Homework 3

David Edwards

2019-09-17

Problem 3

Redo Problem 4 parts a-d from last time using the tidyverse functions and piping.

a. Sensory data from five operators.
   http://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/Sensory.dat
b. Gold Medal performance for Olympic Men's Long Jump, year is coded as 1900=0.
   http://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/LongJumpData.dat
c. Brain weight (g) and body weight (kg) for 62 species.
   http://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/BrainandBodyWeight.dat
d. Triplicate measurements of tomato yield for two varieties of tomatos at three planting densities.
   http://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/tomato.dat

Part a

Here is my code for part a. Didn't really see a need for piping in this one but it was much easier being able to actually read in the data to a table vs parsing it line by line as I had to do the previous week.

```
url <- "http://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/Sensory.dat"
data <- fread(url, fill = TRUE, sep = " ",  skip = 2, col.names = c("Item", "Op1", "Op2",
                                                                    "Op3","Op4","Op5"))
j <-1
for (i in seq(1,dim(data)[1], 3)){
  data[i+1] = c(j, data[i+1,1:5])
  data[i+2] = c(j, data[i+2,1:5])
  j<- j+1
}
data <- gather(data, Operator, Value, -Item)
data
```

```
##     Item Operator Value
## 1      1      Op1   4.3
## 2      1      Op1   4.3
## 3      1      Op1   4.1
## 4      2      Op1   6.0
## 5      2      Op1   4.9
## 6      2      Op1   6.0
## 7      3      Op1   2.4
## 8      3      Op1   3.9
## 9      3      Op1   1.9
## 10     4      Op1   7.4
## 11     4      Op1   7.1
## 12     4      Op1   6.4
## 13     5      Op1   5.7
## 14     5      Op1   5.8
## 15     5      Op1   5.8
## 16     6      Op1   2.2
## 17     6       Op1   3.0
## 18     6       Op1   2.1
## 19     7       Op1   1.2
## 20     7       Op1   1.3
## 21     7       Op1   0.9
## 22     8       Op1   4.2
## 23     8       Op1   3.0
## 24     8       Op1   4.8
## 25     9       Op1   8.0
## 26     9       Op1   9.0
## 27     9       Op1   8.9
## 28    10       Op1   5.0
## 29    10       Op1   5.4
## 30    10       Op1   2.8
## 31     1       Op2   4.9
## 32     1       Op2   4.5
## 33     1       Op2   5.3
## 34     2       Op2   5.3
## 35     2       Op2   6.3
## 36     2       Op2   5.9
## 37     3       Op2   2.5
## 38     3       Op2   3.0
## 39     3       Op2   3.9
## 40     4       Op2   8.2
## 41     4       Op2   7.9
## 42     4       Op2   7.1
## 43     5       Op2   6.3
## 44     5       Op2   5.7
## 45     5       Op2   6.0
## 46     6       Op2   2.4
## 47     6       Op2   1.8
## 48     6       Op2   3.3
## 49     7       Op2   1.5
## 50     7       Op2   2.4
## 51     7       Op2   3.1
## 52     8       Op2   4.8
## 53     8       Op2   4.5
## 54     8       Op2   4.8
## 55     9       Op2   8.6
## 56     9       Op2   7.7
## 57     9       Op2   9.2
## 58    10       Op2   4.8
## 59    10       Op2   5.0
## 60    10       Op2   5.2
## 61     1       Op3   3.3
## 62     1       Op3   4.0
## 63     1       Op3   3.4
## 64     2       Op3   4.5
## 65     2       Op3   4.2
## 66     2       Op3   4.7
## 67     3       Op3   2.3
## 68     3       Op3   2.8
## 69     3       Op3   2.6
## 70     4       Op3   6.4
## 71     4      Op3   5.9
## 72     4      Op3   6.9
## 73     5      Op3   5.4
## 74     5      Op3   5.4
## 75     5      Op3   6.1
## 76     6      Op3   1.7
## 77     6      Op3   2.1
## 78     6      Op3   1.1
## 79     7      Op3   1.2
## 80     7      Op3   0.8
## 81     7      Op3   1.1
## 82     8      Op3   4.5
## 83     8      Op3   4.7
## 84     8      Op3   4.7
## 85     9      Op3   9.0
## 86     9      Op3   6.7
## 87     9      Op3   8.1
## 88    10      Op3   3.9
## 89    10      Op3   3.4
## 90    10      Op3   4.1
## 91     1      Op4   5.3
## 92     1      Op4   5.5
## 93     1      Op4   5.7
## 94     2      Op4   5.9
## 95     2      Op4   5.5
## 96     2      Op4   6.3
## 97     3      Op4   3.1
## 98     3      Op4   2.7
## 99     3      Op4   4.6
## 100    4      Op4   6.8
## 101    4      Op4   7.3
## 102    4      Op4   7.0
## 103    5      Op4   6.1
## 104    5      Op4   6.2
## 105    5      Op4   7.0
## 106    6      Op4   3.4
## 107    6      Op4   4.0
## 108    6      Op4   3.3
## 109    7      Op4   0.9
## 110    7      Op4   1.2
## 111    7      Op4   1.9
## 112    8      Op4   4.6
## 113    8      Op4   4.9
## 114    8      Op4   4.8
## 115    9      Op4   9.4
## 116    9      Op4   9.0
## 117    9      Op4   9.1
## 118   10      Op4   5.5
## 119   10      Op4   4.9
## 120   10      Op4   3.9
## 121    1      Op5   4.4
## 122    1      Op5   3.3
## 123    1      Op5   4.7
## 124    2      Op5   4.7
## 125    2       Op5   4.9
## 126    2       Op5   4.6
## 127    3       Op5   2.4
## 128    3       Op5   1.3
## 129    3       Op5   2.2
## 130    4       Op5   6.0
## 131    4       Op5   6.1
## 132    4       Op5   6.7
## 133    5       Op5   5.9
## 134    5       Op5   6.5
## 135    5       Op5   4.9
## 136    6       Op5   1.7
## 137    6       Op5   1.7
## 138    6       Op5   2.1
## 139    7       Op5   0.7
## 140    7       Op5   1.3
## 141    7       Op5   1.6
## 142    8       Op5   3.2
## 143    8       Op5   4.6
## 144    8       Op5   4.3
## 145    9       Op5   8.8
## 146    9       Op5   7.9
## 147    9       Op5   7.6
## 148   10       Op5   3.8
## 149   10       Op5   4.6
## 150   10       Op5   5.5
```

Part b

I don't remember seeing a function in dplyr or tidyr that would have helped with this one (I'm sure I missed something). However, this time I was able to just read in the file and then combine the columns in a much simpler way than before.

```
url <- "http://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/LongJumpData.dat"
data <- fread(url, fill = TRUE, skip = 1)
mydf <- data.frame("Year"=c(data$V1,data$V3,data$V5,data$V7[!is.na(data$V7)]),
                   "LongJump"=c(data$V2, data$V4, data$V6, data$V8[!is.na(data$V8)]))
mydf
```

```
##    Year LongJump
## 1    -4   249.75
## 2     0   282.88
## 3     4   289.00
## 4     8   294.50
## 5    12   299.25
## 6    20   281.50
## 7    24   293.13
## 8    28   304.75
## 9    32   300.75
## 10   36   317.31
## 11   48   308.00
## 12   52   298.00
## 13   56   308.25
## 14   60   319.75
## 15   64   317.75
## 16   68   350.50
## 17   72   324.50
## 18   76   328.50
## 19   80   336.25
## 20   84   336.25
## 21   88   343.25
## 22   92   342.50
```

Part c

Just like part b as this is almost the same problem (in terms of cleaning the data)

```
url <- "http://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/BrainandBodyWeight.dat"
data <- fread(url, fill = TRUE, skip = 1)
mydf <- data.frame("BodyWt"=c(data$V1,data$V3,data$V5[!is.na(data$V5)]),
                   "BrainWt"=c(data$V2, data$V4, data$V6[!is.na(data$V6)]))

mydf
```

```
##      BodyWt BrainWt
## 1     3.385   44.50
## 2     0.480   15.50
## 3     1.350    8.10
## 4   465.000  423.00
## 5    36.330  119.50
## 6    27.660  115.00
## 7    14.830   98.20
## 8     1.040    5.50
## 9     4.190   58.00
## 10    0.425    6.40
## 11    0.101    4.00
## 12    0.920    5.70
## 13    1.000    6.60
## 14    0.005    0.10
## 15    0.060    1.00
## 16    3.500   10.80
## 17    2.000   12.30
## 18    1.700    6.30
## 19 2547.000 4603.00
## 20    0.023    0.30
## 21  187.100  419.00
## 22  521.000  655.00
## 23    0.785    3.50
## 24   10.000  115.00
## 25    3.300   25.60
## 26    0.200    5.00
## 27    1.410   17.50
## 28  529.000  680.00
## 29  207.000  406.00
## 30   85.000  325.00
## 31    0.750   12.30
## 32   62.000 1320.00
## 33 6654.000 5712.00
## 34    3.500    3.90
## 35    6.800  179.00
## 36   35.000   56.00
## 37    4.050   17.00
## 38    0.120    1.00
## 39    0.023    0.40
## 40    0.010    0.30
## 41    1.400   12.50
## 42  250.000  490.00
## 43    2.500   12.10
## 44   55.500  175.00
## 45  100.000  157.00
## 46   52.160  440.00
## 47   10.550  179.50
## 48    0.550    2.40
## 49   60.000   81.00
## 50    3.600   21.00
## 51    4.288   39.20
## 52    0.280    1.90
## 53    0.075    1.20
## 54    0.122    3.00
## 55    0.048    0.33
## 56  192.000  180.00
## 57    3.000   25.00
## 58  160.000  169.00
## 59    0.900    2.60
## 60    1.620   11.40
## 61    0.104    2.50
## 62    4.235   50.40
```

Part d

For this problem I was able to take advantage of the functions in dplyr and tidyr to help tidy up the data set. The pipes also really helped.

```
url <- "http://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/tomato.dat"
data <- fread(url, fill = TRUE, header = T, skip = 1, col.names = c("Treatment", "10000","20000","30000
mydf <- bind_rows(data[1,],data[2,])
mydf <- mydf %>% gather(Levels,Values,-Treatment)%>% separate(Values,c("V1","V2","V3"), sep = ",") %>%ga
```

```
## Warning: Expected 3 pieces. Additional pieces discarded in 1 rows [2].
```

```
mydf
```

```
##          Treatment Levels Values
## 1          Ife\\#1  10000   16.1
## 2  PusaEarlyDwarf  10000    8.1
## 3          Ife\\#1  20000   16.6
## 4  PusaEarlyDwarf  20000   12.7
## 5          Ife\\#1  30000   20.8
## 6  PusaEarlyDwarf  30000   14.4
## 7          Ife\\#1  10000   15.3
## 8  PusaEarlyDwarf  10000    8.6
## 9          Ife\\#1  20000   19.2
## 10 PusaEarlyDwarf  20000   13.7
## 11         Ife\\#1  30000   18.0
## 12 PusaEarlyDwarf  30000   15.4
## 13         Ife\\#1  10000   17.5
## 14 PusaEarlyDwarf  10000   10.1
## 15        Ife\\#1  20000   18.5
## 16 PusaEarlyDwarf  20000   11.5
## 17        Ife\\#1  30000   21.0
## 18 PusaEarlyDwarf  30000   13.7
```

Problem 4

Finish this homework by pushing your changes to your repo. In general, your workflow for this should be:

1. In terminal: git pull – to make sure you have the most recent local repo

2. In terminal: do some work

3. In terminal: git add – check files you want to commit

4. In terminal: git commit – make message INFORMATIVE and USEFUL

5. In terminal: git push – this pushes your local changes to the repo

If you have difficulty with steps 1-5, git is not correctly or completely setup.

Only submit the .Rmd and .pdf solution files. Names should be formatted HW3_lastname_firstname.Rmd and HW3_lastname_firstname.pdf

Optional preperation for next class:

TBD – could be something sent as a class message