R ASSIGNMENT

Olympia Mugarura

2023-11-09

SECTION A

Reading metaFile.csv into R

```
setwd("C:/Users/OLYMPIA/Downloads/fwdmsb7102bioconductorandrprogrammingassignment")
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.3.2

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library("ggplot2")
```

```
## Warning: package 'ggplot2' was built under R version 4.3.2
```

```
metadata <- read.csv("metaFile.csv")
print(metadata)
```

```
##    sample_id Participant Treatment
## 1   sample_1           C         4
## 2   sample_2           B         3
## 3   sample_3           B         2
## 4   sample_4           C         1
## 5   sample_5           C         3
## 6   sample_6           A         4
## 7   sample_7           B         0
## 8   sample_8           A         1
## 9   sample_9           B         4
## 10 sample_10           C         0
## 11 sample_11           A         2
## 12 sample_12           A         3
## 13 sample_13           B         1
## 14 sample_14           C         2
## 15 sample_15           A         0
## 16 sample_16           C         4
## 17 sample_17           B         3
## 18 sample_18           B         2
## 19 sample_19           C         1
## 20 sample_20           C         3
## 21 sample_21           A         4
## 22 sample_22           B         0
## 23 sample_23           A         1
## 24 sample_24           B         4
## 25 sample_25           C         0
## 26 sample_26           A         2
## 27 sample_27           A         3
## 28 sample_28           B         1
## 29 sample_29           C         2
## 30 sample_30           A         0
## 31 sample_31           C         4
## 32 sample_32           B         3
## 33 sample_33           B         2
## 34 sample_34           C         1
## 35 sample_35           C         3
## 36 sample_36           A         4
## 37 sample_37           B         0
## 38 sample_38           A         1
## 39 sample_39           B         4
## 40 sample_40           C         0
## 41 sample_41           A         2
## 42 sample_42           A         3
## 43 sample_43           B         1
## 44 sample_44           C         2
## 45 sample_45           A         0
## 46 sample_46           C         4
## 47 sample_47           B         3
## 48 sample_48           B         2
## 49 sample_49           C         1
## 50 sample_50           C         3
```

(i) number of variables/columns and records/rows

```
ncol(metadata)
```

```
## [1] 3
```

```
nrow(metadata)
```

```
## [1] 50
```

(ii) number of samples obtained for each level of treatment

```
table(metadata$Treatment)
```

```
## 
##  0  1  2  3  4 
##  9 10 10 11 10
```

(iii) number of samples obtained for each participate

```
table(metadata$Participant)
```

```
## 
##  A  B  C 
## 15 17 18
```

(iv) extracting records for samples obtained at treatment levels 0, 2 and 4

```
metadata[metadata$Treatment %in% c(0, 2, 4),]
```

```
##    sample_id Participant Treatment
## 1   sample_1           C         4
## 3   sample_3           B         2
## 6   sample_6           A         4
## 7   sample_7           B         0
## 9   sample_9           B         4
## 10 sample_10           C         0
## 11 sample_11           A         2
## 14 sample_14           C         2
## 15 sample_15           A         0
## 16 sample_16           C         4
## 18 sample_18           B         2
## 21 sample_21           A         4
## 22 sample_22           B         0
## 24 sample_24           B         4
## 25 sample_25           C         0
## 26 sample_26           A         2
## 29 sample_29           C         2
## 30 sample_30           A         0
## 31 sample_31           C         4
## 33 sample_33           B         2
## 36 sample_36           A         4
## 37 sample_37           B         0
## 39 sample_39           B         4
## 40 sample_40           C         0
## 41 sample_41           A         2
## 44 sample_44           C         2
## 45 sample_45           A         0
## 46 sample_46           C         4
## 48 sample_48           B         2
```

(v) extracting records for samples obtained from participant A and C

```
metadata %>% filter(Participant %in% c("A", "C"))
```

```
##    sample_id Participant Treatment
## 1   sample_1           C         4
## 2   sample_4           C         1
## 3   sample_5           C         3
## 4   sample_6           A         4
## 5   sample_8           A         1
## 6  sample_10           C         0
## 7  sample_11           A         2
## 8  sample_12           A         3
## 9  sample_14           C         2
## 10 sample_15           A         0
## 11 sample_16           C         4
## 12 sample_19           C         1
## 13 sample_20           C         3
## 14 sample_21           A         4
## 15 sample_23           A         1
## 16 sample_25           C         0
## 17 sample_26           A         2
## 18 sample_27           A         3
## 19 sample_29           C         2
## 20 sample_30           A         0
## 21 sample_31           C         4
## 22 sample_34           C         1
## 23 sample_35           C         3
## 24 sample_36           A         4
## 25 sample_38           A         1
## 26 sample_40           C         0
## 27 sample_41           A         2
## 28 sample_42           A         3
## 29 sample_44           C         2
## 30 sample_45           A         0
## 31 sample_46           C         4
## 32 sample_49           C         1
## 33 sample_50           C         3
```

(vi) number of samples under each participant groups per treatment level

```
table(metadata$Participant, metadata$Treatment)
```

```
##
##     0 1 2 3 4
##   A 3 3 3 3 3
##   B 3 3 4 4 3
##   C 3 4 3 4 4
```

SECTION B

```r
setwd("C:/Users/OLYMPIA/Downloads/fwdmsb7102bioconductorandrprogrammingassignment")
Assignmentdata <- read.csv("AssignmentFile.csv")
print(Assignmentdata)
```

```
##      samples      Conc1     Conc2      Conc3
## 1   sample_1 5.93804869 2.2877765  3.2487625
## 2   sample_2 3.39846291 4.0508845  3.2359883
## 3   sample_3 1.94647521 1.2918230  1.2597982
## 4   sample_4 1.75043775 4.5088147  2.6153558
## 5   sample_5 2.90319389 2.6006057  1.6662130
## 6   sample_6 0.74662991 3.5212946  1.3361532
## 7   sample_7 2.61412606 1.7075223  2.9077324
## 8   sample_8 2.86823250 0.6855859  0.7057831
## 9   sample_9 4.25332693 1.5380970  4.1286548
## 10 sample_10 2.53152856 1.4914619  2.8437762
## 11 sample_11 2.02469134 1.0988479  3.0460737
## 12 sample_12 1.68888486 1.5382037  1.5721820
## 13 sample_13 0.07948294 1.8417749  5.4324201
## 14 sample_14 2.35002162 2.8437762  3.5107200
## 15 sample_15 2.39041813 4.7705793  2.6154821
## 16 sample_16 0.61849725 2.2429638  2.3998393
## 17 sample_17 2.67478610 3.2269440  1.4297713
## 18 sample_18 1.31226408 3.2960420  3.3647171
## 19 sample_19 1.51768209 2.1535883  1.6766006
## 20 sample_20 0.40522027 2.5920906  3.1057814
## 21 sample_21 2.11207326 3.4797040  0.4052203
## 22 sample_22 1.41522107 3.1598762  1.9140865
## 23 sample_23 2.48876302 2.8768958  2.2684174
## 24 sample_24 1.88033411 3.1700562  2.3061025
## 25 sample_25 1.60430191 3.2612428  1.4106253
## 26 sample_26 2.74773395 2.8603185  3.4021108
## 27 sample_27 3.10092688 3.8005525  3.0387250
## 28 sample_28 1.50200809 0.8532958  2.9407410
## 29 sample_29 0.96211799 1.7579986  2.1523931
## 30 sample_30 2.14791875 1.9930619  4.9701923
## 31 sample_31 2.29835762 1.8367984  3.2096464
## 32 sample_32 1.79410531 1.8296990  2.4910030
## 33 sample_33 2.38717120 1.7903664  4.3216649
## 34 sample_34 2.18463473 1.4871950  2.1999546
## 35 sample_35 1.97282767 3.3091355 -0.2754937
## 36 sample_36 1.07893276 2.4602617  3.0395745
## 37 sample_37 2.25989664 3.0217533  2.5672832
## 38 sample_38 4.41868056 3.7645799  2.7567790
## 39 sample_39 2.41582396 3.8140188  2.2961438
## 40 sample_40 1.16442633 3.7059658  3.9772471
## 41 sample_41 2.74157875 3.2011782  3.7463223
## 42 sample_42 2.84826466 2.0364330  2.1030553
## 43 sample_43 3.84637857 1.9730300  3.8144462
## 44 sample_44 3.11295419 3.5936637  1.8196195
## 45 sample_45 1.64704709 4.2174458  3.1303307
## 46 sample_46 2.66380292 2.5251532  3.0458905
## 47 sample_47 1.65227360 2.7611225  0.6374504
## 48 sample_48 2.86585757 2.2171705  2.9346651
## 49 sample_49 2.58467415 3.1289636  1.5891398
## 50 sample_50 1.88399429 2.4144523  2.6693480
```

(i) number of variables and records

```
ncol(Assignmentdata)
```

```
## [1] 4
```

```
nrow(Assignmentdata)
```

```
## [1] 50
```

(ii) computing the average concentration for each of the concentrations

```
colMeans(Assignmentdata[, c("Conc1","Conc2","Conc3")])
```

```
##    Conc1    Conc2    Conc3
## 2.235910 2.631801 2.579690
```

(iii) computing the average Conc for each sample

```
rowMeans(Assignmentdata[c("Conc1", "Conc2", "Conc3")])
```

```
##  [1] 3.824863 3.561779 1.499365 2.958203 2.390004 1.868026 2.409794 1.419867
##  [9] 3.306693 2.288922 2.056538 1.599757 2.451226 2.901506 3.258827 1.753767
## [17] 2.443834 2.657674 1.782624 2.034364 1.998999 2.163061 2.544692 2.452164
## [25] 2.092057 3.003388 3.313401 1.765348 1.624170 3.037058 2.448267 2.038269
## [33] 2.833067 1.957261 1.668823 2.192923 2.616311 3.646680 2.841996 2.949213
## [41] 3.229693 2.329251 3.211285 2.842079 2.998275 2.744949 1.683616 2.672564
## [49] 2.434259 2.322598
```

(iv) extending assignmentdata by introducing new variable "Concmean"

```
Assignmentdata$Concmean <- rowMeans(Assignmentdata[c("Conc1", "Conc2", "Conc3")])
print(Assignmentdata)
```

```
##      samples      Conc1     Conc2      Conc3 Concmean
## 1   sample_1 5.93804869 2.2877765  3.2487625 3.824863
## 2   sample_2 3.39846291 4.0508845  3.2359883 3.561779
## 3   sample_3 1.94647521 1.2918230  1.2597982 1.499365
## 4   sample_4 1.75043775 4.5088147  2.6153558 2.958203
## 5   sample_5 2.90319389 2.6006057  1.6662130 2.390004
## 6   sample_6 0.74662991 3.5212946  1.3361532 1.868026
## 7   sample_7 2.61412606 1.7075223  2.9077324 2.409794
## 8   sample_8 2.86823250 0.6855859  0.7057831 1.419867
## 9   sample_9 4.25332693 1.5380970  4.1286548 3.306693
## 10 sample_10 2.53152856 1.4914619  2.8437762 2.288922
## 11 sample_11 2.02469134 1.0988479  3.0460737 2.056538
## 12 sample_12 1.68888486 1.5382037  1.5721820 1.599757
## 13 sample_13 0.07948294 1.8417749  5.4324201 2.451226
## 14 sample_14 2.35002162 2.8437762  3.5107200 2.901506
## 15 sample_15 2.39041813 4.7705793  2.6154821 3.258827
## 16 sample_16 0.61849725 2.2429638  2.3998393 1.753767
## 17 sample_17 2.67478610 3.2269440  1.4297713 2.443834
## 18 sample_18 1.31226408 3.2960420  3.3647171 2.657674
## 19 sample_19 1.51768209 2.1535883  1.6766006 1.782624
## 20 sample_20 0.40522027 2.5920906  3.1057814 2.034364
## 21 sample_21 2.11207326 3.4797040  0.4052203 1.998999
## 22 sample_22 1.41522107 3.1598762  1.9140865 2.163061
## 23 sample_23 2.48876302 2.8768958  2.2684174 2.544692
## 24 sample_24 1.88033411 3.1700562  2.3061025 2.452164
## 25 sample_25 1.60430191 3.2612428  1.4106253 2.092057
## 26 sample_26 2.74773395 2.8603185  3.4021108 3.003388
## 27 sample_27 3.10092688 3.8005525  3.0387250 3.313401
## 28 sample_28 1.50200809 0.8532958  2.9407410 1.765348
## 29 sample_29 0.96211799 1.7579986  2.1523931 1.624170
## 30 sample_30 2.14791875 1.9930619  4.9701923 3.037058
## 31 sample_31 2.29835762 1.8367984  3.2096464 2.448267
## 32 sample_32 1.79410531 1.8296990  2.4910030 2.038269
## 33 sample_33 2.38717120 1.7903664  4.3216649 2.833067
## 34 sample_34 2.18463473 1.4871950  2.1999546 1.957261
## 35 sample_35 1.97282767 3.3091355 -0.2754937 1.668823
## 36 sample_36 1.07893276 2.4602617  3.0395745 2.192923
## 37 sample_37 2.25989664 3.0217533  2.5672832 2.616311
## 38 sample_38 4.41868056 3.7645799  2.7567790 3.646680
## 39 sample_39 2.41582396 3.8140188  2.2961438 2.841996
## 40 sample_40 1.16442633 3.7059658  3.9772471 2.949213
## 41 sample_41 2.74157875 3.2011782  3.7463223 3.229693
## 42 sample_42 2.84826466 2.0364330  2.1030553 2.329251
## 43 sample_43 3.84637857 1.9730300  3.8144462 3.211285
## 44 sample_44 3.11295419 3.5936637  1.8196195 2.842079
## 45 sample_45 1.64704709 4.2174458  3.1303307 2.998275
## 46 sample_46 2.66380292 2.5251532  3.0458905 2.744949
## 47 sample_47 1.65227360 2.7611225  0.6374504 1.683616
## 48 sample_48 2.86585757 2.2171705  2.9346651 2.672564
## 49 sample_49 2.58467415 3.1289636  1.5891398 2.434259
## 50 sample_50 1.88399429 2.4144523  2.6693480 2.322598
```

(vi) obtaining the basic statistics

```
summary(Assignmentdata)
```

```
##    samples              Conc1              Conc2            Conc3        
##  Length:50          Min.   :0.07948   Min.   :0.6856   Min.   :-0.2755  
##  Class :character   1st Qu.:1.64835   1st Qu.:1.8380   1st Qu.: 1.8432  
##  Mode  :character   Median :2.22227   Median :2.5963   Median : 2.6424  
##                     Mean   :2.23591   Mean   :2.6318   Mean   : 2.5797  
##                     3rd Qu.:2.72488   3rd Qu.:3.2873   3rd Qu.: 3.1898  
##                     Max.   :5.93805   Max.   :4.7706   Max.   : 5.4324  
##     Concmean    
##  Min.   :1.420  
##  1st Qu.:2.035
##  Median :2.446
##  Mean   :2.482
##  3rd Qu.:2.937
##  Max.   :3.825
```

```
sapply(Assignmentdata[,2:5], sd)
```

```
##     Conc1     Conc2     Conc3  Concmean
## 1.0384617 0.9653108 1.1169179 0.6018165
```

```
sapply(Assignmentdata[,2:5], var)
```

```
##     Conc1     Conc2     Conc3  Concmean
## 1.0784026 0.9318250 1.2475056 0.3621831
```

SECTION C

(i) removing variables Conc1, Conc2 and Conc3 from the data frame

```
Assignmentdata <- subset(Assignmentdata, select = -c(Conc1, Conc2, Conc3))
print(Assignmentdata)
```

```
##      samples Concmean
## 1   sample_1 3.824863
## 2   sample_2 3.561779
## 3   sample_3 1.499365
## 4   sample_4 2.958203
## 5   sample_5 2.390004
## 6   sample_6 1.868026
## 7   sample_7 2.409794
## 8   sample_8 1.419867
## 9   sample_9 3.306693
## 10 sample_10 2.288922
## 11 sample_11 2.056538
## 12 sample_12 1.599757
## 13 sample_13 2.451226
## 14 sample_14 2.901506
## 15 sample_15 3.258827
## 16 sample_16 1.753767
## 17 sample_17 2.443834
## 18 sample_18 2.657674
## 19 sample_19 1.782624
## 20 sample_20 2.034364
## 21 sample_21 1.998999
## 22 sample_22 2.163061
## 23 sample_23 2.544692
## 24 sample_24 2.452164
## 25 sample_25 2.092057
## 26 sample_26 3.003388
## 27 sample_27 3.313401
## 28 sample_28 1.765348
## 29 sample_29 1.624170
## 30 sample_30 3.037058
## 31 sample_31 2.448267
## 32 sample_32 2.038269
## 33 sample_33 2.833067
## 34 sample_34 1.957261
## 35 sample_35 1.668823
## 36 sample_36 2.192923
## 37 sample_37 2.616311
## 38 sample_38 3.646680
## 39 sample_39 2.841996
## 40 sample_40 2.949213
## 41 sample_41 3.229693
## 42 sample_42 2.329251
## 43 sample_43 3.211285
## 44 sample_44 2.842079
## 45 sample_45 2.998275
## 46 sample_46 2.744949
## 47 sample_47 1.683616
## 48 sample_48 2.672564
## 49 sample_49 2.434259
## 50 sample_50 2.322598
```

(ii) merging information in the two data frames concentration with metadata

```
merged_data <- merge(metadata, Assignmentdata, by.x = "sample_id", by.y = "samples")
print(merged_data)
```

```
##    sample_id Participant Treatment Concmean
## 1   sample_1           C         4 3.824863
## 2  sample_10           C         0 2.288922
## 3  sample_11           A         2 2.056538
## 4  sample_12           A         3 1.599757
## 5  sample_13           B         1 2.451226
## 6  sample_14           C         2 2.901506
## 7  sample_15           A         0 3.258827
## 8  sample_16           C         4 1.753767
## 9  sample_17           B         3 2.443834
## 10 sample_18           B         2 2.657674
## 11 sample_19           C         1 1.782624
## 12  sample_2           B         3 3.561779
## 13 sample_20           C         3 2.034364
## 14 sample_21           A         4 1.998999
## 15 sample_22           B         0 2.163061
## 16 sample_23           A         1 2.544692
## 17 sample_24           B         4 2.452164
## 18 sample_25           C         0 2.092057
## 19 sample_26           A         2 3.003388
## 20 sample_27           A         3 3.313401
## 21 sample_28           B         1 1.765348
## 22 sample_29           C         2 1.624170
## 23  sample_3           B         2 1.499365
## 24 sample_30           A         0 3.037058
## 25 sample_31           C         4 2.448267
## 26 sample_32           B         3 2.038269
## 27 sample_33           B         2 2.833067
## 28 sample_34           C         1 1.957261
## 29 sample_35           C         3 1.668823
## 30 sample_36           A         4 2.192923
## 31 sample_37           B         0 2.616311
## 32 sample_38           A         1 3.646680
## 33 sample_39           B         4 2.841996
## 34  sample_4           C         1 2.958203
## 35 sample_40           C         0 2.949213
## 36 sample_41           A         2 3.229693
## 37 sample_42           A         3 2.329251
## 38 sample_43           B         1 3.211285
## 39 sample_44           C         2 2.842079
## 40 sample_45           A         0 2.998275
## 41 sample_46           C         4 2.744949
## 42 sample_47           B         3 1.683616
## 43 sample_48           B         2 2.672564
## 44 sample_49           C         1 2.434259
## 45  sample_5           C         3 2.390004
## 46 sample_50           C         3 2.322598
## 47  sample_6           A         4 1.868026
## 48  sample_7           B         0 2.409794
## 49  sample_8           A         1 1.419867
## 50  sample_9           B         4 3.306693
```

(iii) checking the size of combined data-frame

```
size <- dim(merged_data)
print(paste("Number of rows:", size[1]))
```

```
## [1] "Number of rows: 50"
```

```
print(paste("Number of columns:", size[2]))
```

```
## [1] "Number of columns: 4"
```

```
print (colnames(merged_data))
```

```
## [1] "sample_id"   "Participant" "Treatment"   "Concmean"
```

```
print(head(merged_data))
```

```
##   sample_id Participant Treatment Concmean
## 1  sample_1           C         4 3.824863
## 2 sample_10           C         0 2.288922
## 3 sample_11           A         2 2.056538
## 4 sample_12           A         3 1.599757
## 5 sample_13           B         1 2.451226
## 6 sample_14           C         2 2.901506
```

(iv) extracting records for samples for participants B and C obtained at treatment levels 2,3 and 4 and Assigning it “participantsBC”

```
participantsBC <- subset(merged_data, Participant %in% c("B", "C") & Treatment %in% c(2, 3, 4))
print(participantsBC)
```

```
##    sample_id Participant Treatment Concmean
## 1   sample_1           C         4 3.824863
## 6  sample_14           C         2 2.901506
## 8  sample_16           C         4 1.753767
## 9  sample_17           B         3 2.443834
## 10 sample_18           B         2 2.657674
## 12  sample_2           B         3 3.561779
## 13 sample_20           C         3 2.034364
## 17 sample_24           B         4 2.452164
## 22 sample_29           C         2 1.624170
## 23  sample_3           B         2 1.499365
## 25 sample_31           C         4 2.448267
## 26 sample_32           B         3 2.038269
## 27 sample_33           B         2 2.833067
## 29 sample_35           C         3 1.668823
## 33 sample_39           B         4 2.841996
## 39 sample_44           C         2 2.842079
## 41 sample_46           C         4 2.744949
## 42 sample_47           B         3 1.683616
## 43 sample_48           B         2 2.672564
## 45  sample_5           C         3 2.390004
## 46 sample_50           C         3 2.322598
## 50  sample_9           B         4 3.306693
```

(v) computing the average concentration for participants B and C under treatments 2, 3 and 4

```
avg_concentration <- participantsBC %>% group_by(Participant, Treatment) %>% summarize(AvgConcentration
```

```
## `summarise()` has grouped output by 'Participant'. You can override using the
## `.groups` argument.
```

```
print(avg_concentration)
```

```
## # A tibble: 6 x 3
## # Groups:   Participant [2]
##   Participant Treatment AvgConcentration
##   <chr>           <int>            <dbl>
## 1 B                   2             2.42
## 2 B                   3             2.43
## 3 B                   4             2.87
## 4 C                   2             2.46
## 5 C                   3             2.10
## 6 C                   4             2.69
```

(vi) changing treatment levels from 0,1,2,3,4 to “very-low”, “low”, “moderate”, “high”, “very-high” using for loop

```
merged_data
```

```
##    sample_id Participant Treatment Concmean
## 1   sample_1           C         4 3.824863
## 2  sample_10           C         0 2.288922
## 3  sample_11           A         2 2.056538
## 4  sample_12           A         3 1.599757
## 5  sample_13           B         1 2.451226
## 6  sample_14           C         2 2.901506
## 7  sample_15           A         0 3.258827
## 8  sample_16           C         4 1.753767
## 9  sample_17           B         3 2.443834
## 10 sample_18           B         2 2.657674
## 11 sample_19           C         1 1.782624
## 12  sample_2           B         3 3.561779
## 13 sample_20           C         3 2.034364
## 14 sample_21           A         4 1.998999
## 15 sample_22           B         0 2.163061
## 16 sample_23           A         1 2.544692
## 17 sample_24           B         4 2.452164
## 18 sample_25           C         0 2.092057
## 19 sample_26           A         2 3.003388
## 20 sample_27           A         3 3.313401
## 21 sample_28           B         1 1.765348
## 22 sample_29           C         2 1.624170
## 23  sample_3           B         2 1.499365
## 24 sample_30           A         0 3.037058
## 25 sample_31           C         4 2.448267
## 26 sample_32           B         3 2.038269
## 27 sample_33           B         2 2.833067
## 28 sample_34           C         1 1.957261
## 29 sample_35           C         3 1.668823
## 30 sample_36           A         4 2.192923
## 31 sample_37           B         0 2.616311
## 32 sample_38           A         1 3.646680
## 33 sample_39           B         4 2.841996
## 34  sample_4           C         1 2.958203
## 35 sample_40           C         0 2.949213
## 36 sample_41           A         2 3.229693
## 37 sample_42           A         3 2.329251
## 38 sample_43           B         1 3.211285
## 39 sample_44           C         2 2.842079
## 40 sample_45           A         0 2.998275
## 41 sample_46           C         4 2.744949
## 42 sample_47           B         3 1.683616
## 43 sample_48           B         2 2.672564
## 44 sample_49           C         1 2.434259
## 45  sample_5           C         3 2.390004
## 46 sample_50           C         3 2.322598
## 47  sample_6           A         4 1.868026
## 48  sample_7           B         0 2.409794
## 49  sample_8           A         1 1.419867
## 50  sample_9           B         4 3.306693
```

```r
for (c in 1:ncol(merged_data)){
  for (r in 1:nrow(merged_data)){
    value = merged_data[r,c]
    if (value == 0){
      merged_data[r, c] = "Very Low"
    } else if (value == 1){
      merged_data[r, c] = "Low"
    } else if (value == 2){
      merged_data[r, c] = "Moderate"
    } else if (value == 3){
      merged_data[r, c] = "High"
    } else if (value == 4){
      merged_data[r, c] = "Very High"
    }
  }
}
merged_data
```

```
##    sample_id Participant Treatment Concmean
## 1   sample_1           C Very High 3.824863
## 2  sample_10           C  Very Low 2.288922
## 3  sample_11           A  Moderate 2.056538
## 4  sample_12           A      High 1.599757
## 5  sample_13           B       Low 2.451226
## 6  sample_14           C  Moderate 2.901506
## 7  sample_15           A  Very Low 3.258827
## 8  sample_16           C Very High 1.753767
## 9  sample_17           B      High 2.443834
## 10 sample_18           B  Moderate 2.657674
## 11 sample_19           C       Low 1.782624
## 12  sample_2           B      High 3.561779
## 13 sample_20           C      High 2.034364
## 14 sample_21           A Very High 1.998999
## 15 sample_22           B  Very Low 2.163061
## 16 sample_23           A       Low 2.544692
## 17 sample_24           B Very High 2.452164
## 18 sample_25           C  Very Low 2.092057
## 19 sample_26           A  Moderate 3.003388
## 20 sample_27           A      High 3.313401
## 21 sample_28           B       Low 1.765348
## 22 sample_29           C  Moderate 1.624170
## 23  sample_3           B  Moderate 1.499365
## 24 sample_30           A  Very Low 3.037058
## 25 sample_31           C Very High 2.448267
## 26 sample_32           B      High 2.038269
## 27 sample_33           B  Moderate 2.833067
## 28 sample_34           C       Low 1.957261
## 29 sample_35           C      High 1.668823
## 30 sample_36           A Very High 2.192923
## 31 sample_37           B  Very Low 2.616311
## 32 sample_38           A       Low 3.646680
## 33 sample_39           B Very High 2.841996
## 34  sample_4           C       Low 2.958203
## 35 sample_40           C  Very Low 2.949213
## 36 sample_41           A  Moderate 3.229693
## 37 sample_42           A      High 2.329251
## 38 sample_43           B       Low 3.211285
## 39 sample_44           C  Moderate 2.842079
## 40 sample_45           A  Very Low 2.998275
## 41 sample_46           C Very High 2.744949
## 42 sample_47           B      High 1.683616
## 43 sample_48           B  Moderate 2.672564
## 44 sample_49           C       Low 2.434259
## 45  sample_5           C      High 2.390004
## 46 sample_50           C      High 2.322598
## 47  sample_6           A Very High 1.868026
## 48  sample_7           B  Very Low 2.409794
## 49  sample_8           A       Low 1.419867
## 50  sample_9           B Very High 3.306693
```

(vii) a sapply() function to change treatment levels from 0,1,2,3,4 to ”very-low”, ”low”, ”moderate”, ”high”, ”very-high” respectively

```
labels_function <- function(x) {
  labels <- ifelse(x==0, "very low",
                   ifelse(x==1, "low",
                          ifelse(x==2, "moderate",
                                 ifelse(x==3, "high",
                                        ifelse(x==4, "very high", NA)))))
  return(labels)
}

merged_data$Treatment <- sapply(merged_data$Treatment, labels_function)
merged_data
```

```
##    sample_id Participant Treatment Concmean
## 1   sample_1           C        NA 3.824863
## 2  sample_10           C        NA 2.288922
## 3  sample_11           A        NA 2.056538
## 4  sample_12           A        NA 1.599757
## 5  sample_13           B        NA 2.451226
## 6  sample_14           C        NA 2.901506
## 7  sample_15           A        NA 3.258827
## 8  sample_16           C        NA 1.753767
## 9  sample_17           B        NA 2.443834
## 10 sample_18           B        NA 2.657674
## 11 sample_19           C        NA 1.782624
## 12  sample_2           B        NA 3.561779
## 13 sample_20           C        NA 2.034364
## 14 sample_21           A        NA 1.998999
## 15 sample_22           B        NA 2.163061
## 16 sample_23           A        NA 2.544692
## 17 sample_24           B        NA 2.452164
## 18 sample_25           C        NA 2.092057
## 19 sample_26           A        NA 3.003388
## 20 sample_27           A        NA 3.313401
## 21 sample_28           B        NA 1.765348
## 22 sample_29           C        NA 1.624170
## 23  sample_3           B        NA 1.499365
## 24 sample_30           A        NA 3.037058
## 25 sample_31           C        NA 2.448267
## 26 sample_32           B        NA 2.038269
## 27 sample_33           B        NA 2.833067
## 28 sample_34           C        NA 1.957261
## 29 sample_35           C        NA 1.668823
## 30 sample_36           A        NA 2.192923
## 31 sample_37           B        NA 2.616311
## 32 sample_38           A        NA 3.646680
## 33 sample_39           B        NA 2.841996
## 34  sample_4           C        NA 2.958203
## 35 sample_40           C        NA 2.949213
## 36 sample_41           A        NA 3.229693
## 37 sample_42           A        NA 2.329251
## 38 sample_43           B        NA 3.211285
## 39 sample_44           C        NA 2.842079
## 40 sample_45           A        NA 2.998275
## 41 sample_46           C        NA 2.744949
## 42 sample_47           B        NA 1.683616
## 43 sample_48           B        NA 2.672564
## 44 sample_49           C        NA 2.434259
## 45  sample_5           C        NA 2.390004
## 46 sample_50           C        NA 2.322598
## 47  sample_6           A        NA 1.868026
## 48  sample_7           B        NA 2.409794
## 49  sample_8           A        NA 1.419867
## 50  sample_9           B        NA 3.306693
```

SECTION D

```
setwd("C:/Users/OLYMPIA/Downloads/fwdmsb7102bioconductorandrprogrammingassignment")
statsFile <- read.csv("statsFile.csv")
print(statsFile)
```

```
##          X      BMI       BP Gender Status Location
## 1     P_ 1 3.306331 1.848332   Male      1        C
## 2     P_ 2 3.144198 1.803189 Female      0        C
## 3     P_ 3 3.268586 1.837923 Female      1        B
## 4     P_ 4 3.175555 1.812009   Male      1        C
## 5     P_ 5 3.084312 1.786221 Female      1        A
## 6     P_ 6 3.227014 1.826389 Female      1        C
## 7     P_ 7 2.810668 1.706505   Male      1        A
## 8     P_ 8 3.080841 1.785232   Male      1        B
## 9     P_ 9 3.127907 1.798589   Male      1        C
## 10   P_ 10 2.986719 1.758213 Female      1        B
## 11   P_ 11 2.877521 1.726326 Female      0        B
## 12   P_ 12 3.059652 1.779186 Female      1        B
## 13   P_ 13 2.984433 1.757551 Female      0        B
## 14   P_ 14 3.425850 1.880905   Male      1        B
## 15   P_ 15 3.439004 1.884455   Male      1        B
## 16   P_ 16 3.428063 1.881503 Female      1        B
## 17   P_ 17 3.122743 1.797128 Female      0        B
## 18   P_ 18 2.978352 1.755790   Male      0        C
## 19   P_ 19 2.938888 1.744319   Male      1        B
## 20   P_ 20 3.179976 1.813249   Male      0        C
## 21   P_ 21 3.127936 1.798597   Male      1        B
## 22   P_ 22 3.274198 1.839474 Female      1        B
## 23   P_ 23 2.963454 1.751469   Male      0        A
## 24   P_ 24 3.445759 1.886276 Female      0        B
## 25   P_ 25 3.228358 1.826763 Female      1        C
## 26   P_ 26 2.840086 1.715256 Female      0        C
## 27   P_ 27 3.281707 1.841548   Male      1        C
## 28   P_ 28 3.273761 1.839354   Male      0        B
## 29   P_ 29 3.279747 1.841007   Male      0        B
## 30   P_ 30 3.257727 1.834917   Male      1        A
## 31   P_ 31 3.005242 1.763563 Female      1        B
## 32   P_ 32 3.253051 1.833622 Female      0        C
## 33   P_ 33 3.192021 1.816623 Female      0        B
## 34   P_ 34 3.307199 1.848571   Male      0        B
## 35   P_ 35 3.057509 1.778573   Male      1        C
## 36   P_ 36 3.829554 1.986925   Male      1        B
## 37   P_ 37 3.237382 1.829273 Female      0        B
## 38   P_ 38 3.272570 1.839025 Female      1        B
## 39   P_ 39 3.260025 1.835554 Female      0        B
## 40   P_ 40 3.330410 1.854941 Female      1        A
## 41   P_ 41 3.501719 1.901288   Male      1        C
## 42   P_ 42 2.959901 1.750436 Female      0        B
## 43   P_ 43 3.225575 1.825989   Male      1        C
## 44   P_ 44 3.110074 1.793540   Male      1        A
## 45   P_ 45 3.185849 1.814895 Female      1        C
## 46   P_ 46 3.278592 1.840688   Male      0        A
## 47   P_ 47 3.036027 1.772420   Male      0        B
## 48   P_ 48 3.417105 1.878541   Male      1        A
## 49   P_ 49 3.060042 1.779298 Female      1        C
## 50   P_ 50 3.307350 1.848612 Female      0        B
## 51   P_ 51 3.328844 1.854512   Male      0        A
## 52   P_ 52 2.946301 1.746479   Male      0        B
## 53   P_ 53 3.291036 1.844121   Male      0        C
## 54   P_ 54 3.045418 1.775113   Male      0        B
## 55   P_ 55 3.260084 1.835570   Male      0        B
## 56   P_ 56 3.216617 1.823493 Female      0        C
## 57   P_ 57 3.219905 1.824409 Female      0        A
## 58   P_ 58 3.356026 1.861946   Male      0        B
## 59   P_ 59 3.244715 1.831309 Female      1        B
## 60   P_ 60 3.058407 1.778830   Male      1        B
## 61   P_ 61 3.107917 1.792929   Male      0        A
## 62   P_ 62 2.950893 1.747816   Male      0        C
## 63   P_ 63 3.543076 1.912306 Female      0        C
## 64   P_ 64 3.272872 1.839108 Female      0        C
## 65   P_ 65 3.350650 1.860478   Male      0        B
## 66   P_ 66 3.027500 1.769971   Male      0        B
## 67   P_ 67 3.291736 1.844314 Female      1        A
## 68   P_ 68 3.061583 1.779738 Female      0        B
## 69   P_ 69 2.796652 1.702319 Female      0        C
## 70   P_ 70 3.074429 1.783405 Female      1        A
## 71   P_ 71 3.055393 1.777968 Female      1        B
## 72   P_ 72 2.605800 1.644249   Male      1        A
## 73   P_ 73 2.942185 1.745280 Female      0        C
## 74   P_ 74 3.260082 1.835570 Female      0        A
## 75   P_ 75 2.982465 1.756981   Male      0        B
## 76   P_ 76 2.984420 1.757547 Female      0        B
## 77   P_ 77 3.313720 1.850363 Female      0        A
## 78   P_ 78 3.374860 1.867079 Female      1        B
## 79   P_ 79 3.456377 1.889133 Female      1        B
## 80   P_ 80 3.172054 1.811026 Female      0        B
## 81   P_ 81 3.061465 1.779704 Female      0        C
## 82   P_ 82 3.123993 1.797482   Male      1        B
## 83   P_ 83 2.818505 1.708840 Female      1        B
## 84   P_ 84 3.495468 1.899617 Female      0        B
## 85   P_ 85 3.187770 1.815433 Female      1        B
## 86   P_ 86 3.138830 1.801674 Female      0        B
## 87   P_ 87 3.542345 1.912112   Male      0        C
## 88   P_ 88 3.319393 1.851920 Female      0        C
## 89   P_ 89 2.752596 1.689095   Male      0        A
## 90   P_ 90 2.682065 1.667701 Female      0        A
## 91   P_ 91 3.459533 1.889982 Female      1        B
## 92   P_ 92 3.314744 1.850644 Female      1        C
## 93   P_ 93 3.291362 1.844211 Female      0        C
## 94   P_ 94 3.369655 1.865662   Male      1        C
## 95   P_ 95 3.182388 1.813925 Female      1        B
## 96   P_ 96 3.234275 1.828409 Female      1        A
## 97   P_ 97 3.295484 1.845347   Male      1        A
## 98   P_ 98 2.883014 1.727944 Female      1        C
## 99   P_ 99 3.249635 1.832675 Female      0        B
## 100 P_ 100 3.107447 1.792795 Female      1        C
## 101 P_ 101 3.401663 1.874360 Female      1        C
## 102 P_ 102 3.245057 1.831404 Female      1        C
## 103 P_ 103 3.106424 1.792505   Male      0        C
## 104 P_ 104 2.997683 1.761382 Female      0        B
## 105 P_ 105 3.190716 1.816258 Female      1        A
## 106 P_ 106 3.044257 1.774780   Male      1        C
## 107 P_ 107 3.147981 1.804255   Male      1        B
## 108 P_ 108 3.027535 1.769981 Female      1        C
## 109 P_ 109 3.136686 1.801069   Male      0        A
## 110 P_ 110 3.324542 1.853333   Male      0        C
## 111 P_ 111 3.001240 1.762409   Male      1        C
## 112 P_ 112 3.313261 1.850237   Male      1        B
## 113 P_ 113 3.113887 1.794621 Female      1        C
## 114 P_ 114 3.216494 1.823459   Male      0        A
## 115 P_ 115 3.098998 1.790397 Female      1        A
## 116 P_ 116 3.250519 1.832920   Male      0        A
## 117 P_ 117 2.947339 1.746782   Male      0        C
## 118 P_ 118 3.518533 1.905775 Female      0        A
## 119 P_ 119 3.079973 1.784985 Female      1        B
## 120 P_ 120 3.206107 1.820560   Male      1        A
## 121 P_ 121 3.329679 1.854741   Male      1        C
## 122 P_ 122 3.365403 1.864504   Male      1        B
## 123 P_ 123 2.983788 1.757364 Female      1        C
## 124 P_ 124 3.258012 1.834996 Female      0        A
## 125 P_ 125 3.380836 1.868705 Female      0        C
## 126 P_ 126 2.865420 1.722755   Male      1        C
## 127 P_ 127 3.414156 1.877743 Female      1        C
## 128 P_ 128 3.324880 1.853425   Male      0        C
## 129 P_ 129 3.138834 1.801676 Female      0        A
## 130 P_ 130 3.567947 1.918901   Male      1        C
## 131 P_ 131 3.209298 1.821451 Female      1        B
## 132 P_ 132 3.571585 1.919864 Female      1        C
## 133 P_ 133 3.248606 1.832389   Male      0        C
## 134 P_ 134 3.307159 1.848560 Female      0        A
## 135 P_ 135 3.330612 1.854996   Male      1        B
## 136 P_ 136 3.028741 1.770328 Female      1        C
## 137 P_ 137 2.978578 1.755856 Female      1        C
## 138 P_ 138 2.960410 1.750584 Female      1        A
## 139 P_ 139 3.083216 1.785909   Male      0        A
## 140 P_ 140 2.751381 1.688729 Female      1        A
## 141 P_ 141 3.460443 1.890227   Male      0        A
## 142 P_ 142 3.170763 1.810664 Female      1        A
## 143 P_ 143 3.201116 1.819166   Male      1        B
## 144 P_ 144 3.140151 1.802047 Female      0        C
## 145 P_ 145 3.299345 1.846410 Female      0        B
## 146 P_ 146 3.106474 1.792519 Female      1        C
## 147 P_ 147 3.002342 1.762727 Female      0        A
## 148 P_ 148 3.187441 1.815341   Male      0        C
## 149 P_ 149 3.637052 1.937106 Female      0        B
## 150 P_ 150 3.347696 1.859671 Female      1        C
## 151 P_ 151 3.108883 1.793202   Male      0        C
## 152 P_ 152 3.336628 1.856644 Female      0        A
## 153 P_ 153 3.136921 1.801136 Female      0        B
## 154 P_ 154 3.285429 1.842575   Male      0        B
## 155 P_ 155 3.079074 1.784729   Male      1        B
## 156 P_ 156 2.837984 1.714632 Female      1        C
## 157 P_ 157 3.040833 1.773798 Female      1        A
## 158 P_ 158 3.074275 1.783361 Female      1        A
## 159 P_ 159 3.020722 1.768023   Male      1        B
## 160 P_ 160 2.850235 1.718264 Female      0        A
## 161 P_ 161 3.061075 1.779593 Female      1        B
## 162 P_ 162 3.091067 1.788143   Male      1        A
## 163 P_ 163 3.061354 1.779673   Male      1        B
## 164 P_ 164 3.145095 1.803442 Female      0        B
## 165 P_ 165 3.645553 1.939333 Female      0        B
## 166 P_ 166 3.427649 1.881391 Female      1        A
## 167 P_ 167 3.165412 1.809160 Female      0        C
## 168 P_ 168 3.401770 1.874389 Female      1        B
## 169 P_ 169 3.199875 1.818819 Female      1        C
## 170 P_ 170 3.277965 1.840515   Male      1        C
## 171 P_ 171 3.387328 1.870469   Male      1        C
## 172 P_ 172 3.484293 1.896626 Female      0        B
## 173 P_ 173 3.197957 1.818283   Male      0        C
## 174 P_ 174 3.366375 1.864768   Male      0        B
## 175 P_ 175 3.103066 1.791552 Female      0        A
## 176 P_ 176 3.203687 1.819885   Male      0        C
## 177 P_ 177 3.211482 1.822061 Female      1        A
## 178 P_ 178 3.114099 1.794681 Female      1        B
## 179 P_ 179 3.191928 1.816597   Male      0        A
## 180 P_ 180 3.375234 1.867181   Male      0        A
## 181 P_ 181 3.254546 1.834036 Female      0        B
## 182 P_ 182 2.968039 1.752800   Male      0        A
## 183 P_ 183 3.172897 1.811263 Female      1        B
## 184 P_ 184 3.092469 1.788542 Female      0        C
## 185 P_ 185 3.162369 1.808305 Female      1        C
## 186 P_ 186 3.273333 1.839236 Female      0        B
## 187 P_ 187 3.314979 1.850708 Female      1        B
## 188 P_ 188 3.200403 1.818967   Male      1        A
## 189 P_ 189 2.989698 1.759074 Female      1        C
## 190 P_ 190 2.841601 1.715705   Male      1        A
## 191 P_ 191 3.207055 1.820825   Male      0        B
## 192 P_ 192 2.842813 1.716064 Female      0        C
## 193 P_ 193 3.162854 1.808442 Female      0        C
## 194 P_ 194 3.369207 1.865540 Female      0        B
## 195 P_ 195 3.221186 1.824766   Male      1        A
## 196 P_ 196 2.970283 1.753451   Male      1        B
## 197 P_ 197 3.024880 1.769218 Female      1        B
## 198 P_ 198 2.876037 1.725888 Female      0        C
## 199 P_ 199 3.064094 1.780455   Male      0        B
## 200 P_ 200 3.021683 1.768299   Male      1        C
## 201 P_ 201 3.427609 1.881380   Male      1        A
## 202 P_ 202 2.900182 1.732992   Male      0        A
## 203 P_ 203 3.422883 1.880104   Male      0        C
## 204 P_ 204 3.188691 1.815691   Male      1        C
## 205 P_ 205 3.052251 1.777069 Female      0        A
## 206 P_ 206 2.957071 1.749614   Male      1        A
## 207 P_ 207 3.174878 1.811819 Female      0        B
## 208 P_ 208 3.359505 1.862895 Female      0        B
## 209 P_ 209 3.036145 1.772454   Male      1        A
## 210 P_ 210 2.989265 1.758949 Female      0        B
## 211 P_ 211 3.458403 1.889678   Male      1        A
## 212 P_ 212 3.101973 1.791242 Female      1        B
## 213 P_ 213 3.069651 1.782042   Male      0        C
## 214 P_ 214 3.051889 1.776966 Female      0        C
## 215 P_ 215 3.083543 1.786002   Male      0        B
## 216 P_ 216 3.262140 1.836140   Male      1        A
## 217 P_ 217 3.300682 1.846778   Male      1        C
## 218 P_ 218 3.487922 1.897598 Female      0        A
## 219 P_ 219 3.094160 1.789023   Male      1        B
## 220 P_ 220 3.583811 1.923096 Female      0        B
## 221 P_ 221 2.826060 1.711089 Female      1        C
## 222 P_ 222 2.894821 1.731417   Male      0        C
## 223 P_ 223 3.267766 1.837696 Female      0        C
## 224 P_ 224 3.071254 1.782499 Female      1        C
## 225 P_ 225 3.373706 1.866765 Female      0        C
## 226 P_ 226 3.354765 1.861602 Female      1        B
## 227 P_ 227 3.325455 1.853583 Female      0        A
## 228 P_ 228 2.981858 1.756806   Male      0        A
## 229 P_ 229 3.600099 1.927393 Female      0        B
## 230 P_ 230 3.157704 1.806993   Male      1        B
## 231 P_ 231 3.125039 1.797778 Female      1        B
## 232 P_ 232 3.162960 1.808471 Female      0        C
## 233 P_ 233 3.012237 1.765580 Female      0        A
## 234 P_ 234 3.166220 1.809388 Female      0        C
## 235 P_ 235 3.208090 1.821114   Male      0        B
## 236 P_ 236 3.137806 1.801385 Female      0        A
## 237 P_ 237 3.117345 1.795600 Female      0        B
## 238 P_ 238 3.096366 1.789649 Female      1        C
## 239 P_ 239 3.044178 1.774757 Female      0        A
## 240 P_ 240 3.202435 1.819535 Female      0        B
## 241 P_ 241 2.744284 1.686588 Female      0        B
## 242 P_ 242 2.982725 1.757057 Female      0        C
## 243 P_ 243 3.054217 1.777632   Male      1        B
## 244 P_ 244 2.928868 1.741394 Female      1        B
## 245 P_ 245 3.389145 1.870963   Male      1        B
## 246 P_ 246 3.263518 1.836521   Male      1        C
## 247 P_ 247 3.051933 1.776978   Male      0        A
## 248 P_ 248 3.194558 1.817333 Female      1        B
## 249 P_ 249 3.157671 1.806984   Male      0        B
## 250 P_ 250 3.260292 1.835628 Female      0        B
## 251 P_ 251 3.097566 1.789990 Female      0        A
## 252 P_ 252 3.285243 1.842524   Male      0        A
## 253 P_ 253 3.300640 1.846766   Male      1        A
## 254 P_ 254 3.011234 1.765291   Male      0        A
## 255 P_ 255 3.145161 1.803460   Male      0        C
## 256 P_ 256 2.683956 1.668278   Male      0        B
## 257 P_ 257 3.444672 1.885983 Female      1        A
## 258 P_ 258 3.363551 1.863999   Male      1        C
## 259 P_ 259 3.041729 1.774055 Female      0        C
## 260 P_ 260 3.277359 1.840348 Female      1        C
## 261 P_ 261 3.454507 1.888630 Female      0        C
## 262 P_ 262 3.029328 1.770497   Male      1        C
## 263 P_ 263 3.169232 1.810234 Female      0        A
## 264 P_ 264 3.565012 1.918124 Female      0        B
## 265 P_ 265 3.283842 1.842137   Male      0        A
## 266 P_ 266 3.241339 1.830372 Female      0        A
## 267 P_ 267 2.671312 1.664415 Female      1        B
## 268 P_ 268 3.152116 1.805420   Male      0        C
## 269 P_ 269 3.383015 1.869297 Female      0        A
## 270 P_ 270 3.529879 1.908797 Female      0        A
## 271 P_ 271 3.116114 1.795252 Female      0        B
## 272 P_ 272 3.319530 1.851958   Male      0        A
## 273 P_ 273 3.386186 1.870159 Female      0        A
## 274 P_ 274 3.115294 1.795020 Female      1        B
## 275 P_ 275 3.360955 1.863291   Male      1        B
## 276 P_ 276 3.455400 1.888871   Male      1        B
## 277 P_ 277 3.291781 1.844327   Male      1        C
## 278 P_ 278 2.846931 1.717285 Female      1        C
## 279 P_ 279 3.503196 1.901683   Male      1        B
## 280 P_ 280 2.924019 1.739976 Female      1        C
## 281 P_ 281 3.351866 1.860810 Female      0        C
## 282 P_ 282 3.373093 1.866598 Female      0        C
## 283 P_ 283 3.409646 1.876523 Female      0        B
## 284 P_ 284 3.025672 1.769446 Female      0        C
## 285 P_ 285 3.206865 1.820772 Female      0        A
## 286 P_ 286 3.352775 1.861058 Female      0        A
## 287 P_ 287 3.117144 1.795544 Female      0        C
## 288 P_ 288 3.301987 1.847137 Female      1        A
## 289 P_ 289 3.440225 1.884784   Male      1        A
## 290 P_ 290 3.299523 1.846459 Female      1        C
## 291 P_ 291 3.351118 1.860606 Female      1        B
## 292 P_ 292 2.828296 1.711754   Male      1        C
## 293 P_ 293 3.003546 1.763074 Female      1        C
## 294 P_ 294 3.184543 1.814529   Male      0        A
## 295 P_ 295 3.347345 1.859575 Female      0        A
## 296 P_ 296 3.323872 1.853149 Female      1        A
## 297 P_ 297 3.328356 1.854378 Female      0        C
## 298 P_ 298 2.935770 1.743409 Female      1        B
## 299 P_ 299 3.156123 1.806548 Female      1        B
## 300 P_ 300 3.357638 1.862386   Male      1        A
## 301 P_ 301 3.411765 1.877097 Female      1        C
## 302 P_ 302 3.159818 1.807588 Female      1        A
## 303 P_ 303 3.177057 1.812430 Female      1        B
## 304 P_ 304 3.564893 1.918092   Male      0        C
## 305 P_ 305 3.172953 1.811278 Female      0        A
## 306 P_ 306 3.011653 1.765411   Male      1        A
## 307 P_ 307 3.028018 1.770120 Female      0        A
## 308 P_ 308 3.321001 1.852361   Male      0        B
## 309 P_ 309 3.183032 1.814105   Male      0        B
## 310 P_ 310 3.430008 1.882028   Male      0        C
## 311 P_ 311 3.410629 1.876789 Female      1        B
## 312 P_ 312 3.450777 1.887627   Male      0        B
## 313 P_ 313 3.258893 1.835240 Female      0        C
## 314 P_ 314 3.252923 1.833586 Female      0        A
## 315 P_ 315 2.700896 1.673440 Female      0        C
## 316 P_ 316 3.286402 1.842844   Male      1        C
## 317 P_ 317 3.249083 1.832521   Male      0        B
## 318 P_ 318 3.206189 1.820583 Female      1        B
## 319 P_ 319 2.991568 1.759615   Male      1        C
## 320 P_ 320 3.438551 1.884333   Male      1        B
```

Question 1

The method and technique used is calculating Pearson correlation coefficient between BMI and BP and comparing the means and variances of BMI and BP.

```r
cor(statsFile$BMI,statsFile$BP)
```

```
## [1] 0.9997509
```

```r
cor.test(statsFile$BMI, statsFile$BP)
```

```
##
##  Pearson's product-moment correlation
##
## data:  statsFile$BMI and statsFile$BP
## t = 798.75, df = 318, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9996895 0.9998001
## sample estimates:
##       cor
## 0.9997509
```

```
cor.test(statsFile$BMI, statsFile$BP, method = "pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  statsFile$BMI and statsFile$BP
## t = 798.75, df = 318, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9996895 0.9998001
## sample estimates:
##       cor
## 0.9997509
```

```
?cor.test
```

```
## starting httpd help server ... done
```

```
cov(statsFile$BMI, statsFile$BP)
```

```
## [1] 0.01065146
```

```
cov(statsFile$BMI, statsFile$BP) / (sd(statsFile$BMI) * sd(statsFile$BP))
```

```
## [1] 0.9997509
```

comparison of means

```
mean(statsFile$BMI)
```

```
## [1] 3.185907
```

```
mean(statsFile$BP)
```

```
## [1] 1.814074
```

```
?t.test
t.test(statsFile$BMI, statsFile$BP)
```

```
##
##  Welch Two Sample t-test
##
## data:  statsFile$BMI and statsFile$BP
## t = 121.36, df = 369.13, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.349605 1.394060
## sample estimates:
## mean of x mean of y
##  3.185907  1.814074
```

comparison of variances

```
var(statsFile$BMI)
```

```
## [1] 0.03789084
```

```
var(statsFile$BP)
```

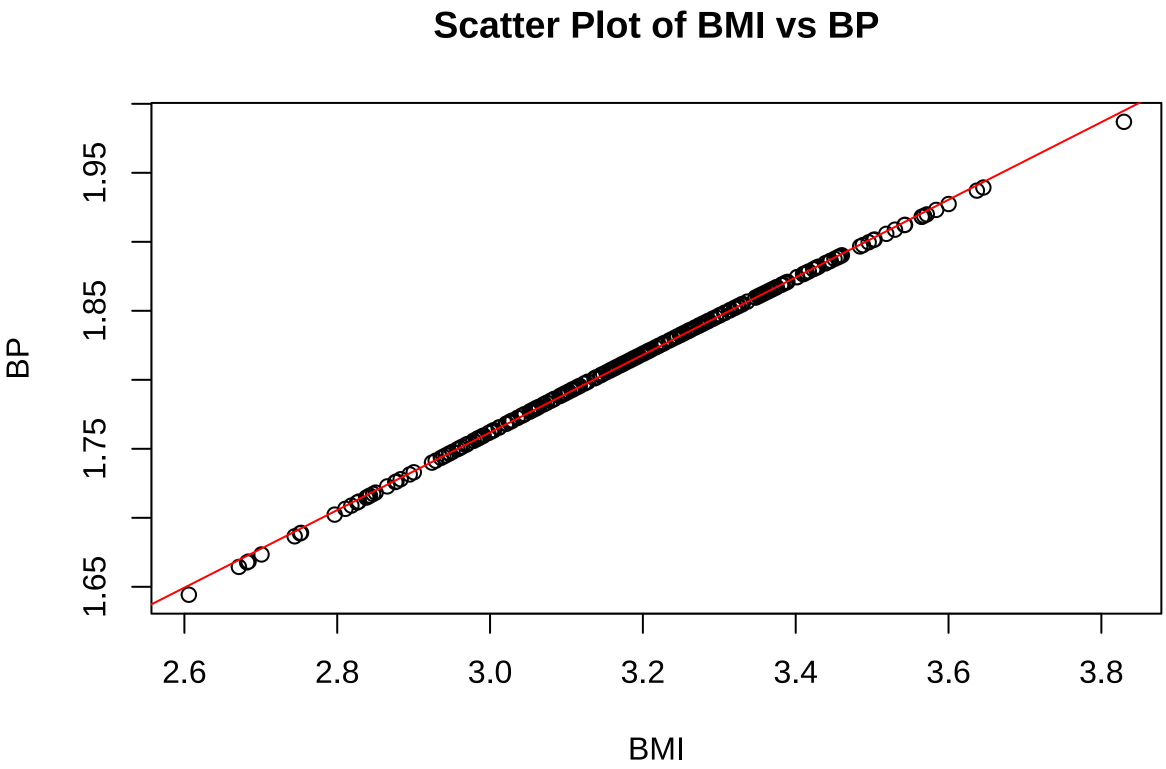
```
## [1] 0.002995712
```

```
?var.test
var.test(statsFile$BMI, statsFile$BP)
```

```
##
##  F test to compare two variances
##
## data:  statsFile$BMI and statsFile$BP
## F = 12.648, num df = 319, denom df = 319, p-value < 2.2e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  10.15190 15.75872
## sample estimates:
## ratio of variances
##           12.64836
```

Visualization is produced by a Scatter plot with regression line

```
plot(statsFile$BMI, statsFile$BP, main = "Scatter Plot of BMI vs BP", xlab = "BMI", ylab = "BP")
abline(lm(statsFile$BP ~ statsFile$BMI), col = "red")
```



Question 2

The method used is Chi-Square Test of Independence and the technique is by conducting a chi-square test to examine the association between disease status and gender.

```
chisq.test(table(statsFile$Status, statsFile$Gender))
```

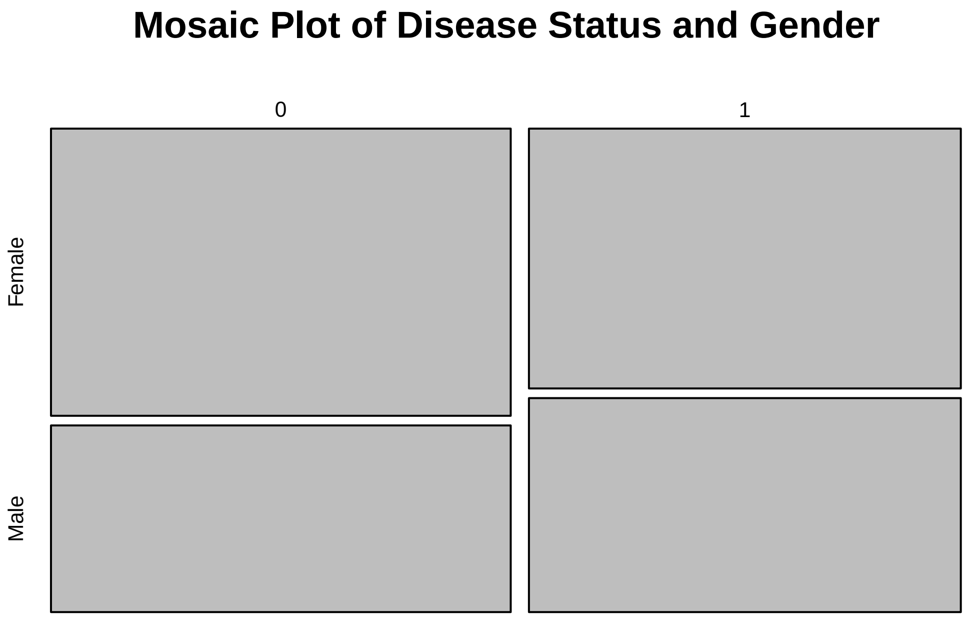
```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(statsFile$Status, statsFile$Gender)
## X-squared = 0.86632, df = 1, p-value = 0.352
```

```
table(statsFile$Status, statsFile$Gender)
```

```
##
##     Female Male
##   0    100   65
##   1     85   70
```

visualization is produced by a mosaic plot

```
mosaicplot(table (statsFile$Status, statsFile$Gender), main = "Mosaic Plot of Disease Status and Gender
```



Question 3

A linear regression model with BMI as the predictor and BP as the response variable is used as a technique.

```r
linear_model <- lm(BP ~ BMI, data = statsFile)
linear_model
```

```
##
## Call:
## lm(formula = BP ~ BMI, data = statsFile)
##
## Coefficients:
## (Intercept)          BMI
##      0.9185       0.2811
```

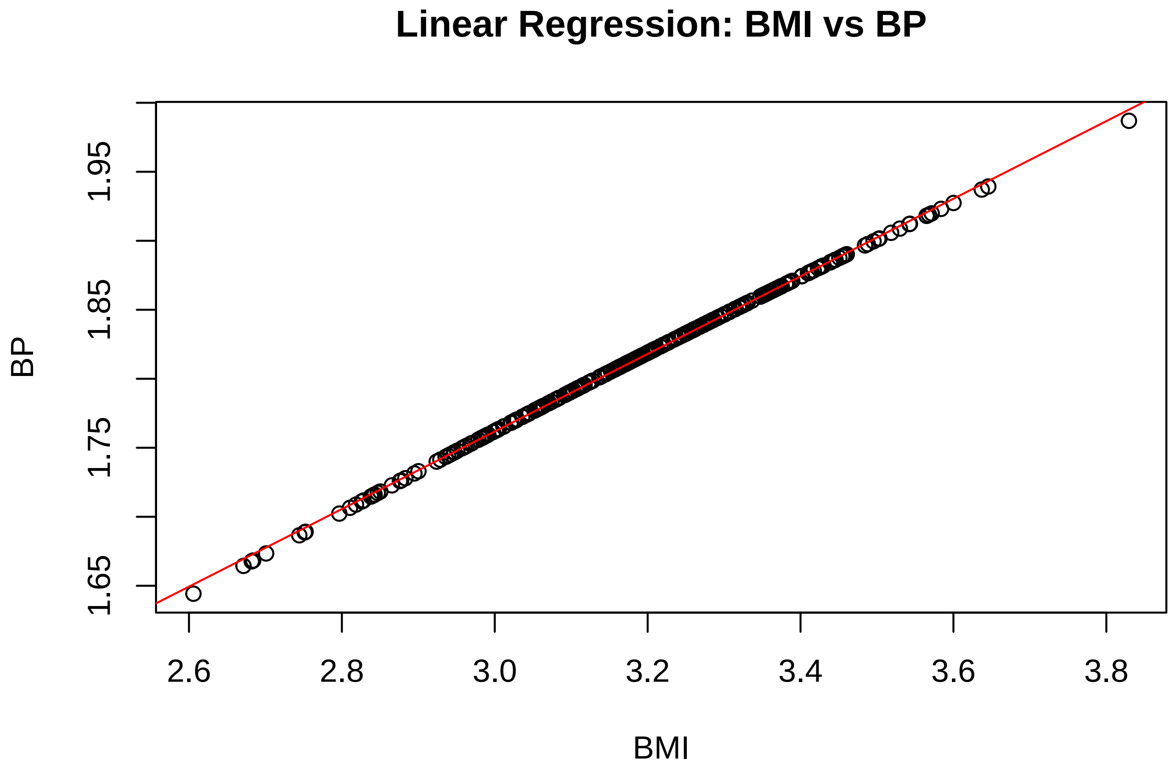
```r
summary(linear_model)
```

```
##
## Call:
## lm(formula = BP ~ BMI, data = statsFile)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0080845 -0.0001705  0.0004430  0.0007507  0.0008477
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9184871  0.0011233   817.6   <2e-16 ***
## BMI         0.2811090  0.0003519   798.7   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.001224 on 318 degrees of freedom
## Multiple R-squared:  0.9995, Adjusted R-squared:  0.9995
## F-statistic: 6.38e+05 on 1 and 318 DF,  p-value: < 2.2e-16
```

visualization is produced by constructing a scatter plot with a regression line

```r
plot(statsFile$BMI, statsFile$BP, main = "Linear Regression: BMI vs BP", xlab = "BMI", ylab = "BP")
abline(linear_model, col = "red")
```



Interpretation

```r
summary(linear_model)
```

```
##
## Call:
## lm(formula = BP ~ BMI, data = statsFile)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0080845 -0.0001705  0.0004430  0.0007507  0.0008477
```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9184871  0.0011233   817.6   <2e-16 ***
## BMI         0.2811090  0.0003519   798.7   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.001224 on 318 degrees of freedom
## Multiple R-squared:  0.9995, Adjusted R-squared:  0.9995
## F-statistic: 6.38e+05 on 1 and 318 DF,  p-value: < 2.2e-16
```

Question 4

Approach 1

mean, median and Variance

```
statsFile %>% group_by(Location) %>% summarize(mean = mean(BMI), median = median(BMI))
```

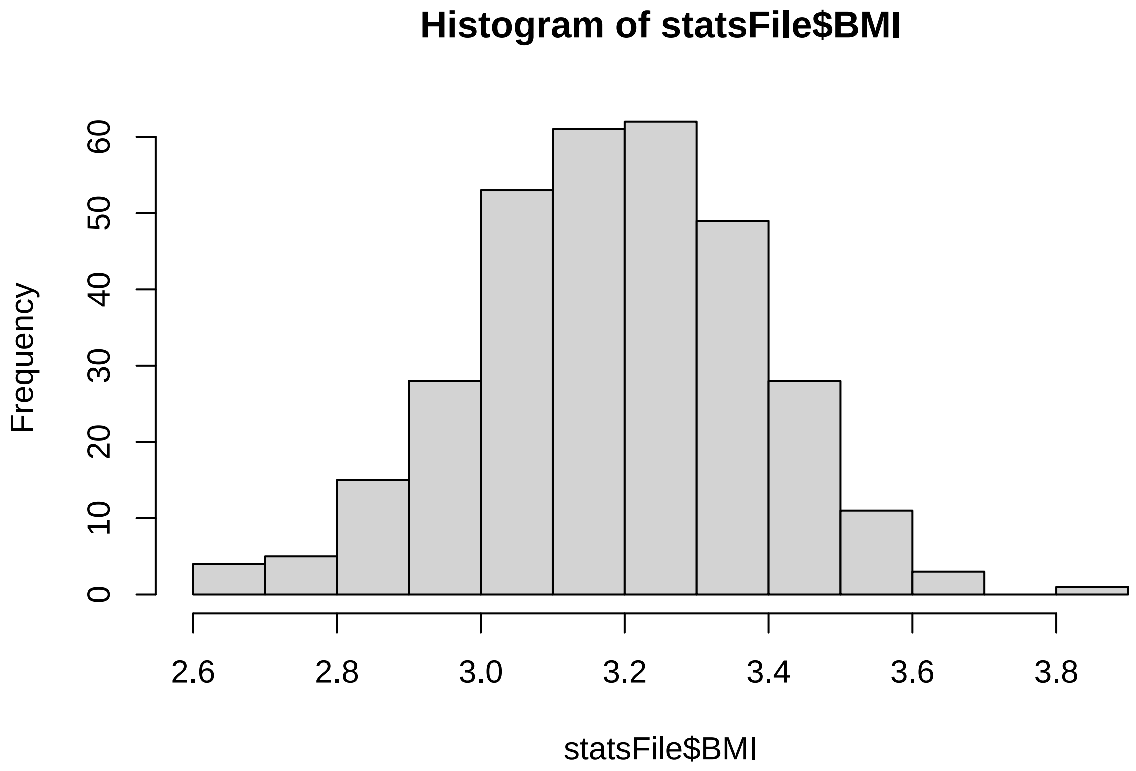
```
## # A tibble: 3 x 3
##   Location  mean median
##   <chr>    <dbl>  <dbl>
## 1 A         3.18   3.21
## 2 B         3.21   3.19
## 3 C         3.17   3.17
```

```
statsFile %>% group_by(Location) %>% summarize(variance = var(BMI))
```

```
## # A tibble: 3 x 2
##   Location variance
##   <chr>       <dbl>
## 1 A          0.0369
## 2 B          0.0395
## 3 C          0.0367
```

visualization (histogram)

```
hist(statsFile$BMI)
```



Approach 2

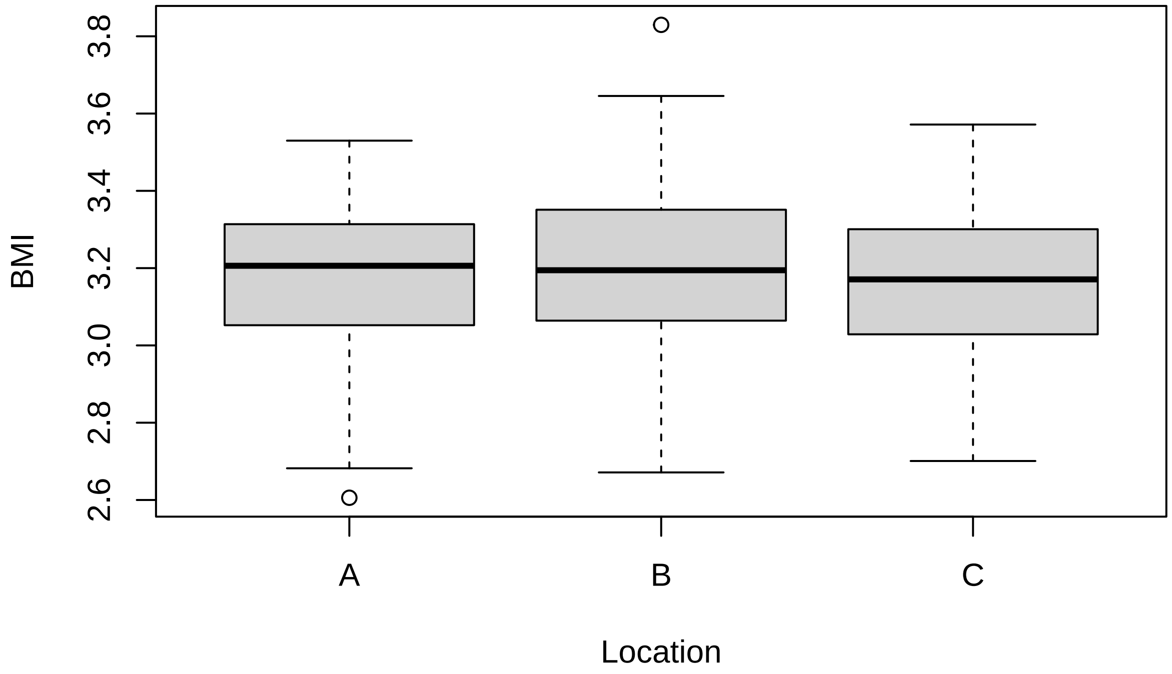
Performing a one-way ANOVA to test whether there is any statistically significant differences in BMI among difference sampling locations.

```
anov_result <- aov(BMI~Location, statsFile)
anov_result
```

```
## Call:
##    aov(formula = BMI ~ Location, data = statsFile)
##
## Terms:
##                  Location Residuals
## Sum of Squares   0.096678 11.990500
## Deg. of Freedom         2       317
##
## Residual standard error: 0.1944863
## Estimated effects may be unbalanced
```

visualization (boxplot)

```
boxplot(BMI ~ Location, data = statsFile)
```



Interpretation of ANOVA results.

```r
summary(anov_result)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Location      2  0.097 0.04834   1.278   0.28
## Residuals   317 11.991 0.03782
```

Question 5

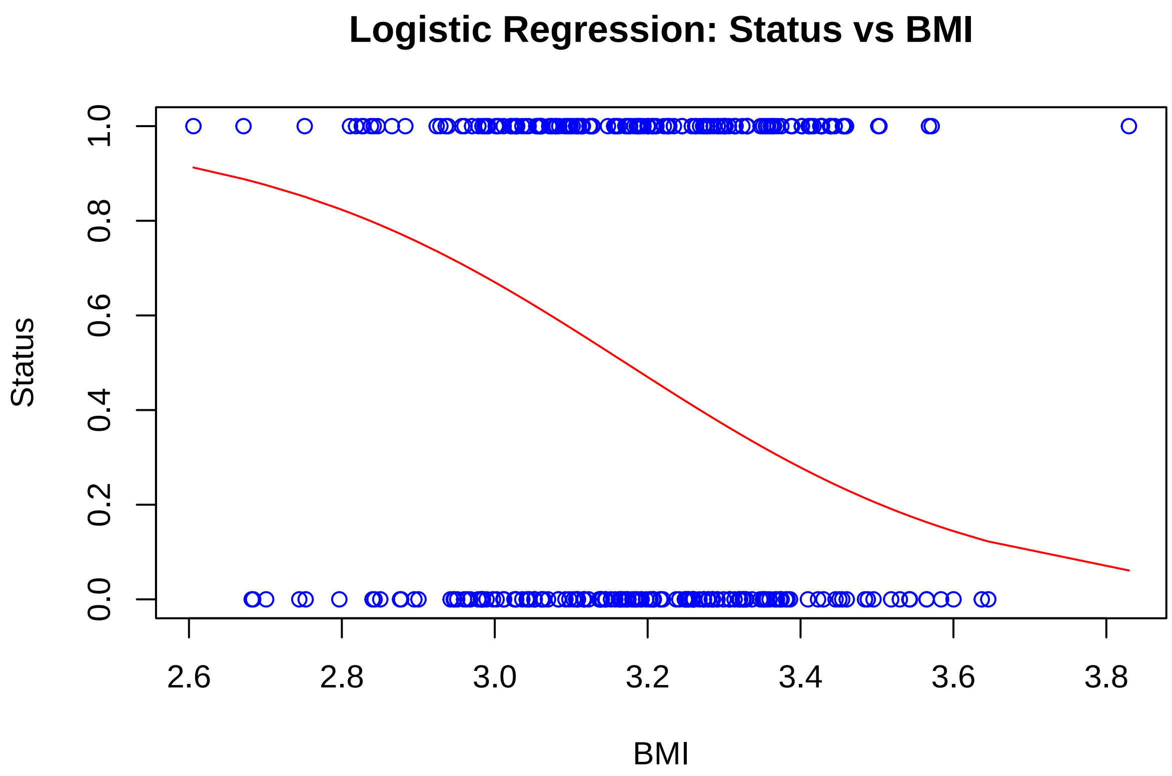
Building a logistic regression model with disease status as the binary response variable and BMI and BP as predictors.

```r
logistic_model <- glm(Status ~ BMI + BP, data = statsFile, family = "binomial")
logistic_model
```

```
##
## Call:  glm(formula = Status ~ BMI + BP, family = "binomial", data = statsFile)
##
## Coefficients:
## (Intercept)          BMI           BP
##      -11.64        -4.15        13.67
##
## Degrees of Freedom: 319 Total (i.e. Null);  317 Residual
## Null Deviance:       443.3
## Residual Deviance: 443   AIC: 449
```

visualization

```
plot(statsFile$BMI, statsFile$Status, main = "Logistic Regression: Status vs BMI", xlab = "BMI", ylab =
lines(sort(statsFile$BMI), predict(logistic_model, newdata = data.frame(BMI = sort(statsFile$BMI), BP =
```



Interpretation

```
summary(logistic_model)
```

```
##
## Call:
## glm(formula = Status ~ BMI + BP, family = "binomial", data = statsFile)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -11.64      84.53  -0.138    0.890
## BMI            -4.15      25.88  -0.160    0.873
## BP             13.67      92.02   0.149    0.882
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 443.30  on 319  degrees of freedom
## Residual deviance: 442.99  on 317  degrees of freedom
## AIC: 448.99
##
## Number of Fisher Scoring iterations: 3
```