Problem Set 1: PSID - Labor Outcomes

Tate Mason

Part 1: Overall Trends:

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

```
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,
    weight = ER16518,
    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,
    weight = ER20394,
    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,
    weight = ER24179,
    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,
    weight = ER28078,
    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,
    weight = ER41069,
    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,
    weight = ER47012,
    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,
    weight = ER52436,
    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,
    weight = ER58257,
    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,
    weight = ER65492,
    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,
    weight = ER71570,
    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) %>%
  filter(
    wage <= 997,
    inc <= 9999997,
    hr_worked <= 112
  ) %>%
  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_
    )
  )
describe(psid_clean)
```

```
              vars     n      mean         sd   median  trimmed      mad
year             1 81283   2008.53       5.71  2009.00  2008.63     5.93
age              2 81283     45.48      21.21    43.00    44.00    17.79
sex              3 81283      1.31       0.46     1.00     1.27     0.00
lfp              4 81259      0.95       0.21     1.00     1.00     0.00
wage             5 81283      5.78      15.26     0.00     3.57     0.00
hr_worked        6 81283     23.42      23.09    30.00    21.69    44.48
cpi_ratio        7 81283      0.81       0.10     0.83     0.82     0.11
educ_HS          8 81283      1.32       1.27     1.00     1.15     0.00
educ_coll        9 81283      1.32       8.36     0.00     0.40     0.00
ind             10 81283    965.18    1857.02   617.00   500.95   444.78
wealth          11 81283 218416.11 1070901.66 30500.00 81265.13 52928.82
weight          12 81283     21.68      18.42    18.52    19.32    18.40
inc             13 81283  15799.99   63254.43     0.00  4704.69     0.00
educ_group*     14 27270      2.19       0.80     2.00     2.23     1.48
ind_group*      15  9741      1.42       0.49     1.00     1.39     0.00
wealth_group*   16 81281      1.88       0.71     2.00     1.84     1.48
wage_real       17 81283      4.71      12.05     0.00     2.84     0.00
inc_real        18 81283  12709.88   48435.40     0.00  3690.52     0.00
log_wage        19 28935      2.38       0.57     2.34     2.37     0.50
```

```
log_hr_worked   20 44894      3.68       0.43     3.69     3.74     0.17
                      min          max        range  skew kurtosis       se
year              1999.00 2.01700e+03 1.80000e+01 -0.11    -1.20     0.02
age                 16.00 9.99000e+02 9.83000e+02 18.14   801.32     0.07
sex                  1.00 2.00000e+00 1.00000e+00  0.80    -1.36     0.00
lfp                  0.00 1.00000e+00 1.00000e+00 -4.38    17.21     0.00
wage                 0.00 9.75000e+02 9.75000e+02 29.35  1540.13     0.05
hr_worked            0.00 1.12000e+02 1.12000e+02  0.27    -1.23     0.08
cpi_ratio            0.64 9.50000e-01 3.00000e-01 -0.30    -1.22     0.00
educ_HS              0.00 9.00000e+00 9.00000e+00  3.93    20.48     0.00
educ_coll            0.00 9.90000e+01 9.90000e+01 11.21   125.80     0.03
ind                  0.00 9.99900e+03 9.99900e+03  3.34    10.22     6.51
wealth        -2699990.00 1.00555e+08 1.03255e+08 36.20  2408.83 3756.21
weight               0.00 1.67680e+02 1.67680e+02  1.20     1.94     0.06
inc                  0.00 5.00000e+06 5.00000e+06 18.82   759.74   221.87
educ_group*          1.00 3.00000e+00 2.00000e+00 -0.35    -1.34     0.00
ind_group*           1.00 2.00000e+00 1.00000e+00  0.34    -1.88     0.00
wealth_group*        1.00 3.00000e+00 2.00000e+00  0.19    -1.04     0.00
wage_real            0.00 8.45320e+02 8.45320e+02 26.13  1337.44     0.04
inc_real             0.00 4.50500e+06 4.50500e+06 21.60  1207.50   169.89
log_wage            -1.58 6.74000e+00 8.32000e+00  0.41     3.38     0.00
log_hr_worked        0.00 4.72000e+00 4.72000e+00 -3.37    18.66     0.00
```

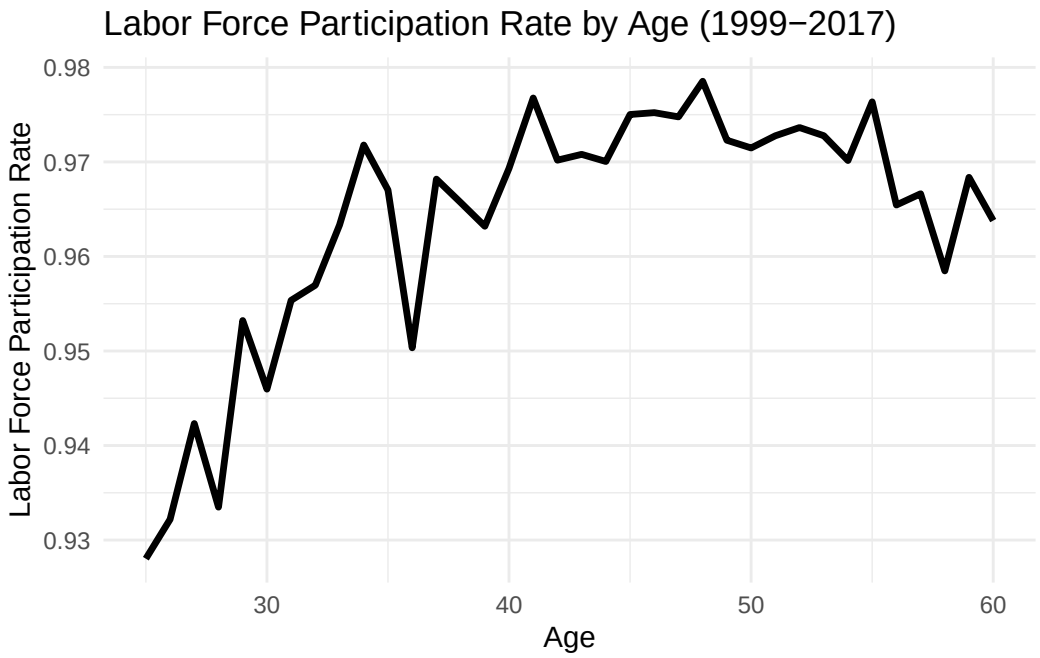
```r
lfp_all <- psid_clean %>%
  filter(
    age >= 25 & age <= 60,
    sex == 1,
  ) %>%
  group_by(age) %>%
  summarise(
    lfp_rate = mean(lfp, na.rm = TRUE),
    wage_rate = mean(wage_real, na.rm = TRUE),
    hr_worked = mean(hr_worked, na.rm = TRUE),
    inc = mean(inc, na.rm = TRUE),
    var_log_wage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

```
               vars  n     mean      sd   median  trimmed     mad     min
age               1 36    42.50   10.54    42.50    42.50   13.34   25.00
```

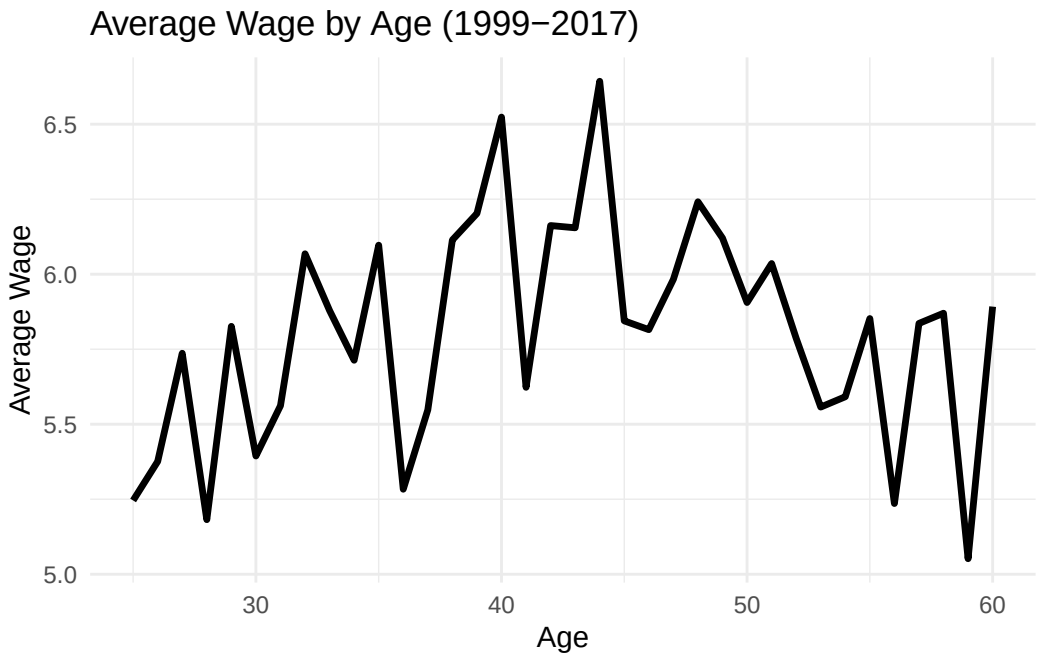
```
lfp_rate             2 36     0.96    0.01     0.97     0.97    0.01    0.93
wage_rate            3 36     5.80    0.37     5.84     5.80    0.39    5.05
hr_worked            4 36    28.16    2.49    27.80    28.16    3.35   24.28
inc                  5 36 23663.71 5792.52 25237.27 24242.94 4079.38 8870.08
var_log_wage         6 36     0.31    0.06     0.31     0.31    0.05    0.20
var_log_hr_worked    7 36     0.12    0.03     0.11     0.12    0.02    0.07
n                    8 36  1195.78  171.06  1204.50  1206.17  163.09  801.00
                       max    range  skew kurtosis     se
age                  60.00    35.00  0.00    -1.30   1.76
lfp_rate              0.98     0.05 -1.28     0.68   0.00
wage_rate             6.64     1.59 -0.03    -0.45   0.06
hr_worked            32.22     7.94  0.07    -1.41   0.42
inc                31761.53 22891.45 -0.98     0.05 965.42
var_log_wage           0.44     0.24 -0.06    -0.54   0.01
var_log_hr_worked      0.20     0.12  0.67    -0.06   0.00
n                   1475.00   674.00 -0.55    -0.39  28.51
```

```r
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()
```

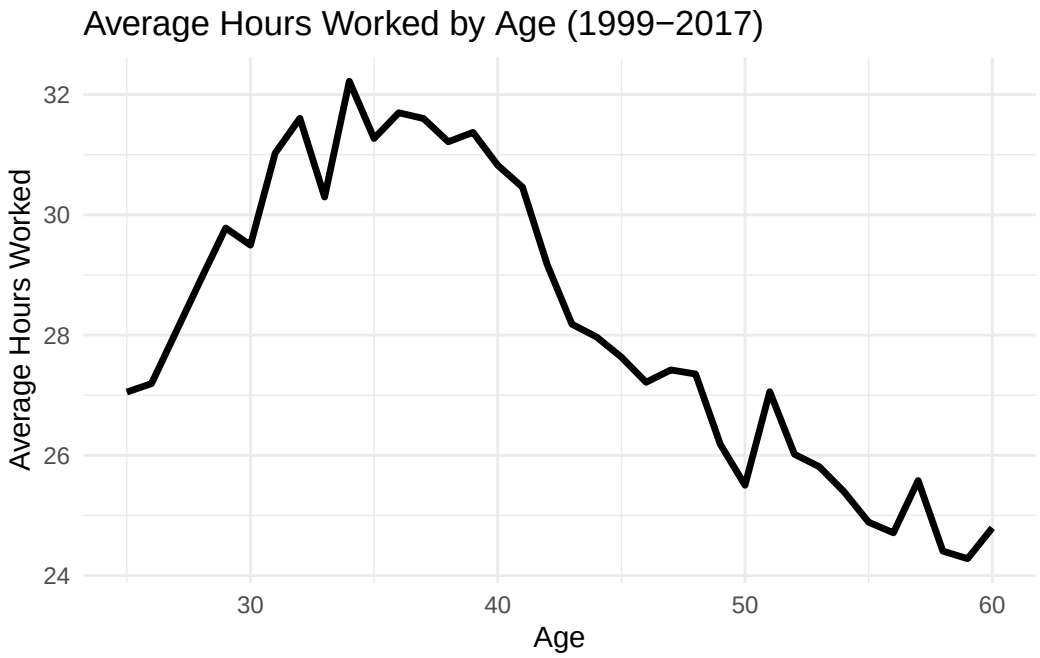
```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```



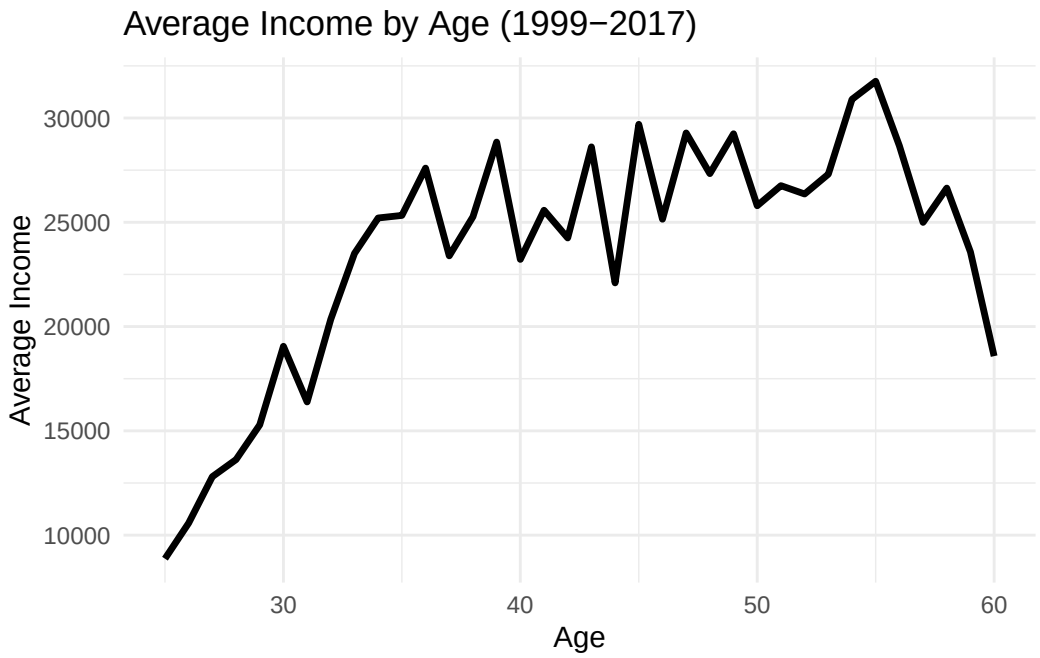

```
ggplot(
  lfp_all,
  aes(x = age, y = wage_rate)
) +
  geom_line(
    size = 1.2
  ) +
  labs(
    x = "Age",
    y = "Average Wage",
    title = "Average Wage by Age (1999-2017)"
  ) +
  theme_minimal()
```



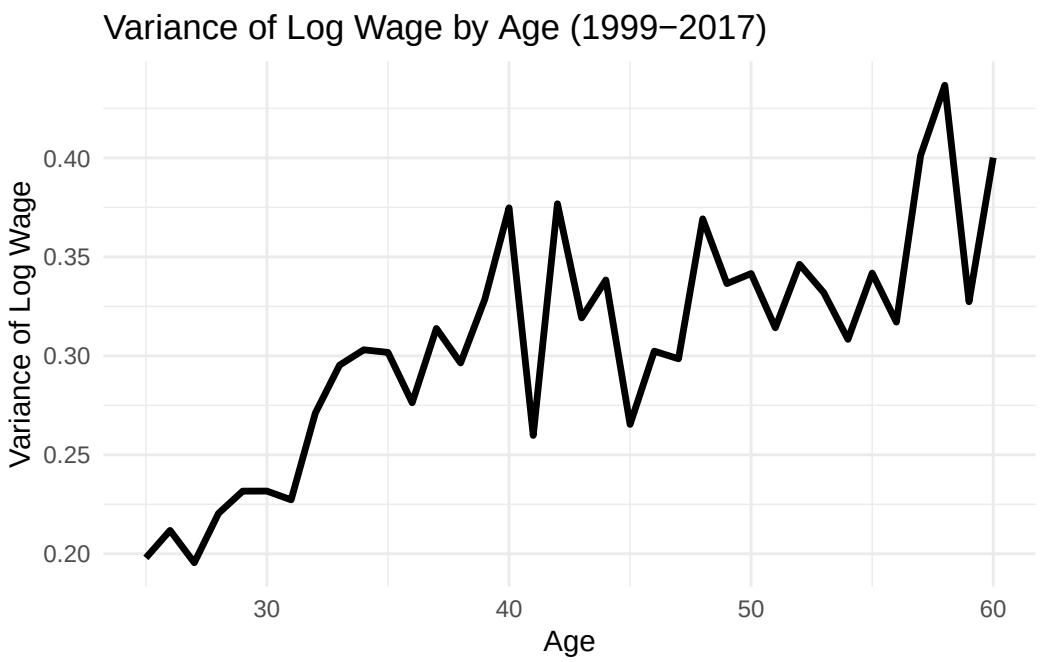

```
ggplot(
  lfp_all,
  aes(x = age, y = hr_worked)
) +
  geom_line(
    size = 1.2
  ) +
  labs(
    x = "Age",
    y = "Average Hours Worked",
    title = "Average Hours Worked by Age (1999-2017)"
  ) +
  theme_minimal()
```



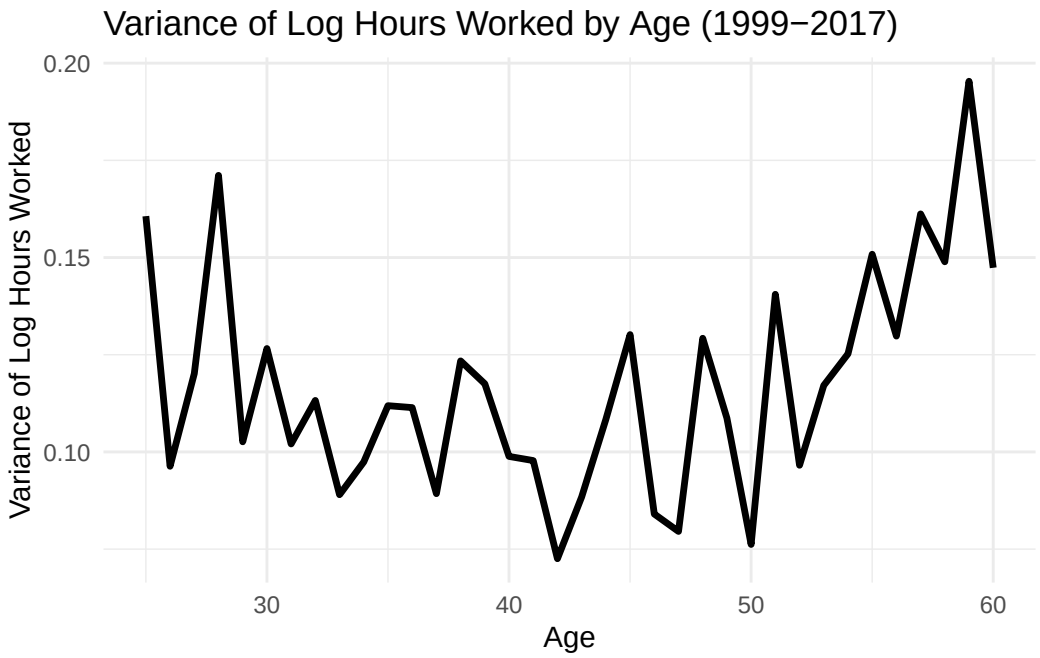

```
ggplot(
  lfp_all,
  aes(x = age, y = inc)
) +
  geom_line(
    size = 1.2
  ) +
  labs(
    x = "Age",
    y = "Average Income",
    title = "Average Income by Age (1999-2017)"
  ) +
  theme_minimal()
```




```
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()
```




```
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()
```

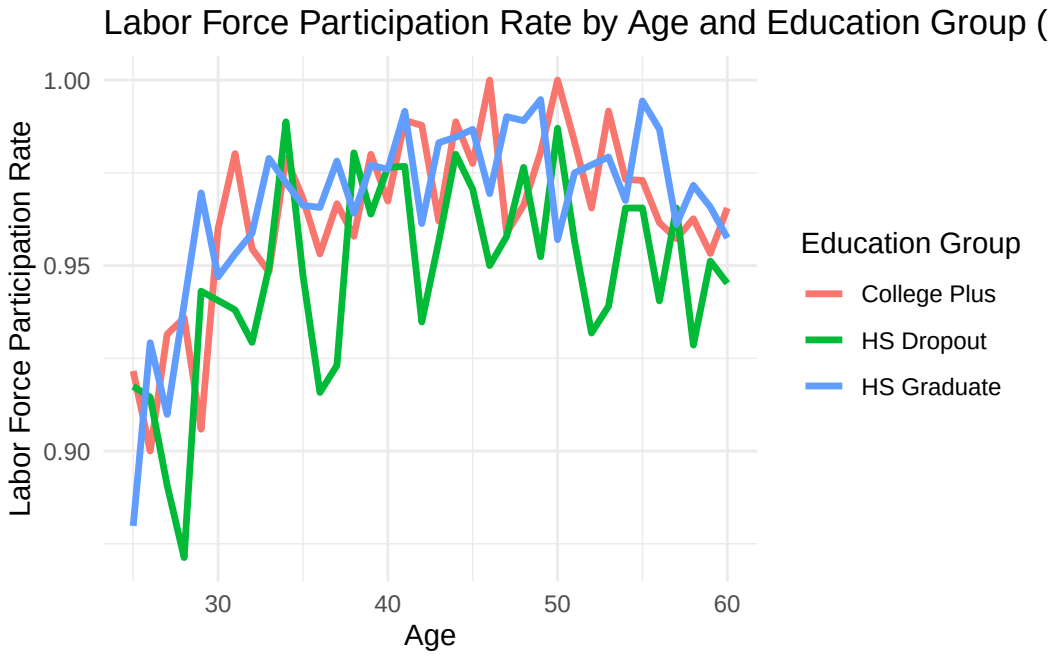


Part 2: Stratify by Education Groups:

```
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()
  )
```

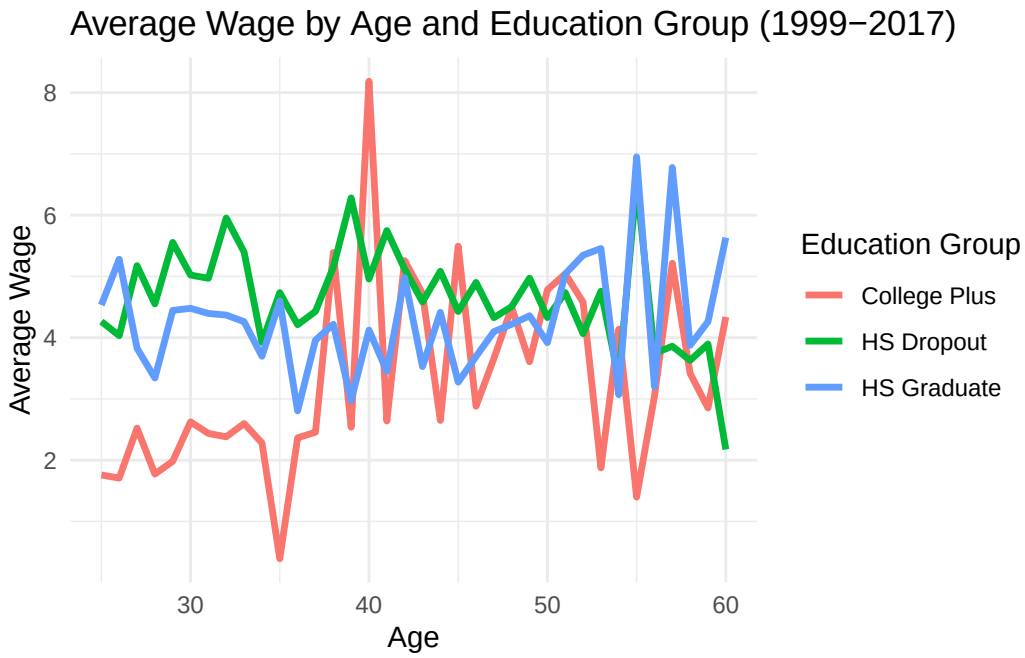
```
`summarise()` has grouped output by 'age'. You can override using the `.groups`
argument.
```

```r
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()
```

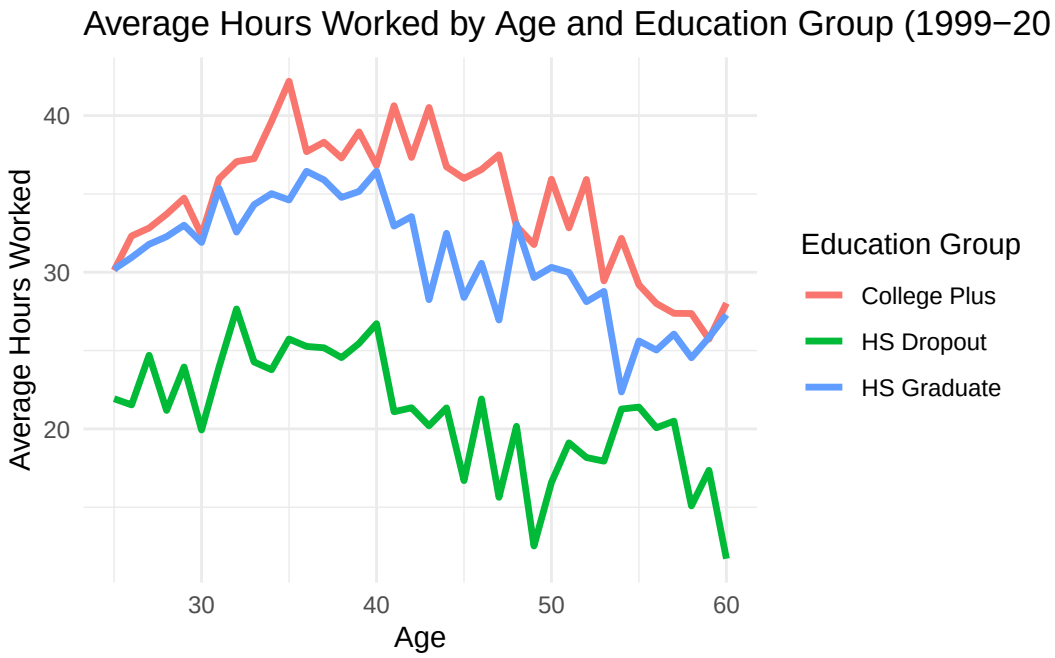



```r
ggplot(
  lfp_edu,
  aes(x = age, y = wage_rate, color = educ_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Average Wage",
    color = "Education Group",
    title = "Average Wage by Age and Education Group (1999-2017)",
```

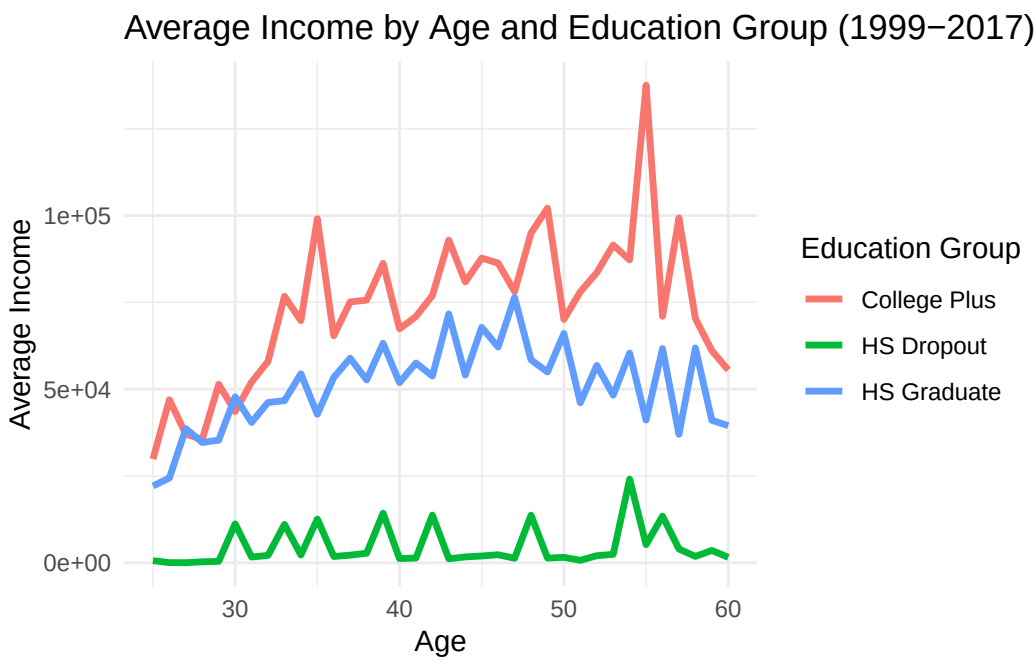
```
  ) +
  theme_minimal()
```



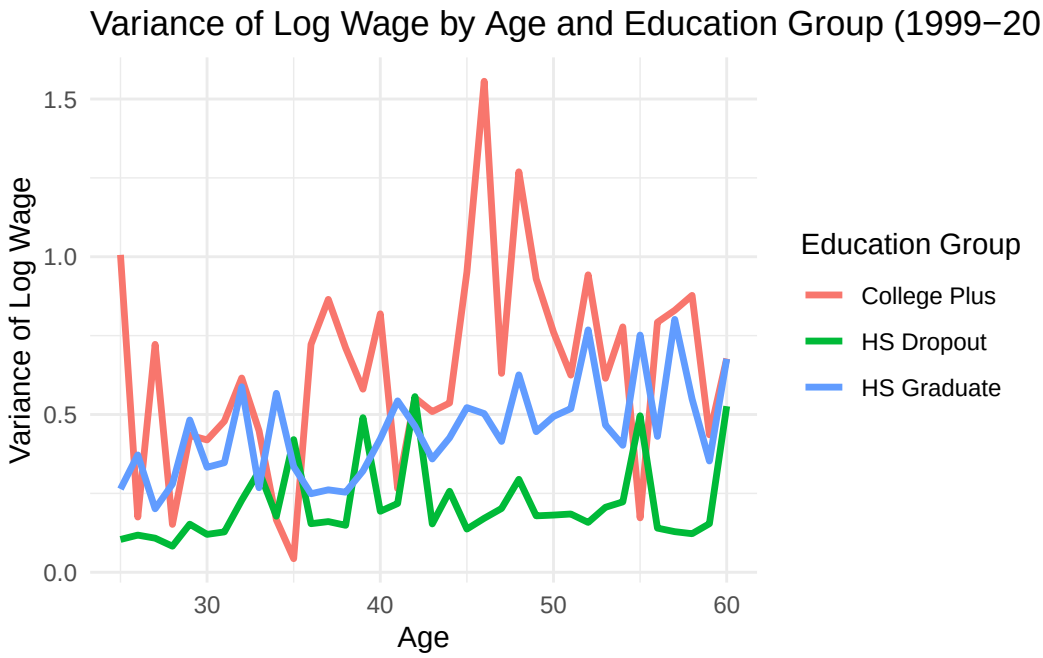

```
ggplot(
  lfp_edu,
  aes(x = age, y = hr_worked, color = educ_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Average Hours Worked",
    color = "Education Group",
    title = "Average Hours Worked by Age and Education Group (1999-2017)",
  ) +
  theme_minimal()
```



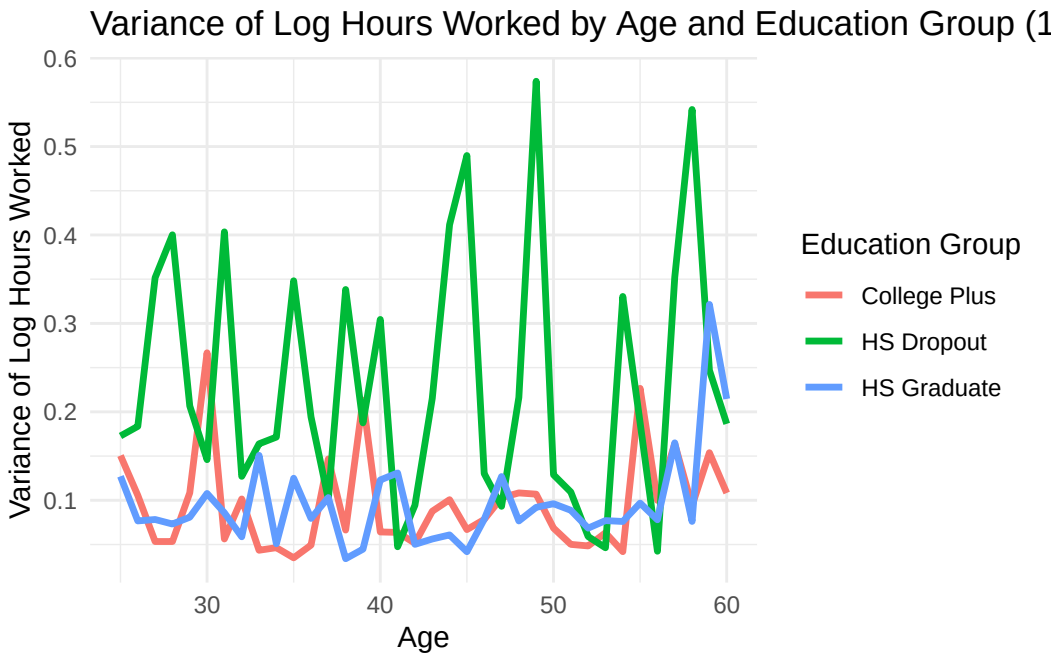

```
ggplot(
  lfp_edu,
  aes(x = age, y = inc, color = educ_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Average Income",
    color = "Education Group",
    title = "Average Income by Age and Education Group (1999-2017)",
  ) +
  theme_minimal()
```




```
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()
```




```
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()
```

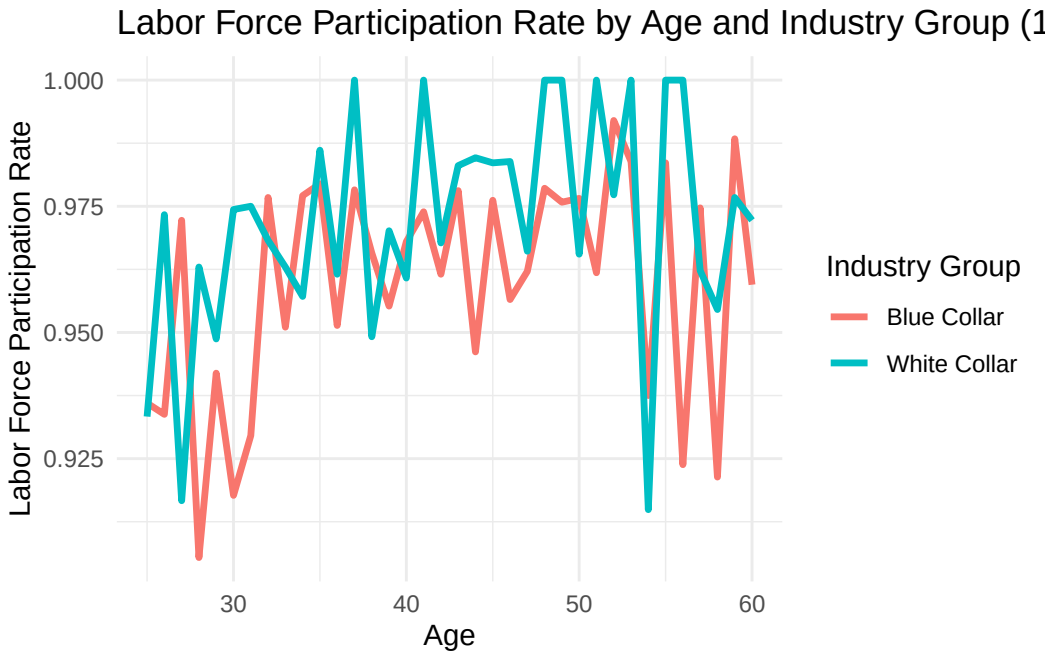


Part 3: Stratify by Industry:

```
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()
  )
```

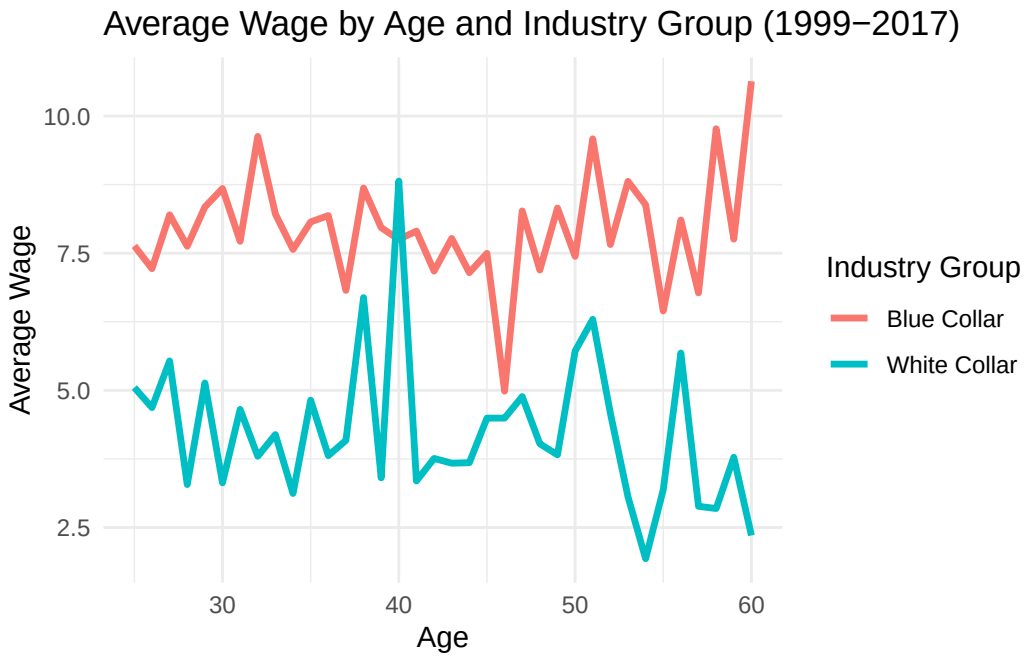
```
`summarise()` has grouped output by 'age'. You can override using the `.groups`
argument.
```

```r
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()
```

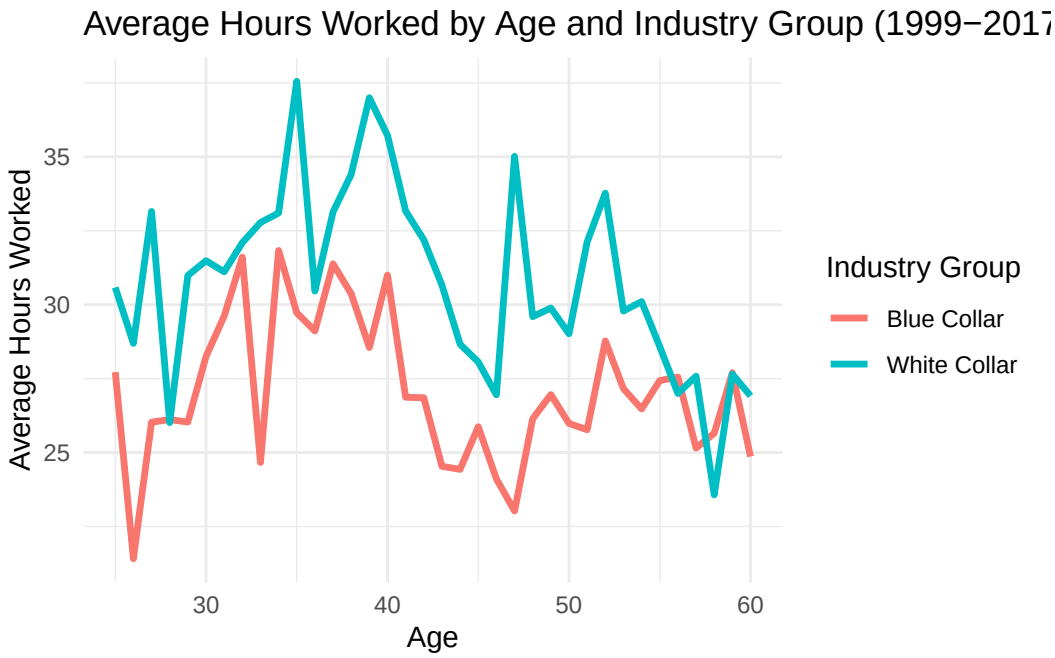



```r
ggplot(
  lfp_ind,
  aes(x = age, y = wage_rate, color = ind_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Average Wage",
    color = "Industry Group",
    title = "Average Wage by Age and Industry Group (1999-2017)",
```

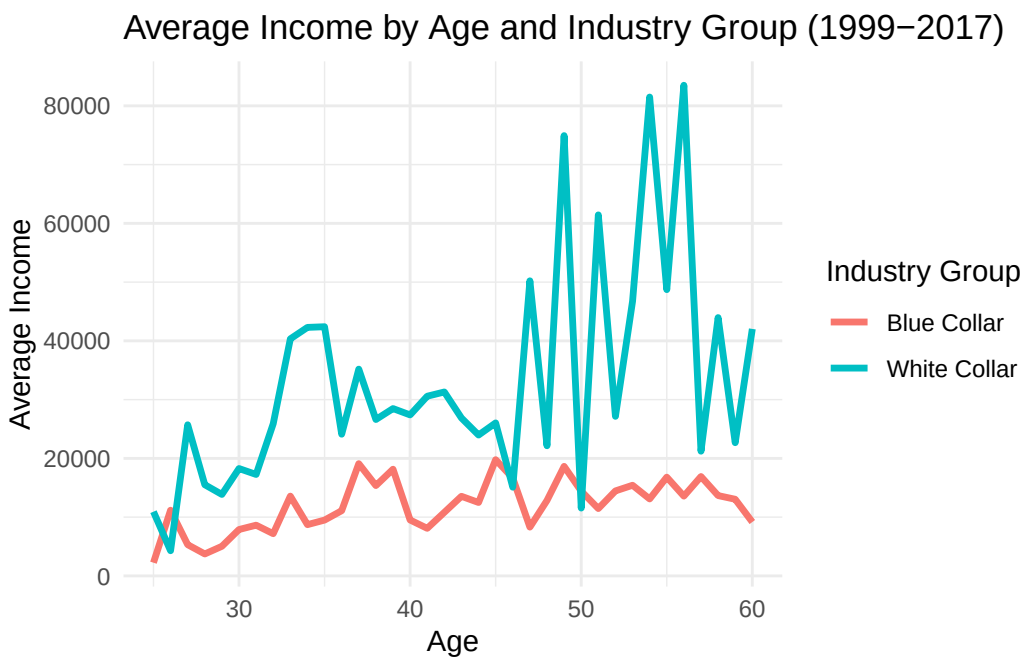
```
  ) +
  theme_minimal()
```



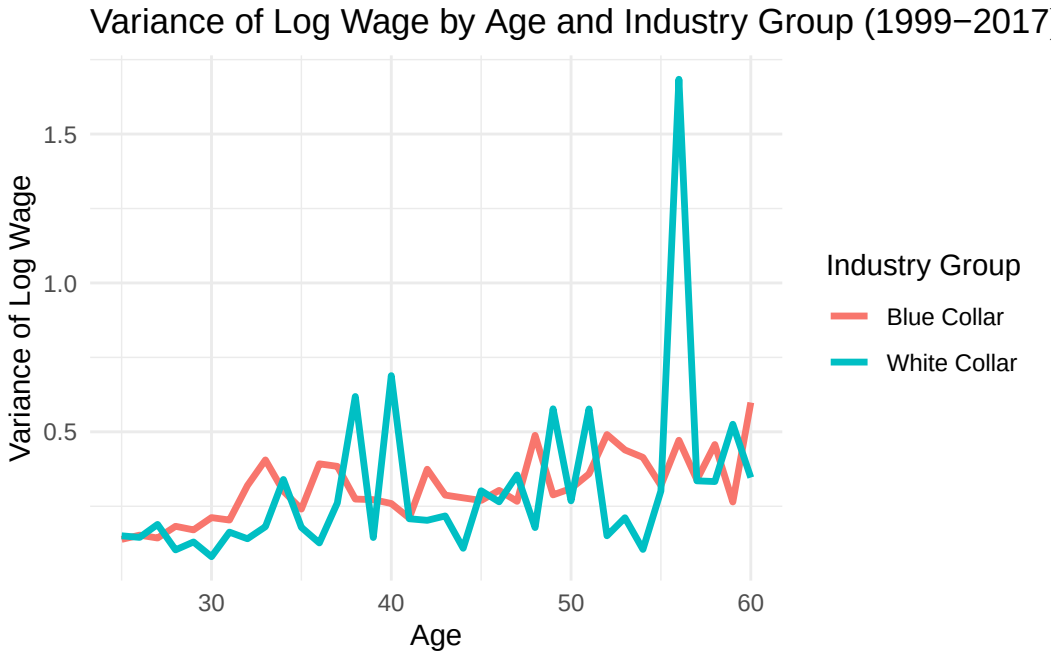

```
ggplot(
  lfp_ind,
  aes(x = age, y = hr_worked, color = ind_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Average Hours Worked",
    color = "Industry Group",
    title = "Average Hours Worked by Age and Industry Group (1999-2017)",
  ) +
  theme_minimal()
```




```
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()
```



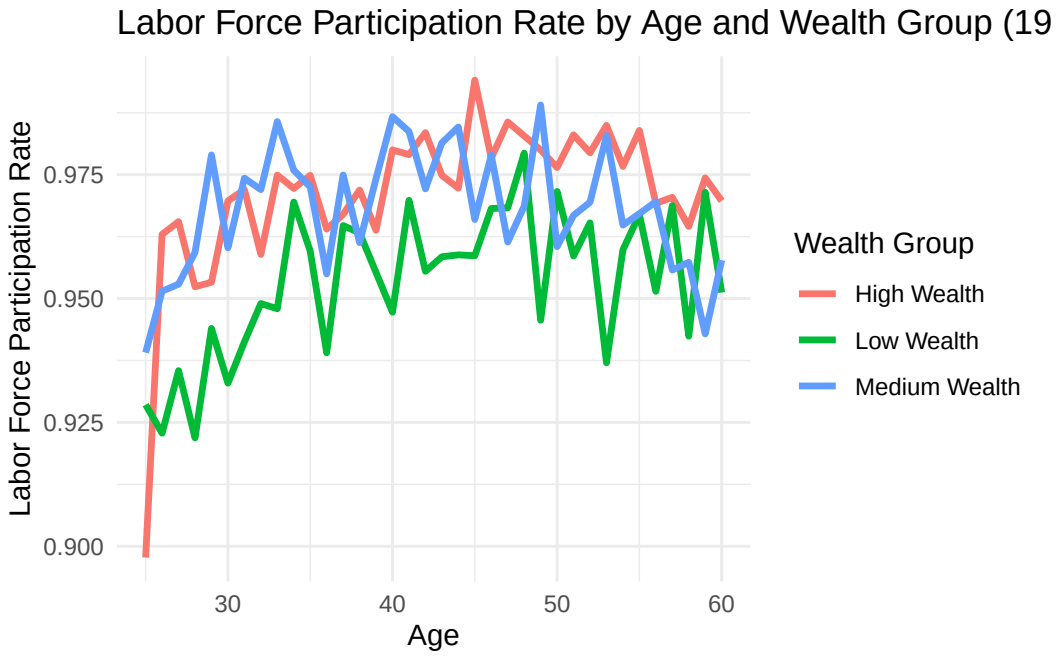

```
ggplot(
  lfp_ind,
  aes(x = age, y = var_log_hr_worked, color = ind_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Variance of Log Hours Worked",
    color = "Industry Group",
    title = "Variance of Log Hours Worked by Age and Industry Group (1999-2017)",
  ) +
  theme_minimal()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

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()
  )
```

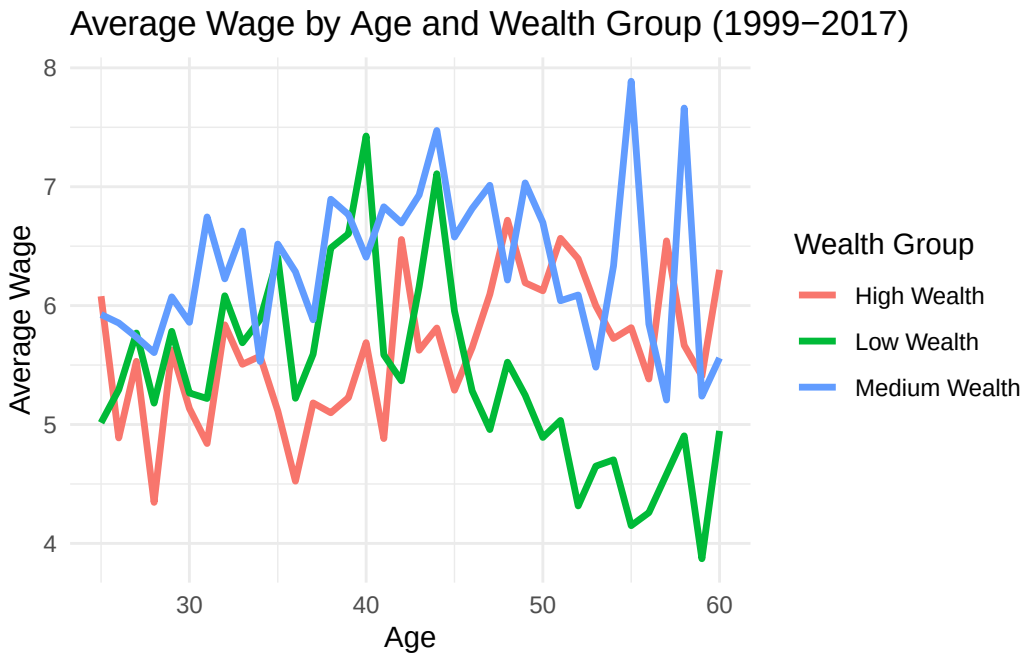
```
`summarise()` has grouped output by 'age'. You can override using the `.groups`
argument.
```

```
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()
```

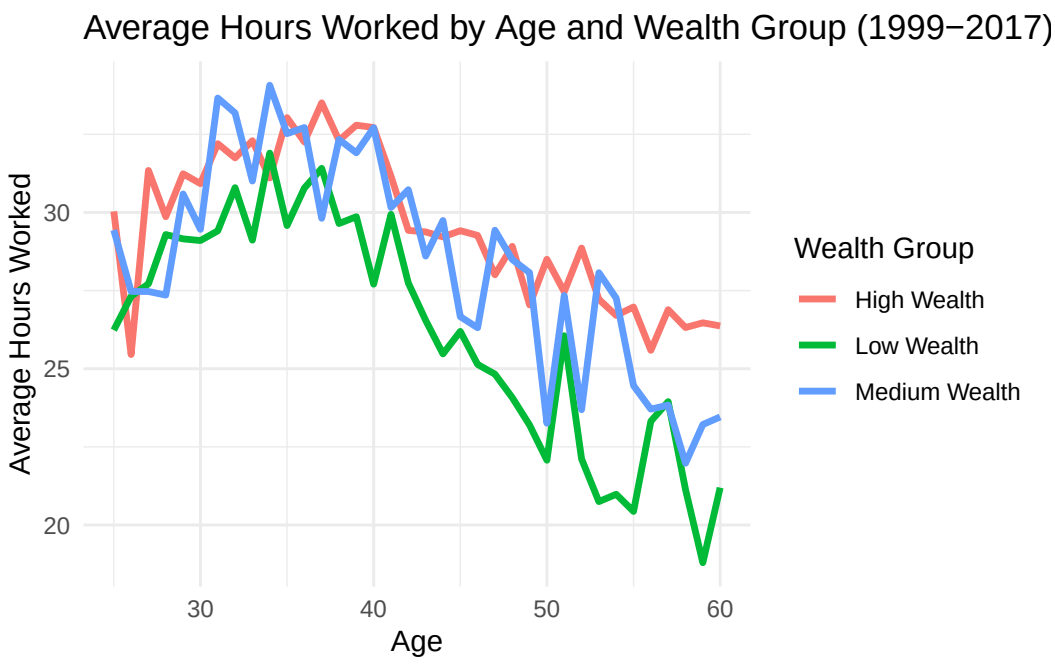



```
ggplot(
  lfp_wealth,
  aes(x = age, y = wage_rate, color = wealth_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Average Wage",
    color = "Wealth Group",
    title = "Average Wage by Age and Wealth Group (1999-2017)",
```

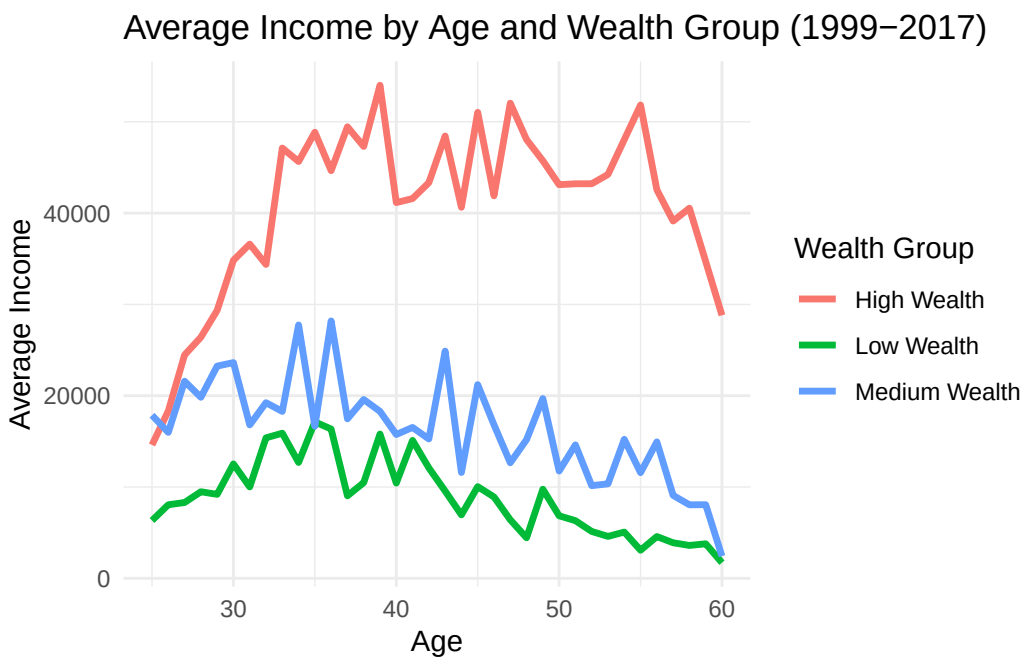
```
  ) +
  theme_minimal()
```



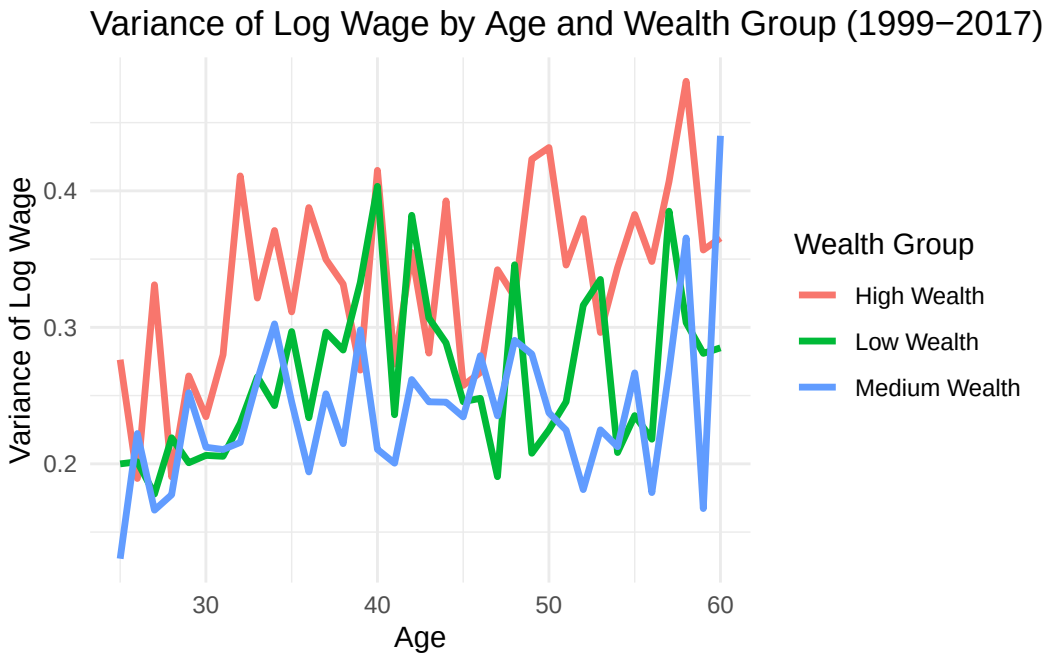

```
ggplot(
  lfp_wealth,
  aes(x = age, y = hr_worked, color = wealth_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Average Hours Worked",
    color = "Wealth Group",
    title = "Average Hours Worked by Age and Wealth Group (1999-2017)",
  ) +
  theme_minimal()
```



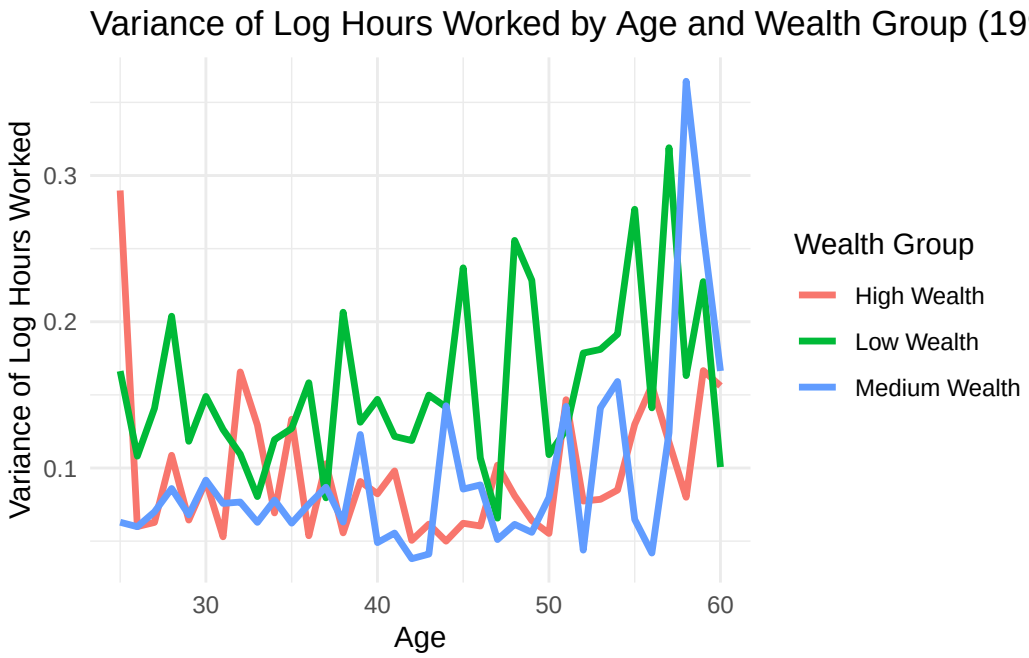

```
ggplot(
  lfp_wealth,
  aes(x = age, y = inc, color = wealth_group)
) +
  geom_line(size = 1.2) +
  labs(
    x = "Age",
    y = "Average Income",
    color = "Wealth Group",
    title = "Average Income by Age and Wealth Group (1999-2017)",
  ) +
  theme_minimal()
```




```
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()
```




```
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 (199
Variance of Log Hours Worked
0.3
0.2
0.1
30
40
50
60
Age
Wealth Group
High Wealth
Low Wealth
Medium Wealth