G5243 Project1

wz2563

2022-09-20

```
philosopy <- read.csv("philosophy_data.csv")
```

First take look at the composition of school

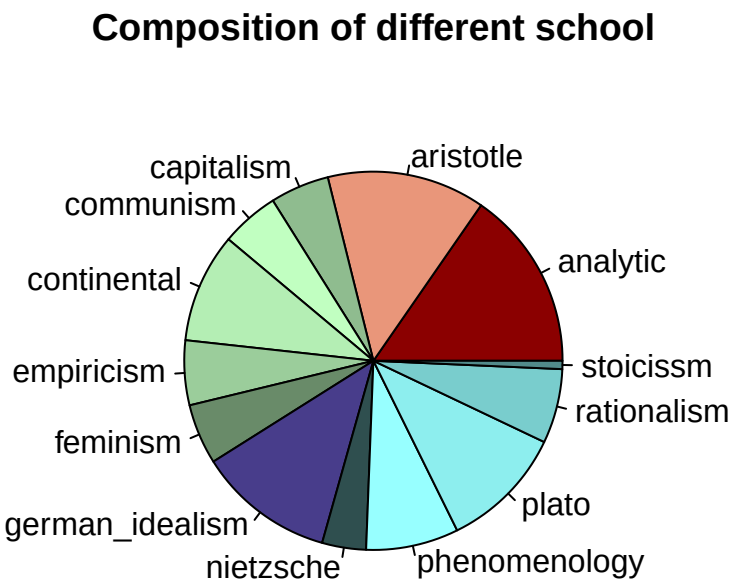
```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
school <- group_by(philosopy,school)%>%
  summarise(count=n())

school_list <- c("analytic","aristotle","capitalism","communism","continental","empiricism","feminism",

school_pie <- pie(school$count,
                  main="Composition of different school",
                  labels = school_list,
                  col=colors()[100:120])
```



From the plot we can see that 'aristotle' & 'analytic' represent the main strem, which is the top 2 school

Next i wanna analyse the emotion in different school type:

```
analytic <- philosopy[philosopy$school=="analytic",]$sentence_str
aristotle <- philosopy[philosopy$school=="aristotle",]$sentence_str
capitalism <- philosopy[philosopy$school=="capitalism",]$sentence_str
communism <- philosopy[philosopy$school=="communism",]$sentence_str
continental <- philosopy[philosopy$school=="continental",]$sentence_str
empiricism <- philosopy[philosopy$school=="empiricism",]$sentence_str
feminism <- philosopy[philosopy$school=="feminism",]$sentence_str
german_idealism <- philosopy[philosopy$school=="german_idealism",]$sentence_str
nietzsche <- philosopy[philosopy$school=="nietzsche",]$sentence_str
phenomenology <- philosopy[philosopy$school=="phenomenology",]$sentence_str
plato <- philosopy[philosopy$school=="plato",]$sentence_str
rationalism <- philosopy[philosopy$school=="rationalism",]$sentence_str
stoicissm <- philosopy[philosopy$school=="stoicism",]$sentence_str
```

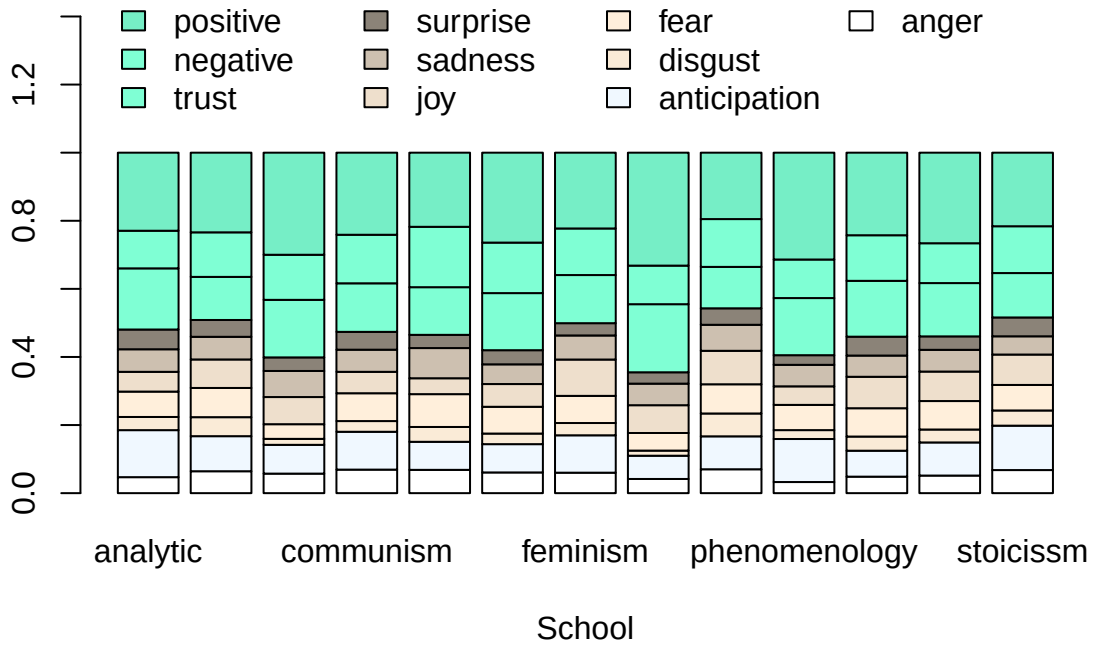
```
library(syuzhet)
```

```
## Warning: package 'syuzhet' was built under R version 4.1.3
```

```
emotion <- as.data.frame(colSums(get_nrc_sentiment(sample(analytic,100,replace = TRUE))),)

emotions <- as.data.frame(colSums(get_nrc_sentiment(sample(analytic,100,replace = TRUE))))%>%
  cbind(colSums(get_nrc_sentiment(sample(aristotle,100,replace = TRUE))))%>%
```

```
  cbind(colSums(get_nrc_sentiment(sample(capitalism,100,replace = TRUE))))%>%
  cbind(colSums(get_nrc_sentiment(sample(communism,100,replace = TRUE))))%>%
  cbind(colSums(get_nrc_sentiment(sample(continental,100,replace = TRUE))))%>%
  cbind(colSums(get_nrc_sentiment(sample(empiricism,100,replace = TRUE))))%>%
  cbind(colSums(get_nrc_sentiment(sample(feminism,100,replace = TRUE))))%>%
  cbind(colSums(get_nrc_sentiment(sample(german_idealism,100,replace = TRUE))))%>%
  cbind(colSums(get_nrc_sentiment(sample(nietzsche,100,replace = TRUE))))%>%
  cbind(colSums(get_nrc_sentiment(sample(phenomenology,100,replace = TRUE))))%>%
  cbind(colSums(get_nrc_sentiment(sample(plato,100,replace = TRUE))))%>%
  cbind(colSums(get_nrc_sentiment(sample(rationalism,100,replace = TRUE))))%>%
  cbind(colSums(get_nrc_sentiment(sample(stoicissm,100,replace = TRUE))))


colnames(emotions) <- c("analytic","aristotle","capitalism","communism","continental","empiricism","femi

emotions <- data.frame(emotions)

prop_emotion <- prop.table(data.matrix(emotions),2)
```

```
library(ggplot2)

stacked_emotions <- barplot(prop_emotion,
                            xlab = "School",
                            ylim = c(0,1.5),
                            col = colors()[1:10],
                            legend=TRUE,
                            args.legend = list(bty = "n",x = "top", ncol = 4))
```

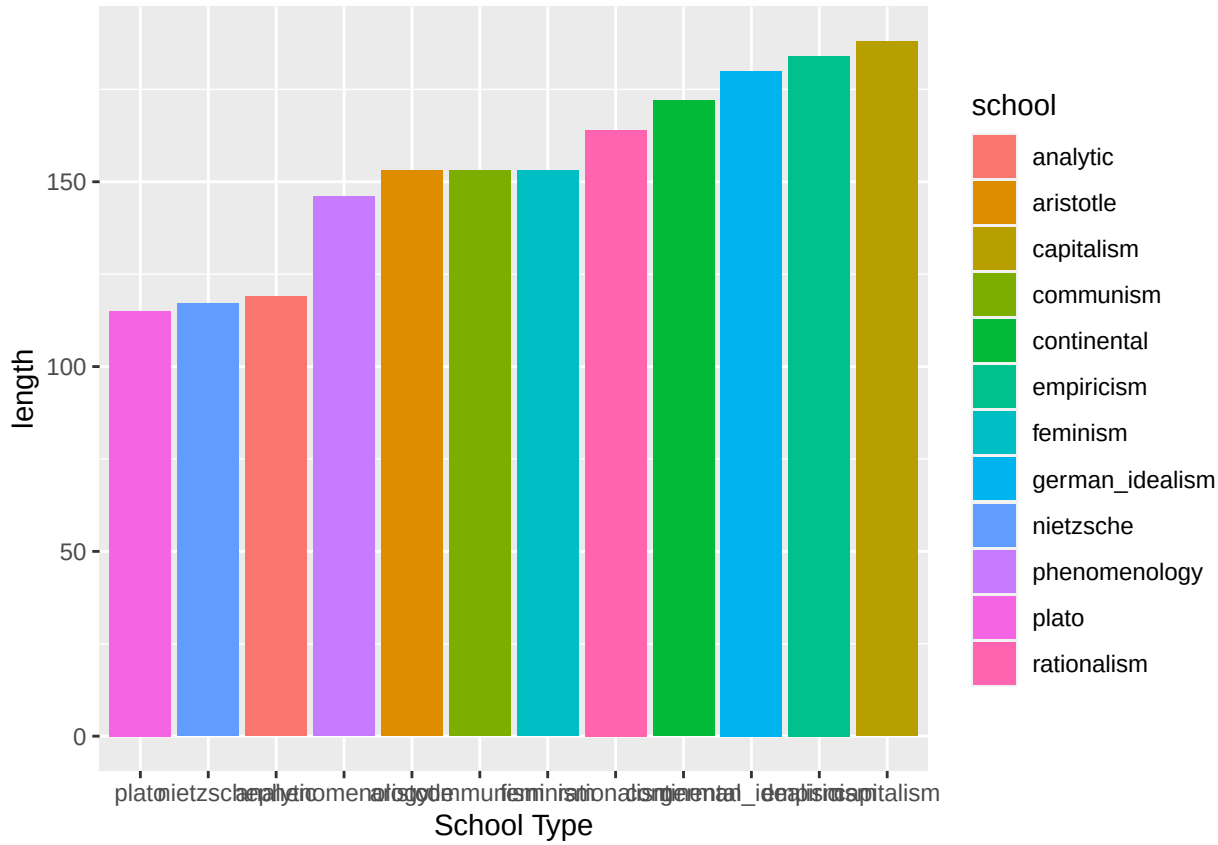


From the plot we can see the different types of composition of emotions in each school. Mostly includes positive,negative,trust emotion.

Next I wanna explore the XXX of each school

```
library(tidyr)
length_of_school <- data.frame()

for (i in 1:13){
  length_of_school <- rbind(length_of_school,c(school_list[i],mean(nchar(philosopy[philosopy$school==sc
}

colnames(length_of_school) <- c("school","length")

length_of_school <- length_of_school[1:12,]

length_of_school$length <- round(as.numeric(length_of_school$length,2))
```

```
library(forcats)
bar_length <- ggplot(length_of_school,aes(fct_reorder(school,length),length,fill=school))+
  geom_bar(stat="identity")+
  labs(x="School Type")
bar_length
```



From the statistic we can see that 'capitalism' has the most words in sentence, which may seem the most verbose. While 'plato' has the shortest sentence in average.