Problem Set 1: PSID - Labor Outcomes

Tate Mason

Part 1: Overall Trends:

```
library(AER)
library(haven)
library(tidyverse)
library(psych)
library(patchwork)
library(broom)
```

```
df <- read_dta("~/SchoolWork/Y2S1/Macro/Data/PSID/PSID.dta")
```

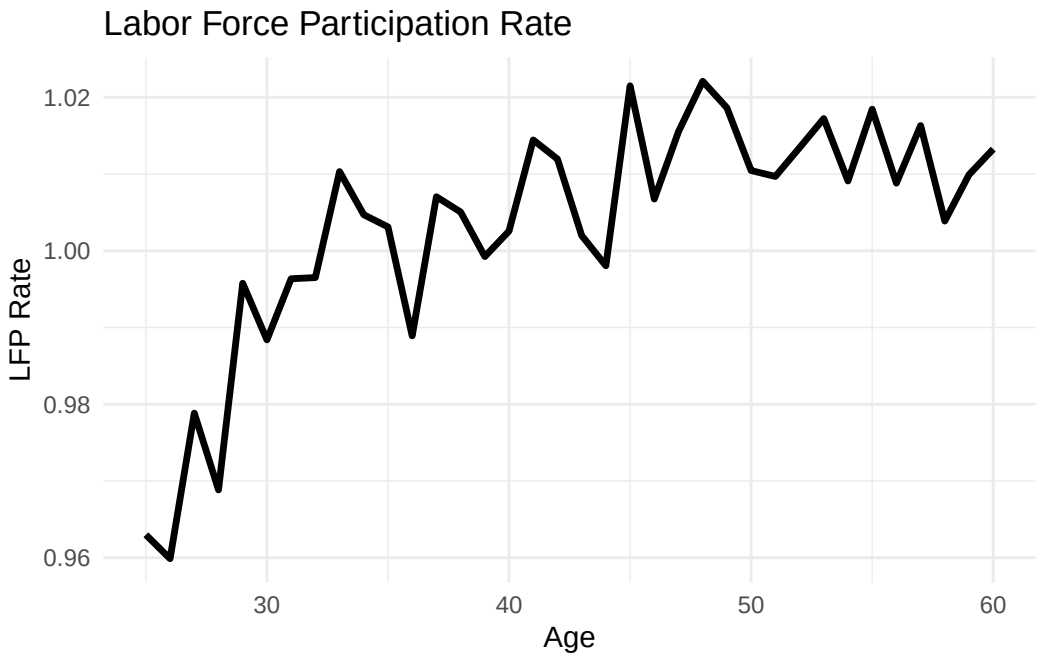
Age profiles

```
wm <- function(x,w) weighted.mean(x, w, na.rm=TRUE)

age_profile_fe <- function(data, y, base_age = 25, w_col = weight) {
  yq <- rlang::enquo(y)
  wq <- rlang::enquo(w_col)

  d <- data %>%
    filter(!is.na(!!yq), !is.na(!!wq)) %>%
    mutate(
      age  = as.integer(age),
      year = as.integer(year)
    )

  if (nrow(d) == 0) return(tibble(age = integer(), y_m = numeric()))
```

```r
  # build formula: response ~ factor(age) + factor(year)
  resp <- rlang::as_name(yq)
  fml  <- stats::as.formula(paste(resp, "~ factor(age) + factor(year)"))

  # evaluate weights as a numeric vector
  wv <- as.numeric(rlang::eval_tidy(wq, d))

  reg <- stats::lm(fml, data = d, weights = wv)

  af <- broom::tidy(reg) %>%
    dplyr::filter(grepl("^factor\\(age\\)", term)) %>%
    dplyr::mutate(age = as.integer(gsub("factor\\(age\\)", "", term))) %>%
    tidyr::complete(age = base_age:60, fill = list(estimate = 0)) %>%
    dplyr::arrange(age) %>%
    dplyr::transmute(age, y_m = estimate)

  mu <- mean(d[[resp]], na.rm = TRUE)
  dplyr::mutate(af, y_m = y_m + mu)
}

var_prof_year_net <- function(data, y, w_col = weight) {
  yq <- rlang::enquo(y)
  wq <- rlang::enquo(w_col)

  d <- data %>%
    filter(!is.na(!!yq), !!yq > 0, !is.na(!!wq)) %>%
    mutate(
      age  = as.integer(age),
      year = as.integer(year),
      ly   = log(!!yq),
      w    = as.numeric(!!wq)   # <- carry weights as a column
    )

  if (nrow(d) == 0) return(tibble(age = integer(), v = numeric()))

  # Remove year effects with weighted regression
  reg <- stats::lm(ly ~ factor(year), data = d, weights = d$w)
  d$res <- stats::resid(reg)

  # Weighted variance within each age using that age group's weights
  d %>%
    group_by(age) %>%
```

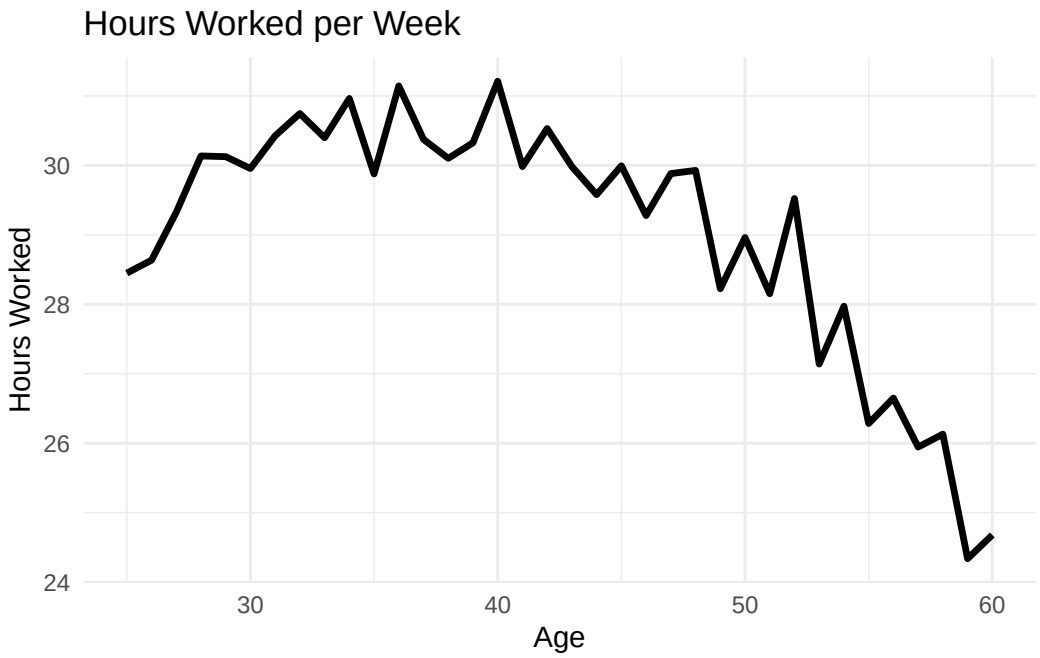
```
    summarise(
      v = {
        wg <- w
        rg <- res
        mu <- weighted.mean(rg, wg, na.rm = TRUE)
        sum(wg * (rg - mu)^2, na.rm = TRUE) / sum(wg, na.rm = TRUE)
      },
      .groups = "drop"
    )
}

lfp_age_fe <- age_profile_fe(psid_m, lfp)
wage_age_fe <- age_profile_fe(psid_m, wage_real)
hr_age_fe <- age_profile_fe(psid_m, hr_worked)
inc_age_fe <- age_profile_fe(psid_m, inc_real)
var_wage_age <- var_prof_year_net(psid_m, wage_real)
var_hr_age <- var_prof_year_net(psid_m, hr_worked)

ggplot(lfp_age_fe, aes(age, y_m)) +
  geom_line(size = 1.2) +
  labs(
    title = "Labor Force Participation Rate",
    x = "Age",
    y = "LFP Rate"
  ) +
  theme_minimal()
```

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```

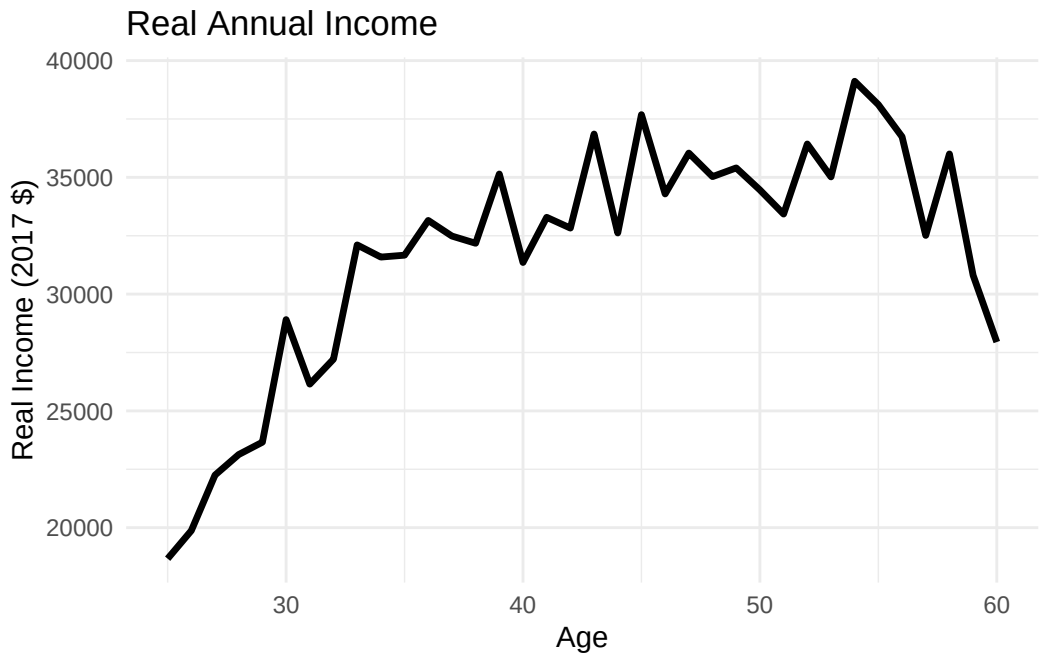



```
ggplot(wage_age_fe, aes(age, y_m)) +
  geom_line(size = 1.2) +
  labs(
    title = "Real Wage",
    x = "Age",
    y = "Real Wage (2017 $)"
  ) +
  theme_minimal()
```

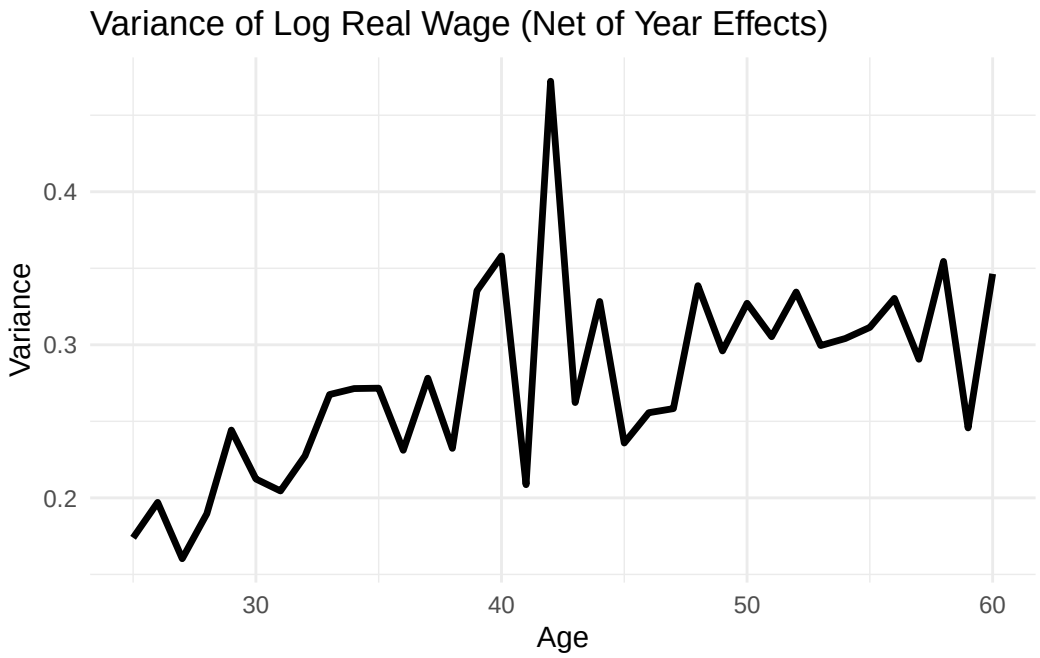

```
ggplot(hr_age_fe, aes(age, y_m)) +
  geom_line(size = 1.2) +
  labs(
    title = "Hours Worked per Week",
    x = "Age",
    y = "Hours Worked"
  ) +
  theme_minimal()
```



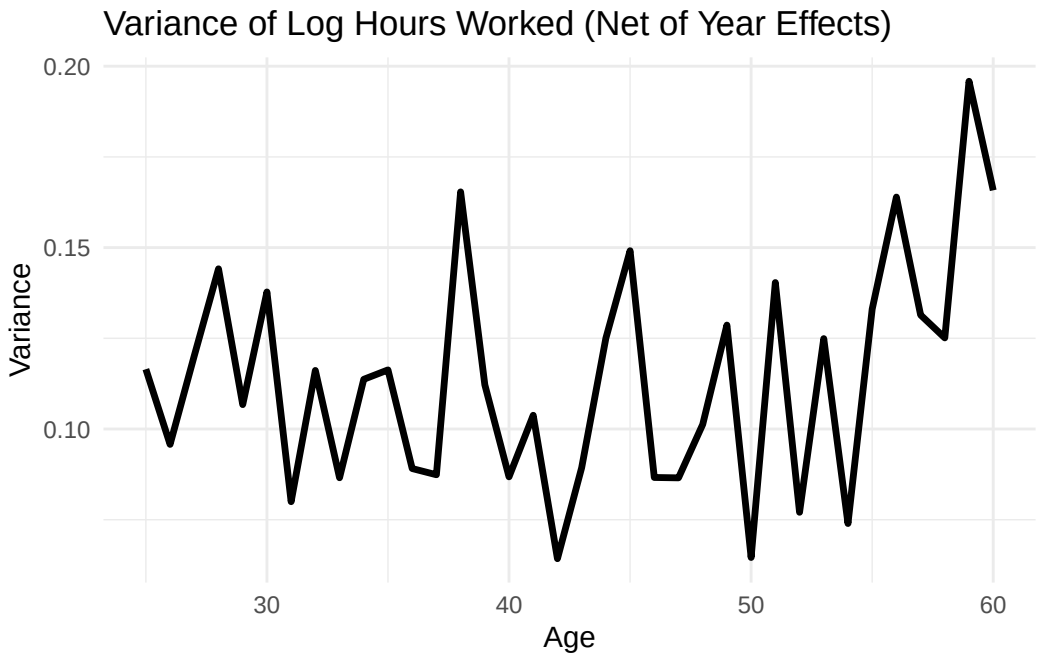

```
ggplot(inc_age_fe, aes(age, y_m)) +
  geom_line(size = 1.2) +
  labs(
    title = "Real Annual Income",
    x = "Age",
    y = "Real Income (2017 $)"
  ) +
  theme_minimal()
```




```
ggplot(var_wage_age, aes(age, v)) +
  geom_line(size = 1.2) +
  labs(
    title = "Variance of Log Real Wage (Net of Year Effects)",
    x = "Age",
    y = "Variance"
  ) +
  theme_minimal()
```




```
ggplot(var_hr_age, aes(age, v)) +
  geom_line(size = 1.2) +
  labs(
    title = "Variance of Log Hours Worked (Net of Year Effects)",
    x = "Age",
    y = "Variance"
  ) +
  theme_minimal()
```

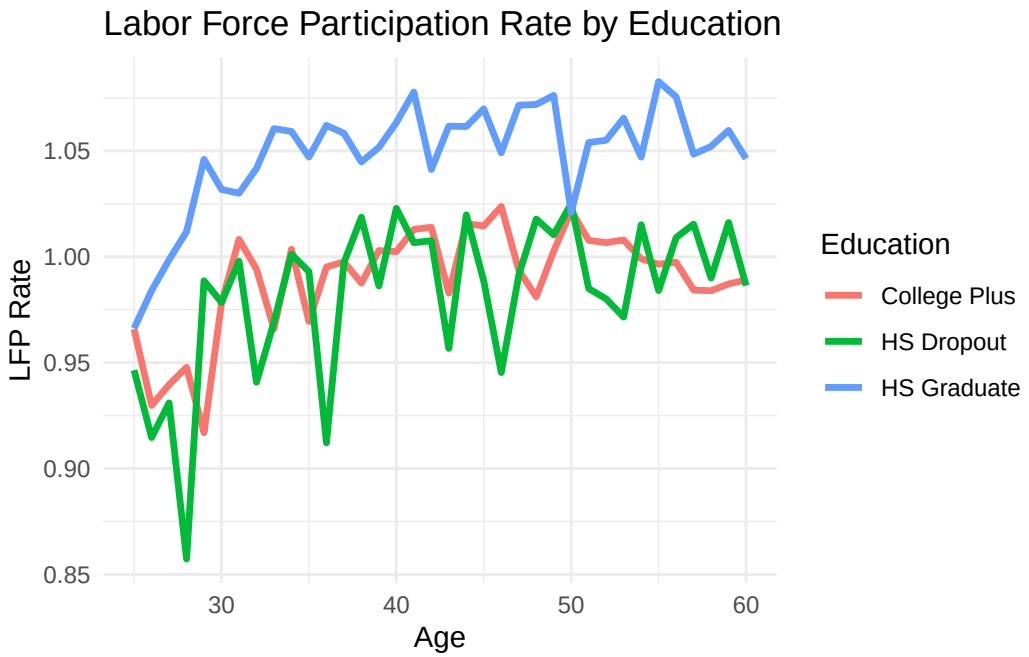


Education groups

```
wage_age_fe_educ <- psid_m %>%
  filter(!is.na(educ_group)) %>%
  group_by(educ_group) %>%
  group_modify(~ age_profile_fe(.x, wage_real)) %>%
  ungroup()

hour_age_fe_educ <- psid_m %>%
  filter(!is.na(educ_group)) %>%
  group_by(educ_group) %>%
  group_modify(~ age_profile_fe(.x, hr_worked)) %>%
  ungroup()

lfp_age_fe_educ <- psid_m %>%
  filter(!is.na(educ_group)) %>%
  group_by(educ_group) %>%
  group_modify(~ age_profile_fe(.x, lfp)) %>%
  ungroup()

var_wage_age_educ <- psid_m %>%
  filter(!is.na(educ_group)) %>%
```

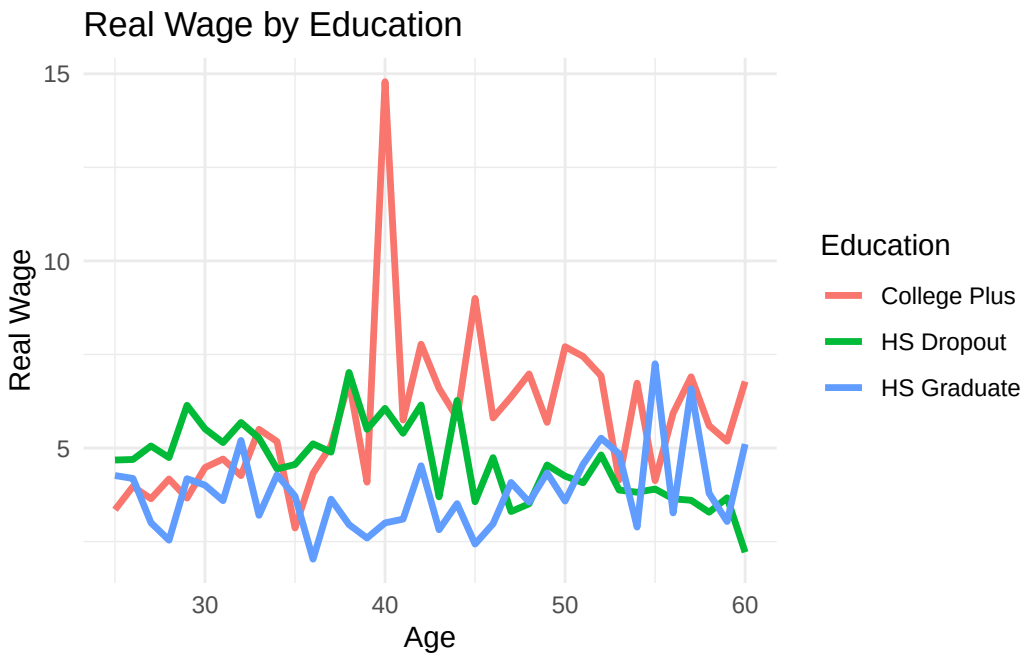
```
  group_by(educ_group) %>%
  group_modify(~ var_prof_year_net(.x, wage_real)) %>%
  ungroup()

var_hr_age_educ <- psid_m %>%
  filter(!is.na(educ_group)) %>%
  group_by(educ_group) %>%
  group_modify(~ var_prof_year_net(.x, hr_worked)) %>%
  ungroup()
```

```
ggplot(lfp_age_fe_educ, aes(age, y_m, color = educ_group)) +
  geom_line(size = 1.2) +
  labs(
    title = "Labor Force Participation Rate by Education",
    x = "Age",
    y = "LFP Rate",
    color = "Education"
  ) +
  theme_minimal()
```

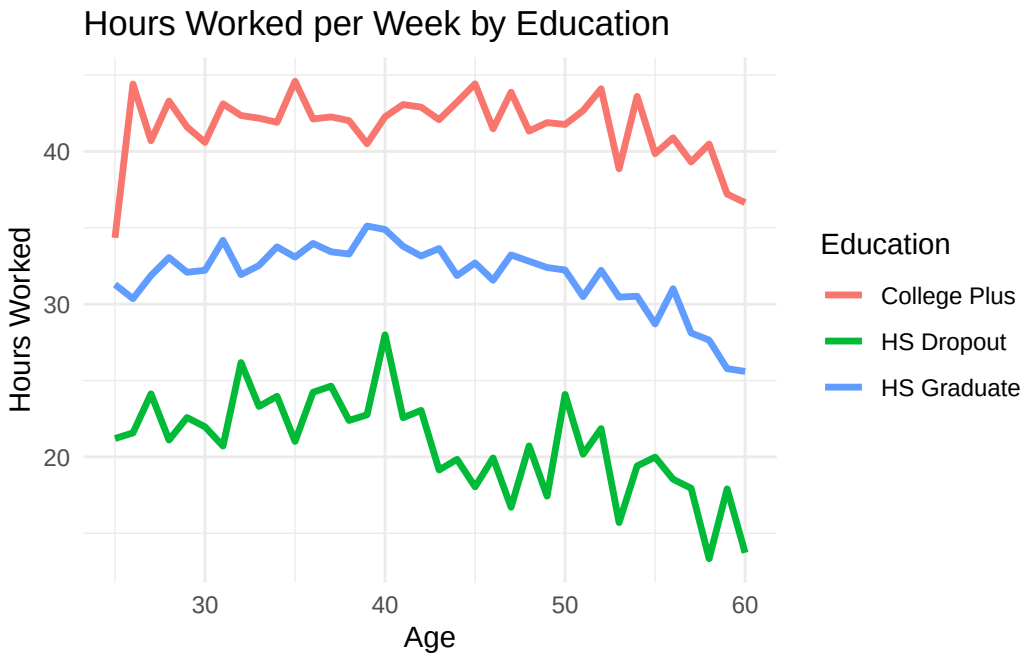



```
ggplot(wage_age_fe_educ, aes(age, y_m, color = educ_group)) +
  geom_line(size = 1.2) +
  labs(
```

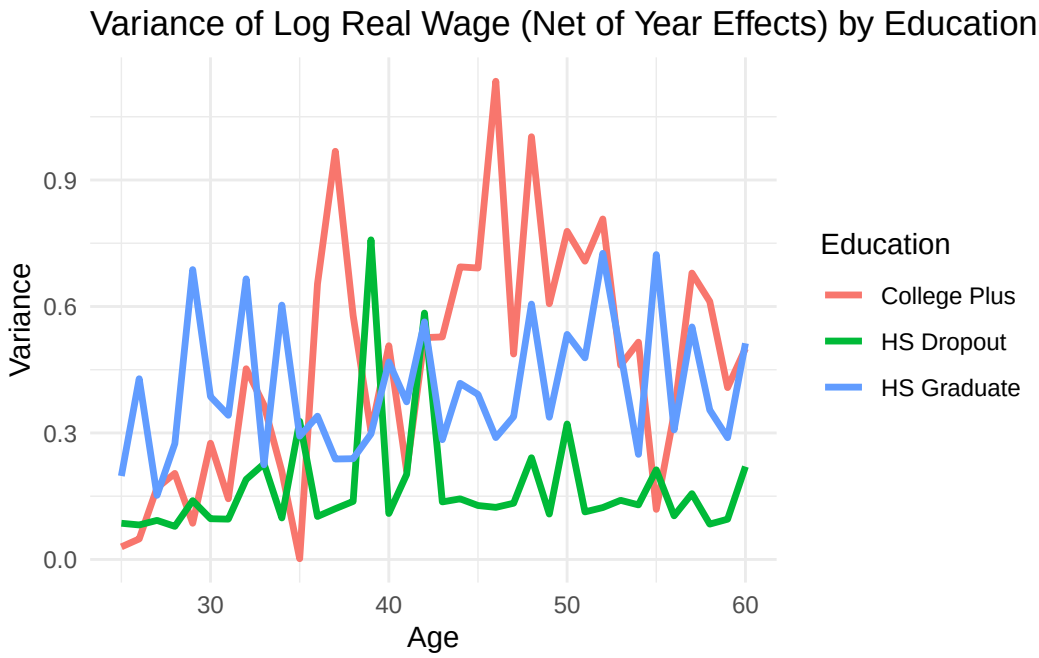
```
    title = "Real Wage by Education",
    x = "Age",
    y = "Real Wage",
    color = "Education"
  ) +
  theme_minimal()
```



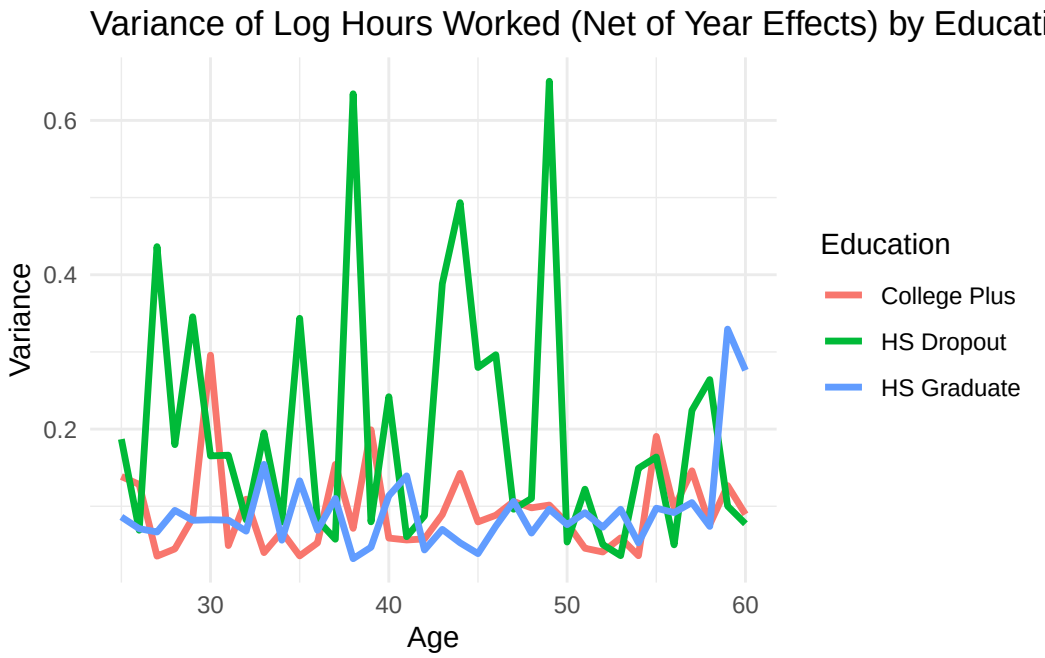

```
ggplot(hour_age_fe_educ, aes(age, y_m, color = educ_group)) +
  geom_line(size = 1.2) +
  labs(
    title = "Hours Worked per Week by Education",
    x = "Age",
    y = "Hours Worked",
    color = "Education"
  ) +
  theme_minimal()
```




```
ggplot(var_wage_age_educ, aes(age, v, color = educ_group)) +
  geom_line(size = 1.2) +
  labs(
    title = "Variance of Log Real Wage (Net of Year Effects) by Education",
    x = "Age",
    y = "Variance",
    color = "Education"
  ) +
  theme_minimal()
```



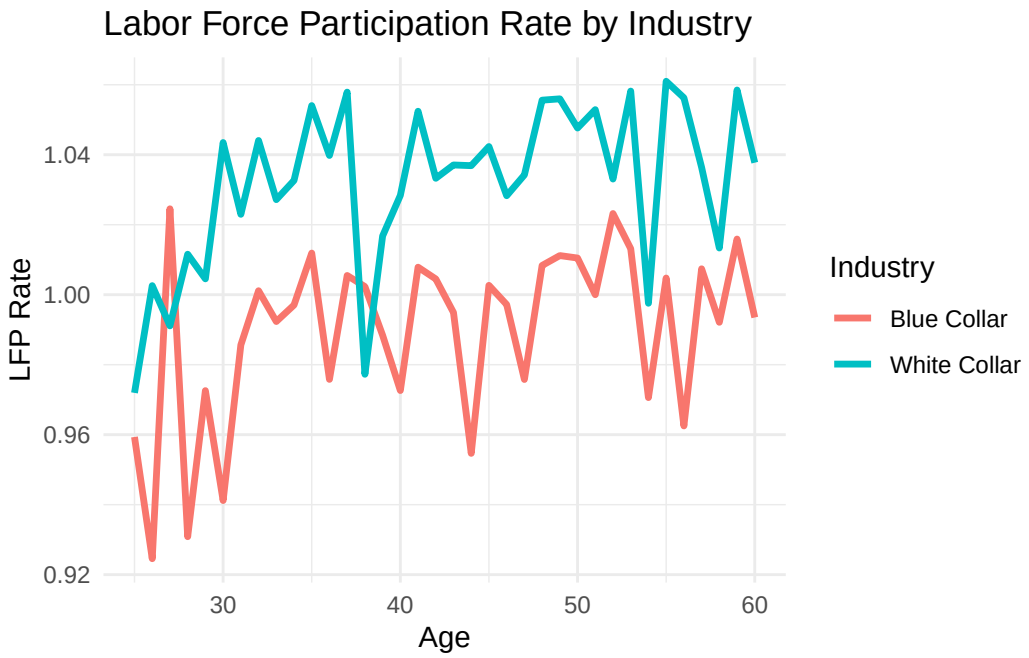

```
ggplot(var_hr_age_educ, aes(age, v, color = educ_group)) +
  geom_line(size = 1.2) +
  labs(
    title = "Variance of Log Hours Worked (Net of Year Effects) by Education",
    x = "Age",
    y = "Variance",
    color = "Education"
  ) +
  theme_minimal()
```



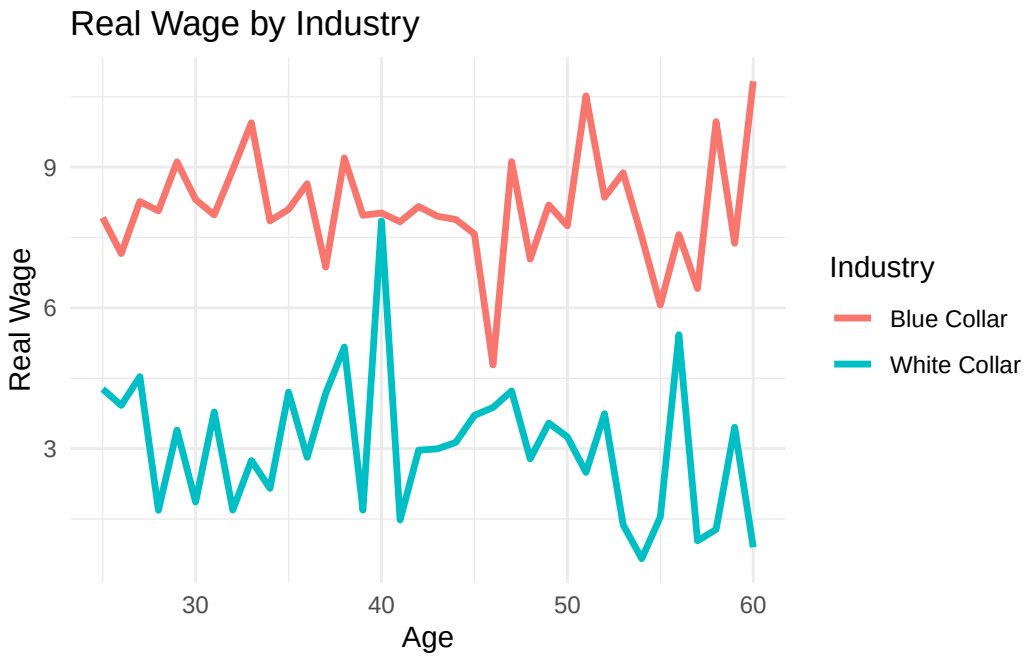
Industry groups

```
wage_age_fe_ind <- psid_m %>%
  filter(!is.na(ind_group)) %>%
  group_by(ind_group) %>%
  group_modify(~ age_profile_fe(.x, wage_real)) %>%
  ungroup()

hour_age_fe_ind <- psid_m %>%
  filter(!is.na(ind_group)) %>%
  group_by(ind_group) %>%
  group_modify(~ age_profile_fe(.x, hr_worked)) %>%
  ungroup()

inc_age_fe_ind <- psid_m %>%
  filter(!is.na(ind_group)) %>%
  group_by(ind_group) %>%
  group_modify(~ age_profile_fe(.x, inc_real)) %>%
  ungroup()

lfp_age_fe_ind <- psid_m %>%
  filter(!is.na(ind_group)) %>%
```

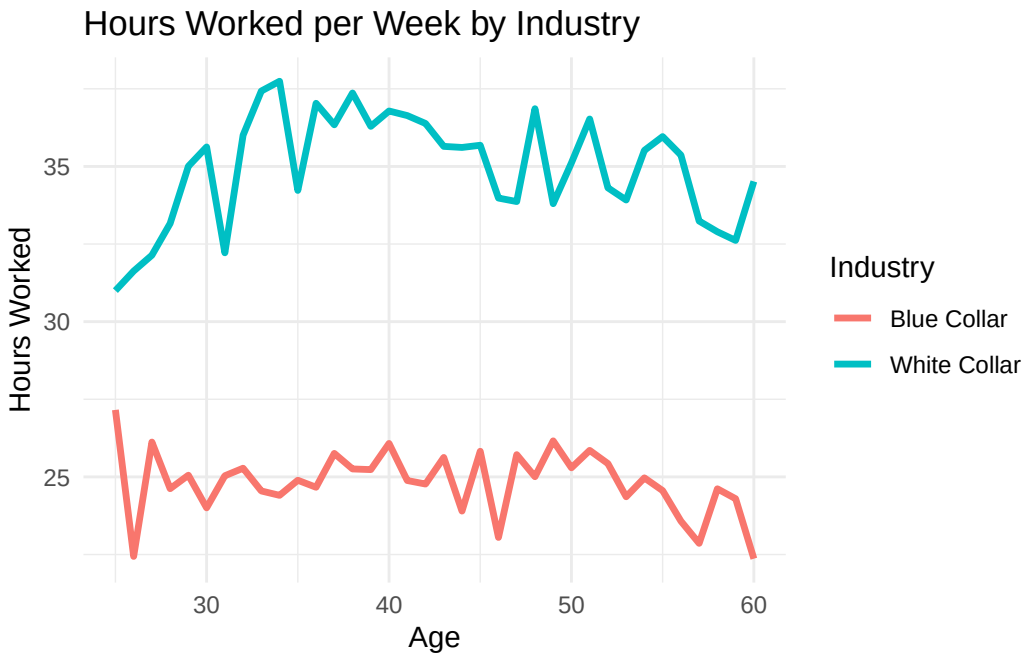
```r
  group_by(ind_group) %>%
  group_modify(~ age_profile_fe(.x, lfp)) %>%
  ungroup()

var_wage_age_ind <- psid_m %>%
  filter(!is.na(ind_group)) %>%
  group_by(ind_group) %>%
  group_modify(~ var_prof_year_net(.x, wage_real)) %>%
  ungroup()

var_hr_age_ind <- psid_m %>%
  filter(!is.na(ind_group)) %>%
  group_by(ind_group) %>%
  group_modify(~ var_prof_year_net(.x, hr_worked)) %>%
  ungroup()

ggplot(lfp_age_fe_ind, aes(age, y_m, color = ind_group)) +
  geom_line(size = 1.2) +
  labs(
    title = "Labor Force Participation Rate by Industry",
    x = "Age",
    y = "LFP Rate",
    color = "Industry"
  ) +
  theme_minimal()
```



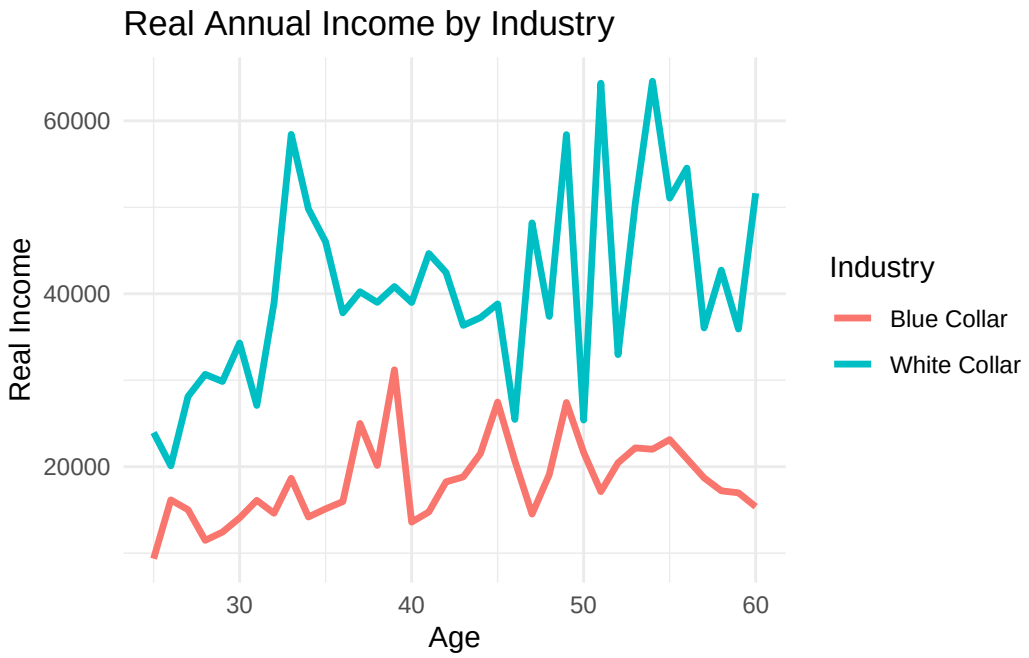

```
ggplot(wage_age_fe_ind, aes(age, y_m, color = ind_group)) +
  geom_line(size = 1.2) +
  labs(
    title = "Real Wage by Industry",
    x = "Age",
    y = "Real Wage",
    color = "Industry"
  ) +
  theme_minimal()
```



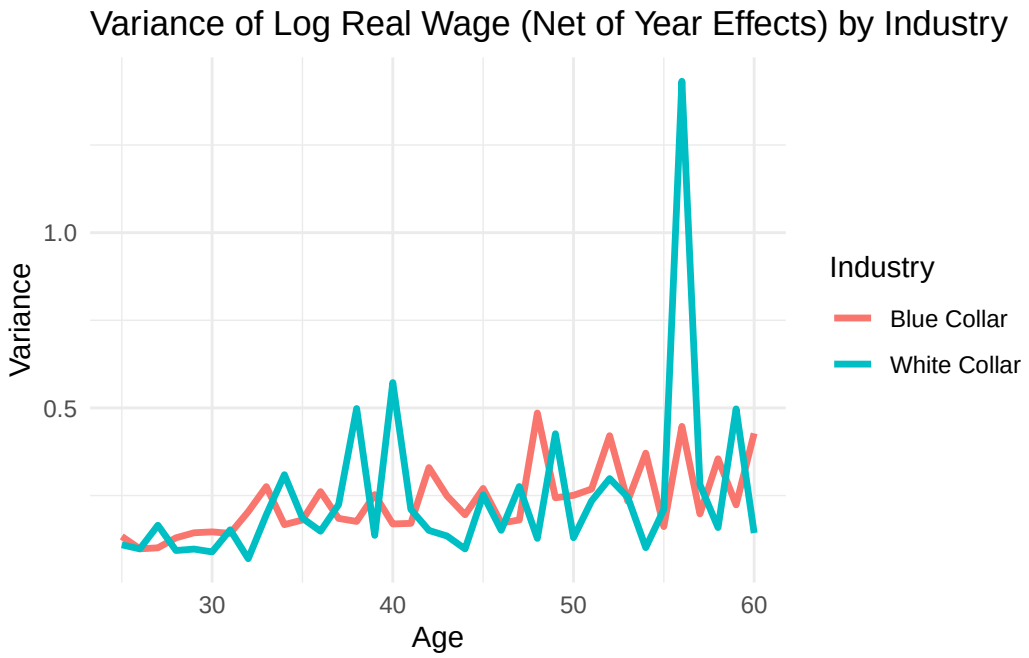

```
ggplot(hour_age_fe_ind, aes(age, y_m, color = ind_group)) +
  geom_line(size = 1.2) +
  labs(
    title = "Hours Worked per Week by Industry",
    x = "Age",
    y = "Hours Worked",
    color = "Industry"
  ) +
  theme_minimal()
```



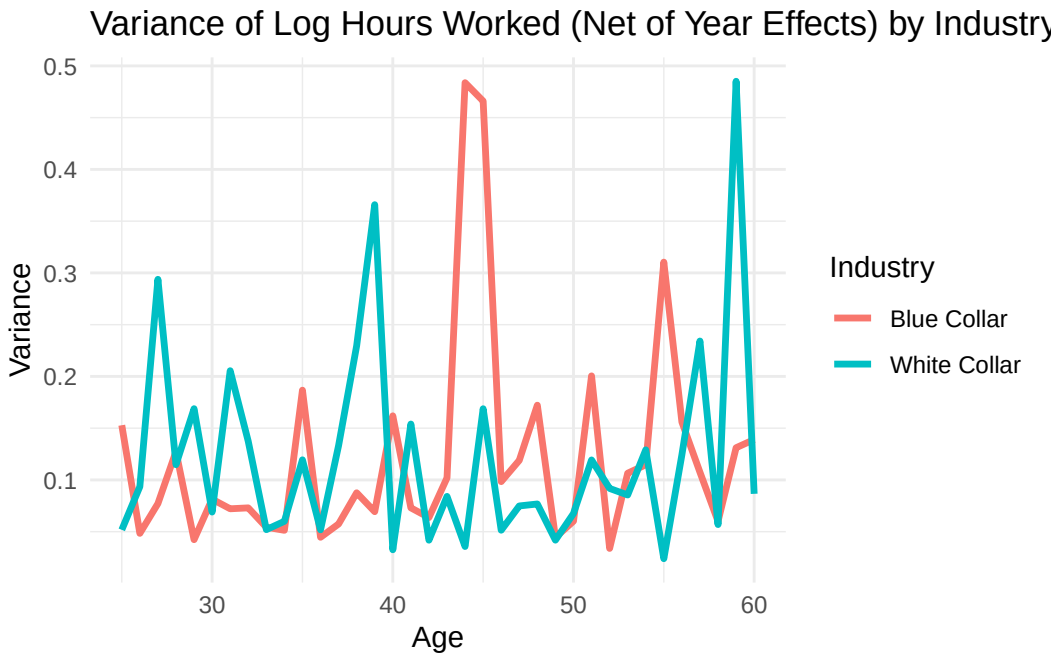

```
ggplot(inc_age_fe_ind, aes(age, y_m, color = ind_group)) +
  geom_line(size = 1.2) +
  labs(
    title = "Real Annual Income by Industry",
    x = "Age",
    y = "Real Income",
    color = "Industry"
  ) +
  theme_minimal()
```




```
ggplot(var_wage_age_ind, aes(age, v, color = ind_group)) +
  geom_line(size = 1.2) +
  labs(
    title = "Variance of Log Real Wage (Net of Year Effects) by Industry",
    x = "Age",
    y = "Variance",
    color = "Industry"
  ) +
  theme_minimal()
```



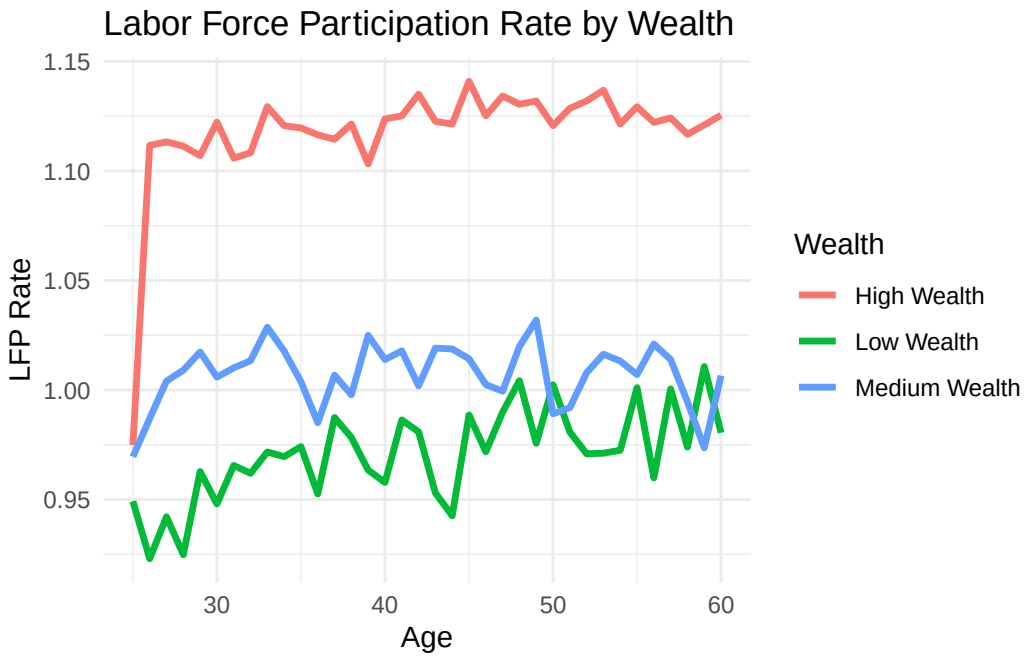

```
ggplot(var_hr_age_ind, aes(age, v, color = ind_group)) +
  geom_line(size = 1.2) +
  labs(
    title = "Variance of Log Hours Worked (Net of Year Effects) by Industry",
    x = "Age",
    y = "Variance",
    color = "Industry"
  ) +
  theme_minimal()
```



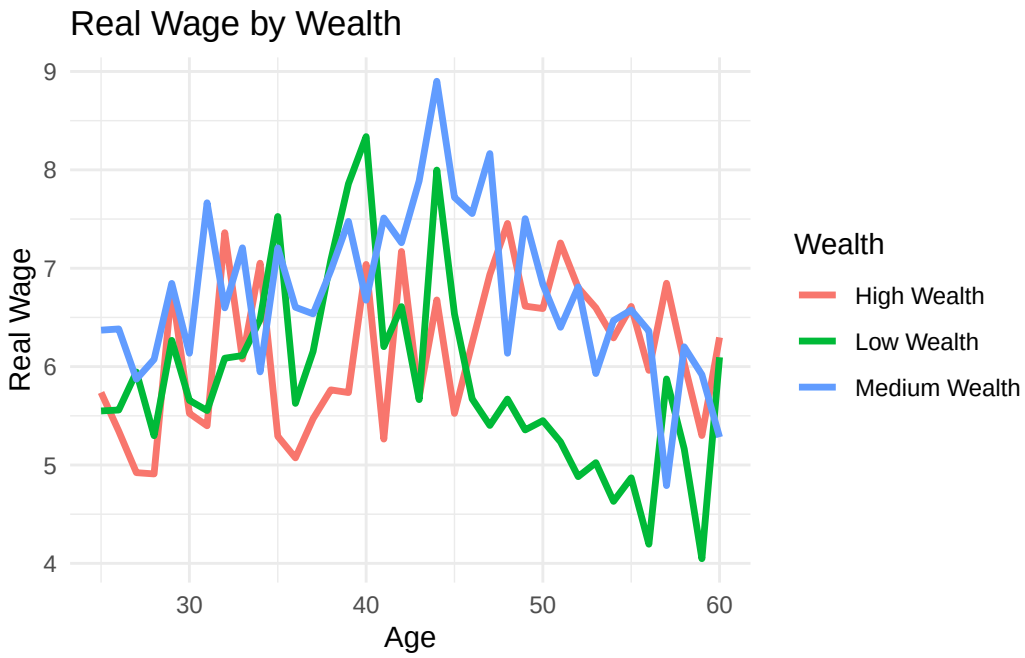
Wealth groups

```
wage_age_fe_wealth <- psid_m %>%
  filter(!is.na(wealth_group)) %>%
  group_by(wealth_group) %>%
  group_modify(~ age_profile_fe(.x, wage_real)) %>%
  ungroup()

hour_age_fe_wealth <- psid_m %>%
  filter(!is.na(wealth_group)) %>%
  group_by(wealth_group) %>%
  group_modify(~ age_profile_fe(.x, hr_worked)) %>%
  ungroup()

inc_age_fe_wealth <- psid_m %>%
  filter(!is.na(wealth_group)) %>%
  group_by(wealth_group) %>%
  group_modify(~ age_profile_fe(.x, inc_real)) %>%
  ungroup()

lfp_age_fe_wealth <- psid_m %>%
  filter(!is.na(wealth_group)) %>%
```

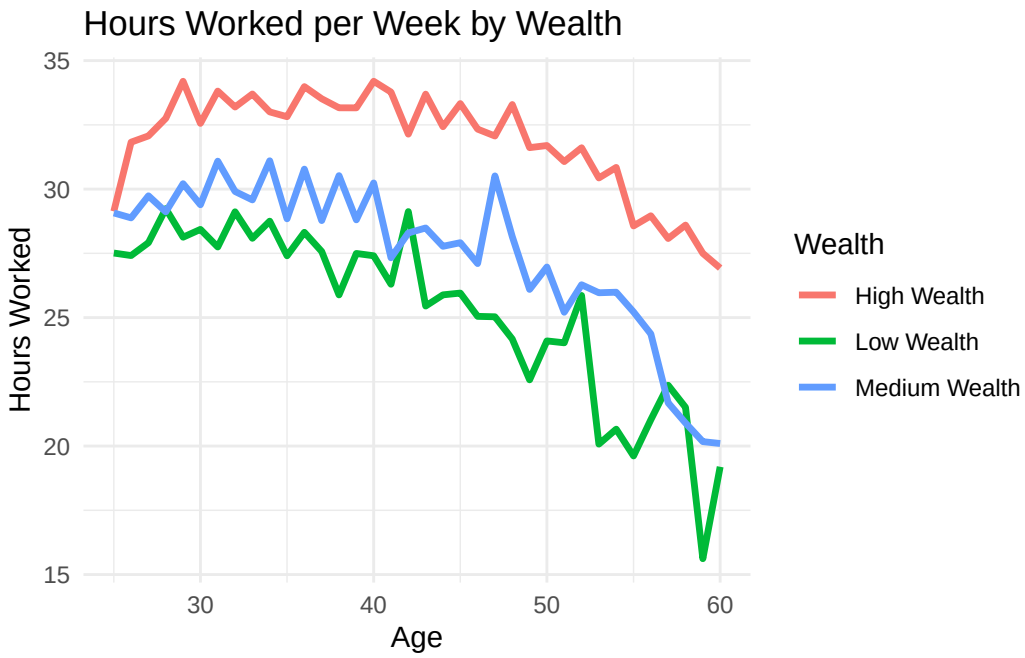
```r
  group_by(wealth_group) %>%
  group_modify(~ age_profile_fe(.x, lfp)) %>%
  ungroup()

var_wage_age_wealth <- psid_m %>%
  filter(!is.na(wealth_group)) %>%
  group_by(wealth_group) %>%
  group_modify(~ var_prof_year_net(.x, wage_real)) %>%
  ungroup()

var_hr_age_wealth <- psid_m %>%
  filter(!is.na(wealth_group)) %>%
  group_by(wealth_group) %>%
  group_modify(~ var_prof_year_net(.x, hr_worked)) %>%
  ungroup()

ggplot(lfp_age_fe_wealth, aes(age, y_m, color = wealth_group)) +
  geom_line(size = 1.2) +
  labs(
    title = "Labor Force Participation Rate by Wealth",
    x = "Age",
    y = "LFP Rate",
    color = "Wealth"
  ) +
  theme_minimal()
```



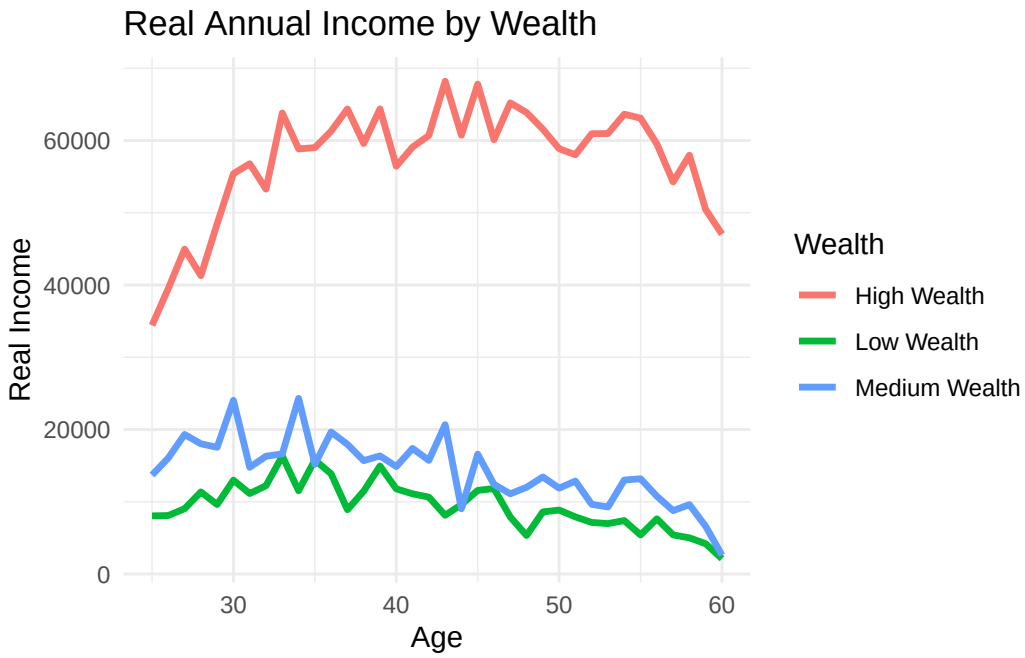

```
ggplot(wage_age_fe_wealth, aes(age, y_m, color = wealth_group)) +
  geom_line(size = 1.2) +
  labs(
    title = "Real Wage by Wealth",
    x = "Age",
    y = "Real Wage",
    color = "Wealth"
  ) +
  theme_minimal()
```



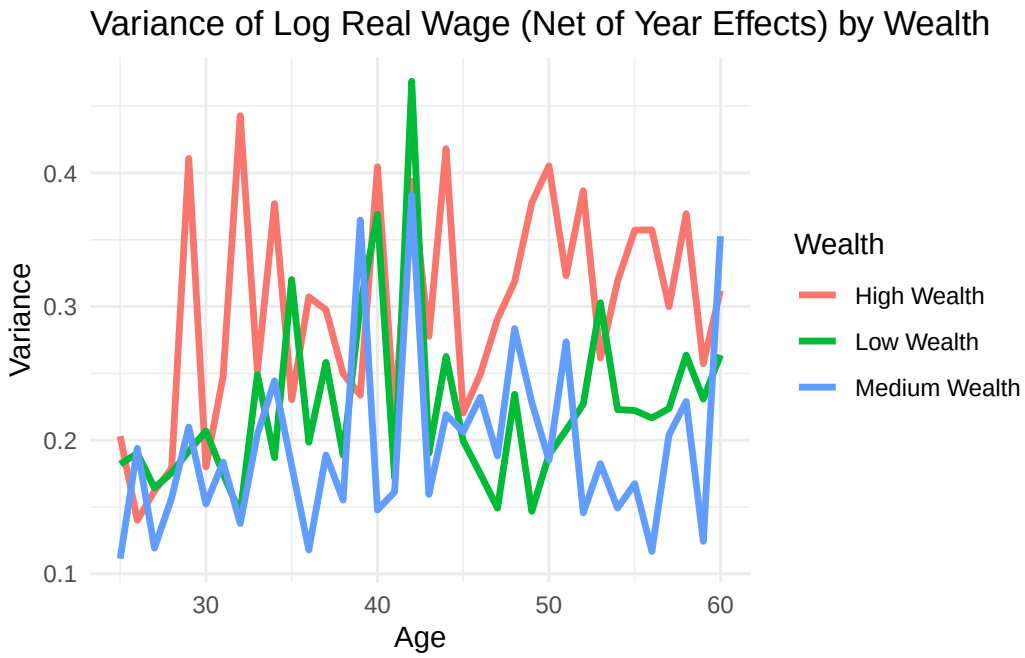

```
ggplot(hour_age_fe_wealth, aes(age, y_m, color = wealth_group)) +
  geom_line(size = 1.2) +
  labs(
    title = "Hours Worked per Week by Wealth",
    x = "Age",
    y = "Hours Worked",
    color = "Wealth"
  ) +
  theme_minimal()
```



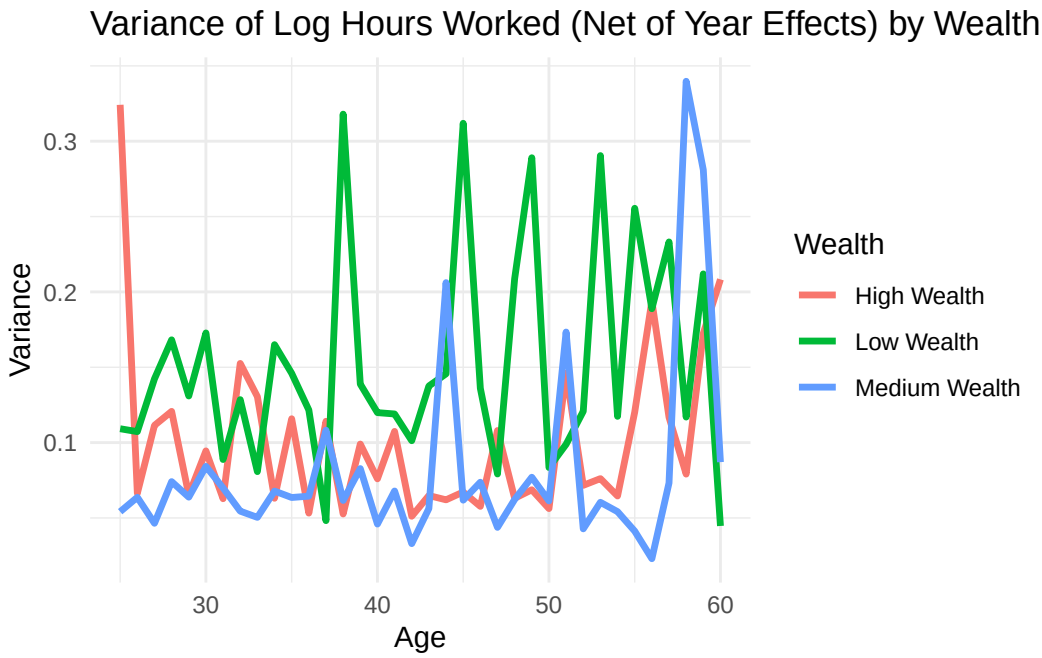

```
ggplot(inc_age_fe_wealth, aes(age, y_m, color = wealth_group)) +
  geom_line(size = 1.2) +
  labs(
    title = "Real Annual Income by Wealth",
    x = "Age",
    y = "Real Income",
    color = "Wealth"
  ) +
  theme_minimal()
```




```
ggplot(var_wage_age_wealth, aes(age, v, color = wealth_group)) +
  geom_line(size = 1.2) +
  labs(
    title = "Variance of Log Real Wage (Net of Year Effects) by Wealth",
    x = "Age",
    y = "Variance",
    color = "Wealth"
  ) +
  theme_minimal()
```




```
ggplot(var_hr_age_wealth, aes(age, v, color = wealth_group)) +
  geom_line(size = 1.2) +
  labs(
    title = "Variance of Log Hours Worked (Net of Year Effects) by Wealth",
    x = "Age",
    y = "Variance",
    color = "Wealth"
  ) +
  theme_minimal()
```


Variance of Log Hours Worked (Net of Year Effects) by Wealth
Variance
0.3
0.2
0.1
30
40
50
60
Age
Wealth
High Wealth
Low Wealth
Medium Wealth