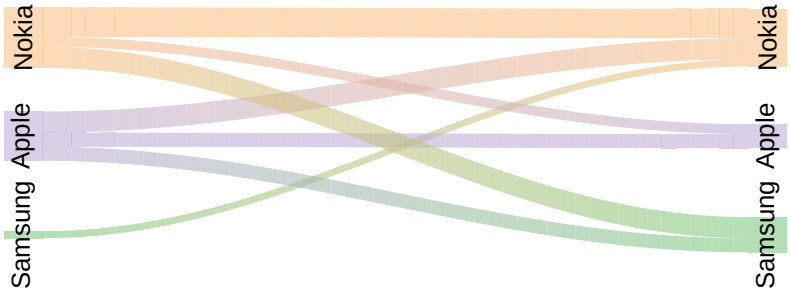
River plots

- A riverplot (also called a Sankey plot) is intended to represent data that flows. But what does that mean?! Sometimes a riverplot can help denote a journey, show changes over time, or can represent the dispersion of resources.
- The code to construct these riverplots is a little cumbersome to work with. All column names must adhere to the format describe below.

```
require(riverplot)
require(RColorBrewer)

nodes<-data.frame(ID=c("Samsung1","Apple1","Nokia1","Samsung2","Apple2","Nokia2"),
                  x=c(0,0,0,10,10,10),
                  y=c(0,5,10,0,5,10),
                  col=c(brewer.pal(3, "Accent"), brewer.pal(3, "Accent")),
                  labels=c("Samsung", "Apple", "Nokia",
                           "Samsung", "Apple", "Nokia"))
nodes$col<-paste(nodes$col, 95, sep="")
edges<-data.frame(N1=c(rep("Samsung1",3),rep("Apple1",3),rep("Nokia1",3)),
                  N2=c("Samsung2","Apple2","Nokia2","Samsung2",
                       "Apple2","Nokia2","Samsung2","Apple2",
                       "Nokia2"),
                  Value=c(0,0,100,200,200,300,300,140,420))
river_data<-makeRiver(nodes, edges)
riverplot(river_data)
text("Cell Phone Brand Loyalty", x=5,y=12) #Adds a title to your exisiting plot
```



- The riverplot can be customized. Look at documentation of 'riverplot' for detailed changes that can be made - including margins, scaled areas, rotation of labels, etc.

- Another example is to consider where tourists within the United Kingdom originate from. There are two tables provided on the site VisitBritain.org that we will use for this example.

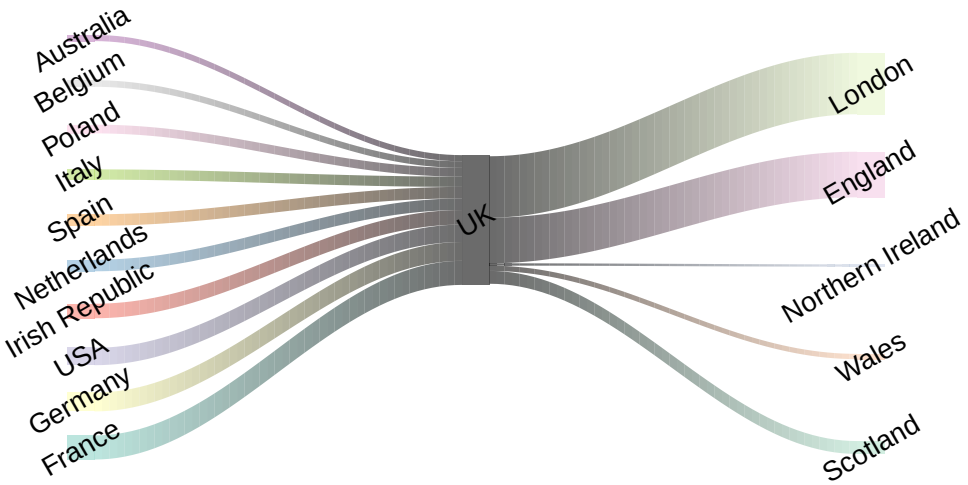
From	Number of Visitors to UK (in thousands)
France	4114
Germany	3220
USA	2976
Irish Republic	2486
Spain	1986
Netherlands	1972
Italy	1757
Poland	1494
Belgium	1122
Australia	1057

Primary Destination of Visitors	Number of Visitors (in thousands)
Scotland	2200
Wales	900
Northern Ireland	400
England	7800
London	10500

- In this example, we do not know how many people from France primarily visited Scotland, or how many people from Australia primarily visited London. We simply know the number of people who visited the UK. And the per region total of visitors.
- Therefore, we must have a plot which first sends all the visitors to the UK and then breaks off into where they visited. This makes the UK a starting location and an ending location.

```
nodes2<-data.frame(ID=c("France","Germany","USA","Irish Republic",
                        "Netherlands","Spain","Italy","Poland",
                        "Belgium","Australia","UK","Scotland",
                        "Wales","Northern Ireland","England",
                        "London"),
                   x=c(rep(0,10),rep(1,1),rep(2,5)),
                   y=c(1:10,6,1,3,5,7,9),
                   col=c(brewer.pal(10,"Set3"),"#000000",
                         brewer.pal(5,"Pastel2")),
                   labels=c("France","Germany","USA","Irish Republic",
                        "Netherlands","Spain","Italy","Poland",
                        "Belgium","Australia","UK","Scotland",
                        "Wales","Northern Ireland","England",
                        "London"))
nodes2$col<-paste(nodes2$col, 95, sep="")
edges2<-data.frame(N1=c("France","Germany","USA","Irish Republic",
                        "Netherlands","Spain","Italy","Poland",
                        "Belgium","Australia",rep("UK",5)),
                   N2=c(rep("UK",10),"Scotland","Wales","Northern Ireland",
                        "England","London"),
                   Value=c(4114,3220,2976,2486,1986,1972,1757,
                           1494,1122,1057,2200,900,400,7800,10500))
river_data<-makeRiver(nodes2, edges2)
riverplot(river_data, lty = 0, srt = 30, default_style = NULL, gravity = "top",
```

```
node_margin = 1, nodewidth = 1, plot_area = 0.5, nsteps = 50,
add_mid_points = NULL, yscale = "auto")
text("Travelers in the UK", x=1,y=13)
```



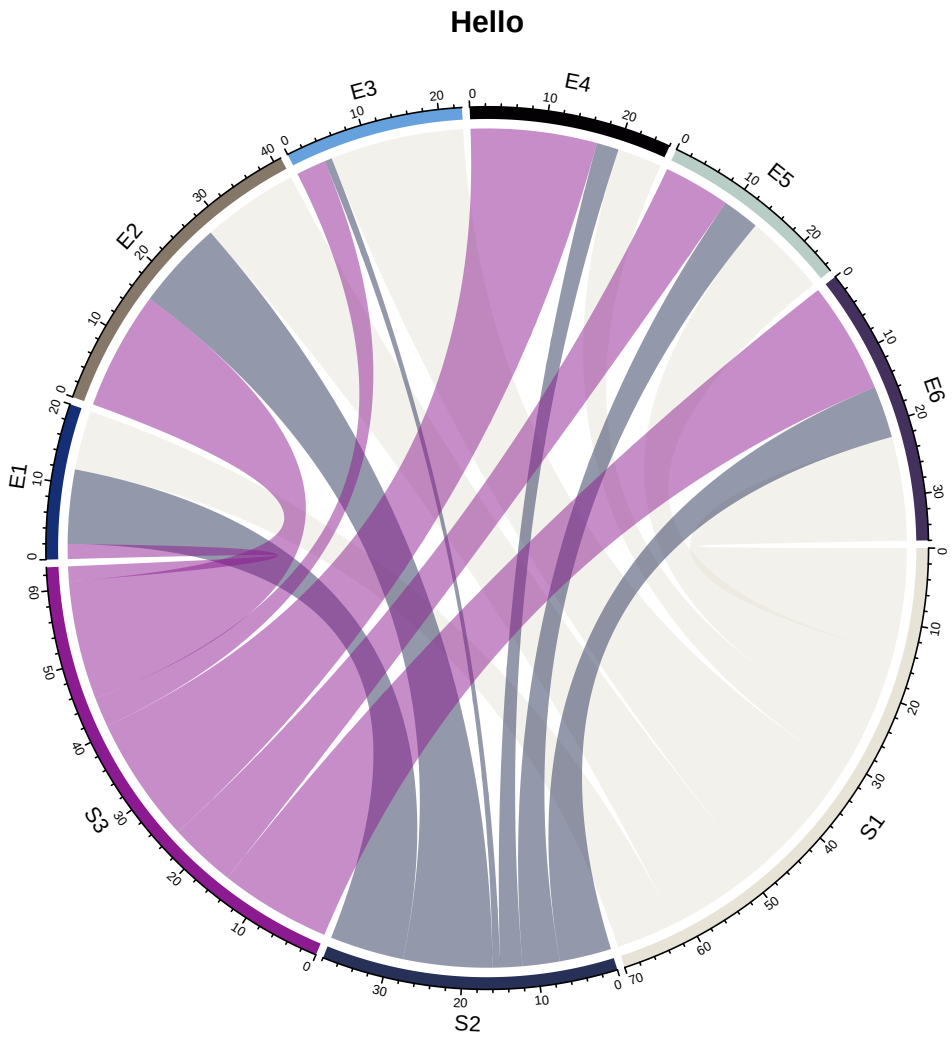
Chord Diagrams

- Chord diagrams are the polarized version of a riverplot. Chord diagrams can only realistically display one starting node and one ending node (and not all the points along a journey).
- Let us take a look at the visualization Where do college graduates work?
  1. What college major do most people who remain in STEM have?
  2. What gender differences do you notice? What ethnicity differences do you notice?
- Now, let us take a look at this visualization: Cell phone loyalty
  1. Spend 5 minutes (or so) going through the interactive visualization.
  2. In what way is this chord diagram different from the previous one?
  3. What customers are the most loyal to their original brand? What customers are the least loyal to their original brand?
  4. What do the colors indicate in this plot?
- Now, let us take a look at this visualization: Congressional Influences
  1. What can you decipher from this chord diagram?
- A chord diagram can display data that can be written in a matrix:

```
##    E1 E2 E3 E4 E5 E6
## S1  8 13 18  6 11 14
## S2 10 12  1  3  5  7
## S3  2 16  4 17  9 15
```

  - Where did most people start from?
  - Where did most people end up?
  - How many people started at E4?
  - People who ended at E6, most likely came from what starting point?

```
require(circlize)
chordDiagram(mat)
title("Hello")
```

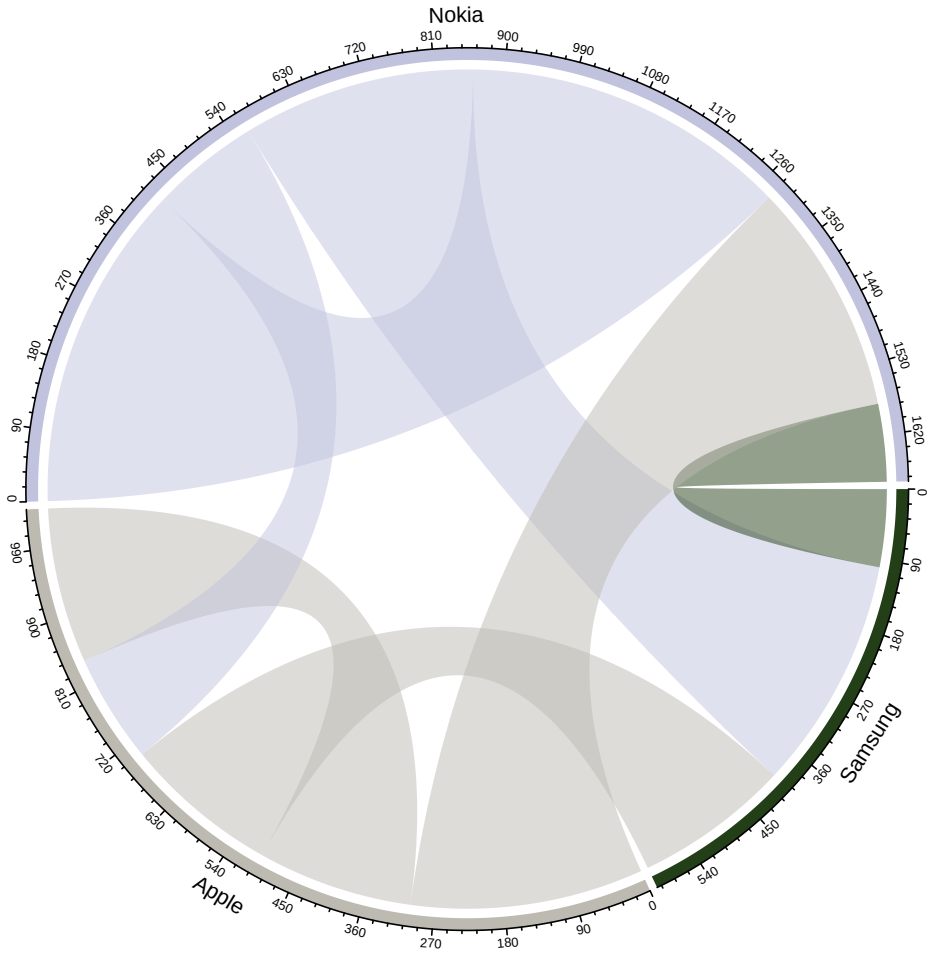


- Notice that our cell phone example from the riverplot section can be written into a matrix:

```
mat2 = matrix(c(0,200,300,0,200,140,100,300,420), nrow=3, ncol=3)
rownames(mat2) = c("Samsung", "Apple", "Nokia")
colnames(mat2) = c("Samsung", "Apple", "Nokia")
mat2

##         Samsung Apple Nokia
## Samsung       0     0   100
## Apple       200   200   300
## Nokia       300   140   420

chordDiagram(mat2)
```

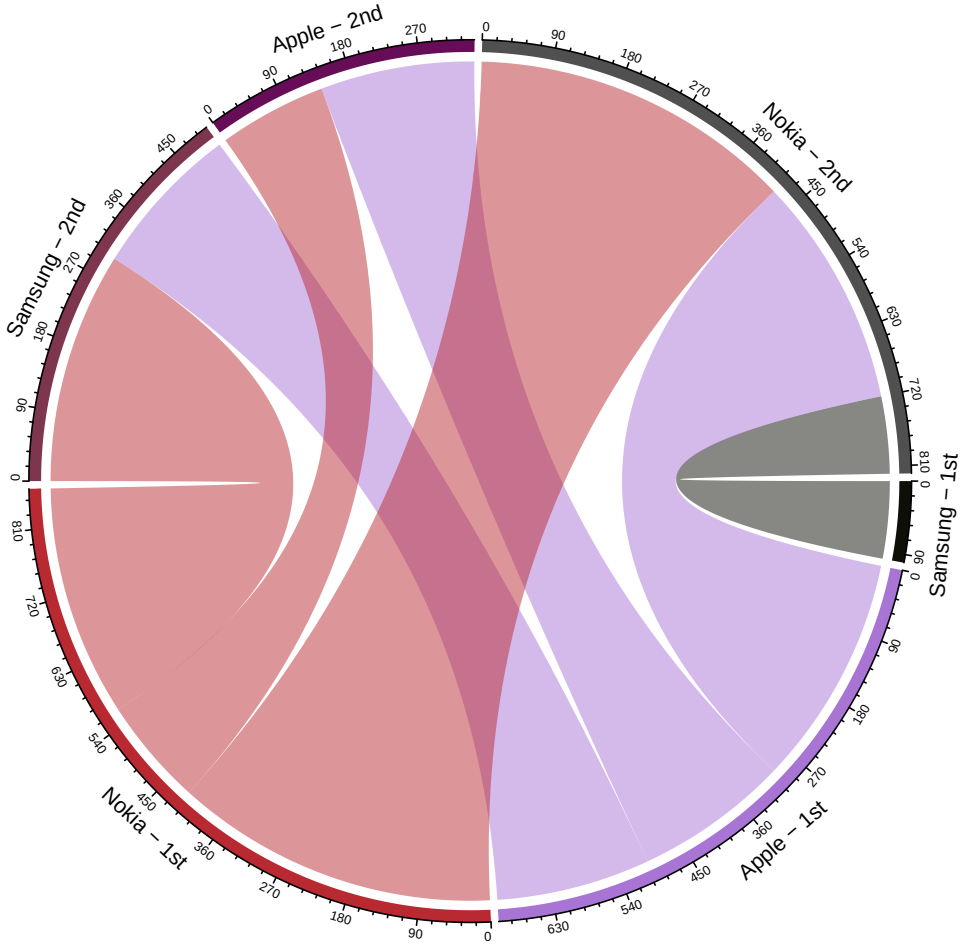


- The chord diagram could also be displayed as follows:

```
mat3 = matrix(c(0,200,300,0,200,140,100,300,420), nrow=3, ncol=3)
rownames(mat3) = c("Samsung - 1st", "Apple - 1st", "Nokia - 1st")
colnames(mat3) = c("Samsung - 2nd", "Apple - 2nd", "Nokia - 2nd")
mat3

##               Samsung - 2nd Apple - 2nd Nokia - 2nd
## Samsung - 1st             0           0         100
## Apple - 1st             200         200         300
## Nokia - 1st             300         140         420

chordDiagram(mat3)
```



- What are your comments on this plot? Which one do you prefer?