Assignment4

yaru peng

4/13/2018

```
##1)
###12.6.1
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 2.2.1     v purrr   0.2.4
## v tibble  1.4.2     v dplyr   0.7.4
## v tidyr   0.8.0     v stringr 1.3.0
## v readr   1.1.1     v forcats 0.3.0
```

```
## -- Conflicts ------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
tidyr::who
```

```
## # A tibble: 7,240 x 60
##    country     iso2  iso3   year new_sp_m014 new_sp_m1524 new_sp_m2534
##    <chr>       <chr> <chr> <int>       <int>        <int>        <int>
##  1 Afghanistan AF    AFG    1980          NA           NA           NA
##  2 Afghanistan AF    AFG    1981          NA           NA           NA
##  3 Afghanistan AF    AFG    1982          NA           NA           NA
##  4 Afghanistan AF    AFG    1983          NA           NA           NA
##  5 Afghanistan AF    AFG    1984          NA           NA           NA
##  6 Afghanistan AF    AFG    1985          NA           NA           NA
##  7 Afghanistan AF    AFG    1986          NA           NA           NA
##  8 Afghanistan AF    AFG    1987          NA           NA           NA
##  9 Afghanistan AF    AFG    1988          NA           NA           NA
## 10 Afghanistan AF    AFG    1989          NA           NA           NA
## # ... with 7,230 more rows, and 53 more variables: new_sp_m3544 <int>,
## #   new_sp_m4554 <int>, new_sp_m5564 <int>, new_sp_m65 <int>,
## #   new_sp_f014 <int>, new_sp_f1524 <int>, new_sp_f2534 <int>,
## #   new_sp_f3544 <int>, new_sp_f4554 <int>, new_sp_f5564 <int>,
## #   new_sp_f65 <int>, new_sn_m014 <int>, new_sn_m1524 <int>,
## #   new_sn_m2534 <int>, new_sn_m3544 <int>, new_sn_m4554 <int>,
## #   new_sn_m5564 <int>, new_sn_m65 <int>, new_sn_f014 <int>,
## #   new_sn_f1524 <int>, new_sn_f2534 <int>, new_sn_f3544 <int>,
## #   new_sn_f4554 <int>, new_sn_f5564 <int>, new_sn_f65 <int>,
## #   new_ep_m014 <int>, new_ep_m1524 <int>, new_ep_m2534 <int>,
## #   new_ep_m3544 <int>, new_ep_m4554 <int>, new_ep_m5564 <int>,
## #   new_ep_m65 <int>, new_ep_f014 <int>, new_ep_f1524 <int>,
## #   new_ep_f2534 <int>, new_ep_f3544 <int>, new_ep_f4554 <int>,
## #   new_ep_f5564 <int>, new_ep_f65 <int>, newrel_m014 <int>,
## #   newrel_m1524 <int>, newrel_m2534 <int>, newrel_m3544 <int>,
## #   newrel_m4554 <int>, newrel_m5564 <int>, newrel_m65 <int>,
## #   newrel_f014 <int>, newrel_f1524 <int>, newrel_f2534 <int>,
## #   newrel_f3544 <int>, newrel_f4554 <int>, newrel_f5564 <int>,
## #   newrel_f65 <int>
```

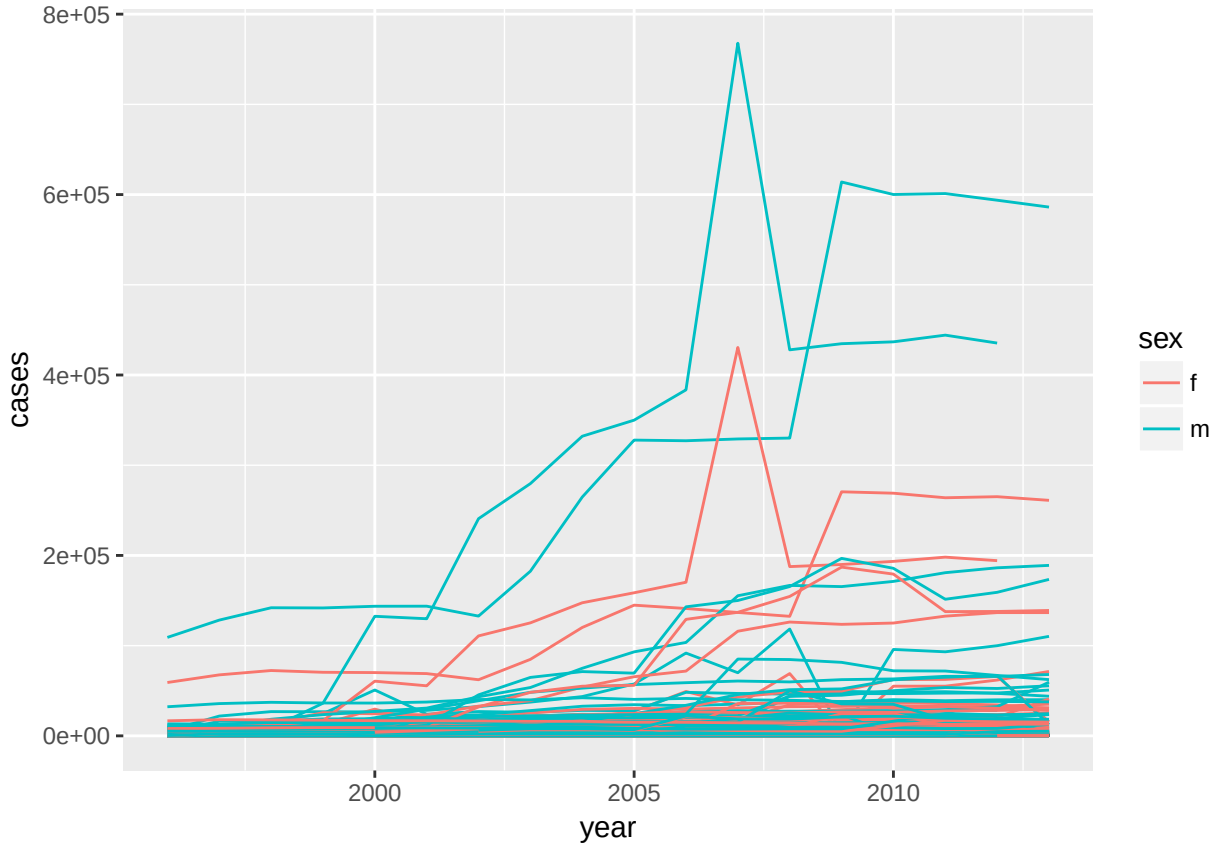
```
who %>%
  gather(code, value, new_sp_m014:newrel_f65, na.rm = TRUE) %>%
  mutate(code = stringr::str_replace(code, "newrel", "new_rel")) %>%
  separate(code, c("new", "var", "sexage")) %>%
  select(-new, -iso2, -iso3) %>%
  separate(sexage, c("sex", "age"), sep = 1)
```

```
## # A tibble: 76,046 x 6
##    country      year var   sex   age   value
##    <chr>       <int> <chr> <chr> <chr> <int>
##  1 Afghanistan  1997 sp    m     014       0
##  2 Afghanistan  1998 sp    m     014      30
##  3 Afghanistan  1999 sp    m     014       8
##  4 Afghanistan  2000 sp    m     014      52
##  5 Afghanistan  2001 sp    m     014     129
##  6 Afghanistan  2002 sp    m     014      90
##  7 Afghanistan  2003 sp    m     014     127
##  8 Afghanistan  2004 sp    m     014     139
##  9 Afghanistan  2005 sp    m     014     151
## 10 Afghanistan  2006 sp    m     014     193
## # ... with 76,036 more rows
```

```
##problem 3
select(who,country, -iso2, -iso3) %>%
  distinct() %>%
  group_by(country) %>%
  filter(n()>1)
```

```
## # A tibble: 0 x 1
## # Groups:   country [0]
## # ... with 1 variable: country <chr>
```

```
##problem 4
who %>%
  gather(code, value, new_sp_m014:newrel_f65, na.rm = TRUE) %>%
  mutate(code = stringr::str_replace(code,"newrel","new_rel")) %>%
  separate(code,c("new","type","sexage")) %>%
  select(-new, -iso2, -iso3) %>%
  separate(sexage, c("sex","age"), sep=1) %>%
  group_by(country, year, sex) %>%
  filter(year > 1995) %>%
  summarise(cases = sum(value)) %>%
  unite(country_sex, country, sex, remove = FALSE) %>%
  ggplot(aes(x = year, y = cases, group = country_sex,
  colour = sex)) + geom_line()
```



#2) #10.5 #problem 5 Enframe converts named atomic vectors or lists to two-column data frames. For unnamed vectors, the natural sequence is used as name column.

```
#3)
  pew <- as.tibble(read.csv("pew.csv"))
  pew.tidy <-pew %>%
  gather(key="Income",value="Frenquency", -religion)
  pew.tidy
```

```
## # A tibble: 180 x 3
##    religion                Income  Frenquency
##    <fct>                   <chr>        <int>
##  1 Agnostic                X..10K.         27
##  2 Atheist                 X..10K.         12
##  3 Buddhist                X..10K.         27
##  4 Catholic                X..10K.        418
##  5 DonÕt know/refused      X..10K.         15
##  6 Evangelical Prot        X..10K.        575
##  7 Hindu                   X..10K.          1
##  8 Historically Black Prot X..10K.        228
##  9 JehovahÕs Witness       X..10K.         20
## 10 Jewish                  X..10K.         19
## # ... with 170 more rows
```

```
##4)
library(tidyverse)
  bb <- as.tibble(read.csv("billboard.csv"))
```

```r
bb.tidy <- bb %>%
   gather(key = "week", value = "rank" , -year, -artist.inverted,-track, -time, -genre,
          -date.entered, ... = -date.peaked)%>%
   select(year, artist=artist.inverted, time, track, date=date.entered, week, rank) %>%
   arrange(track)%>%
   filter(!is.na(rank))%>%
   separate(week, into=c("A","B","C"), sep=c(1:2), convert = F)%>%
   select(-A,-C) %>%
   dplyr::rename(week = B) %>%
   arrange(artist,track) %>%
   mutate(date=as.Date(date) + (as.numeric(week)-1)*7) %>%
   mutate(rank = as.integer(rank))
bb.tidy
```

```
## # A tibble: 5,307 x 7
##     year artist  time  track                         date       week   rank
##    <int> <fct>   <fct> <fct>                         <date>     <chr> <int>
##  1  2000 2 Pac   4:22  Baby Don't Cry (Keep Ya Hea~ 2000-02-26 1        87
##  2  2000 2 Pac   4:22  Baby Don't Cry (Keep Ya Hea~ 2000-03-04 2        82
##  3  2000 2 Pac   4:22  Baby Don't Cry (Keep Ya Hea~ 2000-03-11 3        72
##  4  2000 2 Pac   4:22  Baby Don't Cry (Keep Ya Hea~ 2000-03-18 4        77
##  5  2000 2 Pac   4:22  Baby Don't Cry (Keep Ya Hea~ 2000-03-25 5        87
##  6  2000 2 Pac   4:22  Baby Don't Cry (Keep Ya Hea~ 2000-04-01 6        94
##  7  2000 2 Pac   4:22  Baby Don't Cry (Keep Ya Hea~ 2000-04-08 7        99
##  8  2000 2Ge+her 3:15  The Hardest Part Of Breakin~ 2000-09-02 1        91
##  9  2000 2Ge+her 3:15  The Hardest Part Of Breakin~ 2000-09-09 2        87
## 10  2000 2Ge+her 3:15  The Hardest Part Of Breakin~ 2000-09-16 3        92
## # ... with 5,297 more rows
```