



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

WST 212 Practical 6 -Data Visualization

Instructions:

- Complete the questions which follow and save all your code in a single R script, named Practical6.R •

Submission 1: Code

- Multiple code submissions are allowed and your autograded results will be available shortly after each submission.
- Ensure all variables are named correctly, as incorrectly named variables will not be awarded any marks. (Remember variable names are case sensitive.)
- Ensure your code does not consist of any syntax errors. If your code produces errors when run, the autograder will not be able to mark it.
- Any code commented out (code is commented out when # is typed in front of it) will be considered rough work and will not be marked.
- Once you have completed your submission, ensure the file is submitted on Gradescope, with the correct file name. The autograder will only be able to grade your submission if you use the correct filename.

Guidelines:

- The datasets required to complete the practical have been provided on ClickUP.
- A template that can be used for this assignment has been provided on ClickUP.
- Remember to assign your code to the variables indicated in this document.

In this practical data visualization will be done using the ggplot2 package. All plots should be made using the ggplot2 package.

Question 1

Consider the dataset Penguin, this dataset gives the depths penguins dive to and the recorded heart rate of the penguin when doing this dive (HINT: use a scatterplot)

- Use this dataset to create a plot of the duration of the dive (on the x-axis) against the heart rate of the penguin (on the y-axis). Save your answer as q1a.
- Repeat question 1.a but add the following: Colour the observations by unique bird ID such that each bird's observations are a unique colour, change the shape of the observations to a * and the size to 5. Save your answer as q1b.
- Repeat question 1.b but add the following: Title the plot as 'Heart rate vs Depth of penguin dives', the y-axis as 'Heart Rate' and the x-axis as 'Depth in meters'. Save your answer as q1c.

Question 2

Consider the dataset Penguin, this dataset gives the depths penguins dive to and the recorded heart rate of the penguin when doing this dive. (HINT: use a boxplot)

- Use this dataset to create a plot which can be used to investigate and compare the spread and outliers of the variable Dive heart rate for each bird ID. Plot these graphs on the same axes. Save your answer as q2a.
- Repeat question 2.a but add the following: Fill the graph with the same colour (Blue), outline all the plots with the same colour (red) and set the opacity to 0.2. Save your answer as q2b.
- Repeat question 2.b but add the following: Set the title of the plot to "Spread of Dive heart rates by bird species". Save your answer as q2c.

Question 3

Consider the dataset Influenza, this dataset has the number of recorded cases of Influenza and Appendicitis for the years 1970 to 2005. (HINT: use a line plot)

- Use this dataset to create a plot, with the year on the x-axis and the number of cases on the y-axis, for both Influenza and Appendicitis on the same plot. Colour the graph such that the Influenza is red, and the Appendicitis is blue. Save your answer as q3a.
- Repeat question 3.a but add the following: Fill the area under the graph with the same colour as the graph. Set the opacity of the area under Appendicitis to 0.6. Fill the area under Appendicitis before filling the area under Influenza. Save your answer as q3b.
- Repeat question 3.b but add the following: Title the plot 'Number of Influenza and Appendicitis case from 1970 to 2005', the y-axis as 'Number of cases' and the x-axis as 'Year'. Save your answer as q3c.