People Under Shark's Attack connected with Tourism Economy based on Data Visualization and Multiple Linear Regression

Shiming Xu *

Department of Mathematics and Statistics, Shandong University, No. 27 Shanda South Road, Jinan, 250000, China

*Corresponding author: 202000091080@mail.sdu.edu.cn

Abstract. In the 21st century, big data visualization is becoming a popular method to get insight of social problems and enables people to handle them properly. In the past few years, there was a growing worry about shark attack. Actually, a lot of people are under attack of shark each year, so it is necessary for people to get insight of the tendency and people can deal with such phenomenon correctly. This paper will mainly use data visualization, data mining and multiple linear regression to conduct an analysis. The number of people attacked by shark have risen dramatically since 1900 because of development of transportation and the more personal hobby of diving. Although shark attack accident has a negative influence on tourism especially sea sightseeing, people shouldn't give up the prosperity of tourism. Instead, people are supposed to come up with some ideas to protect people from shark to make travelling safer.

Keywords: Shark attack; Data visualization; Multiple linear regression; Tourism.

1. Introduction

Since 2000, people have already entered a data-driven era, in which there is a lot of data for people to deal with. If the database is small, it will be rather easy for people to get useful information but when the database is large, data analysis is considered as a messy and complex task. So, data visualization and data mining are exceedingly essential for people to deal with data. Besides, data visualization is significant to get insight of social work. Despite the beginning might be late, sociology has applied many visual tools [1]. When people are going to get insight of data about social life, data visualization is taken into consideration. In the past few years more people are anxious about shark because the number of people under shark's attack have risen up rather quickly these years. Besides, tourism is affected by that tendency in some degree. In order to make people travel safely and tourism prospered as well, it is necessary to do some research on the relationship between the tendency and "tourism factors" such as passenger volume or hobby of diving. Finally, advice on how to improve tourism scientifically is provided.

There has been already some paper about the shark attack, but they focus on the spatial distribution of shark attack all over the world and have found out the reason of shark attack [2]. They don't care much about the general reason of the tendency temporally. Some principles of designing data visualization with good efficiency has been described in previous papers [3, 4]. It is becoming popular for people to apply data visualization in the field of sociology. Therefore, this paper will get insight of the social phenomenon based on that which is meaningful and great. Multiple linear regression is proved to have remarkable an effect on finding out the reasons of some social phenomenon [4]. Most of paper focuses on shark product which is a part of economic industry [5]. This paper will focus on describing relation between shark and tourism. Until now, previous research on the shark attack did not include the data from 725 BC to AD 201, and they did not get sight of the data as well.

Therefore, this paper aims to take use of the International Shark Attack File based on the data visualization and data mining to find out temporal and geographical variation of shark attack from 725 BC to AD 2017. Then, the paper will analyze whether shark is a leading killer. Then, possible reasons of the tendency and the relation between the rising number of people under shark attack and

tourism will be analyzed. Finally, this paper will eliminate the panic of the crowds and give some suggestions on how to make tourism safe even shark becomes much danger than before.

2. Methodology

This paper uses International Shark Attack File and put it in order. The World Health Organization (WHO) publishes a report on drowning which provides average number of people dying of drowning each year. On website of ICAO, author gets the number of world international passenger from AD 1945 to AD 2020. It is mainly used to get a general look on the development of airport transportation on international trip and find out its effect on the tendency. On the website of PADI, author gets data about the number of diving certificates from AD 1970 to AD 2011. This dataset reflects the number of people developing a hobby of diving is known. And its influence on the number of people under shark attack could be known.

2.1 Data Visualization and Data Mining

International Shark Attack File is actually a large dataset with thousands of data points. The author uses data visualization based on Tableau so that people can better get insight of the data. People under shark's attack are located all over the world. Therefore, a visualization is made to get knowledge of geographical variation which is helpful to the following research. "Space" and "Time" are two different dimensions which can't be considered together otherwise the analysis would be complex and messy. However, the aim of data visualization is helping people to reduce task that are required in analyzing the data. So, the author decides to take temporal analysis independently. The author takes research on the variation of average year number of people under shark's attack with Tableau as well.

2.2 Comparative Analysis

In order to find out whether the shark is the leading killer of people the paper uses the comparative analysis. The paper compares people under shark's attack with people in air crash and with drowning people based on the standard of average year number. Then, the author can come into conclusion.

2.3 Multiple Regression Analysis

Multiple Regression Analysis is a good method to get insight of relation between Y and X to find out if X has strong effect on Y. X could be a n-dimensional vector. In the field of social life, multiple regression analysis is a good tool to find out the reason of an appearance and then get conclusion [6]. This paper will achieve the that based on function "regress" in MATLAB.

3. Result

3.1 Spatial Variation and Temporal Variation

To get insight of this tendency, the author makes sense of spatial and temporal variation from 725 BC to AD 2017 [7]. In another word the author does research on dimension "time" and "space" separately. The value discussed in this part is the total number (TN) of people under shark attack from 725 BC to AD 2017. Fig. 1 clearly shows that there are only four countries with a lot of people suffering the attack of sharks: America, Australia, South Africa, and Mexico. And people in most countries of world are not attacked by shark through 2793 years.

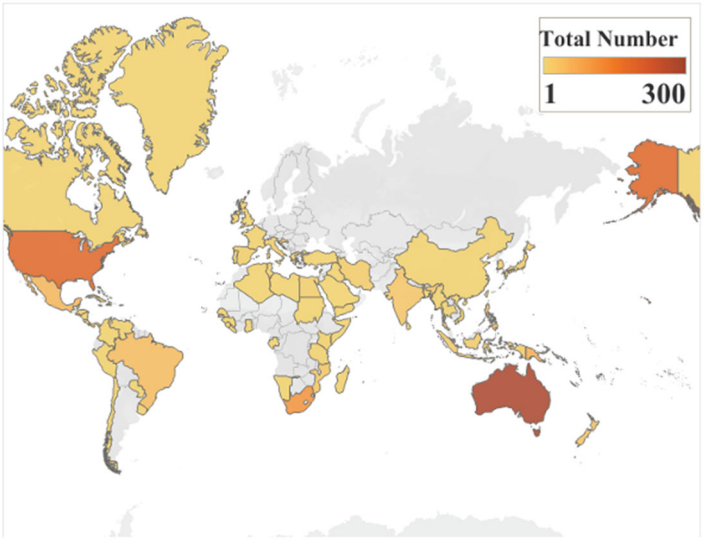

Fig. 1 Geographical variation from 725 BC to AD 2017

Time is another valuable dimension that people should consider when analyzing tendency. Another visualization (Fig. 1) is finished with Tableau. The value discussed in this part is average year number (AYN) of people attacked by sharks all over the world. And before 1800 AYN is almost zero (the difference is within 5). Therefore, only AYN of 1800-2017 is shown. Fig. 2 illustrates that before 1800 there are only a rather small number of people attacked by sharks all over world. Having entered 19th century AYN have risen more quickly than before. Besides there are more people being attacked by sharks after 1900 than that during 1800-1900. Then, during 1998 to 2017, AYN seems stable. So, the data visualization has provided a lot of useful information [8].

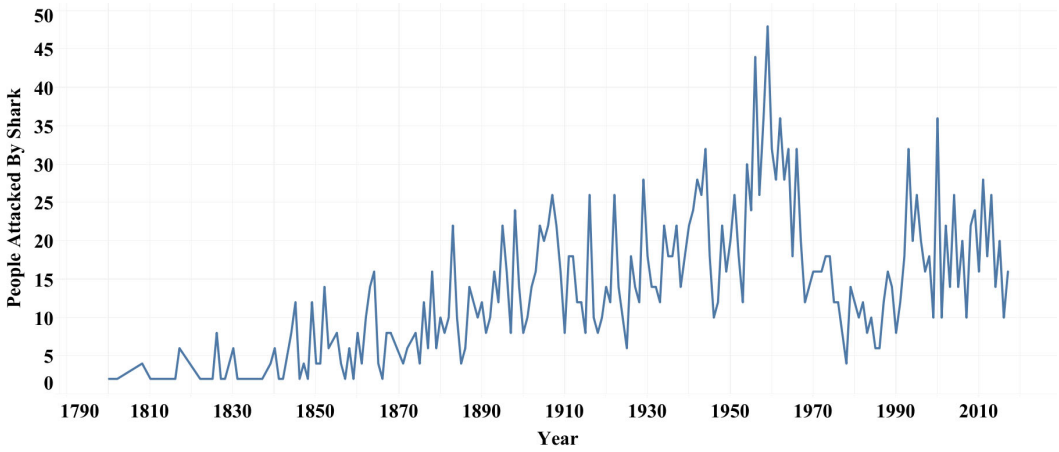

Fig. 2 Temporal variation

3.2 Low-Probability Accident

There is a growing worry about whether shark has become a leading killer of people nowadays. It is true that there are more people attacked by shark every year from 1900 to 2017 than before. But the paper thinks shark is far away from "leading killer of people". In another word, shark haven't killed much people compared with drowning or even air crash. It is just a "freshman" in the field of accident. The author will make comparison "Number of Death Each Year" (NDEY) to make sense of that conclusion.

It is counted by WHO that there are 3720 thousand people dying from drowning. However, there are only 18.05 people killing by shark since 1900 which is the beginning of explosive growth of people attacked by sharks. The difference is very huge. Besides, drowning is becoming the leading

killer of people. So， shark is far away from the term “leading killer of people”. There is no need for people to worry about that. It is just an accident with low probability.

3.3 Reasons

When taking a glance at Fig. 2, people can clearly get a piece of information that having entered 20th century people have been more likely to be attacked by shark every year. Therefore, this period of time is top priority. Besides, there are not that much data before 1970, so this paper will write an analysis on the number of people under shark attack from AD 1970 to AD 2010. Although there are only 40 years, AD 1970 is after World War II. World is peace and people were beginning to make great development. So, the research on this period of time is meaningful.

Regression analysis is a statistical technique for estimating the relationship among variables which have reason and result relation. Main focus of univariate regression is finding out the relationship between a dependent variable and one independent variable and finds out the linear relation equation between dependent and independent variable. Regression models with one dependent variable and more than one independent variable are defined as multilinear regression. By using multiple linear regression, the assumption made before can be examined [6].

The reason of the social and behavioral sciences is complex and if people is going to find out it, it requires adequate prediction or explain. Multiple linear regression and correlation provides powerful tools for analyzing the combined and independent contributions of multiple potential determinants. The method is great and useful when analyzing social problem as well [4]. Author uses multiple linear regression based on a function called “regress” in MATLAB to write an analysis on the correlation [9]. The reason why people have been more likely to be attacked by shark is a kind of social phenomenon, so the method could be used.

In this part, author makes some possible assumptions. Development of airline transportation have made contribution to the tendency. It can be found that after AD 1900 people suffered more from shark every year. And years around AD 1900 is a time when plane was invented and applied. More and more diving certificates are issued all over the world. Having checked the ISAF, people with hobby of diving are more likely to be attacked by shark. Therefore, maybe the increasing popularity is a reason of vital importance.

These two independent variables won’t affect each other. In another word they change respectively. (In most of times, this assumption makes sense) In this part, this paper uses a certain function in MATLAB called “regress” which can achieve multiple linear regression. The dependent variable “Y” is TN and independent variables are “x1” representing the number of diving certificates and “x2” representing international passenger volume. “x1” can show the number of divers all over the world while “x2” could show the development of airplane transportation which is consistent of assumption 1. The result is showing on Fig. 3.

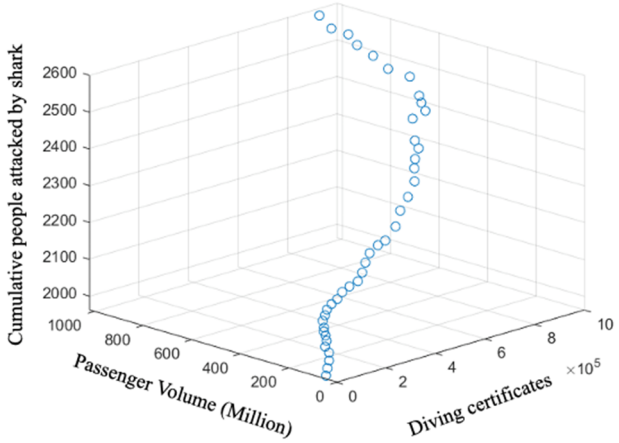

Fig. 3 The relationship between the number attacked by shark and passenger volume and diving certificates

In Fig. 3, (0 200 400……) is x2, (0 2 4 6……) is x1 and z-axis represents Y. The author gets this figure by using a function in MATLAB called "scatter3". This shows distribution of Y corresponding with x1 and x2. Through taking a glance at Fig. 3, people can clearly find out that when x1 and x2 changed respectively, Y performed an almost linear change. Therefore, chances are that multiple linear regression might be meaningful. Then it is time to make a regression by MATLAB and the result is shown. The number of part "stats" illustrates the quality of multiple linear regression. "stats" is 0.989442842845695 1780.72692669947 2.80151365782883e-38 374.437672170330. People should concentrate more on the third number 2.8015e-38 which is almost zero. The "alpha" representing the tolerance in this research is 0.05. And 2.8015e-38 is much lower than 0.05, which is telling people that the regression is rather good and meaningful. It's universally acknowledged that linear relation between dependent variable and independent variable is the strongest relation. Or people can say that the independent variable is the reason of the dependent variable. From what paper get in 3.2.4, the author can conclude that the number of diver and the development of air transportation are two main reasons why TN is increasing much quickly after 1900. Actually, author make another two figures (Fig.4 and Fig.5) to explain.

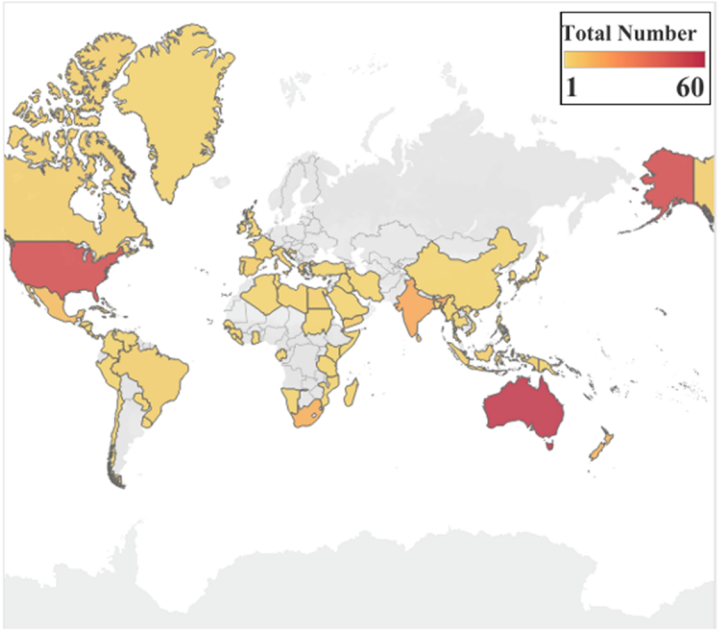

Fig. 4 Geographical variation from AD 1800 to AD 1900

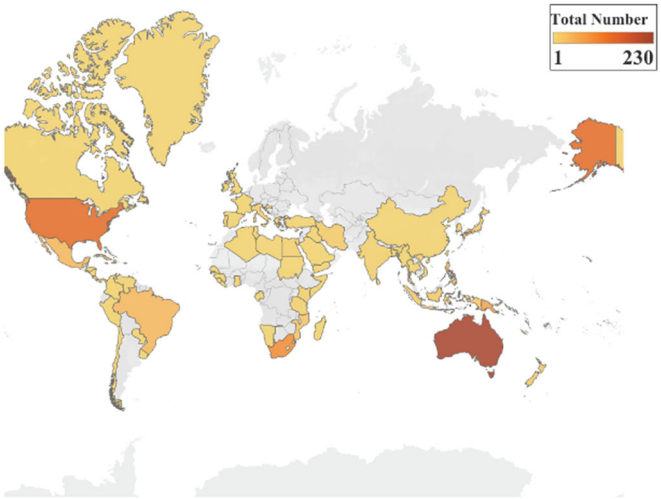

Fig. 5 Geographical variation from AD 1900 to AD 2017

Fig. 4 is the sum of people under shark's attack from 1800 to 1900 and Fig. 5 is the that from 1900-2017. People can find out that no matter what period it is, Australia and America have much larger number of victims than other countries all over the world. So, the reason why Australia and America have more victims is its certain reasons (maybe geographical or humanistic). That is not focus of the paper, it is prepared for following researchers. So, the development of airplane transportation advanced the prosperity of international travel, then more and more people worldwide come to these countries to travel. As a result, possibility of being attacked by shark increased and TN performs a quick change during that period of time.

Aside from that, there were more people developing a hobby of diving after 1970. Human with a diving hobby is more likely to be attacked by shark because their shorter distance to sharks. This is a problem of ratio. For example, there are 1 person loving to dive within 10 persons in the past but now there are 5 persons with diving hobby within 10 persons. And finally， the possibility of being attacked by shark increases. However, the advantage brought by airplane and diving matters more than the disadvantage they brought. People shouldn't give up them though they increase the possibility of being attacked by shark. And that is what author is going to discuss in Part4.

3.4 Tourism

It is true that the increasing number of people attacked by shark will cultivate people's worry about sea, which will finally affect tourism in a certain degree. Firstly, author is going to define diving and visiting seashore as a kind of tourism [10]. No one can deny that taking a trip to sea is tourism. If people go to sea side they will pay for the fee of taking transportation and accommodation. This is an economic perspective on tourism. By using this definition of tourism diving is a kind of tourism as well. If someone go to a sea to dive, he will take transportation and live in a hotel by the sea which will cost a lot of money instead of walking to a beach and just dive. So diving isn't completely free, it can be defined as a kind of tourism. More and more shark attack in modern era will increase the danger of doing these two activities with no doubt. Then there is going to be a group of people canceling the trip considering the danger. Then the industries connected with tourism won't earn that much money because of that. Finally shark attack has a terrible influence on tourism.

However, people shouldn't give up airplane transportation or set a really hard barrier against getting diver certificates, instead they should focus on tourism management [11]. From the result above, shark is not a leading killer of people compared with drowning. And people didn't give up swimming, so it is mad that people give up plane and diving because of shark attack [12]. Although shark attack more people in modern era than before, it is just an accident with rather small possibility. Besides shark is just a small problem faced by people which can be easily solve. On the one hand, people can stay away from shark. For example, the government could set restricted area where there are a lot of sharks appearing to stop visitors diving or sightseeing [13]. On the other hand, some facilities can be applied to keep shark away from people such as special nets around sightseeing area or diving area against shark [14]. All solutions above are just a direction to solve the problem. The author believed that following researchers will solve the problem and give their systematic solutions in the future.

Although shark have just attacked 2718 from 725 BC to AD 2017. It seems that there are no more than even one people dying from shark every year. But everyone's life is invaluable. No one can measure the value of a person's life by money. People can't see shark attack as a reason for stopping develop tourism but it doesn't mean that people will do nothing to shark attack. Therefore, it is necessary to prevent people dying from shark when they are having a sightseeing by a sea or diving. It is cruel for everyone to die when they are performing happy activity.

4. Discussion

Data visualization is an effective method to make temporal and spatial analysis. The data is clear and friendly to people. However, because of the technology author have, the condition of region that

are not country such as Pacific Ocean or other small countries can't be known in the figure. So, it is not strictly a "world size" conclusion and there might be some other result that are not found by author such as the development of fishing and sailing. In the future, software with high technology can be applied and make conclusion complete in world size. Multiple linear regression is proved to be effective to find out the reasons of some social problems and people under shark's attack is a kind of social problem. The result got by this method could be meaningful. However, the dependent variable is not that good. If annual number of people under shark's attack is applied in the analysis it could be best. But in this problem, it is not easy to achieve that, because people being attacked by shark is just an accident with rather small probability, the number of people under shark's attack is very small within a year. If value Y is AYN the regression is not good and that may cause error. Therefore, later researcher can make some operations on data or use a completely new data set to make result more believable.

This paper provides a possible direction on finding relation between shark attack and tourism which is a part of economy. The researcher in the future could find more useful and interesting methods and apply some relative data to make a systematic conclusion. Finally, the assumption that the number of people with diving hobby and development of airplane transportation are independent is strong. They may have effect on each other. In the future researcher can find their detailed relationship and make correction.

5. Conclusion

In summary, this paper proposes data visualization, data mining and multiple linear regression to get insight of the reason of the increasing number of people under shark attack each year since 1900. In terms of geographical variation, Australia, America, and South Africa are the top 3 sufferers. In terms of temporal variation, after AD 1800 and AD 1900, the number of people under shark attack have risen dramatically. The development of airplane transportation driven the prosperity of international travel, leading to that more people worldwide travel to these countries with high probability of shark attack. As a result, TN performs a significant change during that period. Besides, more people have developed a hobby of diving after 1970. Human with a diving hobby is more likely to be attacked by shark due to the distance effect. The government should take measures to be against shark and make sightseeing safe instead of limiting diving certificates or international trip. Some bad effects caused by some accidents will be a barrier, people are expected to focus more on methods to make it safer from perspective of mechanism of accident.

References

[1] Healy K, Moody J. Data visualization in sociology. Annual review of sociology, 2014, 40: 105-128.

[2] Caldicott D G E, Mahajani R, Kuhn M. The anatomy of a shark attack: a case report and review of the literature. Injury, 2001, 32(6): 445-453.

[3] Evergreen S, Metzner C. Design principles for data visualization in evaluation. New directions for evaluation, 2013, 2013(140): 5-20.

[4] Licht M H. Multiple regression and correlation. 1995.

[5] Dent F, Clarke S. State of the global market for shark products. FAO Fisheries and Aquaculture technical paper, 2015 (590): I.

[6] Uyanık G K, Güler N. A study on multiple linear regression analysis. Procedia-Social and Behavioral Sciences, 2013, 106: 234-240.

[7] Keim D, Qu H, Ma K L. Big-data visualization. IEEE Computer Graphics and Applications, 2013, 33(4): 20-21.

[8] Zhu Y. Measuring effective data visualization//International symposium on visual computing. Springer, Berlin, Heidelberg, 2007: 652-661.

[9] Lee Rodgers J, Nicewander W A. Thirteen ways to look at the correlation coefficient. The American Statistician, 1988, 42(1): 59-66.

[10] Tourism and development: Concepts and issues. 2014.

[11] Weaver D, Oppermann M. Tourism management. John Wiley and Sons, 2000.

[12] Eadington W R, Redman M. Economics, and tourism. Annals of Tourism research, 1991, 18(1): 41-56.

[13] Dudley S F J, Cliff G. Some effects of shark nets in the Natal nearshore environment. Environmental Biology of Fishes, 1993, 36(3): 243-255.

[14] Cliff G, Dudley S F J. Protection against shark attack in South Africa, 1952-90. Marine and Freshwater Research, 1992, 43(1): 263-272.