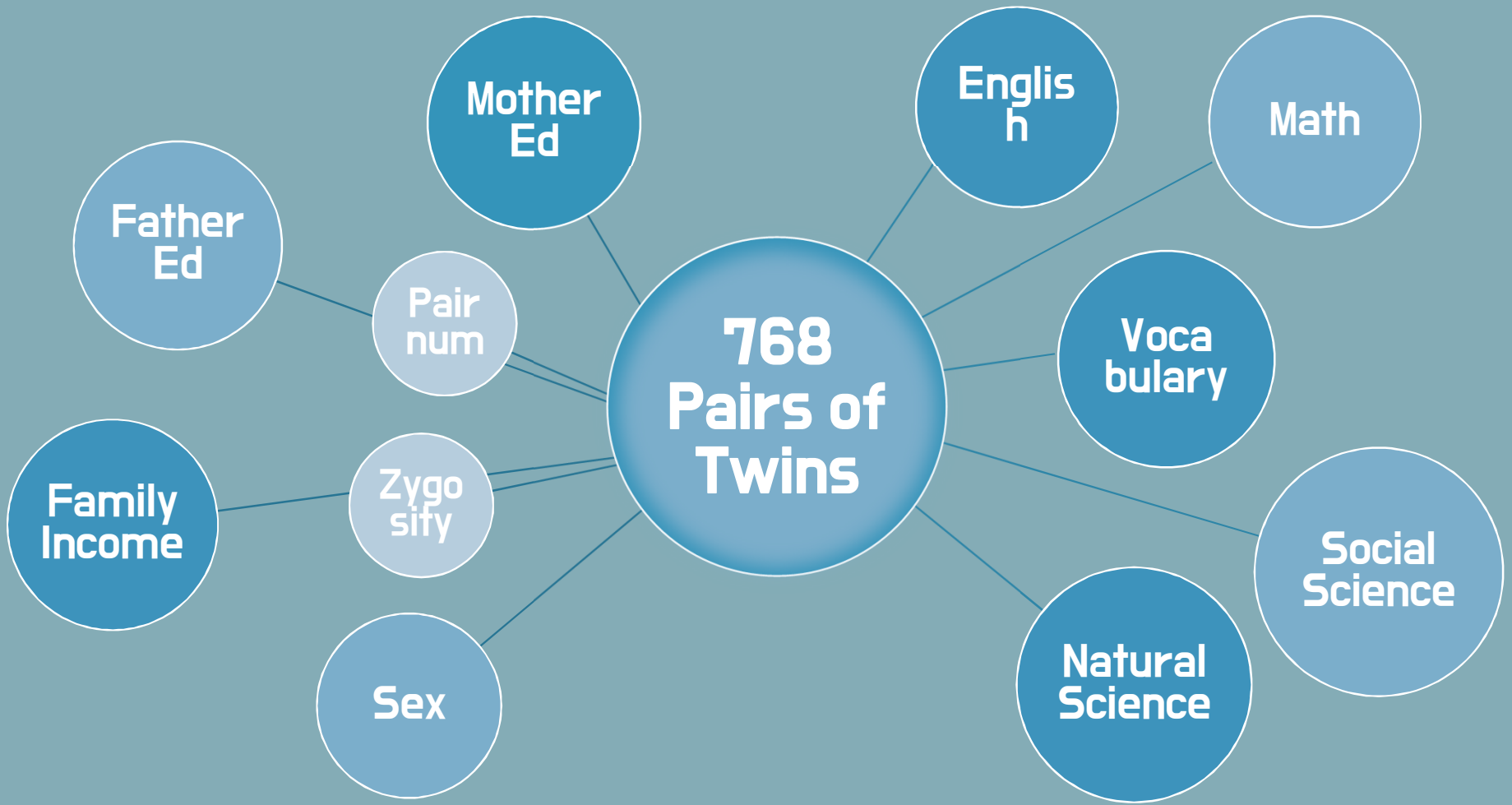
# Twins Analysis

7조

1302085 이응경
1302124 최희원

# Data Description



# Data Description

| pairnum | sex | zygosity | moed | faed | faminc | english | math | socsci | natsci | vocab | even |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 1 | 3 | 4 | 2 | 14 | 13 | 17 | 18 | 14 | FALSE |
| 1 | 2 | 1 | 3 | 4 | 2 | 11 | 14 | 15 | 10 | 12 | TRUE |
| 2 | 2 | 1 | 1 | 1 | 1 | 20 | 20 | 16 | 16 | 13 | FALSE |
| 2 | 2 | 1 | 1 | 1 | 1 | 17 | 19 | 13 | 13 | 14 | TRUE |
| 3 | 2 | 1 | 1 | 1 | 1 | 11 | 8 | 15 | 16 | 12 | FALSE |
| 3 | 2 | 1 | 1 | 1 | 1 | 16 | 13 | 13 | 8 | 15 | TRUE |
| 4 | 1 | 2 | 3 | 2 | 4 | 9 | 19 | 7 | 10 | 6 | FALSE |
| 4 | 1 | 2 | 3 | 2 | 4 | 8 | 16 | 15 | 17 | 11 | TRUE |
| 5 | 1 | 2 | 5 | 4 | 3 | 15 | 23 | 23 | 21 | 21 | FALSE |
| 5 | 1 | 2 | 5 | 4 | 3 | 15 | 13 | 13 | 20 | 19 | TRUE |
| 6 | 2 | 2 | 4 | 2 | 3 | 20 | 17 | 16 | 12 | 12 | FALSE |
| 6 | 2 | 2 | 4 | 2 | 3 | 19 | 18 | 13 | 18 | 15 | TRUE |
| 7 | 1 | 1 | 4 | 6 | 3 | 26 | 20 | 27 | 23 | 28 | FALSE |
| 7 | 1 | 1 | 4 | 6 | 3 | 25 | 20 | 29 | 24 | 30 | TRUE |
| 8 | 1 | 1 | 4 | 6 | 5 | 28 | 31 | 31 | 29 | 30 | FALSE |
| 8 | 1 | 1 | 4 | 6 | 5 | 27 | 28 | 28 | 27 | 30 | TRUE |

# Objectives

## Liberal Arts / Science

- Any relations between each subject?
- If there are, what are the criteria?

## High / Medium / Low Grade

- Any differences in the factors affecting each grade?
- How is it different?

## Well / Low Parents' Educational Level

- Any differences in twins' grades depending on the parents' education?
- Is there a positive correlation?

# Objectives

## High / Low Income

- Any differences in twins' grades depending on family income?
- Is there a positive correlation?

## Male / Female

- Any differences in the factors depending on the sex?
- These factors differ depending on the subject?

## Difference between Twins

- If the home environment is the same, are twins' grades same too?

# Approach

## Summary

- mean
  → average grade of each cluster
- cor
  → relations btw variables after clustering

## PCA

- liberal arts / science

## Clustering

- 3 levels of grades
- high / low parents' ed
- high / low income

## Graph

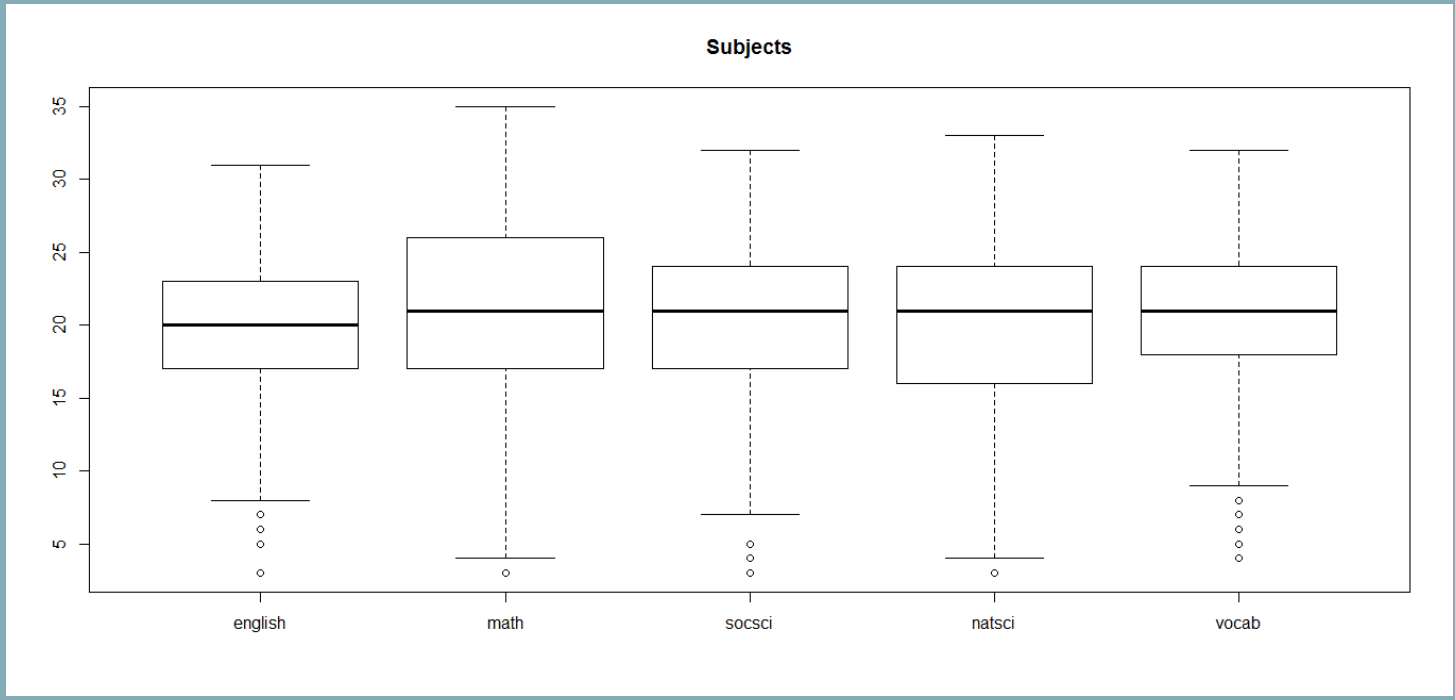
- boxplot
- compare each cluster

## Hotelling's $T^2$ Test

- difference btw twins

# Summary Statistics

```
> summary(twins[,4:11])
      moed            faed           faminc          english          math            socsci          natsci          vocab
 Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   : 3.00   Min.   : 3.00   Min.   : 3.00   Min.   : 3.00   Min.   : 4.00
 1st Qu.:3.000   1st Qu.:3.000   1st Qu.:2.000   1st Qu.:17.00   1st Qu.:17.00   1st Qu.:17.00   1st Qu.:16.00   1st Qu.:18.00
 Median :3.000   Median :4.000   Median :3.000   Median :20.00   Median :21.00   Median :21.00   Median :21.00   Median :21.00
 Mean   :3.424   Mean   :3.587   Mean   :3.241   Mean   :19.76   Mean   :21.25   Mean   :20.69   Mean   :20.09   Mean   :21.06
 3rd Qu.:4.000   3rd Qu.:5.000   3rd Qu.:4.000   3rd Qu.:23.00   3rd Qu.:26.00   3rd Qu.:24.00   3rd Qu.:24.00   3rd Qu.:24.00
 Max.   :6.000   Max.   :6.000   Max.   :7.000   Max.   :31.00   Max.   :35.00   Max.   :32.00   Max.   :33.00   Max.   :32.00
```



# Principal Components Analysis

```
> summary(princomp(subjects))
Importance of components:
                          Comp.1    Comp.2     Comp.3     Comp.4     Comp.5
Standard deviation     9.918813 4.0736321 3.35238377 2.93207343 2.22615435
Proportion of Variance 0.703898 0.1187282 0.08040772 0.06150919 0.03545691
Cumulative Proportion  0.703898 0.8226262 0.90303391 0.96454309 1.00000000
```

```
> princomp(subjects)$loadings

Loadings:
        Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
english -0.365 -0.363 -0.277  0.794 -0.166
math    -0.541  0.742 -0.393
socsci  -0.423 -0.325        -0.494 -0.686
natsci  -0.490         0.845  0.139  0.129
vocab   -0.393 -0.451 -0.231 -0.323  0.696

               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
SS loadings       1.0    1.0    1.0    1.0    1.0
Proportion Var    0.2    0.2    0.2    0.2    0.2
Cumulative Var    0.2    0.4    0.6    0.8    1.0
```

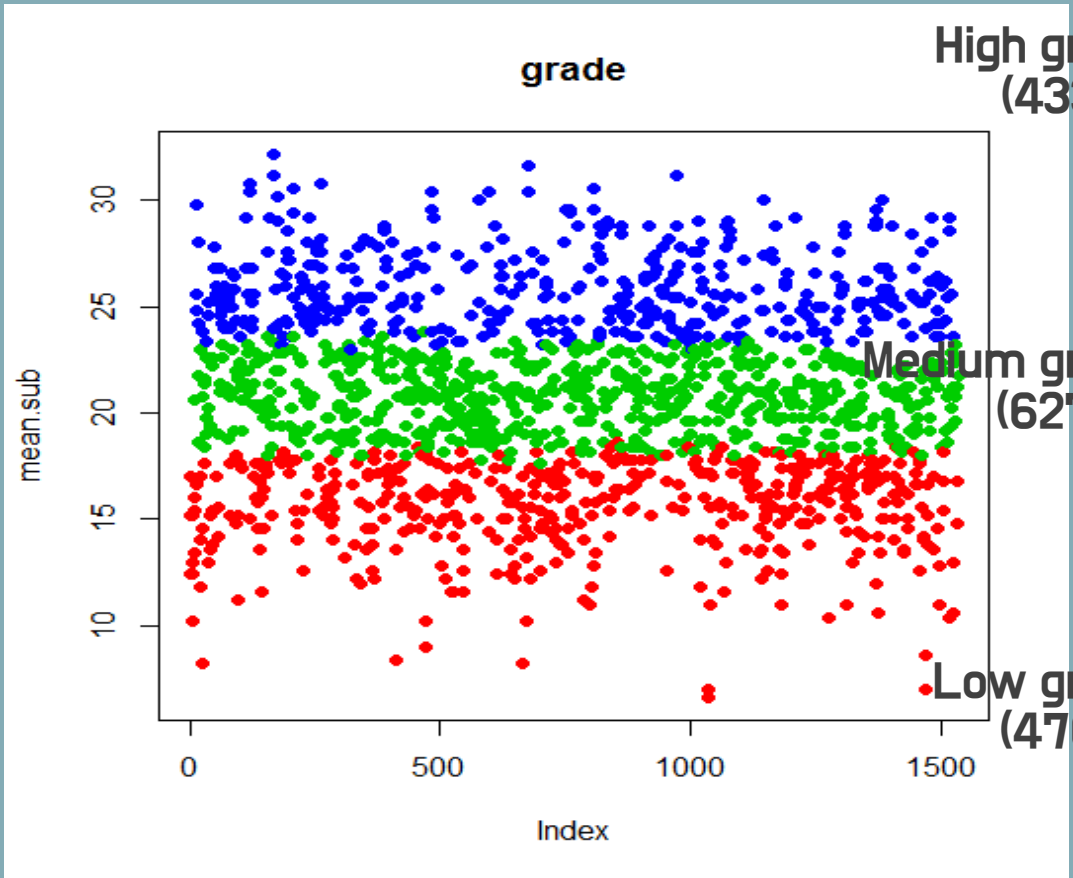
Only natsci has different disposition.
→ cannot divide into liberal arts / science

# Cluster Analysis-Grade

```
> cor(subjects)
          english      math    socsci    natsci     vocab
english 1.0000000 0.5412699 0.6339606 0.5783527 0.6692687
math    0.5412699 1.0000000 0.6086261 0.6583328 0.5556737
socsci  0.6339606 0.6086261 1.0000000 0.6749258 0.7719540
natsci  0.5783527 0.6583328 0.6749258 1.0000000 0.6004781
vocab   0.6692687 0.5556737 0.7719540 0.6004781 1.0000000
```

```
> head(twins)
  pairnum sex zygosity moed faed faminc english math socsci natsci vocab  even mean.sub
1       1   2        1    3    4      2      14   13     17     18    14 FALSE     15.2
2       1   2        1    3    4      2      11   14     15     10    12  TRUE     12.4
3       2   2        1    1    1      1      20   20     16     16    13 FALSE     17.0
4       2   2        1    1    1      1      17   19     13     13    14  TRUE     15.2
5       3   2        1    1    1      1      11    8     15     16    12 FALSE     12.4
6       3   2        1    1    1      1      16   13     13      8    15  TRUE     13.0
```

# Cluster Analysis-Grade




High grade (433)

| english | math | socsci | natsci | vocab |
|---|---|---|---|---|
| Min. :15.00 | Min. :17.00 | Min. :17.00 | Min. :18.00 | Min. :18.00 |
| 1st Qu.:22.00 | 1st Qu.:25.00 | 1st Qu.:24.00 | 1st Qu.:24.00 | 1st Qu.:24.00 |
| Median :24.00 | Median :28.00 | Median :26.00 | Median :26.00 | Median :26.00 |
| Mean :24.08 | Mean :27.88 | Mean :25.94 | Mean :25.85 | Mean :25.92 |
| 3rd Qu.:26.00 | 3rd Qu.:31.00 | 3rd Qu.:28.00 | 3rd Qu.:28.00 | 3rd Qu.:28.00 |
| Max. :31.00 | Max. :35.00 | Max. :32.00 | Max. :33.00 | Max. :32.00 |

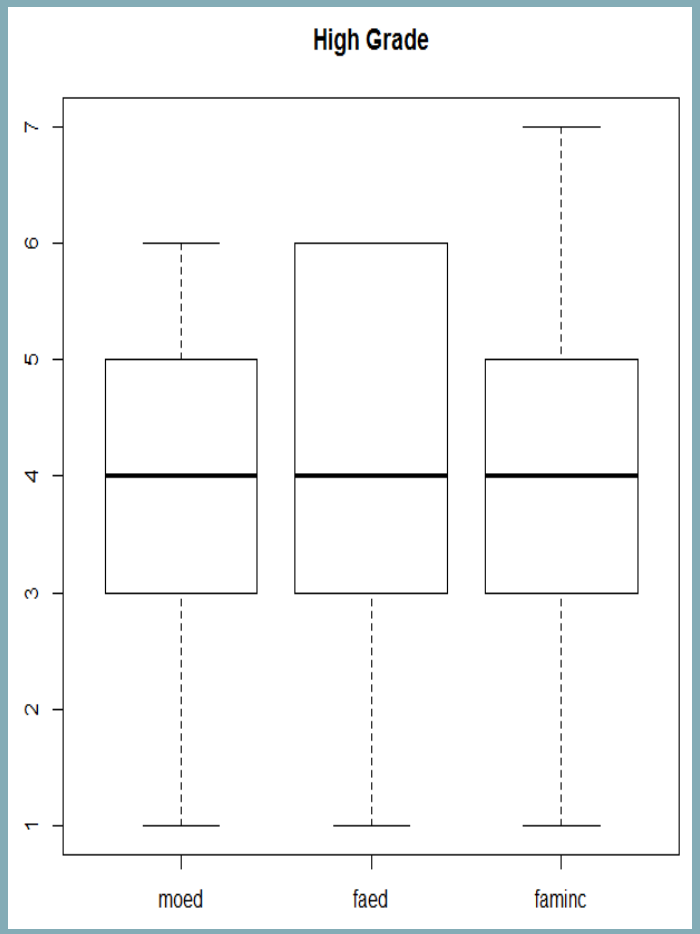
Medium grade (627)

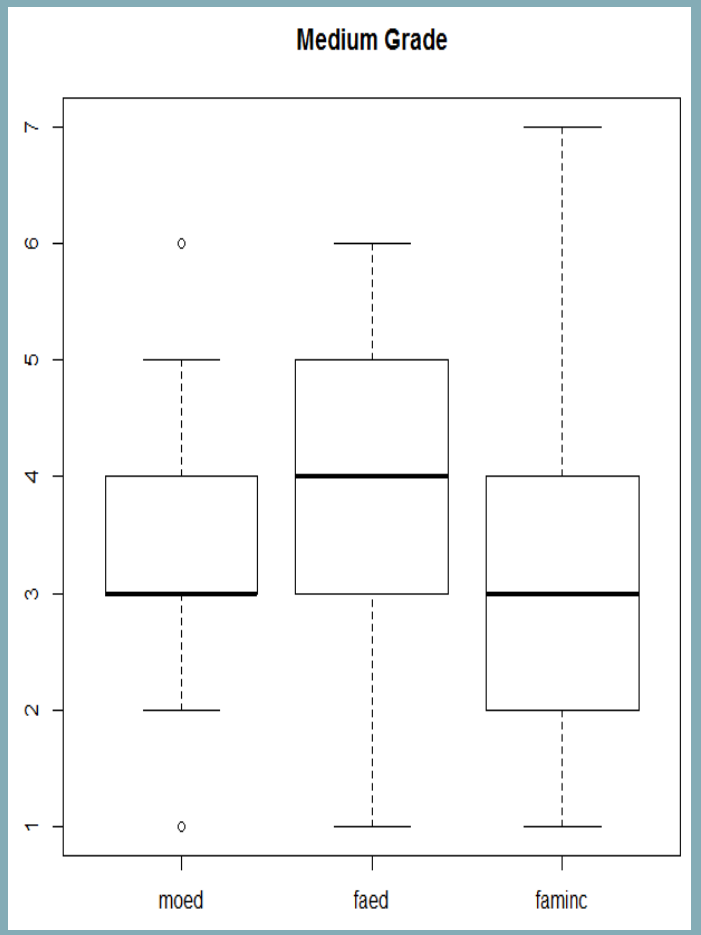
| english | math | socsci | natsci | vocab |
|---|---|---|---|---|
| Min. :10.00 | Min. :10.00 | Min. :11.00 | Min. : 9.00 | Min. :11.00 |
| 1st Qu.:18.00 | 1st Qu.:19.00 | 1st Qu.:19.00 | 1st Qu.:18.00 | 1st Qu.:19.00 |
| Median :20.00 | Median :21.00 | Median :21.00 | Median :21.00 | Median :21.00 |
| Mean :19.87 | Mean :21.28 | Mean :20.69 | Mean :20.56 | Mean :21.14 |
| 3rd Qu.:22.00 | 3rd Qu.:24.00 | 3rd Qu.:23.00 | 3rd Qu.:23.00 | 3rd Qu.:23.00 |
| Max. :28.00 | Max. :33.00 | Max. :30.00 | Max. :29.00 | Max. :32.00 |

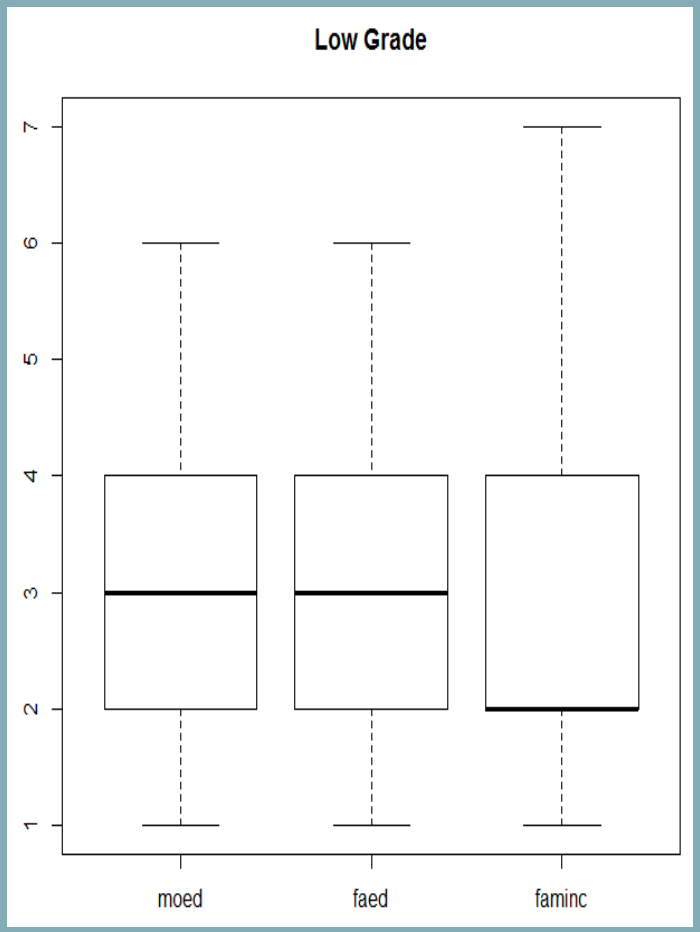
Low grade (476)

| english | math | socsci | natsci | vocab |
|---|---|---|---|---|
| Min. : 3.00 | Min. : 3.00 | Min. : 3.00 | Min. : 3.00 | Min. : 4.00 |
| 1st Qu.:13.00 | 1st Qu.:12.00 | 1st Qu.:14.00 | 1st Qu.:12.00 | 1st Qu.:15.00 |
| Median :16.00 | Median :15.00 | Median :16.00 | Median :14.00 | Median :17.00 |
| Mean :15.68 | Mean :15.19 | Mean :15.91 | Mean :14.24 | Mean :16.54 |
| 3rd Qu.:18.00 | 3rd Qu.:18.00 | 3rd Qu.:18.00 | 3rd Qu.:17.00 | 3rd Qu.:18.00 |
| Max. :24.00 | Max. :26.00 | Max. :24.00 | Max. :23.00 | Max. :28.00 |

# Cluster Analysis-Grade



# Cluster Analysis-Grade

High grade

```
> round(cor(grade3[,4:11]),2)
         moed faed faminc english math socsci natsci vocab
moed     1.00 0.60   0.39    0.20 0.15   0.11   0.13  0.18
faed     0.60 1.00   0.55    0.18 0.18   0.19   0.11  0.19
faminc   0.39 0.55   1.00    0.04 0.18   0.19   0.09  0.18
english  0.20 0.18   0.04    1.00 0.04   0.22   0.19  0.33
math     0.15 0.18   0.18    0.04 1.00   0.14   0.32  0.00
socsci   0.11 0.19   0.19    0.22 0.14   1.00   0.33  0.48
natsci   0.13 0.11   0.09    0.19 0.32   0.33   1.00  0.20
vocab    0.18 0.19   0.18    0.33 0.00   0.48   0.20  1.00
```

Medium grade

```
> round(cor(grade2[,4:11]),2)
          moed  faed faminc english  math socsci natsci vocab
moed      1.00  0.51   0.34    0.03  0.04  -0.02   0.04  0.09
faed      0.51  1.00   0.52    0.09  0.09   0.02  -0.10  0.13
faminc    0.34  0.52   1.00    0.08  0.06   0.00  -0.07  0.12
english   0.03  0.09   0.08    1.00 -0.12   0.05  -0.07  0.27
math      0.04  0.09   0.06   -0.12  1.00  -0.16   0.04 -0.20
socsci   -0.02  0.02   0.00    0.05 -0.16   1.00   0.02  0.43
natsci    0.04 -0.10  -0.07   -0.07  0.04   0.02   1.00 -0.15
vocab     0.09  0.13   0.12    0.27 -0.20   0.43  -0.15  1.00
```

Low grade

```
> round(cor(grade1[,4:11]),2)
          moed  faed faminc english math socsci natsci vocab
moed      1.00  0.51   0.42   -0.05 0.03   0.00  -0.03  0.11
faed      0.51  1.00   0.41   -0.03 0.07   0.04  -0.06  0.18
faminc    0.42  0.41   1.00    0.06 0.13  -0.02   0.01  0.16
english  -0.05 -0.03   0.06    1.00 0.14   0.34   0.16  0.33
math      0.03  0.07   0.13    0.14 1.00   0.06   0.10  0.00
socsci    0.00  0.04  -0.02    0.34 0.06   1.00   0.16  0.40
natsci   -0.03 -0.06   0.01    0.16 0.10   0.16   1.00  0.05
vocab     0.11  0.18   0.16    0.33 0.00   0.40   0.05  1.00
```

# Cluster Analysis-Education Level

value edfmt

1='<= 8th grade'

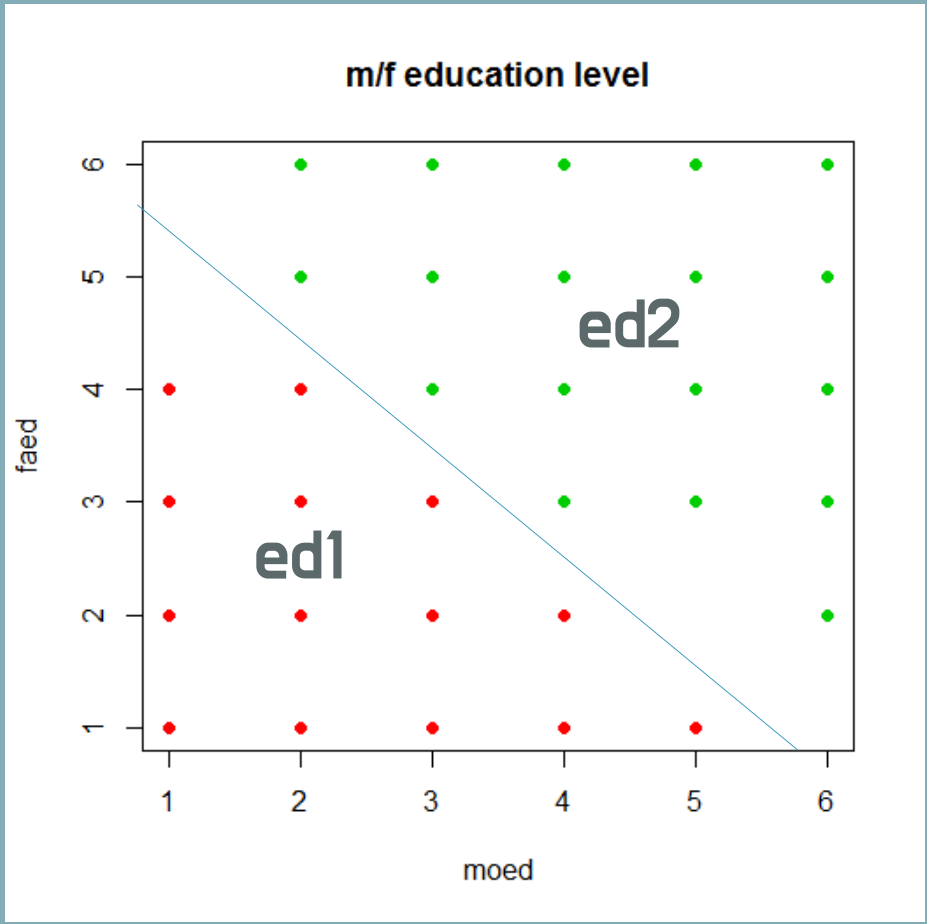
2='part high school'

3='high school grad'

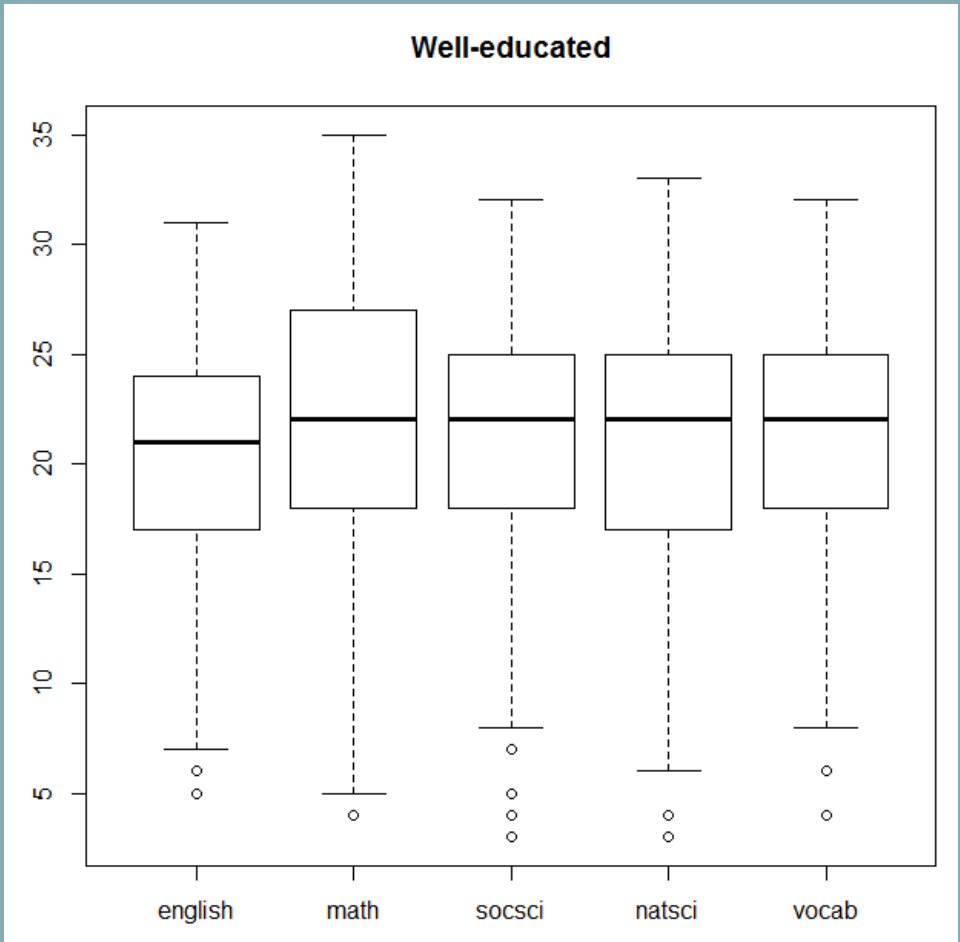
4='part college'

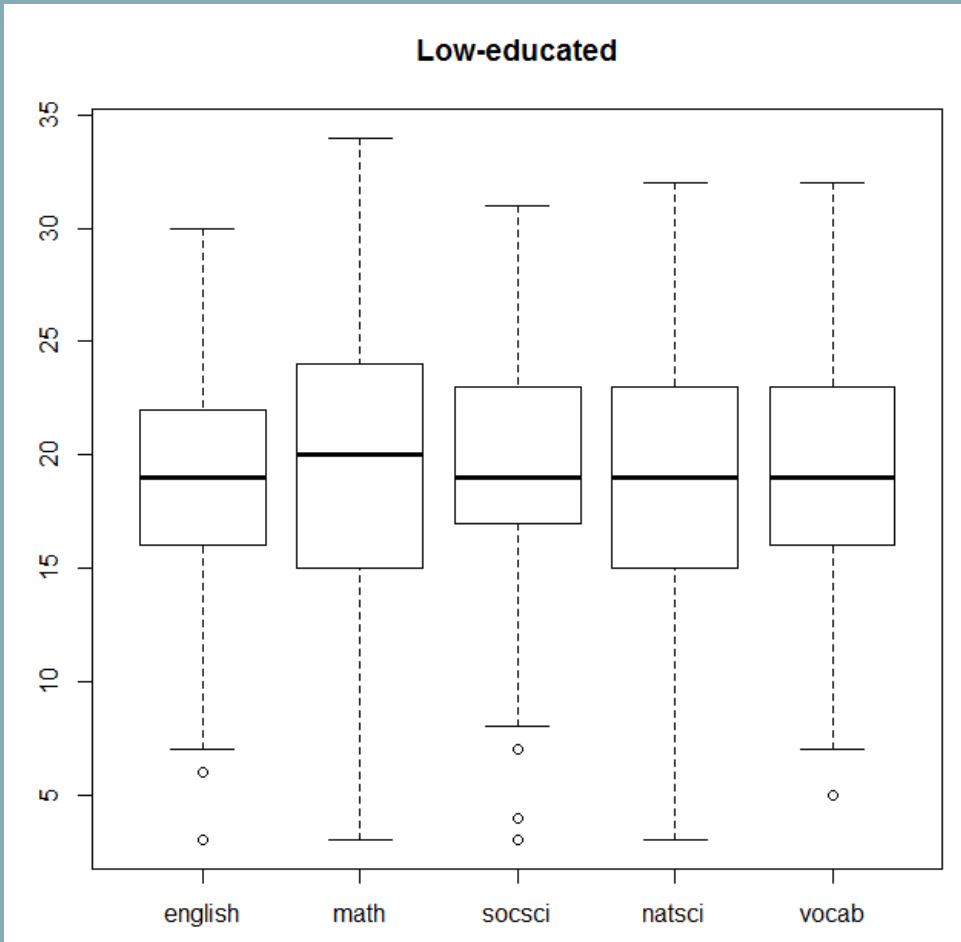
5='college grad'

6='graduate degree';

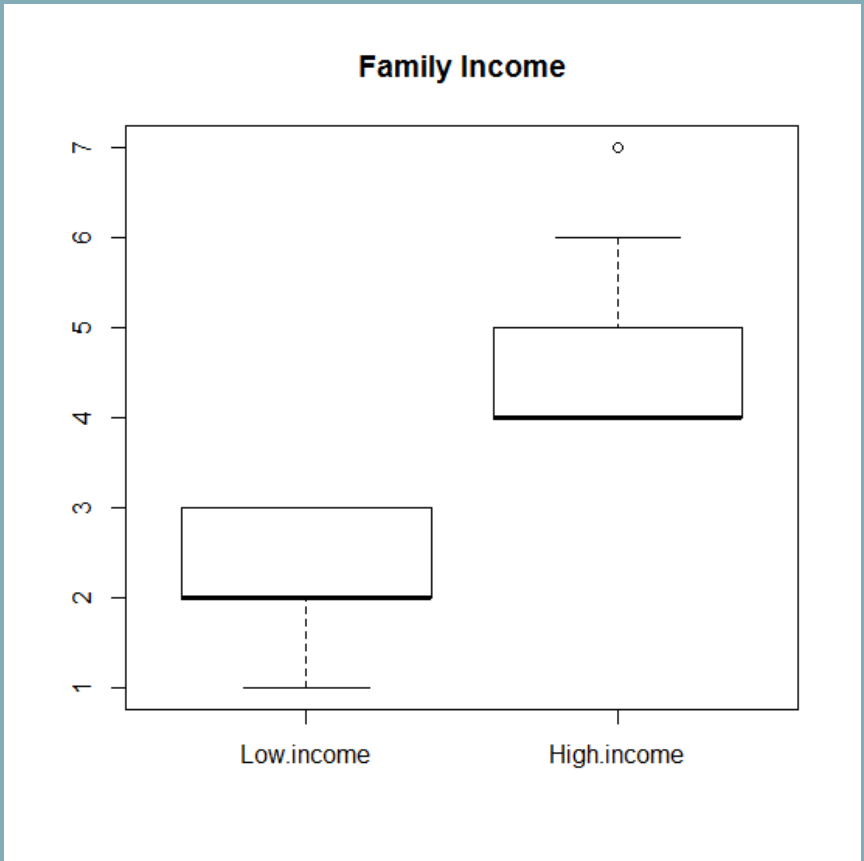


# Cluster Analysis-Education Level



# Cluster Analysis-Family Income

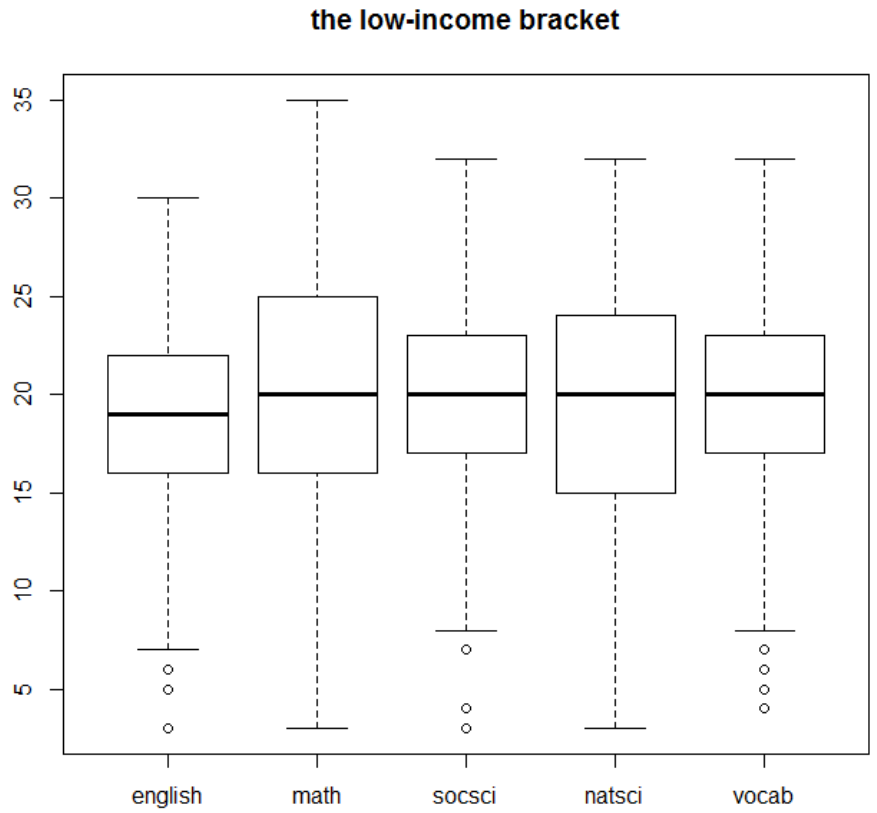


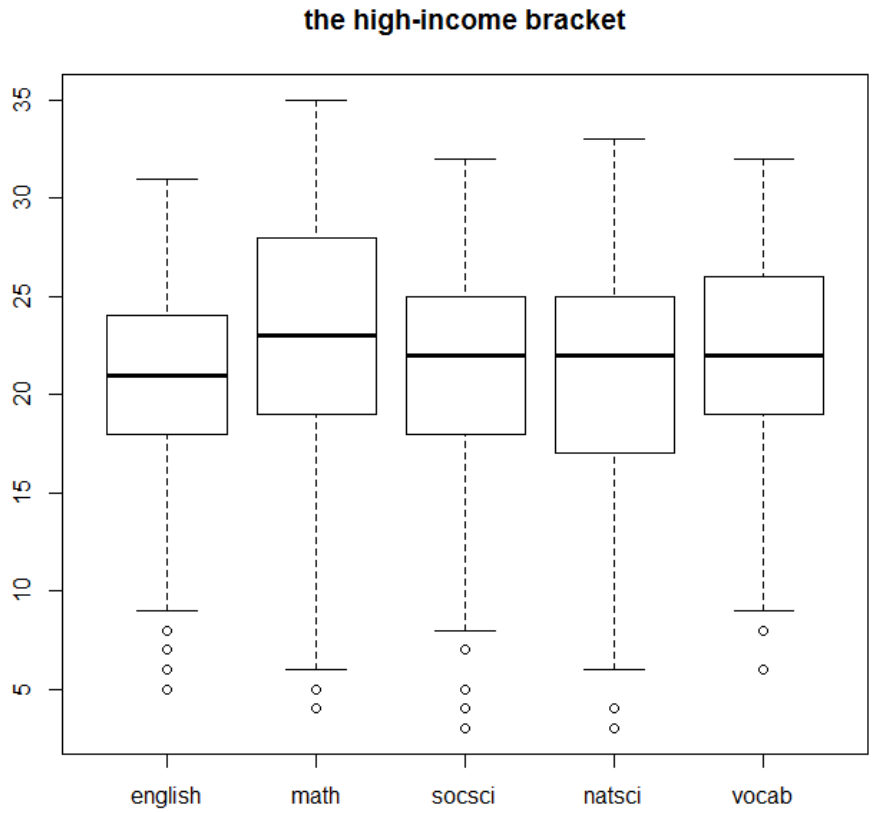

```
value incfmt
    1='< $5000'
    2='$5000 to $7499'
    3='$7500 to $9999'
    4='$10000 to $14999'
    5='$15000 to $19999'
    6='$20000 to $24999'
    7='>= $25000';
```

Inc1

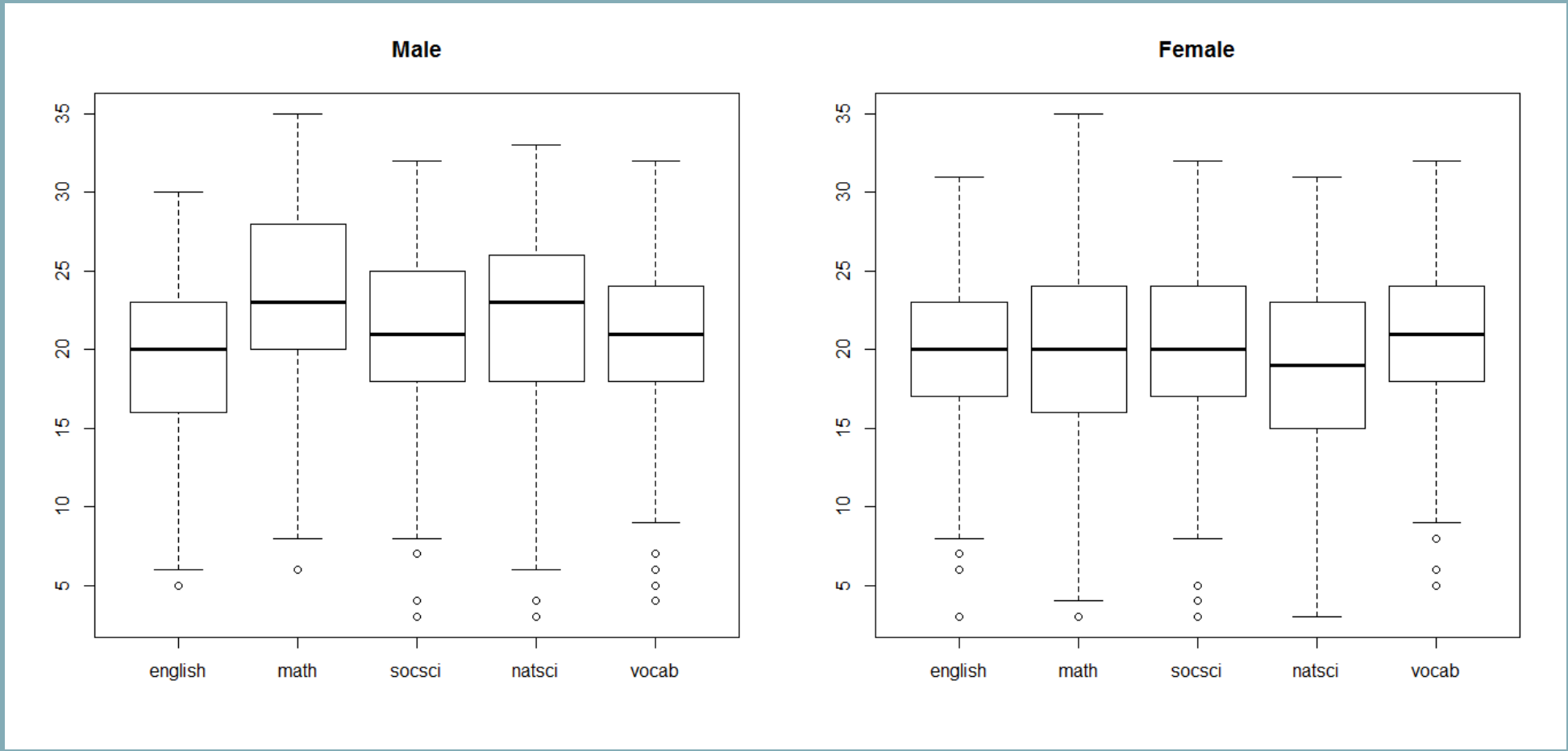
Inc2

# Cluster Analysis-Family Income



# Sexual Education Level



# Correlation by Sex

## Male

```
> cor(cbind(male[,4:6],english))
             moed      faed    faminc   english
moed     1.0000000 0.5508148 0.4437020 0.1915980
faed     0.5508148 1.0000000 0.6370850 0.2290858
faminc   0.4437020 0.6370850 1.0000000 0.2177379
english  0.1915980 0.2290858 0.2177379 1.0000000
> cor(cbind(male[,4:6],math))
             moed      faed    faminc      math
moed     1.0000000 0.5508148 0.443702 0.2103076
faed     0.5508148 1.0000000 0.637085 0.2797060
faminc   0.4437020 0.6370850 1.000000 0.2963390
math     0.2103076 0.2797060 0.296339 1.0000000
> cor(cbind(male[,4:6],socsci))
             moed      faed    faminc    socsci
moed     1.0000000 0.5508148 0.4437020 0.1603984
faed     0.5508148 1.0000000 0.6370850 0.2158116
faminc   0.4437020 0.6370850 1.0000000 0.2243527
socsci   0.1603984 0.2158116 0.2243527 1.0000000
> cor(cbind(male[,4:6],natsci))
             moed      faed    faminc    natsci
moed     1.0000000 0.5508148 0.4437020 0.1884909
faed     0.5508148 1.0000000 0.6370850 0.2037480
faminc   0.4437020 0.6370850 1.0000000 0.2262221
natsci   0.1884909 0.2037480 0.2262221 1.0000000
> cor(cbind(male[,4:6],vocab))
             moed      faed    faminc     vocab
moed     1.0000000 0.5508148 0.4437020 0.2443759
faed     0.5508148 1.0000000 0.6370850 0.2770602
faminc   0.4437020 0.6370850 1.0000000 0.3107039
vocab    0.2443759 0.2770602 0.3107039 1.0000000
```

› Family income

## Female

```
> cor(cbind(female[,4:6],english))
             moed      faed    faminc   english
moed     1.0000000 0.5640613 0.3764427 0.1403705
faed     0.5640613 1.0000000 0.4511499 0.2318194
faminc   0.3764427 0.4511499 1.0000000 0.1650638
english  0.1403705 0.2318194 0.1650638 1.0000000
> cor(cbind(female[,4:6],math))
             moed      faed    faminc      math
moed     1.0000000 0.5640613 0.3764427 0.1863463
faed     0.5640613 1.0000000 0.4511499 0.2817694
faminc   0.3764427 0.4511499 1.0000000 0.2364141
math     0.1863463 0.2817694 0.2364141 1.0000000
> cor(cbind(female[,4:6],socsci))
             moed      faed    faminc    socsci
moed     1.0000000 0.5640613 0.3764427 0.1657139
faed     0.5640613 1.0000000 0.4511499 0.2780162
faminc   0.3764427 0.4511499 1.0000000 0.1913037
socsci   0.1657139 0.2780162 0.1913037 1.0000000
> cor(cbind(female[,4:6],natsci))
             moed      faed    faminc    natsci
moed     1.0000000 0.5640613 0.3764427 0.1662560
faed     0.5640613 1.0000000 0.4511499 0.1816899
faminc   0.3764427 0.4511499 1.0000000 0.1528937
natsci   0.1662560 0.1816899 0.1528937 1.0000000
> cor(cbind(female[,4:6],vocab))
             moed      faed    faminc     vocab
moed     1.0000000 0.5640613 0.3764427 0.2064462
faed     0.5640613 1.0000000 0.4511499 0.3182588
faminc   0.3764427 0.4511499 1.0000000 0.2245227
vocab    0.2064462 0.3182588 0.2245227 1.0000000
```

› Father's education level

# Hotelling's $T^2$-test

```
> twins1<-subset(twins,even==FALSE)
> twins2<-subset(twins,even==TRUE)
> twins.diff<-twins1-twins2
> HotellingsT2(twins.diff[,7:11])

        Hotelling's one sample T2-test

data:  twins.diff[, 7:11]
T.2 = 2.5559, df1 = 5, df2 = 763, p-value = 0.02637
alternative hypothesis: true location is not equal to c(0,0,0,0,0)
```

**Reject H0**
**→ There is a difference between twins' grades**

# Conclusion

## Liberal Arts / Science

- Cannot divide into liberal arts and science

## High / Medium / Low Grade

- The higher grade, the higher level of factors (especially in high grade group)

## Well / Low Parents' Educational Level

- The higher Ed level, the higher overall grade (especially in math)

# Conclusion

## High / Low Income

- The higher income level, the higher overall grade (except natsci) (especially in math)

## Male / Female

- Male's grade of math and natsci is higher then female's
- Faminc is the biggest factor in male
- Faed is the biggest facor in female

## Difference between Twins

- Hoteling's $T^2$-test result
  : reject the H0
  → difference exists

# Thank You!