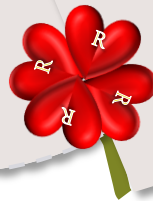


### 3. ggplot2()

r의 기본 그래픽함수보다 더 간단하면서도 다양한 시각적 효과를 표현할 수 있는 ggplot2() 기본내용을 학습합니다.



## ggplot2()패키지 개요

- 데이터를 시각화하는데 있어서 간단한 설정만으로 다양한 그래픽 표현이 가능
- 하나의 데이터를 가지고도 여러 그래픽 표현이 가능
- ggplot2패키지를 이용하기 위해서는 아래와 같은 작업이 필요
  - install.packages("ggplot2")
  - library(ggplot2)

## 1. qplot()

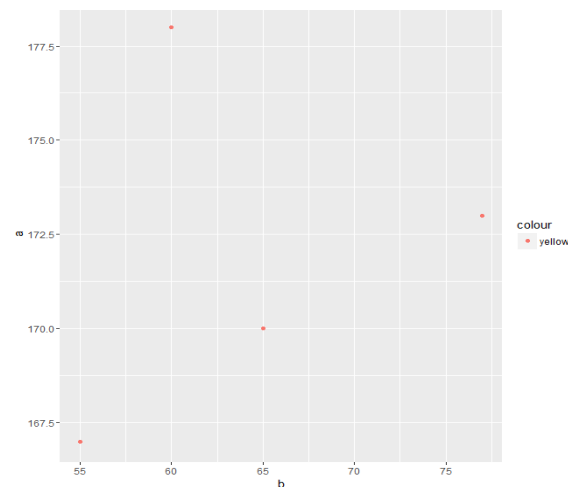
- r에서 기본적으로 사용하는 plot() 함수와 비슷한 구조를 가진 함수
- 형식

•qplot(x,y, data=, color=, shape=, size=, alpha=, geom=, method=, formula=, facets=, xlim=, ylim=, xlab=, ylab=, main=, sub=)



너무 많은 옵션들을 갖고 있으며, 그래픽 핸들링시 인터페이스가 복잡해지는 문제점!!

```
> a<-c(167,170,173,178)
> b<-c(55,65,77,60)
> qplot(b,a,col="yellow")
```



## 2. ggplot()

• `ggplot(data, aes(x=x축 데이터,y=y축 데이터)) + geom` 당신이 원하는 그래프()

## •구성

➢data - 그래픽을 원하는 데이터 집합

➢aes() - 데이터 각 레코드를 그래픽 요소에 매핑하는 작업(미적 요소 매핑)

➢geom()- 각 축에 숫자형 데이터를 입력받아 실제로 시각화 하기 위해 좌표계에 표시하는 작업( 기하객체 (geometric object))으로 표시되는 형태에 따라 점(point), 선(line), 다각형(polygon) 등 많은 객체가 존재

➢stat - 해당 데이터값에 대한 통계적 처리방법



그래픽에 어떤 데이터를 대상으로 하지?



그래프 색상을  
뭘로하지?



그래프를 선으로 그릴까?  
점으로? 아님 막대로?



개수를 보여줘?  
아님 평균으로?  
아님 합계로?



## 기본 예1)

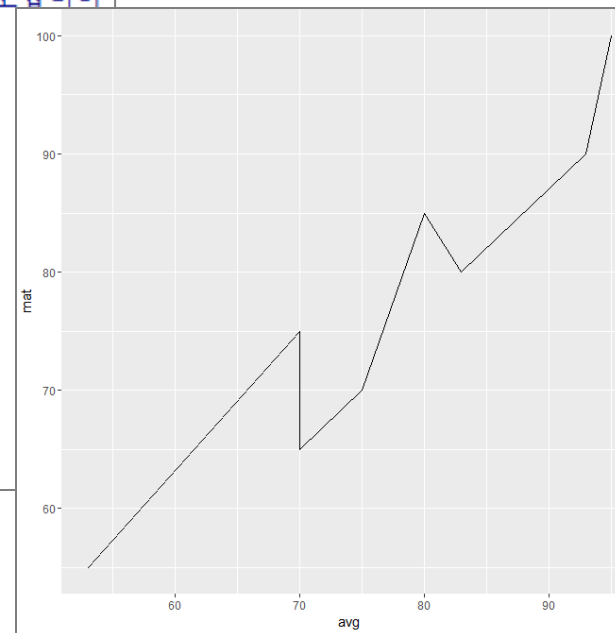
```
> install.packages("ggplot2")
--- 현재 세션에서 사용할 CRAN 미러를 선택해 주세요 ---
URL 'https://rweb.crmdata.ku.edu/cran/bin/windows/contrib/3.2/ggplot2_2.1.0.zip'을 시도합니다
Content type 'application/zip' length 2001438 bytes (1.9 MB)
downloaded 1.9 MB
```

패키지 'ggplot2'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다

다운로드된 바이너리 패키지들은 다음의 위치에 있습니다

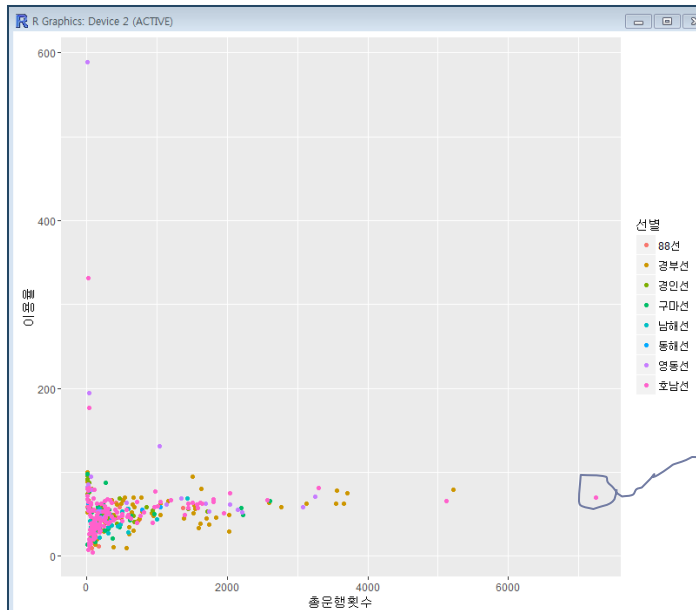
C:\Users\Administrator\AppData\Local\Temp\RtmpSWyGPP\downloaded\_packages

```
> require(ggplot2)
필요한 패키지를 로딩중입니다: ggplot2
Need help? Try the ggplot2 mailing list:
http://groups.google.com/group/ggplot2.
> data<-data.frame(mat=c(55,75,80,65,90,100,70,85),eng=c(65,100,45,50,75,90,90,65),
+ avg=c(53,70,83,70,93,95,75,80),irum=c('김','이','박','최','문','윤','노','정'))
> ggplot(data,aes(x=avg,y=mat))+geom_line()
>
```



## 기본 예2)

```
datainfo<-read.csv("d:\\r1\\highwaybus09.csv",sep=" ",header=TRUE)
datainfo
ggplot(datainfo,aes(x=총운행횟수,y=이용율))+geom_point(aes(colour=선별))
```



운행거리에 따라 점의 크기를  
조정하여 보여주면 시각적으로  
더 강조되겠죠?

```
> ggplot(datainfo,aes(x=총운행횟수,y=이용율))+geom_point(aes(colour=선별,size=거리))
```

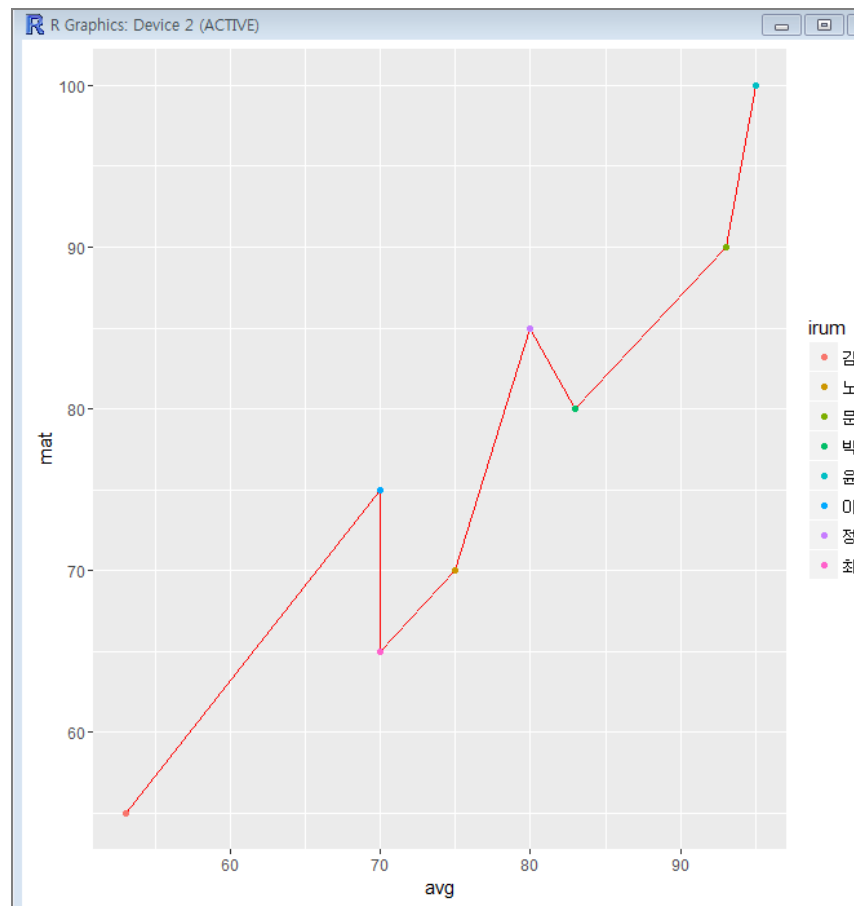
◆레이어 개념을 이용하여 유연한 시각화작업이 가능해요

## 레이어의 기본 예제1)

```
> ggplot(data, aes(x=avg, y=mat)) + geom_line(colour="red") + geom_point(aes(colour=irum))
```

= geom\_point(aes(x=avg, y=mat, colour=irum))  
과 같은 의미입니다.  
irum 으로 각 점의 색상을 구분하겠다는 의미겠죠?

ggplot2의 장점 중 하나는 레이어를 추가하면 작업했던 미적 매핑이 그대로 상속됩니다.



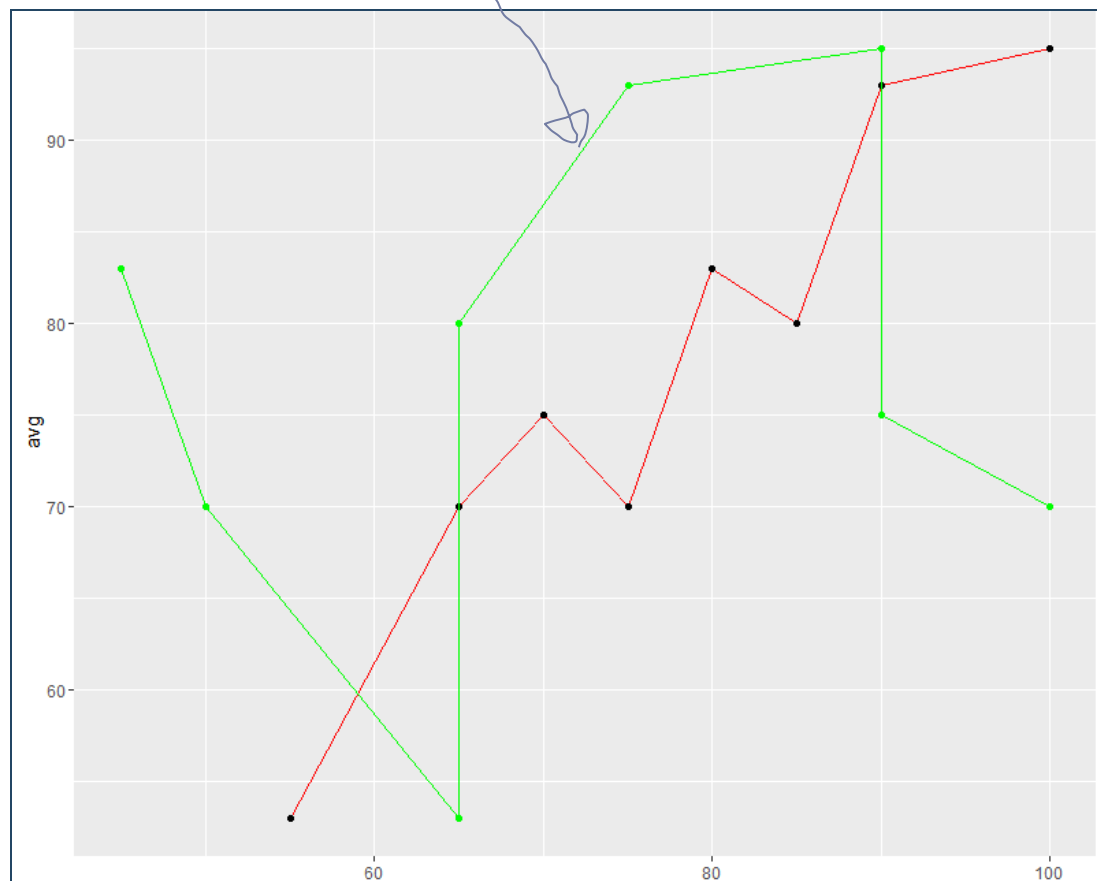
## 레이어의 기본 예제2)

```
> ggplot(data, aes(x=mat, y=avg)) + geom_line(colour="red") + geom_point() +  
+ geom_line(aes(x=eng, y=avg), colour="green") + geom_point(aes(x=eng, y=avg), colour="green")  
>
```

반복?



레이어를 통해 추가된 미적매  
핑은 또 다른 레이어로 상속되  
지 않아요ㅠㅠ

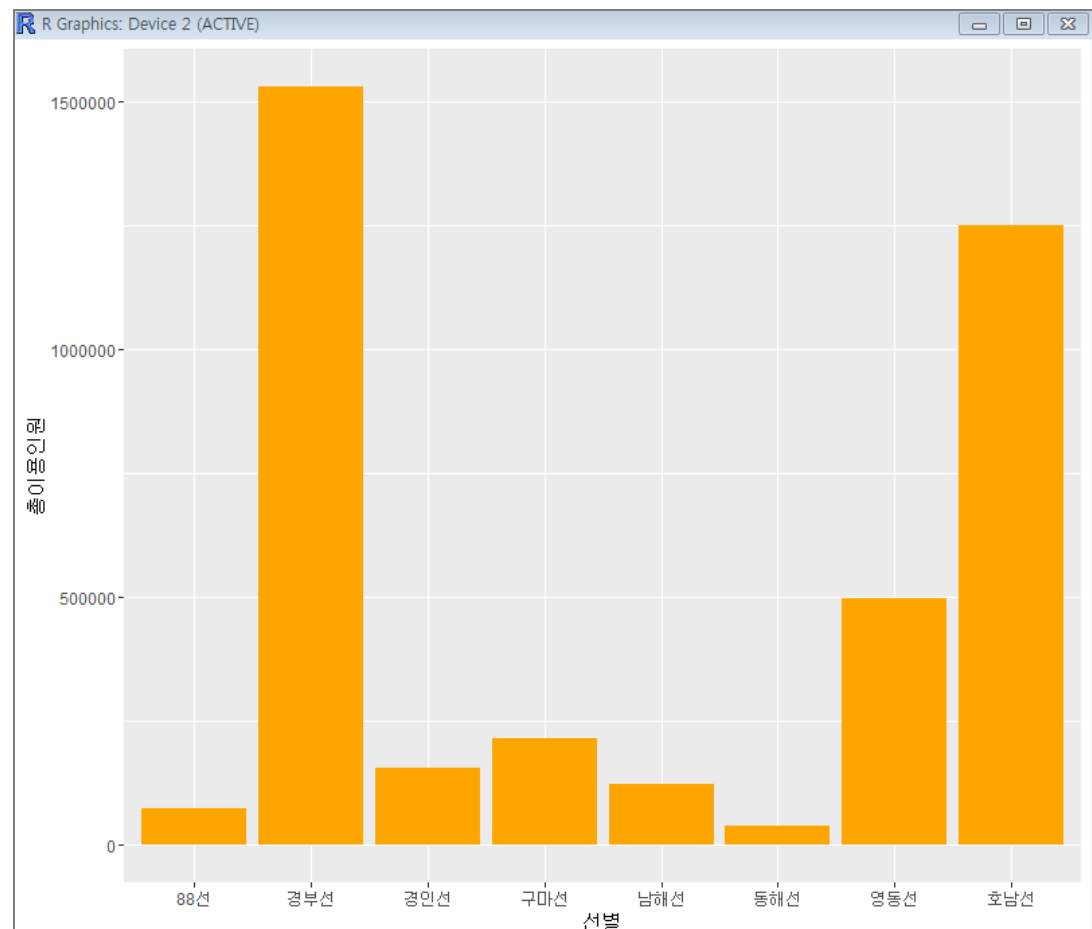


◆geom\_bar()- 막대그래프도 쉽게 할 수 있어요.

기본 예제1)

```
datainfo<-read.csv("d:\\r1\\highwaybus09.csv",sep="," ,header=TRUE)  
ggplot(datainfo,aes(x=선별,y=총이용인원))+geom_bar(stat="identity",fill="orange")
```

geom\_bar()의 stat 요소는 기본적으로 stat\_bin() 즉, 여러분이 표시하려는 항목의 개수를 카운트하여 빈도를 표시하도록 설정되어 있습니다. 따라서, 항목의 계산을 위해서는 stat="identity" 로 셋팅합니다.





## 기본 예제2) 그룹별 평균을 계산하여 막대그래프로 표시하는 작업



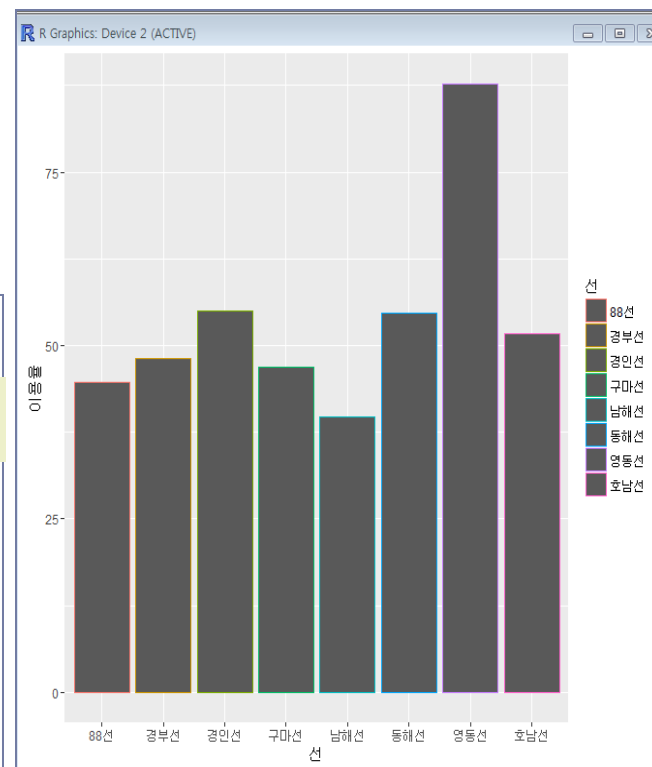
여러분이 방대한 데이터를 그룹별로 요약하여 막대그래프로 표현하고자 한다면 먼저 계산작업을 해야해요. 앞서 배운 `tapply` 함수를 이용하여 함께 응용해봅시다.

```
> datainfo<-read.csv("d:\\r1\\highwaybus09.csv",sep=" ",header=TRUE)
> meandf<-as.data.frame(with(datainfo,tapply(이용율,선별,mean,na.rm=TRUE)))
> meandf$선<-rownames(meandf)
> names(meandf)<-c("이용율","선")
> meandf
```

선별 이용율 평균을 계산한 후 데이터프레임형식으로 변환하여 meandf 에 저장

선	이용율
88선	44.66667
경부선	48.20476
경인선	54.91600
구마선	46.95000
남해선	39.69524
동해선	54.70000
영동선	87.73462
호남선	51.74836

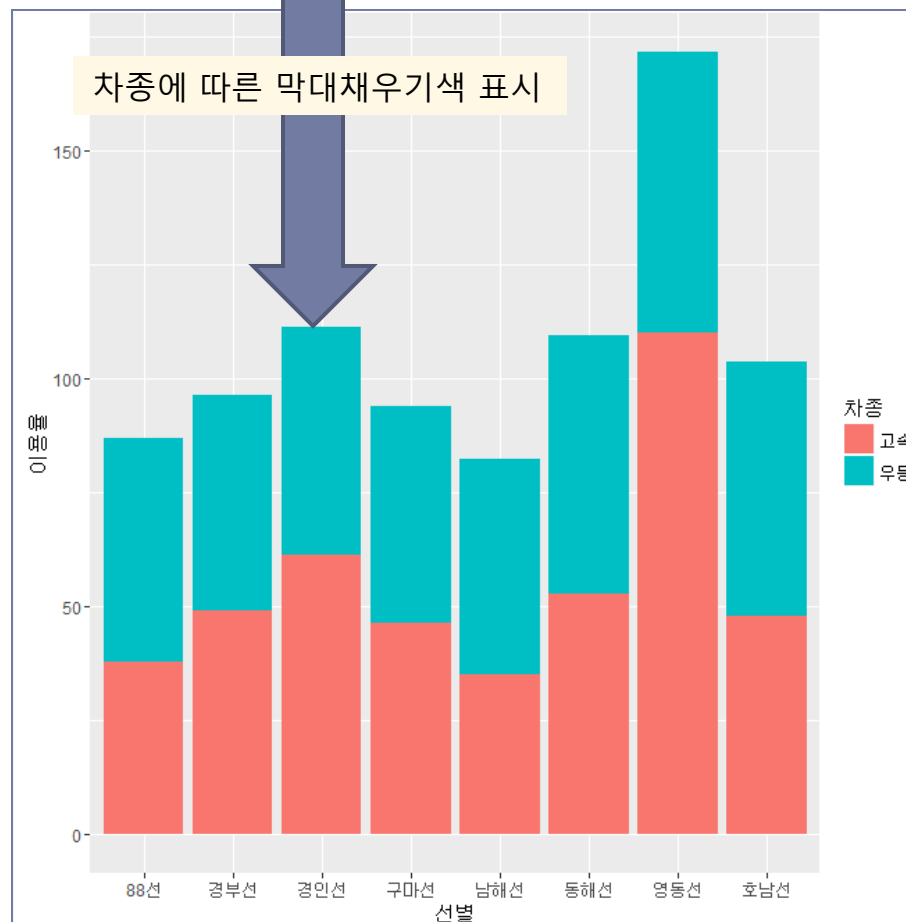
```
> ggplot(meandf,aes(선,이용율))+geom_bar(stat="identity",aes(color=선))
```



## 기본 예제3) 누적막대그래프 작업

선별,차종별 이용률 평균 계산하여 그래픽

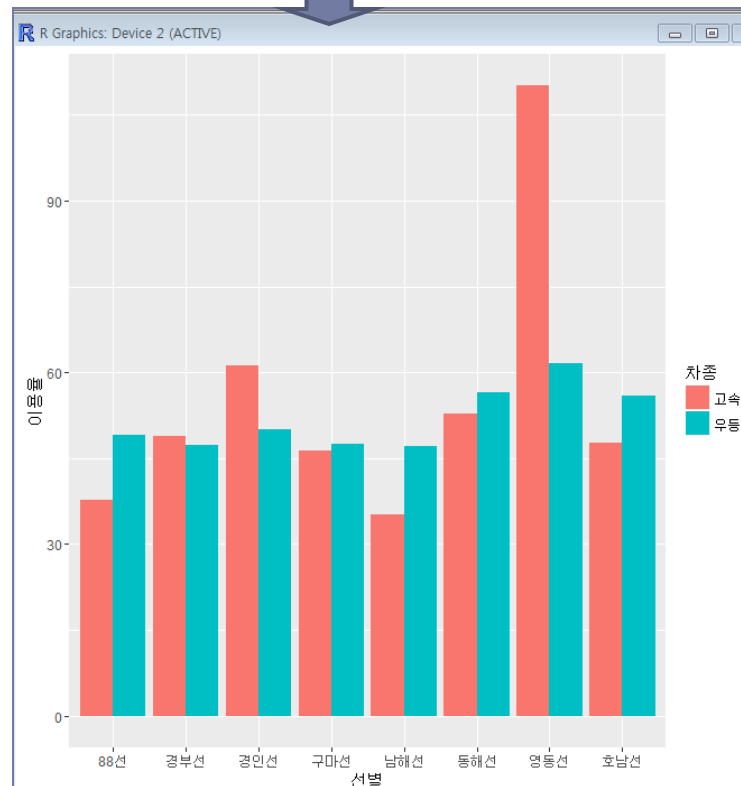
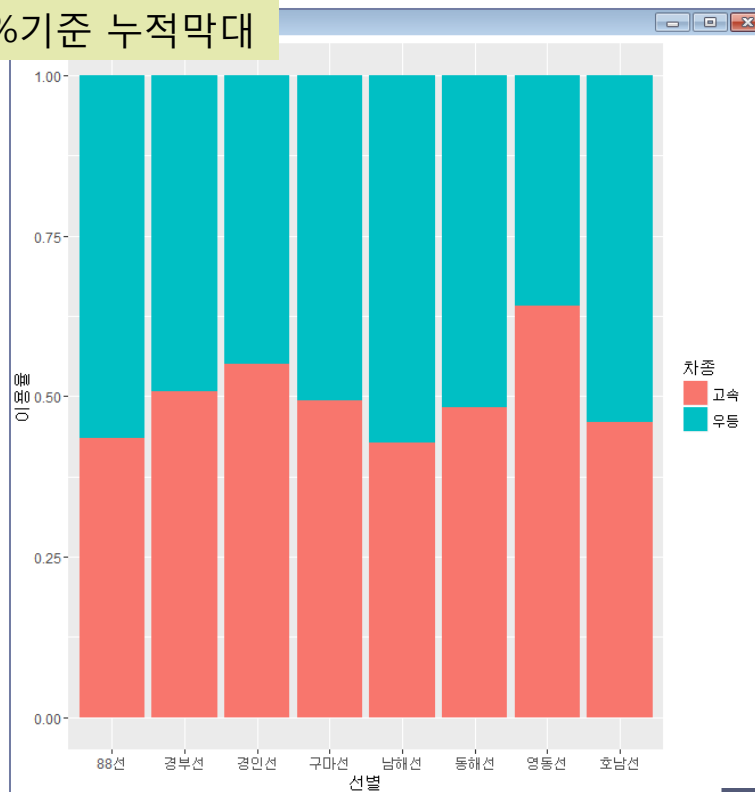
```
> meandf1<-as.data.frame(with(datainfo,aggregate(이용율,list(선별=선별,차종=차종),mean,na.rm=TRUE)))  
> meandf1$이용율<-rownames(meandf1)  
> names(meandf1)<-c("선별","차종","이용율")  
> ggplot(meandf1,aes(선별,이용율))+geom_bar(stat="identity",aes(fill=차종))
```



## 기본 예제4) position요소를 이용하여 여러 형태의 막대형 그래프를 작업

```
> ggplot(meandf1,aes(선별,이용율,fill=차종))+geom_bar(stat="identity",position="dodge")
```

100%기준 누적막대



```
> ggplot(meandf1,aes(선별,이용율,fill=차종))+geom_bar(stat="identity",position="fill")
```