Homework 9

Christopher Brunswick

11/23/2021

1. Load the dplyr package and the murders dataset.

library(dplyr)

## Warning: package 'dplyr' was built under R version 4.0.5

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(dslabs)

## Warning: package 'dslabs' was built under R version 4.0.5

data(murders)

# You can add columns using the dplyr function mutate. This function is aware of the column names and inside the function you can call them unquoted:

murders <- mutate(murders, population\_in\_millions = population / 10^6)  
murders

## state abb region population total  
## 1 Alabama AL South 4779736 135  
## 2 Alaska AK West 710231 19  
## 3 Arizona AZ West 6392017 232  
## 4 Arkansas AR South 2915918 93  
## 5 California CA West 37253956 1257  
## 6 Colorado CO West 5029196 65  
## 7 Connecticut CT Northeast 3574097 97  
## 8 Delaware DE South 897934 38  
## 9 District of Columbia DC South 601723 99  
## 10 Florida FL South 19687653 669  
## 11 Georgia GA South 9920000 376  
## 12 Hawaii HI West 1360301 7  
## 13 Idaho ID West 1567582 12  
## 14 Illinois IL North Central 12830632 364  
## 15 Indiana IN North Central 6483802 142  
## 16 Iowa IA North Central 3046355 21  
## 17 Kansas KS North Central 2853118 63  
## 18 Kentucky KY South 4339367 116  
## 19 Louisiana LA South 4533372 351  
## 20 Maine ME Northeast 1328361 11  
## 21 Maryland MD South 5773552 293  
## 22 Massachusetts MA Northeast 6547629 118  
## 23 Michigan MI North Central 9883640 413  
## 24 Minnesota MN North Central 5303925 53  
## 25 Mississippi MS South 2967297 120  
## 26 Missouri MO North Central 5988927 321  
## 27 Montana MT West 989415 12  
## 28 Nebraska NE North Central 1826341 32  
## 29 Nevada NV West 2700551 84  
## 30 New Hampshire NH Northeast 1316470 5  
## 31 New Jersey NJ Northeast 8791894 246  
## 32 New Mexico NM West 2059179 67  
## 33 New York NY Northeast 19378102 517  
## 34 North Carolina NC South 9535483 286  
## 35 North Dakota ND North Central 672591 4  
## 36 Ohio OH North Central 11536504 310  
## 37 Oklahoma OK South 3751351 111  
## 38 Oregon OR West 3831074 36  
## 39 Pennsylvania PA Northeast 12702379 457  
## 40 Rhode Island RI Northeast 1052567 16  
## 41 South Carolina SC South 4625364 207  
## 42 South Dakota SD North Central 814180 8  
## 43 Tennessee TN South 6346105 219  
## 44 Texas TX South 25145561 805  
## 45 Utah UT West 2763885 22  
## 46 Vermont VT Northeast 625741 2  
## 47 Virginia VA South 8001024 250  
## 48 Washington WA West 6724540 93  
## 49 West Virginia WV South 1852994 27  
## 50 Wisconsin WI North Central 5686986 97  
## 51 Wyoming WY West 563626 5  
## population\_in\_millions  
## 1 4.779736  
## 2 0.710231  
## 3 6.392017  
## 4 2.915918  
## 5 37.253956  
## 6 5.029196  
## 7 3.574097  
## 8 0.897934  
## 9 0.601723  
## 10 19.687653  
## 11 9.920000  
## 12 1.360301  
## 13 1.567582  
## 14 12.830632  
## 15 6.483802  
## 16 3.046355  
## 17 2.853118  
## 18 4.339367  
## 19 4.533372  
## 20 1.328361  
## 21 5.773552  
## 22 6.547629  
## 23 9.883640  
## 24 5.303925  
## 25 2.967297  
## 26 5.988927  
## 27 0.989415  
## 28 1.826341  
## 29 2.700551  
## 30 1.316470  
## 31 8.791894  
## 32 2.059179  
## 33 19.378102  
## 34 9.535483  
## 35 0.672591  
## 36 11.536504  
## 37 3.751351  
## 38 3.831074  
## 39 12.702379  
## 40 1.052567  
## 41 4.625364  
## 42 0.814180  
## 43 6.346105  
## 44 25.145561  
## 45 2.763885  
## 46 0.625741  
## 47 8.001024  
## 48 6.724540  
## 49 1.852994  
## 50 5.686986  
## 51 0.563626

# We can write population rather than murders$population.

#The function mutate knows we are grabbing columns from murders.

murders <- mutate(murders, rate = total/population \* 10^6 , population\_in\_millions = population / 10^6)  
murders

## state abb region population total  
## 1 Alabama AL South 4779736 135  
## 2 Alaska AK West 710231 19  
## 3 Arizona AZ West 6392017 232  
## 4 Arkansas AR South 2915918 93  
## 5 California CA West 37253956 1257  
## 6 Colorado CO West 5029196 65  
## 7 Connecticut CT Northeast 3574097 97  
## 8 Delaware DE South 897934 38  
## 9 District of Columbia DC South 601723 99  
## 10 Florida FL South 19687653 669  
## 11 Georgia GA South 9920000 376  
## 12 Hawaii HI West 1360301 7  
## 13 Idaho ID West 1567582 12  
## 14 Illinois IL North Central 12830632 364  
## 15 Indiana IN North Central 6483802 142  
## 16 Iowa IA North Central 3046355 21  
## 17 Kansas KS North Central 2853118 63  
## 18 Kentucky KY South 4339367 116  
## 19 Louisiana LA South 4533372 351  
## 20 Maine ME Northeast 1328361 11  
## 21 Maryland MD South 5773552 293  
## 22 Massachusetts MA Northeast 6547629 118  
## 23 Michigan MI North Central 9883640 413  
## 24 Minnesota MN North Central 5303925 53  
## 25 Mississippi MS South 2967297 120  
## 26 Missouri MO North Central 5988927 321  
## 27 Montana MT West 989415 12  
## 28 Nebraska NE North Central 1826341 32  
## 29 Nevada NV West 2700551 84  
## 30 New Hampshire NH Northeast 1316470 5  
## 31 New Jersey NJ Northeast 8791894 246  
## 32 New Mexico NM West 2059179 67  
## 33 New York NY Northeast 19378102 517  
## 34 North Carolina NC South 9535483 286  
## 35 North Dakota ND North Central 672591 4  
## 36 Ohio OH North Central 11536504 310  
## 37 Oklahoma OK South 3751351 111  
## 38 Oregon OR West 3831074 36  
## 39 Pennsylvania PA Northeast 12702379 457  
## 40 Rhode Island RI Northeast 1052567 16  
## 41 South Carolina SC South 4625364 207  
## 42 South Dakota SD North Central 814180 8  
## 43 Tennessee TN South 6346105 219  
## 44 Texas TX South 25145561 805  
## 45 Utah UT West 2763885 22  
## 46 Vermont VT Northeast 625741 2  
## 47 Virginia VA South 8001024 250  
## 48 Washington WA West 6724540 93  
## 49 West Virginia WV South 1852994 27  
## 50 Wisconsin WI North Central 5686986 97  
## 51 Wyoming WY West 563626 5  
## population\_in\_millions rate  
## 1 4.779736 28.244238  
## 2 0.710231 26.751860  
## 3 6.392017 36.295273  
## 4 2.915918 31.893901  
## 5 37.253956 33.741383  
## 6 5.029196 12.924531  
## 7 3.574097 27.139722  
## 8 0.897934 42.319369  
## 9 0.601723 164.527532  
## 10 19.687653 33.980688  
## 11 9.920000 37.903226  
## 12 1.360301 5.145920  
## 13 1.567582 7.655102  
## 14 12.830632 28.369608  
## 15 6.483802 21.900730  
## 16 3.046355 6.893484  
## 17 2.853118 22.081106  
## 18 4.339367 26.732010  
## 19 4.533372 77.425810  
## 20 1.328361 8.280881  
## 21 5.773552 50.748655  
## 22 6.547629 18.021791  
## 23 9.883640 41.786225  
## 24 5.303925 9.992600  
## 25 2.967297 40.440846  
## 26 5.988927 53.598917  
## 27 0.989415 12.128379  
## 28 1.826341 17.521372  
## 29 2.700551 31.104763  
## 30 1.316470 3.798036  
## 31 8.791894 27.980319  
## 32 2.059179 32.537239  
## 33 19.378102 26.679599  
## 34 9.535483 29.993237  
## 35 0.672591 5.947151  
## 36 11.536504 26.871225  
## 37 3.751351 29.589340  
## 38 3.831074 9.396843  
## 39 12.702379 35.977513  
## 40 1.052567 15.200933  
## 41 4.625364 44.753235  
## 42 0.814180 9.825837  
## 43 6.346105 34.509357  
## 44 25.145561 32.013603  
## 45 2.763885 7.959810  
## 46 0.625741 3.196211  
## 47 8.001024 31.246001  
## 48 6.724540 13.829942  
## 49 1.852994 14.571013  
## 50 5.686986 17.056487  
## 51 0.563626 8.871131

# 2. If rank(x) gives you the ranks of x from lowest to highest, rank(-x) gives you the ranks from highest to lowest. Use the function mutate to add a column rank containing the rank, from highest to lowest murder rate. Make sure you redefine murders so we can keep using this variable.

murders <- mutate(murders, rate = total / population \* 100000, rank = rank(-rate) , population\_in\_millions = population / 10^6)  
murders

## state abb region population total  
## 1 Alabama AL South 4779736 135  
## 2 Alaska AK West 710231 19  
## 3 Arizona AZ West 6392017 232  
## 4 Arkansas AR South 2915918 93  
## 5 California CA West 37253956 1257  
## 6 Colorado CO West 5029196 65  
## 7 Connecticut CT Northeast 3574097 97  
## 8 Delaware DE South 897934 38  
## 9 District of Columbia DC South 601723 99  
## 10 Florida FL South 19687653 669  
## 11 Georgia GA South 9920000 376  
## 12 Hawaii HI West 1360301 7  
## 13 Idaho ID West 1567582 12  
## 14 Illinois IL North Central 12830632 364  
## 15 Indiana IN North Central 6483802 142  
## 16 Iowa IA North Central 3046355 21  
## 17 Kansas KS North Central 2853118 63  
## 18 Kentucky KY South 4339367 116  
## 19 Louisiana LA South 4533372 351  
## 20 Maine ME Northeast 1328361 11  
## 21 Maryland MD South 5773552 293  
## 22 Massachusetts MA Northeast 6547629 118  
## 23 Michigan MI North Central 9883640 413  
## 24 Minnesota MN North Central 5303925 53  
## 25 Mississippi MS South 2967297 120  
## 26 Missouri MO North Central 5988927 321  
## 27 Montana MT West 989415 12  
## 28 Nebraska NE North Central 1826341 32  
## 29 Nevada NV West 2700551 84  
## 30 New Hampshire NH Northeast 1316470 5  
## 31 New Jersey NJ Northeast 8791894 246  
## 32 New Mexico NM West 2059179 67  
## 33 New York NY Northeast 19378102 517  
## 34 North Carolina NC South 9535483 286  
## 35 North Dakota ND North Central 672591 4  
## 36 Ohio OH North Central 11536504 310  
## 37 Oklahoma OK South 3751351 111  
## 38 Oregon OR West 3831074 36  
## 39 Pennsylvania PA Northeast 12702379 457  
## 40 Rhode Island RI Northeast 1052567 16  
## 41 South Carolina SC South 4625364 207  
## 42 South Dakota SD North Central 814180 8  
## 43 Tennessee TN South 6346105 219  
## 44 Texas TX South 25145561 805  
## 45 Utah UT West 2763885 22  
## 46 Vermont VT Northeast 625741 2  
## 47 Virginia VA South 8001024 250  
## 48 Washington WA West 6724540 93  
## 49 West Virginia WV South 1852994 27  
## 50 Wisconsin WI North Central 5686986 97  
## 51 Wyoming WY West 563626 5  
## population\_in\_millions rate rank  
## 1 4.779736 2.8244238 23  
## 2 0.710231 2.6751860 27  
## 3 6.392017 3.6295273 10  
## 4 2.915918 3.1893901 17  
## 5 37.253956 3.3741383 14  
## 6 5.029196 1.2924531 38  
## 7 3.574097 2.7139722 25  
## 8 0.897934 4.2319369 6  
## 9 0.601723 16.4527532 1  
## 10 19.687653 3.3980688 13  
## 11 9.920000 3.7903226 9  
## 12 1.360301 0.5145920 49  
## 13 1.567582 0.7655102 46  
## 14 12.830632 2.8369608 22  
## 15 6.483802 2.1900730 31  
## 16 3.046355 0.6893484 47  
## 17 2.853118 2.2081106 30  
## 18 4.339367 2.6732010 28  
## 19 4.533372 7.7425810 2  
## 20 1.328361 0.8280881 44  
## 21 5.773552 5.0748655 4  
## 22 6.547629 1.8021791 32  
## 23 9.883640 4.1786225 7  
## 24 5.303925 0.9992600 40  
## 25 2.967297 4.0440846 8  
## 26 5.988927 5.3598917 3  
## 27 0.989415 1.2128379 39  
## 28 1.826341 1.7521372 33  
## 29 2.700551 3.1104763 19  
## 30 1.316470 0.3798036 50  
## 31 8.791894 2.7980319 24  
## 32 2.059179 3.2537239 15  
## 33 19.378102 2.6679599 29  
## 34 9.535483 2.9993237 20  
## 35 0.672591 0.5947151 48  
## 36 11.536504 2.6871225 26  
## 37 3.751351 2.9589340 21  
## 38 3.831074 0.9396843 42  
## 39 12.702379 3.5977513 11  
## 40 1.052567 1.5200933 35  
## 41 4.625364 4.4753235 5  
## 42 0.814180 0.9825837 41  
## 43 6.346105 3.4509357 12  
## 44 25.145561 3.2013603 16  
## 45 2.763885 0.7959810 45  
## 46 0.625741 0.3196211 51  
## 47 8.001024 3.1246001 18  
## 48 6.724540 1.3829942 37  
## 49 1.852994 1.4571013 36  
## 50 5.686986 1.7056487 34  
## 51 0.563626 0.8871131 43

# 3. With dplyr, we can use select to show only certain columns. For example, with this code we would only show the states and population sizes:

select(murders, state, population) %>% head()

## state population  
## 1 Alabama 4779736  
## 2 Alaska 710231  
## 3 Arizona 6392017  
## 4 Arkansas 2915918  
## 5 California 37253956  
## 6 Colorado 5029196

# Use select to show the state names and abbreviations in murders. Do not redefine murders, just show the results.

select(murders, state, abb) %>% head()

## state abb  
## 1 Alabama AL  
## 2 Alaska AK  
## 3 Arizona AZ  
## 4 Arkansas AR  
## 5 California CA  
## 6 Colorado CO

# 4. The dplyr function filter is used to choose specific rows of the data frame to keep. Unlike select which is for columns, filter is for rows. For example, you can show just the New York row like this:

filter(murders, state == "New York")

## state abb region population total population\_in\_millions rate rank  
## 1 New York NY Northeast 19378102 517 19.3781 2.66796 29

# Use filter to show the top 5 states with the highest murder rates. After we add murder rate and rank, do not change the murders dataset, just show the result. Remember that you can filter based on the rank column.

filter(murders, rank <= 5)

## state abb region population total  
## 1 District of Columbia DC South 601723 99  
## 2 Louisiana LA South 4533372 351  
## 3 Maryland MD South 5773552 293  
## 4 Missouri MO North Central 5988927 321  
## 5 South Carolina SC South 4625364 207  
## population\_in\_millions rate rank  
## 1 0.601723 16.452753 1  
## 2 4.533372 7.742581 2  
## 3 5.773552 5.074866 4  
## 4 5.988927 5.359892 3  
## 5 4.625364 4.475323 5

# 5. We can remove rows using the != operator. For example, to remove Florida, we would do this:

no\_florida <- filter(murders, state != "Florida")  
no\_florida

## state abb region population total  
## 1 Alabama AL South 4779736 135  
## 2 Alaska AK West 710231 19  
## 3 Arizona AZ West 6392017 232  
## 4 Arkansas AR South 2915918 93  
## 5 California CA West 37253956 1257  
## 6 Colorado CO West 5029196 65  
## 7 Connecticut CT Northeast 3574097 97  
## 8 Delaware DE South 897934 38  
## 9 District of Columbia DC South 601723 99  
## 10 Georgia GA South 9920000 376  
## 11 Hawaii HI West 1360301 7  
## 12 Idaho ID West 1567582 12  
## 13 Illinois IL North Central 12830632 364  
## 14 Indiana IN North Central 6483802 142  
## 15 Iowa IA North Central 3046355 21  
## 16 Kansas KS North Central 2853118 63  
## 17 Kentucky KY South 4339367 116  
## 18 Louisiana LA South 4533372 351  
## 19 Maine ME Northeast 1328361 11  
## 20 Maryland MD South 5773552 293  
## 21 Massachusetts MA Northeast 6547629 118  
## 22 Michigan MI North Central 9883640 413  
## 23 Minnesota MN North Central 5303925 53  
## 24 Mississippi MS South 2967297 120  
## 25 Missouri MO North Central 5988927 321  
## 26 Montana MT West 989415 12  
## 27 Nebraska NE North Central 1826341 32  
## 28 Nevada NV West 2700551 84  
## 29 New Hampshire NH Northeast 1316470 5  
## 30 New Jersey NJ Northeast 8791894 246  
## 31 New Mexico NM West 2059179 67  
## 32 New York NY Northeast 19378102 517  
## 33 North Carolina NC South 9535483 286  
## 34 North Dakota ND North Central 672591 4  
## 35 Ohio OH North Central 11536504 310  
## 36 Oklahoma OK South 3751351 111  
## 37 Oregon OR West 3831074 36  
## 38 Pennsylvania PA Northeast 12702379 457  
## 39 Rhode Island RI Northeast 1052567 16  
## 40 South Carolina SC South 4625364 207  
## 41 South Dakota SD North Central 814180 8  
## 42 Tennessee TN South 6346105 219  
## 43 Texas TX South 25145561 805  
## 44 Utah UT West 2763885 22  
## 45 Vermont VT Northeast 625741 2  
## 46 Virginia VA South 8001024 250  
## 47 Washington WA West 6724540 93  
## 48 West Virginia WV South 1852994 27  
## 49 Wisconsin WI North Central 5686986 97  
## 50 Wyoming WY West 563626 5  
## population\_in\_millions rate rank  
## 1 4.779736 2.8244238 23  
## 2 0.710231 2.6751860 27  
## 3 6.392017 3.6295273 10  
## 4 2.915918 3.1893901 17  
## 5 37.253956 3.3741383 14  
## 6 5.029196 1.2924531 38  
## 7 3.574097 2.7139722 25  
## 8 0.897934 4.2319369 6  
## 9 0.601723 16.4527532 1  
## 10 9.920000 3.7903226 9  
## 11 1.360301 0.5145920 49  
## 12 1.567582 0.7655102 46  
## 13 12.830632 2.8369608 22  
## 14 6.483802 2.1900730 31  
## 15 3.046355 0.6893484 47  
## 16 2.853118 2.2081106 30  
## 17 4.339367 2.6732010 28  
## 18 4.533372 7.7425810 2  
## 19 1.328361 0.8280881 44  
## 20 5.773552 5.0748655 4  
## 21 6.547629 1.8021791 32  
## 22 9.883640 4.1786225 7  
## 23 5.303925 0.9992600 40  
## 24 2.967297 4.0440846 8  
## 25 5.988927 5.3598917 3  
## 26 0.989415 1.2128379 39  
## 27 1.826341 1.7521372 33  
## 28 2.700551 3.1104763 19  
## 29 1.316470 0.3798036 50  
## 30 8.791894 2.7980319 24  
## 31 2.059179 3.2537239 15  
## 32 19.378102 2.6679599 29  
## 33 9.535483 2.9993237 20  
## 34 0.672591 0.5947151 48  
## 35 11.536504 2.6871225 26  
## 36 3.751351 2.9589340 21  
## 37 3.831074 0.9396843 42  
## 38 12.702379 3.5977513 11  
## 39 1.052567 1.5200933 35  
## 40 4.625364 4.4753235 5  
## 41 0.814180 0.9825837 41  
## 42 6.346105 3.4509357 12  
## 43 25.145561 3.2013603 16  
## 44 2.763885 0.7959810 45  
## 45 0.625741 0.3196211 51  
## 46 8.001024 3.1246001 18  
## 47 6.724540 1.3829942 37  
## 48 1.852994 1.4571013 36  
## 49 5.686986 1.7056487 34  
## 50 0.563626 0.8871131 43

# Create a new data frame called no\_south that removes states from the South region. How many states are in this category? You can use the function nrow for this.

my\_table <- select(murders, state, region)  
no\_south <- filter(my\_table, region != "South")  
nrow(no\_south)

## [1] 34

# 6. We can also use %in% to filter with dplyr. You can therefore see the data from New York and Texas like this:

filter(murders, state %in% c("New York", "Texas"))

## state abb region population total population\_in\_millions rate rank  
## 1 New York NY Northeast 19378102 517 19.37810 2.66796 29  
## 2 Texas TX South 25145561 805 25.14556 3.20136 16

# Create a new data frame called murders\_nw with only the states from the Northeast and the West. How many states are in this category?

murders\_nw <- filter(my\_table, region %in% c("Northeast", "West"))  
murders\_nw

## state region  
## 1 Alaska West  
## 2 Arizona West  
## 3 California West  
## 4 Colorado West  
## 5 Connecticut Northeast  
## 6 Hawaii West  
## 7 Idaho West  
## 8 Maine Northeast  
## 9 Massachusetts Northeast  
## 10 Montana West  
## 11 Nevada West  
## 12 New Hampshire Northeast  
## 13 New Jersey Northeast  
## 14 New Mexico West  
## 15 New York Northeast  
## 16 Oregon West  
## 17 Pennsylvania Northeast  
## 18 Rhode Island Northeast  
## 19 Utah West  
## 20 Vermont Northeast  
## 21 Washington West  
## 22 Wyoming West

nrow(murders\_nw)

## [1] 22

# 7. Suppose you want to live in the Northeast or West and want the murder rate to be less than 1. We want to see the data for the states satisfying these options. Note that you can use logical operators with filter. Here is an example in which we filter to keep only small states in the Northeast region.

filter(murders, population < 5000000 & region == "Northeast")

## state abb region population total population\_in\_millions rate  
## 1 Connecticut CT Northeast 3574097 97 3.574097 2.7139722  
## 2 Maine ME Northeast 1328361 11 1.328361 0.8280881  
## 3 New Hampshire NH Northeast 1316470 5 1.316470 0.3798036  
## 4 Rhode Island RI Northeast 1052567 16 1.052567 1.5200933  
## 5 Vermont VT Northeast 625741 2 0.625741 0.3196211  
## rank  
## 1 25  
## 2 44  
## 3 50  
## 4 35  
## 5 51

# Make sure murders has been defined with rate and rank and still has all states. Create a table called my\_states that contains rows for states satisfying both the conditions: it is in the Northeast or West and the murder rate is less than 1. Use select to show only the state name, the rate, and the rank.

my\_states <- filter(murders, rate < 1 & region %in% c("Northeast", "West"))  
my\_states

## state abb region population total population\_in\_millions rate  
## 1 Hawaii HI West 1360301 7 1.360301 0.5145920  
## 2 Idaho ID West 1567582 12 1.567582 0.7655102  
## 3 Maine ME Northeast 1328361 11 1.328361 0.8280881  
## 4 New Hampshire NH Northeast 1316470 5 1.316470 0.3798036  
## 5 Oregon OR West 3831074 36 3.831074 0.9396843  
## 6 Utah UT West 2763885 22 2.763885 0.7959810  
## 7 Vermont VT Northeast 625741 2 0.625741 0.3196211  
## 8 Wyoming WY West 563626 5 0.563626 0.8871131  
## rank  
## 1 49  
## 2 46  
## 3 44  
## 4 50  
## 5 42  
## 6 45  
## 7 51  
## 8 43

select(my\_states, state, rate, rank)

## state rate rank  
## 1 Hawaii 0.5145920 49  
## 2 Idaho 0.7655102 46  
## 3 Maine 0.8280881 44  
## 4 New Hampshire 0.3798036 50  
## 5 Oregon 0.9396843 42  
## 6 Utah 0.7959810 45  
## 7 Vermont 0.3196211 51  
## 8 Wyoming 0.8871131 43

# 8. The pipe %>% can be used to perform operations sequentially without having to define intermediate objects. Start by redefining murder to include rate and rank.

murders <- mutate(murders, rate = total / population \* 100000,   
 rank = rank(-rate))

# In the solution to the previous exercise, we did the following:

my\_states <- filter(murders, region %in% c("Northeast", "West") &   
 rate < 1)  
select(my\_states, state, rate, rank)

## state rate rank  
## 1 Hawaii 0.5145920 49  
## 2 Idaho 0.7655102 46  
## 3 Maine 0.8280881 44  
## 4 New Hampshire 0.3798036 50  
## 5 Oregon 0.9396843 42  
## 6 Utah 0.7959810 45  
## 7 Vermont 0.3196211 51  
## 8 Wyoming 0.8871131 43

# The pipe %>% permits us to perform both operations sequentially without having to define an intermediate variable my\_states. We therefore could have mutated and selected in the same line like this:

mutate(murders, rate = total / population \* 100000,   
 rank = rank(-rate)) %>%  
 select(state, rate, rank)

## state rate rank  
## 1 Alabama 2.8244238 23  
## 2 Alaska 2.6751860 27  
## 3 Arizona 3.6295273 10  
## 4 Arkansas 3.1893901 17  
## 5 California 3.3741383 14  
## 6 Colorado 1.2924531 38  
## 7 Connecticut 2.7139722 25  
## 8 Delaware 4.2319369 6  
## 9 District of Columbia 16.4527532 1  
## 10 Florida 3.3980688 13  
## 11 Georgia 3.7903226 9  
## 12 Hawaii 0.5145920 49  
## 13 Idaho 0.7655102 46  
## 14 Illinois 2.8369608 22  
## 15 Indiana 2.1900730 31  
## 16 Iowa 0.6893484 47  
## 17 Kansas 2.2081106 30  
## 18 Kentucky 2.6732010 28  
## 19 Louisiana 7.7425810 2  
## 20 Maine 0.8280881 44  
## 21 Maryland 5.0748655 4  
## 22 Massachusetts 1.8021791 32  
## 23 Michigan 4.1786225 7  
## 24 Minnesota 0.9992600 40  
## 25 Mississippi 4.0440846 8  
## 26 Missouri 5.3598917 3  
## 27 Montana 1.2128379 39  
## 28 Nebraska 1.7521372 33  
## 29 Nevada 3.1104763 19  
## 30 New Hampshire 0.3798036 50  
## 31 New Jersey 2.7980319 24  
## 32 New Mexico 3.2537239 15  
## 33 New York 2.6679599 29  
## 34 North Carolina 2.9993237 20  
## 35 North Dakota 0.5947151 48  
## 36 Ohio 2.6871225 26  
## 37 Oklahoma 2.9589340 21  
## 38 Oregon 0.9396843 42  
## 39 Pennsylvania 3.5977513 11  
## 40 Rhode Island 1.5200933 35  
## 41 South Carolina 4.4753235 5  
## 42 South Dakota 0.9825837 41  
## 43 Tennessee 3.4509357 12  
## 44 Texas 3.2013603 16  
## 45 Utah 0.7959810 45  
## 46 Vermont 0.3196211 51  
## 47 Virginia 3.1246001 18  
## 48 Washington 1.3829942 37  
## 49 West Virginia 1.4571013 36  
## 50 Wisconsin 1.7056487 34  
## 51 Wyoming 0.8871131 43

# Notice that select no longer has a data frame as the first argument. The first argument is assumed to be the result of the operation conducted right before the %>%. Repeat the previous exercise, but now instead of creating a new object, show the result and only include the state, rate, and rank columns. Use a pipe %>% to do this in just one line.

mutate(murders, region %in% c("Northeast", "West") & rate < 1) %>% select(state, rate, rank)

## state rate rank  
## 1 Alabama 2.8244238 23  
## 2 Alaska 2.6751860 27  
## 3 Arizona 3.6295273 10  
## 4 Arkansas 3.1893901 17  
## 5 California 3.3741383 14  
## 6 Colorado 1.2924531 38  
## 7 Connecticut 2.7139722 25  
## 8 Delaware 4.2319369 6  
## 9 District of Columbia 16.4527532 1  
## 10 Florida 3.3980688 13  
## 11 Georgia 3.7903226 9  
## 12 Hawaii 0.5145920 49  
## 13 Idaho 0.7655102 46  
## 14 Illinois 2.8369608 22  
## 15 Indiana 2.1900730 31  
## 16 Iowa 0.6893484 47  
## 17 Kansas 2.2081106 30  
## 18 Kentucky 2.6732010 28  
## 19 Louisiana 7.7425810 2  
## 20 Maine 0.8280881 44  
## 21 Maryland 5.0748655 4  
## 22 Massachusetts 1.8021791 32  
## 23 Michigan 4.1786225 7  
## 24 Minnesota 0.9992600 40  
## 25 Mississippi 4.0440846 8  
## 26 Missouri 5.3598917 3  
## 27 Montana 1.2128379 39  
## 28 Nebraska 1.7521372 33  
## 29 Nevada 3.1104763 19  
## 30 New Hampshire 0.3798036 50  
## 31 New Jersey 2.7980319 24  
## 32 New Mexico 3.2537239 15  
## 33 New York 2.6679599 29  
## 34 North Carolina 2.9993237 20  
## 35 North Dakota 0.5947151 48  
## 36 Ohio 2.6871225 26  
## 37 Oklahoma 2.9589340 21  
## 38 Oregon 0.9396843 42  
## 39 Pennsylvania 3.5977513 11  
## 40 Rhode Island 1.5200933 35  
## 41 South Carolina 4.4753235 5  
## 42 South Dakota 0.9825837 41  
## 43 Tennessee 3.4509357 12  
## 44 Texas 3.2013603 16  
## 45 Utah 0.7959810 45  
## 46 Vermont 0.3196211 51  
## 47 Virginia 3.1246001 18  
## 48 Washington 1.3829942 37  
## 49 West Virginia 1.4571013 36  
## 50 Wisconsin 1.7056487 34  
## 51 Wyoming 0.8871131 43

# 9. Reset murders to the original table by using data(murders). Use a pipe to create a new data frame called my\_states that considers only states in the Northeast or West which have a murder rate lower than 1, and contains only the state, rate and rank columns. The pipe should also have four components separated by three %>%. The code should look something like this:

my\_states <- murders %>% mutate(rate= total /   
 population\*100000, rank=rank(-rate)) %>%   
 filter(region %in% c("Northeast", "West") & rate < 1) %>%  
 select(state, rate, rank)  
my\_states

## state rate rank  
## 1 Hawaii 0.5145920 49  
## 2 Idaho 0.7655102 46  
## 3 Maine 0.8280881 44  
## 4 New Hampshire 0.3798036 50  
## 5 Oregon 0.9396843 42  
## 6 Utah 0.7959810 45  
## 7 Vermont 0.3196211 51  
## 8 Wyoming 0.8871131 43