assignment4

Chen Hong

1) 12.6.3

```
who1 <- who %>% gather(new_sp_m014:newrel_f65,key="key",value ="case", na.rm = T)
who2 <- who1 %>% mutate(key=stringr::str_replace(key,"newrel","new_rel"))
who3 <- who2 %>%
separate(key,c("new","type","sexage"),sep = "_")
who3 %>% count(new)
```

```
## # A tibble: 1 x 2
##   new       n
##   <chr> <int>
## 1 new   76046
```

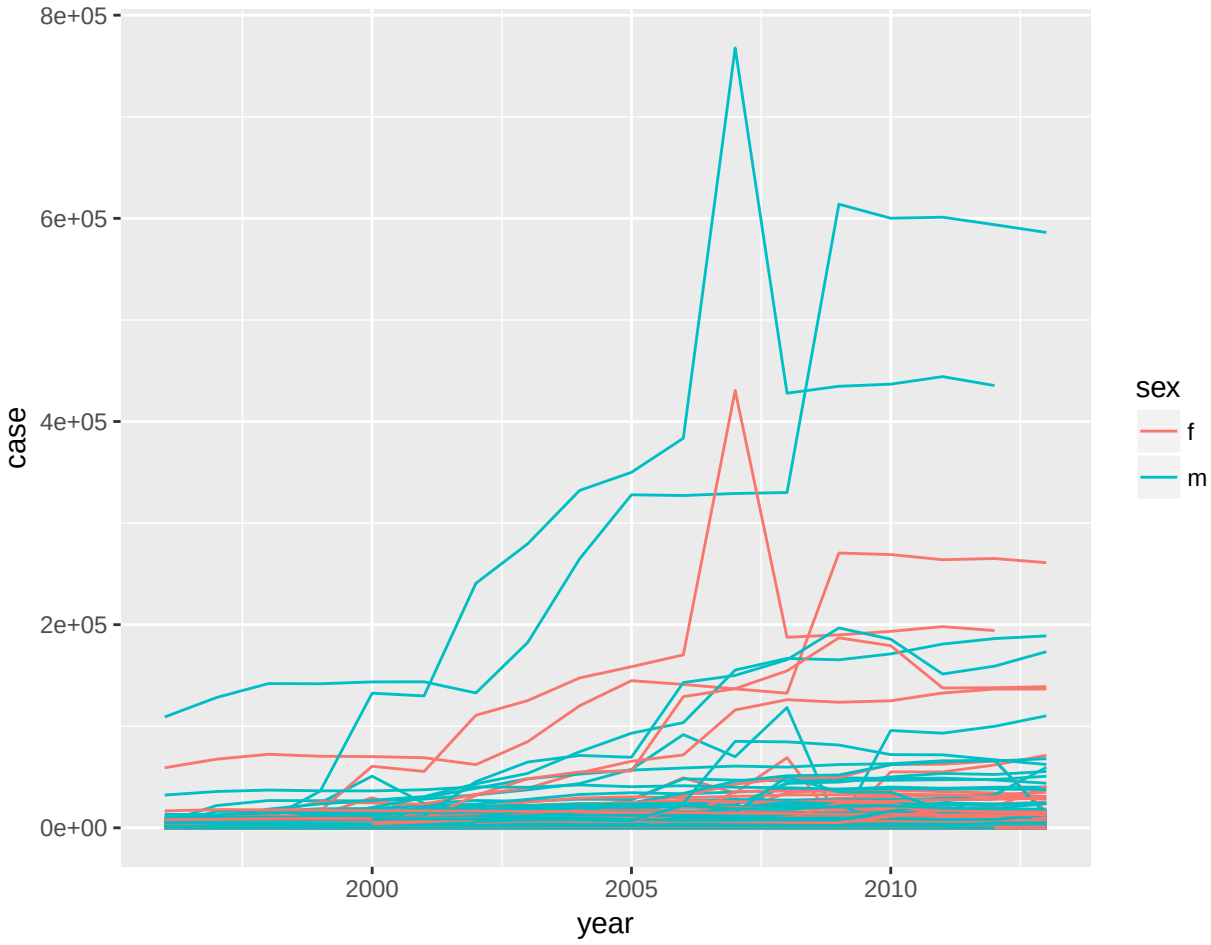
```
who4 <- who3 %>% select(-new,-iso2,-iso3)
who5 <- who4 %>% separate(sexage,c("sex","age"),sep=1)
```

```
select(who3,country,iso2,iso3) %>%
distinct() %>%
group_by(country) %>%
filter(n() >1)
```

```
## # A tibble: 0 x 3
## # Groups:   country [0]
## # ... with 3 variables: country <chr>, iso2 <chr>, iso3 <chr>
```

12.6.1 4)

```
who5 %>% group_by(country,year,sex) %>%
filter(year>1995) %>%
summarise(case=sum(case)) %>%
unite(country_sex,country,sex,remove=FALSE) %>%
ggplot(aes(x = year,y = case, group = country_sex,colour = sex)) + geom_line()
```



2)

enframe()converts names vector or lists to two conlums data format. here is an example.we use it when we want to convert the data into two colums fata set.

```
x <- c(1:10)
enframe(x)
```

```
## # A tibble: 10 x 2
##     name value
##    <int> <int>
##  1     1     1
##  2     2     2
##  3     3     3
##  4     4     4
##  5     5     5
##  6     6     6
##  7     7     7
##  8     8     8
##  9     9     9
## 10    10    10
```

3)

```
table4 <- readRDS("raw4.RDS")
data4 <- as.tibble(table4)
data41 <- data4 %>% gather(key = "income",value = "freq",-religion)
data42 <- data41 %>% arrange(religion)
```

```
data42
```

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

4)

```
bill <- read.csv("billboard.csv")
bill <- as.tibble(bill)
bill1 <- bill %>% gather(key="week",value = "rank",-year,-artist.inverted,-track,-time,-genre,-date.ente
bill2 <- bill1 %>% select(year,artist=artist.inverted,time,track,date=date.entered,week,rank)
bill3 <- bill2 %>% arrange(artist)
bill4 <- bill3 %>% filter(!is.na(rank))
bill5 <- bill4 %>% separate(week,into=c("A","B","C"),sep=c(1,-7),convert = T)
bill6 <- bill5 %>% select(-C,-A) %>% rename(week=B)
bill7 <- bill6 %>% mutate(date=as.Date(date)+(week-1)*7)

bill7
```

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