Sumatran Tiger Project

Jade Holley

2023-05-01

In order to find a data set for the *"Data Analysis & Visualisation"* module that I would be able to understand, I turned towards my interests, thinking that it would be easier to interpret a dataset that I myself found interesting. After some research into conservation, I came across the extinction of Sumatran tigers as a result of deforestation and growing agricultures, such as palm oil and coffee plantations (World Wide Fund for Nature, 2022).

```
tiger_imgpath <- here("..", "Tiger_Images",
                      "Leuser_12A_SW_SD_48 Tiger 2014 02 23 14 08 29.jpg")

knitr::include_graphics(tiger_imgpath)
```

Figure 1: Tiger Captured in Gunung Leuser (Luskin et al., 2017)

Whilst the population of Sumatran Island tigers is declining, there has been an increase in their numbers in national parks across Sumatra as a result of their degraded habitats elsewhere (Luskin et al. 2017). Therefore, this project aims to present the tiger populations in three national parks in Sumatra that are all declared as UNESCO world heritage sites: Gunung Leuser, Bukit Barisan Selatan and Kerinci Seblat.

Although this project allows a colourful presentation of the data, it is important to note the main purpose of this project is to highlight how few Sumatran tigers remain, and their constriction to small national parks due to degraded forests outside of protected areas. Importantly, even in these national parks it has been found that there are only two robust tiger populations (Luskin et al., 2017), and so, more needs to be done to protect and conserve the environment to prevent the total extinction of these beautiful animals which are now critically endangered.

The Data Itself

This project has used difference sources of data. First, the shapefile utilised to create the maps was sourced from GADM.org. This shapefile contains multiple layers of spatial data in the form of polygons and multipolygons. The whole shapefile was used to form the base of the map, but layer two was then extracted to create the IDN_provinces dataframe which was overlaid onto the map, allowing me to then highlight the national parks. The coordinates of the map were then also extracted in order to both zoom into the map, and to pinpoint the central area in each park, to be able to create an interactive areas that revealed how many tigers were in each park.

```
#Import the shapefile and rename it
IDN_shapefile <- st_read(here("..", "Data", "gadm41_IDN_shp"))
```

```
## Multiple layers are present in data source C:\Users\jadeh\OneDrive\Documents\Coding Project\
## Use 'st_layers' to list all layer names and their type in a data source.
## Set the 'layer' argument in 'st_read' to read a particular layer.
## Reading layer 'gadm41_IDN_0' from data source
##   'C:\Users\jadeh\OneDrive\Documents\Coding Project\SumatranTigerProject\Data\gadm41_IDN_shp
##   using driver 'ESRI Shapefile'
## Simple feature collection with 1 feature and 2 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: 95.00971 ymin: -11.00761 xmax: 141.0194 ymax: 6.076941
## Geodetic CRS:  WGS 84
```

```
#Load the layers of the shapefile
st_layers(here("..", "Data", "gadm41_IDN_shp"))
```

```
## Driver: ESRI Shapefile
## Available layers:
##     layer_name geometry_type features fields crs_name
## 1 gadm41_IDN_0       Polygon        1      2   WGS 84
## 2 gadm41_IDN_1       Polygon       34     11   WGS 84
## 3 gadm41_IDN_2       Polygon      502     13   WGS 84
## 4 gadm41_IDN_3       Polygon     6695     16   WGS 84
## 5 gadm41_IDN_4       Polygon    77473     14   WGS 84
```

Secondly, the data pertaining to the tigers was sourced from (Luskin et al. 2017). I contacted the author directly via email to which he responded with the raw data files and a dropbox of tiger images captured during the authors' research. The primary file used was "TigerCapturesSumatra.csv" which contained 7 variables, some of which were wrangled to extract specific information. This dataset contained a vast amount of data that were not required for the analysis, so I have utilised only two of the variables: “Sex”, and “AnimalID”, though some of these variables were wrangled to extract pertinent information. I have then created some new variables e.g. to sum the total number of female/male/unknown gendered tigers sighted for the pie chart.

```
#Import TigerCapturesSumatra dataframe

tiger_capture_data <- read.csv(here("..", "Data", "TigerCapturesSumatra.csv"))
```

Research Aims

Research aims The primary aim of this project is to demonstrate the number and gender of Sumatran tigers inhabiting the three main national parks of Sumatra Island.

Research Questions

Question 1

Where do the majority of Sumatran tigers reside?

Question 2

How many Sumatran tigers have been spotted in each park?

Question 3

Which national park has the largest Sumatran tiger population?

Question 4

What can the gender proportions tell us about tiger populations?

Data Wrangling the shapefile

To use the data from the shapefile I first had to separate out the layers of the shape file to be able to assess their individual contents.

```
#Open and name each  layer of the shapefile
IDN_data_layer1 <- read_sf(
  dsn = "C:\\Users\\jadeh\\OneDrive\\Documents\\Masters Degree\\Data Visualisation Project\\Da
```

```r
IDN_data_layer2 <- read_sf(
  dsn = "C:\\Users\\jadeh\\OneDrive\\Documents\\Masters Degree\\Data Visualisation Project\\Da

IDN_data_layer3 <- read_sf(
  dsn = "C:\\Users\\jadeh\\OneDrive\\Documents\\Masters Degree\\Data Visualisation Project\\Da

IDN_data_layer4 <- read_sf(
  dsn = "C:\\Users\\jadeh\\OneDrive\\Documents\\Masters Degree\\Data Visualisation Project\\Da

#View the contents of layer 2; contains 14 variables
IDN_data_layer2
```

```
## Simple feature collection with 502 features and 13 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: 95.00971 ymin: -11.00761 xmax: 141.0194 ymax: 6.076941
## Geodetic CRS:  WGS 84
## # A tibble: 502 x 14
##    GID_2      GID_0 COUNTRY   GID_1 NAME_1 NL_NA~1 NAME_2 VARNA~2 NL_NA~3 TYPE_2
##    <chr>      <chr> <chr>     <chr> <chr>  <chr>   <chr>  <chr>   <chr>   <chr>
##  1 IDN.1.2_1  IDN   Indonesia IDN.~ Aceh   NA      Aceh ~ NA      NA      Kabup~
##  2 IDN.1.1_1  IDN   Indonesia IDN.~ Aceh   NA      Aceh ~ NA      NA      Kabup~
##  3 IDN.1.3_1  IDN   Indonesia IDN.~ Aceh   NA      Aceh ~ NA      NA      Kabup~
##  4 IDN.1.4_1  IDN   Indonesia IDN.~ Aceh   NA      Aceh ~ NA      NA      Kabup~
##  5 IDN.1.5_1  IDN   Indonesia IDN.~ Aceh   NA      Aceh ~ NA      NA      Kabup~
##  6 IDN.1.6_1  IDN   Indonesia IDN.~ Aceh   NA      Aceh ~ NA      NA      Kabup~
##  7 IDN.1.7_1  IDN   Indonesia IDN.~ Aceh   NA      Aceh ~ NA      NA      Kabup~
##  8 IDN.1.8_1  IDN   Indonesia IDN.~ Aceh   NA      Aceh ~ NA      NA      Kabup~
##  9 IDN.1.9_1  IDN   Indonesia IDN.~ Aceh   NA      Aceh ~ NA      NA      Kabup~
## 10 IDN.1.10_1 IDN   Indonesia IDN.~ Aceh   NA      Aceh ~ NA      NA      Kabup~
## # ... with 492 more rows, 4 more variables: ENGTYPE_2 <chr>, CC_2 <chr>,
## #   HASC_2 <chr>, geometry <MULTIPOLYGON [°]>, and abbreviated variable names
## #   1: NL_NAME_1, 2: VARNAME_2, 3: NL_NAME_2
```

After inspecting the contents of layer 2 it appeared to contain information pertaining to the provinces of Indonesia, under the variable "Name_2", so I decided to use this layer to create a new dataframe which I could use to overlay these provinces on the map.

```r
IDN_provinces <- st_read(
  "C:\\Users\\jadeh\\OneDrive\\Documents\\Masters Degree\\Data Visualisation Project\\Data\\ga
```

```
## Reading layer 'gadm41_IDN_2' from data source
##   'C:\Users\jadeh\OneDrive\Documents\Masters Degree\Data Visualisation Project\Data\gadm41_I
##   using driver 'ESRI Shapefile'
## Simple feature collection with 502 features and 13 fields
## Geometry type: MULTIPOLYGON
```

```
## Dimension:     XY
## Bounding box:  xmin: 95.00971 ymin: -11.00761 xmax: 141.0194 ymax: 6.076941
## Geodetic CRS:  WGS 84
```

I then needed data about the national parks that I could overlay onto the maps, but I could not access this from the shapefile. Instead, I used Google Maps to find the provinces in which the national parks were located, in order to pinpoint these areas on the map as a compromise to not having the exact park boundaries.

```
#Create a new dataframe called National_parks
#use MUTATE to combine province names into areas that make up the national parks
#either Gunung Leuser, Bukit Barisan Selatan, or Kerinci Seblat

National_parks <- IDN_provinces %>%
  mutate(National_Parks = case_when(
    NAME_2 %in% c("Nagan Raya", "Aceh Barat Daya", "Aceh Selatan",

                  "Aceh Tenggara", "Subulussalam", "Aceh Singkil",

                  "Gayo Lues", "Aceh Tamiang", "Aceh Timur",

                  "Bener Meriah") ~ "Gunung Leuser",

    NAME_2 %in% c("Lampung Barat", " Tanggamus", "Bengkulu Selatan",

                  "Kaur") ~ "Bukit Barisan Selatan",

    NAME_2 %in% c("Kerinci", "Solok","Sawahlunto", "Lebong",

                  "Rejang Lebong", "Solok Selatan", "Bungo",

                  "Merangin") ~ "Kerinci Seblat"))
```

Then I wrangled this new data frame to exclude unnecessary rows from the data.

```
#Create a new dataframe which filters out NA values from  National_Parks column
filtered_parks <- dplyr::filter(National_parks, National_Parks %in% c(
  "Gunung Leuser", "Bukit Barisan Selatan", "Kerinci Seblat"))
```

Data wrangling the raw data from Matt Luskin

To use the raw data from Luskin et al.(2017), I first needed to remove irrelevant columns from the dataframe.

```r
#Remove column "SurveyID", "Side" and "StationID"
tiger_capture_tidy <- subset(tiger_capture_data,
                             select = -c(SurveyID, StationID, Side))
```

I then needed to detangle the date and time from the Date.Time column.

```r
#create a new column extracting only the date  from the Date.Time column
#Keep only the month and year values
#use MUTATE to rename the column name
tiger_capture_tidy_date <- tiger_capture_tidy %>%
  mutate(date= str_sub(tiger_capture_tidy$Date.Time, start=1, end=7))

#Now remove the original Date.Time column leaving just the new "date" column
tiger_capture_tidy_date <- subset(tiger_capture_tidy_date,
                                  select = -c(Date.Time))
```

Finally, I needed a column containing information about which national park each tiger had been sighted in, but this data was linked to the AnimalID column, despite there already being an IndividualCode column with each animal ID in it.

```r
#Extract location from AnimalID column
#Keep location information only
#BBS (Bukit Barisan Selatan), LEU (Gunung Leuser), or KER (Kerinci Seblat)
#Rename the column to national-park
tiger_map_data <- tiger_capture_tidy_date %>%
  mutate(national_park= str_sub(
    tiger_capture_tidy_date$AnimalID, start=1, end=3))

#Remove the AnimalID column from the dataset
tiger_map_data <- subset(tiger_map_data, select = -c(AnimalID))
```

Colour Schemes

An important part of data visualisation is coherence, so I used RColBrewer to create both green and orange palettes which I could use throughout the project to create a sense of completeness, and a consistent theme across the visualisations.

```r
#Create an RColBrewer palette to use for the national parks on the map
#Needs to contain different green hues to differentiate each national park
#Green was selected to reflect the natural colours of national parks

map_palette <- brewer.pal(n=3, "Greens")
#Select a green palette from RColBrewer
#n is the number of colors needed in the palette

map_palette #Shows the names of the three green hues to manually add to the map
```

```
## [1] "#E5F5E0" "#A1D99B" "#31A354"
```

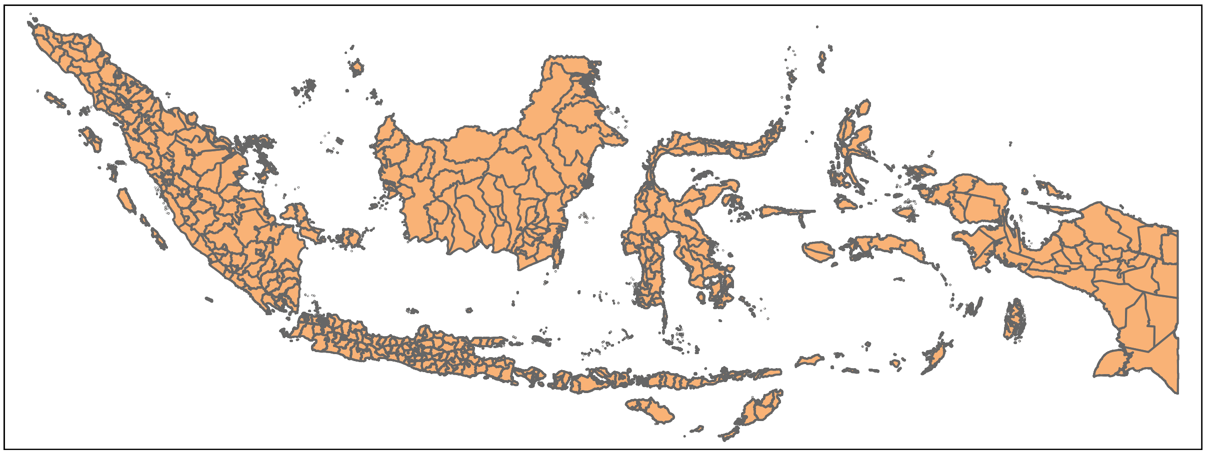
```
#Create an orange palette to use on all graphs and as the basecolor of the map
#Orange was selected to incorporate an overall tiger theme
#Create a RColBrewer palette to use for the graphs with different hues
graph_palette <- brewer.pal(n=3, "Oranges")
#Select orange palette from RColBrewer
#n is the number of colors needed in the map
graph_palette
```

```
## [1] "#FEE6CE" "#FDAE6B" "#E6550D"
```

Map Of Indonesia

```
#Plot a basic map of Indonesia with province borders outlined

map <- tm_shape(IDN_shapefile) + #maps the outline of the shapefile
  tm_polygons() +
  tm_shape(IDN_provinces) + # overlays the provinces dataframe
  tm_polygons("#FDAE6B", alpha = 0.9) +
  #color the map in line with a tiger theme, alpha adjusts transparency
  tm_borders("#A1D99B") + #Add a border to each province
  tm_layout(main.title = "Map of Indonesia", #Add the main title above the map
            fontface = "italic", #Italicize the text
            fontfamily = "serif") #Change font for text
print(map)
```

Map of Indonesia



```r
#save map
tmap_save(
  map, filename =
    "C:\\Users\\jadeh\\OneDrive\\Documents\\Coding Project\\Tmap Figures\\Indonesia.png")
```

This map is included to demonstrate the size of Indonesia, to contextualise how confined the tigers are in the following Visualisations.

Data Visualisation Number 1: Map of Indonesia Highlighting the National Parks in Sumatra

```r
#Plot a basic map of Indonesia with pro}
# Highlight the three national parks on the base map
#To do this overlay the filtered_data df
map_2 <- tm_shape(IDN_shapefile) + #Maps the outline of the shapefule
  tm_polygons() +
  tm_shape(IDN_provinces) + #Overlays each province of Indonesia
  tm_polygons("#FDAE6B", alpha = 0.9) + #set basecolor of the map
  #alpha changes the transparency
```

```
  tm_borders("#A1D99B") + #Add border to each province
  tm_shape(filtered_parks)+ #Highlight the national park df
  tm_fill(col = "National_Parks", #Select which column of the df to add
          palette = c("#E5F5E0", "#A1D99B", "#31A354"),
          #set colors of each national park as a hue from the RcolBrewer palette
          position= c("RIGHT", "BOTTOM")) + #Change the position
  tm_xlab("Longitude")+ #Add x axis label
  tm_ylab("Lattitude")+ #Add y axis label
  tm_compass(type = "4star", size = 0.3, position = c("RIGHT", "TOP")) +
  #Add a compass showing north, adjust its size and position on the map
  tm_scale_bar(width = 0.25, text.size = 0.3, position= c("LEFT", "BOTTOM")) +
  #Add a scale bar, change its width, size and position on the map
  tm_legend(position = c("RIGHT", "TOP"), #change legend position
            legend.outside = TRUE, #Move legend off the map
            legend.text.size = 1) + #Change size of legend text
  tm_credits("Data Source: GADM", #Add credits to source the map
             size= 0.4, align= "right", #Change size and alignment of credits
             position= c("RIGHT", "BOTTOM")) + #Change position of credits
  tm_layout(main.title = "Map of Indonesia Overlaid With National Parks",
             #Add a title to the map
            fontface = "italic", #Change the style of text
            fontfamily = "serif", #Change the font
            legend.width = 1, #Change legend width
            legend.height = 0.9, #Change legend height
            legend.title.size = 1, #Change legend title size
            legend.text.size = 0.6) #Change the text size of the legend
```

```
#Plot a basic map of Indonesia with national parks overlaid
#Add a dashed outline around Sumatra
#This will highlight where the national parks are localised

#To draw a boundary box around Sumatra we need to know the extent of Indonesia
#Find the extent of Indonesia
st_bbox(IDN_shapefile)
```

```
##       xmin       ymin       xmax       ymax
##  95.009705 -11.007615 141.019394   6.076941
```

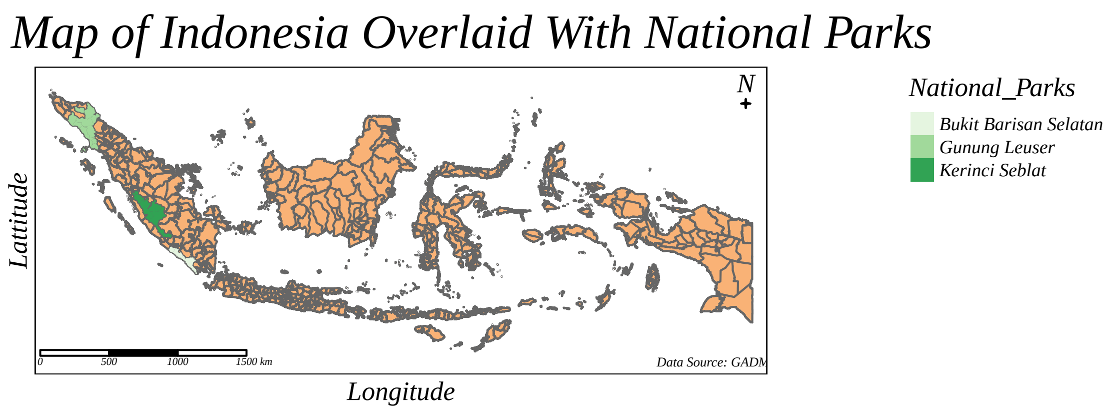
```
#reveals the x-lim, y-lim, x-max, and y-max of the current shapefile

# Use st_bbox to change the extent of the map and zoom into Sumatra
bbox_new <- st_bbox(IDN_shapefile)
bbox_new
```

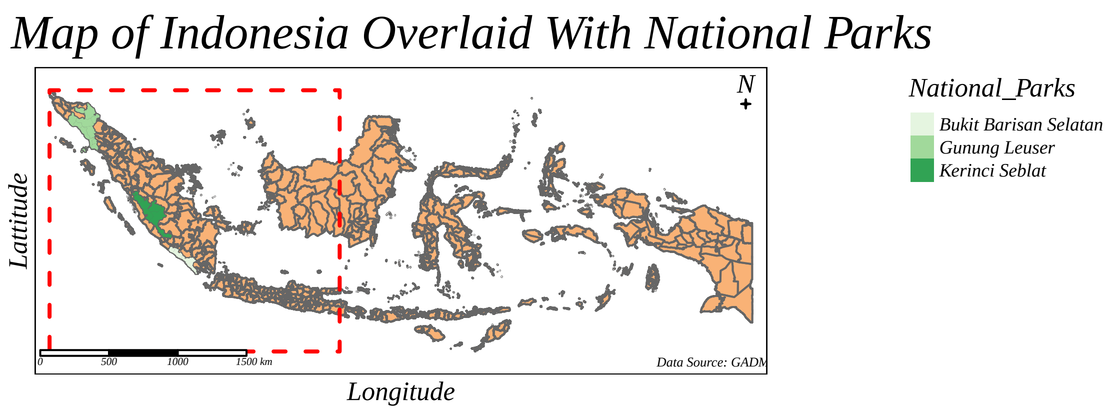
```
##       xmin       ymin       xmax       ymax
##  95.009705 -11.007615 141.019394   6.076941
```

```r
bbox_new[3] <- 114
#change the x-max value to zoom into Sumatra, [3] represents the x-max value
bbox_new <- bbox_new %>%
  st_as_sfc()
#Set it as spatial data

map_3 <- tm_shape(IDN_shapefile) + #Maps the outline of the shapefile
  tm_polygons() +
  tm_shape(IDN_provinces) + #Overlays the provinces of Indonesia onto the map
  tm_polygons("#FDAE6B", alpha = 0.9) + #set the color and transparency of map
  tm_borders("#A1D99B") + #Add borders to the provinces
  tm_shape(filtered_parks)+ #Add the national park dataframe
  tm_fill(col = "National_Parks", #Select which column of the df to add
          palette = c("#E5F5E0", "#A1D99B", "#31A354"),
          #set colors of each national park as a hue from the RcolBrewer palette
          position= c("RIGHT", "BOTTOM")) + #change the position
  tm_xlab("Longitude")+ #label the x axis
  tm_ylab("Lattitude")+ #label the y axis
  tm_compass(type = "4star", size = 0.3, position = c("RIGHT", "TOP")) +
   #Add a compass showing north, adjust its size and position on the map
  tm_scale_bar(width = 0.25, text.size = 0.3, position= c("LEFT", "BOTTOM")) +
   #Add a scale bar, change its width, size and position on the map
  tm_legend(position = c("RIGHT", "TOP"), #change legend position
            legend.outside = TRUE, #Move legend of the map
            legend.text.size = 1) + #Change legend text size
  tm_credits("Data Source: GADM", #Add credits to source the map
             size= 0.4, align= "right", #Change size and alignment of credits
             position= c("RIGHT", "BOTTOM")) + #change position of credits
  tm_layout(main.title = "Map of Indonesia Overlaid With National Parks",
            #Add title to the map
            fontface = "italic", #Change the style of the text
            fontfamily = "serif", #Change the font
            legend.width = 1, #change the legend width
            legend.height = 0.9, #change the legend height
            legend.title.size = 1, #change the size of the legend title
            legend.text.size = 0.6) + #change the size of the legend text
  tm_shape(bbox_new)+ #Add the new x-max value to the map
  tm_borders(col= "red", lwd= 2, lty= "dashed")
#col determines color of the border, lwd is line width, and lty is line style
```

```r
nationalpark <- tmap_arrange(map_2, map_3, #Group map_2 and map_3 into one view
  ncol = NA, #change number of columns
  nrow = NA, # change number of rows
  widths = NA, #change width of visualisation
  heights = NA, #change height of visualisation
  sync = FALSE,
  asp = 0, #Aspect ratio of visualisation
```

```
  outer.margins = 0.02 #Change the outer margins
)

print(nationalpark)
```






```
#save map
tmap_save(map,
  filename =
    "C:\\Users\\jadeh\\OneDrive\\Documents\\Coding Project\\Tmap Figures\\National_Parks.png")
```

Question 1: Where Do The Majority Of Sumatran Tigers Reside?

Through this visualisation, I have been able to highlight the three major national parks in which Sumatran tigers mostly reside. In doing do, this visualisation demonstrates how confined the tigers are in their habitats, as a direct result of deforestation.

This map also uses a bounding box to highlight Sumatra as the region of interest, as this is where **all** remaining Sumatran tigers reside, which emphasises their limitations in movement, and the likelihood of their extinction given the reduced viable habitation options. again emphasises their worldwide confinement an

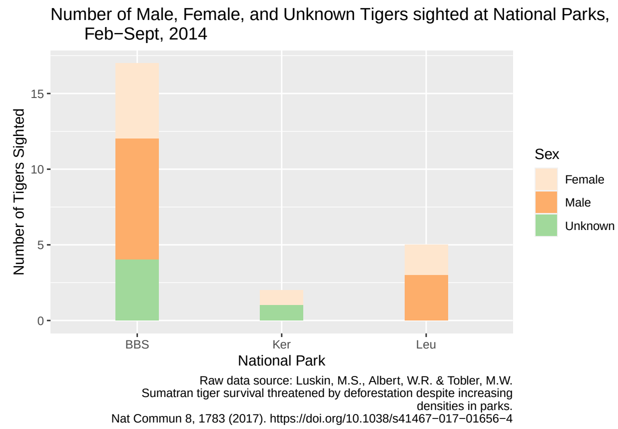
Data Visualisation 2: Zoomed in Map of Sumatra highlighting The Number And Gender Of Tigers Spotted in Each National Park

```
#This incorporates all the for loops into one loop to output the average points

# Making a list of zeros for the x coordinates to be added to
longitude = rep(0,length(Parks))
latitude = rep(0,length(Parks))

for(i in 1:length(Parks)){ # Looping through the different parks
  Specific_Park <- filtered_parks %>%
    filter(National_Parks == Parks[i]) # Selecting the specific park

  geom <- st_geometry(Specific_Park)

  #This allows us to fill in the coordinate data from the loops
  mean_x <- rep(0, nrow(Specific_Park))
  mean_y <- rep(0, nrow(Specific_Park))

  for(j in 1:nrow(Specific_Park)){
    multi <- geom[[j]] #Index the multipolgyon

    coords <- multi[[1]] #get x,y coordinates

    df_coords <- data.frame(coords[1]) #turn it into a dataframe

    means <- colMeans(df_coords) #overall mean x and y coordinates

    mean_x[j] <- means[1] #[1]extracts the average x value

    mean_y[j] <- means[2] #[2]extracts the average y value
  }
  longitude[i] <- mean(mean_x) #create a variable of the mean x coordinate

  latitude[i] <- mean(mean_y) #create a variable of the mean y coordinate
}

geocode <- data.frame(longitude, latitude)

geocode2 <- st_as_sf(geocode, coords= c("longitude", "latitude"), crs= 4326)




map_4 <- tm_shape(IDN_shapefile, bbox= bbox_new) +
  #add new boundary to the whole shapefile
```

```
tm_polygons() +

tm_shape(IDN_provinces, bbox= bbox_new) +
#add new boundary to the provinces df

tm_polygons("#FDAE6B", alpha = 0.9) +
#set colour of regions, change transparency with alpha

tm_borders("#A1D99B", lwd = 2) +
#colour of borders and width of border lines

tm_shape(filtered_parks, bbox= bbox_new)+
#add national park regions and add new boundary to the national parks df

tm_fill(col = "National_Parks", #overlay national park

        palette = c("#E5F5E0", "#A1D99B", "#31A354"),
        #Use RColBrewer palette to colour the national parks with green hues

        position= c("RIGHT", "BOTTOM")) +
#change the position

tm_xlab("Longitude")+ #Label x-axis

tm_ylab("Lattitude")+ #Label y-axis

tm_compass(type = "4star", size = 0.3,
           #Add a compass to indicate direction of North

           position = c("RIGHT", "TOP")) +
#Position where the compass goes on the map

tm_scale_bar(width = 0.25, text.size = 0.3,
             #Add a scale bar and adjust the size of text and  width of bar

             position= c("LEFT", "BOTTOM")) +
#Choose where to place the scale bar

tm_legend(position = c("RIGHT", "TOP"), legend.outside = TRUE,
          #Remove the legend off the map and position it to the right

          legend.text.size = 1) + #change size of legend text

tm_credits("Data Source: GADM", size= 0.4, align= "right",
           #Provide a data source for the map

           position= c("RIGHT", "BOTTOM")) +
```

```
  tm_layout(main.title = "Map of Indonesia Overlaid With National Parks",
            #Title the map

            fontface = "italic", #Set style of text to be italicized

            fontfamily = "serif", #Set font to Serif for aesthetics

            legend.width = 1, #change width of legend

            legend.height = 0.9, #Change the height of the legend bar

            legend.title.size = 1, #Change size of legend title

            legend.text.size = 0.6) + #change size of legend text size

  tm_shape(geocode2)+ #add coordinates to the map

  tm_dots(col="red", size=0.3) #change color of dots and their size

print(map_4) #print the map

#save map
tmap_save(map_4,
  filename =
    "C:\\Users\\jadeh\\OneDrive\\Documents\\Coding Project\\Tmap Figures\\Close_Up_Sumatra.png"

#Create a stacked bar chart
#First create a new column with total number of tigers spotted based on gender
#We need this numeric data to be able to plot the graph and pie chart
#Use case_when to do this
tiger_sighted <- tiger_map_data %>%
  mutate(tigers_spotted = case_when(
    Sex == "Female" &
      national_park== "BBS" ~"5",
    Sex == "Male" &
      national_park== "BBS" ~"8",
    Sex== "Unknown" &
      national_park== "BBS" ~"4",
    Sex == "Female" &
      national_park== "Leu" ~ "2",
    Sex == "Male" &
      national_park== "Leu" ~ "3",
    Sex == "Female" &
      national_park== "Ker" ~ "1",
    Sex == "Unknown" &
      national_park== "Ker" ~ "1"))
```

```
#There were no spotted unknown tigers in Leu and no males spotted at Ker
#so they cannot be included in this mutation


#First remove unnecessary columns
tiger_bar <- subset(tiger_sighted, select = -c(IndividualCode, date))


#Now use unique to find out the total number of each gender tiger in each park
new_tiger_bar <- tiger_bar %>%
  unique()

#Change tigers_spotted from character to numeric data to plot as the y axis
new_tiger_bar$tigers_spotted <- as.numeric(new_tiger_bar$tigers_spotted)


#Create a stacked bar chart to show the number of male vs female tigers spotted
ggplot(data=new_tiger_bar, aes(x=national_park, y=tigers_spotted, fill=Sex)) +
  geom_bar(stat="identity",position="stack", width= 0.3)+
  #stat='identity displays the sum of genders in each bar
  #stack gender bars on top of each other per each park
  scale_y_continuous() + # set y axis as continuous
  scale_fill_manual(values=c("#FEE6CE", "#FDAE6B", "#A1D99b")) +
  #fill bars withe custom palettes
  labs(title =
         "Number of Male, Female, and Unknown Tigers sighted at National Parks,
       Feb-Sept, 2014",
       #add a title to the bar chart
       caption = "Raw data source: Luskin, M.S., Albert, W.R. & Tobler, M.W.
       Sumatran tiger survival threatened by deforestation despite increasing
       densities in parks.
       Nat Commun 8, 1783 (2017). https://doi.org/10.1038/s41467-017-01656-4",
       #cite the data source
       x="National Park", y="Number of Tigers Sighted")
#label x and y axis

#save bar chart
ggsave(here("Ggplot Figures", "barchart.png"))
```



Question 2: How Many Sumatran Tigers Have Been Sighted in Each Park?

This main visualisation demonstrates that across the three national parks, only 24 Sumatran tigers have been spotted in total, with 17 in Bukit Barisan Selatan, 2 in Kerinci Seblat and 5 in Gunung Leuser. This visualisation also demonstrates both how many tigers are in each national park, but also, how confined these tigers are to smaller habitats that they would not typically inhabit. In particular, the tmap of national parks represents the limited number viable habitats across Sumatra, whereby these tiger are forced to migrate to protected heritage sites where the forests remain in tact.

This bar chart provides a good overview of the number and gender of the tigers residing in the national parks.

Question 3: Which National Park Has The Largest Sumatran tiger Population?

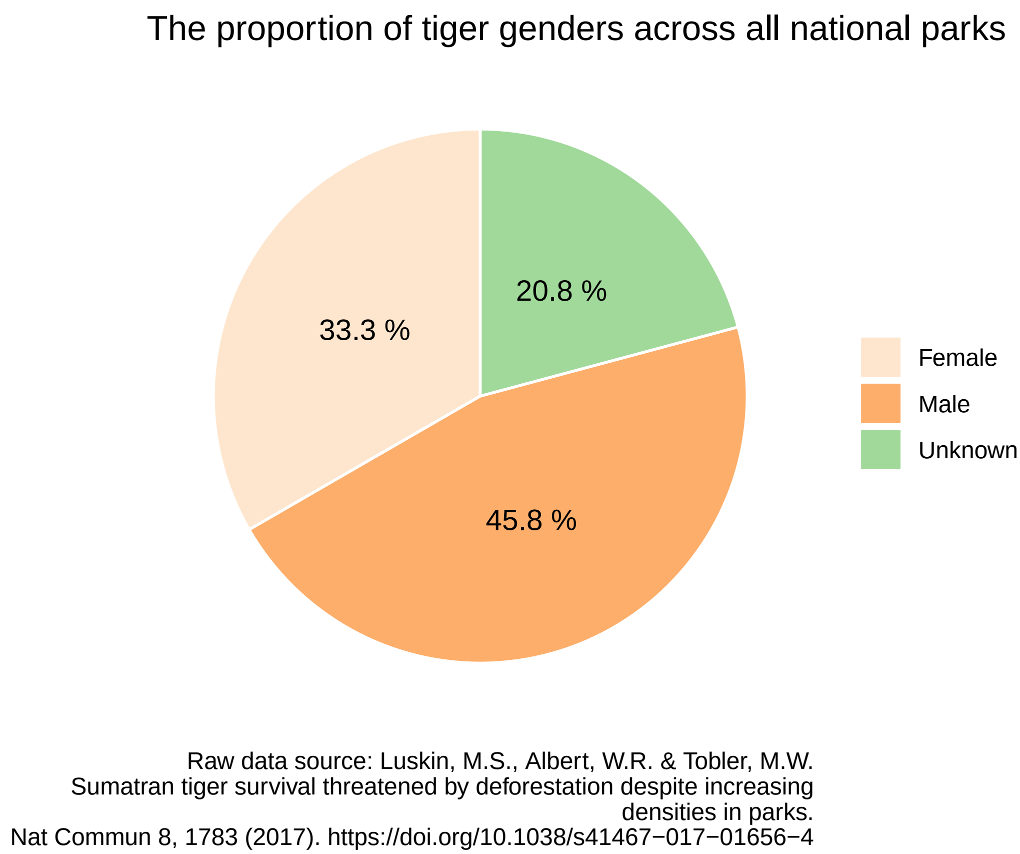
Interestingly, Bukit Barisan Selatan had the largest tiger population, despite being the smallest park. This could suggest that it has a more robust tiger population, with eight male and five female tiger, which increases the likelihood of breeding. However, it could also be that fewer tigers were sighted in the other parks as they are considerably larger in size, making tiger camera captures less likely. This is something that future research could establish with monitoring of tigers over a longer period of time, and with more cameras across the landscape.

Nonetheless, this bar graph demonstrates that ultimately, across eight months of filming, only 24 different tiger were spotted across three national parks, a finding which reiterates the importance of conserving the environment, and protecting these tiger from harm e.g. from poaching.

These two visualisations together are complimentary in meeting the primary aim of the study: elucidating how the number and gender of tigers inhabiting the three main Sumatran national parks.

Data Visualisation Number 3: Pie Chart Showing The Gender Proportions Of Tigers Sighted Across All National Parks

```
#Create a pie chart representing the percentage of each gender
pie_chart <- new_tiger_bar %>%
  group_by(Sex) %>%
  summarise(sum= sum(tigers_spotted)) #sum each gender across all parks


#Create the pie chart
ggplot(pie_chart, aes(x="", y= sum, fill=Sex)) +

  geom_bar(stat="identity", width=1, color="white") +
  coord_polar("y", start=0) + #use coord_polar to turn ggplot into a pie chart
  labs(x = NULL, y = NULL, fill = NULL,
       title= "The proportion of tiger genders across all national parks",
       #remove all labels and add a title
       caption = "Raw data source: Luskin, M.S., Albert, W.R. & Tobler, M.W.
       Sumatran tiger survival threatened by deforestation despite increasing
       densities in parks.
       Nat Commun 8, 1783 (2017). https://doi.org/10.1038/s41467-017-01656-4",
       x="National Park", y="Number of Tigers Sighted")+
  theme_classic() + #set the theme
  theme(axis.line = element_blank(),
        axis.text = element_blank(),
        axis.ticks = element_blank()) +
  scale_fill_manual(values=c("#FEE6CE", "#FDAE6B", "#A1D99b")) +
  #select colour palette of pie chart
  geom_text(aes(label = paste(round(sum / sum(sum) * 100, 1), "%")),
            #add percentages of genders onto the pie chart
            position = position_stack(vjust = 0.5))
```




```
#position the pie chart

#save pie chart
ggsave(here("Ggplot Figures", "piechart.png"))
```

Question 4: What can the gender proportions tell us about tiger populations?

This pie chart is important in elucidating the gender proportions of the entire population of Sumatran tigers in these three national parks. Whilst previous visualisations have demonstrated the gender split in each park, this visualisation, clearly shows that overall there is a larger proportion of males residing in these parks, with only a third of the population being female. This suggests that it may be harder to increase the tiger population and prevent extinction with only a small number of females remaining, and where male tigers in small environments are often in competition for females. However, the gender of 20.8% of the tigers that were sighted could not be identified, which makes it more challenging to comment on what this means for the tiger population more generally, yet, when assessing the visualisations together, it is clear to see that regardless of gender proportions, the tiger population is extremely small.

Conclusions

Overall, it is clear to see that the Sumatran tiger population is decreasing, with only 24 tigers sighted between February-September 2014 across three different national parks. A major contributor to this is the reduction in viable habitats as a result of deforestation and unsustinable agriculture, as represented by Visualisation 1 and 2, which highlight that the most viable habitats at present are three protected national parks. Alongside this, it is clear to see from both visualisation 2 and 3 that there is a larger proportion of male tigers comparatively to females which may exacerbate breeding difficulties and contribute to further decline.

The most important take home message from this project, therefore, is that more needs to be done to protect the habitats of the Sumatran tigers that remain, as their rapidly declining numbers is a direct result of deforestation and displacement, whereby the number of viable habitations is becoming increasingly restricted.

Limitations

A limitation of this project was that I was unable to extract the exact boundaries of the three national parks from the shapefile of Indonesia. This rendered the presentations of the parks less accurate in visualisation 1 and 2, however, the regions in which the parks were located has still been highlighted, and therefore, the visualisations do still reflect the location and approximate size of each park.

A further limitation is that the data pertaining to the number of tigers sighted was collected in 2014 and published in 2017, and since then, the tiger populations of Sumatra have changed. This limitation arose due to the time constraints of the project, whereby it was not possible to wait for the authors with more recently published data to get back to me. It would be interesting to have a more up-to-date representation of the number of remaining tigers to be able to make a comparison to this project. Despite this, the project is still able to convey the scarcity of remaining tigers, and represent how constrained they are in terms of viable habitats.

Future Directions

A second dataset was sent to me by the author Matthew Luskin, containing information about the type of land in which the tigers had been sighted. It would be interesting to expand this project further by assessing these additional variables.

It would also be interesting to make a comparison between the declining rates of Sumatran tigers and the levels of deforestation, as the time constraints of this module did not allow for such an extensive exploration, but is something that would be pertinent to the topic.

References

Flora & Fauna International. (2003). Sumatran Tiger. https://www.fauna-flora.org/species/sumatran-tiger/

Luskin, M.S., Albert, W.R. & Tobler, M.W. (2017) Sumatran tiger survival threatened by deforestation despite increasing densities in parks. *Nat Commun. 8*. 1783. https://doi.org/10.1038/s41467-017-01656-4

World Wide Fund for Nature. (2003). Sunda Tigers. https://www.worldwildlife.org/species/sunda-tiger#:~:text=Habitat%20for%20the%20Sumatran%20tiger,conversion%20are%20out%20of%20control.

UNESCO World Heritage Convention. (2023). World Heritage List. https://whc.unesco.org/en/list/?search=Kerinci+Seblat&order=country