The Relationship Between the Causes of Cardiovascular Diseases

Ruoqi Shi *

School of Economics and Management, Najing University of Science and Technology, Nanjing, China

* Corresponding Author Email: 872306155@qq.com

Abstract. To effectively prevent and improve cardiovascular disease. R and python were used to explore the relationship between various indicators of cardiovascular disease, to predict cardiovascular disease and to analyze netizens' focus on cardiovascular disease. The study found that smoking, drinking, higher age, higher cholesterol, higher blood glucose, and heavier men are more likely to develop cardiovascular disease. Men should pay more attention to cardiovascular disease. With the growth of age, both men and women should control their weight, smoke and drink less, eat healthily, work and rest, and control their cholesterol and blood glucose indicators in an appropriate range.

Keywords: Cardiovascular disease; Indicator relationship; Indicator prediction.

1. Introduction

Cardiovascular disease (CVD) has become the leading cause of death in the world. Especially in China, the absolute number of cardiovascular deaths increased by 46% between 1990 and 2013, mainly due to an ageing population, a figure four and three times higher than in the USA and Western Europe respectively [1, 2]. It is estimated that by 2025, 11% of cardiovascular deaths in China, as well as in the rest of East Asia, will be premature, much higher than in North America (6%) and Western Europe (3%). The economic burden of CVD on China is therefore enormous, including high medical costs and indirect productivity losses due to premature death and disability [3, 4].

The mortality rate from cardiovascular disease increases with age. This is due to the slow and irreversible process of atherosclerosis, where cholesterol builds up in the inner walls of blood vessels causing arteries to gradually harden and blood flow to become lower and lower, which causes inadequate blood supply to vital organs such as the heart, brain and kidneys, leading to conditions such as myocardial infarction, stroke and hypertension [5,6]. Currently, more than 16% of China's population is aged 60 years or older, and this proportion is expected to increase to 35% by 2050.

Therefore, this paper urgently addresses 3 tasks.

Task1: To explore the relationships between indicators of cardiovascular disease.

Task2: To use indicators of cardiovascular disease to predict cardiovascular disease.

Task3: To analyse the focus of netizens on cardiovascular diseases.

The above three tasks can be effective in preventing and improving cardiovascular disease.

2. Data

For task1 and task2, the data were downloaded directly from the Kesci dataset and used for graphical analysis of the relationship between indicators of cardiovascular disease and for predicting cardiovascular disease (CVD). The data included 70,000 records of patients with 12 characteristics: age, gender, height, weight, systolic blood pressure, diastolic blood pressure (SBP), cholesterol, blood glucose (CBG), smoking, drinking, exercise and developing cardiovascular disease. For task3, it is proposed to crawl the post data from netizens and comment data of cardiovascular posting bar for analysis.

3. Data Analysis

This section will cover three aspects, graphical analysis of the relationship between indicators of cardiovascular disease, predicting the presence of cardiovascular disease and discovering the focus of netizens on cardiovascular disease.

3.1. Analysis of the Relationship between Indicators of Cardiovascular Disease

Analysis of the relationships between indicators of cardiovascular disease. Firstly, the loaded data were pre-processed by changing the data format from numeric to str for gender, cholesterol, blood glucose, smoking, alcohol consumption, exercise and developing cardiovascular disease as categorical variables, while the data format for height, weight, systolic blood pressure and diastolic blood pressure remained unchanged as numeric variables; then the ggplot2 package was used to make box plots and violin plots to analyse the relationship between each two categorical The ggcorrplot2 package was then used as a box plot and violin plot to analyse the relationship between every two categorical variables and one numerical variable as a group, a total of 5 x 6 = 30 groups; finally, the ggcorrplot package was used as a relationship plot to analyse the relationship between the indicators of cardiovascular disease.

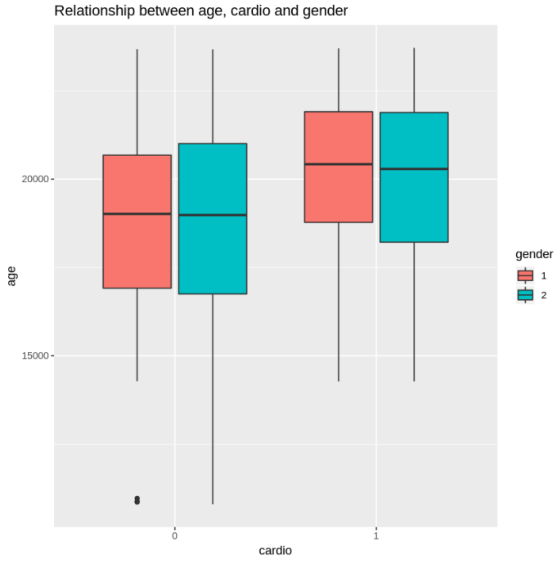

Figure 1. Age, CVD and Gender

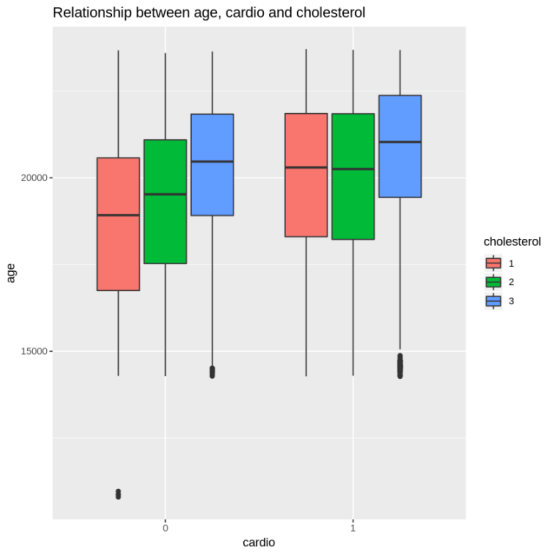

Figure 2. Age, CVD and Cholesterol

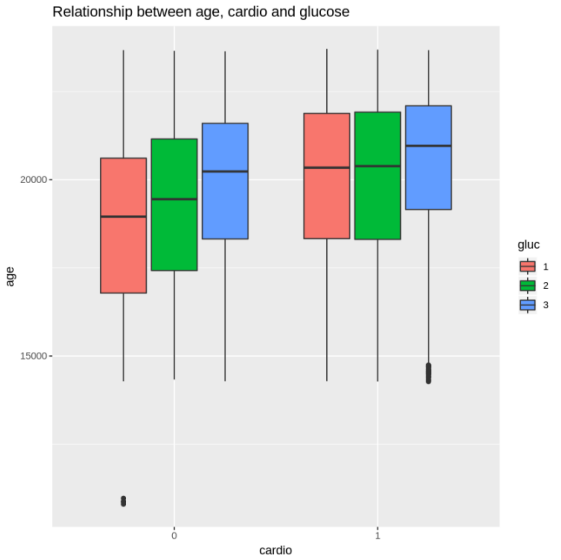

Figure 3. Age, CVD and Blood Glucose

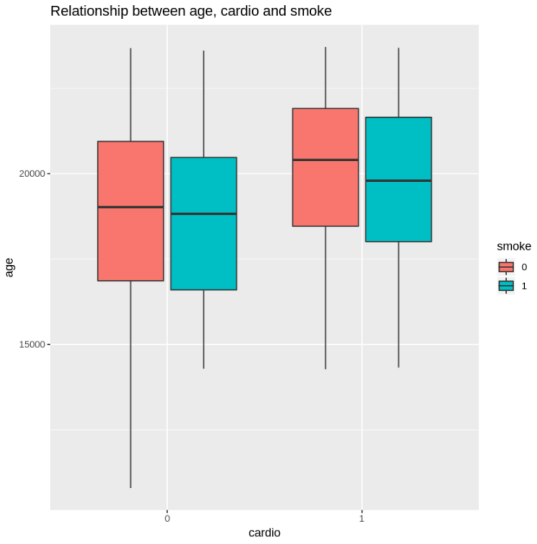

Figure 4. Age, CVD and Smoking

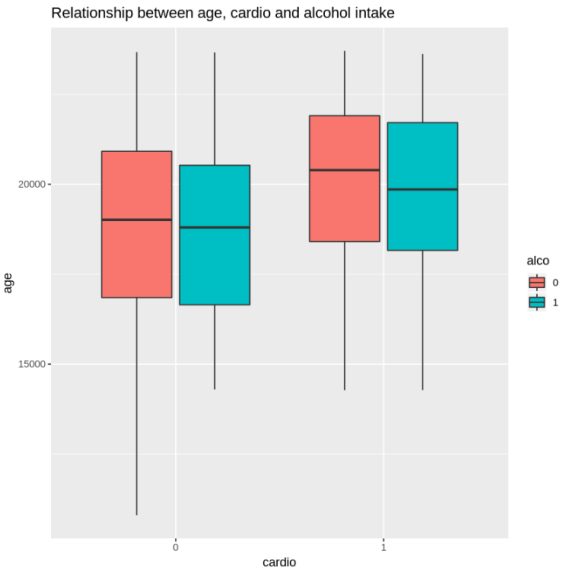

Figure 5. Age, CVD and Drinking

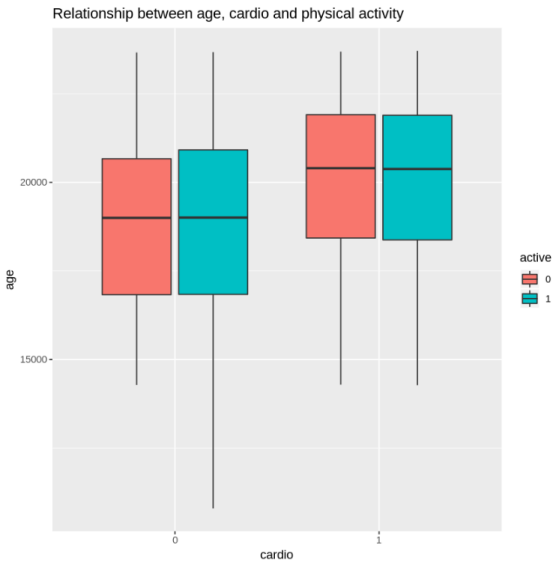

Figure 6. Age, CVD and Exercising

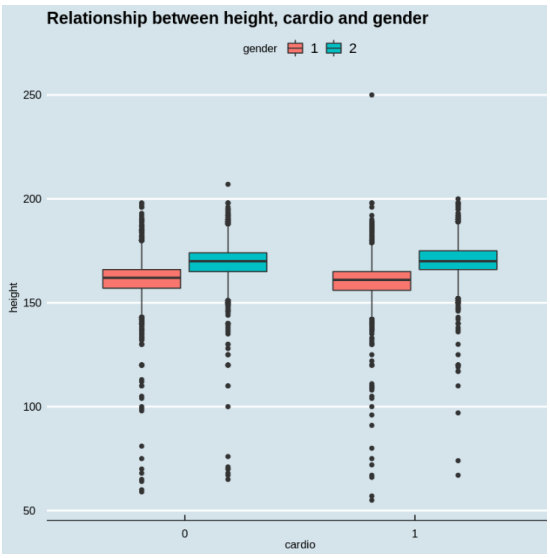

Figure 7. Height, CVD and Gender

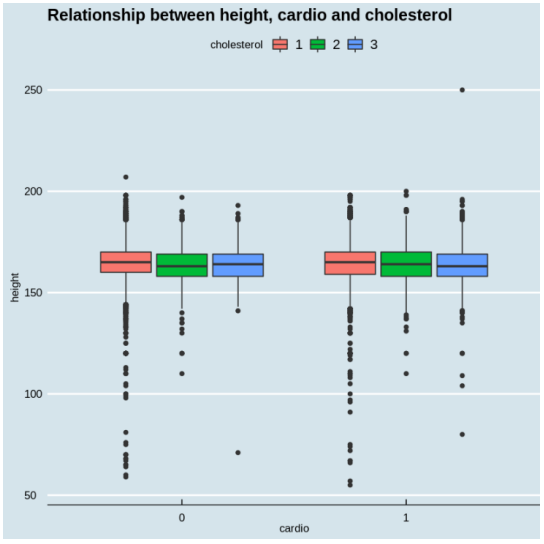

Figure 8. Height, CVD and Cholesterol

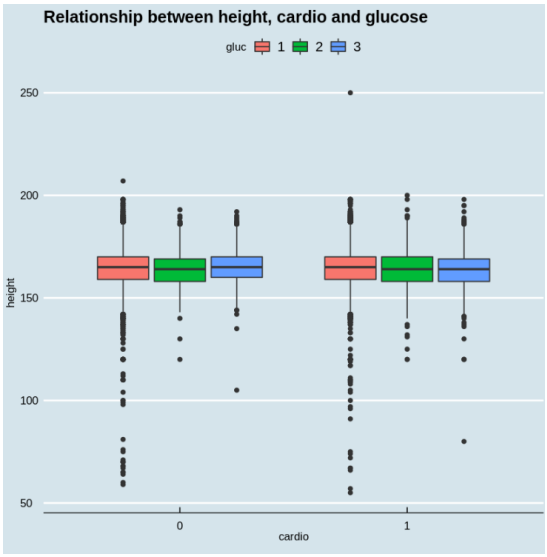

Figure 9. Height, CVD and Blood Glucose

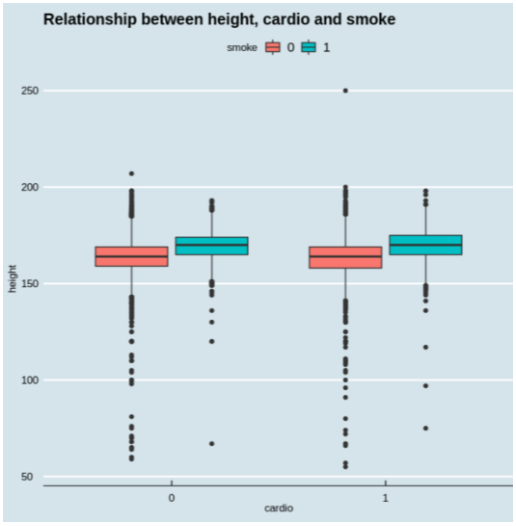

Figure 10. Height, CVD and Smoking

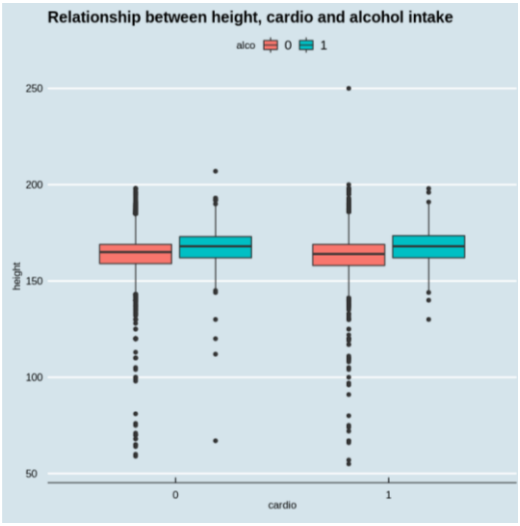

Figure 11. Height, CVD and Driking

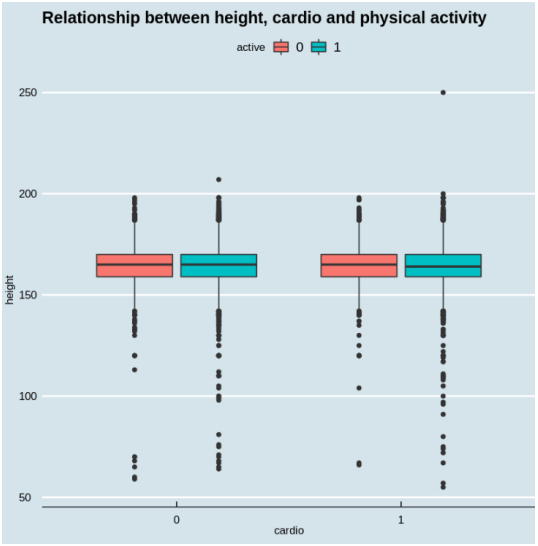

Figure 12. Height, CVD and Exercising

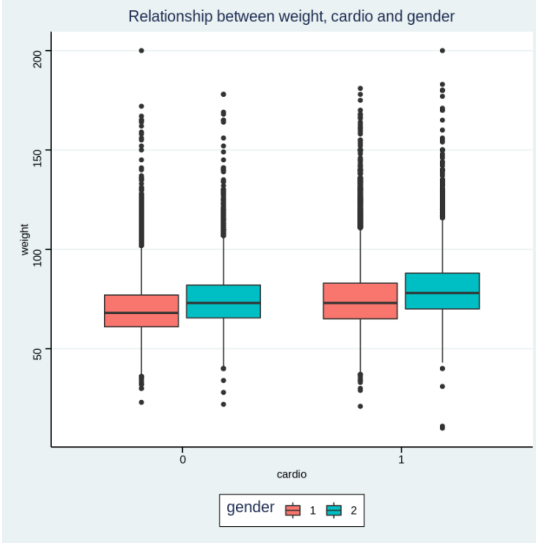

Figure 13. Weight, CVD and Gender

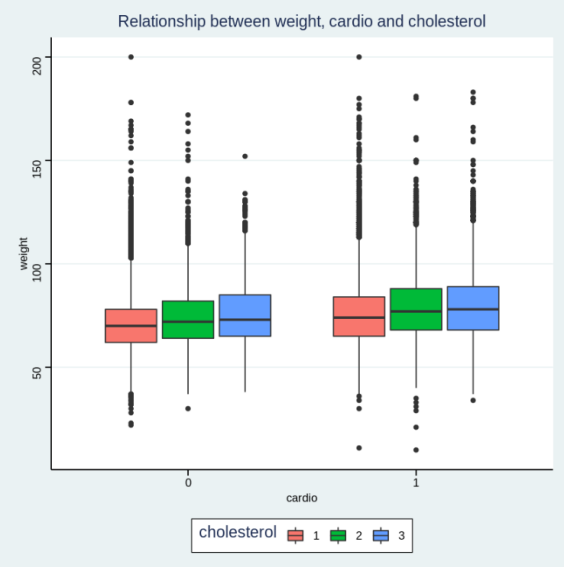

Figure 14. Weight, CVD and Cholesterol

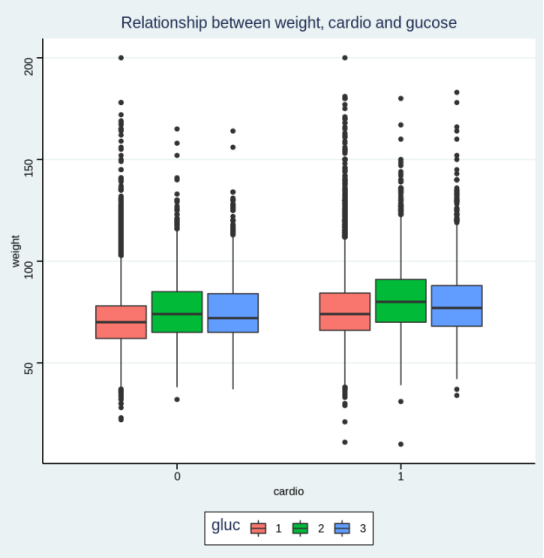

Figure 15. Weight, CVD and Blood Glucose

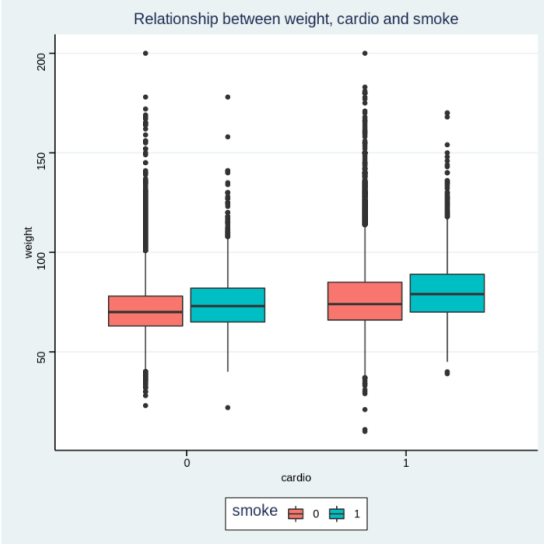

Figure 16. Weight, CVD and Smoking

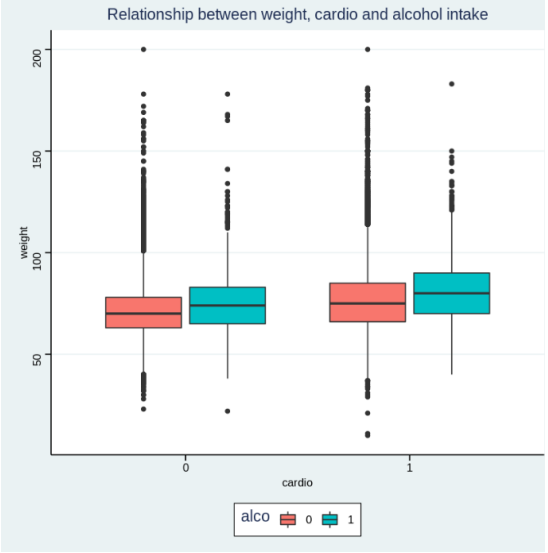

Figure 17. Weight, CVD and Drinking

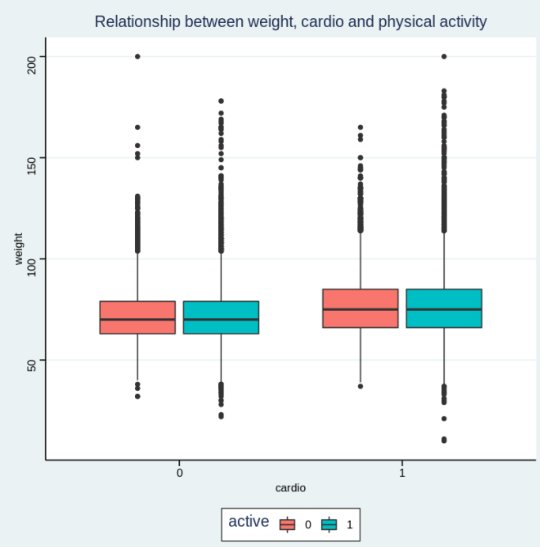

Figure 18. Weight, CVD and Exercising

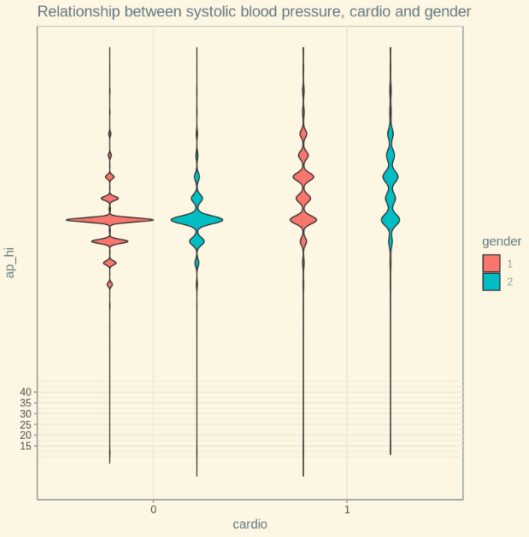

Figure 19. SBP, CVD and Gender

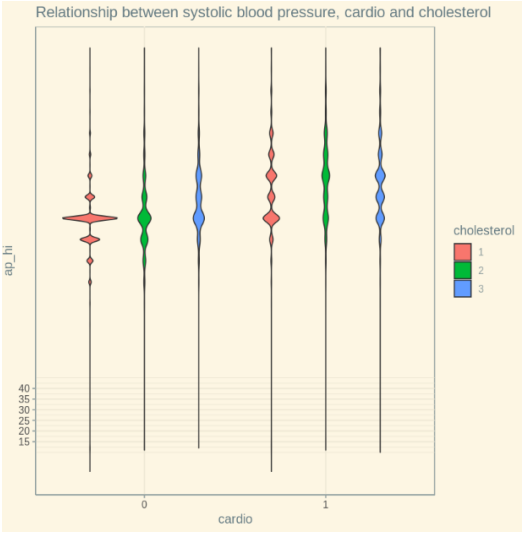

Figure 20. SBP, CVD and Cholesterol

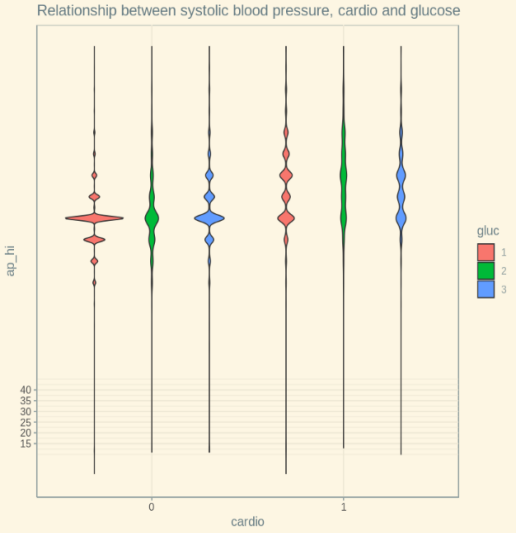

Figure 21. SBP, CVD and Blood Glucose

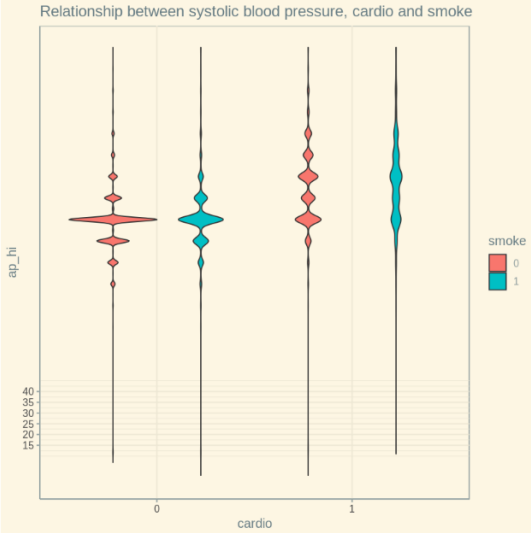

Figure 22. SBP, CVD and Smoking

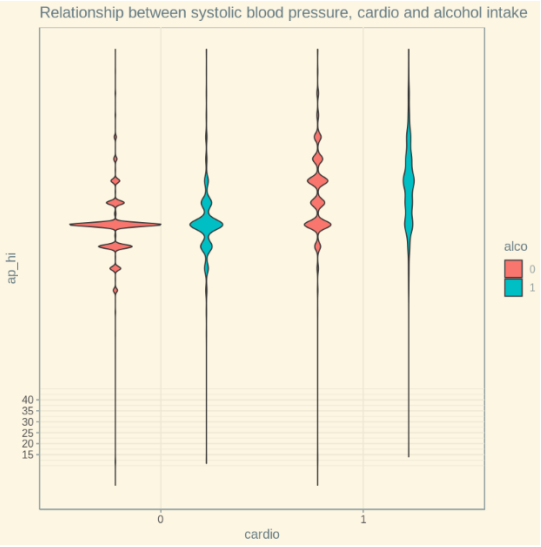

Figure 23. SBP, CVD and Drinking

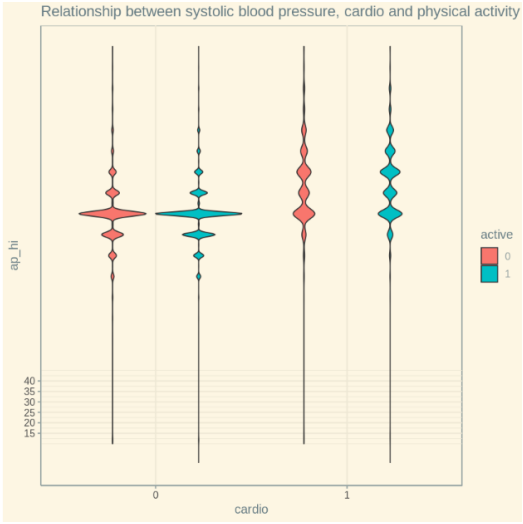

Figure 24. SBP, CVD and Exercising

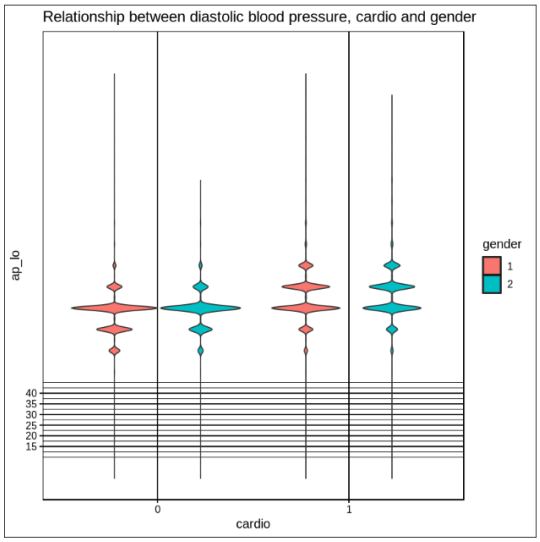

Figure 25. DBP, CVD and Gender

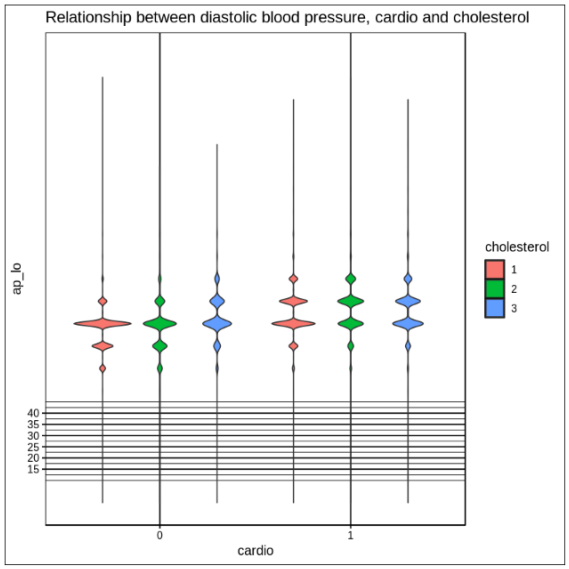

Figure 26. DBP, CVD and Cholesterol

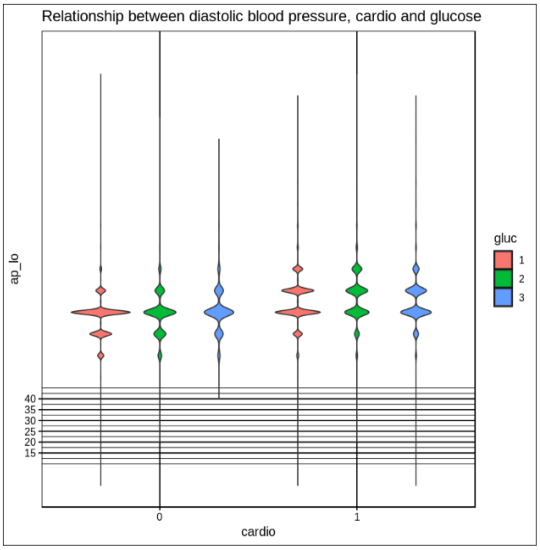

Figure 27. DBP, CVD and Blood Glucose

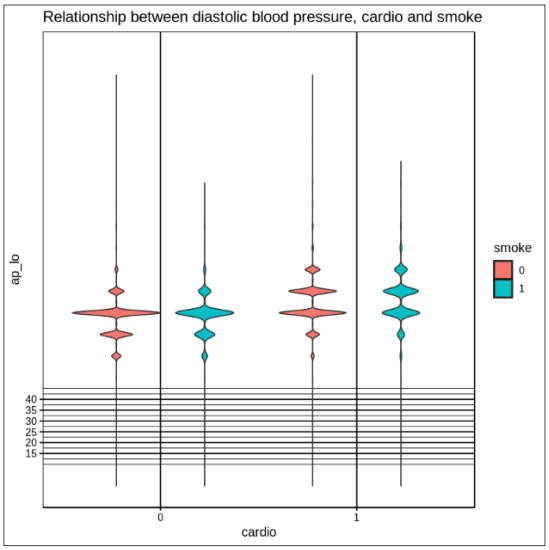

Figure 28. DBP, CVD and Smoking

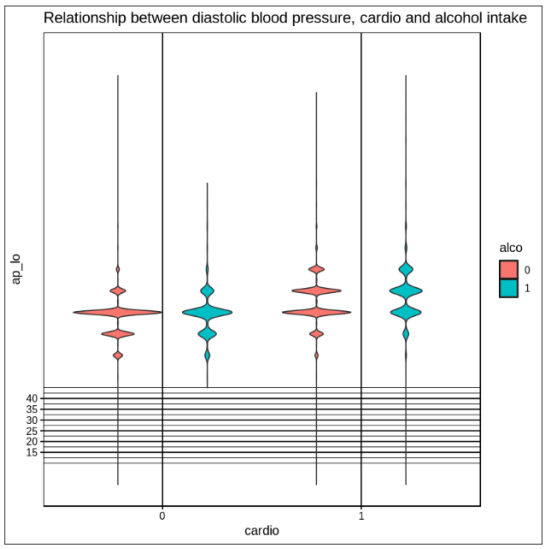

Figure 29. DBP, CVD and Drinking

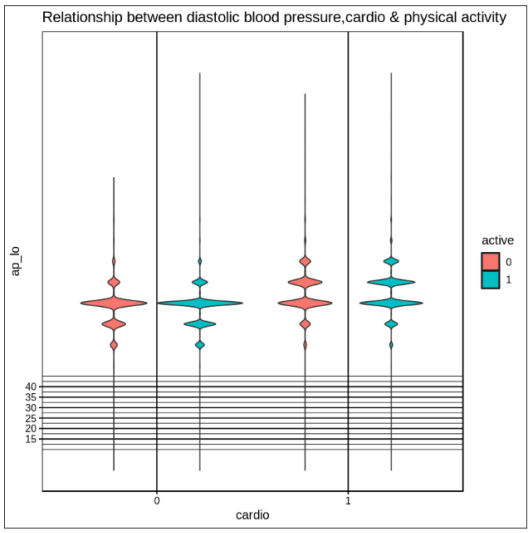

Figure 30. DBP, CVD and Exercising

From the above 30 sets of relationship diagrams, it can be concluded that:

(1) Smoking, drinking, higher age, higher cholesterol, higher blood glucose and heavier men are more likely to develop cardiovascular disease; height has little relationship with whether or not they develop cardiovascular disease.

(2) The older the person, the higher the cholesterol, the higher the blood glucose, the fewer people smoke and the fewer people drink; there is little relationship between age and whether or not they exercised.

(3) People who are taller smoke more and drink more; height has little relationship with cholesterol, blood glucose and whether or not they exercise; and men are taller than women. Since men are taller than women, most people who are taller are men, so people who are taller smoke more and drink more alcohol.

(4) The higher the weight the higher the cholesterol, the higher the blood glucose, the more smokers and the more drinkers; weight has little relationship with whether or not one exercises; men are heavier than women. Since men are heavier than women, most of the heavier people are men, so the heavier people smoke more and drink more; whether men or women, the heavier they are, the higher their cholesterol and blood glucose.

(5) Both systolic and diastolic blood pressure are not significantly related to gender, cholesterol, blood glucose, whether or not they smoke, whether or not they drink, and whether or not they exercise.

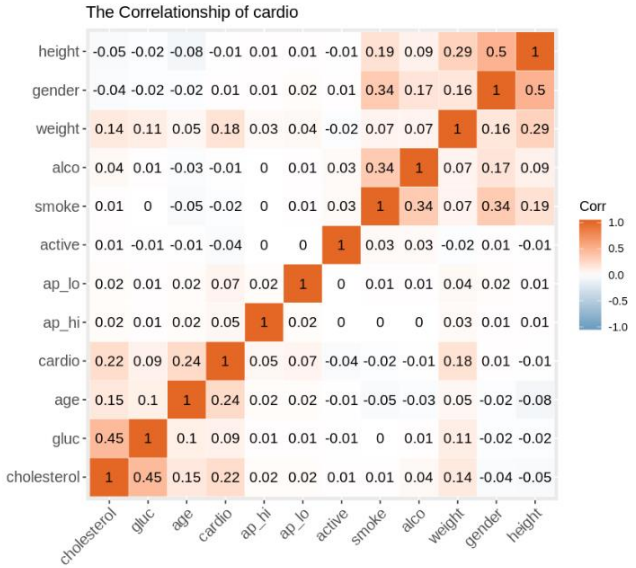

Figure 31. Relationship between indicators of cardiovascular disease

Figure 31 shows that height has the highest correlation with gender at 0.5, followed by cholesterol and blood glucose at 0.45, followed by both gender and smoking at 0.34, followed by smoking and drinking at 0.34, followed by height and weight at 0.29, followed by age and developing cardiovascular disease at 0.24, followed by cholesterol and developing cardiovascular disease at 0.22.

3.2. Predicting Cardiovascular Disease

The lm function in R was used to construct a multiple linear programming model based on various indicators of cardiovascular disease, which was used to predict cardiovascular disease by inputting data on one's height, weight, systolic blood pressure, diastolic blood pressure and so on. The model constructed is:

$$Y=Intercept+X_1*age+X_2*gender+X_3*height+X_4*weight+X_5*ap\_hi+X_6*ap\_lo+X_7*cholesterol+X_8*gluc+X_9*smoke+X_{10}*alco+X_{11}*activ \quad (1)$$

For example, for a person who is 8068 days old, female, 172cm tall, 65kg in weight, with a systolic and diastolic blood pressure of 120 and 60 respectively, normal cholesterol and blood glucose, a non-smoker and a non-drinker, and an exerciser, the predicted result is -0.08792812, which means that the likelihood of developing cardiovascular disease is extremely low, the closer it is to 1 or even greater than 1, the more likely it is to develop cardiovascular disease, and the closer it is to 0 or even less than 0, the less likely it is to develop cardiovascular disease.

3.3. The focus of netizens on cardiovascular disease

Using python's jieba word splitting, a Chinese word splitting component, the crawled posting data and comment data of cardiovascular posting bar netizens is split and deactivated, and the keywords are counted to generate a keyword cloud, which reveal the netizens' focus on cardiovascular disease.

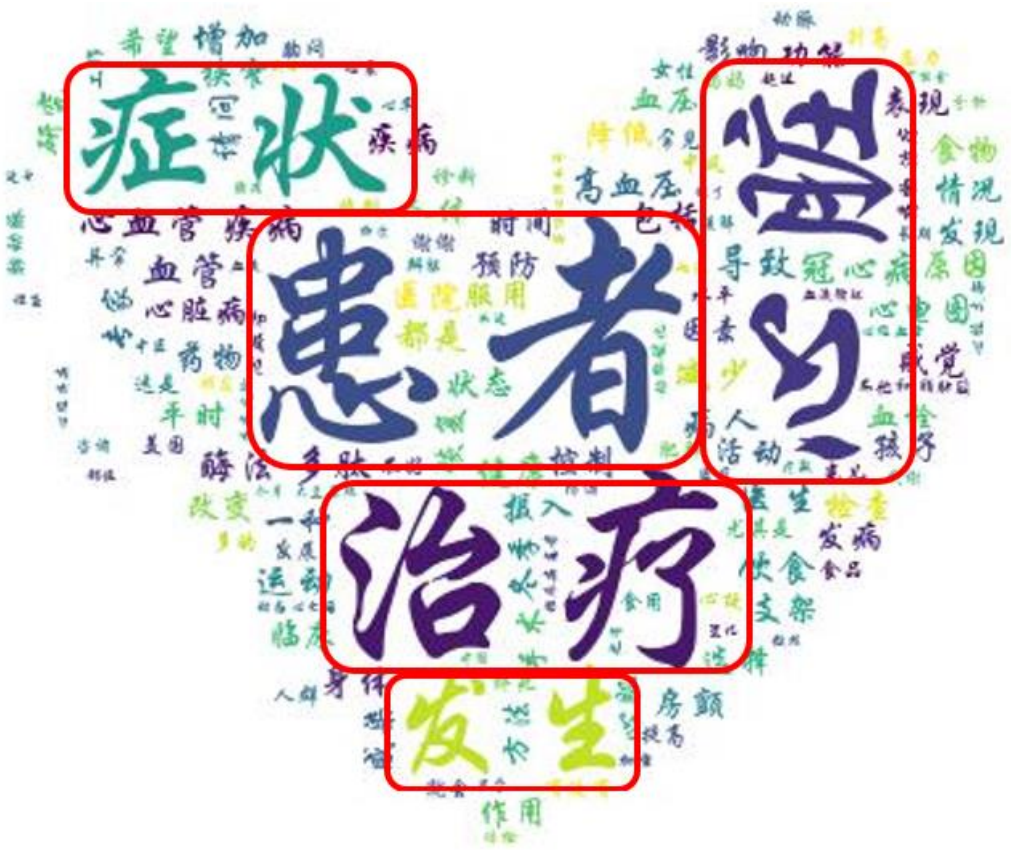

Figure 32. Keyword Cloud

Figure 32 shows that the focus of Netizens on cardiovascular disease is on words such as '患者', '治疗', '心脏', '症状', '发生' and so on, which means 'patient', 'treatment', 'heart', 'symptoms' and

'occurrence', indicating that most of the Netizens of Cardiovascular posting bar are cardiovascular patients or friends and relatives of cardiovascular patients; most of the Netizens of Cardiovascular posting bar are concerned about how cardiovascular disease is treated and how it is treated; whether it causes some diseases of the heart; the symptoms of cardiovascular disease; how cardiovascular disease occurs, etc.

4. Conclusion

This paper studies the indicators of cardiovascular disease, predicts cardiovascular disease and analyses the focus of netizens on cardiovascular disease. The study found that men should pay more attention to cardiovascular disease. with the growth of age, both men and women should control their weight, smoke and drink less, eat healthily, work and rest, and control their cholesterol and blood glucose indicators in an appropriate range; most of the netizens of the cardiovascular posting bar are cardiovascular disease patients or friends and relatives of cardiovascular disease patients; most of the Netizens of Cardiovascular posting bar are concerned about how cardiovascular disease is treated and how it is treated; whether it causes some diseases of the heart; the symptoms of cardiovascular disease; how cardiovascular disease occurs, etc.

The limitations of this paper are that there are few indicators related to cardiovascular disease, only multiple linear regression model has been used and no other models have been used for comparison, no further analysis of posting data and comment data of cardiovascular posting bar netizens has been done, in the future there is a need to expand the indicators related to cardiovascular disease, use more models to predict cardiovascular disease and conduct comparative analysis, analyse posting data and comment data of cardiovascular posting bar netizens for emotional tendencies, hot topics, etc.

References

[1] Lee ChunOn, Li HangLong, Tsoi ManFung, Cheung ChingLung, Cheung Bernard Man Yung. Association between the liver fat score (LFS) and cardiovascular diseases in the national health and nutrition examination survey 1999-2016. [J]. Annals of medicine, 2021, 53(1).

[2] Lee Kayoung. Association Between Food Security and 10-Year Cardiovascular Disease Risk Differs by Gender and Weight Status. [J]. Metabolic syndrome and related disorders, 2021.

[3] Leiherer Andreas, Mündlein Axel, Laaksonen Reijo, Lääperi Mitja, Jylhä Antti, Fraunberger Peter, Drexel Heinz. Comparison of recent ceramide-based coronary risk prediction scores in cardiovascular disease patients. [J]. European journal of preventive cardiology, 2021.

[4] Hejazi Ehsan, Emamat Hadi, Sharafkhah Maryam, Saidpour Atoosa, Poustchi Hossein, Sepanlou Sadaf G, Sotoudeh Masoud, Dawsey Sanford M, Boffetta Paolo, Abnet Christian C, Kamangar Farin, Etemadi Arash, Pourshams Akram, FazeltabarMalekshah Akbar, Brennan Paul, Malekzadeh Reza, Hekmatdoost Azita. Dietary Acid Load and Mortality from All Causes, Cardiovascular Disease, and Cancer: results from the Golestan Cohort Study. [J]. The British journal of nutrition, 2021.

[5] Pakhare Abhijit, Joshi Ankur, Anwar Rasha, Dubey Khushbu, Kumar Sanjeev, Atal Shubham, Tiwari Ishan Raj, Mayank Vipul, Shrivastava Neelesh, Joshi Rajnish. Linkage to primary-care public health facilities for cardiovascular disease prevention: a community-based cohort study from urban slums in India. [J]. BMJ open, 2021, 11(8).

[6] Emmanuel Kelechi E, Nassar Mahmoud, Nso Nso. Prognostic Value of Cardiovascular Testing in Asymptomatic Patients with a History of Cardiovascular Disease: A Review of Contemporary Medical Literature. [J]. Cureus, 2021, 13(8).