NHATS_NCOS_datasets

2025-03-17

1st: NHATS datasets (caregivers/digital tools)

Focusing on Digital tools first:

```
# Load necessary libraries
library(haven)    # For reading SAS files
library(dplyr)    # For data manipulation
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(readr)    # For exporting CSV
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
# Set file paths (Update these based on your file locations)
tracker_data <- read_sas("~/desktop/Digital-tools_caregivers/NHATS_R13_Final_Release_SAS/Tracker_files/
tab_act <- read_sas("~/desktop/Digital-tools_caregivers/NHATS_R13_Final_Release_SAS/Tab_act_files/NHATS_

# Tab_Act => Covers assistive devices, telehealth, vision, hearing, mobility

# Save as CSV
# write.csv(tracker_data, "tracker_data.csv", row.names = FALSE)
# write.csv(tab_act, "tab_act.csv", row.names = FALSE)
```

```r
#Exploratory data analysis:

# # Display all column names
# colnames(tracker_data)
# colnames(tab_act)
#
# #Check Data Structure & Summary
# # View first few rows
# head(tracker_data)
# head(tab_act)
#
# # Summary statistics for numerical variables
# summary(tracker_data)
# summary(tab_act)
#
#
# # Check structure of the datasets
# str(tracker_data)
# str(tab_act)
#
# # Count unique values in selected variables
# table(tab_act$vh13vision)
```

```r
# Ensure participant ID is kept
id_column <- "spid"

# Functional Limitations (Vision & Hearing)
functional_columns <- c("vh13vision", "vh13hearing", "vb135lglasses", "vb135lcontacts", "vb135lothvisaid

# Readmission & Chronic Condition Care (searching for relevant terms)
readmission_columns <- grep("readmit|hospital", names(df), value = TRUE, ignore.case = TRUE)

# Caregiver Support & Digital Tools (searching for relevant terms)
caregiver_columns <- grep("care|support", names(df), value = TRUE, ignore.case = TRUE)

# Cognitive & Functional Ability Measures
cognitive_columns <- c("cb13onbspeed", "cb13onbstdev", "cb13onbacc", "cb13onbcorr", "cb13onberr",
                       "cb13onbcomp", "cb13onbinteg", "cb13dcogbatm")

# Combine all selected columns
selected_columns <- unique(c(id_column, functional_columns, readmission_columns, caregiver_columns, cogn


# Filter the dataset to keep only relevant columns
tab_act_filtered <- tab_act %>% select(all_of(selected_columns))

# View the filtered dataset
head(tab_act_filtered)
```

```
## # A tibble: 6 x 14
##       spid vh13vision vh13hearing vb135lglasses vb135lcontacts vb135lothvisaid
##      <dbl>      <dbl>       <dbl>         <dbl>          <dbl>           <dbl>
## 1 10000008          1           1             2              2               2
```

```
## 2 10000021          1           1             1              2              2
## 3 10000022         -1          -1            -1             -1             -1
## 4 10000036          1           1             2              2              2
## 5 10000041          1           1             1              2              2
## 6 10000043          2           2            -1             -1             -1
## # i 8 more variables: cb13onbspeed <dbl>, cb13onbstdev <dbl>, cb13onbacc <dbl>,
## #   cb13onbcorr <dbl>, cb13onberr <dbl>, cb13onbcomp <dbl>, cb13onbinteg <dbl>,
## #   cb13dcogbatm <dbl>
```

```r
# Rename columns to be more understandable
tab_act_filtered <- tab_act_filtered %>%
  rename(
    participant_id = spid,
    Vision_Status = vh13vision,
    Hearing_Status = vh13hearing,
    Uses_Glasses = vb135lglasses,
    Uses_Contacts = vb135lcontacts,
    Uses_Other_Visual_Aid = vb135lothvisaid,
    Cognitive_Speed = cb13onbspeed,
    Cognitive_Std_Dev = cb13onbstdev,
    Cognitive_Accuracy = cb13onbacc,
    Cognitive_Correct_Responses = cb13onbcorr,
    Cognitive_Errors = cb13onberr,
    Cognitive_Completion = cb13onbcomp,
    Cognitive_Integration = cb13onbinteg,
    Cognitive_Battery_Measure = cb13dcogbatm
  )

# View the renamed dataset
head(tab_act_filtered)
```

```
## # A tibble: 6 x 14
##   participant_id Vision_Status Hearing_Status Uses_Glasses Uses_Contacts
##            <dbl>         <dbl>          <dbl>        <dbl>         <dbl>
## 1       10000008             1              1            2             2
## 2       10000021             1              1            1             2
## 3       10000022            -1             -1           -1            -1
## 4       10000036             1              1            2             2
## 5       10000041             1              1            1             2
## 6       10000043             2              2           -1            -1
## # i 9 more variables: Uses_Other_Visual_Aid <dbl>, Cognitive_Speed <dbl>,
## #   Cognitive_Std_Dev <dbl>, Cognitive_Accuracy <dbl>,
## #   Cognitive_Correct_Responses <dbl>, Cognitive_Errors <dbl>,
## #   Cognitive_Completion <dbl>, Cognitive_Integration <dbl>,
## #   Cognitive_Battery_Measure <dbl>
```

```r
# Define the relevant columns
selected_columns <- c("spid", "r13panel", "r13status", "r13casestdtmt", "r13casestdtyr",
                      "r13spstat1", "r13spstat2", "r13fqstat",
                      "r12status", "r11status")

# Filter the dataset to keep only relevant columns
tracker_data_filtered <- tracker_data %>% select(all_of(selected_columns))
```

```r
# Rename the columns to meaningful names
tracker_data_filtered <- tracker_data_filtered %>%
  rename(
    participant_id = spid,
    Panel_ID = r13panel,
    Overall_Case_Status = r13status,
    Case_Status_Month = r13casestdtmt,
    Case_Status_Year = r13casestdtyr,
    Participant_Interview_Status_1 = r13spstat1,
    Participant_Interview_Status_2 = r13spstat2,
    Facility_Questionnaire_Status = r13fqstat,
    Previous_Year_Status = r12status,
    Two_Years_Ago_Status = r11status
  )

# View renamed dataset
head(tracker_data_filtered)
```

```
## # A tibble: 6 x 10
##   participant_id Panel_ID Overall_Case_Status Case_Status_Month Case_Status_Year
##            <dbl>    <dbl>               <dbl>              <dbl>            <dbl>
## 1       10000001       -1                  -1                 -1               -1
## 2       10000002       -1                  -1                 -1               -1
## 3       10000003       -1                  -1                 -1               -1
## 4       10000004       -1                  -1                 -1               -1
## 5       10000005       -1                  -1                 -1               -1
## 6       10000006       -1                  -1                 -1               -1
## # i 5 more variables: Participant_Interview_Status_1 <dbl>,
## #   Participant_Interview_Status_2 <dbl>, Facility_Questionnaire_Status <dbl>,
## #   Previous_Year_Status <dbl>, Two_Years_Ago_Status <dbl>
```

```r
# We are using tracker data because it helps remove ineligible cases (e.g., deceased participants or re
```

```r
# Merge datasets using left joins on 'participant_id'
merged_digital_tools <- tab_act_filtered %>%
  left_join(tracker_data_filtered, by = "participant_id")

# View summary of merged dataset
glimpse(merged_digital_tools)
```

```
## Rows: 8,597
## Columns: 23
## $ participant_id                <dbl> 10000008, 10000021, 10000022, 10000036,~
## $ Vision_Status                 <dbl> 1, 1, -1, 1, 1, 2, 2, 2, 1, 1, 1, -1, -~
## $ Hearing_Status                <dbl> 1, 1, -1, 1, 1, 2, 2, 2, 1, 1, 1, -1, -~
## $ Uses_Glasses                  <dbl> 2, 1, -1, 2, 1, -1, -9, -1, 2, 2, 1, -1~
## $ Uses_Contacts                 <dbl> 2, 2, -1, 2, 2, -1, -9, -1, 2, 2, 2, -1~
## $ Uses_Other_Visual_Aid         <dbl> 2, 2, -1, 2, 2, -1, -9, -1, 2, 2, 2, -1~
## $ Cognitive_Speed               <dbl> 2.89484, 3.15880, -1.00000, 3.03725, 3.~
## $ Cognitive_Std_Dev             <dbl> 0.14733, 0.14992, -1.00000, 0.15872, 0.~
## $ Cognitive_Accuracy            <dbl> 0.76300, 1.05427, -1.00000, 1.32206, 0.~
## $ Cognitive_Correct_Responses   <dbl> 32, 31, -1, 31, 31, -9, -9, -9, 31, 31,~
```

```
## $ Cognitive_Errors                <dbl> 35, 10, -1, 2, 32, -9, -9, -9, 2, 1, 26~
## $ Cognitive_Completion            <dbl> 1, 1, -1, 1, 1, -9, -9, -9, 1, 1, 1, -1~
## $ Cognitive_Integration           <dbl> 1, 1, -1, 1, 1, -9, -9, -9, 1, 1, 1, -1~
## $ Cognitive_Battery_Measure       <dbl> 6, 6, 1, 6, 6, 4, 4, 4, 6, 6, 6, 1, 1, ~
## $ Panel_ID                        <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ Overall_Case_Status             <dbl> 60, 60, 62, 60, 60, 60, 60, 60, 60, 60,~
## $ Case_Status_Month               <dbl> 8, 6, 6, 6, 7, 7, 9, 6, 8, 8, 7, 7, 8, ~
## $ Case_Status_Year                <dbl> 2023, 2023, 2023, 2023, 2023, 2023, 202~
## $ Participant_Interview_Status_1 <dbl> 20, 20, 20, 20, 20, 20, 20, 20, 20, 20,~
## $ Participant_Interview_Status_2 <dbl> 20, 20, 24, 20, 20, 20, 20, 20, 20, 20,~
## $ Facility_Questionnaire_Status  <dbl> 24, 24, 24, 24, 24, 24, 24, 24, 24, 24,~
## $ Previous_Year_Status            <dbl> 60, 60, 60, 60, 60, 60, 60, 60, 60, 60,~
## $ Two_Years_Ago_Status            <dbl> 60, 60, 60, 60, 60, 60, 60, 60, 60, 60,~
```

```
# Save the final merged dataset as a CSV file
write_csv(merged_digital_tools, "~/desktop/Digital-tools_caregivers/NHATS_Merged_Digital_Tools.csv")
```

Next, working on the remaining datasets that focus more on caregivers, patients, and merge them with the cleaned digital tools dataset:

```
# Load libraries
library(haven)      # To read .sas7bdat files
library(tidyverse)  # For data wrangling
library(janitor)    # For cleaning column names
```

```
##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(skimr)      # For summarizing datasets

# Load NHATS data files
nhats_sp <- read_sas("~/desktop/Digital-tools_caregivers/NHATS_R13_Final_Release_SAS/SP_files/NHATS_Rou
nhats_op <- read_sas("~/desktop/Digital-tools_caregivers/NHATS_R13_Final_Release_SAS/OP_files/NHATS_Rou
merged_digital_tools <- read_csv("~/desktop/Digital-tools_caregivers/NHATS_Merged_Digital_tools.csv") #
```

```
## Rows: 8597 Columns: 23

## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (23): participant_id, Vision_Status, Hearing_Status, Uses_Glasses, Uses_...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
inc_path <- read_sas("~/desktop/Digital-tools_caregivers/NHATS_R13_Final_Release_SAS/Inc_files/NHATS_R13

# Clean column names
nhats_sp <- nhats_sp %>% clean_names()
nhats_op <- nhats_op %>% clean_names()
merged_digital_tools <- merged_digital_tools %>% clean_names()
nhats_inc <- inc_path %>% clean_names()  # Income & Socioeconomic Data

# Check variables
# colnames(nhats_sp)
# colnames(nhats_op)
# colnames(merged_digital_tools)
# colnames(nhats_inc)
```

```r
# Save the final merged dataset as a CSV file
# write_csv(nhats_op, "nhats_op.csv")
# write_csv(nhats_sp, "nhats_sp.csv")
# write_csv(inc_path, "inc_path.csv")
```

Income dataset

```r
# Define the relevant columns
selected_columns_inc_path <- c("spid", "ia13toincimif", "ia13dtoincimi1",
                               "ia13dtoincimi2", "ia13dtoincimi3",
                               "ia13dtoincimi4", "ia13dtoincimi5", "ia13dtoincimreas")

# Filter the dataset to keep only relevant columns
inc_data_filtered <- nhats_inc %>% select(all_of(selected_columns_inc_path))

# View filtered dataset
head(inc_data_filtered)
```

```
## # A tibble: 6 x 8
##       spid ia13toincimif ia13dtoincimi1 ia13dtoincimi2 ia13dtoincimi3
##      <dbl>         <dbl>          <dbl>          <dbl>          <dbl>
## 1 10000008            -1          38000          38000          38000
## 2 10000021            -1          45000          45000          45000
## 3 10000022            -1             -1             -1             -1
## 4 10000036            -1          80000          80000          80000
## 5 10000041            -1           2000           2000           2000
## 6 10000043            -1          12000          12000          12000
## # i 3 more variables: ia13dtoincimi4 <dbl>, ia13dtoincimi5 <dbl>,
## #   ia13dtoincimreas <dbl>
```

```r
# Rename the columns to meaningful names
inc_data_filtered <- inc_data_filtered %>%
  rename(
    participant_id = spid,
    Total_Imputed_Income = ia13toincimif,
    Income_Source_1 = ia13dtoincimi1,
```

```
    Income_Source_2 = ia13dtoincimi2,
    Income_Source_3 = ia13dtoincimi3,
    Income_Source_4 = ia13dtoincimi4,
    Income_Source_5 = ia13dtoincimi5,
    Missing_Income_Reason = ia13dtoincimreas
  )

# View renamed dataset
head(inc_data_filtered)
```

```
## # A tibble: 6 x 8
##   participant_id Total_Imputed_Income Income_Source_1 Income_Source_2
##            <dbl>                <dbl>           <dbl>           <dbl>
## 1       10000008                   -1           38000           38000
## 2       10000021                   -1           45000           45000
## 3       10000022                   -1              -1              -1
## 4       10000036                   -1           80000           80000
## 5       10000041                   -1            2000            2000
## 6       10000043                   -1           12000           12000
## # i 4 more variables: Income_Source_3 <dbl>, Income_Source_4 <dbl>,
## #   Income_Source_5 <dbl>, Missing_Income_Reason <dbl>
```

#SP dataset

```
# Select and rename relevant variables from NHATS_SP (Older Adults Data)
nhats_sp_filtered <- nhats_sp %>%
  select(
    spid,               # Unique respondent ID
    r13dresid,          # Residential status (community, assisted living, nursing home)
    r13dgender,         # Gender of respondent
    r13d2intvrage,      # Age of respondent
    hc13disescn1, hc13disescn2, hc13disescn3, hc13disescn4, hc13disescn5,
    hc13disescn6, hc13disescn7, hc13disescn8, hc13disescn9, hc13disescn10,  # Chronic diseases
    is13proxlivsp,      # Does caregiver live with respondent?
    is13prxyrelat,      # Relationship of the proxy caregiver
    is13prxygendr,      # Gender of the proxy caregiver
    em13paydevce1:em13paydevce6,  # Assistive technology usage
    te13intrntmd2, te13intrntmd3, te13intrntmd4,
    te13computer, te13tablet, te13emailtext, te13online,
    te13shoponli1, te13shoponli2, te13shoponli3, te13socialnet,
#Post acute care related
    hc13hosptstay, hc13hosovrnht,
# NEW: cognitive and functional decline variables
    cp13chgthink1:cp13chgthink8,  # Changes in thinking/memory
    mo13bedslf, mo13bedwout,      # Bed mobility
    rh13impactiv, rh13impcomp, rh13imphh,  # Limitations in daily life
    rh13funcback, rh13funcknees, rh13funcfeet, rh13funcwrist, # Body part limitations
    pc13walk6blks, pc13up10stair, pc13car20pnds # Mobility
  ) %>%

  rename(
    participant_id = spid,
    residence_type = r13dresid,         # Residential status (community, assisted living, nursing home)
```

```
    respondent_gender = r13dgender,      # Gender of respondent
    respondent_age = r13d2intvrage,      # Age of respondent
    has_heart_disease = hc13disescn1,
    has_hypertension = hc13disescn2,
    has_lung_disease = hc13disescn3,
    has_diabetes = hc13disescn4,
    has_cancer = hc13disescn5,
    has_stroke = hc13disescn6,
    has_arthritis = hc13disescn7,
    has_dementia = hc13disescn8,
    has_depression = hc13disescn9,
    has_other_chronic = hc13disescn10,
    caregiver_lives_with_respondent = is13proxlivsp, # Does the caregiver live with respondent?
    caregiver_relationship = is13prxyrelat, # Relationship of the proxy caregiver
    caregiver_gender = is13prxygendr, # Gender of the proxy caregiver
    vision_aids = em13paydevce1, # Glasses, magnifiers
    hearing_aids = em13paydevce2, # Hearing aids
    cane = em13paydevce3, # Cane
    walker = em13paydevce4, # Walker
    wheelchair = em13paydevce5, # Wheelchair
    other_mobility_aid = em13paydevce6, # Grab bars, raised toilet seats
    telehealth_use = te13intrntmd2,          # Used internet for telehealth
    insurance_info_online = te13intrntmd3,   # Looked up insurance info online
    health_info_online = te13intrntmd4,      # Searched for health information online
    has_computer = te13computer,             # Owns and uses a computer
    has_tablet = te13tablet,                 # Owns and uses a tablet
    uses_email_or_text = te13emailtext,      # Uses email or text messaging
    uses_online_services = te13online,       # Uses a computer for online tasks
    orders_groceries_online = te13shoponli1,# Orders groceries online
    online_banking = te13shoponli2,          # Uses online banking
    orders_prescriptions_online = te13shoponli3, # Orders prescription refills online
    social_media_use = te13socialnet,         # Uses social networking sites
    had_hosp_stay_12mo = hc13hosptstay,
    num_hosp_stays = hc13hosovrnht,

    # Cognitive change indicators
    change_memory = cp13chgthink1,
    change_decisionmaking = cp13chgthink2,
    change_following_instructions = cp13chgthink3,
    change_concentration = cp13chgthink4,
    change_language = cp13chgthink5,
    change_organization = cp13chgthink6,
    change_multitasking = cp13chgthink7,
    change_recognition = cp13chgthink8,

    # Functional mobility
    can_get_out_of_bed_alone = mo13bedslf,
    can_get_out_of_bed_with_aid = mo13bedwout,

    # Activity limitations
    needs_help_activities = rh13impactiv,
    needs_help_comprehension = rh13impcomp,
    needs_help_household = rh13imphh,
```

```
    # Physical function impairments
    back_limitation = rh13funcback,
    knee_limitation = rh13funcknees,
    foot_limitation = rh13funcfeet,
    wrist_limitation = rh13funcwrist,

    # Strength/mobility tests
    walk_six_blocks = pc13walk6blks,
    climb_stairs = pc13up10stair,
    lift_20_pounds = pc13car20pnds
  ) %>%
  mutate(
    residence_type = case_when(
      residence_type == 1 ~ "Community",
      residence_type == 2 ~ "Residential Care",
      residence_type == 3 ~ "Nursing Home",
      residence_type == 4 ~ "Assisted Living",
      TRUE ~ NA_character_
    ),
    respondent_gender = case_when(
      respondent_gender == 1 ~ "Male",
      respondent_gender == 2 ~ "Female",
      TRUE ~ NA_character_
    ),
    respondent_age = case_when(
      respondent_age == 1 ~ "65-69",
      respondent_age == 2 ~ "70-74",
      respondent_age == 3 ~ "75-79",
      respondent_age == 4 ~ "80-84",
      respondent_age == 5 ~ "85-89",
      respondent_age == 6 ~ "90+",
      TRUE ~ NA_character_
    ),
    caregiver_lives_with_respondent = case_when(
      caregiver_lives_with_respondent == 1 ~ "Yes",
      caregiver_lives_with_respondent == 2 ~ "No",
      TRUE ~ NA_character_
    ),
    caregiver_relationship = case_when(
      caregiver_relationship == 1 ~ "Spouse/Partner",
      caregiver_relationship == 2 ~ "Child",
      caregiver_relationship == 3 ~ "Sibling",
      caregiver_relationship == 4 ~ "Other Relative",
      caregiver_relationship == 5 ~ "Friend/Neighbor",
      caregiver_relationship == 6 ~ "Paid Caregiver",
      caregiver_relationship == 7 ~ "Other Non-Relative",
      caregiver_relationship == 8 ~ "Unknown",
      TRUE ~ NA_character_
    ),
    caregiver_gender = case_when(
      caregiver_gender == 1 ~ "Male",
      caregiver_gender == 2 ~ "Female",
      TRUE ~ NA_character_
```

```
    ),
    # Convert assistive technology usage (1=Yes, 2=No, -9=Missing)
    across(starts_with("vision_aids"):starts_with("other_mobility_aid"), ~ case_when(
      . == 1 ~ "Yes",
      . == 2 ~ "No",
      TRUE ~ "Missing"
    )),
    # Convert digital tool usage (1=Yes, 2=No, -9=Missing)
    across(starts_with("telehealth_use"):starts_with("social_media_use"), ~ case_when(
      . == 1 ~ "Yes",
      . == 2 ~ "No",
      TRUE ~ "Missing"
    )),
     # Existing binary conversions
  had_hosp_stay_12mo_bin = case_when(
    had_hosp_stay_12mo == 1 ~ 1,
    had_hosp_stay_12mo == 2 ~ 0,
    TRUE ~ NA_real_
  ),
  multiple_hosp_stays = case_when(
    num_hosp_stays >= 2 ~ 1,
    num_hosp_stays < 2 ~ 0,
    TRUE ~ NA_real_
  ),

  # New: Categorized version of number of hospital stays
  hosp_stay_category = case_when(
    num_hosp_stays == 0 ~ "None",
    num_hosp_stays == 1 ~ "One",
    num_hosp_stays >= 2 ~ "Multiple",
    TRUE ~ NA_character_
  )
  )

write_csv(nhats_sp_filtered, "~/desktop/Digital-tools_caregivers/nhats_sp_filtered.csv")
```

#OP dataset

```
# Select and rename relevant variables from NHATS_OP (Caregivers Data)
nhats_op_filtered <- nhats_op %>%
  select(
    spid,              # Unique respondent ID (for merging)
    op13relatnshp,     # Relationship of caregiver to respondent
    op13proxy,         # Whether caregiver answered on behalf of respondent
    op13dage,          # Age of the caregiver
    op13numhrsday,     # Hours of caregiving per day
    op13numdayswk      # Days of caregiving per week
  ) %>%
  rename(
    participant_id = spid,
    caregiver_relationship_to_respondent = op13relatnshp, # Relationship of caregiver to respondent
    caregiver_answered_for_respondent = op13proxy, # Whether caregiver answered on behalf of respondent
    caregiver_age = op13dage, # Age of the caregiver
```

```
    caregiver_hours_per_day = op13numhrsday, # Hours of caregiving per day
    caregiver_days_per_week = op13numdayswk # Days of caregiving per week
  )%>%
  mutate(
    caregiver_relationship_to_respondent = case_when(
      caregiver_relationship_to_respondent == 1 ~ "Spouse/Partner",
      caregiver_relationship_to_respondent == 2 ~ "Child",
      caregiver_relationship_to_respondent == 3 ~ "Sibling",
      caregiver_relationship_to_respondent == 4 ~ "Other Relative",
      caregiver_relationship_to_respondent == 5 ~ "Friend/Neighbor",
      caregiver_relationship_to_respondent == 6 ~ "Paid Caregiver",
      caregiver_relationship_to_respondent == 7 ~ "Other Non-Relative",
      caregiver_relationship_to_respondent == 8 ~ "Unknown",
      TRUE ~ NA_character_
    ),

    caregiver_answered_for_respondent = case_when(
      caregiver_answered_for_respondent == 1 ~ "Yes",
      caregiver_answered_for_respondent == 2 ~ "No",
      TRUE ~ NA_character_
    ),

    caregiver_age = case_when(
      caregiver_age == -9 ~ NA_real_,  # Replace missing values
      TRUE ~ caregiver_age  # Keep actual age values
    ),

    caregiver_hours_per_day = case_when(
      caregiver_hours_per_day == -9 ~ NA_real_,
      caregiver_hours_per_day == -1 ~ NA_real_,
      TRUE ~ caregiver_hours_per_day
    ),

    caregiver_days_per_week = case_when(
      caregiver_days_per_week == -9 ~ NA_real_,
      caregiver_days_per_week == -1 ~ NA_real_,
      TRUE ~ caregiver_days_per_week
    )
  )
```

Merge Caregiver & Digital Tool Use Data

Merge NHATS caregiving data with digital tool usage

```
# Merge NHATS_SP with NHATS_OP (Caregiver + Care Recipient Data)
nhats_merged <- nhats_sp_filtered %>%
  left_join(nhats_op_filtered, by = "participant_id") %>%
  left_join(merged_digital_tools, by = "participant_id") %>%
  left_join(inc_data_filtered, by = "participant_id")

# Check merged dataset
glimpse(nhats_merged)
```

```
## Rows: 62,863
```

```
## Columns: 93
## $ participant_id                      <dbl> 10000008, 10000008, 10000008, 100~
## $ residence_type                      <chr> "Community", "Community", "Commun~
## $ respondent_gender                   <chr> "Male", "Male", "Male", "Male", "~
## $ respondent_age                      <chr> "85-89", "85-89", "85-89", "85-89~
## $ has_heart_disease                   <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~
## $ has_hypertension                    <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~
## $ has_lung_disease                    <dbl> 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, ~
## $ has_diabetes                        <dbl> 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, ~
## $ has_cancer                          <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~
## $ has_stroke                          <dbl> 7, 7, 7, 7, 7, 7, 7, 7, 7, 2, 2, ~
## $ has_arthritis                       <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~
## $ has_dementia                        <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~
## $ has_depression                      <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~
## $ has_other_chronic                   <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~
## $ caregiver_lives_with_respondent     <chr> NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ caregiver_relationship              <chr> NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ caregiver_gender                    <chr> NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ vision_aids                         <chr> "No", "No", "No", "No", "No", "No~
## $ hearing_aids                        <chr> "No", "No", "No", "No", "No", "No~
## $ cane                                <chr> "No", "No", "No", "No", "No", "No~
## $ walker                              <chr> "No", "No", "No", "No", "No", "No~
## $ wheelchair                          <chr> "No", "No", "No", "No", "No", "No~
## $ other_mobility_aid                  <chr> "No", "No", "No", "No", "No", "No~
## $ telehealth_use                      <chr> "Missing", "Missing", "Missing", ~
## $ insurance_info_online               <chr> "Missing", "Missing", "Missing", ~
## $ health_info_online                  <chr> "Missing", "Missing", "Missing", ~
## $ has_computer                        <chr> "No", "No", "No", "No", "No", "No~
## $ has_tablet                          <chr> "No", "No", "No", "No", "No", "No~
## $ uses_email_or_text                  <chr> "No", "No", "No", "No", "No", "No~
## $ uses_online_services                <chr> "No", "No", "No", "No", "No", "No~
## $ orders_groceries_online             <chr> "Missing", "Missing", "Missing", ~
## $ online_banking                      <chr> "Missing", "Missing", "Missing", ~
## $ orders_prescriptions_online         <chr> "Missing", "Missing", "Missing", ~
## $ social_media_use                    <chr> "Missing", "Missing", "Missing", ~
## $ had_hosp_stay_12mo                  <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~
## $ num_hosp_stays                      <dbl> -1, -1, -1, -1, -1, -1, -1, -1, -~
## $ change_memory                       <dbl> -1, -1, -1, -1, -1, -1, -1, -1, -~
## $ change_decisionmaking               <dbl> -1, -1, -1, -1, -1, -1, -1, -1, -~
## $ change_following_instructions       <dbl> -1, -1, -1, -1, -1, -1, -1, -1, -~
## $ change_concentration                <dbl> -1, -1, -1, -1, -1, -1, -1, -1, -~
## $ change_language                     <dbl> -1, -1, -1, -1, -1, -1, -1, -1, -~
## $ change_organization                 <dbl> -1, -1, -1, -1, -1, -1, -1, -1, -~
## $ change_multitasking                 <dbl> -1, -1, -1, -1, -1, -1, -1, -1, -~
## $ change_recognition                  <dbl> -1, -1, -1, -1, -1, -1, -1, -1, -~
## $ can_get_out_of_bed_alone            <dbl> -1, -1, -1, -1, -1, -1, -1, -1, -~
## $ can_get_out_of_bed_with_aid         <dbl> -1, -1, -1, -1, -1, -1, -1, -1, -~
## $ needs_help_activities               <dbl> -1, -1, -1, -1, -1, -1, -1, -1, -~
## $ needs_help_comprehension            <dbl> -1, -1, -1, -1, -1, -1, -1, -1, -~
## $ needs_help_household                <dbl> -1, -1, -1, -1, -1, -1, -1, -1, -~
## $ back_limitation                     <dbl> -1, -1, -1, -1, -1, -1, -1, -1, -~
## $ knee_limitation                     <dbl> -1, -1, -1, -1, -1, -1, -1, -1, -~
## $ foot_limitation                     <dbl> -1, -1, -1, -1, -1, -1, -1, -1, -~
## $ wrist_limitation                    <dbl> -1, -1, -1, -1, -1, -1, -1, -1, -~
```

```
## $ walk_six_blocks                      <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ climb_stairs                         <dbl> -1, -1, -1, -1, -1, -1, -1, -1, -~
## $ lift_20_pounds                       <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ had_hosp_stay_12mo_bin               <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ multiple_hosp_stays                  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ hosp_stay_category                   <chr> NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ caregiver_relationship_to_respondent <chr> NA, NA, "Other Relative", "Other ~
## $ caregiver_answered_for_respondent    <chr> NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ caregiver_age                        <dbl> -1, -1, 10, 9, 9, 8, -1, 7, -8, -~
## $ caregiver_hours_per_day              <dbl> NA, NA, NA, NA, 2, 2, NA, NA, NA,~
## $ caregiver_days_per_week              <dbl> NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ vision_status                        <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ hearing_status                       <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ uses_glasses                         <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 1, ~
## $ uses_contacts                        <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~
## $ uses_other_visual_aid                <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~
## $ cognitive_speed                      <dbl> 2.89484, 2.89484, 2.89484, 2.8948~
## $ cognitive_std_dev                    <dbl> 0.14733, 0.14733, 0.14733, 0.1473~
## $ cognitive_accuracy                   <dbl> 0.76300, 0.76300, 0.76300, 0.7630~
## $ cognitive_correct_responses          <dbl> 32, 32, 32, 32, 32, 32, 32, 32, 3~
## $ cognitive_errors                     <dbl> 35, 35, 35, 35, 35, 35, 35, 35, 3~
## $ cognitive_completion                 <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ cognitive_integration                <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ cognitive_battery_measure            <dbl> 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, ~
## $ panel_id                             <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ overall_case_status                  <dbl> 60, 60, 60, 60, 60, 60, 60, 60, 6~
## $ case_status_month                    <dbl> 8, 8, 8, 8, 8, 8, 8, 8, 8, 6, 6, ~
## $ case_status_year                     <dbl> 2023, 2023, 2023, 2023, 2023, 202~
## $ participant_interview_status_1       <dbl> 20, 20, 20, 20, 20, 20, 20, 20, 2~
## $ participant_interview_status_2       <dbl> 20, 20, 20, 20, 20, 20, 20, 20, 2~
## $ facility_questionnaire_status        <dbl> 24, 24, 24, 24, 24, 24, 24, 24, 2~
## $ previous_year_status                 <dbl> 60, 60, 60, 60, 60, 60, 60, 60, 6~
## $ two_years_ago_status                 <dbl> 60, 60, 60, 60, 60, 60, 60, 60, 6~
## $ Total_Imputed_Income                 <dbl> -1, -1, -1, -1, -1, -1, -1, -1, -~
## $ Income_Source_1                      <dbl> 38000, 38000, 38000, 38000, 38000~
## $ Income_Source_2                      <dbl> 38000, 38000, 38000, 38000, 38000~
## $ Income_Source_3                      <dbl> 38000, 38000, 38000, 38000, 38000~
## $ Income_Source_4                      <dbl> 38000, 38000, 38000, 38000, 38000~
## $ Income_Source_5                      <dbl> 38000, 38000, 38000, 38000, 38000~
## $ Missing_Income_Reason                <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
```

```
write_csv(nhats_merged, "~/desktop/Digital-tools_caregivers/nhats_merged.csv")
```

Work on the NHATS merged dataset to deal with missingness and make it ready to merge with HRRP dataset:

Handling Missing Values

```
# Check for missing values in each column
colSums(is.na(nhats_merged))  # Shows number of missing values per column
```

```
##                     participant_id                     residence_type
##                                  0                               2966
##                  respondent_gender                     respondent_age
##                                  0                               2659
##                  has_heart_disease                  has_hypertension
##                                  0                                  0
##                   has_lung_disease                      has_diabetes
##                                  0                                  0
##                         has_cancer                       has_stroke
##                                  0                                  0
##                       has_arthritis                      has_dementia
##                                  0                                  0
##                      has_depression                  has_other_chronic
##                                  0                                  0
##     caregiver_lives_with_respondent               caregiver_relationship
##                                59619                              58204
##                    caregiver_gender                       vision_aids
##                                56915                                  0
##                       hearing_aids                               cane
##                                  0                                  0
##                             walker                        wheelchair
##                                  0                                  0
##                  other_mobility_aid                    telehealth_use
##                                  0                                  0
##               insurance_info_online                  health_info_online
##                                  0                                  0
##                       has_computer                       has_tablet
##                                  0                                  0
##                  uses_email_or_text                uses_online_services
##                                  0                                  0
##              orders_groceries_online                      online_banking
##                                  0                                  0
##          orders_prescriptions_online                   social_media_use
##                                  0                                  0
##                  had_hosp_stay_12mo                     num_hosp_stays
##                                  0                                  0
##                       change_memory                change_decisionmaking
##                                  0                                  0
##        change_following_instructions                change_concentration
##                                  0                                  0
##                     change_language                 change_organization
##                                  0                                  0
##                 change_multitasking                  change_recognition
##                                  0                                  0
##             can_get_out_of_bed_alone          can_get_out_of_bed_with_aid
##                                  0                                  0
##                needs_help_activities             needs_help_comprehension
##                                  0                                  0
##                 needs_help_household                    back_limitation
##                                  0                                  0
##                      knee_limitation                    foot_limitation
##                                  0                                  0
##                     wrist_limitation                    walk_six_blocks
##                                  0                                  0
```

```
##                         climb_stairs                         lift_20_pounds
##                                    0                                      0
##                 had_hosp_stay_12mo_bin                   multiple_hosp_stays
##                                 3353                                      0
##                    hosp_stay_category caregiver_relationship_to_respondent
##                                50140                                  35182
##      caregiver_answered_for_respondent                          caregiver_age
##                                62136                                    258
##                caregiver_hours_per_day                caregiver_days_per_week
##                                52393                                  59827
##                          vision_status                         hearing_status
##                                    0                                      0
##                           uses_glasses                          uses_contacts
##                                    0                                      0
##                  uses_other_visual_aid                       cognitive_speed
##                                    0                                      0
##                     cognitive_std_dev                    cognitive_accuracy
##                                    0                                      0
##            cognitive_correct_responses                      cognitive_errors
##                                    0                                      0
##                  cognitive_completion                  cognitive_integration
##                                    0                                      0
##              cognitive_battery_measure                               panel_id
##                                    0                                      0
##                   overall_case_status                     case_status_month
##                                    0                                      0
##                     case_status_year        participant_interview_status_1
##                                    0                                      0
##         participant_interview_status_2         facility_questionnaire_status
##                                    0                                      0
##                  previous_year_status                  two_years_ago_status
##                                    0                                      0
##                  Total_Imputed_Income                       Income_Source_1
##                                    0                                      0
##                       Income_Source_2                       Income_Source_3
##                                    0                                      0
##                       Income_Source_4                       Income_Source_5
##                                    0                                      0
##                  Missing_Income_Reason
##                                    0
```

#Handling Missing Categorical Variables #Categorical values -> "Unknown" for missing values

```
# List of categorical columns to replace NAs with "Unknown"
categorical_cols <- c("residence_type", "caregiver_lives_with_respondent",
                      "caregiver_relationship", "caregiver_gender",
                      "caregiver_relationship_to_respondent", "caregiver_answered_for_respondent")

# Replace missing values with "Unknown"
nhats_merged[categorical_cols] <- lapply(nhats_merged[categorical_cols], function(x) ifelse(is.na(x), "
```

#Handling Missing Numeric Variables

```r
# Handling Missing Values in Age Variables
# Keep respondent_age as a character (preserving the range)
nhats_merged$respondent_age <- as.character(nhats_merged$respondent_age)

# keep missing ages them as NA
nhats_merged$respondent_age[nhats_merged$respondent_age == ""] <- NA  # Ensures empty strings are NA
```

```r
# Handling Caregiver Age (Numeric)
# Convert caregiver_age to numeric and replace NA with the median age
nhats_merged$caregiver_age <- as.numeric(nhats_merged$caregiver_age)
nhats_merged$caregiver_age[is.na(nhats_merged$caregiver_age)] <- median(nhats_merged$caregiver_age, na.
```

```r
# Handling Missing Values for Care Hours and Days
nhats_merged$caregiver_hours_per_day[is.na(nhats_merged$caregiver_hours_per_day)] <- median(nhats_merge
nhats_merged$caregiver_days_per_week[is.na(nhats_merged$caregiver_days_per_week)] <- median(nhats_merge

print(colSums(is.na(nhats_merged)))
```

```
##                       participant_id                       residence_type
##                                    0                                    0
##                     respondent_gender                       respondent_age
##                                    0                                 2659
##                    has_heart_disease                    has_hypertension
##                                    0                                    0
##                     has_lung_disease                       has_diabetes
##                                    0                                    0
##                          has_cancer                         has_stroke
##                                    0                                    0
##                        has_arthritis                       has_dementia
##                                    0                                    0
##                       has_depression                    has_other_chronic
##                                    0                                    0
##      caregiver_lives_with_respondent                caregiver_relationship
##                                    0                                    0
##                     caregiver_gender                        vision_aids
##                                    0                                    0
##                          hearing_aids                               cane
##                                    0                                    0
##                                walker                          wheelchair
##                                    0                                    0
##                   other_mobility_aid                     telehealth_use
##                                    0                                    0
##                insurance_info_online                   health_info_online
##                                    0                                    0
##                          has_computer                         has_tablet
##                                    0                                    0
##                    uses_email_or_text                  uses_online_services
##                                    0                                    0
##               orders_groceries_online                       online_banking
##                                    0                                    0
##           orders_prescriptions_online                    social_media_use
##                                    0                                    0
##                   had_hosp_stay_12mo                       num_hosp_stays
```

```
##                                   0                                   0
##                        change_memory                  change_decisionmaking
##                                   0                                   0
##          change_following_instructions                   change_concentration
##                                   0                                   0
##                       change_language                    change_organization
##                                   0                                   0
##                   change_multitasking                     change_recognition
##                                   0                                   0
##              can_get_out_of_bed_alone           can_get_out_of_bed_with_aid
##                                   0                                   0
##                 needs_help_activities              needs_help_comprehension
##                                   0                                   0
##                  needs_help_household                      back_limitation
##                                   0                                   0
##                       knee_limitation                      foot_limitation
##                                   0                                   0
##                     wrist_limitation                      walk_six_blocks
##                                   0                                   0
##                          climb_stairs                       lift_20_pounds
##                                   0                                   0
##                had_hosp_stay_12mo_bin                  multiple_hosp_stays
##                                3353                                   0
##                   hosp_stay_category caregiver_relationship_to_respondent
##                                50140                                   0
##      caregiver_answered_for_respondent                        caregiver_age
##                                   0                                   0
##                caregiver_hours_per_day                caregiver_days_per_week
##                                   0                                   0
##                          vision_status                       hearing_status
##                                   0                                   0
##                           uses_glasses                        uses_contacts
##                                   0                                   0
##                  uses_other_visual_aid                      cognitive_speed
##                                   0                                   0
##                    cognitive_std_dev                    cognitive_accuracy
##                                   0                                   0
##           cognitive_correct_responses                     cognitive_errors
##                                   0                                   0
##                  cognitive_completion                 cognitive_integration
##                                   0                                   0
##              cognitive_battery_measure                            panel_id
##                                   0                                   0
##                   overall_case_status                     case_status_month
##                                   0                                   0
##                      case_status_year         participant_interview_status_1
##                                   0                                   0
##         participant_interview_status_2          facility_questionnaire_status
##                                   0                                   0
##                  previous_year_status                  two_years_ago_status
##                                   0                                   0
##                  Total_Imputed_Income                       Income_Source_1
##                                   0                                   0
##                       Income_Source_2                       Income_Source_3
```

```
##                                   0                                   0
##                       Income_Source_4                     Income_Source_5
##                                   0                                   0
##                 Missing_Income_Reason
##                                   0
```

Handling Special Cases (Total_Imputed_Income has -1)

Replacing -1 with NA (Your code)

```r
# Replace -1 with NA in Total_Imputed_Income
nhats_merged$Total_Imputed_Income[nhats_merged$Total_Imputed_Income == -1] <- NA

head(nhats_merged)
```

```
## # A tibble: 6 x 93
##   participant_id residence_type respondent_gender respondent_age
##            <dbl> <chr>          <chr>             <chr>
## 1       10000008 Community      Male              85-89
## 2       10000008 Community      Male              85-89
## 3       10000008 Community      Male              85-89
## 4       10000008 Community      Male              85-89
## 5       10000008 Community      Male              85-89
## 6       10000008 Community      Male              85-89
## # i 89 more variables: has_heart_disease <dbl>, has_hypertension <dbl>,
## #   has_lung_disease <dbl>, has_diabetes <dbl>, has_cancer <dbl>,
## #   has_stroke <dbl>, has_arthritis <dbl>, has_dementia <dbl>,
## #   has_depression <dbl>, has_other_chronic <dbl>,
## #   caregiver_lives_with_respondent <chr>, caregiver_relationship <chr>,
## #   caregiver_gender <chr>, vision_aids <chr>, hearing_aids <chr>, cane <chr>,
## #   walker <chr>, wheelchair <chr>, other_mobility_aid <chr>, ...
```

```r
#Standardizing Binary Variables
binary_cols <- c("has_heart_disease", "has_hypertension", "has_lung_disease", "has_diabetes",
                 "has_cancer", "has_stroke", "has_arthritis", "has_dementia", "has_depression", "has_ot

nhats_merged[binary_cols] <- lapply(nhats_merged[binary_cols], function(x) {
  x <- as.numeric(x)  # Ensure numeric
  ifelse(x == 1, 1, ifelse(x == 2, 0, NA))  # 1 = Yes, 0 = No, NA for other values
})
```

```r
# Convert income source columns to numeric
income_cols <- c("Income_Source_1", "Income_Source_2", "Income_Source_3", "Income_Source_4", "Income_So
nhats_merged[income_cols] <- lapply(nhats_merged[income_cols], as.numeric)
```

```r
# Handling Duplicates ----
# Check for duplicate participant IDs
duplicate_count <- sum(duplicated(nhats_merged$participant_id))
print(paste("Number of duplicate participant IDs:", duplicate_count))
```

```
## [1] "Number of duplicate participant IDs: 54266"
```

```
# If duplicates exist, remove all but the first occurrence
nhats_merged <- nhats_merged[!duplicated(nhats_merged$participant_id), ]
```

```
# Encoding Categorical Variables
# Convert categorical variables to factors
categorical_cols <- c("residence_type", "caregiver_relationship", "caregiver_gender")
nhats_merged[categorical_cols] <- lapply(nhats_merged[categorical_cols], as.factor)
```

```
#Removing or Transforming Outliers
# Identify outliers in caregiver_hours_per_day using IQR
Q1 <- quantile(nhats_merged$caregiver_hours_per_day, 0.25, na.rm = TRUE)
Q3 <- quantile(nhats_merged$caregiver_hours_per_day, 0.75, na.rm = TRUE)
IQR_value <- Q3 - Q1
upper_bound <- Q3 + 1.5 * IQR_value
lower_bound <- Q1 - 1.5 * IQR_value

# Replace outliers with the median
median_caregiver_hours <- median(nhats_merged$caregiver_hours_per_day, na.rm = TRUE)
nhats_merged$caregiver_hours_per_day[nhats_merged$caregiver_hours_per_day > upper_bound | nhats_merged$
```

```
# Verifying Logical Consistency
# If a participant has no caregiver, ensure caregiver_hours_per_day is 0 or NA
nhats_merged$caregiver_hours_per_day[nhats_merged$caregiver_relationship == "No caregiver"] <- 0

# Final check for missing values
colSums(is.na(nhats_merged))
```

```
##                       participant_id                       residence_type
##                                    0                                    0
##                    respondent_gender                       respondent_age
##                                    0                                  305
##                   has_heart_disease                    has_hypertension
##                                  608                                 1719
##                    has_lung_disease                       has_diabetes
##                                 4327                                 4016
##                          has_cancer                         has_stroke
##                                 2175                                 2103
##                       has_arthritis                       has_dementia
##                                 1690                                  601
##                     has_depression                    has_other_chronic
##                                  900                                  596
##     caregiver_lives_with_respondent                caregiver_relationship
##                                    0                                    0
##                     caregiver_gender                        vision_aids
##                                    0                                    0
##                        hearing_aids                               cane
##                                    0                                    0
##                              walker                           wheelchair
##                                    0                                    0
##                  other_mobility_aid                       telehealth_use
##                                    0                                    0
```

```
##               insurance_info_online                   health_info_online
##                                   0                                    0
##                         has_computer                         has_tablet
##                                   0                                    0
##                   uses_email_or_text                uses_online_services
##                                   0                                    0
##              orders_groceries_online                      online_banking
##                                   0                                    0
##          orders_prescriptions_online                    social_media_use
##                                   0                                    0
##                  had_hosp_stay_12mo                     num_hosp_stays
##                                   0                                    0
##                       change_memory                change_decisionmaking
##                                   0                                    0
##        change_following_instructions                 change_concentration
##                                   0                                    0
##                      change_language                 change_organization
##                                   0                                    0
##                 change_multitasking                  change_recognition
##                                   0                                    0
##             can_get_out_of_bed_alone          can_get_out_of_bed_with_aid
##                                   0                                    0
##                needs_help_activities             needs_help_comprehension
##                                   0                                    0
##                needs_help_household                    back_limitation
##                                   0                                    0
##                     knee_limitation                    foot_limitation
##                                   0                                    0
##                    wrist_limitation                    walk_six_blocks
##                                   0                                    0
##                        climb_stairs                     lift_20_pounds
##                                   0                                    0
##               had_hosp_stay_12mo_bin                 multiple_hosp_stays
##                                 613                                    0
##                  hosp_stay_category caregiver_relationship_to_respondent
##                                6976                                    0
##    caregiver_answered_for_respondent                         caregiver_age
##                                   0                                    0
##              caregiver_hours_per_day              caregiver_days_per_week
##                                   0                                    0
##                       vision_status                     hearing_status
##                                   0                                    0
##                        uses_glasses                      uses_contacts
##                                   0                                    0
##                uses_other_visual_aid                    cognitive_speed
##                                   0                                    0
##                   cognitive_std_dev                  cognitive_accuracy
##                                   0                                    0
##          cognitive_correct_responses                    cognitive_errors
##                                   0                                    0
##                 cognitive_completion                 cognitive_integration
##                                   0                                    0
##             cognitive_battery_measure                             panel_id
##                                   0                                    0
```

```
##                  overall_case_status                    case_status_month
##                                    0                                    0
##                     case_status_year       participant_interview_status_1
##                                    0                                    0
##       participant_interview_status_2        facility_questionnaire_status
##                                    0                                    0
##                 previous_year_status                two_years_ago_status
##                                    0                                    0
##                 Total_Imputed_Income                     Income_Source_1
##                                 6146                                    0
##                      Income_Source_2                     Income_Source_3
##                                    0                                    0
##                      Income_Source_4                     Income_Source_5
##                                    0                                    0
##                Missing_Income_Reason
##                                    0
```

```
write_csv(nhats_merged, "~/desktop/Digital-tools_caregivers/Updated_nhats_merged.csv")
```

Create Multimorbidity Score

=>Counts how many chronic conditions a participant has

```
nhats_merged_cleaned <- nhats_merged %>%
  mutate(
    num_chronic_conditions = rowSums(across(
      starts_with("has_"), ~ . == 1), na.rm = TRUE),
    multimorbidity_flag = ifelse(num_chronic_conditions >= 2, 1, 0)
  )
```

Create Caregiver Burden Score

=> Total hours per week the caregiver provides care

```
nhats_merged_cleaned <- nhats_merged %>%
  mutate(
    caregiver_total_hours = caregiver_hours_per_day * caregiver_days_per_week,
    caregiver_burden_category = case_when(
      caregiver_total_hours < 10 ~ "Low",
      caregiver_total_hours < 30 ~ "Moderate",
      caregiver_total_hours >= 30 ~ "High",
      TRUE ~ NA_character_
    )
  )
```

Create Functional Impairment Index

Counts how many mobility/functional limitations a person reports

```
nhats_merged_cleaned <- nhats_merged %>%
  mutate(
    num_functional_impairments = rowSums(across(
      c(back_limitation, knee_limitation, foot_limitation, wrist_limitation), ~ . == 1), na.rm = TRUE),

    mobility_limitations = rowSums(across(
      c(walk_six_blocks, climb_stairs, lift_20_pounds), ~ . == 2), na.rm = TRUE),  # Assuming 1 = can d

    high_functional_limitation = ifelse(num_functional_impairments >= 2 | mobility_limitations >= 2, 1,
  )
```

Cognitive Change Index

=> Sums self-reported changes in cognition

```
nhats_merged_cleaned <- nhats_merged %>%
  mutate(
    cognitive_change_score = rowSums(across(
      starts_with("change_"), ~ . == 1), na.rm = TRUE),
    cognitive_change_flag = ifelse(cognitive_change_score >= 2, 1, 0)
  )

#write_csv(nhats_merged_cleaned, "~/desktop/Digital-tools_caregivers/nhats_merged_cleaned.csv")
```

```
nhats_merged_cleaned <- nhats_merged_cleaned %>%
  mutate(
    # Cleaned categorical label
    had_hosp_stay_12mo_clean = case_when(
      had_hosp_stay_12mo == 1 ~ "Yes",
      had_hosp_stay_12mo == 2 ~ "No",
      had_hosp_stay_12mo %in% c(-1, -7, -8, -9) ~ NA_character_
    ),

    # Binary version
    had_hosp_stay_12mo_bin = case_when(
      had_hosp_stay_12mo == 1 ~ 1,
      had_hosp_stay_12mo == 2 ~ 0,
      had_hosp_stay_12mo %in% c(-1, -7, -8, -9) ~ NA_real_
    ),

    # Cleaned numeric hospital count
    num_hosp_stays_clean = ifelse(num_hosp_stays < 0, NA, num_hosp_stays),

    # Create hospitalization category
    hosp_stay_category = case_when(
      is.na(num_hosp_stays_clean) ~ NA_character_,
      num_hosp_stays_clean == 0 ~ "None",
      num_hosp_stays_clean == 1 ~ "One",
      num_hosp_stays_clean >= 2 ~ "Multiple"
    )
  )
```

```r
write_csv(nhats_merged_cleaned, "~/desktop/Digital-tools_caregivers/nhats_merged_cleaned.csv")
```