SEIS 631 Final-project

Nassangnan Kamagate

5/11/2022

Intro

In this project,! will gain insight into public health by generating simple graphical and numerical summaries of a data set collected by the collected from different hospitals , community clinics, maternal health cares through the IoT based risk monitoring system.

What

I would like to determine the risk level of pregnant women of 18 and less. To help me,i will be using a data set from Kaggle.

Why

It is important to know if they are on high risk level so that we can make some sensitization and also work on how to find some solution to protect them and fight against those condition

How

I will use R Markdown to determine the rate. I would like to go further by determining Which health conditions are the strongest indications for health risks during pregnancy?

Body

The purpose of this will the be to analyse and calculate the risk level faced by pregnant women under the age of 18 and know which category of the conditions affect them the most . By doing this study we can learn and try to protect their pregnancy .The differ e nt attributes are : Systolic BP: Upp er value of Blood Pressure in mmHg, another significant attribute during pregnancy.

Diastolic BP: Lower value of Blood Pressure in mmHg, another significant attribute during pregnancy.

BS: Blood glucose levels is in terms of a molar concentration, mmol/ L. Heart Rate: A normal resting heart rate in beats per minute .

Risk Level: Predicted Risk Intensity Level during pregnancy considering the previous attribute.

study of the Data

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.1.3
```

```
setwd('C:/Users/kamag/Downloads/')
MAT <- read_excel("MAT.xlsx")
library(Hmisc)
```

```
## Warning: package 'Hmisc' was built under R version 4.1.3
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Warning: package 'survival' was built under R version 4.1.3
```

```
## Loading required package: Formula
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 4.1.3
```

```
##
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
describe(MAT)
```

```
## MAT
##
##  7  Variables      1014  Observations
## ---------------------------------------------------------------------------
## Age
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1014        0       50    0.998    29.87    14.93    13.65    15.00
##      .25      .50      .75      .90      .95
##    19.00    26.00    39.00    50.00    55.00
##
## lowest : 10 12 13 14 15, highest: 62 63 65 66 70
## ---------------------------------------------------------------------------
## SystolicBP
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1014        0       19    0.907    113.2       20       85       90
##      .25      .50      .75      .90      .95
##      100      120      120      140      140
##
```

```
## lowest :  70  75  76  78  80, highest: 129 130 135 140 160
##
## Value         70    75    76    78    80    83    85    90    95    99   100
## Frequency      7     8    16     3     5     2    43   154    12     2    92
## Proportion 0.007 0.008 0.016 0.003 0.005 0.002 0.042 0.152 0.012 0.002 0.091
##
## Value        110   115   120   129   130   135   140   160
## Frequency     19     8   449     1    60     3   120    10
## Proportion 0.019 0.008 0.443 0.001 0.059 0.003 0.118 0.010
## ------------------------------------------------------------------------------
## DiastolicBP
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1014        0       16    0.978    76.46     15.8       60       60
##      .25      .50      .75      .90      .95
##       65       80       90       95      100
##
## lowest :  49  50  60  63  65, highest:  85  89  90  95 100
##
## Value         49    50    60    63    65    68    69    70    75    76    80
## Frequency     25    24   174     8    87     2     1   100    38     3   226
## Proportion 0.025 0.024 0.172 0.008 0.086 0.002 0.001 0.099 0.037 0.003 0.223
##
## Value         85    89    90    95   100
## Frequency     49     1   153    36    87
## Proportion 0.048 0.001 0.151 0.036 0.086
## ------------------------------------------------------------------------------
## BS
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1014        0       29    0.991    8.726    2.979     6.10     6.70
##      .25      .50      .75      .90      .95
##     6.90     7.50     8.00    15.00    17.35
##
## lowest :  6.0  6.1  6.3  6.4  6.5, highest: 15.0 16.0 17.0 18.0 19.0
## ------------------------------------------------------------------------------
## BodyTemp
##        n  missing distinct     Info     Mean      Gmd
##     1014        0        8      0.5    98.67    1.098
##
## lowest :  98.0  98.4  98.6  99.0 100.0, highest:  99.0 100.0 101.0 102.0 103.0
##
## Value       98.0  98.4  98.6  99.0 100.0 101.0 102.0 103.0
## Frequency    804     2     1    10    20    98    66    13
## Proportion 0.793 0.002 0.001 0.010 0.020 0.097 0.065 0.013
## ------------------------------------------------------------------------------
## HeartRate
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1014        0       16    0.975     74.3    8.653       60       66
##      .25      .50      .75      .90      .95
##       70       76       80       86       88
##
## lowest :  7 60 65 66 67, highest: 80 82 86 88 90
##
## Value          7    60    65    66    67    68    70    75    76    77    78
## Frequency      2    74     5    87    12     2   271    19   131    96    46
```

```
## Proportion 0.002 0.073 0.005 0.086 0.012 0.002 0.267 0.019 0.129 0.095 0.045
##
## Value         80    82    86    88    90
## Frequency    117    19    55    59    19
## Proportion 0.115 0.019 0.054 0.058 0.019
## ---------------------------------------------------------------------------
## RiskLevel
##        n  missing distinct
##     1014        0        3
##
## Value      high risk  low risk  mid risk
## Frequency        272       406       336
## Proportion     0.268     0.400     0.331
## ---------------------------------------------------------------------------
```

In this that data we have 7 variable for 1014 Oobservation

Distribution of Risk Level

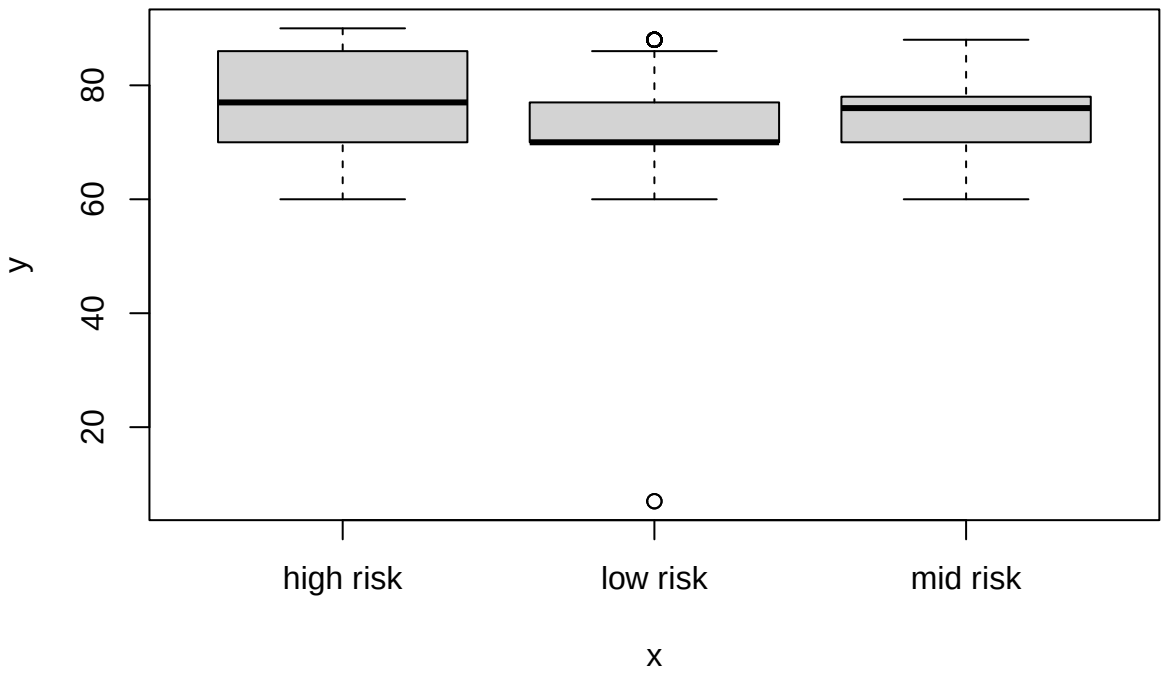
```
table(MAT$RiskLevel)
```

```
##
## high risk  low risk  mid risk
##       272       406       336
```

We know from the distribution that most of the pregnant women have A low risk conditions .

I used this plot to represent the difference between the three different levels .

```
RiskLevel <- table(MAT$RiskLevel)
plot(as.factor(MAT$RiskLevel),MAT$HeartRate)
```



As we can see high risk level have the bigger proportion.

```
anova(aov(HeartRate ~ RiskLevel, data=MAT))
```

```
## Analysis of Variance Table
##
## Response: HeartRate
##             Df Sum Sq Mean Sq F value    Pr(>F)
## RiskLevel    2   2577 1288.67  20.453 1.961e-09 ***
## Residuals 1011  63700   63.01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The function head and tail will help me get the first row and last of the data set.

```
head(MAT)
```

```
## # A tibble: 6 x 7
##     Age SystolicBP DiastolicBP    BS BodyTemp HeartRate RiskLevel
##   <dbl>      <dbl>       <dbl> <dbl>    <dbl>     <dbl> <chr>
## 1    25        130          80 15          98        86 high risk
## 2    35        140          90 13          98        70 high risk
## 3    29         90          70  8         100        80 high risk
## 4    30        140          85  7          98        70 high risk
## 5    35        120          60  6.1        98        76 low risk
## 6    23        140          80  7.01       98        70 high risk
```

```
tail(MAT)
```

```
## # A tibble: 6 x 7
##     Age SystolicBP DiastolicBP    BS BodyTemp HeartRate RiskLevel
##   <dbl>      <dbl>       <dbl> <dbl>    <dbl>     <dbl> <chr>
## 1    48        120          80    11       98        88 high risk
## 2    22        120          60    15       98        80 high risk
## 3    55        120          90    18       98        60 high risk
## 4    35         85          60    19       98        86 high risk
## 5    43        120          90    18       98        70 high risk
## 6    32        120          65     6      101        76 mid risk
```

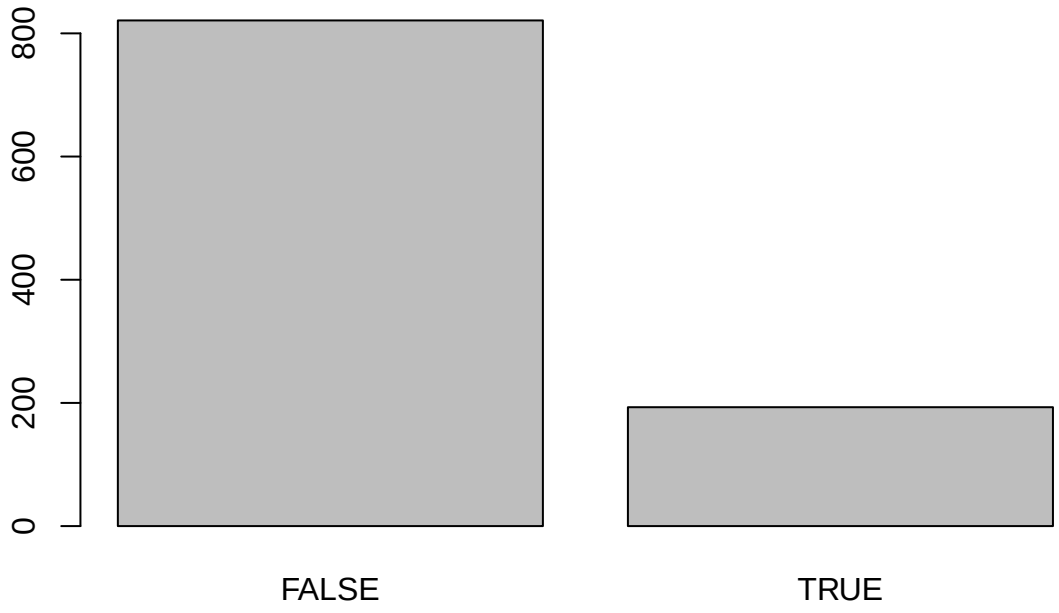
Pregnant women under 18

Let's start by plotting a table of girl under 18.

```
table(MAT$Age<18)
```

```
##
## FALSE  TRUE
##   821   193
```

```
Age <- table(MAT$Age<18)
barplot(Age)
```



We can see from that only 193 girl are under 18.

Pregnant women under 18 with high risk level

Here I will divided the data into different subset.The first one will extracting the group of girl under 18 from the data set.

```
ltage18 <- subset(MAT, Age<18)
ltage18
```

```
## # A tibble: 193 x 7
##      Age SystolicBP DiastolicBP    BS BodyTemp HeartRate RiskLevel
##    <dbl>      <dbl>       <dbl> <dbl>    <dbl>     <dbl> <chr>
##  1    15        120          80  7.01       98        70 low risk
##  2    10         70          50  6.9        98        70 low risk
##  3    16        100          70  7.2        98        80 low risk
##  4    12         95          60  6.1       102        60 low risk
##  5    15         76          49  7.5        98        77 low risk
##  6    15        120          80  7          98        70 low risk
##  7    15         76          49  6.4        98        77 low risk
##  8    15        120          80  7.2        98        70 low risk
##  9    15         80          60  7          98        80 low risk
## 10    12         95          60  7.2        98        77 low risk
## # ... with 183 more rows
```

I will create a subset representing the group of women having high risk condition.

```
MAT$RiskLevel == "high risk"
```

```
##    [1]  TRUE  TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE FALSE
##   [13] FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE
##   [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##   [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##   [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##   [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##   [73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##   [85] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##   [97] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##  [109]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##  [121]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##  [133]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE FALSE FALSE FALSE
##  [145] FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE
##  [157] FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE
##  [169] FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE FALSE
##  [181] FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
##  [193]  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE
##  [205] FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
##  [217] FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
##  [229]  TRUE FALSE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##  [241]  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE FALSE FALSE
##  [253] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE
##  [265] FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE FALSE
```

```
##  [277]  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE
##  [289] FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE
##  [301] FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [313] FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE
##  [325] FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE
##  [337] FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE
##  [349]  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE
##  [361] FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE  TRUE  TRUE FALSE  TRUE
##  [373] FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE
##  [385] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
##  [397] FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [409] FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE  TRUE
##  [421] FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE FALSE  TRUE FALSE FALSE  TRUE
##  [433]  TRUE FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE FALSE
##  [445] FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE
##  [457]  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
##  [469]  TRUE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE
##  [481] FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE
##  [493] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE
##  [505]  TRUE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [517] FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [529] FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE
##  [541] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE
##  [553] FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [565] FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [577] FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE
##  [589] FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE
##  [601]  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE
##  [613] FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
##  [625] FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE
##  [637]  TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE
##  [649] FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE
##  [661] FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [673] FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##  [685]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE
##  [697] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [709] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [721] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [733] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [745] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [757] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [769] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [781] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [793] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [805] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [817] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [829] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [841] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [853] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [865] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [877] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [889] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [901] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [913] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
##  [925] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [937] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [949] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##  [961]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##  [973]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##  [985]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##  [997]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [1009]  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
```

```
hr <- MAT$RiskLevel == "high risk"
hr <- subset(MAT, MAT$RiskLevel == "high risk")
```

Now I will combine both of the subset to create a subset about the women under 18 with high risk condition.

```
p18_highrisk <- subset(ltage18, ltage18$RiskLevel == "high risk")
```

From the research above we can see that 37 seven person out 272 from the list of women under 18 are in high risk . This represent 13% of the list . It's a small percentage but it is still not to be be neglected.

Study of the different condition

I will use the summary function to help me study the different category.

DiastolicBP

```
summary(MAT$DiastolicBP)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   49.00   65.00   80.00   76.46   90.00  100.00
```

From the result we can see that the minimum upper value of blood pressure in 49 and the maximum is 100. The mean which represent the average is 76.46.

systolicBP

```
summary(MAT$SystolicBP)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    70.0   100.0   120.0   113.2   120.0   160.0
```

From the result we can see that the minimum lower value blood pressure in 70 and the maximum is 160. The mean which represent the average is 113.2

BS

```
summary(MAT$BS)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   6.000   6.900   7.500   8.726   8.000  19.000
```

From the result we can see that the minimum blood glucose levels in 6 and the maximum is 19. The mean which represent the average is 8.72.

Heart Rate

```
summary(MAT$HeartRate)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     7.0    70.0    76.0    74.3    80.0    90.0
```

The result indicate that the minimum heart rate is 7 and the maximum is 90. The mean which represent the average is 74.30

Topics From Class

Rmarkdown:

One of my favorite things to use this semester, it is my first semester in Business Analytics I was scared at the beginning but really after having one class I really loved it.

Github:

This is an also something I enjoy learning I did not go yet as far I wanted to go for now but I will like to learn more about it. ## HMISC package: I used the package hmisc to describe the Data have and more incite of the Data.

Anova:

I used this topics from the suggestion of one of the student and it came out great. It really help me have create a great visualization .

Data subsetting:

This one of my favorite I really enjoy using it in my homeworks so I decide to use it for my project and it really me acheive what I was looking for.

Conclusion

This project help me learn a lot , it help me surpass myself and succeed my research. It help me learn new function also go back to what I learned in class and study them more. It help me advance my knowledge This is my first semester ,I was scared at beginning but i learned that nothing is too hard or too easy we just have be ready to work harder.

Citation:

HMISC: Donovan, K. (2019, July 11). Data Analysis and processing with R based on Ibis Data. 9 Documenting your results with R Markdown. Retrieved May 8, 2022, from https://bookdown.org/kdonovan125/ ibis_data_analysis_ r4/ documenting-your-results-with- r-markdown.html