

# Mini-Project-2.R

edrus

2025-11-28

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.2
## v ggplot2    4.0.0      v tibble    3.3.0
## v lubridate  1.9.4      v tidyr     1.3.1
## v purrr      1.1.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggforce)
```

```
## Warning: package 'ggforce' was built under R version 4.5.2
```

```
# This R script contains code generated in collaboration with
# ChatGPT (OpenAI), used as my selected AI tool for this
# assignment. The goal of the project was to reproduce one
# visualization from Lab 6: Chicago Rat Hole and create a new
# visualization based on AI-generated ideas.
```

```
#
```

```
# This file contains the AI-generated R code used to construct these
# visualizations. Per assignment instructions, the HTML files created
# from these plots are submitted separately. I also kept a record of
# all prompts used in a separate document, as required.
```

```
#
```

```
# The plot(s) below correspond to:
```

```
# 1. A reproduced Lab 6-style plot (Convex Hull / Ellipse Variation)
```

```
# 2. A new visualization based on the AI's suggested methods
```

```
#
```

```
# The goal of this project is to evaluate how well an AI tool can assist
# with designing, refining, and coding data visualizations using R, while
# ensuring good design principles and clear comparison of Chicago Rat Hole
# measurements to other species
```

```
# Load data
```

```
df <- read_csv("rsbl20250343_si_002.csv")
```

```
## Rows: 996 Columns: 11
```

```
## -- Column specification -----
```

```
## Delimiter: ","
```

```
## chr (4): Species, Specimen number (AMNH), Sex, Observer
```

```

## dbl (7): Snout-to-tail base length (mm), Forelimb length (mm), Third digit l...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
# Clean and filter required columns
df_clean <- df %>%
  drop_na(`Snout-to-tail base length (mm)`,
    `Third digit length (mm)`,
    Species)

# Separate Chicago Rat Hole from others
hull_data <- df_clean %>%
  filter(Species == "Chicago Rat Hole")

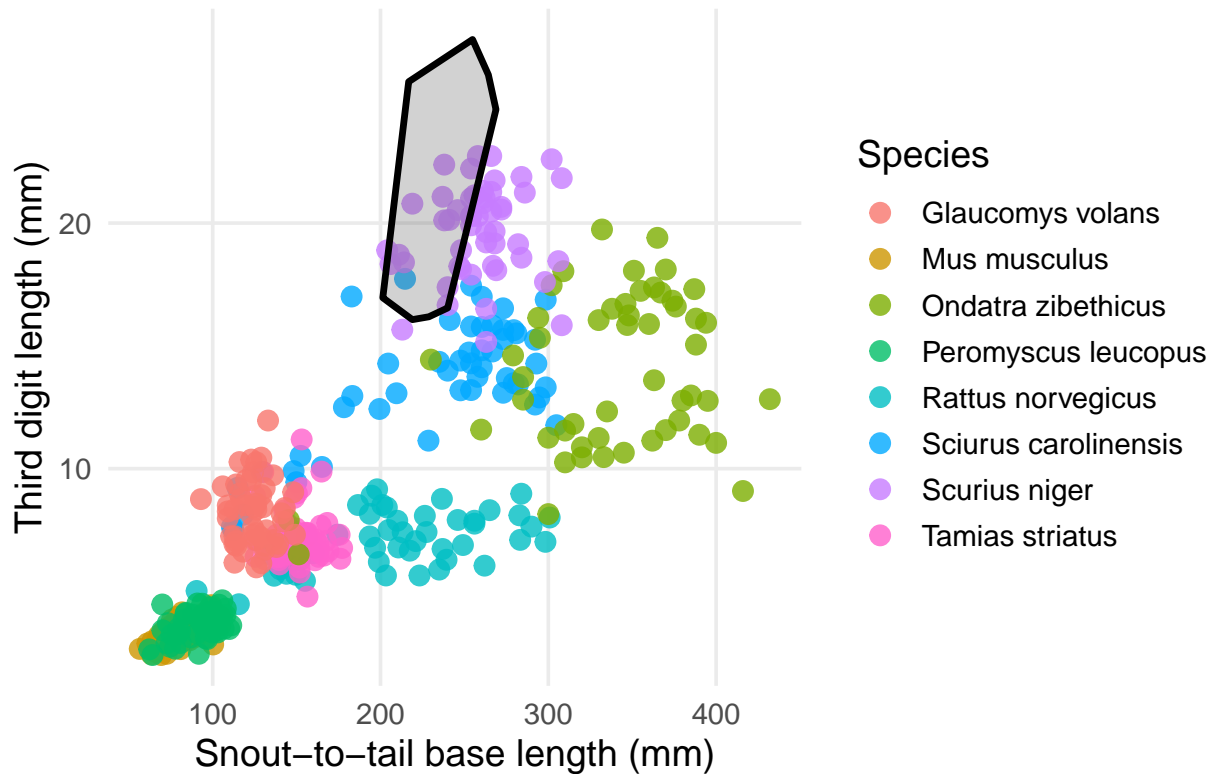
others <- df_clean %>%
  filter(Species != "Chicago Rat Hole")

# Compute convex hull indices if enough points exist
if (nrow(hull_data) >= 3) {
  hull_indices <- chull(
    hull_data$`Snout-to-tail base length (mm)`,
    hull_data$`Third digit length (mm)`
  )
  hull_poly <- hull_data[hull_indices, ]
}

# Plot
ggplot() +
  geom_point(data = others,
    aes(x = `Snout-to-tail base length (mm)`,
      y = `Third digit length (mm)`,
      color = Species),
    size = 3,
    alpha = 0.8) +
  geom_polygon(data = hull_poly,
    aes(x = `Snout-to-tail base length (mm)`,
      y = `Third digit length (mm)`),
    fill = "grey40",
    alpha = 0.3,
    color = "black",
    linewidth = 1.2) +
  labs(
    title = "Third Digit Length vs Snout-to-Tail Base Length",
    x = "Snout-to-tail base length (mm)",
    y = "Third digit length (mm)"
  ) +
  theme_minimal(base_size = 14) +
  theme(
    legend.position = "right",
    panel.grid.minor = element_blank()
  )

```

## Third Digit Length vs Snout-to-Tail Base Length



```
# Split Rat Hole vs others
rh <- df_clean %>% filter(Species == "Chicago Rat Hole")
others <- df_clean %>% filter(Species != "Chicago Rat Hole")

ggplot() +
  # Other species
  geom_point(data = others,
    aes(x = `Snout-to-tail base length (mm)`,
      y = `Third digit length (mm)`,
      color = Species),
    size = 3, alpha = 0.8) +
  # 95% confidence ellipse around Chicago Rat Hole
  stat_ellipse(data = rh,
    aes(x = `Snout-to-tail base length (mm)`,
      y = `Third digit length (mm)`),
    type = "norm",
    level = 0.95,
    fill = "grey60",
    alpha = 0.3,
    color = "black",
    linewidth = 1.2) +
  labs(
    title = "Third Digit Length vs Snout-to-Tail Base Length",
    subtitle = "95% Normal Confidence Ellipse for Chicago Rat Hole",
    x = "Snout-to-tail base length (mm)",
    y = "Third digit length (mm)"
```

```

) +
theme_minimal(base_size = 14) +
theme(
  legend.position = "right",
  panel.grid.minor = element_blank()
)

```

```

## Warning in stat_ellipse(data = rh, aes(x = `Snout-to-tail base length (mm)`), :
## Ignoring unknown parameters: `fill`

```

## Third Digit Length vs Snout-to-Tail Base Length

### 95% Normal Confidence Ellipse for Chicago Rat Hole

