

REPORT



제 목 : 데이터베이스기초 16장

과 목 명 : 데이터베이스기초

담당교수 : 오 세 종 교수님

이 름 : 조 정 민

학 번 : 32164420

제 출 일 : 2019.06.09.



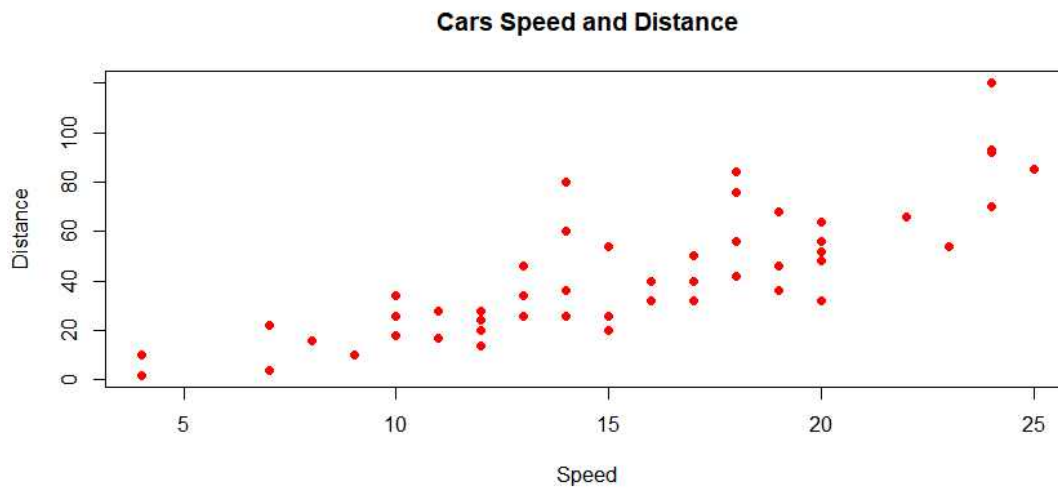
단국대학교
Dankook University

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

```
spe <- cars[, "speed"]
```

```
dis <- cars[, "dist"]
```

```
plot(spe, dis, main="Cars Speed and Distance", xlab="Speed", ylab="Distance", col="red",  
pch=19)
```



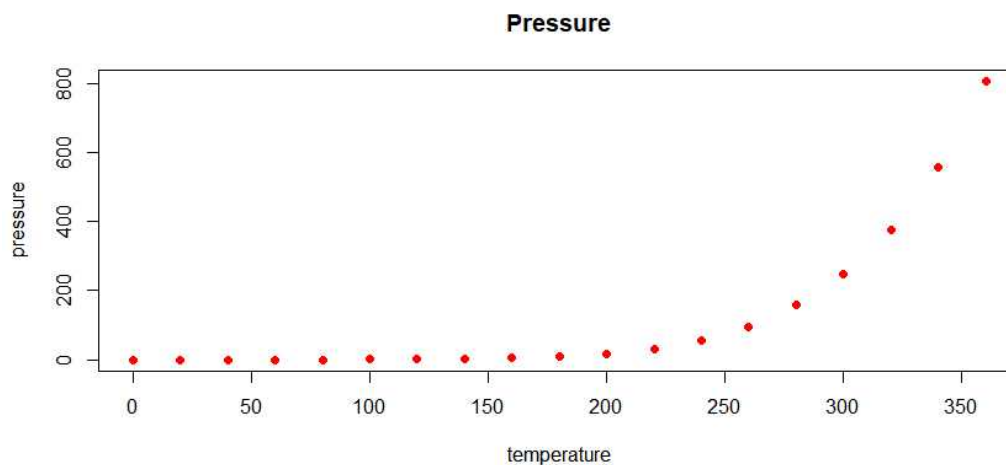
>> Speed가 높을수록 차가 이동하는 Distance가 높아진다.

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

```
temp <- pressure[, "temperature"]
```

```
press <- pressure[, "pressure"]
```

```
plot(temp, press, main="Pressure", xlab="temperature", ylab="pressure", col="red", pch=19)
```



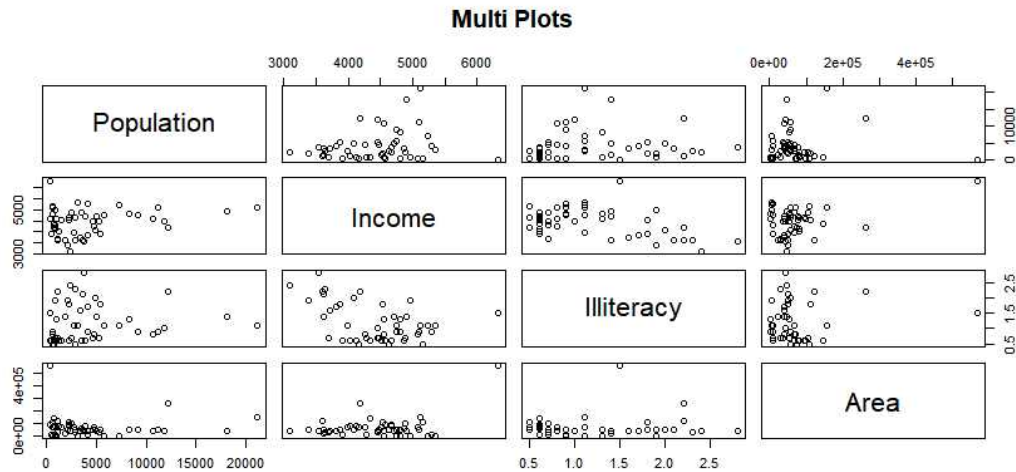
>> 온도가 높을수록 압력도 높아진다.

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

```
a <- c("Population","Income","Illiteracy","Area")
```

```
target <- state.x77[,a]
```

```
pairs(target, main="Multi Plots")
```



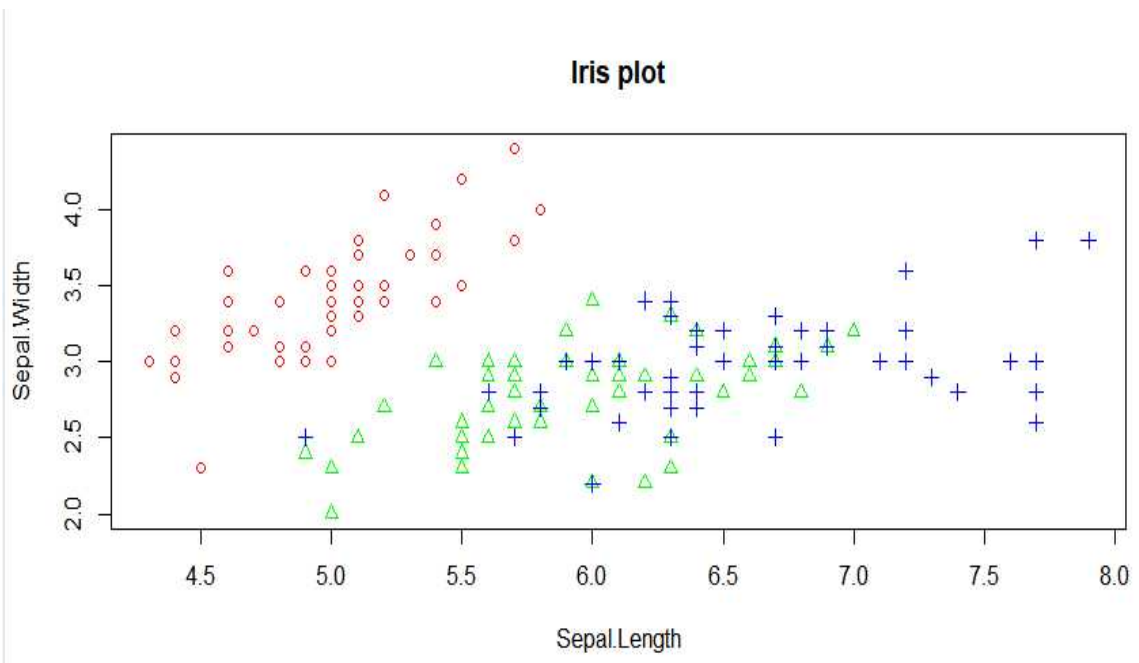
4. iris 데이터셋에서 Species 정보에 따른 Sepal.Length, Sepal.Width (꽃받침의 길이, 폭)의 분포를 알아 보시오

```
iris.2 <- iris[,1:2]
```

```
point <- as.numeric(iris$Species)
```

```
color <- c("red","green","blue")
```

```
plot(iris.2, main="Iris plot", pch=c(point), col=color[point])
```



1. 다음은 10명의 수입과 교육받은 기간을 조사한 표이다. 수입과 교육기간 사이에 어느정도 상관관계가 있는지 조사하시

오 (산점도, 상관계수 구하기)

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

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

```
tbl <- data.frame(cbind(Income, YofE))
```

```
plot(YofE~Income, data=tbl)
```

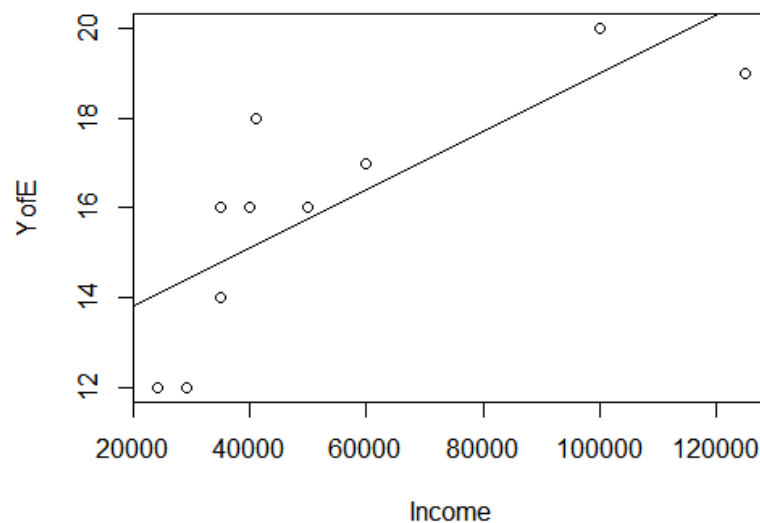
```
res=lm(YofE~Income, data=tbl)
```

```
abline(res)
```

```
cor(Income, YofE)
```

```
> cor(Income, YofE)
```

```
[1] 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)
```

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

```
tbl <- data.frame(cbind(gpa,time))
```

```
plot(time~gpa, data=tbl)
```

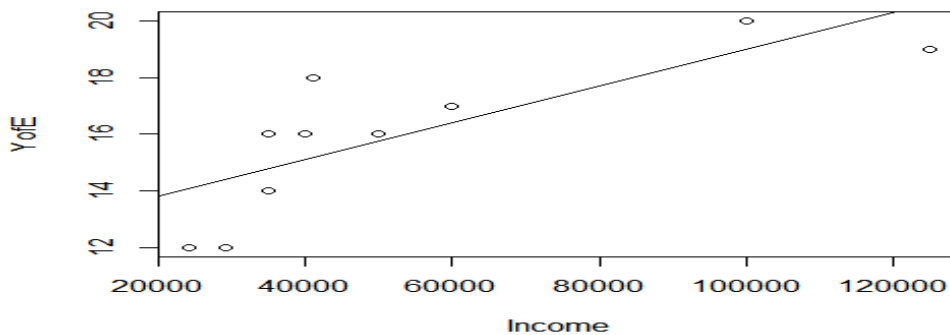
```
res=lm(time~gpa, data=tbl)
```

```
abline(res)
```

```
cor(gpa,time)
```

```
> cor(Income, YofE)
```

```
[1] 0.7887259
```



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

```
res=cor(mtcars$mpg,mtcars[-1])
```

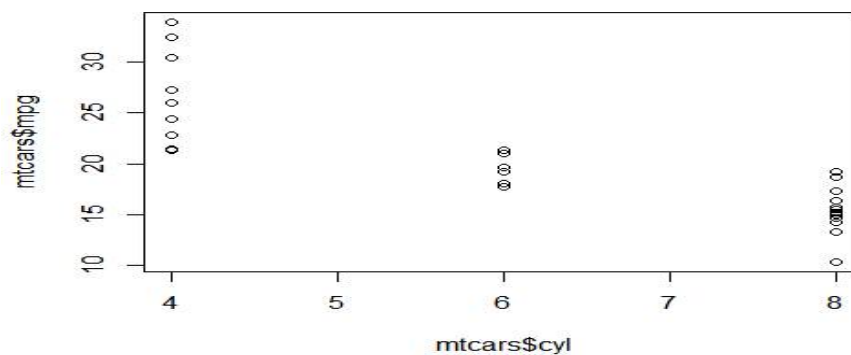
```
res
```

```
plot(mtcars$mpg~mtcars$cyl, data=mtcars)
```

```
> res
```

	cyl	disp	hp	drat	wt
[1,]	-0.852162	-0.8475514	-0.7761684	0.6811719	-0.8676594

	qsec	vs	am	gear	carb
[1,]	0.418684	0.6640389	0.5998324	0.4802848	-0.5509251

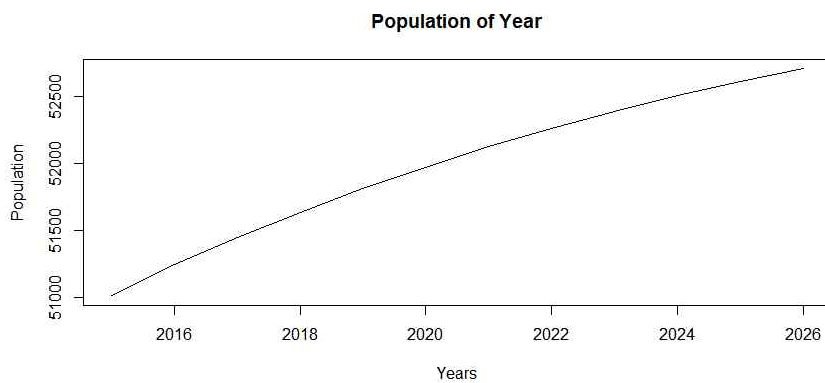


1. 다음은 2015년부터 2026년도까지의 예상 인구수 추계자료이다. 선그래프를 작성하시오.

```
years <- 2015:2026
```

```
pop <- c(51014,51245,51446,51635,51811,51973,52123,52261,52388,52504,52609,52704)
```

```
plot(years,pop,main="Population of Year",type="l", lty=1,lwd=1, xlab="Years", ylab="Population")
```



2. 다음은 2014년 4분기부터 2017년 3분기까지 남,녀의 경제활동참가율 통계이다. 선그래프를 작성하시오 (남,녀를 각각 다른 선으로 표시)

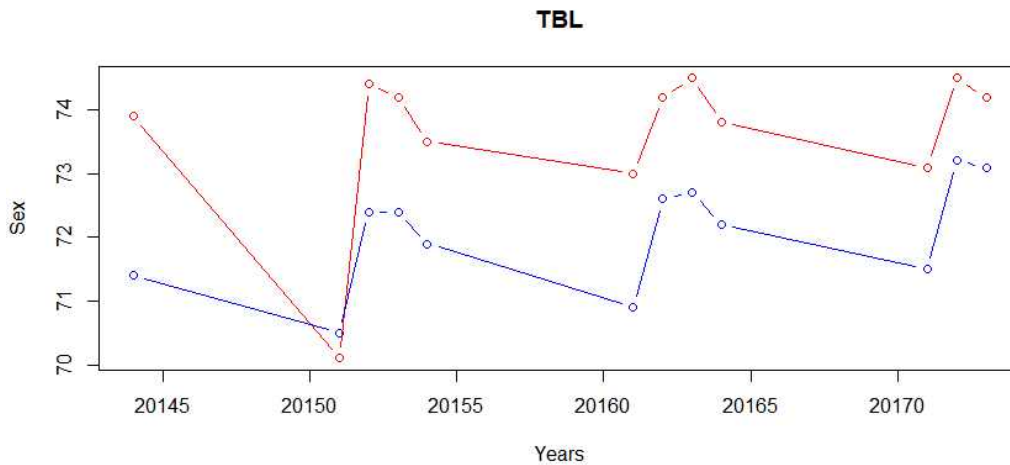
```
years <- c(20144,20151,20152,20153,20154,20161,20162,20163,20164,20171,20172,20173)
```

```
male <- c(73.9,70.1,74.4,74.2,73.5,73.0,74.2,74.5,73.8,73.1,74.5,74.2)
```

```
female <- c(71.4,70.5,72.4,72.4,71.9,70.9,72.6,72.7,72.2,71.5,73.2,73.1)
```

```
plot(years,male,main="TBL",type="b", lty=1,lwd=1,col="red",xlab="Years",ylab="Sex")
```

```
lines(years,female,type="b",col="blue")
```



1. R 에서 제공하는 state.x77 (미국 50개주에 대한 통계)데이터셋을 분석해보시오

● State.x77 에 지역정보(state.region) 추가하여 분석

```
st <- data.frame(state.x77, state.region)
```

```
par(mfrow=c(3,3))
```

```
boxplot(Population~state.region, data = st,
        main = "Region : Population")
```

```
boxplot(Income~state.region, data = st,
        main = "Region : Income")
```

```
boxplot(Illiteracy~state.region, data = st,
        main = "Region : Illiteracy")
```

```
boxplot(Life.Exp~state.region, data = st,
        main = "Region : Life.Exp")
```

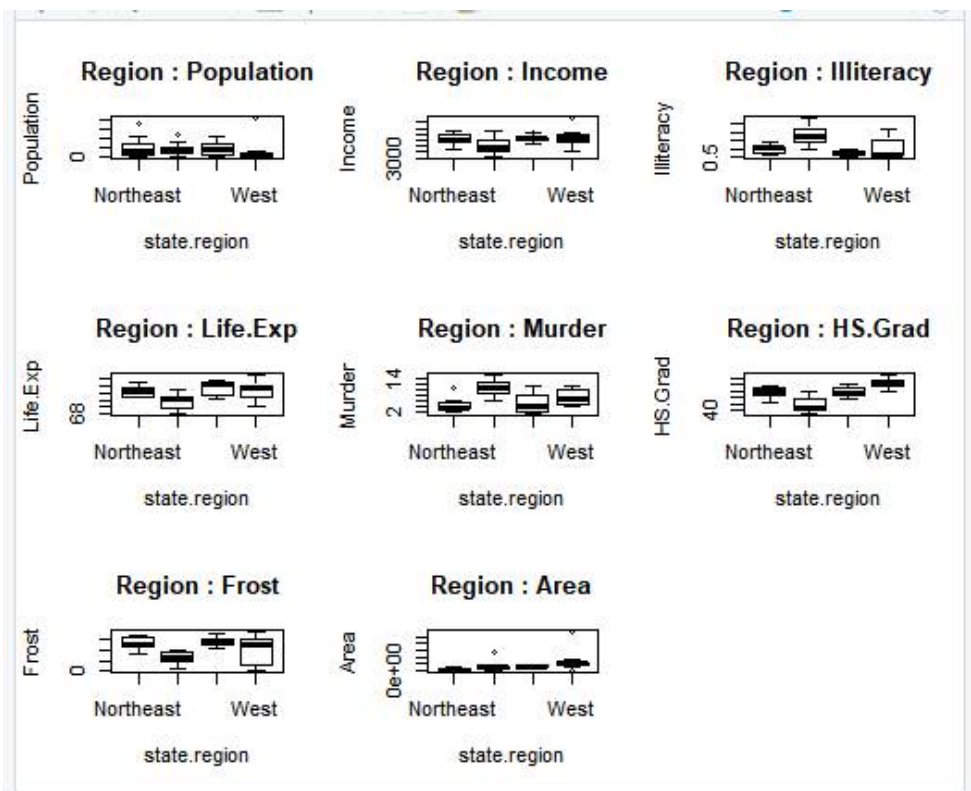
```
boxplot(Murder~state.region, data = st,
        main = "Region : Murder")
```

```
boxplot(HS.Grad~state.region, data = st,
        main = "Region : HS.Grad")
```

```
boxplot(Frost~state.region, data = st,
        main = "Region : Frost")
```

```
boxplot(Area~state.region, data = st,
        main = "Region : Area")
```

p a r (m f r o w = c (1 , 1))



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

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

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

```
emp.high
```

```
> emp.high
```

	EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM
1	7499	ALLEN	SALESMAN	7698	1981-02-20	1600.0	300
2	7566	JONES	MANAGER	7839	1981-04-02	3272.5	NA
3	7698	BLAKE	MANAGER	7839	1981-05-01	2850.0	NA
4	7782	CLARK	MANAGER	7839	1981-06-09	2450.0	NA
5	7788	SCOTT	ANALYST	7566	1982-12-09	3000.0	NA
6	7839	KING	PRESIDENT	NA	1981-11-17	5000.0	NA
7	7844	TURNER	SALESMAN	7698	1981-08-08	1500.0	0
8	7902	FORD	ANALYST	7566	1981-12-03	3000.0	NA

	DEPTNO
1	30
2	20
3	30
4	10
5	20
6	10
7	30
8	20

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

```
data2 <- dbSendQuery(mydb,
```

```
"select ename, sal, hiredate, dname, loc
```

```
from emp e, dept d where e.deptno=d.deptno")
```

```
emp.info <- fetch(data2, n = -1)
```

```
emp.info
```

```
> emp.info
```

	ename	sal	hiredate	dname	loc
1	CLARK	2450.0	1981-06-09	ACCOUNTING	NEW YORK
2	KING	5000.0	1981-11-17	ACCOUNTING	NEW YORK
3	MILLER	1300.0	1982-01-23	ACCOUNTING	NEW YORK
4	smith	960.0	1980-12-17	RESEARCH	DALLAS
5	JONES	3272.5	1981-04-02	RESEARCH	DALLAS
6	SCOTT	3000.0	1982-12-09	RESEARCH	DALLAS
7	ADAMS	1100.0	1983-01-12	RESEARCH	DALLAS
8	FORD	3000.0	1981-12-03	RESEARCH	DALLAS
9	ALLEN	1600.0	1981-02-20	SALES	CHICAGO
10	WARD	1250.0	1981-02-22	SALES	CHICAGO
11	MARTIN	1250.0	1981-08-28	SALES	CHICAGO
12	BLAKE	2850.0	1981-05-01	SALES	CHICAGO
13	TURNER	1500.0	1981-08-08	SALES	CHICAGO
14	JAMES	950.0	1981-12-03	SALES	CHICAGO

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

```
dbSendQuery(mydb,
             "insert into dept values
             (70,'Management','Paris'),
             (80,'Production','London'))"
```

	DEPTNO	DNAME	LOC
	40	OPERATIONS	BOSTON
	60	DB	seoul
	70	Management	Paris
	80	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

```
data <- dbSendQuery(mydb, "select * from emp")
emp <- fetch(data, n=-1)
```

emp

```
> emp
  EMPNO  ENAME      JOB   MGR  HIREDATE    SAL  COMM  DEPTNO
1   7369  smith      CLERK 7902 1980-12-17  960.0   NA     20
2   7499  ALLEN      SALESMAN 7698 1981-02-20 1600.0  300    30
3   7521  WARD        SALESMAN 7698 1981-02-22 1250.0  500    30
4   7566  JONES        MANAGER 7839 1981-04-02 3272.5   NA     20
5   7654  MARTIN       SALESMAN 7698 1981-08-28 1250.0 1400    30
6   7698  BLAKE        MANAGER 7839 1981-05-01 2850.0   NA     30
7   7782  CLARK        MANAGER 7839 1981-06-09 2450.0   NA     10
8   7788  SCOTT        ANALYST 7566 1982-12-09 3000.0   NA     20
9   7839  KING         PRESIDENT   NA 1981-11-17 5000.0   NA     10
10  7844  TURNER       SALESMAN 7698 1981-08-08 1500.0   0      30
11  7876  ADAMS        CLERK   7788 1983-01-12 1100.0   NA     20
12  7900  JAMES        CLERK   7698 1981-12-03  950.0   NA     20
13  7902  FORD         ANALYST 7566 1981-12-03 3000.0   NA     20
14  7934  MILLER       CLERK   7902 1982-01-23 1300.0   NA     10
```

(1) 모든 사원의 이름, 입사일자를 보이시오

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

```
> 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"])
```

```
> sum(emp[, "SAL"])  
[1] 29482.5
```

(3) 모든 사원의 급여를 10% 올리려면 얼마의 비용이 추가로 필요한지 보이시오

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

```
> 0.1 * sum(emp[, "SAL"])  
[1] 2948.25
```

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

```
tb <- cbind(emp[, "ENAME"], emp[, "SAL"], 1.1 * emp[, "SAL"])
```

```
colnames(tb) <- c("ENAME", "SAL", "1.1 * SAL")
```

```
tb
```

```
> tb  
      ENAME      SAL      1.1 * SAL  
[1,] "smith"    "960"    "1056"  
[2,] "ALLEN"    "1600"   "1760"  
[3,] "WARD"     "1250"   "1375"  
[4,] "JONES"    "3272.5"  "3599.75"  
[5,] "MARTIN"   "1250"   "1375"  
[6,] "BLAKE"    "2850"   "3135"  
[7,] "CLARK"    "2450"   "2695"  
[8,] "SCOTT"    "3000"   "3300"  
[9,] "KING"     "5000"   "5500"  
[10,] "TURNER"  "1500"   "1650"  
[11,] "ADAMS"   "1100"   "1210"  
[12,] "JAMES"   "950"    "1045"  
[13,] "FORD"    "3000"   "3300"  
[14,] "MILLER"  "1300"   "1430"
```

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

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

```
> subset(emp[, c("ENAME", "SAL", "JOB")], emp[, "SAL"] >= 2000)  
      ENAME      SAL      JOB  
4 JONES 3272.5  MANAGER  
6 BLAKE 2850.0  MANAGER  
7 CLARK 2450.0  MANAGER  
8 SCOTT 3000.0  ANALYST  
9 KING 5000.0  PRESIDENT  
13 FORD 3000.0  ANALYST
```

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

```
subset(emp[, c("ENAME", "SAL", "JOB", "DEPTNO")],
```

```
      SAL >= 2000 & DEPTNO == 20)
```

```
> subset(emp[, c("ENAME", "SAL", "JOB", "DEPTNO")],  
+      SAL >= 2000 & DEPTNO == 20)  
      ENAME      SAL      JOB DEPTNO  
4 JONES 3272.5  MANAGER      20  
8 SCOTT 3000.0  ANALYST      20  
13 FORD 3000.0  ANALYST      20
```