

SETH L.U.J. & SIR M,V, COLLEGE

PRAC 10-12

AIM:-

10 Creating graphical reports using `ggplot2` (R).

OUTPUT:-

The screenshot shows the RStudio interface with two main panels. The left panel is the Console, displaying R code and its output. The right panel is the Environment pane, showing the global environment and a ggplot2 visualization titled "Top 5 Diseases by Habitat".

Console Output:

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
R 4.5.2 - / -
The downloaded binary packages are in
  C:\Users\Hp\AppData\Local\Temp\Rtmpov244K\downloaded_packages
>
> library(ggplot2)
> 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)
>
> bat_data <- read.csv("C:/Users/Hp/downloads/bats_informations.csv")
>
> str(bat_data)
'data.frame': 30000 obs. of 11 variables:
 $ Shape : chr "Forelimbs adapted as wings (patagium)" "Forelimbs adapted as wings (patagium)" "Forelimbs adapted as wings (patagium)" "Distinctive facial structures (nose-leafs)" ...
 $ Species : chr "Yangochiroptera (Microbat); Example: Big Brown Bat" "Yangochiroptera (Microbat); Example: Little Brown Bat" "Yangochiroptera (Microbat); Example: Brandt's Bat" "Vesperopterochiroptera (Megabat); Example: Horseshoe Bat" ...
 $ Habitat : chr "Rock crevices" "Human-made structures" "Tree hollows" "Caves" ...
 $ Average_Lifespan : chr "15-20 years" "20-25 years" "25-30 years" "10-15 years" ...
 $ Environment : chr "specific temperature/humidity needs" "Adaptation to various climates" "specific temperature/humidity needs" "specific temperature/humidity needs" ...
 $ Food : chr "Nectarivorous" "Carnivorous" "Carnivorous" "Carnivorous" ...
 $ Articironous" : chr "Nectarivorous" "Carnivorous" "Carnivorous" "Carnivorous" ...
 $ Parasitic_infections : chr "Parasitic infections (mites)" "Parasitic infections (mites)" "Parasitic infections (mites)" "Parasitic infections (mites)" ...
 $ Disease : chr "Coronaviruses" "Histoplasmosis" "Parasitic infections (mites)" "ticks" "White-Nose Syndrome" ...
 $ ticks : chr "ticks" "ticks" "ticks" "ticks" "ticks" ...
 $ Global_Environment : chr "Global Environment" "Global Environment" "Global Environment" "Global Environment" "Global Environment" ...
  .. and 6 more variables: top_diseases chr [1:5] "Coronavirus" "Histoplasmosis" "White-Nose Syndrome" "ticks" ...'
```

Environment pane:

Global Environment

top_diseases

Disease

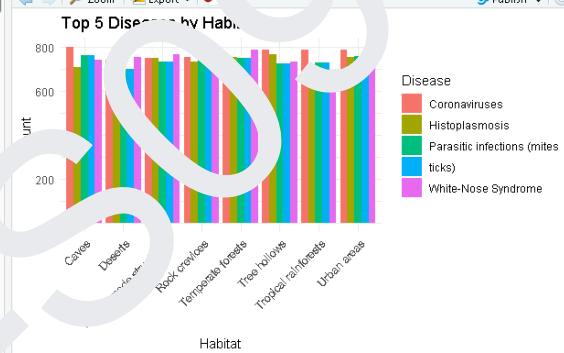
- Coronaviruses
- Histoplasmosis
- Parasitic infections (mites)
- ticks
- White-Nose Syndrome

Unit

Habitat

Top 5 Diseases by Habitat

Habitat	Coronaviruses	Histoplasmosis	Parasitic infections (mites)	ticks	White-Nose Syndrome
Caves	~750	~750	~750	~750	~750
Deserts	~700	~700	~700	~700	~700
Rock crevices	~750	~750	~750	~750	~750
Temperate forests	~750	~750	~750	~750	~750
Tropical rainforests	~750	~750	~750	~750	~750
Urban areas	~750	~750	~750	~750	~750



The figure shows a dual-pane RStudio interface. The left pane displays R code and its output, while the right pane shows a data environment and a generated plot.

Code Output:

```
R - R 4.5.2 - /~/
6 white-Nose Syndrome
>
> print(ggplot(bat_data, aes(x = species)) +
+   geom_bar() +
+   labs(title = "Bat Distribution by Suborder", x = "Bat suborder", y = "Count")
+
+   theme_minimal() +
+   theme(axis.text.x = element_text(angle = 45, hjust = 1)))
>
> print(ggplot(bat_data, aes(x = habitat)) +
+   geom_bar() +
+   labs(title = "Bat Habitat Distribution by Habitat Type", y = "Number of Bats")
+
+   theme_minimal() +
+   theme(axis.text.x = element_text(angle = 45, hjust = 1)))
>
> bat_d_clean <- bat_d %>%
+   mutate(
+     avg_lifespan = as.numeric(gsub("-.*", "", Average.Lifespan)),
+     max_lifespan = as.numeric(gsub("-.*", "", Average.Lifespan)),
+     lifespan_mean = ifelse(is.na(max_lifespan), avg_lifespan, (avg_lifespan + max_lifespan) / 2)
+   )
+
Warning message:
There was 1 warning: i state().
In argument: 'max_life.span = as.numeric(gsub("-.", "", Average.Lifespan))'.
Caused by warning:
! NAs introduced by coercion
>
> print(ggplot(bat_d_clean, aes(x = species, y = lifespan_midpoint)) +
+   geom_boxplot() +
+   labs(title = "Lifespan Distribution by Bat suborder", x = "Bat suborder", y =
"Average Lifespan (Years)")
+
+   theme_minimal())
>
> print(ggplot(bat_data, aes(x = Food)) +
```

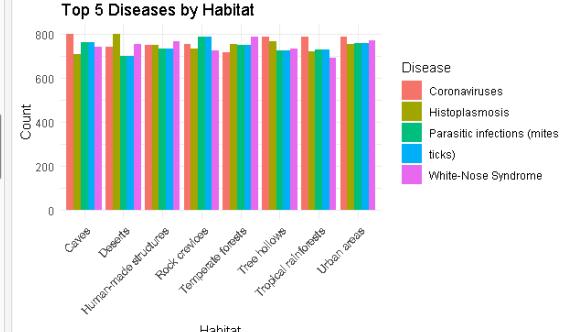
Environment View:

The environment view lists the following objects:

- bat_data_clean: 30000 obs. of 14 variables
- disease_by_habitat: 48 obs. of 3 variables
- disease_by_habitat...: 40 obs. of 3 variables
- disease_counts: 6 obs. of 2 variables
- food_by_suborder: 90 obs. of 3 variables
- protection_counts: 6 obs. of 2 variables
- values
- top_diseases: chr [1:5] "Coronaviruses" "Histoplasmosis" "white-...

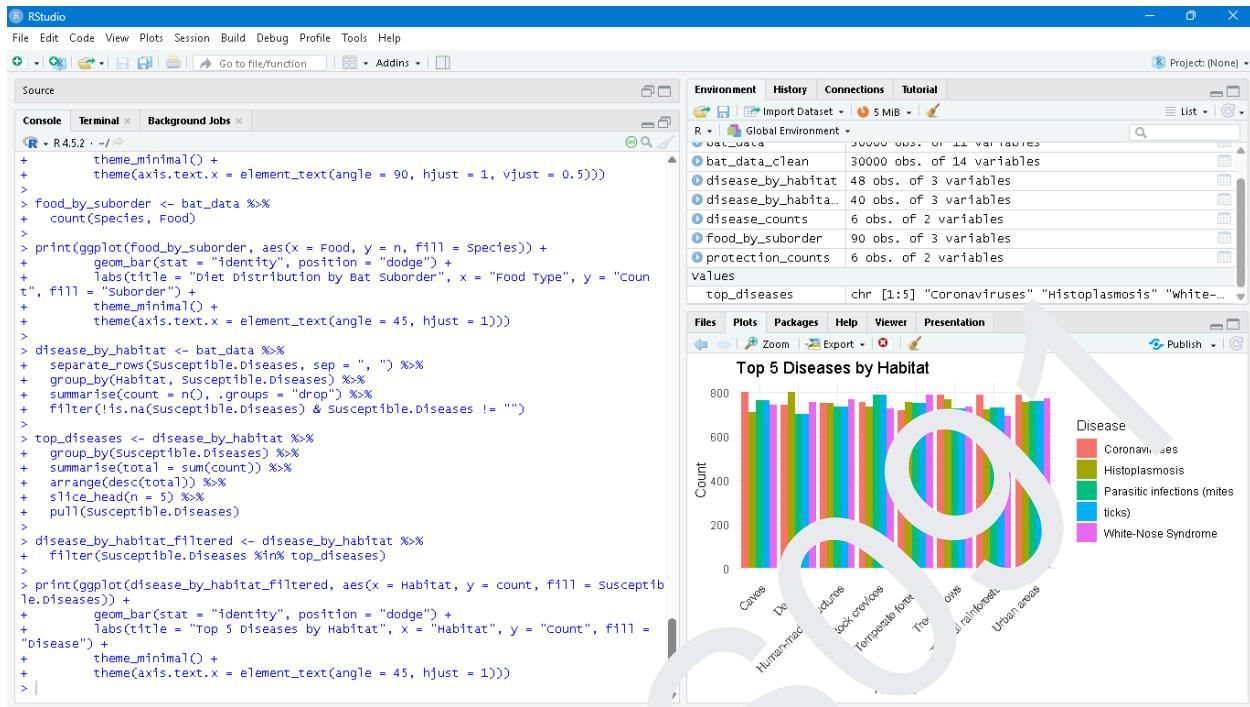
Plots:

A bar chart titled "Top 5 Diseases by Habitat" is displayed. The Y-axis is labeled "Count" and ranges from 0 to 800. The X-axis is labeled "Habitat" and includes categories: Caves, Deserts, Human-made structures, Rock crevices, Temperate forests, Tree hollows, Tropical rainforests, and Urban areas. The legend identifies five diseases: Coronaviruses (red), Histoplasmosis (olive green), Parasitic infections (mites) (teal), ticks (blue), and White-Nose Syndrome (pink). The chart shows varying counts across habitats for each disease.



SETH L.U.J. & SIR M,V, COLLEGE

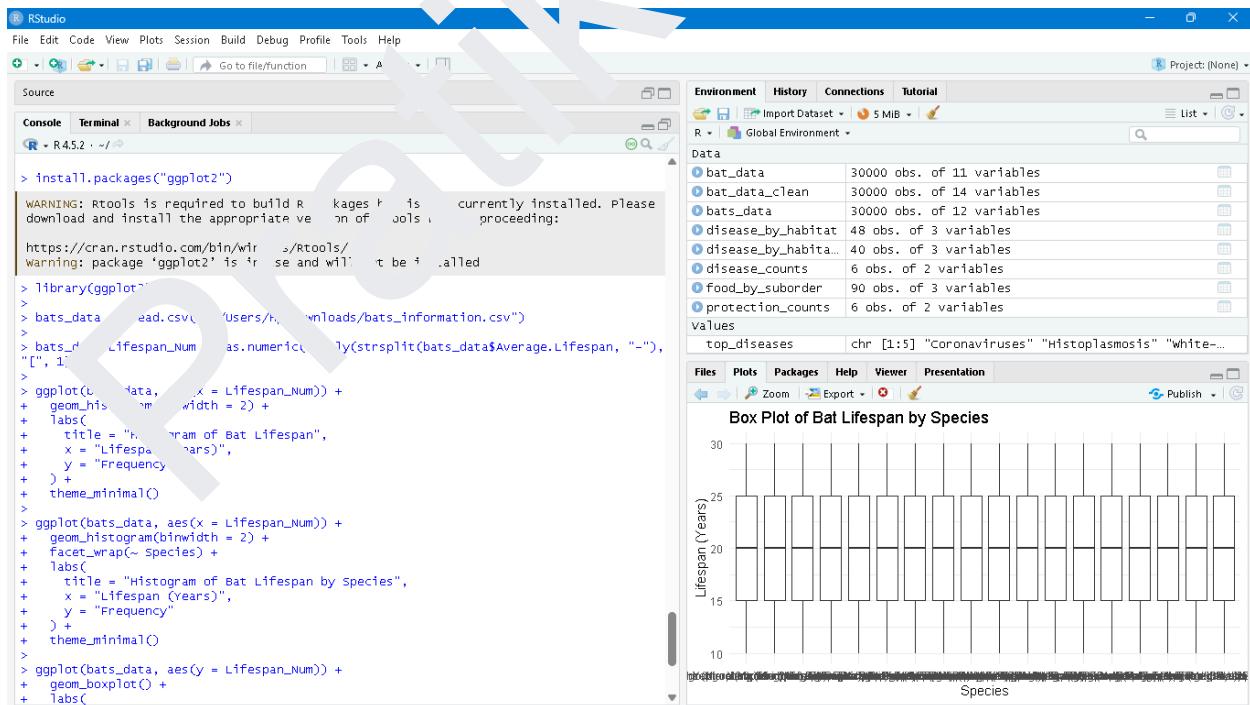
PRAC 10-12



AIM:-

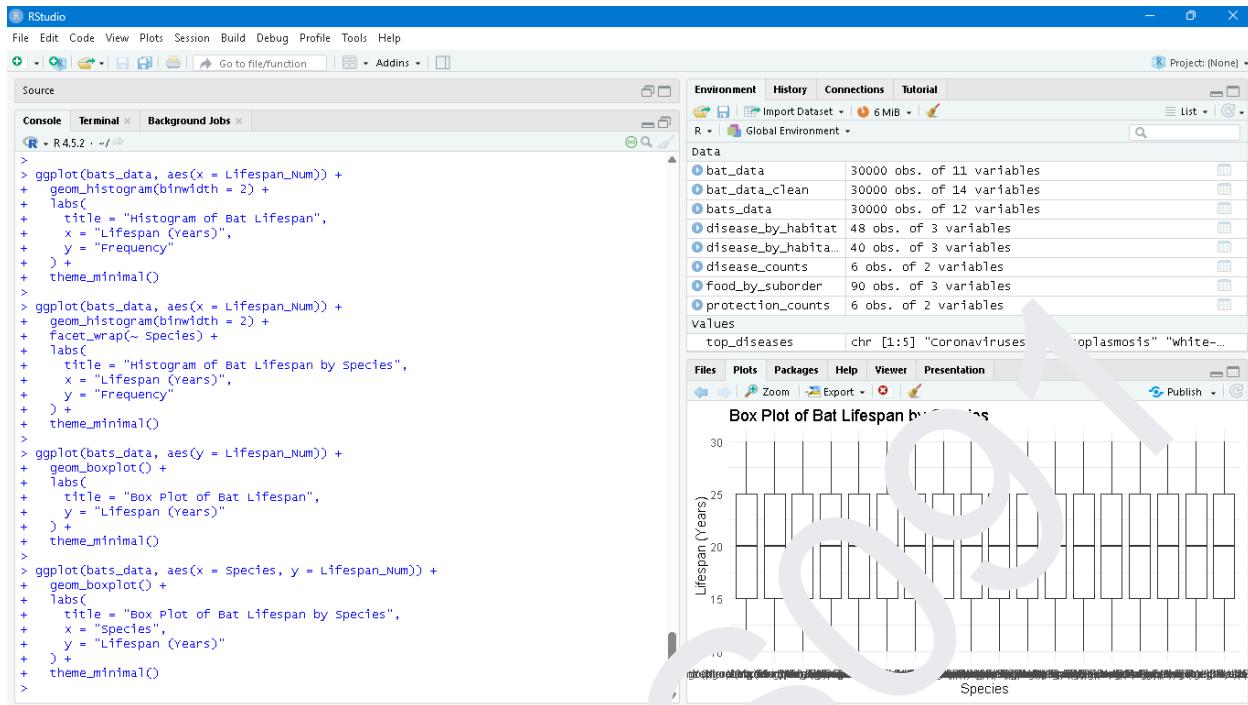
11 Generating histograms and box plots using ggplot2 (R).

OUTPUT:-



SETH L.U.J. & SIR M,V, COLLEGE

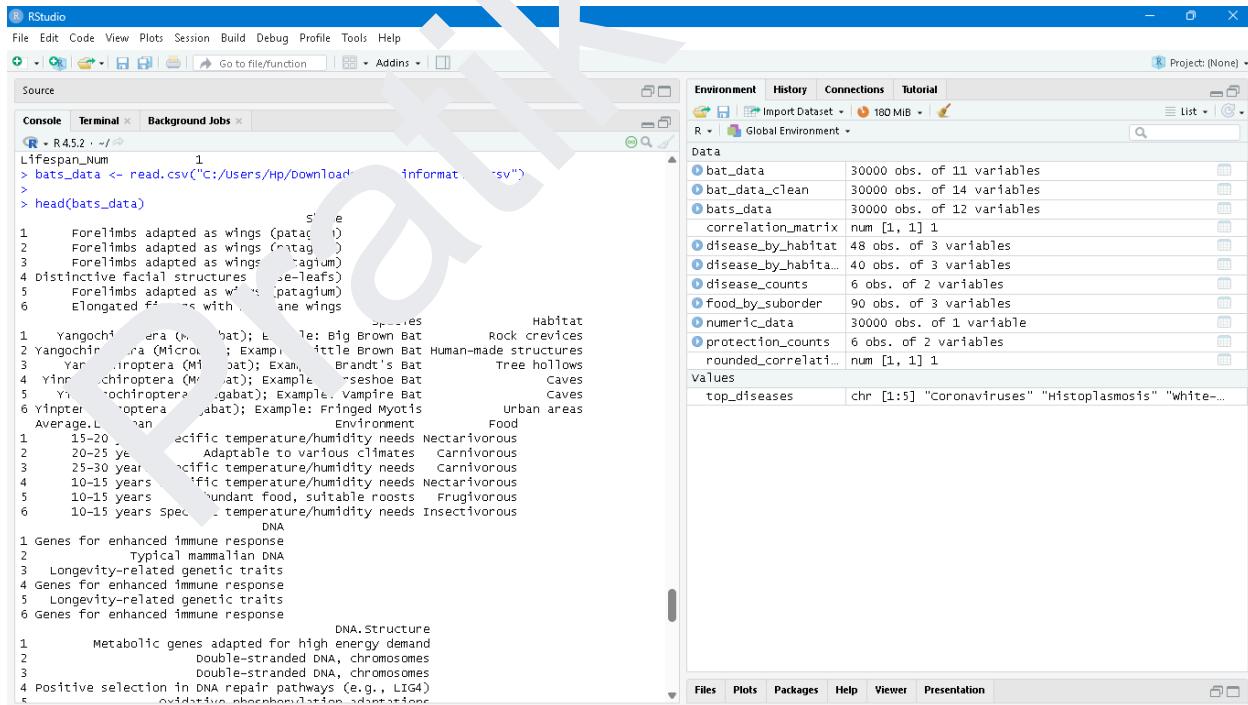
PRAC 10-12



AIM:-

12 Generating correlation matrices using cor() (R).

OUTPUT:-



SETH L.U.J. & SIR M,V, COLLEGE

PRAC 10-12

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Console Terminal Background Jobs
R + R 4.5.2 · ~/
...
\$ DNA.structure : chr "Metabolic genes adapted for high energy demand"
"double-stranded DNA, chromosomes" "double-stranded DNA, chromosomes" "Positive selection in DNA repair pathways (e.g., LIG4)" ...
\$ Protection.from.External.Threats: chr "Hibernation/torpor" "Large colony social structure" "Large colony social structure" "Large colony social structure" ...
\$ Protection.from.Diseases : chr "Genetic variations in pathogen receptors" "High interferon-alpha levels" "Unique immune system, pathogen tolerance" "Efficient DNA repair mechanisms" ...
\$ Susceptible.Diseases : chr "Coronaviruses" "white-Nose syndrome" "white-Nose Syndrome" "white-Nose Syndrome" ...
>
> bats_data\$Lifespan_Num <- as.numeric(sapply(strsplit(bats_data\$Average.Lifespan, "-"),
"[", 1))
>
> numeric_data <- bats_data[sapply(bats_data, is.numeric)]
>
> head(numeric_data)
Lifespan_Num
1 15
2 20
3 25
4 10
5 10
6 10
>
> correlation_matrix <- cor(numeric_data)
>
> print(correlation_matrix)
Lifespan_Num
Lifespan_Num 1
>
> rounded_correlation <- round(correlation_matrix, 2)
> print(rounded_correlation)
Lifespan_Num
Lifespan_Num 1
>
Project: (None)