R Project

Sai Sasank Kothuru & Saikrishna Battula

2023-04-13

Clinical\_Data<- read.csv("Copy of HEAD-NECK-RADIOMICS-HN1 Clinical data updated July 2020.csv")  
View(Clinical\_Data)

head(Clinical\_Data)

## id index\_tumour\_location age\_at\_diagnosis biological\_sex  
## 1 HN1004 oropharynx 56 male  
## 2 HN1006 oropharynx 63 female  
## 3 HN1022 oropharynx 56 male  
## 4 HN1026 larynx 67 male  
## 5 HN1029 oropharynx 54 female  
## 6 HN1046 oropharynx 66 male  
## performance\_status\_ecog overall\_hpv\_p16\_status clin\_t clin\_n clin\_m  
## 1 1 negative 4 2 0  
## 2 1 negative 1 2 0  
## 3 0 positive 1 0 0  
## 4 0 4 0 0  
## 5 0 positive 1 2 0  
## 6 0 negative 2 2 0  
## ajcc\_stage pretreat\_hb\_in\_mmolperlitre cancer\_surgery\_performed  
## 1 iva 7.1 no  
## 2 iva NA no  
## 3 i NA no  
## 4 iva 9.3 no  
## 5 iva 7.4 no  
## 6 iva NA no  
## chemotherapy\_given radiotherapy\_total\_treat\_time  
## 1 concomitant 46  
## 2 none 37  
## 3 none 37  
## 4 none 36  
## 5 none 37  
## 6 none 37  
## radiotherapy\_refgydose\_perfraction\_highriskgtv  
## 1 2  
## 2 2  
## 3 2  
## 4 2  
## 5 2  
## 6 2  
## radiotherapy\_refgydose\_total\_highriskgtv  
## 1 70  
## 2 68  
## 3 68  
## 4 68  
## 5 68  
## 6 68  
## radiotherapy\_number\_fractions\_highriskgtv event\_overall\_survival  
## 1 35 1  
## 2 34 0  
## 3 34 0  
## 4 34 0  
## 5 34 0  
## 6 34 1  
## overall\_survival\_in\_days event\_recurrence\_metastatic\_free\_survival  
## 1 3193 0  
## 2 2805 0  
## 3 3198 0  
## 4 3244 1  
## 5 2925 0  
## 6 828 1  
## recurrence\_metastatic\_free\_survival\_in\_days event\_local\_recurrence  
## 1 3193 0  
## 2 1940 0  
## 3 1269 0  
## 4 600 1  
## 5 2101 0  
## 6 707 0  
## local\_recurrence\_in\_days event\_locoregional\_recurrence  
## 1 3193 0  
## 2 1940 0  
## 3 1269 0  
## 4 600 1  
## 5 2101 0  
## 6 749 0  
## locoregional\_recurrence\_in\_days event\_distant\_metastases  
## 1 3193 0  
## 2 1940 0  
## 3 1269 0  
## 4 600 0  
## 5 2101 0  
## 6 749 1  
## distant\_metastases\_in\_days  
## 1 3193  
## 2 1940  
## 3 1269  
## 4 1315  
## 5 2101  
## 6 707

#installing required packages  
#install.packages("ggplot2")  
#install.packages("dplyr")  
#install.packages("plyr")  
#install.packages("tidyr")  
#install.packages("ggplot")  
#install.packages("ggplot2")  
#library(ggplot2)  
  
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.2.3

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(plyr)

## ------------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.  
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:  
## library(plyr); library(dplyr)

## ------------------------------------------------------------------------------

##   
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':  
##   
## arrange, count, desc, failwith, id, mutate, rename, summarise,  
## summarize

library(tidyr)  
library(ggplot2)

#data cleaning and pre processing  
  
#checking null values  
  
Clinical\_Data1<- data.frame(Clinical\_Data)  
colSums(is.na(Clinical\_Data1))

## id   
## 0   
## index\_tumour\_location   
## 0   
## age\_at\_diagnosis   
## 0   
## biological\_sex   
## 0   
## performance\_status\_ecog   
## 10   
## overall\_hpv\_p16\_status   
## 0   
## clin\_t   
## 0   
## clin\_n   
## 0   
## clin\_m   
## 0   
## ajcc\_stage   
## 0   
## pretreat\_hb\_in\_mmolperlitre   
## 54   
## cancer\_surgery\_performed   
## 0   
## chemotherapy\_given   
## 0   
## radiotherapy\_total\_treat\_time   
## 0   
## radiotherapy\_refgydose\_perfraction\_highriskgtv   
## 0   
## radiotherapy\_refgydose\_total\_highriskgtv   
## 0   
## radiotherapy\_number\_fractions\_highriskgtv   
## 0   
## event\_overall\_survival   
## 0   
## overall\_survival\_in\_days   
## 0   
## event\_recurrence\_metastatic\_free\_survival   
## 0   
## recurrence\_metastatic\_free\_survival\_in\_days   
## 0   
## event\_local\_recurrence   
## 0   
## local\_recurrence\_in\_days   
## 0   
## event\_locoregional\_recurrence   
## 0   
## locoregional\_recurrence\_in\_days   
## 0   
## event\_distant\_metastases   
## 0   
## distant\_metastases\_in\_days   
## 0

#Elemenating null values  
  
Clinical\_Data1<-na.omit(Clinical\_Data1)  
  
#rechecking null values  
colSums(is.na(Clinical\_Data1))

## id   
## 0   
## index\_tumour\_location   
## 0   
## age\_at\_diagnosis   
## 0   
## biological\_sex   
## 0   
## performance\_status\_ecog   
## 0   
## overall\_hpv\_p16\_status   
## 0   
## clin\_t   
## 0   
## clin\_n   
## 0   
## clin\_m   
## 0   
## ajcc\_stage   
## 0   
## pretreat\_hb\_in\_mmolperlitre   
## 0   
## cancer\_surgery\_performed   
## 0   
## chemotherapy\_given   
## 0   
## radiotherapy\_total\_treat\_time   
## 0   
## radiotherapy\_refgydose\_perfraction\_highriskgtv   
## 0   
## radiotherapy\_refgydose\_total\_highriskgtv   
## 0   
## radiotherapy\_number\_fractions\_highriskgtv   
## 0   
## event\_overall\_survival   
## 0   
## overall\_survival\_in\_days   
## 0   
## event\_recurrence\_metastatic\_free\_survival   
## 0   
## recurrence\_metastatic\_free\_survival\_in\_days   
## 0   
## event\_local\_recurrence   
## 0   
## local\_recurrence\_in\_days   
## 0   
## event\_locoregional\_recurrence   
## 0   
## locoregional\_recurrence\_in\_days   
## 0   
## event\_distant\_metastases   
## 0   
## distant\_metastases\_in\_days   
## 0

View(Clinical\_Data1)  
head(Clinical\_Data1)

## id index\_tumour\_location age\_at\_diagnosis biological\_sex  
## 1 HN1004 oropharynx 56 male  
## 4 HN1026 larynx 67 male  
## 5 HN1029 oropharynx 54 female  
## 8 HN1054 oropharynx 64 male  
## 12 HN1067 oropharynx 69 male  
## 15 HN1079 oropharynx 71 male  
## performance\_status\_ecog overall\_hpv\_p16\_status clin\_t clin\_n clin\_m  
## 1 1 negative 4 2 0  
## 4 0 4 0 0  
## 5 0 positive 1 2 0  
## 8 0 positive 2 0 0  
## 12 1 negative 3 0 0  
## 15 1 positive 3 2 0  
## ajcc\_stage pretreat\_hb\_in\_mmolperlitre cancer\_surgery\_performed  
## 1 iva 7.1 no  
## 4 iva 9.3 no  
## 5 iva 7.4 no  
## 8 ii 8.2 no  
## 12 iii 8.9 no  
## 15 iva 8.7 no  
## chemotherapy\_given radiotherapy\_total\_treat\_time  
## 1 concomitant 46  
## 4 none 36  
## 5 none 37  
## 8 none 37  
## 12 concurrent 45  
## 15 none 37  
## radiotherapy\_refgydose\_perfraction\_highriskgtv  
## 1 2  
## 4 2  
## 5 2  
## 8 2  
## 12 2  
## 15 2  
## radiotherapy\_refgydose\_total\_highriskgtv  
## 1 70  
## 4 68  
## 5 68  
## 8 68  
## 12 70  
## 15 68  
## radiotherapy\_number\_fractions\_highriskgtv event\_overall\_survival  
## 1 35 1  
## 4 34 0  
## 5 34 0  
## 8 34 0  
## 12 35 1  
## 15 34 1  
## overall\_survival\_in\_days event\_recurrence\_metastatic\_free\_survival  
## 1 3193 0  
## 4 3244 1  
## 5 2925 0  
## 8 2978 0  
## 12 3927 0  
## 15 743 1  
## recurrence\_metastatic\_free\_survival\_in\_days event\_local\_recurrence  
## 1 3193 0  
## 4 600 1  
## 5 2101 0  
## 8 1049 0  
## 12 2318 0  
## 15 532 0  
## local\_recurrence\_in\_days event\_locoregional\_recurrence  
## 1 3193 0  
## 4 600 1  
## 5 2101 0  
## 8 1049 0  
## 12 2318 0  
## 15 578 0  
## locoregional\_recurrence\_in\_days event\_distant\_metastases  
## 1 3193 0  
## 4 600 0  
## 5 2101 0  
## 8 1049 0  
## 12 2318 0  
## 15 578 1  
## distant\_metastases\_in\_days  
## 1 3193  
## 4 1315  
## 5 2101  
## 8 1049  
## 12 2318  
## 15 532

#summary of the data  
  
summary(Clinical\_Data1)

## id index\_tumour\_location age\_at\_diagnosis biological\_sex   
## Length:78 Length:78 Min. :46.00 Length:78   
## Class :character Class :character 1st Qu.:55.00 Class :character   
## Mode :character Mode :character Median :60.00 Mode :character   
## Mean :60.88   
## 3rd Qu.:66.00   
## Max. :83.00   
## performance\_status\_ecog overall\_hpv\_p16\_status clin\_t clin\_n   
## Min. :0.0000 Length:78 Min. :1.000 Min. :0.000   
## 1st Qu.:0.0000 Class :character 1st Qu.:2.000 1st Qu.:0.000   
## Median :0.5000 Mode :character Median :3.000 Median :2.000   
## Mean :0.5256 Mean :2.987 Mean :1.244   
## 3rd Qu.:1.0000 3rd Qu.:4.000 3rd Qu.:2.000   
## Max. :2.0000 Max. :4.000 Max. :3.000   
## clin\_m ajcc\_stage pretreat\_hb\_in\_mmolperlitre  
## Min. :0.00000 Length:78 Min. : 5.000   
## 1st Qu.:0.00000 Class :character 1st Qu.: 7.925   
## Median :0.00000 Mode :character Median : 8.700   
## Mean :0.01282 Mean : 8.504   
## 3rd Qu.:0.00000 3rd Qu.: 9.200   
## Max. :1.00000 Max. :11.300   
## cancer\_surgery\_performed chemotherapy\_given radiotherapy\_total\_treat\_time  
## Length:78 Length:78 Min. :32.00   
## Class :character Class :character 1st Qu.:37.00   
## Mode :character Mode :character Median :41.00   
## Mean :41.51   
## 3rd Qu.:46.00   
## Max. :56.00   
## radiotherapy\_refgydose\_perfraction\_highriskgtv  
## Min. :2.000   
## 1st Qu.:2.000   
## Median :2.000   
## Mean :2.015   
## 3rd Qu.:2.000   
## Max. :2.400   
## radiotherapy\_refgydose\_total\_highriskgtv  
## Min. :60.00   
## 1st Qu.:68.00   
## Median :68.00   
## Mean :68.59   
## 3rd Qu.:70.00   
## Max. :70.00   
## radiotherapy\_number\_fractions\_highriskgtv event\_overall\_survival  
## Min. :25.0 Min. :0.0000   
## 1st Qu.:34.0 1st Qu.:0.0000   
## Median :34.0 Median :1.0000   
## Mean :34.1 Mean :0.5897   
## 3rd Qu.:35.0 3rd Qu.:1.0000   
## Max. :35.0 Max. :1.0000   
## overall\_survival\_in\_days event\_recurrence\_metastatic\_free\_survival  
## Min. : 55 Min. :0.0000   
## 1st Qu.: 596 1st Qu.:0.0000   
## Median :2681 Median :0.0000   
## Mean :2160 Mean :0.2949   
## 3rd Qu.:3238 3rd Qu.:1.0000   
## Max. :4572 Max. :1.0000   
## recurrence\_metastatic\_free\_survival\_in\_days event\_local\_recurrence  
## Min. : 38.0 Min. :0.0000   
## 1st Qu.: 415.5 1st Qu.:0.0000   
## Median :1099.5 Median :0.0000   
## Mean :1226.3 Mean :0.2051   
## 3rd Qu.:1947.0 3rd Qu.:0.0000   
## Max. :3200.0 Max. :1.0000   
## local\_recurrence\_in\_days event\_locoregional\_recurrence  
## Min. : 38.0 Min. :0.0000   
## 1st Qu.: 437.8 1st Qu.:0.0000   
## Median :1197.0 Median :0.0000   
## Mean :1246.8 Mean :0.2564   
## 3rd Qu.:1947.0 3rd Qu.:0.7500   
## Max. :3200.0 Max. :1.0000   
## locoregional\_recurrence\_in\_days event\_distant\_metastases  
## Min. : 38.0 Min. :0.00000   
## 1st Qu.: 415.5 1st Qu.:0.00000   
## Median :1099.5 Median :0.00000   
## Mean :1226.9 Mean :0.07692   
## 3rd Qu.:1947.0 3rd Qu.:0.00000   
## Max. :3200.0 Max. :1.00000   
## distant\_metastases\_in\_days  
## Min. : 55   
## 1st Qu.: 533   
## Median :1350   
## Mean :1387   
## 3rd Qu.:2096   
## Max. :3254

#number of people diagnosed at different ages  
  
age <-table(Clinical\_Data1$age\_at\_diagnosis)  
  
age

##   
## 46 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 74 76   
## 3 4 3 1 1 7 2 5 5 2 1 7 6 5 3 1 1 3 3 1 1 2 1 1 3 2   
## 77 79 83   
## 1 2 1

#total number of females and males  
  
gender<-table(Clinical\_Data1$biological\_sex)  
  
gender

##   
## female male   
## 16 62

#number of performance status 0,1,2  
  
performance\_status<-table(Clinical\_Data1$performance\_status\_ecog)  
  
performance\_status

##   
## 0 1 2   
## 39 37 2

#number of ph status  
  
ph\_status<-table(Clinical\_Data1$overall\_hpv\_p16\_status)  
  
ph\_status

##   
## negative positive   
## 23 40 15

#number of tumor locations  
  
tumor<-table(Clinical\_Data1$index\_tumour\_location)  
tumor

##   
## larynx oropharynx   
## 21 57

#Types and count of chemotherapy  
  
chemotherapy<-data.frame(table(Clinical\_Data1$chemotherapy\_given))  
chemotherapy

## Var1 Freq  
## 1 concomitant 25  
## 2 concurrent 10  
## 3 none 43

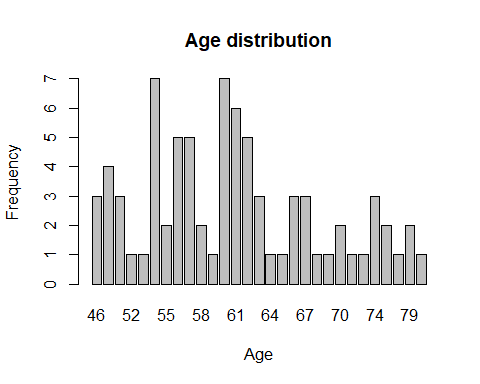
#number of cancer surgery performed  
  
cancer\_surgery<-data.frame(table(Clinical\_Data1$cancer\_surgery\_performed))  
cancer\_surgery

## Var1 Freq  
## 1 no 73  
## 2 yes 5

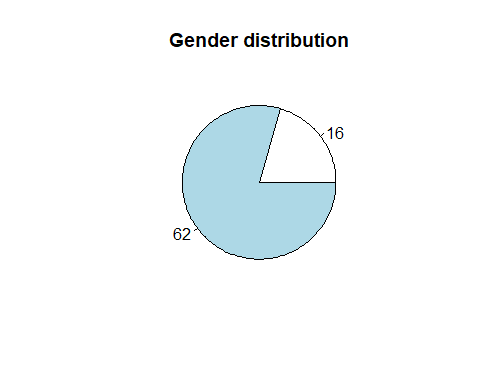
#stages of ajcc  
  
ajcc<-table(Clinical\_Data1$ajcc\_stage)  
ajcc

##   
## i ii iii iva ivb ivc   
## 5 6 9 48 9 1

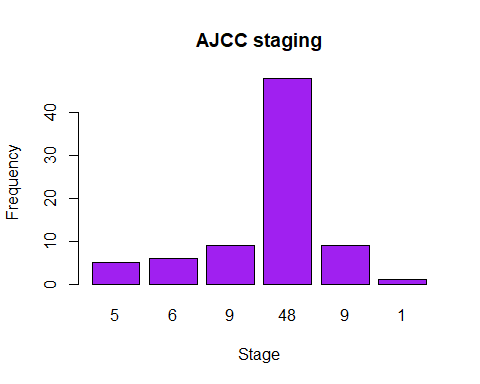
#data vizualization  
#barplot for number of people with same age   
  
barplot(age, main = "Age distribution", xlab = "Age", ylab = "Frequency")



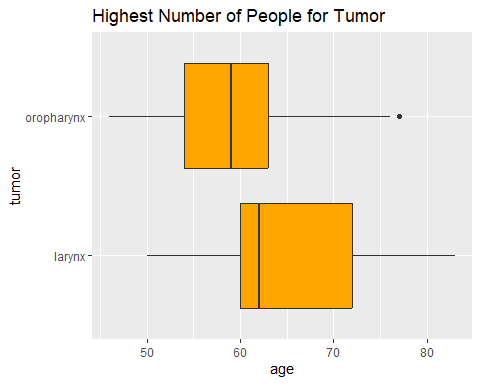
#62-male, 16- female  
pie(gender,labels = gender, main = "Gender distribution")



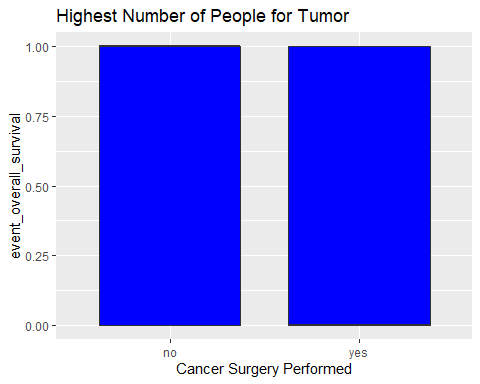
#  
  
barplot(ajcc, main = "AJCC staging", xlab = "Stage", ylab = "Frequency", names.arg = ajcc, col = "purple")



#age,tumor  
  
ggplot(Clinical\_Data1, aes(x = age\_at\_diagnosis, y = index\_tumour\_location)) +  
 geom\_boxplot(fill = "orange") +  
 labs(x = "age", y = "tumor") +  
 ggtitle("Highest Number of People for Tumor")



#yes-treatment- success  
  
#cancer\_surgery  
#chemotherapy  
ggplot(Clinical\_Data1, aes(x = cancer\_surgery\_performed, y = event\_overall\_survival)) +  
 geom\_boxplot(fill = "blue") +  
 labs(x = "Cancer Surgery Performed", y = "event\_overall\_survival") +  
 ggtitle("Highest Number of People for Tumor")



table(Clinical\_Data1$event\_overall\_survival)

##   
## 0 1   
## 32 46