Exercise 1. Data manipulation

Sasha D. Hafner

06 March, 2024

R data.table solution

1. Read in data

```
library(data.table)
air <- fread('../data/air_cleaners.csv')
```

```
air
```

```
##        aircleaner       timestamp elapsed_time   form     mtzr compound_name
##    1: PureAir2000 3/10/2022 13:45     0.000000   CH4O  32.0335      Methanol
##    2: PureAir2000 3/10/2022 13:45     0.000000   C7H8  92.0699       Toluene
##    3: PureAir2000 3/10/2022 13:45     0.000000   C5H8  68.0699      Isoprene
##    4: PureAir2000 3/10/2022 13:45     0.000000  C7H6O 106.0491  Benzaldehyde
##    5: PureAir2000 3/10/2022 13:45     0.000000  C3H6O  58.0491       Acetone
##   ---
## 7115:  NoSmell4.2 3/30/2022 18:05     5.166667   CH4O  32.0335      Methanol
## 7116:  NoSmell4.2 3/30/2022 18:05     5.166667  C7H6O 106.0491  Benzaldehyde
## 7117:  NoSmell4.2 3/30/2022 18:05     5.166667  C3H6O  58.0491       Acetone
## 7118:  NoSmell4.2 3/30/2022 18:05     5.166667 C2H4O2  60.0284   Acetic acid
## 7119:  NoSmell4.2 3/30/2022 18:05     5.166667   C7H8  92.0699       Toluene
##       flow_dir concentration
##    1:       In    971.682900
##    2:       In    528.657950
##    3:       In    443.841800
##    4:       In    673.093200
##    5:       In   1063.301000
##   ---
## 7115:       In    193.325450
## 7116:       In     50.615150
## 7117:       In    120.169100
## 7118:       In     33.904700
## 7119:       In      7.061408
```

2. Check data

```
source('../functions-R/dfsumm.R')
```

```
dfsumm(air)
```

```
##
##  7119 rows and 8 columns
```

```
##  7119 unique rows
##                      aircleaner       timestamp elapsed_time      form    mtzr
## Class                 character       character      numeric character numeric
## Minimum              NoSmell4.2 3/10/2022 13:45            0    C10H16      32
## Maximum             PureAir2000 3/30/2022 18:05         7.98      CH4O     136
## Mean                      <NA>            <NA>         3.44      <NA>    74.3
## Unique (excld. NA)           2             791          480         9       9
## Missing values               0               0            0         0       0
## Sorted                   FALSE            TRUE        FALSE     FALSE   FALSE
##
##                     compound_name  flow_dir concentration
## Class                   character character       numeric
## Minimum              Acetaldehyde        In         0.602
## Maximum                   Toluene       Out          1320
## Mean                         <NA>      <NA>           280
## Unique (excld. NA)              9         2          7116
## Missing values                  0         0             0
## Sorted                      FALSE     FALSE         FALSE
##
```

```
summary(air)
```

```
##   aircleaner         timestamp          elapsed_time       form
##  Length:7119        Length:7119        Min.   :0.000   Length:7119
##  Class :character   Class :character   1st Qu.:1.633   Class :character
##  Mode  :character   Mode  :character   Median :3.283   Mode  :character
##                                        Mean   :3.438
##                                        3rd Qu.:4.933
##                                        Max.   :7.983
##       mtzr        compound_name        flow_dir         concentration
##  Min.   : 32.03   Length:7119        Length:7119        Min.   :   0.6016
##  1st Qu.: 58.05   Class :character   Class :character   1st Qu.:  78.1638
##  Median : 68.07   Mode  :character   Mode  :character   Median : 216.9324
##  Mean   : 74.28                                         Mean   : 279.6238
##  3rd Qu.: 92.07                                         3rd Qu.: 438.5930
##  Max.   :136.13                                         Max.   :1318.6745
```

```
dim(air)
```

```
## [1] 7119    8
```

3. Subset

```
ns <- air[aircleaner == 'NoSmell4.2', ]
dim(ns)
```

```
## [1] 2799    8
```

```
tol <- air[aircleaner == 'NoSmell4.2' & compound_name == 'Toluene' & flow_dir == 'In', ]
tol
```

```
##      aircleaner       timestamp elapsed_time form    mtzr compound_name
##   1: NoSmell4.2 3/30/2022 12:55   0.00000000 C7H8 92.0699       Toluene
##   2: NoSmell4.2 3/30/2022 12:56   0.01666667 C7H8 92.0699       Toluene
##   3: NoSmell4.2 3/30/2022 12:57   0.03333333 C7H8 92.0699       Toluene
##   4: NoSmell4.2 3/30/2022 12:58   0.05000000 C7H8 92.0699       Toluene
```

```
##   5: NoSmell4.2 3/30/2022 12:59   0.06666667 C7H8 92.0699       Toluene
##  ---
## 152: NoSmell4.2 3/30/2022 17:56   5.01666667 C7H8 92.0699       Toluene
## 153: NoSmell4.2 3/30/2022 17:57   5.03333333 C7H8 92.0699       Toluene
## 154: NoSmell4.2 3/30/2022 17:58   5.05000000 C7H8 92.0699       Toluene
## 155: NoSmell4.2 3/30/2022 17:59   5.06666667 C7H8 92.0699       Toluene
## 156: NoSmell4.2 3/30/2022 18:05   5.16666667 C7H8 92.0699       Toluene
##      flow_dir concentration
##   1:       In    625.205200
##   2:       In    636.001050
##   3:       In    636.295600
##   4:       In    630.652100
##   5:       In    617.685400
##  ---
## 152:       In      7.036319
## 153:       In      6.967802
## 154:       In      6.814997
## 155:       In      6.756792
## 156:       In      7.061408
```

```
dim(tol)
```

```
## [1] 156   8
```

4. Merge

```
mm <- fread('../data/mol_mass.csv')
mm
```

```
##      form mol_mass
## 1:   CH4O   32.042
## 2:   C7H8   92.134
## 3:   C5H8   68.114
## 4:  C7H6O  106.118
## 5:  C3H6O   58.078
## 6:  C2H4O   44.052
## 7: C2H4O2   60.052
## 8:  C4H8O   72.104
## 9: C10H16  136.228
```

```
dim(air)
```

```
## [1] 7119    8
```

```
air <- merge(air, mm, by = 'form')
dim(air)
```

```
## [1] 7119    9
```

```
air
```

```
##         form  aircleaner       timestamp elapsed_time     mtzr compound_name
##    1: C10H16 PureAir2000 3/10/2022 13:45   0.00000000 136.1325      Limonene
##    2: C10H16 PureAir2000 3/10/2022 13:46   0.01666667 136.1325      Limonene
##    3: C10H16 PureAir2000 3/10/2022 13:47   0.03333333 136.1325      Limonene
##    4: C10H16 PureAir2000 3/10/2022 13:48   0.05000000 136.1325      Limonene
##    5: C10H16 PureAir2000 3/10/2022 13:49   0.06666667 136.1325      Limonene
```

```
##    ---
## 7115:   CH4O  NoSmell4.2 3/30/2022 18:01   5.10000000  32.0335      Methanol
## 7116:   CH4O  NoSmell4.2 3/30/2022 18:02   5.11666667  32.0335      Methanol
## 7117:   CH4O  NoSmell4.2 3/30/2022 18:03   5.13333333  32.0335      Methanol
## 7118:   CH4O  NoSmell4.2 3/30/2022 18:04   5.15000000  32.0335      Methanol
## 7119:   CH4O  NoSmell4.2 3/30/2022 18:05   5.16666667  32.0335      Methanol
##       flow_dir concentration mol_mass
##    1:       In      639.5513  136.228
##    2:       In      696.0174  136.228
##    3:       In      744.5909  136.228
##    4:       In      748.1602  136.228
##    5:       In      735.1984  136.228
##    ---
## 7115:      Out      196.1996   32.042
## 7116:      Out      196.6986   32.042
## 7117:      Out      197.1389   32.042
## 7118:      Out      195.0970   32.042
## 7119:       In      193.3254   32.042
```

5. Add columns

```
air[, concm := mol_mass * 1.0 * concentration / 1E9 / 8.2057E-5 / (273.15 + 20)]
air
```

```
##         form  aircleaner       timestamp elapsed_time     mtzr compound_name
##    1: C10H16 PureAir2000 3/10/2022 13:45   0.00000000 136.1325      Limonene
##    2: C10H16 PureAir2000 3/10/2022 13:46   0.01666667 136.1325      Limonene
##    3: C10H16 PureAir2000 3/10/2022 13:47   0.03333333 136.1325      Limonene
##    4: C10H16 PureAir2000 3/10/2022 13:48   0.05000000 136.1325      Limonene
##    5: C10H16 PureAir2000 3/10/2022 13:49   0.06666667 136.1325      Limonene
##    ---
## 7115:   CH4O  NoSmell4.2 3/30/2022 18:01   5.10000000  32.0335      Methanol
## 7116:   CH4O  NoSmell4.2 3/30/2022 18:02   5.11666667  32.0335      Methanol
## 7117:   CH4O  NoSmell4.2 3/30/2022 18:03   5.13333333  32.0335      Methanol
## 7118:   CH4O  NoSmell4.2 3/30/2022 18:04   5.15000000  32.0335      Methanol
## 7119:   CH4O  NoSmell4.2 3/30/2022 18:05   5.16666667  32.0335      Methanol
##       flow_dir concentration mol_mass        concm
##    1:       In      639.5513  136.228 0.0036218981
##    2:       In      696.0174  136.228 0.0039416762
##    3:       In      744.5909  136.228 0.0042167570
##    4:       In      748.1602  136.228 0.0042369703
##    5:       In      735.1984  136.228 0.0041635652
##    ---
## 7115:      Out      196.1996   32.042 0.0002613438
## 7116:      Out      196.6986   32.042 0.0002620085
## 7117:      Out      197.1389   32.042 0.0002625950
## 7118:      Out      195.0970   32.042 0.0002598751
## 7119:       In      193.3254   32.042 0.0002575153
```

In g/min.

```
air[, mass_flow := concm * 0.1]
air
```

```
##         form  aircleaner       timestamp elapsed_time     mtzr compound_name
```

```
##    1: C10H16 PureAir2000 3/10/2022 13:45   0.00000000 136.1325      Limonene
##    2: C10H16 PureAir2000 3/10/2022 13:46   0.01666667 136.1325      Limonene
##    3: C10H16 PureAir2000 3/10/2022 13:47   0.03333333 136.1325      Limonene
##    4: C10H16 PureAir2000 3/10/2022 13:48   0.05000000 136.1325      Limonene
##    5: C10H16 PureAir2000 3/10/2022 13:49   0.06666667 136.1325      Limonene
##   ---
## 7115:   CH4O  NoSmell4.2 3/30/2022 18:01   5.10000000  32.0335      Methanol
## 7116:   CH4O  NoSmell4.2 3/30/2022 18:02   5.11666667  32.0335      Methanol
## 7117:   CH4O  NoSmell4.2 3/30/2022 18:03   5.13333333  32.0335      Methanol
## 7118:   CH4O  NoSmell4.2 3/30/2022 18:04   5.15000000  32.0335      Methanol
## 7119:   CH4O  NoSmell4.2 3/30/2022 18:05   5.16666667  32.0335      Methanol
##       flow_dir concentration mol_mass        concm    mass_flow
##    1:       In      639.5513  136.228 0.0036218981 3.621898e-04
##    2:       In      696.0174  136.228 0.0039416762 3.941676e-04
##    3:       In      744.5909  136.228 0.0042167570 4.216757e-04
##    4:       In      748.1602  136.228 0.0042369703 4.236970e-04
##    5:       In      735.1984  136.228 0.0041635652 4.163565e-04
##   ---
## 7115:      Out      196.1996   32.042 0.0002613438 2.613438e-05
## 7116:      Out      196.6986   32.042 0.0002620085 2.620085e-05
## 7117:      Out      197.1389   32.042 0.0002625950 2.625950e-05
## 7118:      Out      195.0970   32.042 0.0002598751 2.598751e-05
## 7119:       In      193.3254   32.042 0.0002575153 2.575153e-05
```

6. Dates and times

```
class(air[, timestamp])
```

```
## [1] "character"
```

```
air[, timestamp := as.POSIXct(timestamp, format = '%m/%d/%Y %H:%M')]
```

```
air[, etime_min := timestamp - as.POSIXct('03/10/2022 13:40', format = '%m/%d/%Y %H:%M')]
air
```

```
##         form  aircleaner           timestamp elapsed_time     mtzr
##    1: C10H16 PureAir2000 2022-03-10 13:45:00   0.00000000 136.1325
##    2: C10H16 PureAir2000 2022-03-10 13:46:00   0.01666667 136.1325
##    3: C10H16 PureAir2000 2022-03-10 13:47:00   0.03333333 136.1325
##    4: C10H16 PureAir2000 2022-03-10 13:48:00   0.05000000 136.1325
##    5: C10H16 PureAir2000 2022-03-10 13:49:00   0.06666667 136.1325
##   ---
## 7115:   CH4O  NoSmell4.2 2022-03-30 18:01:00   5.10000000  32.0335
## 7116:   CH4O  NoSmell4.2 2022-03-30 18:02:00   5.11666667  32.0335
## 7117:   CH4O  NoSmell4.2 2022-03-30 18:03:00   5.13333333  32.0335
## 7118:   CH4O  NoSmell4.2 2022-03-30 18:04:00   5.15000000  32.0335
## 7119:   CH4O  NoSmell4.2 2022-03-30 18:05:00   5.16666667  32.0335
##       compound_name flow_dir concentration mol_mass        concm    mass_flow
##    1:      Limonene       In      639.5513  136.228 0.0036218981 3.621898e-04
##    2:      Limonene       In      696.0174  136.228 0.0039416762 3.941676e-04
##    3:      Limonene       In      744.5909  136.228 0.0042167570 4.216757e-04
##    4:      Limonene       In      748.1602  136.228 0.0042369703 4.236970e-04
##    5:      Limonene       In      735.1984  136.228 0.0041635652 4.163565e-04
##   ---
```

```
## 7115:      Methanol      Out      196.1996   32.042 0.0002613438 2.613438e-05
## 7116:      Methanol      Out      196.6986   32.042 0.0002620085 2.620085e-05
## 7117:      Methanol      Out      197.1389   32.042 0.0002625950 2.625950e-05
## 7118:      Methanol      Out      195.0970   32.042 0.0002598751 2.598751e-05
## 7119:      Methanol       In      193.3254   32.042 0.0002575153 2.575153e-05
##        etime_min
##    1:     5 mins
##    2:     6 mins
##    3:     7 mins
##    4:     8 mins
##    5:     9 mins
##   ---
## 7115: 29001 mins
## 7116: 29002 mins
## 7117: 29003 mins
## 7118: 29004 mins
## 7119: 29005 mins
```

```
air[, etime_min := as.numeric(timestamp - as.POSIXct('03/10/2022 13:40', format = '%m/%d/%Y %H:%M'), uni
air
```

```
##         form  aircleaner           timestamp elapsed_time     mtzr
##    1: C10H16 PureAir2000 2022-03-10 13:45:00   0.00000000 136.1325
##    2: C10H16 PureAir2000 2022-03-10 13:46:00   0.01666667 136.1325
##    3: C10H16 PureAir2000 2022-03-10 13:47:00   0.03333333 136.1325
##    4: C10H16 PureAir2000 2022-03-10 13:48:00   0.05000000 136.1325
##    5: C10H16 PureAir2000 2022-03-10 13:49:00   0.06666667 136.1325
##   ---
## 7115:   CH4O  NoSmell4.2 2022-03-30 18:01:00   5.10000000  32.0335
## 7116:   CH4O  NoSmell4.2 2022-03-30 18:02:00   5.11666667  32.0335
## 7117:   CH4O  NoSmell4.2 2022-03-30 18:03:00   5.13333333  32.0335
## 7118:   CH4O  NoSmell4.2 2022-03-30 18:04:00   5.15000000  32.0335
## 7119:   CH4O  NoSmell4.2 2022-03-30 18:05:00   5.16666667  32.0335
##       compound_name flow_dir concentration mol_mass        concm    mass_flow
##    1:      Limonene       In      639.5513  136.228 0.0036218981 3.621898e-04
##    2:      Limonene       In      696.0174  136.228 0.0039416762 3.941676e-04
##    3:      Limonene       In      744.5909  136.228 0.0042167570 4.216757e-04
##    4:      Limonene       In      748.1602  136.228 0.0042369703 4.236970e-04
##    5:      Limonene       In      735.1984  136.228 0.0041635652 4.163565e-04
##   ---
## 7115:      Methanol      Out      196.1996   32.042 0.0002613438 2.613438e-05
## 7116:      Methanol      Out      196.6986   32.042 0.0002620085 2.620085e-05
## 7117:      Methanol      Out      197.1389   32.042 0.0002625950 2.625950e-05
## 7118:      Methanol      Out      195.0970   32.042 0.0002598751 2.598751e-05
## 7119:      Methanol       In      193.3254   32.042 0.0002575153 2.575153e-05
##       etime_min
##    1:         5
##    2:         6
##    3:         7
##    4:         8
##    5:         9
##   ---
## 7115:     29001
## 7116:     29002
## 7117:     29003
```

```
## 7118:     29004
## 7119:     29005
```

7. Grouped operations

```
summ <- air[, .(conc_mean = mean(concentration), conc_sd = sd(concentration)), by = .(aircleaner, flow_
summ
```

```
##      aircleaner flow_dir compound_name  conc_mean    conc_sd
##  1: PureAir2000       In      Limonene 236.140833 180.664169
##  2: PureAir2000      Out      Limonene 130.510286 103.007851
##  3:  NoSmell4.2       In      Limonene 136.359756 162.623852
##  4:  NoSmell4.2      Out      Limonene   7.639283  10.431104
##  5: PureAir2000       In  Acetaldehyde 449.924469  98.998369
##  6: PureAir2000      Out  Acetaldehyde 454.567072  90.466989
##  7:  NoSmell4.2       In  Acetaldehyde 422.528957 101.064376
##  8:  NoSmell4.2      Out  Acetaldehyde 400.491769  67.328837
##  9: PureAir2000       In   Acetic acid 129.285799  54.802916
## 10: PureAir2000      Out   Acetic acid  99.807736  29.420402
## 11:  NoSmell4.2       In   Acetic acid 103.180419  79.966900
## 12:  NoSmell4.2      Out   Acetic acid  30.311270  19.543211
## 13: PureAir2000       In       Acetone 693.295568 176.724781
## 14: PureAir2000      Out       Acetone 679.441257 140.144185
## 15:  NoSmell4.2       In       Acetone 328.435626 307.730374
## 16:  NoSmell4.2      Out       Acetone 124.188223  14.459754
## 17: PureAir2000       In      Butanone 510.982916 195.353095
## 18: PureAir2000      Out      Butanone 447.162991 110.646108
## 19:  NoSmell4.2       In      Butanone 230.317210 308.215928
## 20:  NoSmell4.2      Out      Butanone  14.837143  11.594848
## 21: PureAir2000       In      Isoprene 273.931101  81.724540
## 22: PureAir2000      Out      Isoprene 262.720859  67.494987
## 23:  NoSmell4.2       In      Isoprene 110.604514 149.514732
## 24:  NoSmell4.2      Out      Isoprene   9.569464   5.493605
## 25: PureAir2000       In  Benzaldehyde 272.506462 170.121305
## 26: PureAir2000      Out  Benzaldehyde 149.062778  92.595240
## 27:  NoSmell4.2       In  Benzaldehyde 214.215377 200.471036
## 28:  NoSmell4.2      Out  Benzaldehyde  16.816555  20.715427
## 29: PureAir2000       In       Toluene 213.520652 136.016226
## 30: PureAir2000      Out       Toluene 140.190740  74.142221
## 31:  NoSmell4.2       In       Toluene 123.564931 163.690328
## 32:  NoSmell4.2      Out       Toluene   5.499191   8.587825
## 33: PureAir2000       In      Methanol 664.936044 162.540508
## 34: PureAir2000      Out      Methanol 663.641502 159.934803
## 35:  NoSmell4.2       In      Methanol 269.553071  55.797548
## 36:  NoSmell4.2      Out      Methanol 261.641510  43.055782
##      aircleaner flow_dir compound_name  conc_mean    conc_sd
```

8. Export

```
fwrite(summ, 'output/cleaner_summary.csv')
```

9. Integrate

```
source('../functions-R/mintegrate.R')
```

```
air[, mass_cum := mintegrate(etime_min, mass_flow, method = 'trap'), by = .(aircleaner, compound_name, f
air
```

```
##         form  aircleaner           timestamp elapsed_time     mtzr
##    1: C10H16 PureAir2000 2022-03-10 13:45:00   0.00000000 136.1325
##    2: C10H16 PureAir2000 2022-03-10 13:46:00   0.01666667 136.1325
##    3: C10H16 PureAir2000 2022-03-10 13:47:00   0.03333333 136.1325
##    4: C10H16 PureAir2000 2022-03-10 13:48:00   0.05000000 136.1325
##    5: C10H16 PureAir2000 2022-03-10 13:49:00   0.06666667 136.1325
##   ---
## 7115:   CH4O  NoSmell4.2 2022-03-30 18:01:00   5.10000000  32.0335
## 7116:   CH4O  NoSmell4.2 2022-03-30 18:02:00   5.11666667  32.0335
## 7117:   CH4O  NoSmell4.2 2022-03-30 18:03:00   5.13333333  32.0335
## 7118:   CH4O  NoSmell4.2 2022-03-30 18:04:00   5.15000000  32.0335
## 7119:   CH4O  NoSmell4.2 2022-03-30 18:05:00   5.16666667  32.0335
##       compound_name flow_dir concentration mol_mass        concm    mass_flow
##    1:      Limonene       In      639.5513  136.228 0.0036218981 3.621898e-04
##    2:      Limonene       In      696.0174  136.228 0.0039416762 3.941676e-04
##    3:      Limonene       In      744.5909  136.228 0.0042167570 4.216757e-04
##    4:      Limonene       In      748.1602  136.228 0.0042369703 4.236970e-04
##    5:      Limonene       In      735.1984  136.228 0.0041635652 4.163565e-04
##   ---
## 7115:      Methanol      Out      196.1996   32.042 0.0002613438 2.613438e-05
## 7116:      Methanol      Out      196.6986   32.042 0.0002620085 2.620085e-05
## 7117:      Methanol      Out      197.1389   32.042 0.0002625950 2.625950e-05
## 7118:      Methanol      Out      195.0970   32.042 0.0002598751 2.598751e-05
## 7119:      Methanol       In      193.3254   32.042 0.0002575153 2.575153e-05
##       etime_min     mass_cum
##    1:         5 0.0000000000
##    2:         6 0.0003781787
##    3:         7 0.0007861004
##    4:         8 0.0012087867
##    5:         9 0.0016288135
##   ---
## 7115:     29001 0.0105637982
## 7116:     29002 0.0105899658
## 7117:     29003 0.0106161960
## 7118:     29004 0.0106423195
## 7119:     29005 0.0110559226
```

```
airtot <- air[, .(mass_tot = mintegrate(etime_min, mass_flow, method = 'trap', value = 'total')), by =
```

10. Reshape

```
head(airtot)
```

```
##     aircleaner compound_name flow_dir    mass_tot
## 1: PureAir2000      Limonene       In 0.062887209
## 2: PureAir2000      Limonene      Out 0.034763664
## 3:  NoSmell4.2      Limonene       In 0.023223945
```

```
## 4:  NoSmell4.2      Limonene      Out 0.001246885
## 5: PureAir2000  Acetaldehyde       In 0.039029157
## 6: PureAir2000  Acetaldehyde      Out 0.039434328
```

```
airw <- dcast(airtot, aircleaner + compound_name ~ flow_dir, value.var = 'mass_tot')
airw
```

```
##       aircleaner compound_name          In          Out
##  1:  NoSmell4.2  Acetaldehyde 0.023791907 0.0224286691
##  2:  NoSmell4.2   Acetic acid 0.007612748 0.0023939442
##  3:  NoSmell4.2       Acetone 0.023817244 0.0091450527
##  4:  NoSmell4.2  Benzaldehyde 0.028422072 0.0022710746
##  5:  NoSmell4.2      Butanone 0.020487881 0.0013117810
##  6:  NoSmell4.2      Isoprene 0.009279196 0.0008125889
##  7:  NoSmell4.2      Limonene 0.023223945 0.0012468854
##  8:  NoSmell4.2      Methanol 0.011055923 0.0106423195
##  9:  NoSmell4.2       Toluene 0.014098526 0.0005959644
## 10: PureAir2000  Acetaldehyde 0.039029157 0.0394343280
## 11: PureAir2000   Acetic acid 0.015048660 0.0118633107
## 12: PureAir2000       Acetone 0.079212838 0.0778036108
## 13: PureAir2000  Benzaldehyde 0.056331657 0.0311351986
## 14: PureAir2000      Butanone 0.072328418 0.0634952583
## 15: PureAir2000      Isoprene 0.036705295 0.0352861632
## 16: PureAir2000      Limonene 0.062887209 0.0347636642
## 17: PureAir2000      Methanol 0.041945483 0.0418636924
## 18: PureAir2000       Toluene 0.038527797 0.0253431198
```

```
head(airw)
```

```
##    aircleaner compound_name          In          Out
## 1: NoSmell4.2  Acetaldehyde 0.023791907 0.0224286691
## 2: NoSmell4.2   Acetic acid 0.007612748 0.0023939442
## 3: NoSmell4.2       Acetone 0.023817244 0.0091450527
## 4: NoSmell4.2  Benzaldehyde 0.028422072 0.0022710746
## 5: NoSmell4.2      Butanone 0.020487881 0.0013117810
## 6: NoSmell4.2      Isoprene 0.009279196 0.0008125889
```

```
airw[, rem_eff := 100 * (1 - Out / In)]
airw
```

```
##       aircleaner compound_name          In          Out    rem_eff
##  1:  NoSmell4.2  Acetaldehyde 0.023791907 0.0224286691  5.7298381
##  2:  NoSmell4.2   Acetic acid 0.007612748 0.0023939442 68.5534818
##  3:  NoSmell4.2       Acetone 0.023817244 0.0091450527 61.6032295
##  4:  NoSmell4.2  Benzaldehyde 0.028422072 0.0022710746 92.0094688
##  5:  NoSmell4.2      Butanone 0.020487881 0.0013117810 93.5972833
##  6:  NoSmell4.2      Isoprene 0.009279196 0.0008125889 91.2428957
##  7:  NoSmell4.2      Limonene 0.023223945 0.0012468854 94.6310351
##  8:  NoSmell4.2      Methanol 0.011055923 0.0106423195  3.7410097
##  9:  NoSmell4.2       Toluene 0.014098526 0.0005959644 95.7728601
## 10: PureAir2000  Acetaldehyde 0.039029157 0.0394343280 -1.0381235
## 11: PureAir2000   Acetic acid 0.015048660 0.0118633107 21.1669960
## 12: PureAir2000       Acetone 0.079212838 0.0778036108  1.7790394
## 13: PureAir2000  Benzaldehyde 0.056331657 0.0311351986 44.7287724
## 14: PureAir2000      Butanone 0.072328418 0.0634952583 12.2125719
## 15: PureAir2000      Isoprene 0.036705295 0.0352861632  3.8662864
## 16: PureAir2000      Limonene 0.062887209 0.0347636642 44.7206124
```

```
## 17: PureAir2000      Methanol 0.041945483 0.0418636924  0.1949935
## 18: PureAir2000       Toluene 0.038527797 0.0253431198 34.2212074
```

```
library(ggplot2)
ggplot(airw, aes(compound_name, rem_eff, colour = aircleaner)) +
  geom_point()
```