```r
#***************************#
#**GH811 Tidy Data Example**#
#***************************#

library(tidyr)

####gather()####

#In this experiment we've given three people two different drugs and recorded their
heart rate

#Generate dataset
messy <- data.frame(
  name = c("Wilbur", "Petunia", "Gregory"),
  a = c(67, 80, 64),
  b = c(56, 90, 50)
)

messy

#>      name  a  b
#> 1  Wilbur 67 56
#> 2 Petunia 80 90
#> 3 Gregory 64 50

#We have three variables (name, drug and heartrate), but only name is currently in a
column.
#We use gather() to gather the a and b columns into key-value pairs of drug and
heartrate:

messy %>%
  gather(drug, heartrate, a:b)

#>      name drug heartrate
#> 1  Wilbur    a        67
#> 2 Petunia    a        80
#> 3 Gregory    a        64
#> 4  Wilbur    b        56
#> 5 Petunia    b        90
#> 6 Gregory    b        50


set.seed(10)
messy <- data.frame(
  id = 1:4,
  trt = sample(rep(c('control', 'treatment'), each = 2)),
  work.T1 = runif(4),
  home.T1 = runif(4),
  work.T2 = runif(4),
  home.T2 = runif(4)
)

####separate()####

#If two variables are clumped together in one column. separate() allows you to tease
them apart

#We have some measurements of how much time people spend on their phones
#Measured at two locations (work and home), at two times
#Each person has been randomly assigned to either treatment or control

#Generate dataset
set.seed(10)
```

```
messy <- data.frame(
  id = 1:4,
  trt = sample(rep(c('control', 'treatment'), each = 2)),
  work.T1 = runif(4),
  home.T1 = runif(4),
  work.T2 = runif(4),
  home.T2 = runif(4)
)

#>  id       trt    work.T1   home.T1   work.T2    home.T2
#>1  1 treatment 0.08513597 0.6158293 0.1135090 0.05190332
#>2  2   control 0.22543662 0.4296715 0.5959253 0.26417767
#>3  3 treatment 0.27453052 0.6516557 0.3580500 0.39879073
#>4  4   control 0.27230507 0.5677378 0.4288094 0.83613414

#First use gather() to turn columns work.T1, home.T1, work.T2 and home.T2 into key-
value pair of key and time.

tidier <- messy %>%
  gather(key, time, -id, -trt)
tidier %>% head(8)

#>  id       trt      key        time
#>1  1 treatment work.T1 0.08513597
#>2  2   control work.T1 0.22543662
#>3  3 treatment work.T1 0.27453052
#>4  4   control work.T1 0.27230507
#>5  1 treatment home.T1 0.61582931
#>6  2   control home.T1 0.42967153
#>7  3 treatment home.T1 0.65165567
#>8  4   control home.T1 0.56773775

#Next we use separate() to split the key into location and period (t1, t2)
#Note a regular expression is used to describe the character that separates them
#might need to install package Rcpp if you get an error message below

tidy <- tidier %>%
  separate(key, into = c("location", "period"), sep = "\\.")
tidy %>% head(8)

#>  id       trt location period      time
#>1  1 treatment     work     T1 0.08513597
#>2  2   control     work     T1 0.22543662
#>3  3 treatment     work     T1 0.27453052
#>4  4   control     work     T1 0.27230507
#>5  1 treatment     home     T1 0.61582931
#>6  2   control     home     T1 0.42967153
#>7  3 treatment     home     T1 0.65165567
#>8  4   control     home     T1 0.56773775

####spread()####

#The last tool, spread(), takes two columns (a key-value pair) and spreads them in to
multiple columns
#Makes "long" data wider
#It is used when you have variables that form rows instead of columns
#You need spread() less frequently than gather() or separate()
#To learn more, refer to Hadley Whickam's tidyr documentation and demos
```