**통계 조사론 과제 # 1**

학과 : 전기공학과

학번 : 201724570

이름 : 정석규

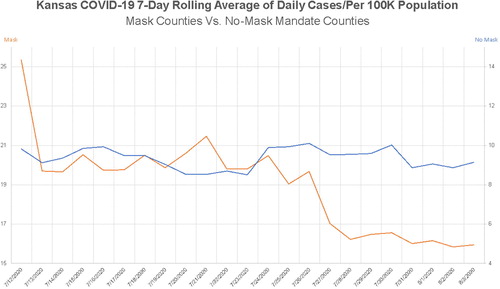
**1. 통계조사와 데이터 시각화에서의 통계오류**

통계조사란 알고자 하는 집단을 구성하고 있는 개인이나 가구, 사업체, 기관 등 개개의 조사단위를 일정한 목적에 따라 관찰함으로써 정보를 객관적이고 체계적으로 얻기 위한 일련의 조사활동을 말한다.

우리가 통계조사를 진행할 때, 다양한 원인에 의해 통계 오류가 발생하게 된다. 이때 발생하는 통계의 오류는 통계 분석 결과의 객관성을 해치는 오류 또는 편향을 의미한다.

데이터 시각화는 데이터 분석결과를 쉽게 이해할 수 있도록 시각적으로 표현하고 전달되는 과정을 말한다. 데이터 시각화의 목적은 도표라는 수단을 통해 정보를 명확하고 효과적으로 전달하는 것이다. 아래의 기사의 예시를 통해 데이터 시각화에서 발생하는 통계조사의 오류를 짚어보고자 한다.

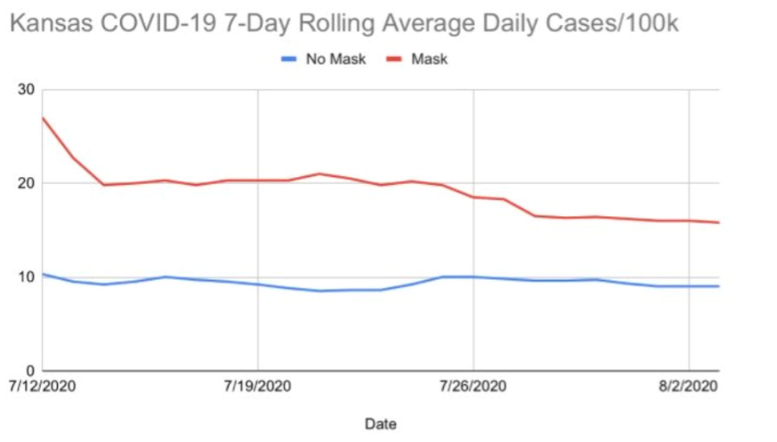
**2. 데이터 시각화에서의 통계조사 오류**



Plot shared by Rachel Maddow on Twitter and live on The Rachel Maddow Show on August 6th, 2020.

위의 차트는 미국 캔자스 주의 마스크 의무화 vs 비의무화 군의 수에 대한 조사 결과에 대한 데이터 시각화이다. 대중들이 보기에 해당 시각화 자료는 특별한 오류가 보이지 않을 수도 있다. 실제로 해당 차트는 매체에 공개되어, 캔자스 주의 마스크 의무화 군의 수가 착용 권고 군의 수에 비해 급감했다는 내용이 보도되었다(Hegeman 2020, August 5, emphasis added).

하지만 차트를 자세하게 확인해 보면, 단일 축이 아닌 두개의 수직 축으로 구성되어 있음을 확인할 수 있다. 실제로 마스크 의무화 군의 수를 표기하는 주황색 선에 대한 축은 왼쪽 축, 마스크 작용 권고 군의 수에 대한 축은 오른쪽 축으로 구성되어 있다. 두개의 수직 축을 개별 적용하여 자료를 해석하면 마스크 의무화 군의 변화는 26개 군에서 16개 군으로, 마스크 착용권고 군의 수10개 군에서 9개 군으로 변화되었다.



위의 차트는 두 개의 축을 하나의 축으로 합쳐 다시 시각화를 한 결과로, 두개의 축을 사용할 경우 왜곡될 가능성이 매우 높음을 확인할 수 있다.

**3. 데이터 시각화 오류 예시에 대한 문제 지적**

앞선 조사의 결과 조사 결과를 서로 다른 y축으로 동시에 표현할 경우 조사 결과에 대한 왜곡가능성이 높음을 확인하였다. 실제로 이러한 유형의 차트를 콤보차트, 이중축차트, 혼합차트라고 부르며, 데이터 값이 개별 항목으로 구분되는 범주형 변수를 하나의 차트에 나타낼 때 사용된다.

하지만 앞의 조사는 두 선그래프가 군의 수라는 동일 도메인으로 존재하기 때문에, 서로 다른 개별 항목으로 구분하여 이중축차트를 사용했다는 것을 문제점으로 지적할 수 있다.

**4. 결론**

데이터 시각화는 도표라는 수단을 통해 정보를 명확하고 효과적으로 전달한다. 하지만 통계적 독해력의 부족이나 기준, 단위 등의 오류가 발생하게 되면 시각화를 하는 과정과 해석하는 일에 있어 오류가 발생하고 왜곡되게 된다. 이는 대중만이 아니라 정확하고 빠른 정보를 전달해야 하는 정부 또한 의도치 않게 결과를 왜곡하는 불상사를 초래할 수 있다. 따라서 통계조사를 진행하는 주체만 아니라 결과를 해석하고 전달하는 매체 또한 통계학에 대한 기본적인 공부가 필요하다고 생각된다.