Assignment 3

okan can

2025-10-20

```
library(ggplot2)
#library(ggmosaic)
library(treemapify)
```

Dataset

The `salary` dataset contains information about the average annual salaries of data professionals across different roles and experience levels. It includes numerical data on salaries (`mean_salary`) in Rupees as well as categorical variables describing the job role (`position`) and experience level (`experience`). The experience levels are ordered from Junior → Intermediate → Senior → Executive, reflecting increasing levels of expertise and responsibility. This dataset can be used to explore which positions and experience levels contribute most to overall salary distribution in the data science field.

Source: Kaggle – Data Science Fields Salary Categorization (accessed October 2025).

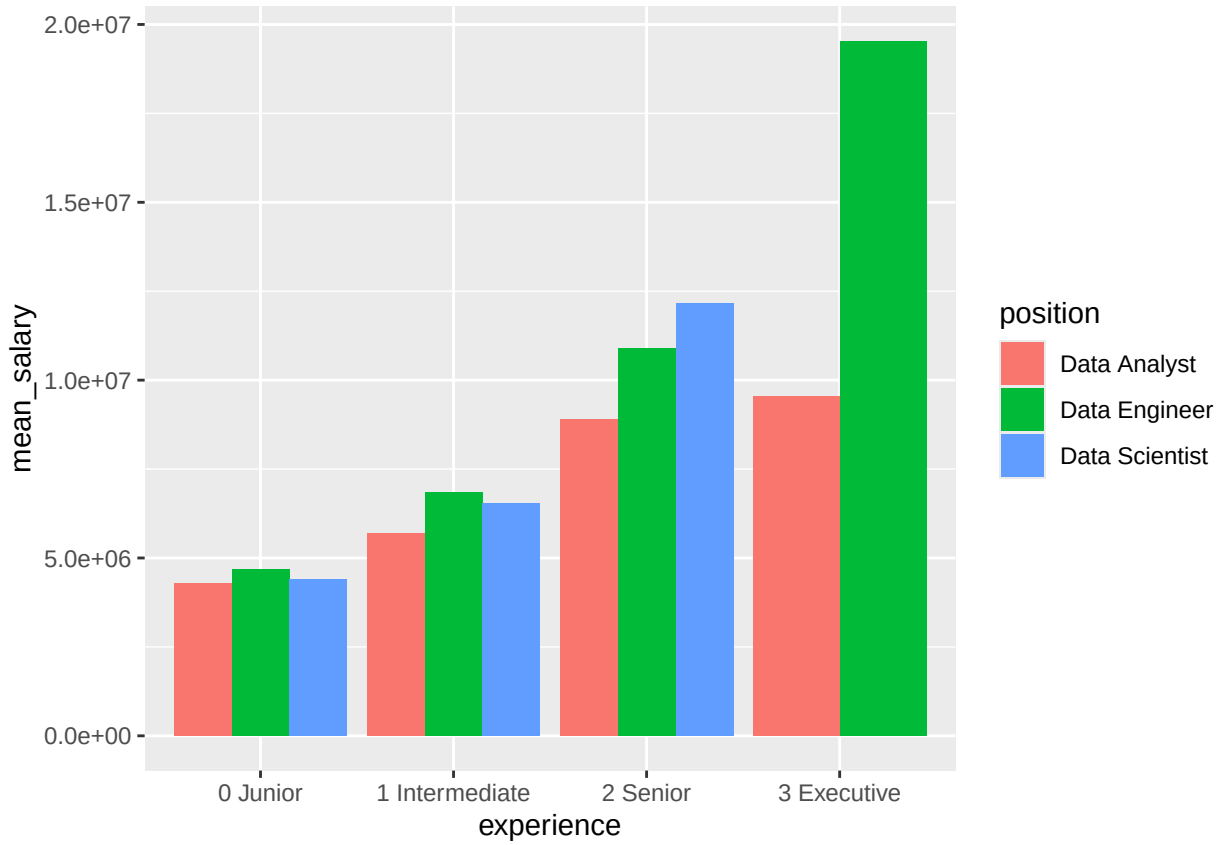
To load the dataset into your environment, just run the code below:

```
salary <- data.frame(
  position    = c("Data Analyst", "Data Analyst", "Data Analyst", "Data Analyst",
                  "Data Engineer", "Data Engineer", "Data Engineer", "Data Engineer",
                  "Data Scientist", "Data Scientist", "Data Scientist"),
  experience  = c("0 Junior", "3 Executive", "1 Intermediate", "2 Senior",
                  "0 Junior", "3 Executive", "1 Intermediate", "2 Senior",
                  "0 Junior", "1 Intermediate", "2 Senior"),
  mean_salary = c(4293623, 9548340, 5705070, 8905628,
                  4689309, 19534312, 6841836, 10903873,
                  4402653, 6527813, 12171827))
```

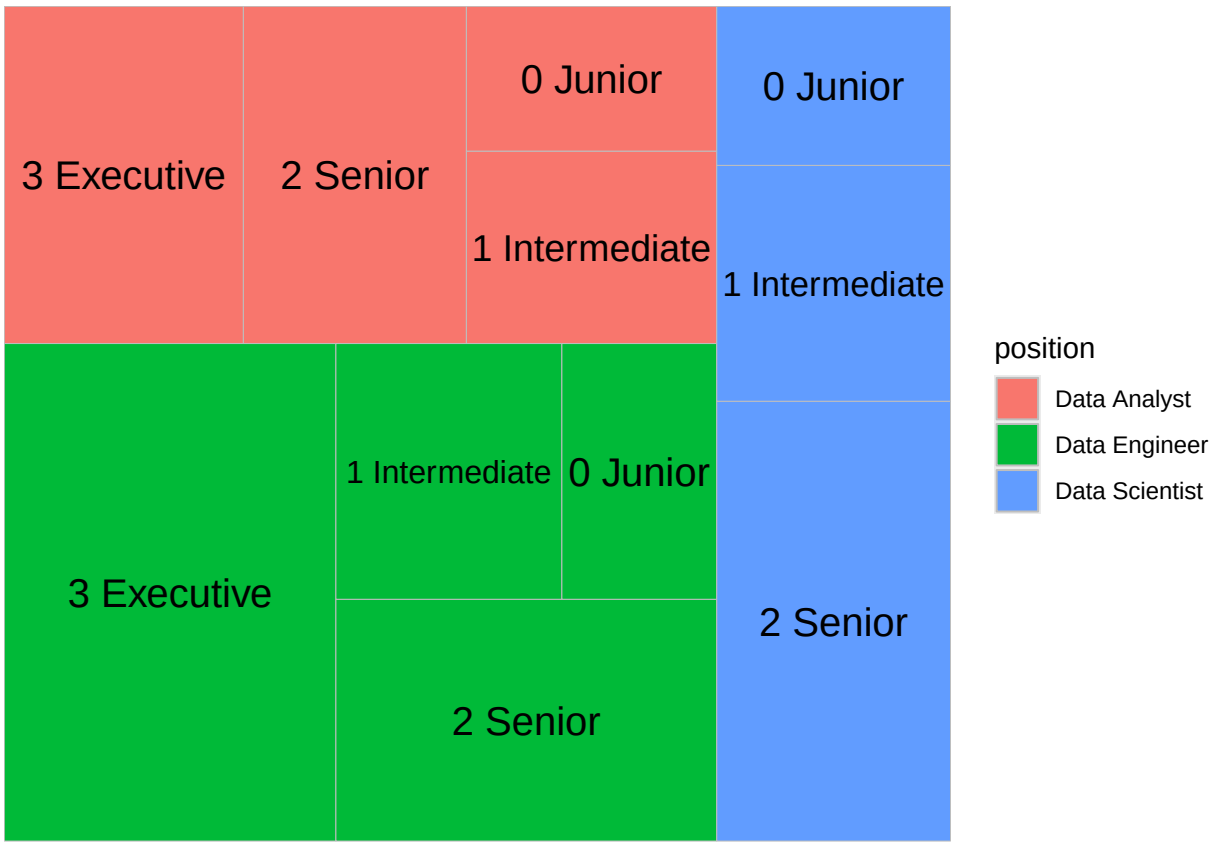
Drawing a plot for proportion

1. Draw a plot showing the proportion of total mean salaries by experience and position. The plot should clearly display the hierarchical contribution of each job title within its experience level.

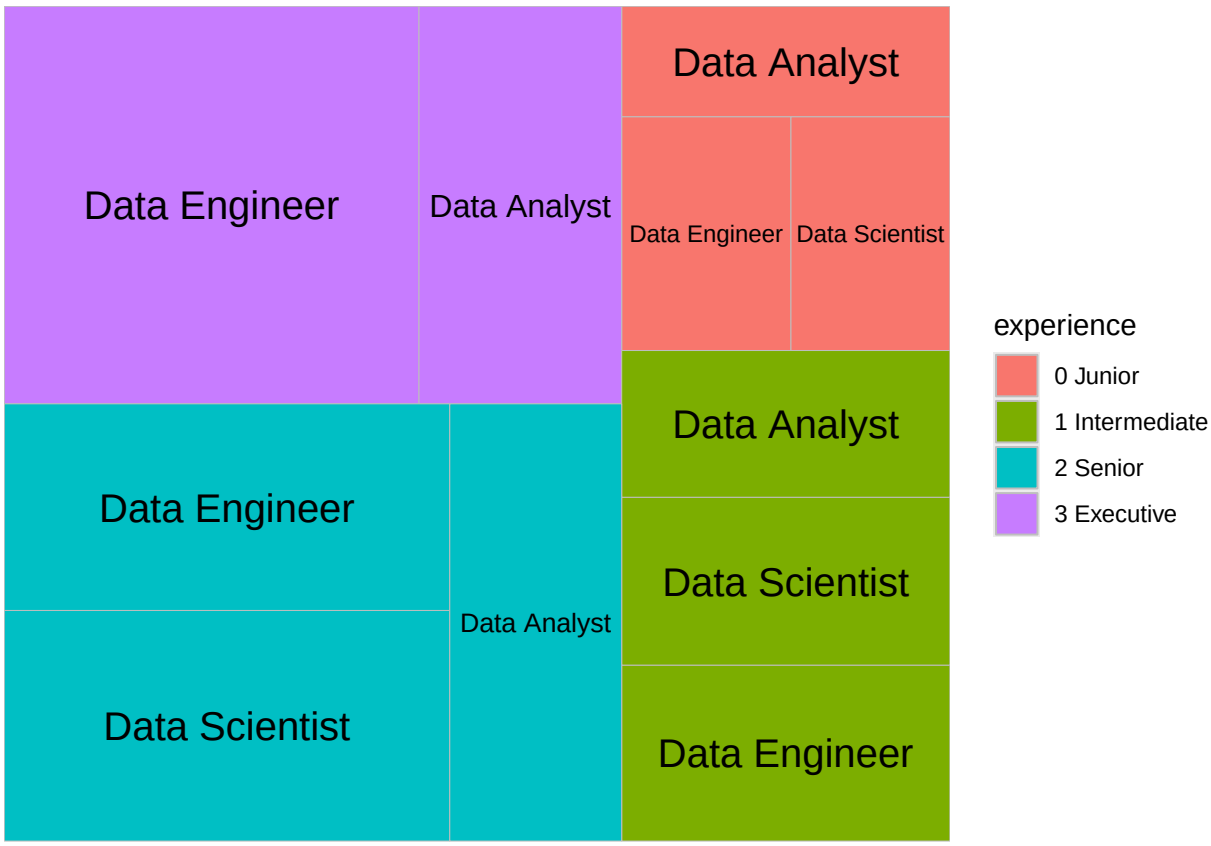
```
ggplot(salary, aes(fill = position,
                       x    = experience,
                       y    = mean_salary)) +
  geom_bar(position = "dodge",
           stat     = "identity")
```




```
ggplot(salary, aes(area     = mean_salary,
                   fill     = position,
                   label    = experience,
                   subgroup = position)) +
  geom_treemap() +
  geom_treemap_text(colour = "black",
                    place  = "centre",
                    size   = 15)
```




```
ggplot(salary, aes(area     = mean_salary,
                   fill     = experience,
                   label    = position,
                   subgroup = experience)) +
  geom_treemap() +
  geom_treemap_text(colour = "black",
                    place  = "centre",
                    size   = 15)
```



2. Interpret the plot (30 pts).

the first conspicuous part is in executive level there is no data scientist we can commit (guess) in executive level of this sector people can prefer to data analyt or data engineer (mostly people of senior class when they promotion they can be swich to data engineer) or there is no need to "data scientist"

second conspicuous is most changeble mean_salary is executive level according to experience salary is more changeble

third in this sector when people take promotion, their salary increasing exponential

ı thin most weird thing is almost senior earn as much as executive

Data engineers earn the least in intermadate level each experince earn it almost same