

# Aquaculture in Vietnam

Minh Chau

2025-01-13

```
#this is a data base about value of aquatic products per hectare. time from 2016 to 2022.
#link for this data: https://www.gso.gov.vn/px-web-2/?pxid=V0603&theme=N%C3%B4ng%2C%20l%C3%A2m%20nghi%E
# Read the Excel file
data_01 <- read_excel("C:/Users/ngoti/Desktop/Economic in Aquaculture/file_01.xlsx")

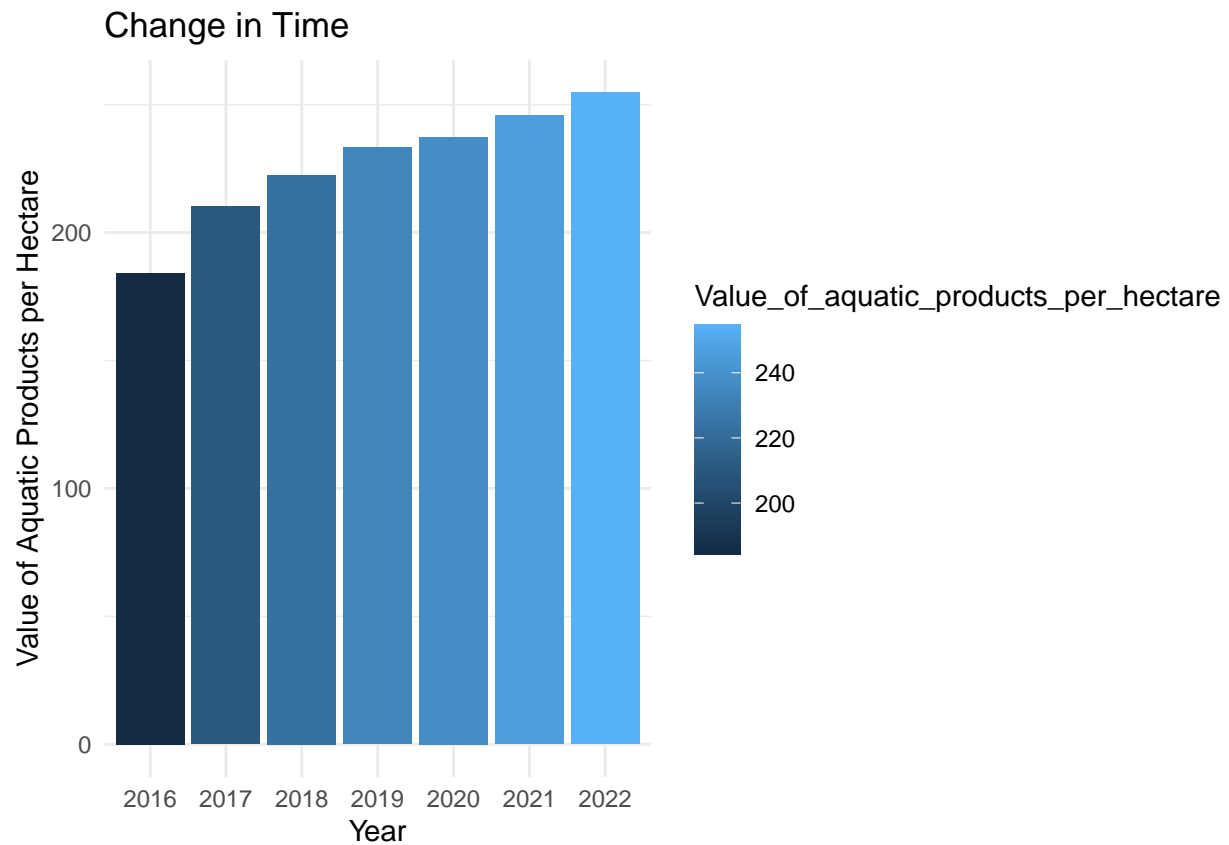
## New names:
## * `` -> `...2`
## * `` -> `...3`

# Subset the data: Select rows where the second column (year) is >= 2016 and choose relevant columns
data_01_c <- data_01 %>%
  filter(...2 >= 2016) %>%
  select(...2, ...3)

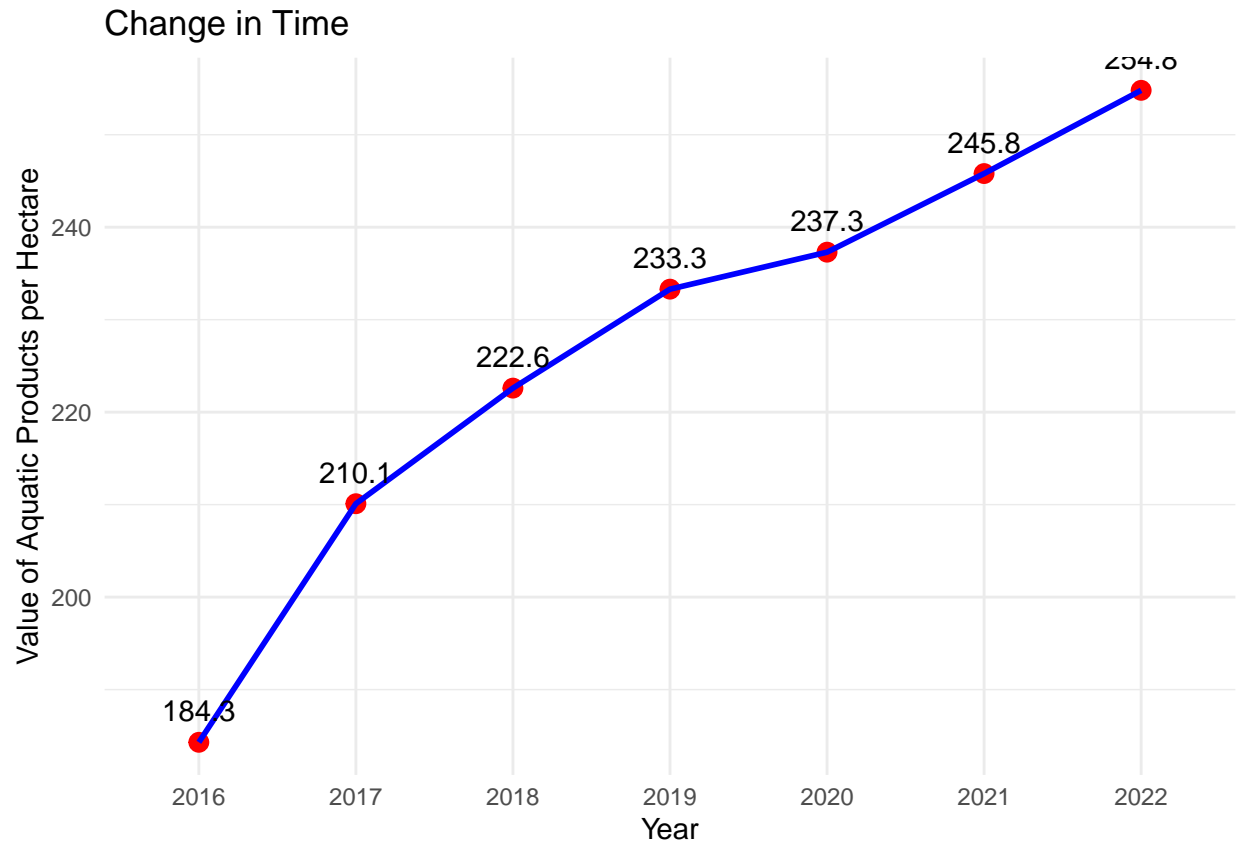
# Rename columns for clarity
colnames(data_01_c) <- c("year", "Value_of_aquatic_products_per_hectare")

# Convert the value column to numeric
data_01_c$Value_of_aquatic_products_per_hectare <- as.numeric(data_01_c$Value_of_aquatic_products_per_h

# --- Create Bar Plot ---
ggplot(data_01_c, aes(x = year, y = Value_of_aquatic_products_per_hectare, fill = Value_of_aquatic_produ
  geom_bar(stat = "identity") +
  labs(
    title = "Change in Time",
    x = "Year",
    y = "Value of Aquatic Products per Hectare"
  ) +
  theme_minimal()
```



```
# --- Create Line Plot ---
ggplot(data_01_c, aes(x = year, y = Value_of_aquatic_products_per_hectare, group = 1)) +
  geom_point(size = 3, color = "red") +
  geom_line(linewidth = 1, color = "blue") +
  geom_text(aes(label = Value_of_aquatic_products_per_hectare), vjust = -1, color = "black") + # Add labels
  labs(
    title = "Change in Time",
    x = "Year",
    y = "Value of Aquatic Products per Hectare"
  ) +
  theme_minimal()
```



```
# --- Compare Each Year with Previous Year (2016 as 100%) ---
data_01_c <- data_01_c %>%
  mutate(increase_lvl = ifelse(row_number() == 1, 1, Value_of_aquatic_products_per_hectare / lag(Value_of_aquatic_products_per_hectare)))

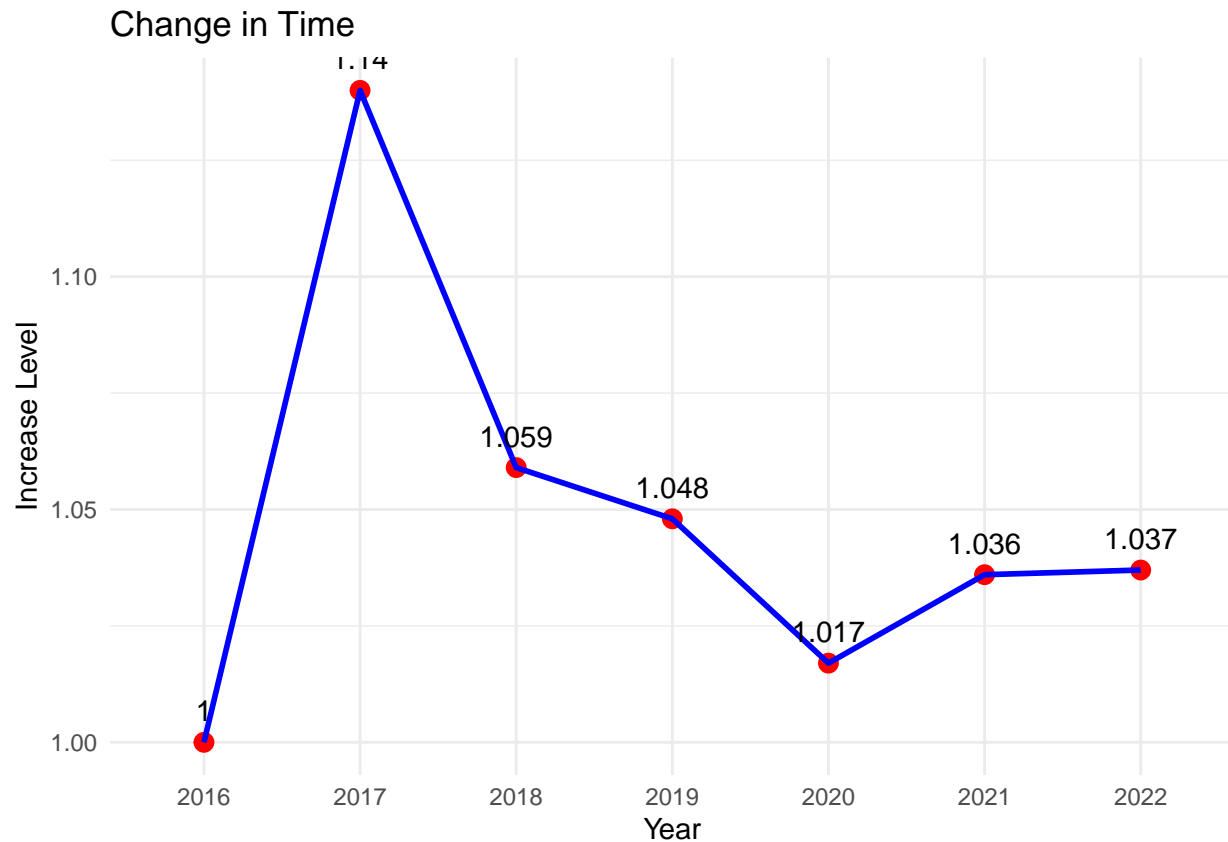
# Convert increase level to formatted numeric values
data_01_c$increase_lvl <- round(as.numeric(data_01_c$increase_lvl), 3)

# Print updated dataset
print(data_01_c)
```

```
## # A tibble: 7 x 3
##   year Value_of_aquatic_products_per_hectare increase_lvl
##   <chr>                                <dbl>         <dbl>
## 1 2016                                184.           1
## 2 2017                                210.          1.14
## 3 2018                                223.          1.06
## 4 2019                                233.          1.05
## 5 2020                                237.          1.02
## 6 2021                                246.          1.04
## 7 2022                                255.          1.04
```

```
# --- Create Line Plot for Increase Level ---
ggplot(data_01_c, aes(x = year, y = increase_lvl, group = 1)) +
  geom_point(size = 3, color = "red") +
  geom_line(linewidth = 1, color = "blue") +
```

```
geom_text(aes(label = increase_lvl), vjust = -1, color = "black") + # Add labels
labs(
  title = "Change in Time",
  x = "Year",
  y = "Increase Level"
) +
theme_minimal()
```



#Summary chunk 1 Nam nao gia tri tren 1 hecta dat nuoi trong thu hoạch thuy san deu tang. Cong thuc tinh la tu san luong tinh ra tien roi chia cho dien tich. Vay thi no se phu thuoc vao 2 yeu to do la san luong thu duoc va dien tich.

Cau hoi dat ra: Voi gia tri deu tang nhu vay thi san luon va dien tich thay doi nhu the nao?

```
# insert data from this link: https://www.gso.gov.vn/pv-web-2/?pxid=V0656&theme=N%C3%B4ng%20%20l%C3%A2m
```

```
#from 2022, the survey is count on the area that harvest
```

```
# Read the Excel file
```

```
data_02 <- read_excel("C:/Users/ngoti/Desktop/Economic in Aquaculture/file_02.xlsx")
```

```
## New names:
```

```
## * `` -> `...2`
```

```
## * `` -> `...3`
```

```
## * `` -> `...4`
```

```
## * `` -> `...5`
```

```
## * `` -> `...6`
## * `` -> `...7`
## * `` -> `...8`
```

```
# Select the relevant rows and columns
data_02_c <- data_02[2:12, 1:8]

# Rename the first cell to "Year"
data_02_c[1, 1] <- "Year"

# Set column names
colnames(data_02_c) <- c("type", 2016:2022)

# Convert text to ASCII format for consistency
data_02_c$type <- stri_trans_general(data_02_c$type, "Latin-ASCII")

# Rename "Year" to avoid conflicts
data_02_c <- data_02_c %>%
  mutate(type = ifelse(type == "Year", "Year_Column", type))

# Convert 'type' column into row names, then transpose the data
data_02_c <- data_02_c %>%
  column_to_rownames(var = "type") %>%
  t() %>%
  as.data.frame() %>%
  rownames_to_column(var = "Year")

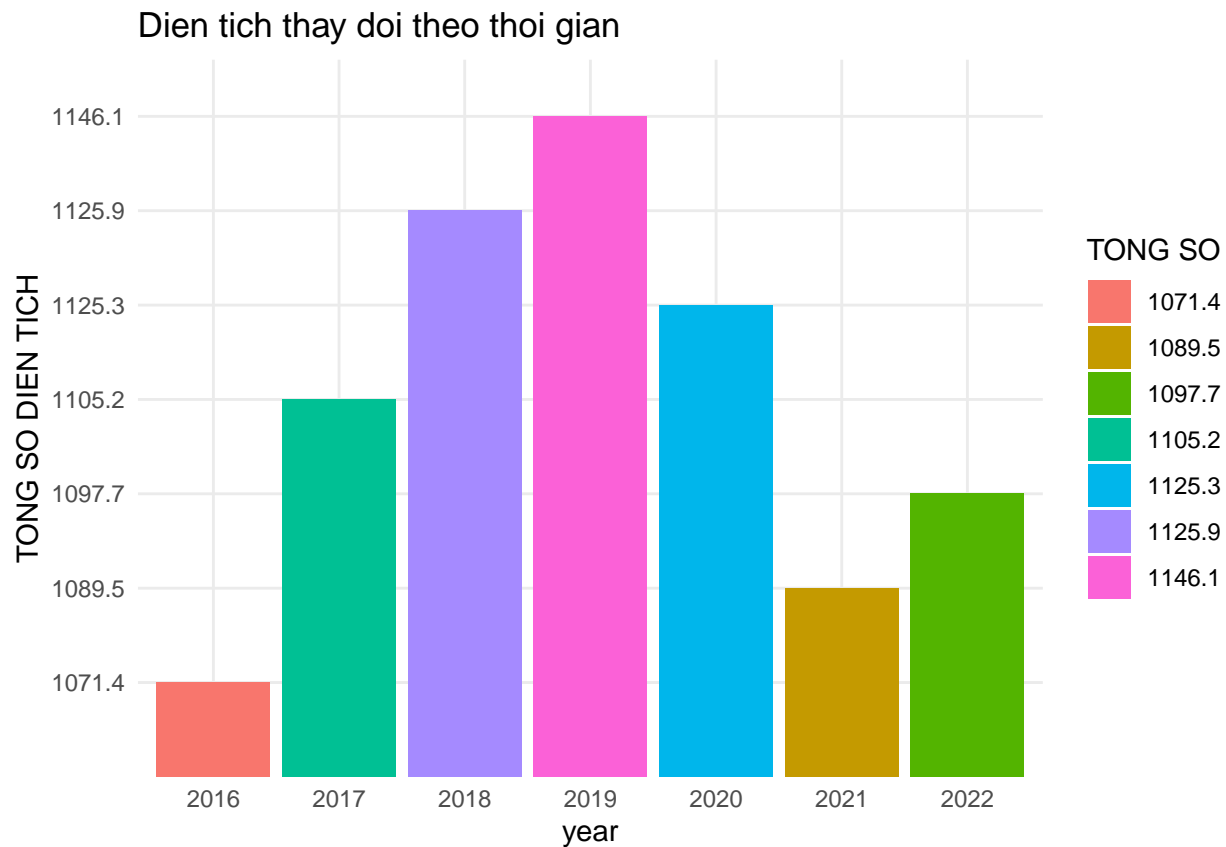
# Remove "Year" column
data_02_c <- select(data_02_c, -Year)
colnames(data_02_c)[3] <- "Thuy san bien"

# Print final cleaned dataset
print(data_02_c)
```

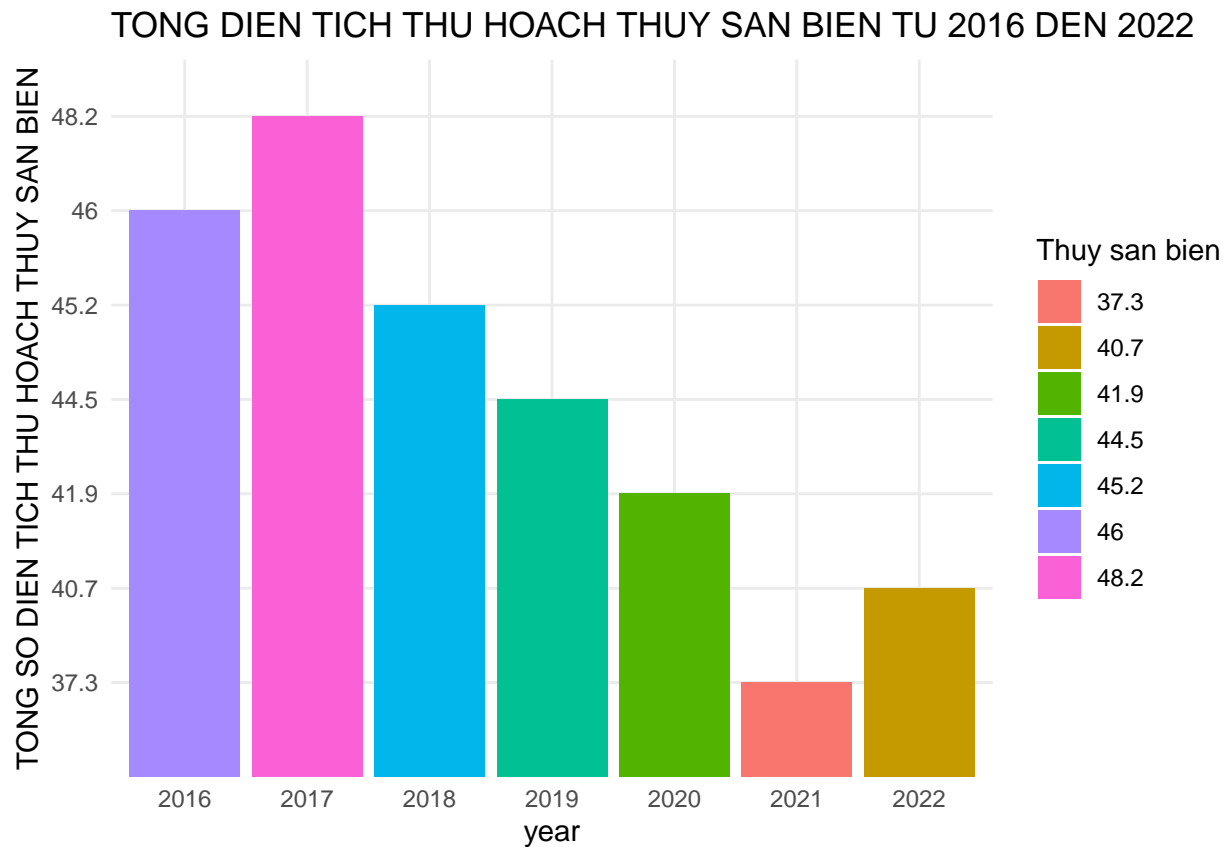
```
##   Year_Column TONG SO Thuy san bien Nuoi ca (thuy san bien)
## 1      2016   1071.4          46              0.6
## 2      2017   1105.2          48.2             0.6
## 3      2018   1125.9          45.2             0.5
## 4      2019   1146.1          44.5             0.6
## 5      2020   1125.3          41.9             0.5
## 6      2021   1089.5          37.3             0.4
## 7      2022   1097.7          40.7             1.1
##   Nuoi tom (thuy san bien) Nuoi hon hop va thuy san khac (thuy san bien)
## 1                3.4                42
## 2                2.5                45.1
## 3                2.4                42.3
## 4                2.4                41.5
## 5                2.5                38.9
## 6                0.3                36.6
## 7                1.9                37.7
##   Thuy san noi dia Nuoi ca (thuy san noi dia) Nuoi tom (thuy san noi dia)
## 1          1020.5                321.6                686.5
## 2          1052.4                320.4                720.7
## 3          1072.8                327.2                734.7
```

```
## 4          1093.5          331.9          749.9
## 5          1076.4          326.7          736.2
## 6          1052.2          316.6          718.5
## 7           1057          326.7          712.8
##   Nuoi hon hop va thuy san khac (thuy san noi dia)
## 1                                     12.4
## 2                                     11.3
## 3                                     10.9
## 4                                     11.7
## 5                                     13.5
## 6                                     17.1
## 7                                     17.5
##   Dien tich uom, nuoi giong thuy san
## 1                                     4.9
## 2                                     4.6
## 3                                     7.9
## 4                                     8.1
## 5                                     7
## 6                                     ..
## 7                                     ..
```

```
#creat a bar plot for TONG SO
ggplot(data_02_c,aes(x=Year_Column,y = `TONG SO`,fill = `TONG SO`)) +
  geom_bar(stat = "identity") +
  labs(title = "Dien tich thay doi theo thoi gian",
    x ="year",
    y = "TONG SO DIEN TICH") +
  theme_minimal()
```

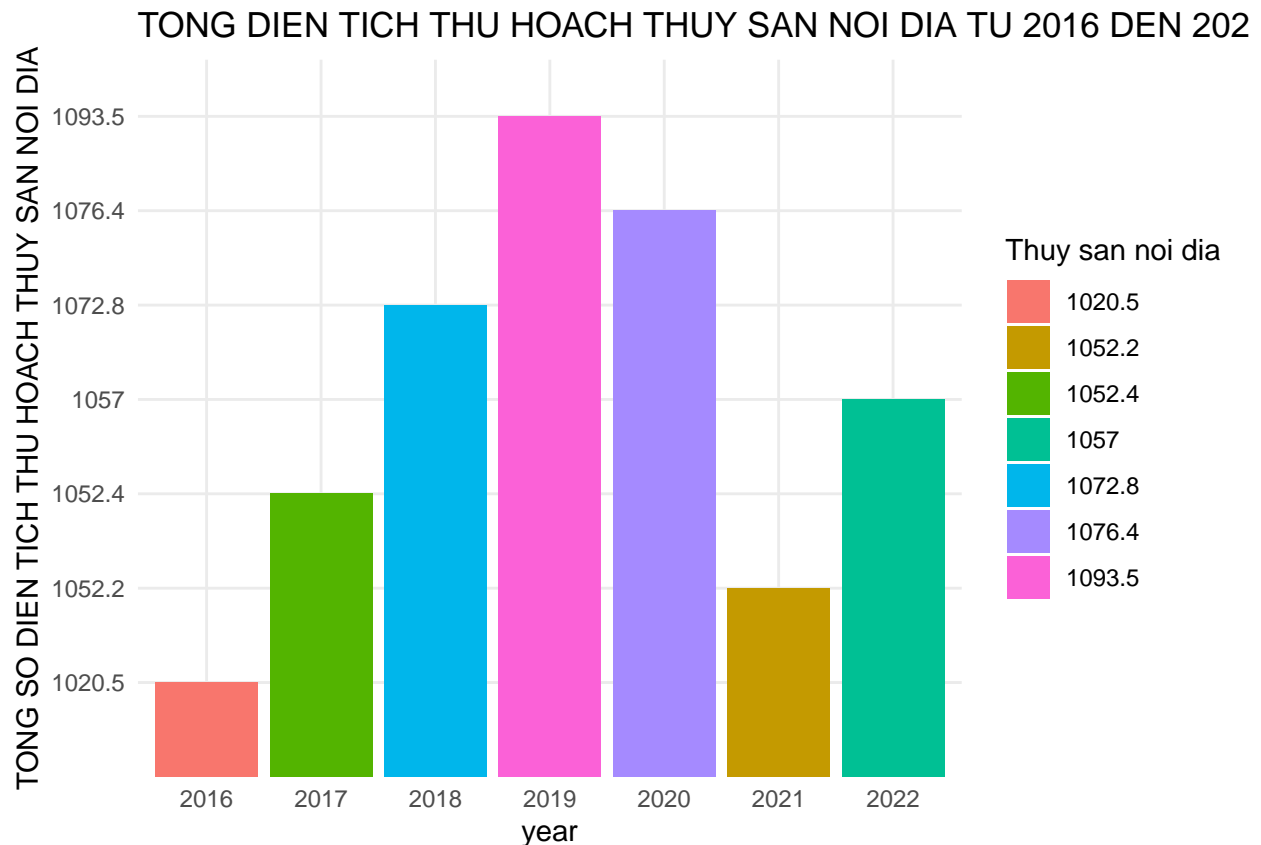


```
#creat a bar plot for Thuy san bien
ggplot(data_02_c,aes(x=Year_Column,y = `Thuy san bien`,fill = `Thuy san bien`)) +
  geom_bar(stat = "identity") +
  labs(title = "TONG DIEN TICH THU HOACH THUY SAN BIEN TU 2016 DEN 2022",
    x ="year",
    y = "TONG SO DIEN TICH THU HOACH THUY SAN BIEN") +
  theme_minimal()
```



```
#creat a bar plot for Thuy san noi dia
ggplot(data_02_c,aes(x=Year_Column,y = `Thuy san noi dia`,fill = `Thuy san noi dia`)) +
  geom_bar(stat = "identity") +
  labs(title = "TONG DIEN TICH THU HOACH THUY SAN NOI DIA TU 2016 DEN 2022",
    x ="year",
    y = "TONG SO DIEN TICH THU HOACH THUY SAN NOI DIA") +
  theme_minimal()
```

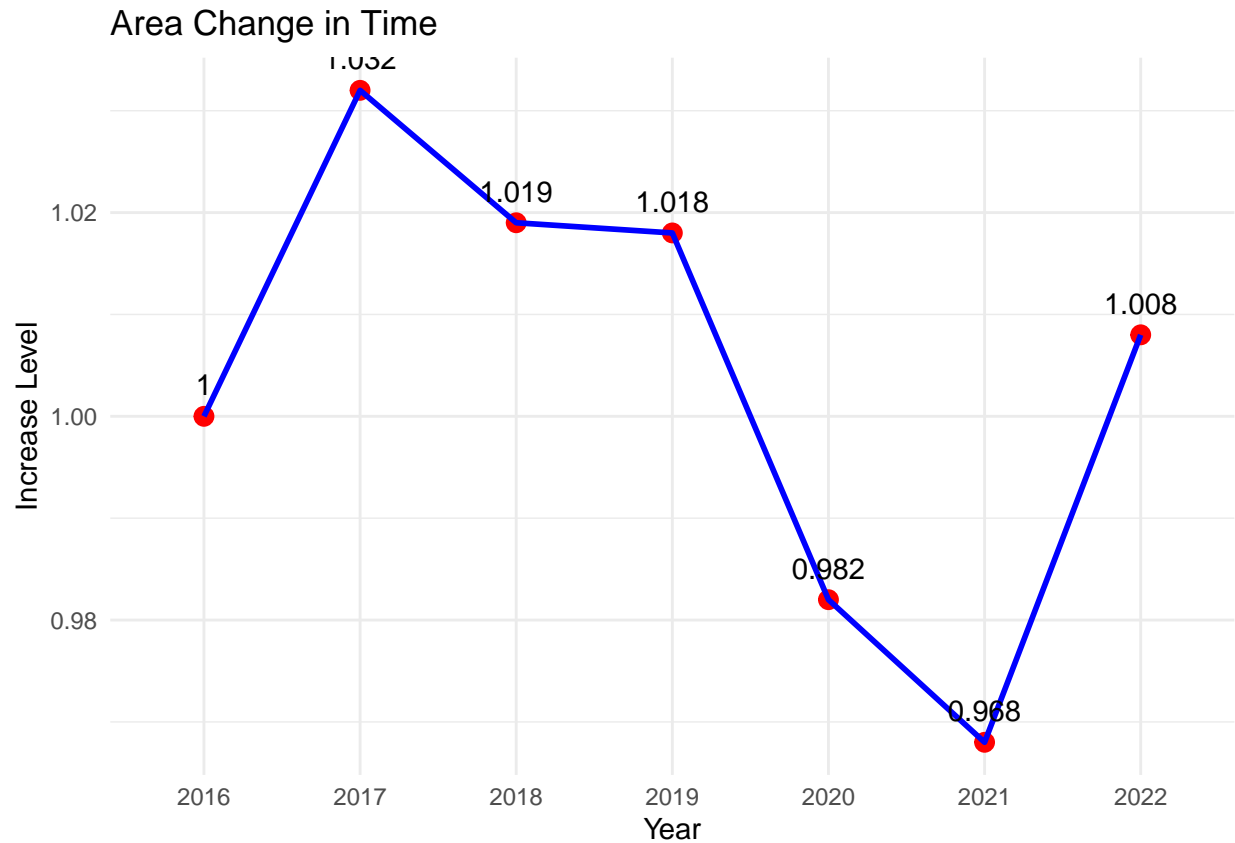




```
#Creat increase level for Tong so
data_02_c$`TONG SO` <- as.numeric(data_02_c$`TONG SO`)
data_02_c <- data_02_c %>%
  mutate(increase_lvl = ifelse(row_number() == 1, 1, `TONG SO` / lag(`TONG SO`)))

# Convert increase level to formatted numeric values
data_02_c$increase_lvl <- round(as.numeric(data_02_c$increase_lvl), 3)

#Creat line plot for increae level in Tong so
ggplot(data_02_c, aes(x = Year_Columnn, y = increase_lvl, group = 1)) +
  geom_point(size = 3, color = "red") +
  geom_line(linewidth = 1, color = "blue") +
  geom_text(aes(label = increase_lvl), vjust = -1, color = "black") + # Add labels
  labs(
    title = "Area Change in Time",
    x = "Year",
    y = "Increase Level"
  ) +
  theme_minimal()
```



```
#creat a data frame for piechart. bring wide to long
data_02_c1 <- data_02_c %>% select(1,3,7,11)
data_02_c1 <- data_02_c1 %>%
  pivot_longer(
    cols = c(`Thuy san bien`, `Thuy san noi dia`, `Dien tich uom, nuoi giong thuy san`),
    names_to = "categories",
    values_to = "Value"
  )
data_02_c1 <- data_02_c1 %>%
  mutate(Value = ifelse(Value == "..", 0, Value)) %>%
  mutate(as.numeric(Value))

#Cal percentage
data_02_c1 <- data_02_c1 %>%
  group_by(Year_Column) %>%
  mutate(
    Value = as.numeric(Value),
    percentage = Value / sum(Value) * 100
  )

#creat a pie plot to show the percentage of category
ggplot(data_02_c1, aes(x = "", y = percentage, fill = categories)) +
  geom_bar(stat = "identity", width = 1) + # Create pie chart
  coord_polar(theta = "y") + # Convert to polar coordinates
  geom_label_repel(
    aes(label = paste0(round(percentge, 1), "%")), # Add percentage labels
    position = position_stack(vjust = 0.5), # Stack labels vertically
  )
```

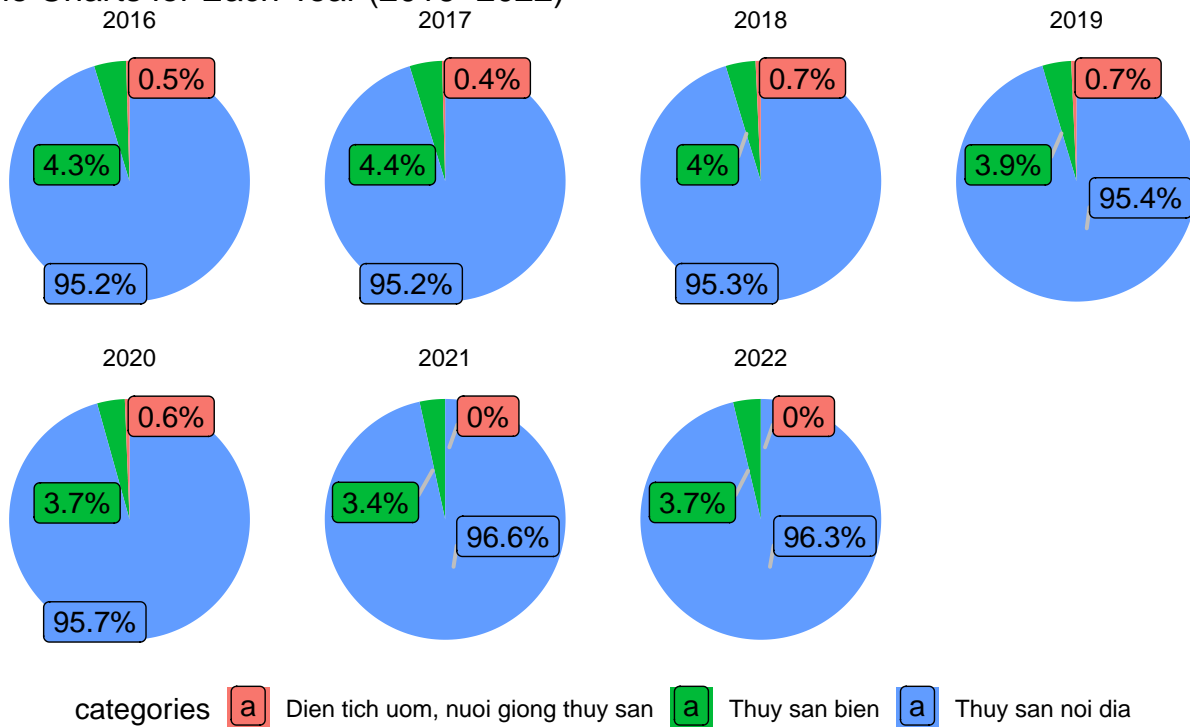
```

box.padding = 0.5,
point.padding = 0.5,
segment.color = "gray",
segment.size = 0.7
) +
facet_wrap(~ Year_Column, ncol = 4) +
labs(title = "Pie Charts for Each Year (2016-2022)", x = NULL, y = NULL) +
theme_void() +
theme(legend.position = "bottom")

```

*# Padding around the text box*  
*# Padding around the arrow point*  
*# Arrow color*  
*# Arrow thickness*  
*# Create facets for each year*  
*# Minimal theme*  
*# Move legend to the bottom*

## Pie Charts for Each Year (2016–2022)



#Summary chunk 2 Có thể thấy rằng diện tích không tăng đều mà có sự thay đổi theo thời gian, hiện tại có giai đoạn từ 2019 tới 2021 diện tích giảm. Note: Diện tích giảm, mà năm nào giá trị cũng tăng, nhưng mà mức tăng không có đột biến. Vậy nên có thể dự đoán rằng sản lượng cũng có giảm, nhưng mà khi tính ra giá trị/ diện tích thì nó vẫn lớn hơn so với các năm trước. Vậy rang cần kiểm tra thêm cả sản lượng thu hoạch từ năm 2016 và năm 2022.

*# Insert data from this link: <https://www.gso.gov.vn/px-web-2/?pxid=V0657&theme=N%C3%B4ng%20%20%C3%A2m>*

*# Read the excel file*

```
data_03 <- read_excel("C:/Users/ngoti/Desktop/Economic in Aquaculture/file_03.xlsx")
```

## New names:

```
## * `` -> `...2`
```

```
## * `` -> `...3`
```

```
## * `` -> `...4`
```

```
## * `` -> `...5`
## * `` -> `...6`
## * `` -> `...7`
## * `` -> `...8`
```

```
print(data_03)
```

```
## # A tibble: 73 x 8
##   Diện tích nuôi trồng/thu hoạch th~1 ...2 ...3 ...4 ...5 ...6 ...7 ...8
##   <chr>                                <chr> <chr> <chr> <chr> <chr> <chr> <chr>
## 1 <NA>                                <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 2 <NA>                                2016 2017 2018 2019 2020 2021 2022
## 3 CẢ NƯỚC                            1066~ 1100~ 1118 1138 1118~ 1089~ 1097~
## 4 Đồng bằng sông Hồng                128.2 134.1 135.3 137  139.3 136.8 138.7
## 5 Hà Nội                             20.9 23.4 24    23.4 23.6 18.8 19.1
## 6 Hà Tây                             ..    ..    ..    ..    ..    ..    ..
## 7 Vĩnh Phúc                           6.9  6.9  7     6.8  6.7  6.5  6.4
## 8 Bắc Ninh                           5.3  5.3  5.2  5.2  4.9  5.2  4.8
## 9 Quảng Ninh                         19   20.1 20.8 21   23.5 27.7 29
## 10 Hải Dương                         11   11.2 11.2 11.9 12    12.3 12.4
## # i 63 more rows
## # i abbreviated name:
## # 1: `Diện tích nuôi trồng/thu hoạch thủy sản phân theo địa phương (*) chia theo Tỉnh, thành phố và
```

```
#Clean data
data_03_c <- data_03[3:73,1:8]
colnames(data_03_c) <- c("area",2016:2022)

# Convert text to ASCII format for consistency
data_03_c$area <- stri_trans_general(data_03_c$area, "Latin-ASCII")

library(tibble)

# Define year range and ranges for each region
years <- 2016:2022
ranges <- list(
  DBSH = 3:14,
  TD_MNPB = 16:29,
  BTB_DHMT = 31:44,
  TN = 46:50,
  DNB = 52:57,
  DBSCL = 59:71
)

# Loop through each year and create a tibble dynamically
for (year in years) {
  # Get the column for the specific year
  year_data <- data_03_c[[as.character(year)]]

  # Find the maximum column length for padding
  max_length <- max(sapply(ranges, function(r) length(r)))

  # Create the tibble for this year
```

```

tibble_data <- as_tibble(lapply(ranges, function(r) {
  c(year_data[r], rep(NA, max_length - length(r)))
}))

# Dynamically assign the tibble to a variable named dt_03_cXX (e.g., dt_03_c16)
assign(paste0("dt_03_c", substr(year, 3, 4)), tibble_data)
}

# Replace "." with 0 across all columns
dt_03_c16 <- dt_03_c16 %>%
  mutate(across(everything(), ~ ifelse(. == ".", 0, .))) %>%
  mutate(across(everything(), ~ ifelse(is.na(.), 0, .)))

# dt_03_16 to long table
dt_03_c16_1 <- dt_03_c16 %>% pivot_longer(
  cols = c(DBSH,TD_MNPB,BTB_DHMT,TN,DNB,DBSCL),
  values_to = "Value",
)

anova_test<-aov( Value ~ name,data = dt_03_c16_1)

```