Final Project

Group Ciao: Luwei Jiang, Qiting Guan, Wangpeng Xie, Yueling Zhang

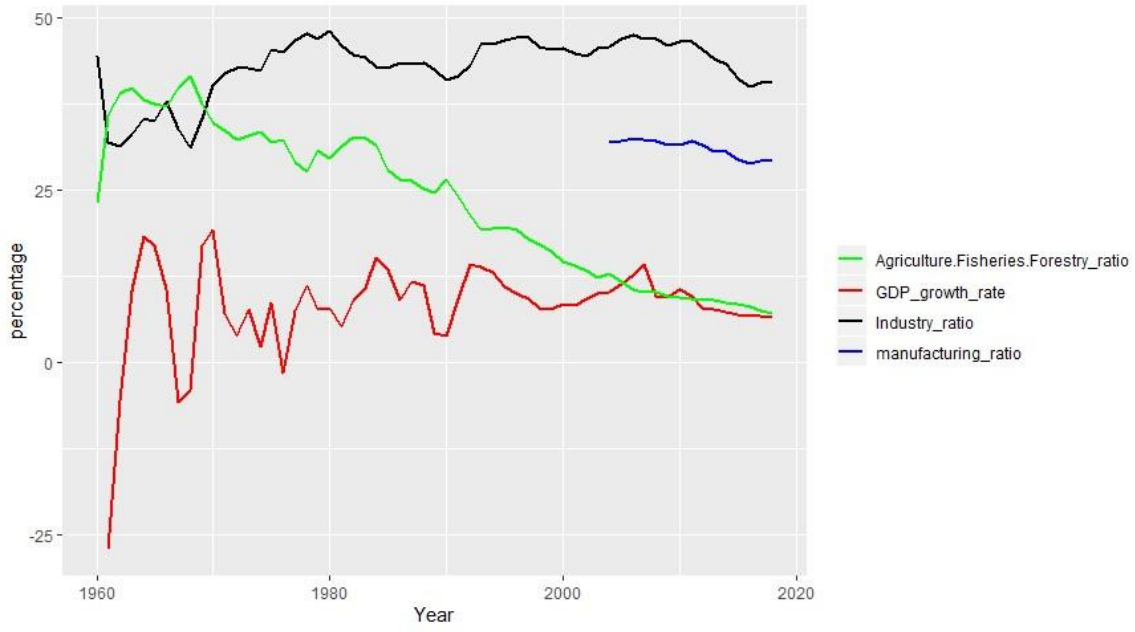
Introduction

To analyze the GDP growth in China, we collected data on China 's GDP, Industry, manufacturing, agriculture, fishery and forestry from 1960 to 2018. In the following analysis, we will analyze China 's Industry, manufacturing, agriculture, fishery and forestry. The impact of development on China's GDP. And establish a suitable regression model to predict the development of China's GDP next.

Analysis

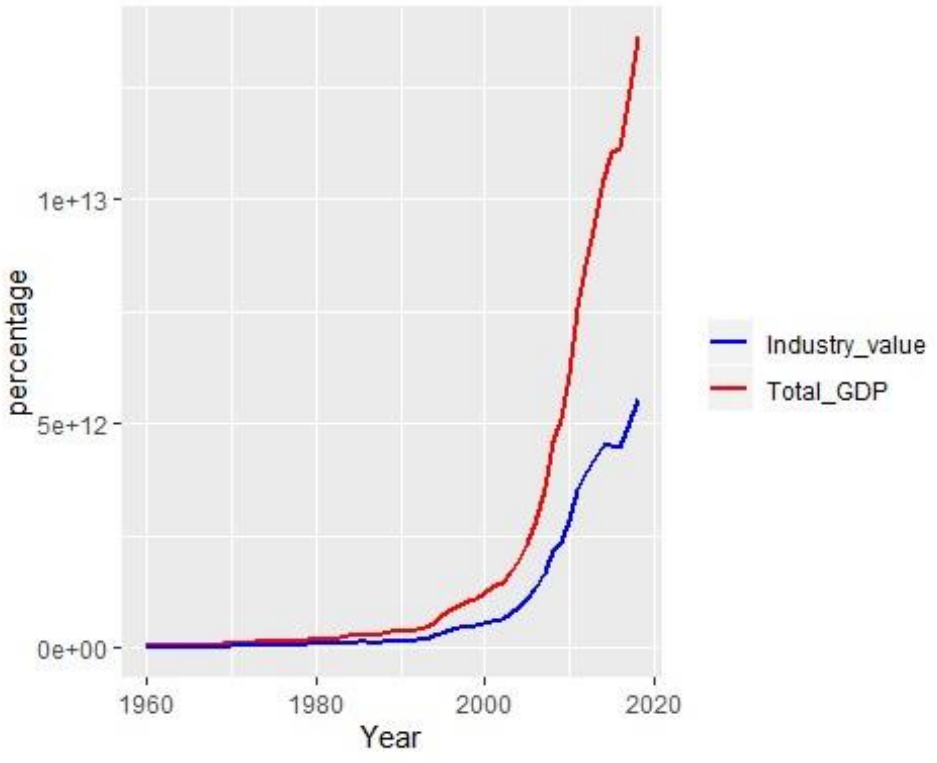
Here we first analyze the first problem： Which industry helps the most to accelerate China GDP growth? In order to solve this problem, we will show the annual GDP growth rate and the proportion of various industries on a graph.

```
library(readxl)

#load the data

data<-read_excel("C:/Users/ibf/Desktop/分析/z200412/总数据.xlsx")

#first problem

library(ggplot2)

ggplot(data = data) +
  geom_line( aes(x = Year, y = GDP_growth_rate, colour = "GDP_growth_ra
te"), size = 1) +  geom_line( aes(x = Year, y = Industry_ratio, colour =
"Industry_ratio
"), size = 1) +
  geom_line( aes(x = Year, y = manufacturing_ratio, colour = "manufactu
ring_ratio"), size = 1) +
  geom_line( aes(x = Year, y = Agriculture.Fisheries.Forestry_ratio, co
lour = "Agriculture.Fisheries.Forestry_ratio"), size = 1)+
scale_colour_manual("", values = c("GDP_growth_rate" = "red", "Indust
ry_ratio" = "black", "manufacturing_ratio" = "blue", "Agriculture.Fishe
ries.Forestry_ratio" = "green"))+  ylab("percentage")
```



Here, due to the lack of manufacturing data, it is difficult to intuitively see the relationship between GDP growth and it. However, from the above line chart, it can be clearly seen that the growth rate of GDP is basically within a positive range after 1980, that is to say, after 1980, China's GDP has been increasing. And the proportion of agriculture, fishery and forestry has been on a downward trend around 1968, which shows that with the growth of GDP, the proportion of agriculture, fishery and forestry's output value is getting lower and lower. It is worth noting here that with the change in GDP growth rate, the proportion of Industry also fluctuates, which shows that the development of Industry has a great impact on the growth of GDP. Next, we will separately analyze the output value of Industry and GDP. The output value is plotted.

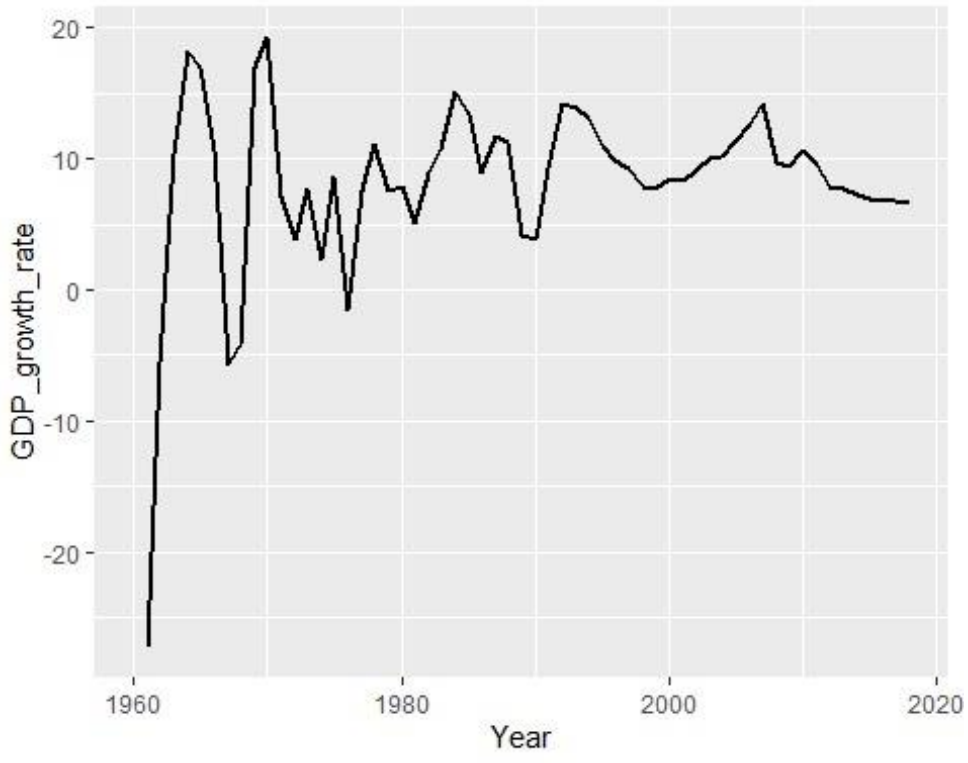
```
ggplot(data = data) +
  geom_line( aes(x = Year, y = Total_GDP, colour = "Total_GDP"), size =
 1) +
  geom_line( aes(x = Year, y = Industry_value, colour = "Industry_value
"), size = 1) +
  scale_colour_manual("", values = c("Total_GDP" = "red","Industry_valu
e" = "blue"))+
  ylab("percentage")
```



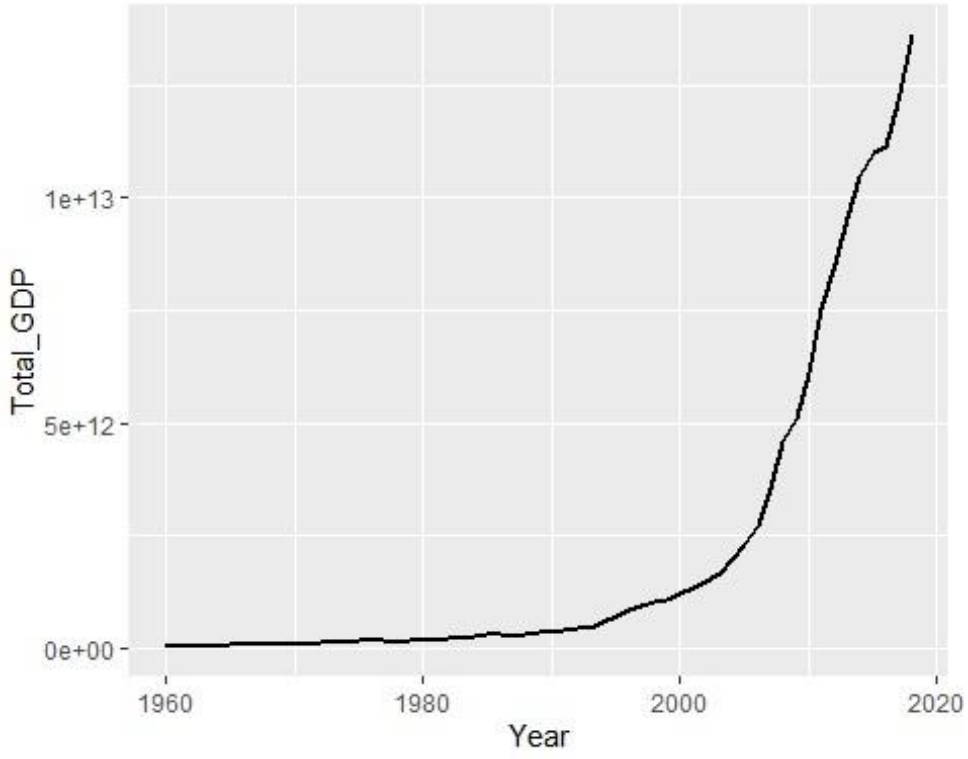
As can be seen from the above chart, Industry's output value growth and GDP's output value growth trend are basically the same, which also shows that Industry's development has a crucial role in GDP growth. Industry-related industries can best promote Chinese GDP increase.

Here we analyze the second problem: How did Chinese GDP trend during these years? Here we make two graphs separately on the growth rate and output value of GDP.

```
ggplot(data = data) +
  geom_line( aes(x = Year, y = GDP_growth_rate), size = 1)
```




```
ggplot(data = data) +
  geom_line( aes(x = Year, y = Total_GDP), size = 1)
```



It can be seen that the growth of GDP output value was relatively slow before 2000, but after 2000, it began to show exponential growth, and the growth rate remained at about 10%. And after 2010, the GDP growth rate began to decline slowly. So it can be guessed that China's GDP will continue to grow in the future, but the growth rate has slowly dropped to another fixed value.

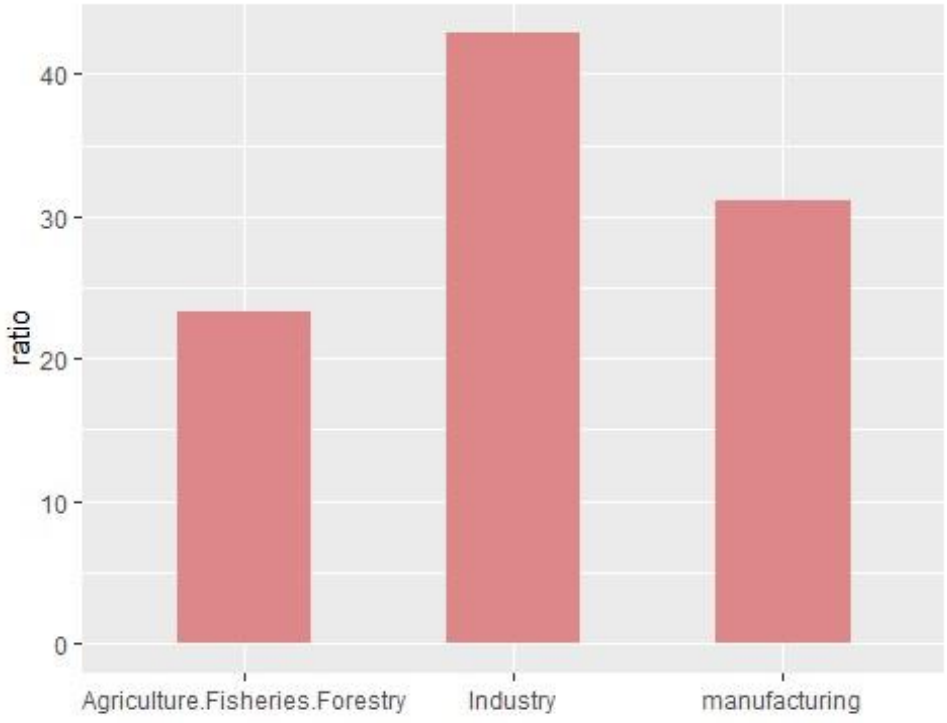
Here we analyze the third problem: Which industry plays the most important role in Chinese Economy? Here, we analyze the descriptive indicators of the proportion of GDP occupied by various industries in these years to determine which industry plays the most important role in the Chinese economy.

```
ratio<-data[,c("Industry_ratio","manufacturing_ratio","Agriculture.Fish
eries.Forestry_ratio")]

summary(ratio)
```

```
##  Industry_ratio  manufacturing_ratio Agriculture.Fisheries.Forestry_
ratio
##  Min.   :31.11   Min.   :28.96       Min.   : 7.19

##  1st Qu.:41.30   1st Qu.:30.01       1st Qu.:12.63

##  Median :44.23   Median :31.60       Median :24.61

##  Mean   :42.91   Mean   :31.08       Mean   :23.29

##  3rd Qu.:46.07   3rd Qu.:32.08       3rd Qu.:32.49

##  Max.   :48.06   Max.   :32.45       Max.   :41.64

##                  NA's   :44
```

In order to display the proportion of each industry more intuitively, we use the average value of these years as a comparative indicator to make a bar chart of the average value of these three industries.

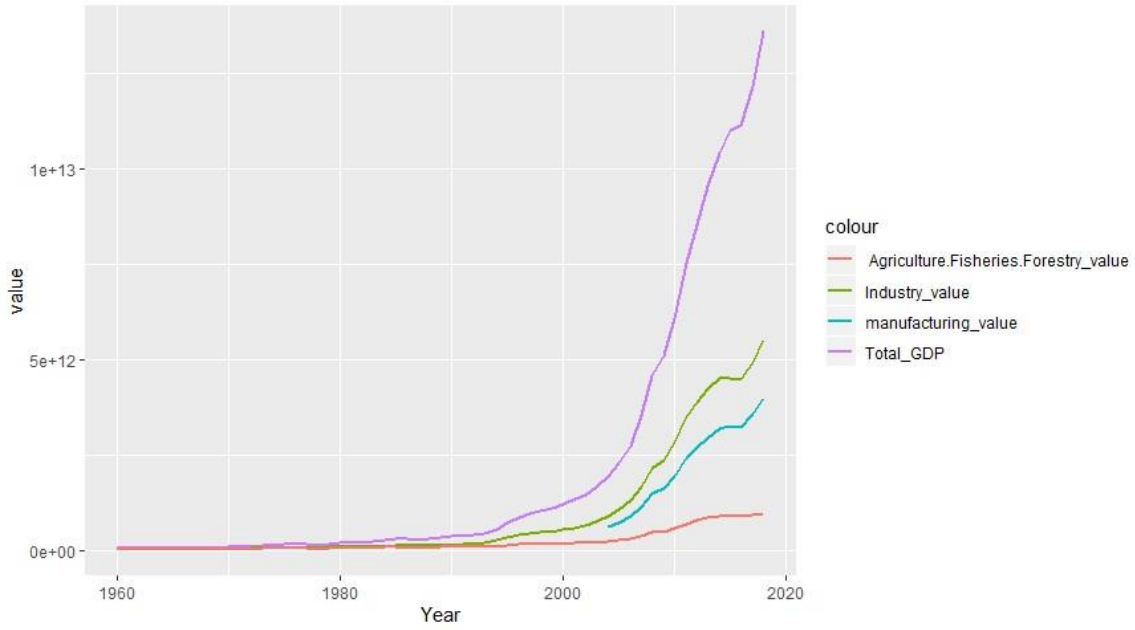
```
mean_ratio<-c(mean(ratio$Industry_ratio),mean(ratio$manufacturing_ratio,
na.rm=TRUE),mean(ratio$Agriculture.Fisheries.Forestry_ratio))
mean_ratio<-data.frame(
  vb=c("Industry","manufacturing","Agriculture.Fisheries.Forestry"),
ratio=mean_ratio
)

ggplot(data = mean_ratio,mapping=aes(x=vb,y=ratio)) +
geom_bar(stat="identity",width=0.5,fill="#DD8888")+   xlab("")
```



From the above results, we can see that Industry has the highest proportion, which means Industry plays the most important role in the Chinese economy.

Here we analyze the fourth problem: What does Chinese Economy look like during these years? Here we analyze the development trends of the output value and GDP of all industries:

```
ggplot(data = data) +
  geom_line( aes(x = Year, y = Total_GDP, colour = "Total_GDP"), size =
 1) +
  geom_line( aes(x = Year, y = Industry_value, colour = "Industry_value
"), size = 1) +
  geom_line( aes(x = Year, y = manufacturing_value, colour = "manufactu
ring_value"), size = 1) +
  geom_line( aes(x = Year, y =Agriculture.Fisheries.Forestry_value, col
our = " Agriculture.Fisheries.Forestry_value"), size = 1)+
ylab("value")
```



Here we analyze the fifth problem: How will Chinese Economy change in the future? Here we obtain the prediction model of GDP by establishing a regression model of the output value of the three industries and the total output value of GDP. Due to the fact that there are many missing values in manufacturing, we establish two regression models here, one is the Industry 1960 ~ 2003 Industry, Agriculture.Fisheries.Forestry regression prediction of Total_GDP, and the other is 2004 ~ 2018 Industry, manufacturing, Agriculture.Fisheries .Forestry's regression prediction of Total_GDP.

```
data_mode1<-data[data$Year<2004,]
mode1<-lm(Total_GDP~Industry_value+Agriculture.Fisheries.Forestry_value,
data=data_mode1) summary(mode1)

## ##
Call:
## lm(formula = Total_GDP ~ Industry_value + Agriculture.Fisheries.Fore
stry_value,
##     data = data_mode1)
##
## Residuals:
##        Min         1Q     Median         3Q        Max  ##
-3.355e+10 -8.646e+09  4.518e+08  7.080e+09  3.364e+10
##
## Coefficients:
##                                      Estimate Std. Error t value P
r(>|t|)
## (Intercept)                         1.432e+10  5.889e+09   2.432
0.0195 *
```

```
 <2e-16 ***
## Agriculture.Fisheries.Forestry_value -1.324e-01  1.642e-01  -0.806
 0.4247    ##
---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 ##
## Residual standard error: 1.481e+10 on 41 degrees of freedom
## Multiple R-squared:  0.9989, Adjusted R-squared:  0.9988  ##
F-statistic: 1.788e+04 on 2 and 41 DF,  p-value: < 2.2e-16

data_mode2<-data[data$Year>=2004,]
mode2<-
lm(Total_GDP~Industry_value+manufacturing_value+Agriculture.Fish
eries.Forestry_value,data=data_mode2) summary(mode2)

##
## Call:
## lm(formula = Total_GDP ~ Industry_value + manufacturing_value +
##     Agriculture.Fisheries.Forestry_value, data = data_mode2) ##
## Residuals:
##        Min         1Q     Median         3Q        Max  ##
-1.125e+11 -5.407e+10 -7.935e+09  3.460e+10  1.509e+11  ##
## Coefficients:
##                                       Estimate Std. Error t value P
r(>|t|)
## (Intercept)                          -1.412e+11  8.550e+10  -1.651
0.127
## Industry_value                       -3.803e+00  4.334e-01  -8.775 2.
68e-06 ***
## manufacturing_value                   8.408e+00  5.251e-01  16.013
72e-09 ***
## Agriculture.Fisheries.Forestry_value  1.167e+00  7.657e-01   1.525
  0.156    ##
---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 ##
## Residual standard error: 8.639e+10 on 11 degrees of freedom
## Multiple R-squared:  0.9996, Adjusted R-squared:  0.9995
## F-statistic:  9393 on 3 and 11 DF,  p-value: < 2.2e-16



## Industry_value                        2.216e+00  4.722e-02  46.931
5.
```

From the output of the model, we can see that the agriculture, fishery and forestry industries have had little effect on GDP. And industry and manufacturing have a very important impact on GDP. By comparing the two regression models, we can find that in recent years, the development of industry has begun to have a negative impact on GDP, and the development of manufacturing has a positive impact on GDP. This shows that the role of industry in GDP growth has gradually weakened in recent years, and manufacturing has begun to become the main industry for GDP growth.

Conclusion

Through the above analysis, we can see that China's GDP has developed rapidly in recent years, among which industry and manufacturing have a very important influence. Industrial output value is the most important part of China's GDP output value. However, in recent years, GDP growth has gradually slowed down, the role of industry in GDP growth has gradually weakened, and manufacturing has begun to become the main industry for GDP growth. Therefore, if we want to continue to develop the economy at high speed in the future, we need to pay attention to the development of the manufacturing industry.