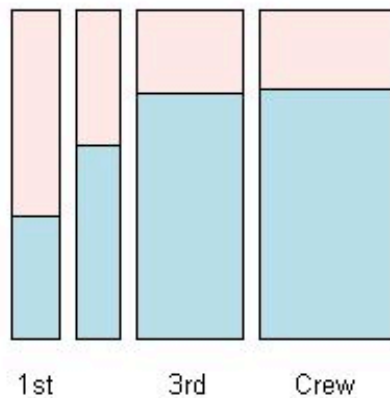

Spine Plot

A graphical display of cross-classified data

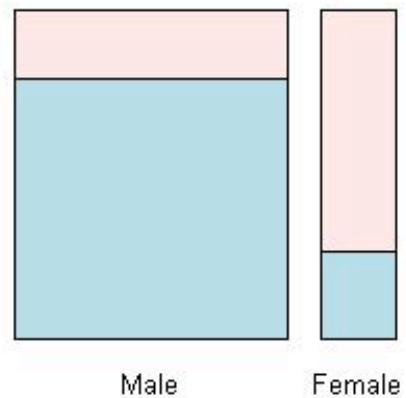
유 충현

블로그 모음 2탄(<http://blog.naver.com/bdboys>) • (주)오픈베이스 • 2012년 10월 3일

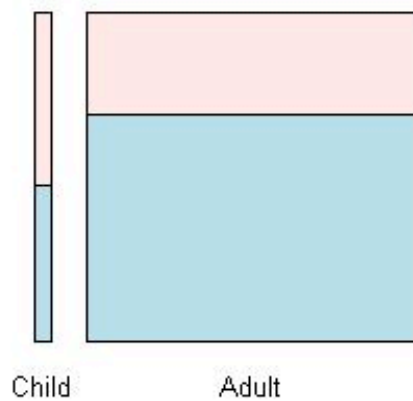
Survived over Economic status (cla



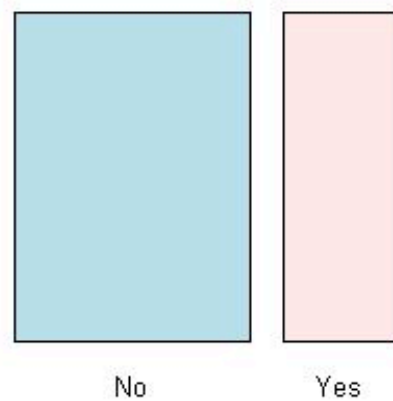
Survived over Sex



Survived over Age



Survived



Spine Plot

Spine Plot은 Barchart의 일종이라 할 수 있다. Barchart는 일변량 범주형 자료에서 Class의 숫수의 비율을 막대의 길이로 표현한 것인 반면 Spine Plot은 막대의 길이는 동일하게 하고 Class의 숫수의 비율을 막대의 폭으로 표현한 것이다.

Mosaic Plot이 Spine Plot을 기본 아이디어로 만들어 졌다고 Mosaic Plot을 이야기할 때 언급한 적이 있었다. Spine Plot은 단지 단변량 Mosaic Plot이라고도 할 수 있다. Spine Plot은 범주들간의 비율차를 이해하는데 장점이 있다.

R의 barplot 함수를 이용해서 다음과 같은 spineplot 함수를 만들어 보았다.

```
spineplot <- function (data, ...)  
{  
  temp=data  
  if (is.null(ncol(data))) {  
    len = length(data)  
    for (i in 1:len)  
      temp[i]=data[i]*(max(data)/data[i])  
    wid=data  
  }  
  else {  
    len = ncol(data)  
    for (i in 1:len)  
      temp[,i]=data[,i]*(max(apply(data,2,sum))/apply(data,2,sum))[i]  
    wid=apply(data,2,sum)  
  }  
  barplot(temp, width=wid, axes=FALSE, ...)  
}
```

그러면 Mosaic Plot에서 사용한 클래식 음악에 대한 자료와 타이타닉 자료를 이용해서 Spine Plot을 그려보면서 Barchart 및 Mosaic Plot과의 차이점을 알아보자.

	교육수준			
나이	고학력		저학력	
	클래식음악 듣기			
	예	아니오	예	아니오
고연령	210	190	170	730
저연령	194	406	110	290

```

> music = c(210, 194, 170, 110, 190, 406, 730, 290)

> dim(music) = c(2, 2, 2)

> dimnames(music) = list(Age = c("Old", "Young"),
                           Education = c("High", "Low"),
                           Listen = c("Yes", "No"))

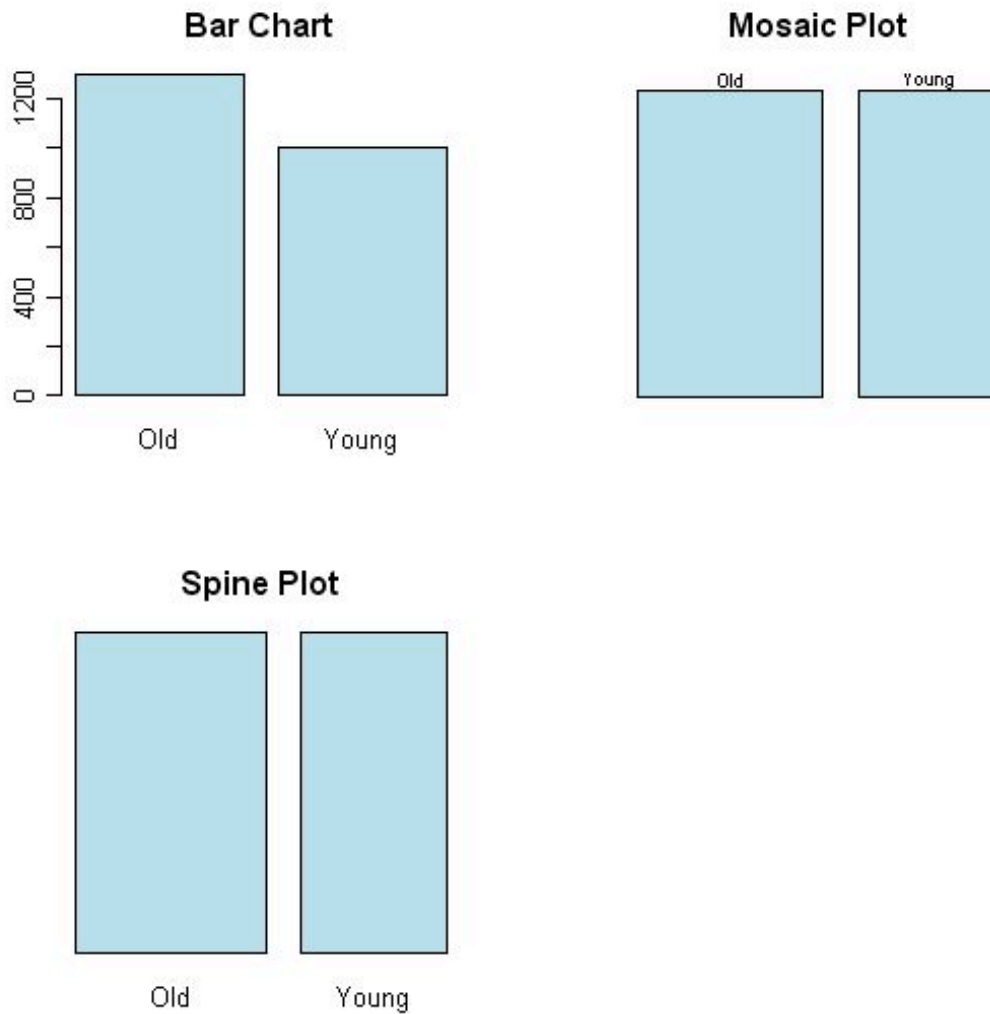
> par(mfrow=c(2,2))

> barplot(apply(music,1,sum), col = "lightblue",
           main="Bar Chart")

> mosaicplot(apply(music,1,sum), col = "lightblue",
             main="Mosaic Plot")

> spineplot(apply(music,1,sum), col = "lightblue",
            main="Spine Plot")

```

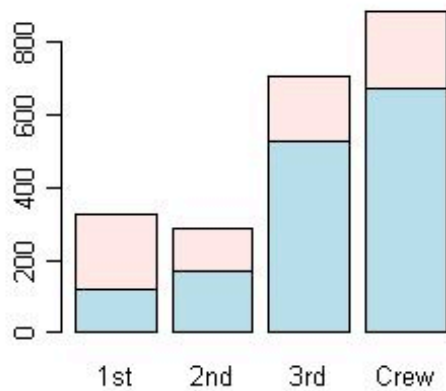


그림에서처럼 일변량일 경우에는 Mosaic Plot과 Spine Plot의 차이점은 없다. 단지 Barchart에서는 막대의 폭은 동일한데 길이의 차이가 있고, Spine Plot은 막대의 길이는 동일한데, 폭의 넓이가 차이가 난다.

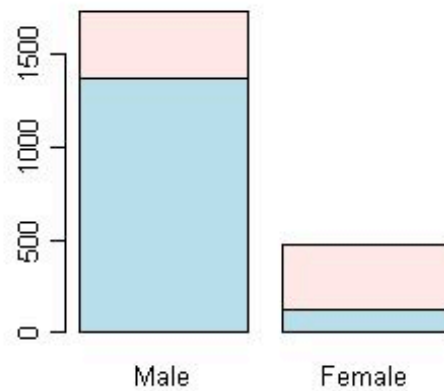
타이타닉의 자료를 이용하여 몇 개의 Chart를 그려 보자.

```
> par(mfrow=c(2,2))
> barplot(apply(Titanic,c(4,1),sum),col=c("lightblue", "mistyrose"),
  main="Survived over Economic status (class)")
> barplot(apply(Titanic,c(4,2),sum),col=c("lightblue", "mistyrose"),
  main="Survived over Sex",)
> barplot(apply(Titanic,c(4,3),sum),col=c("lightblue", "mistyrose"),
  main="Survived over Age",)
> barplot(apply(Titanic,4,sum),col=c("lightblue", "mistyrose"),
  main="Survived")
```

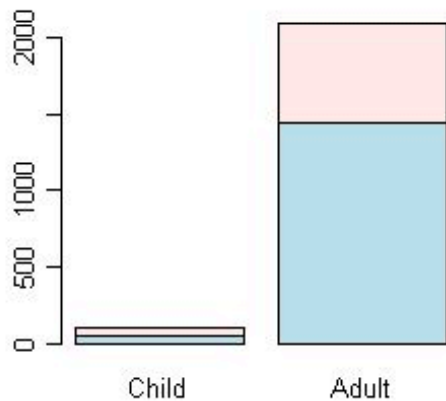
Survived over Economic status (class)



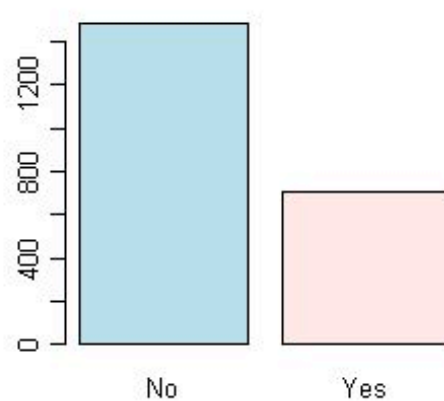
Survived over Sex



Survived over Age

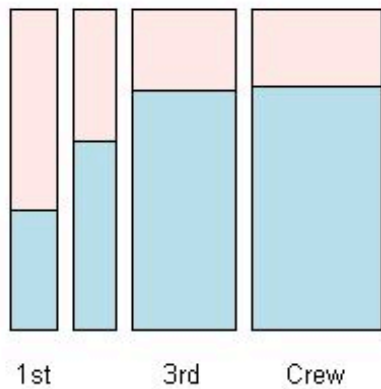


Survived

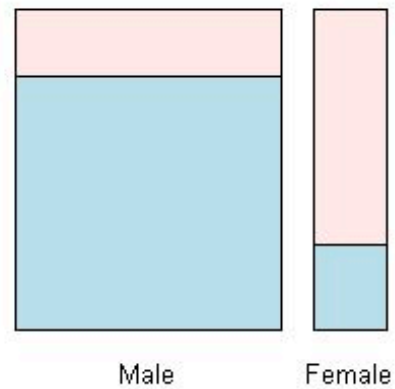


```
> par(mfrow=c(2,2))
> spineplot(apply(Titanic,c(4,1),sum),col=c("lightblue", "mistyrose"),
  main="Survived over Economic status (class)")
> spineplot(apply(Titanic,c(4,2),sum),col=c("lightblue", "mistyrose"),
  main="Survived over Sex",)
> spineplot(apply(Titanic,c(4,3),sum),col=c("lightblue", "mistyrose"),
  main="Survived over Age",)
> spineplot(apply(Titanic,4,sum),col=c("lightblue", "mistyrose"),
  main="Survived")
```

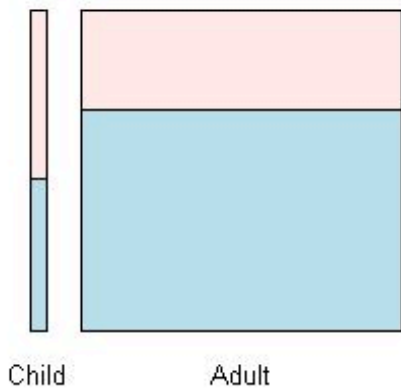
Survived over Economic status (class)



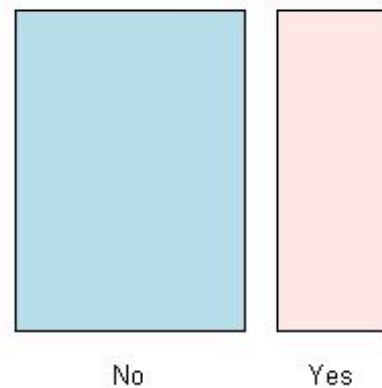
Survived over Sex



Survived over Age

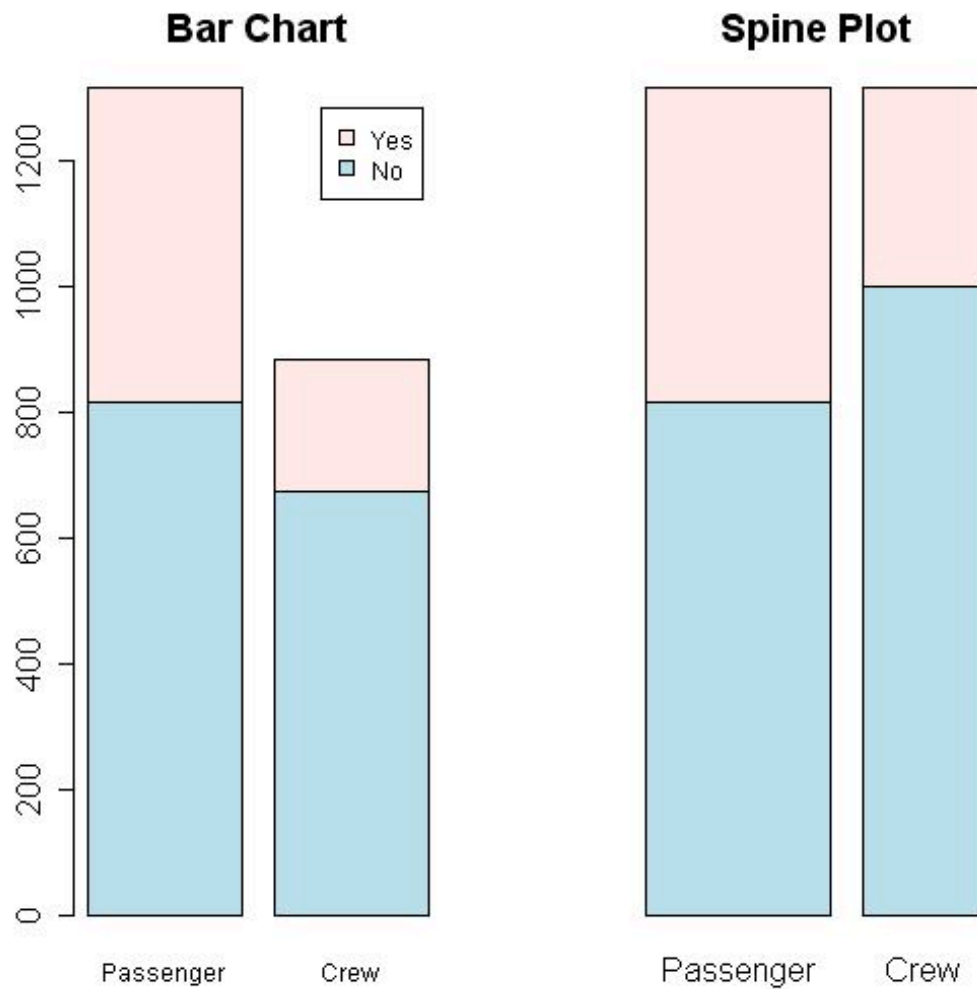


Survived



Bar chart보다 Spine Plot이 범주들간의 비율차를 이해하기가 쉽다.

```
> temp=c(apply(Titanic[dimnames(Titanic)[[1]]!="Crew",,,2:4,sum),
             Titanic[dimnames(Titanic)[[1]]=="Crew",,,])
> dim(temp)=c(2,2,2,2)
> dimnames(temp)=list(Sex = c("Male", "Female"),
                      Age = c("Child", "Adult"),
                      Survived = c("No", "Yes"),
                      Class = c("Passenger", "Crew"))
> par(mfrow=c(1,2))
> barplot(apply(temp,3:4,sum),col=c("lightblue", "mistyrose"),
          main="Bar Chart",
          legend=row.names(apply(temp,3:4,sum)))
> spineplot(apply(temp,3:4,sum),col=c("lightblue", "mistyrose"),
            main="Spine Plot")
```



마지막으로 타이타닉의 승무원과 승객의 생존 비율을 두 가지의 Chart로 그려서 비교해 보았다.