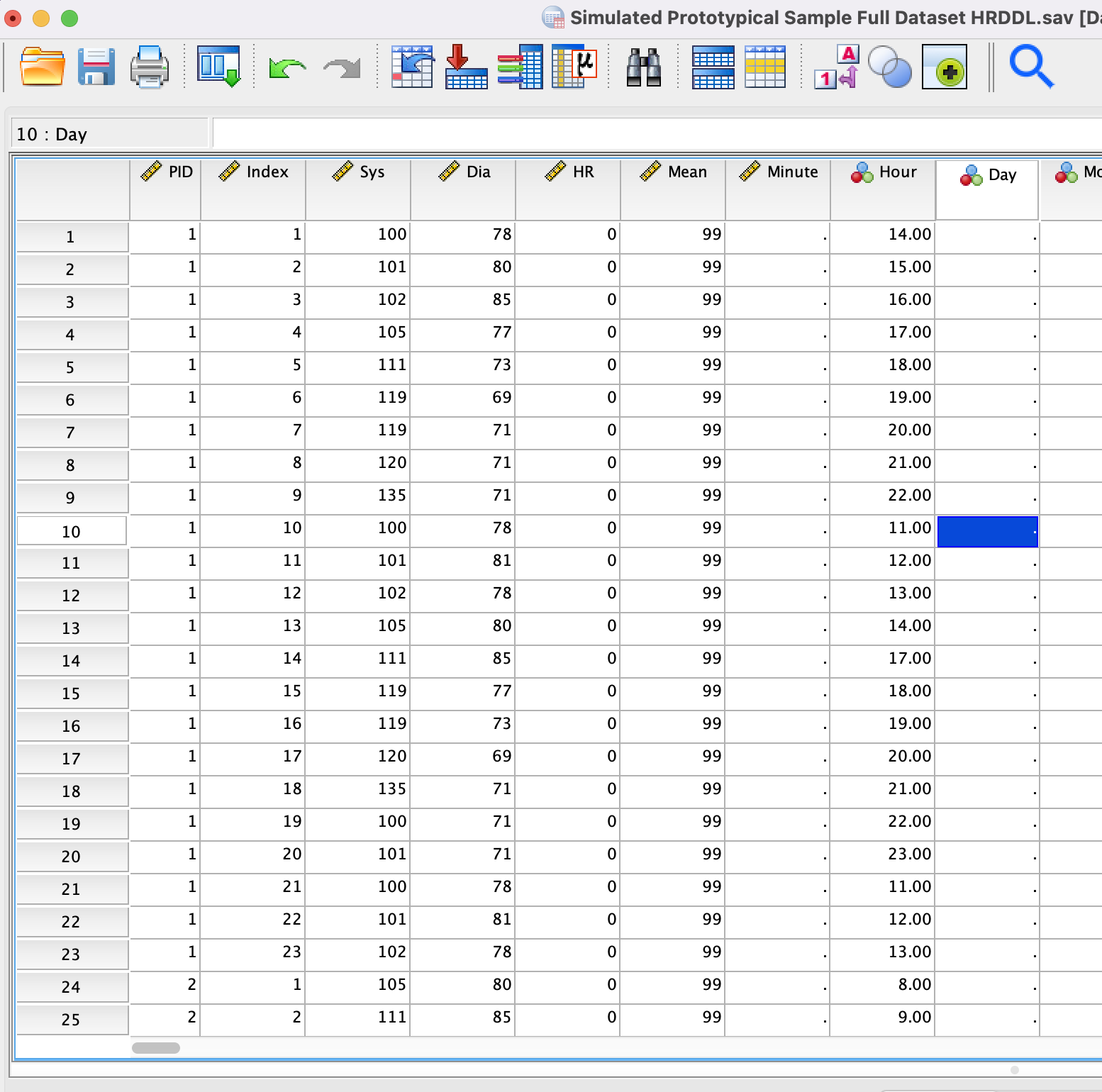
**Sample of What Final Dataset Should Look Like**

Screenshot below - full Excel file with full data structure for 2 sample participants in Google Drive folder - you will see a lot of “null” data in the Excel file - that just means that I didn’t put in mock data for that cell



**Study Protocol**

1. Participant comes in to office to take a survey and get measurements [Baseline Survey]
2. Participant goes home to do the following for four days:
   1. Take a survey every morning upon awakening [BOD survey]
   2. Have their blood pressure readings taken hourly throughout the day and night [Hourly BP]
   3. Take a survey after each hourly blood pressure reading that occurs while they are awake [Hourly EMA]
   4. Take a survey every night before sleep [EOD survey]
   5. Wear the actigraphy watch that collects physical activity and sleep data [Sleep] [Physical Activity]
3. Participant comes in to office to take a final survey [Final Survey]

**Mock files in Excel or .csv: in Google Drive folder**

Note: PID is the variable that contains a number for a specific mock participant that is the same throughout all files

Note: For the mock files, I only focused on creating mock data for the PID and time based variables that you need for the merge - I also put in a sample mock variable column or two beyond the necessary columns for the merge

Notes: I created them to include all the bugs that real files usually have

*Example Bugs that I incorporated into the simulated files*:

A lot of missing data, e.g., people that have baseline survey but no hourlies or no sleep data

Multiple BP readings in the same hour

Multiple EMA surveys taken in the same hour

Mock Data Files Included

Baseline Survey: includes all participants already together in one file

Final Survey: includes all participants already together in one file

Hourly Survey: includes all hours of all participants already together in one file

BOD Survey: includes all days of all participants already together in one file

EOD Survey: includes all days of all participants already together in one file

BP data: each participant has a separate file - each participant’s file contains all his/her BPs

Sleep data: includes all days for all participant

**Primary Merge Need**

Timestamps between EMA and BP files - EMA should have been taken within 10 minutes after BP but please match EMAs that are taken within 50 minutes after the BP or within the same hour

**Variables that Need to be Coded for in order to automatically show up in the final file**

Add the source of the data to the beginning of each variable - i.e., posture should be changed to BODPosture or HourlyPosture depending on whether it came from BOD or Hourly files

For each row of each participant, a variable indicating the Time Difference between the EMA and BP [with a negative number indicating the EMA was taken before the BP and positive number indicating the EMA was taken after the BP, which is what participants were told to do]

Monitoring day variable that indicates which day of the monitoring period it is for that participant

Lagged variables for EMA, BP, and sleep (for example, add the 11 am BP also to the 10 am row, the 12 pm BP also to the 11 am row, etc.- add Lag to the beginning of the variable name to separate it from the regular variable)

A variable for each participant that states the number of hourly surveys they completed

A variable for each participant that states the number of BPs they completed

A variable for each participant that states that number of hourly survey - BP matches they have

A variable for each BP row that indicates whether the BP was taken while awake or asleep [use sleep data that shows bedtime for this]

**Project Synopsis**

