Data visualization

Michael Qian

2025-10-03

```r
#import data
emp_data=read.csv("C:/Users/mqian/Downloads/combined data/Combined_data_pub_Jul2023_to_Jun2025.csv")
```

Since our goal is to investigate the potential influences of the minimum wage increase announced by the Saskatchewan government to the employment rate, we can use some visualization tools to view some relationships between some variables to gain a better understanding about the employment rate.

```r
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
```

```r
#cleaning for variables
alberta_1=emp_data[emp_data$PROV=="48",]
alberta_2=alberta_1[alberta_1$LFSSTAT<=3,]
saskatchewan_1=emp_data[emp_data$PROV=="47",]
saskatchewan_2=saskatchewan_1[saskatchewan_1$LFSSTAT<=3,]
#check for na's
c(sum(is.na(alberta_2$LFSSTAT)), sum(is.na(saskatchewan_2$LFSSTAT)))
```

```
## [1] 0 0
```

```r
#add a new variable indicating whether or not the individual is employed currently
# 1 for employed 0 for umemployed
alberta_2 = alberta_2 %>%
  mutate(Emp = ifelse(LFSSTAT <= 2, 1, 0))
saskatchewan_2 = saskatchewan_2 %>%
  mutate(Emp = ifelse(LFSSTAT <= 2, 1, 0))
```

```r
#create a new data frame contains monthly employment rate by age group for alberta
alberta_youth_emp=c()
alberta_adult_emp=c()
for (i in 7:12){
  g1 = alberta_2[alberta_2$AGE_12<=2 &
                   alberta_2$SURVYEAR==2023 &
                   alberta_2$SURVMNTH==i,]
  rate = (sum(g1$Emp))/(nrow(g1))
  alberta_youth_emp = append(alberta_youth_emp, rate)
  g11 = alberta_2[alberta_2$AGE_12>=3 &
                   alberta_2$SURVYEAR==2023 &
                   alberta_2$SURVMNTH==i,]
  rate1 = (sum(g11$Emp))/(nrow(g11))
  alberta_adult_emp = append(alberta_adult_emp, rate1)
}
for (i in 1:12){
  g2 = alberta_2[alberta_2$AGE_12<=2 &
                   alberta_2$SURVYEAR==2024 &
                   alberta_2$SURVMNTH==i,]
  rate = (sum(g2$Emp))/(nrow(g2))
  alberta_youth_emp = append(alberta_youth_emp, rate)
  g22 = alberta_2[alberta_2$AGE_12>=3 &
                   alberta_2$SURVYEAR==2024 &
                   alberta_2$SURVMNTH==i,]
  rate1 = (sum(g22$Emp))/(nrow(g22))
  alberta_adult_emp = append(alberta_adult_emp, rate1)
}
for (i in 1:6){
  g3 = alberta_2[alberta_2$AGE_12<=2 &
                   alberta_2$SURVYEAR==2025 &
                   alberta_2$SURVMNTH==i,]
  rate = (sum(g3$Emp))/(nrow(g3))
  alberta_youth_emp = append(alberta_youth_emp, rate)
  g33 = alberta_2[alberta_2$AGE_12>=3 &
                   alberta_2$SURVYEAR==2025 &
                   alberta_2$SURVMNTH==i,]
  rate1 = (sum(g33$Emp))/(nrow(g33))
  alberta_adult_emp = append(alberta_adult_emp, rate1)
}
Date=c(2023.07,2023.08,2023.09,2023.10,2023.11,2023.12,2024.01,2024.02,2024.03,
       2024.04,2024.05,2024.06,2024.07,2024.08,2024.09,2024.10,2024.11,2024.12,2025.01,
       2025.02,2025.03,2025.04,2025.05,2025.06)
df_alyouth=data.frame(Date, alberta_youth_emp, c(rep("Alberta", 24)))
df_aladult=data.frame(Date, alberta_adult_emp, c(rep("Alberta", 24)))
colnames(df_alyouth)=c("Date", "Youth Employment rate", "Province")
colnames(df_aladult)=c("Date", "Adult Employment rate", "Province")
```

Visualization 1

First we can use a time series to visualize the employment rate change over the past two years (July 2023 to June 2025) for young employers and adult employers in two provinces Alberta and Saskatchewan as control groups

This is the time series for monthly youth Employment rate in Alberta.

```
#plot the time series
ts_alyouth=ts(df_alyouth$`Youth Employment rate`, start = c(2023, 7), frequency = 12)
plot(ts_alyouth, main = "Monthly Youth Employment rate in Alberta Time Series",
     ylab = "Youth Employment Rate", xlab = "Time")
abline(v = 2024.7, col="red")
```

Monthly Youth Employment rate in Alberta Time Series
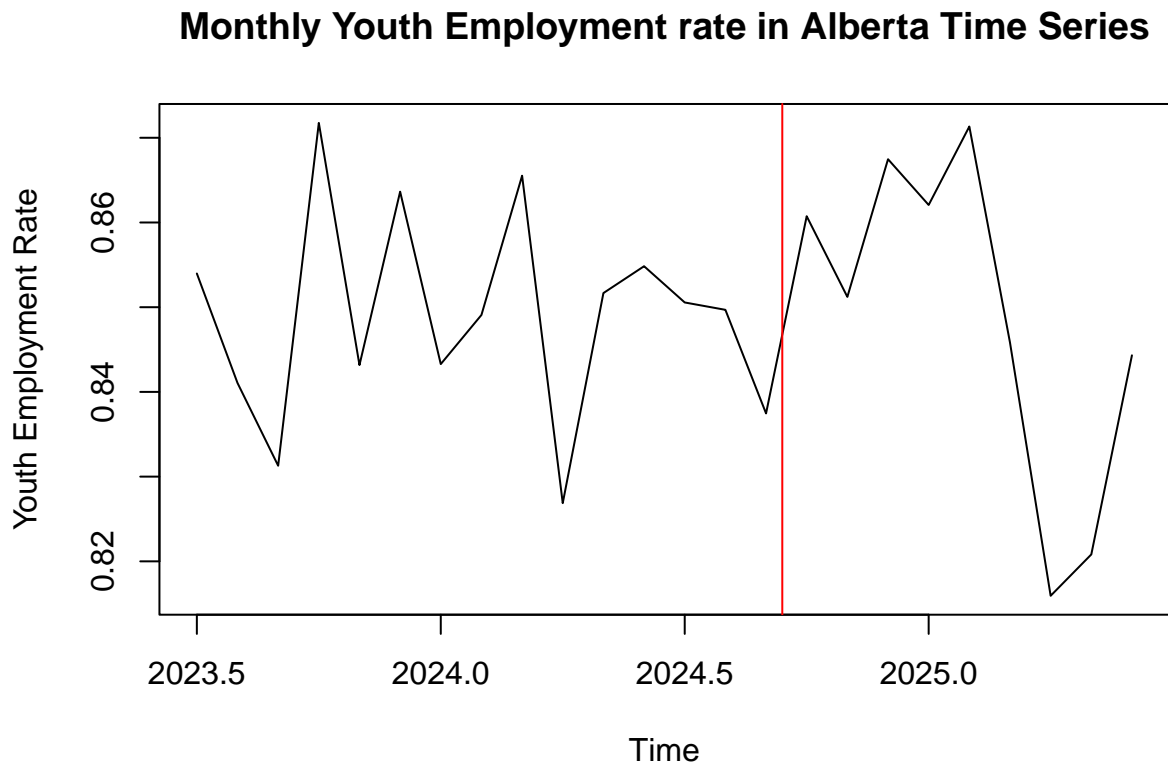
This is the time series for monthly adult Employment rate in Alberta.

```
ts_aladult=ts(df_aladult$`Adult Employment rate`, start = c(2023, 7), frequency = 12)
plot(ts_aladult, main = "Monthly Adult Employment rate in Alberta Time Series",
     ylab = "Adult Employment Rate", xlab = "Time")
abline(v = 2024.7, col="red")
```

Monthly Adult Employment rate in Alberta Time Series
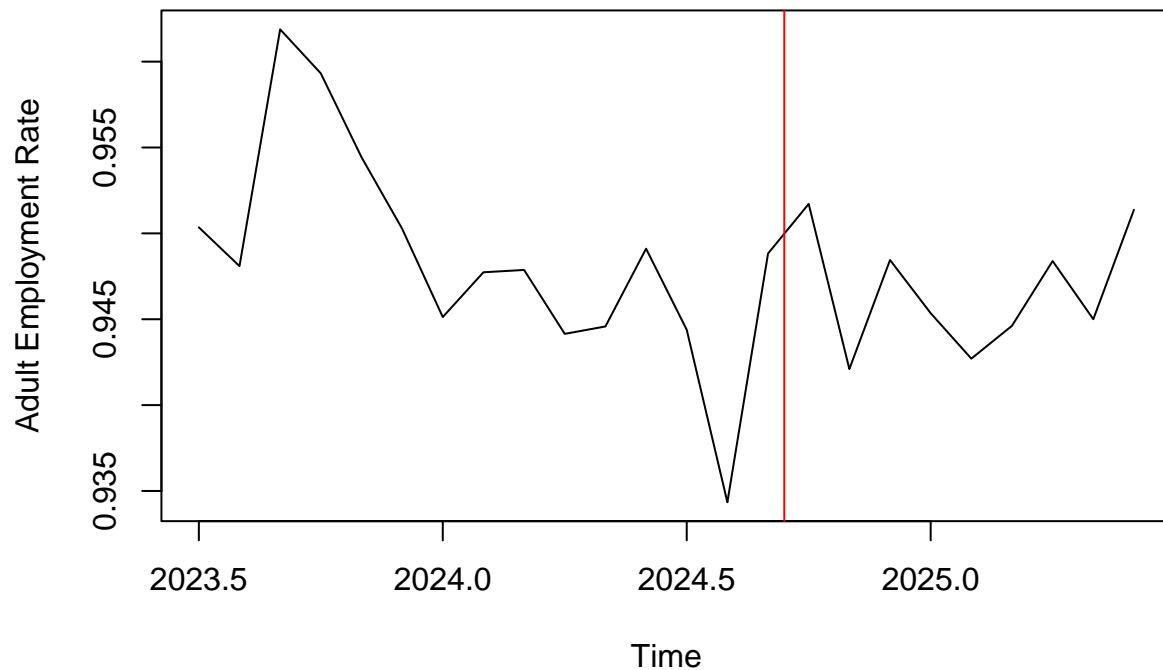

```r
#create a new data frame contains monthly employment rate by age group for Saskatchewan
sas_youth_emp=c()
sas_adult_emp=c()
for (i in 7:12){
  g1 = saskatchewan_2[saskatchewan_2$AGE_12<=2 &
                      saskatchewan_2$SURVYEAR==2023 &
                      saskatchewan_2$SURVMNTH==i,]
  rate = (sum(g1$Emp))/(nrow(g1))
  sas_youth_emp = append(sas_youth_emp, rate)
  g11 = saskatchewan_2[saskatchewan_2$AGE_12>=3 &
                      saskatchewan_2$SURVYEAR==2023 &
                      saskatchewan_2$SURVMNTH==i,]
  rate1 = (sum(g11$Emp))/(nrow(g11))
  sas_adult_emp = append(sas_adult_emp, rate1)
}
for (i in 1:12){
  g2 = saskatchewan_2[saskatchewan_2$AGE_12<=2 &
                      saskatchewan_2$SURVYEAR==2024 &
                      saskatchewan_2$SURVMNTH==i,]
  rate = (sum(g2$Emp))/(nrow(g2))
  sas_youth_emp = append(sas_youth_emp, rate)
  g22 = saskatchewan_2[saskatchewan_2$AGE_12>=3 &
                      saskatchewan_2$SURVYEAR==2024 &
                      saskatchewan_2$SURVMNTH==i,]
  rate1 = (sum(g22$Emp))/(nrow(g22))
```

```
  sas_adult_emp = append(sas_adult_emp, rate1)
}
for (i in 1:6){
  g3 = saskatchewan_2[saskatchewan_2$AGE_12<=2 &
                      saskatchewan_2$SURVYEAR==2025 &
                      saskatchewan_2$SURVMNTH==i,]
  rate = (sum(g3$Emp))/(nrow(g3))
  sas_youth_emp = append(sas_youth_emp, rate)
  g33 = saskatchewan_2[saskatchewan_2$AGE_12>=3 &
                      saskatchewan_2$SURVYEAR==2025 &
                      saskatchewan_2$SURVMNTH==i,]
  rate1 = (sum(g33$Emp))/(nrow(g33))
  sas_adult_emp = append(sas_adult_emp, rate1)
}
df_sasyouth=data.frame(Date, sas_youth_emp, c(rep("Saskatchewan", 24)))
df_sasadult=data.frame(Date, sas_adult_emp, c(rep("Saskatchewan", 24)))
colnames(df_sasyouth)=c("Date", "Youth Employment rate", "Province")
colnames(df_sasadult)=c("Date", "Adult Employment rate", "Province")
```
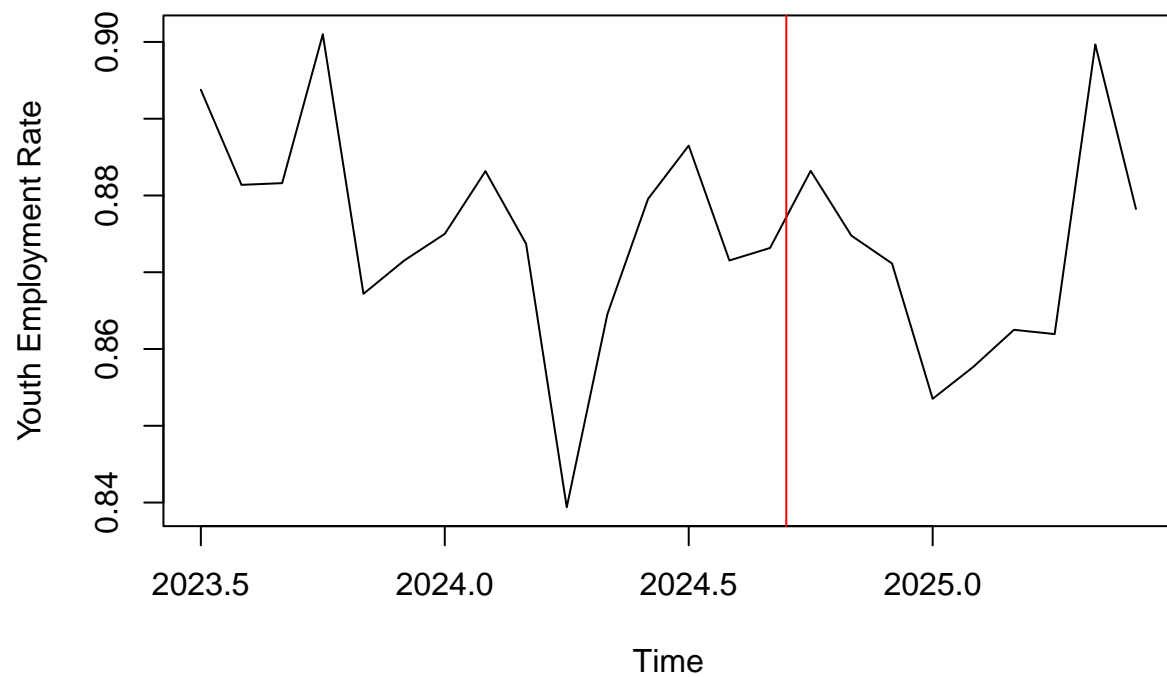
This is the time series for monthly youth Employment rate in Saskatchewan.

```
#plot the time series
ts_sasyouth=ts(df_sasyouth$`Youth Employment rate`, start = c(2023, 7), frequency = 12)
plot(ts_sasyouth, main = "Monthly Youth Employment rate in Saskatchewan Time Series",
     ylab = "Youth Employment Rate", xlab = "Time")
abline(v = 2024.7, col="red")
```

Monthly Youth Employment rate in Saskatchewan Time Series

This is the time series for monthly adult Employment rate in Saskatchewan.

```
ts_sasadult=ts(df_sasadult$`Adult Employment rate`, start = c(2023, 7), frequency = 12)
plot(ts_sasadult, main = "Monthly Adult Employment rate in Saskatchewan Time Series",
     ylab = "Adult Employment Rate", xlab = "Time")
abline(v = 2024.7, col="red")
```

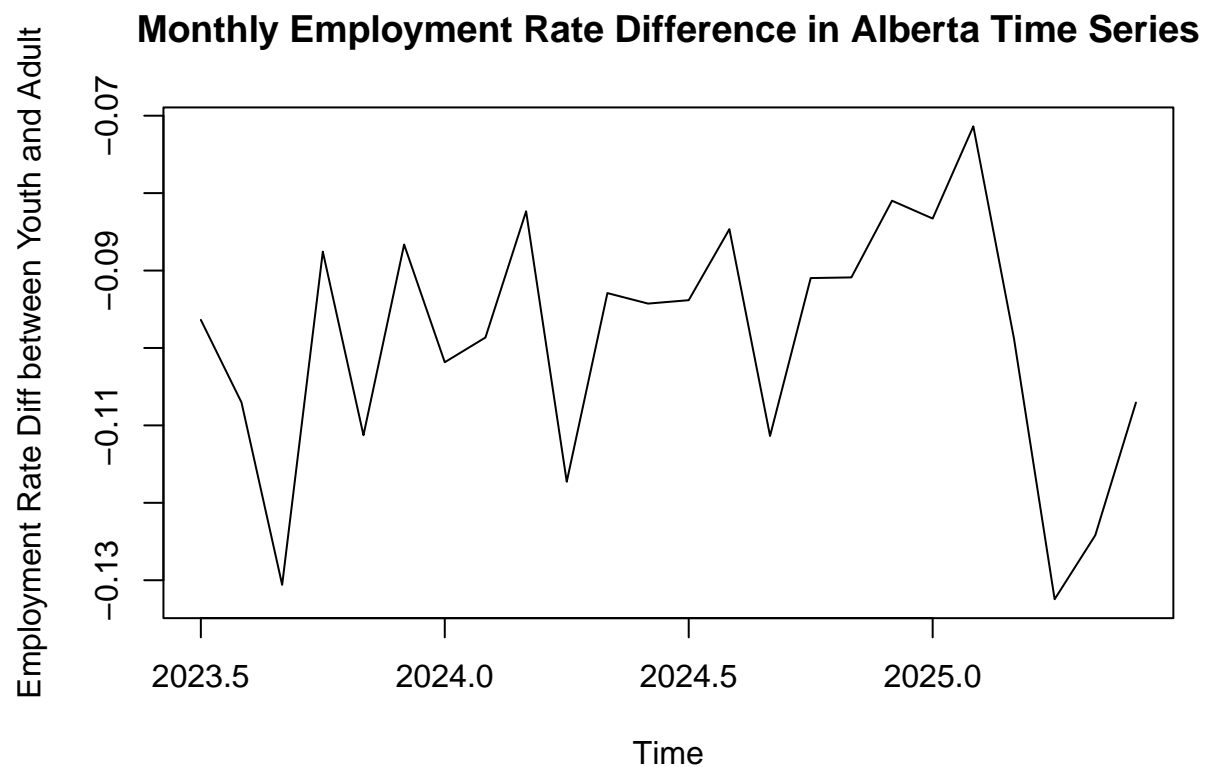## Monthly Adult Employment rate in Saskatchewan Time Series
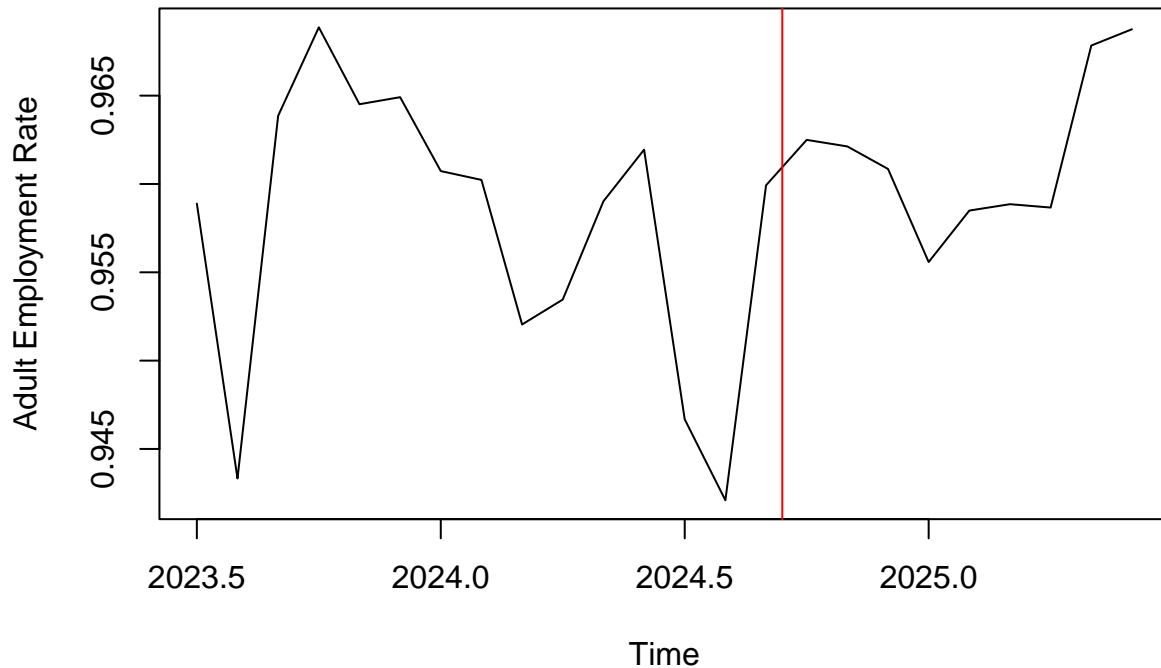
From all four plots showing different employment rate trend, all the employment rate changes over the 24 months period seem not very steady. But the variation is actually not that large. By inspection of these plots, we can see that for both provinces, the overall youth employment rate are generally lower than the adult employment rate. In all these four plots, the red vertical line indicates the time that the government of Saskatchewan announces the minimum wage increase which is around July 2024. We can see that the adult employment rate in Saskatchewan seems to having a positive trend after the announcement, but the youth employment rate seems not to be affected. $Visualization$ 2

For another visualization, we can look at the difference between the youth employment rate and the adult employment rate in both provinces to see a more clear relationship between these employment rates.

This is the time series for monthly Employment rate difference in Alberta.

```r
n=nrow(df_alyouth)
alberta_diff=numeric(n)
for (i in 1:n){
  alberta_diff[i] = df_alyouth$`Youth Employment rate`[i] -
    df_aladult$`Adult Employment rate`[i]
}
ts_aldiff=ts(alberta_diff, start = c(2023, 7), frequency = 12)
plot(ts_aldiff, main = "Monthly Employment Rate Difference in Alberta Time Series",
     ylab = "Employment Rate Diff between Youth and Adult", xlab = "Time")
```

Monthly Employment Rate Difference in Alberta Time Series

This is the time series for monthly Employment rate difference in Saskatchewan.

```
n1=nrow(df_sasyouth)
sas_diff=numeric(n1)
for (i in 1:n){
  sas_diff[i] = df_sasyouth$`Youth Employment rate`[i] -
    df_sasadult$`Adult Employment rate`[i]
}
ts_sasdiff=ts(sas_diff, start = c(2023, 7), frequency = 12)
plot(ts_sasdiff, main = "Monthly Employment Rate Difference in Saskatchewan Time Series",
     ylab = "Employment Rate Diff between Youth and Adult", xlab = "Time")
```

Monthly Employment Rate Difference in Saskatchewan Time Series
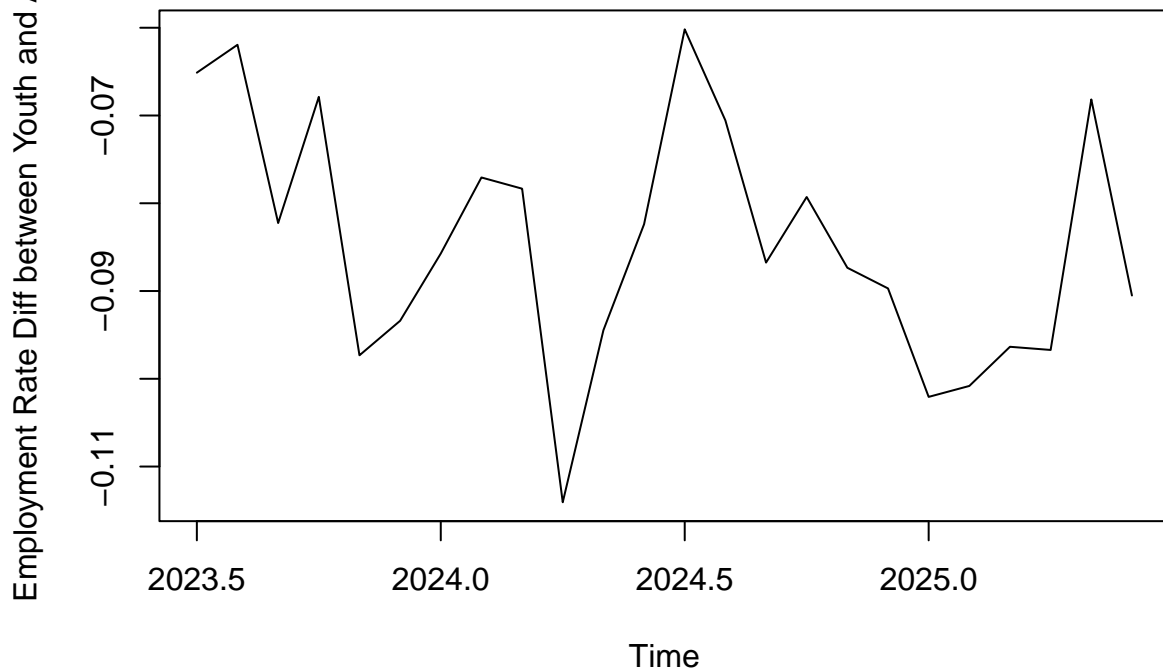
Hence from both plots, we can clearly observe the difference between the youth employment rate and the adult employment rate in both provinces. Both of the trends seems to be in a negative range which is another indication that the youth employment rate is generally lower than the adult employment rate in both provinces over the 24 months period.

Visualization 3

We want to compare the 12-month pre/post minimum wage raise differences between youth employment rate. By using two scatter plots where the left plot shows the hourly wages trend and the youth employment rate trend from the pre 12-month window (July 2023 to June 2024), the right plot shows the hourly wages trend and the youth employment rate trend from the post 12-month window (July 2024 to June 2025).

```
#add a variable monthly average hourly wage to the data set
sas_youth_wage=c()
for (i in 7:12){
  s1 = saskatchewan_2[saskatchewan_2$AGE_12<=2 &
                  saskatchewan_2$SURVYEAR==2023 &
                  saskatchewan_2$SURVMNTH==i,]
  avg1 = mean(s1$HRLYEARN, na.rm = TRUE)
  sas_youth_wage = append(sas_youth_wage, avg1)
}
for (i in 1:12){
  s1 = saskatchewan_2[saskatchewan_2$AGE_12<=2 &
                  saskatchewan_2$SURVYEAR==2024 &
                  saskatchewan_2$SURVMNTH==i,]
  avg1 = mean(s1$HRLYEARN, na.rm = TRUE)
  sas_youth_wage = append(sas_youth_wage, avg1)
```

```
}
for (i in 1:6){
  s1 = saskatchewan_2[saskatchewan_2$AGE_12<=2 &
                      saskatchewan_2$SURVYEAR==2025 &
                      saskatchewan_2$SURVMNTH==i,]
  avg1 = mean(s1$HRLYEARN, na.rm = TRUE)
  sas_youth_wage = append(sas_youth_wage, avg1)
}
df_sasyouth=df_sasyouth %>%
  mutate(Hourly_wage = sas_youth_wage)
```
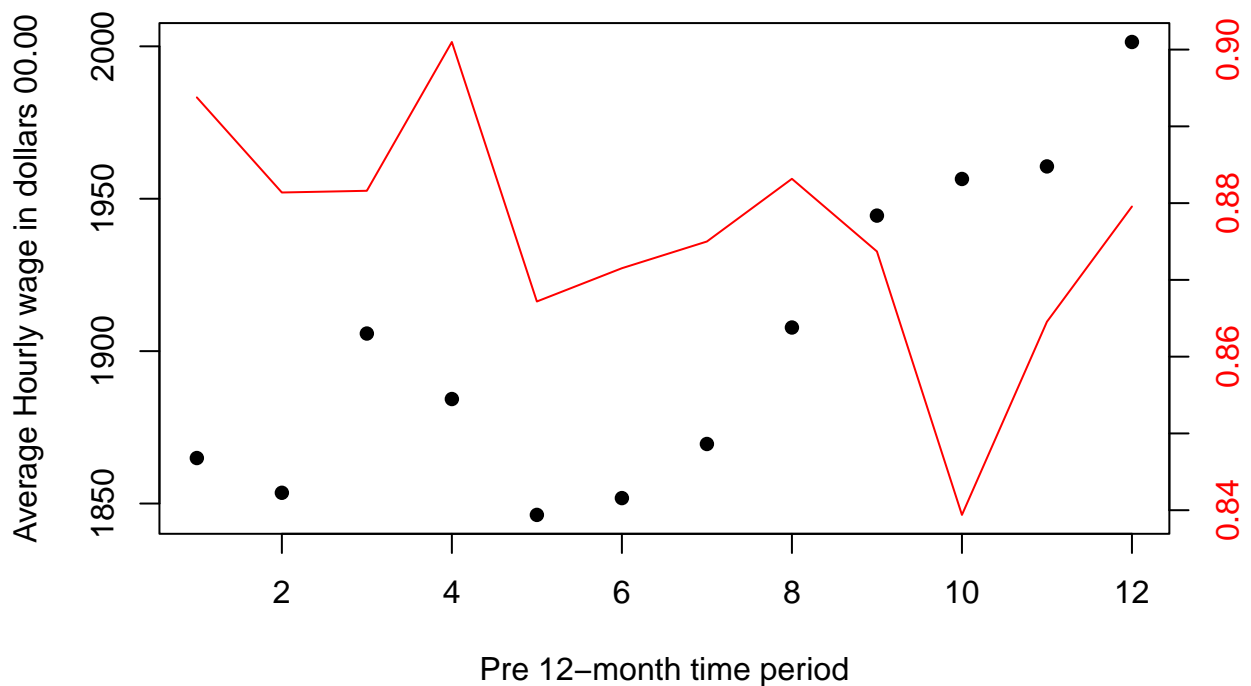
```
#plot the scatter plot and line plot for pre 12-month period
plot(df_sasyouth[1:12,]$Hourly_wage, pch = 16, xlab = "Pre 12-month time period",
     ylab = "Average Hourly wage in dollars 00.00",
     main = "Hourly wage and Employment rate comparison for youth group")
par(new=TRUE)
plot(df_sasyouth[1:12,]$`Youth Employment rate`, type = "l", col = "red", axes = FALSE,
     xlab="", ylab="")
axis(side=4, col.axis="red")
mtext("Youth Employment rate", side=4, line=3, col="red")
```


Hourly wage and Employment rate comparison for youth group

```
#plot the scatter plot and line plot for post 12-month period
plot(df_sasyouth[13:24,]$Hourly_wage, pch = 16, xlab = "Post 12-month time period",
     ylab = "Average Hourly wage in dollars 00.00",
     main = "Hourly wage and Employment rate comparison for youth group")
par(new=TRUE)
plot(df_sasyouth[13:24,]$`Youth Employment rate`, type = "l",
     col = "red", axes = FALSE,
     xlab="", ylab="")
axis(side=4, col.axis="red")
mtext("Youth Employment rate", side=4, line=3, col="red")
```

Hourly wage and Employment rate comparison for youth group
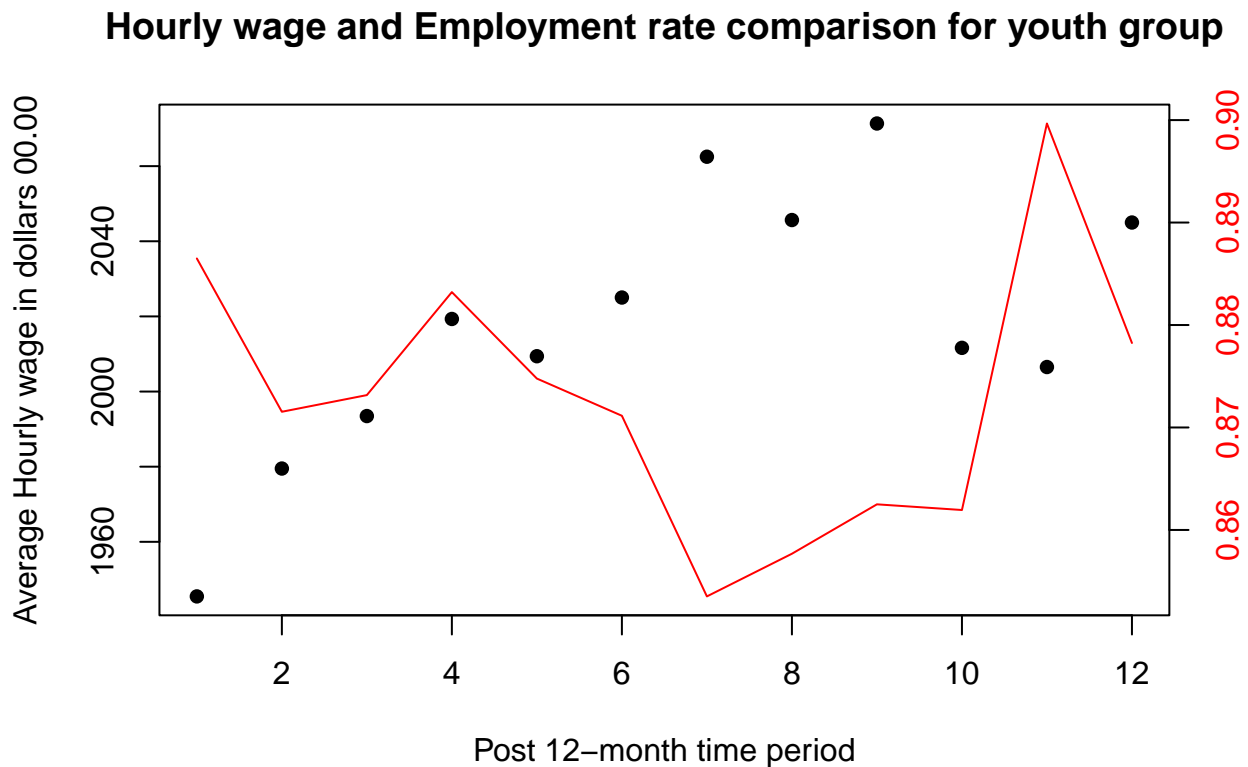

```
par(mfrow=c(1,1))
#full data set
df_alberta=merge(df_aladult, df_alyouth %>% select(-Province), by="Date")
df_saskatchewan=merge(df_sasadult, df_sasyouth %>% select(-Province), by="Date",
                      all = TRUE)
```

By observing these two comparison plots, we can see after the annoucement of minimum wage raise in Saskatchewan, there is a small period of time between January to July when the employment rate of young people seem to have a decreasing trend but rises afterwards.