Chapter 157

Violin Plots

Introduction

When analyzing data, you often need to study the characteristics of a single group of numbers, observations, or measurements. You might want to know the center and the spread about this central value. You might want to investigate extreme values (referred to as outliers) or study the distribution or pattern of the data values. Several plots are available to allow you to study the distribution. One such plot is the violin plot.

Violin Plot

A violin plot (Hintze and Nelson, 1998) was originally designed as an enhancement of Tukey's box plot. The enhancement was to embed the box plot inside a rotated kernel density plot. The name "violin" came from some of the first examples that were tried.

The violin plot is useful because it allows the comparison of several distributional features across groups or variables.
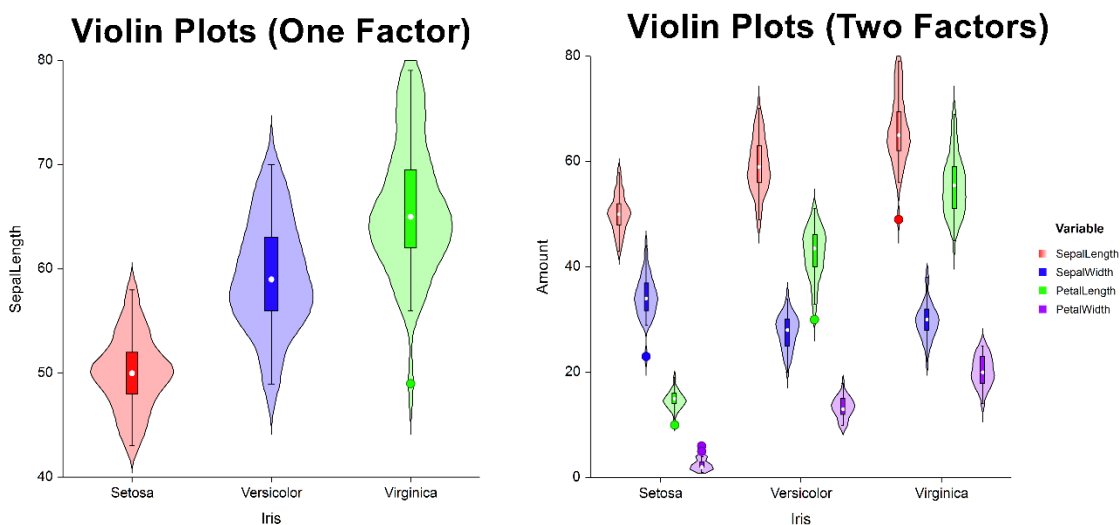

In this application, other plot objects may be used instead of the box plot: dot plot, error-bar plot, percentile plot, and histogram.

Density Trace

The violin plot uses a *density trace* to form the borders of the violin object. Density refers to the relative frequency (concentration) of data points along the data range. Mathematically, the density at a value x is defined as the fraction of data values per unit of measurement that lie in an interval centered at x. Once you choose a suitable interval width, you can calculate the density at any (and every) x value. If you calculate the density at, say, 50 values and connect them, you'll have a density trace.

The interval width is specified as a percentage. As you increase this percentage, you increase the amount of data included in each density calculation. This increases the smoothness of the chart. The following four density traces were made of the same data at increasing percentage smoothness.
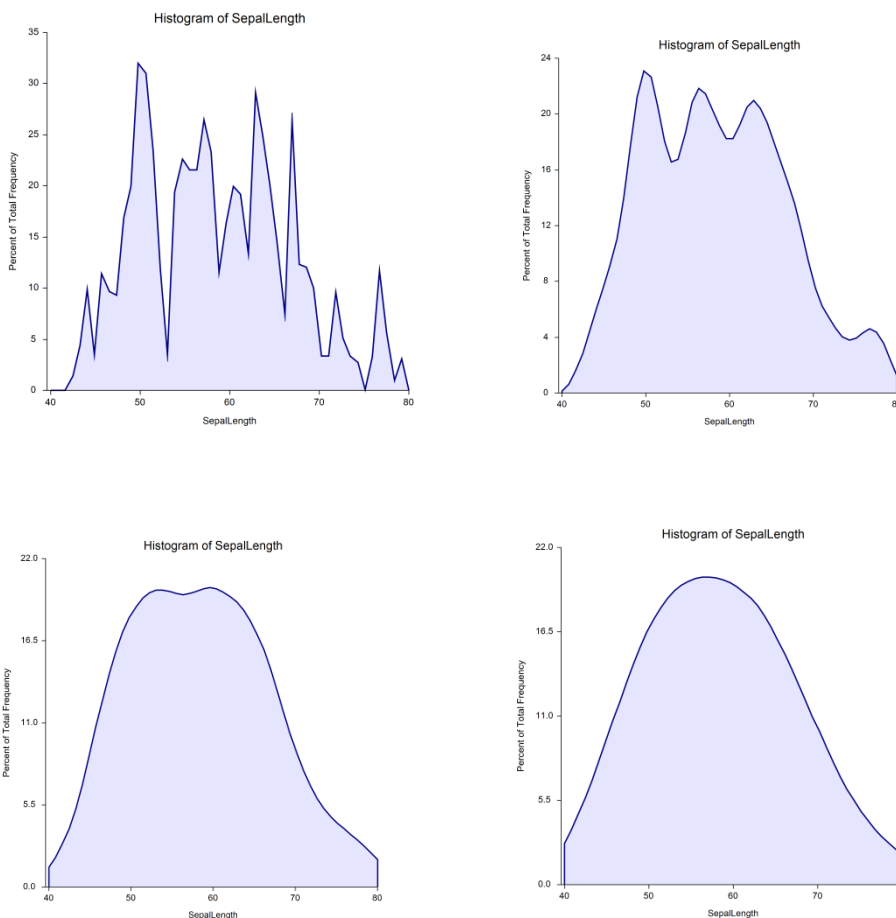
As the interval width is increased, data points further and further from the center value are included. In order to decrease the weight of points that are far removed from the center value, a weighting scheme is used that weights points proportionally to their distance from the center value. The weight function used is half the cosine function with its peak at the center value. It decreases symmetrically to zero, after which a weight of zero is applied. Hence, points have a smaller and smaller impact on the density trace as they are further and further from the center.

Another way to think of the density trace is to imagine that you construct 1000 histograms of the same data using slightly different boundary positions and take the average rectangle height at each of 50 values along the data range. This would give you a *smoothed histogram* that has many of the same properties of the density trace.
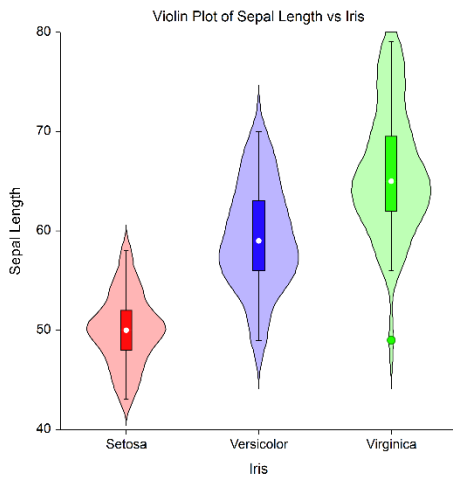
Data Structure

A violin plot is constructed from a numeric variable. A second variable may be used to divide the first variable into groups (e.g., age group or gender). In the two-factor procedure, a third variable may be used to further divide the groups into subgroups.
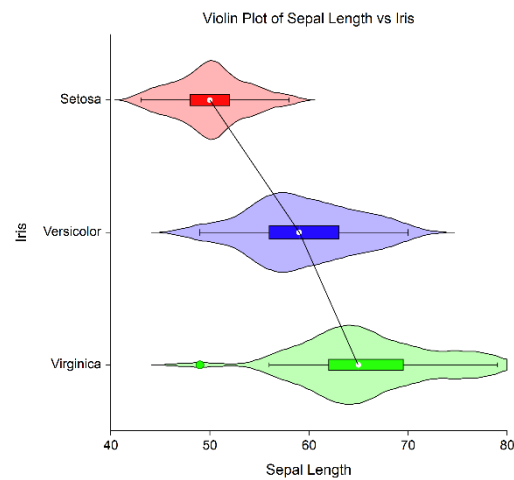
Examples

This section contains examples of violin plots. Instructions on how to create them are given below.

One-Factor Violin Plots

Example 1a

Example 1b

Example 2
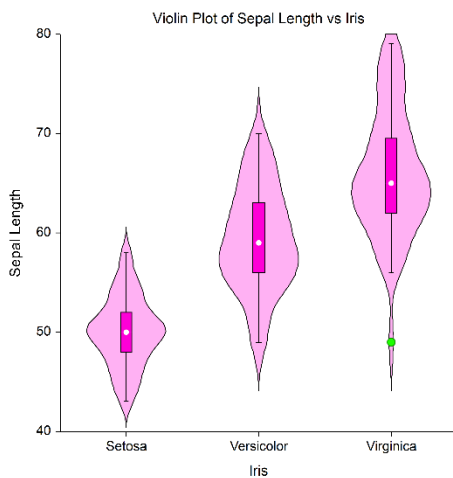
Example 3
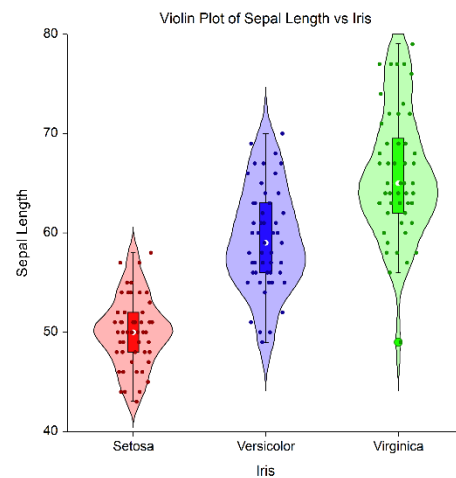
Violin Plots

Example 4
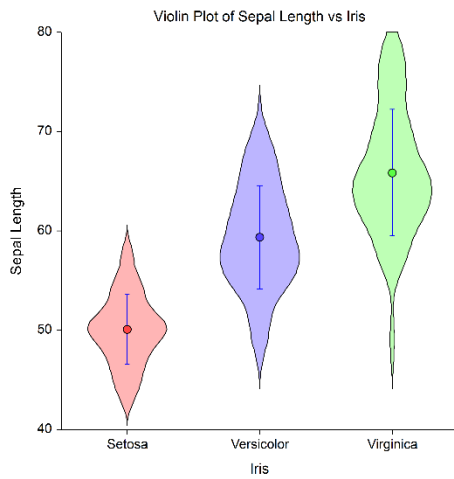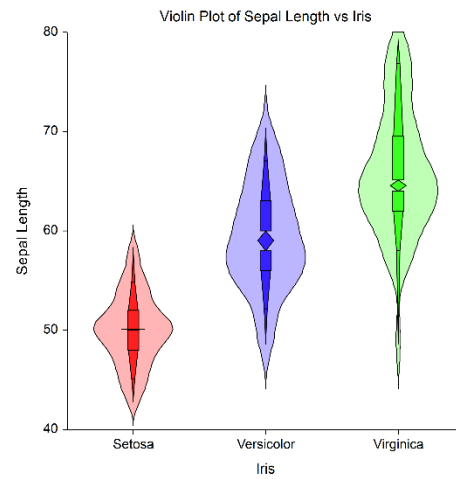
Violin Plot of Sepal Length vs Iris

Example 5

Violin Plot of Sepal Length vs Iris

Example 6

Violin Plot of Sepal Length vs Iris

Two-Factor Violin Plots

Example 7

Violin Plot of Amount vs Variable by Iris

Example 8

Violin Plot of Amount vs Variable by Iris

Example 9

Violin Plot of Amount vs Iris by Variable

Example 10

Violin Plot of Amount vs Variable by Iris
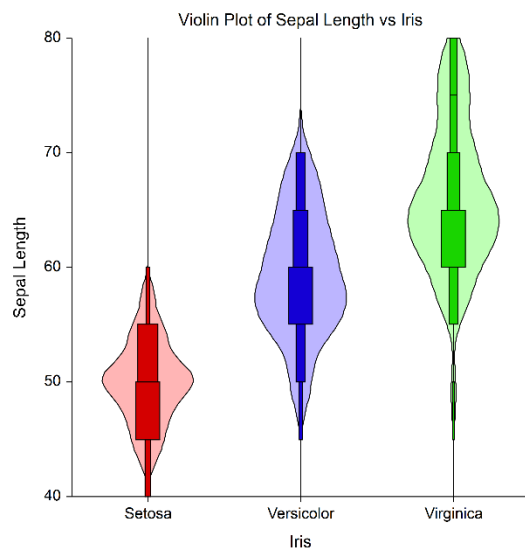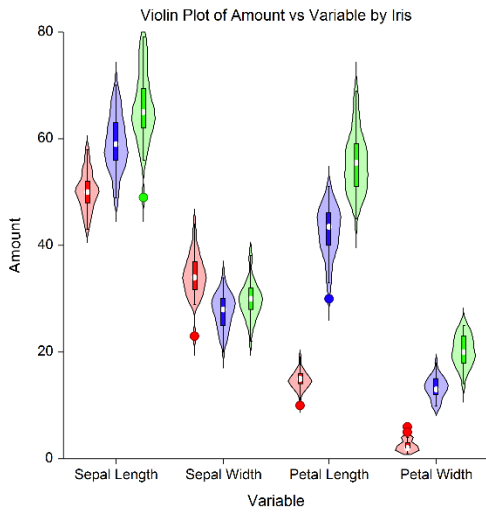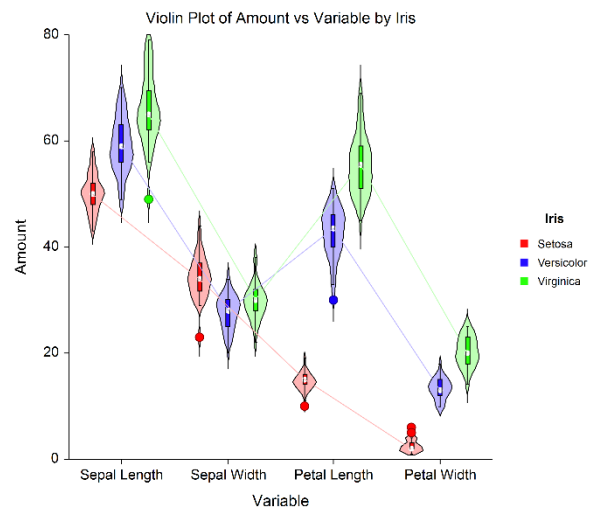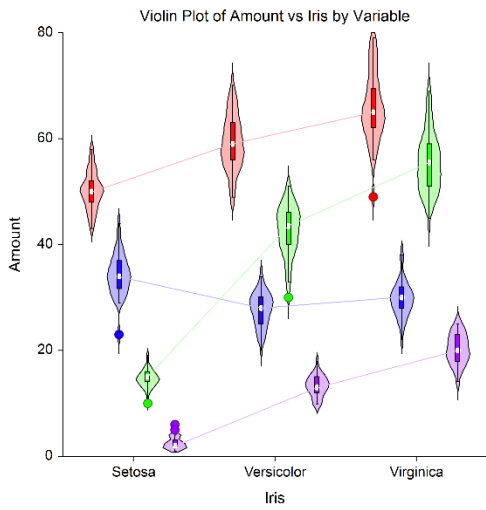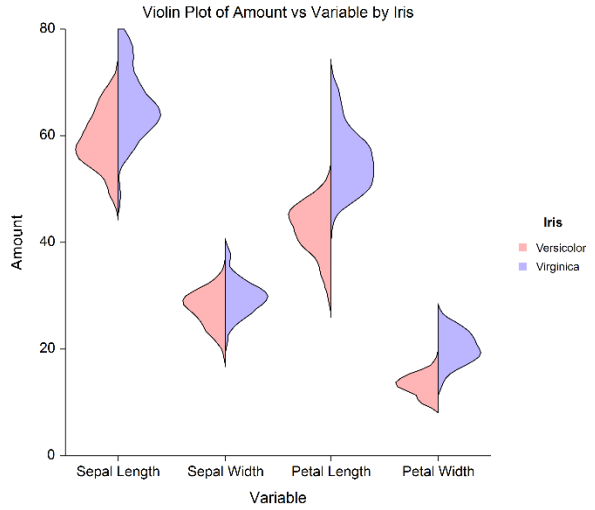
Example 1 – Basic Violin Plot

Here is an example of generating a simple violin plot.

The data are in the Fisher dataset. A set of violin plots are constructed using *SepalLength* as the Data Variable and *Iris* as the Horizontal (Group) Variable.

Setup

To run this example, complete the following steps:

1 Open the Fisher example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Fisher** and click **OK**.

2 Specify the Violin Plots procedure options

- Find and open the **Violin Plots** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1a** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Data Variable(s)...**SepalLength**
Horizontal (Group) Variable ....**Iris**

Report Options (*in the Toolbar*)

Variable Labels...**Column Labels**
Data Labels...**Value Labels**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Output
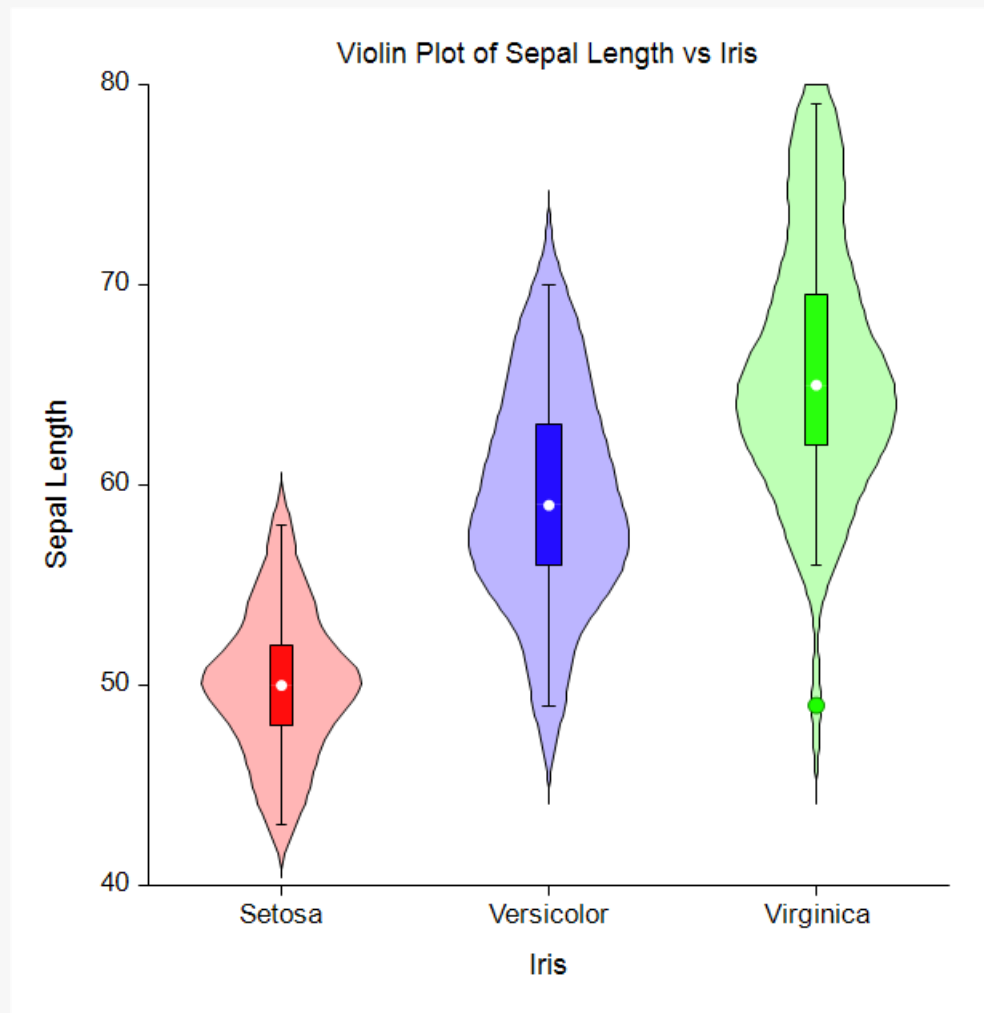
Violin Plots
_____

To show the violins horizontally and add median connecting lines, make the adjustments outlined below.

4   Specify the Violin Plots format options

- The settings for this section are listed below and are stored in the **Example 1b** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab
_____

Violin Plot Format (*Click the Button*)
_____

Layout Tab
Orientation ....**Horizontal**

Connecting Lines Tab
Medians ....**Checked**

5   Run the procedure

- Click the **Run** button to bring up the Violin Plot Format window with your data being shown.

Violin Plots
_____

Violin Plot of Sepal Length vs Iris
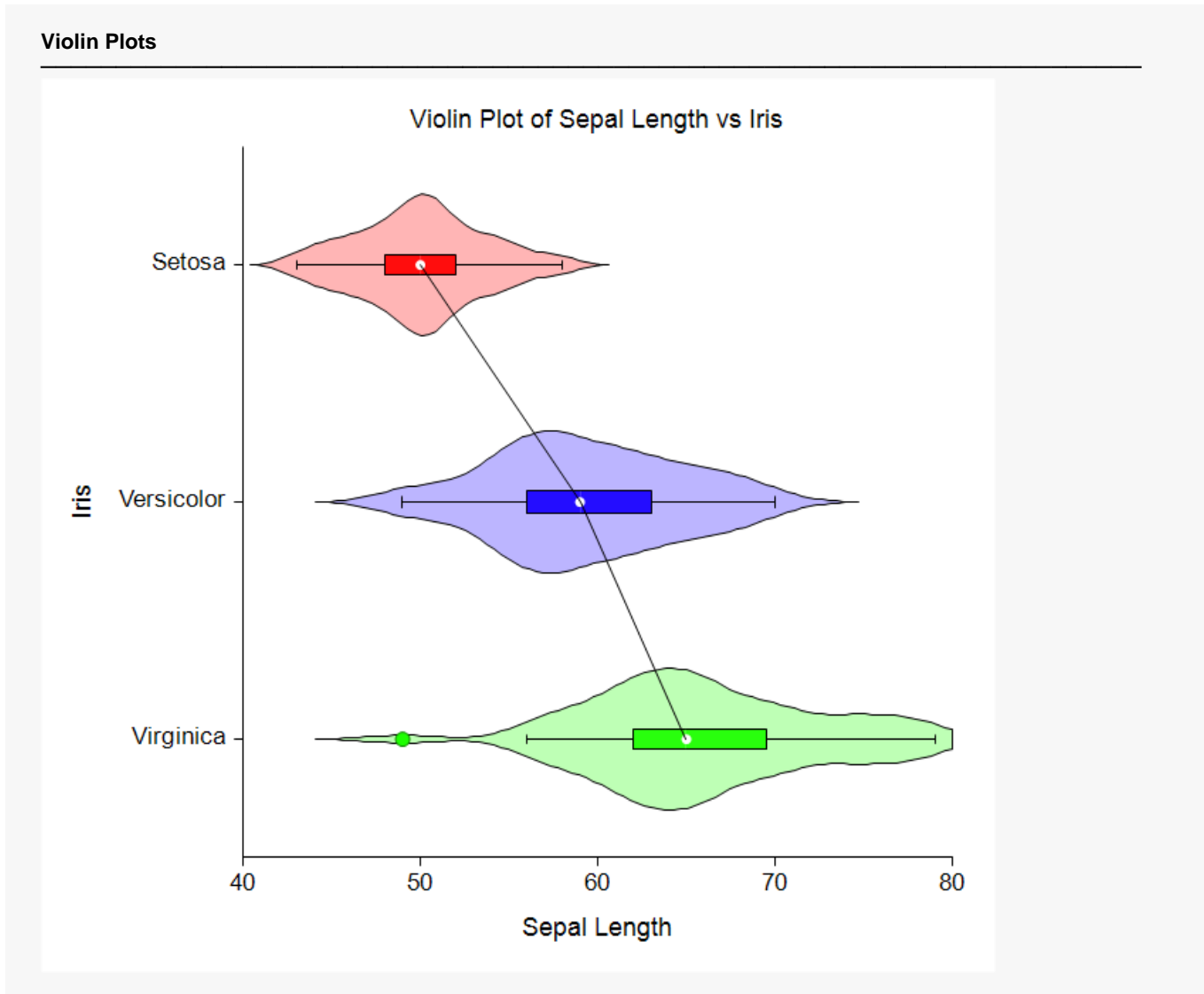
Example 2 – Basic Violin Plot with a Single Color

Here is an example of generating a simple violin plot with a single color for all violins.

The data are in the Fisher dataset. A set of violin plots are constructed using *SepalLength* as the Data Variable and *Iris* as the Horizontal (Group) Variable.

Setup

To run this example, complete the following steps:

1   Open the Fisher example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Fisher** and click **OK**.

2   Specify the Violin Plots procedure options

- Find and open the **Violin Plots** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

> Variables Tab
> _____
>
> Data Variable(s)...**SepalLength**
> Horizontal (Group) Variable ....**Iris**
>
> Violin Plot Format (*Click the Button*)
> _____
>
> *Violin Plot Tab*
> Fill Format (*Click the Button*)
> Fill Mode ....**Single Fill (one fill for All groups)**
> Fill ....**Light Pink** (select by clicking the dropdown)
>
> *Box Plot Tab*
> Fill Format (Boxes) (*Click the Button*)
> Fill Mode ....**Single Fill (one fill for All groups)**
> Fill ....**Pink** (select by clicking the dropdown)
>
> Report Options (*in the Toolbar*)
> _____
>
> Variable Labels...**Column Labels**
> Data Labels..**Value Labels**

3   Run the procedure

- Click the **Run** button to bring up the Violin Plot Format window with your data being shown.

Output

Violin Plots
_____

Violin Plot of Sepal Length vs Iris
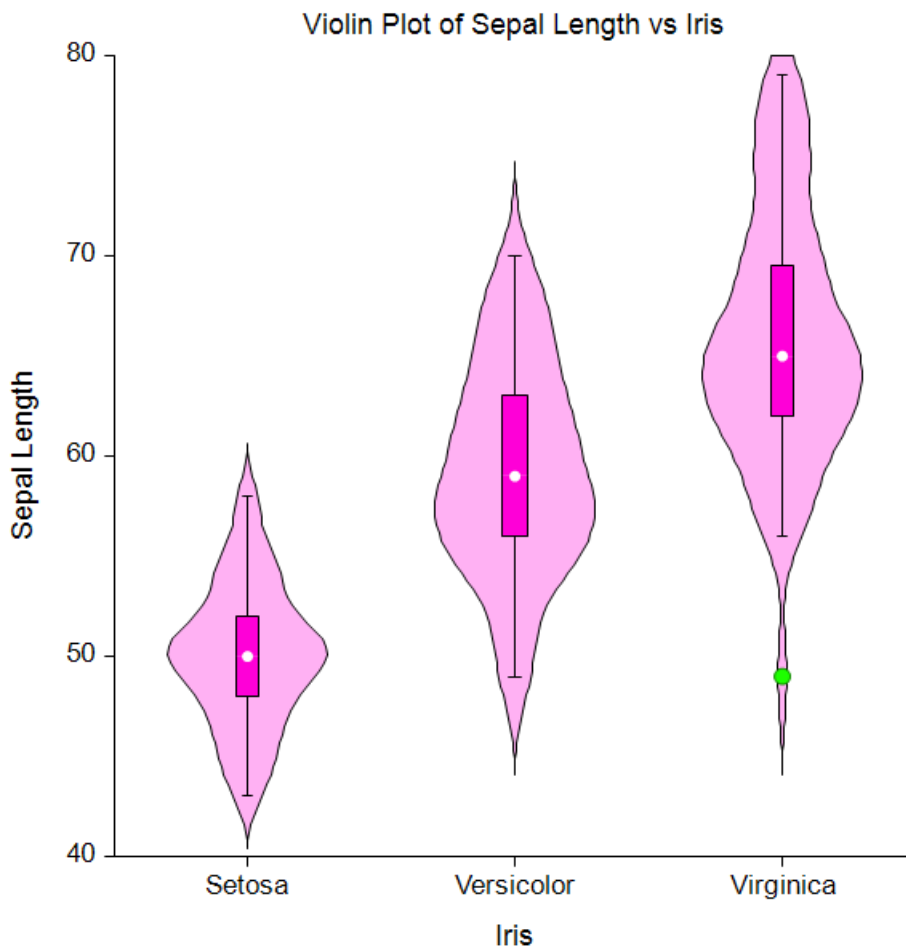
Example 3 – Violin Plot with Embedded Dot Plot

Here is an example of generating a violin plot which includes an overlay of the data.

The data are in the Fisher dataset. A set of violin plots are constructed using *SepalLength* as the Data Variable and *Iris* as the Horizontal (Group) Variable.

Setup

To run this example, complete the following steps:

1 Open the Fisher example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Fisher** and click **OK**.

2 Specify the Violin Plots procedure options

- Find and open the **Violin Plots** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 3** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Data Variable(s)...**SepalLength**
Horizontal (Group) Variable ....**Iris**

Violin Plot Format (*Click the Button*)

Dot Plot Tab
Show Dot Plot...**Checked**

Report Options (*in the Toolbar*)

Variable Labels...**Column Labels**
Data Labels...**Value Labels**

3 Run the procedure

- Click the **Run** button to bring up the Violin Plot Format window with your data being shown.

Output
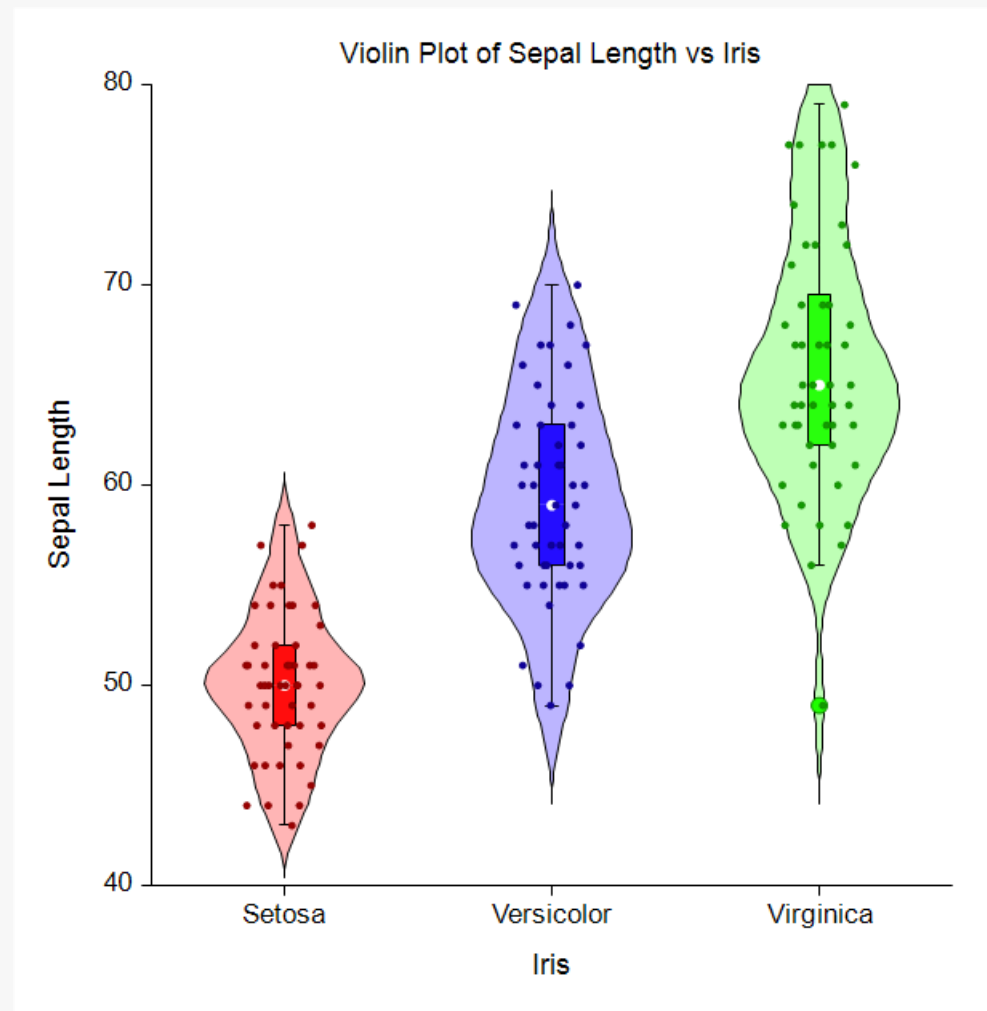
Violin Plots
_____

Example 4 – Violin Plot with Embedded Error-Bar Plot

Here is an example of generating a violin plot which includes an embedded error-bar plot.

The data are in the Fisher dataset. A set of violin plots are constructed using *SepalLength* as the Data Variable and *Iris* as the Horizontal (Group) Variable.

Setup

To run this example, complete the following steps:

1   Open the Fisher example dataset
  - From the File menu of the NCSS Data window, select **Open Example Data**.
  - Select **Fisher** and click **OK**.

2   Specify the Violin Plots procedure options
  - Find and open the **Violin Plots** procedure using the menus or the Procedure Navigator.
  - The settings for this example are listed below and are stored in the **Example 4** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

> Variables Tab
> _____
>
> Data Variable(s)..**SepalLength**
> Horizontal (Group) Variable ....**Iris**
>
> Violin Plot Format (*Click the Button*)
> _____
>
> *Box Plot Tab*
> Show Box Plot ....**Unchecked**
>
> *Error-Bar Chart Tab*
> Show Error-Bar Chart ....**Checked**
>
> Report Options (*in the Toolbar*)
> _____
>
> Variable Labels...**Column Labels**
> Data Labels..**Value Labels**
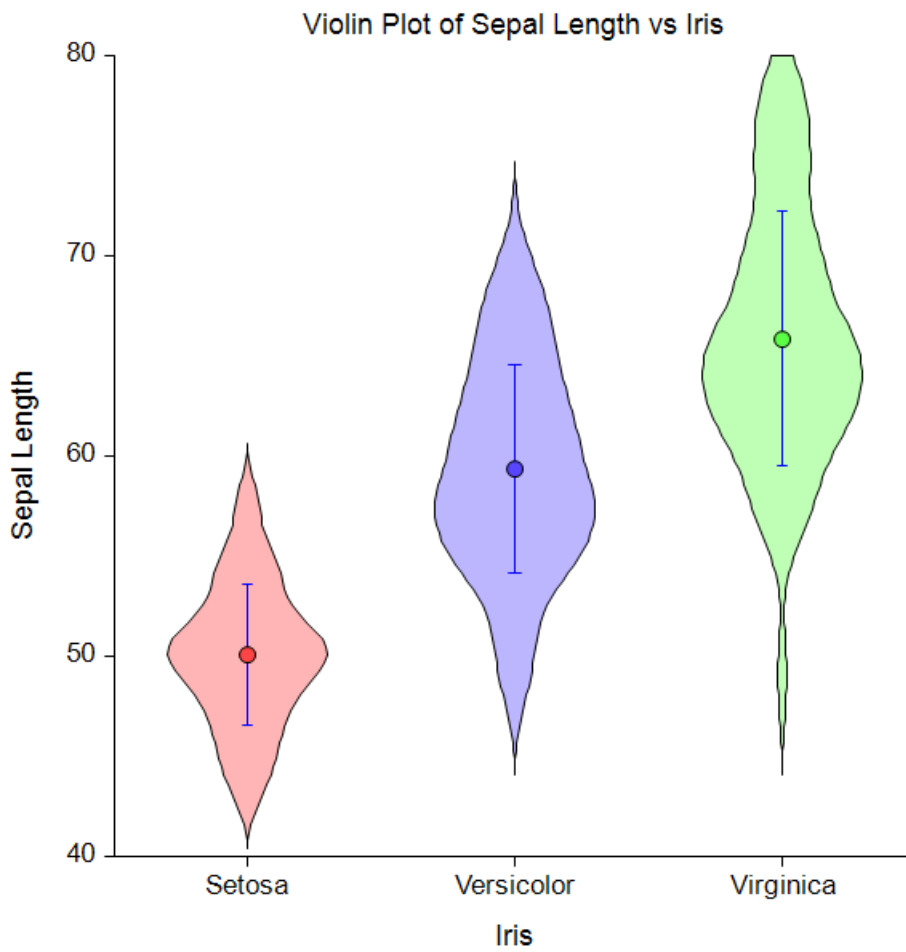
3   Run the procedure
  - Click the **Run** button to bring up the Violin Plot Format window with your data being shown.

Output

Violin Plots
_____

Violin Plot of Sepal Length vs Iris

Example 5 – Violin Plot with Embedded Percentile Plot

Here is an example of generating a violin plot which includes an embedded percentile plot.

The data are in the Fisher dataset. A set of violin plots are constructed using *SepalLength* as the Data Variable and *Iris* as the Horizontal (Group) Variable.

Setup

To run this example, complete the following steps:

1   Open the Fisher example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Fisher** and click **OK**.

2   Specify the Violin Plots procedure options

- Find and open the **Violin Plots** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 5** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Data Variable(s)...**SepalLength**
Horizontal (Group) Variable ....**Iris**

Violin Plot Format (*Click the Button*)

Box Plot Tab
Show Box Plot ....**Unchecked**

Percentile Plot Tab
Show Percentile Plot...**Checked**

Report Options (*in the Toolbar*)

Variable Labels...**Column Labels**
Data Labels..**Value Labels**

3   Run the procedure

- Click the **Run** button to bring up the Violin Plot Format window with your data being shown.

Output

Violin Plots
_____

Violin Plot of Sepal Length vs Iris
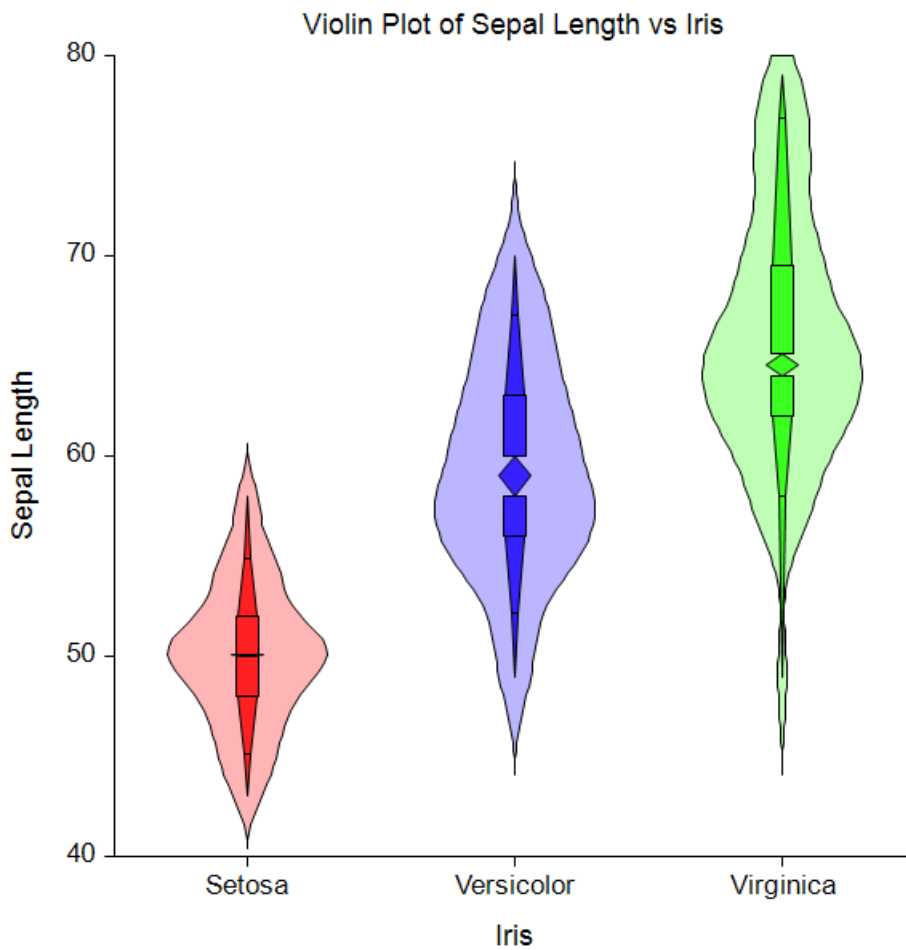
Example 6 – Violin Plot with an Embedded Histogram

Here is an example of generating a violin plot which includes an embedded histogram with reflection.

The data are in the Fisher dataset. A set of violin plots are constructed using *SepalLength* as the Data Variable and *Iris* as the Horizontal (Group) Variable.

Setup

To run this example, complete the following steps:

1   Open the Fisher example dataset
  - From the File menu of the NCSS Data window, select **Open Example Data**.
  - Select **Fisher** and click **OK**.

2   Specify the Violin Plots procedure options
  - Find and open the **Violin Plots** procedure using the menus or the Procedure Navigator.
  - The settings for this example are listed below and are stored in the **Example 6** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Data Variable(s)...**SepalLength**
Horizontal (Group) Variable ....**Iris**

Violin Plot Format (*Click the Button*)

Box Plot Tab
Show Box Plot ....**Unchecked**

Histogram Tab
Show Histogram ....**Checked**

Report Options (*in the Toolbar*)

Variable Labels...**Column Labels**
Data Labels...**Value Labels**

---

3   Run the procedure
  - Click the **Run** button to bring up the Violin Plot Format window with your data being shown.

Output
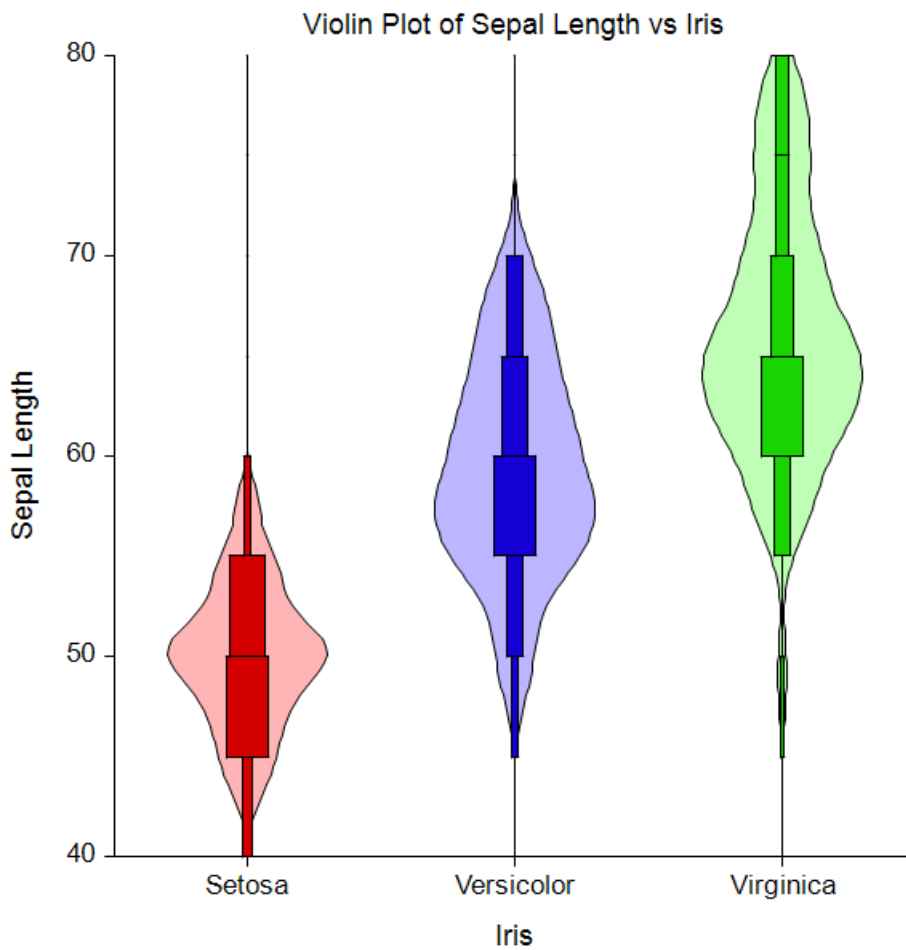
Violin Plots
_____

Note that the embedded histogram also contains its reflection so that it is analogous to the violin plot.

Example 7 – Violin Plot with Two Factors

Here is an example of generating a set of violin plots that display two factors: one is the Data Variable, and the other is the Legend Variable.

The data are in the Fisher dataset. A set of violin plots are constructed using *SepalLength, SepalWidth, PetalLength, PetalWidth* as the Data Variables and *Iris* as the Legend Variable. Note that the Horizontal (Group) Variable is left unspecified.

Setup

To run this example, complete the following steps:

1 Open the Fisher example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Fisher** and click **OK**.

2 Specify the Violin Plots (2 Factors) procedure options

- Find and open the **Violin Plots (2 Factors)** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 7** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Data Variable(s)..**SepalLength**, **SepalWidth**, **PetalLength**, **PetalWidth**
Legend (Subgroup) Variable............................**Iris**

Report Options (*in the Toolbar*)

Variable Labels..**Column Labels**
Data Labels...**Value Labels**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Output
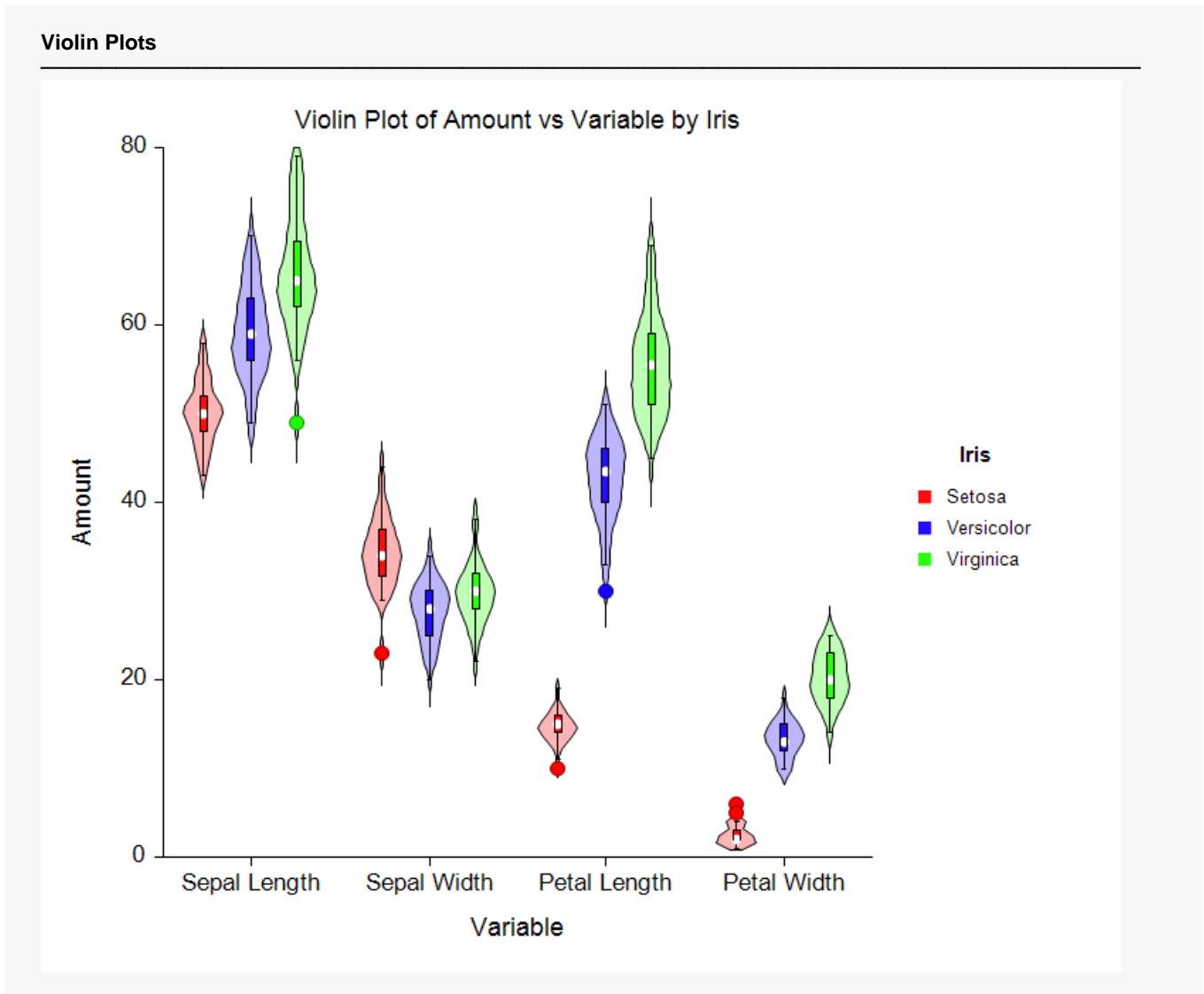
Violin Plots
_____

This plot allows you to see several patterns and relationships within the data.

Example 8 – Violin Plot with Two Factors and Connected Medians

Here is an example of generating a set of violin plots that display two factors: one is the Data Variable and the other is the Legend Variable. The plots of each variety are connected at the medians.

The data are in the Fisher dataset. A set of violin plots are constructed using *SepalLength, SepalWidth, PetalLength, PetalWidth* as the Data Variables and *Iris* as the Legend Variable. Note that the Horizontal (Group) Variable is left unspecified.

Setup

To run this example, complete the following steps:

1   Open the Fisher example dataset
- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Fisher** and click **OK**.

2   Specify the Violin Plots (2 Factors) procedure options
- Find and open the **Violin Plots (2 Factors)** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 8** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Data Variable(s)...**SepalLength, SepalWidth, PetalLength, PetalWidth**
Legend (Subgroup) Variable............................**Iris**

Violin Plot Format (*Click the Button*)

Connecting Lines Tab
Medians (Connect Between Groups)...............**Checked**

Report Options (*in the Toolbar*)

Variable Labels...**Column Labels**
Data Labels..**Value Labels**

3   Run the procedure
- Click the **Run** button to bring up the Violin Plot Format window with your data being shown.

Output
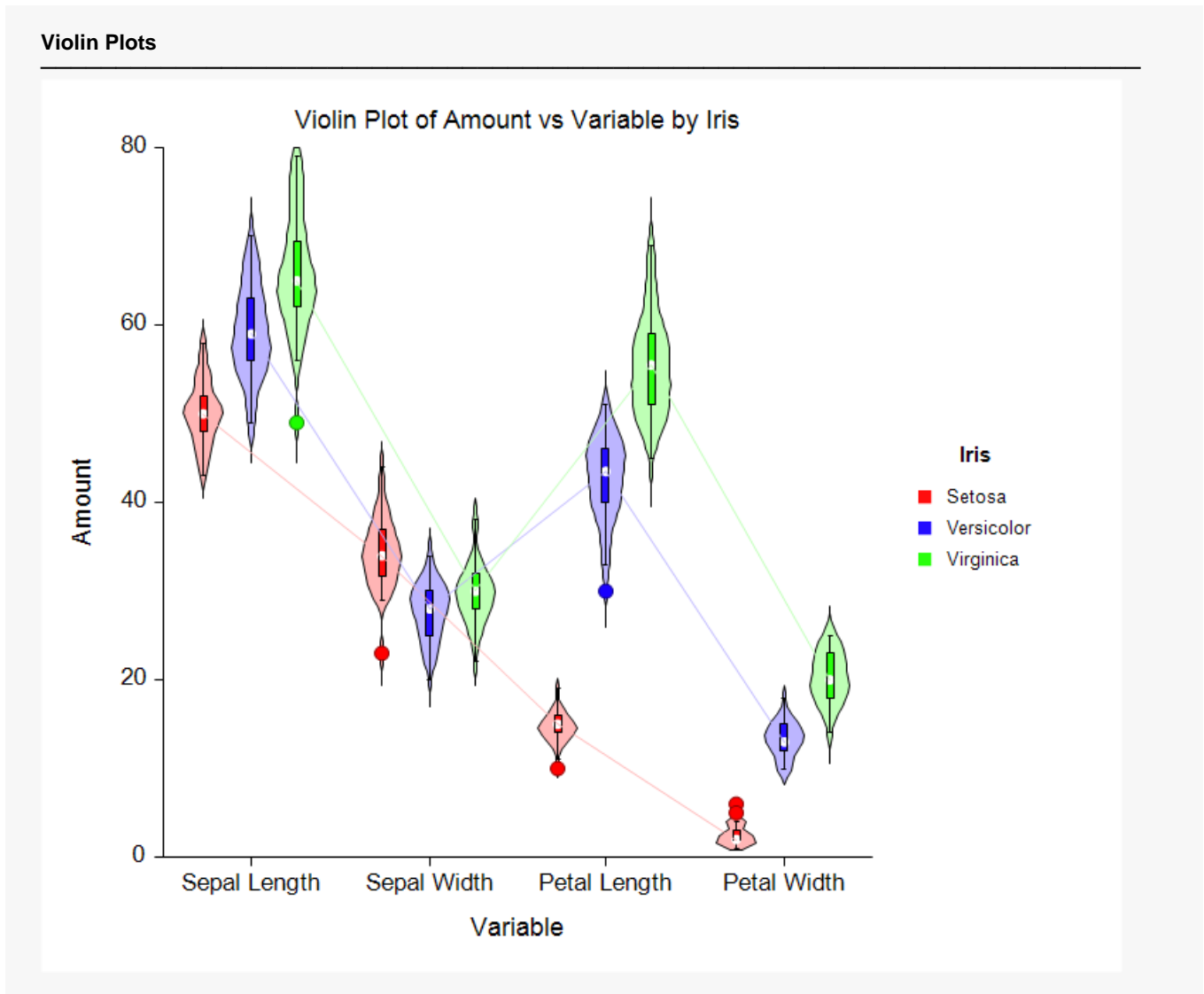
Violin Plots
_____

This plot allows you to see several patterns and relationships within the data.

Example 9 – Violin Plot with Two Factors and Connected Medians with Iris as the Horizontal Variable

Here is an example of generating a set of violin plots that display two factors: one is the Data Variable name and the other is the Horizontal Variable. The plots of each variable are connected at the medians.

The data are in the Fisher Iris dataset. A set of violin plots are constructed using *SepalLength, SepalWidth, PetalLength, PetalWidth* as the Data Variables and *Iris* as the Horizontal Variable. Note that the Legend (Subgroup) Variable is left unspecified.

Setup

To run this example, complete the following steps:

1   Open the Fisher example dataset
- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Fisher** and click **OK**.

2   Specify the Violin Plots (2 Factors) procedure options
- Find and open the **Violin Plots (2 Factors)** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 9** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Data Variable(s)...**SepalLength, SepalWidth, PetalLength, PetalWidth**
Horizontal (Group) Variable ....**Iris**

Violin Plot Format (*Click the Button*)

Connecting Lines Tab
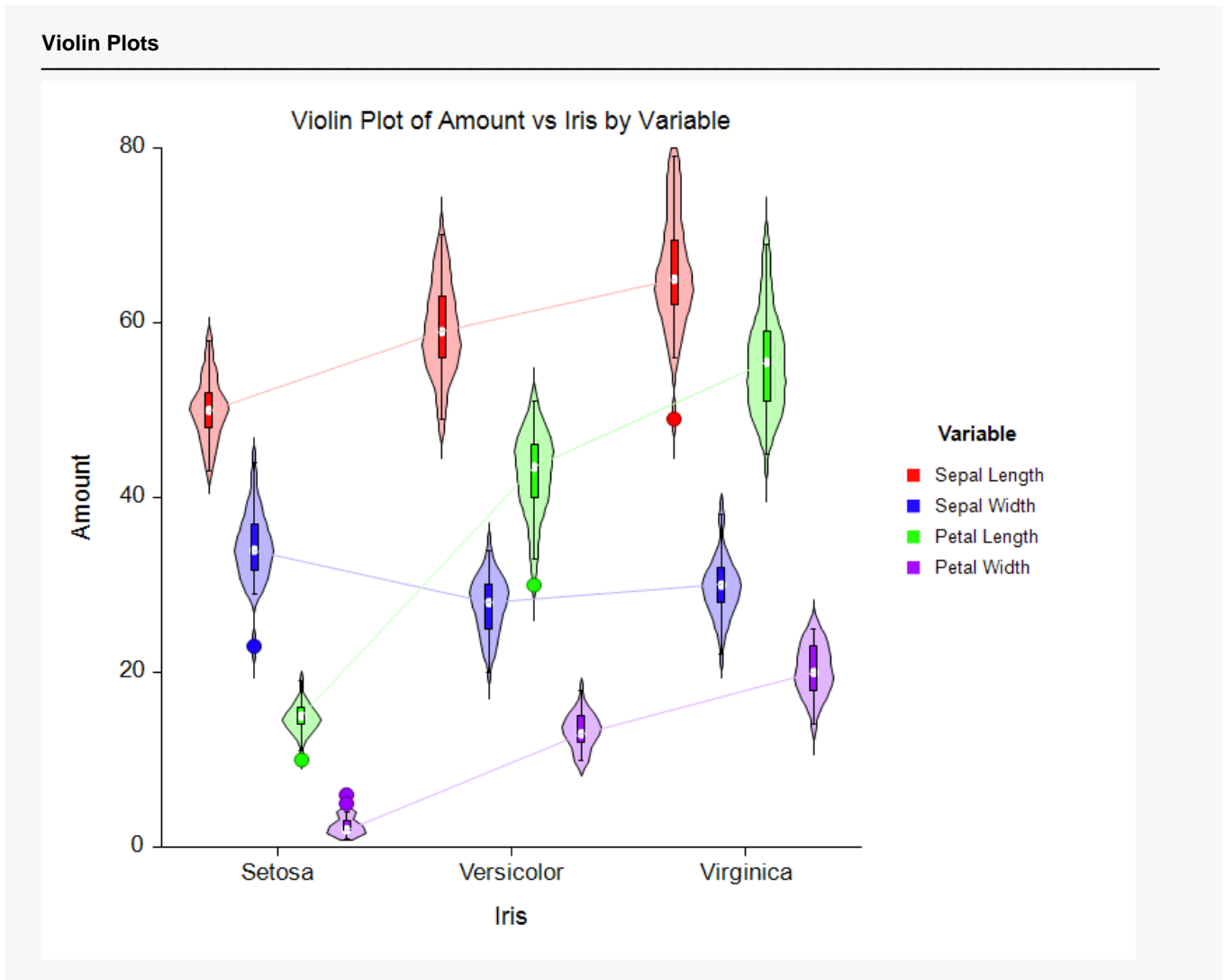Medians (Connect Between Groups)...............**Checked**

Report Options (*in the Toolbar*)

Variable Labels..**Column Labels**
Data Labels...**Value Labels**

3   Run the procedure
- Click the **Run** button to bring up the Violin Plot Format window with your data being shown.

Output

Violin Plots
_____

Notice how the connecting line helps you see the relations among the varieties within the variables.

Example 10 – Side-by-Side Violin Plot with Two Factors

Here is an example of generating a set of side-by-side (split) violin plots that display two factors: one is the Data Variable name (which is shown alone the horizontal axis) and the other is a two-value Legend Variable.

Note that the side-by-side violin plot requires that the legend variable has only two values.

The data are in the Fisher Iris dataset. A set of violin plots are constructed using *SepalLength, SepalWidth, PetalLength, PetalWidth* as the Data Variables and *Iris* as the Legend Variable.

To obtain a data set with only two Iris categories, a filter is set so that rows with Iris = 1 (Setosa) are not used.

Setup

To run this example, complete the following steps:

1   Open the Fisher example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Fisher** and click **OK**.
- Set a filter on the data by entering **2, 3** in column 5 of the row labelled **Filter** in the Column Info Table. This will cause all rows in which Iris = 1 to be ignored.

2   Specify the Violin Plots (2 Factors) procedure options

- Find and open the **Violin Plots (2 Factors)** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 10** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab
_____

Data Variable(s)..**SepalLength, SepalWidth, PetalLength, PetalWidth**
Legend (Subgroup) Variable............................**Iris**

Violin Plot Format (*Click the Button*)
_____

Violin Plot Tab
Shape ....**Alternating left**, **right (or down, up)**

Box Plot Tab
Show Box Plot ....**Unchecked**

Layout Tab
Gap between Subgroups **0%**

Report Options (*in the Toolbar*)
_____

Variable Labels...**Column Labels**
Data Labels...**Value Labels**

---

3   Run the procedure

- Click the **Run** button to bring up the Violin Plot Format window with your data being shown.

Output

Violin Plots

This plot allows you to see several patterns and relationships within the data.