Summary of Work Completed (Joy Bhattacharya)

1. **SMCE**: First, I worked on matching the school names from the dataset to schools from the 2 medio, 8 basico, and 4 basico SMCE datasets.
   1. I created a code that would find the best string match in general and the best match based on location (based on the location of the participant). You can find this code in the Jupyter notebook file called SMCE Matching Code in the Python 🡪 Jupyter Files filepath in dropbox
   2. Using this code, I made an excel document of the best matches called chile\_school\_comparison\_final in the data folder. In this document, I manually went through and found the best match for each participant based on the string matches, personal intuition, and Googling. You can find a key to this document called chile\_school\_comparison\_final\_key
   3. After I finished making the initial matches, I wanted to make sure I made as many matches as possible. So, I created a Jupyter notebook file called Finding Missing School Matches to find any remaining unmatched schools, trying to match them. I added the new matches to the chile\_school\_comparison\_final file. This yielded about a 60% match rate.
   4. With all the matches, I then merged the matched data to the bbdd\_chile\_final.dta dataset using the file called merging\_SMCE\_variables\_code.do. I merged based on school rbd.
   5. Now that I had the matches, I could also merge the SMCE data on schools, such as test scores, socioeconomic status, significances, and etcetera. I also cleaned the merged data by combining redundant variables, creating dummies for significance, location, and cod\_grupo (socioeconomic status), and reordering the data.
   6. I also created a dummy based on if the participant still lived in the same community as their school, called same\_community, using the capital\_comuna.dta file that had the capitalized name of the comuna they lived in. I used the Python file uppercase\_comuna\_s to create the capital\_comuna\_dta file because Stata was giving me trouble.
   7. All this can be found in the merging\_SMCE\_variables\_code.do file. I saved the cleaned data in the bbdd\_chile\_final\_cleaned.dta file.
2. **Employment information**: Next, I merged information on employment and unemployment based on comuna
   1. I merged the muniunemp.dta file with the bbdd\_chile\_final\_data file based on comuna\_s to get the employment and unemployment rates where the participant lives, tasaocup and tasadesocup respectively.
3. **Bio parent information**: Next, I collected information on if participants still lived with their bioparent
   1. In the Python file combining\_parents\_a3, I stored information on if the participant lives with their mom or lives with their dad into the dummies lives\_with\_mom and lives\_with\_dad. I saved this information in a file called chile\_parents.dta which stored idencuestas and the values of the dummies for each person
   2. Then, I merged the chile\_parents.dta file with the bbdd\_chile\_final\_cleaned. Once I had the information on if they currently lived with their mom or dad, I wanted to find if they had lived with their mom or dad in infancy (until age 13). To do this, created two new dummies, lives\_with\_mom\_inf and lives\_with\_dad\_inf, to check if they lived with their mom or dad until age 13 using responses to questions a13 and a15. Then, I created another dummy called lives\_with\_momdad\_inf if they lived with both their mom and dad until at least age 13.
4. **Testing instruments**
   1. I tested all SMCE instruments in the file instrumental\_variables\_SMCE.do
   2. I tested all ENE (employment) instruments in the file instrumental\_variables\_ENE.do
   3. I tested all bio-parent instruments in the file instrumental\_variables\_padresbio.do
   4. For each test, I included tests on relevance and exogeneity when applicable.
   5. I saved all results as tables, which you can find in the Tables🡪Instrumental Variables Aggshock Analysis or Tables🡪 Instrumental Variables Diaggshock Analysis. In the Diaggshock Analysis folder, I separated results based on if jobloss was instrumented or if other shock variables were instrumented.
   6. I summarized all results in the file Summary.xls which you can find in the Tables folder.
   7. In conclusion, none of the instruments I tested were good. Some were relevant and exogenous, but they didn’t lead to significant results.

Summary of Files Created

1. **Python folder:**
   1. Combining\_parents\_a3: This is the file where I scrape information on who they currently live with and store dummies if they currently live with their bio mom or bio dad. I store this information into a .dta file called chile\_parents.dta
   2. Uppercase\_comuna\_s: This is the file where I capitalize the comuna names in Python and store the data in capital\_comuna.dta to later be merged. I did this in Python because Stata wouldn’t let me capitalize with special Spanish characters
   3. SMCEmatching: This is the file where I first start the process of string matching. However, to do actual matched, I used the Jupyter File SMCE matching code because it was more efficient.
   4. Jupyter Files:
      1. SMCE matching code: This is the code where I create string matches and create the excel document that collects best match information, chile\_school\_comparison\_final.xls
      2. Finding missing school matches: In this file, I find any schools that are still unmatched after my initial matching process
2. **Do folder**
   1. Instrumental\_variables\_ENE: In this do file, I merge the data on employment and do the IV analysis on the tasaocup and tasadesocup variables
   2. Instrumental\_variables\_padresbio: In this do file, I merge the data on if they live with their bio parent. I also create dummies if they lived with their bio parent during infancy. I then do the IV analysis on these variables.
   3. Instrumental\_variables\_SMCE: In this do file, I do all the IV analysis on the different SMCE variables.
   4. IV\_mate2m\_difgru: In this file, I do the IV analysis on the variables prom\_mate2m\_rbd and difgru\_mate2m\_rbd that instrument the variable jobloss (in disagg) and individual shocks (in agg).
   5. Merging\_SMCE\_varaibles\_code: This is the file where I merge all the SMCE information with the bbdd\_final\_chile dataset and do all the cleaning and reordering.
3. **Data folder**
   1. **Bbdd\_final\_Chile\_cleaned**: The most important file. Here is the cleaned version of bbdd\_final\_chile with all merged data and dummies.
   2. Capital\_comuna: This is where I stored the capitalized comuna names by idencuesta
   3. Chile\_parents: This file contains idencuesta followed by dummies lives\_with\_mom and lives\_with\_dad to measure if the participant currently lives with their mom or dad
   4. Chile\_school\_comparison\_final: This is where the SMCE school matches are stored. This contains the original participant response (c27\_est) followed by the best match in general and the best match based on location.
   5. Chile\_school\_comparison\_final\_copy: This is a copy of chile\_school\_comparison\_final because I ran into problems with data loss.
   6. Chile\_school\_comparison\_final\_key: This is a key to the variables in chile\_school\_comparison\_final
4. **Tables folder**
   1. **Summary**: This is where I summarize all IV analysis performed along with helpful statistics on significance, relevance, and exogeneity
   2. Instrumental Variables Disaggshock Analysis
      1. Bioparent IVs: Contains the tables of Bioparent IV analysis, separated based on if jobloss was instrumented.
      2. ENE IVs: Contains the tables of employment IV analysis, separated based on if jobloss was instrumented.
      3. SMCE IVs: Contains the tables of SMCE IV analysis, separated based on if jobloss was instrumented.
   3. Instrumental Variables Aggshock Anslysis
      1. Bioparent IVs: Contains the tables of Bioparent IV analysis where individual shocks in the instrument
      2. ENE IVs: Contains the tables of employment IV analysis where individual shocks in the instrument
      3. SMCE IVs: Contains the tables of SMCE IV analysis where individual shocks in the instrument