**library(tidyr; library(dplyr); library(reshape2)**

#- functions gather(), separate() and spread(), from tidyr, with the functions melt(), colsplit() and dcast(), from reshape2.

messy <- data.frame(id = 1:40,

trt = sample(rep(c('control', 'treatment'), each =20)),

work.T1 = round( .... [TRUNCATED]

id trt work.T1 home.T1 work.T2 home.T2

1 1 treatment 0.28 0.86 0.11 0.72

2 2 control 0.23 0.46 0.33 0.52

3 3 control 0.01 0.22 0.32 0.41

gathered.messy <- gather(messy, key, value, -id, -trt)

id trt key value

1 1 treatment work.T1 0.28

2 2 control work.T1 0.23

3 3 control work.T1 0.01

molten.messy <- melt(messy,

variable.name = "key",

value.names = "value",

id.vars = c("id", "trt"))

id trt key value

1 1 treatment work.T1 0.28

2 2 control work.T1 0.23

3 3 control work.T1 0.01

**## Split a column: separate() vs colsplit()**

tidy <- separate(gathered.messy,

key, into = c("location", "time"), sep = "\\.)

id trt location time value

1 1 treatment work T1 0.28

2 2 control work T1 0.23

3 3 control work T1 0.01

res.tidy <- cbind(molten.messy[1:2],

colsplit(molten.messy[, 3], "\\.", c("location", "time")),

molten.messy[4])

id trt location time value

1 1 treatment work T1 0.28

2 2 control work T1 0.23

3 3 control work T1 0.01

tbl\_df(messy) %%

gather( key, value, -id, -trt) %%

separate(key, into = c("location", "time"), sep = "\\.")

Source: local data frame [160 x 5]

id trt location time value

(int) (fctr) (chr) (chr) (dbl)

1 1 treatment work T1 0.28

2 2 control work T1 0.23

3 3 control work T1 0.01

4 4 treatment work T1 0.73

.. ... ... ... ... ...

**#-- From the long to the wide format: spread() vs dcast()**

spread.tidy <- spread(tidy, location, value)

id trt time home work

1 1 treatment T1 0.86 0.28

2 1 treatment T2 0.72 0.11

3 2 control T1 0.46 0.23

cast.tidy <- dcast(res.tidy, formula = id trt time~ location,

value.var = "value")

id trt time home work

1 1 treatment T1 0.86 0.28

2 1 treatment T2 0.72 0.11

3 2 control T1 0.46 0.23

dcast(res.tidy, formula = trt ~ location time, value.var = "value",

fun.aggregate =function(x){

paste0(

round(mea .... [TRUNCATED]

trt home\_T1 home\_T2 work\_T1 work\_T2

1 control 0.47 (0.31) 0.48 (0.27) 0.35 (0.23) 0.47 (0.29)

2 treatment 0.57 (0.23) 0.56 (0.24) 0.44 (0.28) 0.4 (0.33)

head(tips)

total\_bill tip sex smoker day time size

1 16.99 1.01 Female No Sun Dinner 2

2 10.34 1.66 Male No Sun Dinner 3

3 21.01 3.50 Male No Sun Dinner 3

m.tips <- melt(tips)

dcast(m.tips, day time ~ variable, mean3)

day time total\_bill tip size

1 Fri Dinner 19.7 2.9 2.2

2 Fri Lunch 12.8 2.4 2.0

3 Sat Dinner 20.4 3.0 2.5

tbl\_df(m.tips) %%

group\_by(day, time, variable) %%

summarise (mean = mean3(value)) %%

spread(variable, mean)

day time total\_bill tip size

(fctr) (fctr) (dbl) (dbl) (dbl)

1 Fri Dinner 19.7 2.9 2.2

2 Fri Lunch 12.8 2.4 2.0

3 Sat Dinner 20.4 3.0 2.5

tbl\_df(m.tips) %>%

group\_by(day, time, variable) %>%

summarise (mean = mean3(value), sd= sd3(value)) %>%

melt(variable.name = "key") %>%

dcast(day + time ~ variable + key)

Using day, time, variable as id variables

day time total\_bill\_mean total\_bill\_sd tip\_mean tip\_sd size\_mean size\_sd

1 Fri Dinner 19.7 9.5 2.9 1.2 2.2 0.6

2 Fri Lunch 12.8 2.8 2.4 0.7 2.0 0.6

3 Sat Dinner 20.4 9.5 3.0 1.6 2.5 0.8