Labor Market Analysis Task

Anthony Flores

1/6/2022

```
# Load Necessary Packages
if (!require("pacman")) install.packages("pacman")
p_load(
    "tidyverse",
    "haven",
    # Data Visualization
    "ggthemes",
    "scales",
    # For working with date-times
    "lubridate",
    # Weighted Functions
    "Hmisc",
    "weights")
```

```
# Import Data
data <- read_dta(file = "../input/cps_wages_LFP.dta")
```

Introduction

Key Trends for Skilled and Unskilled Workers, 1976 - 2015

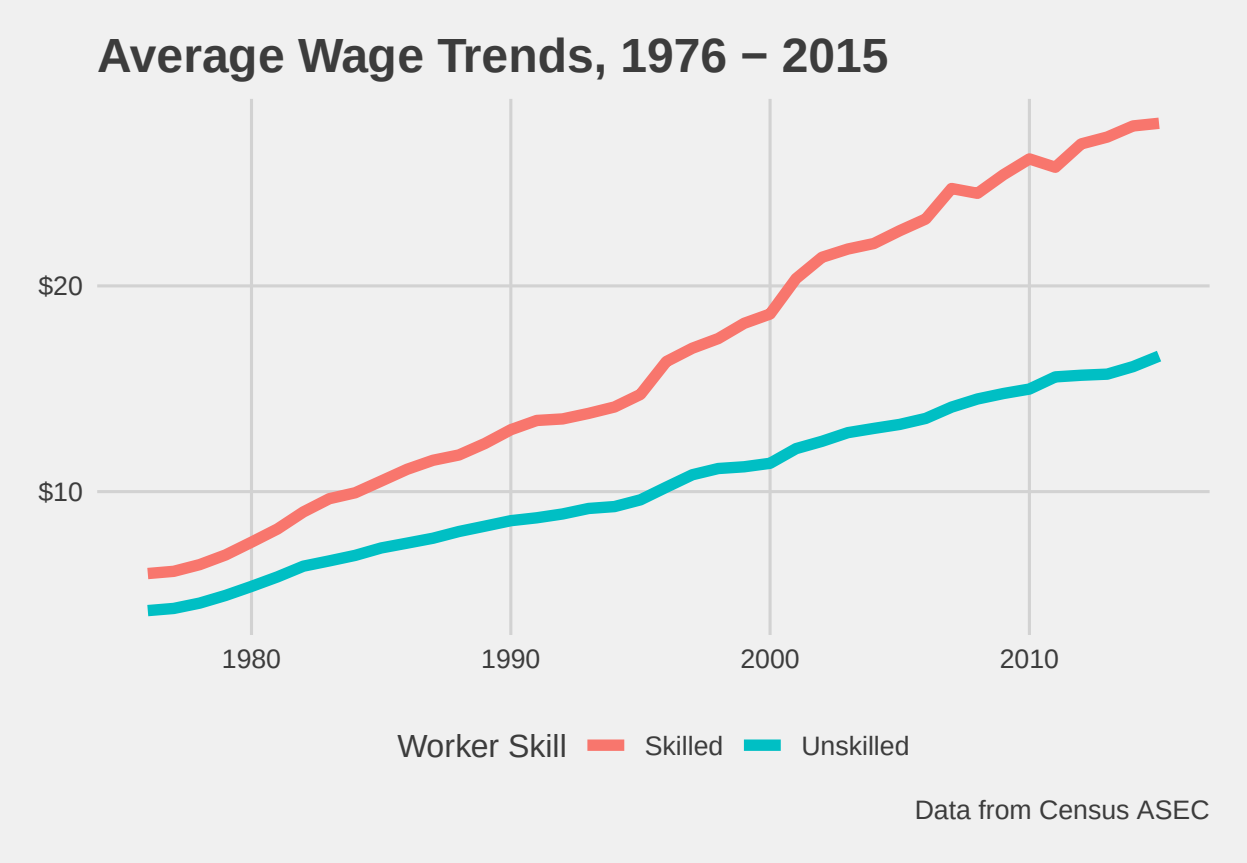
The following results reveal that the wage gap between skilled and unskilled workers has been steadily increasing over the period. The gap in labor force participation between the two skill groups also increased substantially but not as steadily. Most of the increase in the gap occurred in the 90s.

```
compute_trends <- function(data) {
    data %>%
    group_by(year) %>%
    summarise(avg_wage = wtd.mean(wage, weights = wtsupp, na.rm = TRUE),
              lfp = sum(wtsupp * lfp, na.rm = TRUE) /
                    sum(wtsupp[!is.na(lfp)])
              ) %>%
    mutate(date = make_datetime(year))
}
trend_data <-
    # Skilled Workers
    data %>%
    filter(skilled == 1) %>% # Education is high school diploma equivalent or higher.
    compute_trends() %>%
```

```
    mutate(skill = "Skilled") %>%
    # Unskilled Workers
    rbind(data %>%
          # Education is below high school diploma equivalent.
          filter(skilled != 1) %>%
          compute_trends() %>%
          mutate(skill = "Unskilled"))
```

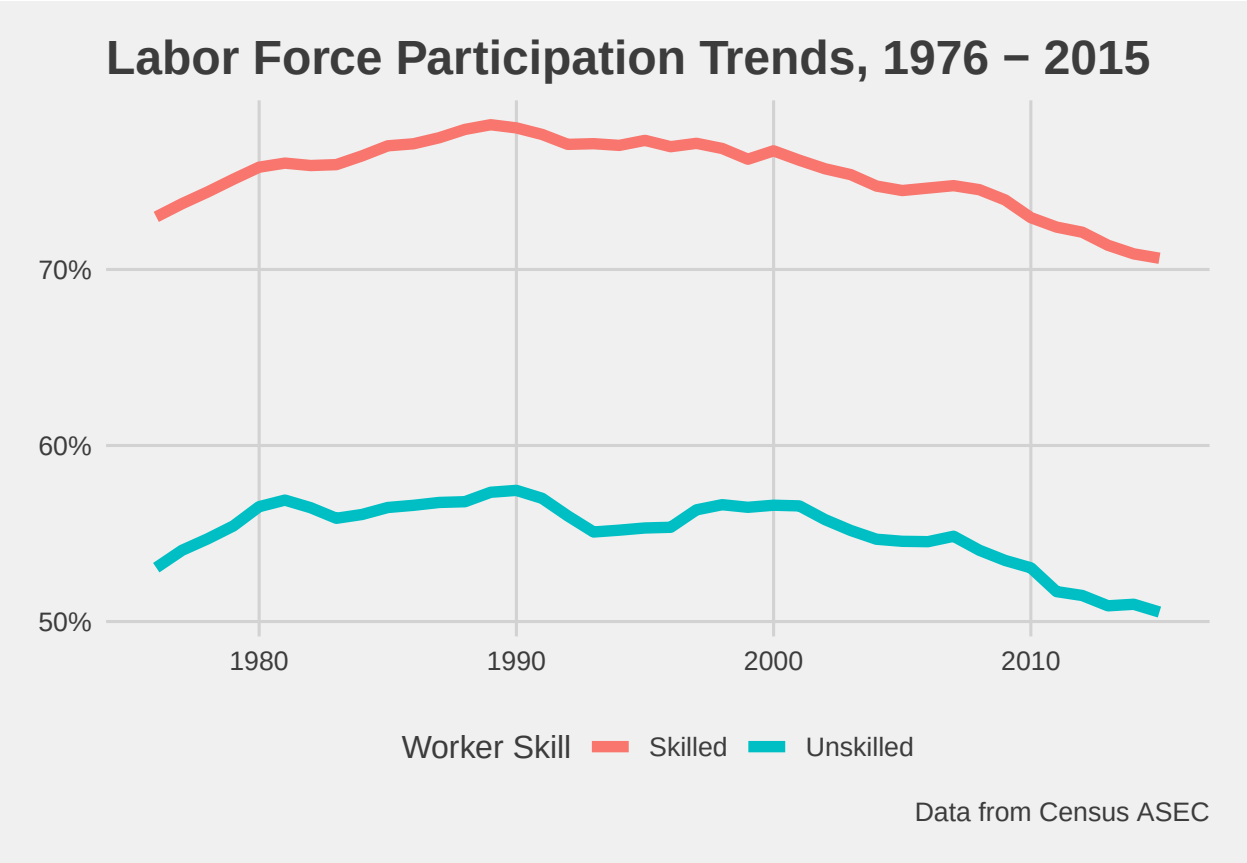
Wage Trends

```
trend_data %>%
    ggplot(aes(x = date, y = avg_wage)) +
    geom_line(aes(color = skill, group = skill), size = 2) +
    labs(title = "Average Wage Trends, 1976 - 2015", color = "Worker Skill",
         caption = "Data from Census ASEC") +
    scale_y_continuous(labels = label_dollar()) +
    theme_fivethirtyeight()
```



Labor Force Participation (LFP) Trends

```
trend_data %>%
    ggplot(aes(x = date, y = lfp)) +
    geom_line(aes(color = skill, group = skill), size = 2) +
    labs(title = "Labor Force Participation Trends, 1976 - 2015", color = "Worker Skill",
         caption = "Data from Census ASEC") +
    scale_y_continuous(labels = label_percent()) +
    theme_fivethirtyeight()
```



Labor Force Participation Trends Among Men Older than 25

The following analysis examines labor force participation trends among men older than 25 in the following subgroups: (1) unskilled and skilled people; (2) people aged 25 and above but below 45, those aged 45 and above but below 65, and those aged 65 and above; (3) non-hispanic white, non-hispanic black, and hispanic people. For those interested, I have also added a comparison of LFP for men and women older than 25.

Our results show that the 1976 - 2015 period, mature men in the skilled subgroup experienced the largest change in LFP, a decrease of 12.4 percentage points. It is also of note that women experienced an even larger change in LFP, an increase of 13.2 pp.

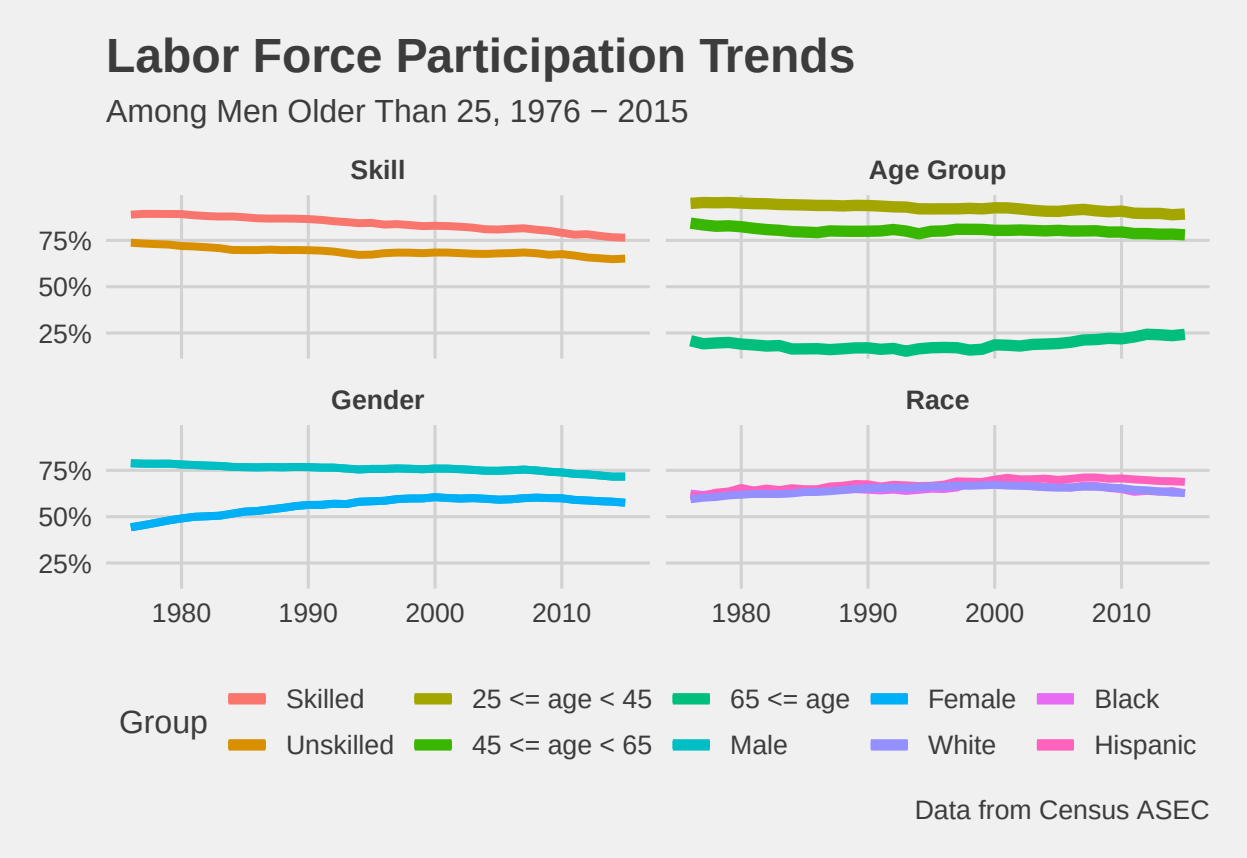
```
men_olderthan25_trend_data <-
    # Skilled Workers Over 25 Years Old
    data %>%
    filter(skilled == 1 & sex == 1 & age_group != 0) %>%
    compute_trends() %>%
```

```
    mutate(skill = "Skilled", chart_group = 1) %>%
    bind_rows(
        # Unskilled Workers Over 25 Years Old
        data %>%
          filter(skilled != 1 & sex == 1 & age_group != 0) %>%
          compute_trends() %>%
          mutate(skill = "Unskilled", chart_group = 1),
        # 25 <= age < 45
        data %>%
          filter(sex == 1 & age_group == 1) %>%
          compute_trends() %>%
          mutate(age_group = "25 <= age < 45", chart_group = 2),
        # 45 <= age < 65
        data %>%
          filter(sex == 1 & age_group == 2) %>%
          compute_trends() %>%
          mutate(age_group = "45 <= age < 65", chart_group = 2),
        # 65 <= age
        data %>%
          filter(sex == 1 & age_group == 3) %>%
          compute_trends() %>%
          mutate(age_group = "65 <= age", chart_group = 2),
        # Male Workers
        data %>%
          filter(sex == 1 & age_group != 0 ) %>%
          compute_trends() %>%
          mutate(gender = "Male", chart_group = 3),
        # Female Workers
        data %>%
          filter(sex == 2 & age_group != 0 ) %>%
          compute_trends() %>%
          mutate(gender = "Female", chart_group = 3),
        # Non-Hispanic White Workers
        data %>%
          filter(race == 100 & hispan == 0 & age > 25) %>%
          compute_trends() %>%
          mutate(race = "White", chart_group = 4),
        # Non-Hispanic Black Workers
        data %>%
          filter(race == 200 & hispan == 0 & age > 25) %>%
          compute_trends() %>%
          mutate(race = "Black", chart_group = 4),
        # Hispanic Workers
        data %>%
          filter(hispan %in% 100:500 & age > 25) %>%
          compute_trends() %>%
          mutate(race = "Hispanic", chart_group = 4)
        ) %>%
    mutate(chart_group = factor(chart_group, labels = c("Skill", "Age Group", "Gender", "Race")))
```

```
men_olderthan25_trend_data %>%
    ggplot(aes(x = date, y = lfp)) +
    geom_line(data = ~ filter(.x, !is.na(skill)),
```

```
            aes(color = skill, group = skill), size = 1.4) +
  geom_line(data = ~ filter(.x, !is.na(age_group)),
            aes(color = age_group, group = age_group), size = 2) +
  geom_line(data = ~ filter(.x, !is.na(gender)),
            aes(color = gender, group = gender), size = 1.4) +
  geom_line(data = ~ filter(.x, !is.na(gender)),
            aes(color = gender, group = gender), size = 1.4) +
  geom_line(data = ~ filter(.x, !is.na(race)),
            aes(color = race, group = race), size = 1.4) +
  labs(title = "Labor Force Participation Trends",
       caption = "Data from Census ASEC",
       subtitle = "Among Men Older Than 25, 1976 - 2015",
       color = "Group") +
  scale_y_continuous(labels = label_percent()) +
  scale_color_discrete(limits = c("Skilled", "Unskilled", "25 <= age < 45",
                                  "45 <= age < 65", "65 <= age", "Male", "Female",
                                  "White", "Black", "Hispanic")) +

  theme_fivethirtyeight() +
  facet_wrap(~chart_group) +
  theme(strip.text = element_text(size = 10, face = "bold"))
```




```
men_olderthan25_trend_data %>%
    ungroup() %>%
    split(f = men_olderthan25_trend_data$chart_group) %>%
```

```
    map(.f = function (x) {
        a = slice(x, 1,41,81) %>% pull(lfp)
        b = slice(x, 40,80,120) %>% pull(lfp)
        result = percent(b - a)
        names(result) <-
            x %>%
            select(skill, age_group, gender, race) %>%
            select(function(y) any(!is.na(y))) %>% # Select Column Without Any NAs
            unique() %>%
            pull()
        result
        }
    )
```

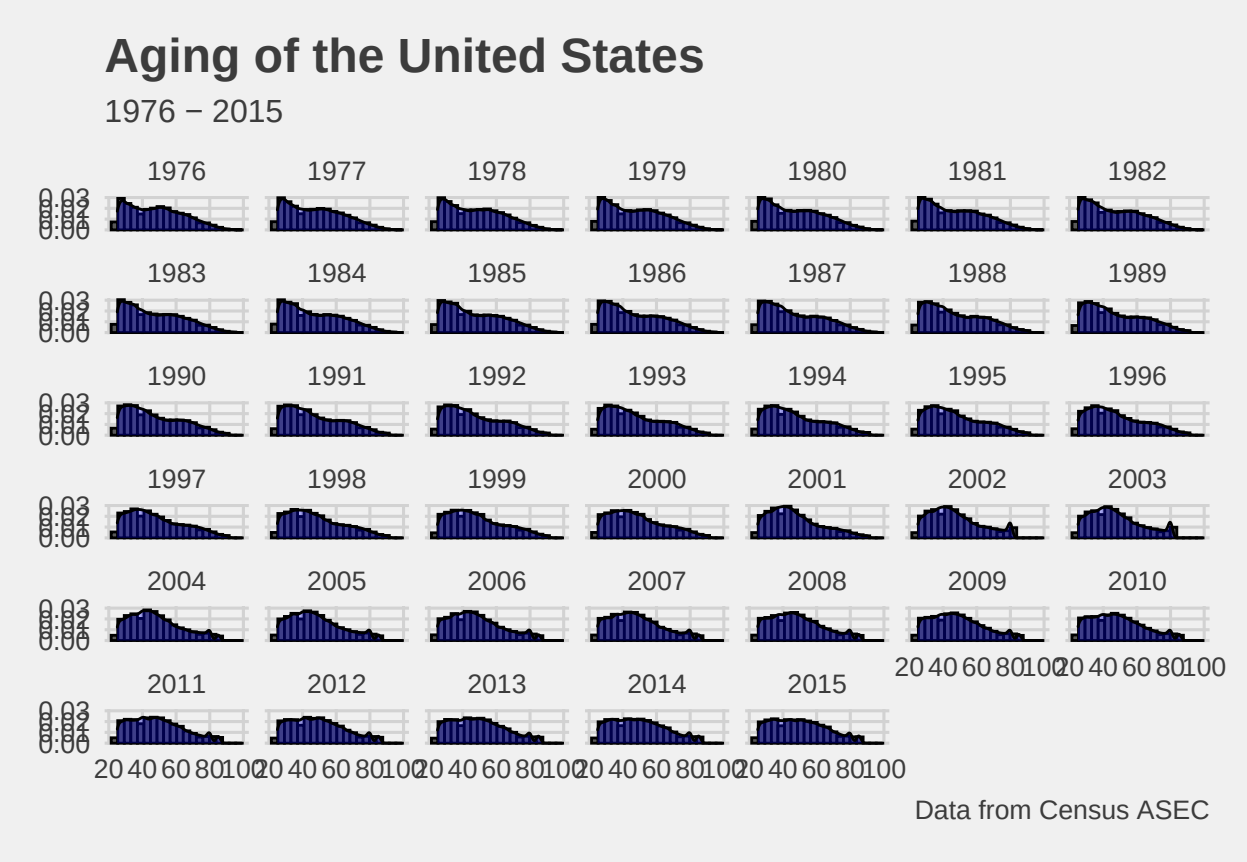
```
## $Skill
##   Skilled Unskilled
##  "-12.4%"   "-8.6%"
##
## $`Age Group`
## 25 <= age < 45 45 <= age < 65      65 <= age
##       "-5.74%"       "-6.30%"        "3.43%"
##
## $Gender
##   Male Female
##  "-7%"  "13%"
##
## $Race
##    White    Black Hispanic
##   "3.4%"   "0.1%"   "7.4%"
```

Discussion

Composition changes in the age distribution may explain why men older than 25 in the skilled subgroup experienced the largest change in LFP. Specifically as men in the U.S. have started to live longer and birth rates have declined, the median age has steadily climbed upwards. Since skill might be highly correlated with age, this would drag down the LFP rate for the skilled subgroup as older people naturally have a significantly lower LFP rate. Luckily we can use the same data set to test this hypothesis by finding the correlation between age and skill.

```
df <-
    data %>%
    filter(age >= 25) %>%
    select(year,month, age, skilled, wtsupp) %>%
    na.omit()
df %>%
    ggplot(aes(x = age)) +
    geom_histogram(aes(y = ..density..), bins = 20, color = "Black") +
    geom_density(alpha=.3, fill = "blue") +
    labs(title = "Aging of the United States",
         caption = "Data from Census ASEC",
         subtitle = "1976 - 2015") +
    theme_fivethirtyeight() +
```

```
    theme(legend.position = "None") +
    facet_wrap(~year)
```




```
# wtd.cors(df$age, df$skilled, df$wtsupp)
# df %>%
#     group_by(year) %>%
#     summarise(correlation = wtd.cors(age, skilled, wtsupp)) %>% view()
```

Another theory is that it is mainly driven by those with mid-level skills rather those whom are highly-skilled. We could test this by breaking up our skilled category to include a new "middle-skilled" category defined as those with at least 1 year of college but with education less than a bachelor's degree.

```
# Skilled is defined as those with at least 1 year of college education.
# head(data$educ)
# identical(data %>% filter(educ %in% 80:125), data %>% filter(skilled == 1))
# identical(data %>% filter(educ %in% 0:73), data %>% filter(skilled == 0))
midskilled_olderthan25_data <-
    # Skilled Workers Over 25 Years Old
    data %>%
    filter(educ %in% 111:125 & sex == 1 & age_group != 0) %>%
    compute_trends() %>%
    mutate(skill = "Highly Skilled", chart_group = 1) %>%
    rbind(
        # Skilled Workers Over 25 Years Old
        data %>%
```

```
          filter(skilled == 1 & sex == 1 & age_group != 0) %>%
          compute_trends() %>%
          mutate(skill = "Skilled", chart_group = 1),
        # Mid-Skilled Workers Over 25 Years Old
        data %>%
          filter(educ %in% 80:110 & sex == 1 & age_group != 0) %>%
          compute_trends() %>%
          mutate(skill = "Mid-Skilled", chart_group = 1),
        # Low-Skilled Workers Over 25 Years Old
        data %>%
          filter(skilled == 0 & sex == 1 & age_group != 0) %>%
          compute_trends() %>%
          mutate(skill = "Low-Skilled", chart_group = 1)
    )
midskilled_olderthan25_data %>%
    ggplot(aes(x = date, y = lfp)) +
    geom_line(data = ~ filter(.x, !is.na(skill)),
              aes(color = skill, group = skill), size = 2) +
    labs(title = "Labor Force Participation Trends, 1976 - 2015",
         caption = "Data from Census ASEC") +
    scale_y_continuous(labels = label_percent()) +
    scale_color_discrete(limits = c("Highly Skilled", "Skilled",
                                    "Mid-Skilled", "Low-Skilled")) +
    theme_fivethirtyeight()
```