# 우편번호 앞 세자리 추출하여 새 컬럼 만들기

rows\_df['nw\_zp\_3'] = list(map(lambda x: x[:3], rows\_df['nw\_zp']))

rows\_df['nw\_zp'] = list(map(lambda x: int(x), rows\_df['nw\_zp']))

rows\_df = rows\_df.sort\_values(by=['nw\_zp'], axis=0).reset\_index(drop=True)

rows\_df[rows\_df['nw\_zp']==8866]

# 서울만 추출

rows\_df = rows\_df[:600838]

# 개인소득금액 chr -> int 변환

rows\_df['indv\_incm\_amt'] = list(map(lambda x: int(x), rows\_df['indv\_incm\_amt']))

# 우편번호 앞 세자리(구 단위)로 grouping

grouped = rows\_df.groupby(rows\_df['nw\_zp\_3'])

# 지역별 개인소득금액 평균(단위 : 만원)

grouped\_mean = grouped['indv\_incm\_amt'].mean()

import matplotlib.pyplot as plt

plt.hist(grouped\_mean)

plt.show()

grouped\_mean\_df = pd.DataFrame(grouped\_mean)

grouped\_mean\_df.rename(columns = {'indv\_incm\_amt':'mean'}, inplace = True)

grouped\_mean\_df

# 지역별 개인소득금액 합(단위 : 만원)

grouped\_sum = grouped['indv\_incm\_amt'].sum()

grouped\_sum

grouped\_sum\_df = pd.DataFrame(grouped\_sum)

grouped\_sum\_df.rename(columns = {'indv\_incm\_amt':'sum'}, inplace = True)

grouped\_sum\_df

# 지역별 row 개수

grouped\_cnt = grouped['indv\_incm\_amt'].count()

grouped\_cnt

grouped\_cnt\_df = pd.DataFrame(grouped\_cnt)

grouped\_cnt\_df.rename(columns = {'indv\_incm\_amt':'count'}, inplace = True)

grouped\_cnt\_df

# 지역별 median

grouped\_med = grouped['indv\_incm\_amt'].median()

grouped\_med

grouped\_med\_df = pd.DataFrame(grouped\_med)

grouped\_med\_df.rename(columns = {'indv\_incm\_amt':'median'}, inplace = True)

grouped\_med\_df

final\_df = pd.concat([grouped\_cnt\_df, grouped\_mean\_df, grouped\_med\_df, grouped\_sum\_df], axis = 1)

final\_df

final\_df.to\_csv('서울시구별개인소득금액데이터.csv')

final\_df[final\_df['median']>4000]

check = rows\_df[rows\_df['nw\_zp\_3']=='067']['indv\_incm\_amt']

plt.boxplot(check)



