Assignment 3

Ceren Yıldız

2025-10-20

```
# Installing necessary package(s)
install.packages("ggplot2")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.5'
## (as 'lib' is unspecified)
```

```
library(ggplot2)
```

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).

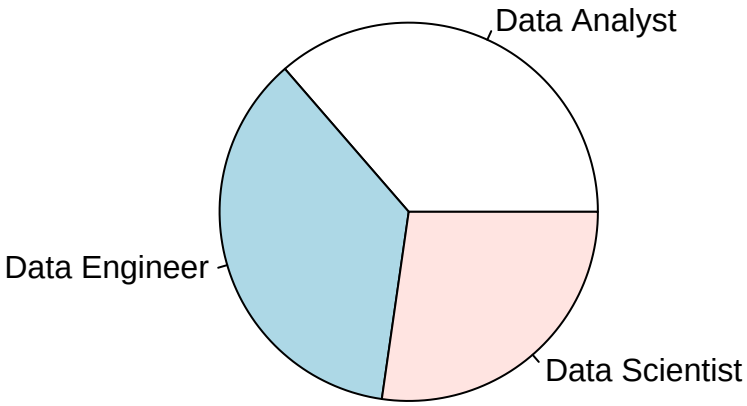
To load the dataset into your environment, just run the code below:

```
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))
```

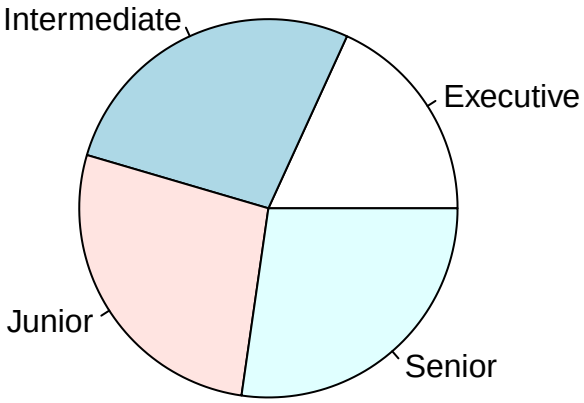
Drawing a plot for proportion

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.

```
pie(table(salary$position))
```




```
pie(table(salary$experience))
```



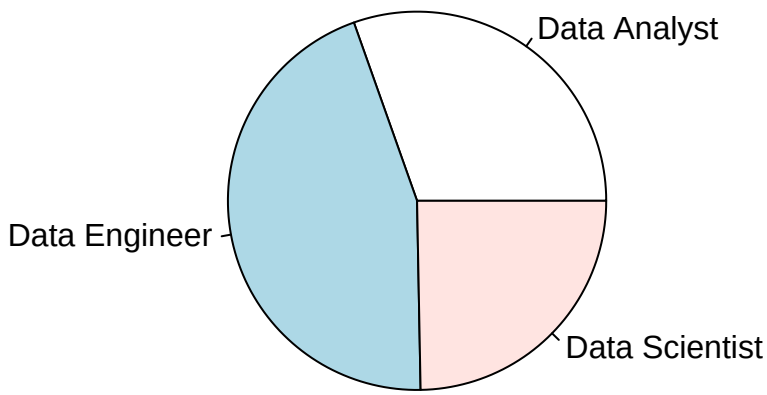

```
#We can use this code but it is not true because it is complicated.Also, Data Scientist,Data Analysist
table(salary$position)
```

```
##
##   Data Analyst  Data Engineer Data Scientist
##              4              4              3
```

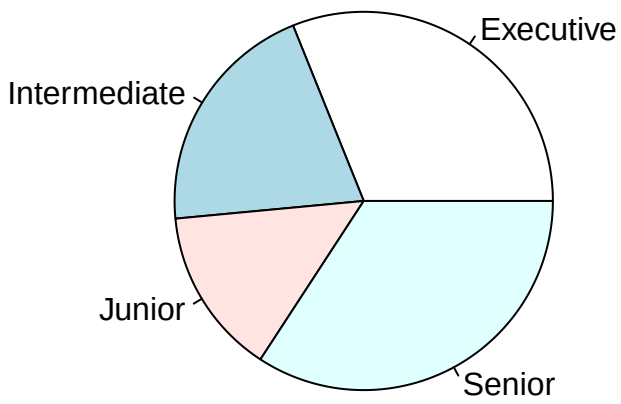
```
tapply(salary$mean_salary,salary$position,sum)
```

```
##   Data Analyst  Data Engineer Data Scientist
##       28452661       41969330       23102293
```

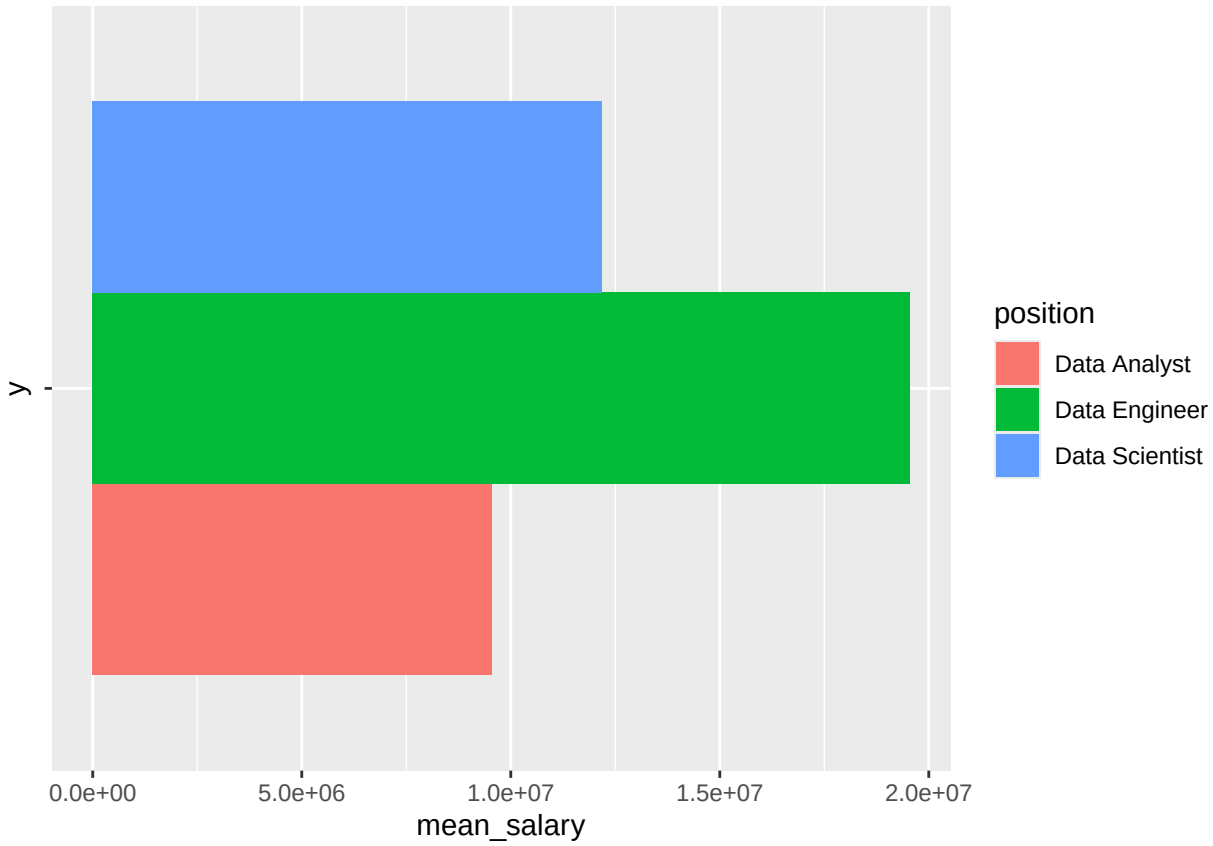
```
pie(tapply(salary$mean_salary,salary$position,sum),main="Mean Salaries of the Position")
```



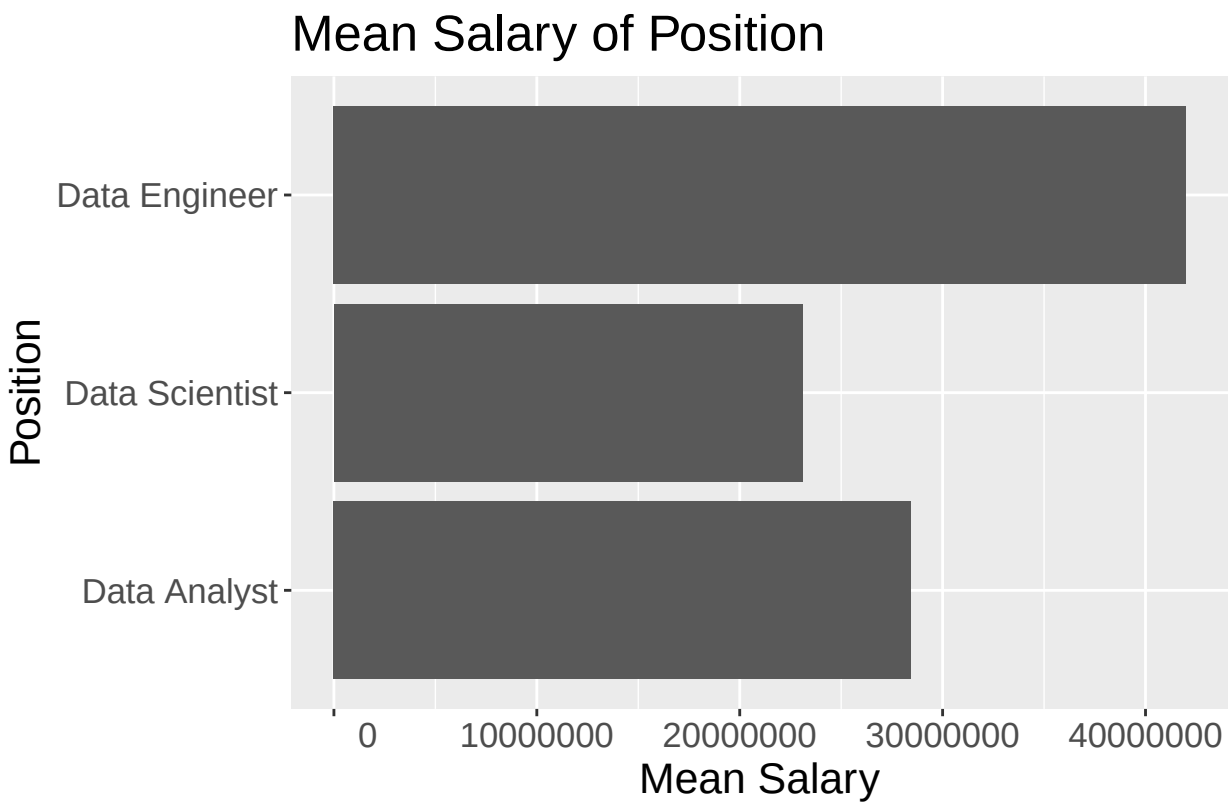

```
pie(tapply(salary$mean_salary,salary$experience,sum),main="Mean Salaries of the Experience")
```



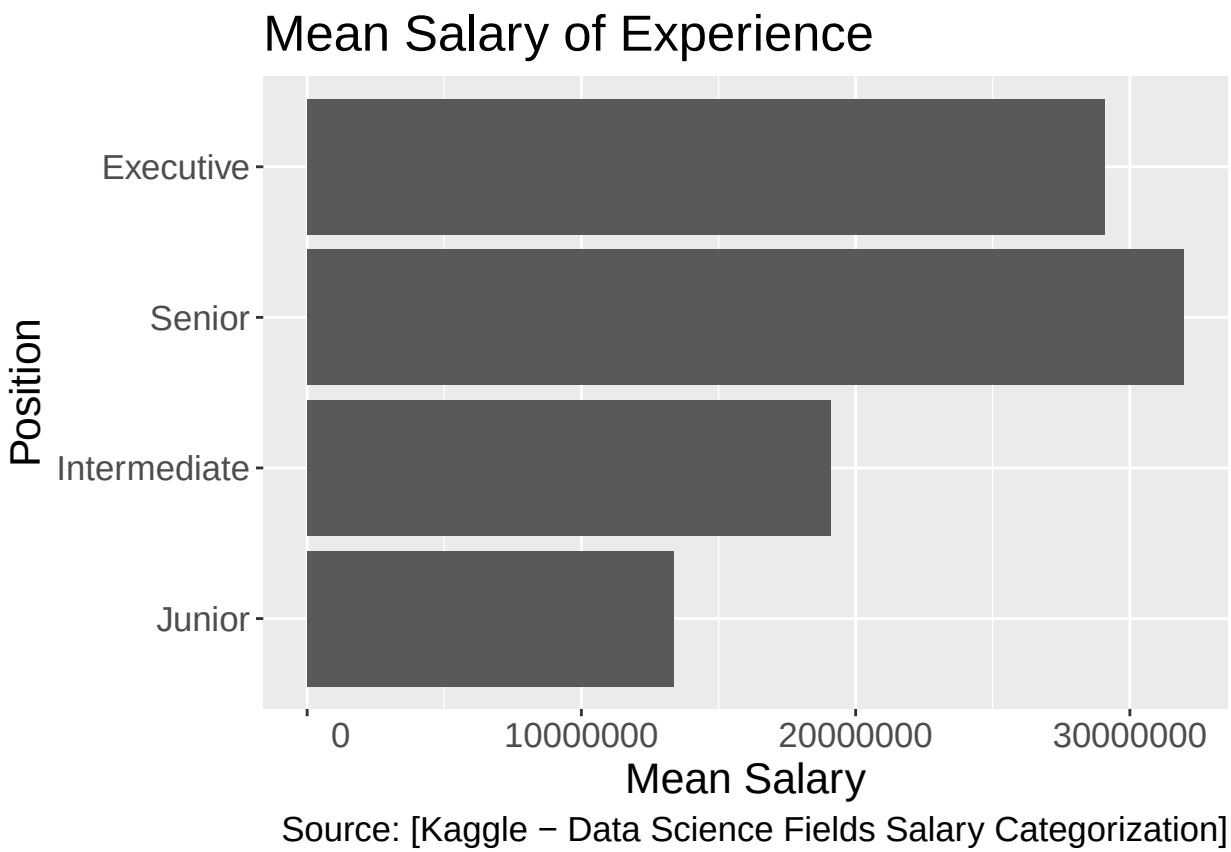

```
ggplot(salary, aes(fill = position,
                   y    = "",
                   x    = mean_salary)) +
  geom_bar(position = "dodge",
           stat     = "identity")
```




```
ggplot(salary, aes(x= reorder(position,mean_salary),y=mean_salary)) +
 geom_bar(stat="identity")+
 scale_y_continuous(labels= function(x) format(x,scientific=FALSE))+
 labs(
 x="Position",
 y="Mean Salary",
 title = "Mean Salary of Position ",
 caption ="Source: [Kaggle - Data Science Fields Salary Categorization]"
 )+
 coord_flip() +
 theme(text= element_text(size=16))
```




```
ggplot(salary, aes(x= reorder(experience,mean_salary),y=mean_salary)) +
 geom_bar(stat="identity")+
 scale_y_continuous(labels= function(x) format(x,scientific=FALSE))+
 labs(
 x="Position",
 y="Mean Salary",
 title = "Mean Salary of Experience ",
 caption ="Source: [Kaggle - Data Science Fields Salary Categorization]"
 )+
 coord_flip() +
 theme(text= element_text(size=16))
```



2. Interpret the plot (30 pts). First of all, we can compare groups mean salaries of the Position.For example,Data engineer is bigger than Data Scientist or Data Analyst.Morever,we can compare groups mean salaries of the Experience.The experience levels are ordered from Junior → Intermediate → Senior → Executive, reflecting increasing levels of expertise and responsibility. For example, mean salaries of junior experience level is smaller than others.So this graphs show that experience levels is very important.