BA - Descriptive Statistics

Sona Khloyan, Rafayel Mnatsakanyan, Perchuhi Proshyan

2022-12-13

```
turnover <- read.csv("turnover_data.csv")
head(turnover)
```

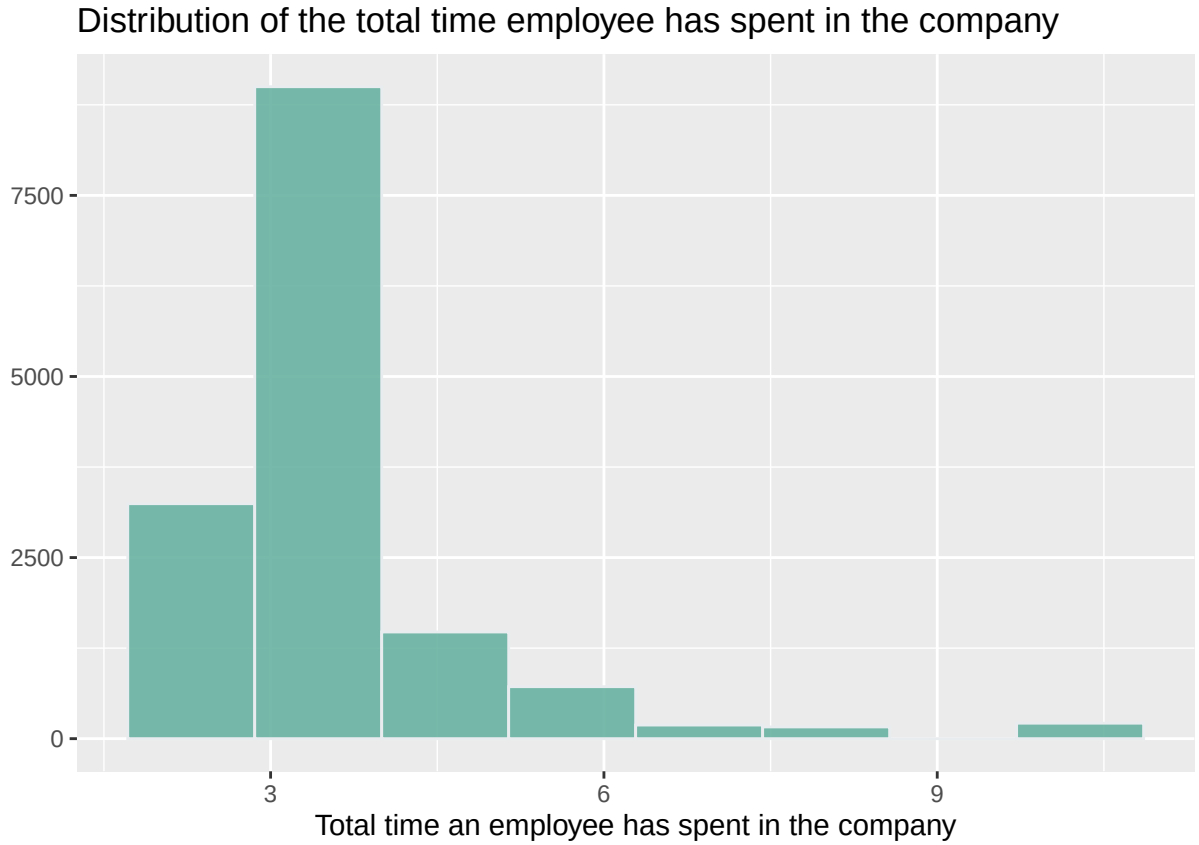
```
##   satisfaction_level last_evaluation number_project average_montly_hours
## 1               0.38            0.53              2                  157
## 2               0.80            0.86              5                  262
## 3               0.11            0.88              7                  272
## 4               0.72            0.87              5                  223
## 5               0.37            0.52              2                  159
## 6               0.41            0.50              2                  153
##   time_spend_company work_accident churn promotion_last_5years department
## 1                  3             0     1                     0      sales
## 2                  6             0     1                     0      sales
## 3                  4             0     1                     0      sales
## 4                  5             0     1                     0      sales
## 5                  3             0     1                     0      sales
## 6                  3             0     1                     0      sales
##   salary
## 1    low
## 2 medium
## 3 medium
## 4    low
## 5    low
## 6    low
```

```
summary(turnover)
```

```
##  satisfaction_level last_evaluation  number_project  average_montly_hours
##  Min.   :0.0900     Min.   :0.3600   Min.   :2.000   Min.   : 96.0
##  1st Qu.:0.4400     1st Qu.:0.5600   1st Qu.:3.000   1st Qu.:156.0
##  Median :0.6400     Median :0.7200   Median :4.000   Median :200.0
##  Mean   :0.6128     Mean   :0.7161   Mean   :3.803   Mean   :201.1
##  3rd Qu.:0.8200     3rd Qu.:0.8700   3rd Qu.:5.000   3rd Qu.:245.0
##  Max.   :1.0000     Max.   :1.0000   Max.   :7.000   Max.   :310.0
##  time_spend_company work_accident        churn        promotion_last_5years
##  Min.   : 2.000     Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.: 3.000     1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median : 3.000     Median :0.0000   Median :0.0000   Median :0.00000
##  Mean   : 3.498     Mean   :0.1446   Mean   :0.2381   Mean   :0.02127
##  3rd Qu.: 4.000     3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :10.000     Max.   :1.0000   Max.   :1.0000   Max.   :1.00000
```

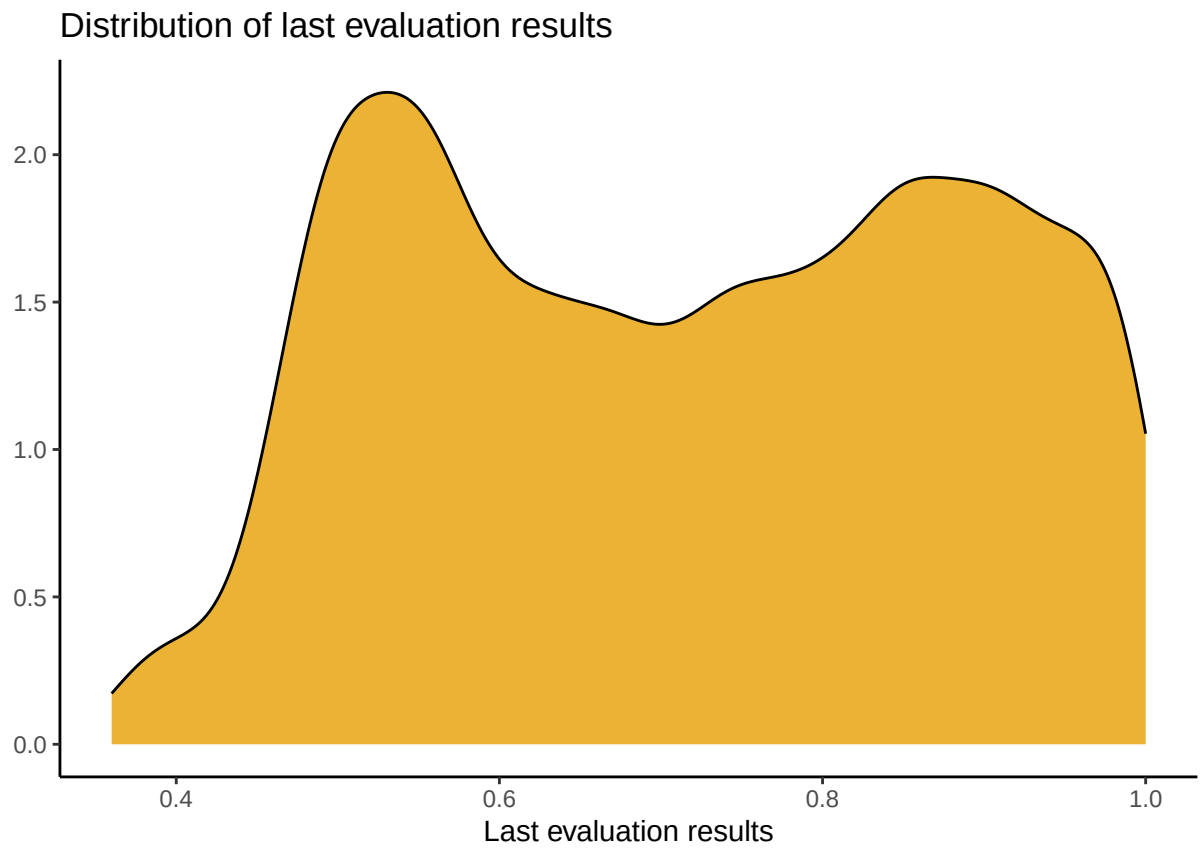
```
##   department           salary
##  Length:14999       Length:14999
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
```

```
ggplot(turnover, aes(x = time_spend_company)) +geom_histogram( bins = 8, fill="#69b3a2", color="#e9ecef
ggtitle("Distribution of the total time employee has spent in the company")
```



By this graph we see and confirm the pattern that employees mostly leave the company after 2-5 years, maybe for personal and career growth. Also, employees do not spend more than 10 years in one company.

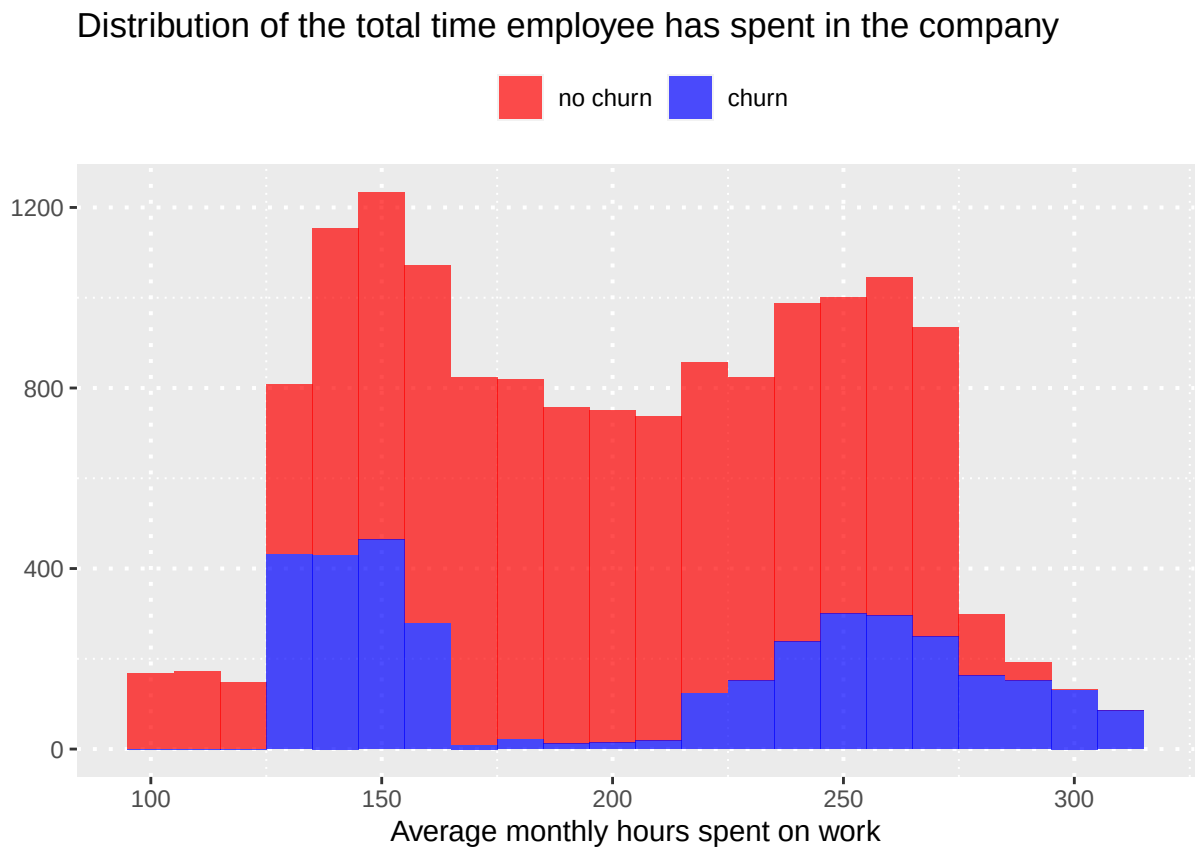
```
ggplot(turnover, aes(x = last_evaluation)) + geom_density(fill="#E69F00", color="black", alpha=0.8) + tl
```



The results of last evaluation are visualized above which mostly lay in range lower than 0.6 . Lower quantity of employees have result of 0.9.

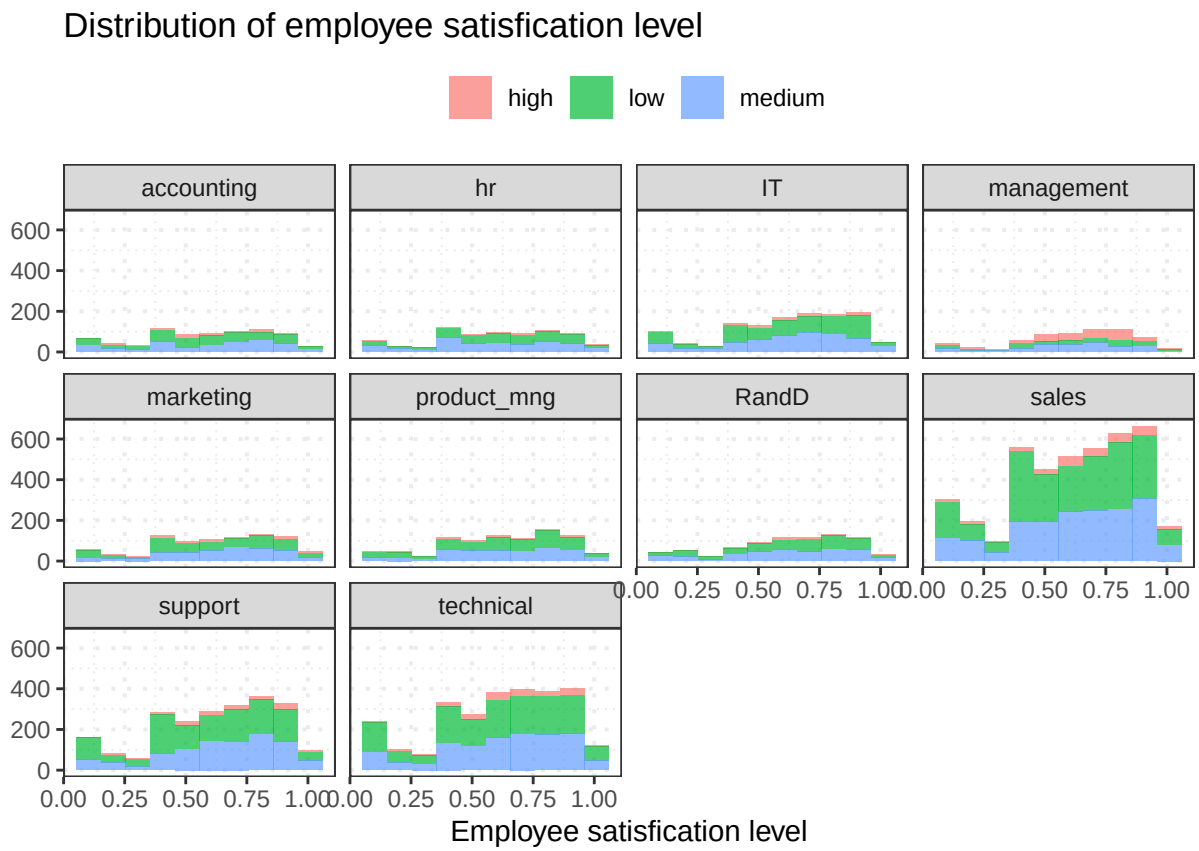
```
ggplot(turnover, aes(x = average_montly_hours, fill = factor(churn))) +
  geom_histogram(binwidth = 10, alpha=0.7) +
  labs(title = "Distribution of the total time employee has spent in the company",
       x = "Average monthly hours spent on work", y = "") +
  scale_fill_manual(values = c("red", "blue"), labels = c("no churn","churn")) +
  theme(legend.title = element_blank(), legend.position = "top",
        legend.direction = "horizontal",
        panel.grid = element_line(linetype = "dotted", size = 0.7))
```

```
## Warning: The ‘size‘ argument of ‘element_line()‘ is deprecated as of ggplot2 3.4.0.
## i Please use the ‘linewidth‘ argument instead.
```



The plot above represents the distribution of average monthly hours employees spend on work, the coloring is done on factor whether the employees churn the company or not. Conclusion is that the more hours employees spend on work the likelier is that they'll stay in the company.

```
ggplot(data = turnover, aes(x = satisfaction_level, fill = salary)) +
  geom_histogram(bins = 10, alpha=0.7) +
  theme_bw() +
  facet_wrap(.~department) +
  labs(x = "Employee satisfication level", y = "",
       title = "Distribution of employee satisfication level") +
  theme(legend.title = element_blank(), legend.position = "top",
        legend.direction = "horizontal",
        panel.grid = element_line(linetype = "dotted", size = 0.7))
```

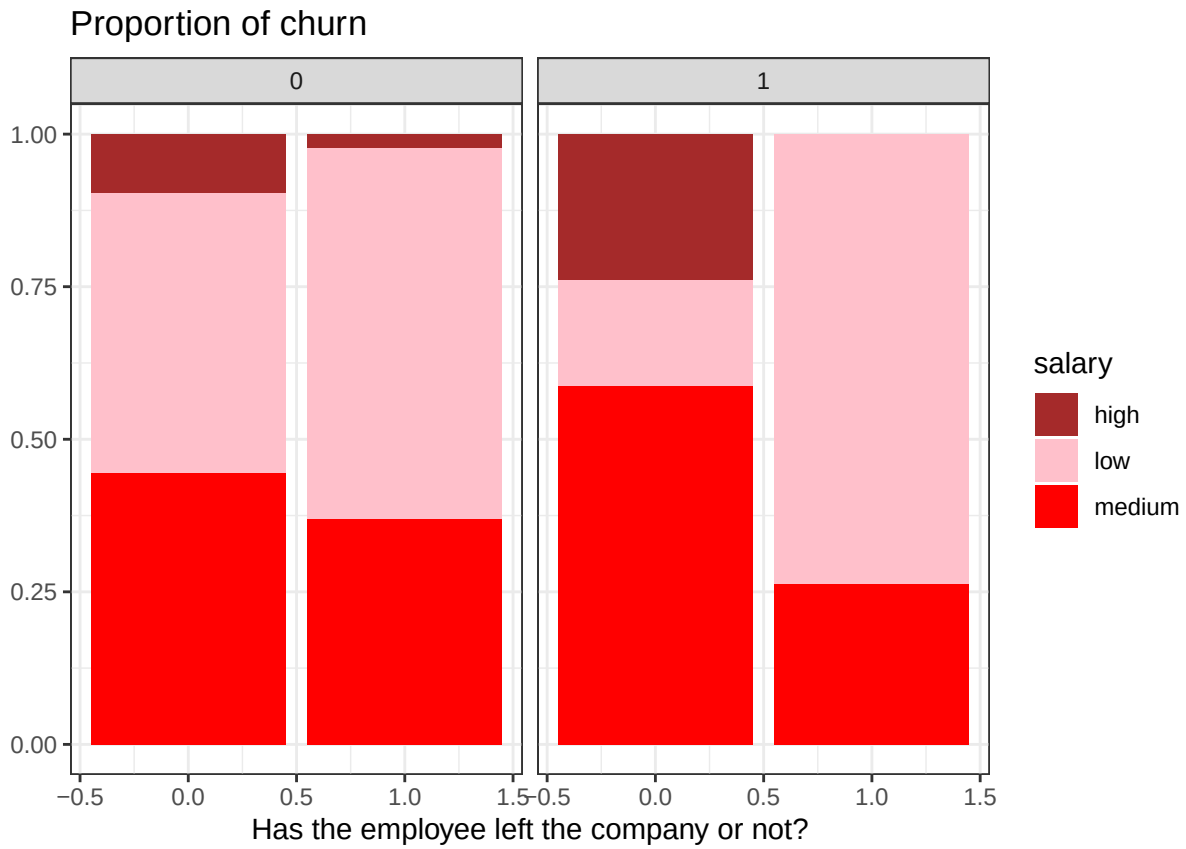


The graph above represents the satisfactory level of employees in different departments. Coloring is done based on the salary level either it is low, medium or high. We can see that mostly people with medium salary have low satisfactory level. Also, we see that people in the management department are paid well, high (which is intuitive). The last thing is that employees in sales, support, and technical departments have comparatively higher satisfactory level.

```
ggplot(turnover, aes(x = number_project, fill = factor(work_accident))) +
  geom_histogram(binwidth = 1) +
  labs(x = "Number of projects", y = "") +
  scale_fill_manual(values = c("red", "blue")) +
  ggtitle("Distribution of the total time employee has spent in the company") +
  theme_clean()+
  theme(legend.title = element_blank(), legend.position = "top",
        legend.direction = "horizontal",
        panel.grid = element_line(linetype = "dotted", size = 0.7))
```

The graph above represent the following: - The distribution of the number of projects can be considered to be normal, though it is a bit right skewed.The average number of projects an employee conducts at a time is 4. - Also, very little amount of employees experience work accidents. The accidents may happen regardless the number of projects employee conducts.

```
ggplot(turnover, aes(x = churn, fill = salary)) +
  geom_bar(position = "fill") +
  facet_grid(.~promotion_last_5years) +
  scale_fill_manual(values = c("brown", "pink", "red")) +
  labs(x = "Has the employee left the company or not?",
       y = "", title = "Proportion of churn") +
  theme_bw() + theme(panel.background = element_rect(color = "black"))
```



The graph above represents the proportion of churns taking into account the fact whether the employee had promotion during last 5 years or not. We can see that in both cases employees with low salary churn the most. Also, among the ones who do not churn are employees with medium salary rate.