

암 사망률과 흡연/음주 관계

이윤정

1. 분석 목적

저는 '서울시의 암과 흡연, 음주 간의 관계'를 분석주제로 결정하였으며, 사망까지 도달 할 수 있는 암이라는 병과 흡연, 음주의 분석을 통해 병에 대한 예방을 목적으로 분석주제를 설정하게 되었습니다. 또한 성별에 따라서 암 사망 발생 비율에 차이가 있는지, 음주와 흡연의 차이가 존재하는지 분석하고자 합니다.

2. 분석 데이터 설명

'서울 열린데이터 광장'(<https://data.seoul.go.kr/>) 의 총 3가지 데이터 '서울시 주요 암 사망원인별 사망(성별/구별)통계(2017)', '서울시 고위험 음주율 통계(2017)', '서울시 현재흡연율 통계(2017)'를 활용하였습니다. 세 가지의 데이터는 모두 성별에 따른 데이터를 가지고 있으며 '구'라는 공통적인 행을 가지고 있습니다.

첫 번째 데이터인 '서울시 주요 암 사망원인별 사망(성별/구별)통계(2017)' 데이터는 각 암별 사망자수와 사망률, 각 암별/성별 사망자수와 사망률을 총 31개의 열로 가지고 있습니다. 두 번째 데이터와 세 번째 데이터를 통합하여 '서울시 고위험 음주율, 흡연율 통계(2017)'라는 파일을 만들었으며, 이는 음주율, 성별 음주율, 흡연율, 성별 흡연율 4가지 열로 이루어져 있습니다.

3. 데이터분석 결과

1) 탐색적 자료 분석 - 막대그래프

폐암	간암	위암	대장암	췌장암	전체
강북구	강북구	종로구	종로구	중구	강북구
용산구	종로구	강북구	중구	강북구	종로구
서대문구	종량구	금천구	강북구	용산구	중구

성별을 배제하고 사망률을 높은 순으로 정렬했을 때의 상위 3개 구 들입니다. 강북구는 모든 열에 속해 있을 만큼 사망률이 높다고 보이며, 종로구, 중구 또한 공통적으로 보입니다.

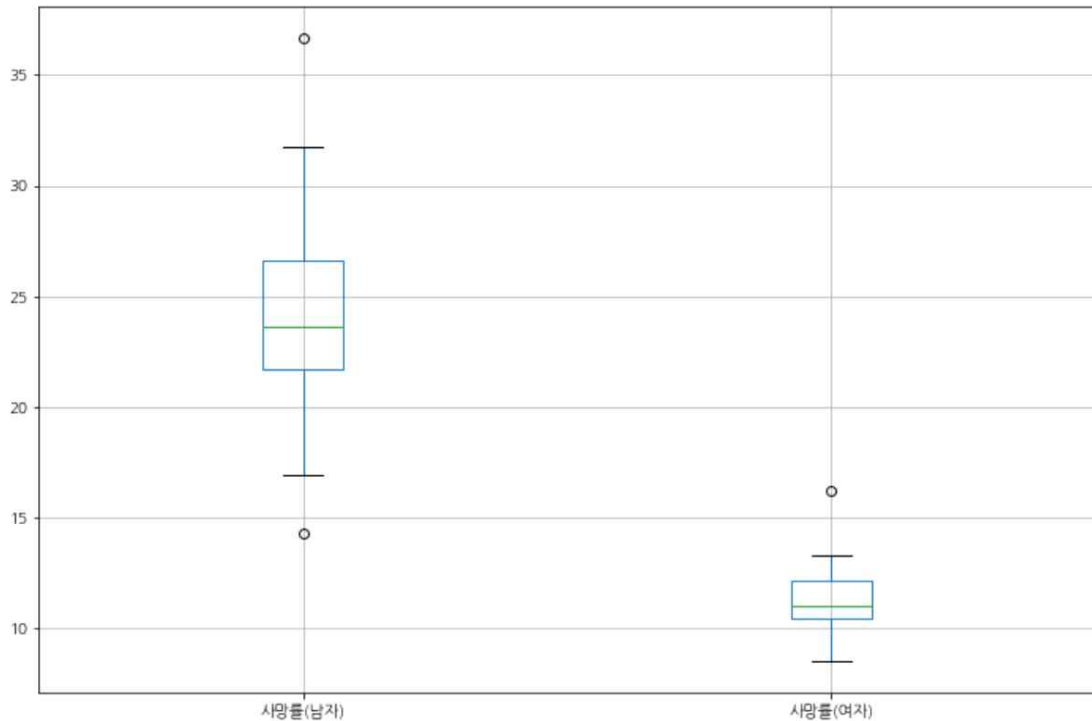
폐암(남자)	간암(남자)	위암(남자)	대장암(남자)	췌장암(남자)	전체(남자)
강북구	강북구	종로구	종로구	중구	강북구
종량구	서대문구	강북구	도봉구	강북구	중구
용산구	종로구	중구	강북구	종량구	종로구

위 표는 성별이 남자일 경우 각 사망률을 높은 순으로 정렬했을 때의 상위 3개 구입니다. 강북구는 이번에도 모든 열에 속해 있으며, 종량구도 살짝 높아 보이며 중구와 종로구 역시 공통적으로 보입니다.

폐암(여자)	간암(여자)	위암(여자)	대장암(여자)	췌장암(여자)	전체(여자)
종로구	금천구	종로구	중구	용산구	종로구
영등포구	성동구	금천구	종로구	종로구	용산구

서대문구	용산구	양천구	강동구	중구	노원구
------	-----	-----	-----	----	-----

위 표는 성별이 여자일 경우 각 사망률을 높은 순으로 정렬했을 때의 상위 3개 구입니다. 앞서 주어진 두 표와는 차이가 있는 것으로 보입니다. 종로구, 용산구, 금천구가 사망률이 높은 것으로 보입니다. 전체 사망률 표가 성별이 남자인 사망률 표와는 구에 대한 차이가 없고, 성별이 여자인 사망률 표와는 구에 대한 차이가 있는 것으로 보아 암 사망률의 비율은 남자가 더 높다는 추측이 가능합니다.



Boxplot 결과 추측과 일치합니다. 남자의 사망 비율이 여자의 사망 비율보다 현저히 높으며 둘의 평균 차이가 존재한다고 예상할 수 있습니다.

또한 암 사망률과 음주와 흡연 간의 관계파악을 위해 위에서 분석한 상위 3개 구의 음주 흡연율을 주목할 필요가 있습니다.

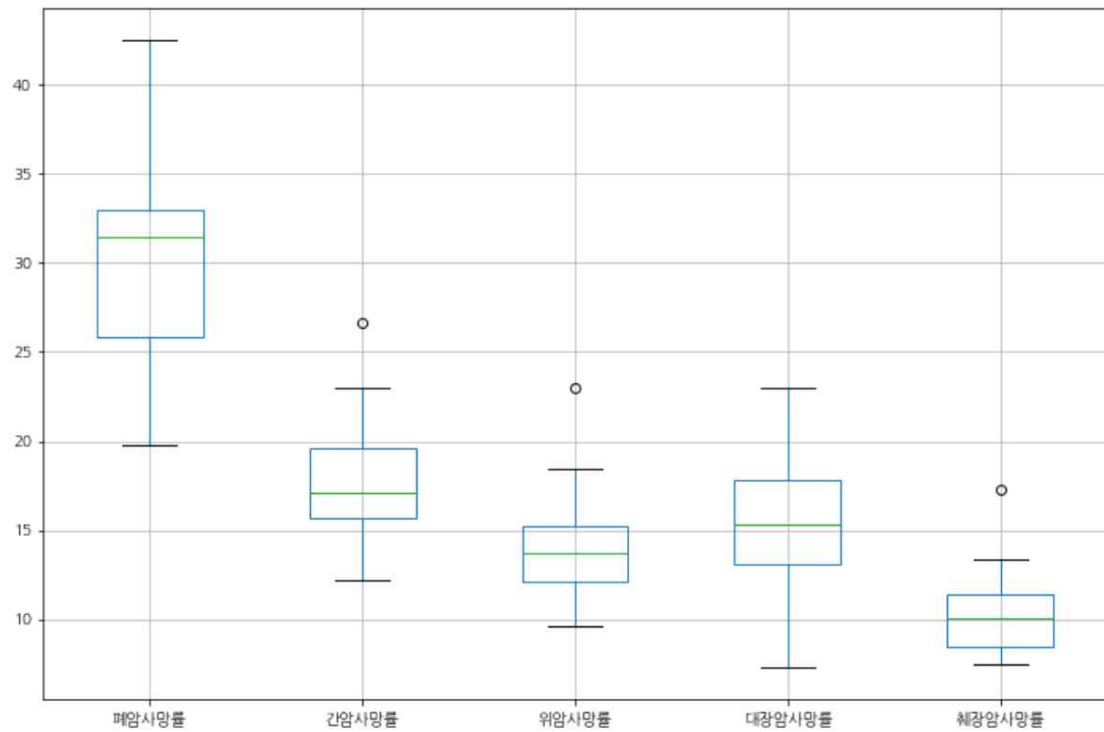
다음은 음주 흡연율 데이터에 관한 탐색적 자료 분석입니다.

음주율	흡연율	음주율(남자)	음주율(여자)	흡연율(남자)	흡연율(여자)
금천구	종량구	금천구	동대문구	종량구	은평구
동대문구	강북구	동대문구	강북구	강북구	동대문구
종량구	동대문구	종량구	종량구	동대문구	강북구

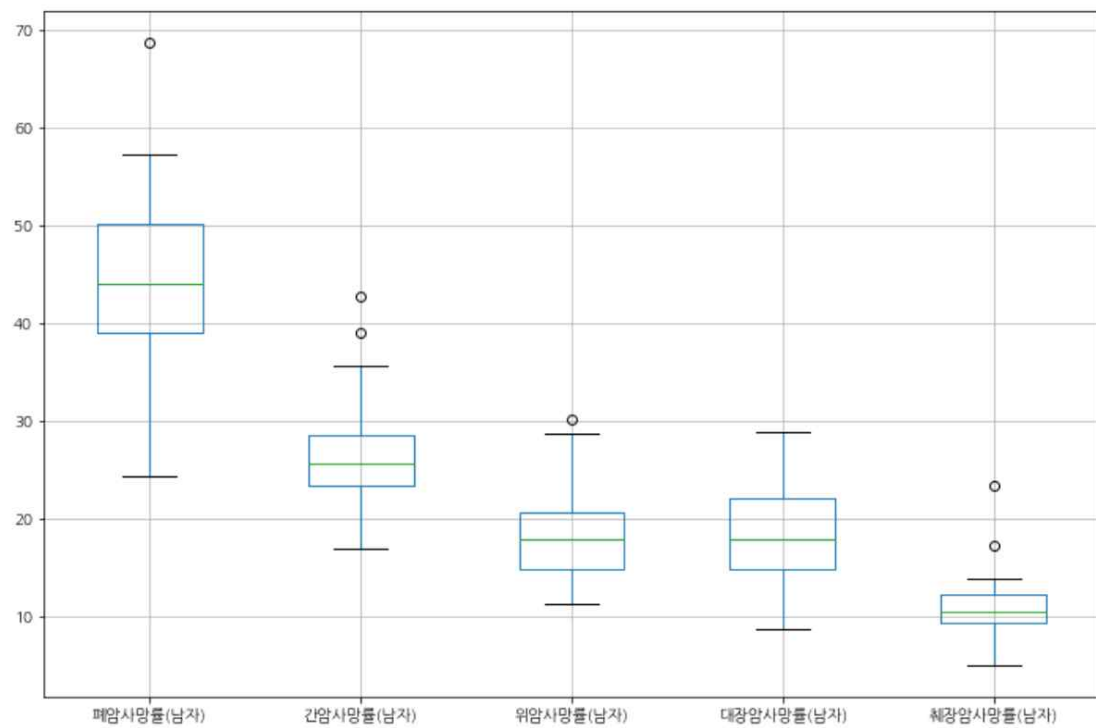
암 사망률이 높은 구인 강북구와 종량구가 음주율과 흡연율에서도 공통적으로 보입니다. 그러나 음주와 흡연이 암 사망률과 상관관계가 존재하는지는 알 수 없습니다.

2) 시각화

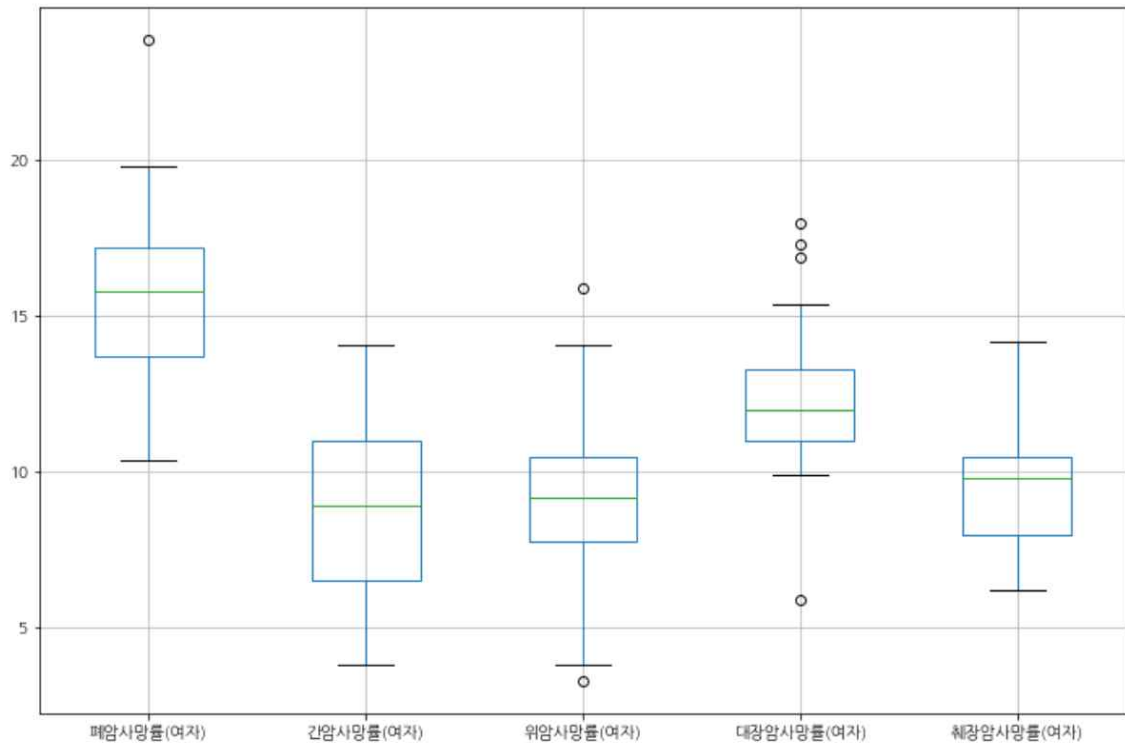
분포 확인



위 그래프는 암 별 사망률의 Boxplot입니다. 폐암 사망률이 가장 높은 비율을 차지 하고 있으며 그 다음으로 간암, 대장암, 위암, 췌장암 순으로 이어짐을 확인할 수 있습니다.

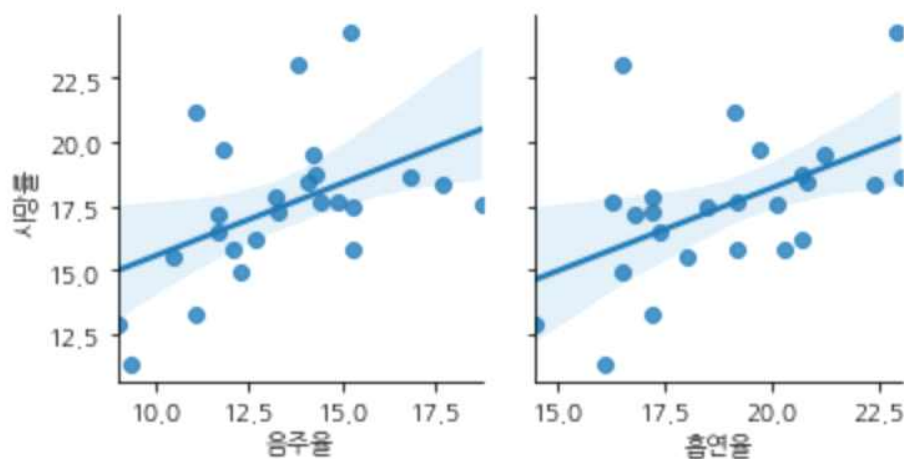


성별이 남자인 경우의 암 별 사망률 Boxplot입니다. 전체 사망률과 같이 폐암 사망률이 가장 높은 비율을 차지하며, 간암, 대장암, 위암, 췌장암 순으로 이어집니다.

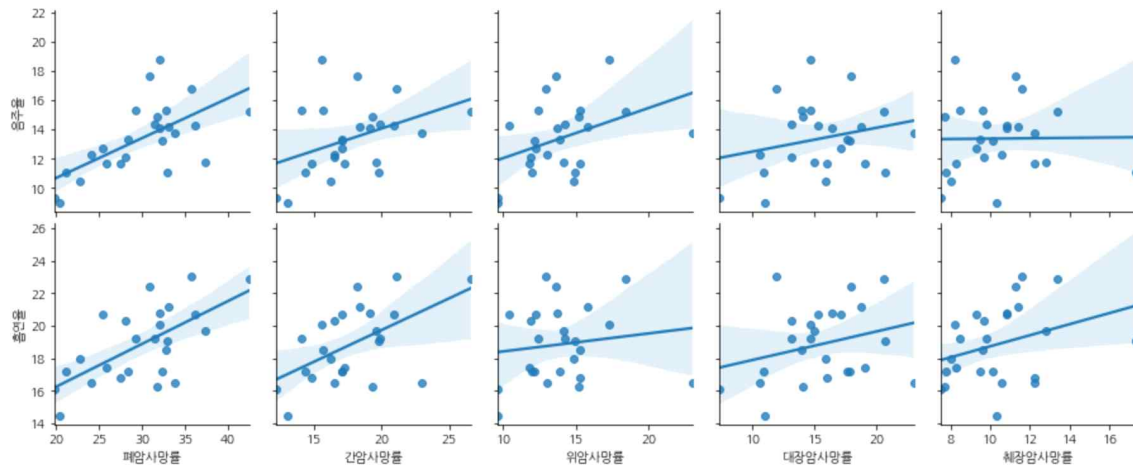


다음은 성별이 여자인 경우 암 별 사망률 Boxplot입니다. 이 역시 폐암이 가장 높으나, 대장암이 두 번째, 간암, 위암, 췌장암이 비슷한 비율선상에 존재합니다.

산점도

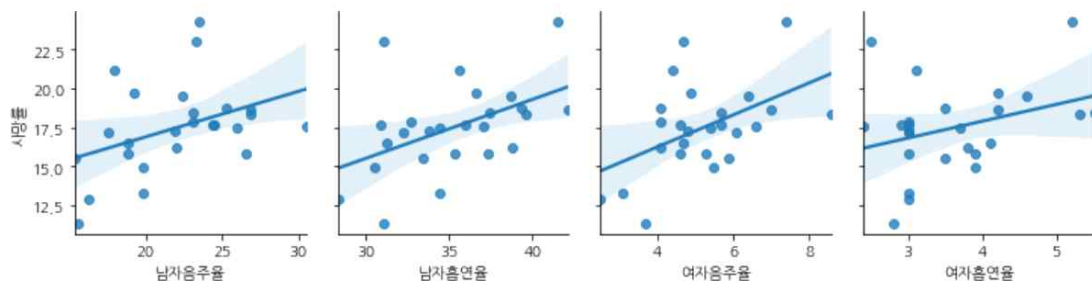


다음은 음주율과 흡연율 그리고 사망률 간의 상관관계입니다. 강한 상관관계를 띄고 있지는 않으나 약한 양의 상관관계를 보이고 있습니다. 더 세부적인 상관관계를 알아보기 위하여 각 암 별로 음주율과 흡연율 간의 상관관계를 분석했습니다.

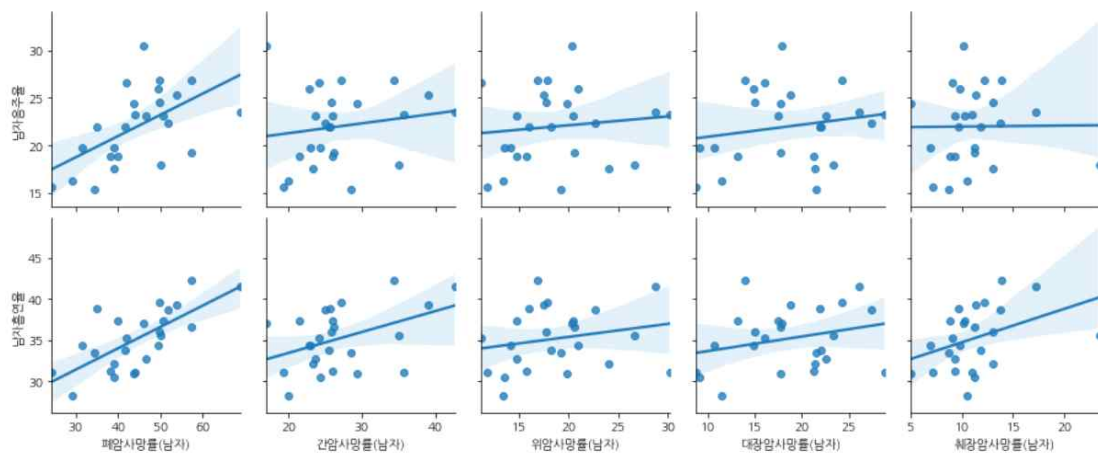


폐암사망률에서만 음주량과 흡연율 간의 양의 상관관계가 존재하며 간암은 흡연율과 약한 양의 상관관계를 보입니다. 그리고 나머지 암들의 사망률은 상관관계가 없는 것으로 보입니다. 따라서 전체 사망률과 음주 흡연율 간의 약한 양의 상관관계가 나타난 이유는 폐암 사망률과 음주 흡연율 간의 양의 상관관계 때문이라고 볼 수 있습니다. 이 결과에 따라 폐암에 걸린 환자들은 회복을 위해 술과 담배를 멀리해야 할 필요가 있습니다.

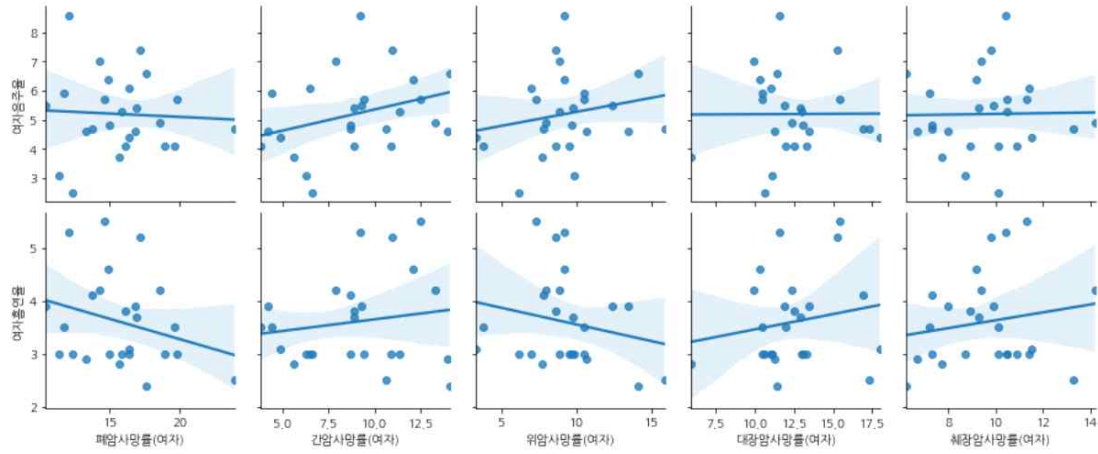
성별에 따른 사망률과 흡연 음주율의 상관관계?



다음은 성별 음주 흡연율과 사망률 간의 상관관계입니다. 4개의 그래프 모두 약한 양의 상관관계 혹은 아무 관계도 가지지 않는 것처럼 보입니다.

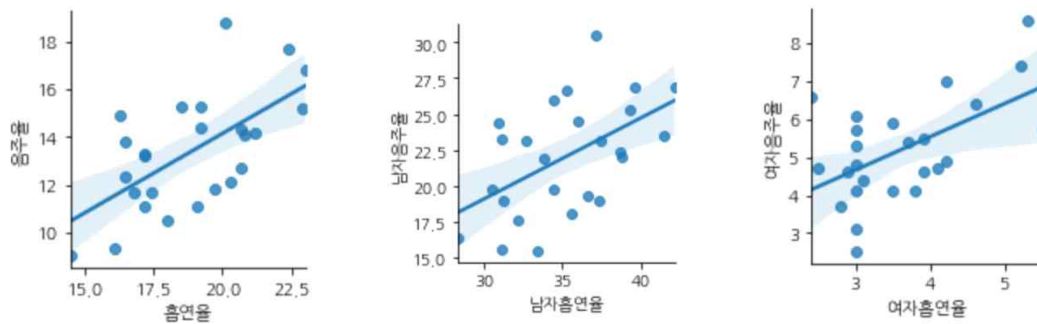


성별이 남자인 경우의 사망률과 음주 흡연율의 상관관계 역시 폐암 사망률만 양의 상관관계가 존재하는 것으로 보입니다. 간암 또한 흡연율과 약한 양의 상관관계를 보이고 나머지 암들은 모두 상관관계가 존재하지 않습니다.



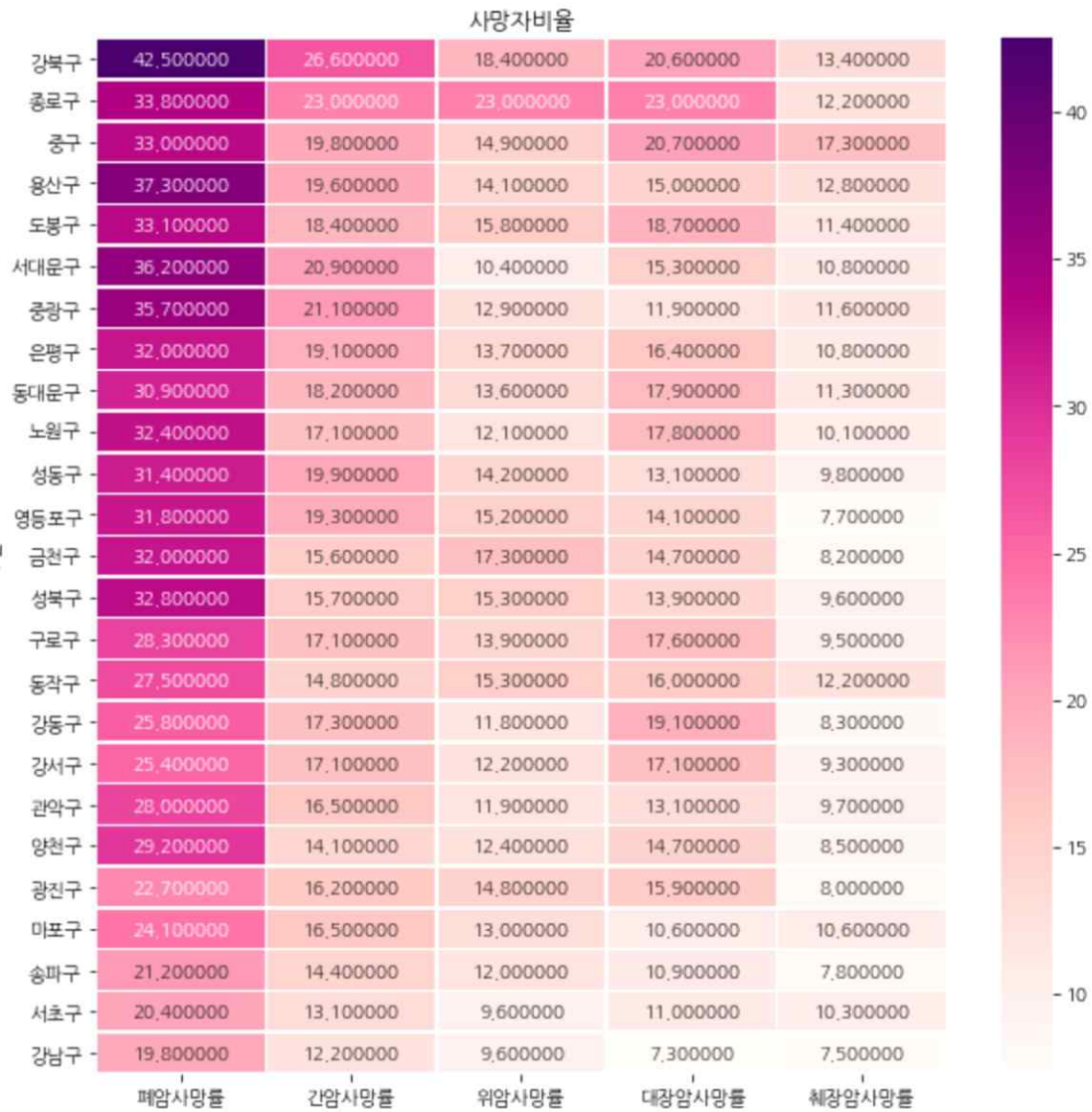
성별이 여자인 경우의 사망률과 음주 흡연율의 상관관계는 전혀 없는 것처럼 보입니다.

흡연과 음주의 상관관계

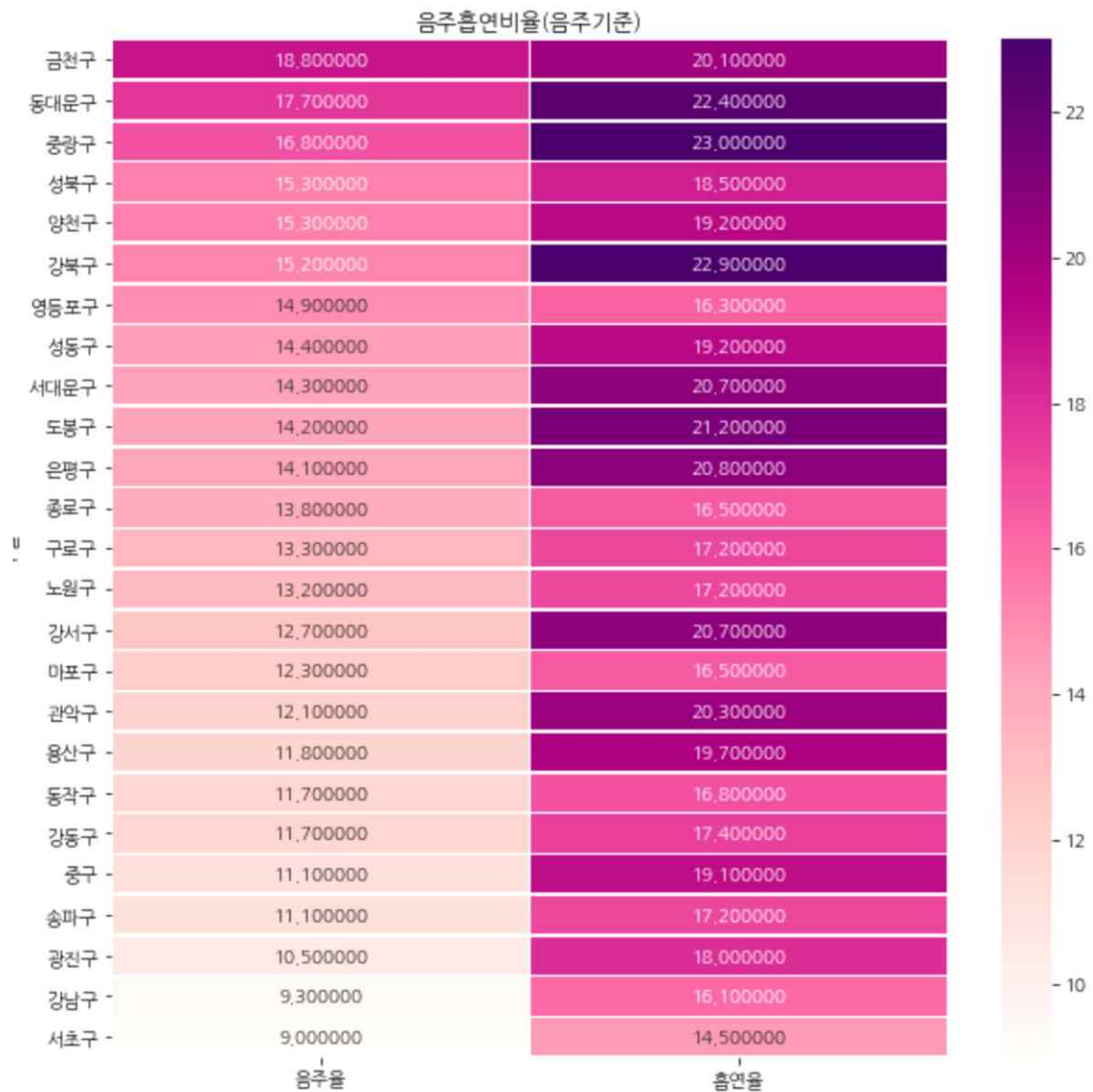


흡연과 음주율은 성별과 전체 모두 양의 상관관계를 가지고 있는 것으로 보입니다.

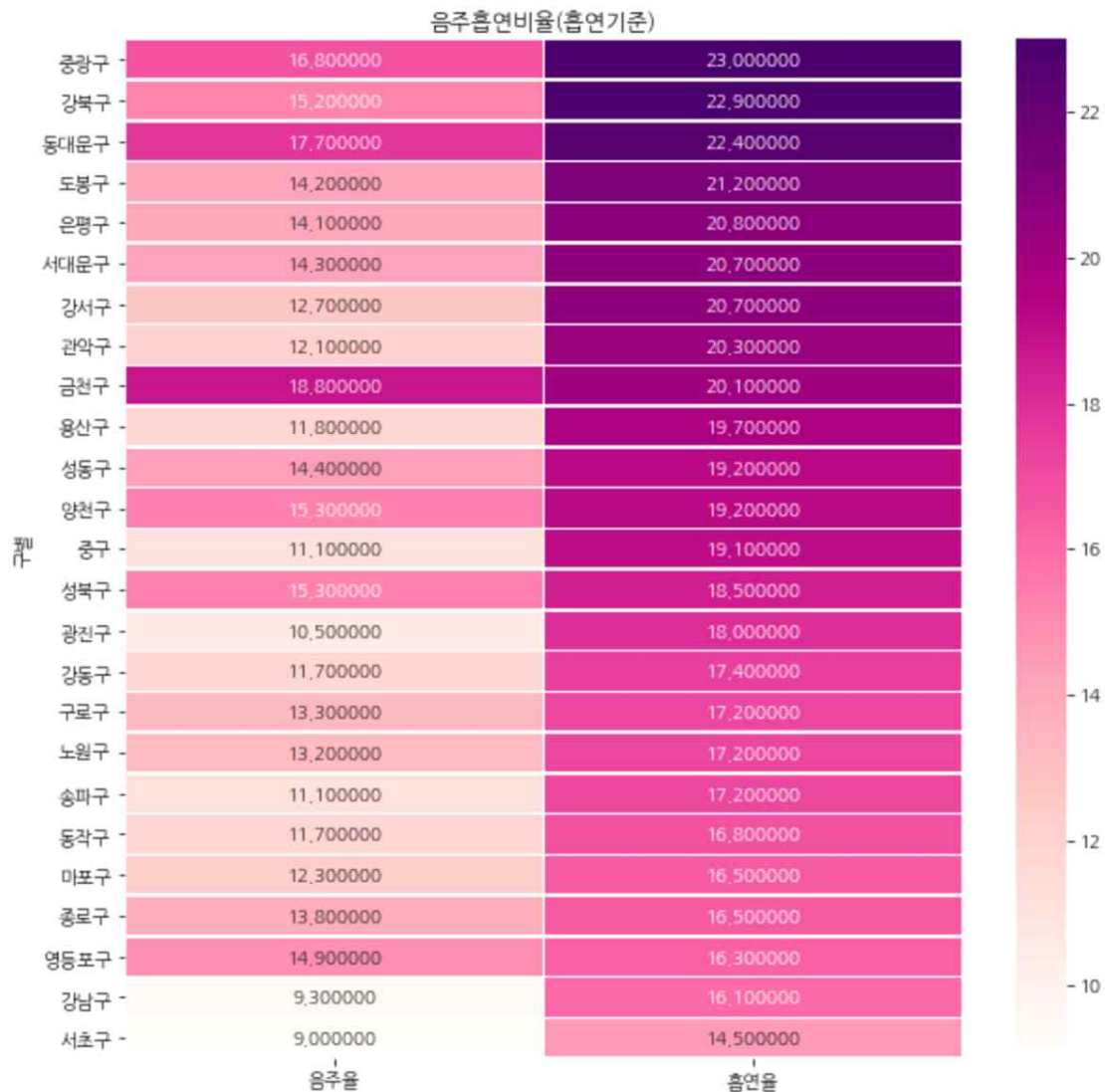
히트맵



위의 그래프를 보면 강북구 종로구 중구 순으로 사망자 비율을 확인할 수 있으며 폐암의 사망자 비율이 가장 높은 것을 볼 수 있습니다.



위는 음주율을 기준으로 한 읍·면·동 흡연 히트맵으로 금천구 동대문구 중랑구 순으로 음주율이 높은 것을 알 수 있습니다. 앞선 사망자 비율 히트맵과 공통되는 부분이 크지 않다고 보입니다.



위는 흡연율을 기준으로 한 음주흡연 히트맵으로 종랑구 강북구 동대문구 순으로 높은 흡연율을 보입니다. 또한 사망자 비율의 상위 10구인 (강북구, 종로구, 중구, 용산구, 도봉구, 서대문구, 중랑구, 은평구, 동대문구, 노원구) 와 (중랑구, 강북구, 동대문구, 도봉구, 은평구, 서대문구, 강서구, 관악구, 금천구, 용산구)가 종로구, 중구, 노원구를 제외하고 일치하며 높은 상관성을 보입니다.