Productive Engagement in the HRS

Peter Sun

January 2, 2022

Contents

1 Load Packages and Data

```
library(tidyverse)
library(haven)
library(sjlabelled)
library(ggpubr)
library(kableExtra)

# Avoid select clashes
select <- dplyr::select
recode <- dplyr::recode
summarize <- dplyr::summarize
```

2 Re-Read Project-Level Environmental Variables

```
readRenviron(".Renviron")
```

3 Import RAND Long Demographics - Wave 7 (2004) and Wave 14 (2018)

```
rand.long <- read_dta(Sys.getenv("HRS_LONG"),
                      col_select = c(hhid, pn,
                                     s7hhidpn, s14hhidpn, # w7, w14 spouse hhidpn
                                     r7agey_e, # age w7
                                     r14agey_e, # age w14
                                     ragender, # gender
                                     raracem, # race
                                     raedyrs, # education
                                     rahispan, # hispanic
                                     raeduc)) %>% # edu categorical summary
  haven::zap_formats() %>%
  sjlabelled::remove_all_labels() %>%
  as_tibble()
```

4 Calculate 2004 Numbers Using RAND Long

```
# Import 2004 RAND Fat File (File name: h04f1c.dta)
rand.04 <- read_dta(Sys.getenv("HRS_2004_FAT"),
                    col_select = c("hhidpn", "hhid", "pn",

                                   # adl helpers
                                   starts_with("jg033_"),

                                   # iadl helpers
                                   starts_with("jg055_"),

                                   # caregiving grandchildren
                                   "je060",

                                   # caregiving parental personal
                                   "jf119",

                                   # caregiving parental errands
                                   "jf139")) %>%
  haven::zap_formats() %>%
  sjlabelled::remove_all_labels() %>%
  as_tibble()

# Identify participants who had an ADL or IADL helper
# spouse_helper_sum counts the number of 2s (spouse helper) in jg033 and jg055
# has_spouse_helper is 1 if there is at least one 2s, 0 if none
rand.04b <- rand.04 %>%
  mutate(spouse_helper_sum = rowSums(
    ifelse(
      select(., starts_with("jg033") | starts_with("jg055")) == 2, 1, 0
    ), na.rm = TRUE)) %>%
  mutate(has_spouse_helper = ifelse(spouse_helper_sum >= 1, 1, 0))

# Extract participants who have a spousal ADL/IADL caregiver
# Merge their spouse PN
# Then create a dataset with hhid and pn of spouse and an indication of
# whether or not they are a spousal caregiver
spousal_caregivers <- rand.04b %>%
  filter(has_spouse_helper == 1) %>%
  select(hhid, pn, has_spouse_helper) %>%
  left_join(rand.long %>% select(hhid, pn, s7hhidpn), by = c("hhid", "pn")) %>%
  select(hhidpn = s7hhidpn, caregiver_spousal = has_spouse_helper)

# Merge the spousal_caregivers data back to the dataset
rand.04c <- rand.04b %>%
  left_join(spousal_caregivers, by = c("hhidpn"))

# Format parental/grandchildren caregivers
# Create caregiver_parental if either personal or errands == 1
rand.04d <- rand.04c %>%
  mutate(across(.cols = c(je060, jf119, jf139),
    ~recode(., `1` = 1, `5` = 0, `8` = 0, `9` = 0, .default = NA_real_))) %>%
```

```
  rename(caregiver_grandchildren = je060,
         caregiver_parental_personal = jf119,
         caregiver_parental_errands = jf139) %>%
  mutate(caregiver_parental =
    ifelse(caregiver_parental_personal == 1 | caregiver_parental_errands == 1,
           1, 0))

# Count
table(rand.04d$caregiver_spousal)
```

```
##
##    1
## 1110
```

```
table(rand.04d$caregiver_grandchildren)
```

```
##
##     0     1
## 10196  4994
```

```
table(rand.04d$caregiver_parental)
```

```
##
##    0    1
## 3915 1957
```

```
table(rand.04d$caregiver_parental_personal)
```

```
##
##    0    1
## 5205  668
```

```
table(rand.04d$caregiver_parental_errands)
```

```
##
##    0    1
## 4064 1808
```

5 Import Geography

```
region.18 <- read_dta(Sys.getenv("HRS_REGION_2018_82"),
                      col_select = c(hhid, pn, beale2013_04, beale2013_18)) %>%
  haven::zap_formats() %>%
  sjlabelled::remove_all_labels() %>%
  as_tibble()
```

6 Calculate Age 51+ Caregiving Rates and Merge/Format Additional Variables

```
rand.04.over51 <- rand.04d %>%
  left_join(rand.long, by = c("hhid", "pn")) %>%
  filter(r7agey_e >= 51) %>%
  mutate(raeduc = recode(raeduc, `1` = "Less than college",
                         `2` = "Less than college",
                         `3` = "Less than college",
                         `4` = "Less than college",
                         `5` = "College degree or more")) %>%
  mutate(rahispan = recode(rahispan, `0` = "Not Hispanic",
                           `1` = "Hispanic")) %>%
  mutate(raracem = recode(raracem, `1` = "White/Caucasian",
                          `2` = "Black/African American", `3` = "Other")) %>%
  mutate(ragender = recode(ragender, `1` = "Male", `2` = "Female")) %>%
  left_join(region.18, by = c("hhid", "pn")) %>%
  mutate(beale2013_04 = recode(beale2013_04, `1` = "Urban",
                         `2` = "Suburban", `3` = "Ex-Urban", #`9` = "No Match",
                         .default = NA_character_))

# Check beale codes
nrow(rand.04.over51)
```

```
## [1] 18954
```

```
rand.04.over51 %>%
  count(beale2013_04) # 575 missing geography
```

```
## # A tibble: 4 x 2
##   beale2013_04     n
##   <chr>        <int>
## 1 Ex-Urban      5040
## 2 Suburban      4372
## 3 Urban         8967
## 4 <NA>           575
```

```
table(rand.04.over51$beale2013_04)
```

```
##
## Ex-Urban Suburban    Urban
##     5040     4372     8967
```

```
nrow(rand.04.over51)
```

```
## [1] 18954
```

```
table(rand.04.over51$caregiver_spousal)
```

```
##
##    1
## 1052
```

```
table(rand.04.over51$caregiver_grandchildren)
```

```
##
##    0    1
```

```
## 9844 4761
```

```
table(rand.04.over51$caregiver_parental)
```

```
##
##    0    1
## 3202 1703
```

```
table(rand.04.over51$caregiver_parental_personal)
```

```
##
##    0    1
## 4313  593
```

```
table(rand.04.over51$caregiver_parental_errands)
```

```
##
##    0    1
## 3334 1571
```

7 Calculate 2018 Numbers Using RAND Long

Repeat above steps for 2018

```
# Import 2018 RAND Fat File (File name: h04f1c.dta)
rand.18 <- read_dta(Sys.getenv("HRS_2018_FAT"),
                    col_select = c("hhidpn", "hhid", "pn",

                                   # adl helpers
                                   starts_with("qg033_"),

                                   # iadl helpers
                                   starts_with("qg055_"),

                                   # caregiving grandchildren
                                   "qe060",

                                   # caregiving parental personal
                                   "qf119",

                                   # caregiving parental errands
                                   "qf139")) %>%
  haven::zap_formats() %>%
  sjlabelled::remove_all_labels() %>%
  as_tibble()

# Identify participants who had an ADL or IADL helper
# spouse_helper_sum counts the number of 2s (spouse helper) in jg033 and jg055
# has_spouse_helper is 1 if there is at least one 2s, 0 if none
rand.18b <- rand.18 %>%
  mutate(spouse_helper_sum = rowSums(
    ifelse(
      select(., starts_with("qg033") | starts_with("qg055")) == 2, 1, 0
    ), na.rm = TRUE)) %>%
  mutate(has_spouse_helper = ifelse(spouse_helper_sum >= 1, 1, 0))

# Extract participants who have a spousal ADL/IADL caregiver
# Merge their spouse PN
# Then create a dataset with hhid and pn of spouse and an indication of
# whether or not they are a spousal caregiver
spousal_caregivers.18 <- rand.18b %>%
  filter(has_spouse_helper == 1) %>%
  select(hhid, pn, has_spouse_helper) %>%
  left_join(rand.long %>% select(hhid, pn, s14hhidpn),
            by = c("hhid", "pn")) %>%
  select(hhidpn = s14hhidpn, caregiver_spousal = has_spouse_helper)

# Merge the spousal_caregivers data back to the dataset
rand.18c <- rand.18b %>%
  left_join(spousal_caregivers.18, by = c("hhidpn"))

# Format parental/grandchildren caregivers
# Create caregiver_parental if either personal or errands == 1
rand.18d <- rand.18c %>%
```

```
  mutate(across(.cols = c(qe060, qf119, qf139),
    ~recode(., `1` = 1, `5` = 0, `8` = 0, `9` = 0, .default = NA_real_))) %>%
  rename(caregiver_grandchildren = qe060,
         caregiver_parental_personal = qf119,
         caregiver_parental_errands = qf139) %>%
  mutate(caregiver_parental =
    ifelse(caregiver_parental_personal == 1 | caregiver_parental_errands == 1,
           1, 0))

# Count
table(rand.18d$caregiver_spousal)
```

```
##
##   1
## 936
```

```
table(rand.18d$caregiver_grandchildren)
```

```
##
##    0    1
## 8786 3493
```

```
table(rand.18d$caregiver_parental)
```

```
##
##    0    1
## 3564 1938
```

```
table(rand.18d$caregiver_parental_personal)
```

```
##
##    0    1
## 4767  734
```

```
table(rand.18d$caregiver_parental_errands)
```

```
##
##    0    1
## 3712 1791
```

8 Calculate Age 51+ Caregiving Rates and Merge/Format Additional Variables

```
rand.18.over51 <- rand.18d %>%
  left_join(rand.long, by = c("hhid", "pn")) %>%
  filter(r14agey_e >= 51) %>%
  mutate(raeduc = recode(raeduc, `1` = "Less than college",
                         `2` = "Less than college",
                         `3` = "Less than college",
                         `4` = "Less than college",
                         `5` = "College degree or more")) %>%
  mutate(rahispan = recode(rahispan, `0` = "Not Hispanic",
                           `1` = "Hispanic")) %>%
  mutate(raracem = recode(raracem, `1` = "White/Caucasian",
                          `2` = "Black/African American", `3` = "Other")) %>%
  mutate(ragender = recode(ragender, `1` = "Male", `2` = "Female")) %>%
  left_join(region.18, by = c("hhid", "pn")) %>%
  mutate(beale2013_18 = recode(beale2013_18, `1` = "Urban",
                               `2` = "Suburban",
                               `3` = "Ex-Urban", #`9` = "No Match",
                               .default = NA_character_))

# Check beale codes
nrow(rand.18.over51)
```

```
## [1] 16584
```

```
rand.18.over51 %>%
  count(beale2013_18) # 384 missing geography
```

```
## # A tibble: 4 x 2
##   beale2013_18     n
##   <chr>        <int>
## 1 Ex-Urban      3758
## 2 Suburban      3515
## 3 Urban         8927
## 4 <NA>           384
```

```
table(rand.18.over51$beale2013_18)
```

```
##
## Ex-Urban Suburban    Urban
##     3758     3515     8927
```

```
nrow(rand.18.over51)
```

```
## [1] 16584
```

```
table(rand.18.over51$caregiver_spousal)
```

```
##
##   1
## 891
```

```
table(rand.18.over51$caregiver_grandchildren)
```

```
##
```

```
##    0    1
## 8613 3393
```

```
table(rand.18.over51$caregiver_parental)
```

```
##
##    0    1
## 3203 1839
```

```
table(rand.18.over51$caregiver_parental_personal)
```

```
##
##    0    1
## 4337  704
```

```
table(rand.18.over51$caregiver_parental_errands)
```

```
##
##    0    1
## 3346 1697
```

9 Merge and Simplify 2004 and 2018 Data

```
r4 <- rand.04.over51 %>%
  select(cs = caregiver_spousal, cp = caregiver_parental,
         cg = caregiver_grandchildren,
         Sex = ragender, Hispanic = rahispan,
         Race = raracem, Education = raeduc,
         Geography = beale2013_04) %>%
  mutate(Year = "2004")
r18 <- rand.18.over51 %>%
  select(cs = caregiver_spousal, cp = caregiver_parental,
         cg = caregiver_grandchildren,
         Sex = ragender, Hispanic = rahispan,
         Race = raracem, Education = raeduc,
         Geography = beale2013_18) %>%
  mutate(Year = "2018")
r.418 <- bind_rows(r4, r18) %>%
  mutate(Caregiver_Sum = rowSums(select(., cs:cg), na.rm = T),
         Caregiver = ifelse(Caregiver_Sum >= 1, 1, 0),
         Caregiver_Multi = ifelse(Caregiver_Sum > 1, 1, 0)) %>%
  select(-Caregiver_Sum)
```

10 Total Sample Sizes (Age 51+)

```
nrow(r4) # 2004
```

```
## [1] 18954
```

```
nrow(r18) # 2018
```

```
## [1] 16584
```

11 Contingency Tables

```
# Function for contingency tables
get_cont <- function(var) {
  r.418 %>%
    count(Year, Geography, {{ var }}) %>%
    filter(!is.na(Geography)) %>%
    #filter(!is.na({{ var }})) %>%
    group_by(Year, Geography) %>%
    mutate(pct = scales::percent(n / sum(n), accuracy = 0.1)) %>%
    kbl(booktabs = T, linesep = "", digits = 2) %>%
    kable_styling(position = "center") %>%
    kable_styling(latex_options = c("striped", "hold_position"))
}

# Get total estimates without cross-tabulation
get_cont(Caregiver)
```

Year	Geography	Caregiver	n	pct
2004	Ex-Urban	0	3165	62.8%
2004	Ex-Urban	1	1875	37.2%
2004	Suburban	0	2851	65.2%
2004	Suburban	1	1521	34.8%
2004	Urban	0	5904	65.8%
2004	Urban	1	3063	34.2%
2018	Ex-Urban	0	2542	67.6%
2018	Ex-Urban	1	1216	32.4%
2018	Suburban	0	2391	68.0%
2018	Suburban	1	1124	32.0%
2018	Urban	0	6023	67.5%
2018	Urban	1	2904	32.5%

```
# Get total estimates without cross-tabulation (MULTI)
get_cont(Caregiver_Multi)
```

Year	Geography	Caregiver_Multi	n	pct
2004	Ex-Urban	0	4795	95.1%
2004	Ex-Urban	1	245	4.9%
2004	Suburban	0	4177	95.5%
2004	Suburban	1	195	4.5%
2004	Urban	0	8595	95.9%
2004	Urban	1	372	4.1%
2018	Ex-Urban	0	3594	95.6%
2018	Ex-Urban	1	164	4.4%
2018	Suburban	0	3373	96.0%
2018	Suburban	1	142	4.0%
2018	Urban	0	8515	95.4%
2018	Urban	1	412	4.6%

```
# Types of caregivers
get_cont(cs) # spousal
```

Year	Geography	cs	n	pct
2004	Ex-Urban	1	329	6.5%
2004	Ex-Urban		4711	93.5%
2004	Suburban	1	249	5.7%
2004	Suburban		4123	94.3%
2004	Urban	1	435	4.9%
2004	Urban		8532	95.1%
2018	Ex-Urban	1	237	6.3%
2018	Ex-Urban		3521	93.7%
2018	Suburban	1	231	6.6%
2018	Suburban		3284	93.4%
2018	Urban	1	406	4.5%
2018	Urban		8521	95.5%

```
get_cont(cp) # parental
```

Year	Geography	cp	n	pct
2004	Ex-Urban	0	822	16.3%
2004	Ex-Urban	1	468	9.3%
2004	Ex-Urban		3750	74.4%
2004	Suburban	0	724	16.6%
2004	Suburban	1	393	9.0%
2004	Suburban		3255	74.5%
2004	Urban	0	1572	17.5%
2004	Urban	1	789	8.8%
2004	Urban		6606	73.7%
2018	Ex-Urban	0	637	17.0%
2018	Ex-Urban	1	363	9.7%
2018	Ex-Urban		2758	73.4%
2018	Suburban	0	603	17.2%
2018	Suburban	1	326	9.3%
2018	Suburban		2586	73.6%
2018	Urban	0	1885	21.1%
2018	Urban	1	1112	12.5%
2018	Urban		5930	66.4%

```
get_cont(cg) # grandchildren
```

Year	Geography	cg	n	pct
2004	Ex-Urban	0	2750	54.6%
2004	Ex-Urban	1	1331	26.4%
2004	Ex-Urban		959	19.0%
2004	Suburban	0	2385	54.6%
2004	Suburban	1	1081	24.7%
2004	Suburban		906	20.7%
2004	Urban	0	4374	48.8%
2004	Urban	1	2217	24.7%
2004	Urban		2376	26.5%
2018	Ex-Urban	0	2174	57.8%
2018	Ex-Urban	1	787	20.9%
2018	Ex-Urban		797	21.2%
2018	Suburban	0	1996	56.8%
2018	Suburban	1	713	20.3%
2018	Suburban		806	22.9%
2018	Urban	0	4228	47.4%
2018	Urban	1	1816	20.3%
2018	Urban		2883	32.3%

12 Plots

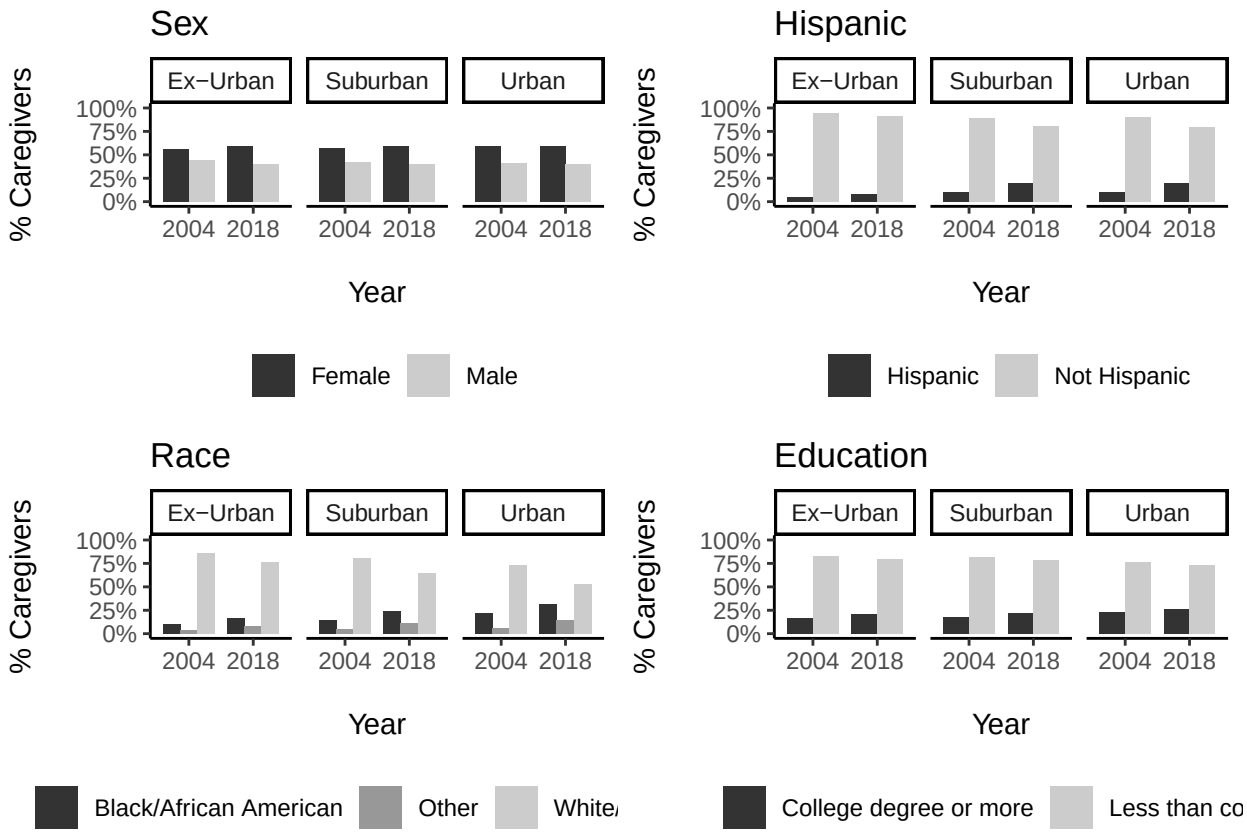
```
get_count <- function(iv, caregiver_type) {
  r.418 %>%
    filter({{ caregiver_type }} == 1) %>%
    count(Year, Geography, {{ iv }}) %>%
    group_by(Year, Geography) %>%
    mutate(pct = n / sum(n)) %>%
    ungroup(Year, Geography) %>%
    filter(!is.na(Geography)) %>%
    filter(!is.na({{ iv }}))
}

create_plot <- function(df, iv, mytitle, upr_limit = 1) {
  kab <- df %>%
    mutate(pct = scales::percent(pct, accuracy = .1)) %>%
    kbl(booktabs = T, linesep = "", digits = 1) %>%
    kable_styling(position = "center") %>%
    kable_styling(latex_options = c("striped", "hold_position"))
  p <- ggplot(df, aes(x = Year, y = pct, fill = {{ iv }})) +
      geom_bar(stat = "identity", width = 0.8, position = "dodge") +
      facet_wrap(~Geography, scales = "fixed") +
      theme_classic() +
      scale_fill_grey() +
      scale_y_continuous(labels = scales::percent, limits = c(0, upr_limit)) +
      labs(y = "% Caregivers\n", x = "\nYear", title = mytitle, fill = "") +
      theme(legend.position = "bottom")
  return(list(kab = kab, p = p))
}
```

13 Caregiving Plot

```
p1 <- create_plot(get_count(Sex, Caregiver), Sex, "Sex")
p2 <- create_plot(get_count(Hispanic, Caregiver), Hispanic, "Hispanic")
p3 <- create_plot(get_count(Race, Caregiver), Race, "Race")
p4 <- create_plot(get_count(Education, Caregiver), Education, "Education")

p5 <- ggpubr::ggarrange(p1$p, p2$p, p3$p, p4$p)
p5
```

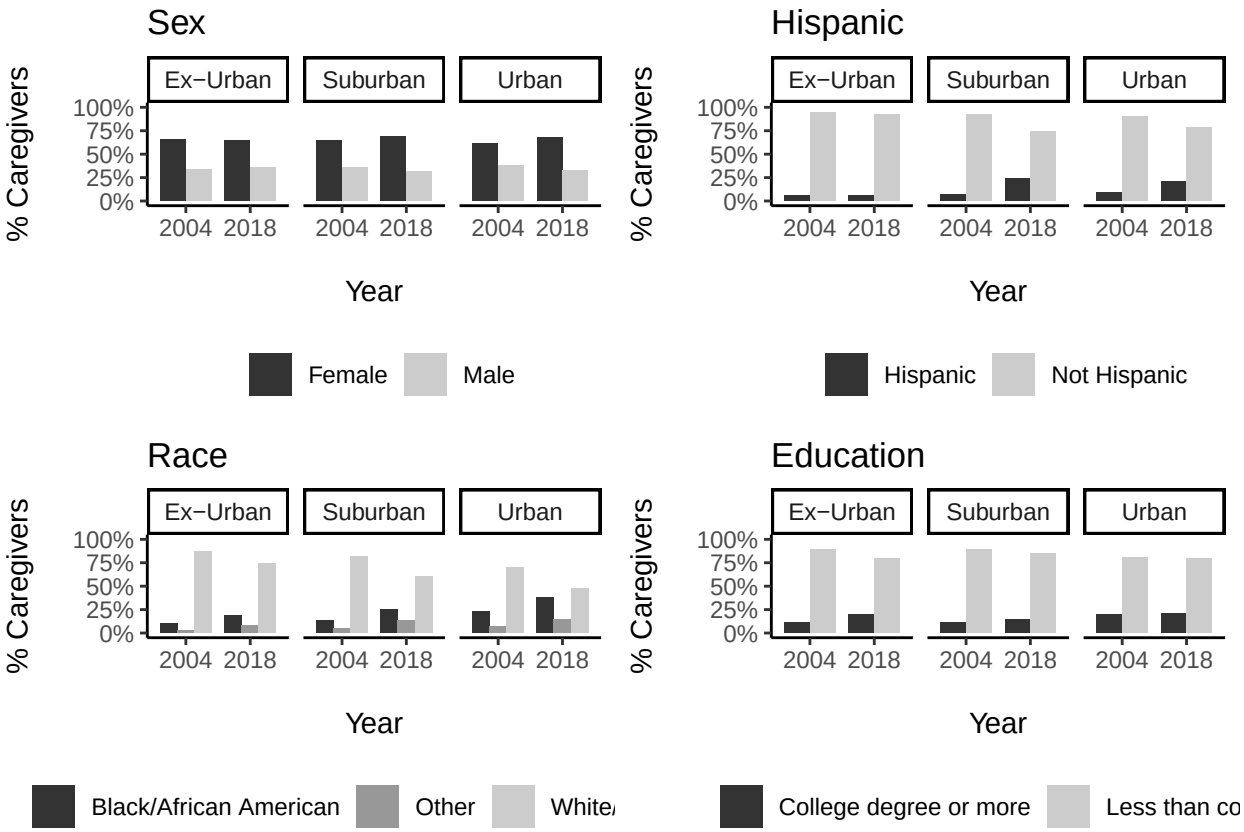



```
ggsave(paste0("figures/", "caregiving.emf"), width = 12, height = 7, plot = p5)
```

14 Multiple Caregiving Roles Plot

```
# Multiple Caregiver Roles
p6 <- create_plot(get_count(Sex, Caregiver_Multi), Sex, "Sex")
p7 <- create_plot(get_count(Hispanic, Caregiver_Multi), Hispanic, "Hispanic")
p8 <- create_plot(get_count(Race, Caregiver_Multi), Race, "Race")
p9 <- create_plot(get_count(Education, Caregiver_Multi), Education, "Education")

p10 <- ggpubr::ggarrange(p6$p, p7$p, p8$p, p9$p)
p10
```



15 Grandchild/Parent/Spousal Caregiving Plots

```
# switch between cs/cp/cg
p11 <- create_plot(get_count(Sex, cg), Sex, "Sex")
p12 <- create_plot(get_count(Hispanic, cg), Hispanic, "Hispanic")
p13 <- create_plot(get_count(Race, cg), Race, "Race")
p14 <- create_plot(get_count(Education, cg), Education, "Education")
p15 <- ggpubr::ggarrange(p11$p, p12$p, p13$p, p14$p)
p15
```