**R Code for Examples in the book**



***“Statistics: The Art and Science of Learning from Data”***

**by Agresti, Franklin and Klingenberg, 5th edition**

**Chapter 2**

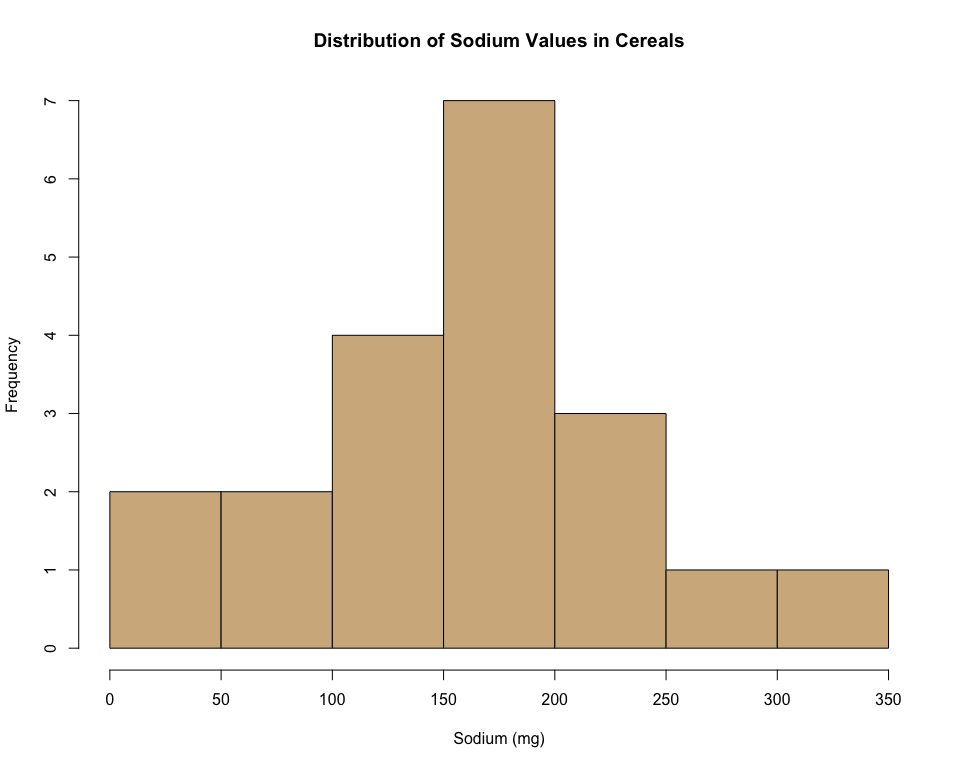
**Example 7: Health Value of Cereals – Histogram for a Continuous Variable**

## Read in Sodium values:

sodium <- c(0, 340, 70, 140, 200, 180, 210, 150, 100, 130,   
 140, 180, 190, 160, 290, 50, 220, 180, 200, 210)

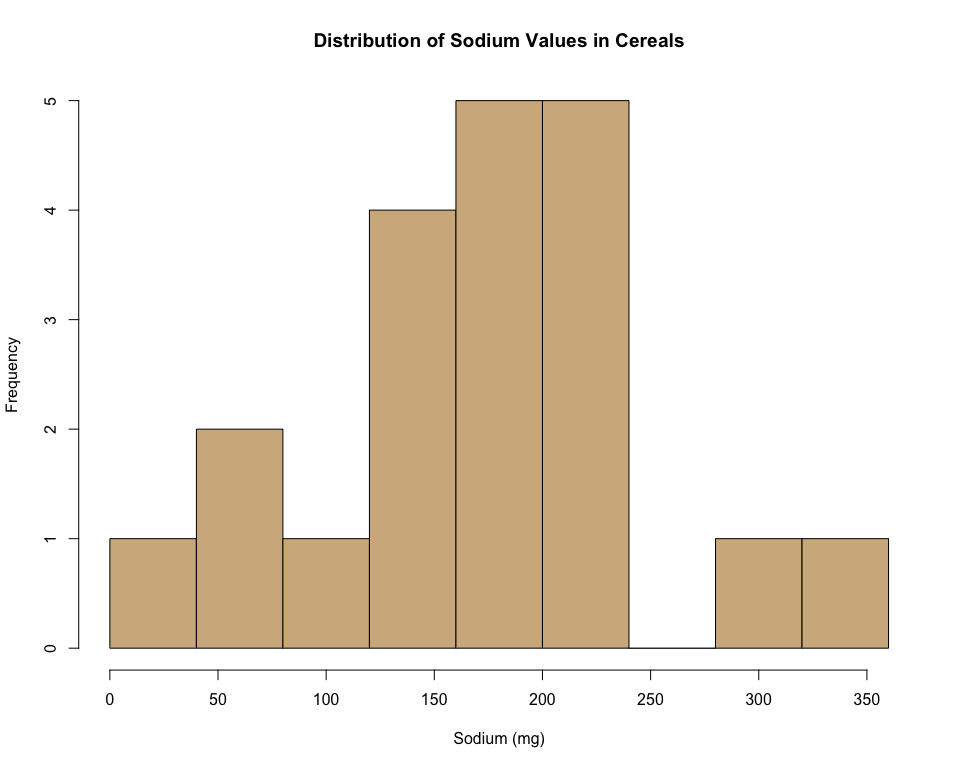
## Create Basic Histogram:

hist(sodium, col = 'tan',   
 main = 'Distribution of Sodium Values in Cereals',  
 xlab = 'Sodium (mg)', ylab = 'Frequency')



## Changing the bins by providing the boundaries. (Note: right = FALSE puts an observation such as 120 in the interval from 120-160 and not 80-120).

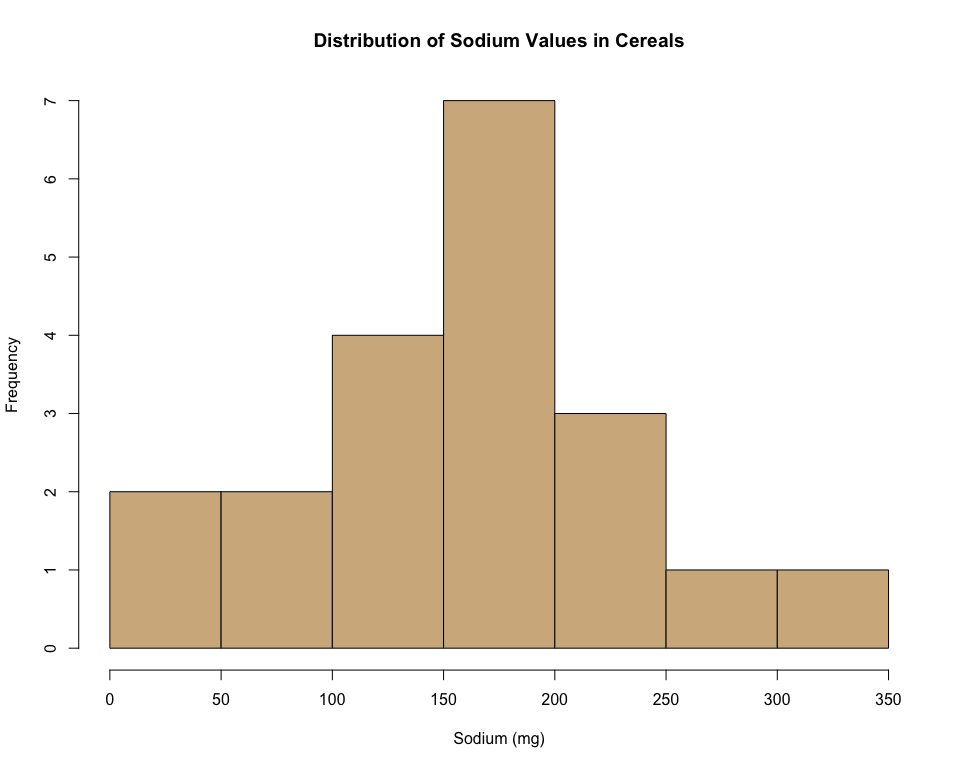
hist(sodium, breaks = seq(0,360,40), right = FALSE, col = 'tan',   
 main = 'Distribution of Sodium Values in Cereals',   
 xlab = 'Sodium (mg)', ylab = 'Frequency')



## 

## Another way to request a certain number of bins:

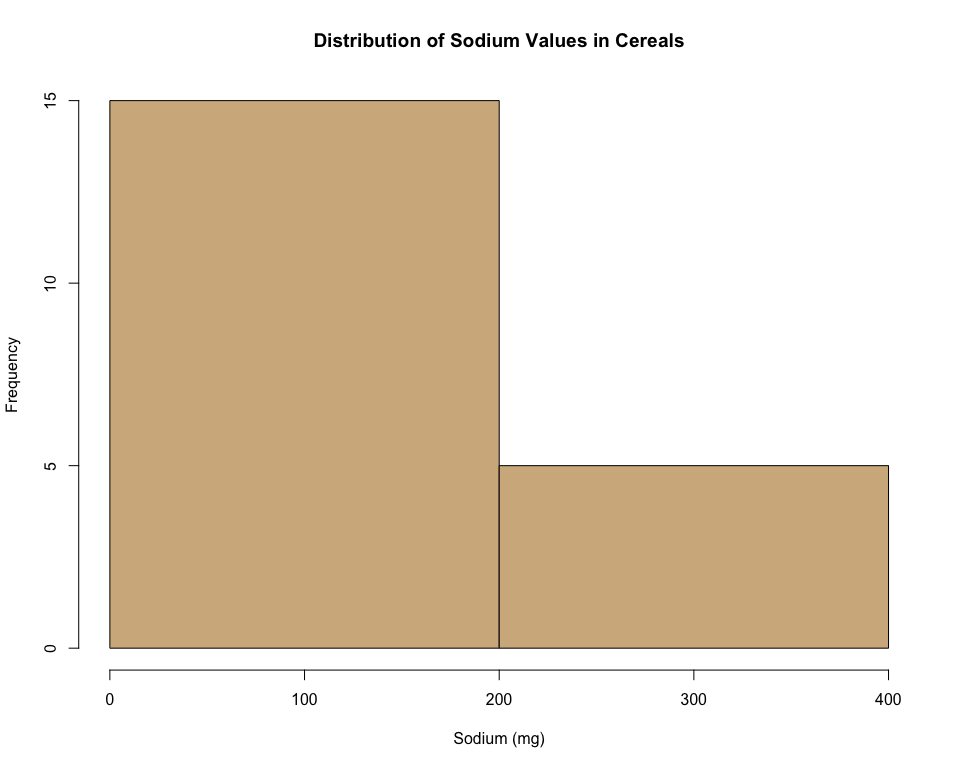
hist(sodium, breaks = 10, col = 'tan',   
 main = 'Distribution of Sodium Values in Cereals',  
 xlab = 'Sodium (mg)', ylab = 'Frequency')



## 

## Too few breaks:

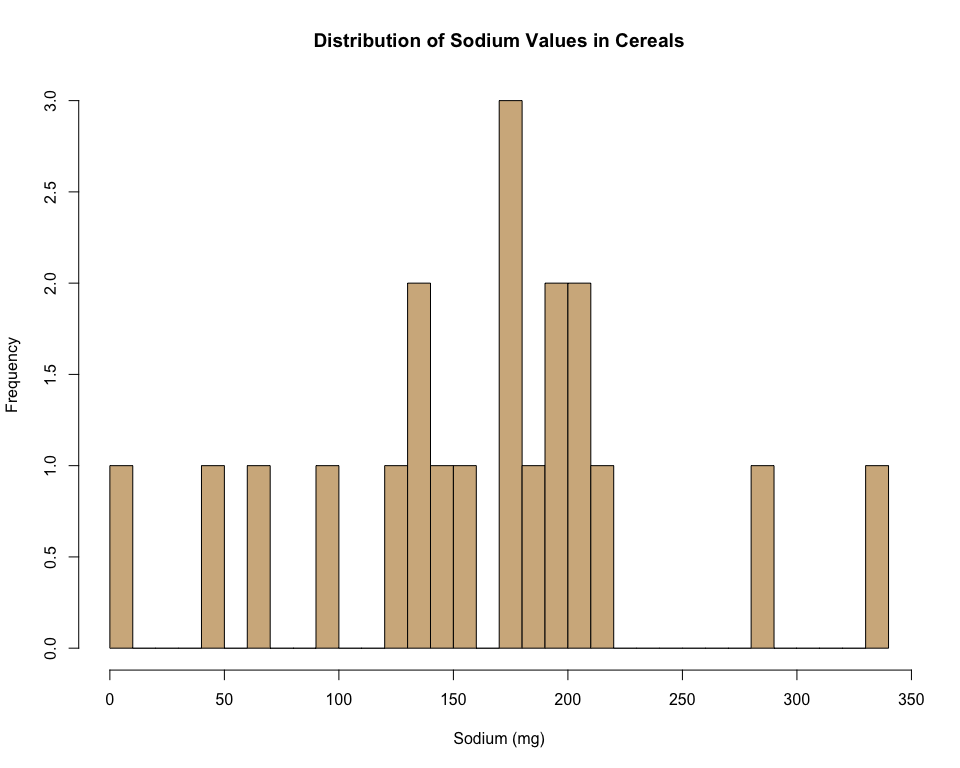
hist(sodium, breaks = 2, col = 'tan',  
 main = 'Distribution of Sodium Values in Cereals',  
 xlab = 'Sodium (mg)', ylab = 'Frequency')



## 

## Too many breaks:

hist(sodium, breaks = 30, col = 'tan',  
 main = 'Distribution of Sodium Values in Cereals',  
 xlab = 'Sodium (mg)', ylab = 'Frequency')



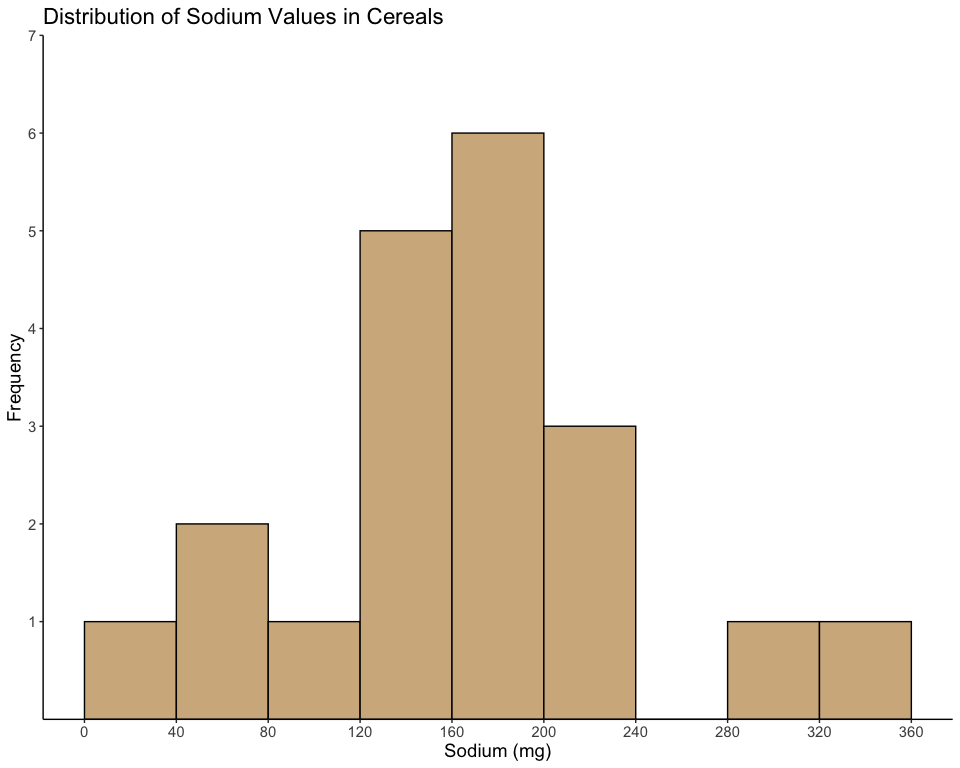
## 

## For more fine tuning, it is better to use the ggplot2 library. If you haven’t installed it already, first type: install.packages(ggplot2).

library(ggplot2)

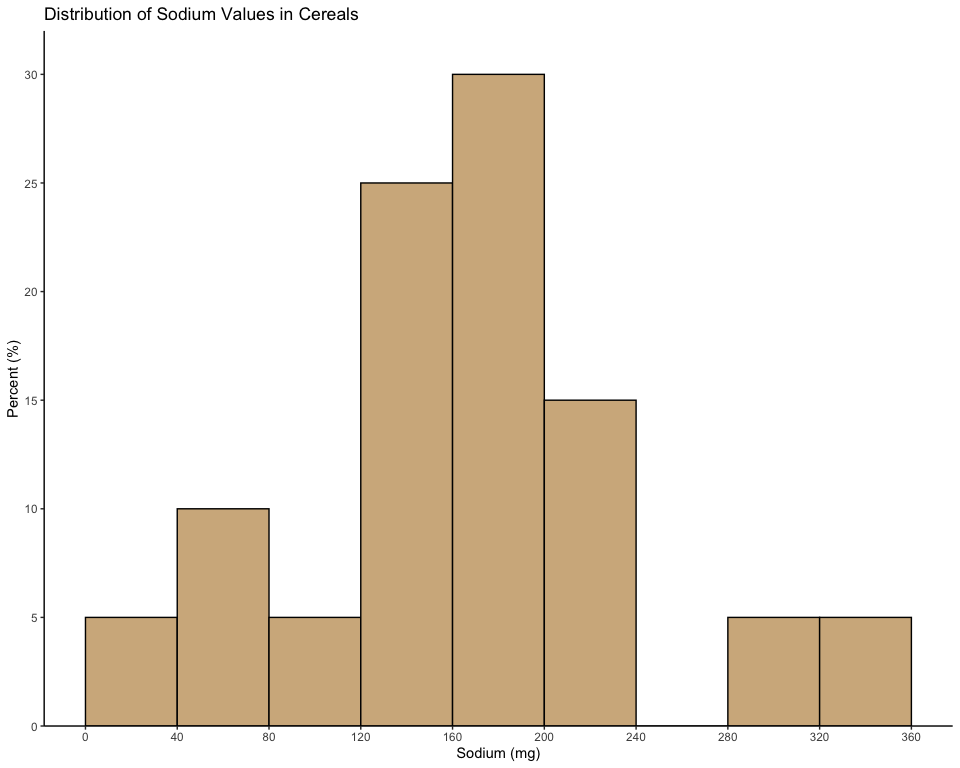
## Adjusting x-axis labels:

ggplot(data.frame(sodium), aes(x = sodium)) +   
 geom\_histogram(breaks = seq(0,360,40), color="black", fill="tan") +  
 labs(x = 'Sodium (mg)', y = 'Frequency',   
 title="Distribution of Sodium Values in Cereals") +  
 scale\_y\_continuous(limit = c(0,7),   
 breaks = 1:7,   
 expand = c(0,0)) +  
 scale\_x\_continuous(breaks = seq(0,360,40)) +  
 theme\_classic() +  
 theme(panel.grid.minor = element\_blank(),   
 text=element\_text(size = 14))



## Plotting percentages rather than counts on the y-axis:

ggplot(data.frame(sodium),   
 aes(x = sodium, y = 100 \* (..count.. / sum(..count..)))) +   
 geom\_histogram(breaks = seq(0,360,40), color = 'black', fill = 'tan') +  
 labs(x = 'Sodium (mg)', y = 'Percent (%)',   
 title = 'Distribution of Sodium Values in Cereals') +  
 scale\_y\_continuous(limit = c(0,32),   
 breaks = seq(0,30,5),   
 expand = c(0,0)) +  
 scale\_x\_continuous(breaks = seq(0,360,40)) +  
 theme\_classic() +  
 theme(panel.grid.minor=element\_blank())



## 

## R actually defines intervals open to the left and closed to the right. To get the histograms perfectly match the ones in the textbook, use closed = 'left':

ggplot(data.frame(sodium),   
 aes(x = sodium, y = 100 \* (..count.. / sum(..count..)))) +   
 geom\_histogram(breaks = seq(0,360,40), closed = 'left', color = 'black', fill = 'tan') +  
 labs(x = 'Sodium (mg)', y = 'Percent (%)',   
 title = 'Distribution of Sodium Values in Cereals') +  
 scale\_y\_continuous(limit = c(0,32),   
 breaks = seq(0,30,5),   
 expand = c(0,0)) +  
 scale\_x\_continuous(breaks = seq(0,360,40)) +  
 theme\_classic() +  
 theme(panel.grid.minor = element\_blank())

