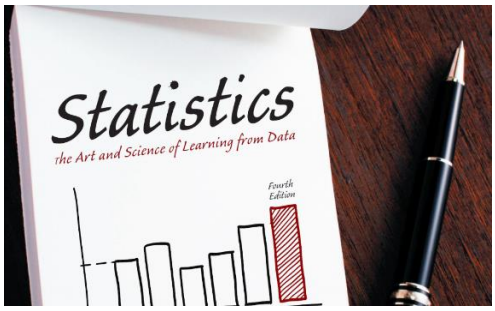
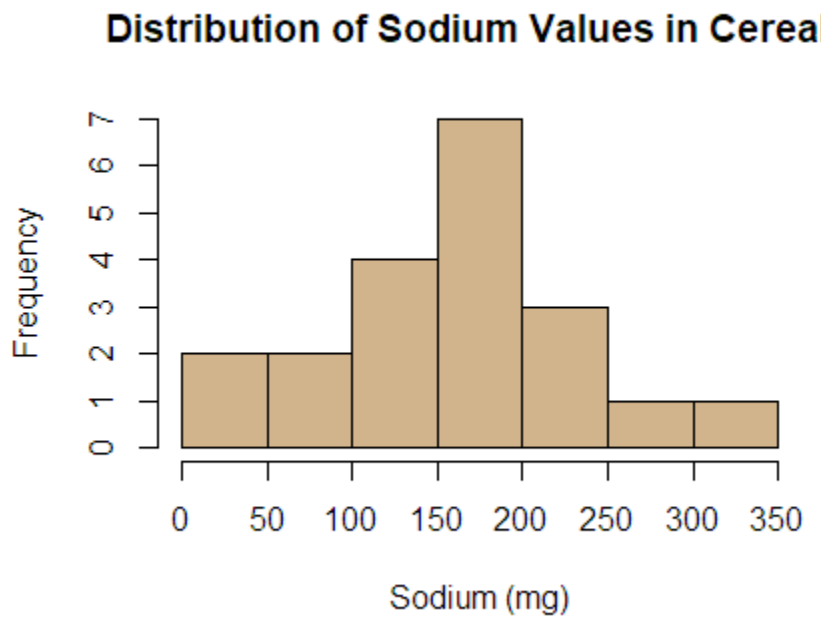


STATISTICS
THE ART & SCIENCE OF LEARNING FROM DATA
AGRESTI · FRANKLIN · KLINGENBERG

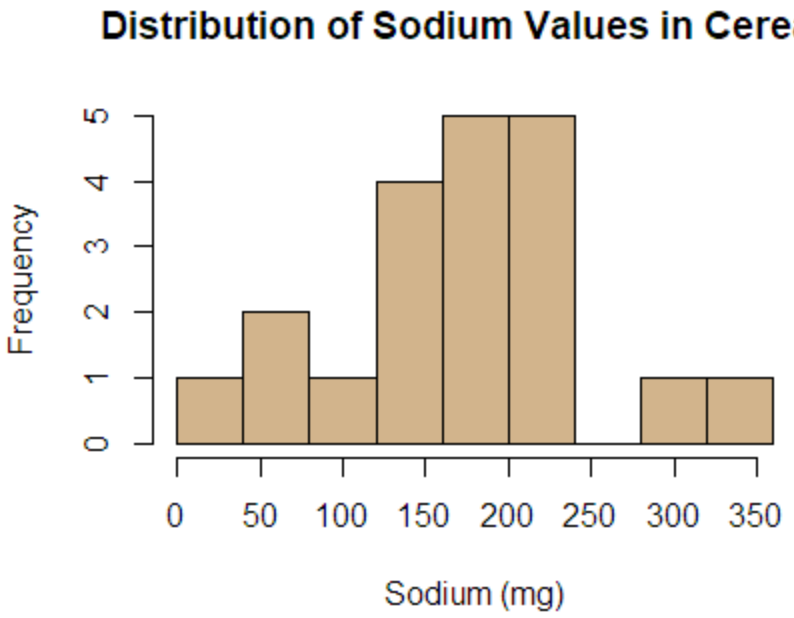
Chapter 2

Example 7: Health Value of Cereals – Histograms

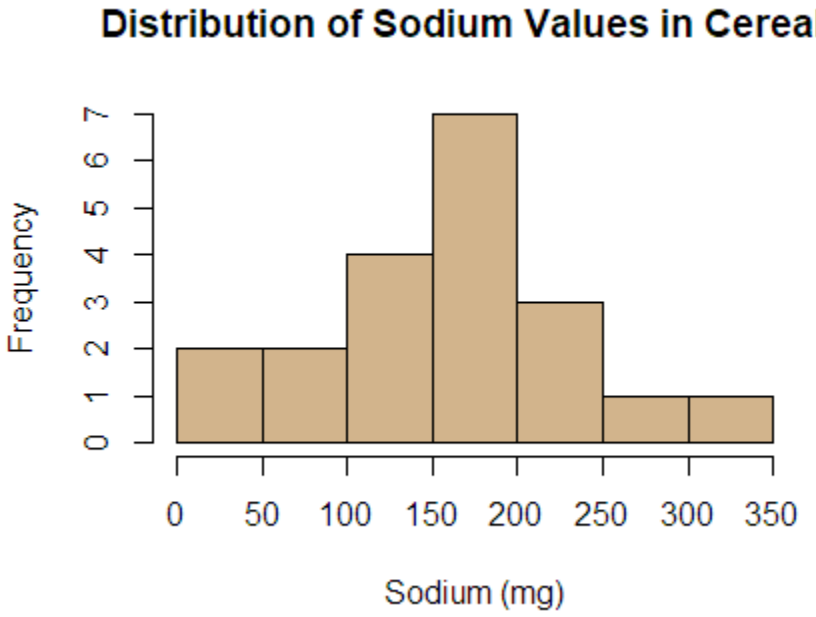
```
> # Read in Sodium values:
> Sodium <- c(0, 340, 70, 140, 200, 180, 210, 150, 100, 130, 140, 180, 190, 160,
290, 50, 220, 180, 200, 210)

> # Create Basic Histogram:
> hist(Sodium, xlab="Sodium (mg)", ylab="Frequency", main="Distribution of Sodiu
m Values in Cereals", col="tan")
```



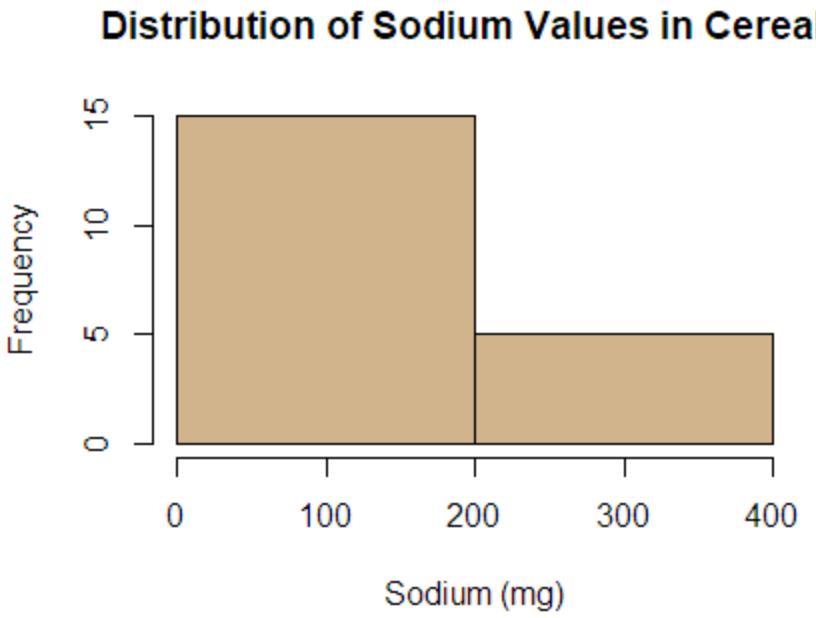

```
> # Changing the bins by providing the boundaries.
> # (Note: right=FALSE puts an observation such as 120 in the interval from 120-
160 and not 80-120)
> hist(Sodium, breaks=seq(0,360,40), right=FALSE, xlab="Sodium (mg)", ylab="Freq
uency", main="Distribution of Sodium Values in Cereals", col="tan")
```



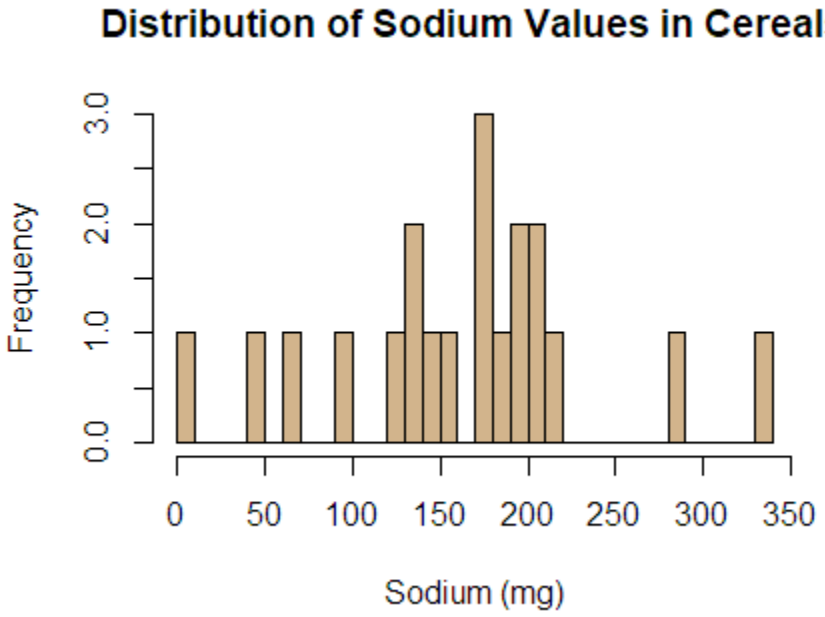

```
> # Another way to request a certain number of bins:
> hist(Sodium, breaks=10, xlab="Sodium (mg)", ylab="Frequency", main="Distributi
on of Sodium Values in Cereals", col="tan")
```



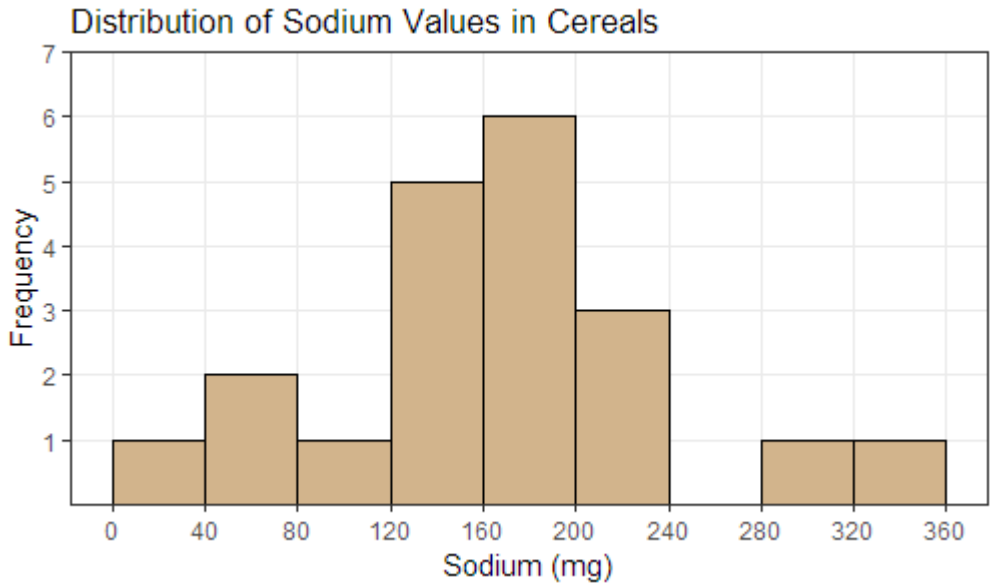

```
> # Too few breaks:
> hist(Sodium, breaks=2, xlab="Sodium (mg)", ylab="Frequency", main="Distributio
n of Sodium Values in Cereals", col="tan")
```



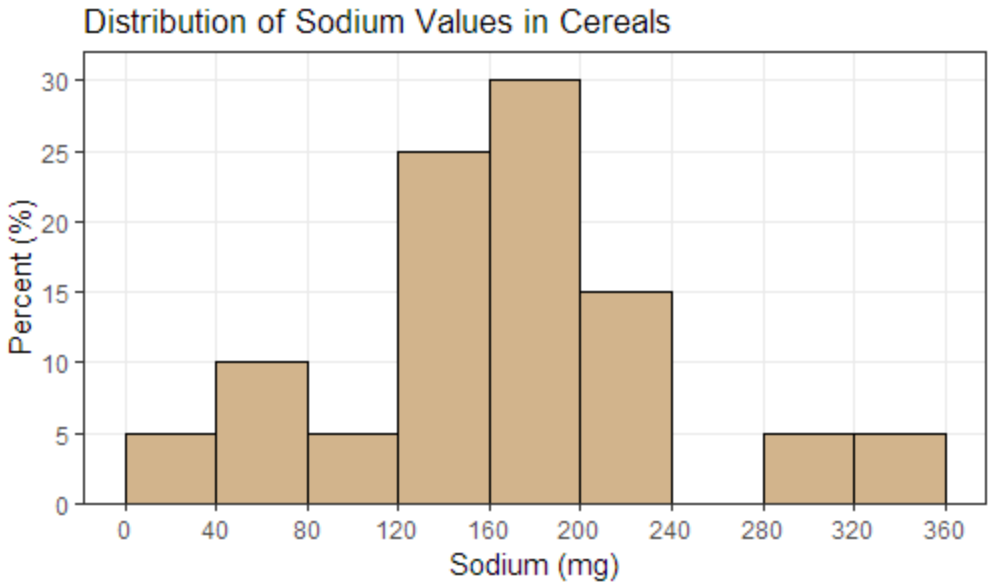

```
> # Too many breaks:
> hist(Sodium, breaks=30, xlab="Sodium (mg)", ylab="Frequency", main="Distributi
on of Sodium Values in Cereals", col="tan")
```



```
> # 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_x_continuous(breaks=seq(0,360,40)) +
+   scale_y_continuous(limit=c(0,7), breaks=1:7, expand=c(0,0)) +
+   theme_bw() +
+   theme(panel.grid.minor=element_blank())
```




```
> # 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_x_continuous(breaks=seq(0,360,40)) +
+   scale_y_continuous(limit=c(0,32), breaks=seq(0,30,5), expand=c(0,0)) +
+   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_x_continuous(breaks=seq(0,360,40)) +
+   scale_y_continuous(limit=c(0,27), breaks=seq(0,25,5), expand=c(0,0)) +
+   theme_classic() +
+   theme(panel.grid.minor=element_blank())
```