

Project 1

Calvin Hoang

1/28/2023

Heart Disease Dataset

Introduction

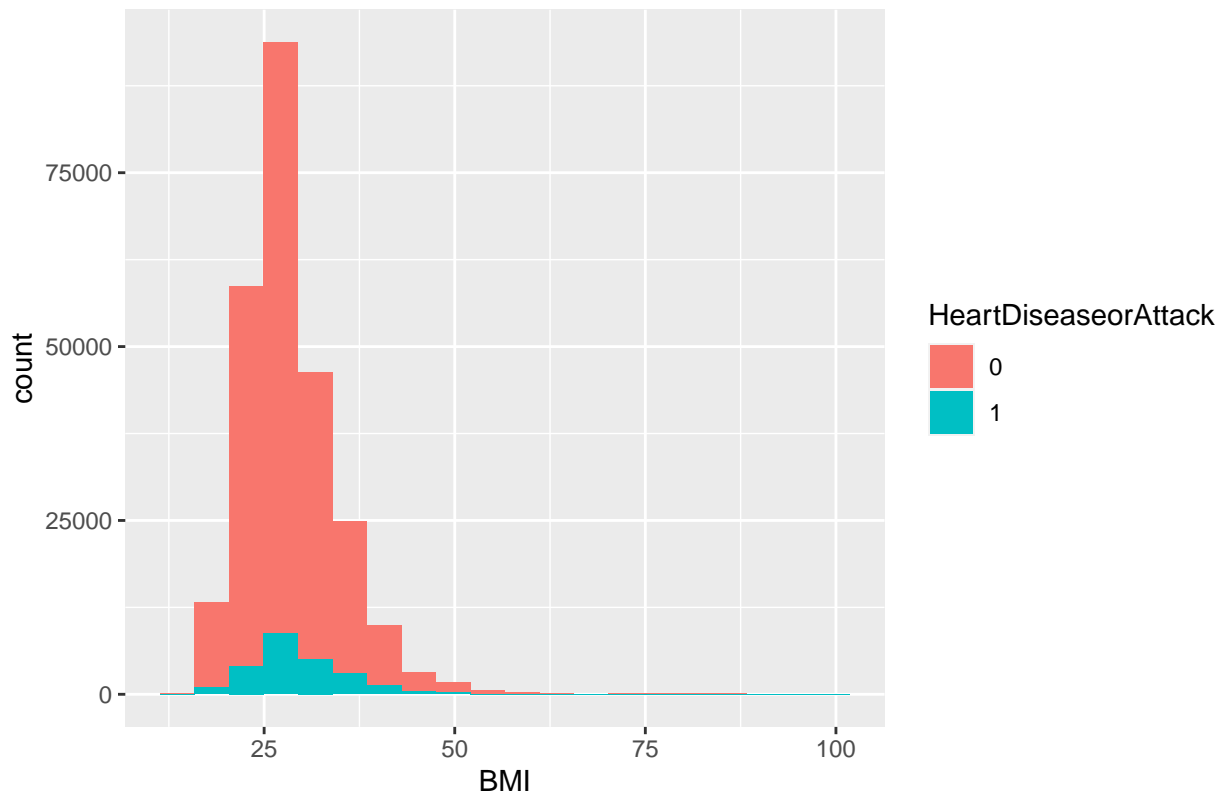
The Heart Disease Health Indicators Dataset was created from surveys which collect responses from over 400,000 Americans on health-related risk behaviors. In this dataset, I will be analyzing each factor to see which one has effects toward Heart Disease. Furthermore, I will compare which factor has the most and least effects to Heart Disease. However, I will be looking at factors that I believe have certain effects toward Heart Disease.

Histogram and Analysis

1) Heart Disease vs BMI

```
## Warning: package 'ggplot2' was built under R version 4.2.2
```

Histogram of Heart Disease vs BMI (stack)

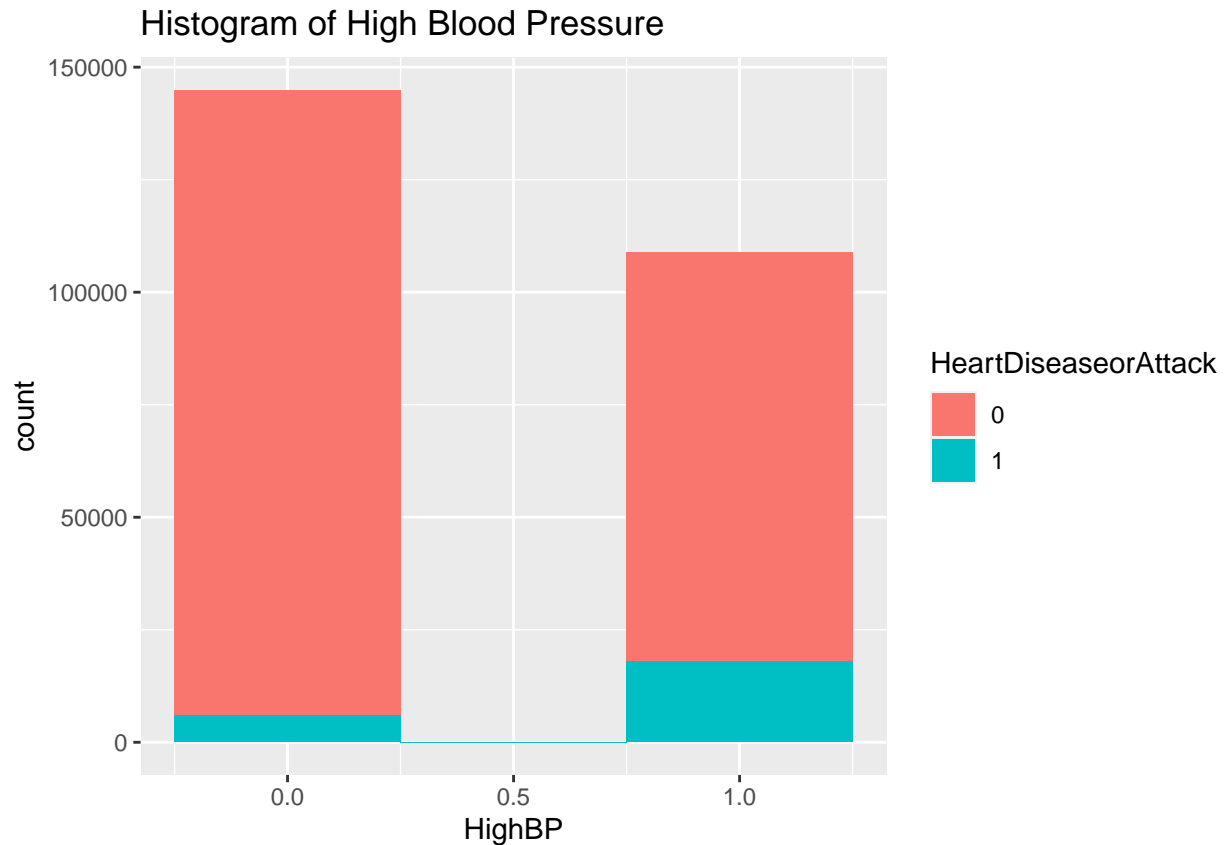


```
## [1] 12.00000 21.55556 31.11111 40.66667 50.22222 59.77778 69.33333 78.88889
## [9] 88.44444 98.00000
```

```
##          22      31      41      50      60      69      79      88      98
## FALSE 21758.00 154272.0 43695.0 8138.00 1135.0 232.00 236.00 246.00 75.00
## TRUE   1519.00  15128.0  5829.0 1166.00  172.0  37.00  16.00  19.00  7.00
## Total 23277.00 169400.0 49524.0 9304.00 1307.0 269.00 252.00 265.00 82.00
## Odds   14.32    10.2     7.5     6.98    6.6    6.27  14.75  12.95 10.71
##          Total
## FALSE 229787.00
## TRUE   23893.00
## Total 253680.00
## Odds    9.62
```

Based on my observation of this comparison between Heart Disease and BMI, the histogram indicates that the chances of getting heart disease for people who don't have BMI is more likely compared to the people who have BMI. However based on the data chart, we can see that more than half of the odds of each bar indicates the percentage higher than the total odds, while the rest are lower. Furthermore, it shows that the rate of getting heart disease is increasing. The odd ratio between each odds and the odd total is significant. Therefore, BMI has a big effect toward the Heart Disease.

2) Heart Disease vs High Blood Pressure



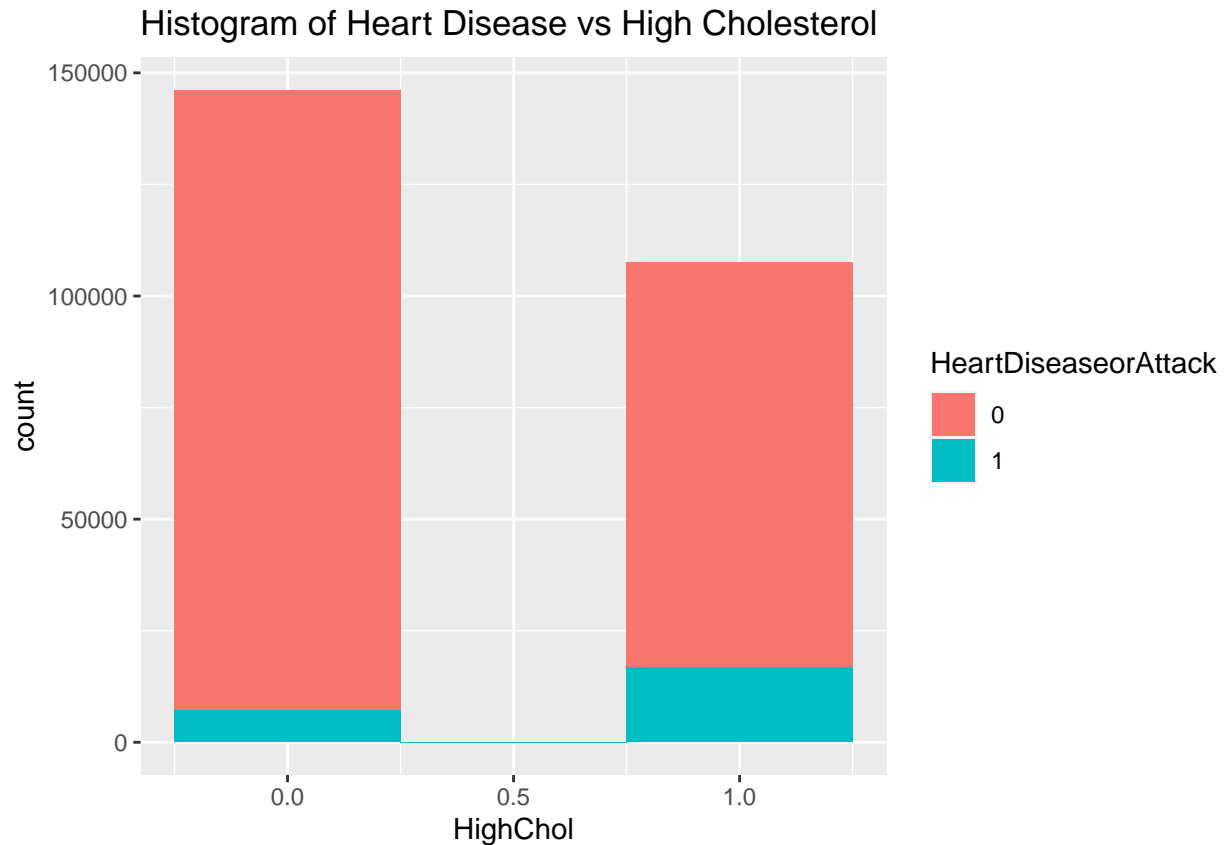
```
## [1] 0.0000000 0.1111111 0.2222222 0.3333333 0.4444444 0.5555556 0.6666667
## [8] 0.7777778 0.8888889 1.0000000
```

```
##           0    0    0    0    1    1    1    1           1    Total
## FALSE 138886.00    0    0    0    0    0    0    0  90901.00 229787.00
## TRUE   5965.00    0    0    0    0    0    0    0  17928.00  23893.00
## Total 144851.00    0    0    0    0    0    0    0 108829.00 253680.00
## Odds      23.28 NaN NaN NaN NaN NaN NaN NaN      5.07      9.62
```

- No High Blood Pressure: 23%
- High Blood Pressure: 5%
- There is 4 times difference between these.

For this histogram, it is different from the one above as it only indicates two bars. Based on the histogram, we can see that there is more data of people who don't have high blood pressure. Furthermore, based on the data of the chart, we see that there is more data of people who don't have high blood pressure, and there is a change by 4 times. For high blood pressure, people who have high blood pressure has higher rate of getting heart disease than people who don't have high blood pressure. The odd ratio doesn't seem to be significant.

3) Heart Disease vs High Cholesterol



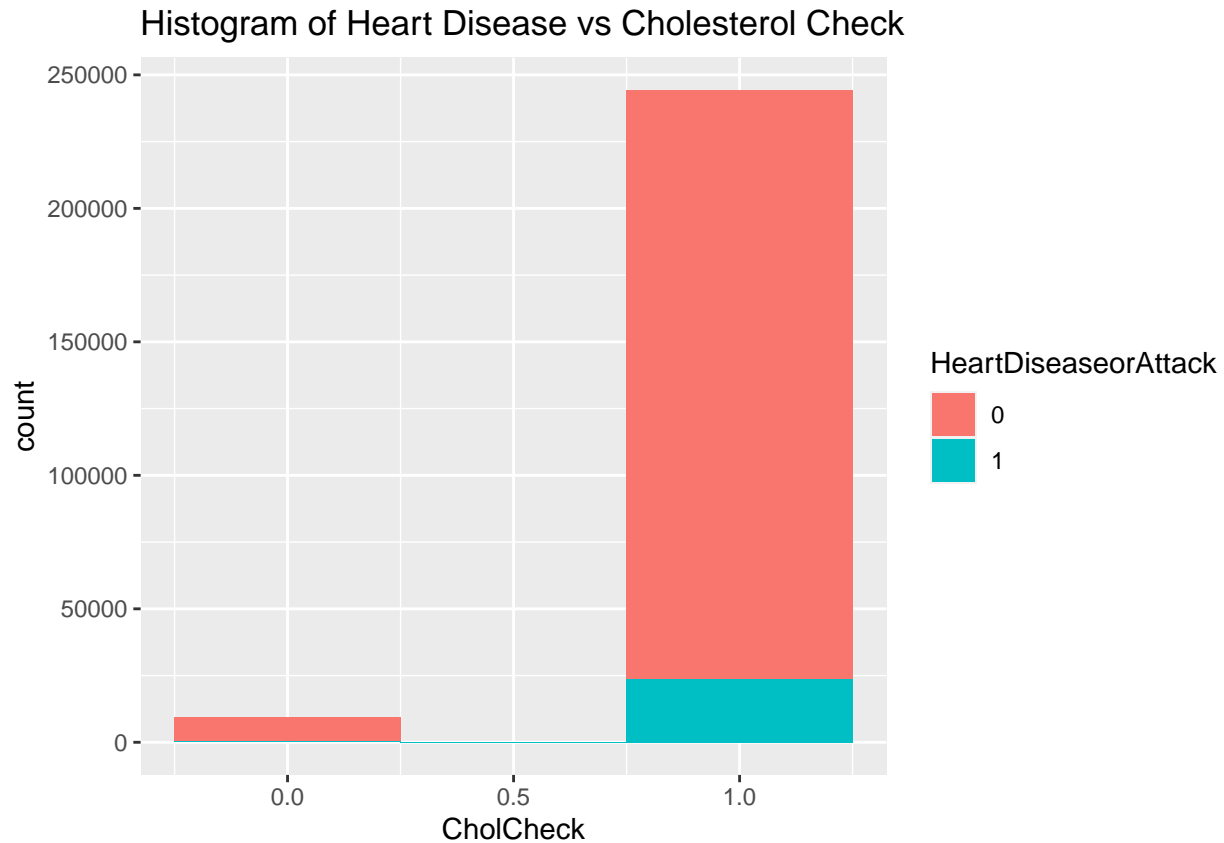
```
## [1] 0.0000000 0.1111111 0.2222222 0.3333333 0.4444444 0.5555556 0.6666667
## [8] 0.7777778 0.8888889 1.0000000
```

```
##           0    0    0    0    1    1    1    1           1    Total
## FALSE 138949.00    0    0    0    0    0    0    0  90838.00 229787.00
## TRUE   7140.00    0    0    0    0    0    0    0  16753.00  23893.00
## Total 146089.00    0    0    0    0    0    0    0 107591.00 253680.00
## Odds   19.46 NaN NaN NaN NaN NaN NaN NaN    5.42    9.62
```

- No High Cholesterol: 19%
- High Cholesterol: 5%
- There is 4 times difference between these.

For this histogram, it is similar to the one above. Based on the histogram, we can see that there is more data of people who don't have high cholesterol. Furthermore, based on the data of the chart, we see that there is more data of people who don't have high cholesterol, and there is a change by 4 times. For high cholesterol, people who have high cholesterol has higher rate of getting heart disease than people who don't have high cholesterol. The odd ratio doesn't seem to be significant.

4) Heart Disease vs Cholesterol Check



```
## [1] 0.0000000 0.1111111 0.2222222 0.3333333 0.4444444 0.5555556 0.6666667
## [8] 0.7777778 0.8888889 1.0000000
```

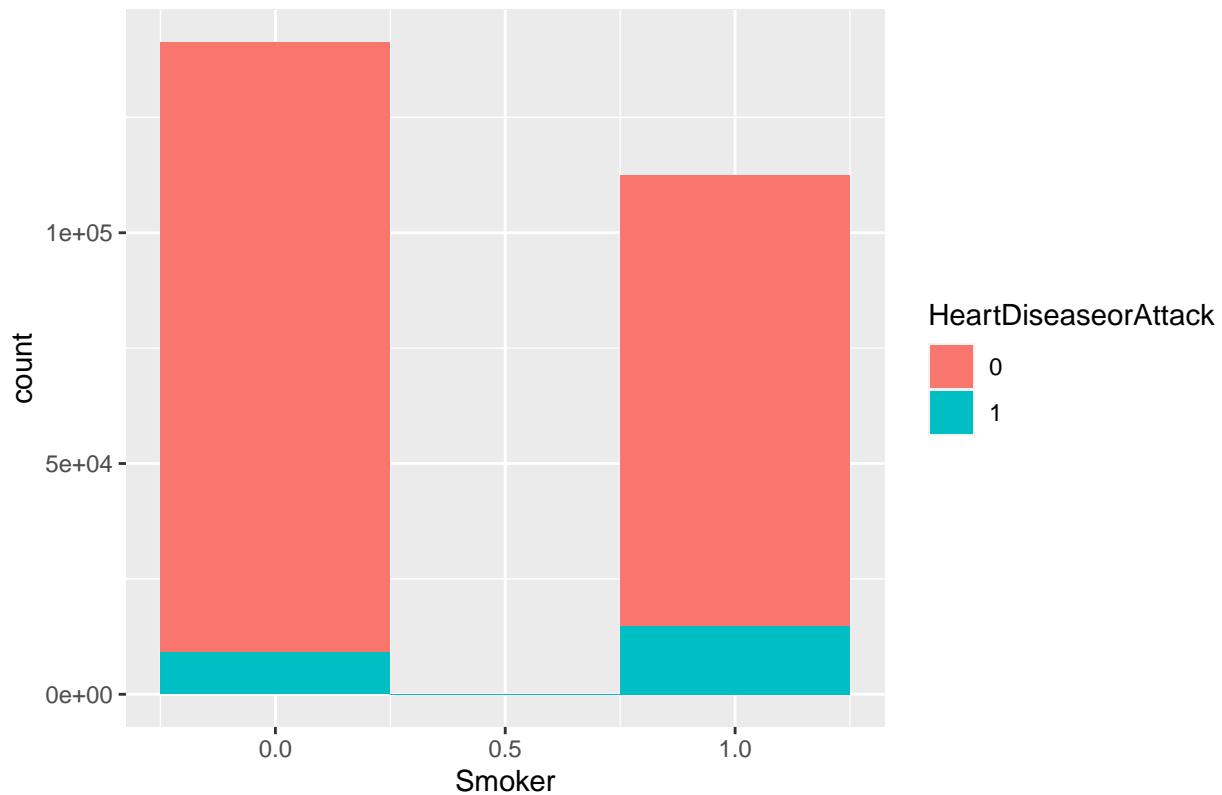
```
##           0    0    0    0    1    1    1    1           1    Total
## FALSE 9199.00    0    0    0    0    0    0    0 220588.00 229787.00
## TRUE  271.00    0    0    0    0    0    0    0  23622.00  23893.00
## Total 9470.00    0    0    0    0    0    0    0 244210.00 253680.00
## Odds   33.94 NaN NaN NaN NaN NaN NaN NaN    9.34    9.62
```

- No Cholesterol Check: 34%
- Cholesterol Check: 9%
- There is a 4 times difference between these.

Based on the histogram, we can see that there is more data of people who have cholesterol check. Furthermore, based on the data of the chart, we see that there is more data of people who have high cholesterol, and there is a change by 4 times. For cholesterol check, people who have cholesterol check has higher rate of getting heart disease than people who don't have cholesterol check. The odd ratio doesn't seem to be significant.

5) Heart Disease vs Smoker

Histogram of Heart Disease vs Smoker



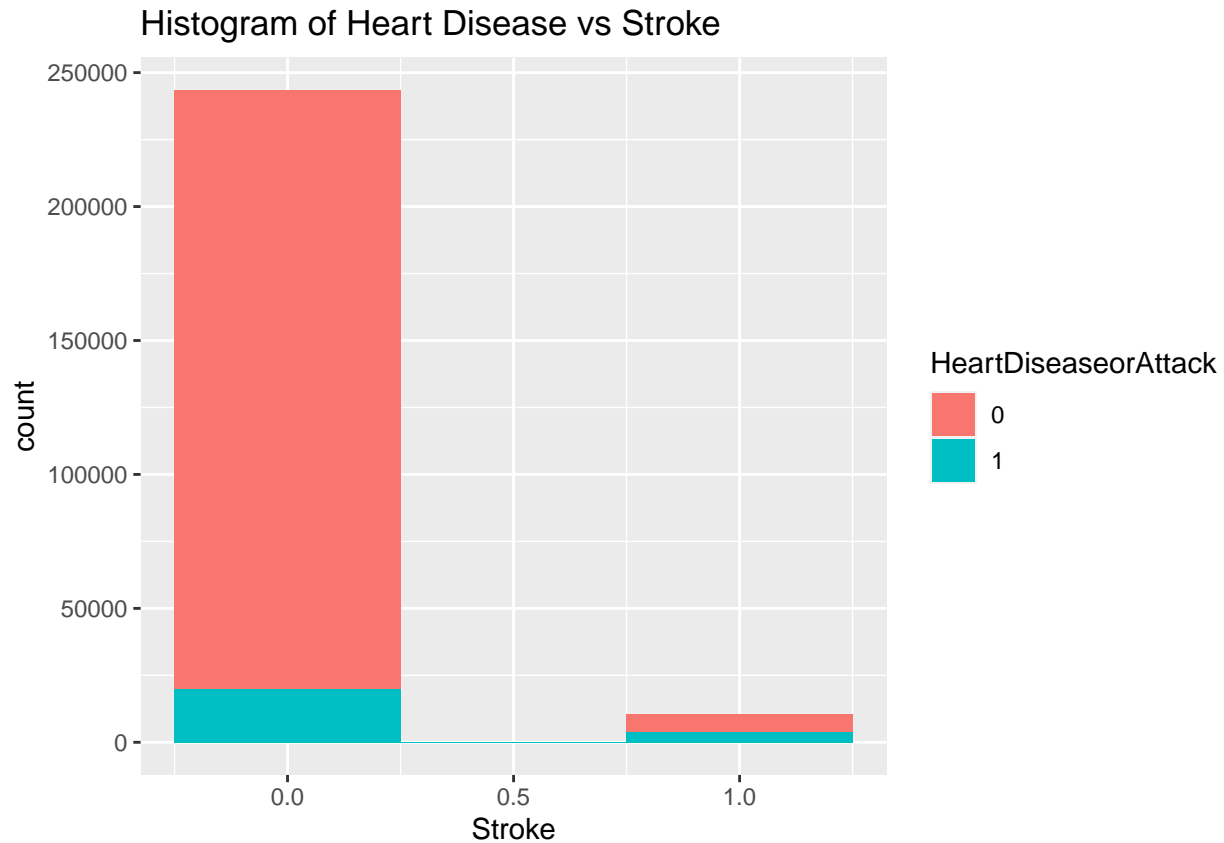
```
## [1] 0.0000000 0.1111111 0.2222222 0.3333333 0.4444444 0.5555556 0.6666667
## [8] 0.7777778 0.8888889 1.0000000
```

```
##           0    0    0    0    1    1    1    1           1    Total
## FALSE 132165.00    0    0    0    0    0    0    0  97622.0 229787.00
## TRUE   9092.00    0    0    0    0    0    0    0  14801.0 23893.00
## Total 141257.00    0    0    0    0    0    0    0 112423.0 253680.00
## Odds      14.54 NaN NaN NaN NaN NaN NaN NaN      6.6      9.62
```

- No Smoke: 15%
- Smoke: 6%
- There is a 2 times difference between these.

Based on the histogram, we can see that there is more data of people who don't smoke. Furthermore, based on the data of the chart, we see that there is more data of people who don't have high cholesterol, and there is an change by 2 times. People who smoke has higher rate of getting heart disease than people who don't smoke. The odd ratio doesn't seem to be significant.

6) Heart Disease vs Stroke



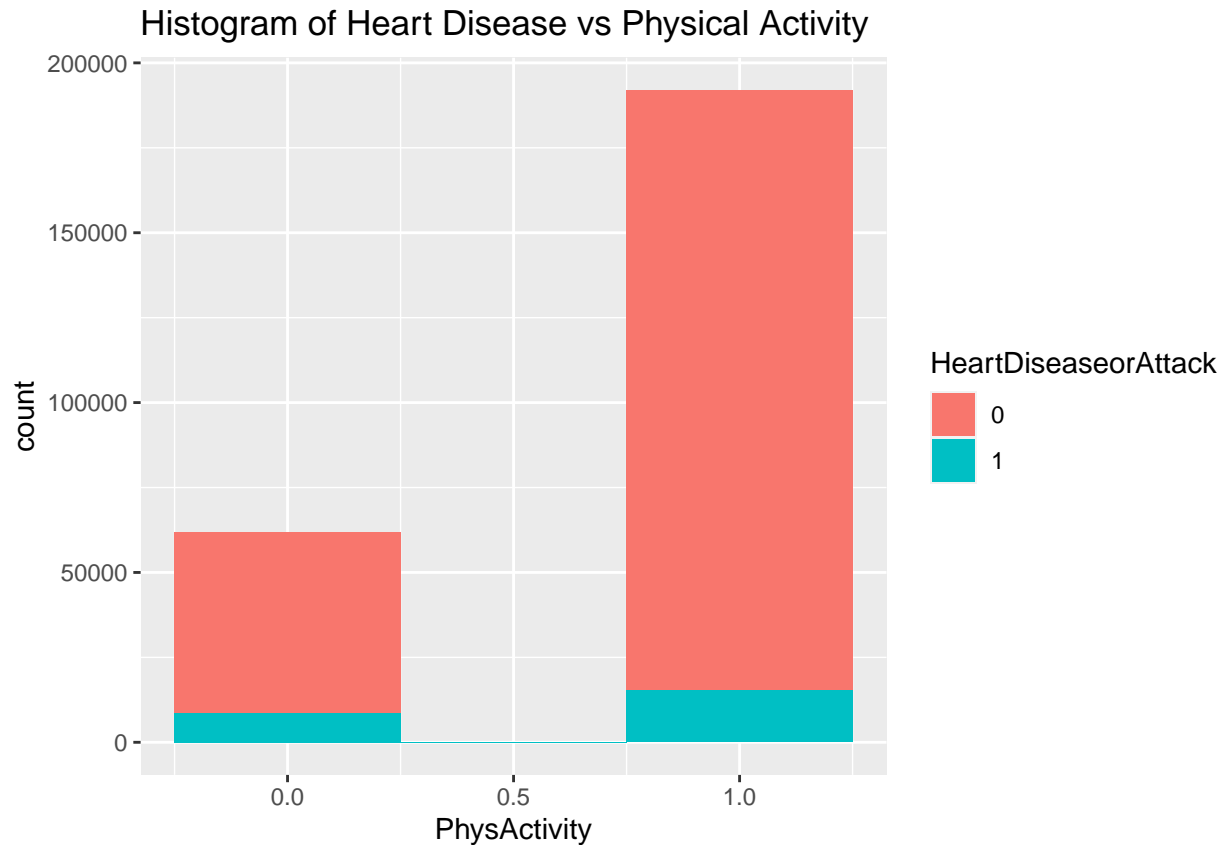
```
## [1] 0.0000000 0.1111111 0.2222222 0.3333333 0.4444444 0.5555556 0.6666667
## [8] 0.7777778 0.8888889 1.0000000
```

```
##           0    0    0    0    1    1    1    1           1    Total
## FALSE 223432.0    0    0    0    0    0    0    0  6355.00 229787.00
## TRUE  19956.0    0    0    0    0    0    0    0  3937.00  23893.00
## Total 243388.0    0    0    0    0    0    0    0 10292.00 253680.00
## Odds   11.2 NaN NaN NaN NaN NaN NaN NaN   1.61    9.62
```

- No Stroke: 11%
- Stroke: 2%
- There is a 5 times difference between these.

Based on the histogram, we can see that there is more data of people who don't stroke. Furthermore, based on the data of the chart, we see that there is more data of people who don't stroke, and there is a change by 5 times. People who stroke has higher rate of getting heart disease than people who don't stroke. The odd ratio doesn't seem to be significant.

7) Heart Disease vs Physical Activity



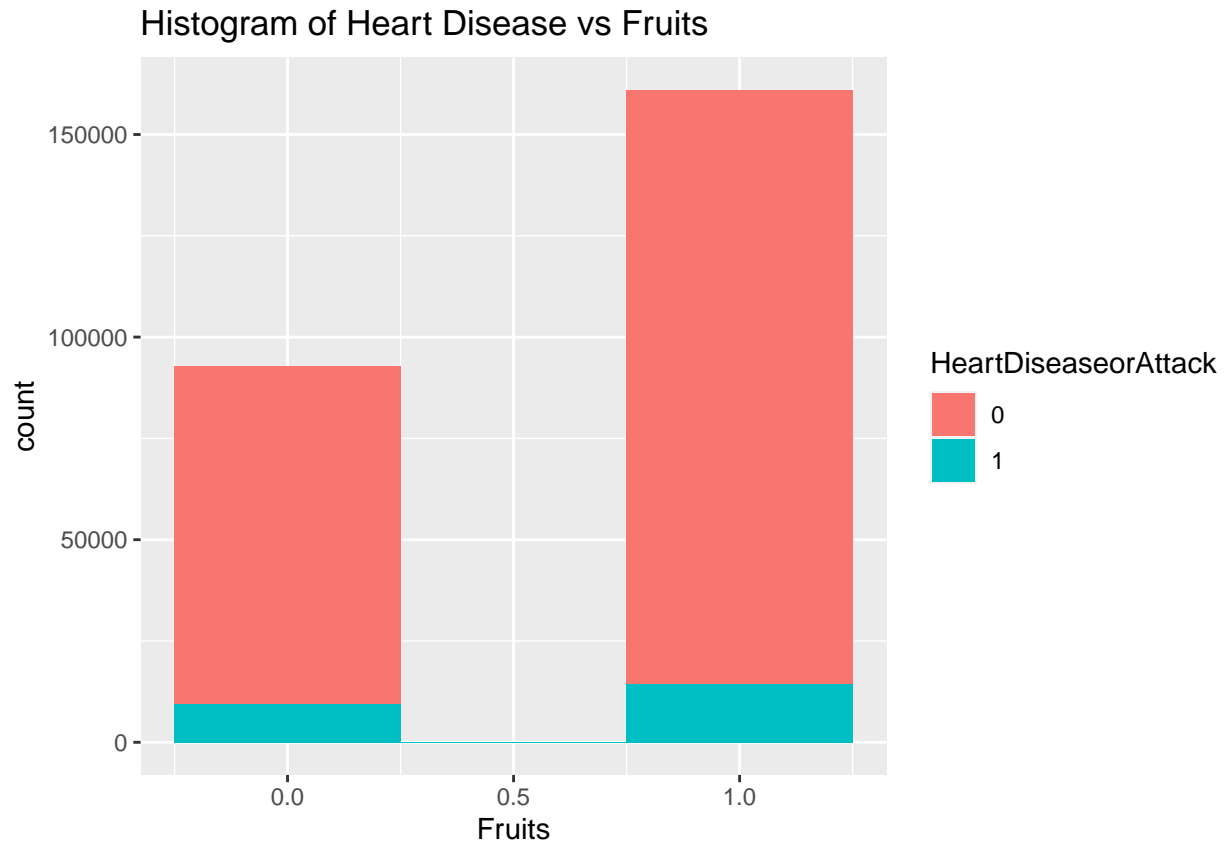
```
## [1] 0.0000000 0.1111111 0.2222222 0.3333333 0.4444444 0.5555556 0.6666667
## [8] 0.7777778 0.8888889 1.0000000
```

```
##           0    0    0    0    1    1    1    1           1    Total
## FALSE 53167.00    0    0    0    0    0    0    0 176620.00 229787.00
## TRUE  8593.00    0    0    0    0    0    0    0  15300.00  23893.00
## Total 61760.00    0    0    0    0    0    0    0 191920.00 253680.00
## Odds   6.19 NaN NaN NaN NaN NaN NaN NaN   11.54    9.62
```

- No Physical Activity: 6%
- Physical Activity: 11%
- There is a 2 times difference between these.

Based on the histogram, we can see that there is more data of people who have physical activity. Furthermore, based on the data of the chart, we see that there is more data of people who have physical activity, and there is a change by 2 times. People who don't have physical activity has higher rate of getting heart disease than people who have. The odd ratio doesn't seem to be significant.

8) Heart Disease vs Fruits

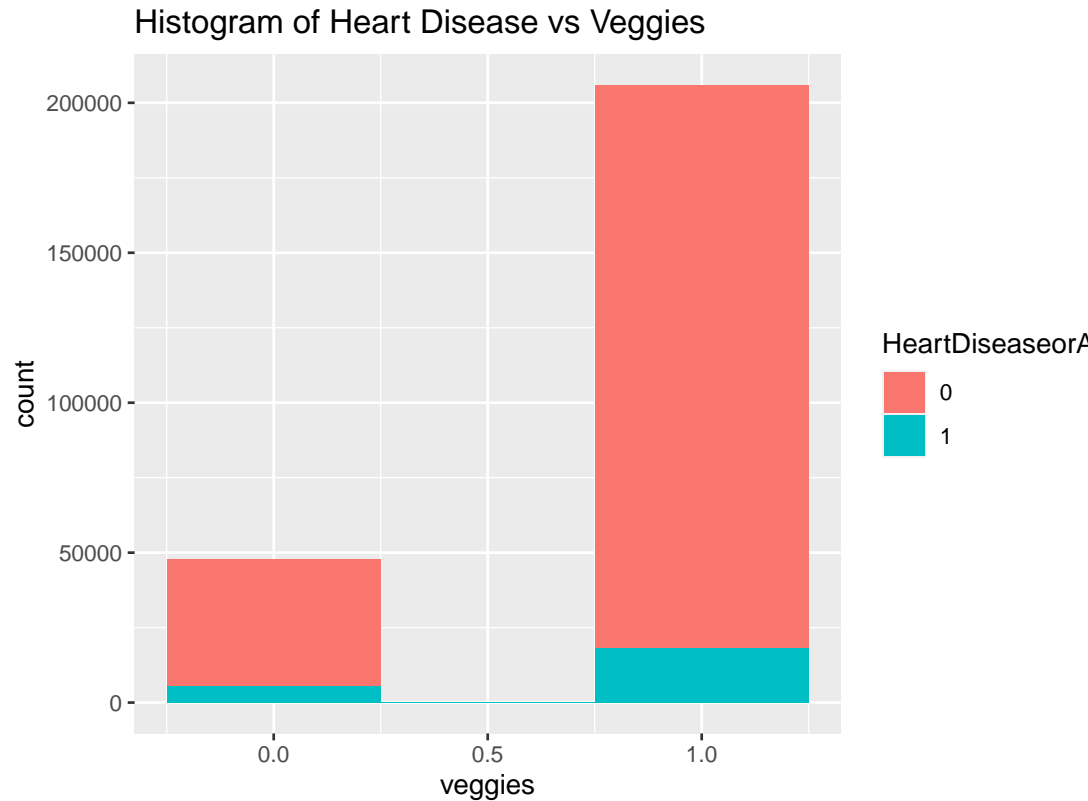


```
## [1] 0.0000000 0.1111111 0.2222222 0.3333333 0.4444444 0.5555556 0.6666667
## [8] 0.7777778 0.8888889 1.0000000
```

```
##           0    0    0    0    1    1    1    1           1    Total
## FALSE 83337.00    0    0    0    0    0    0    0 146450.00 229787.00
## TRUE   9445.00    0    0    0    0    0    0    0  14448.00  23893.00
## Total 92782.00    0    0    0    0    0    0    0 160898.00 253680.00
## Odds      8.82 NaN NaN NaN NaN NaN NaN NaN    10.14     9.62
```

- No Fruits: 9%
- Fruits: 10%
- There is a 2 times difference between these.

Based on the histogram, we can see that there is more data of people who have fruits. Furthermore, based on the data of the chart, we see that there is more data of people who have fruits, and there is a change by 2 times. People who don't have fruits has higher rate of getting heart disease than people who have fruits. The odd ratio doesn't seem to be significant.



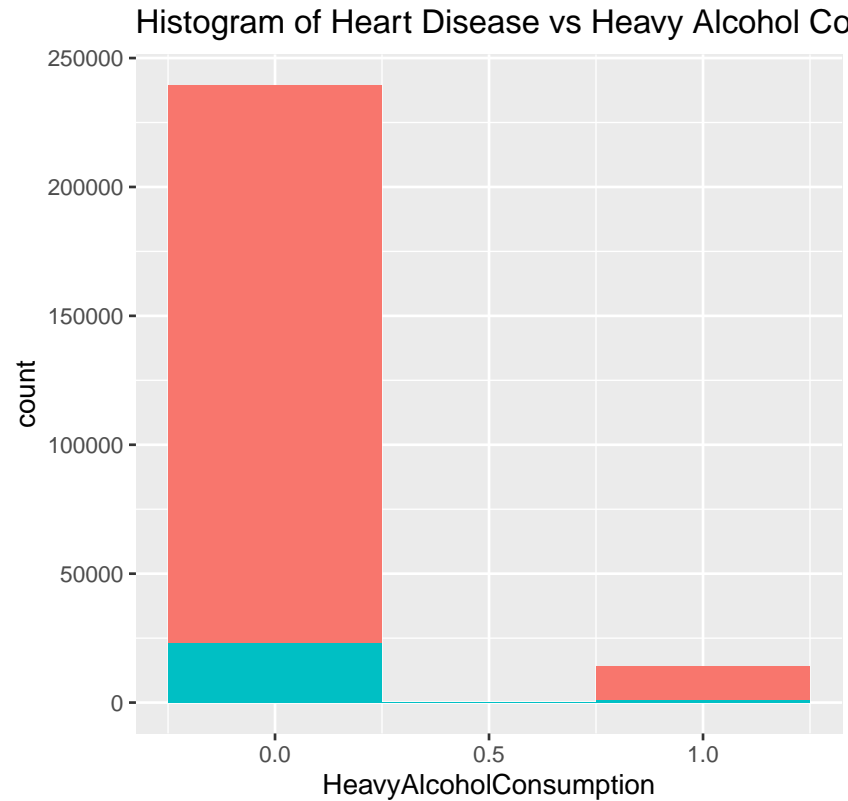
9) Heart Disease vs Veggies

```
## [1] 0.0000000 0.1111111 0.2222222 0.3333333 0.4444444 0.5555556 0.6666667
## [8] 0.7777778 0.8888889 1.0000000
```

```
##           0    0    0    0    1    1    1    1           1    Total
## FALSE 42198.00  0    0    0    0    0    0    0 187589.00 229787.00
## TRUE  5641.00  0    0    0    0    0    0    0  18252.00  23893.00
## Total 47839.00  0    0    0    0    0    0    0 205841.00 253680.00
## Odds      7.48 NaN NaN NaN NaN NaN NaN NaN    10.28     9.62
```

- No Veggies: 7%
- Veggies: 10%
- There is a 2 times difference between these.

Based on the histogram, we can see that there is more data of people who have veggies. Furthermore, based on the data of the chart, we see that there is more data of people who have veggies, and there is a change by 2 times. People who don't have veggies has higher rate of getting heart disease than people who have veggies. The odd ratio doesn't seem to be significant.



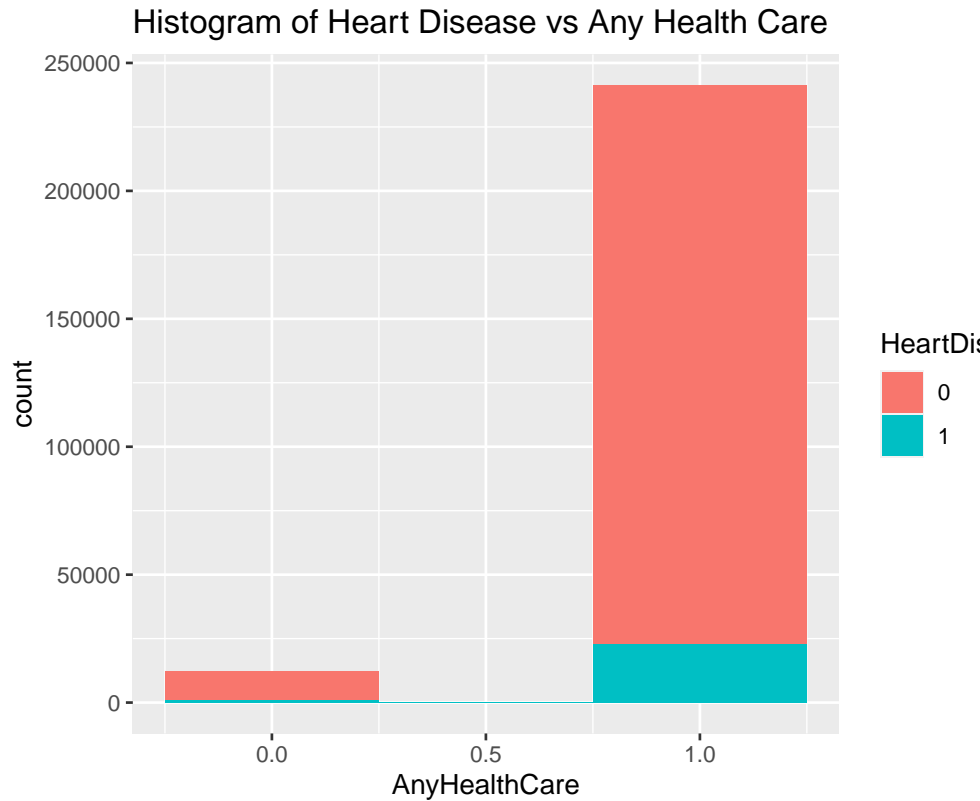
10) Heart Disease vs Heavy Alcohol Consumption

```
## [1] 0.0000000 0.1111111 0.2222222 0.3333333 0.4444444 0.5555556 0.6666667
## [8] 0.7777778 0.8888889 1.0000000
```

```
##           0    0    0    0    1    1    1    1           1    Total
## FALSE 216379.00  0    0    0    0    0    0    0 13408.00 229787.00
## TRUE  23045.00  0    0    0    0    0    0    0   848.00  23893.00
## Total 239424.00  0    0    0    0    0    0    0 14256.00 253680.00
## Odds      9.39 NaN NaN NaN NaN NaN NaN NaN   15.81    9.62
```

- No Heavy Alcohol Consumption: 9%
- Heavy Alcohol Consumption: 16%
- There is a 2 times difference between these.

Based on the histogram, we can see that there is more data of people who have heavy alcohol consumption. Furthermore, based on the data of the chart, we see that there is more data of people who have heavy alcohol consumption, and there is a change by 2 times. People who don't have heavy alcohol consumption has higher rate of getting heart disease than people who have veggies. The odd ratio doesn't seem to be significant.



11) Heart Disease vs Any Health Care

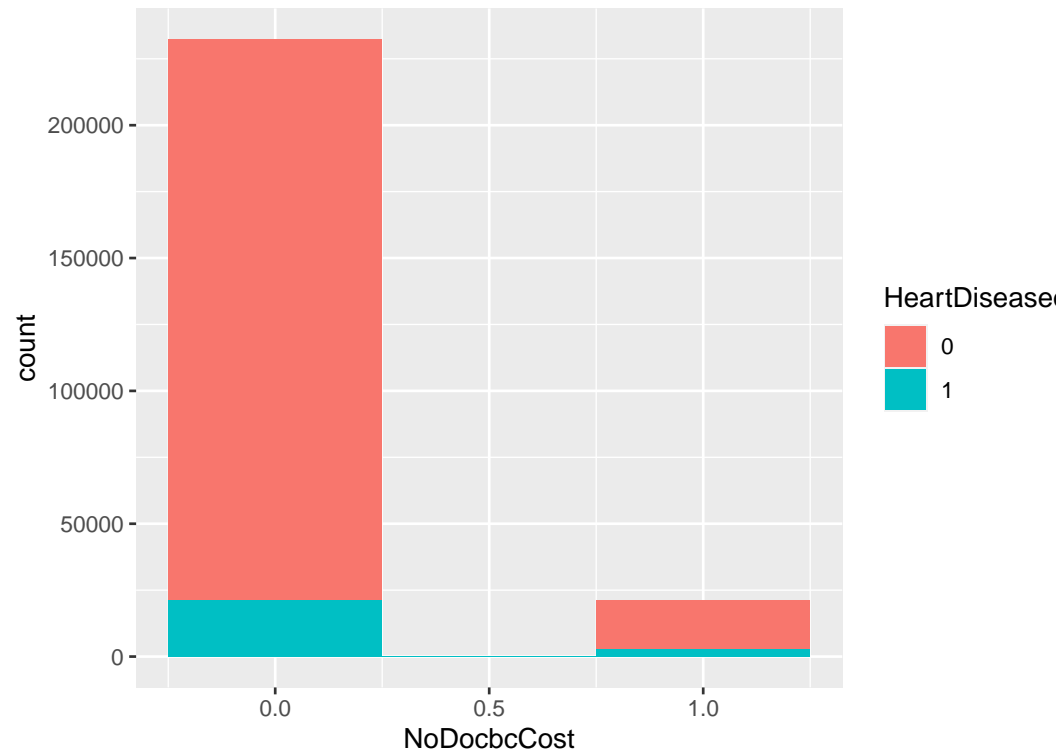
```
## [1] 0.0000000 0.1111111 0.2222222 0.3333333 0.4444444 0.5555556 0.6666667
## [8] 0.7777778 0.8888889 1.0000000
```

```
##           0    0    0    0    1    1    1    1           1    Total
## FALSE 11547.00    0    0    0    0    0    0    0 218240.00 229787.00
## TRUE   870.00    0    0    0    0    0    0    0  23023.00  23893.00
## Total 12417.00    0    0    0    0    0    0    0 241263.00 253680.00
## Odds   13.27 NaN NaN NaN NaN NaN NaN NaN    9.48    9.62
```

- No Health Care: 13%
- Any Health Care: 9%
- There is a 2 times difference between these.

Based on the histogram, we can see that there is more data of people who have any health care. Furthermore, based on the data of the chart, we see that there is more data of people who have any health care, and there is a change by 2 times. People who have any health care has higher rate of getting heart disease than people who don't. The odd ratio doesn't seem to be significant.

Histogram of Heart Disease vs No Doctor because of Cost



12) Heart Disease vs No Doctor

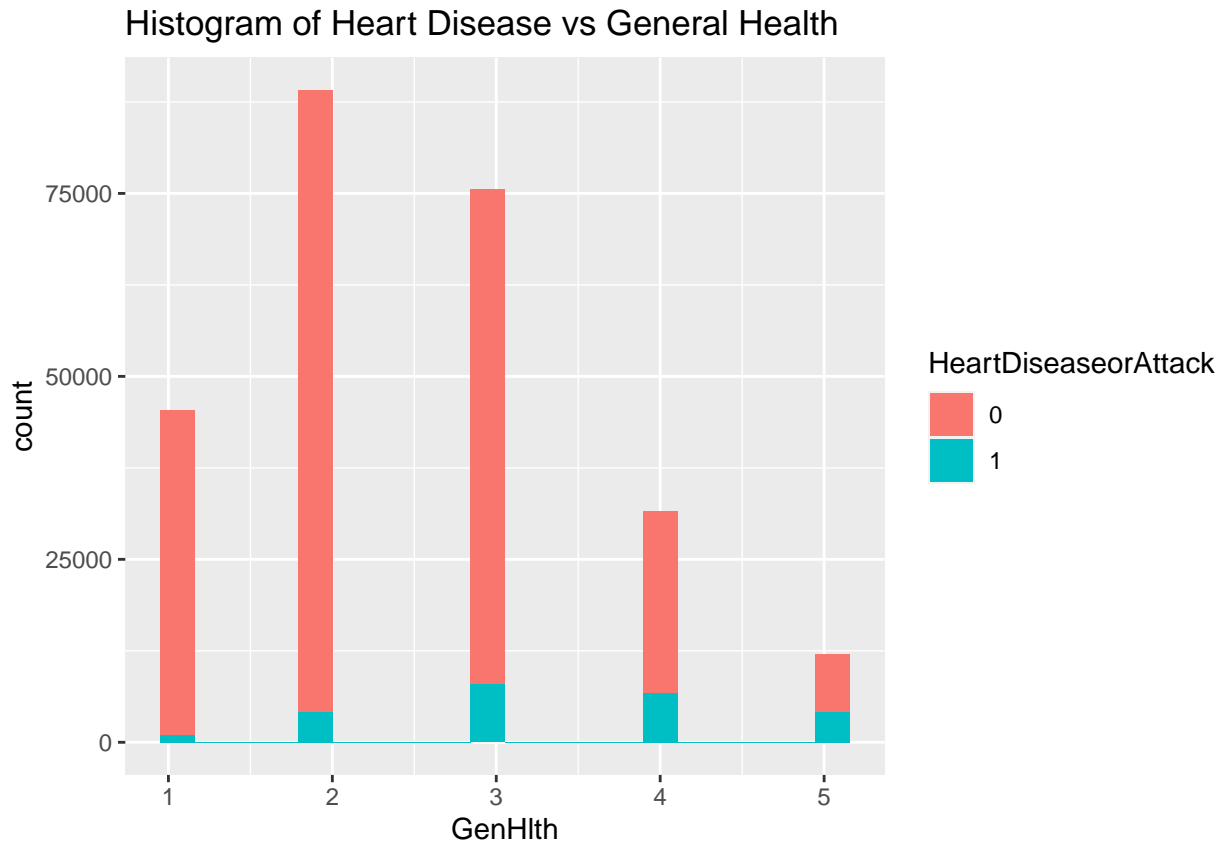
```
## [1] 0.0000000 0.1111111 0.2222222 0.3333333 0.4444444 0.5555556 0.6666667
## [8] 0.7777778 0.8888889 1.0000000
```

```
##           0    0    0    0    1    1    1    1           1    Total
## FALSE 211082.00  0    0    0    0    0    0    0 18705.00 229787.00
## TRUE  21244.00  0    0    0    0    0    0    0  2649.00  23893.00
## Total 232326.00  0    0    0    0    0    0    0 21354.00 253680.00
## Odds      9.94 NaN NaN NaN NaN NaN NaN NaN      7.06      9.62
```

- Have Doctor: 10%
- No Doctor: 7%
- There is a 2 times difference between these.

Based on the histogram, we can see that there is more data of people who have doctors. Furthermore, based on the data of the chart, we see that there is more data of people who have doctors, and there is a change by 2 times. People who don't have a doctor has higher rate of getting heart disease than people who has. The odd ratio doesn't seem to be significant.

13) Heart Disease vs General Health

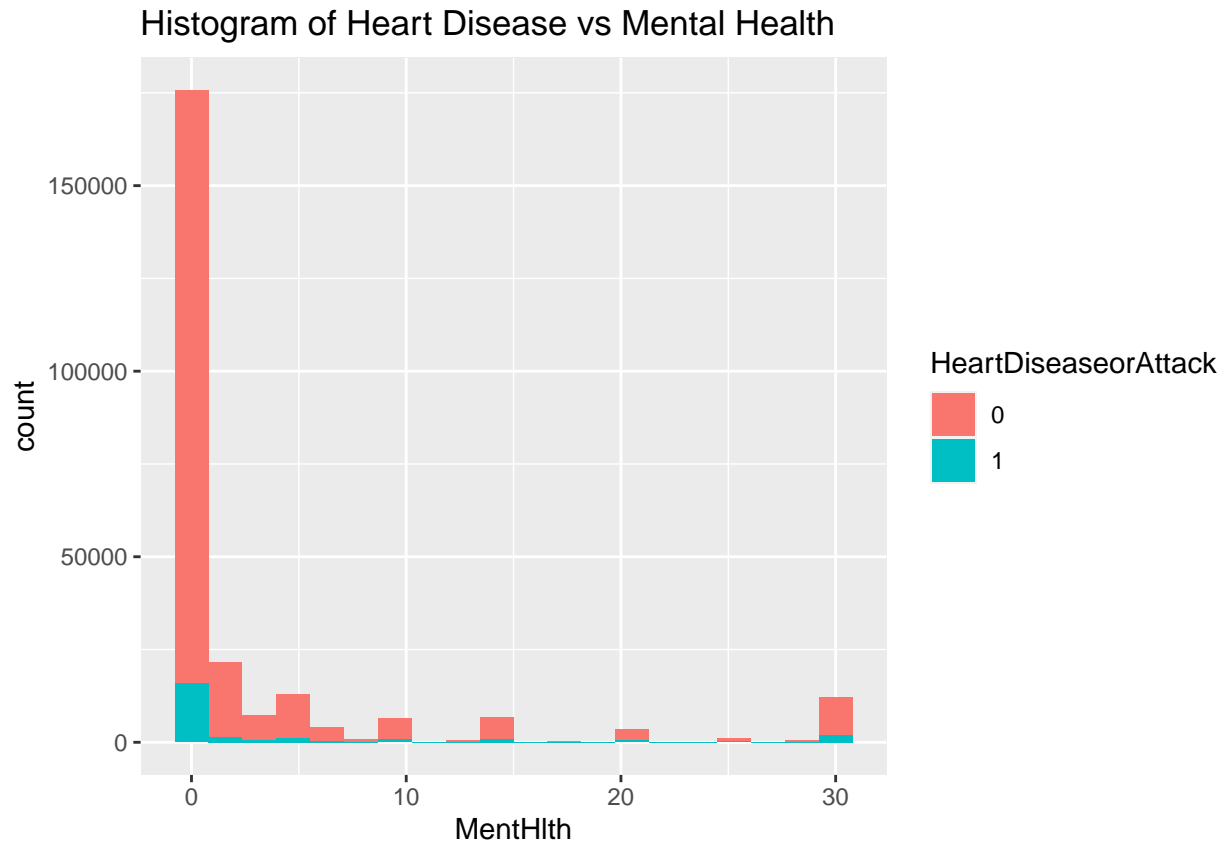


```
## [1] 1.000000 1.444444 1.888889 2.333333 2.777778 3.222222 3.666667 4.111111
## [9] 4.555556 5.000000
```

```
##          1    2      2    3          3    4          4    5          5    Total
## FALSE 44283.00  0 84956.00  0 67732.00  0 24842.00  0 7974.00 229787.00
## TRUE  1016.00  0  4128.00  0  7914.00  0  6728.00  0  4107.00  23893.00
## Total 45299.00  0 89084.00  0 75646.00  0 31570.00  0 12081.00 253680.00
## Odds   43.59 NaN   20.58 NaN    8.56 NaN    3.69 NaN    1.94    9.62
```

Based on my observation of this comparison between Heart Disease and General Health, the histogram indicates that the chances of getting heart disease for people who don't have General Health is more likely compared to the people who have General Health. However based on the data chart, we can see that more than half of the odds of each bar indicates the percentage higher than the total odds, while the rest are lower. Furthermore, it shows that the rate of getting heart disease is increasing. The odd ratio between each odds and the odd total seems to be significant. Therefore, General Health has a big effect toward the Heart Disease.

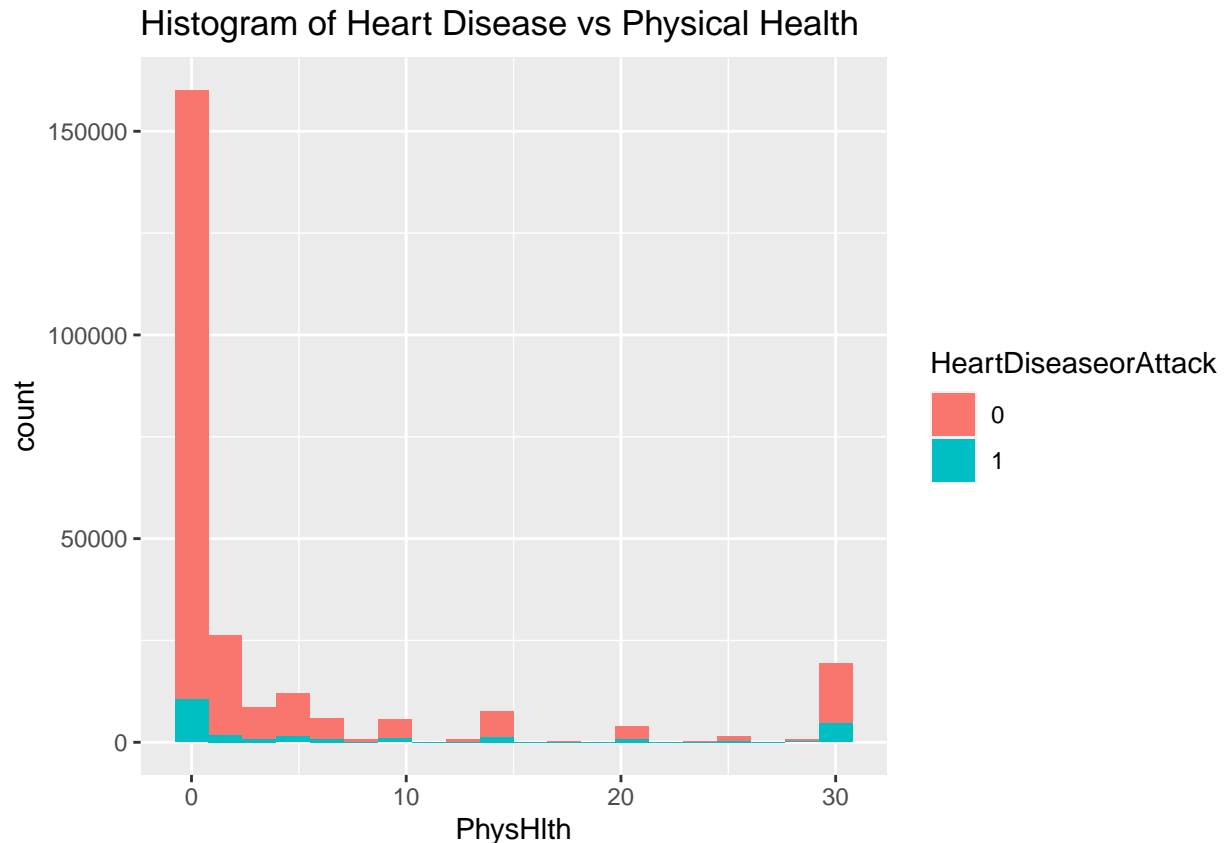
14) Heart Disease vs Mental Health



```
## [1] 0.000000 3.333333 6.666667 10.000000 13.333333 16.666667 20.000000
## [8] 23.333333 26.666667 30.000000
```

```
##          3          7          10          13          17          20          23          27
## FALSE 186802.00 12556.00  9164.00  423.00  5895.00  3037.00  277.00 1072.00
## TRUE   17851.00  1251.00  1039.00   57.00   865.00   494.00   51.00  194.00
## Total 204653.00 13807.00 10203.00  480.00  6760.00  3531.00  328.00 1266.00
## Odds   10.46   10.04    8.82    7.42    6.82    6.15    5.43    5.53
##          30      Total
## FALSE 10561.00 229787.00
## TRUE   2091.00  23893.00
## Total 12652.00 253680.00
## Odds    5.05    9.62
```

Based on my observation of this comparison between Heart Disease and Mental Health, the histogram indicates that the chances of getting heart disease for people who have Mental Health is more likely compared to the people who don't have Mental Health. Furthermore, based on the data chart, we can see that more than half of the odds of each bar indicates the percentage lower than the total odds. Furthermore, it shows that the rate of getting heart disease is decreasing. The odd ratio between each odds and the odd total is not significant. 15) Heart Disease vs Physical Health

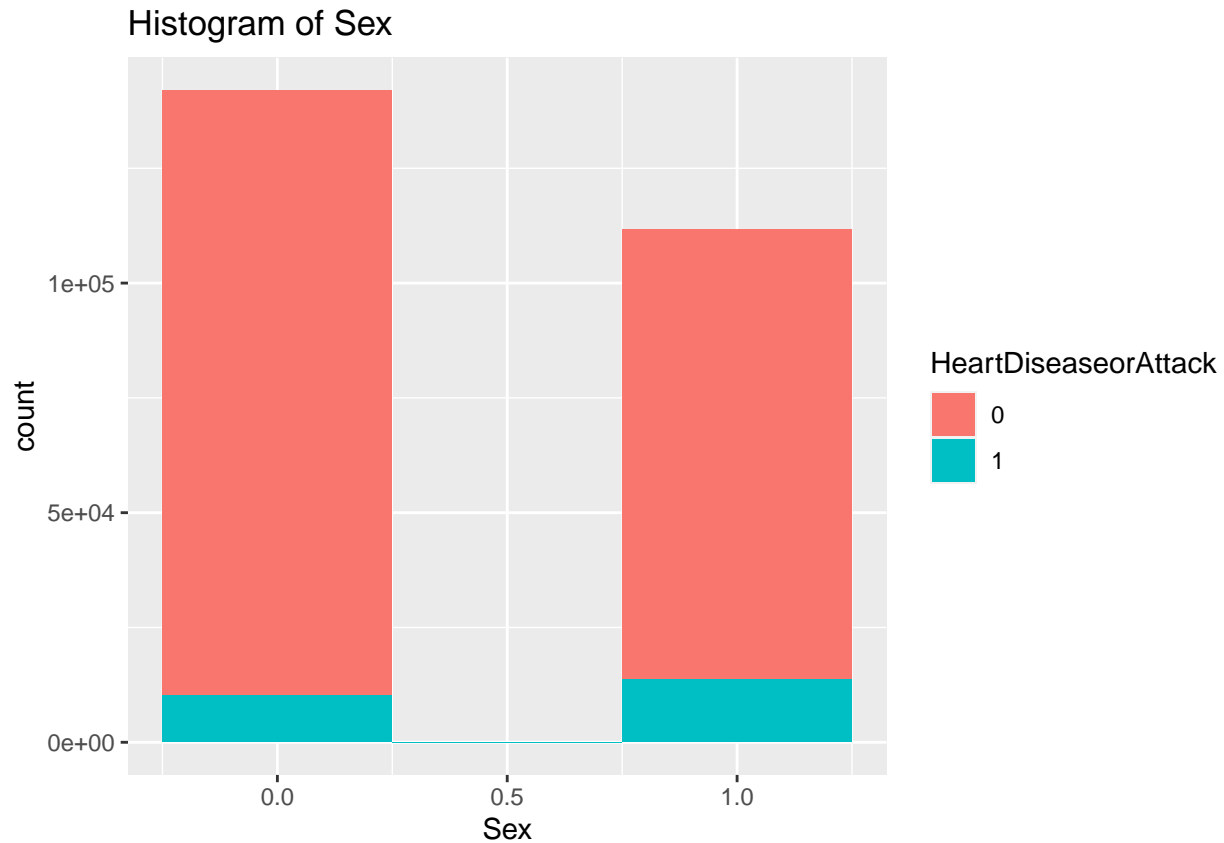


```
## [1] 0.000000 3.333333 6.666667 10.000000 13.333333 16.666667 20.000000
## [8] 23.333333 26.666667 30.000000

##           3           7           10           13           17           20           23           27           30
## FALSE 181532.00 11828.0  9525.00  590.00  6293.00  2811.00  665.00  1151.00 15392.00
## TRUE   13167.00  1666.0  1596.00  116.00  1322.00  732.00  124.00  326.00  4844.00
## Total 194699.00 13494.0 11121.00  706.00  7615.00  3543.00  789.00 1477.00 20236.00
## Odds   13.79     7.1    5.97    5.09    4.76    3.84    5.36    3.53    3.18
##      Total
## FALSE 229787.00
## TRUE   23893.00
## Total 253680.00
## Odds    9.62
```

This histogram is also similar to the one above. Based on my observation of this comparison between Heart Disease and Physical Health, the histogram indicates that the chances of getting heart disease for people who have Physical Health is more likely compared to the people who don't have Physical Health. Furthermore, based on the data chart, we can see that more than half of the odds of each bar indicates the percentage lower than the total odds. Furthermore, it shows that the rate of getting heart disease is decreasing. The odd ratio between each odds and the odd total is not significant.

16) Heart Disease vs Sex



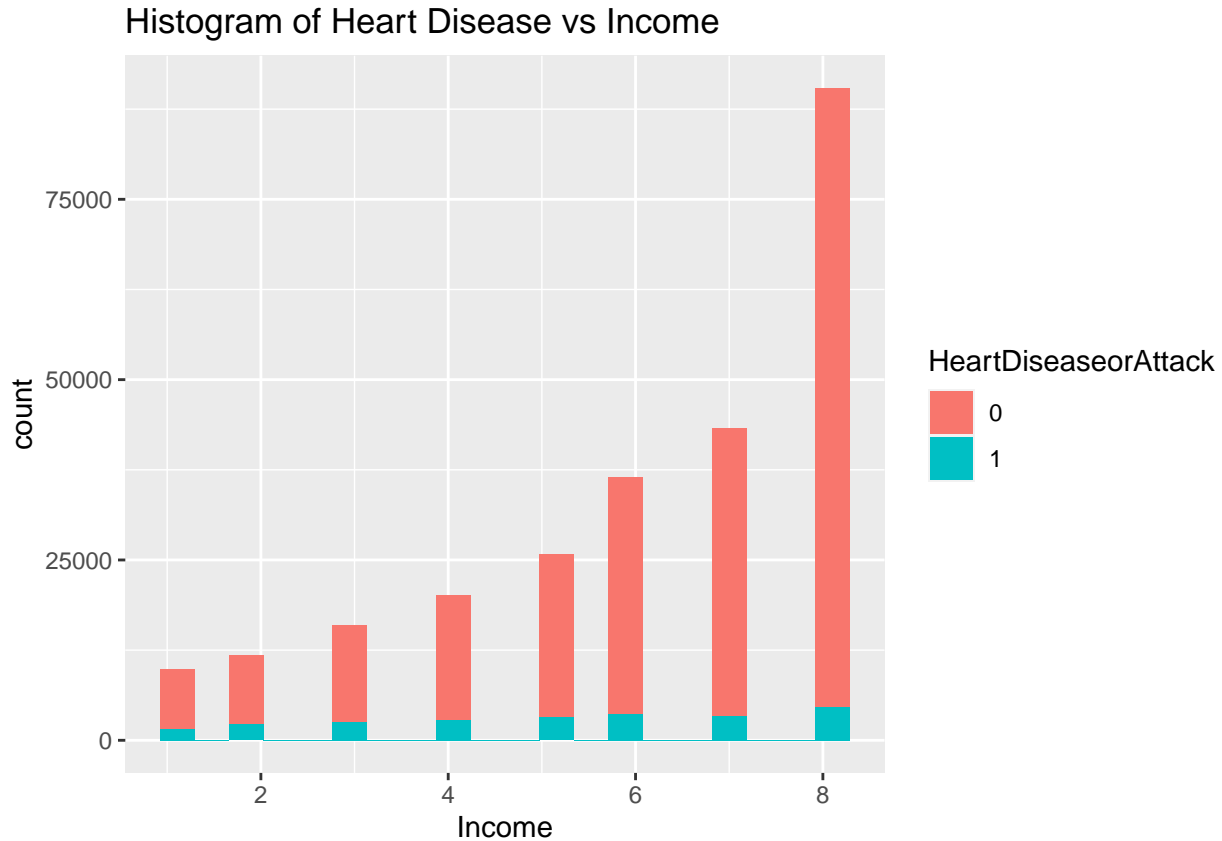
```
## [1] 0.0000000 0.1111111 0.2222222 0.3333333 0.4444444 0.5555556 0.6666667
## [8] 0.7777778 0.8888889 1.0000000
```

```
##           0    0    0    0    1    1    1    1           1    Total
## FALSE 131769.00    0    0    0    0    0    0    0  98018.00 229787.00
## TRUE   10205.00    0    0    0    0    0    0    0  13688.00  23893.00
## Total 141974.00    0    0    0    0    0    0    0 111706.00 253680.00
## Odds    12.91 NaN NaN NaN NaN NaN NaN NaN    7.16    9.62
```

- No Health Care: 13%
- Any Health Care: 7%
- There is a 2 times difference between these.

Based on the histogram, we can see that there is more data of people who don't have sex. Furthermore, based on the data of the chart, we see that there is more data of people who don't have sex, and there is a change by 2 times. People who have sex has higher rate of getting heart disease than people who don't. The odd ratio doesn't seem to be significant.

17) Heart Disease vs Income



```
## [1] 1.000000 1.777778 2.555556 3.333333 4.111111 4.888889 5.666667 6.444444
## [9] 7.222222 8.000000

##           2           3           3           4           5           6           6           7           8
## FALSE 8258.00  9586.00 13475.00 17307.00  0 22722.00 32824 39815.0 85800.00
## TRUE  1553.00  2197.00  2519.00  2828.00  0  3161.00  3646  3404.0  4585.00
## Total 9811.00 11783.00 15994.00 20135.00  0 25883.00 36470 43219.0 90385.00
## Odds   5.32   4.36   5.35   6.12 NaN   7.19   9   11.7   18.71
##           Total
## FALSE 229787.00
## TRUE  23893.00
## Total 253680.00
## Odds   9.62
```

Based on my observation of this comparison between Heart Disease and Income, the histogram indicates that the chances of getting heart disease for people who don't have Income is more likely compared to the people who have Income. Furthermore, based on the data chart, we can see that more than half of the odds of each bar indicates the percentage lower than the total odds. However, it shows that the rate of getting heart disease is increasing. The odd ratio between each odds and the odd total is not significant.

Conclusion

After going through comparing each factors to the heart disease, I conclude that only the BMI and General Health have significant effects to the Heart Disease since they have significant odd ratio. Furthermore, although the rest of the factors don't have significant effects to the Heart Disease, it can be seen that they still have effects at a certain degree.

Code Appendix

```
heart_disease <- read.csv("C:/Users/hoang/Downloads/STA 106/heart_disease_health_indicators_BRFSS2015.csv")
heart_disease$HeartDiseaseorAttack=as.factor(heart_disease$HeartDiseaseorAttack)
library(ggplot2)
ggplot(data=heart_disease, aes(x=BMI,fill = HeartDiseaseorAttack)) +
  geom_histogram(bins=20,position='stack')+
  labs(x ='BMI', title ='Histogram of Heart Disease vs BMI (stack)')
variable_name='BMI'
response_name='HeartDiseaseorAttack'

ax=seq(min(heart_disease[,variable_name]),max(heart_disease[,variable_name]),length.out=10)
ax
hgA = hist(heart_disease[heart_disease[,response_name]==0,variable_name], breaks = ax, plot = FALSE)
hgB = hist(heart_disease[heart_disease[,response_name]==1,variable_name], breaks = ax, plot = FALSE)

table1=rbind('FALSE'=hgA$counts,'TRUE'=hgB$counts)

colnames(table1)=round(hgA$breaks[-1],0)
# add row sum
table1=cbind(table1, 'Total'=apply(table1,1,sum))
# add column sum
table1=rbind(table1, 'Total'=apply(table1,2,sum))
# add odds
table1=rbind(table1, 'Odds'=round(table1[1,]/table1[2,],2))
to_print='Odds'=round(table1[1,]/table1[2,],2)
to_print_ratio='ratio'=round(table1[2,]/table1[1,],2)

table1
library(ggplot2)
ggplot(data=heart_disease, aes(x=HighBP,fill = HeartDiseaseorAttack)) +
  geom_histogram(bins=3,position='stack')+
  labs(x ='HighBP', title ='Histogram of High Blood Pressure')
variable_name='HighBP'
response_name='HeartDiseaseorAttack'

ax=seq(min(heart_disease[,variable_name]),max(heart_disease[,variable_name]),length.out=10)
ax
hgA = hist(heart_disease[heart_disease[,response_name]==0,variable_name], breaks = ax, plot = FALSE)
hgB = hist(heart_disease[heart_disease[,response_name]==1,variable_name], breaks = ax, plot = FALSE)

table1=rbind('FALSE'=hgA$counts,'TRUE'=hgB$counts)

colnames(table1)=round(hgA$breaks[-1],0)
# add row sum
table1=cbind(table1, 'Total'=apply(table1,1,sum))
# add column sum
table1=rbind(table1, 'Total'=apply(table1,2,sum))
# add odds
table1=rbind(table1, 'Odds'=round(table1[1,]/table1[2,],2))
to_print='Odds'=round(table1[1,]/table1[2,],2)
to_print_ratio='ratio'=round(table1[2,]/table1[1,],2)
```

```

table1
library(ggplot2)
ggplot(data=heart_disease, aes(x=HighChol, fill = HeartDiseaseorAttack)) +
  geom_histogram(bins=3, position='stack') +
  labs(x = 'HighChol', title = 'Histogram of Heart Disease vs High Cholesterol')
variable_name='HighChol'
response_name='HeartDiseaseorAttack'

ax=seq(min(heart_disease[,variable_name]),max(heart_disease[,variable_name]),length.out=10)
ax
hgA = hist(heart_disease[heart_disease[,response_name]==0,variable_name], breaks = ax, plot = FALSE)
hgB = hist(heart_disease[heart_disease[,response_name]==1,variable_name], breaks = ax, plot = FALSE)

table1=rbind('FALSE'=hgA$counts,'TRUE'=hgB$counts)

colnames(table1)=round(hgA$breaks[-1],0)
# add row sum
table1=cbind(table1, 'Total'=apply(table1,1,sum))
# add column sum
table1=rbind(table1, 'Total'=apply(table1,2,sum))
# add odds
table1=rbind(table1, 'Odds'=round(table1[1,]/table1[2,],2))
to_print='Odds'=round(table1[1,]/table1[2,],2)
to_print_ratio='ratio'=round(table1[2,]/table1[1,],2)

table1
library(ggplot2)
ggplot(data=heart_disease, aes(x=CholCheck, fill = HeartDiseaseorAttack)) +
  geom_histogram(bins=3, position='stack') +
  labs(x = 'CholCheck', title = 'Histogram of Heart Disease vs Cholesterol Check')
variable_name='CholCheck'
response_name='HeartDiseaseorAttack'

ax=seq(min(heart_disease[,variable_name]),max(heart_disease[,variable_name]),length.out=10)
ax
hgA = hist(heart_disease[heart_disease[,response_name]==0,variable_name], breaks = ax, plot = FALSE)
hgB = hist(heart_disease[heart_disease[,response_name]==1,variable_name], breaks = ax, plot = FALSE)

table1=rbind('FALSE'=hgA$counts,'TRUE'=hgB$counts)

colnames(table1)=round(hgA$breaks[-1],0)
# add row sum
table1=cbind(table1, 'Total'=apply(table1,1,sum))
# add column sum
table1=rbind(table1, 'Total'=apply(table1,2,sum))
# add odds
table1=rbind(table1, 'Odds'=round(table1[1,]/table1[2,],2))
to_print='Odds'=round(table1[1,]/table1[2,],2)
to_print_ratio='ratio'=round(table1[2,]/table1[1,],2)

table1
library(ggplot2)
ggplot(data=heart_disease, aes(x=Smoker, fill = HeartDiseaseorAttack)) +

```

```

    geom_histogram(bins=3,position='stack')+
    labs(x='Smoker', title='Histogram of Heart Disease vs Smoker')
variable_name='Smoker'
response_name='HeartDiseaseorAttack'

ax=seq(min(heart_disease[,variable_name]),max(heart_disease[,variable_name]),length.out=10)
ax
hgA = hist(heart_disease[heart_disease[,response_name]==0,variable_name], breaks = ax, plot = FALSE)
hgB = hist(heart_disease[heart_disease[,response_name]==1,variable_name], breaks = ax, plot = FALSE)

table1=rbind('FALSE'=hgA$counts,'TRUE'=hgB$counts)

colnames(table1)=round(hgA$breaks[-1],0)
# add row sum
table1=cbind(table1, 'Total'=apply(table1,1,sum))
# add column sum
table1=rbind(table1, 'Total'=apply(table1,2,sum))
# add odds
table1=rbind(table1, 'Odds'=round(table1[1,]/table1[2,],2))
to_print='Odds'=round(table1[1,]/table1[2,],2)
to_print_ratio='ratio'=round(table1[2,]/table1[1,],2)

table1
library(ggplot2)
ggplot(data=heart_disease, aes(x=Stroke,fill = HeartDiseaseorAttack)) +
  geom_histogram(bins=3,position='stack')+
  labs(x='Stroke', title='Histogram of Heart Disease vs Stroke')
variable_name='Stroke'
response_name='HeartDiseaseorAttack'

ax=seq(min(heart_disease[,variable_name]),max(heart_disease[,variable_name]),length.out=10)
ax
hgA = hist(heart_disease[heart_disease[,response_name]==0,variable_name], breaks = ax, plot = FALSE)
hgB = hist(heart_disease[heart_disease[,response_name]==1,variable_name], breaks = ax, plot = FALSE)

table1=rbind('FALSE'=hgA$counts,'TRUE'=hgB$counts)

colnames(table1)=round(hgA$breaks[-1],0)
# add row sum
table1=cbind(table1, 'Total'=apply(table1,1,sum))
# add column sum
table1=rbind(table1, 'Total'=apply(table1,2,sum))
# add odds
table1=rbind(table1, 'Odds'=round(table1[1,]/table1[2,],2))
to_print='Odds'=round(table1[1,]/table1[2,],2)
to_print_ratio='ratio'=round(table1[2,]/table1[1,],2)

table1
library(ggplot2)
ggplot(data=heart_disease, aes(x=PhysActivity,fill = HeartDiseaseorAttack)) +
  geom_histogram(bins=3,position='stack')+
  labs(x='PhysActivity', title='Histogram of Heart Disease vs Physical Activity')
variable_name='PhysActivity'

```

```

response_name='HeartDiseaseorAttack'

ax=seq(min(heart_disease[,variable_name]),max(heart_disease[,variable_name]),length.out=10)
ax
hgA = hist(heart_disease[heart_disease[,response_name]==0,variable_name], breaks = ax, plot = FALSE)
hgB = hist(heart_disease[heart_disease[,response_name]==1,variable_name], breaks = ax, plot = FALSE)

table1=rbind('FALSE'=hgA$counts,'TRUE'=hgB$counts)

colnames(table1)=round(hgA$breaks[-1],0)
# add row sum
table1=cbind(table1, 'Total'=apply(table1,1,sum))
# add column sum
table1=rbind(table1, 'Total'=apply(table1,2,sum))
# add odds
table1=rbind(table1, 'Odds'=round(table1[1,]/table1[2,],2))
to_print='Odds'=round(table1[1,]/table1[2,],2)
to_print_ratio='ratio'=round(table1[2,]/table1[1,],2)

table1
library(ggplot2)
ggplot(data=heart_disease, aes(x=Fruits,fill = HeartDiseaseorAttack)) +
  geom_histogram(bins=3,position='stack')+
  labs(x = 'Fruits', title = 'Histogram of Heart Disease vs Fruits')
variable_name='Fruits'
response_name='HeartDiseaseorAttack'

ax=seq(min(heart_disease[,variable_name]),max(heart_disease[,variable_name]),length.out=10)
ax
hgA = hist(heart_disease[heart_disease[,response_name]==0,variable_name], breaks = ax, plot = FALSE)
hgB = hist(heart_disease[heart_disease[,response_name]==1,variable_name], breaks = ax, plot = FALSE)

table1=rbind('FALSE'=hgA$counts,'TRUE'=hgB$counts)

colnames(table1)=round(hgA$breaks[-1],0)
# add row sum
table1=cbind(table1, 'Total'=apply(table1,1,sum))
# add column sum
table1=rbind(table1, 'Total'=apply(table1,2,sum))
# add odds
table1=rbind(table1, 'Odds'=round(table1[1,]/table1[2,],2))
to_print='Odds'=round(table1[1,]/table1[2,],2)
to_print_ratio='ratio'=round(table1[2,]/table1[1,],2)

table1
library(ggplot2)
ggplot(data=heart_disease, aes(x=Veggies,fill = HeartDiseaseorAttack)) +
  geom_histogram(bins=3,position='stack')+
  labs(x = 'veggies', title = 'Histogram of Heart Disease vs Veggies')
variable_name='Veggies'
response_name='HeartDiseaseorAttack'

ax=seq(min(heart_disease[,variable_name]),max(heart_disease[,variable_name]),length.out=10)

```

```

ax
hgA = hist(heart_disease[heart_disease[,response_name]==0,variable_name], breaks = ax, plot = FALSE)
hgB = hist(heart_disease[heart_disease[,response_name]==1,variable_name], breaks = ax, plot = FALSE)

table1=rbind('FALSE'=hgA$counts,'TRUE'=hgB$counts)

colnames(table1)=round(hgA$breaks[-1],0)
# add row sum
table1=cbind(table1, 'Total'=apply(table1,1,sum))
# add column sum
table1=rbind(table1, 'Total'=apply(table1,2,sum))
# add odds
table1=rbind(table1, 'Odds'=round(table1[1,]/table1[2,],2))
to_print='Odds'=round(table1[1,]/table1[2,],2)
to_print_ratio='ratio'=round(table1[2,]/table1[1,],2)

table1
library(ggplot2)
ggplot(data=heart_disease, aes(x=HvyAlcoholConsump,fill = HeartDiseaseorAttack)) +
  geom_histogram(bins=3,position='stack')+
  labs(x = 'HeavyAlcoholConsumption', title = 'Histogram of Heart Disease vs Heavy Alcohol Consumption')
variable_name='HvyAlcoholConsump'
response_name='HeartDiseaseorAttack'

ax=seq(min(heart_disease[,variable_name]),max(heart_disease[,variable_name]),length.out=10)
ax
hgA = hist(heart_disease[heart_disease[,response_name]==0,variable_name], breaks = ax, plot = FALSE)
hgB = hist(heart_disease[heart_disease[,response_name]==1,variable_name], breaks = ax, plot = FALSE)

table1=rbind('FALSE'=hgA$counts,'TRUE'=hgB$counts)

colnames(table1)=round(hgA$breaks[-1],0)
# add row sum
table1=cbind(table1, 'Total'=apply(table1,1,sum))
# add column sum
table1=rbind(table1, 'Total'=apply(table1,2,sum))
# add odds
table1=rbind(table1, 'Odds'=round(table1[1,]/table1[2,],2))
to_print='Odds'=round(table1[1,]/table1[2,],2)
to_print_ratio='ratio'=round(table1[2,]/table1[1,],2)

table1
library(ggplot2)
ggplot(data=heart_disease, aes(x=AnyHealthcare,fill = HeartDiseaseorAttack)) +
  geom_histogram(bins=3,position='stack')+
  labs(x = 'AnyHealthCare', title = 'Histogram of Heart Disease vs Any Health Care')
variable_name='AnyHealthcare'
response_name='HeartDiseaseorAttack'

ax=seq(min(heart_disease[,variable_name]),max(heart_disease[,variable_name]),length.out=10)
ax
hgA = hist(heart_disease[heart_disease[,response_name]==0,variable_name], breaks = ax, plot = FALSE)
hgB = hist(heart_disease[heart_disease[,response_name]==1,variable_name], breaks = ax, plot = FALSE)

```

```

table1=rbind('FALSE'=hgA$counts,'TRUE'=hgB$counts)

colnames(table1)=round(hgA$breaks[-1],0)
# add row sum
table1=cbind(table1, 'Total'=apply(table1,1,sum))
# add column sum
table1=rbind(table1, 'Total'=apply(table1,2,sum))
# add odds
table1=rbind(table1, 'Odds'=round(table1[1,]/table1[2,],2))
to_print='Odds'=round(table1[1,]/table1[2,],2)
to_print_ratio='ratio'=round(table1[2,]/table1[1,],2)

table1
library(ggplot2)
ggplot(data=heart_disease, aes(x=NoDocbcCost,fill = HeartDiseaseorAttack)) +
  geom_histogram(bins=3,position='stack')+
  labs(x = 'NoDocbcCost', title = 'Histogram of Heart Disease vs No Doctor because of Cost')
variable_name='NoDocbcCost'
response_name='HeartDiseaseorAttack'

ax=seq(min(heart_disease[,variable_name]),max(heart_disease[,variable_name]),length.out=10)
ax
hgA = hist(heart_disease[heart_disease[,response_name]==0,variable_name], breaks = ax, plot = FALSE)
hgB = hist(heart_disease[heart_disease[,response_name]==1,variable_name], breaks = ax, plot = FALSE)

table1=rbind('FALSE'=hgA$counts,'TRUE'=hgB$counts)

colnames(table1)=round(hgA$breaks[-1],0)
# add row sum
table1=cbind(table1, 'Total'=apply(table1,1,sum))
# add column sum
table1=rbind(table1, 'Total'=apply(table1,2,sum))
# add odds
table1=rbind(table1, 'Odds'=round(table1[1,]/table1[2,],2))
to_print='Odds'=round(table1[1,]/table1[2,],2)
to_print_ratio='ratio'=round(table1[2,]/table1[1,],2)

table1
library(ggplot2)
ggplot(data=heart_disease, aes(x=GenHlth,fill = HeartDiseaseorAttack)) +
  geom_histogram(bins=20,position='stack')+
  labs(x = 'GenHlth', title = 'Histogram of Heart Disease vs General Health')
variable_name='GenHlth'
response_name='HeartDiseaseorAttack'

ax=seq(min(heart_disease[,variable_name]),max(heart_disease[,variable_name]),length.out=10)
ax
hgA = hist(heart_disease[heart_disease[,response_name]==0,variable_name], breaks = ax, plot = FALSE)
hgB = hist(heart_disease[heart_disease[,response_name]==1,variable_name], breaks = ax, plot = FALSE)

table1=rbind('FALSE'=hgA$counts,'TRUE'=hgB$counts)

colnames(table1)=round(hgA$breaks[-1],0)

```



```

# add row sum
table1=cbind(table1, 'Total'=apply(table1,1,sum))
# add column sum
table1=rbind(table1, 'Total'=apply(table1,2,sum))
# add odds
table1=rbind(table1, 'Odds'=round(table1[1,]/table1[2,],2))
to_print='Odds'=round(table1[1,]/table1[2,],2)
to_print_ratio='ratio'=round(table1[2,]/table1[1,],2)

table1
library(ggplot2)
ggplot(data=heart_disease, aes(x=MentHlth,fill = HeartDiseaseorAttack)) +
  geom_histogram(bins=20,position='stack')+
  labs(x = 'MentHlth', title = 'Histogram of Heart Disease vs Mental Health')
variable_name='MentHlth'
response_name='HeartDiseaseorAttack'

ax=seq(min(heart_disease[,variable_name]),max(heart_disease[,variable_name]),length.out=10)
ax
hgA = hist(heart_disease[heart_disease[,response_name]==0,variable_name], breaks = ax, plot = FALSE)
hgB = hist(heart_disease[heart_disease[,response_name]==1,variable_name], breaks = ax, plot = FALSE)

table1=rbind('FALSE'=hgA$counts,'TRUE'=hgB$counts)

colnames(table1)=round(hgA$breaks[-1],0)
# add row sum
table1=cbind(table1, 'Total'=apply(table1,1,sum))
# add column sum
table1=rbind(table1, 'Total'=apply(table1,2,sum))
# add odds
table1=rbind(table1, 'Odds'=round(table1[1,]/table1[2,],2))
to_print='Odds'=round(table1[1,]/table1[2,],2)
to_print_ratio='ratio'=round(table1[2,]/table1[1,],2)

table1
library(ggplot2)
ggplot(data=heart_disease, aes(x=PhysHlth,fill = HeartDiseaseorAttack)) +
  geom_histogram(bins=20,position='stack')+
  labs(x = 'PhysHlth', title = 'Histogram of Heart Disease vs Physical Health')
variable_name='PhysHlth'
response_name='HeartDiseaseorAttack'

ax=seq(min(heart_disease[,variable_name]),max(heart_disease[,variable_name]),length.out=10)
ax
hgA = hist(heart_disease[heart_disease[,response_name]==0,variable_name], breaks = ax, plot = FALSE)
hgB = hist(heart_disease[heart_disease[,response_name]==1,variable_name], breaks = ax, plot = FALSE)

table1=rbind('FALSE'=hgA$counts,'TRUE'=hgB$counts)

colnames(table1)=round(hgA$breaks[-1],0)
# add row sum
table1=cbind(table1, 'Total'=apply(table1,1,sum))
# add column sum

```

```

table1=rbind(table1, 'Total'=apply(table1,2,sum))
# add odds
table1=rbind(table1, 'Odds'=round(table1[1,]/table1[2,],2))
to_print='Odds'=round(table1[1,]/table1[2,],2)
to_print_ratio='ratio'=round(table1[2,]/table1[1,],2)

table1
library(ggplot2)
ggplot(data=heart_disease, aes(x=Sex,fill = HeartDiseaseorAttack)) +
  geom_histogram(bins=3,position='stack')+
  labs(x = 'Sex', title = 'Histogram of Sex')
variable_name='Sex'
response_name='HeartDiseaseorAttack'

ax=seq(min(heart_disease[,variable_name]),max(heart_disease[,variable_name]),length.out=10)
ax
hgA = hist(heart_disease[heart_disease[,response_name]==0,variable_name], breaks = ax, plot = FALSE)
hgB = hist(heart_disease[heart_disease[,response_name]==1,variable_name], breaks = ax, plot = FALSE)

table1=rbind('FALSE'=hgA$counts,'TRUE'=hgB$counts)

colnames(table1)=round(hgA$breaks[-1],0)
# add row sum
table1=cbind(table1, 'Total'=apply(table1,1,sum))
# add column sum
table1=rbind(table1, 'Total'=apply(table1,2,sum))
# add odds
table1=rbind(table1, 'Odds'=round(table1[1,]/table1[2,],2))
to_print='Odds'=round(table1[1,]/table1[2,],2)
to_print_ratio='ratio'=round(table1[2,]/table1[1,],2)

table1
library(ggplot2)
ggplot(data=heart_disease, aes(x=Income,fill = HeartDiseaseorAttack)) +
  geom_histogram(bins=20,position='stack')+
  labs(x = 'Income', title = 'Histogram of Heart Disease vs Income')
variable_name='Income'
response_name='HeartDiseaseorAttack'

ax=seq(min(heart_disease[,variable_name]),max(heart_disease[,variable_name]),length.out=10)
ax
hgA = hist(heart_disease[heart_disease[,response_name]==0,variable_name], breaks = ax, plot = FALSE)
hgB = hist(heart_disease[heart_disease[,response_name]==1,variable_name], breaks = ax, plot = FALSE)

table1=rbind('FALSE'=hgA$counts,'TRUE'=hgB$counts)

colnames(table1)=round(hgA$breaks[-1],0)
# add row sum
table1=cbind(table1, 'Total'=apply(table1,1,sum))
# add column sum
table1=rbind(table1, 'Total'=apply(table1,2,sum))
# add odds
table1=rbind(table1, 'Odds'=round(table1[1,]/table1[2,],2))

```

```
to_print='Odds'=round(table1[1,]/table1[2,],2)
to_print_ratio='ratio'=round(table1[2,]/table1[1,],2)

table1
```