Assessment 2

Danko Castaño Duro

2024-05-19

#Assignment. End-to-end data analysis in R. In this assignment, we will use real-world marine data collected as part of Queensland fisheries QFISH database to develop a report (either written on your website or via an R Markdown report embedded in your website) suitable for reporting to your manager, a community group or for your own research.

# First, obtain your data.

#ALL DONE #Open dataset

setwd("C:/Users/Danko Castaño Duro/Desktop/Australia 3 semestre/Modulos 1/INTRODUCTION TO R 4/MB5370\_Mod04\_Danko/MB5370\_Mod04\_Danko")  
  
  
shark <- read.csv("shark.csv")  
  
View(shark)

# Clean data

#Delete all rows that have the word “total…”

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.2.3

## Warning: package 'ggplot2' was built under R version 4.2.3

## Warning: package 'tidyr' was built under R version 4.2.3

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ forcats 1.0.0 ✔ readr 2.1.4  
## ✔ ggplot2 3.5.1 ✔ stringr 1.5.0  
## ✔ lubridate 1.9.2 ✔ tibble 3.1.8  
## ✔ purrr 1.0.1 ✔ tidyr 1.3.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force all conflicts to become errors

rows\_with\_total <- apply(shark, 1, function(row) any(grepl("total", row, ignore.case = TRUE)))  
  
shark\_cleaned <- shark[!rows\_with\_total, ]  
  
view(shark\_cleaned)

#Explanation #apply(shark, 1, function(row) …): This applies the function to each row (MARGIN = 1) of the data frame shark.

#grepl(“total”, row, ignore.case = TRUE): This checks if the word “total” (case insensitive) is present in any element of the row.

#any(grepl(…)): This returns TRUE if any element in the row contains the word “total”.

#rows\_with\_total: This logical vector indicates which rows contain the word “total”.

#shark[!rows\_with\_total, ]: This subsets the data frame to include only rows that do not contain the word “total”.

# Check structure of data set

str(shark\_cleaned)

## 'data.frame': 785 obs. of 4 variables:  
## $ CalendarYear : chr "2001" "2001" "2001" "2001" ...  
## $ Area : chr "Bundaberg" "Bundaberg" "Bundaberg" "Cairns" ...  
## $ SpeciesGroup : chr "Other" "Shark" "Turtle" "Mammal" ...  
## $ NumberCaught.Total: int 1 62 12 1 16 79 4 3 87 32 ...

# Plot data

#Filter the data by Making new tiny datset

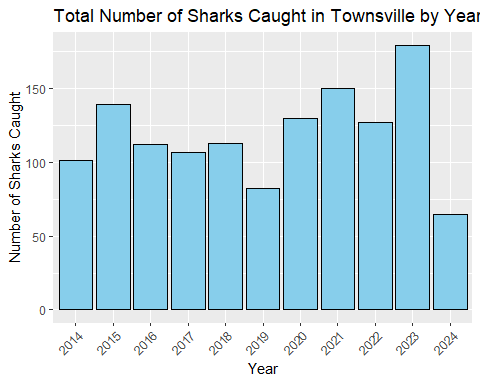
just\_shark\_Townsville <- shark\_cleaned %>%  
 filter(Area == "Townsville", SpeciesGroup == "Shark", CalendarYear >= "2014", CalendarYear <= "2024")  
view(just\_shark\_Townsville)

#Plot the “easy” dataset

str(just\_shark\_Townsville)

## 'data.frame': 11 obs. of 4 variables:  
## $ CalendarYear : chr "2014" "2015" "2016" "2017" ...  
## $ Area : chr "Townsville" "Townsville" "Townsville" "Townsville" ...  
## $ SpeciesGroup : chr "Shark" "Shark" "Shark" "Shark" ...  
## $ NumberCaught.Total: int 101 139 112 107 113 82 130 150 127 179 ...

ggplot(data = just\_shark\_Townsville, aes(x = CalendarYear, y = NumberCaught.Total)) +  
 geom\_bar(stat = "identity", fill = "skyblue", color = "black") +  
 labs(x = "Year", y = "Number of Sharks Caught", title = "Total Number of Sharks Caught in Townsville by Year") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1)) # Rotate x-axis labels for better readability



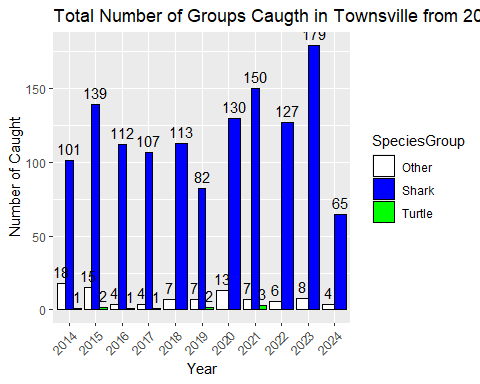
#Plot for all species in townsvile

#Filter data

just\_Townsville <- shark\_cleaned %>%  
 filter(Area =="Townsville", CalendarYear >="2014", CalendarYear <= "2024")  
  
view(just\_Townsville)

# Plot FINAL data

ggplot(data = just\_Townsville, aes(x = CalendarYear, y = NumberCaught.Total, fill = SpeciesGroup)) +  
 geom\_bar(stat = "identity", position = "dodge", color = "black") +  
 geom\_text(aes(label = NumberCaught.Total), position = position\_dodge(width = 0.9), vjust = -0.5) +  
 labs(x = "Year", y = "Number of Caught", title = "Total Number of Groups Caugth in Townsville from 2014-2024") +  
 scale\_fill\_manual(values = c("Shark" = "blue", "Turtle" = "green", "Other" = "white")) +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1)) # Rotate x-axis labels for better readability



#Final just\_Townsville plot displays the Number of Caught Groups on Townsville from 2014 to 2024