

# Assignment 1

2013-11086 김찬규

In [6]:

```
import numpy as np
import pandas as pd
import matplotlib
from sklearn.cluster import KMeans
from matplotlib import pyplot as plt
%matplotlib inline
matplotlib.rcParams['font.family'] = "AppleGothic" # for representing Korean in
the graph
```

In [2]:

```
# Load data
broadcast = pd.read_csv("data/broadcast.csv", encoding="euc-kr") # encoding="CP
949" also works
broadcast.head()
```

Out[2]:

	프로그램명	시청률	시청연령
0	SF8	9.2	25
1	뉴스데스크	3.5	39
2	뉴스룸	3.4	25
3	더쇼	3.5	16
4	뮤직뱅크	3.0	14

In [3]:

```
# Verify data
print(f>Data shape: {broadcast.shape}\n")
for key in broadcast.keys():
    print(f>Data count for {key}: {broadcast[key].count()})
```

Data shape: (15, 3)

Data count for 프로그램명: 15

Data count for 시청률: 15

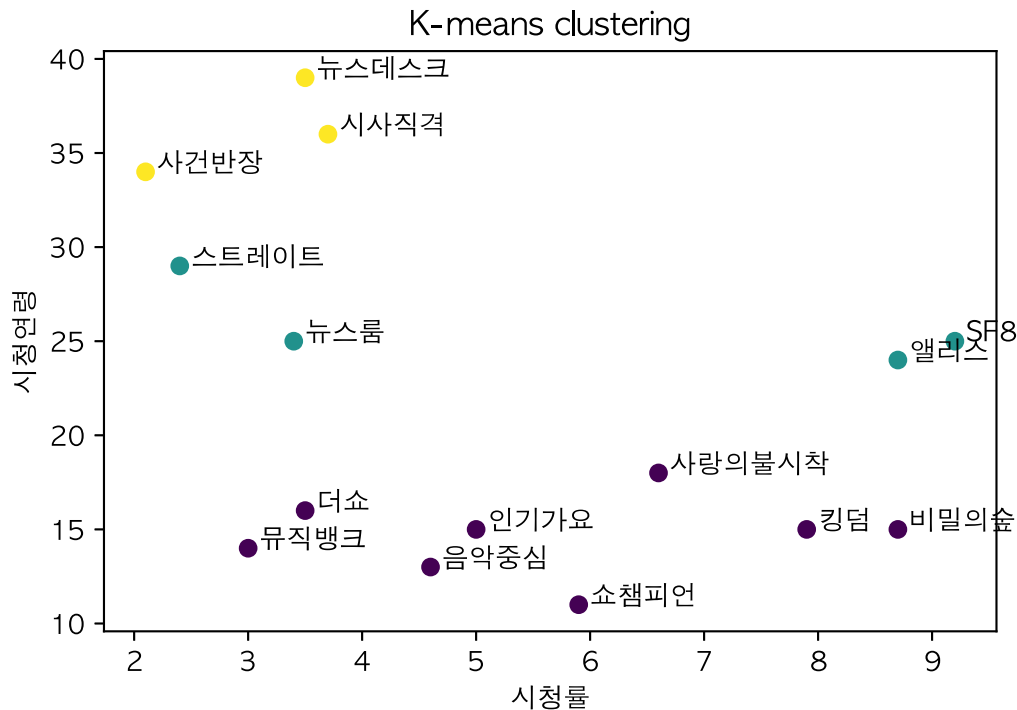
Data count for 시청연령: 15

In [4]:

```
# Cluster data
data = broadcast[["시청률", "시청연령"]]
k_means = KMeans(n_clusters=3, random_state=0)
k_means.fit(data)
result = k_means.predict(data)
```

In [5]:

```
# Visualize clustering result
plt.scatter(broadcast["시청률"], broadcast["시청연령"], c=result)
for idx, row in broadcast.iterrows():
    # print(row["프로그램명"])
    plt.annotate(row["프로그램명"], (row["시청률"] + 0.1, row["시청연령"]))
plt.xlabel("시청률")
plt.ylabel("시청연령")
plt.title("K-means clustering")
plt.show()
```



## Problem answer

1. 가로축을 시청률, 세로축을 시청연령으로 하여 군집화 결과를 시각화하여라.  
위 그래프를 통해 시각화하였음.
2. 군집화 결과, <뮤직뱅크>와 <음악중심>은 같은 군집으로 지정되었는가?  
위 그래프에 따르면 <뮤직뱅크>와 <음악중심>은 같은 군집으로 지정되었음.

In [ ]: