Modul 6

Dio Cahyo Saputra

3/11/2020

Soal Nomor 1

Gunakan paket dplyr dan dataset “US murders”.

```
library(dslabs)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.6.3
```

```
##
## 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
```

```
data(murders)
```

Tambahkan kolom baru dengan nama ‘rate’ menggunakan fungsi mutate pada paket dplyr seperti pada contoh kode di bawah ini

```
rate <- mutate(murders, population_in_millions = population / 10^6)
```

```
murders = mutate(murders, rate = total / population * 100000)
murders
```

```
##                   state abb        region population total       rate
## 1               Alabama  AL         South    4779736   135  2.8244238
## 2                Alaska  AK          West     710231    19  2.6751860
## 3               Arizona  AZ          West    6392017   232  3.6295273
## 4              Arkansas  AR         South    2915918    93  3.1893901
## 5            California  CA          West   37253956  1257  3.3741383
## 6              Colorado  CO          West    5029196    65  1.2924531
## 7           Connecticut  CT     Northeast    3574097    97  2.7139722
## 8              Delaware  DE         South     897934    38  4.2319369
## 9  District of Columbia  DC         South     601723    99 16.4527532
## 10              Florida  FL         South   19687653   669  3.3980688
## 11              Georgia  GA         South    9920000   376  3.7903226
## 12               Hawaii  HI          West    1360301     7  0.5145920
## 13                Idaho  ID          West    1567582    12  0.7655102
```

```
## 14             Illinois  IL North Central   12830632   364  2.8369608
## 15              Indiana  IN North Central    6483802   142  2.1900730
## 16                 Iowa  IA North Central    3046355    21  0.6893484
## 17               Kansas  KS North Central    2853118    63  2.2081106
## 18             Kentucky  KY         South    4339367   116  2.6732010
## 19            Louisiana  LA         South    4533372   351  7.7425810
## 20                Maine  ME     Northeast    1328361    11  0.8280881
## 21             Maryland  MD         South    5773552   293  5.0748655
## 22        Massachusetts  MA     Northeast    6547629   118  1.8021791
## 23             Michigan  MI North Central    9883640   413  4.1786225
## 24            Minnesota  MN North Central    5303925    53  0.9992600
## 25          Mississippi  MS         South    2967297   120  4.0440846
## 26             Missouri  MO North Central    5988927   321  5.3598917
## 27              Montana  MT          West     989415    12  1.2128379
## 28             Nebraska  NE North Central    1826341    32  1.7521372
## 29               Nevada  NV          West    2700551    84  3.1104763
## 30        New Hampshire  NH     Northeast    1316470     5  0.3798036
## 31           New Jersey  NJ     Northeast    8791894   246  2.7980319
## 32           New Mexico  NM          West    2059179    67  3.2537239
## 33             New York  NY     Northeast   19378102   517  2.6679599
## 34       North Carolina  NC         South    9535483   286  2.9993237
## 35         North Dakota  ND North Central     672591     4  0.5947151
## 36                 Ohio  OH North Central   11536504   310  2.6871225
## 37             Oklahoma  OK         South    3751351   111  2.9589340
## 38               Oregon  OR          West    3831074    36  0.9396843
## 39         Pennsylvania  PA     Northeast   12702379   457  3.5977513
## 40         Rhode Island  RI     Northeast    1052567    16  1.5200933
## 41       South Carolina  SC         South    4625364   207  4.4753235
## 42         South Dakota  SD North Central     814180     8  0.9825837
## 43            Tennessee  TN         South    6346105   219  3.4509357
## 44                Texas  TX         South   25145561   805  3.2013603
## 45                 Utah  UT          West    2763885    22  0.7959810
## 46              Vermont  VT     Northeast     625741     2  0.3196211
## 47             Virginia  VA         South    8001024   250  3.1246001
## 48           Washington  WA          West    6724540    93  1.3829942
## 49        West Virginia  WV         South    1852994    27  1.4571013
## 50            Wisconsin  WI North Central    5686986    97  1.7056487
## 51              Wyoming  WY          West     563626     5  0.8871131
```

Soal Nomor 2

rank(x) menghasilkan pemeringkatan ‘x’ dari nilai terendah ke tertinggi. Gunakan fungsi mutate untuk menambahkan kolom baru yang berisi hasil pemeringkatan dari nilai tingkat pembunuhan tertinggi ke terendah.

```
murders = mutate(murders, rank = rank(rate))
murders
```

```
##                   state abb        region population total       rate rank
## 1               Alabama  AL         South    4779736   135  2.8244238   29
```

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

Soal Nomor 3

Dengan dplyr, kita dapat menggunakan fungsi select untuk menampilkan kolom tertentu saja. Misalnya dengan contoh script ini, kita hanya akan menampilkan kolom state dan population:

```
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
```

Gunakan select untuk menampilkan nama negara (state) dan singkatan (abb) dalam dataset "US murders".

```
select(murders, state, abb)
```

```
##                   state abb
## 1               Alabama  AL
## 2                Alaska  AK
## 3               Arizona  AZ
## 4              Arkansas  AR
## 5            California  CA
## 6              Colorado  CO
## 7           Connecticut  CT
## 8              Delaware  DE
## 9  District of Columbia  DC
## 10              Florida  FL
## 11              Georgia  GA
## 12               Hawaii  HI
## 13                Idaho  ID
## 14             Illinois  IL
## 15              Indiana  IN
## 16                 Iowa  IA
## 17               Kansas  KS
## 18             Kentucky  KY
## 19            Louisiana  LA
## 20                Maine  ME
## 21             Maryland  MD
## 22        Massachusetts  MA
## 23             Michigan  MI
## 24            Minnesota  MN
## 25          Mississippi  MS
## 26             Missouri  MO
## 27              Montana  MT
## 28             Nebraska  NE
## 29               Nevada  NV
```

```
## 30        New Hampshire  NH
## 31           New Jersey  NJ
## 32           New Mexico  NM
## 33             New York  NY
## 34       North Carolina  NC
## 35         North Dakota  ND
## 36                 Ohio  OH
## 37             Oklahoma  OK
## 38               Oregon  OR
## 39         Pennsylvania  PA
## 40         Rhode Island  RI
## 41       South Carolina  SC
## 42         South Dakota  SD
## 43            Tennessee  TN
## 44                Texas  TX
## 45                 Utah  UT
## 46              Vermont  VT
## 47             Virginia  VA
## 48           Washington  WA
## 49        West Virginia  WV
## 50            Wisconsin  WI
## 51              Wyoming  WY
```

Soal Nomor 4

Fungsi filter pada dplyr dapat digunakan untuk memilih baris tertentu dari data frame yang akan disimpan. Berbeda dengan select yang digunakan untuk memilih tampilan kolom, filter digunakan untuk memilih tampilan baris. Misalnya, kita ingin hanya menampilkan baris yang berisi dengan state = New York seperti contoh ini.

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

```
##      state abb    region population total    rate rank
## 1 New York  NY Northeast   19378102   517 2.66796   23
```

Gunakan filter untuk menampilkan 5 negara bagian teratas dengan tingkat pembunuhan tertinggi.

```
arrange(murders, desc(rate))
```

```
##                   state abb        region population total       rate rank
## 1  District of Columbia  DC         South     601723    99 16.4527532   51
## 2             Louisiana  LA         South    4533372   351  7.7425810   50
## 3              Missouri  MO North Central    5988927   321  5.3598917   49
## 4              Maryland  MD         South    5773552   293  5.0748655   48
## 5        South Carolina  SC         South    4625364   207  4.4753235   47
## 6              Delaware  DE         South     897934    38  4.2319369   46
## 7              Michigan  MI North Central    9883640   413  4.1786225   45
## 8           Mississippi  MS         South    2967297   120  4.0440846   44
## 9               Georgia  GA         South    9920000   376  3.7903226   43
## 10              Arizona  AZ          West    6392017   232  3.6295273   42
```

```
## 11          Pennsylvania  PA     Northeast   12702379   457  3.5977513   41
## 12             Tennessee  TN         South    6346105   219  3.4509357   40
## 13               Florida  FL         South   19687653   669  3.3980688   39
## 14            California  CA          West   37253956  1257  3.3741383   38
## 15            New Mexico  NM          West    2059179    67  3.2537239   37
## 16                 Texas  TX         South   25145561   805  3.2013603   36
## 17              Arkansas  AR         South    2915918    93  3.1893901   35
## 18              Virginia  VA         South    8001024   250  3.1246001   34
## 19                Nevada  NV          West    2700551    84  3.1104763   33
## 20        North Carolina  NC         South    9535483   286  2.9993237   32
## 21              Oklahoma  OK         South    3751351   111  2.9589340   31
## 22              Illinois  IL North Central   12830632   364  2.8369608   30
## 23               Alabama  AL         South    4779736   135  2.8244238   29
## 24            New Jersey  NJ     Northeast    8791894   246  2.7980319   28
## 25           Connecticut  CT     Northeast    3574097    97  2.7139722   27
## 26                  Ohio  OH North Central   11536504   310  2.6871225   26
## 27                Alaska  AK          West     710231    19  2.6751860   25
## 28              Kentucky  KY         South    4339367   116  2.6732010   24
## 29              New York  NY     Northeast   19378102   517  2.6679599   23
## 30                Kansas  KS North Central    2853118    63  2.2081106   22
## 31               Indiana  IN North Central    6483802   142  2.1900730   21
## 32         Massachusetts  MA     Northeast    6547629   118  1.8021791   20
## 33              Nebraska  NE North Central    1826341    32  1.7521372   19
## 34             Wisconsin  WI North Central    5686986    97  1.7056487   18
## 35          Rhode Island  RI     Northeast    1052567    16  1.5200933   17
## 36         West Virginia  WV         South    1852994    27  1.4571013   16
## 37            Washington  WA          West    6724540    93  1.3829942   15
## 38              Colorado  CO          West    5029196    65  1.2924531   14
## 39               Montana  MT          West     989415    12  1.2128379   13
## 40             Minnesota  MN North Central    5303925    53  0.9992600   12
## 41          South Dakota  SD North Central     814180     8  0.9825837   11
## 42                Oregon  OR          West    3831074    36  0.9396843   10
## 43               Wyoming  WY          West     563626     5  0.8871131    9
## 44                 Maine  ME     Northeast    1328361    11  0.8280881    8
## 45                  Utah  UT          West    2763885    22  0.7959810    7
## 46                 Idaho  ID          West    1567582    12  0.7655102    6
## 47                  Iowa  IA North Central    3046355    21  0.6893484    5
## 48          North Dakota  ND North Central     672591     4  0.5947151    4
## 49                Hawaii  HI          West    1360301     7  0.5145920    3
## 50         New Hampshire  NH     Northeast    1316470     5  0.3798036    2
## 51               Vermont  VT     Northeast     625741     2  0.3196211    1
```

```
filter(murders, rate > 4.4)
```

```
##                   state abb        region population total      rate rank
## 1 District of Columbia  DC         South     601723    99 16.452753   51
## 2             Louisiana  LA         South    4533372   351  7.742581   50
## 3              Maryland  MD         South    5773552   293  5.074866   48
## 4              Missouri  MO North Central    5988927   321  5.359892   49
## 5        South Carolina  SC         South    4625364   207  4.475323   47
```

Soal Nomor 5

Buat script yang dapat menampilkan hasil sesuai kondisi berikut: seseorang ingin tinggal di regional Northeast atau West dan ingin calon tempat tinggal yang dipilih memiliki tingkat pembunuhan kurang dari 1. Gunakan filter untuk hanya menampilkan hasil yang terdiri dari: state, rate, dan peringkatnya.

```
murders %>% filter(region == 'Northeast', rate < 1) %>%
select(state,rate,rank)
```

```
##           state      rate rank
## 1         Maine 0.8280881    8
## 2 New Hampshire 0.3798036    2
## 3       Vermont 0.3196211    1
```

```
murders %>% filter(region == 'West', rate < 1) %>% select(state,rate,rank)
```

```
##     state      rate rank
## 1  Hawaii 0.5145920    3
## 2   Idaho 0.7655102    6
## 3  Oregon 0.9396843   10
## 4    Utah 0.7959810    7
## 5 Wyoming 0.8871131    9
```

Operator Pipe

Nomor 1

Reset dataset “US murders” ke tabel aslinya dengan melakukan update dengan perintah: data(murders). Gunakan operator pipe untuk membuat data frame baru dengan nama ‘my_states’ yang hanya berisi negara-negara di regional Northeast atau Eastwest yang memiliki tingkat pembunuhan kurang dari 1, dan hanya menampilkan kolom: state, tingkat, dan rate. Script yang dibuat seharusnya terdiri dari empat komponen yang dipisahkan oleh tiga %>%. Seperti contoh kerangka ini:

```
data("murders")
my_states <- murders %>%
  mutate(rate = total / population * 100000) %>%
  filter(region == 'Northeast' , rate < 1) %>%
  select(state,total,rate)
my_states
```

```
##           state total      rate
## 1         Maine    11 0.8280881
## 2 New Hampshire     5 0.3798036
## 3       Vermont     2 0.3196211
```