PREDOC Sample Data Task

The file cps_women_lfp.csv[1] (also available as a .dta file) contains individual-level data from the U.S. Current Population Survey since 1994. The data set includes information on demographics and labor market outcomes. Please use the provided data to answer each question below in complete sentences. Please submit a pdf of your written responses and any graphs and/or tables that support your responses, as well as your code. You can use any programming language of your choice, though you may find the workshop more beneficial if you use R or STATA.

Please spend no more than 6 hours on this task. Many of these questions are open-ended and do not have a single "correct" answer. The goal is not to produce the most complex analysis, but rather to practice using data to explore labor market patterns and interpreting them.

Part 1: Labor Force Participation

Note: Please spend no more than 4 hours on this section.

1. How has female labor force participation evolved since 1994? Please provide graphs and/or tables to support your answer.

2. Among women older than 25, which groups (race, age, income percentile, etc.) of people had the biggest changes in labor force participation since 1994? Please provide at least three graphs and/or tables to support your answer.

3. Use the data to examine trends among women older than 25 for each of the following factors from 1994 to 2024:

   (a) Wage and salary income

   (b) Social insurance income

   (c) Education attainment

   Based on these trends, what factors could be driving the patterns you found in Questions 1 and 2?

4. Between 1994 and 2024, which year had the steepest increase in female labor force participation relative to the previous year? What factors do you think are driving this pattern? Support your answers by using the data, referencing major events that happened around this time period, and/or citing previous studies.

[1] Extracted from https://cps.ipums.org/cps/

5. How has labor force participation for college-educated and not college-educated women evolved since 1994? Please provide graphs and/or tables to support your answer.

6. Create an alternative measure of labor force participation that excludes individuals from the labor force if they are self-employed in their main job (*lfp = 0* if self-employed in main job). Using the new measure, describe how labor force participation for college-educated and not college-educated women has evolved since 1994. Please provide graphs and/or tables to support your answer.

7. How does our labor market analysis change when we use the new measure? Which measure do you prefer? Explain.

Part 2: Telework

Note: Please spend no more than 2 hours on this section.

1. Since the rise of telework in 2020, how have wages, employment, and labor force participation changed for women who had telework from 2020-2024 and women who did not? Please provide at least three graphs and/or tables to support your answer.

2. For which groups of women older than 25 was telework due to the pandemic most common in 2021? Based on these patterns, what can you infer about the relationship between economic well-being and the ability to telework between 2021? Please provide at least three graphs and/or tables to support your answer.

3. Predict what trends in wages, employment, and labor force participation for college-educated women from 2020 to 2024 would have looked like if telework was not an option. What does this tell you about the economic impacts of telework during the COVID-19 pandemic? Please support your answer with graphs and/or tables.

   Hint: Look at trends from previous years that had similar economic contexts. Also, feel free to explore the variables you haven't used yet.