# Final Project

Group 1

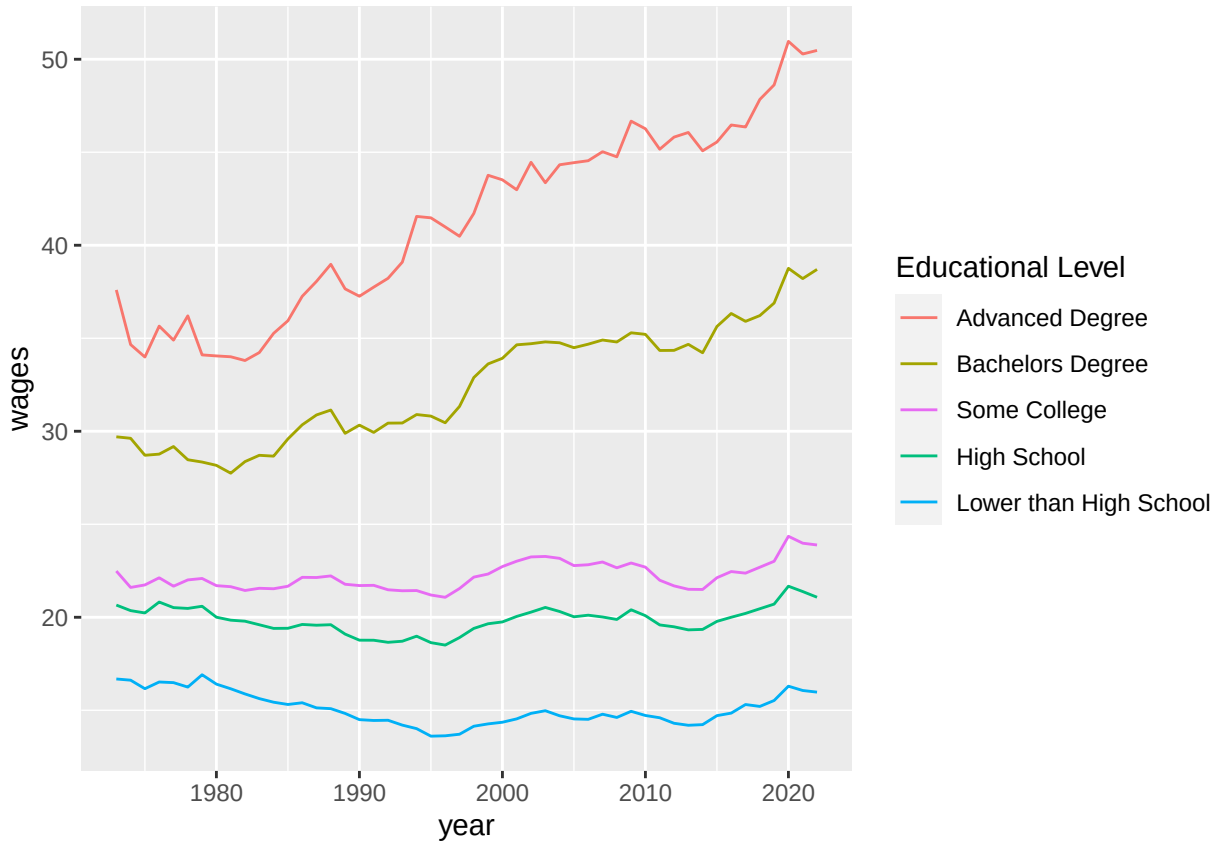
2023-11-17

##Junyoung

```
overall_income <- read.csv("wages_sep.csv") %>%
  select(year, demographic, wages)
```

```
aggregated_overall_income <- aggregate(wages ~ year + demographic, data = overall_income, FUN = mean)
```

```
library(ggplot2)
aggregated_overall_income %>%
  ggplot() +
  geom_line(
    mapping = aes(
      x = year,
      y = wages,
      color = demographic
    )) + scale_color_discrete(name = "Educational Level",
                              breaks =c("advanced_degree",
                                        "bachelors_degree",
                                        "some_college",
                                        "high_school",
                                        "less_than_hs"),
                              labels = c("Advanced Degree",
                                         "Bachelors Degree",
                                         "Some College",
                                         "High School",
                                         "Lower than High School"))
```




## Areum

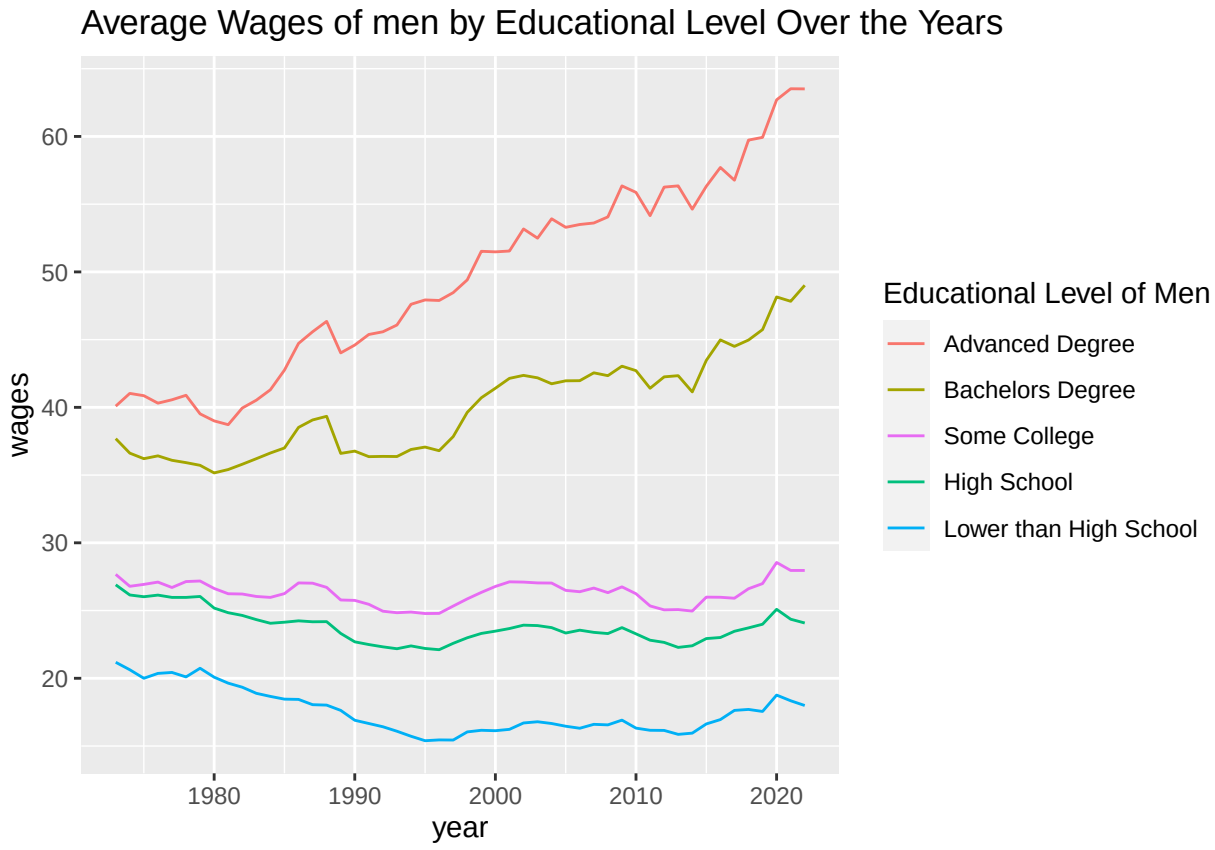
```
wbe_men <- wages_by_education %>%
  select(year,men_less_than_hs:men_advanced_degree)
```

```
wbe_women <- wages_by_education %>%
  select(year,women_less_than_hs:women_advanced_degree)
```

```
wbe_men2 <-wbe_men %>%
  pivot_longer(cols =2:men_advanced_degree, names_to = 'educational_level',
               values_to = 'wages' )
```

```
wbe_men2 %>%
  ggplot()+
  geom_line(
    mapping = aes(x = year, y = wages,
                  color = educational_level)
  )+
  scale_color_discrete(name = "Educational Level of Men",
                       breaks =c("men_advanced_degree",
                                 "men_bachelors_degree",
                                 "men_some_college",
                                 "men_high_school",
```

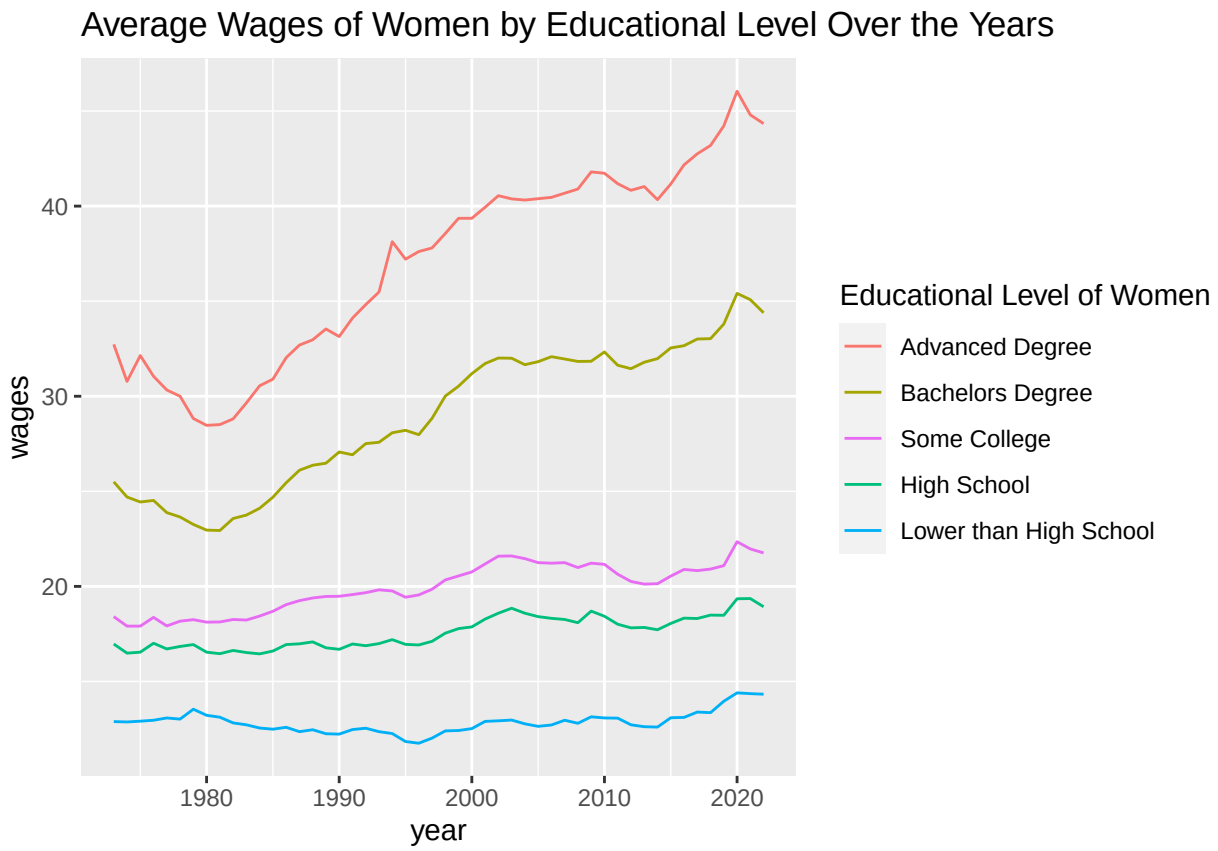
```
                                "men_less_than_hs"),
                     labels = c("Advanced Degree",
                                "Bachelors Degree",
                                "Some College",
                                "High School",
                                "Lower than High School"))+
  labs(title="Average Wages of men by Educational Level Over the Years"
  )
```




```
wbe_women2 <-wbe_women %>%
  pivot_longer(cols =2:women_advanced_degree, names_to = 'educational_level',
               values_to = 'wages' )
```

```
wbe_women2 %>%
  ggplot()+
  geom_line(
    mapping = aes(x = year, y = wages,
                  color = educational_level))+
  scale_color_discrete(name = "Educational Level of Women",
                       breaks =c("women_advanced_degree",
                                 "women_bachelors_degree",
                                 "women_some_college",
                                 "women_high_school",
                                 "women_less_than_hs"),
                       labels = c("Advanced Degree",
```

```
                                "Bachelors Degree",
                                "Some College",
                                "High School",
                                "Lower than High School"))+
  labs(title="Average Wages of Women by Educational Level Over the Years"
  )
```




## Ikjoo

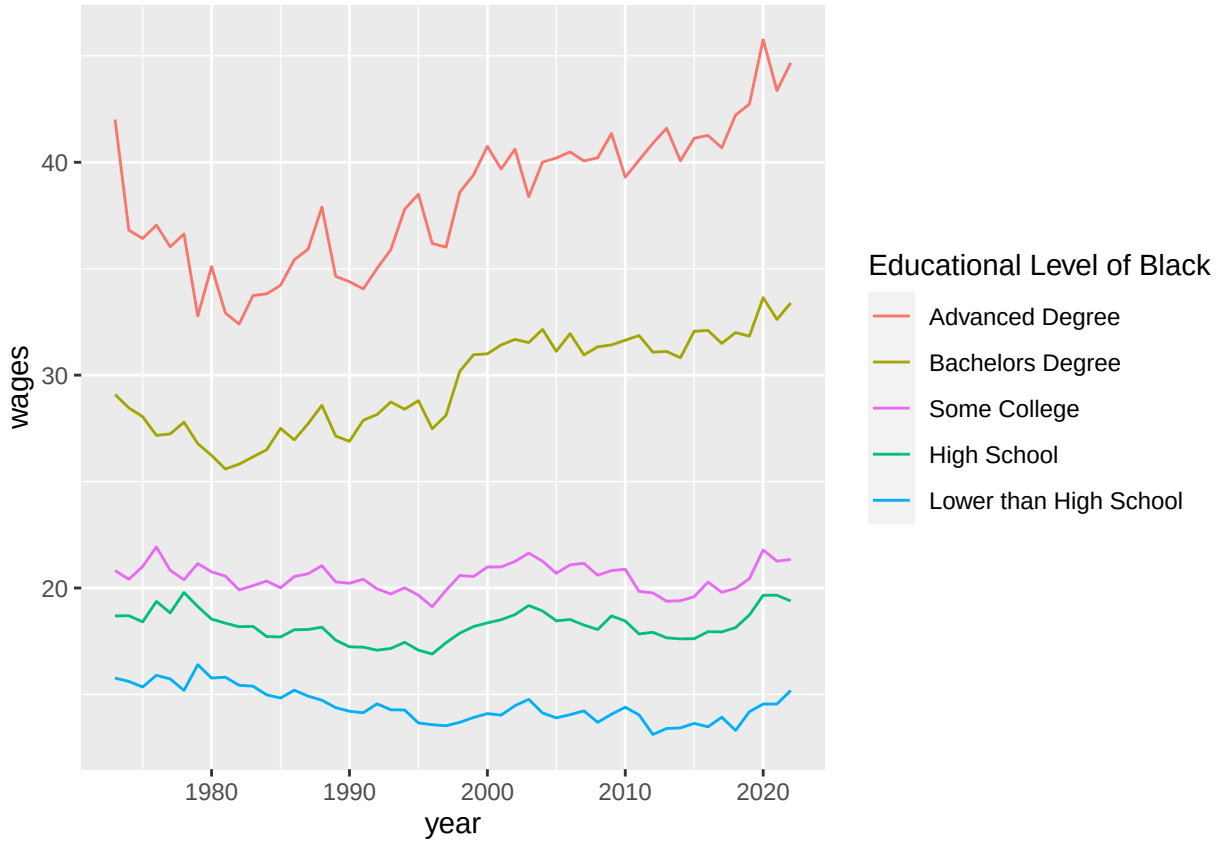
```
# Data of White people
wbe_white <- wages_by_education %>%
  select(year,white_less_than_hs:white_advanced_degree)
```

```
# Data of Black people
wbe_black <- wages_by_education %>%
  select(year,black_less_than_hs:black_advanced_degree)
```

```
# Data of Hispanic People
wbe_hispanic <- wages_by_education %>%
  select(year,hispanic_less_than_hs:hispanic_advanced_degree)
```

```
# Black people divided in their level of education
wbe_black_education <-wbe_black %>%
pivot_longer(cols =2:black_advanced_degree, names_to = 'educational_level',
             values_to = 'wages')
```

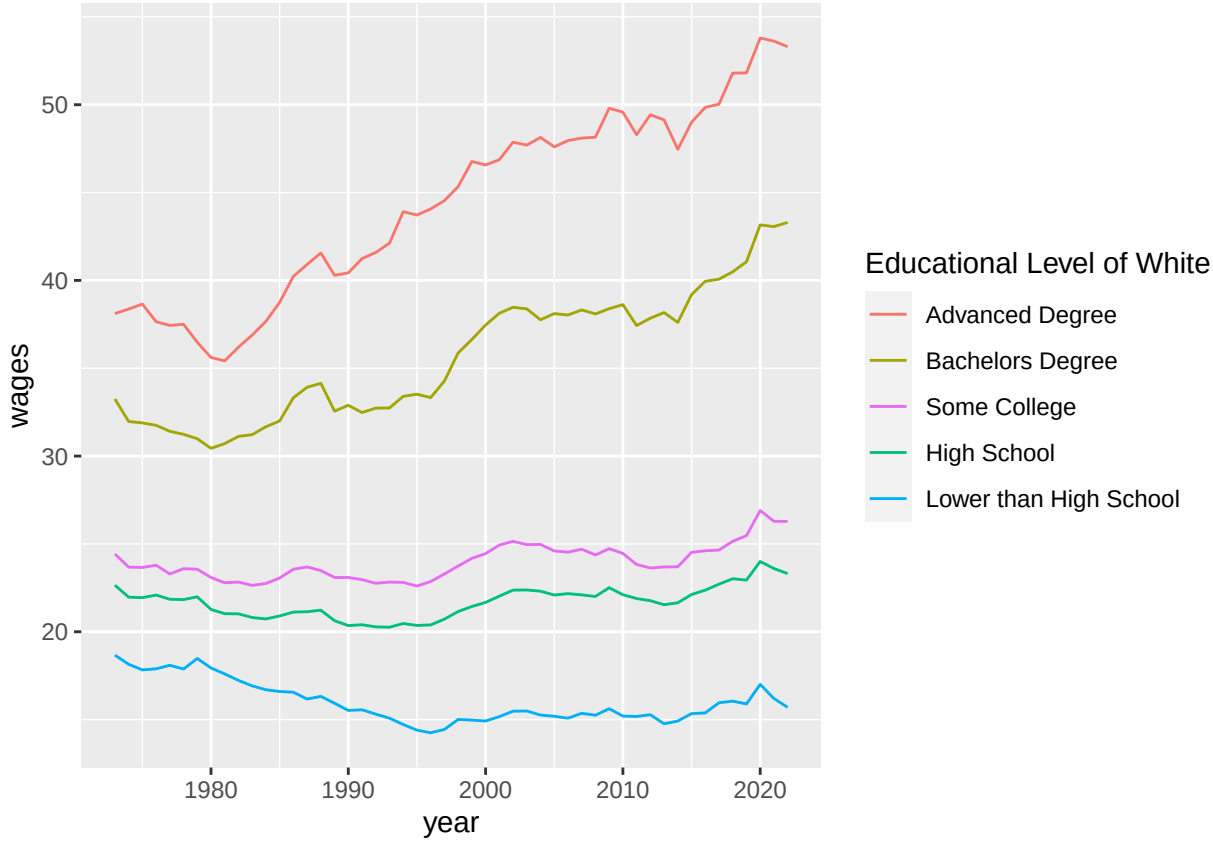
```
wbe_black_education %>%
  ggplot()+
  geom_line(mapping = aes(x = year,
                          y = wages,
                          color = educational_level)) +
  scale_color_discrete(name = "Educational Level of Black",
                       breaks =c("black_advanced_degree",
                                 "black_bachelors_degree",
                                 "black_some_college",
                                 "black_high_school",
                                 "black_less_than_hs"),
                       labels = c("Advanced Degree",
                                  "Bachelors Degree",
                                  "Some College",
                                  "High School",
                                  "Lower than High School"))
```




```
# White people divided in their level of education
wbe_white_education <-wbe_white %>%
```

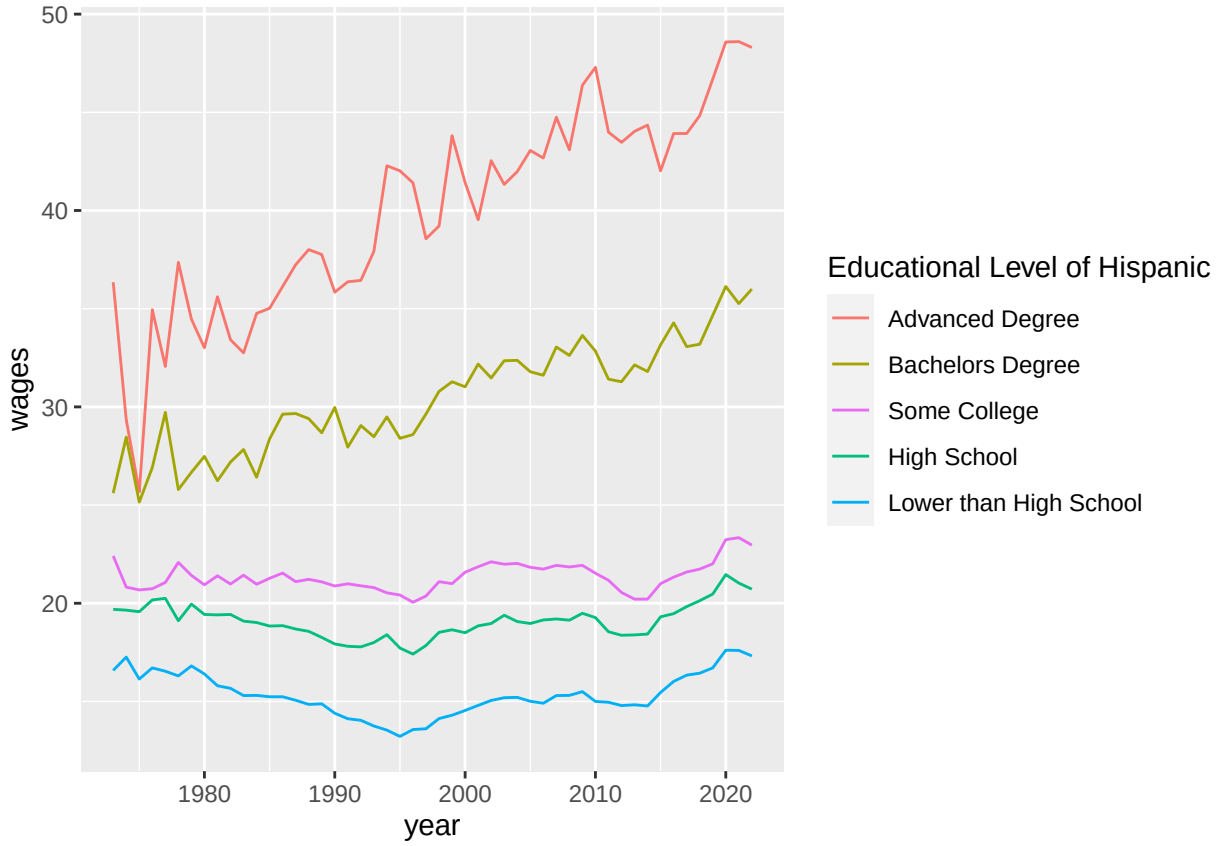
```
pivot_longer(cols =2:white_advanced_degree, names_to = 'educational_level',
             values_to = 'wages')
```

```
wbe_white_education %>%
  ggplot()+
  geom_line(mapping = aes(x = year,
                          y = wages,
                          color = educational_level))+
              scale_color_discrete(name = "Educational Level of White",
                                   breaks =c("white_advanced_degree",
                                             "white_bachelors_degree",
                                             "white_some_college",
                                             "white_high_school",
                                             "white_less_than_hs"),
                                   labels = c("Advanced Degree",
                                              "Bachelors Degree",
                                              "Some College",
                                              "High School",
                                              "Lower than High School"))
```




```
# Hispanic people divided in their level of education
wbe_hispanic_education <-wbe_hispanic %>%
pivot_longer(cols =2:hispanic_advanced_degree, names_to
             ='educational_level',
             values_to = 'wages')
```

```
wbe_hispanic_education %>%
  ggplot()+
  geom_line(mapping = aes(x = year,
                          y = wages,
                          color = educational_level))+
  scale_color_discrete(name = "Educational Level of Hispanic",
                                breaks =c("hispanic_advanced_degree",
                                          "hispanic_bachelors_degree",
                                          "hispanic_some_college",
                                          "hispanic_high_school",
                                          "hispanic_less_than_hs"),
                                labels = c("Advanced Degree",
                                           "Bachelors Degree",
                                           "Some College",
                                           "High School",
                                           "Lower than High School"))
```




## Tyson

```
wages_ethnicity <- wages_tidy %>%
 filter(grepl(c("black|hispanic|white"), demographic)) %>%
 separate(demographic, into = c("ethnicity", "demographic"),
          sep = "_", extra = "merge") %>%
```

```
 arrange(ethnicity, demographic, year)

wages_ethnicity_inverse <- wages_tidy %>%
 filter(!grepl(c("black|hispanic|white"), demographic)) %>%
 mutate(ethnicity = NA) %>%
 relocate(ethnicity, .after = year)
wages_ethnicity_combined <- wages_ethnicity %>%
 bind_rows(wages_ethnicity_inverse) %>%
 arrange(ethnicity, demographic, year)
```

```
wages_gender <- wages_ethnicity_combined %>%
 filter(grepl(c("men|women"), demographic)) %>%
 separate(demographic, into = c("gender", "demographic"),
          sep = "_", extra = "merge") %>%
 arrange(ethnicity, gender, demographic, year)
wages_gender_inverse <- wages_ethnicity_combined %>%
 filter(!grepl(c("men|women"), demographic)) %>%
 mutate(gender = NA) %>%
 relocate(gender, .after = ethnicity)
wages_gender_combined <- wages_gender %>%
 bind_rows(wages_gender_inverse) %>%
 arrange(ethnicity, gender, demographic, year)
```

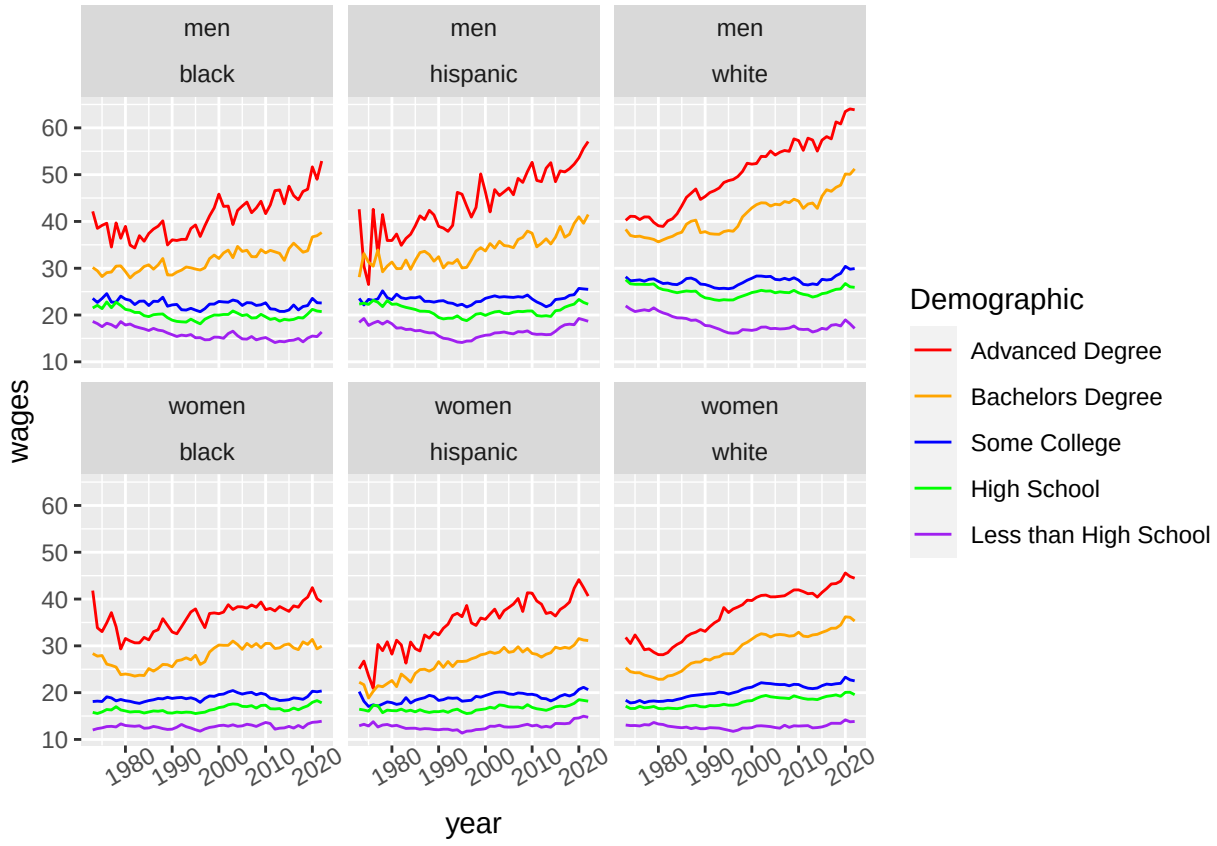
```
wages_sep <- wages_gender_combined
wages_sep %>%
 write.csv("wages_sep.csv", row.names=FALSE)
```

```
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")
  )
```




## 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
```