



Visualization of Complex Data

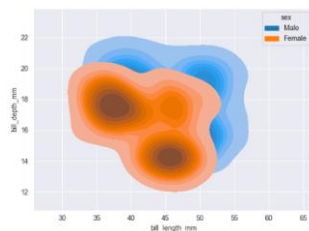
DATS 6401

Homework # 3

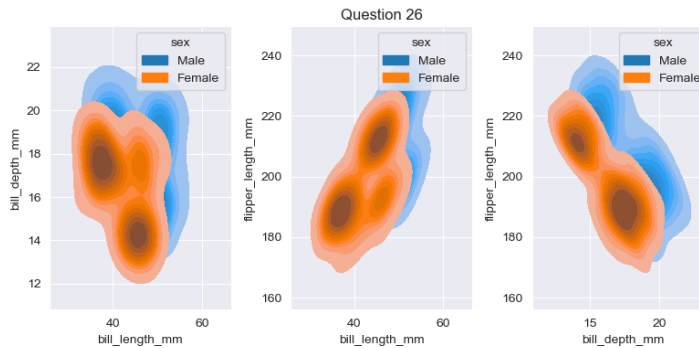
In this assignment you will practice data visualization using seaborn package. The dataset for this LAB is “penguins”. All figures in the LAB must have an appropriate title [question number], x-label, y-label, and legend [if applicable]. All calculated numbers must be displayed with 2-digit precisions. Add the question number as the title of each graph for this assignment. **Question 1 weighs 2pts and the rest of questions weigh 3.5 each.**

1. Load the ‘penguins’ dataset from the seaborn package. Display the last five observations. Display the dataset statistics. Hint: describe ()
2. Dataset cleaning: Write a python program that checks if the dataset is cleaned [with both isna() & isnull()]. If not, remove the missing observations from the data set. Display the portion of the code that performs the task here. Display the results that confirm the dataset is clean.
3. Using the seaborn package graph the histogram plot “flipper_length_mm” with the kernel density estimate on top of the histogram plot. Use the ‘darkgrid’ style as seaborn the theme. Write down your observation about the graph.
4. Change the bin width in the previous question to three and replot the graph. Write down your observation about the graph. Hint: binwidth
5. Change the bin numbers to thirty in the previous question and replot the graph. Write down your observation about the graph on the console. Hint: bins
6. Using the seaborn “displot”, graph the histogram of “flipper_length_mm” per the species. Hint: You need to use the ‘hue’. Write down your observation about the graph.
7. Re-graph the plot in the previous question with element=‘step’.
8. Using the seaborn package graph the ‘stacked’ histogram plot of ‘flipper_length_mm’ with respect to species. Hint: multiple = ‘stack’. Write down your observation about the graph.
9. Using the seaborn package and ‘displot’, graph the histogram plot of ‘flipper_length_mm’ with respect to ‘sex’ and use the option “dodge”. Write down your observation about the graph. Hint: multiple = ‘dodge’.

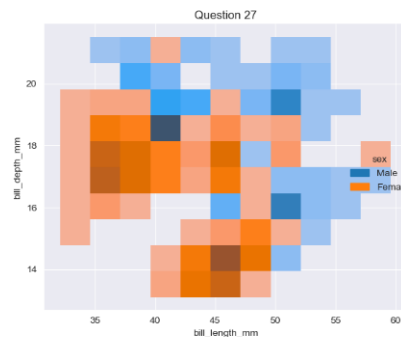
10. Using the seaborn package and 'displot', graph the histogram plot of 'flipper_length_mm' in two separate figures (not shared axis) but in one single graph (one row two columns). What is the most frequent range of flipper length in mm for male and female penguins?
11. Using the seaborn package compare the distribution of 'flipper_length_mm' with respect to species in one graph (shared axis) in a normalized fashion. Which species has a larger flipper length and what is the approximate range? Hint: Use stat = 'density'
12. Using the seaborn package compare the distribution of 'flipper_length_mm' with respect to sex in one graph (shared axis) in a normalized fashion. Which sex has the larger flipper length and what is the approximate flipper length? Hint: Use stat = 'density' and KDE to be true.
13. Using the seaborn package compare the distribution of 'flipper_length_mm' with respect to species in one graph (shared axis) in a normalized fashion that the bars height sums to 1. Which flipper length and species is more probable? Hint: Use stat = 'probability' and KDE to be true.
14. Using the seaborn package, estimate the underlying density function of flipper length with respect to 'species' and the kernel density estimation. Plot the result. Hint: hue = 'species', kind = 'kde'
15. Using the seaborn package estimate the underlying density function of flipper length with respect to 'sex' and the kernel density estimation. Plot the result. Hint: hue = 'sex', kind = 'kde'
16. Repeat question 14 with argument multiple = 'stack'
17. Repeat question 15 with argument multiple = 'stack'
18. Repeat question 14 with argument fill = True. Write down your observations about the graph.
19. Repeat question 15 with argument fill = True. Write down your observations about the graph.
20. Plot the scatter plot and the regression line in one graph for the x-axis is 'bill_length_mm' and y-axis is 'bill_depth_mm'. How do the 'bill_length_mm' and 'bill_depth_mm' correlated?
21. Using the count plot, display the bar plot of the number penguins in different islands using the hue = species. Write down your observations about the graph?
22. Using the count plot, display the bar plot of the number of male and female penguins [in the dataset] using the hue = species. Write down your observations about the graph?
23. Plot the bivariate distribution between 'bill_length_mm' versus 'bill_depth_mm' for male and female. Write your observations about correlation coefficient. Final plot:



24. Plot the bivariate distribution between 'bill_length_mm' versus 'flipper_length_mm' for male and female. Write your observations about correlation coefficient. Final plot like question 23.
25. Plot the bivariate distribution between 'flipper_length_mm' versus 'bill_depth_mm' for male and female. Write your observations about correlation coefficient. Final plot like question 23.
26. Using subplot, plot the last three questions in one graph with one row and three columns. Figure size = (8,4). Write down your observations about the plot in the last three questions.



27. Graph the bivariate distributions between "bill_length_mm" versus "bill_depth_mm" for male and female. Final plot:



28. Graph the bivariate distributions between 'bill_length_mm' versus 'flipper_length_mm' for male and female. The final plot should be like question 27.
29. Graph the bivariate distributions between 'flipper_length_mm' versus 'bill_depth_mm' for male and female. The final plot should be like question 27.

Upload the solution manual **report (as a single pdf)** plus **the .py file** through Canvas by the due date.

