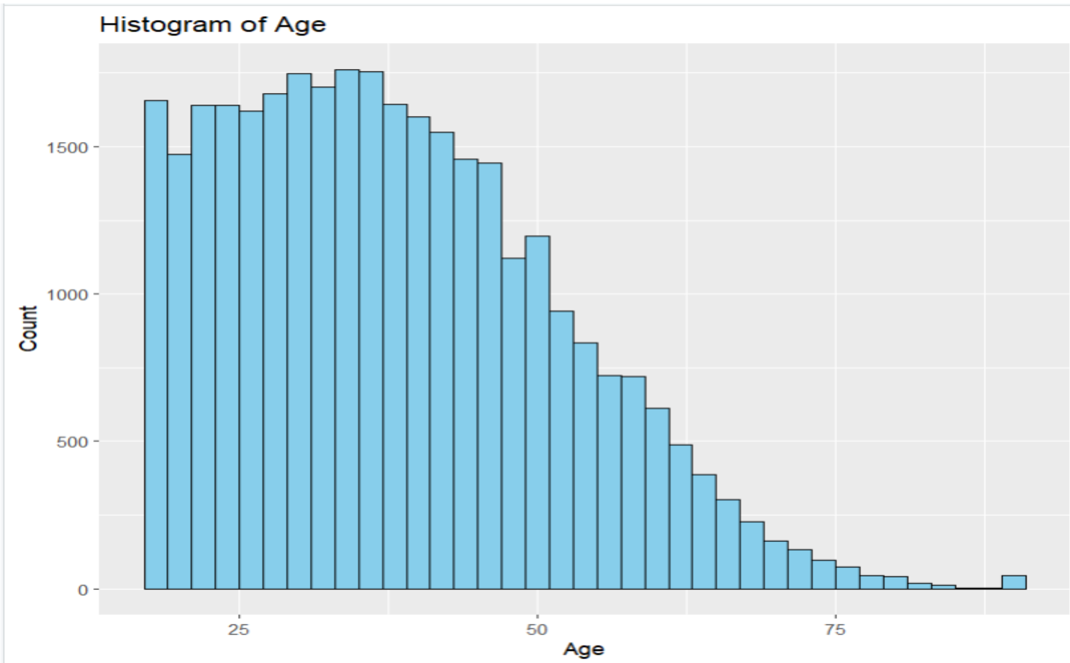
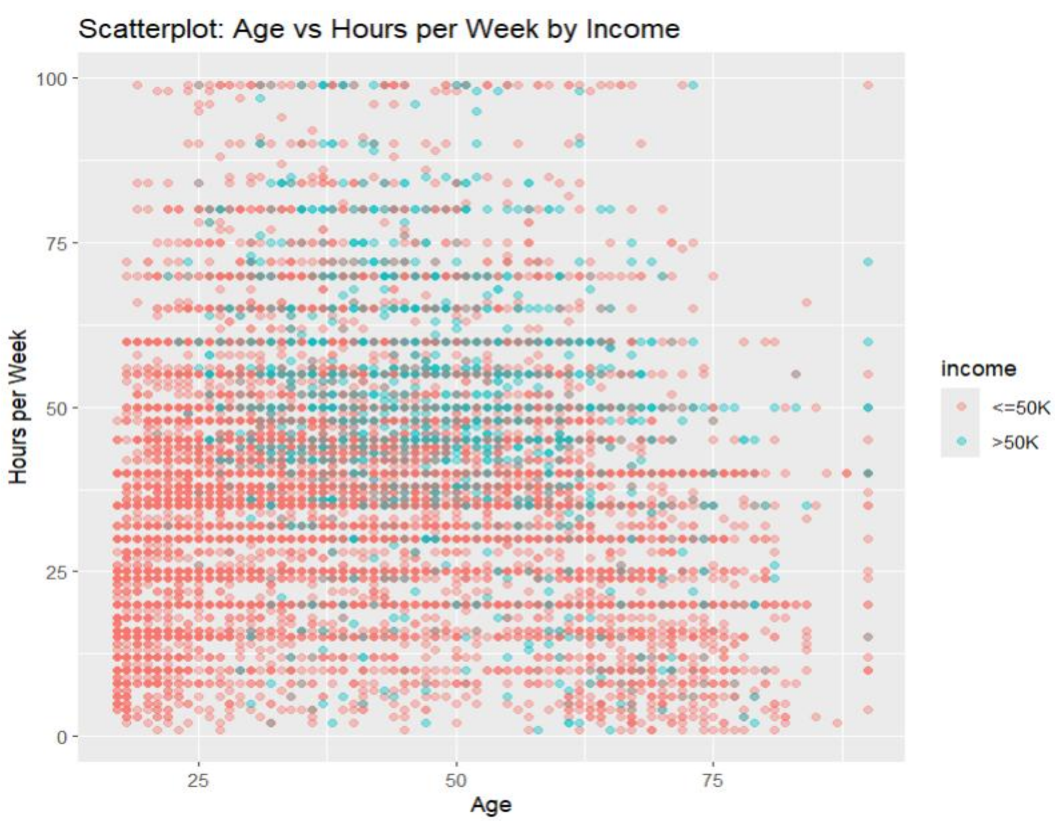
Task-1 Output:

Histogram of Age Column:

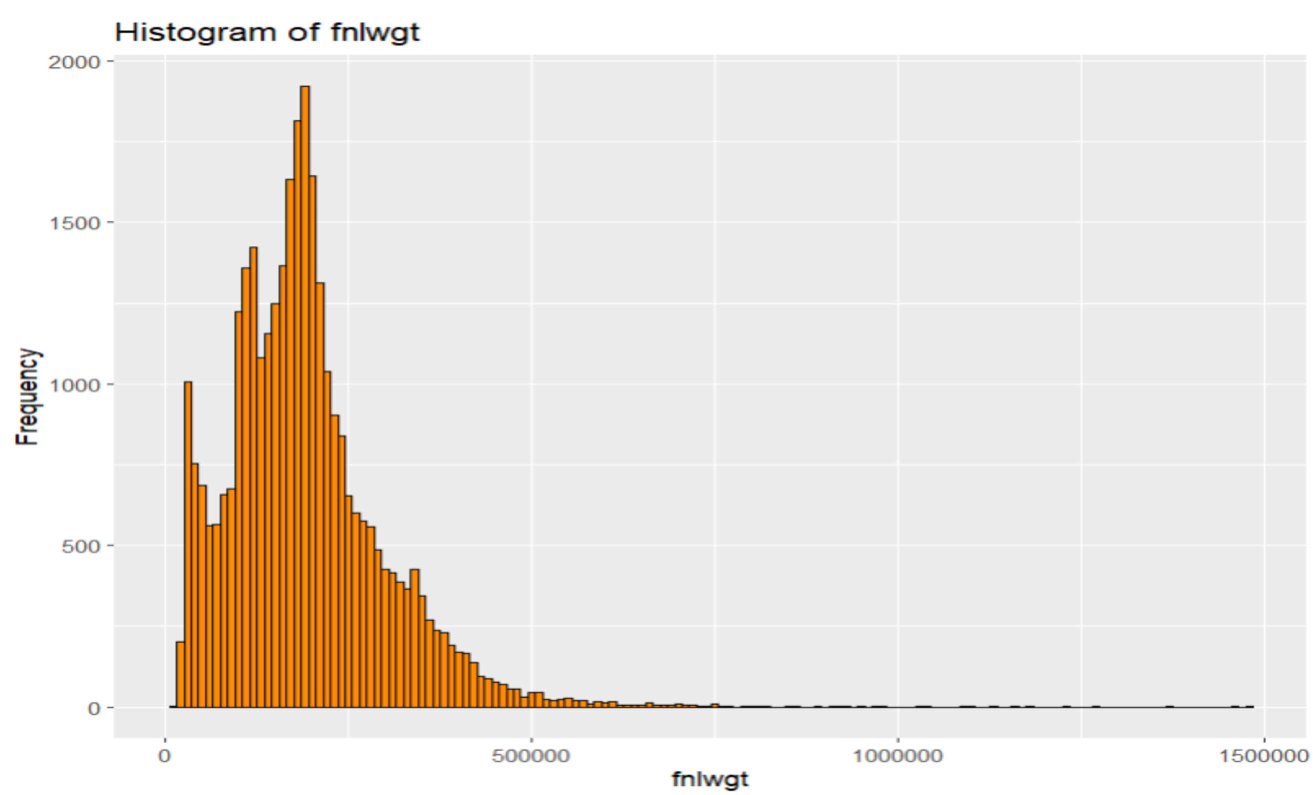


Scatterplot — Age vs Hours per Week, Colored by Income:

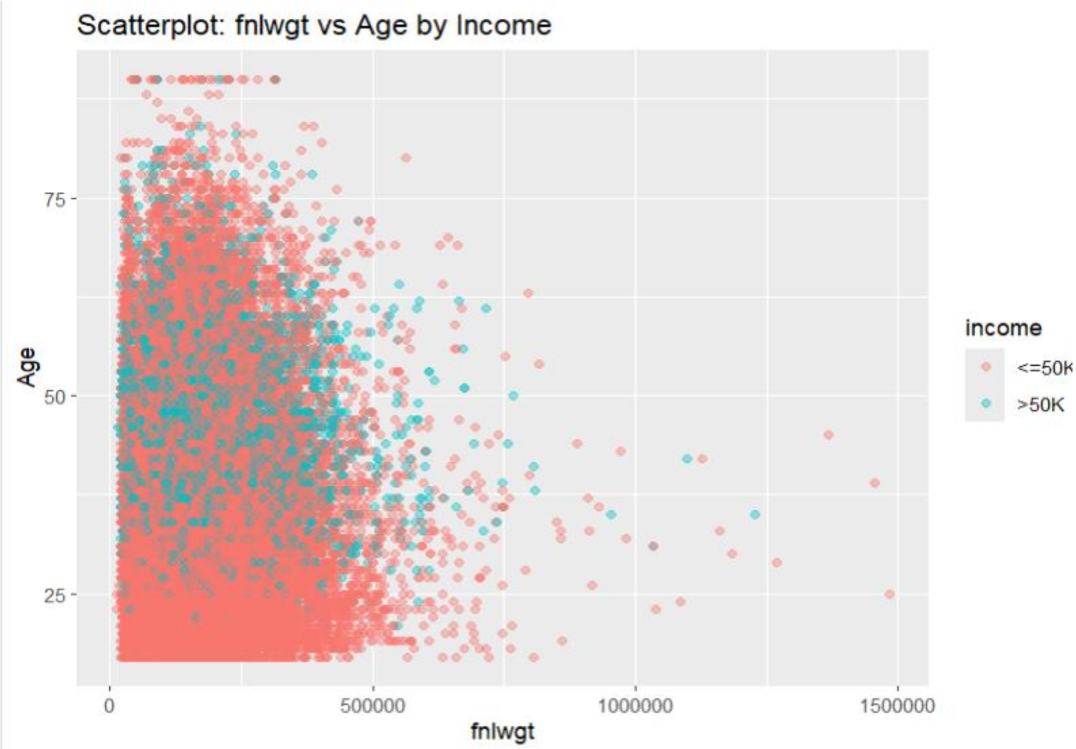


Violin Plot — Age by Income:

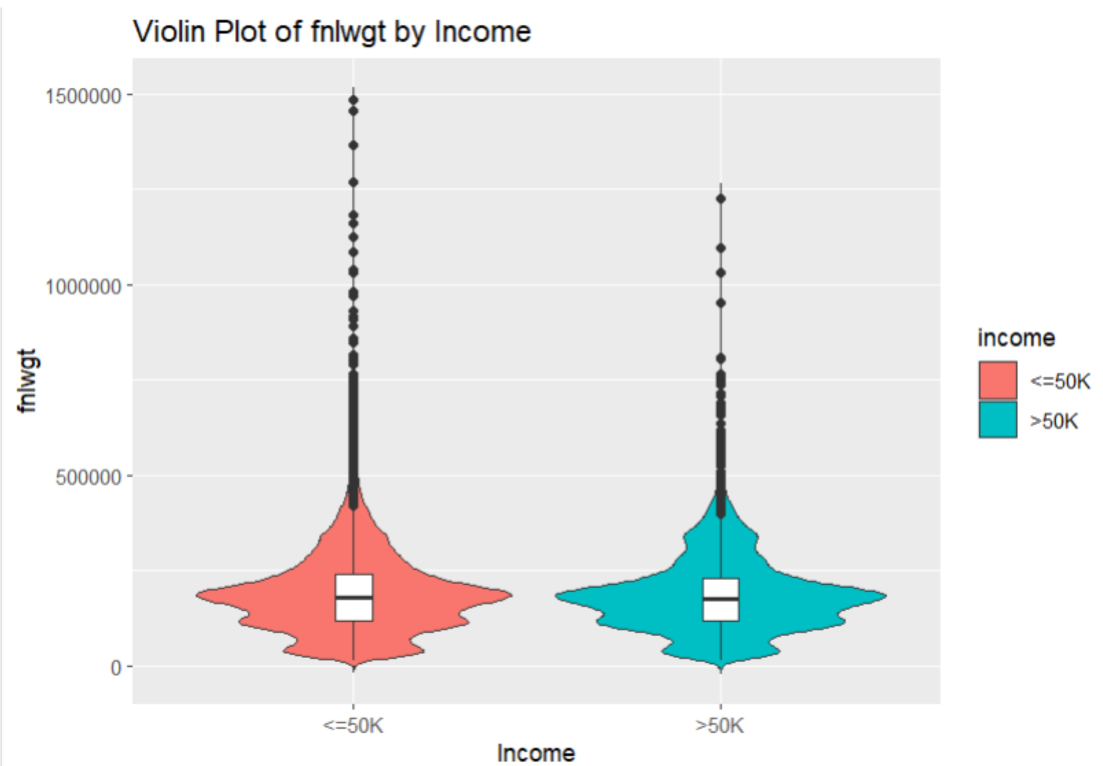
Histogram of fnlwgt:



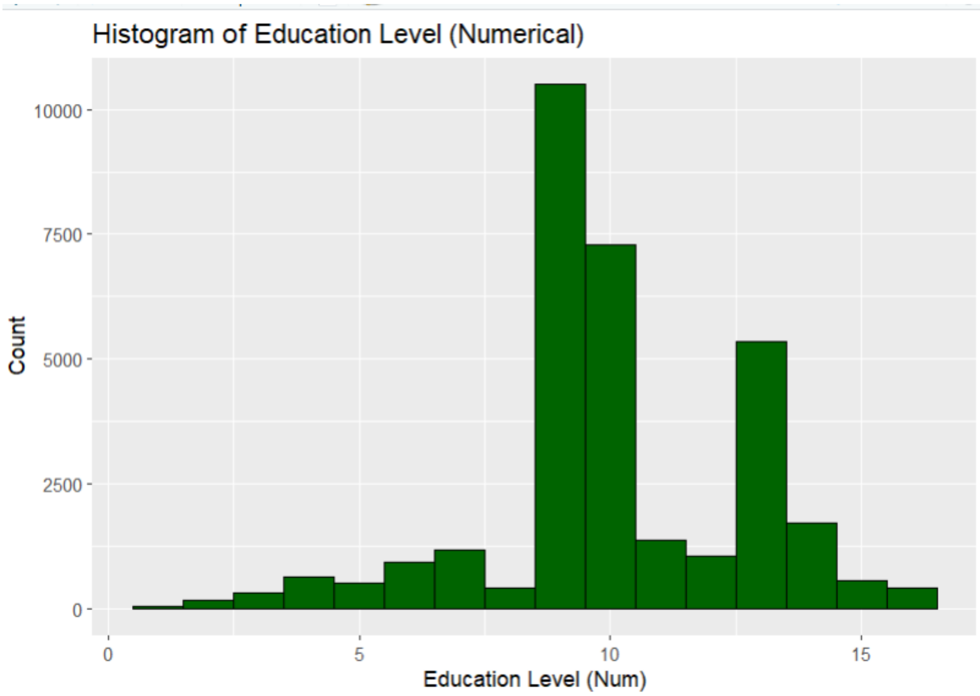
Scatterplot — fnlwgt vs Age, Colored by Income:



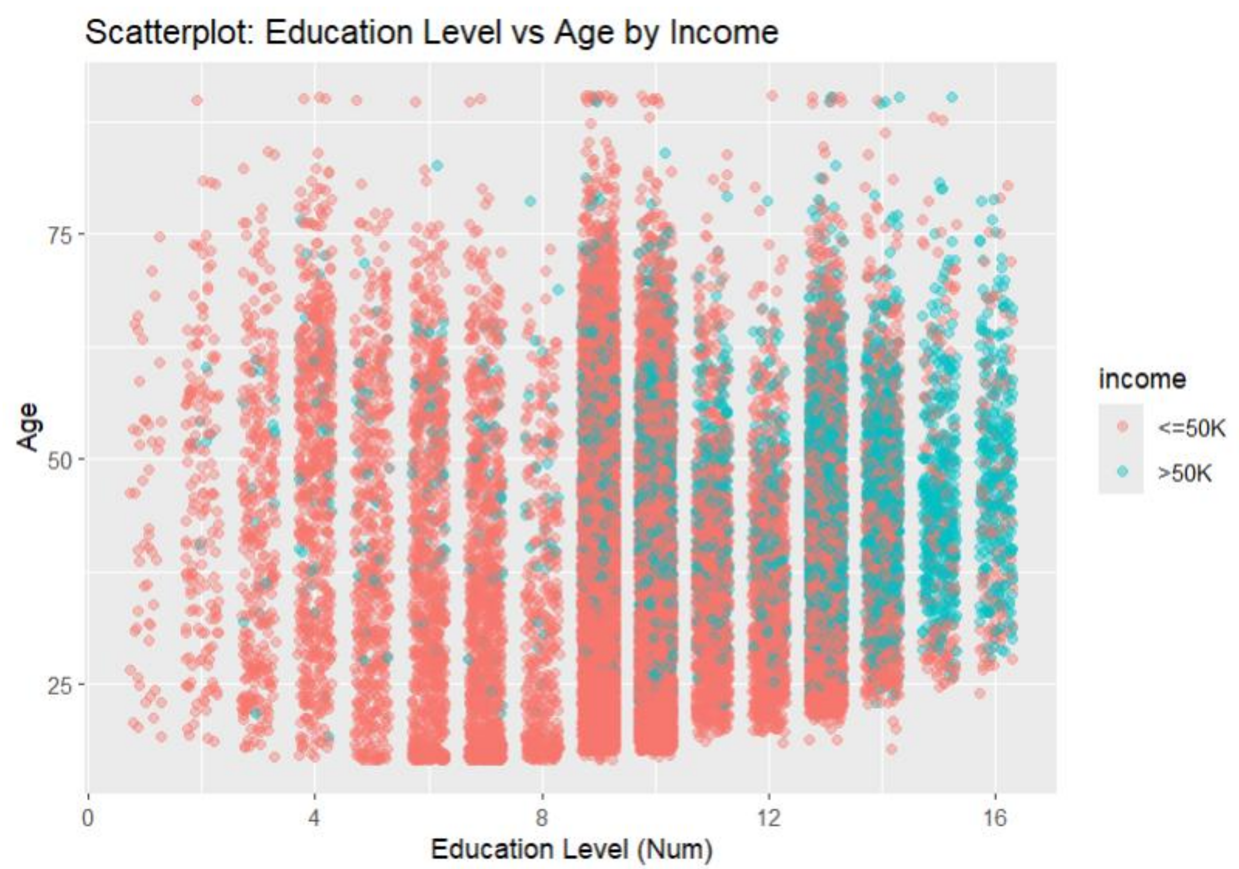
Violin Plot — fnlwgt by Income:



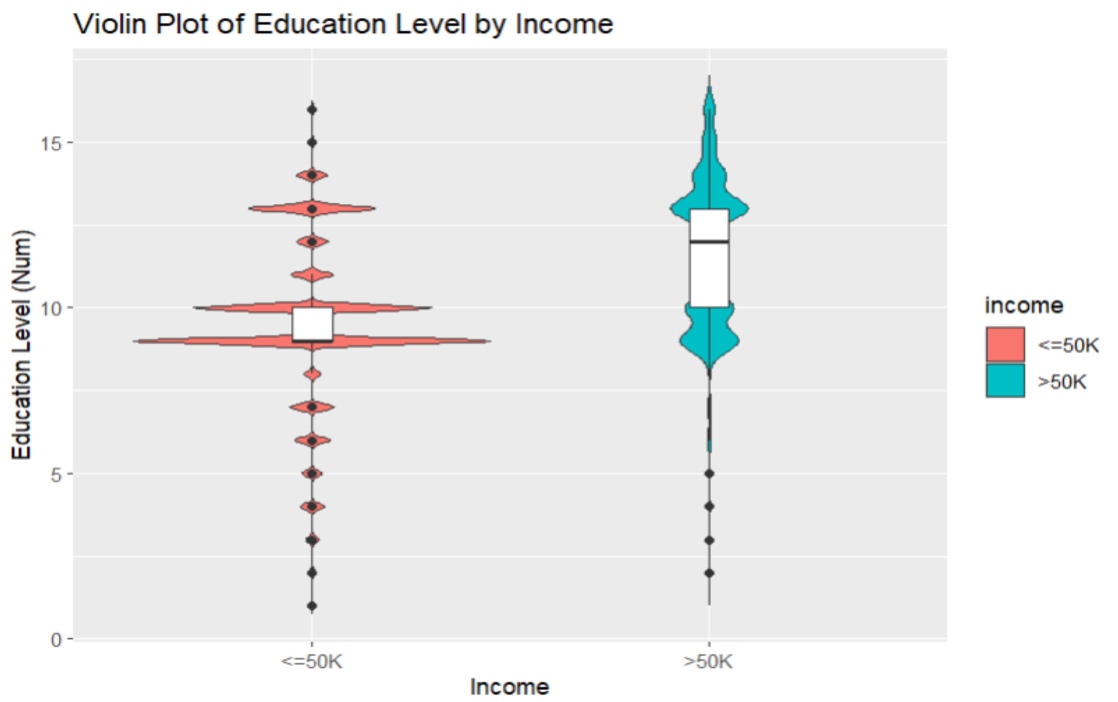
Histogram of education.num:



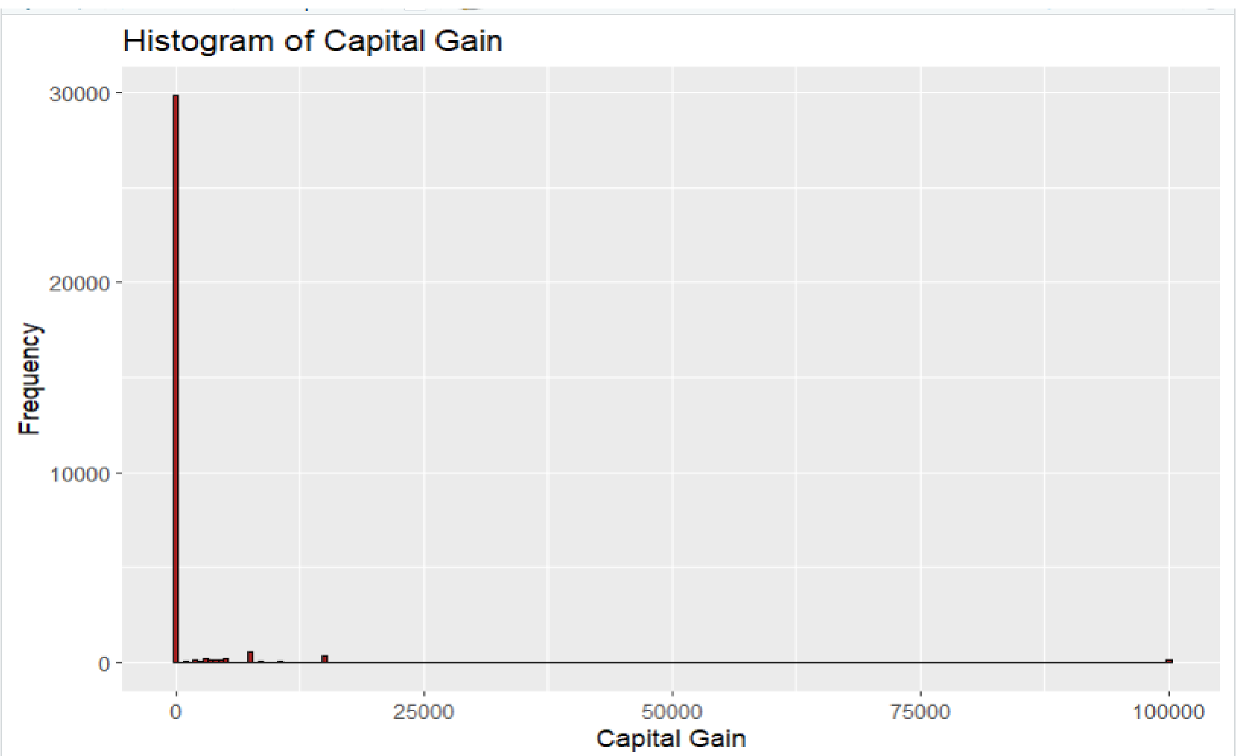
Scatterplot — education.num vs Age, Colored by Income:



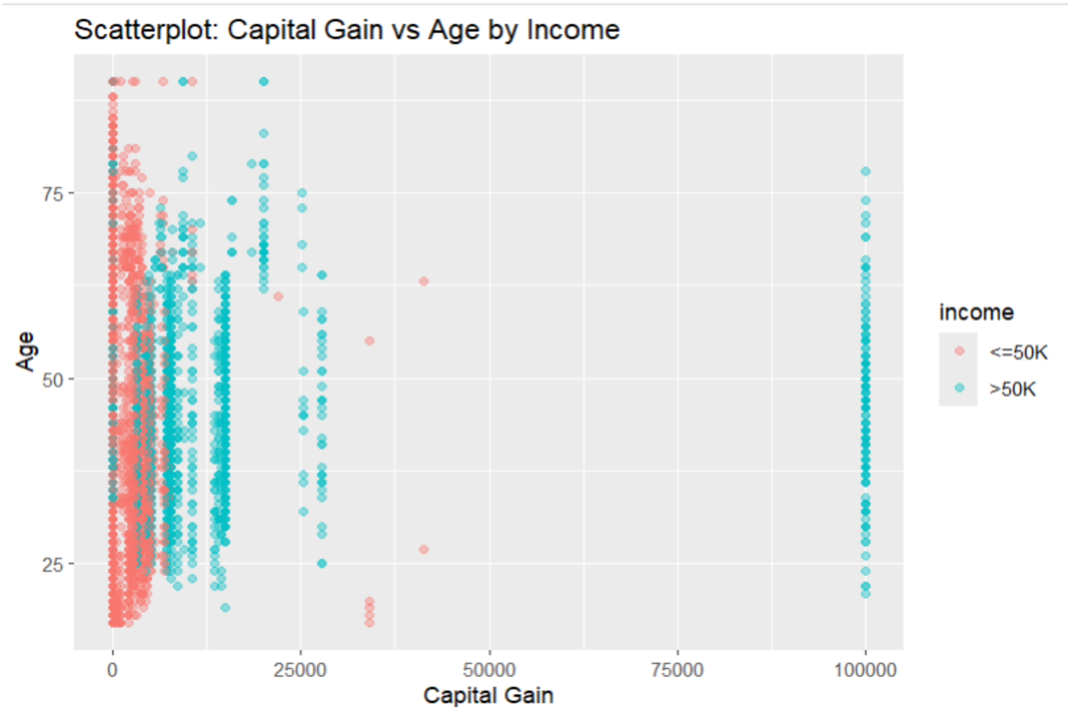
Violin Plot — education.num by Income:



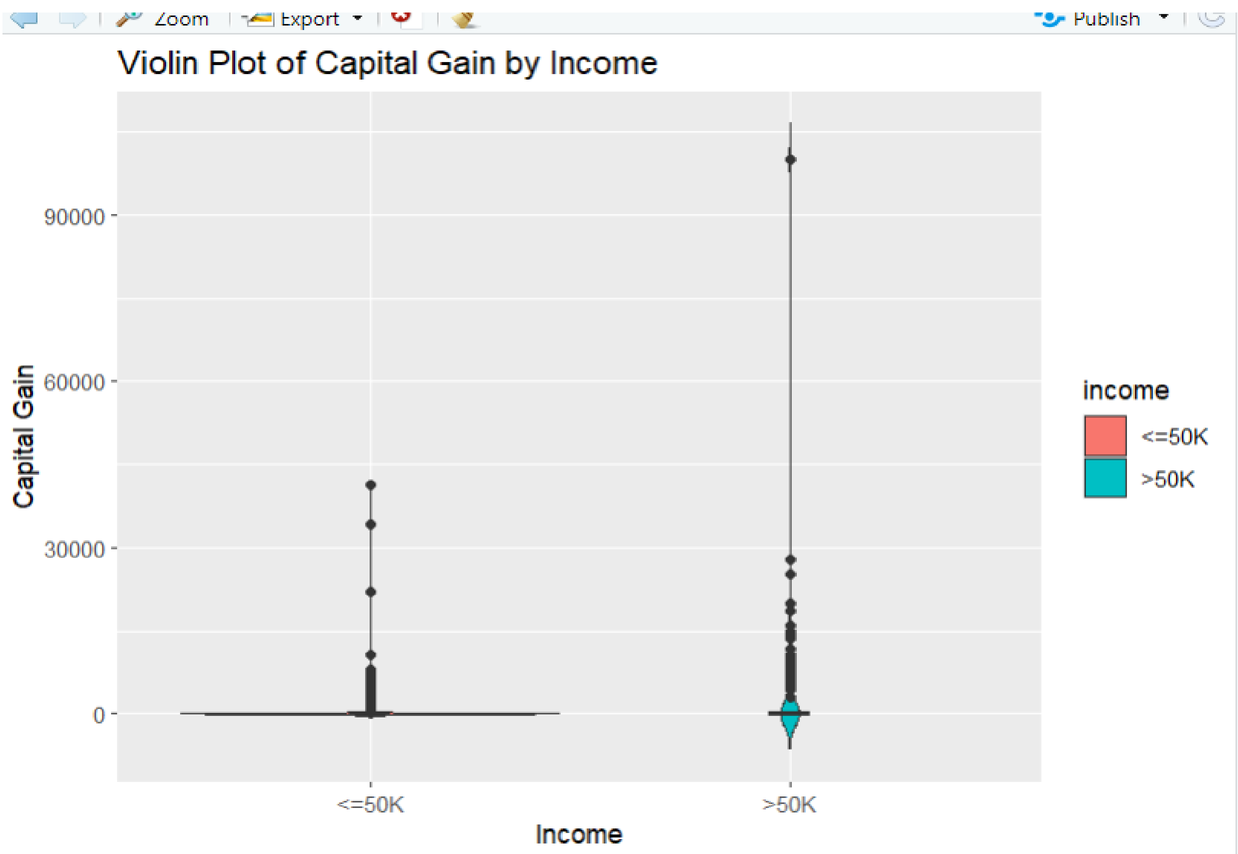
Histogram of Capital Gain:



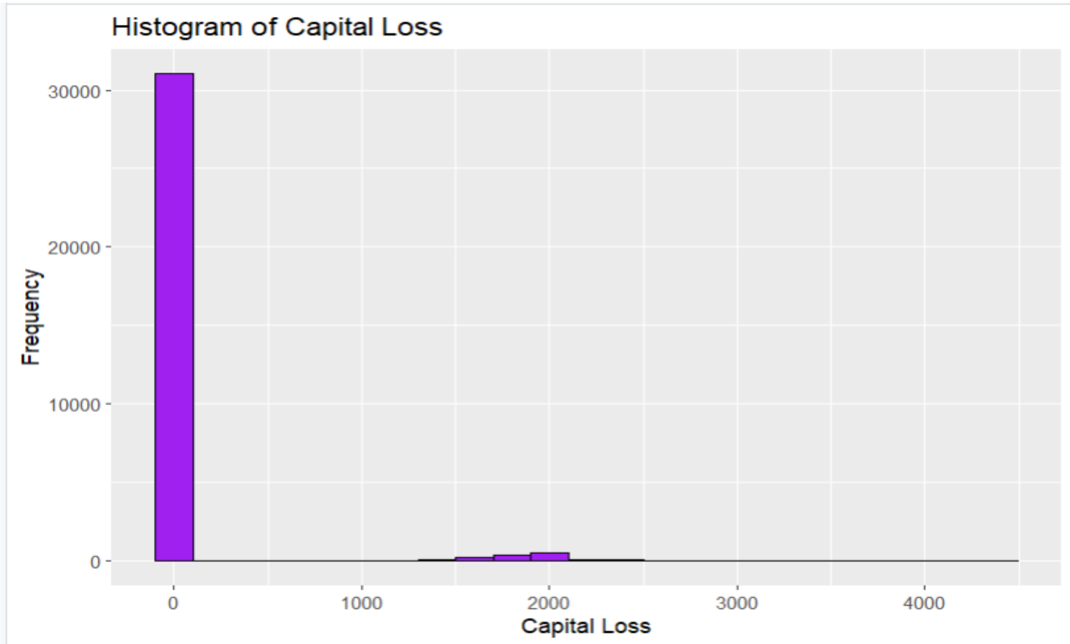
Scatterplot — Capital Gain vs Age by Income:



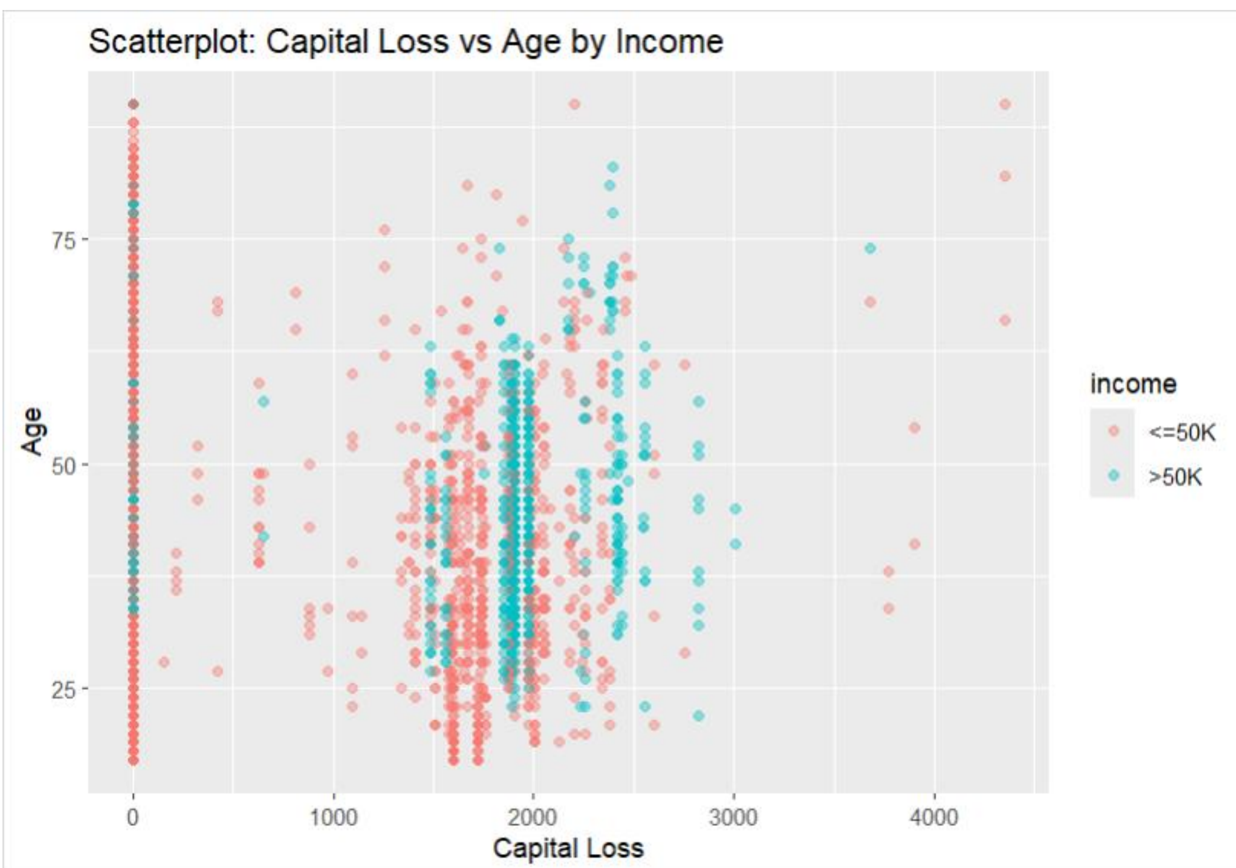
Violin Plot — Capital Gain by Income:



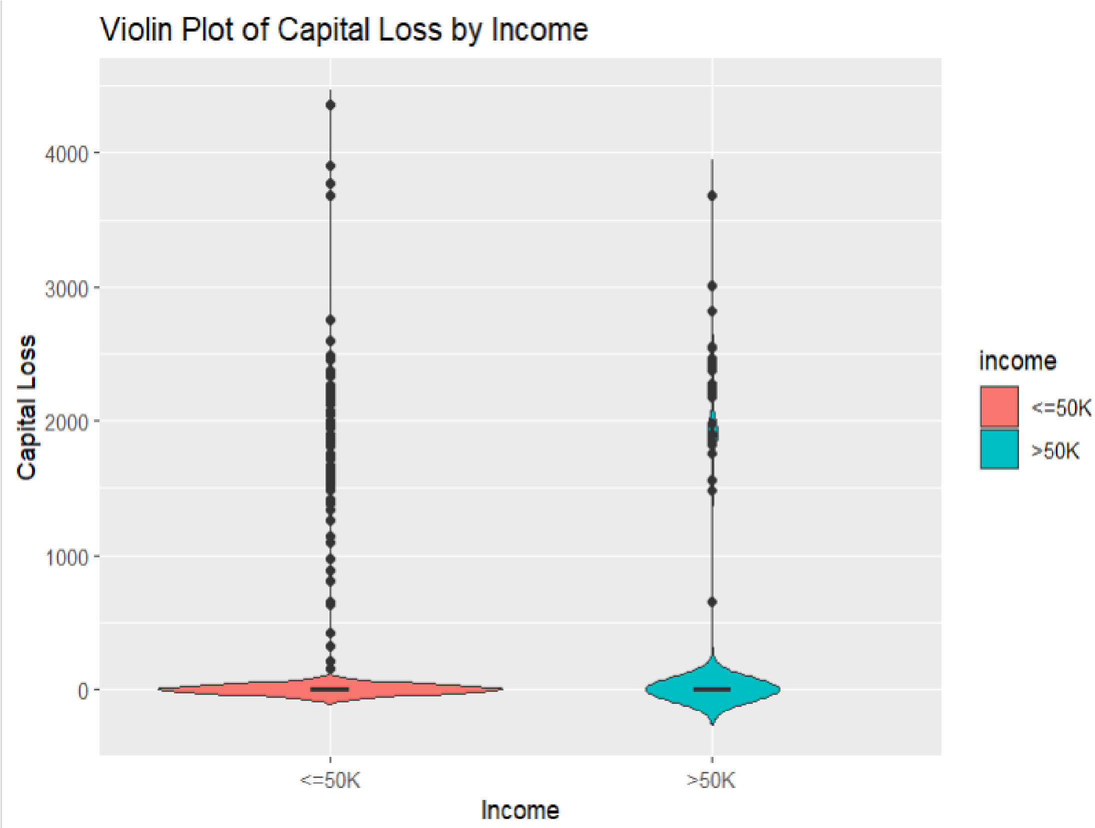
Histogram of Capital Loss:



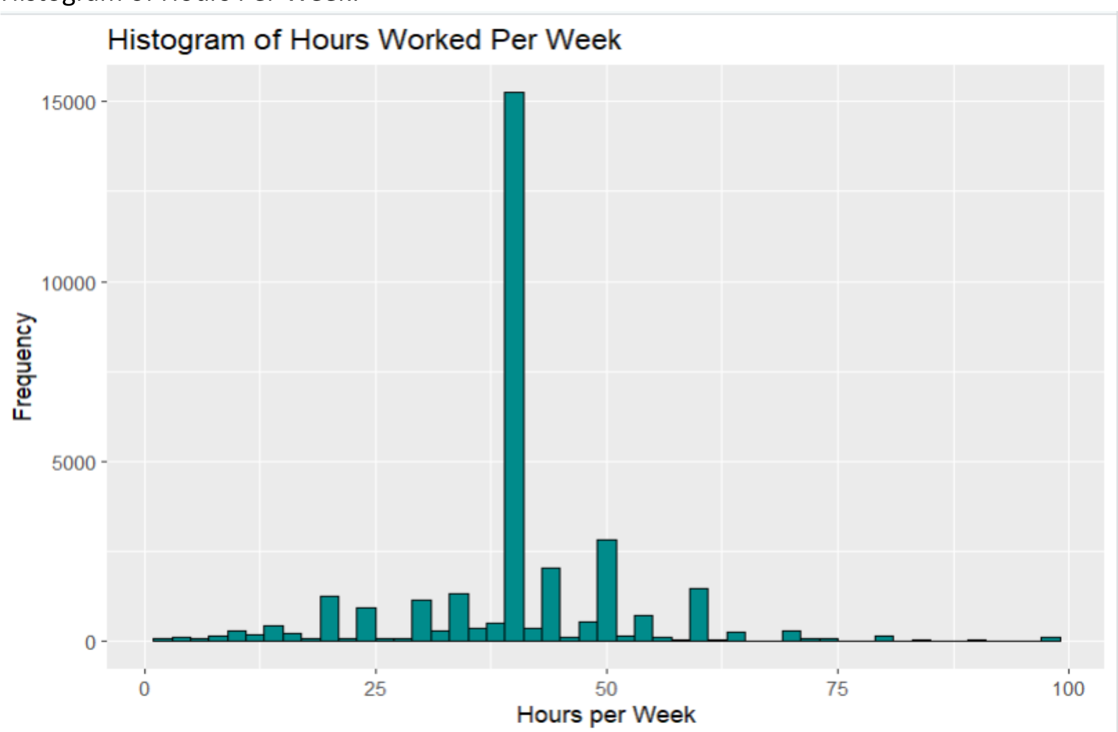
Scatterplot — Capital Loss vs Age by Income:



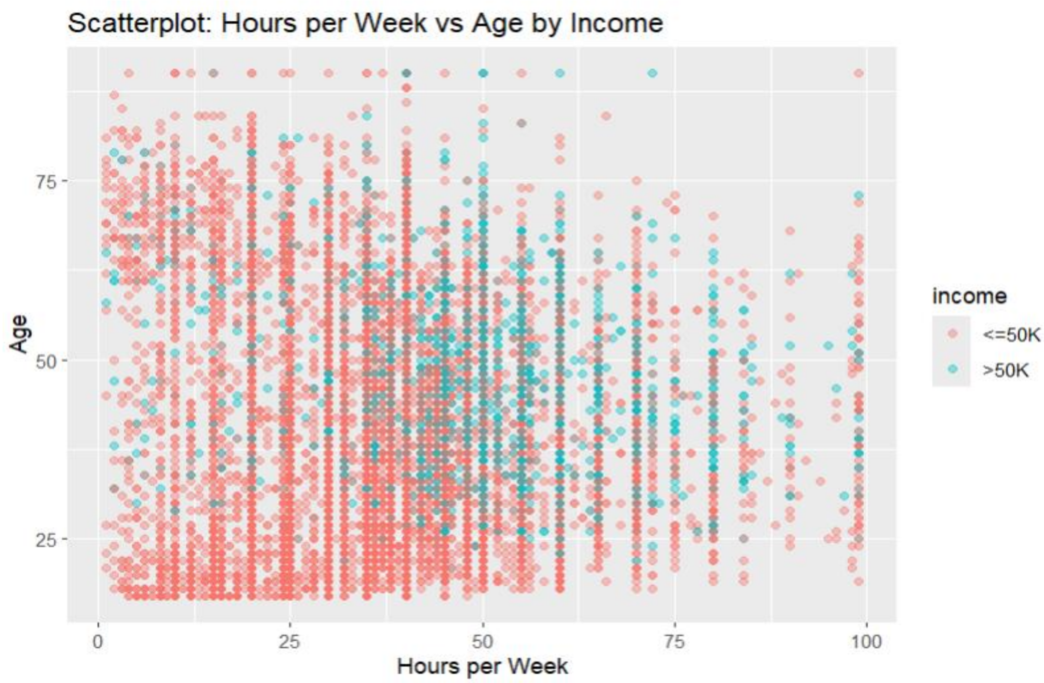
Violin Plot — Capital Loss by Income:



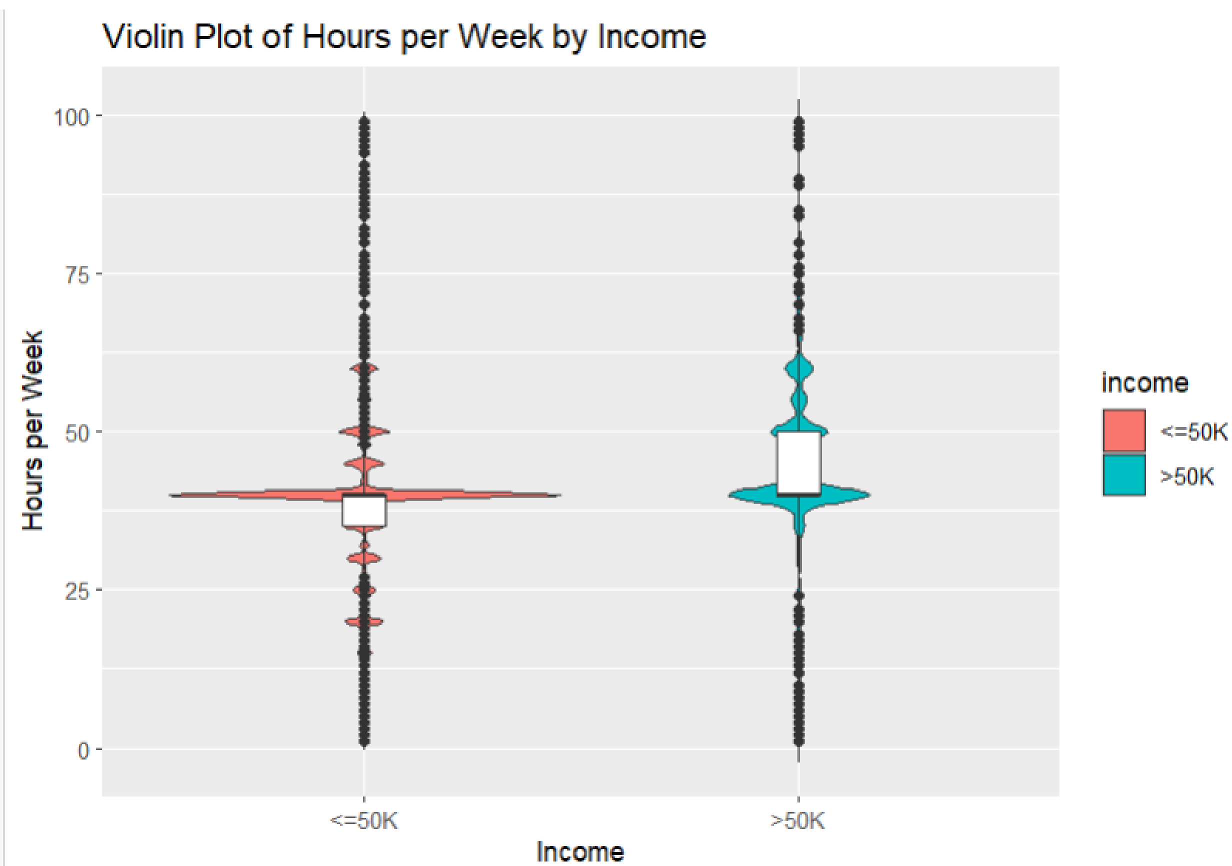
Histogram of Hours Per Week:



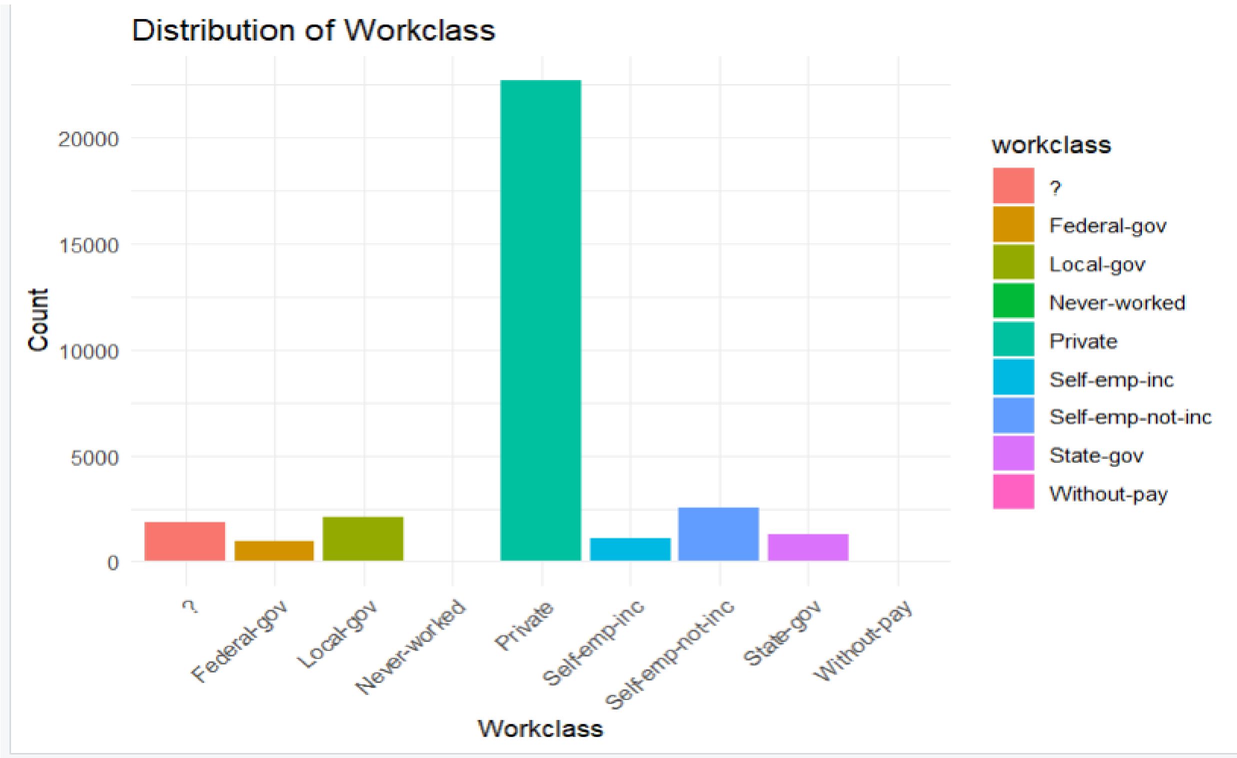
Scatterplot — Hours Per Week vs Age by Income:



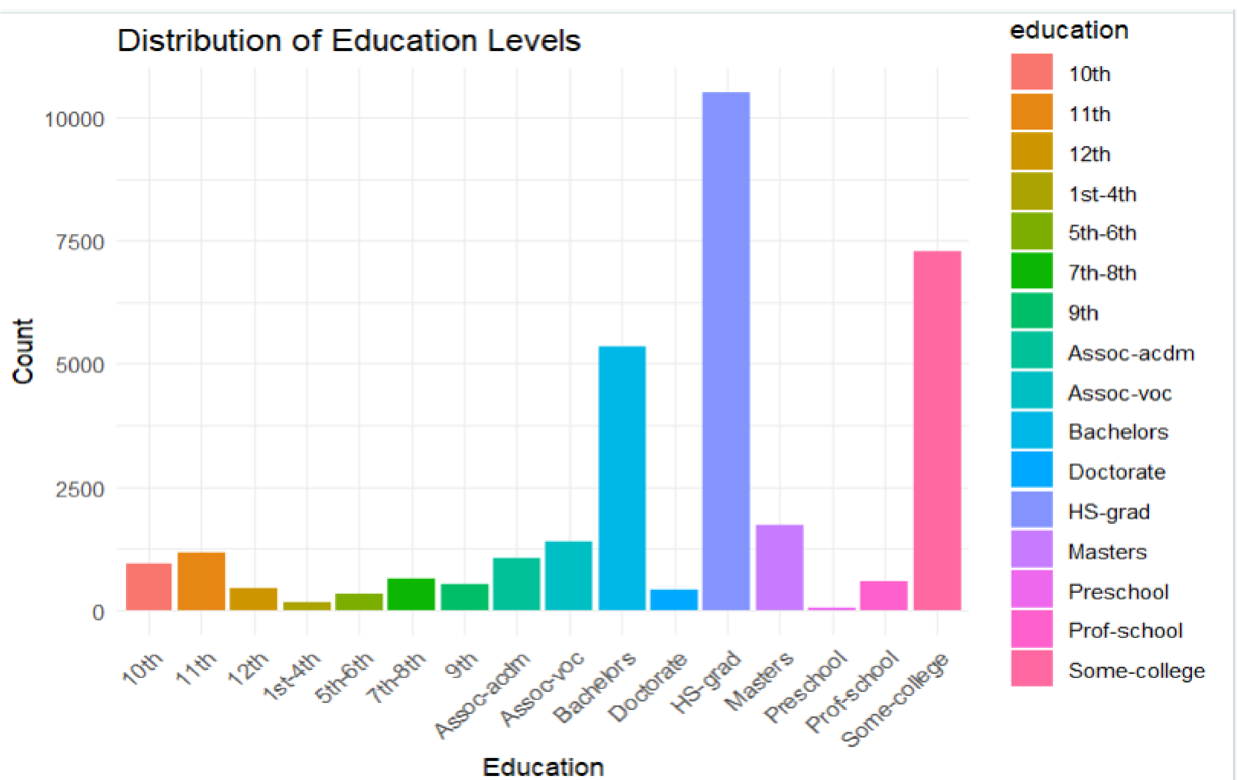
Violin Plot — Hours Per Week by Income:



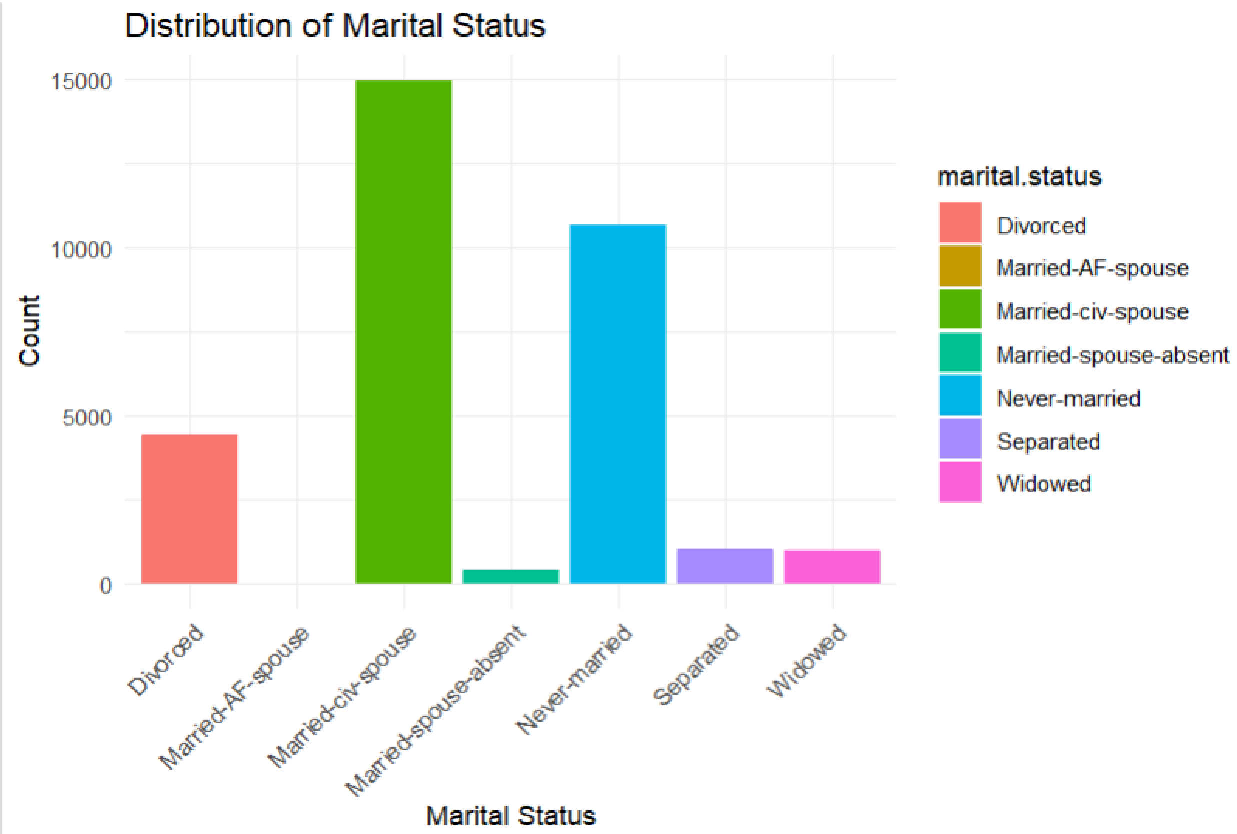
Bar Graph of Workclass:



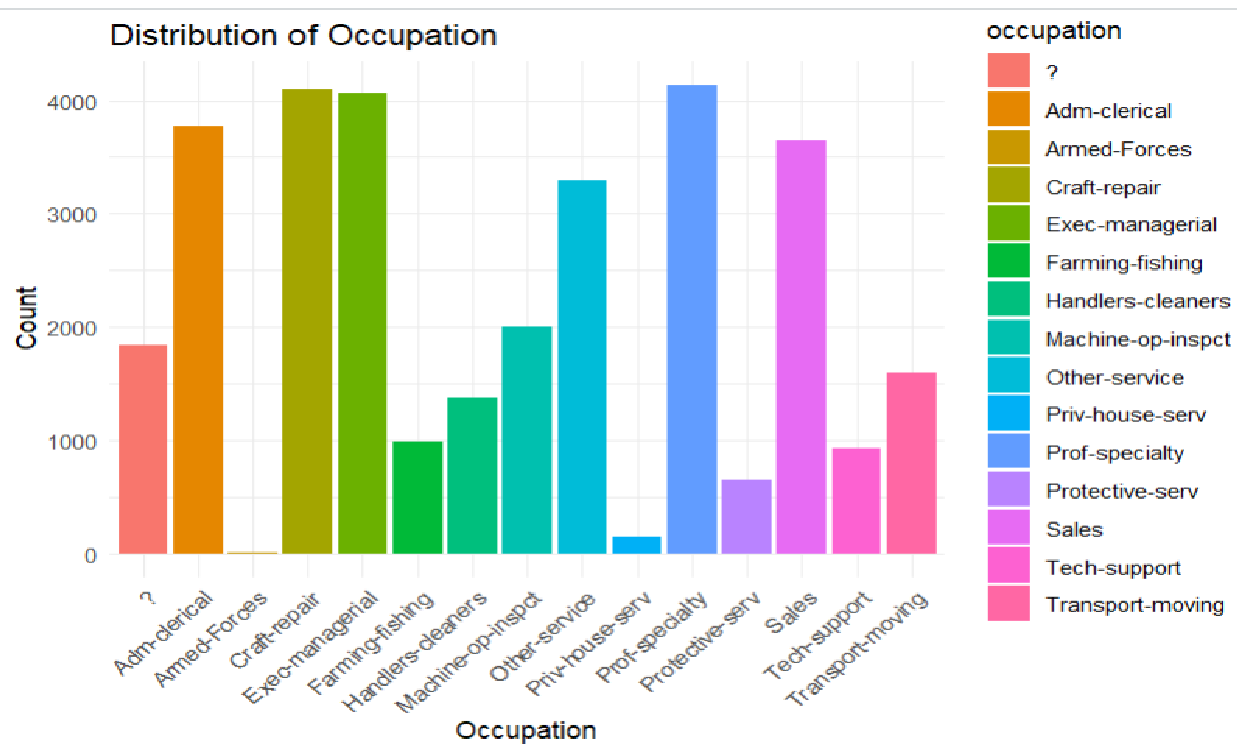
Bar Graph of Education:



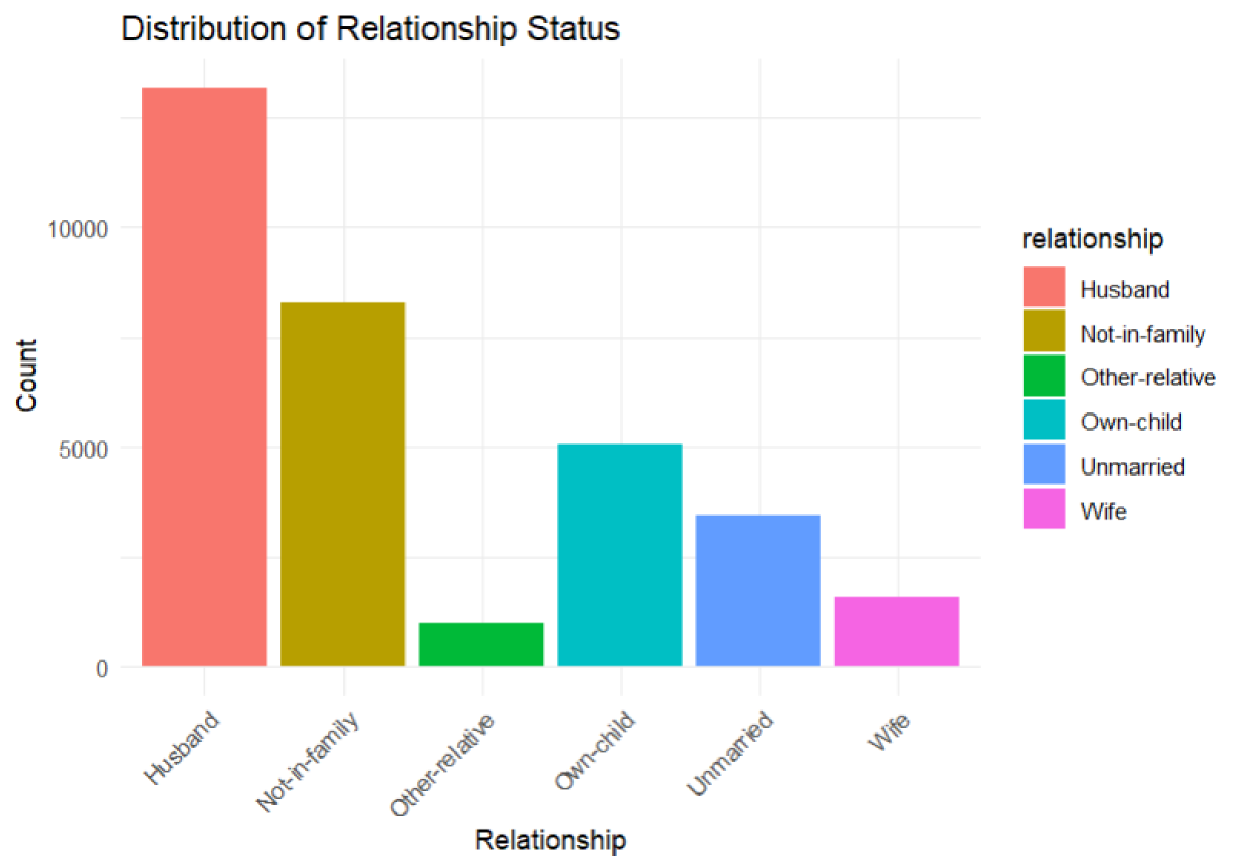
Bar Graph of Marital Status:



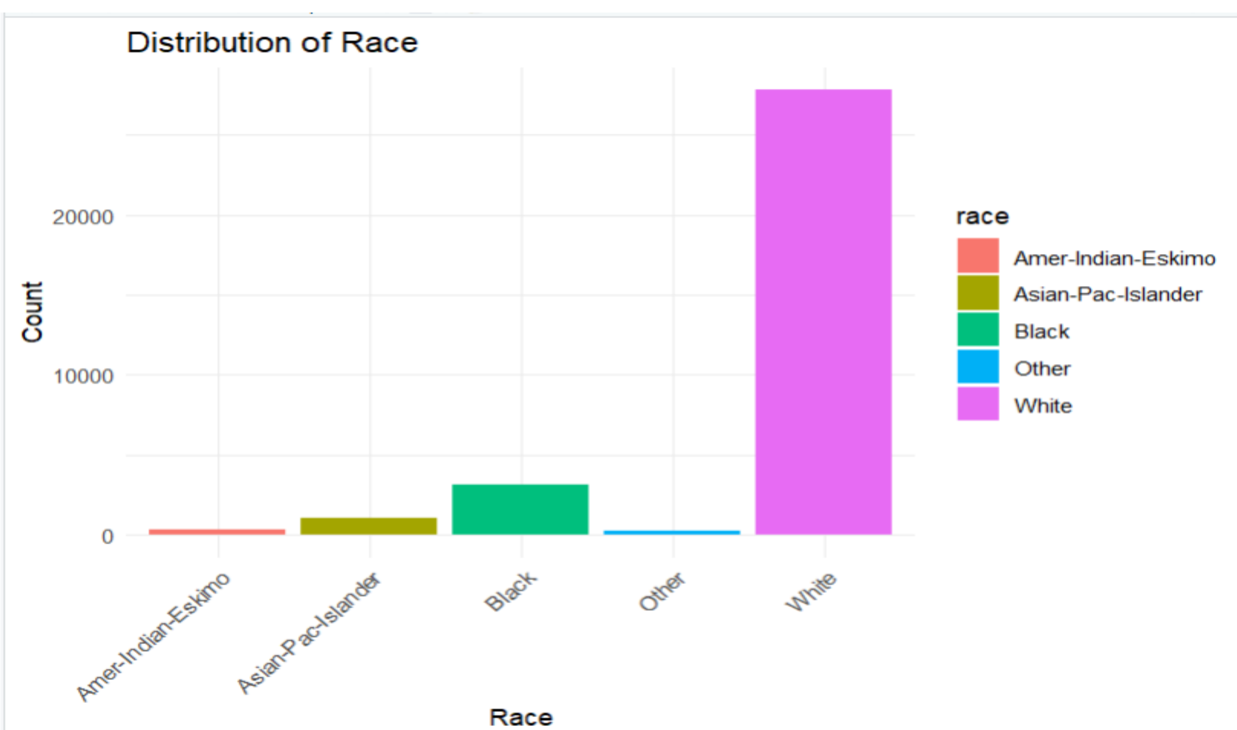
Bar Graph of Occupation:



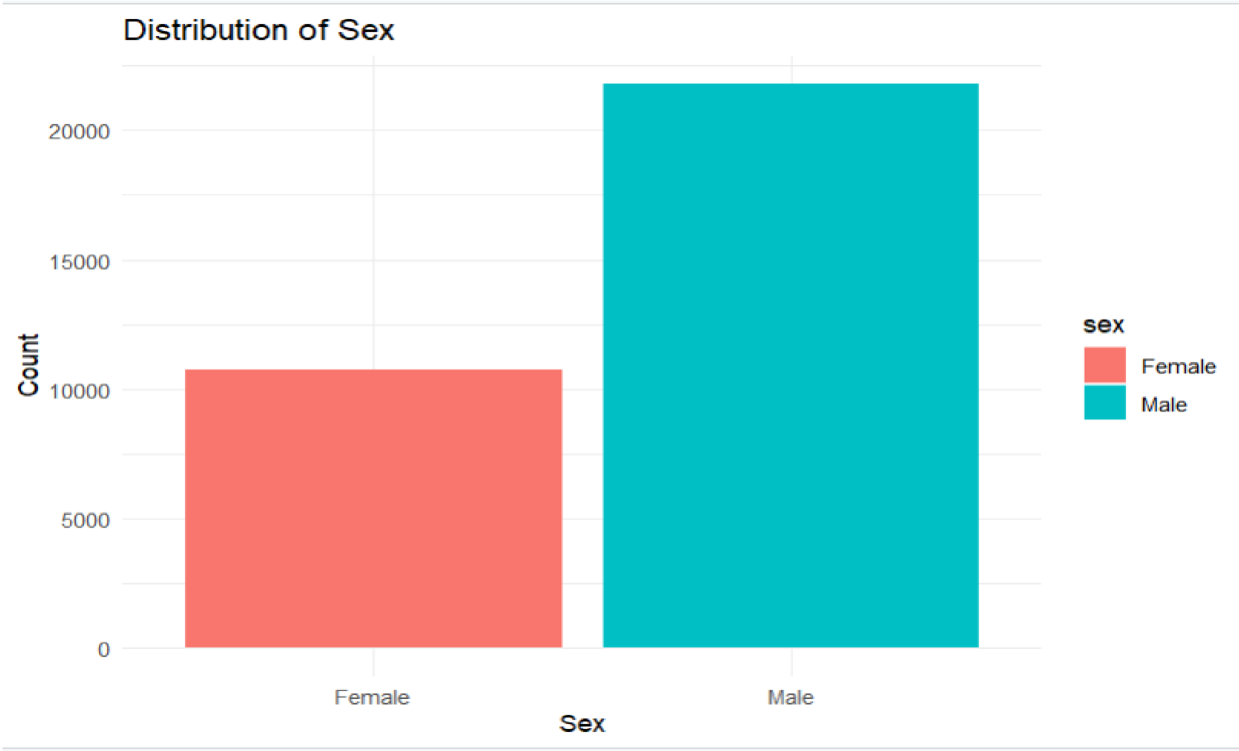
Bar Graph of Relationship:



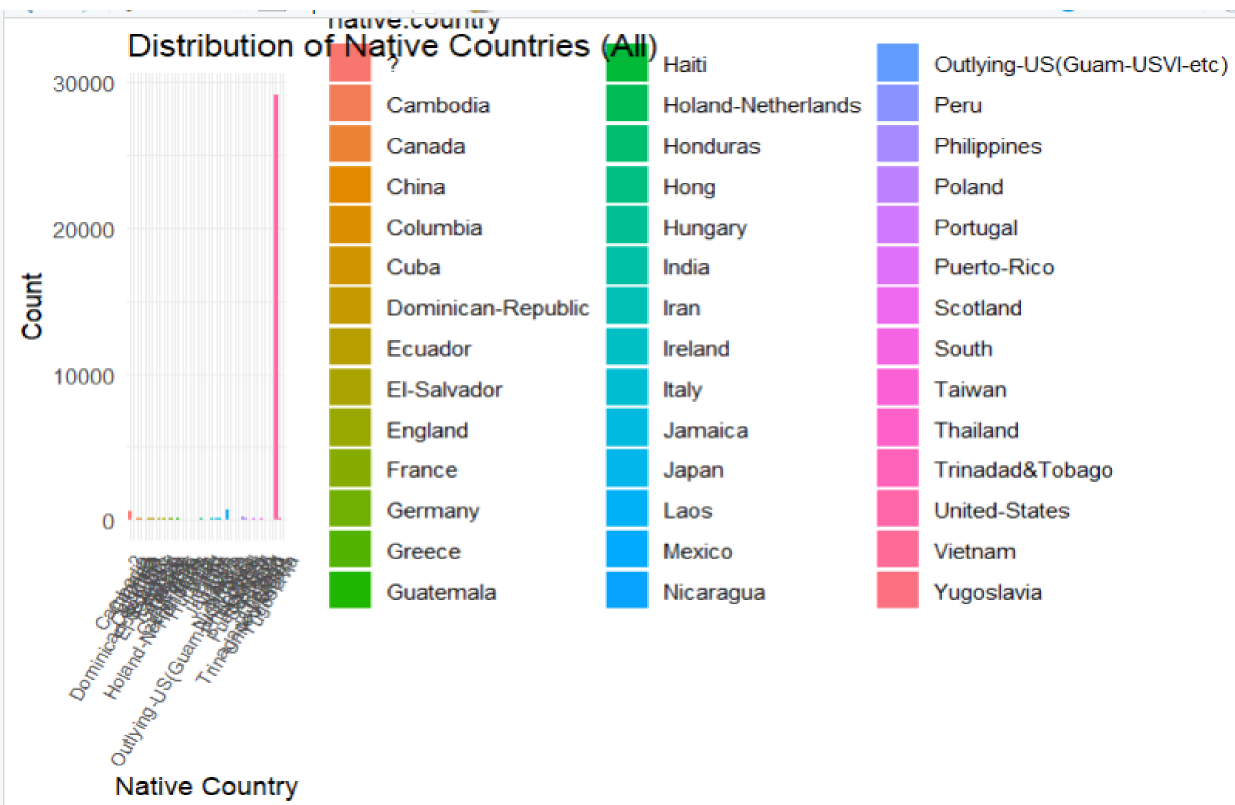
Bar Graph of Race:



Bar Graph of Sex:



Bar Graph of Native Countries:



Bar Graph of Income:

Task-2:

Sample output of dataset:

```
> head(adult_data)
  age workclass fnlwgt    education education.num marital.status        occupation  relationship
1  90         ?  77053      HS-grad             9        Widowed                 ? Not-in-family
2  82   Private 132870      HS-grad             9        Widowed   Exec-managerial Not-in-family
3  66         ? 186061 Some-college            10        Widowed                 ?     Unmarried
4  54   Private 140359      7th-8th             4       Divorced Machine-op-inspct     Unmarried
5  41   Private 264663 Some-college            10      Separated    Prof-specialty     Own-child
6  34   Private 216864      HS-grad             9       Divorced     Other-service     Unmarried
   race    sex capital.gain capital.loss hours.per.week native.country income
1 White Female            0         4356             40  United-States  <=50K
2 White Female            0         4356             18  United-States  <=50K
3 Black Female            0         4356             40  United-States  <=50K
4 White Female            0         3900             40  United-States  <=50K
5 White Female            0         3900             40  United-States  <=50K
6 White Female            0         3770             45  United-States  <=50K
>
```

Apply ANOVA and Kandall's Rank Correlation Tau for all numeric column:

For Age Column,

ANOVA:

```
> anova_result <- aov(age ~ income, data = adult_data)
> summary(anova_result)
               Df  Sum Sq Mean Sq F value Pr(>F)    
income          1  331826  331826    1887 <2e-16 ***
Residuals   32559 5726333     176                   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Kandall's Rank Correlation:

```
> cor.test(adult_data$age, adult_data$income_binary, method = "kendall")

        Kendall's rank correlation tau

data:  adult_data$age and adult_data$income_binary
z = 49.254, p-value < 2.2e-16
alternative hypothesis: true tau is not equal to 0
sample estimates:
      tau 
0.2252346 
```

For Age Fnlwgt,

Anova:

```
> anova_fnlwgt <- aov(fnlwgt ~ income, data = adult_data)
> summary(anova_fnlwgt)
               Df    Sum Sq   Mean Sq F value Pr(>F)  
income          1 3.248e+10 3.248e+10   2.916 0.0877 .
Residuals   32559 3.627e+14 1.114e+10                 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Kandall's Rank Correlation:

```
> cor.test(adult_data$fnlwgt, adult_data$income_binary, method = "kendall")

	Kendall's rank correlation tau

data:  adult_data$fnlwgt and adult_data$income_binary
z = -1.9375, p-value = 0.05268
alternative hypothesis: true tau is not equal to 0
sample estimates:
         tau 
-0.008767448 
```

For Age Capital.loss,

Anova:

```
> anova_loss <- aov(capital.loss ~ income, data = adult_data)
> summary(anova_loss)
               Df    Sum Sq   Mean Sq F value Pr(>F)    
income          1 1.198e+08 119793591   754.8 <2e-16 ***
Residuals   32559 5.167e+09    158703                   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Kandall's Rank Correlation:

```
> cor.test(adult_data$capital.loss, adult_data$income_binary, method = "kendall")

	Kendall's rank correlation tau

data:  adult_data$capital.loss and adult_data$income_binary
z = 25.45, p-value < 2.2e-16
alternative hypothesis: true tau is not equal to 0
sample estimates:
      tau 
0.1394871 
```

For Age hours.per.week,

Anova:

```
> anova_hours <- aov(hours.per.week ~ income, data = adult_data)
> summary(anova_hours)
               Df  Sum Sq Mean Sq F value Pr(>F)    
income          1  261890  261890    1813 <2e-16 ***
Residuals   32559 4702175     144                   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Kandall's Rank Correlation:

```
> cor.test(adult_data$hours.per.week, adult_data$income_binary, method = "kendall")

        Kendall's rank correlation tau

data:  adult_data$hours.per.week and adult_data$income_binary
z = 48.553, p-value < 2.2e-16
alternative hypothesis: true tau is not equal to 0
sample estimates:
      tau
0.2382733
```

Apply Chi-Squared and Mutual Information for all numeric column:

For workclass:

Chi-Squared:

```
> tbl_workclass <- table(adult_data$workclass, adult_data$income)
> chisq.test(tbl_workclass, simulate.p.value = TRUE, B = 10000)

        Pearson's Chi-squared test with simulated p-value (based on 10000 replicates)

data:  tbl_workclass
X-squared = 1045.7, df = NA, p-value = 9.999e-05
```

Mutual Information:

```
> adult_data$workclass <- as.factor(adult_data$workclass)
> adult_data$income <- as.factor(adult_data$income)
>
> information_gain(income ~ workclass, data = adult_data)
  attributes importance
1  workclass 0.01495229
>
```

For education:

Chi-Squared:

```
> tbl_education <- table(adult_data$education, adult_data$income)
> chisq.test(tbl_education)

        Pearson's Chi-squared test

data:  tbl_education
X-squared = 4429.7, df = 15, p-value < 2.2e-16
```

Mutual Information:

```
> adult_data$education <- as.factor(adult_data$education)
> adult_data$income <- as.factor(adult_data$income)
>
> information_gain(income ~ education, data = adult_data)
  attributes importance
1  education 0.06487223
```

For marital.status:

Chi-Squared:

```
> tbl_marital <- table(adult_data$marital.status, adult_data$income)
> chisq.test(tbl_marital)

	Pearson's Chi-squared test

data:  tbl_marital
X-squared = 6517.7, df = 6, p-value < 2.2e-16
```

Mutual Information:

```
> adult_data$marital.status <- as.factor(adult_data$marital.status)
> adult_data$income <- as.factor(adult_data$income)
> information_gain(income ~ marital.status, data = adult_data)
      attributes importance
1 marital.status  0.1084968
```

For relationship:

Chi-Squared:

```
> tbl_relationship <- table(adult_data$relationship, adult_data$income)
> chisq.test(tbl_relationship)

	Pearson's Chi-squared test

data:  tbl_relationship
X-squared = 6699.1, df = 5, p-value < 2.2e-16
```

Mutual Information:

```
> adult_data$relationship <- as.factor(adult_data$relationship)
> adult_data$income <- as.factor(adult_data$income)
>
> information_gain(income ~ relationship, data = adult_data)
    attributes importance
1 relationship  0.1146228
```

For native.country:

Chi-Squared:

```
> tbl_country <- table(adult_data$`native.country`, adult_data$income)
> chisq.test(tbl_country, simulate.p.value = TRUE, B = 10000)

	Pearson's Chi-squared test with simulated p-value (based on 10000 replicates)

data:  tbl_country
X-squared = 317.23, df = NA, p-value = 9.999e-05
```

Mutual Information:

```
> adult_data$`native.country` <- as.factor(adult_data$`native.country`)
> adult_data$income <- as.factor(adult_data$income)
>
> information_gain(income ~ `native.country`, data = adult_data)
      attributes  importance
1 native.country 0.006027152
>
```