# Problem Statement

1. What should be the output of the following Script?

v <- c( 2,5.5,6) t <- c(8, 3, 4) print(v%/%t)

> v <- c( 2,5.5,6)

> v

[1] 2.0 5.5 6.0

> t <- c(8, 3, 4)

> t

[1] 8 3 4

> print(v%/%t)

[1] 0 1 1

1. Q You have 25 excel files with names as xx\_1.xlsx, xx\_2.xlsx,……..xx\_25.xlsx in a dir.

Write a program to extract the contents of each excel sheet and make it one df.

##2. You have 25 excel files with names as xx\_1.xlsx, xx\_2.xlsx,....xx\_25.xlsx in a dir.

#Write a program to extract the contents of each excel sheet and make it one df.

getwd()

setwd("D:\\BIG DATA\\DATA ANALYTICS WITH R, EXCEL & TABLEAU\\1 INTRODUCTION")

getwd()

library(xlsx)

library(readxl)

**file.list <- list.files(pattern = "\*.xlsx")**

> file.list <- list.files(pattern = "\*.xlsx")

**file.list**

> file.list

[1] "xx\_1.xlsx" "xx\_10.xlsx" "xx\_2.xlsx" "xx\_3.xlsx" "xx\_4.xlsx" "xx\_5.xlsx"

[7] "xx\_6.xlsx" "xx\_7.xlsx" "xx\_8.xlsx" "xx\_9.xlsx"

**df.list <- lapply(file.list, read\_excel)**

**df.list**

|  |
| --- |
| [[1]]  # A tibble: 6 x 3  X Y Z  *<dbl>* *<dbl>* *<dbl>*  1 1.00 3.00 5.00  2 2.00 4.00 6.00  3 3.00 5.00 7.00  4 4.00 6.00 8.00  5 5.00 7.00 9.00  6 6.00 8.00 10.0  [[2]]  # A tibble: 6 x 3  A B C  *<dbl>* *<dbl>* *<dbl>*  1 1000 1100 1200  2 1001 1101 1201  3 1002 1102 1202  4 1003 1103 1203  5 1004 1104 1204  6 1005 1105 1205  [[3]]  # A tibble: 6 x 3  A B C  *<dbl>* *<dbl>* *<dbl>*  1 11.0 13.0 15.0  2 12.0 14.0 16.0  3 13.0 15.0 17.0  4 14.0 16.0 18.0  5 15.0 17.0 19.0  6 16.0 18.0 20.0  [[4]]  # A tibble: 6 x 3  X Y Z  *<dbl>* *<dbl>* *<dbl>*  1 21.0 27.0 33.0  2 22.0 28.0 34.0  3 23.0 29.0 35.0  4 24.0 30.0 36.0  5 25.0 31.0 37.0  6 26.0 32.0 38.0  [[5]]  # A tibble: 6 x 3  A B C  *<dbl>* *<dbl>* *<dbl>*  1 110 111 112  2 111 112 113  3 112 113 114  4 113 114 115  5 114 115 116  6 115 116 117  [[6]]  # A tibble: 6 x 3  X Y Z  *<dbl>* *<dbl>* *<dbl>*  1 31.0 32.0 33.0  2 32.0 33.0 34.0  3 33.0 34.0 35.0  4 34.0 35.0 36.0  5 35.0 36.0 37.0  6 36.0 37.0 38.0  [[7]]  # A tibble: 6 x 3  A B C  *<dbl>* *<dbl>* *<dbl>*  1 151 141 131  2 152 142 132  3 153 143 133  4 154 144 134  5 155 145 135  6 156 146 136  [[8]]  # A tibble: 6 x 3  X Y Z  *<dbl>* *<dbl>* *<dbl>*  1 41.0 42.0 43.0  2 51.0 52.0 53.0  3 61.0 62.0 63.0  4 45.0 46.0 47.0  5 55.0 56.0 57.0  6 65.0 66.0 67.0  [[9]]  # A tibble: 6 x 3  A B C  *<dbl>* *<dbl>* *<dbl>*  1 112 113 114  2 121 122 123  3 132 133 134  4 143 144 145  5 154 155 156  6 165 166 167  [[10]]  # A tibble: 6 x 3  X Y Z  *<dbl>* *<dbl>* *<dbl>*  1 61.0 70.0 81.0  2 62.0 71.0 82.0  3 63.0 72.0 83.0  4 64.0 73.0 84.0  5 65.0 74.0 85.0  6 66.0 75.0 86.0 |
|  |
| |  | | --- | | > | |

**df1 <- rbind.data.frame(df.list,idrow=TRUE)**

**df1**

|  |
| --- |
| > df1 <- rbind.data.frame(df.list,idrow=TRUE)  > df1  X Y Z NA. A B C NA..1 A.1 B.1 C.1 NA..2 X.1 Y.1 Z.1 NA..3 A.2 B.2 C.2  1 1 3 5 TRUE 1000 1100 1200 TRUE 11 13 15 TRUE 21 27 33 TRUE 110 111 112  2 2 4 6 TRUE 1001 1101 1201 TRUE 12 14 16 TRUE 22 28 34 TRUE 111 112 113  3 3 5 7 TRUE 1002 1102 1202 TRUE 13 15 17 TRUE 23 29 35 TRUE 112 113 114  4 4 6 8 TRUE 1003 1103 1203 TRUE 14 16 18 TRUE 24 30 36 TRUE 113 114 115  5 5 7 9 TRUE 1004 1104 1204 TRUE 15 17 19 TRUE 25 31 37 TRUE 114 115 116  6 6 8 10 TRUE 1005 1105 1205 TRUE 16 18 20 TRUE 26 32 38 TRUE 115 116 117  NA..4 X.2 Y.2 Z.2 NA..5 A.3 B.3 C.3 NA..6 X.3 Y.3 Z.3 NA..7 A.4 B.4 C.4 NA..8 X.4  1 TRUE 31 32 33 TRUE 151 141 131 TRUE 41 42 43 TRUE 112 113 114 TRUE 61  2 TRUE 32 33 34 TRUE 152 142 132 TRUE 51 52 53 TRUE 121 122 123 TRUE 62  3 TRUE 33 34 35 TRUE 153 143 133 TRUE 61 62 63 TRUE 132 133 134 TRUE 63  4 TRUE 34 35 36 TRUE 154 144 134 TRUE 45 46 47 TRUE 143 144 145 TRUE 64  5 TRUE 35 36 37 TRUE 155 145 135 TRUE 55 56 57 TRUE 154 155 156 TRUE 65  6 TRUE 36 37 38 TRUE 156 146 136 TRUE 65 66 67 TRUE 165 166 167 TRUE 66  Y.4 Z.4 NA..9  1 70 81 TRUE  2 71 82 TRUE  3 72 83 TRUE  4 73 84 TRUE  5 74 85 TRUE  6 75 86 TRUE |
|  |
| |  | | --- | | > | |

**df2 <- cbind.data.frame(df.list, idcol=TRUE)**

**df2**

|  |
| --- |
| > df2 <- cbind.data.frame(df.list, idcol=TRUE)  > df2  X Y Z A B C A B C X Y Z A B C X Y Z A B C X Y Z  1 1 3 5 1000 1100 1200 11 13 15 21 27 33 110 111 112 31 32 33 151 141 131 41 42 43  2 2 4 6 1001 1101 1201 12 14 16 22 28 34 111 112 113 32 33 34 152 142 132 51 52 53  3 3 5 7 1002 1102 1202 13 15 17 23 29 35 112 113 114 33 34 35 153 143 133 61 62 63  4 4 6 8 1003 1103 1203 14 16 18 24 30 36 113 114 115 34 35 36 154 144 134 45 46 47  5 5 7 9 1004 1104 1204 15 17 19 25 31 37 114 115 116 35 36 37 155 145 135 55 56 57  6 6 8 10 1005 1105 1205 16 18 20 26 32 38 115 116 117 36 37 38 156 146 136 65 66 67  A B C X Y Z idcol  1 112 113 114 61 70 81 TRUE  2 121 122 123 62 71 82 TRUE  3 132 133 134 63 72 83 TRUE  4 143 144 145 64 73 84 TRUE  5 154 155 156 65 74 85 TRUE  6 165 166 167 66 75 86 TRUE |
|  |
| |  | | --- | | > | |

1. If the above 25 files were csv files, what would be your script to read?

getwd()

setwd("D:\\BIG DATA\\DATA ANALYTICS WITH R, EXCEL & TABLEAU\\1 INTRODUCTION")

getwd()

library(readxl)

**temp = list.files(pattern="\*.csv")**

**temp**

> temp = list.files(pattern="\*.csv")

> temp

[1] "xx\_1.csv" "xx\_10.csv" "xx\_2.csv" "xx\_3.csv" "xx\_4.csv" "xx\_5.csv"

[7] "x

**myfiles = lapply(temp, read.delim)**

**myfiles**

|  |
| --- |
| > myfiles = lapply(temp, read.delim)  > myfiles  [[1]]  ï..X.Y.Z  1 1,3,5  2 2,4,6  3 3,5,7  4 4,6,8  5 5,7,9  6 6,8,10  [[2]]  ï..A.B.C  1 1000,1100,1200  2 1001,1101,1201  3 1002,1102,1202  4 1003,1103,1203  5 1004,1104,1204  6 1005,1105,1205  [[3]]  ï..A.B.C  1 11,13,15  2 12,14,16  3 13,15,17  4 14,16,18  5 15,17,19  6 16,18,20  [[4]]  ï..X.Y.Z  1 21,27,33  2 22,28,34  3 23,29,35  4 24,30,36  5 25,31,37  6 26,32,38  [[5]]  ï..A.B.C  1 110,111,112  2 111,112,113  3 112,113,114  4 113,114,115  5 114,115,116  6 115,116,117  [[6]]  ï..X.Y.Z  1 31,32,33  2 32,33,34  3 33,34,35  4 34,35,36  5 35,36,37  6 36,37,38  [[7]]  ï..A.B.C  1 151,141,131  2 152,142,132  3 153,143,133  4 154,144,134  5 155,145,135  6 156,146,136  [[8]]  ï..X.Y.Z  1 41,42,43  2 51,52,53  3 61,62,63  4 45,46,47  5 55,56,57  6 65,66,67  [[9]]  ï..A.B.C  1 112,113,114  2 121,122,123  3 132,133,134  4 143,144,145  5 154,155,156  6 165,166,167  [[10]]  ï..X.Y.Z  1 61,70,81  2 62,71,82  3 63,72,83  4 64,73,84  5 65,74,85  6 66,75,86 |
|  |
| |  | | --- | | > | |

**df1 <- rbind.data.frame(myfiles, idcol=TRUE)**

**df1**

|  |
| --- |
| > df1 <- rbind.data.frame(myfiles, idcol=TRUE)  > df1  ï..X.Y.Z NA. ï..A.B.C NA..1 ï..A.B.C.1 NA..2 ï..X.Y.Z.1 NA..3 ï..A.B.C.2  1 1,3,5 TRUE 1000,1100,1200 TRUE 11,13,15 TRUE 21,27,33 TRUE 110,111,112  2 2,4,6 TRUE 1001,1101,1201 TRUE 12,14,16 TRUE 22,28,34 TRUE 111,112,113  3 3,5,7 TRUE 1002,1102,1202 TRUE 13,15,17 TRUE 23,29,35 TRUE 112,113,114  4 4,6,8 TRUE 1003,1103,1203 TRUE 14,16,18 TRUE 24,30,36 TRUE 113,114,115  5 5,7,9 TRUE 1004,1104,1204 TRUE 15,17,19 TRUE 25,31,37 TRUE 114,115,116  6 6,8,10 TRUE 1005,1105,1205 TRUE 16,18,20 TRUE 26,32,38 TRUE 115,116,117  NA..4 ï..X.Y.Z.2 NA..5 ï..A.B.C.3 NA..6 ï..X.Y.Z.3 NA..7 ï..A.B.C.4 NA..8  1 TRUE 31,32,33 TRUE 151,141,131 TRUE 41,42,43 TRUE 112,113,114 TRUE  2 TRUE 32,33,34 TRUE 152,142,132 TRUE 51,52,53 TRUE 121,122,123 TRUE  3 TRUE 33,34,35 TRUE 153,143,133 TRUE 61,62,63 TRUE 132,133,134 TRUE  4 TRUE 34,35,36 TRUE 154,144,134 TRUE 45,46,47 TRUE 143,144,145 TRUE  5 TRUE 35,36,37 TRUE 155,145,135 TRUE 55,56,57 TRUE 154,155,156 TRUE  6 TRUE 36,37,38 TRUE 156,146,136 TRUE 65,66,67 TRUE 165,166,167 TRUE  ï..X.Y.Z.4 NA..9  1 61,70,81 TRUE  2 62,71,82 TRUE  3 63,72,83 TRUE  4 64,73,84 TRUE  5 65,74,85 TRUE  6 66,75,86 TRUE |
|  |
| |  | | --- | | > | |

**df2 <- cbind.data.frame(myfiles, idcol=TRUE)**

**df2**

|  |
| --- |
| > df2 <- cbind.data.frame(myfiles, idcol=TRUE)  > df2  ï..X.Y.Z ï..A.B.C ï..A.B.C ï..X.Y.Z ï..A.B.C ï..X.Y.Z ï..A.B.C  1 1,3,5 1000,1100,1200 11,13,15 21,27,33 110,111,112 31,32,33 151,141,131  2 2,4,6 1001,1101,1201 12,14,16 22,28,34 111,112,113 32,33,34 152,142,132  3 3,5,7 1002,1102,1202 13,15,17 23,29,35 112,113,114 33,34,35 153,143,133  4 4,6,8 1003,1103,1203 14,16,18 24,30,36 113,114,115 34,35,36 154,144,134  5 5,7,9 1004,1104,1204 15,17,19 25,31,37 114,115,116 35,36,37 155,145,135  6 6,8,10 1005,1105,1205 16,18,20 26,32,38 115,116,117 36,37,38 156,146,136  ï..X.Y.Z ï..A.B.C ï..X.Y.Z idcol  1 41,42,43 112,113,114 61,70,81 TRUE  2 51,52,53 121,122,123 62,71,82 TRUE  3 61,62,63 132,133,134 63,72,83 TRUE  4 45,46,47 143,144,145 64,73,84 TRUE  5 55,56,57 154,155,156 65,74,85 TRUE  6 65,66,67 165,166,167 66,75,86 TRUE |
|  |
| |  | | --- | | > | |