# Plots for Week 4

```
MyTheme_transparent <- theme(
    panel.background = element_rect(fill = "transparent"), # bg of the panel
    legend.background = element_rect(fill = "transparent"), # get rid of legend bg
    legend.box.background = element_rect(fill = "transparent"), # get rid of legend panel bg
    legend.key = element_rect(fill = "transparent", colour = NA) # get rid of key legend fill, and of t
)
```

Histogram example

```
tacos.df <- data.frame(tacos = floor(rnorm(100, 4.5, 2)))
tacos.df$tacos[tacos.df$tacos<0] <- 0
write.csv(tacos.df, "../data/wk04_tacos.csv", row.names = FALSE)

tacos.brk <- c((0:5)*2-0.5)
g <- ggplot(tacos.df, aes(x=tacos))
g <- g + geom_histogram(breaks=tacos.brk, color="black", fill="cadetblue")
g <- g + theme_bw() + MyTheme_transparent
g <- g + scale_x_continuous(breaks=(0:5)*2)
g <- g + labs(x="Number of tacos eaten per week",
              y="Frequency")

g
```
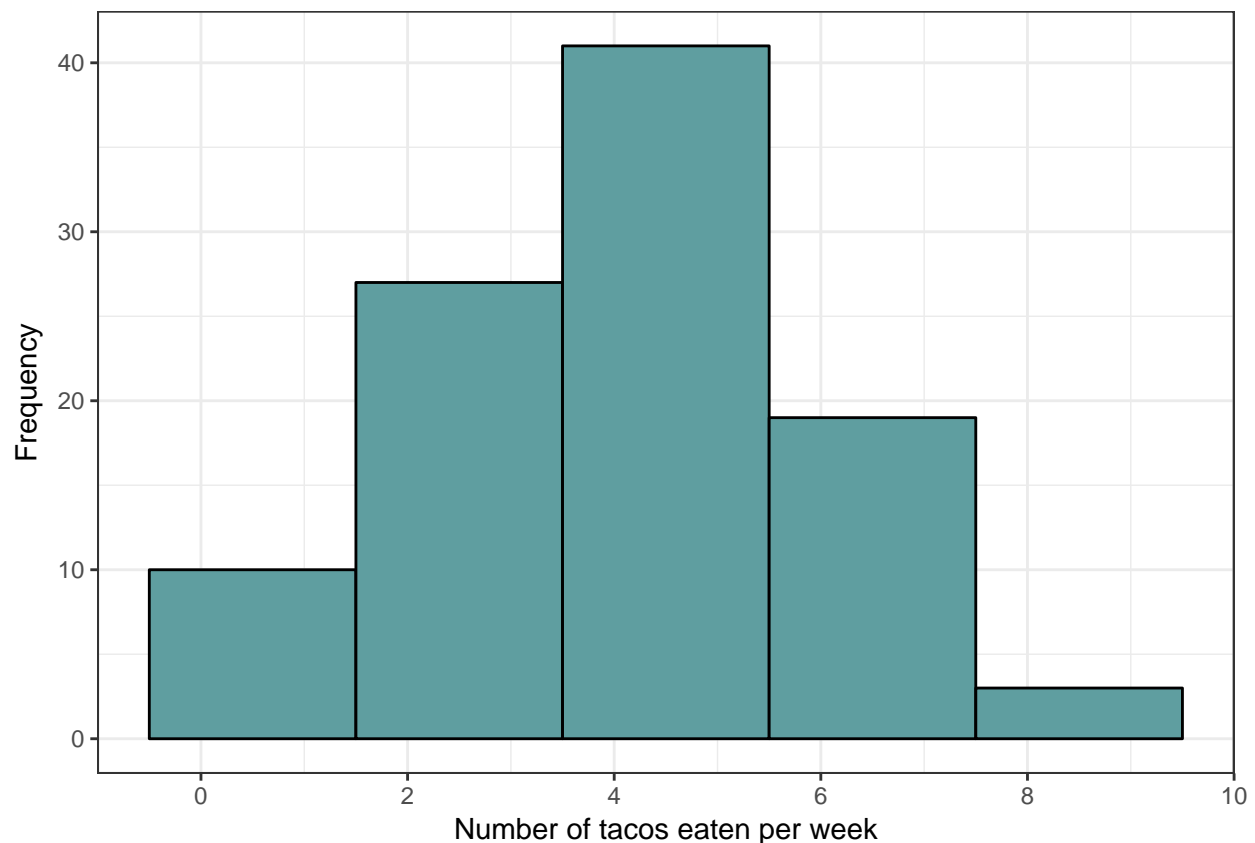
```
ggsave("../wk04_taco_hist.png", width=3.25, height=1.75, dev=png(bg = "transparent", type="Xlib"))

tacos.cut <- cut(tacos.df$tacos, breaks=tacos.brk)

xtable(table(tacos.cut))
```

```
% latex table generated in R 3.4.4 by xtable 1.8-2 package
% Sun Feb 10 18:09:18 2019
\begin{table}[ht]
\centering
\begin{tabular}{rr}
  \hline
 & tacos.cut \\
  \hline
(-0.5,1.5] &  10 \\
  (1.5,3.5] &  27 \\
  (3.5,5.5] &  41 \\
  (5.5,7.5] &  19 \\
  (7.5,9.5] &   3 \\
   \hline
\end{tabular}
\end{table}
```
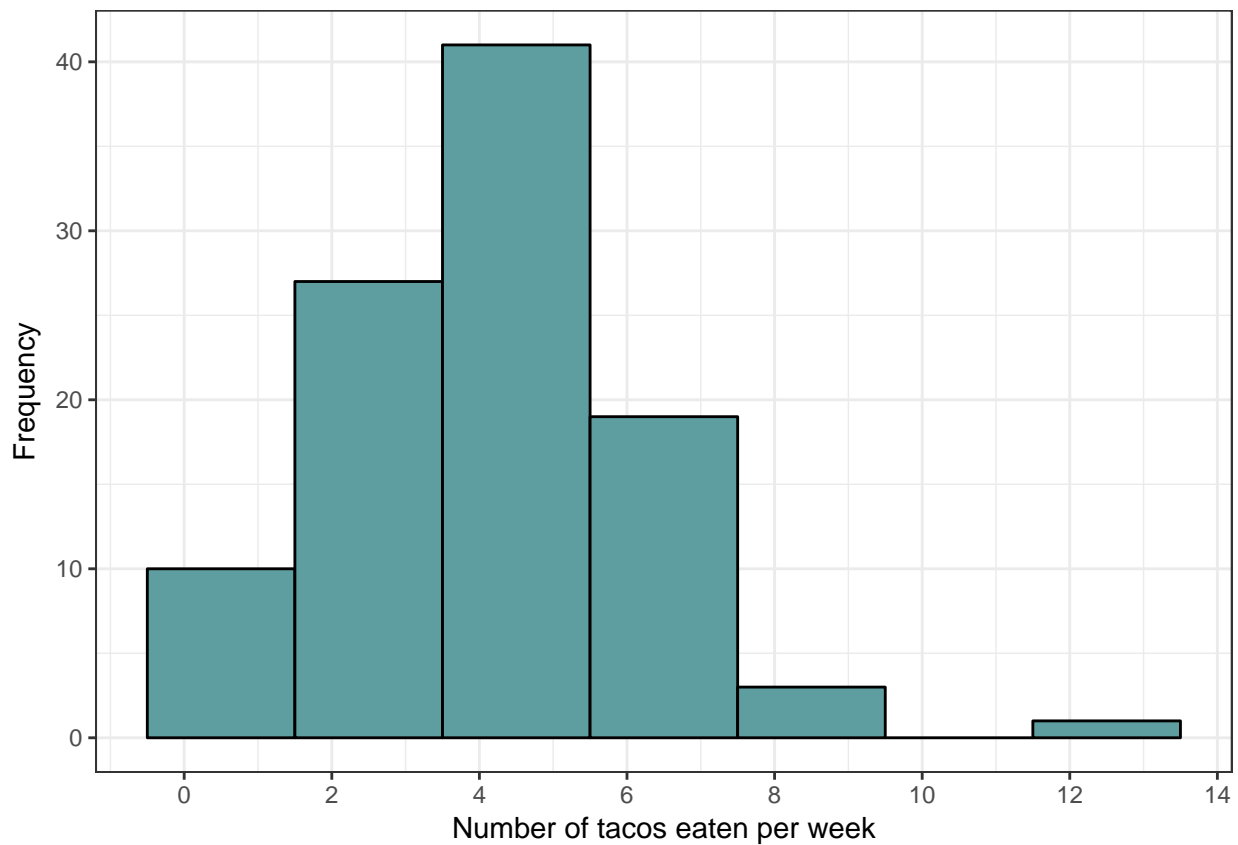
Outliers

```
tacos.out.df <- rbind(tacos.df, list(tacos=13))
write.csv(tacos.out.df, "../data/wk04_tacos_out.csv", row.names = FALSE)

tacos.out.brk <- c((0:7)*2-0.5)
g <- ggplot(tacos.out.df, aes(x=tacos))
g <- g + geom_histogram(breaks=tacos.out.brk, color="black", fill="cadetblue")
g <- g + theme_bw() + scale_x_continuous(breaks=(0:7)*2)
g <- g + labs(x="Number of tacos eaten per week",
              y="Frequency")


g
```

```
ggsave("../wk04_taco_out_hist.png", width=3.25, height=1.75)
```

```
tacos.out.cut <- cut(tacos.out.df$tacos, breaks=tacos.out.brk)
```

```
xtable(table(tacos.out.cut))
```

```
% latex table generated in R 3.4.4 by xtable 1.8-2 package
% Sun Feb 10 18:09:19 2019
\begin{table}[ht]
\centering
\begin{tabular}{rr}
  \hline
 & tacos.out.cut \\
  \hline
(-0.5,1.5] &  10 \\
  (1.5,3.5] &  27 \\
  (3.5,5.5] &  41 \\
  (5.5,7.5] &  19 \\
  (7.5,9.5] &   3 \\
  (9.5,11.5] &   0 \\
  (11.5,13.5] &   1 \\
   \hline
\end{tabular}
\end{table}
```
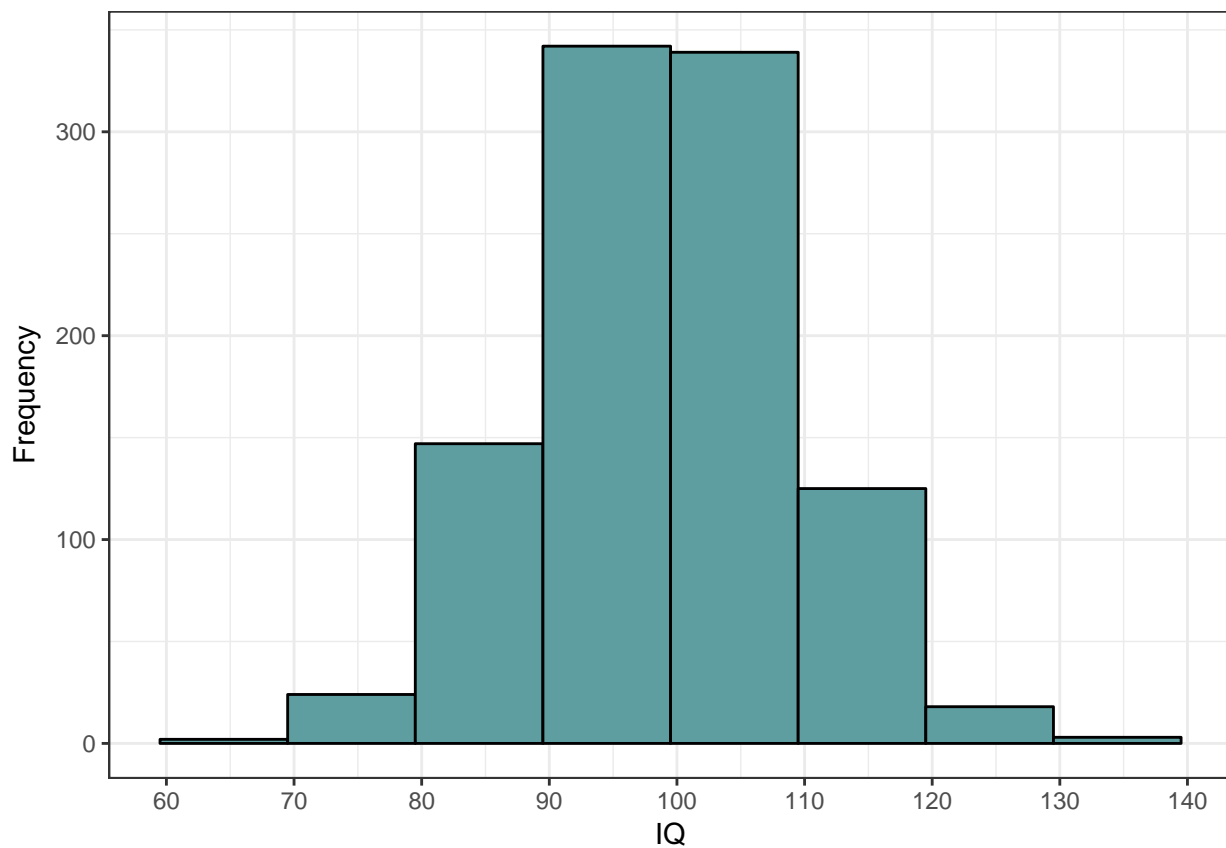
Normal and skewed distributions

Normal

```
tmp <- rnorm(50)
iqs.df <- data.frame(iq = floor(rnorm(1000, 100, 10)))
write.csv(iqs.df, "../data/wk04_iqs.csv", row.names = FALSE)

iqs.brk <- c((6:14)*10-0.5)
g <- ggplot(iqs.df, aes(x=iq))
g <- g + geom_histogram(breaks=iqs.brk, color="black", fill="cadetblue")
g <- g + theme_bw() + MyTheme_transparent
g <- g + scale_x_continuous(breaks=(6:14)*10)
g <- g + labs(x="IQ",
              y="Frequency")

g
```



```
ggsave("../wk04_iq_hist.png", width=3.25, height=1.75, dev=png(bg = "transparent", type="Xlib"))
```

```
iqs.cut <- cut(iqs.df$iq, breaks=iqs.brk)
```

```
xtable(table(iqs.cut))
```

```
% latex table generated in R 3.4.4 by xtable 1.8-2 package
% Sun Feb 10 18:09:20 2019
\begin{table}[ht]
```

```
\centering
\begin{tabular}{rr}
  \hline
 & iqs.cut \\
  \hline
(59.5,69.5] &   2 \\
  (69.5,79.5] &  24 \\
  (79.5,89.5] & 147 \\
  (89.5,99.5] & 342 \\
  (99.5,110] & 339 \\
  (110,120] & 125 \\
  (120,130] &  18 \\
  (130,140] &   3 \\
   \hline
\end{tabular}
\end{table}
```

Right skew

```
tmp <- rnorm(100)
iqs.right.df <- data.frame(iq = rnorm(1000, 95, 10))
iqs.right.max <- max(iqs.right.df$iq)
iqs.right.mean <- mean(iqs.right.df$iq)

iqs.right.df$iq <- floor(iqs.right.df$iq*ifelse(iqs.right.df$iq<= iqs.right.mean, 1,
                                  1 + 0.3*(iqs.right.df$iq-iqs.right.mean)/(iqs.right.max-iqs.r

write.csv(iqs.right.df, "../data/wk04_iqs_right.csv", row.names = FALSE)

range(iqs.right.df$iq)
```
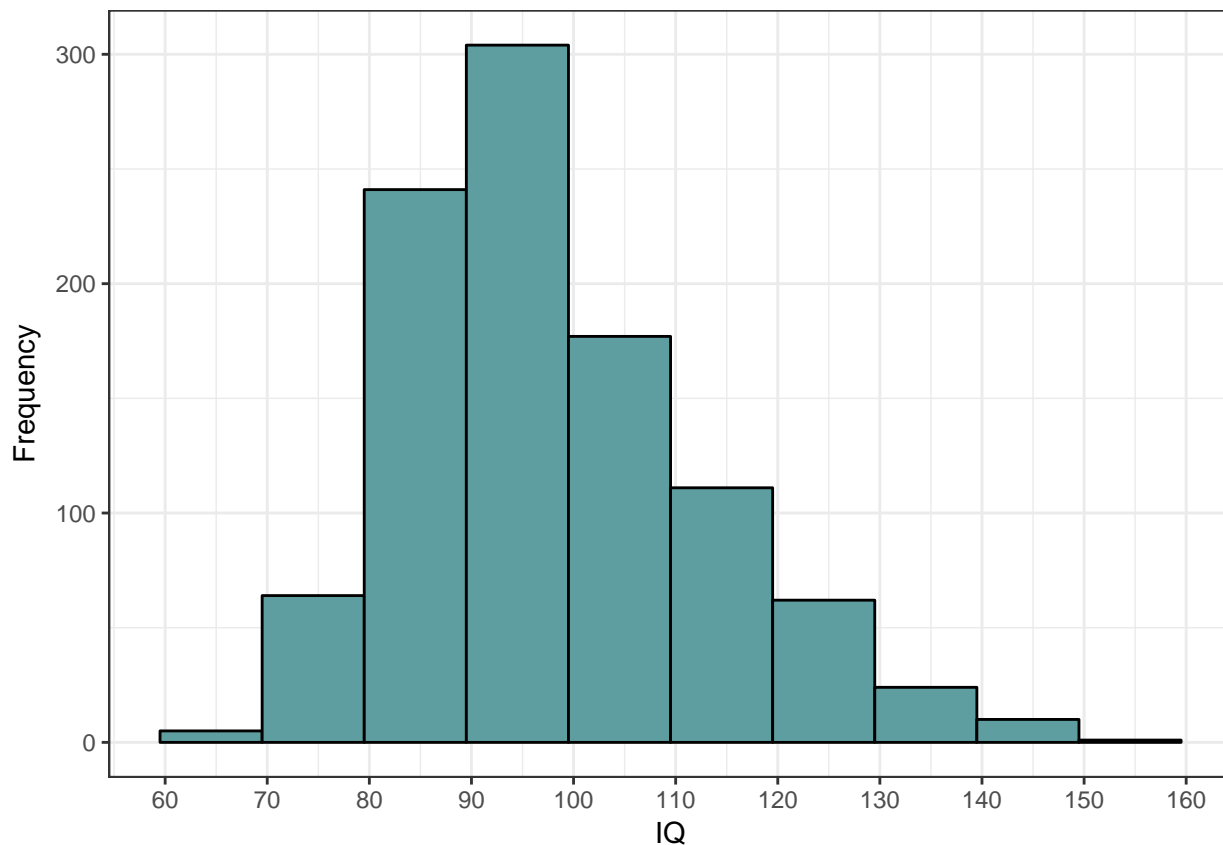
```
## [1]  66 170
```

```
iqs.right.brk <- c((6:16)*10-0.5)
g <- ggplot(iqs.right.df, aes(x=iq))
g <- g + geom_histogram(breaks=iqs.right.brk, color="black", fill="cadetblue")
g <- g + theme_bw() + MyTheme_transparent
g <- g + scale_x_continuous(breaks=iqs.right.brk + 0.5)
g <- g + labs(x="IQ",
              y="Frequency")

g
```

```
ggsave("../wk04_iq_right_hist.png", width=3.25, height=1.75, dev=png(bg = "transparent", type="Xlib"))
```

```
iqs.right.cut <- cut(iqs.right.df$iq, breaks=iqs.right.brk)
```

```
xtable(table(iqs.right.cut))
```

```
% latex table generated in R 3.4.4 by xtable 1.8-2 package
% Sun Feb 10 18:09:20 2019
\begin{table}[ht]
\centering
\begin{tabular}{rr}
  \hline
 & iqs.right.cut \\
  \hline
(59.5,69.5] &   5 \\
  (69.5,79.5] &  64 \\
  (79.5,89.5] & 241 \\
  (89.5,99.5] & 304 \\
  (99.5,110] & 177 \\
  (110,120] & 111 \\
  (120,130] &  62 \\
  (130,140] &  24 \\
  (140,150] &  10 \\
  (150,160] &   1 \\
   \hline
\end{tabular}
\end{table}
```

Left skew

```r
tmp <- rnorm(100)
iqs.left.df <- data.frame(iq = rnorm(1000, 95, 10))
iqs.left.min <- min(iqs.left.df$iq)
iqs.left.mean <- mean(iqs.left.df$iq)

iqs.left.df$iq <- floor(iqs.left.df$iq*ifelse(iqs.left.df$iq >= iqs.left.mean, 1,
                                  1 - 0.3*(iqs.left.mean-iqs.left.df$iq)/(iqs.left.mean-iqs.left

write.csv(iqs.left.df, "../data/wk04_iqs_left.csv", row.names = FALSE)

range(iqs.left.df$iq)
```
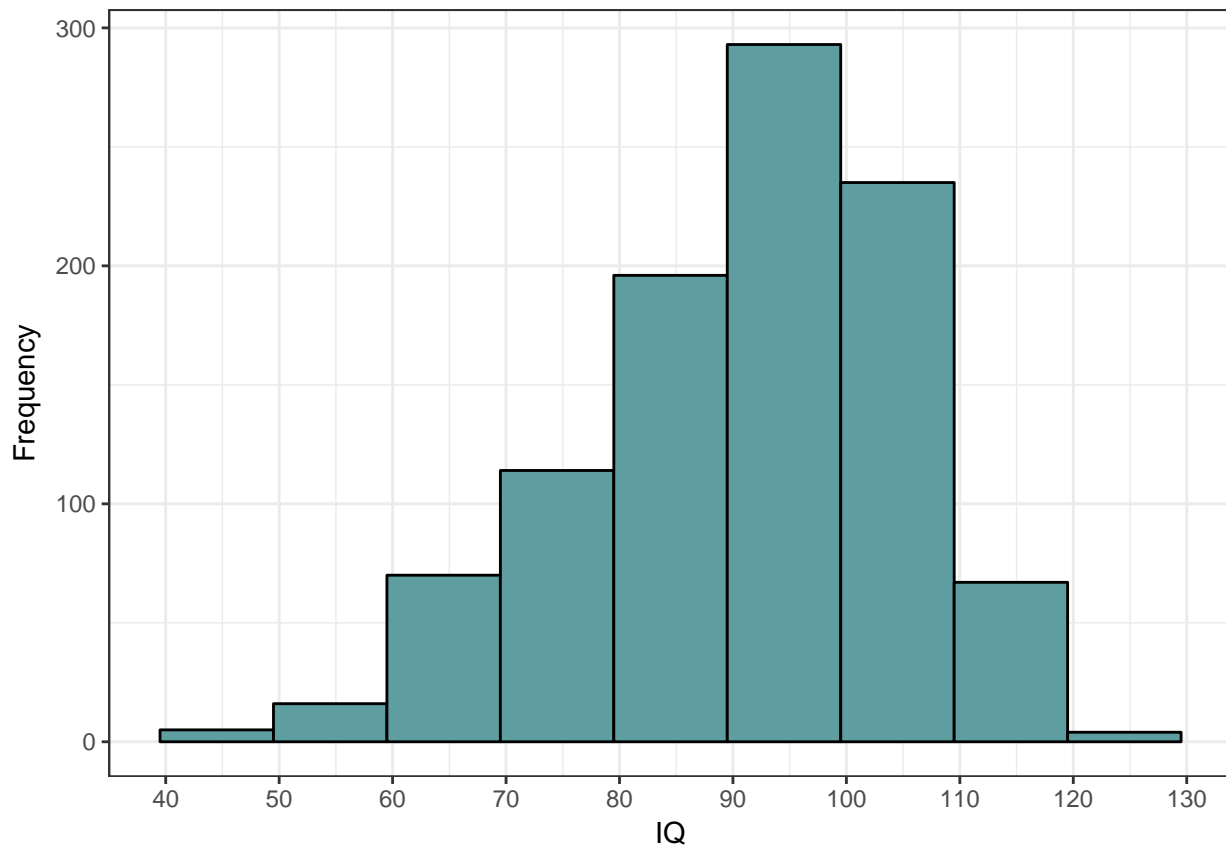
```
## [1]  44 126
```

```r
iqs.left.brk <- c((4:13)*10-0.5)
g <- ggplot(iqs.left.df, aes(x=iq))
g <- g + geom_histogram(breaks=iqs.left.brk, color="black", fill="cadetblue")
g <- g + theme_bw() + MyTheme_transparent
g <- g + scale_x_continuous(breaks=iqs.left.brk + 0.5)
g <- g + labs(x="IQ",
              y="Frequency")

g
```

```
ggsave("../wk04_iq_left_hist.png", width=3.25, height=1.75, dev=png(bg = "transparent", type="Xlib"))

iqs.left.cut <- cut(iqs.left.df$iq, breaks=iqs.left.brk)

xtable(table(iqs.left.cut))
```

```
% latex table generated in R 3.4.4 by xtable 1.8-2 package
% Sun Feb 10 18:09:21 2019
\begin{table}[ht]
\centering
\begin{tabular}{rr}
  \hline
 & iqs.left.cut \\
  \hline
(39.5,49.5] &    5 \\
  (49.5,59.5] &   16 \\
  (59.5,69.5] &   70 \\
  (69.5,79.5] &  114 \\
  (79.5,89.5] &  196 \\
  (89.5,99.5] &  293 \\
  (99.5,110] &  235 \\
  (110,120] &   67 \\
  (120,130] &    4 \\
   \hline
\end{tabular}
\end{table}
```

Bimodal distribution

```
tacos.bi.df <- data.frame(tacos = c(round(rnorm(50, 4.5, 1.5)),
                                    round(rnorm(50, 13.5, 1.5))))
tacos.bi.df$tacos[tacos.bi.df$tacos<0] <- 0
write.csv(tacos.df, "../data/wk04_tacos_bi.csv", row.names = FALSE)
range(tacos.bi.df$tacos)
```
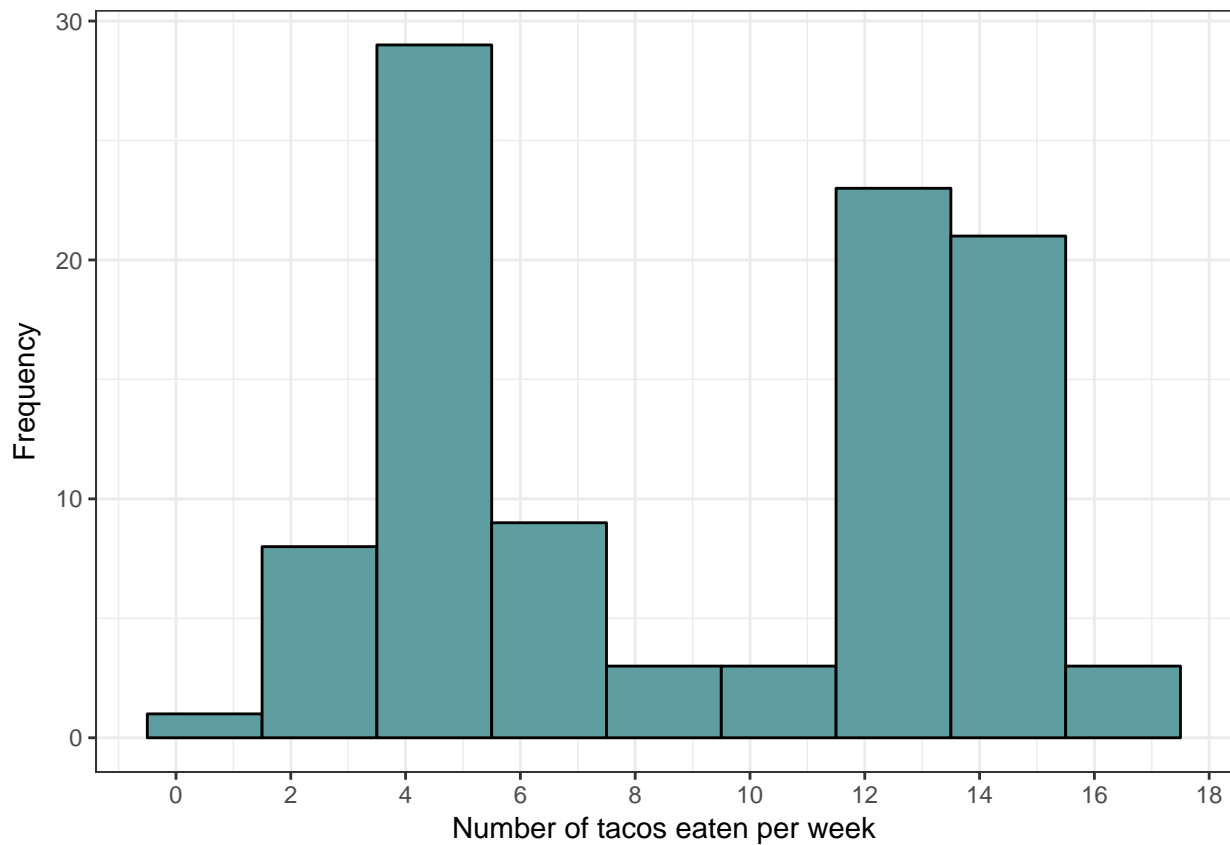
```
## [1]  1 17
```

```
tacos.bi.brk <- c((0:9)*2-0.5)
g <- ggplot(tacos.bi.df, aes(x=tacos))
g <- g + geom_histogram(breaks=tacos.bi.brk, color="black", fill="cadetblue")
g <- g + theme_bw() + MyTheme_transparent
g <- g + scale_x_continuous(breaks=tacos.bi.brk+0.5)
g <- g + labs(x="Number of tacos eaten per week",
              y="Frequency")


g
```

```
ggsave("../wk04_taco_bi_hist.png", width=3.25, height=1.75, dev=png(bg = "transparent", type="Xlib"))
```

```
tacos.bi.cut <- cut(tacos.bi.df$tacos, breaks=tacos.bi.brk)
```

```
xtable(table(tacos.bi.cut))
```

```
% latex table generated in R 3.4.4 by xtable 1.8-2 package
% Sun Feb 10 18:09:21 2019
\begin{table}[ht]
\centering
\begin{tabular}{rr}
  \hline
 & tacos.bi.cut \\
  \hline
(-0.5,1.5] &    1 \\
  (1.5,3.5] &    8 \\
  (3.5,5.5] &   29 \\
  (5.5,7.5] &    9 \\
  (7.5,9.5] &    3 \\
  (9.5,11.5] &    3 \\
  (11.5,13.5] &   23 \\
  (13.5,15.5] &   21 \\
  (15.5,17.5] &    3 \\
   \hline
\end{tabular}
\end{table}
```