3.2-Basic Data Wrangling

Igor Luciano de Paula

February 15, 2018

In [ ]:
```
> murders <- mutate(murders, rate = total/population * 100000)
> names(murders)
[1] "state"      "abb"        "region"     "population" "total"
[6] "rate"
> head(mmurders)
Error: object 'mmurders' not found
> head(murders)
       state abb region population total     rate
1    Alabama  AL  South    4779736   135 2.824424
2     Alaska  AK   West     710231    19 2.675186
3    Arizona  AZ   West    6392017   232 3.629527
4   Arkansas  AR  South    2915918    93 3.189390
5 California  CA   West   37253956  1257 3.374138
6   Colorado  CO   West    5029196    65 1.292453
> filter(murders, rate <= 0.71)
          state abb        region population total      rate
1        Hawaii  HI          West    1360301     7 0.5145920
2          Iowa  IA North Central    3046355    21 0.6893484
3 New Hampshire  NH     Northeast    1316470     5 0.3798036
4  North Dakota  ND North Central     672591     4 0.5947151
5       Vermont  VT     Northeast     625741     2 0.3196211
> new_table <- select(murders, states, region, rate)
Error: object 'states' not found
> new_table <- select(murders, state, region, rate)
> new_table
                   state        region       rate
1                Alabama         South  2.8244238
2                 Alaska          West  2.6751860
3                Arizona          West  3.6295273
4               Arkansas         South  3.1893901
5             California          West  3.3741383
6               Colorado          West  1.2924531
7            Connecticut     Northeast  2.7139722
8               Delaware         South  4.2319369
9  District of Columbia         South 16.4527532
10                Florida         South  3.3980688
11                Georgia         South  3.7903226
```

```
12                Hawaii          West  0.5145920
13                 Idaho          West  0.7655102
14              Illinois North Central  2.8369608
15               Indiana North Central  2.1900730
16                  Iowa North Central  0.6893484
17                Kansas North Central  2.2081106
18              Kentucky         South  2.6732010
19             Louisiana         South  7.7425810
20                 Maine     Northeast  0.8280881
21              Maryland         South  5.0748655
22         Massachusetts     Northeast  1.8021791
23              Michigan North Central  4.1786225
24             Minnesota North Central  0.9992600
25           Mississippi         South  4.0440846
26              Missouri North Central  5.3598917
27               Montana          West  1.2128379
28              Nebraska North Central  1.7521372
29                Nevada          West  3.1104763
30         New Hampshire     Northeast  0.3798036
31            New Jersey     Northeast  2.7980319
32            New Mexico          West  3.2537239
33              New York     Northeast  2.6679599
34        North Carolina         South  2.9993237
35          North Dakota North Central  0.5947151
36                  Ohio North Central  2.6871225
37              Oklahoma         South  2.9589340
38                Oregon          West  0.9396843
39          Pennsylvania     Northeast  3.5977513
40          Rhode Island     Northeast  1.5200933
41        South Carolina         South  4.4753235
42          South Dakota North Central  0.9825837
43             Tennessee         South  3.4509357
44                 Texas         South  3.2013603
45                  Utah          West  0.7959810
46               Vermont     Northeast  0.3196211
47              Virginia         South  3.1246001
48            Washington          West  1.3829942
49         West Virginia         South  1.4571013
50             Wisconsin North Central  1.7056487
51               Wyoming          West  0.8871131
> str(new_table)
'data.frame':        51 obs. of  3 variables:
 $ state : chr  "Alabama" "Alaska" "Arizona" "Arkansas" ...
 $ region: Factor w/ 4 levels "Northeast","South",..: 2 4 4 2 4 4 1 2 2 2 ...
 $ rate  : num  2.82 2.68 3.63 3.19 3.37 ...
> filter(new_table, rate <= 0.71)
          state        region      rate
```

```
1        Hawaii          West 0.5145920
2          Iowa North Central 0.6893484
3 New Hampshire     Northeast 0.3798036
4  North Dakota North Central 0.5947151
5       Vermont     Northeast 0.3196211
> murders %>% select (state, region, rate) %>% filter(rate <= 0.71)
         state        region      rate
1        Hawaii          West 0.5145920
2          Iowa North Central 0.6893484
3 New Hampshire     Northeast 0.3798036
4  North Dakota North Central 0.5947151
5       Vermont     Northeast 0.3196211
>
> grades <- data.frame(names = c("John", "Juan", "Jean"), exam_1 = c(95, 80, 70), exam_2
> class(grades$names)
[1] "character"
> grades
  names exam_1 exam_2
1  John     95     90
2  Juan     80     85
3  Jean     70     90
>
```