

COVID-19와 전염병간 연관성

| | |
|-----|--------------------------|
| 과목 | 탐색적 자료분석 |
| 교수님 | 주용성 교수님 |
| 팀 | 18Team |
| 전공 | 통계학과 통계학과 |
| 학번 | 2017110515 2017110522 |
| 이름 | 이용휘 이정호 |

목차

I. 서론

II. 본론

- 계절별 감염자수 추이
- 감염재생산지수를 통한 감염병의 유행여부 파악
- 각 감염병의 성과 연령별 차이여부 파악
- 코로나19와 감염 추세가 비슷한 감염병의 존재 유무
- 감염병들 간의 상관관계 파악

III. 결론

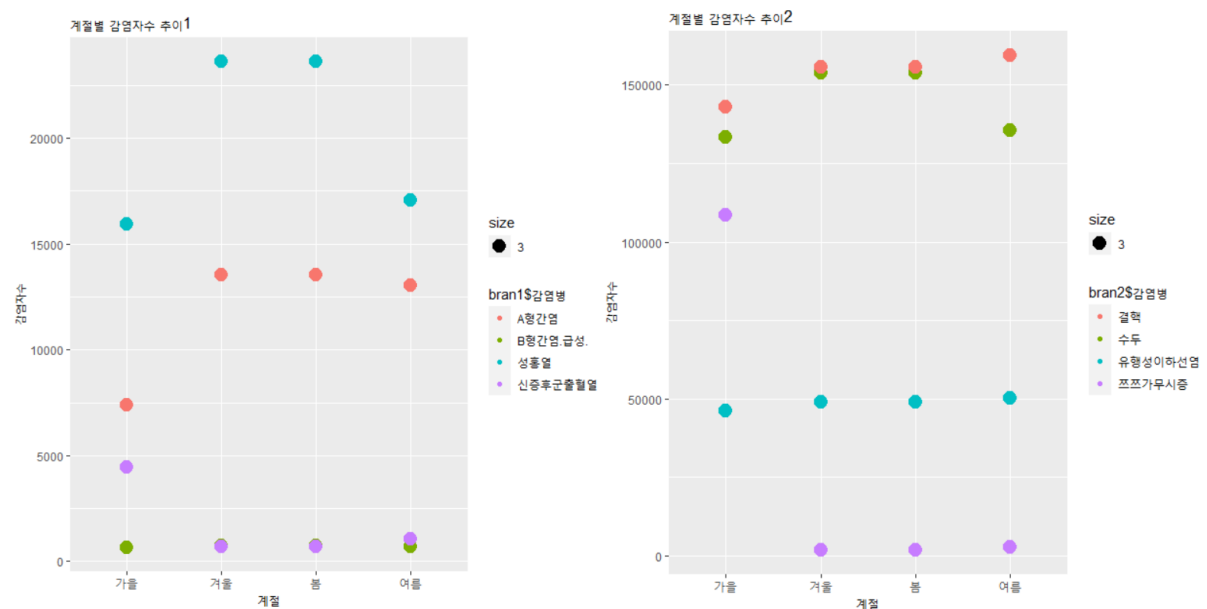
IV. 참고문헌

I. 서론

코로나가 유행인 가운데, 다른 감염병에 흥미가 생겼다. 그래서 지속적으로 발생하고 있는 감염병들의 데이터를 수집해, 각 감염병이 계절 별로 감염자 수에 차이가 있는지, 각 질병이 유행한 시기가 있었는지, 각 감염병이 성과 연령별로 감염자 수에 차이가 있는지, 코로나19와 감염 추세가 비슷한 감염병이 있는지, 감염병들 간에 상관관계가 있는지를 알아보려고 한다. 수집한 자료는 A형 간염, B형 간염(급성), 유행성이하선염, 수두, 성홍열, 쯤쯤가무시증, 신증후군출혈열, 후천성면역결핍증, 코로나19의 감염자 수이다. A형 간염~후천성면역결핍증의 감염자 수는 연도별, 월별, 성과 연령별로 각각 수집하였고, A형 간염과 B형 간염(급성)은 2011년~2019년, 수두는 2005년~2019년, 코로나19는 2020년 2월 13일~2020년 12월 5일까지의 자료를 수집하였다. 코로나19는 수집한 시기에 발생한 감염자 수만 월별, 일별로 수집하였다.

II-1. 계절별 감염자수 추이

계절별로 감염병의 감염자 수에 차이가 있는지 파악하기 위해 해당 자료를 분석하였다. 계절별 감염자 수 추이는 감염자 수가 각 계절마다 5만여명 이하인 그룹과 그렇지 않은 그룹으로 분류하여 보기 편하게 하였다. 다음은 감염병의 감염자 수를 계절별로 구분하여 qplot으로 나타낸 것이다.



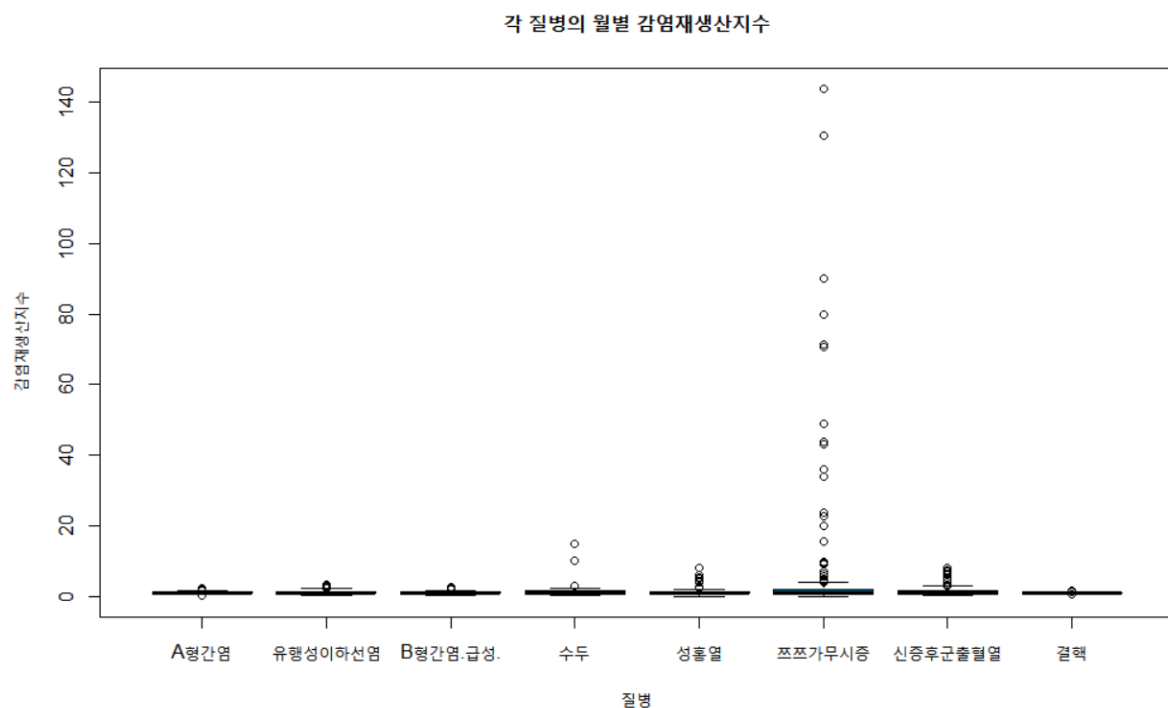
B형 간염(급성)은 유행성이하선염 계절마다 거의 차이가 없고, 성홍열과 수두는 겨울과 봄에 더 많은 감염자가 나타났다. A형 간염과 결핵은 타 계절에 비해 가을에 감염자 수가 적은 한편, 신증후군출혈열과 쯤쯤가무시증은 반대로 가을에 유독 더 많은 감염자가 나타났다.

감염원을 크게 RNA바이러스, DNA바이러스, 그리고 세균으로 보았을 때, RNA바이러스가 원인인

질병에는 A형간염, 유행성이하선염, 신증후군출혈열이 있다. DNA바이러스가 원인인 질병에는 B형간염(급성), 수두가 있으며 세균이 원인인 질병에는 성홍열, 쯔쯔가무시증, 결핵이 있다. 해당자료에서는 계절별 감염자 수가 감염원 별로 특정한 차이를 보이지 않고 있어, 감염원이 계절별로 나타나는 감염자 수 차이에 영향을 준다고 보기에는 어렵다고 판단된다.

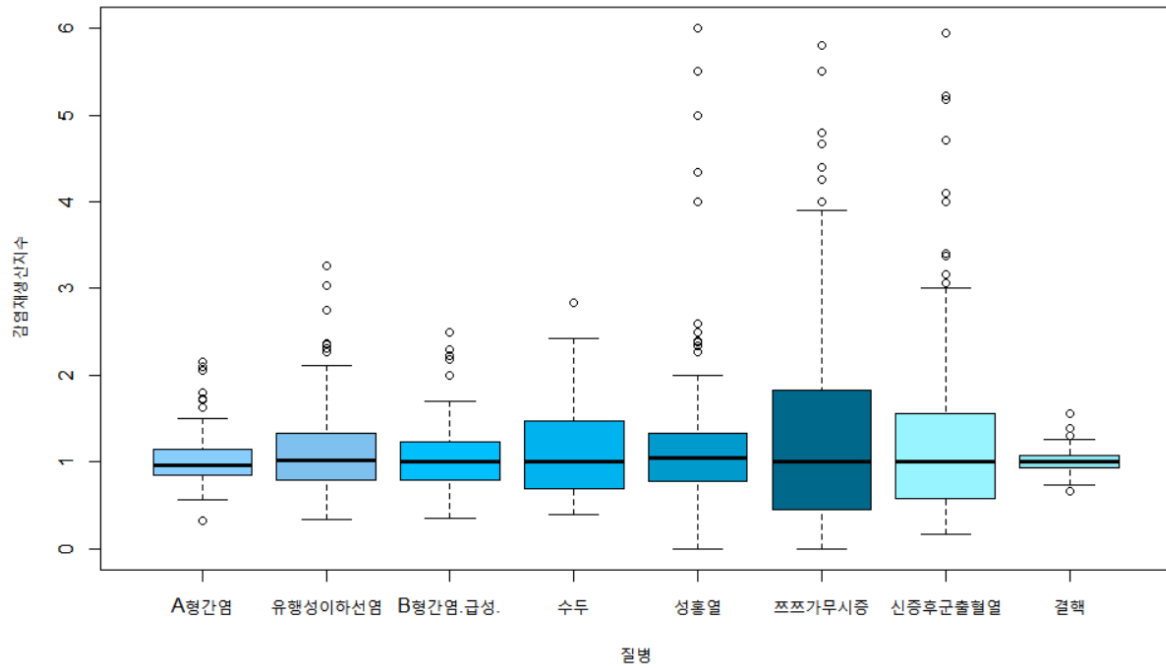
II-2. 감염재생산지수를 통한 감염병의 유행여부 파악

각 감염병이 유행한 때가 있었는 지를 알아보기 위해 감염재생산지수를 이용하였다. 감염재생산지수란 확진자 한 명이 몇 명을 감염시킬 수 있는지를 보여주는 지표이다. 예를 들어 감염재생산지수가 2라면 확진자 한 명이 두 명을 감염시키고 있다는 뜻이다. 다음은 감염재생산지수를 월단위로 구하여 boxplot으로 나타낸 것이다. (후천성면역결핍증은 월별 자료가 거의 존재하지 않아서 빼고 분석하였다.)

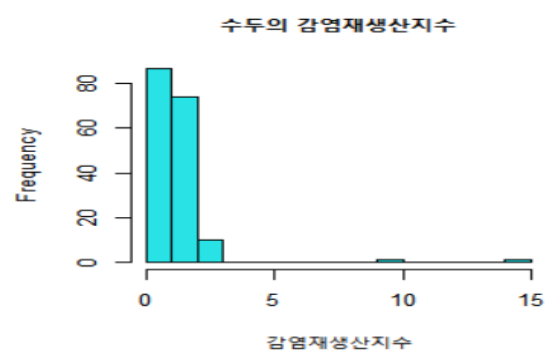
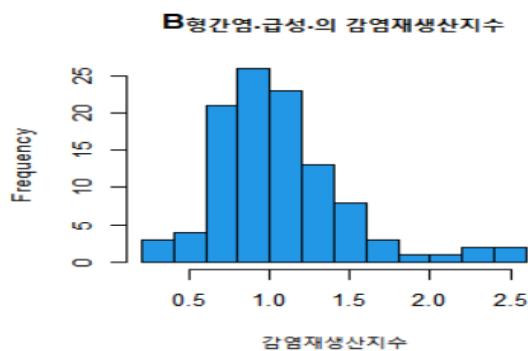
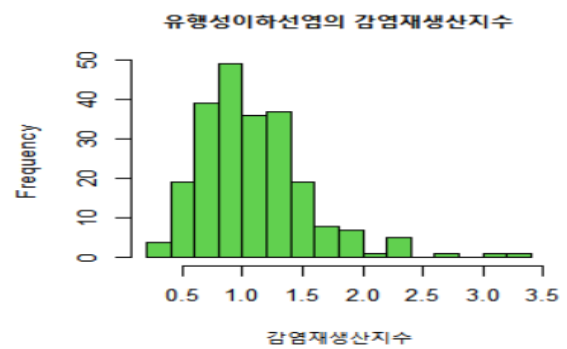
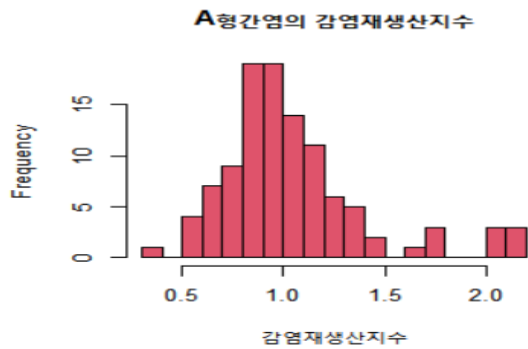


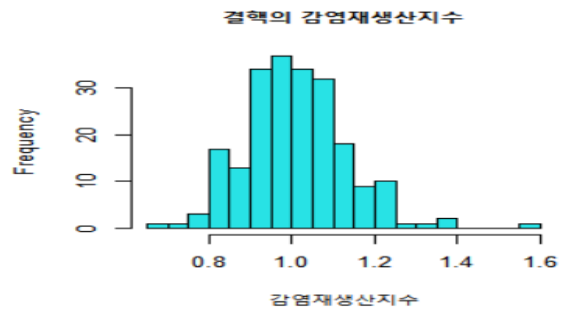
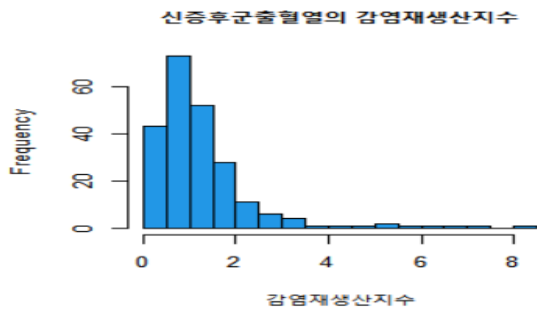
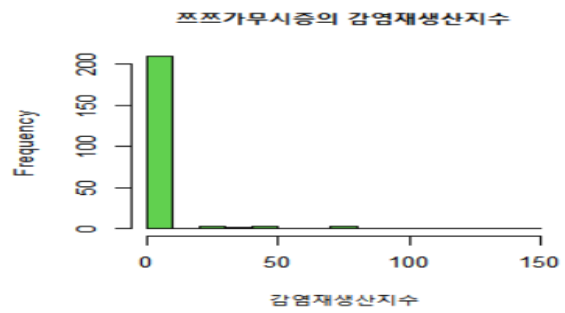
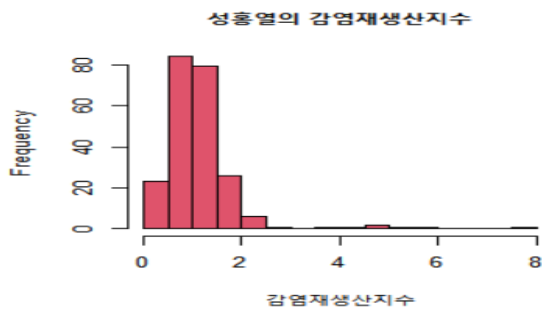
쯔쯔가무시증의 이상치 때문에 자료를 잘 파악할 수 없다고 판단하여 y축을 축소시켰다.

각 질병의 월별 감염재생산지수



각 질병의 감염재생산지수의 median은 1에 가깝다. 즉, 각 질병은 주로 꾸준히 발생하고 있다고 볼 수 있다. 이어서 히스토그램을 통해 더 자세히 분석해 보겠다.

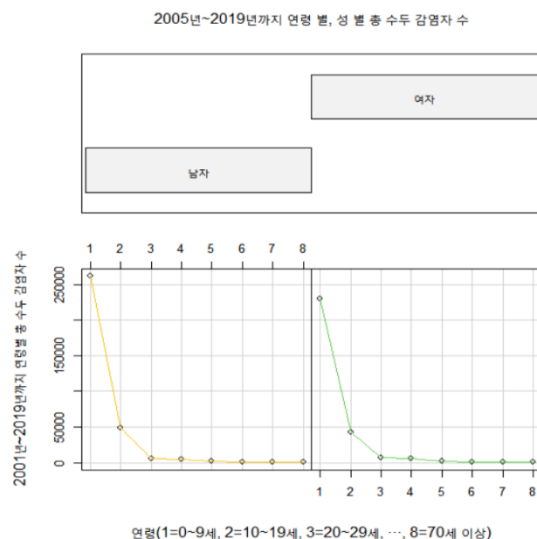
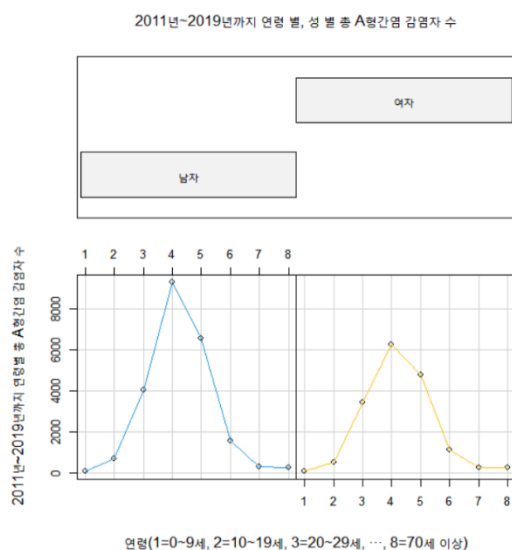


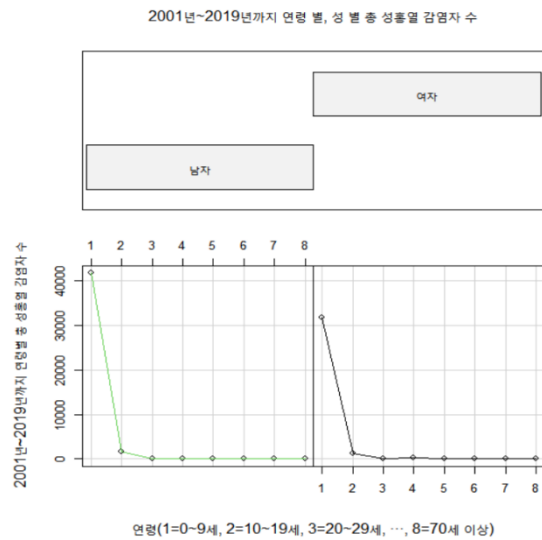
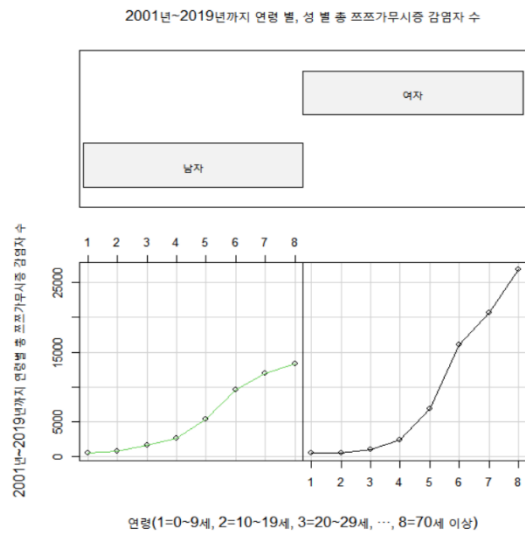


해당 월의 감염재생산지수가 2이상이면 전월 대비 2배이상의 감염자수가 증가한 것이라고 할 수 있기에 그 수치가 높을수록 해당 월에 해당 감염병이 심각(유행)했다고 볼 수 있다. 때문에 위 히스토그램을 보면 수두, 성홍열, 프쯔가무시증, 신증후군출혈열은 유행했을 때가 있었다고 판단된다. 특히 프쯔가무시증은 감염자 수가 매우 급격하게 증가한 시기(크게 유행한 시기)가 있었다는 것을 알 수 있다.

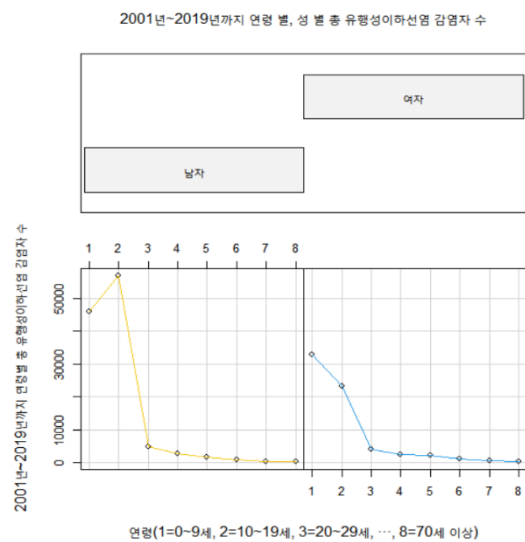
II-3. 성별, 연령별 각 질병의 차이 파악

각 질병에 대해서 성과 연령에 따라 감염자 수의 차이가 있는지 파악하기 위하여 각 질병마다 성, 연령별 coplot을 그려서 분석해 보았다.

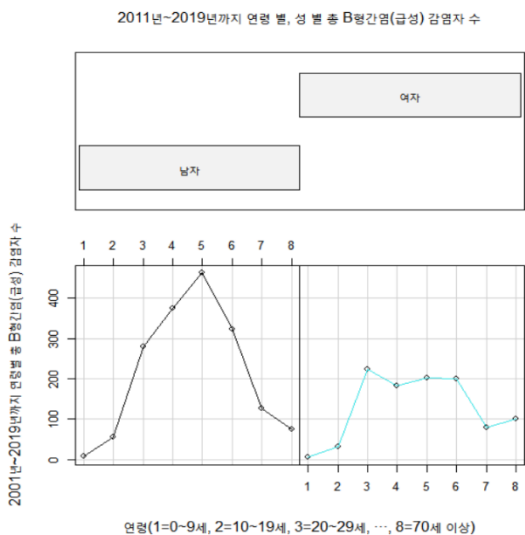




A형 간염, 수두, 성홍열, 쯤쯤가무시증의 감염자 수는 성별에 따라 수에 차이는 있지만 성과 연령 별로 모두 비슷한 추세를 보인다. 또한 수두와 성홍열은 0~9세가 특히 많이 걸리는 질병이며 두 감염병의 그래프의 개형이 매우 유사하다.

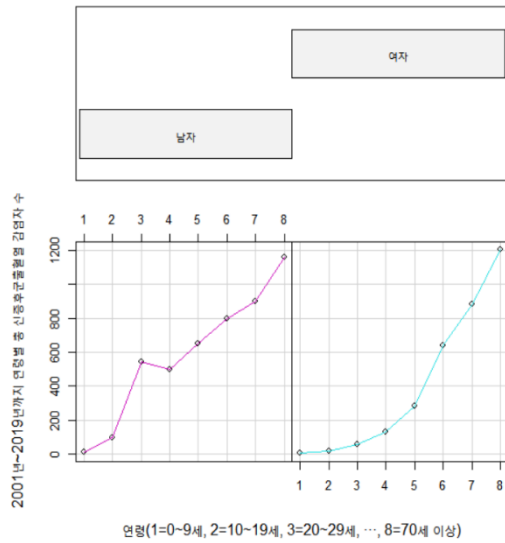


유행성이하선염의 남성 감염자 수는 10대에서 증가하는 추세를 보이지만 여성의 경우는 모든 구간에서 꾸준히 감소하는 추세를 보인다. 또한 0~19세의 남성이 같은 연령대의 여성보다 감염자 수가 훨씬 더 많은 것을 알 수 있다. 따라서 유행성이하선염은 0~19세의 남성이 같은 연령대의 여성보다 더 조심해야 하는 감염병이다.



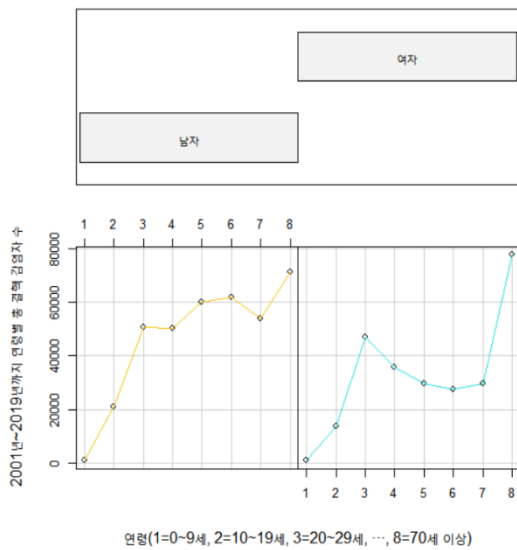
B형 간염의 감염자 수는 20대까지는 남성과 여성 모두 비슷하게 증가하는 추세이지만 30, 40대의 경우 남성은 증가, 여성은 감소하거나 그 전 연령층과 비슷한 추세를 보인다. 이후 구간에서는 다시 비슷한 추세를 보인다. 따라서 B형 간염은 30~40대의 남성이 같은 연령대의 여성보다 더 조심해야 하는 감염병이다.

2001년~2019년까지 연령 별, 성 별 총 신증후군출혈열 감염자 수



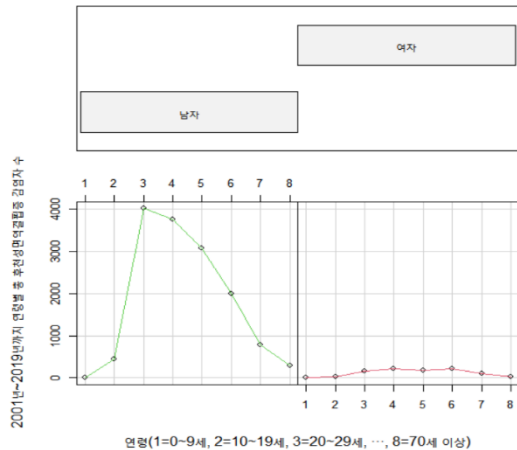
신증후군출혈열의 남성 감염자수는 20대에서 급격히 증가하고, 30대에서 감소한 뒤 40대부터 다시 증가하는 추세를 보이지만, 여성 감염자수는 40대 까지는 원만하게 증가하는 추세를 보이다가 50대부터 급격히 증가하는 추세를 보인다. 따라서 신증후군출혈열은 20대의 남성이 같은 연령대의 여성보다 더 조심해야 하며, 50대 이상부터는 남성과 여성 모두 조심해야 하는 감염병이다.

2001년~2019년까지 연령 별, 성 별 총 결핵 감염자 수



결핵의 감염자 수는 20대까지는 남성과 여성 모두 급격하게 증가하는 추세를 보인다. 그 이후 남성은 같은 수를 유지하거나 증가하다가 60대에서 감소한 후 70세 이상에서 다시 급격한 증가를 보인다. 여성은 20대 이후로 꾸준히 감소하다가 60대에서 약간 증가한 후 70세 이상에서는 급격한 증가를 보인다. 따라서 결핵은 20대 이상 남성과 20대와 70세 이상 여성이 특히 조심해야 하는 감염병이다.

2001년~2019년까지 연령 별, 성 별 총 후천성면역결핍증 감염자 수

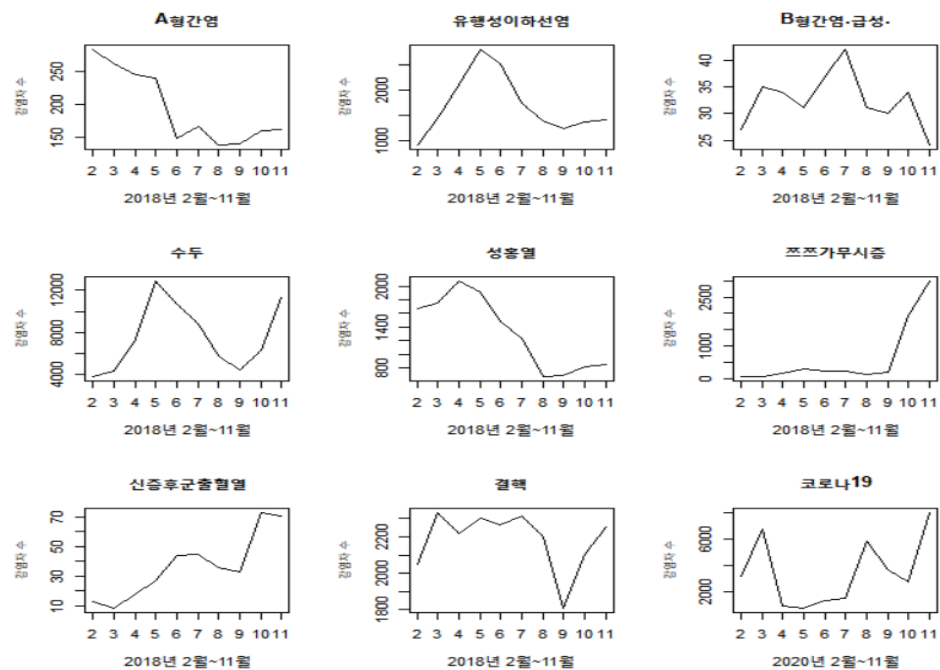


후천성면역결핍증의 감염자 수는 남성이 압도적으로 많다. 특히 20~50대 남성이 같은 연령대의 여성보다 훨씬 많은 것을 볼 수 있다. 남성은 20대로 넘어가면서 매우 급격히 증가하지만 여성은 모든 연령층이 고르게 분포하고 있다. 따라서 후천성면역결핍증은 20~50대 남성이 특히 조심해야 하는 감염병이다.

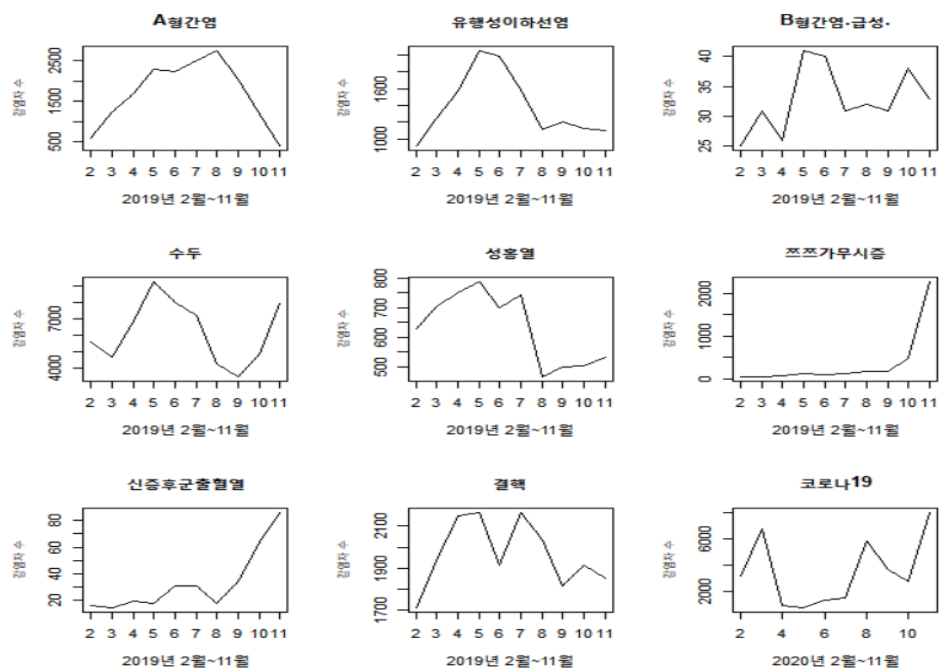
II-4. 코로나19와 감염 추세가 비슷한 감염병이 있는지 확인

현재 전 세계가 코로나19에 관심을 가지고 있을 것이다. 코로나19의 앞으로의 감염 추세를 다른 질병을 통해서 예측해 볼 수도 있다는 생각이 들었다. 2020년 2월~11월까지 코로나19의 월별 감염자 수를 알 수 있는 데이터를 수집하였다. 따라서 다른 감염병 중 최근 2년(수집한 데이터가 2019년까지이기 때문에 2018년, 2019년)동안 코로나19와 비슷한 감염 추세를 보인 것이 있는지 분석하여 보고, 없다면 다른 연도도 분석해 볼 것이다. 아래는 각 감염병과 코로나19의 감염 추세를 나타낸 그래프이다.

<2018년>



<2019년>



2018년과 2019년에는 코로나19와 비슷한 감염 추세를 보이는 감염병은 없다고 판단된다. 나머지 분석 가능한 모든 연도 내에서 같은 방식으로 분석해 보았지만 비슷한 추세는 나타나지 않았다. 현재 수집한 감염병 데이터로는 코로나19의 감염 추세를 예측하는 것은 힘들 것으로 판단된다.

II-5. 각 감염병별 상관관계 파악

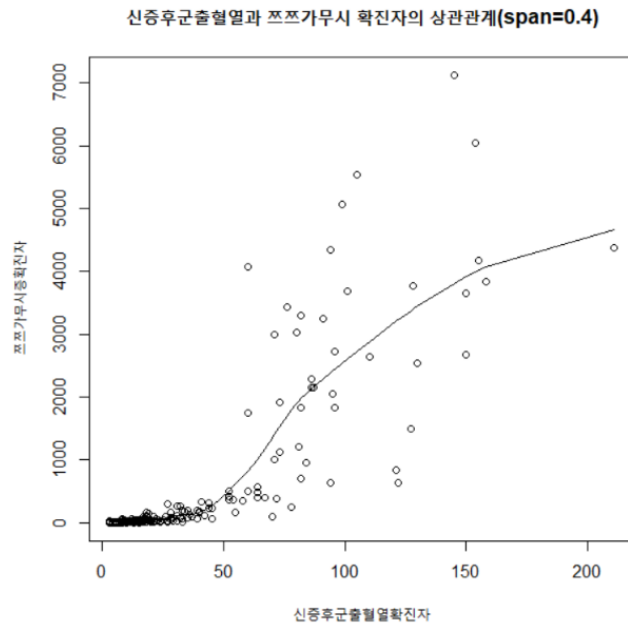
위의 계절별 감염자수 추이를 비교해봤을 때, 두 개의 감염병씩 비슷한 모양새를 보이고 있어, 해당 감염병끼리 관계가 있는지 파악해 보았다.

| | A형간염 | 유행성이하선염 | B형간염, 급성 | 수두 |
|----------|-------------|-------------|-------------|-------------|
| A형간염 | 1.00000000 | -0.09865772 | 0.43322579 | 0.1571681 |
| 유행성이하선염 | -0.09865772 | 1.00000000 | -0.14032607 | 0.1850545 |
| B형간염, 급성 | 0.43322579 | -0.14032607 | 1.00000000 | 0.5917499 |
| 수두 | 0.15716808 | 0.18505452 | 0.59174990 | 1.00000000 |
| 성홍열 | 0.02988086 | 0.04606612 | 0.64661347 | 0.5161427 |
| 쯔쯔가무시증 | -0.14051161 | 0.17866317 | 0.01167142 | 0.1551491 |
| 신증후군출혈열 | -0.12286687 | 0.17981419 | 0.12939783 | 0.2775757 |
| 결핵 | -0.46806278 | 0.29499930 | -0.64102630 | -0.4466204 |
| | 성홍열 | 쯔쯔가무시증 | 신증후군출혈열 | 결핵 |
| A형간염 | 0.02988086 | -0.14051161 | -0.12286687 | -0.46806278 |
| 유행성이하선염 | 0.04606612 | 0.17866317 | 0.17981419 | 0.29499930 |
| B형간염, 급성 | 0.64661347 | 0.01167142 | 0.12939783 | -0.64102630 |
| 수두 | 0.51614271 | 0.15514907 | 0.27757566 | -0.44662045 |
| 성홍열 | 1.00000000 | 0.01169004 | 0.06333032 | -0.32643919 |
| 쯔쯔가무시증 | 0.01169004 | 1.00000000 | 0.86475264 | -0.05679794 |
| 신증후군출혈열 | 0.06333032 | 0.86475264 | 1.00000000 | -0.12648517 |
| 결핵 | -0.32643919 | -0.05679794 | -0.12648517 | 1.00000000 |

좌측의 표는 2013년~2019년 자료를 통해 구한 상관계수를 나타낸 것이다. (2001년부터 하지 않은 이유는 2013년까지 결측값이 있는 감염병이 있어 우선 2013년부터 파악한 뒤, 수치가 높은 감염병들의 자료를 조사했다.)

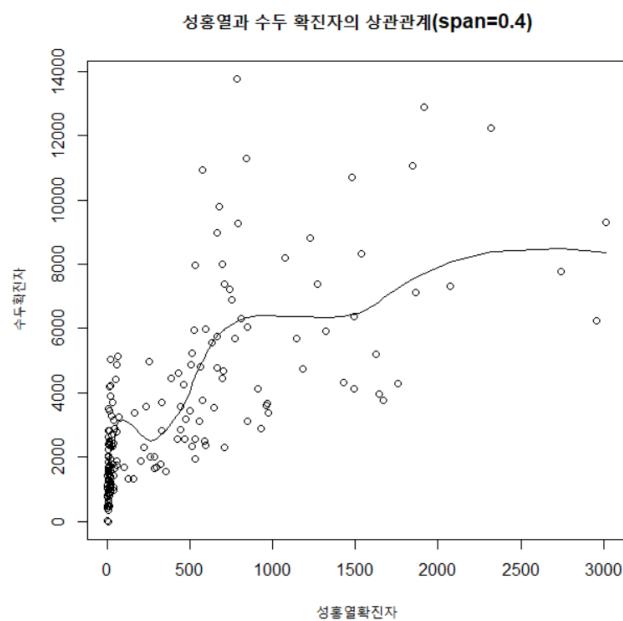
위의 표에서 상관계수가 0.5이상인 것들은 수집한 자료를 최대한 사용하여 상관관계가 있는지 알아보았다. 상관계수가 0.5이상인 질병은 B형 간염(급성)과 수두, B형 간염(급성)과 성홍열, B형 간염(급성)과 결핵, 수두와 성홍열, 쯔쯔가무시증과 신증후군출혈열로 총 5개이다.

| | |
|--|----------------------------------|
| <pre>> cor(bg[,1],bg[,2])</pre> | 첫 번째는 B형 간염과 결핵의 상관계수로 약-0.15였다. |
| <pre>[1] -0.1574361</pre> | |
| <pre>> cor(bs[,1],bs[,2])</pre> | 두 번째는 B형 간염과 수두의 상관계수로 약0.35였다. |
| <pre>[1] 0.3567798</pre> | |
| <pre>> cor(bs_1[,1],bs_1[,2])</pre> | 세 번째는 B형 간염과 성홍열의 상관계수로 약0.29였다 |
| <pre>[1] 0.2949553</pre> | |
| <pre>> cor(ss[,1],ss[,2])</pre> | 네 번째는 수두와 성홍열의 상관계수로 약0.69였다. |
| <pre>[1] 0.6997198</pre> | |
| <pre>> cor(zs[,1],zs[,2])</pre> | 다섯 번째는 쯔쯔가무시증과 신증후군출혈열의 상관계수로 약 |
| <pre>[1] 0.8295355</pre> | 0.82였다. |



좌측의 그래프는 신증후군출혈열의 확진자(x축)와 쯔쯔가무시증 확진자(y축)의 scatter plot에 LOESS를 통해 선을 그린 것이다.

쯔쯔가무시증은 쯔쯔가무시라는 세균을 가지고 있는 진드기의 유충에 물릴 때 감염되고, 신증후군출혈열은 한타바이러스를 가진 쥐의 배설물, 타액 속 바이러스가 공기 중에 건조되어 감염된다. 둘 다 야외활동이 잦고, 날이 건조한 가을에 감염이 많이 되기 때문에 상관계수가 높게 나온 것으로 추측된다.



좌측은 성홍열의 확진자(x축)와 수두의 확진자(y축)의 scatter plot에 LOESS를 통해 선을 그린 것이다.

상관계수가 위의 쯔쯔가무시증과 신증후군출혈열의 상관계수보다 낮다. 이는 전파방식에 영향을 받는 것이 아닌 감염되는 연령대가 영향을 미친다고 추측했다. 성홍열과 수두는 0~9세의 아이들이 많이 감염되고 있었는데, 때문에 신생아가 많이 태어나는 시기에 해당 감염병들이 더 높은 감염율을 나타낼 것이라고 판단된다.

많은 감염병들이 비말을 통해 감염이 되기도 했는데 비말 감염은 많은 감염병들이 가지는 특징으로 체내의 면역력을 제외하고는 모두 외부의 영향을 비슷하게 받아 상관관계에 영향을 주지는 않는 것으로 판단된다. 상관관계를 파악한 결과, 해당 감염병들은 서로가 직접적인 영향을 끼치는 것은 아니다. 다만, 상관관계를 나타나는 원인이 제3의 요인으로 전파방식이 서로 비슷하거나, 감염율이 높은 연령대나 성별이 있는 등의 요인이 존재하는 것으로 파악했다. 또한, 수집한 자료를 통해 제3의 요인에 감염원(RNA, DNA, 세균)이 포함되지 않는다고 파악했다. (모든 감염병에 주요한지는 단정하기 힘들다.)

III. 결론

초기 자료 조사 시 군별로 법정 감염병을 파악했을 때, 굉장히 다양한 감염병이 존재했다. 이 중에서 걸릴 위험이 높을 것 같은 감염병을 위주로 분석하기 위하여 감염자 수가 다른 질병에 비해 많은 감염병을 선택했다. 우리는 현재 코로나19에만 집중하여 다른 감염병에는 신경을 쓰지 못하고 있을 수 있다. 그러나 해당 프로젝트를 진행한 결과, 감염병마다 유행하는 시기, 유행하는 연령대, 성별이 존재하며 코로나19 사태가 진정이 되어도 또다른 팬데믹이 될 수 있는 감염병들도 항상 주위에 있다는 것을 파악했다. 어떤 감염병의 특징이 언제 어떻게 변화할지 모르기 때문에 항상 주의하며 제2의 코로나19 사태를 방지하기 위한 노력이 필요하다.

IV. 참고문헌

A형 간염: <https://terms.naver.com/entry.nhn?docId=926901&cid=51007&categoryId=51007>

유행성이하선염:

<https://terms.naver.com/entry.nhn?docId=5550485&cid=51004&categoryId=51004>

B형 간염: <https://terms.naver.com/entry.nhn?docId=926902&cid=51007&categoryId=51007>

수두: <https://terms.naver.com/entry.nhn?docId=926600&cid=51007&categoryId=51007>

성홍열: <https://terms.naver.com/entry.nhn?docId=5550507&cid=51004&categoryId=51004>

프프가무시증:

<https://terms.naver.com/entry.nhn?docId=926611&cid=51007&categoryId=51007>

신증후군출혈열:

<https://terms.naver.com/entry.nhn?docId=2119899&cid=51004&categoryId=51004>

결핵: <https://terms.naver.com/entry.nhn?docId=926990&cid=51007&categoryId=51007>