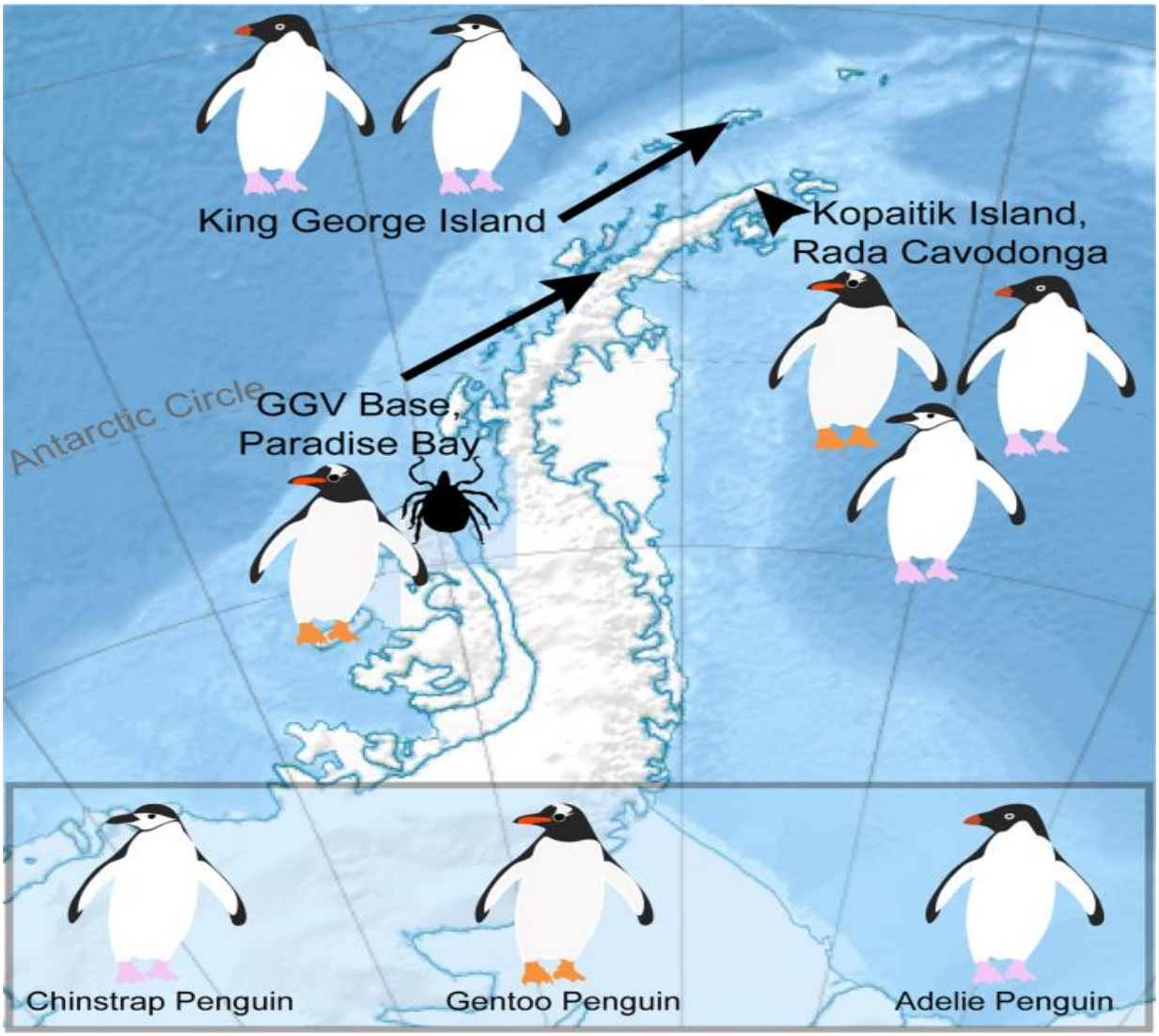
R Programming Capstone Project

The Vice-Chancellor of Thomas Adewumi University, Prof. Francisca Oladipo, a known enthusiast of penguins, has provided you with a dataset for analysis. The data, collected by Dr. Samuel Farohunbi in collaboration with the Palmer Station, Antarctica LTER, a member of the Long Term Ecological Research Network, is to be analyzed using R programming.

Penguins Pictures



Source: Wikipedia, developed by user Kikos

Load and Display Data

Read the file penguins.csv and store the result in a dataframe called penguin

Data Cleaning

Examine missing values (NA) in this data. How many are they? (Hint: You may use sum(is.na(dataframe_name)) or inspectdf::inspect_na(dataframe_name)

Derived Variable

Assuming body_mass_g and flipper_length_mm are the weight and height of Penguins. Use mutate() function in the dplyr package to calculate the BMI for each Penguins. Please save your resulting dataframe as penguins_bmi.

Analyze Data

- Create descriptive statistics for all variables in penguins_bmi dataframe
- Find and display the number of each Penguin specie in the penguins_bmi dataframe (hint: you can use count() function in dplyr to do this)
- Create and display a dataframe that shows the number of penguins of each species on each island (hint: you can use count() function to do this or a combination of group_by() and summarize(n = n()).

Data Visualization

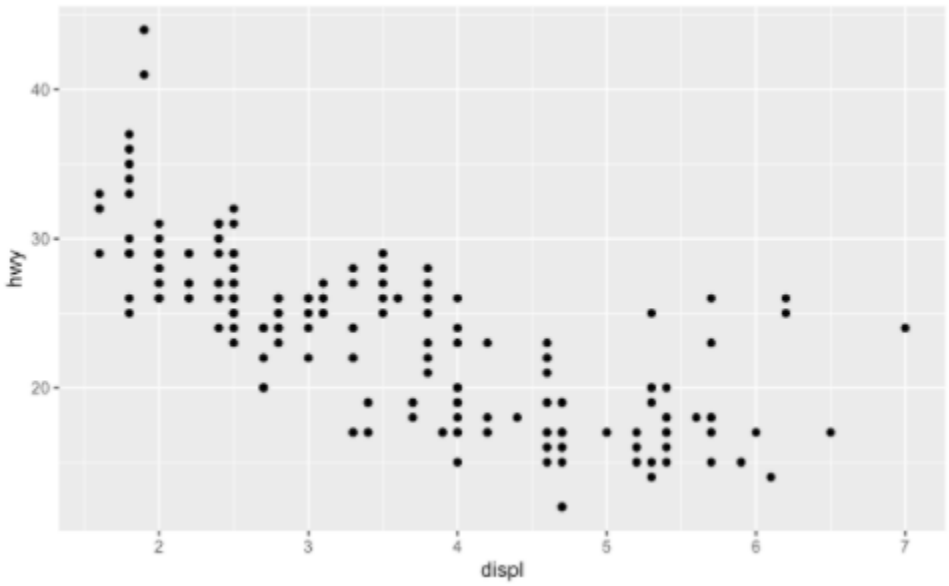
You are going to use ggplot2 package to plot some data in this section. For example, let's say I have a dataframe called mpg (the mpg dataframe is available when you load tidyverse or ggplot2 package). If I want to plot a scatter diagram where displ variable will be in the x-axis and hwy variable will be in the y-axis I will use:

```
ggplot(mpg, aes(x = displ, y = hwy)) +
  geom_point()
```

or with pipe:

```
mpg %>% ggplot(aes(x = displ, y = hwy)) +
  geom_point()
```

The result of this is shown below:



You can read more on data visualization with ggplot2 via https://r4ds.had.co.nz/data-visualisation.html and https://r4ds.had.co.nz/graphics-for-communication.html

Your Task:

What is the relationship between body mass and flipper length? You can answer this question by using a scatter plot. Please use penguins_bmi dataframe for this task.

Presentation

Create a PowerPoint presentation for your findings in this analysis.

To create an effective PowerPoint presentation for your findings from the penguin data analysis, please follow these steps:

Slide 1: Title Slide

- **Title**: Penguin Data Analysis
- **Subtitle**: An Analysis Using R Programming
- **Presenter Name**: [Your Name]
- **Date**: [Date of Presentation]

Slide 2: Introduction

- **Title**: Background
- **Content**:

- Overview of the data collection by Dr. Samuel Farohunbi and the Palmer Station, Antarctica LTER.
- Objective of the analysis: To understand key patterns and insights from the penguin dataset.

Slide 3: Data Overview

- **Title**: Dataset Description
- **Content**:
  - Brief description of the dataset (e.g., number of observations, key variables).
  - Mention the metadata and data dictionary that accompanied the dataset.

Slide 4: Methodology

- **Title**: Analysis Approach
- **Content**:
  - Outline the steps taken in R programming to analyze the data.
  - Mention any specific libraries or packages used (e.g., ggplot2, dplyr).

Slide 5: Data Visualization

- **Title**: Exploratory Data Analysis (EDA)
- **Content**:
  - Present key visualizations (e.g., distribution of penguin species, correlation between variables).
  - Provide a brief interpretation of each visualization.

Slide 6: Statistical Analysis

- **Title**: Key Findings
- **Content**:
  - Highlight significant statistical findings from the analysis.
  - Include any interesting patterns or anomalies discovered in the data.

Slide 7: Conclusion

- **Title**: Summary of Insights
- **Content**:
  - Recap the key insights gained from the analysis.
  - Discuss the potential implications of these findings.

Slide 8: Recommendations

- **Title**: Recommendations
- **Content**:

    - Offer suggestions based on the analysis, such as potential areas for further research or conservation strategies.

Slide 9: Acknowledgments

- **Title**: Acknowledgments
- **Content**:
    - Thank Prof. Francisca Oladipo, Dr. Samuel Farohunbi, and the AI General Assembly trainers
    - Mention any other individuals that supported the analysis.