

Report

Tianyu Zhang

CPS: Project Management, Northeastern University

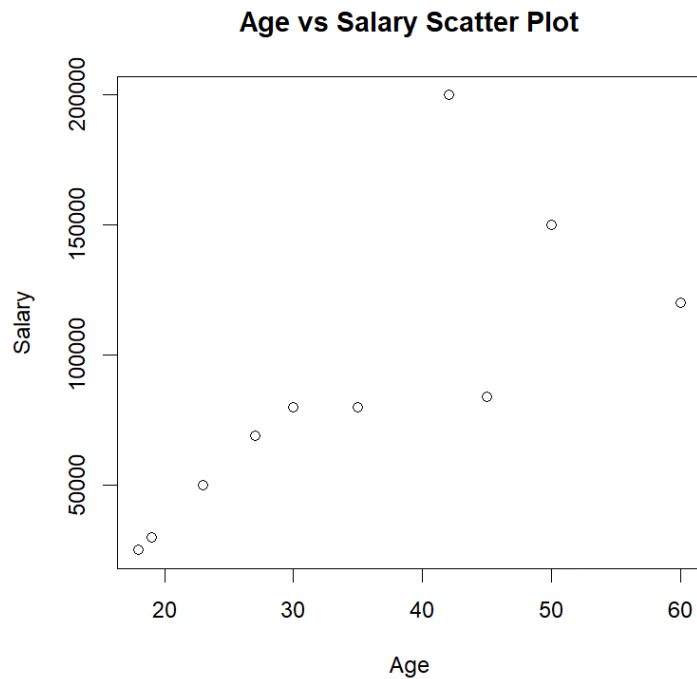
ALY 6000: Introduction to Analytics

Professor: Richard He

January 13, 2023,

Key findings based on instruction

a. A scatter plot of the age ~ salary data



b. The mean and median salary

```
mean(salary) 88800
```

```
median(salary) 80000
```

c. Display the data after steps 6 and 7

```
# 6. Delete the 6th element from the salary vector
```

```
[1] 69000 80000 50000 84000 80000 30000 25000 150000 120000
```

```
# 7. Insert 150000 as the 6th element into the salary vector
```

```
[1] 69000 80000 50000 84000 80000 150000 30000 25000 150000
```

```
[10] 120000 NA
```

d. Display the movies vector

```
movies
```

```
[1] "Lord of Ring" "Harry Pottery" "Top Gun"
```

e. Display the 7 row by 5 column matrix of 35 integers from 1 to 35

```
matrix
```

```
      [,1] [,2] [,3] [,4] [,5]
[1,]  1   8  15  22  29
[2,]  2   9  16  23  30
[3,]  3  10  17  24  31
[4,]  4  11  18  25  32
[5,]  5  12  19  26  33
[6,]  6  13  20  27  34
[7,]  7  14  21  28  35
```

f. Display the employee data frame

```
employee
  age salary
1  27 69000
2  30 80000
3  23 50000
4  45 84000
5  35 80000
6  42 150000
7  19 30000
8  18 25000
9  50 150000
10 60 120000
```

g. Display the structure and summary of the employee data frame

```
str(employee)
$ age : num 27 30 23 45 35 42 19 18 50 60
$ salary: num 69000 80000 50000 84000 80000 150000 30000 25000 150000 120000
summary(employee)
   age      salary 
Min. :18.00 Min. : 25000
1st Qu.:24.00 1st Qu.: 54750
Median :32.50 Median : 80000
Mean   :34.90 Mean   : 83800
3rd Qu.:44.25 3rd Qu.:111000
Max.   :60.00 Max.   :150000
```

h. Display the variable names only from the bank.csv data set.

```
[1] "age"      "job"      "marital"  "education" "default"
[6] "balance" "housing"  "loan"     "contact"  "day"
[11] "month"    "duration" "campaign" "pdays"   "previous"
[16] "poutcome" "y"
```

i. A summary of the information you learned about the data sets based on the instructions you followed.

From this module's learning, I acquired a lot of knowledge about how to install packages and load the library, learn how to use some functions to set plots, insert or delete some data in the vector, and create frames for the dataset. In addition, I search a lot of material on the Internet about how to import the data into the RStudio and how to display only the variables. It is interesting for me to explore the method to solve these problems.

Bibliography

- Brydon, M. (2020, April 3). *Basic analytics in R*. Lesson 2 Importing Data. Retrieved January 15, 2023, from <https://www.sfu.ca/~mjbrydon/tutorials/BAinR/data.html>
- Person. (2020, May 25). *Read_csv - could not find function*. RStudio Community. Retrieved January 15, 2023, from <https://community.rstudio.com/t/read-csv-could-not-find-function/67381>
- R programming: Create a Dataframes which contain details of 5 employees and display the details*. w3resource. (n.d.). Retrieved January 15, 2023, from <https://www.w3resource.com/r-programming-exercises/basic/r-programming-basic-exercise-25.php>
- YouTube. (2019, June 6). *R programming tutorial - LEARN THE BASICS of statistical computing*. YouTube. Retrieved January 15, 2023, from https://www.youtube.com/watch?v=_V8eKsto3Ug
- Zach. (2021, October 21). *How to fix in R: Arguments imply differing number of rows*. Statology. Retrieved January 15, 2023, from <https://www.statology.org/r-arguments-imply-differing-number-of-rows/>

Appendix

1. Print your name at the top of the script

```
print("Tianyu Zhang")
```

2. Install the vcd package

```
install.packages("vcd")
```

3. Import the vcd library

```
library(grid)
```

```
library(vcd)
```

```
library(readr)
```

4. Plot an age ~ salary scatter plot using the data below

```
age <- c(27,30,23,45,35,42,19,18,50,60)
```

```
salary <- c(69000,80000,50000,84000,80000,200000,30000,25000,150000,120000)
```

```
plot(age, salary, xlab = "Age", ylab = "Salary", main = "Age vs Salary Scatter Plot")
```

5. Find the mean and median salary

```
mean_salary <- mean(salary)
```

```
mean(salary)
```

```
median_salary <- median(salary)
```

```
median(salary)
```

6. Delete the 6th element from the salary vector

```
salary <- salary[-6]
```

```
salary
```

7. Insert 150000 as the 6th element into the salary vector

```
salary <- c(salary[1:5], 150000, salary[6:10])
```

```
salary
```

8. Create a vector <movies> with elements Lord of Ring, Harry Pottery, Top Gun

```
movies <- c("Lord of Ring", "Harry Pottery", "Top Gun")
```

```
movies
```

9. Create a 7 row and 5 column matrix of 35 integers from 1 to 35

```
matrix <- matrix(1:35, nrow = 7, ncol = 5)
```

```
matrix
```

10. Create a data frame <employee> with age and salary attributes

```
employee <- data.frame(age = c(27,30,23,45,35,42,19,18,50,60) ,  
                        salary  
                        c(69000,80000,50000,84000,80000,150000,30000,25000,150000,120000) )  
employee
```

11. Display the data frame structure and summary of the employee data frame

```
str(employee)
```

```
summary(employee)
```

12. Import the dataset bank.csv

```
bank <- read_csv("bank.csv")
```

13. Display only the variable names of the bank.csv dataset

```
colnames(bank)
```