

Midterm Project

Bao Hu

2024-04-03

To make Y to one dimensional variable:

Categorize the output (Y) to 8 classes, a 8*3 matrix

Ideas:

- we can compare the group of people with each response outcome only
- try different combinations to find information or pattern
- how to differentiate different groups
- whether gender play some role for Y

```
df <- read.csv("heart_disease_health_indicators_BRFSS2015.csv")
```

```
df_dimensions <- dim(df)
df_dimensions
```

```
## [1] 253680      22
```

```
# get name of column and find the target variables
```

```
names <- colnames(df)
names
```

```
## [1] "HeartDiseaseorAttack" "HighBP" "HighChol"
## [4] "CholCheck" "BMI" "Smoker"
## [7] "Stroke" "Diabetes" "PhysActivity"
## [10] "Fruits" "Veggies" "HvyAlcoholConsump"
## [13] "AnyHealthcare" "NoDocbcCost" "GenHlth"
## [16] "MentHlth" "PhysHlth" "DiffWalk"
## [19] "Sex" "Age" "Education"
## [22] "Income"
```

```
dia <- df["Diabetes"]
stro <- df["Stroke"]
heart <- df["HeartDiseaseorAttack"]
```

```
## [1] FALSE
```

```
## [1] 0
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.