Assignment 3

Yağmur Nisan AKTAŞ

2025-10-20

```
# Installing necessary package(s)
```

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("Junior", "Executive", "Intermediate", "Senior",
                  "Junior", "Executive", "Intermediate", "Senior",
                  "Junior", "Intermediate", "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.

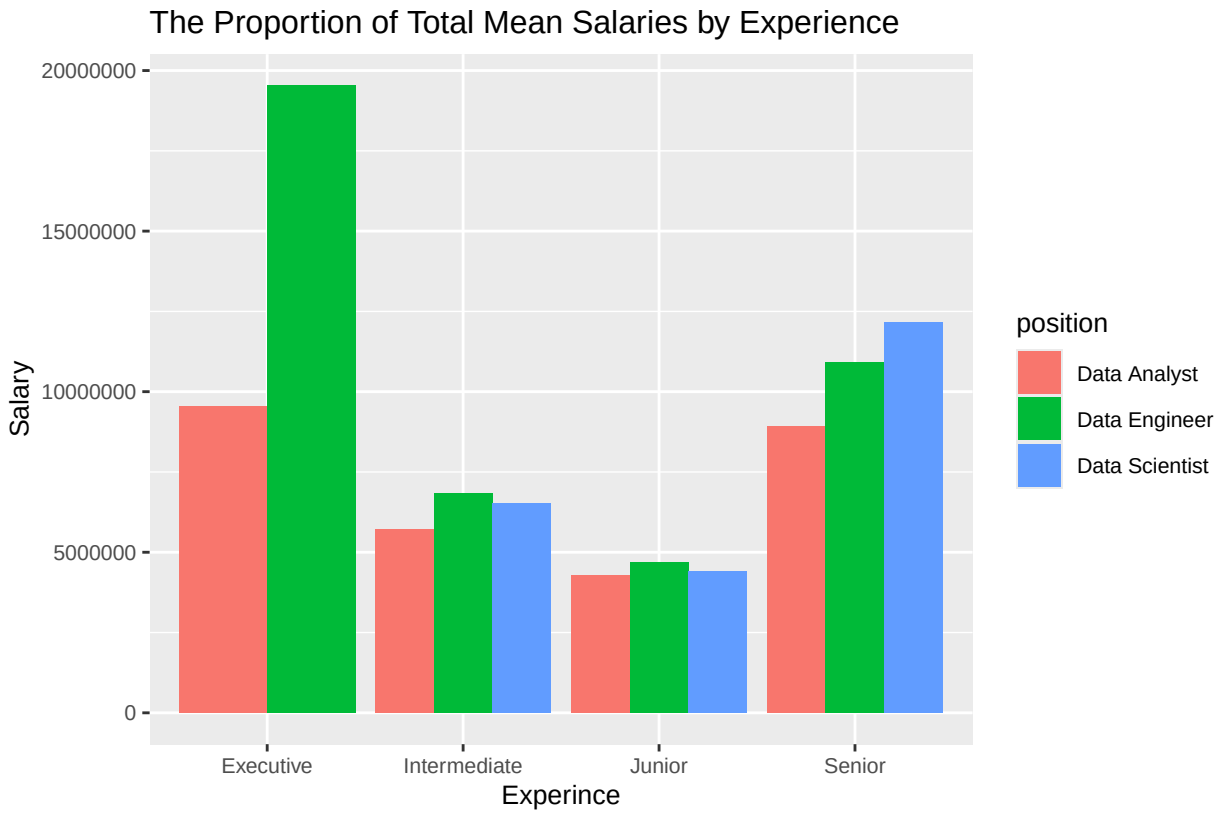
```
installed.packages("ggplot2")
```

```
##      Package LibPath Version Priority Depends Imports LinkingTo Suggests
##      Enhances License License_is_FOSS License_restricts_use OS_type Archs
##      MD5sum NeedsCompilation Built
```

```
library(ggplot2)
```

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

```
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE)) +
  labs(x       = "Experince",
       y       = "Salary",
       title   = "The Proportion of Total Mean Salaries by Experience",
       caption = "") +
  theme(text = element_text(size = 10))
```



2. Interpret the plot (30 pts).

When we look at the graph, as the title of the data engineer increases, their salary also increases and they also receive the highest salary.