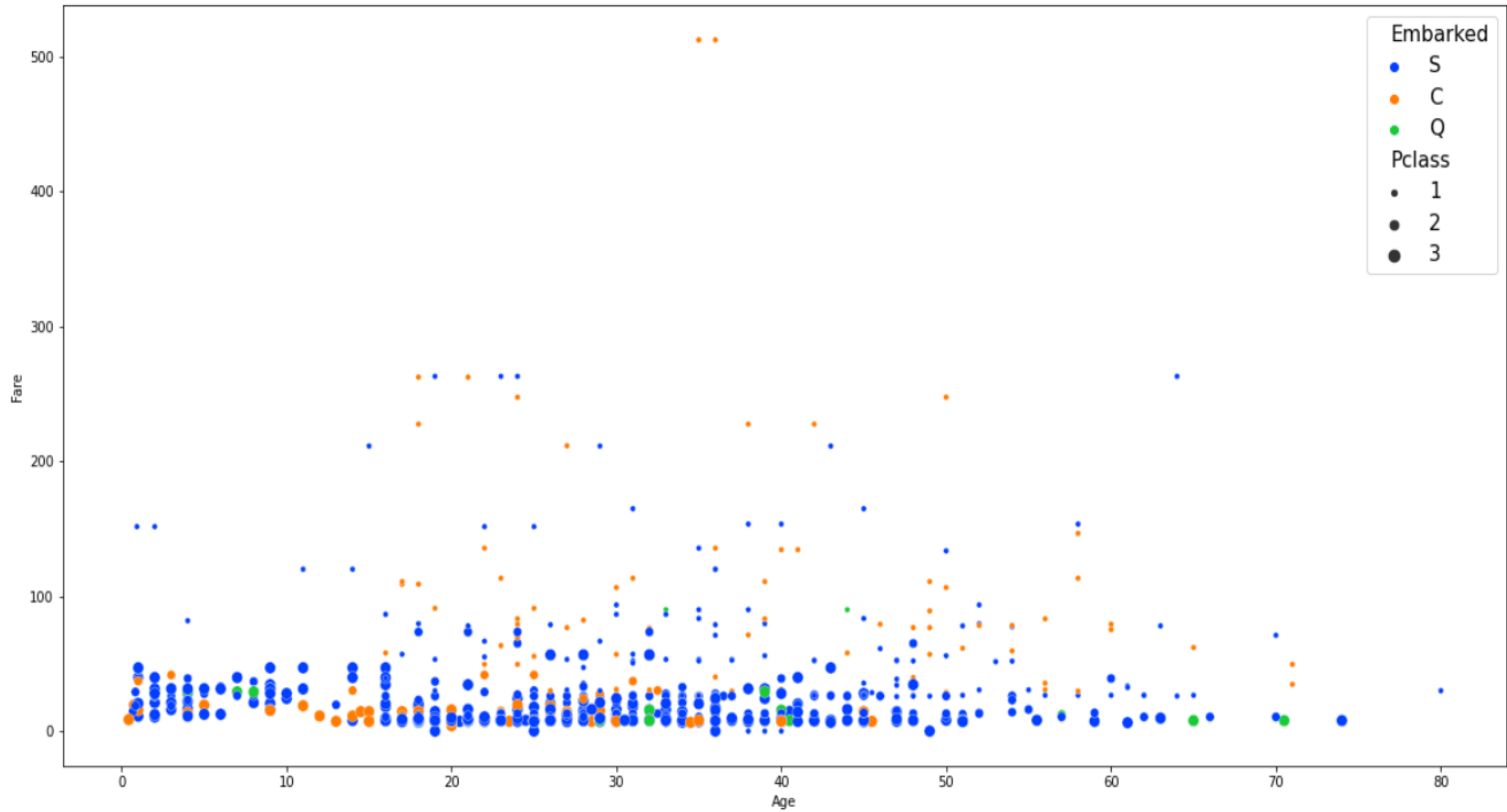
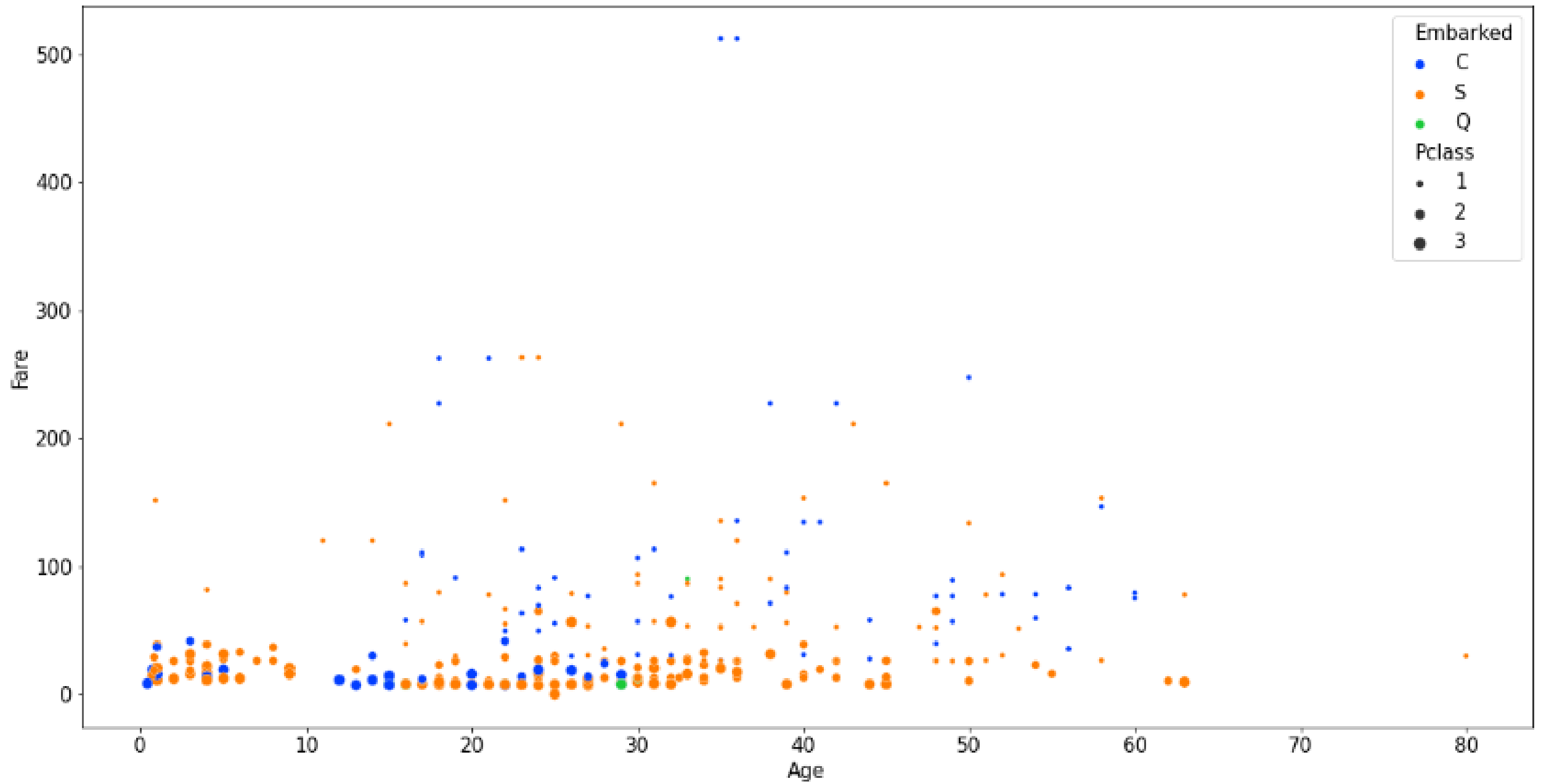


0202 시각화 과제

22기 김수환





- 해당 자료는 titanic dataset의 인원들을 Embarked 별로 분류한 것이며, 상대적으로 주목받지 않았던 'Embarked' 컬럼을 통해 승객을 분류한 것을 한 눈에 알아보기 위해 시각화를 한 자료이다. 해당 자료는 seaborn과 matplotlib을 통하여 시각화를 하였으며, 선정 사유는 해당 자료가 그렇게 size가 크지 않고, colab으로 구현하였기 때문이다. 예상독자는 나와 같이 데이터 분석을 처음 접하는 학생들로 생각하였다.
- Plot은 bubble chart를 선정하였으며, 선정 이유는 2가지가 있다. 첫째, bubble chart는 단순한 2차원 형태의 그래프에 색, 점의 크기 등을 추가 변수로 설정하여 다양한 기준들을 한 눈에 알아볼 수 있다. 이를 통해 찾으려는 Embarked 데이터와 다른 데이터가 어떤 관련성이 있는지 확인해 볼 수 있을 것이라 생각하였다. 둘째는 Embarked가 object 자료형이다 보니 상관계수와 같은 수치를 통한 분석이 불가능했다. 물론 가공을 하면 가능했겠지만, 이 상태 그대로 분석을 해보고 싶어 bubble chart를 선택하게 되었다.
- 첫번째 그래프를 통해 대부분의 사람들이 S(사우스햄프턴)에서 탑승했음을 알 수 있고, 20~40대가 대부분의 탑승객임을 알 수 있다. 또한 0~10세 사이의 아동들은 거의 사우스햄프턴에서 탑승했음 또한 알 수 있다.
- 두번째 그래프는 첫번째 그래프와 달리, 생존자들에 한해 같은 시각화 과정을 거친 그래프이다. 앞선 그래프와의 비교를 통해, 40대 이상의 승객들이 대부분 사망하였으며, Pclass 숫자가 클수록 사망률이 높다는 것을 알 수 있었다.