Labor Market Analysis

Exploratory Data Analysis

Marjorie Blanco, Joe Thomson, Haodi Tu

Data

We used data from the 2016 ACS for Puerto Rico to examine wage gaps between individuals with different education levels. Our research questions are: 1) How do earnings vary by education level? 2) How does the premium for education vary by gender? The 2016 ACS is a nationally representative sample of 5194. The household survey includes questions pertaining to each household member's demographic characteristics and labor market activity.

We restrict our sample to these three racial groups: White, Black and Other. In addition, given our goal of examining earning differences by gender and marital status and the reporting of earnings in the ACS on an annual basis (wages, salary, commissions, bonuses, tips, and self-employment income during the past 12 months), we restrict our sample to full-time year-round (FTYR) workers. We define FTYR workers as individuals who report positive earnings over the past year, who worked at least 40 of the past 52 weeks, and who worked at least 35 hours per week in a usual work week over this period.

EDA Insights:

For our exploratory analysis we looked at population breakdowns by education, age, marital status, gender, race, earnings, and work hours. We applied filters on education (HS diploma or above), age (18-64), and work hours (>35/week).

An earnings histogram identified a default maximum amount of earnings (189k) which we also filtered out of the data. The earning distribution is progressive above the median, but drops off sharply below the median, likely indicating the presence of a minimum wage. The correlation between age and earnings is very weak (.23). Likewise, earnings is very weakly correlated with hours worked among those who work more than 35 hours per week. However, white individuals appear to have an earnings premium over other races, and both married and divorced individuals appear to have an earnings premium over those who have never been married. Given that the correlation between age and earnings was weak, this may be due to other qualitative factors possessed by those who get married. Married was recategoried to married and not married.

Earnings appear positively correlated with how well people speak English, as well as with higher levels of education. It should be noted that, on the island, fluency in the English language is neither a requirement, nor really needed. As a Latin American destination, the predominant language spoken and used (e.g. street signs, day-to-day communications) is Spanish, with English only coming into use in the tourism industry, or in those industies or companies imported from the mainland. This may help to explain the earning difference, as those are likely to pay more than local businesses (note that tourism is the largest industry on the island and the source of most of the island's GDP). Men also appear to earn a small premium over women.

The age distribution of full time workers is skewed towards older adults, possibly indicating that younger workers have trouble finding full-time work, wait to enter the workforce, or are leaving the territory.

Task 1:

Examine the first 10 or 20 observations (rows of data) corresponding to variables of interest (columns) and compare the observed values to the data dictionary for person records.

Earnings	Sex	Age	Race	Marital Status	Education	Work Week	Work Hou
34000	Male	47	White	Married	Associate's degree	50 to 52	
13000	Male	58	Black or African American	Never married	High school diploma	50 to 52	
18000	Male	50	White	Married	Master's degree	50 to 52	
10300	Female	39	White	Married	Bachelor's degree	50 to 52	
28600	Female	39	Black or African American	Married	Bachelor's degree	50 to 52	
24800	Male	37	Black or African American	Married	Bachelor's degree	50 to 52	
22000	Female	47	Some Other	Never married	Associate's degree	50 to 52	
19000	Female	60	Black or African American	Never married	High school diploma	50 to 52	
87000	Female	58	White	Divorced	Associate's degree	50 to 52	
22900	Male	61	White	Divorced	High school diploma	50 to 52	
19000	Male	39	Black or African American	Married	Associate's degree	50 to 52	
19600	Female	36	Black or African American	Married	Bachelor's degree	50 to 52	
48000	Male	30	Two or More Races	Divorced	Some college	50 to 52	
40000	Female	30	White	Never married	Some college	50 to 52	
15600	Female	41	White	Never married	High school diploma	50 to 52	
12100	Male	46	White	Divorced	High school diploma	50 to 52	
14000	Male	53	White	Married	High school diploma	50 to 52	
80000	Male	38	White	Never married	Bachelor's degree	50 to 52	
15100	Female	26	Some Other	Never married	High school diploma	50 to 52	
84000	Female	60	White	Married	Doctorate degree	50 to 52	

Task 2:

Compute and examine descriptive statistics including the minimum, maximum, mean, and median for quantitative variables of interest

Total person's earnings

	ss16ppr (N = 5,194)
Minimum	10000.00
Maximum	125000.00
Median	24000.00
Mean	29278.84

Age

	ss16ppr (N = 5,194)
Minimum	18.00
Maximum	64.00
Median	43.00
Mean	29278.84

Hours worked

	ss16ppr (N = 5,194)
Minimum	35.00000
Maximum	99.00000
Median	40.00000
Mean	41.21544

Race: White

In Puerto Rico, the majority of people identify themselves as white. Minority races including American Indian, Alaska Native, Asian, Native Hawaiian and Other Pacific Islander can be eliminated.

RACWHT	Count
No	1401
Yes	3793

Race: Black

RACBLK	Count
No	4417
Yes	777

Race: Other

RACOTHER	Count
No	4381
Yes	813

Marital status

MAR	Count
Married	2580
Widowed	65
Divorced	944
Separated	105
Never married	1500

MAR1	Count
No	2614
Yes	2580

MAR2	Count
Married	2645
Divorced	1049
Never married	1500

Educational attainment

- We will categorize people into seven education levels
- HS diploma or equivalent
- Some college
- Associate's degree
- Bachelor's degree
- Professional degree
- Doctorate degree

SCHL	Count
High school diploma	1148
Some college	791
Associate's degree	818
Bachelor's degree	1726
Master's degree	505
Professional degree	113
Doctorate degree	93

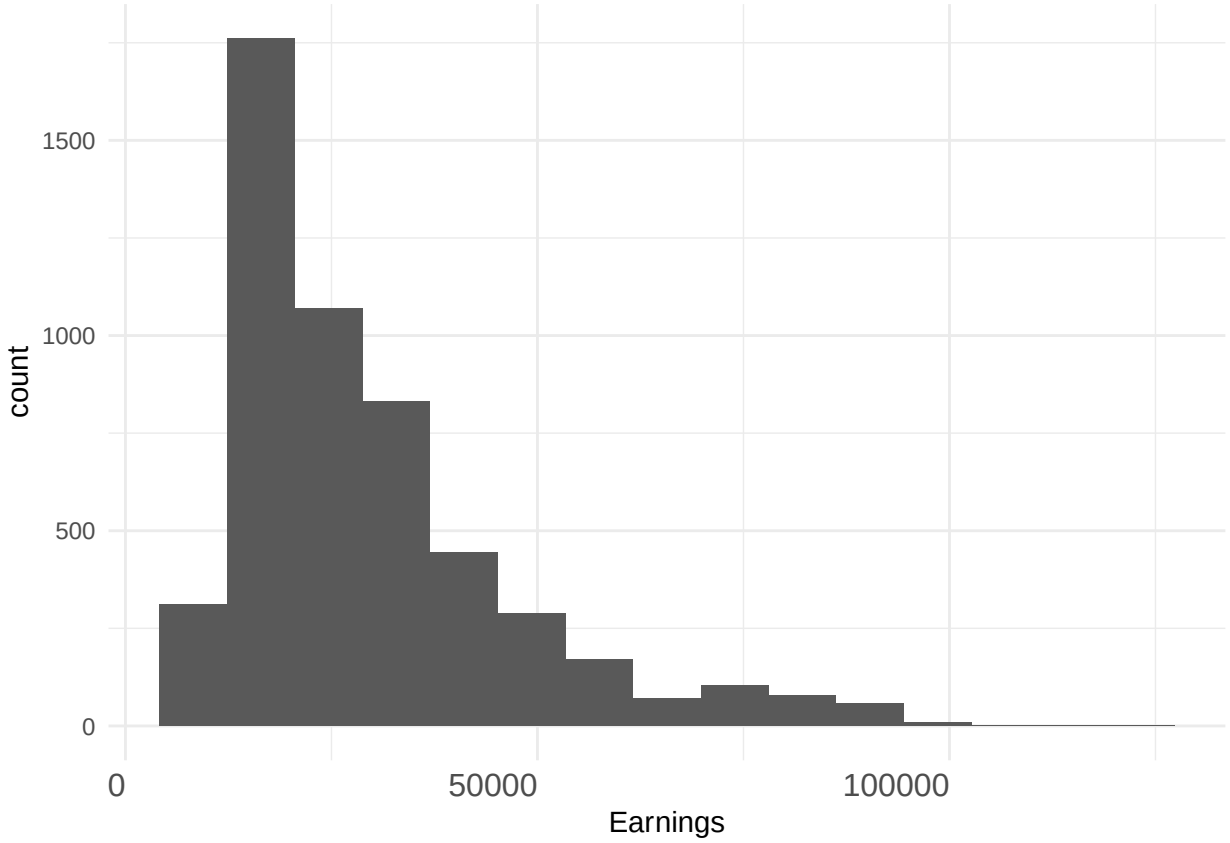
Sex

SEX	Count
Male	2627
Female	2567

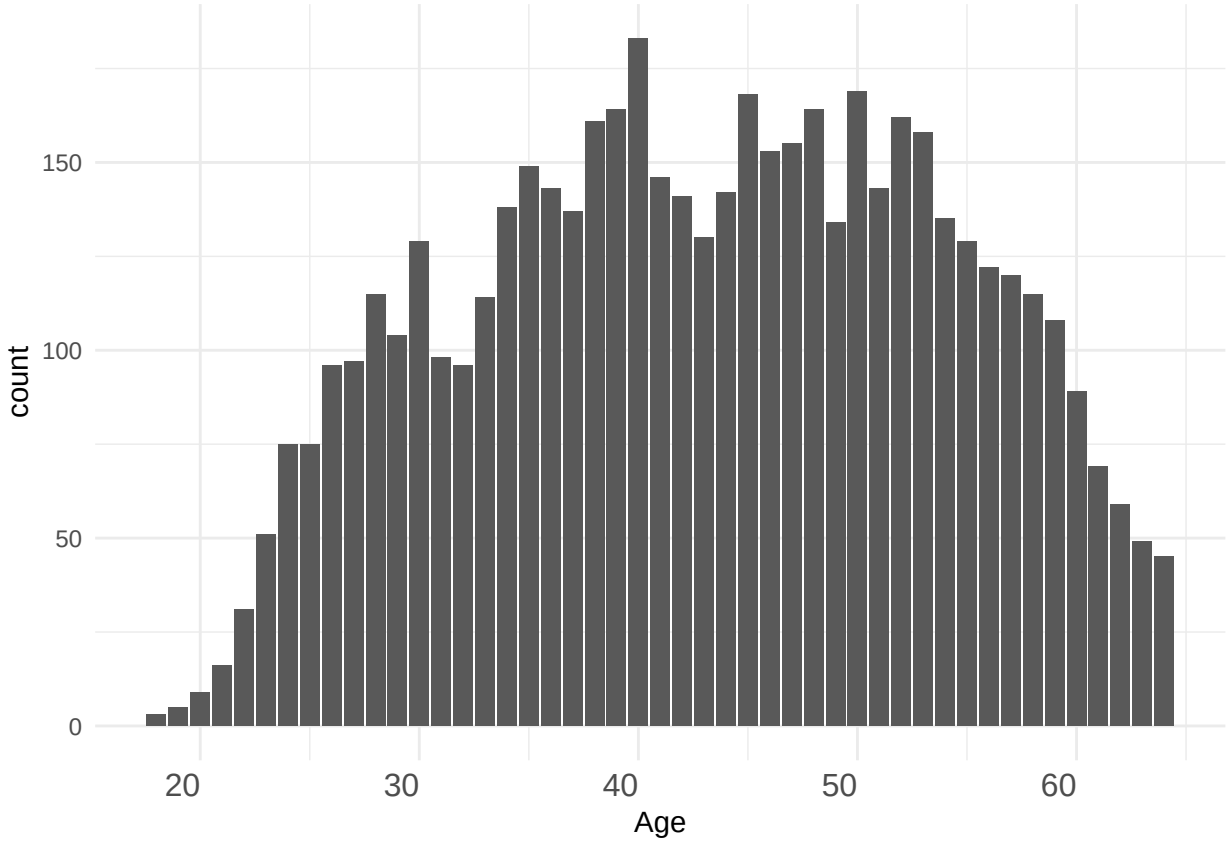
Task 3:

Generate and examine histograms for quantitative variables of interest

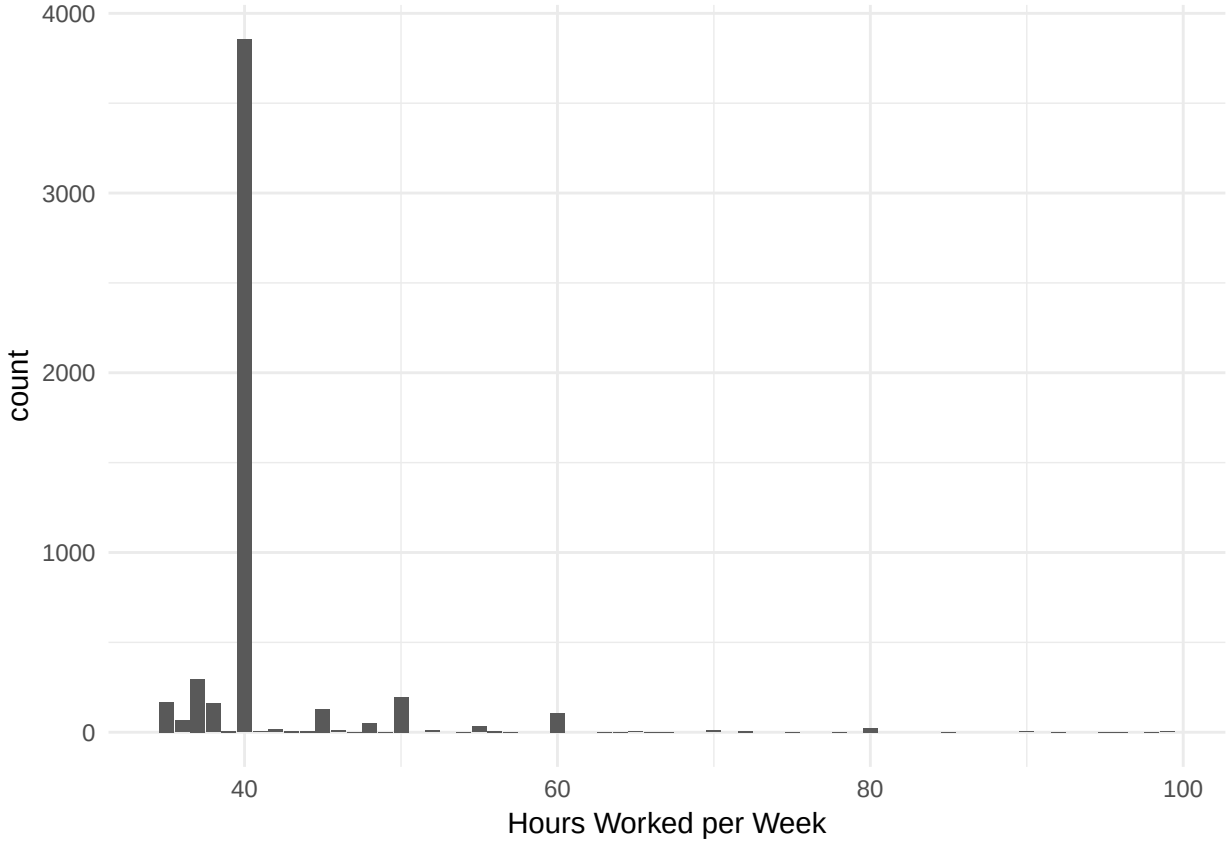
Total person's earnings



Age



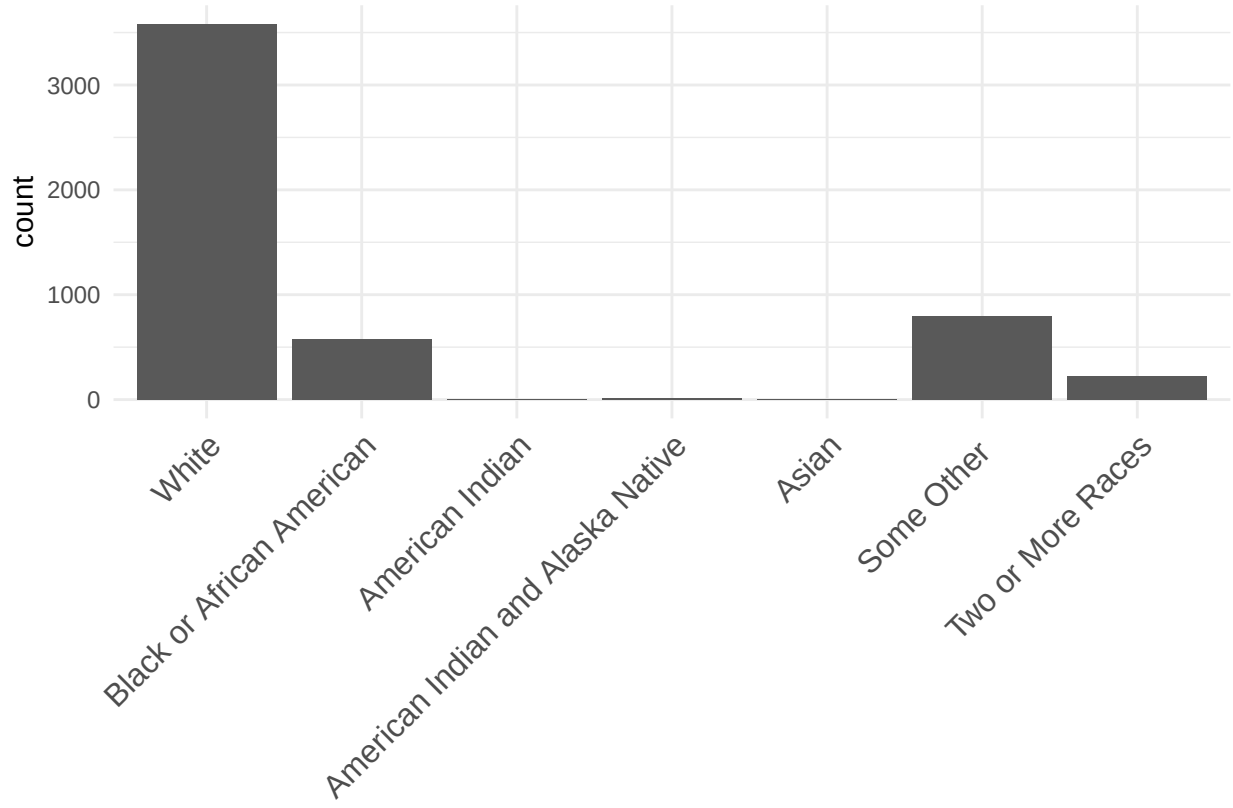
Hours worked

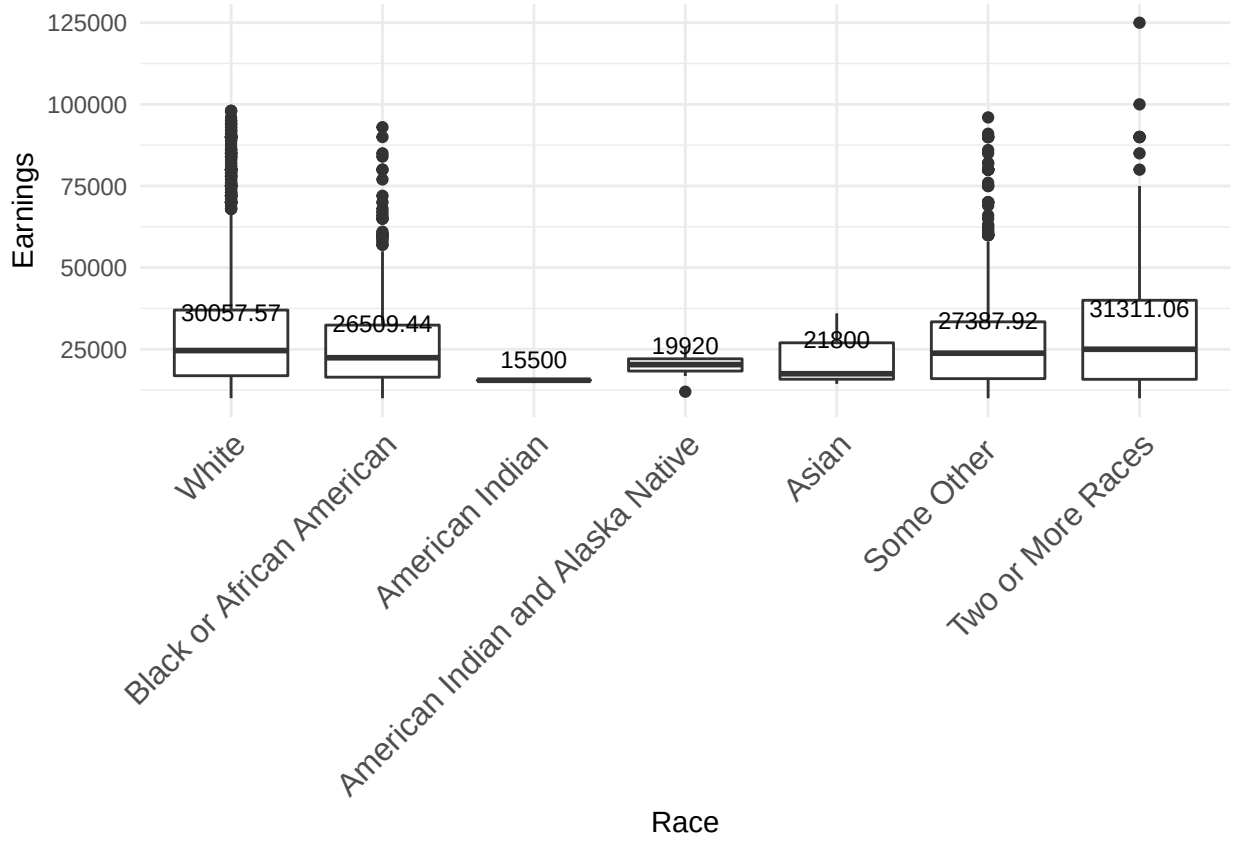


Task 4:

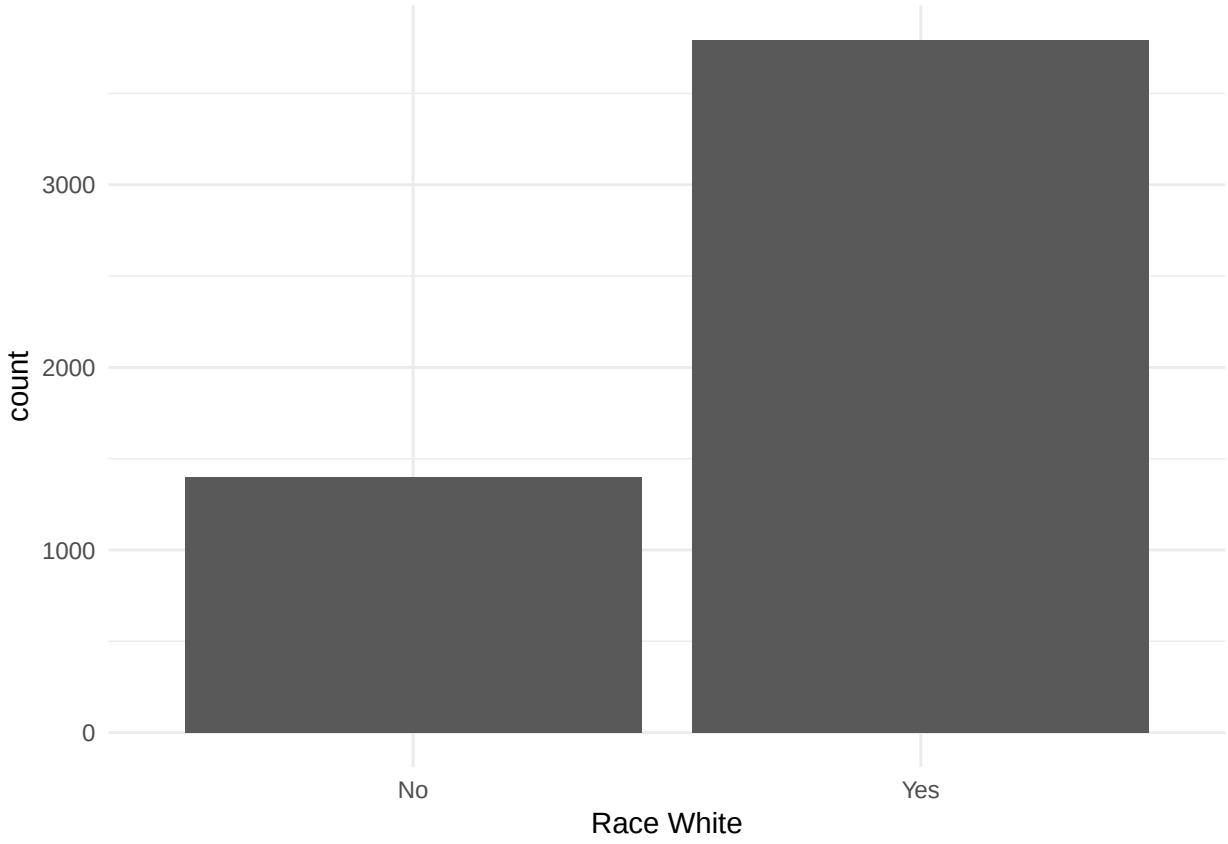
Generate and examine bar charts/graphs for qualitative variables of interest

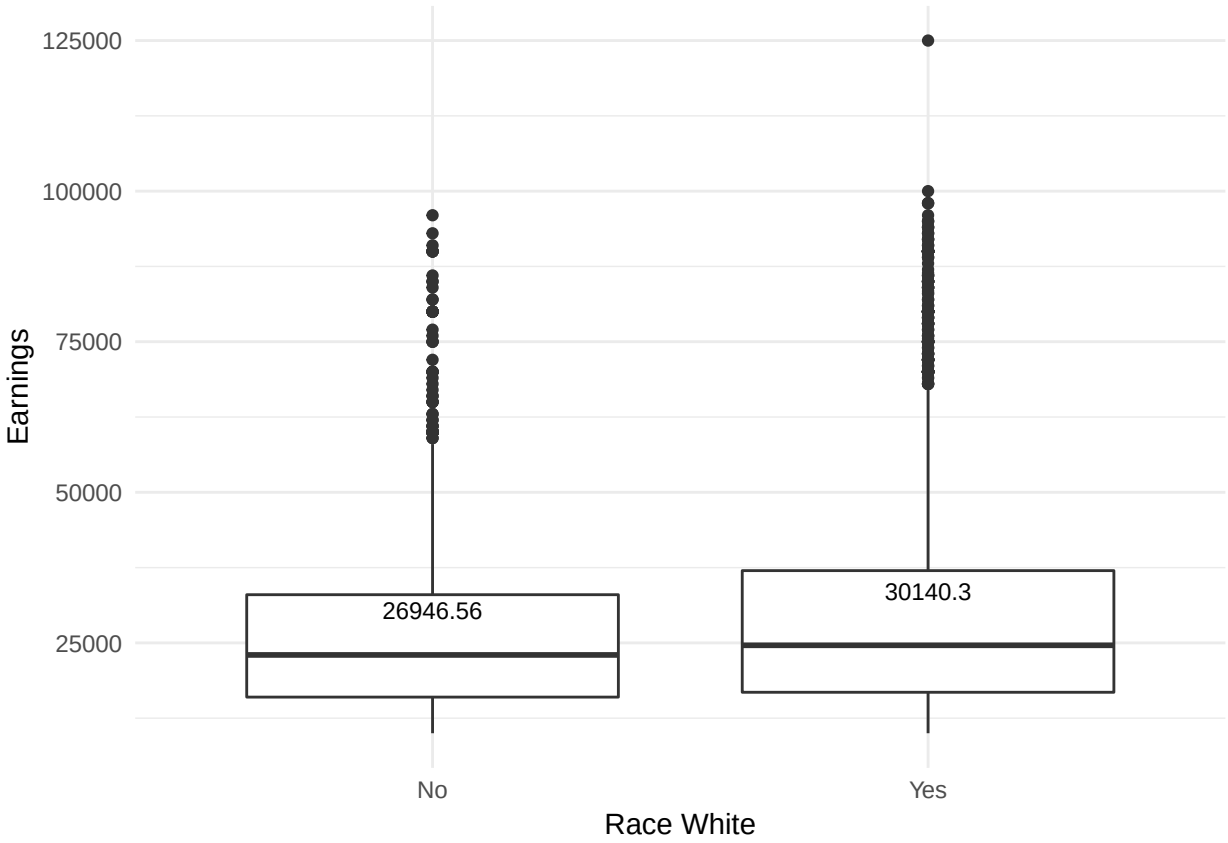
Race


count
3000
2000
1000
0
White
Black or African American
American Indian
American Indian and Alaska Native
Asian
Some Other
Two or More Races
Race

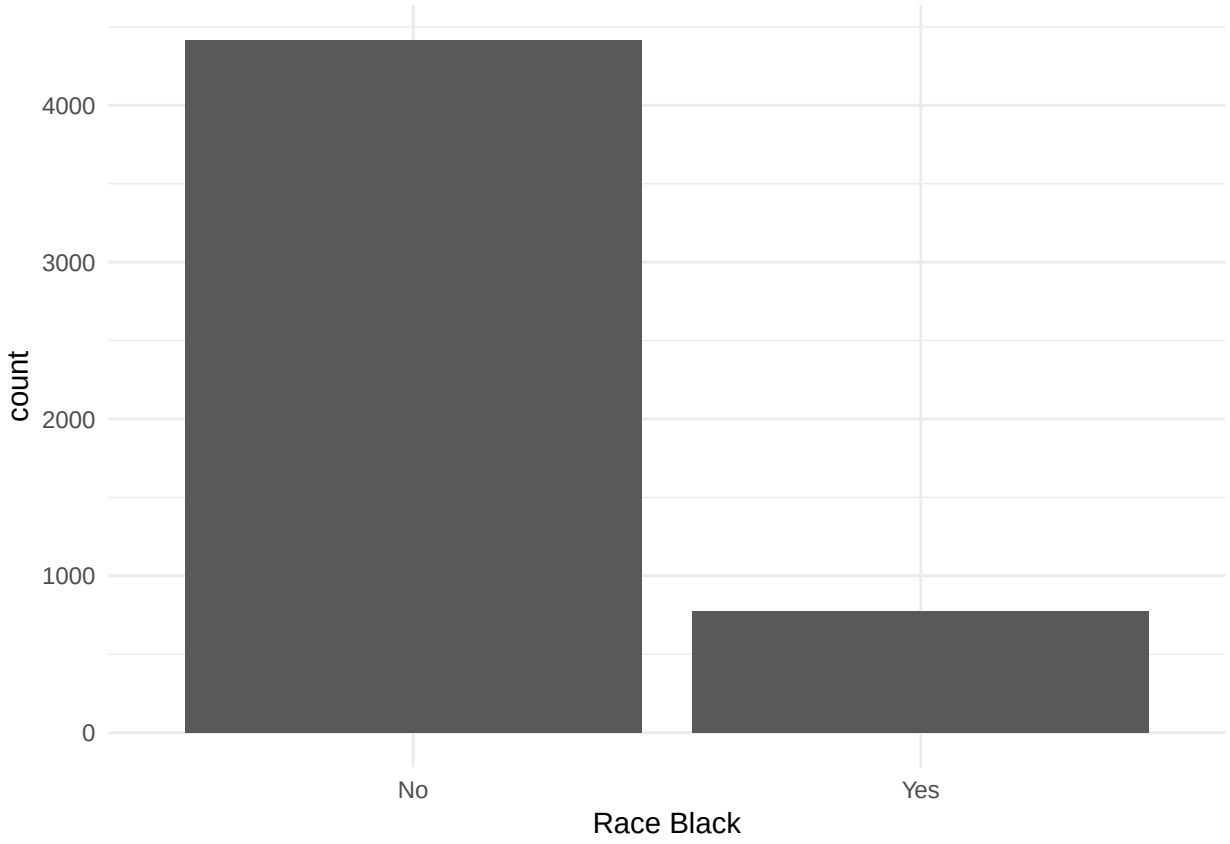

Earnings
125000
100000
75000
50000
25000
30057.57
26509.44
15500
19920
21800
27387.92
31311.06
White
Black or African American
American Indian
American Indian and Alaska Native
Asian
Some Other
Two or More Races
Race

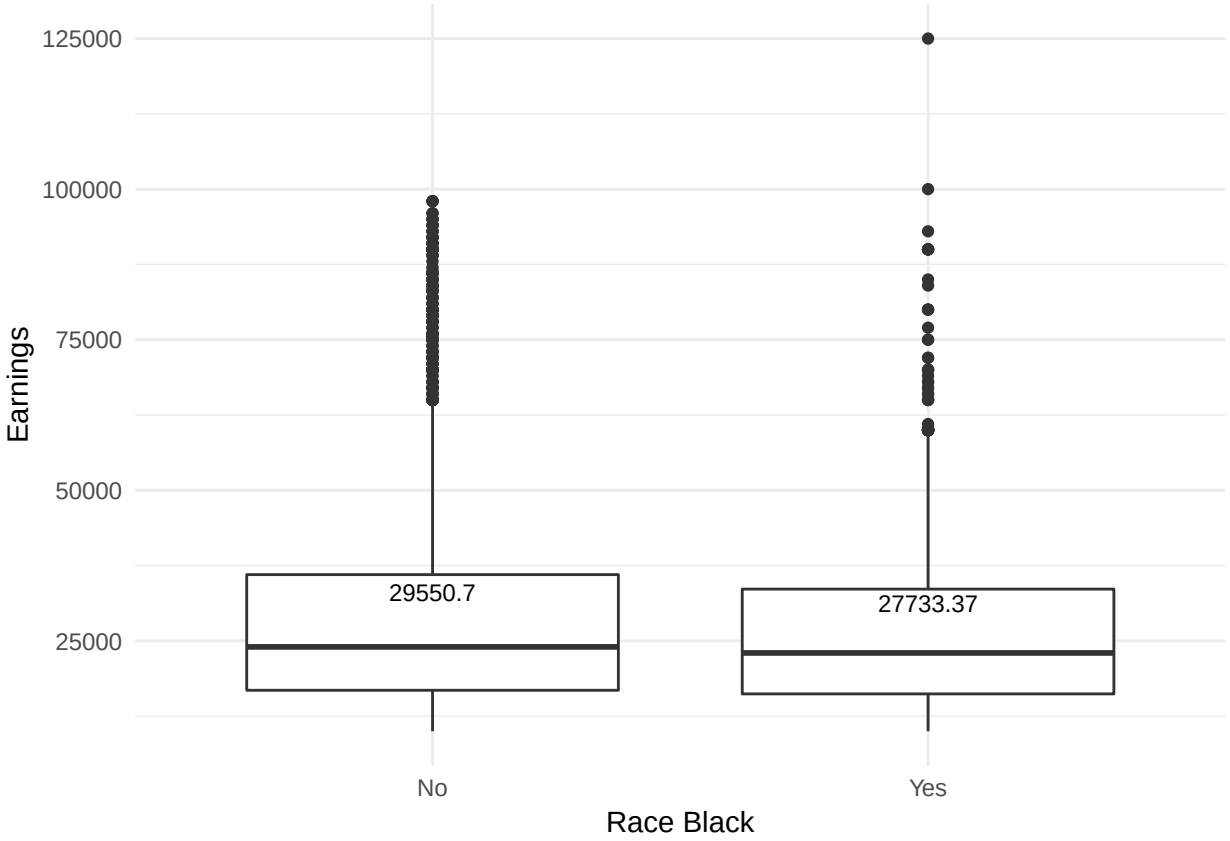
Race: White


3000
2000
1000
0
count
No
Yes
Race White

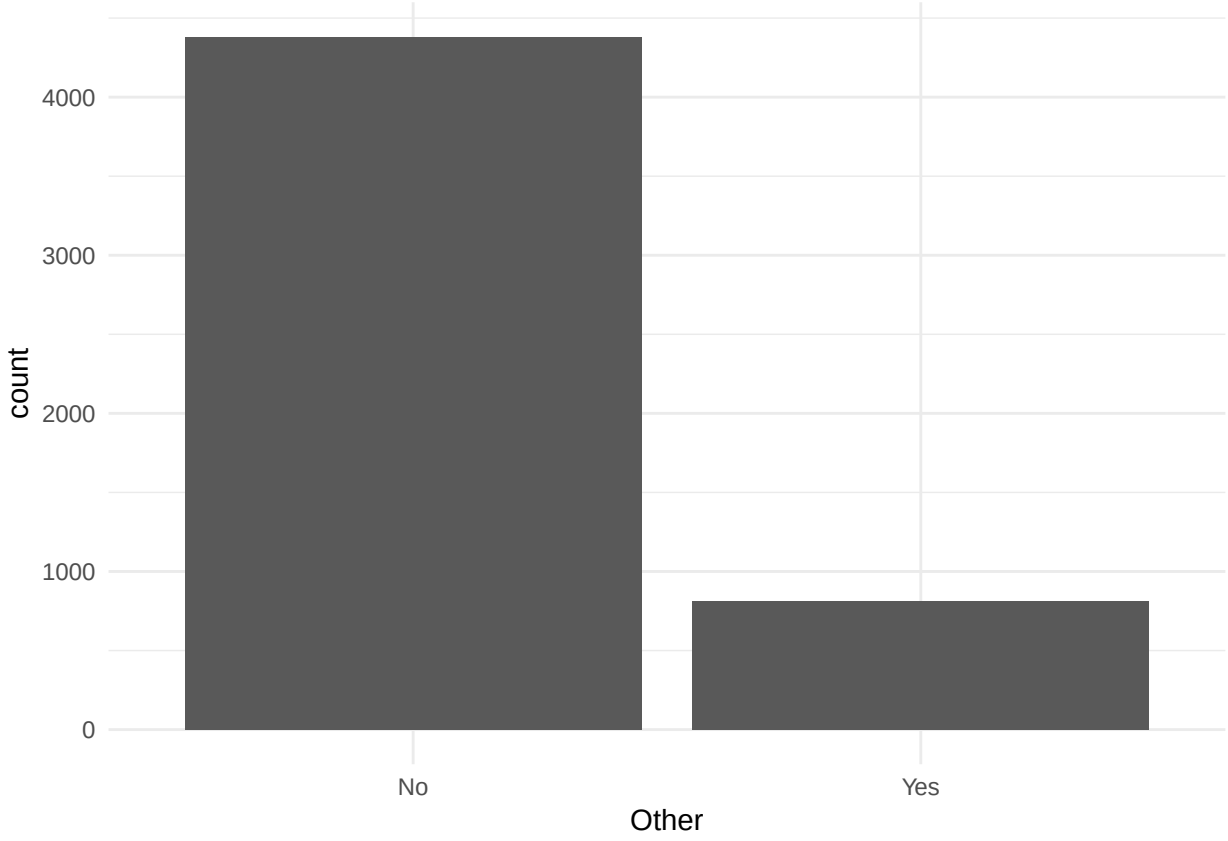

125000
100000
75000
50000
25000
Earnings
26946.56
30140.3
No
Yes
Race White

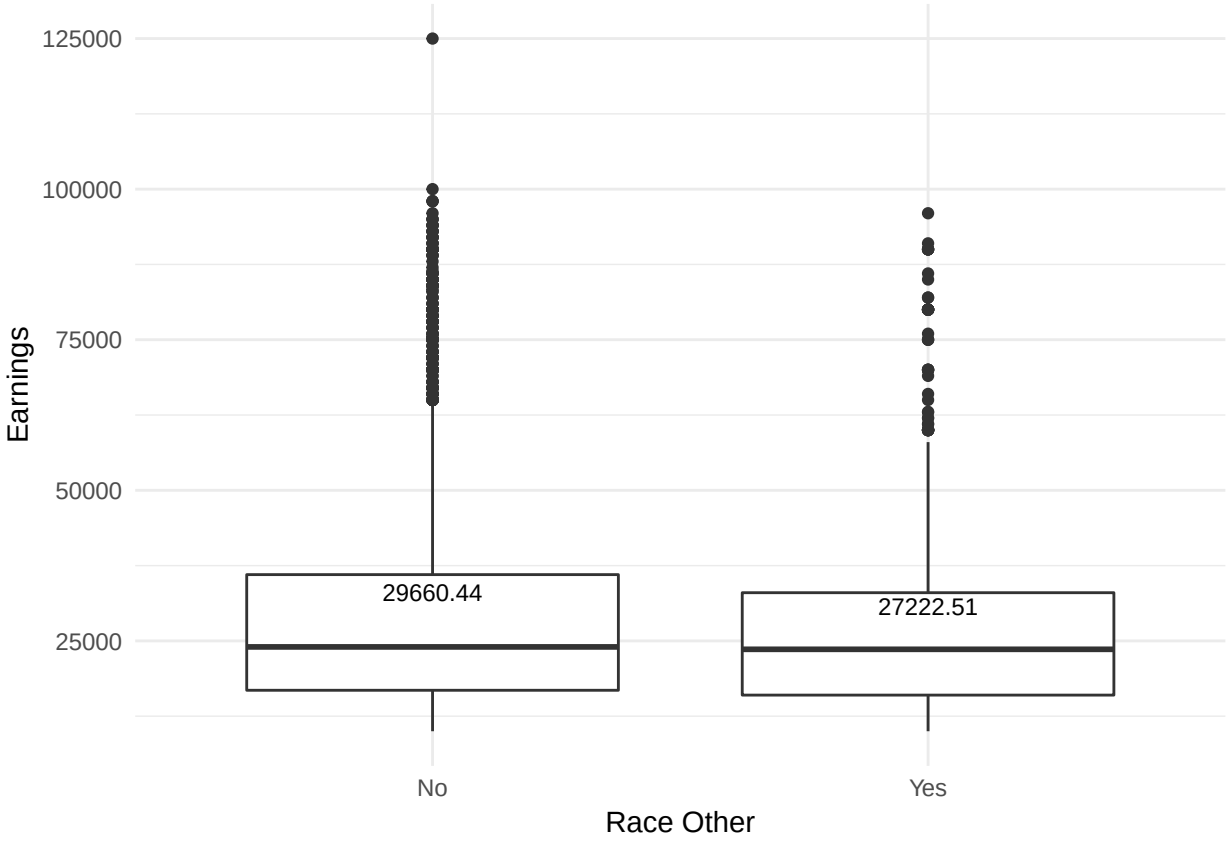
Race: Black


4000
3000
2000
1000
0
count
No
Yes
Race Black

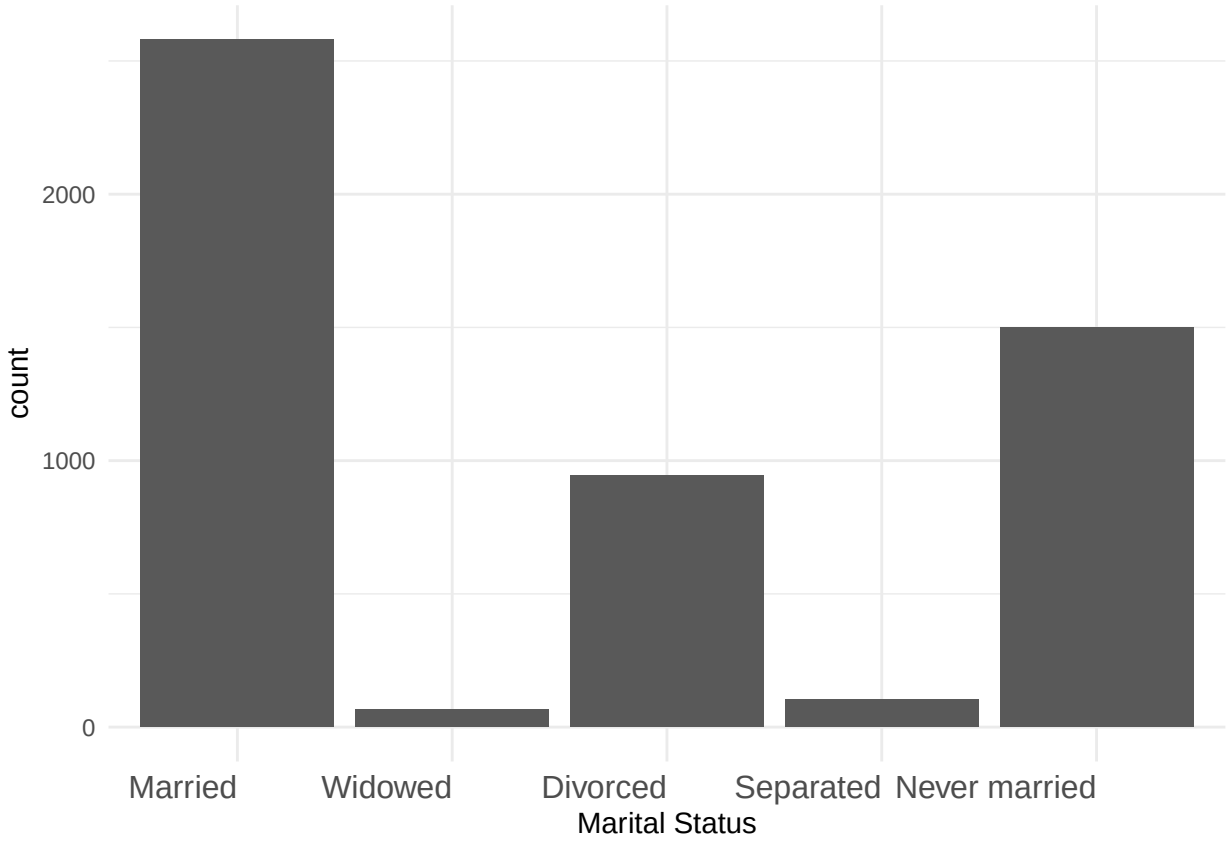

125000
100000
75000
50000
25000
Earnings
29550.7
27733.37
No
Yes
Race Black

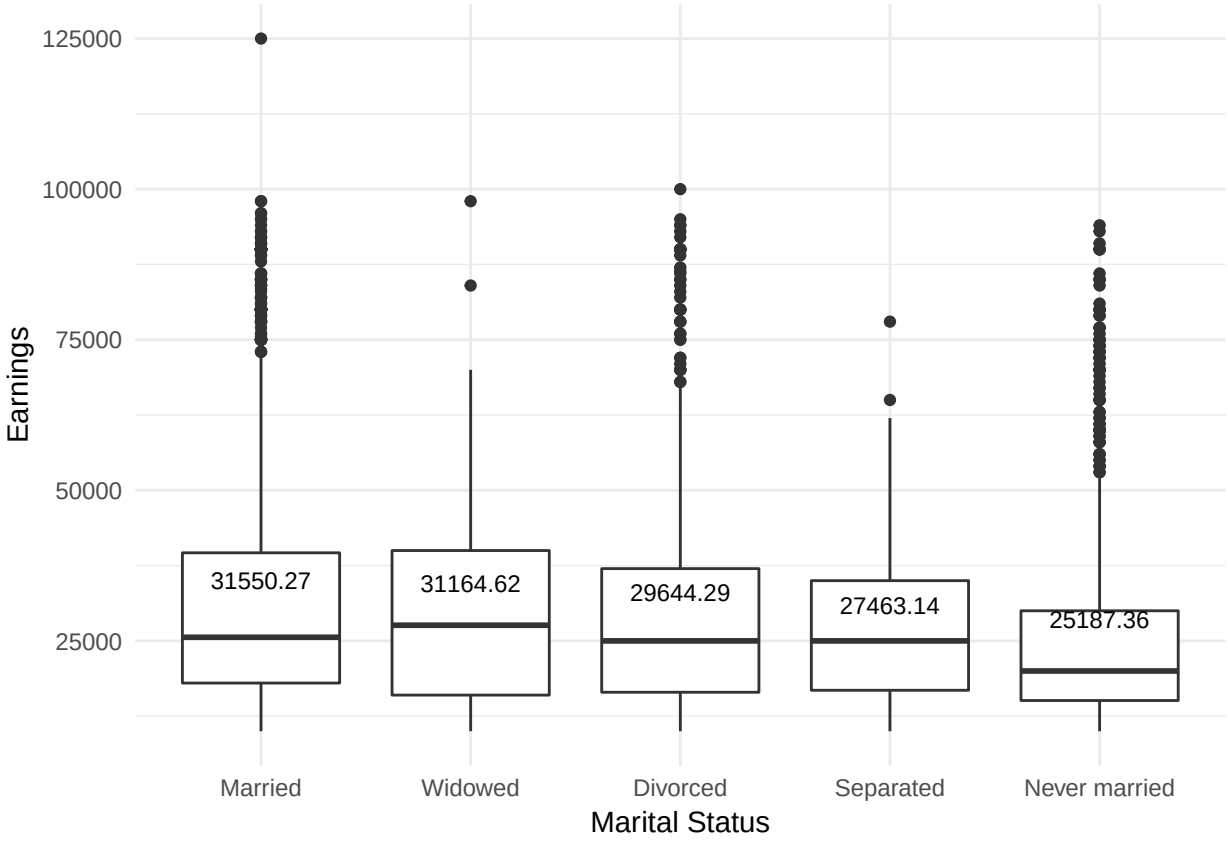
Race: Other

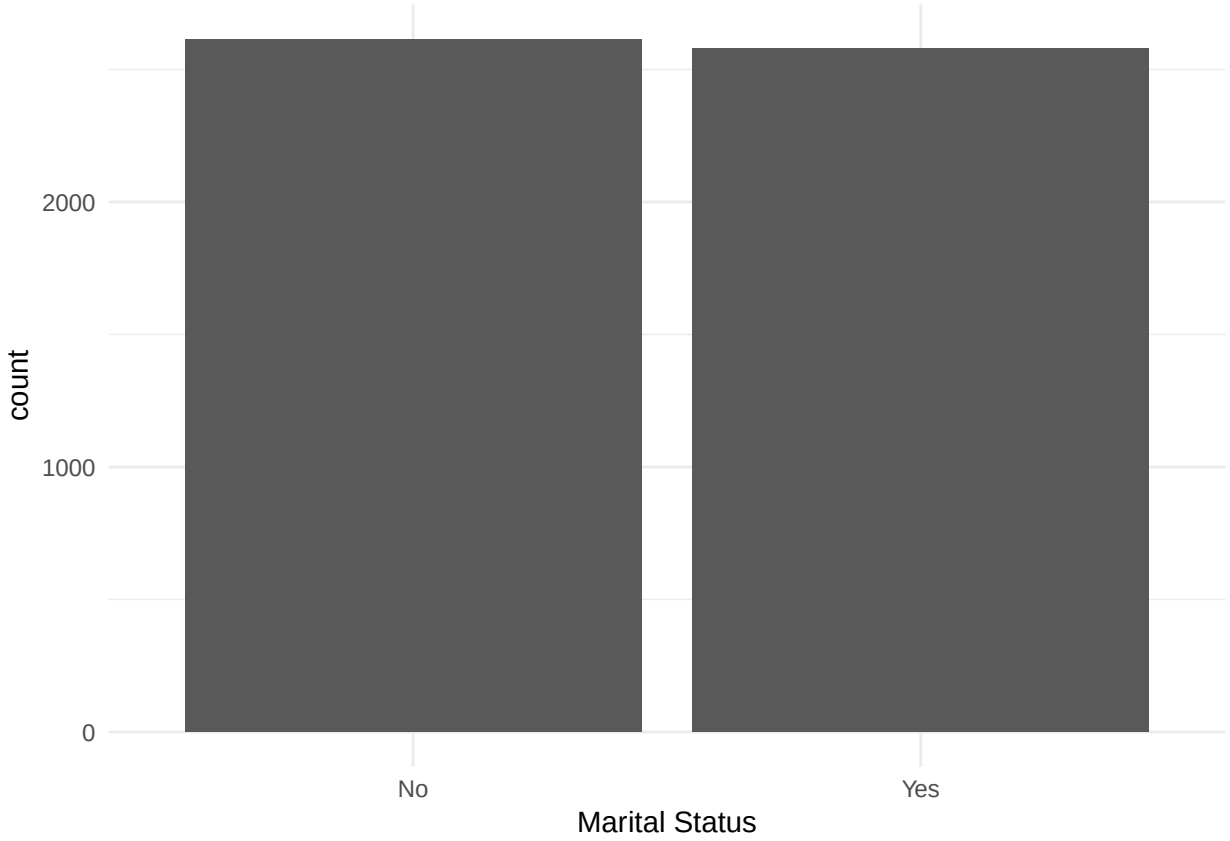

4000
3000
2000
1000
0
count
No
Yes
Other

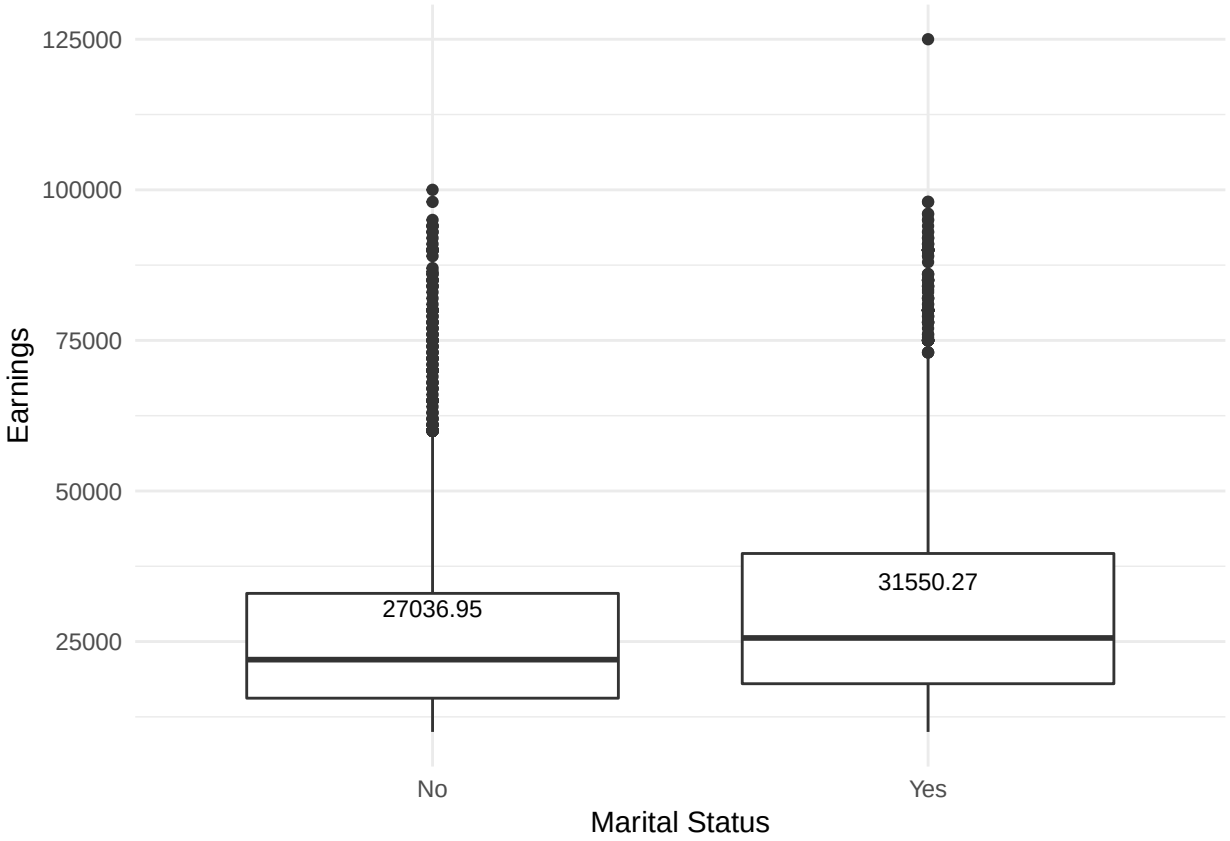

125000
100000
75000
50000
25000
Earnings
29660.44
27222.51
No
Yes
Race Other

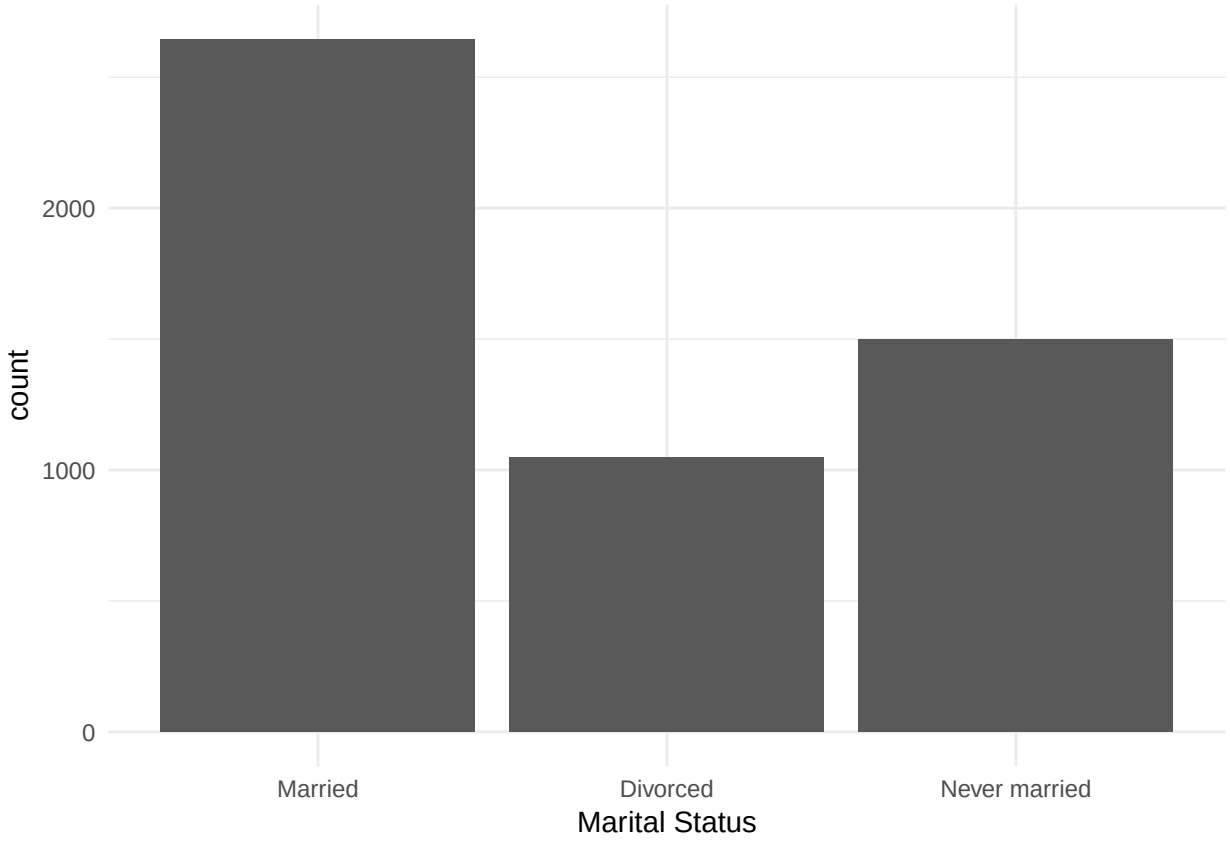
Marital Status

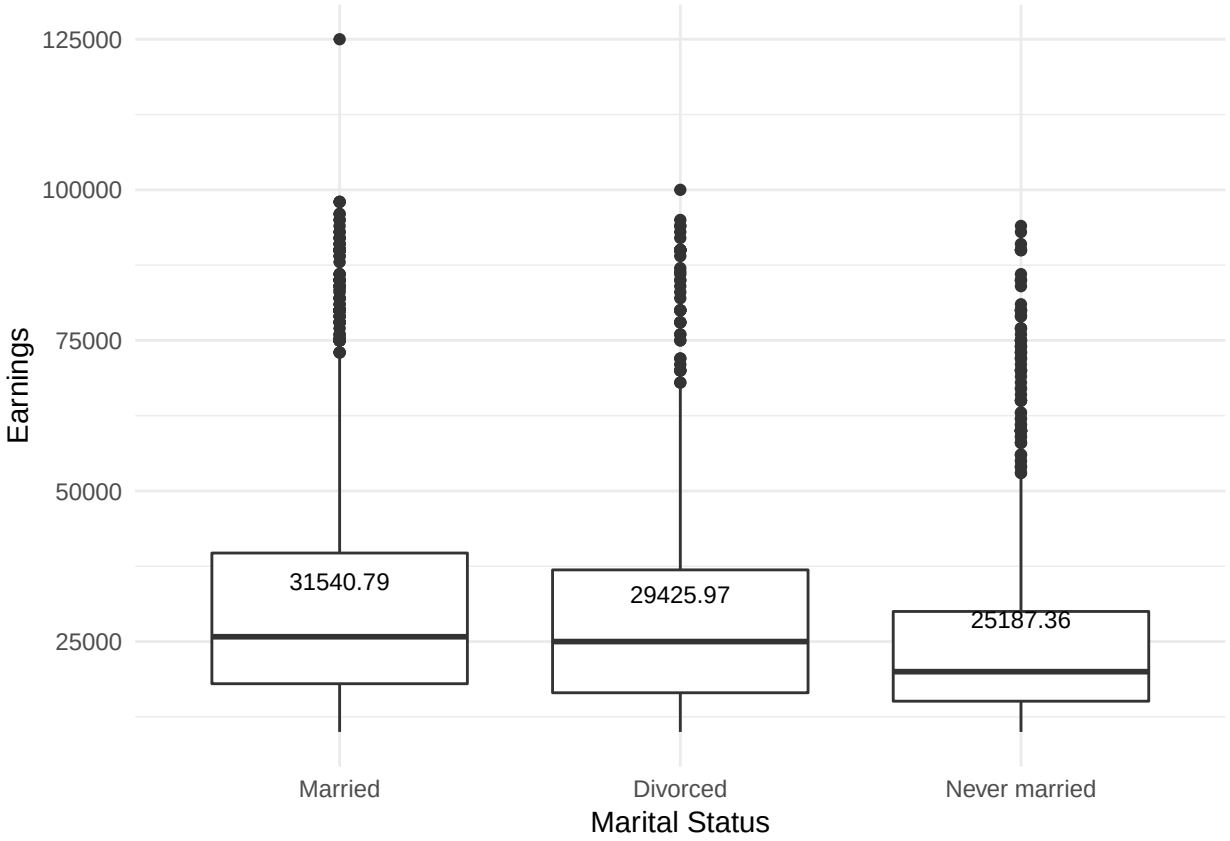




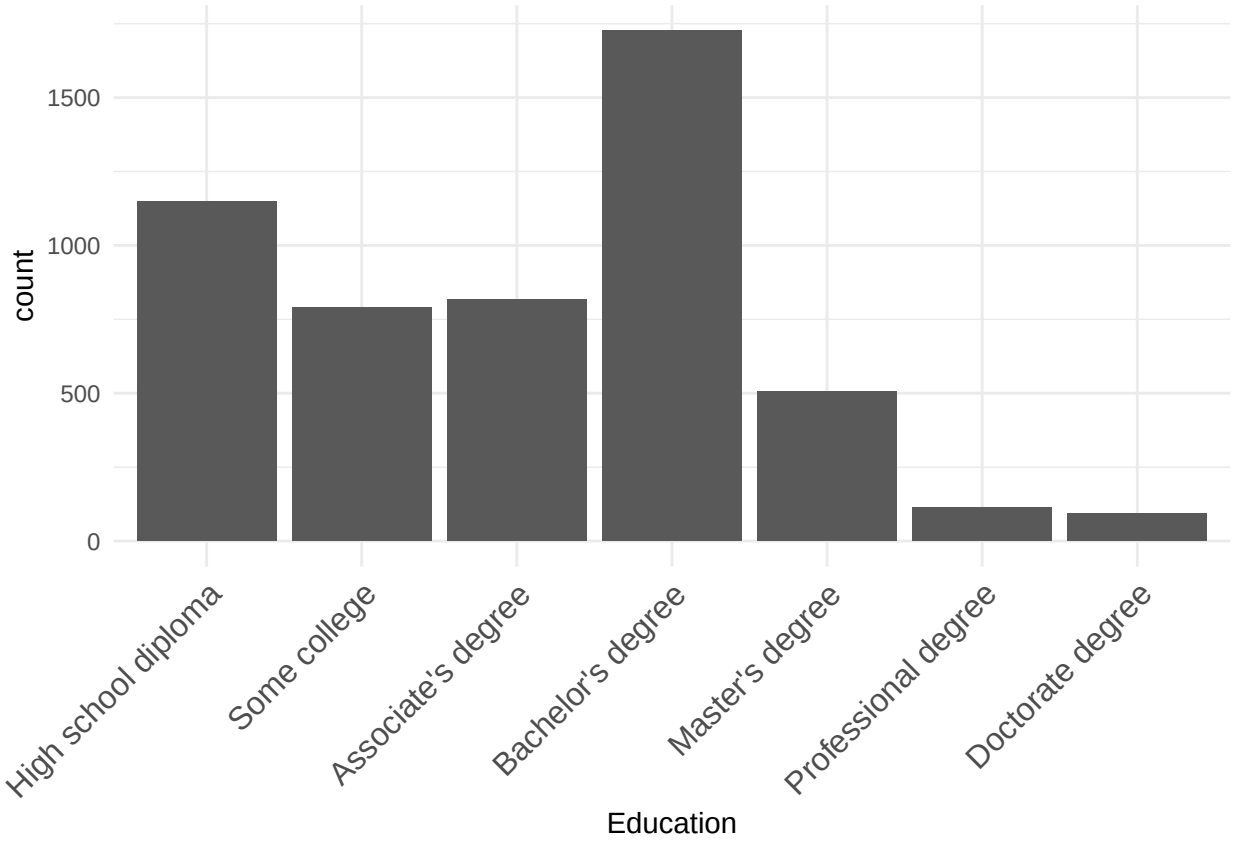

2000
1000
0
count
No
Yes
Marital Status

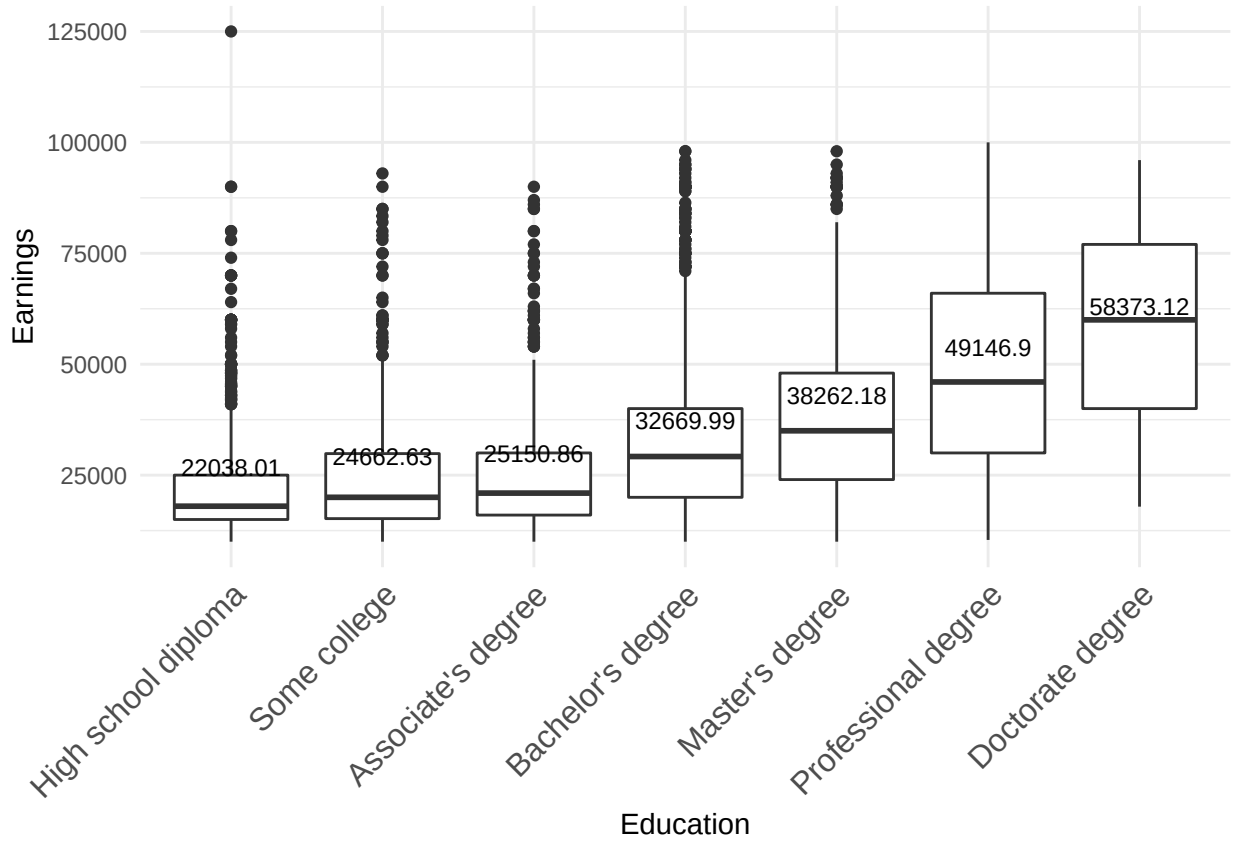

125000
100000
75000
50000
25000
Earnings
27036.95
31550.27
No
Yes
Marital Status


2000
1000
0
count
Married
Divorced
Never married
Marital Status


125000
100000
75000
50000
25000
Earnings
31540.79
29425.97
25187.36
Married
Divorced
Never married
Marital Status

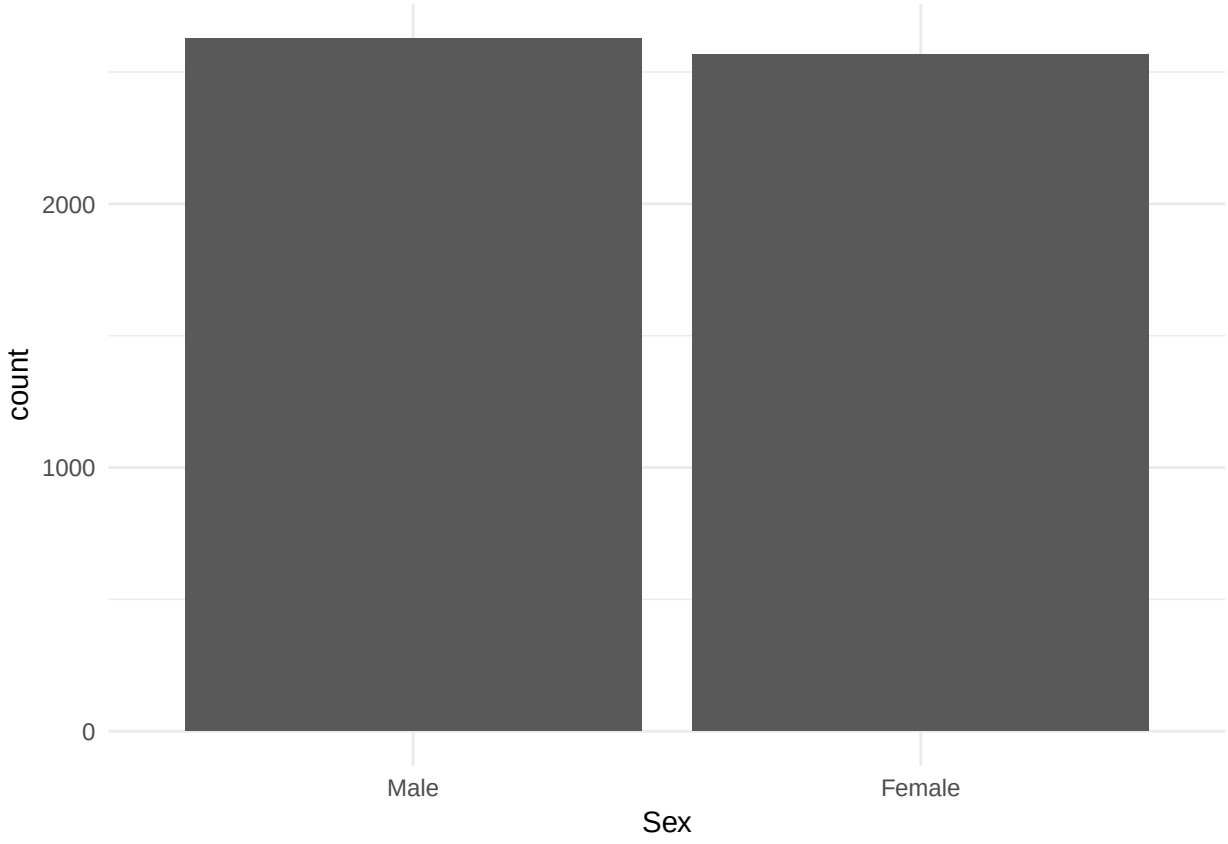
Educational attainment

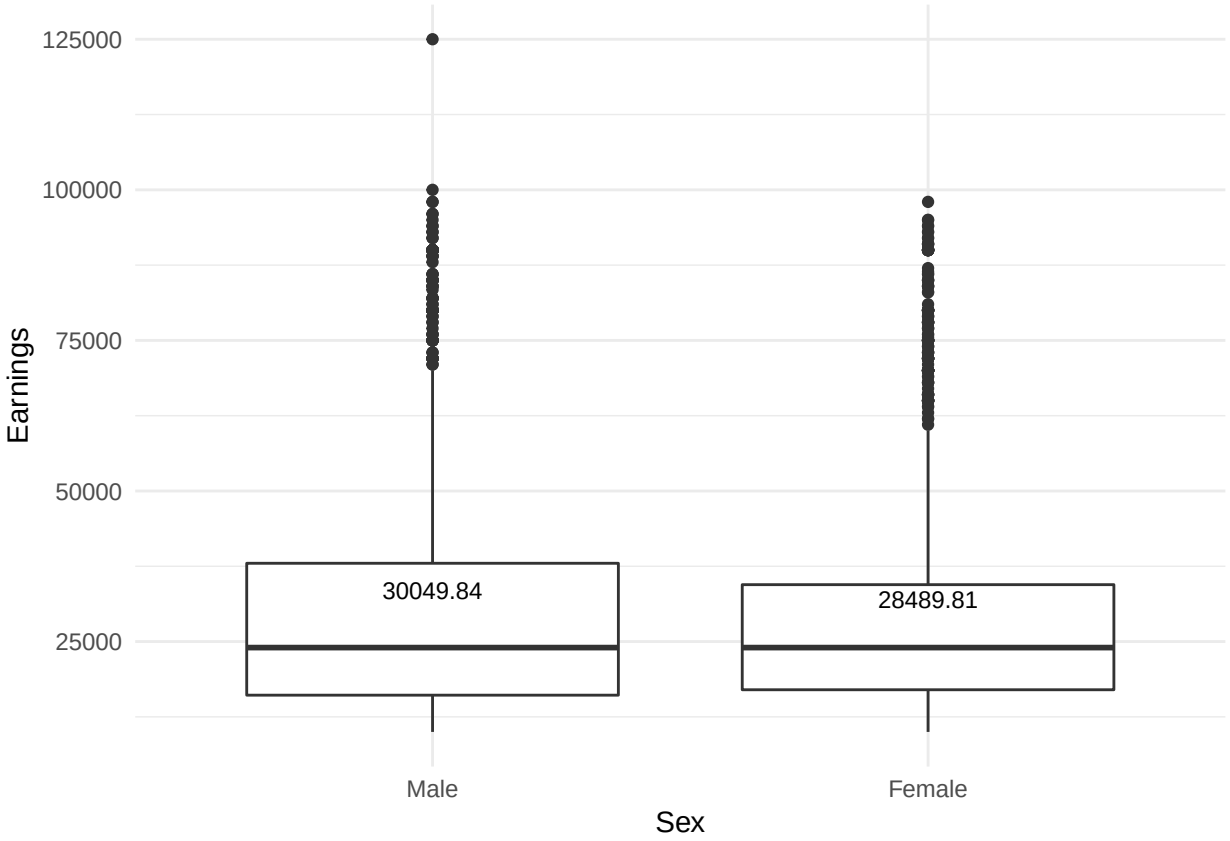




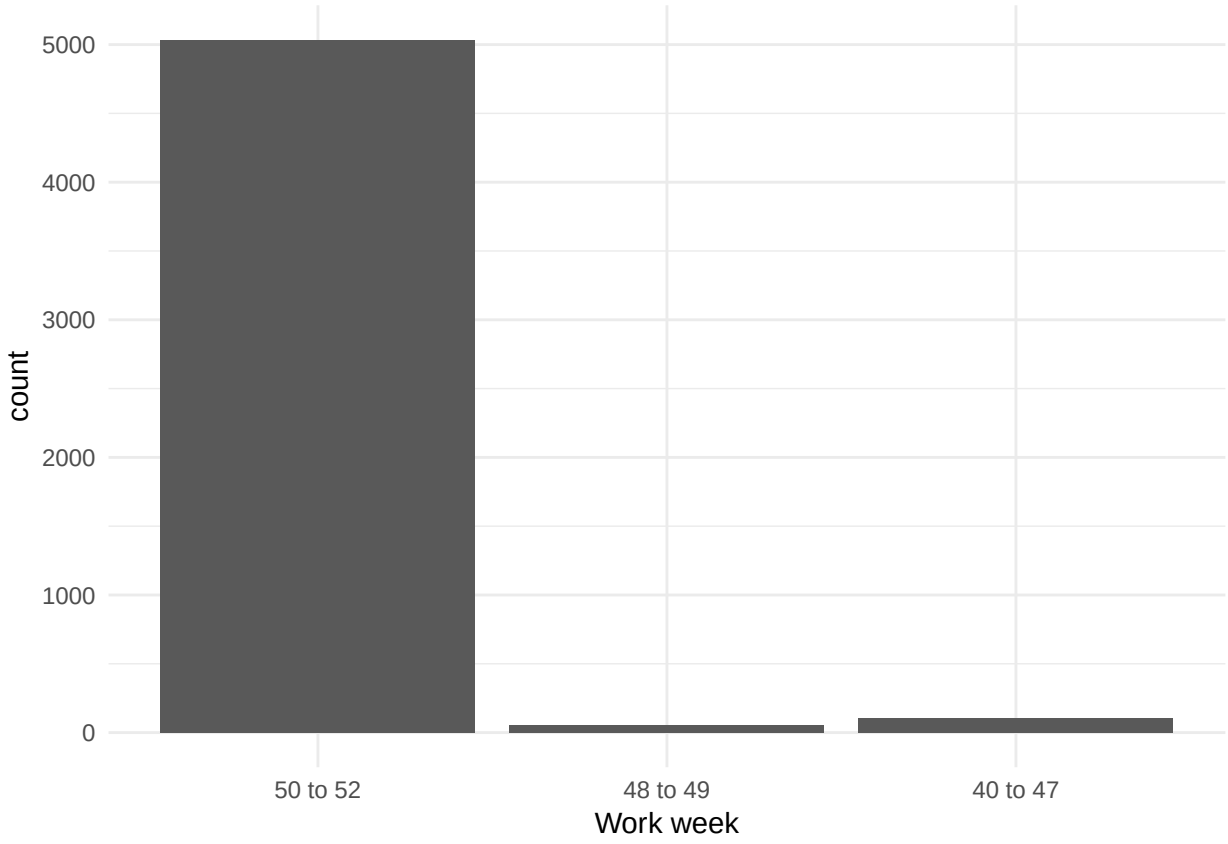
Sex

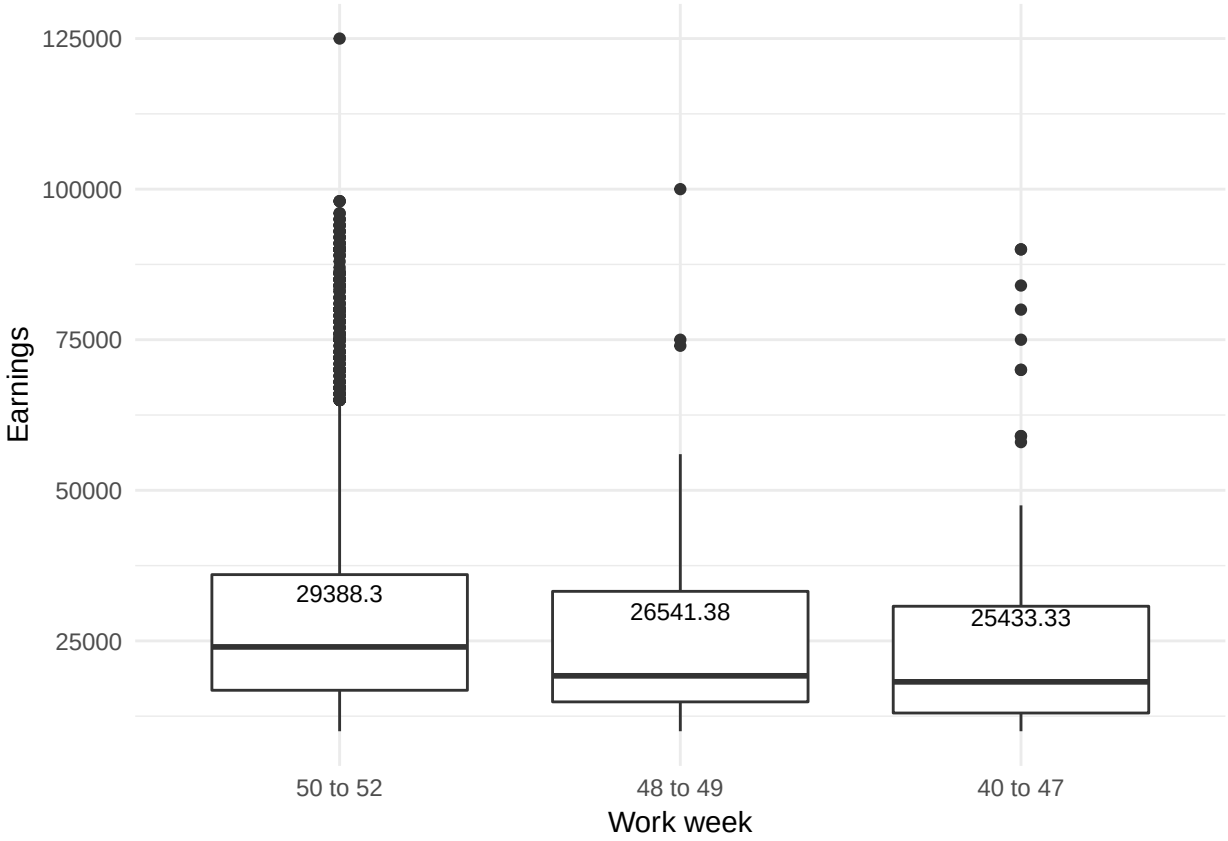
Gender is nearly equalized in Puerto Rico




125000
100000
75000
50000
25000
Earnings
30049.84
28489.81
Male
Female
Sex

Work week

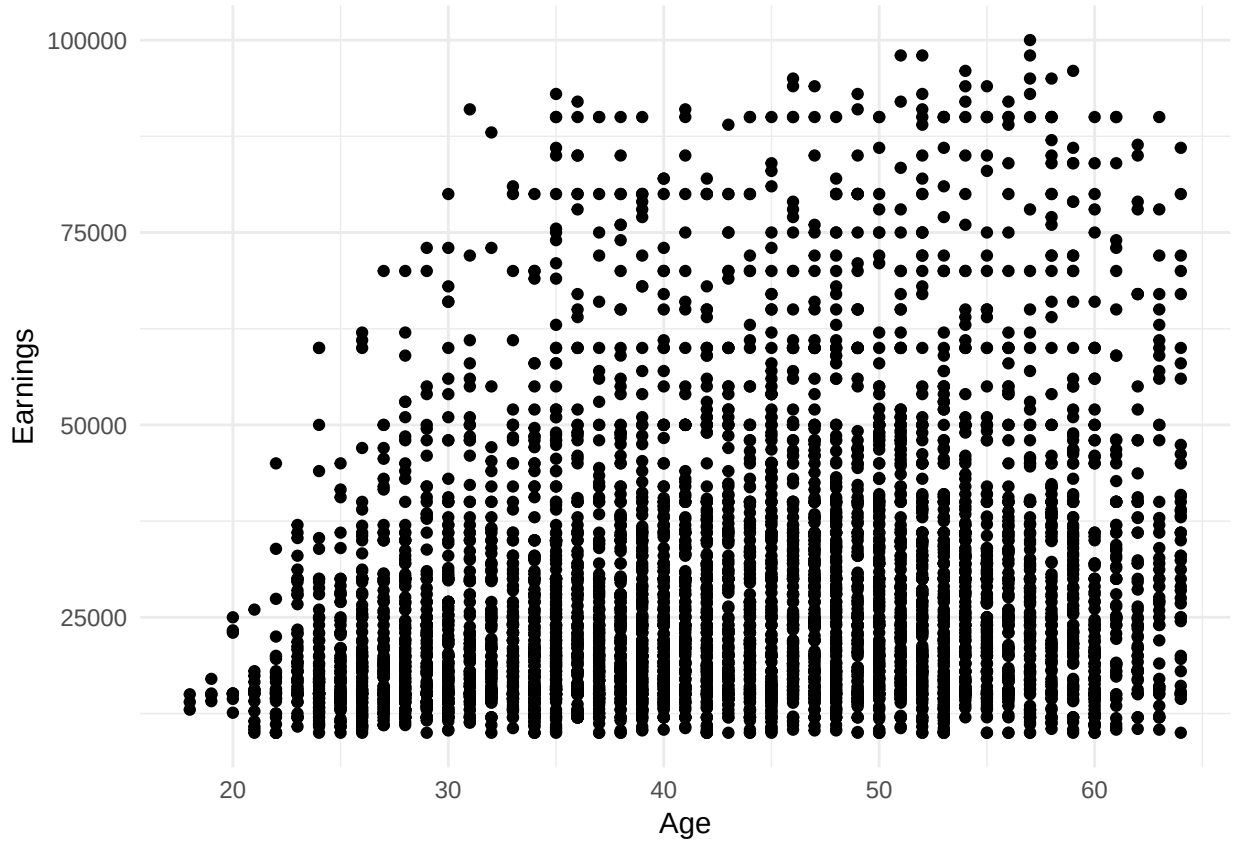

5000
4000
3000
2000
1000
0
count
50 to 52
48 to 49
40 to 47
Work week

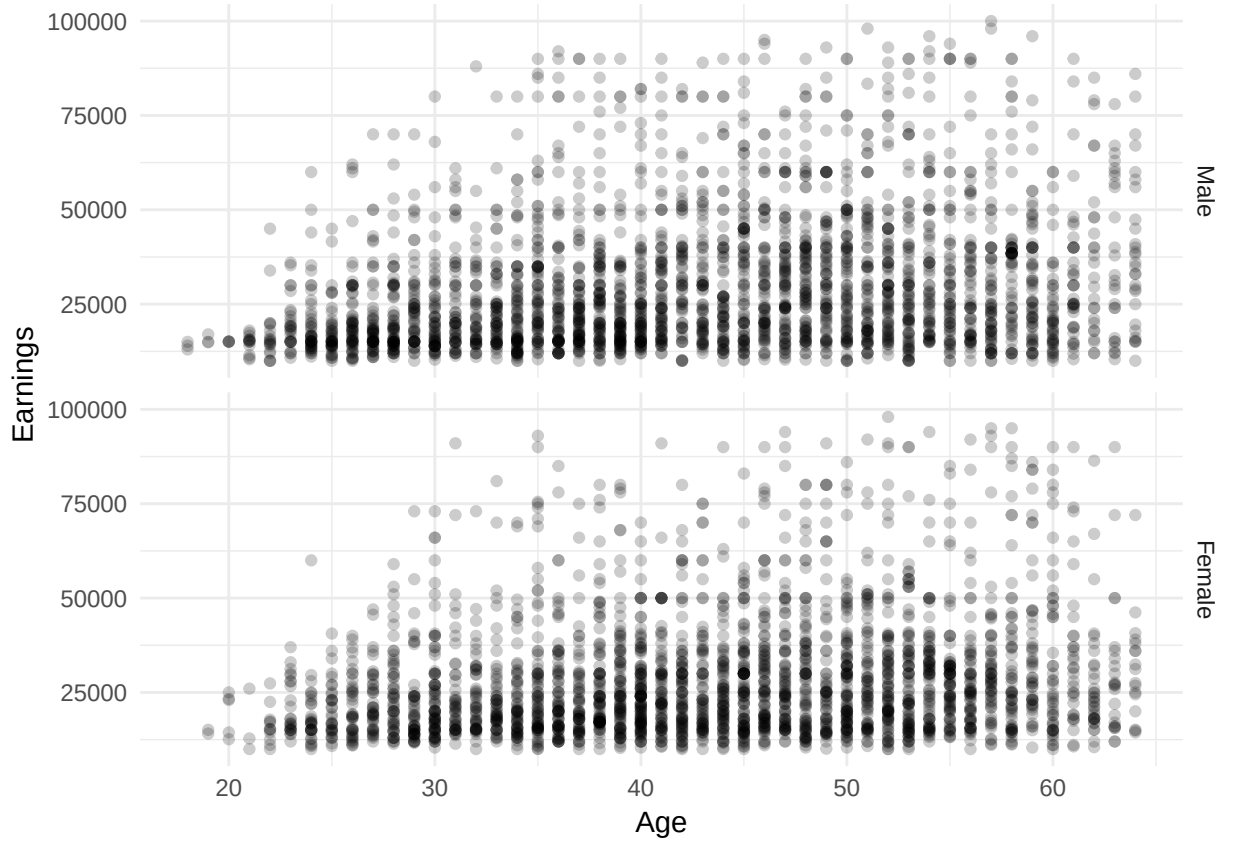

125000
100000
75000
50000
25000
Earnings
29388.3
26541.38
25433.33
50 to 52
48 to 49
40 to 47
Work week

Task 5:

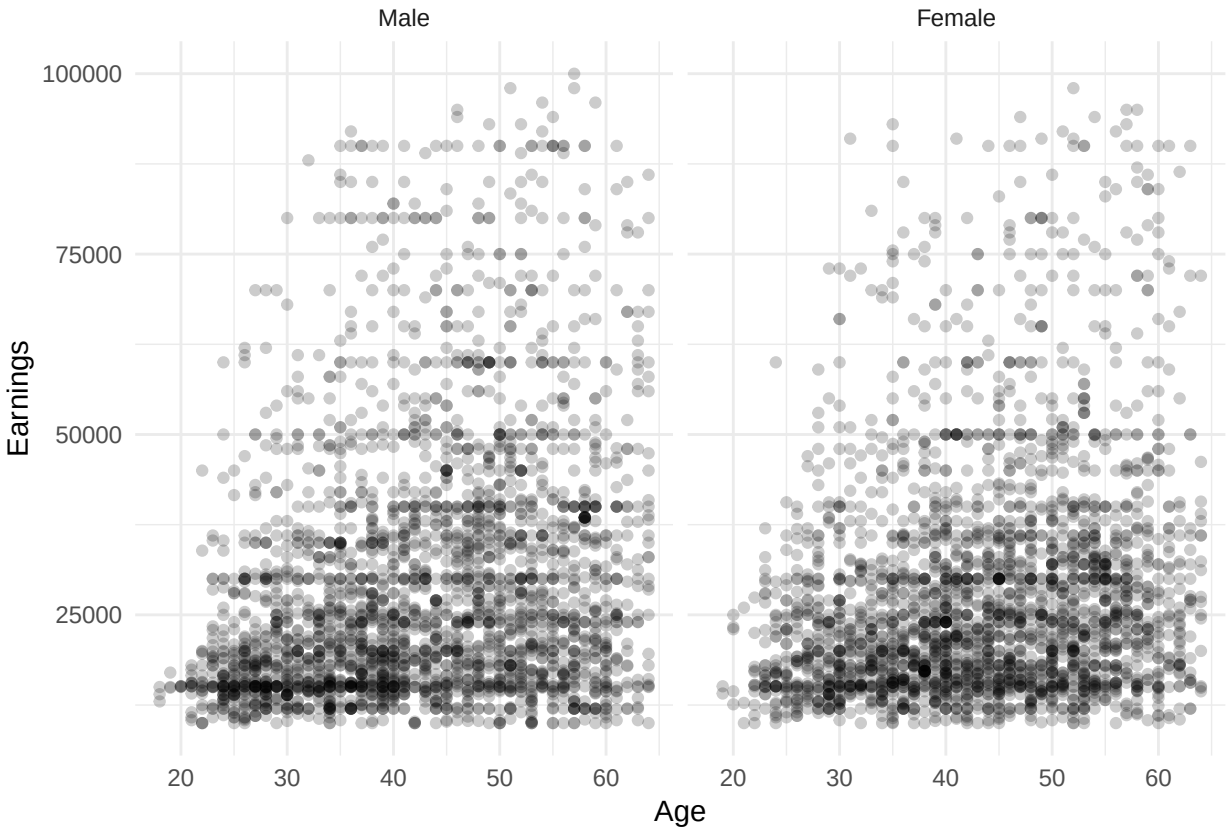
Generate and examine cross tabulations, scatterplots, and/or correlation coefficients of interest

Age vs. Total person's earnings



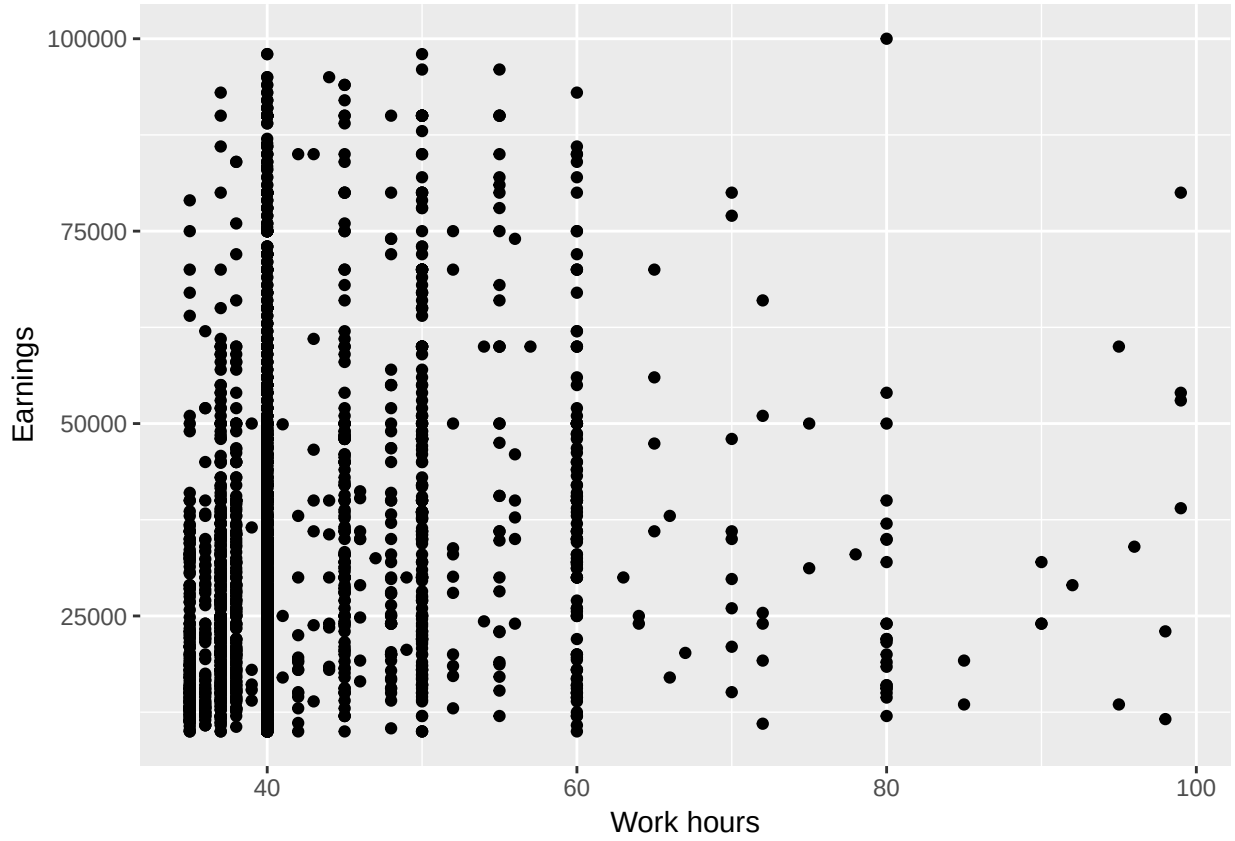



High school diploma
Some college
Associate's degree
Bachelor's degree
Master's degree
Professional degree
Doctorate degree
Earnings
100000
75000
50000
25000
20
30
40
50
60
Age



The correlation of 0.22 for age and earnings indicates a very weak relationship. Age is neither a primary reason for differences in earnings, nor a clear proxy for some other variable.

Work hours vs. Total person's earnings



The correlation of 0.18 for earnings and work hours is also very weak. No doubt it would be strong if the data were not filtered to those working more than 35 hours per week. Interestingly, earnings appear to drop for those working more than 60 hours per week.

Race: White

	RACWHT: No (N = 1,401)	RACWHT: Yes (N = 3,792)
Minimum	10000	10000
Maximum	96000	100000
Median	23000	24600
Mean	26946.56	30115.28

Race: Black

	RACBLK: No (N = 4,417)	RACBLK: Yes (N = 776)
Minimum	10000	10000
Maximum	98000	100000
Median	24000	23000
Mean	29550.70	27608.03

Race: Other

	RACOTHER: No (N = 4,380)	RACOTHER: Yes (N = 813)
Minimum	10000	10000
Maximum	100000	96000
Median	24000	23600
Mean	29638.67	27222.51

Marital Status

	MAR1: No (N = 2,614)	MAR1: Yes (N = 2,579)
Minimum	10000	10000
Maximum	100000	98000
Median	22000	25600
Mean	27036.95	31514.04

Educational attainment

	SCHL: High school diploma (N = 1,147)	SCHL: Some college (N = 791)	SCHL: Associate's degree (N =
Minimum	10000	10000	10000
Maximum	90000	93000	90000
Median	18000	20000	20950
Mean	21948.24	24662.63	25150.86

Preliminary Econometric Estimates

First Model:

$Earning = \beta_0 + Divorced * \beta_1 + NeverMarried * \beta_2 + Female * \beta_3 + RaceBlack * \beta_4 + RaceOther * \beta_5 + SomeCollege * \beta_6 + Associat$

```
##
## Call:
## lm(formula = PERNP ~ Divorced + NeverMarried + Female + RaceBlack +
##     RaceOther + SomeCollege + Associate + Bachelor + Master +
##     Professional + Doctoral + AGEP, data = ss16ppr)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -43686  -9198  -2921   5348  63456
##
## Coefficients:
##              Estimate Std. Error t value             Pr(>|t|)
## (Intercept)  12519.82    1104.43  11.336 < 0.0000000000000002 ***
## Divorced     -1146.91     546.10  -2.100             0.035760 *
## NeverMarried -2959.79     517.30  -5.722       0.000000011144 ***
## Female       -4824.55     429.48 -11.234 < 0.0000000000000002 ***
## RaceBlack    -1301.52     589.11  -2.209             0.027196 *
## RaceOther    -2132.00     577.36  -3.693             0.000224 ***
## SomeCollege   4311.57     691.75   6.233       0.000000000494 ***
## Associate     4232.24     685.40   6.175       0.000000000713 ***
```

```
## Bachelor     12406.91     584.22  21.237 < 0.0000000000000002 ***
## Master       17855.61     808.79  22.077 < 0.0000000000000002 ***
## Professional 28203.02    1469.19  19.196 < 0.0000000000000002 ***
## Doctoral     35732.87    1609.22  22.205 < 0.0000000000000002 ***
## AGEP           285.13      20.95  13.609 < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14860 on 5180 degrees of freedom
## Multiple R-squared:  0.2484, Adjusted R-squared:  0.2467
## F-statistic: 142.7 on 12 and 5180 DF,  p-value: < 0.00000000000000022
```

- Coefficients Explanation
  - Holding gender, race, education and age constant, married or widowed people makes $1146.91 more than people who divorced or separated on average.
  - Holding gender, race, education and age constant, married or widowed people makes $2959.79 more than people who never married on average.
  - Holding marriage, race, education and age constant, male makes $4824.55 more than female on average.
  - Holding marriage, gender, education and age constant, White makes $1301.52 more than Black on average.
  - Holding marriage, gender, education and age constant, White makes $2132 more than Other race on average.
  - Holding marriage, gender, race and age constant, people have high school education makes $4311.57 less than people have some college education on average.
  - Holding marriage, gender, race and age constant, people have high school education makes $4232.24 less than people have associate education on average.
  - Holding marriage, gender, race and age constant, people have high school education makes $12406.91 less than people have bachelor's degree on average.
  - Holding marriage, gender, race and age constant, people have high school education makes $17855.61 less than people have master's degree on average.
  - Holding marriage, gender, race and age constant, people have high school education makes $28203.02 less than people have Professional education on average.
  - Holding marriage, gender, race and age constant, people have high school education makes $35732.87 less than people have doctor's degree on average.
  - Holding marriage, gender, race and education constant, people make $285.13 more as age increases on average between the age of 18 to 64.

Second Model:

$$Earning = \beta_0 + Female * \beta_1 + SomeCollege * \beta_2 + Associate * \beta_3 + Bachelor * \beta_4 + Master * \beta_5 + Professional * \beta_6 + Doctoral * \beta_7$$

```
##
## Call:
## lm(formula = PERNP ~ Female + SomeCollege + Associate + Bachelor +
##     Master + Professional + Doctoral, data = ss16ppr)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -38660  -9524  -3508   5802  67102
##
## Coefficients:
```

```
##              Estimate Std. Error t value             Pr(>|t|)
## (Intercept)   23324.4      471.0  49.521 < 0.0000000000000002 ***
## Female        -4656.1      440.7 -10.566 < 0.0000000000000002 ***
## SomeCollege    3339.6      711.0   4.697          0.000002704 ***
## Associate      4000.8      705.6   5.670          0.000000015 ***
## Bachelor      12229.4      601.2  20.343 < 0.0000000000000002 ***
## Master        17934.3      832.9  21.532 < 0.0000000000000002 ***
## Professional  28336.0     1515.3  18.700 < 0.0000000000000002 ***
## Doctoral      37602.1     1656.5  22.699 < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15330 on 5185 degrees of freedom
## Multiple R-squared:  0.1989, Adjusted R-squared:  0.1979
## F-statistic:   184 on 7 and 5185 DF,  p-value: < 0.00000000000000022
```

- Coefficients Explanation
  - Holding education constant, male makes $4656 more than female on average.
  - Holding gender constant, people have high school education makes $4311.57 less than people have some college education on average.
  - Holding gender constant, people have high school education makes $4232.24 less than people have associate education on average.
  - Holding gender constant, people have high school education makes $12406.91 less than people have bachelor's degree on average.
  - Holding gender constant, people have high school education makes $17855.61 less than people have master's degree on average.
  - Holding gender constant, people have high school education makes $28203.02 less than people have Professional education on average.
  - Holding gender constant, people have high school education makes $35732.87 less than people have doctor's degree on average.