

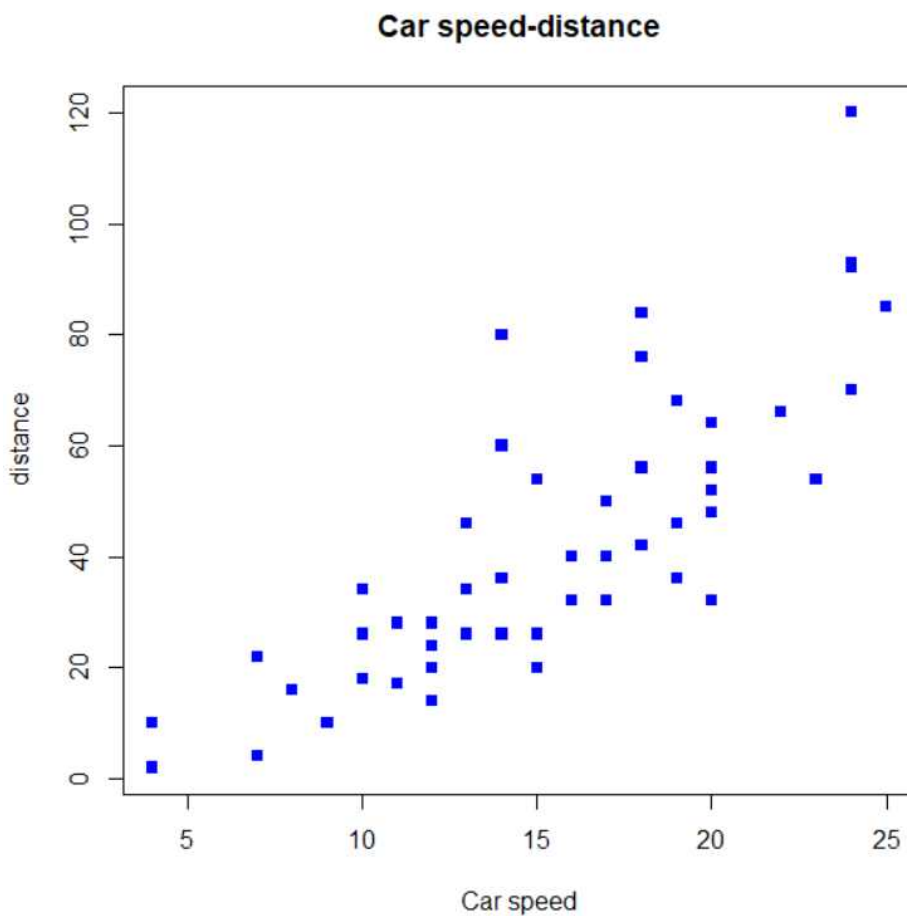
# 데이터베이스 기초

## 16장 과제

## <16장\_연습1>

1. R에서 제공하는 cars 데이터셋을 이용해서 speed 와 dist 에 대한 산점도를 그리시오 (x축이 speed). speed 와 dist (제동거리)에 대한 상관관계를 설명해 보시오

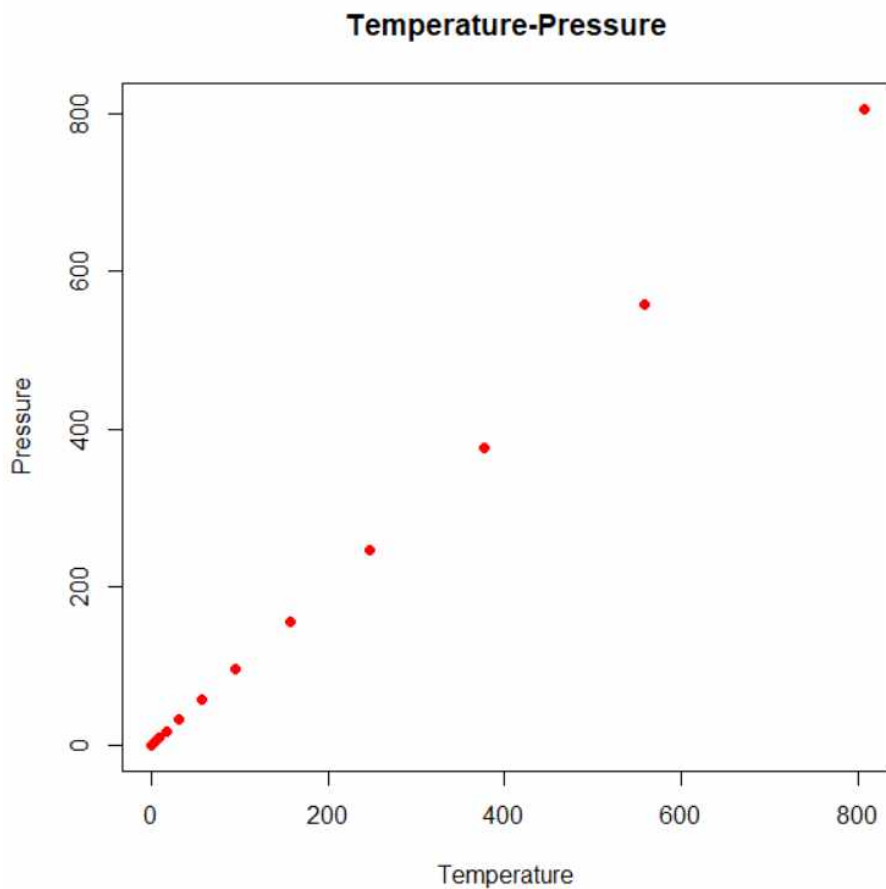
```
speed <- cars$speed
dist <- cars$dist
plot(speed, dist,
      main = "Car speed-distance",
      xlab = "Car speed",
      ylab = "distance",
      col = "blue",
      pch = 15)
```



속도가 빠를수록 멈출 때까지의 제동거리가 길다는 것을 확인할 수 있다.

2. R에서 제공하는 pressure 데이터셋을 이용해서 temperature 와 pressure 에 대한 산점도를 그리시오 (x축이 temperature). 두 변수간 상관관계를 설명해 보시오

```
temperature <- pressure$temperature
pressure <- pressure$pressure
plot(temperature, pressure,
     main = "Temperature-Pressure",
     xlab = "Temperature",
     ylab = "Pressure",
     col = "red",
     pch = 19)
```



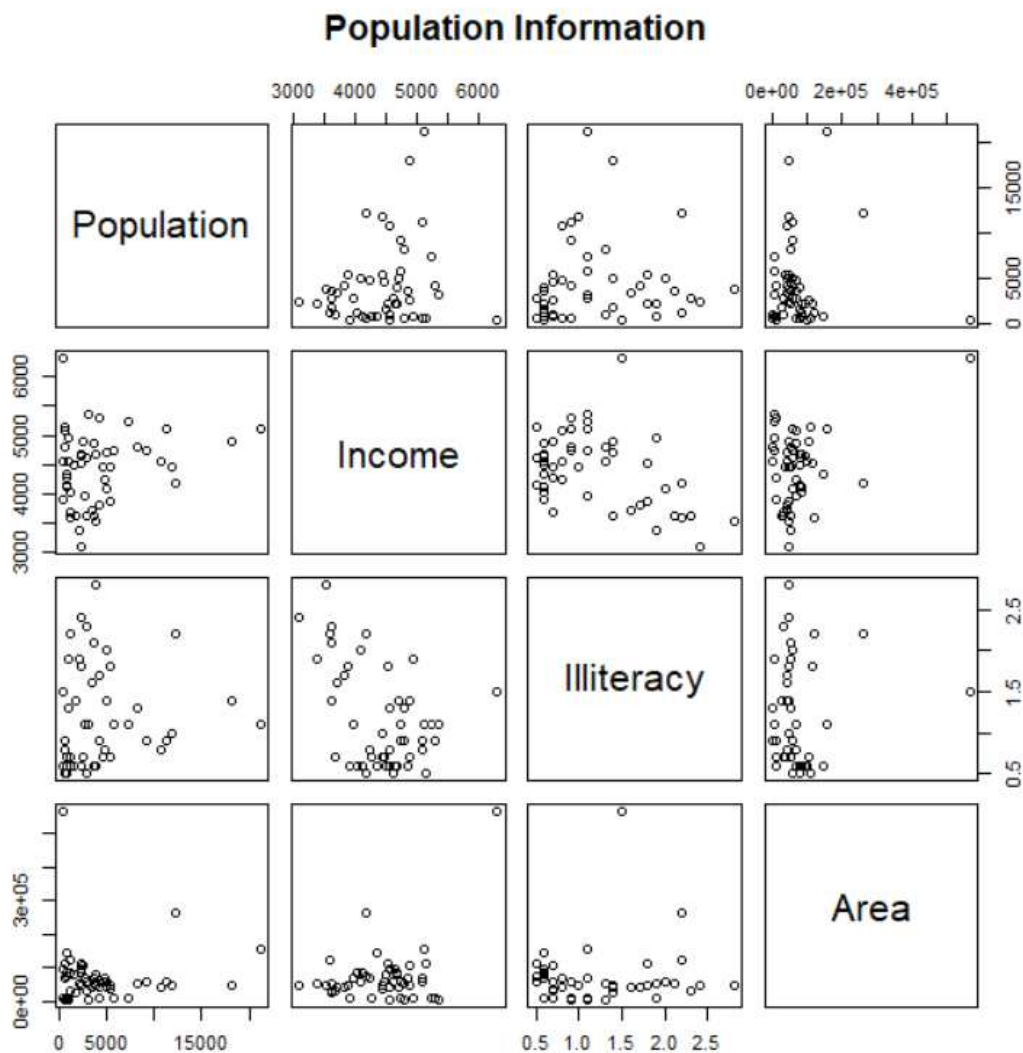
온도가 높아질수록 압력이 높아진다는 것이 완벽히 일치하는 것을 알 수 있다.

3. R에서 제공하는 state.x77 데이터셋에서 Population, Income, Illiteracy, Area 변수간산점도를 그려 상관관계를 관찰하시오 (pairs() 함수 이용)

```
country <- state.x77
```

```
target <- country[, c("Population", "Income", "Illiteracy", "Area")]
```

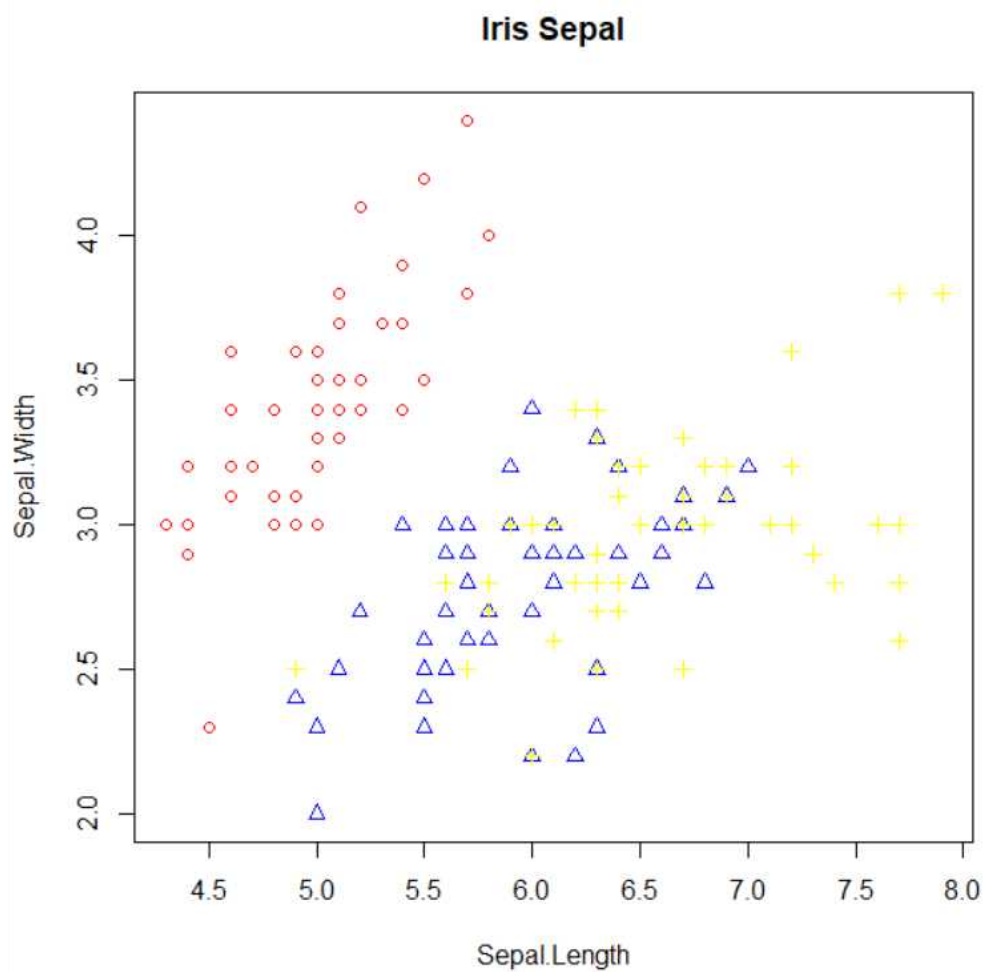
```
pairs(target, main = "Population Information")
```



전체적으로 높은 상관 관계를 보이는 변수의 조합은 관찰이 되지 않으나 Illiteracy 가 증가 하면 income 이 감소하는 현상이 보임

4. iris 데이터셋에서 Species 정보에 따른 Sepal.Length, Sepal.Width (꽃받침의 길이, 폭)의 분포를 알아보시오

```
iris.2 <- iris[, 1:2]  
point <- as.numeric(iris$Species)  
plot(iris.2,  
      main = "Iris Sepal",  
      pch = c(point),  
      col = color[point])
```



<16장\_연습문제2>

1. 다음은 10명의 수입과 교육받은 기간을 조사한 표이다. 수입과 교육기간 사이에 어느 정도 상관관계가 있는지 조사하시오 (산점도, 상관계수 구하기)

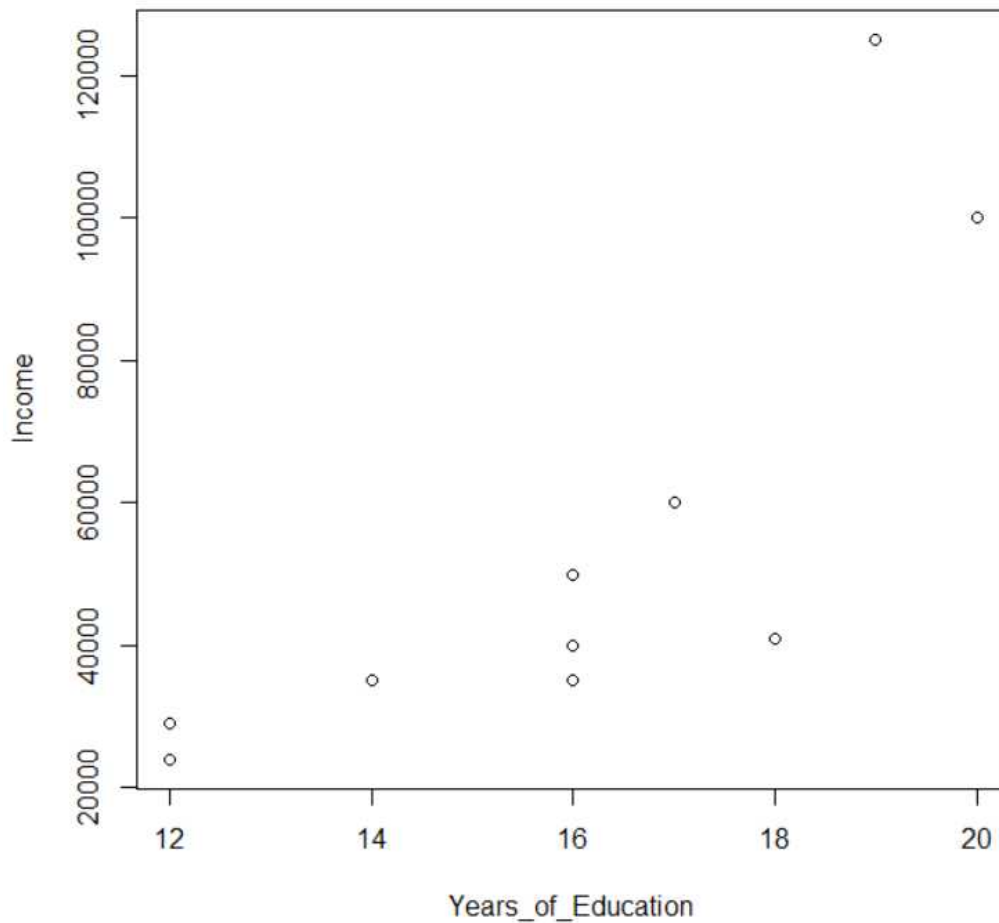
```
Income <- c(125000, 100000, 40000, 35000, 41000, 29000, 35000, 24000, 50000, 60000)
```

```
Years_of_Education <- c(19, 20, 16, 16, 18, 12, 14, 12, 16, 17)
```

```
table <- data.frame(Income, Years_of_Education)
```

```
plot(Income~Years_of_Education, data = table)
```

```
cor(Income, Years_of_Education)      상관계수 : 0.7887259
```



2. 다음은 학생 10명의 성적과 TV 시청시간을 조사한 표이다. 성적과 TV시청시간 사이의 상관관계를 조사하시오. (산점도, 상관계수 구하기)

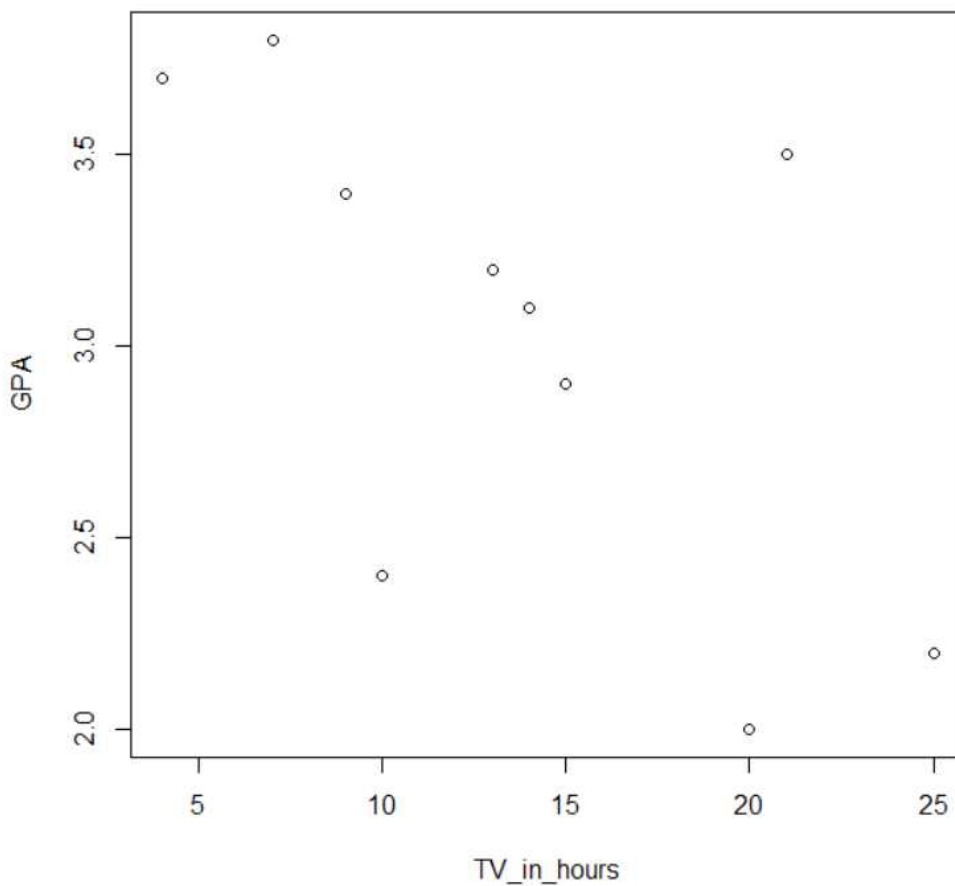
```
GPA <- c(3.1, 2.4, 2.0, 3.8, 2.2, 3.4, 2.9, 3.2, 3.7, 3.5)
```

```
TV_in_hours <- c(14, 10, 20, 7, 25, 9, 15, 13, 4, 21)
```

```
tbl <- data.frame(GPA, TV_in_hours)
```

```
plot(GPA~TV_in_hours, data = tbl)
```

```
cor(GPA, TV_in_hours)      상관계수 : -0.6283671
```



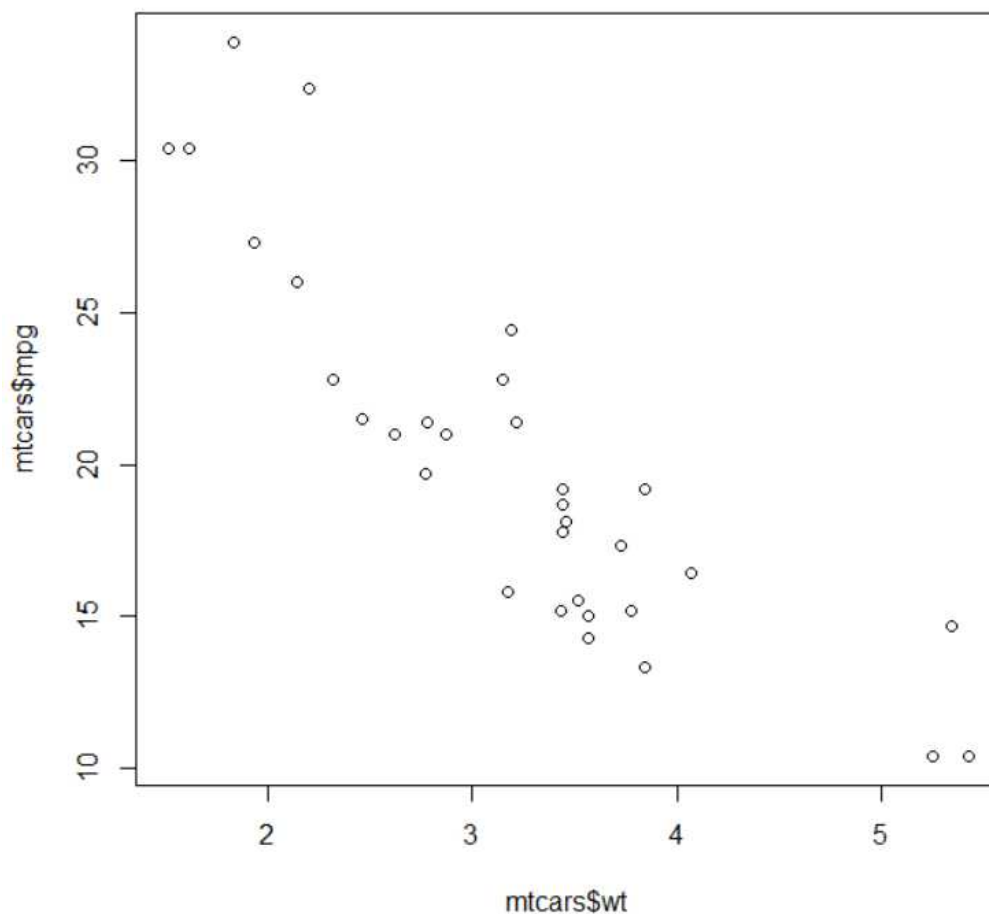
3. R에서 제공하는 mtcars 데이터셋에서 mpg 와 다른 변수들 간의 상관 계수를 구하시오. 어느 변수가 mpg 와 가장 상관성 이 높은지 산점도와 함께 제시하시오.

```
cor(mtcars[, 1:11])["mpg",]
```

```
> cor(mtcars[, 1:11])["mpg",]  
      mpg      cyl      disp      hp      drat      wt  
1.0000000 -0.8521620 -0.8475514 -0.7761684  0.6811719 -0.8676594  
      qsec      vs      am      gear      carb  
0.4186840  0.6640389  0.5998324  0.4802848 -0.5509251
```

wt 가 가장 높음

```
plot(mtcars$mpg~mtcars$wt)
```

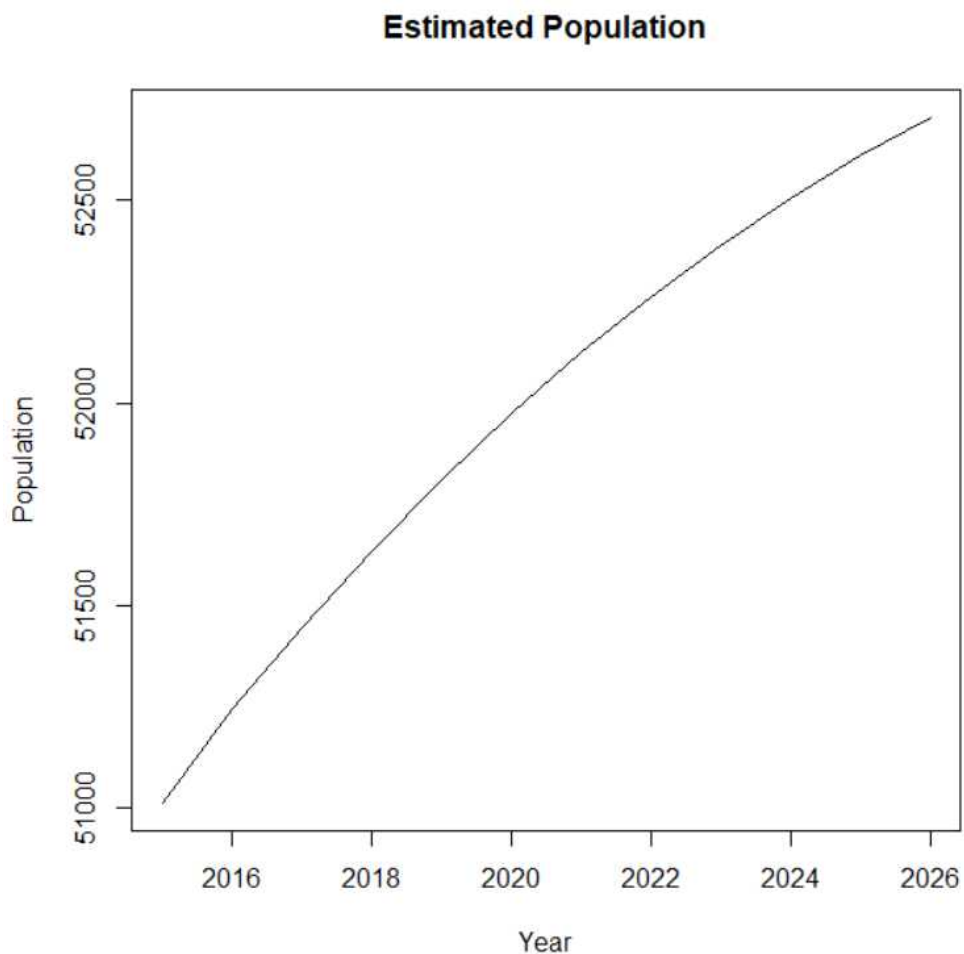




### <16장\_연습3>

1. 다음은 2015년부터 2026년도까지의 예상 인구수 추계자료이다. 선그래프를 작성하시오.

```
year <- 2015:2026
population <- c(51014, 51245, 51446, 51635, 51811, 51973, 52123, 52261,
52388, 52504, 52609, 52704)
plot(year, population,
      main = "Estimated Population",
      type = "l",
      lty = 1,
      lwd = 1,
      xlab = "Year",
      ylab = "Population")
```

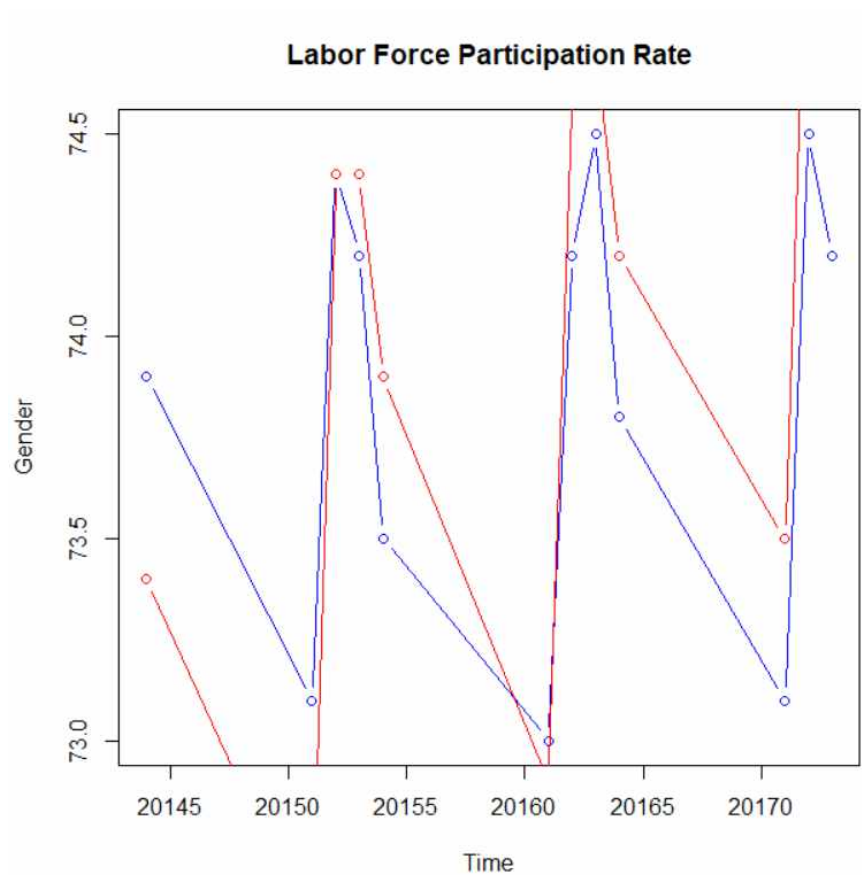


2. 다음은 2014년 4분기부터 2017년 3분기까지 남,녀의 경제활동참가율 통계이다. 선그래프를 작성하시오 (남,녀를 각각 다른 선으로 표시) (2014년 4분기는 20144, 2015년1분기는 20151 과 같이 입력한다)

```
time <- c(20144, 20151:20154, 20161:20164, 20171:20173)
men <- c(73.9, 73.1, 74.4, 74.2, 73.5, 73.0, 74.2, 74.5, 73.8, 73.1, 74.5,
74.2)
women <- c(51.4, 50.5, 52.4, 52.4, 51.9, 50.9, 52.6, 52.7, 52.2, 51.5, 53.2,
53.1)
women <- (women + 22)
```

```
plot(time, men,
      main = "Labor Force Participation Rate",
      type = "b",
      lty = 1,
      col = "blue",
      xlab = "Time",
      ylab = "Gender")
```

```
lines(time, women,
      type = "b",
      col = "red")
```

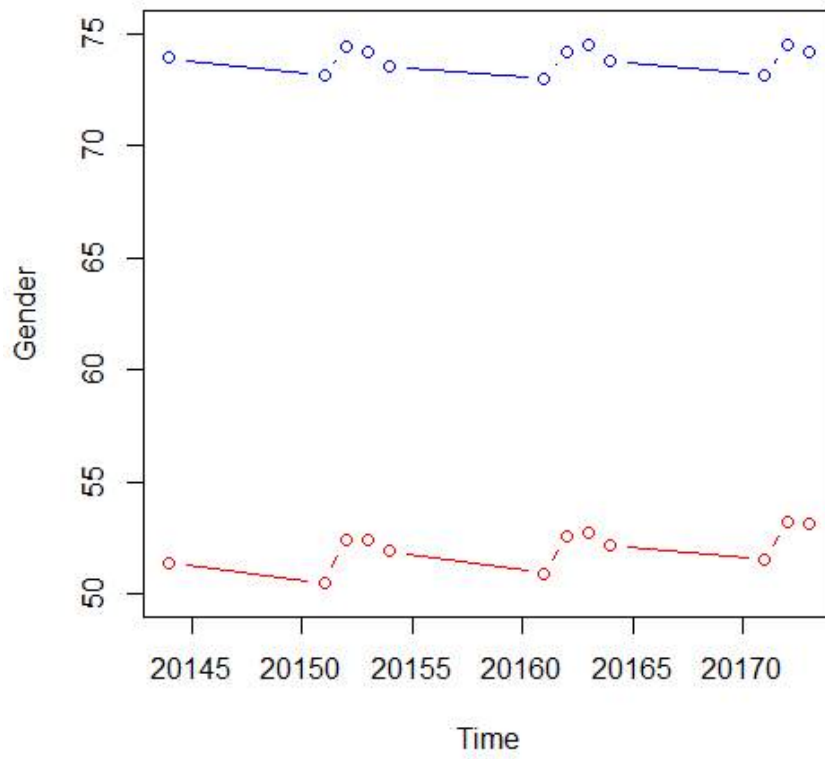


※ 수업시간에 배우지 않은 방법

```
plot(time, men,
      main = "Labor Force Participation Rate",
      type = "b",
      lty = 1,
      col = "blue",
      xlab = "Time",
      ylab = "Gender",
      ylim=c(50,75))
```

```
lines(time, women,
      type = "b",
      col = "red")
```

### Labor Force Participation Rate



#### <16장\_연습4>

- R 에서 MySQL에 접속후 다음 문제를 해결하시오

1. Emp 테이블에서 연봉을 1500 이상 받는 직원들의 모든 정보를 가져와 emp.high 데이터프레임에 저장하시오. emp.high의 내용을 보이시오

```
result <- dbSendQuery (mydb, "select * from emp where sal >= 1500")
```

```
emp.high <- fetch(result, n=-1)
```

```
emp.high
```

	EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
1	7499	ALLEN	SALESMAN	7698	1981-02-20	1600	300	30
2	7566	JONES	MANAGER	7839	1981-04-02	2975	NA	20
3	7698	BLAKE	MANAGER	7839	1981-05-01	2850	NA	30
4	7782	CLARK	MANAGER	7839	1981-06-09	2450	NA	10
5	7788	SCOTT	ANALYST	7566	1982-12-09	3000	NA	20
6	7839	KING	PRESIDENT	NA	1981-11-17	5000	NA	10
7	7844	TURNER	SALESMAN	7698	1981-08-08	1500	0	30
8	7902	FORD	ANALYST	7566	1981-12-03	3000	NA	20

2. 모든 사원의 사원번호, 이름, 급여, 입사일, 부서명, 부서위치를 emp.info 데이터프레임에 저장하시오. emp.info 의 내용을 보이시오

```
dbClearResult (result)
```

```
result <- dbSendQuery (mydb, "select empno, ename, sal, hiredate, dname, loc  
from emp, dept where emp.deptno = dept.deptno")
```

```
emp.info <- fetch(result, n=-1)
```

```
emp.info
```

	empno	ename	sal	hiredate	dname	loc
1	7782	CLARK	2450	1981-06-09	ACCOUNTING	NEW YORK
2	7839	KING	5000	1981-11-17	ACCOUNTING	NEW YORK
3	7934	MILLER	1300	1982-01-23	ACCOUNTING	NEW YORK
4	7369	SMITH	800	1980-12-17	RESEARCH	DALLAS
5	7566	JONES	2975	1981-04-02	RESEARCH	DALLAS
6	7788	SCOTT	3000	1982-12-09	RESEARCH	DALLAS
7	7876	ADAMS	1100	1983-01-12	RESEARCH	DALLAS
8	7902	FORD	3000	1981-12-03	RESEARCH	DALLAS
9	7499	ALLEN	1600	1981-02-20	SALES	CHICAGO
10	7521	WARD	1250	1981-02-22	SALES	CHICAGO
11	7654	MARTIN	1250	1981-08-28	SALES	CHICAGO
12	7698	BLAKE	2850	1981-05-01	SALES	CHICAGO
13	7844	TURNER	1500	1981-08-08	SALES	CHICAGO
14	7900	JAMES	950	1981-12-03	SALES	CHICAGO

3. Dept 테이블에 아래와 같이 2개의 부서를 추가하시오. Workbench 를 통해 2개의 부서가 추가되었는지 확인하시오.

```
dbSendQuery (mydb, "insert into dept values(50, 'Management', 'Paris')")
```

```
dbSendQuery (mydb, "insert into dept values(60, 'Production', 'London')")
```

	DEPTNO	DNAME	LOC
▶	10	ACCOUNTING	NEW YORK
	20	RESEARCH	DALLAS
	30	SALES	CHICAGO
	40	OPERATIONS	BOSTON
	50	Management	Paris
	60	Production	London
•	NULL	NULL	NULL

4. emp 테이블의 모든 정보를 가져와 emp 데이터프레임에 저장하시오. emp 데이터프레임에 대해 다음 작업을 R 로 수행하고 그 결과를 보이시오

```
result <- dbSendQuery (mydb, "select * from emp")
```

```
emp <- data.frame(result)
```

(1) 모든 사원의 이름, 입사일자를 보이시오

```
emp[ , c("ENAME", "HIREDATE")]
```

	ENAME	HIREDATE
1	SMITH	1980-12-17
2	ALLEN	1981-02-20
3	WARD	1981-02-22
4	JONES	1981-04-02
5	MARTIN	1981-08-28
6	BLAKE	1981-05-01
7	CLARK	1981-06-09
8	SCOTT	1982-12-09
9	KING	1981-11-17
10	TURNER	1981-08-08
11	ADAMS	1983-01-12
12	JAMES	1981-12-03
13	FORD	1981-12-03
14	MILLER	1982-01-23



(2) 모든 사원의 급여합계를 보이시오

```
sum(emp[ , "SAL"])
```

결과 : 29025

(3) 모든 사원의 급여를10% 올리려면 얼마의 비용이 추가로 필요한지 보이 시오

```
sum(emp[ , "SAL"])*0.1
```

결과 : 2902.5

(4) 모든 사원의 이름, 급여, 10%올린 급여를 보이시오(힌트. cbind이용)

```
increase <- emp[ , "SAL"]*1.1
```

```
cbind(emp[ , c("ENAME", "SAL")], increase)
```

	ENAME	SAL	increase
1	SMITH	800	880.0
2	ALLEN	1600	1760.0
3	WARD	1250	1375.0
4	JONES	2975	3272.5
5	MARTIN	1250	1375.0
6	BLAKE	2850	3135.0
7	CLARK	2450	2695.0
8	SCOTT	3000	3300.0
9	KING	5000	5500.0
10	TURNER	1500	1650.0
11	ADAMS	1100	1210.0
12	JAMES	950	1045.0
13	FORD	3000	3300.0
14	MILLER	1300	1430.0

(5) 급여를 2000 이상 받는 사원의 이름, 급여, 담당업무를 보이시오

```
subset(emp[ , c("ENAME", "SAL", "JOB")], SAL >= 2000)
```

또는

```
subset(emp, SAL >= 2000)[ , c("ENAME", "SAL", "JOB")]
```

	ENAME	SAL	JOB
4	JONES	2975	MANAGER
6	BLAKE	2850	MANAGER
7	CLARK	2450	MANAGER
8	SCOTT	3000	ANALYST
9	KING	5000	PRESIDENT
13	FORD	3000	ANALYST

(6) 급여를 2000 이상 받고 부서번호가 20 인 사원의 이름, 급여, 입사일자를 보이시오

```
subset(emp[ , c("ENAME", "SAL", "HIREDATE", "DEPTNO")], SAL>=2000 &  
DEPTNO==20)
```

	ENAME	SAL	HIREDATE	DEPTNO
4	JONES	2975	1981-04-02	20
8	SCOTT	3000	1982-12-09	20
13	FORD	3000	1981-12-03	20