

CSCE 587 Homework Assign1

Due: Jan. 30, 2024

Start by downloading the data for Kaggle NFL salaries from the HW1 folder in the Class Materials folder in the Teams General Channel. (See **VM Terminal Sessions.pptx** for instructions on how to upload the data set to your VM as well as how to download files from the VM to your local machine.)

Alternatively, you can use wget to download the data from

https://cse.sc.edu/~rose/587/CSV/Kaggle_football_salaries.csv

Note: While it is possible to answer some of the questions by looking at the data set, you will lose points if you do not use R commands to answer the questions. BE SURE TO INCLUDE R CODE TO ANSWER EACH QUESTION. To make it clear what problem each R code statement solves, include a comment line.

1 Find the subset of players that earned **at least** \$10,000,000 (field avg_year). Assign those rows to the variable BigBucks.

(a) How many players are in this list?

(b) Using the hist() method, draw a histogram of the salaries of the Players in the BigBucks category.

- In order to improve the readability of the histogram, scale the data so that the salaries are in millions of dollars. For example, Aaron Rogers average annual salary is \$33,500,000. You would depict this as 33.5.
- Of course you need to provide an appropriate label for the x-axis so that it is clear that this is \$ 33.5 million and not thirty three dollars and 50 cents ;-)
- Set the number of bins in the histogram to 40 using the “breaks” parameter. (see ?hist for details)
- Be careful to provide a meaning plot title.
- Save the plot to a PDF
- Open the PDF file to make sure that you succeeded.
- N.B.: You will lose major points if your plot does not have a meaningful title and/or the x axis is not clearly labelled to indicate the units (millions of dollars).

2. Analyze the entire dataset to find out how many players make the lowest salary. Start by finding the minimum salary and the find those players that make that minimum salary.

3. How many players earn more than 20 million?

4. Compare the salaries (avg_year)of the Kansas City Chiefs (Chiefs) to the salaries of the San Fransico 49ers (49ers). Create two smaller datasets called **Chiefs** and **FortyNiners** that contain the data for just these teams.

You may assume that the column labelled team indicates what team someone plays for.

- What is the total salary for each of these teams?
- What is the largest salary on each team and who makes this amount?
- What is the average salaries on both teams?
- Draw a histogram of the salaries of the two teams. Make a PDF of each of these histograms. Give them sensible titles and label the axes.

- N.B.: You will lose major points if your plots do not have meaningful titles and/or the x axis is not clearly labelled to indicate the units (millions of dollars).

Graduate Students Only

5. Which position (first column of data set) right-tackle or left-tackle on average over the entire dataset earns a higher annual salary (the salary value in the avg_year column)? Obviously, you will have to calculate the average salary each of these two positions. **What is this salary? Be sure to show your code for how you arrived at this answer.** No points for eyeballing the answer.

Turn in via Teams:

1. Your R code for all problems
2. A separate PDF for each histogram.