

Análisis Estadístico de Enfermedades en el Corazón

Presenta: Pablo César Rodríguez Aguayo

Fecha: 19/Junio/2019

Materia: Fundamentos de Estadística

Parte I: Introducción

Heart Disease UCI (Enfermedades del Corazón UCI)

Esta base de datos contiene 76 atributos, pero todos los experimentos publicados se refieren al uso de un subconjunto de 14 de ellos. En particular, la base de datos de Cleveland es la única que ha sido utilizada por los investigadores de Machine Learning hasta la fecha. El campo "objetivo" se refiere a la presencia de una enfermedad cardíaca en el paciente. Es un valor entero de 0 (sin presencia) a 4. Los experimentos con la base de datos de Cleveland se han concentrado en simplemente intentar distinguir la presencia (valores 1,2,3,4) de la ausencia (valor 0). Los nombres y números de seguridad social de los pacientes se eliminaron recientemente de la base de datos y se reemplazaron con valores ficticios.

Problema Presentado

Dentro de la perspectiva social-sanitaria uno de los mayores retos que enfrentan las organizaciones de salud es la prestación de servicios los cuales permitan al paciente un costo bajo así como accesibilidad, con una calidad superior la cual garantice un diagnóstico objetivo. Un diagnóstico preciso comprende una toma de decisiones correcta la cual permita generar un diagnóstico preciso para el paciente. Una decisión médica tomada de una forma no adecuada puede conducir a problemas mayores de salud donde cada vez se vuelven más costosos y difíciles de tratar, ya en última instancia pueden traer al paciente consecuencias desastrosas incluso llegando a la muerte. Habitualmente, las decisiones tomadas por el personal clínico a menudo son tomadas basándose en la intuición de los médicos y experiencia en lugar de las observaciones estadísticas de la información presentada, lo cual termina por llevar a errores de tratamiento, costos excesivos y efectos en la calidad de los servicios médicos de los pacientes.

Significado del Diagnóstico

El diagnóstico de enfermedad cardíaca se realiza mediante una combinación de signos clínicos y resultados de pruebas. Los tipos de pruebas que se realizarán se elegirán en función de lo que el médico cree que está pasando, que va desde los electrocardiogramas y las tomografías computarizadas (TC) cardíacas hasta las pruebas de sangre y las pruebas de esfuerzo con ejercicios.

Alcance del Estudio

Para propósitos de esta investigación, se pretende realizar un análisis exploratorio de los factores que influyen sobre las enfermedades cardíacas. Lo cual permitirá posteriormente aplicar diversas técnicas de minería de datos para descubrir patrones dentro de los datos que nos permitan obtener mediciones y descripciones de los datos a través de la estadística descriptiva.

Análisis Exploratorio de Datos

El análisis exploratorio de datos permitirá resumir las características que rodean el fenómeno en cuestión. La importancia del análisis exploratorio permite proporcionar y definir la percepción del conjunto de datos utilizado. Permitiendo definir los objetivos a través del análisis de los datos mediante resúmenes numéricos los cuales se enfocan en el aspecto estadístico. De igual manera, permiten enfocarse en los objetivos de la investigación a través de la delimitación de los datos no necesarios.

Recolección de Datos y Preprocesamiento

La recolección de datos y preprocesamiento estuvo a cargo de:

1. Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D.
2. University Hospital, Zurich, Switzerland: William Steinbrunn, M.D.
3. University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D.
4. V.A. Medical Center, Long Beach and Cleveland Clinic Foundation: Robert Detrano, M.D., Ph.D.

Siendo donado por: David W. Aha (aha '@' ics.uci.edu) (714) 856-8779

Por lo que se desconocen los métodos de recolección y preprocesamiento empleados.

Parte II: Análisis Exploratorio

Análisis de las Características

Librerías Necesarias

```
In [91]:
library(knitr)
library(moments)
library(ggplot2)
library(GGally)

options(warn=-1)
```

Descripción

```
In [68]:
data <- read.csv(file = "../data/heart.csv", header = TRUE, sep = ",", stringsAsFactors=FALSE)
heart data <- data.frame(data)
```

Dentro de las características del conjunto de datos, se presentan las siguientes:

- **age:** Edad en años.
- **sex:** Sexo como variable dicotómica (1 = masculino, 0 = femenino)
- **cp:** Tipo de dolor en el pecho.
- **trestbps:** Presión arterial en reposo (en mmHg al ingresar en el hospital)
- **chol:** Colesterol sérico en mg/dl
- **fbs:** Azúcar en la sangre en ayunas >120mg/dl (1 = verdadero, 0 = falso)
- **restecg:** Resultados electrocardiográficos en reposo.
- **thalach:** Ritmo Cardíaco Máximo Alcanzado
- **exang:** Angina Inducida por el Ejercicio
- **oldpeak:** Depresión ST inducida por el ejercicio en relación con el descanso.
- **slope:** La pendiente del segmento pico del ejercicio ST.
- **ca:** Número de vasos principales (0-3) coloreados por fluoroscopia
- **thal:** 3 = normal, 6 = defecto fijo, 7 = defecto reversible
- **target:** Enfermo = 1, no enfermo = 0.

Cabecera

```
In [3]:
head(data)
```

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
57	1	0	140	192	0	1	148	0	0.4	1	0	1	1

Cola

In [4]:

```
tail(data, 5)
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
299	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
300	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
301	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0
302	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
303	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

El conjunto de datos está conformado por 14 características y 303 observaciones.

Máximos y Mínimos

Como se ha observado, el conjunto de datos tiene un número considerable de valores por lo que una de las metas es describir estas características y obtener una conclusión. El uso de máximos y mínimos permite el cálculo de otras medidas como el rango el cual se verá posteriormente. Además permite determinar los valores atípicos dentro del conjunto de datos, ya sean los valores atípicos menores o mayores.

In [5]:

```
# --- age ---
women_age <- subset(heart_data, sex == 0)[c(1)]
men_age <- subset(heart_data, sex == 1)[c(1)]

# --- trestbps ---
women_trestbps <- subset(heart_data, sex == 0)[c(4)]
men_trestbps <- subset(heart_data, sex == 1)[c(4)]

# --- chol ---
women_chol <- subset(heart_data, sex == 0)[c(5)]
men_chol <- subset(heart_data, sex == 1)[c(5)]

# --- thalach ---
women_thalach <- subset(heart_data, sex == 0)[c(8)]
men_thalach <- subset(heart_data, sex == 1)[c(8)]

# --- oldpeak ---
women_oldpeak <- subset(heart_data, sex == 0)[c(10)]
men_oldpeak <- subset(heart_data, sex == 1)[c(10)]
```

Máximos

A continuación se hace un análisis de los valores máximos de las variables: edad, presión arterial, colesterol, ritmo cardiaco y depresión de los pacientes. Separado por sexo, haciendo una comparación con el análisis en general.

In [6]:

```
max_vals_men <- c(max(men_age$age), max(men_trestbps$trestbps), max(men_chol$chol), max(men_thalach$thalach)
, max(men_oldpeak$oldpeak))
max_vals_women <- c(max(women_age$age), max(women_trestbps$trestbps), max(women_chol$chol), max(women_thalach$thalach), max(women_oldpeak$oldpeak))
max_vals_general <- c(max(heart_data$age), max(heart_data$trestbps), max(heart_data$chol), max(heart_data$thalach), max(heart_data$oldpeak))

data_gen <- cbind(max_vals_general ,max_vals_men, max_vals_women)
df.max_val <- data.frame(data_gen)
colnames(df.max_val) <- sex_ <- c('General', 'Hombre', 'Mujer')
rownames(df.max_val) <- c('Edad', 'Presión Arterial', 'Colesterol', 'Ritmo Cardíaco', 'Depresión ST')

df.max_val
```

	General	Hombre	Mujer
Edad	77.0	77.0	76.0
Presión Arterial	200.0	192.0	200.0
Colesterol	564.0	353.0	564.0
Ritmo Cardíaco	202.0	202.0	192.0
Depresión ST	6.2	5.6	6.2

Comentarios: Cabe destacar que de lo anterior visto, la *edad* máxima es perteneciente a los hombres por un año. En cuanto a *presión arterial*, se puede ver que el valor mayor corresponde a una mujer tomando el valor de 200, siendo el máximo en hombres menor por 8 unidades. Así mismo pasa con el *colesterol*, siendo el valor máximo encontrado en el subconjunto de las mujeres tomando el valor de 564 y en hombres de 353. Dada la naturaleza aún no estudiada del colesterol, hasta este punto pudiera ser que el colesterol máximo encontrado es un valor atípico. Respecto al *ritmo cardíaco*, se puede apreciar que es mayor en los hombres con un valor de 202, mientras que en mujeres es 10 unidades menor. Por último, se puede ver que en la depresión post-entrenamiento es ligeramente mayor en mujeres por 0.6, tomando las mujeres el valor máximo nuevamente. Hasta este punto, basandose en los aspectos analizados se puede ver que las mujeres tienen los valores máximos en tres de cinco características analizadas, siendo una de ellas un probable valor atípico.

Mínimos

Se define como el valor matemático menor en el conjunto de datos.

In [7]:

```
min_vals_men <- c(min(men_age$age), min(men_trestbps$trestbps), min(men_chol$chol), min(men_thalach$thalach)
, min(men_oldpeak$oldpeak))
min_vals_women <- c(min(women_age$age), min(women_trestbps$trestbps), min(women_chol$chol), min(women_thalach$thalach), min(women_oldpeak$oldpeak))
min_vals_general <- c(min(heart_data$age), min(heart_data$trestbps), min(heart_data$chol), min(heart_data$thalach), min(heart_data$oldpeak))

data_gen <- cbind(min_vals_general ,min_vals_men, min_vals_women)
df.min_val <- data.frame(data_gen)
colnames(df.min_val) <- sex_ <- c('General', 'Hombre', 'Mujer')
rownames(df.min_val) <- c('Edad', 'Presión Arterial', 'Colesterol', 'Ritmo Cardíaco', 'Depresión ST')

df.min_val
```

	General	Hombre	Mujer
Edad	29	29	34
Presión Arterial	94	94	94
Colesterol	126	126	141
Ritmo Cardíaco	71	71	96
Depresión ST	0	0	0

Comentarios: De acuerdo a los valores mínimos, respecto a la *edad* podemos decir que siendo el valor mínimo perteneciente a un hombre con 29 años y el valor máximo a un hombre igualmente, se puede decir que es más ámplio el rango de edades en los varones. Para este caso, la *presión arterial*, presentó un caso con un valor mínimo de 94 tanto para hombres como para mujeres. Para el *colesterol* se puede ver que el valor mínimo se encuentra en el subconjunto de los hombres, con un valor de 126, mientras que las mujeres poseen un valor mínimo de 141. Para el caso del *ritmo cardíaco*, la presencia en hombres es con un valor mínimo de 71, mientras que en el subconjunto de las mujeres es de 96. Por último, se puede ver que para la depresión post-entrenamiento, el valor para todos es 0, lo cual puede ser un posible valor atípico encontrado dentro del conjunto de datos.

Cuartiles, Percentiles y Valores Atípicos

Cuartiles

En términos generales, los cuartiles dividen el conjunto de datos en cuatro partes iguales, con observaciones sobre el tercer cuartil que constituyen el cuarto superior de los datos del conjunto, siendo el segundo cuartil idéntico a la mediana, y el primer cuartil separado al cuarto más bajo de los tres cuartos superiores. Los valores comunmente esperados se encuentran entre el primer cuartil y el tercero, mientras que los valores atípicos se encuentran ya sea por encima del tercer cuartil o bien por debajo del primer cuartil. A continuación se hace el análisis de los cuartiles de los atributos.

Cuartiles Mujeres

```
In [8]:
q_vals_women <- data.frame(
  quantile(women_age$age),
  quantile(women_trestbps$trestbps),
  quantile(women_chol$chol),
  quantile(women_thalach$thalach),
  quantile(women_oldpeak$oldpeak)
)

colnames(q_vals_women) <- c('Edad', 'Presión Arterial', 'Colesterol', 'Ritmo Cardíaco', 'Depresión ST')
q_vals_women
```

	Edad	Presión Arterial	Colesterol	Ritmo Cardíaco	Depresión ST
0%	34.00	94	141.00	96.00	0.0
25%	49.75	120	214.75	141.25	0.0
50%	57.00	131	253.00	157.00	0.6
75%	63.00	140	296.75	165.00	1.4
100%	76.00	200	564.00	192.00	6.2

Cuartiles Hombres

```
In [9]:
q_vals_men <- data.frame(
  quantile(men_age$age),
  quantile(men_trestbps$trestbps),
  quantile(men_chol$chol),
  quantile(men_thalach$thalach),
  quantile(men_oldpeak$oldpeak)
)

colnames(q_vals_men) <- c('Edad', 'Presión Arterial', 'Colesterol', 'Ritmo Cardíaco', 'Depresión ST')
q_vals_men
```

	Edad	Presión Arterial	Colesterol	Ritmo Cardíaco	Depresión ST
0%	29.0	94	126	71	0.0
25%	47.0	120	208	132	0.0
50%	54.0	130	235	151	0.8
75%	59.5	140	268	168	1.8
100%	77.0	192	353	202	5.6

Cuartiles de Diagnóstico en General

In [10]:

```
q_vals_general <- data.frame(
  quantile(heart_data$age),
  quantile(heart_data$trestbps),
  quantile(heart_data$chol),
  quantile(heart_data$thalach),
  quantile(heart_data$oldpeak)
)

colnames(q_vals_general) <- c('Edad', 'Presión Arterial', 'Colesterol', 'Ritmo Cardíaco', 'Depresión ST')
q_vals_general
```

	Edad	Presión Arterial	Colesterol	Ritmo Cardíaco	Depresión ST
0%	29.0	94	126.0	71.0	0.0
25%	47.5	120	211.0	133.5	0.0
50%	55.0	130	240.0	153.0	0.8
75%	61.0	140	274.5	166.0	1.6
100%	77.0	200	564.0	202.0	6.2

Percentiles

Del mismo modo, un conjunto de datos se puede dividir aún más finamente usando percentiles; el percentil 99 separa el 1% más alto del 99% inferior, y así sucesivamente. A menos que el número de observaciones sea múltiplo de 100.

Nota: El percentil en la posición 50, el segundo cuartil, la mediana y el quinto decil son el mismo valor el cual está representado mediante la siguiente fórmula.

$$P_{50} = \tilde{x} = Q_2 = D_5$$

A continuación analizaremos los percentiles desde la perspectiva de un deciles.

Percentiles en Mujeres

In [11]:

```
q_vals_women <- data.frame(
  quantile(women_age$age, c(.10, .20, .30, .40, .50, .60, .70, .80, .90, 1.0)),
  quantile(women_trestbps$trestbps, c(.10, .20, .30, .40, .50, .60, .70, .80, .90, 1.0)),
  quantile(women_chol$chol, c(.10, .20, .30, .40, .50, .60, .70, .80, .90, 1.0)),
  quantile(women_thalach$thalach, c(.10, .20, .30, .40, .50, .60, .70, .80, .90, 1.0)),
  quantile(women_oldpeak$oldpeak, c(.10, .20, .30, .40, .50, .60, .70, .80, .90, 1.0))
)

colnames(q_vals_women) <- c('Edad', 'Presión Arterial', 'Colesterol', 'Ritmo Cardíaco', 'Depresión ST')
q_vals_women
```

	Edad	Presión Arterial	Colesterol	Ritmo Cardíaco	Depresión ST
10%	42.0	109.0	197.0	121.5	0.0
20%	46.0	120.0	209.0	133.0	0.0
30%	51.0	121.0	224.0	144.0	0.0
40%	54.0	130.0	240.0	152.0	0.2
50%	57.0	131.0	253.0	157.0	0.6
60%	59.0	136.0	268.0	160.0	0.9
70%	62.0	140.0	285.5	163.0	1.2
80%	64.0	146.0	305.0	169.0	1.5
90%	66.5	157.5	335.0	172.0	2.0
100%	76.0	200.0	564.0	192.0	6.2

Percentiles en Hombres

In [12]:

```
q_vals_men <- data.frame(
  quantile(men_age$age, c(.10, .20, .30, .40, .50, .60, .70, .80, .90, 1.0)),
  quantile(men_trestbps$trestbps, c(.10, .20, .30, .40, .50, .60, .70, .80, .90, 1.0)),
  quantile(men_chol$chol, c(.10, .20, .30, .40, .50, .60, .70, .80, .90, 1.0)),
  quantile(men_thalach$thalach, c(.10, .20, .30, .40, .50, .60, .70, .80, .90, 1.0)),
  quantile(men_oldpeak$oldpeak, c(.10, .20, .30, .40, .50, .60, .70, .80, .90, 1.0))
)

colnames(q_vals_men) <- c('Edad', 'Presión Arterial', 'Colesterol', 'Ritmo Cardíaco', 'Depresión ST')
q_vals_men
```

	Edad	Presión Arterial	Colesterol	Ritmo Cardíaco	Depresión ST
10%	42	110	185.6	113.6	0.0
20%	45	120	203.2	127.2	0.0
30%	49	120	213.8	140.0	0.0
40%	52	125	227.4	144.0	0.4
50%	54	130	235.0	151.0	0.8
60%	57	132	247.0	158.0	1.2
70%	59	140	260.0	163.0	1.6
80%	61	142	275.8	171.0	2.0
90%	65	152	299.0	178.4	2.8
100%	77	192	353.0	202.0	5.6

Percentiles de Diagnóstico en General

In [13]:

```
q_vals_general <- data.frame(
  quantile(heart_data$age, c(.10, .20, .30, .40, .50, .60, .70, .80, .90, 1.0)),
  quantile(heart_data$trestbps, c(.10, .20, .30, .40, .50, .60, .70, .80, .90, 1.0)),
  quantile(heart_data$chol, c(.10, .20, .30, .40, .50, .60, .70, .80, .90, 1.0)),
  quantile(heart_data$thalach, c(.10, .20, .30, .40, .50, .60, .70, .80, .90, 1.0)),
  quantile(heart_data$oldpeak, c(.10, .20, .30, .40, .50, .60, .70, .80, .90, 1.0))
)

colnames(q_vals_general) <- c('Edad', 'Presión Arterial', 'Colesterol', 'Ritmo Cardíaco', 'Depresión ST')
q_vals_general
```

	Edad	Presión Arterial	Colesterol	Ritmo Cardíaco	Depresión ST
10%	42	110	188.0	116.0	0.00
20%	45	120	204.0	130.0	0.00
30%	50	120	217.6	140.6	0.00
40%	53	126	230.0	146.0	0.38
50%	55	130	240.0	153.0	0.80
60%	58	134	254.0	159.0	1.12
70%	59	140	268.0	163.0	1.40
80%	62	144	285.2	170.0	1.90
90%	66	152	308.8	176.6	2.80
100%	77	200	564.0	202.0	6.20

Valores Atípicos

Se definen como los valores mucho mayores o menores que los demás valores en el conjunto de datos. O bien, se puede definir como el valor que radica por fuera del conjunto de datos establecido las cuales se distinguen por su distancia a los valores normales.

Regla del Pulgar

Siendo una práctica común la *regla del pulgar*, consiste en designar un valor atípico en base a la experiencia obtenida. Para esto, los valores atípicos se calculan mediante dos fórmulas. La primera permite calcular los valores atípicos mínimos, es decir los valores que están por debajo de lo esperado, para esto se emplea la siguiente fórmula:

$$V_{ai} = Q_1 - 1.5 * IQR$$

Donde:

- V_{ai} son los valores atípicos inferiores.
- Q_1 es el valor del cuartil 1.
- IQR es el valor intercuartil.

Mientras que para calcular los valores atípicos por encima de lo esperado, se emplea la siguiente fórmula:

$$V_{as} = Q_3 + 1.5 * IQR$$

Donde:

- V_{as} son los valores atípicos superiores.
- Q_3 es el valor del cuartil 3.
- IQR es el valor intercuartil.

Si bien estas medidas no son utilizadas como la forma adecuada para calcular los valores atípicos, dentro de la industria son ampliamente aceptadas. Sin embargo, estos métodos solo son usados si el analista de datos es considerado de grado midlevel o senior. Para este caso se utilizará el diagrama de caja para ver la distribución de los datos. Para visualizar los valores atípicos se utilizará el diagrama de caja, este permitirá ver el comportamiento dentro del atributo de los datos.

Rango

El rango permite conocer la diferencia entre el valor máximo y mínimo de la característica seleccionada. Esta medida es considerada como medida de dispersión y permite conocer el intervalo que se haya entre la distribución. Se calcula mediante la siguiente fórmula:

$$R_{var} = \max(x) - \min(x)$$

Indicaciones Generales

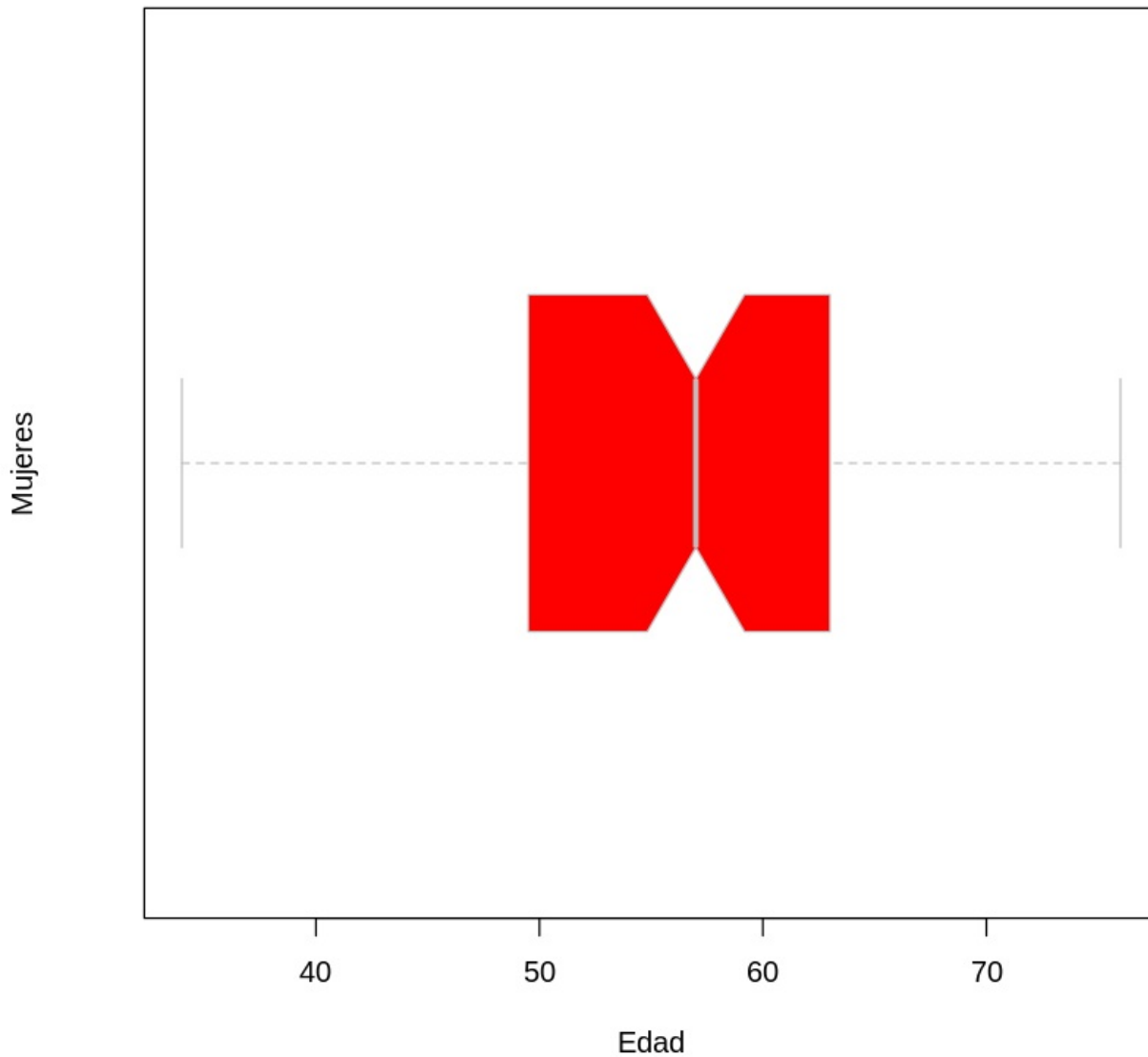
A continuación se realizará un análisis para las características: edad y presión arterial. Mostrando los valores mínimos, máximo, rango segundo cuartil, obteniendo un breve análisis de los datos obtenidos.

Edad

In [14]:

```
boxplot(women_age$age,  
  main = "Edad en Mujeres",  
  xlab = "Edad",  
  ylab = "Mujeres",  
  col = "red",  
  border = "gray",  
  horizontal = TRUE,  
  notch = TRUE  
)
```

Edad en Mujeres

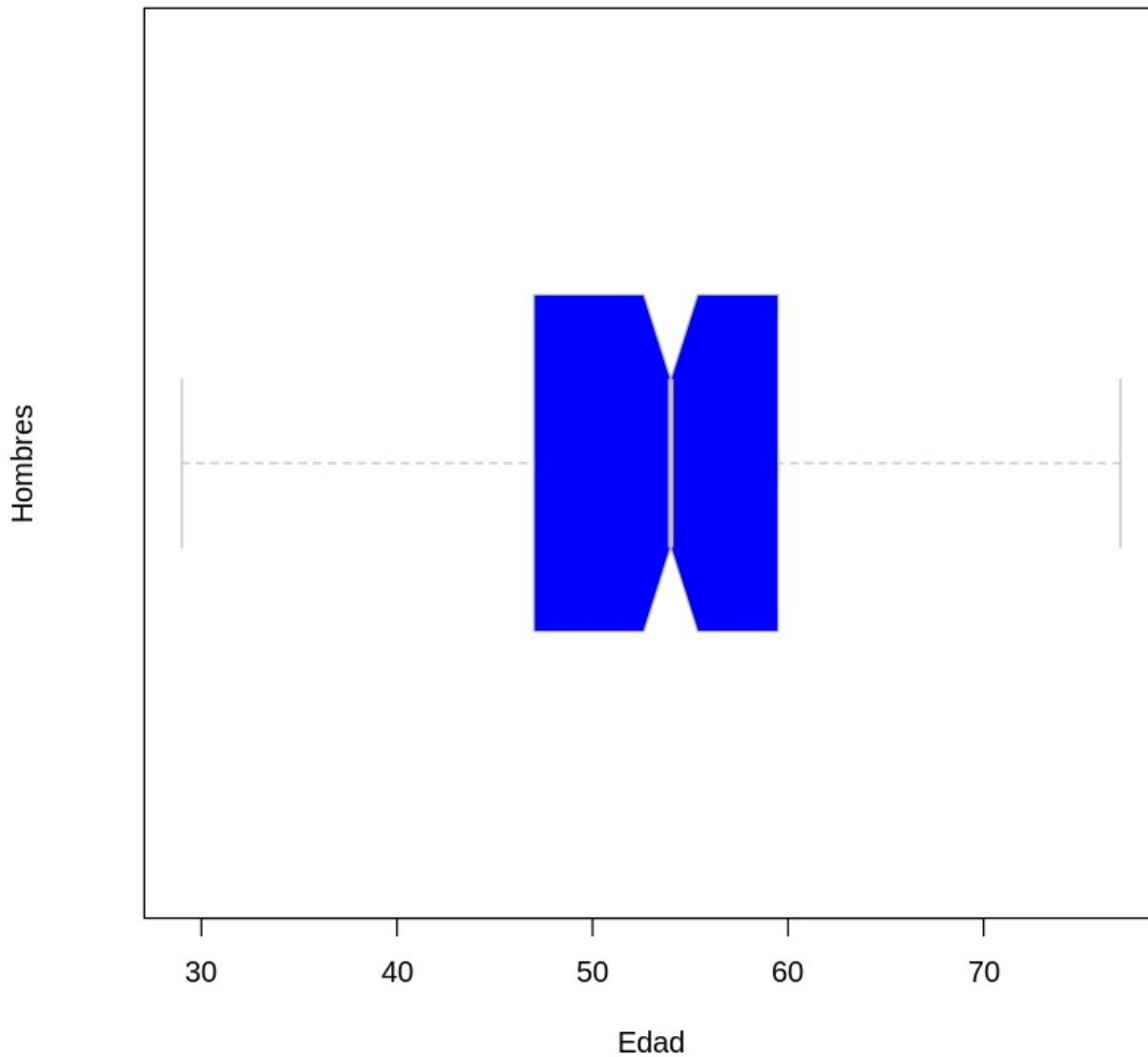


Comentarios: De acuerdo a la gráfica anterior se puede apreciar que el valor mínimo está en 34 mientras que su máximo es de 76, con un rango de 42. A su vez se puede apreciar que no hay valores atípicos y que su segundo cuartil es de 57.

In [15]:

```
boxplot(men_age$age,  
  main = "Edad en Hombres",  
  xlab = "Edad",  
  ylab = "Hombres",  
  col = "blue",  
  border = "gray",  
  horizontal = TRUE,  
  notch = TRUE  
)
```

Edad en Hombres

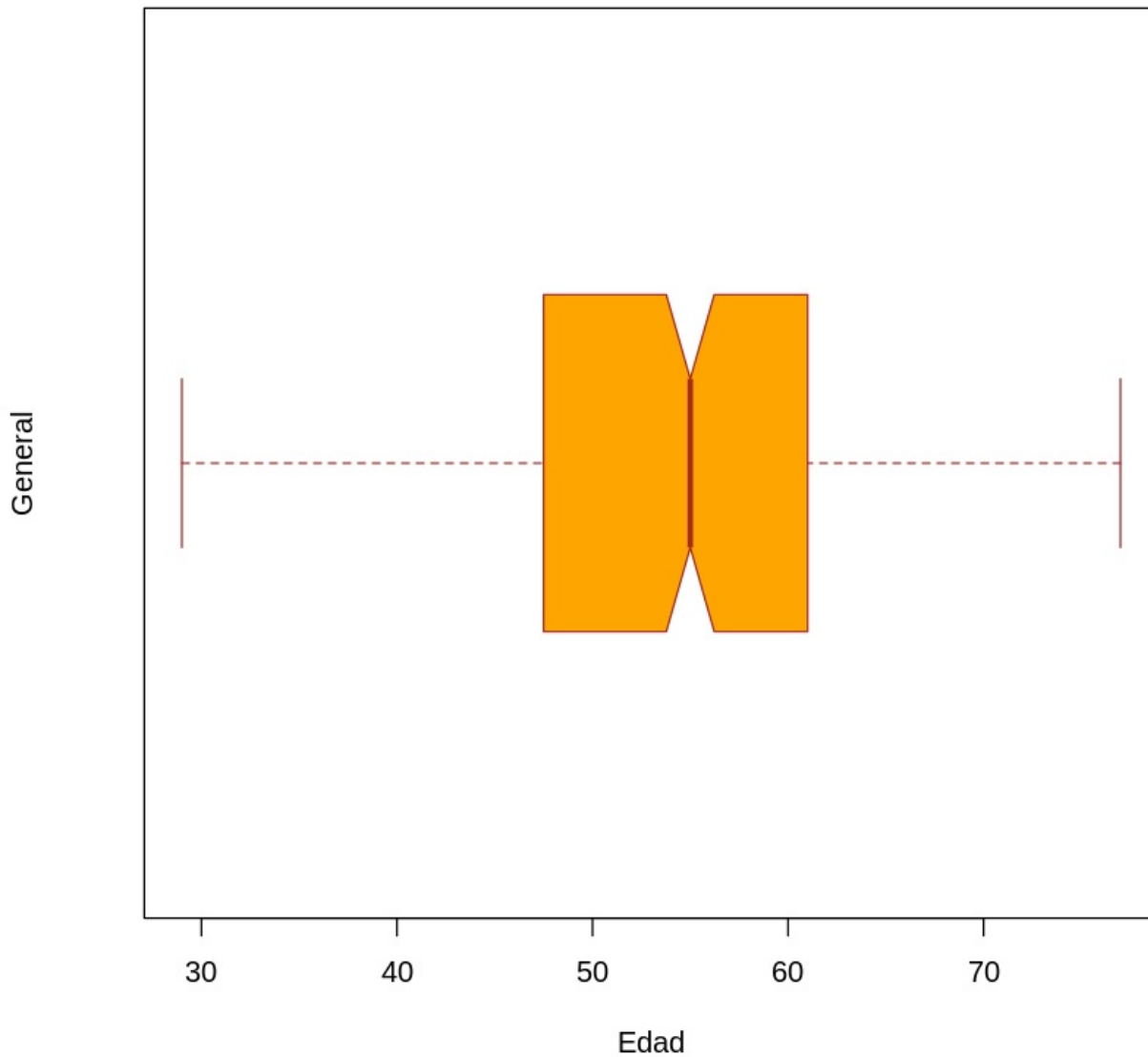


Comentarios: Para el conjunto de los hombres se puede apreciar un valor mínimo de 29 y un máximo de 77, donde su rango es de 48. A su vez se puede apreciar que no hay valores atípicos y que su segundo cuartil es de 54, esto debido al mayor rango de los datos.

In [16]:

```
boxplot(heart_data$age,
  main = "Edad en Pacientes",
  xlab = "Edad",
  ylab = "General",
  col = "orange",
  border = "brown",
  horizontal = TRUE,
  notch = TRUE
)
```

Edad en Pacientes

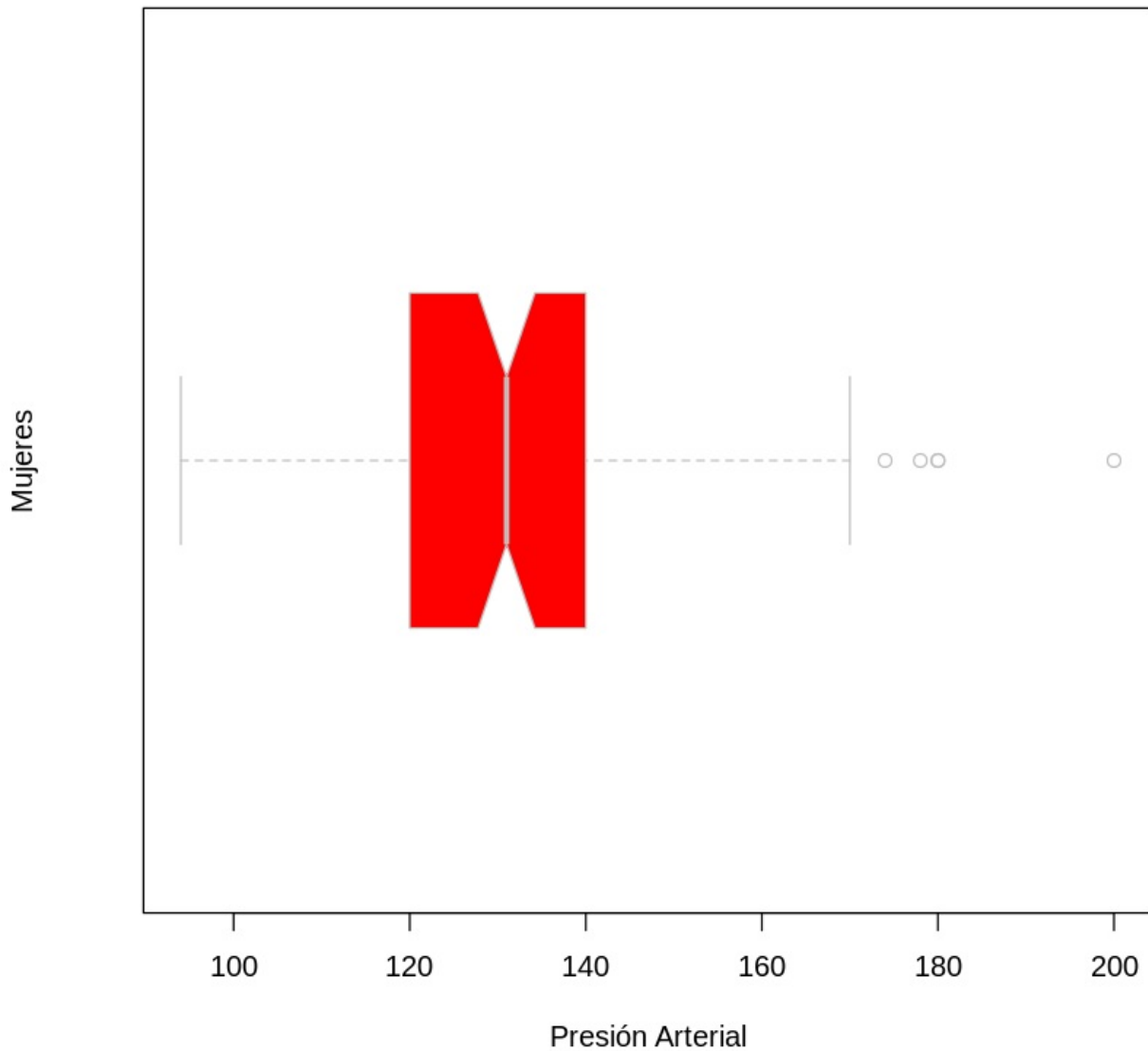


Comentarios: Para el conjunto general se puede apreciar un valor mínimo de 29 y un máximo de 77, donde su rango es de 42. A su vez se puede apreciar que no hay valores atípicos y que su segundo cuartil es de 55, debido a la suma de ambos conjuntos y.

In [17]:

```
boxplot(women_trestbps$trestbps,  
  main = "Presión Arterial en Mujeres",  
  xlab = "Presión Arterial",  
  ylab = "Mujeres",  
  col = "red",  
  border = "gray",  
  horizontal = TRUE,  
  notch = TRUE  
)
```

Presión Arterial en Mujeres

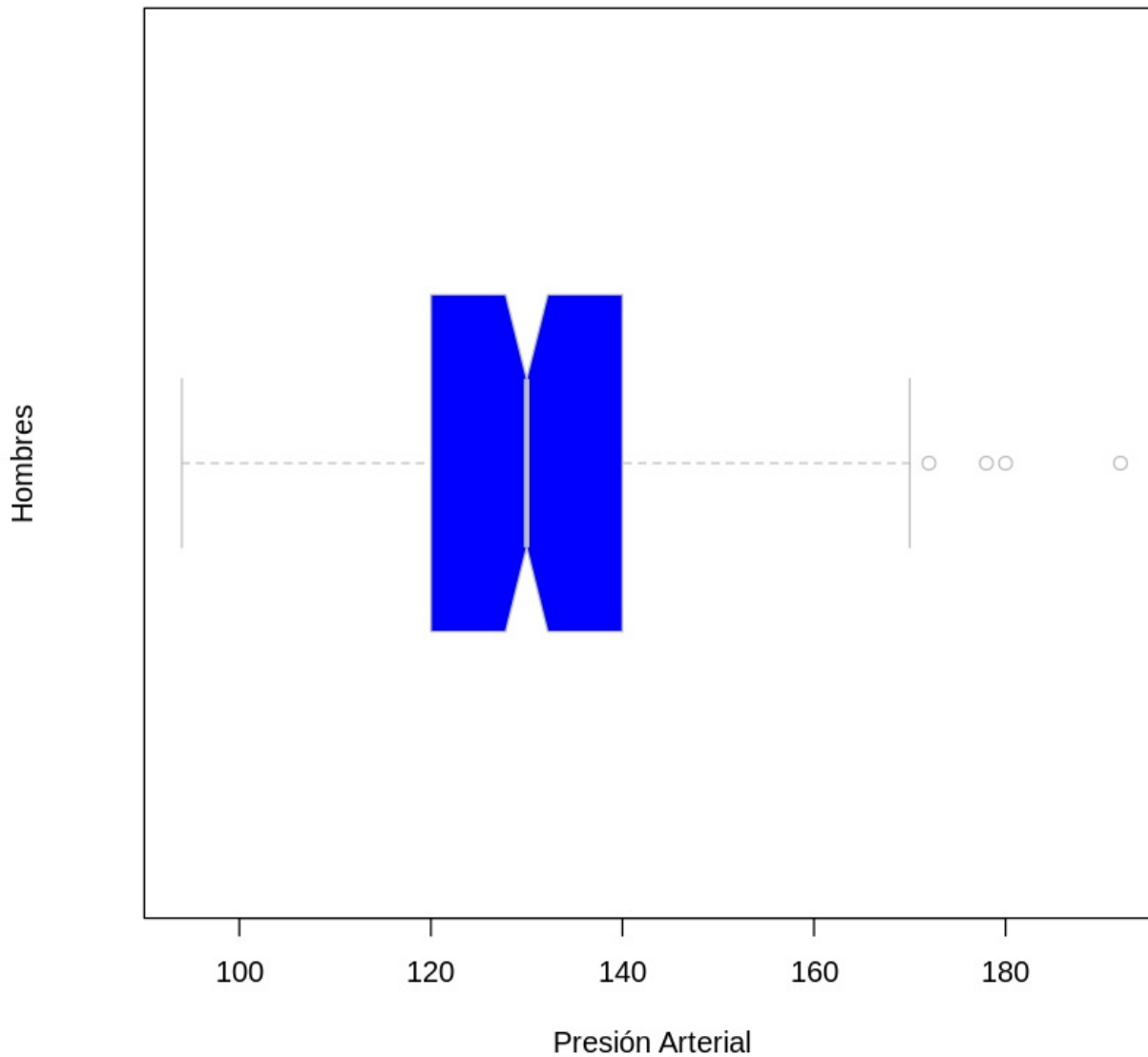


Comentarios: En el subconjunto de las mujeres se puede apreciar un valor mínimo de 94 y un máximo de 200, donde su rango es de 106. Presentando valores atípicos después del tercer cuartil y con segundo cuartil de 131 indicando la mitad de los datos.

In [18]:

```
# --- trestbps ---  
boxplot(men_trestbps$trestbps,  
  main = "Presión Arterial en Hombres",  
  xlab = "Presión Arterial",  
  ylab = "Hombres",  
  col = "blue",  
  border = "gray",  
  horizontal = TRUE,  
  notch = TRUE  
)
```

Presión Arterial en Hombres

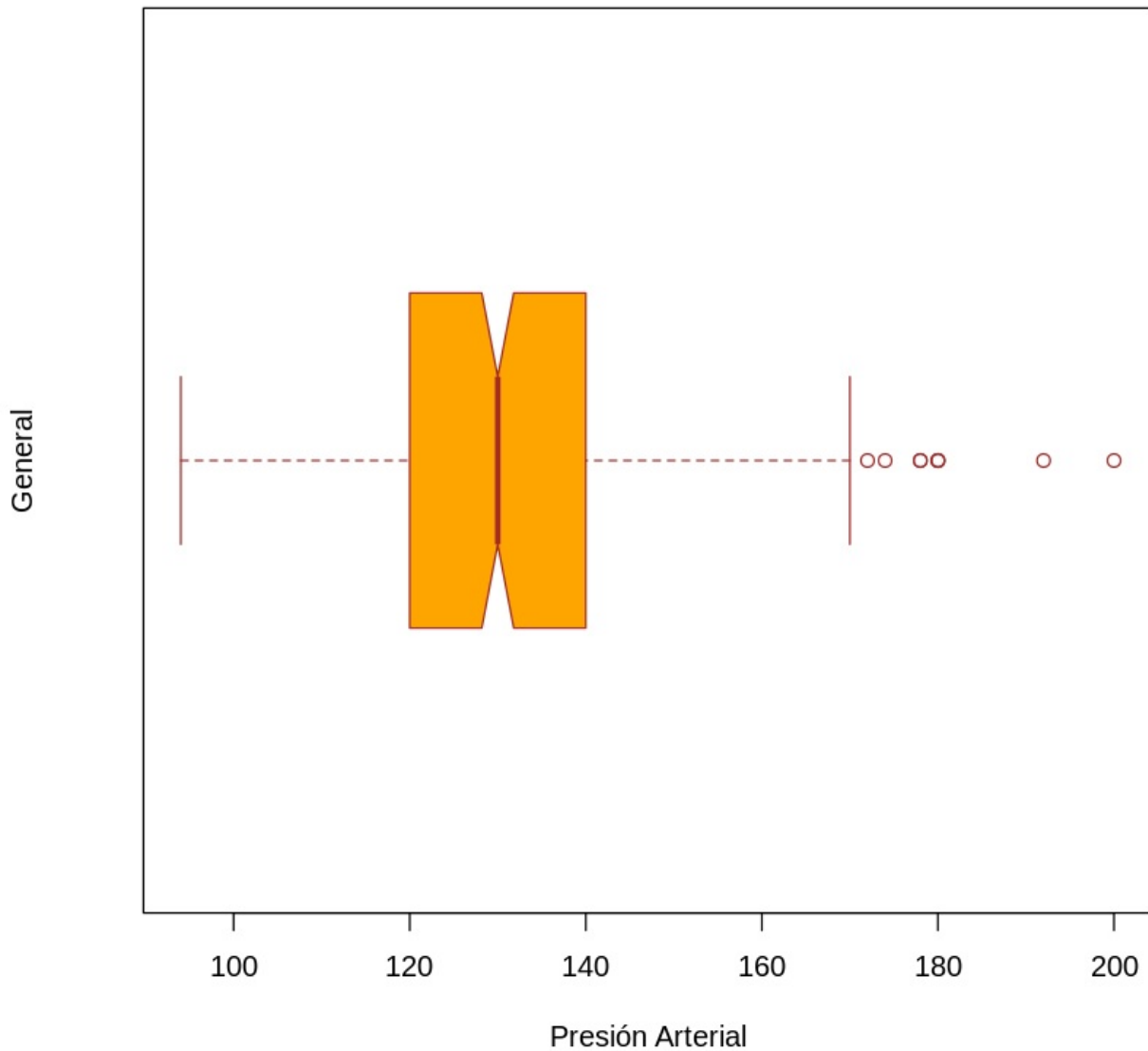


Comentarios: En el subconjunto de los hombres se puede apreciar un valor mínimo de 94 y un máximo de 192 siendo que menor el rango que el de las mujeres por 8, donde su rango es de 98. Presentando valores atípicos despues del tercer cuartil y un segundo cuartil de 131 indicando la mitad de los datos.

In [19]:

```
# --- trestbps ---  
boxplot(heart data$trestbps,  
  main = "Presión Arterial en Pacientes",  
  xlab = "Presión Arterial",  
  ylab = "General",  
  col = "orange",  
  border = "brown",  
  horizontal = TRUE,  
  notch = TRUE  
)
```

Presión Arterial en Pacientes



Comentarios: Por último, en el conjunto general se puede ver un valor mínimo de 94 y un máximo de 200 con un rango de 106. Presentando valores atípicos después del tercer cuartil y un segundo cuartil de 130.

Medidas de Tendencia Central

Las medidas de tendencia central permiten obtener indicadores preeliminares que brindan una interpretación numérica de las medidas obtenidas.

Media

Primer momento alrededor del cero es la media o valor esperado de la variable aleatoria y se denota por μ . La media de una variable aleatoria se considera como una cantidad numérica alrededor de la cual los valores de la variable aleatoria tienden a agruparse. Por lo tanto, la media es una medida de tendencia central. Para el cálculo de la media se utiliza:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Mediana

La mediana en un conjunto de datos es el valor medio cuando los valores son ordenados de manera ascendente o descendente. Igualmente, en este punto será calculada la mediana para futuras referencias, si n es par entonces se calcula mediante:

$$\tilde{x} = x_{(n + 1)/2}$$

Por otro lado si n es impar se calcula:

$$\tilde{x} = x_{(n + 1)/2} + x_{n/2}$$

Moda

La moda es un término estadístico que se refiere al número que aparece con mayor frecuencia en un conjunto de números. Esta se encuentra al recopilar y organizar datos para contar la frecuencia de cada resultado. Donde radica sobre su recuento más alto de ocurrencias del conjunto, igualmente se le conoce como el valor modal.

$$\text{Varlor repetido con más frecuencia}$$

Conclusiones Generales

Para concluir las medidas de tendencia central se mostrará un histograma de frecuencias el cual contendrá un diagrama de densidad, que nos permite ver la inclinación de la curva, el promedio o media de los datos, la mediana y la desviación estándar como medida de referencia.

Colesterol

In [20]:

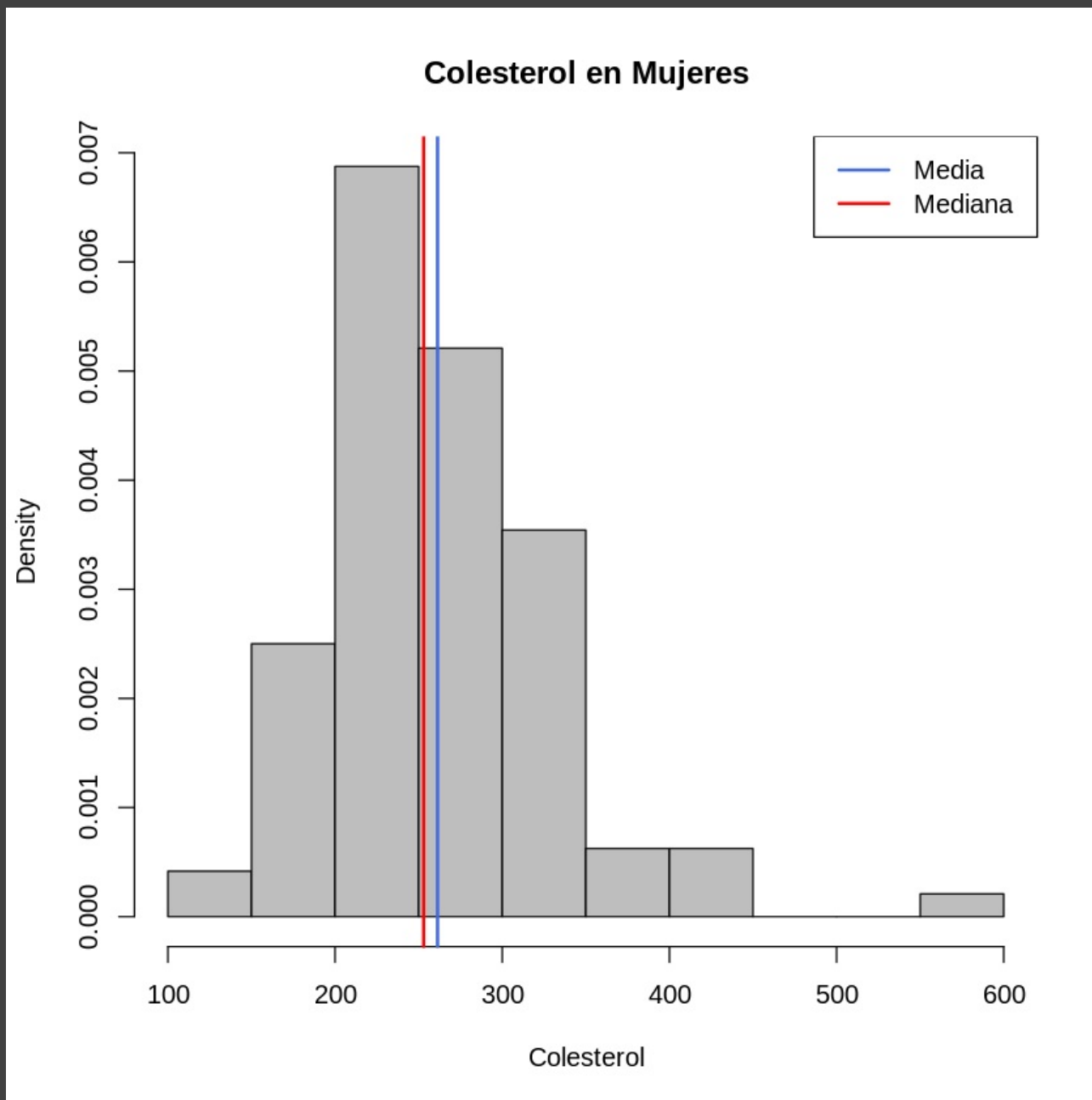
```
Mode <- function(x) {
  ta = table(x)
  tam = max(ta)
  if (all(ta == tam))
    mod = NA
  else
    if(is.numeric(x))
      mod = as.numeric(names(ta)[ta == tam])
    else
      mod = names(ta)[ta == tam]
  return(mod)
}

# --- trestbps ---
hist(women_chol$chol,
  col = "gray",
  border = "black",
  prob = TRUE,
  xlab = "Colesterol",
  main = "Colesterol en Mujeres")
abline(v = mean(women_chol$chol), col = "royalblue", lwd = 2)
abline(v = median(women_chol$chol), col = "red", lwd = 2)
legend(x = "topright",
  c("Media", "Mediana"),
  col = c("royalblue", "red", "yellow"),
  lwd = c(2, 2, 2)
)

# --- Tabla ---
table_women_chol <- data.frame(
  mean(women_chol$chol),
  median(women_chol$chol)
)

colnames(table_women_chol) <- c('Media', 'Mediana')
table_women_chol
```


Media	Mediana
261.3021	253



Comentarios: Para esta gráfica se puede apreciar que se obtuvo una mediana media muy cercanas, donde la media tuvo un valor de 261.3, mientras que la mediana tuvo un valor de 253. Se puede observar que la diferencia entre ambas medidas sólo fue de 12 unidades.

In [21]:

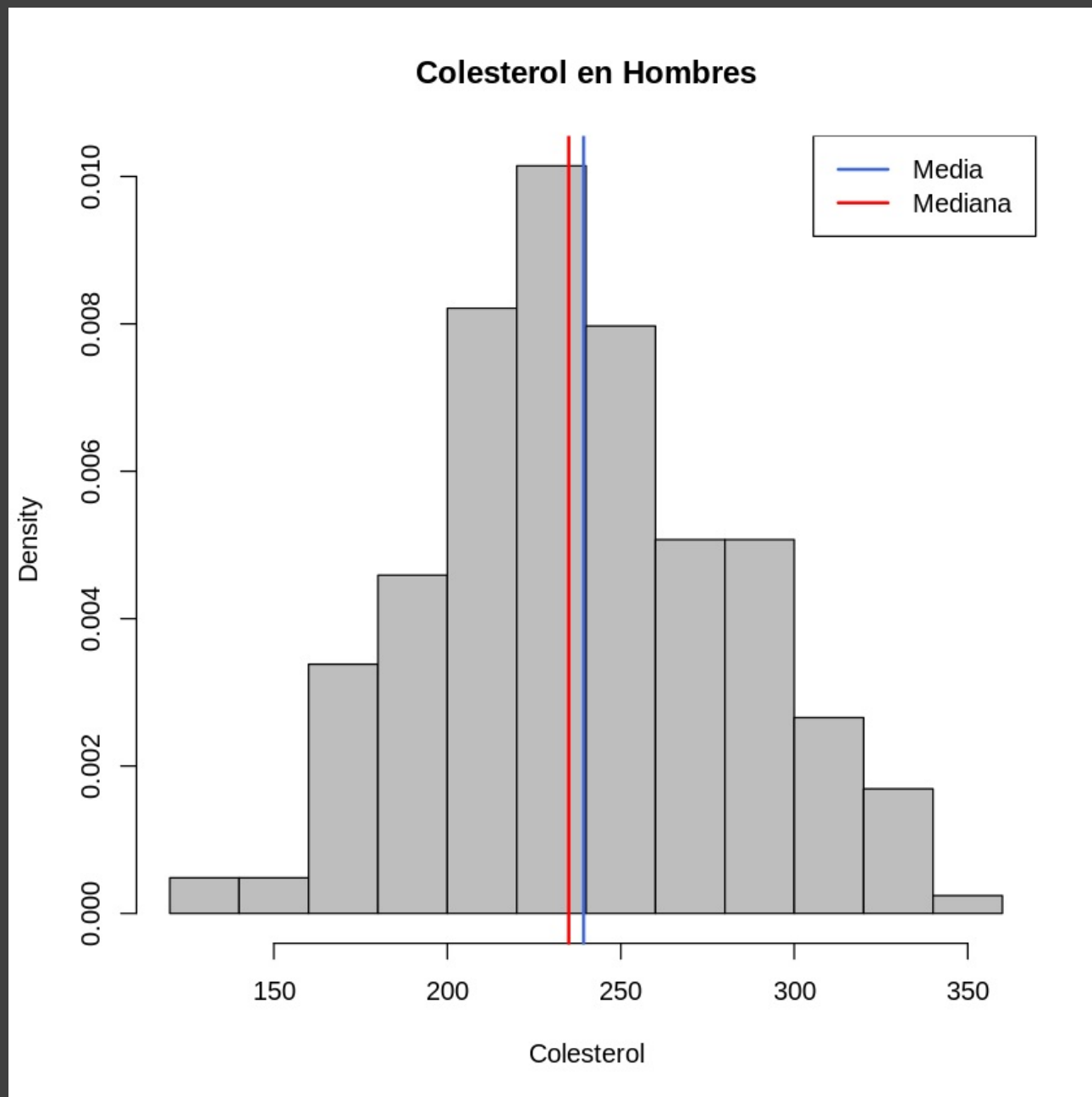
```
# --- trestbps ---
hist(men_chol$chol,
     col = "gray",
     border = "black",
     prob = TRUE,
     xlab = "Colesterol",
     main = "Colesterol en Hombres")
abline(v = mean(men_chol$chol), col = "royalblue", lwd = 2)
abline(v = median(men_chol$chol), col = "red", lwd = 2)

legend(x = "topright",
      c("Media", "Mediana"),
      col = c("royalblue", "red", "yellow"),
      lwd = c(2, 2, 2)
    )

# --- Tabla ---
table_men_chol <- data.frame(
  mean(men_chol$chol),
  median(men_chol$chol)
)

colnames(table_men_chol) <- c('Media', 'Mediana')
table_men_chol
```

Media	Mediana
239.2899	235



Comentarios: Para esta gráfica se puede apreciar que se obtuvo una mediana media aún más cercanas, donde la media tuvo un valor de 239.28, por otro lado la mediana tuvo un valor de 235. Se puede observar que la diferencia entre ambas medidas sólo fue de 4.28 unidades. Cabe mencionar que esta gráfica se parece más a la distribución normal, la cual será vista posteriormente.

In [22]:

```
hist(heart_data$chol, # histogram
     col = "gray", # column color
     border = "black",
     prob = TRUE, # show densities instead of frequencies
     xlab = "Colesterol",
     main = "Colesterol en Pacientes")

abline(v = mean(heart_data$chol), col = "royalblue", lwd = 2)
abline(v = median(heart_data$chol), col = "red", lwd = 2)

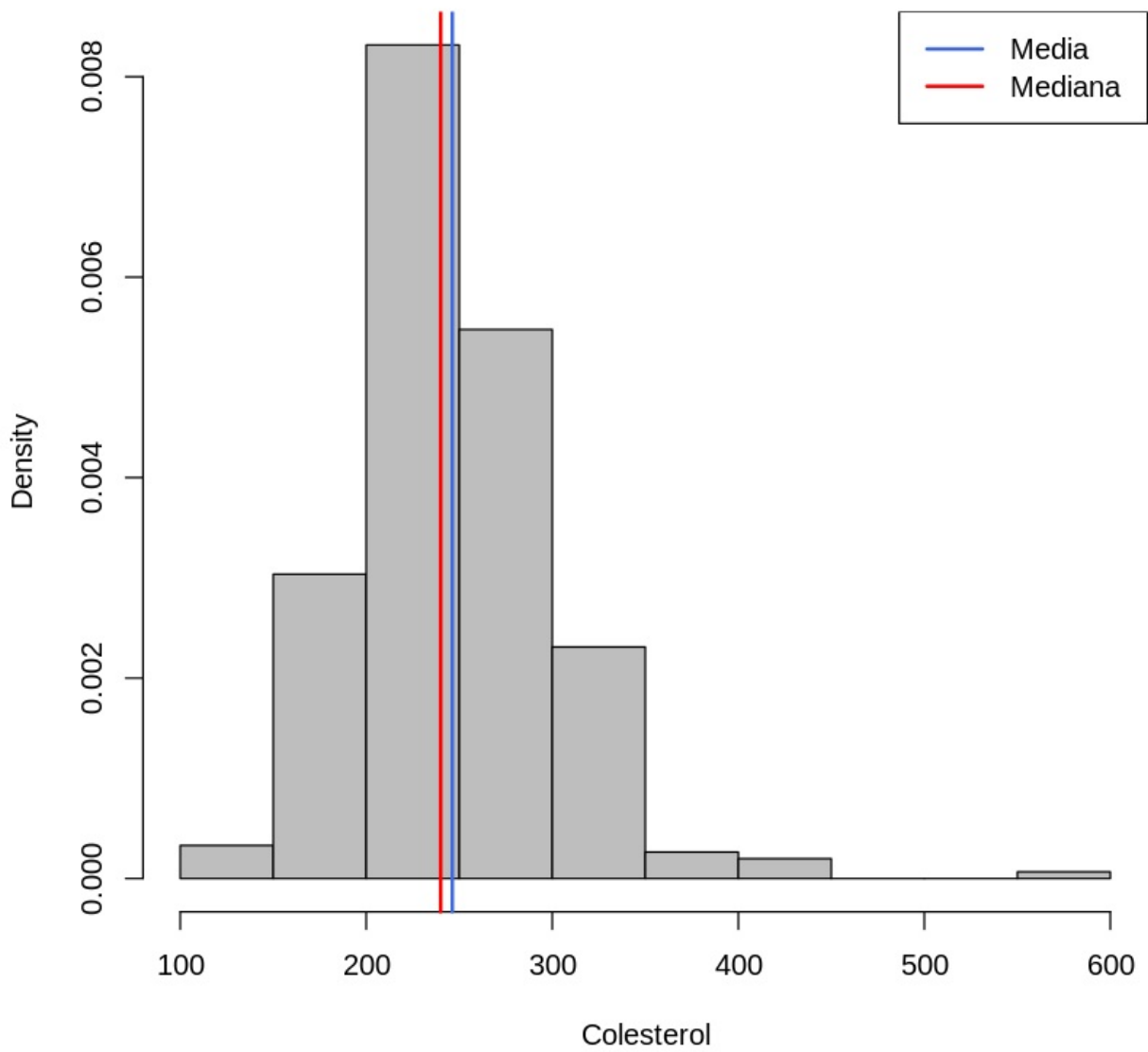
legend(x = "topright",
       c("Media", "Mediana"),
       col = c("royalblue", "red", "yellow"),
       lwd = c(2, 2, 2)
      )

# --- Tabla ---
table_men_chol <- data.frame(
  mean(heart_data$chol),
  median(heart_data$chol)
)

colnames(table_men_chol) <- c('Media', 'Mediana')
table_men_chol
```

Media	Mediana
246.264	240

Colesterol en Pacientes



Comentarios: A partir del conjunto de datos en general, se puede apreciar que de acuerdo a la media la mediana es un valor por debajo por sólo 6.26 unidades. Donde la media fue superior en los tres casos presentados.

Medidas de Dispersión

Las medidas distribución, permiten determinar que tan estirados o compactos están los datos del conjunto. Se puede ver que las medidas de dispersión estadística son la varianza y la desviación estándar son las más comunmente usadas. Estas medidas de dispersión contrastan la ubicación o tendencia central de los datos utilizados.

Varianza

El segundo momento central, alrededor de la media, recibe el nombre de varianza de la variable aleatoria. La varianza de una variable aleatoria es una medida de la dispersión de la distribución de probabilidad de esta. La varianza puede representarse mediante la siguiente fórmula:

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

Desviación Estándar

La desviación estándar es una medida estadística muestra la dispersión de un conjunto de datos en relación con su media y se calcula como la raíz cuadrada de la varianza. El cálculo realizado a través de la *varianza* permite determinar la variación entre cada punto de datos en relación con la media. Si los puntos de datos están más alejados de la media, hay una mayor desviación dentro del conjunto de datos; por lo tanto, cuanto más dispersos estén los datos, mayor será la desviación estándar. Su representación es la siguiente:

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

O bien: $s = \sqrt{s^2}$

A continuación se realizará un análisis del ritmo cardíaco máximo alcanzado, donde se analizarán las medidas de dispersión y de tendencia central.

In [23]:

```
# --- trestbps ---
hist(women_thalach$thalach, # histogram
     col = "gray", # column color
     border = "black",
     prob = TRUE, # show densities instead of frequencies
     xlab = "Ritmo Cardíaco Máximo Alcanzado",
     main = "Ritmo Cardíaco Máximo Alcanzado en Mujeres")

lines(density(women_thalach$thalach), # density plot
     lwd = 2, # thickness of line
     col = "orange")

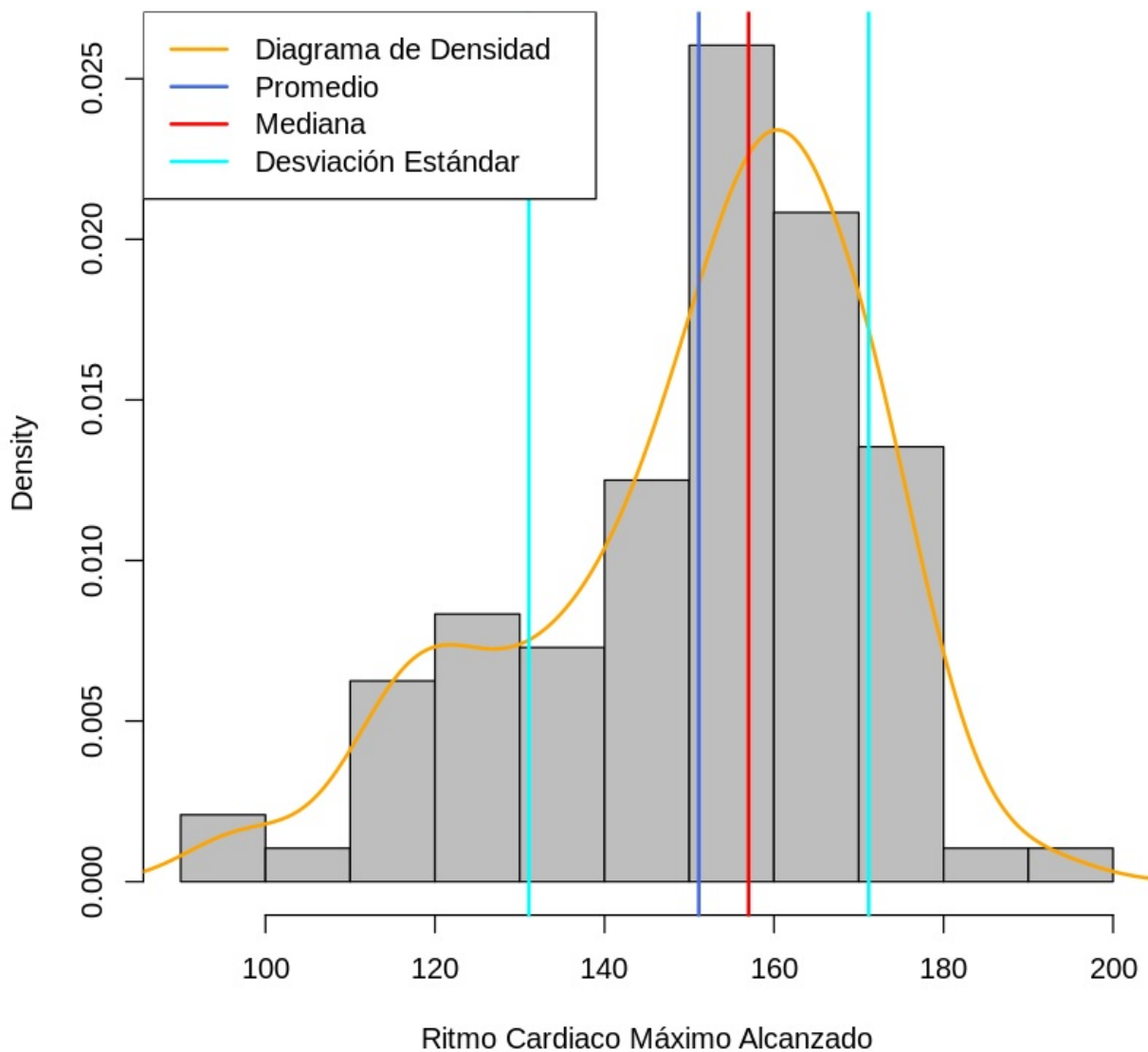
abline(v = mean(women_thalach$thalach), col = "royalblue", lwd = 2)
abline(v = median(women_thalach$thalach), col = "red", lwd = 2)
abline(v = mean(women_thalach$thalach) + sd(women_thalach$thalach), col = "cyan", lwd = 2)
abline(v = mean(women_thalach$thalach) - sd(women_thalach$thalach), col = "cyan", lwd = 2)

legend(x = "topleft", # location of legend within plot area
     c("Diagrama de Densidad", "Promedio", "Mediana", "Desviación Estándar"),
     col = c("orange", "royalblue", "red", "cyan"),
     lwd = c(2, 2, 2, 2))

# --- Tabla ---
table_md_women_thalach <- data.frame(
  mean(women_thalach$thalach),
  median(women_thalach$thalach),
  sd(women_thalach$thalach)
)

colnames(table_md_women_thalach) <- c('Media', 'Mediana', 'Desviación Estándar')
table_md_women_thalach
```

Ritmo Cardíaco Máximo Alcanzado en Mujeres



Comentarios: La desviación estándar ayuda a determinar el tamaño de la dispersión de los datos en comparación con el valor medio. A medida que aumenta la variación, se producen más variaciones en los valores de los datos, y puede haber una brecha mayor entre un valor de datos y otro. Para este caso se puede ver que hay una desviación estándar de 20.04, lo cual indica una dispersión respecto a los datos amplia.

In [24]:

```
# --- trestbps ---
hist(men_thalach$thalach, # histogram
col = "gray", # column color
border = "black",
prob = TRUE, # show densities instead of frequencies
xlab = "Ritmo Cardíaco Máximo Alcanzado",
main = "Ritmo Cardíaco Máximo Alcanzado en Hombres")

lines(density(men_thalach$thalach), # density plot
lwd = 2, # thickness of line
col = "orange")

abline(v = mean(men_thalach$thalach), col = "royalblue", lwd = 2)
abline(v = median(men_thalach$thalach), col = "red", lwd = 2)
abline(v = mean(men_thalach$thalach) + sd(men_thalach$thalach), col = "cyan", lwd = 2)
abline(v = mean(men_thalach$thalach) - sd(men_thalach$thalach), col = "cyan", lwd = 2)

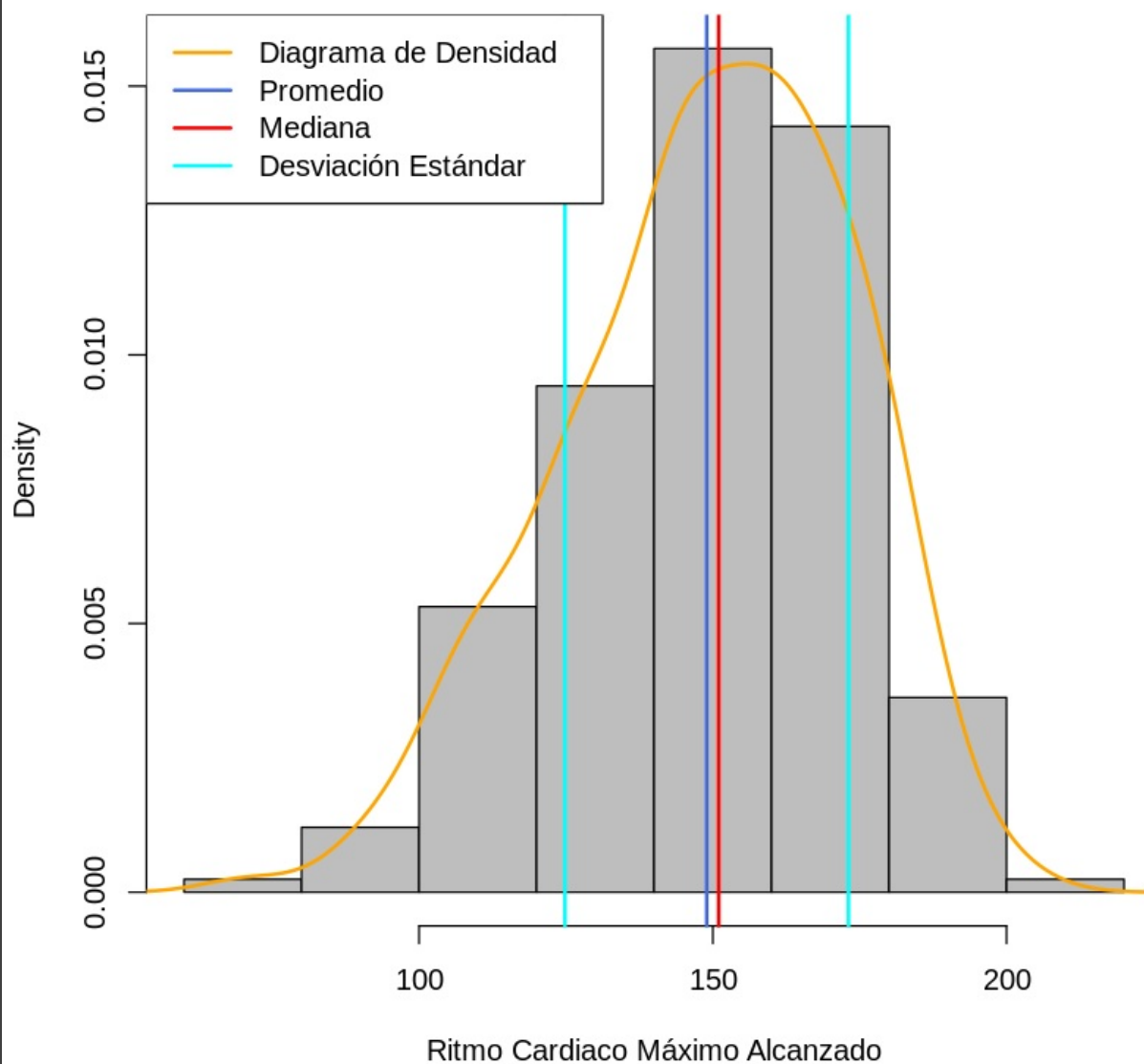
legend(x = "topleft", # location of legend within plot area
c("Diagrama de Densidad", "Promedio", "Mediana", "Desviación Estándar"),
col = c("orange", "royalblue", "red", "cyan"),
lwd = c(2, 2, 2, 2))

# --- Tabla ---
table_men_thalach_ <- data.frame(
  mean(men_thalach$thalach),
  median(men_thalach$thalach),
  sd(men_thalach$thalach)
)

colnames(table_men_thalach_) <- c('Media', 'Mediana', 'Desviación Estándar')
table_men_thalach_
```

Media	Mediana	Desviación Estándar
148.9614	151	24.13088

Ritmo Cardíaco Máximo Alcanzado en Hombres



Comentarios: Para este segundo caso, se puede observar que la desviación estándar es de 24.13, siendo aún más alta en hombres que en mujeres. Por otro lado, se puede observar que el diagrama de densidad está más centralizado donde incluso se puede ver que la diferencia entre la media y la mediana es muy baja, con un valor de 6 unidades.

In [25]:

```
# --- trestbps ---
hist(heart_data$thalach, # histogram
     col = "gray", # column color
     border = "black",
     prob = TRUE, # show densities instead of frequencies
     xlab = "Ritmo Cardíaco Máximo Alcanzado",
     main = "Ritmo Cardíaco Máximo Alcanzado en Pacientes")

lines(density(heart_data$thalach), lwd = 2, col = "orange")

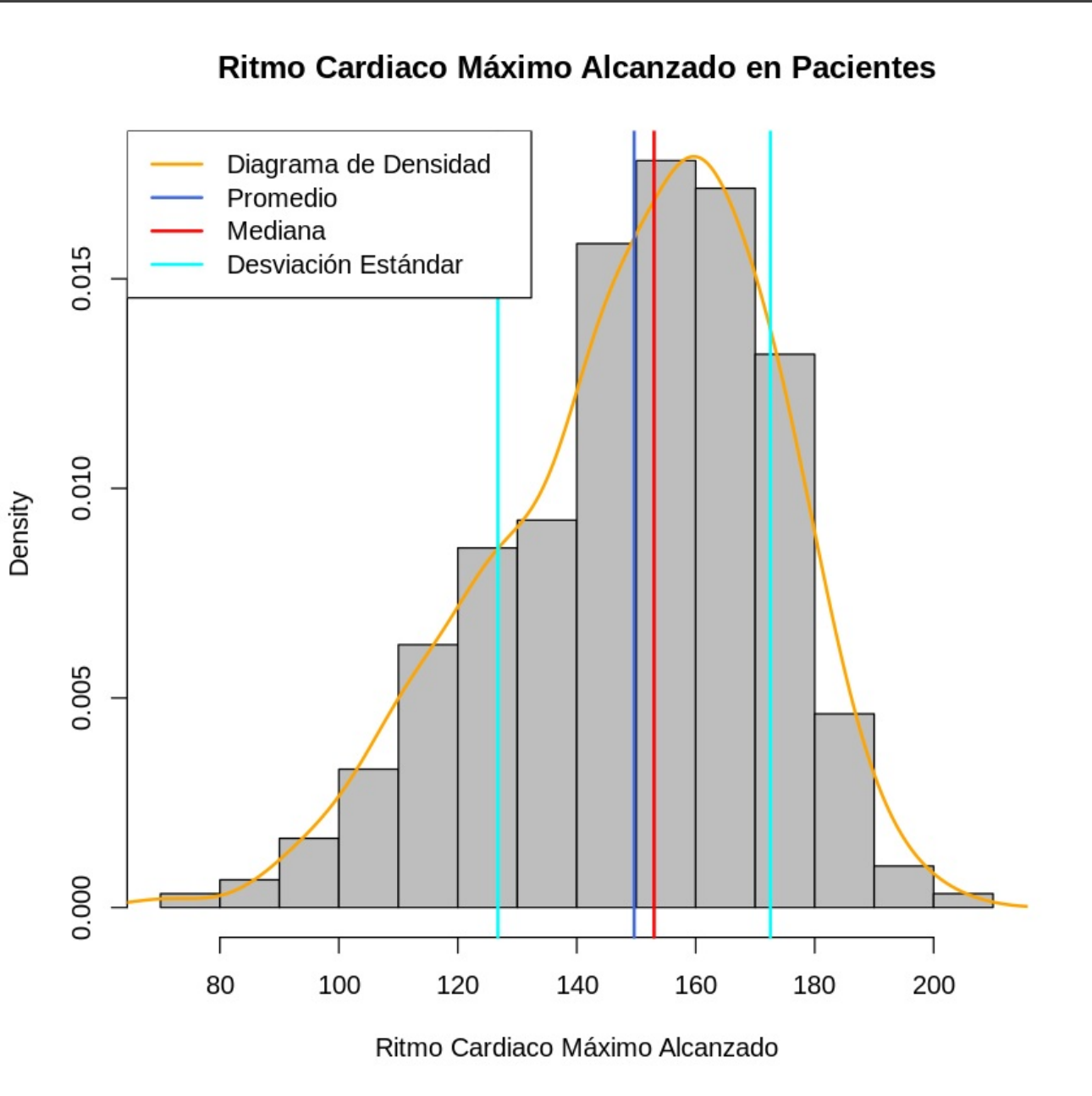
abline(v = mean(heart_data$thalach), col = "royalblue", lwd = 2)
abline(v = median(heart_data$thalach), col = "red", lwd = 2)
abline(v = mean(heart_data$thalach) + sd(heart_data$thalach), col = "cyan", lwd = 2)
abline(v = mean(heart_data$thalach) - sd(heart_data$thalach), col = "cyan", lwd = 2)

legend(x = "topleft",
       c("Diagrama de Densidad", "Promedio", "Mediana", "Desviación Estándar"),
       col = c("orange", "royalblue", "red", "cyan"),
       lwd = c(2, 2, 2, 2))
)

# --- Tabla ---
table_men_chol <- data.frame(
  mean(heart_data$thalach),
  median(heart_data$thalach),
  sd(heart_data$thalach)
)

colnames(table_men_chol) <- c('Media', 'Mediana', 'Desviación Estándar')
table_men_chol
```

Media	Mediana	Desviación Estándar
149.6469	153	22.90516



Comentarios: Por último, se puede observar que para los pacientes en general hay una desviación estándar de 22.9 unidades, con una mediana de 153 y una media de 149.64.

Curtosis y Sesgo

Curtosis

Las curtosis permite examinar la proporción de la varianza que se explica por la combinación de datos extremos respecto a la media en contraposición con datos poco alejados de la misma. Una mayor curtosis implica una mayor concentración de datos cerca de la media de la distribución elevando la frecuencia de datos muy alejados de la misma.

$$\alpha = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n^4}$$

A continuación se analizará la curtosis de los pacientes con el *ritmo cardíaco máximo* y sus subconjuntos de hombres y mujeres.

In [26]:

```
#install.packages('moments')
library(moments)
kurt_pac_thalach <- kurtosis(heart_data$thalach)
kurt_men_thalach <- kurtosis(men_thalach$thalach)
kurt_wom_thalach <- kurtosis(women_thalach$thalach)

data_kurt <- cbind(kurt_pac_thalach, kurt_men_thalach, kurt_wom_thalach)
df.kurt_val <- data.frame(data_kurt)
colnames(df.kurt_val) <- sex <- c('General', 'Hombre', 'Mujer')
rownames(df.kurt_val) <- c('Curtosis')

df.kurt_val
```

	General	Hombre	Mujer
Curtosis	2.919311	2.810824	2.995223

Comentarios: Para este caso las tres curtosis son menores a tres, lo que representa una curtosis mesocurtica o platicurtica. Lo cual indica que tiene una distribución con menores registros de valores en los extremos, esto significa que la probabilidad de eventos extremos es mayor que lo que implica la curva normal. Es decir, la distribución mesocurtica tiene colas más claras, y su probabilidad de eventos extremos es menor que lo que implica la curva normal.

Sesgo

Esta forma de distribución permite identificar la manera como los datos tienden a reunirse de acuerdo con la frecuencia con que se hallen dentro de la distribución. Las medidas de asimetría son indicadores que permiten establecer el grado de simetría que presentan una distribución de probabilidad de una variable aleatoria sin tener que hacer su representación gráfica.

$$s_A = \frac{3(\bar{x} - \tilde{x})}{s}$$

In [27]:

```
skew_pac_thalach <- skewness(heart_data$thalach)
skew_men_thalach <- skewness(men_thalach$thalach)
skew_wom_thalach <- skewness(women_thalach$thalach)

data_skew <- cbind(skew_pac_thalach, skew_men_thalach, skew_wom_thalach)
df.skew_val <- data.frame(data_skew)
colnames(df.skew_val) <- sex <- c('General', 'Hombre', 'Mujer')
rownames(df.skew_val) <- c('Sesgo')

df.skew_val
```

	General	Hombre	Mujer
Sesgo	-0.5347455	-0.44657	-0.7496368

Comentarios: El análisis del sesgo para los pacientes en general indica que los datos están sesgados moderadamente, ya que su valor es menor que -0.5. Por otro lado, se puede apreciar que el sesgo para los hombres indica una distribución parcial de los datos ya que su valor es mayor a -0.5 y menor que 0.5. Por último, para el caso de las mujeres se puede apreciar que los datos están sesgados moderadamente ya que su valor es menor a -0.5, lo cual indica una distribución mayor de los datos.

Correlación

La correlación mide la forma en la que dos medidas se relacionan, para esto se puede ver que hay tanto correlación positiva como negativa o también llamada inversa. La correlación positiva describe la relación entre dos variables que cambian juntas, mientras que una correlación inversa describe la relación entre dos variables que cambian en direcciones opuestas. La correlación inversa se describe a veces como correlación negativa, que describe el mismo tipo de relación entre las variables.

In [29]:

```
cormat <- round(cor(heart_data),2)
library(reshape2)
melted_cormat <- melt(cormat)
#head(melted_cormat)

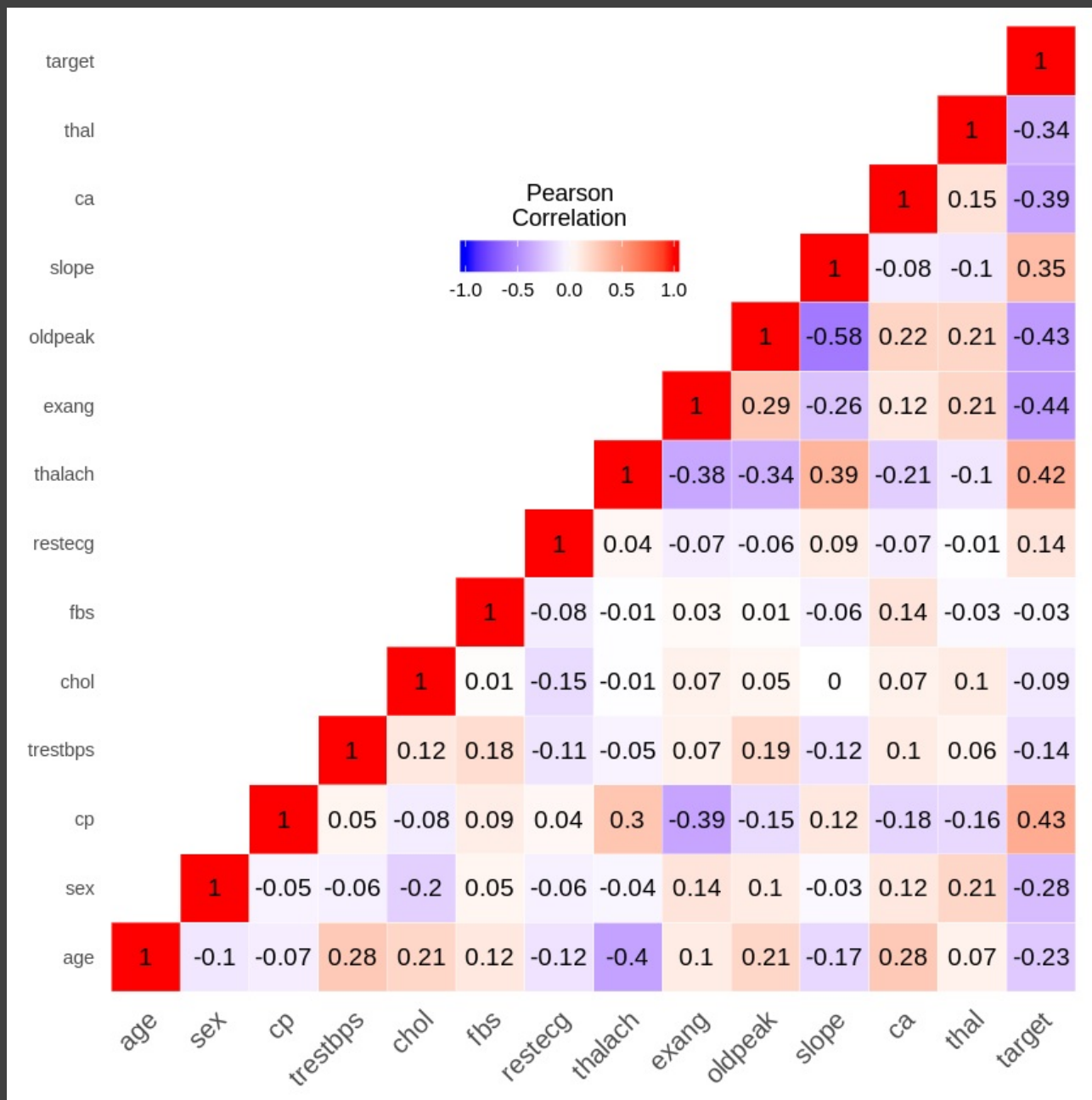
# Get lower triangle of the correlation matrix
get_lower_tri<-function(cormat){
  cormat[upper.tri(cormat)] <- NA
  return(cormat)
}
# Get upper triangle of the correlation matrix
get_upper_tri <- function(cormat){
  cormat[lower.tri(cormat)]<- NA
  return(cormat)
}

upper_tri <- get_upper_tri(cormat)

library(reshape2)
melted_cormat <- melt(upper_tri, na.rm = TRUE)

# Create a ggheatmap
ggheatmap <- ggplot(melted_cormat, aes(Var2, Var1, fill = value))+
  geom_tile(color = "white")+
  scale_fill_gradient2(low = "blue", high = "red", mid = "white",
    midpoint = 0, limit = c(-1,1), space = "Lab",
    name="Pearson\nCorrelation") +
  theme_minimal()+ # minimal theme
  theme(axis.text.x = element_text(angle = 45, vjust = 1,
    size = 12, hjust = 1))+
  coord_fixed()

ggheatmap +
  geom_text(aes(Var2, Var1, label = value), color = "black", size = 4) +
  theme(
    axis.title.x = element_blank(),
    axis.title.y = element_blank(),
    panel.grid.major = element_blank(),
    panel.border = element_blank(),
    panel.background = element_blank(),
    axis.ticks = element_blank(),
    legend.justification = c(1, 0),
    legend.position = c(0.6, 0.7),
    legend.direction = "horizontal")+
  guides(fill = guide_colorbar(barwidth = 7, barheight = 1,
    title.position = "top", title.hjust = 0.5))
```



Comentarios: Para este caso se puede observar tanto correlaciones negativas como positivas. A continuación analizaremos la relación entre algunas de las variables:

- age-sex: En esta correlación se puede apreciar que hay una correlación negativa por lo que indica que ambas variables no cambian juntas.
- thalach-cp: Estas variables indican una correlación positiva, sin embargo dado su valor se puede ver que no es una correlación fuerte.

Cabe destacar que la correlación no necesariamente implica una causa, las variables A y B pueden subir y bajar juntas, o A puede aumentar a medida que B cae, pero no siempre es cierto que el aumento de un factor influye directamente en el aumento o la caída del otro. Ambos pueden ser causados por un tercer factor subyacente, como una variable C, o la relación aparente entre las variables podría ser una coincidencia.

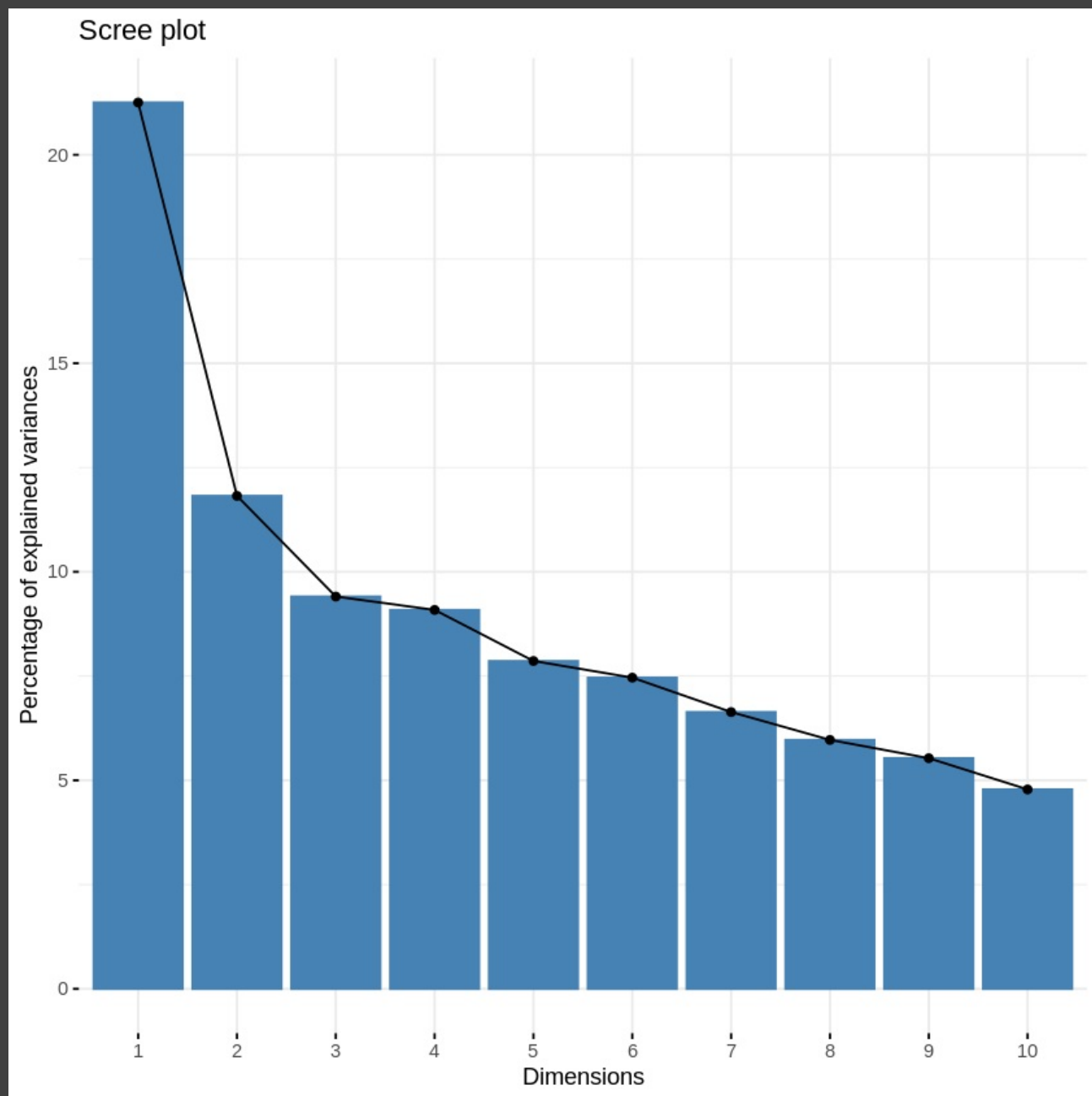
Principal Component Analysis

Graph of individuals. Individuals with a similar profile are grouped together.

In [31]:

```
library(factoextra)
res.pca <- prcomp(heart_data[c(1:length(heart_data) - 1)], scale = TRUE)
fviz_eig(res.pca)
```

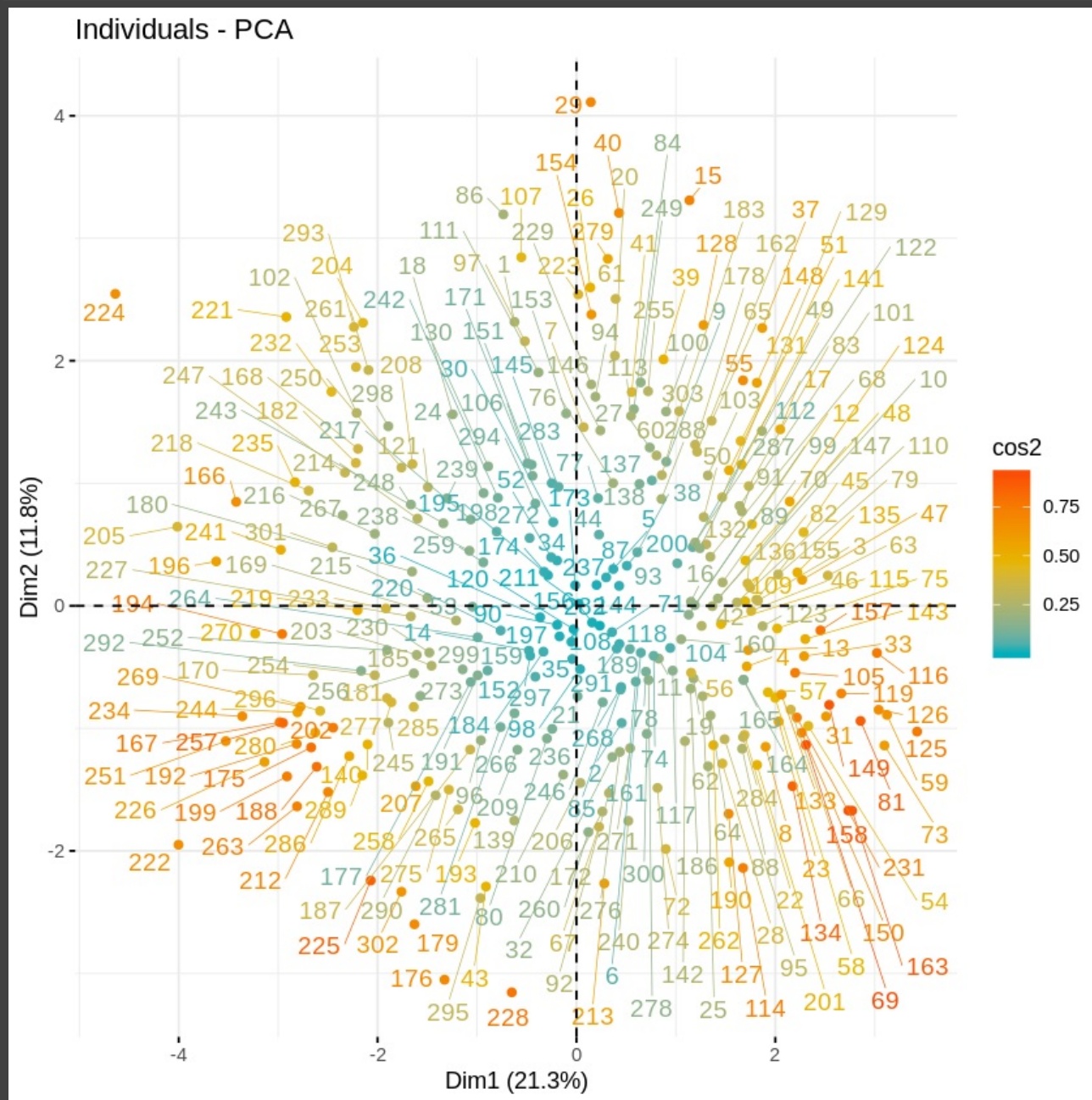
Welcome! Related Books: `Practical Guide To Cluster Analysis in R` at <https://goo.gl/13EFCZ>



Graph of variables. Positive correlated variables point to the same side of the plot. Negative correlated variables point to opposite sides of the graph.

In [32]:

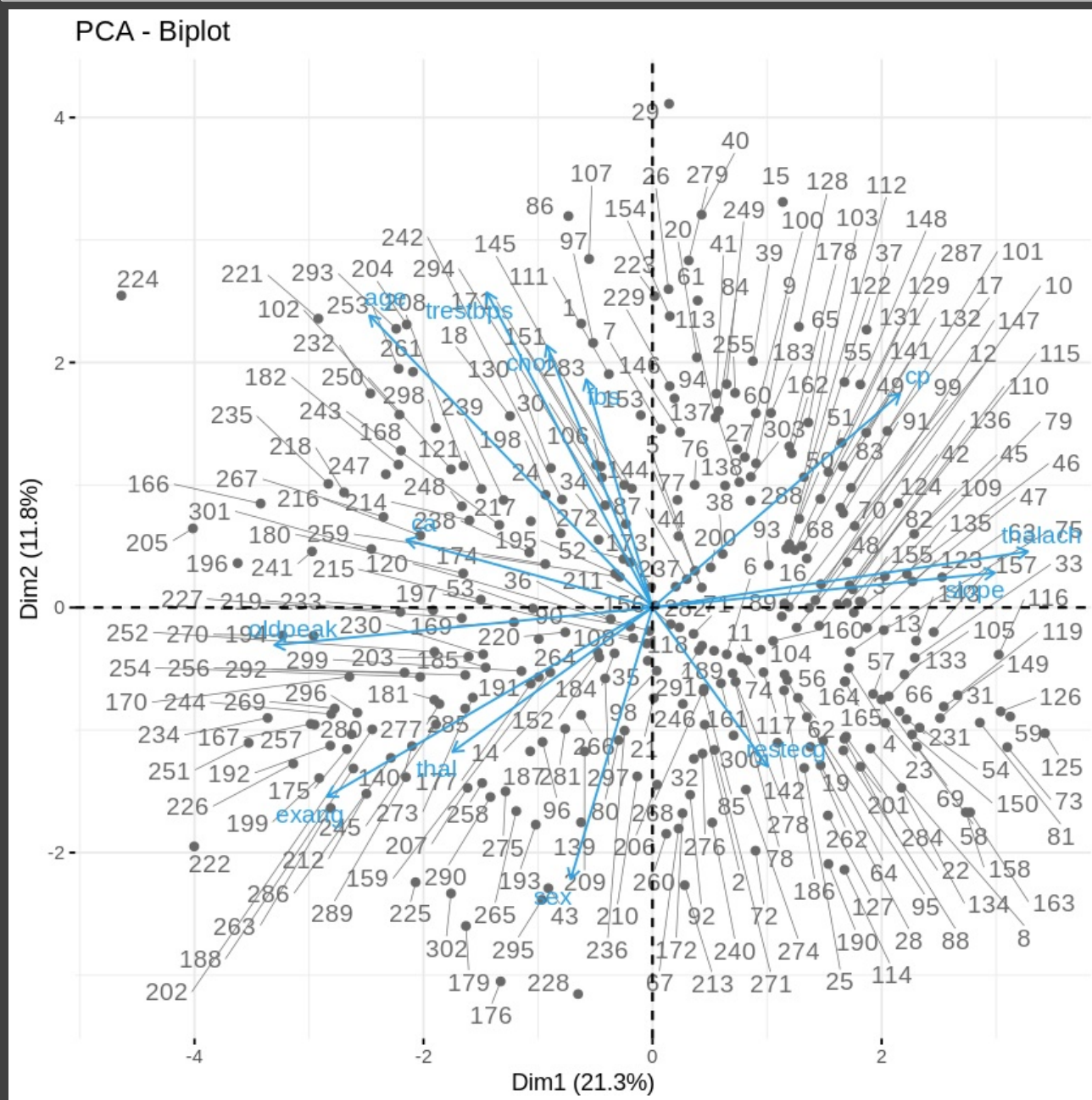
```
fviz_pca_ind(res.pca,
  col.ind = "cos2", # Color by the quality of representation
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE      # Avoid text overlapping
)
```



Biplot of individuals and variables

In [33]:

```
fviz_pca_biplot(res.pca, repel = TRUE,  
  col.var = "#2E9FDF", # Variables color  
  col.ind = "#696969" # Individuals color  
)
```



Access to the PCA results

In [34]:

```
# Eigenvalues  
eig.val <- get_eigenvalue(res.pca)  
eig.val
```


	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	2.7630269	21.254053	21.25405
Dim.2	1.5366920	11.820708	33.07476
Dim.3	1.2228343	9.406418	42.48118
Dim.4	1.1811455	9.085735	51.56691
Dim.5	1.0219665	7.861281	59.42819
Dim.6	0.9700159	7.461661	66.88985
Dim.7	0.8627699	6.636692	73.52655
Dim.8	0.7759454	5.968811	79.49536
Dim.9	0.7189255	5.530196	85.02555
Dim.10	0.6215702	4.781309	89.80686
Dim.11	0.5301048	4.077729	93.88459
Dim.12	0.4231424	3.254941	97.13953
Dim.13	0.3718607	2.860467	100.00000

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8	
age	-0.5222786	0.50347568	-0.10403165	0.02245535	-0.31050834	0.12635808	-0.20781270	0.23121029	0.321
sex	-0.1509939	-0.46832338	0.61356239	0.27747086	0.05125827	-0.05413838	-0.15094513	0.15502736	0.168
cp	0.4564623	0.36850110	0.39474876	-0.31289166	0.16496196	0.19048947	-0.20006550	-0.04223802	0.297
trestbps	-0.3057187	0.54319112	0.22542031	-0.02456292	0.19019324	0.17674887	0.30908860	0.52441625	-0.297
chol	-0.1951050	0.45186391	-0.45098036	0.37321957	0.32356298	0.10314751	0.04581978	-0.32802166	0.130
fbs	-0.1224070	0.39350103	0.53271279	0.07456062	-0.23599187	-0.24584293	0.47447559	-0.38129934	0.150
restecg	0.2123139	-0.27381256	-0.09862886	-0.28919402	-0.39796753	0.65674039	0.36865838	-0.08794753	0.032
thalach	0.6923180	0.09653795	0.17500161	0.20010843	0.32681574	0.11915684	0.09425327	-0.12637183	-0.315
exang	-0.6005116	-0.32616984	-0.13972690	0.12504368	0.03491293	-0.22721416	0.41790872	0.09919259	0.049
oldpeak	-0.6975388	-0.06477703	0.12201921	-0.35462015	0.25331650	0.16751058	-0.10485631	-0.16941320	-0.198
slope	0.6312709	0.05996623	-0.08162973	0.53780439	-0.24951895	0.06310151	0.05112224	0.23061996	0.024
ca	-0.4542253	0.11670806	0.20299435	0.35648984	-0.44012091	0.17935655	-0.31358672	-0.22874481	-0.411
thal	-0.3690558	-0.24881981	0.13823961	0.42297521	0.33557659	0.50116967	0.05124006	-0.03025708	0.240

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8	Di
age	9.8723226	16.4956780	0.8850409	0.04269101	9.4343041	1.6459901	5.0055194	6.8894280	14.36411
sex	0.8251512	14.2726576	30.7857576	6.51825495	0.2570936	0.3021563	2.6408467	3.0973161	3.95712
cp	7.5409275	8.8367129	12.7430660	8.28866463	2.6627532	3.7407879	4.6392673	0.2299196	12.35046
trestbps	3.3826637	19.2007630	4.1554538	0.05108065	3.5395941	3.2205827	11.0731446	35.4422370	12.27744
chol	1.3776899	13.2870472	16.6321213	11.79303041	10.2442693	1.0968284	0.2433386	13.8667247	2.37083
fbs	0.5422848	10.0763888	23.2069802	0.47066905	5.4495094	6.2306968	26.0935245	18.7370387	3.13305
restecg	1.6314420	4.8788775	0.7955005	7.08068391	15.4973918	44.4640093	15.7526353	0.9968185	0.14672
thalach	17.3470677	0.6064700	2.5044738	3.39021587	10.4512745	1.4637237	1.0296694	2.0581140	13.84171
exang	13.0514172	6.9231028	1.5965864	1.32379306	0.1192713	5.3222093	20.2426733	1.2680236	0.34222
oldpeak	17.6096880	0.2730582	1.2175556	10.64690595	6.2789969	2.8927151	1.2743659	3.6988213	5.45703
slope	14.4226939	0.2340058	0.5449154	24.48754714	6.0921472	0.4104882	0.3029178	6.8542924	0.08125
ca	7.4671967	0.8863697	3.3697703	10.75947060	18.9542817	3.3163140	11.3977816	6.7432821	23.60098
thal	4.9294546	4.0288685	1.5627783	15.14699276	11.0191129	25.8934981	0.3043156	0.1179839	8.07702

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8
age	0.27277493	0.253487765	0.010822583	0.0005042429	0.096415431	0.015966365	0.043186116	0.0534581985
sex	0.02279915	0.219326789	0.376458803	0.0769900759	0.002627411	0.002930964	0.022784431	0.0240334808
cp	0.20835786	0.135793060	0.155826582	0.0979011903	0.027212447	0.036286236	0.040026203	0.0017840506
trestbps	0.09346391	0.295056589	0.050814314	0.0006033369	0.036173467	0.031240163	0.095535761	0.2750124000
chol	0.03806594	0.204180992	0.203383285	0.1392928495	0.104693004	0.010639409	0.002099452	0.1075982096
fb	0.01498348	0.154843061	0.283782914	0.0055592863	0.055692163	0.060438748	0.225127082	0.1453891856
restecg	0.04507718	0.074973320	0.009727653	0.0836331803	0.158378159	0.431307945	0.135909000	0.0077347673
thalach	0.47930415	0.009319576	0.030625564	0.0400433827	0.106808529	0.014198352	0.008883678	0.0159698406
exang	0.36061417	0.106386767	0.019523606	0.0156359223	0.001218913	0.051626275	0.174647698	0.0098391707
oldpeak	0.48656042	0.004196064	0.014888687	0.1257554518	0.064169247	0.028059796	0.010994845	0.0287008331
slope	0.39850291	0.003595949	0.006663413	0.2892335642	0.062259706	0.003981800	0.002613483	0.0531855648
ca	0.20632066	0.013620772	0.041206707	0.1270850042	0.193706417	0.032168772	0.098336632	0.0523241859
thal	0.13620216	0.061911300	0.019110189	0.1789080252	0.112611646	0.251171040	0.002625544	0.0009154908

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8	D
1	-0.62307999	2.31743663	2.47058555	-2.67181961	0.37463354	-1.71073661	-0.12015287	-0.73534031	0.6378
2	0.45523490	-0.95576989	1.13771273	-2.42283034	2.27001191	0.78655946	-0.08137873	-1.53284832	-1.2397
3	1.82578459	0.04281395	-0.45148191	-0.40574371	0.86636898	-0.76628302	-0.10467873	0.06341831	-1.2910
4	1.71317202	-0.49451926	0.03058043	0.11197637	-0.23520879	0.50183446	-0.14251850	0.17716929	0.1527
5	0.37074312	0.30065881	-2.83637722	0.80770555	0.01137015	0.08535335	1.41325773	-0.60613163	0.1860
6	0.64779584	-0.38225001	-0.07666423	-1.49151361	-1.13566901	-0.70781257	0.10571962	1.13295226	-0.6190
7	0.07253341	1.45780795	-1.21976909	-0.71586330	1.13290986	-0.48918550	-0.36256224	-0.09952172	-0.3150
8	1.90277805	-1.15009215	0.03603569	1.08384433	0.63629204	1.07436743	0.32451757	-0.25588930	0.4384
9	0.90423692	1.17607951	2.73099630	0.29570094	-0.01626214	1.09192367	2.18120529	0.92465258	0.4588
10	1.42216820	0.05994488	1.29824820	-0.91088660	-0.09281419	0.96900010	0.03313757	1.55905019	-0.3623
11	0.82787972	-0.43039967	-0.17208236	0.12338223	-0.26033416	0.46040864	0.37914623	0.86965918	-0.4500
12	1.76545000	0.66498196	-1.24841350	-0.63241903	-0.24717540	0.58142596	0.30699001	-0.07524513	0.4027
13	1.72753584	-0.36152440	-0.08410139	0.28186822	0.15209741	0.49962349	0.23664448	0.17575168	-0.0961
14	-0.47778915	-0.36549625	0.60053034	-1.18694021	0.50741068	-1.05936860	-1.10475251	0.20762335	1.6495
15	1.13716882	3.30957543	0.92039458	-0.20270778	0.42721083	-0.59923499	0.94247599	-0.51321449	0.6743
16	1.15115297	0.03382832	-0.56061795	-2.02975472	0.20294816	0.59566475	-0.12657273	-0.73031729	-0.1168
17	2.04950593	1.44066053	-1.77068091	0.15536136	0.13134457	0.89671076	0.09841362	-0.76624625	0.7171
18	-1.24454687	1.56304651	-0.17602508	-3.70235381	0.18033173	1.14031737	-0.53501382	0.33877693	0.7799
19	1.17704082	-0.59230852	0.09965677	0.14259104	0.48835645	0.52555719	0.86703638	0.71513956	-1.3241
20	0.38715853	2.04171929	-0.12657064	-0.82666498	-1.25111658	1.74898825	-1.02624091	0.23523667	-0.1540
21	0.01093038	-0.73831219	0.02170127	0.14544671	0.29179428	1.09963575	0.22962044	0.50908974	0.3861
22	1.69008792	-1.05304011	0.34054164	0.13831149	0.41878583	0.07161138	1.13228655	0.44459406	-0.0274
23	2.03348089	-0.94325625	0.06483101	0.49187919	0.06029596	0.18391936	0.85879842	0.70228429	-1.0403
24	-0.93401447	0.92008769	1.55729534	-0.89919654	-0.59797957	-0.40208011	2.20286632	0.06654351	1.3242
25	1.32445078	-1.30938351	1.40392102	-0.04132613	1.35335214	1.20842779	1.16794166	0.65077766	0.1245
26	0.13941429	2.59825876	-1.15935666	0.60930785	-1.17828109	1.58740410	-0.03382630	0.77523772	-0.8610
27	0.73747489	1.29219856	2.16630608	-0.55886184	-0.78405326	0.29559339	1.38451917	0.19197484	0.6248

28	1.46742523	-1.28692655	0.39299029	-0.95554959	-1.21139712	0.07321957	-0.70005206	0.46262110	1.2367
29	0.14447958	4.11119074	-0.54264128	1.32967133	0.21309463	-0.39694476	0.58368513	-1.75612346	0.8701
30	-0.23306081	0.68273388	2.45065079	-1.67394624	0.50448699	-1.58511352	0.22192901	-1.11483719	0.7257
:	:	:	:	:	:	:	:	:	
274	0.8152840	-1.4859302	-0.38759183	1.3842905	-1.033020723	0.9244189	-0.63391258	-0.43939783	0.77657
275	-1.2836869	-1.5000630	-0.84077136	0.3718852	0.037864951	-1.8088468	-0.36650358	-0.56269256	0.21695
276	0.4362096	-1.1919897	0.48619635	1.3569622	-0.759911211	1.4243233	-0.39718889	-0.08428858	-0.88969
277	-1.8612323	-0.7865629	0.22096055	-0.5279537	-0.547590643	1.2677562	-0.27633247	0.81688008	0.20986
278	0.7253126	-0.6051624	-0.22897715	0.7534695	-0.159708122	1.1698201	-0.09266754	0.42069861	1.35454
279	0.3154472	2.8320387	-0.02247219	1.4539467	-1.025325647	-0.8876986	0.52114775	-1.43376828	-0.20656
280	-2.6272350	-1.0342010	0.49277046	-1.0289439	-0.148421748	-1.1269798	-0.50563766	1.10097955	-0.71884
281	-0.7630684	-0.9892299	-1.23041479	-1.0389562	0.139301412	-1.1195727	1.23608344	-0.16818567	-0.58323
282	0.2349479	-0.1621491	0.73282851	-1.7103604	-1.902811092	-2.7697663	2.29141192	-0.57403598	-0.64566
283	-0.1803919	0.9709500	1.83337748	-1.9595981	-1.766950877	-0.7284368	0.26507689	-1.24383774	0.38951
284	1.6755545	-1.0716031	0.47353328	1.1117505	0.822637829	1.1193309	1.23628413	0.99848808	-0.95703
285	-1.6365841	-0.8245045	0.20605508	1.2610576	0.084430167	-0.2689683	-0.02637910	1.44322804	0.00430
286	-2.2849405	-1.2270638	-0.46105694	0.8111165	0.184495549	0.9617126	0.70739429	-0.47682442	-0.44869
287	1.2418140	0.4701786	1.35427007	-0.1562651	-1.145174635	1.2692701	-1.12981275	0.46842304	-0.10608
288	0.8555346	0.8720629	0.62324887	0.9830020	0.023627330	-0.4253135	-0.59499325	1.54506513	-0.65627
289	-2.1535021	-1.3837784	-1.01909608	0.5607041	0.648207792	0.9778051	0.20541224	-1.43858925	0.54105
290	-1.4158223	-1.5462924	-1.31136931	-1.2057230	-1.150560991	2.0419095	1.61559180	-0.27184089	-0.22989
291	0.4053892	-0.3490172	0.40820887	1.1921613	-0.762009490	1.4070824	0.20529153	1.48463681	-0.21040
292	-2.1656948	-0.5305697	-0.56706155	-2.1560895	-1.545266891	1.5368666	-1.04164653	-2.61496647	-1.75699
293	-2.2380574	2.2752467	0.63775798	-0.9625993	-0.947682171	-1.9647919	1.76495377	-0.11310801	-2.35741
294	-0.4134739	0.8346719	1.12342761	-0.2286554	1.042078016	0.4347537	-0.91589480	1.59156415	1.04151
295	-0.9660627	-2.3852398	0.05778334	-2.9124522	-0.117036390	-1.4071777	0.64191280	-0.31406212	-1.16990
296	-2.5775912	-0.8577670	0.76324603	0.9119308	0.005444774	0.2362637	-0.60242087	1.00535370	-0.96668
297	-0.4136300	-0.5785253	-1.74047389	-1.0754744	-1.228178569	-0.4406689	1.04133269	0.67466405	0.40308
298	-1.8924235	1.4654424	1.84483212	-0.9186031	-2.492015276	-2.2974258	0.20035262	0.85102742	-0.76727
299	-1.1447781	-0.5186715	-1.70519914	-0.3432780	-0.180074366	0.5189217	1.54765435	0.72045638	0.59884
300	0.7074226	-1.0440248	0.54391023	-0.9221354	0.898440762	1.2182480	-0.69392414	-1.03337782	1.74228
301	-2.4549443	0.4774720	2.20456939	-0.2589481	-1.332485524	1.2105776	0.54289419	-0.70893817	-0.25169
302	-1.7598441	-2.3329569	0.45400154	-0.5359201	-1.180654313	0.3586843	0.44169649	1.17948821	0.35239
303	0.8586364	1.0667509	-0.68893461	-0.2300728	0.223745865	-0.5960070	-0.74284931	-0.16349972	-0.80045

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.
1	4.637243e-02	1.1534168060	1.6473639400	1.9946570291	4.532460e-02	0.995737428	0.0055224354	0.22998683
2	2.475388e-02	0.1961901631	0.3493453725	1.6402119569	1.664091e+00	0.210494938	0.0025332861	0.99936603
3	3.981726e-01	0.0003936780	0.0550136515	0.0459999703	2.423966e-01	0.199782269	0.0041915945	0.00171062
4	3.505696e-01	0.0525214806	0.0002523929	0.0035035375	1.786602e-02	0.085683951	0.0077697109	0.01335067
5	1.641793e-02	0.0194141509	2.1712895261	0.1822888493	4.174970e-05	0.002478676	0.7640206710	0.15626442
6	5.012433e-02	0.0313809110	0.0015862633	0.6215962821	4.165089e-01	0.170457094	0.0042753686	0.54594568
7	6.284178e-04	0.4564268315	0.4015549705	0.1431904865	4.144875e-01	0.081418986	0.0502837322	0.00421271
8	4.324626e-01	0.2840768342	0.0003504737	0.3282371910	1.307476e-01	0.392721068	0.0402845754	0.02785033
9	9.766450e-02	0.2970598148	2.0129453459	0.0244320208	8.540359e-05	0.405660845	1.8199332400	0.36364983
	2.415873e-							

10	01	0.0007717467	0.4548887695	0.2318369159	2.781951e-03	0.319467129	0.0004200525	1.03382353
11	8.186660e-02	0.0397845451	0.0079921285	0.0042536250	2.188684e-02	0.072121619	0.0549890030	0.32168024
12	3.722915e-01	0.0949708364	0.4206362228	0.1117541925	1.973019e-02	0.115018412	0.0360504635	0.00240814
13	3.564728e-01	0.0280702163	0.0019089568	0.0221996529	7.470752e-03	0.084930608	0.0214217587	0.01313787
14	2.726747e-02	0.0286903865	0.0973329019	0.3936512690	8.314572e-02	0.381832357	0.4668661313	0.01833490
15	1.544622e-01	2.3524213254	0.2286324477	0.0114813985	5.893933e-02	0.122172333	0.3397840320	0.11202734
16	1.582845e-01	0.0002457714	0.0848249778	1.1511743010	1.330120e-02	0.120720863	0.0061283375	0.22685553
17	5.017306e-01	0.4457525678	0.8461948342	0.0067443482	5.571151e-03	0.273579337	0.0037048685	0.24972553
18	1.850098e-01	0.5247039626	0.0083625521	3.8300972954	1.050183e-02	0.442415172	0.1094945186	0.04881506
19	1.654837e-01	0.0753470937	0.0026804230	0.0056811786	7.701841e-02	0.093976333	0.2875656511	0.21752432
20	1.790399e-02	0.8952876699	0.0043237019	0.1909470903	5.054943e-01	1.040764127	0.4028663757	0.02353619
21	1.427061e-05	0.1170713472	0.0001271042	0.0059110107	2.749630e-02	0.411411315	0.0201689587	0.11023383
22	3.411857e-01	0.2381553619	0.0312989815	0.0053452805	5.663757e-02	0.001744788	0.4904277919	0.08407241
23	4.939153e-01	0.1910864627	0.0011343717	0.0676037633	1.174079e-03	0.011508890	0.2821271237	0.20977422
24	1.042028e-01	0.1818146937	0.6545332734	0.2259244431	1.154764e-01	0.055005195	1.8562593713	0.00188337
25	2.095289e-01	0.3682173695	0.5319552667	0.0004772034	5.914831e-01	0.496843881	0.5218007193	0.18013225
26	2.321597e-03	1.4498909510	0.3627637480	0.1037355251	4.483514e-01	0.857340875	0.0004376949	0.25562069
27	6.496310e-02	0.3586154927	1.2665705880	0.0872695779	1.985235e-01	0.029728156	0.7332639285	0.01567526
28	2.572079e-01	0.3556952455	0.0416824659	0.2551293548	4.739077e-01	0.001824034	0.1874662288	0.09102844
29	2.493361e-03	3.6299930444	0.0794722324	0.4940177754	1.466445e-02	0.053609121	0.1303225690	1.31170541
30	6.487999e-03	0.1001090749	1.6208865500	0.7829553182	8.219031e-02	0.854868422	0.0188404177	0.52862673
:	:	:	:	:	:	:	:	:
274	0.079394438	0.474206374	0.0405451572	0.535437094	3.446186e-01	0.29074761	0.1537167221	0.08211869
275	0.196829628	0.483269641	0.1907855954	0.038643072	4.630155e-04	1.11322286	0.0513829234	0.13466928
276	0.022728082	0.305151509	0.0637988942	0.514504873	1.864861e-01	0.69023293	0.0603471206	0.00302178
277	0.413783716	0.132873214	0.0131770821	0.077883519	9.683508e-02	0.54682705	0.0292096409	0.28381989
278	0.062838048	0.078652846	0.0141505742	0.158630019	8.237107e-03	0.46560399	0.0032848656	0.07527805
279	0.011885724	1.722538378	0.0001362952	0.590678162	3.395036e-01	0.26810792	0.1038924781	0.87434766
280	0.824460944	0.229710293	0.0655358723	0.295826695	7.114031e-03	0.43212638	0.0978005264	0.51556654
281	0.069550334	0.210167248	0.4085948116	0.301611896	6.266596e-03	0.42646474	0.5844642431	0.01203106
282	0.006593493	0.005646764	0.1449420749	0.817389805	1.169263e+00	2.61014590	2.0084855291	0.14015366
283	0.003886927	0.202471676	0.9071801576	1.072970956	1.008253e+00	0.18053536	0.0268785662	0.65804210
284	0.335343107	0.246625779	0.0605188638	0.345357277	2.185436e-01	0.42628056	0.5846540508	0.42404504
285	0.319925546	0.146001273	0.0114592556	0.444348566	2.302057e-03	0.02461393	0.0002661842	0.88592332
286	0.623622898	0.323373807	0.0573718542	0.183831724	1.099240e-02	0.31467998	0.1914191870	0.09670372
287	0.184198188	0.047478419	0.4949944372	0.006823038	4.235105e-01	0.54813384	0.4882871704	0.09332601
288	0.087427366	0.163330073	0.1048365593	0.269999384	1.802810e-04	0.06154558	0.1354211372	1.01535941
289	0.553940186	0.411247823	0.2802980127	0.087845871	1.356904e-01	0.32529932	0.0161404262	0.88023745
290	0.239436158	0.513515787	0.4641301639	0.406208535	4.275038e-01	1.41857245	0.9984484577	0.03143082

291	0.019629836	0.026161594	0.0449732874	0.397122030	1.875173e-01	0.67362403	0.0161214630	0.93749005		
292	0.560230517	0.060458233	0.0867860966	1.298935309	7.711302e-01	0.80362024	0.4150525778	2.90842901		
293	0.598294056	1.111802150	0.1097745237	0.258907762	2.900321e-01	1.31344431	1.1915960140	0.00544142		
294	0.020420595	0.149624286	0.3406276866	0.014608874	3.506882e-01	0.06430802	0.3208880621	1.07739391		
295	0.111476377	1.221897089	0.0009011458	2.370126727	4.423464e-03	0.67371533	0.1576211317	0.04195245		
296	0.793597591	0.158018974	0.1572240090	0.232368781	9.573714e-06	0.01899207	0.1388233080	0.42989657		
297	0.020436022	0.071881180	0.8175696963	0.323187157	4.871287e-01	0.06606984	0.4148025083	0.19359822		
298	0.427768632	0.461219939	0.9185513933	0.235781508	2.005497e+00	1.79581466	0.0153550923	0.30804440		
299	0.156536238	0.057777038	0.7847656207	0.032926531	1.047188e-02	0.09161830	0.9162422532	0.22077078		
300	0.059776452	0.234095018	0.0798443533	0.237598304	2.606751e-01	0.50495180	0.1841986190	0.45419723		
301	0.719872455	0.048962822	1.3117084174	0.018736106	5.733842e-01	0.49861324	0.1127438230	0.21376812		
302	0.369930944	1.168917784	0.0556294057	0.080251672	4.501593e-01	0.04377270	0.0746294543	0.59171615		
303	0.088062460	0.244397385	0.1280990082	0.014790558	1.616705e-02	0.12085961	0.2110881338	0.01136999		
	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8		
1	1.581614e-02	0.2187906353	0.2486639281	0.2908220013	5.717765e-03	0.1192282838	0.0005881407	0.0220287442		
2	1.053331e-02	0.0464301935	0.0657898718	0.2983594593	2.619088e-01	0.0314453858	0.0003366012	0.1194241527		
3	3.536748e-01	0.0001944803	0.0216264736	0.0174666084	7.963625e-02	0.0622993343	0.0011625774	0.0004267115		
4	5.119396e-01	0.0426563041	0.0001631188	0.0021871048	9.649924e-03	0.0439276282	0.0035429032	0.0054751208		
5	1.048567e-02	0.0068960127	0.6137311578	0.0497687039	9.862406e-06	0.0005557652	0.1523677065	0.0280274829		
6	4.175036e-02	0.0145371382	0.0005847489	0.2213290159	1.283180e-01	0.0498448690	0.0011119752	0.1277047730		
7	8.502953e-04	0.3434742301	0.2404632863	0.0828235286	2.074361e-01	0.0386759231	0.0212450943	0.0016007713		
8	4.346590e-01	0.1587953057	0.0001558971	0.1410282771	4.860549e-02	0.1385728173	0.0126429610	0.0078609788		
9	4.570422e-02	0.0773152491	0.4169020373	0.0048876176	1.478248e-05	0.0666463473	0.2659407433	0.0477913137		
10	2.134957e-01	0.0003793075	0.1779109411	0.0875821950	9.093183e-04	0.0991139644	0.0001159120	0.2565708997		
11	1.371603e-01	0.0370712890	0.0059260666	0.0030464851	1.356301e-02	0.0424209863	0.0287678433	0.1513533390		
12	3.954678e-01	0.0561073805	0.1977503130	0.0507469713	7.751946e-03	0.0428932648	0.0119577302	0.0007183846		
13	5.780950e-01	0.0253174687	0.0013700973	0.0153899525	4.481142e-03	0.0483538168	0.0108477037	0.0059833450		
14	1.845837e-02	0.0108015618	0.0291602072	0.1139143359	2.081804e-02	0.0907433668	0.0986848966	0.0034855642		
15	7.602609e-02	0.6439576692	0.0498036140	0.0024157588	1.072994e-02	0.0211109163	0.0522219813	0.0154849774		
16	1.861713e-01	0.0001607709	0.0441550820	0.5788070589	5.786513e-03	0.0498483167	0.0022507487	0.0749324386		
17	3.512355e-01	0.1735494940	0.2621685839	0.0020183016	1.442528e-03	0.0672364535	0.0008098610	0.0490949437		
18	7.540973e-02	0.1189456875	0.0015085301	0.6673608658	1.583249e-03	0.0633076767	0.0139358962	0.0055876933		
19	1.686490e-01	0.0427067538	0.0012089662	0.0024750562	2.903183e-02	0.0336233174	0.0915114071	0.0622561413		
20	1.108253e-02	0.3082146280	0.0011844785	0.0505266216	1.157329e-01	0.2261699827	0.0778681456	0.0040913933		
21	2.058509e-05	0.0939209592	0.0000811433	0.0036449386	1.467020e-02	0.2083435894	0.0090845527	0.0446551075		
22	2.993863e-01	0.1162259585	0.0121549663	0.0020050714	1.838219e-02	0.0005374987	0.1343773317	0.0207176567		
23	4.547435e-01	0.0978465855	0.0004622234	0.0266074466	3.998184e-04	0.0037199850	0.0811089468	0.0542390660		
24	6.624011e-	0.0642794700	0.1841432560	0.0613936002	2.715104e-	0.0122754953	0.3684598198	0.0003362213		

02					02				
25	1.147864e-01	0.1121895311	0.1289745303	0.0001117553	1.198506e-01	0.0955564514	0.0892608262	0.0277130212	
26	1.379224e-03	0.4790545832	0.0953793417	0.0263447294	9.851855e-02	0.1788113973	0.0000811950	0.0426471480	
27	4.414558e-02	0.1355349234	0.3809187459	0.0253513806	4.989808e-02	0.0070922192	0.1555932153	0.0029914473	
28	2.221558e-01	0.1708650306	0.0159334313	0.0942002666	1.513975e-01	0.0005530954	0.0505599154	0.0220798719	
29	8.613110e-04	0.6974005940	0.0121499003	0.0729517423	1.873666e-03	0.0065014049	0.0140573779	0.1272498126	
30	3.490299e-03	0.0299520848	0.3859108657	0.1800558181	1.635400e-02	0.1614525430	0.0031648440	0.0798632625	
:	:	:	:	:	:	:	:	:	
274	0.071629470	0.237941811	1.618912e-02	0.206503969	1.149984e-01	0.092089801	4.330446e-02	0.0208060542	
275	0.169464222	0.231408220	7.269686e-02	0.014222551	1.474466e-04	0.336483205	1.381390e-02	0.0325613311	
276	0.024255013	0.181115642	3.013244e-02	0.234718058	7.360998e-02	0.258599760	2.010968e-02	0.0009056252	
277	0.340858021	0.060875043	4.803988e-03	0.027426095	2.950424e-02	0.158140766	7.513395e-03	0.0656582034	
278	0.094886379	0.066053730	9.456659e-03	0.102396439	4.600526e-03	0.246826398	1.548848e-03	0.0319224050	
279	0.006582537	0.530564649	3.340648e-05	0.139841763	6.954463e-02	0.052127998	1.796643e-02	0.1359871514	
280	0.478902385	0.074209428	1.684762e-02	0.073456895	1.528425e-03	0.088121406	1.773895e-02	0.0841022616	
281	0.038393131	0.064523967	9.982278e-02	0.071173949	1.279492e-03	0.082647774	1.007447e-01	0.0018651104	
282	0.002125247	0.001012267	2.067620e-02	0.112626686	1.393983e-01	0.295360427	2.021494e-01	0.0126865718	
283	0.002074537	0.060100873	2.142845e-01	0.244805428	1.990380e-01	0.033827568	4.479513e-03	0.0986312425	
284	0.229221883	0.093757686	1.830796e-02	0.100914505	5.525306e-02	0.102295362	1.247888e-01	0.0814000589	
285	0.297716859	0.075563640	4.719472e-03	0.176765124	7.923591e-04	0.008041353	7.734758e-05	0.2315243536	
286	0.423263201	0.122066085	1.723336e-02	0.053336831	2.759513e-03	0.074981011	4.056802e-02	0.0184322221	
287	0.156735752	0.022468842	1.864084e-01	0.002481869	1.332903e-01	0.163743119	1.297382e-01	0.0223013667	
288	0.114480830	0.118946932	6.075480e-02	0.151135482	8.731461e-05	0.028292801	5.537092e-02	0.3733801358	
289	0.341233177	0.140894392	7.641716e-02	0.023132790	3.091642e-02	0.070350187	3.104654e-03	0.1522769805	
290	0.118345590	0.141162019	1.015277e-01	0.085828136	7.815448e-02	0.246154392	1.540983e-01	0.0043627887	
291	0.019968819	0.014801360	2.024757e-02	0.172694284	7.055515e-02	0.240573058	5.120944e-03	0.2678233258	
292	0.129702722	0.007784657	8.892318e-03	0.128554754	6.603308e-02	0.065317104	3.000512e-02	0.1890979559	
293	0.194064830	0.200567876	1.575855e-02	0.035900088	3.479604e-02	0.149567604	1.206899e-01	0.0004956683	
294	0.018827455	0.076723259	1.389907e-01	0.005757823	1.195903e-01	0.020815273	9.238186e-02	0.2789611928	
295	0.044516284	0.271376607	1.592625e-04	0.404600037	6.533564e-04	0.094450924	1.965443e-02	0.0047047812	
296	0.383421126	0.042460651	3.361840e-02	0.047992323	1.710833e-06	0.003221379	2.094344e-02	0.0583291597	
297	0.016444946	0.032170144	2.911678e-01	0.111175343	1.449878e-01	0.018665218	1.042287e-01	0.0437505473	
298	0.130697301	0.078373181	1.242063e-01	0.030795390	2.266373e-01	0.192625119	1.464941e-03	0.0264312104	

299	0.125472171	0.025756658	2.783909e-01	0.011282278	3.104621e-03	0.025781518	2.293257e-01	0.0496958535
300	0.044268466	0.096418025	2.616924e-02	0.075218725	7.140284e-02	0.131282849	4.259519e-02	0.0944615066
301	0.306744273	0.011603502	2.473664e-01	0.003412861	9.036873e-02	0.074589627	1.500112e-02	0.0255805503
302	0.229997603	0.404192575	1.530698e-02	0.021329223	1.035191e-01	0.009554317	1.448848e-02	0.1033146985
303	0.111161863	0.171578663	7.156378e-02	0.007981194	7.548268e-03	0.053559921	8.320295e-02	0.0040306148

In [36]:

```
# Results for Variables
#res.var <- get_pca_var(res.pca)
#res.var$coord      # Coordinates
#res.var$contrib     # Contributions to the PCs
#res.var$cos2       # Quality of representation

# Results for individuals
#res.ind <- get_pca_ind(res.pca)
#res.ind$coord      # Coordinates
#res.ind$contrib     # Contributions to the PCs

head(res.ind$cos2)    # Quality of representation
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8	
1	0.01581614	0.2187906353	0.2486639281	0.290822001	5.717765e-03	0.1192282838	0.0005881407	0.0220287442	0.0
2	0.01053331	0.0464301935	0.0657898718	0.298359459	2.619088e-01	0.0314453858	0.0003366012	0.1194241527	0.0
3	0.35367476	0.0001944803	0.0216264736	0.017466608	7.963625e-02	0.0622993343	0.0011625774	0.0004267115	0.1
4	0.51193963	0.0426563041	0.0001631188	0.002187105	9.649924e-03	0.0439276282	0.0035429032	0.0054751208	0.0
5	0.01048567	0.0068960127	0.6137311578	0.049768704	9.862406e-06	0.0005557652	0.1523677065	0.0280274829	0.0
6	0.04175036	0.0145371382	0.0005847489	0.221329016	1.283180e-01	0.0498448690	0.0011119752	0.1277047730	0.0

In []:

Naive-Bayes

Naive Bayes es un modelo estadístico basado en el teorema de Bayes que se utiliza para resolver problemas de clasificación siguiendo un enfoque probabilístico. Se basa en la idea de que las variables predictoras en un modelo de aprendizaje automático son independientes entre sí. Lo que significa que el resultado de un modelo depende de un conjunto de variables independientes que no tienen nada que ver entre sí.

Las variables predictoras no siempre son independientes entre sí, siempre hay algunas correlaciones entre ellas. Dado que Naive Bayes considera que cada variable predictiva es independiente de cualquier otra variable en el modelo, se le llama *Naive*.

El principio detrás de Naive Bayes es el teorema de Bayes, también conocido como la Regla de Bayes. El teorema se usa para calcular la probabilidad condicional, que no es más que la probabilidad de que ocurra un evento en base a la información sobre los eventos del pasado. Estadísticamente se represeta como:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Donde:

- P(A|B): probabilidad condicional de que ocurra el evento A, dado el evento B
 - P(A): Probabilidad de que ocurra el evento A
 - P(B): Probabilidad de que ocurra el evento B
 - P(B|A): probabilidad condicional de que ocurra el evento B, dado el evento A
- Formalmente, las terminologías del Teorema bayesiano son las siguientes:

A se conoce como la proposición y B es la evidencia.

- P(A) representa la probabilidad previa de la proposición
- P(B) representa la probabilidad previa de evidencia
- P(A|B) se llama la posterior
- P(B|A) es la probabilidad

Por lo tanto, el teorema de Bayes se puede resumir como:

$$\text{Posterior} = (\text{Probabilidad}) (\text{Probabilidad previa de la proposición}) / \text{Probabilidad previa probatoria}$$

También se puede considerar de la siguiente manera:

Dada una hipótesis H y evidencia E, el teorema de Bayes establece que la relación entre la probabilidad de hipótesis antes de obtener la evidencia P(H) y la probabilidad de la hipótesis después de obtener la evidencia P(H|E) es:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

Derivación del teorema de Bayes El objetivo principal del Teorema de Bayes es calcular la probabilidad condicional. La regla de Bayes se puede derivar de las siguientes dos ecuaciones:

La siguiente ecuación representa la probabilidad condicional de A, dada B:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

La siguiente ecuación representa la probabilidad condicional de B, dada A:

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

Por lo tanto, al combinar las dos ecuaciones anteriores obtenemos el Teorema de Bayes:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

La ecuación anterior fue para una sola variable predictiva, sin embargo, en las aplicaciones del mundo real, hay más de una variable predictiva y para un problema de clasificación, hay más de una clase de salida. Las clases se pueden representar como, C1, C2, ..., Ck y las variables predictoras se pueden representar como un vector, x1, x2, ..., xn.

El objetivo de un algoritmo Naive Bayes es medir la probabilidad condicional de un evento con un vector de características x1, x2, ..., xn perteneciente a una clase particular Ci,

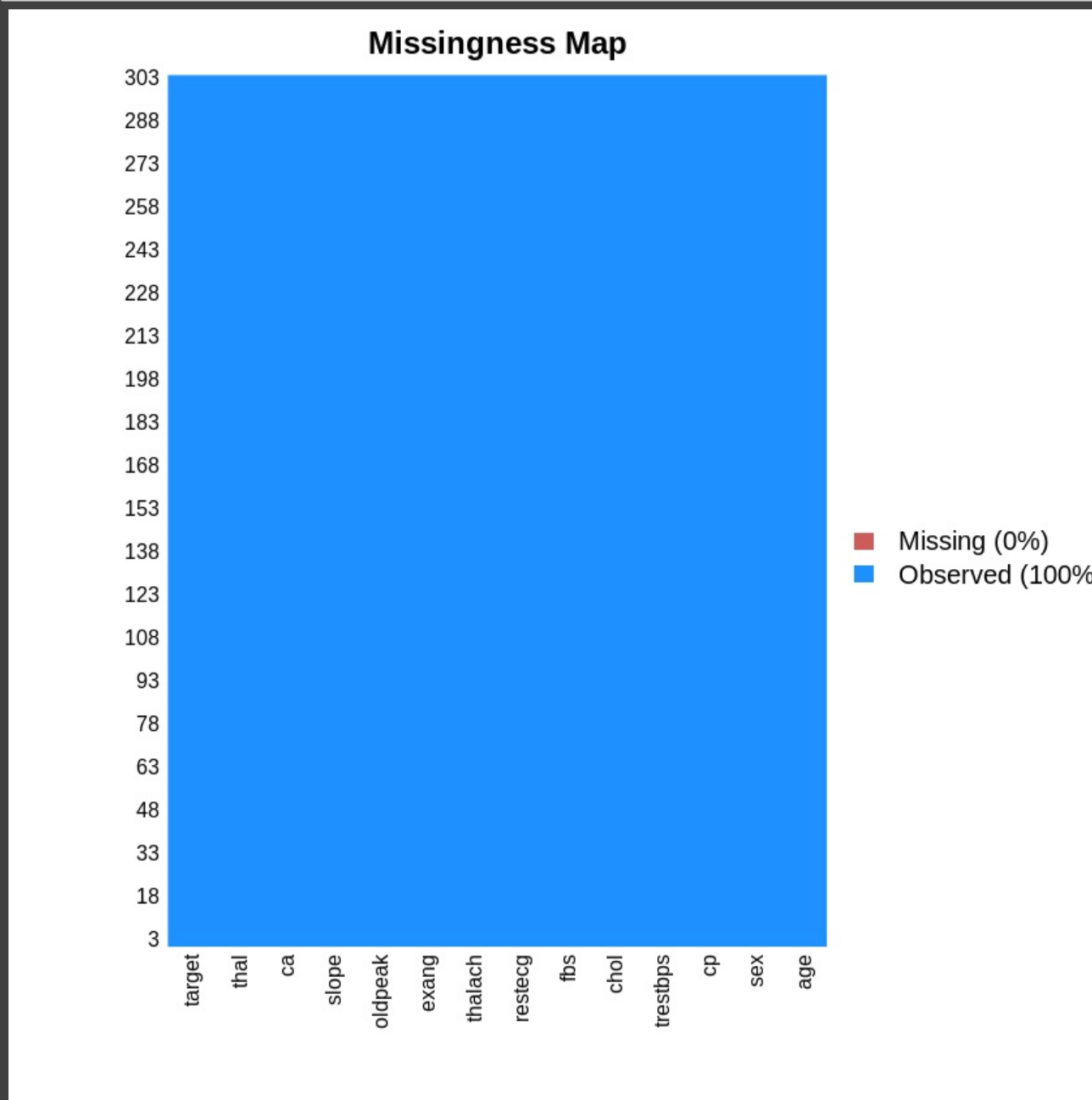
$$P(C_i|x_1, ..., x_n) = \frac{P(x_1, ..., x_n)P(C_i)}{P(x_1, ..., x_n)}$$

Implementación de Naive-Bayes

Preprocesamiento

In [97]:

```
library(Rcpp)
library(Amelia)
missmap(heart_data)
```



Primero verificamos que no haya valores faltantes

Selección de Características

In [80]:

```
library(caret)
heart_data$target <- factor(heart_data$target)
indxTrain <- createDataPartition(y = heart_data$target, p = 0.80, list = FALSE)
training <- data[indxTrain,]
testing <- data[-indxTrain,]
```

In [81]:

```
prop.table(table(heart_data$target)) * 100
```

```
      0      1
45.54455 54.45545
```

In [82]:

```
prop.table(table(training$target)) * 100
```

```
      0      1  
45.67901 54.32099
```

In [83]:

```
prop.table(table(testing$target)) * 100
```

```
      0      1  
45      55
```

In [89]:

```
x = training[,-14]  
y = factor(training$target)
```

In [100]:

```
model = train(x,y,'nb',trControl=trainControl(method='cv',number=10))  
model
```

Naive Bayes

243 samples
13 predictor
2 classes: '0', '1'

No pre-processing

Resampling: Cross-Validated (10 fold)

Summary of sample sizes: 218, 219, 219, 219, 219, 219, ...

Resampling results across tuning parameters:

usekernel	Accuracy	Kappa
FALSE	0.8235000	0.6424545
TRUE	0.8363333	0.6669133

Tuning parameter 'fL' was held constant at a value of 0

Tuning

parameter 'adjust' was held constant at a value of 1

Accuracy was used to select the optimal model using the largest value.

The final values used for the model were fL = 0, usekernel = TRUE and adjust = 1.

Model Evaluation

In [94]:

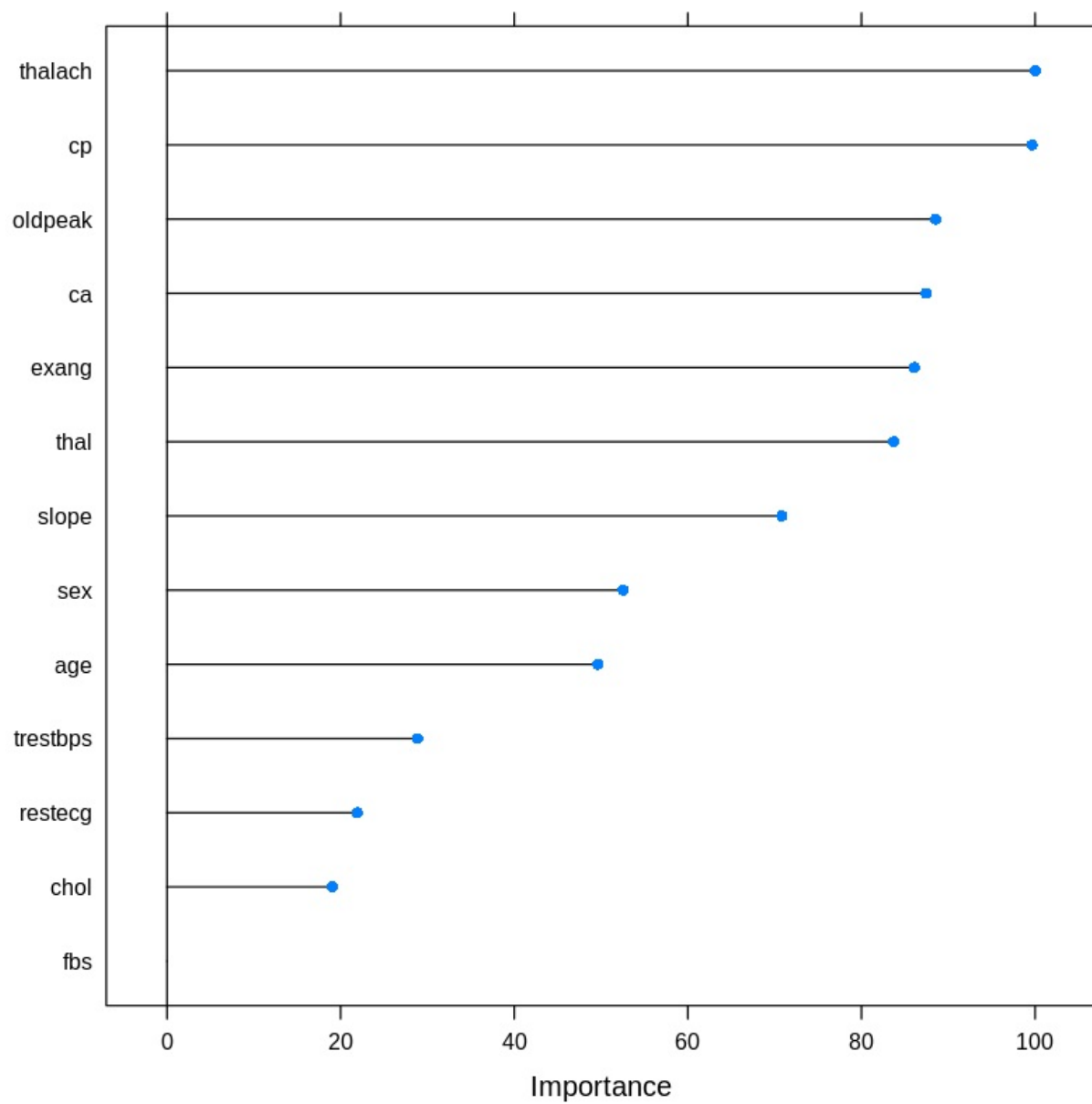
```
pred <- predict(model,newdata = testing )  
#Get the confusion matrix to see accuracy value and other parameter values  
confusionMatrix(pred, factor(testing$target) )
```

Confusion Matrix and Statistics

```
      Reference  
Prediction 0  1  
0  20  3  
1   7 30  
  
      Accuracy : 0.8333  
      95% CI : (0.7148, 0.9171)  
No Information Rate : 0.55  
P-Value [Acc > NIR] : 3.483e-06  
  
      Kappa : 0.6587  
McNemar's Test P-Value : 0.3428  
  
      Sensitivity : 0.7407  
      Specificity : 0.9091  
Pos Pred Value : 0.8696  
Neg Pred Value : 0.8108  
Prevalence : 0.4500  
Detection Rate : 0.3333  
Detection Prevalence : 0.3833  
Balanced Accuracy : 0.8249  
  
      'Positive' Class : 0
```

In [96]:

```
#Plot Variable performance  
X <- varImp(model)  
plot(X)
```



In []: