Alluvial Plots

- An alluvial plot can be used to plot a special type of riverplot. These plots work well in the following scenarios:
  - If the focus of the visualization is on the evolving classification of a variable (think of repeated categorical measurements).
  - If we want to visualize how a particular variable is divided among a series of subsequent variables.
- One benefit to an alluvial plot is that you can more easily maintain summary information about the data. (In a riverplot you typically do not see counts, though there may be a legend that describes what each edge thickness represents).
- An alluvial plot is less flexible since you have to have the same variable flowing from one node to another. (that is, our focus will be on the flow of a single variable – usually over time). Think of how this differs from the Traveler to the UK example we looked at last class.
- The sytanx for ggalluvial will depend on the format of your data.
  - Typically, you will want to make sure your data is in long format.
  - You will have to pay attention to whether you have 'individual-level' data or aggregated data. If you have aggregated data, you will want to utilize the 'weight' argument below that contains your counts. If you have individual level data, you will not need to use this argument.
  - You will also have to think about your stratum. Are they the same at each time point or are they different?
    - Suppose that you want to see how someone's major changes over time. Then, the stratum at each time point is major.
    - Suppose that you want to see how someone's geographic location relates to major choice, which in turn relates to career choice. Then the stratum change over time. (NOTE: Even though the stratum change over time – the focus will be on the left most variable.)
  - If the stratum are the same at each time point, utilize the code in the first (or second) example. If they differ, follow the code in the third example.
- Let us revisit the cell-phone loyalty example and plot it as an alluvial diagram instead.

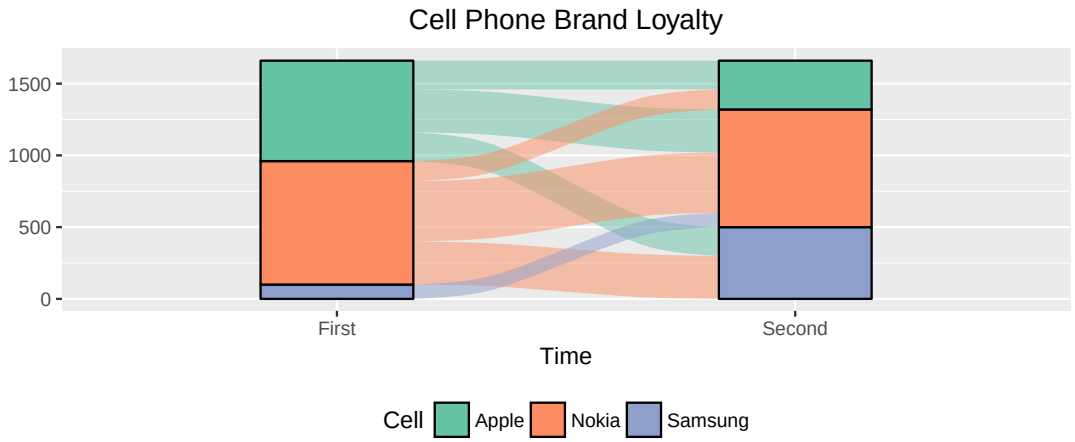
```
require(ggalluvial)
require(reshape2)

cell_data<-data.frame(First=c(rep("Samsung",3),rep("Apple",3),rep("Nokia",3)),
                      Second=c("Samsung","Apple","Nokia","Samsung",
                           "Apple","Nokia","Samsung","Apple",
                           "Nokia"),
                      Value=c(0,0,100,200,200,300,300,140,420),
                      id=1:9)

cell_long<-melt(cell_data, id=c("id","Value"))
names(cell_long)<-c("id","Freq","Time","Cell")

ggplot(cell_long,
        aes(x = Time, stratum = Cell, alluvium = id,
            weight=Freq, fill = Cell))+
```

```
geom_flow() +
geom_stratum() +
theme(legend.position = "bottom") +
ggtitle("Cell Phone Brand Loyalty")+
scale_fill_brewer(palette = "Set2")+
theme(plot.title = element_text(hjust = 0.5))
```

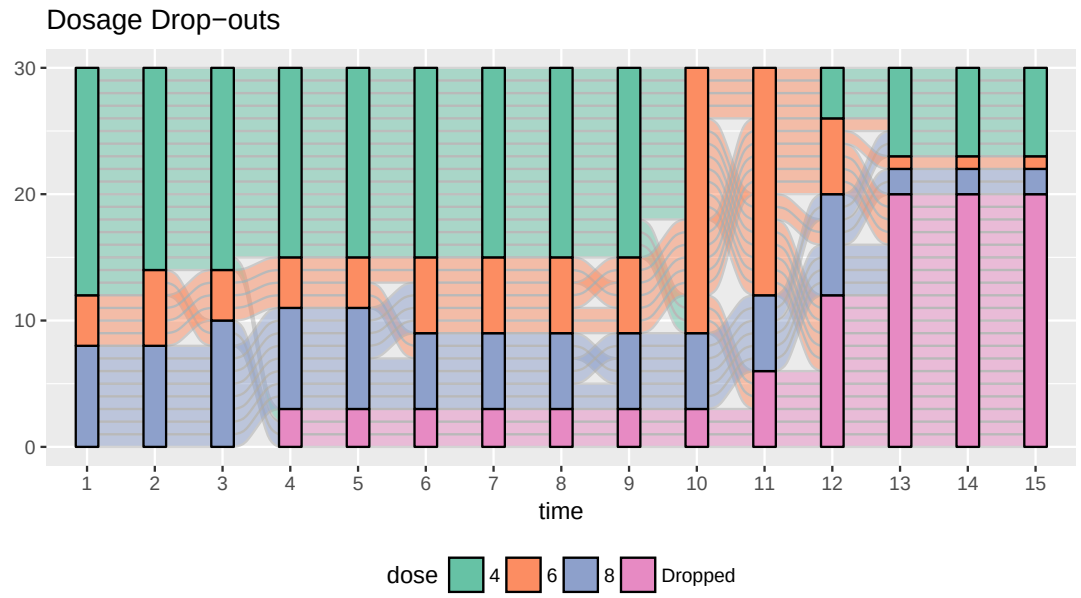


- It looks pretty similar to our riverplot from before. I would argue it looks even nicer here too as it maintains more information about the counts of customers.

- Pharmaceutical Example: Each week a patient entered the clinic and either kept the dose of medicine they were on or switched to a higher or lower dose. Clinicians wanted to examine how dosages were re-assigned over time and to see how they may be related to drop-out of patients.

```
##Import pharma.csv data set

##The plot
ggplot(pharma, aes(x = time, stratum = dose,
                   alluvium = patient,
                   fill = dose)) +
  geom_flow(stat = "alluvium", color = "darkgrey") +
  geom_stratum() +
  scale_fill_brewer(palette = "Set2") +
  theme(legend.position = "bottom") +
  ggtitle("Dosage Drop-outs")
```



- The other example we considered last class was Travelers to the UK. This example does not conform to our needs of an an alluvial diagram. We would need to know more information for this type of plot to work. I am artificially making up the number of travellers from each country that visited each part of the UK.

```
ggplot(uk_data_fake,
       aes(weight = totals, axis1 = Origin, axis2 =destination)) +
  geom_alluvium(aes(fill = Origin), width = 1/12) +
  geom_stratum(width = 1/12, color = "grey")+
  geom_label(stat = "stratum", label.strata = TRUE, size=2)+
  scale_x_continuous(breaks = 1:2, labels = c("Country of Origin", "Destination"))+
  ylab("Number of Visitors")
```