A Comprehensive Analysis of Snakebite Envenomation Rates and Venom Diversity Across Regions

Mirza Ali Murtuza Ahmadi

2024-10-04

**Introduction**

Global snakebite envenoming is a leading cause of death among neglected tropical diseases (GBD 2019 Snakebite Envenomation Collaborators, 2022). Approximately 2.7 million people are envenomed by snakes each year, resulting in an estimated 81,000 to 138,000 deaths (Willyard, 2023). In addition, approximately 400,000 individuals experience permanent disabilities due to envenomation, which not only causes excruciating pain but can also result in severe swelling, kidney damage, and tissue necrosis, often necessitating amputation or skin grafting (Willyard, 2023). Small children are especially vulnerable to these effects (Willyard, 2023).

Due to the severity of snakebite envenomation, preventing death and injury is critical. One of the most effective treatments is antivenom, which neutralizes the toxins in snake venom (Leon et al., 2013). However, since different snake families produce different venom types, it is essential that countries most affected by envenomations have access to antivenoms in the correct proportions, tailored to the snake species prevalent in their regions (Leon et al., 2013). Therefore, using BOLD sampling data for venomous snakes in different regions, envenoming incidence rate data by country (sourced from Kasturiratne et al., 2008), and venom protein proportion data in different types of snake venom (sourced from Tasoulis and Isbister, 2017) , this investigation looks to answer the following question: **How does the geographic distribution of envenomation rates correlate with the prevalence of different venomous snake families, and how can this information guide the regional prioritization of antivenom distribution?** In this study, we will be focusing on the snake families Viperidae (vipers) and Elapidae (elapids) as they are two of the most medically significant venomous snake families and are responsible for the majority of envenomings worldwide (Tasoulis and Isbister, 2017).

*Below is the code and the resulting figures that illustrate the envenomation incidence rate by country, the proportions of vipers and elapids per country, and the variations in venom protein proportions and venom types between vipers and elapids.*

library(BiocManager)

library(remotes)

library(bold)

library(tidyverse)

library(readxl)

library(maps)

library(ggplot2)  
  
library(dplyr)  
  
library(tidyr)  
  
library(RColorBrewer)  
  
library(viridis)

library(ggfortify)

# Loading and filtering data ----  
raw\_external\_envenomations\_df <- read\_excel("../data/Estimation of the total number of snake bite envenomings by country.xlsx")

# Data inspection is done frequently to ensure the structure, type, organization of data set is understood before working with the data  
dim(raw\_external\_envenomations\_df)

## [1] 233 9

names(raw\_external\_envenomations\_df)

## [1] "Global Burden Region"   
## [2] "Country"   
## [3] "Incidence per 100 000 population (Lower)"   
## [4] "Source...4"   
## [5] "Incidence per 100 000 population (Higher)"  
## [6] "Source...6"   
## [7] "Current population"   
## [8] "Number of cases (Low)"   
## [9] "Number of cases (High)"

# Only keep 'Country' and 'Incidence' columns because they are relevant for analysis  
external\_envenomations\_df <- raw\_external\_envenomations\_df %>%  
 select(  
 Country,  
 `Incidence per 100 000 population (Lower)`,  
 `Incidence per 100 000 population (Higher)`  
 )  
names(external\_envenomations\_df)

## [1] "Country"   
## [2] "Incidence per 100 000 population (Lower)"   
## [3] "Incidence per 100 000 population (Higher)"

# Create average incidence rate from the higher and lower bounds in order to have a single, representative measure of envenoming incidence for each measure  
filtered\_external\_envenomations\_df <- external\_envenomations\_df %>%  
 mutate(Average\_Incidence\_rate\_per\_100k = (`Incidence per 100 000 population (Lower)` + `Incidence per 100 000 population (Higher)`) / 2) %>%  
 select(Country, Average\_Incidence\_rate\_per\_100k)  
names(filtered\_external\_envenomations\_df)

## [1] "Country" "Average\_Incidence\_rate\_per\_100k"

# squamata\_df <- as\_tibble(bold\_specimens(taxon='squamata'))  
squamata\_df <- read\_tsv("../data/squamata\_data.tsv")

dim(squamata\_df)

## [1] 21980 68

names(squamata\_df)

## [1] "processid" "sampleid"   
## [3] "recordID" "catalognum"   
## [5] "fieldnum" "institution\_storing"   
## [7] "collection\_code" "bin\_uri"   
## [9] "phylum\_taxID" "phylum\_name"   
## [11] "class\_taxID" "class\_name"   
## [13] "order\_taxID" "order\_name"   
## [15] "family\_taxID" "family\_name"   
## [17] "subfamily\_taxID" "subfamily\_name"   
## [19] "genus\_taxID" "genus\_name"   
## [21] "species\_taxID" "species\_name"   
## [23] "subspecies\_taxID" "subspecies\_name"   
## [25] "identification\_provided\_by" "identification\_method"   
## [27] "identification\_reference" "tax\_note"   
## [29] "voucher\_status" "tissue\_type"   
## [31] "collection\_event\_id" "collectors"   
## [33] "collectiondate\_start" "collectiondate\_end"   
## [35] "collectiontime" "collection\_note"   
## [37] "site\_code" "sampling\_protocol"   
## [39] "lifestage" "sex"   
## [41] "reproduction" "habitat"   
## [43] "associated\_specimens" "associated\_taxa"   
## [45] "extrainfo" "notes"   
## [47] "lat" "lon"   
## [49] "coord\_source" "coord\_accuracy"   
## [51] "elev" "depth"   
## [53] "elev\_accuracy" "depth\_accuracy"   
## [55] "country" "province\_state"   
## [57] "region" "sector"   
## [59] "exactsite" "image\_ids"   
## [61] "image\_urls" "media\_descriptors"   
## [63] "captions" "copyright\_holders"   
## [65] "copyright\_years" "copyright\_licenses"   
## [67] "copyright\_institutions" "photographers"

# Subset data to only include relevant venomous snake families for analysis  
vipers\_elapids\_df <- subset(squamata\_df, family\_name == c("Viperidae", "Elapidae"))  
dim(vipers\_elapids\_df)

## [1] 847 68

# Filter data to include only country and family\_name for analyzing distinct family sampling by country  
filtered\_countries\_from\_BOLD <- vipers\_elapids\_df %>%  
 select(country, family\_name) %>%  
 distinct(country) %>%  
 mutate(across(where(is.character), ~ na\_if(.x, ""))) %>%  
 filter(!is.na(country)) # *Filter NA values as they are not useful for geographic mapping and visualizations related to geographic distance* names(filtered\_countries\_from\_BOLD)

## [1] "country"

# Load in venom protein proportion data in order to distinguish differing venom types of different snake families in analysis  
unique\_species\_df <- vipers\_elapids\_df %>%  
 distinct(species\_name, .keep\_all = TRUE)  
dim(unique\_species\_df)

## [1] 178 68

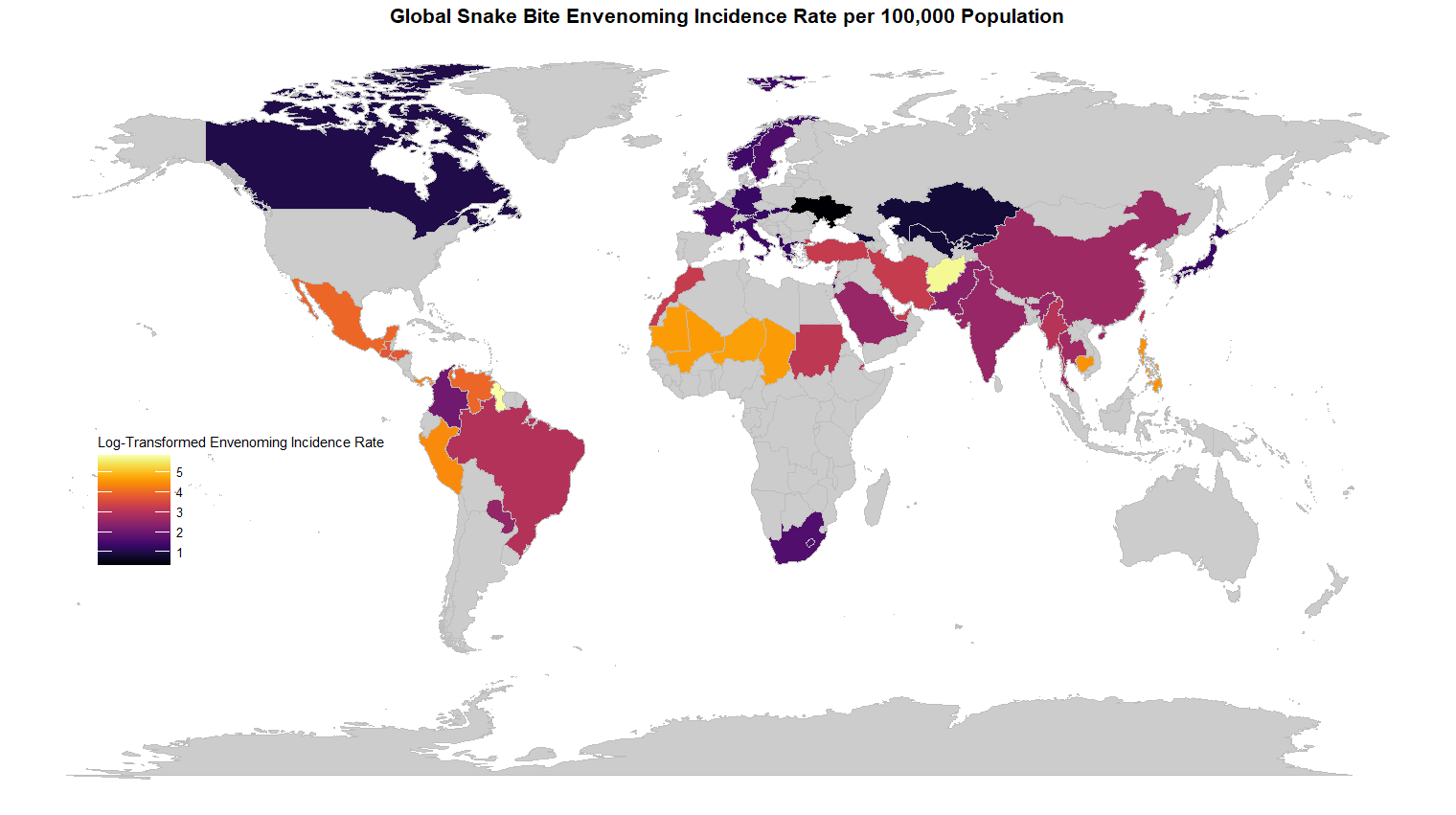
venom\_proportions\_df <- read\_excel("../data/Venom Protein Proportion Data.xlsx") %>%  
 inner\_join(unique\_species\_df, by = c("SPECIES" = "species\_name")) %>%  
 select(SPECIES, PLA2, SVSP, SVMP, `3FT`, LAAO, CRiSP, `CTL/SNACLEC`, DIS, NP, KUN, VEGF, CYS, DEF, MPi, VT, family\_name)  
dim(venom\_proportions\_df)

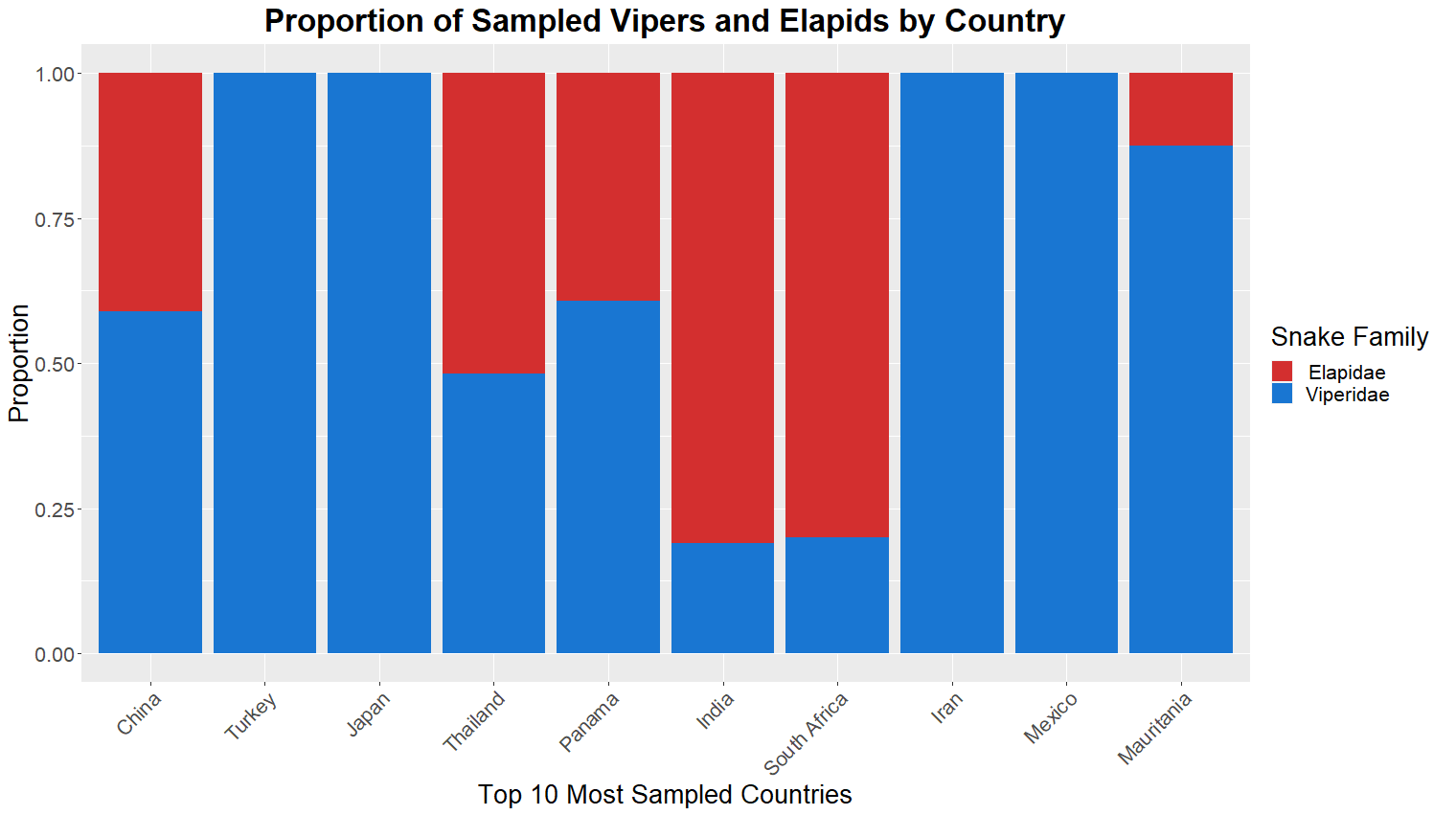
## [1] 68 17

names(venom\_proportions\_df)

## [1] "SPECIES" "PLA2" "SVSP" "SVMP" "3FT"   
## [6] "LAAO" "CRiSP" "CTL/SNACLEC" "DIS" "NP"   
## [11] "KUN" "VEGF" "CYS" "DEF" "MPi"   
## [16] "VT" "family\_name"

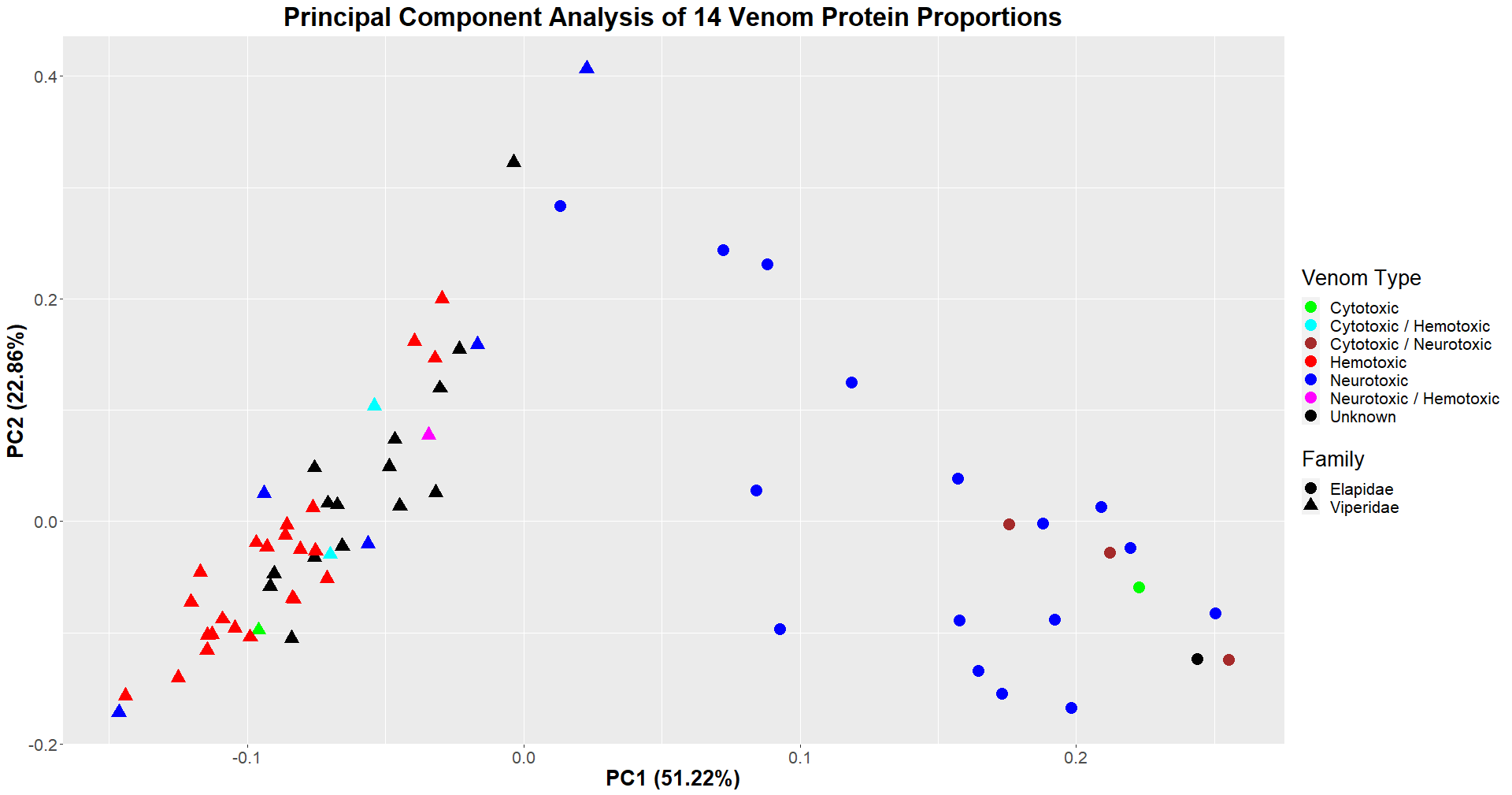
Figure 1: Create envenoming incidence map by country ----  
merged\_country\_incidence\_df <- filtered\_external\_envenomations\_df %>%  
 inner\_join(filtered\_countries\_from\_BOLD, by = c("Country" = "country"))  
  
# Merge country lat and long data to create a map of countries in order to visually represent envenoming incidence per country  
world\_map <- map\_data("world")  
merged\_map\_data <- world\_map %>%  
 left\_join(merged\_country\_incidence\_df, by = c("region" = "Country")) %>%  
 select(-subregion)  
  
# The envenoming incidence rate is log-transformed because it makes it easier to interpret changes in incidence rates  
map <- ggplot(data = merged\_map\_data, aes(x = long, y = lat, group = group, fill = log(Average\_Incidence\_rate\_per\_100k + 1))) +  
 geom\_polygon(color = "grey") +  
 scale\_fill\_viridis(option = "B", na.value = "grey80") +  
 labs(  
 title = "Global Snake Bite Envenoming Incidence Rate per 100,000 Population",  
 fill = "Log-Transformed Envenoming Incidence Rate"  
 ) +  
 theme\_void() +  
 theme(  
 plot.title = element\_text(hjust = 0.5, size = 16, face = "bold"),  
 legend.position = c(0.17, 0.4),  
 legend.title = element\_text(size = 11, face = "plain", hjust = 0, margin = margin(r = 10)),  
 legend.text = element\_text(size = 10, hjust = 0.5),  
 legend.key.width = unit(2, "cm"),  
 legend.box.margin = margin(10, 10, 10, 10)  
 )  
  
print(map)

# Figure 2: Stacked bar chart of the most prevalent snake families in the top ten countries with the highest envenoming incidence ----  
BOLD\_sampling\_per\_country\_df <- vipers\_elapids\_df %>%  
 group\_by(country) %>%  
 summarise(Count = n()) %>%  
 na.omit(country)  
  
# Select only the 10 most sampled countries to compare viper and elapid proportions in order to keep the graph clear and uncluttered  
most\_sampled\_countries <- BOLD\_sampling\_per\_country\_df %>%  
 arrange(desc(Count)) %>%  
 head(15) # The top 15 countries are specified, but only 10 appear in the figure because 5 of them are missing from the BOLD viper/elapid data set.  
  
  
family\_proportions\_per\_country\_df <- vipers\_elapids\_df %>%  
 select(country, family\_name) %>%  
 mutate(across(where(is.character), ~ na\_if(.x, ""))) %>%  
 filter(!is.na(country)) %>%  
 group\_by(country) %>%  
 summarise(  
 Proportion\_of\_vipers\_sampled = sum(family\_name == "Viperidae") / (sum(family\_name == "Viperidae") + sum(family\_name == "Elapidae")),  
 Proportion\_of\_elapids\_sampled = sum(family\_name == "Elapidae") / (sum(family\_name == "Elapidae") + sum(family\_name == "Viperidae"))  
 ) %>%  
 ungroup() %>%  
 inner\_join(merged\_country\_incidence\_df, by = c("country" = "Country")) %>%  
 arrange(desc(Average\_Incidence\_rate\_per\_100k)) %>%  
 select(-Average\_Incidence\_rate\_per\_100k) %>%  
 filter(country %in% most\_sampled\_countries$country) %>%  
 mutate(country = factor(country, levels = most\_sampled\_countries$country))  
  
  
# Reshape data from wide to long format is done because it makes the data more manageable and ensuring compatibility with analytical tools and visualization libraries, leading to clearer insight  
data\_long <- family\_proportions\_per\_country\_df %>%  
 pivot\_longer(  
 cols = c(Proportion\_of\_vipers\_sampled, Proportion\_of\_elapids\_sampled),  
 names\_to = "Snake\_Family",  
 values\_to = "Proportion"  
 ) %>%  
 mutate(country = factor(country, levels = most\_sampled\_countries$country)) # Set country as a factor to maintain the same order when displayed on x axis  
  
  
# Create the stacked bar chart  
ggplot(data\_long, aes(x = country, y = Proportion, fill = Snake\_Family)) +  
 geom\_bar(stat = "identity") +  
 scale\_fill\_manual(  
 values = c("#D32F2F", "#1976D2"),  
 labels = c("Elapidae", "Viperidae")  
 ) + # Set colors for viper and elapid  
 labs(  
 title = "Proportion of Sampled Vipers and Elapids by Country",  
 x = "Top 10 Most Sampled Countries",  
 y = "Proportion"  
 fill = "Snake Family"  
 ) +  
 theme(  
 axis.text.x = element\_text(angle = 45, hjust = 1, size = 16),  
 axis.text.y = element\_text(family = "bold", size = 16),  
 axis.title.x = element\_text(size = 20),  
 axis.title.y = element\_text(size = 20),  
 plot.title = element\_text(hjust = 0.5, size = 25, face = "bold")  
 legend.title = element\_text(size = 20, face = "plain", hjust = 0, margin = margin(r = 10)),   
 legend.text = element\_text(size = 16, hjust = 0.5),   
 )



# Figure 3: Create PCA plot with venom protein proportion data ----  
# Centering is important for PCA as it prevents bias towards variables with larger means, ensuring accurate representation of variance.  
venom\_protein\_proportions\_df.pca <- prcomp(venom\_proportions\_df[, c(2:15)], center = TRUE) # Filter to retain only numeric columns, as PCA requires numeric input.  
  
summary(venom\_protein\_proportions\_df.pca)

## Importance of components:  
## PC1 PC2 PC3 PC4 PC5 PC6 PC7  
## Standard deviation 32.7474 21.8771 16.9081 9.63688 7.95439 6.23216 4.6888  
## Proportion of Variance 0.5122 0.2286 0.1366 0.04436 0.03022 0.01855 0.0105  
## Cumulative Proportion 0.5122 0.7409 0.8774 0.92176 0.95199 0.97054 0.9810  
## PC8 PC9 PC10 PC11 PC12 PC13 PC14  
## Standard deviation 3.83688 3.44344 2.74087 1.5823 1.23166 1.09145 0.62538  
## Proportion of Variance 0.00703 0.00566 0.00359 0.0012 0.00072 0.00057 0.00019  
## Cumulative Proportion 0.98807 0.99374 0.99732 0.9985 0.99924 0.99981 1.00000

pca\_scores <- as.data.frame(venom\_protein\_proportions\_df.pca$x)  
  
# ensure the pca\_scores data frame has an according species column as each point in the plot will represent a viper or elapid species  
pca\_scores$SPECIES <- venom\_proportions\_df$SPECIES  
  
# Different colours are chosen to represent various snake venom types  
autoplot(venom\_protein\_proportions\_df.pca,  
 data = venom\_proportions\_df,  
 colour = "VT",  
 shape = "family\_name"  
) +  
 geom\_point(aes(color = venom\_proportions\_df$VT, shape = venom\_proportions\_df$family\_name), size = 3, alpha = 0.8) +  
 scale\_color\_manual(values = c("green", "cyan", "brown", "red", "blue", "magenta", "black"), name = "Venom Type") +  
 scale\_shape(name = "Family") +  
 ggtitle("Principal Component Analysis of 14 Venom Protein Proportions") +  
 theme(  
 plot.title = element\_text(hjust = 0.5, face = "bold", size = 16),  
 axis.title.x = element\_text(face = "bold", size = 14),  
 axis.title.y = element\_text(face = "bold", size = 14)  
 )

**Results and Discussion**

After analyzing data from multiple sources, I gained valuable insights into global snake venom distribution and the differences in venom types between vipers and elapids. The first figure illustrates the global incidence rate of snakebite envenomation, with varying colors representing different levels of envenomation. As expected, tropical countries near the equator, such as those in South America, Asia, and Africa, show the highest rates, likely due to increased species richness in these regions (Brown, 2014). The top three countries with the highest incidence rates are Guyana, Afghanistan, and Trinidad and Tobago. The second figure shows the proportions of vipers and elapids sampled in the top ten most sampled countries from the BOLD database. Vipers dominate in most countries and are ubiquitously sampled in Turkey, Japan, Iran, and Mexico. However, elapids are more prevalent in regions like India and South Africa, while some countries show a balanced proportion of both snake families. Lastly, the PCA plot illustrates significant differences in venom protein proportions and venom type between vipers and elapids. The distinct clustering of species reflects their evolutionary divergence, which was expected because vipers and elapids have been separate groups for the last 40-50 million years, enabling them to develop different venom types during this time (Reyes-Valesco et al., 2014). The PCA plot illustrates that most elapids possess either purely neurotoxic venom or a neurotoxic mix that targets the nervous system, including the brain (Tasoulis and Isbister, 2017). In contrast, most vipers exhibit either purely hemotoxic venom or a hemotoxic mix that affects the cardiovascular system, particularly the heart and blood (Tasoulis and Isbister, 2017).

These visualizations highlight crucial data for understanding snake venom distribution and envenomation risks, emphasizing the need to ensure an adequate supply of antivenoms in regions with high envenomation rates. Matching the supply of antivenoms to venom types of local snake species can prevent countless deaths and disabilities. While there was adequate data to address my original question, additional information regarding sampling, envenomation incidence rates, and viper/elapid proportions across countries would have significantly enhanced the scope of our project. This would enable a deeper understanding of the distribution and relative proportions of venomous snakes across a broader range of countries. Future work could focus on expanding data collection, including smaller venomous species outside of Viperidae and Elapidae, and extending this approach to other venomous animals, such as lizards, scorpions and insects, to better assess antivenom needs globally for a range of venomous taxa.

**Acknowledgements**

I am grateful for the support and advice from my classmates, which helped me complete a project I am proud of. Thomas Tekle guided me in refining my research question before I began coding, allowing me to approach it from different angles and ensure it was significant and data-rich. He also assisted me in correctly importing data from BOLD when I encountered errors. Isha Baxi and Dylan Harding provided valuable feedback on my figures, enhancing their readability and impact. For instance, Isha recommended reducing the number of countries sampled in Figure two, which improved clarity and highlighted the most relevant data. Additionally, after struggling for 45 minutes to load my TSV file of BOLD data into R, Dylan took only 30 seconds to identify that I had downloaded it as a .txt file instead, saving me potentially an hour of troubleshooting.

**References**

Brown J. H. (2014). Why are there so many species in the tropics?. Journal of biogeography, 41(1), 8–22. <https://doi.org/10.1111/jbi.12228>

GBD 2019 Snakebite Envenomation Collaborators. Global mortality of snakebite envenoming between 1990 and 2019. Nat Commun 13, 6160 (2022). <https://doi.org/10.1038/s41467-022-33627-9>

Kasturiratne, A., Wickremasinghe, A. R., de Silva, N., Gunawardena, N. K., Pathmeswaran, A., Premaratna, R., Savioli, L., Lalloo, D. G., & de Silva, H. J. (2008). The global burden of snakebite: a literature analysis and modelling based on regional estimates of envenoming and deaths. *PLoS medicine*, *5*(11), e218. <https://doi.org/10.1371/journal.pmed.0050218>

Reyes-Velasco, J., Card, D. C., Andrew, A. L., Shaney, K. J., Adams, R. H., Schield, D. R., Casewell, N. R., Mackessy, S. P., & Castoe, T. A. (2015). Expression of venom gene homologs in diverse python tissues suggests a new model for the evolution of snake venom. Molecular biology and evolution, 32(1), 173–183. <https://doi.org/10.1093/molbev/msu294>

Tasoulis, T., & Isbister, G. K. (2017). A Review and Database of Snake Venom Proteomes. Toxins, 9(9), 290. <https://doi.org/10.3390/toxins9090290>

Willyard, C. (2023, September 20). Global mortality of snakebite envenoming between 1990 and 2019. Nature News. <https://www.nature.com/articles/s41467-022-33627-9>