

Sankey Plots (or Ribbon Plots)

using the
Riverplots package

EdinbR User Group

15 Nov 2017

Nevil Hopley

Newbie Notes

- R user since January 2017
- Teaching in Secondary Schools since 1993
- Teaching S6 AH Statistics since 1999
- Previous EdinbR Talks:
 - Breakdown Plots (June 2017)
 - Postcodes on Maps (Sep 2017)

Nevil Hopley

Sources of Code and Help

<http://www.exegetic.biz/blog/2014/08/plotting-flows-with-riverplot/>

<https://www.rdocumentation.org/packages/riverplot/versions/0.5/topics/makeRiver>

<https://stackoverflow.com/questions/9052189/r-how-to-reshape-a-table-into-vectors>

R Packages Used

riverplot

reshape2
(for 'melt' function)

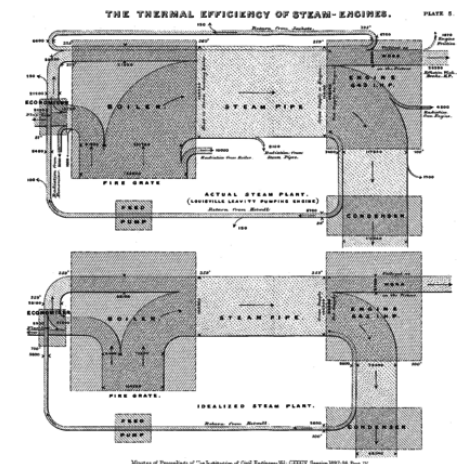
RColorBrewer

A Brief Wiki-History

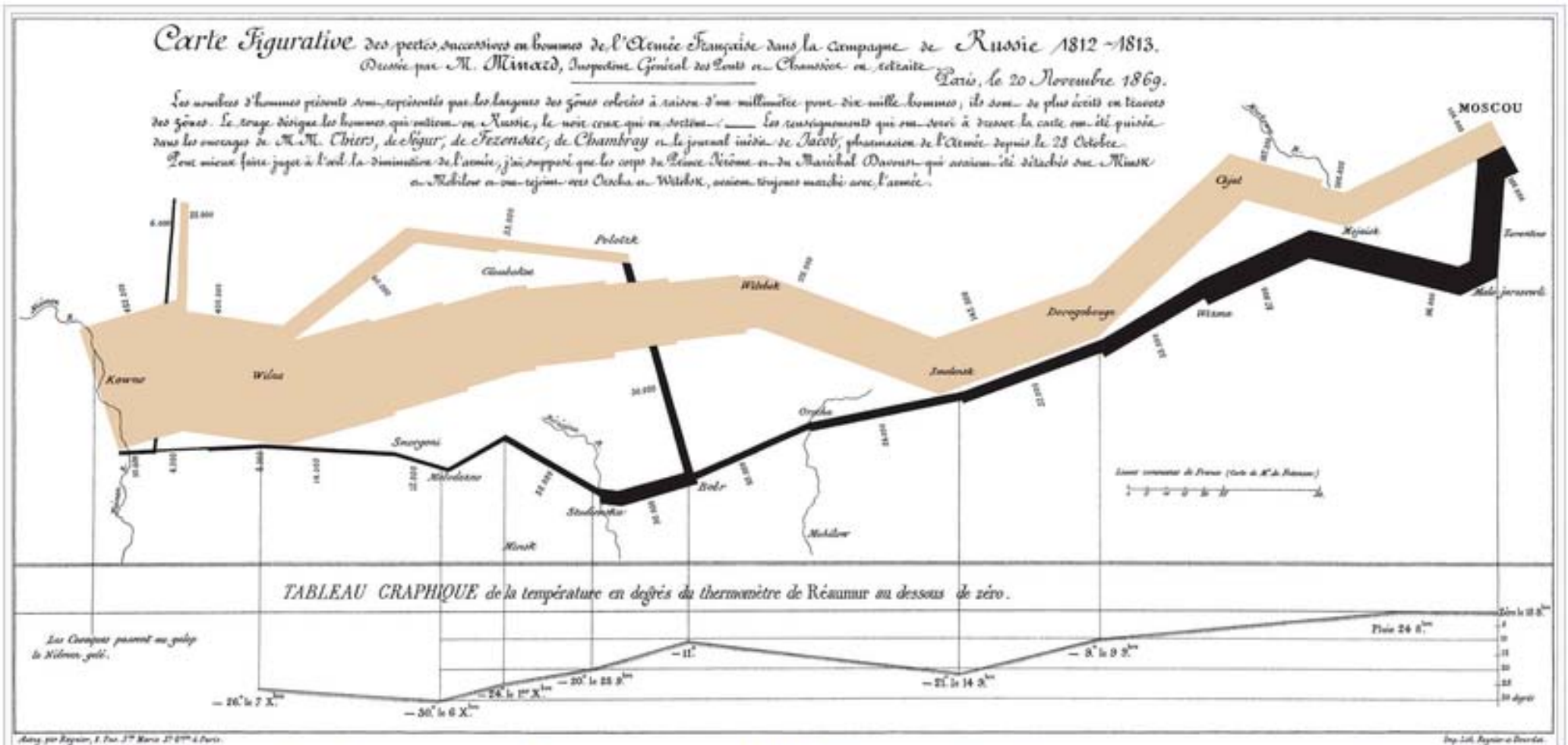
Sankey diagrams are a specific type of **flow diagram**, in which the width of the arrows is shown proportionally to the flow quantity.

Sankey diagrams are named after Irish Captain **Matthew Henry Phineas Riall Sankey**, who used this type of **diagram** in 1898 in a classic figure (see panel on the right) showing the **energy efficiency** of a **steam engine**. While the first charts in black and white were merely used to display one type of flow (e.g. steam), using colors for different types of flows has added more degrees of freedom to Sankey diagrams.

One of the most famous Sankey diagrams is **Charles Minard's** Map of Napoleon's Russian Campaign of 1812. It is a **flow map**, overlaying a Sankey diagram onto a geographical map. It was created in 1869, so it actually predates Sankey's 'first' Sankey diagram of 1898.



Charles Joseph Minard (French: [minard](#); 27 March 1781 – 24 October 1870) was a French [civil engineer](#) recognized for his significant contribution in the field of [information graphics](#) in civil engineering and statistics. Minard was, among other things, noted for his representation of numerical data on geographic maps.



Charles Minard's map of [Napoleon's disastrous Russian campaign of 1812](#). The graphic is notable for its representation in two dimensions of six types of data: the number of Napoleon's troops; distance; temperature; the latitude and longitude; direction of travel; and location relative to specific dates.^[2]

Key to SQA Exam Abbreviations

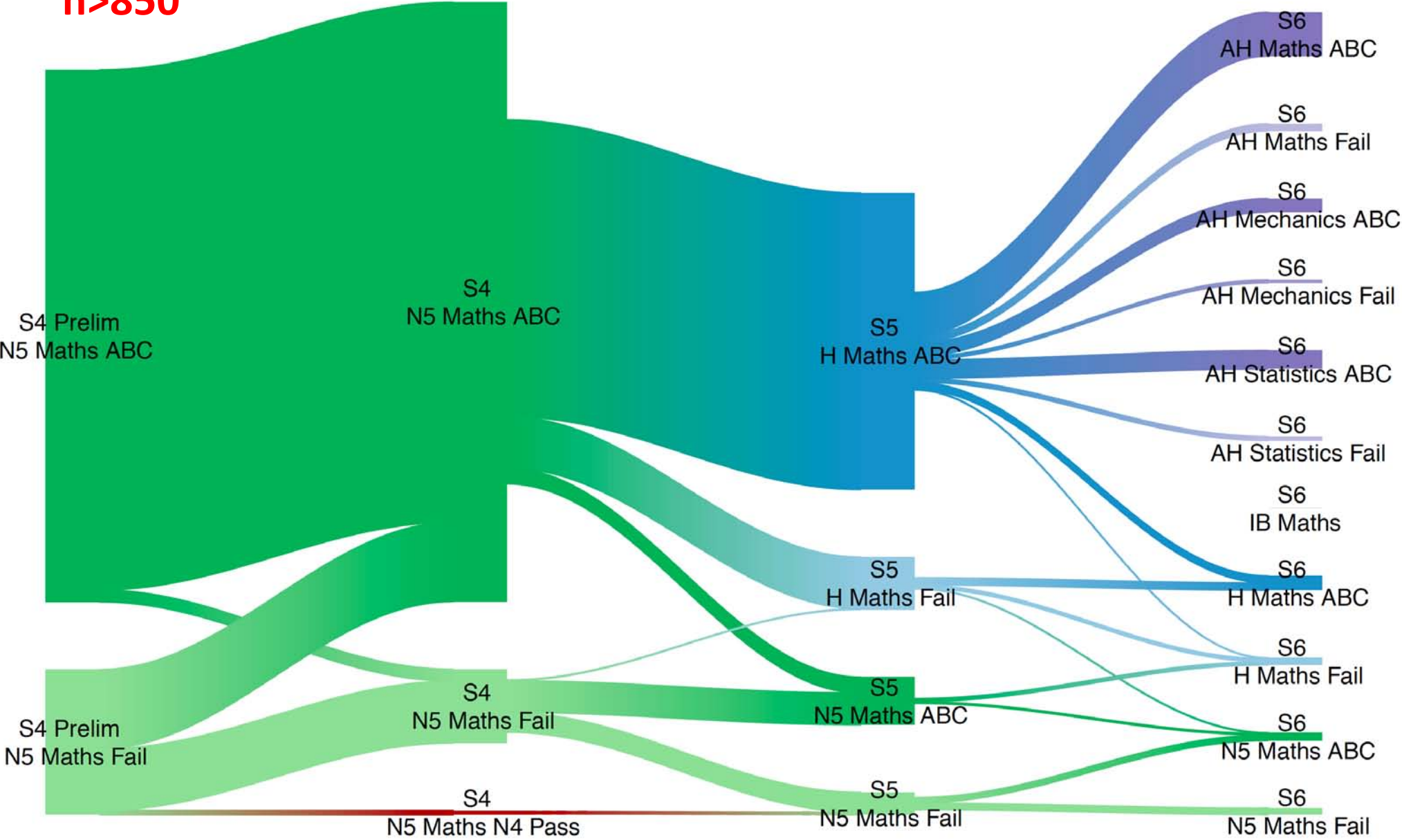
N5 = National 5

H = Higher

AH = Advanced Higher

<u>Grade</u>	<u>Band</u>	<u>Typical %</u>
A	A1	85% or more
A	A2	70%-84%
B	B3	65%-69%
B	B4	60%-64%
C	C5	55%-59%
C	C6	50%-54%
D	D7	45%-49%
NA	NA8 or NA9	44% or less
wd	Withdrawn from Exam	

n>850



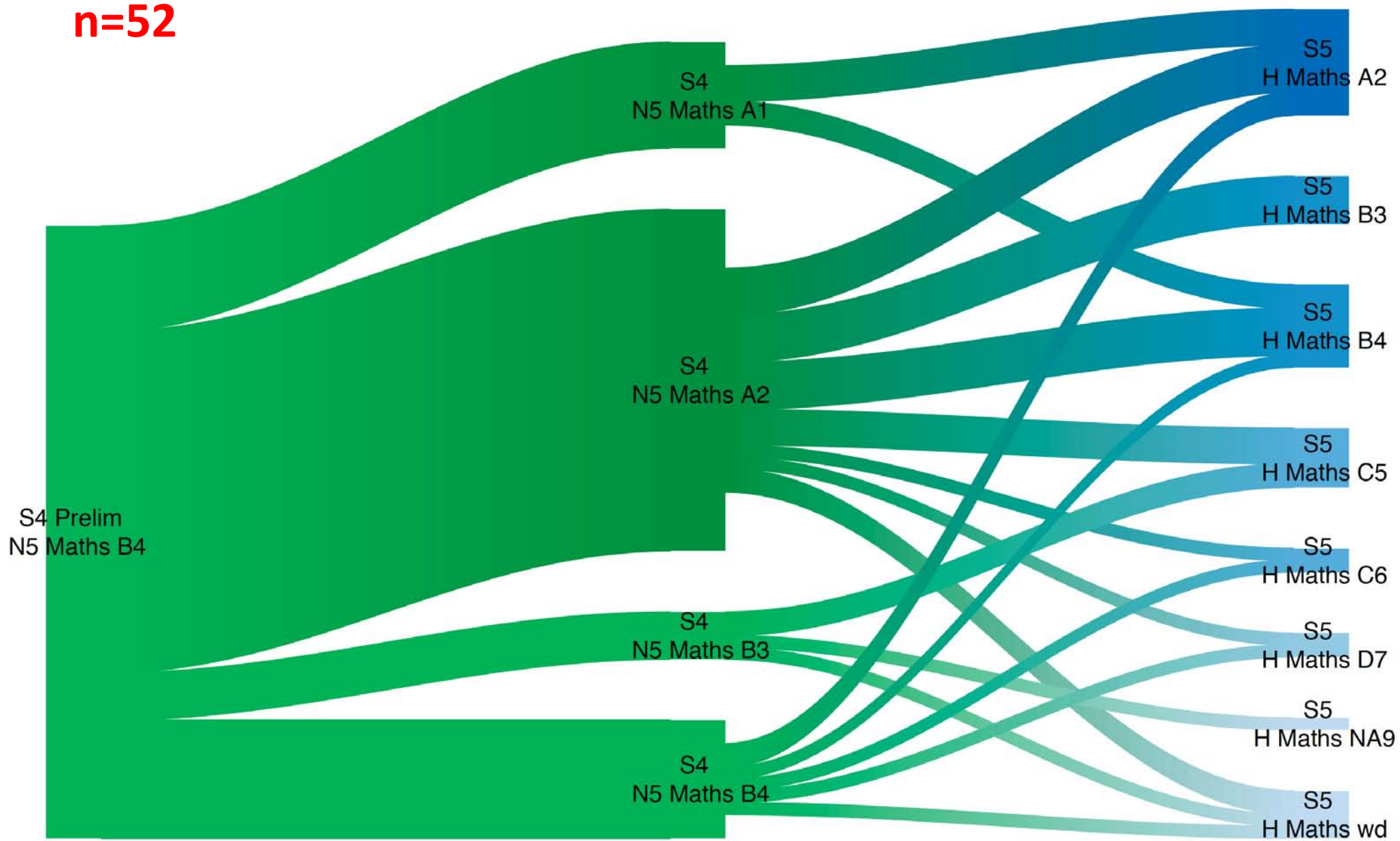
node	colour_expression	y_pos
H Maths A1	brewer.pal(9,'Blues')[7]	20.1
H Maths A	brewer.pal(9,'Blues')[7]	20
H Maths ABC	brewer.pal(9,'Blues')[6]	20
H Maths A2	brewer.pal(9,'Blues')[7]	19.9
H Maths B3	brewer.pal(9,'Blues')[6]	19.1
H Maths B	brewer.pal(9,'Blues')[6]	19
H Maths B4	brewer.pal(9,'Blues')[6]	18.9
H Maths C5	brewer.pal(9,'Blues')[5]	18.1



N5 Maths C5	brewer.pal(9,'Greens')[5]	11.1
N5 Maths C	brewer.pal(9,'Greens')[5]	11
N5 Maths C6	brewer.pal(9,'Greens')[5]	10.9
N5 Maths D	brewer.pal(9,'Greens')[4]	10
N5 Maths D7	brewer.pal(9,'Greens')[4]	10
N5 Maths Fail	brewer.pal(9,'Greens')[4]	10
N5 Maths NA8	brewer.pal(9,'Greens')[3]	9.1
N5 Maths NA	brewer.pal(9,'Greens')[3]	9



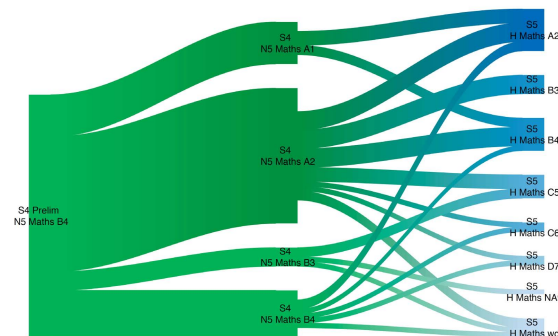
n=52



```

riverplot(r_plot,
  srt = 0, #angle of rotation of node labels, in degrees
  textcex = 1, # reduce node font size
  default_style = NULL,
  gravity = "bottom",
  # node_margin = 0.05,
  nodewidth = 0.5,
  plot_area = 0.9,
  mar = NULL, # leave margins untouched
  # nsteps = 100, # number of polygons to use
  # add_mid_points = NULL, # default setting for
  yscale = "auto",
  fix.pdf = TRUE
)

```



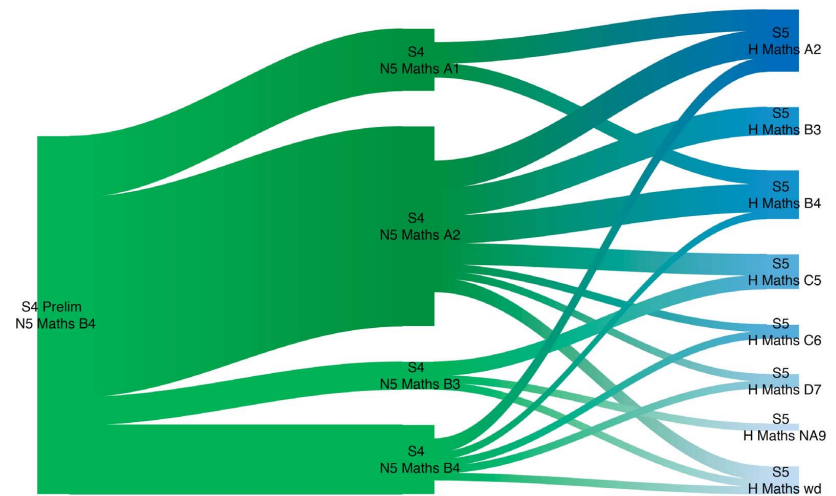
```
r_plot <- makeRiver(nodes, edges)
```

Description

Create a new riverplot object

Usage

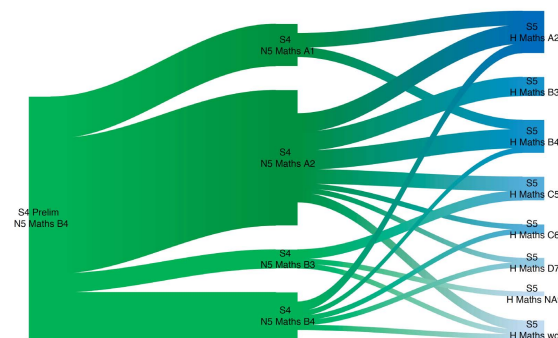
```
makeRiver(nodes, edges, node_labels = NULL, node_xpos = NULL,  
  node_ypos = NULL, node_styles = NULL, edge_styles = NULL,  
  default_style = NULL)
```



```
> class(r_plot)  
[1] "list"      "riverplot"
```

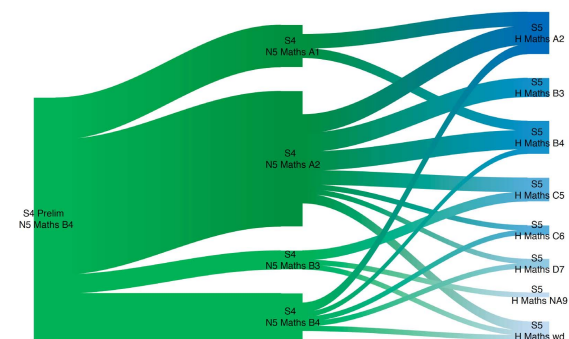

> nodes

		ID	x	col		labels	my_order
4		S4 \n N5 Maths B4	13	#41AB5D		S4 \n N5 Maths B4	11.9
5	S4 Prelim \n N5 Maths B4	12	#41AB5D		S4 Prelim \n N5 Maths B4		11.9
3		S4 \n N5 Maths B3	13	#41AB5D		S4 \n N5 Maths B3	12.1
2		S4 \n N5 Maths A2	13	#238B45		S4 \n N5 Maths A2	12.9
1		S4 \n N5 Maths A1	13	#238B45		S4 \n N5 Maths A1	13.1
13		S5 \n H Maths wd	14	#C6DBEF		S5 \n H Maths wd	15.8
12		S5 \n H Maths NA9	14	#C6DBEF		S5 \n H Maths NA9	15.9
11		S5 \n H Maths D7	14	#9ECAE1		S5 \n H Maths D7	17.0
10		S5 \n H Maths C6	14	#6BAED6		S5 \n H Maths C6	17.9
9		S5 \n H Maths C5	14	#6BAED6		S5 \n H Maths C5	18.1
8		S5 \n H Maths B4	14	#4292C6		S5 \n H Maths B4	18.9
7		S5 \n H Maths B3	14	#4292C6		S5 \n H Maths B3	19.1
6		S5 \n H Maths A2	14	#2171B5		S5 \n H Maths A2	19.9



> edges

				N1		N2	Value
32		S4	\n	N5 Maths B4	S5	\n H Maths wd	1
31		S4	\n	N5 Maths B3	S5	\n H Maths wd	1
30		S4	\n	N5 Maths A2	S5	\n H Maths wd	2
27		S4	\n	N5 Maths B3	S5	\n H Maths NA9	1
24		S4	\n	N5 Maths B4	S5	\n H Maths D7	1
22		S4	\n	N5 Maths A2	S5	\n H Maths D7	1
20		S4	\n	N5 Maths B4	S5	\n H Maths C6	1
18		S4	\n	N5 Maths A2	S5	\n H Maths C6	1
15		S4	\n	N5 Maths B3	S5	\n H Maths C5	2
14		S4	\n	N5 Maths A2	S5	\n H Maths C5	3
12		S4	\n	N5 Maths B4	S5	\n H Maths B4	1
10		S4	\n	N5 Maths A2	S5	\n H Maths B4	4
9		S4	\n	N5 Maths A1	S5	\n H Maths B4	2
6		S4	\n	N5 Maths A2	S5	\n H Maths B3	4
41		S4	\n	N5 Maths B4	S5	\n H Maths A2	2
21		S4	\n	N5 Maths A2	S5	\n H Maths A2	4
11		S4	\n	N5 Maths A1	S5	\n H Maths A2	3
4	S4 Prelim	\n	N5 Maths B4	S4	\n N5 Maths B4		10
3	S4 Prelim	\n	N5 Maths B4	S4	\n N5 Maths B3		4
2	S4 Prelim	\n	N5 Maths B4	S4	\n N5 Maths A2		29
1	S4 Prelim	\n	N5 Maths B4	S4	\n N5 Maths A1		9



Contingency Table (ct) for each 'flow'

	S4 \n N5 Maths A1	S4 \n N5 Maths A2	S4 \n N5 Maths B3	S4 \n N5 Maths B4
S4 Prelim \n N5 Maths B4	9	29	4	10

	S5 \n H Maths A2	S5 \n H Maths B3	S5 \n H Maths B4	S5 \n H Maths C5	S5 \n H Maths C6	S5 \n H Maths D7	S5 \n H Maths NA9
S4 \n N5 Maths A1	3	0	2	0	0	0	0
S4 \n N5 Maths A2	4	4	4	3	1	1	0
S4 \n N5 Maths B3	0	0	0	2	0	0	1
S4 \n N5 Maths B4	2	0	1	0	1	1	0

Create Data Frames for Edges and Nodes

```
# melt the table into a data frame of edges
edges_df <- melt(ct, varnames=c("N1", "N2"), value.name="Value")
# remove all rows with 0 frequency or 'NA NA'
edges_df <- subset(edges_df, subset = (Value != 0))
```

```
nodes_df <- data.frame(ID = c(rownames(ct), colnames(ct)))
```

```
> raw_data
```

	S4_Exam_Course_Prelim	S4_Exam_Course_Actual	S5_Exam_Course_Actual
709	S4 Prelim \n N5 Maths B4	S4 \n N5 Maths B3	S5 \n H Maths wd
712	S4 Prelim \n N5 Maths B4	S4 \n N5 Maths A2	S5 \n H Maths B4
721	S4 Prelim \n N5 Maths B4	S4 \n N5 Maths B3	S5 \n H Maths C5
728	S4 Prelim \n N5 Maths B4	S4 \n N5 Maths A2	S5 \n H Maths B4
731	S4 Prelim \n N5 Maths B4	S4 \n N5 Maths A2	S5 \n H Maths C5
744	S4 Prelim \n N5 Maths B4	S4 \n N5 Maths A2	S5 \n H Maths C6
750	S4 Prelim \n N5 Maths B4	S4 \n N5 Maths A2	S5 \n H Maths B3
758	S4 Prelim \n N5 Maths B4	S4 \n N5 Maths A2	S5 \n H Maths C5
760	S4 Prelim \n N5 Maths B4	S4 \n N5 Maths B4	S5 \n H Maths B4
766	S4 Prelim \n N5 Maths B4	S4 \n N5 Maths A2	S5 \n H Maths wd
805	S4 Prelim \n N5 Maths B4	S4 \n N5 Maths B3	S5 \n H Maths NA9
3531	S4 Prelim \n N5 Maths B4	S4 \n N5 Maths B4	S5 \n H Maths A2
3547	S4 Prelim \n N5 Maths B4	S4 \n N5 Maths A2	S5 \n H Maths A2
3557	S4 Prelim \n N5 Maths B4	S4 \n N5 Maths A2	S5 \n H Maths B4
3559	S4 Prelim \n N5 Maths B4	S4 \n N5 Maths A2	S5 \n H Maths B3
3560	S4 Prelim \n N5 Maths B4	S4 \n N5 Maths A2	S5 \n H Maths A2
3575	S4 Prelim \n N5 Maths B4	S4 \n N5 Maths A2	S5 \n H Maths C5
3616	S4 Prelim \n N5 Maths B4	S4 \n N5 Maths A2	S5 \n H Maths D7

```
ct <- table(df$S4_Exam_Course_Prelim, df$S4_Exam_Course_Actual)
```

```
ct <- table(df$S4_Exam_Course_Actual, df$S5_Exam_Course_Actual)
```

The Mouth of the River

Comments?

Questions?

Suggestions?

Criticisms?

Advice?

...all welcome!

Nevil Hopley
nevil@hopley.me