

In [1]:

```
df <- read.table('babies.txt', head=TRUE)
head(df, 10)
```

A data.frame: 10 × 7

	bwt	gestation	parity	age	height	weight	smoke
	<int>	<int>	<int>	<int>	<int>	<int>	<int>
1	120	284	0	27	62	100	0
2	113	282	0	33	64	135	0
3	128	279	0	28	64	115	1
4	123	999	0	36	69	190	0
5	108	282	0	23	67	125	1
6	136	286	0	25	62	93	0
7	138	244	0	33	62	178	0
8	132	245	0	23	65	140	0
9	120	289	0	25	62	125	0
10	143	299	0	30	66	136	1

In [2]:

```
nrow(df) # number of raw data
```

1236

In [3]:

```
unique(df$smoke)
```

0 1 9

In [4]:

```
df2 <- subset(df, smoke!=9) # cleared data
```

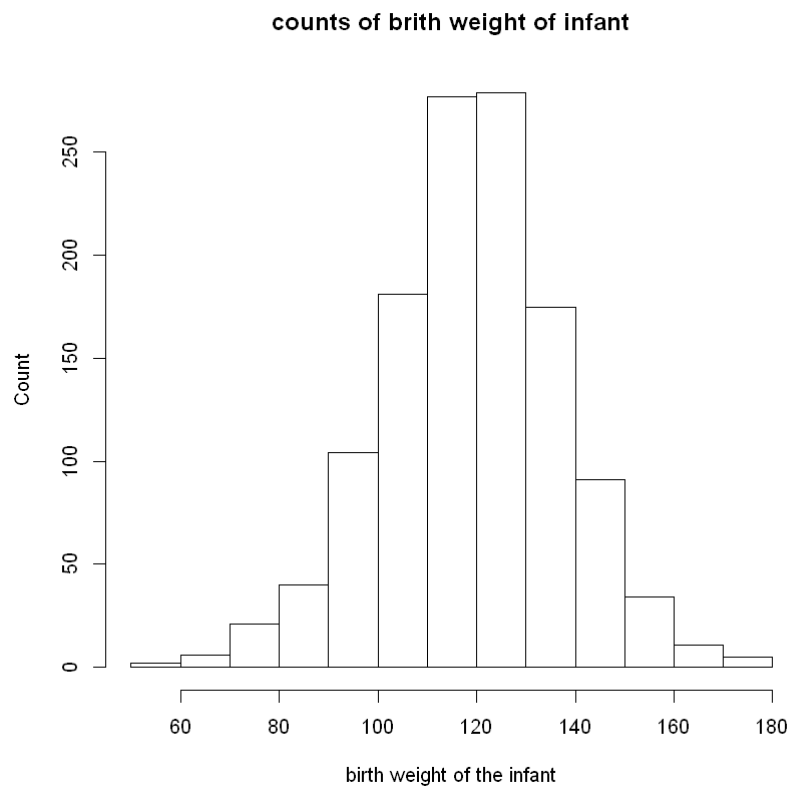
In [5]:

```
nrow(df2) #number of valid data
```

1226

In [6]:

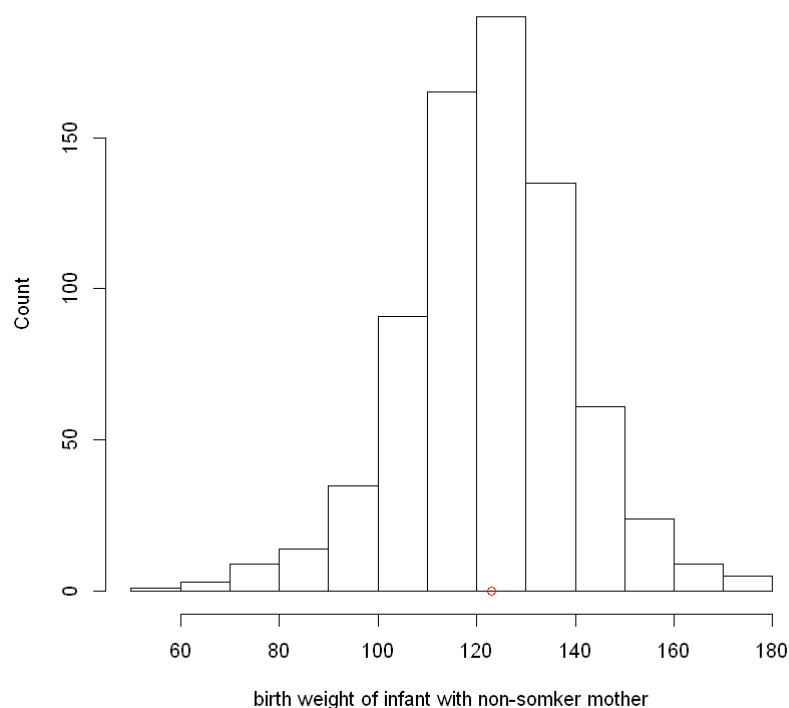
```
hist(df2$bwt,xlab = 'birth weight of the infant',ylab = 'Count',main='counts of brith w  
eight of infant')
```



In [7]:

```
non_smoker <- subset(df2,smoke==0)
h1<-hist(non_smoker$bwt,xlab='birth weight of infant with non-smoker mother',ylab='Count',main=
        'counts of birth weight of infant with non-smoker mother')
points(x=mean(non_smoker$bwt),y=0,col='red')
```

counts of birth weight of infant with non-smoker mother



In [8]:

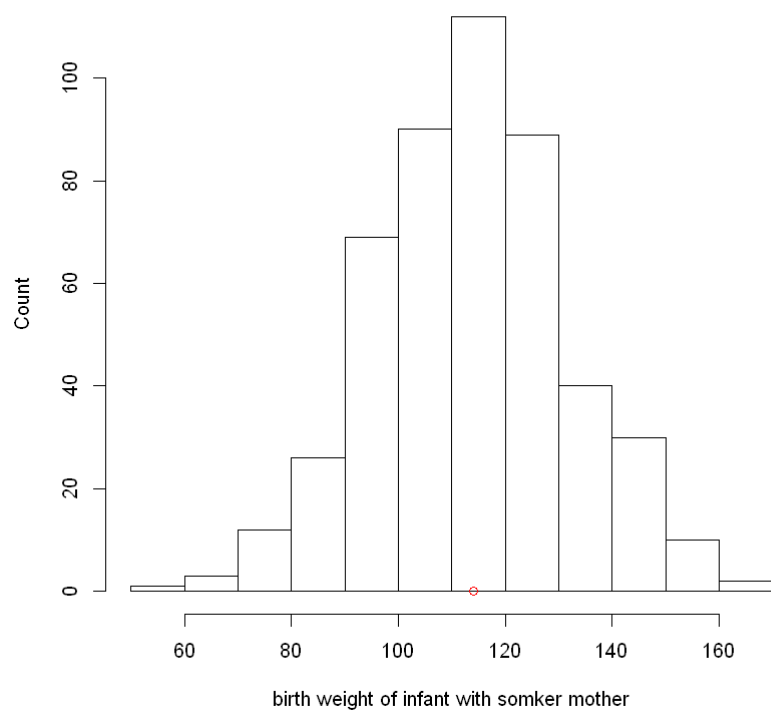
```
mean(non_smoker$bwt)
```

123.047169811321

In [9]:

```
smoker <- subset(df2,smoke==1)
h2<-hist(smoker$bwt,xlab='birth weight of infant with somker mother',ylab='Count',main=
        'counts of brith weight of infant with smoker mother')
points(x=mean(smoker$bwt),y=0,col='red')
```

counts of brith weight of infant with smoker mother



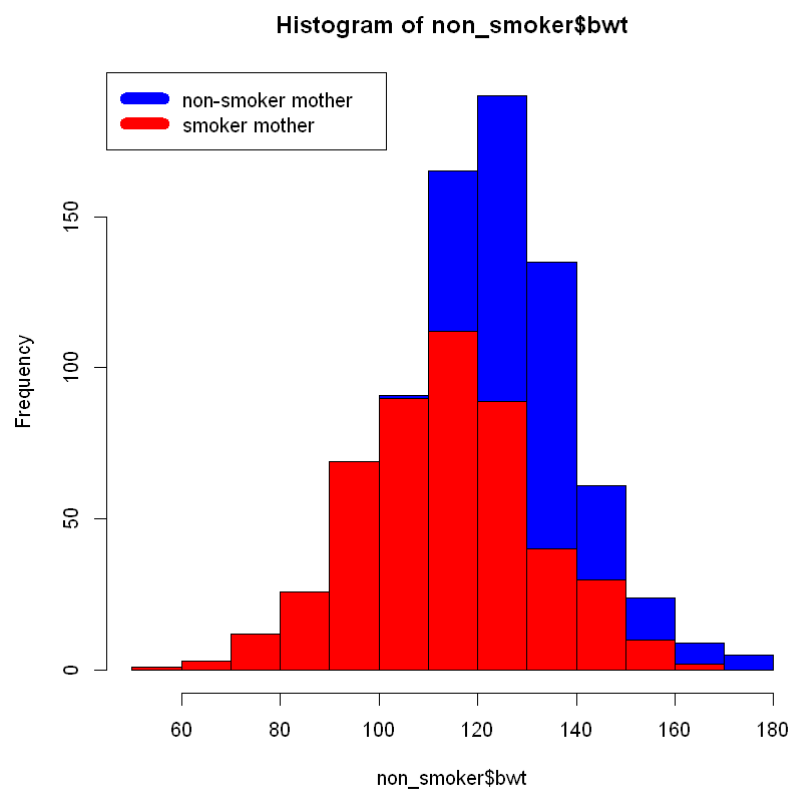
In [10]:

```
mean(smoker$bwt)
```

114.109504132231

In [11]:

```
plot(h1,col='blue')  
plot(h2,add=T,col='red')  
legend('topleft',c('non-smoker mother','smoker mother'),col=c('blue','red'),lwd=10)
```



In [12]:

```
num_non_smoker <- nrow(non_smoker)  
num_non_smoker
```

742

In [13]:

```
num_smoker <- nrow(smoker)  
num_smoker
```

484

In [14]:

```
permutation_test <- function(dataframe){  
  perm <- sample(df2$bwt)  
  mean_smoker <- mean(perm[1:num_smoker])  
  mean_non_smoker <- mean(perm[(num_smoker+1):(  
    num_smoker+num_non_smoker)])  
  res <- mean_non_smoker - mean_smoker  
  return(res)  
}  
diff <- mean(non_smoker$bwt)-mean(smoker$bwt)
```

In [15]:

```
permutation_diffs <- replicate(1000,permutation_test(df2))
```

In [16]:

```
hist(permutation_diffs,xlim = c(-10,10),main='result of permutation test',xlab='differe  
nce from permutation',ylab='Counts')  
abline(v=diff,col='red')
```

