EDUCATION32           -0.920189   5.350146  -0.172 0.863444
## EDUCATION33            0.552802   5.008664   0.110 0.912117
## EDUCATION34            0.121929   4.854079   0.025 0.979960
## EDUCATION35            2.289700   4.793955   0.478 0.632921
## EDUCATION36            2.535934   4.798319   0.529 0.597152
## EDUCATION37            2.600007   4.783508   0.544 0.586764
## EDUCATION38            1.997473   4.869605   0.410 0.681667
## EDUCATION39            2.442142   4.727900   0.517 0.605481
## EDUCATION40            2.734943   4.730433   0.578 0.563159
## EDUCATION41            2.904728   4.753544   0.611 0.541159
## EDUCATION42            3.370644   4.745877   0.710 0.477568
## EDUCATION43            3.152271   4.728820   0.667 0.505026
## EDUCATION44            3.711467   4.736366   0.784 0.433273
## EDUCATION45            2.220185   4.800841   0.462 0.643755
## EDUCATION46            4.422777   4.790847   0.923 0.355922
## SPOUSE2                0.659941   0.252988   2.609 0.009095 **
## CHILDREN              -0.225396   0.118229  -1.906 0.056600 .
## METROPOLITAN_STATUS2 -0.768093   0.326995  -2.349 0.018831 *
## METROPOLITAN_STATUS3 -0.903681   1.315475  -0.687 0.492110
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 24.96 on 46358 degrees of freedom
## Multiple R-squared:  0.007061,   Adjusted R-squared:  0.006483
## F-statistic: 12.21 on 27 and 46358 DF,  p-value: < 2.2e-16
```

6) Volunteer.

```
Volunteer_regr <- lm(Volunteer ~Computer_leisure+LABOUR_FORCE_STATUS+AGE+HISP
ANIC+SEX+EDUCATION+SPOUSE+CHILDREN+METROPOLITAN_STATUS,data=ATUS_phase4)
summary(Volunteer_regr)

##
## Call:
## lm(formula = Volunteer ~ Computer_leisure + LABOUR_FORCE_STATUS +
##     AGE + HISPANIC + SEX + EDUCATION + SPOUSE + CHILDREN + METROPOLITAN_ST
ATUS,
##     data = ATUS_phase4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -28.31  -12.71   -9.02   -5.30 1120.41
##
## Coefficients:
```

```
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)           -10.523797   9.695707  -1.085  0.27775
## Computer_leisure        0.005886   0.021537   0.273  0.78464
## LABOUR_FORCE_STATUS2    0.131229   1.444783   0.091  0.92763
## LABOUR_FORCE_STATUS3    0.067187   3.426778   0.020  0.98436
## LABOUR_FORCE_STATUS4    5.675414   1.179684   4.811 1.51e-06 ***
## LABOUR_FORCE_STATUS5    4.059436   0.588048   6.903 5.15e-12 ***
## AGE                     0.137034   0.018220   7.521 5.53e-14 ***
## HISPANIC2               2.785533   0.713826   3.902 9.54e-05 ***
## SEX2                    0.504757   0.486437   1.038  0.29943
## EDUCATION32             1.753097  10.918998   0.161  0.87244
## EDUCATION33             3.602859  10.222074   0.352  0.72450
## EDUCATION34             5.144855   9.906586   0.519  0.60353
## EDUCATION35             8.653160   9.783879   0.884  0.37647
## EDUCATION36             9.557077   9.792786   0.976  0.32910
## EDUCATION37             8.399918   9.762558   0.860  0.38956
## EDUCATION38             5.352710   9.938272   0.539  0.59017
## EDUCATION39             5.641411   9.649069   0.585  0.55878
## EDUCATION40             7.413261   9.654237   0.768  0.44256
## EDUCATION41             7.717152   9.701404   0.795  0.42635
## EDUCATION42            10.703965   9.685759   1.105  0.26911
## EDUCATION43            12.329166   9.650947   1.278  0.20143
## EDUCATION44            14.315511   9.666348   1.481  0.13862
## EDUCATION45            11.745573   9.797933   1.199  0.23062
## EDUCATION46            18.124334   9.777536   1.854  0.06379 .
## SPOUSE2                -1.171858   0.516318  -2.270  0.02323 *
## CHILDREN                1.012740   0.241291   4.197 2.71e-05 ***
## METROPOLITAN_STATUS2    2.217992   0.667357   3.324  0.00089 ***
## METROPOLITAN_STATUS3    7.537340   2.684725   2.807  0.00500 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 50.94 on 46358 degrees of freedom
## Multiple R-squared:  0.009514,   Adjusted R-squared:  0.008937
## F-statistic: 16.49 on 27 and 46358 DF,  p-value: < 2.2e-16
```

7) Religion.

```
Religion_regr <- lm(Religion ~Computer_leisure+LABOUR_FORCE_STATUS+AGE+HISPAN
IC+SEX+EDUCATION+SPOUSE+CHILDREN+METROPOLITAN_STATUS,data=ATUS_phase4)
summary(Religion_regr)
```

```
##
## Call:
## lm(formula = Religion ~ Computer_leisure + LABOUR_FORCE_STATUS +
```

```
##     AGE + HISPANIC + SEX + EDUCATION + SPOUSE + CHILDREN + METROPOLITAN_ST
ATUS,
##     data = ATUS_phase4)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -30.48 -15.10 -11.35  -7.54 999.87
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             5.78504    8.63326   0.670 0.502806
## Computer_leisure       -0.01347    0.01918  -0.702 0.482571
## LABOUR_FORCE_STATUS2    1.87921    1.28647   1.461 0.144090
## LABOUR_FORCE_STATUS3   -3.97776    3.05128  -1.304 0.192363
## LABOUR_FORCE_STATUS4    5.84392    1.05042   5.563 2.66e-08 ***
## LABOUR_FORCE_STATUS5    4.30945    0.52361   8.230  < 2e-16 ***
## AGE                     0.21302    0.01622  13.130  < 2e-16 ***
## HISPANIC2              -0.24351    0.63561  -0.383 0.701639
## SEX2                    3.86619    0.43313   8.926  < 2e-16 ***
## EDUCATION32            -3.70919    9.72251  -0.382 0.702830
## EDUCATION33            -3.89496    9.10195  -0.428 0.668707
## EDUCATION34            -4.46131    8.82103  -0.506 0.613029
## EDUCATION35            -4.14655    8.71177  -0.476 0.634097
## EDUCATION36            -5.43777    8.71970  -0.624 0.532881
## EDUCATION37            -5.55273    8.69279  -0.639 0.522973
## EDUCATION38            -1.17560    8.84925  -0.133 0.894314
## EDUCATION39            -6.95199    8.59174  -0.809 0.418434
## EDUCATION40            -6.76436    8.59634  -0.787 0.431351
## EDUCATION41            -3.92544    8.63834  -0.454 0.649528
## EDUCATION42            -7.35886    8.62440  -0.853 0.393520
## EDUCATION43            -6.06532    8.59341  -0.706 0.480310
## EDUCATION44            -7.01563    8.60712  -0.815 0.415022
## EDUCATION45            -9.37695    8.72429  -1.075 0.282465
## EDUCATION46            -8.72770    8.70612  -1.002 0.316118
## SPOUSE2                -1.35196    0.45974  -2.941 0.003276 **
## CHILDREN                0.78598    0.21485   3.658 0.000254 ***
## METROPOLITAN_STATUS2    1.33369    0.59423   2.244 0.024811 *
## METROPOLITAN_STATUS3    5.42464    2.39054   2.269 0.023260 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 45.36 on 46358 degrees of freedom
```

```
## Multiple R-squared:  0.01165,    Adjusted R-squared:  0.01107
## F-statistic: 20.24 on 27 and 46358 DF,  p-value: < 2.2e-16
```

8) Helping Non Household members.

```
HelpingNonHH_regr <- lm(Helping_NONHH ~Computer_leisure+LABOUR_FORCE_STATUS+A
GE+HISPANIC+SEX+EDUCATION+SPOUSE+CHILDREN+METROPOLITAN_STATUS,data=ATUS_phase
4)
summary(HelpingNonHH_regr)
```

```
##
## Call:
## lm(formula = Helping_NONHH ~ Computer_leisure + LABOUR_FORCE_STATUS +
##     AGE + HISPANIC + SEX + EDUCATION + SPOUSE + CHILDREN + METROPOLITAN_ST
ATUS,
##     data = ATUS_phase4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -23.38  -11.39   -7.98   -4.31 1239.26
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           0.18721    8.62191   0.022  0.98268
## Computer_leisure      0.01364    0.01915   0.712  0.47630
## LABOUR_FORCE_STATUS2  3.61682    1.28477   2.815  0.00488 **
## LABOUR_FORCE_STATUS3  9.73262    3.04726   3.194  0.00140 **
## LABOUR_FORCE_STATUS4  6.34568    1.04903   6.049 1.47e-09 ***
## LABOUR_FORCE_STATUS5  2.32958    0.52292   4.455 8.41e-06 ***
## AGE                   0.07651    0.01620   4.722 2.34e-06 ***
## HISPANIC2             1.35144    0.63477   2.129  0.03326 *
## SEX2                  1.06075    0.43256   2.452  0.01420 *
## EDUCATION32          -1.81824    9.70972  -0.187  0.85146
## EDUCATION33           0.84621    9.08998   0.093  0.92583
## EDUCATION34           5.51931    8.80943   0.627  0.53098
## EDUCATION35           3.51145    8.70032   0.404  0.68651
## EDUCATION36           4.98336    8.70824   0.572  0.56715
## EDUCATION37           2.54240    8.68136   0.293  0.76963
## EDUCATION38           1.01330    8.83761   0.115  0.90872
## EDUCATION39           5.23069    8.58044   0.610  0.54213
## EDUCATION40           3.97305    8.58503   0.463  0.64352
## EDUCATION41           3.93462    8.62698   0.456  0.64833
## EDUCATION42           2.78793    8.61306   0.324  0.74618
## EDUCATION43           1.50292    8.58211   0.175  0.86098
## EDUCATION44           2.39258    8.59580   0.278  0.78075
```

```
## EDUCATION45           1.83740    8.71281   0.211  0.83298
## EDUCATION46           1.26105    8.69468   0.145  0.88468
## SPOUSE2               1.01464    0.45914   2.210  0.02712 *
## CHILDREN             -2.08541    0.21457  -9.719  < 2e-16 ***
## METROPOLITAN_STATUS2  1.78741    0.59345   3.012  0.00260 **
## METROPOLITAN_STATUS3  1.93642    2.38739   0.811  0.41731
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 45.3 on 46358 degrees of freedom
## Multiple R-squared:  0.008829,   Adjusted R-squared:  0.008252
## F-statistic: 15.29 on 27 and 46358 DF,  p-value: < 2.2e-16
```

9) Eating and drinking.

```
EatDrink_regr <- lm(Eat_Drink ~Computer_leisure+LABOUR_FORCE_STATUS+AGE+HISPA
NIC+SEX+EDUCATION+SPOUSE+CHILDREN+METROPOLITAN_STATUS,data=ATUS_phase4)
summary(EatDrink_regr)
```

```
##
## Call:
## lm(formula = Eat_Drink ~ Computer_leisure + LABOUR_FORCE_STATUS +
##     AGE + HISPANIC + SEX + EDUCATION + SPOUSE + CHILDREN + METROPOLITAN_ST
ATUS,
##     data = ATUS_phase4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -95.70  -35.12   -9.67   23.13 1249.82
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           73.678879   9.881645   7.456 9.06e-14 ***
## Computer_leisure      -0.004003   0.021951  -0.182  0.85530
## LABOUR_FORCE_STATUS2   4.731723   1.472490   3.213  0.00131 **
## LABOUR_FORCE_STATUS3   2.432409   3.492495   0.696  0.48614
## LABOUR_FORCE_STATUS4  -0.230297   1.202308  -0.192  0.84810
## LABOUR_FORCE_STATUS5   5.041665   0.599325   8.412  < 2e-16 ***
## AGE                    0.174925   0.018570   9.420  < 2e-16 ***
## HISPANIC2             -5.525589   0.727515  -7.595 3.13e-14 ***
## SEX2                  -4.922587   0.495765  -9.929  < 2e-16 ***
## EDUCATION32          -11.405577  11.128396  -1.025  0.30541
## EDUCATION33          -13.599346  10.418108  -1.305  0.19178
## EDUCATION34           -9.119648  10.096568  -0.903  0.36640
## EDUCATION35           -4.913339   9.971509  -0.493  0.62220
```

```
## EDUCATION36           -7.928925   9.980587  -0.794  0.42695
## EDUCATION37          -11.121797   9.949778  -1.118  0.26366
## EDUCATION38          -12.357319  10.128863  -1.220  0.22247
## EDUCATION39           -8.574480   9.834113  -0.872  0.38326
## EDUCATION40           -6.235378   9.839381  -0.634  0.52627
## EDUCATION41           -4.900370   9.887452  -0.496  0.62017
## EDUCATION42           -2.365896   9.871507  -0.240  0.81059
## EDUCATION43            3.039960   9.836027   0.309  0.75727
## EDUCATION44            5.953103   9.851723   0.604  0.54567
## EDUCATION45            7.325287   9.985832   0.734  0.46322
## EDUCATION46           10.955722   9.965044   1.099  0.27159
## SPOUSE2               -9.037094   0.526219 -17.174  < 2e-16 ***
## CHILDREN              -0.594915   0.245918  -2.419  0.01556 *
## METROPOLITAN_STATUS2  -3.062712   0.680156  -4.503 6.72e-06 ***
## METROPOLITAN_STATUS3  -1.026643   2.736212  -0.375  0.70751
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 51.92 on 46358 degrees of freedom
## Multiple R-squared:  0.03321,    Adjusted R-squared:  0.03265
## F-statistic: 58.99 on 27 and 46358 DF,  p-value: < 2.2e-16
```

10) Helping household members.

```
HelpingHH_regr <- lm(Helping_HH~Computer_leisure+LABOUR_FORCE_STATUS+AGE+HISP
ANIC+SEX+EDUCATION+SPOUSE+CHILDREN+METROPOLITAN_STATUS,data=ATUS_phase4)
summary(EatDrink_regr)

##
## Call:
## lm(formula = Helping_HH ~ Computer_leisure + LABOUR_FORCE_STATUS +

    AGE + HISPANIC + SEX + EDUCATION + SPOUSE + CHILDREN + METROPOLITAN_STATU
S,

    data = ATUS_phase4)


##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -287.11  -32.42  -10.05    9.27 1209.77


##
## Coefficients:
```

```
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              31.69530   11.25199   2.817  0.00485 **
## Computer_leisure         -0.09164    0.03078  -2.977  0.00291 **
## LABOUR_FORCE_STATUS2     18.21030    2.09714   8.683  < 2e-16 ***
## LABOUR_FORCE_STATUS3     20.47601    4.90537   4.174 3.00e-05 ***
## LABOUR_FORCE_STATUS4     13.87108    1.65778   8.367  < 2e-16 ***
## LABOUR_FORCE_STATUS5     19.41277    0.84648  22.934  < 2e-16 ***
## AGE                      -0.88515    0.02592 -34.146  < 2e-16 ***
## HISPANIC2                 5.48552    1.03930   5.278 1.31e-07 ***
## SEX2                     17.02742    0.70084  24.296  < 2e-16 ***
## EDUCATION32              -0.96745   12.81416  -0.075  0.93982
## EDUCATION33               0.89896   11.94732   0.075  0.94002
## EDUCATION34             -36.48976   11.59576  -3.147  0.00165 **
## EDUCATION35             -46.03952   11.43978  -4.025 5.72e-05 ***
## EDUCATION36             -43.40110   11.39632  -3.808  0.00014 ***
## EDUCATION37             -21.95017   11.37567  -1.930  0.05367 .
## EDUCATION38              -8.47427   11.69719  -0.724  0.46878
## EDUCATION39               9.89062   11.15223   0.887  0.37515
## EDUCATION40              12.98027   11.16114   1.163  0.24484
## EDUCATION41              13.07489   11.24634   1.163  0.24500
## EDUCATION42              16.38372   11.21819   1.460  0.14417
## EDUCATION43              20.99804   11.15916   1.882  0.05988 .
## EDUCATION44              23.95599   11.18612   2.142  0.03223 *
## EDUCATION45              21.63426   11.42676   1.893  0.05832 .
## EDUCATION46              31.25670   11.38953   2.744  0.00607 **
## SPOUSE2                 -21.84961    0.74088 -29.491  < 2e-16 ***
## CHILDREN                 24.78393    0.34889  71.036  < 2e-16 ***
## METROPOLITAN_STATUS2     -2.23902    0.96605  -2.318  0.02047 *
## METROPOLITAN_STATUS3      0.28271    3.89557   0.073  0.94215

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
Residual standard error: 73.72 on 46961 degrees of freedom Multiple R-squared
:  0.2266,  Adjusted R-squared:  0.2261 F-statistic: 509.6 on 27 and 46961 DF
,  p-value: < 2.2e-16
```

11) Sports.

```
Sports_regr <- lm(Sports~Computer_leisure+LABOUR_FORCE_STATUS+AGE+HISPANIC+SE
X+EDUCATION+SPOUSE+CHILDREN+METROPOLITAN_STATUS,data=ATUS_phase4)
summary(Sports_regr)

##
## Call:
```

```
## lm(formula = Sports ~ Computer_leisure + LABOUR_FORCE_STATUS +
##     AGE + HISPANIC + SEX + EDUCATION + SPOUSE + CHILDREN + METROPOLITAN_ST
ATUS,
##     data = ATUS_phase4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -57.86  -24.02  -16.77   -7.54 1100.84
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            30.64364   11.45459   2.675 0.007470 **
## Computer_leisure       -0.06359    0.02544  -2.499 0.012452 *
## LABOUR_FORCE_STATUS2    7.57107    1.70688   4.436 9.20e-06 ***
## LABOUR_FORCE_STATUS3    3.79069    4.04842   0.936 0.349104
## LABOUR_FORCE_STATUS4    4.96739    1.39369   3.564 0.000365 ***
## LABOUR_FORCE_STATUS5    4.46388    0.69472   6.425 1.33e-10 ***
## AGE                    -0.31784    0.02153 -14.766  < 2e-16 ***
## HISPANIC2               3.72505    0.84332   4.417 1.00e-05 ***
## SEX2                  -11.36297    0.57468 -19.773  < 2e-16 ***
## EDUCATION32             9.49269   12.89979   0.736 0.461808
## EDUCATION33            -0.19208   12.07644  -0.016 0.987310
## EDUCATION34            13.33359   11.70372   1.139 0.254600
## EDUCATION35            18.25352   11.55875   1.579 0.114298
## EDUCATION36            17.84302   11.56928   1.542 0.123013
## EDUCATION37             9.85538   11.53357   0.854 0.392835
## EDUCATION38             3.49886   11.74116   0.298 0.765705
## EDUCATION39             1.16377   11.39949   0.102 0.918686
## EDUCATION40             2.87588   11.40559   0.252 0.800929
## EDUCATION41             5.71710   11.46132   0.499 0.617911
## EDUCATION42             5.06023   11.44283   0.442 0.658333
## EDUCATION43             9.74863   11.40171   0.855 0.392547
## EDUCATION44            11.71749   11.41990   1.026 0.304869
## EDUCATION45            11.73634   11.57536   1.014 0.310632
## EDUCATION46            15.56385   11.55126   1.347 0.177867
## SPOUSE2                 0.21178    0.60998   0.347 0.728454
## CHILDREN               -1.25514    0.28506  -4.403 1.07e-05 ***
## METROPOLITAN_STATUS2    1.70062    0.78842   2.157 0.031011 *
## METROPOLITAN_STATUS3    7.31001    3.17176   2.305 0.021187 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 60.19 on 46358 degrees of freedom
```

```
## Multiple R-squared:  0.02251,    Adjusted R-squared:  0.02194
## F-statistic: 39.54 on 27 and 46358 DF,  p-value: < 2.2e-16
```

12) Education.

```
Education_regr <- lm(Education~Computer_leisure+LABOUR_FORCE_STATUS+AGE+HISPA
NIC+SEX+EDUCATION+SPOUSE+CHILDREN+METROPOLITAN_STATUS,data=ATUS_phase4)
summary(Education_regr)

##
## Call:
## lm(formula = Education ~ Computer_leisure + LABOUR_FORCE_STATUS +
##     AGE + HISPANIC + SEX + EDUCATION + SPOUSE + CHILDREN + METROPOLITAN_ST
ATUS,
##     data = ATUS_phase4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -138.26  -22.91   -7.44    5.34  923.16
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          68.45566   14.23423   4.809 1.52e-06 ***
## Computer_leisure     -0.10285    0.03162  -3.253 0.001144 **
## LABOUR_FORCE_STATUS2  6.04355    2.12108   2.849 0.004384 **
## LABOUR_FORCE_STATUS3  9.50453    5.03084   1.889 0.058864 .
## LABOUR_FORCE_STATUS4 14.90933    1.73189   8.609  < 2e-16 ***
## LABOUR_FORCE_STATUS5 26.22225    0.86331  30.374  < 2e-16 ***
## AGE                  -1.15878    0.02675 -43.321  < 2e-16 ***
## HISPANIC2             6.09696    1.04796   5.818 6.00e-09 ***
## SEX2                 -1.67755    0.71414  -2.349 0.018825 *
## EDUCATION32          -8.05645   16.03014  -0.503 0.615261
## EDUCATION33          -8.09162   15.00699  -0.539 0.589758
## EDUCATION34          21.17369   14.54382   1.456 0.145440
## EDUCATION35          44.17917   14.36367   3.076 0.002101 **
## EDUCATION36          43.51628   14.37675   3.027 0.002473 **
## EDUCATION37          26.44001   14.33237   1.845 0.065077 .
## EDUCATION38         -11.74610   14.59034  -0.805 0.420789
## EDUCATION39         -25.50331   14.16576  -1.800 0.071812 .
## EDUCATION40         -14.62025   14.17335  -1.032 0.302297
## EDUCATION41         -23.72475   14.24259  -1.666 0.095768 .
## EDUCATION42         -17.89997   14.21962  -1.259 0.208101
## EDUCATION43         -19.01219   14.16852  -1.342 0.179647
## EDUCATION44         -17.80353   14.19113  -1.255 0.209647
## EDUCATION45         -16.71701   14.38431  -1.162 0.245172
```

```
## EDUCATION46          -17.98549   14.35436  -1.253 0.210225    
## SPOUSE2               12.39407    0.75800  16.351  < 2e-16 ***
## CHILDREN              -1.70619    0.35424  -4.817 1.47e-06 ***
## METROPOLITAN_STATUS2  -3.28416    0.97974  -3.352 0.000803 ***
## METROPOLITAN_STATUS3   4.35972    3.94143   1.106 0.268678    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 74.79 on 46358 degrees of freedom
## Multiple R-squared:  0.1381, Adjusted R-squared:  0.1376 
## F-statistic: 275.1 on 27 and 46358 DF,  p-value: < 2.2e-16
```

13) Household Activities.

```
HHActivities_regr <- lm(HH_Activities~Computer_leisure+LABOUR_FORCE_STATUS+AG
E+HISPANIC+SEX+EDUCATION+SPOUSE+CHILDREN+METROPOLITAN_STATUS,data=ATUS_phase4
)
summary(HHActivities_regr)

## 
## Call:
## lm(formula = HH_Activities ~ Computer_leisure + LABOUR_FORCE_STATUS + 
##     AGE + HISPANIC + SEX + EDUCATION + SPOUSE + CHILDREN + METROPOLITAN_ST
ATUS, 
##     data = ATUS_phase4)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -218.12  -89.02  -41.04   50.84 1284.12 
## 
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)    
## (Intercept)            2.88004   25.56725   0.113 0.910312    
## Computer_leisure      -0.18831    0.05679  -3.316 0.000915 ***
## LABOUR_FORCE_STATUS2  28.10212    3.80984   7.376 1.66e-13 ***
## LABOUR_FORCE_STATUS3  70.09582    9.03630   7.757 8.86e-15 ***
## LABOUR_FORCE_STATUS4  43.01551    3.11079  13.828  < 2e-16 ***
## LABOUR_FORCE_STATUS5  28.32895    1.55066  18.269  < 2e-16 ***
## AGE                    1.20776    0.04805  25.138  < 2e-16 ***
## HISPANIC2             -9.89574    1.88233  -5.257 1.47e-07 ***
## SEX2                  40.81200    1.28272  31.817  < 2e-16 ***
## EDUCATION32           32.44136   28.79303   1.127 0.259871    
## EDUCATION33           25.53738   26.95527   0.947 0.343441    
## EDUCATION34           16.10612   26.12333   0.617 0.537540    
## EDUCATION35           13.56580   25.79976   0.526 0.599022    
```

```
## EDUCATION36            8.39975   25.82325   0.325 0.744972
## EDUCATION37           12.97222   25.74353   0.504 0.614333
## EDUCATION38           33.36739   26.20689   1.273 0.202943
## EDUCATION39           41.42855   25.44427   1.628 0.103488
## EDUCATION40           38.40894   25.45790   1.509 0.131376
## EDUCATION41           48.29710   25.58228   1.888 0.059044 .
## EDUCATION42           43.75837   25.54102   1.713 0.086672 .
## EDUCATION43           39.66081   25.44922   1.558 0.119138
## EDUCATION44           35.52917   25.48983   1.394 0.163368
## EDUCATION45           23.65970   25.83682   0.916 0.359810
## EDUCATION46           36.78948   25.78303   1.427 0.153619
## SPOUSE2              -24.84821    1.36151 -18.250  < 2e-16 ***
## CHILDREN               7.84275    0.63628  12.326  < 2e-16 ***
## METROPOLITAN_STATUS2  14.39822    1.75980   8.182 2.87e-16 ***
## METROPOLITAN_STATUS3   5.91030    7.07953   0.835 0.403810
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 134.3 on 46358 degrees of freedom
## Multiple R-squared:  0.07691,    Adjusted R-squared:  0.07637
## F-statistic: 143.1 on 27 and 46358 DF,  p-value: < 2.2e-16
```

14) Travel.

```
Travel_regr <- lm(Travel~Computer_leisure+LABOUR_FORCE_STATUS+AGE+HISPANIC+SE
X+EDUCATION+SPOUSE+CHILDREN+METROPOLITAN_STATUS,data=ATUS_phase4)
summary(Travel_regr)
```

```
##
## Call:
## lm(formula = Travel ~ Computer_leisure + LABOUR_FORCE_STATUS +
##     AGE + HISPANIC + SEX + EDUCATION + SPOUSE + CHILDREN + METROPOLITAN_ST
ATUS,
##     data = ATUS_phase4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -0.586  -0.118  -0.063  -0.009 239.517
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           0.1013897  0.5271055   0.192 0.847468
## Computer_leisure     -0.0010751  0.0011709  -0.918 0.358503
## LABOUR_FORCE_STATUS2  0.2892774  0.0785454   3.683 0.000231 ***
## LABOUR_FORCE_STATUS3 -0.0646419  0.1862962  -0.347 0.728604
```

```
## LABOUR_FORCE_STATUS4 -0.0184783  0.0641333  -0.288 0.773253
## LABOUR_FORCE_STATUS5 -0.0023566  0.0319691  -0.074 0.941238
## AGE                  -0.0025848  0.0009905  -2.609 0.009071 **
## HISPANIC2            -0.0246209  0.0388070  -0.634 0.525794
## SEX2                  0.0440503  0.0264450   1.666 0.095774 .
## EDUCATION32          -0.0104452  0.5936095  -0.018 0.985961
## EDUCATION33          -0.0104005  0.5557214  -0.019 0.985068
## EDUCATION34          -0.0237548  0.5385699  -0.044 0.964819
## EDUCATION35          -0.0312461  0.5318989  -0.059 0.953156
## EDUCATION36           0.0921066  0.5323832   0.173 0.862646
## EDUCATION37           0.1289516  0.5307398   0.243 0.808033
## EDUCATION38           0.1639813  0.5402925   0.304 0.761507
## EDUCATION39           0.0566443  0.5245700   0.108 0.914010
## EDUCATION40           0.0637512  0.5248510   0.121 0.903323
## EDUCATION41           0.0660404  0.5274152   0.125 0.900354
## EDUCATION42           0.1479065  0.5265646   0.281 0.778796
## EDUCATION43           0.1308674  0.5246721   0.249 0.803032
## EDUCATION44           0.0945749  0.5255094   0.180 0.857178
## EDUCATION45           0.2092545  0.5326630   0.393 0.694435
## EDUCATION46           0.0331943  0.5315541   0.062 0.950207
## SPOUSE2              -0.0348883  0.0280695  -1.243 0.213902
## CHILDREN              0.0110021  0.0131177   0.839 0.401629
## METROPOLITAN_STATUS2 -0.0096205  0.0362808  -0.265 0.790881
## METROPOLITAN_STATUS3 -0.0704329  0.1459546  -0.483 0.629405
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.77 on 46358 degrees of freedom
## Multiple R-squared:  0.0009812,  Adjusted R-squared:  0.0003994
## F-statistic: 1.686 on 27 and 46358 DF,  p-value: 0.01431
```

15) Personal Care (incl. sleep).

```
PersCare_Sleep_regr <- lm(Personal_Care_Sleep~Computer_leisure+LABOUR_FORCE_S
TATUS+AGE+HISPANIC+SEX+EDUCATION+SPOUSE+CHILDREN+METROPOLITAN_STATUS,data=ATU
S_phase4)
summary(PersCare_Sleep_regr)

##
## Call:
## lm(formula = Personal_Care_Sleep ~ Computer_leisure + LABOUR_FORCE_STATUS
+
##     AGE + HISPANIC + SEX + EDUCATION + SPOUSE + CHILDREN + METROPOLITAN_ST
ATUS,
##     data = ATUS_phase4)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -650.83  -80.18   -8.76   70.04  895.68
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           663.73078   26.23350  25.301  < 2e-16 ***
## Computer_leisure       -0.28394    0.05827  -4.872 1.11e-06 ***
## LABOUR_FORCE_STATUS2   37.94432    3.90912   9.707  < 2e-16 ***
## LABOUR_FORCE_STATUS3   35.31501    9.27177   3.809 0.000140 ***
## LABOUR_FORCE_STATUS4   27.74010    3.19185   8.691  < 2e-16 ***
## LABOUR_FORCE_STATUS5   44.27410    1.59107  27.827  < 2e-16 ***
## AGE                    -0.72524    0.04930 -14.711  < 2e-16 ***
## HISPANIC2             -14.93907    1.93139  -7.735 1.06e-14 ***
## SEX2                   22.72991    1.31614  17.270  < 2e-16 ***
## EDUCATION32           -47.51111   29.54334  -1.608 0.107802
## EDUCATION33           -38.50794   27.65769  -1.392 0.163837
## EDUCATION34           -28.90426   26.80407  -1.078 0.280882
## EDUCATION35           -40.58884   26.47207  -1.533 0.125216
## EDUCATION36           -31.99196   26.49617  -1.207 0.227277
## EDUCATION37           -36.14246   26.41438  -1.368 0.171229
## EDUCATION38           -40.25957   26.88981  -1.497 0.134347
## EDUCATION39           -58.37928   26.10732  -2.236 0.025348 *
## EDUCATION40           -66.67045   26.12130  -2.552 0.010703 *
## EDUCATION41           -62.34605   26.24892  -2.375 0.017544 *
## EDUCATION42           -72.52484   26.20659  -2.767 0.005652 **
## EDUCATION43           -76.85211   26.11240  -2.943 0.003251 **
## EDUCATION44           -82.24086   26.15407  -3.144 0.001665 **
## EDUCATION45           -87.66677   26.51009  -3.307 0.000944 ***
## EDUCATION46           -81.05780   26.45491  -3.064 0.002185 **
## SPOUSE2                12.80282    1.39699   9.165  < 2e-16 ***
## CHILDREN               -9.03159    0.65286 -13.834  < 2e-16 ***
## METROPOLITAN_STATUS2   -2.70819    1.80566  -1.500 0.133663
## METROPOLITAN_STATUS3   -9.60051    7.26401  -1.322 0.186290
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 137.8 on 46358 degrees of freedom
## Multiple R-squared:  0.0561, Adjusted R-squared:  0.05555
## F-statistic:   102 on 27 and 46358 DF,  p-value: < 2.2e-16
```

16) Work Activities.

```
Work_regr <- glm(Work~Computer_leisure+LABOUR_FORCE_STATUS+AGE+HISPANIC+SEX+E
DUCATION+SPOUSE+CHILDREN+METROPOLITAN_STATUS,data=ATUS_phase4)
summary(Work_regr)
```

```
##
## Call:
## glm(formula = Work ~ Computer_leisure + LABOUR_FORCE_STATUS +
##     AGE + HISPANIC + SEX + EDUCATION + SPOUSE + CHILDREN + METROPOLITAN_ST
ATUS,
##     data = ATUS_phase4)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -327.48  -244.72    -1.32   170.04  1112.93
##
## Coefficients:
##                        Estimate Std. Error  t value Pr(>|t|)
## (Intercept)           271.49946   41.11723    6.603 4.07e-11 ***
## Computer_leisure       -0.97216    0.09134  -10.644  < 2e-16 ***
## LABOUR_FORCE_STATUS2 -227.34000    6.12699  -37.105  < 2e-16 ***
## LABOUR_FORCE_STATUS3 -236.82157   14.53217  -16.296  < 2e-16 ***
## LABOUR_FORCE_STATUS4 -229.91456    5.00277  -45.957  < 2e-16 ***
## LABOUR_FORCE_STATUS5 -253.47991    2.49377 -101.645  < 2e-16 ***
## AGE                    -0.28153    0.07727   -3.644 0.000269 ***
## HISPANIC2              -2.67225    3.02717   -0.883 0.377374
## SEX2                  -31.93841    2.06286  -15.483  < 2e-16 ***
## EDUCATION32            54.23390   46.30493    1.171 0.241511
## EDUCATION33            25.37352   43.34944    0.585 0.558332
## EDUCATION34             4.56631   42.01152    0.109 0.913447
## EDUCATION35             1.94289   41.49115    0.047 0.962652
## EDUCATION36           -10.16298   41.52893   -0.245 0.806674
## EDUCATION37            -4.11387   41.40073   -0.099 0.920847
## EDUCATION38             8.32966   42.14590    0.198 0.843329
## EDUCATION39            27.66457   40.91946    0.676 0.498997
## EDUCATION40            30.87080   40.94137    0.754 0.450838
## EDUCATION41            24.82447   41.14140    0.603 0.546250
## EDUCATION42            31.14679   41.07505    0.758 0.448281
## EDUCATION43            28.07113   40.92742    0.686 0.492795
## EDUCATION44            32.53611   40.99273    0.794 0.427372
## EDUCATION45            62.22286   41.55075    1.498 0.134266
## EDUCATION46            41.63738   41.46425    1.004 0.315299
## SPOUSE2                 1.92002    2.18958    0.877 0.380551
## CHILDREN               -2.28737    1.02326   -2.235 0.025397 *
## METROPOLITAN_STATUS2    1.08938    2.83011    0.385 0.700295
```

```
## METROPOLITAN_STATUS3    5.95303   11.38530    0.523 0.601067
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 46673.32)
##
##     Null deviance: 2892482706  on 46385  degrees of freedom
## Residual deviance: 2163681574  on 46358  degrees of freedom
## AIC: 630360
##
## Number of Fisher Scoring iterations: 2
```

17) Leisure (excluding computer).

```
LeisureNoComputer_regr <- glm(Leisure_Excl_Computer~Computer_leisure+LABOUR_F
ORCE_STATUS+AGE+HISPANIC+SEX+EDUCATION+SPOUSE+CHILDREN+METROPOLITAN_STATUS,da
ta=ATUS_phase4)
summary(LeisureNoComputer_regr)

##
## Call:
## glm(formula = Leisure_Excl_Computer ~ Computer_leisure + LABOUR_FORCE_STAT
US +
##     AGE + HISPANIC + SEX + EDUCATION + SPOUSE + CHILDREN + METROPOLITAN_ST
ATUS,
##     data = ATUS_phase4)
##
## Deviance Residuals:
##     Min       1Q    Median       3Q      Max
## -208.77   -42.46   -35.42     3.68  1118.31
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             48.01643   16.83673   2.852 0.004348 **
## Computer_leisure         0.55118    0.03740  14.737  < 2e-16 ***
## LABOUR_FORCE_STATUS2 21.39678    2.50889   8.528  < 2e-16 ***
## LABOUR_FORCE_STATUS3 16.76275    5.95065   2.817 0.004850 **
## LABOUR_FORCE_STATUS4 17.65933    2.04854   8.620  < 2e-16 ***
## LABOUR_FORCE_STATUS5 11.91278    1.02115  11.666  < 2e-16 ***
## AGE                    -0.32336    0.03164 -10.220  < 2e-16 ***
## HISPANIC2              -2.53103    1.23957  -2.042 0.041171 *
## SEX2                    2.26743    0.84470   2.684 0.007271 **
## EDUCATION32            -3.85020   18.96099  -0.203 0.839090
## EDUCATION33             8.28116   17.75077   0.467 0.640843
## EDUCATION34            -3.59966   17.20292  -0.209 0.834256
```

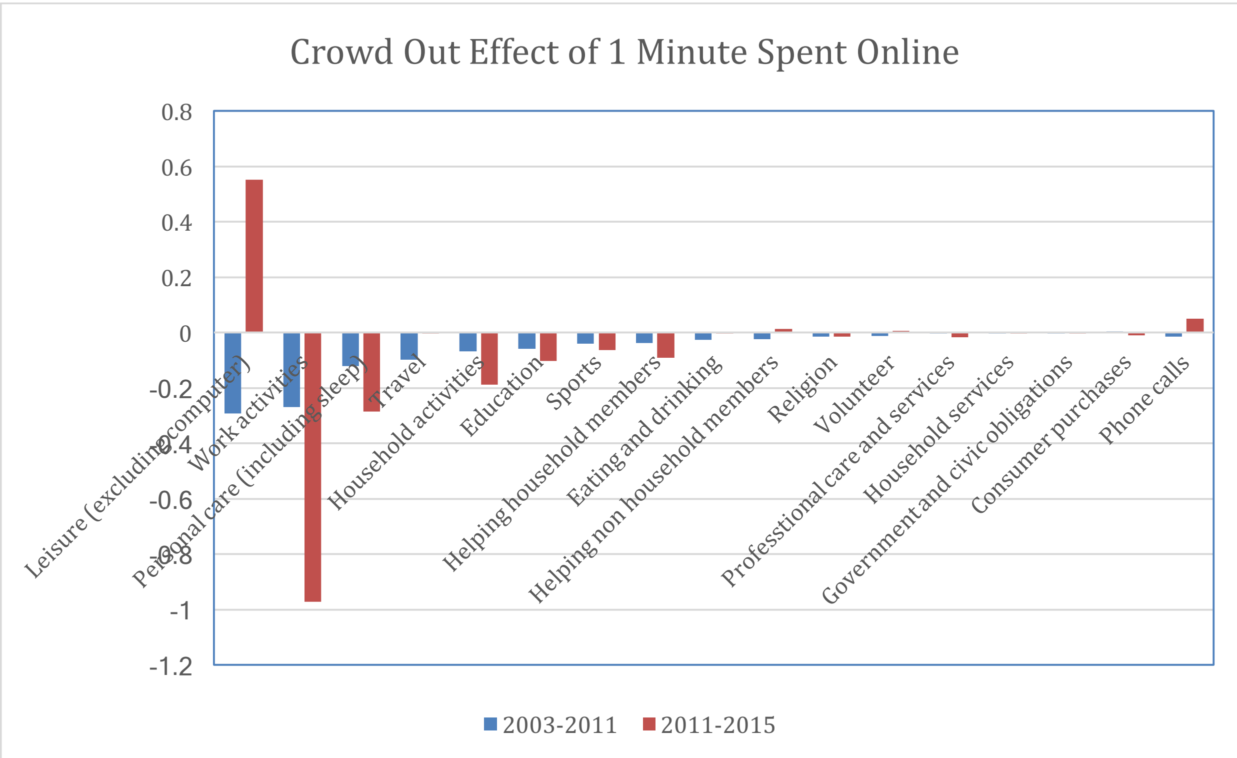
```
## EDUCATION35          2.66633   16.98984   0.157 0.875296    
## EDUCATION36          4.60540   17.00531   0.271 0.786529    
## EDUCATION37          4.75894   16.95281   0.281 0.778929    
## EDUCATION38          5.90865   17.25795   0.342 0.732072    
## EDUCATION39          6.28187   16.75574   0.375 0.707730    
## EDUCATION40          5.59617   16.76472   0.334 0.738527    
## EDUCATION41          6.10545   16.84662   0.362 0.717044    
## EDUCATION42          5.85707   16.81945   0.348 0.727667    
## EDUCATION43          6.62350   16.75900   0.395 0.692682    
## EDUCATION44          7.46676   16.78574   0.445 0.656446    
## EDUCATION45          5.67144   17.01424   0.333 0.738883    
## EDUCATION46          1.61435   16.97883   0.095 0.924252    
## SPOUSE2             -0.19846    0.89659  -0.221 0.824825    
## CHILDREN            -2.05867    0.41900  -4.913 8.99e-07 ***
## METROPOLITAN_STATUS2  4.01316    1.15888   3.463 0.000535 ***
## METROPOLITAN_STATUS3  2.04729    4.66206   0.439 0.660564    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for gaussian family taken to be 7825.934)
## 
##     Null deviance: 367607670  on 46385  degrees of freedom
## Residual deviance: 362794671  on 46358  degrees of freedom
## AIC: 547527
## 
## Number of Fisher Scoring iterations: 2
```

Final Findings and Conclusion

Estimated Crowdout Effects of Computer Leisure on Major Categories

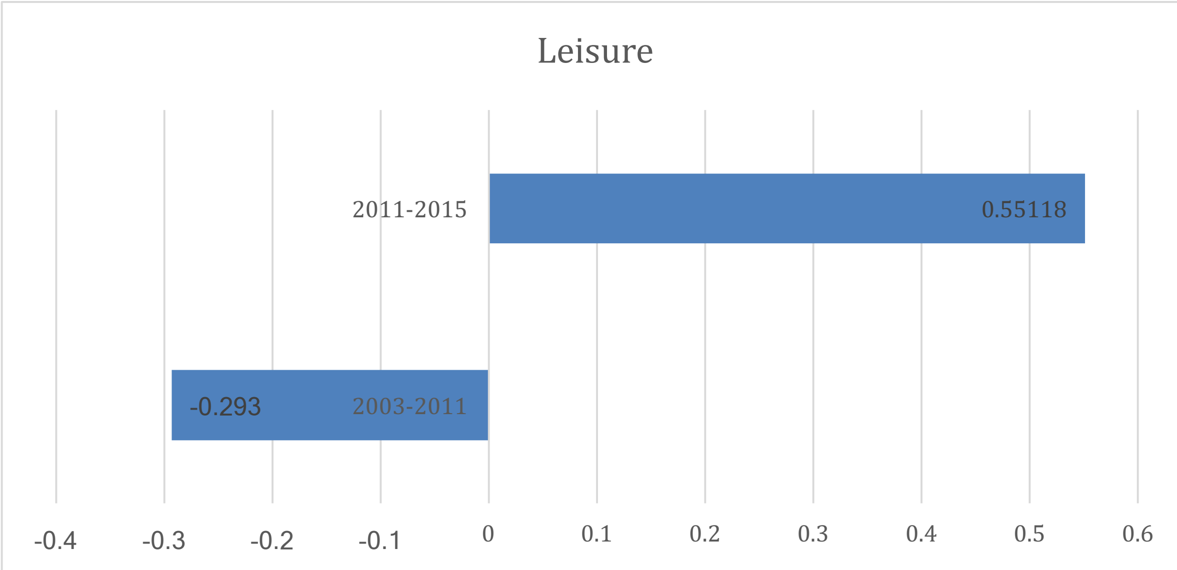
Category	2003-2011 coefficients as per Wallsten's findings	2011-2015 coefficients
Leisure (excluding computer)	-0.293***	0.55118***
	(22.34)	(14.737)
Work activities	-0.268***	-0.97216***
	(19.38)	(10.644)
Personal care (including sleep)	-0.121***	-0.28394***
	(12.36)	(4.872)
Travel	-0.0969***	-0.0010751
	(17.36)	(0.918)
Household activities	-0.0667***	-0.18831***
	(7.149)	(3.316)
Education	-0.0574***	-0.10285*
	(8.560)	(3.253)
Sports	-0.0397***	-0.06359*
	(9.17)	(2.499)
Helping household members	-0.0368***	-0.09164**
	(7.589)	(2.977)
Eating and drinking	-0.0254***	-0.004003
	(6.991)	(0.182)
Helping non-household members	-0.0232***	0.01364
	(6.763)	(0.712)
Religion	-0.0146***	-0.01347
	(5.758)	(-0.702)
Volunteer	-0.0120***	0.005886
	(3.503)	0.273
Professional care and services	-0.00360*	-0.016877
	(1.896)	(1.599)
Household services	-0.00129	-0.001102
	(1.583)	(0.283)
Government and civic obligations	-0.000177	-0.0030381
	(0.303)	(0.868)
Consumer purchases	0.00368	-0.008901

	(1.025)	(0.412)
Phone calls	0.0134***	0.050564***
	(7.433)	(4.983)
Absolute t-statistics in parentheses		
***p<0.01, **p<0.05, *p<0.1		



We notice that most of the correlations are similar between years 2003-2011 and 2011-2015. However, we notice a few opposite correlations. We will first explore those, then we explore those with a similar correlation. We only focus on major changes in our analysis below.

Leisure Time

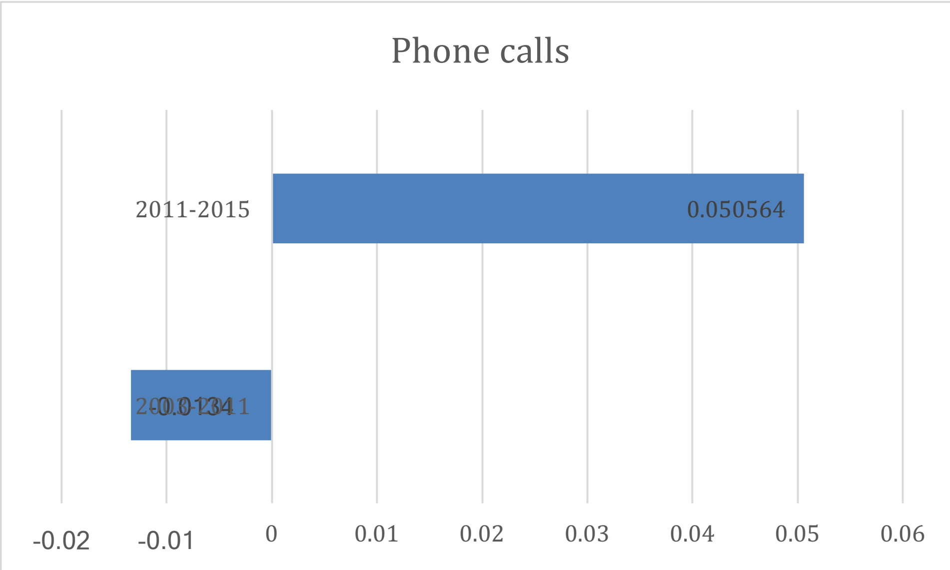


Leisure category includes social communication and hosting or attending social functions, watching television, reading, relaxing or thinking, playing or listening to music or other activies such as attending arts. We have not included leisure activities on the computer in this category (i.e. playing on the computer).

- From 2003 to 2011: 1 minute of online leisure activity translated into 0.29 fewer minutes spent on all other types of leisures
- From 2011 to 2015: 1 minute of online leisure activity translates into 0.55 more minutes spent on all other types of leisure.

An explanation of this positive correlation could be due to some of the subcategories, such as watching tv, reading or listening to music, are now activities being done through the internet. It would require further research in order for us to reach to a conclusion regarding this point.

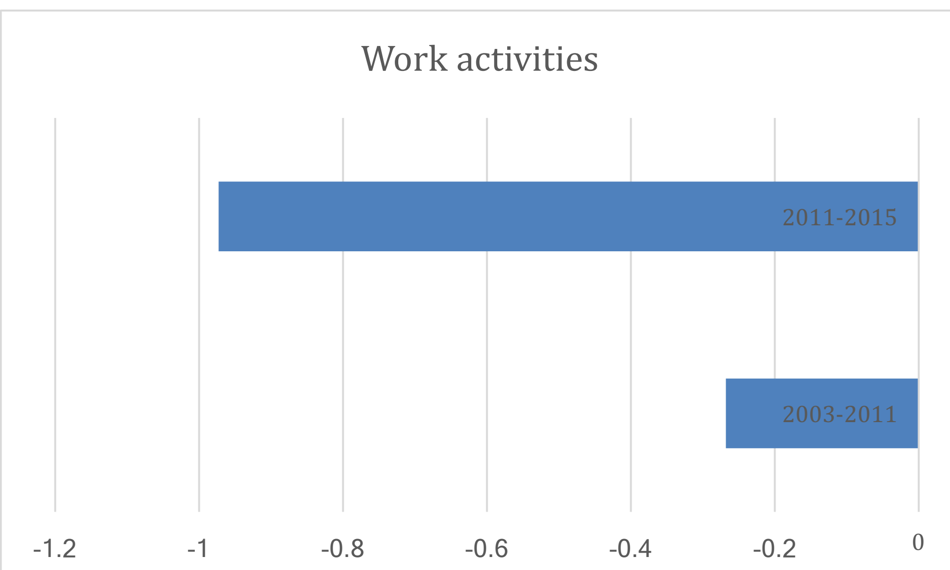
Phone calls



This category includes time spent in telephone communication, texting and Internet voice and video calling.

- From 2003 to 2011: 1 minute of online leisure activity translated into 0.0134 fewer minutes spent on all other types of leisures
- From 2011 to 2015: 1 minute of online leisure activity translates into 0.050564 more minutes spent on all other types of leisure.

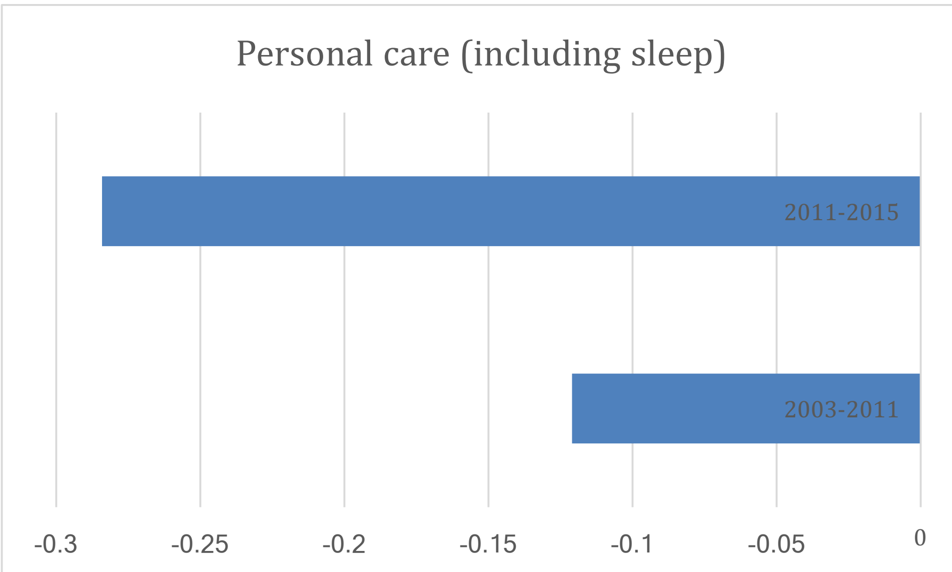
Work activities



This category includes time spent in working, doing activities as part of one's job, engaging in income-generating activities not as part of one's job, and job search activities.

- From 2003 to 2011: 1 minute of online leisure activity translated into 0.268 fewer minutes spent on all other types of leisures
- From 2011 to 2015: 1 minute of online leisure activity translates into 0.97216 fewer minutes spent on all other types of leisure.

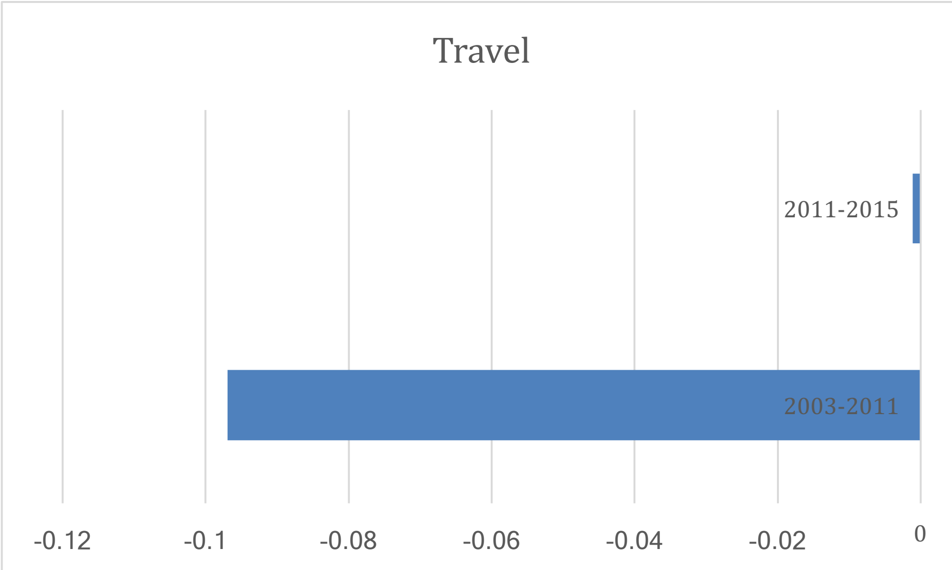
Personal care and sleep



This category includes time spent sleeping, grooming (bathing or dressing), health-related, self-care, and personal or private activities.

- From 2003 to 2011: 1 minute of online leisure activity translated into 0.121 fewer minutes spent on all other types of leisures
- From 2011 to 2015: 1 minute of online leisure activity translates into 0.28394 fewer minutes spent on all other types of leisure.

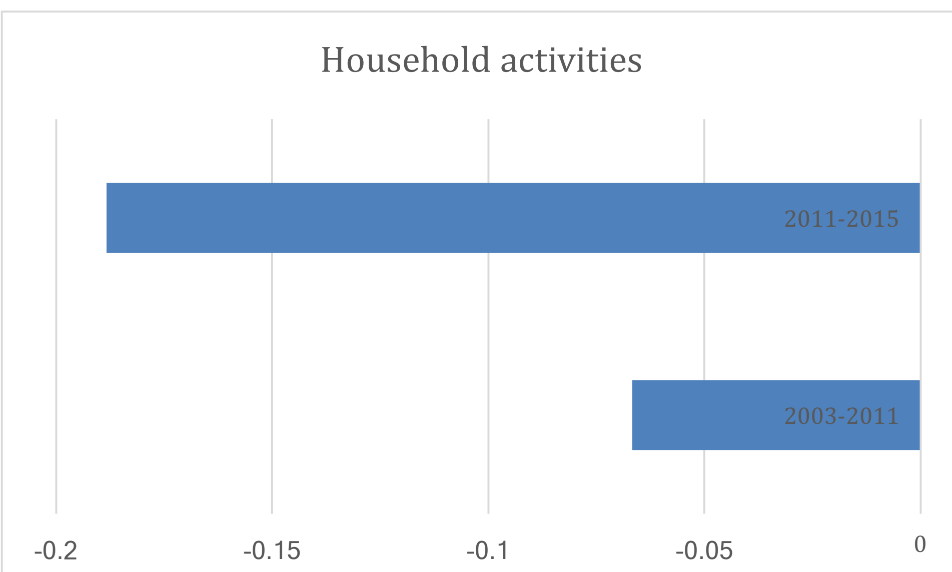
Travel



This category includes any travel time. If an activity is being done while travelling, for example: reading on the subway, the survey accounts for both leisure reading and travel time.

- From 2003 to 2011: 1 minute of online leisure activity translated into 0.10 fewer minutes spent on all other types of leisures
- From 2011 to 2015: 1 minute of online leisure activity translates into 0.001 fewer minutes spent on all other types of leisure.

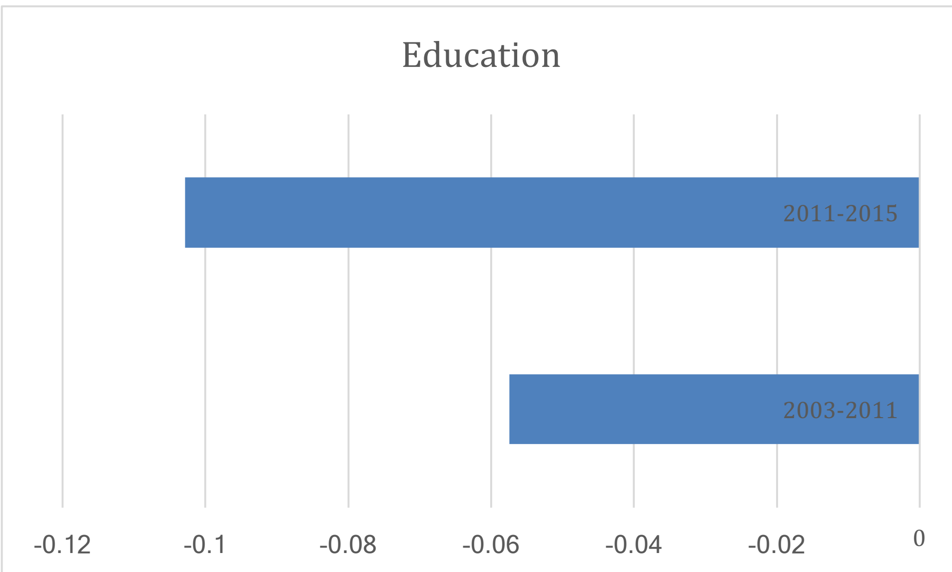
Household activities



This category includes time spent by people to maintain their households. For example: housework, cooking, lawn and garden, pet care, vehicle maintenance and repair etc.

- From 2003 to 2011: 1 minute of online leisure activity translated into 0.07 fewer minutes spent on all other types of leisures
- From 2011 to 2015: 1 minute of online leisure activity translates into 0.18831 fewer minutes spent on all other types of leisure.

Education



This category includes time spent taking classes for a degree or for personal interest (including taking internet or other distance-learning courses), time spent doing research and homework, and time spent taking care of administrative tasks related to education.

- From 2003 to 2011: 1 minute of online leisure activity translated into 0.0574 fewer minutes spent on all other types of leisures
- From 2011 to 2015: 1 minute of online leisure activity translates into 0.10285 fewer minutes spent on all other types of leisure.