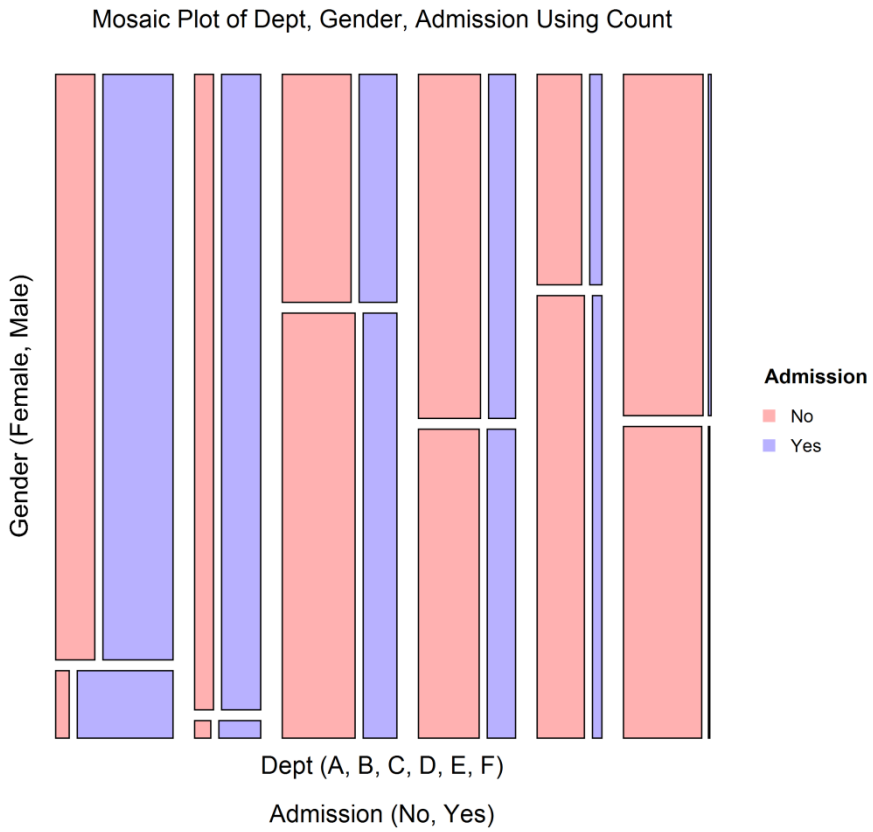
Chapter 147

Mosaic Plots

Introduction

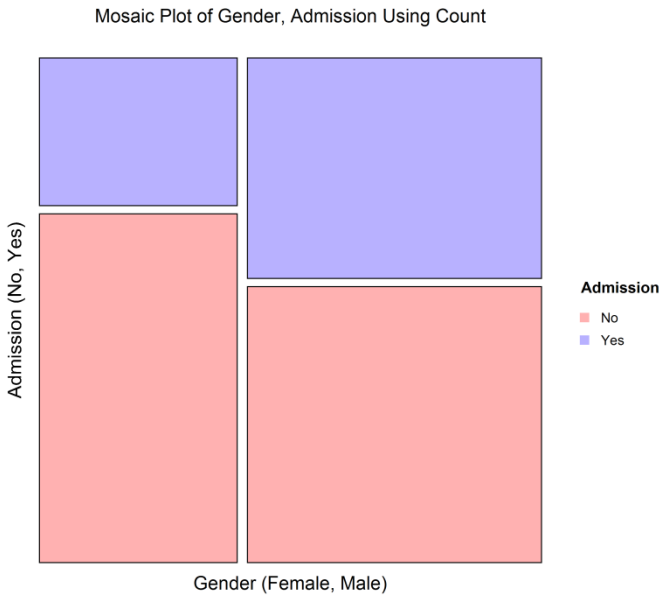
A mosaic plot is a graphical display of the cell frequencies of a contingency table in which the area of boxes of the plot are proportional to the cell frequencies of the contingency table. This procedure can construct mosaic plots for up to four-way contingency tables.

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.



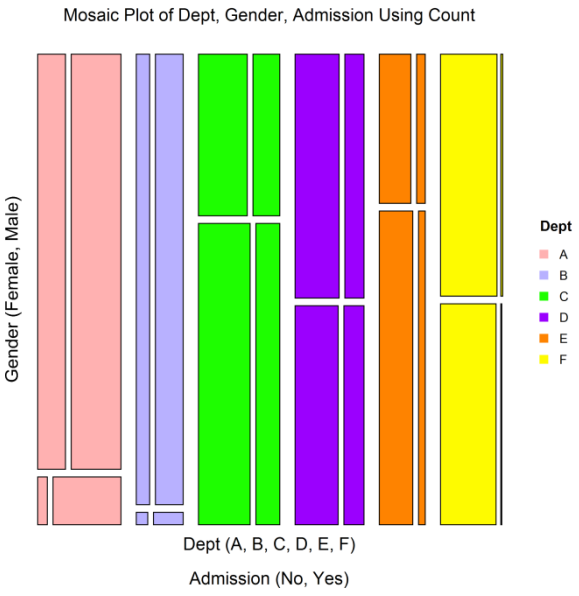
The widths of the boxes are proportional to the percentage of females and males, respectively. In fact, 41% of applicants were female and 59% were male.

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.

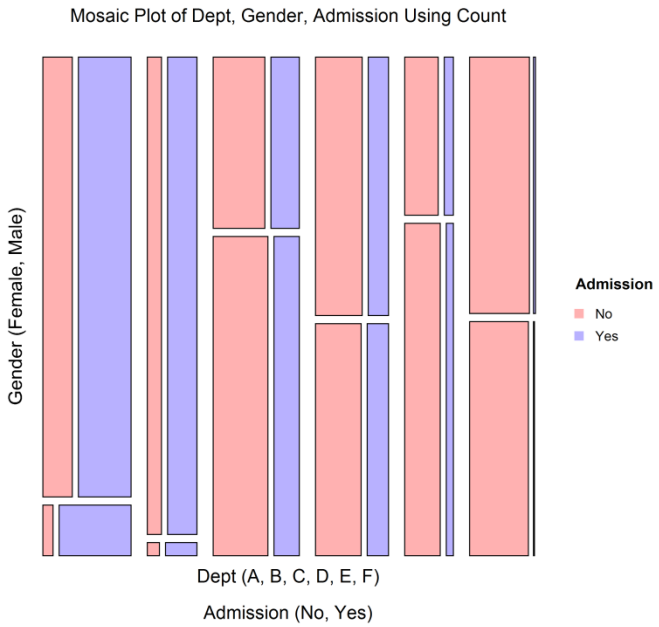
It is easy to see that females' blue box on the left is much shorter than the males' blue box on the right.

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.

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 show 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 to Interpretation of Mosaic Plots

1. The categories of each new factor divide each box either horizontally (1st and 3rd factor) or vertically (2nd and 4th 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
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

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.

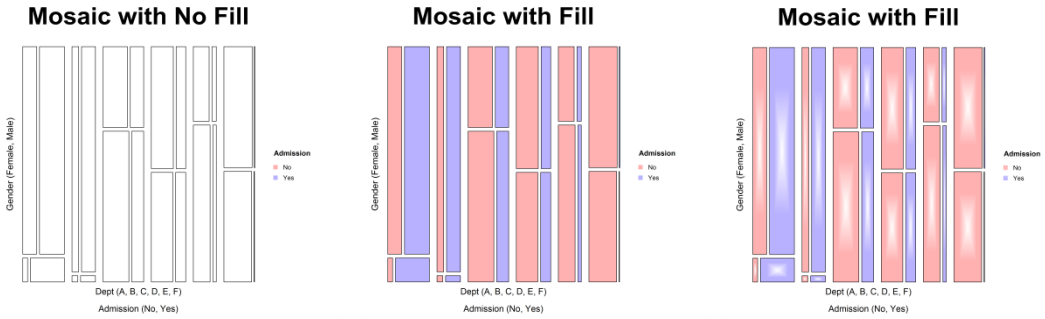
Mosaic Plot Tab

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

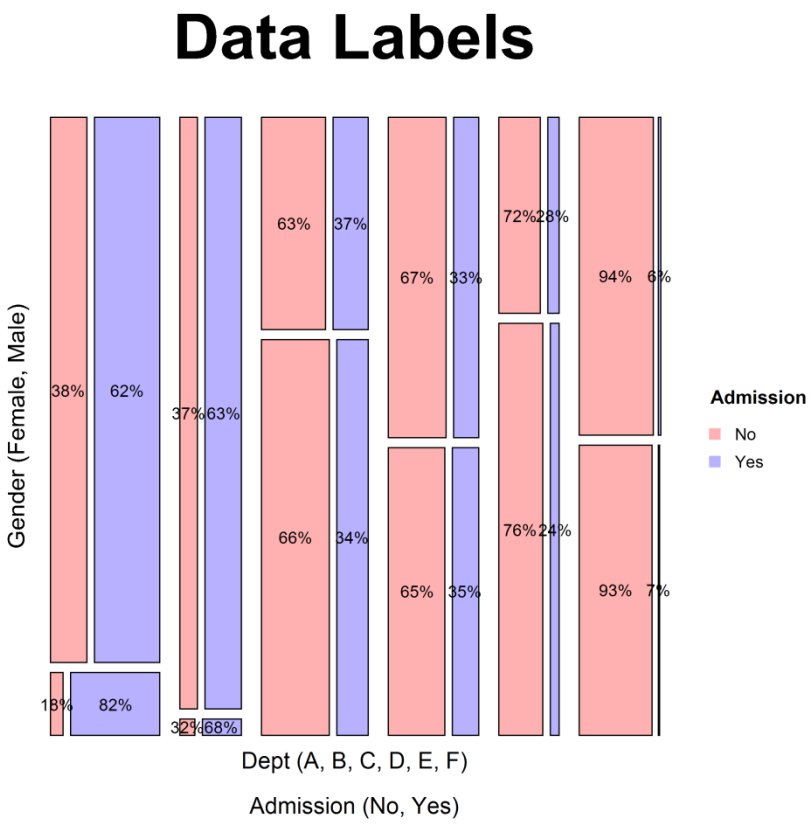
This option controls the colors of the borders. The colors are applied according to the colorizing factor, which is usually the last factor specified.

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.

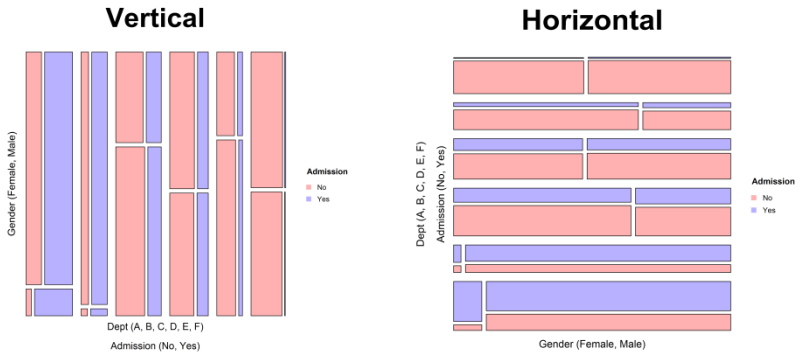


Position

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

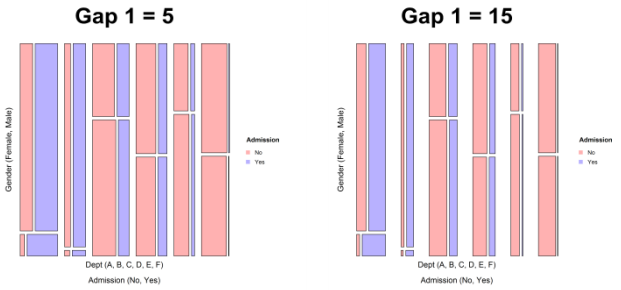
Orientation Section

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



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.

Setup

To run this example, complete the following steps:

1 Open the Berkeley 1973 Admissions example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select Berkeley 1973 Admissions and click OK.

2 Specify the Mosaic Plots procedure options

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

Variables Tab	
Variable 1	**Dept**
Variable 2	**Gender**
Variable 3	**Admission**
Frequencies	**Count**
Data Summary Report	**Checked**

3 Run the procedure

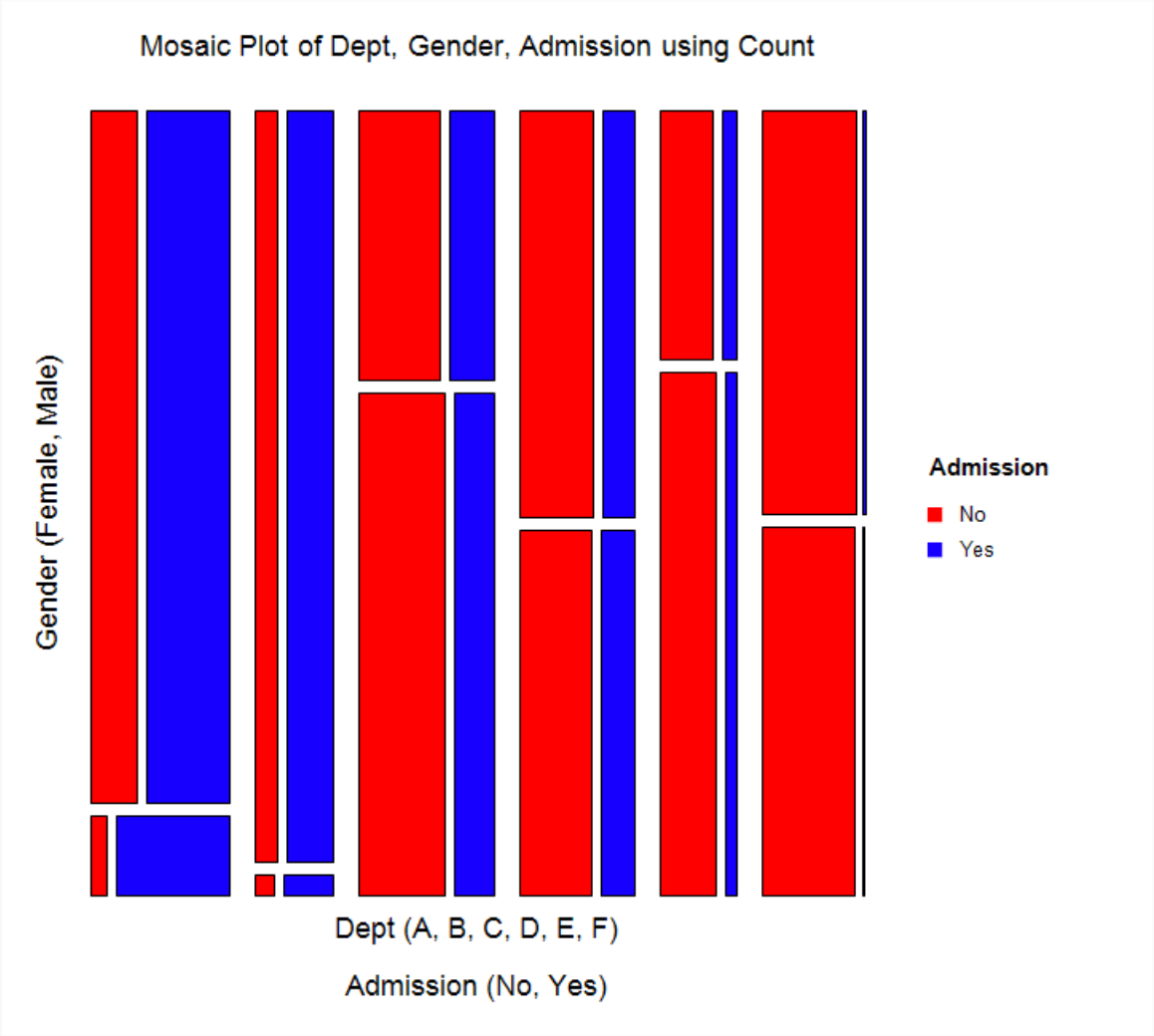
- Click the **Run** button to perform the calculations and generate the output.

Output

Data Summary

Admission	Gender	Dept	Count
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

Mosaic Plot of Dept, Gender, Admission using Count



The Data Summary displays the data from which the percentages are calculated. The Mosaic plot is displayed next.