Homework 4

Evan Yacek ety78

This homework is due on Feb. 16, 2016 at 11:59pm. Please submit as a PDF file on Canvas.

Question 1: (4 pts) *Recall the* `HairEyeColor` *data set from an earlier in-class exercise. This data set contains counts of males and females with different combinations of hair and eye color.*

```
HairEyeColor
```

```
## , , Sex = Male
##
##        Eye
## Hair    Brown Blue Hazel Green
##   Black    32   11    10     3
##   Brown    53   50    25    15
##   Red      10   10     7     7
##   Blond     3   30     5     8
##
## , , Sex = Female
##
##        Eye
## Hair    Brown Blue Hazel Green
##   Black    36    9     5     2
##   Brown    66   34    29    14
##   Red      16    7     7     7
##   Blond     4   64     5     8
```

I have split the data set into two data-frames, one `male` *and one* `female`*. Using the* `dplyr` *and* `tidyr` *packages, make these data-frames tidy and then combine them into a single data-frame. Make sure that your final data-frame has a* `sex` *column indicating which data-frame the observations originally came from.* ***HINT****: You can use the* `bind_rows` *function to add rows from one data-frame onto another as long as both data-frames have identical column names.*

```
male <- read.table(text="
Hair    Brown Blue Hazel Green
  Black    32   11    10     3
  Brown    53   50    25    15
  Red      10   10     7     7
  Blond     3   30     5     8
", head=T)

m <- male %>% select(Hair, Brown, Blue, Hazel, Green) %>%
  gather(eye_color, number_of_people, Brown:Green) %>%
  arrange(Hair, eye_color)
m["Sex"] <- "M"
m
```

```
##     Hair eye_color number_of_people Sex
## 1  Black      Blue               11   M
## 2  Black     Brown               32   M
## 3  Black     Green                3   M
## 4  Black     Hazel               10   M
## 5  Blond      Blue               30   M
## 6  Blond     Brown                3   M
## 7  Blond     Green                8   M
## 8  Blond     Hazel                5   M
## 9  Brown      Blue               50   M
## 10 Brown     Brown               53   M
## 11 Brown     Green               15   M
## 12 Brown     Hazel               25   M
## 13   Red      Blue               10   M
## 14   Red     Brown               10   M
## 15   Red     Green                7   M
## 16   Red     Hazel                7   M
```

```
female <- read.table(text="
Hair    Brown Blue Hazel Green
  Black    36    9     5     2
  Brown    66   34    29    14
  Red      16    7     7     7
  Blond     4   64     5     8
", head=T)

f <- female %>% select(Hair, Brown, Blue, Hazel, Green) %>%
  gather(eye_color, number_of_people, Brown:Green) %>%
  arrange(Hair, eye_color)

f["Sex"] <- "F"
f
```

```
##     Hair eye_color number_of_people Sex
## 1  Black      Blue                9   F
## 2  Black     Brown               36   F
## 3  Black     Green                2   F
## 4  Black     Hazel                5   F
## 5  Blond      Blue               64   F
## 6  Blond     Brown                4   F
## 7  Blond     Green                8   F
## 8  Blond     Hazel                5   F
## 9  Brown      Blue               34   F
## 10 Brown     Brown               66   F
## 11 Brown     Green               14   F
## 12 Brown     Hazel               29   F
## 13   Red      Blue                7   F
## 14   Red     Brown               16   F
## 15   Red     Green                7   F
## 16   Red     Hazel                7   F
```

```
combineMF <- rbind(m, f)
combineMF
```

```
##     Hair eye_color number_of_people Sex
## 1  Black      Blue               11   M
## 2  Black     Brown               32   M
## 3  Black     Green                3   M
## 4  Black     Hazel               10   M
## 5  Blond      Blue               30   M
## 6  Blond     Brown                3   M
## 7  Blond     Green                8   M
## 8  Blond     Hazel                5   M
## 9  Brown      Blue               50   M
## 10 Brown     Brown               53   M
## 11 Brown     Green               15   M
## 12 Brown     Hazel               25   M
## 13   Red      Blue               10   M
## 14   Red     Brown               10   M
## 15   Red     Green                7   M
## 16   Red     Hazel                7   M
## 17 Black      Blue                9   F
## 18 Black     Brown               36   F
## 19 Black     Green                2   F
## 20 Black     Hazel                5   F
## 21 Blond      Blue               64   F
## 22 Blond     Brown                4   F
## 23 Blond     Green                8   F
## 24 Blond     Hazel                5   F
## 25 Brown      Blue               34   F
## 26 Brown     Brown               66   F
## 27 Brown     Green               14   F
## 28 Brown     Hazel               29   F
## 29   Red      Blue                7   F
## 30   Red     Brown               16   F
## 31   Red     Green                7   F
## 32   Red     Hazel                7   F
```

Using the data-frame you created above, compute the total counts for each sex (i.e., the sum of the counts for each sex).

```
groups <- group_by(combineMF,Hair,eye_color,Sex)
count_table <- summarize(groups,count=number_of_people)
count_table
```

```
## Source: local data frame [32 x 4]
## Groups: Hair, eye_color [?]
##
##      Hair eye_color   Sex count
##    (fctr)     (chr) (chr) (int)
## 1   Black      Blue     F     9
## 2   Black      Blue     M    11
## 3   Black     Brown     F    36
## 4   Black     Brown     M    32
## 5   Black     Green     F     2
## 6   Black     Green     M     3
## 7   Black     Hazel     F     5
## 8   Black     Hazel     M    10
## 9   Blond      Blue     F    64
## 10  Blond      Blue     M    30
## ..    ...       ...   ...   ...
```

Question 2: (3 pts) *Recall that the `InsectSprays` data-set contains information about the effectiveness of different insecticides. The different insecticides are labeled A through F in the `spray` column. I have created a new data-frame (`spray_names`), that contains the full names of each insecticide. Using one of the `dplyr` join functions, combine the two data-frames so that there is an additional `full_name` column and all of the original columns and rows in `InsectSprays` are retained. Which join function is most appropriate to use and why?*

```
head(InsectSprays)
```

```
##   count spray
## 1    10     A
## 2     7     A
## 3    20     A
## 4    14     A
## 5    14     A
## 6    12     A
```

```
spray_names <- read.table(text="
spray full_name
A Brigade
B Dimilin
C Movento
D Synapse
E Timectin
F Tombstone
", head=T)

nameSprays <- left_join(InsectSprays, spray_names)
```

```
## Joining by: "spray"
```

```
nameSprays
```

```
##    count spray full_name
## 1     10     A   Brigade
## 2      7     A   Brigade
## 3     20     A   Brigade
## 4     14     A   Brigade
## 5     14     A   Brigade
## 6     12     A   Brigade
## 7     10     A   Brigade
## 8     23     A   Brigade
## 9     17     A   Brigade
## 10    20     A   Brigade
## 11    14     A   Brigade
## 12    13     A   Brigade
## 13    11     B   Dimilin
## 14    17     B   Dimilin
## 15    21     B   Dimilin
## 16    11     B   Dimilin
## 17    16     B   Dimilin
## 18    14     B   Dimilin
## 19    17     B   Dimilin
## 20    17     B   Dimilin
## 21    19     B   Dimilin
## 22    21     B   Dimilin
## 23     7     B   Dimilin
## 24    13     B   Dimilin
## 25     0     C   Movento
## 26     1     C   Movento
## 27     7     C   Movento
## 28     2     C   Movento
## 29     3     C   Movento
## 30     1     C   Movento
## 31     2     C   Movento
## 32     1     C   Movento
## 33     3     C   Movento
## 34     0     C   Movento
## 35     1     C   Movento
## 36     4     C   Movento
## 37     3     D   Synapse
## 38     5     D   Synapse
## 39    12     D   Synapse
## 40     6     D   Synapse
## 41     4     D   Synapse
## 42     3     D   Synapse
## 43     5     D   Synapse
## 44     5     D   Synapse
## 45     5     D   Synapse
## 46     5     D   Synapse
```

```
## 47     2     D   Synapse
## 48     4     D   Synapse
## 49     3     E  Timectin
## 50     5     E  Timectin
## 51     3     E  Timectin
## 52     5     E  Timectin
## 53     3     E  Timectin
## 54     6     E  Timectin
## 55     1     E  Timectin
## 56     1     E  Timectin
## 57     3     E  Timectin
## 58     2     E  Timectin
## 59     6     E  Timectin
## 60     4     E  Timectin
## 61    11     F Tombstone
## 62     9     F Tombstone
## 63    15     F Tombstone
## 64    22     F Tombstone
## 65    15     F Tombstone
## 66    16     F Tombstone
## 67    13     F Tombstone
## 68    10     F Tombstone
## 69    26     F Tombstone
## 70    26     F Tombstone
## 71    24     F Tombstone
## 72    13     F Tombstone
```

left_join() is the approriate function because it combines two tables duplicating values from the second table where necessary.

Question 3: (3 pts) *Make up your own data set which is **not** tidy. First, explain why it is not tidy. Then, using* `dplyr` *and/or* `tidyr`*, convert it into a tidy data set.*

```
messyData <- data.frame(
  name = c("Evan", "Adam", "Doug"),
  druga = c(74, 80, 73),
  drugb = c(80, 90, 82)
)
messyData
```

```
##   name druga drugb
## 1 Evan    74    80
## 2 Adam    80    90
## 3 Doug    73    82
```

The data frame is messy because we have 3 variables(name,drug,and heartreate), but only the variable name is in a correct tidy column.

```
messyData %>%
  gather(drug, heartrate, druga:drugb)
```

```
##   name  drug heartrate
## 1 Evan druga        74
## 2 Adam druga        80
## 3 Doug druga        73
## 4 Evan drugb        80
## 5 Adam drugb        90
## 6 Doug drugb        82
```

```
messyData
```

```
##   name druga drugb
## 1 Evan    74    80
## 2 Adam    80    90
## 3 Doug    73    82
```