

# MSMS 105 : Practical

Ananda Biswas

25 November, 2024

## ➡ Question

A class of 15 students has their test scores recorded in four subjects - Math, Science, English and History. Analyze the pairwise Pearson Correlation coefficients between these subjects to determine the relationships between them. Present the data and results of the analysis.

Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Math	67	88	77	92	81	76	89	72	95	68	74	85	80	90	79
Science	93	63	96	85	72	88	78	91	84	75	89	80	87	94	76
English	83	87	85	92	76	88	82	74	86	81	79	84	90	77	80
History	87	85	75	80	79	86	83	82	91	88	77	90	78	84	89

## ➡ R Program, Plot and Interpretation

```
df <- read.csv('https://raw.githubusercontent.com/sakunisgithub/data_sets/refs/heads/master/test_scores.csv')
```

```
my_mean <- function(vec){  
  s <- 0  
  for (i in 1:length(vec)) {  
    s <- s + vec[i]  
  }  
  return(s/length(vec))  
}
```

```
my_cov <- function(vec1, vec2){  
  s <- 0  
  for (i in 1:length(vec1)) {  
    s <- s + (vec1[i] - my_mean(vec1)) * (vec2[i] - my_mean(vec2))  
  }  
  return(s/length(vec1))  
}
```

```
my_var <- function(vec){
  s <- 0
  for (i in 1:length(vec)) {
    s <- s + (vec[i] - my_mean(vec))^2
  }
  return(s/length(vec))
}
```

```
my_corr <- function(vec1, vec2){
  temp <- my_cov(vec1, vec2) / sqrt(my_var(vec1) * my_var(vec2))
  return(temp)
}
```

```
for (i in 1:4) {
  temp <- paste("Average score in", names(df)[i+1], "is", my_mean(df[,i+1]), sep = " ")
  print(temp)
}
```

```
## [1] "Average score in Mathematics is 80.9333333333333"
## [1] "Average score in Science is 83.4"
## [1] "Average score in English is 82.9333333333333"
## [1] "Average score in History is 83.6"
```

```
for (i in 2:4) {
  for (j in (i+1):5) {
    temp <- paste("Correlation between", names(df)[i], "and", names(df)[j], "is",
                  my_corr(df[,i], df[,j]), sep = " ")
    print(temp)
  }
}
```

```
## [1] "Correlation between Mathematics and Science is -0.251550054947875"
## [1] "Correlation between Mathematics and English is 0.333064725095778"
## [1] "Correlation between Mathematics and History is 0.125201003384908"
## [1] "Correlation between Science and English is -0.0378386398631776"
## [1] "Correlation between Science and History is -0.312589056194899"
## [1] "Correlation between English and History is 0.00986489121169494"
```

```
library(tidyverse)
```

```
library(ggcorrplot)

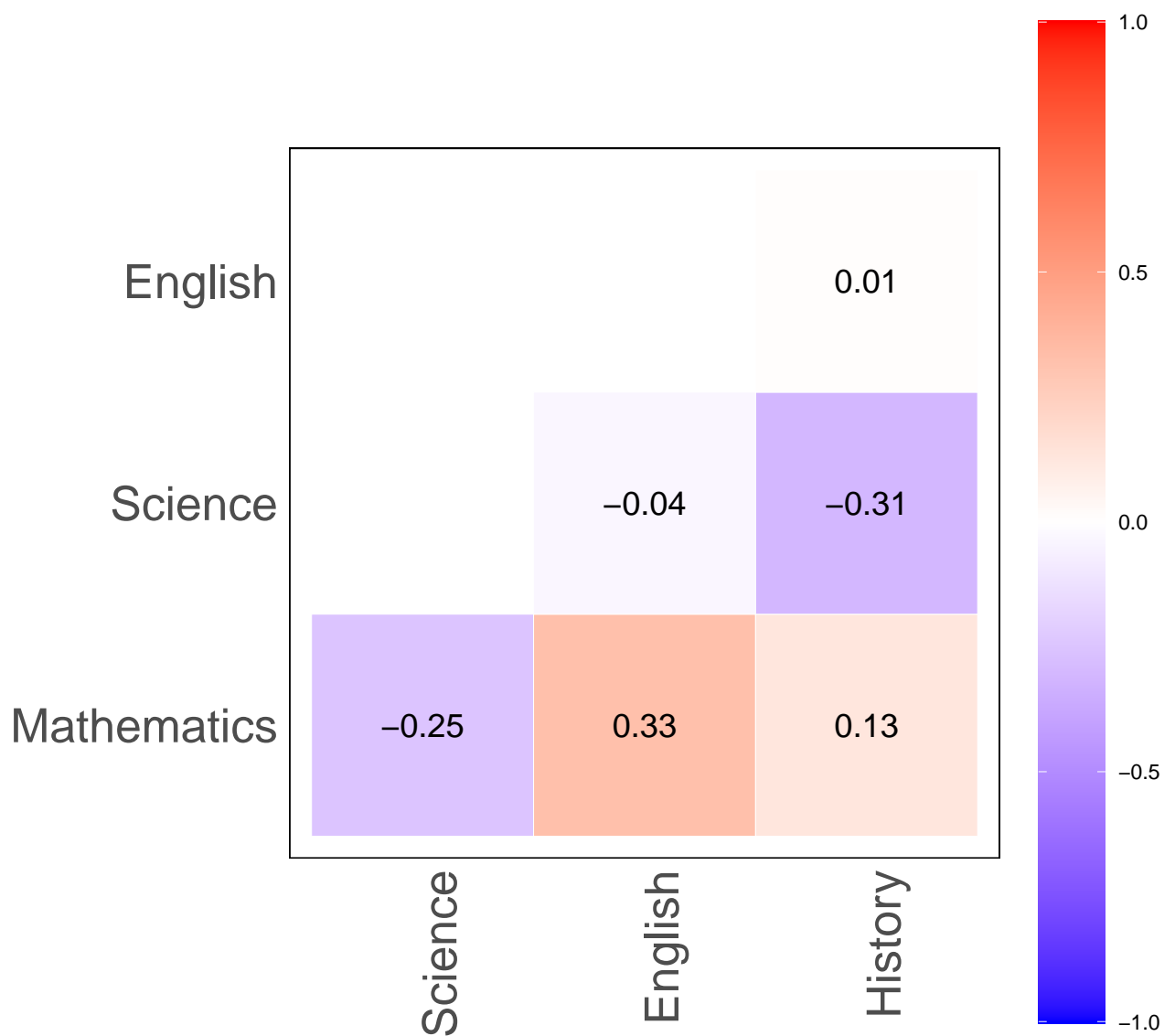
cor_df <- cor(df[,2:5])

ggcorrplot(cor_df,
            method = "square",
```

```

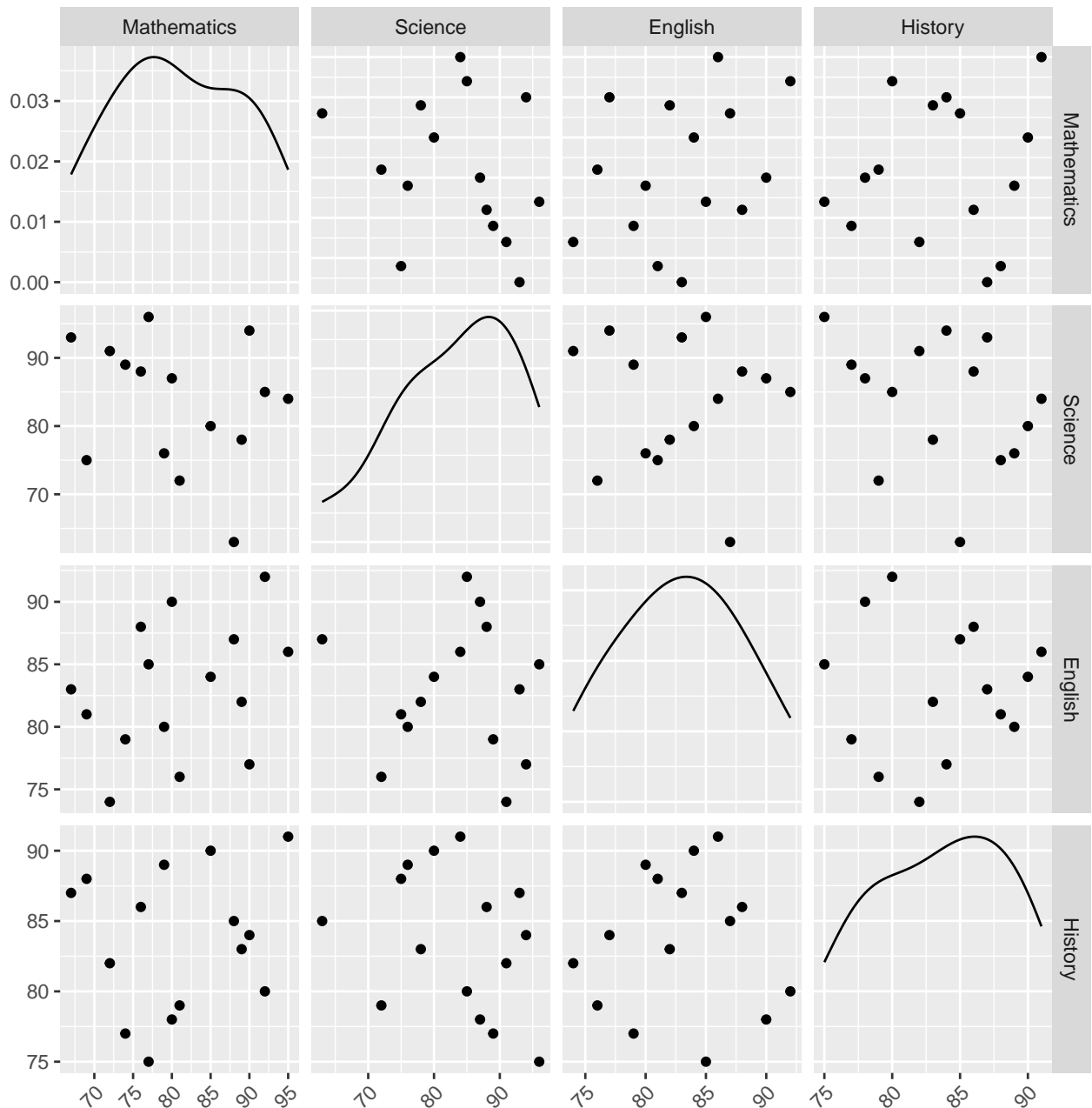
    type = "lower",
    lab = TRUE,
    lab_col = "black",
    lab_size = 5,
    outline.color = "white",
    tl.cex = 20,
    legend.title = "") +
theme(axis.text.x = element_text(angle = 90, hjust = 1),
      panel.background = element_rect(fill = "white"),
      panel.grid.major = element_blank(),
      panel.border = element_blank(),
      legend.key.height = unit(3, 'cm'),
      legend.key.width = unit(1, 'cm'))

```



```
library(GGally)
```

```
ggpairs(df,
  columns = 2:5,
  upper = list(continuous = "points", legend = c(1,1)) +
  scale_alpha(guide = "none") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



## ➡ Conclusion

 No significant linear or curvi-linear pattern is found in the scatterplots of the scores of different subjects.