

5243 Project 1 - Zhejing Shi

The Brief Introduction to Capitalism and Communism.



Capitalism is defined as an economic system in which a country's trade, industry, and profits are controlled only by private companies. The United States and most of the nations around the world are capitalist countries, but capitalism is not the only economic system available. Communism, on the other hand, advocating that goods are owned in common and are available to all as needed, means of production are owned communally.

**This report aims to 1.)help people understand the basic concepts of communism.
2.)explore insights between capitalism and communism.**

```
library(RColorBrewer)
library(tm)
library(wordcloud)
library(NLP)
library(tidyverse)
library(syuzhet)
```

[Hide](#)

```
##### Separate data sets into: capitalism, communism, Communist_Manifesto, lenin, Capital
docs = read.csv("philosophy_data.csv",stringsAsFactors = F)
capitalism =
  docs %>% filter(school == "capitalism")
communism =
  docs %>% filter(school == "communism")
Communist_Manifesto =
  docs %>% filter(title == "The Communist Manifesto")
lenin =
  docs %>% filter(title == "Essential Works Of Lenin")
Capital =
  docs %>% filter(title == "Capital")
```

1. Scratch the Surface of Capitalism

[Hide](#)

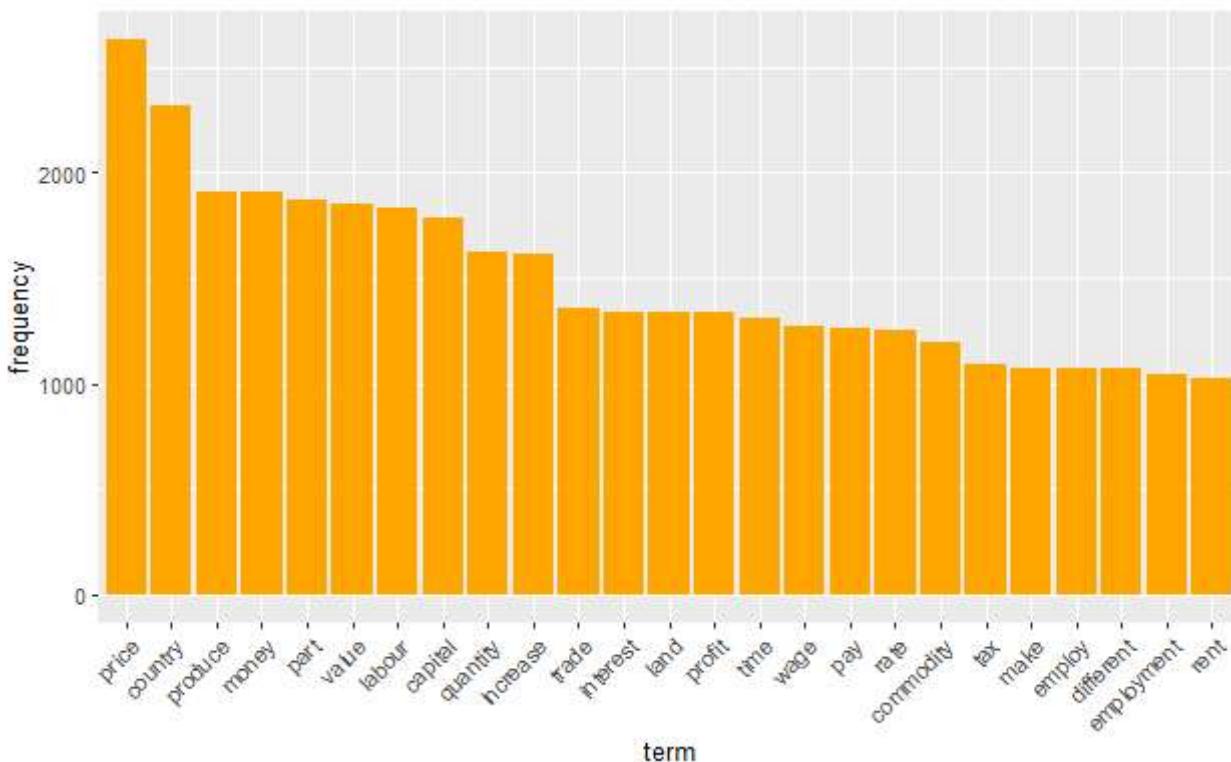
```
#####
##### Capitalism #####
#####

## data processing
words = Corpus(VectorSource(capitalism$lemmatized_str))
words = words %>%
  tm_map( removeNumbers)%>%
  tm_map( removePunctuation)%>%
  tm_map( removeWords, c("the", "and", "PRON", stopwords("english")) )%>% # Remove PRON
  tm_map( removeWords, c("must", "however", "two", "one", "upon", "much", "every", "either", "mu-
st", "can", "may", "therefore", "will", "good", "great", stopwords("english")))%>%
  tm_map( stripWhitespace)

### Document matrix 1
words_dtm <- DocumentTermMatrix(words)
### Document matrix 2 based on [TF-IDF] which measures the relative importance of a word.
words_dtm_tfidf <- DocumentTermMatrix(words, control = list(weighting = weightTfIdf))
words_dtm_tfidf = removeSparseTerms(words_dtm_tfidf, 0.95)

## creating freq data frame
freq = data.frame(sort(colSums(as.matrix(words_dtm)), decreasing=TRUE))
table = data.frame(term = row.names(freq), frequency= freq)
colnames(table) = c('term', 'frequency')

# Bar chart
capitalism_chart<-ggplot(data= subset(table, frequency >1000), aes(x=reorder(term, -frequency),
y=frequency)) +
  geom_bar(stat="identity", fill="orange") + theme(axis.text.x = element_text(angle = 45,hjust=1)
) + labs(x ="term")
capitalism_chart
```



```
## First word cloud
wordcloud(rownames(freq), freq[,1], max.words=50, colors=brewer.pal(1, "Dark2"))
```



```
### Second Word Cloud based on [TF-IDF] which measures the relative importance of a word.
freq.capitalism = data.frame(sort(colSums(as.matrix(words_dtm_tfidf))), decreasing=TRUE)
wordcloud(rownames(freq.capitalism), freq.capitalism[,1], max.words=50, colors=brewer.pal(1, "Dark2"), scale=c(3,0.25))
```



[Hide](#)

NA
NA
NA
NA

The first frequency graph shows the words appears more than 1000 times. The second graph is a overall wordcloud, while the last graph is a wordcloud based on [TF-IDF] which measures the relative importance of a word.

Words such as capital, employment, commodity, labour, profit, money, and etc. are heavily mentionly in the School of Capitalism. Most of those words also happen to appearin the School of Communism, and we will introduce what do communists think of these terms later in this report.

As we mentioned earlier, most of the nations in the world are capilist. However, it is worthy to get to known other systems too. In order to better understand communism, which aspect should we pay attention to? We will start with some of the most frequent termnologies appears in communist studies.

2. Take a Overview of School of Communism

```
#####
##### Communism: #####
#####

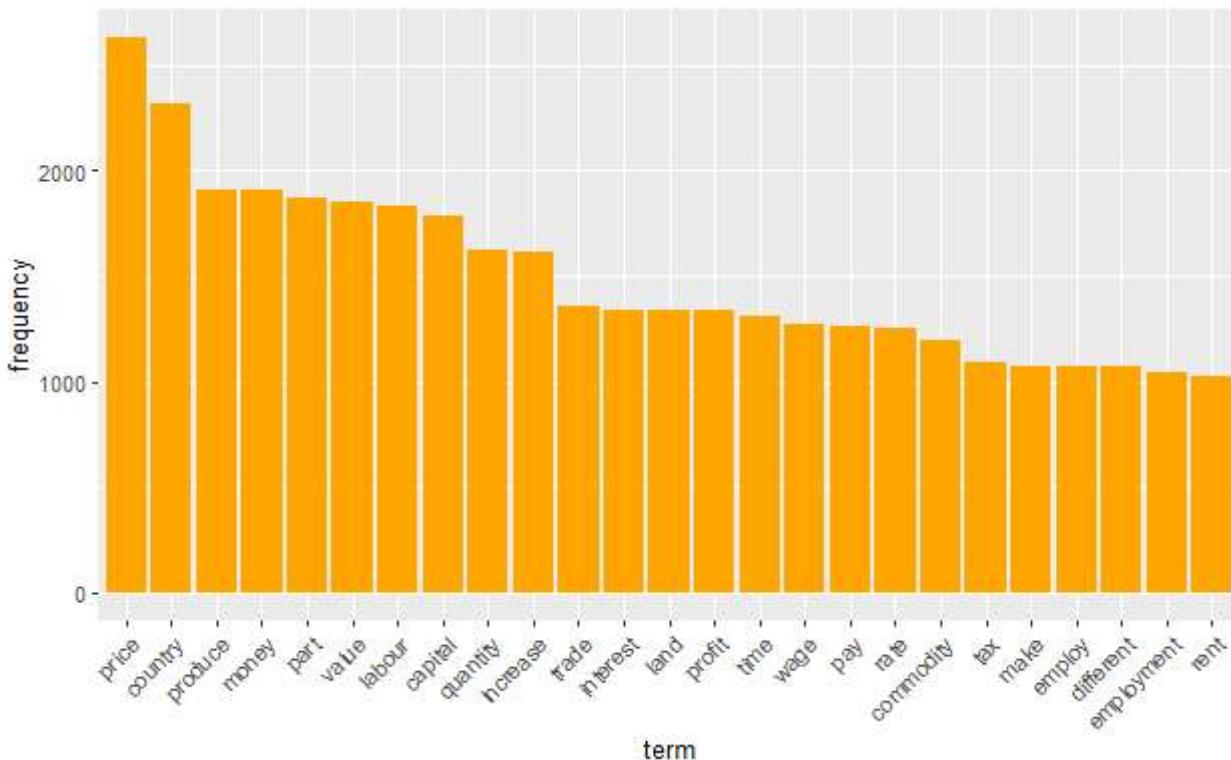
## data processing
words = Corpus(VectorSource(capitalism$lemmatized_str))
words = tm_map(words, removeNumbers)
words = tm_map(words, removePunctuation)
words = tm_map(words, removeWords, c("the", "and", "PRON", stopwords("english"))) # Remove PRON
words = tm_map(words, removeWords, c("must", "however", "two", "one", "upon", "much", "every",
"either", "must", "can", "may", "therefore", "will", "good", "great", stopwords("english")))
words = tm_map(words, removeWords, c("even", "always", "whole", stopwords("english")))
words = tm_map(words, stripWhitespace)

### Document matrix1
words_dtm <- DocumentTermMatrix(words)

### Document matrix 2 based on [TF-IDF] which measures the relative importance of a word.
words_dtm_tfidf <- DocumentTermMatrix(words, control = list(weighting = weightTfIdf))
words_dtm_tfidf = removeSparseTerms(words_dtm_tfidf, 0.95)

## creating freq data frame
freq = data.frame(sort(colSums(as.matrix(words_dtm)), decreasing=TRUE))
table = data.frame(term = row.names(freq), frequency= freq)
colnames(table) = c('term', 'frequency')

# Bar chart
communism_chart<-ggplot(data= subset(table, frequency >1000), aes(x=reorder(term, -frequency), y
=frequency)) +
  geom_bar(stat="identity", fill="orange") + theme(axis.text.x = element_text(angle = 45,hjust=1)
) + labs(x ="term")
communism_chart
```



Hide

```
## First word cloud  
wordcloud(rownames(freq), freq[,1], max.words=50, colors=brewer.pal(1, "Dark2"))
```



Hide

```
### Second Word Cloud, based on [TF-IDF] which measures the relative importance of a word.  
freq.communism = data.frame(sort(colSums(as.matrix(words_dtm_tfidf))), decreasing=TRUE))  
wordcloud(rownames(freq.communism), freq.communism[,1], max.words=100, colors=brewer.pal(1, "Dark2"), scale=c(3,0.25))
```



Hide

NA
NA
NA
NA

Both frequency bar chart and wordcloud graphs reveal us some frequent terms such as Land, Capital, Money, commodity. In order to understand Communism, it is important to ask the following question.

Question: what does communism think of these terms?

- Money: There is no government or private property or currency, and the wealth is divided among citizens equally or according to individual need
- Capital: Marxists define capital as “a social, economic relation” between people (rather than between people and things). In this sense they seek to abolish capital.
- Land: Under communism, land is seen as property. All property is communally owned, and each person receives a portion based on what they need.
- Commodity: In Marx’s theory, a commodity is something that is bought and sold, or exchanged in a relationship of trade. It has value, which represents a quantity of

human labor.

- Capitalism: Karl Marx saw capitalism as a progressive historical stage that would eventually be followed by socialism. Class conflict derives in capitalism because the proletariat are exploited in the system.

Frequency Table of Book Titles

[Hide](#)

```
communism %>%
  group_by(title)%>%
  summarise(count = n())%>%
  arrange(desc(count))
```

title	count
<chr>	<int>
Capital	12996
Essential Works Of Lenin	4469
The Communist Manifesto	493
3 rows	

Above table shows number of sentences in 3 books introducing communism. Since all 3 books have different length, one overall wordcloud graph may not be representative. Thus, we need to explore more!!!

2.1 Check the First Communism Related Book: Communist_Manifesto (共产党宣言)

[Hide](#)

```
#####
##### Communist_Manifesto #####
#####

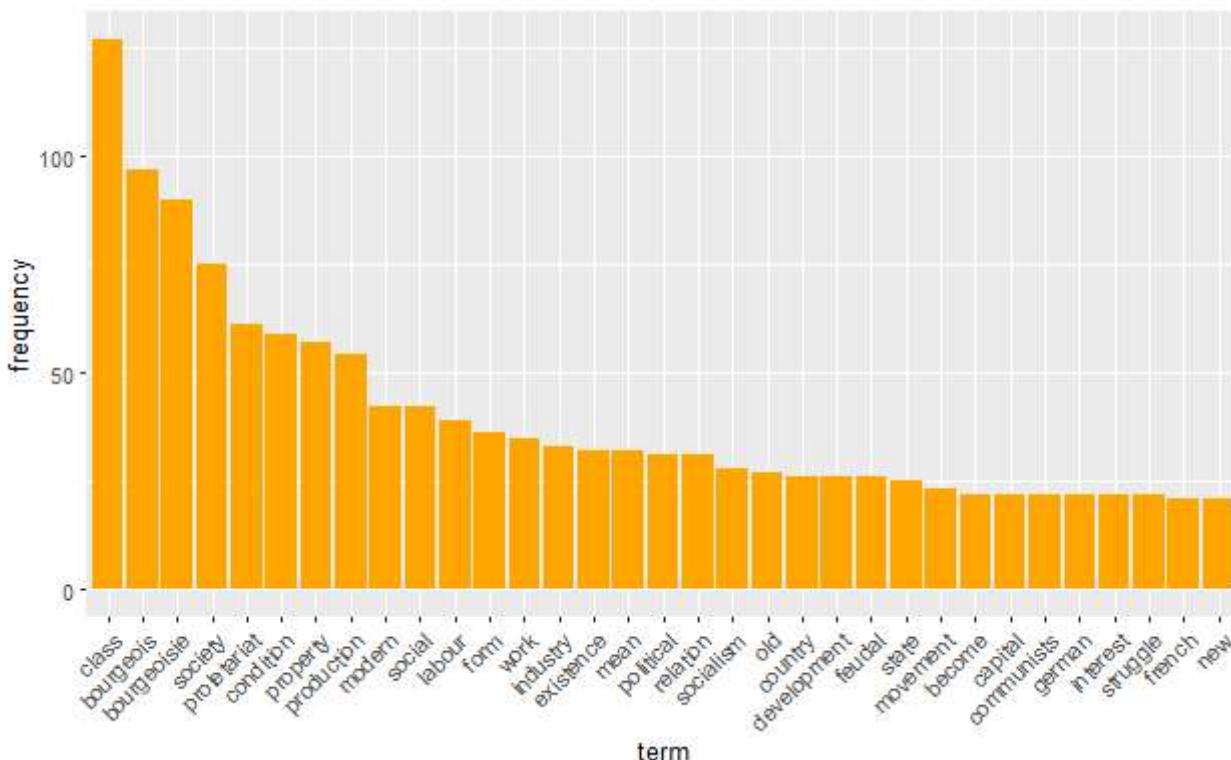
## data preprocessing
words = Corpus(VectorSource(Communist_Manifesto$lemmatized_str))
words = words %>%
  tm_map( removeNumbers)%>%
  tm_map( removePunctuation)%>%
  tm_map( removeWords, c("the", "and", "PRON", stopwords("english")) )%>% # Remove PRON
  tm_map( removeWords, c("must", "however", "two", "one", "upon", "much", "every", "either", "mu
st", "can", "may", "therefore", "will", "good", "great", stopwords("english")))%>%
  tm_map(removeWords, c("even", "always", "whole", stopwords("english")))%>%
  tm_map(stripWhitespace)

### Document matrix
words_dtm <- DocumentTermMatrix(words)

### Document matrix 2 based on [TF-IDF] which measures the relative importance of a word.
words_dtm_tfidf <- DocumentTermMatrix(words, control = list(weighting = weightTfIdf))
words_dtm_tfidf = removeSparseTerms(words_dtm_tfidf, 0.95)
words_dtm_tfidf

## creating freq data frame
freq = data.frame(sort(colSums(as.matrix(words_dtm)), decreasing=TRUE))
table = data.frame(term = row.names(freq), frequency= freq)
colnames(table) = c('term', 'frequency')

# Bar chart
communism_chart<-ggplot(data= subset(table, frequency >20), aes(x=reorder(term, -frequency), y=f
requency)) +
  geom_bar(stat="identity", fill="orange") + theme(axis.text.x = element_text(angle = 45,hjust=1)
) + labs(x ="term")
communism_chart
```



Hide

```
## First word cloud  
wordcloud(rownames(freq), freq[,1], max.words=50, colors=brewer.pal(1, "Dark2"))
```



Hide

```
### Second Word Cloud based on [TF-IDF] which measures the relative importance of a word.  
freq.Manifesto = data.frame(sort(colSums(as.matrix(words_dtm_tfidf))), decreasing=TRUE)  
wordcloud(rownames(freq.Manifesto), freq.Manifesto[,1], max.words=100, colors=brewer.pal(1, "Dark2"), scale=c(3,0.25))
```



Besides what we discussed previously, Communist Manifesto revealed us some other frequent words such as [bourgeois, proletariat, class]. In order to make everyone better prepared to read this book. The following terminologies are worth knowing.

- bourgeois: bourgeoisie, the social order that is dominated by the so-called middle class. In Marxist theory, the bourgeoisie plays a heroic role by revolutionizing industry and modernizing society. However, it also seeks to monopolize the benefits of this modernization by exploiting the propertyless proletariat
- proletariat: the lowest or one of the lowest economic and social classes in a society. The proletariat is that class in society which lives entirely from the sale of its labor and does not draw profit from any kind of capital
- class: The two classes are the proletariat (the working class), who make up the majority of the population within society and must work to survive, and the bourgeoisie (the capitalist class), a small minority who derives profit from employing the working class—(Wikipedia)

2.2 Check the Second Communism Related Book:Essential Works of Lenin

[Hide](#)

```
#####
##### Lenin #####
#####

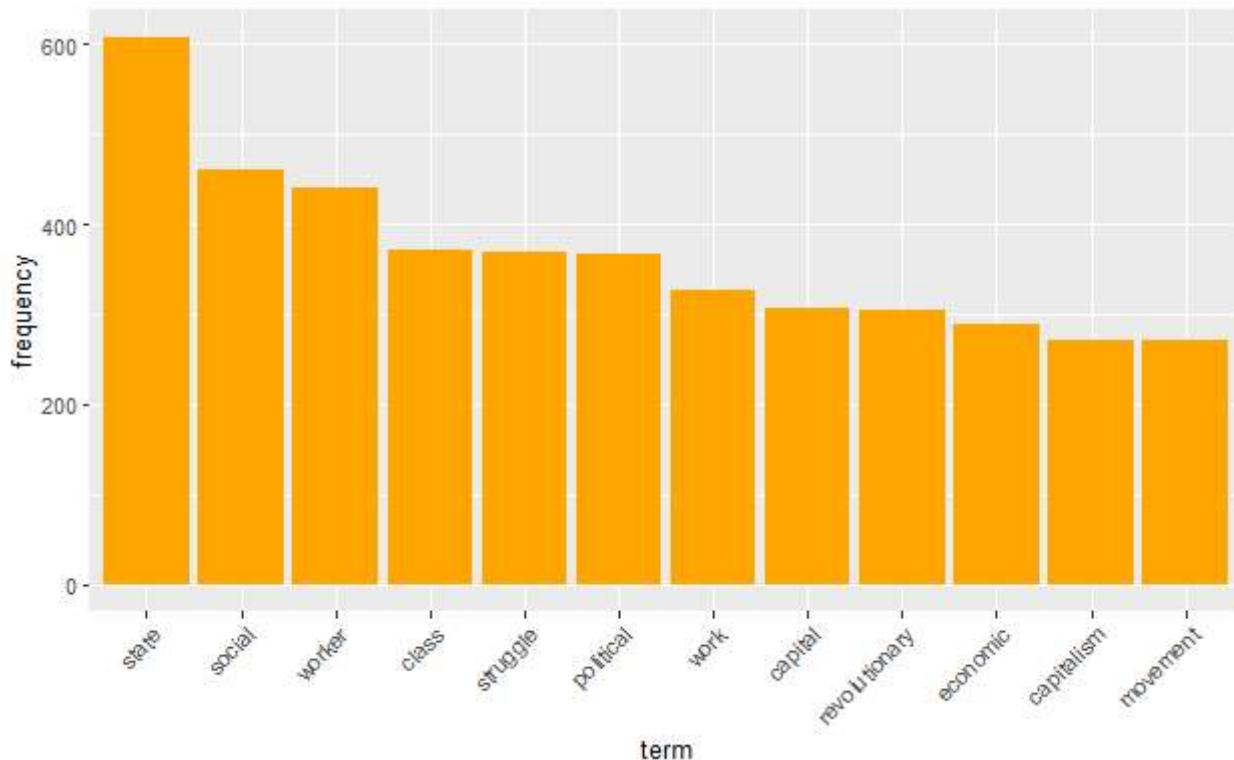
## data processing
words = Corpus(VectorSource(lenin$lemmatized_str))
words = words %>%
  tm_map( removeNumbers)%>%
  tm_map( removePunctuation)%>%
  tm_map( removeWords, c("the", "and", "PRON", stopwords("english")) )%>% # Remove PRON
  tm_map( removeWords, c("must", "however", "two", "one", "upon", "much", "every", "either", "must", "can", "may", "therefore", "will", "good", "great", stopwords("english")))%>%
  tm_map(removeWords, c("even", "always", "whole", stopwords("english")))%>%
  tm_map(stripWhitespace)

### Document matrix
words_dtm <- DocumentTermMatrix(words)

### Document matrix 2 based on [TF-IDF] which measures the relative importance of a word.
words_dtm_tfidf <- DocumentTermMatrix(words, control = list(weighting = weightTfIdf))
words_dtm_tfidf = removeSparseTerms(words_dtm_tfidf, 0.95)

## creating freq data frame
freq = data.frame(sort(colSums(as.matrix(words_dtm)), decreasing=TRUE))
table = data.frame(term = row.names(freq), frequency= freq)
colnames(table) = c('term', 'frequency')

# Bar chart
lenin_chart<-ggplot(data= subset(table, frequency >270), aes(x=reorder(term, -frequency), y=freq))
+ geom_bar(stat="identity", fill="orange")+
  theme(axis.text.x = element_text(angle = 45,hjust=1))
+ labs(x ="term")
lenin_chart
```

[Hide](#)

```
## First word cloud
wordcloud(rownames(freq), freq[,1], max.words=50, colors=brewer.pal(1, "Dark2"))
```

[Hide](#)

```
### Second Word Cloud based on [TF-IDF] which measures the relative importance of a word.
freq.lenin = data.frame(sort(colSums(as.matrix(words_dtm_tfidf))), decreasing=TRUE)
wordcloud(rownames(freq.lenin), freq.lenin[,1], max.words=100, colors=brewer.pal(1, "Dark2"))
```



Essential Works Of Lenin revealed us some other frequent words such as [imperialism, monopoly, revolutionary]. In order to make everyone better prepared to read this book. The following terminologies are worth knowing.

- Imperialism: Lenin thought that imperialism was the “highest stage” of capitalism. He says that modern imperialism is different from ancient and medieval imperialism. Modern imperialism is mainly subject to economic considerations which is a method of exploitation.
- Monopoly: A monopoly is defined as a single seller or producer that excludes competition from providing the same product. A monopoly can dictate price changes and creates barriers for competitors to enter the marketplace. Communists oppose monopoly.
- Work class: There are two classes, proletariat (the working class) and capital class. Communists concern the struggle of the working class
- revolution: Lenin exhorted all oppressed nations of the world to unite against their oppressors and play the role of the revolutionary proletariat against oppressive, capitalist nations. Lenin wanted Communist political parties in every country to lead the revolution. The Russian Revolution (1917) established the modern world’s first socialist state – the Union of Socialist Soviet Republics.

2.3 Check the Third Communism Related Book: Capital: (资本论)

[Hide](#)

```
#####
##### Capital #####
#####

## data reprocessing

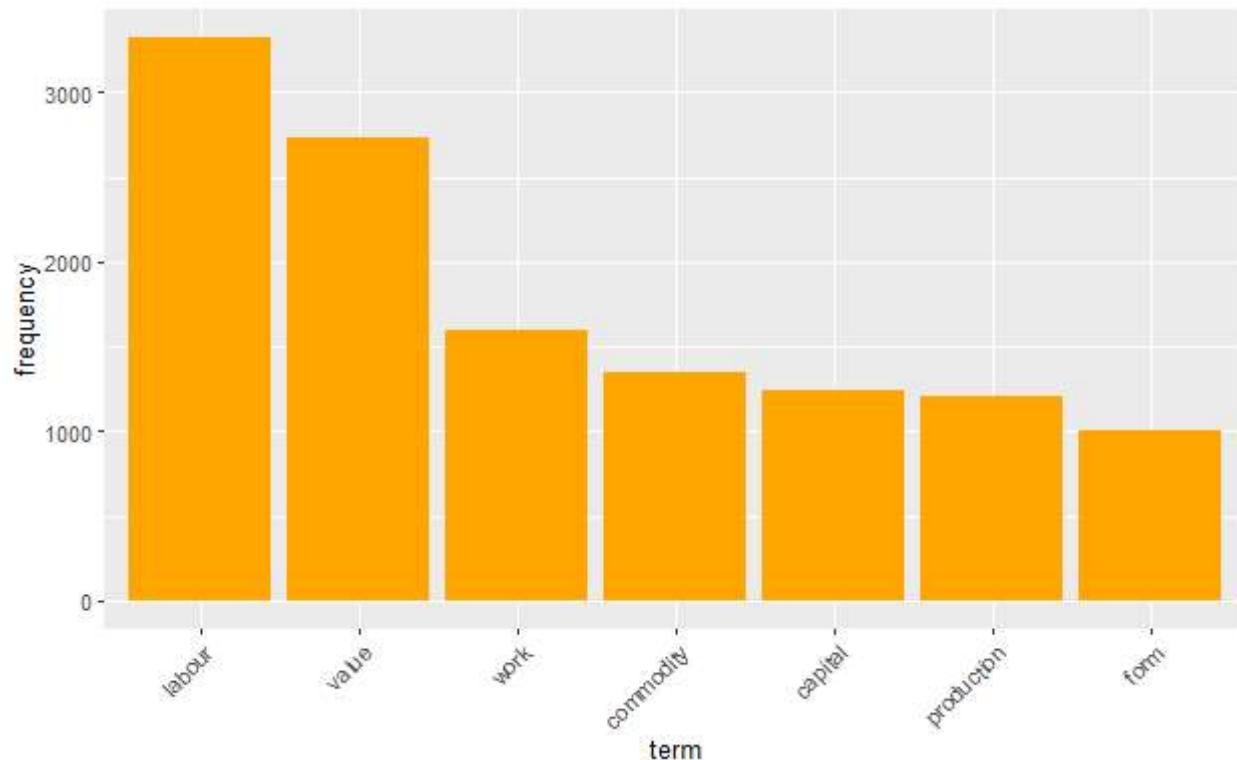
words = Corpus(VectorSource(Capital$lemmatized_str))
words = words %>%
  tm_map( removeNumbers)%>%
  tm_map( removePunctuation)%>%
  tm_map( removeWords, c("the", "and", "PRON", stopwords("english")) )%>% # Remove PRON
  tm_map( removeWords, c("must", "however", "two", "one", "upon", "much", "every", "either", "mu
st", "can", "may", "therefore", "will", "good", "great", stopwords("english")))%>%
  tm_map(removeWords, c("even", "always", "whole", stopwords("english")))%>%
  tm_map(stripWhitespace)

### Document matrix 1
words_dtm <- DocumentTermMatrix(words)

### Document matrix 2 based on [TF-IDF] which measures the relative importance of a word.
words_dtm_tfidf <- DocumentTermMatrix(words, control = list(weighting = weightTfIdf))
words_dtm_tfidf = removeSparseTerms(words_dtm_tfidf, 0.95)

## creating freq data frame
freq = data.frame(sort(colSums(as.matrix(words_dtm)), decreasing=TRUE))
table = data.frame(term = row.names(freq), frequency= freq)
colnames(table) = c('term', 'frequency')

# Bar chart
Capital_chart<-ggplot(data= subset(table, frequency >1000), aes(x=reorder(term, -frequency), y=f
requency)) +
  geom_bar(stat="identity", fill="orange") + theme(axis.text.x = element_text(angle = 45,hjust=1)
) + labs(x ="term")
Capital_chart
```

[Hide](#)

```
## First word cloud  
wordcloud(rownames(freq), freq[,1], max.words=50, colors=brewer.pal(1, "Dark2"))
```

[Hide](#)

```
### Second Word Cloud  
freq.Capital = data.frame(sort(colSums(as.matrix(words_dtm_tfidf)), decreasing=TRUE))  
wordcloud(rownames(freq.Capital), freq.Capital[,1], max.words=100, colors=brewer.pal(1, "Dark2"), scale=c(3,0.25))
```



Based on the word cloud graph generated from book Capital, the book Capital shares similar terms with previous books. However, Capital has a lot of sentences related to the word **Labour/Labourer**

- Labourer: Marx held that, in capitalism, labour was merely a commodity: in exchange for work, a labourer would receive a subsistence wage.

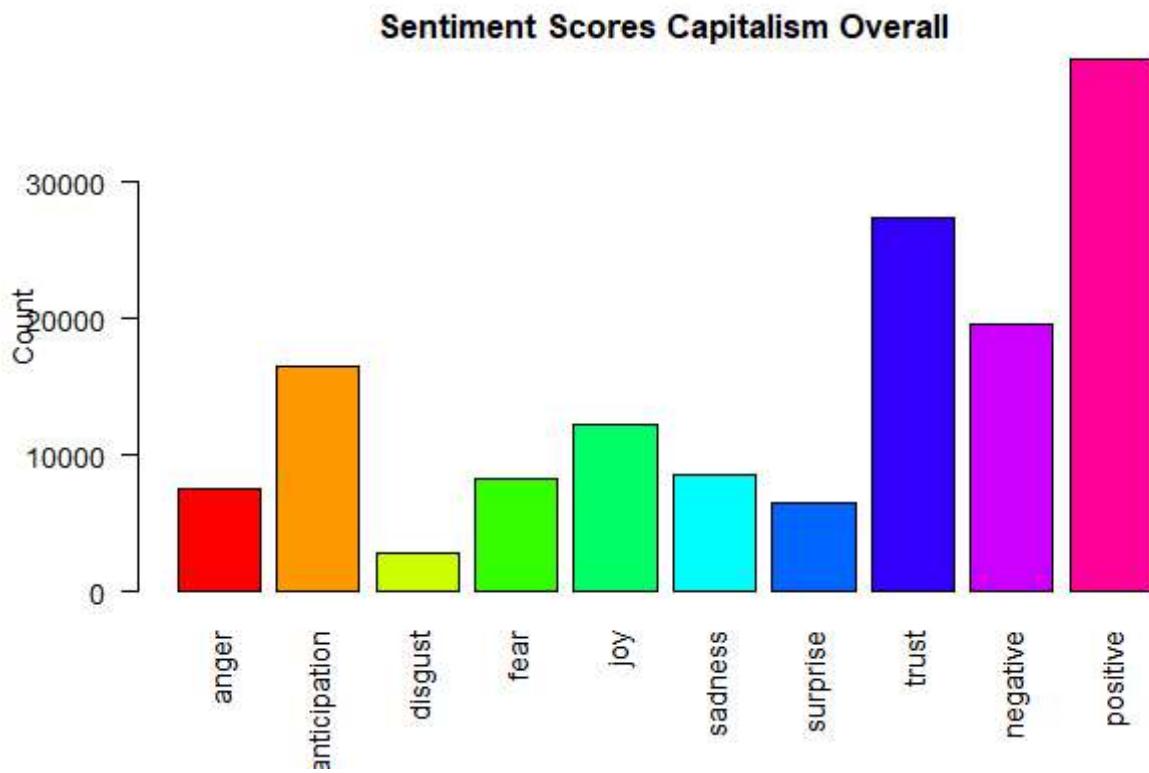
3. Sentiment analysis

Sentiment comparison between capitalism and Communism

[Hide](#)

```
#####
##### sentiments score for capitalism #####
capitalism.score <- iconv(capitalism$lemmatized_str)
capitalism.sent <- get_nrc_sentiment(capitalism.score)

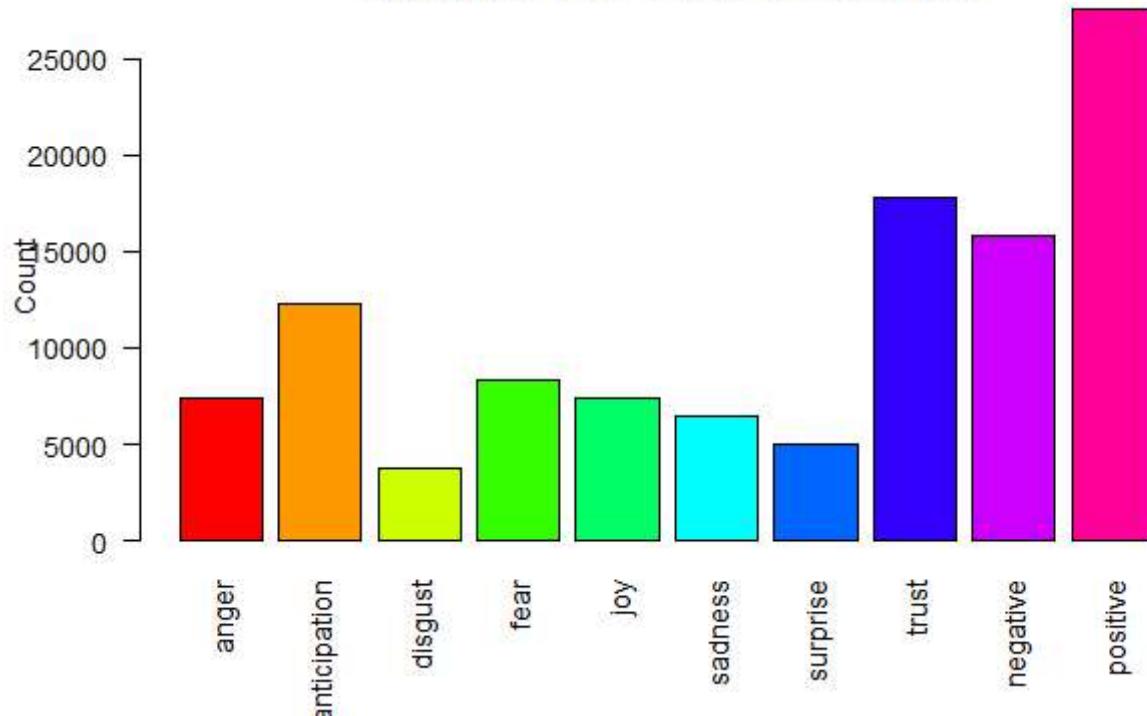
barplot(colSums(capitalism.sent),
       las = 2,
       col = rainbow(10),
       ylab = 'Count',
       main = 'Sentiment Scores Capitalism Overall')
```

[Hide](#)

```
#####
##### sentiment score #####
communism_score <- iconv(communism$lemmatized_str)
communism_sent<- get_nrc_sentiment(communism_score)

barplot(colSums(communism_sent),
       las = 2,
       col = rainbow(10),
       ylab = 'Count',
       main = 'Sentiment Scores Communism Overall')
```

Sentiment Scores Communism Overall

[Hide](#)

NA
NA
NA

[Hide](#)

```
capitalism_rate = as.numeric(colSums(capitalism.sent)[9]/colSums(capitalism.sent)[10])
capitalism_rate ## counts of negative / counts of positive
```

[1] 0.5009868

[Hide](#)

```
communism_rate = as.numeric(colSums(communism_sent)[9]/colSums(communism_sent)[10])
communism_rate ## counts of negative / counts of positive
```

[1] 0.5724233

After comparing the proportions of counts of negative words. We see that, overall, books from school of communism has higher negative rate

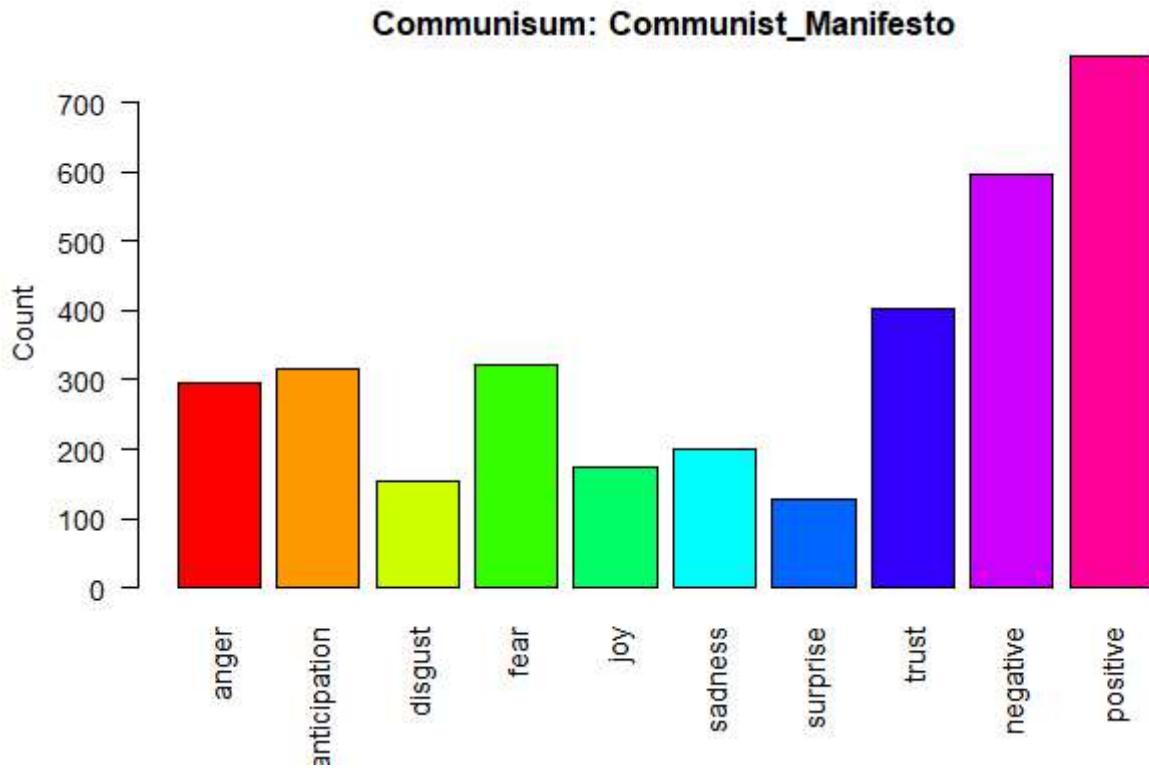
Sentiment Comparison Among Communism Books

[Hide](#)

```
#####
##### Communist_Manifesto #####
freq.Manifesto_score <- iconv(Communist_Manifesto$lemmatized_str)
Manifesto.sent<- get_nrc_sentiment(freq.Manifesto_score)

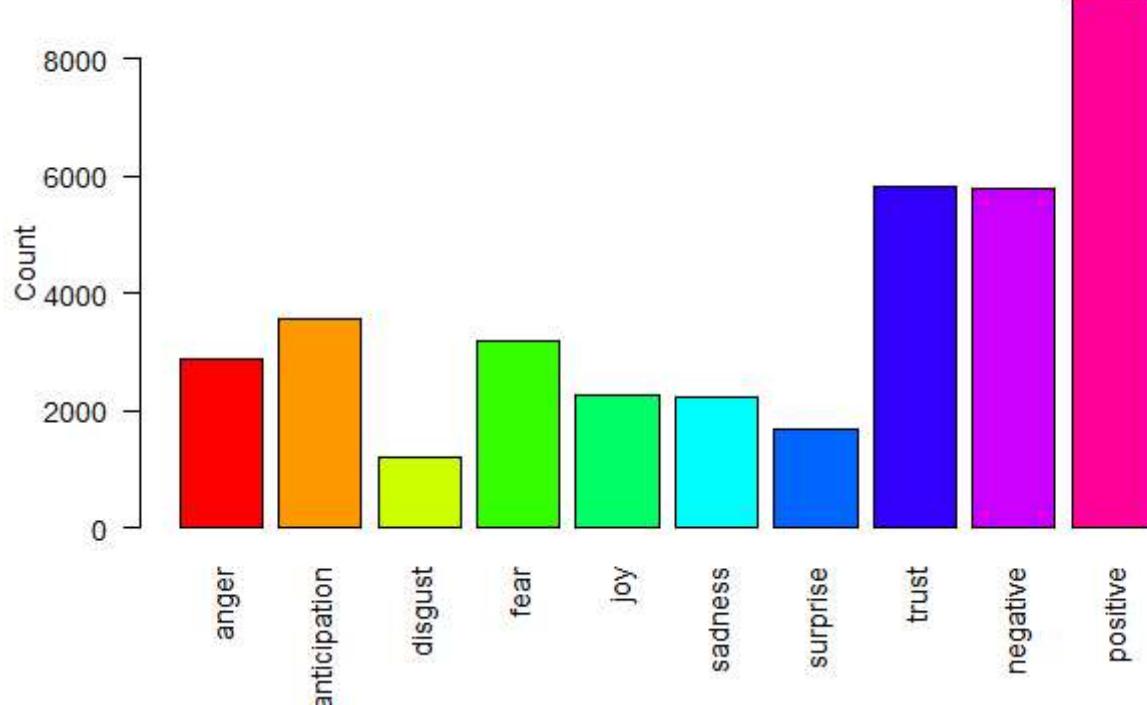
#####
##### lenin #####
lenin.score <- iconv(lenin$lemmatized_str)
lenin.sent<- get_nrc_sentiment(lenin.score)
#####
##### Capital #####
#### sentiment score
Capital.score <- iconv(Capital$lemmatized_str)
Capital.sent<- get_nrc_sentiment(Capital.score)

#####
##### Communist_Manifesto #####
barplot(colSums(Manifesto.sent),
       las = 2,
       col = rainbow(10),
       ylab = 'Count',
       main = 'Communism: Communist_Manifesto')
```

[Hide](#)

```
#####
##### lenin #####
barplot(colSums(lenin.sent),
       las = 2,
       col = rainbow(10),
       ylab = 'Count',
       main = 'SCommunism: Essential Work of Lene')
```

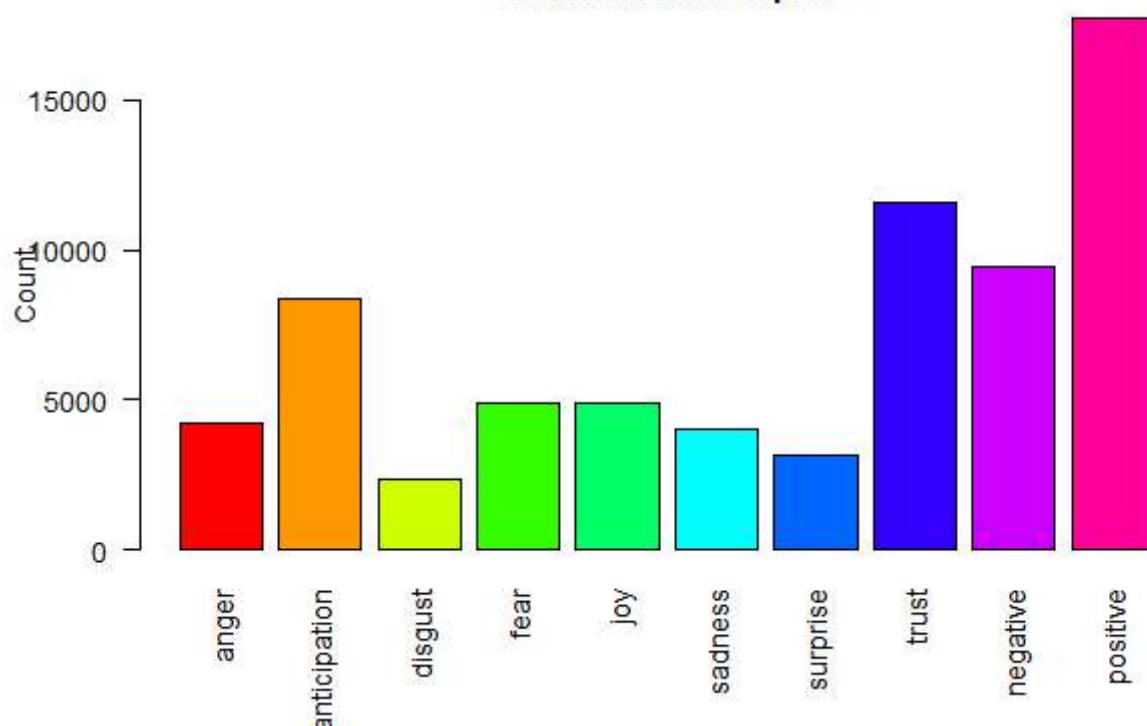
Communism: Essential Work of Lene



[Hide](#)

```
#####
# barplot(colSums(Capital.sent),
#         las = 2,
#         col = rainbow(10),
#         ylab = 'Count',
#         main = 'Communism: Capital')
```

Communism: Capital



```
Manifesto_rate = as.numeric(colSums(Manifesto.sent)[9]/colSums(Manifesto.sent)[10])
Manifesto_rate ## counts of negative / counts of positive
```

[1] 0.7760417

```
lenin_rate = as.numeric(colSums(lenin.sent)[9]/colSums(lenin.sent)[10])
lenin_rate ## counts of negative / counts of positive
```

[1] 0.6376268

```
Capital_rate = as.numeric(colSums(Capital.sent)[9]/colSums(Capital.sent)[10])
Capital_rate ## counts of negative / counts of positive
```

[1] 0.530274

```
data.frame( title = c("Communism Manifesto", "Essential Work of Lenin", "Capital"), negative_rate
e= c(Manifesto_rate, lenin_rate, Capital_rate))
```

title	negative_rate
<chr>	<dbl>
Communism Manifesto	0.7760417
Essential Work of Lenin	0.6376268
Capital	0.5302740
3 rows	

It seems like sentiments are distributed similarly among three different books. However, Communism Manifesto has the highest negative rate. (counts of negative/counts of positive)

Take a brief look on which sentences in Communist Manifesto express anger and joy

```
### Take a brief look on which sentences in Communist Manifesto express anger and joy
Communist_Manifesto[Manifesto.sent$anger >5,]$sentence_spacy[1]
```

[1] "It proved, incontrovertibly, the disastrous effects of machinery and division of labour; the concentration of capital and land in a few hands; overproduction and crises; it pointed out the inevitable ruin of the petty bourgeois and peasant, the misery of the proletariat, the anarchy in production, the crying inequalities in the distribution of wealth, the industrial war of extermination between nations, the dissolution of old moral bonds, of the old family relations, of the old nationalities."

[Hide](#)

Communist_Manifesto[Manifesto.sent\$joy >3,]\$sentence_spacy[1]

[1] "By this, the long wished for opportunity was offered to True Socialism of confronting the political movement with the Socialist demands, of hurling the traditional anathemas against liberalism, against representative government, against bourgeois competition, bourgeois freedom of the press, bourgeois legislation, bourgeois liberty and equality, and of preaching to the masses that they had nothing to gain, and everything to lose, by this bourgeois movement."

[Hide](#)

Take a brief look on which sentences in Essential Works of Lenin express anger and joy
lenin[lenin.sent\$anger >8,]\$sentence_spacy[1]

[1] "During the lifetime of great revolutionaries, the oppressing classes relentlessly persecute them, and treat their teachings with malicious hostility, the most furious hatred and the most unscrupulous campaign of lies and slanders."

[Hide](#)

lenin[lenin.sent\$joy >5,]\$sentence_spacy[2]

[1] "In these pamphlets, legal and illegal material should be collected and grouped systematically, on conditions of labor in a given trade, on the various conditions prevailing in the various parts of Russia, on the principal demands advanced by the workers in a given trade, on the defects of the laws in relation to that trade, on the outstanding cases of workers' economic struggle in this trade, on the rudiments, the present state and the requirements of their trade union organizations, etc."

[Hide](#)

Take a brief look on which sentences in the book Capital express anger and joy
Capital[Capital.sent\$anger >9,]\$sentence_spacy[1]

[1] "We have seen how this absolute contradiction between the technical necessities of Modern Industry, and the social character inherent in its capitalistic form, dispels all fixity and security in the situation of the labourer; how it constantly threatens, by taking away the instrument s of labour, to snatch from his hands his means of subsistence, and, by suppressing his detail function, to make him superfluous, We have seen, too, how this antagonism vents its rage in the creation of that monstrosity, an industrial reserve army, kept in misery in order to be always at the disposal of capital; in the incessant human sacrifices from among the working class, in the most reckless squandering of labour power and in the devastation caused by a social anarchy which turns every economic progress into a social calamity."

Hide

Capital[Capital.sent\$joy >4,]\$sentence_spacy[1]

[1] "Side by side with the masses thus occupied with one and the same work, we find the chief inhabitant, who is judge, police, and tax gatherer in one; the book keeper, who keeps the accounts of the tillage and registers everything relating thereto; another official, who prosecutes criminals, protects strangers travelling through and escorts them to the next village; the boundary man, who guards the boundaries against neighbouring communities; the water overseer, who distributes the water from the common tanks for irrigation; the Brahmin, who conducts the religious services; the schoolmaster, who on the sand teaches the children reading and writing; the calendar Brahmin, or astrologer, who makes known the lucky or unlucky days for seed time and harvest, and for every other kind of agricultural work; a smith and a carpenter, who make and repair all the agricultural implements; the potter, who makes all the pottery of the village; the barber, the washerman, who washes clothes, the silversmith, here and there the poet, who in some communities replaces the silversmith, in others the schoolmaster."

- Wrote at the end, it is hard to believe that there is a perfect social system works perfectly for all. However, it's worth time to learn from different social system and maintain the good communication. We are living in world with a diversity of species, people, cultures, and social systems. Wish us a harmonies, peaceful, and non-violent world

