UsualWeeklyWages

2023-07-19

Introduction to Usual Weekly Wages

This is an analysis of the Usual Weekly Earnings of Wage and Salary Workers from the Bureau of Labor Statistics (BLS), which is data that comes from the Current Population Survey (CPS). Earnings data is collected from 1/4 of the CPS monthly sample. Full time workers are defined as working 35 or more hours per week. All wages in this analysis are for full-time workers only.

In order to improve the reliability of the estimates, a 4 quarter moving average (12-month moving average) for all the data is used. Otherwise, the data is simply too noisy, since the sample size of the CPS monthly survey that includes earnings data (just 1/4 of the sample) is too small.

The analysis includes all of the available data from Table 5, includes data from the 1st quarter of 2000 to the 2nd quarter of 2023: https://www.bls.gov/webapps/legacy/cpswktab5.htm

Wages are adjusted for inflation using the CPIAUCSL, which is the Consumer Price Index for Urban Consumers. https://fred.stlouisfed.org/series/CPIAUCSL

Important notice about race and ethnicity: "White", "Black", and "Asian" includes workers of all ethnicities, including those who are Hispanic or Latino. Most Hispanic or Latino workers are included under the "White" category (in many data sets many Hispanic or Latino are included under the "Some Other Race" category, but there isn't such a category here). Hispanic or Latino may be of any race.

Takeaways from the following charts:

1) Wage growth from 2019-2022 were stronger for lower wage workers than higher wage workers.

2) The gender wage gap has declined for all percentiles from 2000-2022.

3) The growth of female full-time workers with advanced degrees is much faster than the growth of male full-time workers with advanced degrees.

4) Hispanic and Asian wages have been growing faster than average. Black and White wages have not grown faster than average or slower than average.

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
##
## 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(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.2.3

library(tidyr)

## Warning: package 'tidyr' was built under R version 4.2.3

data = readxl::read_excel("Data/SeriesReport-20230719035728_307369.xlsx", skip = 3)
inflation = read.csv("Data/CPIAUCSL (1).csv",
                     col.names = c("Date", "Factor"))
inflation$Date = as.Date(inflation$Date)
names(data) = c("ID", as.character(seq(as.Date("2000/1/1"), as.Date("2023/12/1"), by="quarter")))
names

## function (x)  .Primitive("names")

recode_names = expand.grid(Type = c("Count","P10","P25","P50","P75","P90"), Next = c("Total","Male","Fer

recode_names_v2 = expand.grid(Type = c("Count","P10","P25","P50","P75","P90"), Race=
c("Total25","LessThanHS","HS","SomeCollege","Bachelor's+","Bachelor'sOnly","AdvancedDegree"), Next =
c("Total", "Male", "Female"))

names = bind_rows(recode_names, recode_names_v2)


data$Type = names$Type
data$Race = names$Race
data$Sex = names$Next

data = data %>%
  pivot_longer(cols = 2:97, names_to = "Date", values_to = "Value")
data$Date = as.Date(data$Date)

data = data %>%
  left_join(inflation, by = "Date") %>%
  mutate(InflationAdj = Value/Factor)
data = data %>%
  mutate(InflationAdj = ifelse(Type == "Count", Value, InflationAdj)) %>%
  group_by(Type, Race, Sex) %>%
  mutate(MovAverage = zoo::rollapply(InflationAdj, 4, mean, align = 'right', fill = NA))

relative_to_white = data %>%
  filter(!is.na(MovAverage)) %>%
  group_by(Type, Sex, Date) %>%
  summarize(Asian = 100*MovAverage[Race == "Asian"]/MovAverage[Race == "White"],
            Black = 100*MovAverage[Race == "Black"]/MovAverage[Race == "White"],
            All = 100*MovAverage[Race == "All"]/MovAverage[Race == "White"],
            Hispanic = 100*MovAverage[Race == "Hispanic"]/MovAverage[Race == "White"])
```

```
## 'summarise()' has grouped output by 'Type', 'Sex'. You can override using the
## '.groups' argument.

relative_to_date = data %>%
  filter(!is.na(MovAverage)) %>%
  group_by(Type, Sex, Race) %>%
  summarize(Index = 100*MovAverage[Date == "2023-04-01"]/MovAverage[Date == "2019-10-01"])


## 'summarise()' has grouped output by 'Type', 'Sex'. You can override using the
## '.groups' argument.

growth = relative_to_date %>%
  filter(Type %in%  c("P10","P25","P50","P75","P90") & Sex != "Total" &
           Race %in% c("White", "Asian","Black","Hispanic","All")) %>%
  ggplot(aes(x = Type, y = Index, color = Race, linetype = Sex,
             group = interaction(Race, Sex))) + geom_line(linewidth = 1) +
  geom_point(size = 2) + xlab("Percentile") +
  ylab("Index of Usual Weekly Wages (2019 = 100)") +
  scale_y_continuous(breaks = seq(98, 110, 2)) + theme_bw() +
  ggtitle("Real Wages (2022 Q3-2023 Q2) as a Percent of 2019 by Race and Sex")


relative_to_date2 = data %>%
  filter(!is.na(MovAverage)) %>%
  group_by(Type, Sex, Race) %>%
  summarize(Index = 100*MovAverage[Date == "2023-04-01"]/MovAverage[Date == "2007-10-01"])


## 'summarise()' has grouped output by 'Type', 'Sex'. You can override using the
## '.groups' argument.

growth2 = relative_to_date2 %>%
  filter(Type %in%  c("P10","P25","P50","P75","P90") & Sex != "Total" &
           Race %in% c("White", "Asian","Black","Hispanic","All")) %>%
  ggplot(aes(x = Type, y = Index, color = Race, linetype = Sex,
             group = interaction(Race, Sex))) + geom_line(linewidth = 1) +
  geom_point(size = 2) + xlab("Percentile") +
  ylab("Index of Usual Weekly Wages (2007 = 100)") +
  scale_y_continuous(breaks = seq(98, 130, 4)) + theme_bw() +
  ggtitle("Real Wages (2022 Q3-2023 Q2) as a Percent of 2007 by Race and Sex")

okay = expand.grid(Type = c("P10","P25","P50","P75","P90"), Race = c("Asian","Black","Hispanic","All"))
nice = list()

plot_this = function(type, race){
  title = paste0("Percent of White Wage for ", race, " for ", type, " Percentile")
  relative_to_white %>%
  filter(Type == type) %>%
  ggplot(aes(x = Date, y = !!sym(race), color = Sex)) +
  geom_line(size = 1) + theme_bw() + scale_x_date(date_breaks = "2 years", date_labels = "%Y") +
  xlab("Date") + ylab("Percent of White Wage") +
  ggtitle(title)
```
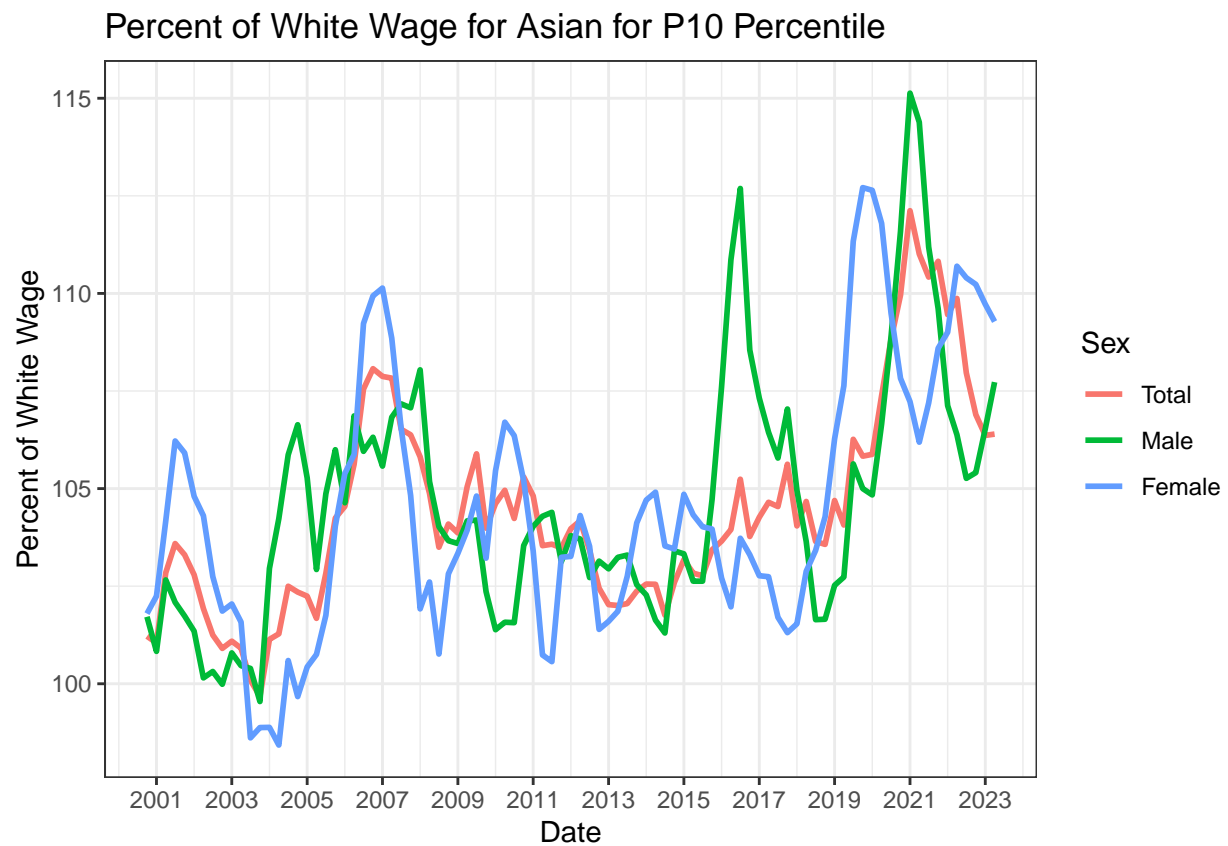
```
}

for(i in seq_along(okay[[1]])){
  nice[[i]] = plot_this(as.character(okay[i, 1]), as.character(okay[i, 2]))
}
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```
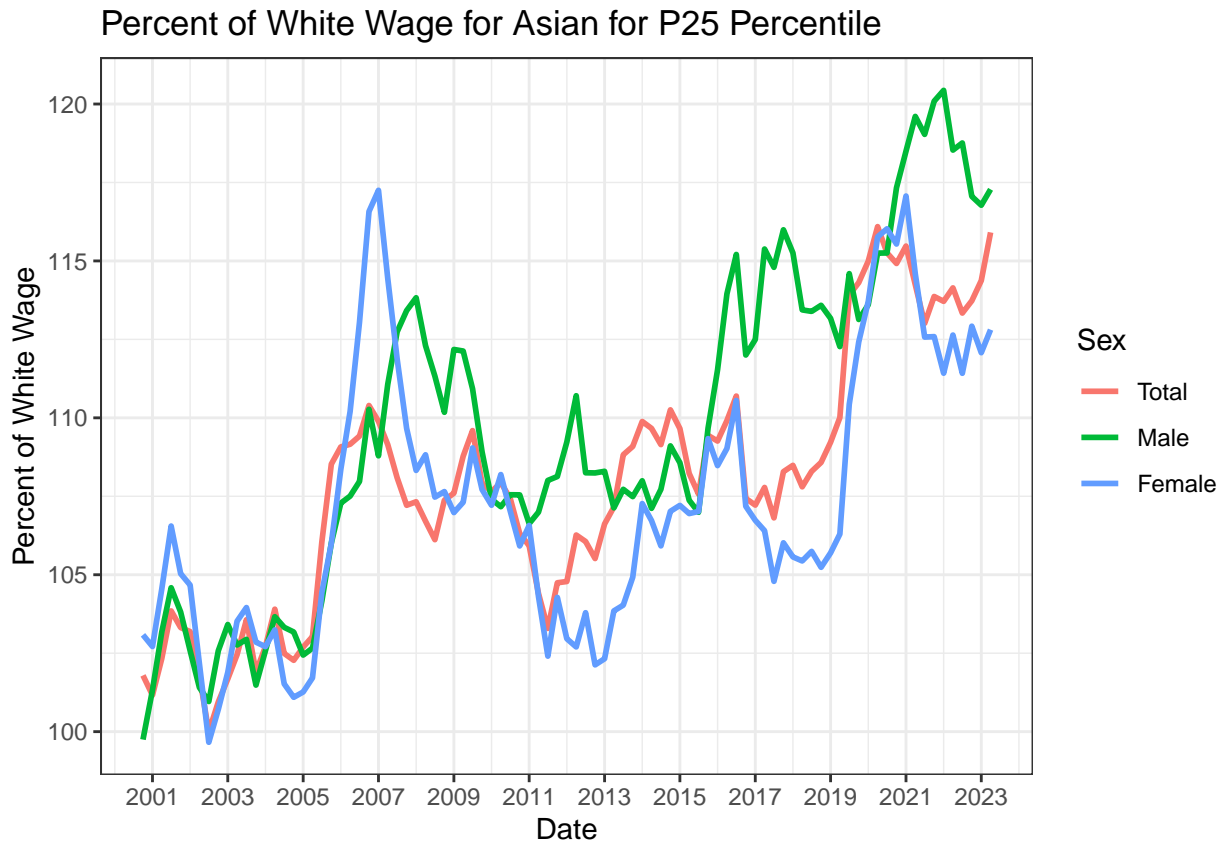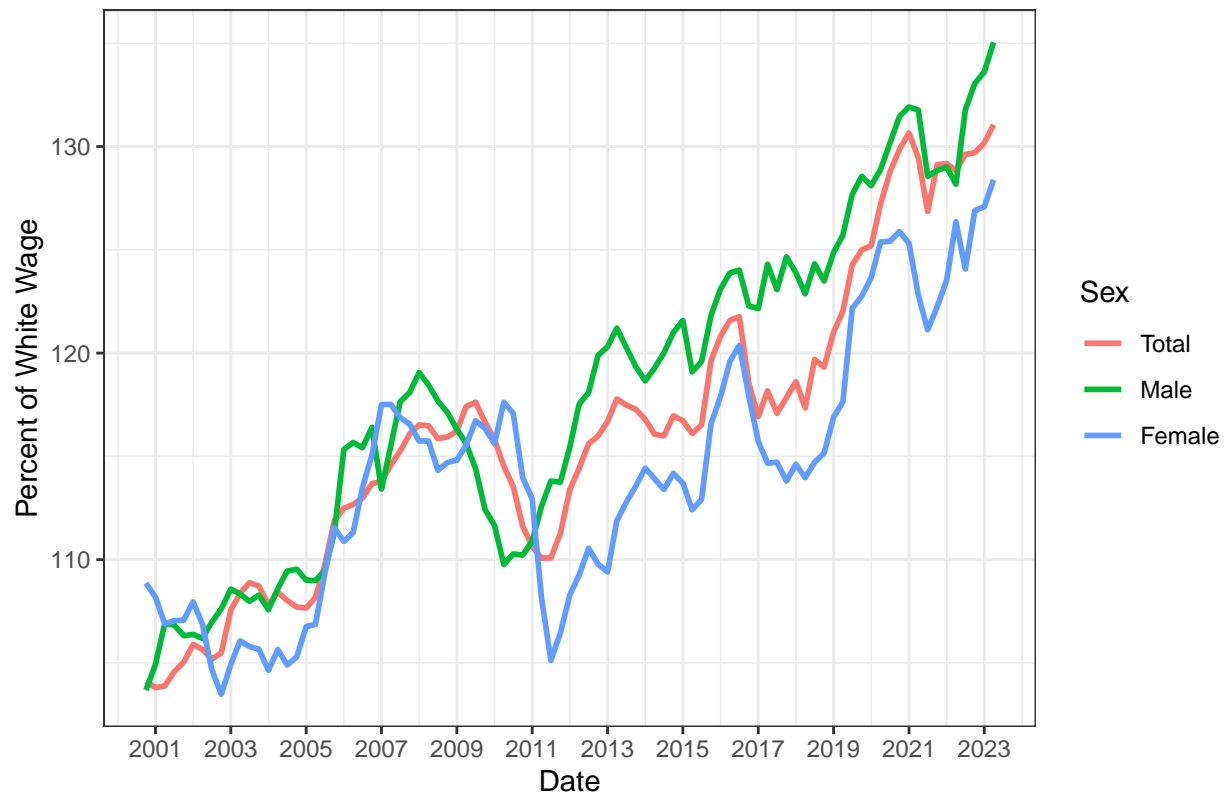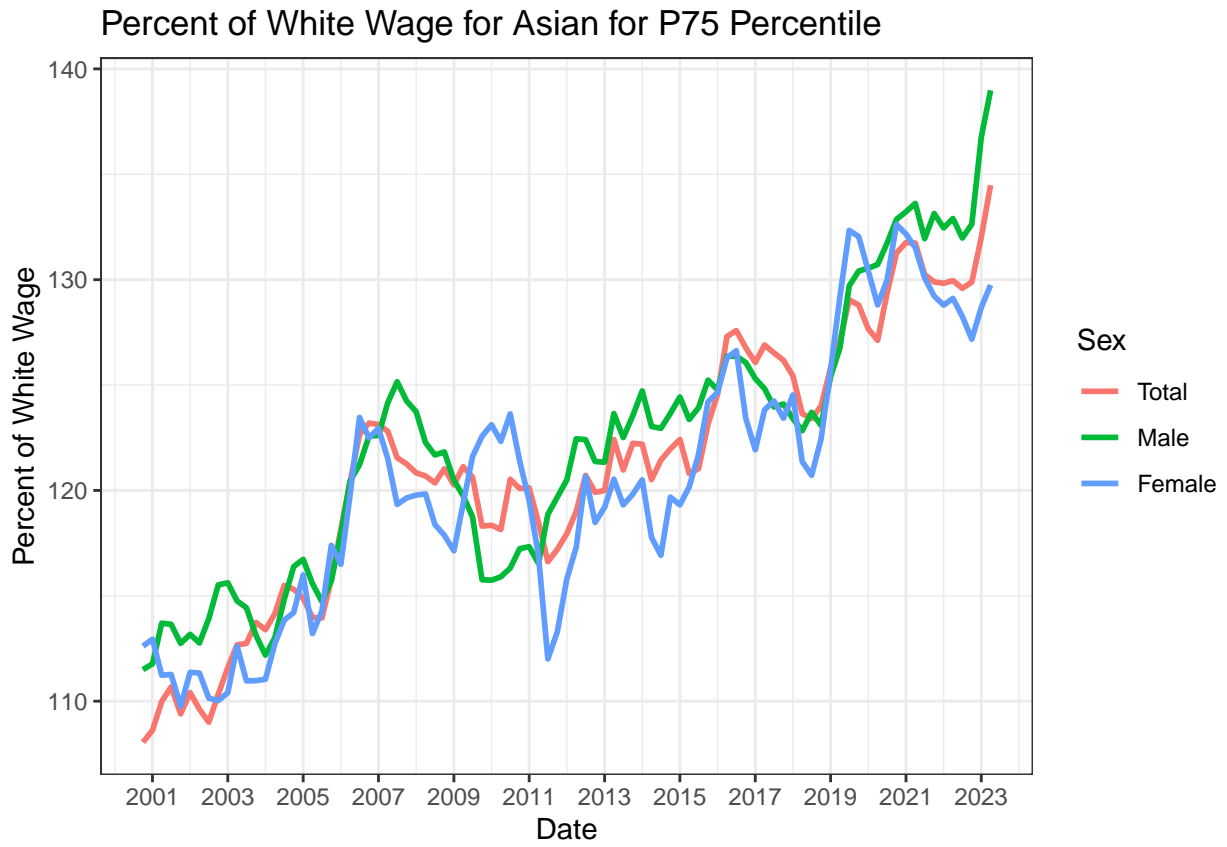
```
nice
```

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

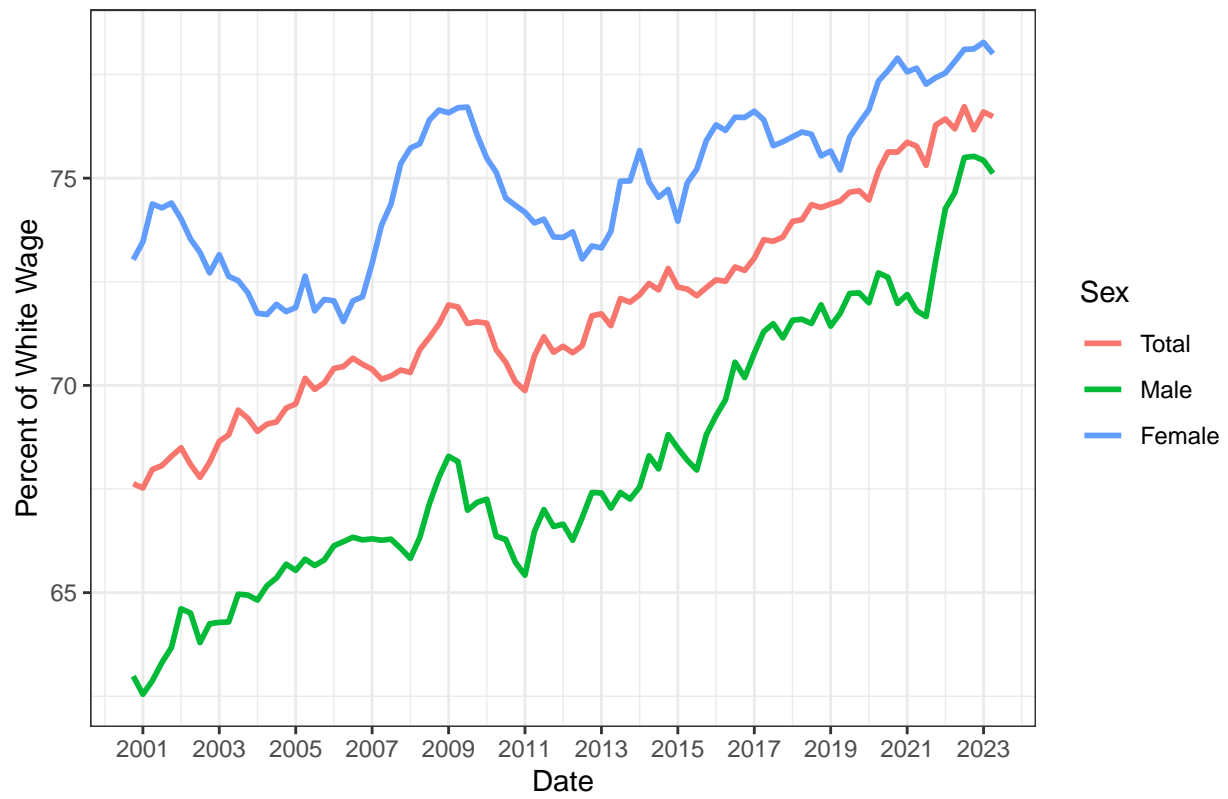### Percent of White Wage for Asian for P10 Percentile


```
##
## [[2]]
```

Percent of White Wage for Asian for P25 Percentile
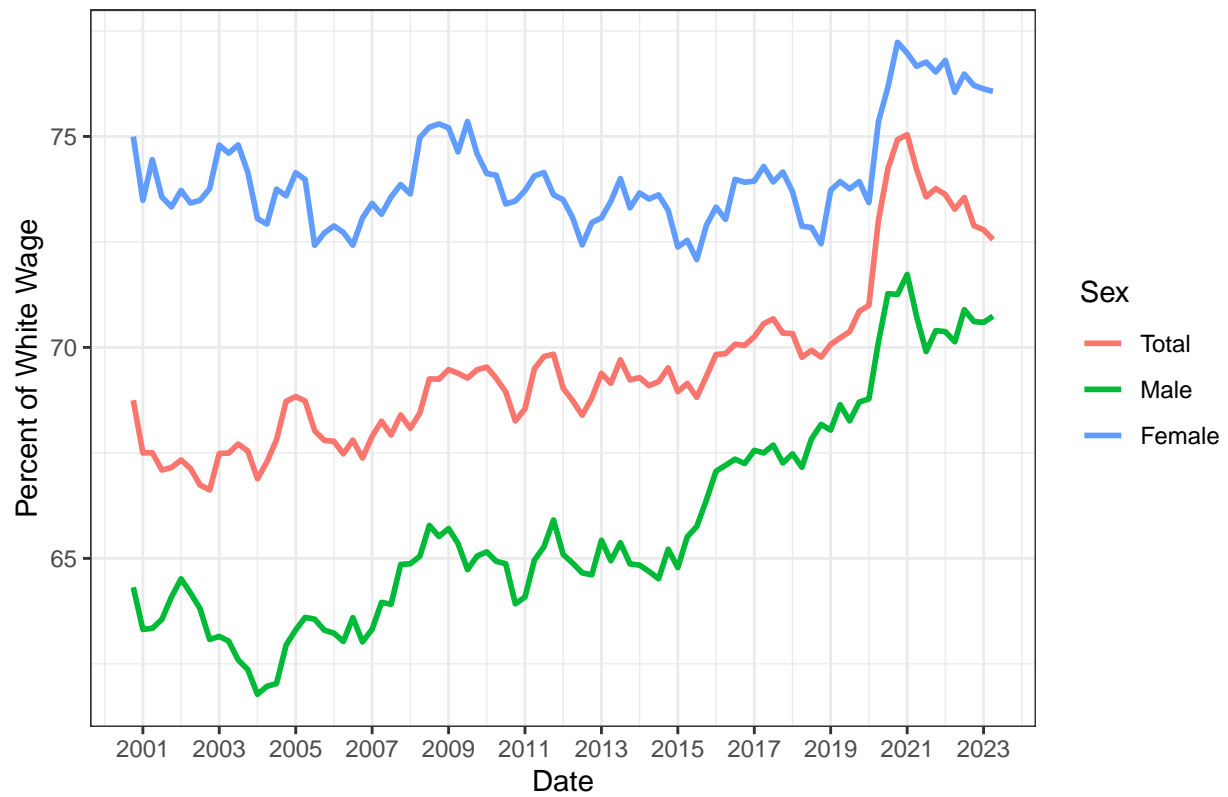


```
## 
## [[3]]
```
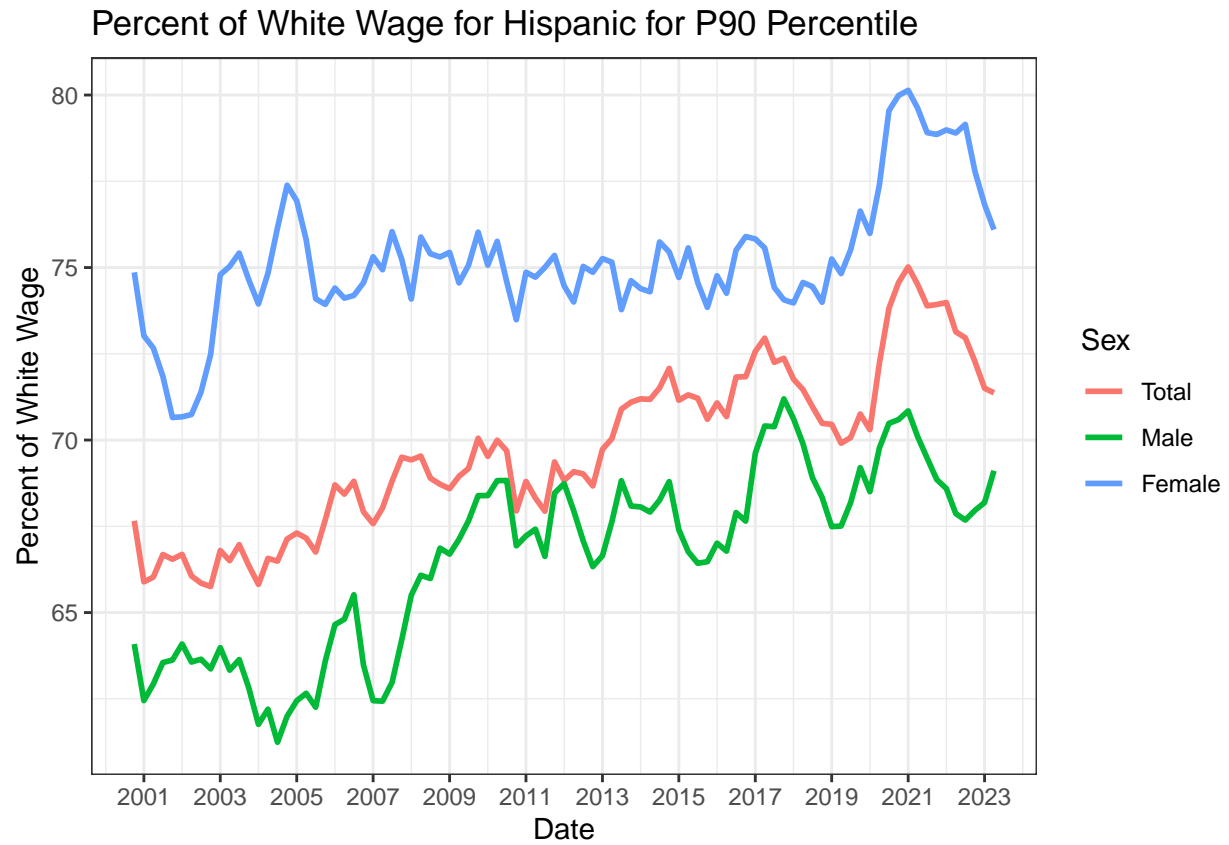
Percent of White Wage for Asian for P50 Percentile


```
## 
## [[4]]
```
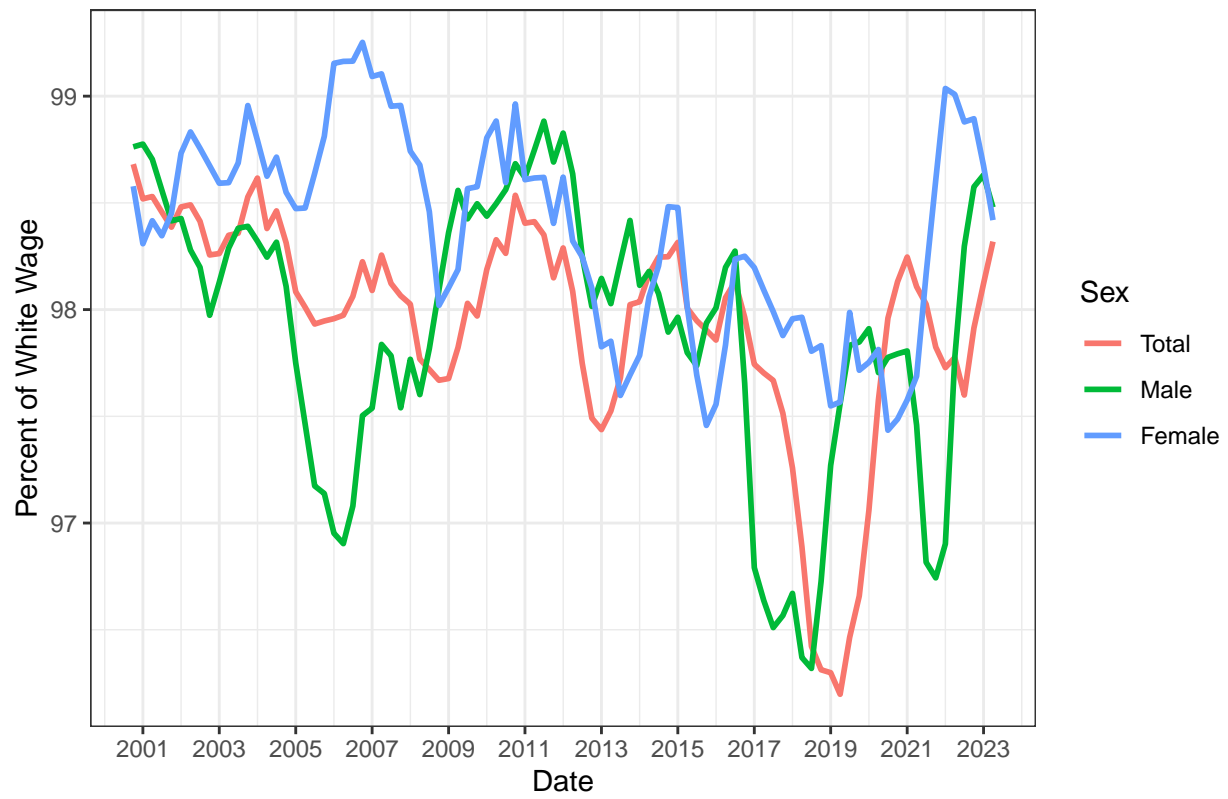
Percent of White Wage for Asian for P75 Percentile


```
## 
## [[5]]
```
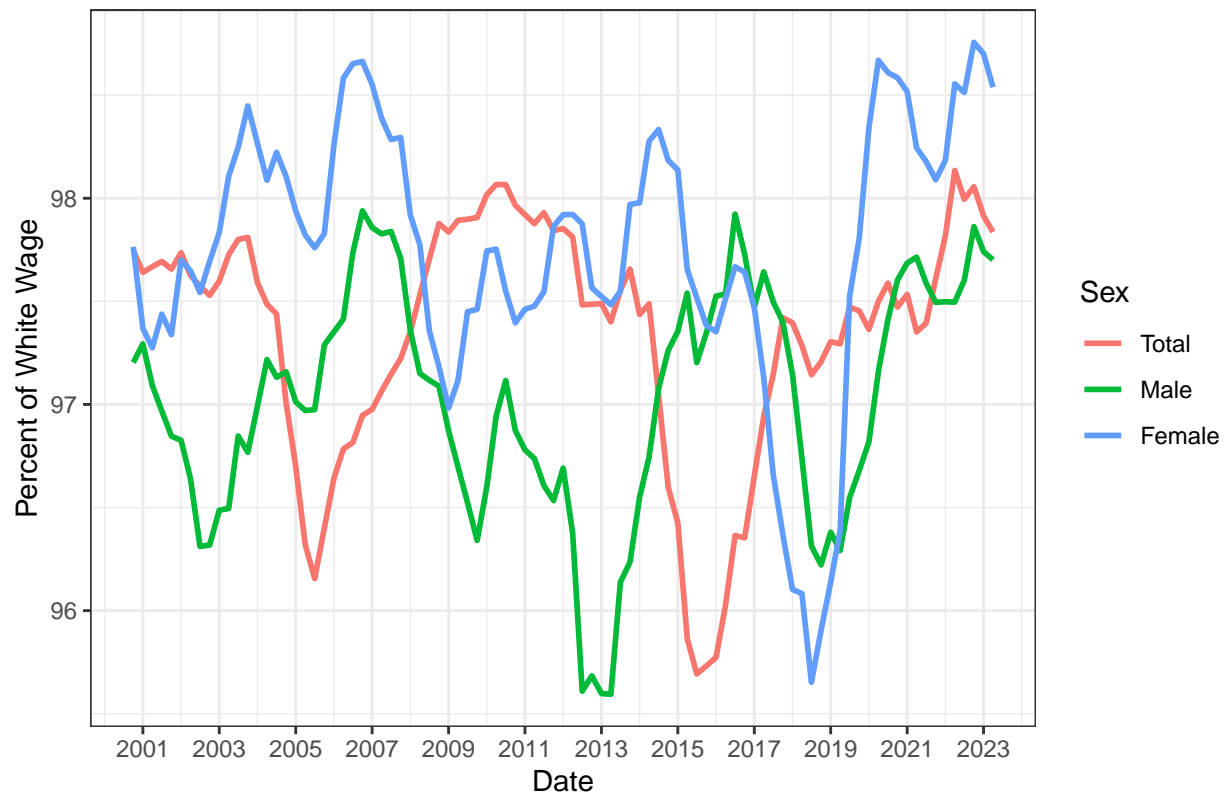
Percent of White Wage for Asian for P90 Percentile


```
##
## [[6]]
```
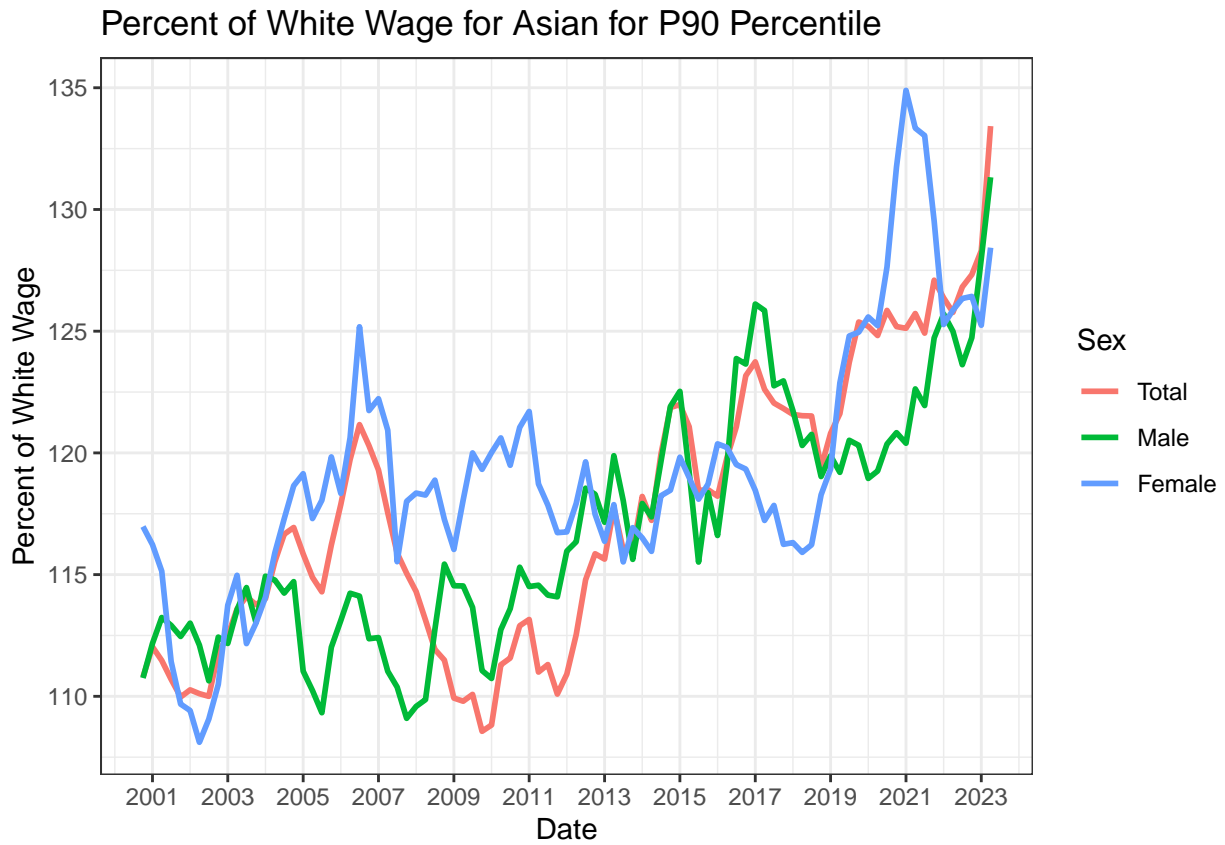
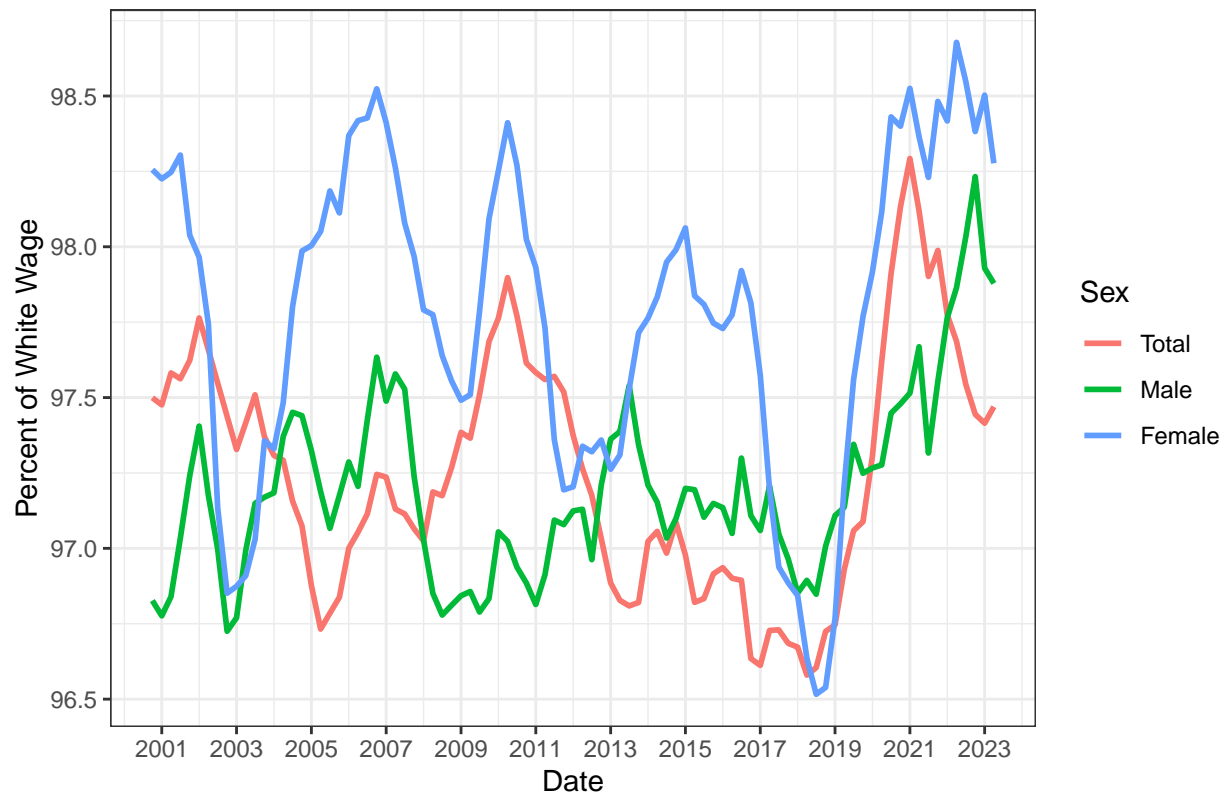## Percent of White Wage for Black for P10 Percentile


```
##
## [[7]]
```
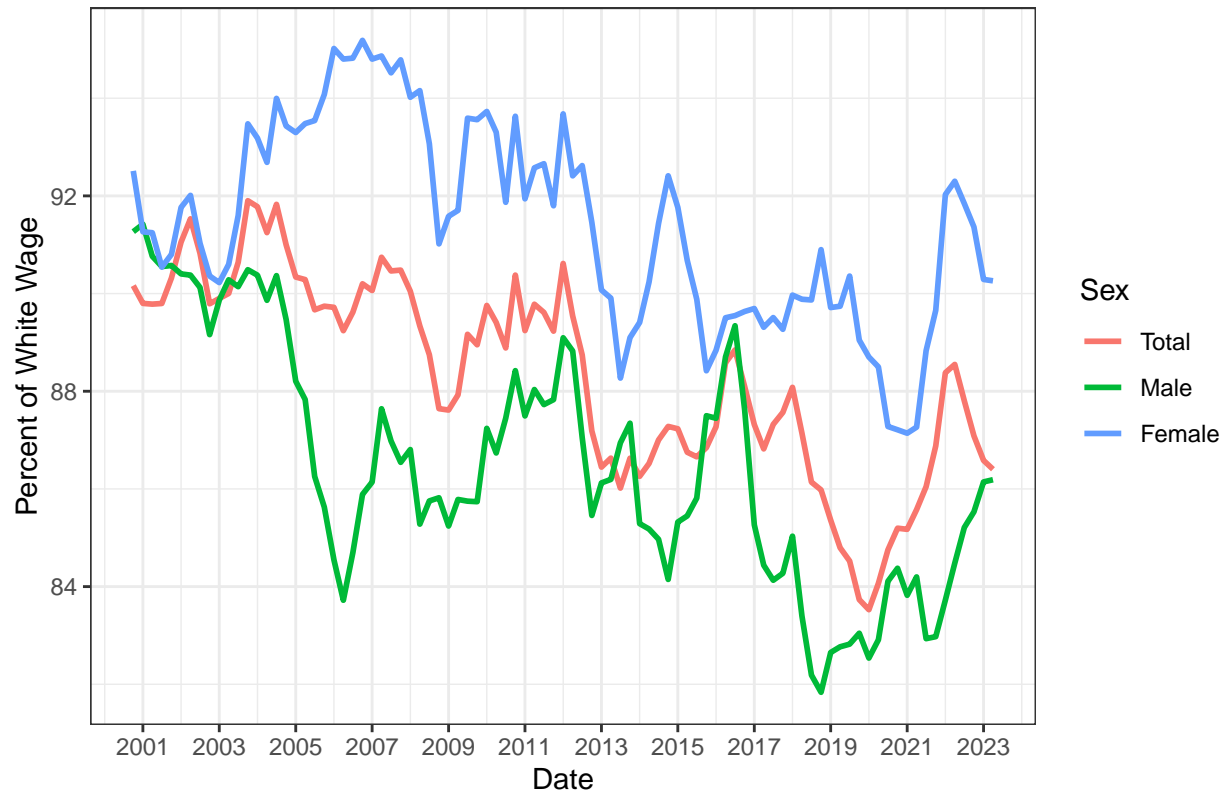
Percent of White Wage for Black for P25 Percentile


```
## 
## [[8]]
```
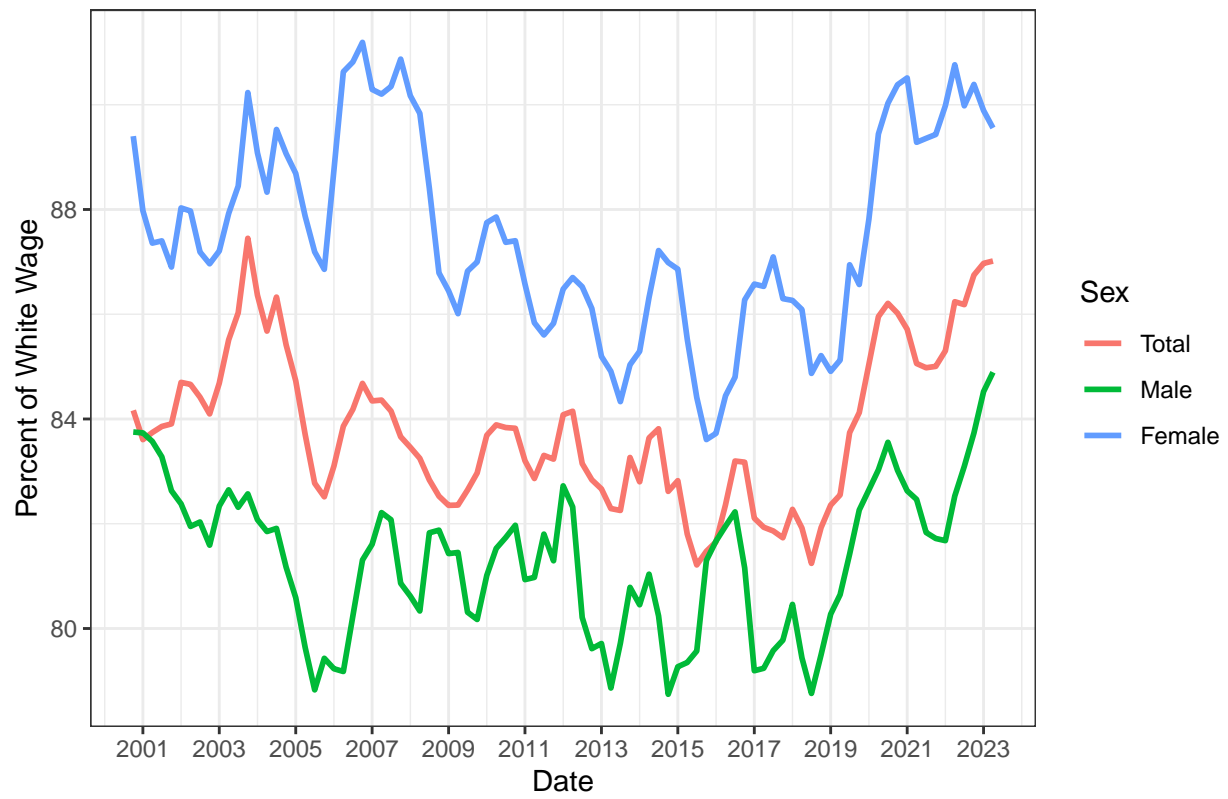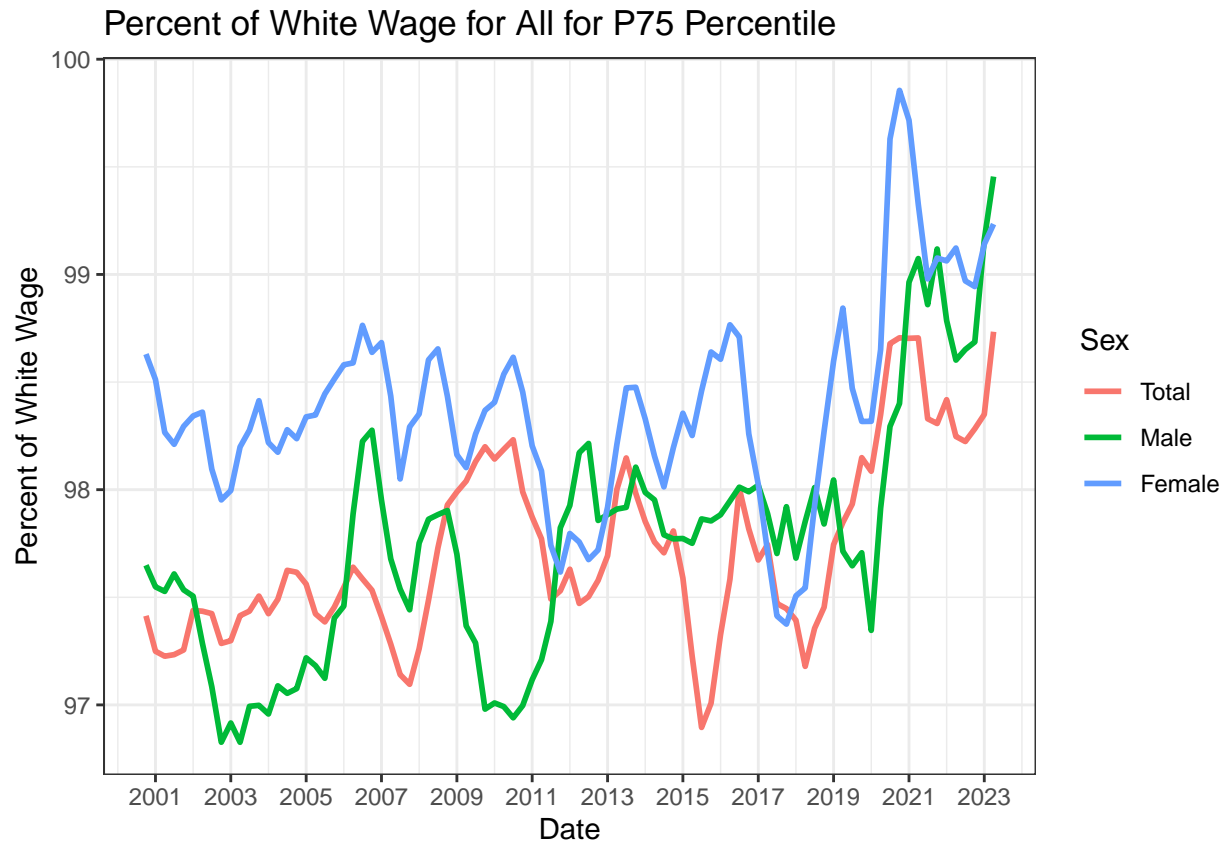
Percent of White Wage for Black for P50 Percentile


```
## 
## [[9]]
```

Percent of White Wage for Black for P75 Percentile

```
##
## [[10]]
```

Percent of White Wage for Black for P90 Percentile


```
##
## [[11]]
```

Percent of White Wage for Hispanic for P10 Percentile
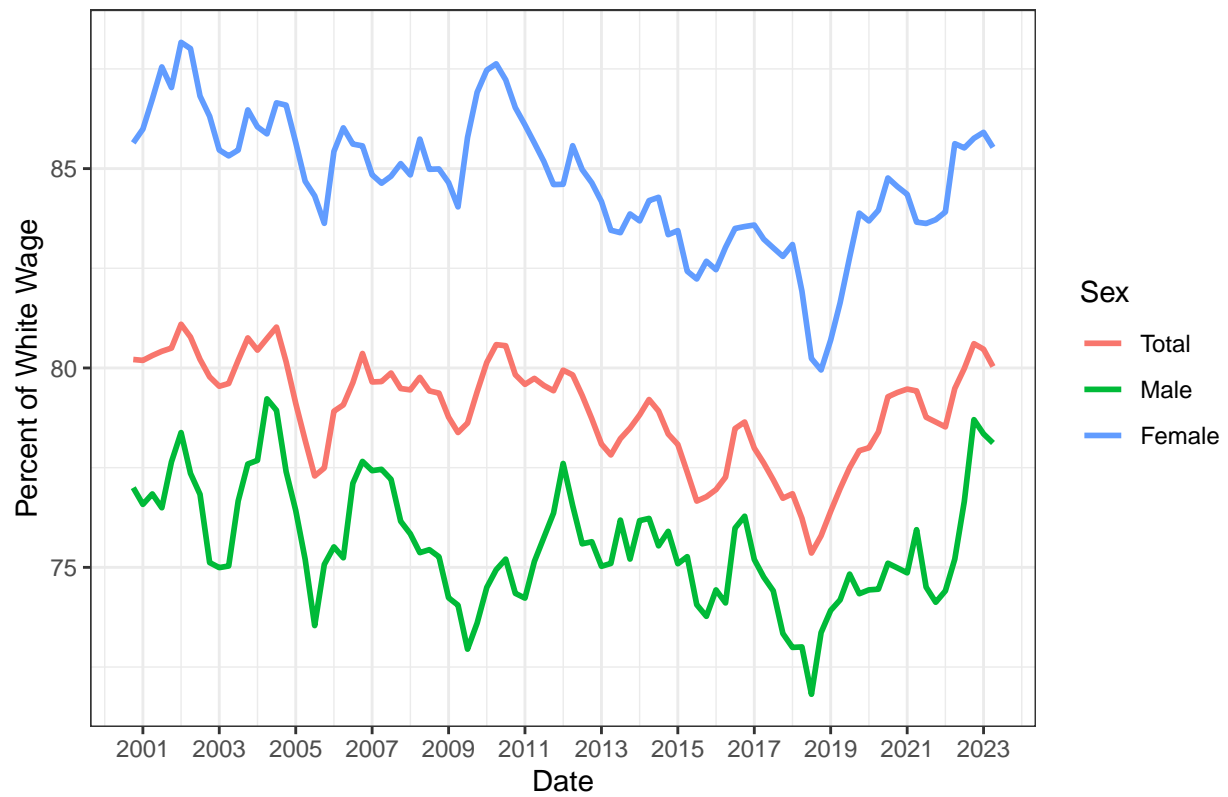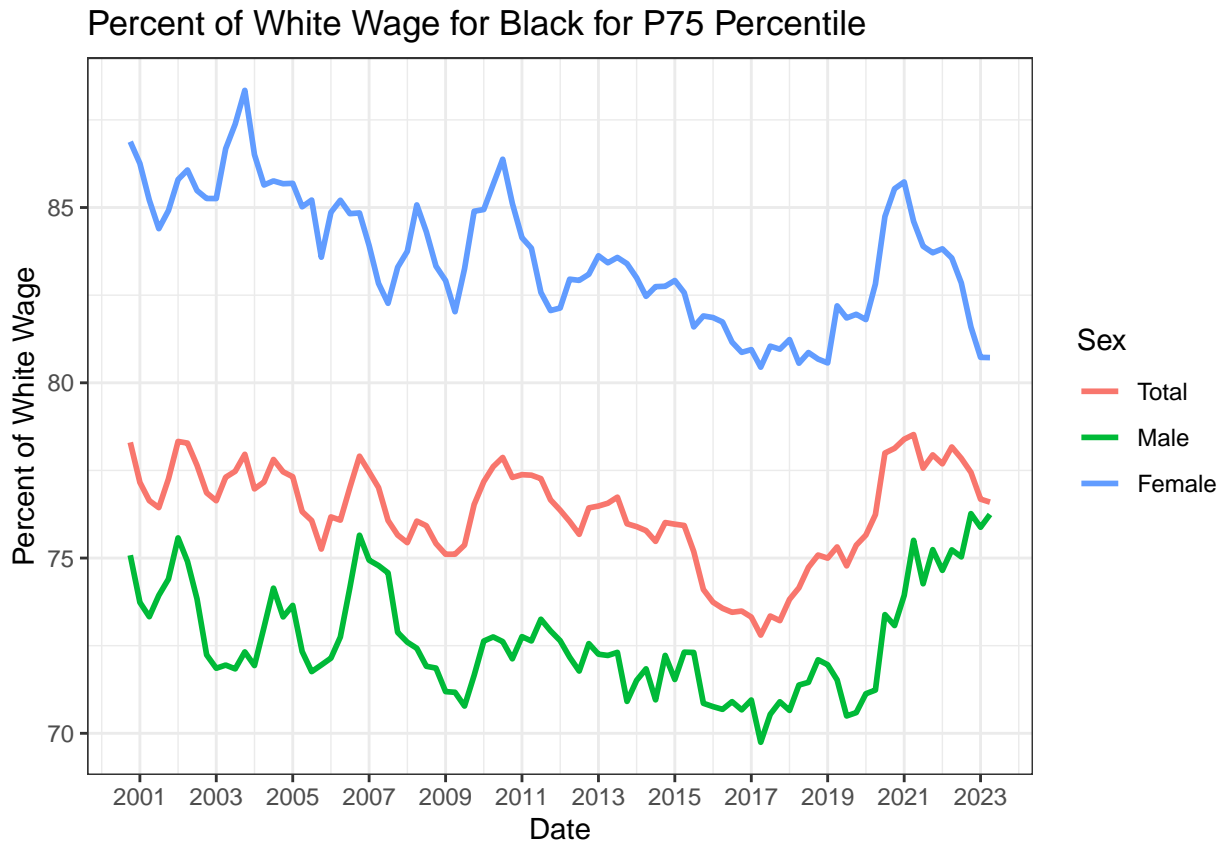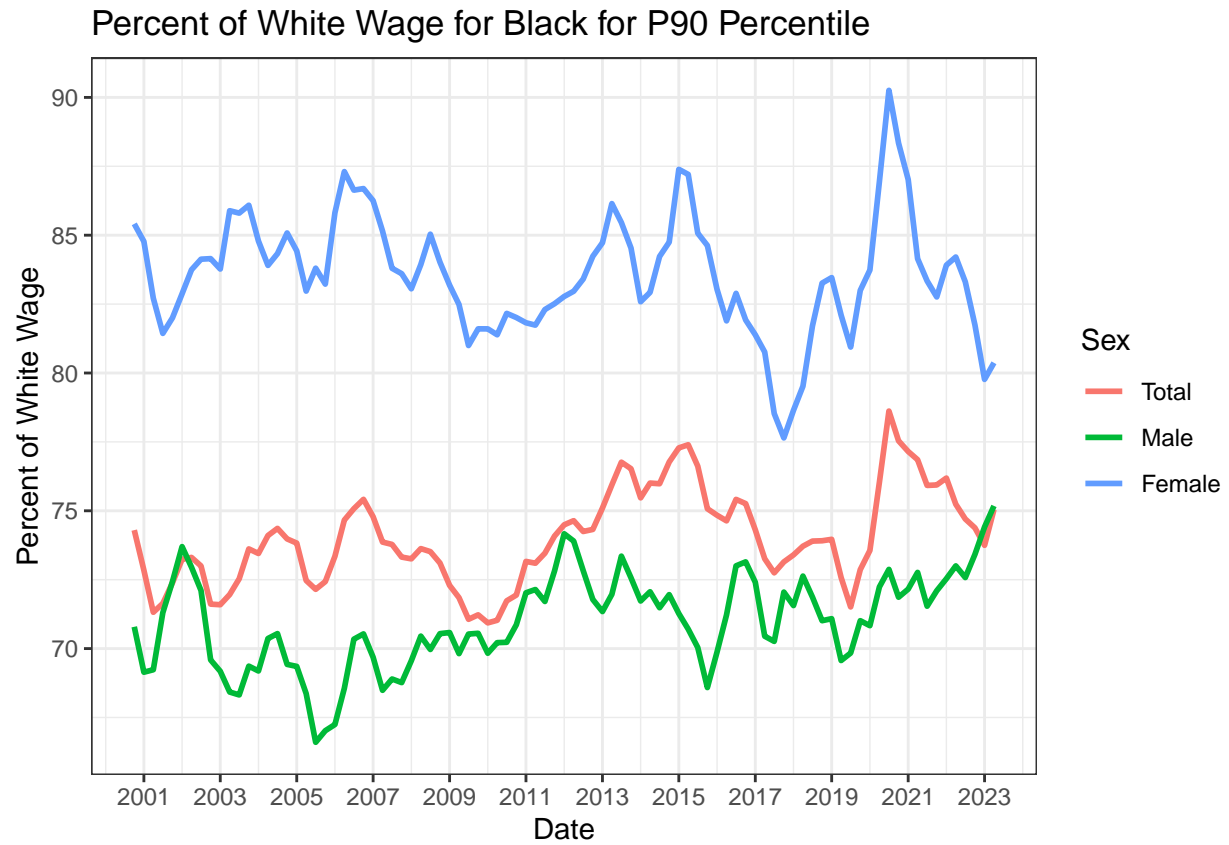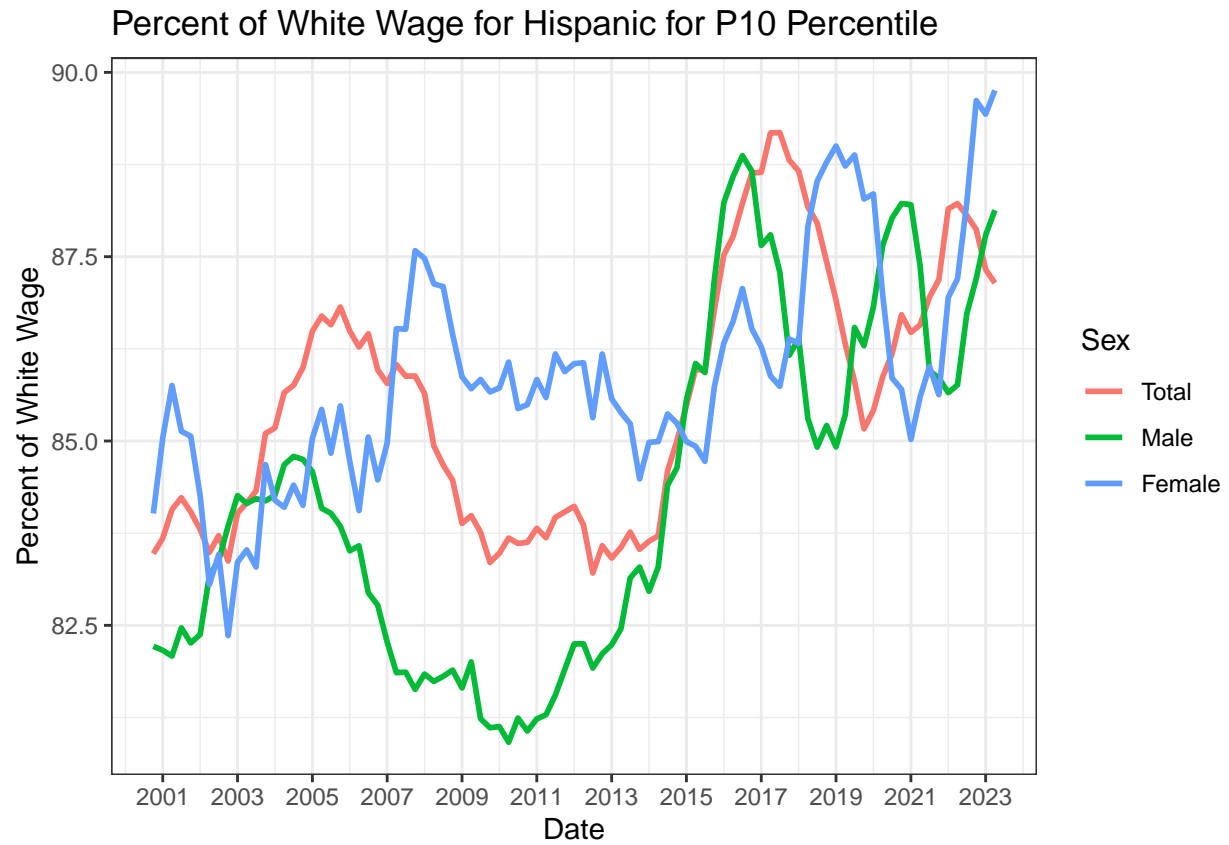

```
## 
## [[12]]
```

Percent of White Wage for Hispanic for P25 Percentile
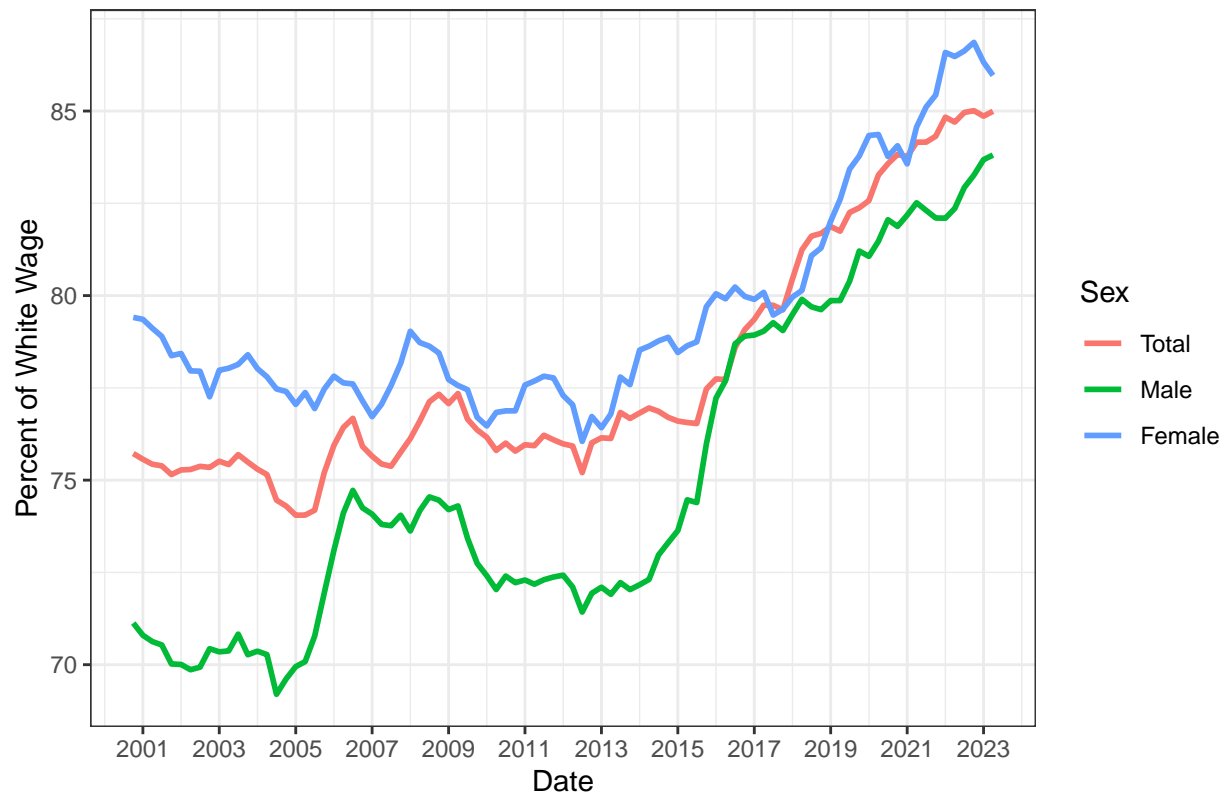

```
##
## [[13]]
```

Percent of White Wage for Hispanic for P50 Percentile


```
## 
## [[14]]
```

Percent of White Wage for Hispanic for P75 Percentile


```
##
## [[15]]
```

Percent of White Wage for Hispanic for P90 Percentile


```
##
## [[16]]
```

Percent of White Wage for All for P10 Percentile


```
## 
## [[17]]
```
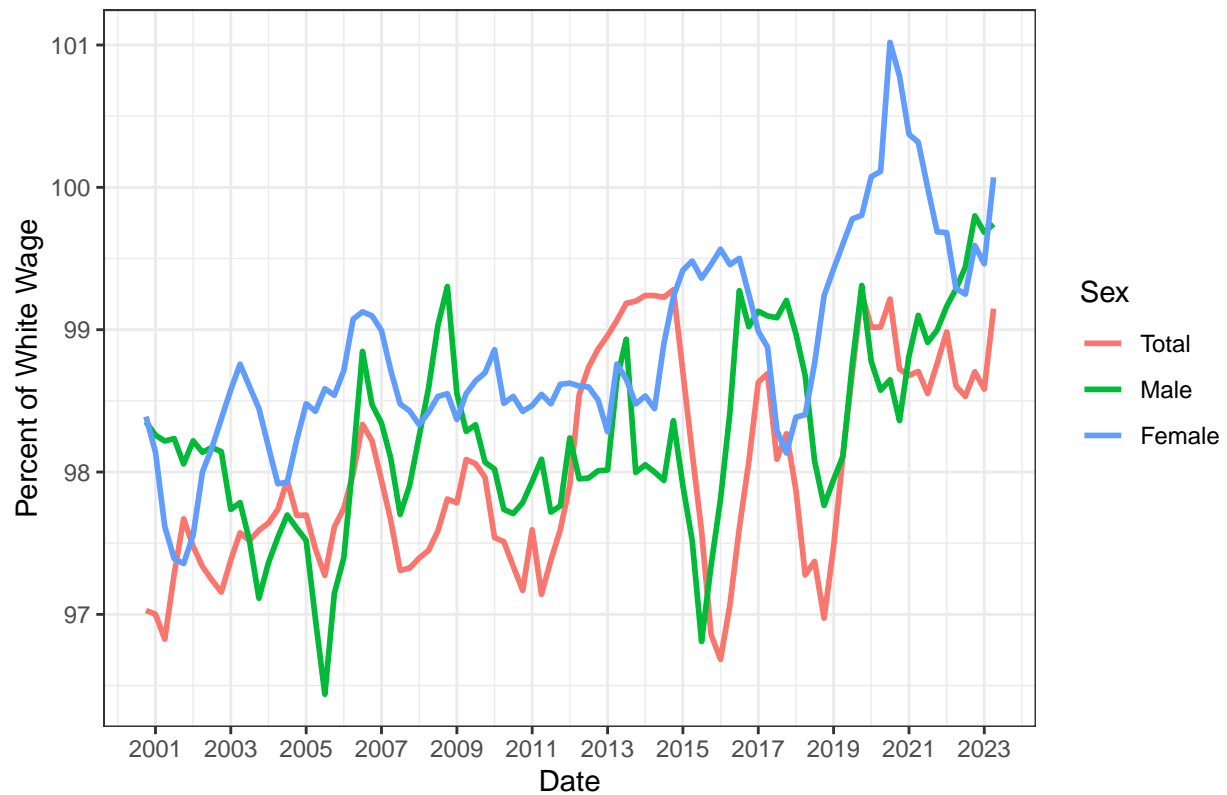
Percent of White Wage for All for P25 Percentile


```
##
## [[18]]
```

Percent of White Wage for All for P50 Percentile


```
##
## [[19]]
```
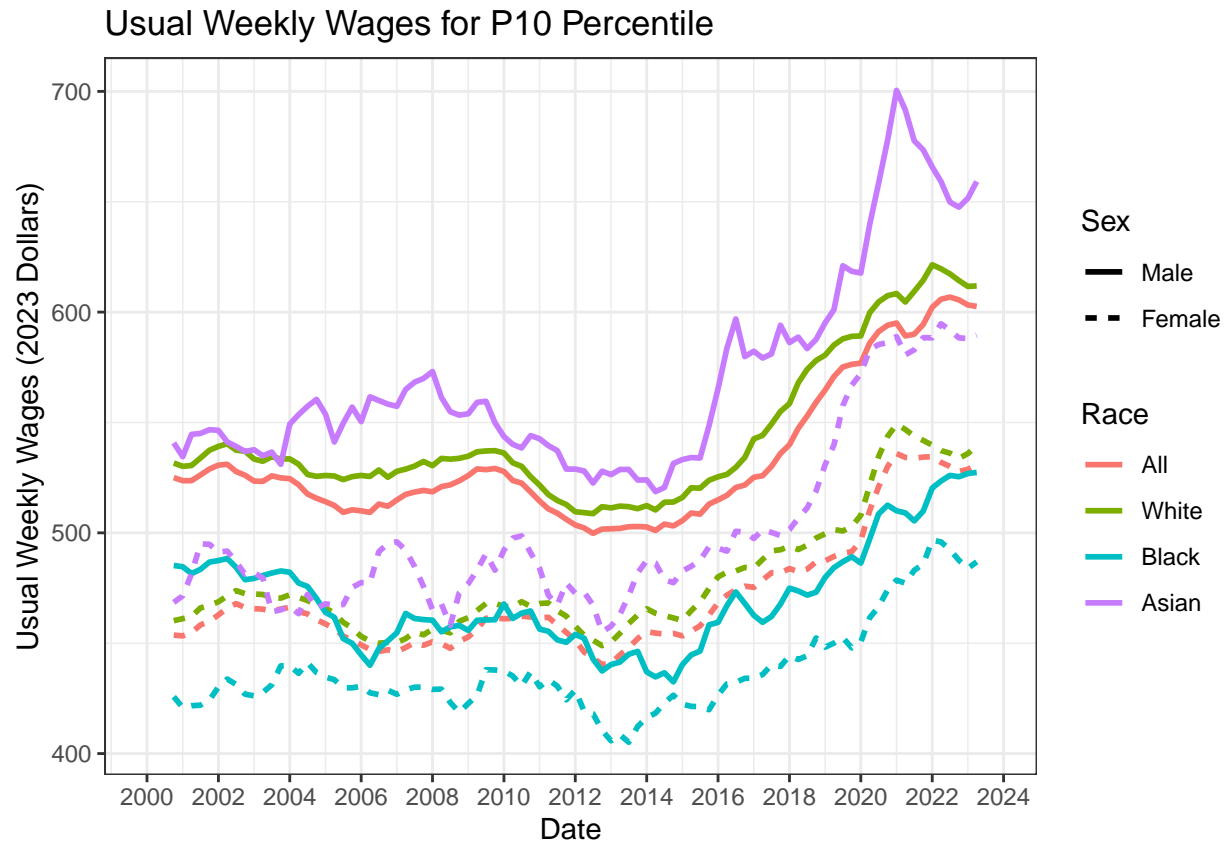
Percent of White Wage for All for P75 Percentile


```
## 
## [[20]]
```

Percent of White Wage for All for P90 Percentile


```
okay = c("P10","P25","P50","P75","P90")
plot_percentile = function(type){
  title = paste0("Usual Weekly Wages for ", type, " Percentile")
  data %>%
    filter(Sex != "Total" & Type == type & Race %in% c("White","All", "Asian","Black")) %>%
    ggplot(aes(x = Date, y = MovAverage, color = Race, linetype = Sex)) +
      geom_line(size = 1) + theme_bw() +   scale_x_date(date_breaks = "2 years", date_labels = "%Y") +
      ylab("Usual Weekly Wages (2023 Dollars)") +
      ggtitle(title)
}
percentile_plots = lapply(okay, plot_percentile)
```
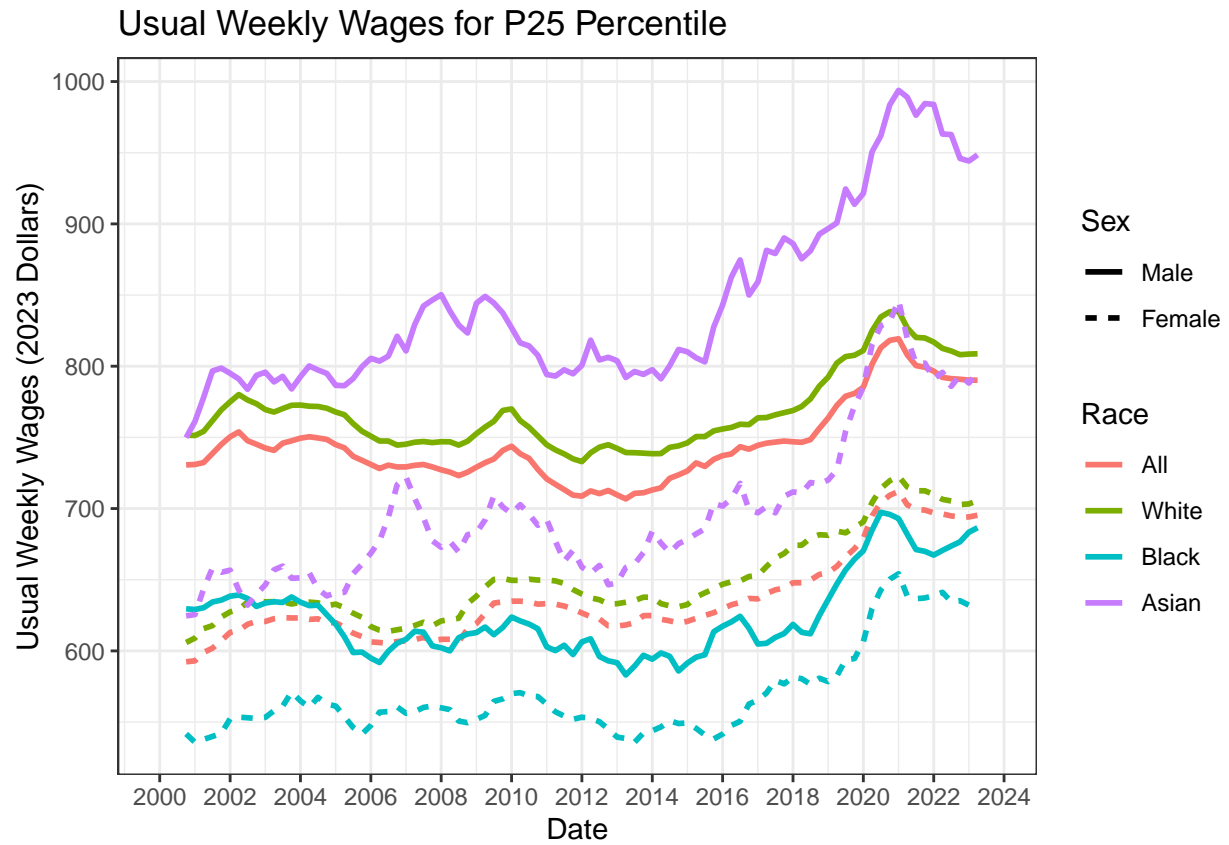
```
percentile_plots
```

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

```
## Warning: Removed 40 rows containing missing values (`geom_line()`).
```

Usual Weekly Wages for P10 Percentile


```
##
## [[2]]
```

```
## Warning: Removed 40 rows containing missing values ('geom_line()').
```

Usual Weekly Wages for P25 Percentile
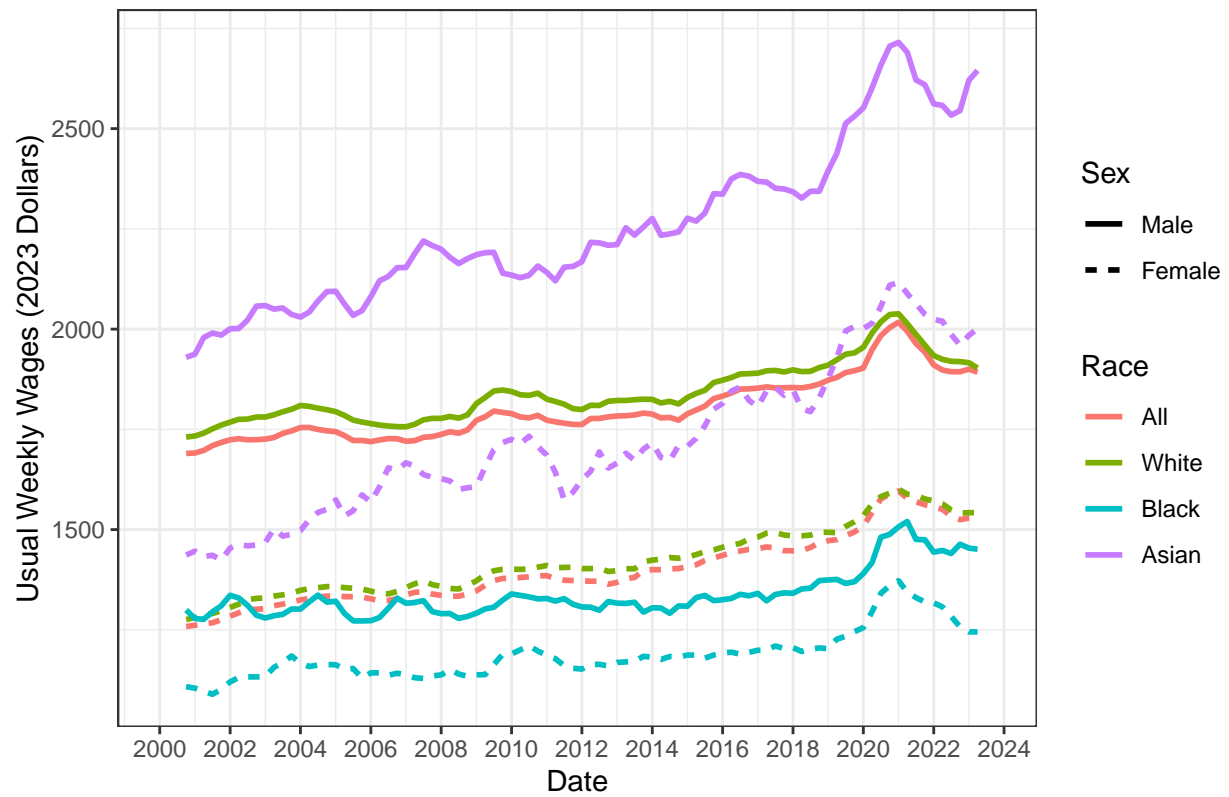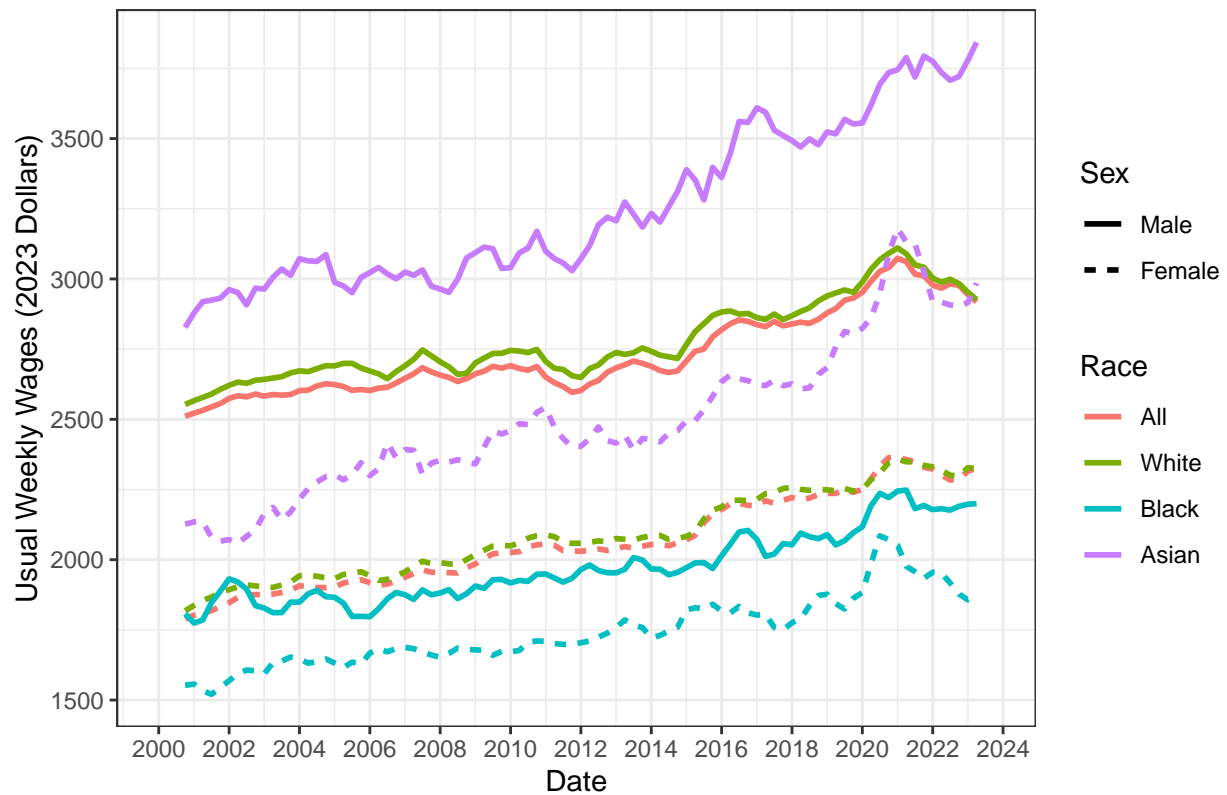
```
##
## [[3]]

## Warning: Removed 40 rows containing missing values (`geom_line()`).
```

Usual Weekly Wages for P50 Percentile

```
##
## [[4]]

## Warning: Removed 40 rows containing missing values ('geom_line()').
```

Usual Weekly Wages for P75 Percentile

```
##
## [[5]]

## Warning: Removed 40 rows containing missing values (`geom_line()`).
```
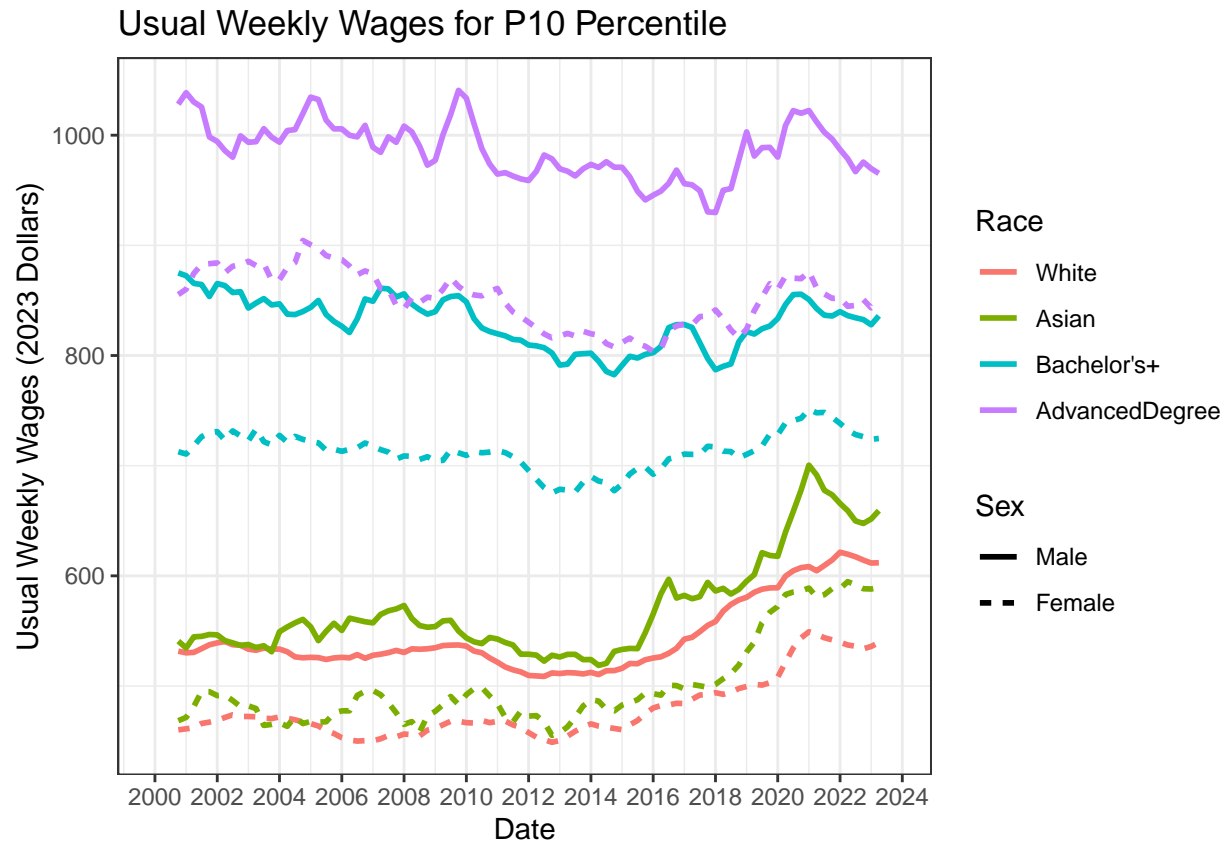
Usual Weekly Wages for P90 Percentile


```r
okay = c("P10","P25","P50","P75","P90")
plot_percentile2 = function(type){
  title = paste0("Usual Weekly Wages for ", type, " Percentile")
  data %>%
    filter(Sex != "Total" & Type == type & Race %in% c("White","Asian","Bachelor's+","AdvancedDegree"))
    ggplot(aes(x = Date, y = MovAverage, color = Race, linetype = Sex)) +
      geom_line(size = 1) + theme_bw() +  scale_x_date(date_breaks = "2 years", date_labels = "%Y") +
      ylab("Usual Weekly Wages (2023 Dollars)") +
      ggtitle(title)
}
percentile_plots2 = lapply(okay, plot_percentile2)
```

```r
percentile_plots2
```
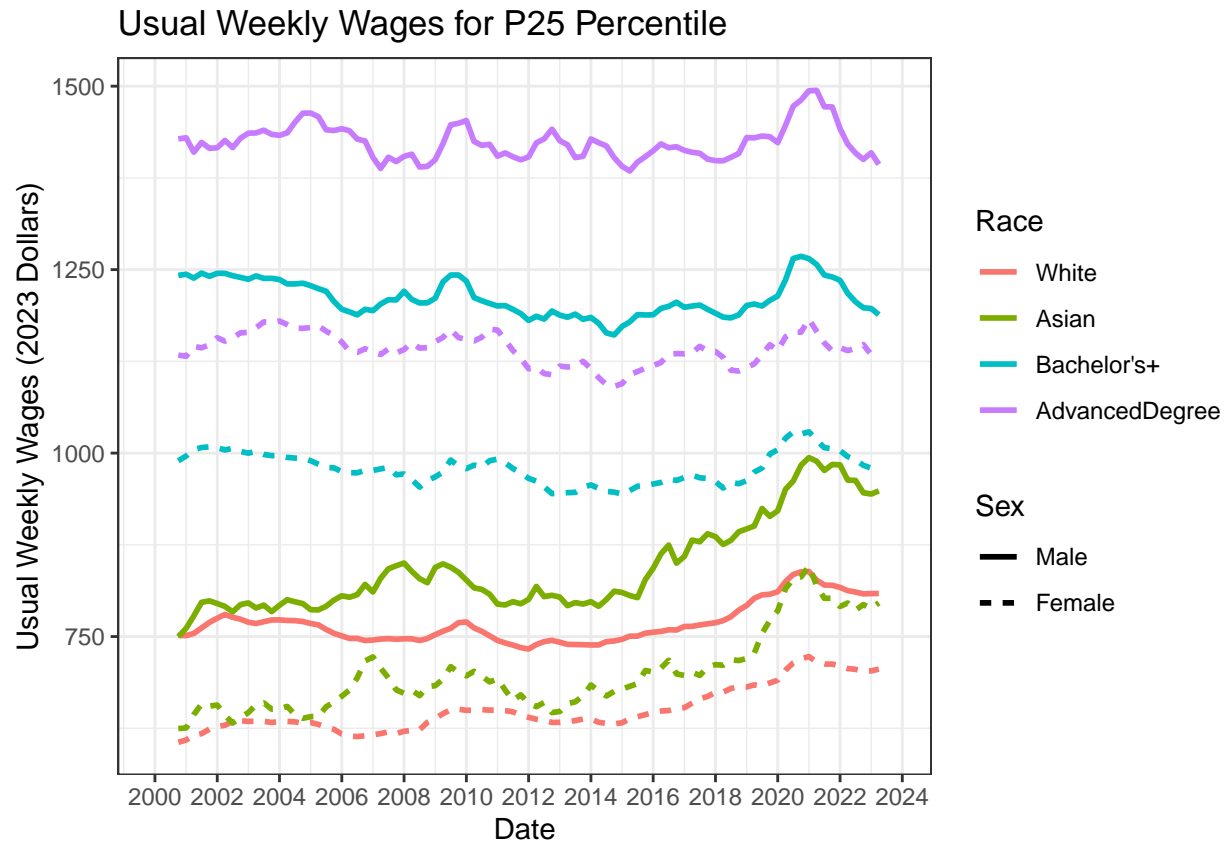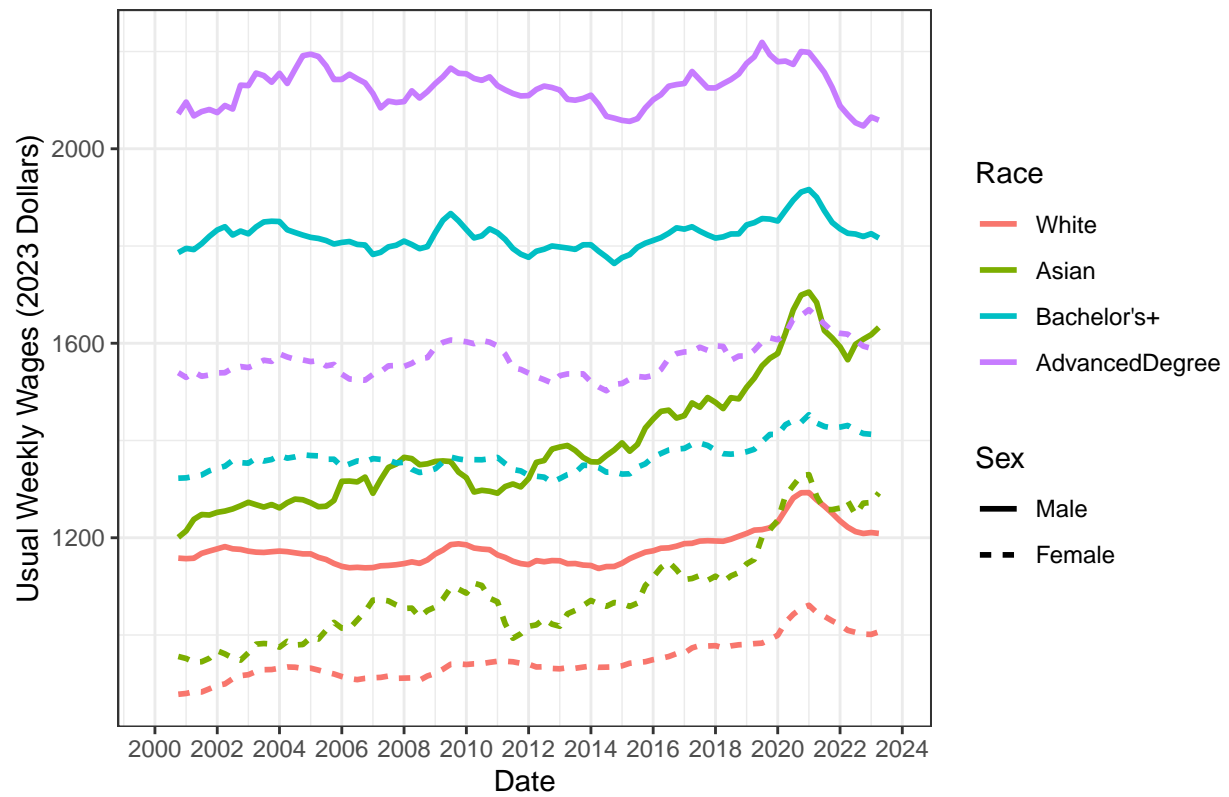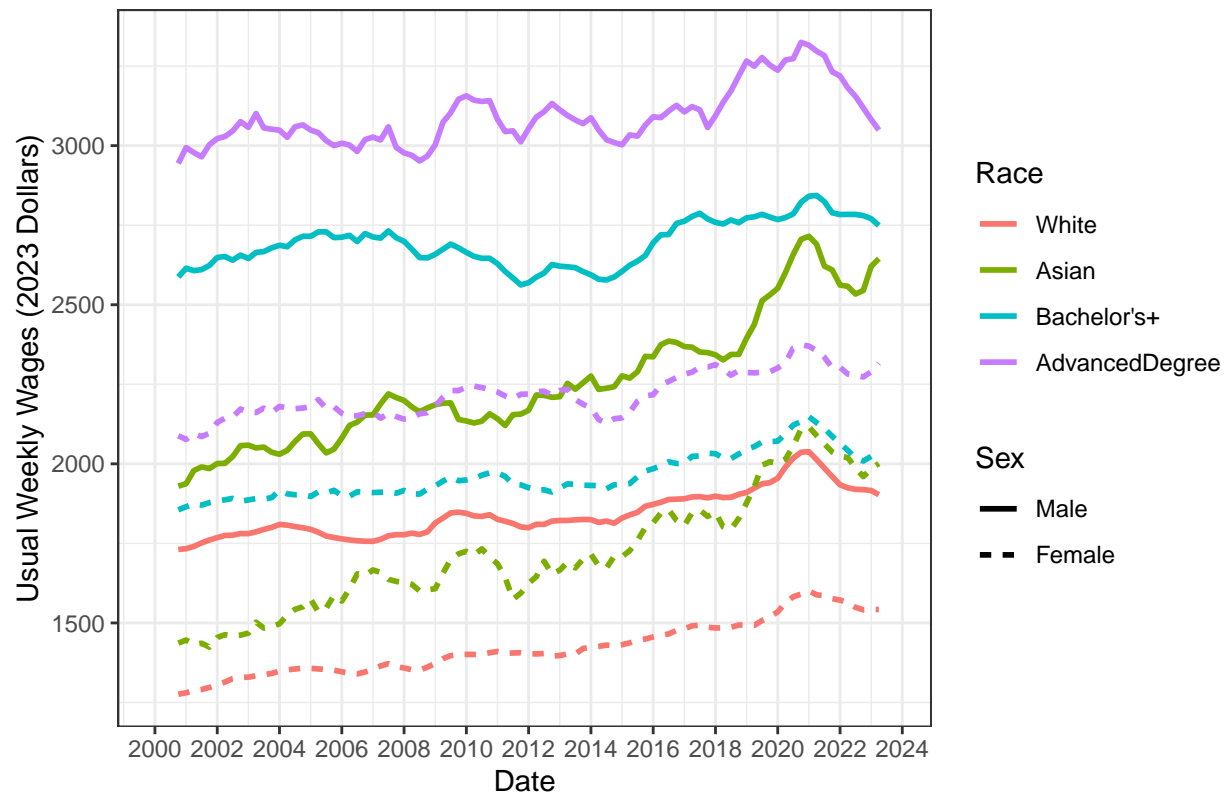
```
## [[1]]
```

```
## Warning: Removed 40 rows containing missing values ('geom_line()').
```

Usual Weekly Wages for P10 Percentile


```
##
## [[2]]

## Warning: Removed 40 rows containing missing values (`geom_line()`).
```

Usual Weekly Wages for P25 Percentile


```
##
## [[3]]

## Warning: Removed 40 rows containing missing values (`geom_line()`).
```

Usual Weekly Wages for P50 Percentile


```
##
## [[4]]

## Warning: Removed 40 rows containing missing values (`geom_line()`).
```
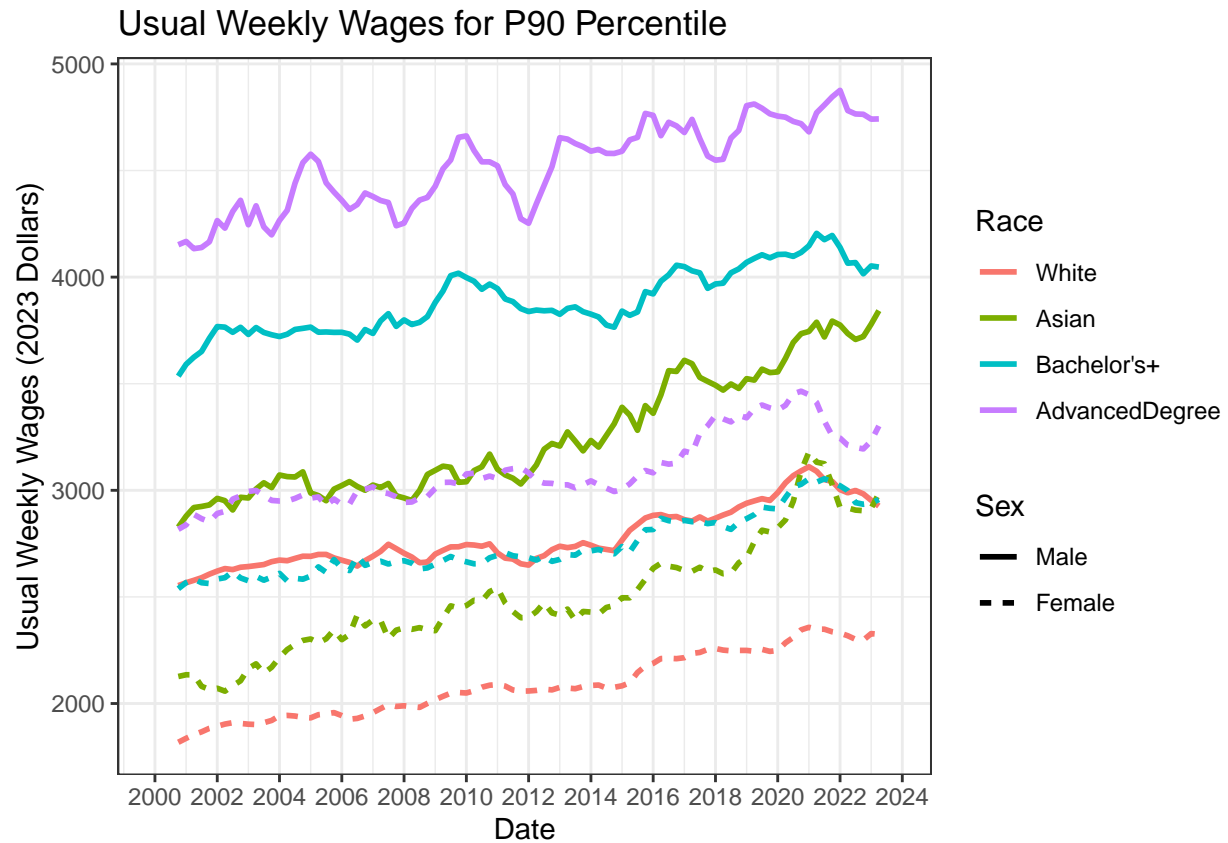
Usual Weekly Wages for P75 Percentile


```
## 
## [[5]]

## Warning: Removed 40 rows containing missing values (`geom_line()`).
```

## Usual Weekly Wages for P90 Percentile


```
okay = unique(data$Race)
plot_count = function(race){
  title = paste0("Number of Full-Time Workers for ", race)
  data %>%
    filter(Sex != "Total" & Type == "Count" & Race == race) %>%
    ggplot(aes(x = Date, y = MovAverage, color = Sex)) +
      geom_line(size = 1) + theme_bw() +  scale_x_date(date_breaks = "2 years", date_labels = "%Y") +
      ylab("Number of Full-Time Workers") +
      ggtitle(title)
}
count_plots = lapply(okay, plot_count)
```
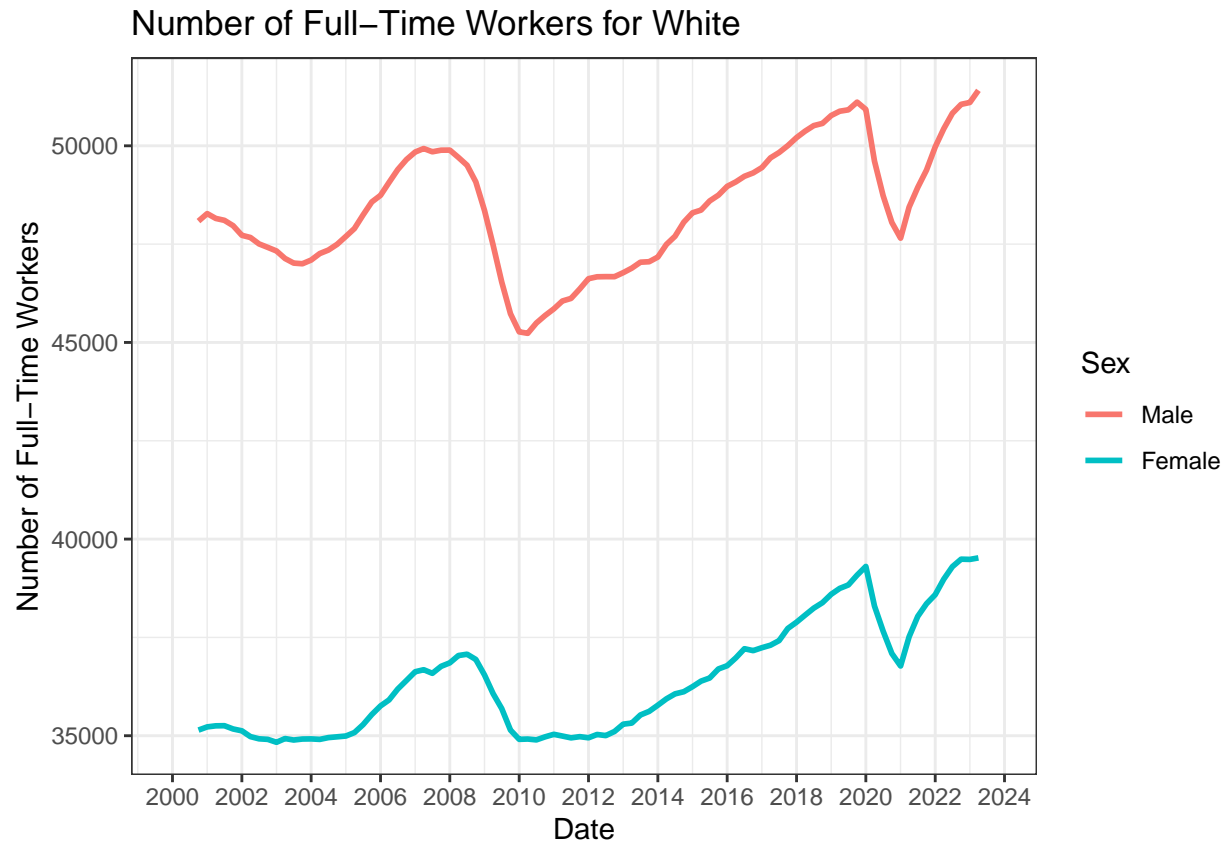
```
count_plots
```

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

```
## Warning: Removed 10 rows containing missing values (`geom_line()`).
```
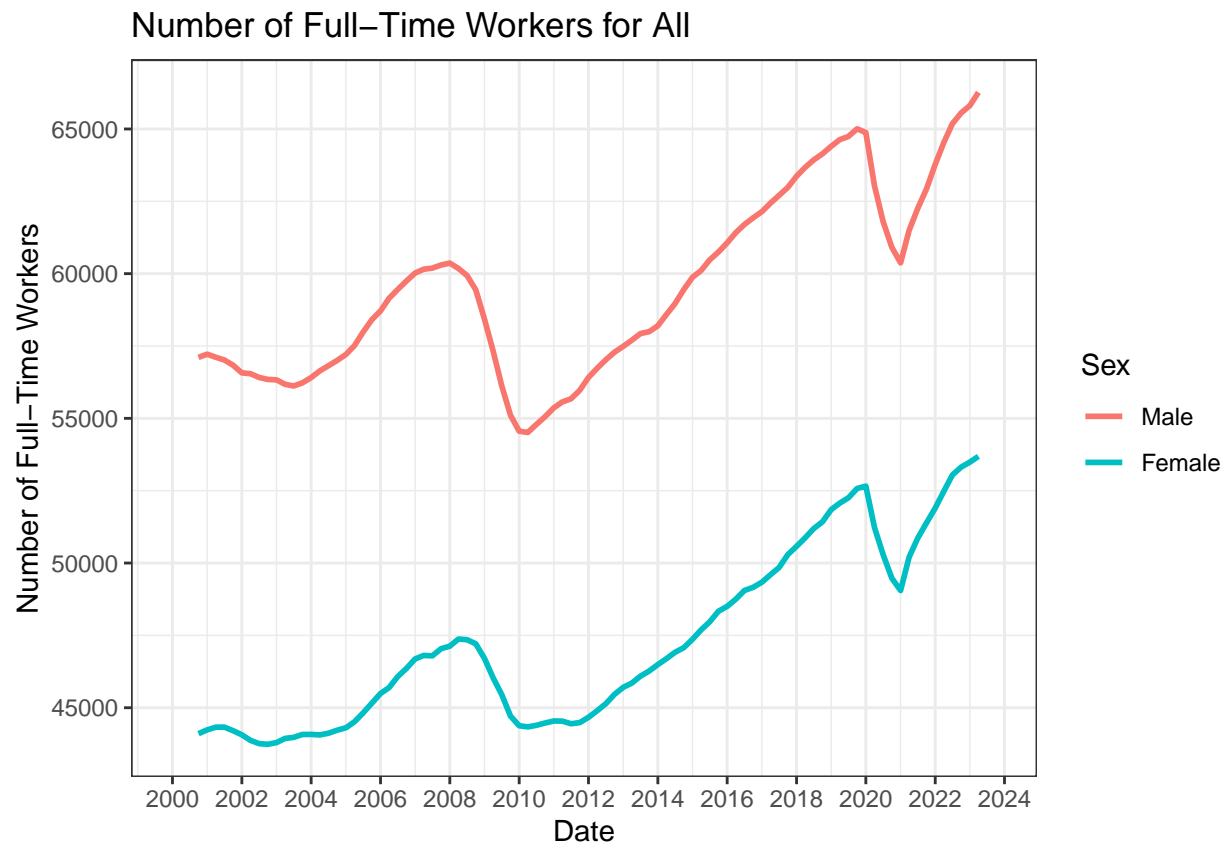
Number of Full−Time Workers for All
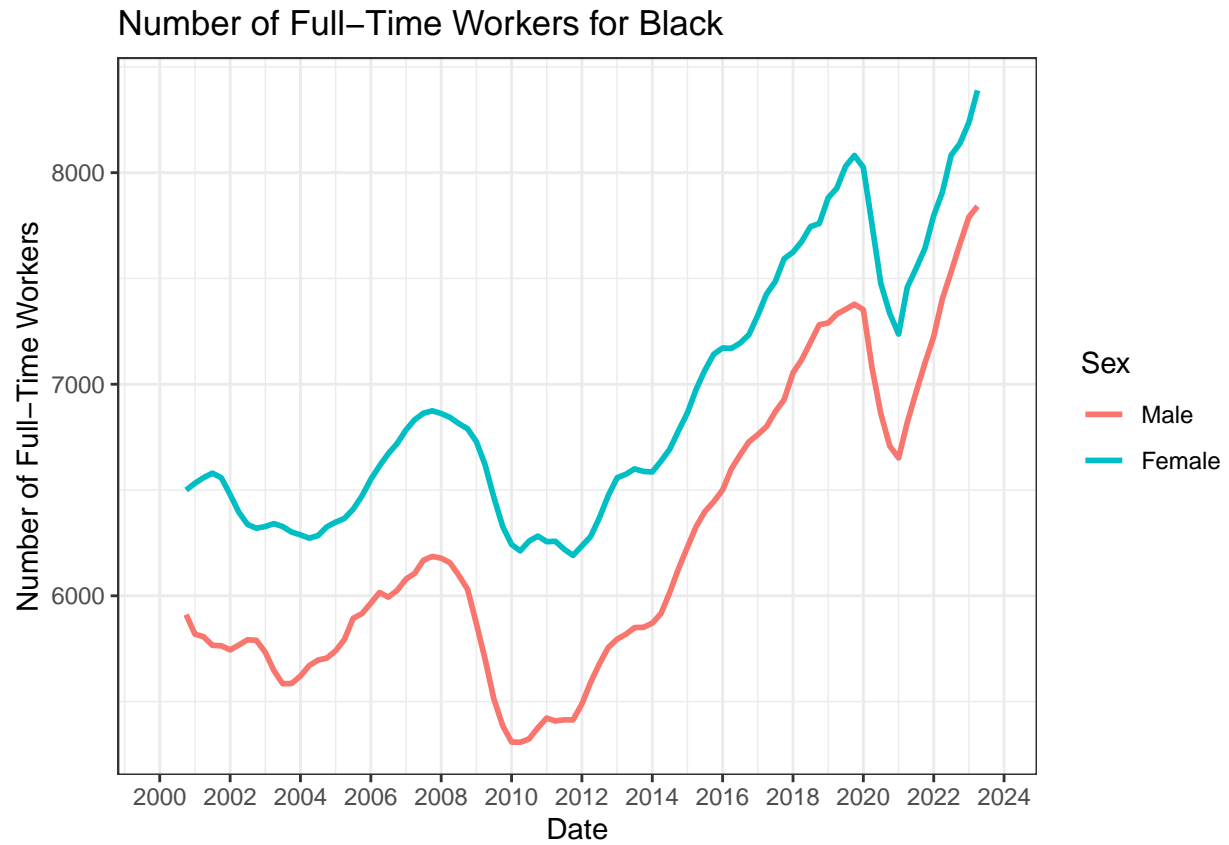


```
##
## [[2]]
```

```
## Warning: Removed 10 rows containing missing values (`geom_line()`).
```

Number of Full−Time Workers for White


```
## 
## [[3]]

## Warning: Removed 10 rows containing missing values ('geom_line()').
```

Number of Full−Time Workers for Black


```
## 
## [[4]]
```

```
## Warning: Removed 10 rows containing missing values ('geom_line()').
```

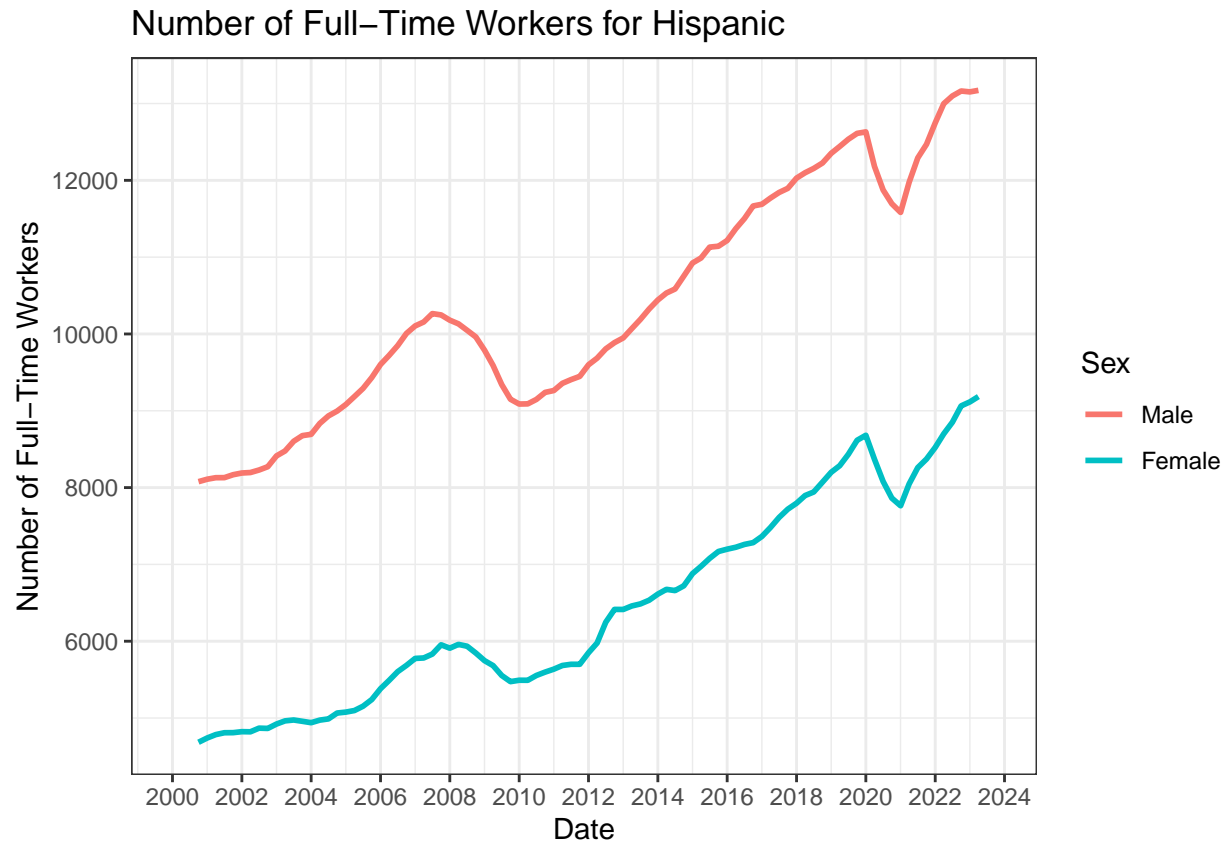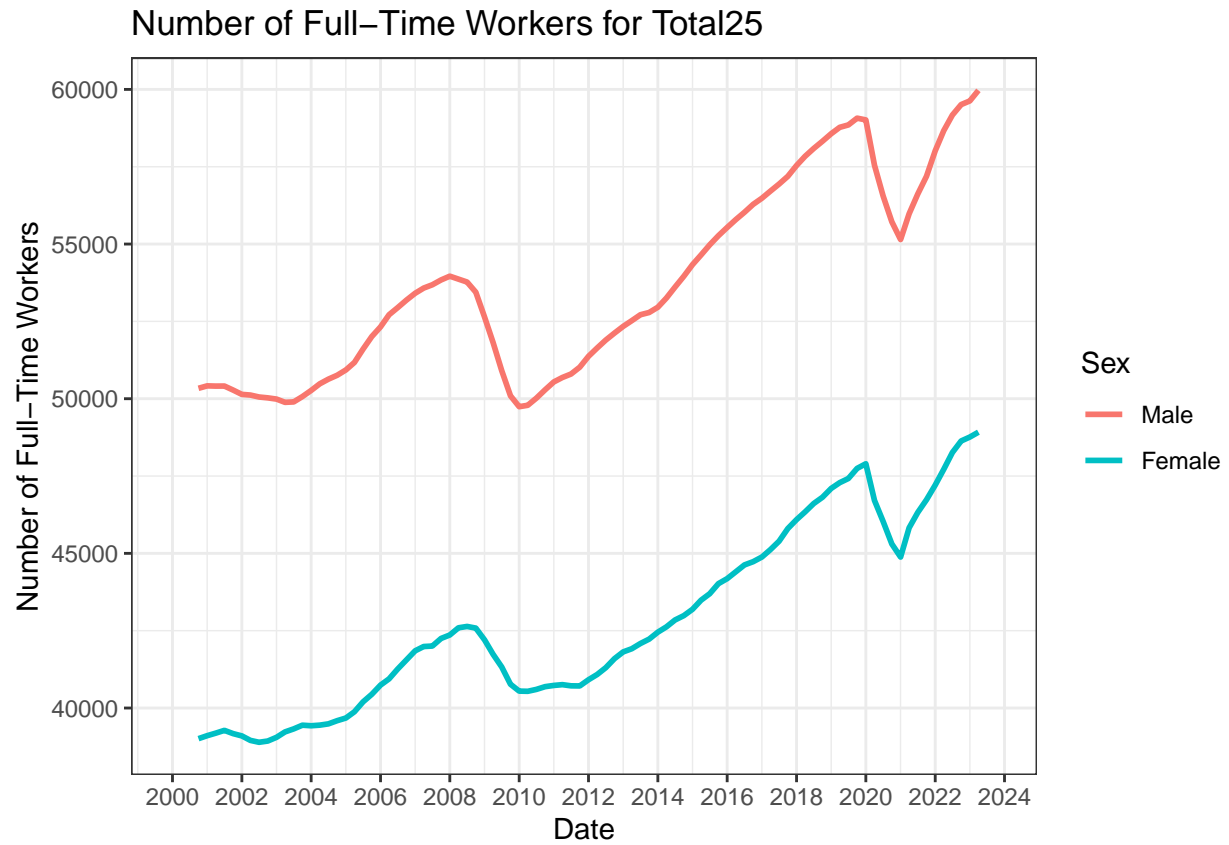## Number of Full−Time Workers for Asian


```
##
## [[5]]
```

```
## Warning: Removed 10 rows containing missing values ('geom_line()').
```

Number of Full-Time Workers for Hispanic
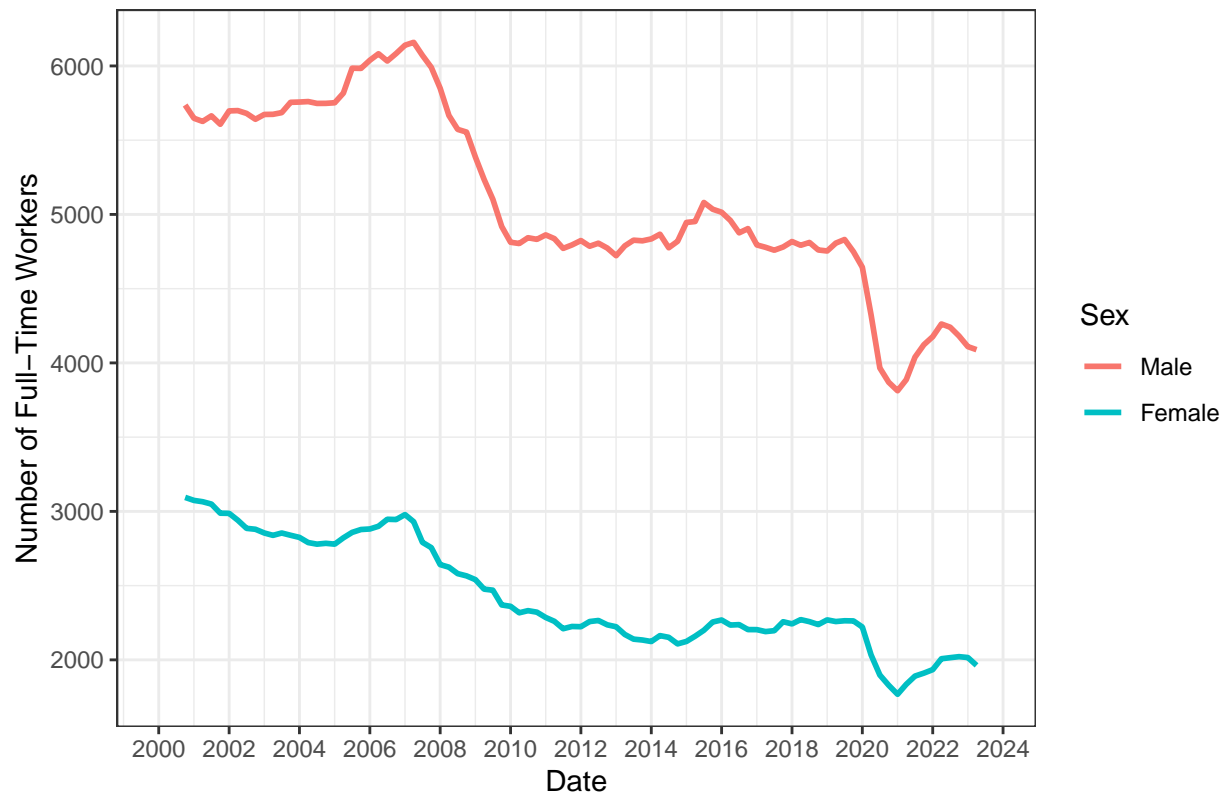


```
##
## [[6]]
```

```
## Warning: Removed 10 rows containing missing values ('geom_line()').
```
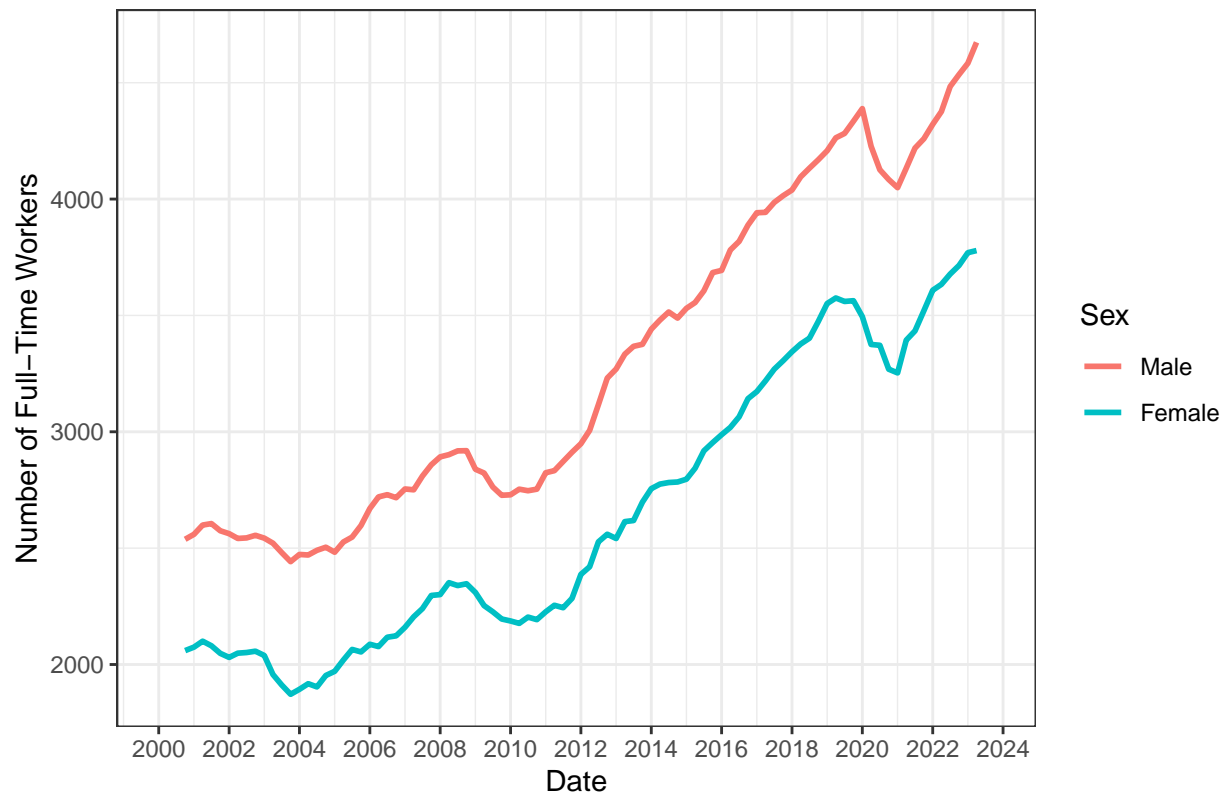
Number of Full–Time Workers for Total25


```
##
## [[7]]

## Warning: Removed 10 rows containing missing values (`geom_line()`).
```

Number of Full–Time Workers for LessThanHS


```
##
## [[8]]
```
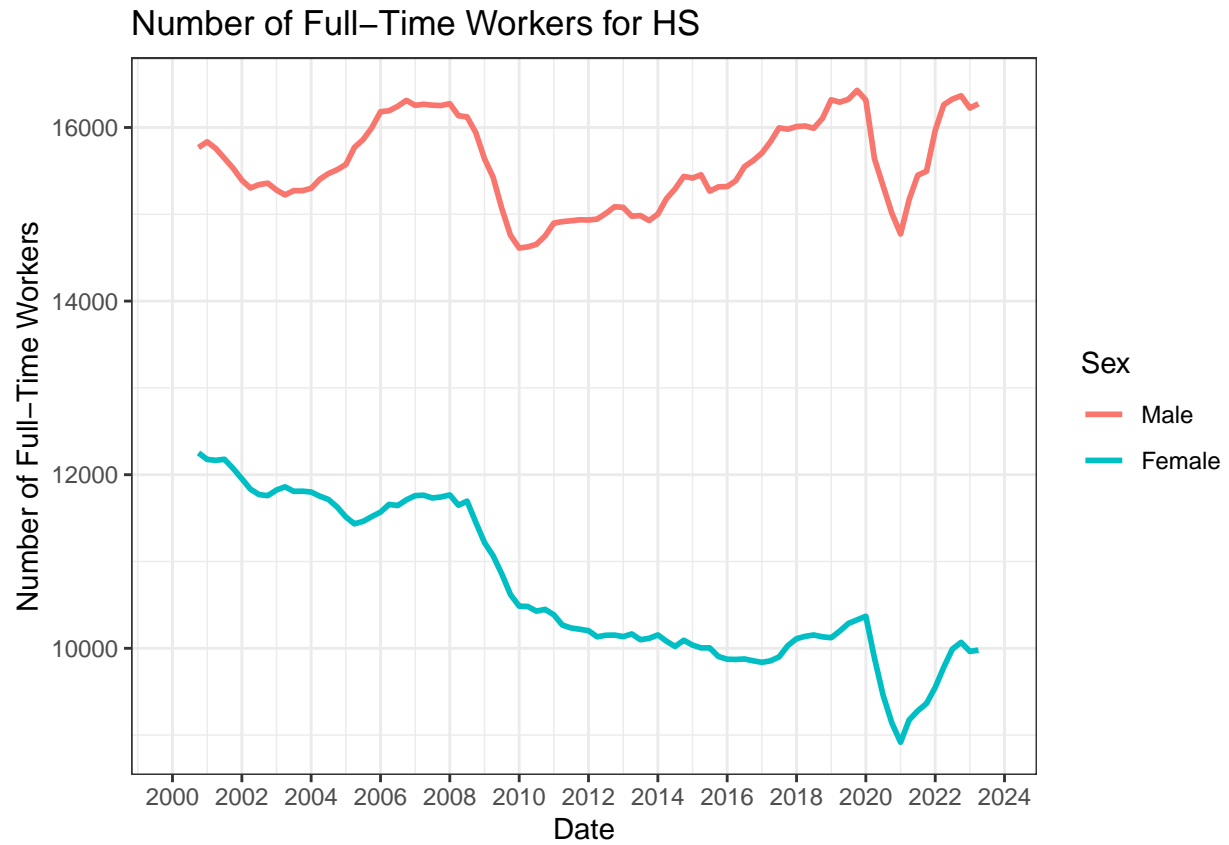
```
## Warning: Removed 10 rows containing missing values ('geom_line()').
```
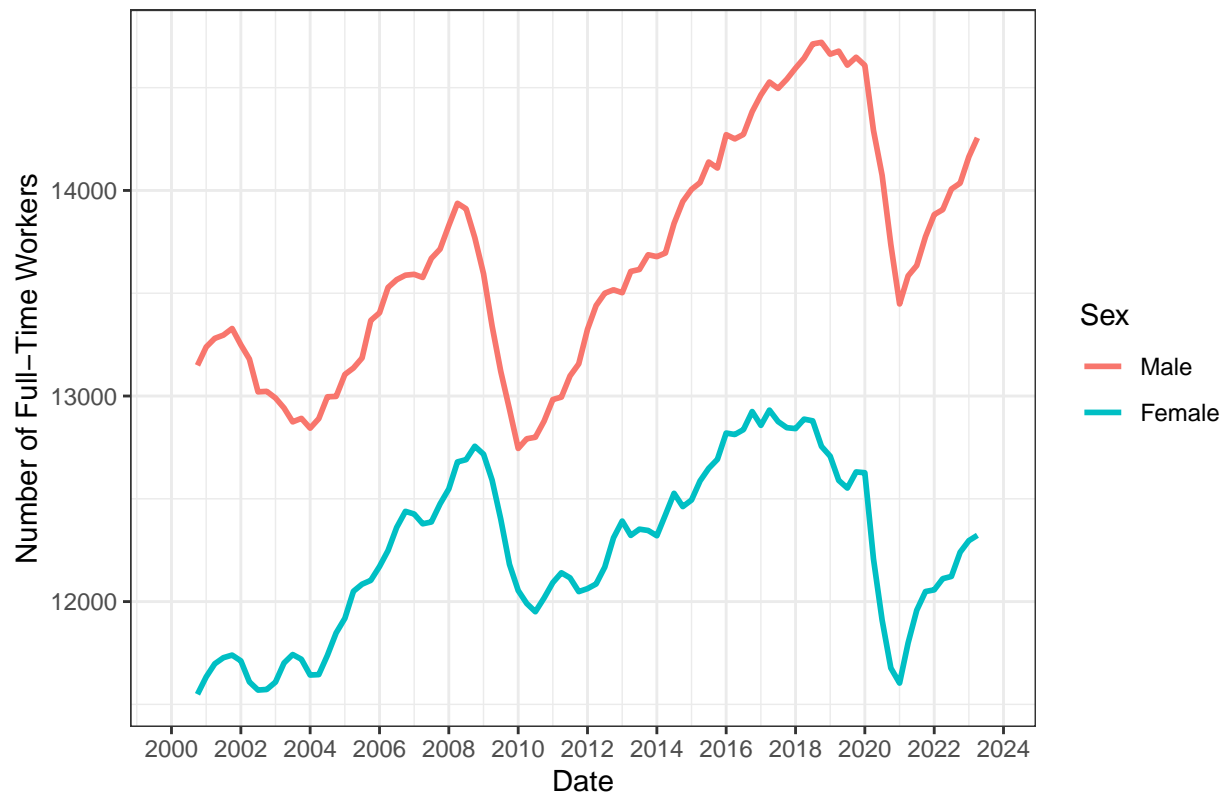
Number of Full–Time Workers for HS


```
##
## [[9]]
```

```
## Warning: Removed 10 rows containing missing values (`geom_line()`).
```
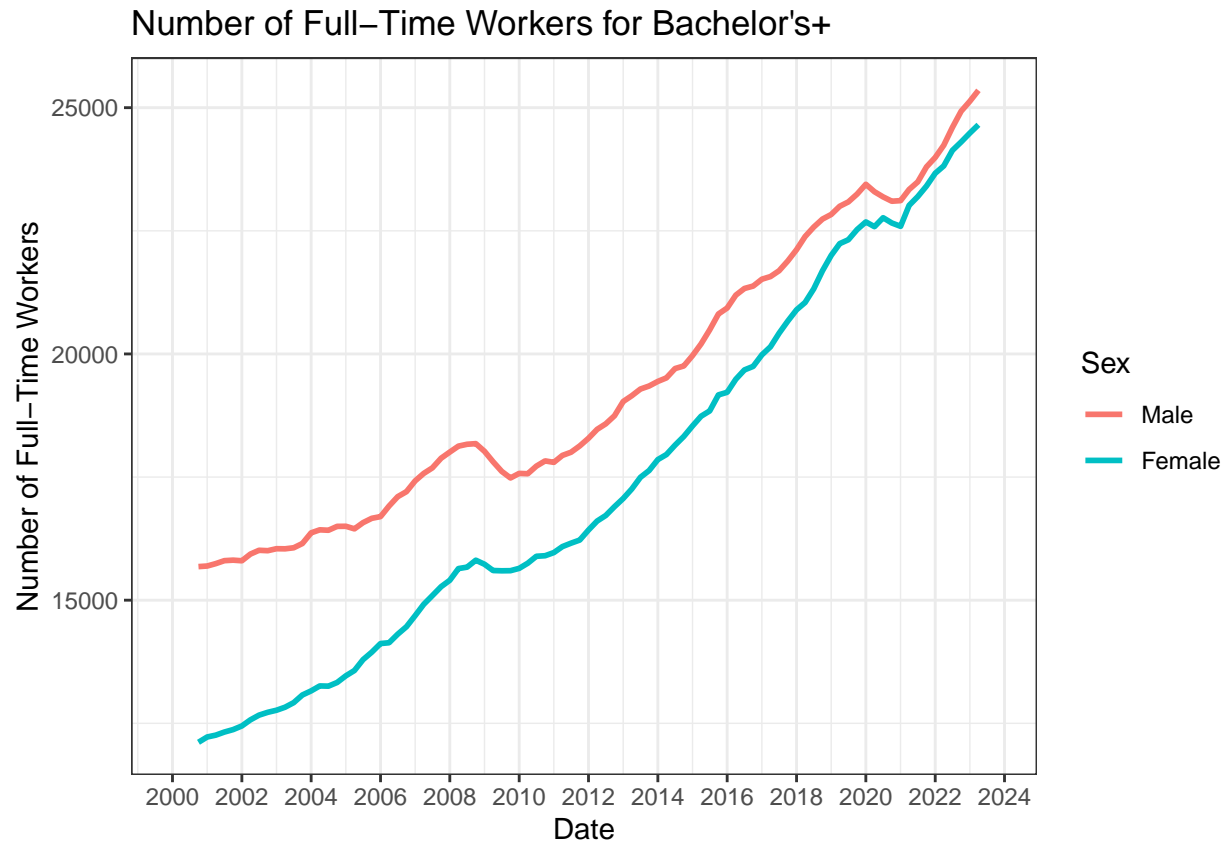
# Number of Full−Time Workers for SomeCollege


```
## 
## [[10]]
```

```
## Warning: Removed 10 rows containing missing values (`geom_line()`).
```

Number of Full−Time Workers for Bachelor's+
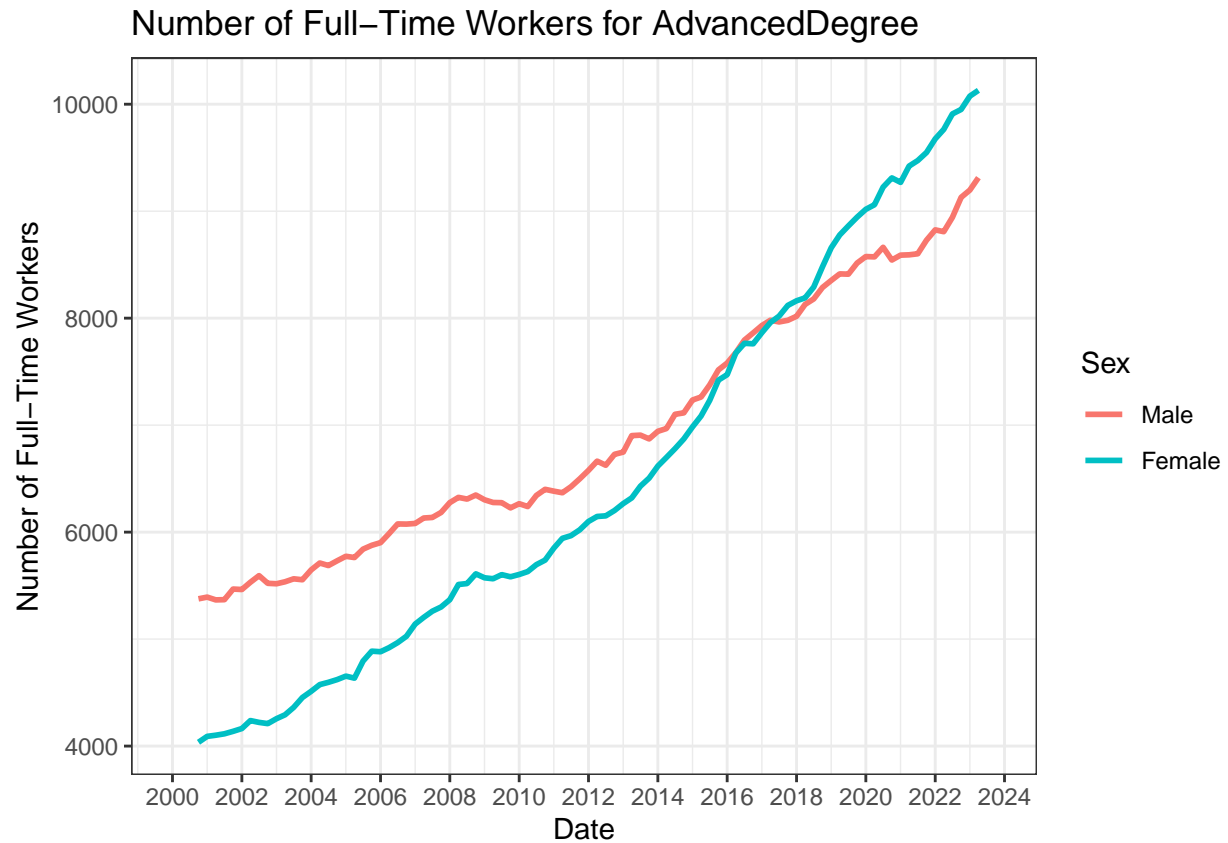


```
##
## [[11]]
```

```
## Warning: Removed 10 rows containing missing values (`geom_line()`).
```

Number of Full–Time Workers for Bachelor'sOnly


```
## 
## [[12]]
```

```
## Warning: Removed 10 rows containing missing values (`geom_line()`).
```

Number of Full−Time Workers for AdvancedDegree
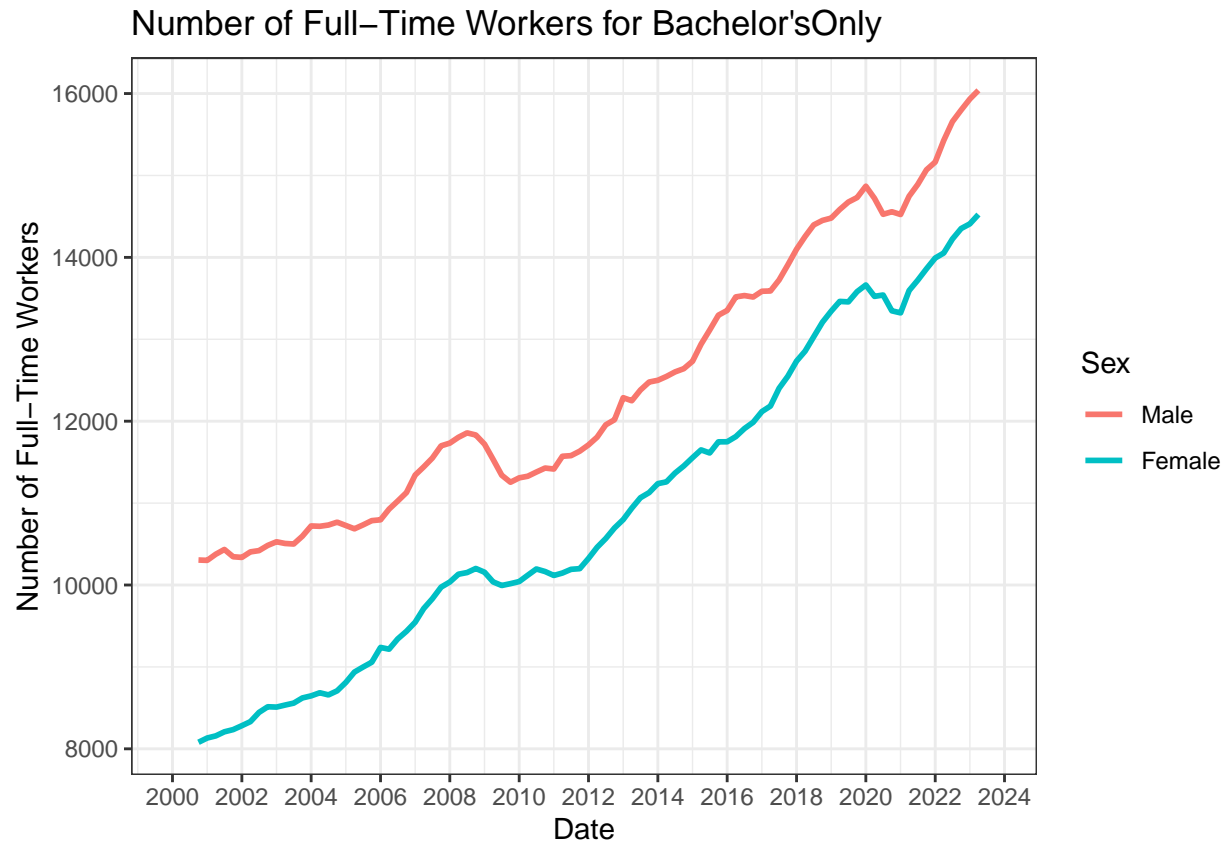

```r
data %>%
  filter(Sex != "Total" & Type != "Count" & Race == "AdvancedDegree") %>%
  ggplot(aes(x = Date, y = MovAverage, color = Sex, linetype = Type)) +
    geom_line(size = 1) + theme_bw() +  scale_x_date(date_breaks = "2 years", date_labels = "%Y") +
    ylab("Usual Weekly Wages (2023 Dollars)") +
    ggtitle("Usual Weekly Wages for Workers with an Advanced Degree by Percentile")
```

```
## Warning: Removed 50 rows containing missing values (`geom_line()`).
```

Usual Weekly Wages for Workers with an Advanced Degree by Percentile
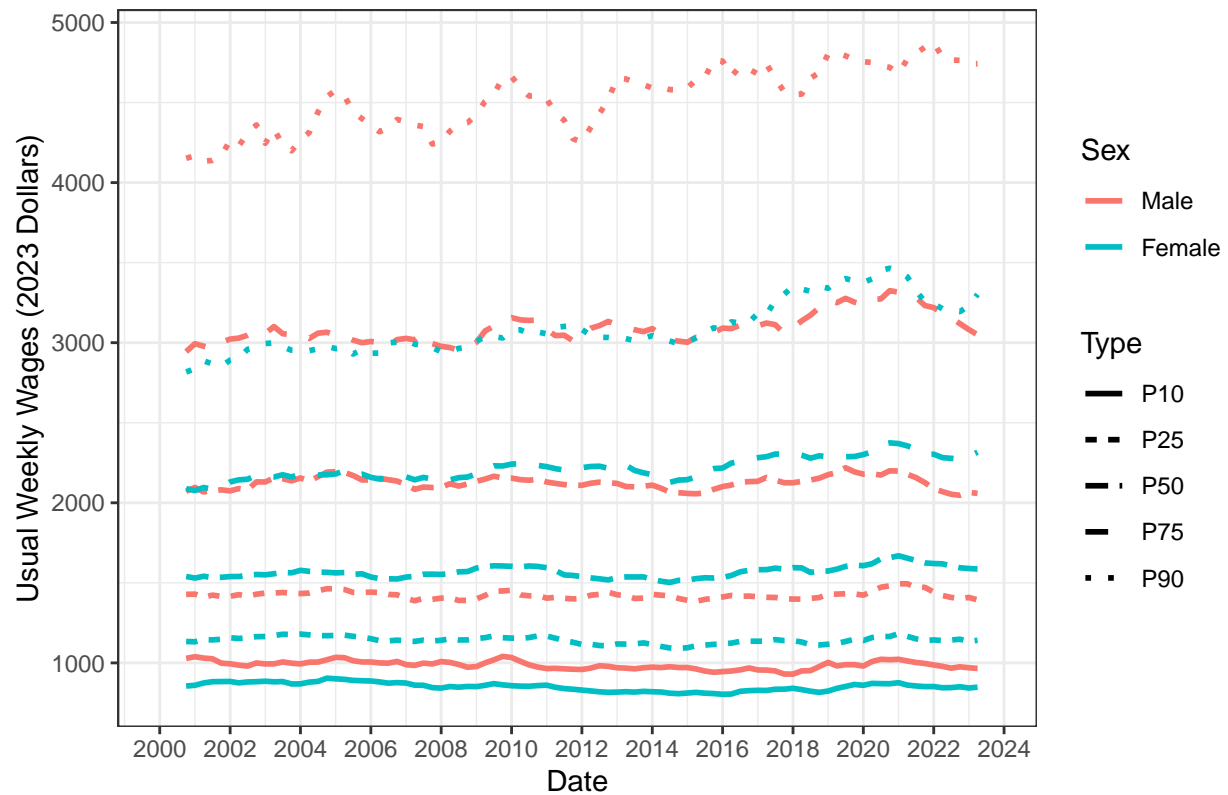

```
data %>%
  filter(Sex != "Total" & Type != "Count" & Race == "Bachelor's+") %>%
  ggplot(aes(x = Date, y = MovAverage, color = Sex, linetype = Type)) +
    geom_line(size = 1) + theme_bw() + scale_x_date(date_breaks = "2 years", date_labels = "%Y") +
    ylab("Usual Weekly Wages (2023 Dollars)") +
    ggtitle("Usual Weekly Wages for Workers with an Bachelor's Degree by Percentile")
```

```
## Warning: Removed 50 rows containing missing values (`geom_line()`).
```

Usual Weekly Wages for Workers with an Bachelor's Degree by Percentile
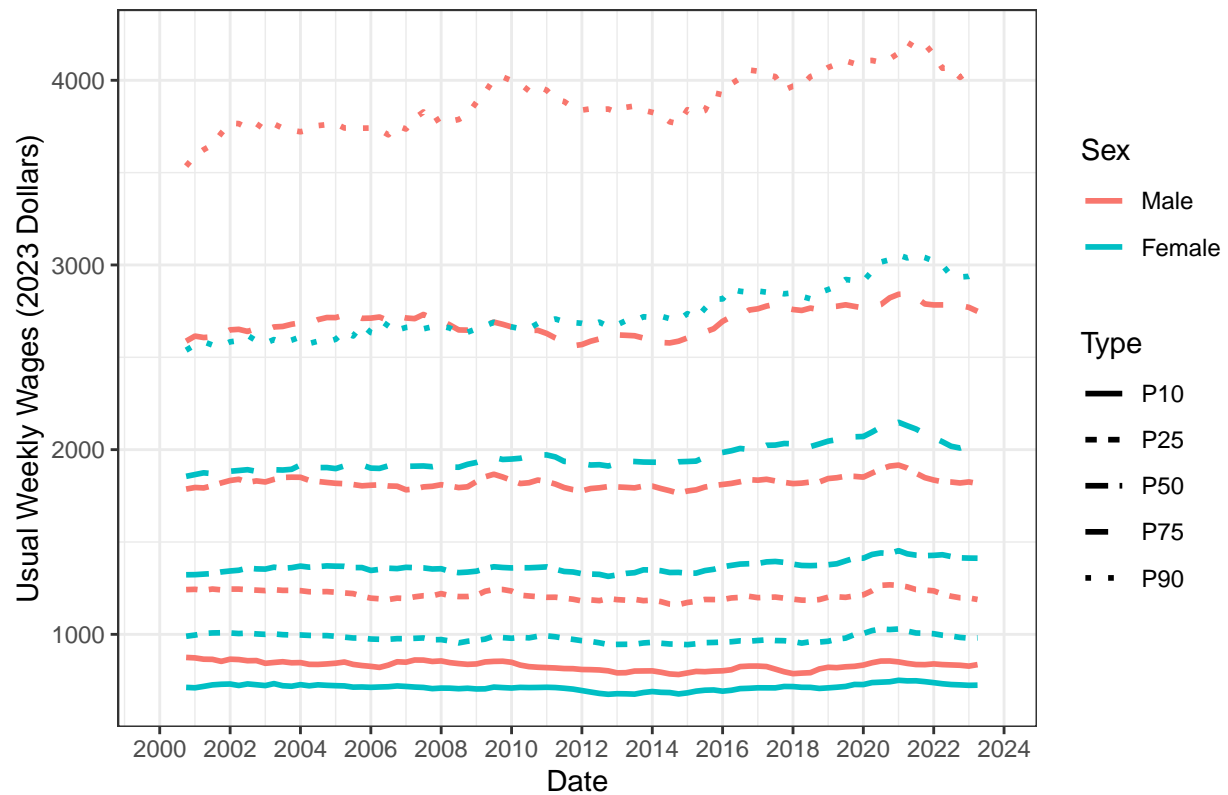


```r
relative_to_male = data %>%
  ungroup() %>%
  filter(!is.na(MovAverage)) %>%
  group_by(Type, Race, Date) %>%
  summarize(Female = MovAverage[Sex == "Female"]/MovAverage[Sex == "Male"])
```

```
## `summarise()` has grouped output by 'Type', 'Race'. You can override using the
## `.groups` argument.
```
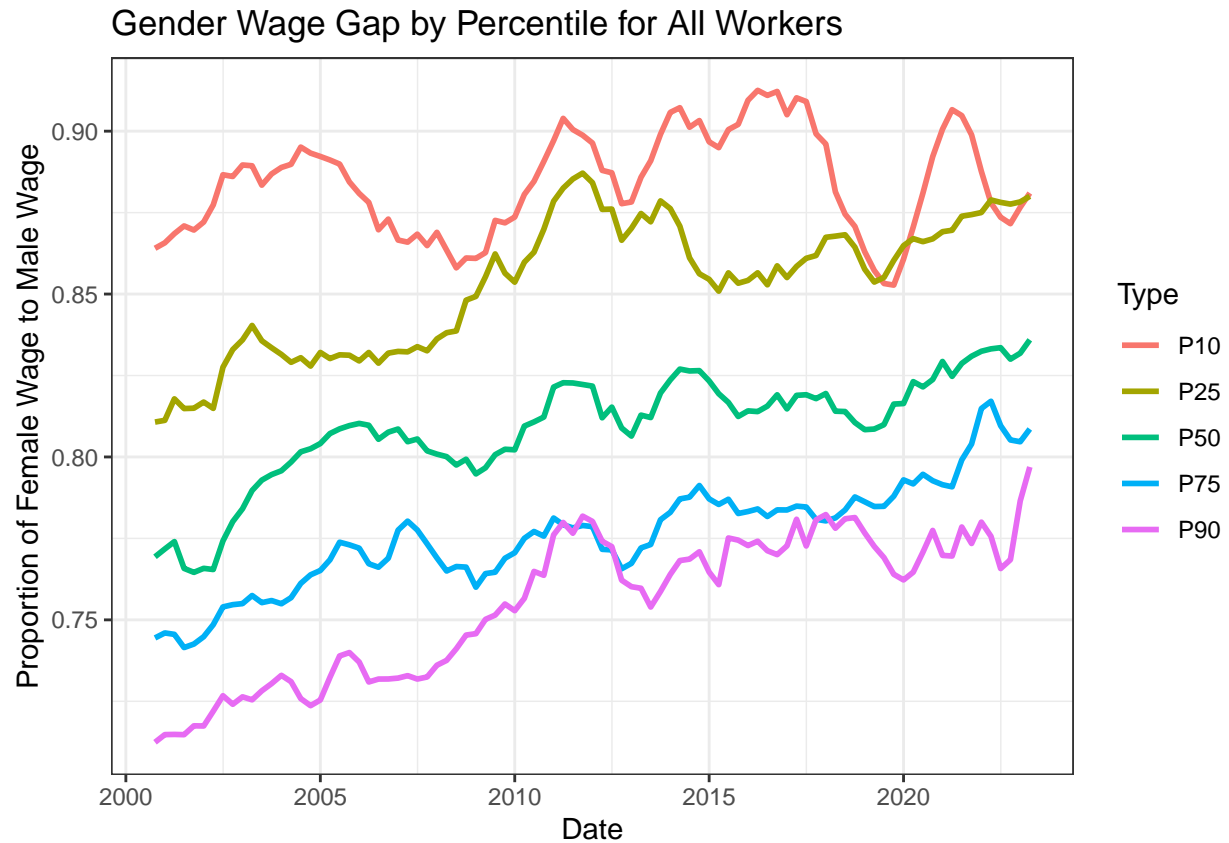
```r
relative_to_male %>%
  filter(Race == "AdvancedDegree" & Type != "Count") %>%
  ggplot(aes(x = Date, y = Female, color = Type)) +
    geom_line(size = 1) + theme_bw() +
    ylab("Proportion of Female Wage to Male Wage") +
    ggtitle("Gender Wage Gap by Percentile for Advanced Degree")
```

# Gender Wage Gap by Percentile for Advanced Degree
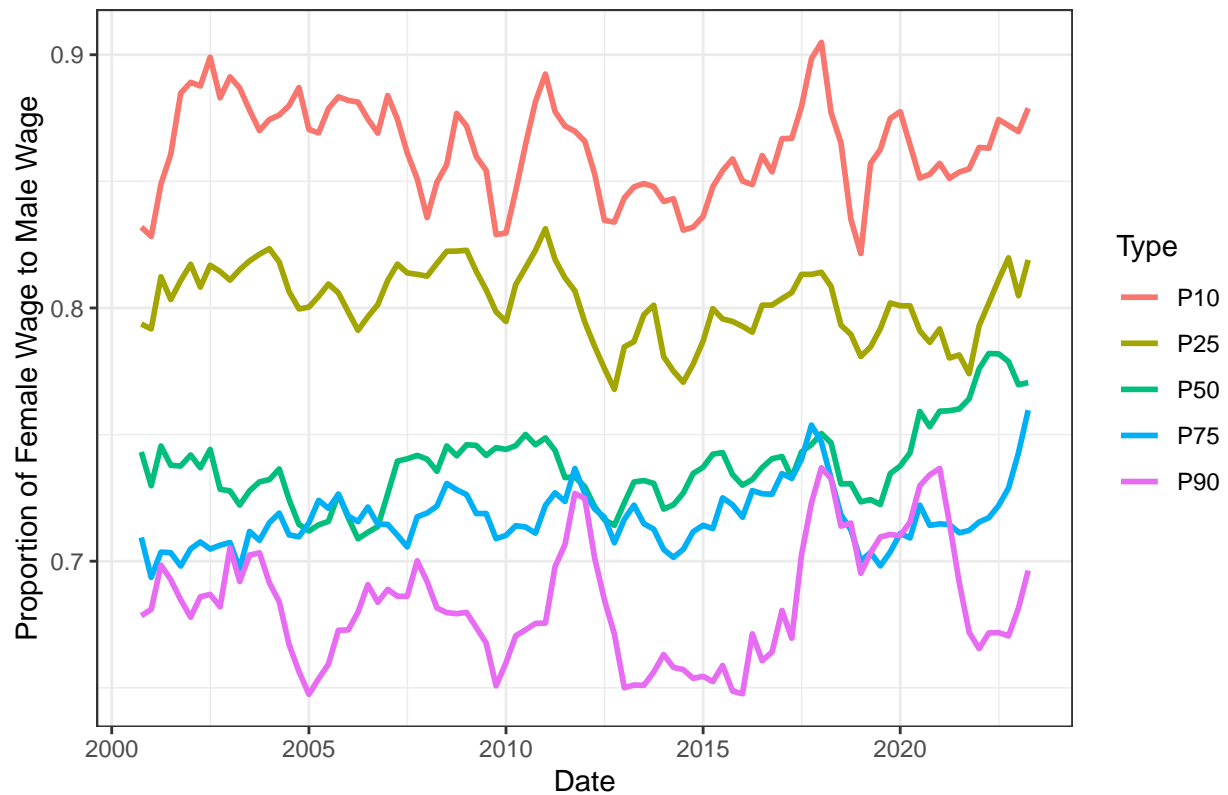

```
relative_to_male %>%
  filter(Race == "All" & Type != "Count") %>%
  ggplot(aes(x = Date, y = Female, color = Type)) +
    geom_line(size = 1) + theme_bw() +
    ylab("Proportion of Female Wage to Male Wage") +
    ggtitle("Gender Wage Gap by Percentile for All Workers")
```

Gender Wage Gap by Percentile for All Workers


```r
relative_to_male %>%
  filter(Race == "Bachelor's+" & Type != "Count") %>%
  ggplot(aes(x = Date, y = Female, color = Type)) +
    geom_line(size = 1) + theme_bw() +
    ylab("Proportion of Female Wage to Male Wage") +
    ggtitle("Gender Wage Gap by Percentile for Bachelor's Degree+")
```

Gender Wage Gap by Percentile for Bachelor's Degree+
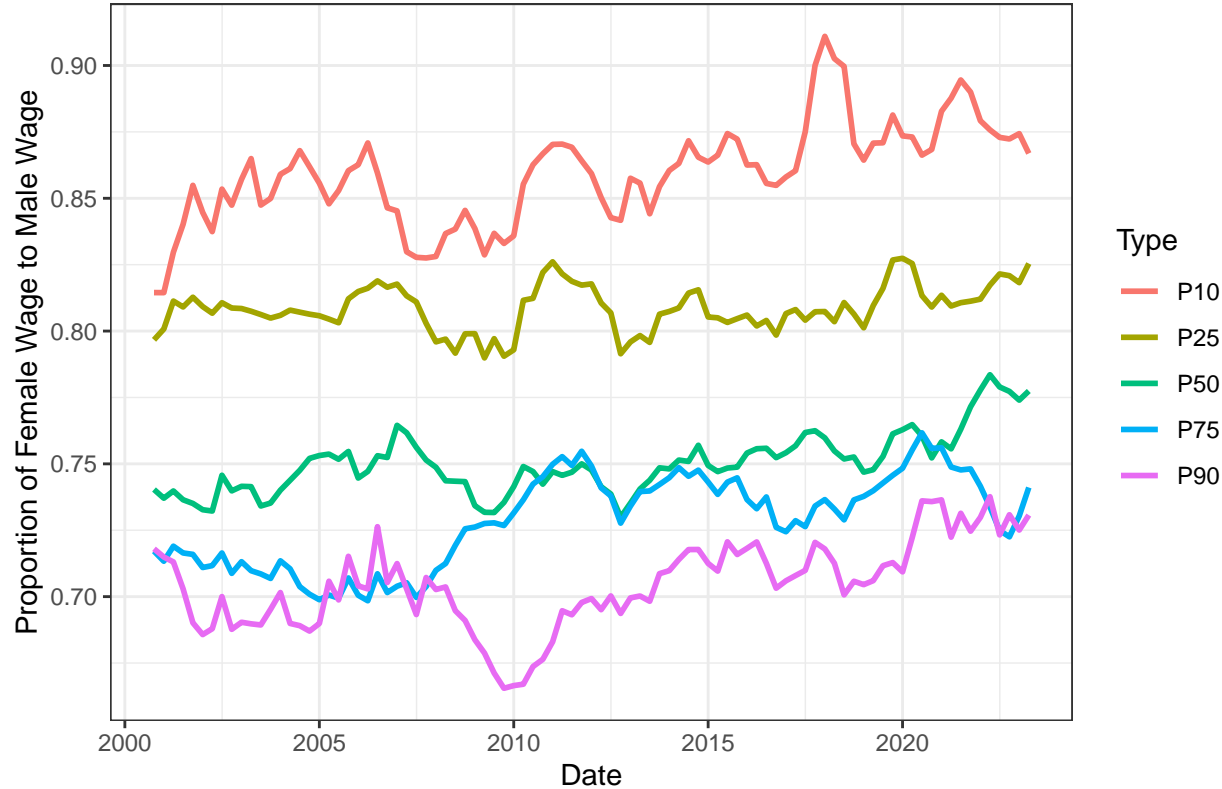
growth

Real Wages (2022 Q3–2023 Q2) as a Percent of 2019 by Race and Sex
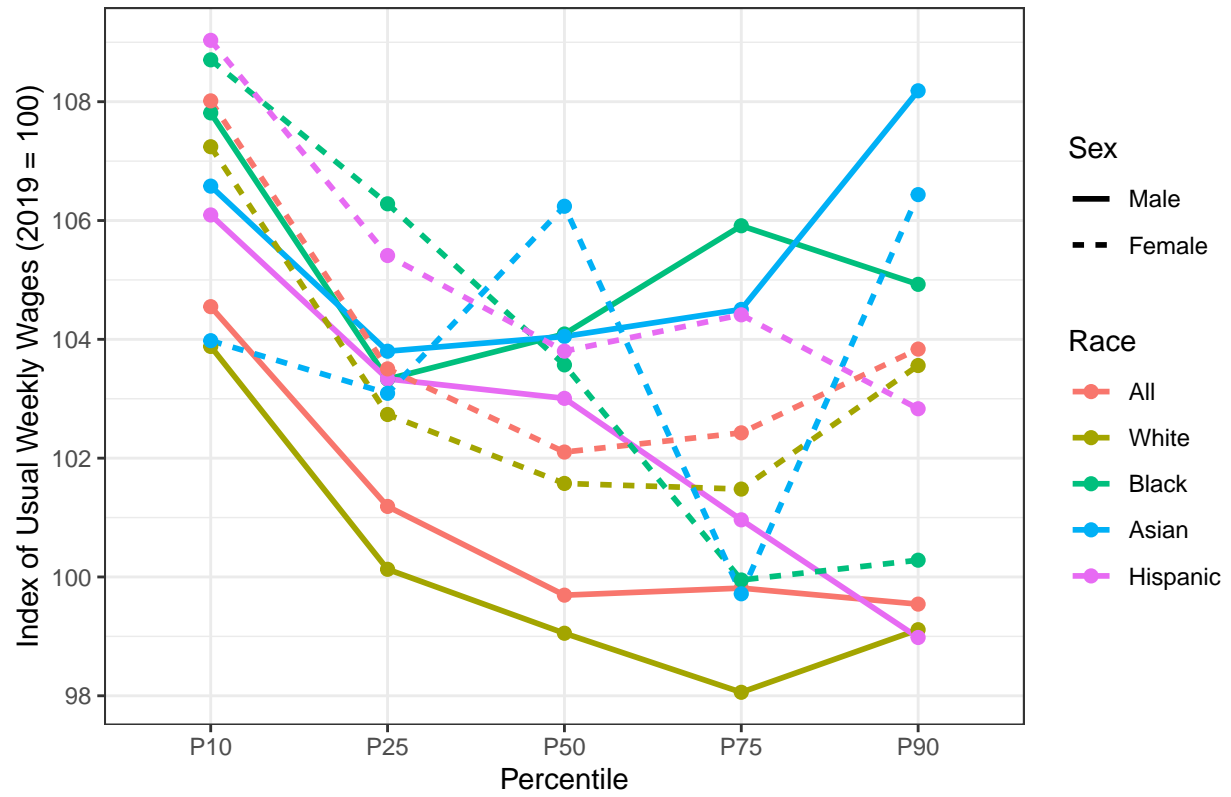
growth2

Real Wages (2022 Q3–2023 Q2) as a Percent of 2007 by Race and Sex