

# The Relationship Between The Number of Visitors and The Water Quality of Southern District Beaches

—based on the data from 2019 BWP report

## Group 1

YANG YANJIAO(Leader)

WANG YITING

CHEN YUQING

WANG CHANGHAO

YANG CHENXI

TIAN JINYUE

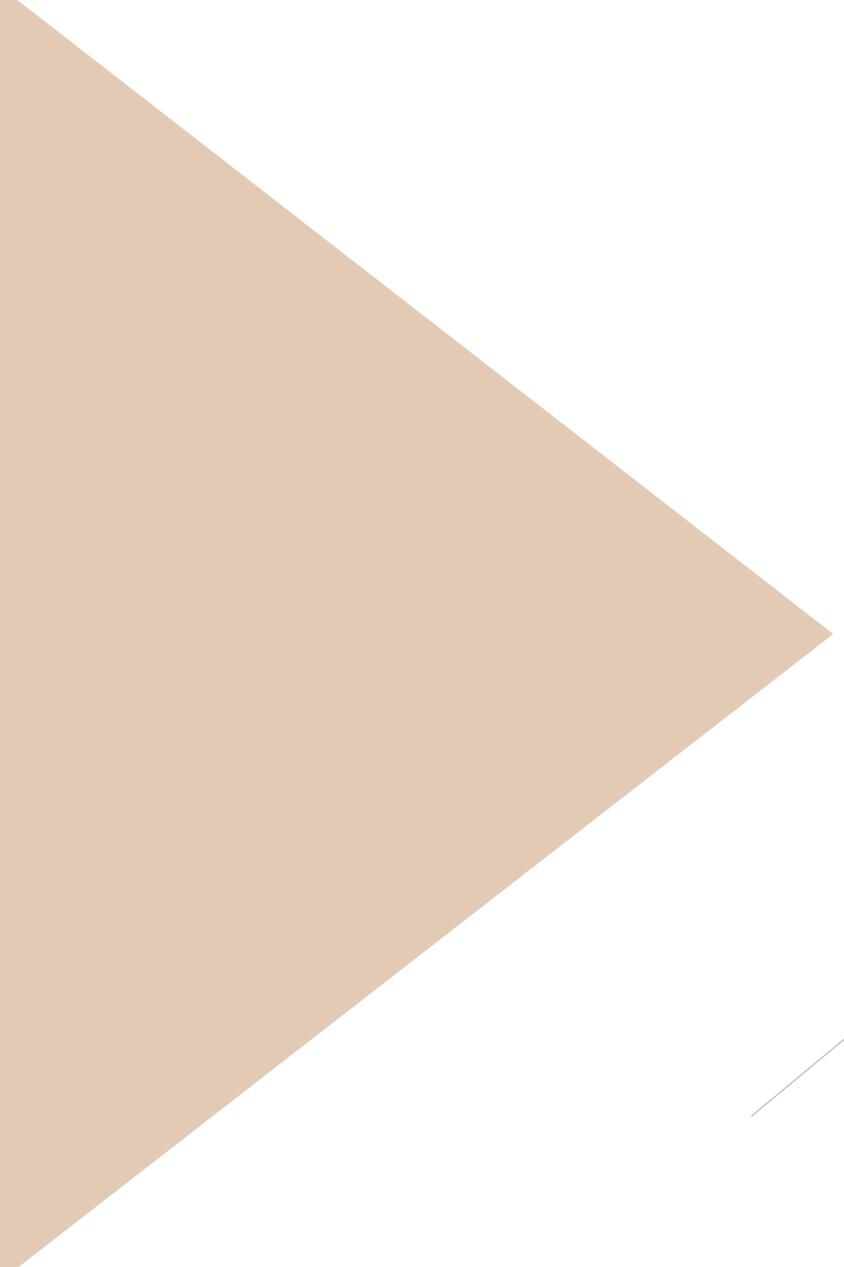
CUI MANQIAN

HU LEI

HUANG HAICHUAN

# CONTENTS

- 1 Problem Introduction  
Summaries & Importance
- 2 Quantitative Approaches  
Methods
- 3 Data Collection Commenting  
Source & Sample & Rationality
- 4 Exploratory Analysis' Ways  
Ways
- 5 Regression Model  
Procedure & Results
- 6 Assumptions' Verification  
Valid or Invalid
- 7 Conclusion  
Explanation & Recommendation

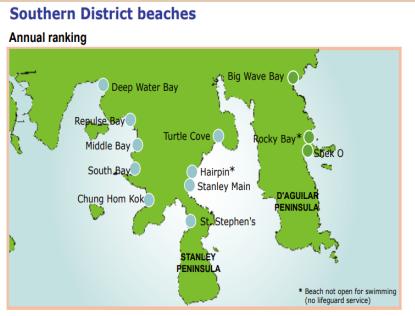


# PART 01

## Problem Introduction

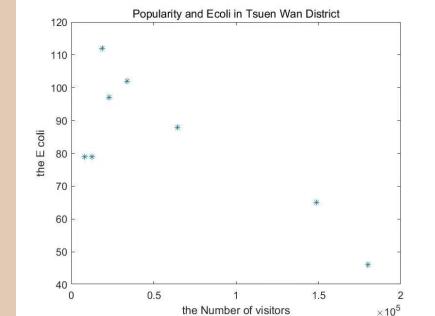
Summaries & Importance

# • Summaries •



## Goal

Southern District : Whether the number of visitors affects the water quality of beaches



## Analysis

Look for a linear relationship between water quality and number of visitors



## Scientific Method

- Quantitative Approach
- Data Collection & Analysis
- Linear Regression Model
- Assumptions
- Conclusions and explanations are based on the results
- Recommendations from ecological and economic aspect

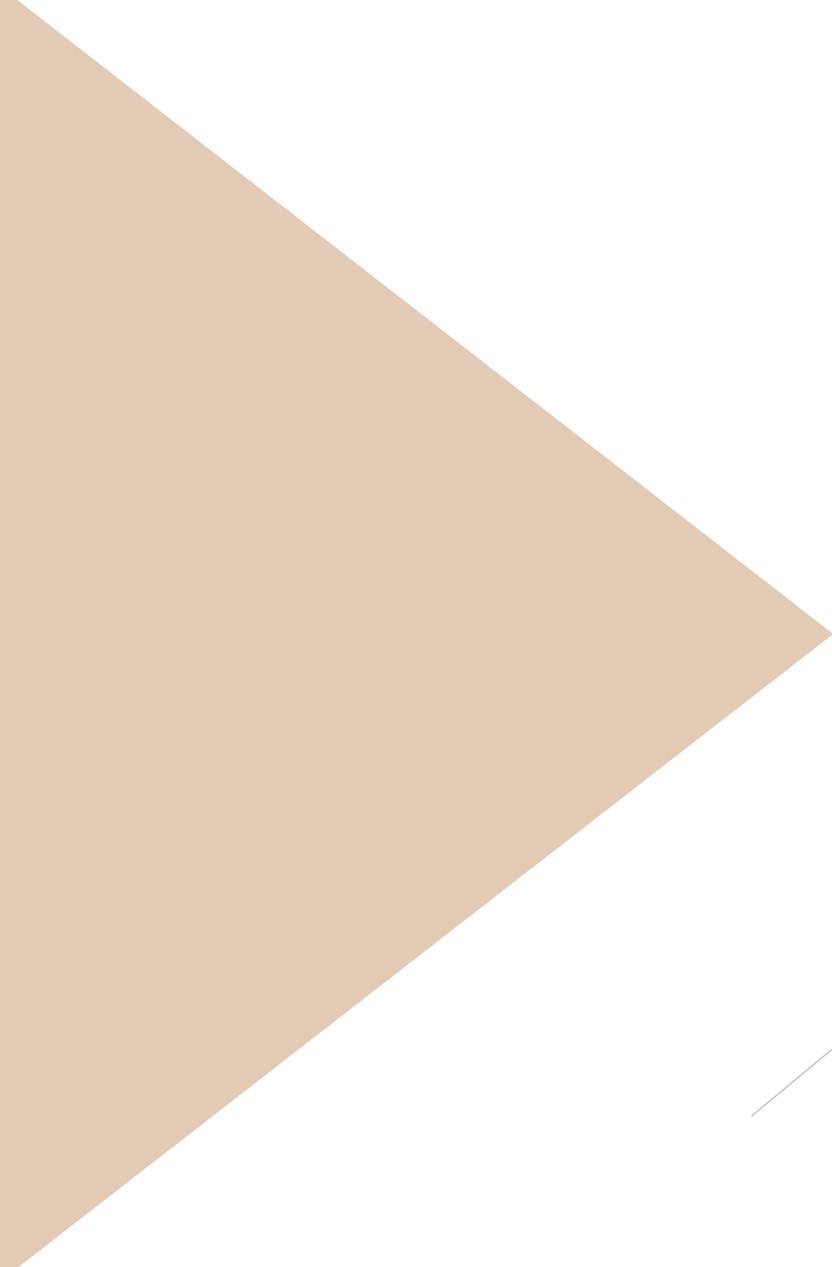


## Conclusion

# • Significance of studying the problem •

- Beach water quality: an issue for years
- Having clean bathing water is important to public health
- Government: To better improve water quality, ensure health of visitors and enhance the tourist experience
- Beach: A balance between visitors and water quality means long-term benefits both ecologically and economically
- A sustainable development maximizes profits





# PART 02

## Quantitative Approaches

Methods



# Preliminary Survey

At the very begining of the whole survey, we used some scientific methods to try to analyze current situation without any data management.

## Defining the problem

We define the problem as the possible relationship between the number of visitors and the water quality in the sea.



01

## Asking a question

Whether there is a relationship between the number of tourists and the quality of water at beaches



02

## Formulating hypotheses

According to the previous researches, We assume: if there are more and more tourists, the water quality will be much worse



03

## Operationalizing

- Number of visitors: can be accounted straightforwardly
- Water quality: level of E.coli



04

## Deducing a prediction



05

If our hypothesis is true, then we expect to see a positive correlation between these 2 factors in our data management.

# Data Management

01



Arithmetic mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Indicates where the centre of the water quality lies in the southern distract in 2019

02



Median

50<sup>th</sup> percentile

Illucidates the most common data among all the statistics

03



Geometric Mean

$$\bar{x} = \sqrt[n]{(x_1 \cdot x_2 \cdots \cdot x_n)}$$

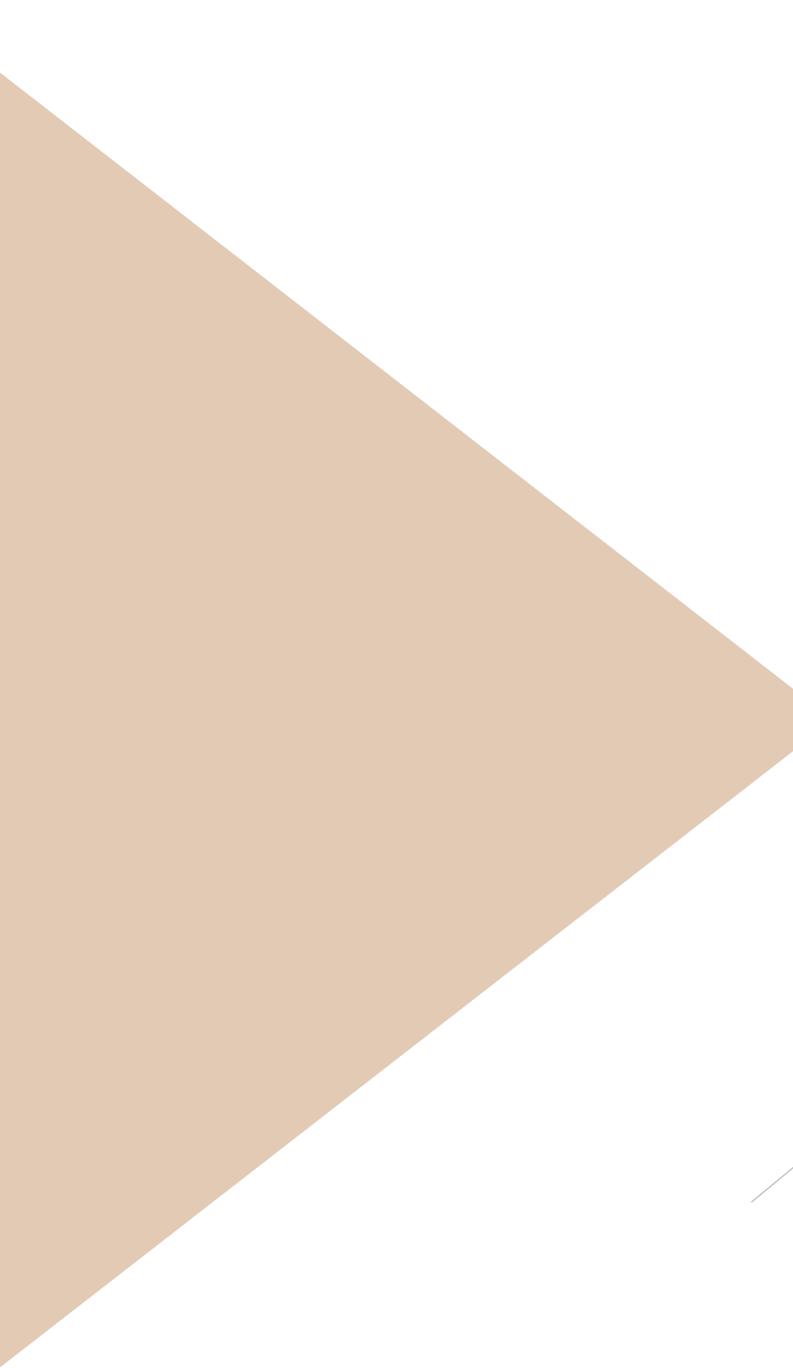
Used to represent the averaging water quality and the rate of change

04



Range and IQR

Range = maximum – minimum To see the most significant difference value



# PART 03

## Data Collection Commenting

Source & Sample & Rationality

# CONTENTS

1

Source of data

2

Rationality of data

# Source of data

( Retrospective study)



## Report from website

From the official website of Hong Kong, the ‘Beach Water Quality in Hong Kong’ report in 2019



## Appendix of the report

Tables in the appendix give us information such as the annual E. Coli level and visitor numbers



## Thesis from the internet

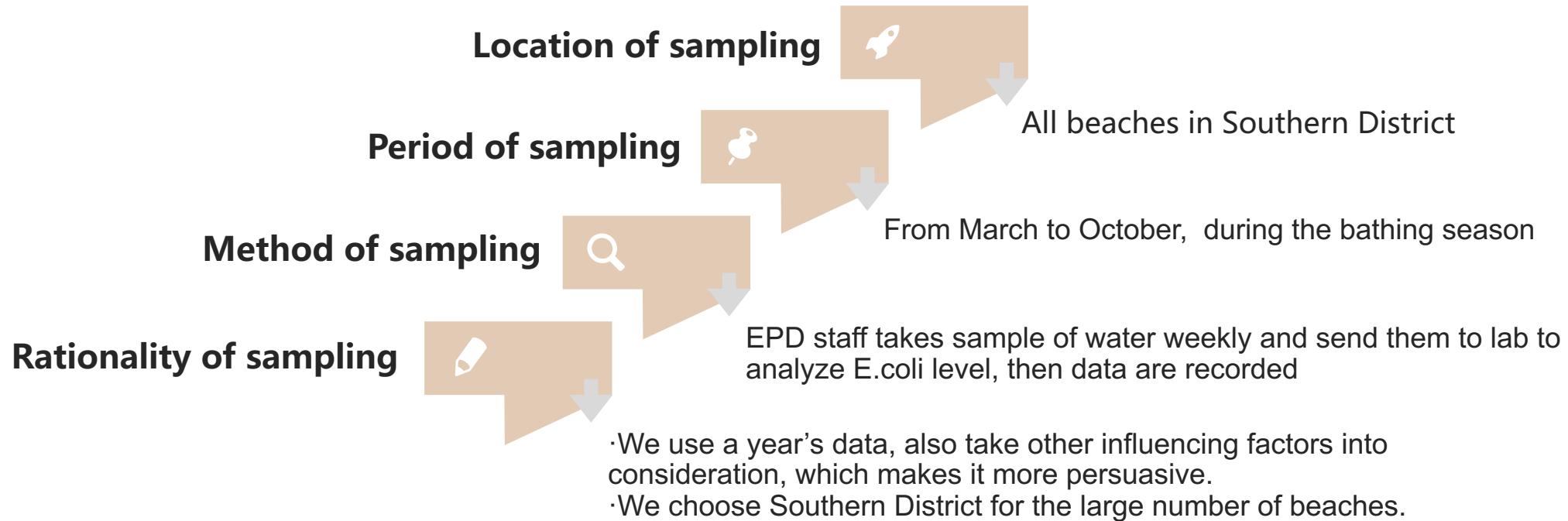
Learn the background knowledge of the relationship between E. coli and water quality



Beach Water Quality in Hong Kong 2019

Data in Southern District of 2019

# Rationality of data





## PART 04

### Initial exploratory analysis of data

Insights





## PART 04

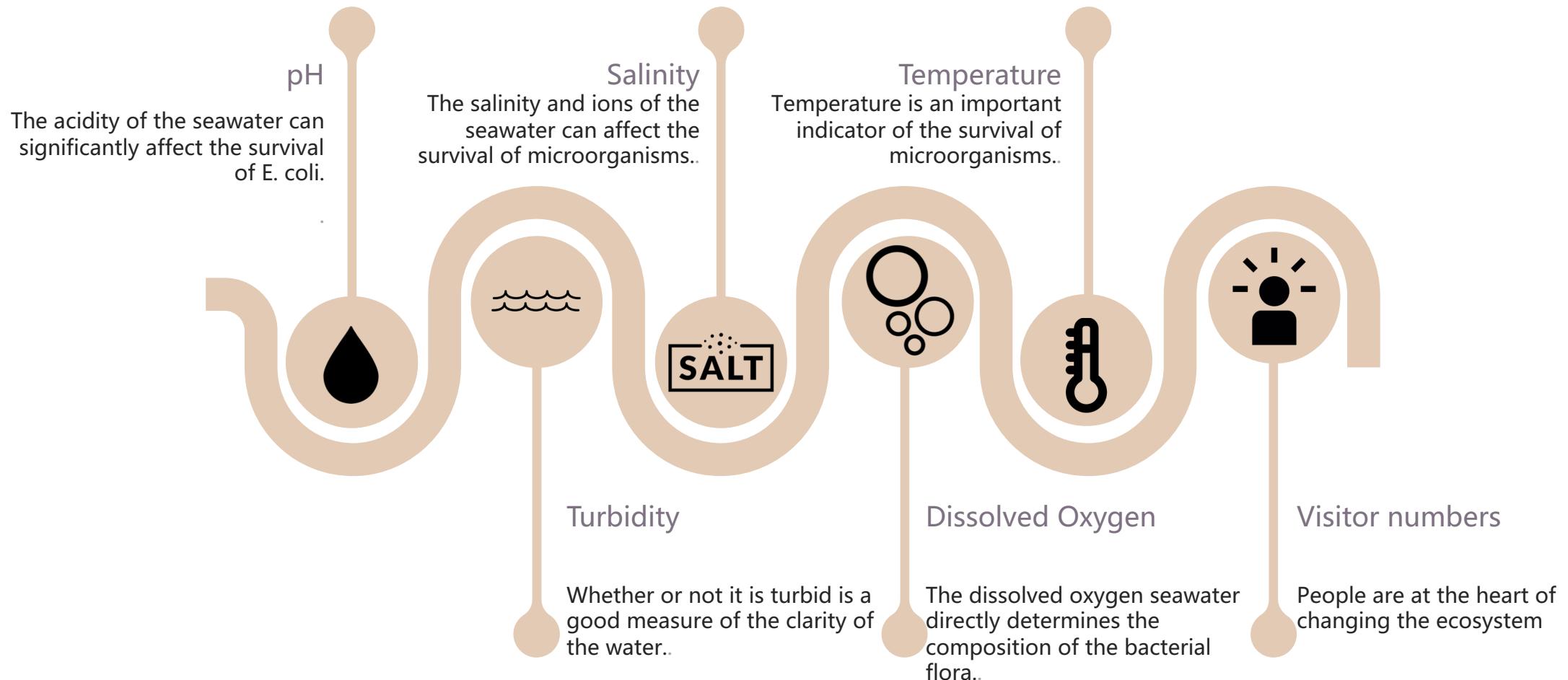
### Initial exploratory analysis of data

Insights

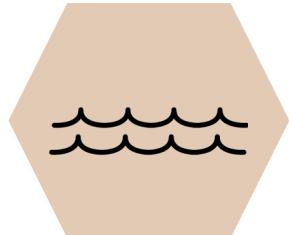


# Potential Impact Factors

Six factors that affect the growth of E. coli are presented



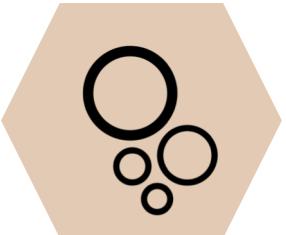
- Is it necessary to consider all the factors?



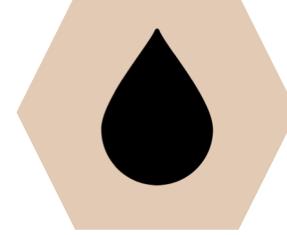
**Turbidity**



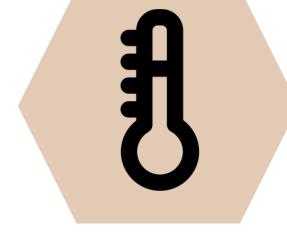
**Salinity**



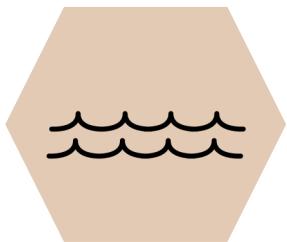
**Dissolved Oxygen**



**pH**



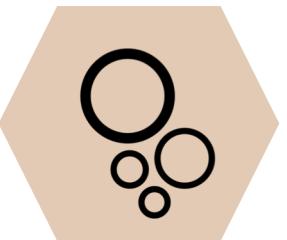
**Temperature**



## Turbidity



## Salinity



## Oxygen

# Turbidity, Salinity & Oxygen

- According to the reference, the turbidity is closely associated with E. coli
- As a class of microorganisms with high viability, E.coli can survive in water bodies with the turbidity of 1-56 NTU [1]
- Within in the range → ignore the factor

- Stable data in the report: all are round 30 PSU
- According to the reference, the salinity of the water body for E. coli is roughly between 30 and 36.6 PSU [2]
- Therefore, the water quality of Hong Kong provides the conditions for the growth of E. coli

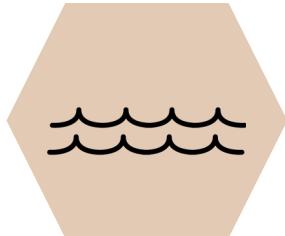
- Stable data in the report
- E. coli: facultative anaerobe [3]

## References:

- [1] Jozić, S., & Šolić, M. (2017). Effect of Environmental Conditions on Escherichia coli Survival in Seawater. *Escherichia Coli - Recent Advances on Physiology, Pathogenesis and Biotechnological Applications*. doi:10.5772/67912
- [2] Smith, R. P., Paiba, G. A., & Ellis-Iversen, J. (2008). Short Communication: Turbidity as an Indicator of Escherichia coli Presence in Water Troughs on Cattle Farms. *Journal of Dairy Science*, 91(5), 2082–2085. doi:10.3168/jds.2007-0597
- [3] Goldberg, J. J., Bramley, A. J., Sjogren, R. E., & Pankey, J. W. (1994). Effects of Temperature and Oxygen Tension on Growth of Escherichia coli in Milk. *Journal of Dairy Science*, 77(11), 3338–3346. doi:10.3168/jds.s0022-0302(94)77275-1

# Condition

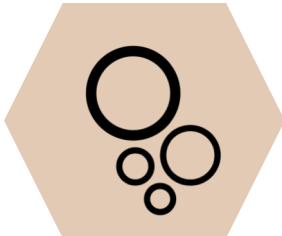
Is it necessary to consider all the factors?



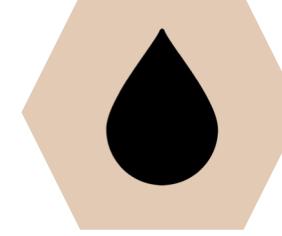
**Turbidity**



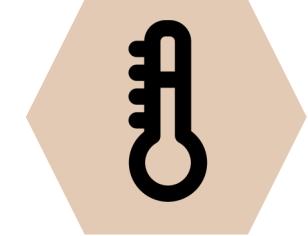
**Salinity**



**Dissolved Oxygen**



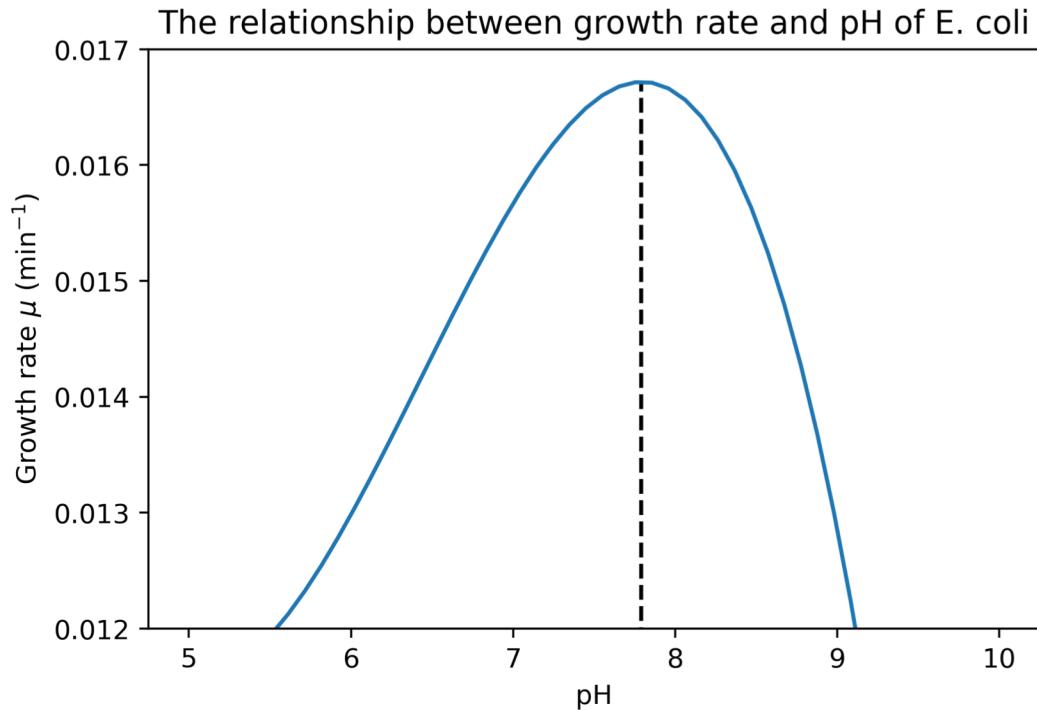
**pH**



**Temperature**

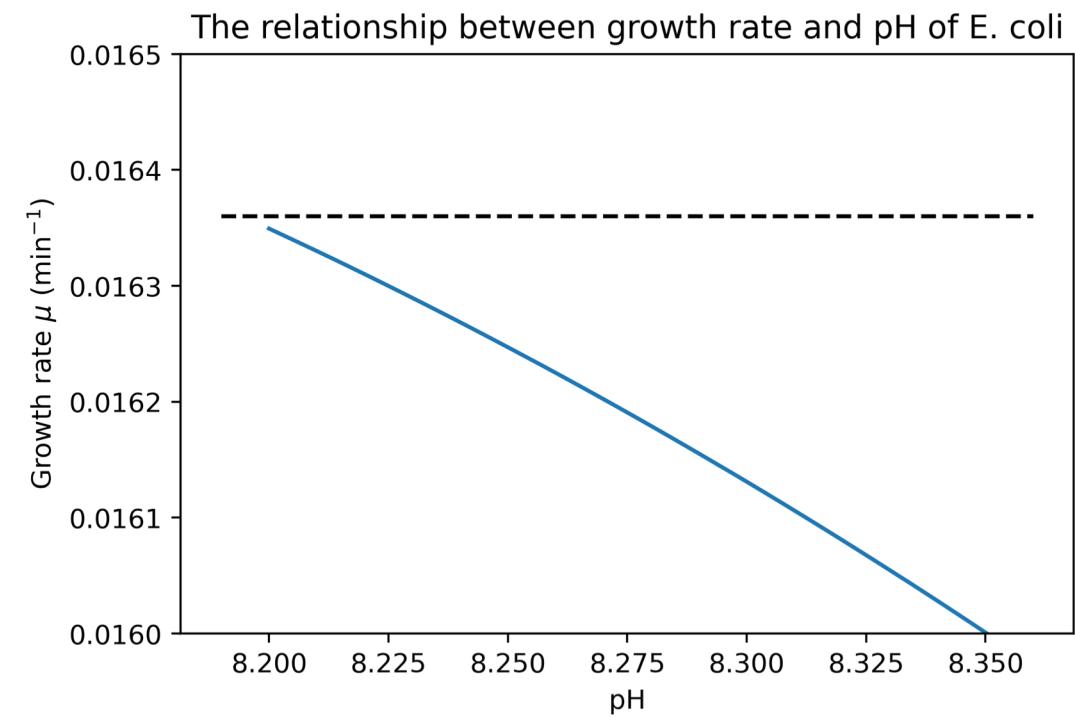
# pH

Does seawater acidity change substantially? Will this change affect the survival of E. coli in seawater?



Biological effects of pH on the growth rate of Escherichia coli.

- the pH varies little : between 8.2-8.35
- Ref: Presser, K. A., Ross, T., & Ratkowsky, D. A. (1998). Modelling the Growth Limits (Growth/No Growth Interface) of Escherichia coli as a Function of Temperature, pH, Lactic Acid Concentration, and Water Activity. *Applied and Environmental Microbiology*, 64(5), 1773–1779. doi:10.1128/aem.64.5.1773-1779.1998



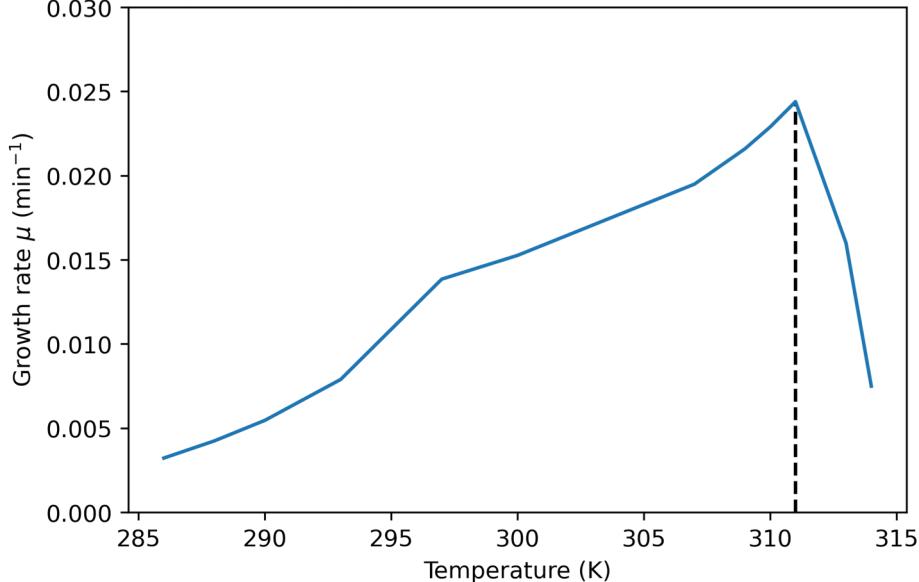
Effects of pH changes on Hong Kong beaches.

- the change in pH brings only 0.004 change in the growth rate.

# Temperature

Does seawater temperature change substantially? Will this change affect the survival of E. coli in seawater?

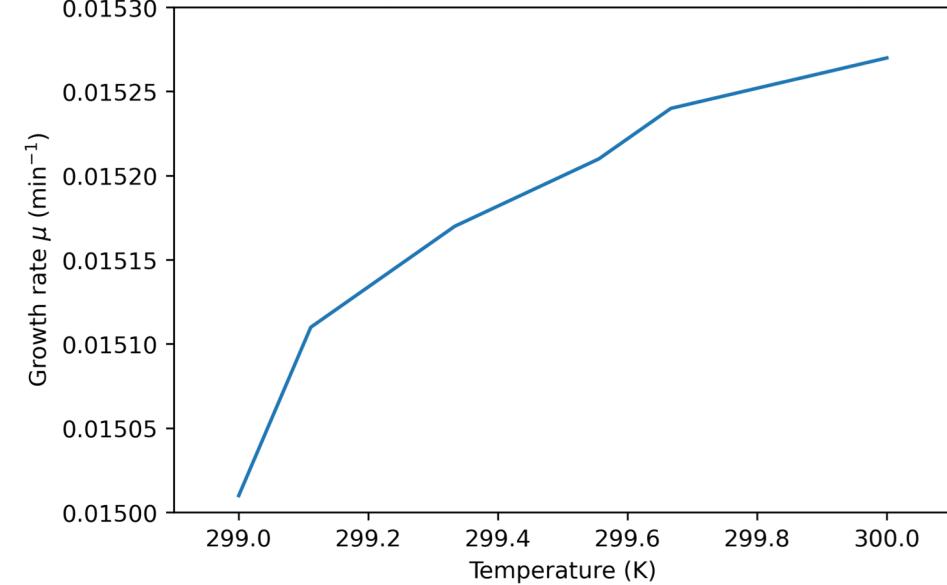
The relationship between growth rate and environment temperature of E. coli



Biological effects of temperature on the growth rate of E.coli

- The seawater temperature fluctuates between 29-30 degrees Celsius
- The maximum change to the growth rate is only 0.0003
- Ref: Li Xiangyang, Shao Weihua, Diao Enjie, Zhang Honglin. Calorimetric study of temperature, pH and drug's inhibitory effect on Escherichia coli [J]. Food Science, 2007(06): 252-255

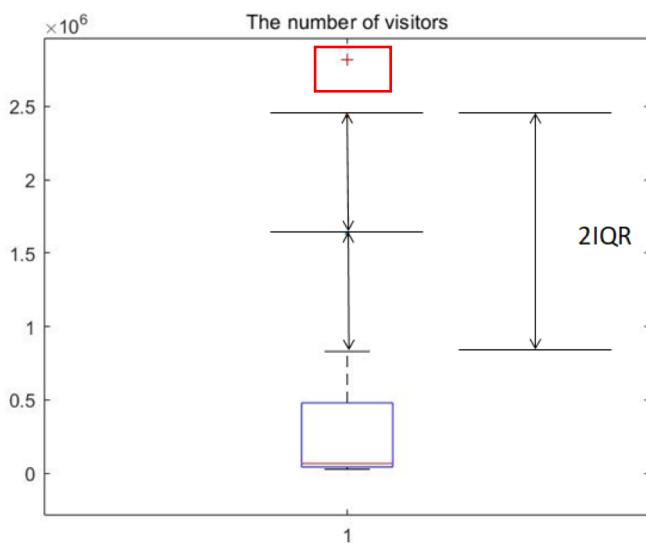
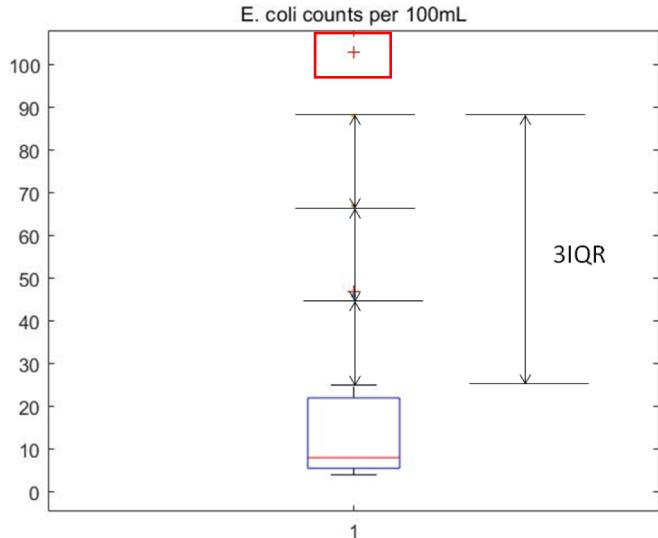
The relationship between growth rate and environment temperature of E. coli



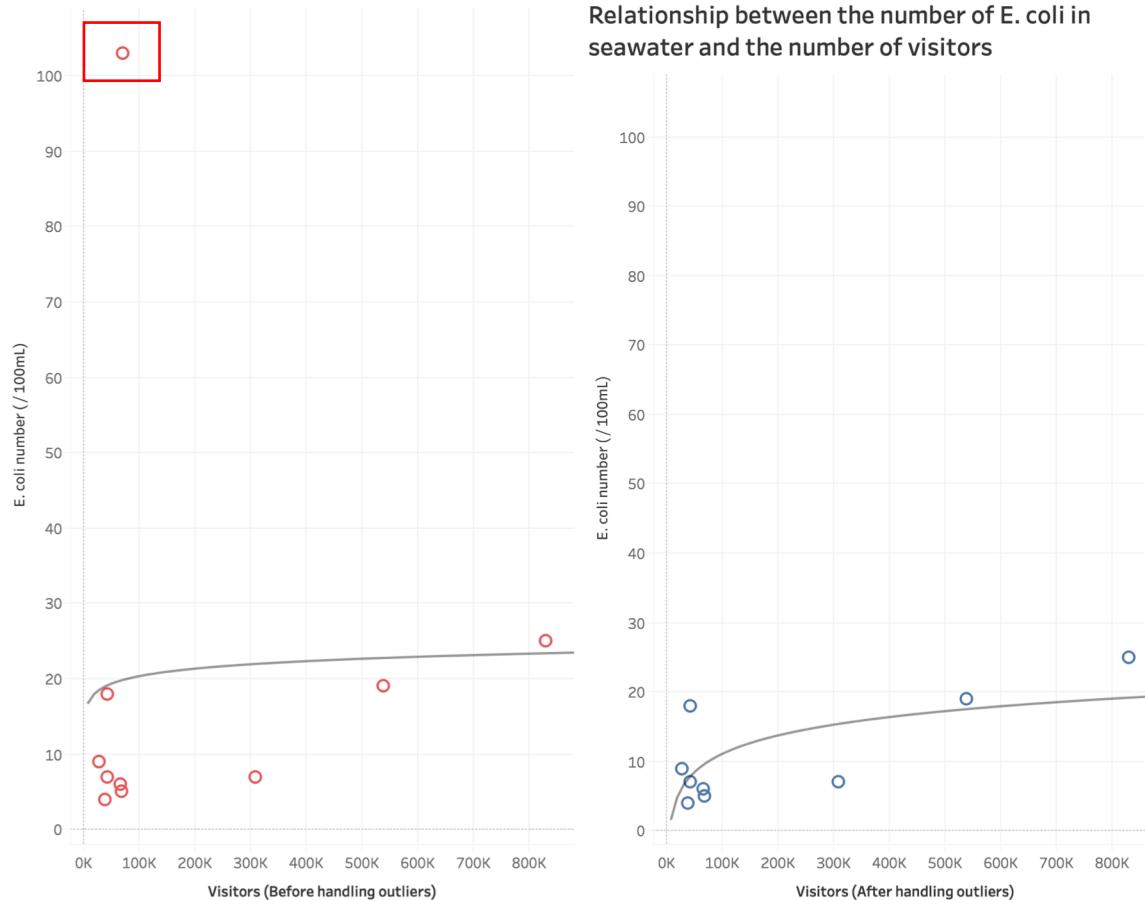
Effects of temperature changes on Hong Kong beaches

## • Data pre-processing •

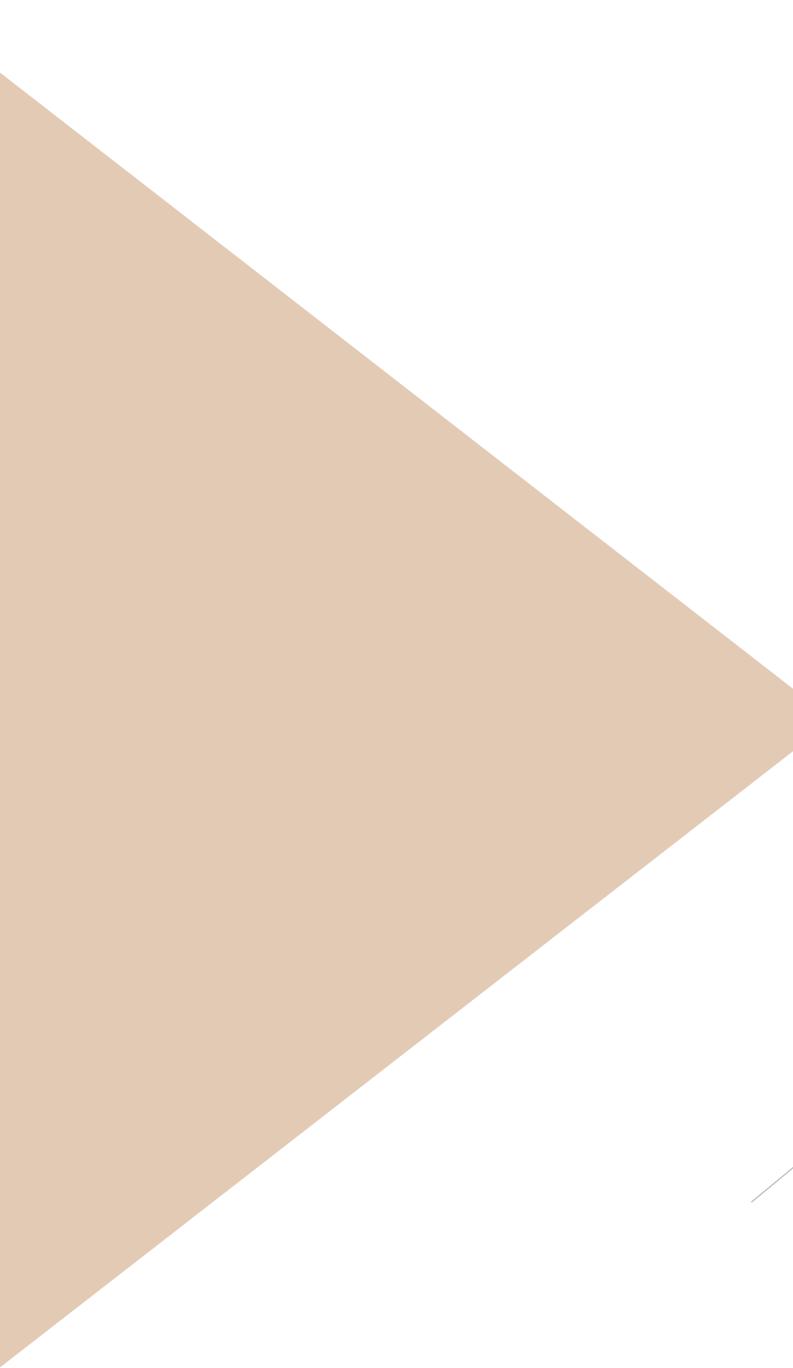
Is it necessary to remove outliers?



The outlier analysis



Comparison of before and after removal



# PART 05

## Regression Model

Procedure & Results

# • Modelling •

- **Variable : X**
- **Y**
- **Least squares**

Visitor numbers during the bathing season in different beaches of Southern District 2019

Annual geometric mean E. coli level in different beaches of Southern District 2019

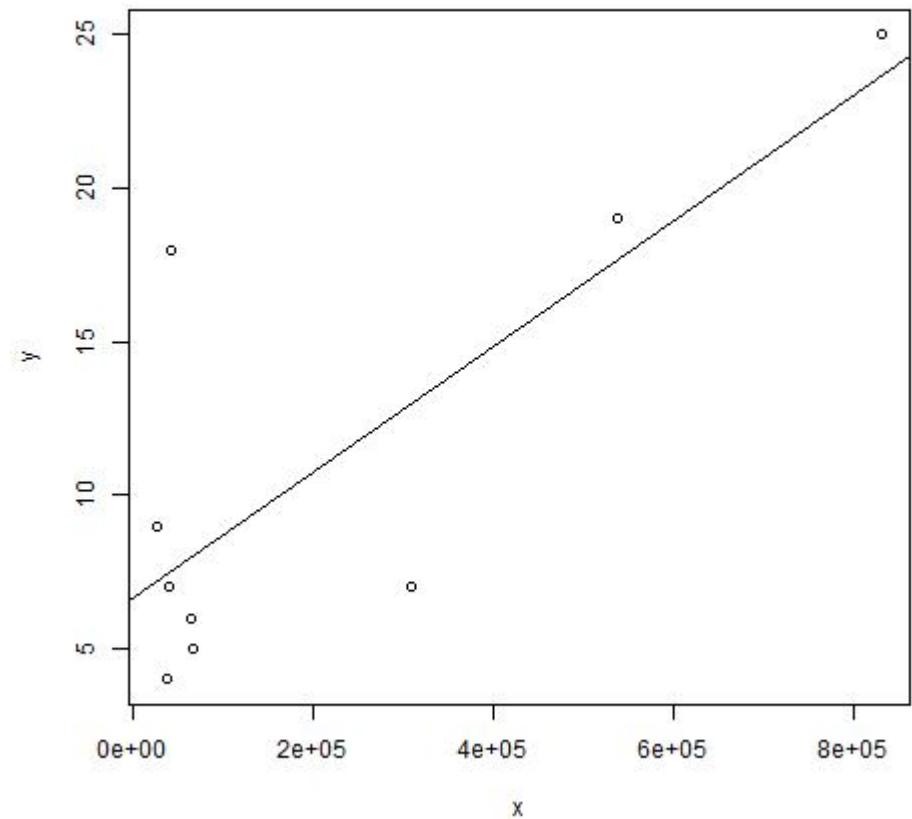
The “best-fit” line is defined as the one that minimises the sum of squared errors, i.e.  $\sum_{i=1}^n \varepsilon_i^2$

# • Modelling •

Using the method of least squares, we find the best-fit line through the sample.

Using R, we have:

```
Call:  
lm(formula = y ~ x, data = data_res)  
  
Residuals:  
    Min      1Q  Median      3Q     Max  
-5.9658 -3.0335 -0.5079  1.3482 10.4832  
  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) 6.653e+00 2.154e+00  3.089  0.0176 *  
x           2.048e-05 6.195e-06  3.306  0.0130 *  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 5.037 on 7 degrees of freedom  
Multiple R-squared:  0.6095,   Adjusted R-squared:  0.5538  
F-statistic: 10.93 on 1 and 7 DF,  p-value: 0.01302
```

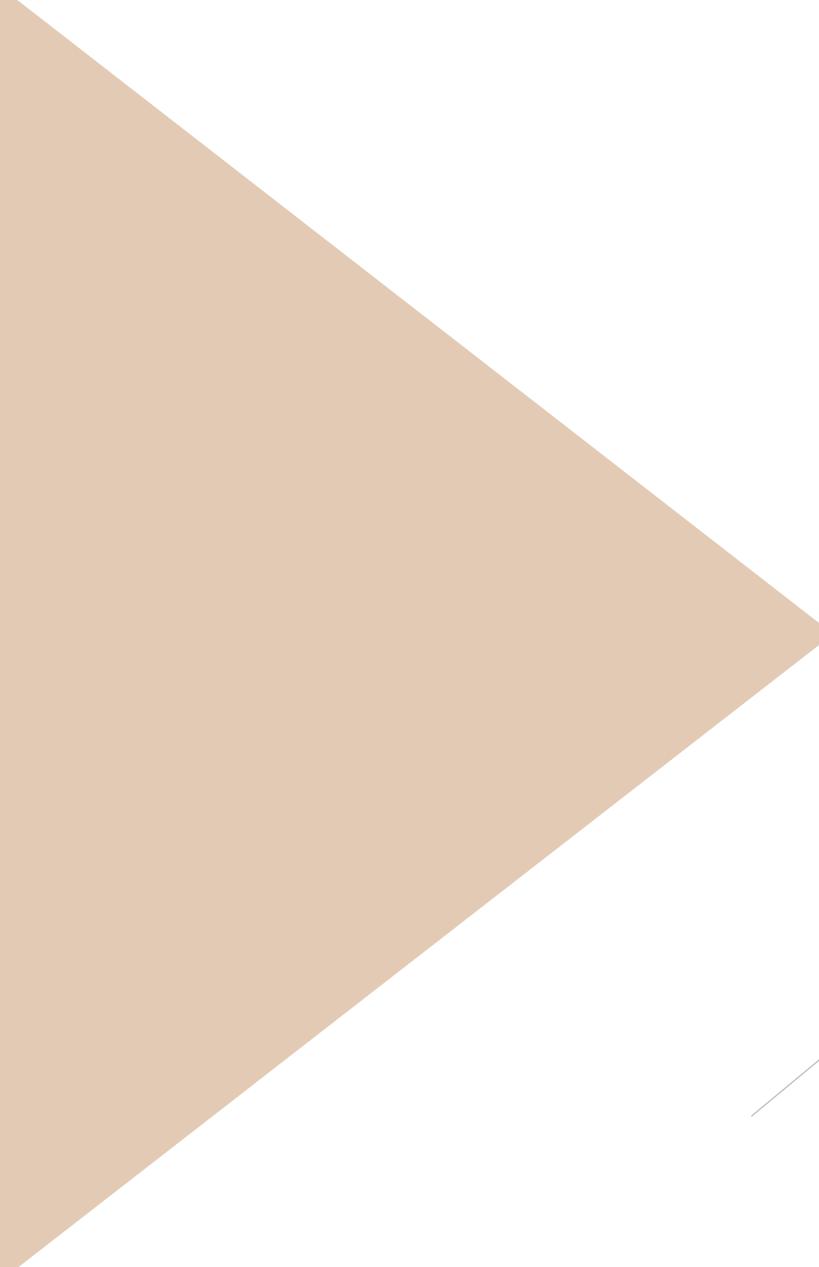


**Result :**  $y = 2.048\text{e-}5x + 6.653 + \varepsilon$

---

• Analysis •

- $H_0$  There is no relationship between water quality and the popularity of the beach
  - $H_1$  There is a relationship between water quality and the popularity of the beach
- Result:** Because the p-value is 0.0132 , we have more than 95% confidence to reject  $H_0$



# PART 06

## Assumptions' Verification

Valid or Invalid

# Assumptions

Error term is normally distributed

Observation of the error term are uncorrelated with each other  
Error term: 2, -1.3, 0.5, -10, -1.3, 3.5, 3, 6, -2

The error term has a population mean of 0 and has a constant variance

The model is linear in the coefficients  $y = mx + c + \varepsilon$

The independent variable, X, is uncorrelated with the error term,  $\varepsilon$



# PART 07

## Conclusion

Explanation & Recommandation

# Conclusion

```
call:  
lm(formula = y ~ x, data = data_res)  
  
Residuals:  
    Min      1Q  Median      3Q     Max  
-5.9658 -3.0335 -0.5079  1.3482 10.4832  
  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) 6.653e+00 2.154e+00 3.089  0.0176 *  
x           2.048e-05 6.195e-06 3.306  0.0130 *  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 5.037 on 7 degrees of freedom  
Multiple R-squared:  0.6095,   Adjusted R-squared:  0.5538  
F-statistic: 10.93 on 1 and 7 DF,  p-value: 0.01302
```

## The result of linear regression model (2019)

```
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) 6.744e+00 1.610e+00 4.188 0.000286 ***  
x           1.602e-05 4.836e-06 3.311 0.002730 **  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 6.656 on 26 degrees of freedom  
Multiple R-squared:  0.2966,   Adjusted R-squared:  0.2696  
F-statistic: 10.97 on 1 and 26 DF,  p-value: 0.00273
```

## The result of linear regression model (2017-2019)

### Basic Conclusions and explanations:

- The number of visitors affects the water quality of beaches in Southern District
- More visitors will worsen the water quality
- This is likely due to the human activities on the beach

### Further Analysis

- Why don't we use data from 2017-2019?
- Both results support the hypothesis

# Conclusion

- Other factors have influence on the water quality
  - Rainfall
  - Sewage emissions
  - Sunlight
  - Policy
- A balance between visitors and water quality will make a sustainable development of beaches
  - Ecological environment
  - Economic profits



- Strength of our work
  - Preliminary research
  - Data cleaning: outliers
  - Validity of assumptions
  - Convincing results
- Weakness & Promotions
  - Data collection: retrospective study
  - Model

# THANKS FOR YOUR LISTENING

