ISCG8026 Introduction to Data Science
Semester 1 – 2020
R Programming Assignment A
Due Date: 23 April 2020, 8:30am

Part 1

Write a function named 'pollutantmean' that calculates the mean of a pollutant (sulfate or nitrate) across a specified list of monitors. The function 'pollutantmean' takes three arguments: 'directory', 'pollutant', and 'id'. Given a vector monitor ID numbers, 'pollutantmean' reads that monitors' particulate matter data from the directory specified in the 'directory' argument and returns the mean of the pollutant across all of the monitors, ignoring any missing values coded as NA.

Solution: source("pollutantmean.R")

```
 pollutantmean("specdata", "sulfate", 1:10)
[1] 4.064128

 pollutantmean("specdata", "nitrate", 23)
[1] 1.280833
```

Part 2

Write a function that reads a directory full of files and reports the number of completely observed cases in each data file. The function should return a data frame where the first column is the name of the file and the second column is the number of complete cases.

Solution: source("complete.R")

```
  complete("specdata", 1)
  id nobs
1 1  117

 complete("specdata", 30:25)
  id nobs
1 30  932
2 29  711
3 28  475
4 27  338
5 26  586
6 25  463
```

ISCG8026 Introduction to Data Science
Semester 1 – 2020
R Programming Assignment A
Due Date: 23 April 2020, 8:30am

Part 3

Write a function that takes a directory of data files and a threshold for complete cases and calculates the correlation between sulfate and nitrate for monitor locations where the number of completely observed cases (on all variables) is greater than the threshold. The function should return a vector of correlations for the monitors that meet the threshold requirement. If no monitors meet the threshold requirement, then the function should return a numeric vector of length 0.

Solution: source("corr.R") (source("complete.R") is inside corr.R)

```
cr <- corr("specdata", 150)
cr
  [1] -0.018957541 -0.140512544 -0.043897372 -0.068159562 -0.123506666
  [6] -0.075888144 -0.159673652 -0.086841940  0.763128837 -0.157828603
 [11] -0.156998919 -0.044898818  0.117249264  0.259057178  0.133274607
 [16]  0.366201078  0.580751264  0.006863930  0.726693888  0.057741676
 [21]  0.115338086  0.465754012  0.515804375  0.412693537  0.375631176
 [26]  0.315725317  0.244560561  0.594426499  0.553514976  0.614340566
 [31]  0.460513619  0.405022501  0.434789780  0.088421364  0.118136697
 [36] -0.091022820 -0.033091304  0.440660466 -0.029683708  0.268525390
 [41]  0.277220958 -0.049108453  0.322627410  0.091139374 -0.025750053
 [46]  0.120521602 -0.061746831  0.041306963 -0.146202136 -0.162485185
 [51] -0.097254393  0.089262856  0.568403991  0.711864008  0.268203237
 [56]  0.190644585  0.227222983  0.229238882  0.005635506  0.018628108
 [61] -0.064750174  0.096614297  0.002864405  0.107184775  0.128477284
 [66] -0.042533572 -0.137041337  0.136609030  0.118975253  0.098073855
 [71]  0.066928310  0.100212474 -0.063984344 -0.066525489 -0.129245884
 [76] -0.111066409 -0.089441210 -0.114090325 -0.106280702 -0.176855164
 [81] -0.116984680  0.019138583  0.100643502 -0.073858484  0.036665921
 [86] -0.107957809  0.296744105  0.347421569  0.146528765  0.362414577
 [91]  0.093330832  0.198915192  0.164602262  0.180626975  0.176508543
 [96]  0.139158631  0.231984399  0.227615918  0.275903634  0.299630040
[101]  0.248143145  0.298344178 -0.056325366 -0.178114558  0.002032940
[106] -0.022802183 -0.001202233  0.085217423 -0.076409023  0.010021716
[111]  0.016411646 -0.038785934 -0.075297768  0.041917773  0.193324040
[116]  0.596929143  0.113596590 -0.143750037 -0.017703373  0.284905360
[121]  0.305506111  0.150031306  0.134895077  0.172850003  0.286076203
[126] -0.106687748  0.244744168  0.337120085  0.424798956  0.095921881
[131]  0.022899033  0.143330735  0.087196218  0.408741028  0.425176879
[136]  0.361728434 -0.035090337 -0.082388453 -0.094742313 -0.087573726
[141] -0.060405837 -0.092398269 -0.183197353  0.124650112 -0.053001162
[146] -0.039911536  0.010158287  0.451828854  0.295793699  0.615268727
[151] -0.075214053  0.132207405  0.089547098 -0.019086127 -0.045552626
[156]  0.211599525 -0.073972834  0.112668377  0.138387891 -0.003207550
[161] -0.052643174  0.042168144 -0.067460173 -0.030882797  0.017805647
[166]  0.026138073 -0.050287543  0.016535643  0.199919014  0.482158286
[171]  0.355110474  0.589606340  0.368038099 -0.029094866 -0.074495323
[176]  0.262101561 -0.005386993  0.258826380  0.144110820  0.101915017
[181]  0.023020993  0.074594252  0.256665139  0.162401158 -0.003454405
[186]  0.190141976  0.184581239  0.120596460 -0.176233152 -0.144699131
[191]  0.147074115  0.273520382  0.109557323 -0.092863394 -0.182752126
[196] -0.008836513  0.356592359 -0.089133895 -0.017185129 -0.156323514
[201] -0.042538204  0.010235676 -0.009912754 -0.042910367 -0.210567709
[206] -0.155957816  0.046211272 -0.060808231  0.160865053  0.615095781
[211]  0.598343330  0.506535631  0.191834811 -0.024723462 -0.150627164
[216] -0.002500089 -0.166201361  0.619349867  0.531380642  0.520115665
[221]  0.466673962  0.518820173  0.394191512  0.379446208 -0.123172036
[226] -0.061565518 -0.180133963  0.253978075  0.139867175  0.316429404
[231]  0.268780500  0.279397143  0.267260662  0.287133842
```

ISCG8026 Introduction to Data Science
Semester 1 – 2020
R Programming Assignment A
Due Date: 23 April 2020, 8:30am

```
head(cr)

[1] -0.01895754 -0.14051254 -0.04389737 -0.06815956 -0.12350667 -0.07588814

summary(cr)

    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-0.21057 -0.04999  0.09463  0.12525  0.26844  0.76313
```

Part 4

Write a function named 'pollutantvector' that returns a vector of those pollutants (sulfate or nitrate) whose values are greater than 'p', across a specified list of monitors. The function 'pollutantvector' takes four arguments: 'directory', 'pollutant', 'id' and 'p'. Given a vector monitor ID numbers, 'pollutantvector' reads that monitors' particulate matter data from the directory specified in the 'directory' argument and returns the ones more than a certain value ('p') across all of the monitors, ignoring any missing values coded as NA.

```
source("pollutantvector.R")

pollutantvector("specdata", "sulfate", 1:35, 0.5)

[1]  7.210  5.990  4.680  3.470  2.420  1.430  2.760  3.410  1.300  3.150
[11]  2.870  2.270  2.330  1.840  7.130  2.050  2.050  2.580  3.260  3.540
[21]  4.400  2.040  4.360  3.460  2.990  2.000  2.420  2.760  1.480  3.540
[31]  6.630 10.900  1.630  5.760  5.520 12.200  4.980  4.390  1.650  4.530
[41]  4.300 15.800  5.440  3.090  3.610  4.190  7.370  2.680  2.210  1.990
[51]  1.780  2.030  2.110  1.820  2.370  3.410  4.960  2.300  4.670  2.870
[61]  1.530  1.280  4.170  2.750  2.830  1.220  4.190  1.920  2.550  2.180
[71]  2.930  2.790  2.740  2.020  4.730  5.840  4.640  4.790  6.650  2.090
[81]  4.850  7.250  3.220  3.810  2.850  1.360  1.850 13.000  2.570  2.680
[91]  5.420  9.520  7.680  6.650 19.100  6.040  2.800  5.520  1.820  4.880
[101] 2.280  5.510  2.410  1.630  1.670  0.959  2.740  0.613  2.370  2.240
[111] 2.530  3.500  2.600  2.700  3.740  2.640  4.820  5.990  2.300  2.190
[121] 1.200  2.230  4.150  3.000  4.040  4.790  2.050  7.930  7.740  2.630
[131] 3.800  3.520  3.700  3.900  2.720  5.950  2.110  6.400  2.670  3.920
[141] 2.060  1.890  7.560  6.870  3.000  3.420  3.760  3.110 11.100  4.600
[151]10.300  4.700  3.540  5.290  1.510  1.500  4.380  3.200  2.780 10.700
[161] 3.630  8.120  6.830  6.570  3.920  2.890  3.150  6.430  6.040 12.200
[171]12.100 13.400 10.500 13.500  2.670  4.630  6.280  5.850  3.740  8.360
[181]11.400  4.410  9.620 10.800  3.290  3.640  3.110  4.360  4.330 12.500
[191] 8.550  6.850  5.730  3.520  4.380  8.260  2.060  1.850  1.050  4.300
[201] 2.840  4.610  4.520  4.250  7.310  4.710  4.940  7.380  1.900  5.170
[211] 2.510  3.170  0.555  1.190  3.650  1.750  3.610  2.840  4.220  1.570
[221] 2.600  3.280  1.900  1.870  2.720  4.730  4.520  2.540  1.990  3.270
[231] 1.870  3.140  5.550  3.040  5.660  8.720  3.330  3.580  3.860  4.990
[241] 3.660  3.690  3.980  6.100  2.630  2.230  2.320  6.720  7.300  6.410
[251] 6.570  2.740  4.770  4.130  6.880  4.800  2.220  6.330 11.500  4.900
[261] 4.070  6.130  3.710  6.290  5.210  4.040  1.780  8.610  7.540 11.200
[271] 4.000  6.520  3.660  9.770  5.280  3.410  3.440 13.700 10.700 11.400
[281] 4.080  5.010  4.300  6.340  4.470  9.180  5.890  5.360 11.000  8.310
[291]18.500  9.590  5.430 13.600  8.620  4.780  8.700  7.410  5.990  2.830
[301] 3.520 11.000  4.340  4.610  4.680  2.830  4.560  3.730  4.600  4.050
[311] 3.000  5.150  3.830  3.620  4.140  1.970  3.720  3.680  1.650  2.180
[321] 4.330  3.790  0.844  1.650  0.917  3.200  1.610  2.600  6.480  4.550
[331] 1.070  3.210  2.640  5.660  5.400  4.500  6.080  3.490  6.580  3.060
[341] 3.820  7.540  2.690  3.750  4.210  3.310  2.460  5.600  5.580  4.710
[351] 7.560  2.890  3.490  3.810  9.780  8.810  5.230  4.270  5.950  1.030
```

ISCG8026 Introduction to Data Science

Semester 1 – 2020

R Programming Assignment A

Due Date: 23 April 2020, 8:30am

```
[361]  4.070  1.710  5.380 12.700  6.540  2.500  5.190  6.960  7.120  3.940
[371]  1.970  5.040  6.590  8.620  9.130  5.120  2.310  2.990  3.000  3.710
[381]  2.550  5.010  2.580  4.330  8.360  3.570  5.550  3.940  7.650 10.800
[391]  1.660  8.490  9.460  4.270 16.800  6.220  2.190  3.980 11.200 13.900
[401]  7.740  2.500 11.300  6.120  6.980  2.980  3.780  4.610  5.960  8.240
[411]  4.070  1.480  4.150  4.310  4.730  2.740  2.560  6.080  3.170  3.220
[421]  1.870  1.540  4.350  1.610  3.370  1.960  1.290  1.980  2.900  5.070
[431]  2.920  3.070  1.690  1.440  1.870  3.470  1.840  3.560  1.340  6.290
[441]  3.970  4.700  2.280  2.710  2.140  3.000  4.940  2.160  3.270  2.340
[451]  2.560  3.880  6.380  3.820  1.610  1.950  1.180  4.600  2.320  3.690
[461]  4.520  2.640  3.020  1.820  2.840  2.670  3.250  3.780  4.060  5.310
[471]  3.620  3.450  2.300  3.760  4.170  4.780  6.760  5.000  3.770  2.620
[481]  6.330  4.730  4.360  4.480  3.710 12.100  7.290  7.550  4.700  2.840
[491]  9.050  2.060  1.950  3.370  3.510  4.790  3.830  4.210  3.660  3.650
[501]13.700  7.600  4.090 10.300  5.460 13.900  7.140  3.540  5.320 13.300
[511]11.100  2.100  5.030 12.400  6.160  5.390  5.310  5.520  5.320  4.260
[521]  3.030 10.300  8.890  5.050  6.330  4.790  2.680  1.140  2.660 11.200
[531]  3.070  8.200  8.960  2.530  2.220  1.510  2.800  1.490  4.840  3.230
[541]  3.190  2.020  1.730  3.640  3.090  1.820  2.310  2.330  2.300  1.650
[551]  2.320  1.620  3.040  2.150  3.250  5.330  1.140  3.620  1.730  4.160
[561]  2.510  2.520  5.420  3.530  3.970  1.810  1.870  3.610  1.450  6.070
[571]  6.890  5.020  1.830  2.030  2.410  1.400  2.770  4.530  6.170  1.540
[581]  6.030  3.250  1.030  2.650  1.930  4.470  2.140  5.140  7.490  3.540
[591]  1.250  3.050  4.890  4.590 14.700  8.580  7.230  4.260  7.030  7.830
[601]  3.980  7.770  3.860  3.190  1.160  5.700 10.600 18.400  8.550 11.300
[611]  4.970  2.410 10.400  4.690  3.380  3.550  6.340 14.100 14.600 13.300
[621]  9.080 17.400  3.760 11.100  8.210  7.220  9.350  5.860 11.000 10.600
[631]  4.830 10.600 23.300  4.120  9.690  2.390  2.590  6.520  8.140  5.620
[641]  3.510  6.410  9.390  1.860  5.010  2.190  4.320  6.580  3.800  4.690
[651]  4.340  3.030  2.850  1.910  3.700  0.991  3.720  1.210  1.780  6.250
[661]  0.817  1.840  2.300  2.150  2.730  3.280  3.380  1.770  3.280  1.140
[671]  1.150  2.410  1.810  1.330  1.560  1.310  1.230  3.230  2.740  2.370
[681]  3.030  3.080  2.200  3.010  4.690  3.320  4.080  3.190  1.840  6.010
[691]  2.280  1.800  4.290  4.560  6.140  3.480  3.320  2.390  3.250  5.730
[701]  4.230  3.740  6.840  6.560  5.370  5.610  9.150  1.530  1.790  2.660
[711]  6.140  3.980  9.780 10.400  4.550  6.830 12.500 10.400  5.480  6.770
[721]  6.520  4.810  7.670 10.600  8.050  7.220  2.830  6.530 10.700  4.460
[731]  3.680  2.980  6.520  4.110  5.280  4.770  6.700  4.420 12.000  8.840
[741]17.900  2.600  3.370  5.450  8.090  6.490 10.500  5.040  9.610  2.520
[751]  4.140  4.020  9.370  3.470  4.630  2.810  1.840  2.590  1.830  2.960
[761]  1.860  1.880  3.440  4.220  2.220  6.480  2.050  4.120  3.070  1.350
[771]  1.640  2.380  4.560  6.500  6.780  1.590  3.400  1.270  3.610  3.790
[781]  2.100  1.790  4.100  1.600  3.170  2.650  1.180  3.650  0.877  2.660
[791]  2.360  4.010  1.880  1.900  1.630  1.660  3.160  1.480  8.530  6.160
[801]  2.850  2.790  2.100  5.190  5.320  5.810  6.310  4.550  1.970  1.560
[811]  2.800  5.070 11.100  2.660  2.140  4.920  7.460 11.500  4.330  6.560
[821]  5.320  3.340 10.800  3.860  6.190  8.840 11.300  2.140  2.360  7.890
[831]  7.220  6.570  2.510  8.310  5.400  7.930  4.240  5.340  6.980  3.510
[841]  4.300  4.340  2.730  5.710  5.700  3.240  8.190 18.800  5.330  4.860
[851]11.200  8.430  6.560 11.300  6.120  9.470 27.900  8.160  4.720 11.500
[861]  6.320  4.300  2.650  7.820  3.500  3.570  4.780  3.880  4.360  4.760
[871]  2.390  0.769  2.850  3.250  6.560  6.100  3.900  5.480  2.400  4.030
[881]  2.160  1.320  2.850  4.620  1.760  6.250  2.740  1.560  3.020  1.390
[891]  1.780  0.775  1.820  1.710  2.390  1.210  2.700  2.080  1.460  2.180
[901]  1.700  1.200  1.210  3.210  1.560  1.390  4.800  1.560  2.720  3.670
[911]  2.340  4.590  2.310  4.330  3.720  6.680  5.510  6.800  4.480  1.380
[921]  4.740  2.750  5.340  2.950  2.490  2.660  1.800  5.120  2.170  1.840
[931]  2.820  2.610  5.310  3.480  7.780  4.220  2.520  3.710  5.810  2.020
[941]  2.790  5.600  4.750  2.420  1.960  4.650  2.720  4.180  3.650  2.320
[951]12.600 11.000  3.270  1.820  4.370  4.200  3.170  1.960  4.530  7.650
[961]  9.160 11.000  6.320 11.700  7.560  5.790  4.710  4.310  3.130  2.590
[971]  7.990  4.820  4.240  6.260  3.870  6.750  2.920  2.270  4.990  3.260
```

ISCG8026 Introduction to Data Science

Semester 1 – 2020

R Programming Assignment A

Due Date: 23 April 2020, 8:30am

```
[981] 6.580  0.858  4.030  6.520  4.110  0.718  3.330  5.270  1.550  2.870
[991] 5.440  1.960  1.890  1.850  0.725  2.220  2.750  3.150  1.750  2.670

[ reached getOption("max.print") -- omitted 11200 entries ]
```