**REMOVING NAs**

> test <- read.csv("hw1\_data.csv", na.string = "Not Available", stringsAsFactors = FALSE)

> ncol(test)

[1] 6

> nrow(test)

[1] 153

> colnames(test)

[1] "Ozone" "Solar.R" "Wind" "Temp" "Month" "Day"

>

head(test)

Ozone Solar.R Wind Temp Month Day

1 41 190 7.4 67 5 1

2 36 118 8.0 72 5 2

3 12 149 12.6 74 5 3

4 18 313 11.5 62 5 4

5 NA NA 14.3 56 5 5

6 28 NA 14.9 66 5 6

> tail(test)

Ozone Solar.R Wind Temp Month Day

148 14 20 16.6 63 9 25

149 30 193 6.9 70 9 26

150 NA 145 13.2 77 9 27

151 14 191 14.3 75 9 28

152 18 131 8.0 76 9 29

153 20 223 11.5 68 9 30

> str(test)

'data.frame': 153 obs. of 6 variables:

$ Ozone : int 41 36 12 18 NA 28 23 19 8 NA ...

$ Solar.R: int 190 118 149 313 NA NA 299 99 19 194 ...

$ Wind : num 7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 8.6 ...

$ Temp : int 67 72 74 62 56 66 65 59 61 69 ...

$ Month : int 5 5 5 5 5 5 5 5 5 5 ...

$ Day : int 1 2 3 4 5 6 7 8 9 10 ...

summary(test)

Ozone Solar.R Wind Temp Month

Min. : 1.00 Min. : 7.0 Min. : 1.700 Min. :56.00 Min. :5.000

1st Qu.: 18.00 1st Qu.:115.8 1st Qu.: 7.400 1st Qu.:72.00 1st Qu.:6.000

Median : 31.50 Median :205.0 Median : 9.700 Median :79.00 Median :7.000

Mean : 42.13 Mean :185.9 Mean : 9.958 Mean :77.88 Mean :6.993

3rd Qu.: 63.25 3rd Qu.:258.8 3rd Qu.:11.500 3rd Qu.:85.00 3rd Qu.:8.000

Max. :168.00 Max. :334.0 Max. :20.700 Max. :97.00 Max. :9.000

NA's :37 NA's :7

Day

Min. : 1.0

1st Qu.: 8.0

Median :16.0

Mean :15.8

3rd Qu.:23.0

Max. :31.0

> sapply(test, class)

Ozone Solar.R Wind Temp Month Day

"integer" "integer" "numeric" "integer" "integer" "integer"

> test\_Ozone <- test[, "Ozone"]

> test\_Ozone

[1] 41 36 12 18 NA 28 23 19 8 NA 7 16 11 14 18 14 34 6 30 11

[21] 1 11 4 32 NA NA NA 23 45 115 37 NA NA NA NA NA NA 29 NA 71

[41] 39 NA NA 23 NA NA 21 37 20 12 13 NA NA NA NA NA NA NA NA NA

[61] NA 135 49 32 NA 64 40 77 97 97 85 NA 10 27 NA 7 48 35 61 79

[81] 63 16 NA NA 80 108 20 52 82 50 64 59 39 9 16 78 35 66 122 89

[101] 110 NA NA 44 28 65 NA 22 59 23 31 44 21 9 NA 45 168 73 NA 76

[121] 118 84 85 96 78 73 91 47 32 20 23 21 24 44 21 28 9 13 46 18

[141] 13 24 16 13 23 36 7 14 30 NA 14 18 20

> test\_Solar.R <- test[, "Solar.R"]

> test\_Solar.R

[1] 190 118 149 313 NA NA 299 99 19 194 NA 256 290 274 65 334 307 78 322 44

[21] 8 320 25 92 66 266 NA 13 252 223 279 286 287 242 186 220 264 127 273 291

[41] 323 259 250 148 332 322 191 284 37 120 137 150 59 91 250 135 127 47 98 31

[61] 138 269 248 236 101 175 314 276 267 272 175 139 264 175 291 48 260 274 285 187

[81] 220 7 258 295 294 223 81 82 213 275 253 254 83 24 77 NA NA NA 255 229

[101] 207 222 137 192 273 157 64 71 51 115 244 190 259 36 255 212 238 215 153 203

[121] 225 237 188 167 197 183 189 95 92 252 220 230 259 236 259 238 24 112 237 224

[141] 27 238 201 238 14 139 49 20 193 145 191 131 223

> test\_Wind <- test[, "Wind"]

> test\_Wind

[1] 7.4 8.0 12.6 11.5 14.3 14.9 8.6 13.8 20.1 8.6 6.9 9.7 9.2 10.9 13.2 11.5

[17] 12.0 18.4 11.5 9.7 9.7 16.6 9.7 12.0 16.6 14.9 8.0 12.0 14.9 5.7 7.4 8.6

[33] 9.7 16.1 9.2 8.6 14.3 9.7 6.9 13.8 11.5 10.9 9.2 8.0 13.8 11.5 14.9 20.7

[49] 9.2 11.5 10.3 6.3 1.7 4.6 6.3 8.0 8.0 10.3 11.5 14.9 8.0 4.1 9.2 9.2

[65] 10.9 4.6 10.9 5.1 6.3 5.7 7.4 8.6 14.3 14.9 14.9 14.3 6.9 10.3 6.3 5.1

[81] 11.5 6.9 9.7 11.5 8.6 8.0 8.6 12.0 7.4 7.4 7.4 9.2 6.9 13.8 7.4 6.9

[97] 7.4 4.6 4.0 10.3 8.0 8.6 11.5 11.5 11.5 9.7 11.5 10.3 6.3 7.4 10.9 10.3

[113] 15.5 14.3 12.6 9.7 3.4 8.0 5.7 9.7 2.3 6.3 6.3 6.9 5.1 2.8 4.6 7.4

[129] 15.5 10.9 10.3 10.9 9.7 14.9 15.5 6.3 10.9 11.5 6.9 13.8 10.3 10.3 8.0 12.6

[145] 9.2 10.3 10.3 16.6 6.9 13.2 14.3 8.0 11.5

> test\_Month <- test[, "Month"]

> test\_Month

[1] 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 6 6 6 6

[42] 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7

[83] 7 7 7 7 7 7 7 7 7 7 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8

[124] 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9

> test\_Day <- test[, "Day"]

> test\_Day

[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

[28] 28 29 30 31 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

[55] 24 25 26 27 28 29 30 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

[82] 21 22 23 24 25 26 27 28 29 30 31 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

[109] 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 1 2 3 4 5 6 7 8 9 10 11 12

[136] 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

Sometimes R reads a column as a vector of characters rather than integers....a subtle difference with big consequences

> class(test$Ozone)

[1] "character"

> sapply(test, class)

Ozone Solar.R Wind Temp Month Day

"character" "character" "numeric" "integer" "integer" "integer"

|  |
| --- |
| > test\_Ozone <- test[, "Ozone"]  > test\_Ozone  [1] "41" "36" "12" "18" "NA" "28" "23" "19" "8" "NA" "7" "16" "11"  [14] "14" "18" "14" "34" "6" "30" "11" "1" "11" "4" "32" "NA" "NA"  [27] "NA" "23" "45" "115" "37" "NA" "NA" "NA" "NA" "NA" "NA" "29" "NA"  [40] "71" "39" "NA" "NA" "23" "NA" "NA" "21" "37" "20" "12" "13" "NA"  [53] "NA" "NA" "NA" "NA" "NA" "NA" "NA" "NA" "NA" "135" "49" "32" "NA"  [66] "64" "40" "77" "97" "97" "85" "NA" "10" "27" "NA" "7" "48" "35"  [79] "61" "79" "63" "16" "NA" "NA" "80" "108" "20" "52" "82" "50" "64"  [92] "59" "39" "9" "16" "78" "35" "66" "122" "89" "110" "NA" "NA" "44"  [105] "28" "65" "NA" "22" "59" "23" "31" "44" "21" "9" "NA" "45" "168"  [118] "73" "NA" "76" "118" "84" "85" "96" "78" "73" "91" "47" "32" "20"  [131] "23" "21" "24" "44" "21" "28" "9" "13" "46" "18" "13" "24" "16"  [144] "13" "23" "36" "7" "14" "30" "NA" "14" "18" "20" |

NAs are not recognized as NAs but rather as “NA”s.

is.na returns as false

If R reads a column as a vector of characters rather than integers.... is.na doesn’t work

> class(test$Ozone)

[1] "character"

> test\_Ozone\_NA <- is.na(test[, "Ozone"])

> test\_Ozone\_NA

[1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[14] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[27] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[40] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[53] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[66] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[79] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[92] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[105] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[118] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[131] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

[144] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

> test\_Ozone\_NA <- length(is.na(test$Ozone))

NAs are not recognized as NAs but rather as “NA”s.

is.na returns as false

> test\_Ozone\_NA

[1] 153

> myNAs <- sapply(test, is.na)

> myNAs

Ozone Solar.R Wind Temp Month Day

[1,] FALSE FALSE FALSE FALSE FALSE FALSE

[2,] FALSE FALSE FALSE FALSE FALSE FALSE

[3,] FALSE FALSE FALSE FALSE FALSE FALSE

[4,] FALSE FALSE FALSE FALSE FALSE FALSE

...

[152,] FALSE FALSE FALSE FALSE FALSE FALSE

[153,] FALSE FALSE FALSE FALSE FALSE FALSE

> sum(is.na(test$Ozone))

[1] 0

> any(is.na(test$Ozone))

[1] FALSE

> all(test$Ozone > 0)

[1] TRUE

> colSums(is.na(test))

Ozone Solar.R Wind Temp Month Day

0 0 0 0 0 0

When R reads a column as a vectors of integers or numerics rather than character.... NAs are recognized as NA

> test\_Ozone <- test[, "Ozone"]

> test\_Ozone

[1] 41 36 12 18 NA 28 23 19 8 NA 7 16 11 14 18 14 34 6 30 11

[21] 1 11 4 32 NA NA NA 23 45 115 37 NA NA NA NA NA NA 29 NA 71

[41] 39 NA NA 23 NA NA 21 37 20 12 13 NA NA NA NA NA NA NA NA NA

[61] NA 135 49 32 NA 64 40 77 97 97 85 NA 10 27 NA 7 48 35 61 79

[81] 63 16 NA NA 80 108 20 52 82 50 64 59 39 9 16 78 35 66 122 89

[101] 110 NA NA 44 28 65 NA 22 59 23 31 44 21 9 NA 45 168 73 NA 76

[121] 118 84 85 96 78 73 91 47 32 20 23 21 24 44 21 28 9 13 46 18

[141] 13 24 16 13 23 36 7 14 30 NA 14 18 20

Bad acts as template from which we can remove the NAs in that column in the data frame test

> bad <- is.na(test\_Ozone)

> clean\_test\_Ozone <- test\_Ozone[!bad]

> clean\_test\_Ozone

[1] 41 36 12 18 28 23 19 8 7 16 11 14 18 14 34 6 30 11 1 11

[21] 4 32 23 45 115 37 29 71 39 23 21 37 20 12 13 135 49 32 64 40

[41] 77 97 97 85 10 27 7 48 35 61 79 63 16 80 108 20 52 82 50 64

[61] 59 39 9 16 78 35 66 122 89 110 44 28 65 22 59 23 31 44 21 9

[81] 45 168 73 76 118 84 85 96 78 73 91 47 32 20 23 21 24 44 21 28

[101] 9 13 46 18 13 24 16 13 23 36 7 14 30 14 18 20

> myNAs <- sapply(test, is.na)

> myNAs

Ozone Solar.R Wind Temp Month Day

When vectors are numerics or integers, NAs are nor recognized as NAs

[1,] FALSE FALSE FALSE FALSE FALSE FALSE

...

[5,] TRUE TRUE FALSE FALSE FALSE FALSE

...

[9,] FALSE FALSE FALSE FALSE FALSE FALSE

[10,] TRUE FALSE FALSE FALSE FALSE FALSE

....

...

[25,] TRUE FALSE FALSE FALSE FALSE FALSE

[26,] TRUE FALSE FALSE FALSE FALSE FALSE

[27,] TRUE TRUE FALSE FALSE FALSE FALSE

...

[150,] TRUE FALSE FALSE FALSE FALSE FALSE

[151,] FALSE FALSE FALSE FALSE FALSE FALSE

[152,] FALSE FALSE FALSE FALSE FALSE FALSE

[153,] FALSE FALSE FALSE FALSE FALSE FALSE

When R reads a column as a vectors of integers or numerics rather than character.... NAs are recognized as NA

> sapply(test, class)

Ozone Solar.R Wind Temp Month Day

"character" "character" "numeric" "integer" "integer" "integer"

> sapply(test, function(a) length(a[is.na(a)]))

Ozone Solar.R Wind Temp Month Day

0 0 0 0 0 0

> sapply(test, function(a) length(a[!is.na(a)]))

Ozone Solar.R Wind Temp Month Day

153 153 153 153 153 153

> sapply(test, class)

Ozone Solar.R Wind Temp Month Day

"integer" "integer" "numeric" "integer" "integer" "integer"

> sapply(test, function(a) length(a[is.na(a)]))

Ozone Solar.R Wind Temp Month Day

37 7 0 0 0 0

> sapply(test, function(a) length(a[!is.na(a)]))

Ozone Solar.R Wind Temp Month Day

116 146 153 153 153 153

> sapply(test, function(a) a[!is.na(a)])

Character is denoted by double brackets - ”character”. NAs are recognized as character “NA”s and not as missing values.

Ozone Solar.R Wind Temp Month Day

[1,] "41" "190" "7.4" "67" "5" "1"

[2,] "36" "118" "8" "72" "5" "2"

[3,] "12" "149" "12.6" "74" "5" "3"

[4,] "18" "313" "11.5" "62" "5" "4"

[5,] "NA" "NA" "14.3" "56" "5" "5"

[6,] "28" "NA" "14.9" "66" "5" "6"

[7,] "23" "299" "8.6" "65" "5" "7"

[8,] "19" "99" "13.8" "59" "5" "8"

[9,] "8" "19" "20.1" "61" "5" "9"

Originally there were 7 NAs in the column Solar.R but we removed them, along with 2 NAs in the Ozone column which coincided with the columns of NAs in Solar.R

[10,] "NA" "194" "8.6" "69" "5" "10"

[11,] "7" "NA" "6.9" "74" "5" "11"

[12,] "16" "256" "9.7" "69" "5" "12"

> sapply(test, function(a) length(a[is.na(a)]))

Ozone Solar.R Wind Temp Month Day

37 7 0 0 0 0

> sapply(clean\_Solar.R, function(a) length(a[is.na(a)]))

Ozone Solar.R Wind Temp Month Day

35 0 0 0 0 0

> sapply(test, function(a) length(a))

Originally there were 153 columns in the test data frame but we removed 7 rows to where the 7 NAs in the Solar.R column was located

Ozone Solar.R Wind Temp Month Day

153 153 153 153 153 153

> sapply(clean\_Solar.R, function(a) length(a))

Ozone Solar.R Wind Temp Month Day

146 146 146 146 146 146

Comparing is.na with:

na. omit, and complete.cases...

Remember !

Clean\_Solar.R is the output we got when we removed the NAs from the original data frame **test** using the logical vector **bad** as template. A total of 7 rows were removed as there were as much NAs in that column

> sapply(test, length)

Ozone Solar.R Wind Temp Month Day

153 153 153 153 153 153

> sapply(test, function(a) length(a[is.na(a)]))

Ozone Solar.R Wind Temp Month Day

37 7 0 0 0 0

> sapply(clean\_Solar.R, function(a) length(a[is.na(a)]))

Ozone Solar.R Wind Temp Month Day

35 0 0 0 0 0

> sapply(clean\_Solar.R, function(a) length(a))

Ozone Solar.R Wind Temp Month Day

146 146 146 146 146 146

na. omit

> sapply(test, function(a) length(a))

Ozone Solar.R Wind Temp Month Day

153 153 153 153 153 153

> clean <- sapply(test, na.omit)

> clean

$Ozone

[1] 41 36 12 18 28 23 19 8 7 16 11 14 18 14 34 6 30 11 1 11

[21] 4 32 23 45 115 37 29 71 39 23 21 37 20 12 13 135 49 32 64 40

[41] 77 97 97 85 10 27 7 48 35 61 79 63 16 80 108 20 52 82 50 64

[61] 59 39 9 16 78 35 66 122 89 110 44 28 65 22 59 23 31 44 21 9

[81] 45 168 73 76 118 84 85 96 78 73 91 47 32 20 23 21 24 44 21 28

[101] 9 13 46 18 13 24 16 13 23 36 7 14 30 14 18 20

attr(,"na.action")

[1] 5 10 25 26 27 32 33 34 35 36 37 39 42 43 45 46 52 53 54 55

[21] 56 57 58 59 60 61 65 72 75 83 84 102 103 107 115 119 150

attr(,"class")

[1] "omit"

$Solar.R

[1] 190 118 149 313 299 99 19 194 256 290 274 65 334 307 78 322 44 8 320 25

[21] 92 66 266 13 252 223 279 286 287 242 186 220 264 127 273 291 323 259 250 148

[41] 332 322 191 284 37 120 137 150 59 91 250 135 127 47 98 31 138 269 248 236

[61] 101 175 314 276 267 272 175 139 264 175 291 48 260 274 285 187 220 7 258 295

[81] 294 223 81 82 213 275 253 254 83 24 77 255 229 207 222 137 192 273 157 64

[101] 71 51 115 244 190 259 36 255 212 238 215 153 203 225 237 188 167 197 183 189

[121] 95 92 252 220 230 259 236 259 238 24 112 237 224 27 238 201 238 14 139 49

[141] 20 193 145 191 131 223

attr(,"na.action")

[1] 5 6 11 27 96 97 98

attr(,"class")

[1] "omit"

$Wind

[1] 7.4 8.0 12.6 11.5 14.3 14.9 8.6 13.8 20.1 8.6 6.9 9.7 9.2 10.9 13.2 11.5

[17] 12.0 18.4 11.5 9.7 9.7 16.6 9.7 12.0 16.6 14.9 8.0 12.0 14.9 5.7 7.4 8.6

[33] 9.7 16.1 9.2 8.6 14.3 9.7 6.9 13.8 11.5 10.9 9.2 8.0 13.8 11.5 14.9 20.7

[49] 9.2 11.5 10.3 6.3 1.7 4.6 6.3 8.0 8.0 10.3 11.5 14.9 8.0 4.1 9.2 9.2

[65] 10.9 4.6 10.9 5.1 6.3 5.7 7.4 8.6 14.3 14.9 14.9 14.3 6.9 10.3 6.3 5.1

[81] 11.5 6.9 9.7 11.5 8.6 8.0 8.6 12.0 7.4 7.4 7.4 9.2 6.9 13.8 7.4 6.9

[97] 7.4 4.6 4.0 10.3 8.0 8.6 11.5 11.5 11.5 9.7 11.5 10.3 6.3 7.4 10.9 10.3

[113] 15.5 14.3 12.6 9.7 3.4 8.0 5.7 9.7 2.3 6.3 6.3 6.9 5.1 2.8 4.6 7.4

[129] 15.5 10.9 10.3 10.9 9.7 14.9 15.5 6.3 10.9 11.5 6.9 13.8 10.3 10.3 8.0 12.6

[145] 9.2 10.3 10.3 16.6 6.9 13.2 14.3 8.0 11.5

$Temp

[1] 67 72 74 62 56 66 65 59 61 69 74 69 66 68 58 64 66 57 68 62 59 73 61 61 57 58 57

[28] 67 81 79 76 78 74 67 84 85 79 82 87 90 87 93 92 82 80 79 77 72 65 73 76 77 76 76

[55] 76 75 78 73 80 77 83 84 85 81 84 83 83 88 92 92 89 82 73 81 91 80 81 82 84 87 85

[82] 74 81 82 86 85 82 86 88 86 83 81 81 81 82 86 85 87 89 90 90 92 86 86 82 80 79 77

[109] 79 76 78 78 77 72 75 79 81 86 88 97 94 96 94 91 92 93 93 87 84 80 78 75 73 81 76

[136] 77 71 71 78 67 76 68 82 64 71 81 69 63 70 77 75 76 68

$Month

[1] 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 6 6 6 6

[42] 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7

[83] 7 7 7 7 7 7 7 7 7 7 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8

[124] 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9

$Day

[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

[28] 28 29 30 31 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

[55] 24 25 26 27 28 29 30 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

[82] 21 22 23 24 25 26 27 28 29 30 31 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

[109] 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 1 2 3 4 5 6 7 8 9 10 11 12

[136] 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

Using na.omit in sapply will remove the NAs in your data frame but It will return a list.

> class(clean)

[1] "list"

> clean$Ozone

[1] 41 36 12 18 28 23 19 8 7 16 11 14 18 14 34 6 30 11 1 11

[21] 4 32 23 45 115 37 29 71 39 23 21 37 20 12 13 135 49 32 64 40

[41] 77 97 97 85 10 27 7 48 35 61 79 63 16 80 108 20 52 82 50 64

[61] 59 39 9 16 78 35 66 122 89 110 44 28 65 22 59 23 31 44 21 9

[81] 45 168 73 76 118 84 85 96 78 73 91 47 32 20 23 21 24 44 21 28

[101] 9 13 46 18 13 24 16 13 23 36 7 14 30 14 18 20

attr(,"na.action")

[1] 5 10 25 26 27 32 33 34 35 36 37 39 42 43 45 46 52 53 54 55

[21] 56 57 58 59 60 61 65 72 75 83 84 102 103 107 115 119 150

attr(,"class")

[1] "omit"

> names(clean)

[1] "Ozone" "Solar.R" "Wind" "Temp" "Month" "Day"

> sapply(clean, class)

Ozone Solar.R Wind Temp Month Day

"integer" "integer" "numeric" "integer" "integer" "integer"

Notice too that the elements of each element of the list are of different lengths since the original columns where they derived contained different numbers of Nas . The other values in the original columns are preserved in the list. Only the NAs are removed.

sapply(clean, length)

Ozone Solar.R Wind Temp Month Day

116 146 153 153 153 153

> sapply(test, complete.cases)

Using complete.cases in sapply will not remove the NAs in your data frame. It will just point to the location of the NAs in the columns of your data frame by returning a dataframe with elements of a logical vector of TRUE and FALSE. These will show you which rows contain NAs , those rows with FALSE.

If we use

> clean <- test[complete.cases(test), ]

We remove all the rows with NAs

> clean

Ozone Solar.R Wind Temp Month Day

1 41 190 7.4 67 5 1

2 36 118 8.0 72 5 2

3 12 149 12.6 74 5 3

4 18 313 11.5 62 5 4

7 23 299 8.6 65 5 7

8 19 99 13.8 59 5 8

9 8 19 20.1 61 5 9

12 16 256 9.7 69 5 12

13 11 290 9.2 66 5 13

14 14 274 10.9 68 5 14

15 18 65 13.2 58 5 15

16 14 334 11.5 64 5 16

17 34 307 12.0 66 5 17

18 6 78 18.4 57 5 18

19 30 322 11.5 68 5 19

20 11 44 9.7 62 5 20

21 1 8 9.7 59 5 21

22 11 320 16.6 73 5 22

23 4 25 9.7 61 5 23

24 32 92 12.0 61 5 24

28 23 13 12.0 67 5 28

29 45 252 14.9 81 5 29

30 115 223 5.7 79 5 30

31 37 279 7.4 76 5 31

38 29 127 9.7 82 6 7

40 71 291 13.8 90 6 9

41 39 323 11.5 87 6 10

44 23 148 8.0 82 6 13

47 21 191 14.9 77 6 16

48 37 284 20.7 72 6 17

49 20 37 9.2 65 6 18

50 12 120 11.5 73 6 19

...

Unintended effect - It also removes values that are not NAs that fall in rows with other NAs in other columns

> clean <- test[complete.cases(test), ]

> sapply(clean, length)

Ozone Solar.R Wind Temp Month Day

111 111 111 111 111 111

Only 7 NAs were present in the column Solar.R. Using complete .cases removed 42 rows.

Ozone Solar.R Wind Temp Month Day

[1,] TRUE TRUE TRUE TRUE TRUE TRUE

[2,] TRUE TRUE TRUE TRUE TRUE TRUE

[3,] TRUE TRUE TRUE TRUE TRUE TRUE

[4,] TRUE TRUE TRUE TRUE TRUE TRUE

[5,] FALSE FALSE TRUE TRUE TRUE TRUE

[6,] TRUE FALSE TRUE TRUE TRUE TRUE

[7,] TRUE TRUE TRUE TRUE TRUE TRUE

[8,] TRUE TRUE TRUE TRUE TRUE TRUE

[9,] TRUE TRUE TRUE TRUE TRUE TRUE

[10,] FALSE TRUE TRUE TRUE TRUE TRUE

[11,] TRUE FALSE TRUE TRUE TRUE TRUE

...

[25,] FALSE TRUE TRUE TRUE TRUE TRUE

[26,] FALSE TRUE TRUE TRUE TRUE TRUE

[27,] FALSE FALSE TRUE TRUE TRUE TRUE

[28,] TRUE TRUE TRUE TRUE TRUE TRUE

[29,] TRUE TRUE TRUE TRUE TRUE TRUE

[30,] TRUE TRUE TRUE TRUE TRUE TRUE

[31,] TRUE TRUE TRUE TRUE TRUE TRUE

[32,] FALSE TRUE TRUE TRUE TRUE TRUE

[33,] FALSE TRUE TRUE TRUE TRUE TRUE

[34,] FALSE TRUE TRUE TRUE TRUE TRUE

[35,] FALSE TRUE TRUE TRUE TRUE TRUE

[36,] FALSE TRUE TRUE TRUE TRUE TRUE

[37,] FALSE TRUE TRUE TRUE TRUE TRUE

[38,] TRUE TRUE TRUE TRUE TRUE TRUE

[39,] FALSE TRUE TRUE TRUE TRUE TRUE

[40,] TRUE TRUE TRUE TRUE TRUE TRUE

[41,] TRUE TRUE TRUE TRUE TRUE TRUE

[42,] FALSE TRUE TRUE TRUE TRUE TRUE

[43,] FALSE TRUE TRUE TRUE TRUE TRUE

[44,] TRUE TRUE TRUE TRUE TRUE TRUE

[45,] FALSE TRUE TRUE TRUE TRUE TRUE

[46,] FALSE TRUE TRUE TRUE TRUE TRUE

[47,] TRUE TRUE TRUE TRUE TRUE TRUE

[48,] TRUE TRUE TRUE TRUE TRUE TRUE

[49,] TRUE TRUE TRUE TRUE TRUE TRUE

[50,] TRUE TRUE TRUE TRUE TRUE TRUE

[51,] TRUE TRUE TRUE TRUE TRUE TRUE

[52,] FALSE TRUE TRUE TRUE TRUE TRUE

[53,] FALSE TRUE TRUE TRUE TRUE TRUE

[54,] FALSE TRUE TRUE TRUE TRUE TRUE

[55,] FALSE TRUE TRUE TRUE TRUE TRUE

[56,] FALSE TRUE TRUE TRUE TRUE TRUE

[57,] FALSE TRUE TRUE TRUE TRUE TRUE

[58,] FALSE TRUE TRUE TRUE TRUE TRUE

[59,] FALSE TRUE TRUE TRUE TRUE TRUE

[60,] FALSE TRUE TRUE TRUE TRUE TRUE

[61,] FALSE TRUE TRUE TRUE TRUE TRUE

[62,] TRUE TRUE TRUE TRUE TRUE TRUE

...

[150,] FALSE TRUE TRUE TRUE TRUE TRUE

[151,] TRUE TRUE TRUE TRUE TRUE TRUE

[152,] TRUE TRUE TRUE TRUE TRUE TRUE

[153,] TRUE TRUE TRUE TRUE TRUE TRUE