

# Federal Reserve Bank of St. Louis Economic Data Report

## Project 2

Alicia Fels

```
library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(ggthemes)
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --
## v tibble  3.1.0      v purrr   0.3.4
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(Hmisc)

## Loading required package: lattice
## Loading required package: survival
## Loading required package: Formula
##
## Attaching package: 'Hmisc'
## The following objects are masked from 'package:dplyr':
##
##   src, summarize
## The following objects are masked from 'package:base':
##
##   format.pval, units

library(skimr)
library(ggribes)
library(scales)
```

```
##
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':
##
##   discard

## The following object is masked from 'package:readr':
##
##   col_factor

library(cowplot)

##
## Attaching package: 'cowplot'

## The following object is masked from 'package:ggthemes':
##
##   theme_map

library(gghighlight)
fredDataDF <- read.csv("fredData.csv")
fredDataDF$date <- as.Date(fredDataDF$date, format = "%Y-%m-%d") #have to change the date format to make
Hmisc::describe(fredDataDF)

## fredDataDF
##
## 16 Variables      879 Observations
## -----
## X
##      n missing distinct    Info    Mean      Gmd      .05      .10
##    879      0      879      1    440    293.3    44.9    88.8
##    .25    .50    .75    .90    .95
##   220.5   440.0   659.5   791.2   835.1
##
## lowest :    1    2    3    4    5, highest: 875 876 877 878 879
## -----
## medFamIncome
##      n missing distinct    Info    Mean      Gmd      .05      .10
##    52    827      52      1   40253   24721   10097   12136
##    .25    .50    .75    .90    .95
##   22047   37870   58827   66516   74250
##
## lowest :  8632  9433  9867 10285 11116, highest: 70697 72707 76135 78646 86011
## -----
## medHouseIncome
##      n missing distinct    Info    Mean      Gmd      .05      .10
##    36    843      36      1   42188   14114   24577   26643
##    .25    .50    .75    .90    .95
##   31090   42318   50250   57778   61647
##
## lowest : 22415 23618 24897 26061 27225, highest: 56516 59039 61136 63179 68703
## -----
## medPersIncome
##      n missing distinct    Info    Mean      Gmd      .05      .10
##    46    833      46      1   18559   10046    6109    7034
##    .25    .50    .75    .90    .95
```

```

##      11142      18172      26165      29498      31746
##
## lowest : 5335 5664 6002 6429 6813, highest: 30240 31099 31962 33706 35977
## -----
## fedMinWage
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      54      825      12      0.972      4.443      2.228      1.600      1.600
##      .25      .50      .75      .90      .95
##      3.350      4.250      5.675      7.250      7.250
##
## lowest : 1.60 2.30 2.90 3.10 3.35, highest: 4.75 5.15 5.85 6.55 7.25
##
## Value      1.60 2.30 2.90 3.10 3.35 3.80 4.25 4.75 5.15 5.85 6.55
## Frequency      8      3      1      1      10      1      5      1      10      1      1
## Proportion 0.148 0.056 0.019 0.019 0.185 0.019 0.093 0.019 0.185 0.019 0.019
##
## Value      7.25
## Frequency      12
## Proportion 0.222
## -----
## date
##      n missing distinct      Info      Mean      Gmd      .05
##      879      0      879      1 1984-07-31      8928 1951-08-28
##      .10      .25      .50      .75      .90      .95
## 1955-04-25 1966-04-16 1984-08-01 2002-11-16 2013-11-07 2017-07-04
##
## lowest : 1948-01-01 1948-02-01 1948-03-01 1948-04-01 1948-05-01
## highest: 2020-11-01 2020-12-01 2021-01-01 2021-02-01 2021-03-01
## -----
## lfprAll
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      879      0      108      1      62.86      3.492      58.20      58.80
##      .25      .50      .75      .90      .95
##      59.95      63.10      65.85      66.70      67.10
##
## lowest : 57.2 57.4 57.5 57.6 57.7, highest: 67.6 67.7 67.8 67.9 68.1
## -----
## lfprMen
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      879      0      202      1      77.23      5.951      68.9      69.8
##      .25      .50      .75      .90      .95
##      74.1      76.4      80.7      85.4      86.2
##
## lowest : 65.9 66.8 67.0 67.1 67.2, highest: 88.1 88.2 88.3 88.5 88.7
## -----
## lfprWomen
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      879      0      236      1      49.82      10.45      34.10      35.60
##      .25      .50      .75      .90      .95
##      40.10      53.90      58.10      59.60      59.81
##
## lowest : 30.7 31.4 31.8 32.0 32.1, highest: 60.0 60.1 60.2 60.3 60.4
## -----
## lfpr16_19yrs

```

```

##          n missing distinct      Info      Mean      Gmd      .05      .10
##          879         0       316         1     48.02     10.27     32.70     33.98
##          .25        .50        .75        .90        .95
##          41.95     48.20     52.40     61.42     64.60
##
## lowest : 28.1 30.0 30.4 30.8 31.1, highest: 69.6 69.8 70.7 70.9 71.8
## -----
## lfpr20_24yrs
##          n missing distinct      Info      Mean      Gmd      .05      .10
##          879         0       221         1     71.85      6.396     62.39     63.80
##          .25        .50        .75        .90        .95
##          66.85     72.90     76.40     78.30     80.30
##
## lowest : 58.5 59.5 60.0 60.1 60.3, highest: 82.5 82.7 82.8 82.9 83.8
## -----
## lfpr25_54yrs
##          n missing distinct      Info      Mean      Gmd      .05      .10
##          879         0       182         1     77.02      7.204     66.29     67.50
##          .25        .50        .75        .90        .95
##          70.20     80.60     82.90     83.60     83.90
##
## lowest : 63.6 64.1 64.4 64.6 64.7, highest: 84.1 84.2 84.3 84.4 84.5
## -----
## lfpr55yrsPlus
##          n missing distinct      Info      Mean      Gmd      .05      .10
##          879         0       149         1     36.68      5.071     29.80     30.10
##          .25        .50        .75        .90        .95
##          32.10     38.60     40.20     41.92     42.80
##
## lowest : 28.9 29.1 29.3 29.4 29.5, highest: 43.7 43.8 43.9 44.0 44.1
## -----
## shareNetWorthBott50
##          n missing distinct      Info      Mean      Gmd      .05      .10
##          126        753        38     0.998      2.365      1.352      0.600      0.800
##          .25        .50        .75        .90        .95
##          1.225      2.300      3.500      3.750      3.975
##
## lowest : 0.4 0.5 0.6 0.7 0.8, highest: 3.8 3.9 4.0 4.1 4.3
## -----
## shareNetWorth50_90
##          n missing distinct      Info      Mean      Gmd      .05      .10
##          126        753        61         1     32.83      3.048     28.83     29.10
##          .25        .50        .75        .90        .95
##          30.52     33.75     35.38     36.10     36.20
##
## lowest : 28.3 28.6 28.7 28.8 28.9, highest: 36.0 36.1 36.2 36.3 36.6
## -----
## shareNetWorthTop1
##          n missing distinct      Info      Mean      Gmd      .05      .10
##          126        753        65     0.999     28.15      2.604     23.70     24.40
##          .25        .50        .75        .90        .95
##          26.65     28.30     29.98     31.10     31.20
##
## lowest : 23.1 23.3 23.4 23.5 23.6, highest: 31.1 31.2 31.3 31.4 31.5

```

## -----

## Executive Summary

The overall sample size for Federal Reserves Economics data set was 879 observations at different time points from 1948 to 2021. Data was collected on median family income, median household income, median personal income, federal minimum wage, labor force participation rate over all working-age individuals aged 16+, labor force participation rate for working age men, labor force participation rate for working age women, labor force participation rate for 16-19, 20-24, 25-54, and 55+ years old; share of the nation's net worth held by the bottom 50%, by the middle 50-90%, and held by the top 1% at each time point. 7 of the collected variables are missing observations, the proportion of numbers of missing cases/data: median family income has a 0.94 proportion, median household income has a 0.96 proportion, median personal income has a 0.95 proportion, federal minimum wage has a 0.94 proportion, all the net worth variables (bottom 50%, middle 50-90%, and top 1%) have a 0.86 proportion of missing cases. However, these variables were only first reported later than the beginning of the time period for the labor force participation rate variables and they are also reported at different intervals as well. For example, both Median Family Income and Federal Minimum Wage were first reported in January 1968 and reported once every year thereafter. Median Personal Income was first reported in January 1974 (also reported once every year afterwards) and Median Household Income was only first reported in January 1984, though all three median income variables were not reported after January 2019. Lastly, all three net worth variables were first reported in July 1989, with follow up at the beginning of every new quarter (January, April, July, and October) with the last net worth reported in October 2020.

We found: - The share of net worth of the Top 1% can most likely predict the share of the 50-90% because the two variables are negatively correlated to each other. - The average shared net worth for the Bottom 50%, 50-90%, and the Top 1% all differed between 1990s and 2000s. - Median personal income and median household income increased over time and are positively related to each other. - The labor force participation rate for women increased over time, while the participation rate of men decreased over time. - The federal minimum wage also increased as personal and household income increased. - The distribution of labor force participation rate of women before January 1980 was lower than the overall participation rate. - The labor force participation rate changed over time in different directions depending on the age group that is looked at. - Lastly, the average labor force participation rate of those 55 years and older differed between the 1970s and the 1990s.

**Question: What happens to the share of the net worth of the 50-90% as that of the top 1% increases? Is it possible to predict the share of net worth of the 50-90 from the top 1%?**

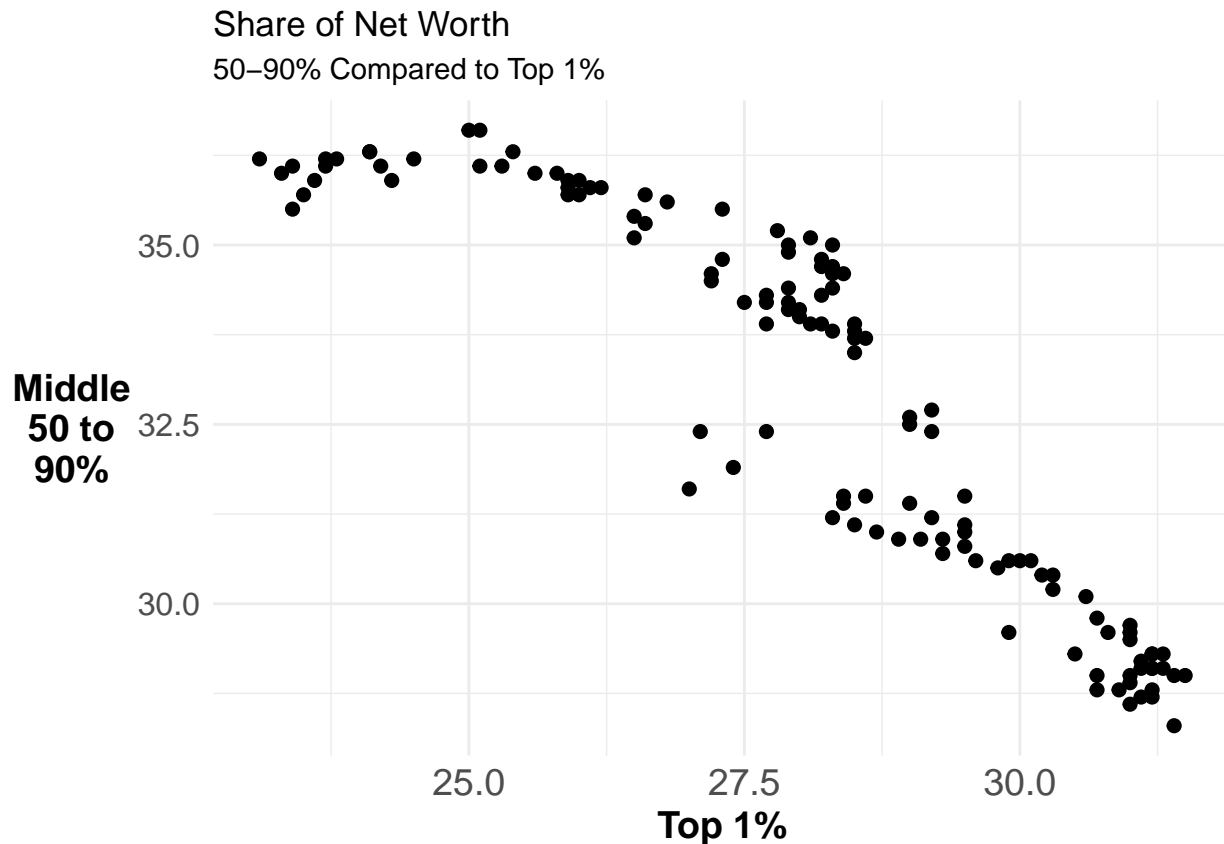
The graph below depicts the relationship of the share of the net worth of the 50-90% as the top 1% increases. On the x-axis is the net worth for the Top 1% and the y-axis is the net worth of the 50-90%. We can see that there is a negative relationship between the Top 1% share of net worth and the 50-90% share. Thus, as the Top 1% share of the nation's net worth increases, the 50-90% share decreases. While the share of net worth has only been reported every quarter from July 1989 to October 2020, there are 126 reported cases which is a large sample size, we can predict what the share of net worth might be for the 50-90% from the Top 1%. We can predict this because in general, when the share of the net worth for the Top 1% is around 32.5, the share of the net worth for the 50-90% is around 27.5.

```
fredDataDF %>%
  filter(!is.na(shareNetWorth50_90), !is.na(shareNetWorthTop1)) %>%
  ggplot(aes(x = shareNetWorthTop1, y = shareNetWorth50_90)) +
  geom_point(size = 2) +
  labs(title = "Share of Net Worth",
       subtitle = "50-90% Compared to Top 1%",
       x = "Top 1%",
```

```

y = "Middle\n50 to\n90%" +
theme_minimal() +
theme(axis.title.x = element_text(face = "bold", color = "#000000", size = 14),
      axis.text.x = element_text(angle = 0, size = 14),
      axis.text.y = element_text(size = 12),
      axis.title.y = element_text(angle = 0, size = 14, face = "bold", vjust = .5))

```



**Question: Is the average shared Net Worth for the Bottom 50% different in the 1990s versus the 2000s?**

In the graph below, we can see that the average share of net worth for the Bottom 50% is different in the 1990s compared to the 2000s. As we can see there was a significant decrease in the the share of net worth for the Bottom 50% from the 1990s to the 2000s. In the 1990s the average share of net worth was 3.70, while in the 2000s the average share of net worth was 2.14.

```

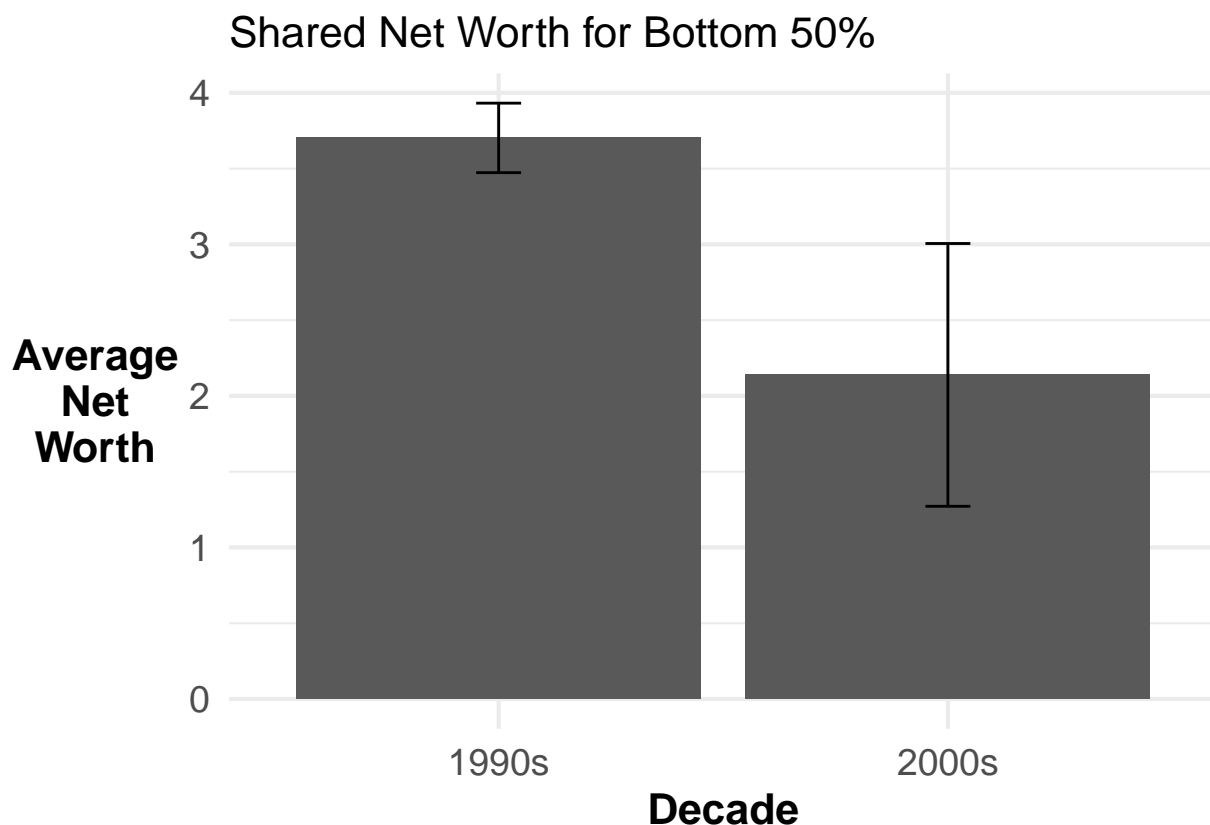
bottom50DF <- fredDataDF %>%
  filter(!is.na(shareNetWorthBott50)) %>%
  mutate(
    decade = ifelse(date > as.Date("1989-12-31") & date < as.Date("1999-12-31"), "1990s",
                    ifelse(date > as.Date("1999-12-31") & date < as.Date("2010-12-31"), "2000s", NA))) %>%
  group_by(decade) %>%
  summarise(n = n(),
            mean = mean(shareNetWorthBott50),
            sd = sd(shareNetWorthBott50),
            se = sd / sqrt(n))
bottom50DF %>%

```

```

filter(!is.na(decade))%>%
ggplot(aes(x = decade, y = mean, group = 1)) +
geom_bar(stat = "identity") +
geom_errorbar(aes(ymin = mean - sd,
                  ymax = mean + sd),
              width = .1) +
labs(title = "Shared Net Worth for Bottom 50%",
     x = "Decade",
     y = "Average\nNet\nWorth") +
theme_minimal(base_size = 16) +
theme(axis.title.x = element_text(face = "bold", color = "#000000", size = 16),
      axis.text.x = element_text(size = 14),
      axis.text.y = element_text(angle = 0, size = 14),
      axis.title.y = element_text(angle = 0, size = 16, face = "bold", vjust = .5),
      plot.title = element_text(angle = 0, size = 16, vjust = .5))

```



### Subquestion: What about the middle 50-90%?

We can see, in the graph below, that the average share of net worth for the 50-90% is different in the 1990s compared to the 2000s. As we can see there was a slight decrease in the the share of net worth for the 50-90% from the 1990s to the 2000s. In the 1990s the average share of net worth was 35.26, while in the 2000s the average share of net worth was 33.45.

```

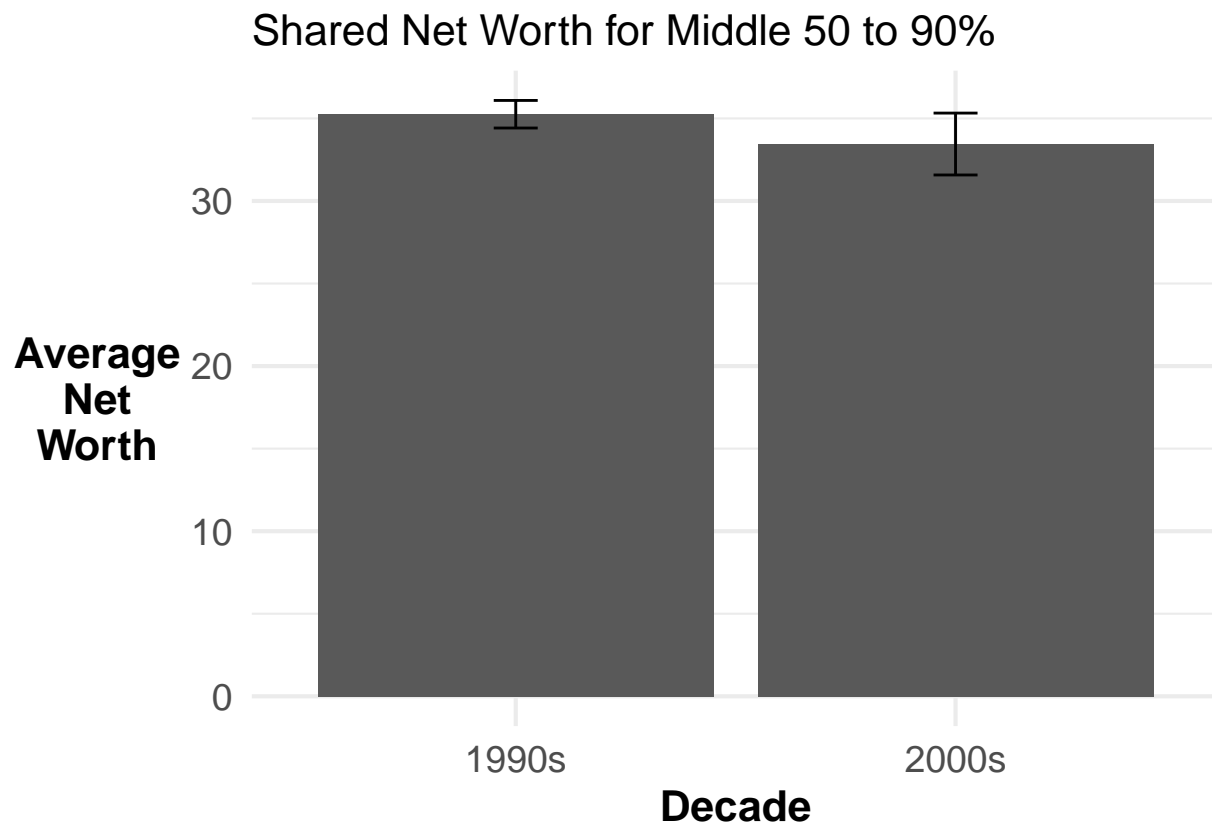
middle50_90DF <- fredDataDF %>%
  filter(!is.na(shareNetWorth50_90)) %>%
  mutate(
    decade = ifelse(date > as.Date("1989-12-31") & date < as.Date("1999-12-31"), "1990s",
                    ifelse(date > as.Date("1999-12-31") & date < as.Date("2010-12-31"), "2000s", NA))) %>%

```

```

group_by(decade) %>%
  summarise(n = n(),
            mean = mean(shareNetWorth50_90),
            sd = sd(shareNetWorth50_90),
            se = sd / sqrt(n))
middle50_90DF %>%
  filter(!is.na(decade))%>%
  ggplot(aes(x = decade, y = mean, group = 1)) +
  geom_bar(stat = "identity") +
  geom_errorbar(aes(ymin = mean - sd,
                  ymax = mean + sd),
              width = .1) +
  labs(title = "Shared Net Worth for Middle 50 to 90%",
       x = "Decade",
       y = "Average\nNet\nWorth") +
  theme_minimal(base_size = 16) +
  theme(axis.title.x = element_text(face = "bold", color = "#000000", size = 16),
        axis.text.x = element_text(size = 14),
        axis.text.y = element_text(angle = 0, size = 14),
        axis.title.y = element_text(angle = 0, size = 16, face = "bold", vjust = .5),
        plot.title = element_text(angle = 0, size = 16, vjust = .5))

```



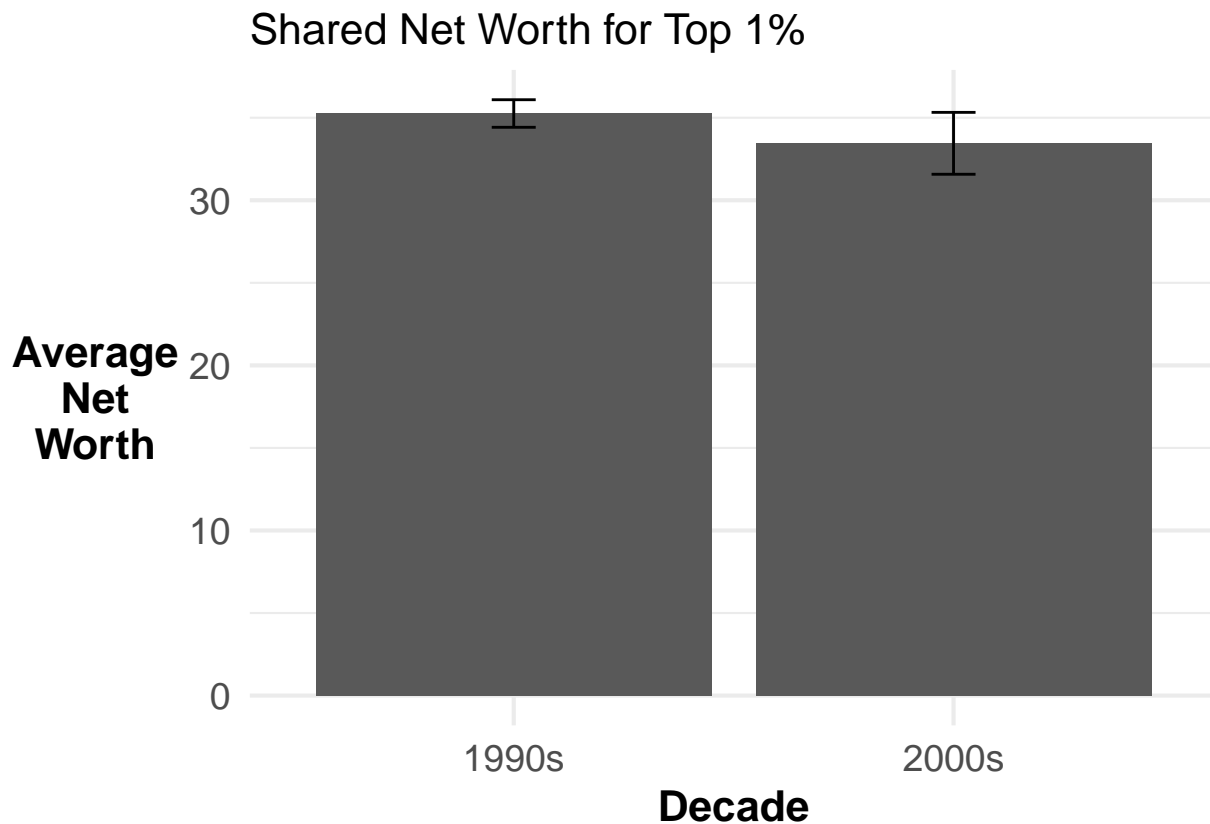
### Subquestion: What about the Top 1%?

The graph below visualizes a difference in the average share of total net worth for the Top 1% in the 1990s compared to the 2000s. As we can see there was a slight decrease in the the share of net worth for the Top 1% from the 1990s to the 2000s. In the 1990s the average share of net worth was 26.42, while in the 2000s



the average share of net worth was 27.68.

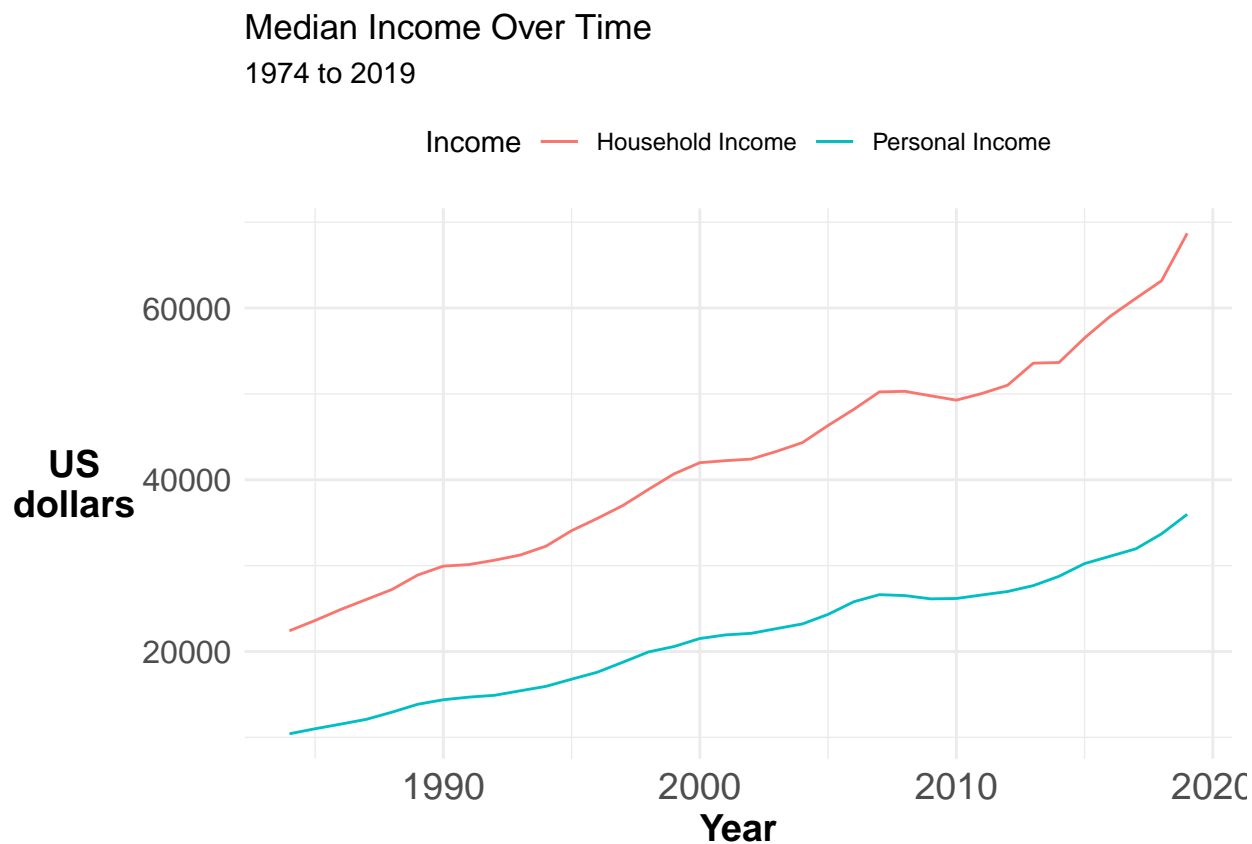
```
top1DF <- fredDataDF %>%
  filter(!is.na(shareNetWorthTop1)) %>%
  mutate(
    decade = ifelse(date > as.Date("1989-12-31") & date < as.Date("1999-12-31"), "1990s",
                     ifelse(date > as.Date("1999-12-31") & date < as.Date("2010-12-31"), "2000s", NA))) %>%
  group_by(decade) %>%
  summarise(n = n(),
            mean = mean(shareNetWorthTop1),
            sd = sd(shareNetWorthTop1),
            se = sd / sqrt(n))
middle50_90DF %>%
  filter(!is.na(decade))%>%
  ggplot(aes(x = decade, y = mean, group = 1)) +
  geom_bar(stat = "identity") +
  geom_errorbar(aes(ymin = mean - sd,
                    ymax = mean + sd),
                width = .1) +
  labs(title = "Shared Net Worth for Top 1%",
        x = "Decade",
        y = "Average\nNet\nWorth") +
  theme_minimal(base_size = 16) +
  theme(axis.title.x = element_text(face = "bold", color = "#000000", size = 16),
        axis.text.x = element_text(size = 14),
        axis.text.y = element_text(angle = 0, size = 14),
        axis.title.y = element_text(angle = 0, size = 16, face = "bold", vjust = .5),
        plot.title = element_text(angle = 0, size = 16, vjust = .5))
```



## Question: How has the median personal income and the median household income change over time?

The graph below depicts the change of median personal (blue line) and household (red line) income over time from 1974 for Personal income and 1984 for household income to 2019. As we can see both personal and household income increased over time, however personal income was overall lower than household income, which makes sense since household income is the sum of all income in one household.

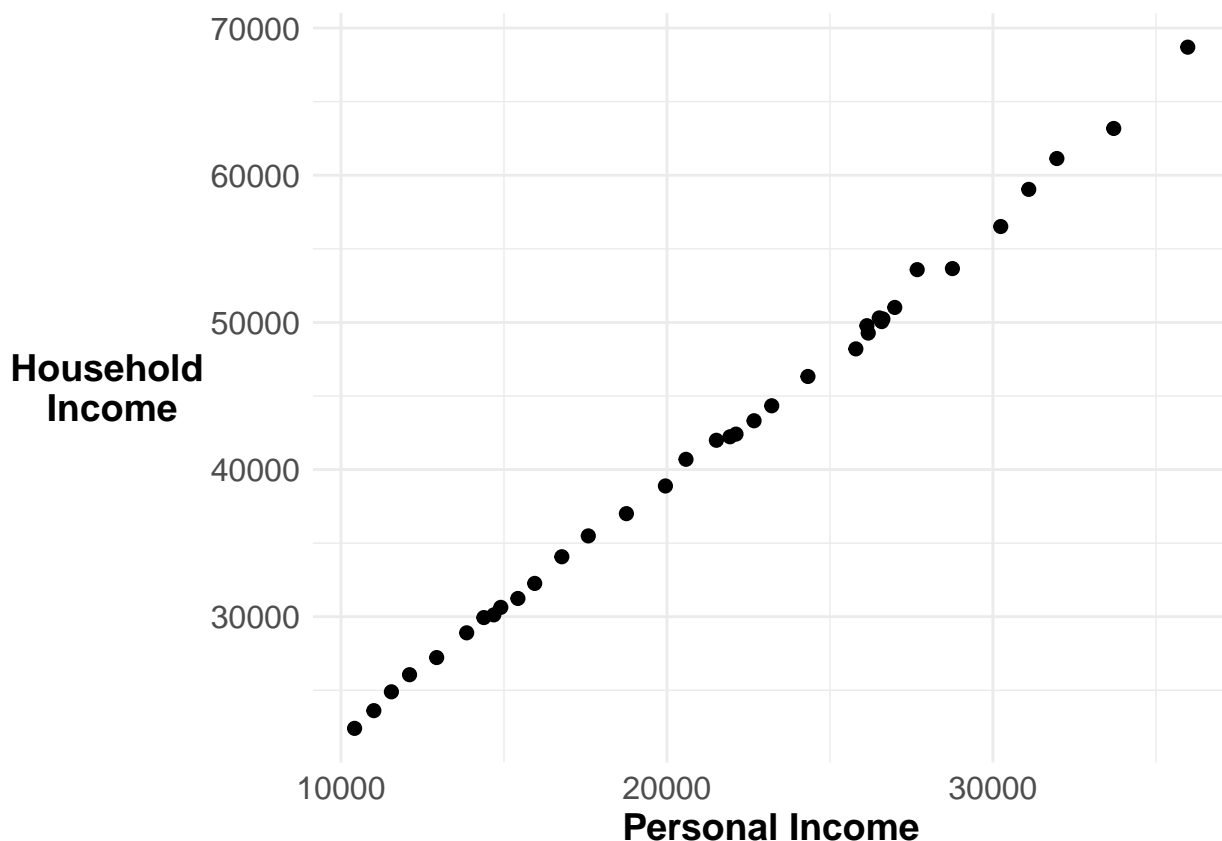
```
fredDataDF %>%
  filter(!is.na(medPersIncome), !is.na(medHouseIncome)) %>%
  ggplot(aes(x = date)) +
  geom_line(aes(y = medPersIncome, color = "Personal Income")) +
  geom_line(aes(y = medHouseIncome, color = "Household Income")) +
  labs(title = "Median Income Over Time",
       subtitle = "1974 to 2019",
       x = "Year",
       y = "US\ndollars") +
  theme_minimal() +
  theme(axis.title.x = element_text(face = "bold", color = "#000000", size = 14),
        axis.text.x = element_text(angle = 0, size = 14),
        axis.text.y = element_text(size = 12),
        axis.title.y = element_text(angle = 0, size = 14, face = "bold", vjust = .5)) +
  scale_color_discrete("Income") +
  theme(legend.position = "top")
```



**Subquestion: Do they always vary together? How might we explain divergence or gaps between household income and personal income?**

Personal and household income always directly vary together, which is visualized in the graph below. As personal income increases household income increases as well, which makes sense since median household income relies on the personal income of each household member. One possible explanation for this divergence between household and personal income could be household size, which was not reported in this data set. Since household income is made up of the individual income of each member of a household, the size of the household and how many are working will increase the gap between median household and median personal income.

```
fredDataDF %>%
  filter(!is.na(medPersIncome), !is.na(medHouseIncome)) %>%
  ggplot(aes(x = medPersIncome, y = medHouseIncome)) +
  geom_point(size = 2) +
  labs(x = "Personal Income", y = "Household\n Income") +
  theme_minimal() +
  theme(axis.title.x = element_text(face = "bold", color = "#000000", size = 14),
        axis.text.x = element_text(angle = 0, size = 12),
        axis.text.y = element_text(size = 12),
        axis.title.y = element_text(angle = 0, size = 14, face = "bold", vjust = .5))
```

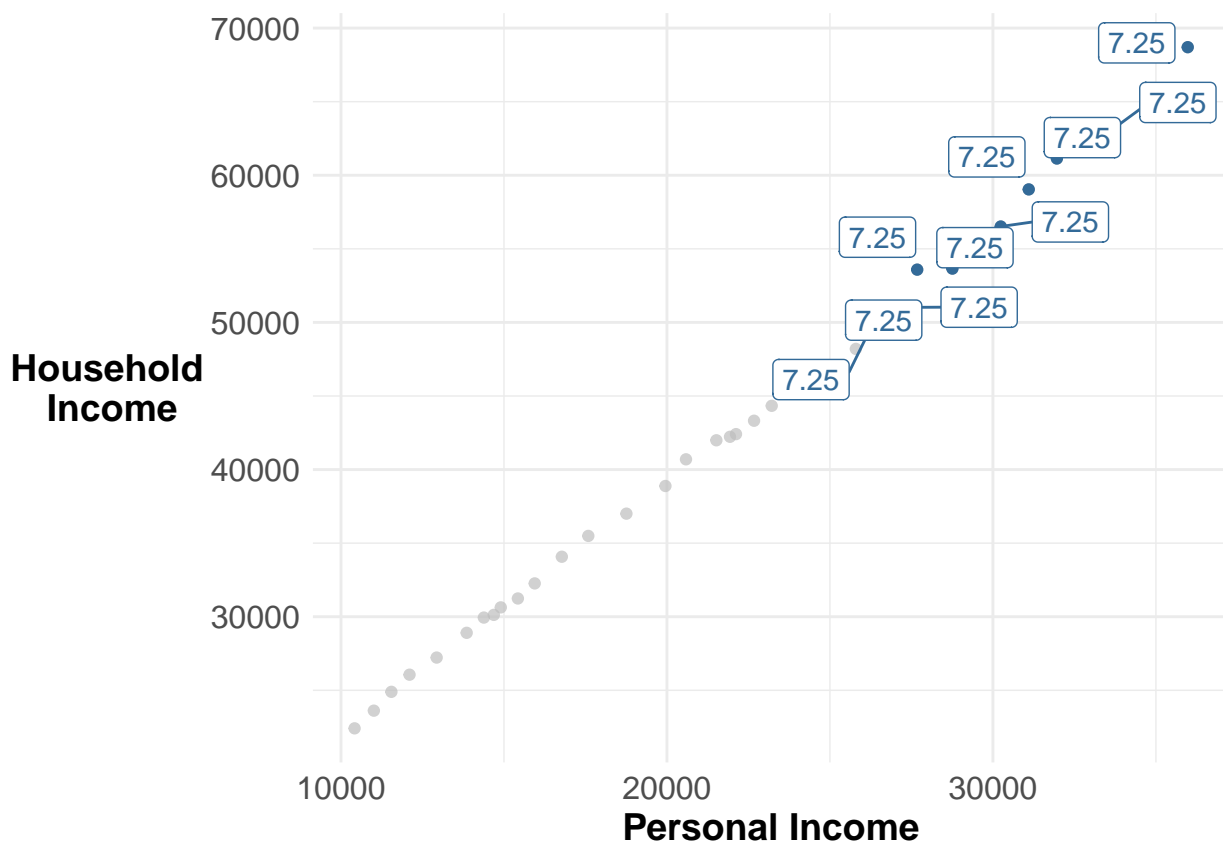


**Question: How is median personal income related to household income with a focus on federal minimum wage greater than \$7.00?**

As established in the prior question, median personal and median household income are positively related (personal income increases then household income increases). However, when we highlight all of the federal

minimum wage greater than \$7 (visualized below), we can see that the highest federal minimum wage reported was \$7.25, which were also the highest reported household (greater than \$50,000) and personal income (greater than \$20,000). This means that as household and personal income increased, federal minimum wage looks to increase as well.

```
fredDataDF %>%
  filter(!is.na(medPersIncome), !is.na(medHouseIncome)) %>%
  ggplot(aes(x = medPersIncome, y = medHouseIncome, colour = fedMinWage, use_group_by=FALSE)) +
  geom_point() +
  gghighlight(fedMinWage > 7, label_key = fedMinWage) + # here I am highlighting the data points where
  labs(x = "Personal Income", y = "Household Income") +
  theme_minimal() +
  theme(axis.title.x = element_text(face = "bold", color = "#000000", size = 14),
        axis.text.x = element_text(angle = 0, size = 12),
        axis.title.y = element_text(size = 12),
        axis.text.y = element_text(angle = 0, size = 14, face = "bold", vjust = .5))
```

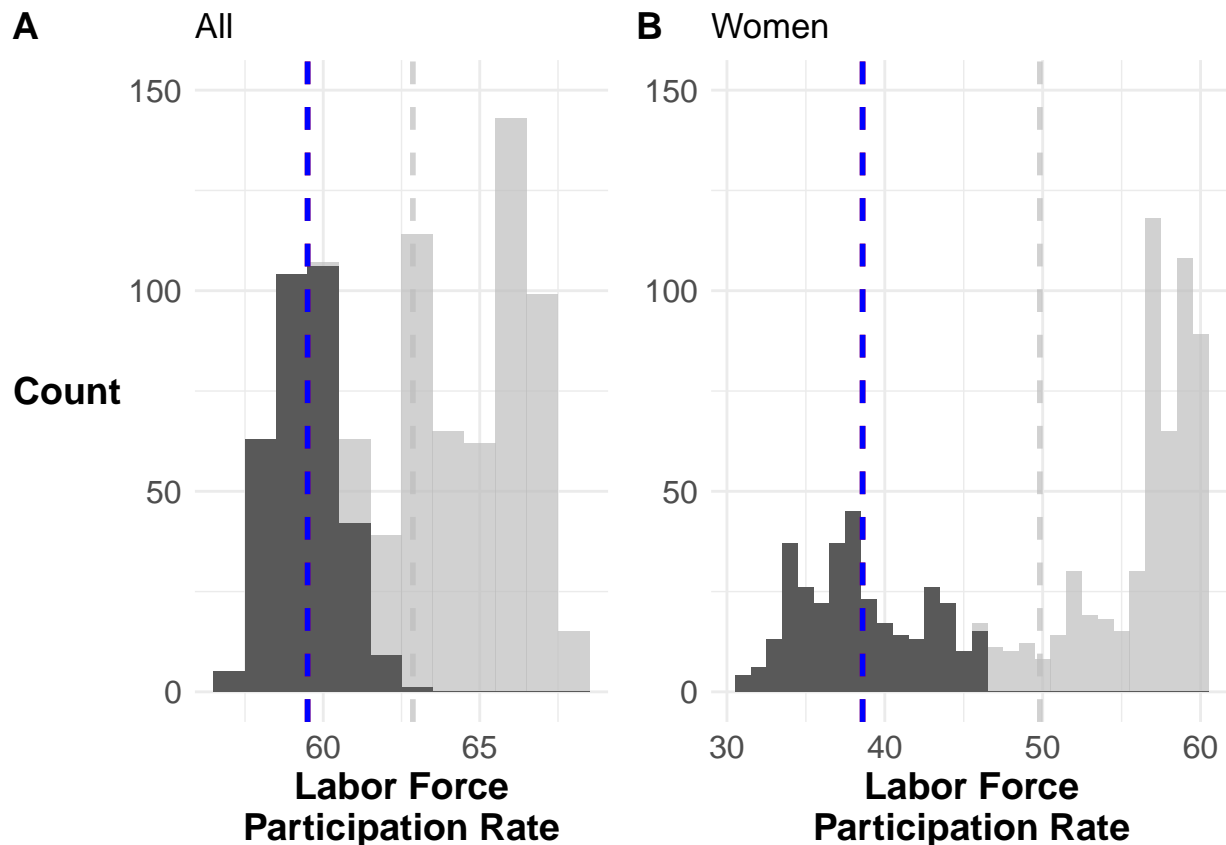


**Question: What is the distribution of labor force participation overall compared to the overall distribution of labor force participation for women? What is the distribution for both earlier than 1980?**

The figure below visualizes the distribution of the total labor force participation rate in the left graph and the distribution of labor force participation rate of women in the right graph. The light gray sections of both graphs is the distribution of labor force participation after January 1980, while the dark gray is the labor force participation before January 1980. The gray dashed line in both graphs indicates the overall average participation rate from 1948 to 2021, while the blue dashed lines indicates the average participation rate

before January 1980. When we compare the distributions in both graphs we can see that before January 1980 the participation rate of women was lower than the overall participation rate. We can also see that the average participation rate for women before 1980 is approximately 38, while the overall average participation rate was approximately 59. Lastly, we can see that after January 1980 the participation rate of women increased significantly, but is still less than the overall participation rate; the average overall participation rate is 62.86 while the average participation rate of women is 49.82. We can also see that that entire distribution of participation rate of women is negatively skewed, while the overall participation rate seems to be normally distributed.

```
women <- ggplot(fredDataDF, aes(x=lfprWomen)) +
  geom_histogram(binwidth=1) +
  ylim(0, 150) +
  geom_vline(aes(xintercept=mean(lfprWomen)),
    color="red", linetype="dashed", size=1) +
  gghighlight(date < 1980-01-01) +
  labs(title = "Women",
    x = "Labor Force\nParticipation Rate",
    y = "") +
  theme_minimal() +
  theme(axis.title.x = element_text(face = "bold", color = "#000000", size = 14),
    axis.text.x = element_text(angle = 0, size = 12),
    axis.text.y = element_text(size = 12),
    axis.title.y = element_text(angle = 0, size = 14, face = "bold", vjust = .5)) +
  geom_vline(aes(xintercept=mean(lfprWomen)),
    color="blue", linetype="dashed", size=1) # the mean is 49.82036
all <- ggplot(fredDataDF, aes(x=lfprAll)) +
  geom_histogram(binwidth=1) +
  ylim(0, 150) +
  geom_vline(aes(xintercept=mean(lfprAll)),
    color="red", linetype="dashed", size=1) +
  gghighlight(date < 1980-01-01) +
  labs(title = "All",
    x = "Labor Force\nParticipation Rate",
    y = "Count") +
  theme_minimal() +
  theme(axis.title.x = element_text(face = "bold", color = "#000000", size = 14),
    axis.text.x = element_text(angle = 0, size = 12),
    axis.text.y = element_text(size = 12),
    axis.title.y = element_text(angle = 0, size = 14, face = "bold", vjust = .5)) +
  geom_vline(aes(xintercept=mean(lfprAll)),
    color="blue", linetype="dashed", size=1) # the mean is 62.86303
plot_grid(all, women, labels = "AUTO")
```

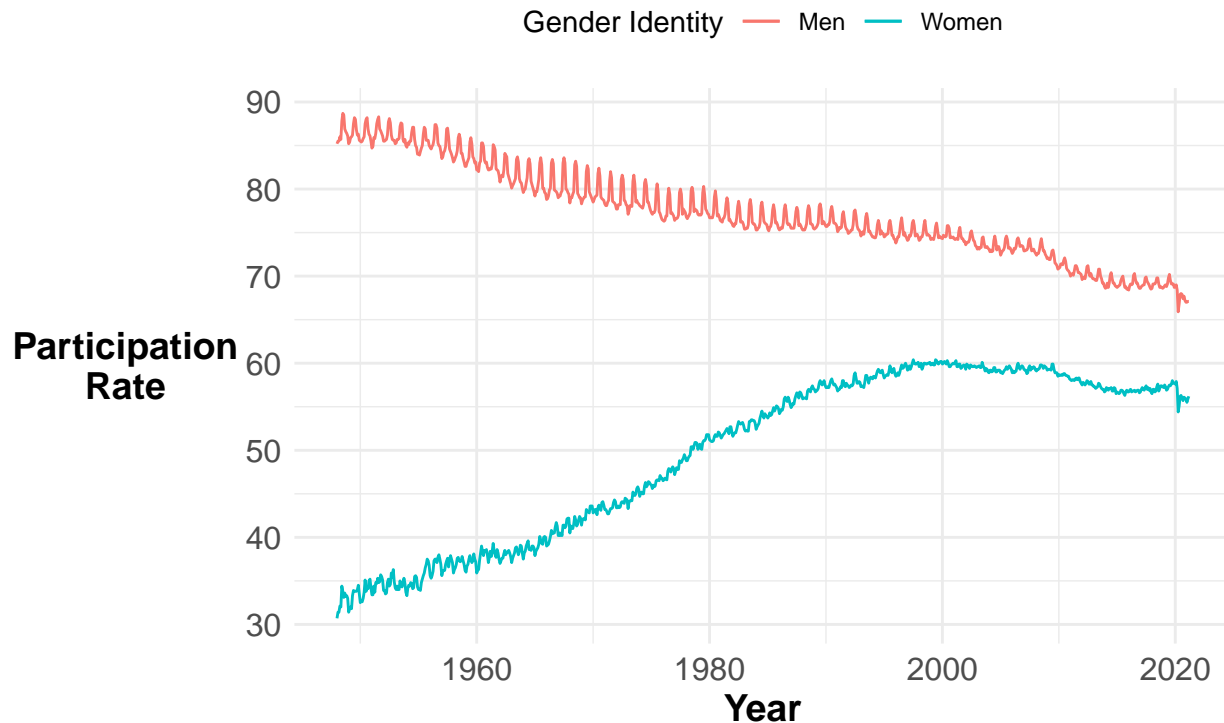


**Question: How has the labor force participation rate of men and women in the United States change over time?**

The graph below visualizes change in labor force participation of men (the red line) and women (the blue line) over time, from 1948 to 2021. We can see that there were sharp dips and spikes in participation rate for both genders, but overall women's participation rate increased over time while men's participation rate decreased over time. We can also see a slight decrease of women and men participating in the labor force between 2010 and 2020, with the largest dip between 2020 and 2021, which makes sense due to the COVID-19 pandemic causing a lot of people in the US to lose their jobs.

```
ggplot(fredDataDF, aes(x = date)) +
  geom_line(aes(y = lfprMen, color = "Men")) + #Since we are looking at how both change over time and t
  geom_line(aes(y = lfprWomen, color = "Women")) +
  labs(title = "Labor Force Participation Rate Over Time",
        subtitle = "1948 to 2021",
        x = "Year",
        y = "Participation\nRate") +
  theme_minimal() +
  theme(axis.title.x = element_text(face = "bold", color = "#000000", size = 14),
        axis.text.x = element_text(angle = 0, size = 12),
        axis.text.y = element_text(size = 12),
        axis.title.y = element_text(angle = 0, size = 14, face = "bold", vjust = .5)) +
  scale_color_discrete("Gender Identity") +
  theme(legend.position = "top")
```

## Labor Force Participation Rate Over Time 1948 to 2021



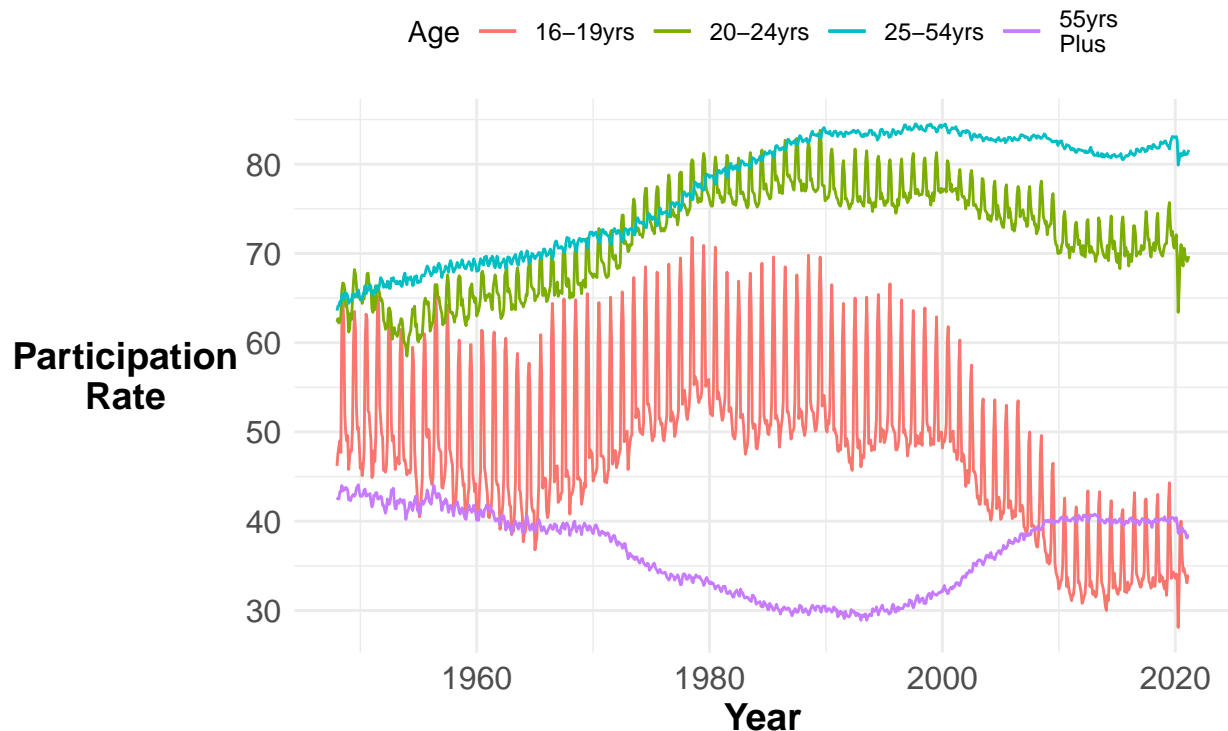
**Question: How has the rate of labor force participation changed over time for each age group?**

Visualized in the graph below is the change in labor force participation rate by age group over time (16-19 year group is the red line, 20-24 year group is the green line, 25-54 year group is the blue line, and those over 55 group is the purple line). As we can see each age group, like with what was seen with the two genders, there were sharp dips and spikes, but each group seems to have an overall trend. For the 16-19 year old participation rate seemed to decrease slightly before increasing until 2000, where the participation rate decreased again before leveling out between 2010 and 2020. For the 20-24 year olds, their participation rate increased over time before decreasing slightly between 2010 and 2020. The 25-54 year group seems to follow a similar trend as the 20-24 year olds, but this age group did not show a decrease in participation until 2020. Lastly, the participation rate of individuals over 55 years old decreased slightly between 1948 and 1970, before significantly decreasing from 1970 to 1990. This age group also increased in participation between 1990 and 2010 before leveling off.

```
ggplot(fredDataDF, aes(x = date)) +
  geom_line(aes(y = lfpr16_19yrs, color = "16-19yrs")) +
  geom_line(aes(y = lfpr20_24yrs, color = "20-24yrs")) +
  geom_line(aes(y = lfpr25_54yrs, color = "25-54yrs")) +
  geom_line(aes(y = lfpr55yrsPlus, color = "55yrs\nPlus")) +
  labs(title = "Labor Force Participation Rate Over Time",
        subtitle = "1948 to 2021",
        x = "Year",
        y = "Participation\nRate") +
  theme_minimal() +
  theme(axis.title.x = element_text(face = "bold", color = "#000000", size = 14),
```

```
axis.text.x = element_text(angle = 0, size = 12),
axis.text.y = element_text(size = 12),
axis.title.y = element_text(angle = 0, size = 14, face = "bold", vjust = .5)) +
scale_color_discrete("Age") +
theme(legend.position = "top")
```

## Labor Force Participation Rate Over Time 1948 to 2021



**Question:** Is the average labor force participation rate of people over the age of 55 different in the 1970s versus the 1990s?

In this last graph below I have visualized the average participation rate of individuals over 55 years old in the 1970s and 1990s. We can see that on average individuals over 55 years old had a higher participation rate in the 1970s compared to their participation rate in the 1990s. Their average participation rate was 35.50 in the 1970s while their average participation rate was 30.31 in the 1990s.

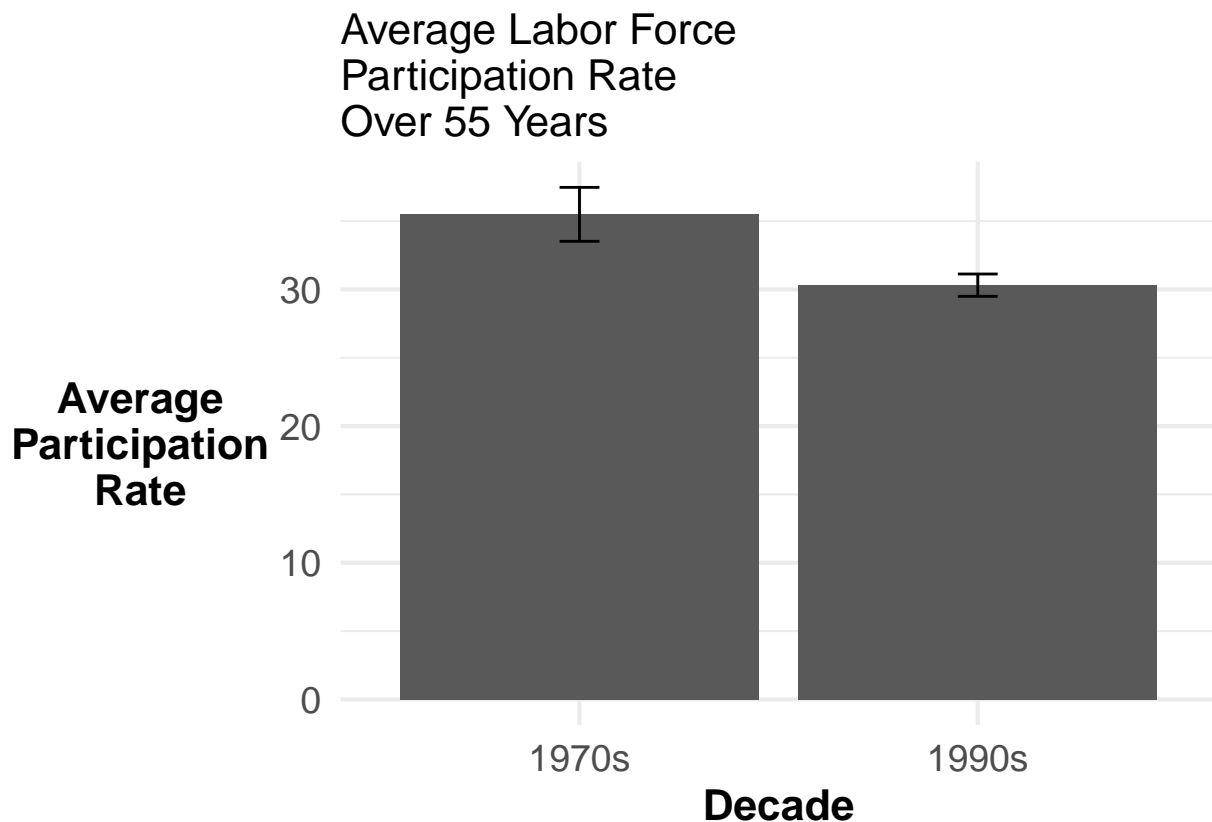
```
lfpr55DF <- fredDataDF %>%
  mutate(
    decade = ifelse(date > as.Date("1969-12-31") & date < as.Date("1979-12-31"), "1970s",
                     ifelse(date > as.Date("1989-12-31") & date < as.Date("1999-12-31"), "1990s", NA))) %>%
  group_by(decade) %>%
  summarise(n = n(),
            mean = mean(lfpr55yrsPlus),
            sd = sd(lfpr55yrsPlus),
            se = sd / sqrt(n))
lfpr55DF %>%
  filter(!is.na(decade)) %>%
  ggplot(aes(x = decade, y = mean, group = 1)) +
```



```

geom_bar(stat = "identity") +
  geom_errorbar(aes(ymin = mean - sd,
                    ymax = mean + sd),
                width = .1) +
labs(title = "Average Labor Force\nParticipation Rate\nOver 55 Years",
     x = "Decade",
     y = "Average\nParticipation\nRate") +
theme_minimal(base_size = 16) +
theme(axis.title.x = element_text(face = "bold", color = "#000000", size = 16),
      axis.text.x = element_text(size = 14),
      axis.text.y = element_text(angle = 0, size = 14),
      axis.title.y = element_text(angle = 0, size = 16, face = "bold", vjust = .5),
      plot.title = element_text(angle = 0, size = 16, vjust = .5))

```



## Conclusion

The sample size for this data set is fairly large and represents different economic variables between 1948 and 2021. Due to the longitudinal style of this data set and the amount of data that was able to be collected the results are likely stable. From this we can predict that median personal income and median household income is positively related to each other and they both increased over time. We can also conclude that the share of net worth of the Top 1% can most likely predict the share of the 50-90% because the two variables are negatively correlated to each other. We can also see that the labor force participation rate for women increased over time, while the participation rate of men decreased over time.