At L-IFT, we do financial diaries research. In essence, we survey individuals asking them various questions over a period of time. The difference between a survey as you might think and diaries, is that we ask the same questions every two weeks, for a period of between six months and a year. Every two weeks therefore we receive a new dataset (with answers to the same questions as the previous interview two weeks ago, i.e. the last biweekly) which has to be combined with the overall dataset so that we can do various analyses on these data, over time.

Please find several sheets attached. You will find three files containing survey results (biweeklies), one codebook, and one file containing segmentation variables. Segmentation variables include respondent’s information like age, gender, ppi score (what we use as a wealth proxy), etc. You will use a combination of these files to conduct the below tasks.

1. Merge the three datasets. The result should be a data.frame or data.table where each of the questions are organized under each other. Merge by Respondent ID, or name.
2. Subset the dataframe: use the codebook to find question A\_4 about amount of income earned to select only the relevant questions from the dataframe. (This is a multiple choice question and each answer option is a separate column.)
3. Create a frequency table of question A\_4 to see how many respondents answered each answer option. You can chose to first melt the data.table and then make the table.
4. Make a histogram/bar graph of question A\_4 with answer options on the x axis and count of respondents on the y axis. Label the title as “Count of respondents having an income earning activity”.
5. Find question Q\_12: how much respondents have added in each savings tool. Find the median and mean savings for the different types of savings tools used.
6. Given the segmentation file and the biweekly files that you have combine into one data frame find if level of stress (Q\_91) is associated with ppi score (variable named "ppicut" in this file).
7. Plot a graph to visualize above task 6.
8. Using variables VEnd and VStart find the average time each surveyor (variable “Srvyr”) used for interviews.
9. Using a plot, find which savings tools men and women are more likely to use.
10. Use one biweekly of your choice to plot a map of where our respondents are located.
11. Use any column of choice to make a pie chart and explain what the pie chart shows
12. Use dplyr in any question of choice to:
    1. Filter
    2. Select
    3. Groupby
    4. Summarise
    5. Graph (ggplot)
13. Explain various sampling methods that you know and explain how you would use R to implement them
14. Show two ways of dealing with Null values in R - explain the choice of method used.