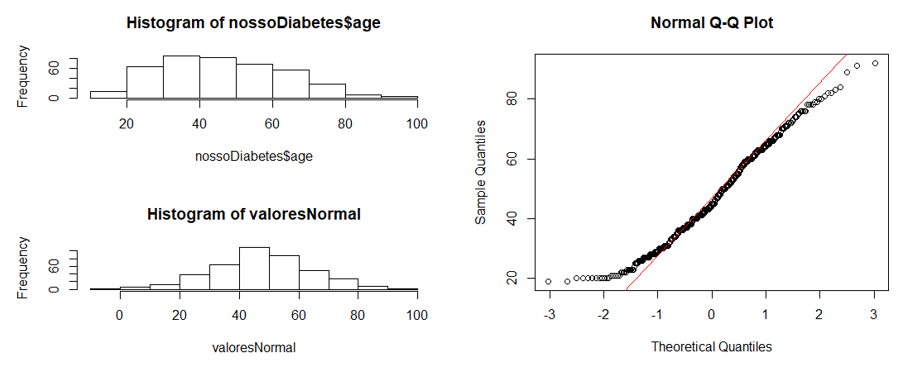
# Análise exploratória

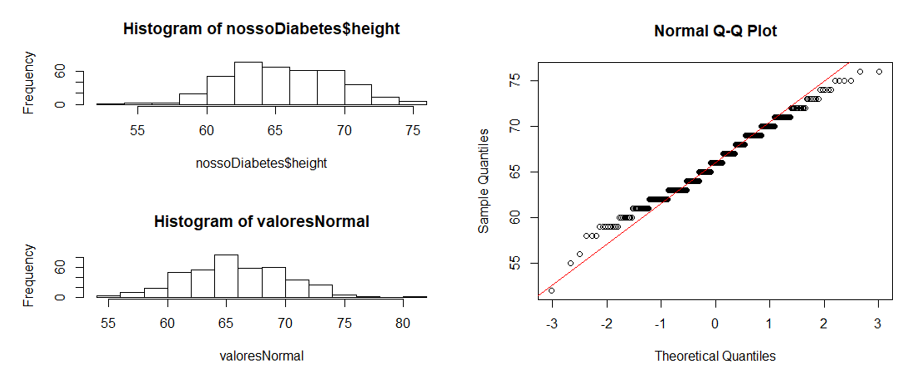
# Teste à normalidade

## Idade (age)



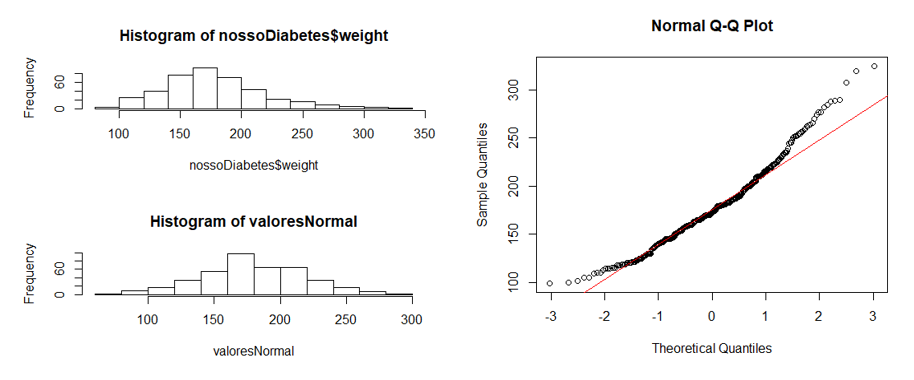
No histograma podemos ver bastantes parecenças com uma distribuição normal apesar desta estar mais concentrada do lado esquerdo do histograma. Já no gráfico podemos observar que na sua maioria (parte central) se aproxima da normalidade, contudo nos extremos diverge um pouco.

## Altura (height)



Quanto ao histograma da altura, é possível verificar que se claramente do histograma normalmente distribuído apresentado abaixo. Já no gráfico podemos concluir que apesar de pequenos desvios, praticamente todos os valores se aproximam de uma distribuição normal.

## Peso (weight)



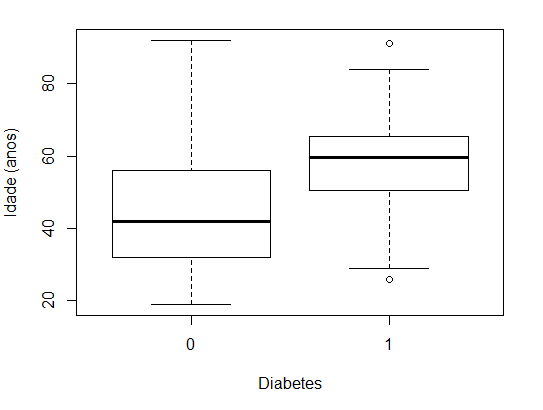
Comparando os histogramas podemos observar que apesar do histograma do peso concentrar a sua distribuição mais à esquerda que o histograma da distribuição normal, os dois são muito semelhantes. Por sua vez, o gráfico mostra convergência nos valores centrais e afasta-se um pouco nos extremos, em especial nos valores mais à direita.

Concluímos então que os preditores que se aproximam da distribuição normal são:

* age
* height
* weight

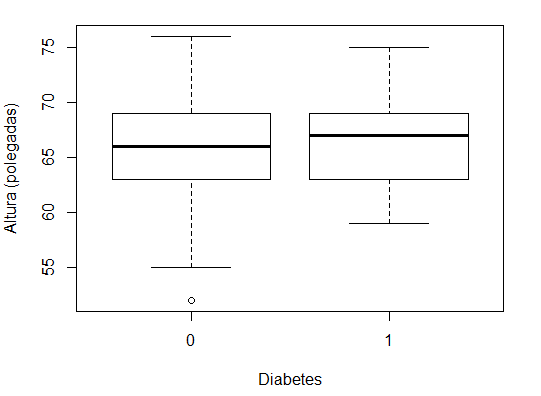
# Teste à influência

## Idade (age)



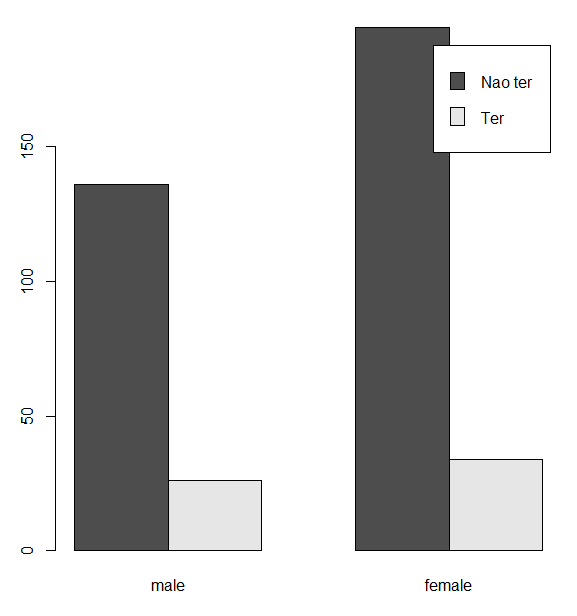
Analisando o boxplot, podemos verificar uma elevada relação entre a idade e o número de indivíduos com diabetes, de tal modo que à medida que a idade aumenta, aumenta também a taxa de deteção de pessoas com diabetes.

## Altura (height)



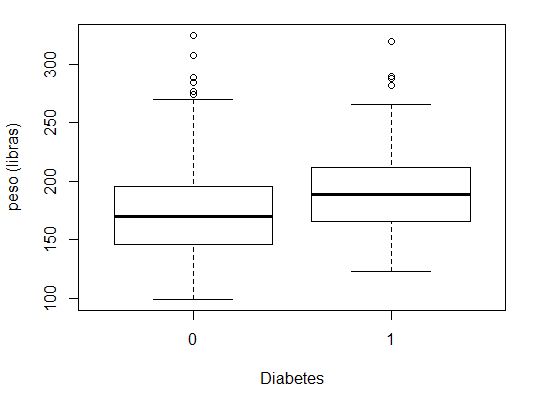
Segundo o boxplot, não parece existir uma relação significativa entre a altura e o número de indivíduos com diabetes, ou seja, a altura de um indivíduo não parece influenciar a presença ou ausência de diabetes no mesmo.

## Sexo (gender)



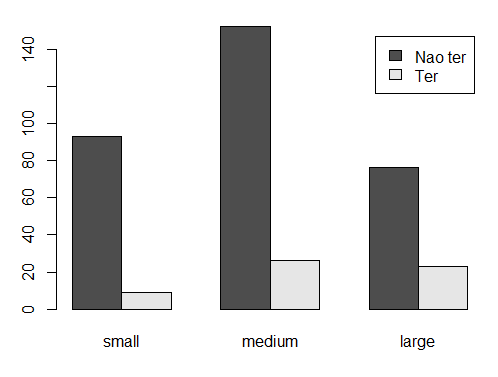
Analisando o gráfico de barras podemos verificar que apesar de se ter detetado mais ocorrências de diabetes em mulheres, foram também realizados testes a mais mulheres que homens. Analisando as percentagens obtivemos uma taxa de ocorrência de diabetes de 16% para os homens e de 15% para as mulheres. Deste modo a presença de diabetes cão parece ser afetada pelo sexo do indivíduo.

## Peso (weight)



Relativamente ao peso, parece existir uma ligeira relação entre este e o número de indivíduos com diabetes, onde a taxa de deteção de diabetes aumenta à medida que o peso também aumenta.

## Frame



Analisando o gráfico podemos verificar que parece existir uma relação entre o frame e o número de indivíduos com diabetes. Analisando as percentagens, obtivemos uma taxa de ocorrência de diabetes de 8.8% para o frame pequeno, de 14.6% para o frame médio e de 23.2% para o frame elevado. Vimos assim que existe uma forte relação entre o frame e o número de indivíduos com diabetes.

## Ranking

Após analisarmos a influência dos vários preditores sobre a variável de resposta, organizamos os mesmos num ranking, do mais influente até ao menos influente:

1. Age
2. Frame
3. Weight
4. Height
5. Gender