

Problem Set 1: PSID - Labor Outcomes

Tate Mason

Part 1: Overall Trends:

```
library(AER)
library(haven)
library(tidyverse)
```

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

```
psid_99_clean <- df %>%
  transmute(
    year = 1999,
    age = ER13010,
    sex = ER13011,
    lfp = ER13601,
    wage = ER13224,
    hr_worked = ER13363,
    cpi_ratio = 0.644,
    educ_HS = ER15937,
    educ_coll = ER15953,
    ind = ER13216,
    wealth = S417,
    inc = ER13218
  )
psid_01_clean <- df %>%
  transmute(
    year = 2001,
    age = ER17013,
    sex = ER17014,
    lfp = ER17657,
    wage = ER17235,
```

```

    hr_worked = ER17393,
    cpi_ratio = 0.685,
    educ_HS = ER19998,
    educ_coll = ER20014,
    ind = ER17227,
    wealth = S517,
    inc = ER17229
  )
psid_03_clean <- df %>%
  transmute(
    year = 2003,
    age = ER21017,
    sex = ER21018,
    lfp = ER21339,
    wage = ER21159,
    hr_worked = ER21356,
    cpi_ratio = 0.711,
    educ_HS = ER23435,
    educ_coll = ER23451,
    ind = ER21146,
    wealth = S617,
    inc = ER21153
  )
psid_05_clean <- df %>%
  transmute(
    year = 2005,
    age = ER25017,
    sex = ER25018,
    lfp = ER25328,
    wage = ER25148,
    hr_worked = ER25345,
    cpi_ratio = 0.755,
    educ_HS = ER27402,
    educ_coll = ER27418,
    ind = ER25128,
    wealth = S717,
    inc = ER25142
  )
psid_07_clean <- df %>%
  transmute(
    year = 2007,
    age = ER36017,

```

```

    sex = ER36018,
    lfp = ER36333,
    wage = ER36153,
    hr_worked = ER36350,
    cpi_ratio = 0.802,
    educ_HS = ER40574,
    educ_coll = ER40590,
    ind = ER36133,
    wealth = S817,
    inc = ER36147
  )
psid_09_clean <- df %>%
  transmute(
    year = 2009,
    age = ER42017,
    sex = ER42018,
    lfp = ER42360,
    wage = ER42188,
    hr_worked = ER42148,
    cpi_ratio = 0.829,
    educ_HS = ER46552,
    educ_coll = ER46568,
    ind = ER42168,
    wealth = ER46970,
    inc = ER42182
  )
psid_11_clean <- df %>%
  transmute(
    year = 2011,
    age = ER47317,
    sex = ER47318,
    lfp = ER47673,
    wage = ER47501,
    hr_worked = ER47456,
    cpi_ratio = 0.867,
    educ_HS = ER51913,
    educ_coll = ER51929,
    ind = ER47480,
    wealth = ER52394,
    inc = ER47495
  )
psid_13_clean <- df %>%

```

```

transmute(
  year = 2013,
  age = ER53017,
  sex = ER53018,
  lfp = ER53636,
  wage = ER53201,
  hr_worked = ER53156,
  cpi_ratio = 0.901,
  educ_HS = ER57669,
  educ_coll = ER57685,
  ind = ER53180,
  wealth = ER58211,
  inc = ER53195
)
psid_15_clean <- df %>%
  transmute(
    year = 2015,
    age = ER60017,
    sex = ER60018,
    lfp = ER60388,
    wage = ER60216,
    hr_worked = ER60171,
    cpi_ratio = 0.916,
    educ_HS = ER64821,
    educ_coll = ER64837,
    ind = ER60195,
    wealth = ER65408,
    inc = ER60210
  )
psid_17_clean <- df %>%
  transmute(
    year = 2017,
    age = ER66017,
    sex = ER66018,
    lfp = ER66666,
    wage = ER66492,
    hr_worked = ER66172,
    cpi_ratio = 0.947,
    educ_HS = ER70755,
    educ_coll = ER70909,
    ind = ER66196,
    wealth = ER71485,

```

```

    inc = ER66211
  )
psid_clean <- bind_rows(
  psid_99_clean,
  psid_01_clean,
  psid_03_clean,
  psid_05_clean,
  psid_07_clean,
  psid_09_clean,
  psid_11_clean,
  psid_13_clean,
  psid_15_clean,
  psid_17_clean
)

psid_clean <- psid_clean %>%
  group_by(year) %>%
  mutate(
    educ_group = case_when(
      educ_HS == 3 ~ "HS Dropout",
      educ_HS == 1 & educ_coll == 2 ~ "HS Graduate",
      educ_coll >= 2 & educ_coll <= 7 ~ "College Plus",
      TRUE ~ NA_character_
    ),
    ind_group = case_when(
      ind %in% c(range(17:28), range(47:57), range(67:77), range(107:398)) ~ "Blue Collar",
      ind %in% c(range(407:479), range(507:698), range(707:718), range(727:759),
        range(769:798), range(807:809), range(828:897), range(907:937)) ~ "White Collar",
      TRUE ~ NA_character_
    ),
    wealth_group = case_when(
      wealth > -99999999 & wealth < 25000 ~ "Low Wealth",
      wealth >= 25000 & wealth < 100000 ~ "Medium Wealth",
      wealth >= 100000 & wealth <= 99999998 ~ "High Wealth",
      TRUE ~ NA_character_
    ),
    wage_real = wage * cpi_ratio,
    inc_real = inc * cpi_ratio,
    log_wage = if_else(wage > 0, log(wage_real), NA_real_),
    log_hr_worked = if_else(hr_worked > 0, log(hr_worked), NA_real_),
    lfp = case_when(
      lfp == 0 ~ 1,

```

```

    lfp >= 1 & lfp <= 52 ~ 0,
    TRUE ~ NA_real_
  )
)

lfp_all <- psid_clean %>%
  filter(
    age >= 25 & age <= 60,
    sex == 1,
  ) %>%
  group_by(age) %>%
  summarise(
    lfp_rate = mean(lfp, na.rm = TRUE),
    wage_rate = mean(wage_real, na.rm = TRUE),
    hr_worked = mean(hr_worked, na.rm = TRUE),
    inc = mean(inc, na.rm = TRUE),
    var_log_wage = var(log_wage, na.rm = TRUE),
    var_log_hr_worked = var(log_hr_worked, na.rm = TRUE),
    n = n()
  )

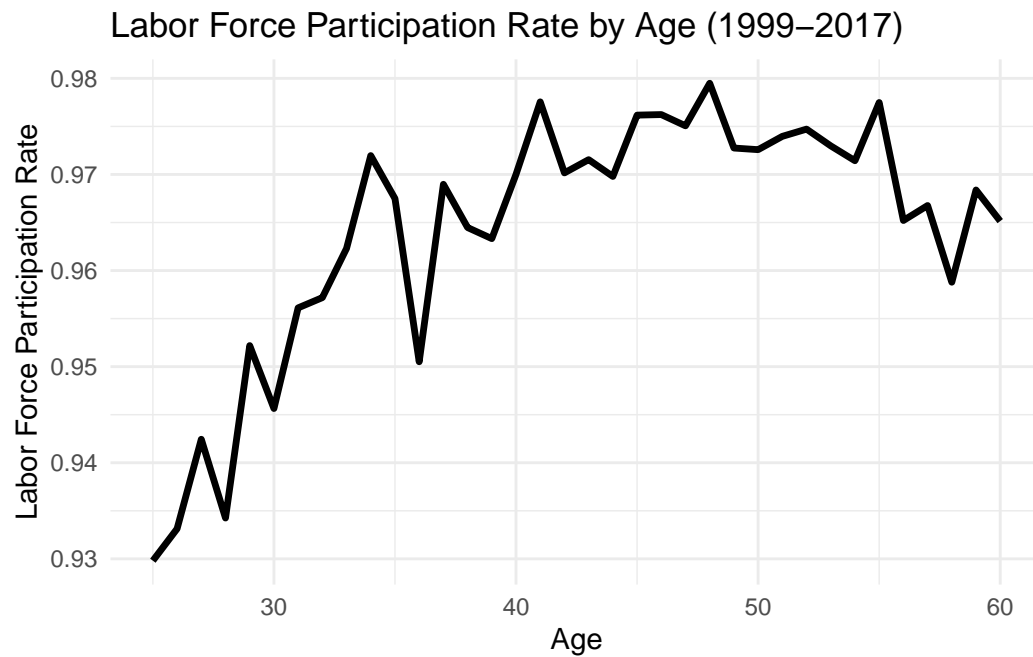
```

```

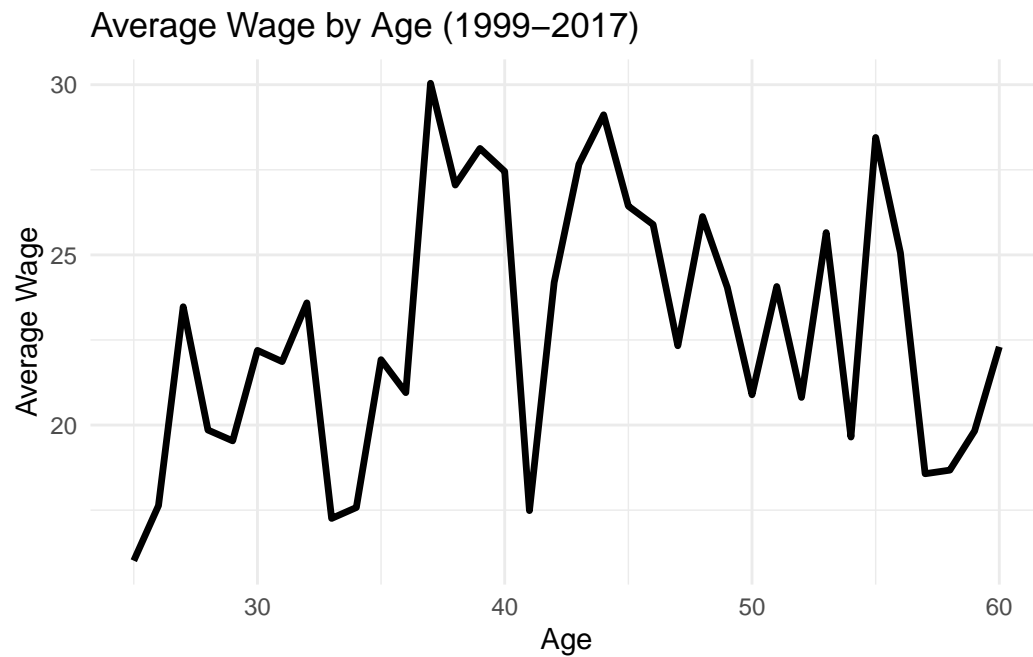
ggplot(
  lfp_all,
  aes(x = age, y = lfp_rate)
) +
  geom_line(
    size = 1.2
  ) +
  labs(
    x = "Age",
    y = "Labor Force Participation Rate",
    title = "Labor Force Participation Rate by Age (1999-2017)"
  ) +
  theme_minimal()

```

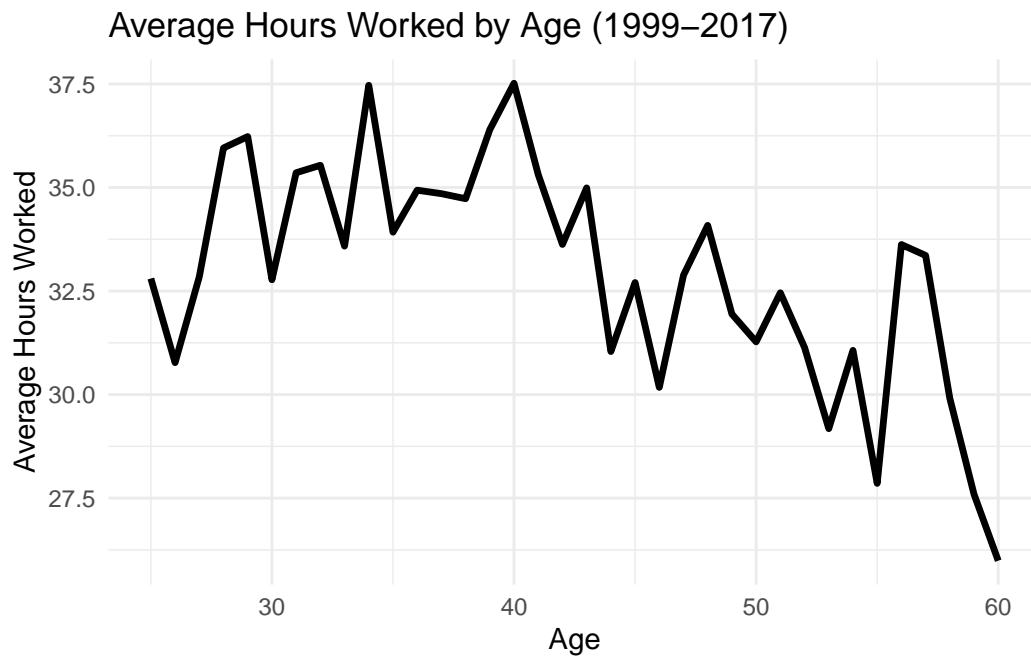
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
 i Please use `linewidth` instead.



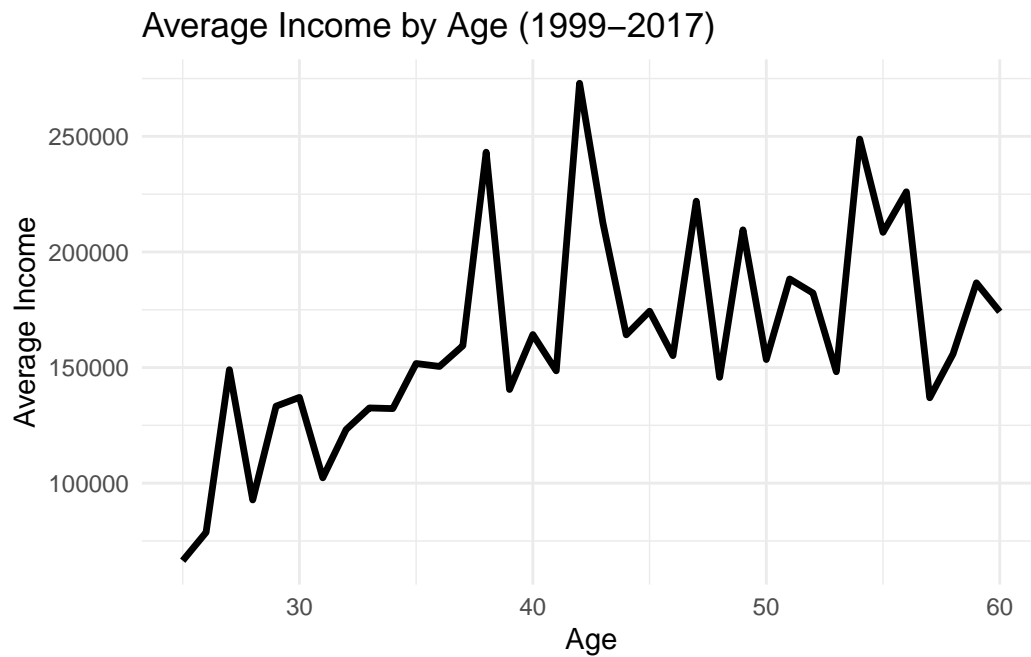
```
ggplot(  
  lfp_all,  
  aes(x = age, y = wage_rate)  
) +  
  geom_line(  
    size = 1.2  
) +  
  labs(  
    x = "Age",  
    y = "Average Wage",  
    title = "Average Wage by Age (1999–2017)"  
) +  
  theme_minimal()
```



```
ggplot(  
  lfp_all,  
  aes(x = age, y = hr_worked)  
) +  
  geom_line(  
    size = 1.2  
) +  
  labs(  
    x = "Age",  
    y = "Average Hours Worked",  
    title = "Average Hours Worked by Age (1999–2017)"  
) +  
  theme_minimal()
```

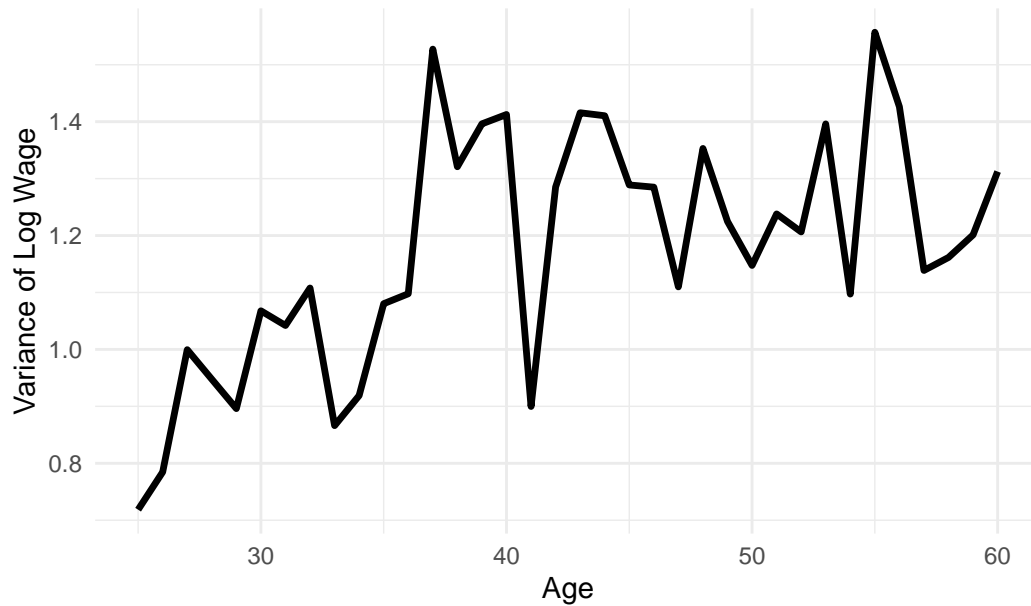



```
ggplot(  
  lfp_all,  
  aes(x = age, y = inc)  
) +  
  geom_line(  
    size = 1.2  
) +  
  labs(  
    x = "Age",  
    y = "Average Income",  
    title = "Average Income by Age (1999–2017)"  
) +  
  theme_minimal()
```

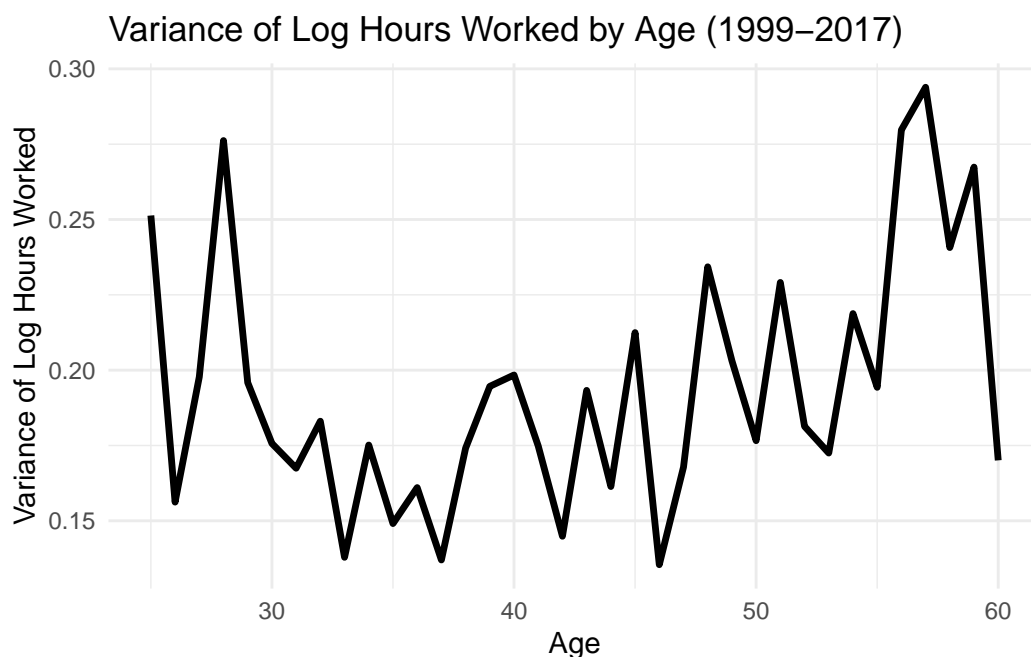


```
ggplot(  
  lfp_all,  
  aes(x = age, y = var_log_wage)  
) +  
  geom_line(  
    size = 1.2  
) +  
  labs(  
    x = "Age",  
    y = "Variance of Log Wage",  
    title = "Variance of Log Wage by Age (1999-2017)"  
) +  
  theme_minimal()
```

Variance of Log Wage by Age (1999–2017)



```
ggplot(  
  lfp_all,  
  aes(x = age, y = var_log_hr_worked)  
) +  
  geom_line(  
    size = 1.2  
) +  
  labs(  
    x = "Age",  
    y = "Variance of Log Hours Worked",  
    title = "Variance of Log Hours Worked by Age (1999–2017)"  
) +  
  theme_minimal()
```

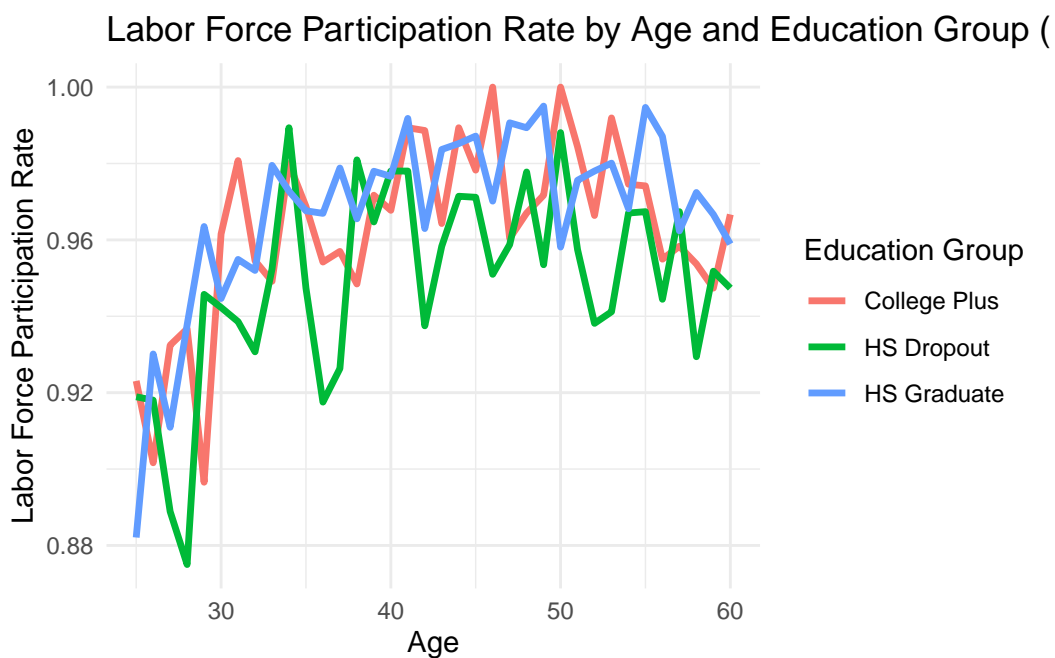


Part 2: Stratify by Education Groups:

```
lfp_edu <- psid_clean %>%
  filter(
    age >= 25 & age <= 60,
    sex == 1,
    !is.na(educ_group)
  ) %>%
  group_by(age, educ_group) %>%
  summarise(
    lfp_rate_drop = mean(lfp, na.rm = TRUE),
    wage_rate = mean(wage_real, na.rm = TRUE),
    hr_worked = mean(hr_worked, na.rm = TRUE),
    inc = mean(inc, na.rm = TRUE),
    var_log_wage = var(log_wage, na.rm = TRUE),
    var_log_hr_worked = var(log_hr_worked, na.rm = TRUE),
    n = n()
  )
```

`summarise()` has grouped output by 'age'. You can override using the `.groups` argument.

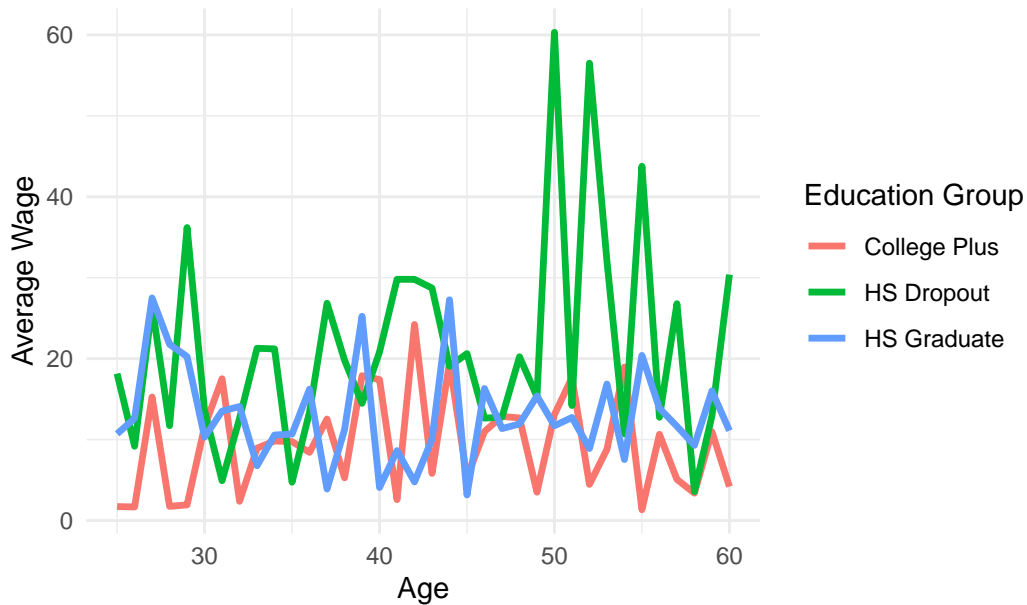
```
ggplot(
  lfp_edu,
  aes(x = age, y = lfp_rate_drop, color = educ_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Labor Force Participation Rate",
    color = "Education Group",
    title = "Labor Force Participation Rate by Age and Education Group (1999-2017)",
  ) +
  theme_minimal()
```



```
ggplot(
  lfp_edu,
  aes(x = age, y = wage_rate, color = educ_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Average Wage",
    color = "Education Group",
    title = "Average Wage by Age and Education Group (1999-2017)",
  )
```

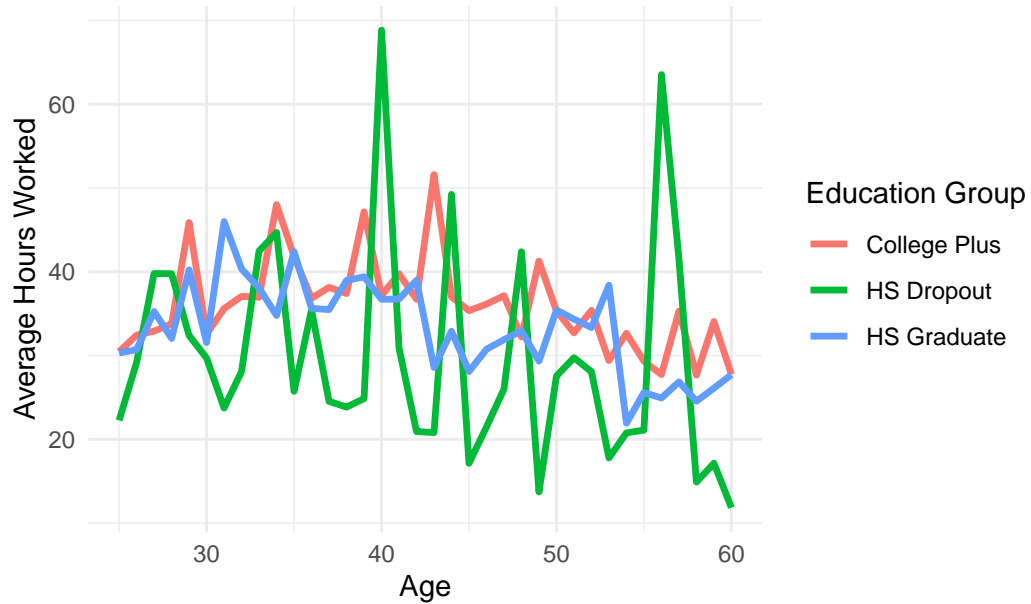
```
) +  
theme_minimal()
```

Average Wage by Age and Education Group (1999–2017)

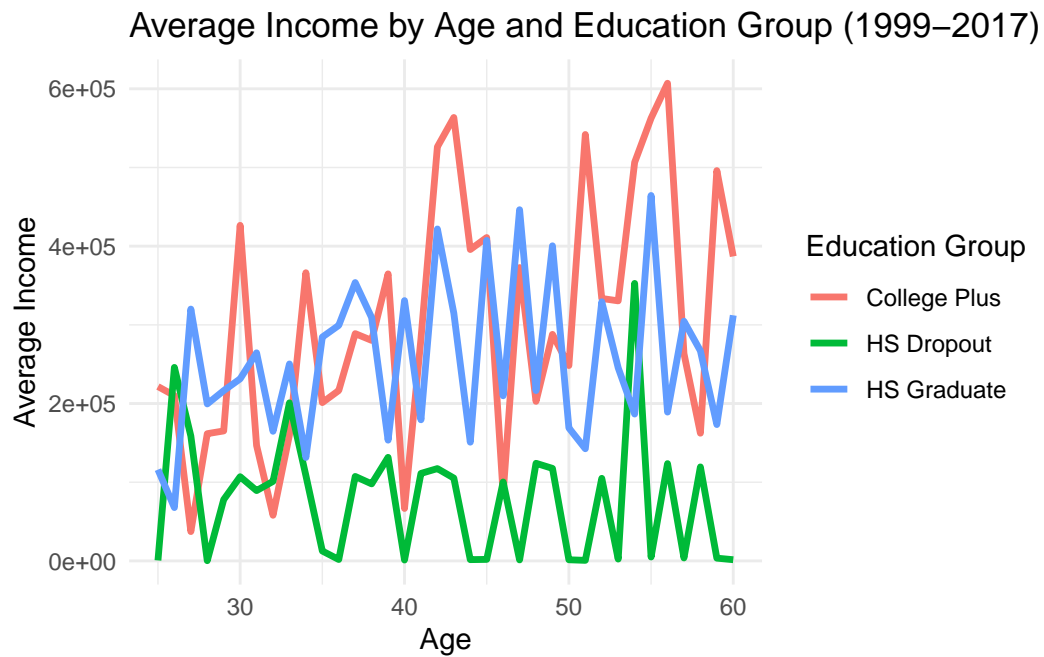


```
ggplot(  
  lfp_edu,  
  aes(x = age, y = hr_worked, color = educ_group)  
) +  
  geom_line(size = 1.2) +  
  labs(  
    x = "Age",  
    y = "Average Hours Worked",  
    color = "Education Group",  
    title = "Average Hours Worked by Age and Education Group (1999–2017)",  
  ) +  
  theme_minimal()
```

Average Hours Worked by Age and Education Group (1999–20

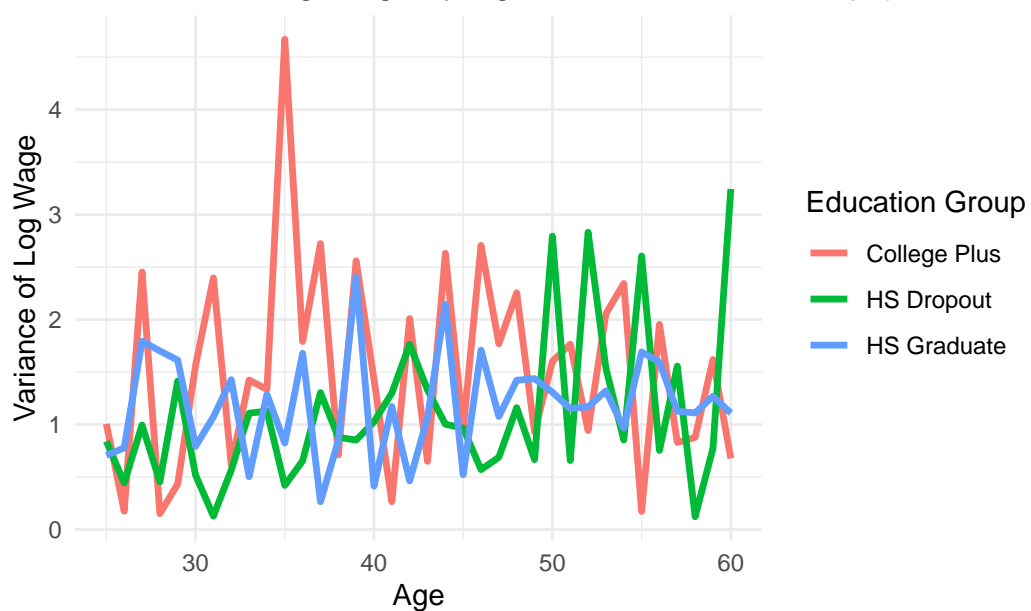


```
ggplot(
  lfp_edu,
  aes(x = age, y = inc, color = educ_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Average Income",
    color = "Education Group",
    title = "Average Income by Age and Education Group (1999-2017)",
  ) +
  theme_minimal()
```

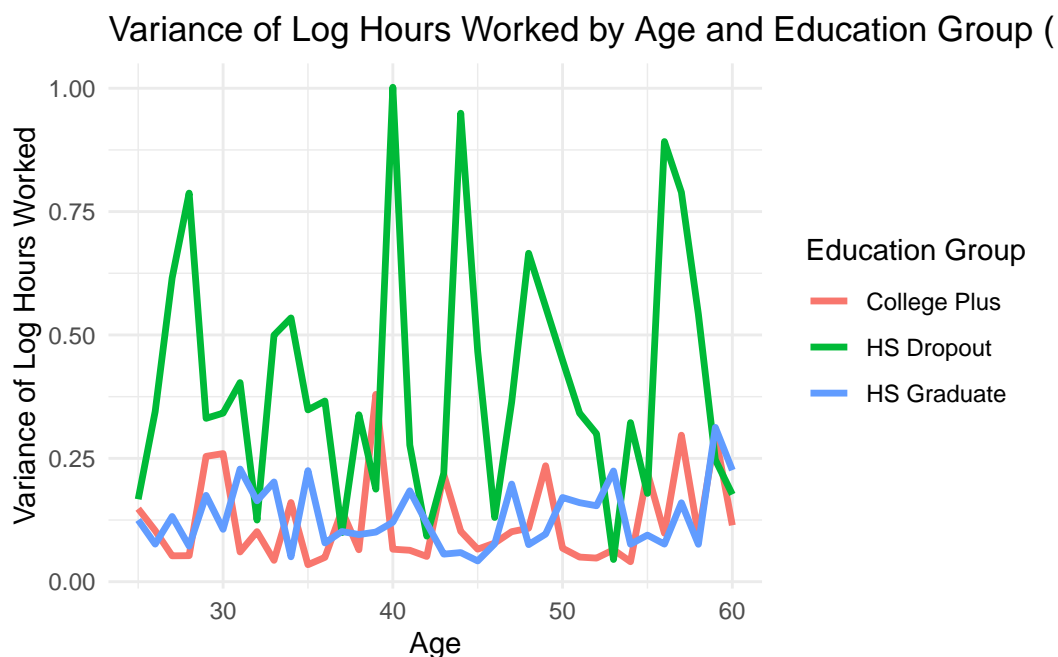


```
ggplot(
  lfp_edu,
  aes(x = age, y = var_log_wage, color = educ_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Variance of Log Wage",
    color = "Education Group",
    title = "Variance of Log Wage by Age and Education Group (1999–2017)",
  ) +
  theme_minimal()
```


Variance of Log Wage by Age and Education Group (1999–2017)



```
ggplot(
  lfp_edu,
  aes(x = age, y = var_log_hr_worked, color = educ_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Variance of Log Hours Worked",
    color = "Education Group",
    title = "Variance of Log Hours Worked by Age and Education Group (1999–2017)",
  ) +
  theme_minimal()
```

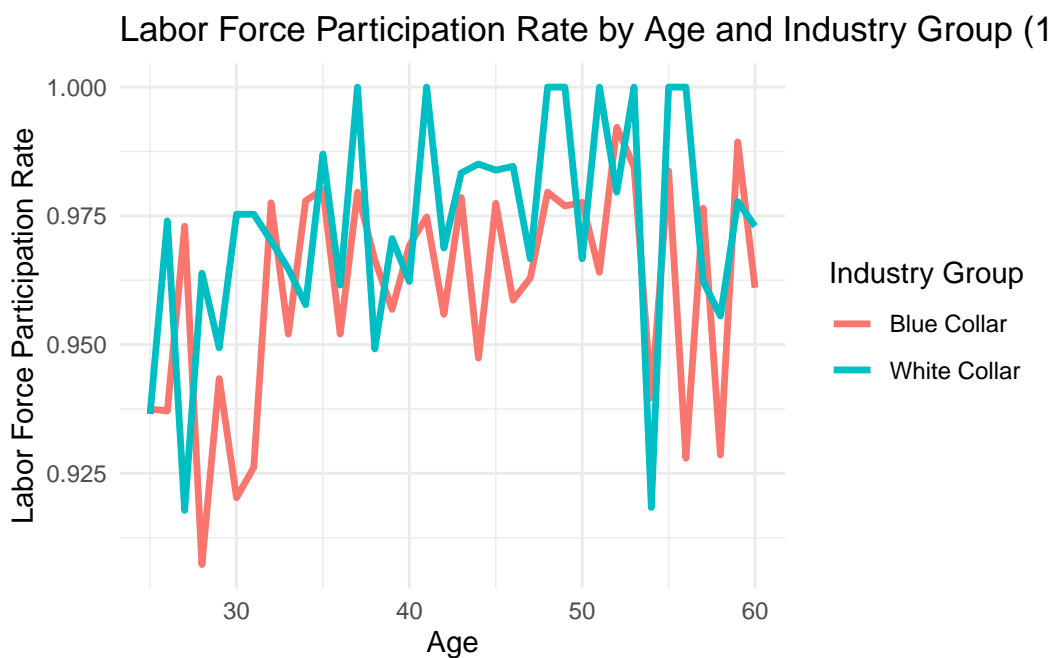


Part 3: Stratify by Industry:

```
lfp_ind <- psid_clean %>%
  filter(
    age >= 25 & age <= 60,
    sex == 1,
    !is.na(ind_group)
  ) %>%
  group_by(age, ind_group) %>%
  summarise(
    lfp_rate = mean(lfp, na.rm = TRUE),
    wage_rate = mean(wage_real, na.rm = TRUE),
    hr_worked = mean(hr_worked, na.rm = TRUE),
    inc = mean(inc, na.rm = TRUE),
    var_log_wage = var(log_wage, na.rm = TRUE),
    var_log_hr_worked = var(log_hr_worked, na.rm = TRUE),
    n = n()
  )
```

`summarise()` has grouped output by 'age'. You can override using the `.groups` argument.

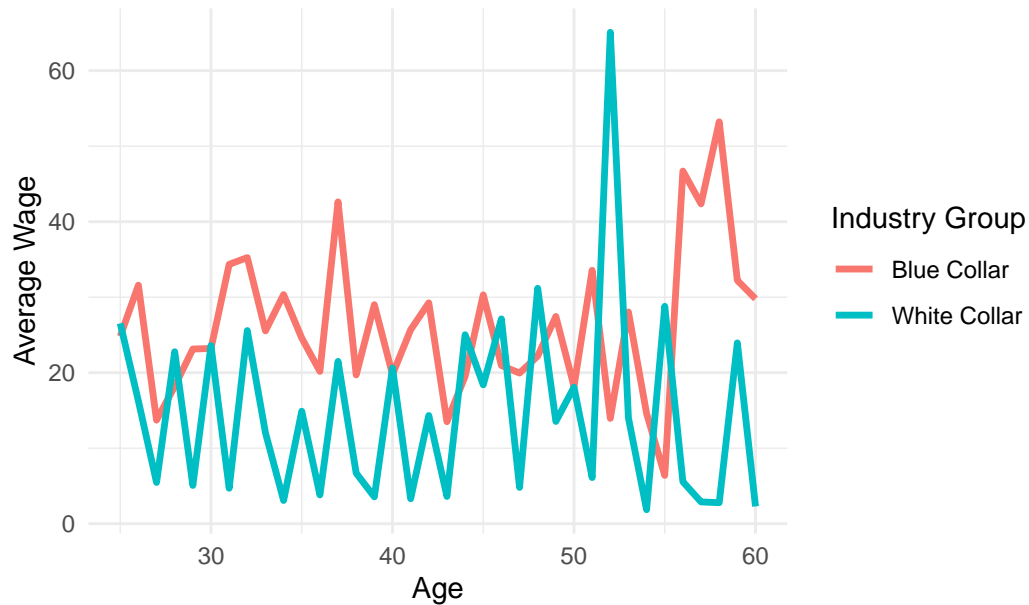
```
ggplot(
  lfp_ind,
  aes(x = age, y = lfp_rate, color = ind_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Labor Force Participation Rate",
    color = "Industry Group",
    title = "Labor Force Participation Rate by Age and Industry Group (1999-2017)",
  ) +
  theme_minimal()
```



```
ggplot(
  lfp_ind,
  aes(x = age, y = wage_rate, color = ind_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Average Wage",
    color = "Industry Group",
    title = "Average Wage by Age and Industry Group (1999-2017)",
  )
```

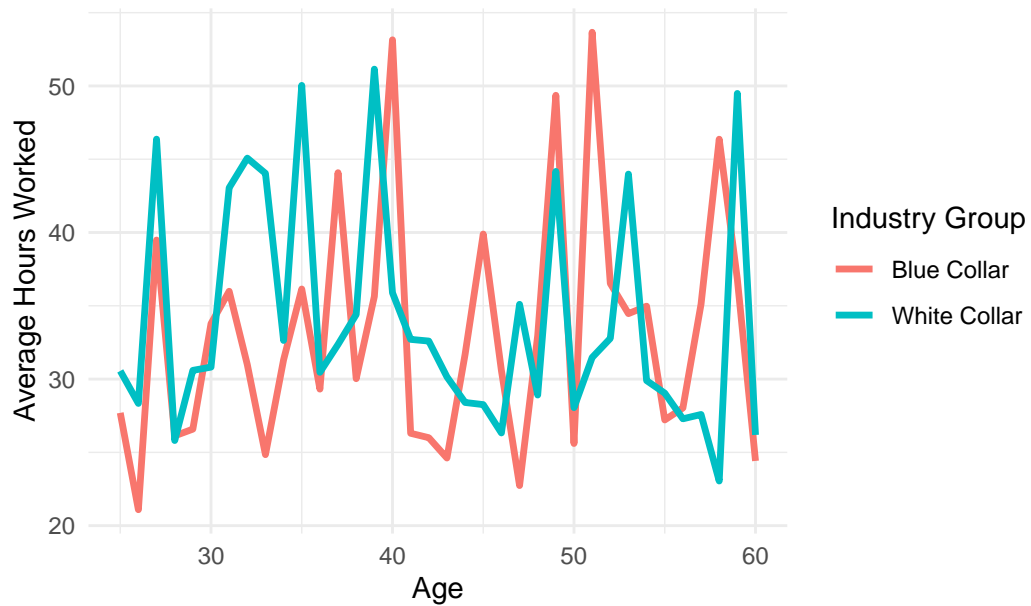
```
) +  
theme_minimal()
```

Average Wage by Age and Industry Group (1999–2017)



```
ggplot(  
  lfp_ind,  
  aes(x = age, y = hr_worked, color = ind_group)  
) +  
  geom_line(size = 1.2) +  
  labs(  
    x = "Age",  
    y = "Average Hours Worked",  
    color = "Industry Group",  
    title = "Average Hours Worked by Age and Industry Group (1999–2017)",  
  ) +  
  theme_minimal()
```

Average Hours Worked by Age and Industry Group (1999–2017)

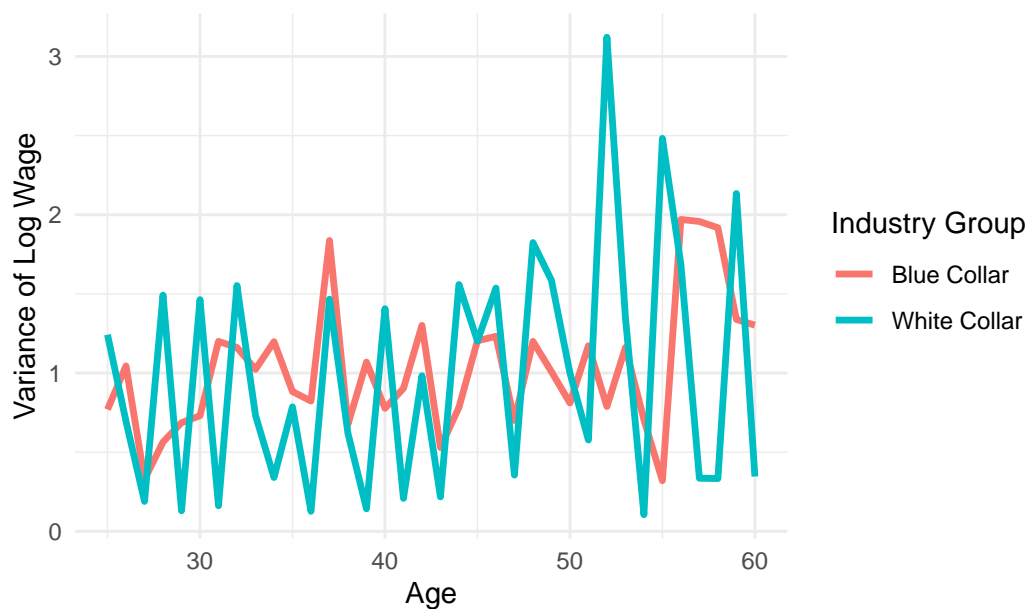


```
ggplot(  
  lfp_ind,  
  aes(x = age, y = inc, color = ind_group)  
) +  
  geom_line(size = 1.2) +  
  labs(  
    x = "Age",  
    y = "Average Income",  
    color = "Industry Group",  
    title = "Average Income by Age and Industry Group (1999–2017)",  
  ) +  
  theme_minimal()
```



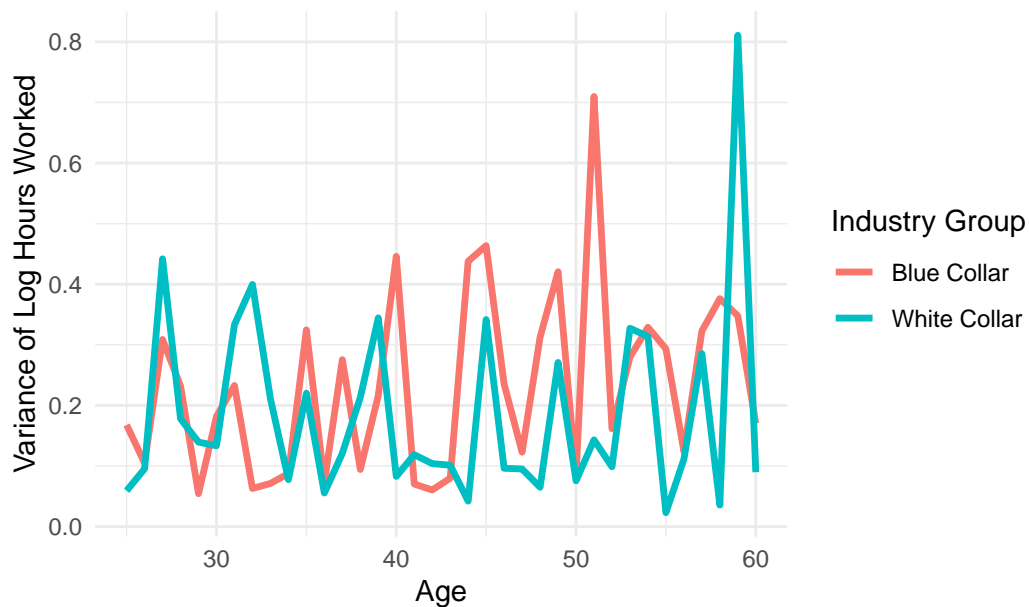
```
ggplot(
  lfp_ind,
  aes(x = age, y = var_log_wage, color = ind_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Variance of Log Wage",
    color = "Industry Group",
    title = "Variance of Log Wage by Age and Industry Group (1999–2017)",
  ) +
  theme_minimal()
```

Variance of Log Wage by Age and Industry Group (1999–2017)



```
ggplot(
  lfp_ind,
  aes(x = age, y = var_log_hr_worked, color = ind_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Variance of Log Hours Worked",
    color = "Industry Group",
    title = "Variance of Log Hours Worked by Age and Industry Group (1999–2017)",
  ) +
  theme_minimal()
```

Variance of Log Hours Worked by Age and Industry Group (1990)



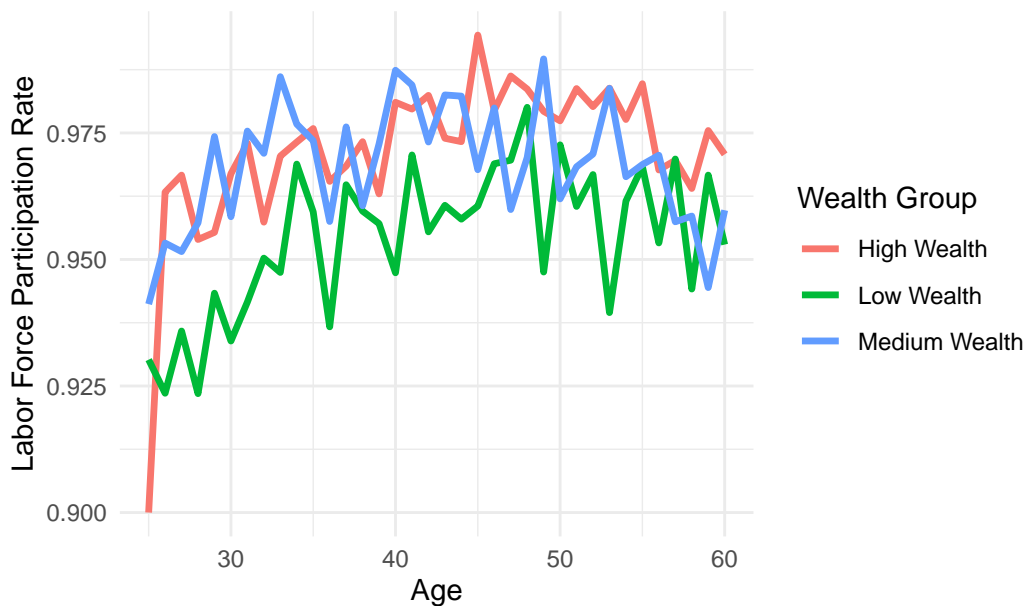
Part 4: Stratify by Wealth Quartiles:

```
lfp_wealth <- psid_clean %>%
  filter(
    age >= 25 & age <= 60,
    sex == 1,
    !is.na(wealth)
  ) %>%
  group_by(age, wealth_group) %>%
  summarise(
    lfp_rate = mean(lfp, na.rm = TRUE),
    wage_rate = mean(wage_real, na.rm = TRUE),
    hr_worked = mean(hr_worked, na.rm = TRUE),
    inc = mean(inc, na.rm = TRUE),
    var_log_wage = var(log_wage, na.rm = TRUE),
    var_log_hr_worked = var(log_hr_worked, na.rm = TRUE),
    n = n()
  )
```

`summarise()` has grouped output by 'age'. You can override using the `.groups` argument.

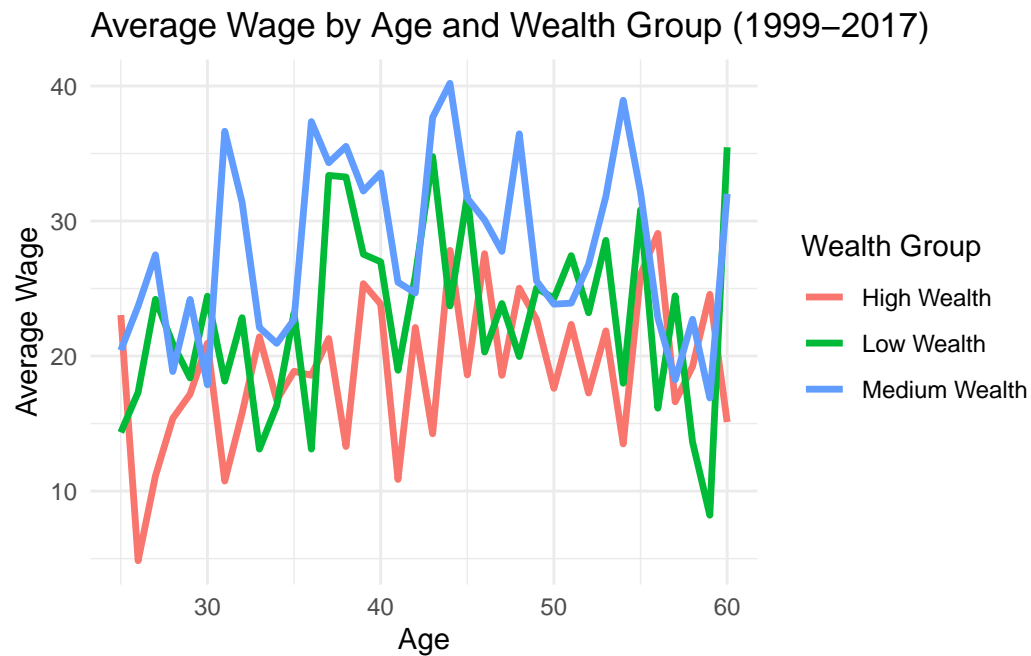

```
ggplot(
  lfp_wealth,
  aes(x = age, y = lfp_rate, color = wealth_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Labor Force Participation Rate",
    color = "Wealth Group",
    title = "Labor Force Participation Rate by Age and Wealth Group (1999-2017)",
  ) +
  theme_minimal()
```

Labor Force Participation Rate by Age and Wealth Group (19



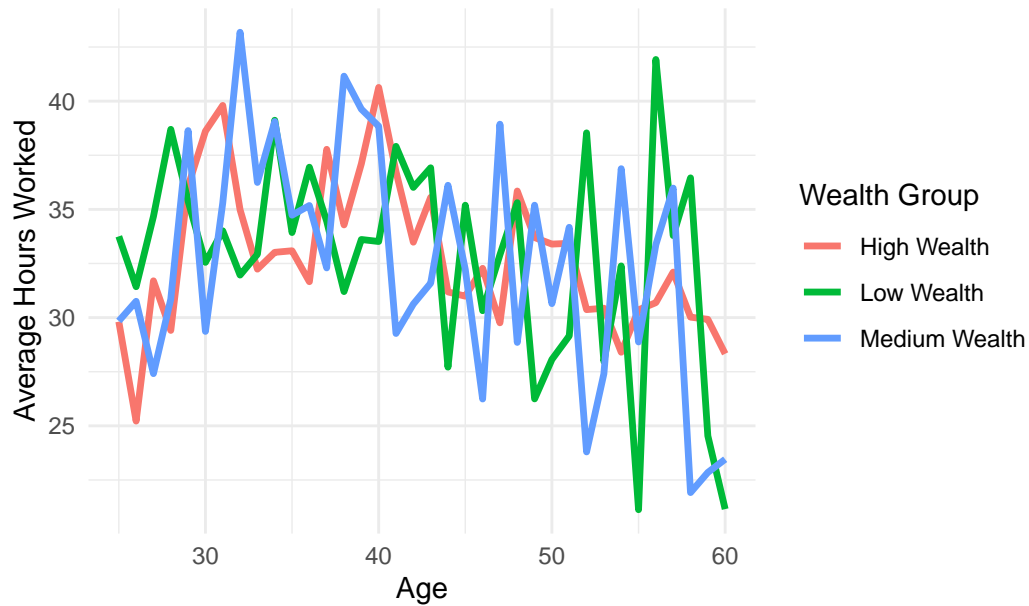
```
ggplot(
  lfp_wealth,
  aes(x = age, y = wage_rate, color = wealth_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Average Wage",
    color = "Wealth Group",
    title = "Average Wage by Age and Wealth Group (1999-2017)",
  )
```

```
) +  
theme_minimal()
```

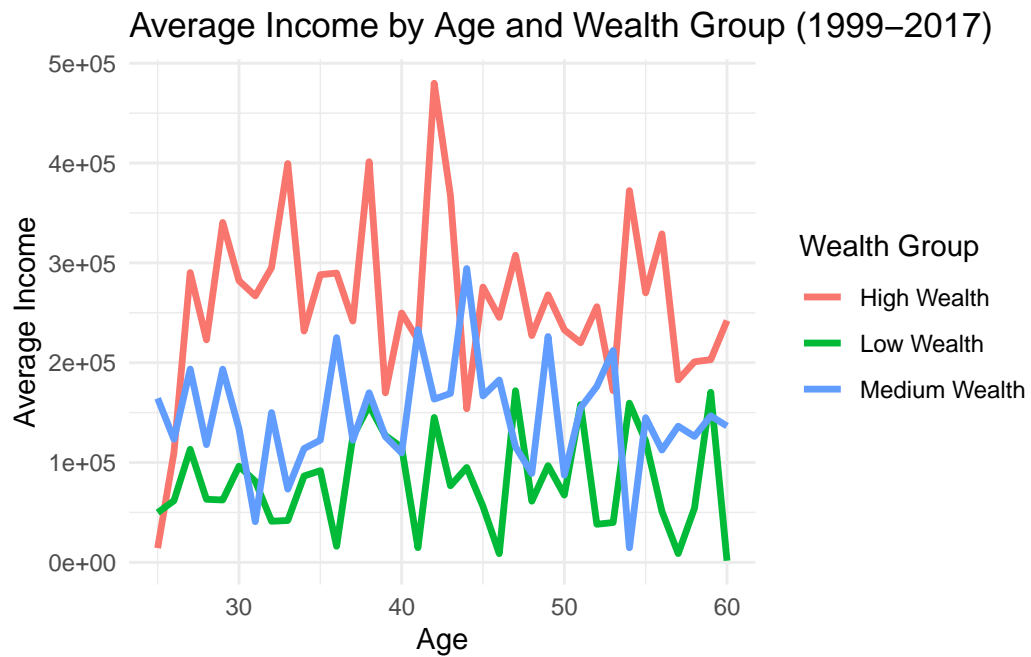


```
ggplot(  
  lfp_wealth,  
  aes(x = age, y = hr_worked, color = wealth_group)  
) +  
  geom_line(size = 1.2) +  
  labs(  
    x = "Age",  
    y = "Average Hours Worked",  
    color = "Wealth Group",  
    title = "Average Hours Worked by Age and Wealth Group (1999–2017)",  
  ) +  
  theme_minimal()
```

Average Hours Worked by Age and Wealth Group (1999–2017)

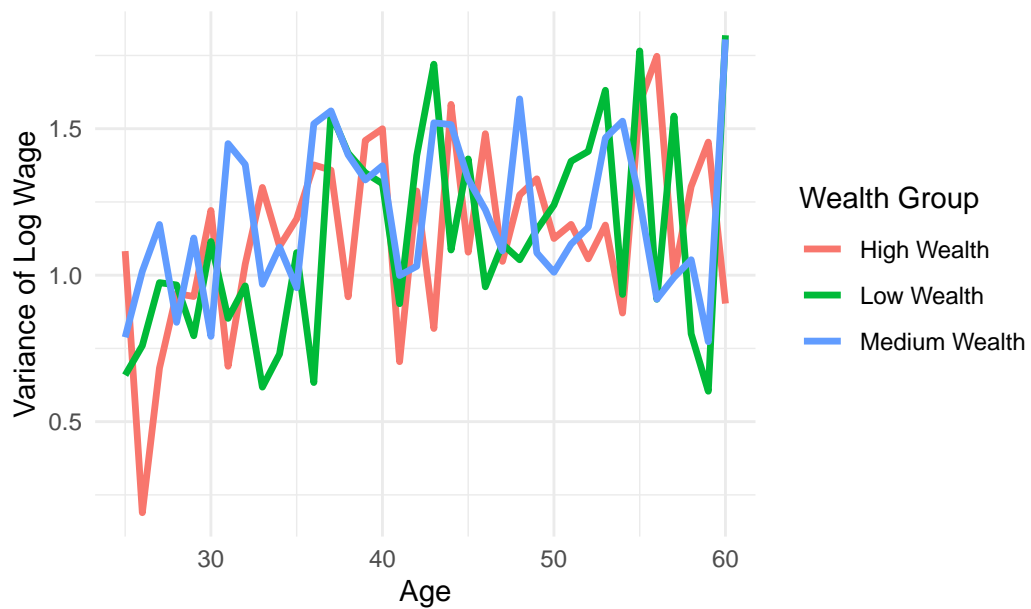


```
ggplot(
  lfp_wealth,
  aes(x = age, y = inc, color = wealth_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Average Income",
    color = "Wealth Group",
    title = "Average Income by Age and Wealth Group (1999–2017)",
  ) +
  theme_minimal()
```



```
ggplot(
  lfp_wealth,
  aes(x = age, y = var_log_wage, color = wealth_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Variance of Log Wage",
    color = "Wealth Group",
    title = "Variance of Log Wage by Age and Wealth Group (1999–2017)",
  ) +
  theme_minimal()
```

Variance of Log Wage by Age and Wealth Group (1999–2017)



```
ggplot(
  lfp_wealth,
  aes(x = age, y = var_log_hr_worked, color = wealth_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Variance of Log Hours Worked",
    color = "Wealth Group",
    title = "Variance of Log Hours Worked by Age and Wealth Group (1999–2017)",
  ) +
  theme_minimal()
```

Variance of Log Hours Worked by Age and Wealth Group (1990)

