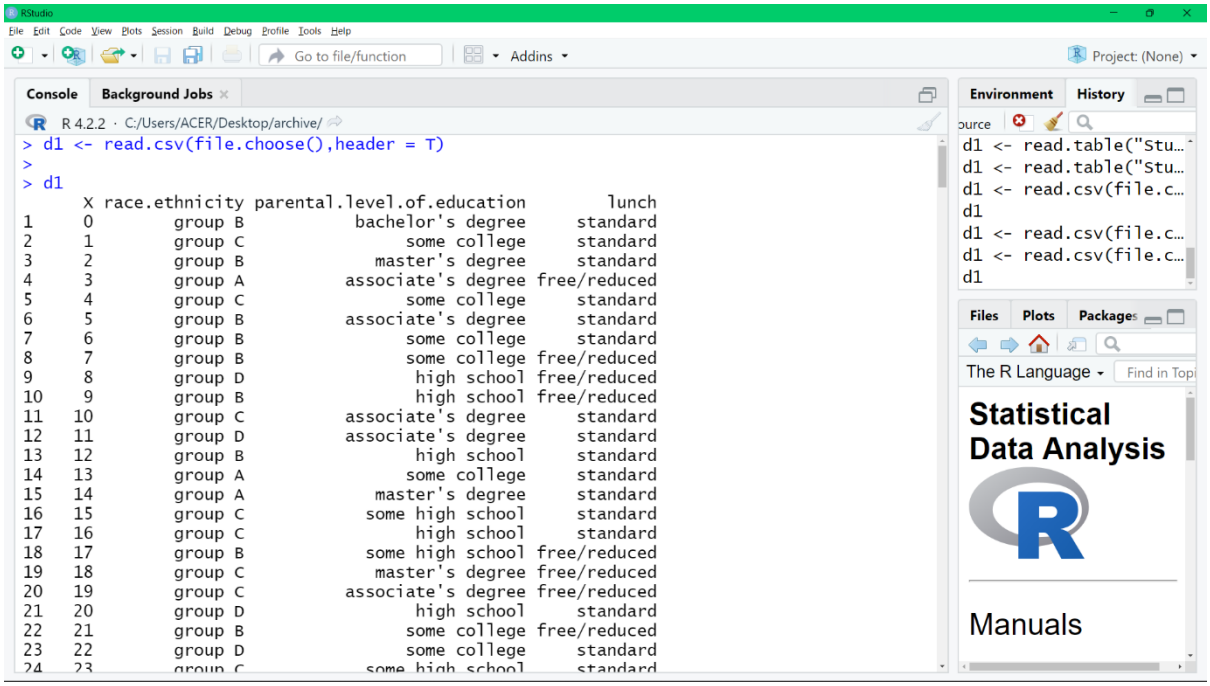
SCS2211 - LABORATORY II

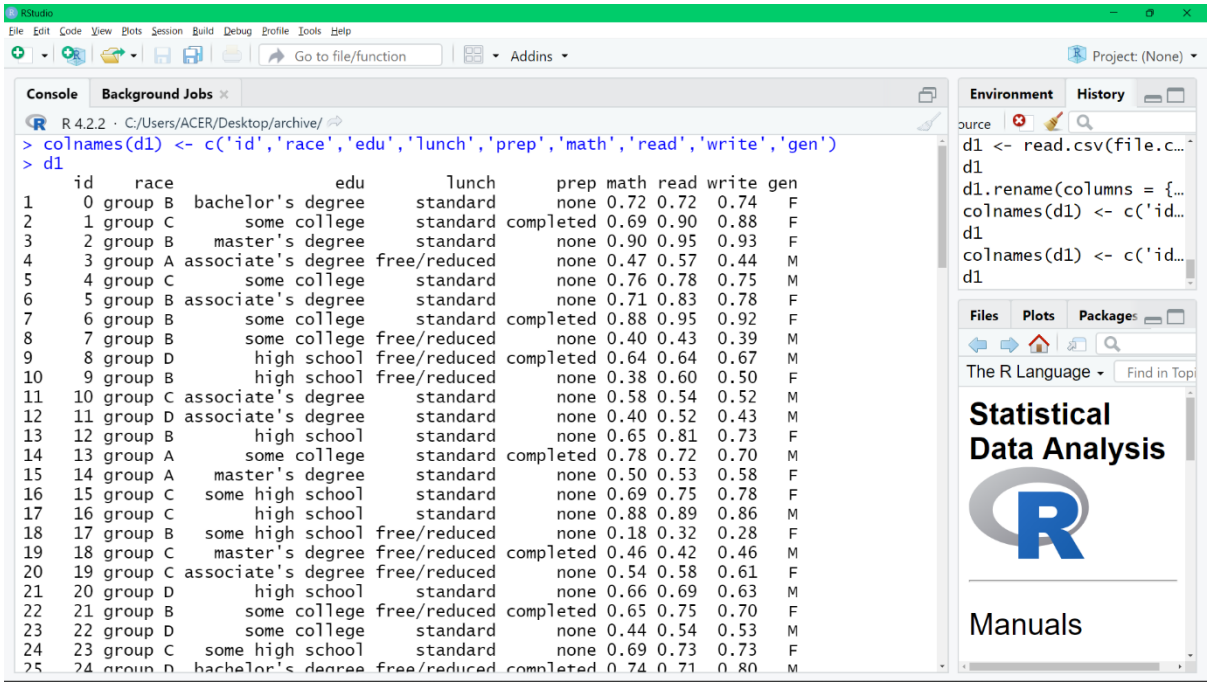
20000111

M.A.D.V.S.ANAWARATHNA

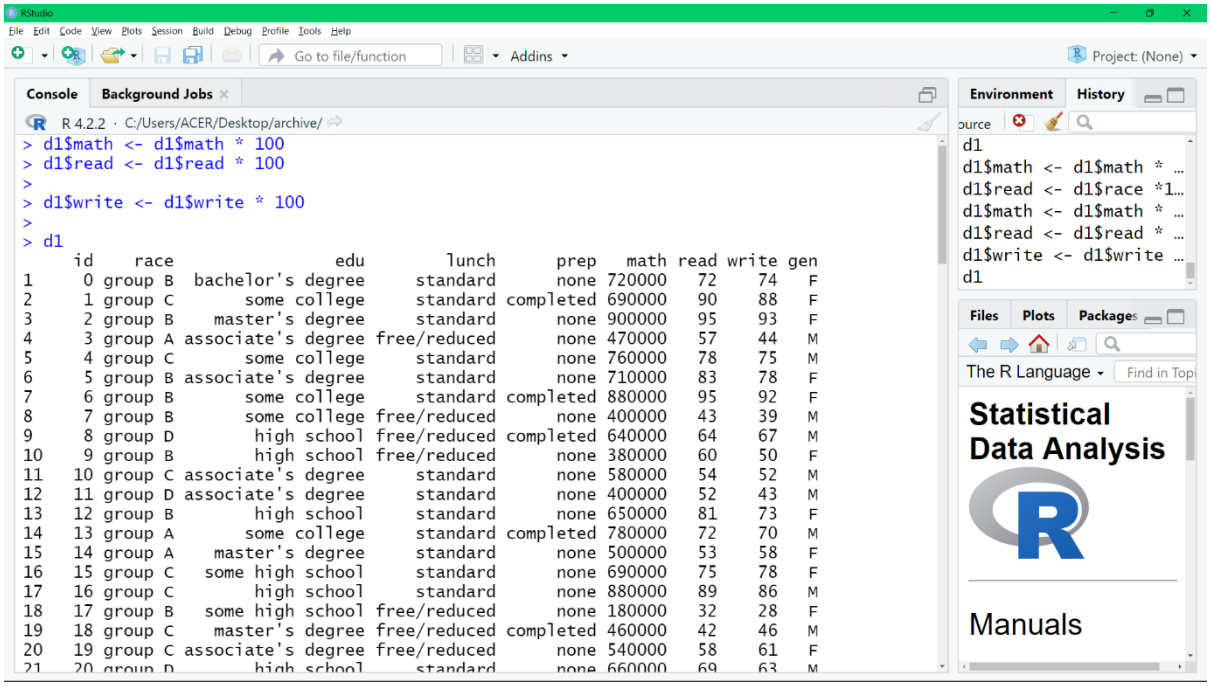
Question - 01



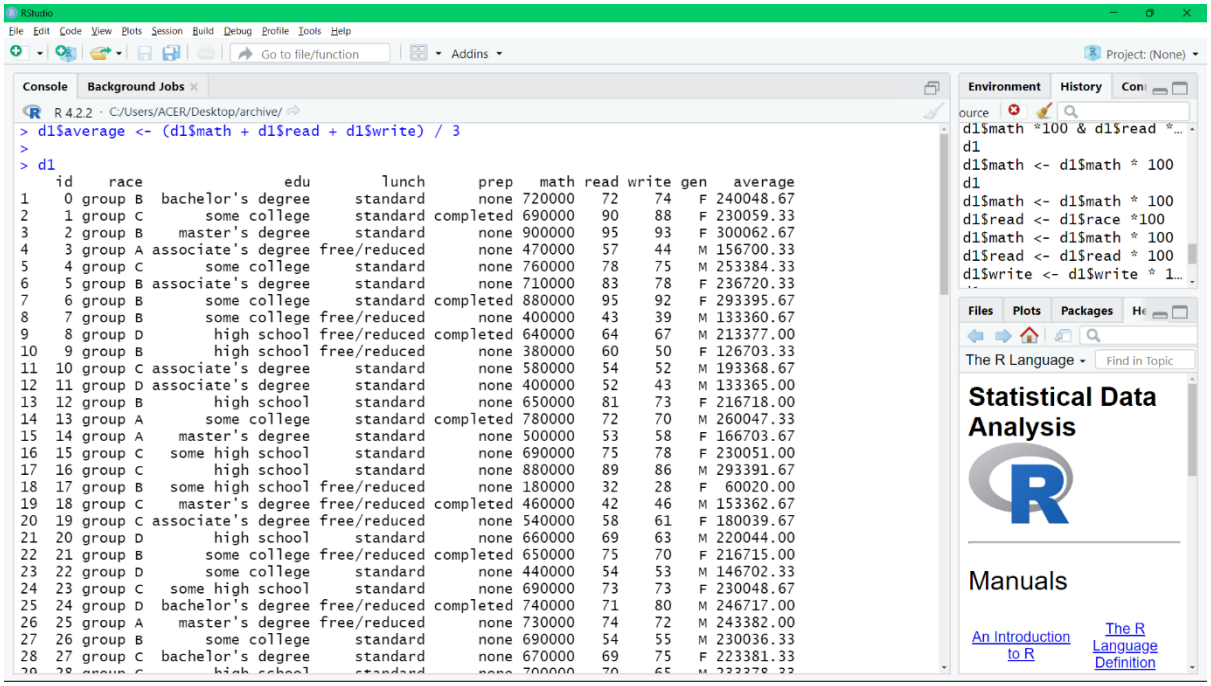
```
> d1 <- read.csv(file.choose(),header = T)
>
> d1
    X race.ethnicity parental.level.of.education        lunch
1   0        group B           bachelor's degree     standard
2   1        group C                some college     standard
3   2        group B             master's degree     standard
4   3        group A          associate's degree free/reduced
5   4        group C                some college     standard
6   5        group B          associate's degree     standard
7   6        group B                some college     standard
8   7        group B                some college free/reduced
9   8        group D                 high school free/reduced
10  9        group B                 high school free/reduced
11 10        group C          associate's degree     standard
12 11        group D          associate's degree     standard
13 12        group B                 high school     standard
14 13        group A                some college     standard
15 14        group A             master's degree     standard
16 15        group C              some high school     standard
17 16        group C                 high school     standard
18 17        group B              some high school free/reduced
19 18        group C             master's degree free/reduced
20 19        group C          associate's degree free/reduced
21 20        group D                 high school     standard
22 21        group B                some college free/reduced
23 22        group D                some college     standard
24 23        group C              some high school     standard
```



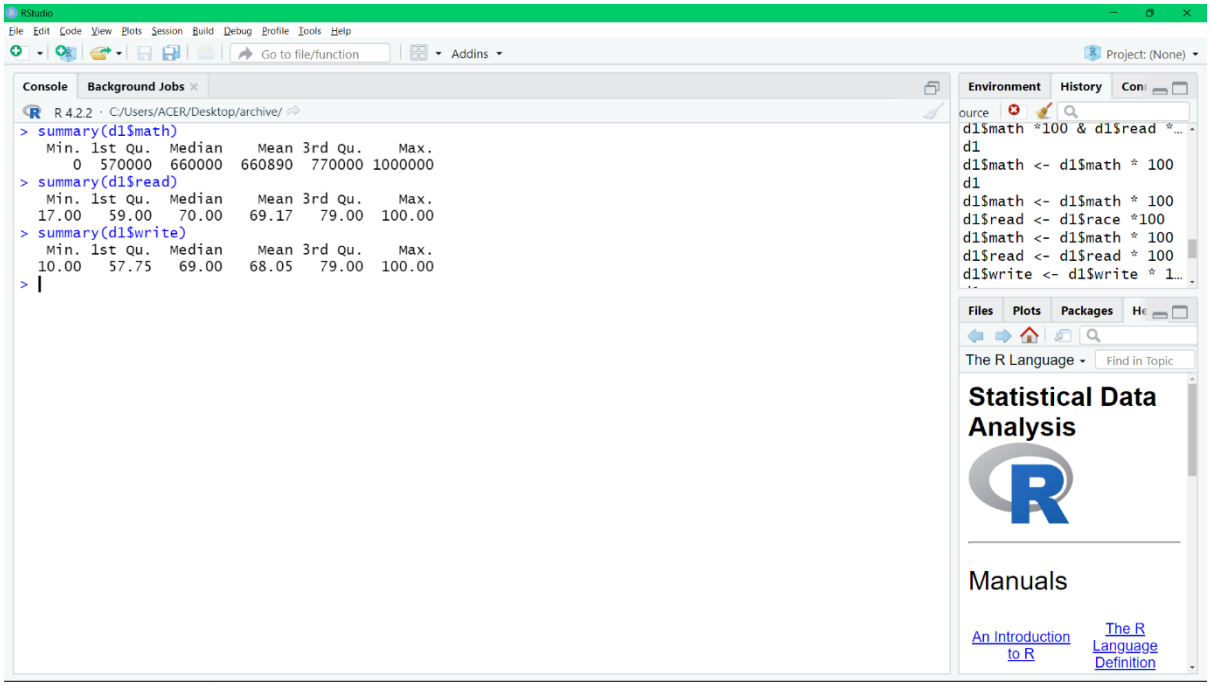
```
> colnames(d1) <- c('id','race','edu','lunch','prep','math','read','write','gen')
> d1
   id    race                edu        lunch      prep math read write gen
1   0 group B  bachelor's degree     standard      none 0.72 0.72  0.74   F
2   1 group C       some college     standard completed 0.69 0.90  0.88   F
3   2 group B    master's degree     standard      none 0.90 0.95  0.93   F
4   3 group A associate's degree free/reduced      none 0.47 0.57  0.44   M
5   4 group C       some college     standard      none 0.76 0.78  0.75   M
6   5 group B associate's degree     standard      none 0.71 0.83  0.78   F
7   6 group B       some college     standard completed 0.88 0.95  0.92   F
8   7 group B       some college free/reduced      none 0.40 0.43  0.39   M
9   8 group D        high school free/reduced completed 0.64 0.64  0.67   M
10  9 group B        high school free/reduced      none 0.38 0.60  0.50   F
11 10 group C associate's degree     standard      none 0.58 0.54  0.52   M
12 11 group D associate's degree     standard      none 0.40 0.52  0.43   M
13 12 group B        high school     standard      none 0.65 0.81  0.73   F
14 13 group A       some college     standard completed 0.78 0.72  0.70   M
15 14 group A    master's degree     standard      none 0.50 0.53  0.58   F
16 15 group C   some high school     standard      none 0.69 0.75  0.78   F
17 16 group C        high school     standard      none 0.88 0.89  0.86   M
18 17 group B   some high school free/reduced      none 0.18 0.32  0.28   F
19 18 group C    master's degree free/reduced completed 0.46 0.42  0.46   M
20 19 group C associate's degree free/reduced      none 0.54 0.58  0.61   F
21 20 group D        high school     standard      none 0.66 0.69  0.63   M
22 21 group B       some college free/reduced completed 0.65 0.75  0.70   F
23 22 group D       some college     standard      none 0.44 0.54  0.53   M
24 23 group C   some high school     standard      none 0.69 0.73  0.73   F
```

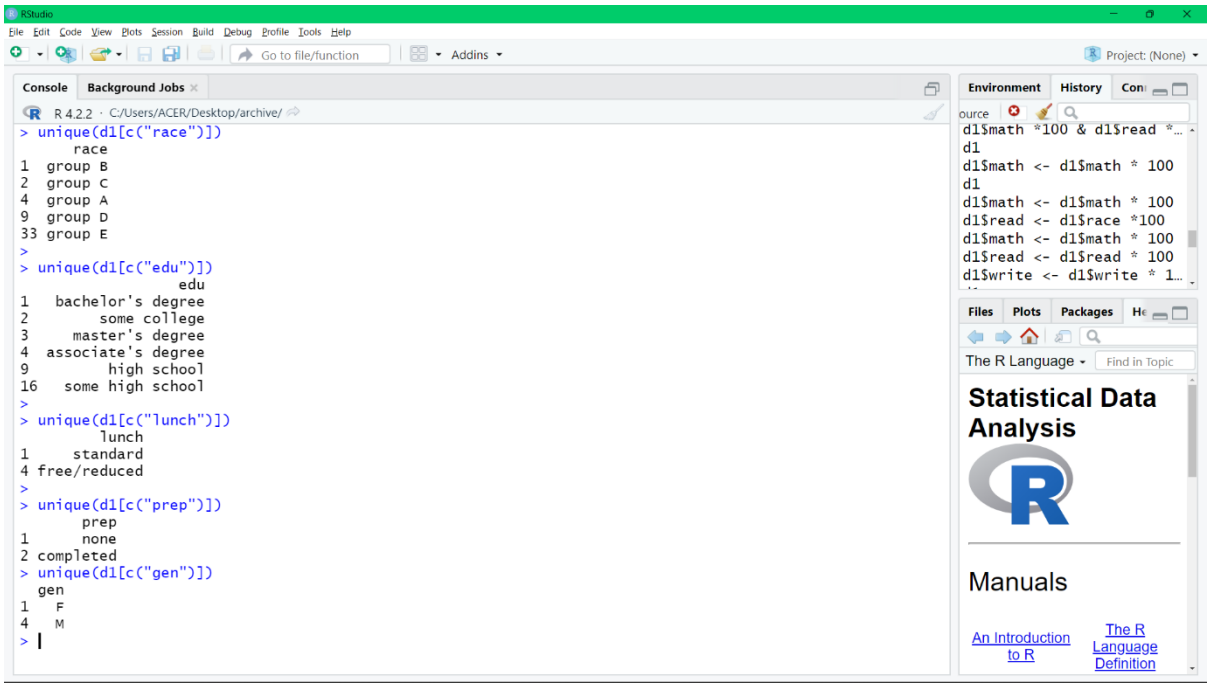


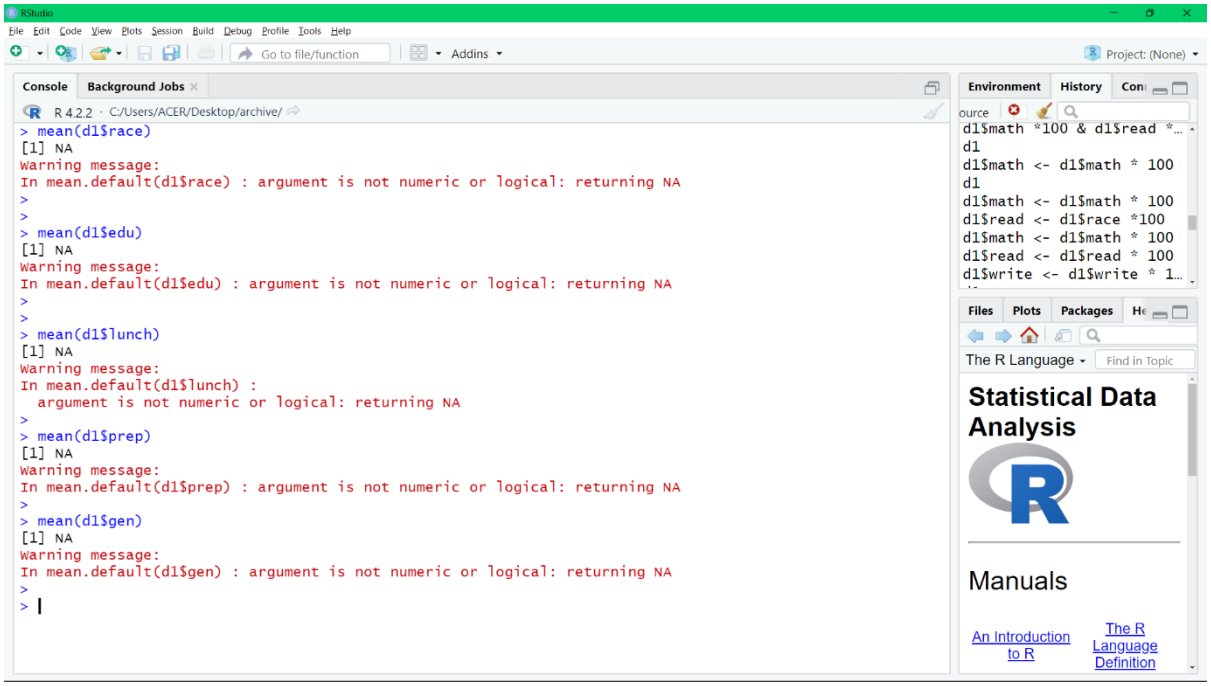
```
> d1$math <- d1$math * 100
> d1$read <- d1$read * 100
>
> d1$write <- d1$write * 100
>
> d1
    id    race                edu        lunch      prep   math read write gen
1    0 group B  bachelor's degree     standard      none 720000   72    74   F
2    1 group C       some college     standard completed 690000   90    88   F
3    2 group B    master's degree     standard      none 900000   95    93   F
4    3 group A associate's degree free/reduced      none 470000   57    44   M
5    4 group C       some college     standard      none 760000   78    75   M
6    5 group B associate's degree     standard      none 710000   83    78   F
7    6 group B       some college     standard completed 880000   95    92   F
8    7 group B       some college free/reduced      none 400000   43    39   M
9    8 group D        high school free/reduced completed 640000   64    67   M
10   9 group B        high school free/reduced      none 380000   60    50   F
11  10 group C associate's degree     standard      none 580000   54    52   M
12  11 group D associate's degree     standard      none 400000   52    43   M
13  12 group B        high school     standard      none 650000   81    73   F
14  13 group A       some college     standard completed 780000   72    70   M
15  14 group A    master's degree     standard      none 500000   53    58   F
16  15 group C   some high school     standard      none 690000   75    78   F
17  16 group C        high school     standard      none 880000   89    86   M
18  17 group B   some high school free/reduced      none 180000   32    28   F
19  18 group C    master's degree free/reduced completed 460000   42    46   M
20  19 group C associate's degree free/reduced      none 540000   58    61   F
21  20 group D        high school     standard      none 660000   69    63   M
```



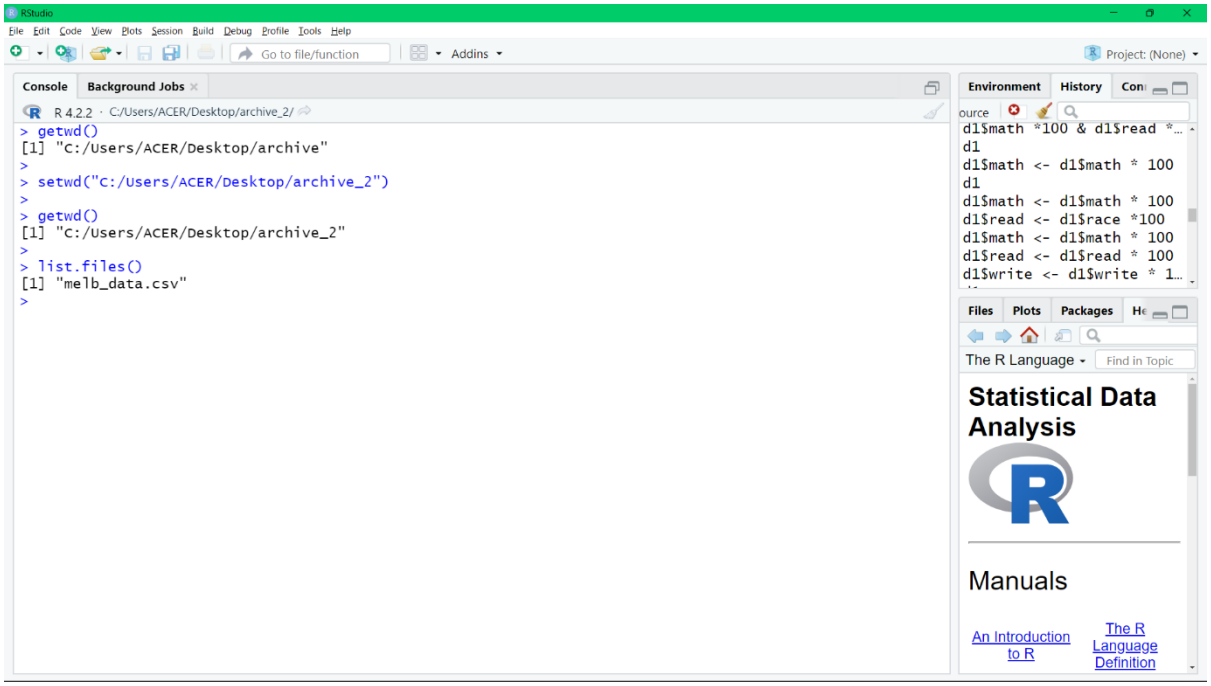
```
> d1$average <- (d1$math + d1$read + d1$write) / 3
>
> d1
    id    race                edu        lunch      prep   math read write gen   average
1    0 group B  bachelor's degree     standard      none 720000   72    74   F 240048.67
2    1 group C       some college     standard completed 690000   90    88   F 230059.33
3    2 group B    master's degree     standard      none 900000   95    93   F 300062.67
4    3 group A associate's degree free/reduced      none 470000   57    44   M 156700.33
5    4 group C       some college     standard      none 760000   78    75   M 253384.33
6    5 group B associate's degree     standard      none 710000   83    78   F 236720.33
7    6 group B       some college     standard completed 880000   95    92   F 293395.67
8    7 group B       some college free/reduced      none 400000   43    39   M 133360.67
9    8 group D        high school free/reduced completed 640000   64    67   M 213377.00
10   9 group B        high school free/reduced      none 380000   60    50   F 126703.33
11  10 group C associate's degree     standard      none 580000   54    52   M 193368.67
12  11 group D associate's degree     standard      none 400000   52    43   M 133365.00
13  12 group B        high school     standard      none 650000   81    73   F 216718.00
14  13 group A       some college     standard completed 780000   72    70   M 260047.33
15  14 group A    master's degree     standard      none 500000   53    58   F 166703.67
16  15 group C   some high school     standard      none 690000   75    78   F 230051.00
17  16 group C        high school     standard      none 880000   89    86   M 293391.67
18  17 group B   some high school free/reduced      none 180000   32    28   F  60020.00
19  18 group C    master's degree free/reduced completed 460000   42    46   M 153362.67
20  19 group C associate's degree free/reduced      none 540000   58    61   F 180039.67
21  20 group D        high school     standard      none 660000   69    63   M 220044.00
22  21 group B       some college free/reduced completed 650000   75    70   F 216715.00
23  22 group D       some college     standard      none 440000   54    53   M 146702.33
24  23 group C   some high school     standard      none 690000   73    73   F 230048.67
25  24 group D  bachelor's degree free/reduced completed 740000   71    80   M 246717.00
26  25 group A    master's degree free/reduced      none 730000   74    72   M 243382.00
27  26 group B       some college     standard      none 690000   54    55   M 230036.33
28  27 group C  bachelor's degree     standard      none 670000   69    75   F 223381.33
```

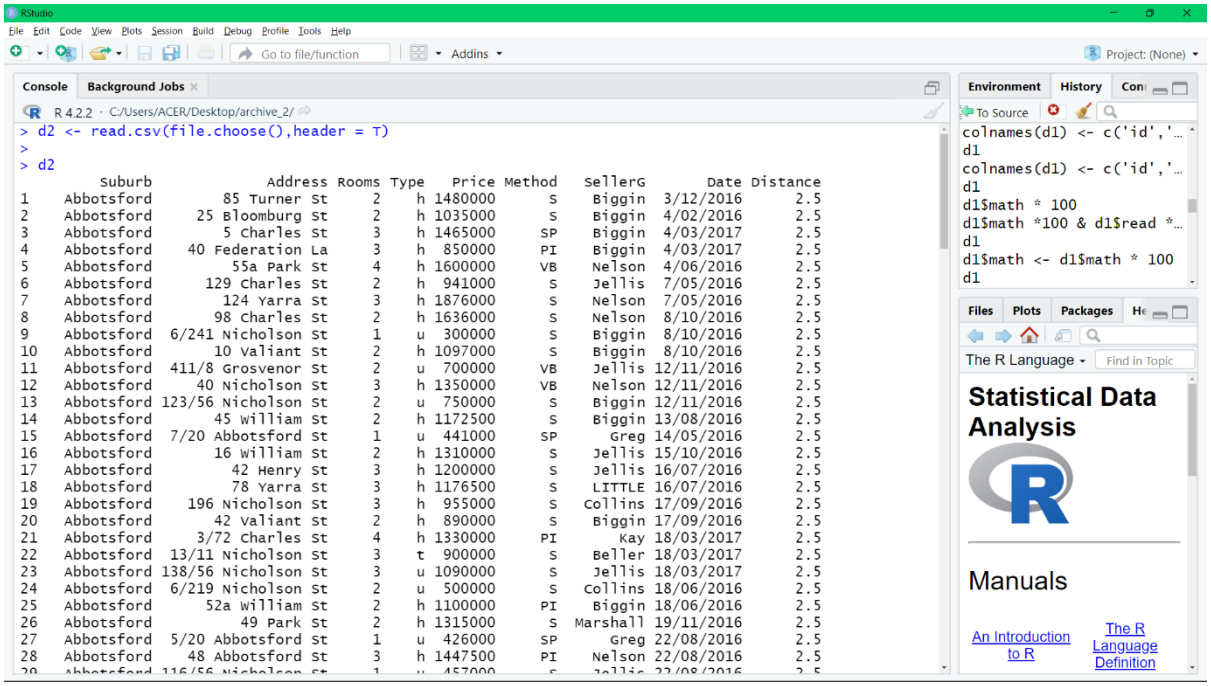

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Console
Background Jobs
R 4.2.2 · C:/Users/ACER/Desktop/archive/
> summary(d1$math)
Min. 1st Qu. Median Mean 3rd Qu. Max.
0 570000 660000 660890 770000 1000000
> summary(d1$read)
Min. 1st Qu. Median Mean 3rd Qu. Max.
17.00 59.00 70.00 69.17 79.00 100.00
> summary(d1$write)
Min. 1st Qu. Median Mean 3rd Qu. Max.
10.00 57.75 69.00 68.05 79.00 100.00
>
Environment
History
d1$math *100 & d1$read *...
d1
d1$math <- d1$math * 100
d1
d1$math <- d1$math * 100
d1$read <- d1$race *100
d1$math <- d1$math * 100
d1$read <- d1$read * 100
d1$write <- d1$write * 1...
Files
Plots
Packages
The R Language
Find in Topic
Statistical Data Analysis
Manuals
An Introduction to R
The R Language Definition

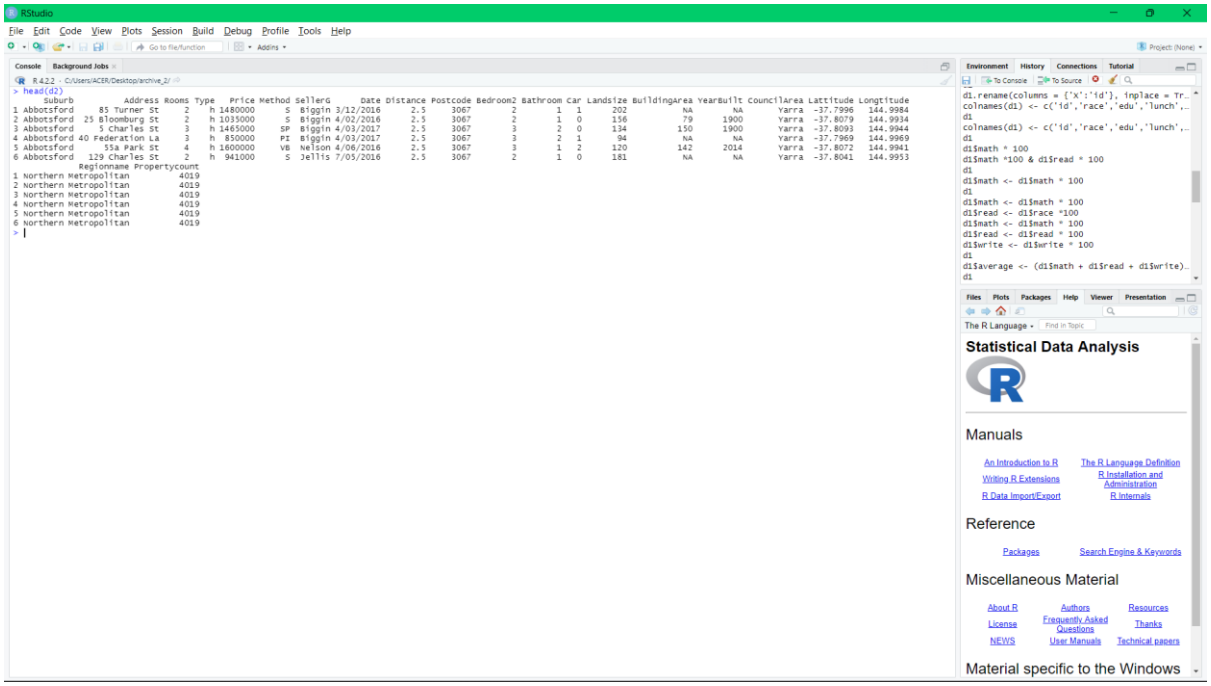

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Console
Background Jobs
R 4.2.2 · C:/Users/ACER/Desktop/archive/
> unique(d1[c("race")])
race
1 group B
2 group C
4 group A
9 group D
33 group E
>
> unique(d1[c("edu")])
edu
1 bachelor's degree
2 some college
3 master's degree
4 associate's degree
9 high school
16 some high school
>
> unique(d1[c("lunch")])
lunch
1 standard
4 free/reduced
>
> unique(d1[c("prep")])
prep
1 none
2 completed
> unique(d1[c("gen")])
gen
1 F
4 M
>
Environment
History
d1$math *100 & d1$read *...
d1
d1$math <- d1$math * 100
d1
d1$math <- d1$math * 100
d1$read <- d1$race *100
d1$math <- d1$math * 100
d1$read <- d1$read * 100
d1$write <- d1$write * 1...
Files
Plots
Packages
The R Language
Find in Topic
Statistical Data Analysis
Manuals
An Introduction to R
The R Language Definition

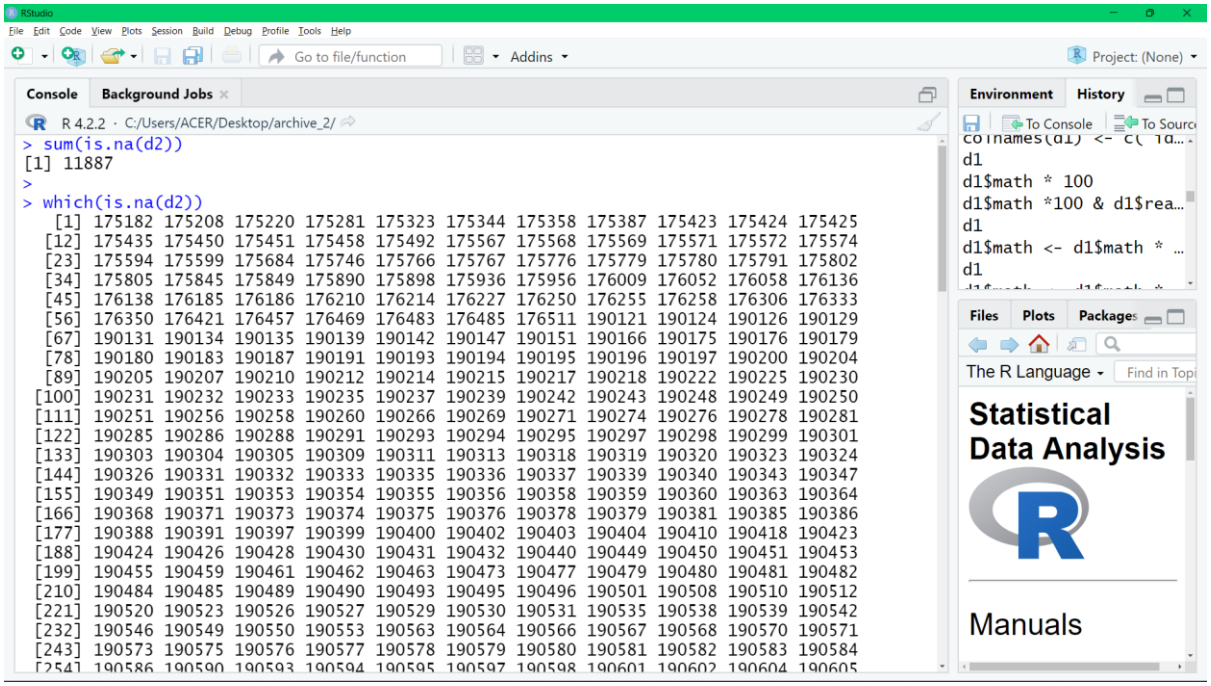

> mean(d1$race)
[1] NA
Warning message:
In mean.default(d1$race) : argument is not numeric or logical: returning NA
> mean(d1$edu)
[1] NA
Warning message:
In mean.default(d1$edu) : argument is not numeric or logical: returning NA
> mean(d1$lunch)
[1] NA
Warning message:
In mean.default(d1$lunch) :
argument is not numeric or logical: returning NA
> mean(d1$prep)
[1] NA
Warning message:
In mean.default(d1$prep) : argument is not numeric or logical: returning NA
> mean(d1$gen)
[1] NA
Warning message:
In mean.default(d1$gen) : argument is not numeric or logical: returning NA
Statistical Data Analysis
Manuals
An Introduction to R
The R Language Definition

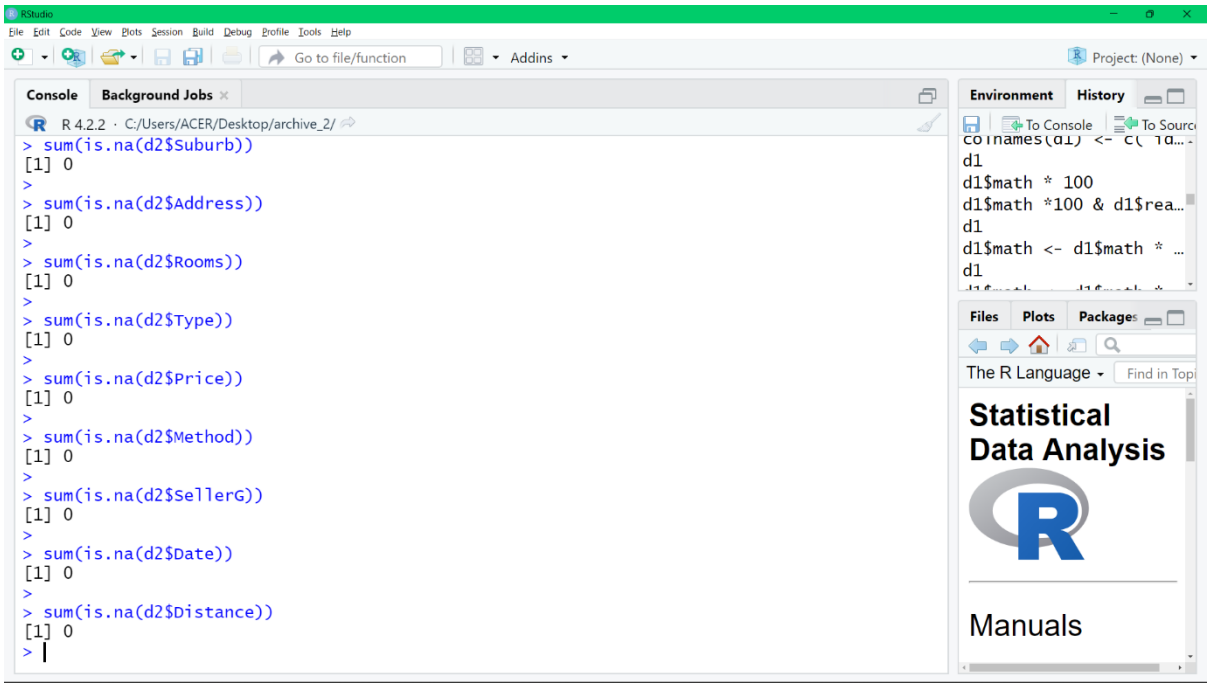
Question - 02

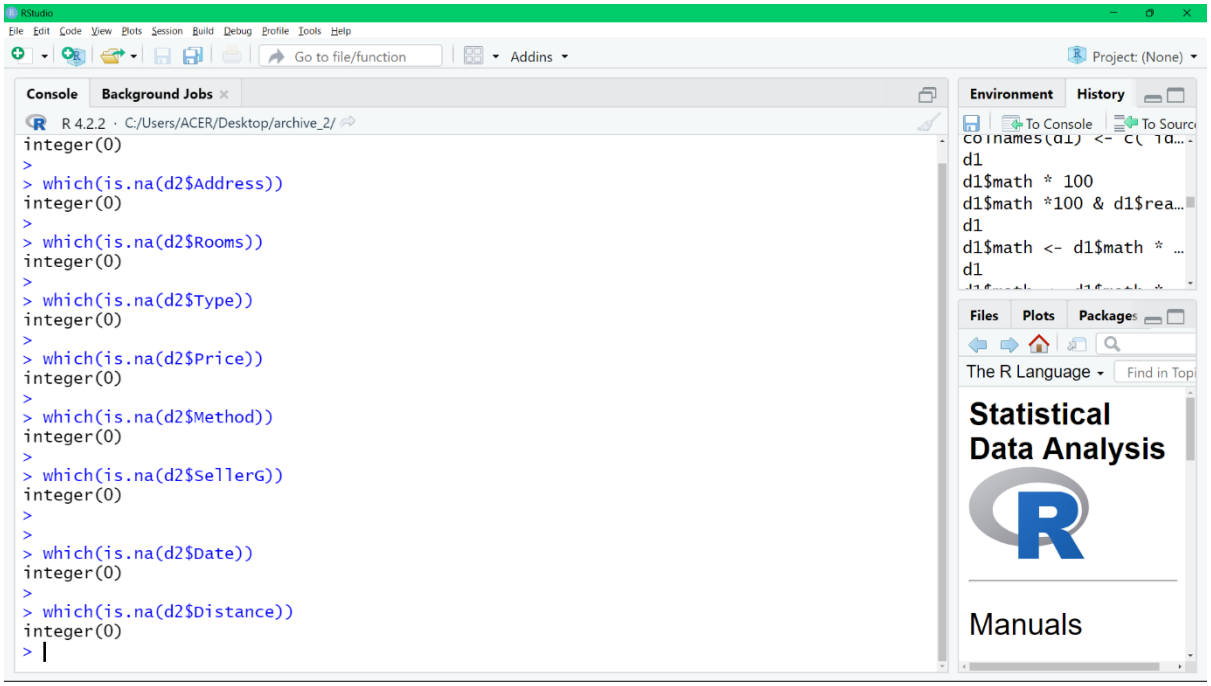

> getwd()
[1] "C:/Users/ACER/Desktop/archive"
> setwd("C:/Users/ACER/Desktop/archive_2")
> getwd()
[1] "C:/Users/ACER/Desktop/archive_2"
> list.files()
[1] "melb_data.csv"
Statistical Data Analysis
Manuals
An Introduction to R
The R Language Definition

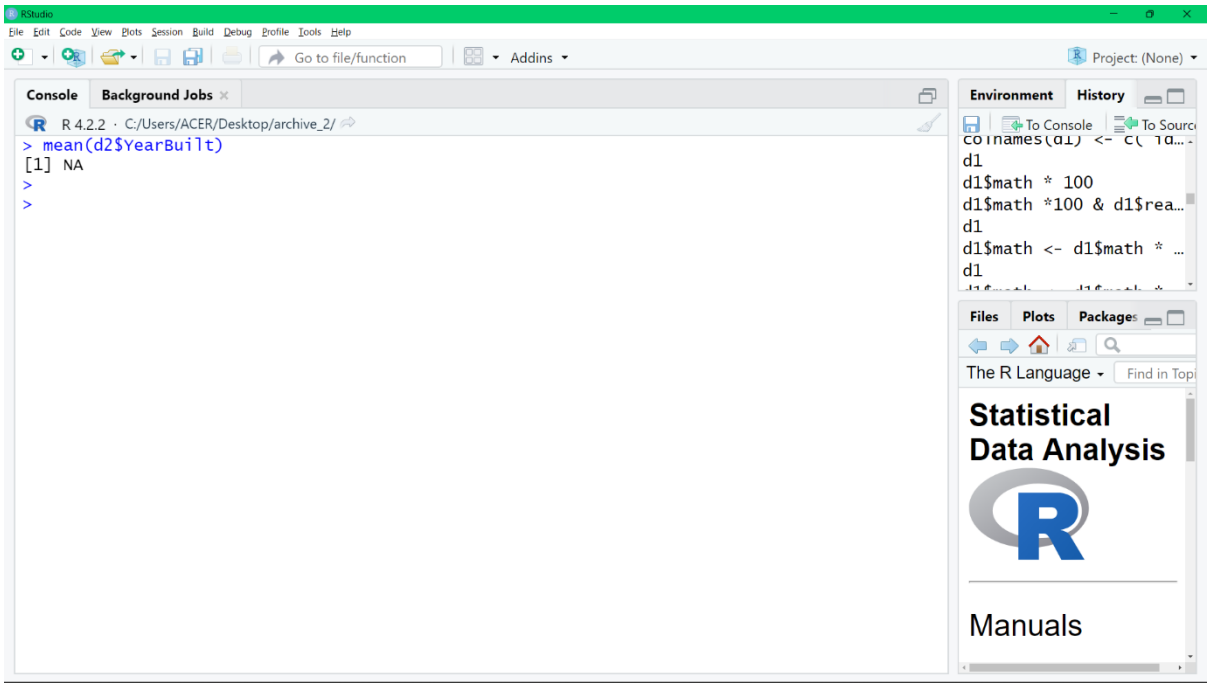

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Console
Background Jobs
R 4.2.2 · C:/Users/ACER/Desktop/archive_2/
> d2 <- read.csv(file.choose(),header = T)
> d2
Environment
History
Files
Plots
Packages
The R Language
Statistical Data Analysis
Manuals
An Introduction to R
The R Language Definition


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
> head(d2)
> d1.rename(columns = {'X':'id'}, inplace = Tr...
> colnames(d1) <- c('id','race','edu','lunch',...
> d1
> colnames(d1) <- c('id','race','edu','lunch',...
> d1
> d1$math * 100
> d1$math *100 & d1$read * 100
> d1
> d1$math <- d1$math * 100
> d1
> d1$math <- d1$math * 100
> d1$read <- d1$race *100
> d1$math <- d1$math * 100
> d1$read <- d1$read * 100
> d1$write <- d1$write * 100
> d1
> d1$average <- (d1$math + d1$read + d1$write)...
> d1
Statistical Data Analysis
Manuals
An Introduction to R
The R Language Definition
Writing R Extensions
R Installation and Administration
R Data Import/Export
R Internals
Reference
Packages
Search Engine & Keywords
Miscellaneous Material
About R
Authors
Resources
License
Frequently Asked Questions
Thanks
NEWS
User Manuals
Technical papers
Material specific to the Windows


> sum(is.na(d2))
[1] 11887
>
> which(is.na(d2))
[1] 175182 175208 175220 175281 175323 175344 175358 175387 175423 175424 175425
[12] 175435 175450 175451 175458 175492 175567 175568 175569 175571 175572 175574
[23] 175594 175599 175684 175746 175766 175767 175776 175779 175780 175791 175802
[34] 175805 175845 175849 175890 175898 175936 175956 176009 176052 176058 176136
[45] 176138 176185 176186 176210 176214 176227 176250 176255 176258 176306 176333
[56] 176350 176421 176457 176469 176483 176485 176511 190121 190124 190126 190129
[67] 190131 190134 190135 190139 190142 190147 190151 190166 190175 190176 190179
[78] 190180 190183 190187 190191 190193 190194 190195 190196 190197 190200 190204
[89] 190205 190207 190210 190212 190214 190215 190217 190218 190222 190225 190230
[100] 190231 190232 190233 190235 190237 190239 190242 190243 190248 190249 190250
[111] 190251 190256 190258 190260 190266 190269 190271 190274 190276 190278 190281
[122] 190285 190286 190288 190291 190293 190294 190295 190297 190298 190299 190301
[133] 190303 190304 190305 190309 190311 190313 190318 190319 190320 190323 190324
[144] 190326 190331 190332 190333 190335 190336 190337 190339 190340 190343 190347
[155] 190349 190351 190353 190354 190355 190356 190358 190359 190360 190363 190364
[166] 190368 190371 190373 190374 190375 190376 190378 190379 190381 190385 190386
[177] 190388 190391 190397 190399 190400 190402 190403 190404 190410 190418 190423
[188] 190424 190426 190428 190430 190431 190432 190440 190449 190450 190451 190453
[199] 190455 190459 190461 190462 190463 190473 190477 190479 190480 190481 190482
[210] 190484 190485 190489 190490 190493 190495 190496 190501 190508 190510 190512
[221] 190520 190523 190526 190527 190529 190530 190531 190535 190538 190539 190542
[232] 190546 190549 190550 190553 190563 190564 190566 190567 190568 190570 190571
[243] 190573 190575 190576 190577 190578 190579 190580 190581 190582 190583 190584
Console
Background Jobs
R 4.2.2 · C:/Users/ACER/Desktop/archive_2/
Environment
History
Files
Plots
Packages
The R Language
Statistical Data Analysis
Manuals


> sum(is.na(d2$Suburb))
[1] 0
>
> sum(is.na(d2$Address))
[1] 0
>
> sum(is.na(d2$Rooms))
[1] 0
>
> sum(is.na(d2$Type))
[1] 0
>
> sum(is.na(d2$Price))
[1] 0
>
> sum(is.na(d2$Method))
[1] 0
>
> sum(is.na(d2$SellerG))
[1] 0
>
> sum(is.na(d2$Date))
[1] 0
>
> sum(is.na(d2$Distance))
[1] 0
>
Console
Background Jobs
R 4.2.2 · C:/Users/ACER/Desktop/archive_2/
Environment
History
Files
Plots
Packages
The R Language
Statistical Data Analysis
Manuals


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Console
Background Jobs
R 4.2.2 · C:/Users/ACER/Desktop/archive_2/
integer(0)
> which(is.na(d2$Address))
integer(0)
> which(is.na(d2$Rooms))
integer(0)
> which(is.na(d2$Type))
integer(0)
> which(is.na(d2$Price))
integer(0)
> which(is.na(d2$Method))
integer(0)
> which(is.na(d2$SellerG))
integer(0)
> which(is.na(d2$Date))
integer(0)
> which(is.na(d2$Distance))
integer(0)
Environment
History
To Console
d1
d1$math * 100
d1$math *100 & d1$rea...
d1$math <- d1$math * ...
Files
Plots
Packages
The R Language
Statistical
Data Analysis
Manuals


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Console
Background Jobs
R 4.2.2 · C:/Users/ACER/Desktop/archive_2/
> mean(d2$YearBuilt)
[1] NA
Environment
History
To Console
d1
d1$math * 100
d1$math *100 & d1$rea...
d1$math <- d1$math * ...
Files
Plots
Packages
The R Language
Statistical
Data Analysis
Manuals

*****