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 earnings 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 30224. 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 earnings differences by ability to speak English 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), work weeks ($>$ 40 weeks) and work hours ($>$35 hours/week).

An earnings histogram identified a default maximum amount of earnings (189k) which we also filtered out of the data. The wage 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 wage 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.

ID	WAGP	SEX	AGEP	RAC1P	RACWHT	RACBLK	HISP	MAR
10	34000	1	47	White	Yes	No	03	Married
20	13000	1	58	Black or African American	No	Yes	03	Never married
43	18000	1	50	White	Yes	No	03	Married
44	10300	2	39	White	Yes	No	03	Married
49	28600	2	39	Black or African American	No	Yes	03	Married
50	24800	1	37	Black or African American	No	Yes	05	Married
53	22000	2	47	Some Other	No	No	03	Never married
58	19000	2	60	Black or African American	No	Yes	03	Never married
61	87000	2	58	White	Yes	No	03	Divorced
62	8000	1	33	White	Yes	No	03	Never married
64	22900	1	61	White	Yes	No	03	Divorced
70	19000	1	39	Black or African American	No	Yes	03	Married
71	19600	2	36	Black or African American	No	Yes	03	Married
74	48000	1	30	Two or More Races	Yes	Yes	03	Divorced
77	40000	2	30	White	Yes	No	03	Never married
78	15600	2	41	White	Yes	No	03	Never married
82	12100	1	46	White	Yes	No	03	Divorced
85	14000	1	53	White	Yes	No	03	Married
94	80000	1	38	White	Yes	No	03	Never married
108	15100	2	26	Some Other	No	No	03	Never married

ID	SCHL	ENG	COW	WKW	WKHP
10	Associate's degree	Very well	1	50 to 52	40
20	High school diploma	Well	1	50 to 52	40
43	Master's degree	Well	3	50 to 52	40
44	Bachelor's degree	Well	1	50 to 52	40
49	Bachelor's degree	NA	1	50 to 52	45
50	Bachelor's degree	Very well	1	50 to 52	46
53	Associate's degree	Not well	3	50 to 52	40
58	High school diploma	Very well	1	50 to 52	40
61	Associate's degree	NA	5	50 to 52	40
62	Bachelor's degree	NA	1	50 to 52	35
64	High school diploma	Not well	5	50 to 52	40
70	Associate's degree	Not well	1	50 to 52	38
71	Bachelor's degree	Not well	5	50 to 52	40
74	Some college	NA	5	50 to 52	40
77	Some college	NA	1	50 to 52	40
78	High school diploma	Not well	1	50 to 52	40
82	High school diploma	Not well	1	50 to 52	40
85	High school diploma	Well	1	50 to 52	40
94	Bachelor's degree	NA	1	50 to 52	40
108	High school diploma	Not well	1	50 to 52	40

Task 2:

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

Wage

	ss16ppr (N = 5,057)
Minimum	150.00
Maximum	98000.00
Median	24000.00
Mean	28609.17

Age

	ss16ppr (N = 5,057)
Minimum	18.00
Maximum	64.00
Median	43.00
Mean	28609.17

Hours worked

	ss16ppr (N = 5,057)
Minimum	35.00000
Maximum	99.00000
Median	40.00000
Mean	40.93277

Ethicity: Hispanic

RACHISP	Count
No	4962
Yes	95

Race: White

In Puerto Rico, the majority of people identify themselves as white, though this should be interpreted as skin color (i.e. not Caucasian), rather than ethnicity, as the terminology is used in the mainland US as in the mainland, all Puerto Ricans would be identified as Latin regardless of skin color. Minority races including American Indian, Alaska Native, Asian, Native Hawaiian and Other Pacific Islander can be eliminated.

RACWHT	Count
No	1376
Yes	3681

Race: Black

RACBLK	Count
No	4297
Yes	760

Race: Other

RACOTHER	Count
No	4264
Yes	793

Marital status

MAR	Count
Married	2458
Widowed	66
Divorced	917
Separated	100
Never married	1516

MAR1	Count
No	2599
Yes	2458

Ability to speak English

ENG	Count
NA	1572
Very well	1025
Well	1213
Not well	1247

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	1116
Some college	783
Associate's degree	789
Bachelor's degree	1701
Master's degree	488
Professional degree	94
Doctorate degree	86

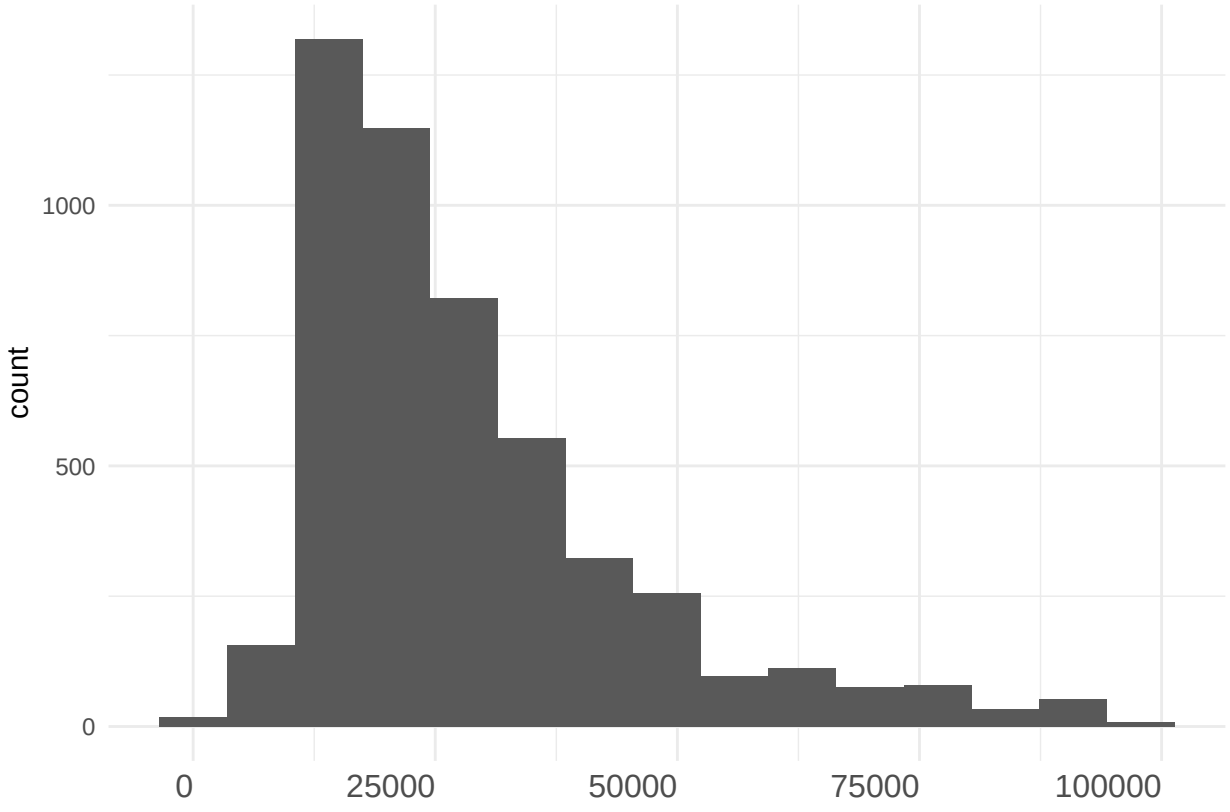
Sex

SEX	Count
Male	2504
Female	2553

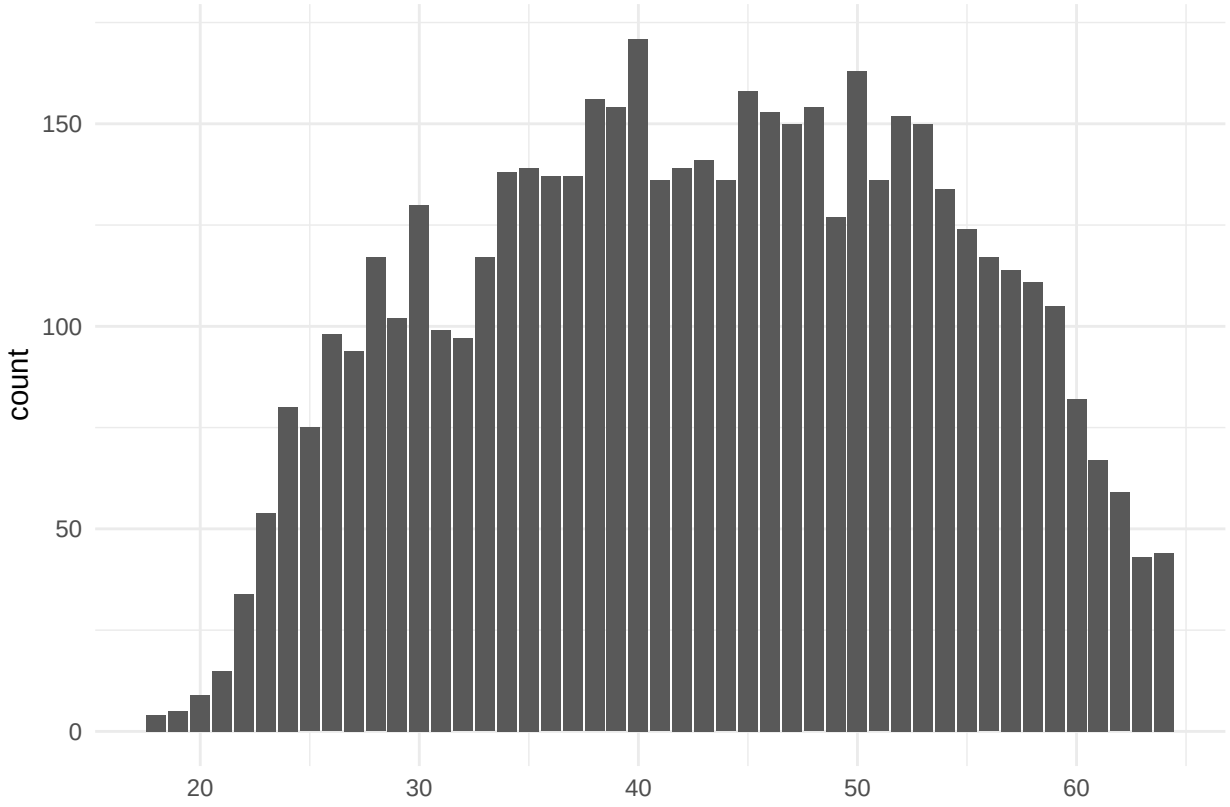
Task 3:

Generate and examine histograms for quantitative variables of interest

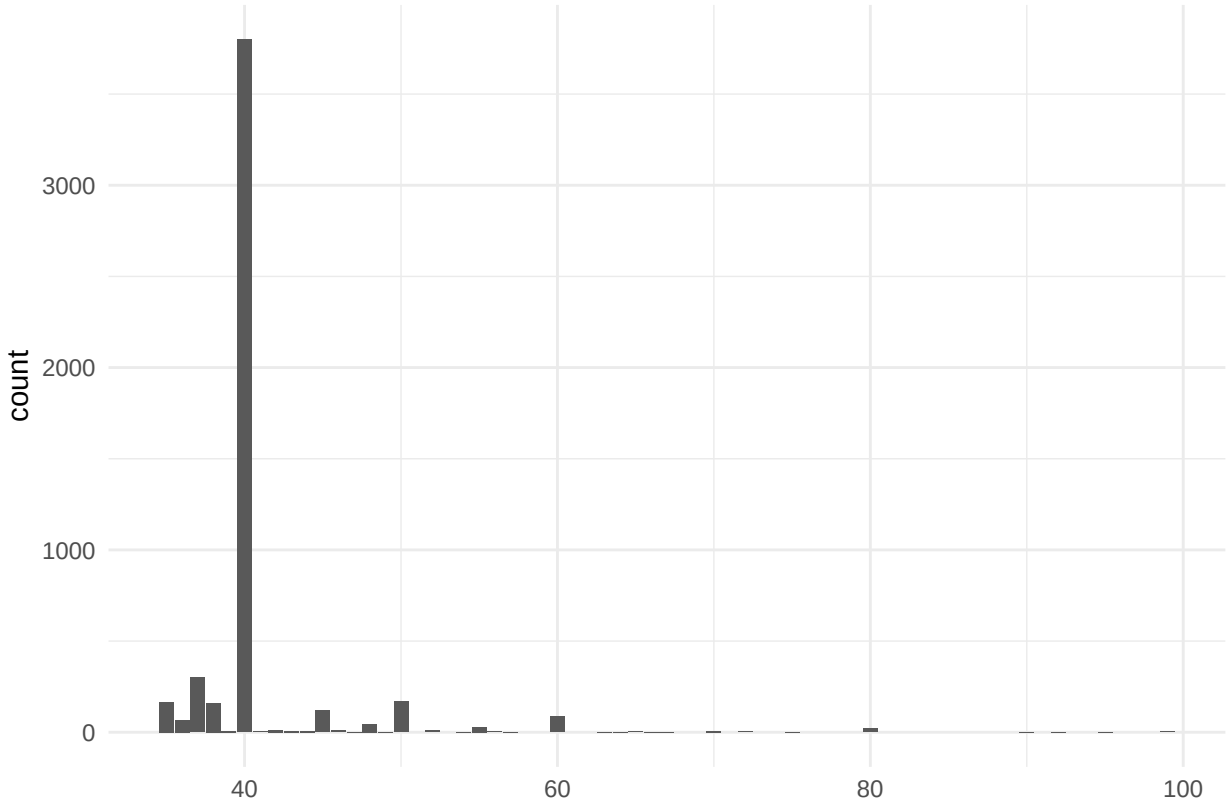
Wage



Age



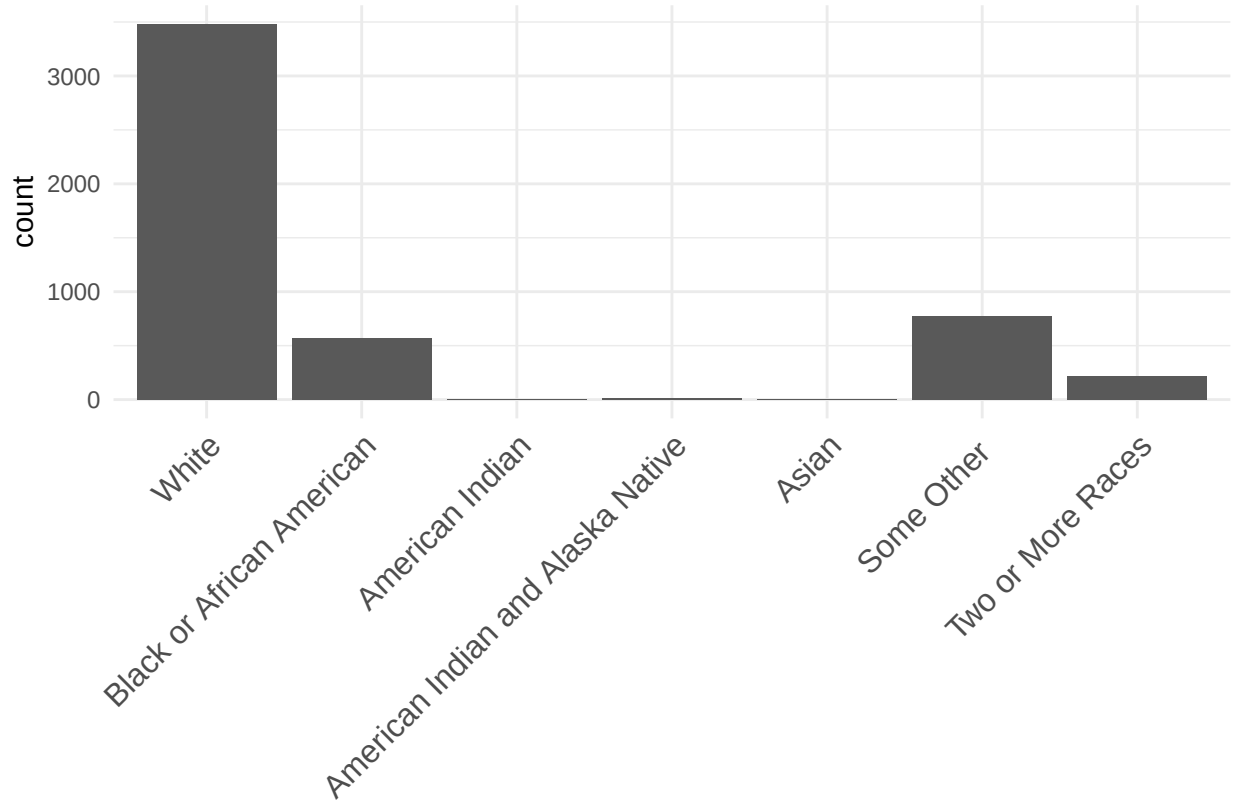
Hours worked

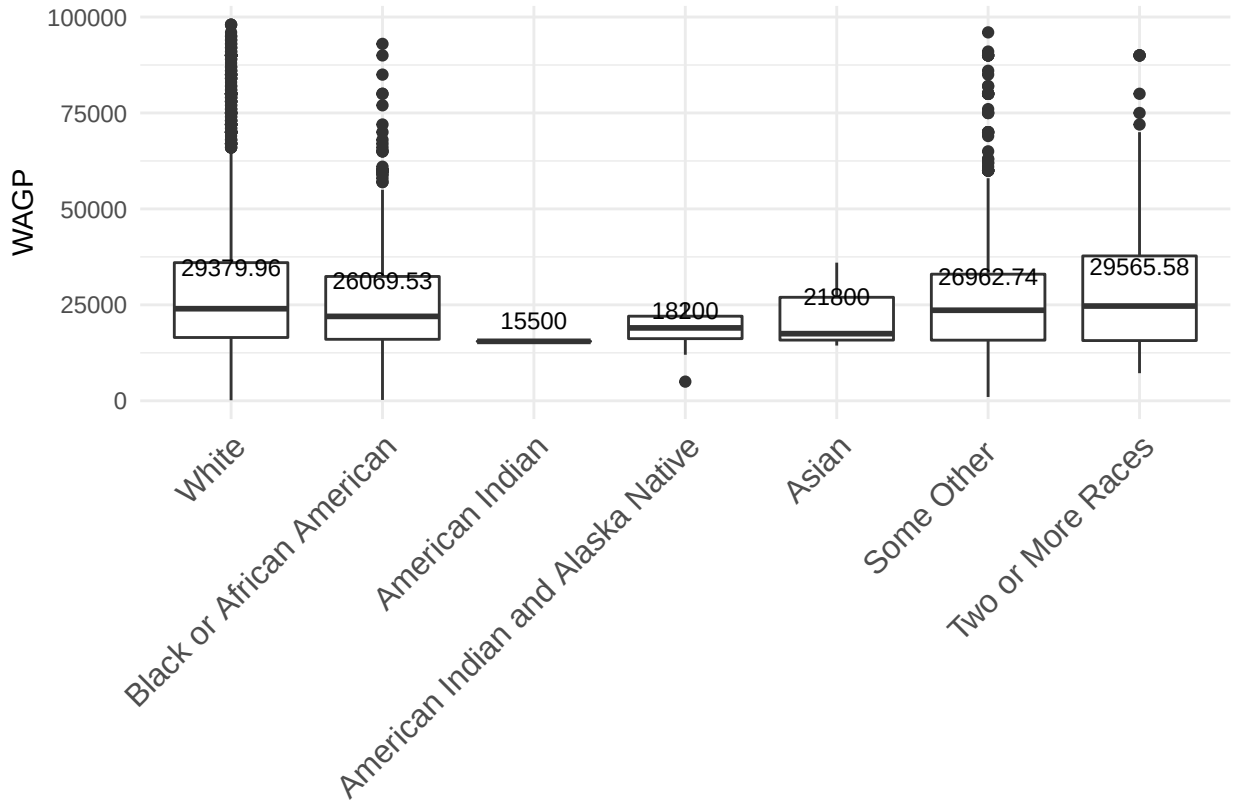


Task 4:

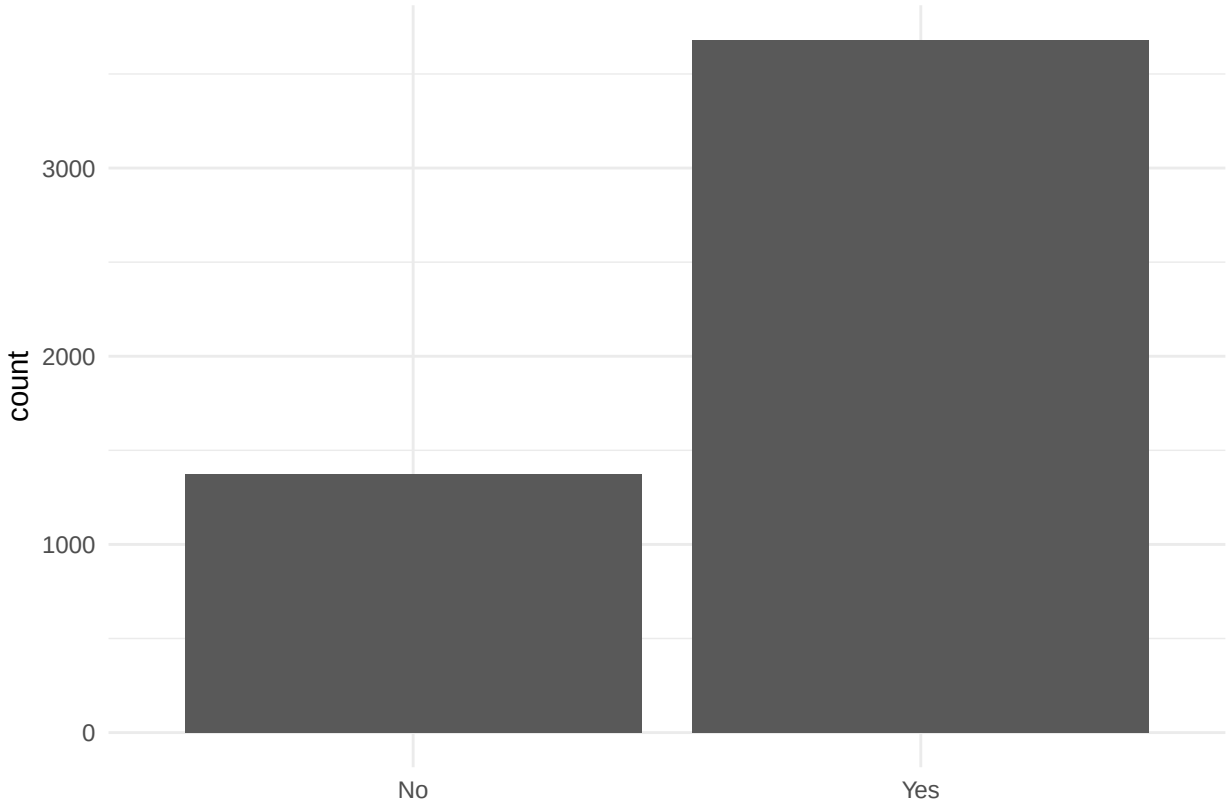
Generate and examine bar charts/graphs for qualitative variables of interest

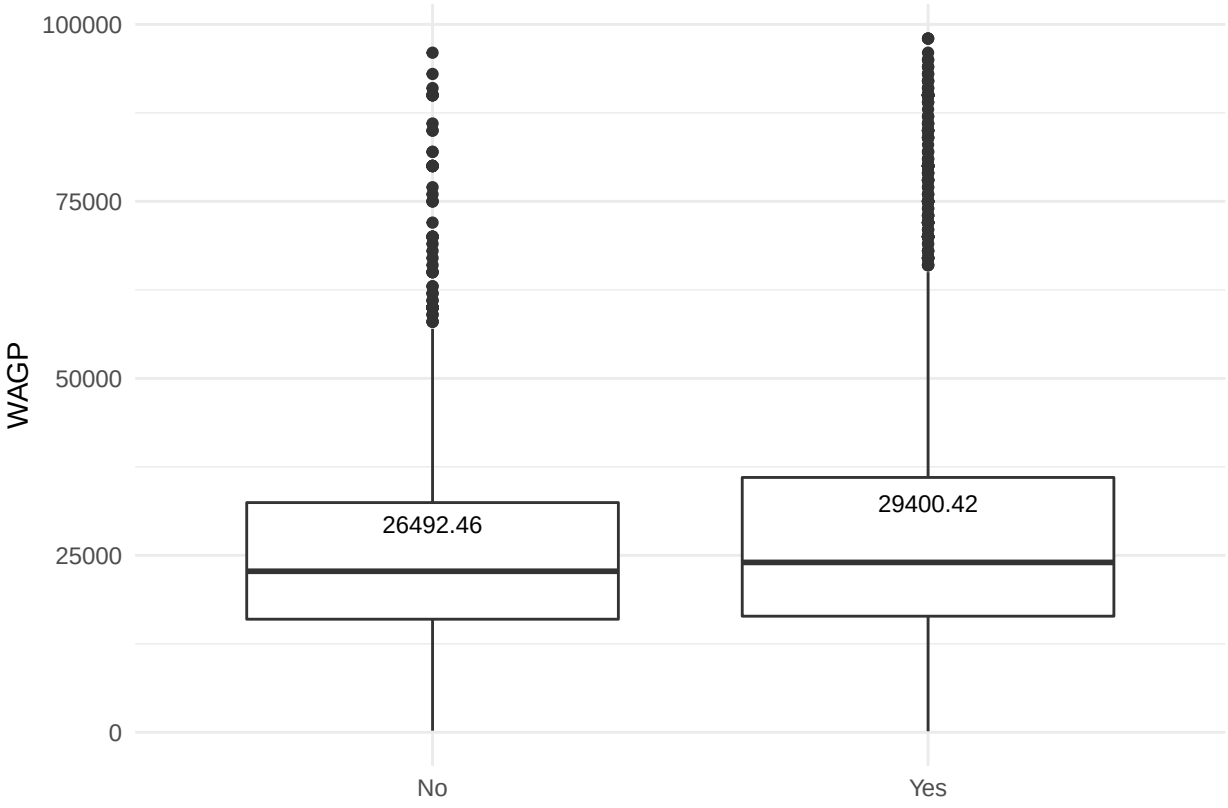
Race


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

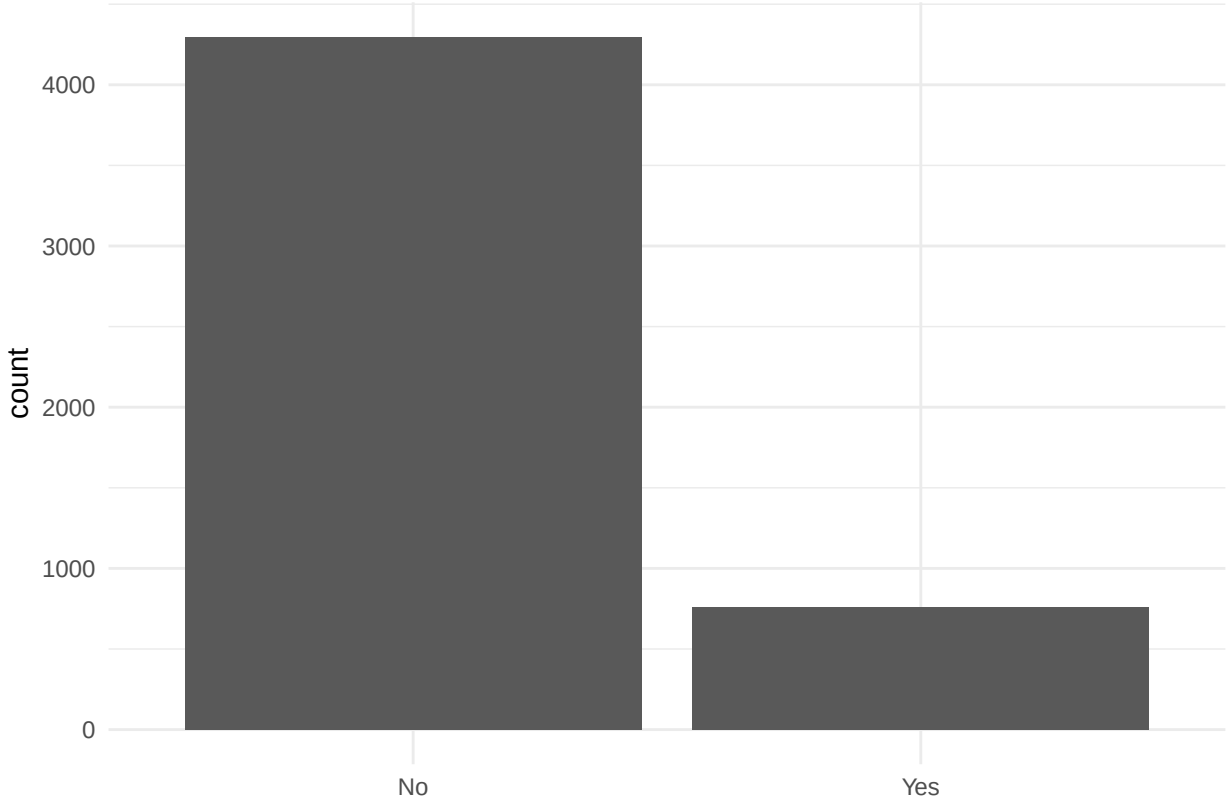

WAGP
100000
75000
50000
25000
0
29379.96
26069.53
15500
18200
21800
26962.74
29565.58
White
Black or African American
American Indian
American Indian and Alaska Native
Asian
Some Other
Two or More Races

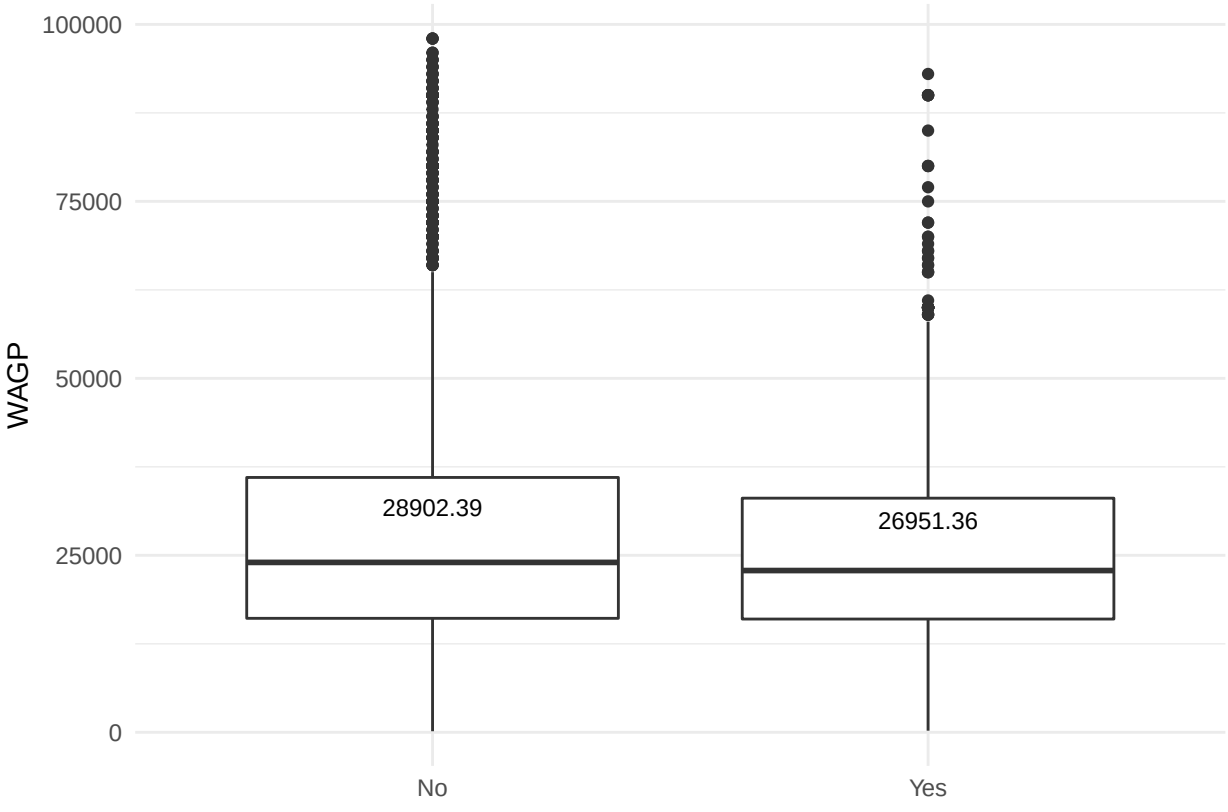
Race: White


count
3000
2000
1000
0
No
Yes

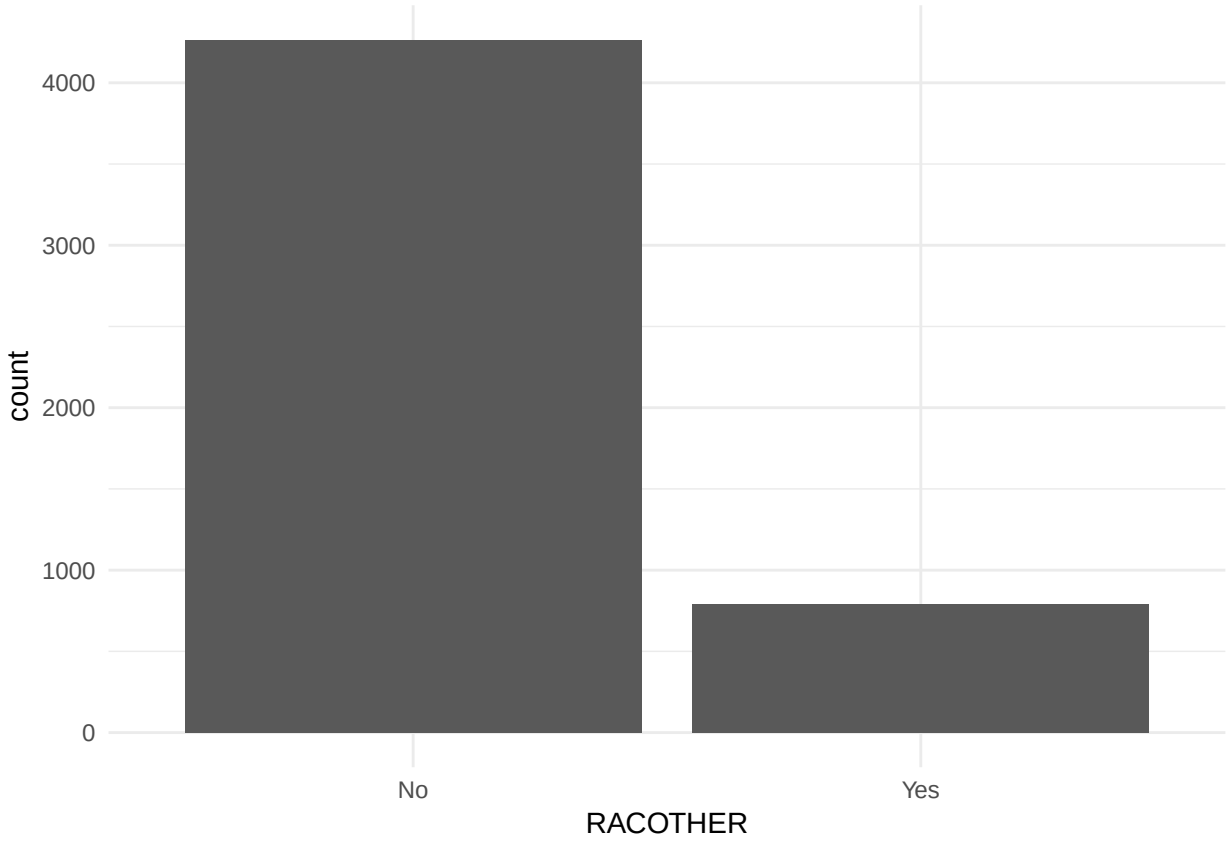

WAGP
100000
75000
50000
25000
0
26492.46
29400.42
No
Yes

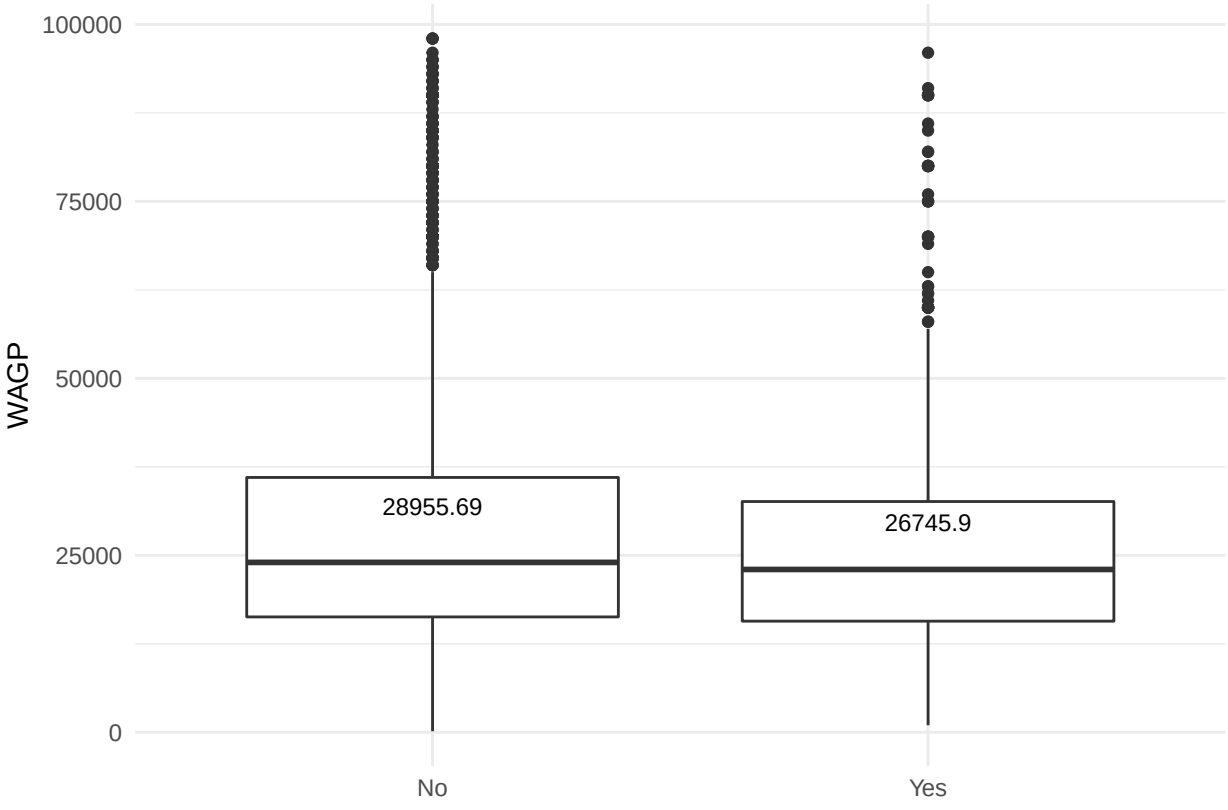
Race: Black


4000
3000
2000
1000
0
count
No
Yes

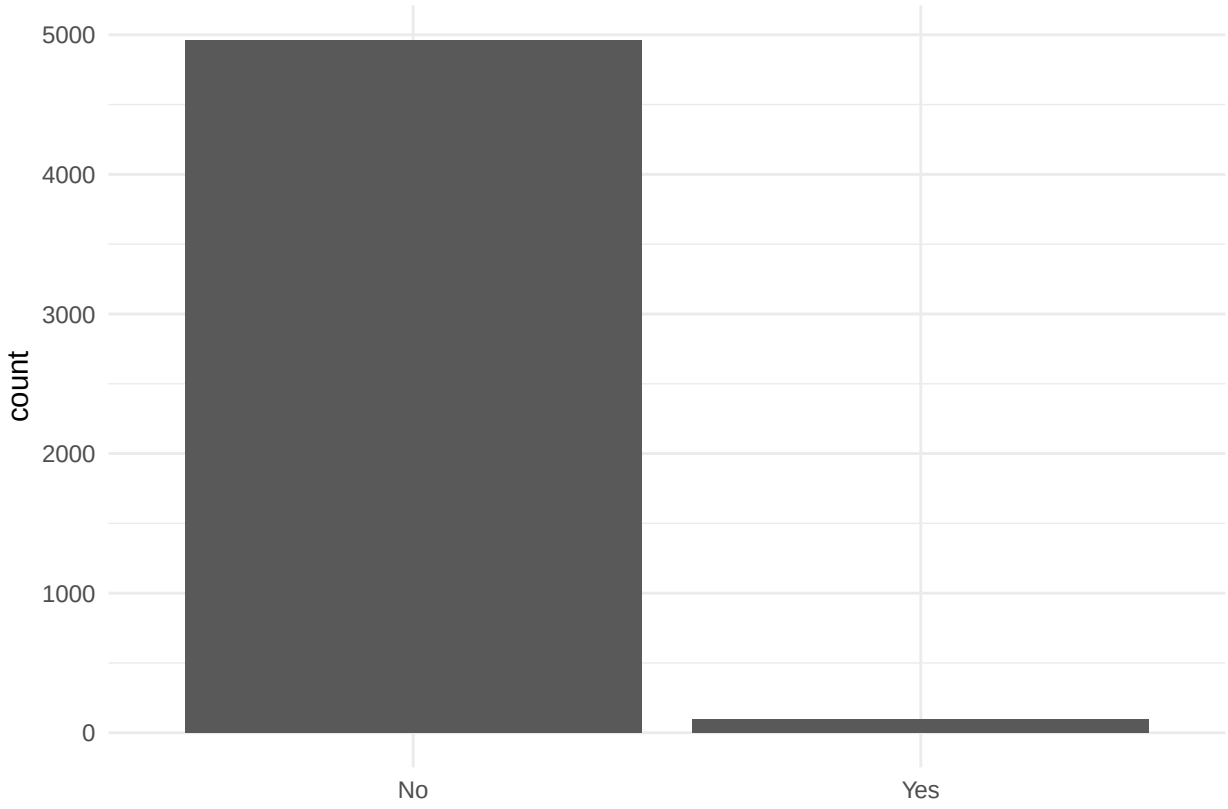

100000
75000
50000
25000
0
WAGP
28902.39
26951.36
No
Yes

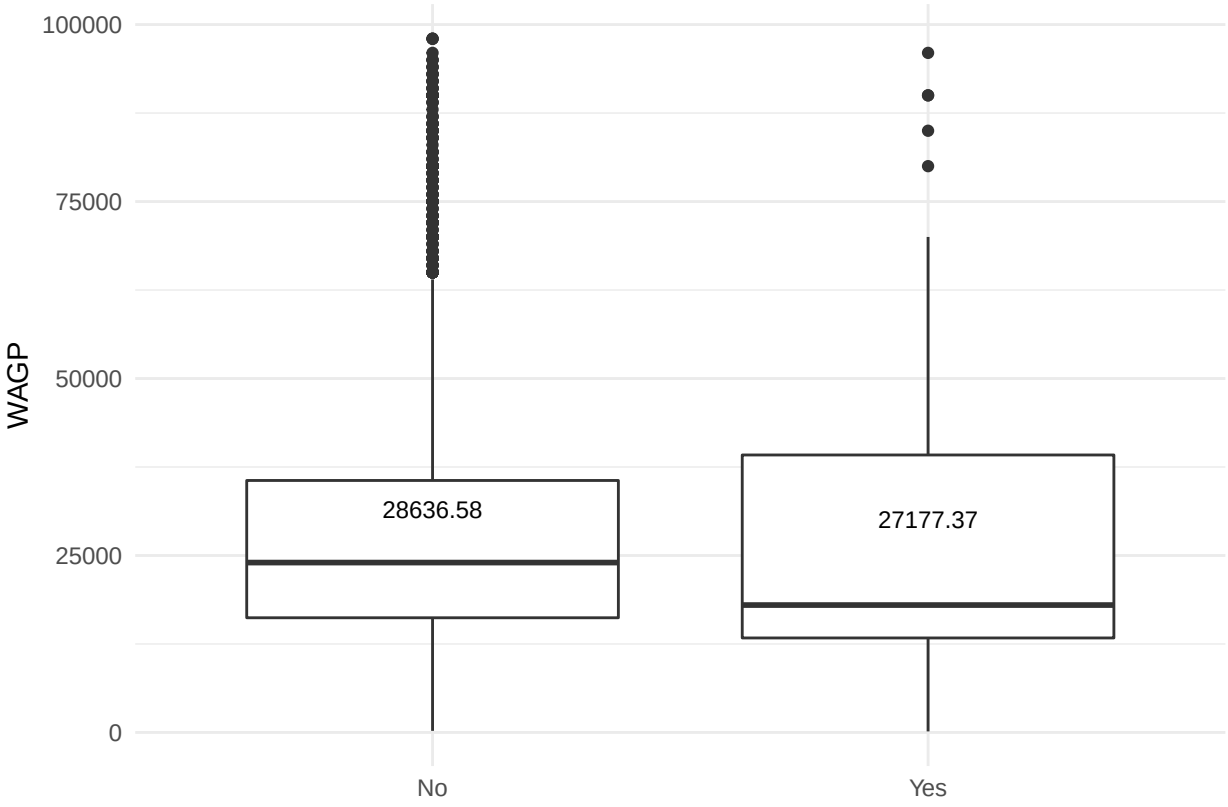
Race: Other


4000
3000
2000
1000
0
count
No
Yes
RACOTHER

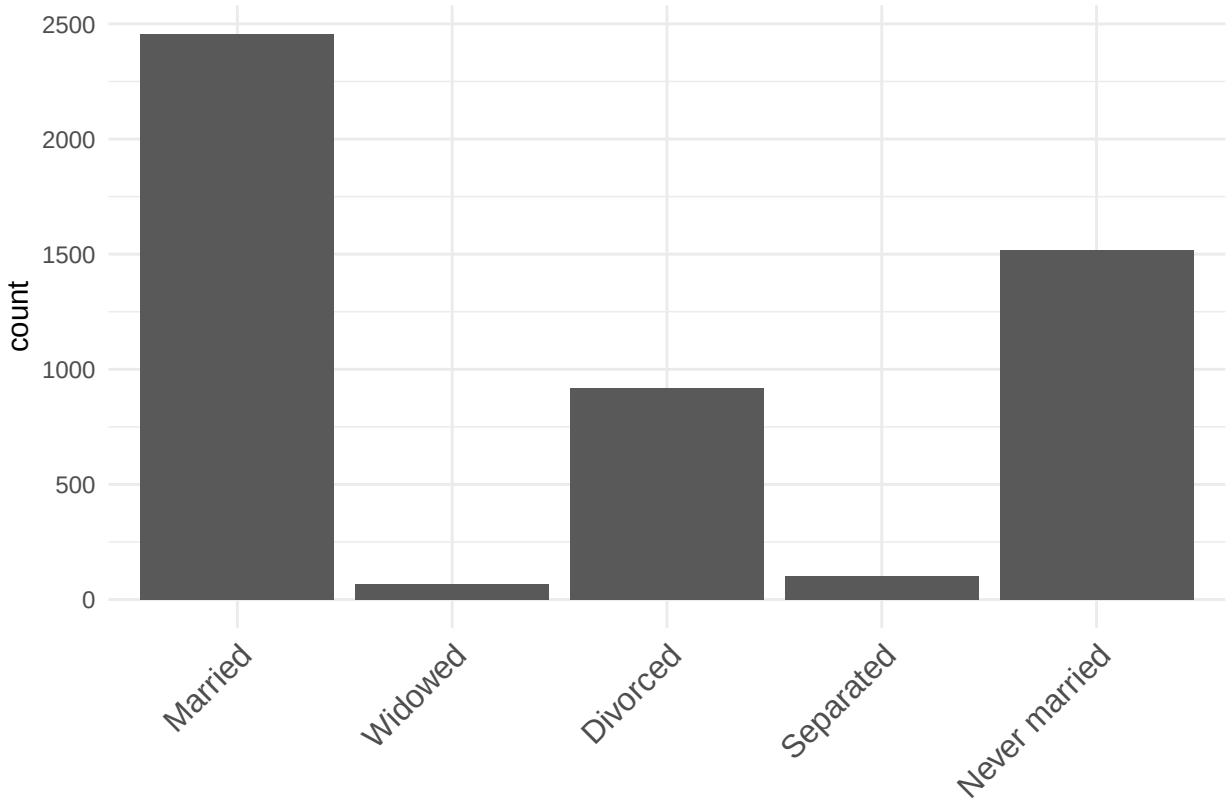

100000
75000
50000
25000
0
WAGP
28955.69
26745.9
No
Yes

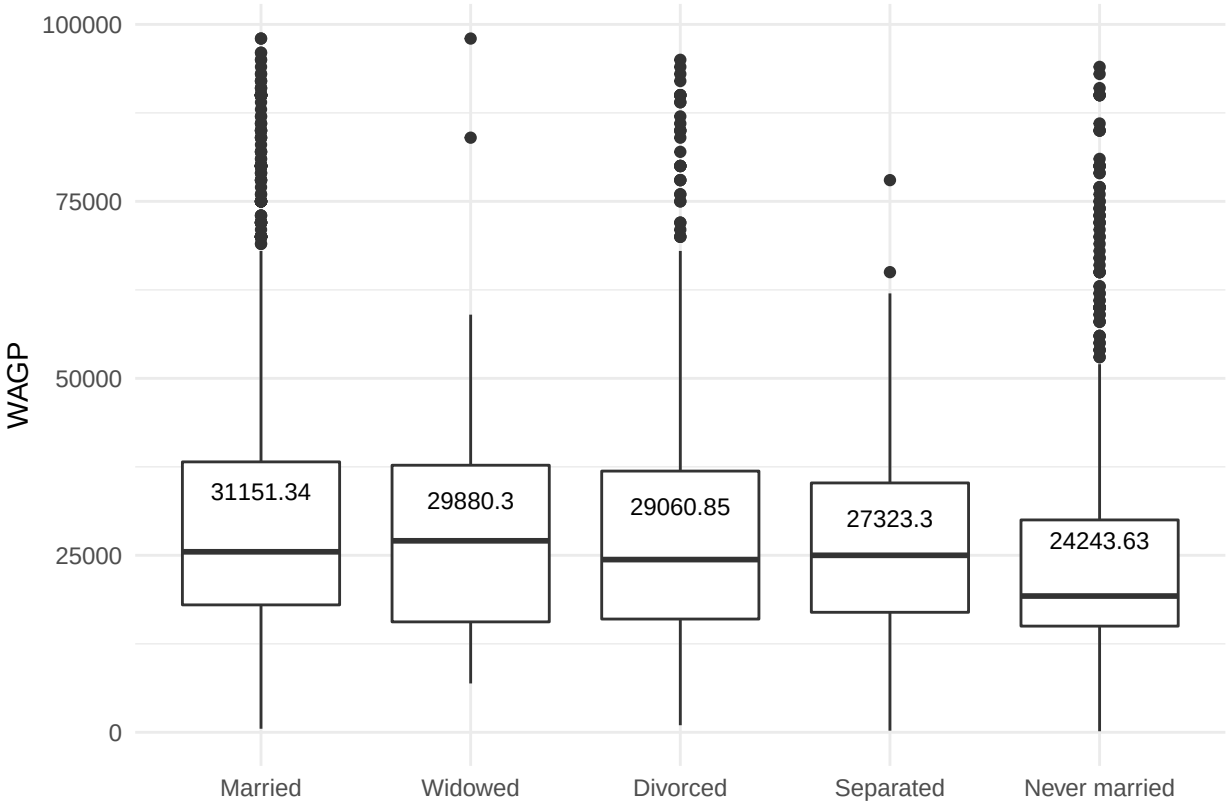
Ethnicity: Hispanic

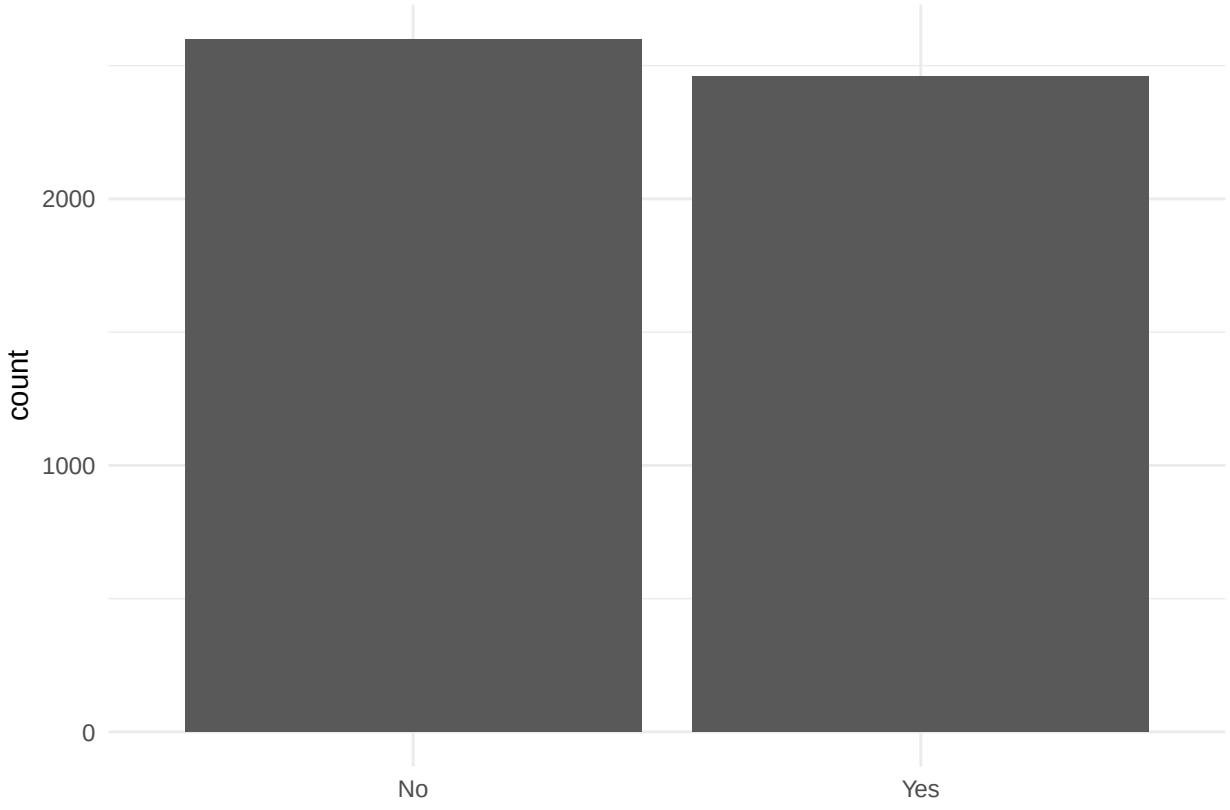


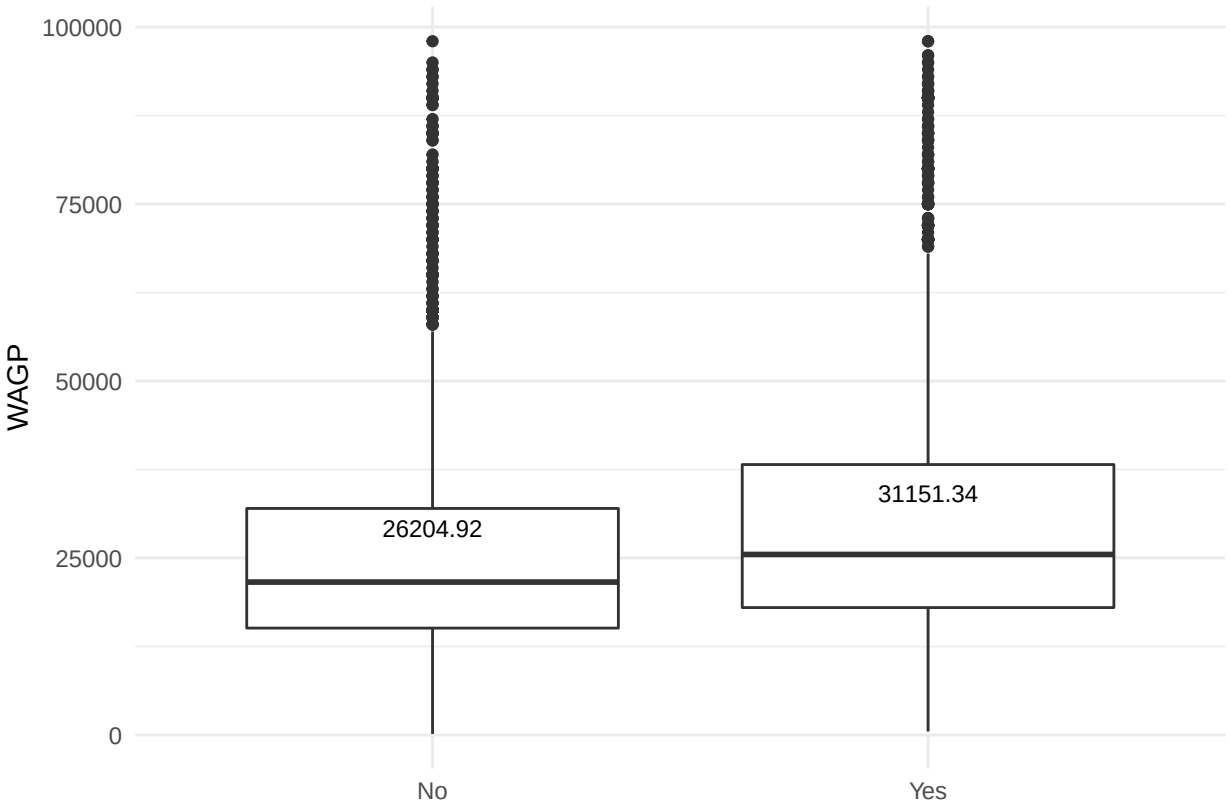


Marital Status



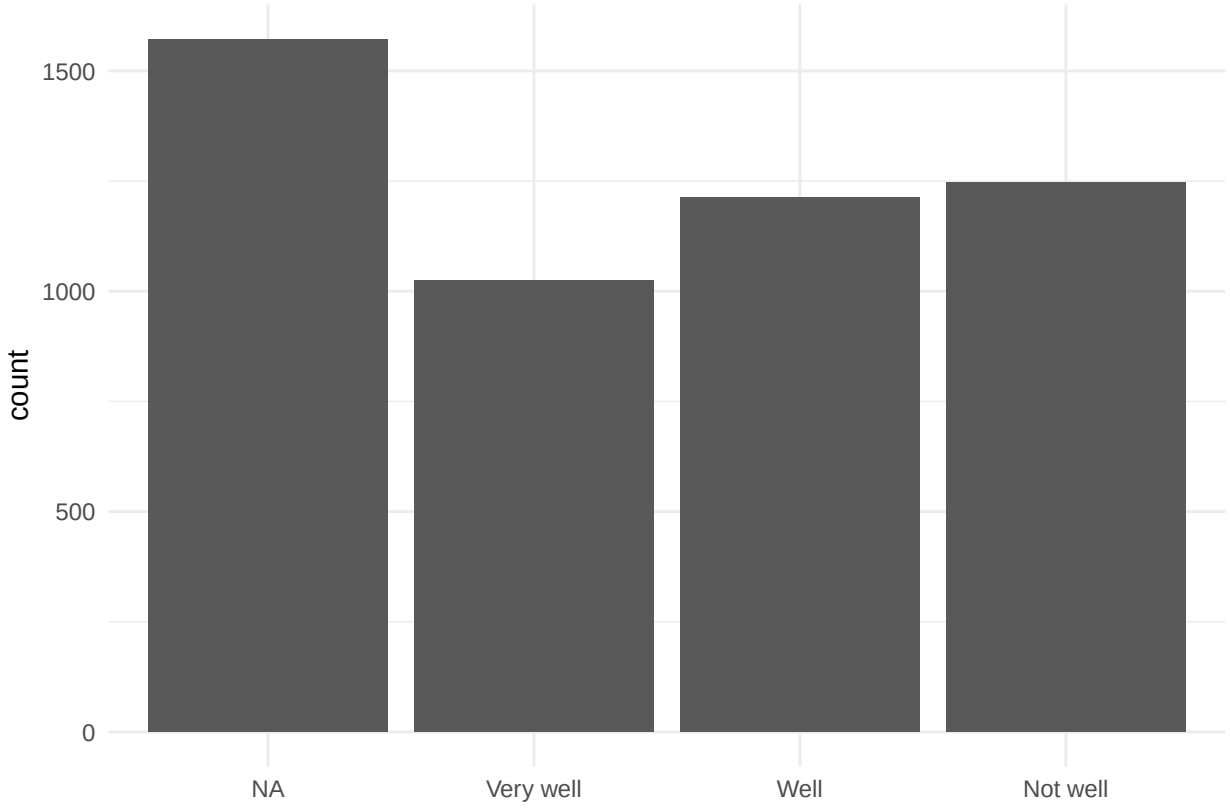


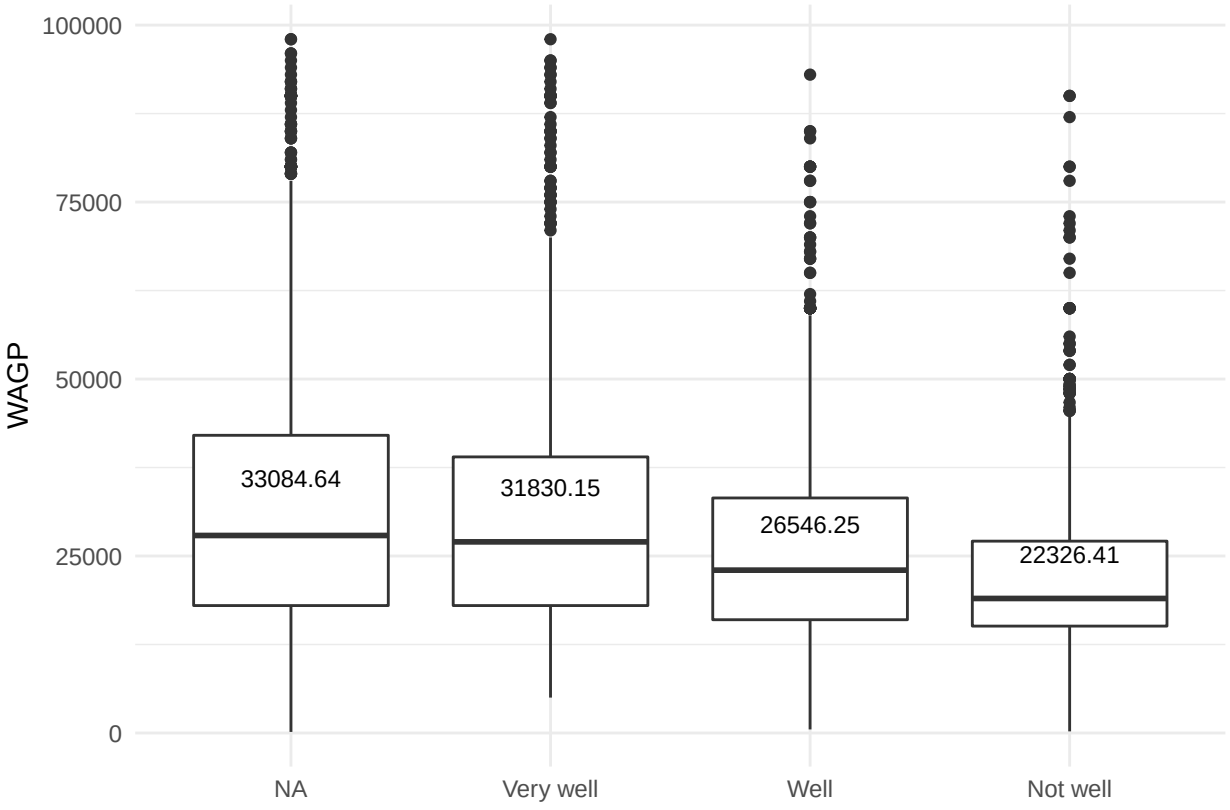

count
2000
1000
0
No
Yes



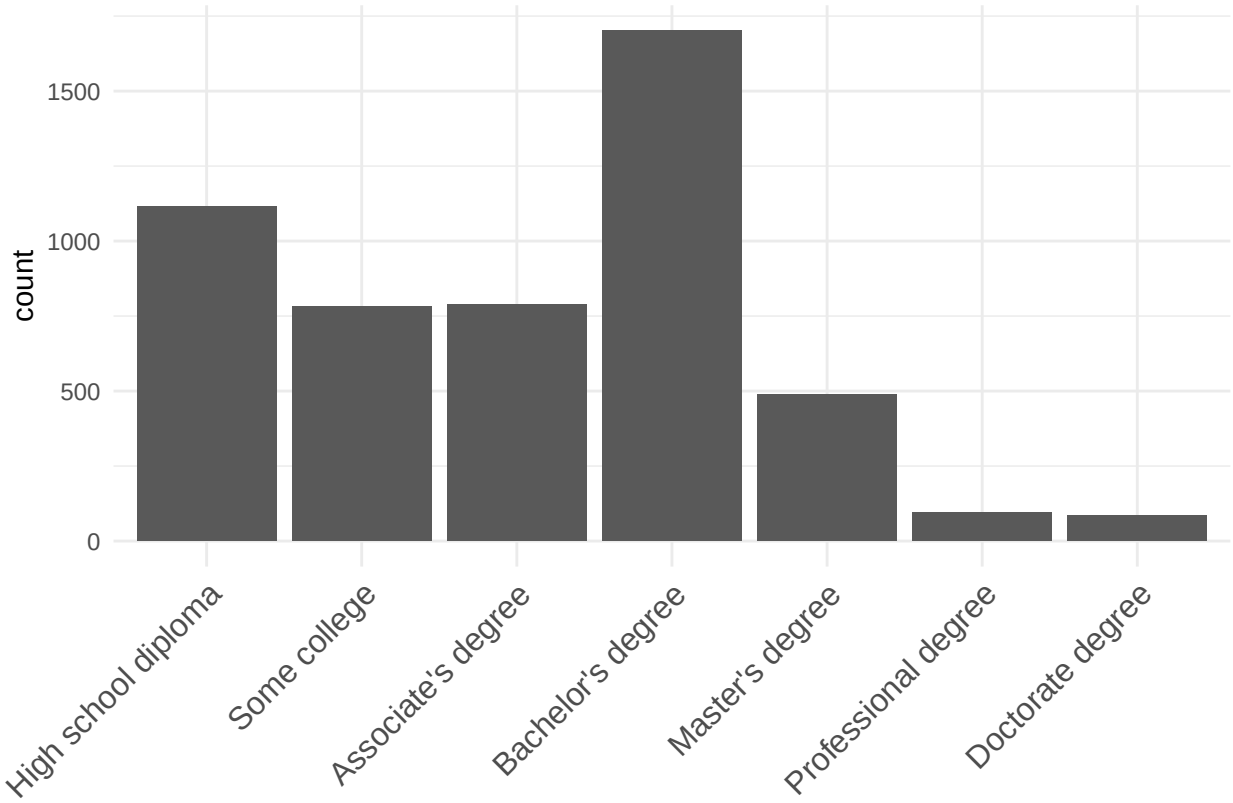
Ability to speak English

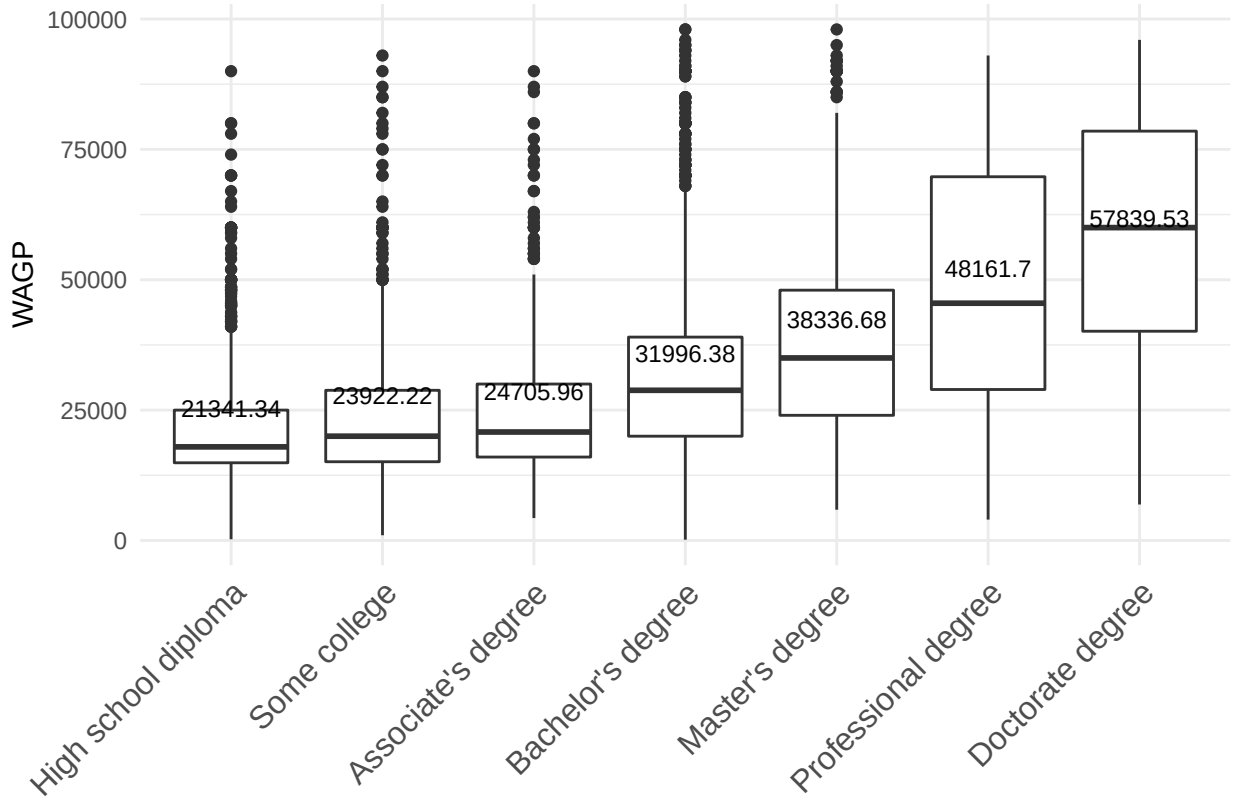
About half of population does not speak English well


1500
1000
500
0
count
NA
Very well
Well
Not well


100000
75000
50000
25000
0
WAGP
33084.64
31830.15
26546.25
22326.41
NA
Very well
Well
Not well

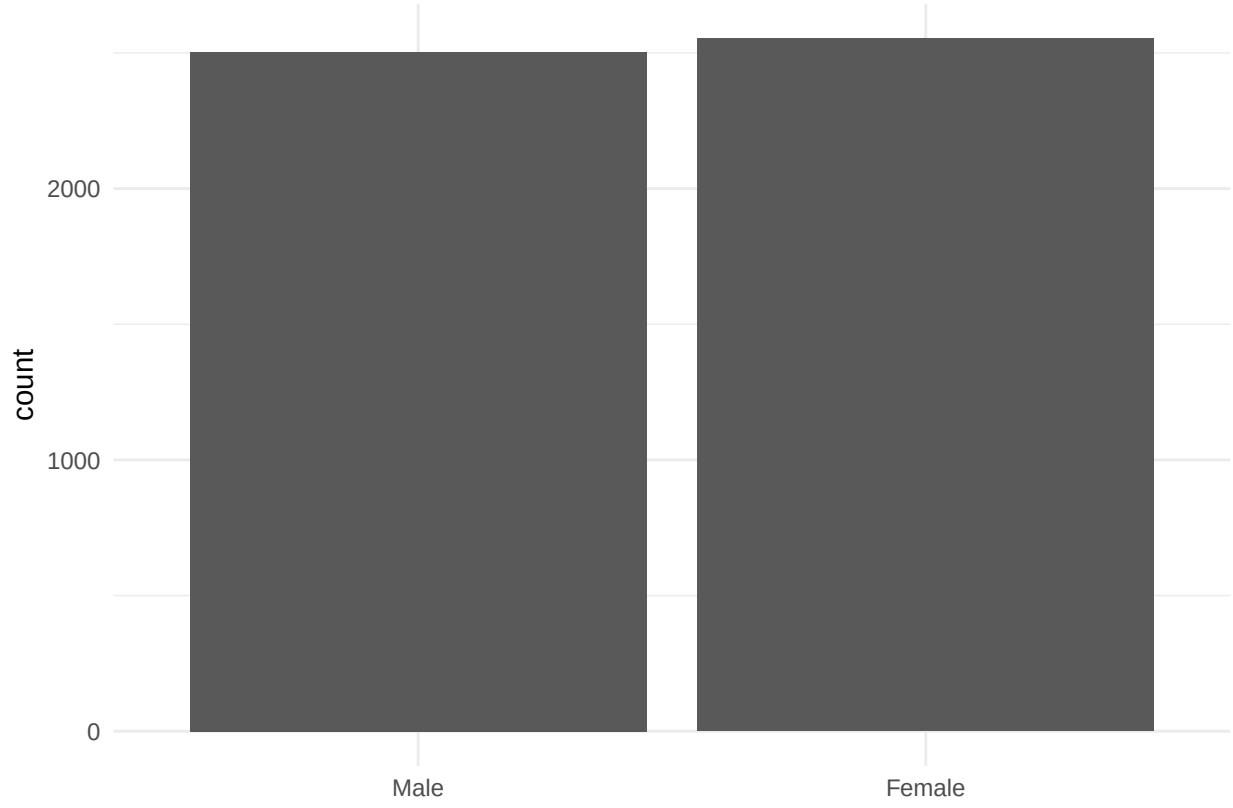
Educational attainment

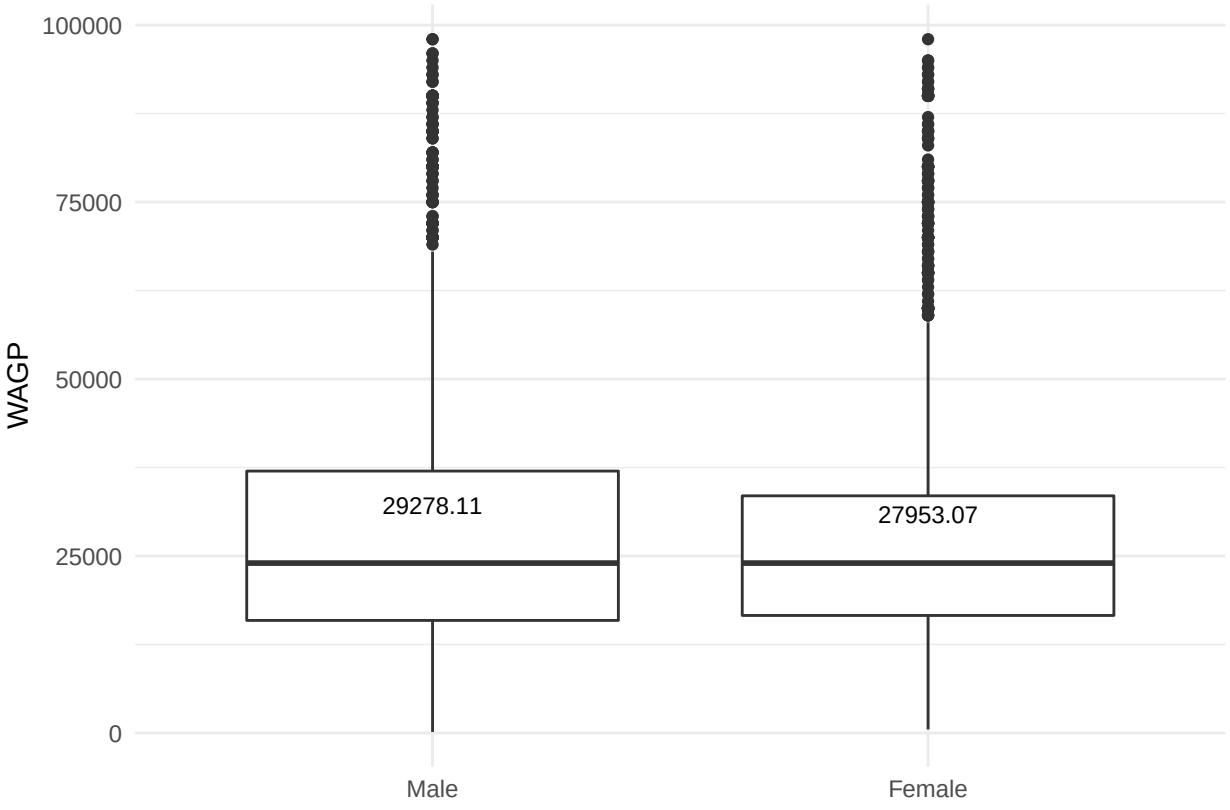




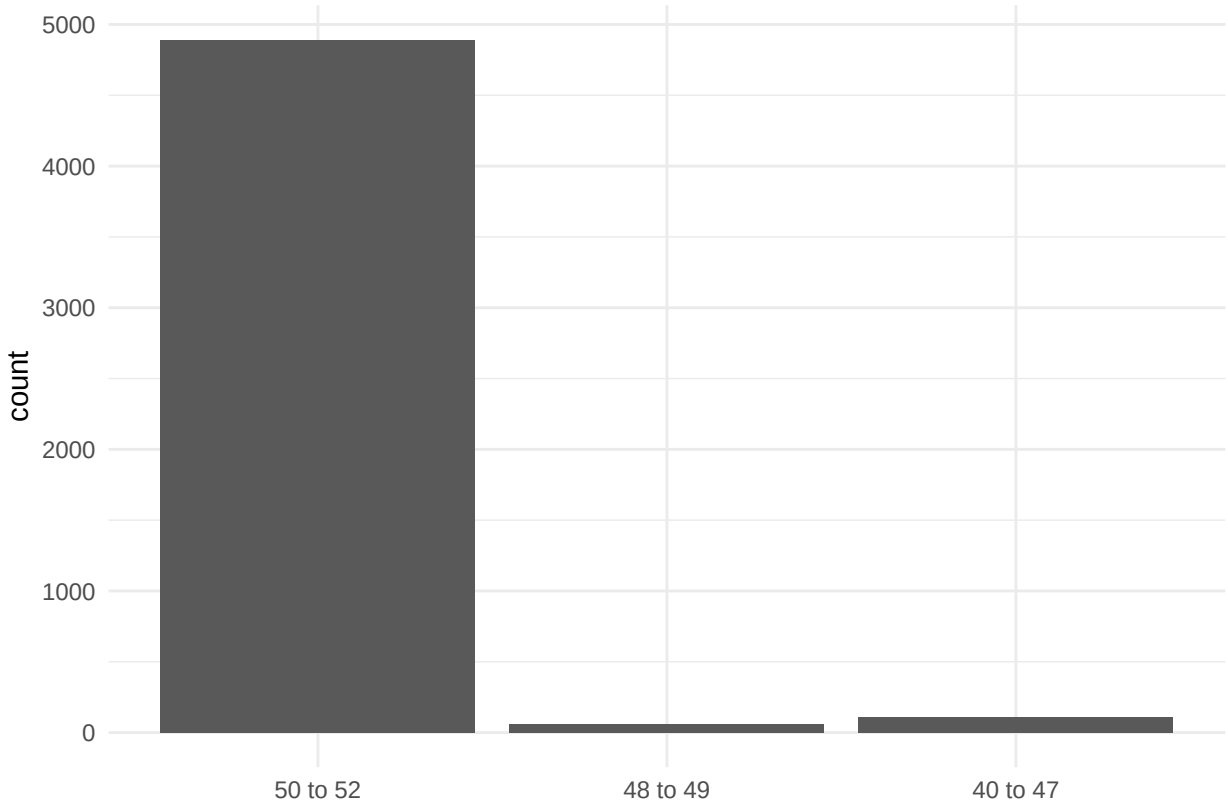
Sex

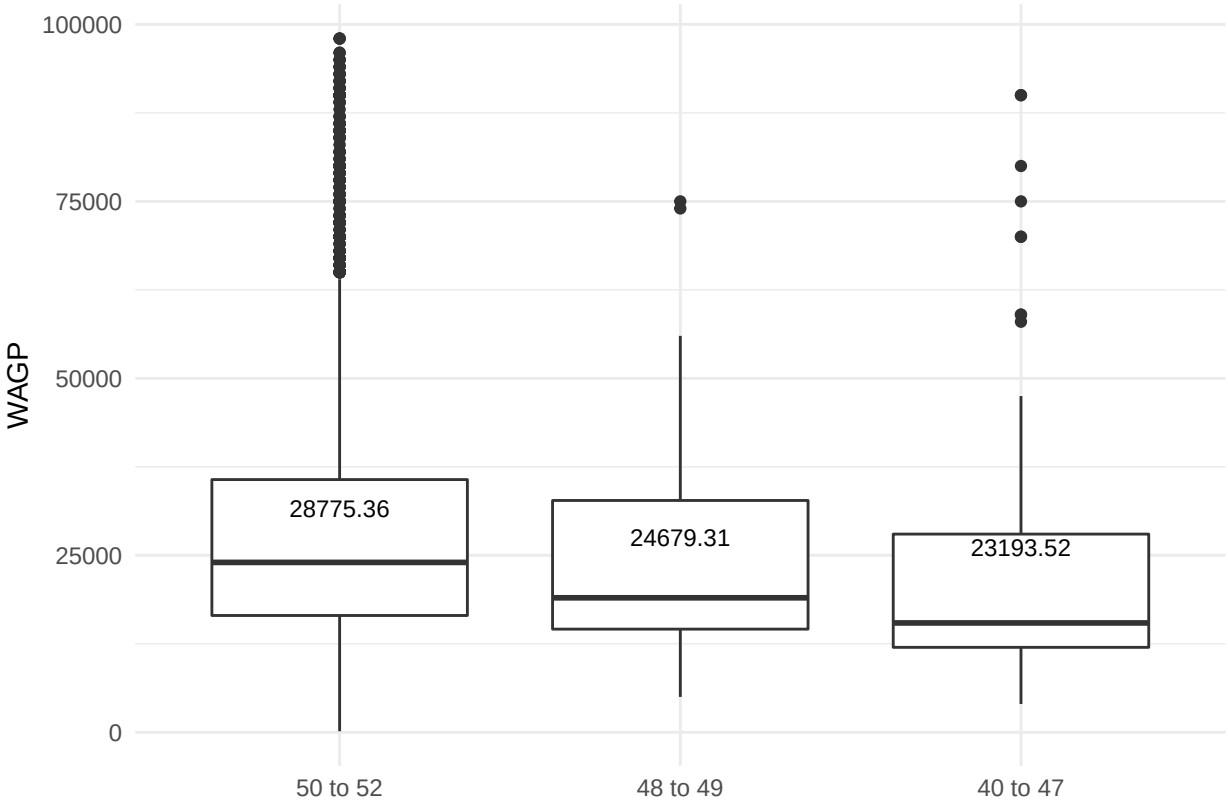
Gender is nearly equalized in Puerto Rico




100000
75000
50000
25000
0
WAGP
29278.11
27953.07
Male
Female

Work week

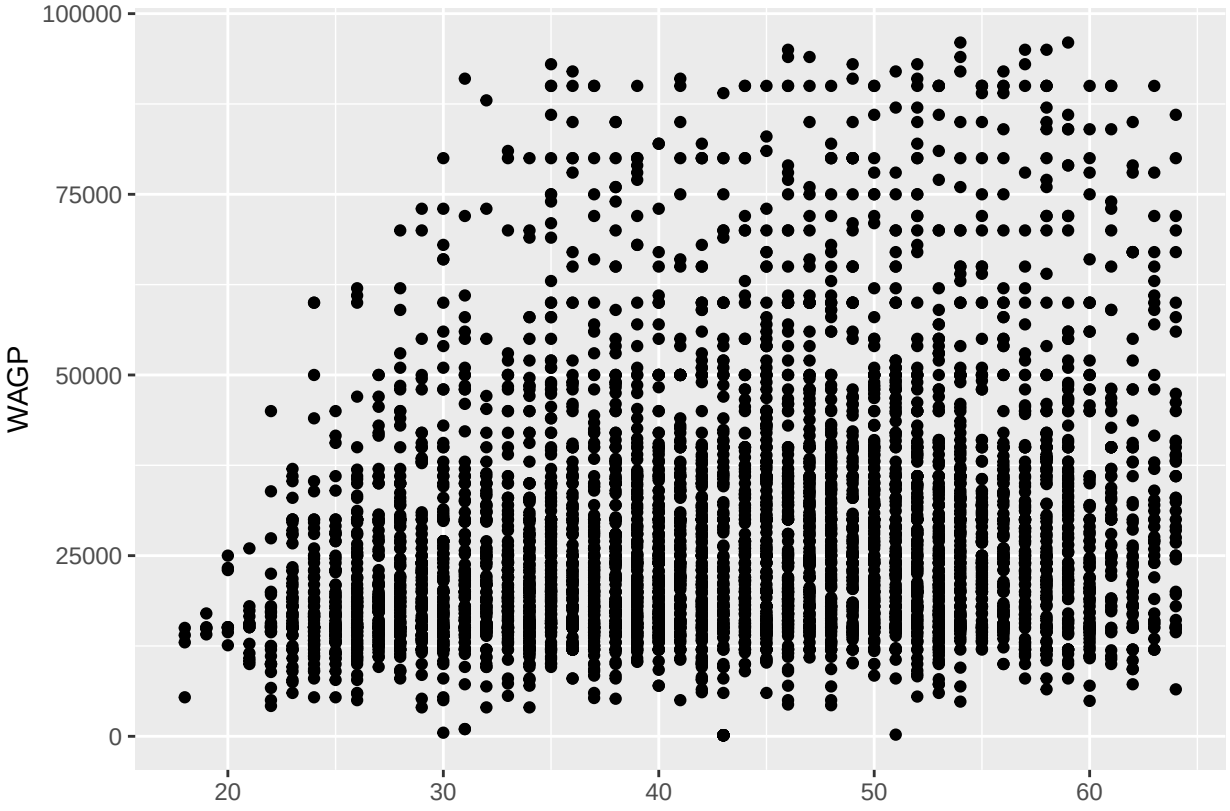

5000
4000
3000
2000
1000
0
count
50 to 52
48 to 49
40 to 47


100000
75000
50000
25000
0
WAGP
28775.36
24679.31
23193.52
50 to 52
48 to 49
40 to 47

Task 5:

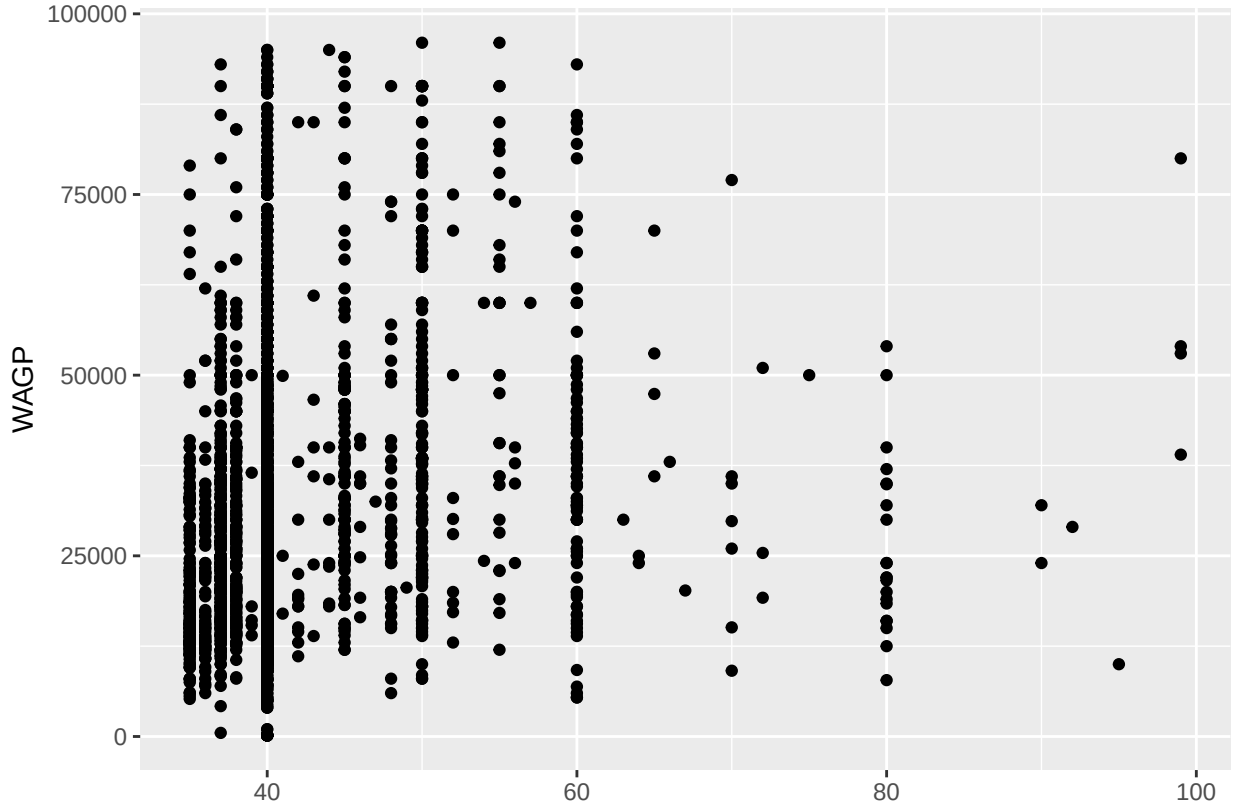
Generate and examine cross tabulations, scatterplots, and/or correlation coefficients of interest

Age vs. Wage



The correlation of 0.23 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. Wage



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.

Ethinicity: Hispanic

	RACHISP: No (N = 4,959)	RACHISP: Yes (N = 95)
Minimum	230	150
Maximum	96000	96000
Median	24000	18000
Mean	28594.62	27177.37

Race: White

	RACWHT: No (N = 1,376)	RACWHT: Yes (N = 3,678)
Minimum	230	150
Maximum	96000	96000
Median	22750	24000
Mean	26492.46	29344.47

Race: Black

	RACBLK: No (N = 4,294)	RACBLK: Yes (N = 760)
Minimum	150	230
Maximum	96000	93000
Median	24000	22850
Mean	28854.11	26951.36

Race: Other

	RACOTHER: No (N = 4,261)	RACOTHER: Yes (N = 793)
Minimum	150	1000
Maximum	96000	96000
Median	24000	23000
Mean	28907.08	26745.90

Marital Status

	MAR1: No (N = 2,598)	MAR1: Yes (N = 2,456)
Minimum	150	500
Maximum	95000	96000
Median	21600	25450
Mean	26177.28	31096.91

Educational attainment

	SCHL: High school diploma (N = 1,116)	SCHL: Some college (N = 783)	SCHL: Associate's degree (N =
Minimum	230	1000	4300
Maximum	90000	93000	90000
Median	17950	20000	20800
Mean	21341.34	23922.22	24705.96