Tidy Data Article Exercises

Xiru Lyu

2/23/2018

```r
# Read required package
library(tidyverse)
```

1. Table 4 & Table 6

```r
# Load table 4
tab4 <- readRDS("./Data/tab4_raw.rds")

# view first ten rows of table 4
head(tab4,n=10)
```

```
##                      religion <$10k $10-20k $20-30k $30-40k $40-50k $50-75k
## 1                    Agnostic    27      34      60      81      76     137
## 2                     Atheist    12      27      37      52      35      70
## 3                    Buddhist    27      21      30      34      33      58
## 4                    Catholic   418     617     732     670     638    1116
## 5         Don't know/refused     15      14      15      11      10      35
## 6            Evangelical Prot   575     869    1064     982     881    1486
## 7                       Hindu     1       9       7       9      11      34
## 8     Historically Black Prot   228     244     236     238     197     223
## 9           Jehovah's Witness    20      27      24      24      21      30
## 10                     Jewish    19      19      25      25      30      95
##    $75-100k $100-150k >150k Don't know/refused
## 1       122       109    84                 96
## 2        73        59    74                 76
## 3        62        39    53                 54
## 4       949       792   633               1489
## 5        21        17    18                116
## 6       949       723   414               1529
## 7        47        48    54                 37
## 8       131        81    78                339
## 9        15        11     6                 37
## 10       69        87   151                162
```

```r
# convert the data frame to a tibble object
tab4 <- tab4 %>% as.tibble

# tidy the data to be table 6
tab6 <- tab4 %>% gather(key='income',value='freq',-religion) %>%
  arrange(religion)

# view first ten rows of table 6
print(tab6,n=10)
```

```
## # A tibble: 180 x 3
##    religion income            freq
##    <chr>    <chr>            <int>
```

1

```
##  1 Agnostic <$10k                  27
##  2 Agnostic $10-20k                34
##  3 Agnostic $20-30k                60
##  4 Agnostic $30-40k                81
##  5 Agnostic $40-50k                76
##  6 Agnostic $50-75k               137
##  7 Agnostic $75-100k              122
##  8 Agnostic $100-150k             109
##  9 Agnostic >150k                  84
## 10 Agnostic Don't know/refused     96
## # ... with 170 more rows
```

2. Table 7 & Table 8

```r
# Load table 7
tab7 <- readRDS("./Data/tab7_raw.rds")

# view first 8 rows and 8 columns of table 7
print(tab7[1:8,1:8])
```

```
##    year       artist                          track time date.entered wk1
## 1 2000        2 Pac              Baby Don't Cry 4:22   2000-02-26  87
## 2 2000      2Ge+her The Hardest Part Of Breaking Up 3:15   2000-09-02  91
## 3 2000 3 Doors Down                     Kryptonite 3:53   2000-04-08  81
## 4 2000 3 Doors Down                          Loser 4:24   2000-10-21  76
## 5 2000      504 Boyz                  Wobble Wobble 3:35   2000-04-15  57
## 6 2000          98^0      Give Me Just One Night 3:24   2000-08-19  51
## 7 2000      A*Teens                  Dancing Queen 3:44   2000-07-08  97
## 8 2000      Aaliyah                  I Don't Wanna 4:15   2000-01-29  84
##   wk2 wk3
## 1  82  72
## 2  87  92
## 3  70  68
## 4  76  72
## 5  34  25
## 6  39  34
## 7  97  96
## 8  62  51
```

```r
# convert the data frame to a tibble object
tab7 <- as.tibble(tab7)

# tidy the data to be table 8
tab8 <- tab7 %>% gather(week,rank,wk1:wk76,na.rm=TRUE) %>%
  separate(week,into=c('wk','week'),sep=2) %>%
  mutate(date = as.Date(date.entered) + (as.numeric(week)-1)*7) %>%
  select(year,artist,track,time,date,week,rank) %>%
  arrange(artist,track)

# view first 15 rows of table 8
print(tab8,n=15)
```

```
## # A tibble: 5,307 x 7
##     year artist     track                time  date      week   rank
```

```
##    <int> <chr>         <chr>                    <fct> <date>        <chr> <int>
## 1   2000 2 Pac         Baby Don't Cry            4:22 2000-02-26 1          87
## 2   2000 2 Pac         Baby Don't Cry            4:22 2000-03-04 2          82
## 3   2000 2 Pac         Baby Don't Cry            4:22 2000-03-11 3          72
## 4   2000 2 Pac         Baby Don't Cry            4:22 2000-03-18 4          77
## 5   2000 2 Pac         Baby Don't Cry            4:22 2000-03-25 5          87
## 6   2000 2 Pac         Baby Don't Cry            4:22 2000-04-01 6          94
## 7   2000 2 Pac         Baby Don't Cry            4:22 2000-04-08 7          99
## 8   2000 2Ge+her       The Hardest Part Of Br~   3:15 2000-09-02 1          91
## 9   2000 2Ge+her       The Hardest Part Of Br~   3:15 2000-09-09 2          87
## 10  2000 2Ge+her       The Hardest Part Of Br~   3:15 2000-09-16 3          92
## 11  2000 3 Doors Down  Kryptonite                3:53 2000-04-08 1          81
## 12  2000 3 Doors Down  Kryptonite                3:53 2000-04-15 2          70
## 13  2000 3 Doors Down  Kryptonite                3:53 2000-04-22 3          68
## 14  2000 3 Doors Down  Kryptonite                3:53 2000-04-29 4          67
## 15  2000 3 Doors Down  Kryptonite                3:53 2000-05-06 5          66
## # ... with 5,292 more rows
```