Assignment 8

Sathish Prasad V T, sp702

2023-08-14

Instructions

*This is an **individual** assignment.*

The objective of this assignment is to give you experience composing programs in R.

Please meet the specific requirements exactly. Do not change or remove any of the instructions; just add you name and NetID in the author metadata area above and code in the code cells as indicated. Please make sure to test your programs before submitting your solutions on Canvas!

Problem 1: Employee Salary Analysis

The file `employee_data.csv` contains employee data (excluding executives) for a non-profit trade association; the first five lines of this file are shown below. Please use this data to answer the following questions.

```
Name,Salary,Department
Angela Coston,108300,Finance
Jerome Vinson,172900,Economics and Strategic Planning
Jamar Southerland,158800,Legal
Tom Banks,150500,Governmental Affairs
```

What are the lowest and highest salaries?

```
# TODO: Insert your code here.
data = read.csv('employee_data.csv')
max(data$Salary)
```

```
## [1] 185800
```

```
min(data$Salary)
```

```
## [1] 73400
```

What is the mean salary for employees in the Operations department?

```
# TODO: Insert your code here.
op_emp = data[data$Department=='Operations',]
mean(op_emp$Salary)
```

```
## [1] 102572.2
```

Who has is the highest salary for employees in the Legal department?

```
# TODO: Insert your code here.
leg_salary = data[data$Department=='Legal',]
subset(leg_salary, Salary == max(Salary))
```

```
##                  Name Salary Department
## 68 Frances Battaglia 166500      Legal
```

How many employees have a salary under 80,000?

```
# TODO: Insert your code here.
less = data[data$Salary<80000,]
nrow(less)
```

```
## [1] 14
```

The company is considering merging the “Product Development” and “Sales” departments into a single department. Assuming there are no layoffs or additional hires as part of this process, **how many employees would there be in this new combined department?**

```
# TODO: Insert your code here.
nrow(data[data$Department=='Sales',]) + nrow(data[data$Department=='Product Development',])
```

```
## [1] 9
```

Problem 2: Cutting, or binning data

In Assignment 6, Problem 1, you were given code to add a new `age.group` column to a data frame of customer information. Your task now is to repeat that manipulation in R using the base `cut` method. Your code should load data locally from `customers.csv` and output the resulting data frame to `customers_age_groups.csv`. The first ten lines of the output file should look as follows.

```
customer.id,customer.name,age,gender,age.group
1,Rick Hill,45,male,35-54
2,Don Baum,35,male,35-54
3,Thomas Anderson,47,male,35-54
4,Douglas Gaffney,42,male,35-54
5,Crystal Auten,33,female,18-34
6,John Canty,43,male,35-54
7,Pauline Mullan,45,female,35-54
8,Edward Hall,36,male,35-54
9,Ebony Blackburn,48,female,35-54
```

```
# TODO: Insert your code here.
library(readr)
data2 = read.csv('Customers.csv')
age_bin = c(18,35,55,100)
age_label = c("18-34","35-54","55+")
data2$age.group = cut(data2$age, breaks=age_bin,label=age_label,right=FALSE)
write_csv(data2,"customers_age_groups.csv",col_names = TRUE)
```