

Venn and Euler Diagrams



Revised: 8/11/2019



Summary	1
Data Input.....	3
Tables and Graphs.....	4
Analysis Summary	4
R Diagram.....	5
Statgraphics Diagram.....	6
Diagnostics.....	7
References.....	8

Summary

The **Venn and Euler Diagrams** procedure creates diagrams that display the relative frequency of occurrence of discrete events. They consist of circular regions that represent the frequency of specific events, where the overlap of the circles indicates the simultaneous occurrence of more than one event.

The diagram is created by the “VennEuler” package in R. To run the procedure, R must be installed on your computer together with that package. For information on downloading and installing R, refer to the document titled “R – Installation and Configuration”.

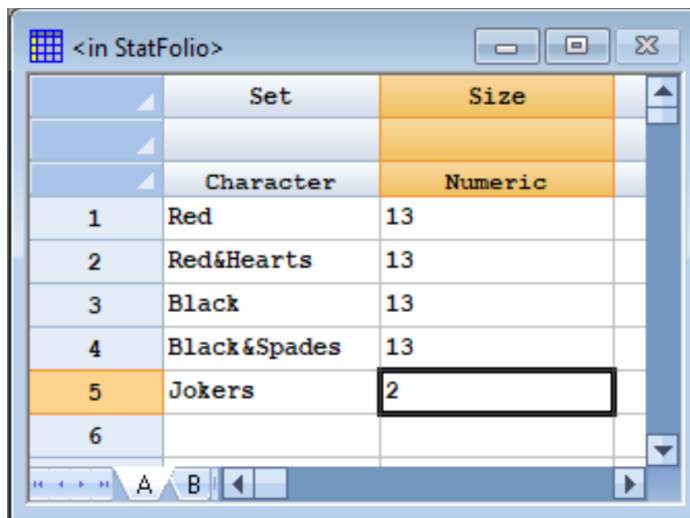
Sample StatFolio: *venneuler.sgp*

Sample Data

A Venn diagram will be created to represent the members of a standard deck of cards. The deck consists of:

- 13 red cards in the suit of diamonds
- 13 red cards in the suit of hearts
- 13 black cards in the suit of clubs
- 13 black cards in the suit of spades
- 2 jokers that do not belong to any suit or have any color

To represent the members of the deck, the following data set will be created:

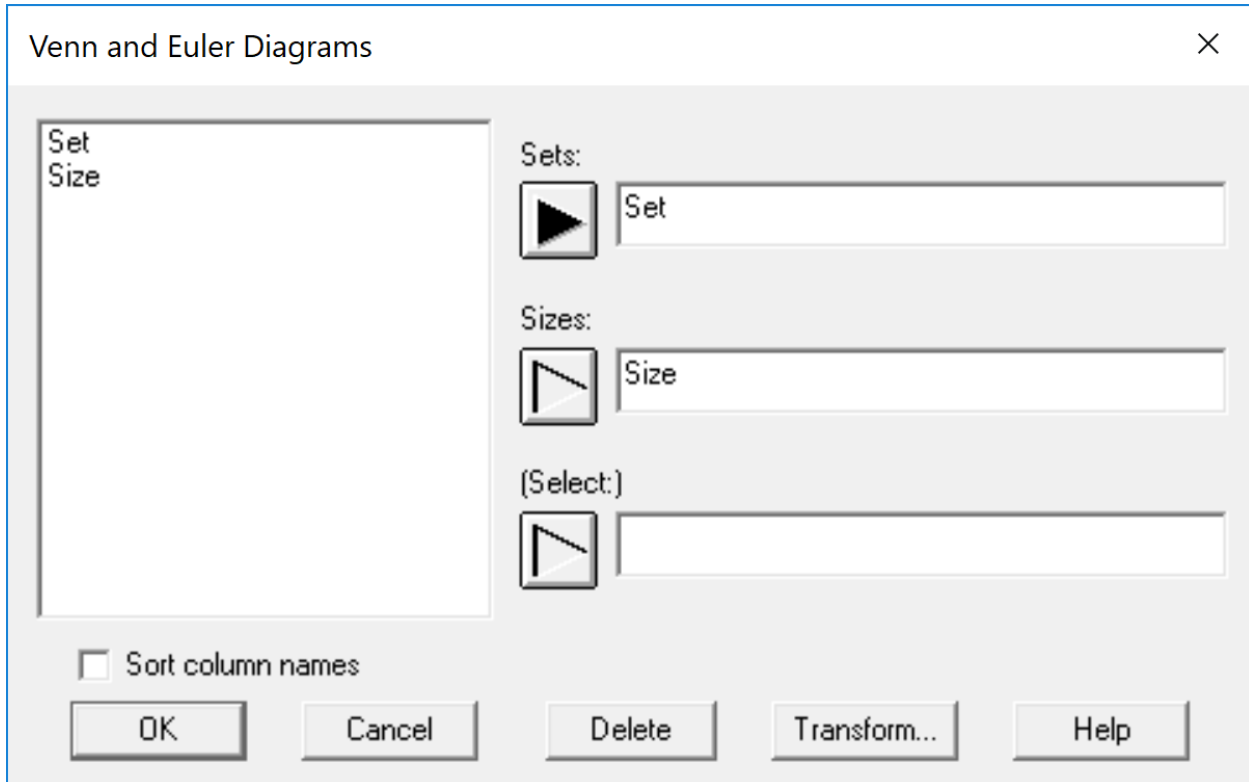


	Set	Size
	Character	Numeric
1	Red	13
2	Red&Hearts	13
3	Black	13
4	Black&Spades	13
5	Jokers	2
6		

A row without an ampersand (&) such as “Red” indicates how many members belong to that set but no others. A row with an ampersand such as “Red&Hearts” indicates how many members belong to both sets but no others. Members belonging to 3 sets would be indicated by an expression such as “A&B&C”.

Data Input

When the procedure is first selected, a data input dialog box is displayed requesting the names of columns identifying each set and the frequency associated with each set:



- **Sets:** the name of a character column identifying each set or intersection of sets to be included in the diagram. Valid entries includes strings such as “A”, “A&B”, “A&B&C”, and so forth where A, B, and C represent distinct characteristics.
- **Sizes:** numeric column indicating the size of each set. Sizes may be absolute (counts) or relative (proportions).
- **(Select:)** optional subset selection.

Tables and Graphs

The following tables and graphs may be created:

Tables and Graphs		×
TABLES	GRAPHS	<input type="button" value="OK"/>
<input checked="" type="checkbox"/> Analysis Summary	<input checked="" type="checkbox"/> Statgraphics Diagram	<input type="button" value="Cancel"/>
<input checked="" type="checkbox"/> R Diagram		<input type="button" value="All"/>
<input checked="" type="checkbox"/> Diagnostics		<input type="button" value="Store"/>
		<input type="button" value="Help"/>

Analysis Summary

The *Analysis Summary* displays the R commands that were executed.

```
Venn and Euler Diagrams

d<-
read.csv("C:\\\\Users\\NEIL~1.STA\\AppData\\Local\\Temp\\sets.csv",dec=".",
sep=",")
setwd("C:\\Users\\NEIL~1.STA\\AppData\\Local\\Temp\\")
str(d)

## 'data.frame':    5 obs. of  2 variables:
## $ set : Factor w/ 5 levels "Black","Black&Spades",...: 4 5 1 2 3
## $ size: int  13 13 13 13 2

library("venneuler")

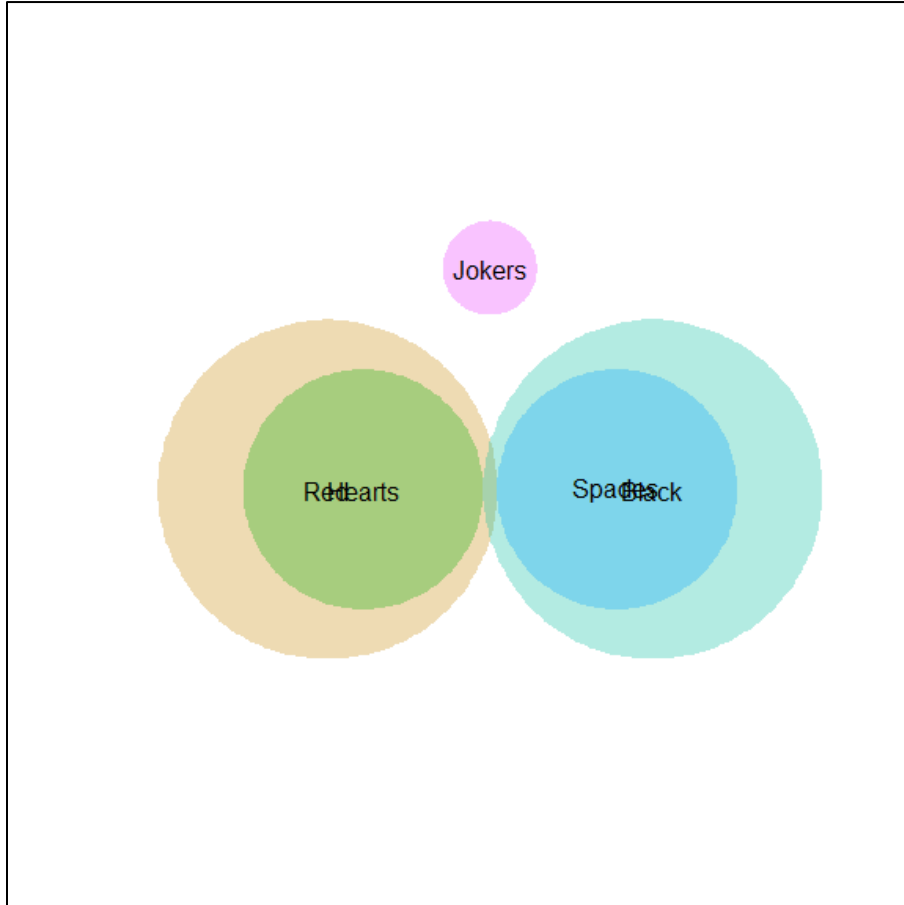
## Loading required package: rJava

library("rJava")
vd<-venneuler(as.character(d$set),d$size)
plot(vd)

#return diagram information
write(vd$centers,file="C:\\Users\\NEIL~1.STA\\AppData\\Local\\Temp\\circles.csv",sep=
",",ncolumns=1)
write(vd$residuals,file="C:\\Users\\NEIL~1.STA\\AppData\\Local\\Temp\\regions.csv",se
p=",",ncolumns=1)
write(vd$diameters,file="C:\\Users\\NEIL~1.STA\\AppData\\Local\\Temp\\diameters.csv",
sep=",",ncolumns=1)
write(vd$colors,file="C:\\Users\\NEIL~1.STA\\AppData\\Local\\Temp\\colors.csv",sep="
",ncolumns=1)
write(vd$labels,file="C:\\Users\\NEIL~1.STA\\AppData\\Local\\Temp\\labels.csv",sep="
",ncolumns=1)
write(names(vd$residuals),file="C:\\Users\\NEIL~1.STA\\AppData\\Local\\Temp\\rlabels.
csv",sep=",",ncolumns=1)
```

R Diagram

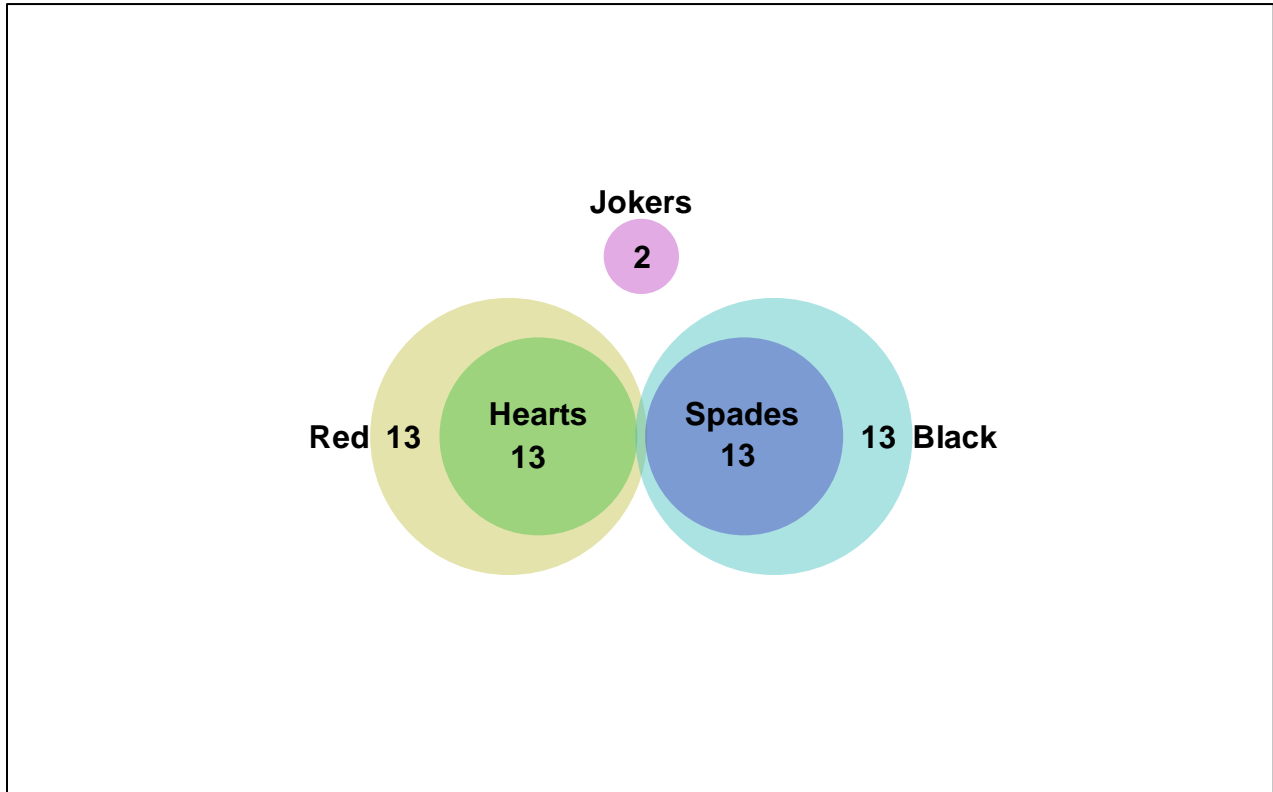
This pane shows the diagram as generated by R:



Each circle represents one of the input sets. The size of the circles is scaled to represent the relative size of each set (to the extent possible). Since this diagram is generated by R, it cannot be edited.

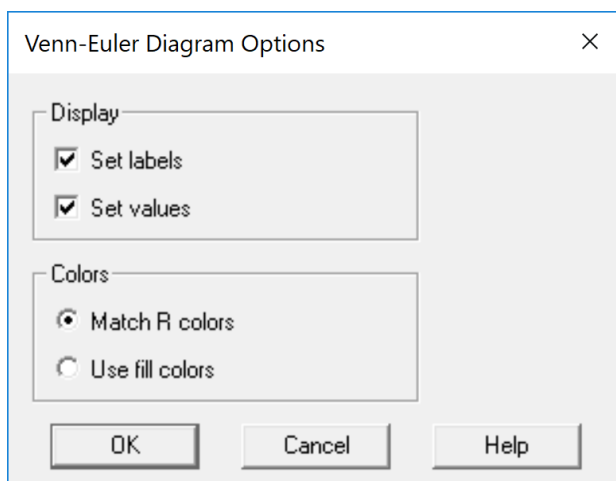
Statgraphics Diagram

This graph redraws the diagram using the Statgraphics graphics facilities:



The graph labels each of the circles. It also shows how often events occur in each section of the diagram. Labels may be relocated by dragging them to their desired positions.

Pane Options



- **Display:** whether to display labels for each set and/or the values associated with each main set and intersection.
- **Colors:** whether to match the colors used in the R diagram or use the standard Statgraphics fill colors (which may be modified using Graphics Options).

Diagnostics

This pane shows how well the relative sizes of each set have been represented:

Diagnostics	
<i>Set</i>	<i>Residual</i>
Jokers	-0.0657327
Spades	0.00495667
Spades&Jokers	0.0
Black	-0.0985805
Black&Jokers	0.0
Black&Spades	0.103637
Black&Spades&Jokers	0.0
Hearts	0.00495667
Hearts&Jokers	0.0
Hearts&Spades	4.32223E-78
Hearts&Spades&Jokers	0.0
Hearts&Black	0.000970537
Hearts&Black&Jokers	0.0
Hearts&Black&Spades	0.000901213
Hearts&Black&Spades&Jokers	0.0
Red	-0.0985805
Red&Jokers	0.0
Red&Spades	0.000970537
Red&Spades&Jokers	0.0
Red&Black	0.594154
Red&Black&Jokers	0.0
Red&Black&Spades	0.00039938
Red&Black&Spades&Jokers	0.0
Red&Hearts	0.103637
Red&Hearts&Jokers	0.0
Red&Hearts&Spades	0.000901213
Red&Hearts&Spades&Jokers	0.0
Red&Hearts&Black	0.00039938
Red&Hearts&Black&Jokers	0.0
Red&Hearts&Black&Spades	0.00187175
Red&Hearts&Black&Spades&Jokers	0.0

The *Residual* column shows the percentage difference between the actual size of each set and its size as represented in the diagram. For example, “Black&Spades” has a residual value equal to 0.103637, which indicates that the circle drawn for that intersection is about 0.1% larger than it should be. Note that it is rarely possible to draw each region in exact proportion to how frequently it occurs. In this case, all residuals are less than 1%.

References

R Package “venneuler” (2015) - <https://cran.r-project.org/web/packages/venneuler/venneuler.pdf>

Wilkinson, Leland (2012) “Exact and Approximate Area-proportional Circular Venn and Euler Diagrams”. IEEE Transactions on Visualization and Computer Graphics. 18(2):321-31.