Distributions

The term **distribution** refers to the manner in which some characteristic (such as height) is *distributed* in a given population.

Distributions

The term **distribution** refers to the manner in which some characteristic (such as height) is *distributed* in a given population.

A considerable amount of information is required to specify a distribution.

One very efficient way to convey this information is a kind of graph called a **histogram**

Histograms

The horizontal scale on a histogram is divided into ranges called **bins**.

The bins may or may not be of equal width.

Histograms

The horizontal scale on a histogram is divided into ranges called **bins**.

The bins may or may not be of equal width.

Some statisticians prefer to avoid histograms with unequal bin sizes because they can be difficult to construct.
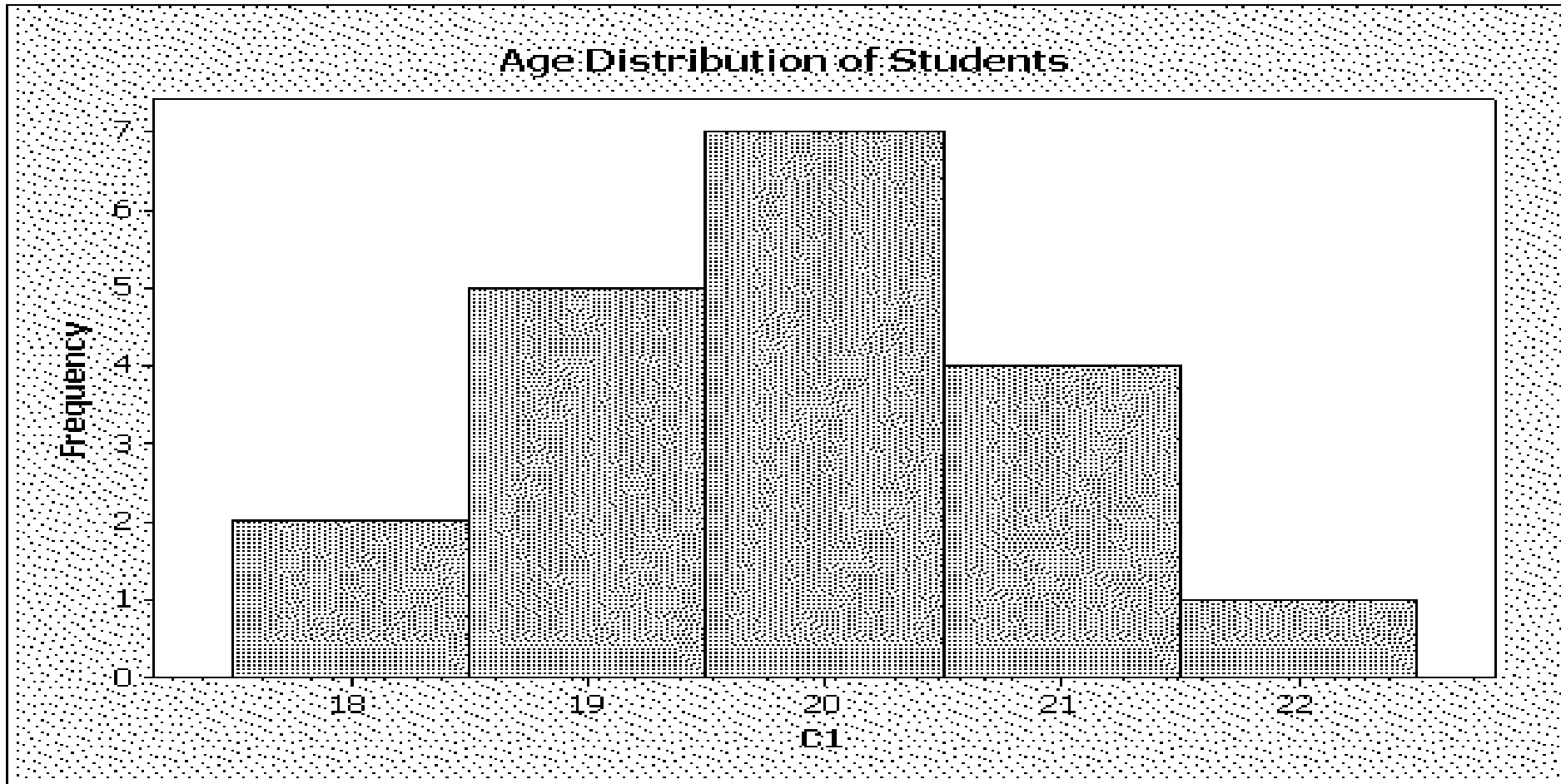
Histograms

There are several varieties of histograms. Perhaps the most common is the *frequency histogram* in which the height of each bar reflects the count of individuals that fall into that bin.

Histograms

There are several varieties of histograms. Perhaps the most common is the *frequency histogram* in which the height of each bar reflects the count of individuals that fall into that bin.

Suppose the distribution of ages in a course with 19 students is:

- 2 students are age 18
- 5 students are age 19
- 7 students are age 20
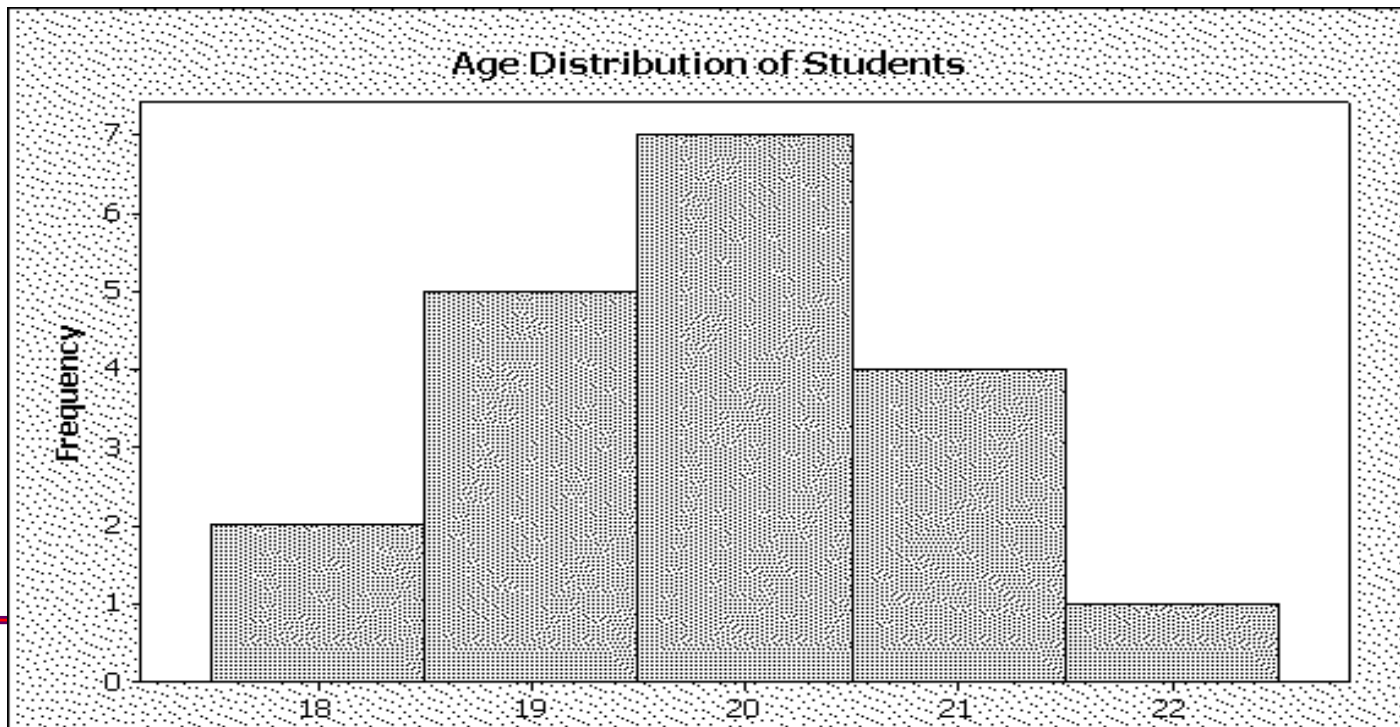- 4 students are age 21
- 1 student is age 22

Histograms

The following frequency histogram depicts the distribution of ages in the class:
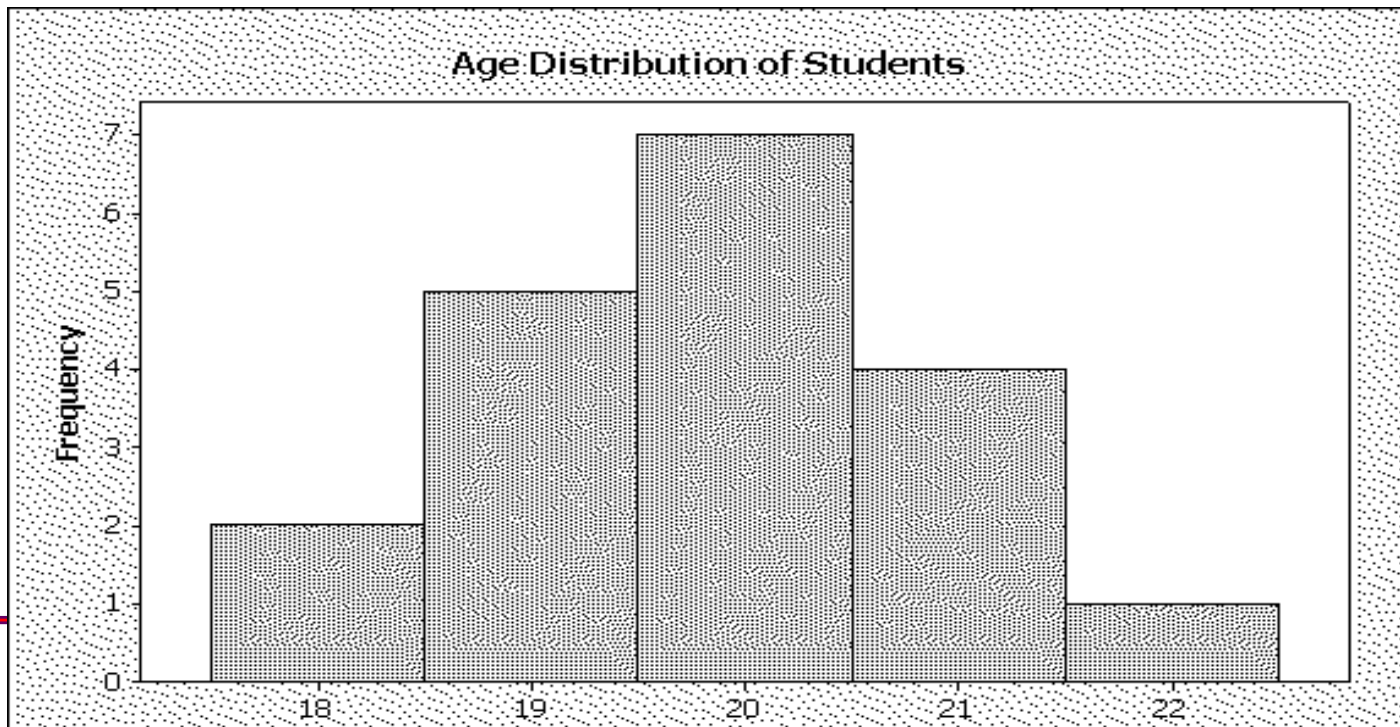
Histograms

This histogram is easy to interpret:

- The bin a student is counted in is determined by the students age.

Age Distribution of Students

Histograms

This histogram is easy to interpret:

- The bin a student is counted in is determined by the students age.

- The height of the rectangle over each bin is the number of students in that bin.

Age Distribution of Students

Histograms

The advantage of a histogram is that it conveys a great deal of information about the distribution of some characteristic (age in this case) over the population at a single glance.
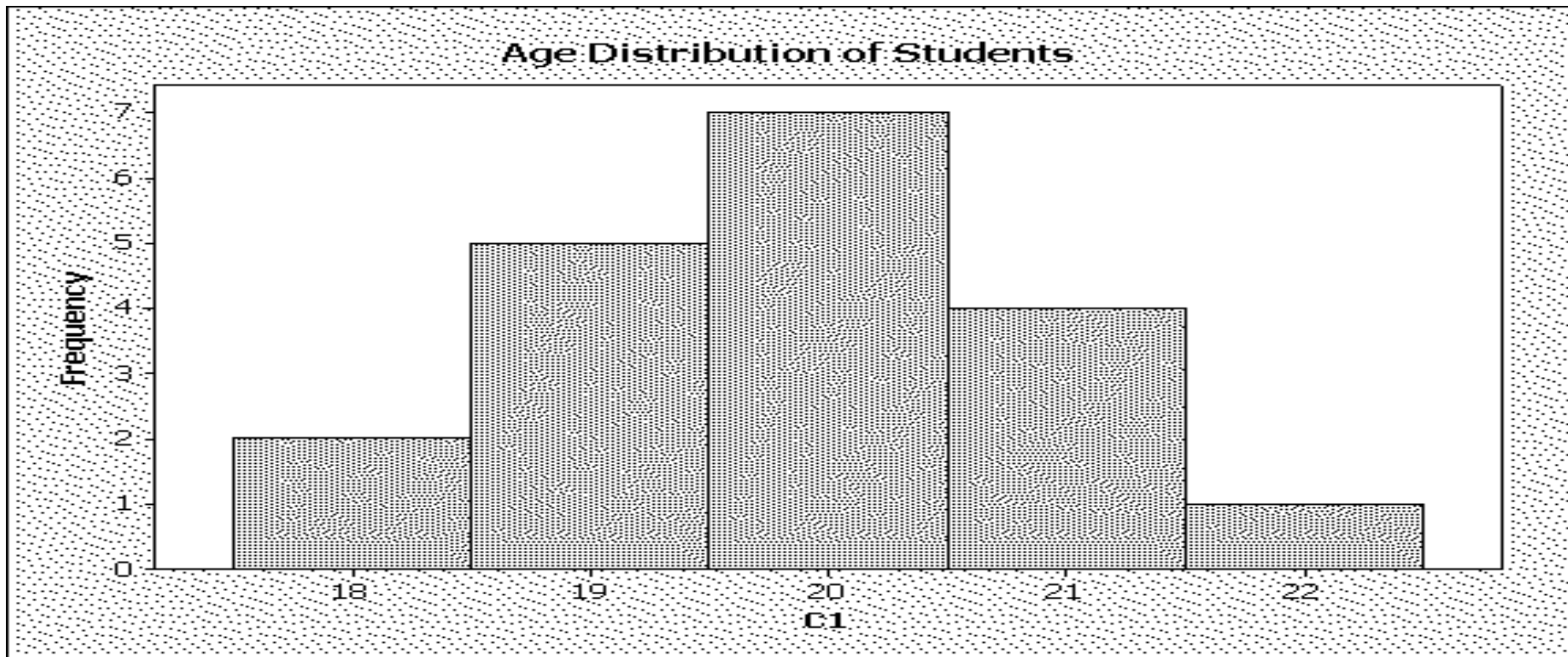
Age Distribution of Students

MINITAB - Types of Histograms

Some software products like MINITAB can produce several kinds of histograms.
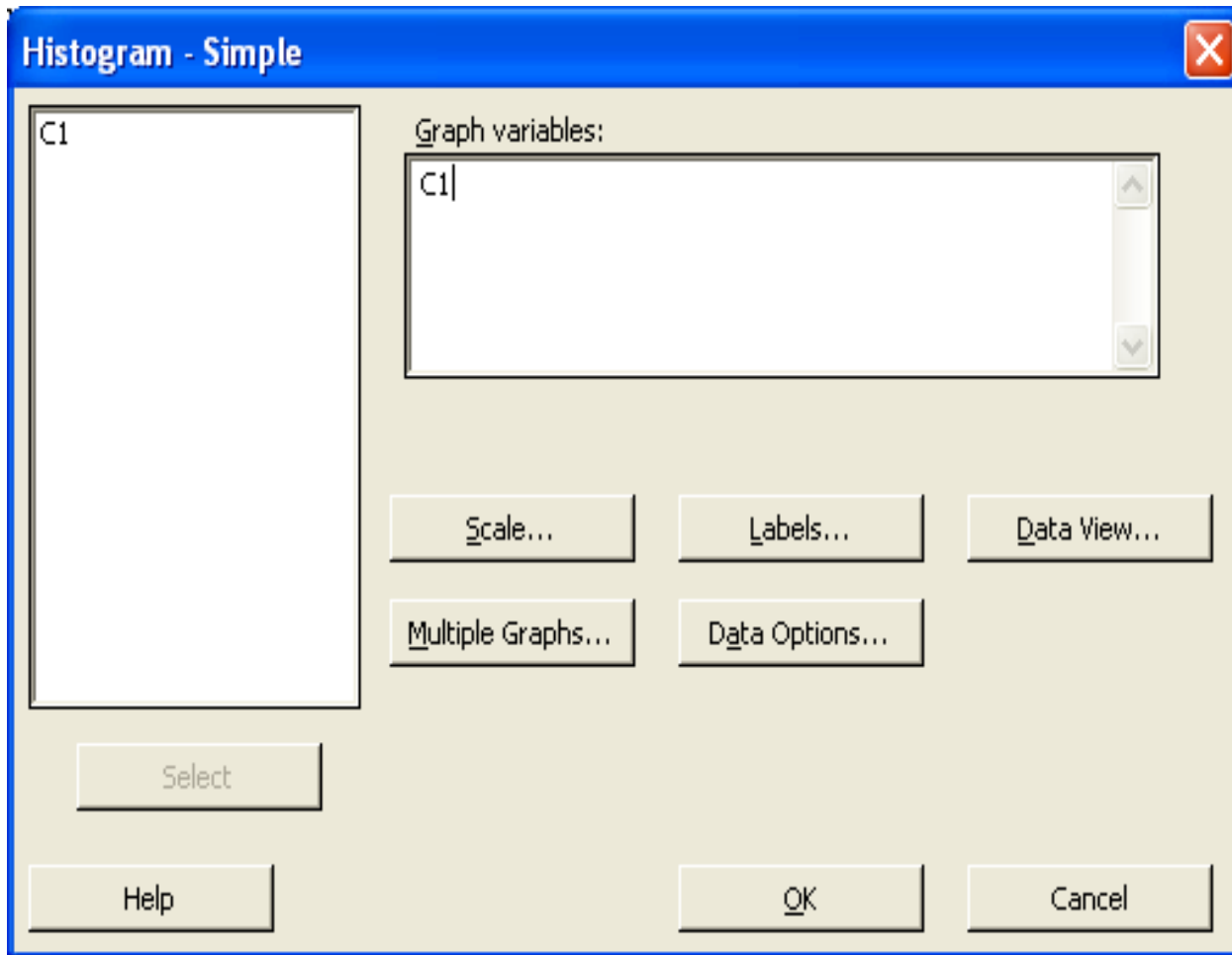
MINITAB - Types of Histograms

Some software products like MINITAB can produce several kinds of histograms.

With MINITAB, the difference is in the way the vertical scale is labelled:

- in a **Frequency** histogram, the numbers on the vertical scale are *counts*

- in a **Percent** histogram, the numbers on the vertical scale are *percentages*

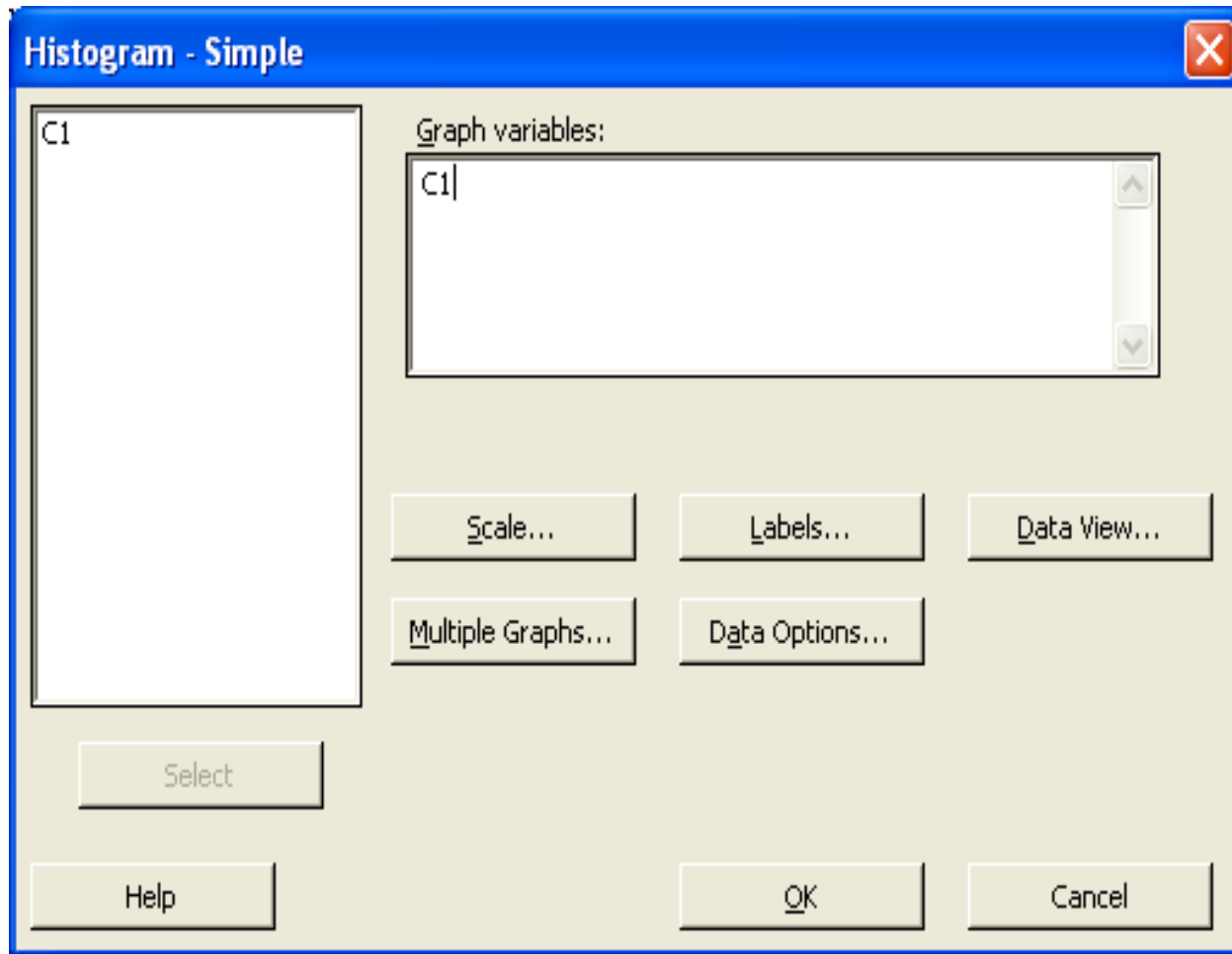- in a **Density** histogram, the numbers on the vertical scale are *proportions*

MINITAB - Types of Histograms

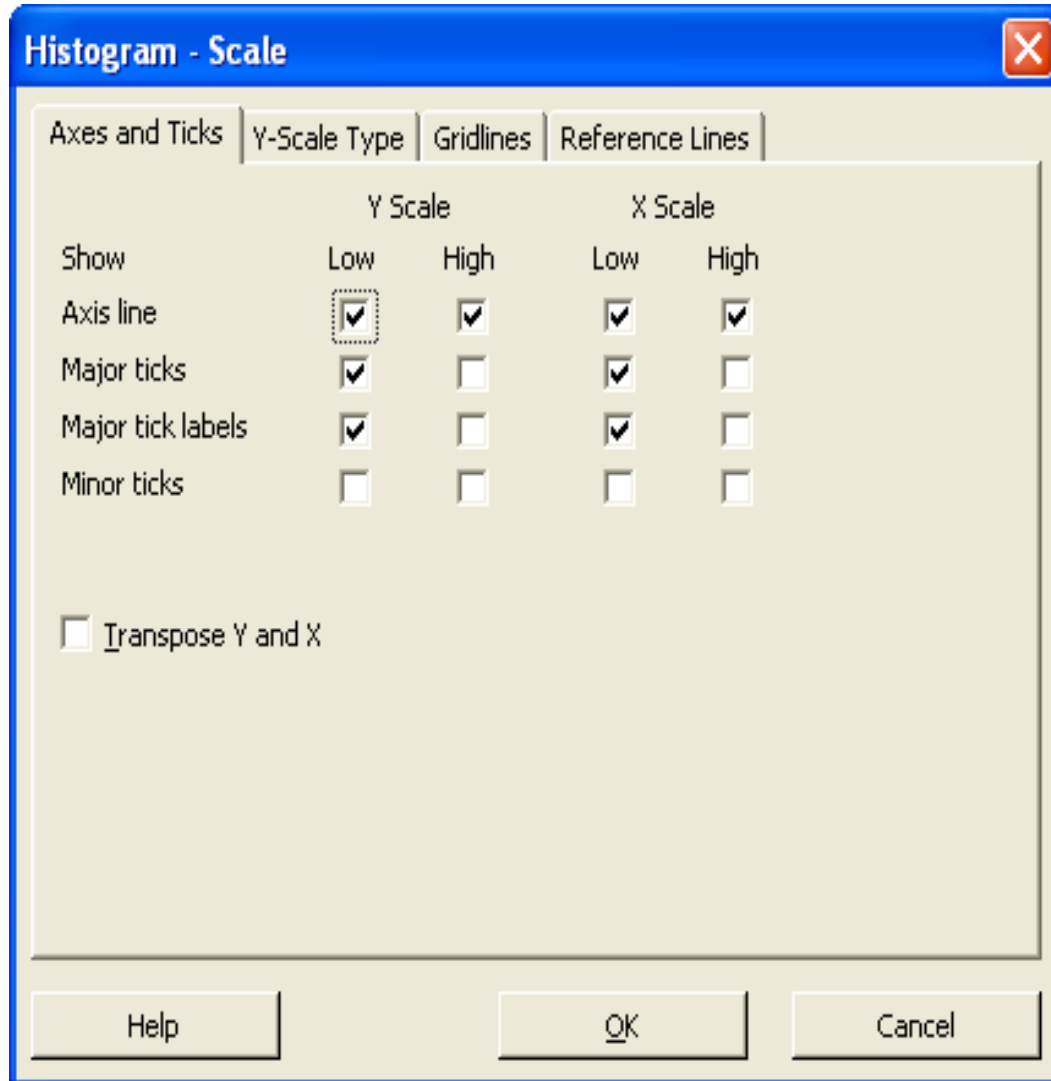The type of histogram produced is determined from the Histogram dialogue box:

MINITAB - Types of Histograms
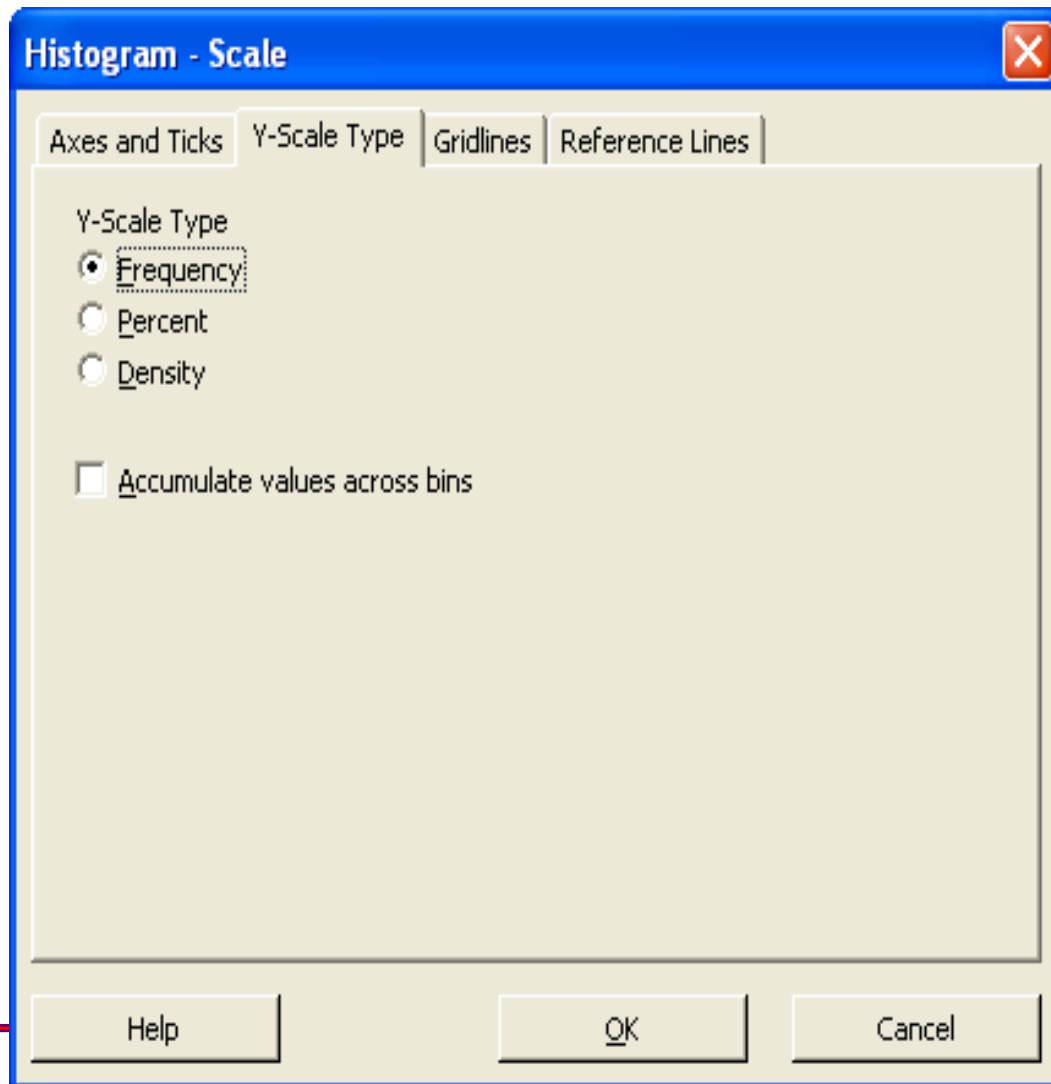
Click on the **S**cale... box.

MINITAB - Types of Histograms

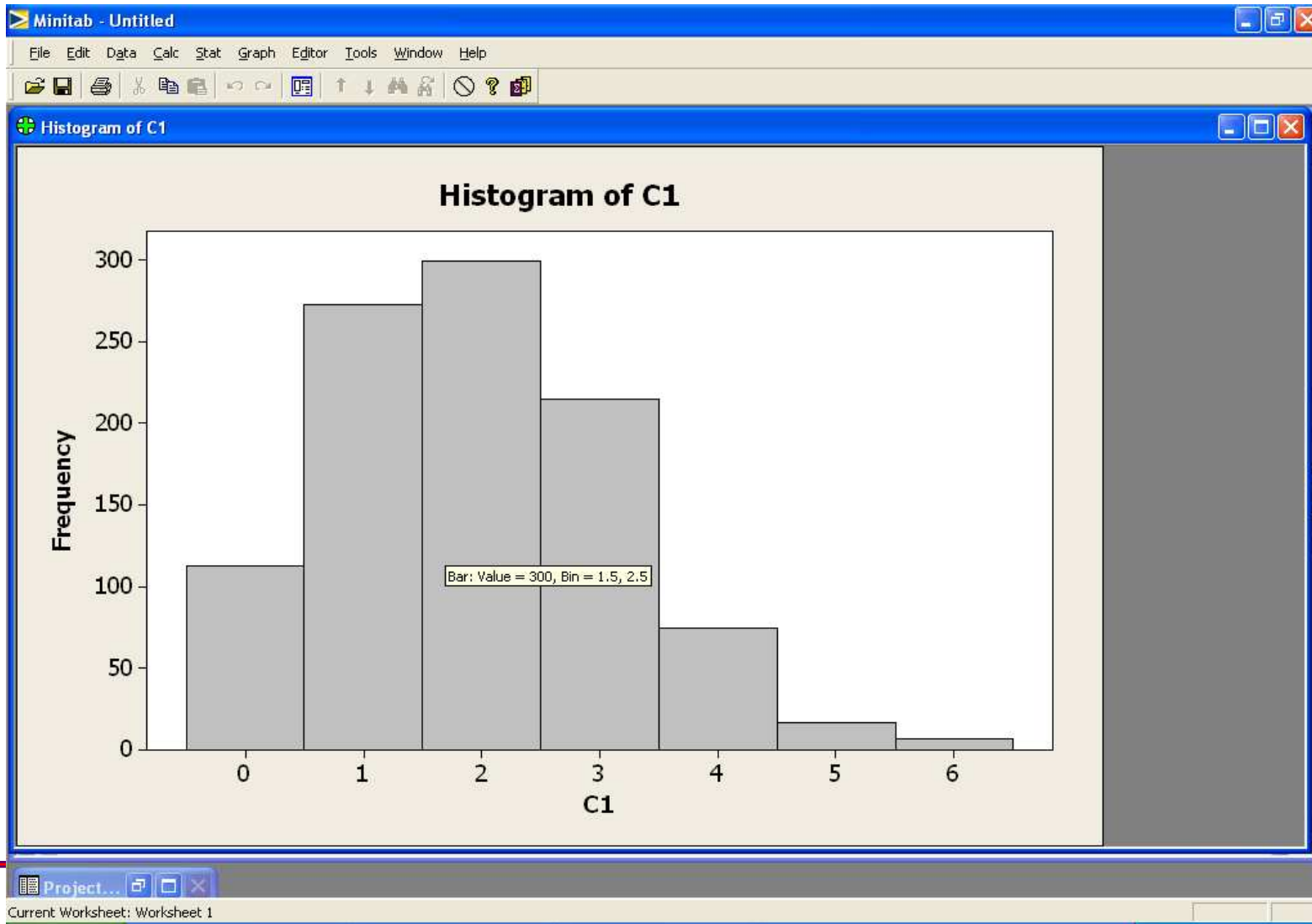The following dialog box should appear:

MINITAB - Types of Histograms

Select the Y-Scale Type tab, and the following dialogue box appears:
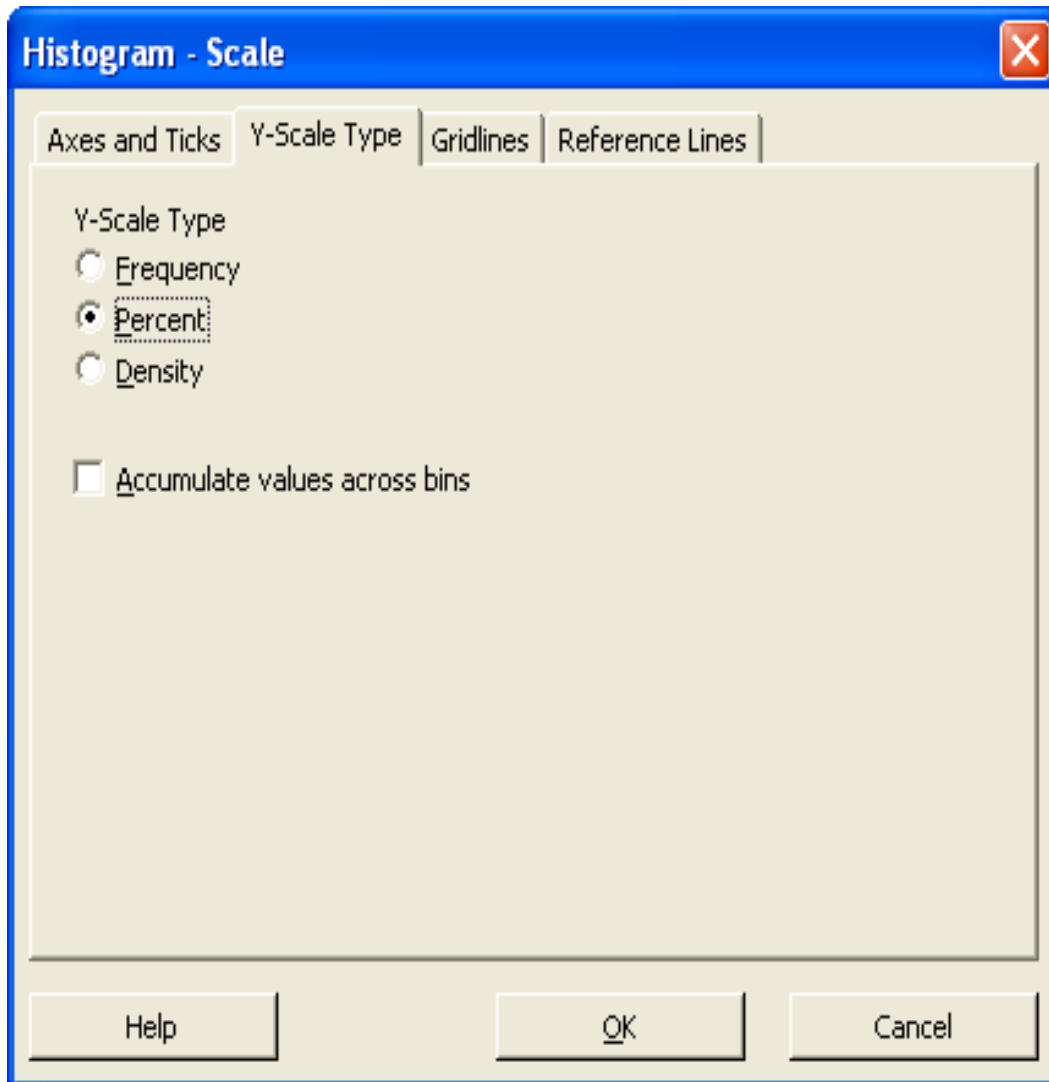
MINITAB - Types of Histograms

In this case the **Frequency** button was chosen, which produces the following histogram:
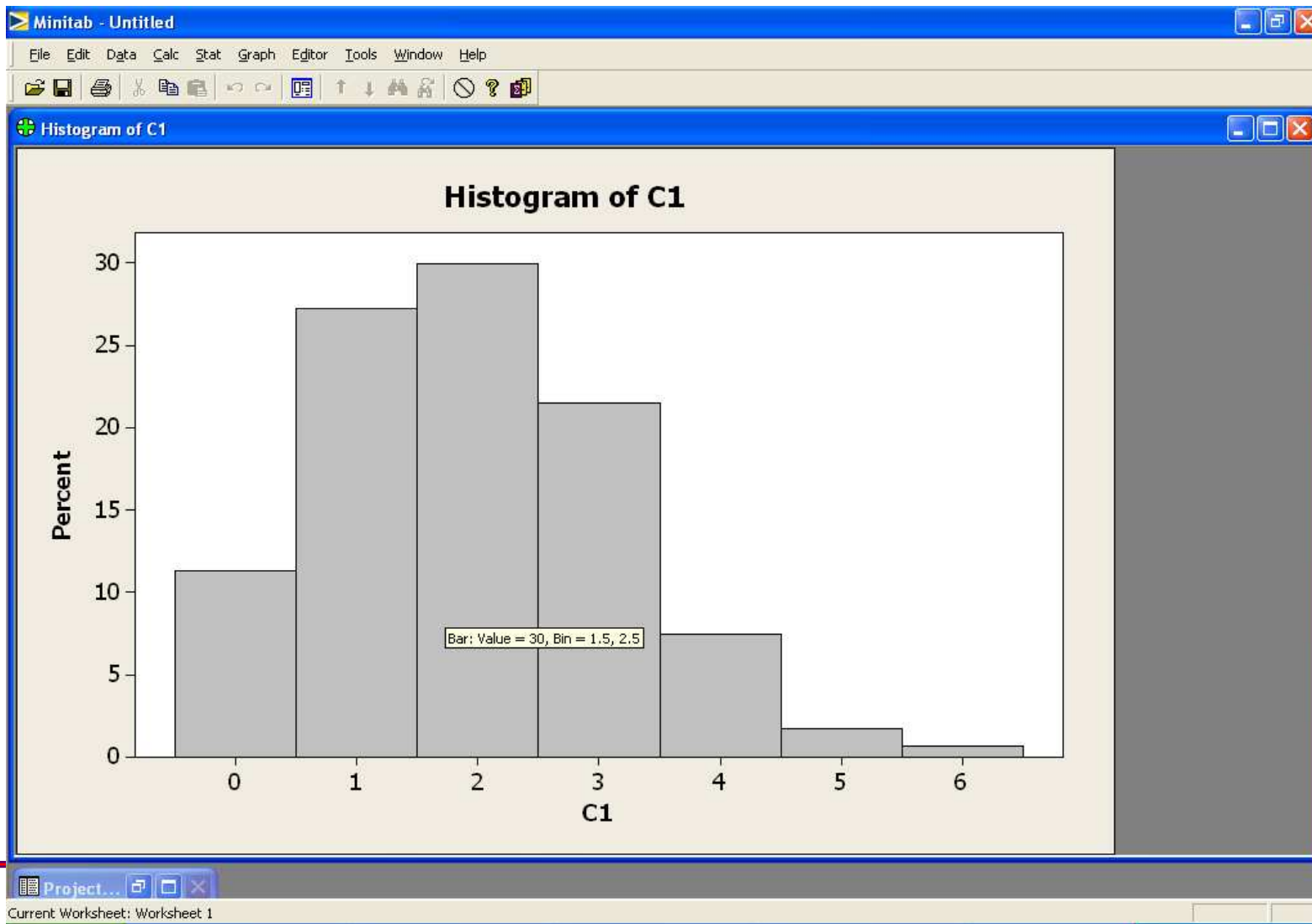
MINITAB - Types of Histograms

For a percentage histogram, select the **Percent** button.

MINITAB - Types of Histograms
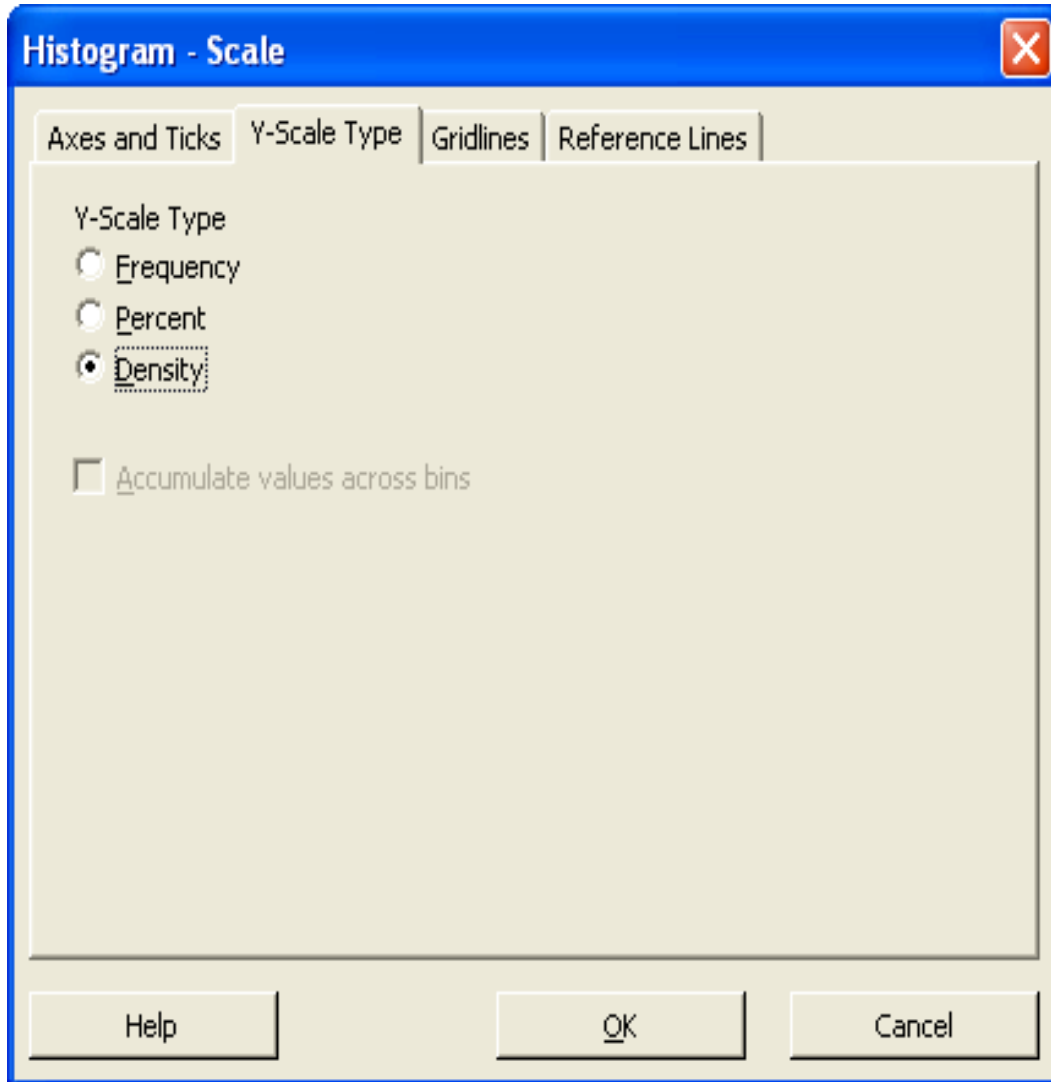
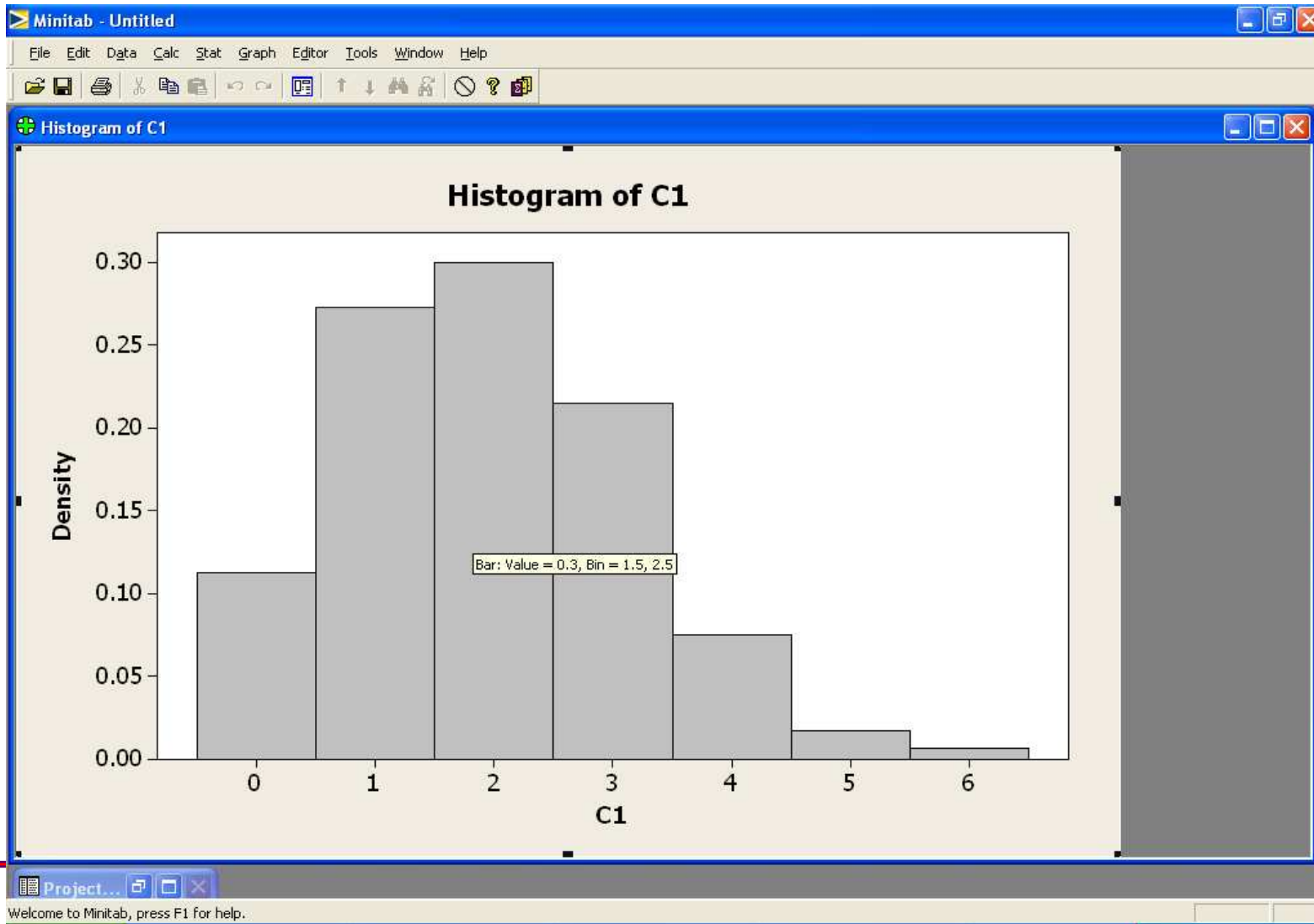Selecting the **Percent** button produces the following histogram:

MINITAB - Types of Histograms

For a density histogram, select the **Density** button.

MINITAB - Types of Histograms

Selecting the **Density** button produces the following histogram:

Histograms

Situations where one would consider unequal bin sizes arise occasionally and there is a correct way to handle them.

Histograms

Situations where one would consider unequal bin sizes arise occasionally and there is a correct way to handle them.

Suppose we wanted to combine ages 21 and 22 in the class into a single bin:

- 2 students are age 18
- 5 students are age 19
- 7 students are age 20
- 5 students are age 21-22

How should the histogram be drawn in this case?

Histograms

The convention for unequal bin sizes for frequency histograms is as follows:

The **width** of each bin on the horizontal scale represents the range of values in that bin.

Histograms

The convention for unequal bin sizes for frequency histograms is as follows:

The **width** of each bin on the horizontal scale represents the range of values in that bin.

In our example,

- The bin for students age 18 has width 1
- The bin for students age 19 has width 1
- The bin for students age 20 has width 1
- The bin for students age 21-22 has width **2**

Histograms

The **height** of each bin on the vertical scale is chosen so that the **area** of the rectangle represents the count of individuals in the bin.

Histograms

The **height** of each bin on the vertical scale is chosen so that the **area** of the rectangle represents the count of individuals in the bin.

In our example,

- The rectangle over the bin for students age 18 has area 2

- The rectangle over the bin for students age 19 has area 5

- The rectangle over the bin for students age 20 has area 7

- The rectangle over the bin for students age 21-22 has area 5

Histograms

Now based on the area and the width of each bin, we determine the height of the rectangles

Histograms

Now based on the area and the width of each bin, we determine the height of the rectangles

In our example,

- For students age 18, the width is 1 and the area is 2, so the height is 2

- For students age 19, the width is 1 and the area is 5, so the height is 5

- For students age 20. the width is 1 and the area is 7, so the height is 7

- For students age 21-22, the width is **2** and the area is 5, so the height is **2.5**