Course 1 Section 2.17 - Filter Data

Jiaying Wu

14/09/2020

```r
library(tidyverse)
library(here)
tb_long <- read_rds(here("data","tb_long.rds"))
tb_long
```

```
## # A tibble: 157,820 x 7
##    country     iso3  year type   sex   age_group count
##    <chr>       <chr> <dbl> <chr>  <chr> <chr>     <dbl>
##  1 Afghanistan AFG    1980 new_sp m     04           NA
##  2 Afghanistan AFG    1981 new_sp m     04           NA
##  3 Afghanistan AFG    1982 new_sp m     04           NA
##  4 Afghanistan AFG    1983 new_sp m     04           NA
##  5 Afghanistan AFG    1984 new_sp m     04           NA
##  6 Afghanistan AFG    1985 new_sp m     04           NA
##  7 Afghanistan AFG    1986 new_sp m     04           NA
##  8 Afghanistan AFG    1987 new_sp m     04           NA
##  9 Afghanistan AFG    1988 new_sp m     04           NA
## 10 Afghanistan AFG    1989 new_sp m     04           NA
## # ... with 157,810 more rows
```

Use the verb filter to select TB cases only from:

1.the year 2012

```r
filter(tb_long, year == "2012")
```

```
## # A tibble: 4,340 x 7
##    country             iso3   year type   sex   age_group count
##    <chr>               <chr> <dbl> <chr>  <chr> <chr>     <dbl>
##  1 Afghanistan         AFG    2012 new_sp m     04            0
##  2 Albania             ALB    2012 new_sp m     04            0
##  3 Algeria             DZA    2012 new_sp m     04           NA
##  4 American Samoa      ASM    2012 new_sp m     04           NA
##  5 Andorra             AND    2012 new_sp m     04            0
##  6 Angola              AGO    2012 new_sp m     04           58
##  7 Anguilla            AIA    2012 new_sp m     04            0
##  8 Antigua and Barbuda ATG    2012 new_sp m     04            0
##  9 Argentina           ARG    2012 new_sp m     04            8
## 10 Armenia             ARM    2012 new_sp m     04            0
## # ... with 4,330 more rows
```

2.the years 2013 - 2018

```r
filter(tb_long, year >= "2013", year <= "2018")
```

```
## # A tibble: 17,360 x 7
##    country     iso3  year type   sex   age_group count
##    <chr>       <chr> <dbl> <chr> <chr> <chr>     <dbl>
##  1 Afghanistan AFG    2013 new_sp m    04           NA
##  2 Afghanistan AFG    2014 new_sp m    04           NA
##  3 Afghanistan AFG    2015 new_sp m    04           NA
##  4 Afghanistan AFG    2016 new_sp m    04           NA
##  5 Albania     ALB    2013 new_sp m    04           NA
##  6 Albania     ALB    2014 new_sp m    04           NA
##  7 Albania     ALB    2015 new_sp m    04           NA
##  8 Albania     ALB    2016 new_sp m    04           NA
##  9 Algeria     DZA    2013 new_sp m    04           NA
## 10 Algeria     DZA    2014 new_sp m    04           NA
## # ... with 17,350 more rows
```

3.countries in North America

```r
tb_north <- filter(tb_long, iso3 %in% c("AIA", "ATG", "ABW", "BHS", "BRB", "BLZ", "BMU", "BES", "VGB",

tb_north
```

```
## # A tibble: 24,700 x 7
##    country  iso3   year type   sex   age_group count
##    <chr>    <chr> <dbl> <chr> <chr> <chr>     <dbl>
##  1 Anguilla AIA    1980 new_sp m    04           NA
##  2 Anguilla AIA    1981 new_sp m    04           NA
##  3 Anguilla AIA    1982 new_sp m    04           NA
##  4 Anguilla AIA    1983 new_sp m    04           NA
##  5 Anguilla AIA    1984 new_sp m    04           NA
##  6 Anguilla AIA    1985 new_sp m    04           NA
##  7 Anguilla AIA    1986 new_sp m    04           NA
##  8 Anguilla AIA    1987 new_sp m    04           NA
##  9 Anguilla AIA    1988 new_sp m    04           NA
## 10 Anguilla AIA    1989 new_sp m    04           NA
## # ... with 24,690 more rows
```

4.countries in North America but not ages 0-4, 0-14 or u.

```r
unique(factor(tb_north$age_group))
```

```
##  [1] 04   514  014  1524 2534 3544 4554 5564 65   u
## Levels: 014 04 1524 2534 3544 4554 514 5564 65 u
```

```r
filter(tb_north, !age_group %in% c("04", "014", "u"))
```

```
## # A tibble: 17,290 x 7
##    country  iso3   year type   sex   age_group count
##    <chr>    <chr> <dbl> <chr>  <chr> <chr>     <dbl>
##  1 Anguilla AIA    1980 new_sp m     514          NA
##  2 Anguilla AIA    1981 new_sp m     514          NA
##  3 Anguilla AIA    1982 new_sp m     514          NA
##  4 Anguilla AIA    1983 new_sp m     514          NA
##  5 Anguilla AIA    1984 new_sp m     514          NA
##  6 Anguilla AIA    1985 new_sp m     514          NA
##  7 Anguilla AIA    1986 new_sp m     514          NA
##  8 Anguilla AIA    1987 new_sp m     514          NA
##  9 Anguilla AIA    1988 new_sp m     514          NA
## 10 Anguilla AIA    1989 new_sp m     514          NA
## # ... with 17,280 more rows
```