

# Labor Market Analysis

## Exploratory Data Analysis

*Marjorie Blanco, Joe Thompson, Haodi Tu*

### Research Questions and Geographic area

#### Data

We use data from the 2016 ACS for Puerto Rico to examine earnings gaps between individuals who are TBD. The 2016 ACS is a nationally representative sample of 30224 households with a response rate of 97.6% (TBD). The household survey includes questions pertaining to each household member's demographic characteristics, and labor market activity. Our analysis relies on data collected on individuals living in households but not on those living in group quarters.

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 selfemployment 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 hr per week in a usual work week over this period.

#### Sample selection criteria

##### 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.

WAGP	SEX	AGEP	RAC1P	RACWHT	RACBLK	HISP	MAR	SCHL	ENG	COW	WKW	WKHP
34000	1	47	1	1	0	03	1	20	2	1	1	40
13000	1	58	2	0	1	03	5	16	3	1	1	40
18000	1	50	1	1	0	03	1	22	3	3	1	40
10300	2	39	1	1	0	03	1	21	3	1	1	40
28600	2	39	2	0	1	03	1	21	1	1	1	45
24800	1	37	2	0	1	05	1	21	2	1	1	46
22000	2	47	8	0	0	03	5	20	4	3	1	40
19000	2	60	2	0	1	03	5	16	2	1	1	40
87000	2	58	1	1	0	03	3	20	1	5	1	40
8000	1	33	1	1	0	03	5	21	1	1	1	35
22900	1	61	1	1	0	03	3	16	4	5	1	40
19000	1	39	2	0	1	03	1	20	4	1	1	38
19600	2	36	2	0	1	03	1	21	4	5	1	40
48000	1	30	9	1	1	03	3	18	1	5	1	40
40000	2	30	1	1	0	03	5	18	1	1	1	40
15600	2	41	1	1	0	03	5	16	4	1	1	40
12100	1	46	1	1	0	03	3	16	4	1	1	40
14000	1	53	1	1	0	03	1	16	3	1	1	40
80000	1	38	1	1	0	03	5	21	1	1	1	40
15100	2	26	8	0	0	03	5	16	4	1	1	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
Maximum	98000
Median	24000
mean (sd)	28,609.17 \$\pm\$ 17,072.31

### **Age**

	ss16ppr (N = 5,057)
Minimum	18
Maximum	64
Median	43
mean (sd)	42.95 \$\pm\$ 10.84

### **Hours worked**

	ss16ppr (N = 5,057)
Minimum	35
Maximum	99
Median	40
mean (sd)	40.93 \$\pm\$ 5.38

### **Ethicity: Hispanic**

RACHISP	Count
No	4962
Yes	95

### **Race: White**

In Puerto Rico, majority of people identify themselves as white. Minority races include American Indian, American Indian and 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

## Marital status

RACBLK	Count
No	5057

## Marital status

MAR1	Count
No	2599
Yes	2458

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

## Ability to speak English

ENG	n()
NA	1572
Very well	1025
Well	1213
Not well	1247

## Educational attainment

TODO: Please update. This is not the case

- We will category people in four education level
  - People in sec 0 - 15 don't have diploma and are considered as no education
  - People in sec 16 - 17 have high school education
  - People in sec 18 - 21 have college education
  - People in sec 22 - 24 have after bachelor's education

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

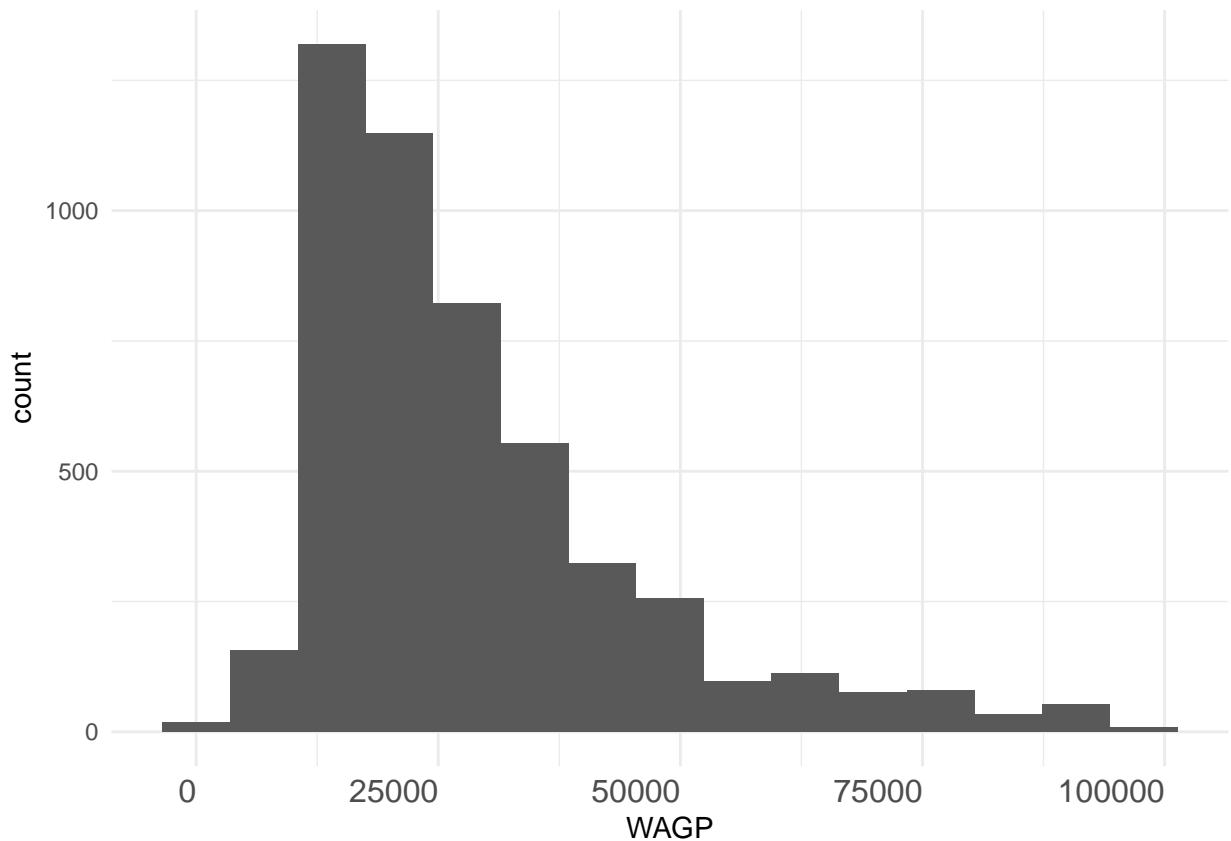
## Sex

SEX	Count
Male	2504
Female	2553

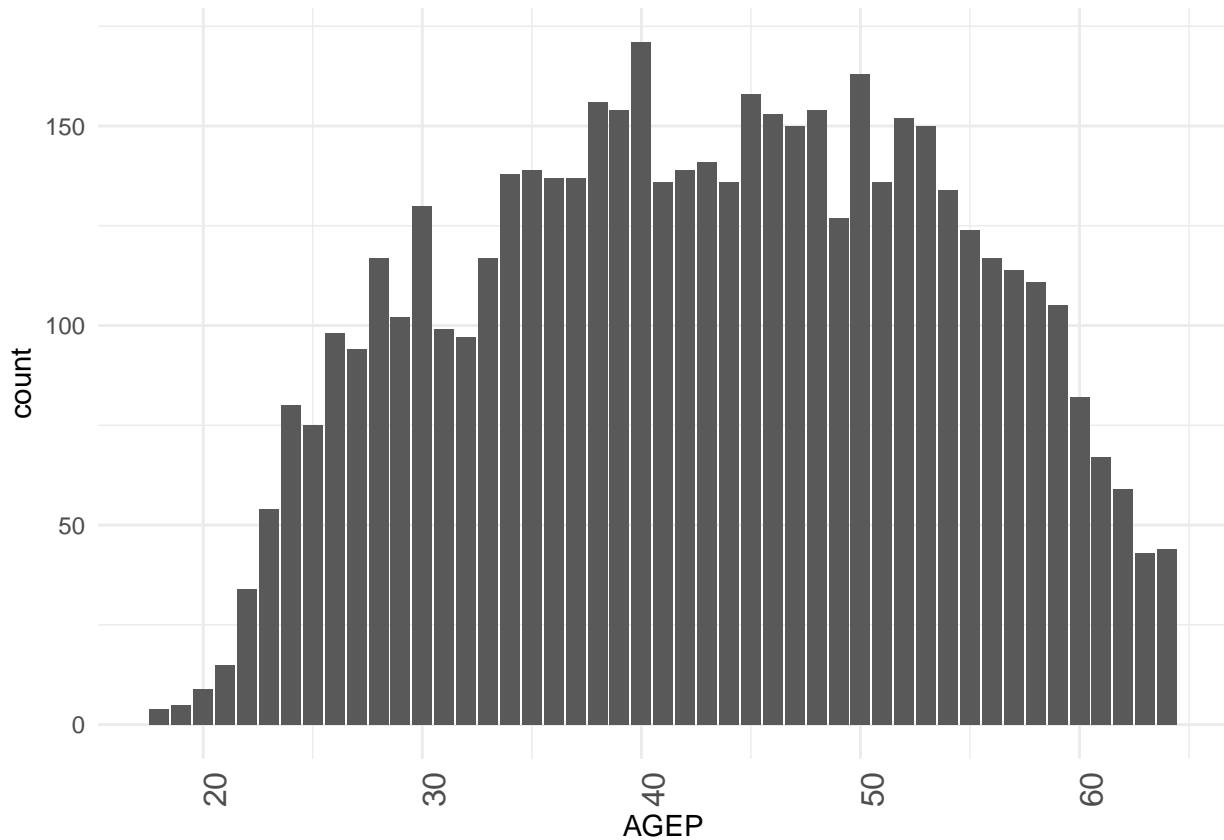
## Task 3:

Generate and examine histograms for quantitative variables of interest

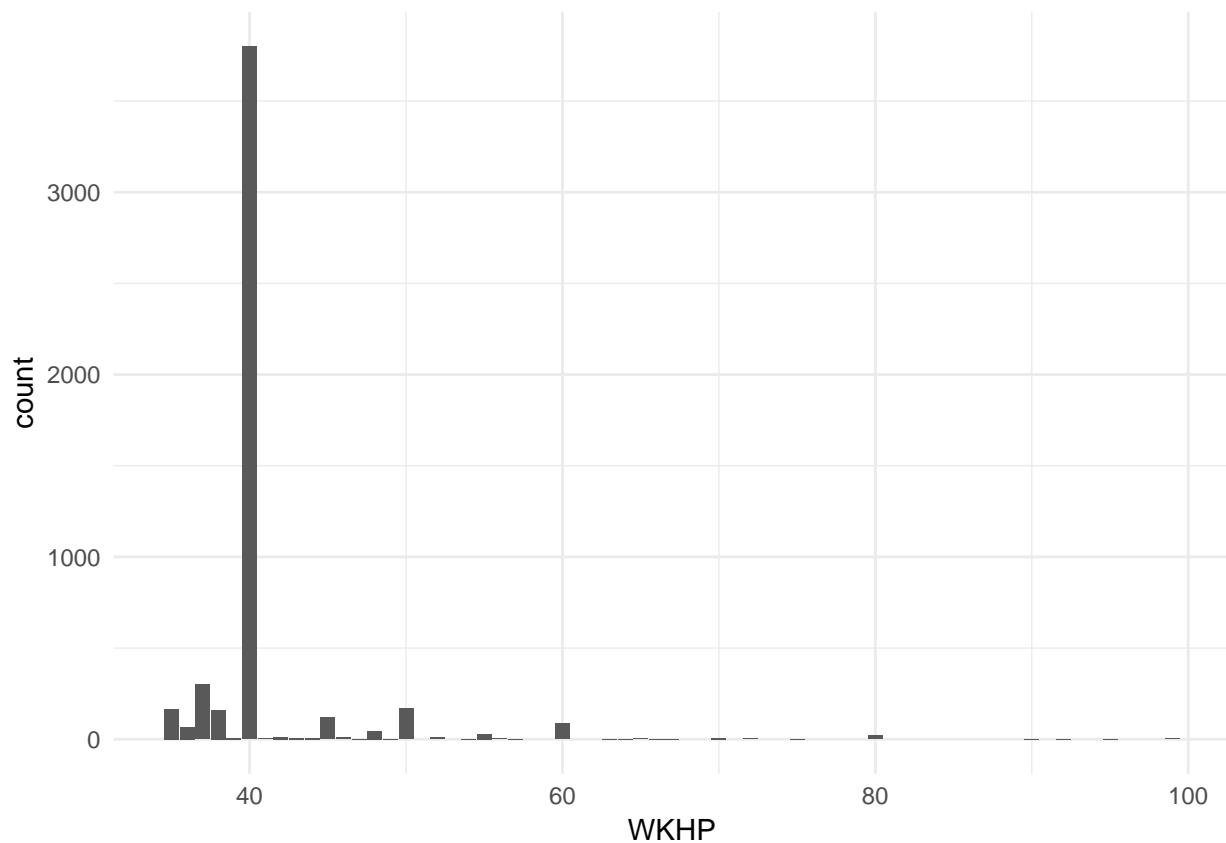
## Wage



## Age



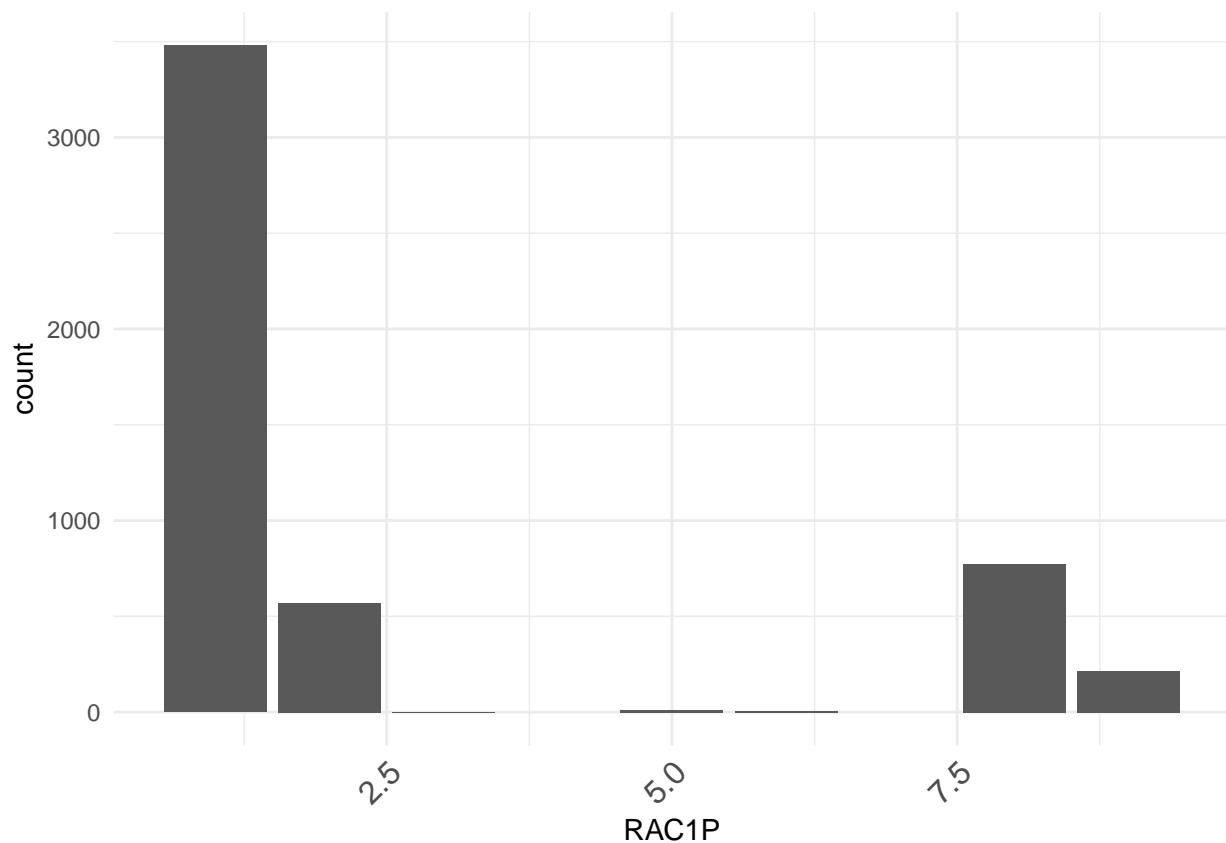
### Hours worked



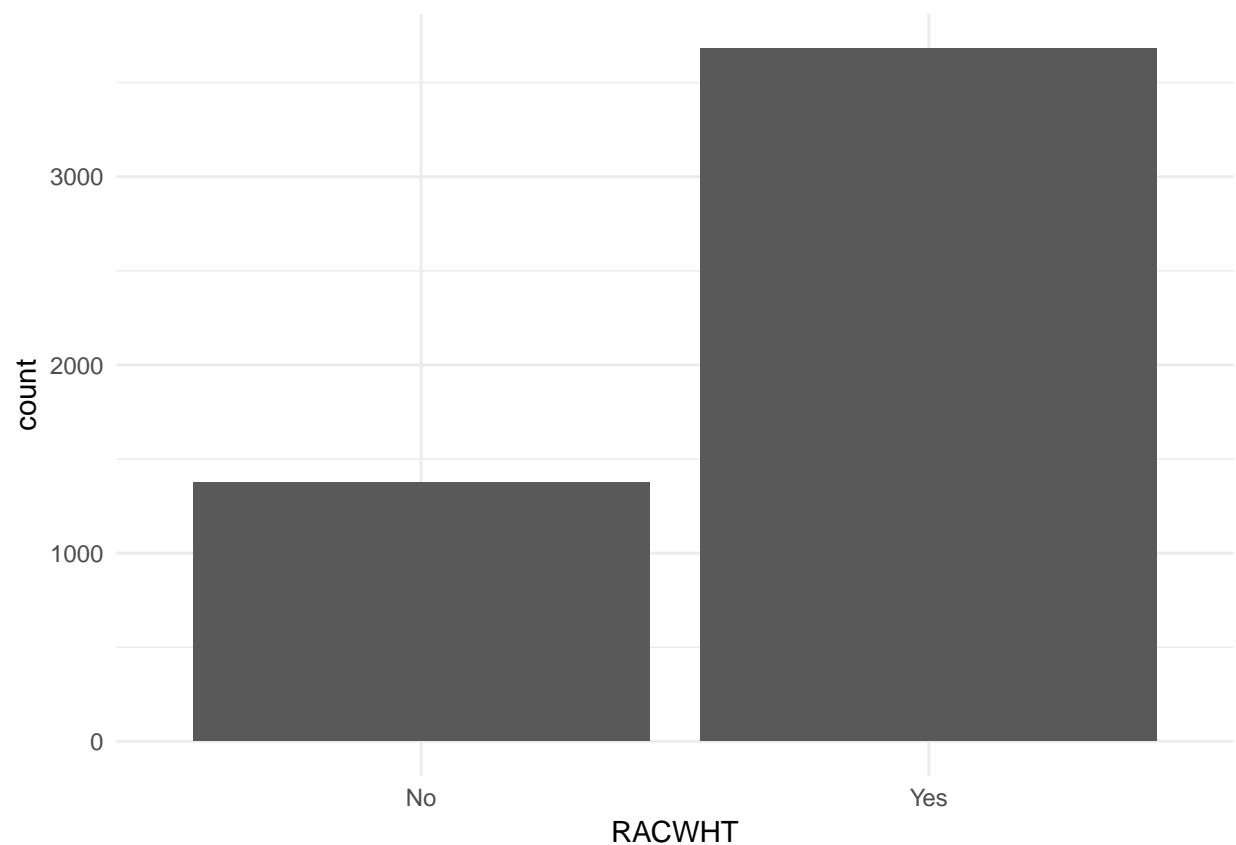
### Task 4:

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

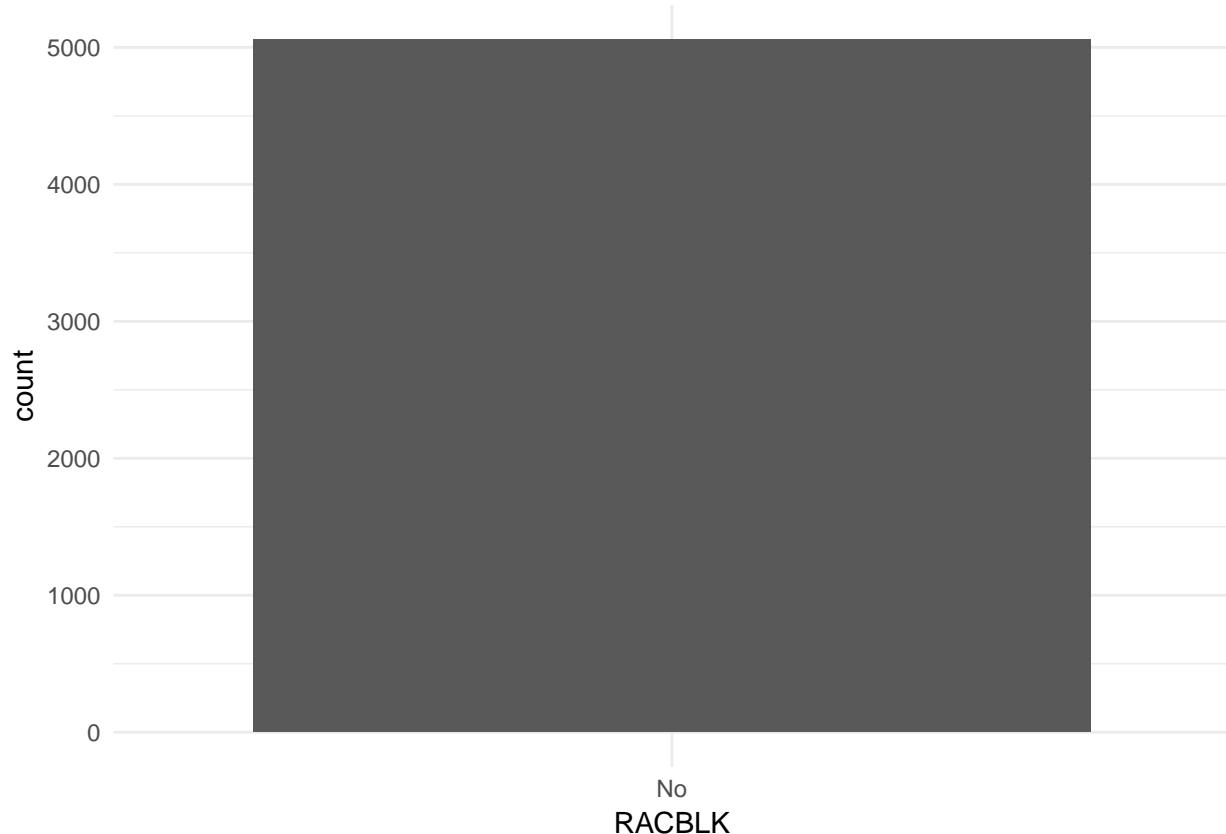
## Race



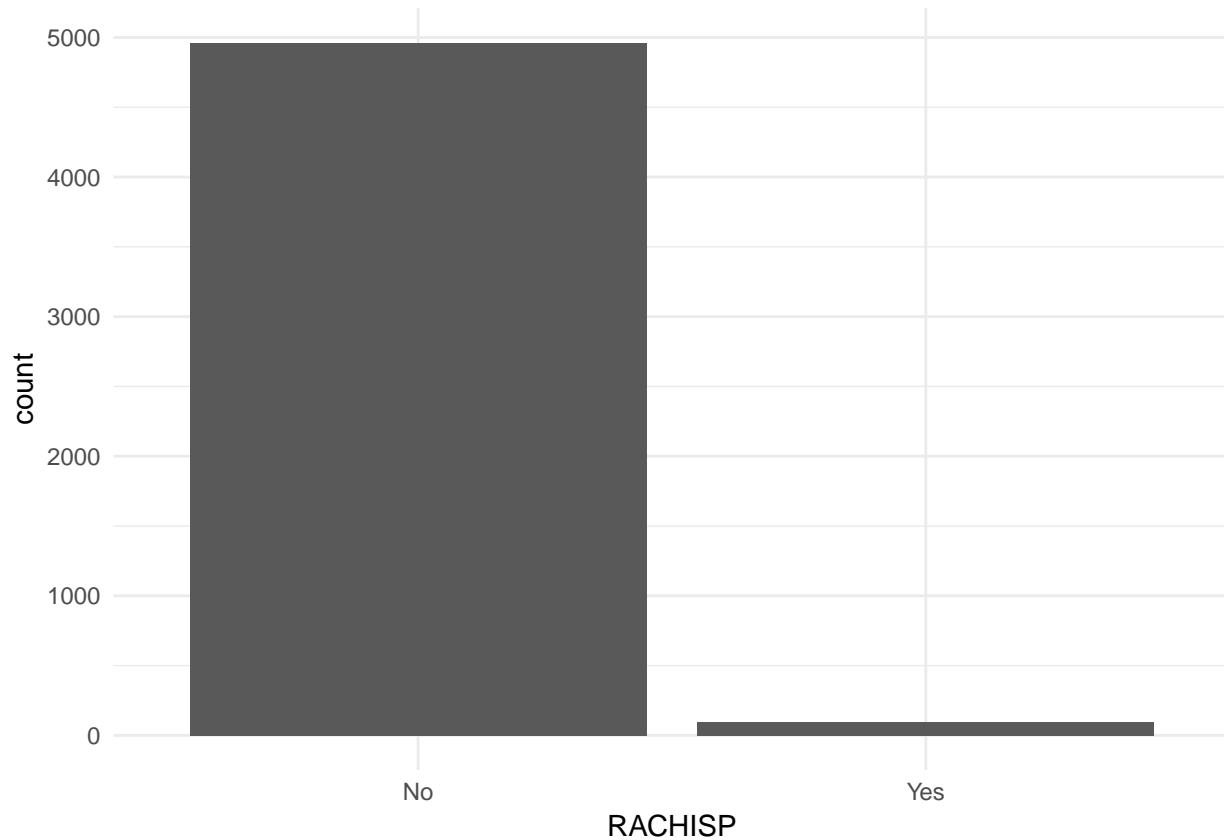
**White**



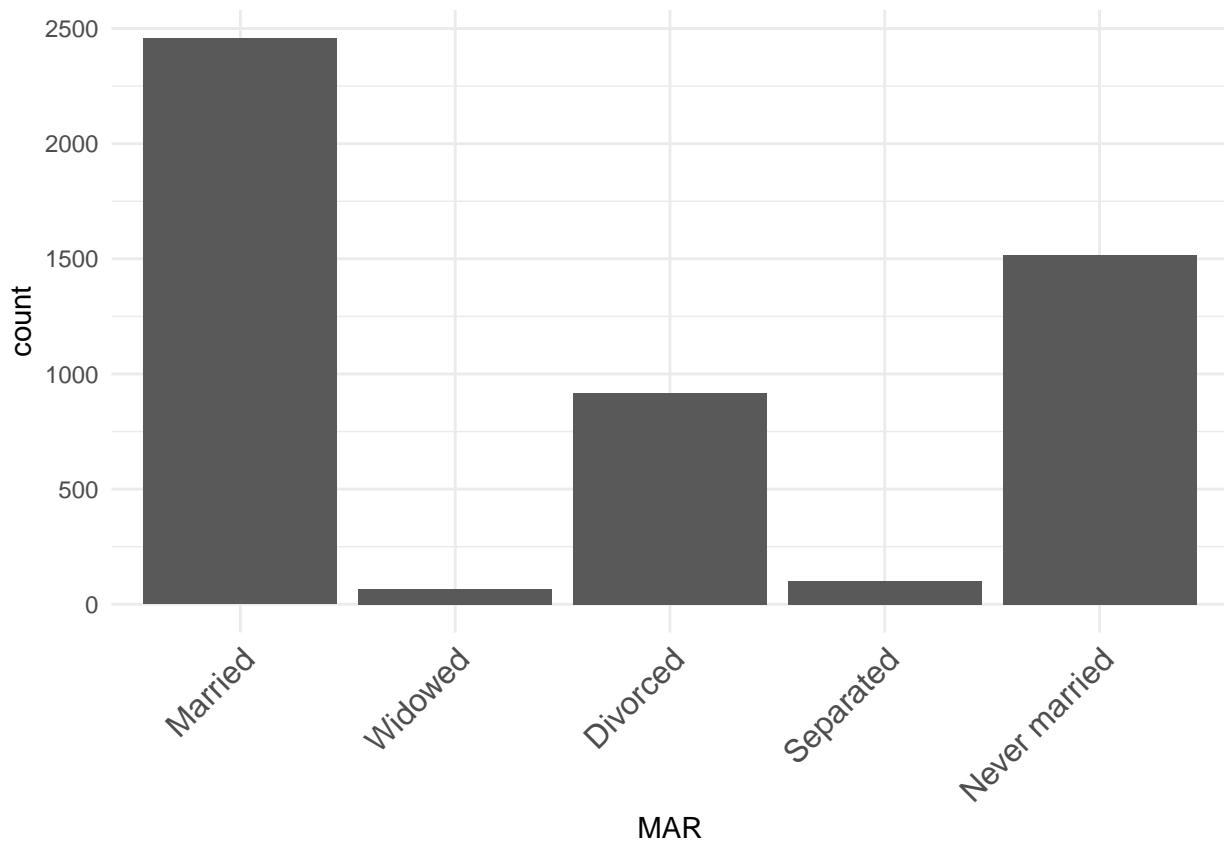
**Black**

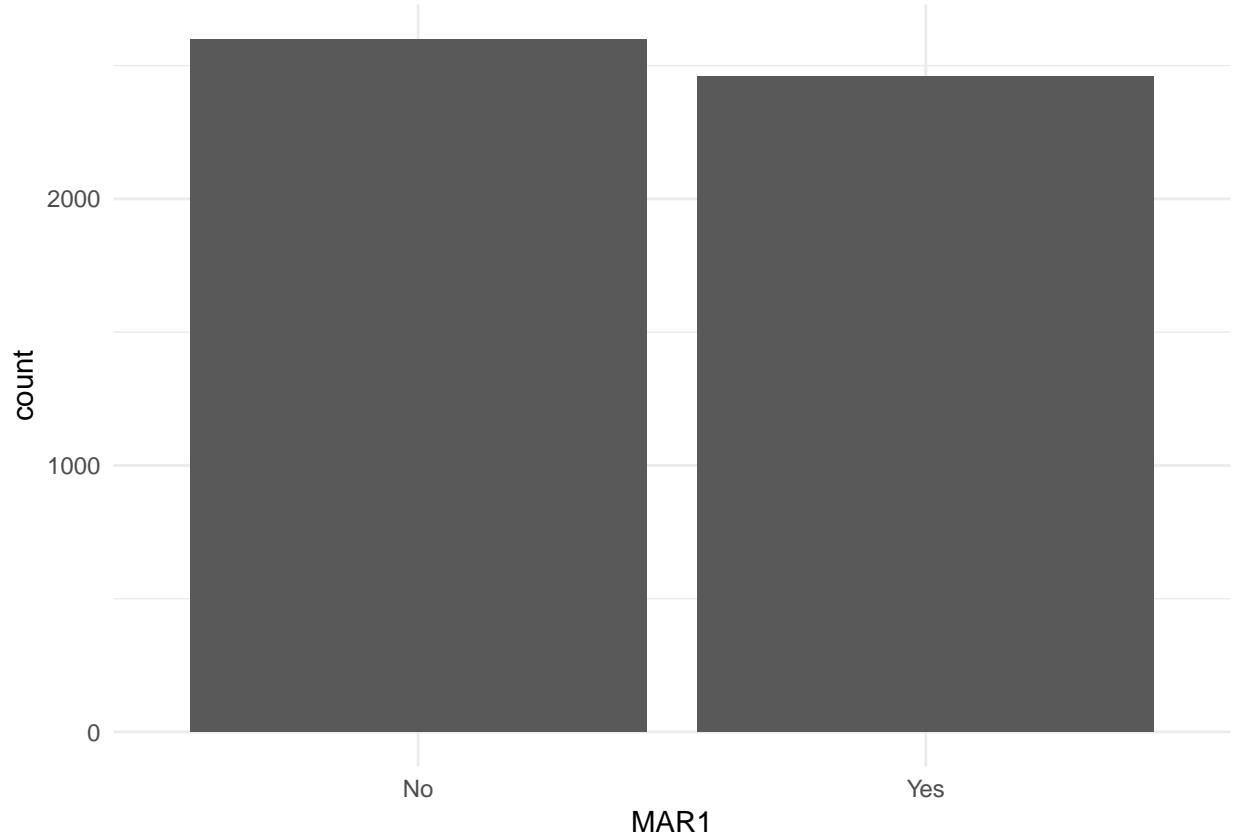


### Hispanic



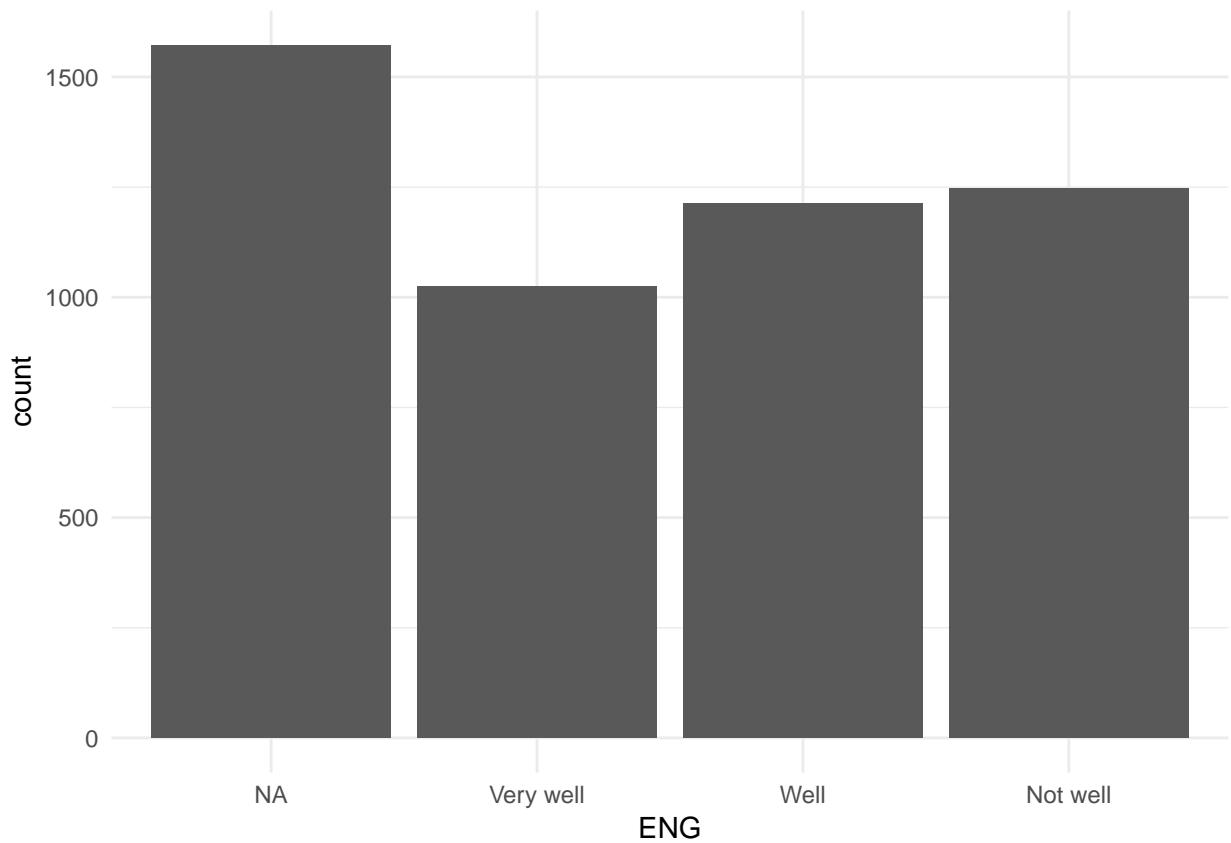
### Marital Status



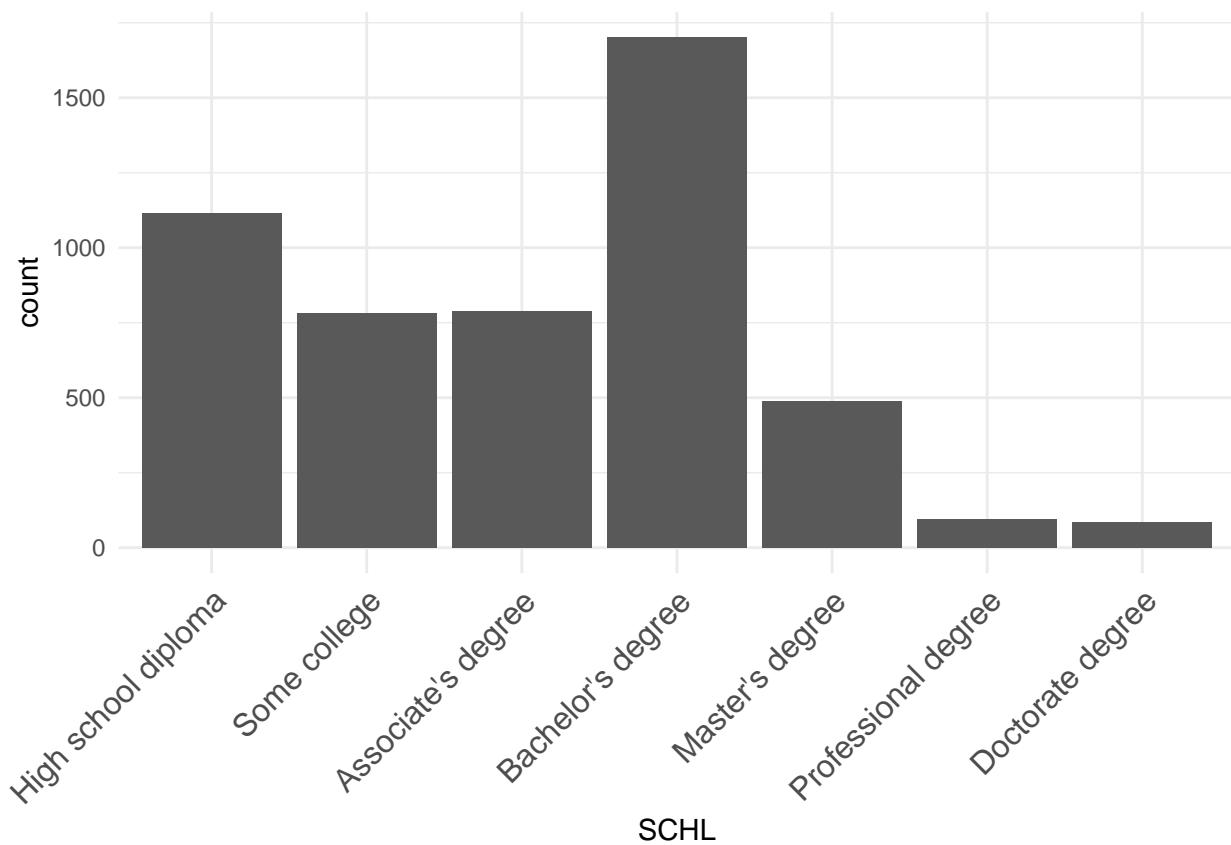


### Ability to speak English

About half of population does not speak English well

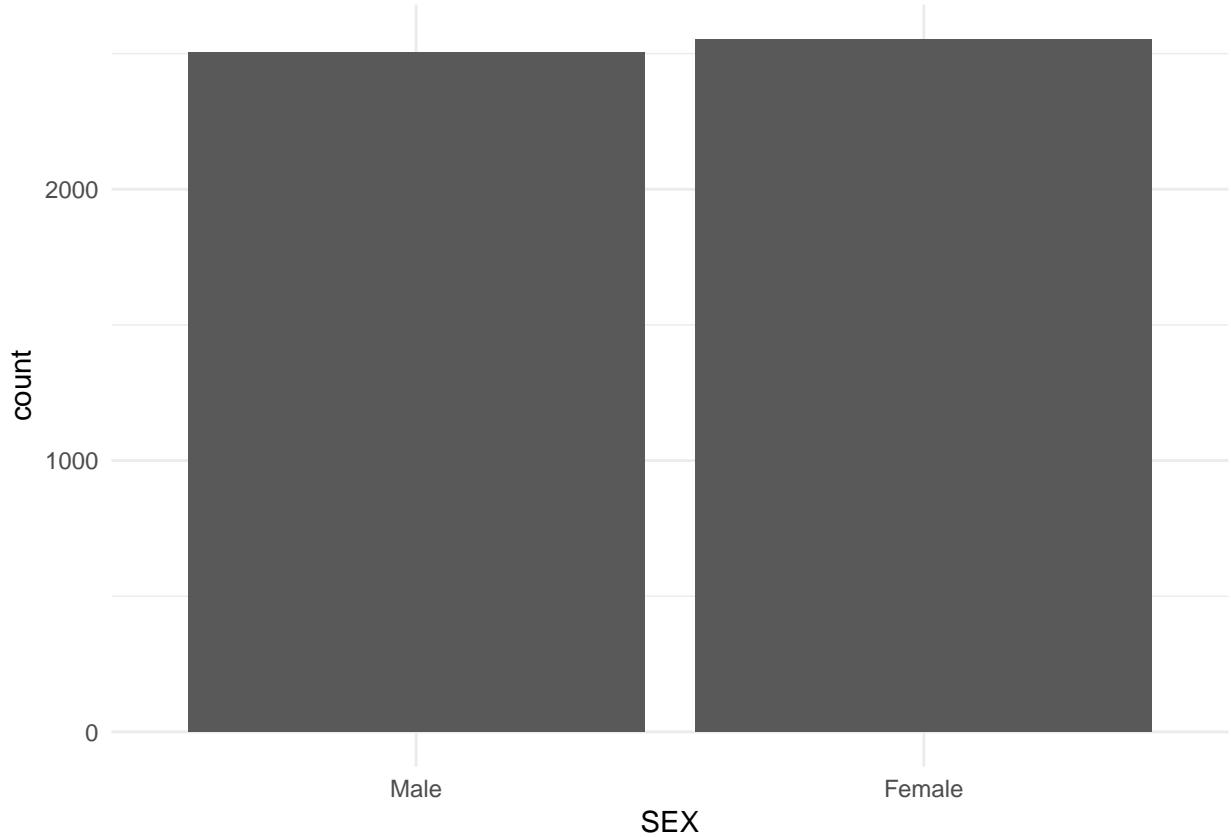


## Educational attainment

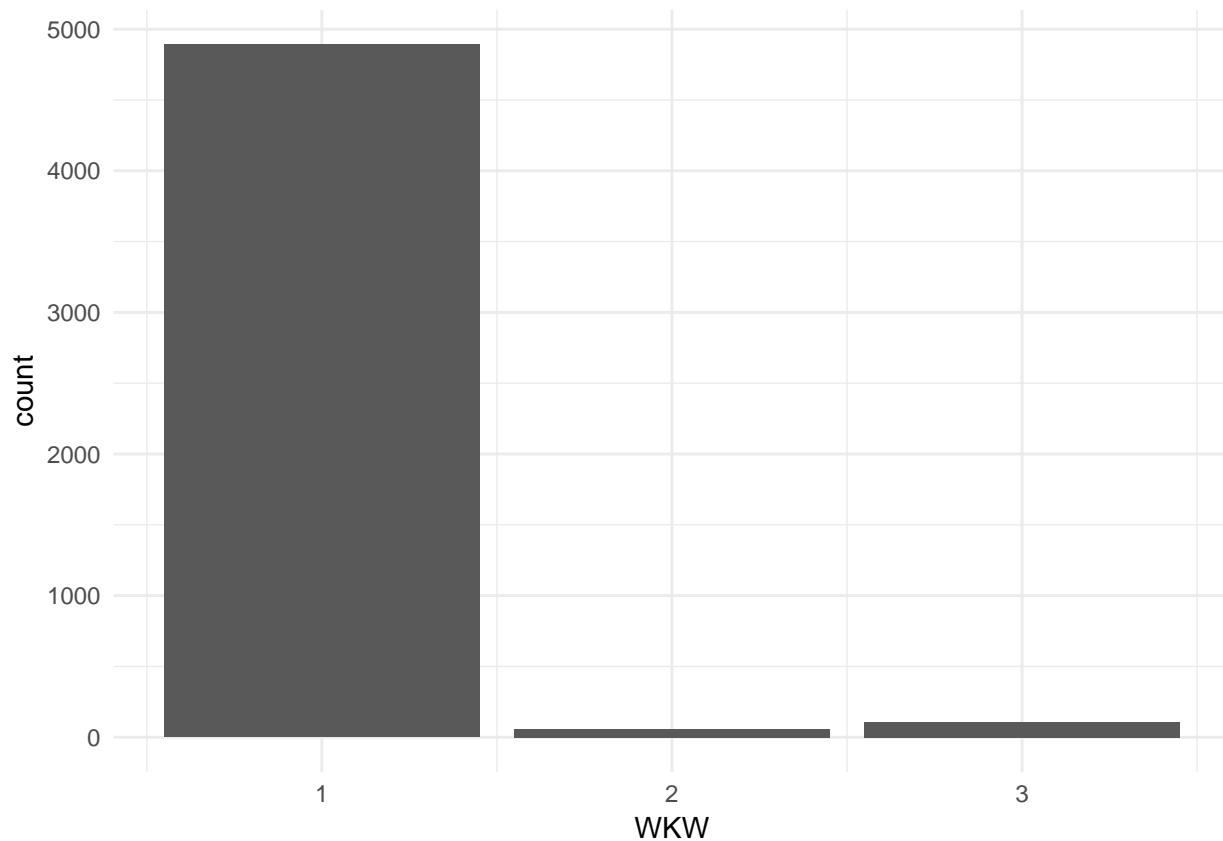


## Sex

Gender is nearly equalized in Puerto Rico

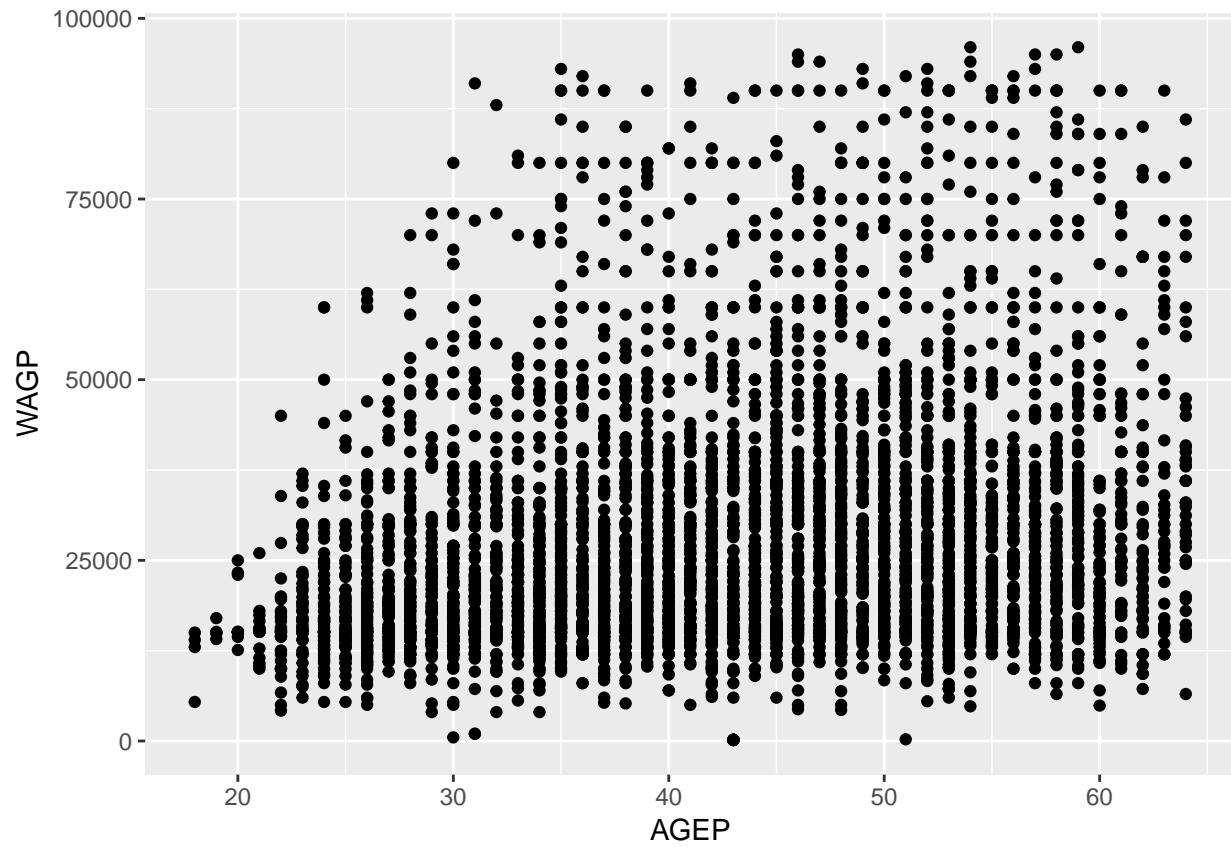


## Work week



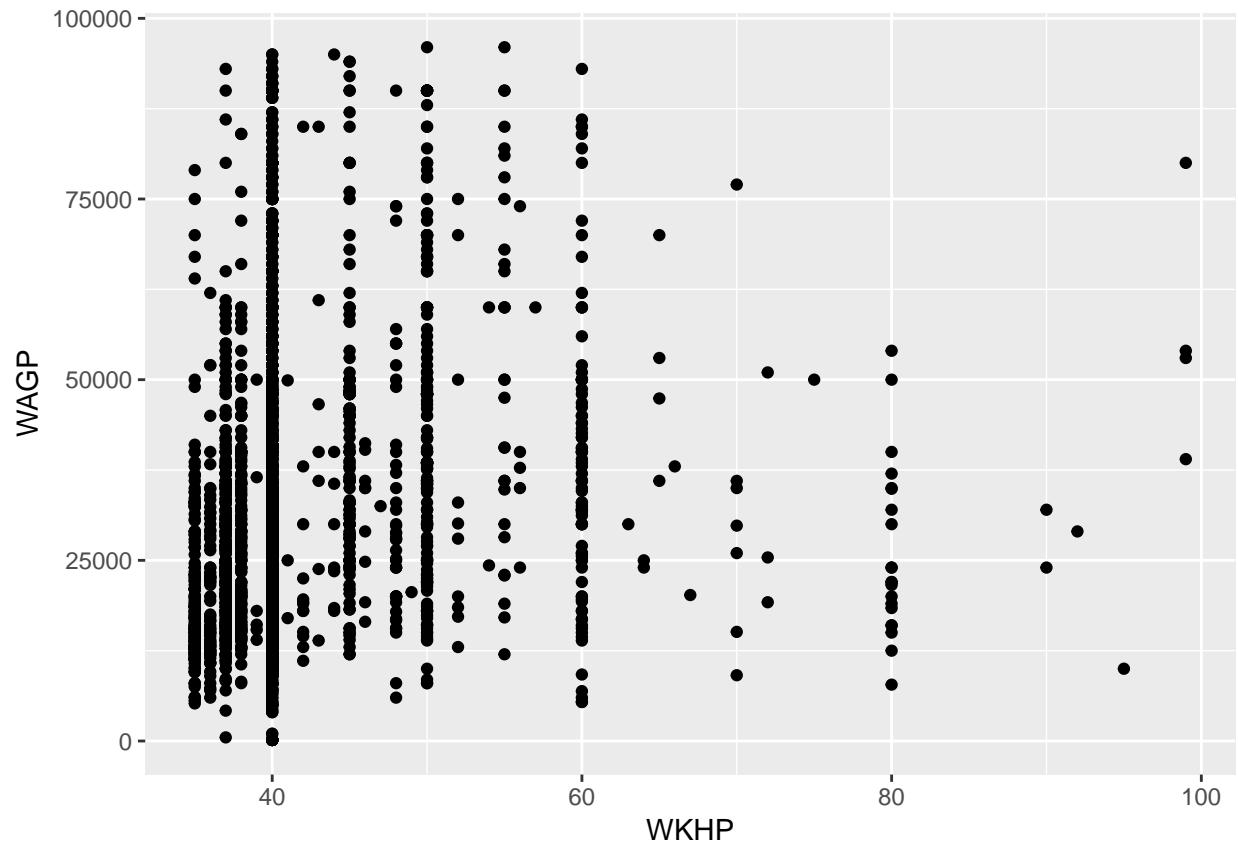
Task 4:

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



```
## [1] 0.2348047
```

TODO: Describe the correlation



```
## [1] 0.1829235
```

TODO: Describe the correlation