빅데이터 R 분석

김경민

해결문제

고속도로 역주행 교통사고 와 일반 교통사고 분석

- https://www.news1.kr/articles/?4695261
- 최근 3년간 역주행 교통사고의 치명률이 10.2%로 일반 교통사고(4.7%) 보다 2.3배 높은 것으로 나타났다고 한다. 다음 주어진 데이터를 활용하여 이를 분석해 본다.
  - 년도별 치명률 구하기
  - 3년 평균 사고건수, 사망자수, 치명률(사망자수/사고건수) 구하기

구분	2019년사고	2019년사망	2020년사고	2020년사망	2021년사고	2021년사망
전체	4223	206	4039	223	4883	191
역주행	28	5	33	3	27	2

해결문제

	년도	구분	사고	사망
1	2019	전체	4223	206
2	2019	역주행	28	5
3	2020	전체	4039	223
4	2020	역주행	33	3
5	2021	전체	4883	191
6	2021	역주행	27	2

1. 나누기

Filter

	년도	구분	사고	사망
1	2019	전체	4223	206
2	2020	전체	4039	223
3	2021	전체	4883	191

	년도	구분	사고	사망
1	2019	역주행	28	5
2	2020	역주행	33	3
3	2021	역주행	27	2

2. 일반 교통사고 데이터 만들기

	년도	구분	사고	사망
1	2019	일반	4223	206
2	2020	일반	4039	223
3	2021	일반	4883	191

	년도	구분	사고	사망	치명률
1	2019	역주행	28	5	17.857143
2	2020	역주행			
3	2021	역주행			

	년도	구분	사고	사망	치명률
1	2019	일반	4223	206	4.878049
2			4039	223	5.521169
3			4883	191	3.911530

3. 치명률계산

4. 치명률 평균 계산

기본 그래프

plot(벡터) : 가장 기본적인 함수

- main : 그래프 제목
- xlab , ylab : x, y축 제목
- xlim, ylim : c(a, b) x, y축 범위
- pch : 마커종류
- col : 색상 (1 ~ 8)
- black, red, green, blue, cyan, magenta, yellow, gray
- type : 연결(p, l, b, c, o, h, s, S, n)
- lty : 선종류 (0~6)
- blank, soild, dashed, dotted, dotdash, longdash, twodash

barplot(벡터) : 막대 그래프

시각화 그래프

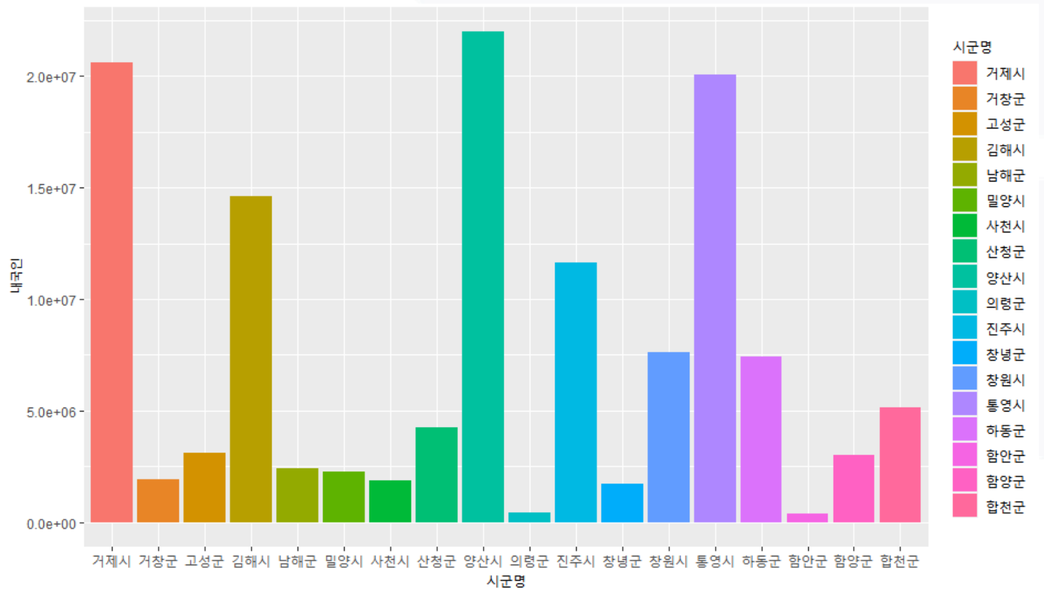
ggplot2 패키지

- 축 그림
  - ggplot(데이터명,aes(x=변수1,y=변수2))
- 그래프 그림
  - geom_bar( ): 막대도표
  - geom_histogram( ): 히스토그램
  - geom_boxplot( ): 박스플롯
  - geom_line( ):선 그래프
- 범례, 제목, 글씨 등 기타 옵션을 수정
  - labs( ) : 범례 제목 수정
  - xlabs( ), ylabs( ): x축 y축 이름 수정

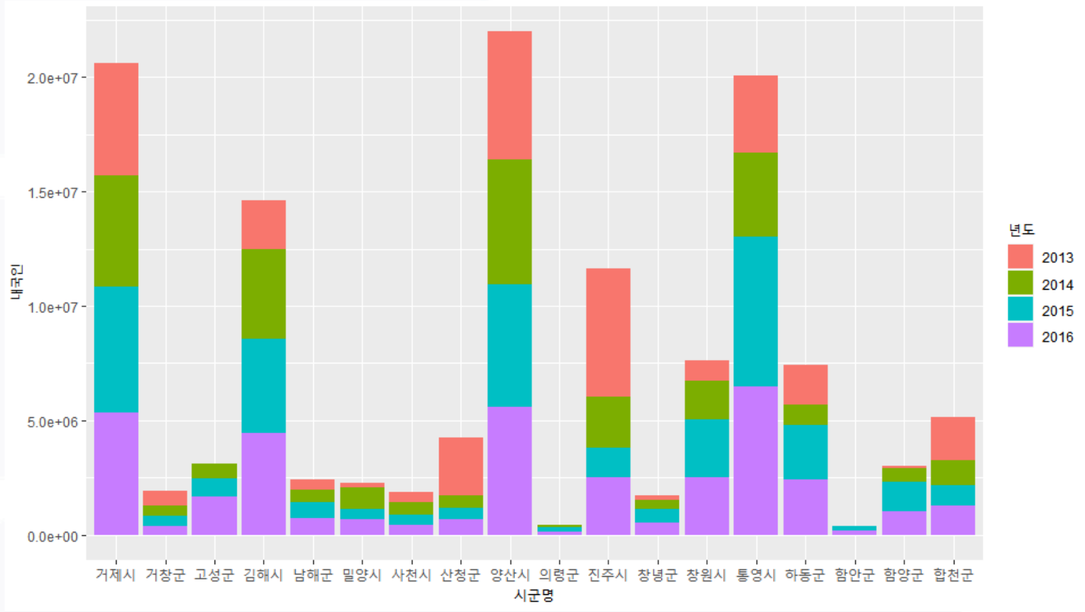
```
ggplot(mapping =aes(x=년도, y=사고, fill=구분), data=dfxl) +
  geom_bar(stat="identity", position=position_dodge())  +
  ggtitle('년도별 사고건수')+
  theme(plot.title = element_text(hjust = 0.5,size=20,face='bold'))
```

ggplot2 막대그래프

```
ggplot(data=df, aes(x=시군명, y=내국인, fill=시군명)) +
  geom_col()
```

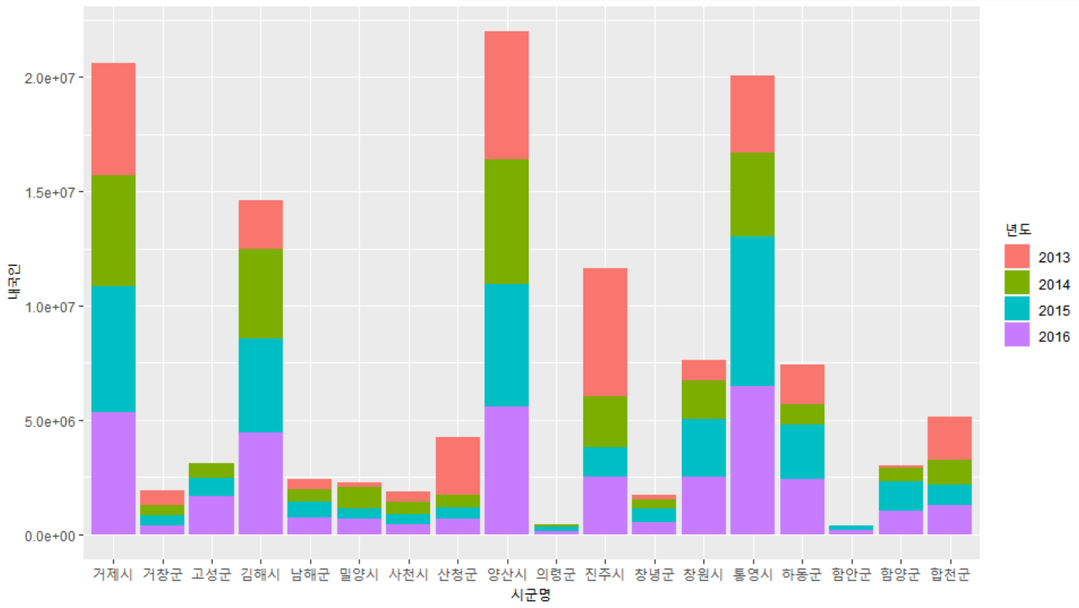



```
ggplot(data=df, aes(x=시군명, y=내국인, fill=년도)) +
  geom_bar(stat = "identity", position='dodge')
```

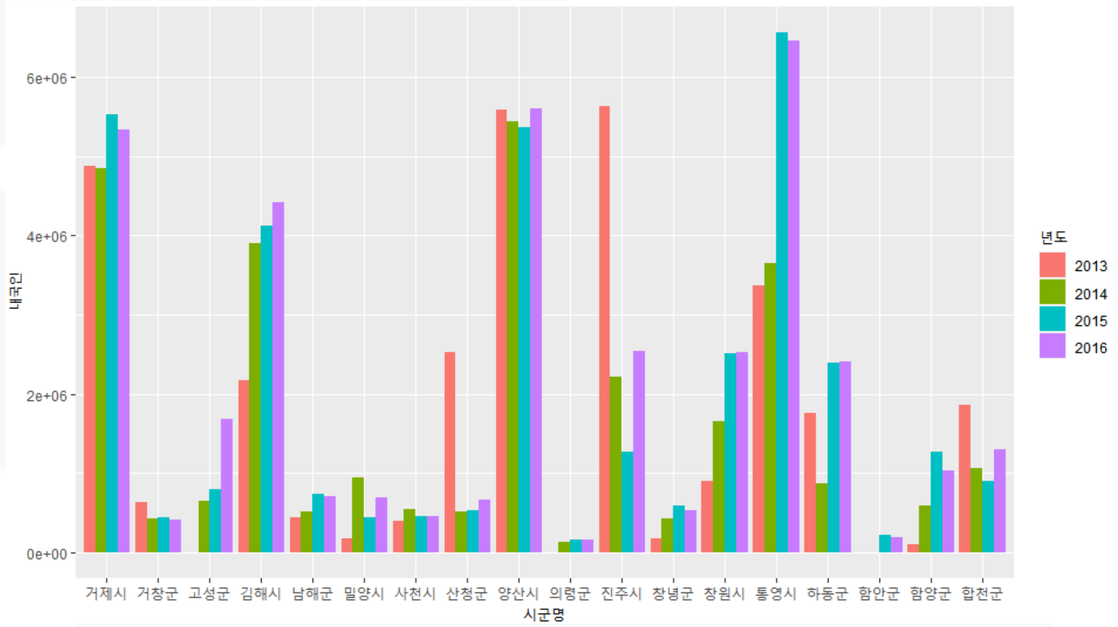


ggplot2 막대그래프

```
ggplot(data=df, aes(x=시군명, y=내국인, fill=년도)) +
  geom_bar(stat = "identity", position='dodge')
```

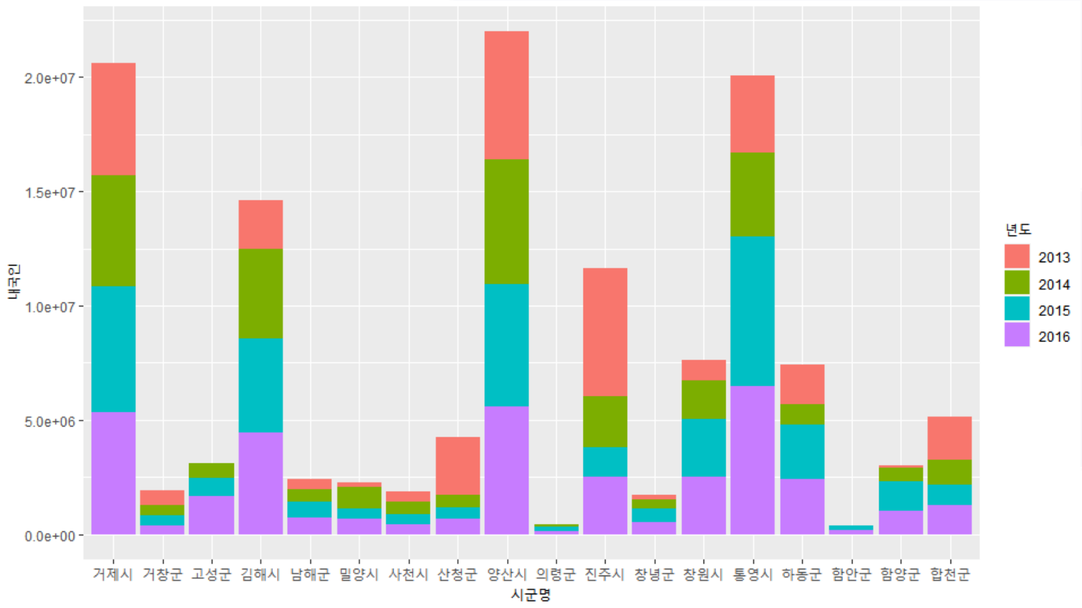



```
ggplot(data=df, aes(x=시군명, y=내국인, fill=년도)) +
  geom_bar(stat = "identity", position='dodge')
```

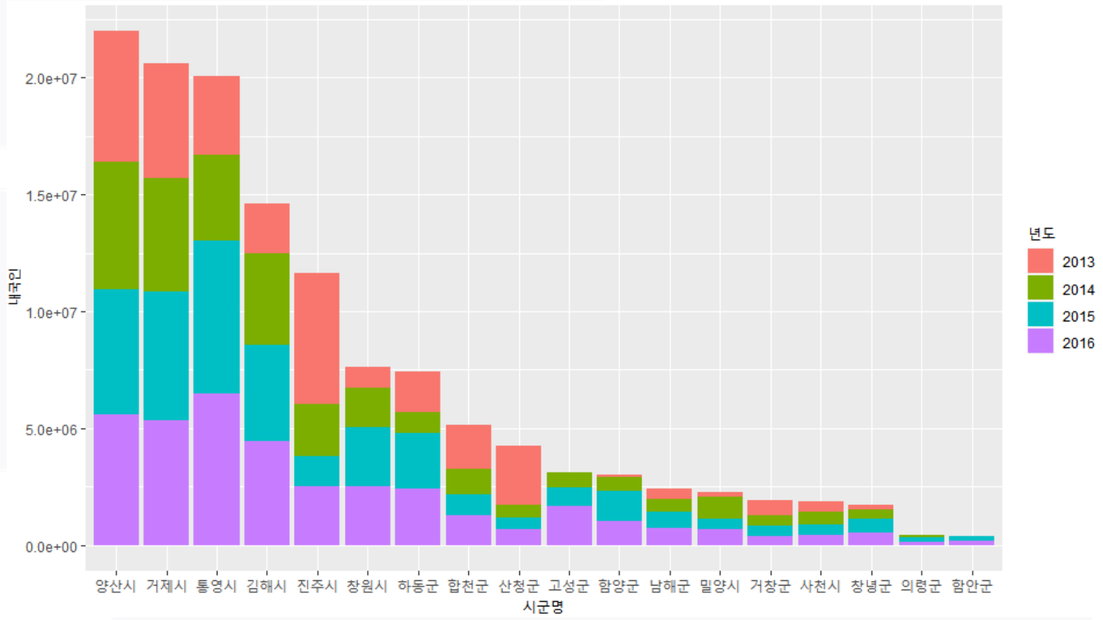


ggplot2 막대그래프

```
ggplot(data=df, aes(x=시군명, y=내국인, fill=년도)) +
  geom_bar(stat = "identity", position='dodge')
```

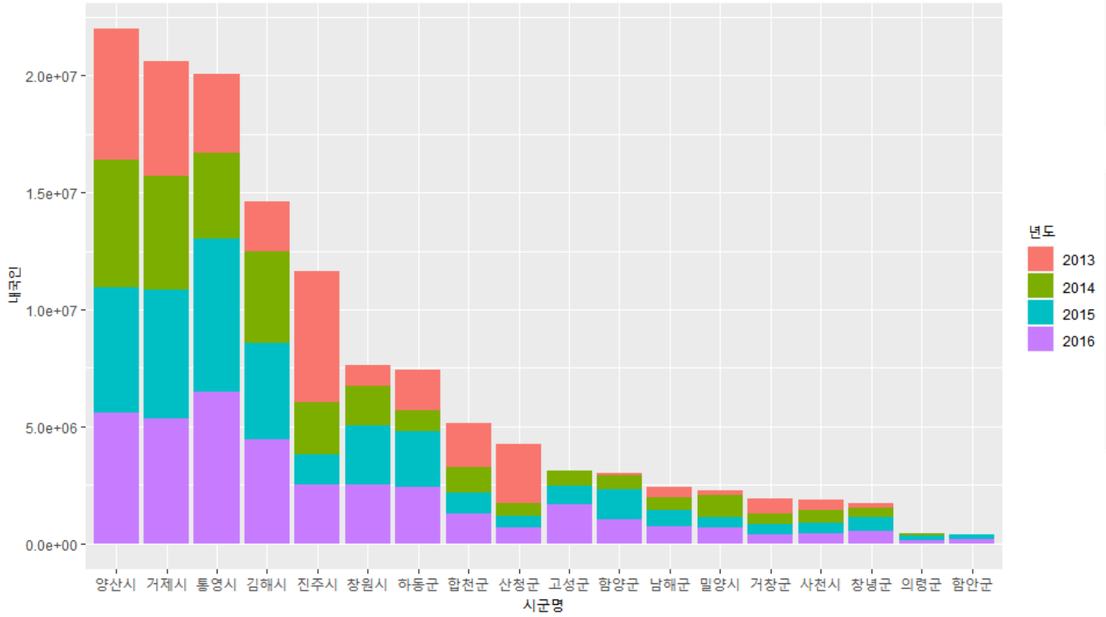



```
ggplot(data=df, aes(x=reorder(시군명, -내국인), y=내국인, fill=년도)) +
  geom_bar(stat = "identity")  +
  labs(x = "시군명")
```

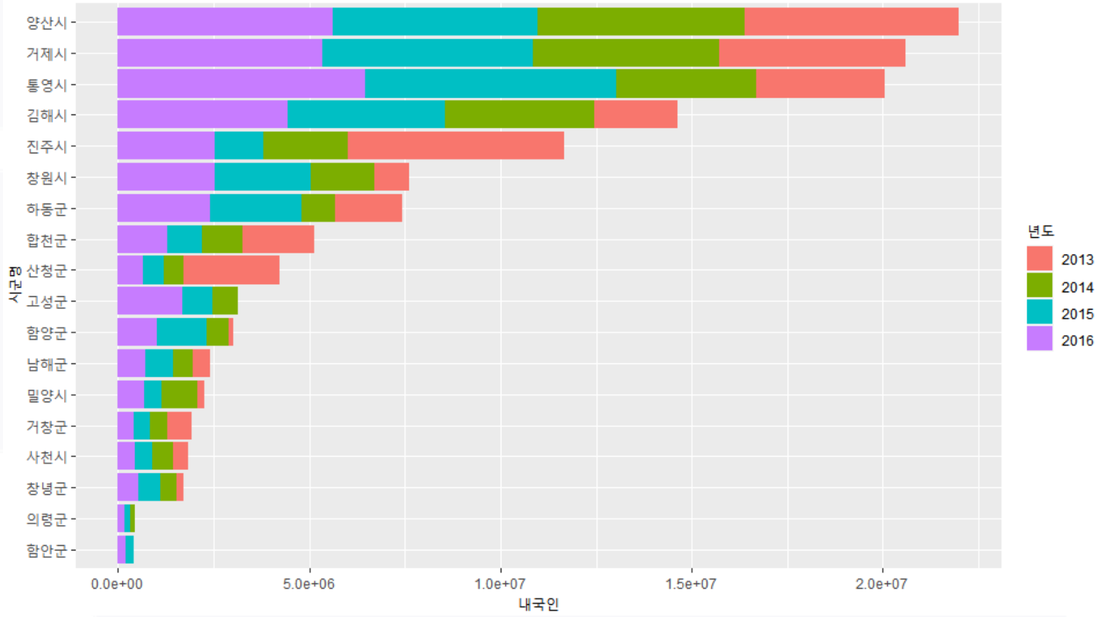


ggplot2 막대그래프

```
ggplot(data=df, aes(x=reorder(시군명, -내국인), y=내국인, fill=년도)) +
  geom_bar(stat = "identity")  +
  labs(x = "시군명")
```



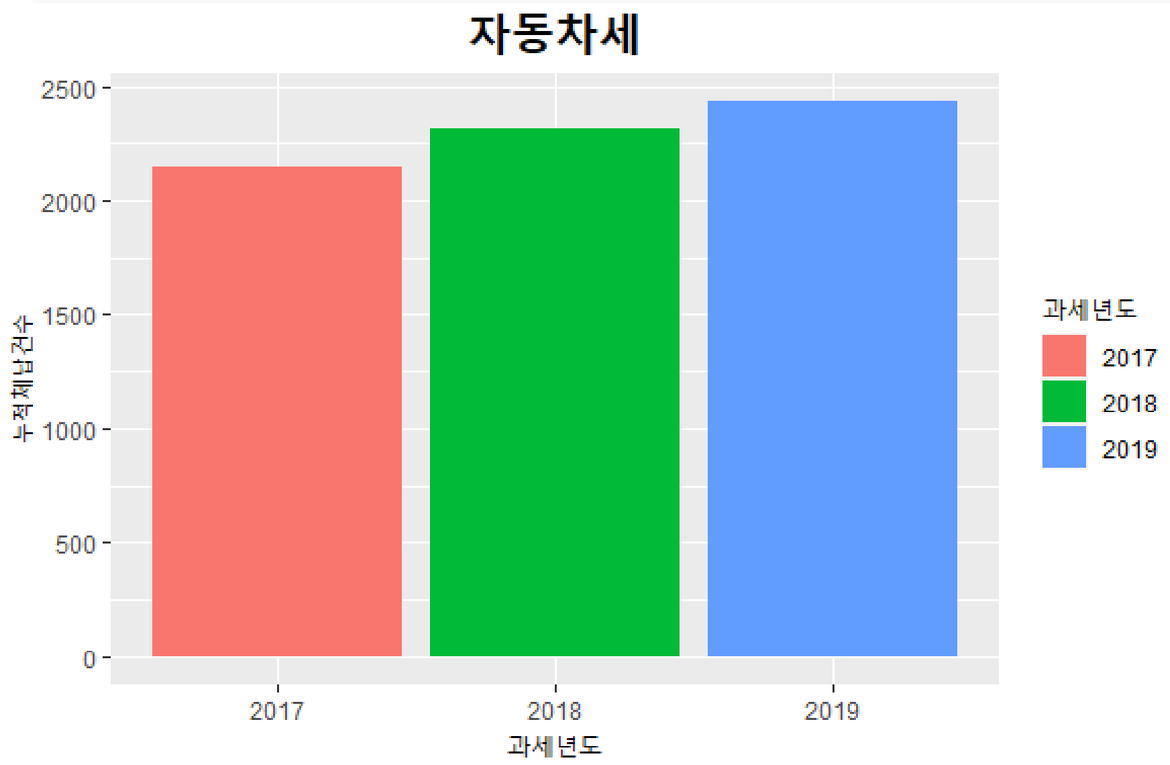

```
ggplot(data=df, aes(x=reorder(시군명, 내국인), y=내국인, fill=년도)) +
  geom_bar(stat = "identity")  +
  labs(x = "시군명") +
  coord_flip()
```



해결문제

부산시 체납현황 분석

- https://www.data.go.kr/data/15079162/fileData.do#tab-layer-file
- 3년간 세목별을 키로 누적 체납건수와 누적 체납금액



해결문제

기상개황 자료를 분석하여 월별 불쾌지수와 단계

- https://kosis.kr/statHtml/statHtml.do?orgId=735&tblId=DT_A1040&vw_cd=MT_ZTITLE&list_id=215_215A_735_73503_A&seqNo=&lang_mode=ko&language=kor&obj_var_id=&itm_id=&conn_path=MT_ZTITLE
- 불쾌지수 공식
  - DI = 0.81 * Ta + 0.01 * RH(0.99 * Ta − 14.3) + 46.3
    - DI: 불쾌지수
    - Ta: 건구온도(평균기온)
    - RH: 상대습도(평균상대습도)
- 불쾌지수 단계
  - 매우높음: 80이상
  - 높음: 75이상 80미만
  - 보통: 68이상 75미만
  - 낮음: 68미만