

문제해결시나리오 평가 문제지

훈련과정	[디지털컨버전스] 공공데이터를 융합한 자바스프링&부트 웹개발자 과정 B		
교 과 목	R을 이용한 빅데이터 분석 및 시각화		
능력단위	빅데이터 분석 결과 시각화	출제자	조 성 현 (인)
평가일시	2023.07.26. (5-7 교시)	성 명	
배점기준	각 문제 난이도별 개별 점수	점 수	

※ 아래에 제시된 지문에 따라 데이터 분석을 실행하고 결과를 도출하시오.

1. 아래의 분석절차에 따라 데이터를 분석하고 결과를 도출하시오. (총 45점)

데이터셋(dataset) : ggplot2 패키지에 내장된 “midwest”

1) 데이터를 확인하고 특징을 파악하시오. (10점)

① 데이터의 개요 및 변수들에 대해 설명하시오. (부분점수 5점)

② 데이터프레임 형식으로 데이터를 가져와 특징을 파악하시오. (부분점수 5점)

2) 데이터 변수명을 “poptotal” -> “total” , “popasian” -> “asian” 으로
수정하시오. (5점)

3) 데이터 분석을 위해 위의 두 변수를 이용해 “전체 인구 대비 아시아 인구 백분율” 인
파생변수(“ratio”) 추가하여 히스토그램의 그래프로 만들어 분포를 확인하시오. (10점)

4) “전체 인구 대비 아시아 인구 백분율” 평균을 구하고, 평균보다 크면 “large”,
그 외에는 “small”의 값을 가지는 파생변수(“group”)을 생성하시오. (10점)

5) 파생 변수 “group”의 “large”, “small”에 해당되는 지역의 빈도표를 생성하고,
막대그래프를 만들어 분포를 확인하시오.(10점)

2. 아래 질문에 따라 분석하고 인터랙티브의 시계열 그래프로 시각화 하시오. (총 55점)

데이터셋(dataset) : ggplot2 패키지에 내장된 “economics”

1) 실업자수 변수(“unemploy”)를 시간 순서 속성을 지니는 타입인 “eco”으로 변경하고 날짜와 범위 선택 기능이 포함된 시계열 그래프로 표현하시오. (15점)

2) 두 변수(실업자수“unemploy”, 저축률“psavert”)의 비교가 용이하도록 변수 단위를 조정하여 날짜와 범위 선택 기능이 포함된 여러 값의 시계열 그래프를 표현하시오. (20점)

3) 위의 2)에 생성된 인터랙티브 그래프를 html 파일로 저장하시오. (5점)

4) 위의 3)에 생성된 인터랙티브 그래프를 기반으로 데이터 스토리텔링을 하시오. (15점)

[채점기준표]

데이터 분석			데이터 시각화	데이터 스토리텔링	합계(총점)
데이터 파악	데이터가공	데이터 분석	시각화	데이터 스토리텔링	
10	10	30	35	15	