Shark Control Program Analysis

Pratul Singh Raghava

University of Sydney

Contents

Executive Summary

- The aim of this report is to analyze the trends observed in Shark Control Program Data regarding Shark catch numbers on the basis of Length, Location, Month and Season.

- The main discoveries are :

  1) The Dusky Whaler species of Sharks shows the highest median length (2.7 m) and also the maximum range (1.4 m) of lengths among the sharks caught.
  2) The Tiger Shark is the most caught Shark species in the Program with 207 catches in total. A majority (32) of these catches took place at Harbour Beach.
  3) A majority of Tiger Shark catches happened in warmer seasons and Winter season saw the lowest catch numbers compared to Autumn which saw the highest numbers.

Full Report

Initial Data Analysis (IDA)

- The data came from the Agriculture & Fisheries Department of the State of Queensland. The dataset contains data from each day of 2016 and was collected as part of the Queensland Shark Control Program, a Government initiative to prevent shark bites.

- The data is valid because it is official Government data available openly to the public.

- Possible issues include :

  1) False reporting of certain features of the catch, for example, its species or length, and also the coordinates of the place where it was caught.

  2) Inaccurate measurement of the water temperature since sharks are caught at varying depths and we only measure the water surface temperature, this field may not be a much accurate indicator.

- Potential stakeholders include marine life environmentalists, other people who live near the seas and even tourists who visit these waters. Also, the government and healthcare providers whose interests lie in the safety of the people.

- Each row represents one shark caught by the Program and contains all possible information about the shark.

- Each column represents fields of data about the shark caught by the Program. These are Species Name, Date, Area, Location, Latitude, Longitude, Length(m), Water Temperature(C), Month, Day of Week.

- The key variables are species, length, month, location.

  1) Species, a character variable, holds the value of one of the 24 different types of sharks. It helps us to classify the shark so that they can be distinctly categorised for further analysis.

  2) Length, a floating point variable, holds the value of the length of the shark and is a primary characteristic of the shark. It enables us to categorise sharks into different categories of sizes for easy classification.

  3) Month, a character variable, holds the value of the month that the shark was caught in and helps us to do analysis of shark catches categorised by months and seasons,

  4) Location, a character variable, holds the value of the name of the site where the shark was caught and helps us to categorize the sharks by the place where they were found.

```
## read in data from file
file=read.csv("sharks.csv")

##Store attributes in variables
species=file$Species.Name
lengthm=file$Length..m.
month=file$Month
location=file$Location

## show classification of variables
str(file)
```

```
## 'data.frame':    532 obs. of  10 variables:
##  $ Species.Name  : chr  "AUSTRALIAN BLACKTIP" "BLACKTIP REEF WHALER" "BLACKTIP REEF WHALER" "BLACKTIP
##  $ Date          : chr  "2016-11-16" "2016-01-02" "2016-01-02" "2016-01-05" ...
##  $ Area          : chr  "Cairns" "Cairns" "Cairns" "Mackay" ...
##  $ Location      : chr  "Holloways Beach" "Buchans Point Beach" "Ellis Beach" "Harbour Beach" ...
##  $ Latitude      : chr  "-16Â°49.82" "-16Â°43.56" "-16Â°43.3" "-21Â°7.08" ...
##  $ Longitude     : chr  "145Â°44.85" "145Â°39.78" "145Â°39.01" "149Â°13.62" ...
##  $ Length..m.    : num  1 0.7 1.5 2.2 1.7 1.2 0.75 1.2 0.8 1.3 ...
##  $ Water.Temp..C.: int  27 27 27 26 26 29 30 31 29 29 ...
##  $ Month         : chr  "November" "January" "January" "January" ...
##  $ Day.of.Week   : chr  "Wednesday" "Saturday" "Saturday" "Tuesday" ...
```

Research Question 1 : How do the different species of sharks vary in length among themselves and compared to others?

We aim to analyse the trends of heights of the sharks caught in the Shark Control Program in 2016. We achieve this by producing multiple charts that look at both the median heights of species compared to each other and the full range of heights of those species with the highest median values.

First, we produce a bar plot to analyse the trends observed in the median heights of species of sharks caught. We represent Length(m) on the X-axis and the different Species on the Y-Axis such that we get horizontal bars for each species with the length of the bars signifying the median height in meters of that species of shark.

```
#Extracting all unique names of species
uniq_spec=unique(species)

#Initialising a blank vector to hold the median length values for each of the 24 species
med_length=c(rep(0,length(uniq_spec)))

#Appending the median lengths into the blank vector
for (i in 1:length(uniq_spec)){
  med_length[i] <- median( lengthm[ species==uniq_spec[i] ])
  }

#Producing the Barplot
barplot(med_length,main = "Median height for each species caught in 2016",horiz=TRUE, col = c("light blu
```

# Median height for each species caught in 2016
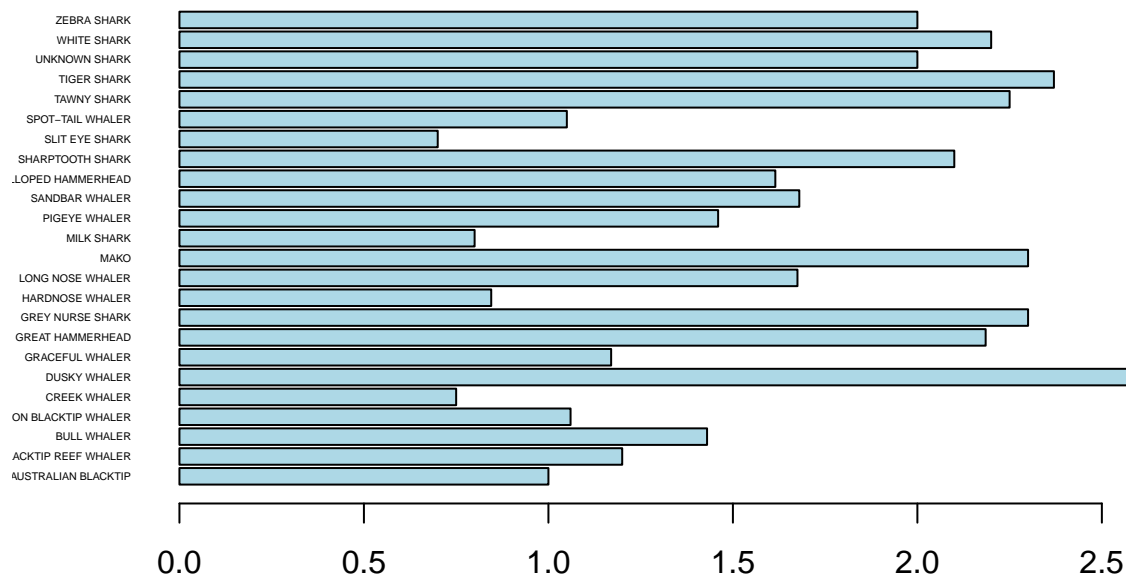
Summary : On producing this bar plot, we observe that the Dusky Whaler, Tiger Shark, Mako and Grey Nurse Shark species show the maximum median lengths among all species of shark observed. Moreover, we see that the range of median lengths varies quite a bit from 0.7 metres for the Slit Eye Shark to 2.7 for the Dusky Whaler.

Second, we aim to observe the ranges in which the lengths of the species vary. We do so by producing multiple boxplots for the species with highest median lengths and comparing them.

```
#Producing 4 Boxplots next to each other
par(mfrow=c(1,4)) #1 row 4 columns

boxplot(lengthm[species=="DUSKY WHALER"],col="light blue",las=1,main = "DUSKY WHALER")
boxplot(lengthm[species=="TIGER SHARK"],col="light green",las=1,main = "TIGER SHARK")
boxplot(lengthm[species=="MAKO"],col="yellow",las=1,main = "MAKO")
boxplot(lengthm[species=="GREY NURSE SHARK"],col="light green",las=1,main = "GREY NURSE SHARK")
```
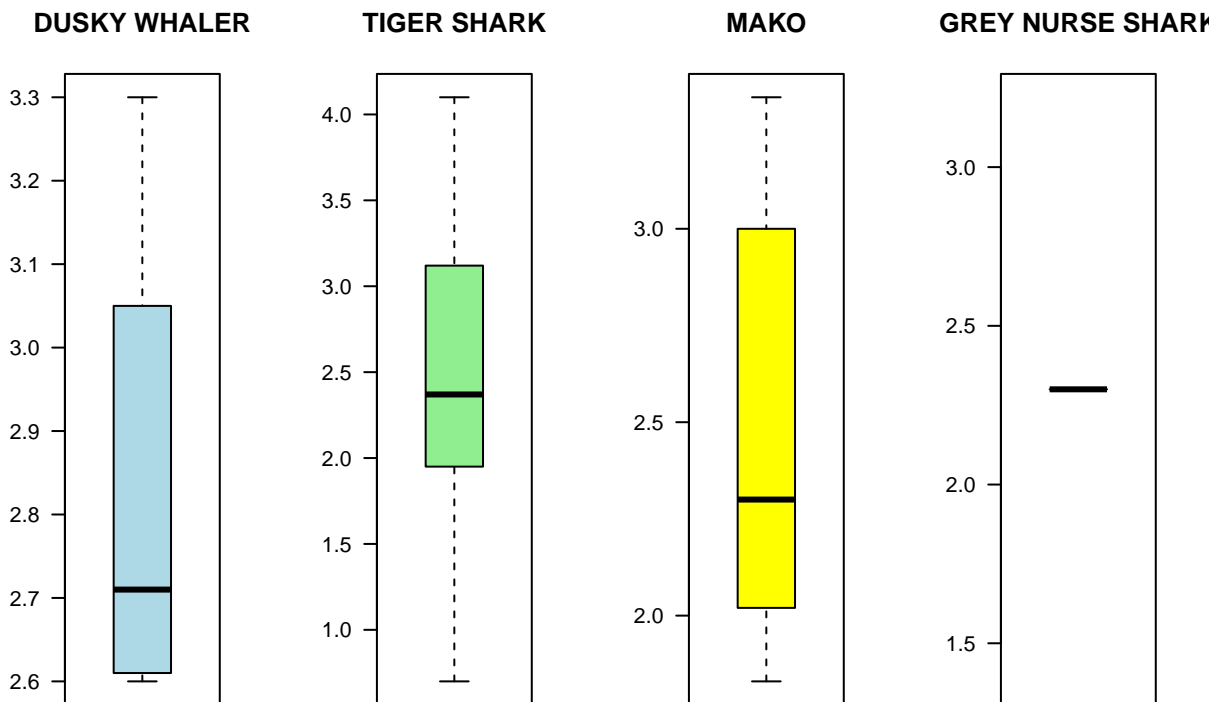

Summary: While the Dusky Whaler shows the maximum range in heights, the Grey Nurse Shark doesn't produce a range at all. This is because only one Grey Nurse Shark was caught in the whole Program with a height greater than the median heights of most other species. Due to this, it appears in our analysis as a sole data point and sort of weakens any conclusions we make from these charts since we are unable to make a true comparison.

Research Question 2 : How do catches of different Shark species vary with location and season?

We aim to observe the trends seen in Sharks with respect to the sites where they are caught and seasons they are more commonly caught in. We do so by producing multiple charts that first see the geographical distribution of all shark catches and then the seasonal distribution for the same. A hypothesis test is also performed regarding the same.

First, we find the most caught shark species to perform our analysis on. By simple analysis, we find that :

```
#Finding the most caught shark species
g=table(species)
g4=sort(g,decreasing=TRUE)
g4[1] #The top valued species
```

```
## TIGER SHARK
##          207
```

Tiger Shark is the most caught species in the Shark Control Program in 2016.

We now further analyse the characteristics of Tiger Shark catches by producing a barplot to see the most common sites where it was caught. We represent the different sites on the Y-Axis and number of catches on the X-Axis such that we get horizontal bars for each site signifying the corresponding number of Tiger Shark catches in that site.

```
#Fetching all sites where Tiger Shark was caught
w=location[species=="TIGER SHARK"]

#Sorting the sites in decreasing order
w4=table(w)
w42=sort(w4,decreasing=TRUE)

#Producing the Barplot
barplot(w42,main = "Most common sites for Tiger Shark catchings",horiz=TRUE, col = c("lightgreen") ,las=
```

Most common sites for Tiger Shark catchings
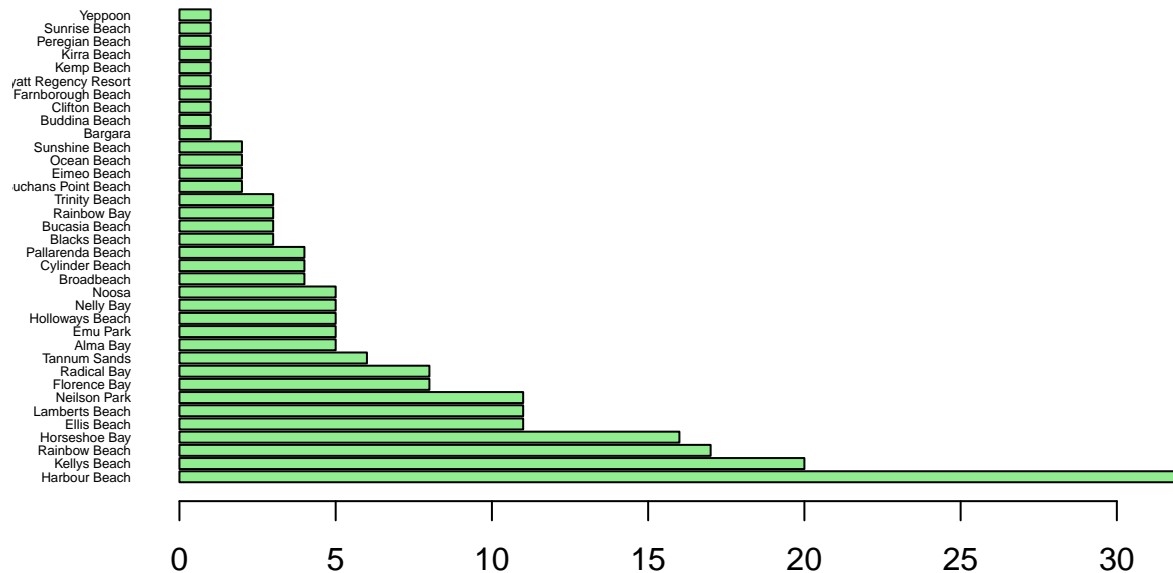
Summary : We find that Harbour Beach sees more Tiger Shark catches than any other site by a huge margin over the nearest names like Kellys Beach, Rainbow Beach and Horseshoe Bay.

We now aim to analyse the seasonal distribution of Tiger Shark catches. We do so by categorising the catch numbers by months and grouping them by seasons as a broader category.

We produce a barplot to see whether climatic conditions over a period of time affect the catch numbers or not. We represent the months on the X-Axis and the catch numbers on the Y-Axis such that we get vertical bars for each month with the height of the month signifying the Tiger Shark catches for that month.

```r
month_names=c("January", "February", "March", "April", "May", "June", "July","August","September","Octob

#Ordering the months as per convention
ordermonth <- ordered(month, levels=month_names)

#Fetching Tiger shark catches through the year
z=ordermonth[species=="TIGER SHARK"]

#Assigning colours to months as per season
month_colors=c("red","red","light green","light green","light green","light blue","light blue","light bl

#Producing the Barplot
barplot(table(z),names=month_names,las=2,col=month_colors,main="Tiger Shark catch numbers by Months")

#Creating the legend for the Barplot
legend("topright",y=NULL,legend=c("Summer","Autumn","Winter","Spring"),fill=c("red","light green","light
```
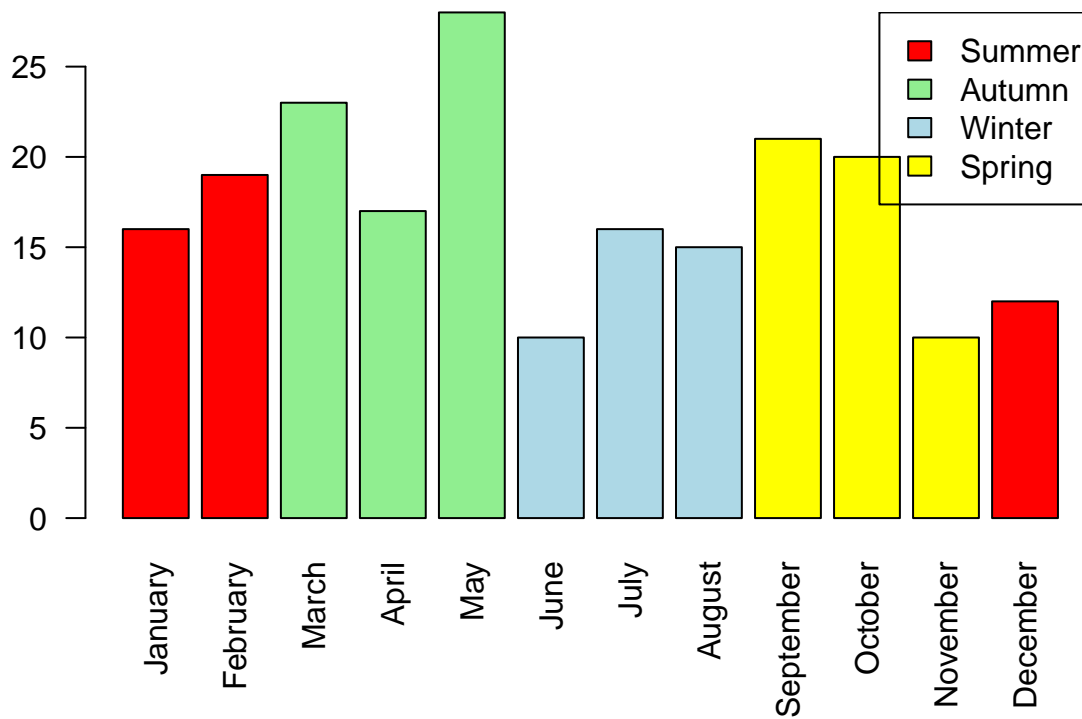
Tiger Shark catch numbers by Months

Summary : We observe that the catch numbers vary from month to month but the observable trend is that Tiger Shark catches reduce in the Winter season compared to other warmer seasons. Autumn comes out to be the period with the highest catches. This finding is in consensus with a report by ABC Science which concludes that "Shark sightings are more common during summer than winter". Although, it is to be noted that this article dates back to 2010 whereas our analysis is on data from 2016. At the time of writing this report in 2021, several factors may have changed and the findings may even be inconsistent with reality.

We also see that the numbers in Summer season are a bit less than those in transitional seasons like Spring and Autumn. This finding seems to be on par with a BBC report which discusses "Why sharks like it hot - but not too hot" and relates its findings to season and water temperature.

Hypothesis Test : Majority of Tiger Sharks are caught in Warmer Seasons

- **Significance Level** : By convention, we use $\alpha = 0.05$.

- **Hypothesis** : if p is the proportion of Tiger Sharks caught in warmer seasons than Winter, we will test $H_0 : p = 0.75$ vs $H_1 : p \neq 0.75$.

- **Assumptions** : For the Z test, we need to make two assumptions:

  1. The observations are independent of each other, which is reasonable as the Tiger Sharks once caught are not released back into the seas.
  2. The sample size is large enough for the Central Limit Theorem. The sample size is 182 and if the data is not greatly skewed, this assumption is satisfied.

- **Test statistic** : Assuming $p = 0.75$, we have that for the box:

Mean, Standard Deviation

```
mean = 0.75
sd = (1-0) * sqrt(0.75 * 0.25)
c(mean,sd)
```

```
## [1] 0.7500000 0.4330127
```

And for the **sum** of a sample:

Expected Value, Standard Error

```
n = 207
EV = mean * n
SE = sd * sqrt(n)
c(EV, SE)
```

```
## [1] 155.250000   6.229968
```

Thus, the test statistic is

```
OV = 166
test.stat = (OV - EV)/SE
test.stat
```

```
## [1] 1.725531
```

That is, our observed value (166) is 0.07 standard errors away from the expected value (assuming $p = 0.8$).

- **p-value** : We calculate the p-value :

```
pnorm(test.stat,lower.tail = TRUE)
```

```
## [1] 0.957784
```

- A p-value of 0.95 means that we can expect to observe a test statistic of size 1.72 or more approximately 96% of times.
- **Conclusion** : Since our p-value is so high and nearly equal to 1,

  1. We approve the null hypothesis
  2. We conclude that the data doesn't provide enough evidence that the proportion of Tiger Sharks caught in warmer seasons (other than Winter) is not equal to 80%.

References

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Climate Glossary - Seasons - Bureau of Meteorology. http://www.bom.gov.au/climate/glossary/seasons.shtml

(2021, September 26). Shark bites 'statistically more likely' as swimmer protection lags, expert says https://www.abc.net.au/news/2021-09-26/nsw-shark-control-program-triples-spend-doubles-qld-efforts/100456268

(2010, November 04). Are we seeing more sharks than usual at this time of year? https://www.abc.net.au/science/articles/2010/11/04/3056893.htm

(2018, March 13). Why sharks like it hot - but not too hot. https://www.bbc.com/news/science-environment-43377026