chad\_huntebrinker\_hw3

Chad Huntebrinker

2024-09-14

Part 1: Basic Analysis of Cancer Patients Data Using Lists

Question 1: create a list object Create a list named patient\_data to represent information about cancer patients. Include the given 5 patients, within the following four vectors on the list:

PatientID (Numeric) : 1, 2, 3, 4, 5 Age (Numeric) : 45, 32, 58, 41, 50 CancerType (Character) : “Breast”, “Lung”, “Colon”, “Breast”, “Lung” TreatmentStatus (Logical) : TRUE, FALSE, TRUE, TRUE, FALSE

Display the patient\_data.

patient\_data <- list(  
 PatientID = c(1, 2, 3, 4, 5),  
 Age = c(45, 32, 58, 41, 50),  
 CancerType = c("Breast", "Lung", "Colon", "Breast", "Lung"),  
 TreatmentStatus = c(TRUE, FALSE, TRUE, TRUE, FALSE))  
patient\_data

## $PatientID  
## [1] 1 2 3 4 5  
##   
## $Age  
## [1] 45 32 58 41 50  
##   
## $CancerType  
## [1] "Breast" "Lung" "Colon" "Breast" "Lung"   
##   
## $TreatmentStatus  
## [1] TRUE FALSE TRUE TRUE FALSE

Question 2 : Perform Basic Operations on the list object Find the average age of patients.

attributes(patient\_data)

## $names  
## [1] "PatientID" "Age" "CancerType" "TreatmentStatus"

length(patient\_data)

## [1] 4

class(patient\_data)

## [1] "list"

print(mean(patient\_data$Age))

## [1] 45.2

Retrieve the patientIDs with “Breast” cancer type.

index <- which(patient\_data$CancerType == "Breast")  
ids <- patient\_data$PatientID[index]  
print(ids)

## [1] 1 4

Update the treatment status for the “Colon” cancer patient to FALSE.

patient\_data$TreatmentStatus[3] <- FALSE  
patient\_data$TreatmentStatus[patient\_data$CancerType=="Colon"]

## [1] FALSE

Display the updated patient\_data.

print(patient\_data)

## $PatientID  
## [1] 1 2 3 4 5  
##   
## $Age  
## [1] 45 32 58 41 50  
##   
## $CancerType  
## [1] "Breast" "Lung" "Colon" "Breast" "Lung"   
##   
## $TreatmentStatus  
## [1] TRUE FALSE FALSE TRUE FALSE

Calculate the percentage of patients undergoing treatment

num\_treatments <- which(patient\_data$TreatmentStatus == TRUE)  
final\_percentage <- (length(num\_treatments) / 5) \* 100  
cat(toString(final\_percentage), "%", sep = "")

## 40%

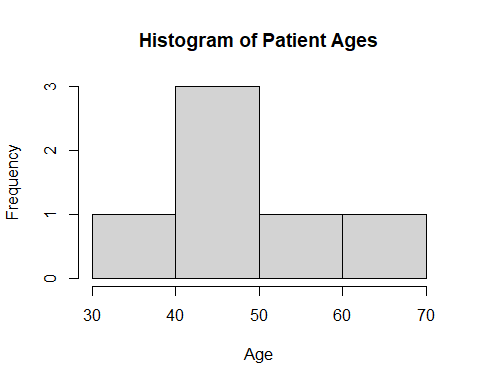
Question 3: Update the patient\_data by adding another patient with the following details. Print the updated list

patient\_data$PatientID[6] <- 6   
patient\_data$Age[6] <- 68  
patient\_data$CancerType[6] <- "Brain"  
patient\_data$TreatmentStatus[6] <- TRUE  
print(patient\_data)

## $PatientID  
## [1] 1 2 3 4 5 6  
##   
## $Age  
## [1] 45 32 58 41 50 68  
##   
## $CancerType  
## [1] "Breast" "Lung" "Colon" "Breast" "Lung" "Brain"   
##   
## $TreatmentStatus  
## [1] TRUE FALSE FALSE TRUE FALSE TRUE

Question 4: plot a histogram for the age of patients

hist(patient\_data$Age, xlab = "Age", main = "Histogram of Patient Ages")



——————————————End of Part 1 ——————————————————–

Part 2: Analysis of Breast cancer data set You are provided with a breast cancer data set (BC\_data.cvs) and you have to work on this table as per the instructions provided.

Question 1: Read the Breast cancer data into a dataframe called “Bc.data”

rm(list = ls())  
Bc.data <- read.csv("BC\_data.csv", header = TRUE, row.names = 1)

The rownames should be the patient\_IDs.

rownames(Bc.data)

## [1] "MB-0115" "MB-3046" "MB-5173" "MB-0020" "MB-0062" "MB-0079" "MB-0100"  
## [8] "MB-0127" "MB-0149" "MB-0157" "MB-0158" "MB-0164" "MB-0174" "MB-0179"  
## [15] "MB-0188" "MB-0191" "MB-0200" "MB-0206" "MB-0214" "MB-0220" "MB-0221"  
## [22] "MB-0238" "MB-0249" "MB-0265" "MB-0269" "MB-0289" "MB-0292" "MB-0294"  
## [29] "MB-0303" "MB-0316" "MB-0318" "MB-0340" "MB-0352" "MB-0354" "MB-0372"  
## [36] "MB-0396" "MB-0399" "MB-0400" "MB-0401" "MB-0414" "MB-0420" "MB-0424"  
## [43] "MB-0432" "MB-0435" "MB-0436" "MB-0446" "MB-0453" "MB-0464" "MB-0470"  
## [50] "MB-0481" "MB-0489" "MB-0494" "MB-0495" "MB-0500" "MB-0502" "MB-0516"  
## [57] "MB-0525" "MB-0540" "MB-0558" "MB-0581" "MB-0582" "MB-0588" "MB-0608"  
## [64] "MB-0613" "MB-0617" "MB-0635" "MB-0639" "MB-0653" "MB-0658" "MB-0660"  
## [71] "MB-0664" "MB-0869" "MB-0874" "MB-0893" "MB-0901" "MB-0906" "MB-2556"  
## [78] "MB-2643" "MB-2718" "MB-2724" "MB-2753" "MB-2754" "MB-2771" "MB-2827"  
## [85] "MB-2833" "MB-2834" "MB-2842" "MB-2849" "MB-2850" "MB-2857" "MB-2904"  
## [92] "MB-2912" "MB-2917" "MB-2922" "MB-2929" "MB-2957" "MB-2963" "MB-2993"  
## [99] "MB-3001" "MB-3014" "MB-3057" "MB-3058" "MB-3062" "MB-3063" "MB-3067"  
## [106] "MB-3123" "MB-3153" "MB-3211" "MB-3218" "MB-3271" "MB-3277" "MB-3292"  
## [113] "MB-3297" "MB-3383" "MB-3395" "MB-3396" "MB-3453" "MB-3476" "MB-3500"  
## [120] "MB-3502" "MB-3567" "MB-3702" "MB-3706" "MB-3752" "MB-4024" "MB-4146"  
## [127] "MB-4303" "MB-4332" "MB-4351" "MB-4407" "MB-4416" "MB-4417" "MB-4621"  
## [134] "MB-4622" "MB-4640" "MB-4660" "MB-4667" "MB-4679" "MB-4707" "MB-4714"  
## [141] "MB-4715" "MB-4717" "MB-4732" "MB-4733" "MB-4769" "MB-4792" "MB-4793"  
## [148] "MB-4859" "MB-4865" "MB-4880" "MB-4888" "MB-4893" "MB-4911" "MB-4928"  
## [155] "MB-4931" "MB-4938" "MB-4942" "MB-4974" "MB-4992" "MB-4993" "MB-5008"  
## [162] "MB-5041" "MB-5052" "MB-5057" "MB-5058" "MB-5065" "MB-5072" "MB-5076"  
## [169] "MB-5100" "MB-5102" "MB-5109" "MB-5115" "MB-5135" "MB-5137" "MB-5145"  
## [176] "MB-5148" "MB-5155" "MB-5162" "MB-5208" "MB-5209" "MB-5213" "MB-5222"  
## [183] "MB-5223" "MB-5225" "MB-5232" "MB-5235" "MB-5236" "MB-5255" "MB-5258"  
## [190] "MB-5281" "MB-5295" "MB-5298" "MB-5299" "MB-5323" "MB-5335" "MB-5348"  
## [197] "MB-5390" "MB-5392" "MB-5408" "MB-5421" "MB-5427" "MB-5440" "MB-5446"  
## [204] "MB-5450" "MB-5465" "MB-5526" "MB-5529" "MB-5531" "MB-5547" "MB-5551"  
## [211] "MB-5560" "MB-5572" "MB-5577" "MB-5616" "MB-5633" "MB-5634" "MB-5651"  
## [218] "MB-5655" "MB-6143" "MB-6223" "MB-6237" "MB-6251"

Perform all the basic functions on the dataframe and get familiarize with the data.

dim(Bc.data)

## [1] 222 9

class(Bc.data)

## [1] "data.frame"

str(Bc.data)

## 'data.frame': 222 obs. of 9 variables:  
## $ Survival\_Status: chr "DECEASED" "DECEASED" "DECEASED" "DECEASED" ...  
## $ Survival\_Months: num 66.7 73.8 22.1 22.4 154 ...  
## $ Mutation\_Count : int 5 5 8 0 3 4 1 5 2 1 ...  
## $ Tumor\_Size : num 25 35 50 65 17 40 39 13 39 30 ...  
## $ Hormone\_Therapy: chr "NO" "YES" "YES" "NO" ...  
## $ Surgery\_Type : chr "MASTECTOMY" "MASTECTOMY" "MASTECTOMY" "MASTECTOMY" ...  
## $ Oncotree\_Code : chr "IDC" "IDC" "IDC" "ILC" ...  
## $ Tumor\_Stage : int 2 1 2 3 1 2 2 2 2 2 ...  
## $ Chemotherapy : chr "YES" "NO" "NO" "YES" ...

ncol(Bc.data)

## [1] 9

colnames(Bc.data)

## [1] "Survival\_Status" "Survival\_Months" "Mutation\_Count" "Tumor\_Size"   
## [5] "Hormone\_Therapy" "Surgery\_Type" "Oncotree\_Code" "Tumor\_Stage"   
## [9] "Chemotherapy"

Question 2: Check if NA values are available.

any(is.na(Bc.data))

## [1] FALSE

create a vector called “Surv\_Mnths” with the data from “Survival\_Months” column

Surv\_Mnths <- Bc.data$Survival\_Months

create a vector called “Tum\_Size” with the data from “Tumor\_Size” column

Tum\_Size <- Bc.data$Tumor\_Size

check the correlation between “Surv\_Mnths” and “Tum\_Size” and interpret the relation based on the correlation value. Write your interpretation in the comments. Clue: There is a built-in function for correlation

cor(Surv\_Mnths, Tum\_Size)

## [1] -0.1989186

#The cor function returns a value of -0.1989186  
#Thus, there is a negative linear relationship between the two

Calculate the mean and median of the “Survival\_Months” column.

mean(Surv\_Mnths)

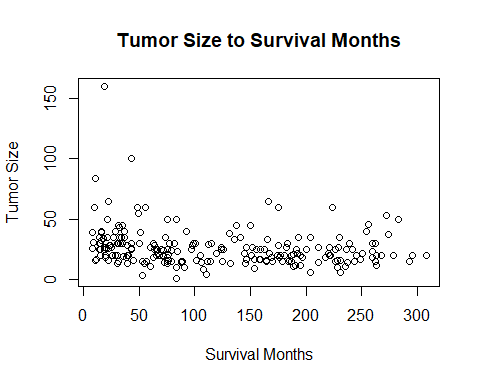
## [1] 118.7036

median(Surv\_Mnths)

## [1] 98.2

generate a scatter plot “Surv\_Mnths” Vs “Tum\_Size” and present with proper axis labels

plot(Surv\_Mnths, Tum\_Size, ylab= "Tumor Size", xlab = "Survival Months", main = "Tumor Size to Survival Months")



Question 3: Print a table to show how many patients are DECEASED and how many LIVING

table(Bc.data$Survival\_Status)

##   
## DECEASED LIVING   
## 120 102

Create a new dataframe called “living\_patient” with only living patients.

living\_patients <- subset(Bc.data, Bc.data$Survival\_Status=="LIVING")

In these living patients, how many are undergoing Chemotherapy and how many undergoing Hormone\_Therapy

table(living\_patients$Chemotherapy)

##   
## NO YES   
## 47 55

table(living\_patients$Hormone\_Therapy)

##   
## NO YES   
## 75 27

Question 4:

Create an additional column in the BC.data dataframe with the name ‘Full\_Info’ by concatenating the values of Survival\_Status with Tumor\_Stage seperated by underscore.

Bc.data$Full\_Info <- paste(Bc.data$Survival\_Status, Bc.data$Tumor\_Stage, sep = "\_")

Create additional column ‘Survival\_Status\_Short’ by replacing the the Survival\_Status values “DECEASED” with “D” and “LIVING” with “L”.

Bc.data$Survival\_Status\_Short[Bc.data$Survival\_Status == "LIVING"] <- "L"  
Bc.data$Survival\_Status\_Short[Bc.data$Survival\_Status == "DECEASED"] <- "D"

Check the factor levels of Survival\_Status, Tumor\_Stage and Surgery\_Type in BC.data dataframe

factor(Bc.data$Survival\_Status)

## [1] DECEASED DECEASED DECEASED DECEASED LIVING DECEASED DECEASED LIVING   
## [9] DECEASED LIVING LIVING LIVING LIVING DECEASED DECEASED DECEASED  
## [17] DECEASED LIVING LIVING DECEASED DECEASED LIVING LIVING LIVING   
## [25] LIVING DECEASED DECEASED LIVING LIVING LIVING LIVING LIVING   
## [33] DECEASED DECEASED DECEASED LIVING DECEASED DECEASED DECEASED LIVING   
## [41] LIVING DECEASED LIVING LIVING DECEASED LIVING DECEASED DECEASED  
## [49] LIVING DECEASED LIVING DECEASED DECEASED LIVING LIVING LIVING   
## [57] LIVING LIVING DECEASED DECEASED DECEASED LIVING DECEASED LIVING   
## [65] DECEASED LIVING LIVING DECEASED LIVING DECEASED DECEASED DECEASED  
## [73] DECEASED LIVING LIVING DECEASED DECEASED DECEASED LIVING DECEASED  
## [81] LIVING DECEASED DECEASED LIVING LIVING DECEASED DECEASED DECEASED  
## [89] DECEASED LIVING DECEASED LIVING DECEASED DECEASED DECEASED LIVING   
## [97] DECEASED LIVING LIVING LIVING DECEASED LIVING LIVING DECEASED  
## [105] LIVING LIVING LIVING LIVING LIVING DECEASED DECEASED LIVING   
## [113] LIVING DECEASED LIVING LIVING DECEASED DECEASED LIVING LIVING   
## [121] LIVING LIVING DECEASED DECEASED LIVING DECEASED DECEASED LIVING   
## [129] DECEASED DECEASED DECEASED DECEASED LIVING DECEASED LIVING DECEASED  
## [137] DECEASED DECEASED LIVING DECEASED DECEASED DECEASED LIVING DECEASED  
## [145] DECEASED DECEASED DECEASED DECEASED LIVING LIVING LIVING LIVING   
## [153] DECEASED LIVING DECEASED DECEASED DECEASED DECEASED LIVING DECEASED  
## [161] LIVING LIVING DECEASED DECEASED DECEASED LIVING DECEASED DECEASED  
## [169] DECEASED DECEASED DECEASED LIVING DECEASED LIVING DECEASED DECEASED  
## [177] LIVING LIVING LIVING DECEASED DECEASED DECEASED DECEASED DECEASED  
## [185] DECEASED DECEASED LIVING LIVING DECEASED LIVING DECEASED DECEASED  
## [193] DECEASED LIVING DECEASED LIVING DECEASED LIVING LIVING LIVING   
## [201] DECEASED LIVING DECEASED LIVING DECEASED LIVING DECEASED LIVING   
## [209] LIVING LIVING LIVING LIVING DECEASED LIVING DECEASED DECEASED  
## [217] DECEASED LIVING LIVING LIVING DECEASED DECEASED  
## Levels: DECEASED LIVING

factor(Bc.data$Tumor\_Stage)

## [1] 2 1 2 3 1 2 2 2 2 2 2 2 2 3 2 2 2 2 2 2 2 2 1 2 2 2 2 2 1 2 1 2 1 2 1 2 3  
## [38] 1 1 2 1 3 2 1 1 2 3 3 1 1 1 3 2 2 2 1 2 1 2 2 2 2 2 3 2 1 1 2 2 1 3 2 2 2  
## [75] 2 2 2 1 2 1 2 1 1 2 2 1 2 1 3 2 2 3 2 3 2 2 3 1 2 2 2 2 1 3 2 3 2 1 2 1 2  
## [112] 1 2 3 1 2 2 1 1 2 2 2 1 1 2 2 1 1 1 1 2 2 3 2 2 2 1 1 2 2 2 2 2 2 2 2 2 1  
## [149] 2 2 1 2 2 2 2 2 2 1 2 2 3 1 2 1 2 1 2 2 2 2 1 2 2 2 2 1 2 1 2 2 3 1 2 2 1  
## [186] 1 3 1 2 1 2 2 2 1 2 2 3 1 1 1 2 2 2 1 2 1 2 1 2 2 1 1 3 2 3 2 2 2 2 2 1 2  
## Levels: 1 2 3

factor(Bc.data$Surgery\_Type)

## [1] MASTECTOMY MASTECTOMY MASTECTOMY MASTECTOMY   
## [5] MASTECTOMY MASTECTOMY MASTECTOMY BREAST CONSERVING  
## [9] BREAST CONSERVING BREAST CONSERVING MASTECTOMY MASTECTOMY   
## [13] BREAST CONSERVING MASTECTOMY MASTECTOMY BREAST CONSERVING  
## [17] MASTECTOMY MASTECTOMY BREAST CONSERVING MASTECTOMY   
## [21] MASTECTOMY MASTECTOMY BREAST CONSERVING BREAST CONSERVING  
## [25] MASTECTOMY MASTECTOMY MASTECTOMY MASTECTOMY   
## [29] MASTECTOMY BREAST CONSERVING MASTECTOMY MASTECTOMY   
## [33] MASTECTOMY MASTECTOMY BREAST CONSERVING BREAST CONSERVING  
## [37] MASTECTOMY BREAST CONSERVING BREAST CONSERVING BREAST CONSERVING  
## [41] BREAST CONSERVING MASTECTOMY MASTECTOMY MASTECTOMY   
## [45] BREAST CONSERVING MASTECTOMY MASTECTOMY MASTECTOMY   
## [49] BREAST CONSERVING BREAST CONSERVING MASTECTOMY MASTECTOMY   
## [53] MASTECTOMY BREAST CONSERVING BREAST CONSERVING BREAST CONSERVING  
## [57] BREAST CONSERVING BREAST CONSERVING MASTECTOMY MASTECTOMY   
## [61] MASTECTOMY MASTECTOMY MASTECTOMY MASTECTOMY   
## [65] MASTECTOMY BREAST CONSERVING BREAST CONSERVING BREAST CONSERVING  
## [69] BREAST CONSERVING MASTECTOMY MASTECTOMY BREAST CONSERVING  
## [73] MASTECTOMY MASTECTOMY BREAST CONSERVING MASTECTOMY   
## [77] MASTECTOMY BREAST CONSERVING MASTECTOMY BREAST CONSERVING  
## [81] BREAST CONSERVING MASTECTOMY BREAST CONSERVING MASTECTOMY   
## [85] MASTECTOMY BREAST CONSERVING MASTECTOMY MASTECTOMY   
## [89] MASTECTOMY MASTECTOMY MASTECTOMY BREAST CONSERVING  
## [93] MASTECTOMY MASTECTOMY MASTECTOMY MASTECTOMY   
## [97] BREAST CONSERVING MASTECTOMY MASTECTOMY BREAST CONSERVING  
## [101] MASTECTOMY MASTECTOMY BREAST CONSERVING BREAST CONSERVING  
## [105] MASTECTOMY MASTECTOMY MASTECTOMY BREAST CONSERVING  
## [109] MASTECTOMY BREAST CONSERVING MASTECTOMY BREAST CONSERVING  
## [113] MASTECTOMY MASTECTOMY MASTECTOMY MASTECTOMY   
## [117] MASTECTOMY BREAST CONSERVING MASTECTOMY MASTECTOMY   
## [121] MASTECTOMY MASTECTOMY BREAST CONSERVING BREAST CONSERVING  
## [125] BREAST CONSERVING MASTECTOMY MASTECTOMY MASTECTOMY   
## [129] BREAST CONSERVING MASTECTOMY BREAST CONSERVING MASTECTOMY   
## [133] MASTECTOMY MASTECTOMY BREAST CONSERVING BREAST CONSERVING  
## [137] BREAST CONSERVING MASTECTOMY BREAST CONSERVING MASTECTOMY   
## [141] MASTECTOMY MASTECTOMY BREAST CONSERVING BREAST CONSERVING  
## [145] BREAST CONSERVING BREAST CONSERVING BREAST CONSERVING BREAST CONSERVING  
## [149] MASTECTOMY MASTECTOMY BREAST CONSERVING MASTECTOMY   
## [153] MASTECTOMY MASTECTOMY MASTECTOMY BREAST CONSERVING  
## [157] BREAST CONSERVING BREAST CONSERVING MASTECTOMY BREAST CONSERVING  
## [161] BREAST CONSERVING BREAST CONSERVING MASTECTOMY BREAST CONSERVING  
## [165] MASTECTOMY MASTECTOMY MASTECTOMY BREAST CONSERVING  
## [169] BREAST CONSERVING BREAST CONSERVING BREAST CONSERVING BREAST CONSERVING  
## [173] BREAST CONSERVING BREAST CONSERVING BREAST CONSERVING BREAST CONSERVING  
## [177] BREAST CONSERVING MASTECTOMY MASTECTOMY MASTECTOMY   
## [181] MASTECTOMY BREAST CONSERVING BREAST CONSERVING BREAST CONSERVING  
## [185] BREAST CONSERVING BREAST CONSERVING BREAST CONSERVING BREAST CONSERVING  
## [189] MASTECTOMY BREAST CONSERVING BREAST CONSERVING BREAST CONSERVING  
## [193] BREAST CONSERVING MASTECTOMY MASTECTOMY MASTECTOMY   
## [197] BREAST CONSERVING MASTECTOMY BREAST CONSERVING BREAST CONSERVING  
## [201] BREAST CONSERVING BREAST CONSERVING MASTECTOMY BREAST CONSERVING  
## [205] BREAST CONSERVING BREAST CONSERVING BREAST CONSERVING BREAST CONSERVING  
## [209] BREAST CONSERVING BREAST CONSERVING BREAST CONSERVING BREAST CONSERVING  
## [213] MASTECTOMY BREAST CONSERVING MASTECTOMY MASTECTOMY   
## [217] BREAST CONSERVING MASTECTOMY BREAST CONSERVING BREAST CONSERVING  
## [221] MASTECTOMY MASTECTOMY   
## Levels: BREAST CONSERVING MASTECTOMY