

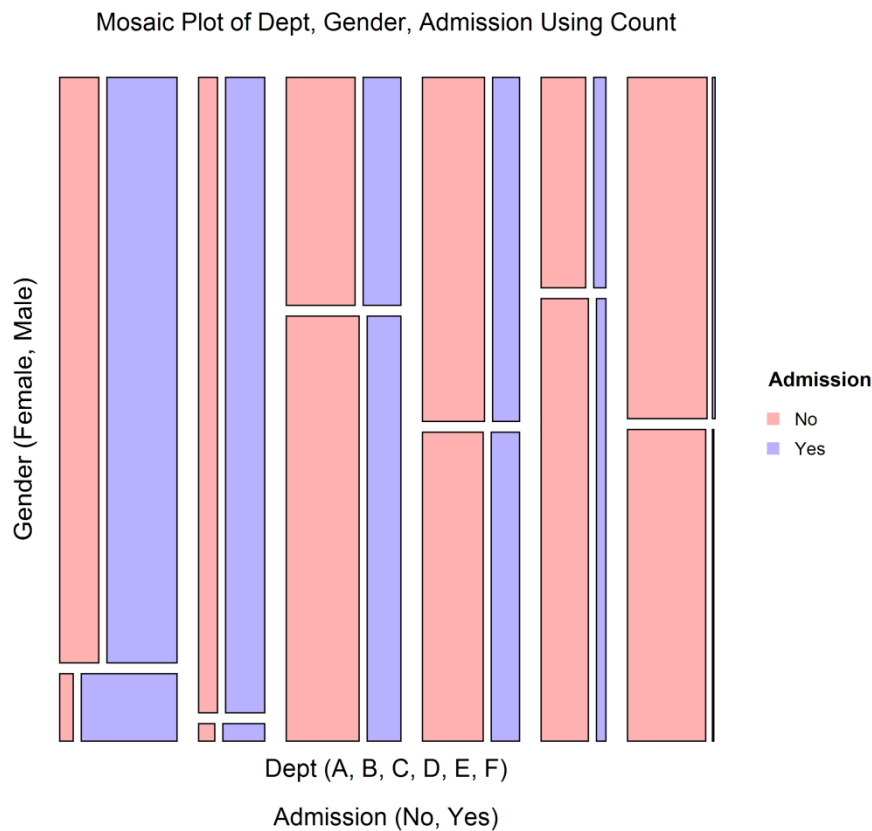
## Chapter 147

# Mosaic Plots

### Introduction

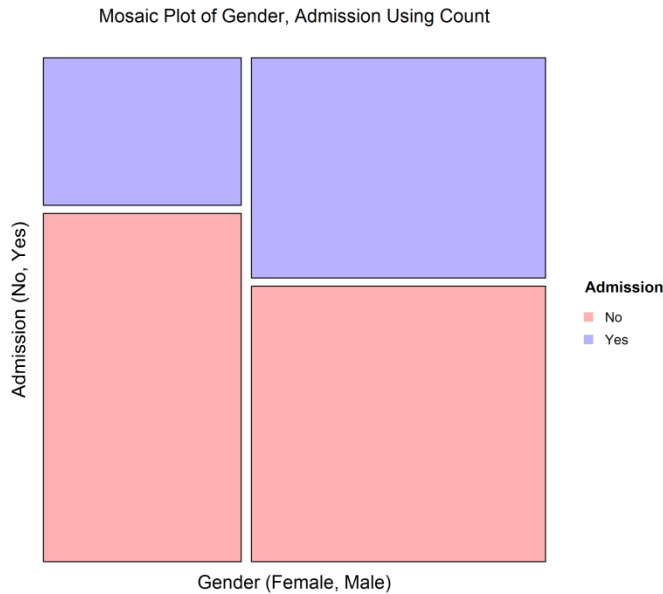
A mosaic plot is a graphical display of the cell frequencies of a contingency table in which the area of boxes of the plot are proportional to the cell frequencies of the contingency table. This procedure can construct mosaic plots for up to four-way contingency tables.

Here is an example of a three-way mosaic plot of the 1973 Berkeley Admissions data.



## Mosaic Plot Construction

Since the mosaic plot is based on conditional probabilities, to understand and interpret it, you must understand how it is created. To do this, we will use the famous 1973 Berkeley admissions data contained in the *Berkeley 1973 Admissions* dataset show below. These data are of interest, because, initially, they were used to show that males were admitted at a higher rate than females. The following chart seems to show this.



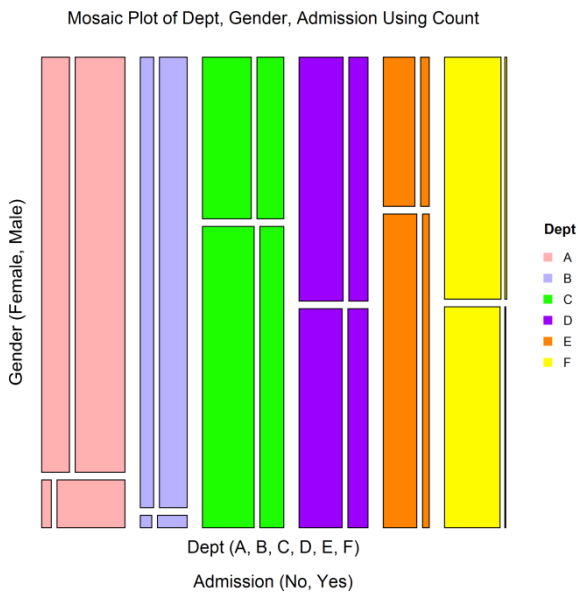
The widths of the boxes are proportional to the percentage of females and males, respectively. In fact, 41% of applicants were female and 59% were male.

The heights of the boxes are proportional to percent admitted. In fact, 45% of the male applicants were admitted, while only 30% of the female applicants were admitted. This seems to show a large gender-bias in admission.

To make the plot easier to interpret, the boxes for admitted females and males are colored blue while the not admitted females and males are colored pink.

It is easy to see that females' blue box on the left is much shorter than the males' blue box on the right.

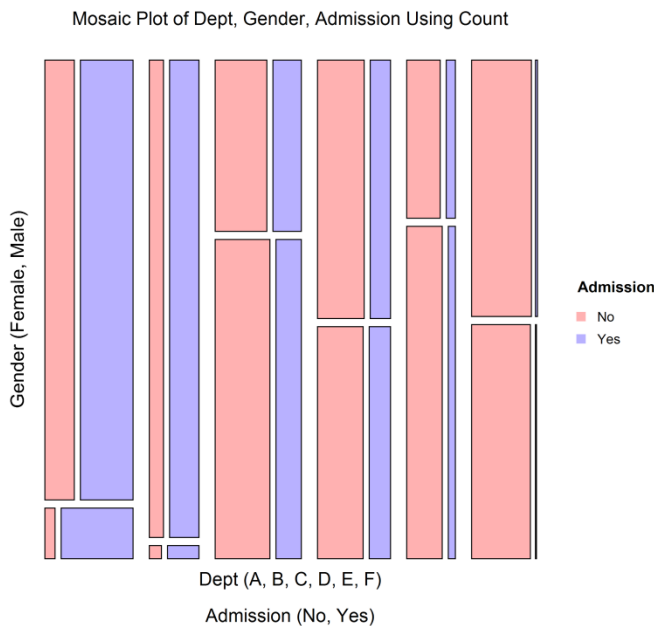
To understand this admission pattern further, the university department of application was considered.



In the following plot, the departments are shown across the plot in different colors, from department A on the left in pink to department F on the right in yellow. The percentage of applicants to each department is proportional to the width of the bars. It is obvious that departments A and C have the largest number of applicants and departments B and E have the smallest.

### Mosaic Plots

Finally, the color of the boxes is changed so that those that were admitted are shown as blue and those that were not admitted are shown as red.



By construction, the percent admitted within each gender-by-department combination is the width of the corresponding box.

For example, the percentage of females that were admitted to department A (shown by the width of blue box at the lower left) is much larger than that of the males (shown by the width of the long blue box directly above the female box).

If you consider each department in turn by scanning from left to right across the plot, the width of the blue box on the bottom appears to be quite similar to the box directly above it. This indicates that in most departments the percent of females admitted is about the same as that of males admitted.

### Keys of Interpretation of Mosaic Plots

1. The categories of each new factor divide each box either horizontally (1<sup>st</sup> and 3<sup>rd</sup> factor) or vertically (2<sup>nd</sup> and 4<sup>th</sup> factor).
2. If two factors are independent, the gaps between the corresponding sets of boxes will align.
3. The area of each box is proportional to the corresponding cell frequency.

## Data Structure

Data for a mosaic plot are entered in columns. Up to four factor variables may be used followed by an optional variable containing the counts (frequencies) for that cell. The program will tabulate data, so you do not have to use the Count variable.

Following are the data for the 1973 Berkeley Admissions dataset.

### Berkeley 1973 Admissions dataset

Dept	Gender	Admission	Count
A	Male	Yes	512
A	Male	No	313
A	Female	Yes	89
A	Female	No	19
B	Male	Yes	353
B	Male	No	207
B	Female	Yes	17
B	Female	No	8
C	Male	Yes	120
C	Male	No	205
C	Female	Yes	202
C	Female	No	391

## Mosaic Plots

D	Male	Yes	138
D	Male	No	279
D	Female	Yes	131
D	Female	No	244
E	Male	Yes	53
E	Male	No	138
E	Female	Yes	94
E	Female	No	299
F	Male	Yes	22
F	Male	No	351
F	Female	Yes	24
F	Female	No	317

---

## Procedure Options

This section describes the options available.

---

### Variables Tab

Specify the variables (columns) used to make a simple bar chart.

---

#### Factors

##### Variables 1 - 4

These variables contain the factor variables. Each variable holds the categories for a single factor. The categories may be text or numeric values. Numeric values must contain only a few unique values.

---

#### Frequency Variable

##### Frequencies

Specify an optional frequency (count) variable. This variable contains integers that represent the number of observations (frequency) associated with each row of the dataset.

If this option is left blank, each dataset row has a frequency of one. This variable lets you modify that frequency. This may be useful when your data are tabulated and you want to enter counts.

---

#### Labels

##### Names

This option specifies whether the variable names or labels are used.

##### Values

Value Labels may be used to make reports more legible by assigning meaningful labels to numbers and codes.

##### Data Values

All data are displayed in their original format, regardless of whether a value label has been set or not.

##### Value Labels

All values of variables that have a value label variable designated are converted to their corresponding value label when they are output. This does not modify their value during computation.

## Mosaic Plots

### Both

Both data value and value label are displayed.

### Example

A variable named GENDER (used as a factor variable) contains 1's and 2's. By specifying a value label for GENDER, the report can display *Male* instead of 1 and *Female* instead of 2. This option specifies whether (and how) to use the value labels.

---

## Report

### Data Summary Report

Check this box to display a numeric report of the summary table from which the plots are generated. Label Variable

---

## Plots

### Which Plots

This option designates which set of plots is generated.

### Single Plot

Generate a mosaic plot using the factors in the order given.

### Several Plots - Colored by Each Factor

Generate mosaic plots using the factors in the order given. The coloring of the boxes in each plot uses categories of a particular factor: the first plot is colored using the first factor, the second plot is colored using the second factor, and so on.

### Several Plots - One for Each Factor Ordering

The shape of a mosaic plot is greatly influenced by the order of the factors. This option generates a separate plot for each permutation of the factors. In all cases, the plot is colored by the categories of the last factor.

### Edit During Run

Checking this option will cause the mosaic plot format window to appear when the procedure is run. This allows you to modify the format of the graph with the actual data.

## Mosaic Plot Window Options

This section describes the specific options available on the Mosaic Plot window, which is displayed when the Mosaic Plot button is clicked. Common options, such as axes, labels, legends, and titles are documented in the Graphics Components chapter.

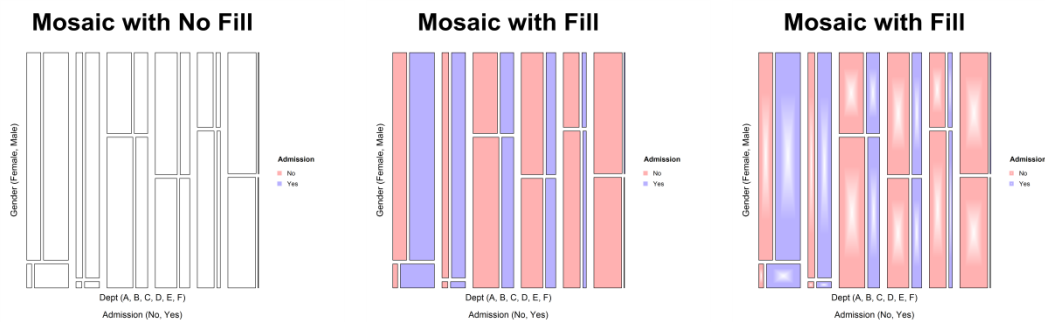
### Mosaic Plot Tab

#### Rectangles Section

You can control the borders and the fills using these options.

##### Rectangle Fill

This option controls the colors and gradients that are used to fill the boxes. The colors are applied according to the colorizing factor, which is usually the last factor specified.



##### Rectangle Borders

This option controls the colors of the borders. The colors are applied according to the colorizing factor, which is usually the last factor specified.

## Mosaic Plots

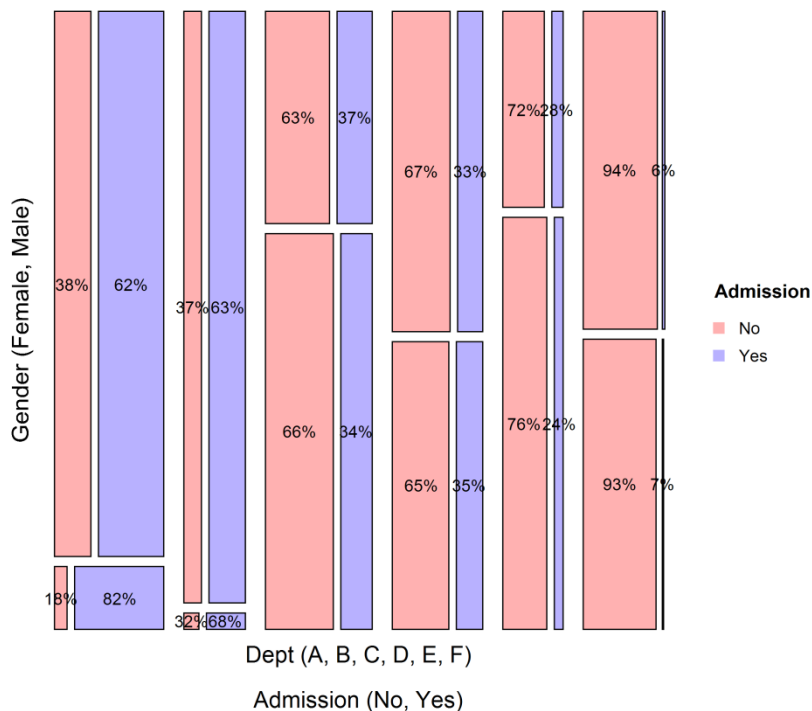
### Data Values Section

You can control whether the percentages are displayed along with their style using these options.

#### Values

This option controls whether the percentage values are displayed inside each box.

## Data Labels

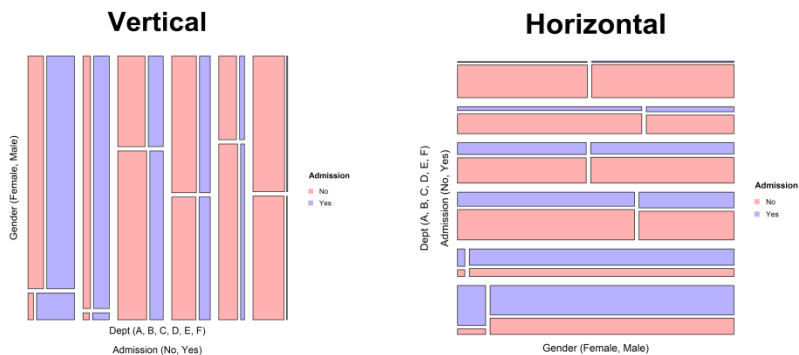


#### Position

This option controls the vertical position of the data label within each box.

### Orientation Section

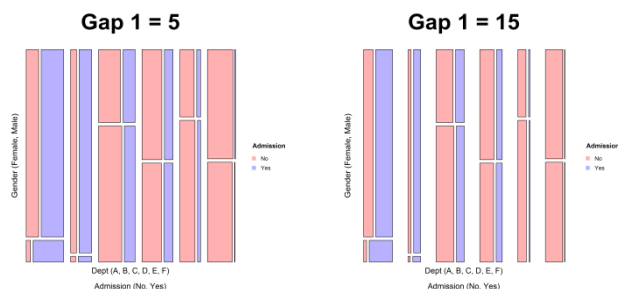
This option controls whether the boxes for factor 1 are *Vertical* or *Horizontal*.



## Mosaic Plots

## Spacing Section

You can change the space between the boxes for each category.



## Titles, Legend, Numeric Axis, Group Axis, Grid Lines, and Background Tabs

Details on setting the options in these tabs are given in the Graphics Components chapter.

### Example 1 – Creating a Mosaic Plot

This section presents an example of how to create a mosaic plot of the data stored in the *Berkeley 1973 Admissions* dataset.

You may follow along here by making the appropriate entries or load the completed template **Example 1** by clicking on Open Example Template from the File menu of the Mosaic Plots window.

#### 1 Open the Berkeley 1973 Admissions dataset.

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **Berkeley 1973 Admissions.NCSS**.
- Click **Open**.

#### 2 Open the Mosaic Plot window.

- Using the Graphics menu or the Procedure Navigator, find and select the **Mosaic Plots** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

#### 3 Specify the Variables.

- Double-click in the Variable 1 text box. This will bring up the variable selection window.
- Select **Dept** from the list of variables and then click **Ok**. “Dept” will appear in the Factor 1 box.
- Double-click in the Variable 2 text box. This will bring up the variable selection window.
- Select **Gender** from the list of variables and then click **Ok**. “Gender” will appear in the Factor 2 box.
- Double-click in the Variable 3 text box. This will bring up the variable selection window.
- Select **Admission** from the list of variables and then click **Ok**. “Admission” will appear in the Factor 3 box.
- Double-click in the Frequencies text box. This will bring up the variable selection window.
- Select **Count** from the list of variables and then click **Ok**. “Count” will appear in the Frequencies box.
- Check the **Data Summary Report** box.

#### 4 Run the procedure.

- From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.

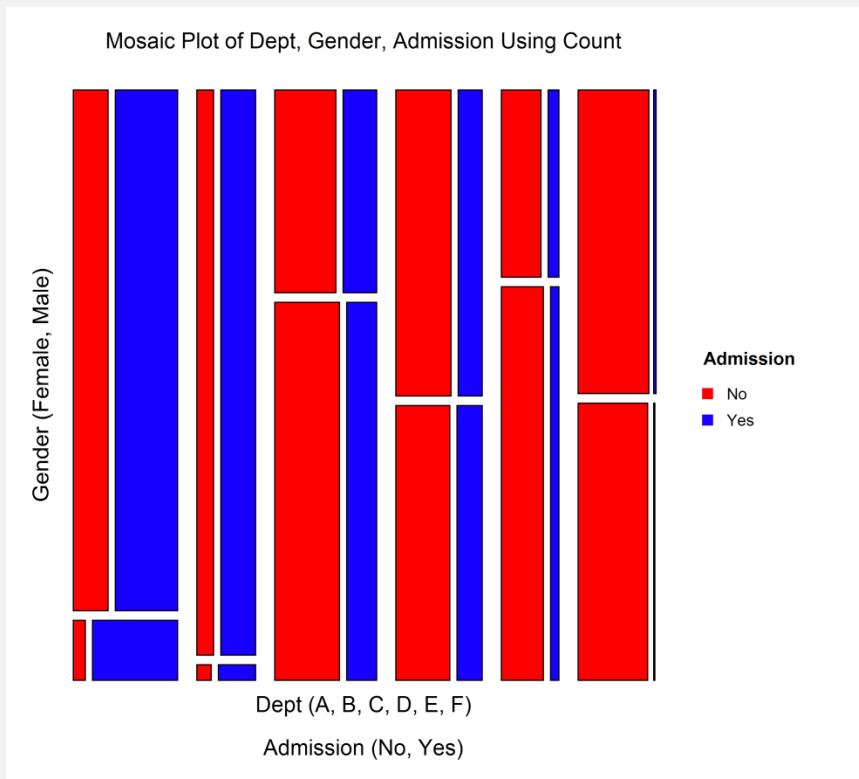


## Output Report

### Data Table Report

Admission	Gender	Dept	Actual
No	Female	A	19
No	Female	B	8
No	Female	C	391
No	Female	D	244
No	Female	E	299
No	Female	F	317
No	Male	A	313
No	Male	B	207
No	Male	C	205
No	Male	D	279
No	Male	E	138
No	Male	F	351
Yes	Female	A	89
Yes	Female	B	17
Yes	Female	C	202
Yes	Female	D	131
Yes	Female	E	94
Yes	Female	F	24
Yes	Male	A	512
Yes	Male	B	353
Yes	Male	C	120
Yes	Male	D	138
Yes	Male	E	53
Yes	Male	F	22

### Dept, Gender, Admission Using Count Mosaic Plot



The Data Table Report gives the summarized data from which the percentages are calculated. The Mosaic plot is displayed next.