

American Time Use Survey 2008

STAT 440

Fall 2011

Luis S. Lin

Contents

1. Introduction	1
1.1 Purpose	1
1.2 Source	1
1.3 Data Description	1
2. Methods	2
2.1 Reading the raw data file	2
2.2 Deriving new variables via calculations or recoding	2
2.3 Dealing with special cases	3
2.4 Creating formats and labels	3
2.5 Checking the data for errors	3
2.6 Further data preparation techniques	5
3. Results	8
3.1 Set 1: Daily participation rates (% of population engaged in activity)	8
3.2 Set 2: Average hours per day by weekday and weekend/holiday	9
3.3 Set 3: Average hours per day by sex	10
3.4 Set 4: Average work hours per day of participants by age and presence of child	11
3.5 Set 5: Average educational hours per day of participants by education and labor force status	11
3.6 Set 6: Average hours per day by race	13
4. Summary	14

1. Introduction

1.1 Purpose

The objective of this report was to apply data preparation techniques to the 2008 American Time Use Survey (ATUS) in order to prepare the data for analysis. The main goal of the analysis was to determine how Americans spend their time. More specifically, what was the participation rate of the population for different activities? How did the distribution of awake time during weekdays compare to that of weekends and holidays? Did the average hours per day spent in an activity vary among different demographics groups? The data preparation and summary techniques discussed in this report were aimed at answering these research questions.

1.2 Source

The dataset was obtained from the Inter-university Consortium for Political and Social Research (ICPSR) website¹. The ICPSR is an international consortium of about 700 academic institutions and research organizations that provides access to more than 500,000 files of research in the social science, making ICPSR the world's largest archive of digital social science data. The mission of ICPSR is to facilitate the access to data by improving usability and taking a standard-based approach to ensure digital preservation.

1.3 Data Description

The 2008 American Time Use Survey (ATUS) was conducted by United States Bureau of Labor Statistics. The survey collects demographic data and information on the amount of time people spent doing various activities in 2008. Respondents were interviewed only once and were randomly selected from a subset of households that have completed their final month of interviews for the Current Population Survey. The ATUS data is unweighted and is based on a stratified random sample. As the documentation points out, “some demographic groups are oversampled to ensure adequate sample size for detailed estimates”.

The raw data file (26149-0001-Data.txt) is called the “Respondent and Activity Summary File” and contains demographic information about respondents and the total time (in minutes) they spent doing each activity that day. The file is in fixed-column format, containing 572 variables (columns) and 12,723 records (rows). The record length is 4,590 and there is one record per unique respondent. There are no missing values in the raw data file.

In addition to the demographic variables, the file also contains a unique identifier (named TUCASEID in the codebook) corresponding to each unique respondent, and a variable for the statistical final weight (named TUFINLWGT in the codebook) for each record. The activity variables correspond to the total number of minutes that each respondent spent doing each 6-digit activity. The 6-digit coding can be found in the documentation provided (26149-0001-Documentation-Activity_Lexicon.pdf).

2. Methods

The data preparation and summary techniques were done using the statistical software SAS version 9.2. All outputs, formats and relevant datasets were stored in a permanent library called “st” with a reference to the physical location ‘C:\STAT 440’. See the SAS file for details of the code.

2.1 Reading the raw data file

The first dataset created was “st.start”. This dataset input the selected variables using fixed-column format from the raw data file. The codebook (26149-0001-Codebook.pdf) contains information on the column location of each variable. Since the record is greater than 250 bytes (default), the option LRECL=4590 was used to tell SAS to read a line up to 4590 characters.

The following variables were selected for the analysis: ATUS Case ID, ATUS final weight, day of week of diary, and the indicator if day of diary is a holiday. The demographics variables selected were: age, sex, race, educational attainment, presence of household children under 18 years, and labor force status. As described in the previous section, the activity variables in the raw data file refer to a 6-digit coded activity. The first two digits correspond to a major activity category, the following two digits to tier 2 and the last two digits to tier 3. For example, the variable “T130126” refers to the activity “Playing Soccer”, since “13” refers to the “Sports, Exercise, and Recreation” category, “01” to the 2nd-tier “Participating in Sports, Exercise, or Recreation”, and “26” to the 3rd-tier “Playing Soccer”. Therefore, inputting and combining corresponding variables are needed to generate a desired activity variable (e.g. the activity “Participating in Sports, Exercise, or Recreation” is the sum of all variables starting with the code “T1301”). The selected activity variables were: sleeping, personal care (not counting sleeping), household, work and work-related, education, consumer purchases, eating and drinking, watching TV, and Participating in Sports, Exercise, or Recreation. All corresponding variables needed to generate these activities were input from the raw data file under the same variable name (e.g. T130126) as listed in the codebook. The rest of the variable (e.g. demographics) names differ from the ones listed in the codebook. Out of the 512 variables in the raw data file, only 132 variables were input to “st.start”.

2.2 Deriving new variables via calculations or recoding

The main dataset called “st.main” was created from “st.start” to generate the variables needed for the analysis by combining variables, changing to appropriate units, dealing with special responses, and applying labels and formats.

The first step was to combine activity variables to obtain the desired variables. This was done by simply summing up the corresponding activity variables in the “st.start” dataset. For example, the variable “HOUSEHOLD” was created by summing up all activity variables starting with “T02”. The variable “WEEKDAY”, which refers to whether the day is a weekday or weekend/holiday, was created using conditionals statements for the variables “DIARYDAY” (the day of week of diary) and “HOLIDAY” (the indicator if day of diary is a holiday). The activity variables were changed from minutes to hours for better interpretability. After combining old variables, “st.main” contained 20 variables.

2.3 Dealing with special cases

The raw data file does not contain any missing value. However, since all variables are numeric, special responses such as “Refused” (-3), “Don’t Know” (-2) and blank (-1) were set to missing (.) to simplify subsetting and analysis in SAS. Note that this only applies to the demographic variables, since the minimum value for the activity variables is zero, meaning the participant spent no time in that activity.

2.4 Creating formats and labels

Permanent formats were created for the demographics variables following a similar scheme as the codebook. However, the formats differ in that responses are aggregated (e.g. for variable “EDU”, responses for Masters, Professional and Doctoral degrees are formatted as one category) and there are no Refused/Don’t Know/Blank categories. This scheme simplifies the analysis and plotting of the data. Appropriate permanent labels for all variables were also applied in the “st.main” dataset.

2.5 Checking the data for errors

In order to ensure that the raw data file was read properly, and that the correct formats and labels were applied, the CONTENTS, PRINT and FREQ procedures for the dataset “st.main” were used. The selected output from CONTENTS procedure with a variable list from “st.main” is shown below.

Variables in Creation Order					
#	Variable	Type	Len	Format	Label
1	ID	Num	8		ATUS Case ID
2	WEIGHT	Num	8		ATUS final weight
3	AGE	Num	8	AGEF.	Age
4	SEX	Num	8	SEXF.	Sex
5	RACE	Num	8	RACEF.	Race
6	CHILD	Num	8	CHILDF.	Presence of household children < 18
7	STATUS	Num	8	STATUSF.	Labor force status
8	DEGREE	Num	8	DEGREEF.	Education Attainment
9	WEEKDAY	Num	8	WEEKDAYF.	Weekday (non-holiday)
10	SLEEP	Num	8	4.2	Sleeping
11	PERSONAL	Num	8	4.2	Personal Care
12	HOUSEHOLD	Num	8	4.2	Household activities
13	WORK	Num	8	4.2	Working and work-related activities
14	EDU	Num	8	4.2	Educational activities
15	SHOP	Num	8	4.2	Purchasing goods and services
16	EAT	Num	8	4.2	Eating and drinking
17	SOCIAL	Num	8	4.2	Socializing and communicating
18	TV	Num	8	4.2	Watching television
19	REC	Num	8	4.2	Participating in sports, exercise, and recreation
20	OTHER	Num	8	4.2	Other activities

Table 1. PROC CONTENTS st.main : Variable List

The output from the PRINT procedure is shown below. The output is limited to the first 10 observations and only includes the “SLEEP” variable of all the activity variables. As it can be seen, the formatting and labeling of variables worked properly.

ATUS Case ID	ATUS final weight	Age	Sex	Race	Presence of household children < 18	Labor force status	Education Attainment	Weekday (non-holiday)	Sleep
2.01E+13	17424963	55 to 64 years	Male	White only	No	Employed	No High School Diploma	Yes	8.63
2.01E+13	5437896	65 to 74 years	Male	Black only	No	Not in labor force	High School Diploma	Yes	23.7
2.01E+13	2932996	75 years and over	Female	White only	No	Not in labor force	No High School Diploma	No	6
2.01E+13	3255620	25 to 34 years	Female	White only	Yes	Employed	High School Diploma	No	10
2.01E+13	1549552	25 to 34 years	Female	White only	Yes	Unemployed	High School Diploma	No	10
2.01E+13	9269868	25 to 34 years	Female	Black only	Yes	Employed	High School Diploma	Yes	9.25
2.01E+13	6196263	25 to 34 years	Male	White only	No	Employed	No High School Diploma	No	9
2.01E+13	16104499	20 to 24 years	Female	Black only	No	Employed	No High School Diploma	Yes	11.5
2.01E+13	3580695	35 to 44 years	Female	White only	Yes	Employed	High School Diploma	No	6.47
2.01E+13	13941762	25 to 34 years	Male	White only	No	Employed	No High School Diploma	No	15

Table 2. PROC PRINT st.main : 1-10 obs

The output of FREQ procedure for the unweighted and weighted cases are shown in Table 3, which lists the count and percentages by sex and age. The table on the left shows the actual percentage of respondents interviewed for the different categories. It can be noted that the percentage for males is 44% while for females is 56%. On the other hand, the percentages are 48% and 52% for males and females respectively after weighting (table on the right). This shows why weights are needed in order to have a better representation of the population and avoid computing statistics with misleading results. As the documentation notes, “men have lower response rates than do women, so the weights for male respondents are larger, on average, than those for female respondents. The weights ensure that groups and days of the week are correctly represented in spite of differing response rates”. The output also shows that there are no missing observations and that the code correctly read all records (12,723).

Table of AGE by SEX				Table of AGE by SEX			
AGE(Age)	SEX(Sex)			AGE(Age)	SEX(Sex)		
Frequency	Male	Female	Total	Frequency	Male	Female	Total
Percent				Percent			
15 to 19 years	436	425	861	15 to 19 years	3.95E+09	3.84E+09	7.79E+09
	3.43	3.34	6.77		4.54	4.41	8.95
20 to 24 years	247	289	536	20 to 24 years	3.77E+09	3.70E+09	7.47E+09
	1.94	2.27	4.21		4.32	4.25	8.57
25 to 34 years	882	1207	2089	25 to 34 years	7.32E+09	7.32E+09	1.46E+10
	6.93	9.49	16.42		8.41	8.4	16.8
35 to 44 years	1236	1500	2736	35 to 44 years	7.53E+09	7.73E+09	1.53E+10
	9.71	11.79	21.5		8.64	8.88	17.52
45 to 54 years	1116	1254	2370	45 to 54 years	7.88E+09	8.21E+09	1.61E+10
	8.77	9.86	18.63		9.05	9.42	18.47
55 to 64 years	859	1031	1890	55 to 64 years	5.90E+09	6.35E+09	1.23E+10
	6.75	8.1	14.85		6.78	7.29	14.07
65 to 74 years	524	692	1216	65 to 74 years	3.35E+09	3.93E+09	7.28E+09
	4.12	5.44	9.56		3.85	4.51	8.35
75 years and over	345	680	1025	75 years and over	2.51E+09	3.82E+09	6.32E+09
	2.71	5.34	8.06		2.88	4.38	7.26
Total	5645	7078	12723	Total	4.22E+10	4.49E+10	8.71E+10
	44.37	55.63	100		48.46	51.54	100

Table 3. PROC FREQ st.main by sex and age: Unweighted (Left) and Weighted (Right)

2.6 Further data preparation techniques

In order to prepare the data for analysis, the dataset “st.main” was used as a starting point to create relevant datasets for the research question in mind. Additional data preparation methods, such as subsetting data, recoding variables, and rotating datasets, were used to prepare the data that would facilitate building appropriate tables and charts.

The dataset is currently arranged with the respondents as the observations, and the time of each activity as a different variable. However, in order to subset activities, and new variable called “ACTIVITY”, taking the values of the name of each activity was created. In addition, to simplify the analysis and because PROC MEANS easily computes weighted statistics, a new variable called “TIME” containing the weighted mean of the time of each activity was created.

In order to answer the research questions posed in the previous section, the analysis was divided into “sets”. In each set, the appropriate weighted means are computed, a new dataset is generated with the weighted results, and charts are created. Additional intermediate steps are needed depending on the research question. Figure 1 shows a graphical outline for the data preparation of the different sets for the analysis.

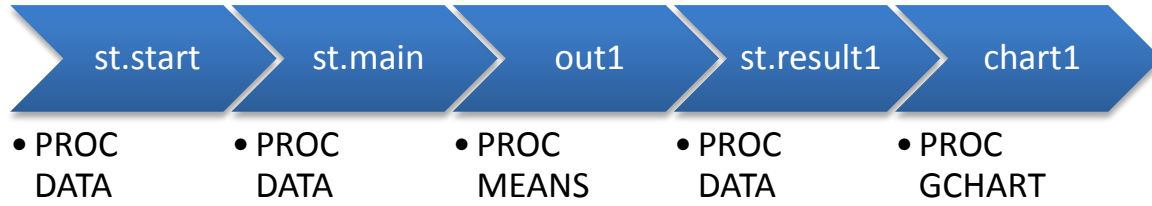


Figure 1. Outline for data preparation and summary for each set

2.6.1 Set 1: Daily participation rates (% of population engaged in activity)

In order to compute the daily participation rate, a new dataset called “st.set1” was created by recoding the values of the times for each activity as follows: 1 if respondent engaged in activity (time spent is different from zero), 0 if respondent did not engaged in activity (time spent is zero). Then, PROC MEANS for all activity variables was used to compute the weighted participation rate and its output was stored as “out1”. PROC DATA was used on “out1” to create a new data set called “st.result1” containing the computed daily participation rate for each activity. Finally, PROC GCHART was used to generate a vertical bar chart. Note: daily participation rate is defined as:

$$P_j = \frac{\sum_i Weight_i \cdot I_{ij}}{\sum_i Weight_i}$$

where P_j is defined as the daily participation rate in activity j , and I_{ij} is equal to 1 if respondent i engaged in activity j and equal to zero otherwise.

2.6.2 Set 2: Average hours per day by weekday and weekend/holiday

In order to compute the average hours per day by weekday and weekend/holiday, PROC MEANS for all activity variables by variable WEEKDAY was used to compute the weighted average hours and its output was stored as “out2”. PROC DATA was used on “out2” to create a new data set called “st.result2” containing the computed weighted averages for each activity by WEEKDAY. The percent of each activity was also calculated and the time values were formatted as HH:MM for improved readability. Finally, PROC GCHART was used to generate a pie chart by WEEKDAY of distribution of hours during awake time. Note: the statistic average hours per day is defined as:

$$\bar{T}_j = \frac{\sum_i Weight_i \cdot T_{ij}}{\sum_i Weight_i}$$

where T_{ij} is defined as hours spent by respondent i in activity j .

2.6.3 Set 3: Average hours per day by sex

Same approach as 2.6.2 with SEX as the class variable. A vertical bar chart is created with selected variables TV, HOUSEHOLD, REC and PERSONAL.

2.6.4 Set 4: Average work hours per day of participants by age and presence of child

Same approach as 2.6.2 with AGE and CHILD as the class variables. Since average hours per day of participants in the activity are computed, the variable WORK was subset (time is different from zero) in PROC MEANS. A vertical bar chart is created. Note: the statistic average hours per day of participants is defined as:

$$\bar{T}_j^p = \frac{\sum_i Weight_i \cdot T_{ij} \cdot I_{ij}}{\sum_i Weight_i \cdot I_{ij}}$$

where T_{ij} is defined as hours spent by respondent i in activity j , and I_{ij} is equal to 1 if respondent i engaged in activity j and equal to zero otherwise.

2.6.5 Set 5: Average educational hours per day of participants by education and labor force status

Same approach as 2.6.4 with DEGREE and STATUS as the class variables. Variable EDU was subset (time is different from zero) in PROC MEANS.

2.6.6 Set 6: Average hours per day by race

Same approach as 2.6.2 with RACE as the class variable. A horizontal bar chart is created with selected variables TV, SOCIAL and REC.

All tables were exported to a document file called 'Output.rtf' using a 'ods' statement. The images were exported manually using the File-Export menu in order to preserve the highest resolution.

3. Results

3.1 Set 1: Daily participation rates (% of population engaged in activity)

Table 4 shows the mean weighted daily participation rates for all activities. As expected, the indispensable activities (sleep and eat) have almost 100% participation rate. An unusual result is the high participation rate of watching TV (81%), which about the same as personal care activities. Figure 2 shows a graphical display of the results ordered by decreasing participation rate of activities.

Variable	Label	Mean
SLEEP	Sleeping	1.00
PERSONAL	Personal Care	0.81
HOUSEHOLD	Household activities	0.76
WORK	Working and work-related activities	0.47
EDU	Educational activities	0.08
SHOP	Purchasing goods and services	0.41
EAT	Eating and drinking	0.96
SOCIAL	Socializing and communicating	0.38
TV	Watching television	0.81
REC	Participating in sports, exercise, and recreation	0.18
OTHER	Other activities	0.97

Table 4. Daily Participation Rate by Activity

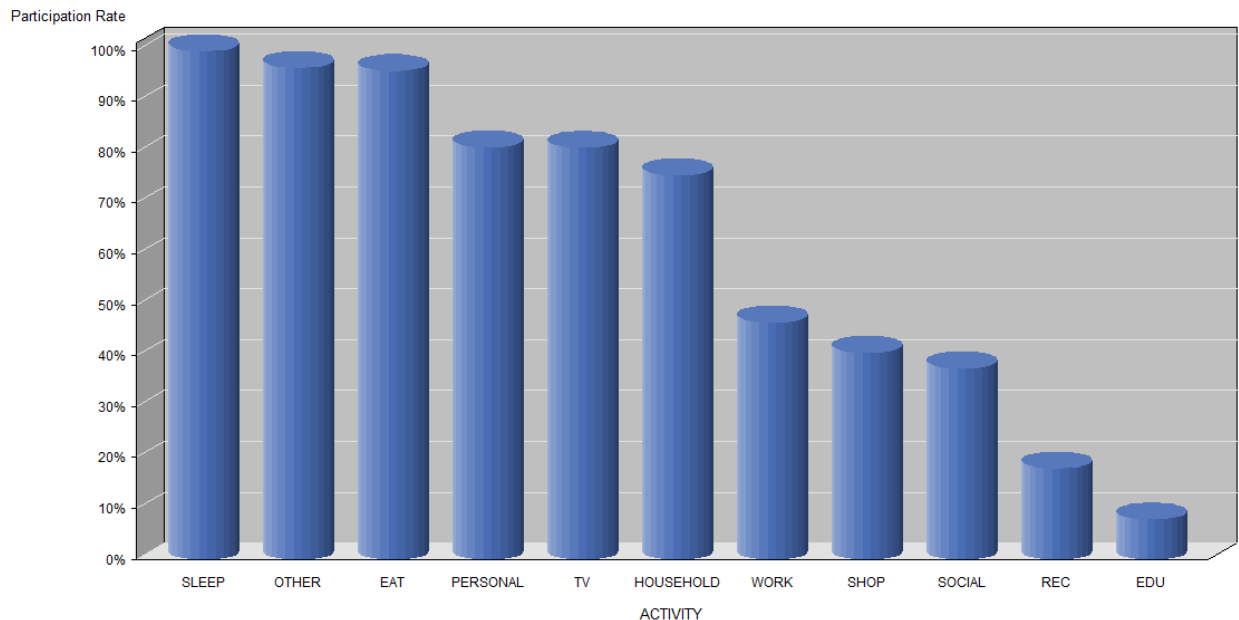


Figure 2. Daily Participation Rate by Activity

3.2 Set 2: Average hours per day by weekday and weekend/holiday

Table 5 shows the mean weighted hours spent in each activity by type of day. The results for awake time are summarized in Figure 3. As expected, Americans spend more time sleeping during weekends and non-holiday compared to weekdays. The difference is about an hour. An interesting result is again the high number of hours spent watching in TV, which is 3:17 and 2:33 hours during weekends/holidays and weekdays respectively. In terms of percentages, Americans spend on average 22% of their awake time during weekends watching TV. Another interesting result is that Americans about an hour (8.73% of their awake time) doing work-related activities during weekends/holidays.

Weekday (non-holiday)	Variable	Label	Mean
No	SLEEP	Sleeping	9.31
	PERSONAL	Personal Care	0.75
	HOUSEHOLD	Household activities	2.10
	WORK	Working and work-related activities	1.28
	EDU	Educational activities	0.13
	SHOP	Purchasing goods and services	0.53
	EAT	Eating and drinking	1.20
	SOCIAL	Socializing and communicating	0.94
	TV	Watching television	3.29
	REC	Participating in sports, exercise, and recreation	0.35
	OTHER	Other activities	4.13
Yes	SLEEP	Sleeping	8.30
	PERSONAL	Personal Care	0.79
	HOUSEHOLD	Household activities	1.63
	WORK	Working and work-related activities	4.38
	EDU	Educational activities	0.57
	SHOP	Purchasing goods and services	0.32
	EAT	Eating and drinking	1.07
	SOCIAL	Socializing and communicating	0.51
	TV	Watching television	2.55
	REC	Participating in sports, exercise, and recreation	0.27
	OTHER	Other activities	3.61

Table 5. Average hours per day by type of day

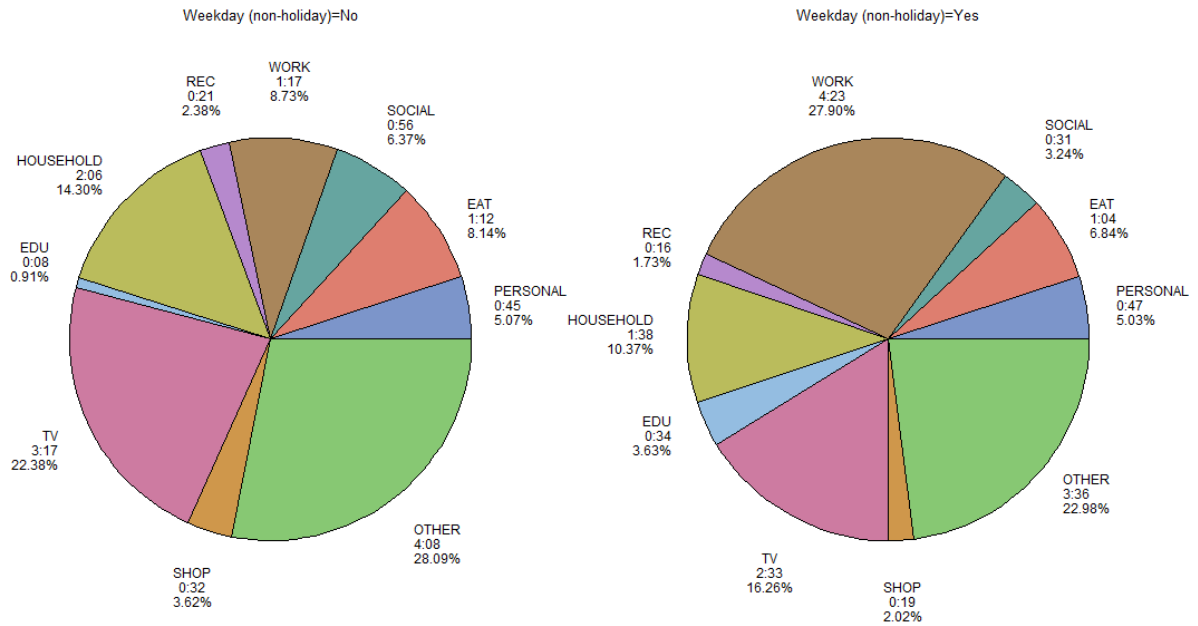


Figure 3. Average hours per day by type of day

3.3 Set 3: Average hours per day by sex

Figure 4 shows the average hours per day of selected activities by sex in decreasing order of time spent in an activity. As it can be seen, females and males have the same relative distribution of hours spent by activity (i.e. both groups spend more time watching TV, followed by household activities, personal care and recreational activities). However, within each activity, the time spent differs by group. For example, males spend about 30 minutes more watching TV, while women spend about one more hour doing household activities. Similarly, males spend more time in recreational activities and less time in personal care activities than females do. All these relative differences are expected.

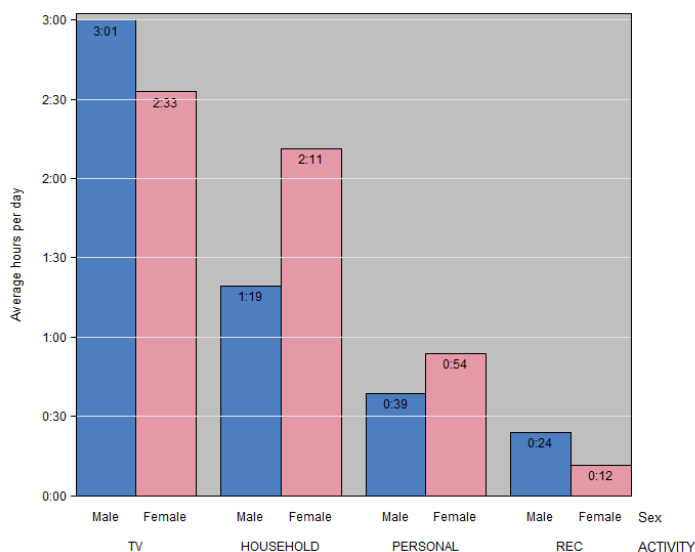


Figure 4. Average hours per day by sex

3.4 Set 4: Average work hours per day of participants by age and presence of child

Figure 4 shows the distribution of the average work hours per day of participants for selected activities. The data is ordered by increasing age and grouped by the presence of a child under 18 years in the household. As it can be seen, the same relative distribution of time is followed by the two different groups (presence of child). For example, the average work-related hours increases with age up to a certain age, and then decreases for both groups. However, there seems to be a difference in the actual time spent in work-related activities for the extreme range of age. For instance, if there is a child under 18 in a household, Americans 75 and older spend 2 hours in work-related activities, compared to 5 hours if there is no child under 18 in the household. Intuitively this makes sense since Americans 75 and older are more likely to spend time taking care of a child in the household, and thus, spend less time in work-related activities.

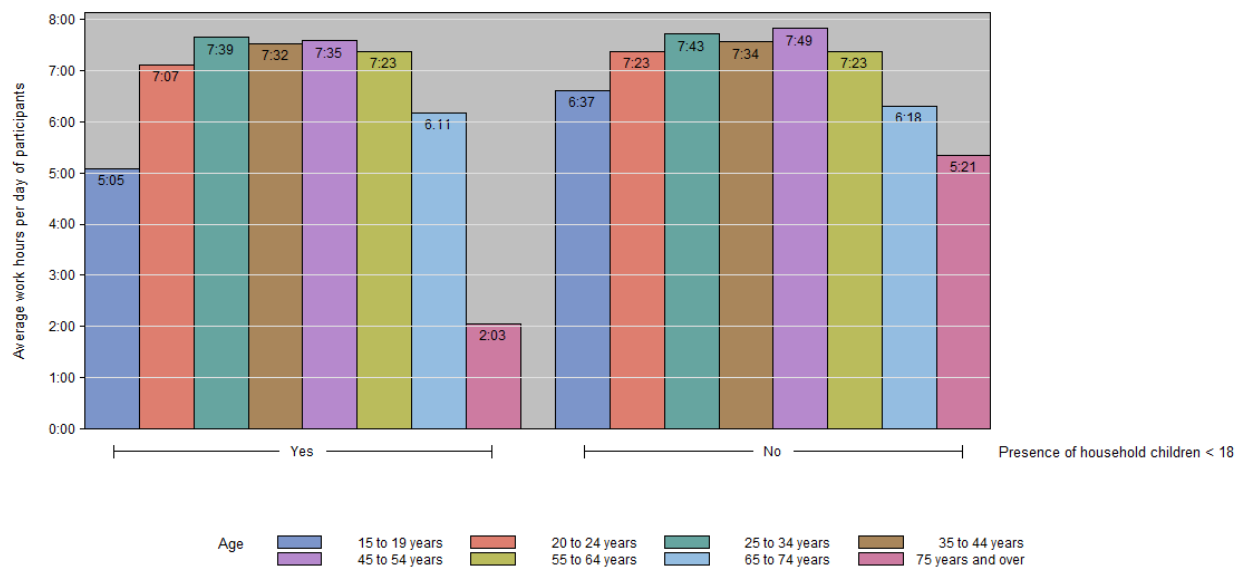


Figure 5. Average work hours per day of participants by age and presence of child

3.5 Set 5: Average educational hours per day of participants by education and labor force status

Table 6 shows the mean weighted hours spent in educational activities of participants by labor force status and degree. The results are graphically summarized in Figure 6, ordered by ascending level of education and grouped by labor force status. As it can be seen, there seems to be a trend across groups: as the educational level increases, the time spent in educational activities decreases. For the group not in the labor force, Americans spend on average more time on educational activities than the other two groups. In addition, for the unemployed and not in labor force groups, Americans with Bachelor's as the highest degree level spend more time in educational activities than Americans with a lower educational degree but with a high school diploma. The highest time spend on educational-related activities is about 6 to 7 hours for Americans with no high school diplomas that are not in the labor force or are unemployed.

Analysis Variable : EDU Educational activities		
Labor force status	Education Attainment	Mean
Employed	No High School Diploma	6.10
	High School Diploma	4.86
	Bachelors degree	3.99
	Masters, Professional or Doctoral degree	3.26
Unemployed	No High School Diploma	6.78
	High School Diploma	4.38
	Bachelors degree	5.20
	Masters, Professional or Doctoral degree	2.44
Not in labor force	No High School Diploma	6.52
	High School Diploma	5.45
	Bachelors degree	6.27
	Masters, Professional or Doctoral degree	2.97

Table 6. Average educational hours of participants by labor force status and degree

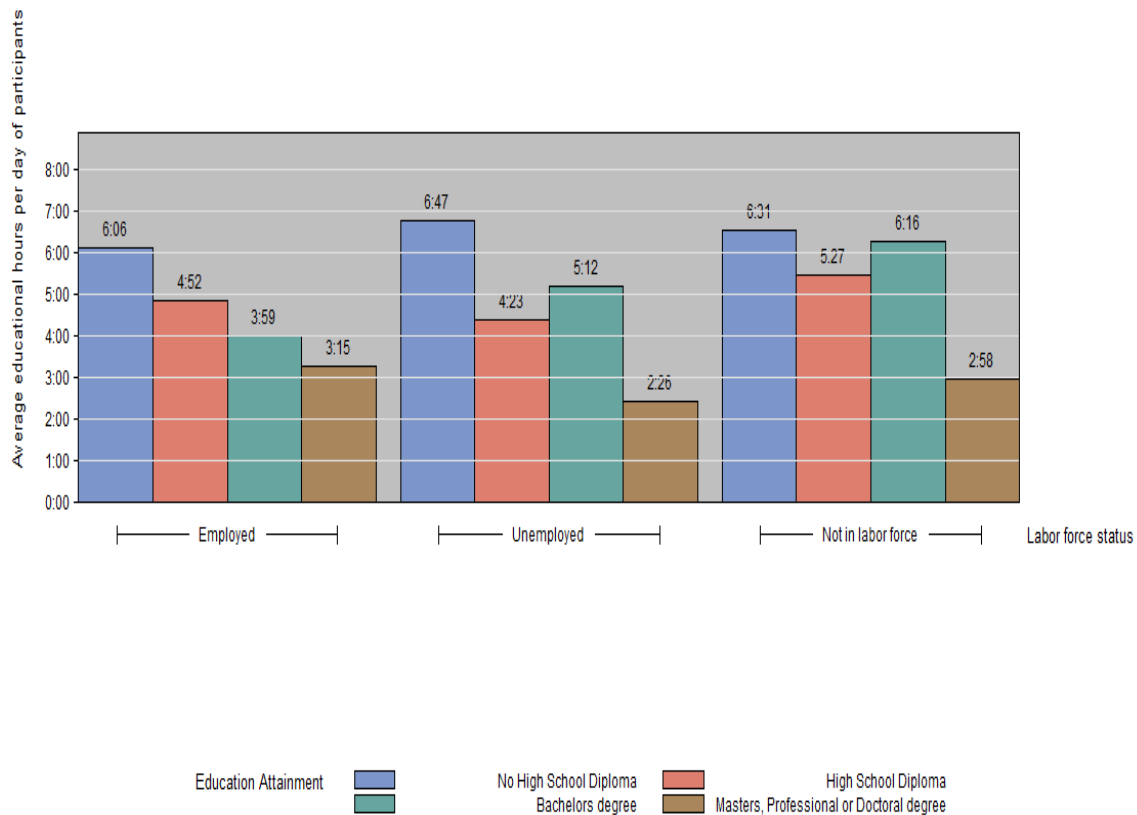


Figure 6. Average educational hours of participants by labor force status and degree

3.6 Set 6: Average hours per day by race

Figure 7 shows the average hours per day of selected activities by race in decreasing order (moving up in the vertical axis) of time spent in an activity. As it can be seen, the different groups of races have the same relative distribution of hours spent by activity (i.e. all groups spend more time watching TV, followed by social activities and recreational activities). Within each activity, the relative difference among different races seems to be maintained. For instance, African Americans spend the most time watching TV, followed by Caucasians, Native Americans, Mixed or Other, and Asian Americans. A similar order applies to the time spent on social activities. For recreational activities, there does not seem to be a significant difference among groups. Regarding watching TV, an interesting observation is that African Americans spend on average about 1 more hour daily watching TV than Caucasians do.

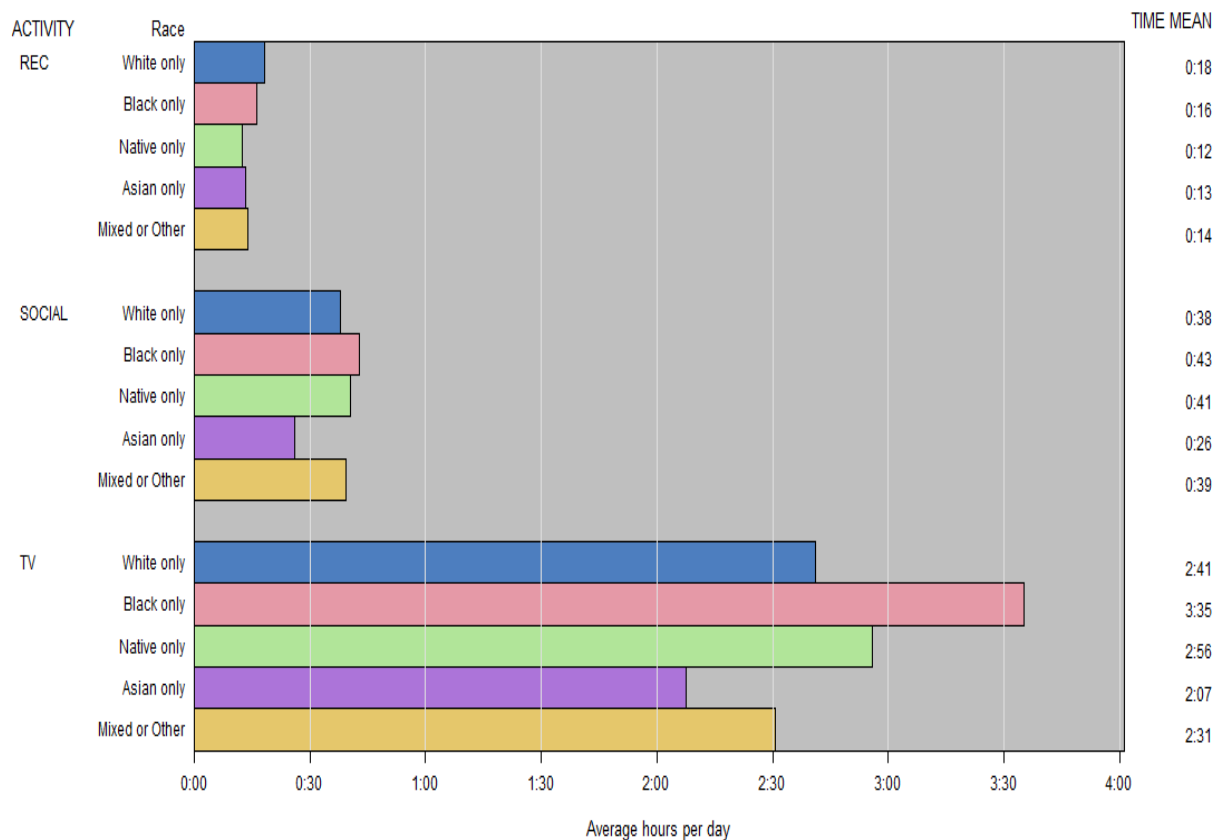


Figure 7. Average hours per day by race

4. Summary

The objective of this report was to apply data preparation techniques to the 2008 American Time Use Survey (ATUS) in order to prepare the data for analysis. The data analysis techniques used included reading raw data files, deriving new variables by recoding or calculation, subsetting, dealing with special responses, creating labels and formats, and checking for errors in the datasets.

A detailed discussion of the data preparation techniques was given in order to prepare the data for addressing the research questions. The results section gives tables and charts generated from the datasets discussed in the methods sections.

After preparing and summarizing the data, future steps involve a statistical analysis to determine , for example, correlations, relationship between variables or if there are any significant differences among demographic groups.

ⁱ http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/26149?classification=ICPSR.IV.*&sortBy=7&paging.rows=25