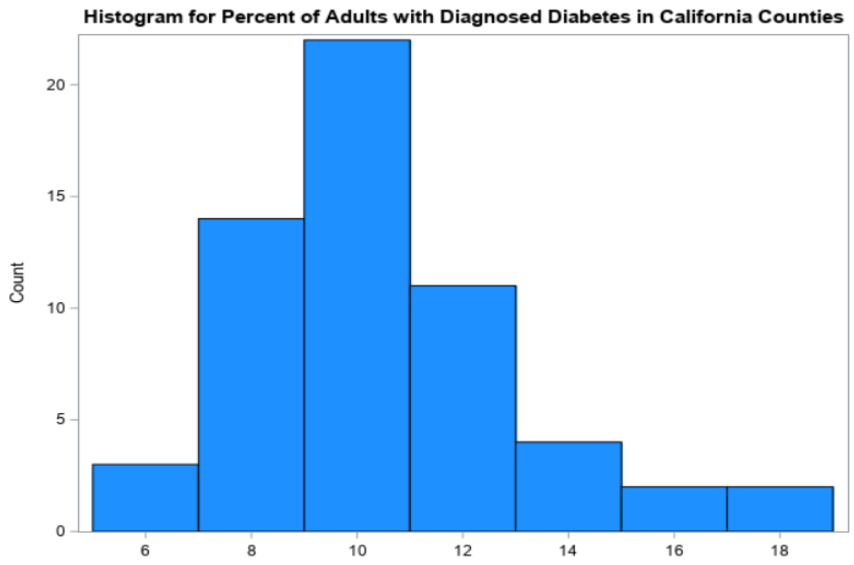
3.2 Histogram

The distribution of values of a continuous random variable may be depicted in the form of a **histogram**. The range of values is subdivided into adjacent subintervals of equal lengths, and bars are constructed above each subinterval, which heights represent proportion of observations that lie inside the corresponding subintervals.
The shape of the distribution seen on a histogram is traditionally described as **unimodal** (one peak), **bimodal** (two peaks), or **multimodal** (three or more peaks. Also, unimodal shapes are also described as **symmetric** (roughly identical on both sides of a line of symmetry), **right-skewed** (having a long right tail), or **left-skewed** (having a long left tail).

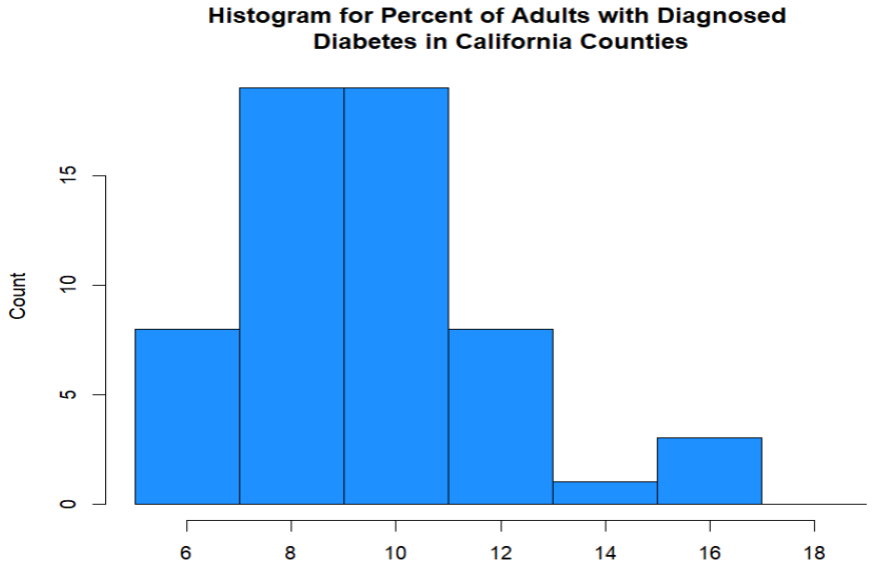
Example. In this example, the data represent percentage of adults aged 20 and above with diagnosed diabetes in 2017 by county in California. The codes below plot histograms for the data. In SAS:

```
proc import out=diabetes2 datafile="./diabetes_data2.csv" dbms=csv replace;
run;
title 'Histogram for Percent of Adults with Diagnosed Diabetes
in California Counties';
proc sgplot data=diabetes2;
 histogram percent/ scale=count binstart=6 binwidth=2
  showbins fillattrs=(color=dodgerblue);
   xaxis display=(nolabel);
run;
```



The histogram is unimodal and right-skewed. In R:

```
diabetes2<- read.csv(file="./diabetes_data2.csv", header=TRUE, sep=",")
hist(diabetes2$percent, main='Histogram for Percent of Adults with Diagnosed
Diabetes in California Counties', breaks=seq(5,19,by=2), col="dodgerblue",
xlab="", ylab="Count")
```



□