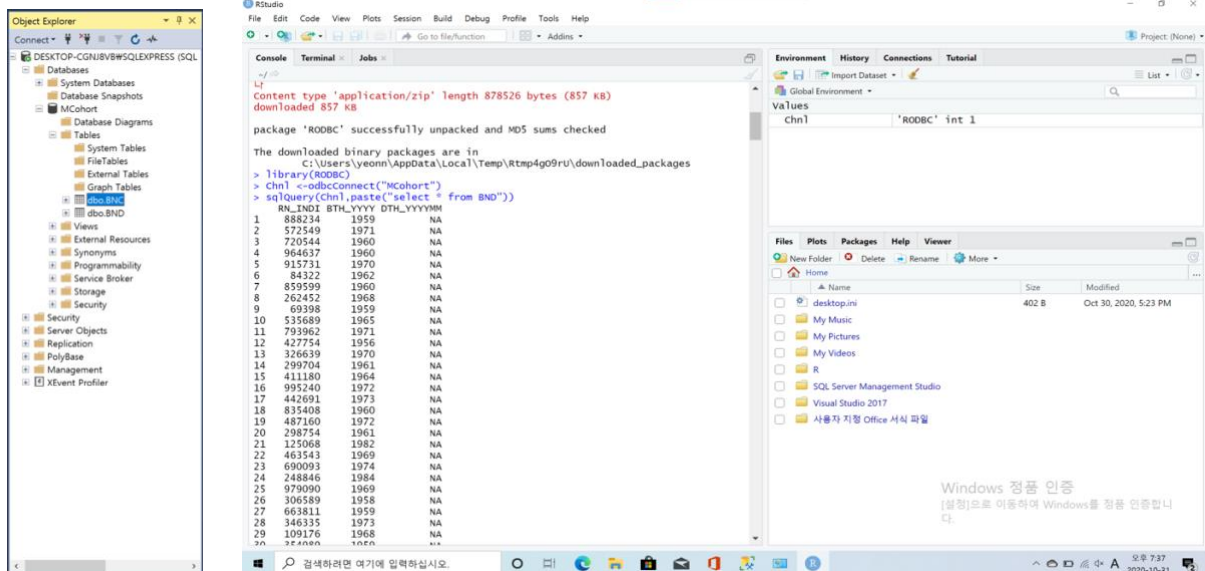


## 데이터문제해결1 과제2

컴퓨터공학과 17011599 안정연

먼저, SQL Server 에서 코호트 파일 DB에 탑재한 후 R에서 ODBC를 통해 불러온다.



### 1. 1000명 중 J209를 한 번이라도 앓았던 사람은 몇 명인가?

- M40 데이터 중 J209 기록이 있는 데이터만 table1 에 저장
- BNC 데이터와 table1 데이터를 join 한 다음 table2에 저장
- 중복제거

```
> sqlQuery(Chn1, paste("select * into table1 from M40 where MCEX_SICK_SYM = 'J209'"))
character(0)
```

```
> sqlQuery(Chn1, paste("select t1.RN_INDI, t1.SEX, t2.MCEX_SICK_SYM from BNC t1 join
table1 t2 on t1.RN_INDI = t2.RN_INDI"))
```

```
> sqlQuery(Chn1, paste("select distinct * from table2"))
```

	RN_INDI	SEX	MCEX_SICK_SYM
1	537873	1	J209
2	483693	2	J209
3	900980	2	J209
4	674034	1	J209
5	440295	1	J209

```
> sqlQuery(Chn1, paste("select distinct * into table3 from table2"))
character(0)
```

```
> sqlQuery(Chn1, paste("select count(*) from table3"))
```

1 843

- 1000명 중 843명이 J209를 한 번이라도 앓았음을 알 수 있다.

2. 해당 사람들의 성별 구성 숫자와 비율을 구하시오.

- Table2에서 SEX=1인 경우의 컬럼수를 구한다.
- Table2에서 SEX=2인 경우의 컬럼수를 구한다.

```
> sqlQuery(Chnl,paste("select count(*) from table3 where SEX=1"))
```

```
1 400
```

```
> sqlQuery(Chnl,paste("select count(*) from table3 where SEX=2"))
```

```
1 443
```

- 여자는 443명 , 남자는 400 명이 J209를 앓았다.

3. 해당 사람들의 성별 그룹별 평균 출생년도를 구하시오.

- BND 데이터와 table2 를 join
- Table3에서 BTH\_YYYY 가 문자형이므로 숫자형으로 변환
- 성별 그룹별 평균 출생년도 출력

```
> sqlQuery(Chnl,paste("select t1.RN_INDI,t1.BTH_YYYY, t2.SEX, t2.MCEX_SICK_SYM into table4 from BND t1 join table3 t2 on t1.RN_INDI=t2.RN_INDI"))
```

```
> sqlQuery(Chnl,paste("select RN_INDI,SEX,LEFT(BTH_YYYY,4) as BTH into table5 from table4"))
```

```
> sqlQuery(Chnl,paste("select RN_INDI,SEX,convert(INT,BTH) as BTH into table6 from table5"))
```

```
> sqlQuery(Chnl,paste("select SEX, avg(BTH) as AVG_BTH from table6 group by SEX"))
```

```
SEX AVG_BTH
1 2 1973
2 1 1976
```

- 남자는 평균 1976년, 여자는 평균1973년에 태어났다.