Report

Khai báo các thư viện sử dụng

library(dplyr)

## Warning: package 'dplyr' was built under R version 4.0.3

##   
## 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(statsr)

## Warning: package 'statsr' was built under R version 4.0.3

1. Thông tin của tập dữ liệu và hiển thị sơ bộ

#Nguồn tham khảo: Tên: High School and Beyond survey Tổ chức sở hữu: OpenIntro Link: <https://www.openintro.org/data/index.php?data=hsb2> Tổ chức thu thập dữ liệu: Đại học California tại Los Angeles (UCLA) Link của tổ chức thu thập dữ liệu: UCLA Academic Technology Services, <http://www.ats.ucla.edu/stat/data/hsb2.csv> File dữ liệu: hsb2.csv

# Giới thiệu

Hai trăm quan sát được lấy mẫu ngẫu nhiên từ cuộc khảo sát Trung học, một cuộc Nghiên cứu được thực hiện trên học sinh của Trung tâm Thống kê Giáo dục Quốc gia.

# Hiển thị sơ bộ  
students\_scores <- read.csv("hsb2.csv")  
students\_scores

## id gender race ses schtyp prog read write math  
## 1 70 male white low public general 57 52 41  
## 2 121 female white middle public vocational 68 59 53  
## 3 86 male white high public general 44 33 54  
## 4 141 male white high public vocational 63 44 47  
## 5 172 male white middle public academic 47 52 57  
## 6 113 male white middle public academic 44 52 51  
## 7 50 male african american middle public general 50 59 42  
## 8 11 male hispanic middle public academic 34 46 45  
## 9 84 male white middle public general 63 57 54  
## 10 48 male african american middle public academic 57 55 52  
## 11 75 male white middle public vocational 60 46 51  
## 12 60 male white middle public academic 57 65 51  
## 13 95 male white high public academic 73 60 71  
## 14 104 male white high public academic 54 63 57  
## 15 38 male african american low public academic 45 57 50  
## 16 115 male white low public general 42 49 43  
## 17 76 male white high public academic 47 52 51  
## 18 195 male white middle private general 57 57 60  
## 19 114 male white high public academic 68 65 62  
## 20 85 male white middle public general 55 39 57  
## 21 167 male white middle public general 63 49 35  
## 22 143 male white middle public vocational 63 63 75  
## 23 41 male african american middle public academic 50 40 45  
## 24 20 male hispanic high public academic 60 52 57  
## 25 12 male hispanic middle public vocational 37 44 45  
## 26 53 male african american middle public vocational 34 37 46  
## 27 154 male white high public academic 65 65 66  
## 28 178 male white middle private vocational 47 57 57  
## 29 196 male white high private academic 44 38 49  
## 30 29 male asian low public general 52 44 49  
## 31 126 male white middle public general 42 31 57  
## 32 103 male white high public academic 76 52 64  
## 33 192 male white high private academic 65 67 63  
## 34 150 male white middle public vocational 42 41 57  
## 35 199 male white high private academic 52 59 50  
## 36 144 male white high public general 60 65 58  
## 37 200 male white middle private academic 68 54 75  
## 38 80 male white high public academic 65 62 68  
## 39 16 male hispanic low public vocational 47 31 44  
## 40 153 male white middle public vocational 39 31 40  
## 41 176 male white middle private academic 47 47 41  
## 42 177 male white middle private academic 55 59 62  
## 43 168 male white middle public academic 52 54 57  
## 44 40 male african american low public general 42 41 43  
## 45 62 male white high public general 65 65 48  
## 46 169 male white low public general 55 59 63  
## 47 49 male african american high public vocational 50 40 39  
## 48 136 male white middle public academic 65 59 70  
## 49 189 male white middle private academic 47 59 63  
## 50 7 male hispanic middle public academic 57 54 59  
## 51 27 male asian middle public academic 53 61 61  
## 52 128 male white high public academic 39 33 38  
## 53 21 male hispanic middle public general 44 44 61  
## 54 183 male white middle private academic 63 59 49  
## 55 132 male white middle public academic 73 62 73  
## 56 15 male hispanic high public vocational 39 39 44  
## 57 67 male white low public vocational 37 37 42  
## 58 22 male hispanic middle public vocational 42 39 39  
## 59 185 male white middle private academic 63 57 55  
## 60 9 male hispanic middle public vocational 48 49 52  
## 61 181 male white middle private academic 50 46 45  
## 62 170 male white high public academic 47 62 61  
## 63 134 male white low public general 44 44 39  
## 64 108 male white middle public general 34 33 41  
## 65 197 male white high private academic 50 42 50  
## 66 140 male white middle public vocational 44 41 40  
## 67 171 male white middle public academic 60 54 60  
## 68 107 male white low public vocational 47 39 47  
## 69 81 male white low public academic 63 43 59  
## 70 18 male hispanic middle public vocational 50 33 49  
## 71 155 male white middle public general 44 44 46  
## 72 97 male white high public academic 60 54 58  
## 73 68 male white middle public academic 73 67 71  
## 74 157 male white middle public general 68 59 58  
## 75 56 male white middle public vocational 55 45 46  
## 76 5 male hispanic low public academic 47 40 43  
## 77 159 male white high public academic 55 61 54  
## 78 123 male white high public general 68 59 56  
## 79 164 male white middle public vocational 31 36 46  
## 80 14 male hispanic high public academic 47 41 54  
## 81 127 male white high public academic 63 59 57  
## 82 165 male white low public vocational 36 49 54  
## 83 174 male white middle private academic 68 59 71  
## 84 3 male hispanic low public academic 63 65 48  
## 85 58 male white middle public vocational 55 41 40  
## 86 146 male white high public academic 55 62 64  
## 87 102 male white high public academic 52 41 51  
## 88 117 male white high public vocational 34 49 39  
## 89 133 male white middle public vocational 50 31 40  
## 90 94 male white high public academic 55 49 61  
## 91 24 male asian middle public academic 52 62 66  
## 92 149 male white low public general 63 49 49  
## 93 82 female white high public academic 68 62 65  
## 94 8 female hispanic low public academic 39 44 52  
## 95 129 female white low public general 44 44 46  
## 96 173 female white low public general 50 62 61  
## 97 57 female white middle public academic 71 65 72  
## 98 100 female white high public academic 63 65 71  
## 99 1 female hispanic low public vocational 34 44 40  
## 100 194 female white high private academic 63 63 69  
## 101 88 female white high public academic 68 60 64  
## 102 99 female white high public general 47 59 56  
## 103 47 female african american low public academic 47 46 49  
## 104 120 female white high public academic 63 52 54  
## 105 166 female white middle public academic 52 59 53  
## 106 65 female white middle public academic 55 54 66  
## 107 101 female white high public academic 60 62 67  
## 108 89 female white low public vocational 35 35 40  
## 109 54 female african american low private general 47 54 46  
## 110 180 female white high private academic 71 65 69  
## 111 162 female white middle public vocational 57 52 40  
## 112 4 female hispanic low public academic 44 50 41  
## 113 131 female white high public academic 65 59 57  
## 114 125 female white low public academic 68 65 58  
## 115 34 female hispanic high private academic 73 61 57  
## 116 106 female white middle public vocational 36 44 37  
## 117 130 female white high public general 43 54 55  
## 118 93 female white high public academic 73 67 62  
## 119 163 female white low public academic 52 57 64  
## 120 37 female african american low public vocational 41 47 40  
## 121 35 female hispanic low private general 60 54 50  
## 122 87 female white middle public general 50 52 46  
## 123 73 female white middle public academic 50 52 53  
## 124 151 female white middle public vocational 47 46 52  
## 125 44 female african american low public vocational 47 62 45  
## 126 152 female white high public academic 55 57 56  
## 127 105 female white middle public academic 50 41 45  
## 128 28 female asian middle public general 39 53 54  
## 129 91 female white high public vocational 50 49 56  
## 130 45 female african american low public vocational 34 35 41  
## 131 116 female white middle public academic 57 59 54  
## 132 33 female asian low public academic 57 65 72  
## 133 66 female white middle public vocational 68 62 56  
## 134 72 female white middle public vocational 42 54 47  
## 135 77 female white low public academic 61 59 49  
## 136 61 female white high public academic 76 63 60  
## 137 190 female white middle private academic 47 59 54  
## 138 42 female african american middle public vocational 46 52 55  
## 139 2 female hispanic middle public vocational 39 41 33  
## 140 55 female african american middle private academic 52 49 49  
## 141 19 female hispanic low public general 28 46 43  
## 142 90 female white high public academic 42 54 50  
## 143 142 female white middle public vocational 47 42 52  
## 144 17 female hispanic middle public academic 47 57 48  
## 145 122 female white middle public academic 52 59 58  
## 146 191 female white high private academic 47 52 43  
## 147 83 female white middle public vocational 50 62 41  
## 148 182 female white middle private academic 44 52 43  
## 149 6 female hispanic low public academic 47 41 46  
## 150 46 female african american low public academic 45 55 44  
## 151 43 female african american low public academic 47 37 43  
## 152 96 female white high public academic 65 54 61  
## 153 138 female white middle public vocational 43 57 40  
## 154 10 female hispanic middle public general 47 54 49  
## 155 71 female white middle public general 57 62 56  
## 156 139 female white middle public academic 68 59 61  
## 157 110 female white middle public vocational 52 55 50  
## 158 148 female white middle public vocational 42 57 51  
## 159 109 female white middle public general 42 39 42  
## 160 39 female african american high public academic 66 67 67  
## 161 147 female white low public academic 47 62 53  
## 162 74 female white middle public academic 57 50 50  
## 163 198 female white high private academic 47 61 51  
## 164 161 female white low public academic 57 62 72  
## 165 112 female white middle public academic 52 59 48  
## 166 69 female white low public vocational 44 44 40  
## 167 156 female white middle public academic 50 59 53  
## 168 111 female white low public general 39 54 39  
## 169 186 female white middle private academic 57 62 63  
## 170 98 female white low public vocational 57 60 51  
## 171 119 female white low public general 42 57 45  
## 172 13 female hispanic middle public vocational 47 46 39  
## 173 51 female african american high public general 42 36 42  
## 174 26 female asian high public academic 60 59 62  
## 175 36 female african american low public general 44 49 44  
## 176 135 female white low public academic 63 60 65  
## 177 59 female white middle public academic 65 67 63  
## 178 78 female white middle public academic 39 54 54  
## 179 64 female white high public vocational 50 52 45  
## 180 63 female white low public general 52 65 60  
## 181 79 female white middle public academic 60 62 49  
## 182 193 female white middle private academic 44 49 48  
## 183 92 female white high public general 52 67 57  
## 184 160 female white middle public academic 55 65 55  
## 185 32 female asian high public vocational 50 67 66  
## 186 23 female asian low public academic 65 65 64  
## 187 158 female white middle public general 52 54 55  
## 188 25 female asian middle public general 47 44 42  
## 189 188 female white high private academic 63 62 56  
## 190 52 female african american low public academic 50 46 53  
## 191 124 female white low public vocational 42 54 41  
## 192 175 female white high private general 36 57 42  
## 193 184 female white middle private vocational 50 52 53  
## 194 30 female asian high public academic 41 59 42  
## 195 179 female white middle private academic 47 65 60  
## 196 31 female asian middle private general 55 59 52  
## 197 145 female white middle public vocational 42 46 38  
## 198 187 female white middle private general 57 41 57  
## 199 118 female white middle public general 55 62 58  
## 200 137 female white high public academic 63 65 65  
## science socst  
## 1 47 57  
## 2 63 61  
## 3 58 31  
## 4 53 56  
## 5 53 61  
## 6 63 61  
## 7 53 61  
## 8 39 36  
## 9 58 51  
## 10 50 51  
## 11 53 61  
## 12 63 61  
## 13 61 71  
## 14 55 46  
## 15 31 56  
## 16 50 56  
## 17 50 56  
## 18 58 56  
## 19 55 61  
## 20 53 46  
## 21 66 41  
## 22 72 66  
## 23 55 56  
## 24 61 61  
## 25 39 46  
## 26 39 31  
## 27 61 66  
## 28 58 46  
## 29 39 46  
## 30 55 41  
## 31 47 51  
## 32 64 61  
## 33 66 71  
## 34 72 31  
## 35 61 61  
## 36 61 66  
## 37 66 66  
## 38 66 66  
## 39 36 36  
## 40 39 51  
## 41 42 51  
## 42 58 51  
## 43 55 51  
## 44 50 41  
## 45 63 66  
## 46 69 46  
## 47 49 47  
## 48 63 51  
## 49 53 46  
## 50 47 51  
## 51 57 56  
## 52 47 41  
## 53 50 46  
## 54 55 71  
## 55 69 66  
## 56 26 42  
## 57 33 32  
## 58 56 46  
## 59 58 41  
## 60 44 51  
## 61 58 61  
## 62 69 66  
## 63 34 46  
## 64 36 36  
## 65 36 61  
## 66 50 26  
## 67 55 66  
## 68 42 26  
## 69 65 44  
## 70 44 36  
## 71 39 51  
## 72 58 61  
## 73 63 66  
## 74 74 66  
## 75 58 51  
## 76 45 31  
## 77 49 61  
## 78 63 66  
## 79 39 46  
## 80 42 56  
## 81 55 56  
## 82 61 36  
## 83 66 56  
## 84 63 56  
## 85 44 41  
## 86 63 66  
## 87 53 56  
## 88 42 56  
## 89 34 31  
## 90 61 56  
## 91 47 46  
## 92 66 46  
## 93 69 61  
## 94 44 48  
## 95 47 51  
## 96 63 51  
## 97 66 56  
## 98 69 71  
## 99 39 41  
## 100 61 61  
## 101 69 66  
## 102 66 61  
## 103 33 41  
## 104 50 51  
## 105 61 51  
## 106 42 56  
## 107 50 56  
## 108 51 33  
## 109 50 56  
## 110 58 71  
## 111 61 56  
## 112 39 51  
## 113 46 66  
## 114 59 56  
## 115 55 66  
## 116 42 41  
## 117 55 46  
## 118 58 66  
## 119 58 56  
## 120 39 51  
## 121 50 51  
## 122 50 56  
## 123 39 56  
## 124 48 46  
## 125 34 46  
## 126 58 61  
## 127 44 56  
## 128 50 41  
## 129 47 46  
## 130 29 26  
## 131 50 56  
## 132 54 56  
## 133 50 51  
## 134 47 46  
## 135 44 66  
## 136 67 66  
## 137 58 46  
## 138 44 56  
## 139 42 41  
## 140 44 61  
## 141 44 51  
## 142 50 52  
## 143 39 51  
## 144 44 41  
## 145 53 66  
## 146 48 61  
## 147 55 31  
## 148 44 51  
## 149 40 41  
## 150 34 41  
## 151 42 46  
## 152 58 56  
## 153 50 51  
## 154 53 61  
## 155 58 66  
## 156 55 71  
## 157 54 61  
## 158 47 61  
## 159 42 41  
## 160 61 66  
## 161 53 61  
## 162 51 58  
## 163 63 31  
## 164 61 61  
## 165 55 61  
## 166 40 31  
## 167 61 61  
## 168 47 36  
## 169 55 41  
## 170 53 37  
## 171 50 43  
## 172 47 61  
## 173 31 39  
## 174 61 51  
## 175 35 51  
## 176 54 66  
## 177 55 71  
## 178 53 41  
## 179 58 36  
## 180 56 51  
## 181 50 51  
## 182 39 51  
## 183 63 61  
## 184 50 61  
## 185 66 56  
## 186 58 71  
## 187 53 51  
## 188 42 36  
## 189 55 61  
## 190 53 66  
## 191 42 41  
## 192 50 41  
## 193 55 56  
## 194 34 51  
## 195 50 56  
## 196 42 56  
## 197 36 46  
## 198 55 52  
## 199 58 61  
## 200 53 61

Các cột điểm read, write, math, science, socst dưới đây đều tính theo cột điểm 100

|  |  |
| --- | --- |
| Biến | Mô tả |
| id | Mã số học sinh |
| gender | Giới tính gồm 2 loại: male và female |
| race | Chủng tộc gồm 4 loại: african american, asian, hispanic và white |
| ses | Điều kiện kinh tế của học sinh gồm 3 loại: low, middle và high |
| schtyp | Loại trường gồm 2 loại: public and private |
| prog | Chương trình giảng dạy gồm 3 loại: general, academic, and vocational |
| read | Điểm đọc |
| write | Điểm viết |
| math | Điểm toán |
| science | Điểm khoa học |
| socst | Điểm công tác xã hội |

glimpse(students\_scores)

## Rows: 200  
## Columns: 11  
## $ id <int> 70, 121, 86, 141, 172, 113, 50, 11, 84, 48, 75, 60, 95, 104...  
## $ gender <chr> "male", "female", "male", "male", "male", "male", "male", "...  
## $ race <chr> "white", "white", "white", "white", "white", "white", "afri...  
## $ ses <chr> "low", "middle", "high", "high", "middle", "middle", "middl...  
## $ schtyp <chr> "public", "public", "public", "public", "public", "public",...  
## $ prog <chr> "general", "vocational", "general", "vocational", "academic...  
## $ read <int> 57, 68, 44, 63, 47, 44, 50, 34, 63, 57, 60, 57, 73, 54, 45,...  
## $ write <int> 52, 59, 33, 44, 52, 52, 59, 46, 57, 55, 46, 65, 60, 63, 57,...  
## $ math <int> 41, 53, 54, 47, 57, 51, 42, 45, 54, 52, 51, 51, 71, 57, 50,...  
## $ science <int> 47, 63, 58, 53, 53, 63, 53, 39, 58, 50, 53, 63, 61, 55, 31,...  
## $ socst <int> 57, 61, 31, 56, 61, 61, 61, 36, 51, 51, 61, 61, 71, 46, 56,...

1. Phân tích tổng quan các biến trong tập dữ liệu (EDA)

Các biến phân loại gồm:

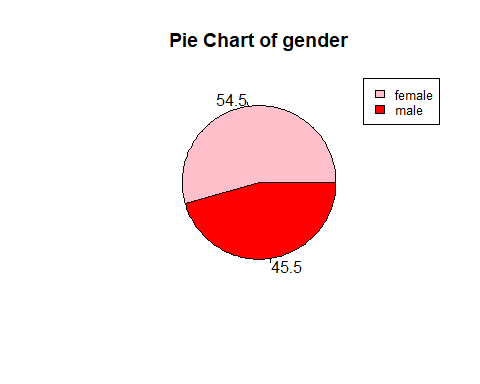
|  |  |
| --- | --- |
| Biến | Mô tả |
| gender | Giới tính gồm 2 loại: male và female |
| race | Chủng tộc gồm 4 loại: african american, asian, hispanic và white |
| ses | Điều kiện kinh tế của học sinh gồm 3 loại: low, middle và high |
| schtyp | Loại trường gồm 2 loại: public and private |
| prog | Chương trình giảng dạy gồm 3 loại: general, academic, and vocational |

Vẽ biểu đồ cho các biến phân loại:

# Biểu đồ cho cột gender  
gender\_data <- students\_scores %>%  
 count(students\_scores$gender)  
gender\_data

## students\_scores$gender n  
## 1 female 109  
## 2 male 91

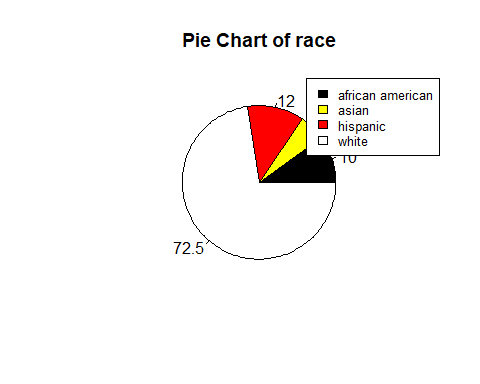
piepercent<- round(100\*gender\_data$n/sum(gender\_data$n), 1)  
colors\_data <- c("pink","red")  
pie(gender\_data$n, labels = piepercent, col = colors\_data, main="Pie Chart of gender")  
legend("topright", gender\_data$`students\_scores$gender`,   
 cex = 0.8, fill = colors\_data)



# Biểu đồ cho cột race  
race\_data <- students\_scores %>%  
 count(students\_scores$race)  
race\_data

## students\_scores$race n  
## 1 african american 20  
## 2 asian 11  
## 3 hispanic 24  
## 4 white 145

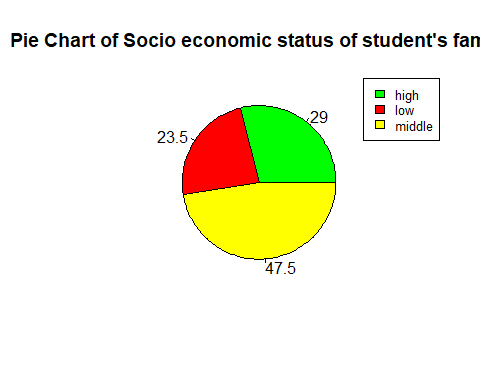
piepercent<- round(100\*race\_data$n/sum(race\_data$n), 1)  
colors\_data <- c("black","yellow","red","white")  
pie(race\_data$n, labels = piepercent, col = colors\_data, main="Pie Chart of race")  
legend("topright", race\_data$`students\_scores$race`,   
 cex = 0.8, fill = colors\_data)



#Biểu đồ cho cột ses   
ses\_data <- students\_scores %>%  
 count(students\_scores$ses)  
ses\_data

## students\_scores$ses n  
## 1 high 58  
## 2 low 47  
## 3 middle 95

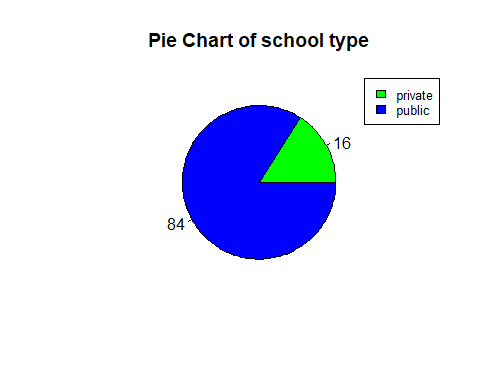
piepercent<- round(100\*ses\_data$n/sum(ses\_data$n), 1)  
colors\_data <- c("green","red","yellow")  
pie(ses\_data$n, labels = piepercent, col = colors\_data, main="Pie Chart of Socio economic status of student's family")  
legend("topright", ses\_data$`students\_scores$ses`,   
 cex = 0.8, fill = colors\_data)



#Biểu đồ cho cột schtyp  
schtyp\_data <- students\_scores %>%  
 count(students\_scores$schtyp)  
schtyp\_data

## students\_scores$schtyp n  
## 1 private 32  
## 2 public 168

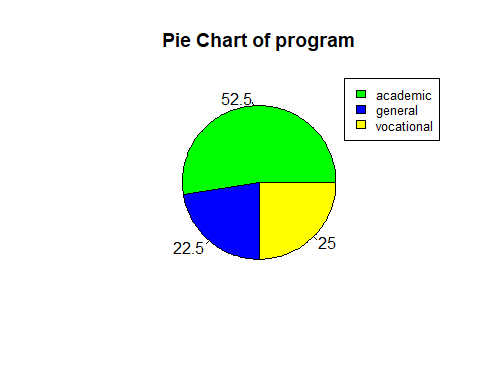
piepercent<- round(100\*schtyp\_data$n/sum(schtyp\_data$n), 1)  
colors\_data <- c("green","blue")  
pie(schtyp\_data$n, labels = piepercent, col = colors\_data, main="Pie Chart of school type")  
legend("topright", schtyp\_data$`students\_scores$schtyp`,   
 cex = 0.8, fill = colors\_data)



#Biểu đồ cho cột prog  
prog\_data <- students\_scores %>%  
 count(students\_scores$prog)  
prog\_data

## students\_scores$prog n  
## 1 academic 105  
## 2 general 45  
## 3 vocational 50

piepercent<- round(100\*prog\_data$n/sum(prog\_data$n), 1)  
colors\_data <- c("green","blue","yellow")  
pie(prog\_data$n, labels = piepercent, col = colors\_data, main="Pie Chart of program")  
legend("topright", prog\_data$`students\_scores$prog`,   
 cex = 0.8, fill = colors\_data)

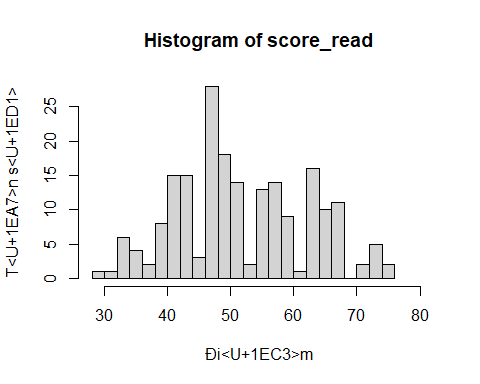


Các biến số gồm:

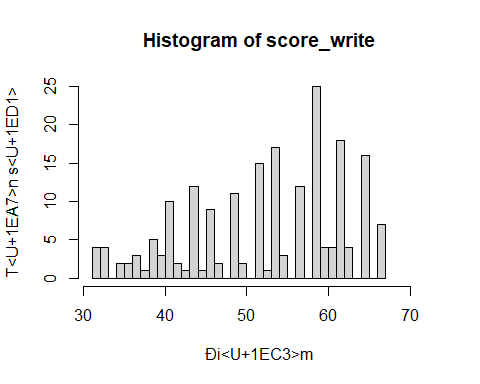
|  |  |
| --- | --- |
| Biến | Mô tả |
| read | Điểm đọc |
| write | Điểm viết |
| math | Điểm toán |
| science | Điểm khoa học |
| socst | Điểm công tác xã hội |

Vẽ biểu đồ cho các biến số và tính các summary statistics:

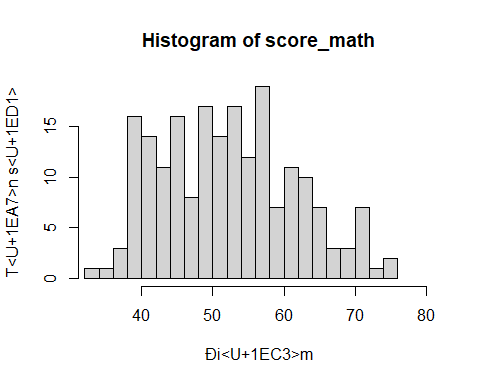
#Biểu đồ cho cột read  
read\_data <- students\_scores %>%  
 count(students\_scores$read) %>%  
 nrow()  
hist(students\_scores$read,  
 include.lowest = TRUE, right = TRUE, breaks = read\_data,  
 main = paste("Histogram of score\_read"),  
 xlim = range(min(students\_scores$read),max(students\_scores$read)+5), ylim = NULL,  
 xlab = "Điểm", ylab = "Tần số")



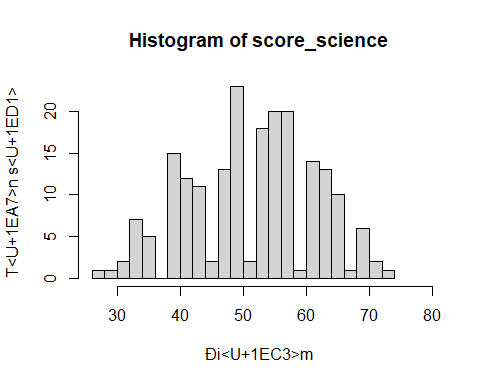
#Biểu đồ cho cột write  
write\_data <- students\_scores %>%  
 count(students\_scores$write) %>%  
 nrow()  
hist(students\_scores$write,  
 include.lowest = TRUE, right = TRUE, breaks = write\_data,  
 main = paste("Histogram of score\_write"),  
 xlim = range(min(students\_scores$write),max(students\_scores$write)+5), ylim = NULL,  
 xlab = "Điểm", ylab = "Tần số")



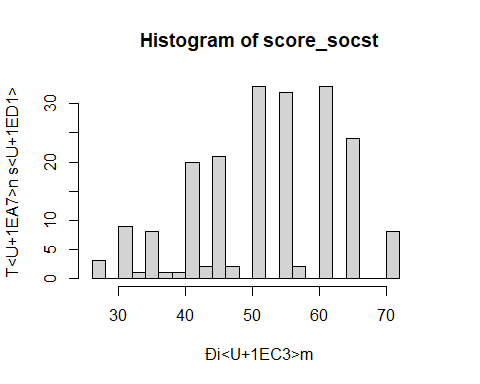
#Biểu đồ cho cột math  
math\_data <- students\_scores %>%  
 count(students\_scores$math) %>%  
 nrow()  
hist(students\_scores$math,  
 include.lowest = TRUE, right = TRUE, breaks = write\_data,  
 main = paste("Histogram of score\_math"),  
 xlim = range(min(students\_scores$math),max(students\_scores$math)+5), ylim = NULL,  
 xlab = "Điểm", ylab = "Tần số")



#Biểu đồ cho cột science  
science\_data <- students\_scores %>%  
 count(students\_scores$science) %>%  
 nrow()  
hist(students\_scores$science,  
 include.lowest = TRUE, right = TRUE, breaks = science\_data,  
 main = paste("Histogram of score\_science"),  
 xlim = range(min(students\_scores$science),max(students\_scores$science)+5), ylim = NULL,  
 xlab = "Điểm", ylab = "Tần số")



#Biểu đồ cho cột socst  
socst\_data <- students\_scores %>%  
 count(students\_scores$socst) %>%  
 nrow()  
hist(students\_scores$socst,  
 include.lowest = TRUE, right = TRUE, breaks = socst\_data,  
 main = paste("Histogram of score\_socst"),  
 xlim = range(min(students\_scores$socst),max(students\_scores$socst)+5), ylim = NULL,  
 xlab = "Điểm", ylab = "Tần số")



1. Chọn 3 biến phân loại và 1 biến số: 3 Biến phân loại được chọn là race, schtyp và gender Biến số dùng để phân tích là math

Tính và so sánh kích cỡ hai mẫu chia theo loại trường

students\_scores %>%  
 count(students\_scores$schtyp)

## students\_scores$schtyp n  
## 1 private 32  
## 2 public 168

Tính và so sánh kích cỡ hai mẫu chia theo giới tính

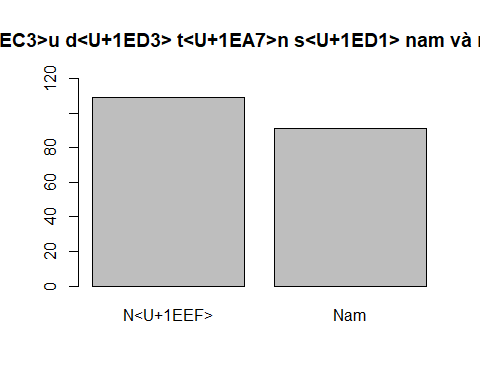
divide\_data <- students\_scores %>%  
 count(students\_scores$gender)  
divide\_data

## students\_scores$gender n  
## 1 female 109  
## 2 male 91

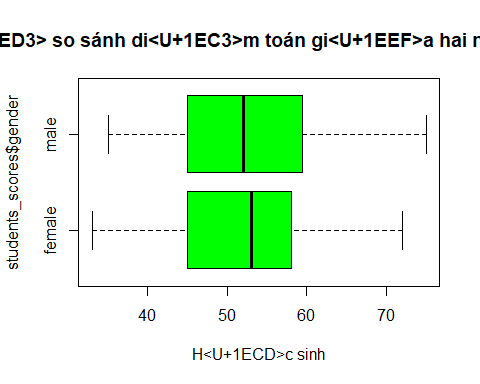
Do việc chia mẫu dựa trên biến phân loại theo giới tính sẽ có kích thước gần giống nhau nên ta sẽ chọn biến phân loại là giới tính đễ tạo ra hai mẫu.

Biểu đồ thanh (bar chart) của hai mẫu được chia dựa trên biến phân loại giới tính

barplot(table(students\_scores$gender),ylim=c(0, 120), main="Biểu đồ tần số nam và nữ", names = c("Nữ","Nam"))



boxplot(students\_scores$math ~ students\_scores$gender, horizontal=TRUE, main="Biểu đồ so sánh điểm toán giữa hai nhóm nam và nữ", xlab="Học sinh", col = "green")



students\_scores %>%   
group\_by(gender) %>%  
summarise(min = min(math), max = max(math),  
mean = mean(math), med = median(math),  
sd = sd(math),  
q1 = quantile(math, probs = 0.25),  
q3 = quantile(math, probs = 0.75), .groups = 'drop')

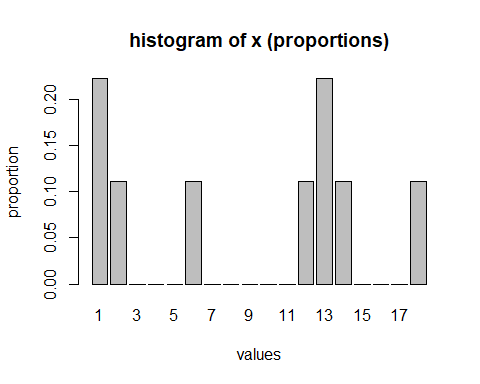
## # A tibble: 2 x 8  
## gender min max mean med sd q1 q3  
## <chr> <int> <int> <dbl> <int> <dbl> <dbl> <dbl>  
## 1 female 33 72 52.4 53 9.15 45 58   
## 2 male 35 75 52.9 52 9.66 45 59.5

Ta có =91,=52.94505, sd = 9.664784 =109,=52.39450, sd = 9.151015 Xét hai mẫu có bằng nhau hay không: giả thuyết: = đối thuyết: != Ta có giá trị kiểm định Z = 0.410958 vậy chỉ khi giá trị -0.410958 < < -0.410958 thì hai mẫu bằng nhau và ngược lại

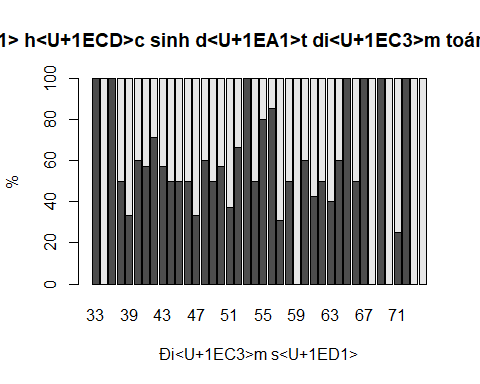
table\_Gender\_Math<-table(students\_scores$gender,students\_scores$math)  
#barplot(table(students\_scores$gender,students\_scores$math), main="Biểu đồ #thanh về học sinh đạt điểm toán theo tỉ lệ nam nữ ", xlab = "Tỉ lệ nam #nữ",freq=FALSE)  
table\_Gender\_Math

##   
## 33 35 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57  
## female 1 0 1 1 2 6 4 5 4 2 4 4 1 3 5 4 3 4 7 5 4 6 4  
## male 0 1 0 1 4 4 3 2 3 2 4 4 2 2 5 3 5 2 0 5 1 1 9  
##   
## 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 75  
## female 3 0 3 3 2 2 3 3 2 2 0 2 0 1 3 0 0  
## male 3 2 2 4 2 3 2 0 2 0 1 0 1 3 0 1 2

table\_Gender\_Math<- table(table\_Gender\_Math)  
barplot(height = table(factor(table\_Gender\_Math, levels=min(table\_Gender\_Math):max(table\_Gender\_Math)))/length(table\_Gender\_Math),  
 ylab = "proportion",  
 xlab = "values",  
 main = "histogram of x (proportions)")



barplot((prop.table(table(students\_scores$gender,students\_scores$math),margin = 2)\*100), main='Biểu đồ thanh về học sinh đạt điểm toán theo tỉ lệ nam nữ ', xlab = "Điểm số", ylab = "%")

 Khi nhìn vaod biểu đồ ta thấy rằng tỉ lệ nữ chiếm điện tích nhiều hơn so với tỉ lện, và điểm số của nữ cũng cao hơn so với nam Chon

prop.table(table(students\_scores$schtyp,students\_scores$math))\*100

##   
## 33 35 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51  
## private 0.0 0.0 0.0 0.0 0.0 0.0 0.5 0.5 1.0 0.0 0.5 0.5 0.0 0.5 1.5 1.5 0.5  
## public 0.5 0.5 0.5 1.0 3.0 5.0 3.0 3.0 2.5 2.0 3.5 3.5 1.5 2.0 3.5 2.0 3.5  
##   
## 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68  
## private 0.5 0.5 0.5 0.5 0.5 1.5 0.0 0.0 1.0 0.0 0.5 1.5 0.0 0.0 0.0 0.0 0.0  
## public 2.5 3.0 4.5 2.0 3.0 5.0 3.0 1.0 1.5 3.5 1.5 1.0 2.5 1.5 2.0 1.0 0.5  
##   
## 69 70 71 72 73 75  
## private 1.0 0.0 0.5 0.0 0.0 0.5  
## public 0.0 0.5 1.5 1.5 0.5 0.5

table(students\_scores$gender,students\_scores$math,students\_scores$schtyp)

## , , = private  
##   
##   
## 33 35 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57  
## female 0 0 0 0 0 0 0 1 2 0 0 1 0 1 1 1 1 1 1 1 0 1 2  
## male 0 0 0 0 0 0 1 0 0 0 1 0 0 0 2 2 0 0 0 0 1 0 1  
##   
## 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 75  
## female 0 0 1 0 0 1 0 0 0 0 0 2 0 0 0 0 0  
## male 0 0 1 0 1 2 0 0 0 0 0 0 0 1 0 0 1  
##   
## , , = public  
##   
##   
## 33 35 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57  
## female 1 0 1 1 2 6 4 4 2 2 4 3 1 2 4 3 2 3 6 4 4 5 2  
## male 0 1 0 1 4 4 2 2 3 2 3 4 2 2 3 1 5 2 0 5 0 1 8  
##   
## 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 75  
## female 3 0 2 3 2 1 3 3 2 2 0 0 0 1 3 0 0  
## male 3 2 1 4 1 1 2 0 2 0 1 0 1 2 0 1 1

1. Kiểm định giả thuyết và xây dựng khoảng tin cậy

table\_tgender<-table(students\_scores$gender,students\_scores$math)  
t.test(table\_tgender)

##   
## One Sample t-test  
##   
## data: table\_tgender  
## t = 12.095, df = 79, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 2.08859 2.91141  
## sample estimates:  
## mean of x   
## 2.5

table\_tschtyp<-table(students\_scores$schtyp,students\_scores$math)  
t.test(table\_tschtyp)

##   
## One Sample t-test  
##   
## data: table\_tschtyp  
## t = 8.5845, df = 79, p-value = 6.305e-13  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 1.920338 3.079662  
## sample estimates:  
## mean of x   
## 2.5

t.test(table\_tgender,table\_tschtyp)

##   
## Welch Two Sample t-test  
##   
## data: table\_tgender and table\_tschtyp  
## t = 0, df = 142.48, p-value = 1  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.7059298 0.7059298  
## sample estimates:  
## mean of x mean of y   
## 2.5 2.5