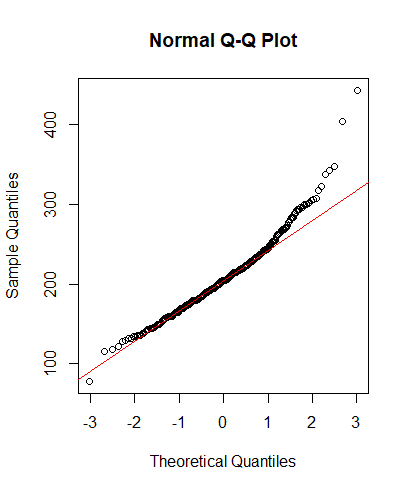
# Análise exploratória

De seguida apresentamos uma análise exploratória sobre as seguintes variáveis do dataset: colesterol, stab.glu, ratio, age,location.

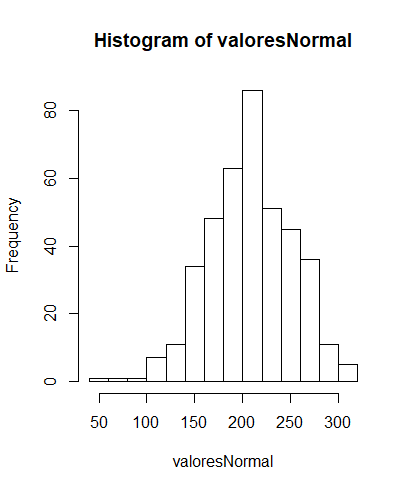
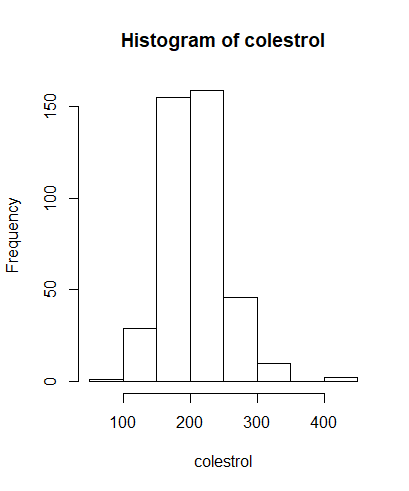
## Colestrol

### Teste a normalidade

Pela análise do gráfico Q-Q plot podemos verificar que esta variável é relativamente normal, visto que se aproxima bastante da linha.

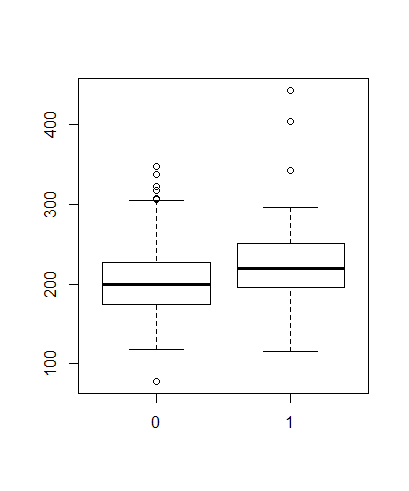


Pela análise do histograma da variável e do histograma da distribuição simulado com os valores da variável podemos verificar que são relativamente simétricos e aparentam representar o comportamento de uma variável normal.



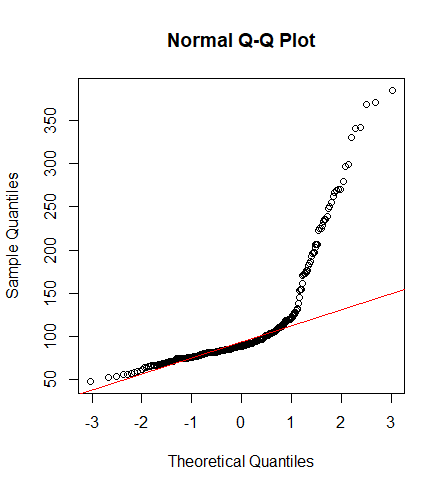
### Teste à influência

Pela análise do boxplot que opõe o colesterol à variável resposta derivada da glyhb , podemos observar que esta variável parece ter alguma importância na variável resposta. Quanto maior o colesterol maior o número de casos de indivíduos com diabetes

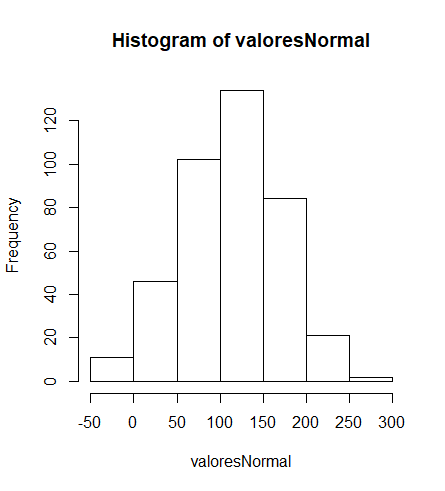
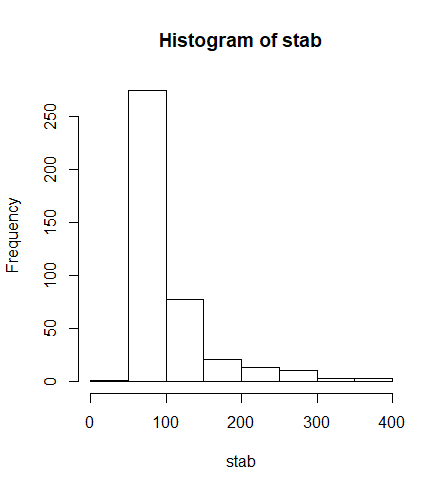


## Stab.glu

### Teste a normalidade

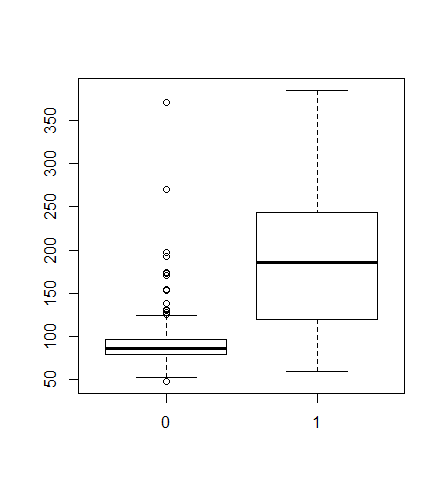
Pela análise do gráfico Q-Q plot podemos verificar que esta variável claramente não apresenta o comportamento de uma variável normal

Pela análise do histograma da variável e do histograma da distribuição simulado com os valores da variável podemos verificar que o histograma da variável é bastante assimétrico e em nada se assemelha ao histograma simulado, confirmando que não segue uma distribuição normal.



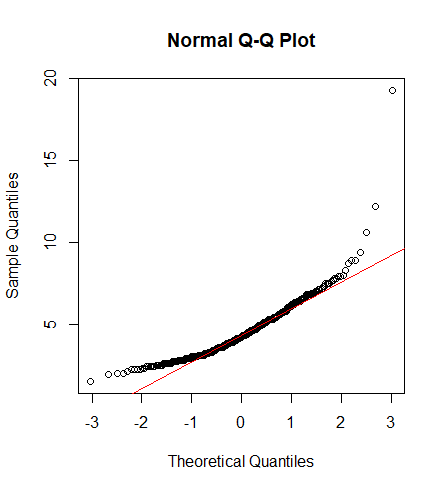
### Teste à influência

### 

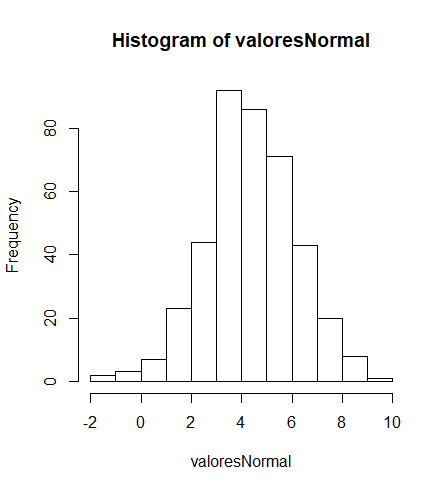
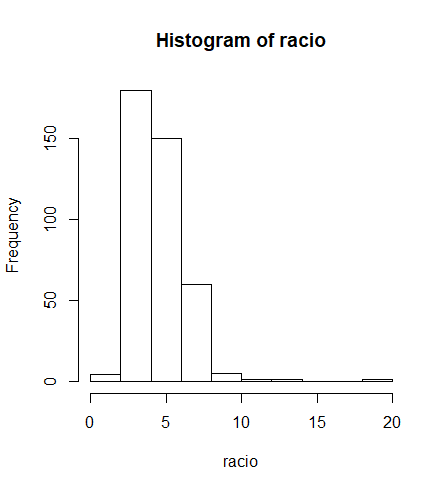
Pela análise do boxplot é facilmente observável que esta variável tem uma importância extrema na variável de resposta. Claramente que nos indivíduos com um valor da stab.glu elevado existe um valor de diabéticos muito superior ao outro grupo.

## Ratio

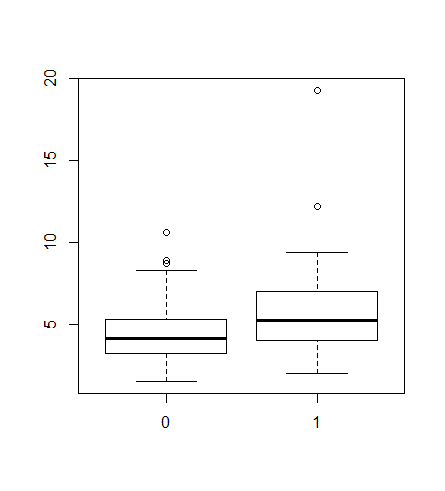
### Teste a normalidade

Pela análise do gráfico Q-Q plot podemos verificar que esta variável é relativamente normal, visto que em certas partes se aproxima bastante da linha, apesar de noutras se afastar bastante.

Pela análise do histograma da variável e do histograma da distribuição simulado com os valores da variável podemos verificar que o histograma da variável é assimétrico e não se assemelha muito ao histograma simulado.



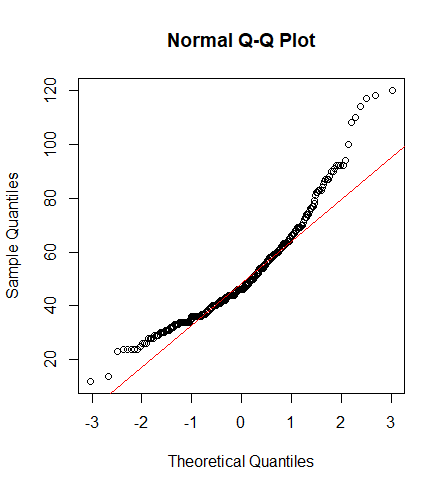
### Teste à influência

Pela análise deste boxplot, verificamos que existe uma ligeira tendência de diabetes nos indivíduos que tem o valor do ratio mais elevado.

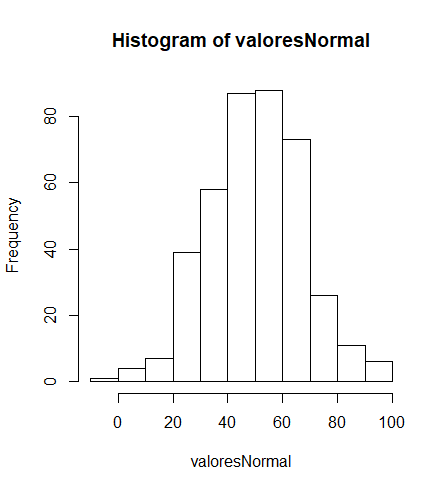
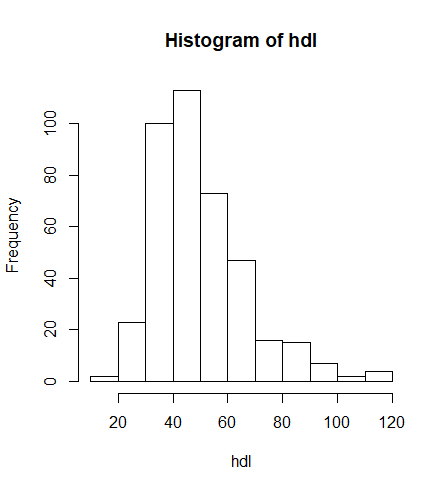
## Hdl

### Teste à normalidade

Pela análise do gráfico Q-Q plot podemos verificar que esta variável não parece apresentar um comportamento muito normal.

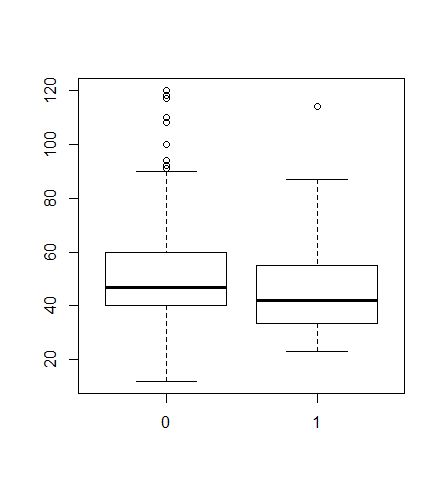


Pela análise do histograma da variável podemos observar que existem alguma simetrias que podem indicar que esta variável se comporta como uma variável normal. O Histograma simulado também parece evidenciar isso.



### Teste à influência

Pela análise deste boxplot, verificamos que existe uma ligeira relação inversa entre o valor do hdl e o número de indivíduos com diabetes.



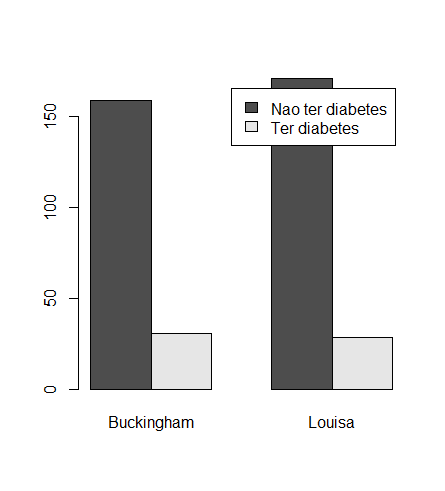
## Location

### Teste à normalidade

Como esta variável é qualitativa este teste não se aplicar.

### Teste à influência

Pela análise deste barplot fica claro que a localização não tem importância neste caso na explicação da variável resposta.



## Ranking

Após efetuados os testes mencionados em cima e feita uma análise simples sobre o resultado dos mesmos criamos um ranking de maneira a ilustrar as conclusões retiradas destes testes, em relação as cinco variáveis estudadas.

1. Stab.glu
2. HDL
3. Colestrol
4. Ratio
5. Location