Summary Report

Shixuan Cao

E-mail: S.Cao15@lse.ac.uk

The data processing steps:

I used the data in enut2021_diario.dta and enut2021_cautal.xlsx to implement the process of data processing.

First, the Excel column names were renamed to be more in line with the R language naming convention and to make them easier to understand and subsequently process. Next, activity information was organized by using pivot_longer() to convert multiple activity columns (activad_1, actividad_2, etc.) into a long format to better handle multiple simultaneous activities while preserving the original column name information and removing null rows to reduce redundant data. The data was then made more readable by mapping activity codes to more descriptive activity categories using left_join(). Then, time-dependent columns were generated to calculate the start and end time of each time slot and standardize the duration to 10 minutes for subsequent analysis. Finally, the cleaned data is inspected and saved to ensure that it is correct and then stored in .dta format using write_dta() for subsequent analysis.

In this way we will be able to generate cleaned data saved as a “diary”, where each line should indicate a time slot and its duration and the classification of the activity at the individual-diary level

Graphs and Key Insights

For visualization, I will focus on **Paid Work, Unpaid Work, and Personal Activity** since these three categories represent the primary types of human activities and collectively account for an individual's total time.

I will categorize the data based on two combinations: **sex and age group** as well as **sex and marriage status**. This approach aims to analyze time allocation across these three key activities from a gender perspective and derive meaningful insights.

I plan to analyze and interpret the data by **“Average Social Time”** and deliberately chose not to use **"Average Time per Participant"** as average because my focus is on understanding time allocation from the perspective of the entire male and female population, rather than just those who participated in a specific event, even if I coded both average value in code part as well as relevant graphs.

Average Social Time Allocation by Sex & Average Social Time Spent in Different Events (Paid Work, Unpaid Work, Personal Activity) by Age Group and Sex

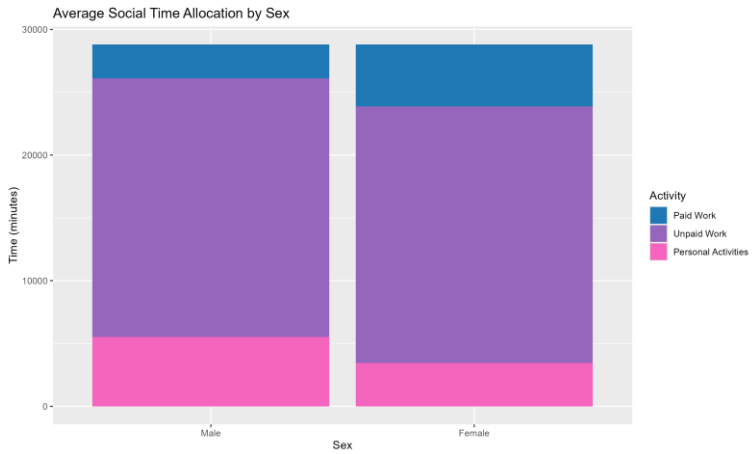

Figure 1: Average Social Time Allocation by Sex

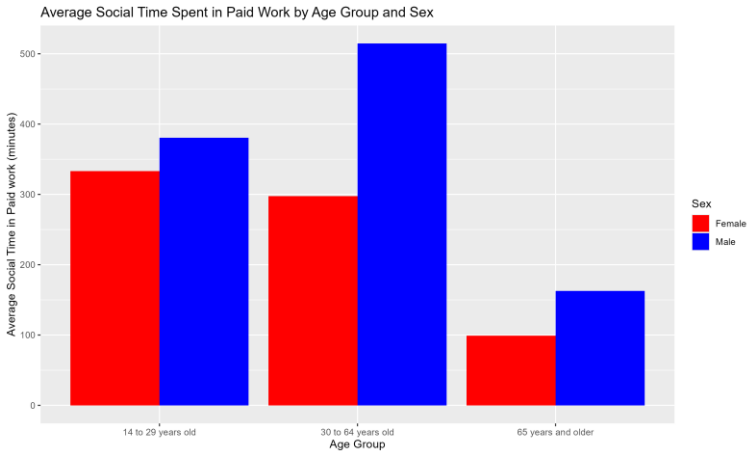

Figure 2: Average Social Time Spent in Paid Work by Age Group and Sex

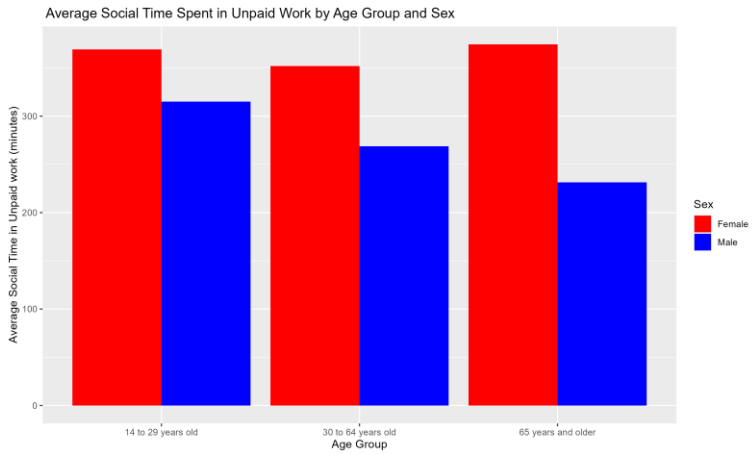

Figure 3: Average Social Time Spent in Unpaid Work by Age Group and Sex

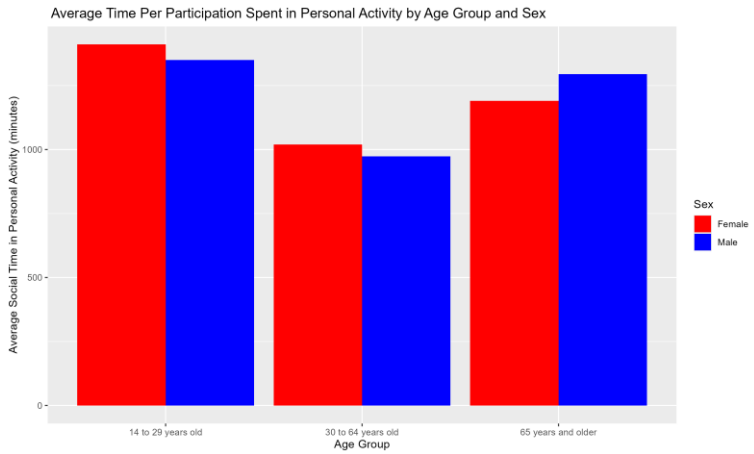

Figure 4: Average Social Time Spent in Personal Activity by Age Group and Sex

- Overall, in the age groups of 14–29, 30–64, and 65+, men spend significantly more time on paid work than women.
- Overall, in the age groups of 14–29, 30–64, and 65+, women spend significantly more time on unpaid work than men on paid work.

  Insight 1: Overall, men spend more time on paid work, while women spend more time on unpaid work and personal activities. This may reflect the influence of traditional gender roles in the allocation of family responsibilities. Women, while taking on more family responsibilities (such as childcare and household chores), are unable to devote as much time to paid work. This may lead to (1) a gender income gap: men spend more time on paid work and earn higher incomes, resulting in an income disparity between men and women; and (2) economic dependency: women spending more time on unpaid work may make them more economically dependent on men, further exacerbating gender inequality.
- Overall, in the age groups of 14–29 and 30–64, women spend slightly more time on personal activities than men. However, in the 65+ age group, men spend slightly more time on personal activities than women.

  Insight 2: This may reflect changes in gender roles before and after retirement:

  - Ages 14–29 and 30–64 (pre-retirement): Women spend more time on personal activities (such as leisure, socializing, and self-care), possibly because they place greater emphasis on personal health and social relationships. Meanwhile, men in these age groups may focus more on paid work and family responsibilities, resulting in less time for personal activities.
  - Age 65+ (post-retirement): Men spend more time on personal activities, possibly because they have more free time after retirement, while women may still take on certain family responsibilities (such as caring for a spouse or grandchildren).

Average Social Time Spent in Different Events (Paid Work, Unpaid Work, Personal Activity) by Marriage Status and Sex

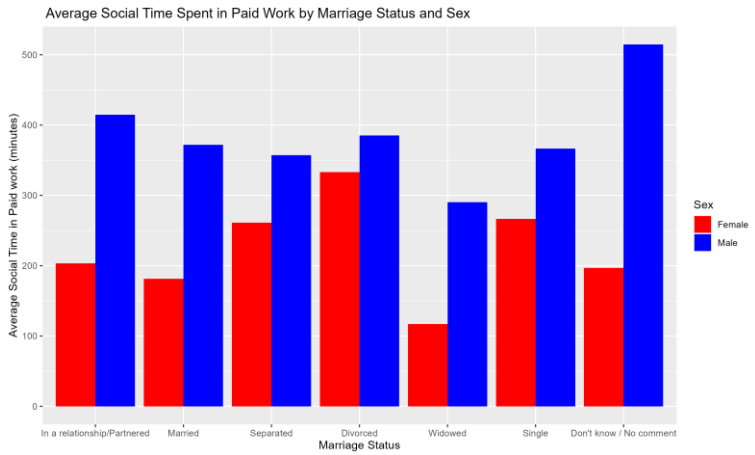

Figure 5: Average Social Time Spent in Paid Work by Marriage Status and Sex

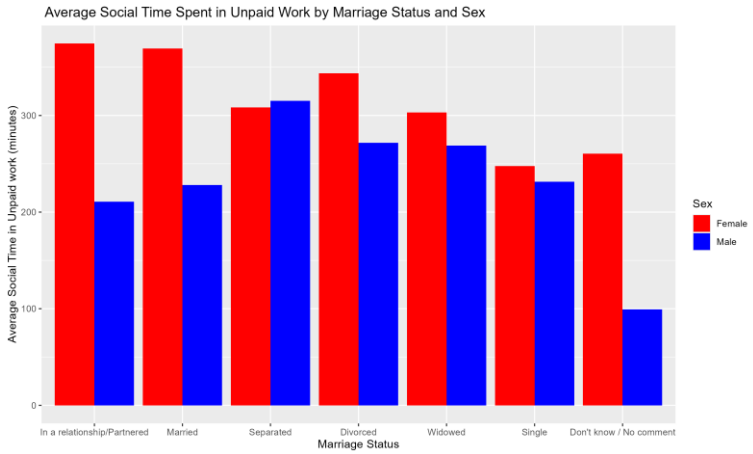

Figure 6: Average Social Time Spent in Unpaid Work by Marriage Status and Sex

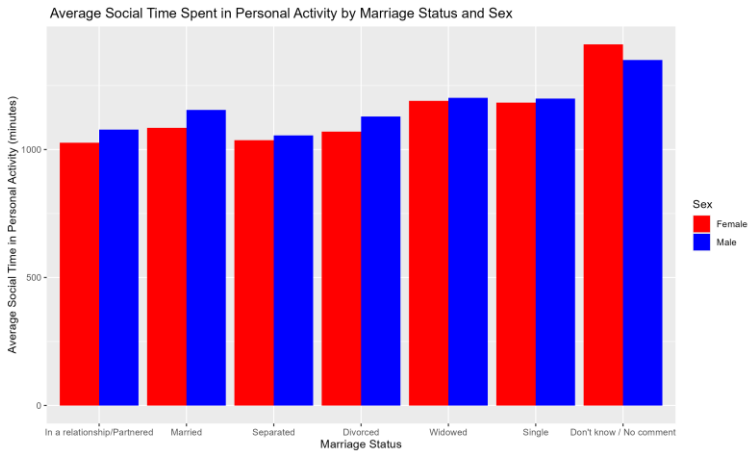

Figure 7: Average Social Time Spent in Personal Activity by Marriage Status and Sex

- Overall, in each marriage status, men spend significantly more time on paid work than women, with this gap being particularly pronounced in the states of "In a relationship/Partnered," "Married," and "Widowed."
- Overall, except for the "Separated" status, women spend significantly more time on unpaid work than men. In

the "Separated" status, men spend slightly more time on unpaid work than women. This gap is especially noticeable in the "In a relationship/Partnered" and "Married" states.

Insight 1: The above data characteristics may reflect the influence of marital status (whether stable) on gender roles and personal time allocation. In the states of "In a relationship/Partnered," "Married," and "Widowed," men may be more inclined to take on the role of the family's economic provider, while women tend to take on more household responsibilities. In these states, men still spend more time on paid work than women, but the gap may be relatively smaller. This could be because, in these states, women face greater economic pressure and are more inclined to work independently or take on more economic responsibilities. In the "Separated" status, men spend slightly more time on unpaid work than women, possibly because, after separation, men need to participate more in household affairs (such as caring for children or managing household chores).

- Overall, in each marriage status, the gap between men and women in the time spent on personal activities is very small.

  Insight 2: Despite the significant gender differences in paid and unpaid work, the time allocation for personal activities is relatively equal between men and women. This equality may reflect that personal activities are equally important to both men and women and are not significantly influenced by marital status.

Clear documentation of the cleaning process, including the steps taken and the reasons for each step：

1. **Import packages:**
   Import haven (for reading **.dta** files) **readxl** (for reading Excel files), **dplyr** and **tidyr** (for data organization and conversion)

2. **Read data**
   diario_data <- read_dta("enut2021_diario.dta") # Extract diary related data
   cautal_data <- read_excel("enut2021_cautal.xlsx",range = "A3:B60")#Read the activity classification(group) and corresponding activity code(code) in the excel table to facilitate the subsequent mapping process

3. **Processing excel column names**
   Rename the columns of cautal_data to actividad_type and actividad_code, the original corresponding columns are in Spanish and may not conform to R naming conventions, so renaming them will make it easier for me to understand and work with the data later.

4. **Processing activity information:**
   Use **pivot_longer()** to convert actividad_1, actividad_2, actividad_3, etc. columns to a long format to make it easier to deal with multiple simultaneous activities, and create a column so that time ID, slot, activity code, and acitivity type correspond to each other (here I plan to use multiple columns to represent different activities in a time slot for a single ID).

cols = starts_with(“actividad_”) & !starts_with(“actividad_hora”) & !starts_with(“actividad_minuto”): here, we select columns starting with actividad_ excluding actividad_hora and actividad_minuto. hora, because these two columns store time information other than the activity code(actividad_1, actividad_2, actividad_3) I want to process.

names_to = “actividad_source”: stores the original column names into the actividad_source column to keep track of the source of the activity code.

values_to = “individad_code”: stores the activity code into the actividad_code column to represent as well as correspond to the activity categorization.

Given the possibility of null values in the individad_2 and actividad_3 columns (which means that only one or two activities were going on at the time), we use **filter(!is.na(actividad_code))** to remove rows with null values for actividad_code to prevent redundant and to prevent redundant and meaningless lines from appearing

5. **Maps actividad_code to actividad_type:**
   Merge diario_data_long with cautal_data using **left_join()** to match the activity type (actividad_type) based on actividad_code. This step is to match the activity code to the more descriptive activity type name for task requirements.

6. **Generate time slot related columns:**
   Generate the start time, and duration of each time slot for calculation using mutate().
   start_hora and start_minuto: extracted directly from actividad_hora and actividad_minuto, representing the start time of the time slot of the time_slot.
   end_hora and end_minuto: calculate the end time of the time slot.
   time_slot_duracion_minuto: fixes the time slot length to 10 minutes.

7. **View cleaned dataset:**
   View the cleaned dataset using the **View()** function to check if the data is correct.

8. **Save the cleaned dataset:**
   Use **write_dta()** to save the cleaned dataset as a .dta file for subsequent analysis.