**시군구 데이터 불러오기**

%let lib=C:\Json\;

%let String01=NjdhZTg3ZTM1OGEzZGMyOGIyZWE0ZmIxZTBiMDg0ZTg=;/\*apiKey\*/

%let var\_want=PRD\_DE/\*연도\*/ C1 C1\_NM C3\_NM/\*연령\*/ DT/\*수\*/;/\*원하는 변수만 남기기\*/

libname json "&lib";

**%macro** json(data\_final, String02, String03, String04, String05, String06, String07, String08, String09, String10, String11, String12, String13, String14, date\_s, date\_e);

data "&lib.&data\_final";run;

%do date\_want=&date\_s %to &date\_e;

%let url=http://kosis.kr/openapi/Param/statisticsParameterData.do?method=getList&apiKey=&String01&itmId=&String12&objL1=&String04&objL2=&String05&objL3=&String06&objL4=&String07&objL5=&String08&objL6=&String09&objL7=&String10&objL8=&String11&format=json&jsonVD=Y&prdSe=&String14&startPrdDe=&date\_want&endPrdDe=&date\_want&loadGubun=&String13&orgId=&String02&tblId=&String03;

filename out temp;proc http url="&url" method="get" out=out;run;

libname raw json fileref=out;

data "&lib.temp";set raw.alldata;if p1='TBL\_NM' then group+**1**;/\*자동으로 p1의 첫번째 값을 'TBL\_NM'대신 들어가도록 하는 방법을 못 찾겠습니다.\*/run;

proc transpose data="&lib.temp" out="&lib.data\_one"(drop=\_:);by group;id P1;var Value;run;

data "&lib.data\_one";set "&lib.data\_one"(keep=&var\_want);run;

data "&lib.&data\_final";set "&lib.&data\_final" "&lib.data\_one";run;

%end;

data "&lib.&data\_final";set "&lib.&data\_final";if \_n\_=**1** then delete;run;

**%mend**;

%***json***(data\_final=data\_state, /\*orgId\*/String02=**101**, /\*tblId\*/ String03=DT\_1B040M1, /\*objL1\*/String04=**00**+**11**+**21**+**22**+**23**+**24**+**25**+**26**+**29**+**31**+**32**+**33**+**34**+**35**+**36**+**37**+**38**+**39**, /\*objL2\*/String05=**0**, /\*objL3\*/String06=ALL, /\*objL4\*/String07=,/\*objL5\*/String08=, /\*objL6\*/String09=, /\*objL7\*/String10=, /\*objL8\*/String11=, /\*itmId\*/String12=all, /\*loadGubun\*/String13=**2**, /\*prdSe\*/String14=Y, date\_s=**1993**, date\_e=**2017**);

%***json***(data\_final=data\_city, /\*orgId\*/String02=**101**, /\*tblId\*/ String03=DT\_1B040M1, /\*objL1\*/String04=all, /\*objL2\*/String05=**0**, /\*objL3\*/String06=ALL, /\*objL4\*/String07=,/\*objL5\*/String08=, /\*objL6\*/String09=, /\*objL7\*/String10=, /\*objL8\*/String11=, /\*itmId\*/String12=all, /\*loadGubun\*/String13=**2**, /\*prdSe\*/String14=Y, date\_s=**1993**, date\_e=**2017**);

/\*변수명, 변수 형태 변경\_start\*/

**data** json.data\_city;

set json.data\_city;

Year=PRD\_DE+**0**;/\*연도\*/

Div\_num=C1;

Div=C1\_NM;

Old=C3\_NM;

Num=DT+**0**;/\*수\*/

keep Year Div\_num Div Old Num;

**run**;

/\*변수명, 변수 형태 변경\_end\*/

**시군구 데이터 합치기**

**data** w1\_raw;

set json.data\_city;

if length(div\_num)<**3** then delete;

**run**;

**data** compare\_all;**run**;

**%macro** ***homework03***;

%do year=**1993** %to **2017**;

proc sql;

create table w1 as

select \*

from w1\_raw where year=&year;

quit;

run;

proc sql;

create table w2 as

select \*

from json.data\_state where year=&year;

quit;

run;

proc sql;

create table w1\_01 as

select \*, substr(div\_num, **1**, **2**) as div\_state, sum(num) as sum

from w1 where substr(div\_num, **5**,**1**)="0"

group by div\_state, old;

quit;

run;

proc sort data = w1\_01 out=w1\_02 nodupkey;

by div\_state old;

run;

proc sql;

create table compare as

select a.year, a.old, a.num, a.div\_state, a.sum, b.year as year\_original, b.old as old\_original, b.div as div\_original, b.div\_num as div\_num\_original, b.num as sum\_original

from w1\_02 as a left join w2 as b

on a.div\_state=b.div\_num and a.old=b.old and a.year=b.year;

quit;

run;

proc sql;

create table adjust\_01 as

select \*,

sum(case when (select count(\*) from compare where old="100세 이상")^=**0** then **1** else **0** end) as up100,

sum(case when (select count(\*) from compare where old="85세 이상")^=**0** then **1** else **0** end) as up85

from compare;

quit;

data adjust\_02;

set adjust\_01;

if up85>**0** and old="80세 이상" then delete;

if up100>**0** and old="80세 이상" then delete;

if up100>**0** and old="85세 이상" then delete;

run;

data compare\_all;

set compare\_all adjust\_02;

run;

%end;

**%mend** homework03;

%***homework03***

**data** compare\_all;set compare\_all;if \_n\_<**7** then delete;**run**;

**data** compare\_all;

set compare\_all;

if sum=sum\_original then compare=**0**; else compare=**1**;

**run**;

**data** semi\_final;

set compare\_all;

if old^='계' then old\_num=put(input(scan(old, **1**, '세'),best.),Z3.);

**run**;

**proc** **sort** data=semi\_final;

by year div\_state old\_num;

**run**;

**data** final;

set semi\_final;

drop up100 up85 old\_num;

**run**;

**proc** **means** data=final;

var compare;

**run**;

**결론**

불러온 시군구 데이터를 시도 및 전국 자료를 다 없애고, 지역코드 앞부분 두자리 수로 묶은 후, 이를 불러온 시도 데이터와 비교했습니다. 이때, 시군구 데이터를 통해 만들어진 시도 데이터(w1\_02)는 시군구 자료가 없는 년도의 자료(1993년부터 1997년)는 생기지 않았습니다. 이 경우 시도 데이터와 비교할 필요가 없기에 시군구 데이터를 통해 만들어진 데이터에 시도 데이터를 left join(compare)하였습니다. Left join한 데이터(compare)를 기반으로 100세 이상이 있거나 85세 이상이 있는 년도에는 80세 이상은 지우고, 100세 이상이 있는 경우에는 85세 이상을 지워서, 중복되는 ~세 이상 값을 처리했습니다.

처리한 데이터들을 연도별로 전부 묶어서 시군구 데이터를 통해 만들어진 자료와 시도 데이터가 일치하는지 확인하기위해 두 자료의 나이별 인구수가 같을 경우 compare 변수에 0을, 다를 경우 compare변수에 1을 할당하였습니다. 이후 연도, 지역, 나이에 따라 정렬하여 마지막으로 Proc means를 통해 compare 변수를 확인한 결과 합이 0으로 두 자료가 모두 같은 값을 가짐을 확인하였습니다.