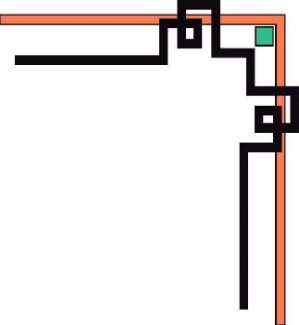
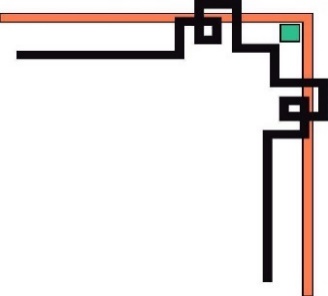
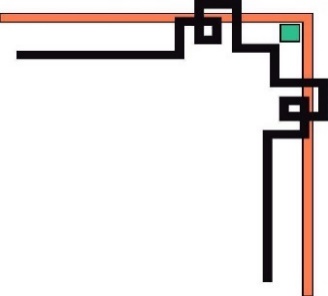
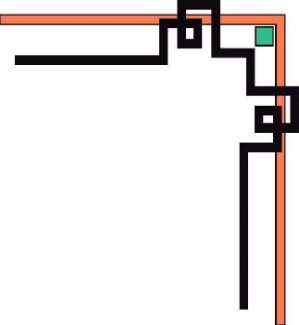
TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HỒ CHÍ MINH



KHOA CÔNG NGHỆ THÔNG TIN

-----🙞🙜🕮🙞🙜-----

****

**ĐỀ TÀI**

**PHÂN TÍCH NGUỒN NHÂN LỰC**

**Môn: Hệ thống thông tin quản lý**

**Giảng viên hướng dẫn:** TS.Nguyễn Hoàng Long

**Thành viên nhóm 4:**

Trần Thị Lý 14110111

Trần Nhật Tuấn 14110220

Phạm Thị Minh Huệ 14110074

Nguyễn Thị Phương Uyên 14110231

# ĐẶT VẤN ĐỀ

Ở các công ty nhân sự luôn luôn là một vấn đề đau đầu đối với các lãnh đạo. Tìm được người phù hợp đã khó, giữ người còn khó hơn. Vì thế bài phân tích này của nhóm sẽ nhằm mục đích tìm ra một số điểm “đáng lưu ý” khi nhân viên ra đi và tại sao những nhân viên tốt nhất và giàu kinh nghiệm lại nghỉ việc?

Nhóm sẽ sử dùng tập dữ liệu [Human Resources A](https://bambo.vn/wp-content/uploads/2017/11/HR_comma_sep.csv)nalytics là tập dữ liệu được cung cấp trên [Kaggle](https://www.kaggle.com/yassineghouzam/don-t-know-why-employees-leave-read-this/data) để phân tích.

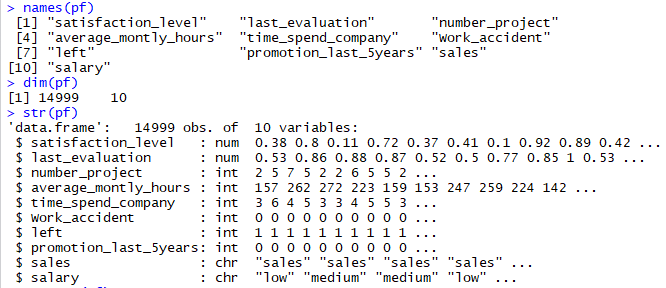
RStudio dùng ngôn ngữ R là một trong những công cụ mạnh dùng để phân tích dữ liệu. Nhóm chúng tôi sẽ sử dụng công cụ này để làm rõ những vấn đề trên.

# NỘI DUNG

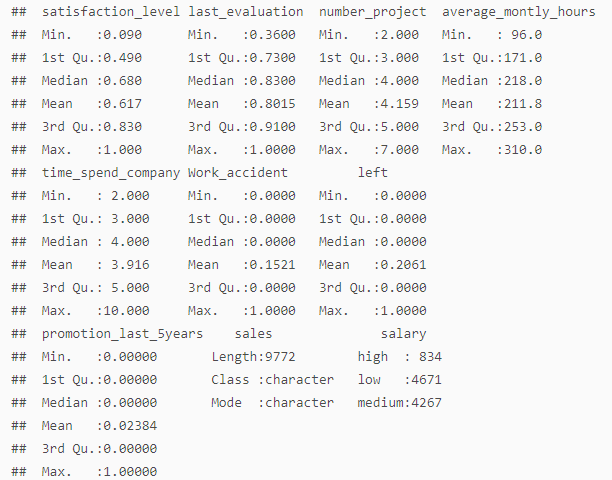
## 1. Giới thiệu

* [Human Resources A](https://bambo.vn/wp-content/uploads/2017/11/HR_comma_sep.csv)nalytics là tập dữ liệu được cung cấp trên [Kaggle](https://www.kaggle.com/yassineghouzam/don-t-know-why-employees-leave-read-this/data).
* Bộ dữ liệu là tập hợp thông tin về công việc của các nhân viên của một công ty.
* Dataset có 14999 dòng tương ứng với thông tin của mỗi nhân viên và thông tin về công việc được thể hiện qua 10 thuộc tính :
* Satisfaction Level : Mức độ hài lòng của nhân viên đối với công ty.
* Last evaluation: Đánh giá nhân viên (có thể được điền bởi người quản lý của nhân viên).
* Number of projects: Số dự án mà nhân viên làm.
* Average monthly hours: Thời gian làm việc trung bình trong tháng.
* Time spent at the company: Thời gian ở công ty
* Work\_accident: Tai nạn nghề nghiệp.
* Promoted\_last\_5years: Khả năng thăng tiến trong 5 năm.
* Departments (column sales): Phòng ban (có 10 phòng ban).
* Salary: Lương (có 3 bậc: thấp, trung bình và cao).
* Left: Đánh dấu nhân viên đã rời đi hay ở lại.
* Thuộc tính quan trọng nhất là “left” biểu diễn:
* 0: nhân viên ở lại
* 1: nhân viên nghỉ việc

Để trả lời cho các câu hỏi tại sao nhân viên lại nghỉ việc? Những nhân viên có kinh nghiệm và giỏi lại rời đi? Chúng ta sẽ đi vào phân tích các tính chất của mẫu dữ liệu bằng ngôn ngữ R và dự đoán kết quả.



> summary(pf)



Tổng quan về dữ liệu

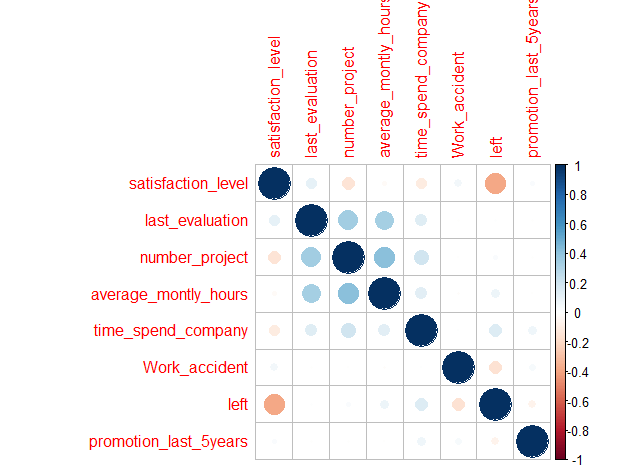
## 2. Phân tích chi tiết dữ liệu

Trước tiên, để tổng quan về các mỗi quan hệ của thuộc tính. Ta sẽ xem qua mức độ tương quan giữa các tính chất với nhau.

### Ma trận độ tương quan giữa các thuộc tính:

|  |
| --- |
| > library(corrplot)  > cor(pf[,1:8])  satisfaction\_level last\_evaluation number\_project  satisfaction\_level 1.00000000 0.105021214 -0.142969586  last\_evaluation 0.10502121 1.000000000 0.349332589  number\_project -0.14296959 0.349332589 1.000000000  average\_montly\_hours -0.02004811 0.339741800 0.417210634  time\_spend\_company -0.10086607 0.131590722 0.196785891  Work\_accident 0.05869724 -0.007104289 -0.004740548  left -0.38837498 0.006567120 0.023787185  promotion\_last\_5years 0.02560519 -0.008683768 -0.006063958  average\_montly\_hours time\_spend\_company Work\_accident  satisfaction\_level -0.020048113 -0.100866073 0.058697241  last\_evaluation 0.339741800 0.131590722 -0.007104289  number\_project 0.417210634 0.196785891 -0.004740548  average\_montly\_hours 1.000000000 0.127754910 -0.010142888  time\_spend\_company 0.127754910 1.000000000 0.002120418  Work\_accident -0.010142888 0.002120418 1.000000000  left 0.071287179 0.144822175 -0.154621634  promotion\_last\_5years -0.003544414 0.067432925 0.039245435  left promotion\_last\_5years  satisfaction\_level -0.38837498 0.025605186  last\_evaluation 0.00656712 -0.008683768  number\_project 0.02378719 -0.006063958  average\_montly\_hours 0.07128718 -0.003544414  time\_spend\_company 0.14482217 0.067432925  Work\_accident -0.15462163 0.039245435  left 1.00000000 -0.061788107  promotion\_last\_5years -0.06178811 1.000000000 |

|  |
| --- |
| > corrplot(cor(pf[,1:8]), method="circle") |
|  |
| |  | | --- | |  | |



H.2.1

Qua ma trận tương quan trên, ta có thể thấy các mỗi quan hệ:

- Những nhân viên có số dự án nhiều thì thời gian làm việc trung bình trong tháng cũng nhiều, và cũng được đánh giá cao hơn.

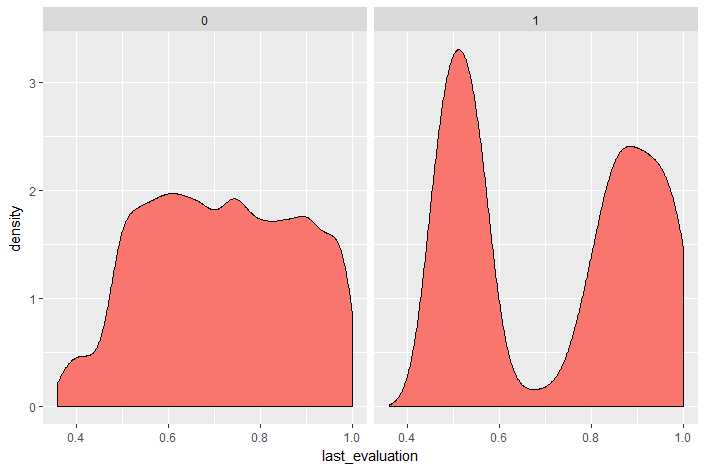
- Nhân viên làm việc lâu năm thì càng tương quan với việc ra đi hơn.

- Mối quan hệ giữa nghỉ việc và mức độ hài lòng là tương quan nghịch (negative correlation). Ta có thể thấy trên độ tương quan lớn là -0.38, những nhân viên càng hài lòng (satisfaction\_level lớn) với công việc thì sẽ ở lại (left = 0) và ngược lại, những nhân viên có mức độ hài lòng càng thấp khả năng ra đi sẽ rất cao.

Để rõ hơn về các mối quan hệ, ta sẽ phân tích thông qua các biểu đổ dưới đây.

### 2.1 Nghỉ việc với đánh giá nhân viên

|  |
| --- |
| > ggplot(pf,aes(last\_evaluation)) +  + geom\_density(fill = "#F8766D") +  + facet\_wrap( ~ left, ncol = 2) |
|  |
| |  | | --- | | > | |

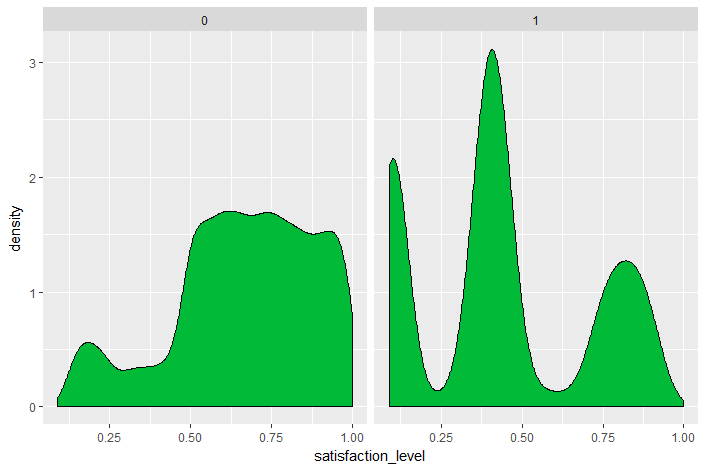


H.2.1.1

* Có thể thấy những nhân viên nghỉ việc nằm trong vùng đánh giá quá thấp thì nghỉ việc nhiều.
* Nhân viên nằm trong vùng đánh giá cao cũng ra đi.
* Phần lớn các nhân viên nằm trong vùng 0.6 – 0,8 thì ở lại với công ty.
* Ta có thể giả định, có thể những người được đánh giá 0.85 – 1 ra đi nhiều vì họ cảm thấy công ty hiện tại không phù hợp với trình độ của họ ra đi với mong muốn tìm được nơi có thể phát triển hết năng lực và những nhân viên có trình độ thấp cũng vậy.
* Và có thể việc đánh giá trình độ của bản thân quá thấp sẽ khiến cho nhân viên có cảm giác chán nản và ít có động lực, mục tiêu để tiếp tục công việc.
* Đôi khi đối với những nhân viên chúng ta cần thì không nên đánh giá họ một cách chân thực. Vì như vậy họ có thể rời đi vì công ty không thể phát triển tài năng của họ hoặc rời đi vì không có mục đích, chủ động.

### 2.2 Nghỉ việc với mức độ hài lòng

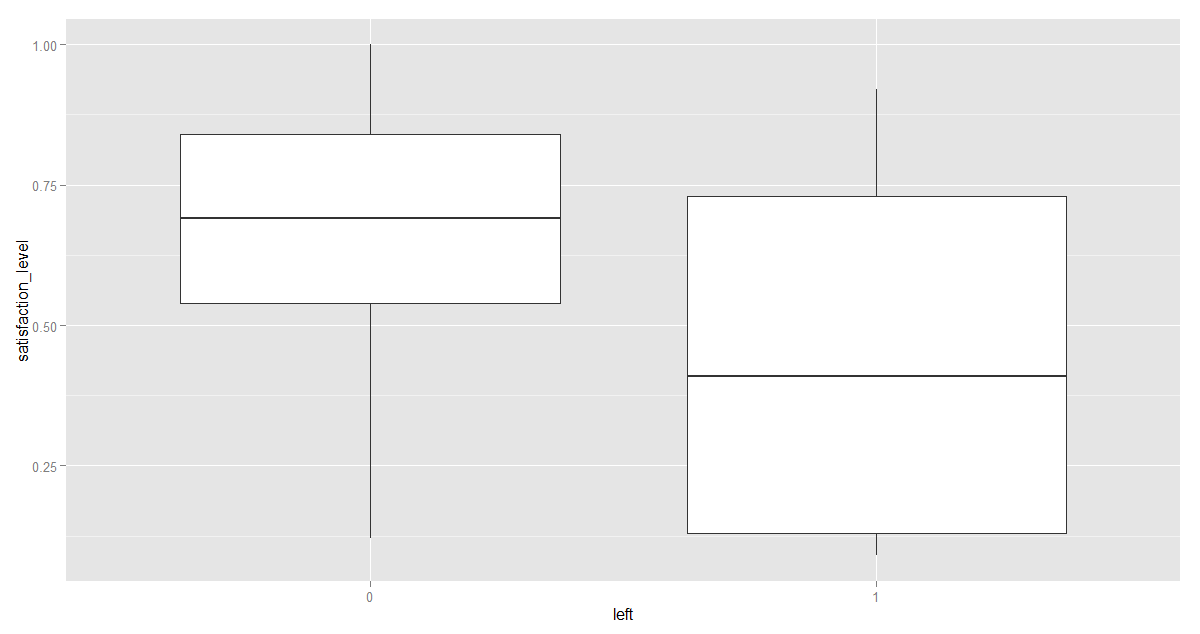
|  |
| --- |
| > ggplot(pf,aes(satisfaction\_level)) +  + geom\_density(fill = "#00BA38") +  + facet\_wrap( ~ left, ncol = 2) |
|  |
| |  | | --- | | > | |



H.2.2.1

> ggplot(pf) +

+ geom\_boxplot(aes(x=left,y=satisfaction\_level))

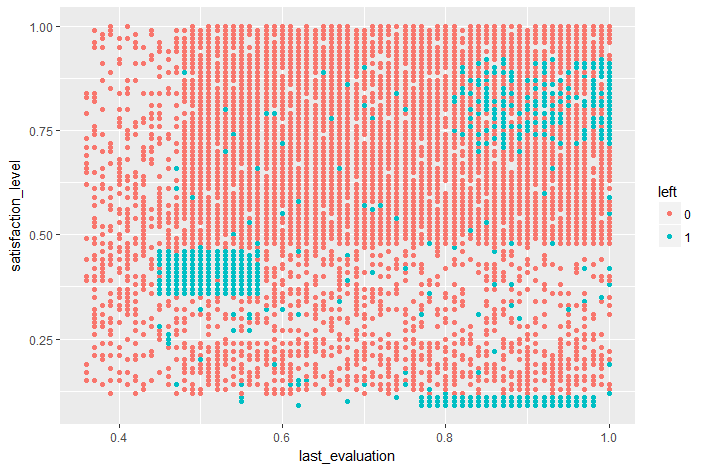


H.2.2.2

* Từ biểu đồ vê ̀mức độ hài lòng của công ty ảnh hưởng tới việc ở lại làm cho công ty. Chúng ta cũng có thể dể hiểu rằng: những nhân viên có mức độ hài lòng thấp (<0.2) và mức thấp (0.3 – 0.5) sẽ rời đi rất nhiều là điều rất bình thường.
* Nếu như môi trường làm việc của bạn không thoải mái, chán nản, không tạo hứng thú trong công việc thì làm sao có thể tiếp tục công việc nữa. Và những người tiếp tục làm ở đây cũng có thể có hoàn cảnh khác nên phải tiếp tục duy trì. Ví dụ như về lương bổng, trình độ chỉ phù hợp làm việc tại đây, tìm công việc khó...
* Và có một đáng chú ý ở đây là những nhân viên có mức hài lòng về công ty khá cao. Vậy thì lý do gì đã làm họ ra đi? Là vì năng lực của họ yếu? Ta kết hợp 2 thuộc tính về mức độ hài lòng và đánh giá trình độ của nhân viên để làm rõ lý do.

### 2.3 Đánh giá với mức độ hài lòng của nhân viên

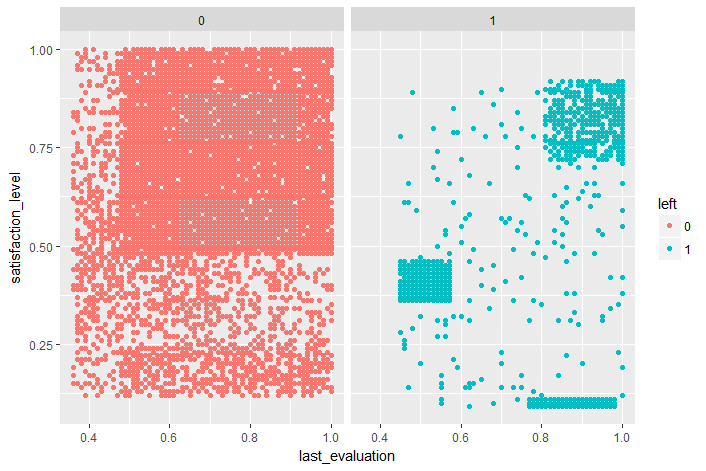
|  |
| --- |
| > pf$left <- as.factor(pf$left)  > ggplot(pf,aes(last\_evaluation,satisfaction\_level,colour= left ))+  + geom\_point() |
|  |
| |  | | --- | |  | |



H.2.3.1

Có rõ ràng 3 nhóm riêng biệt rời đi và nhóm ở lại tập trung đông ở mức hài lòng trung bình, cao về công ty. Để rõ hơn, ta phân biệt thành 2 biểu đồ riêng biệt.

|  |
| --- |
| > pf$left <- as.factor(pf$left)  > ggplot(pf,aes(last\_evaluation,satisfaction\_level,colour= left ))+  + geom\_point()+  + facet\_wrap( ~ left, ncol = 2) |
|  |
| |  | | --- | |  | |



H.2.3.2

Có 3 nhóm nhân viên có xu hướng rời công ty:

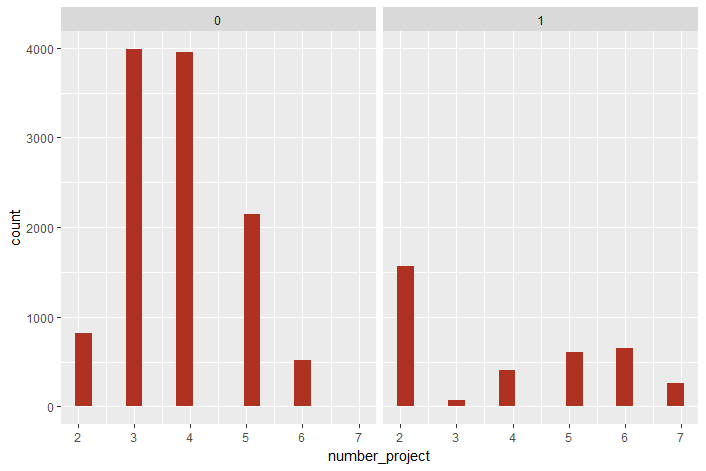
* Nhóm nhân viên có mức độ hài lòng thấp (0.2) và được đánh giá cao (0.75)
  + Những nhân viên có những tố chất rất tốt, làm việc hiệu quả, nhưng cảm thấy chán nản với công việc mình đang làm.
  + Điều gì là lý do khiến các nhân viên tốt cảm thấy buồn chán, có phải vì họ phải làm việc quá nhiều hay bản thân nội dung công việc đó nhàm chán ?
* Nhóm nhân viên được đánh giá thấp, độ thoả mãn thấp.
* Đây là nhóm nhân viên kém của công ty.
* Nhóm nhân viên đặc biệt có mức độ hài lòng cao, từ 0.7~1.0 và được đánh giá trên 0.8.
* Đây là nhóm nhân viên tốt của công ty, nhưng họ lại rời đi.
  + Có phải họ rời đi vì họ tìm thấy công việc khác tốt hơn ? Chúng ta sẽ dựa vào phân tích các đặc điểm của nhóm đối tượng này trên các thuộc tính khác để hiểu rõ hơn.

- Lấy dữ liệu của những nhân viên rời đi:

|  |
| --- |
| > dt\_left\_1 <- subset(pf, left=='1');  > str(dt\_left\_1)  'data.frame': 3571 obs. of 10 variables:  $ satisfaction\_level : num 0.38 0.8 0.11 0.72 0.37 0.41 0.1 0.92 0.89 0.42 ...  $ last\_evaluation : num 0.53 0.86 0.88 0.87 0.52 0.5 0.77 0.85 1 0.53 ...  $ number\_project : int 2 5 7 5 2 2 6 5 5 2 ...  $ average\_montly\_hours : int 157 262 272 223 159 153 247 259 224 142 ...  $ time\_spend\_company : int 3 6 4 5 3 3 4 5 5 3 ...  $ Work\_accident : int 0 0 0 0 0 0 0 0 0 0 ...  $ left : int 1 1 1 1 1 1 1 1 1 1 ...  $ promotion\_last\_5years: int 0 0 0 0 0 0 0 0 0 0 ...  $ sales : chr "sales" "sales" "sales" "sales" ...  $ salary : chr "low" "medium" "medium" "low" ... |
|  |
| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | > dim(dt\_left\_1)  [1] 3571 10   |  |  |  |  |  | | --- | --- | --- | --- | --- | | |  | | --- | | > dim(datatemp)  [1] 943 10 **2.3.1 Mức lương nhân viên** > ggplot (datatemp, aes(salary,fill = "left "))+  + geom\_bar() | |  | | |  | | --- | | > | |     H.2.3.1.1 **2.3.2 Thời gian làm việc tại công ty** | | |  | | --- | | > ggplot (datatemp, aes(time\_spend\_company,fill="left " ))+  + geom\_bar() | |  | | |  | | --- | | > | |     H.2.3.2.1 |  **2.3.3 Khả năng thăng tiến trong công việc**  |  | | --- | | > ggplot (datatemp, aes(promotion\_last\_5years,fill="left "))+  + geom\_bar() | |  | | |  | | --- | | > | |     H.2.3.3.1 **2.3.4 Ngành nghề**  |  | | --- | | > ggplot (datatemp, aes(sales,fill="left "))+  + geom\_bar() | |  | | |  | | --- | | > | |     H.2.3.4.1  Qua các biểu đồ, nhóm đối tượng được đánh giá cao và mức đồ hài lòng cao này không ở lại có thể vì những lý do sau:   * Lương: đã số những nhân viên rời đi lương rất và trung bình rất nhiều. Có thể họ không thể thõa mãn số lương đó được. * Khả năng thăng tiến rất thấp, cứ khoảng 900 ra đi chỉ có 1 -2 người có khả năng thăng tiến trong 5 năm. Và đã số họ đều có thời gian kinh nghiệm 5 năm trở lên. * Lời khuyên cho công ty rằng nên xác định rõ ràng những ai có năng lực và có chiến lược giữ lại khác. Còn đối với những nhân viên không có năng lực cộng thêm mức độ hài lòng về công ty thấp thì không nên vì ít ai có thể làm việc trong môi trường mình không hứng thú mà mang lại hiệu suất cao được. | |

### 2.4 Nghỉ việc với số lượng project đã tham gia

|  |
| --- |
| > ggplot(pf,aes(x=number\_project)) +  + geom\_histogram(binwidth=.3, alpha = 1, fill = "#F8766D")+  + facet\_wrap( ~ left,ncol=2) |
|  |
| |  | | --- | | > | |



H.2.4.1

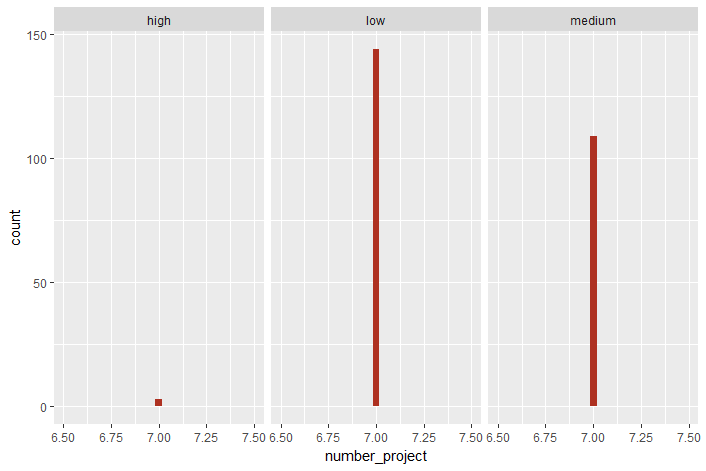
* Hơn nửa số nhân viên ra đi khi đã làm 2, 6 projects.
* Phần lớn số nhân viên không rời công ty khi làm 3, 4, 5 project.
* Còn những người vừa vào làm được 2 project thì họ lại nghỉ, có thể là công việc quá dễ hoặc quá khó với trình độ của họ lúc đó, nên họ chán nản không muốn tiếp tục ở lại.
* Những người làm 6 projects thì chắc họ là những nhân viên lâu năm, họ đã học được nhiều kiến thức và kinh nghiệm, họ không có gì hứng thú để ở lại.
* Số nhân viên làm 7 projects ra đi hoàn toàn. Tìm hiểu mối quan hệ với lương để xác định nhân viên có được đối xử công bằng không?
* Có dấu hiệu tăng của tỉ lệ ra đi khi càng làm nhiều project.

> SLPR <- subset(pf,number\_project=="7")

> ggplot(SLPR,aes(x=number\_project)) +

+ geom\_histogram(alpha = 1, fill = "#AE3121") +

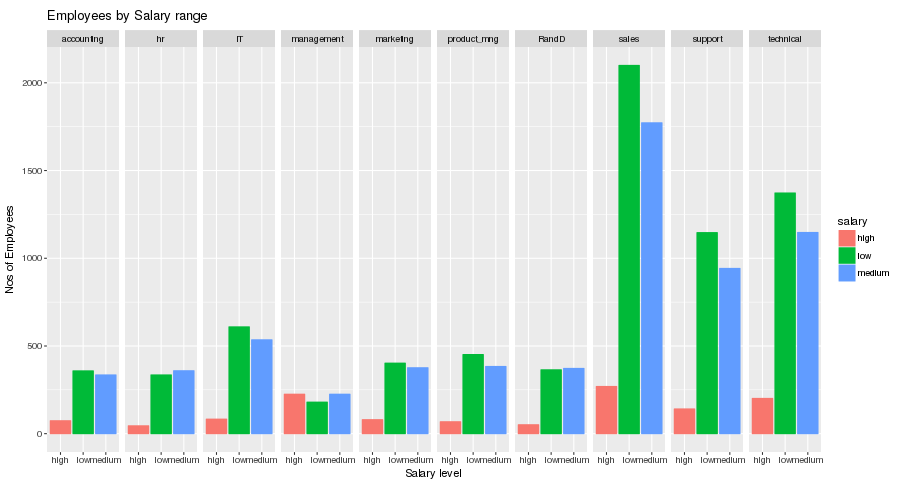
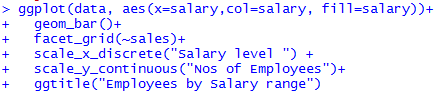
+ facet\_wrap( ~ salary, ncol = 3)



H.2.4.2

Nhân viên làm việc với số dự án nhiều nhưng lương thấp và trung bình lại rất nhiều. Có thể có sự không công bằng ở đây nên nhân viên đã rời đi.

### 2.5 Bộ phận và lương

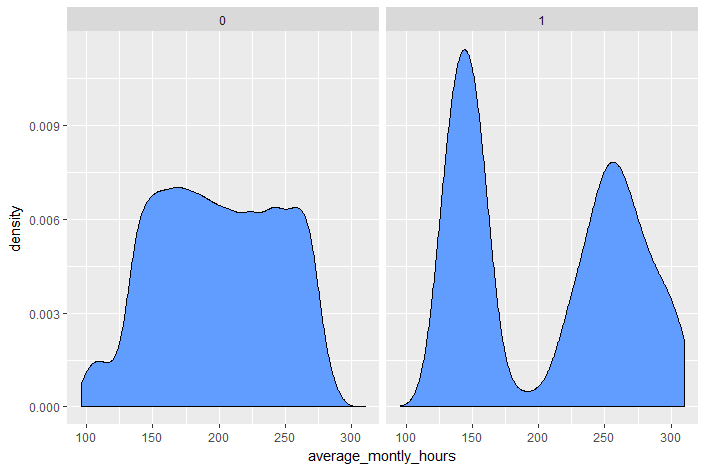


H.2.5.1

* Số lượng nhân viên làm việc ở sale, support và technical rất nhiều nhưng lương không cao
* Số lượng nhân viên ở management tuy không nhiều nhưng không có sự chênh lệch cao giữa các mức lương.
* Các bộ phận có người nghỉ nhiều nhất là sales, technical và support. Có thể thấy rằng sales và technical là các bộ phận thường xuyên chịu áp lực lớn về doanh số cũng như deadline.

### 2.6 Nghỉ việc với thời gian trung bình làm việc trong tháng

|  |
| --- |
| > ggplot(pf, aes(x = average\_montly\_hours)) + |
| + geom\_freqpoly(binwidth = 10, aes(color = left)) |
| |  | | --- | |  | |

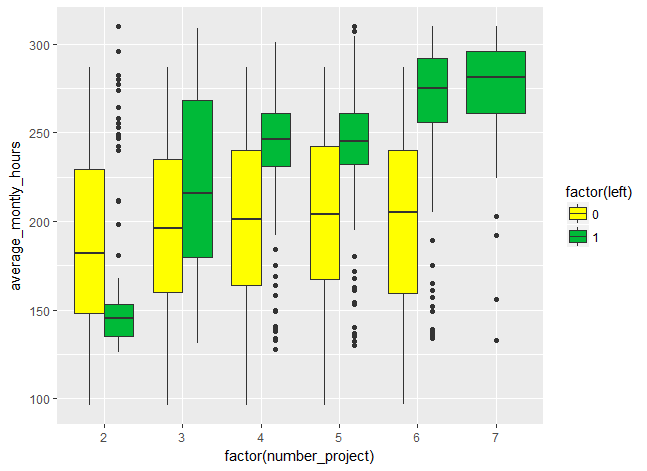


H.2.6.1

* Nhóm nhân viên nổi trội về số giờ làm việc trên 250 giờ và 140 giờ nghỉ việc vì thời gian làm việc quá ít và quá nhiều.

### 2.7 Số lượng project với thời gian làm việc trung bình trong tháng

|  |
| --- |
| > ggplot(pf, aes(x = factor(number\_project), y = average\_montly\_hours, fill = factor(left))) +  + geom\_boxplot() + scale\_fill\_manual(values = c("yellow", "#00BA38")) |
|  |
| |  | | --- | | > | |



H.2.7.1

🡪  Từ biểu đồ tương quan ở trên, có thể nhận ra được số giờ làm việc tỉ lệ thuận với số đồ án của nhân viên nên không thấy làm lạ khi những nhân viên làm nhiều ra đi.

### 2.8 Lương với nghỉ việc

> pf<-read.csv('HR\_comma\_sep.csv', stringsAsFactors = F)

> ggplot(pf,aes(x=salary)) +

+ geom\_histogram( aes(fill = as.factor(salary)))+

+ facet\_wrap(~left, ncol=2)



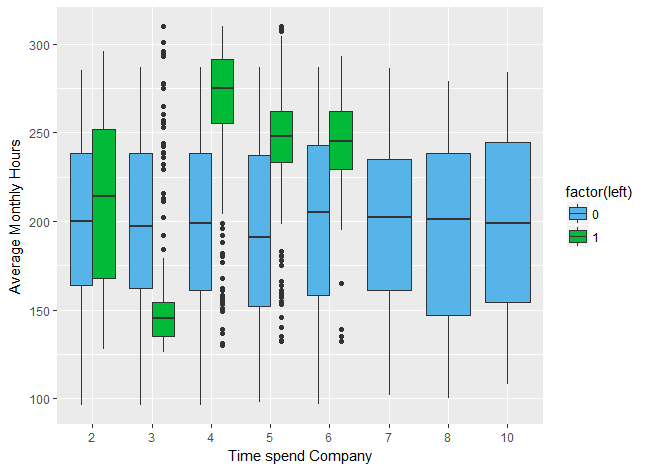
H.2.8.1

* Phần lớn nhân viên ra đi có lương thấp hoặc trung bình.
* Những nhân viên có lương cao ít ra đi.

🡪 Vẫn có những nhân viên lương cao ra đi.

### 2.9 Thời gian làm việc trung bình trong tháng và thời gian ở công ty

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| |  | | --- | | > ggplot(pf, aes(x = factor(time\_spend\_company), y = average\_montly\_hours, fill = factor(left))) +  + geom\_boxplot() + scale\_fill\_manual(values = c("#56B4E9", "#00BA38")) + xlab("Time spend Company") + ylab("Average Monthly Hours") | |  | | |  | | --- | | > | | |
|  |
|  |



H.2.9.1

* Những nhân viên có thời gian làm việc càng nhiều và số năm kinh nghiệm 4~6 thì càng muốn rời đi. Nhưng nếu họ đã làm lâu năm ở công ty (>6 năm) thì sẽ ở lại. Có thể những nhân viên đã gắn bó với công ty này quá lâu rồi nên họ sẽ quyết định gắn bó lâu dài hơn hoặc có thể sẽ chán nản khi bắt đầu một công việc mới, một công ty mới nên họ quyết định chọn việc ở lại.

### 2.10 Mối tương quan giữa thời gian làm việc và mức lương nhân viên

> library(gridExtra)

> pf$time\_spend\_company <- as.factor(pf$time\_spend\_company)

> g1 <- ggplot(data = subset(pf, left =="1"), aes(x =salary , fill = salary))+

+ geom\_bar()+

+ facet\_grid(~time\_spend\_company)+

+ scale\_x\_discrete("Time spend in company") +ggtitle("Left = 0")

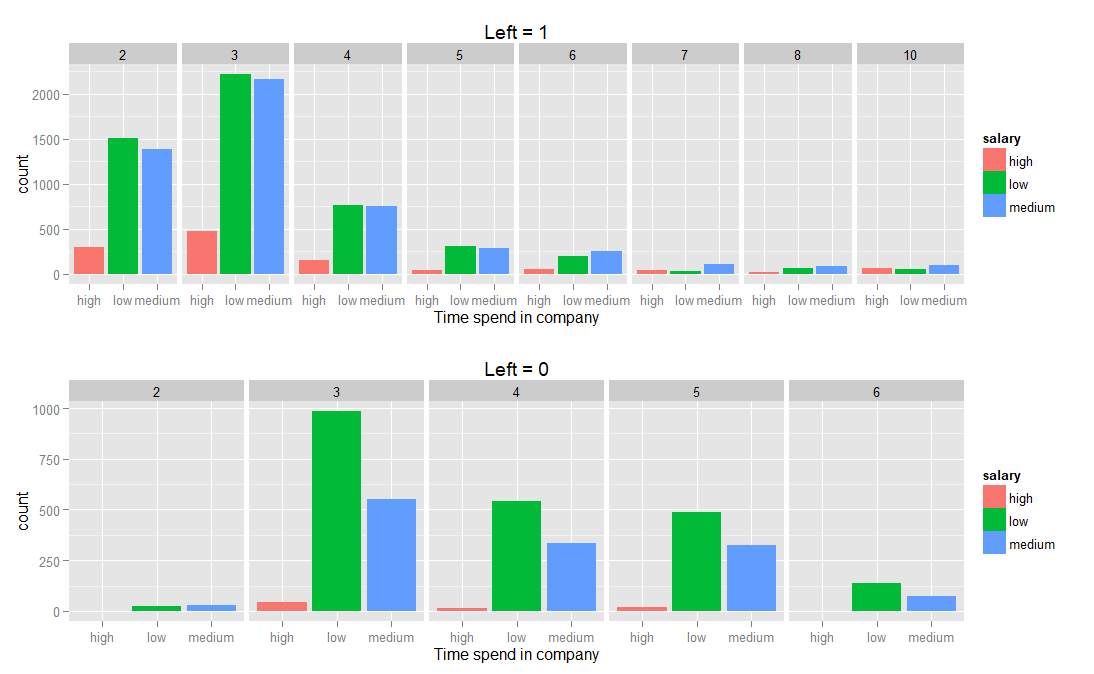
> g2 <- ggplot(data = subset(pf, left =="0"), aes(x =salary , fill = salary))+

+ geom\_bar()+

+ facet\_grid(~time\_spend\_company)+

+ scale\_x\_discrete("Time spend in company") +ggtitle("Left = 1")

> grid.arrange(g2,g1,ncol =1)



**KẾT LUẬN**

* Nhân viên làm ít việc quá sẽ nghỉ việc (ít hơn 150h/tháng hay 6h/ngày).
* Nhân viên làm nhiều việc quá sẽ nghỉ việc (nhiều hơn 250h/tháng hay 10h/ngày).
* Nhân viên được đánh giá quá thấp hay quá cao đều nghỉ.
* Lương thấp và trung bình cũng dễ dẫn tới nghỉ.
* Nhân viên tại các project thứ 2, 6, 7 có nguy cơ cao là nghỉ.
* Nhóm nhân viên tốt nhất khi làm việc từ 5 năm trở lên sẽ có xu hướng ra đi.
* Những người rời đi đều có kinh nghiêm,mức độ thỏa mãn thấp,dành nhiều thời gian trong công việc.

# TÀI LIỆU THAM KHẢO

* **(**Use R) Hadley Wickham (auth.)-ggplot2\_ Elegant Graphics for Data Analysis 2nd ed-Springer-Verlag New York (2016)
* Jiawei Han, Micheline Kamber, Jian Pei Data Mining. Concepts and Techniques, 3rd Edition 2011
* ggplot2-cheatsheet-2.0-Vietnamese.pdf
* <https://www.kaggle.com/sanjaym0410/hr-analytics-complete-analysis-and-prediction>
* <https://www.kaggle.com/ludobenistant/hr-analytics-1>