CDS101 Final Project

Tyson Johnson

2023-11-17

Summary Table

```
wages_tidy %>%
  group_by(demographic) %>%
  summarize(
    mean = mean(wages, na.rm=TRUE),
    median = median(wages, na.rm=TRUE),
    min = min(wages, na.rm=TRUE),
    max = max(wages, na.rm=TRUE),
    sd = sd(wages, na.rm=TRUE),
    iqr = IQR(wages, na.rm=TRUE)
  )
```

```
## # A tibble: 60 x 7
##    demographic              mean median   min   max    sd   iqr
##    <chr>                   <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
##  1 advanced_degree          43.9   44.1  35.3  53.7  5.31  8.80
##  2 bachelors_degree         34.8   34.2  30.0  41.6  3.31  5.25
##  3 black_advanced_degree    38.3   38.5  32.4  45.8  3.27  4.76
##  4 black_bachelors_degree   29.6   29.6  25.6  33.6  2.27  3.96
##  5 black_high_school        18.2   18.2  16.9  19.8 0.712 0.985
##  6 black_less_than_hs       14.5   14.3  13.1  16.4 0.803  1.21
##  7 black_men_advanced_degree 41.2   40.7  34.3  52.9  4.53  6.38
##  8 black_men_bachelors_degree 31.7   31.9  27.9  37.6  2.43  3.91
##  9 black_men_high_school    20.0   20.0  18.1  22.8  1.11  1.53
## 10 black_men_less_than_hs   16.0   15.5  14.1  18.6  1.29  1.97
## # i 50 more rows
```

```
wages_final <- wages_sep

wages_final$demographic <- factor(wages_final$demographic,
                                  levels = c("advanced_degree",
                                             "bachelors_degree",
                                             "some_college",
                                             "high_school",
                                             "less_than_hs"),
                                  labels = c("Advanced Degree",
                                             "Bachelors Degree",
                                             "Some College",
                                             "High School",
                                             "Less than High School"),
                                  ordered = TRUE)
```

```
na.exclude(wages_final) %>%
  ggplot() +
  geom_line(
    mapping = aes(
      x = year,
      y = wages,
      color = demographic
    )
  ) +
  facet_wrap(
    gender ~ ethnicity,
    ncol = 3
  ) +
  theme(
    axis.text.x = element_text(angle = 30)
  ) +
  labs(
    color = "Demographic"
  ) +
  scale_color_manual(
    values = c("red", "orange", "blue", "green", "purple")
  )
```
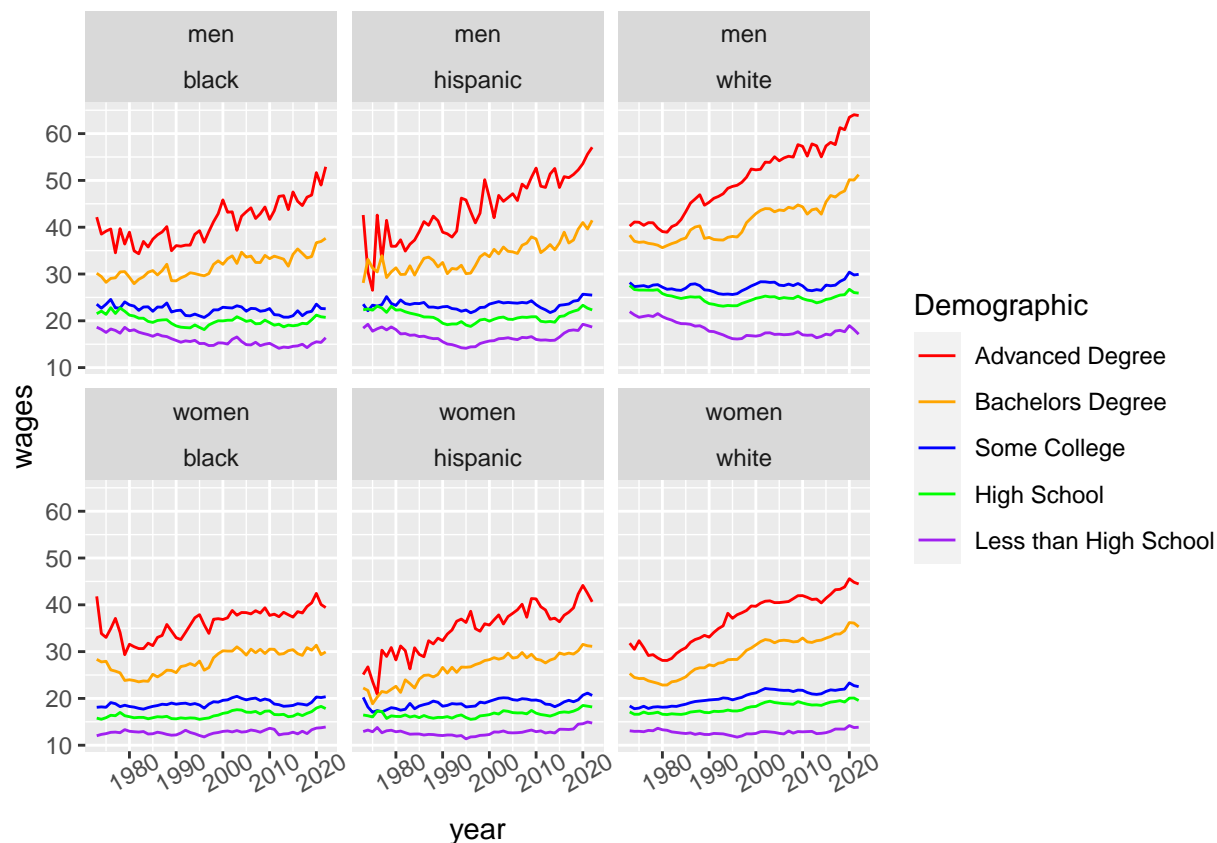
Assumptions for analysis

   i. Students with a bachelors or advanced degree will use 4-year school tuition rates."Some college" will use 2-year school tuition rates for two years.*

   ii. Due to "advanced degree" being broad, we will assume three years on average in addition to the four years required for a bachelors.

   iii. We will assume that individuals will pay off tuition at a rate of 10% of the total per year when they begin working. i.e., an individual who begins working in 1980 with a bachelors will pay 10% of the combined tuition of 1977, 1978, 1979 and 1980 every year, effectively reducing their income by that amount. This will mean all tuition will be paid off at the end of the 10th year of working.

   iv. We will use an estimated 2000 hours worked annually.

* Tuition data dates back to 1969, meaning an advanced degree holder's student debt in 1973 will use 4-year rates for 1969, 1969, 1969, 1970, 1971, 1972, 1973. This will affect the analysis of advanced degree holders from 1973 and 1974.

Advanced Degree Total Tuition

```
advanced_tuition_total <- tuition_fees %>%
  select(!two_year) %>%
  mutate(seven_year_total = 0)
```

```r
for(i in 1:54) {
  if(i %in% 1:4) {
    advanced_tuition_total$seven_year_total[i] = NA
  }
  else if(i == 5) {
    advanced_tuition_total$seven_year_total[i] =
      3 * tuition_fees$four_year[i-4] +
      tuition_fees$four_year[i-3] +
      tuition_fees$four_year[i-2] +
      tuition_fees$four_year[i-1] +
      tuition_fees$four_year[i]
  }
  else if(i == 6) {
    advanced_tuition_total$seven_year_total[i] =
      2 * tuition_fees$four_year[i-5] +
      tuition_fees$four_year[i-4] +
      tuition_fees$four_year[i-3] +
      tuition_fees$four_year[i-2] +
      tuition_fees$four_year[i-1] +
      tuition_fees$four_year[i]
  }
  else {
    advanced_tuition_total$seven_year_total[i] =
      tuition_fees$four_year[i-6] +
      tuition_fees$four_year[i-5] +
      tuition_fees$four_year[i-4] +
      tuition_fees$four_year[i-3] +
      tuition_fees$four_year[i-2] +
      tuition_fees$four_year[i-1] +
      tuition_fees$four_year[i]
  }
}
```

Bachelors Degree Total Tuition

```r
bachelors_tuition_total <- tuition_fees %>%
  select(!two_year) %>%
  mutate(four_year_total = 0)

for(i in 1:54) {
  if(i %in% 1:4) {
    bachelors_tuition_total$four_year_total[i] = NA
  }
  else {
    bachelors_tuition_total$four_year_total[i] =
      tuition_fees$four_year[i-3] +
      tuition_fees$four_year[i-2] +
      tuition_fees$four_year[i-1] +
      tuition_fees$four_year[i]
  }
}
```

Some College Total Tuition

```r
some_college_tuition_total <- tuition_fees %>%
  select(!four_year) %>%
  mutate(two_year_total = 0)

for(i in 1:54) {
  if(i %in% 1:4) {
    some_college_tuition_total$two_year_total[i] = NA
  }
  else {
    some_college_tuition_total$two_year_total[i] =
      tuition_fees$two_year[i-1] +
      tuition_fees$two_year[i]
  }
}
```

Advanced Degree Tuition Total

```r
advanced_tuition_total %>%
  ggplot() +
  geom_line(
    mapping = aes(
      x = year,
      y = seven_year_total
    ),
    na.rm = TRUE
  ) +
  labs(
    x = "Year",
    y = "Total Tuition",
    title = "Tuition Prices for Advanced Degrees"
  )
```

Tuition Prices for Advanced Degrees
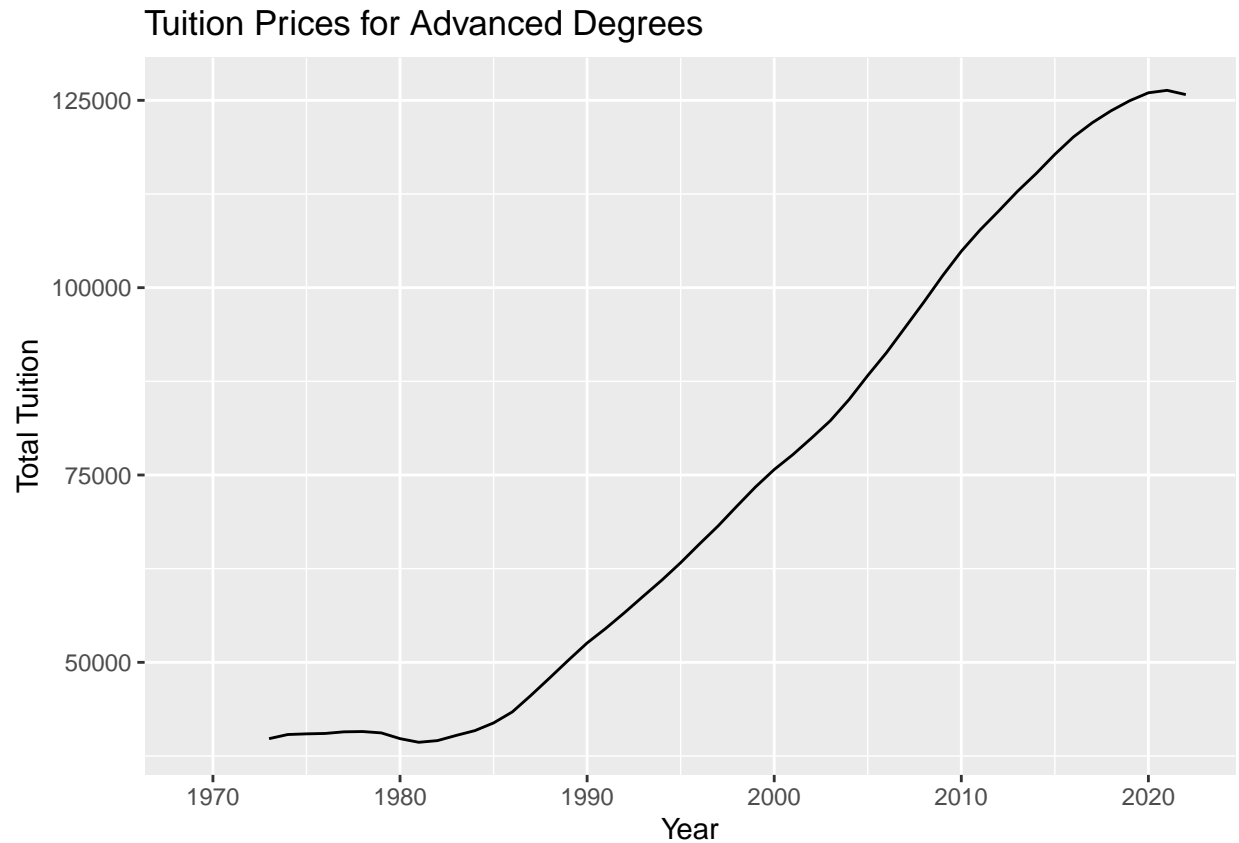
Advanced Degree Tuition-Adjusted Salary

```r
advanced_sig_years <- wages_sep %>%
  filter(demographic == "advanced_degree") %>%
  mutate(
    yearly_tuition_payments =
      advanced_tuition_total$seven_year_total[year-1968] * 0.1
  ) %>%
  mutate(
    salary = wages * 2000
  ) %>%
  mutate(
    adj_salary_ten_year = NA
  )

for (m in 1:591) {
  advanced_sig_years$adj_salary_ten_year[m] = 0
  if(advanced_sig_years$year[m] < advanced_sig_years$year[m+9]) {
    for (n in 0:9) {
      advanced_sig_years$adj_salary_ten_year[m] =
        advanced_sig_years$salary[m+n] +
        advanced_sig_years$adj_salary_ten_year[m]
    }
    advanced_sig_years$adj_salary_ten_year[m] =
```

```
      advanced_sig_years$adj_salary_ten_year[m] -
        (advanced_sig_years$yearly_tuition_payments[m] * 10)
  }
  else {
    advanced_sig_years$adj_salary_ten_year[m] = NA
  }
}
```

Bachelors Degree Tuition-Adjusted Salary

```
bachelors_sig_years <- wages_sep %>%
  filter(demographic == "bachelors_degree") %>%
  mutate(
    yearly_tuition_payments =
      bachelors_tuition_total$four_year_total[year-1968] * 0.1
  ) %>%
  mutate(
    salary = wages * 2000
  ) %>%
  mutate(
    adj_salary_thirteen_year = NA
  )

for (m in 1:588) {
  bachelors_sig_years$adj_salary_thirteen_year[m] = 0
  if(bachelors_sig_years$year[m] < bachelors_sig_years$year[m+12]) {
    for (n in 0:12) {
      bachelors_sig_years$adj_salary_thirteen_year[m] =
        bachelors_sig_years$salary[m+n] +
        bachelors_sig_years$adj_salary_thirteen_year[m]
    }
    bachelors_sig_years$adj_salary_thirteen_year[m] =
      bachelors_sig_years$adj_salary_thirteen_year[m] -
      (bachelors_sig_years$yearly_tuition_payments[m] * 10)
  }
  else {
    bachelors_sig_years$adj_salary_thirteen_year[m] = NA
  }
}
```

Merge Advanced and Bachelors for Graphing

```
adv_bac_sig_bind <- bind_rows(bachelors_sig_years, advanced_sig_years)

adv_bac_sig_graph <- adv_bac_sig_bind

adv_bac_sig_graph$demographic_order <- factor(
  adv_bac_sig_bind$demographic,
  levels = c("bachelors_degree", "advanced_degree"),
  labels = c("Bachelors Degree", "Advanced Degree"),
```

```
    ordered = TRUE
)

adv_bac_sig_graph$gender <-
  replace_na(adv_bac_sig_graph$gender, "coal")

adv_bac_sig_graph$gender_order <- factor(
  adv_bac_sig_graph$gender,
  levels = c("men", "women", "coal"),
  labels = c("Men", "Women", "Coalescent"),
  ordered = TRUE
)


adv_bac_sig_graph$ethnicity <-
  replace_na(adv_bac_sig_graph$ethnicity, "coal")

adv_bac_sig_graph$ethnicity_order <- factor(
  adv_bac_sig_graph$ethnicity,
  levels = c("black", "hispanic", "white", "coal"),
  labels = c("Black", "Hispanic", "White", "Coalescent"),
  ordered = TRUE
)
```

Advanced and Bachelors Graph

```
adv_sal_line <- geom_line(
    data = filter(adv_bac_sig_graph,
                  demographic_order == "Advanced Degree"
                  & !is.na(adj_salary_ten_year)),
    mapping = aes(
      x = year,
      y = adj_salary_ten_year,
      color = demographic_order
    )
  )

bac_sal_line <- geom_line(
    data = filter(adv_bac_sig_graph,
                  demographic_order == "Bachelors Degree"
                  & !is.na(adj_salary_thirteen_year)),
    mapping = aes(
      x = year,
      y = adj_salary_thirteen_year,
      color = demographic_order
    )
  )

graph1 <- ggplot() +
  adv_sal_line +
  bac_sal_line +
  facet_wrap(
    gender_order ~ ethnicity_order,
```

```r
    ncol = 4,
    scales = "free"
  ) +
  theme(
    axis.text.x = element_text(angle = 45),
    axis.text.y = element_text(angle = 45)
  ) +
  scale_y_continuous(
    labels = function(x) format(x, scientific = FALSE)
  ) +
  labs(
    x = "Year",
    y = "Earnings Minus Tuition",
    title = "Earnings Minus Tuition by Graduating Year",
    color = "Degree Level",
    caption = "*Bachelors holders are given 13 years (vs 10) to account for additional time in the work
  ) +
  scale_color_manual(
    values = c("Advanced Degree" = "red",
               "Bachelors Degree" = "blue"),
    labels = c("Advanced Degree\n(10-year earnings)",
               "Bachelors Degree\n(13-year earnings)")
  )

graph1
```

Earnings Minus Tuition by Graduating Year

*Bachelors holders are given 13 years (vs 10) to account for additional time in the workforce.
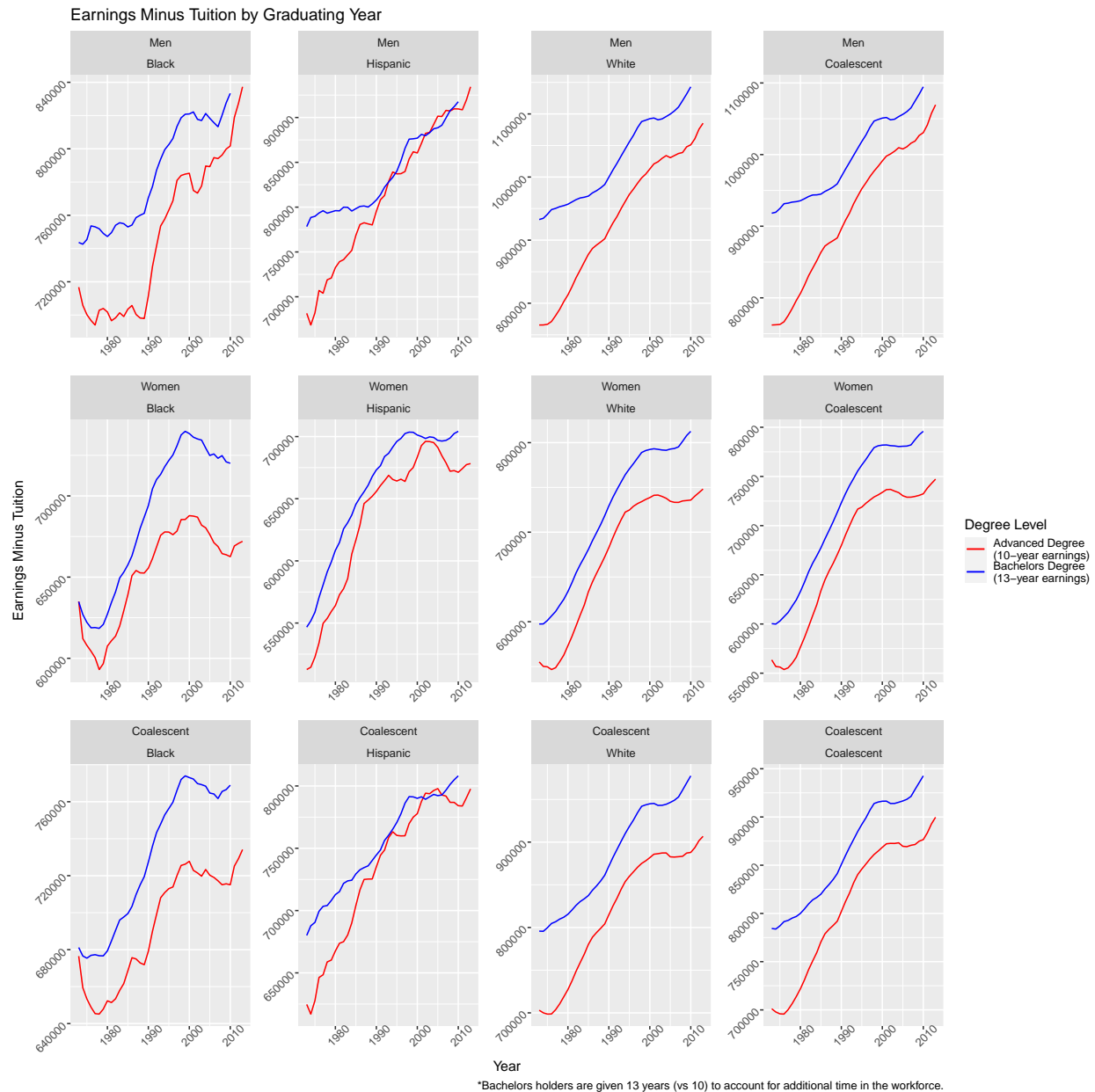
Some College Tuition-Adjusted Salary

```r
some_college_sig_years <- wages_sep %>%
  filter(demographic == "some_college") %>%
  mutate(
    yearly_tuition_payments =
      some_college_tuition_total$two_year_total[year-1968] * 0.1
  ) %>%
  mutate(
    salary = wages * 2000
  ) %>%
  mutate(
```

```
    adj_salary_fifteen_year = NA
  )

for (m in 1:586) {
  some_college_sig_years$adj_salary_fifteen_year[m] = 0
  if(some_college_sig_years$year[m] < some_college_sig_years$year[m+14]) {
    for (n in 0:14) {
      some_college_sig_years$adj_salary_fifteen_year[m] =
        some_college_sig_years$salary[m+n] +
        some_college_sig_years$adj_salary_fifteen_year[m]
    }
    some_college_sig_years$adj_salary_fifteen_year[m] =
      some_college_sig_years$adj_salary_fifteen_year[m] -
      (some_college_sig_years$yearly_tuition_payments[m] * 10)
  }
  else {
    some_college_sig_years$adj_salary_fifteen_year[m] = NA
  }
}
```

High School and Less Than High School Earnings

```
hs_sig_years <- wages_sep %>%
  filter(demographic %in% c("high_school", "less_than_hs")) %>%
  mutate(
    yearly_tuition_payments = 0
  ) %>%
  mutate(
    salary = wages * 2000
  ) %>%
  mutate(
    adj_salary_seventeen_year = NA
  )

for (m in 1:1184) {
  hs_sig_years$adj_salary_seventeen_year[m] = 0
  if(hs_sig_years$year[m] < hs_sig_years$year[m+16]) {
    for (n in 0:16) {
      hs_sig_years$adj_salary_seventeen_year[m] =
        hs_sig_years$salary[m+n] +
        hs_sig_years$adj_salary_seventeen_year[m]
    }
    hs_sig_years$adj_salary_seventeen_year[m] =
      hs_sig_years$adj_salary_seventeen_year[m] -
      (hs_sig_years$yearly_tuition_payments[m] * 10)
  }
  else {
    hs_sig_years$adj_salary_seventeen_year[m] = NA
  }
}
```

```
some_college_hs_sig_bind <- bind_rows(hs_sig_years, some_college_sig_years)

some_college_hs_sig_graph <- some_college_hs_sig_bind

some_college_hs_sig_graph$demographic_order <- factor(
  some_college_hs_sig_bind$demographic,
  levels = c("less_than_hs", "high_school", "some_college"),
  labels = c("Less than High School", "High School", "Some College"),
  ordered = TRUE
)

some_college_hs_sig_graph$gender <-
  replace_na(some_college_hs_sig_graph$gender, "coal")

some_college_hs_sig_graph$gender_order <- factor(
  some_college_hs_sig_graph$gender,
  levels = c("men", "women", "coal"),
  labels = c("Men", "Women", "Coalescent"),
  ordered = TRUE
)

some_college_hs_sig_graph$ethnicity <-
  replace_na(some_college_hs_sig_graph$ethnicity, "coal")

some_college_hs_sig_graph$ethnicity_order <- factor(
  some_college_hs_sig_graph$ethnicity,
  levels = c("black", "hispanic", "white", "coal"),
  labels = c("Black", "Hispanic", "White", "Coalescent"),
  ordered = TRUE
)
```

Some College, High School, Less than High School Graph

```
some_college_sal_line <- geom_line(
    data = filter(some_college_hs_sig_graph,
                  demographic_order == "Some College"
                  & !is.na(adj_salary_fifteen_year)
    ),
    mapping = aes(
      x = year,
      y = adj_salary_fifteen_year,
      color = demographic_order
    )
  )

high_school_sal_line <- geom_line(
    data = filter(some_college_hs_sig_graph,
                  demographic_order == "High School"
                  & !is.na(adj_salary_seventeen_year)
    ),
    mapping = aes(
      x = year,
```

```
      y = adj_salary_seventeen_year,
      color = demographic_order
    )
  )

less_than_hs_sal_line <- geom_line(
    data = filter(some_college_hs_sig_graph,
                  demographic_order == "Less than High School"
                  & !is.na(adj_salary_seventeen_year)
    ),
    mapping = aes(
      x = year,
      y = adj_salary_seventeen_year,
      color = demographic_order
    )
  )

graph2 <- ggplot() +
  some_college_sal_line +
  high_school_sal_line +
  less_than_hs_sal_line +
  facet_wrap(
    gender_order ~ ethnicity_order,
    ncol = 4,
    scales = "free"
  ) +
  theme(
    axis.text.x = element_text(angle = 45),
    axis.text.y = element_text(angle = 45)
  ) +
  scale_y_continuous(
    labels = function(x) format(x, scientific = FALSE)
  ) +
  labs(
    x = "Year",
    y = "Earnings Minus Tuition",
    title = "Earnings Minus Tuition by Graduating Year",
    color = "Degree Level",
    caption = "*Additional years are given to account for additional time in the workforce."
  ) +
  scale_color_manual(
    values = c("green",
               "purple",
               "orange"),
    labels = c(
      "Some College\n(15-year earnings)",
      "High School\n(17-year earnings)",
      "Less than High School\n(17-year earnings)"
    ),
    breaks = c("Some College", "High School", "Less than High School")
  )

graph2
```

Earnings Minus Tuition by Graduating Year

*Additional years are given to account for additional time in the workforce.
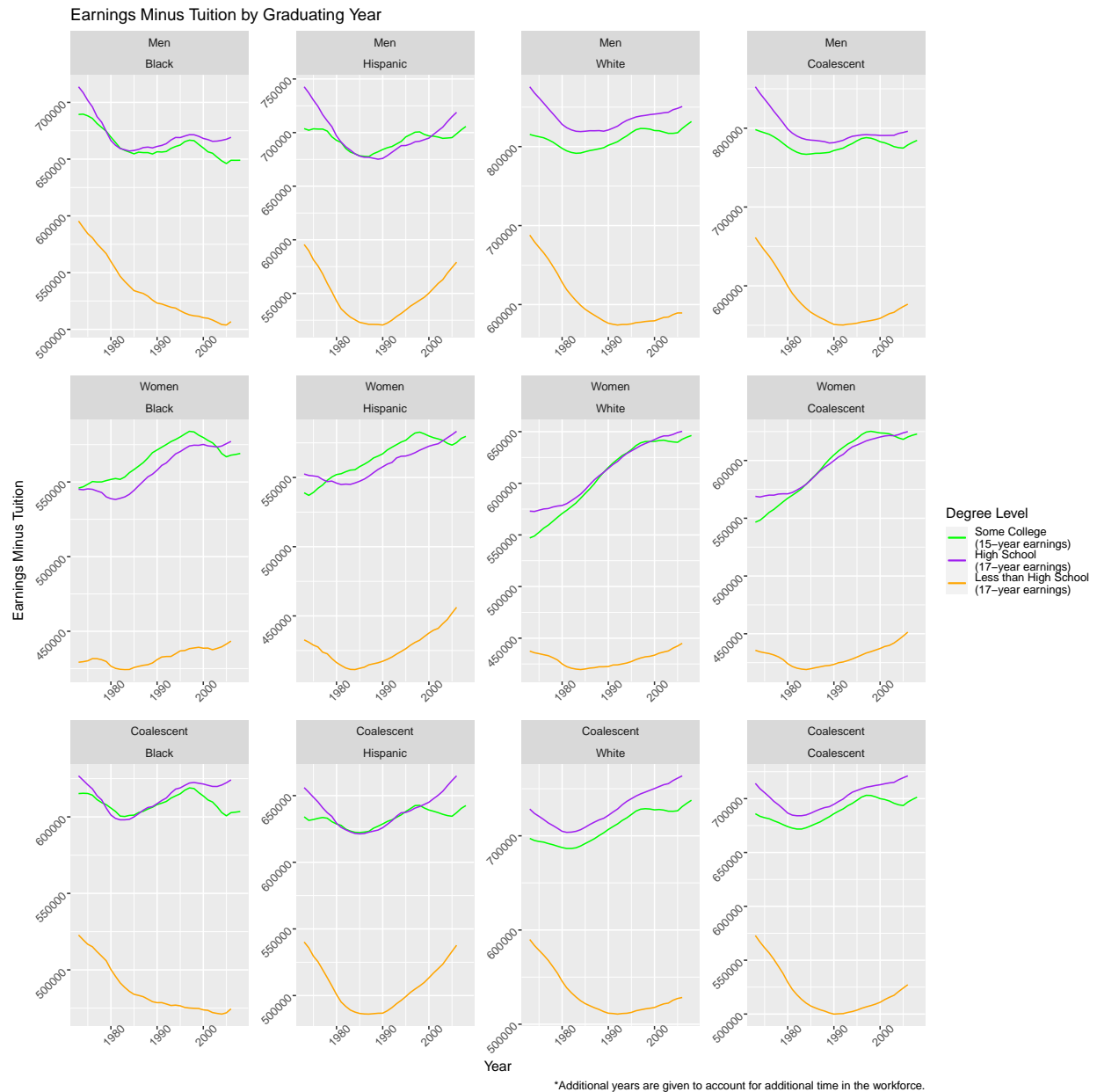
Combined Graph

```
combined_graph <- ggplot() +
  adv_sal_line +
  bac_sal_line +
  some_college_sal_line +
  high_school_sal_line +
  less_than_hs_sal_line +
  facet_wrap(
    gender_order ~ ethnicity_order,
    ncol = 4,
    scales = "free"
```

```
  ) +
  scale_y_continuous(
    labels = function(x) format(x, scientific = FALSE)
  ) +
  labs(
    x = "Year",
    y = "Earnings Minus Tuition",
    title = "Earnings Minus Tuition by Graduating Year",
    color = "Degree Level",
    caption = "*Additional years are given to account for additional time in the workforce."
  ) +
  scale_color_manual(
    values = c("Advanced Degree" = "red",
               "Bachelors Degree" = "blue",
               "Some College" = "green",
               "High School" = "purple",
               "Less than High School" = "orange"),
    labels = c(
      "Advanced Degree\n(10-year earnings)",
      "Bachelors Degree\n(13-year earnings)",
      "Some College\n(15-year earnings)",
      "High School\n(17-year earnings)",
      "Less than High School\n(17-year earnings)"
    ),
    breaks = c(
      "Advanced Degree",
      "Bachelors Degree",
      "Some College",
      "High School",
      "Less than High School"
    )
  )

combined_graph
```

Earnings Minus Tuition by Graduating Year

*Additional years are given to account for additional time in the workforce.

Degree Level

- Advanced Degree (10–year earnings)
- Bachelors Degree (13–year earnings)
- Some College (15–year earnings)
- High School (17–year earnings)
- Less than High School (17–year earnings)

16