Environment Changing and Coral Bleaching: How the Thermal Stress Anomaly relates to the global statistic percentage of the bleaching area of coral reefs?

Idris Sun, Qianning Lian, Krystal Miao, Zixiu Meng

December 2020

1 Problem Description and Research Question

- Background: Coral bleaching is not the latest news in the world. The environmental conditions that coral reefs live are controlled by the temperature and many other conditions. Unprecedented global warming may cause corals to "expel their symbiotic micro algae and bleach on a local scale" (Eakin C.M., Lough J.M., Heron S.F., Liu G., 2018). At this point, coral reef bleaching happens when it loses its vibrant colour and dies. According to Eakin C.M.'s article, ocean temperature caused coral bleaching spread widely.

- Context: The coral reefs are one of the most important ecosystems that has protected the coastlines from many disasters like erosion and storms. They also provide food and jobs for tropical countries. However, due to global warming, the coral reefs ecosystem is influenced to a certain degree by bleaching. Heron (2017) mentions the word zooxanthellae when coral reefs expel the microscopic algae, and zooxanthellae causes coral's calcium carbonate skeleton to become visible through the now translucent tissue layer. Coral bleaching does not mean coral death. However, a constant high-temperature environment will cause coral death.

- Motivation: Environment anomaly is an international issue that has a wide range of effects on our ecosystem. The reason why we have chosen this topic is every time we talk about global warming, El nino, or such news, we first think about forests and creatures on the land and usually ignore what happened under the sea, like coral reefs. In order to pay attention to the issue of coral bleaching, Obura(2017) claims that there are many benefits that humans can get from coral reefs. For example, "coastal protection, food security and poverty alleviation". The coral reef, as an underwater ecosystem built by clusters of coral bodies, are especially

fragile to the temperature change in water. Consequently, connecting coral reefs bleaching to coral reefs to temperature changes can help us understand the influence of environment anomaly in the ecosystem.

- **Research question: How the Thermal Stress Anomaly relates to the global statistic percentage of the bleaching area of coral reefs?**

2 Description of Relevant Data set

- Name of Dataset: Global Bleaching and Environmental Data
  Sources: https://www.bco-dmo.org/dataset/773466
  Citation: Van Woesik, R., Burkepile, D. (2019)
  Bleaching and environmental data for global coral reef sites from 1998-2017. Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2019-07-18 [if applicable, indicate subset used]. doi:10.1575/1912/bco-dmo. 773466.1

- Format: CSV

- Description: There are five categories of data.

  First, there are 10 columns which represent the location of these coral, they are ID, Latitude Degrees, Longitude, Degrees, Ocean, Realm, Ecoregion, Country Name, State Island Province, City Town, City Town 2, City Town 3.

  Second, there are 3 columns which represent dates and the depth when we measure other data of these coral.

  Third, there is a column which represents the bleaching degree of these coral. It is Average Bleaching rate of the coral reefs.

  Fourth, there are 20 columns which have various expressions of temperature. (Here are the terminology that may appear below: SST: Sea Surface Termperature, SSTA:Sea Surface Temperature Anomoly, TSA: Theraml Stress Anomoly, DHW:Degree Heating Week)They are ClimSST, Temperature Kelvin, Temperature Mean, Temperature Minimum, Temperature Maximum, Temperature Kelvin Standard Deviation, Windspeed, SSTA, SSTA Standard Deviation, SSTA Mean, SSTA Minimum, SSTA Maximum, SSTA Frequency, SSTA Frequency Standard Deviation, SSTA FrequencyMax, SSTA FrequencyMean, SSTA DHW, SSTA DHW Standard Deviation, SSTA DHWMax, SSTA DHWMean(SSTA: Sea Surface Temperature Anomaly).

  Fifth, there are 13 column of the Thermal Stress Anomaly (TSA) of these coral: TSA, TSA Standard Deviation, TSA Minimum, TSA Maximum, TSA Mean, TSA Frequency, TSA Frequency Standard Deviation, TSA FrequencyMax, TSA FrequencyMean, TSA DHW, TSA DHW Standard Deviation, TSA DHWMax, TSA DHWMean. Generally speaking, as the

temperature changes, the level of Thermal stress occurs highly (Eakin, 2018).

And we only use ecoregion, date2, average bleaching, TSA_Frequency, TSA_DHW for our project.

3 Computational Plan

Processing data:

- We first do a filtering to the data by classify data according to ecoregions. To make sure that the data is universal so that it reflects the relationship between temperature and bleaching on a more general scale, we only pick ecological regions where data exceeds 400 pairs.

- The dates also vary greatly in our dataset. It is reported that El nino has a strong influence on the coral bleaching which happens in 2019, 2015, 2008... because of its influence on water temperature. To make sure the data we take only includes the ones that is measured after the effect of El nino. We only accept data that is collected after 2009.1.1.

- Second, we do a transformation to the average bleaching part of the data. According its severity, we import the transformation from A new, high-resolution global mass coral bleaching database (Simon D. Donner,Gregory J. M. Rickbeil,Scott F. Heron, 2017). When there is no coral bleaching, we give a severity 0, when there is 1-10 per cent of bleaching, we consider it mild bleaching and a severity 1, when there is 11- 50 per cent of bleaching, we consider it moderate bleaching and a severity of 2, when there is more than 50 percent of bleaching, we call it severe and a security of 3. This may be helpful when it comes to determine dependence of these two variables.

Level	Severity
0	No Bleaching
1	Mild(0-10% Bleaching)
2	Moderate(10-50% Bleaching)
3	Severe(>50% Bleaching)

(Terminology Introduction: TSA: thermal stress anomaly, DHW: Degree Heating Week)

- Third, we read the data-set and extract the corresponding column we want, which are TSA frequency, TSA DHW, Average Bleaching, and ecoregions as well(why we extract them will be explained in the modeling picture section). We store these variables, respectively, in one set.

- Last, we import scikit-learn library to do the final pre-processing for us. Firstly, according to the data description, we got a warning `A few`

`sampling events do not have associated environmental data, this missing data is denoted with "nd".`, and when converting into python floats, it may become nan, which causes errors, so we import the feature of imputing missing values to implement the data that if there is no complete data. The function we use is simpleimputer.fit and simpleimputer. transform which takes a more complete list of data to learn the rules and it will do the transform to the list of data with missing values. And then we will use We use the feature of Generating polynomial features (we use the function: Polynomial Features (2))so that when inputting the sublists in the dictionary for instance [X1, X2], we get a list of [(1, X1, X2, X2 1 , X1X2, X2 2)]. The function we use is PolynomialFeatures(2).fit_transform which takes in a 2D array(or list) and return an array of value as described above. Then we convert the array back to a list. In this way, in our future operation of the data, we can access the more possibilities of relationships instead of simply linear ones. We stops a power of 2 because bleaching percentage may exceed 100 with higher powers and with trials using sample data, we discover the rule may be more linear than higher powers.

In conclusion, after we finish the pre-processing of the data, we are supposed to get a list of two lists, mapping from ecoregion to a list of sublists. In our model, we let TSA frequency or TSA DHW to be independent variable X and average bleaching to be dependent variable Y. And the sublist will be in the form of [1, X, Y, X2 , XY, Y2].

In our main operation, We will first use the chi-square test for getting the relevance of temperature and percentage of coral bleaching. The chi-square is a well-known hypothesis test by using curve fitting using the standard formula: the number of degrees of freedom in the chi-square should be related to the number of locations. In this case, we use the following tabular to determine their relevance. Our goal is to determine whether there is a relationship between the existence of Thermal Stress Anomaly and Bleaching, so we may write function to get a result to determine their relevance.

Variable	Is TSA	No TSA	Totals
Bleaching Exists	a	b	a+b
Bleaching Do not Exist	c	d	c+d
Total	a+c	b+d	a+b+c+d=N

$$x^2 = \frac{(ad-bc)^2(a+b+c+d)}{(a+b)(c+d)(b+d)(a+c)}$$

Modeling Pictures:

- TSA_Frequency and TSA_DHW are two environmental ways to represent Thermal Stress Anomaly in a given region. However, we do not know for sure that which one has a more exact(clear) relationship between it and average bleaching, we will first plot out the points and try to find

some connection. Because locating at one specific independent variable of TSA, there exists many possibilities of average bleaching situation, so we take average of these possibilities and get a more general representation of bleaching under that specific thermal stress anomaly level.

- After ploting these points, we pick a representation to do the following work. We import numpy, plotly.graph_objects, matplotlib.pyplot to do the work for us. Numpy is a library that deals with array. We want to get a line that can best represent the relationship between these points. numpy.polyfit is a method to get the coefficient of the line of best fit. [m, b]=numpy.polyfit, m is the slope of the line and b is the intercept of x_axis and the line. After we get the coefficient of the line, we can estimate all coral bleaching percentages with TSA-frequency values. np.poly1d is a method that takes in the coefficient of the line and returns the figure. Plotly.graph_object as go: This can plot a figure with TSA-frequency as independent variable and bleaching percentage as dependent variable. matplotlib.pyplot: We use it to draw to figure. It takes the x and y coordinates and a line. We can also customize the format of the line. Also, plt.xlim and plt.ylim shows the range of x-coordinates and y-coordinates.

- Finally, we calculate regression and select the best model. R is a number that evaluates the relevance of our model and the actual data and we calculate R for different regions. After getting a relationship or a function, We may use regression to evaluate whether our model fits the actual data best using the formula $X^2 = \sum_{i=1}^{k} \frac{(x_i - m_i)^2}{m_i}$

Finally, we get an approximately picture of the relationship between TSA and temperature.

4 Instructions

Thanks to dear TAs for viewing our project, here are the instructions for obtaining datasets and running our program. I hope you will enjoy your work.

- First, download the dataset from the `https://www.bco-dmo.org/dataset/773466` (Note that choose to download csv file please)

- Then download the libraries listed in the requirement list.

- Download all of our python files we submitted and run the main.py.

- To get the result of chi-square, type result_chi to get the exact number of our chi-square calculation

- The plotted points of Average Bleaching with TSA_Frequency and TSA_DHW should come out instantly. You can view the the coordinates of the points when you put your mouse onto it. The x-axis represents TSA and the

y-axis represents average bleaching. The images you may see is inserted at the end of the report.

- You will also get a small window of Figure which show you a graph with plotted points and a line estimating the relationship between them. The meaning of it will be discussed in the discussion part. The image will also be inserted below.

- To get the model, you may type in model and you will get a string, representing the model we have obtained.

5 Changes we have made

- First, we clear out all the mistakes TAs have pointed out for us in the proposal.

- Another important change is that we decided not to clarify data using SSTA because from the research we have done the relationship seems to be not very clear. So we use TSA and our independent variable.

- A third change is that we decide to do data filtering and missing value implementation using scikit learn to fulfill the requirement of using more advanced third-party library.

6 Discussion

- First, we discuss the result of the chi-square, the result we get is 2.656213361039152. Considering the situation that at a specific time point, bleaching could be just happening or coral could have already died out, the part for (Bleaching Exists, IS TSA) and (Bleaching Do not Exist, No TSA) should be larger and as a result, chi-square could be larger in reality. And according to the chi-square table of values, when we set the degree of freedom to be one, we have approximately 90 per cent of chance to say that the existence of bleaching is related to existence of thermal stress anomaly.

- Second, we discuss the result of plotting the points, from the graph that is expected to see, we see that the graph with TSA_Frequency has an obvious increase of bleaching with the increase of TSA, however, when it comes to TSA_DHW, we find that the relationship between these variables is not obvious, the pattern without a specific rule may increase even more. So finally, we decide to choose TSA_Frequency as our independent variable to model our line. Simply from the data points, we can guess that these two variables have a positive correlation.

- Third, we try the relation between powers of independent and dependent variables and use our function to find the best model, (which means the model function with the least r_value) and we plot it in the figure 1. We

find that that the first part of the line is concrete and the latter is an imaginary line. This is mainly due to the reason that situations with such a high TSA_Frequency is not common, plus the fact that under these circumstances, bleaching may be so severe that most of the coral reefs may die out. This is a limitation due to the dataset we found, if possible, we would like to get more and more detailed data on this topic and finish drawing the line of relationship.

- Above all, in this project, we obtained a general relationship between coral bleaching percentage and Thermal Stress Anomaly and asserted that the higher the TSA, the more the bleaching will be. Also, with an r_value larger than 0.2, we have a relatively great prediction model in this part. This model can be used to predict how severe the bleaching may be with the measurement of TSA_Frequency and take measures to protect them before the bleaching got even worse. Also, our function of recording severity may help at this circumstance may as well, controlling bleaching at 0 and 1 is the best for our environment as well as for our humans.

- Finally, to summary, environment anomaly is indeed a severe problem that needs global attention. No matter it is the global warming that brings the temperature to go unreasonably high for corals to live, or the El nino that cause anomaly temperature for corals to live on continuously. Coral Reefs are fragile to environment changes, that's true, but us human, even the whole ecosystem may be more fragile, one missing coral could break the whole ecosystem in one go. It is our responsibility to save corals or even to save ourselves.

- For future exploration, we do not want to stop at a global scale, we want to go to more specific regions and use their data to predict the more exact relationship with TSA and Average Bleaching. Moreover, we want to step even further to discuss about predicting coral bleaching by modeling even more issues, like carbon dioxide level, sea temperature and wind speed. The more issue we take into consideration, the more closer the modeling could be to the reality, the more coral reefs could be saved.

- Thanks for reading! Why not join us and save the coral from bleaching together?

7 References

Bruno JF, Selig ER, Casey KS, Page CA, Willis BL, Harvell CD, et al. (2007) Thermal Stress and Coral Cover as Drivers of Coral Disease Outbreaks. PLoS Biol 5(6): e124.
https://doi.org/10.1371/journal.pbio.0050124

Donner SD, Rickbeil GJM, Heron SF (2017) A new, high-resolution global mass coral bleaching database. PLoS ONE 12(4): e0175490.

https://doi.org/10.1371/journal.pone.01

Eakin C.M., Lough J.M., Heron S.F., Liu G. (2018), Climate Variability and Change: Monitoring Data and Evidence for Increased Coral Bleaching Stress. In: van Oppen M., Lough J. (eds) Coral Bleaching. Ecological Studies (Analysis and Synthesis), vol 233. Springer, Cham. https://doi-org.myaccess.library.utoronto.ca/10.1007/978-3-319-75393-5_4

Ferreira, B.P., Costa, M.B.S.F., Coxey, M.S. et al (2013). The effects of sea surface temperature anomalies on oceanic coral reef systems in the southwestern tropical Atlantic. Coral Reefs 32, 441–454. https://link.springer.com/article/10.1007/s00338-012-0992-y

Heron et al. (2017). Impacts of climate change on world heritage coral reefs: A first global scientific assessment. Paris, France: UNESCO World Heritage Centre. http://whc.unesco.org/en/news/1676

Obura D (2017). Refilling the coral reefs glass. Science, 22 Sep 2017: Vol. 357, Issue 6357, pp. 1215, https://science.sciencemag.org/content/357/6357/1215/tab-article-info DOI: 10.1126/science.aao5002

Van Woesik, R., Burkepile, D. (2019), Bleaching and environmental data for global coral reef sites from 1998-2017. Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2019-07-18 [if applicable, indicate subset used]. https://www.bco-dmo.org/dataset/773466doi:10.1575/1912/bco-dmo.773466.

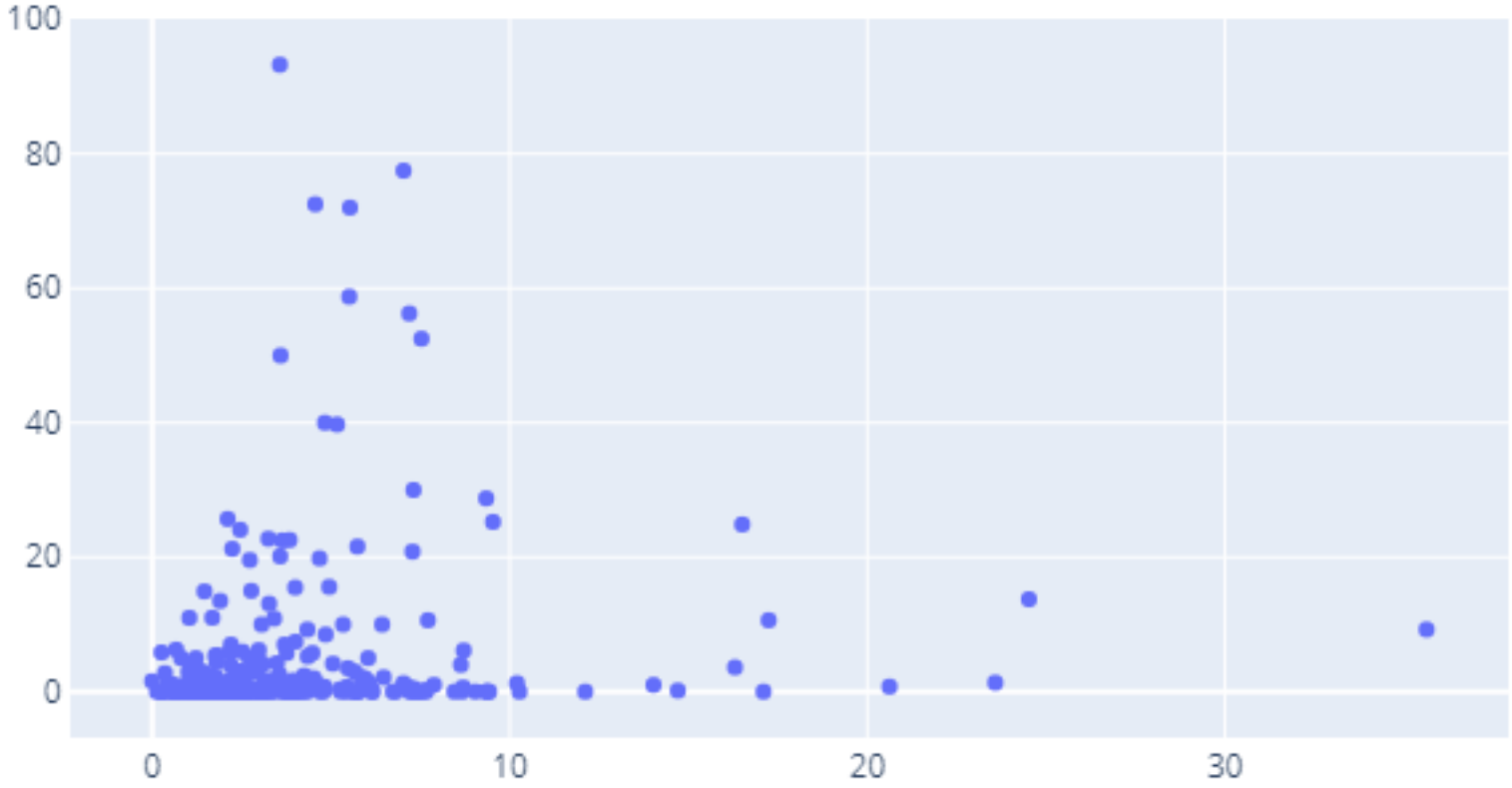

Figure 1: This is an image of ploting points for TSA_degree heating week with average bleaching.

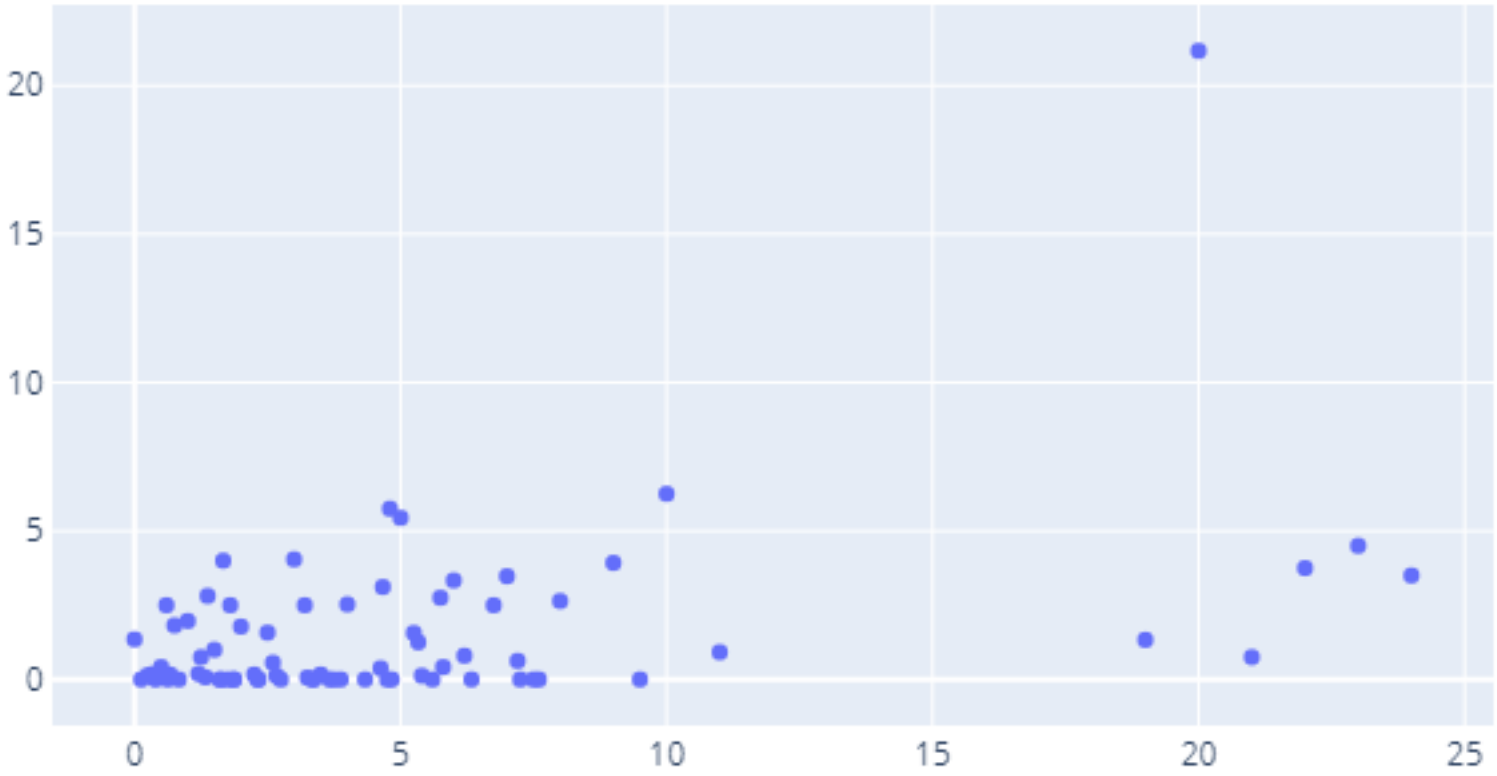

Figure 2: This is an image of ploting points for TSA_frequency with average bleaching.

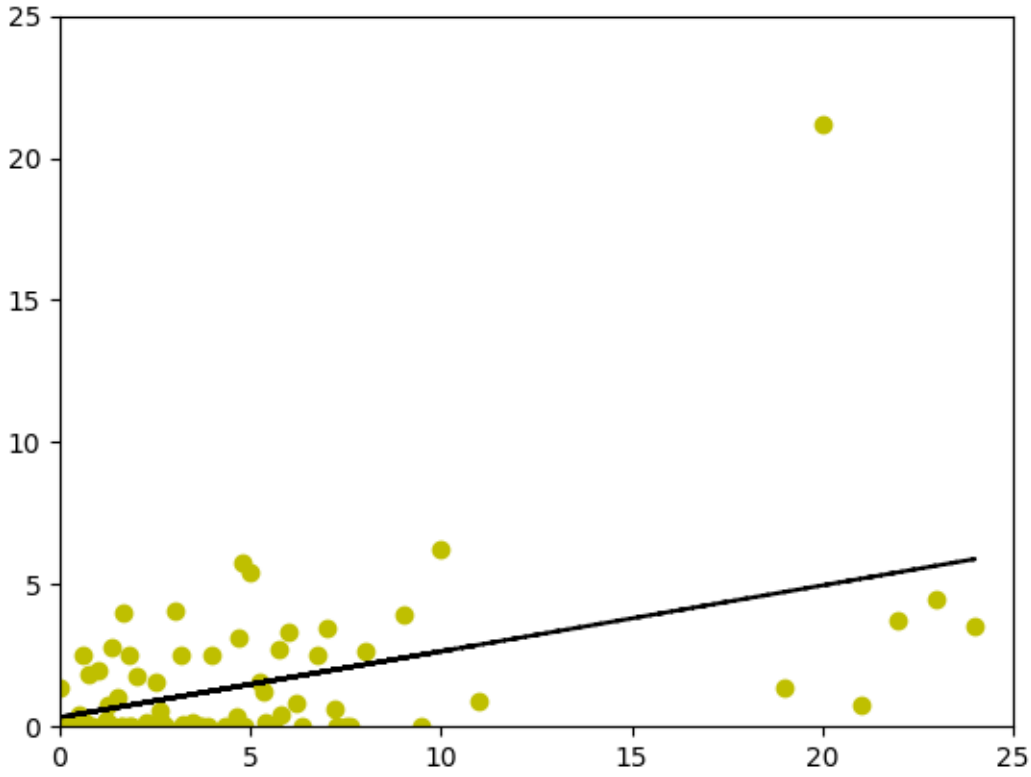

Figure 3: The is the model of pattern prediction we get