How Americans' Time-Use Patterns Have Changed From 2003 to 2017

Group July

26 November 2018

Introduction

The aim of this report is to discover how Americans' time-use patterns have changed over the last 15 years through discovering and analysing long-term trends within this period. The datasets [1] used to tackle this comprise some of the results of the American Time Use Survey (ATUS) [2]. ATUS respondents were interviewed once about how they spent their time on the previous day, where they were, and whom they were with. The survey records the time spent on 431 different activities that are grouped into 17 different categories. In the analysis these categories will often be referred to by a variable beginning with `tu` followed by a two-digit *category number*. The meaning of each category number and identifications of the rest of the variables can be found in the lexicon and data dictionary files which are obtained from the ATUS website.

The Bureau of Labor Statistics website provides a number of simple charts looking at the 2017 annual average for these categories. This report provides a more definitive conclusion on how time-use patterns changed over the period, evidenced with statistical investigation of the data. The findings explain in depth where some of the biggest fluctuations are and briefly note possible reasons for this. Limitations to the analysis are noted where appropriate and these must be taken into account when performing any further analysis or summaries; mentioned alongside any conclusions.

Usually when analysing data, statisticians will split the dataset into training data (around 90%) and validation data (around 10%). This ensures that any exploratory data analysis (EDA) conclusions from the training data can be verified using the remaining data. Whilst the analysis performed for this report follows this technique, it splits the data approximately evenly. This is to ensure that seasonality does not cause incorrect rejections of hypotheses generated in the EDA. More specifically, EDA was performed on every other month which doesn't include July.

The Compelling Change in Caring for & Helping Non-HH Members

Exploratory Data Analysis

The ATUS dataset records time spent on activities in minutes. This report initially looks at how the proportion of American's partaking in the 17 different groups of activities has changed. The following table gives a summary of some of these changes. The activities included are those with a percentage change of over 10% and a variance greater than 0.5.

Table 1: Change in Participation of Activities

Measure	tu04	tu08	tu13	tu14	tu16
Variance	2.54	0.89	1.44	0.60	1.93
% Change	-32.11	-26.46	10.17	12.32	-24.08

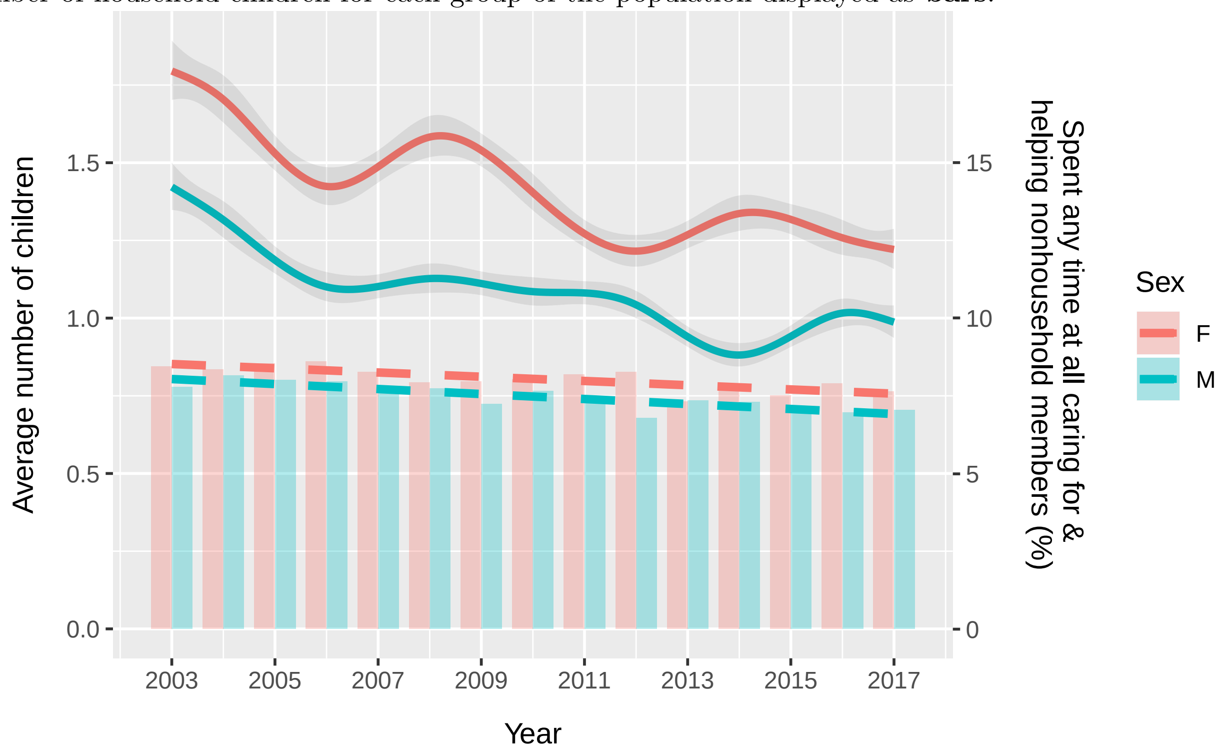
It is immediately clear that the most compelling change out of the 17 activity groups over the period is in *Caring for & Helping Non-household (NonHH) Members* (tu04).

When fitting a linear model, the method of best subsets, with Year forced in, tells us that the most significant variables to predict the proportion of American's spending time *Caring for & Helping*

Nonhousehold (NonHH) Members is **Sex** and **Number of Household Children**. A linear model is not sufficient to accurately predict tu04 participation, however it does give a good idea as to what is causing the change of the period 2003-2017.

The continuous activity data is non-negative and non-linear. Therefore a suitable model, using **Year** and **Sex** as explanatory variables, is a natural cubic spline GLM with multiplicative errors and log link. In fact it gives a D-squared accuracy score ≈ 1. An F test shows this model is a significant improvement over the same model which does not use **Sex** as a variable, confirming the belief that is has a significant effect on participation in caring for & helping nonhousehold members.

The plot below is a graphical interpretation of the model, built on the training data, with the average number of household children for each group of the population displayed as **bars**.



The key observation is that, as the table suggested, over the period the participation in caring for & helping non-household members has decreased for both men and women, and there appears to be a link between this and the number of household children. Updating the model to include the number of household children and then testing this hypothesis with a t-test confirms this is the case. Looking at the graph, fluctuations in the average number of household children, on the whole seem to be followed by but if we check the correlations between them, they show this link is fairly weak (0.65 for Men and 0.49 for Women), suggesting there must be further reasons for this change; possibly not measured in this dataset.

Gender Roles

A major news topic in recent years has been the division between genders in terms of roles and responsibilities. This debate has been fueled by a new wave of feminism, as well as a number of different campaigns, including *'#MeToo'* and *'Times Up'*. Both of these movements aim at addressing commonplace sexual harassment and discrimination and empowering women to take control. Given this backdrop, it was decided that an interesting area of the ATUS data to focus on would be the change in time use for different genders. Specifically, areas of time use where pre-existing gender roles are present were examined.

Talcott Parsons' [**input BibTex reference here**] view of gender roles, developed in the 1950s, compared traditional gender roles with more liberal alternatives. A simplified version of the strictly traditional view of gender roles is detailed below.

Table 3: Generations

Generation	Birth Years
Silent Generation	1928 - 1945
Baby Boomers	1946 - 1964
Generation X	1965 - 1980
Millennials	1981 - 1996

Table 2: Traditionalist Gender Actitivies

Male Activites	Female Activities
Working	Housework
House Maintenance	Cooking
Vehicle Maintenance	Childcare

The liberal viewpoint discussed by Parsons suggests an equal balance of time for the genders in these roles. Whilst this study was developed over 50 years ago, a preliminary look at the data confirmed that, in 2003, a complete transition to the liberal viewpoint had not occurred. Therefore, it was decided to investigate whether the time use of these specific activities was heading to an equal split between the genders or whether it is stationary. Of course, it should be noted that 15 years is not a huge length of time to measure this sort of social change, as only a few sets of generations responses are recorded, however, it presents an interesting topic.

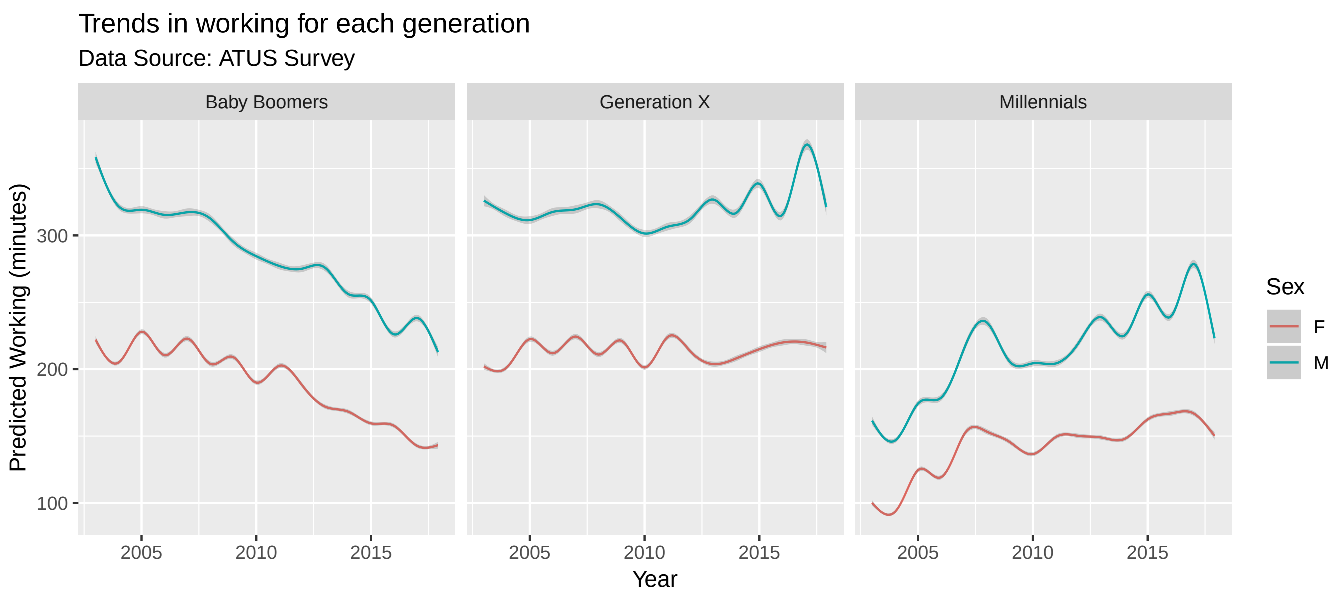
Following, some initial exploration of the data, it was also decided that looking at different generations [insert reference here] presented a different look at the activities, with each generations spending different lengths of time on activities. Essentially, this provides a split in the data by age, but using generations provides a way of tracking a population over time and without having to arbitraily pick age categories. These generations are defined in the table below. A small number of participants fell outside of these ranges, however, these represented only a small number of participants and so were added to the next nearest generation (e.g. if a participant was born in 1998, they would be added to the *Millennials* generation).

For each of the activities, a suitable generalized linear model was developed using sex, year and generation as parameters. The values used to predict each model were constructed using the weighted mean formula provided in the *ATUS User's Guide*, available online in the ATUS Survey Documentation. Splines were added to the model for each year and graphical representations were produced.

For the generation classifications, we need to reference http://www.pewresearch.org/fact-tank/2018/03/01/millennials-overtake-baby-boomers/

Working

The plots below show the change in working patterns between three different generations. The *Silent Generation* have been excluded from this plot as, by 2010, the youngest of this generation will be 65. At this age, most will have retired and as such the plot is not very informative. Furthermore, the same scale is used for each graph (for purpose of easier comparison), so this affects the quality of comparison between the more informative generations.




```
##
## Call:
## glm(formula = weighted.Working ~ TESEX + TESEX:ns(TUDIARYDATE,
##     knots = c(as.Date(Dates))) - 1, family = quasi(link = "log",
##     variance = "mu^2"), data = Gender.Roles.Data, mustart = rep(250,
##     n))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.3926  -0.2231   0.1289   0.2804   0.7687
##
## Coefficients:
##                                                  Estimate Std. Error
## TESEXF                                           5.135814   0.007312
## TESEXM                                           5.560234   0.008324
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))1  0.104058   0.010943
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))1 -0.052514   0.012419
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))2 -0.100624   0.013780
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))2  0.001024   0.015792
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))3  0.082806   0.012737
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))3 -0.052315   0.014687
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))4 -0.031428   0.013588
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))4 -0.013406   0.015403
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))5 -0.017277   0.013032
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))5 -0.024767   0.014805
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))6 -0.122641   0.013180
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))6 -0.151025   0.014943
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))7  0.010372   0.013222
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))7 -0.081187   0.015014
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))8 -0.059924   0.013454
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))8 -0.095472   0.015204
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))9 -0.078993   0.013577
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))9 -0.072931   0.015305
```

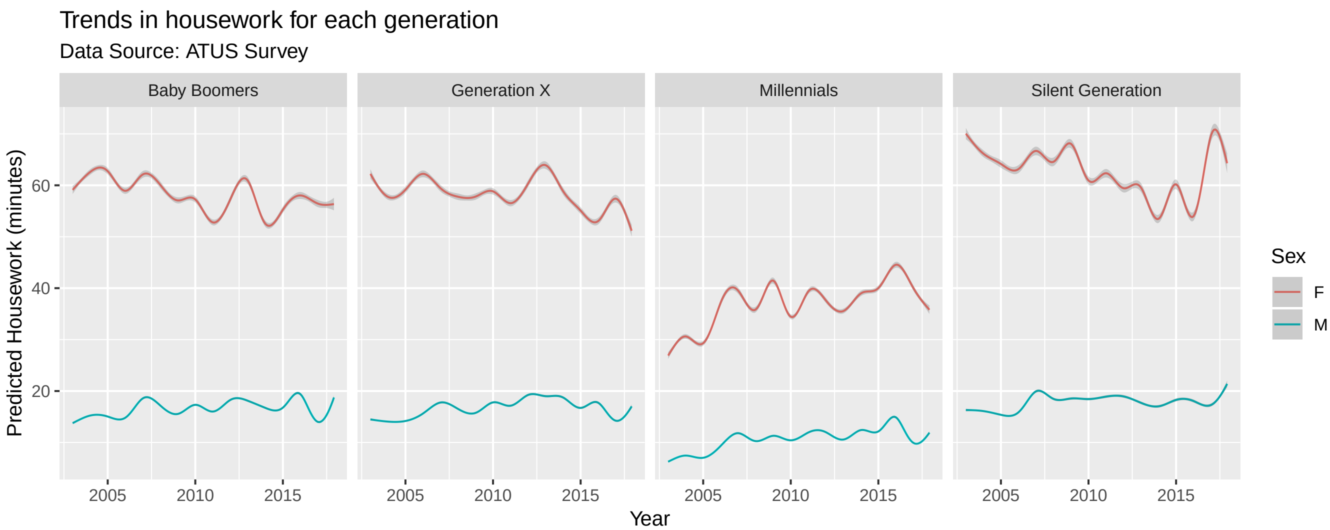
```
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))10 -0.126703   0.013838
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))10 -0.152983   0.015623
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))11 -0.042360   0.013889
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))11 -0.031391   0.015689
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))12 -0.099014   0.014193
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))12 -0.177917   0.016003
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))13 -0.014295   0.012911
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))13  0.030670   0.014294
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))14 -0.229069   0.019789
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))14 -0.188767   0.022367
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))15 -0.044557   0.011963
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))15 -0.145739   0.013037
##                                                     t value Pr(>|t|)    
## TESEXF                                              702.415  < 2e-16 ***
## TESEXM                                              668.015  < 2e-16 ***
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))1    9.509  < 2e-16 ***
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))1   -4.229 2.35e-05 ***
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))2   -7.302 2.85e-13 ***
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))2    0.065 0.948314    
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))3    6.501 7.97e-11 ***
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))3   -3.562 0.000368 ***
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))4   -2.313 0.020729 *  
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))4   -0.870 0.384101    
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))5   -1.326 0.184950    
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))5   -1.673 0.094352 .  
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))6   -9.305  < 2e-16 ***
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))6  -10.106  < 2e-16 ***
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))7    0.784 0.432810    
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))7   -5.408 6.40e-08 ***
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))8   -4.454 8.43e-06 ***
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))8   -6.279 3.41e-10 ***
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))9   -5.818 5.96e-09 ***
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))9   -4.765 1.89e-06 ***
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))10  -9.156  < 2e-16 ***
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))10  -9.792  < 2e-16 ***
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))11  -3.050 0.002290 ** 
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))11  -2.001 0.045414 *  
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))12  -6.976 3.04e-12 ***
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))12 -11.118  < 2e-16 ***
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))13  -1.107 0.268236    
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))13   2.146 0.031899 *  
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))14 -11.575  < 2e-16 ***
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))14  -8.440  < 2e-16 ***
## TESEXF:ns(TUDIARYDATE, knots = c(as.Date(Dates)))15  -3.725 0.000196 ***
## TESEXM:ns(TUDIARYDATE, knots = c(as.Date(Dates)))15 -11.179  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for quasi family taken to be 0.1780519)
## 
```

```
##     Null deviance: 73357226  on 191558  degrees of freedom
## Residual deviance:    68913  on 191526  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

House Maintenance and Vehicle Maintenance

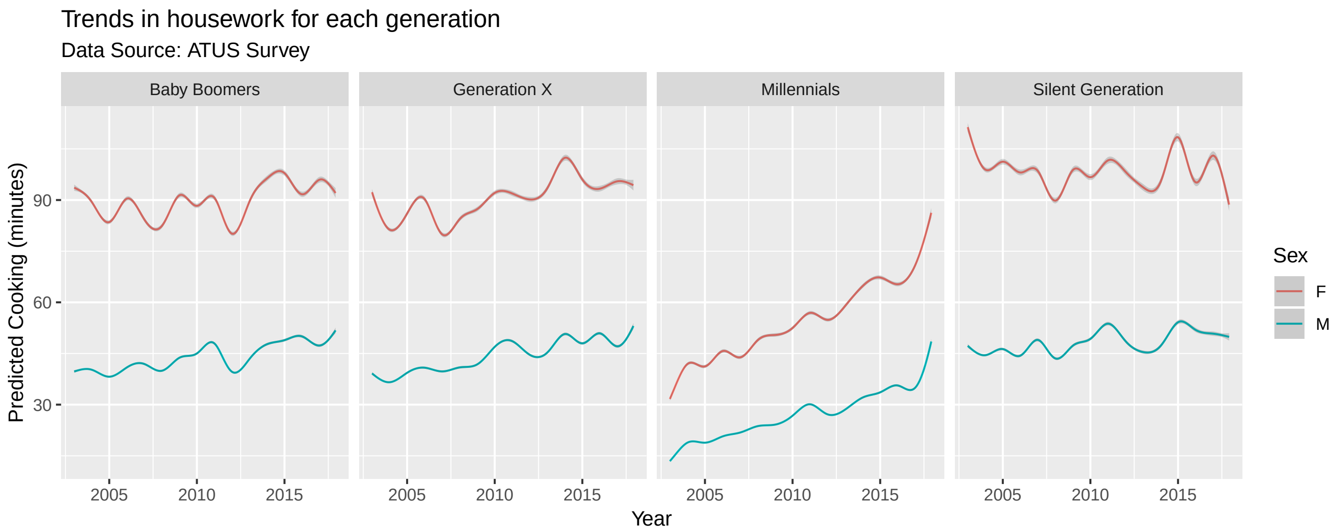
For both of these variables, it was decided that the participation rate was too low to warrant deeper analysis. Whilst the findings represented that there existed a separation in gender, with men spending more time on both of these activities, the participation rates of around 3% for both refelcted that these were more uncommon activities. As such, it was decided that there was not enough data to reflect the time spent on these activities in a suitable linear model.

Housework



The graphs above show that for all generations, except for *Millennials*, there has been a decrease in the gap between men and women. This is as a result of men spending more time and women spending less time on housework, although the decrease for women is sharper than the increase for men. This trend is reflected strongest in *Generation X*, the second youngest generation, however, there is also strong evidence for this trend in the *Silent Generation* despite the slight rise in women in 2016. On the other hand, the gap seems to have increased slightly for *Millennials* - both sexes are increasing the amount of time spent on housework, but the increase is not as severe for men. Notably, *Millennials* also spend substantially less time doing housework than the other generations, with the highest point being around 45 minutes per day (women in 2016), compared to around 62 in *Baby Boomers* and *Generation X* and 72 in the *Silent Generation*. Therefore, whilst the gap is increasing for *Millennials*, it is still smaller than all of the other generations at around 25 minutes. Comparatively, the *Silent Generation's* gap started at over 50 minutes in 2003 and has decreased to just under 45.

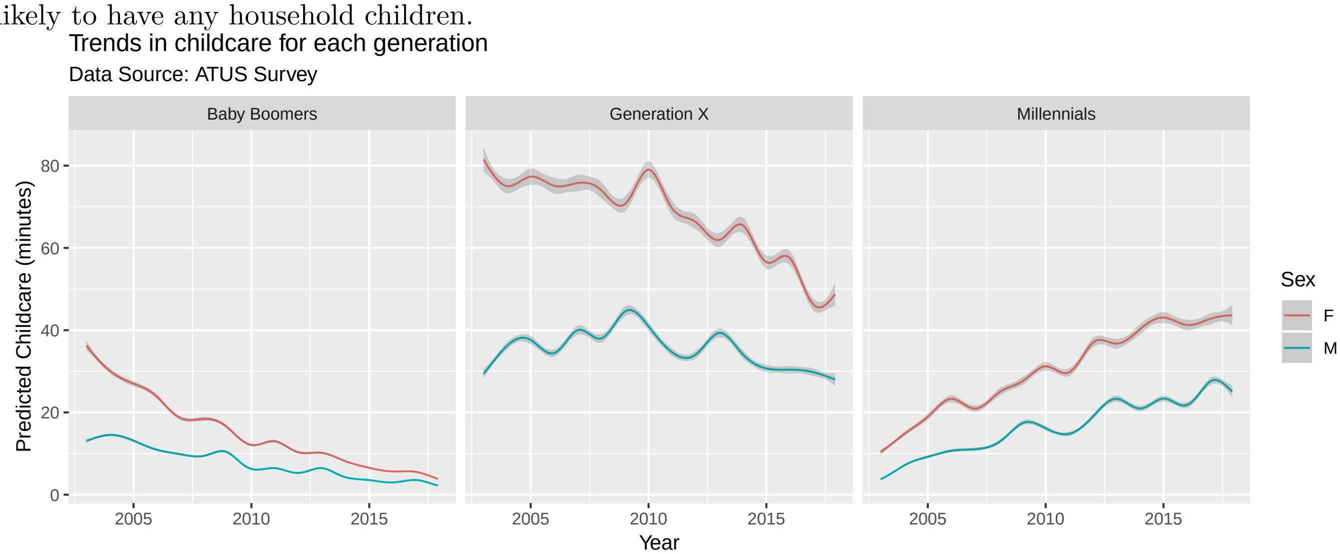
Cooking



The graphs above show a very similar trend to houseworkin the sense that the gap is decreasing and that all generations, apart from *Millennials* spend a similar amount of time in general on this activity. However, the difference with this activity is that for *Baby Boomers* and *Generation X*, women are actually spending more time on this activity than they were previously - the decrease in the gap is as a result of a sharper increase for men. For *Millenails*, this is not true: both sexes are spending more time on this activity but the increase for both is at an equal rate and so the gap is remaining equal. The *Silent Generation* is the generation showing the biggest decrease in the gap between sexes; this is due to the decrease for women and an increase for men, although this generation started off with the highest values for both sexes.

Childcare

Again for similar reasons to working, the *Silent Generation* have been excluded from this. Most are unlikely to have any household children.



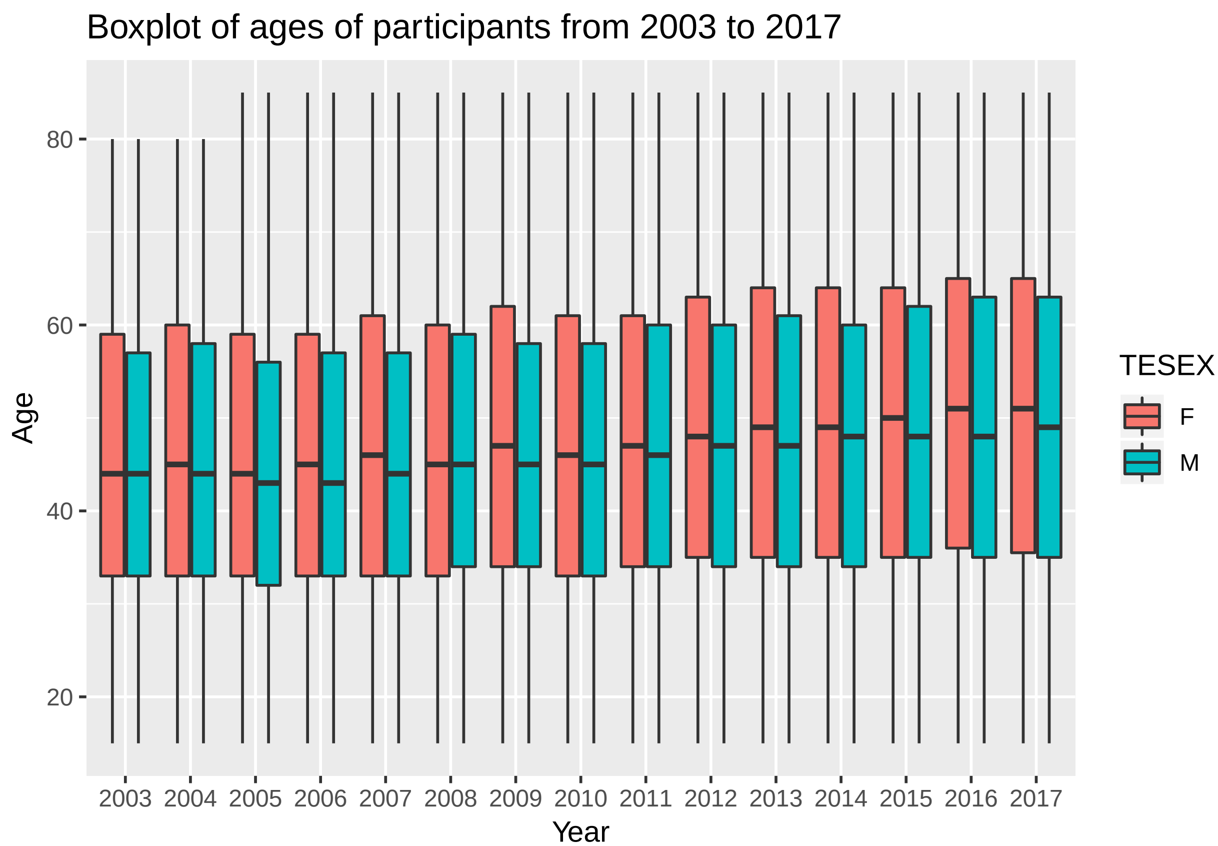
These graphs show perhaps the greatest change in gender roles. The data for *Baby Boomers* and the *Silent Generation* shows very little, most likely as the variables included in childcare relate only to caring for household children, where a child is defined as someone below the age of 18. It is unlikely that many people within these generations would have children living with them at this age (in fact checking the data confirms that, in 2017, only 9% met this criteria). Furthermore at this age, their children are likely to be older and require less constant care as they become more independent. Therefore, the other generations, *Millennials* and *Generation X*, reflect more interesting trends. For reasons similar to the older generations, the drop off in childcare at the end of *Generation X* can likely be explained as children mature; however, the drop off for women is sharper than it is for men, leading to a convergence in the lines. The gap has decreased substantially, from 50 minutes per day in 2003 to less than half that, around 20 minutes in 2017. This could further be attributed to the decrease in working time for this generation, noted above. The *Millennials* on the other hand relfect an increase year on year for childcaring activities. Whilst the gap increases slightly over time, it doesn't ever reach a similar point to that of *Generation X*, with the biggest difference being around 40% compared to that of 60%.

Limitations of the Data

15 years is a short period. Sporadic subset size due to the number of filters required to get down to the demographics we actually want to investigate.

The code below is to check the distribution of age group in those 15 years. It would not be a problem if there is a participant who have joined the survey several times, as her/his data would be regarded as independent samples in different years.

```
#save Dan's tidy dataset as another dataset, so no need to run the code every time opening r stud
write.csv(Data_minus_columns,'Dataset.csv')
Dataset <- import('Dataset.csv')
library(dplyr)
require(ggplot2)
library(dplyr)
age <- select(Dataset, TUYEAR, TEAGE,TESEX)
age$TUYEAR <- as.factor(age$TUYEAR)
p <- ggplot(data = age , aes(x=TUYEAR, y=TEAGE,fill=TESEX)) + geom_boxplot(width=0.8)
p + ggtitle('Boxplot of ages of participants from 2003 to 2017') + xlab('Year') + ylab('Age')
```



It is clear that the amount of participants is not well distributed and should be considered in the analysis. However, it could lead to the birth rate and death rate of the country.

References

[1] “ATUS datasets.” https://www.bls.gov/tus/datafiles_0317.htm.

[2] Bureau of Labor Statistics, “The american time use survey.” https://www.bls.gov/tus/, 2017.