Final Project

S. McGowan

2025-07-29

Ideal Cardiovascular Health: An Analysis of Survey Data from Myanmar, 2021

R Coding Final Project

Simone McGowan

Georgia Tech, Summer 2025

GitHub Profile: "Simone-engineers" ; repository: "ECON-4001"

These data were obtained from the distributor, the **Inter-university Consortium for Political and Social Research**. It was collected in a survey format, to evaluate Metabolic Syndrome. The information file in the hyperlink describes factor variable value assignments. The raw data will be used to assess our outcome of interest, "ideal cardiovascular health" status based on metrics from the American Heart Association.

Table of Contents

Literature Review

Background

Metabolic syndrome is a reversible state comprised of having 3 of the 5 health conditions (elevated waist circumference, high triglycerides, high blood pressure, low HDL cholesterol, high fasting glucose) that puts a person at risk of cardiovascular disease (CD), stroke, and type 2 diabetes (Cleveland Clinic, 2023). Though the data was intended to evaluate risk of these three health outcomes, the purpose of the present analysis is to hone in on current cardiovascular health. In Myanmar, Stroke and Ischaemic Heart Disease were in the top 4 leading causes of death amongst men and women in 2021 (World Health Organization, n.d). It is therefore important to identify patterns in the population as a whole, as well as identify vulnerable populations, so that action can be taken on a community and policy/public health level (Lloyd-Jones et al., 2010). The presence of Ideal Heart health is linked with lower rates of mortality due to cardiovascular disease, as well as lower mortality rates due to diseases such as cancer. It is also helpful in reducing risk of cardiac events and stroke in pediatric populations later in life. An Ideal Heart health score will be created using ranges provided by the American Heart Association; relevant metrics include: BMI, total cholesterol, smoking status, physical activity, healthy diet, fasting plasma glucose, and blood pressure, with age determining cutoff values. (Lloyd-Jones et al., 2010)

Socioeconomic status in Myanmar

According to the website Humanitarian Action the absolute poverty line is \$2.15 per person per day in Myanmar, which is (2.15 USD per day x 4402.32 kyat/USD x 30 days) around 283949.64 kyats per person per month using today's exchange rate (OCHA, 2024). The socioeconomic status will be assessed by dividing monthly income of a person by number of children plus them self, to determine if they are at, above or below the poverty line. There is not enough information to assess marital contribution to household income.

Data Description

The data sources are linked in the abstract, coming from **Inter-university Consortium for Political and Social Research** hosted by the University of Michigan. The research was done by Su Su Maw from the University of Nursing in Yangon, Myanmar.

Data Cleaning

```r
#load needed libraries and import raw data file
library(readr)
library(dplyr)
library(ggplot2)

setwd('C:/Users/smdot/OneDrive/Desktop/R-Summer25/Data/raw')
mydata=read.csv("MetS data to upload.csv", header=TRUE, stringsAsFactors =FALSE)
```

```r
#kept only variables in new data frame called new_data
#rather than removing a lot of variables that were unusable upon examination or context
new_data_indeces=c(7,17,18,21,22,23,24,25,26,27,28,29,31,32,33,45,46,47,50,58)
new_data=mydata[new_data_indeces]
#confirmed that 20 variables that I want to keep remain.
#renaming variables and removing NULL values no longer applicable because
#everything I have kept has a value and proper name
```

```r
#verify ChildrenNo and income are numeric
#monthly income/(number children+self) and rename column adjusted income.
class(new_data$ChildrenNo)
```

```
## [1] "integer"
```

```r
class(new_data$Income)
```

```
## [1] "integer"
```

```r
names(new_data)[1]="adjustedIncome"
new_data$adjustedIncome=new_data$adjustedIncome/(new_data$ChildrenNo+1)
```

```r
# create a filter comparing adjusted income with the poverty line (283949.64 kyats/person).
#If more than 10% greater Assign 3 for above, 1  for 10% below poverty line and 2 for those
#around the poverty line (arbitrary plus or minus 10%). The new factor variable will be
#called social status.

upperClassrows=which(new_data$adjustedIncome>1.1*283949.64)
belowPovertyrows=which(new_data$adjustedIncome<0.9*283949.64)
aroundPovertyrows=which(0.9*283949.64<=new_data$adjustedIncome &
new_data$adjustedIncome<=1.1*283949.64)
new_data$socialstatus=0
new_data$socialstatus[upperClassrows]=3
new_data$socialstatus[aroundPovertyrows]=2
new_data$socialstatus[belowPovertyrows]=1
as.factor(new_data$socialstatus)
```

```r
#create new dataframe with lifestyle factor variables only
#switch values of VegFruit and Physical Exercise so that yes=1 and no=0 as with
#other lifestyle factor variables

factors_df=new_data[,c(16,17,18,19,20)]
yesVegFruit=which(new_data$VegFruit==0)
noVegFruit=which(new_data$VegFruit==1)
factors_df$VegFruit[noVegFruit]=0
factors_df$VegFruit[yesVegFruit]=1

yesExercise=which(new_data$PhysicalExercise==0)
noExercise=which(new_data$PhysicalExercise==1)
factors_df$PhysicalExercise[noExercise]=0
factors_df$PhysicalExercise[yesExercise]=1

as.factor(factors_df$Oilymeal)
as.factor(factors_df$Saltymeal)
as.factor(factors_df$VegFruit)
as.factor(factors_df$Pastsmoking)
as.factor(factors_df$PhysicalExercise)
```

```r
#healthyChol, healthyBP, getsExercise, healthyEater, nonSmoker,
#healthyFBS (fasting blood sugar), healthyBMI will have values of 1 if
#they have met AHA health standards in that category. idealHeartscore will
#be the sum of these with a value of 0-7.
#use dplyr syntax to create filter for each metric to apply in each variable
```

Midterm Goals Status Summary

Summary Statistics for numeric variables of interest

```
summary_stats=summary(new_data[c(1:15)])
summary_stats
```

```
## adjustedIncome         Age          ChildrenNo        HbA1c
## Min.   :      0   Min.   :18.00   Min.   : 0.00   Min.   : 4.600
## 1st Qu.:      0   1st Qu.:33.00   1st Qu.: 1.00   1st Qu.: 5.500
## Median :  20000   Median :43.00   Median : 2.00   Median : 5.900
## Mean   :  59421   Mean   :44.07   Mean   : 1.95   Mean   : 6.124
## 3rd Qu.: 100000   3rd Qu.:55.00   3rd Qu.: 3.00   3rd Qu.: 6.400
## Max.   :1400000   Max.   :83.00   Max.   :14.00   Max.   :12.900
##      FBS         TotalCholesterol Triglycerides        HDL
## Min.   : 66.0   Min.   : 96.0    Min.   : 36.0    Min.   : 28.00
## 1st Qu.: 87.0   1st Qu.:165.2    1st Qu.: 88.0    1st Qu.: 48.00
## Median : 97.0   Median :191.5    Median :120.5    Median : 57.00
## Mean   :107.5   Mean   :193.9    Mean   :142.2    Mean   : 57.88
## 3rd Qu.:108.0   3rd Qu.:222.8    3rd Qu.:167.0    3rd Qu.: 64.00
## Max.   :389.0   Max.   :303.0    Max.   :593.0    Max.   :252.00
##      LDL          Bodyweight        Height           BMI
## Min.   : 28.0   Min.   : 31.50   Min.   :1.850   Min.   :15.13
## 1st Qu.: 84.5   1st Qu.: 52.90   1st Qu.:2.280   1st Qu.:22.02
## Median :107.0   Median : 61.10   Median :2.430   Median :25.20
## Mean   :108.4   Mean   : 62.45   Mean   :2.446   Mean   :25.54
## 3rd Qu.:130.0   3rd Qu.: 70.90   3rd Qu.:2.590   3rd Qu.:28.95
## Max.   :240.0   Max.   :100.90   Max.   :3.310   Max.   :40.57
##   SystolicBP     DiastolicBP     Sleepduration
## Min.   : 95.0   Min.   : 49.00   Min.   : 4.000
## 1st Qu.:116.0   1st Qu.: 75.00   1st Qu.: 7.000
## Median :127.0   Median : 82.00   Median : 8.000
## Mean   :128.4   Mean   : 82.58   Mean   : 7.912
## 3rd Qu.:138.8   3rd Qu.: 89.00   3rd Qu.: 9.000
## Max.   :192.0   Max.   :125.00   Max.   :12.000
```

Analysis

You can also embed plots, for example:
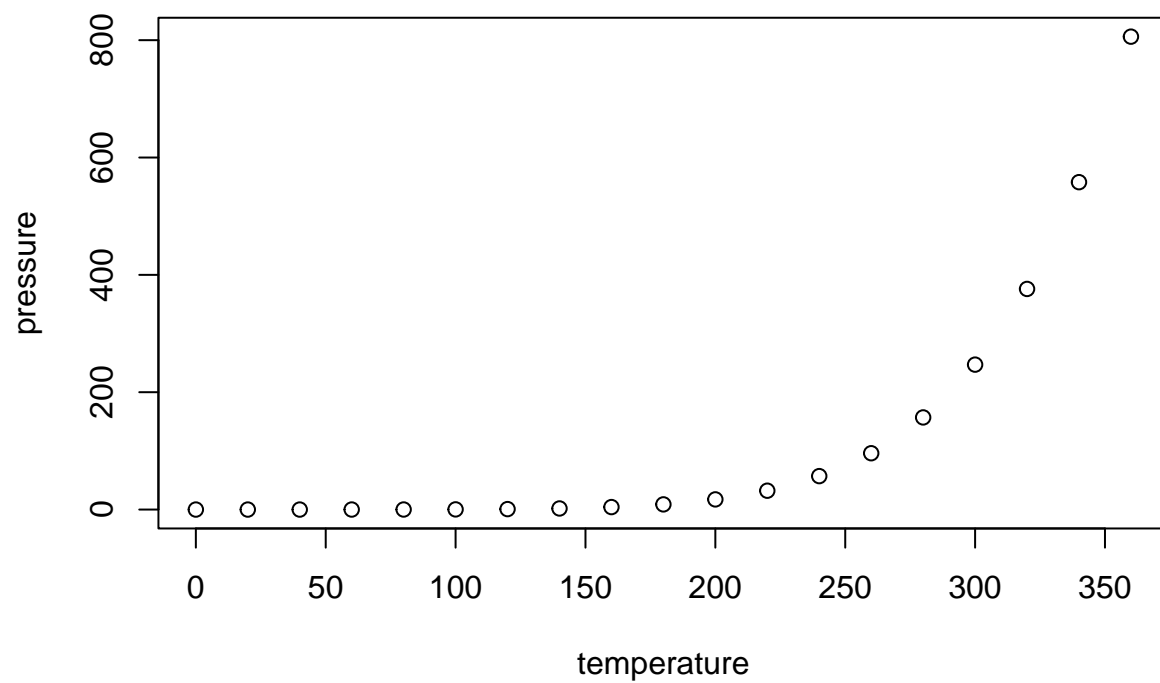
Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

#Beyond Our R

#References copy and paste revised bibliography