**비즈니스 모델링 1**

**Assignment 5**

**Due: 5/3, 2020 (일요일), 11:59PM**

**Instructions:** 파일명은 각자의 이름을 사용하여 ‘숙제5\_김일주.docx’등의 이름을 사용하여 저장, 제출하시오. R 코드 부분은 ‘HW5\_김일주.r’로 저장하여 제출하시오. 늦은 제출은 받지 않으니, 제출시한에 맞추도록 유의하시오. 자료의 글을 (거의) 그대로 베끼는 것은 감점요인임. 반드시 자신들의 이해를 충분히 보이는 방법으로, 되도록 자신들의 Words를 사용하여 답안을 작성하시오. R코드의 경우 문제 번호 및 설명을 위해 최소한 한 번 이상의 주석을 사용하시오.

1. **(2.5pts) 다음 각 그래프가 어떠한 용도에 더 유용하게 사용되는지에 대해 간단히 설명하시오. 수업시간의 내용 및 (필요에 따라) 조사한 내용을 이용하여, 자신의 이해를 보이도록 답안을 작성하시오.**
2. **Bar Chart (막대차트): 평균, 개수, 비율과 같은 단일 통계치를 그룹별로 비교하는 데에 유용하게 사용된다.**

**Eg) 집이 강 주변에 있을 때와 없을 때의 평균 집 값을 분석하는 그래프,**

**나이대별 일년 연봉 분석하는 그래프**

1. **Line Graph (선 그래프): 주로 시계열(연속적인 시간의 흐름)을 보여주는 데에 유용하게 사용된다.**

**Eg) 1년동안 매달 손님의 수를 확인하는 그래프,**

**분기별 판매량을 확인하는 그래프**

1. **Scatter Plot (산점도): 결과변수는 y, 예측변수는 x. 두 가지 수치형 변수 간의 정보 중복이나 군집 발견과 같은 연관성을 밝히는 데에 유용하게 사용된다.**

**Eg) 저소득층과 집가격의 중위값의 관계를 나타내는 그래프,**

1. **Histogram (Bar Chart와 구분하여 설명): Box Chart는 그래프 값들 사이에 연속성이 전혀 없지만 Histogram은 연속성이 있다. 연속성이 있는데 그 범위를 나눈 것이다.**

**Eg) 집가격의 중위값과 그 개수를 나타내는 그래프는 중위값의 범위를 나눈 것이지만 강 주변에 집이 있냐 없냐를 확인하는 Bar Chart의 그래프에서는 두 값의 연속성이 전혀 없다.**

1. **Boxplot: 그래프의 분포를 확인할 수 있기 때문에 평균만 볼 때보다 데이터의 특성을 자세히 이해할 수 있다. 수치형 변수의 잠재성 평가를 하는 데에 유용하게 사용된다. 특히 병렬boxplot이 유용하게 사용된다. 병렬로 나열함으로써 평가하는 데에 있어서 더 유용하다.**

**Eg) 집값의 중위값이 30000미만/초과일 때 NOX, LSTAT, PTRATIO, INDUS에서의 데이터 분포를 나타내는 그래프**

1. **(2.5 pts) ApplianceShipments.csv파일에는 1985년부터 1989년 사이의 미국 가전제품 분기별 Shipments (출하량, 단위: 백만 달러) 데이터가 들어있다. 다음의 문제에서 요구한대로 답안을 작성하여 제출할 것. R 코딩에 대한 부분은 R 스크립트 파일에 작성하여 워드파일과 함께 제출할 것. R 스크립트 파일의 이름은 “HW5\_김일주.r”의 형식을 따를 것. 문제 번호 및 설명을 위해 반드시 한 번 이상의 주석을 사용할 것.**
2. **R을 사용하여 시계열 차트를 그리시오.**
3. **R 코드를 통해 y축의 값을 3500~5000범위로 바꾸시오.**
4. **분기별 차이점이나 패턴이 보이는지, 보인다면 어떠한 패턴이 보이는지 설명해 보시오.**

**1분기에서 2분기까지는 대체로 출하량이 증가하는 반면 3분기에서 4분기까지는 출하량이 감소하는 패턴이 보인다. 하지만 2분기에서 3분기로 넘어가는 출하량은 년도에 따라 증가, 감소가 달라지기 때문에 년도에 따른 패턴은 보이지 않는다.**