**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**

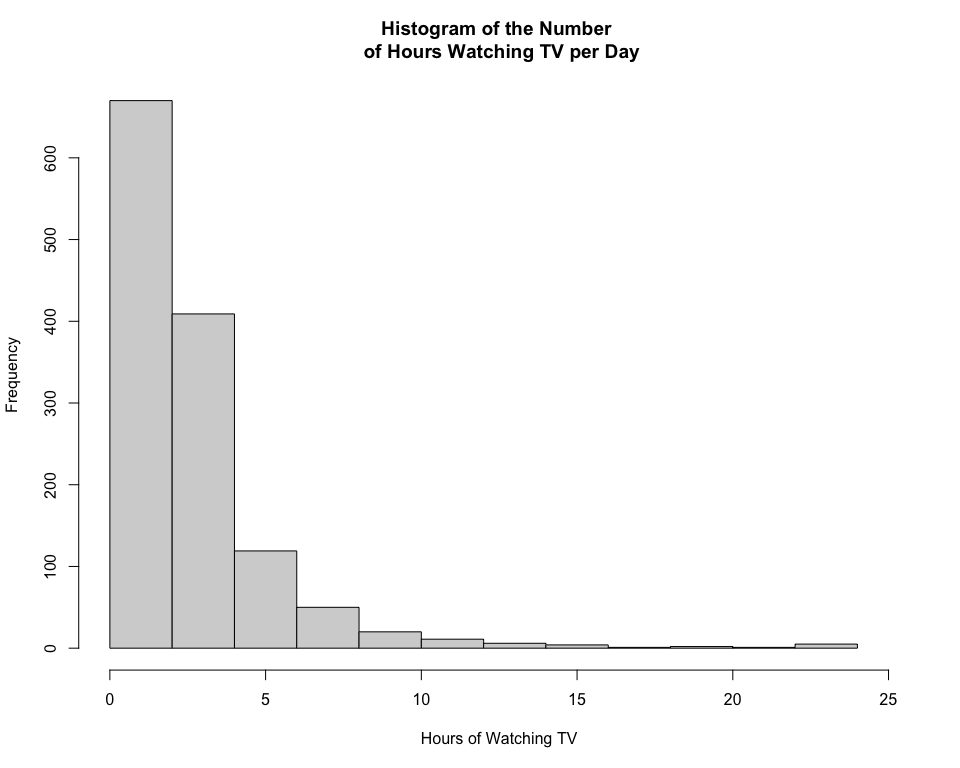
**Example 6: TV Watching – Histogram for a Discrete Variable**

## Read in GSS data

gss <- read.csv(file='https://raw.githubusercontent.com/artofstat/data/master/Chapter2/TVhours.csv')

## Create Basic Histogram

hist(gss$tvhours, xlim = c(0, 25),  
 main = 'Histogram of the Number \n of Hours Watching TV per Day',   
 xlab = 'Hours of Watching TV')

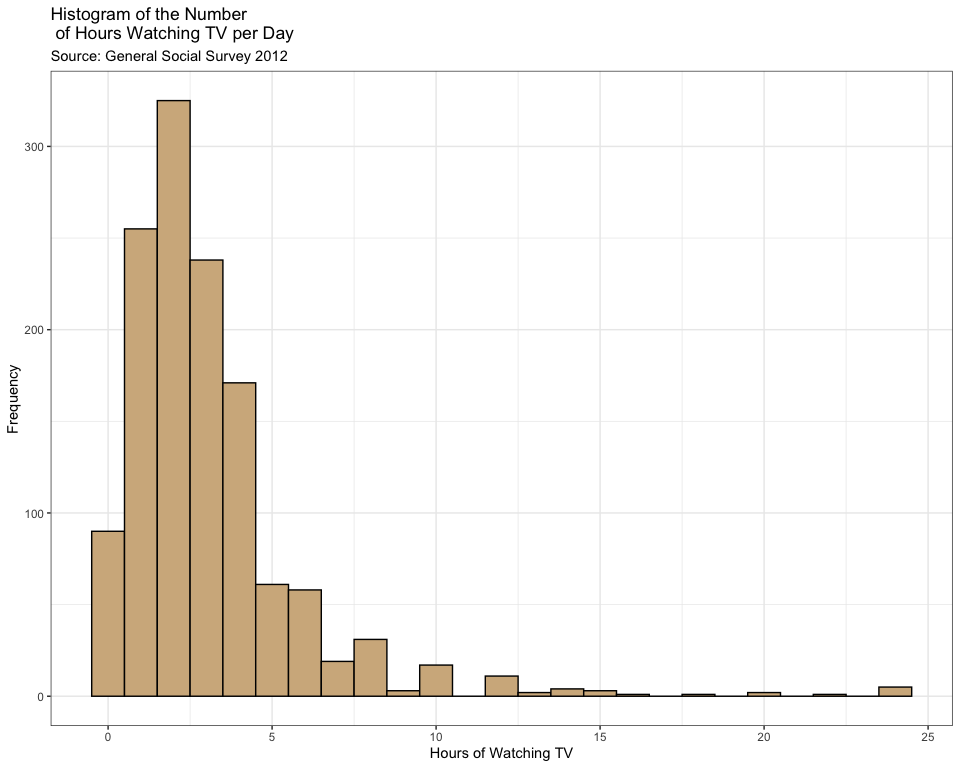


## For more fine tuning, it is better to use the ggplot2 library. If you haven’t installed it already, first type: install.packages(ggplot2).

library(ggplot2)

## Basic Frequency (Count) Plot

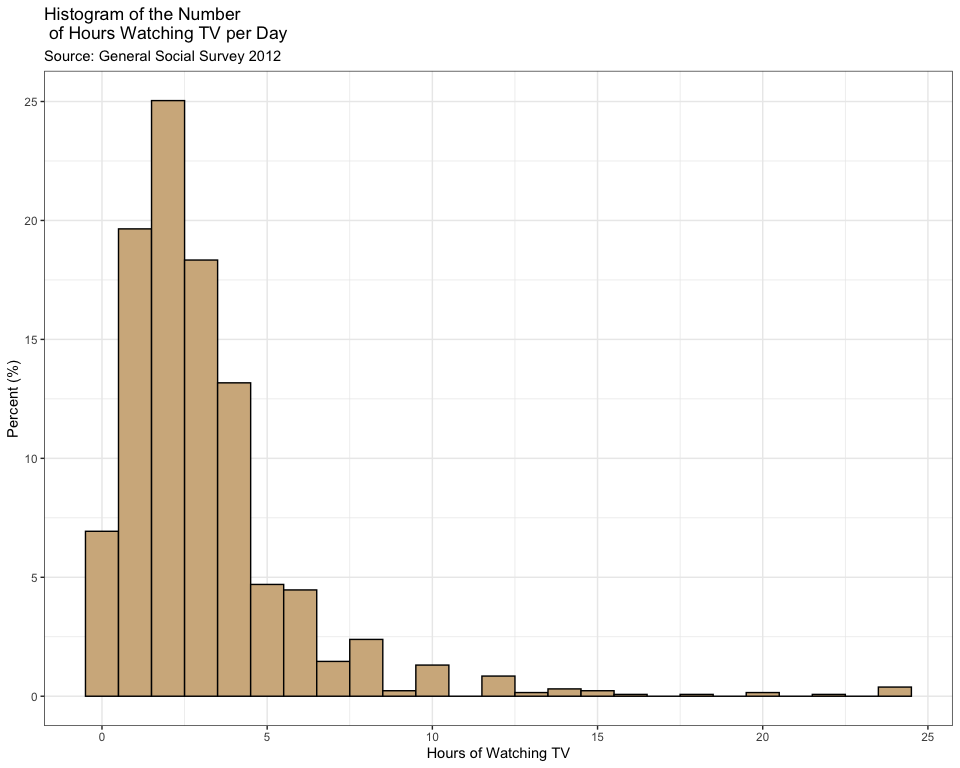
ggplot(gss, aes(x=tvhours)) +   
 geom\_histogram(binwidth=1, color = 'black', fill = 'tan') +  
 labs(x = 'Hours of Watching TV', y='Frequency',   
 title = 'Histogram of the Number \n of Hours Watching TV per Day',   
 subtitle = 'Source: General Social Survey 2012') +  
 theme\_bw()



## 

## Basic Relative Frequency (Percent) Plot

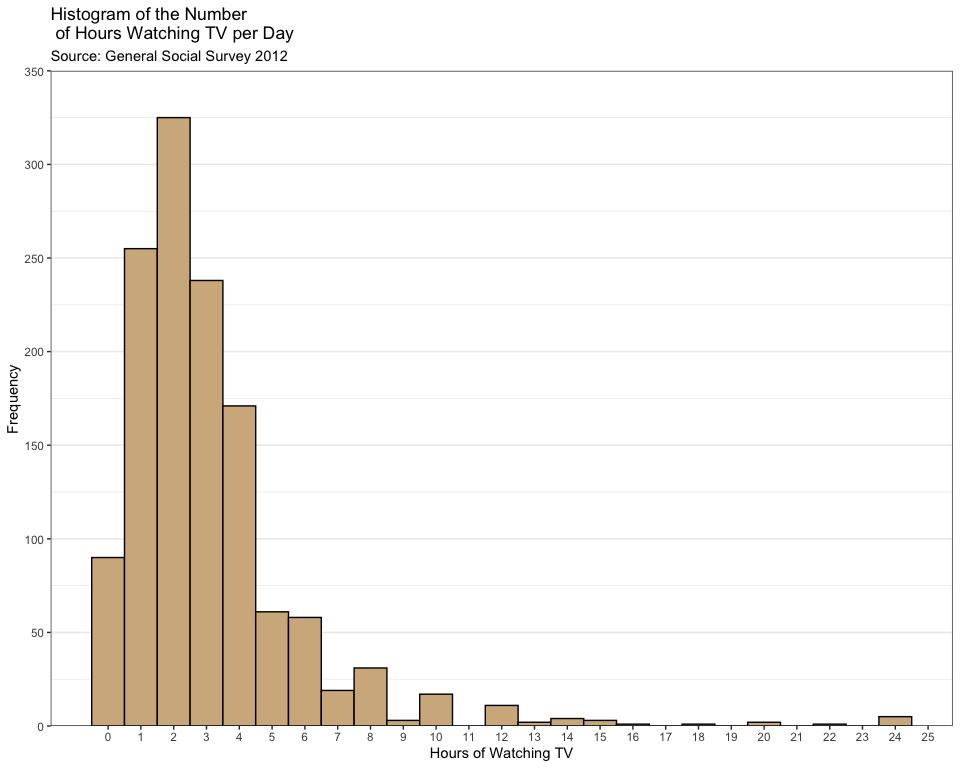
ggplot(gss, aes(x = tvhours, y = 100 \* (..count.. / sum(..count..)))) +   
 geom\_histogram(binwidth = 1, color = 'black', fill = 'tan') +  
 labs(x = 'Hours of Watching TV', y = 'Percent (%)',   
 title = 'Histogram of the Number \n of Hours Watching TV per Day',   
 subtitle = 'Source: General Social Survey 2012') +  
 theme\_bw()



## 

## More fine-tuning such as labels on x-axis

ggplot(gss, aes(x = tvhours)) +   
 geom\_histogram(center = 0, binwidth = 1, color = 'black', fill = 'tan') +  
 labs(x = 'Hours of Watching TV', y = 'Frequency',   
 title = 'Histogram of the Number \n of Hours Watching TV per Day',   
 subtitle = 'Source: General Social Survey 2012') +  
 theme\_bw() +  
 scale\_y\_continuous(limits = c(0,350),   
 breaks = seq(0,350,50),   
 expand = c(0,0)) +  
 scale\_x\_continuous(breaks = seq(0,25,1)) +  
 theme(panel.grid.major.x = element\_blank(),   
 panel.grid.minor.x = element\_blank())



## 

## with percentages rather than counts

ggplot(gss, aes(x = tvhours, y = 100 \* (..count.. / sum(..count..)))) +   
 geom\_histogram(center = 0, binwidth = 1, color = 'black', fill = 'tan') +  
 labs(x = 'Hours of Watching TV', y = 'Percent (%)',   
 title = 'Histogram of the Number \n of Hours Watching TV per Day',   
 subtitle = 'Source: General Social Survey 2012') +  
 theme\_bw() +  
 scale\_y\_continuous(limits = c(0,26),   
 breaks = seq(0,25,5),   
 expand = c(0,0)) +  
 scale\_x\_continuous(breaks = seq(0,25,1)) +  
 theme(panel.grid.major.x = element\_blank(),   
 panel.grid.minor.x = element\_blank())

