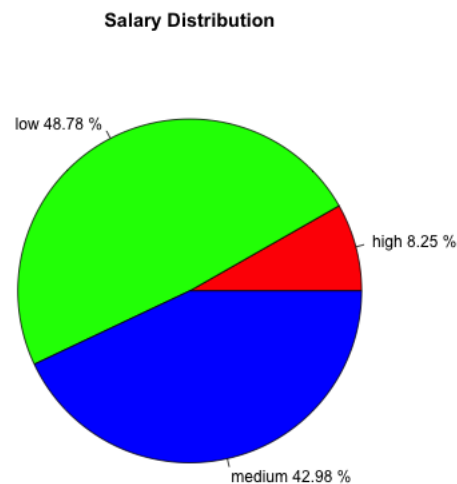


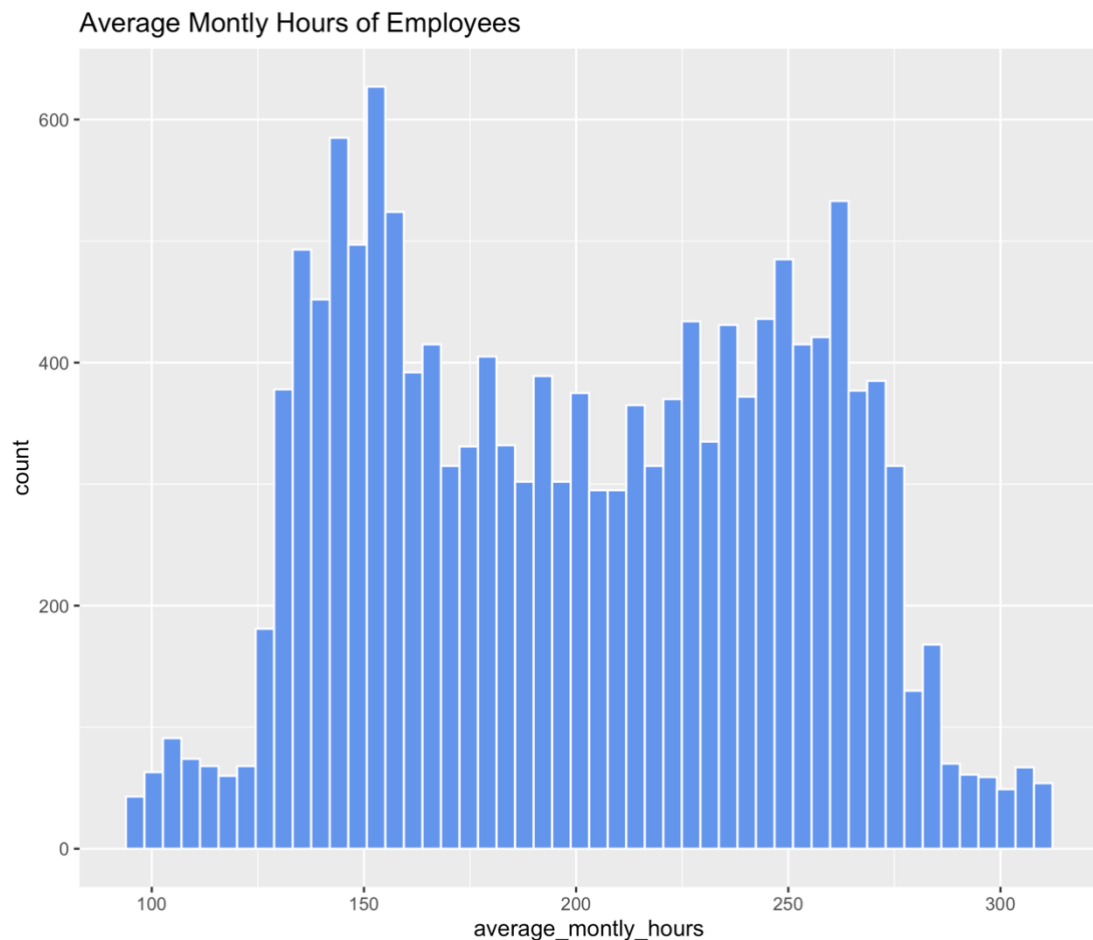
1. Univariable graph – Categorical Data
I chose “Salary” variable.



- a) **Missing Data Treatment**
The missing data has a percentage of around 4.99%, and we know nothing about the distribution of Salary among all employees. Therefore, I used deletion for missing data.
- b) **Reasons for Graph Choice**
“Salary” variable has 3 possible values: “low”, “medium”, and “high”. Therefore, it is a (ordered) categorical variable, suitable for pie charts to reveal the percentage of each category.
- c) **Strengths and Weaknesses**
Strengths: Pie charts are easy to understand, and users can compare the amount of each category according to the percentage shown on the pie chart.
Weaknesses: Pie charts only show the percentage of each category, but not the absolute numbers of categories since the total number of employees is missing on pie charts.
- d) **Audience**
This pie chart targets on those who require an overview of the salary distribution among employees, especially jobseekers and collaborators of this company.
- e) **Data Story**
According to the distribution of salary among all employees, most employees are of low and medium salary, and only a small part of employees are of high salary. The number of low and medium salary employees are similar. This chart might be evidence or supports for employees’ attritions, since most employees don’t obtain high salary. They might leave this company due to dissatisfaction of salary, causing a relatively high attrition rate, though more information is needed to accept or reject this hypothesis.

2. Univariable graph – Quantitative Data

I chose “Average Monthly Hours” variable.



a) Missing Data Treatment

The missing data has a percentage of around 4.99%, and we know nothing about the distribution of Average Monthly Hours among all employees. Therefore, I used deletion for missing data.

b) Reasons for Graph Choice

“Average Monthly Hours” is a continuous quantitative (numerical) variable. Therefore, it is suitable for histograms to reveal the frequency of certain range of average monthly hours.

c) Strengths and Weaknesses

Strengths: Histograms are easy to understand, with absolute frequency within a small range of average monthly hours. Users can directly compare the frequency of different ranges by comparing the height of bars.

Weaknesses: Histograms fail to reveal the distribution of “Average Monthly Hours” within each small range, and it requires users to try out what is the suitable range intervals to demonstrate the overall “Average Monthly Hours” distribution among all employees.

d) Audience

This histogram targets on those who require an overview of the average

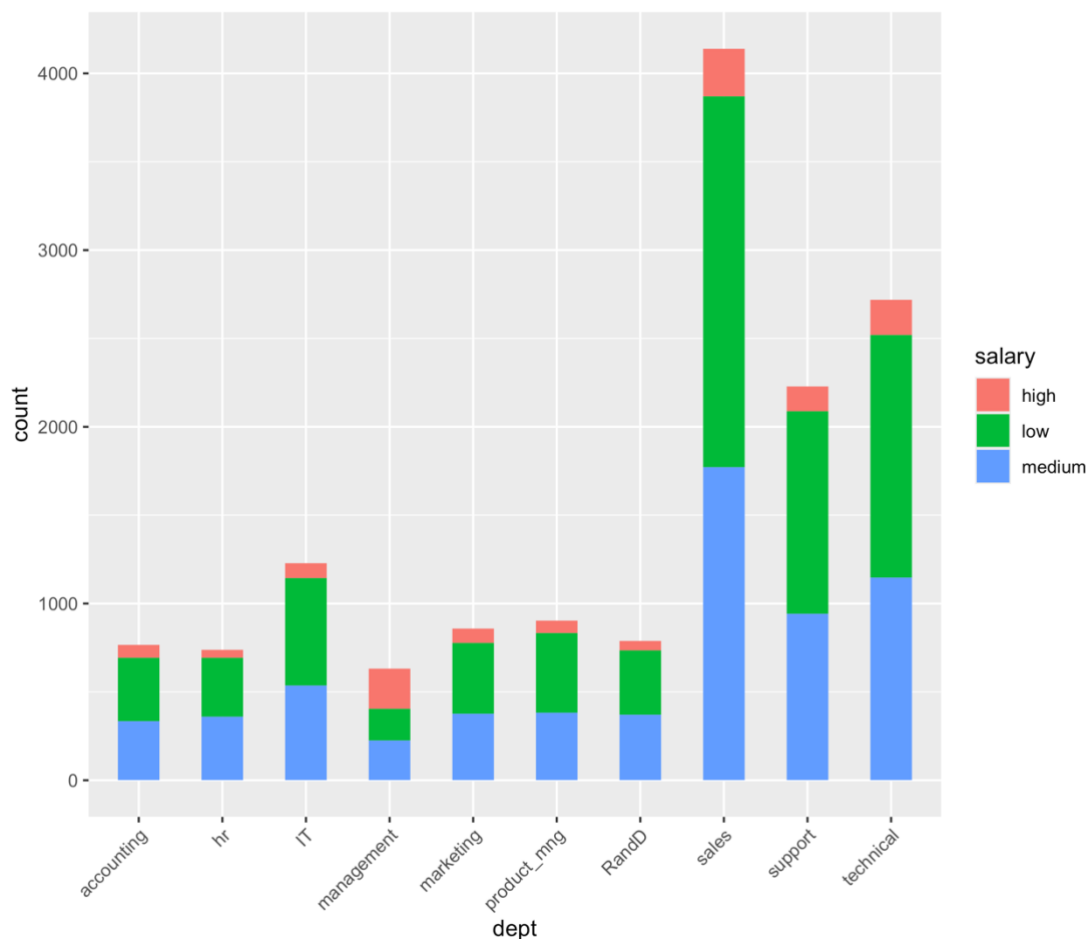
monthly time distribution among employees, especially jobseekers, managers, and collaborators of this company.

e) Data Story

According to the distribution of “Average Monthly Hours” among all employees, most employees are within the range of 150 and 250, with mean value around 200. This chart might be evidence or supports for employees’ attritions, since excessive or insufficient working hours could cause satisfaction issues. They might leave this company due to dissatisfaction of excessive working hours compared to insufficient working hours of other employees, causing a relatively high attrition rate, though more information is needed to accept or reject this hypothesis.

3. Bivariable Graph – Categorical- Categorical

I chose “Salary” and “dept” variables.



a) Missing Data Treatment

The missing data has a percentage of around 4.99%, and we know nothing about the distribution of “Salary” and “dept” among all employees. Therefore, I used deletion for missing data.

b) Reasons for Graph Choice

Both “Salary” and “dept” are categorical variables. Therefore, it is suitable

for stacked bar chart to reveal the relationship between them, with one variable's frequency represented by the height of bars, the other's frequency contained inside each bar.

c) Strengths and Weaknesses

Strengths: Stacked bar charts are easy to understand, with absolute frequency of one categorical variable (working department) shown by bars. Users can also visualize the distribution of Salary inside each bar, revealing the salary information of each individual department.

Weaknesses: Like pie charts, only the relative amount of Salary is shown, but lack of absolute frequency of each Salary class. Furthermore, the actual percentage is ambiguous for the variable contained in the bar.

d) Audience

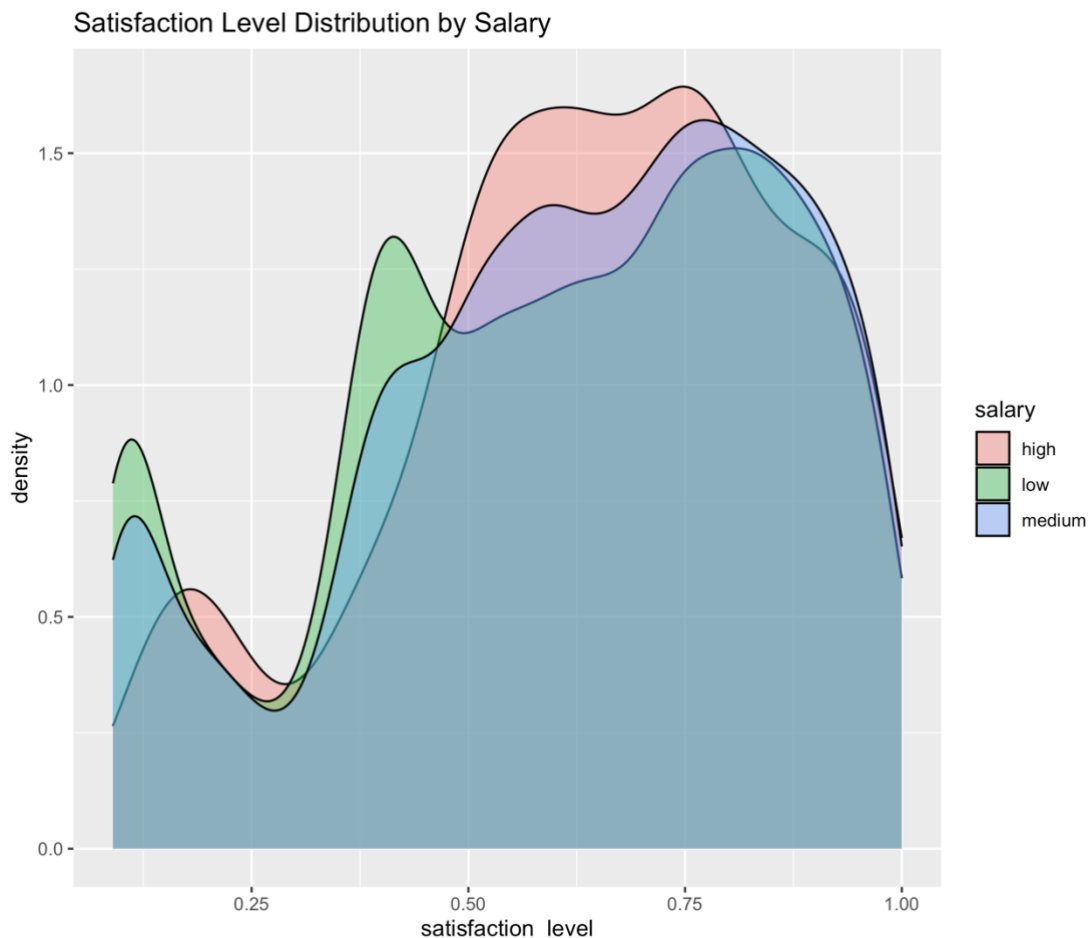
This stacked bar chart targets on those who require an overview of the relationship between departments and salary among employees, especially jobseekers with diverse background or multiple expected departments, managers, and collaborators of this company.

e) Data Story

The distribution of Salary is similar among each individual department, showing that department and salary are independent to some extent when it comes to satisfaction levels or employee attritions. In other words, the attrition rate and satisfaction levels should be similar across departments.

4. Bivariable Graph – Quantitative-Categorical

I chose "Salary" and "satisfaction levels" variables.



- a) **Missing Data Treatment**
The missing data has a percentage of around 4.99%, and we know nothing about the distribution of “satisfaction levels” and “salary” among all employees. Therefore, I used deletion for missing data.
- b) **Reasons for Graph Choice**
“Salary” and “dept” is a categorical variable, and “satisfaction levels” is a quantitative (numeric) variable. Therefore, it is suitable to use grouped kernel density plots, revealing the distribution of “satisfaction levels” across each “Salary” class.
- c) **Strengths and Weaknesses**
Strengths: Grouped kernel density plots are easy to understand, with a continuous density distribution showing absolute frequency of “satisfaction levels”. By using different colors, users can also visualize the frequency of “satisfaction levels” under certain Salary level (low, medium, and high), directly comparing the differences of “satisfaction levels” among Salary levels.
Weaknesses: No percentage or absolute frequency of each Salary level is revealed, lack of the impact of Salary distribution on “satisfaction levels”.
- d) **Audience**
This grouped kernel density plot targets on those who require an overview of the relationship between “satisfaction levels” and salary among employees,

especially jobseekers with concerns of the impact of low salary on employees' "satisfaction levels", managers requiring decision-makings, and collaborators of this company.

e) Data Story

The distribution of "satisfaction levels" is similar among each individual Salary level, showing that "satisfaction levels" and salary are independent to some extent when it comes to employee satisfactions and attritions. In other words, the attrition rate and satisfaction levels should not be affected by employees' salary.