**Year : 2020-2021**

**Data Processing using RStudio**

**Technical Internship Report at ECS Informatique Company**

**Made by : Ghita BOUZIAN**

**Supervised by : Ms. Karima Ammahmoud**

**Mr. Ziat Mohammed**

**Sommaire**

# Starter Code……………………………..3

# Women’s positions…………………………..5

# Recruitment source…………………………………6

# Female’s Percentage per Departments…………11

# Gender distribution………………………………..14

# Race Distribution…………………………………..17

# Citizenship distribution……………………………19

# Absences per Department………………………..20

# Starter Code:

**> Format1<-read.csv2(file.choose(), header=T, sep=",", quote="\"")**

**> head(Format1)**

**Execution Results:**

ï..Employee\_Name EmpID MarriedID MaritalStatusID

1 Adinolfi, Wilson K 10026 0 0

2 Ait Sidi, Karthikeyan 10084 1 1

3 Akinkuolie, Sarah 10196 1 1

4 Alagbe,Trina 10088 1 1

5 Anderson, Carol 10069 0 2

6 Anderson, Linda 10002 0 0

GenderID EmpStatusID DeptID PerfScoreID FromDiversityJobFairID

1 1 1 5 4 0

2 1 5 3 3 0

3 0 5 5 3 0

4 0 1 5 3 0

5 0 5 5 3 0

6 0 1 5 4 0

Salary Termd PositionID Position State Zip

1 62506 0 19 Production Technician I MA 1960

2 104437 1 27 Sr. DBA MA 2148

3 64955 1 20 Production Technician II MA 1810

4 64991 0 19 Production Technician I MA 1886

5 50825 1 19 Production Technician I MA 2169

6 57568 0 19 Production Technician I MA 1844

DOB Sex MaritalDesc CitizenDesc HispanicLatino RaceDesc

1 07/10/83 M Single US Citizen No White

2 05/05/75 M Married US Citizen No White

3 09/19/88 F Married US Citizen No White

4 09/27/88 F Married US Citizen No White

5 09/08/89 F Divorced US Citizen No White

6 05/22/77 F Single US Citizen No White

DateofHire DateofTermination TermReason

1 7/5/2011 N/A-StillEmployed

2 3/30/2015 6/16/2016 career change

3 7/5/2011 9/24/2012 hours

4 1/7/2008 N/A-StillEmployed

5 7/11/2011 9/6/2016 return to school

6 1/9/2012 N/A-StillEmployed

EmploymentStatus Department ManagerName

1 Active Production Michael Albert

2 Voluntarily Terminated IT/IS Simon Roup

3 Voluntarily Terminated Production Kissy Sullivan

4 Active Production Elijiah Gray

5 Voluntarily Terminated Production Webster Butler

6 Active Production Amy Dunn

ManagerID RecruitmentSource PerformanceScore EngagementSurvey

1 22 LinkedIn Exceeds 4.60

2 4 Indeed Fully Meets 4.96

3 20 LinkedIn Fully Meets 3.02

4 16 Indeed Fully Meets 4.84

5 39 Google Search Fully Meets 5.00

6 11 LinkedIn Exceeds 5.00

EmpSatisfaction SpecialProjectsCount

1 5 0

2 3 6

3 3 0

4 5 0

5 4 0

6 5 0

LastPerformanceReview\_Date DaysLateLast30 Absences

1 1/17/2019 0 1

2 2/24/2016 0 17

3 5/15/2012 0 3

4 1/3/2019 0 15

5 2/1/2016 0 2

6 1/7/2019 0 15

**> colnames(Format1)**

**Execution Results:**

[1] "ï..Employee\_Name" "EmpID"

[3] "MarriedID" "MaritalStatusID"

[5] "GenderID" "EmpStatusID"

[7] "DeptID" "PerfScoreID"

[9] "FromDiversityJobFairID" "Salary"

[11] "Termd" "PositionID"

[13] "Position" "State"

[15] "Zip" "DOB"

[17] "Sex" "MaritalDesc"

[19] "CitizenDesc" "HispanicLatino"

[21] "RaceDesc" "DateofHire"

[23] "DateofTermination" "TermReason"

[25] "EmploymentStatus" "Department"

[27] "ManagerName" "ManagerID"

[29] "RecruitmentSource" "PerformanceScore"

[31] "EngagementSurvey" "EmpSatisfaction"

[33] "SpecialProjectsCount" "LastPerformanceReview\_Date"

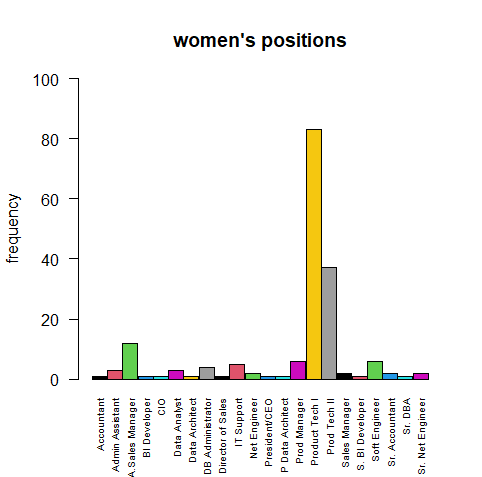
[35] "DaysLateLast30" "Absences"

# Women’s positions:

**> femposition<-table(Format1$Position[Format1$Sex=="F"])**

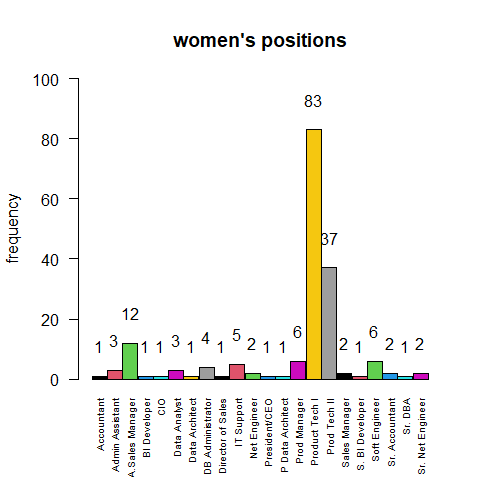
**> barplot(femposition, ylab="frequency", main="women's positions",col=c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22),names.arg=c("Accountant","Admin Assistant","A.Sales Manager","BI Developer","CIO","Data Analyst","Data Architect","DB Administrator","Director of Sales","IT Support","Net Engineer","President/CEO","P Data Architect","Prod Manager","Product Tech I","Prod Tech II","Sales Manager","S. BI Developer","Soft Engineer","Sr. Accountant","Sr. DBA","Sr. Net Engineer"),las=2,cex.names=0.6, space=0.02, ylim=c(0,100))**

**Execution Results:**



**>bar<-barplot(femposition, ylab="frequency", main="women's positions",col=c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22),names.arg=c("Accountant","Admin Assistant","A.Sales Manager","BI Developer","CIO","Data Analyst","Data Architect","DB Administrator","Director of Sales","IT Support","Net Engineer","President/CEO","P Data Architect","Prod Manager","Product Tech I","Prod Tech II","Sales Manager","S. BI Developer","Soft Engineer","Sr. Accountant","Sr. DBA","Sr. Net Engineer"),las=2,cex.names=0.6, space=0.02, ylim=c(0,100))**

**> for (i in 1:nrow(femposition)) text(bar[i], femposition[i]+10, labels=as.character(femposition[i]))**



# Recruitment source:

**>sourcederecrut<-table(Format1$RecruitmentSource)**

**>sourcederecrut**

**Execution Results:**

CareerBuilder Diversity Job Fair

23 29

Employee Referral Google Search

31 49

Indeed LinkedIn

87 76

On-line Web application Other

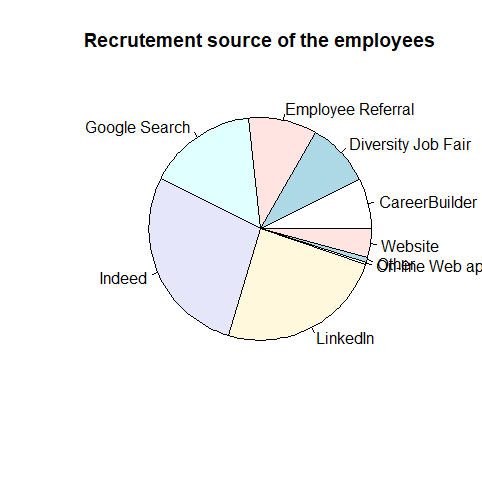
1 2

Website

13

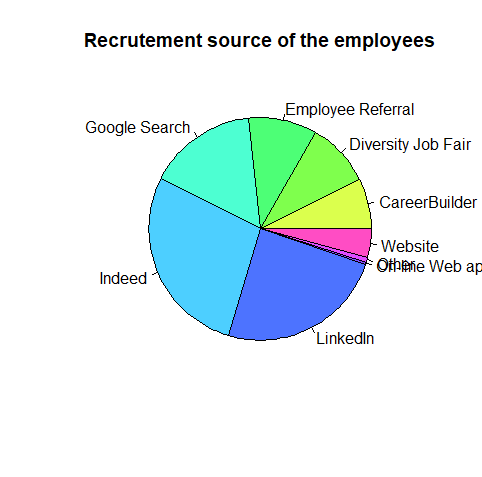
**> pie(sourcederecrut,main="Recrutement source of the employees")**

**Execution Results:**



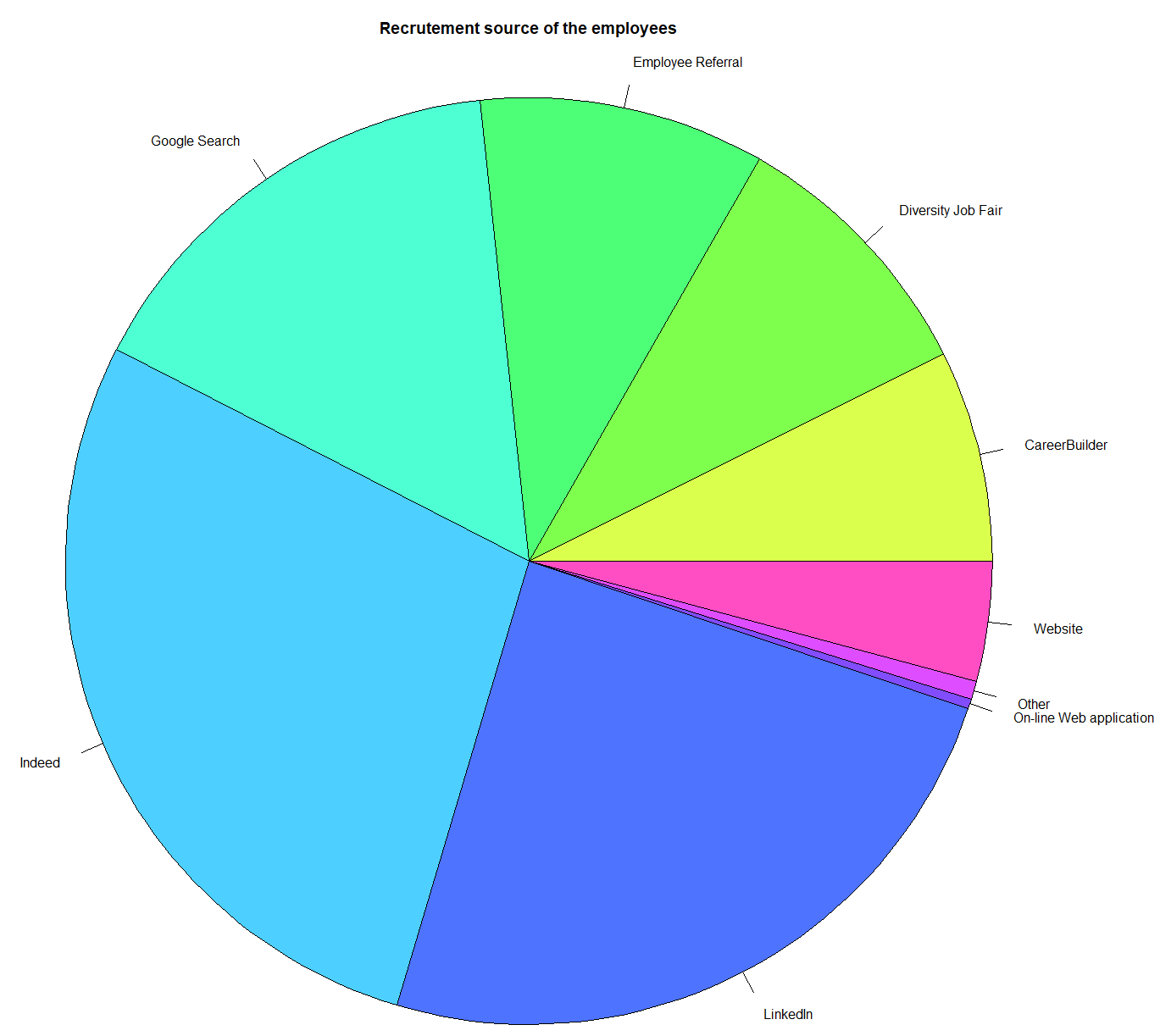
**>pie(sourcederecrut,main="Recrutement source of the employees", init.angle=0,clockwise=FALSE, col=rainbow(9, s=0.7, start=0.2))**

**Execution Results:**

****

**>pie(sourcederecrut,main="Recrutement source of the employees", init.angle=0,clockwise=FALSE, col=rainbow(9, s=0.7, start=0.2), radius=1)**

**Execution Results:**



**> total=dim(Format1)[1] #Total des valeurs**

**> valeurs=c() #Initialisées à liste vide**

**> for (i in 1:nrow(sourcederecrut)){valeurs[i]=sourcederecrut[i]} #Affectation des valeurs à la liste**

**> pourcentages = valeurs/total\*100 #Calcul des pourcentages**

**> percentages #Résultats des pourcentages obtenus que l’on peut arrondir**

**[1] 7.3954984 9.3247588**

**[3] 9.9678457 15.7556270**

**[5] 27.9742765 24.4372990**

**[7] 0.3215434 0.6430868**

**[9] 4.1800643**

**> text\_pie = function(vector,labels=c(),cex=1) { #Fonction pour écrire les pourcentages sur le**

**#Camembert**

**+ vector = vector/sum(vector)\*2\*pi**

**+ temp = c()**

**+ j = 0**

**+ i = 0**

**+ for (i in 1:length(vector)) {**

**+ k = vector[i]/2**

**+ j = j+i+k**

**+ i = k**

**+ text(cos(j)/2,sin(j)/2,labels[i],cex=cex)**

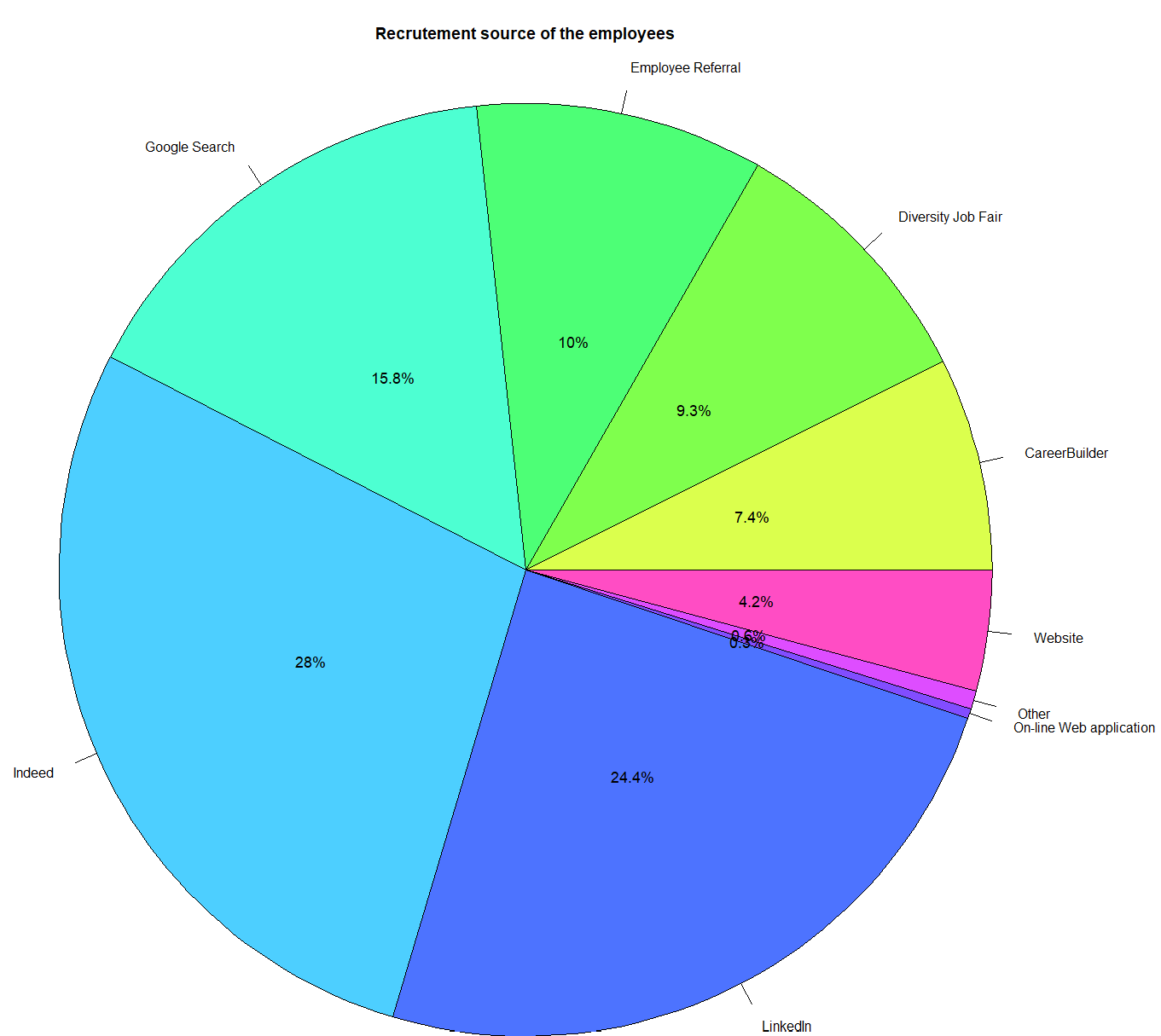
**+ }**

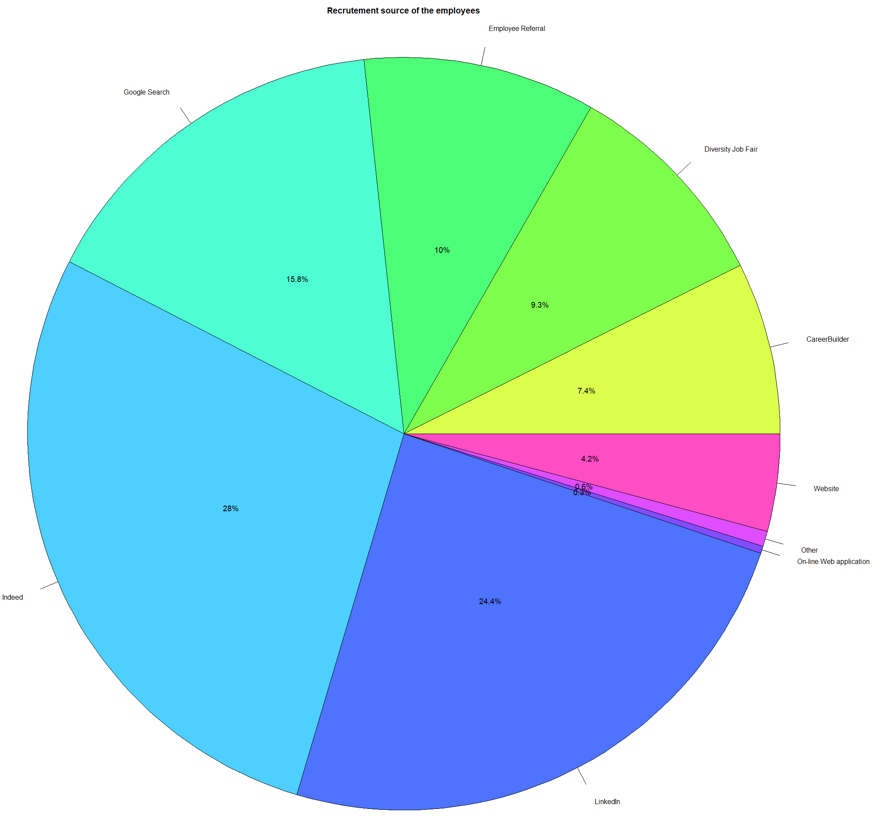
**+ vector = temp**

**+ }**

**> text\_pie(pourcentages,c("7.4%","9.3%","10%", "15.8%", "28%", "24.4%", "0.3%", "0.6%", "4.2%"),cex=1.1) #Appel de la fonction pour écrire les pourcentages arrondis**

**Execution Results:**



****

# Female’s Percentage per Departments:

**> dep<-table(Format1$Department)**

**> dep**

Admin Offices

9

Executive Office

1

IT/IS

50

Production

209

Sales

31

Software Engineering

11

**> femdep<-table(Format1$Department[Format1$Sex=="F"])**

**> femdep**

Admin Offices

6

Executive Office

1

IT/IS

22

Production

126

Sales

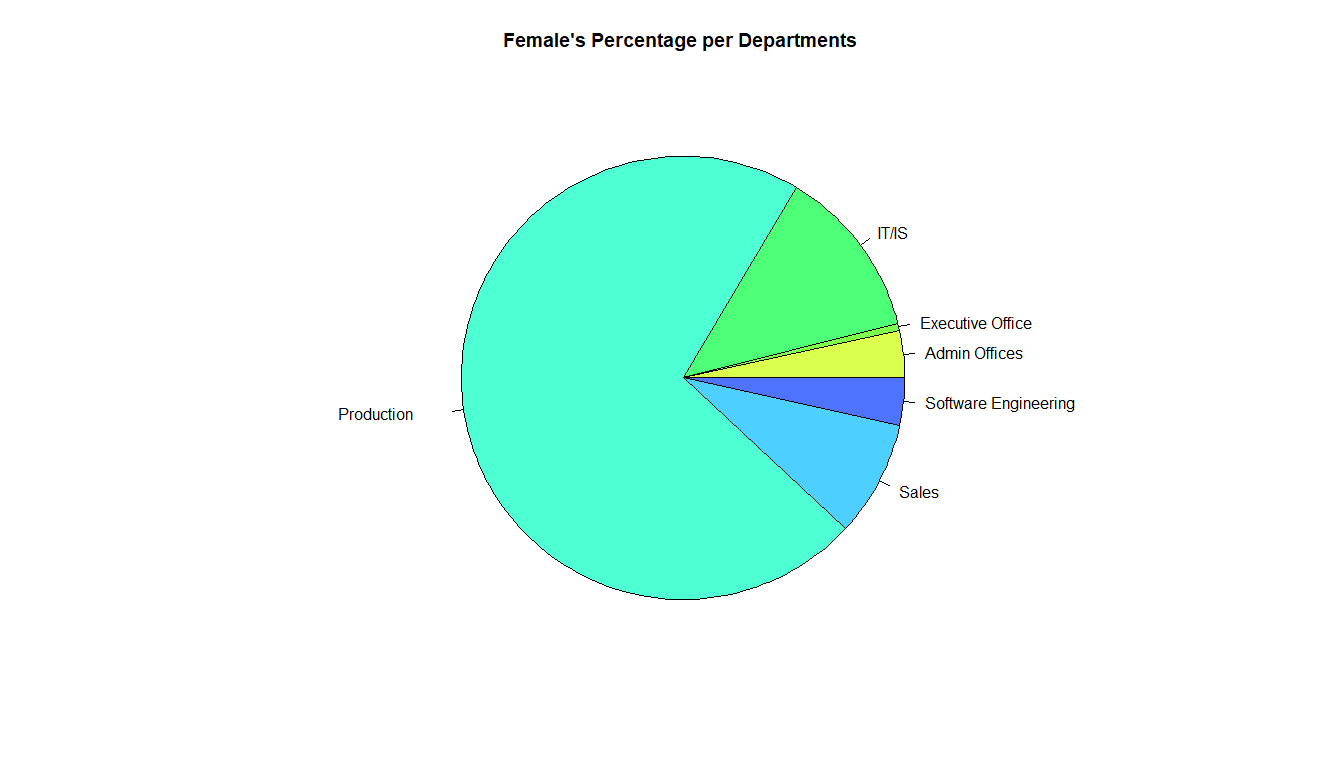
15

Software Engineering

6

**>pie(femdep,main="Female's Percentage per Departments ", init.angle=0,clockwise=FALSE, col=rainbow(9, s=0.7, start=0.2))**

**Execution Results:**

****

**> valeurs=c() #Initialisées à liste vide**

**> for (i in 1:nrow(femdep)){valeurs[i]=femdep[i]} #Affectation des valeurs à la liste**

**> valeurs**

**>** [1] 6 1 22 126 15 6

**> total=0**

**> for (i in 1:nrow(femdep)){total=total+femdep[i]} #Total**

**> total**

Admin Offices

176

**> pourcentages = valeurs/total\*100 #Calcul des pourcentages**

**> percentages**

[1] 3.4090909 0.5681818 12.5000000 71.5909091 8.5227273

[6] 3.4090909

**> text\_pie = function(vector,labels=c(),cex=1) { #Fonction pour écrire les pourcentages sur le**

**#Camembert**

**+ vector = vector/sum(vector)\*2\*pi**

**+ temp = c()**

**+ j = 0**

**+ i = 0**

**+ for (i in 1:length(vector)) {**

**+ k = vector[i]/2**

**+ j = j+i+k**

**+ i = k**

**+ text(cos(j)/2,sin(j)/2,labels[i],cex=cex)**

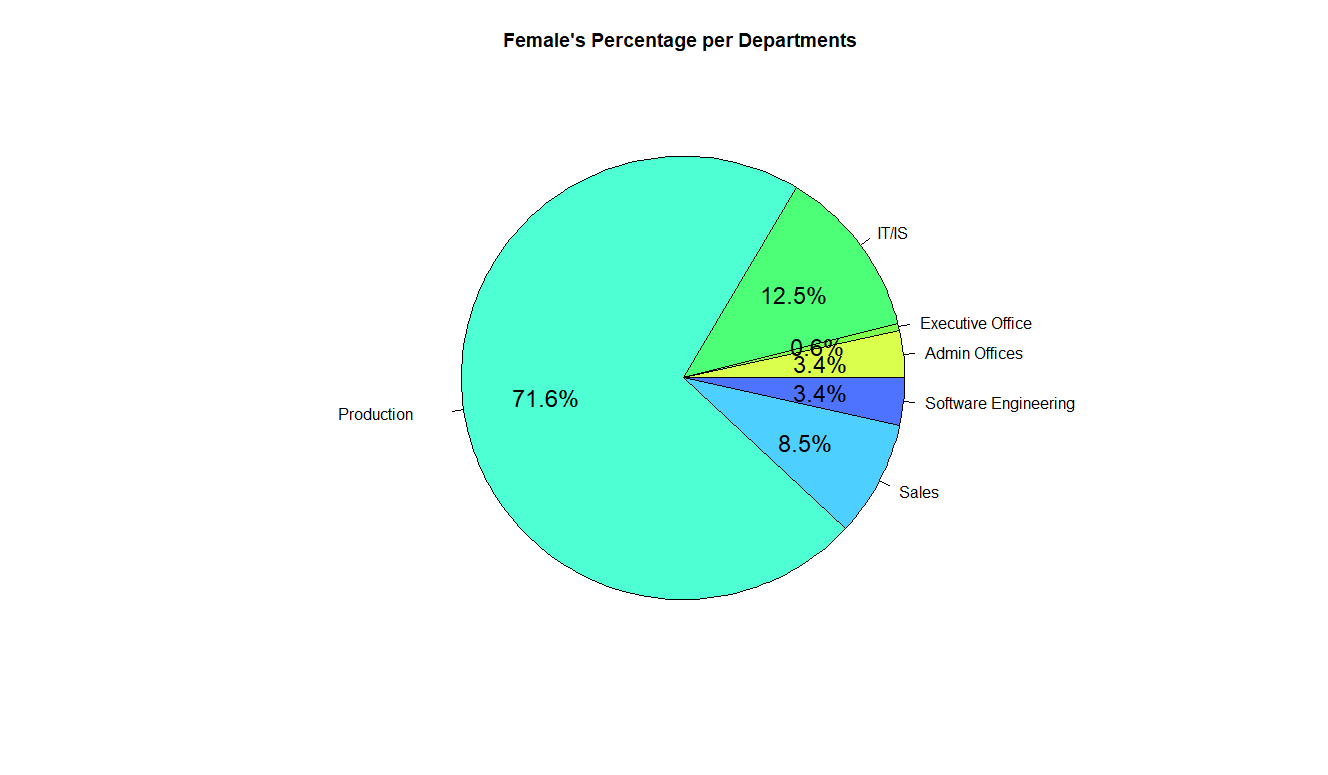
**+ }**

**+ vector = temp**

**+ }**

**> text\_pie(pourcentages,c("3.4%","0.6%","12.5%", "71.6%", "8.5%", "3.4%"),cex=1.5)**

**Execution Results:**



# Gender distribution

**> install.packages("lessR")**

package ‘mime’ successfully unpacked and MD5 sums checked

package ‘glue’ successfully unpacked and MD5 sums checked

package ‘magrittr’ successfully unpacked and MD5 sums checked

package ‘png’ successfully unpacked and MD5 sums checked

package ‘jpeg’ successfully unpacked and MD5 sums checked

package ‘RColorBrewer’ successfully unpacked and MD5 sums checked

package ‘DEoptimR’ successfully unpacked and MD5 sums checked

package ‘Rcpp’ successfully unpacked and MD5 sums checked

package ‘zip’ successfully unpacked and MD5 sums checked

package ‘stringi’ successfully unpacked and MD5 sums checked

package ‘evaluate’ successfully unpacked and MD5 sums checked

package ‘highr’ successfully unpacked and MD5 sums checked

package ‘markdown’ successfully unpacked and MD5 sums checked

package ‘stringr’ successfully unpacked and MD5 sums checked

package ‘yaml’ successfully unpacked and MD5 sums checked

package ‘xfun’ successfully unpacked and MD5 sums checked

package ‘latticeExtra’ successfully unpacked and MD5 sums checked

package ‘robustbase’ successfully unpacked and MD5 sums checked

package ‘ellipse’ successfully unpacked and MD5 sums checked

package ‘leaps’ successfully unpacked and MD5 sums checked

package ‘openxlsx’ successfully unpacked and MD5 sums checked

package ‘colorspace’ successfully unpacked and MD5 sums checked

package ‘knitr’ successfully unpacked and MD5 sums checked

package ‘viridisLite’ successfully unpacked and MD5 sums checked

package ‘lessR’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\fatim\AppData\Local\Temp\Rtmp8EvXcy\downloaded\_packages

**> library(lessR)**

lessR 4.0.0 feedback: gerbing@pdx.edu web: lessRstats.com/new

---------------------------------------------------------------

> d <- Read("") Read text, Excel, SPSS, SAS, or R data file

d is default data frame, data= in analysis routines optional

Learn about reading, writing, and manipulating data, graphics,

testing means and proportions, regression, factor analysis,

customization, and descriptive statistics from pivot tables.

Enter: browseVignettes("lessR")

View changes in this new version of lessR.

Enter: help(package=lessR) Click: Package NEWS

**> gender <- data.frame(gend = Format1$Sex)**

**> PieChart(gend, hole = 0, values = "%", data = gender,**

**+ fill = c(“pink”,"lightblue"), main = "Gender distribution", cex=2)**

>>> Suggestions

PieChart(gend, hole=0) # traditional pie chart

PieChart(gend, values="%") # display %'s on the chart

BarChart(gend) # bar chart

Plot(gend) # bubble plot

Plot(gend, values="count") # lollipop plot

--- gend ---

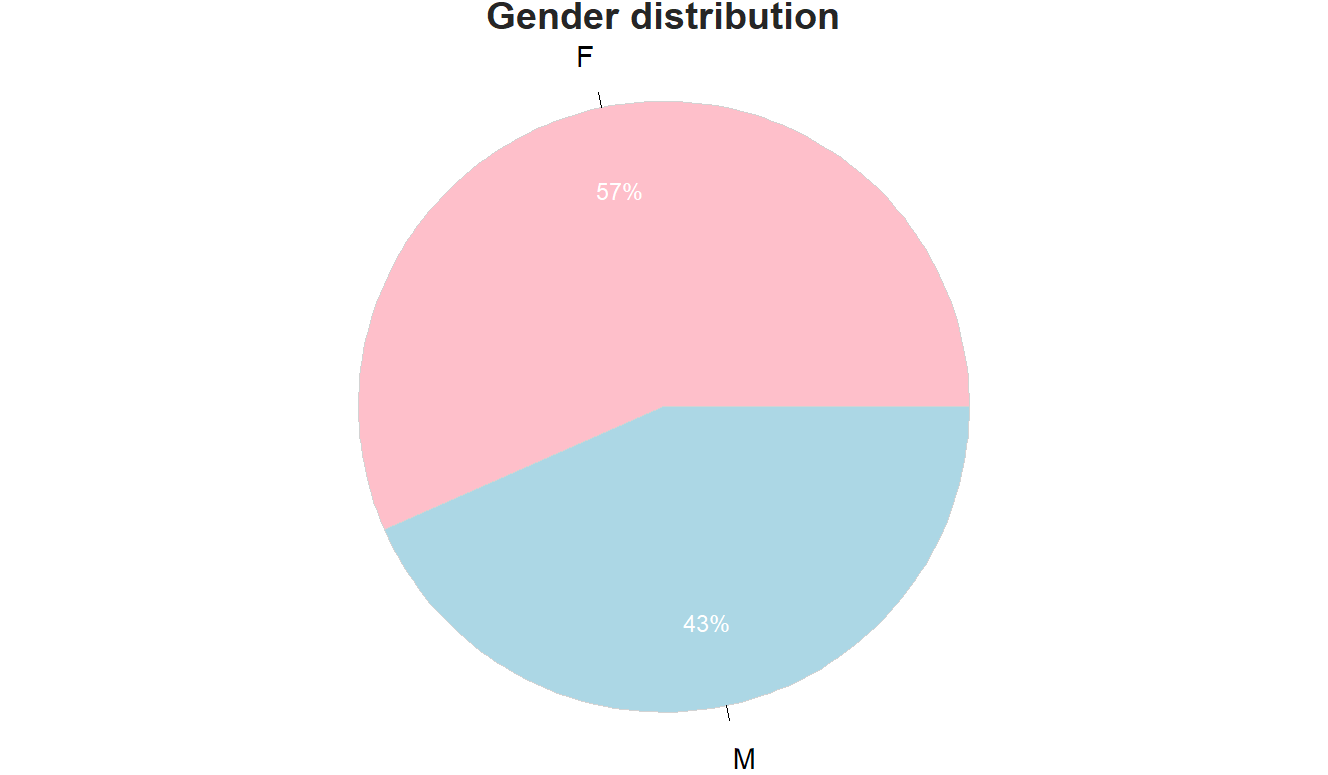
F M Total

Frequencies: 176 135 311

Proportions: 0.566 0.434 1.000

Chi-squared test of null hypothesis of equal probabilities

Chisq = 5.405, df = 1, p-value = 0.020

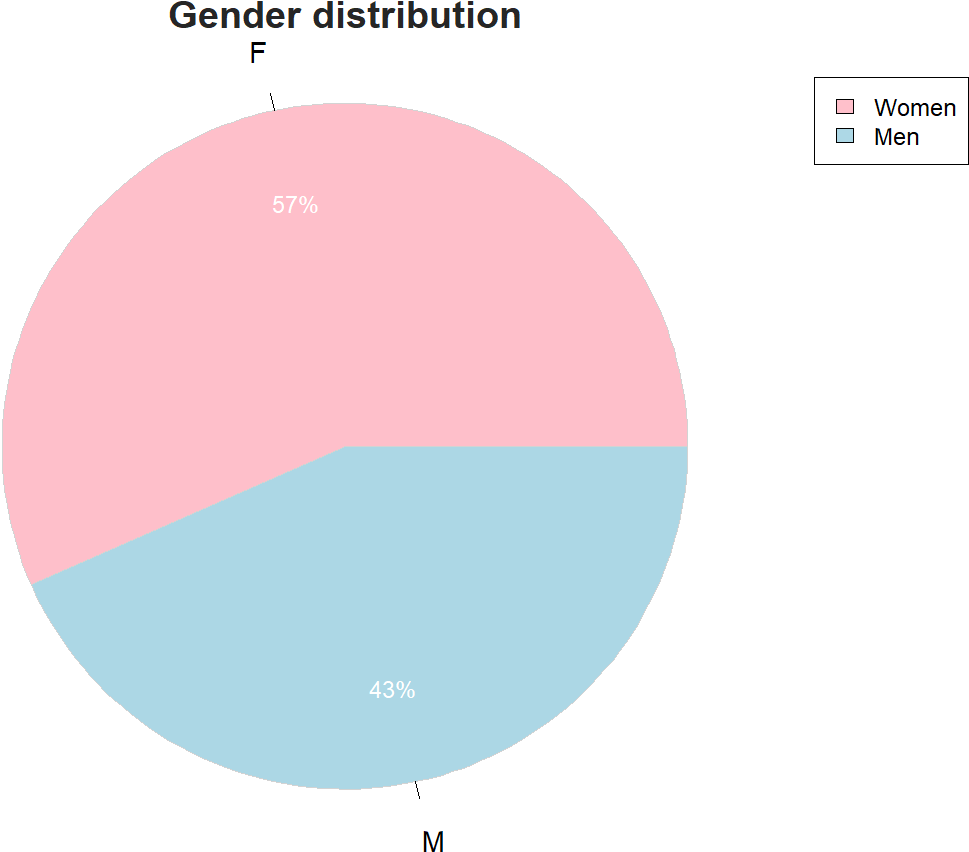


**> legend(x="topright", legend=c("Women","Men"), cex=1.5,fill=c("pink","lightblue"))**

**> #x : position du cadre de légende (ici en bas à droite)**

**> #cex : taille de la police**

**Execution Results: Avec en plus une Légende sur la Figure (Women et Men)**

****

# Race Distribution

**> racedesc<-table(Format1$RaceDesc)**

**> racedesc <- data.frame(race = Format1$RaceDesc)**

**> PieChart(race, hole = 0, values = "%", data = racedesc,**

**+ fill = rainbow(7), main = "Race distribution", cex=1.3)**

**>>>** Suggestions

PieChart(race, hole=0) # traditional pie chart

PieChart(race, values="%") # display %'s on the chart

BarChart(race) # bar chart

Plot(race) # bubble plot

Plot(race, values="count") # lollipop plot

--- race ---

race Count Prop

----------------------------------

AmericnIndnorAlskNtv 3 0.010

Asian 29 0.093

BlackorAfricanAmercn 80 0.257

Hispanic 1 0.003

Two or more races 11 0.035

White 187 0.601

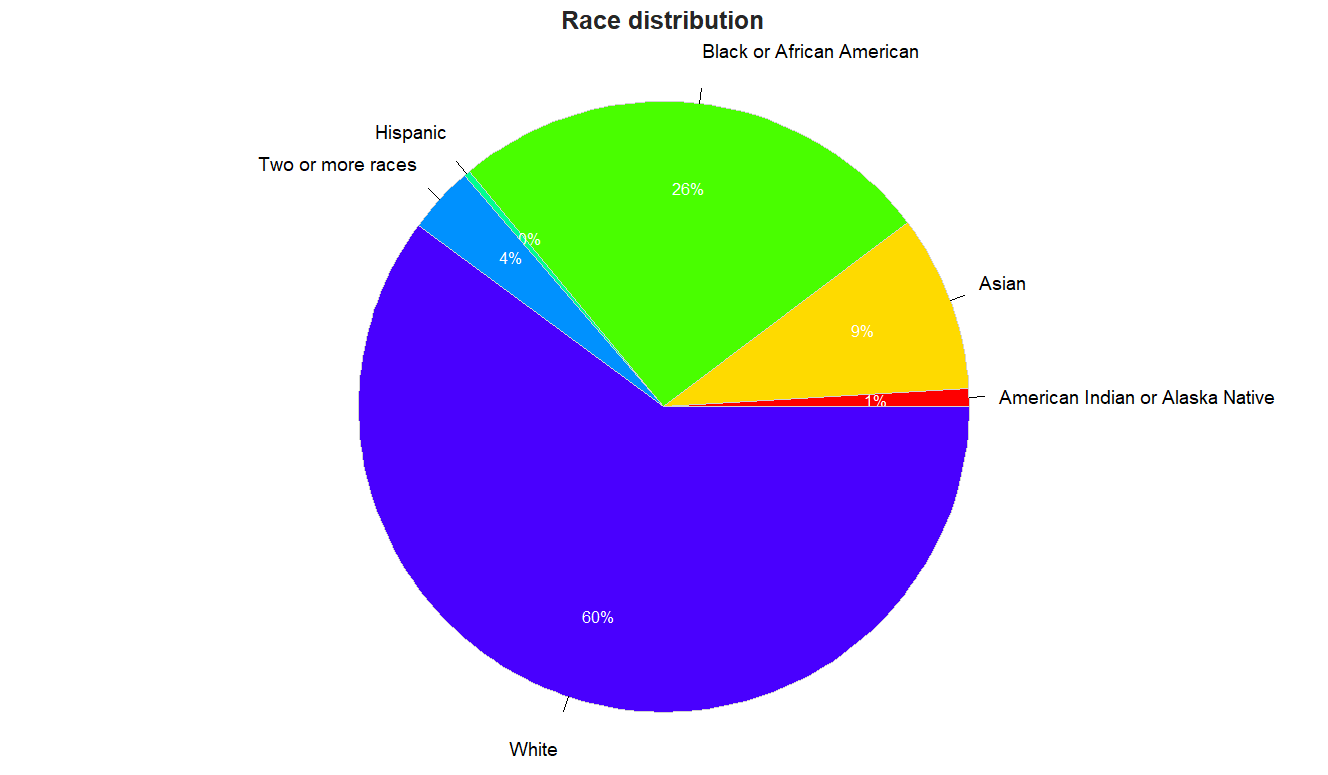
----------------------------------

Total 311 1.000

Chi-squared test of null hypothesis of equal probabilities

Chisq = 505.868, df = 5, p-value = 0.000

**Execution Results:**



# Citizenship distribution

**> citizendesc<-table(Format1$CitizenDesc)**

**> citdesc <- data.frame(citizen = Format1$CitizenDesc)**

**> PieChart(citizen, hole = 0, values = "%", data = citdesc, fill=rainbow(14), main = "Citizenship distribution",radius=0.8, labels.cex=1.5)**

>>> Suggestions

PieChart(citizen, hole=0) # traditional pie chart

PieChart(citizen, values="%") # display %'s on the chart

BarChart(citizen) # bar chart

Plot(citizen) # bubble plot

Plot(citizen, values="count") # lollipop plot

--- citizen ---

citizen Count Prop

---------------------------------

Eligible NonCitizen 12 0.039

Non-Citizen 4 0.013

US Citizen 295 0.949

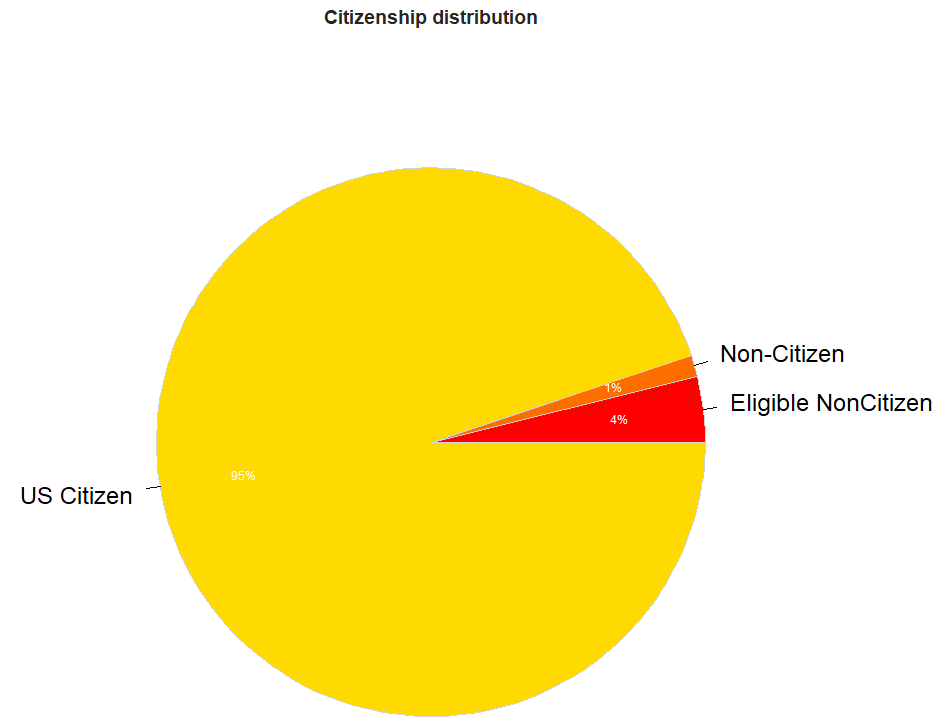
---------------------------------

Total 311 1.000

Chi-squared test of null hypothesis of equal probabilities

Chisq = 530.013, df = 2, p-value = 0.000

**Execution Results:**

****

# Absences per Department:

**> bar1<-barplot(freq, ylab="Frequency", main="Absences",col=rainbow(20), ylim=c(0,30),cex.names=0.7, space=0.02)**

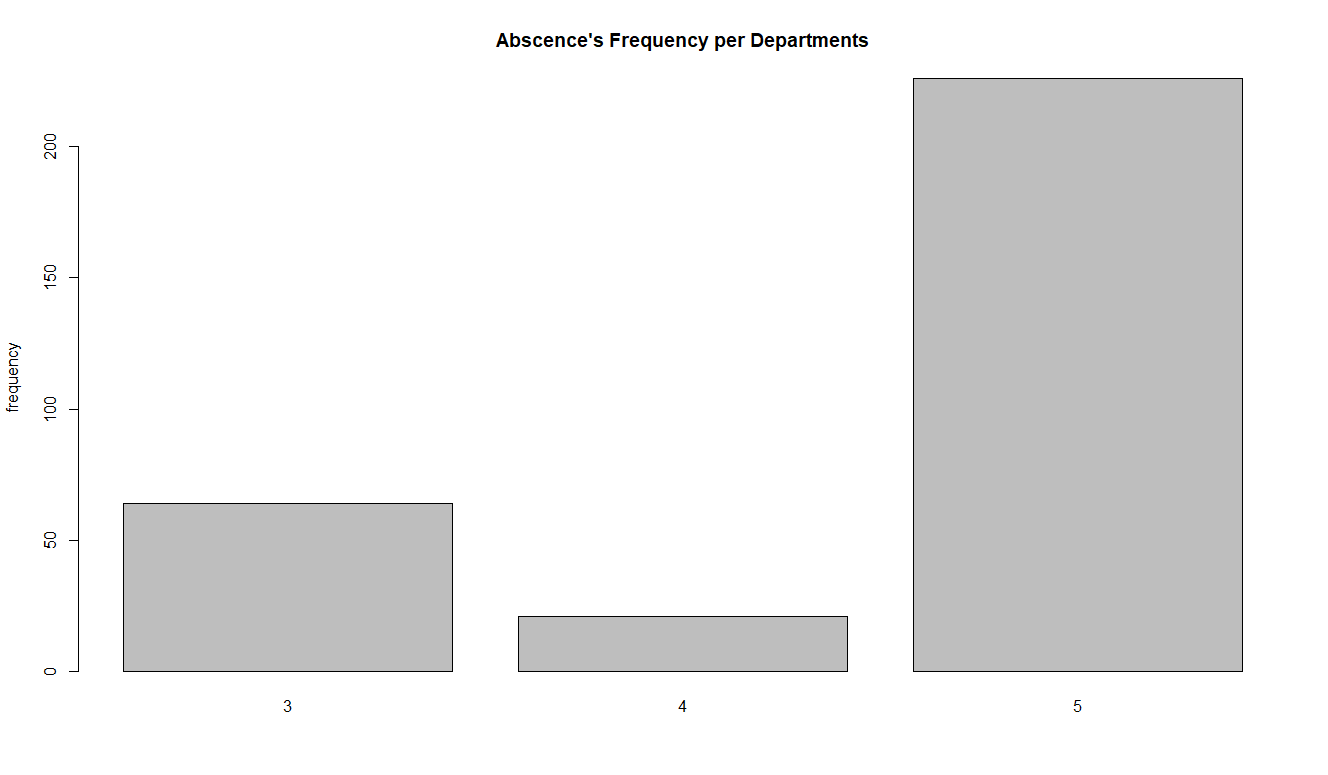
**> for (i in 1:nrow(freq)) text(bar1[i], freq[i]+10,labels=as.character(freq[i]))**

**> Format1<-read.csv2(file.choose(), header=T, sep=",", quote="\"")**

**> abscenceDep<-table(Format1$DeptID[Format1$Absences])**

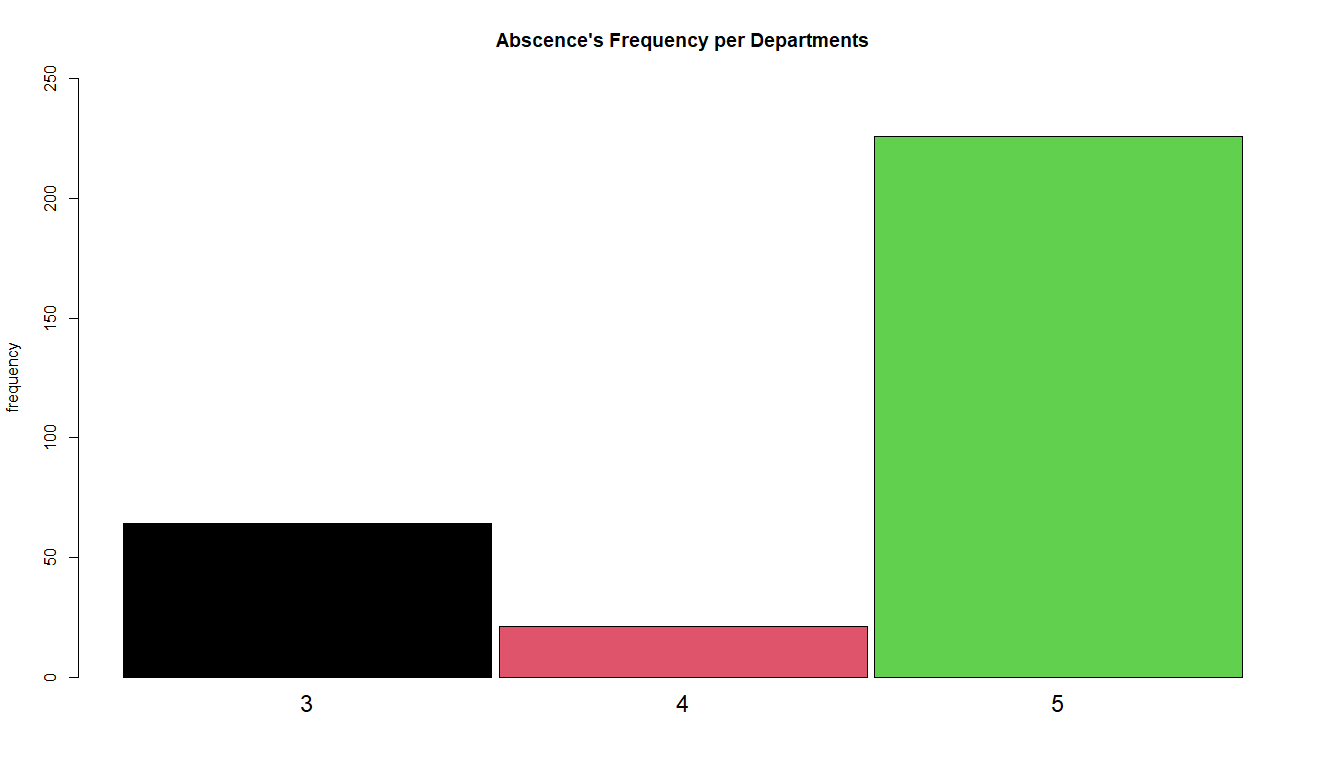
**> barplot(abscenceDep, ylab="frequency", main="Abscence's Frequency per Departments")**

**Execution Results:**

****

**> barplot(abscenceDep, ylab="frequency", main="Abscence's Frequency per Departments",col=c(1,2,3),names.arg=c("3","4","5"),cex.names=1.5, space=0.02, ylim=c(0,250))**

**Execution Results:**



**> for (i in 1:nrow(abscenceDep)) text(bar[i], abscenceDep [i]+10, labels=as.character(abscenceDep [i]))**

**Execution Results:**

