Data Insights

Step 1

Initially, the population records for the state of New Hampshire are imported. The Serial Number for identifying Households and Individuals is the same for years before 2018. So, for our analysis, we will only use the population records. We begin with creating a new dataset containing the following variables.

SERIALNO - Individual Household IDs

ST and PUMA - State ID and PUMA

AGEP - Age

SEX - Sex

RAC1P - Race

POVPIP - Income as measured relative to Federal Poverty Line

HICOV - Health Insurance Coverage

HINS1 - Insurance through Employer

HINS3 - Medicare Eligibility

HINS4 - Medicaid or other government assistance

HINS5 − TRICARE

HINS6 − VA Coverage

Step 2

The Race and Sex variables are transformed into dummy variables, White and Female.
A new variable called tax family measures the number of people filing a joint tax return.
The variable is computed by grouping and comparing the SERIALNO and POVPIP.
Observations have been dropped where the size of the tax family is greater than six.
The dataset has been filtered to include eligible individuals for Premium Tax Credit
Program. The following individuals have been excluded:

Those who have employer sponsored coverage

Those who have Medicare or are eligible for Medicare with age

Those who have TRICARE

Those who have VA Coverage

Those who have Medicaid

Those who are under 19 and their household income is below 318% poverty line

Those who are 65 or older

Coverage Family variable has been calculated by grouping and comparing the Serial
Number and Tax Family.

Step 3

Since numerous observations are missing individual and family income data in the
PUMS dataset, we calculated the expected family income using the Federal Poverty
Line for 2017 and Income relative Federal Poverty Line (POVPIP).
To compute the family income, the following formula has been used.

$$Family\ Income = \frac{POV\ PIP \ * \ FPL_t}{400}$$

FPL = Federal Poverty Line, t = corresponding family size

Family Size	Federal Poverty Line for 2017 (USD)
1	11,880
2	16,020
3	20,160
4	24,300
5	28,440
6	32,580

Source: Internal Revenue Services

To compute the contribution amount, we have extracted the expected percentage bracket across the income relative to Federal Poverty Line for the year 2017.

Income Level	Contribution Percentage for 2017
Up to 133% FPL	2.04
133-150% FPL	3.06 - 4.08
150-200% FPL	4.08 - 6.43
200-250% FPL	6.43 - 8.21
250-300% FPL	8.21 - 9.69
300-400% FPL	9.69

Source: Internal Revenue Services

An example to calculate the contribution amount has been presented below:

POVPIP = 140

Assigned Income Level Bracket Relative to FPL = 133% - 150%

Expected Contribution Level = 3.06% to 4.08%

$$Expected\ Contribution = Family\ Income * \frac{(140 - 133) * (4.08{-}3.06)}{100 * (150 - 133)} + 3.06$$

After computing the contribution amount, subsidy (Premium Tax Credit) has been

calculated using the following formula:

$$Subsidy = Cost\ of\ Second\ Lowest\ Silver\ Plan - Maximum\ Contribution\ (CAT)$$

In our analysis, the second lowest silver plan dataset for 2017 is imported and merged across the associated age groups to calculate subsidy. The second lowest silver plans have been added across the coverage family.

The average subsidy level across coverage families has been measured by AVGCOVERAGEFAM.

Step 4

A csv file has been exported containing the following variables:

SERIALNO, PUMA, STATE, AGE, RACE, FEMALE, POVPIP, SUBSIDY, AVGCOVERAGEFAM

Individual Level Analysis

Total Number of Individuals = 785

	Subsidy Level (USD)			
Demographic Groups	10th percentile	90th percentile	Mean	Std. Deviation
All Demographics (No Filter)	2,085.53	10,488.19	5,624.87	3,339.04
Female	2,149.76	10,842.93	5,728.25	3,302.94
Male	1,982.68	10,317.47	5,526.89	3,374.09
White	2,052.33	10,646.69	5,649.55	3,382.38
Other Races	2,397.54	9,359.05	5,332.03	2,780.40

| Statistical Output	Age Distribution
10th percentile	22
90th percentile	61
Mean	42
Std. Deviation	15

Family Level Analysis

Total Number of Families = 585

Average Coverage Family Size = 1.32

Statistical Output	Subsidy at Family Level (USD)
10th percentile	1,951.58
90th percentile	7,822.79
Mean	4,795.63
Std. Deviation	2,874.92

Subsidy Level Relative to Age Group

Subsidy Level by Age Group
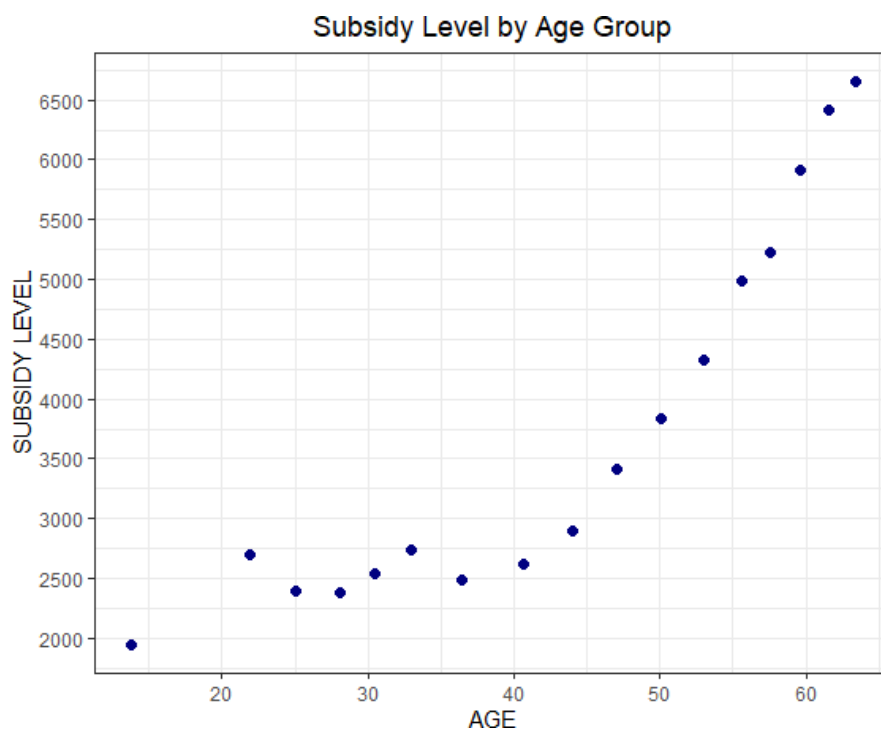
Regression Output between Subsidy Level and Age:

Variables	Coefficients	Std.Error	t-statistic	p-value
(Intercept)	-12.7615	114.7736	-0.11119	0.911495
AGEP	87.91867	2.581195	34.06123	1.05E-156

The subsidy level increases as people grow older. Moreover, the steepness of the slope rises as people reach their 50s. The t-statistic for age is at a statistically significant level.

Subsidy Level Relative to Income Level

Subsidy Level by Federal Poverty Line Income Level
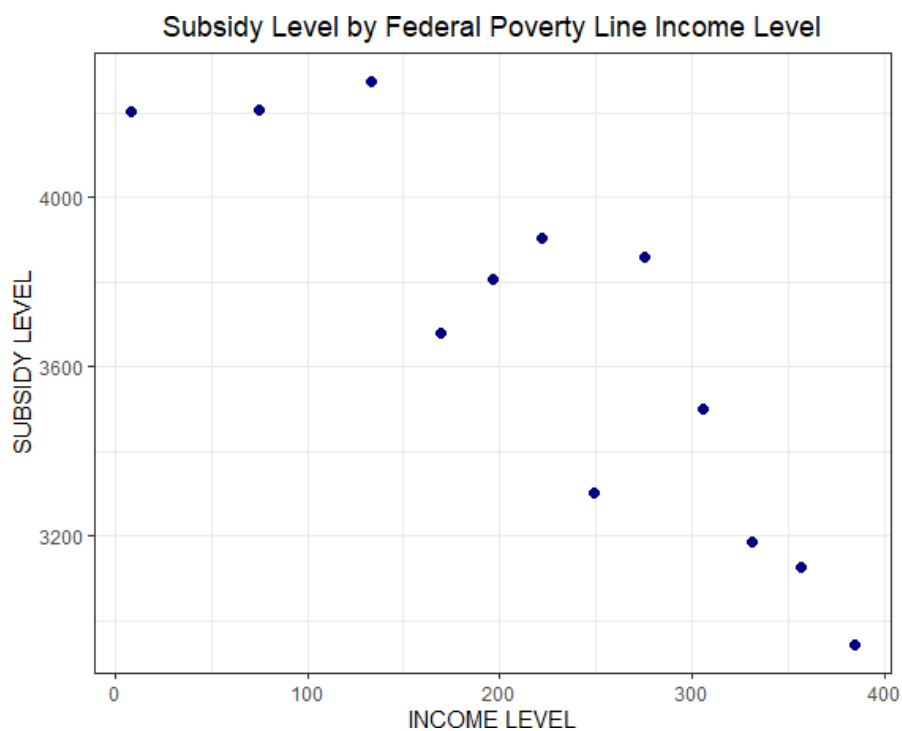
Regression Output between Subsidy Level and Income:

| Variables	Coefficients	Std.Error	t-statistic	p-value
(Intercept)	4451.756	133.6993	33.29678	3.92E-152
POVPIP	-3.49339	0.533383	-6.5495	1.04E-10

The subsidy has an inverse relationship with income. The subsidy level constantly decreases with the rise in income. The t-statistic for income is at a statistically significant level.

Subsidy Level by Sex

Subsidy Level by Sex
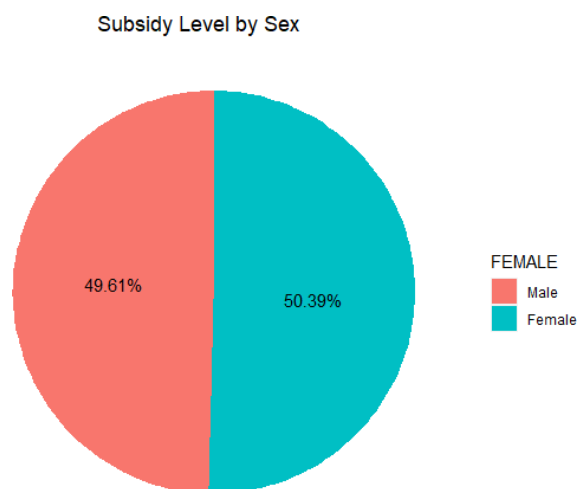
The subsidy level is equally distributed across both sexes. The regression output does not show statistically significant relation between sex and subsidy level.

Subsidy Level by Race

Subsidy Level by Race
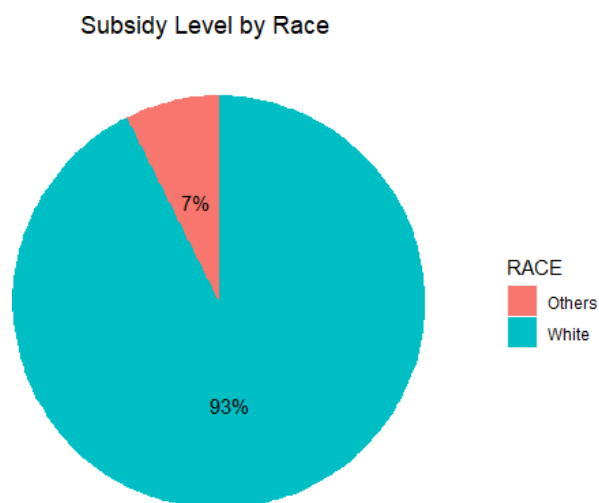
The white population covers 93% of the total subsidy amount than other races. In the year 2017, NH had a 93% White population. Comparing with this fact, subsidy seems to be equally distributed between white and other race groups. The regression output does not show statistically significant relation between race and subsidy level.

Families with Highest and Lowest 10% Subsidy Levels

We categorized the dataset into highest 10% and lowest 10% by subsidy levels.

The highest 10% subsidy population consists of an almost equal number of sexes. Most of the individuals are in their 50s and 60s. The median age is 58. The group is dominated by households with couples and couples with children. The average coverage family size is 2.73. The individual racial population is 97% White. The majority of families are of a single race composition.

In the lowest 10% subsidy level, the male population is slightly higher at 57%. Most of the individuals are in their 20s and 30s. The individuals are primarily single tax filers. The average coverage family size is 1.08. The individual racial composition is 93% White.

Regression Estimate

$$Y_{Subsidy} = \alpha + \beta_1 Age + \beta_2 FamilySize + \mu$$

It seems the subsidy amount is heavily dependent on the age group. People in their 50s and 60s are likely to receive higher subsidy amounts than other age groups. Moreover, the second-lowest silver plan steadily increases with the increment in age.

Family size can also play a vital role in determining the subsidy amount. An increment in the family size can add up to the second-lowest silver plan against the family contribution. The other demographic factors like race and sex do not seem to impact the subsidy levels significantly. Therefore, these variables have been omitted from the regression model.