Human Resource Analytics

#### Loading The required packages

library(ggplot2)  
library(dplyr)  
library(tidyr)

#loading the Dataset  
  
hrm<-read.csv('HR\_comma\_sep.csv')  
  
#Structure of the Dataset  
str(hrm)

## 'data.frame': 14999 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 : Factor w/ 10 levels "accounting","hr",..: 8 8 8 8 8 8 8 8 8 8 ...  
## $ salary : Factor w/ 3 levels "high","low","medium": 2 3 3 2 2 2 2 2 2 2 ...

attach(hrm)  
  
#converting left variable to factor variable   
hrm$left<-ifelse(left==1,'True','False')  
  
hrm$left<-factor(hrm$left,levels=c("True","False"))  
table(hrm$left)

##   
## True False   
## 3571 11428

#Summary Statistics of the dataset  
summary(hrm)

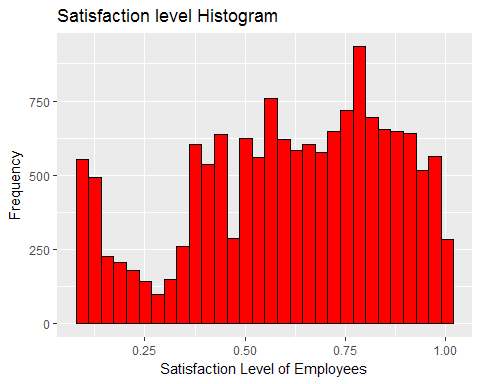
## satisfaction\_level last\_evaluation number\_project average\_montly\_hours  
## Min. :0.0900 Min. :0.3600 Min. :2.000 Min. : 96.0   
## 1st Qu.:0.4400 1st Qu.:0.5600 1st Qu.:3.000 1st Qu.:156.0   
## Median :0.6400 Median :0.7200 Median :4.000 Median :200.0   
## Mean :0.6128 Mean :0.7161 Mean :3.803 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 left promotion\_last\_5years  
## Min. : 2.000 Min. :0.0000 True : 3571 Min. :0.00000   
## 1st Qu.: 3.000 1st Qu.:0.0000 False:11428 1st Qu.:0.00000   
## Median : 3.000 Median :0.0000 Median :0.00000   
## Mean : 3.498 Mean :0.1446 Mean :0.02127   
## 3rd Qu.: 4.000 3rd Qu.:0.0000 3rd Qu.:0.00000   
## Max. :10.000 Max. :1.0000 Max. :1.00000   
##   
## sales salary   
## sales :4140 high :1237   
## technical :2720 low :7316   
## support :2229 medium:6446   
## IT :1227   
## product\_mng: 902   
## marketing : 858   
## (Other) :2923

Satisfaction level statistics splitted by salary ranges

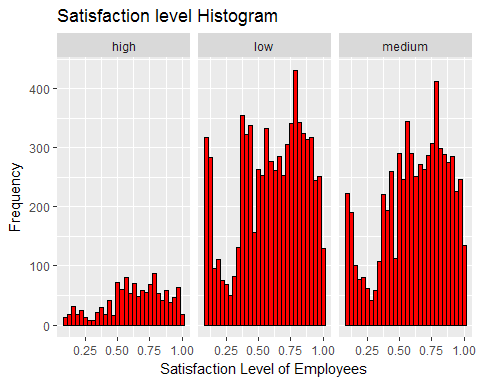
by(hrm$satisfaction\_level,hrm$salary,summary)

## hrm$salary: high  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0900 0.5000 0.6600 0.6375 0.8100 1.0000   
## --------------------------------------------------------   
## hrm$salary: low  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0900 0.4200 0.6300 0.6008 0.8100 1.0000   
## --------------------------------------------------------   
## hrm$salary: medium  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0900 0.4500 0.6600 0.6218 0.8200 1.0000

#Histogram  
  
p1<-ggplot(aes(x=satisfaction\_level),data=hrm) +   
 geom\_histogram(color="black",fill="red",bins = 30) +  
 labs(title="Satisfaction level Histogram",x='Satisfaction Level of Employees', y="Frequency")  
  
p1



#Satisfaction level histogram facetted by sallary classes  
  
p2 = p1 + facet\_wrap(~salary)  
p2



The disrtibution of satisfaction level for each class of Salary Ranges is almost same.

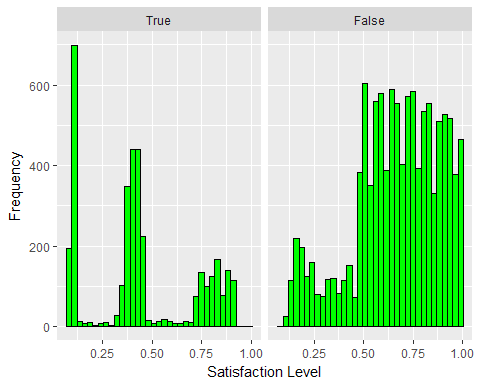
The High salary employees have a little bit higher mean satisfaction level.

Employees with High salary are less in number.

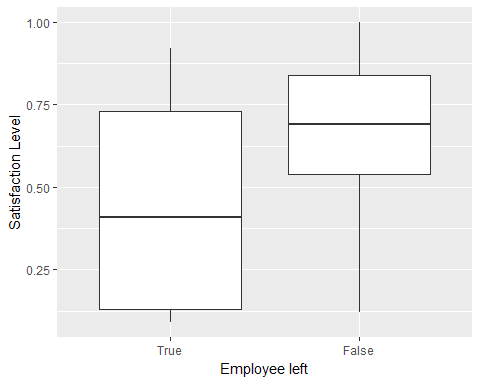
by(satisfaction\_level,left,summary)

## left: 0  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.1200 0.5400 0.6900 0.6668 0.8400 1.0000   
## --------------------------------------------------------   
## left: 1  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0900 0.1300 0.4100 0.4401 0.7300 0.9200

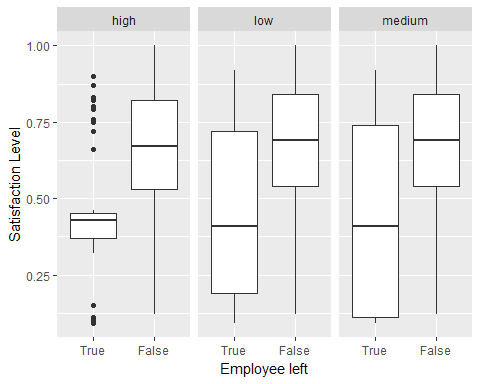
#As peedicted the satifaction level of employees who left was lower  
  
#Sstisfaction level vs left  
ggplot(aes(x = satisfaction\_level),data=hrm) +   
 geom\_histogram(color='black',fill='green',bins=35) +  
 xlab('Satisfaction Level') +   
 ylab("Frequency") +   
 facet\_wrap(~left)



#Boxplot for Satisfaction level vs left  
ggplot(aes(x = left,y=satisfaction\_level),data= hrm) +   
 geom\_boxplot() +   
 ylab('Satisfaction Level') +   
 xlab("Employee left") +   
 labs(fill="Salary Classes")



#Boxplot for Satisfaction level vs left facetted by Salary Ranges  
ggplot(aes(x = left,y=satisfaction\_level),data= hrm) +   
 geom\_boxplot() +   
 ylab('Satisfaction Level') +   
 xlab("Employee left") +   
 facet\_wrap(~salary)



table(hrm$left , salary)

## salary  
## high low medium  
## True 82 2172 1317  
## False 1155 5144 5129

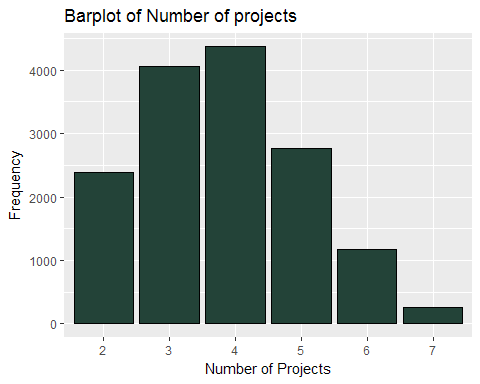
#Testing for the dependence between left and salary Ranges  
#Both are categorial variables so we use Chisq Test statistic  
chisq.test(left,salary)

##   
## Pearson's Chi-squared test  
##   
## data: left and salary  
## X-squared = 381.23, df = 2, p-value < 2.2e-16

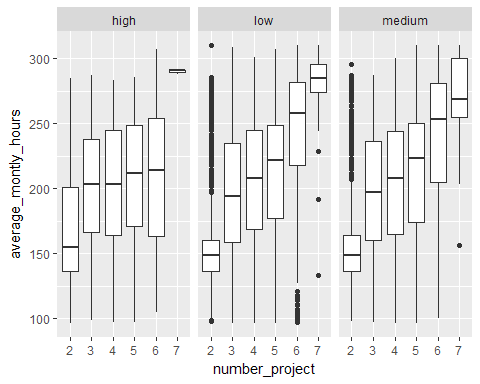
### X-squared value is high and p-value is less i.e results are significant.Both variables are related

## Analysis on number of Projects

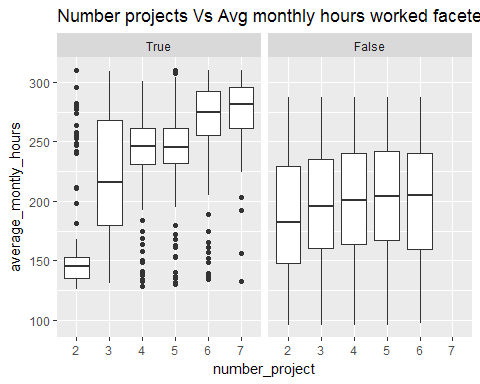
hrm$number\_project<-factor(hrm$number\_project)  
  
ggplot(aes(x=number\_project),data = hrm) +   
 geom\_bar(color='black',fill='#234338') +  
 xlab("Number of Projects") +   
 ylab("Frequency") +   
 labs(title="Barplot of Number of projects")



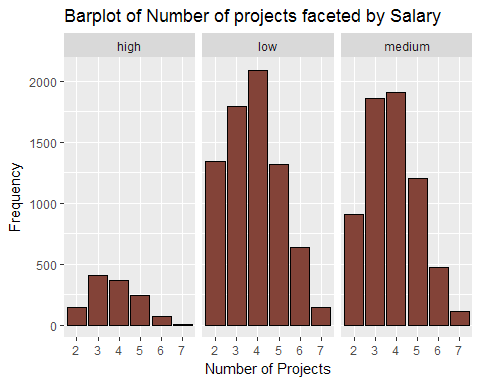
#boxplot of number of projects vs Average monthly hours at workplace of employees  
p3=ggplot(aes(x=number\_project, y = average\_montly\_hours),data=hrm)+  
 geom\_boxplot()  
  
p4=p3+facet\_wrap(~salary)  
p4



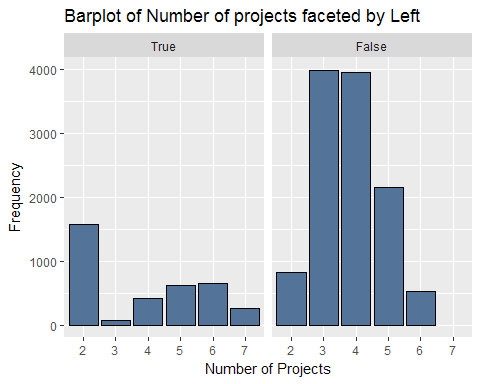
p5=p3+facet\_wrap(~left) + labs(title="Number projects Vs Avg monthly hours worked faceted by Left")  
p5



#facetted by salary  
ggplot(aes(x=number\_project),data = hrm) +   
 geom\_bar(color='black',fill='#834338') +  
 xlab("Number of Projects") +   
 ylab("Frequency") +   
 labs(title="Barplot of Number of projects faceted by Salary") +  
 facet\_wrap(~salary)



#faceted by If a employee left or not  
ggplot(aes(x=number\_project),data = hrm) +   
 geom\_bar(color='black',fill='#547398') +  
 xlab("Number of Projects") +   
 ylab("Frequency") +   
 labs(title="Barplot of Number of projects faceted by Left")+   
 facet\_wrap(~left)

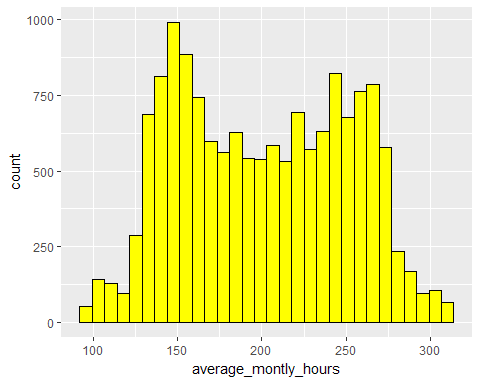


## Analysis on Average Number of Hours a Employee works

#Analysis of average monthly hours  
summary(average\_montly\_hours)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 96.0 156.0 200.0 201.1 245.0 310.0

#Somewhat Normally distributed  
  
ggplot(aes(x= average\_montly\_hours),data = hrm)+  
 geom\_histogram(color='black',fill="yellow",bins = 30)

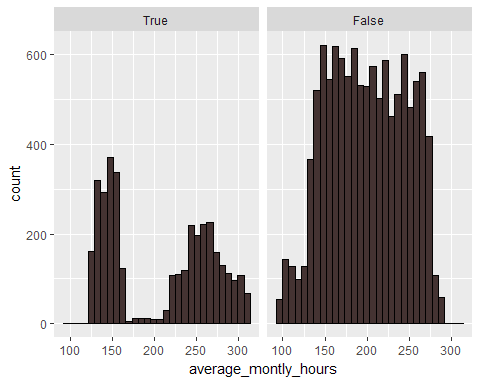


cor.test(satisfaction\_level,average\_montly\_hours)

##   
## Pearson's product-moment correlation  
##   
## data: satisfaction\_level and average\_montly\_hours  
## t = -2.4556, df = 14997, p-value = 0.01408  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.036040356 -0.004045605  
## sample estimates:  
## cor   
## -0.02004811

#No relation between both the variables - as r is eqv to 0

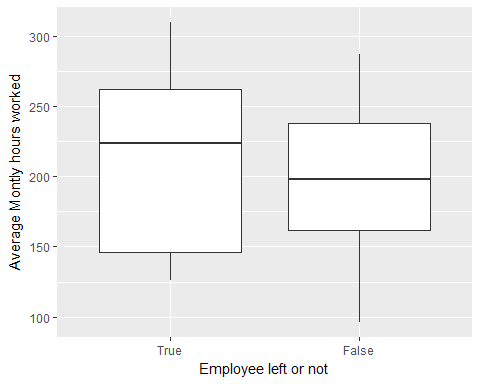
ggplot(aes(x = average\_montly\_hours),data =hrm ) +   
 geom\_histogram(color='black',fill='#443332',bins = 30) +   
 facet\_wrap(~left)



by(average\_montly\_hours , hrm$left ,summary)

## hrm$left: True  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 126.0 146.0 224.0 207.4 262.0 310.0   
## --------------------------------------------------------   
## hrm$left: False  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 96.0 162.0 198.0 199.1 238.0 287.0

ggplot(aes(y = average\_montly\_hours, x = hrm$left),data=hrm)+  
 geom\_boxplot() +   
 xlab("Employee left or not") +   
 ylab("Average Montly hours worked")



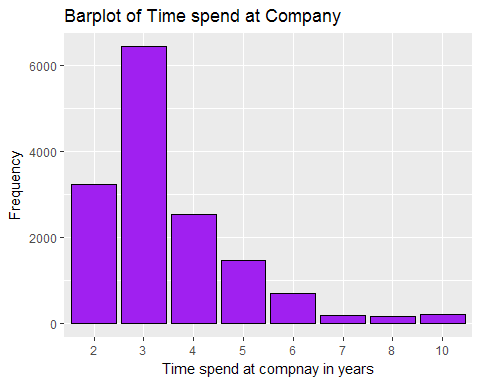
### A thing to notice is that employee who left the company worked more hours than those who did not leave, hence it might be possible that they left bacause they were over pressurized by their peers or bosses or over worked or stressed with lots of work

## Anslysis for variable Time spend at company

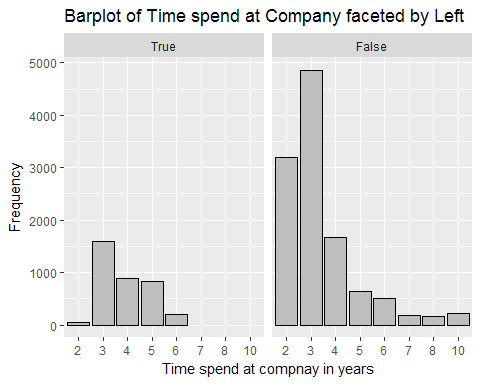
table(hrm$time\_spend\_company)

##   
## 2 3 4 5 6 7 8 10   
## 3244 6443 2557 1473 718 188 162 214

ggplot(aes(x = factor(time\_spend\_company)),data = hrm) +   
 geom\_bar(fill = 'purple',color='black') +   
 xlab("Time spend at compnay in years") +   
 ylab("Frequency")+  
 labs(title = "Barplot of Time spend at Company")



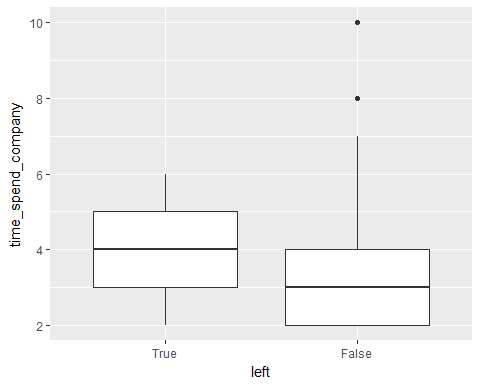
#Time spend at company vs Left or not  
  
ggplot(aes(x = factor(time\_spend\_company)),data = hrm) +   
 geom\_bar(fill = 'grey',color='black') +   
 xlab("Time spend at compnay in years") +   
 ylab("Frequency")+  
 labs(title = "Barplot of Time spend at Company faceted by Left") +  
 facet\_wrap(~left)



by(time\_spend\_company , left , summary)

## left: 0  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.00 2.00 3.00 3.38 4.00 10.00   
## --------------------------------------------------------   
## left: 1  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 4.000 3.877 5.000 6.000

ggplot(aes(x = left , y = time\_spend\_company),data = hrm)+  
 geom\_boxplot()



### Time Spend at company vs Satisfaaction level

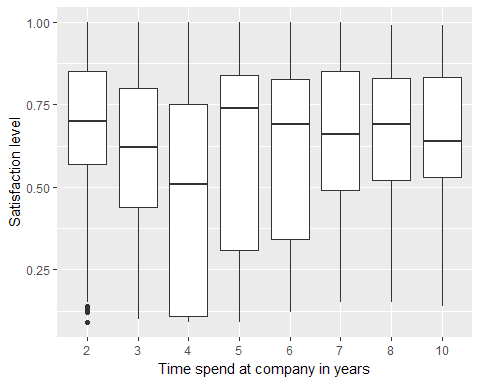
by(satisfaction\_level,factor(time\_spend\_company),summary)

## factor(time\_spend\_company): 2  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0900 0.5700 0.7000 0.6971 0.8500 1.0000   
## --------------------------------------------------------   
## factor(time\_spend\_company): 3  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.1000 0.4400 0.6200 0.6263 0.8000 1.0000   
## --------------------------------------------------------   
## factor(time\_spend\_company): 4  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0900 0.1100 0.5100 0.4675 0.7500 1.0000   
## --------------------------------------------------------   
## factor(time\_spend\_company): 5  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0900 0.3100 0.7400 0.6103 0.8400 1.0000   
## --------------------------------------------------------   
## factor(time\_spend\_company): 6  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.1200 0.3425 0.6900 0.6034 0.8275 1.0000   
## --------------------------------------------------------   
## factor(time\_spend\_company): 7  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.150 0.490 0.660 0.636 0.850 1.000   
## --------------------------------------------------------   
## factor(time\_spend\_company): 8  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.1500 0.5200 0.6900 0.6651 0.8300 0.9900   
## --------------------------------------------------------   
## factor(time\_spend\_company): 10  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.1400 0.5300 0.6400 0.6553 0.8325 0.9900

cor.test(satisfaction\_level,time\_spend\_company)

##   
## Pearson's product-moment correlation  
##   
## data: satisfaction\_level and time\_spend\_company  
## t = -12.416, df = 14997, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.11668153 -0.08499948  
## sample estimates:  
## cor   
## -0.1008661

#both have a negetive correlation  
  
#plots vs Time spend and Satisfaction level  
ggplot(aes(x=factor(time\_spend\_company),y=satisfaction\_level),data=hrm)+  
 geom\_boxplot() +   
 xlab("Time spend at company in years")+   
 ylab("Satisfaction level")

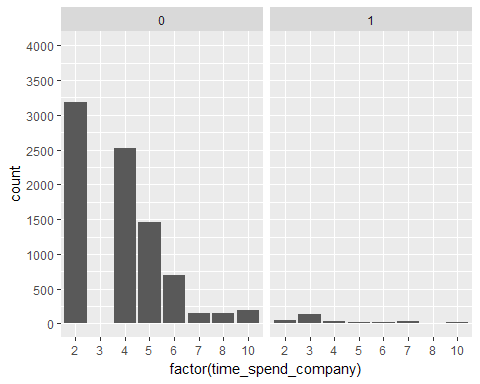


#Time spend at compnay vs Promotion in last 5 years  
  
table(Promotion=promotion\_last\_5years,Time\_Spend=factor(time\_spend\_company))

## Time\_Spend  
## Promotion 2 3 4 5 6 7 8 10  
## 0 3190 6309 2522 1456 701 152 152 198  
## 1 54 134 35 17 17 36 10 16

#Employees who have had promotion are very less  
  
  
ggplot(aes(x = factor(time\_spend\_company)),data = hrm)+  
 geom\_bar()+  
 facet\_wrap(~promotion\_last\_5years) +   
 scale\_y\_continuous(limits=c(0,4000),breaks=seq(0,4000,500))

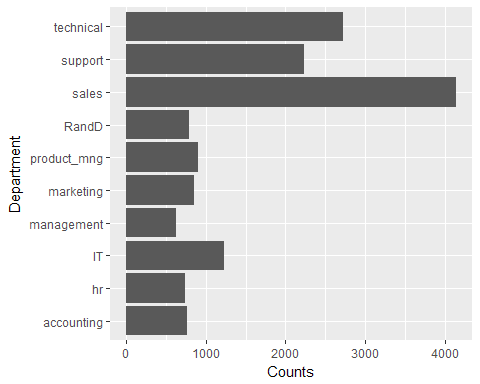
## Warning: Removed 1 rows containing missing values (geom\_bar).



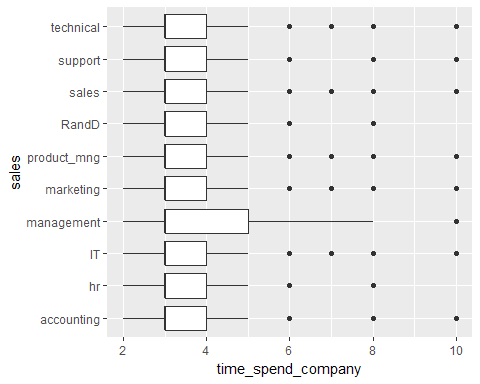
#Time spend vs Department of Work  
  
by(time\_spend\_company,sales,summary)

## sales: accounting  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 3.000 3.523 4.000 10.000   
## --------------------------------------------------------   
## sales: hr  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 3.000 3.356 4.000 8.000   
## --------------------------------------------------------   
## sales: IT  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 3.000 3.469 4.000 10.000   
## --------------------------------------------------------   
## sales: management  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 3.000 4.303 5.000 10.000   
## --------------------------------------------------------   
## sales: marketing  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.00 3.00 3.00 3.57 4.00 10.00   
## --------------------------------------------------------   
## sales: product\_mng  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 3.000 3.476 4.000 10.000   
## --------------------------------------------------------   
## sales: RandD  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 3.000 3.367 4.000 8.000   
## --------------------------------------------------------   
## sales: sales  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 3.000 3.534 4.000 10.000   
## --------------------------------------------------------   
## sales: support  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 3.000 3.393 4.000 10.000   
## --------------------------------------------------------   
## sales: technical  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 3.000 3.411 4.000 10.000

ggplot(aes(x =sales),data = hrm ) +  
 geom\_bar() +  
 xlab('Department') +   
 ylab('Counts') +  
 coord\_flip()

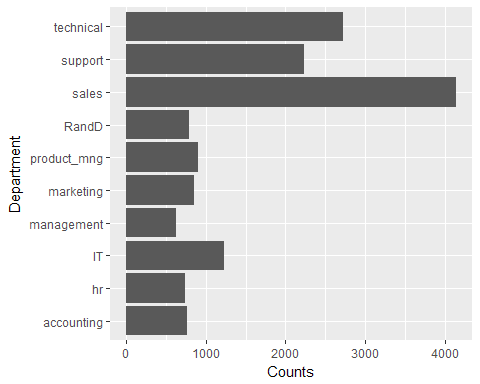


#highest count is for Sales department then Technical and least for   
#Management  
  
ggplot(aes(x = sales,y = time\_spend\_company),data = hrm) +   
 geom\_boxplot() +   
 coord\_flip()



## Analysis of Department of Work

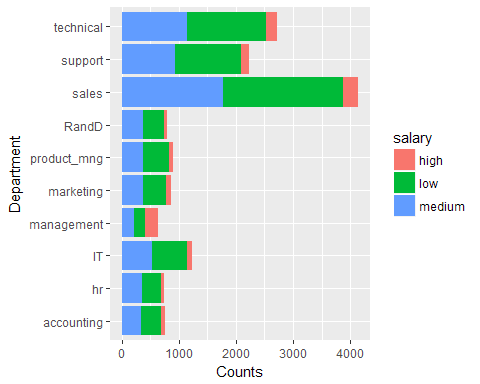
ggplot(aes(x =sales),data = hrm ) +  
 geom\_bar() +  
 xlab('Department') +   
 ylab('Counts') +  
 coord\_flip()



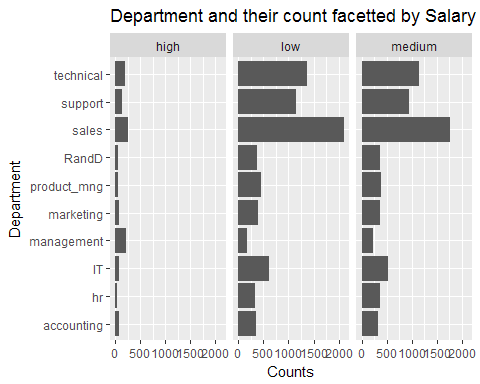
#highest count is for Sales department then Technical and least for   
#Management  
  
  
#Department vs sallary  
  
table(Dept = sales , Salary = salary)

## Salary  
## Dept high low medium  
## accounting 74 358 335  
## hr 45 335 359  
## IT 83 609 535  
## management 225 180 225  
## marketing 80 402 376  
## product\_mng 68 451 383  
## RandD 51 364 372  
## sales 269 2099 1772  
## support 141 1146 942  
## technical 201 1372 1147

ggplot(aes(x =sales),data = hrm ) +  
 geom\_bar(aes(fill=salary)) +  
 xlab('Department') +   
 ylab('Counts') +  
 coord\_flip()



ggplot(aes(x =sales),data = hrm ) +  
 geom\_bar() +  
 xlab('Department') +   
 ylab('Counts') +  
 labs(title = "Department and their count facetted by Salary ranges")+  
 facet\_wrap(~salary) +   
 coord\_flip()



chisq.test(sales,salary)

##   
## Pearson's Chi-squared test  
##   
## data: sales and salary  
## X-squared = 700.92, df = 18, p-value < 2.2e-16

#Department and Salary is dependent on each other .   
  
  
  
  
#finding proportions  
prop.table(table(Dept = sales , left = left))\*100

## left  
## Dept 0 1  
## accounting 3.7535836 1.3600907  
## hr 3.4935662 1.4334289  
## IT 6.3604240 1.8201213  
## management 3.5935729 0.6067071  
## marketing 4.3669578 1.3534236  
## product\_mng 4.6936462 1.3200880  
## RandD 4.4402960 0.8067204  
## sales 20.8413894 6.7604507  
## support 11.1607440 3.7002467  
## technical 13.4875658 4.6469765

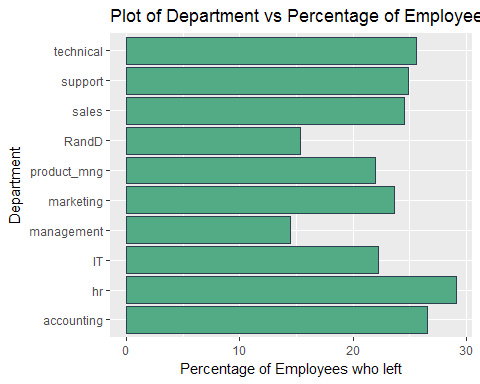
as.data.frame(table(sales , left))->deptdf  
deptdf

## sales left Freq  
## 1 accounting 0 563  
## 2 hr 0 524  
## 3 IT 0 954  
## 4 management 0 539  
## 5 marketing 0 655  
## 6 product\_mng 0 704  
## 7 RandD 0 666  
## 8 sales 0 3126  
## 9 support 0 1674  
## 10 technical 0 2023  
## 11 accounting 1 204  
## 12 hr 1 215  
## 13 IT 1 273  
## 14 management 1 91  
## 15 marketing 1 203  
## 16 product\_mng 1 198  
## 17 RandD 1 121  
## 18 sales 1 1014  
## 19 support 1 555  
## 20 technical 1 697

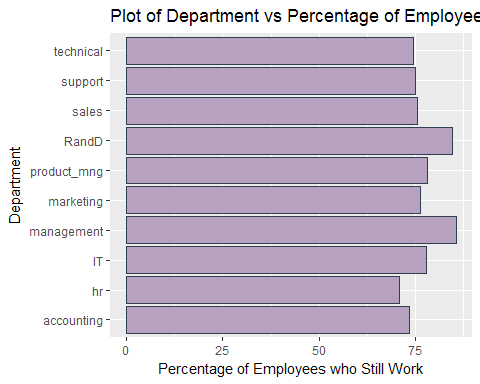
deptdf<-hrm %>% group\_by(sales,left) %>%   
 summarise(count=n())  
  
#making a data frame of Departments and the count of workers who left or not  
deptdf<-spread(deptdf,left,count)  
  
deptdf<-transform(deptdf,Perleft=(True/(True+False))\*100 , PerWork=(False/(True+False))\*100)  
deptdf

## sales True False Perleft PerWork  
## 1 accounting 204 563 26.59713 73.40287  
## 2 hr 215 524 29.09337 70.90663  
## 3 IT 273 954 22.24939 77.75061  
## 4 management 91 539 14.44444 85.55556  
## 5 marketing 203 655 23.65967 76.34033  
## 6 product\_mng 198 704 21.95122 78.04878  
## 7 RandD 121 666 15.37484 84.62516  
## 8 sales 1014 3126 24.49275 75.50725  
## 9 support 555 1674 24.89906 75.10094  
## 10 technical 697 2023 25.62500 74.37500

#Plot of Department vs Percentage of Employees who left  
ggplot(aes(x=sales, y = Perleft),data = deptdf) +   
 geom\_col(fill='#53ab85',color='#2f3f52') +   
 coord\_flip()+  
 xlab("Department") +   
 ylab("Percentage of Employees who left") +   
 labs(title="Plot of Department vs Percentage of Employee left")



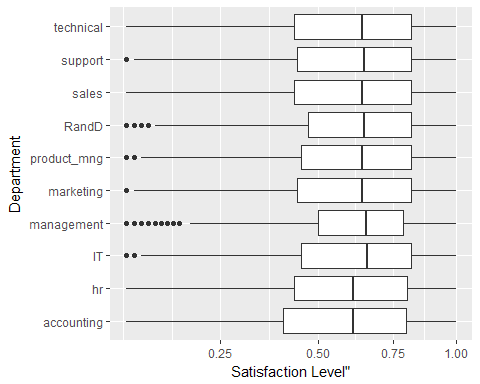
#highest percentage of employees belonged to HR dept then accounting  
# least for management dept who left  
  
  
  
  
  
#Plot of Department vs Percentage of People Working  
ggplot(aes(x=sales, y = PerWork),data = deptdf) +   
 geom\_col(fill='#b6a2bf',color='#2f3f52') +   
 coord\_flip()+  
 xlab("Department") +   
 ylab("Percentage of Employees who Still Work") +   
 labs(title="Plot of Department vs Percentage of Employees Working")



#Department vs Satisfaction level  
  
by(satisfaction\_level,sales,summary)

## sales: accounting  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0900 0.4000 0.6100 0.5822 0.8000 1.0000   
## --------------------------------------------------------   
## sales: hr  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0900 0.4300 0.6100 0.5988 0.8050 1.0000   
## --------------------------------------------------------   
## sales: IT  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0900 0.4500 0.6600 0.6181 0.8200 1.0000   
## --------------------------------------------------------   
## sales: management  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0900 0.5000 0.6550 0.6213 0.7900 1.0000   
## --------------------------------------------------------   
## sales: marketing  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0900 0.4400 0.6400 0.6186 0.8200 1.0000   
## --------------------------------------------------------   
## sales: product\_mng  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0900 0.4500 0.6400 0.6196 0.8200 1.0000   
## --------------------------------------------------------   
## sales: RandD  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0900 0.4700 0.6500 0.6198 0.8200 1.0000   
## --------------------------------------------------------   
## sales: sales  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0900 0.4300 0.6400 0.6144 0.8200 1.0000   
## --------------------------------------------------------   
## sales: support  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0900 0.4400 0.6500 0.6183 0.8200 1.0000   
## --------------------------------------------------------   
## sales: technical  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0900 0.4300 0.6400 0.6079 0.8200 1.0000

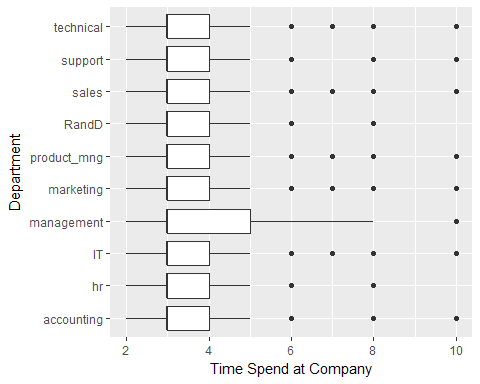
#highest mean satisfaction for R&D and Management Dept  
  
ggplot(aes(x = sales, y = satisfaction\_level),data = hrm)+  
 geom\_boxplot() +   
 scale\_y\_sqrt()+  
 xlab('Department') +   
 ylab('Satisfaction Level"') +   
 coord\_flip()



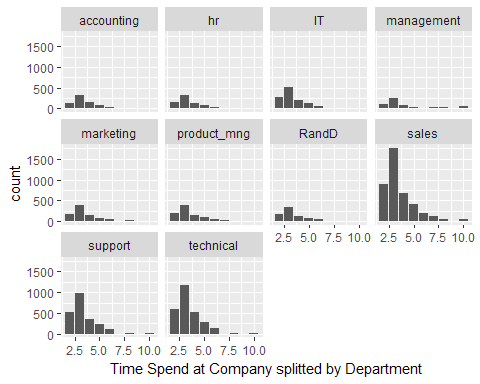
#Highest Median Satisfaction for IT dept, R&D and , Management  
#Least Median Satifaction level for HR and Accounting  
  
  
  
  
  
  
  
#Analysis of Department vs Time spend at company  
  
by(time\_spend\_company,sales,summary)

## sales: accounting  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 3.000 3.523 4.000 10.000   
## --------------------------------------------------------   
## sales: hr  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 3.000 3.356 4.000 8.000   
## --------------------------------------------------------   
## sales: IT  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 3.000 3.469 4.000 10.000   
## --------------------------------------------------------   
## sales: management  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 3.000 4.303 5.000 10.000   
## --------------------------------------------------------   
## sales: marketing  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.00 3.00 3.00 3.57 4.00 10.00   
## --------------------------------------------------------   
## sales: product\_mng  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 3.000 3.476 4.000 10.000   
## --------------------------------------------------------   
## sales: RandD  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 3.000 3.367 4.000 8.000   
## --------------------------------------------------------   
## sales: sales  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 3.000 3.534 4.000 10.000   
## --------------------------------------------------------   
## sales: support  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 3.000 3.393 4.000 10.000   
## --------------------------------------------------------   
## sales: technical  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 3.000 3.411 4.000 10.000

#Maximum Mean Time spent by Managaement Employees  
  
ggplot(aes(x = sales,y = time\_spend\_company),data = hrm) +   
 geom\_boxplot() +  
 xlab('Department') +   
 ylab("Time Spend at Company") +   
 coord\_flip()



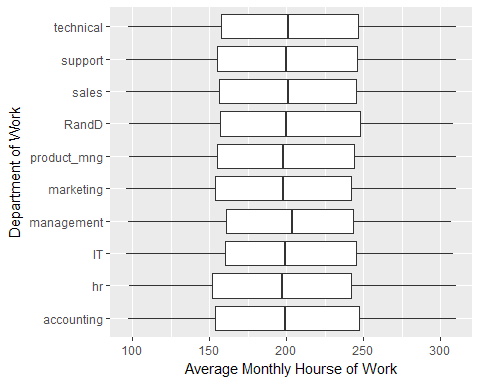
ggplot(aes(x = time\_spend\_company),data = hrm) +   
 geom\_bar() +  
 xlab("Time Spend at Company splitted by Department") +   
facet\_wrap(~sales)



#In every department there is very less count of Employees  
# working for over 5 years  
  
  
  
  
  
  
#Department vs Time average monthly hours  
  
by(average\_montly\_hours,sales , summary)

## sales: accounting  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 97.0 153.5 199.0 201.2 247.0 310.0   
## --------------------------------------------------------   
## sales: hr  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 98.0 152.0 197.0 198.7 242.0 310.0   
## --------------------------------------------------------   
## sales: IT  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 96.0 160.0 199.0 202.2 245.0 308.0   
## --------------------------------------------------------   
## sales: management  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 97.0 161.0 204.0 201.2 243.0 307.0   
## --------------------------------------------------------   
## sales: marketing  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 96.0 154.0 198.0 199.4 242.0 310.0   
## --------------------------------------------------------   
## sales: product\_mng  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 98 155 198 200 244 310   
## --------------------------------------------------------   
## sales: RandD  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 98.0 157.0 200.0 200.8 248.0 308.0   
## --------------------------------------------------------   
## sales: sales  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 96.0 156.0 201.0 200.9 245.0 310.0   
## --------------------------------------------------------   
## sales: support  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 96.0 155.0 200.0 200.8 246.0 310.0   
## --------------------------------------------------------   
## sales: technical  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 97.0 157.8 201.0 202.5 246.2 310.0

#Highest average working time for IT and Technical departments  
  
ggplot(aes(x = sales , y = average\_montly\_hours),data =hrm) +  
 geom\_boxplot() +   
 xlab('Department of Work') +   
 ylab('Average Monthly Hourse of Work') +   
 coord\_flip()



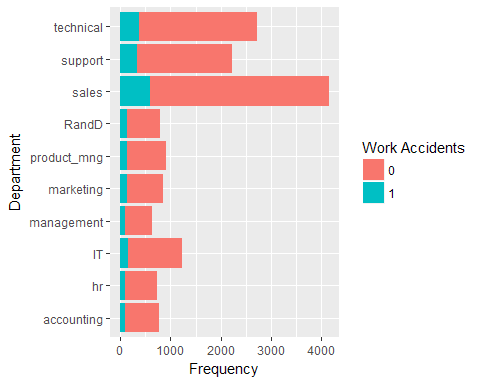
#Highest Median working time of Management department  
  
  
  
  
  
#Department vs Work Accident  
  
table(Work\_accident)

## Work\_accident  
## 0 1   
## 12830 2169

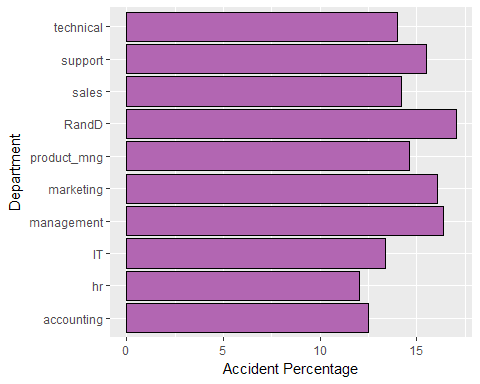
table(sales,Work\_accident)

## Work\_accident  
## sales 0 1  
## accounting 671 96  
## hr 650 89  
## IT 1063 164  
## management 527 103  
## marketing 720 138  
## product\_mng 770 132  
## RandD 653 134  
## sales 3553 587  
## support 1884 345  
## technical 2339 381

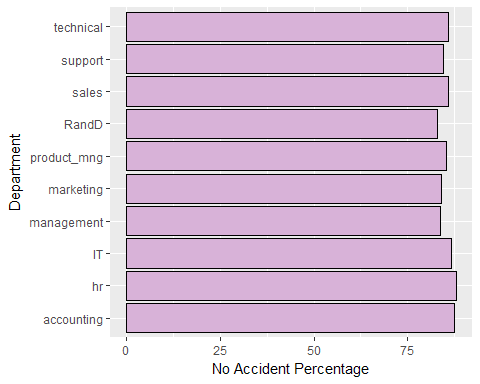
ggplot(aes(x = sales),data = hrm) +  
 geom\_bar(aes(fill=factor(Work\_accident))) +   
 coord\_flip() +   
 labs(x = "Department",y ="Frequency", fill="Work Accidents" )



hrm$Work\_accident<-factor(Work\_accident,labels = c('False','True'))  
  
accidentdf<-hrm %>% group\_by(sales,Work\_accident) %>%  
 summarise(Count= n())  
  
accidentdf<-spread(accidentdf,Work\_accident,Count)  
  
accidentdf<-transform(accidentdf,TrueRate=(True/(True+False))\*100,FalseRate=(False/(True+False))\*100)  
  
#Plot of Departent vs Accidental Rate   
ggplot(aes(x = sales,y = TrueRate),data = accidentdf) +   
 geom\_col(color='black',fill="#b266b2") +   
 xlab('Department') +   
 ylab('Accident Percentage') +   
 coord\_flip()



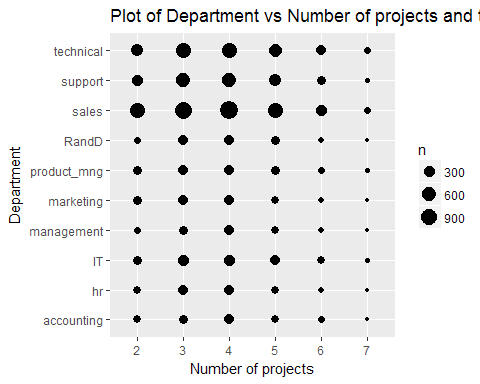
#Highest number of accidents in R and D department  
  
  
  
ggplot(aes(x = sales,y = FalseRate),data = accidentdf) +   
 geom\_col(color='black',fill="#d8b2d8") +   
 xlab('Department') +   
 ylab('No Accident Percentage') +   
 coord\_flip()



#Maximum for HR department  
  
  
  
  
  
#Department vs number\_projects made  
  
by(number\_project,sales,summary)

## sales: accounting  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 4.000 3.825 5.000 7.000   
## --------------------------------------------------------   
## sales: hr  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 4.000 3.655 4.000 7.000   
## --------------------------------------------------------   
## sales: IT  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 4.000 3.817 5.000 7.000   
## --------------------------------------------------------   
## sales: management  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.00 3.00 4.00 3.86 5.00 7.00   
## --------------------------------------------------------   
## sales: marketing  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 4.000 3.688 4.000 7.000   
## --------------------------------------------------------   
## sales: product\_mng  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 4.000 3.807 5.000 7.000   
## --------------------------------------------------------   
## sales: RandD  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 4.000 3.854 5.000 7.000   
## --------------------------------------------------------   
## sales: sales  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 4.000 3.776 5.000 7.000   
## --------------------------------------------------------   
## sales: support  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 4.000 3.804 5.000 7.000   
## --------------------------------------------------------   
## sales: technical  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 4.000 3.878 5.000 7.000

ggplot(aes(x = sales, y =factor(number\_project)),data = hrm) +  
 geom\_count() +  
 xlab("Department") +   
 ylab("Number of projects") +   
 labs(title = "Plot of Department vs Number of projects and their count ") +   
 coord\_flip()



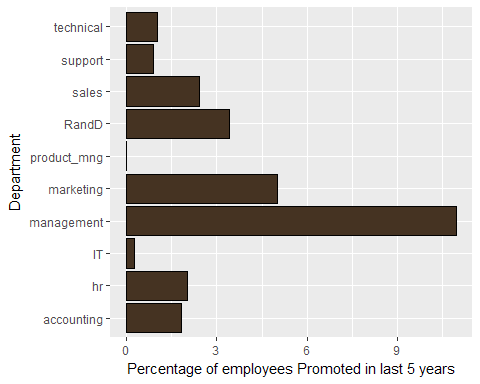
#Department vs Promotion in last 5 years  
  
table(sales , hrm$promotion\_last\_5years)

##   
## sales 0 1  
## accounting 753 14  
## hr 724 15  
## IT 1224 3  
## management 561 69  
## marketing 815 43  
## product\_mng 902 0  
## RandD 760 27  
## sales 4040 100  
## support 2209 20  
## technical 2692 28

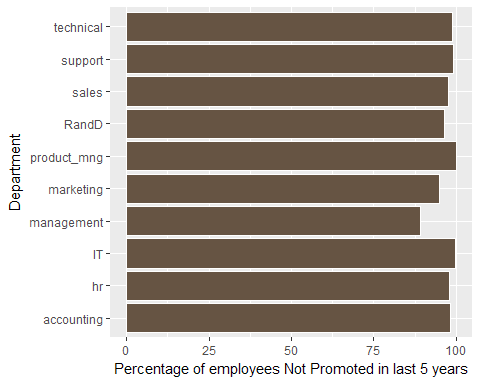
#Transforming Promotion Column to Factor with True and False values  
hrm$promotion\_last\_5years<-factor(promotion\_last\_5years,labels=c('False',"True"))  
  
#Generating a promotions Data frame  
promotiondf<-hrm %>% group\_by(sales,promotion\_last\_5years) %>%  
 summarise(Count = n())  
  
#Spreading the data  
promotiondf<-promotiondf %>% spread(promotion\_last\_5years,Count)  
  
#changing the names of columns  
names(promotiondf)<-c("Department","Nopromotion","Promotion")  
  
#replacing NA value with 0  
promotiondf[is.na(promotiondf)]<-0  
  
promotiondf<-promotiondf %>% transform(PerPromotion=(Promotion/(Promotion+Nopromotion))\*100,  
 PerNopromotion = (Nopromotion/(Promotion + Nopromotion))\*100)  
#Most number of Promotions done in Management and Marketing Departments  
#Least in IT , Technical and Product Manager  
promotiondf

## Department Nopromotion Promotion PerPromotion PerNopromotion  
## 1 accounting 753 14 1.8252934 98.17471  
## 2 hr 724 15 2.0297700 97.97023  
## 3 IT 1224 3 0.2444988 99.75550  
## 4 management 561 69 10.9523810 89.04762  
## 5 marketing 815 43 5.0116550 94.98834  
## 6 product\_mng 902 0 0.0000000 100.00000  
## 7 RandD 760 27 3.4307497 96.56925  
## 8 sales 4040 100 2.4154589 97.58454  
## 9 support 2209 20 0.8972633 99.10274  
## 10 technical 2692 28 1.0294118 98.97059

#Plotting Department vs Promotion Percentage  
ggplot(aes(x =Department, y =PerPromotion ),data = promotiondf) +   
 geom\_col(color='black',fill = '#453322') +   
 xlab("Department") +   
 ylab("Percentage of employees Promoted in last 5 years") +   
 coord\_flip()



#Highest in Management Department  
  
#Plotting Department vs No Promotion Percentage  
ggplot(aes(x =Department, y =PerNopromotion ),data = promotiondf) +   
 geom\_col(color="white",fill = "#665443") +   
 xlab("Department") +   
 ylab("Percentage of employees Not Promoted in last 5 years") +   
 coord\_flip()



#No promotion in IT and Product Management Dept