

# DESCRIBING & COMPARING DISTRIBUTIONS

Mr. Merrick · September 23, 2025

## SOCS Checklist

**S — Shape:** modality (uni/bi/multi), symmetry vs. skew, clusters/gaps. *ECDF tips:* steep = high density, flat = gap/tail, early rise = right-skew, late rise = left-skew.

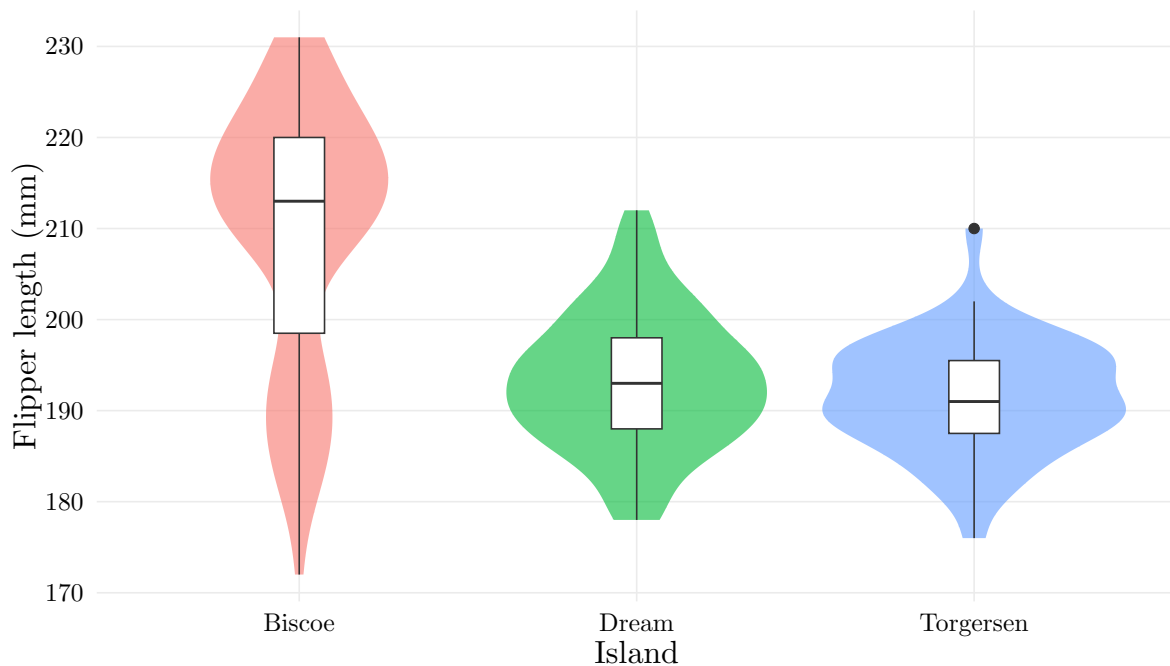
**O — Outliers:** unusual or extreme values, isolated points or small clusters. Gaps.

**C — Center:** Use median: ECDF at  $F(x) = 0.5$ , or boxplot/violin median.

**S — Spread:** Use range as a single number.

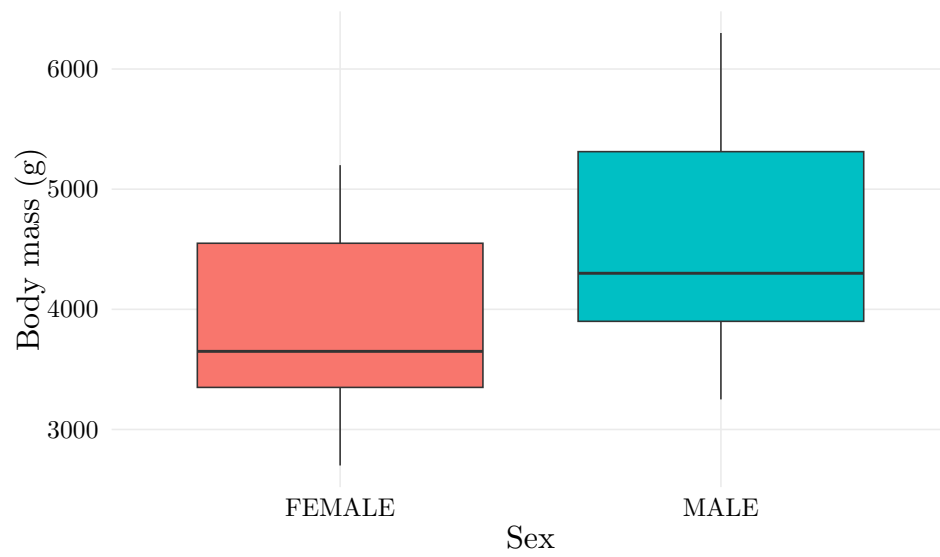
## 1. Flipper Length by Island (Penguins)

*Task.* Describe and compare the distributions of flipper length (mm) for the three islands.



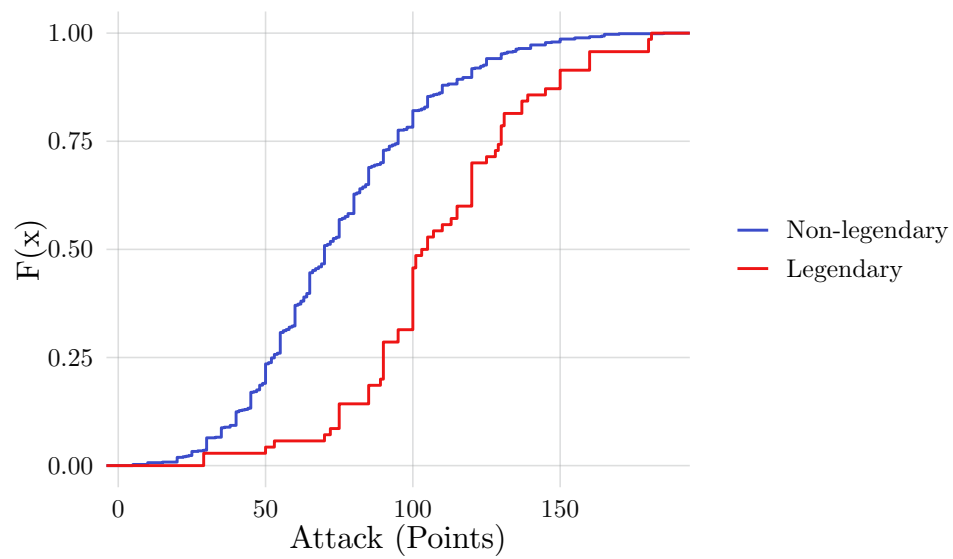
## 2. Body Mass by Sex (Penguins)

*Task.* Compare male vs. female body mass distributions.



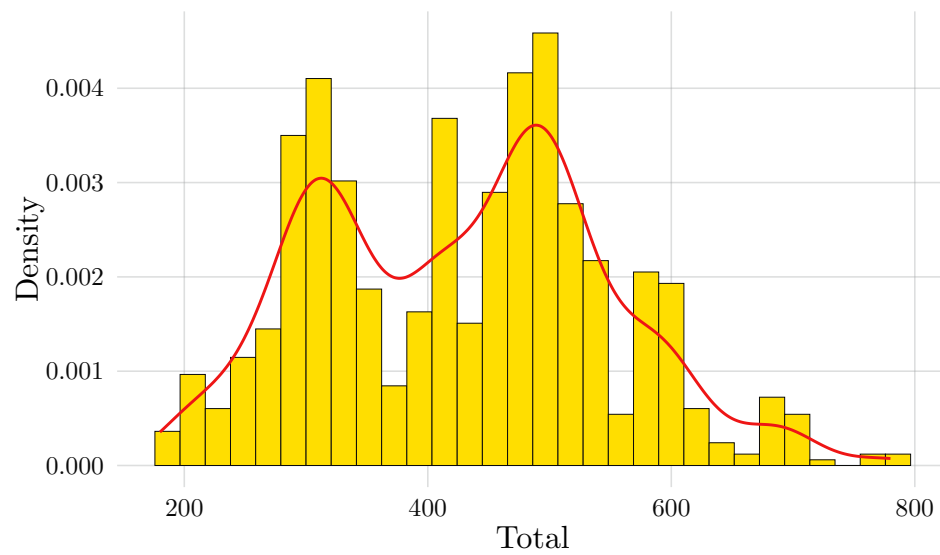
### 3. Empirical CDF of Attack (Pokémon)

*Task.* Using the ECDF, describe and compare the distributions of Attack for Legendary vs. Non-legendary Pokémon.



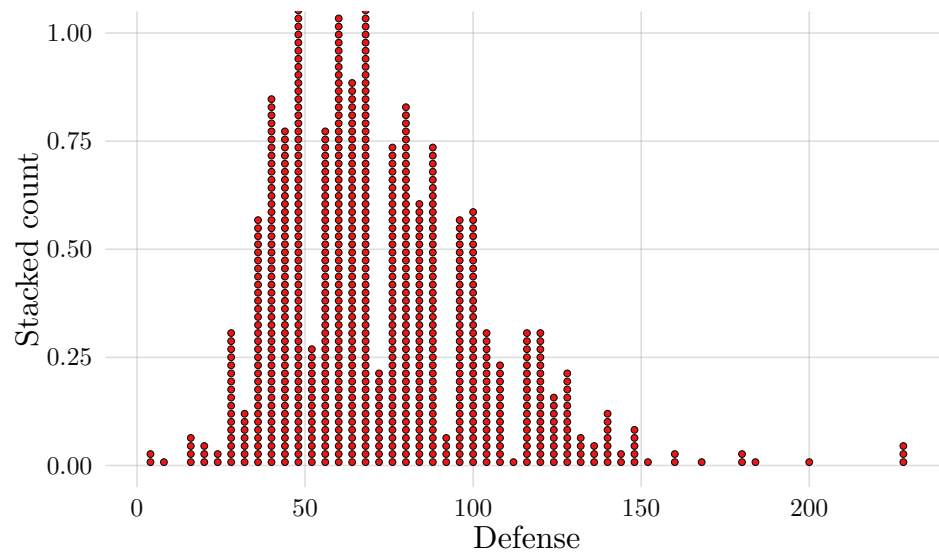
## 5. Histogram of Total (Pokémon) Score with KDE

*Task.* Describe the distribution of the overall Total Pokémon scores.



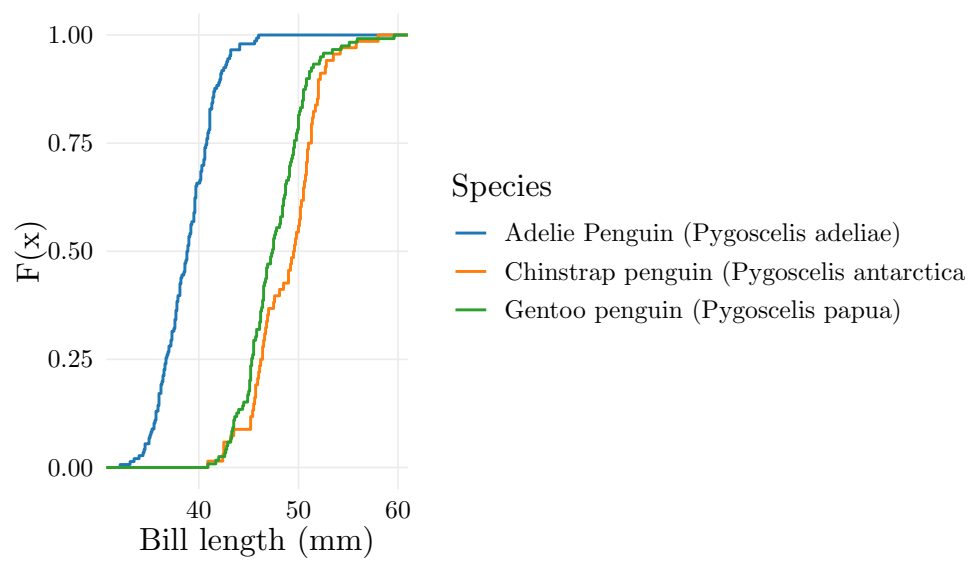
## 6. Dotplot of Defense (Pokémon)

*Task.* Describe the distribution for Pokemon defense scores.



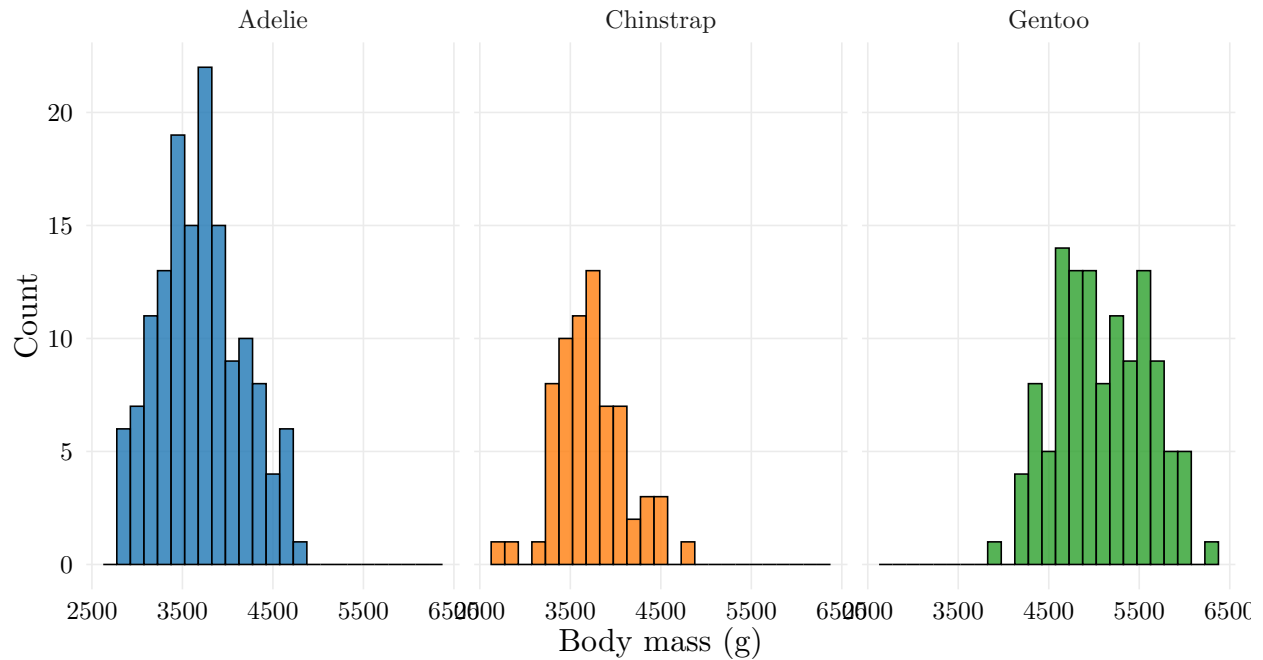
## 7. ECDF of Bill Length (Penguins) by Species

*Task.* Use the ECDF to describe and compare the distributions of bill length across species.



## 9. Body Mass Histograms by Species (Facets)

*Task.* Compare the distributions of body mass for *Adélie*, *Chinstrap*, and *Gentoo* penguins.



## Data Sources

- **Pokémon with stats** (Kaggle): <https://www.kaggle.com/datasets/abcsds/pokemon>.
- **Palmer Archipelago (Antarctica) penguin data** (Kaggle mirror of the palmerpenguins dataset): <https://www.kaggle.com/datasets/parulpandey/palmer-archipelago-antarctica-penguin-data>.