**HW4 (Due date : July.15 PM 19:59, upload in web)**

**Submit the homework named as HW4\_이름.doc, HW4\_이름.R (코드도 첨부)**

**문제**

1. **상관분석에 대한 설명으로 옳지 않은 것은? 3**
2. **절대값이 0에 가까울수록 상관관계가 없다.**
3. **절대값이 1에 가까울수록 강한 상관성이 있다.**
4. **상관계수 값의 범위는 0~1사이이다.**
5. **상관계수를 통해 두 변수의 선형 관계를 확인할 수 있다.**
6. **회귀분석에서 결정 계수란 무엇인가?(SST와 SSR의 정의 설명 필요)**

전체제곱합(SST)에 대한 회귀제곱합(SSR)의 비율, 즉 모형으로 설명 할 수 있는 부분의 비율이다. SST란 전체 제곱합 이며, SSR은 회귀 제곱합이다.

1. **괄호 안을 채워 넣어라.**

**( )이란 웹페이지, 이메일, 소셜네트워크 기록 등 전자문서 파일로부터 특정 연관성(동시적으로 빈도가 높은 단어 추출)을 분석하는 방법이다.**

**텍스트 마이닝**

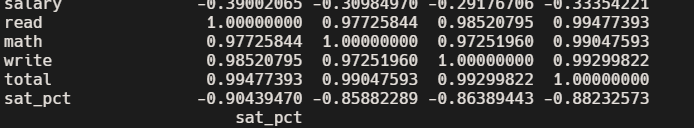
**실습**

**데이터 설명 : SAT\_2010 data (State SAT score from 2010)**

1. **상기 데이터에 대해 분석을 진행하라.**
   1. **해당 데이터에서 중복 데이터가 존재하는가? 존재한다면 해당 변수명을 명시하라.**

**없다.**

* 1. **해당 데이터에서 문자 변수는 총 몇 개가 있는가? 1개**
  2. **결측치가 있는가? 있다면 몇 개가 있는가? 없다**
  3. **변수 read와 변수 math와의 상관계수가 0.9보다 큰 변수를 모두 찾아라.**

****

**Total과 Write**

**데이터설명 : happiness\_2020.csv**

country name: 국가명 (153개 data, 19개의 columhn)

regional indicator: 국가가 속한 지역

ladder score: 표본집단이 응답한 행복지수(-10에 가까울수록 행복하고, 0에 가까울수록 행복하지 않음을 의미)

logged GDP per capita: 1인당 GDP가 행복지수 계산에 기여하는 정도

Social support: 사회적 지원이 행복지수 계산에 기여하는 정도

Healthy life expectancy: 기대 수명이 행복지수계산에 기여하는 정도

Freedom to make life choices: 자유의사결정 정도가 행복지수계산에 기여하는 정도

Generosity: 사회적 관용정도가 행복지수계산에 기여하는 정도

Perceptions of corruption: 부패 인식 지수가 행복지수계산에 기여하는 정도

1. **상기 데이터에 대해 분석을 진행하라.**
   1. **결측치가 있는가? 결측치가 있는지 확인하여라. 있다면 몇 개가 있는가?**

**없다.**

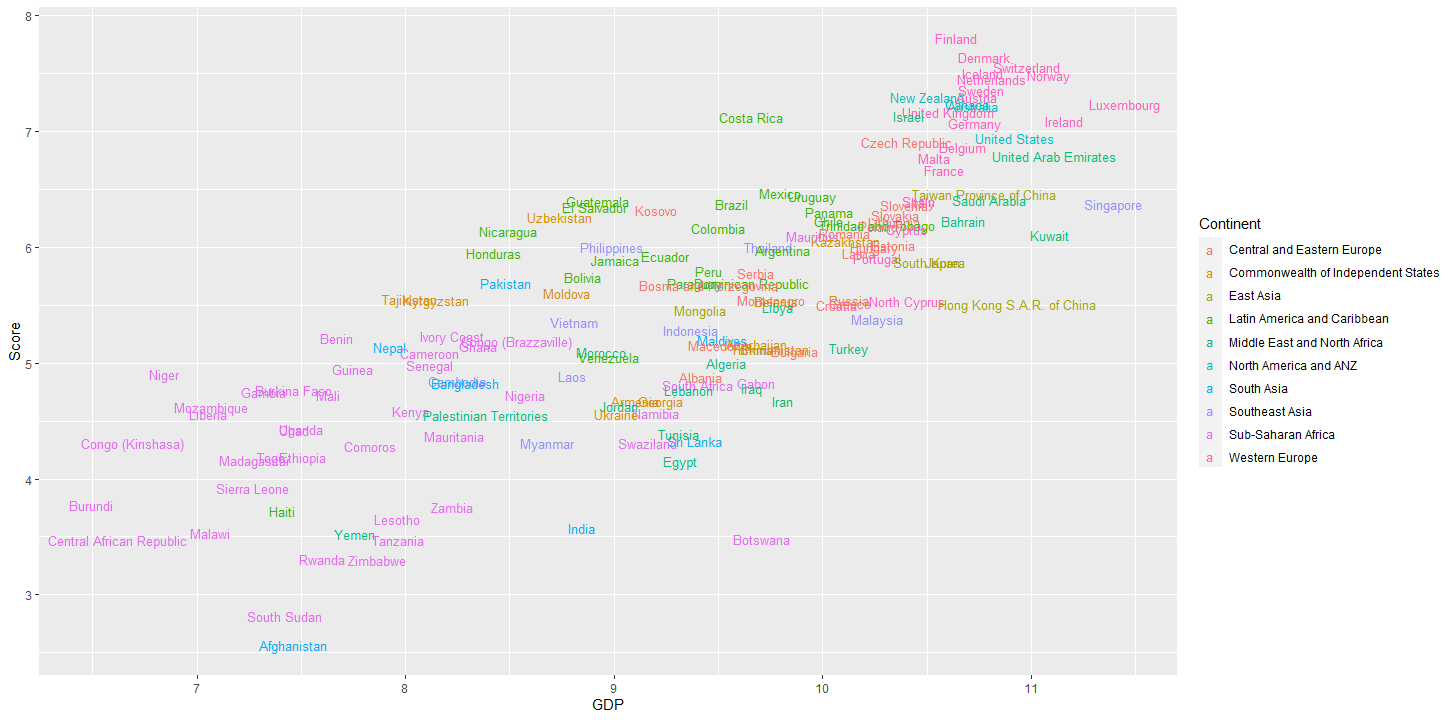
* 1. **행복도(ladder score)에 가장 큰 영향을 미치는 요소 세가지를 상관 분석을 통해서 도출해내라(corrplot 사용 시도 추천). 도출후 각각의 산점도와 상관계수를 명시하라.**
  2. **행복지수가 가장 높은 나라와 낮은 나라는 어디이며 각각의 행복지수는 몇인가?**

가장 낮은 나라는 "Afghanistan"이며, 2.5669이다.

가장 높은 나라는 "Finland"이며, 7.8087이다.

* 1. **아래의 그래프를 해석하라.**

**GDP 기준으로 높은 나라들은 행복지수가 높은 편이다. 핑크색상인 서 유럽이 이런 경향이 강하게 나타난다는 것을 알 수 있다. 반대로 핑크색상인 sub-saharan Africa쪽은 그 반대의 성향을 띈다. 낮은 gdp와 낮은 행복지수가 인상 깊다.**

****

* 1. **A에서 구한 세가지 요소를 이용하여 종속변수를 행복도로 두고 회귀모형을 수행하고 해석하라**

1. **해당 아래 코드를 참고하여 word cloud를 출력하라. (추가 문제)**

filePath <- "http://www.sthda.com/sthda/RDoc/example-files/martin-luther-king-i-have-a-dream-speech.txt"

text <- readLines(filePath)