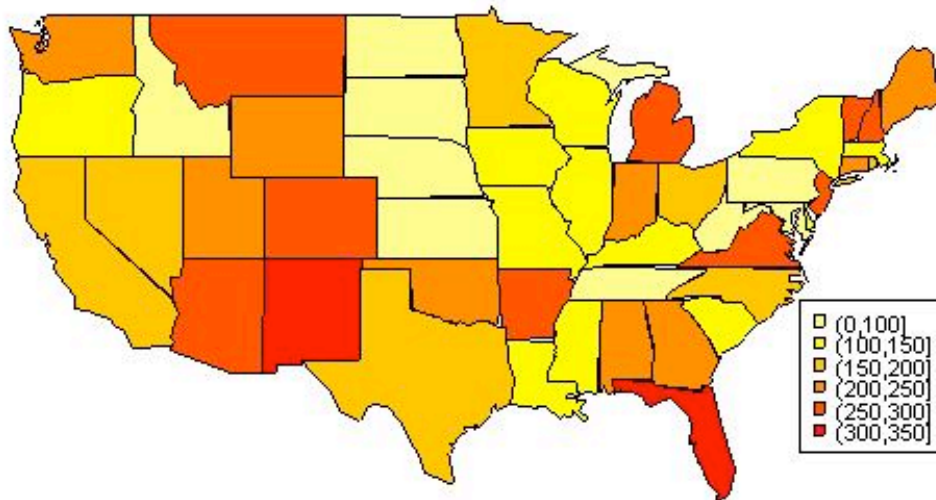

Graphics - maps

유 증현

블로그 모음 7탄(<http://blog.naver.com/bdboys>) • (주)오픈베이스 • 2010년 10월 5일

USA Assault map



Graphics - maps

요즘 Geographics + Statistics의 합성어가 낯설지 않다. 이를 지원하는 R 패키지 지도 많이 나와 있다. 사실 선진국은 예전부터 이 분야에 많은 투자를 해왔으며 국내에서는 자료가 빈곤한 실정이다. Geographics + Statistics를 위해서는 공개된 국내 map 데이터가 필요한데, 찾아보기 어려운 실정이다.

map 데이터에 통계량을 가미해서 시각화 시킨 map chart는 지역의 특성을 쉽게 파악할 수 있는 장점이 있는 반면, map 데이터가 절대적으로 필요하다는 전제조건이 있다. 그러면 R의 maps와 mapdata 패키지를 통해 Geographics + Statistics를 느껴보자.

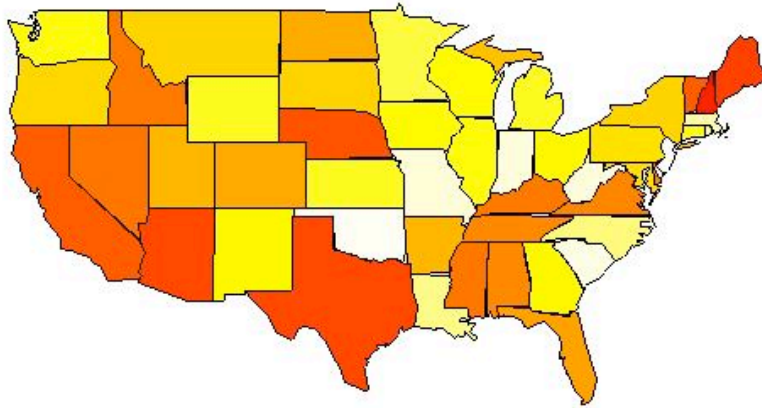
먼저 R의 세계에서 쉽게 예제로 사용되는 USArrests 데이터를 가지고 미국의 1973년 미국의 48개 주의 폭력(강간)범 들의 수를 지도에 표현해 보자. 자료에 대한 자세한 설명은 USArrests의 도움말을 참고하자.

```
> library(maps)
>
> # 본토에서 떨어진 Alaska, Hawaii 데이터 제외
> sub.usa = subset(USArrests, !rownames(USArrests) %in% c
("Alaska", "Hawaii"))
>
> # 폭행범수(100,000명중)의 range+1를 구한다.
> assault.diff=max(sub.usa$Assault)-min(sub.usa$Assault)+1
>
> # range+1개의 heat 색상 테이블을 구한다.
> col.level=heat.colors(max(sub.usa$Assault)-min(sub.usa$Assault)+1)
>
> # 색상 테이블의 색깔 갯수
> cnt.level=length(col.level)
>
> # 폭행범 수가 max에 가까울 수록 붉은 색,
> # min에 가까울 수록 옅은 노란색의 색상 부여
> stat.col=abs(max(sub.usa$Assault)-sub.usa$Assault)+1
```

```

>
> # 주이름, 폭행범수, 색상을 갖는 48 by 3의 matrix생성
> usa.data=cbind(rownames(sub.usa), sub.usa$Assault, stat.col)
>
> # map 그래프로 출력
> map('state', region = usa.data$states, fill=TRUE, col=col.level
[as.integer(usa.data[,3])])

```



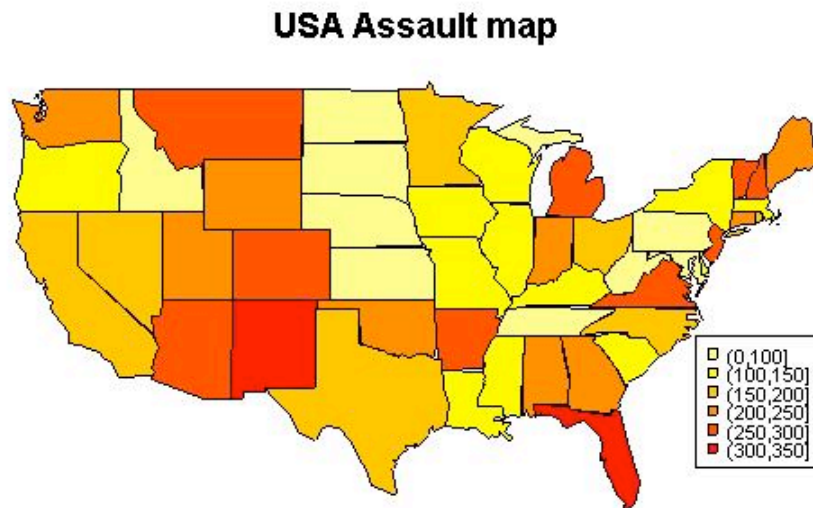
이 그림에서 붉은색에 가까운 주는 폭력(강간범)의 수가 많은 주이고, 노랑고 옅은 색의 주는 폭력(강간범)의 수가 적은 주를 의미한다. 하와이와 알래스카 주는 제외하였다. 지리적으로 멀리 떨어져 있기 때문에 데이터에서 누락시켰다. 단순히 숫자 테이블로 보는 자료보다 이해의 폭이 클 수 있다. 이것이 비록 사소한 것이지만 Geographics + Statistics의 한 예임에는 틀림이 없다.

앞서서 그린 그림은 색상의 레벨이 300정도다. 다음의 예는 6개의 구간으로 나누었다. 마찬가지로 붉은색에 가까운 주는 폭력(강간범)의 수가 많은 주이고, 노랑고 옅은 색의 주는 폭력(강간범)의 수가 적은 주를 의미한다.

```

> col.level=cut(USArrests[,2],c(0,100,150,200,250,300,350))
> legends=levels(col.level)
> levels(col.level)=sort(heat.colors(6),decreasing=TRUE)
> col.level=as.character(col.level)
> map('state', region = usa.data$states, fill=TRUE, col=col.level)
> title("USA Assault map")
> legend(-76,35,legends,fill=sort(heat.colors(6),decreasing=TRUE),
cex=0.7)

```

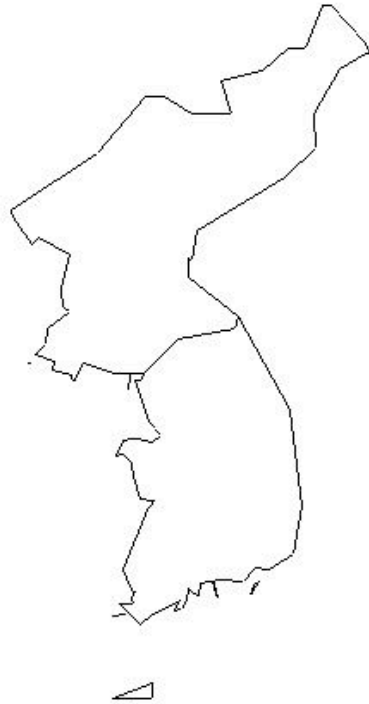


다음은 우리 나라 지도를 그려보자.

```

> map('world',region=c('South Korea','North Korea'))

```



여기서 제공하는 map 데이터는 1980년대 미국에서 만든 CIA World DataBank II라는 라이브러리의 자료를 R에서 사용할 수 있게 Conversion한 것이다. 그래서 미국은 주 및 County자료까지 있다. 그러나 국내 지도는 전국지도 달랑 하나만 있다. 강의 자료라든가, 도시의 자료도 없다. 중국은 별도의 map도 존재한다.

이 map은 데이터의 양이 축소해서 미려하지 못하다. mapdata 패키지에 있는 데이터로 다시 그려 보자.

```
> library(mapdata)
> map('worldHires',region=c('South Korea','North Korea'))
```



훨씬 보기가 좋다.

내친 김에 일본과 중국도 그려 보자. 일본은 검정색으로 그려 보았다. 속이 워낙 검어서....

이상으로 maps 패키지에 대해서 간단하게 살펴 보았다.

```
> map('worldHires',region=c('South Korea','North  
Korea','Japan','China'))  
> map('worldHires',region=c('South Korea'),col='blue',add=TRUE,  
fill=TRUE)  
> map('worldHires',region=c('North Korea'),col='red',add=TRUE,  
fill=TRUE)  
> map('worldHires',region=c('Japan'),col='black',add=TRUE, fill=TRUE)  
> map('worldHires',region=c('China'),col='yellow',add=TRUE, fill=TRUE)
```

