dplyr-and-tidyr-like functions written in base r

pv71u98h1

April 13, 2021

1 Introduction

R-package m61r gathers functions similar to the ones present in dplyr and tidyr, but only written in base r, and without requiring any dependencies.
The purpose of this package is informative.
All the functions only work with data.frames.

2 filter

```
> tmp <- filter_(CO2,~Plant=="Qn1")
> head(tmp)

  Plant   Type  Treatment conc uptake
1   Qn1 Quebec nonchilled   95   16.0
2   Qn1 Quebec nonchilled  175   30.4
3   Qn1 Quebec nonchilled  250   34.8
4   Qn1 Quebec nonchilled  350   37.2
5   Qn1 Quebec nonchilled  500   35.3
6   Qn1 Quebec nonchilled  675   39.2

>

> tmp <- filter_(CO2,~Type=="Quebec")
> head(tmp)

  Plant   Type  Treatment conc uptake
1   Qn1 Quebec nonchilled   95   16.0
2   Qn1 Quebec nonchilled  175   30.4
3   Qn1 Quebec nonchilled  250   34.8
4   Qn1 Quebec nonchilled  350   37.2
5   Qn1 Quebec nonchilled  500   35.3
6   Qn1 Quebec nonchilled  675   39.2

>
```

3 select

```
> tmp <- select_(CO2,~Type)
> head(tmp)
```

```
    Type
1 Quebec
2 Quebec
3 Quebec
4 Quebec
5 Quebec
6 Quebec
```

```
>
```

```
> tmp <- select_(CO2,~c(Plant,Type))
> head(tmp)
```

```
  Plant   Type
1   Qn1 Quebec
2   Qn1 Quebec
3   Qn1 Quebec
4   Qn1 Quebec
5   Qn1 Quebec
6   Qn1 Quebec
```

```
>
```

```
> tmp <- select_(CO2,~-Type)
> head(tmp)
```

```
  Plant  Treatment conc uptake
1   Qn1 nonchilled   95   16.0
2   Qn1 nonchilled  175   30.4
3   Qn1 nonchilled  250   34.8
4   Qn1 nonchilled  350   37.2
5   Qn1 nonchilled  500   35.3
6   Qn1 nonchilled  675   39.2
```

```
>
```

```
> tmp <- select_(CO2,variable=~-(Plant:Treatment))
> head(tmp)
```

```
  conc uptake
1   95   16.0
```

```
2  175   30.4
3  250   34.8
4  350   37.2
5  500   35.3
6  675   39.2

>
```

4 mutate/transmutate

```
> tmp <- mutate_(CO2,z=~conc/uptake)
> head(tmp)

  Plant   Type  Treatment conc uptake         z
1   Qn1 Quebec nonchilled   95   16.0  5.937500
2   Qn1 Quebec nonchilled  175   30.4  5.756579
3   Qn1 Quebec nonchilled  250   34.8  7.183908
4   Qn1 Quebec nonchilled  350   37.2  9.408602
5   Qn1 Quebec nonchilled  500   35.3 14.164306
6   Qn1 Quebec nonchilled  675   39.2 17.219388

>

> tmp <- mutate_(CO2,mean=~mean(uptake))
> head(tmp)

  Plant   Type  Treatment conc uptake mean
1   Qn1 Quebec nonchilled   95   16.0   NA
2   Qn1 Quebec nonchilled  175   30.4   NA
3   Qn1 Quebec nonchilled  250   34.8   NA
4   Qn1 Quebec nonchilled  350   37.2   NA
5   Qn1 Quebec nonchilled  500   35.3   NA
6   Qn1 Quebec nonchilled  675   39.2   NA

>

> tmp <- mutate_(CO2,z1=~uptake/conc,y=~conc/100)
> head(tmp)

  Plant   Type  Treatment conc uptake         z1    y
1   Qn1 Quebec nonchilled   95   16.0 0.16842105 0.95
2   Qn1 Quebec nonchilled  175   30.4 0.17371429 1.75
3   Qn1 Quebec nonchilled  250   34.8 0.13920000 2.50
4   Qn1 Quebec nonchilled  350   37.2 0.10628571 3.50
5   Qn1 Quebec nonchilled  500   35.3 0.07060000 5.00
6   Qn1 Quebec nonchilled  675   39.2 0.05807407 6.75
```

```
>
```

```
> tmp <- transmutate_(CO2,z2=~uptake/conc,y2=~conc/100)
> head(tmp)
```

```
          z2   y2
1 0.16842105 0.95
2 0.17371429 1.75
3 0.13920000 2.50
4 0.10628571 3.50
5 0.07060000 5.00
6 0.05807407 6.75
```

```
>
```

5 summarise

```
> tmp <- summarise_(CO2,mean=~mean(uptake),sd=~sd(uptake))
> tmp
```

```
    mean       sd
1 27.2131 10.81441
```

```
>
```

```
> tmp <- summarise_(CO2, group=~c(Type,Treatment),mean=~mean(uptake),sd=~sd(uptake))
> tmp
```

```
         Type  Treatment     mean       sd
1      Quebec nonchilled 35.33333 9.596371
2      Quebec    chilled 25.95238 7.402136
3 Mississippi nonchilled 31.75238 9.644823
4 Mississippi    chilled 15.81429 4.058976
```

```
>
```

6 arrange/desange

```
> tmp <- arrange_(CO2,~c(conc))
> head(tmp)
```

```
  Plant   Type  Treatment conc uptake
1   Qn1 Quebec nonchilled   95   16.0
```

```
2   Qn2 Quebec nonchilled   95   13.6
3   Qn3 Quebec nonchilled   95   16.2
4   Qc1 Quebec    chilled   95   14.2
5   Qc2 Quebec    chilled   95    9.3
6   Qc3 Quebec    chilled   95   15.1

>

> tmp <- arrange_(CO2,~c(Treatment,conc,uptake))
> head(tmp)

  Plant        Type  Treatment conc uptake
1   Mn1 Mississippi nonchilled   95   10.6
2   Mn3 Mississippi nonchilled   95   11.3
3   Mn2 Mississippi nonchilled   95   12.0
4   Qn2      Quebec nonchilled   95   13.6
5   Qn1      Quebec nonchilled   95   16.0
6   Qn3      Quebec nonchilled   95   16.2

>

> tmp <- desange_(CO2,~c(Treatment,conc,uptake))
> head(tmp)

  Plant        Type Treatment conc uptake
1   Qc2      Quebec   chilled 1000   42.4
2   Qc3      Quebec   chilled 1000   41.4
3   Qc1      Quebec   chilled 1000   38.7
4   Mc1 Mississippi   chilled 1000   21.9
5   Mc3 Mississippi   chilled 1000   19.9
6   Mc2 Mississippi   chilled 1000   14.4

>
```

7 join

```
> authors <- data.frame(
+              surname = I(c("Tukey", "Venables", "Tierney", "Ripley", "McNeil")),
+              nationality = c("US", "Australia", "US", "UK", "Australia"),
+              deceased = c("yes", rep("no", 4)))
> books <- data.frame(
+            name = I(c("Tukey", "Venables", "Tierney","Ripley",
+                  "Ripley", "McNeil", "R Core")),
+            title = c("Exploratory Data Analysis",
+                  "Modern Applied Statistics ...",
```

```
+                  "LISP-STAT",
+                  "Spatial Statistics", "Stochastic Simulation",
+                  "Interactive Data Analysis",
+                  "An Introduction to R"),
+             other.author = c(NA, "Ripley", NA, NA, NA, NA,"Venables & Smith"))
```

7.1 inner join

```
> authors <- data.frame(
+              surname = I(c("Tukey", "Venables", "Tierney", "Ripley", "McNeil")),
+              nationality = c("US", "Australia", "US", "UK", "Australia"),
+              deceased = c("yes", rep("no", 4)))
> books <- data.frame(
+            name = I(c("Tukey", "Venables", "Tierney","Ripley",
+                  "Ripley", "McNeil", "R Core")),
+            title = c("Exploratory Data Analysis",
+                  "Modern Applied Statistics ...",
+                  "LISP-STAT",
+                  "Spatial Statistics", "Stochastic Simulation",
+                  "Interactive Data Analysis",
+                  "An Introduction to R"),
+             other.author = c(NA, "Ripley", NA, NA, NA, NA,"Venables & Smith"))
> tmp <- inner_join_(authors,books, by.x = "surname", by.y = "name")
> tmp

   surname nationality deceased                         title other.author
1   McNeil   Australia       no     Interactive Data Analysis         <NA>
2   Ripley          UK       no            Spatial Statistics         <NA>
3   Ripley          UK       no         Stochastic Simulation         <NA>
4  Tierney          US       no                     LISP-STAT         <NA>
5    Tukey          US      yes     Exploratory Data Analysis         <NA>
6 Venables   Australia       no Modern Applied Statistics ...       Ripley

>
```

7.2 left join

```
> authors <- data.frame(
+              surname = I(c("Tukey", "Venables", "Tierney", "Ripley", "McNeil")),
+              nationality = c("US", "Australia", "US", "UK", "Australia"),
+              deceased = c("yes", rep("no", 4)))
> books <- data.frame(
+            name = I(c("Tukey", "Venables", "Tierney","Ripley",
+                  "Ripley", "McNeil", "R Core")),
+            title = c("Exploratory Data Analysis",
+                  "Modern Applied Statistics ...",
+                  "LISP-STAT",
```

```
+                  "Spatial Statistics", "Stochastic Simulation",
+                  "Interactive Data Analysis",
+                  "An Introduction to R"),
+             other.author = c(NA, "Ripley", NA, NA, NA, NA,"Venables & Smith"))
> tmp <- left_join_(authors,books, by.x = "surname", by.y = "name")
> tmp

   surname nationality deceased                         title other.author
1   McNeil   Australia       no     Interactive Data Analysis         <NA>
2   Ripley          UK       no            Spatial Statistics         <NA>
3   Ripley          UK       no         Stochastic Simulation         <NA>
4  Tierney          US       no                      LISP-STAT         <NA>
5    Tukey          US      yes     Exploratory Data Analysis         <NA>
6 Venables   Australia       no Modern Applied Statistics ...       Ripley

>
```

7.3 right join

```
> authors <- data.frame(
+              surname = I(c("Tukey", "Venables", "Tierney", "Ripley", "McNeil")),
+              nationality = c("US", "Australia", "US", "UK", "Australia"),
+              deceased = c("yes", rep("no", 4)))
> books <- data.frame(
+            name = I(c("Tukey", "Venables", "Tierney","Ripley",
+                  "Ripley", "McNeil", "R Core")),
+            title = c("Exploratory Data Analysis",
+                  "Modern Applied Statistics ...",
+                  "LISP-STAT",
+                  "Spatial Statistics", "Stochastic Simulation",
+                  "Interactive Data Analysis",
+                  "An Introduction to R"),
+             other.author = c(NA, "Ripley", NA, NA, NA, NA,"Venables & Smith"))
> tmp <- right_join_(authors,books, by.x = "surname", by.y = "name")
> tmp

   surname nationality deceased                         title     other.author
1   McNeil   Australia       no     Interactive Data Analysis             <NA>
2   R Core        <NA>     <NA>           An Introduction to R Venables & Smith
3   Ripley          UK       no            Spatial Statistics             <NA>
4   Ripley          UK       no         Stochastic Simulation             <NA>
5  Tierney          US       no                      LISP-STAT             <NA>
6    Tukey          US      yes     Exploratory Data Analysis             <NA>
7 Venables   Australia       no Modern Applied Statistics ...           Ripley

>
```

7.4 full join

```
> authors <- data.frame(
+              surname = I(c("Tukey", "Venables", "Tierney", "Ripley", "McNeil")),
+              nationality = c("US", "Australia", "US", "UK", "Australia"),
+              deceased = c("yes", rep("no", 4)))
> books <- data.frame(
+            name = I(c("Tukey", "Venables", "Tierney","Ripley",
+                  "Ripley", "McNeil", "R Core")),
+            title = c("Exploratory Data Analysis",
+                  "Modern Applied Statistics ...",
+                  "LISP-STAT",
+                  "Spatial Statistics", "Stochastic Simulation",
+                  "Interactive Data Analysis",
+                  "An Introduction to R"),
+             other.author = c(NA, "Ripley", NA, NA, NA, NA,"Venables & Smith"))
> tmp <- full_join_(authors,books, by.x = "surname", by.y = "name")
> tmp

   surname nationality deceased                         title     other.author
1   McNeil   Australia       no     Interactive Data Analysis             <NA>
2   R Core        <NA>     <NA>           An Introduction to R Venables & Smith
3   Ripley          UK       no             Spatial Statistics             <NA>
4   Ripley          UK       no          Stochastic Simulation             <NA>
5  Tierney          US       no                       LISP-STAT             <NA>
6    Tukey          US      yes     Exploratory Data Analysis             <NA>
7 Venables   Australia       no Modern Applied Statistics ...           Ripley

>
```

7.5 semi join

```
> authors <- data.frame(
+              surname = I(c("Tukey", "Venables", "Tierney", "Ripley", "McNeil")),
+              nationality = c("US", "Australia", "US", "UK", "Australia"),
+              deceased = c("yes", rep("no", 4)))
> books <- data.frame(
+            name = I(c("Tukey", "Venables", "Tierney","Ripley",
+                  "Ripley", "McNeil", "R Core")),
+            title = c("Exploratory Data Analysis",
+                  "Modern Applied Statistics ...",
+                  "LISP-STAT",
+                  "Spatial Statistics", "Stochastic Simulation",
+                  "Interactive Data Analysis",
+                  "An Introduction to R"),
+             other.author = c(NA, "Ripley", NA, NA, NA, NA,"Venables & Smith"))
> tmp <- semi_join_(authors,books, by.x = "surname", by.y = "name")
> tmp
```

```
   surname nationality deceased
1    Tukey          US      yes
2 Venables   Australia       no
3  Tierney          US       no
4   Ripley          UK       no
5   McNeil   Australia       no

>
```

7.6 anti join

```
> authors <- data.frame(
+              surname = I(c("Tukey", "Venables", "Tierney", "Ripley", "McNeil")),
+              nationality = c("US", "Australia", "US", "UK", "Australia"),
+              deceased = c("yes", rep("no", 4)))
> books <- data.frame(
+            name = I(c("Tukey", "Venables", "Tierney","Ripley",
+                  "Ripley", "McNeil", "R Core")),
+            title = c("Exploratory Data Analysis",
+                  "Modern Applied Statistics ...",
+                  "LISP-STAT",
+                  "Spatial Statistics", "Stochastic Simulation",
+                  "Interactive Data Analysis",
+                  "An Introduction to R"),
+             other.author = c(NA, "Ripley", NA, NA, NA, NA,"Venables & Smith"))
> tmp <- anti_join_(authors,books, by.x = "surname", by.y = "name")
> tmp

[1] surname     nationality deceased
<0 rows> (or 0-length row.names)

> tmp <- anti_join_(books,authors, by.x = "name", by.y = "surname")
> tmp

    name                title     other.author
7 R Core An Introduction to R Venables & Smith

>
```

8 reshape: merge/spread

8.1 merge

```
> df3 <- data.frame(id = 1:4,
+                   age = c(40,50,60,50),
```

```
+                   dose.a1 = c(1,2,1,2),
+                   dose.a2 = c(2,1,2,1),
+                   dose.a14 = c(3,3,3,3))
> df3
```

```
  id age dose.a1 dose.a2 dose.a14
1  1  40       1       2        3
2  2  50       2       1        3
3  3  60       1       2        3
4  4  50       2       1        3
```

```
> gather_(df3,pivot = c("id","age"))
```

```
   id age parameters values
1   1  40    dose.a1      1
2   2  50    dose.a1      2
3   3  60    dose.a1      1
4   4  50    dose.a1      2
5   1  40    dose.a2      2
6   2  50    dose.a2      1
7   3  60    dose.a2      2
8   4  50    dose.a2      1
9   1  40   dose.a14      3
10  2  50   dose.a14      3
11  3  60   dose.a14      3
12  4  50   dose.a14      3
```

```
>
```

8.2 spread

```
> df3 <- data.frame(id = 1:4,
+                   age = c(40,50,60,50),
+                   dose.a1 = c(1,2,1,2),
+                   dose.a2 = c(2,1,2,1),
+                   dose.a14 = c(3,3,3,3))
> df3
```

```
  id age dose.a1 dose.a2 dose.a14
1  1  40       1       2        3
2  2  50       2       1        3
3  3  60       1       2        3
4  4  50       2       1        3
```

```
> gather_(df3,pivot = c("id","age"))
```

```
   id age parameters values
1   1  40    dose.a1      1
2   2  50    dose.a1      2
3   3  60    dose.a1      1
4   4  50    dose.a1      2
5   1  40    dose.a2      2
6   2  50    dose.a2      1
7   3  60    dose.a2      2
8   4  50    dose.a2      1
9   1  40   dose.a14      3
10  2  50   dose.a14      3
11  3  60   dose.a14      3
12  4  50   dose.a14      3
```

```
> df4 <- gather_(df3,pivot = c("id","age"))
> df5 <- rbind(df4,
+   data.frame(id=5, age=20,parameters="dose.a14",values=8),
+   data.frame(id=6, age=10,parameters="dose.a1",values=5))
> df5
```

```
   id age parameters values
1   1  40    dose.a1      1
2   2  50    dose.a1      2
3   3  60    dose.a1      1
4   4  50    dose.a1      2
5   1  40    dose.a2      2
6   2  50    dose.a2      1
7   3  60    dose.a2      2
8   4  50    dose.a2      1
9   1  40   dose.a14      3
10  2  50   dose.a14      3
11  3  60   dose.a14      3
12  4  50   dose.a14      3
13  5  20   dose.a14      8
14  6  10    dose.a1      5
```

```
> spread_(df5,col_name="parameters",col_values="values",pivot=c("id","age"))
```

```
  id age dose.a1 dose.a2 dose.a14
1  1  40       1       2        3
2  2  50       2       1        3
3  3  60       1       2        3
4  4  50       2       1        3
5  5  20      NA      NA        8
6  6  10       5      NA       NA
```