Labor PS

Question 1: Data Preparation and Analysis

(a) Use the ACS data set that you have downloaded and cleaned up from IPUMS and plot the yearly mean wages, mean hours worked (unconditional and conditional) and mean employment rates of women aged 15-65 in the downloaded sample

In this question we will load the data of the ACS dataset from IPUMS. The data was cleaned using **Stata** and the do-file can be seen in the folder of this problem set. Since we used the same cleaned dataset to be able to do comparisions, we won't talk any further about the process of cleaning the data.

Now we start by loading some packages in **R** and the data in the code chunk below. Note that we have some comments that are worth reading.

```
# Packages
library(tidyverse) # Package for everything
library(haven)     # Package for reading dta files
library(ggthemes)  # Package for themes
library(lubridate) # Converts to date format
library(np)        # Package for non parametric and semiparametric
library(purrr)
library(ks)
library(Matrix)   # Faster computations of matrices

# set.seed(666)
set.seed(666)

# Importing data ------------------------------------------------------------

# Important note: we are dealing with a database that has already been cleaned.
# We will convert the file in a Rdata format so that we can load faster the
```

```
#data.

# Only use this option if you dont't have acess to the Rdata format and use the
# haven package
# data_ps <- read_dta(file = "data_PS1.dta")

# Convert to Rdata the data_ps
# saveRDS(data_ps, file = "data_ps1.rds")

# Load data
data_ps1 <- as_tibble(readRDS(file =  "data_ps1.rds"))
```

We now create new variables that we will use for the graphs of this problem. We create the variables *real_hhincome, real_wage, labor_par*, and *non_labor_income*. Respectively, each represents the real household income, the real anual wage, the labor participation, and the non labor income.

```
# Creates real values of wages
data_ps1 <- data_ps1 %>%
  mutate(real_hhincome = (hhincome*100) / Price_Index,
         real_wage = (incwage * 100) / Price_Index,
         labor_par = ifelse(empstat %in% c(1,2),1, ifelse( empstat == 3 ,0, NA)),
         non_labor_income = real_hhincome - real_wage)
```

And finally, with the following code chunk below, we generate the graphs that are below the code chunk. We won't explain much how to graph because the ggplot grammar of graphs is very easy to understand and do not require much explanation.

```
# For women women aged 15-65

# yearly mean wages total and per hour
year_mean_wage <- data_ps1 %>%
  filter(age %in% 15:65, sex == 2, uhrswork > 0,
         incwage < 999998, real_wage >=0 ) %>%
  group_by(year) %>%
  summarise(mean_wage = mean(real_wage, na.rm = T),
            mean_wage_hour = mean(wage_hour, na.rm = T))

# Plot mean_wage
plot_wages_year <- ggplot(year_mean_wage, aes(x = year
                            , y = mean_wage ))+
  geom_line(color = "#967BB6")+
```

```r
  geom_point(color = "#7B1FA2")+
  theme_few()+
  labs(title = "Yearly mean real wage - Total",
       subtitle = "Women aged 15-65",
       x = "Year",
       y = "Mean wage")+
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle= element_text(hjust = 0.5),
        legend.position = "bottom") +
  scale_y_continuous(n.breaks = 10)+
  scale_x_continuous(n.breaks = 2019-2005)

# Plot mean_wage_hour
plot_wages_hour <- ggplot(year_mean_wage, aes(x = year
                                              , y = mean_wage_hour ))+
  geom_line(color = "#967BB6")+
  geom_point(color = "#7B1FA2")+
  theme_few()+
  labs(title = "Yearly mean real wage - per hour",
       subtitle = "Women aged 15-65",
       x = "Year",
       y = "Mean hourly wage")+
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle= element_text(hjust = 0.5),
        legend.position = "bottom") +
  scale_y_continuous(n.breaks = 10)+
  scale_x_continuous(n.breaks = 2019-2005)

# mean hours worked (unconditional and conditional)
  mean_hour_conditional <- data_ps1 %>%
    filter(age %in% 15:65, sex == 2, empstat == 1) %>%
    mutate(hour_worked = wkswork2*uhrswork) %>%
    group_by(year) %>%
    summarise(mean_hours = mean(hour_worked, na.rm = T))

  mean_hour_unconditional <- data_ps1 %>%
      filter(age %in% 15:65, sex == 2) %>%
      mutate(hour_worked = wkswork2*uhrswork) %>%
    group_by(year) %>%
      summarise(mean_hours = mean(hour_worked, na.rm = T))
```

```r
  # Plot hours worked
  plot_hour_worked_Cond <- ggplot(mean_hour_conditional, aes(x = year
                                                             , y = mean_hours ))+
    geom_line(color = "#967BB6")+
    geom_point(color = "#7B1FA2")+
    theme_few()+
    labs(title = "Yearly hours worked - Conditional",
         subtitle = "Women aged 15-65",
         x = "Year",
         y = "Mean hours worked")+
    theme(plot.title = element_text(hjust = 0.5),
          plot.subtitle= element_text(hjust = 0.5),
          legend.position = "bottom") +
    scale_y_continuous(n.breaks = 10)+
    scale_x_continuous(n.breaks = 2019-2005)


  plot_hour_worked_uncond <- ggplot(mean_hour_unconditional, aes(x = year
                                                                 , y = mean_hours ))+
    geom_line(color = "#967BB6")+
    geom_point(color = "#7B1FA2")+
    theme_few()+
    labs(title = "Yearly hours worked - Unconditional",
         subtitle = "Women aged 15-65",
         x = "Year",
         y = "Mean hours worked")+
    theme(plot.title = element_text(hjust = 0.5),
          plot.subtitle= element_text(hjust = 0.5),
          legend.position = "bottom") +
    scale_y_continuous(n.breaks = 10)+
    scale_x_continuous(n.breaks = 2019-2005)



# mean employment rates of women
  mean_employment <- data_ps1 %>%
    filter(age %in% 15:65, sex == 2) %>%
    group_by(year) %>%
    summarise(mean_employed = mean(labor_par, na.rm = T))

# Plot of labor participation of women
```

```r
plot_employment <-  ggplot(mean_employment, aes(x = year, y = mean_employed))+
   geom_line(color = "#967BB6")+
   geom_point(color = "#7B1FA2")+
   theme_few()+
   labs(title = "Yearly Employment Rate",
        subtitle = "Women aged 15-65",
        x = "Year",
        y = "Percent employed")+
   theme(plot.title = element_text(hjust = 0.5),
         plot.subtitle= element_text(hjust = 0.5),
         legend.position = "bottom") +
   scale_y_continuous(n.breaks = 10, labels = scales::percent)+
   scale_x_continuous(n.breaks = 2019-2005)
```

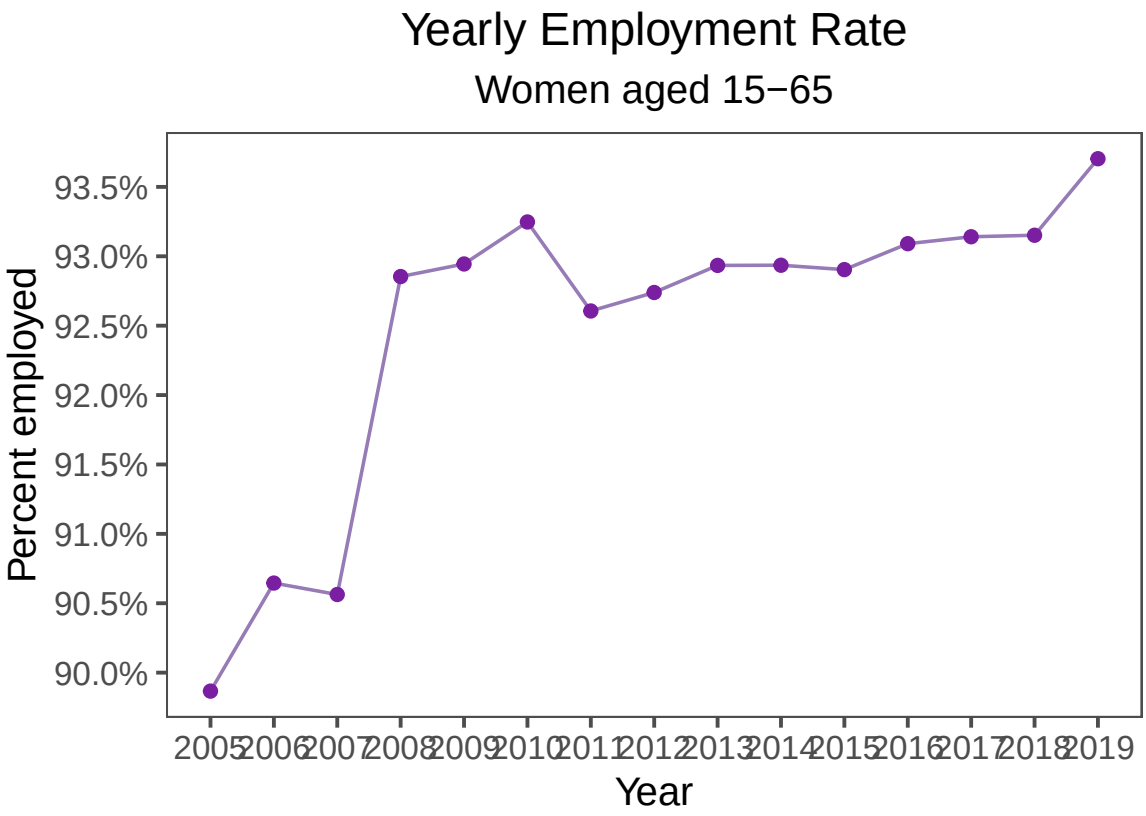

Figure 1: Mean employment - women 15-65

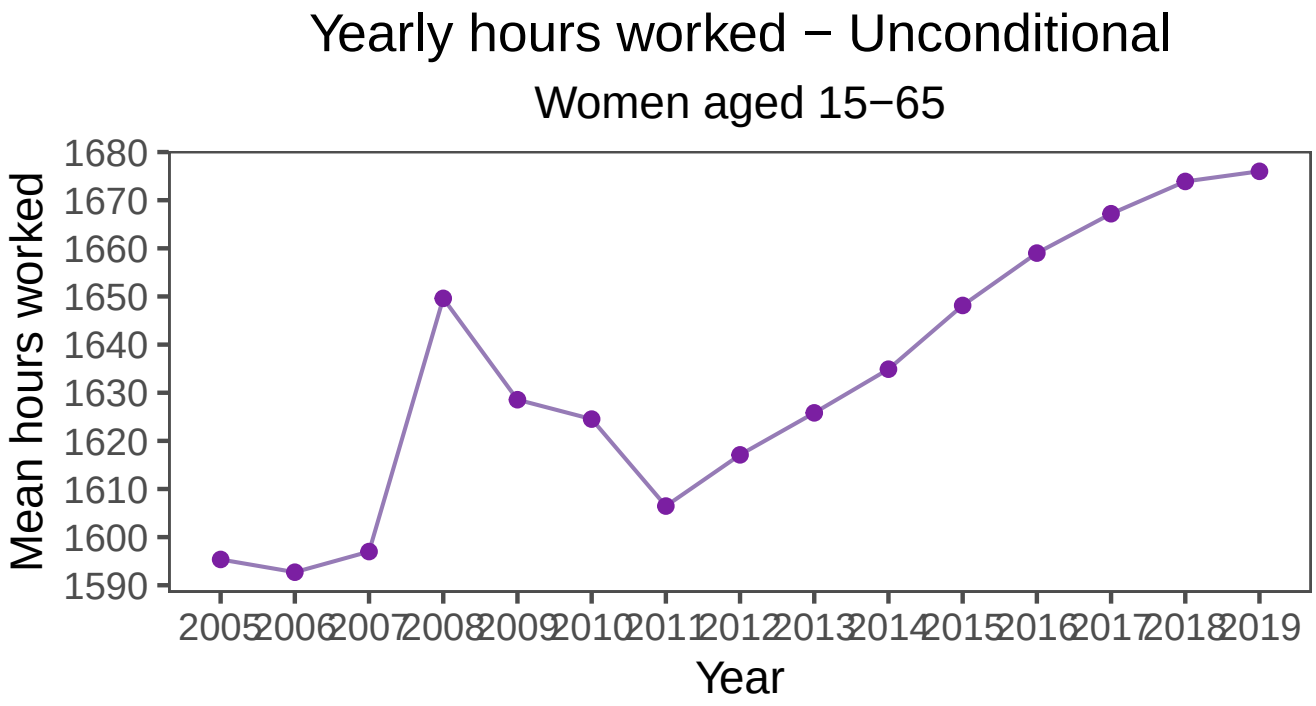

Figure 2: Hours worked per year -Unconditional - women 15-65

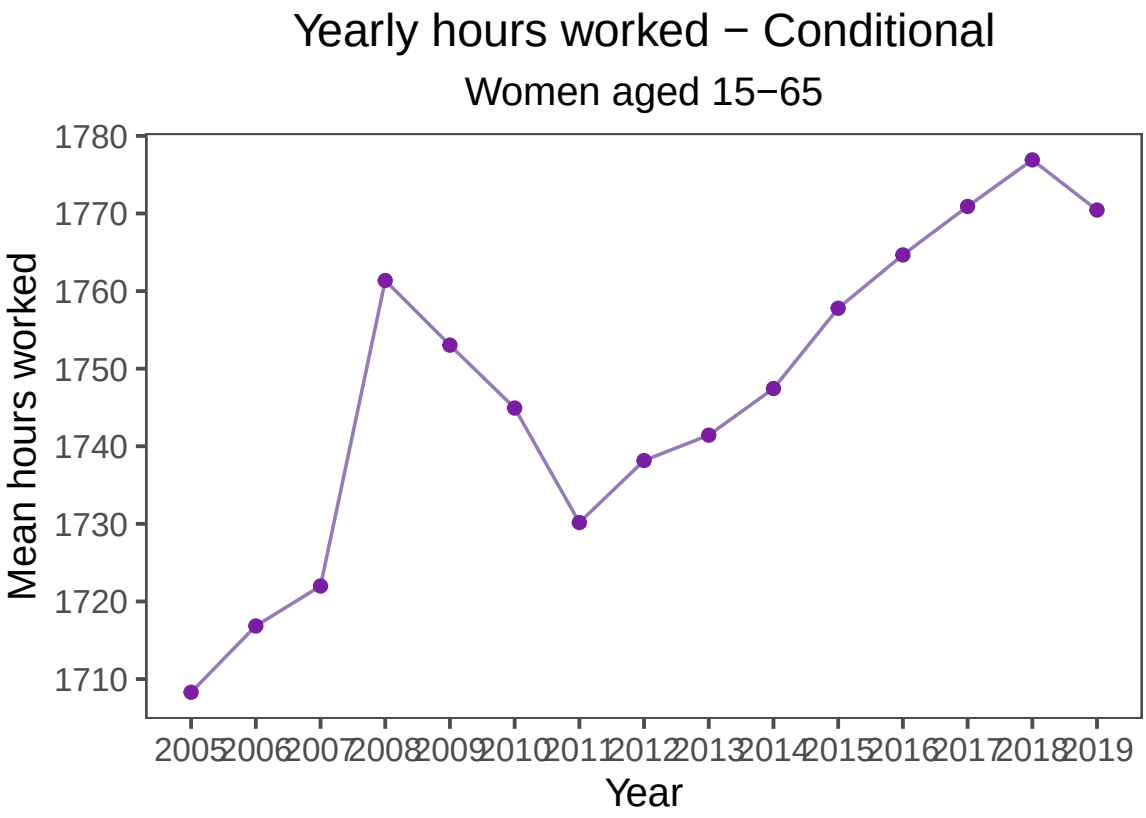

Figure 3: Hours worked per year -Unconditional - women 15-65

Figure 4: Hourly wage - women 15-65

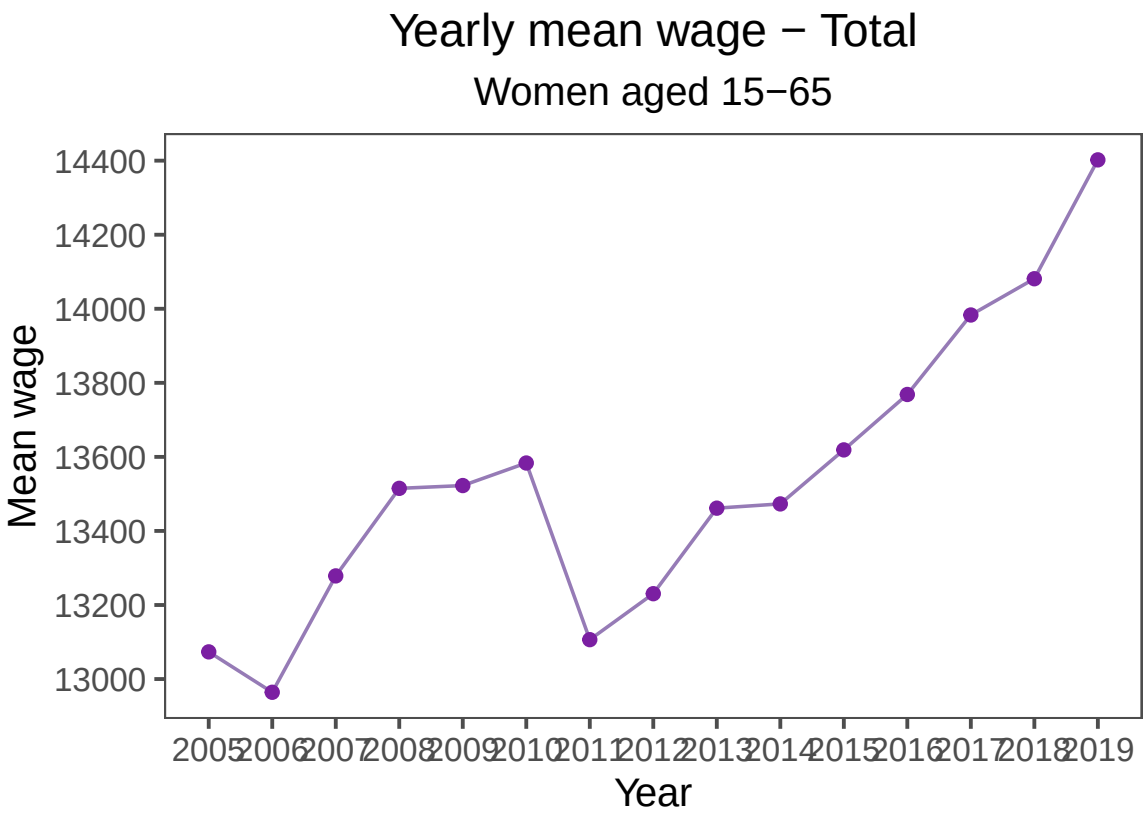

Figure 5: Yearly total wage - women 15-65

(b) Replicate Table 2 in Mincer (1962). Note: Make sure to impose comparable sample restrictions.