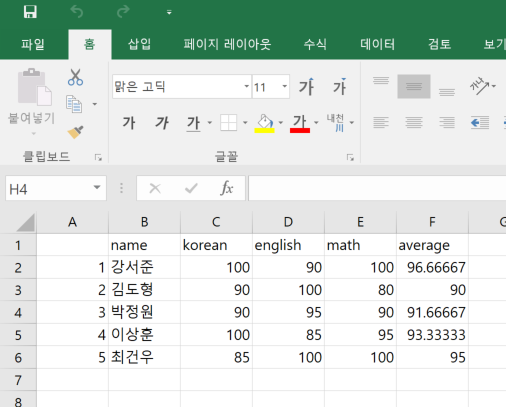
**Mid-term Exam**

April. 20, 2021

**Q1. [20점] gapminder 데이터를 사용해, 아시아 주요 공업 생산국인 대한민국, 중국, 일본을 선택하고, 대표적인 중동의 원유생산국 쿠웨이트, 사우디아라비아, 이란, 이라크를 선택한 후, 하나의 GDP 그래프를 그려 비교하는 R code를 제시하시오.**

**단, x축을 year로, y축은 log10을 취한 GDP로 하는 꺽은선 그래프를 그리고, 국가별 색을 달리하는 범례가 있어야 합니다.**

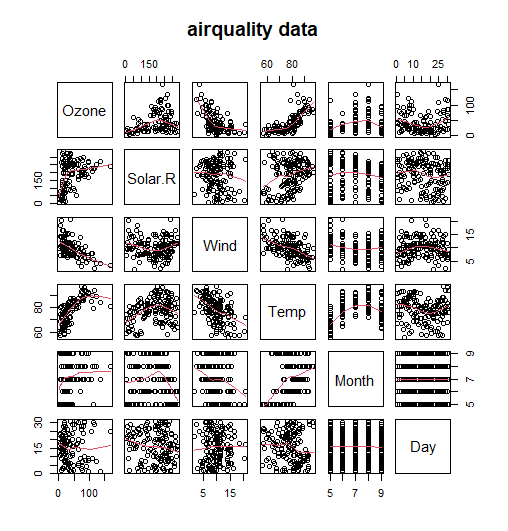
**Q2. [10점] 이미 upload된 데이터 중 하나인 students.txt 파일을 ./Data 폴더로 부터 읽어드린 후, 각 학생의 평균을 구해 오른쪽에 average 열로 추가하여 다음과 같은 output.csv를 ./Data 폴더에 저장하는 R code를 제시하시오.**

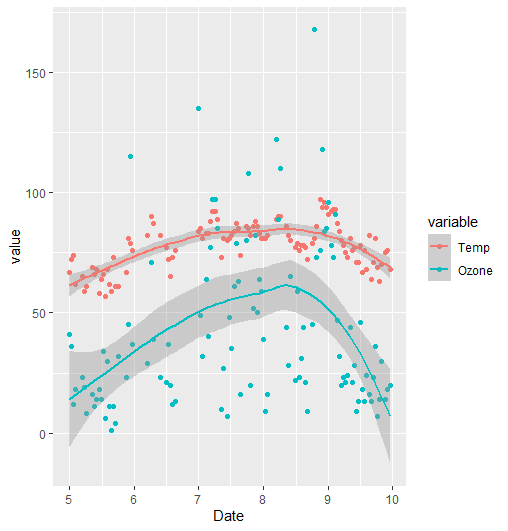


**Q3. [20점] install.packages("hflights")와 library(hflights) 명령어를 사용하여 hflights 데이터를 활용할 수 있게 하십시오.**

1. **[5점] hflights 데이터에는 총 몇 개의 NA가 존재합니까? 이를 찾아내는 R code를 제시하시오.**
2. **[5점] hflights 데이터에서 비행시간이 가장 긴 데이터는 몇 시간 몇 분인지를 알아내는 R code를 제시하십시오.**
3. **[5점] hflights 데이터에서 비행거리가 가장 긴 데이터는 몇 마일인지 알아내는 R code를 제시하십시오.**
4. **[5점] hflights 데이터에서 비행편이 취소된 건수는 몇 건인지 알아내는 R code를 제시하십시오**

**Q4. [40점] 미국 뉴욕의 공기질 데이터를 사용하여 아래와 같은 그래프를 그리는 R code를 제시하십시오. 뉴욕 공기질 데이터는 library(datasets)와 data(airquality) 명령어로 사용할 수 있습니다. (각 20점)**

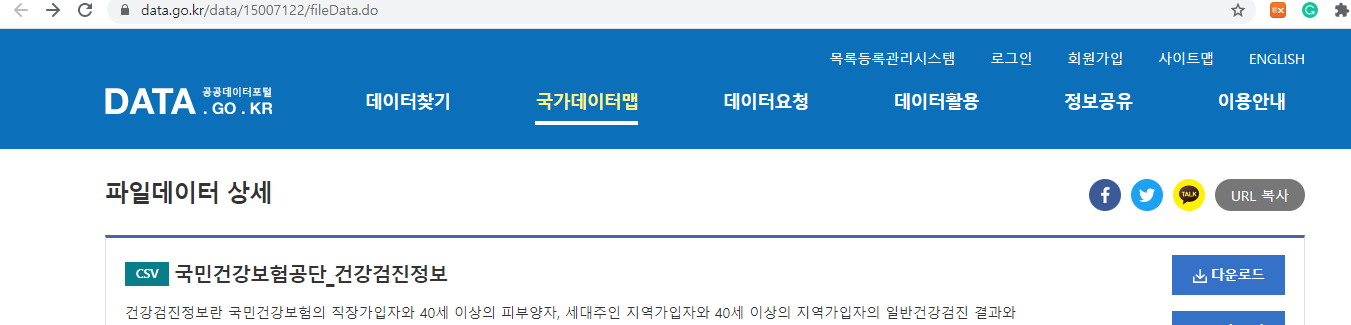




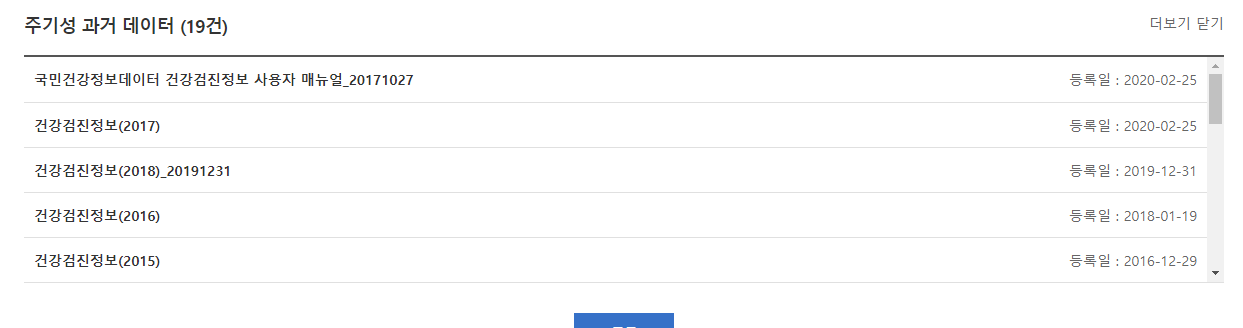
**Extra point 문제: 아래 문제를 모두 맞추면 추가 10점을 드립니다.**

**Q5. [20점] 공공데이터 포털** [**www.data.go.kr**](http://www.data.go.kr)**은 공공기관이 생성, 취득하여 관리하는 공공데이터를 파일, 오픈API, 시각화 등 다양한 방식으로 데이터를 얻을 수 있는 창구입니다.**

**하단은 URL** <https://www.data.go.kr/data/15007122/fileData.do> **의 화면**



**…**



**위 웹페이지에서 2002~2016년의 관측치 데이터가 NHIS\_OPEN\_GJ\_2002\_v1.csv, … NHIS\_OPEN\_GJ\_2016\_v1.csv 이라는 파일명으로 보관되어 있습니다. 데이터를 종합적으로 탐색하려면, 각 연도별 데이터를 download 받고, R로 읽어들여 정제하고 적절히 정리하여 통합하는 가공 과정이 필요합니다.**

1. **[4점] 연도별 데이터 파일을 읽어들여 NHIS2002, … NHIS2016 변수 이름으로 저장하는 R code를 작성하십시오.**
2. **[4점] 다운로드 받을 수 있는 [국민건강정보데이터 건강검진정보 사용자 매뉴얼 20171027] 을 참고하여 결측값을 제거하고 각 열의 데이터형을 알맞게 변환하여 정리하는 R code를 작성하십시오.**
3. **[4점] 2002년 데이터에서 성별, 시도, 연령대 순으로 요약된 통계값을 구하는 R code를 작성하십시오.**
4. **[4점] 2003~2016년 데이터에서 Q5. c 문항과 같은 방법으로 요약된 통계값을 구하는 R code를 작성하십시오.**
5. **[4점] 전체 관측 기간에 대한 요약 통계값을 데이터프레임 변수 하나에 병합하되, 특정 년도의 누락된 검사항목은 결측값으로 표시하는 R code를 작성하십시오.**