

Visualizing Non-Hierarchical Cluster Allocation Using Alluvial Diagrams in R

Joseph Retzer



R-Users Chicago Meet-Up

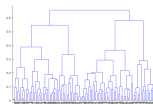
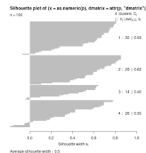
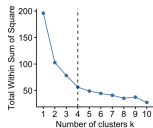
June 16th 2021

“ Visualization is daydreaming with a purpose.”

Bo Bennett

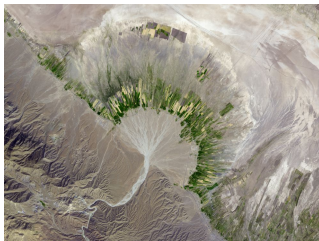
Choosing the Number of Segments

- Elbow plot (numeric data)
- Quality measures (e.g. silhouette plot)
- Hierarchical clustering dendrogram
What if algorithm used is not hierarchical?



Alluvial Diagrams

- Alluvial diagrams are a type of flow diagram originally developed to represent changes in network structure over time.
- Alluvial plots are similar to Sankey diagrams and visualise categorical data over multiple dimensions as flows.
- In allusion to both their visual appearance and their emphasis on flow, alluvial diagrams are named after alluvial fans that are naturally formed by the soil deposited from streaming water.



An alluvial fan is a triangle-shaped deposit of gravel, sand, and smaller materials called alluvium.

Visualize Multiple Cross-Tabulations

- **Titanic Data Set:** Visualize the following cross tabs simultaneously

	Male	Female
1 st	180	145
2 nd	179	106
3 rd	510	196
Crew	862	23

	Child	Adult
Male	64	1667
Female	45	425

- **Partitions Data Set:** Visualize the following cross tabs simultaneously

	1	2	3	4
1	5823	0	17	3019
2	0	4893	32	2472
3	39	0	5124	75

	1	2	3	4	5
1	5732	0	68	0	62
2	0	3989	0	4	900
3	2	5	4225	1	940
4	71	4	1	4138	1352

	1	2	3	4	5	6
1	3396	0	98	0	4	2307
2	0	3907	71	12	8	0
3	0	0	4272	0	5	17
4	1	0	12	3598	1	531
5	0	1	30	38	3185	0

etc.

R Package: easyalluvial

- 'easyalluvial' provides a simple interface that allows the creation of a reasonable / interpretable alluvial plot from any data frame in either long or wide format from a single line of code.
- It also handles continuous data, i.e., numerical variables are scaled, centered and Yeo-Johnson transformed and then binned.
- It is meant to allow a quick visualisation of entire data frames with a focus on different colouring options.

Alluvial Diagrams in R (easyalluvial)

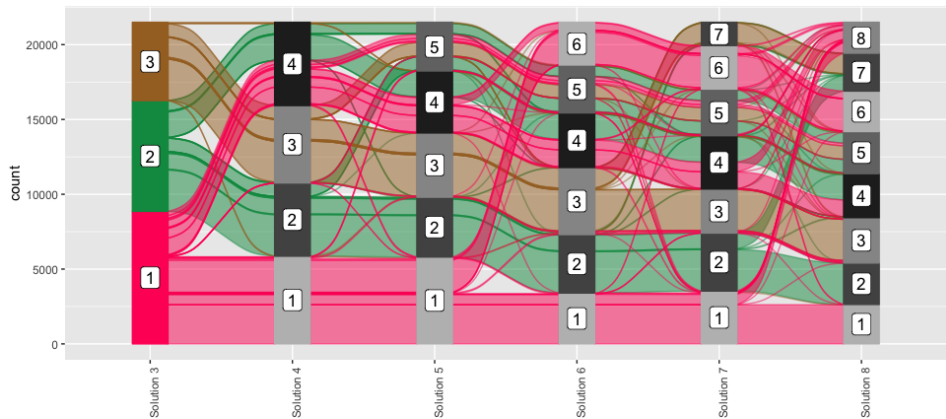
```
> head(myParts)
  Solution 3 Solution 4 Solution 5 Solution 6 Solution 7 Solution 8
1         3         3         3         3         3         3
2         1         4         4         4         4         8
3         1         1         1         1         1         1
4         1         4         4         4         4         8
5         1         4         5         5         1         8
6         3         3         3         3         7         7
```

(Wide-form, each column in graph is a column in the data table)

R-code for easyalluvial plot

```
1 library(easyalluvial)
2 alluvial_wide( data = myParts
3               , max_variables = 6
4               , fill_by = 'first_variable' )
```

easyalluvial Output



Number of flows: 107
Original Dataframe reduced to 0.5 %
Maximum weight of a single flow 12.9 %

R Package: ggalluvial

- The ggalluvial package is a ggplot2 extension for producing alluvial plots in a tidyverse framework.
- The 'ggalluvial' package translates alluvial plot grammar into 'ggplot2' syntax and therefore provides many options to tweak the appearance of an alluvial plot.
- Data in “lode” (long) form, graph column id's are in a single column in data frame.

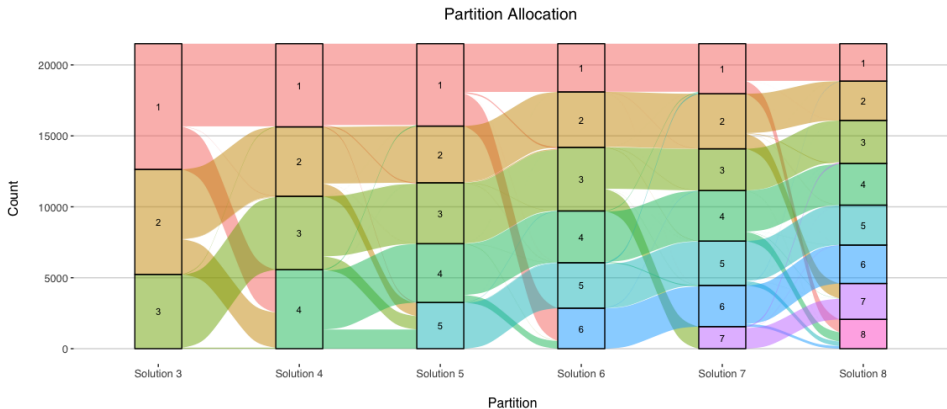
```
> head(myPartsGG)
  respondent variable value
1          1 Solution 3    3
2          2 Solution 3    1
3          3 Solution 3    1
4          4 Solution 3    1
5          5 Solution 3    1
6          6 Solution 3    3
```


Alluvial Diagrams in R (ggalluvial)

R-code for ggalluvial plot

```
1 library(ggalluvial)
2 library(ggthemes)
3
4 ggplot(myPartsGG,
5       aes(x=variable, stratum=value, alluvium=respondent,
6           fill=value, label=value)) +
7   scale_x_discrete(expand = c(.1,.1)) +
8   geom_flow() +
9   labs(x = "\n Partition", y = "Count \n") +
10  geom_stratum(alpha = .5) +
11  geom_text(stat = "stratum", size = 3) +
12  theme_hc() +
13  ggtitle("Partition Allocation") +
14  theme(legend.position="none",
15        plot.title=element_text(hjust=.5))
```

ggalluvial Output



Titanic Alluvial Diagram

