DAT181_Final

Nicolas Son

2023-11-11

```r
#load the data set
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(lattice) #for xyplot

wages <- read.csv("wages_by_education.csv")

wages %>% colnames()
```

```
##  [1] "year"                        "less_than_hs"
##  [3] "high_school"                 "some_college"
##  [5] "bachelors_degree"            "advanced_degree"
##  [7] "men_less_than_hs"            "men_high_school"
##  [9] "men_some_college"            "men_bachelors_degree"
## [11] "men_advanced_degree"         "women_less_than_hs"
## [13] "women_high_school"           "women_some_college"
## [15] "women_bachelors_degree"      "women_advanced_degree"
## [17] "white_less_than_hs"          "white_high_school"
## [19] "white_some_college"          "white_bachelors_degree"
## [21] "white_advanced_degree"       "black_less_than_hs"
## [23] "black_high_school"           "black_some_college"
## [25] "black_bachelors_degree"      "black_advanced_degree"
## [27] "hispanic_less_than_hs"       "hispanic_high_school"
## [29] "hispanic_some_college"       "hispanic_bachelors_degree"
## [31] "hispanic_advanced_degree"    "white_men_less_than_hs"
## [33] "white_men_high_school"       "white_men_some_college"
## [35] "white_men_bachelors_degree"  "white_men_advanced_degree"
## [37] "black_men_less_than_hs"      "black_men_high_school"
## [39] "black_men_some_college"      "black_men_bachelors_degree"
## [41] "black_men_advanced_degree"   "hispanic_men_less_than_hs"
## [43] "hispanic_men_high_school"    "hispanic_men_some_college"
## [45] "hispanic_men_bachelors_degree"  "hispanic_men_advanced_degree"
```

```
## [47] "white_women_less_than_hs"        "white_women_high_school"
## [49] "white_women_some_college"        "white_women_bachelors_degree"
## [51] "white_women_advanced_degree"     "black_women_less_than_hs"
## [53] "black_women_high_school"         "black_women_some_college"
## [55] "black_women_bachelors_degree"    "black_women_advanced_degree"
## [57] "hispanic_women_less_than_hs"     "hispanic_women_high_school"
## [59] "hispanic_women_some_college"     "hispanic_women_bachelors_degree"
## [61] "hispanic_women_advanced_degree"
```

```r
#Smaller dataset, so I can visually check and clean data where need be. If given
#a large dataset, I would use wages %>% na.omit(), drop unrelated columns, check
#data types to make sure it'd match, etc.

#I wanted to do all columns given, but couldnt exceed 50 rows, so settling for
#average pay, pay for white people, and
wage_data <- data.frame(years = c(wages$year),
                        average_less_than_hs = c(wages$less_than_hs),
                        average_high_school = c(wages$high_school),
                        average_some_college = c(wages$some_college),
                        average_bachelors_degree = c(wages$bachelors_degree),
                        average_advanced_degree = c(wages$advanced_degree),

                        men_less_than_hs = c(wages$men_less_than_hs),
                        men_high_school = c(wages$men_high_school),
                        men_some_college = c(wages$men_some_college),
                        men_bachelors_degree = c(wages$men_bachelors_degree),
                        men_advanced_degree = c(wages$men_advanced_degree),

                        women_less_than_hs = c(wages$women_less_than_hs),
                        women_high_school = c(wages$women_high_school),
                        women_some_college = c(wages$women_some_college),
                        women_bachelors_degree = c(wages$women_bachelors_degree),
                        women_advanced_degree = c(wages$women_advanced_degree))
```
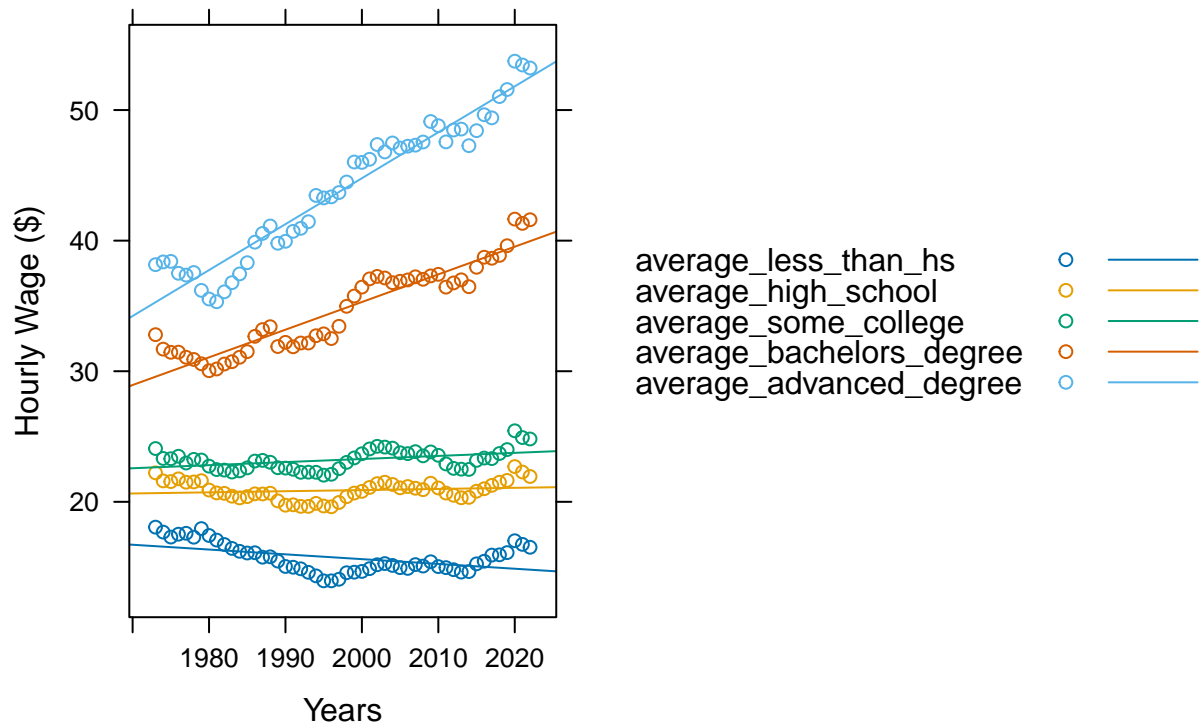
```r
#Y is left of ~, X is right of ~, wage_data is the DF and auto.key is legend.
xyplot(average_less_than_hs + average_high_school + average_some_college +
       average_bachelors_degree + average_advanced_degree ~
       years, wage_data, auto.key = TRUE,
       xlab = "Years",
       ylab = "Hourly Wage ($)",
       main = "Average Hourly Wage from 1973 to 2022",
       #points, regression line
       type = c("p", "r"))
```
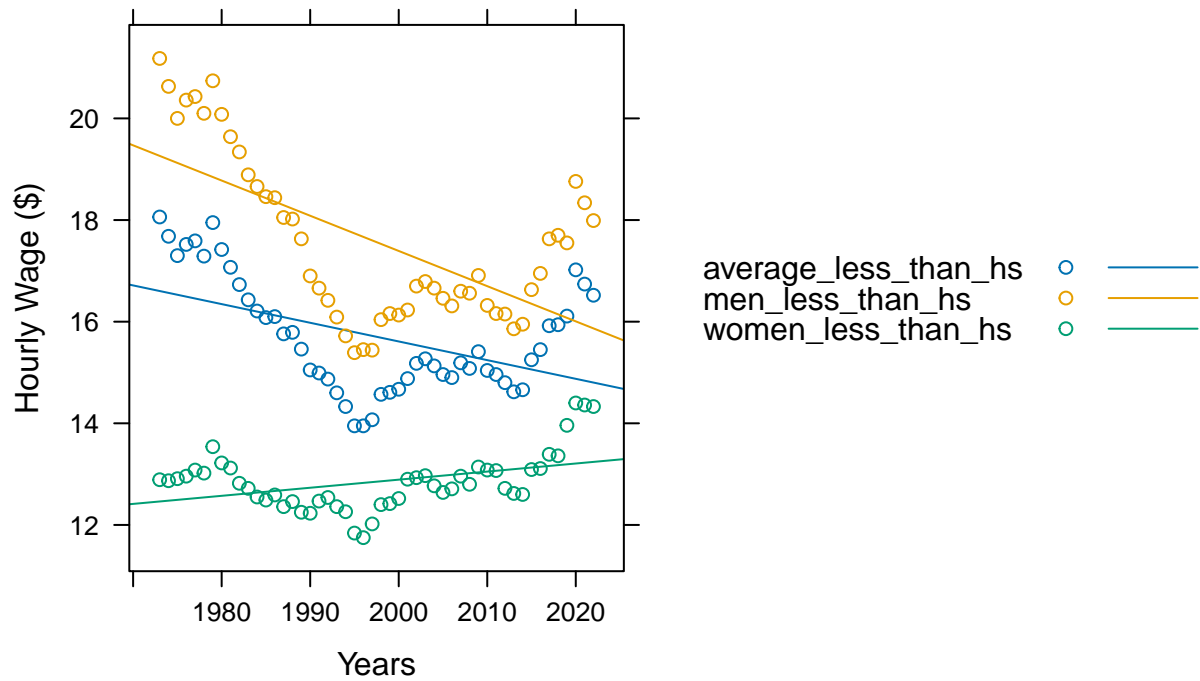
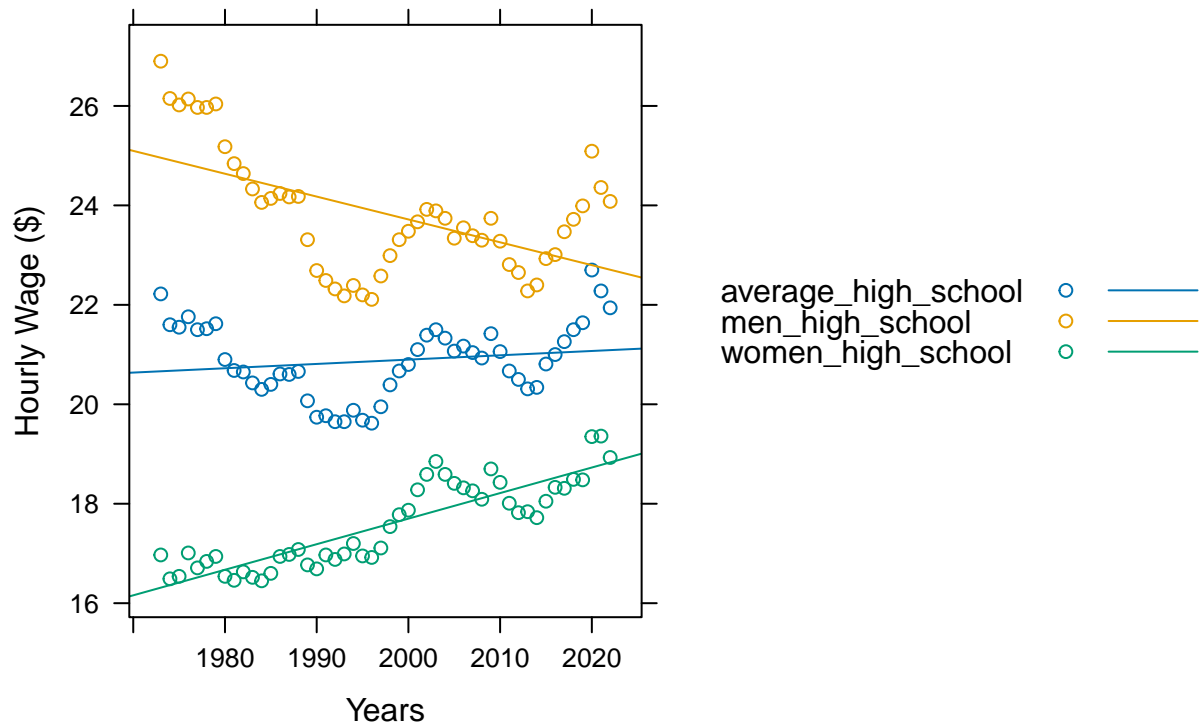# Average Hourly Wage from 1973 to 2022


```
xyplot(average_less_than_hs + men_less_than_hs + women_less_than_hs ~
    years, wage_data, auto.key = TRUE,
    xlab = "Years",
    ylab = "Hourly Wage ($)",
    main = "Average, Male, Female, Less Than High School Hourly Wage
    from 1973 to 2022",
    #points, regression line
    type = c("p", "r"))
```

Average, Male, Female, Less Than High School Hourly Wage
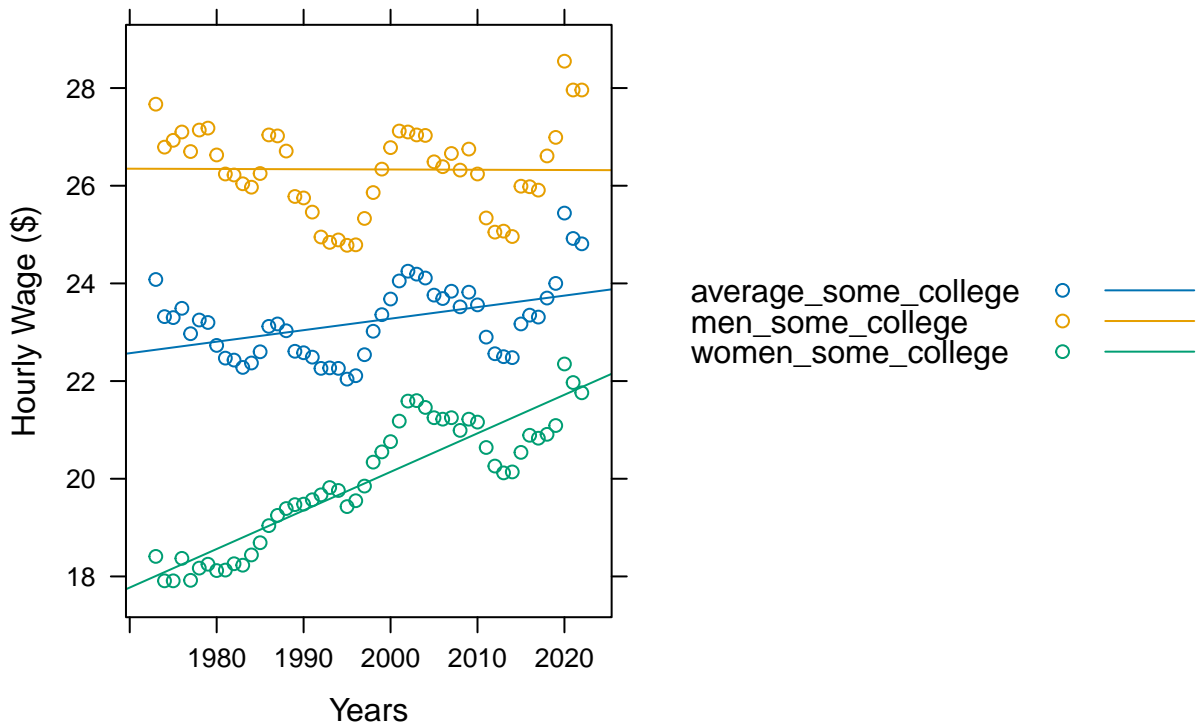# from 1973 to 2022


```
xyplot(average_high_school + men_high_school + women_high_school ~
    years, wage_data, auto.key = TRUE,
    xlab = "Years",
    ylab = "Hourly Wage ($)",
    main = "Average, Male, Female, High School Hourly Wage from 1973 to 2022",
    #points, regression line
    type = c("p", "r"))
```

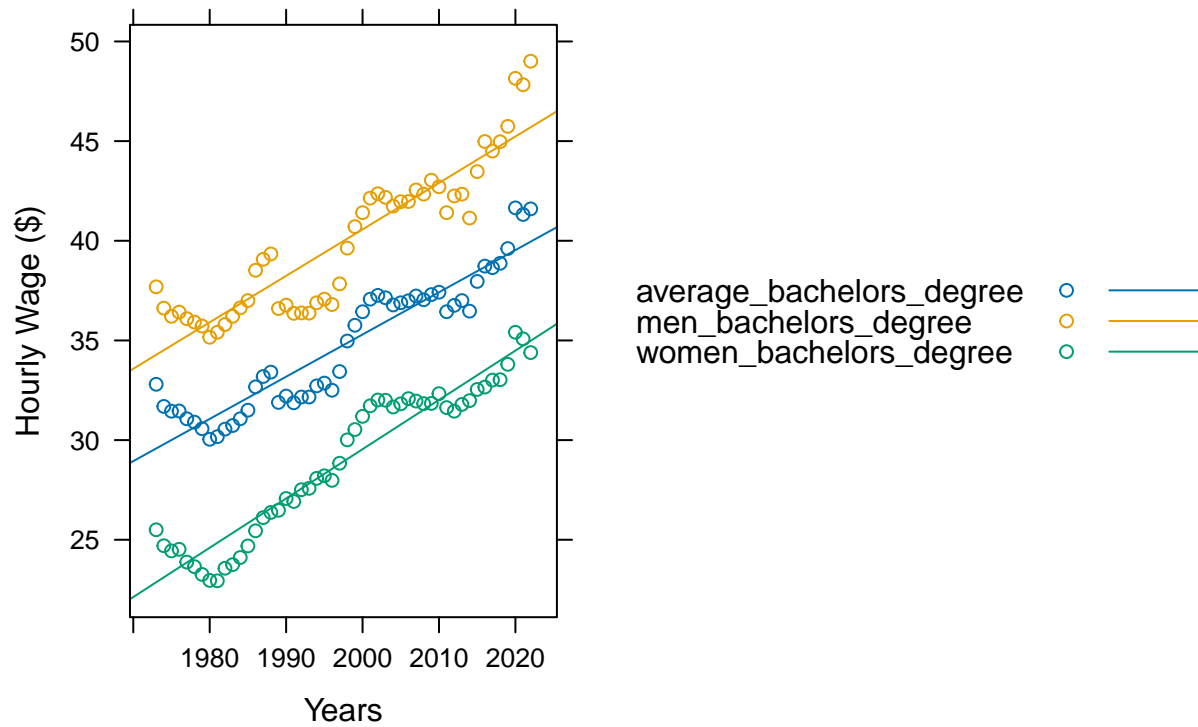# Average, Male, Female, High School Hourly Wage from 1973 to 2022


```r
xyplot(average_some_college + men_some_college + women_some_college ~ years,
       wage_data, auto.key = TRUE,
       xlab = "Years",
       ylab = "Hourly Wage ($)",
       main = "Average, Male, Female, Some College Hourly Wage from 1973 to 2022",
       #points, regression line
       type = c("p", "r"))
```

Average, Male, Female, Some College Hourly Wage from 1973 to 2022
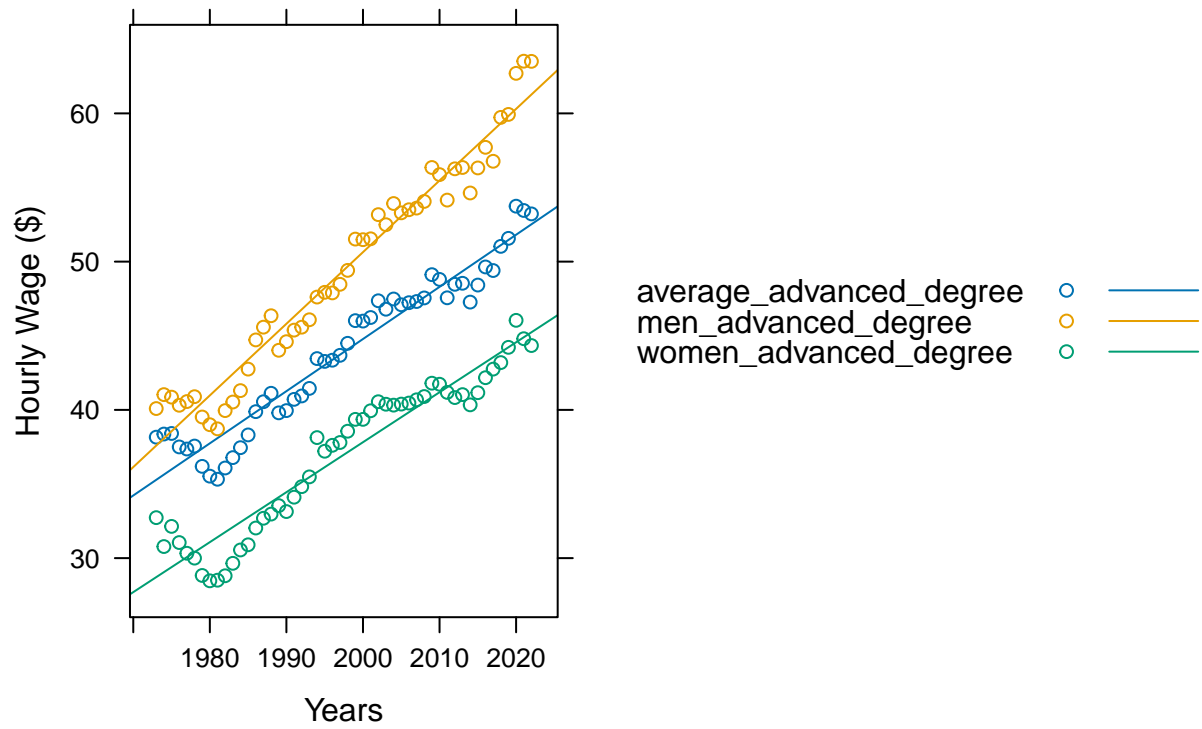


```
xyplot(average_bachelors_degree + men_bachelors_degree + women_bachelors_degree
    ~ years, wage_data, auto.key = TRUE,
    xlab = "Years",
    ylab = "Hourly Wage ($)",
    main = "Average, Male, Female, Bachelors Degree Hourly Wage from 1973 to 2022",
    #points, regression line
    type = c("p", "r"))
```

verage, Male, Female, Bachelors Degree Hourly Wage from 1973 to 202


```
xyplot(average_advanced_degree + men_advanced_degree + women_advanced_degree ~
    years, wage_data, auto.key = TRUE,
    xlab = "Years",
    ylab = "Hourly Wage ($)",
    main="Average, Male, Female, Advanced Degree Hourly Wage from 1973 to 2022",
    #points, regression line
    type = c("p", "r"))
```

```
#?xyplot
```