2022학년도 1학기 기말시험(과제물)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **교과목명** | : | 데이터시각화 |
|  | **학번** | : | 202135-368864 |
|  | **성명** | : | 홍원표 |
|  | **연락처** | : | 010-5343-4341 |

1. 런던 콜레라지도(교재 p.4)를 조사하고 데이터 시각화의 필요성을 설명하시오.(7점)

<https://www.erc.re.kr/webzine/vol33/sub24.jsp>

<http://reportplus.kr/past_press/%ED%95%9C-%EC%9E%A5%EC%9D%98-%EC%A7%80%EB%8F%84-%EC%84%B8%EC%83%81%EC%9D%84-%EB%B0%94%EA%BE%B8%EB%8B%A4_ytn-%ED%95%A8%ED%98%95%EA%B1%B4-%EA%B8%B0%EC%9E%90/>

2. 아름답고 의미 있는 데이터 시각화 사례를 발굴하고 그 이유를 정리하시오.(7점)

3. R 패키지 “vcd”에 내장된 “Arthritis” 데이터셋은 류마티스 관절염 환자를 대상으로 한 임상시험 결과 데이터이다. 각 행은 각 환자를 나타내며, 변수 Treatment는 그룹 (Treated = 새로운 치료제를 투약한 그룹, Placebo = 위약을 받은 그룹)을 나타낸다. 변수 Sex는 성별을, Improved는 치료 결과(None = 차도 없음, Some = 약간 좋아짐, Marked = 매우 좋아짐)를 나타낸다. 새로운 치료제 투약 여부가 치료 결과와 연관이 있는지, 성별과 치료 결과 간에 연관이 있는지를 데이터 시각화를 통해서 탐구하시오. (18점)

4. R에 내장된 “airquality” 데이터셋은 1973년 5월부터 9월까지 뉴욕의 대기질에 관한 데이터셋이다. 변수 Ozone은 대기 중 오존의 양, Solar.R은 태양방사선의 양, Wind는 풍속, Temp는 기온을 나타낸다. 이 네가지 변수(Ozone, Solar.R, Wind, Temp)에 대한 산점도 행렬을 그리고, 이 산점도 행렬에서 알 수 있는 변수들 간의 관계에 대하여 서술하시오. (18점)

<http://www.jisikworld.com/report-view/1168679>