CH3

2023-07-18

1. tidyverse

```
library(tidyverse)
dat <- read.csv("C:\\Users\\phl02\\Desktop\\P\\bio\\ch3\\Ch3_chb.csv")
```

```
dat1<-dat
class(dat1)
```

```
## [1] "data.frame"
```

```
dim(dat1)
```

```
## [1] 24 26
```

```
colnames(dat1)
```

```
##  [1] "id"         "index_date" "gender"     "age"        "treat_gr"
##  [6] "lc"         "hcc"        "hcc_yr"     "b_alt"      "b_bil"
## [11] "b_inr"      "b_cr"       "b_plt"      "b_alb"      "m6_alt"
## [16] "m6_bil"     "m6_inr"     "m6_cr"      "m6_plt"     "m6_alb"
## [21] "m12_alt"    "m12_bil"    "m12_inr"    "m12_cr"     "m12_plt"
## [26] "m12_alb"
```

2. Select

원하는 변수 선택

R에서 기본으로 사용하는 슬라이싱 코드와 비교

```
dat1 %>%
  select(id, b_alt, b_bil, b_alb)
```

```
##    id b_alt b_bil b_alb
## 1   1    67   1.2   3.8
## 2   2    32   1.1   3.8
## 3   3   106   1.8   3.9
## 4   4   159   1.4   4.2
## 5   5    94   0.8   4.2
## 6   6    32   0.7   3.4
## 7   7   104   1.1    NA
## 8   8   143   1.5   3.5
## 9   9    31   1.3   3.9
## 10 10   239   1.3   3.2
## 11 11   359   0.9   4.1
## 12 12   486   0.7   2.8
## 13 13    56   1.3   4.2
## 14 14    34   1.7    NA
## 15 15   107   2.1   2.1
## 16 16    40   1.3   3.4
## 17 17    48   1.0   4.0
## 18 18    61   2.1   2.7
## 19 19    34   1.1   4.3
## 20 20    31   0.9   3.1
## 21 21    57   1.1   4.1
## 22 22    81   1.0   4.1
## 23 23    35   1.9   4.4
## 24 24    84   1.1   4.2
```

```r
dat1[ , c('id','b_alt','b_bil','b_alb')]
```

```
##    id b_alt b_bil b_alb
## 1   1    67   1.2   3.8
## 2   2    32   1.1   3.8
## 3   3   106   1.8   3.9
## 4   4   159   1.4   4.2
## 5   5    94   0.8   4.2
## 6   6    32   0.7   3.4
## 7   7   104   1.1    NA
## 8   8   143   1.5   3.5
## 9   9    31   1.3   3.9
## 10 10   239   1.3   3.2
## 11 11   359   0.9   4.1
## 12 12   486   0.7   2.8
## 13 13    56   1.3   4.2
## 14 14    34   1.7    NA
## 15 15   107   2.1   2.1
## 16 16    40   1.3   3.4
## 17 17    48   1.0   4.0
## 18 18    61   2.1   2.7
## 19 19    34   1.1   4.3
## 20 20    31   0.9   3.1
## 21 21    57   1.1   4.1
## 22 22    81   1.0   4.1
## 23 23    35   1.9   4.4
## 24 24    84   1.1   4.2
```

숫자를 사용해서도 가능

```
dat1 %>%
  select(1, 8, 9, 13)
```

```
##    id    hcc_yr b_alt b_plt
## 1   1  7.641667    67   110
## 2   2  9.763889    32   187
## 3   3 10.594444   106   133
## 4   4  6.016667   159   164
## 5   5  2.355556    94   153
## 6   6  6.069444    32   170
## 7   7  8.069444   104   162
## 8   8 10.716667   143   165
## 9   9  2.163889    31   157
## 10 10 10.766667   239   166
## 11 11  3.713889   359   202
## 12 12  3.269444   486   159
## 13 13 10.655556    56   186
## 14 14  4.488889    34   159
## 15 15  5.869444   107    66
## 16 16  4.008333    40   118
## 17 17  3.513889    48   105
## 18 18  2.372222    61    55
## 19 19 10.677778    34    NA
## 20 20  5.030556    31   163
## 21 21  4.725000    57    91
## 22 22 10.761111    81   292
## 23 23 10.558333    35   204
## 24 24  3.994444    84   181
```

다양하게도 선택 가능

```
dat1 %>%
  select(1:4, 7:9, 13)
```

```
##    id index_date gender age hcc    hcc_yr b_alt b_plt
## 1   1 2007/01/05      M  54   1  7.641667    67   110
## 2   2 2007/01/10      F  45   0  9.763889    32   187
## 3   3 2007/01/11      M  49   0 10.594444   106   133
## 4   4 2007/01/12      M  26   0  6.016667   159   164
## 5   5 2007/01/18      M  50   0  2.355556    94   153
## 6   6 2007/01/18      M  33   0  6.069444    32   170
## 7   7 2007/01/26      M  49   1  8.069444   104   162
## 8   8 2007/01/31      F  50   0 10.716667   143   165
## 9   9 2007/01/31      F  49   0  2.163889    31   157
## 10 10 2007/02/01      M  50   0 10.766667   239   166
## 11 11 2007/02/01      M  32   0  3.713889   359   202
## 12 12 2007/02/01      M  48   1  3.269444   486   159
## 13 13 2007/02/01      M  51   0 10.655556    56   186
## 14 14 2007/02/02      M  41   0  4.488889    34   159
## 15 15 2007/02/08      M  50   1  5.869444   107    66
## 16 16 2007/02/08      M  51   0  4.008333    40   118
## 17 17 2007/02/09      F  62   0  3.513889    48   105
## 18 18 2007/02/14      F  51   0  2.372222    61    55
## 19 19 2007/02/15      M  48   0 10.677778    34    NA
## 20 20 2007/02/15      F  50   0  5.030556    31   163
## 21 21 2007/02/15      M  52   0  4.725000    57    91
## 22 22 2007/02/16      M  30   0 10.761111    81   292
## 23 23 2007/02/16      M  29   0 10.558333    35   204
## 24 24 2007/02/20      M  43   0  3.994444    84   181
```

콜론도 가능

```
dat1 %>%
  select(m6_alt:m6_plt)
```

```
##    m6_alt m6_bil m6_inr m6_cr m6_plt
## 1      35    1.0   1.05   0.9     96
## 2      16    1.1   0.97   0.7    194
## 3     108    1.8   1.03   0.9    120
## 4      24    1.7   1.11   0.9    149
## 5      57    0.9   1.10   1.0    145
## 6      26    0.8   0.97   1.4     NA
## 7      51    1.0   1.05   0.8    128
## 8      24    1.4   1.04   0.8    184
## 9      25    1.3   1.02   0.8    163
## 10     30    1.4   1.00   1.0    158
## 11     23    1.1   0.98   1.0    171
## 12     34    0.9   0.96   0.8    215
## 13     18    1.3     NA   1.1     NA
## 14     23    1.6   0.97   0.9    164
## 15     42    1.0   1.15   1.0     80
## 16     33    1.1   1.09   1.0    129
## 17     28    0.9   1.03   0.7    121
## 18     25    1.2   1.21   0.7     67
## 19     13    0.8     NA   1.1     NA
## 20     29    1.3   1.19   0.7     40
## 21     39    1.3   1.11   0.8     74
## 22     50    1.0   0.93   0.9    264
## 23     26    2.3   1.00   0.9     NA
## 24     63    0.8   1.02   0.9    144
```

특정 변수 제외하기

```
dat1 %>%
  select(-id, -index_date) %>%
  colnames()
```

```
##  [1] "gender"   "age"      "treat_gr" "lc"       "hcc"      "hcc_yr"
##  [7] "b_alt"    "b_bil"    "b_inr"    "b_cr"     "b_plt"    "b_alb"
## [13] "m6_alt"   "m6_bil"   "m6_inr"   "m6_cr"    "m6_plt"   "m6_alb"
## [19] "m12_alt"  "m12_bil"  "m12_inr"  "m12_cr"   "m12_plt"  "m12_alb"
```

변수 이름 삭제 확인

```
colnames(dat1)
```

```
##  [1] "id"         "index_date" "gender"     "age"        "treat_gr"
##  [6] "lc"         "hcc"        "hcc_yr"     "b_alt"      "b_bil"
## [11] "b_inr"      "b_cr"       "b_plt"      "b_alb"      "m6_alt"
## [16] "m6_bil"     "m6_inr"     "m6_cr"      "m6_plt"     "m6_alb"
## [21] "m12_alt"    "m12_bil"    "m12_inr"    "m12_cr"     "m12_plt"
## [26] "m12_alb"
```

순서를 사용해도 가능

```
dat1 %>%
  select(-1, -2) %>%
  colnames()
```

```
##  [1] "gender"   "age"      "treat_gr" "lc"       "hcc"      "hcc_yr"
##  [7] "b_alt"    "b_bil"    "b_inr"    "b_cr"     "b_plt"    "b_alb"
## [13] "m6_alt"   "m6_bil"   "m6_inr"   "m6_cr"    "m6_plt"   "m6_alb"
## [19] "m12_alt"  "m12_bil"  "m12_inr"  "m12_cr"   "m12_plt"  "m12_alb"
```

특정 변수를 조합해서 제외

```
dat1 %>%
  select(-(1:4), -(10:13)) %>%
  ncol()
```

```
## [1] 18
```

같은 특성을 가진 변수만 선택

```
dat1 %>%
  select(b_alt, m6_alt, m12_alt)
```

```
##    b_alt m6_alt m12_alt
## 1     67     35      37
## 2     32     16      18
## 3    106    108      27
## 4    159     24      29
## 5     94     57      35
## 6     32     26      17
## 7    104     51      42
## 8    143     24      16
## 9     31     25      24
## 10   239     30      34
## 11   359     23      19
## 12   486     34      24
## 13    56     18      17
## 14    34     23      19
## 15   107     42      36
## 16    40     33      30
## 17    48     28      31
## 18    61     25      24
## 19    34     13      13
## 20    31     29      31
## 21    57     39      34
## 22    81     50      63
## 23    35     26      28
## 24    84     63      52
```

포함된 단어가 있으면 이 방법이 더 유용

```
dat1 %>%
  select(contains('alt'))
```

```
##    b_alt m6_alt m12_alt
## 1     67     35      37
## 2     32     16      18
## 3    106    108      27
## 4    159     24      29
## 5     94     57      35
## 6     32     26      17
## 7    104     51      42
## 8    143     24      16
## 9     31     25      24
## 10   239     30      34
## 11   359     23      19
## 12   486     34      24
## 13    56     18      17
## 14    34     23      19
## 15   107     42      36
## 16    40     33      30
## 17    48     28      31
## 18    61     25      24
## 19    34     13      13
## 20    31     29      31
## 21    57     39      34
## 22    81     50      63
## 23    35     26      28
## 24    84     63      52
```

6개월째 측정한 변수만 선택

```
dat1 %>%
  select(contains('m6'))
```

```
##    m6_alt m6_bil m6_inr m6_cr m6_plt m6_alb
## 1      35    1.0   1.05   0.9     96    3.6
## 2      16    1.1   0.97   0.7    194    3.8
## 3     108    1.8   1.03   0.9    120    4.3
## 4      24    1.7   1.11   0.9    149    4.2
## 5      57    0.9   1.10   1.0    145    4.6
## 6      26    0.8   0.97   1.4     NA    3.4
## 7      51    1.0   1.05   0.8    128    3.9
## 8      24    1.4   1.04   0.8    184    3.7
## 9      25    1.3   1.02   0.8    163    4.2
## 10     30    1.4   1.00   1.0    158    3.8
## 11     23    1.1   0.98   1.0    171    4.4
## 12     34    0.9   0.96   0.8    215    3.3
## 13     18    1.3     NA   1.1     NA    4.2
## 14     23    1.6   0.97   0.9    164    3.7
## 15     42    1.0   1.15   1.0     80    3.7
## 16     33    1.1   1.09   1.0    129    4.0
## 17     28    0.9   1.03   0.7    121    3.9
## 18     25    1.2   1.21   0.7     67    3.5
## 19     13    0.8     NA   1.1     NA    4.3
## 20     29    1.3   1.19   0.7     40    3.3
## 21     39    1.3   1.11   0.8     74    4.0
## 22     50    1.0   0.93   0.9    264    4.4
## 23     26    2.3   1.00   0.9     NA    4.1
## 24     63    0.8   1.02   0.9    144    4.5
```

m6으로 시작하는 변수만 select 하는 방법도 있음

```
dat1 %>%
  select(starts_with('m6'))
```

```
##    m6_alt m6_bil m6_inr m6_cr m6_plt m6_alb
## 1      35    1.0   1.05   0.9     96    3.6
## 2      16    1.1   0.97   0.7    194    3.8
## 3     108    1.8   1.03   0.9    120    4.3
## 4      24    1.7   1.11   0.9    149    4.2
## 5      57    0.9   1.10   1.0    145    4.6
## 6      26    0.8   0.97   1.4     NA    3.4
## 7      51    1.0   1.05   0.8    128    3.9
## 8      24    1.4   1.04   0.8    184    3.7
## 9      25    1.3   1.02   0.8    163    4.2
## 10     30    1.4   1.00   1.0    158    3.8
## 11     23    1.1   0.98   1.0    171    4.4
## 12     34    0.9   0.96   0.8    215    3.3
## 13     18    1.3     NA   1.1     NA    4.2
## 14     23    1.6   0.97   0.9    164    3.7
## 15     42    1.0   1.15   1.0     80    3.7
## 16     33    1.1   1.09   1.0    129    4.0
## 17     28    0.9   1.03   0.7    121    3.9
## 18     25    1.2   1.21   0.7     67    3.5
## 19     13    0.8     NA   1.1     NA    4.3
## 20     29    1.3   1.19   0.7     40    3.3
## 21     39    1.3   1.11   0.8     74    4.0
## 22     50    1.0   0.93   0.9    264    4.4
## 23     26    2.3   1.00   0.9     NA    4.1
## 24     63    0.8   1.02   0.9    144    4.5
```

cr로 끝나는 변수

```
dat1 %>%
  select(ends_with('cr'))
```

```
##    b_cr m6_cr m12_cr
## 1  1.00   0.9    1.0
## 2  0.60   0.7    0.6
## 3  1.05   0.9    0.9
## 4  0.93   0.9    0.8
## 5  0.90   1.0    1.0
## 6  1.20   1.4    1.9
## 7  0.91   0.8     NA
## 8  0.70   0.8    0.7
## 9  0.73   0.8    0.7
## 10 1.20   1.0    1.0
## 11 1.10   1.0    1.1
## 12 0.81   0.8    0.8
## 13 1.00   1.1    1.1
## 14 0.69   0.9    0.8
## 15 1.00   1.0    1.1
## 16 0.92   1.0     NA
## 17 0.81   0.7     NA
## 18 0.60   0.7    0.6
## 19 1.00   1.1    1.1
## 20 0.63   0.7    0.7
## 21 0.80   0.8    0.8
## 22 0.99   0.9    1.0
## 23 0.91   0.9    0.9
## 24 0.89   0.9    1.0
```

변수 이름 변경

변경한 변수만 보임

```
dat1 %>%
  select(sex = gender)
```

```
##    sex
## 1    M
## 2    F
## 3    M
## 4    M
## 5    M
## 6    M
## 7    M
## 8    F
## 9    F
## 10   M
## 11   M
## 12   M
## 13   M
## 14   M
## 15   M
## 16   M
## 17   F
## 18   F
## 19   M
## 20   F
## 21   M
## 22   M
## 23   M
## 24   M
```

이름 변경 후 다른 변수들은 그대로 유지하고 싶을 때

```
dat1 %>%
  rename(sex = gender)
```

```
##    id index_date sex age treat_gr lc hcc    hcc_yr b_alt b_bil b_inr b_cr b_plt
## 1   1 2007/01/05   M  54      ETV  1   1  7.641667    67   1.2  1.17 1.00   110
## 2   2 2007/01/10   F  45      ETV  0   0  9.763889    32   1.1  0.90 0.60   187
## 3   3 2007/01/11   M  49      TDF  1   0 10.594444   106   1.8  1.03 1.05   133
## 4   4 2007/01/12   M  26      ETV  0   0  6.016667   159   1.4  1.04 0.93   164
## 5   5 2007/01/18   M  50      TDF  1   0  2.355556    94   0.8  1.12 0.90   153
## 6   6 2007/01/18   M  33      ETV  1   0  6.069444    32   0.7  0.88 1.20   170
## 7   7 2007/01/26   M  49      TDF  1   1  8.069444   104   1.1  1.10 0.91   162
## 8   8 2007/01/31   F  50      TDF  0   0 10.716667   143   1.5  1.10 0.70   165
## 9   9 2007/01/31   F  49      ETV  1   0  2.163889    31   1.3    NA 0.73   157
## 10 10 2007/02/01   M  50      ETV  0   0 10.766667   239   1.3  1.03 1.20   166
## 11 11 2007/02/01   M  32      ETV  0   0  3.713889   359   0.9  0.97 1.10   202
## 12 12 2007/02/01   M  48      TDF  0   1  3.269444   486   0.7  1.02 0.81   159
## 13 13 2007/02/01   M  51      ETV  0   0 10.655556    56   1.3  1.10 1.00   186
## 14 14 2007/02/02   M  41      ETV  1   0  4.488889    34   1.7  1.08 0.69   159
## 15 15 2007/02/08   M  50      ETV  1   1  5.869444   107   2.1  1.57 1.00    66
## 16 16 2007/02/08   M  51      ETV  1   0  4.008333    40   1.3  1.15 0.92   118
## 17 17 2007/02/09   F  62      TDF  1   0  3.513889    48   1.0  1.02 0.81   105
## 18 18 2007/02/14   F  51      ETV  1   0  2.372222    61   2.1  1.43 0.60    55
## 19 19 2007/02/15   M  48      TDF  0   0 10.677778    34   1.1  1.10 1.00    NA
## 20 20 2007/02/15   F  50      ETV  1   0  5.030556    31   0.9  1.25 0.63   163
## 21 21 2007/02/15   M  52      TDF  1   0  4.725000    57   1.1  1.11 0.80    91
## 22 22 2007/02/16   M  30      ETV  0   0 10.761111    81   1.0  0.97 0.99   292
## 23 23 2007/02/16   M  29      TDF  0   0 10.558333    35   1.9  1.03 0.91   204
## 24 24 2007/02/20   M  43      ETV  1   0  3.994444    84   1.1  1.07 0.89   181
##    b_alb m6_alt m6_bil m6_inr m6_cr m6_plt m6_alb m12_alt m12_bil m12_inr
## 1    3.8     35    1.0   1.05   0.9     96    3.6      37     1.2      NA
## 2    3.8     16    1.1   0.97   0.7    194    3.8      18     1.1    0.96
## 3    3.9    108    1.8   1.03   0.9    120    4.3      27     1.5    1.02
## 4    4.2     24    1.7   1.11   0.9    149    4.2      29     1.2    0.98
## 5    4.2     57    0.9   1.10   1.0    145    4.6      35     1.2    1.05
## 6    3.4     26    0.8   0.97   1.4     NA    3.4      17     0.6    0.91
## 7     NA     51    1.0   1.05   0.8    128    3.9      42     0.7    1.04
## 8    3.5     24    1.4   1.04   0.8    184    3.7      16     0.9    1.06
## 9    3.9     25    1.3   1.02   0.8    163    4.2      24     0.9    1.04
## 10   3.2     30    1.4   1.00   1.0    158    3.8      34     1.3    1.03
## 11   4.1     23    1.1   0.98   1.0    171    4.4      19     0.9    0.94
## 12   2.8     34    0.9   0.96   0.8    215    3.3      24     0.8    0.95
## 13   4.2     18    1.3     NA   1.1     NA    4.2      17     1.4    1.01
## 14    NA     23    1.6   0.97   0.9    164    3.7      19     1.2    0.97
## 15   2.1     42    1.0   1.15   1.0     80    3.7      36     1.1    1.09
## 16   3.4     33    1.1   1.09   1.0    129    4.0      30     1.1    0.98
## 17   4.0     28    0.9   1.03   0.7    121    3.9      31     0.9    1.06
## 18   2.7     25    1.2   1.21   0.7     67    3.5      24     0.7    1.15
## 19   4.3     13    0.8     NA   1.1     NA    4.3      13     1.1    1.02
## 20   3.1     29    1.3   1.19   0.7     40    3.3      31     1.4    1.17
## 21   4.1     39    1.3   1.11   0.8     74    4.0      34     1.3    1.04
## 22   4.1     50    1.0   0.93   0.9    264    4.4      63     0.8    0.93
## 23   4.4     26    2.3   1.00   0.9     NA    4.1      28     1.3    0.93
```

```
## 24   4.2     63    0.8   1.02   0.9    144    4.5      52     0.9    1.02
##    m12_cr m12_plt m12_alb
## 1     1.0     111     4.0
## 2     0.6     239     4.0
## 3     0.9     120     4.2
## 4     0.8     185     4.1
## 5     1.0     141     4.5
## 6     1.9     179     3.0
## 7      NA      NA     3.9
## 8     0.7     138     3.4
## 9     0.7     148     3.9
## 10    1.0     151     3.7
## 11    1.1      NA     4.3
## 12    0.8     272     3.4
## 13    1.1     143     4.3
## 14    0.8     175     3.8
## 15    1.1     103     4.2
## 16     NA      NA     4.1
## 17     NA      NA     3.9
## 18    0.6      61     3.5
## 19    1.1     143     4.2
## 20    0.7      36     3.4
## 21    0.8      64     3.9
## 22    1.0     286     4.5
## 23    0.9     113     4.0
## 24    1.0     136     4.3
```

여러 개도 가능

```
dat1 %>%
  rename(emr_id = id, baseline_date = index_date, sex = gender) %>%
  colnames()
```

```
##  [1] "emr_id"        "baseline_date" "sex"           "age"
##  [5] "treat_gr"      "lc"            "hcc"           "hcc_yr"
##  [9] "b_alt"         "b_bil"         "b_inr"         "b_cr"
## [13] "b_plt"         "b_alb"         "m6_alt"        "m6_bil"
## [17] "m6_inr"        "m6_cr"         "m6_plt"        "m6_alb"
## [21] "m12_alt"       "m12_bil"       "m12_inr"       "m12_cr"
## [25] "m12_plt"       "m12_alb"
```

변수의 순서 변경

```
dat1 %>%
  select(treat_gr, hcc, id, everything()) %>%
  colnames()
```

```
##  [1] "treat_gr"   "hcc"        "id"         "index_date" "gender"
##  [6] "age"        "lc"         "hcc_yr"     "b_alt"      "b_bil"
## [11] "b_inr"      "b_cr"       "b_plt"      "b_alb"      "m6_alt"
## [16] "m6_bil"     "m6_inr"     "m6_cr"      "m6_plt"     "m6_alb"
## [21] "m12_alt"    "m12_bil"    "m12_inr"    "m12_cr"     "m12_plt"
## [26] "m12_alb"
```

동시에 기능 사용하기

어떤 코드가 더 가독성이 좋은지 생각해보기
3가지 방법의 결과는 같음

```
dat1 %>%
  select(id, gender, contains('m12')) %>%
  rename(sex = gender) %>%
  select(sex, everything())
```

```
##    sex id m12_alt m12_bil m12_inr m12_cr m12_plt m12_alb
## 1    M  1      37     1.2      NA    1.0     111     4.0
## 2    F  2      18     1.1    0.96    0.6     239     4.0
## 3    M  3      27     1.5    1.02    0.9     120     4.2
## 4    M  4      29     1.2    0.98    0.8     185     4.1
## 5    M  5      35     1.2    1.05    1.0     141     4.5
## 6    M  6      17     0.6    0.91    1.9     179     3.0
## 7    M  7      42     0.7    1.04     NA      NA     3.9
## 8    F  8      16     0.9    1.06    0.7     138     3.4
## 9    F  9      24     0.9    1.04    0.7     148     3.9
## 10   M 10      34     1.3    1.03    1.0     151     3.7
## 11   M 11      19     0.9    0.94    1.1      NA     4.3
## 12   M 12      24     0.8    0.95    0.8     272     3.4
## 13   M 13      17     1.4    1.01    1.1     143     4.3
## 14   M 14      19     1.2    0.97    0.8     175     3.8
## 15   M 15      36     1.1    1.09    1.1     103     4.2
## 16   M 16      30     1.1    0.98     NA      NA     4.1
## 17   F 17      31     0.9    1.06     NA      NA     3.9
## 18   F 18      24     0.7    1.15    0.6      61     3.5
## 19   M 19      13     1.1    1.02    1.1     143     4.2
## 20   F 20      31     1.4    1.17    0.7      36     3.4
## 21   M 21      34     1.3    1.04    0.8      64     3.9
## 22   M 22      63     0.8    0.93    1.0     286     4.5
## 23   M 23      28     1.3    0.93    0.9     113     4.0
## 24   M 24      52     0.9    1.02    1.0     136     4.3
```

```r
dat1 %>%
  select(sex = gender, id, contains('m12'))
```

```
##    sex id m12_alt m12_bil m12_inr m12_cr m12_plt m12_alb
## 1    M  1      37     1.2      NA    1.0     111     4.0
## 2    F  2      18     1.1    0.96    0.6     239     4.0
## 3    M  3      27     1.5    1.02    0.9     120     4.2
## 4    M  4      29     1.2    0.98    0.8     185     4.1
## 5    M  5      35     1.2    1.05    1.0     141     4.5
## 6    M  6      17     0.6    0.91    1.9     179     3.0
## 7    M  7      42     0.7    1.04     NA      NA     3.9
## 8    F  8      16     0.9    1.06    0.7     138     3.4
## 9    F  9      24     0.9    1.04    0.7     148     3.9
## 10   M 10      34     1.3    1.03    1.0     151     3.7
## 11   M 11      19     0.9    0.94    1.1      NA     4.3
## 12   M 12      24     0.8    0.95    0.8     272     3.4
## 13   M 13      17     1.4    1.01    1.1     143     4.3
## 14   M 14      19     1.2    0.97    0.8     175     3.8
## 15   M 15      36     1.1    1.09    1.1     103     4.2
## 16   M 16      30     1.1    0.98     NA      NA     4.1
## 17   F 17      31     0.9    1.06     NA      NA     3.9
## 18   F 18      24     0.7    1.15    0.6      61     3.5
## 19   M 19      13     1.1    1.02    1.1     143     4.2
## 20   F 20      31     1.4    1.17    0.7      36     3.4
## 21   M 21      34     1.3    1.04    0.8      64     3.9
## 22   M 22      63     0.8    0.93    1.0     286     4.5
## 23   M 23      28     1.3    0.93    0.9     113     4.0
## 24   M 24      52     0.9    1.02    1.0     136     4.3
```

```
dat1 %>%
  select(sex = gender,
         id,
         contains('m12'))
```

```
##    sex id m12_alt m12_bil m12_inr m12_cr m12_plt m12_alb
## 1    M  1      37     1.2      NA    1.0     111     4.0
## 2    F  2      18     1.1    0.96    0.6     239     4.0
## 3    M  3      27     1.5    1.02    0.9     120     4.2
## 4    M  4      29     1.2    0.98    0.8     185     4.1
## 5    M  5      35     1.2    1.05    1.0     141     4.5
## 6    M  6      17     0.6    0.91    1.9     179     3.0
## 7    M  7      42     0.7    1.04     NA      NA     3.9
## 8    F  8      16     0.9    1.06    0.7     138     3.4
## 9    F  9      24     0.9    1.04    0.7     148     3.9
## 10   M 10      34     1.3    1.03    1.0     151     3.7
## 11   M 11      19     0.9    0.94    1.1      NA     4.3
## 12   M 12      24     0.8    0.95    0.8     272     3.4
## 13   M 13      17     1.4    1.01    1.1     143     4.3
## 14   M 14      19     1.2    0.97    0.8     175     3.8
## 15   M 15      36     1.1    1.09    1.1     103     4.2
## 16   M 16      30     1.1    0.98     NA      NA     4.1
## 17   F 17      31     0.9    1.06     NA      NA     3.9
## 18   F 18      24     0.7    1.15    0.6      61     3.5
## 19   M 19      13     1.1    1.02    1.1     143     4.2
## 20   F 20      31     1.4    1.17    0.7      36     3.4
## 21   M 21      34     1.3    1.04    0.8      64     3.9
## 22   M 22      63     0.8    0.93    1.0     286     4.5
## 23   M 23      28     1.3    0.93    0.9     113     4.0
## 24   M 24      52     0.9    1.02    1.0     136     4.3
```

흔히 범하는 오류

4번째에서 7번째까지 변수 제외를 원함

```
dat1 %>%
  select(-4:7) %>%
  colnames()
```

```
## [1] "id"         "index_date" "gender"     "age"        "treat_gr"
## [6] "lc"         "hcc"
```

첫 번째에서 7번째까지 선택됨 아래와 같이 코드를 짜야 원하는 결과를 얻음

```
dat1 %>%
  select(-c(4:7)) %>%
  colnames()
```

```
##  [1] "id"         "index_date" "gender"     "hcc_yr"     "b_alt"
##  [6] "b_bil"      "b_inr"      "b_cr"       "b_plt"      "b_alb"
## [11] "m6_alt"     "m6_bil"     "m6_inr"     "m6_cr"      "m6_plt"
## [16] "m6_alb"     "m12_alt"    "m12_bil"    "m12_inr"    "m12_cr"
## [21] "m12_plt"    "m12_alb"
```

3. Filter

특정 조건을 만족하는 환자(row) 선택하기

```
dat1 %>%
  filter(id<=20)
```

```
##    id index_date gender age treat_gr lc hcc    hcc_yr b_alt b_bil b_inr b_cr
## 1   1 2007/01/05      M  54      ETV  1   1  7.641667    67   1.2  1.17 1.00
## 2   2 2007/01/10      F  45      ETV  0   0  9.763889    32   1.1  0.90 0.60
## 3   3 2007/01/11      M  49      TDF  1   0 10.594444   106   1.8  1.03 1.05
## 4   4 2007/01/12      M  26      ETV  0   0  6.016667   159   1.4  1.04 0.93
## 5   5 2007/01/18      M  50      TDF  1   0  2.355556    94   0.8  1.12 0.90
## 6   6 2007/01/18      M  33      ETV  1   0  6.069444    32   0.7  0.88 1.20
## 7   7 2007/01/26      M  49      TDF  1   1  8.069444   104   1.1  1.10 0.91
## 8   8 2007/01/31      F  50      TDF  0   0 10.716667   143   1.5  1.10 0.70
## 9   9 2007/01/31      F  49      ETV  1   0  2.163889    31   1.3    NA 0.73
## 10 10 2007/02/01      M  50      ETV  0   0 10.766667   239   1.3  1.03 1.20
## 11 11 2007/02/01      M  32      ETV  0   0  3.713889   359   0.9  0.97 1.10
## 12 12 2007/02/01      M  48      TDF  0   1  3.269444   486   0.7  1.02 0.81
## 13 13 2007/02/01      M  51      ETV  0   0 10.655556    56   1.3  1.10 1.00
## 14 14 2007/02/02      M  41      ETV  1   0  4.488889    34   1.7  1.08 0.69
## 15 15 2007/02/08      M  50      ETV  1   1  5.869444   107   2.1  1.57 1.00
## 16 16 2007/02/08      M  51      ETV  1   0  4.008333    40   1.3  1.15 0.92
## 17 17 2007/02/09      F  62      TDF  1   0  3.513889    48   1.0  1.02 0.81
## 18 18 2007/02/14      F  51      ETV  1   0  2.372222    61   2.1  1.43 0.60
## 19 19 2007/02/15      M  48      TDF  0   0 10.677778    34   1.1  1.10 1.00
## 20 20 2007/02/15      F  50      ETV  1   0  5.030556    31   0.9  1.25 0.63
##    b_plt b_alb m6_alt m6_bil m6_inr m6_cr m6_plt m6_alb m12_alt m12_bil m12_inr
## 1    110   3.8     35    1.0   1.05   0.9     96    3.6      37     1.2      NA
## 2    187   3.8     16    1.1   0.97   0.7    194    3.8      18     1.1    0.96
## 3    133   3.9    108    1.8   1.03   0.9    120    4.3      27     1.5    1.02
## 4    164   4.2     24    1.7   1.11   0.9    149    4.2      29     1.2    0.98
## 5    153   4.2     57    0.9   1.10   1.0    145    4.6      35     1.2    1.05
## 6    170   3.4     26    0.8   0.97   1.4     NA    3.4      17     0.6    0.91
## 7    162    NA     51    1.0   1.05   0.8    128    3.9      42     0.7    1.04
## 8    165   3.5     24    1.4   1.04   0.8    184    3.7      16     0.9    1.06
## 9    157   3.9     25    1.3   1.02   0.8    163    4.2      24     0.9    1.04
## 10   166   3.2     30    1.4   1.00   1.0    158    3.8      34     1.3    1.03
## 11   202   4.1     23    1.1   0.98   1.0    171    4.4      19     0.9    0.94
## 12   159   2.8     34    0.9   0.96   0.8    215    3.3      24     0.8    0.95
## 13   186   4.2     18    1.3     NA   1.1     NA    4.2      17     1.4    1.01
## 14   159    NA     23    1.6   0.97   0.9    164    3.7      19     1.2    0.97
## 15    66   2.1     42    1.0   1.15   1.0     80    3.7      36     1.1    1.09
## 16   118   3.4     33    1.1   1.09   1.0    129    4.0      30     1.1    0.98
## 17   105   4.0     28    0.9   1.03   0.7    121    3.9      31     0.9    1.06
## 18    55   2.7     25    1.2   1.21   0.7     67    3.5      24     0.7    1.15
## 19    NA   4.3     13    0.8     NA   1.1     NA    4.3      13     1.1    1.02
## 20   163   3.1     29    1.3   1.19   0.7     40    3.3      31     1.4    1.17
##    m12_cr m12_plt m12_alb
## 1     1.0     111     4.0
## 2     0.6     239     4.0
```

```
## 3     0.9     120     4.2
## 4     0.8     185     4.1
## 5     1.0     141     4.5
## 6     1.9     179     3.0
## 7      NA      NA     3.9
## 8     0.7     138     3.4
## 9     0.7     148     3.9
## 10    1.0     151     3.7
## 11    1.1      NA     4.3
## 12    0.8     272     3.4
## 13    1.1     143     4.3
## 14    0.8     175     3.8
## 15    1.1     103     4.2
## 16     NA      NA     4.1
## 17     NA      NA     3.9
## 18    0.6      61     3.5
## 19    1.1     143     4.2
## 20    0.7      36     3.4
```

select기능도 같이 사용해보기

```
dat1 %>%
  filter(id<=20) %>%
  select(id, starts_with('b_'))
```

```
##    id b_alt b_bil b_inr b_cr b_plt b_alb
## 1   1    67   1.2  1.17 1.00   110   3.8
## 2   2    32   1.1  0.90 0.60   187   3.8
## 3   3   106   1.8  1.03 1.05   133   3.9
## 4   4   159   1.4  1.04 0.93   164   4.2
## 5   5    94   0.8  1.12 0.90   153   4.2
## 6   6    32   0.7  0.88 1.20   170   3.4
## 7   7   104   1.1  1.10 0.91   162    NA
## 8   8   143   1.5  1.10 0.70   165   3.5
## 9   9    31   1.3    NA 0.73   157   3.9
## 10 10   239   1.3  1.03 1.20   166   3.2
## 11 11   359   0.9  0.97 1.10   202   4.1
## 12 12   486   0.7  1.02 0.81   159   2.8
## 13 13    56   1.3  1.10 1.00   186   4.2
## 14 14    34   1.7  1.08 0.69   159    NA
## 15 15   107   2.1  1.57 1.00    66   2.1
## 16 16    40   1.3  1.15 0.92   118   3.4
## 17 17    48   1.0  1.02 0.81   105   4.0
## 18 18    61   2.1  1.43 0.60    55   2.7
## 19 19    34   1.1  1.10 1.00    NA   4.3
## 20 20    31   0.9  1.25 0.63   163   3.1
```

남자 환자만 선택

```
dat1 %>%
  filter(gender =='M')
```

```
##    id index_date gender age treat_gr lc hcc    hcc_yr b_alt b_bil b_inr b_cr
## 1   1 2007/01/05      M  54      ETV  1   1  7.641667    67   1.2  1.17 1.00
## 2   3 2007/01/11      M  49      TDF  1   0 10.594444   106   1.8  1.03 1.05
## 3   4 2007/01/12      M  26      ETV  0   0  6.016667   159   1.4  1.04 0.93
## 4   5 2007/01/18      M  50      TDF  1   0  2.355556    94   0.8  1.12 0.90
## 5   6 2007/01/18      M  33      ETV  1   0  6.069444    32   0.7  0.88 1.20
## 6   7 2007/01/26      M  49      TDF  1   1  8.069444   104   1.1  1.10 0.91
## 7  10 2007/02/01      M  50      ETV  0   0 10.766667   239   1.3  1.03 1.20
## 8  11 2007/02/01      M  32      ETV  0   0  3.713889   359   0.9  0.97 1.10
## 9  12 2007/02/01      M  48      TDF  0   1  3.269444   486   0.7  1.02 0.81
## 10 13 2007/02/01      M  51      ETV  0   0 10.655556    56   1.3  1.10 1.00
## 11 14 2007/02/02      M  41      ETV  1   0  4.488889    34   1.7  1.08 0.69
## 12 15 2007/02/08      M  50      ETV  1   1  5.869444   107   2.1  1.57 1.00
## 13 16 2007/02/08      M  51      ETV  1   0  4.008333    40   1.3  1.15 0.92
## 14 19 2007/02/15      M  48      TDF  0   0 10.677778    34   1.1  1.10 1.00
## 15 21 2007/02/15      M  52      TDF  1   0  4.725000    57   1.1  1.11 0.80
## 16 22 2007/02/16      M  30      ETV  0   0 10.761111    81   1.0  0.97 0.99
## 17 23 2007/02/16      M  29      TDF  0   0 10.558333    35   1.9  1.03 0.91
## 18 24 2007/02/20      M  43      ETV  1   0  3.994444    84   1.1  1.07 0.89
##    b_plt b_alb m6_alt m6_bil m6_inr m6_cr m6_plt m6_alb m12_alt m12_bil m12_inr
## 1    110   3.8     35    1.0   1.05   0.9     96    3.6      37     1.2      NA
## 2    133   3.9    108    1.8   1.03   0.9    120    4.3      27     1.5    1.02
## 3    164   4.2     24    1.7   1.11   0.9    149    4.2      29     1.2    0.98
## 4    153   4.2     57    0.9   1.10   1.0    145    4.6      35     1.2    1.05
## 5    170   3.4     26    0.8   0.97   1.4     NA    3.4      17     0.6    0.91
## 6    162    NA     51    1.0   1.05   0.8    128    3.9      42     0.7    1.04
## 7    166   3.2     30    1.4   1.00   1.0    158    3.8      34     1.3    1.03
## 8    202   4.1     23    1.1   0.98   1.0    171    4.4      19     0.9    0.94
## 9    159   2.8     34    0.9   0.96   0.8    215    3.3      24     0.8    0.95
## 10   186   4.2     18    1.3     NA   1.1     NA    4.2      17     1.4    1.01
## 11   159    NA     23    1.6   0.97   0.9    164    3.7      19     1.2    0.97
## 12    66   2.1     42    1.0   1.15   1.0     80    3.7      36     1.1    1.09
## 13   118   3.4     33    1.1   1.09   1.0    129    4.0      30     1.1    0.98
## 14    NA   4.3     13    0.8     NA   1.1     NA    4.3      13     1.1    1.02
## 15    91   4.1     39    1.3   1.11   0.8     74    4.0      34     1.3    1.04
## 16   292   4.1     50    1.0   0.93   0.9    264    4.4      63     0.8    0.93
## 17   204   4.4     26    2.3   1.00   0.9     NA    4.1      28     1.3    0.93
## 18   181   4.2     63    0.8   1.02   0.9    144    4.5      52     0.9    1.02
##    m12_cr m12_plt m12_alb
## 1     1.0     111     4.0
## 2     0.9     120     4.2
## 3     0.8     185     4.1
## 4     1.0     141     4.5
## 5     1.9     179     3.0
## 6      NA      NA     3.9
## 7     1.0     151     3.7
## 8     1.1      NA     4.3
## 9     0.8     272     3.4
## 10    1.1     143     4.3
```

```
## 11    0.8     175     3.8
## 12    1.1     103     4.2
## 13     NA      NA     4.1
## 14    1.1     143     4.2
## 15    0.8      64     3.9
## 16    1.0     286     4.5
## 17    0.9     113     4.0
## 18    1.0     136     4.3
```

복합조건 사용하기

50세 이상의 남자 환자 선택

```
dat1 %>%
  filter(age>=50, gender=='M')
```

```
##   id index_date gender age treat_gr lc hcc    hcc_yr b_alt b_bil b_inr b_cr
## 1  1 2007/01/05      M  54      ETV  1   1  7.641667    67   1.2  1.17 1.00
## 2  5 2007/01/18      M  50      TDF  1   0  2.355556    94   0.8  1.12 0.90
## 3 10 2007/02/01      M  50      ETV  0   0 10.766667   239   1.3  1.03 1.20
## 4 13 2007/02/01      M  51      ETV  0   0 10.655556    56   1.3  1.10 1.00
## 5 15 2007/02/08      M  50      ETV  1   1  5.869444   107   2.1  1.57 1.00
## 6 16 2007/02/08      M  51      ETV  1   0  4.008333    40   1.3  1.15 0.92
## 7 21 2007/02/15      M  52      TDF  1   0  4.725000    57   1.1  1.11 0.80
##   b_plt b_alb m6_alt m6_bil m6_inr m6_cr m6_plt m6_alb m12_alt m12_bil m12_inr
## 1   110   3.8     35    1.0   1.05   0.9     96    3.6      37     1.2      NA
## 2   153   4.2     57    0.9   1.10   1.0    145    4.6      35     1.2    1.05
## 3   166   3.2     30    1.4   1.00   1.0    158    3.8      34     1.3    1.03
## 4   186   4.2     18    1.3     NA   1.1     NA    4.2      17     1.4    1.01
## 5    66   2.1     42    1.0   1.15   1.0     80    3.7      36     1.1    1.09
## 6   118   3.4     33    1.1   1.09   1.0    129    4.0      30     1.1    0.98
## 7    91   4.1     39    1.3   1.11   0.8     74    4.0      34     1.3    1.04
##   m12_cr m12_plt m12_alb
## 1    1.0     111     4.0
## 2    1.0     141     4.5
## 3    1.0     151     3.7
## 4    1.1     143     4.3
## 5    1.1     103     4.2
## 6     NA      NA     4.1
## 7    0.8      64     3.9
```

30세 이상 60세 이하

```
dat1 %>%
  filter(age >=30 & age<=60)
```

```
##    id index_date gender age treat_gr lc hcc    hcc_yr b_alt b_bil b_inr b_cr
## 1   1 2007/01/05      M  54      ETV  1   1  7.641667    67   1.2  1.17 1.00
## 2   2 2007/01/10      F  45      ETV  0   0  9.763889    32   1.1  0.90 0.60
## 3   3 2007/01/11      M  49      TDF  1   0 10.594444   106   1.8  1.03 1.05
## 4   5 2007/01/18      M  50      TDF  1   0  2.355556    94   0.8  1.12 0.90
## 5   6 2007/01/18      M  33      ETV  1   0  6.069444    32   0.7  0.88 1.20
## 6   7 2007/01/26      M  49      TDF  1   1  8.069444   104   1.1  1.10 0.91
## 7   8 2007/01/31      F  50      TDF  0   0 10.716667   143   1.5  1.10 0.70
## 8   9 2007/01/31      F  49      ETV  1   0  2.163889    31   1.3    NA 0.73
## 9  10 2007/02/01      M  50      ETV  0   0 10.766667   239   1.3  1.03 1.20
## 10 11 2007/02/01      M  32      ETV  0   0  3.713889   359   0.9  0.97 1.10
## 11 12 2007/02/01      M  48      TDF  0   1  3.269444   486   0.7  1.02 0.81
## 12 13 2007/02/01      M  51      ETV  0   0 10.655556    56   1.3  1.10 1.00
## 13 14 2007/02/02      M  41      ETV  1   0  4.488889    34   1.7  1.08 0.69
## 14 15 2007/02/08      M  50      ETV  1   1  5.869444   107   2.1  1.57 1.00
## 15 16 2007/02/08      M  51      ETV  1   0  4.008333    40   1.3  1.15 0.92
## 16 18 2007/02/14      F  51      ETV  1   0  2.372222    61   2.1  1.43 0.60
## 17 19 2007/02/15      M  48      TDF  0   0 10.677778    34   1.1  1.10 1.00
## 18 20 2007/02/15      F  50      ETV  1   0  5.030556    31   0.9  1.25 0.63
## 19 21 2007/02/15      M  52      TDF  1   0  4.725000    57   1.1  1.11 0.80
## 20 22 2007/02/16      M  30      ETV  0   0 10.761111    81   1.0  0.97 0.99
## 21 24 2007/02/20      M  43      ETV  1   0  3.994444    84   1.1  1.07 0.89
##    b_plt b_alb m6_alt m6_bil m6_inr m6_cr m6_plt m6_alb m12_alt m12_bil m12_inr
## 1    110   3.8     35    1.0   1.05   0.9     96    3.6      37     1.2      NA
## 2    187   3.8     16    1.1   0.97   0.7    194    3.8      18     1.1    0.96
## 3    133   3.9    108    1.8   1.03   0.9    120    4.3      27     1.5    1.02
## 4    153   4.2     57    0.9   1.10   1.0    145    4.6      35     1.2    1.05
## 5    170   3.4     26    0.8   0.97   1.4     NA    3.4      17     0.6    0.91
## 6    162    NA     51    1.0   1.05   0.8    128    3.9      42     0.7    1.04
## 7    165   3.5     24    1.4   1.04   0.8    184    3.7      16     0.9    1.06
## 8    157   3.9     25    1.3   1.02   0.8    163    4.2      24     0.9    1.04
## 9    166   3.2     30    1.4   1.00   1.0    158    3.8      34     1.3    1.03
## 10   202   4.1     23    1.1   0.98   1.0    171    4.4      19     0.9    0.94
## 11   159   2.8     34    0.9   0.96   0.8    215    3.3      24     0.8    0.95
## 12   186   4.2     18    1.3     NA   1.1     NA    4.2      17     1.4    1.01
## 13   159    NA     23    1.6   0.97   0.9    164    3.7      19     1.2    0.97
## 14    66   2.1     42    1.0   1.15   1.0     80    3.7      36     1.1    1.09
## 15   118   3.4     33    1.1   1.09   1.0    129    4.0      30     1.1    0.98
## 16    55   2.7     25    1.2   1.21   0.7     67    3.5      24     0.7    1.15
## 17    NA   4.3     13    0.8     NA   1.1     NA    4.3      13     1.1    1.02
## 18   163   3.1     29    1.3   1.19   0.7     40    3.3      31     1.4    1.17
## 19    91   4.1     39    1.3   1.11   0.8     74    4.0      34     1.3    1.04
## 20   292   4.1     50    1.0   0.93   0.9    264    4.4      63     0.8    0.93
## 21   181   4.2     63    0.8   1.02   0.9    144    4.5      52     0.9    1.02
##    m12_cr m12_plt m12_alb
## 1     1.0     111     4.0
## 2     0.6     239     4.0
## 3     0.9     120     4.2
## 4     1.0     141     4.5
```

```
## 5     1.9     179     3.0
## 6      NA      NA     3.9
## 7     0.7     138     3.4
## 8     0.7     148     3.9
## 9     1.0     151     3.7
## 10    1.1      NA     4.3
## 11    0.8     272     3.4
## 12    1.1     143     4.3
## 13    0.8     175     3.8
## 14    1.1     103     4.2
## 15     NA      NA     4.1
## 16    0.6      61     3.5
## 17    1.1     143     4.2
## 18    0.7      36     3.4
## 19    0.8      64     3.9
## 20    1.0     286     4.5
## 21    1.0     136     4.3
```

between 사용

```
dat1 %>%
  filter(between (age, 30, 60))
```

```
##    id index_date gender age treat_gr lc hcc    hcc_yr b_alt b_bil b_inr b_cr
## 1   1 2007/01/05      M  54      ETV  1   1  7.641667    67   1.2  1.17 1.00
## 2   2 2007/01/10      F  45      ETV  0   0  9.763889    32   1.1  0.90 0.60
## 3   3 2007/01/11      M  49      TDF  1   0 10.594444   106   1.8  1.03 1.05
## 4   5 2007/01/18      M  50      TDF  1   0  2.355556    94   0.8  1.12 0.90
## 5   6 2007/01/18      M  33      ETV  1   0  6.069444    32   0.7  0.88 1.20
## 6   7 2007/01/26      M  49      TDF  1   1  8.069444   104   1.1  1.10 0.91
## 7   8 2007/01/31      F  50      TDF  0   0 10.716667   143   1.5  1.10 0.70
## 8   9 2007/01/31      F  49      ETV  1   0  2.163889    31   1.3    NA 0.73
## 9  10 2007/02/01      M  50      ETV  0   0 10.766667   239   1.3  1.03 1.20
## 10 11 2007/02/01      M  32      ETV  0   0  3.713889   359   0.9  0.97 1.10
## 11 12 2007/02/01      M  48      TDF  0   1  3.269444   486   0.7  1.02 0.81
## 12 13 2007/02/01      M  51      ETV  0   0 10.655556    56   1.3  1.10 1.00
## 13 14 2007/02/02      M  41      ETV  1   0  4.488889    34   1.7  1.08 0.69
## 14 15 2007/02/08      M  50      ETV  1   1  5.869444   107   2.1  1.57 1.00
## 15 16 2007/02/08      M  51      ETV  1   0  4.008333    40   1.3  1.15 0.92
## 16 18 2007/02/14      F  51      ETV  1   0  2.372222    61   2.1  1.43 0.60
## 17 19 2007/02/15      M  48      TDF  0   0 10.677778    34   1.1  1.10 1.00
## 18 20 2007/02/15      F  50      ETV  1   0  5.030556    31   0.9  1.25 0.63
## 19 21 2007/02/15      M  52      TDF  1   0  4.725000    57   1.1  1.11 0.80
## 20 22 2007/02/16      M  30      ETV  0   0 10.761111    81   1.0  0.97 0.99
## 21 24 2007/02/20      M  43      ETV  1   0  3.994444    84   1.1  1.07 0.89
##    b_plt b_alb m6_alt m6_bil m6_inr m6_cr m6_plt m6_alb m12_alt m12_bil m12_inr
## 1    110   3.8     35    1.0   1.05   0.9     96    3.6      37     1.2      NA
## 2    187   3.8     16    1.1   0.97   0.7    194    3.8      18     1.1    0.96
## 3    133   3.9    108    1.8   1.03   0.9    120    4.3      27     1.5    1.02
## 4    153   4.2     57    0.9   1.10   1.0    145    4.6      35     1.2    1.05
## 5    170   3.4     26    0.8   0.97   1.4     NA    3.4      17     0.6    0.91
## 6    162    NA     51    1.0   1.05   0.8    128    3.9      42     0.7    1.04
## 7    165   3.5     24    1.4   1.04   0.8    184    3.7      16     0.9    1.06
## 8    157   3.9     25    1.3   1.02   0.8    163    4.2      24     0.9    1.04
## 9    166   3.2     30    1.4   1.00   1.0    158    3.8      34     1.3    1.03
## 10   202   4.1     23    1.1   0.98   1.0    171    4.4      19     0.9    0.94
## 11   159   2.8     34    0.9   0.96   0.8    215    3.3      24     0.8    0.95
## 12   186   4.2     18    1.3     NA   1.1     NA    4.2      17     1.4    1.01
## 13   159    NA     23    1.6   0.97   0.9    164    3.7      19     1.2    0.97
## 14    66   2.1     42    1.0   1.15   1.0     80    3.7      36     1.1    1.09
## 15   118   3.4     33    1.1   1.09   1.0    129    4.0      30     1.1    0.98
## 16    55   2.7     25    1.2   1.21   0.7     67    3.5      24     0.7    1.15
## 17    NA   4.3     13    0.8     NA   1.1     NA    4.3      13     1.1    1.02
## 18   163   3.1     29    1.3   1.19   0.7     40    3.3      31     1.4    1.17
## 19    91   4.1     39    1.3   1.11   0.8     74    4.0      34     1.3    1.04
## 20   292   4.1     50    1.0   0.93   0.9    264    4.4      63     0.8    0.93
## 21   181   4.2     63    0.8   1.02   0.9    144    4.5      52     0.9    1.02
##    m12_cr m12_plt m12_alb
## 1     1.0     111     4.0
## 2     0.6     239     4.0
## 3     0.9     120     4.2
## 4     1.0     141     4.5
```

```
## 5     1.9     179     3.0
## 6      NA      NA     3.9
## 7     0.7     138     3.4
## 8     0.7     148     3.9
## 9     1.0     151     3.7
## 10    1.1      NA     4.3
## 11    0.8     272     3.4
## 12    1.1     143     4.3
## 13    0.8     175     3.8
## 14    1.1     103     4.2
## 15     NA      NA     4.1
## 16    0.6      61     3.5
## 17    1.1     143     4.2
## 18    0.7      36     3.4
## 19    0.8      64     3.9
## 20    1.0     286     4.5
## 21    1.0     136     4.3
```

50세 이상이거나 간경변증이 있는 환자

```
dat1 %>%
  filter(age >=50 | lc==1)
```

```
##    id index_date gender age treat_gr lc hcc    hcc_yr b_alt b_bil b_inr b_cr
## 1   1 2007/01/05      M  54      ETV  1   1  7.641667    67   1.2  1.17 1.00
## 2   3 2007/01/11      M  49      TDF  1   0 10.594444   106   1.8  1.03 1.05
## 3   5 2007/01/18      M  50      TDF  1   0  2.355556    94   0.8  1.12 0.90
## 4   6 2007/01/18      M  33      ETV  1   0  6.069444    32   0.7  0.88 1.20
## 5   7 2007/01/26      M  49      TDF  1   1  8.069444   104   1.1  1.10 0.91
## 6   8 2007/01/31      F  50      TDF  0   0 10.716667   143   1.5  1.10 0.70
## 7   9 2007/01/31      F  49      ETV  1   0  2.163889    31   1.3    NA 0.73
## 8  10 2007/02/01      M  50      ETV  0   0 10.766667   239   1.3  1.03 1.20
## 9  13 2007/02/01      M  51      ETV  0   0 10.655556    56   1.3  1.10 1.00
## 10 14 2007/02/02      M  41      ETV  1   0  4.488889    34   1.7  1.08 0.69
## 11 15 2007/02/08      M  50      ETV  1   1  5.869444   107   2.1  1.57 1.00
## 12 16 2007/02/08      M  51      ETV  1   0  4.008333    40   1.3  1.15 0.92
## 13 17 2007/02/09      F  62      TDF  1   0  3.513889    48   1.0  1.02 0.81
## 14 18 2007/02/14      F  51      ETV  1   0  2.372222    61   2.1  1.43 0.60
## 15 20 2007/02/15      F  50      ETV  1   0  5.030556    31   0.9  1.25 0.63
## 16 21 2007/02/15      M  52      TDF  1   0  4.725000    57   1.1  1.11 0.80
## 17 24 2007/02/20      M  43      ETV  1   0  3.994444    84   1.1  1.07 0.89
##    b_plt b_alb m6_alt m6_bil m6_inr m6_cr m6_plt m6_alb m12_alt m12_bil m12_inr
## 1    110   3.8     35    1.0   1.05   0.9     96    3.6      37     1.2      NA
## 2    133   3.9    108    1.8   1.03   0.9    120    4.3      27     1.5    1.02
## 3    153   4.2     57    0.9   1.10   1.0    145    4.6      35     1.2    1.05
## 4    170   3.4     26    0.8   0.97   1.4     NA    3.4      17     0.6    0.91
## 5    162    NA     51    1.0   1.05   0.8    128    3.9      42     0.7    1.04
## 6    165   3.5     24    1.4   1.04   0.8    184    3.7      16     0.9    1.06
## 7    157   3.9     25    1.3   1.02   0.8    163    4.2      24     0.9    1.04
## 8    166   3.2     30    1.4   1.00   1.0    158    3.8      34     1.3    1.03
## 9    186   4.2     18    1.3     NA   1.1     NA    4.2      17     1.4    1.01
## 10   159    NA     23    1.6   0.97   0.9    164    3.7      19     1.2    0.97
## 11    66   2.1     42    1.0   1.15   1.0     80    3.7      36     1.1    1.09
## 12   118   3.4     33    1.1   1.09   1.0    129    4.0      30     1.1    0.98
## 13   105   4.0     28    0.9   1.03   0.7    121    3.9      31     0.9    1.06
## 14    55   2.7     25    1.2   1.21   0.7     67    3.5      24     0.7    1.15
## 15   163   3.1     29    1.3   1.19   0.7     40    3.3      31     1.4    1.17
## 16    91   4.1     39    1.3   1.11   0.8     74    4.0      34     1.3    1.04
## 17   181   4.2     63    0.8   1.02   0.9    144    4.5      52     0.9    1.02
##    m12_cr m12_plt m12_alb
## 1     1.0     111     4.0
## 2     0.9     120     4.2
## 3     1.0     141     4.5
## 4     1.9     179     3.0
## 5      NA      NA     3.9
## 6     0.7     138     3.4
## 7     0.7     148     3.9
## 8     1.0     151     3.7
## 9     1.1     143     4.3
## 10    0.8     175     3.8
## 11    1.1     103     4.2
## 12     NA      NA     4.1
```

```
## 13     NA      NA     3.9
## 14    0.6      61     3.5
## 15    0.7      36     3.4
## 16    0.8      64     3.9
## 17    1.0     136     4.3
```

and와 or을 같이 사용

```
dat1 %>%
  filter( (age>=60 & gender=="M") | (age<=50 & gender=="F"))
```

```
##   id index_date gender age treat_gr lc hcc    hcc_yr b_alt b_bil b_inr b_cr
## 1  2 2007/01/10      F  45      ETV  0   0  9.763889    32   1.1  0.90 0.60
## 2  8 2007/01/31      F  50      TDF  0   0 10.716667   143   1.5  1.10 0.70
## 3  9 2007/01/31      F  49      ETV  1   0  2.163889    31   1.3    NA 0.73
## 4 20 2007/02/15      F  50      ETV  1   0  5.030556    31   0.9  1.25 0.63
##   b_plt b_alb m6_alt m6_bil m6_inr m6_cr m6_plt m6_alb m12_alt m12_bil m12_inr
## 1   187   3.8     16    1.1   0.97   0.7    194    3.8      18     1.1    0.96
## 2   165   3.5     24    1.4   1.04   0.8    184    3.7      16     0.9    1.06
## 3   157   3.9     25    1.3   1.02   0.8    163    4.2      24     0.9    1.04
## 4   163   3.1     29    1.3   1.19   0.7     40    3.3      31     1.4    1.17
##   m12_cr m12_plt m12_alb
## 1    0.6     239     4.0
## 2    0.7     138     3.4
## 3    0.7     148     3.9
## 4    0.7      36     3.4
```

조건을 만족하지 않는 경우 골라내기

여자 환자만

```
dat1 %>%
  filter(gender!="M")
```

```
##   id index_date gender age treat_gr lc hcc    hcc_yr b_alt b_bil b_inr b_cr
## 1  2 2007/01/10      F  45      ETV  0   0  9.763889    32   1.1  0.90 0.60
## 2  8 2007/01/31      F  50      TDF  0   0 10.716667   143   1.5  1.10 0.70
## 3  9 2007/01/31      F  49      ETV  1   0  2.163889    31   1.3    NA 0.73
## 4 17 2007/02/09      F  62      TDF  1   0  3.513889    48   1.0  1.02 0.81
## 5 18 2007/02/14      F  51      ETV  1   0  2.372222    61   2.1  1.43 0.60
## 6 20 2007/02/15      F  50      ETV  1   0  5.030556    31   0.9  1.25 0.63
##   b_plt b_alb m6_alt m6_bil m6_inr m6_cr m6_plt m6_alb m12_alt m12_bil m12_inr
## 1   187   3.8     16    1.1   0.97   0.7    194    3.8      18     1.1    0.96
## 2   165   3.5     24    1.4   1.04   0.8    184    3.7      16     0.9    1.06
## 3   157   3.9     25    1.3   1.02   0.8    163    4.2      24     0.9    1.04
## 4   105   4.0     28    0.9   1.03   0.7    121    3.9      31     0.9    1.06
## 5    55   2.7     25    1.2   1.21   0.7     67    3.5      24     0.7    1.15
## 6   163   3.1     29    1.3   1.19   0.7     40    3.3      31     1.4    1.17
##   m12_cr m12_plt m12_alb
## 1    0.6     239     4.0
## 2    0.7     138     3.4
## 3    0.7     148     3.9
## 4     NA      NA     3.9
## 5    0.6      61     3.5
## 6    0.7      36     3.4
```

간경변증이 없는 50세 이상의 환자

```
dat1 %>%
  filter(lc !=1 & age >=50)
```

```
##   id index_date gender age treat_gr lc hcc   hcc_yr b_alt b_bil b_inr b_cr
## 1  8 2007/01/31      F  50      TDF  0   0 10.71667   143   1.5  1.10  0.7
## 2 10 2007/02/01      M  50      ETV  0   0 10.76667   239   1.3  1.03  1.2
## 3 13 2007/02/01      M  51      ETV  0   0 10.65556    56   1.3  1.10  1.0
##   b_plt b_alb m6_alt m6_bil m6_inr m6_cr m6_plt m6_alb m12_alt m12_bil m12_inr
## 1   165   3.5     24    1.4   1.04   0.8    184    3.7      16     0.9    1.06
## 2   166   3.2     30    1.4   1.00   1.0    158    3.8      34     1.3    1.03
## 3   186   4.2     18    1.3     NA   1.1     NA    4.2      17     1.4    1.01
##   m12_cr m12_plt m12_alb
## 1    0.7     138     3.4
## 2    1.0     151     3.7
## 3    1.1     143     4.3
```

count기능

성별에 따라 몇 명?

```
dat1 %>%
  count(gender)
```

```
##   gender  n
## 1      F  6
## 2      M 18
```

남성이면서 간암이 있는 환자는?

```
dat1 %>%
  filter(gender == 'M' & hcc ==1) %>%
  count()
```

```
##   n
## 1 4
```

남성이면서 간경변증이 있는 환자 중 hcc가 발생한 환자는?

```
dat1 %>%
  filter(gender == 'M' & lc ==1) %>%
  count(hcc)
```

```
##   hcc n
## 1   0 7
## 2   1 3
```

남성이면서 간경변증이 있는 환자 중 hcc가 발생한 환자는 항바이러스 종류에 따라 몇 명?

```
dat1 %>%
  filter(gender == 'M' & lc ==1) %>%
  count(treat_gr, hcc)
```

```
##   treat_gr hcc n
## 1      ETV   0 4
## 2      ETV   1 2
## 3      TDF   0 3
## 4      TDF   1 1
```

결측값 다루기

결측값

```
dat1 %>%
  filter(is.na(b_inr)) %>%
  count()
```

```
##   n
## 1 1
```

결측값이 없는 경우

```
dat1 %>%
  filter(!is.na(b_inr))
```

```
##    id index_date gender age treat_gr lc hcc    hcc_yr b_alt b_bil b_inr b_cr
## 1   1 2007/01/05      M  54      ETV  1   1  7.641667    67   1.2  1.17 1.00
## 2   2 2007/01/10      F  45      ETV  0   0  9.763889    32   1.1  0.90 0.60
## 3   3 2007/01/11      M  49      TDF  1   0 10.594444   106   1.8  1.03 1.05
## 4   4 2007/01/12      M  26      ETV  0   0  6.016667   159   1.4  1.04 0.93
## 5   5 2007/01/18      M  50      TDF  1   0  2.355556    94   0.8  1.12 0.90
## 6   6 2007/01/18      M  33      ETV  1   0  6.069444    32   0.7  0.88 1.20
## 7   7 2007/01/26      M  49      TDF  1   1  8.069444   104   1.1  1.10 0.91
## 8   8 2007/01/31      F  50      TDF  0   0 10.716667   143   1.5  1.10 0.70
## 9  10 2007/02/01      M  50      ETV  0   0 10.766667   239   1.3  1.03 1.20
## 10 11 2007/02/01      M  32      ETV  0   0  3.713889   359   0.9  0.97 1.10
## 11 12 2007/02/01      M  48      TDF  0   1  3.269444   486   0.7  1.02 0.81
## 12 13 2007/02/01      M  51      ETV  0   0 10.655556    56   1.3  1.10 1.00
## 13 14 2007/02/02      M  41      ETV  1   0  4.488889    34   1.7  1.08 0.69
## 14 15 2007/02/08      M  50      ETV  1   1  5.869444   107   2.1  1.57 1.00
## 15 16 2007/02/08      M  51      ETV  1   0  4.008333    40   1.3  1.15 0.92
## 16 17 2007/02/09      F  62      TDF  1   0  3.513889    48   1.0  1.02 0.81
## 17 18 2007/02/14      F  51      ETV  1   0  2.372222    61   2.1  1.43 0.60
## 18 19 2007/02/15      M  48      TDF  0   0 10.677778    34   1.1  1.10 1.00
## 19 20 2007/02/15      F  50      ETV  1   0  5.030556    31   0.9  1.25 0.63
## 20 21 2007/02/15      M  52      TDF  1   0  4.725000    57   1.1  1.11 0.80
## 21 22 2007/02/16      M  30      ETV  0   0 10.761111    81   1.0  0.97 0.99
## 22 23 2007/02/16      M  29      TDF  0   0 10.558333    35   1.9  1.03 0.91
## 23 24 2007/02/20      M  43      ETV  1   0  3.994444    84   1.1  1.07 0.89
##    b_plt b_alb m6_alt m6_bil m6_inr m6_cr m6_plt m6_alb m12_alt m12_bil m12_inr
## 1    110   3.8     35    1.0   1.05   0.9     96    3.6      37     1.2      NA
## 2    187   3.8     16    1.1   0.97   0.7    194    3.8      18     1.1    0.96
## 3    133   3.9    108    1.8   1.03   0.9    120    4.3      27     1.5    1.02
## 4    164   4.2     24    1.7   1.11   0.9    149    4.2      29     1.2    0.98
## 5    153   4.2     57    0.9   1.10   1.0    145    4.6      35     1.2    1.05
## 6    170   3.4     26    0.8   0.97   1.4     NA    3.4      17     0.6    0.91
## 7    162    NA     51    1.0   1.05   0.8    128    3.9      42     0.7    1.04
## 8    165   3.5     24    1.4   1.04   0.8    184    3.7      16     0.9    1.06
## 9    166   3.2     30    1.4   1.00   1.0    158    3.8      34     1.3    1.03
## 10   202   4.1     23    1.1   0.98   1.0    171    4.4      19     0.9    0.94
## 11   159   2.8     34    0.9   0.96   0.8    215    3.3      24     0.8    0.95
## 12   186   4.2     18    1.3     NA   1.1     NA    4.2      17     1.4    1.01
```

```
## 13   159    NA     23    1.6   0.97   0.9    164    3.7      19     1.2    0.97
## 14    66   2.1     42    1.0   1.15   1.0     80    3.7      36     1.1    1.09
## 15   118   3.4     33    1.1   1.09   1.0    129    4.0      30     1.1    0.98
## 16   105   4.0     28    0.9   1.03   0.7    121    3.9      31     0.9    1.06
## 17    55   2.7     25    1.2   1.21   0.7     67    3.5      24     0.7    1.15
## 18    NA   4.3     13    0.8     NA   1.1     NA    4.3      13     1.1    1.02
## 19   163   3.1     29    1.3   1.19   0.7     40    3.3      31     1.4    1.17
## 20    91   4.1     39    1.3   1.11   0.8     74    4.0      34     1.3    1.04
## 21   292   4.1     50    1.0   0.93   0.9    264    4.4      63     0.8    0.93
## 22   204   4.4     26    2.3   1.00   0.9     NA    4.1      28     1.3    0.93
## 23   181   4.2     63    0.8   1.02   0.9    144    4.5      52     0.9    1.02
##    m12_cr m12_plt m12_alb
## 1     1.0     111     4.0
## 2     0.6     239     4.0
## 3     0.9     120     4.2
## 4     0.8     185     4.1
## 5     1.0     141     4.5
## 6     1.9     179     3.0
## 7      NA      NA     3.9
## 8     0.7     138     3.4
## 9     1.0     151     3.7
## 10    1.1      NA     4.3
## 11    0.8     272     3.4
## 12    1.1     143     4.3
## 13    0.8     175     3.8
## 14    1.1     103     4.2
## 15     NA      NA     4.1
## 16     NA      NA     3.9
## 17    0.6      61     3.5
## 18    1.1     143     4.2
## 19    0.7      36     3.4
## 20    0.8      64     3.9
## 21    1.0     286     4.5
## 22    0.9     113     4.0
## 23    1.0     136     4.3
```

여러 변수에서 동시에 확인

```
dat1 %>%
  filter(!is.na(b_inr),
         !is.na(b_alt),
         !is.na(b_plt)) %>%
  count()
```

```
##    n
## 1 22
```

결측값이 존재하지 않는 케이스만 남기기

특정 변수에서 결측값이 없는 환자

```
dat1 %>%
  drop_na(b_inr, b_alt, b_plt) %>%
  count()
```

```
##    n
## 1 22
```

모든 변수에서 결측값이 없는 환자

```
dat1 %>%
  drop_na() %>%
  count()
```

```
##    n
## 1 13
```

모든 변수에서 결측값이 없는 환자

```
na.count <- apply(dat1, 2, function(x)sum(is.na(x)))
na.count
```

```
##         id index_date     gender        age   treat_gr         lc        hcc
##          0          0          0          0          0          0          0
##     hcc_yr      b_alt      b_bil      b_inr      b_cr      b_plt      b_alb
##          0          0          0          1          0          1          2
##     m6_alt     m6_bil     m6_inr     m6_cr     m6_plt     m6_alb    m12_alt
##          0          0          2          0          4          0          0
##    m12_bil    m12_inr     m12_cr    m12_plt    m12_alb
##          0          1          3          4          0
```

4. Mutate

기존 변수를 이용해서 새로운 변수 만들기

```
dat1 %>%
  mutate(alt_plt = b_alt / b_plt) %>%
  select(b_alt, b_plt, alt_plt)
```

```
##    b_alt b_plt   alt_plt
## 1     67   110 0.6090909
## 2     32   187 0.1711230
## 3    106   133 0.7969925
## 4    159   164 0.9695122
## 5     94   153 0.6143791
## 6     32   170 0.1882353
## 7    104   162 0.6419753
## 8    143   165 0.8666667
## 9     31   157 0.1974522
## 10   239   166 1.4397590
## 11   359   202 1.7772277
## 12   486   159 3.0566038
## 13    56   186 0.3010753
## 14    34   159 0.2138365
## 15   107    66 1.6212121
## 16    40   118 0.3389831
## 17    48   105 0.4571429
## 18    61    55 1.1090909
## 19    34    NA        NA
## 20    31   163 0.1901840
## 21    57    91 0.6263736
## 22    81   292 0.2773973
## 23    35   204 0.1715686
## 24    84   181 0.4640884
```

결측값이 있는 경우 새로운 변수 값이 어떻게 나오는지 확인

```
dat1 %>%
  mutate(alt_plt = b_alt /b_plt) %>%
  select(id, b_alt, b_plt, alt_plt) %>%
  filter(is.na(b_alt) | is.na(b_plt))
```

```
##   id b_alt b_plt alt_plt
## 1 19    34    NA      NA
```

결측값이 있으면 계산이 불가능하므로 결측값이 있는 경우 제거

```
dat1 %>%
  drop_na(b_alt, b_plt) %>%
  mutate(alt_plt = b_alt /b_plt) %>%
  select(id, b_alt, b_plt, alt_plt)
```

```
##    id b_alt b_plt   alt_plt
## 1   1    67   110 0.6090909
## 2   2    32   187 0.1711230
## 3   3   106   133 0.7969925
## 4   4   159   164 0.9695122
## 5   5    94   153 0.6143791
## 6   6    32   170 0.1882353
## 7   7   104   162 0.6419753
## 8   8   143   165 0.8666667
## 9   9    31   157 0.1974522
## 10 10   239   166 1.4397590
## 11 11   359   202 1.7772277
## 12 12   486   159 3.0566038
## 13 13    56   186 0.3010753
## 14 14    34   159 0.2138365
## 15 15   107    66 1.6212121
## 16 16    40   118 0.3389831
## 17 17    48   105 0.4571429
## 18 18    61    55 1.1090909
## 19 20    31   163 0.1901840
## 20 21    57    91 0.6263736
## 21 22    81   292 0.2773973
## 22 23    35   204 0.1715686
## 23 24    84   181 0.4640884
```

조건을 이용해서 새로운 변수 만들기

```
dat1 %>%
  mutate(age_gr=ifelse(age >=50,'above_50','below_50')) %>%
  select(id, age, age_gr)
```

```
##    id age   age_gr
## 1   1  54 above_50
## 2   2  45 below_50
## 3   3  49 below_50
## 4   4  26 below_50
## 5   5  50 above_50
## 6   6  33 below_50
## 7   7  49 below_50
## 8   8  50 above_50
## 9   9  49 below_50
## 10 10  50 above_50
## 11 11  32 below_50
## 12 12  48 below_50
## 13 13  51 above_50
## 14 14  41 below_50
## 15 15  50 above_50
## 16 16  51 above_50
## 17 17  62 above_50
## 18 18  51 above_50
## 19 19  48 below_50
## 20 20  50 above_50
## 21 21  52 above_50
## 22 22  30 below_50
## 23 23  29 below_50
## 24 24  43 below_50
```

```
dat1 %>%
  mutate(age_gr=ifelse(age >=50,'above_50','below_50'))%>%
  count(age_gr)
```

```
##     age_gr  n
## 1 above_50 11
## 2 below_50 13
```

```
dat1 %>%
  mutate(bil_gr=ifelse(b_bil<2,'A',ifelse(b_bil<3,'B','C'))) %>%
  select(id, b_bil, bil_gr)
```

```
##    id b_bil bil_gr
## 1   1   1.2      A
## 2   2   1.1      A
## 3   3   1.8      A
## 4   4   1.4      A
## 5   5   0.8      A
## 6   6   0.7      A
## 7   7   1.1      A
## 8   8   1.5      A
## 9   9   1.3      A
## 10 10   1.3      A
## 11 11   0.9      A
## 12 12   0.7      A
## 13 13   1.3      A
## 14 14   1.7      A
## 15 15   2.1      B
## 16 16   1.3      A
## 17 17   1.0      A
## 18 18   2.1      B
## 19 19   1.1      A
## 20 20   0.9      A
## 21 21   1.1      A
## 22 22   1.0      A
## 23 23   1.9      A
## 24 24   1.1      A
```

```
dat1 %>%
  mutate(bil_gr=ifelse(b_bil<2,'A',
                       ifelse(b_bil<3,'B','C'))) %>%
  count(bil_gr)
```

```
##   bil_gr  n
## 1      A 22
## 2      B  2
```

```
dat1 %>%
  mutate(risk_gr=ifelse(age>=50 & lc==1,'high_risk',
                        ifelse(age<50 & lc==0,
                        'low_risk','intermediate_risk')))
```

```
##    id index_date gender age treat_gr lc hcc    hcc_yr b_alt b_bil b_inr b_cr
## 1   1 2007/01/05      M  54      ETV  1   1  7.641667    67   1.2  1.17 1.00
## 2   2 2007/01/10      F  45      ETV  0   0  9.763889    32   1.1  0.90 0.60
## 3   3 2007/01/11      M  49      TDF  1   0 10.594444   106   1.8  1.03 1.05
## 4   4 2007/01/12      M  26      ETV  0   0  6.016667   159   1.4  1.04 0.93
## 5   5 2007/01/18      M  50      TDF  1   0  2.355556    94   0.8  1.12 0.90
## 6   6 2007/01/18      M  33      ETV  1   0  6.069444    32   0.7  0.88 1.20
## 7   7 2007/01/26      M  49      TDF  1   1  8.069444   104   1.1  1.10 0.91
## 8   8 2007/01/31      F  50      TDF  0   0 10.716667   143   1.5  1.10 0.70
## 9   9 2007/01/31      F  49      ETV  1   0  2.163889    31   1.3    NA 0.73
## 10 10 2007/02/01      M  50      ETV  0   0 10.766667   239   1.3  1.03 1.20
## 11 11 2007/02/01      M  32      ETV  0   0  3.713889   359   0.9  0.97 1.10
## 12 12 2007/02/01      M  48      TDF  0   1  3.269444   486   0.7  1.02 0.81
## 13 13 2007/02/01      M  51      ETV  0   0 10.655556    56   1.3  1.10 1.00
## 14 14 2007/02/02      M  41      ETV  1   0  4.488889    34   1.7  1.08 0.69
## 15 15 2007/02/08      M  50      ETV  1   1  5.869444   107   2.1  1.57 1.00
## 16 16 2007/02/08      M  51      ETV  1   0  4.008333    40   1.3  1.15 0.92
## 17 17 2007/02/09      F  62      TDF  1   0  3.513889    48   1.0  1.02 0.81
## 18 18 2007/02/14      F  51      ETV  1   0  2.372222    61   2.1  1.43 0.60
## 19 19 2007/02/15      M  48      TDF  0   0 10.677778    34   1.1  1.10 1.00
## 20 20 2007/02/15      F  50      ETV  1   0  5.030556    31   0.9  1.25 0.63
## 21 21 2007/02/15      M  52      TDF  1   0  4.725000    57   1.1  1.11 0.80
## 22 22 2007/02/16      M  30      ETV  0   0 10.761111    81   1.0  0.97 0.99
## 23 23 2007/02/16      M  29      TDF  0   0 10.558333    35   1.9  1.03 0.91
## 24 24 2007/02/20      M  43      ETV  1   0  3.994444    84   1.1  1.07 0.89
##    b_plt b_alb m6_alt m6_bil m6_inr m6_cr m6_plt m6_alb m12_alt m12_bil m12_inr
## 1    110   3.8     35    1.0   1.05   0.9     96    3.6      37     1.2      NA
## 2    187   3.8     16    1.1   0.97   0.7    194    3.8      18     1.1    0.96
## 3    133   3.9    108    1.8   1.03   0.9    120    4.3      27     1.5    1.02
## 4    164   4.2     24    1.7   1.11   0.9    149    4.2      29     1.2    0.98
## 5    153   4.2     57    0.9   1.10   1.0    145    4.6      35     1.2    1.05
## 6    170   3.4     26    0.8   0.97   1.4     NA    3.4      17     0.6    0.91
## 7    162    NA     51    1.0   1.05   0.8    128    3.9      42     0.7    1.04
## 8    165   3.5     24    1.4   1.04   0.8    184    3.7      16     0.9    1.06
## 9    157   3.9     25    1.3   1.02   0.8    163    4.2      24     0.9    1.04
## 10   166   3.2     30    1.4   1.00   1.0    158    3.8      34     1.3    1.03
## 11   202   4.1     23    1.1   0.98   1.0    171    4.4      19     0.9    0.94
## 12   159   2.8     34    0.9   0.96   0.8    215    3.3      24     0.8    0.95
## 13   186   4.2     18    1.3     NA   1.1     NA    4.2      17     1.4    1.01
## 14   159    NA     23    1.6   0.97   0.9    164    3.7      19     1.2    0.97
## 15    66   2.1     42    1.0   1.15   1.0     80    3.7      36     1.1    1.09
## 16   118   3.4     33    1.1   1.09   1.0    129    4.0      30     1.1    0.98
## 17   105   4.0     28    0.9   1.03   0.7    121    3.9      31     0.9    1.06
## 18    55   2.7     25    1.2   1.21   0.7     67    3.5      24     0.7    1.15
## 19    NA   4.3     13    0.8     NA   1.1     NA    4.3      13     1.1    1.02
## 20   163   3.1     29    1.3   1.19   0.7     40    3.3      31     1.4    1.17
## 21    91   4.1     39    1.3   1.11   0.8     74    4.0      34     1.3    1.04
## 22   292   4.1     50    1.0   0.93   0.9    264    4.4      63     0.8    0.93
```

```
## 23   204   4.4     26    2.3   1.00   0.9     NA    4.1      28     1.3    0.93
## 24   181   4.2     63    0.8   1.02   0.9    144    4.5      52     0.9    1.02
##    m12_cr m12_plt m12_alb           risk_gr
## 1     1.0     111     4.0         high_risk
## 2     0.6     239     4.0          low_risk
## 3     0.9     120     4.2 intermediate_risk
## 4     0.8     185     4.1          low_risk
## 5     1.0     141     4.5         high_risk
## 6     1.9     179     3.0 intermediate_risk
## 7      NA      NA     3.9 intermediate_risk
## 8     0.7     138     3.4 intermediate_risk
## 9     0.7     148     3.9 intermediate_risk
## 10    1.0     151     3.7 intermediate_risk
## 11    1.1      NA     4.3          low_risk
## 12    0.8     272     3.4          low_risk
## 13    1.1     143     4.3 intermediate_risk
## 14    0.8     175     3.8 intermediate_risk
## 15    1.1     103     4.2         high_risk
## 16     NA      NA     4.1         high_risk
## 17     NA      NA     3.9         high_risk
## 18    0.6      61     3.5         high_risk
## 19    1.1     143     4.2          low_risk
## 20    0.7      36     3.4         high_risk
## 21    0.8      64     3.9         high_risk
## 22    1.0     286     4.5          low_risk
## 23    0.9     113     4.0          low_risk
## 24    1.0     136     4.3 intermediate_risk
```

```
dat1 %>%
  mutate(risk_gr=ifelse(age>=50 & lc==1,'high_risk',
                        ifelse(age<50 & lc==0,
                        'low_risk','intermediate_risk'))) %>%
  count(risk_gr)
```

```
##             risk_gr n
## 1         high_risk 8
## 2 intermediate_risk 9
## 3          low_risk 7
```

새로운 변수 만들고 나머지 변수 제거

```
dat1 %>%
  transmute(risk_gr=ifelse(age>=50 & lc==1,'high_risk',
                           ifelse(age<50 & lc==0, 'low_risk', 'intermediate_risk')))
```

```
##              risk_gr
## 1          high_risk
## 2           low_risk
## 3  intermediate_risk
## 4           low_risk
## 5          high_risk
## 6  intermediate_risk
## 7  intermediate_risk
## 8  intermediate_risk
## 9  intermediate_risk
## 10 intermediate_risk
## 11          low_risk
## 12          low_risk
## 13 intermediate_risk
## 14 intermediate_risk
## 15         high_risk
## 16         high_risk
## 17         high_risk
## 18         high_risk
## 19          low_risk
## 20         high_risk
## 21         high_risk
## 22          low_risk
## 23          low_risk
## 24 intermediate_risk
```

변수값의 순서를 새로운 변수로 만들기

```
dat1 %>%
  mutate(age_rank = min_rank(age)) %>%
  select(id, age, age_rank)
```

```
##    id age age_rank
## 1   1  54       23
## 2   2  45        8
## 3   3  49       11
## 4   4  26        1
## 5   5  50       14
## 6   6  33        5
## 7   7  49       11
## 8   8  50       14
## 9   9  49       11
## 10 10  50       14
## 11 11  32        4
## 12 12  48        9
## 13 13  51       19
## 14 14  41        6
## 15 15  50       14
## 16 16  51       19
## 17 17  62       24
## 18 18  51       19
## 19 19  48        9
## 20 20  50       14
## 21 21  52       22
## 22 22  30        3
## 23 23  29        2
## 24 24  43        7
```

내림차순

```
dat1 %>%
  mutate(age_rank = min_rank(desc(age))) %>%
  select(id, age, age_rank)
```

```
##    id age age_rank
## 1   1  54        2
## 2   2  45       17
## 3   3  49       12
## 4   4  26       24
## 5   5  50        7
## 6   6  33       20
## 7   7  49       12
## 8   8  50        7
## 9   9  49       12
## 10 10  50        7
## 11 11  32       21
## 12 12  48       15
## 13 13  51        4
## 14 14  41       19
## 15 15  50        7
## 16 16  51        4
## 17 17  62        1
## 18 18  51        4
## 19 19  48       15
## 20 20  50        7
## 21 21  52        3
## 22 22  30       22
## 23 23  29       23
## 24 24  43       18
```

순서를 정할 때 중간에 gap없이 새로운 변수 만들기

min_rank와 차이는 1등이 2명인 경우 3등이 아니라 2등으로 나옴

```
dat1 %>%
  mutate(age_rank = dense_rank(age)) %>%
  select(id, age, age_rank)
```

```
##    id age age_rank
## 1   1  54       14
## 2   2  45        8
## 3   3  49       10
## 4   4  26        1
## 5   5  50       11
## 6   6  33        5
## 7   7  49       10
## 8   8  50       11
## 9   9  49       10
## 10 10  50       11
## 11 11  32        4
## 12 12  48        9
## 13 13  51       12
## 14 14  41        6
## 15 15  50       11
## 16 16  51       12
## 17 17  62       15
## 18 18  51       12
## 19 19  48        9
## 20 20  50       11
## 21 21  52       13
## 22 22  30        3
## 23 23  29        2
## 24 24  43        7
```

퍼센트 순위로 새로운 변수 만들기

```
dat1 %>%
  mutate(age_rank = percent_rank(age)) %>%
  select(id, age, age_rank) %>%
  arrange(age_rank)
```

```
##    id age   age_rank
## 1   4  26 0.00000000
## 2  23  29 0.04347826
## 3  22  30 0.08695652
## 4  11  32 0.13043478
## 5   6  33 0.17391304
## 6  14  41 0.21739130
## 7  24  43 0.26086957
## 8   2  45 0.30434783
## 9  12  48 0.34782609
## 10 19  48 0.34782609
## 11  3  49 0.43478261
## 12  7  49 0.43478261
## 13  9  49 0.43478261
## 14  5  50 0.56521739
## 15  8  50 0.56521739
## 16 10  50 0.56521739
## 17 15  50 0.56521739
## 18 20  50 0.56521739
## 19 13  51 0.78260870
## 20 16  51 0.78260870
## 21 18  51 0.78260870
## 22 21  52 0.91304348
## 23  1  54 0.95652174
## 24 17  62 1.00000000
```

누적합계를 변수로 만들기

```
dat1 %>%
  mutate(id_sum = cumsum(id)) %>%
  select(id, id_sum)
```

```
##    id id_sum
## 1   1      1
## 2   2      3
## 3   3      6
## 4   4     10
## 5   5     15
## 6   6     21
## 7   7     28
## 8   8     36
## 9   9     45
## 10 10     55
## 11 11     66
## 12 12     78
## 13 13     91
## 14 14    105
## 15 15    120
## 16 16    136
## 17 17    153
## 18 18    171
## 19 19    190
## 20 20    210
## 21 21    231
## 22 22    253
## 23 23    276
## 24 24    300
```

동시에 여러 개의 새로운 변수 만들기

```
dat1 %>%
  mutate(age_gr=ifelse(age>=50,'above 50','below 50')) %>%
  mutate(bil_gr=ifelse(b_bil<2,'A',
                       ifelse(b_bil>3,'C','B'))) %>%
  select(age, age_gr, b_bil, bil_gr)
```

```
##    age   age_gr b_bil bil_gr
## 1   54 above 50   1.2      A
## 2   45 below 50   1.1      A
## 3   49 below 50   1.8      A
## 4   26 below 50   1.4      A
## 5   50 above 50   0.8      A
## 6   33 below 50   0.7      A
## 7   49 below 50   1.1      A
## 8   50 above 50   1.5      A
## 9   49 below 50   1.3      A
## 10  50 above 50   1.3      A
## 11  32 below 50   0.9      A
## 12  48 below 50   0.7      A
## 13  51 above 50   1.3      A
## 14  41 below 50   1.7      A
## 15  50 above 50   2.1      B
## 16  51 above 50   1.3      A
## 17  62 above 50   1.0      A
## 18  51 above 50   2.1      B
## 19  48 below 50   1.1      A
## 20  50 above 50   0.9      A
## 21  52 above 50   1.1      A
## 22  30 below 50   1.0      A
## 23  29 below 50   1.9      A
## 24  43 below 50   1.1      A
```

연속형 변수를 일정 범위마다 그룹화하기

```
dat1 %>%
  mutate(age_gr=ifelse(age>=60,'>=60',
                       ifelse(age>=50,'>=50',
                              ifelse(age>=40,'>=40',
                                     ifelse(age>=30,'>=30','<30'))))) %>%
  select(age, age_gr)
```

```
##    age age_gr
## 1   54   >=50
## 2   45   >=40
## 3   49   >=40
## 4   26    <30
## 5   50   >=50
## 6   33   >=30
## 7   49   >=40
## 8   50   >=50
## 9   49   >=40
## 10  50   >=50
## 11  32   >=30
## 12  48   >=40
## 13  51   >=50
## 14  41   >=40
## 15  50   >=50
## 16  51   >=50
## 17  62   >=60
## 18  51   >=50
## 19  48   >=40
## 20  50   >=50
## 21  52   >=50
## 22  30   >=30
## 23  29    <30
## 24  43   >=40
```

간단하게

```
dat1 %>%
  mutate(age_gr=cut(age,
                    c(-Inf,30,40,50,60,Inf),
                    c('<30','>=30','>=40','>=50','>=60'))) %>%
  select(age, age_gr)
```

```
##    age age_gr
## 1   54   >=50
## 2   45   >=40
## 3   49   >=40
## 4   26    <30
## 5   50   >=40
## 6   33   >=30
## 7   49   >=40
## 8   50   >=40
## 9   49   >=40
## 10  50   >=40
## 11  32   >=30
## 12  48   >=40
## 13  51   >=50
## 14  41   >=40
## 15  50   >=40
## 16  51   >=50
## 17  62   >=60
## 18  51   >=50
## 19  48   >=40
## 20  50   >=40
## 21  52   >=50
## 22  30    <30
## 23  29    <30
## 24  43   >=40
```

10세 차이를 기준으로 분류하기

```
dat1 %>%
  mutate(age_gr=cut_width(age, width=10)) %>%
  select(age, age_gr)
```

```
##    age  age_gr
## 1   54 (45,55]
## 2   45 (35,45]
## 3   49 (45,55]
## 4   26 [25,35]
## 5   50 (45,55]
## 6   33 [25,35]
## 7   49 (45,55]
## 8   50 (45,55]
## 9   49 (45,55]
## 10  50 (45,55]
## 11  32 [25,35]
## 12  48 (45,55]
## 13  51 (45,55]
## 14  41 (35,45]
## 15  50 (45,55]
## 16  51 (45,55]
## 17  62 (55,65]
## 18  51 (45,55]
## 19  48 (45,55]
## 20  50 (45,55]
## 21  52 (45,55]
## 22  30 [25,35]
## 23  29 [25,35]
## 24  43 (35,45]
```

구간 범위에 옵션주기

```
dat1 %>%
  mutate(age_gr=cut_width(age, width=10, boundary=0)) %>%
  mutate(age_gr2=cut_width(age, width=10, boundary=9)) %>%
  select(age, age_gr, age_gr2)
```

```
##    age  age_gr age_gr2
## 1   54 (50,60] (49,59]
## 2   45 (40,50] (39,49]
## 3   49 (40,50] (39,49]
## 4   26 [20,30] [19,29]
## 5   50 (40,50] (49,59]
## 6   33 (30,40] (29,39]
## 7   49 (40,50] (39,49]
## 8   50 (40,50] (49,59]
## 9   49 (40,50] (39,49]
## 10  50 (40,50] (49,59]
## 11  32 (30,40] (29,39]
## 12  48 (40,50] (39,49]
## 13  51 (50,60] (49,59]
## 14  41 (40,50] (39,49]
## 15  50 (40,50] (49,59]
## 16  51 (50,60] (49,59]
## 17  62 (60,70] (59,69]
## 18  51 (50,60] (49,59]
## 19  48 (40,50] (39,49]
## 20  50 (40,50] (49,59]
## 21  52 (50,60] (49,59]
## 22  30 [20,30] (29,39]
## 23  29 [20,30] [19,29]
## 24  43 (40,50] (39,49]
```

연속형 변수를 균일한 범위를 가지는 그룹으로 나누기

```
dat1 %>%
  mutate(age_gr=cut_interval(age, n=4)) %>%
  select(age, age_gr) %>%
  count(age_gr)
```

```
##    age_gr  n
## 1 [26,35]  5
## 2 (35,44]  2
## 3 (44,53] 15
## 4 (53,62]  2
```

연속형 변수를 원하는 개수의 그룹으로 나누기

```
dat1 %>%
  mutate(age_gr = cut_number(age, n=5)) %>%
  select(age, age_gr)
```

```
##    age      age_gr
## 1   54     (51,62]
## 2   45 (37.8,48.2]
## 3   49   (48.2,50]
## 4   26   [26,37.8]
## 5   50   (48.2,50]
## 6   33   [26,37.8]
## 7   49   (48.2,50]
## 8   50   (48.2,50]
## 9   49   (48.2,50]
## 10  50   (48.2,50]
## 11  32   [26,37.8]
## 12  48 (37.8,48.2]
## 13  51     (50,51]
## 14  41 (37.8,48.2]
## 15  50   (48.2,50]
## 16  51     (50,51]
## 17  62     (51,62]
## 18  51     (50,51]
## 19  48 (37.8,48.2]
## 20  50   (48.2,50]
## 21  52     (51,62]
## 22  30   [26,37.8]
## 23  29   [26,37.8]
## 24  43 (37.8,48.2]
```

```
dat1 %>%
  mutate(age_gr = cut_number(age, n=5)) %>%
  select(age, age_gr) %>%
  count(age_gr)
```

```
##        age_gr n
## 1   [26,37.8] 5
## 2 (37.8,48.2] 5
## 3   (48.2,50] 8
## 4     (50,51] 3
## 5     (51,62] 3
```

5. Arrange

오름차순

```
dat1 %>%
  arrange(age)
```

```
##    id index_date gender age treat_gr lc hcc    hcc_yr b_alt b_bil b_inr b_cr
## 1   4 2007/01/12      M  26      ETV  0   0  6.016667   159   1.4  1.04 0.93
## 2  23 2007/02/16      M  29      TDF  0   0 10.558333    35   1.9  1.03 0.91
## 3  22 2007/02/16      M  30      ETV  0   0 10.761111    81   1.0  0.97 0.99
## 4  11 2007/02/01      M  32      ETV  0   0  3.713889   359   0.9  0.97 1.10
## 5   6 2007/01/18      M  33      ETV  1   0  6.069444    32   0.7  0.88 1.20
## 6  14 2007/02/02      M  41      ETV  1   0  4.488889    34   1.7  1.08 0.69
## 7  24 2007/02/20      M  43      ETV  1   0  3.994444    84   1.1  1.07 0.89
## 8   2 2007/01/10      F  45      ETV  0   0  9.763889    32   1.1  0.90 0.60
## 9  12 2007/02/01      M  48      TDF  0   1  3.269444   486   0.7  1.02 0.81
## 10 19 2007/02/15      M  48      TDF  0   0 10.677778    34   1.1  1.10 1.00
## 11  3 2007/01/11      M  49      TDF  1   0 10.594444   106   1.8  1.03 1.05
## 12  7 2007/01/26      M  49      TDF  1   1  8.069444   104   1.1  1.10 0.91
## 13  9 2007/01/31      F  49      ETV  1   0  2.163889    31   1.3    NA 0.73
## 14  5 2007/01/18      M  50      TDF  1   0  2.355556    94   0.8  1.12 0.90
## 15  8 2007/01/31      F  50      TDF  0   0 10.716667   143   1.5  1.10 0.70
## 16 10 2007/02/01      M  50      ETV  0   0 10.766667   239   1.3  1.03 1.20
## 17 15 2007/02/08      M  50      ETV  1   1  5.869444   107   2.1  1.57 1.00
## 18 20 2007/02/15      F  50      ETV  1   0  5.030556    31   0.9  1.25 0.63
## 19 13 2007/02/01      M  51      ETV  0   0 10.655556    56   1.3  1.10 1.00
## 20 16 2007/02/08      M  51      ETV  1   0  4.008333    40   1.3  1.15 0.92
## 21 18 2007/02/14      F  51      ETV  1   0  2.372222    61   2.1  1.43 0.60
## 22 21 2007/02/15      M  52      TDF  1   0  4.725000    57   1.1  1.11 0.80
## 23  1 2007/01/05      M  54      ETV  1   1  7.641667    67   1.2  1.17 1.00
## 24 17 2007/02/09      F  62      TDF  1   0  3.513889    48   1.0  1.02 0.81
##    b_plt b_alb m6_alt m6_bil m6_inr m6_cr m6_plt m6_alb m12_alt m12_bil m12_inr
## 1    164   4.2     24    1.7   1.11   0.9    149    4.2      29     1.2    0.98
## 2    204   4.4     26    2.3   1.00   0.9     NA    4.1      28     1.3    0.93
## 3    292   4.1     50    1.0   0.93   0.9    264    4.4      63     0.8    0.93
## 4    202   4.1     23    1.1   0.98   1.0    171    4.4      19     0.9    0.94
## 5    170   3.4     26    0.8   0.97   1.4     NA    3.4      17     0.6    0.91
## 6    159    NA     23    1.6   0.97   0.9    164    3.7      19     1.2    0.97
## 7    181   4.2     63    0.8   1.02   0.9    144    4.5      52     0.9    1.02
## 8    187   3.8     16    1.1   0.97   0.7    194    3.8      18     1.1    0.96
## 9    159   2.8     34    0.9   0.96   0.8    215    3.3      24     0.8    0.95
## 10    NA   4.3     13    0.8     NA   1.1     NA    4.3      13     1.1    1.02
## 11   133   3.9    108    1.8   1.03   0.9    120    4.3      27     1.5    1.02
## 12   162    NA     51    1.0   1.05   0.8    128    3.9      42     0.7    1.04
## 13   157   3.9     25    1.3   1.02   0.8    163    4.2      24     0.9    1.04
## 14   153   4.2     57    0.9   1.10   1.0    145    4.6      35     1.2    1.05
## 15   165   3.5     24    1.4   1.04   0.8    184    3.7      16     0.9    1.06
## 16   166   3.2     30    1.4   1.00   1.0    158    3.8      34     1.3    1.03
## 17    66   2.1     42    1.0   1.15   1.0     80    3.7      36     1.1    1.09
## 18   163   3.1     29    1.3   1.19   0.7     40    3.3      31     1.4    1.17
## 19   186   4.2     18    1.3     NA   1.1     NA    4.2      17     1.4    1.01
```

```
## 20   118   3.4     33    1.1   1.09   1.0    129    4.0      30     1.1    0.98
## 21    55   2.7     25    1.2   1.21   0.7     67    3.5      24     0.7    1.15
## 22    91   4.1     39    1.3   1.11   0.8     74    4.0      34     1.3    1.04
## 23   110   3.8     35    1.0   1.05   0.9     96    3.6      37     1.2      NA
## 24   105   4.0     28    0.9   1.03   0.7    121    3.9      31     0.9    1.06
##    m12_cr m12_plt m12_alb
## 1     0.8     185     4.1
## 2     0.9     113     4.0
## 3     1.0     286     4.5
## 4     1.1      NA     4.3
## 5     1.9     179     3.0
## 6     0.8     175     3.8
## 7     1.0     136     4.3
## 8     0.6     239     4.0
## 9     0.8     272     3.4
## 10    1.1     143     4.2
## 11    0.9     120     4.2
## 12     NA      NA     3.9
## 13    0.7     148     3.9
## 14    1.0     141     4.5
## 15    0.7     138     3.4
## 16    1.0     151     3.7
## 17    1.1     103     4.2
## 18    0.7      36     3.4
## 19    1.1     143     4.3
## 20     NA      NA     4.1
## 21    0.6      61     3.5
## 22    0.8      64     3.9
## 23    1.0     111     4.0
## 24     NA      NA     3.9
```

내림차순

```
dat1 %>%
  arrange(desc(age))
```

```
##    id index_date gender age treat_gr lc hcc    hcc_yr b_alt b_bil b_inr b_cr
## 1  17 2007/02/09      F  62      TDF  1   0  3.513889    48   1.0  1.02 0.81
## 2   1 2007/01/05      M  54      ETV  1   1  7.641667    67   1.2  1.17 1.00
## 3  21 2007/02/15      M  52      TDF  1   0  4.725000    57   1.1  1.11 0.80
## 4  13 2007/02/01      M  51      ETV  0   0 10.655556    56   1.3  1.10 1.00
## 5  16 2007/02/08      M  51      ETV  1   0  4.008333    40   1.3  1.15 0.92
## 6  18 2007/02/14      F  51      ETV  1   0  2.372222    61   2.1  1.43 0.60
## 7   5 2007/01/18      M  50      TDF  1   0  2.355556    94   0.8  1.12 0.90
## 8   8 2007/01/31      F  50      TDF  0   0 10.716667   143   1.5  1.10 0.70
## 9  10 2007/02/01      M  50      ETV  0   0 10.766667   239   1.3  1.03 1.20
## 10 15 2007/02/08      M  50      ETV  1   1  5.869444   107   2.1  1.57 1.00
## 11 20 2007/02/15      F  50      ETV  1   0  5.030556    31   0.9  1.25 0.63
## 12  3 2007/01/11      M  49      TDF  1   0 10.594444   106   1.8  1.03 1.05
## 13  7 2007/01/26      M  49      TDF  1   1  8.069444   104   1.1  1.10 0.91
## 14  9 2007/01/31      F  49      ETV  1   0  2.163889    31   1.3    NA 0.73
## 15 12 2007/02/01      M  48      TDF  0   1  3.269444   486   0.7  1.02 0.81
## 16 19 2007/02/15      M  48      TDF  0   0 10.677778    34   1.1  1.10 1.00
## 17  2 2007/01/10      F  45      ETV  0   0  9.763889    32   1.1  0.90 0.60
## 18 24 2007/02/20      M  43      ETV  1   0  3.994444    84   1.1  1.07 0.89
## 19 14 2007/02/02      M  41      ETV  1   0  4.488889    34   1.7  1.08 0.69
## 20  6 2007/01/18      M  33      ETV  1   0  6.069444    32   0.7  0.88 1.20
## 21 11 2007/02/01      M  32      ETV  0   0  3.713889   359   0.9  0.97 1.10
## 22 22 2007/02/16      M  30      ETV  0   0 10.761111    81   1.0  0.97 0.99
## 23 23 2007/02/16      M  29      TDF  0   0 10.558333    35   1.9  1.03 0.91
## 24  4 2007/01/12      M  26      ETV  0   0  6.016667   159   1.4  1.04 0.93
##    b_plt b_alb m6_alt m6_bil m6_inr m6_cr m6_plt m6_alb m12_alt m12_bil m12_inr
## 1    105   4.0     28    0.9   1.03   0.7    121    3.9      31     0.9    1.06
## 2    110   3.8     35    1.0   1.05   0.9     96    3.6      37     1.2      NA
## 3     91   4.1     39    1.3   1.11   0.8     74    4.0      34     1.3    1.04
## 4    186   4.2     18    1.3     NA   1.1     NA    4.2      17     1.4    1.01
## 5    118   3.4     33    1.1   1.09   1.0    129    4.0      30     1.1    0.98
## 6     55   2.7     25    1.2   1.21   0.7     67    3.5      24     0.7    1.15
## 7    153   4.2     57    0.9   1.10   1.0    145    4.6      35     1.2    1.05
## 8    165   3.5     24    1.4   1.04   0.8    184    3.7      16     0.9    1.06
## 9    166   3.2     30    1.4   1.00   1.0    158    3.8      34     1.3    1.03
## 10    66   2.1     42    1.0   1.15   1.0     80    3.7      36     1.1    1.09
## 11   163   3.1     29    1.3   1.19   0.7     40    3.3      31     1.4    1.17
## 12   133   3.9    108    1.8   1.03   0.9    120    4.3      27     1.5    1.02
## 13   162    NA     51    1.0   1.05   0.8    128    3.9      42     0.7    1.04
## 14   157   3.9     25    1.3   1.02   0.8    163    4.2      24     0.9    1.04
## 15   159   2.8     34    0.9   0.96   0.8    215    3.3      24     0.8    0.95
## 16    NA   4.3     13    0.8     NA   1.1     NA    4.3      13     1.1    1.02
## 17   187   3.8     16    1.1   0.97   0.7    194    3.8      18     1.1    0.96
## 18   181   4.2     63    0.8   1.02   0.9    144    4.5      52     0.9    1.02
## 19   159    NA     23    1.6   0.97   0.9    164    3.7      19     1.2    0.97
## 20   170   3.4     26    0.8   0.97   1.4     NA    3.4      17     0.6    0.91
## 21   202   4.1     23    1.1   0.98   1.0    171    4.4      19     0.9    0.94
## 22   292   4.1     50    1.0   0.93   0.9    264    4.4      63     0.8    0.93
## 23   204   4.4     26    2.3   1.00   0.9     NA    4.1      28     1.3    0.93
```

```
## 24   164   4.2     24    1.7   1.11   0.9    149    4.2      29     1.2    0.98
##    m12_cr m12_plt m12_alb
## 1      NA      NA     3.9
## 2     1.0     111     4.0
## 3     0.8      64     3.9
## 4     1.1     143     4.3
## 5      NA      NA     4.1
## 6     0.6      61     3.5
## 7     1.0     141     4.5
## 8     0.7     138     3.4
## 9     1.0     151     3.7
## 10    1.1     103     4.2
## 11    0.7      36     3.4
## 12    0.9     120     4.2
## 13     NA      NA     3.9
## 14    0.7     148     3.9
## 15    0.8     272     3.4
## 16    1.1     143     4.2
## 17    0.6     239     4.0
## 18    1.0     136     4.3
## 19    0.8     175     3.8
## 20    1.9     179     3.0
## 21    1.1      NA     4.3
## 22    1.0     286     4.5
## 23    0.9     113     4.0
## 24    0.8     185     4.1
```

여러 개 변수를 오름차순

```
dat1 %>%
  arrange(age, b_alt) %>%
  select(age, b_alt)
```

```
##    age b_alt
## 1   26   159
## 2   29    35
## 3   30    81
## 4   32   359
## 5   33    32
## 6   41    34
## 7   43    84
## 8   45    32
## 9   48    34
## 10  48   486
## 11  49    31
## 12  49   104
## 13  49   106
## 14  50    31
## 15  50    94
## 16  50   107
## 17  50   143
## 18  50   239
## 19  51    40
## 20  51    56
## 21  51    61
## 22  52    57
## 23  54    67
## 24  62    48
```

6. Summarise

원하는 변수의 평균 구하기

```
dat1 %>%
  summarise(mean_age = mean(age))
```

```
##   mean_age
## 1 45.54167
```

여러 개의 요약정보 동시에 구하기

```
dat1 %>%
  summarise(mean_age = mean(age),
            median_age = median(age),
            iqr_age = IQR(age))
```

```
##   mean_age median_age iqr_age
## 1 45.54167         49    7.75
```

고유값 개수 확인

```
dat1 %>%
  summarise(n_distinct(age))
```

```
##   n_distinct(age)
## 1              15
```

```
unique(dat1$age)
```

```
##  [1] 54 45 49 26 50 33 32 48 51 41 62 52 30 29 43
```

```
length(unique(dat1$age))
```

```
## [1] 15
```

발생률 구하기

```
dat1 %>%
  summarise( patient_number = n(),
             event = sum(hcc),
             person_year = sum(hcc_yr),
             incidence_rate = event / person_year)
```

```
##   patient_number event person_year incidence_rate
## 1             24     4    157.7972     0.02534899
```

7. Group_by

원하는 변수로 나누어 각각 값 계산

```
dat1 %>%
  group_by(gender) %>%
  summarise(mean_age = mean(age))
```

```
## # A tibble: 2 x 2
##   gender mean_age
##   <chr>     <dbl>
## 1 F          51.2
## 2 M          43.7
```

해당 조건을 만족하는 환자의 숫자를 변수로 만들기

```
dat1 %>%
  group_by(gender, lc) %>%
  summarise(patient_no = n(),
            mean_age = mean(age))
```

```
## `summarise()` has grouped output by 'gender'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 4 x 4
## # Groups:   gender [2]
##   gender    lc patient_no mean_age
##   <chr>  <int>      <int>    <dbl>
## 1 F          0          2     47.5
## 2 F          1          4     53
## 3 M          0          8     39.2
## 4 M          1         10     47.2
```

```
dat1 %>%
  group_by(gender, lc) %>%
  summarise(patient_no = n(),
            mean_inr = mean(b_inr))
```

```
## `summarise()` has grouped output by 'gender'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 4 x 4
## # Groups:   gender [2]
##   gender    lc patient_no mean_inr
##   <chr>  <int>      <int>    <dbl>
## 1 F          0          2     1
## 2 F          1          4    NA
## 3 M          0          8     1.03
## 4 M          1         10     1.13
```

```
dat1 %>%
  group_by(gender, lc) %>%
  summarise(patient_no = n(),
            mean_inr = mean(b_inr, na.rm=T))
```

```
## `summarise()` has grouped output by 'gender'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 4 x 4
## # Groups:   gender [2]
##   gender    lc patient_no mean_inr
##   <chr>  <int>      <int>    <dbl>
## 1 F          0          2     1
## 2 F          1          4     1.23
## 3 M          0          8     1.03
## 4 M          1         10     1.13
```

summary table 만들기

```
dat1 %>%
  group_by(gender, lc) %>%
  summarise( N_alt = sum(!is.na(b_alt)),
             MEAN_alt = mean(b_alt, na.rm=T),
             MEDIAN_alt = median(b_alt, na.rm=T),
             MIN_alt = min(b_alt, na.rm=T),
             MAX_alt = max(b_alt, na.rm=T))
```

```
## `summarise()` has grouped output by 'gender'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 4 x 7
## # Groups:   gender [2]
##   gender    lc N_alt MEAN_alt MEDIAN_alt MIN_alt MAX_alt
##   <chr>  <int> <int>    <dbl>      <dbl>   <int>   <int>
## 1 F          0     2     87.5       87.5      32     143
## 2 F          1     4     42.8       39.5      31      61
## 3 M          0     8    181.       120        34     486
## 4 M          1    10     72.5       75.5      32     107
```