EDA

```
categorical_columns = c('d_ira', 'cor', 'has', 'dm', 'icc',
                        'asma_dpoc', 'hiv', 'ca_ativo', 'avc',
                        'dvp', 'drogadicao', 'ventilacao_mecanica',
                        'diureticos', 'vasopressina', 'israa',
                        'causa_ira', 'kdigo', 'criterio_ira',
                        'pos_operatorio', 'mortalidade_intra_hospitalar')

daily_columns = c('diurese',
                  'bh',
                  'sofa',
                  'su',
                  'scr',
                  'sna',
                  'sk',
                  'sosm',
                  'uu',
                  'ucr',
                  'una',
                  'uk',
                  'volume_urinario',
                  'uvu24h',
                  'feu',
                  'uu_ucr',
                  'una_ucr',
                  'uvna24h',
                  'uosm',
                  'uosm_estimada',
                  'una+uk',
                  'uu_su',
                  'fena',
                  'fek',
                  'su_scr',
                  'delta_scr')
```

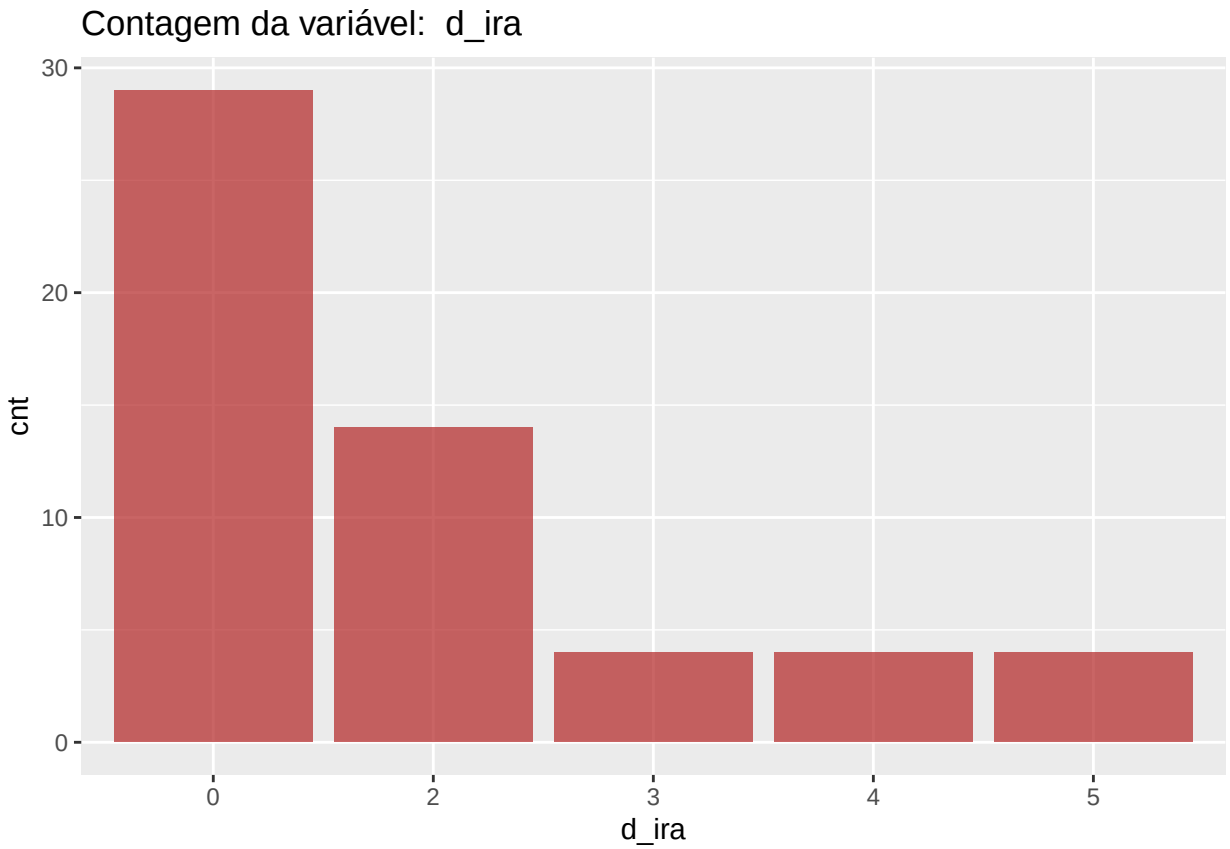
```
df = readRDS('./data/processed_data_mean.rds') %>%
    mutate_at(categorical_columns, list(~factor(.)))

columns = colnames(df)

head(df)
```

```
## # A tibble: 6 x 59
##   numero d_ira idade cor    peso altura   imc has   dm    icc   asma_dpoc hiv
##    <dbl> <fct> <dbl> <fct> <dbl>  <dbl> <dbl> <fct> <fct> <fct> <fct>     <fct>
## 1      4 2        51 1      58.5   1.75  19.1 0     0     0     0         0
## 2      7 2        82 1      72     1.75  23.5 1     0     0     0         0
## 3     23 2        50 1     110     1.7   38.1 1     1     1     0         0
```

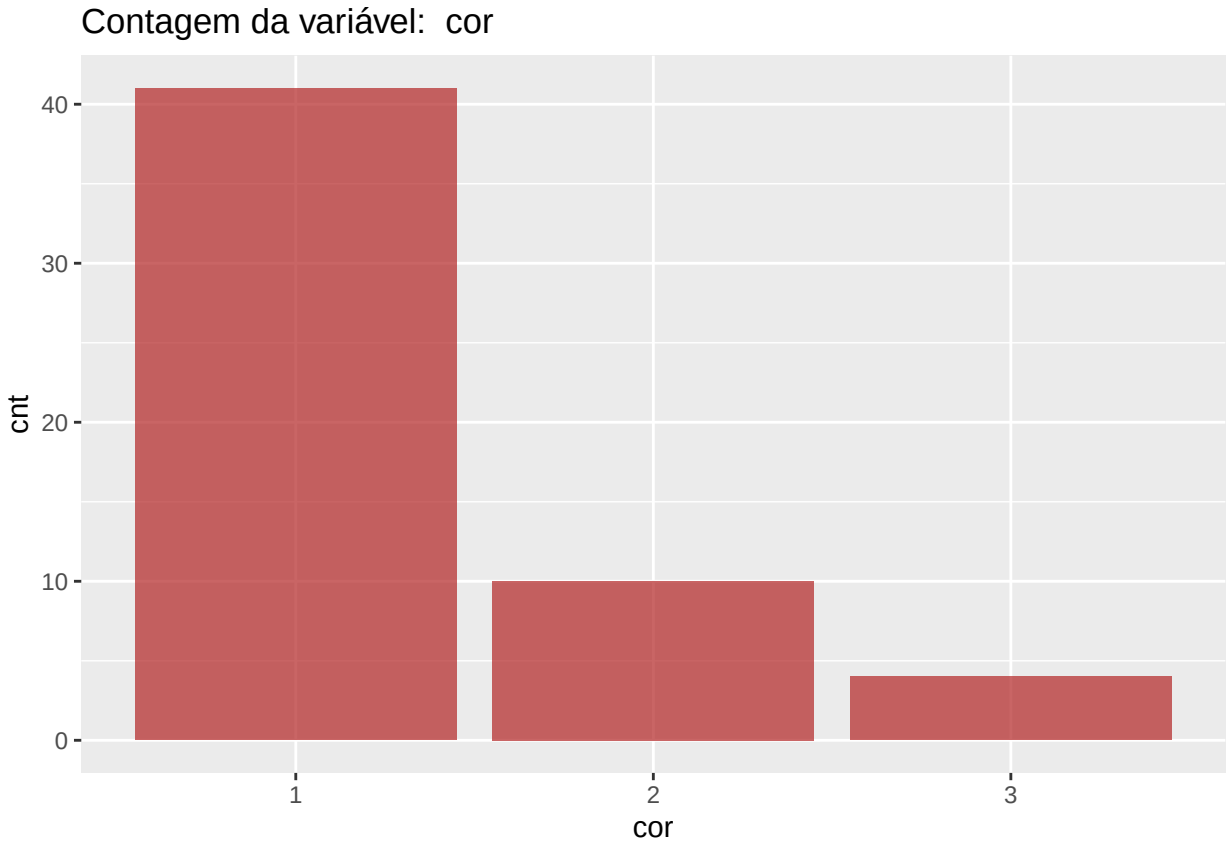
```
## 4     26 2        79 1      46     1.55  19.1 0     0     0     0         0
## 5     31 2        26 1      60     1.75  19.6 0     0     0     0         0
## 6     46 2        54 2      55     1.65  20.2 1     0     0     0         0
## # ... with 47 more variables: ca_ativo <fct>, avc <fct>, dvp <fct>,
## #   drogadicao <fct>, score_clinico <dbl>, ventilacao_mecanica <fct>,
## #   dva_mcg_kg_min <dbl>, diureticos <fct>, vasopressina <fct>, israa <fct>,
## #   saps3 <dbl>, egfr_basal <dbl>, scr_basal <dbl>, causa_ira <fct>,
## #   kdigo <fct>, criterio_ira <fct>, pos_operatorio <fct>,
## #   mortalidade_intra_hospitalar <fct>, tempo_internacao_hospitalar <dbl>,
## #   ira <fct>, kdigo_agrupado <chr>, diurese <dbl>, bh <dbl>, sofa <dbl>,
## #   su <dbl>, scr <dbl>, sna <dbl>, sk <dbl>, sosm <dbl>, uu <dbl>, ucr <dbl>,
## #   una <dbl>, uk <dbl>, volume_urinario <dbl>, uvu24h <dbl>, feu <dbl>,
## #   uu_ucr <dbl>, una_ucr <dbl>, uvna24h <dbl>, uosm <dbl>,
## #   uosm_estimada <dbl>, `una+uk` <dbl>, uu_su <dbl>, fena <dbl>, fek <dbl>,
## #   su_scr <dbl>, delta_scr <dbl>
```

```
for(column in categorical_columns){
    p = df %>%
        group_by(!!sym(column)) %>%
        summarise(cnt = n()) %>%
        ggplot(aes(x=!!sym(column), y=cnt)) +
            geom_col(fill='firebrick', alpha=0.7) +
            labs(title=paste('Contagem da variável: ', column))
    print(p)
}
```

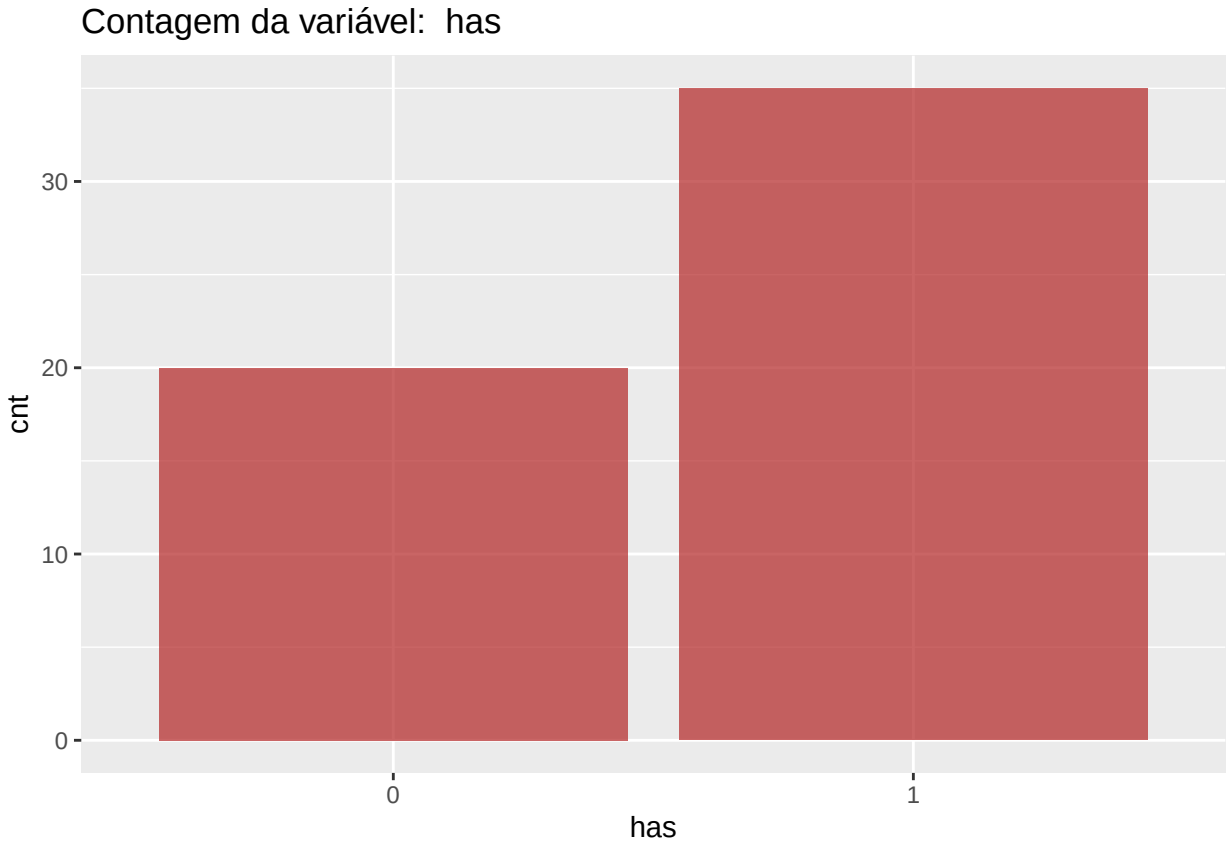
```
## `summarise()` ungrouping output (override with `.groups` argument)
## `summarise()` ungrouping output (override with `.groups` argument)
```



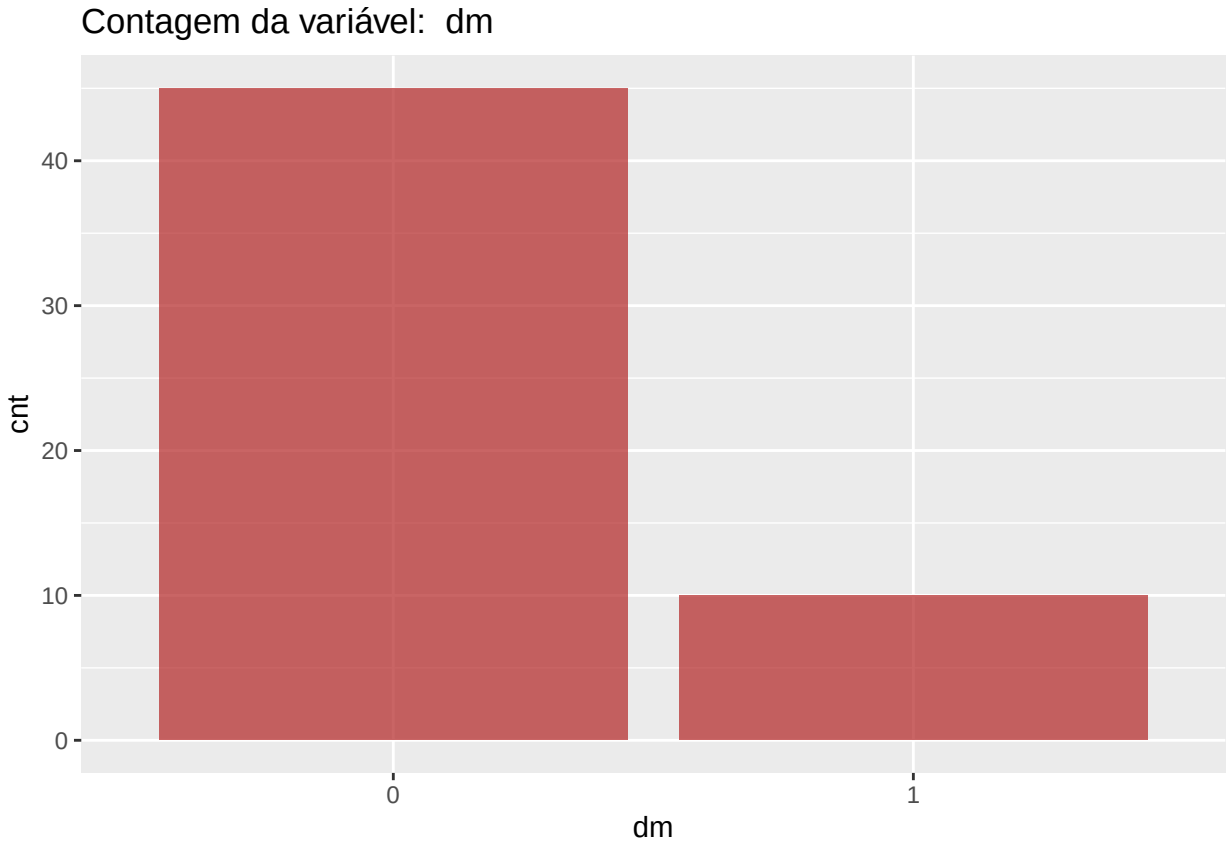

```
## `summarise()` ungrouping output (override with `.groups` argument)
```



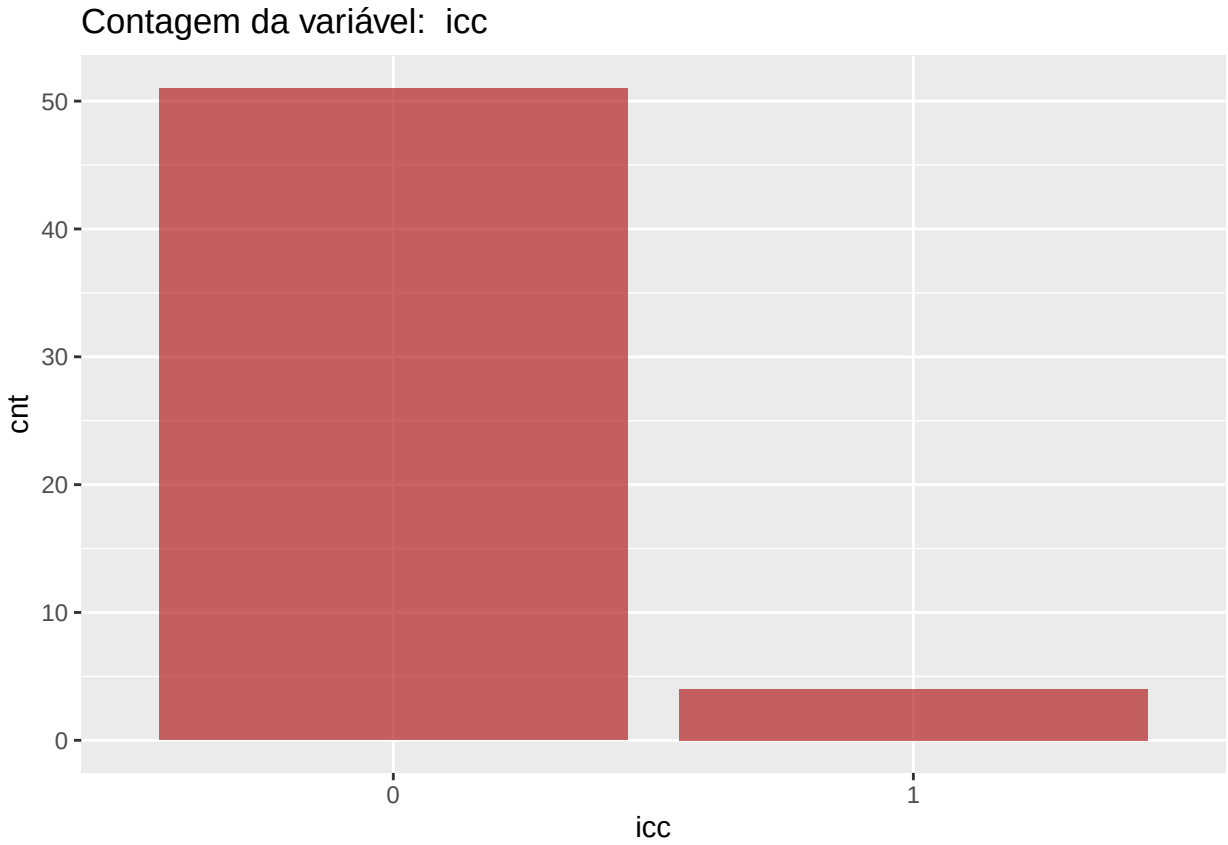

```
## `summarise()` ungrouping output (override with `.groups` argument)
```



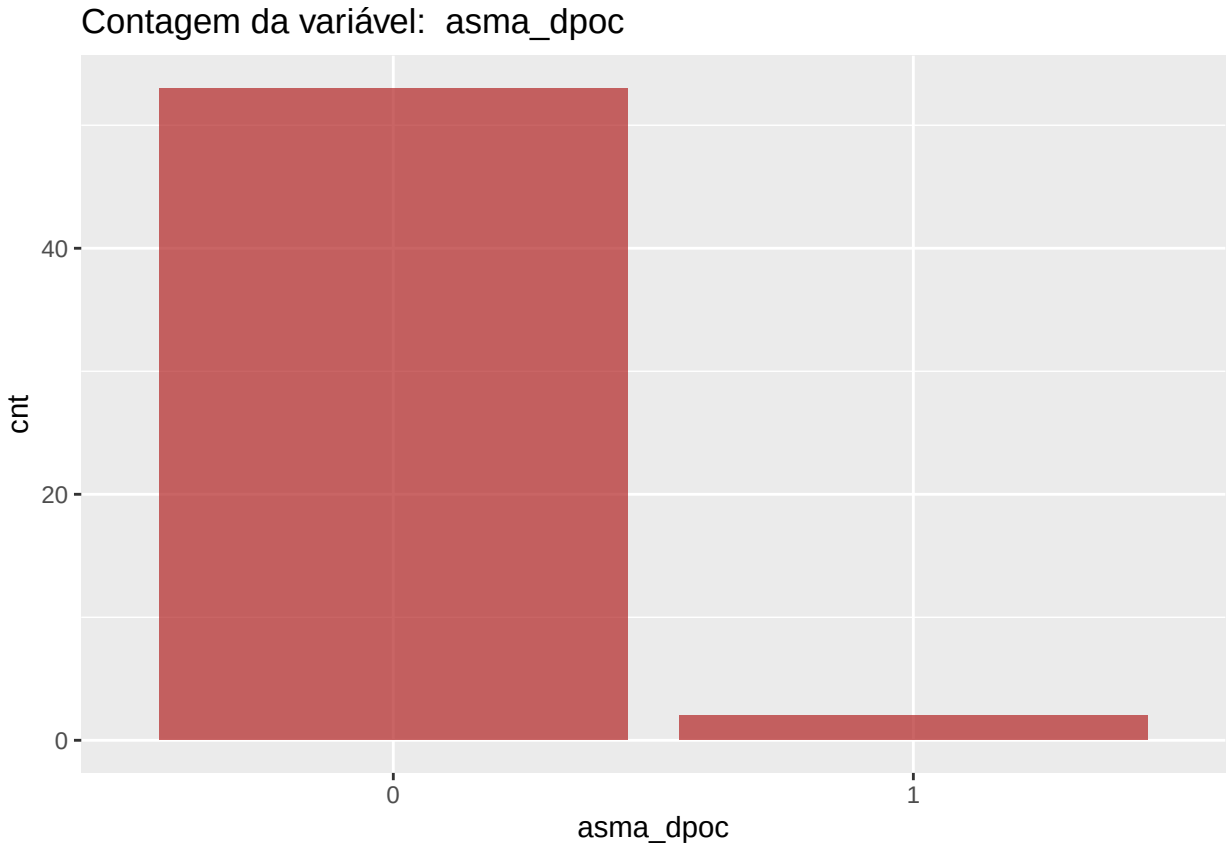

```
## `summarise()` ungrouping output (override with `.groups` argument)
```



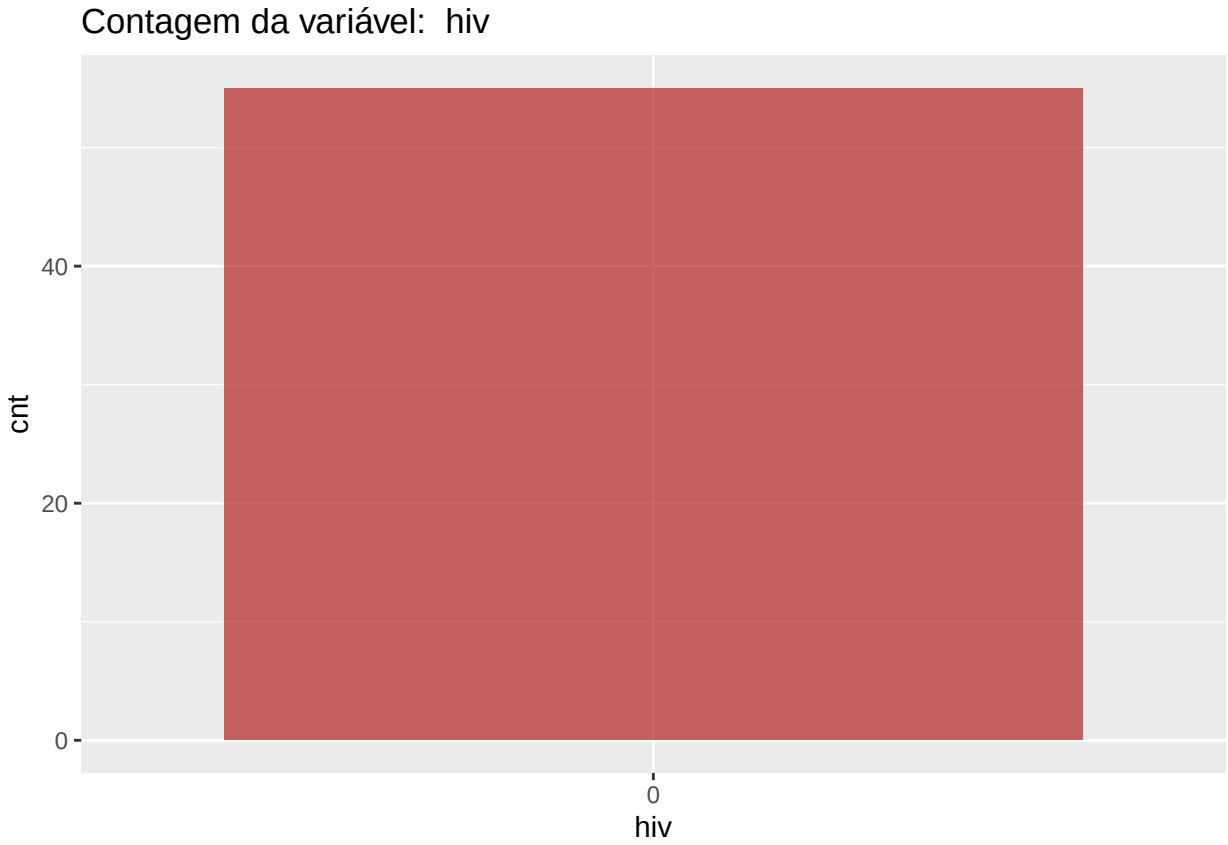

```
## `summarise()` ungrouping output (override with `.groups` argument)
```



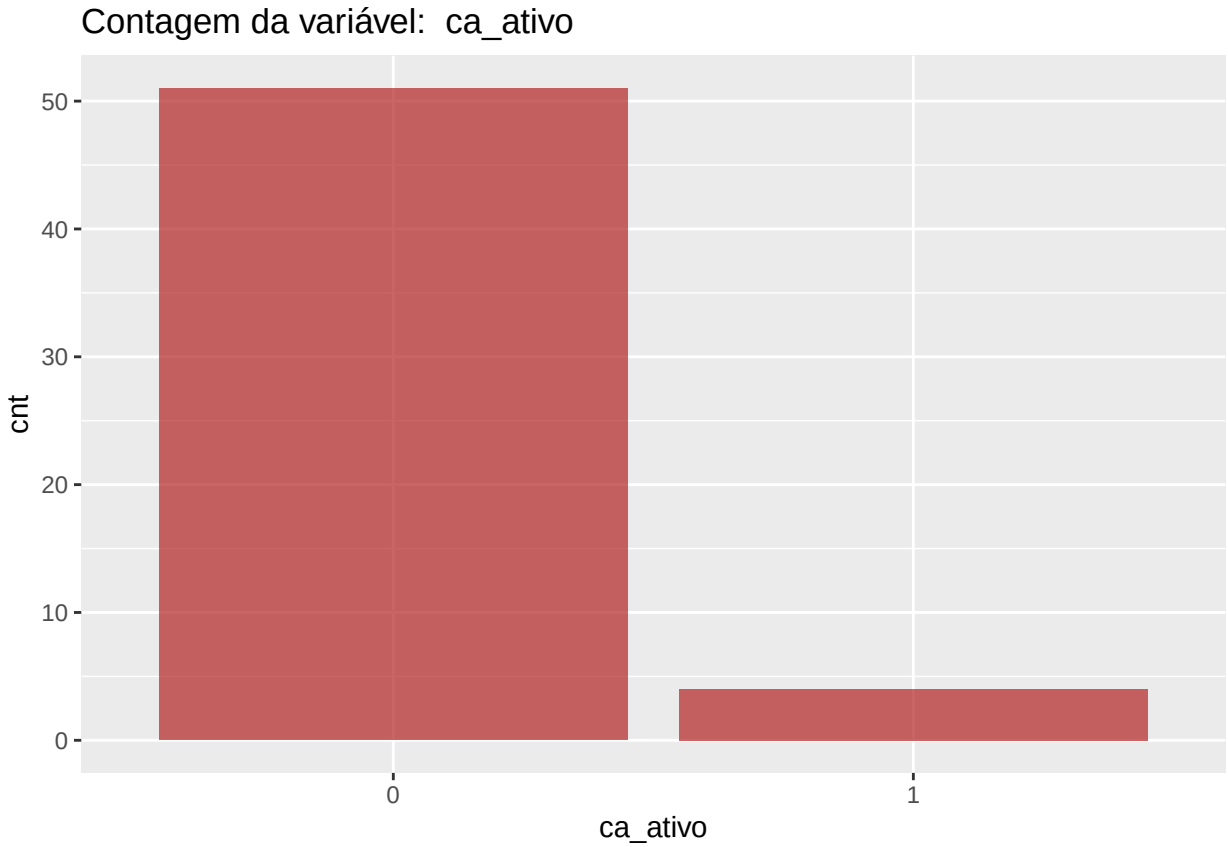

```
## `summarise()` ungrouping output (override with `.groups` argument)
```



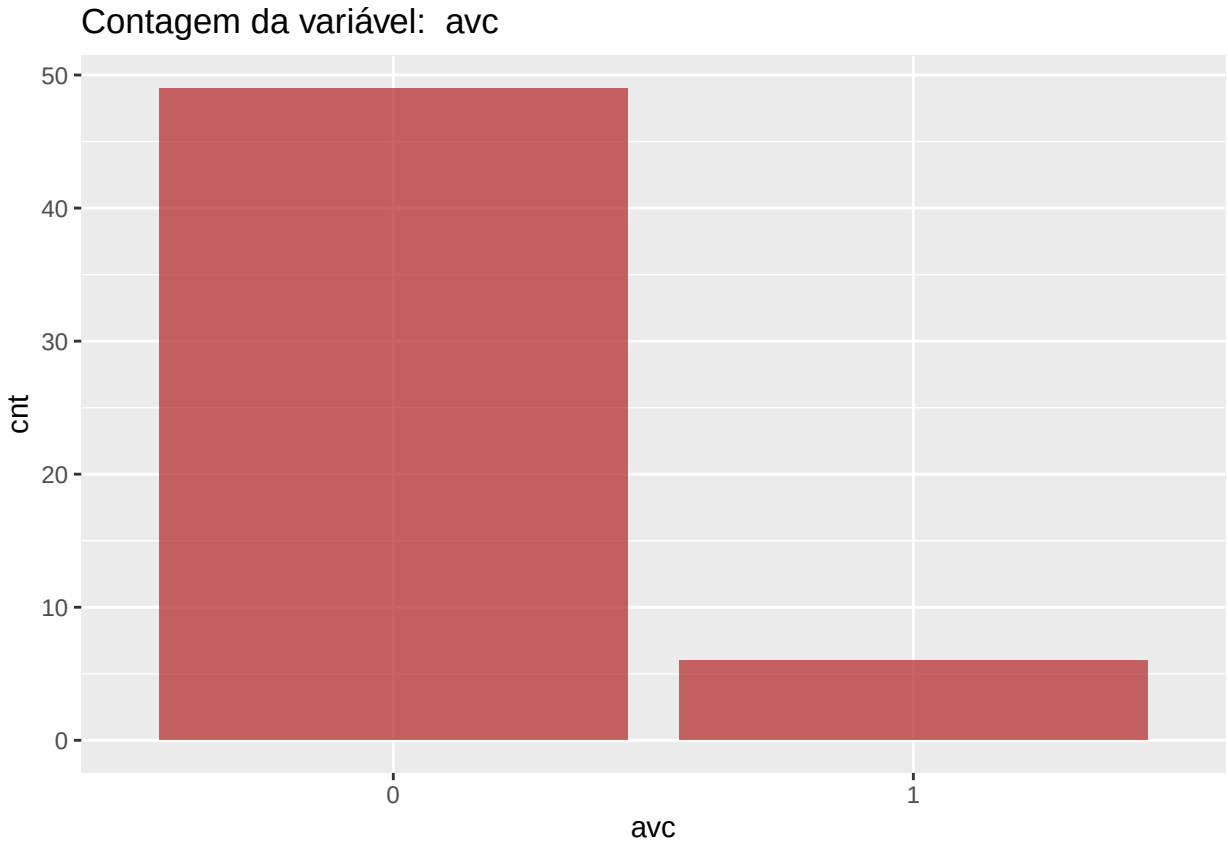

```
## `summarise()` ungrouping output (override with `.groups` argument)
```



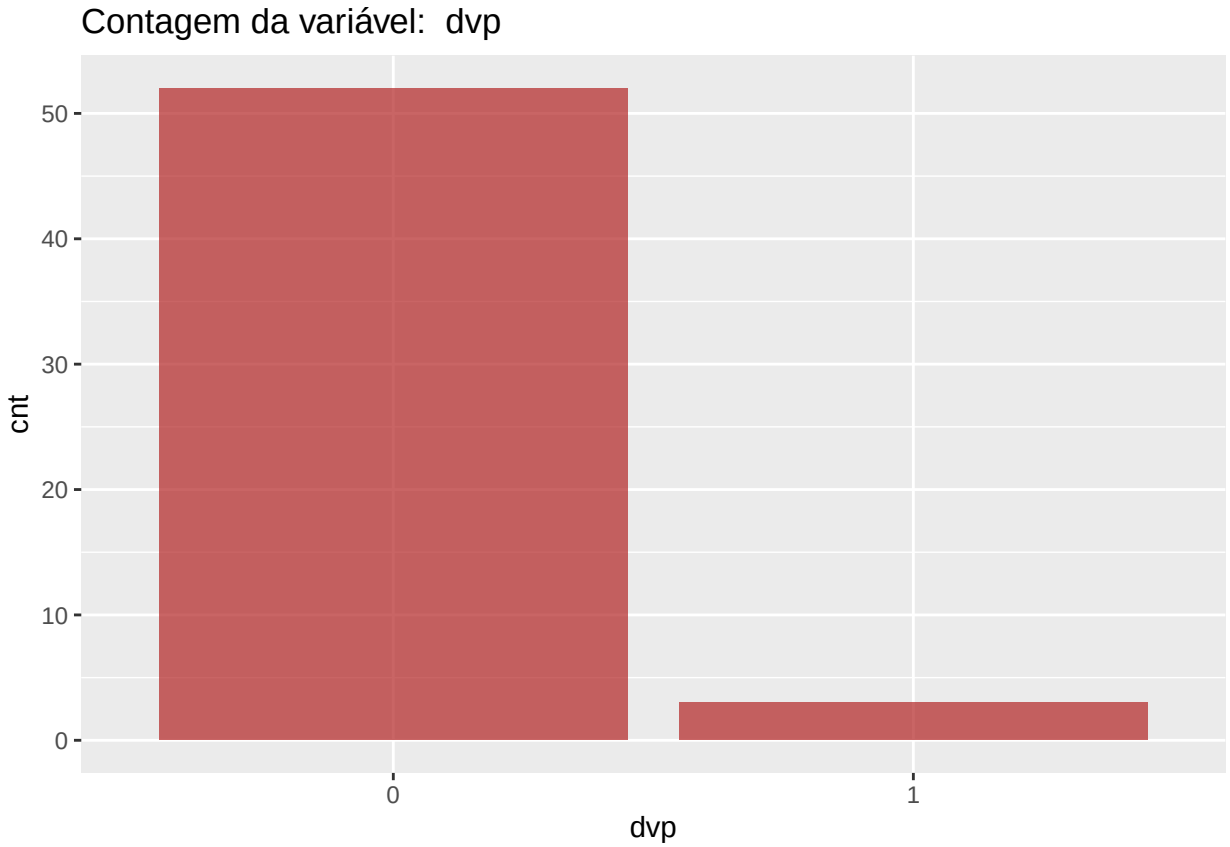

```
## `summarise()` ungrouping output (override with `.groups` argument)
```



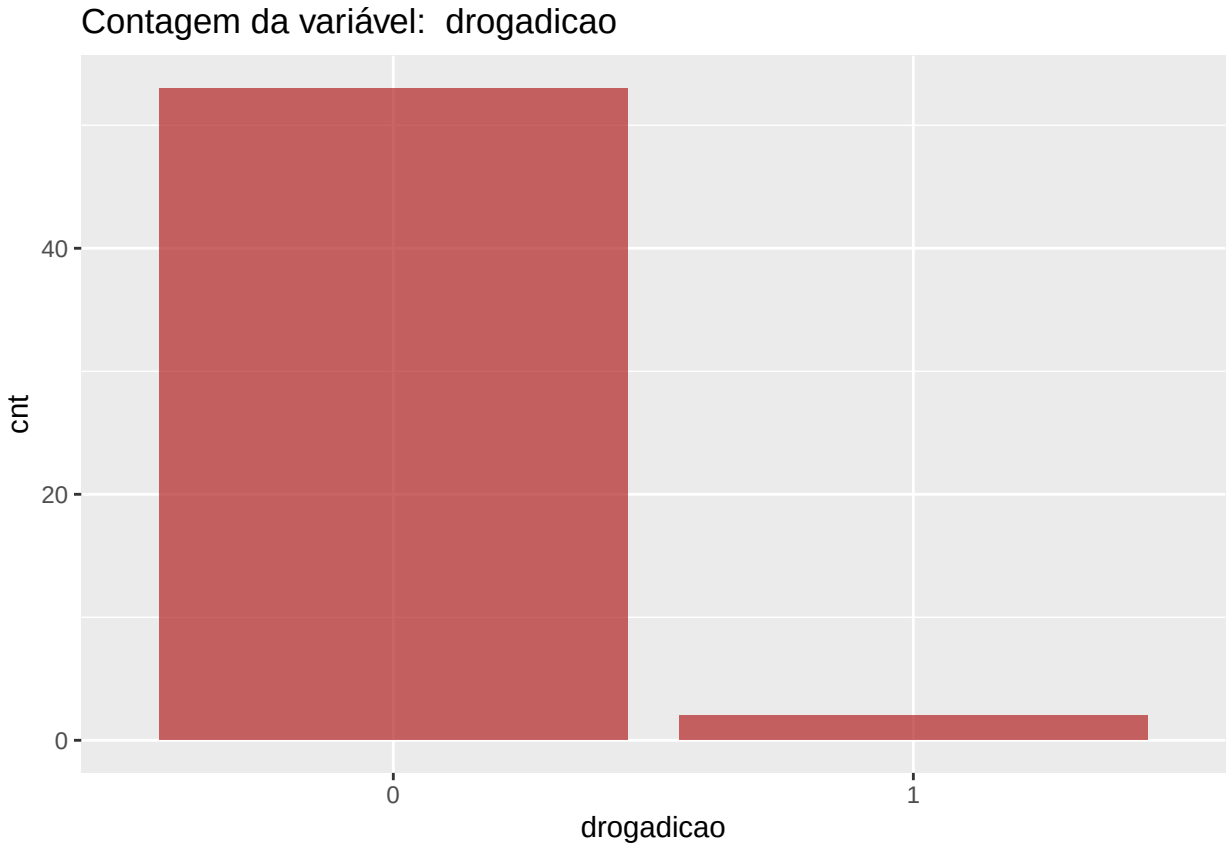

```
## `summarise()` ungrouping output (override with `.groups` argument)
```



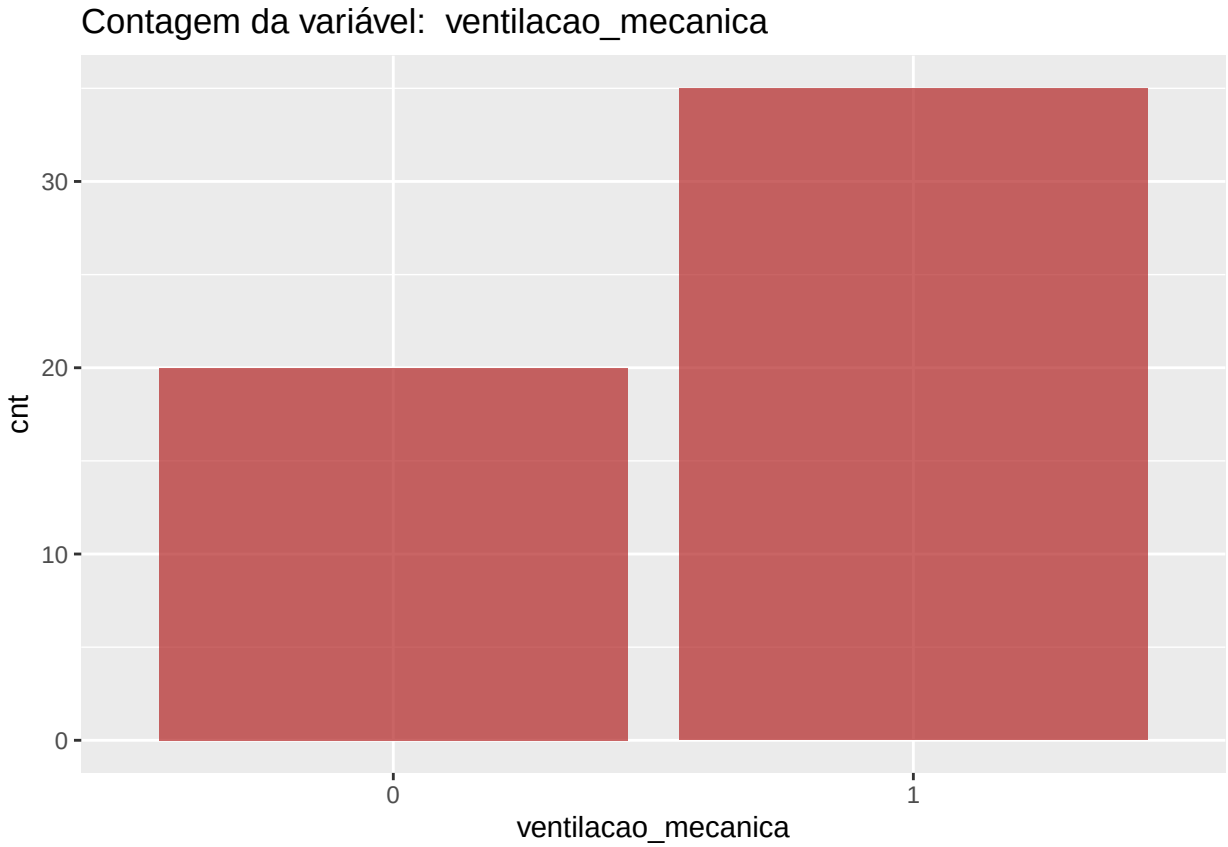

```
## `summarise()` ungrouping output (override with `.groups` argument)
```



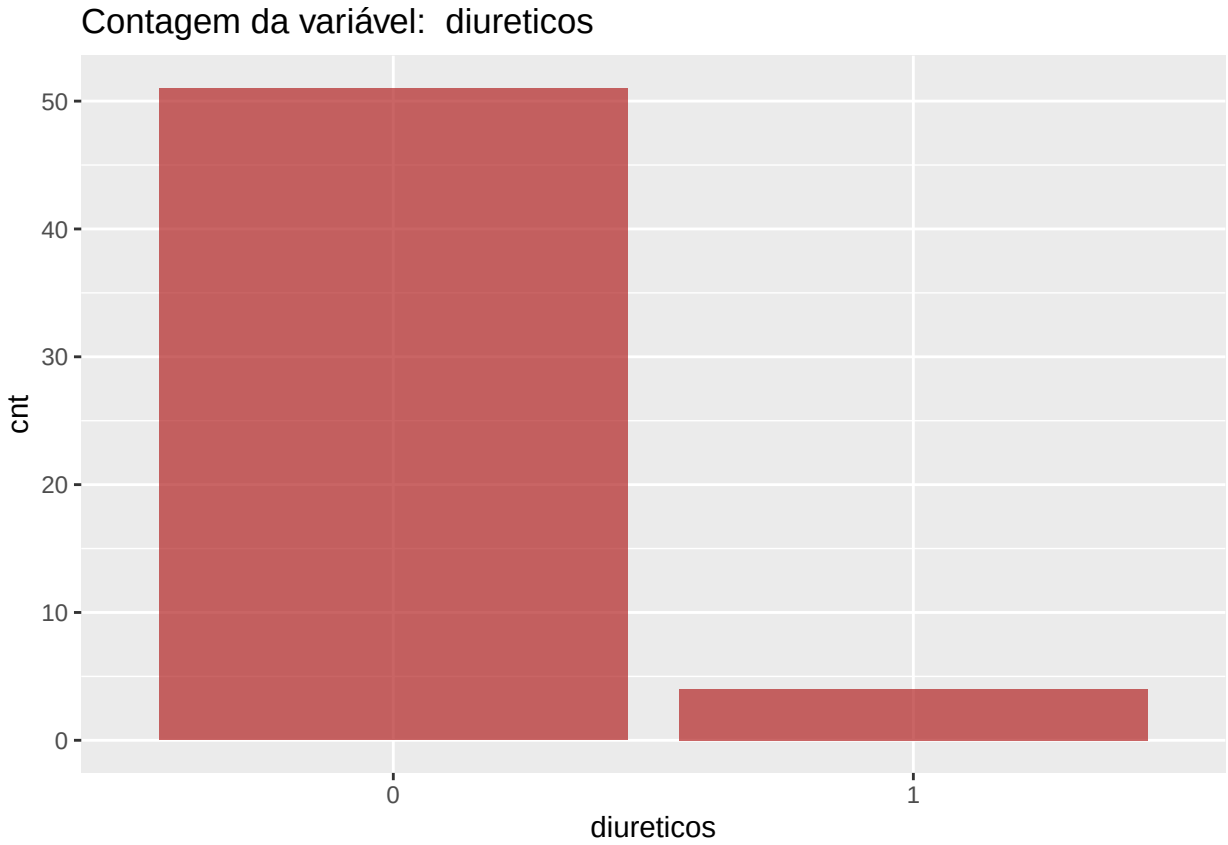

```
## `summarise()` ungrouping output (override with `.groups` argument)
```



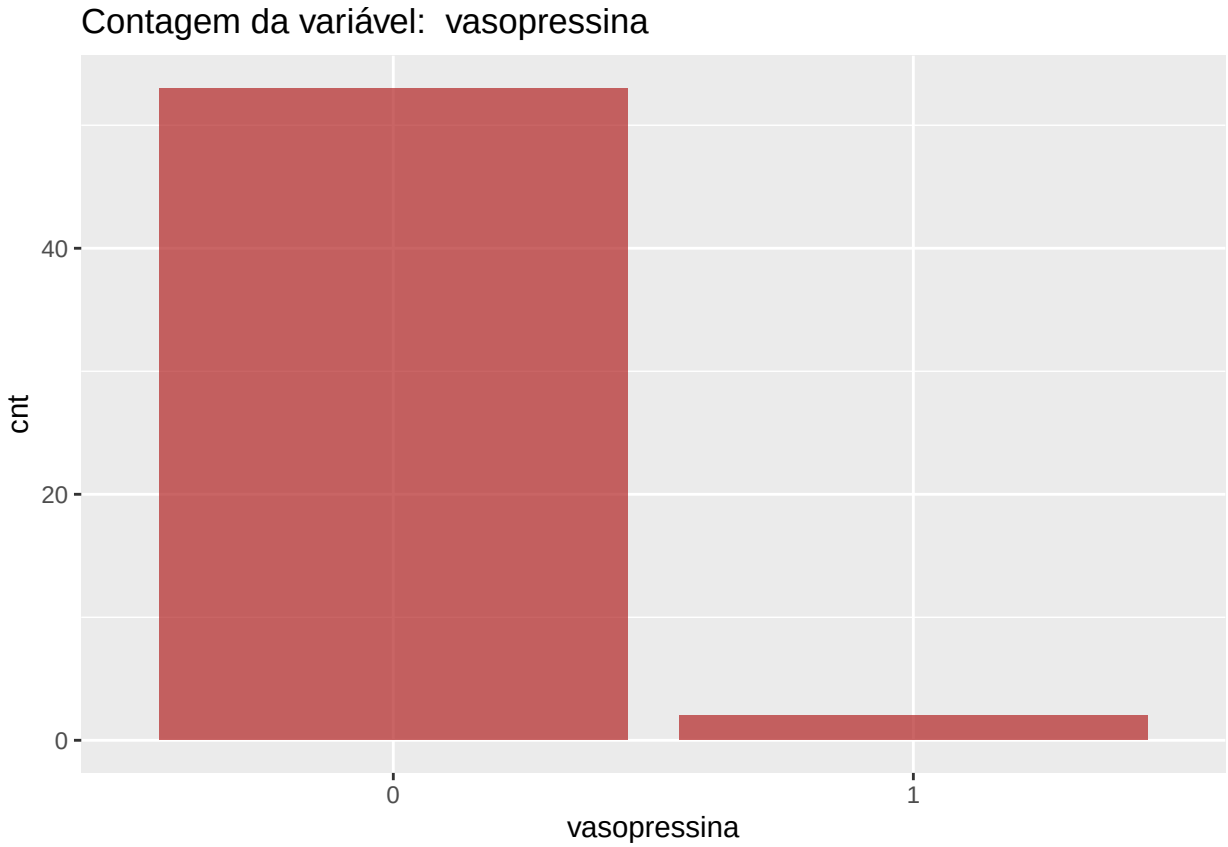

```
## `summarise()` ungrouping output (override with `.groups` argument)
```



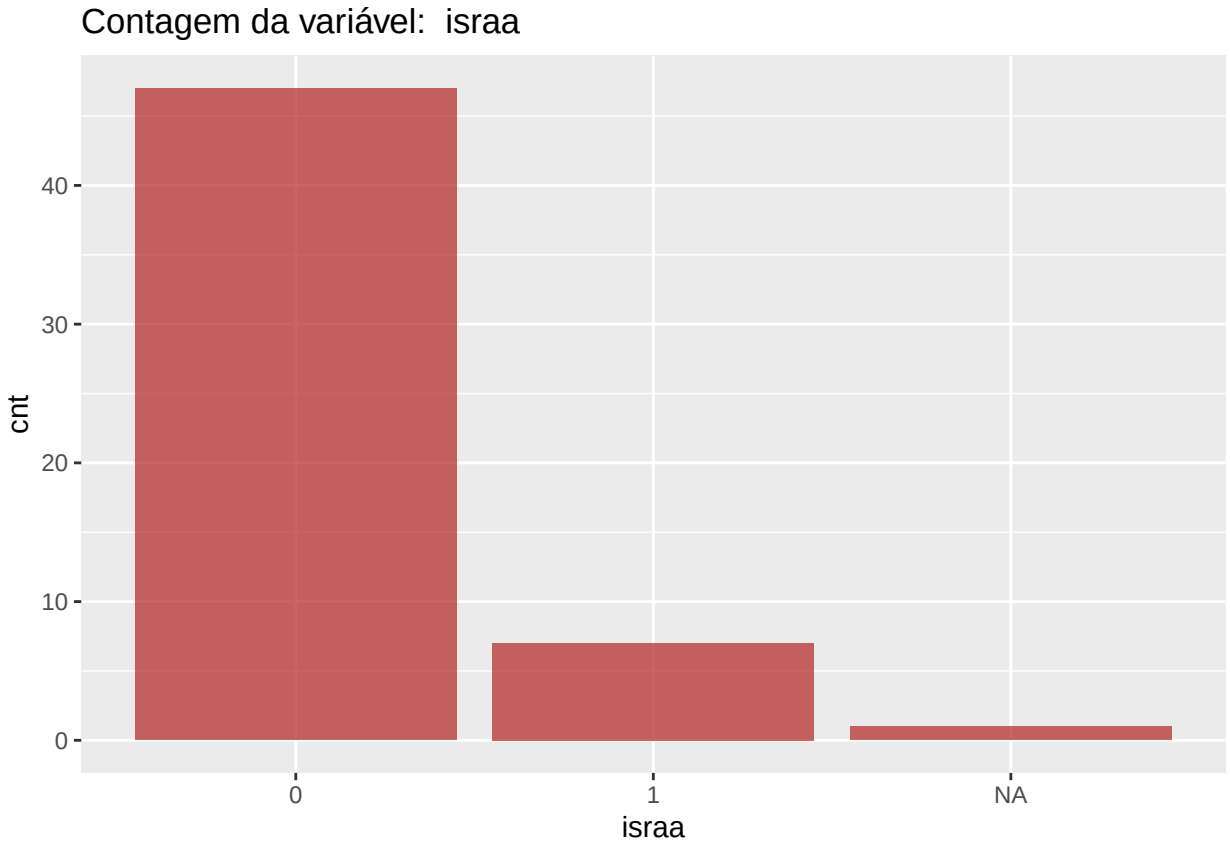

```
## `summarise()` ungrouping output (override with `.groups` argument)
```



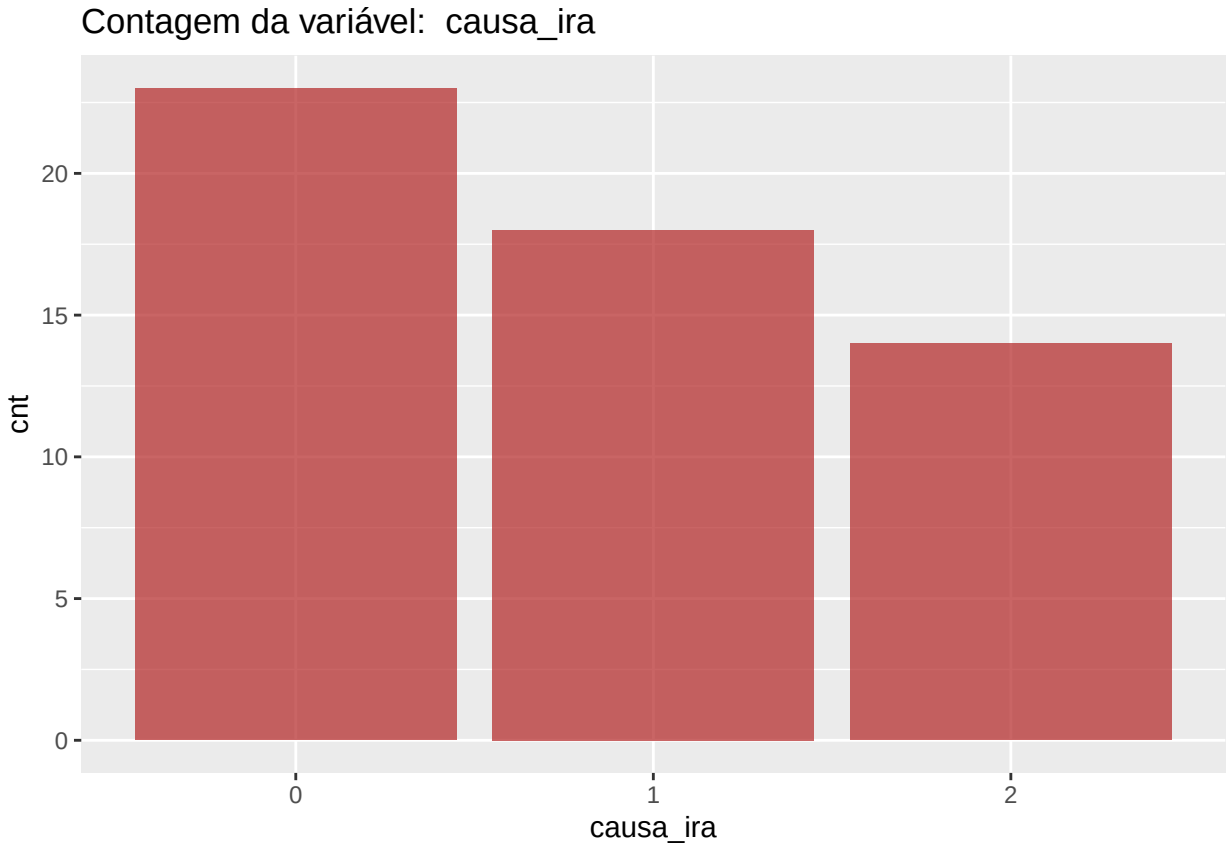

```
## `summarise()` ungrouping output (override with `.groups` argument)
```



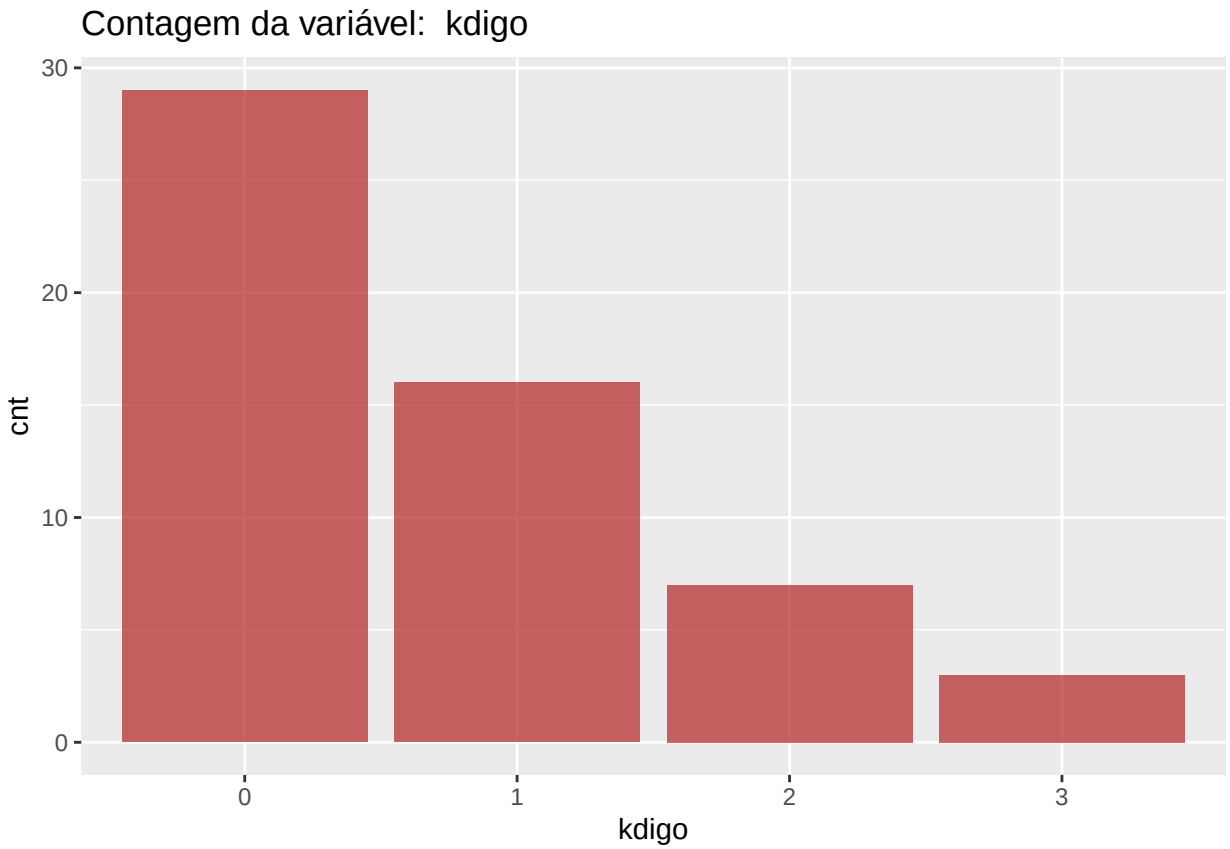

```
## `summarise()` ungrouping output (override with `.groups` argument)
```



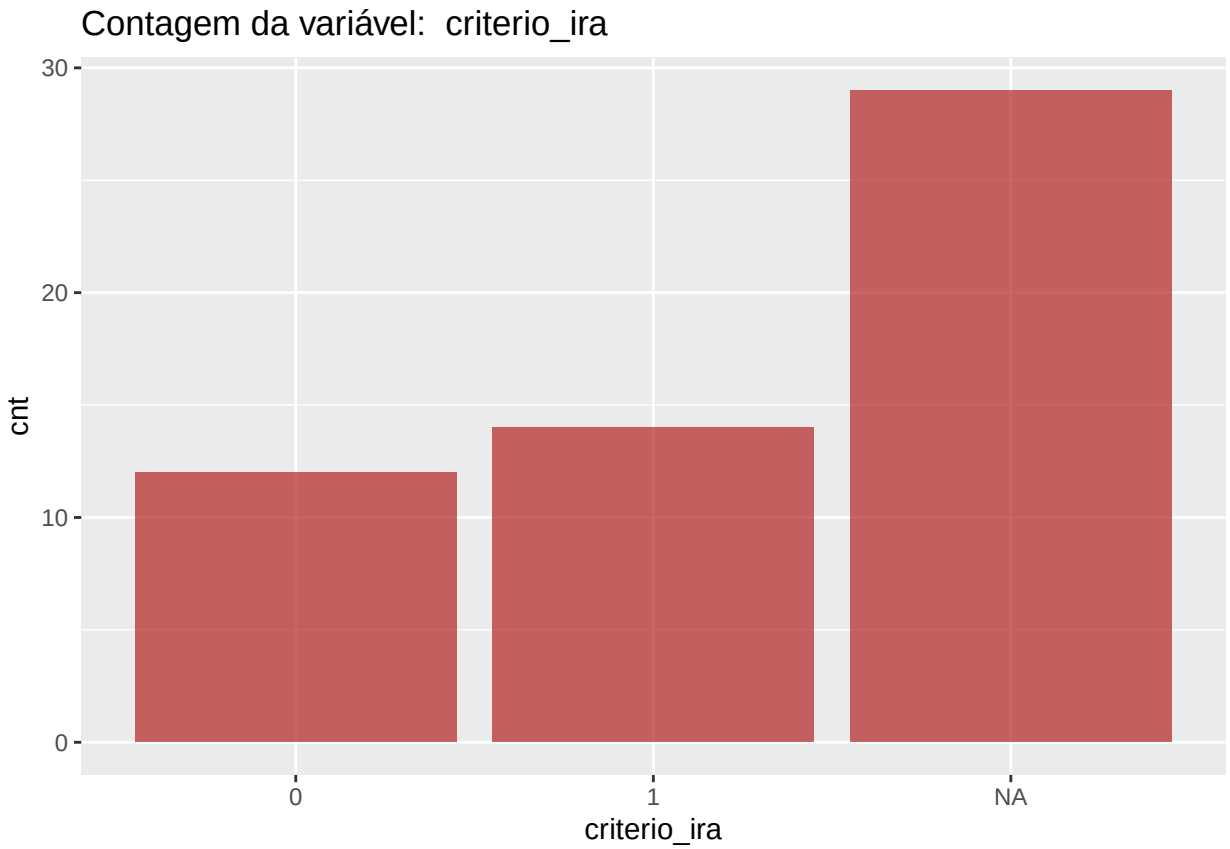

```
## `summarise()` ungrouping output (override with `.groups` argument)
```



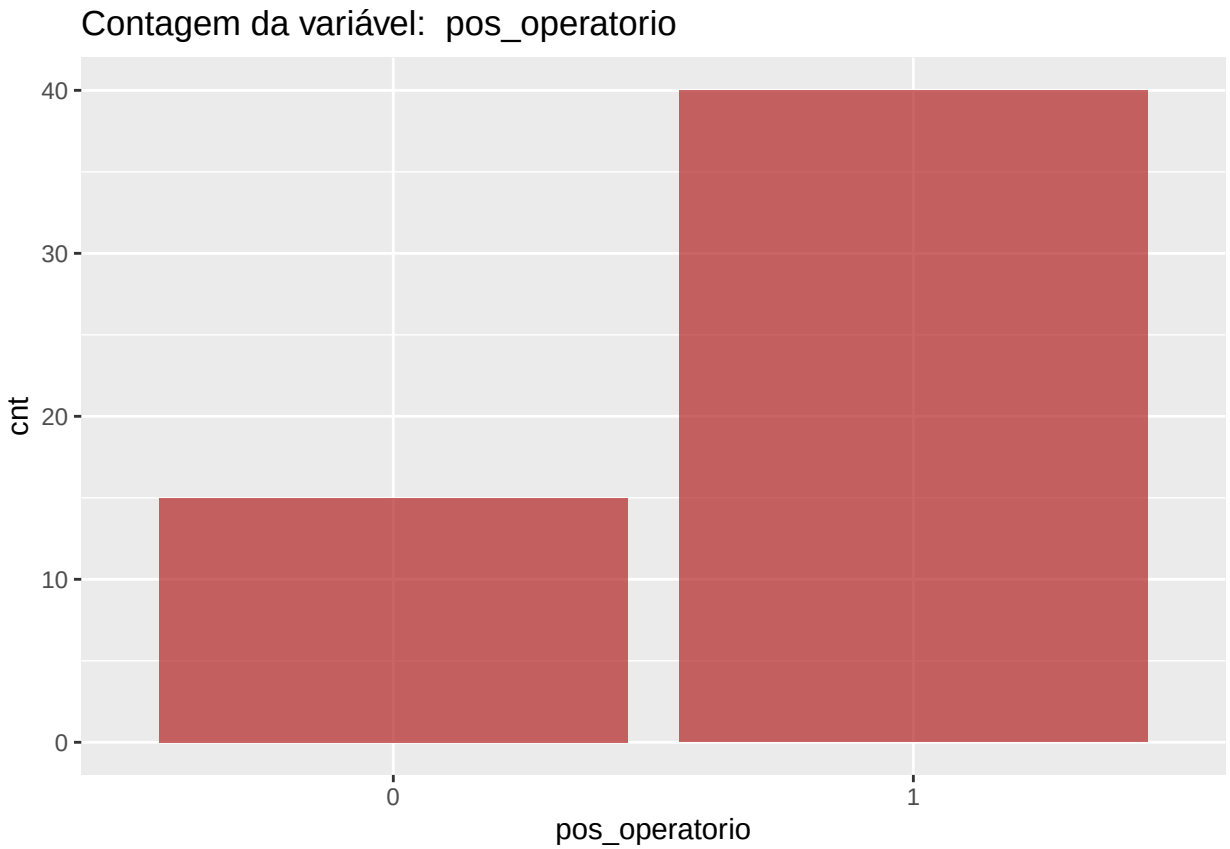

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

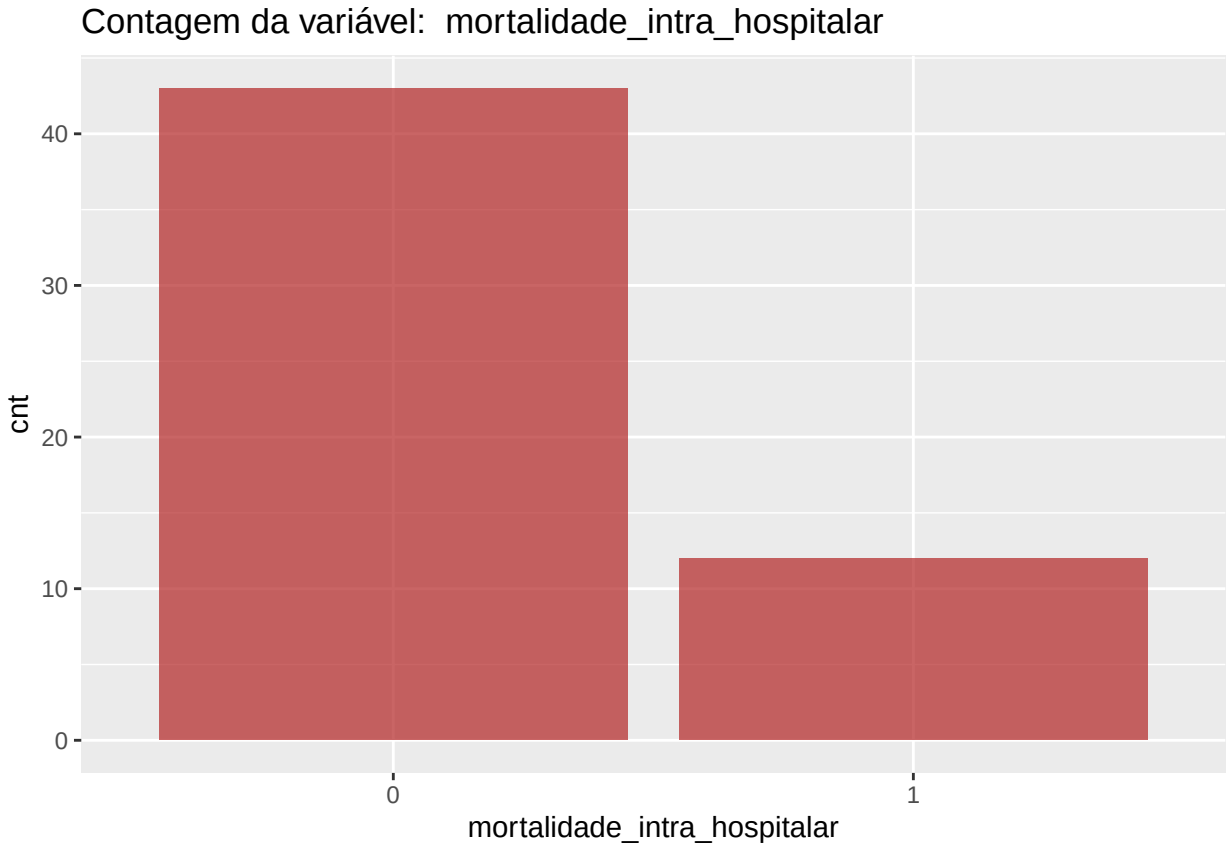



```
## `summarise()` ungrouping output (override with `.groups` argument)
```




```
## `summarise()` ungrouping output (override with `.groups` argument)
```

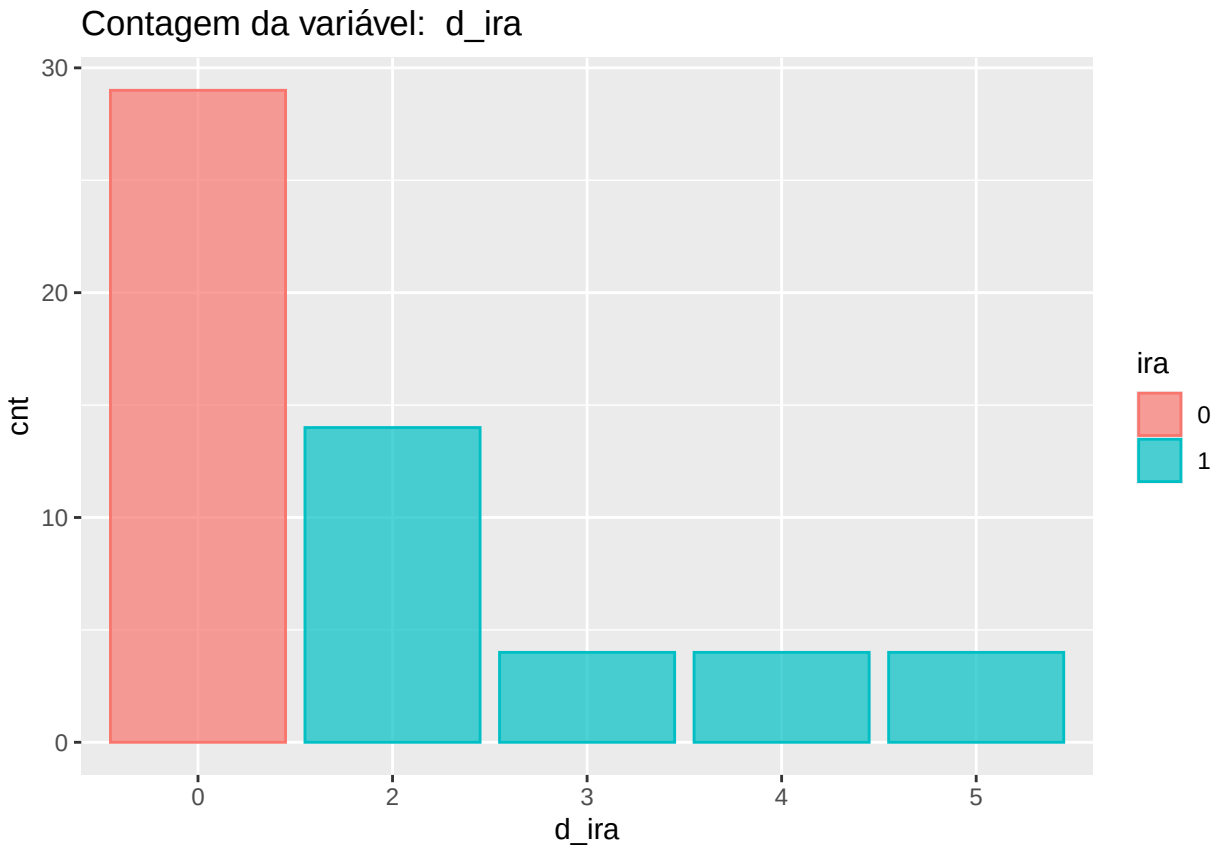

Contagem da variável: pos_operatorio
cnt
40
30
20
10
0
0
1
pos_operatorio



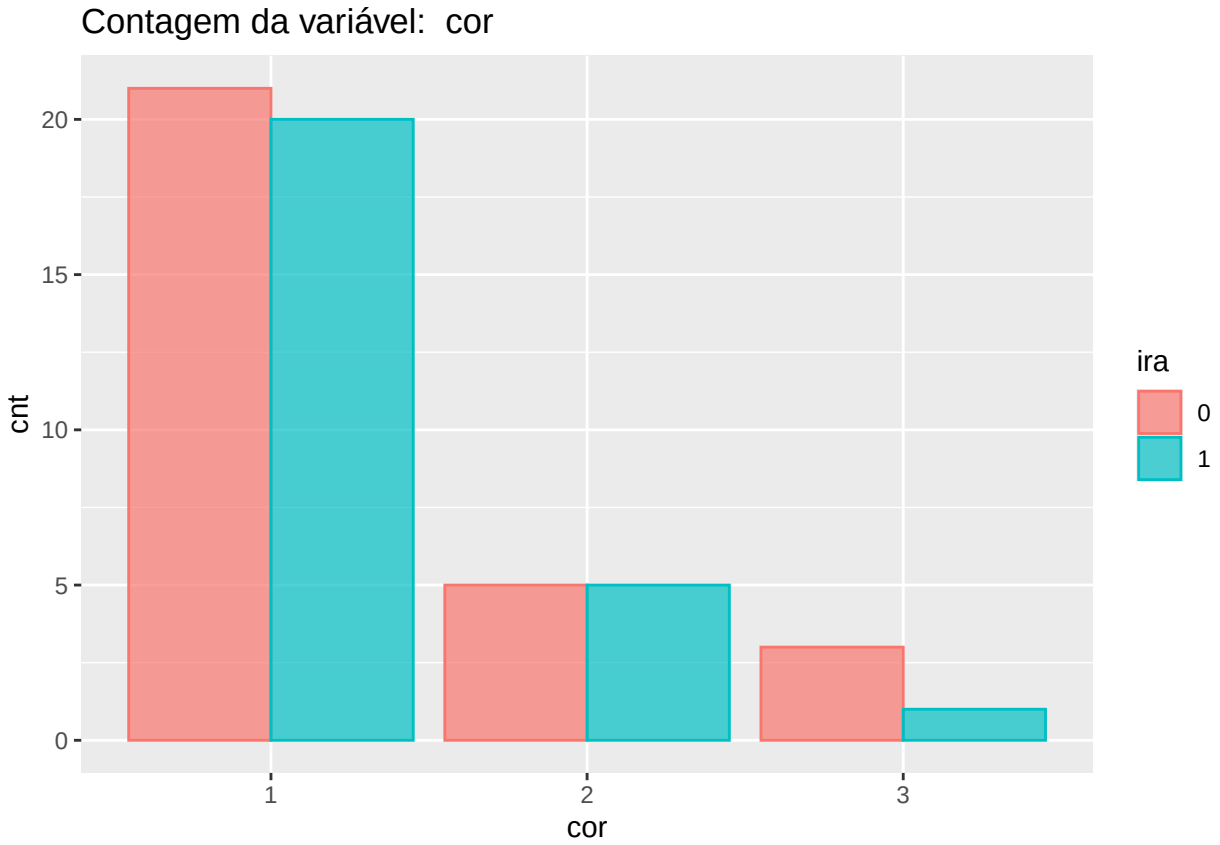

```
for(column in categorical_columns[categorical_columns != 'D_IRA']){
    p = df %>%
        group_by(!!sym(column), ira) %>%
        summarise(cnt = n()) %>%
        ggplot(aes(x=!!sym(column), y=cnt,
                   color=ira, fill=ira)) +
            geom_col(alpha=0.7, position='dodge') +
            labs(title=paste('Contagem da variável: ', column))
    print(p)
}
```

```
## `summarise()` regrouping output by 'd_ira' (override with `.groups` argument)
```

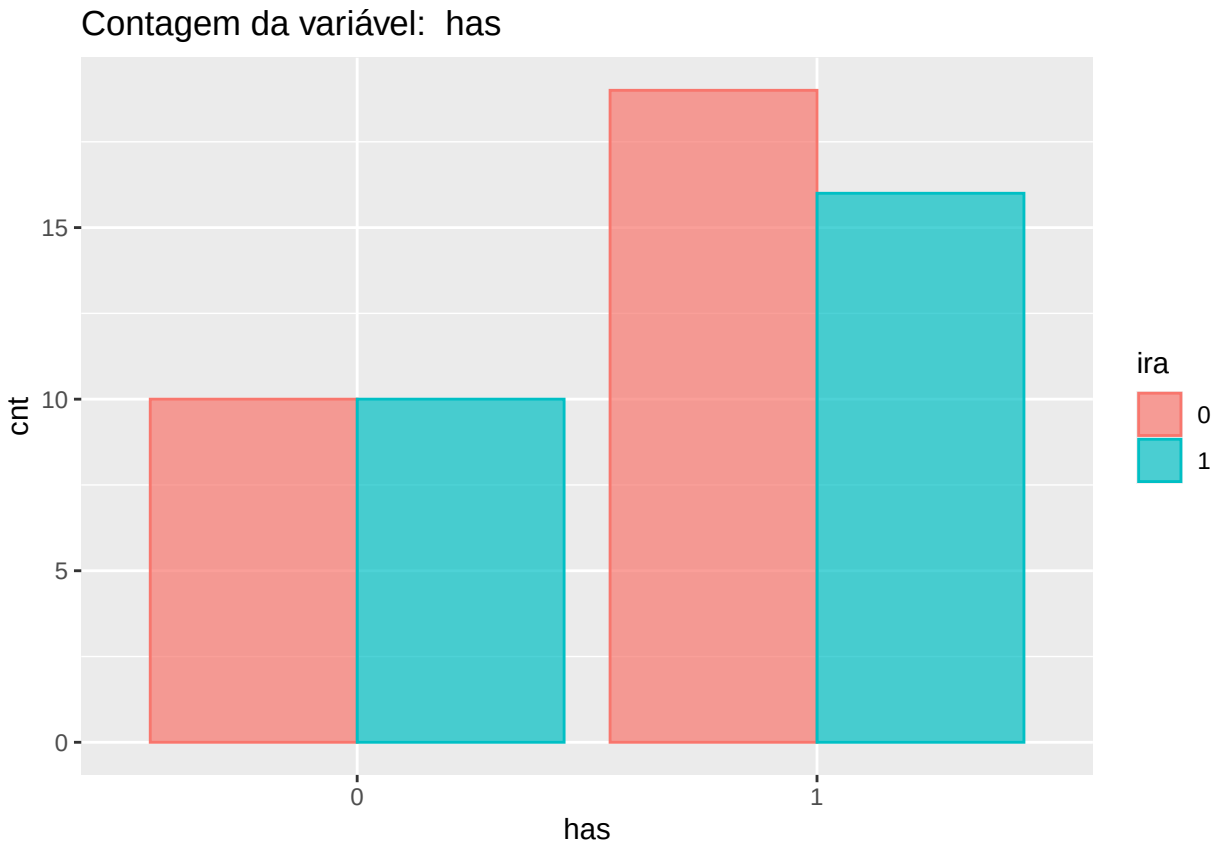
```
## `summarise()` regrouping output by 'cor' (override with `.groups` argument)
```



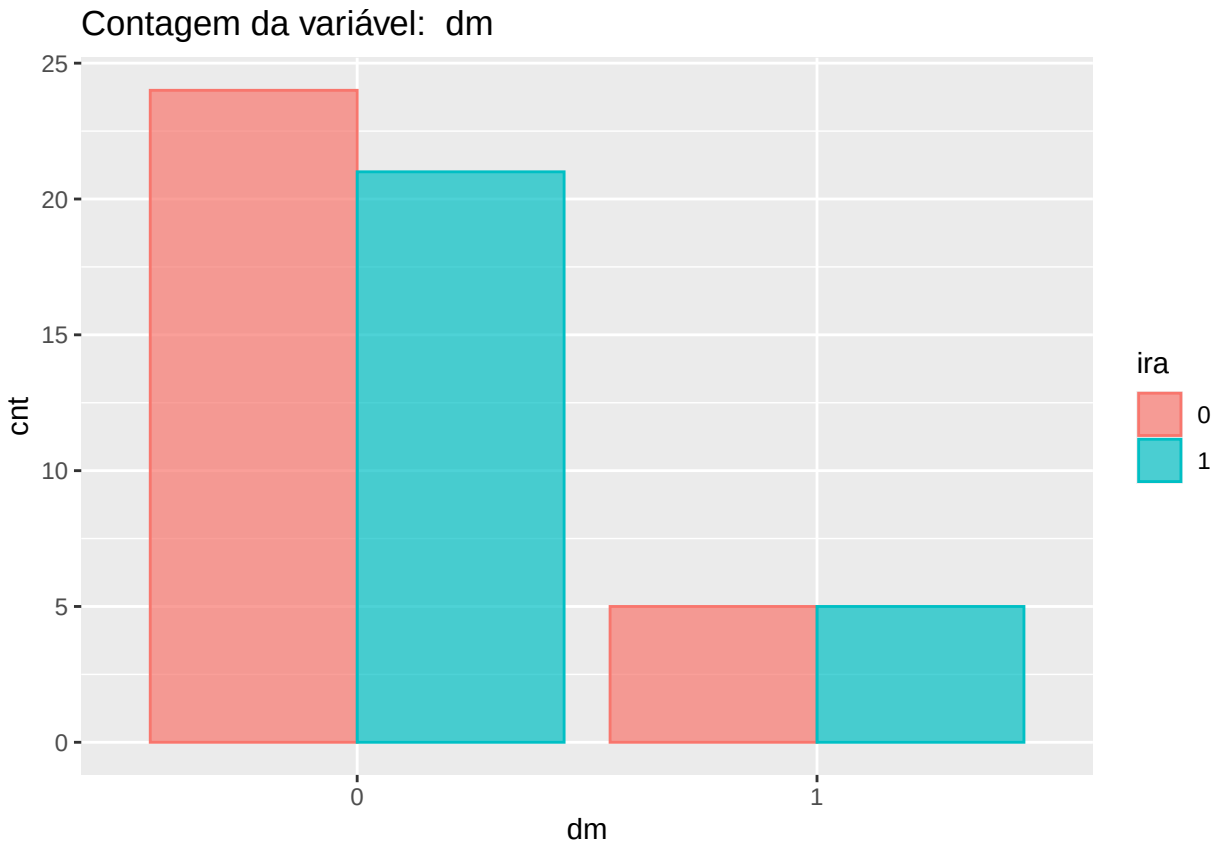

```
## `summarise()` regrouping output by 'has' (override with `.groups` argument)
```



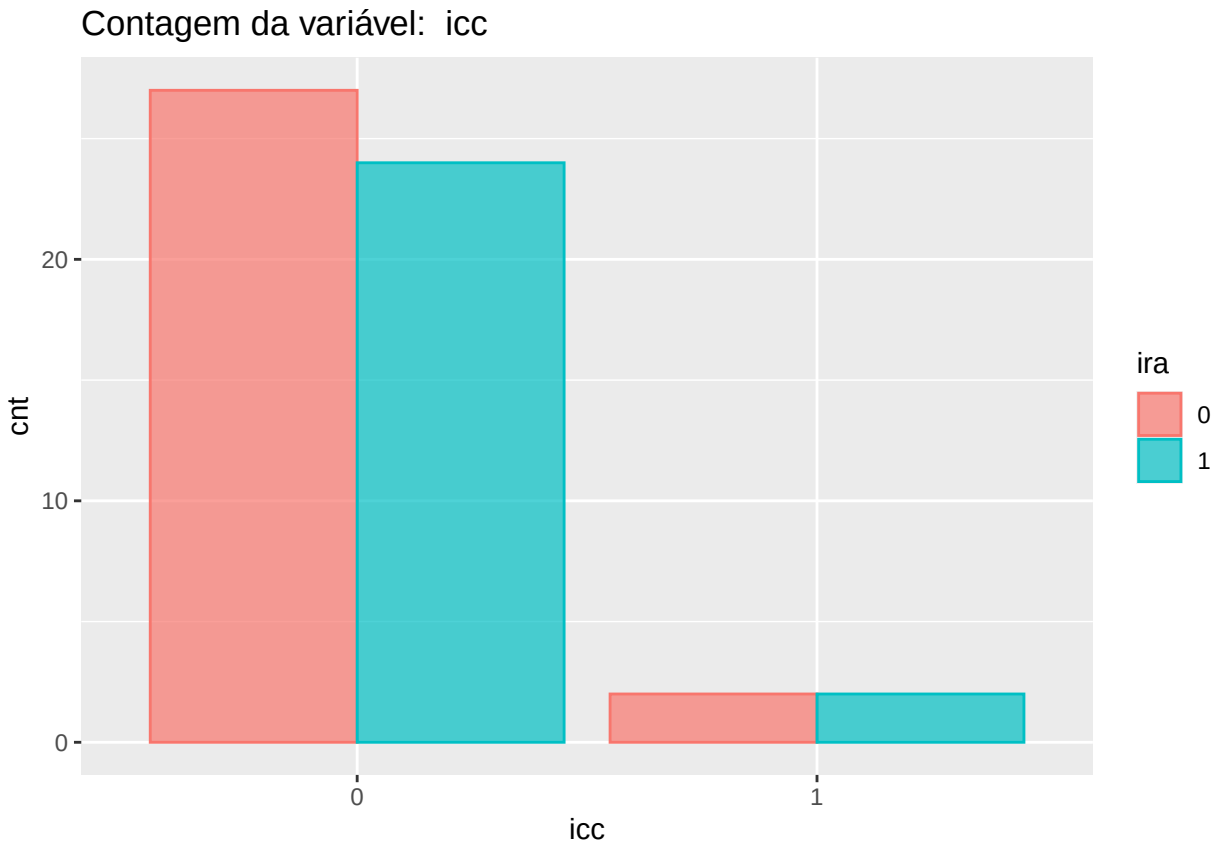

```
## `summarise()` regrouping output by 'dm' (override with `.groups` argument)
```



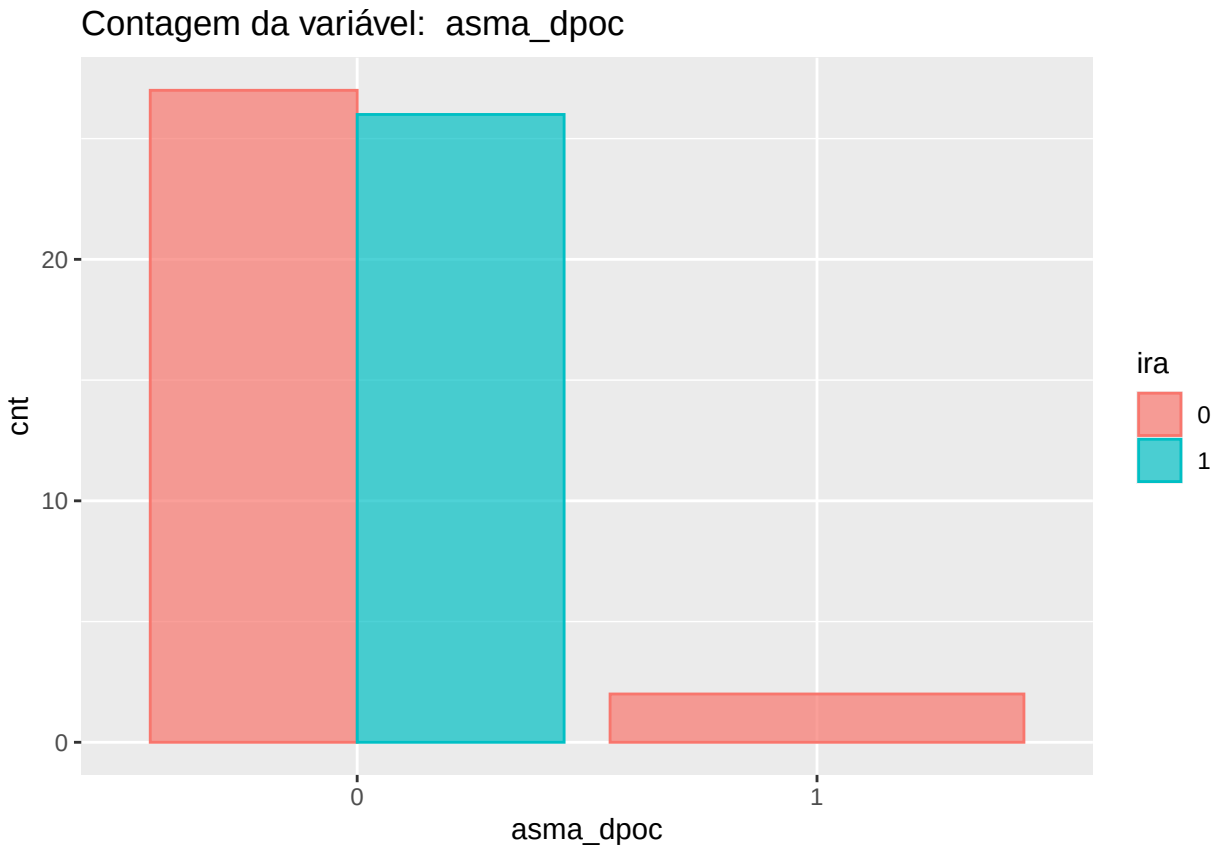

```
## `summarise()` regrouping output by 'icc' (override with `.groups` argument)
```



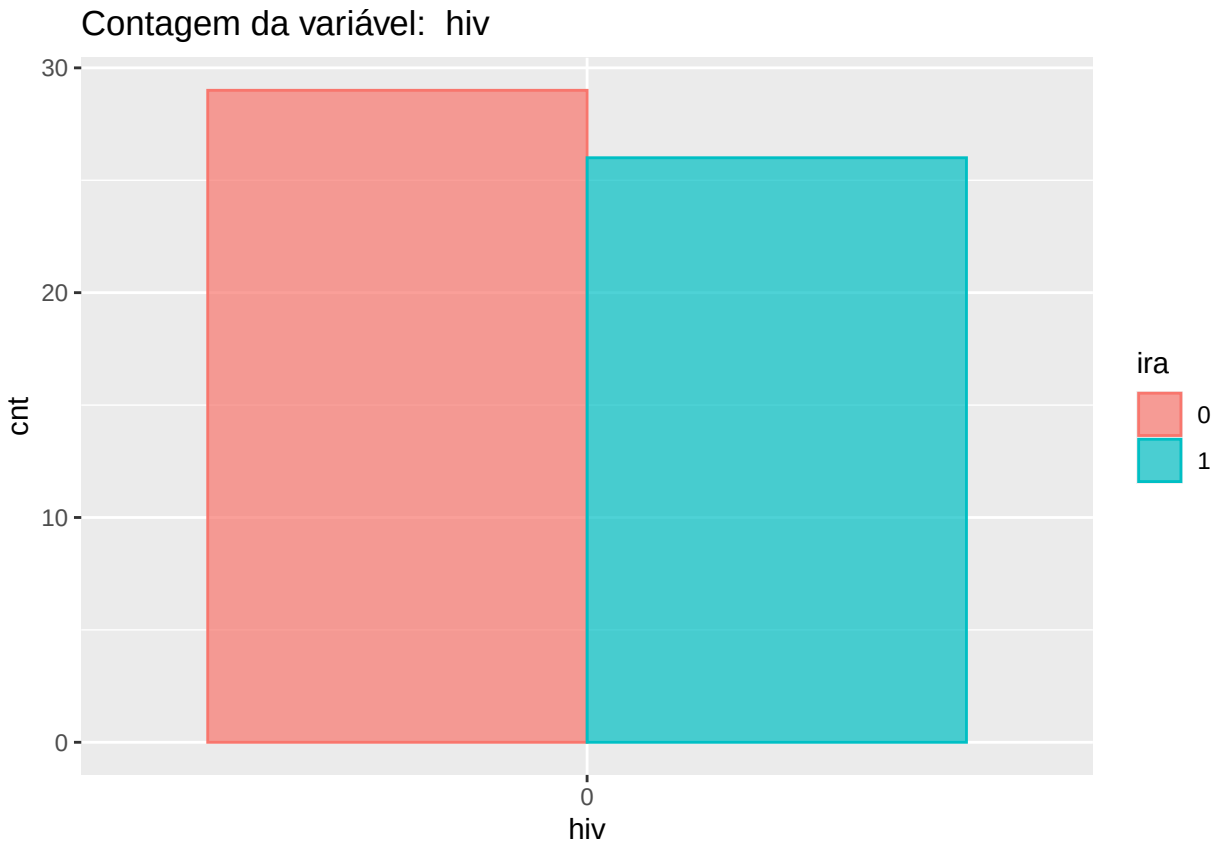

```
## `summarise()` regrouping output by 'asma_dpoc' (override with `.groups` argument)
```



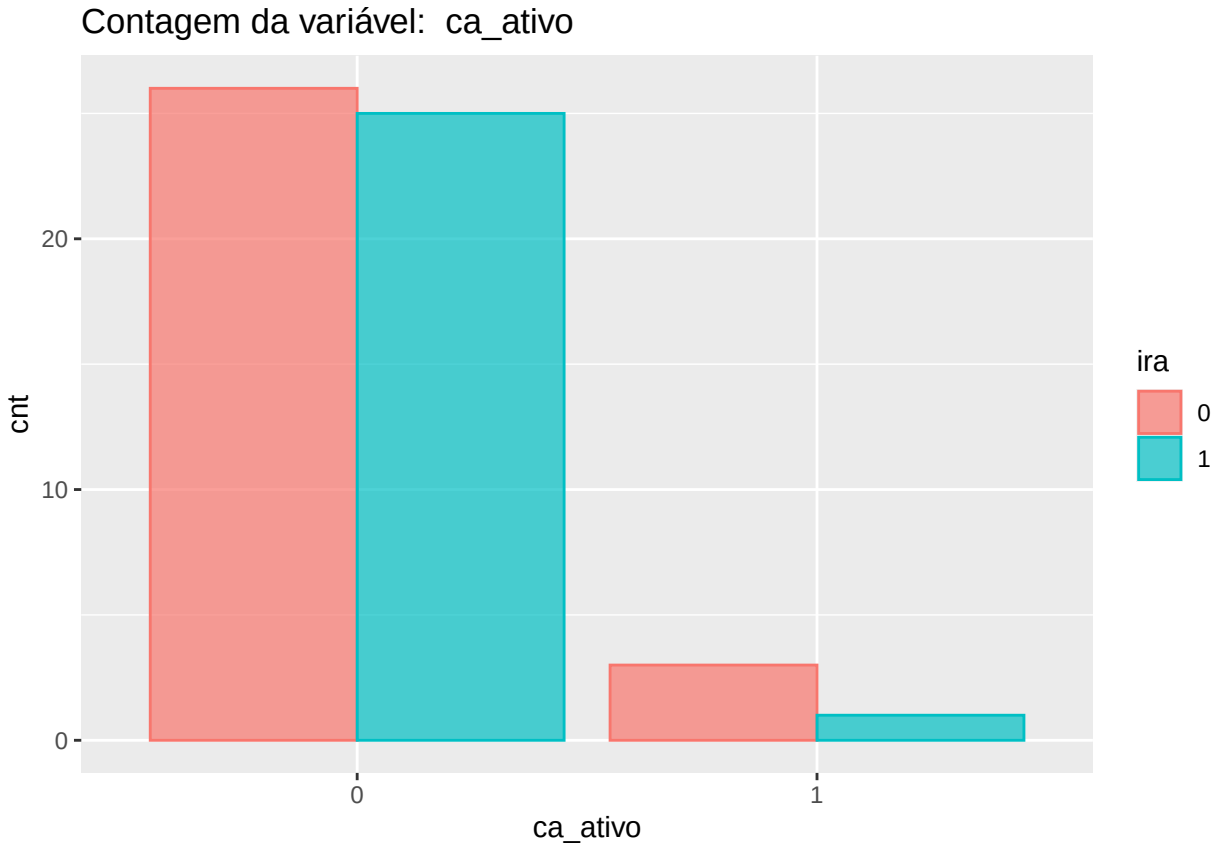

```
## `summarise()` regrouping output by 'hiv' (override with `.groups` argument)
```



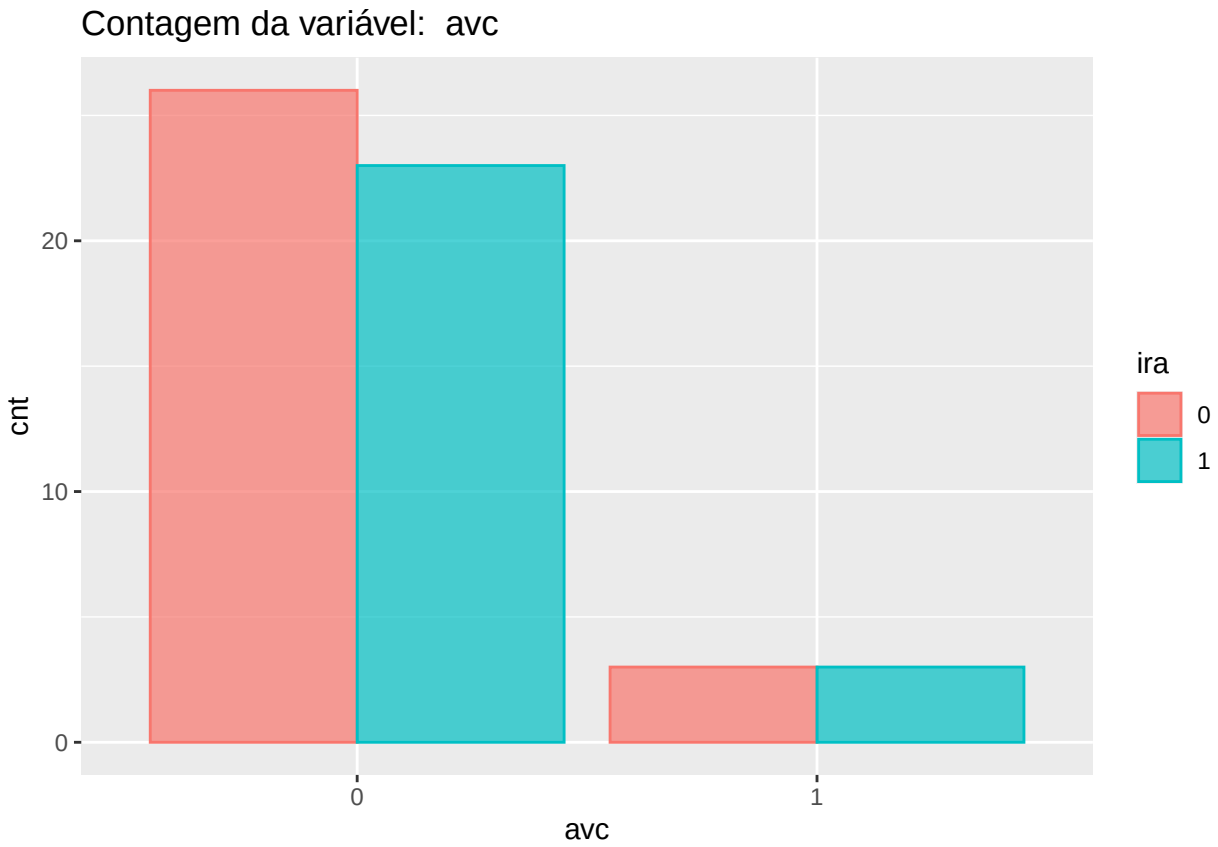

```
## `summarise()` regrouping output by 'ca_ativo' (override with `.groups` argument)
```



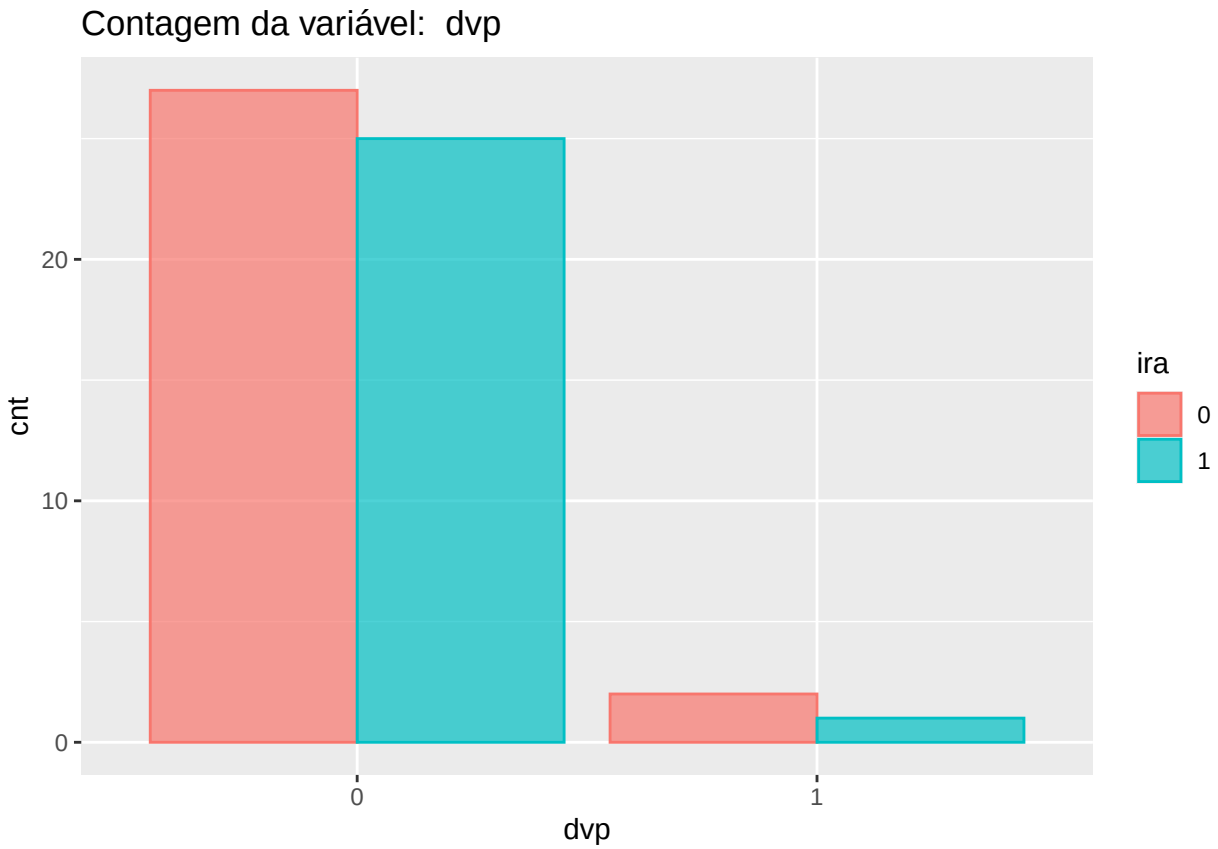

```
## `summarise()` regrouping output by 'avc' (override with `.groups` argument)
```



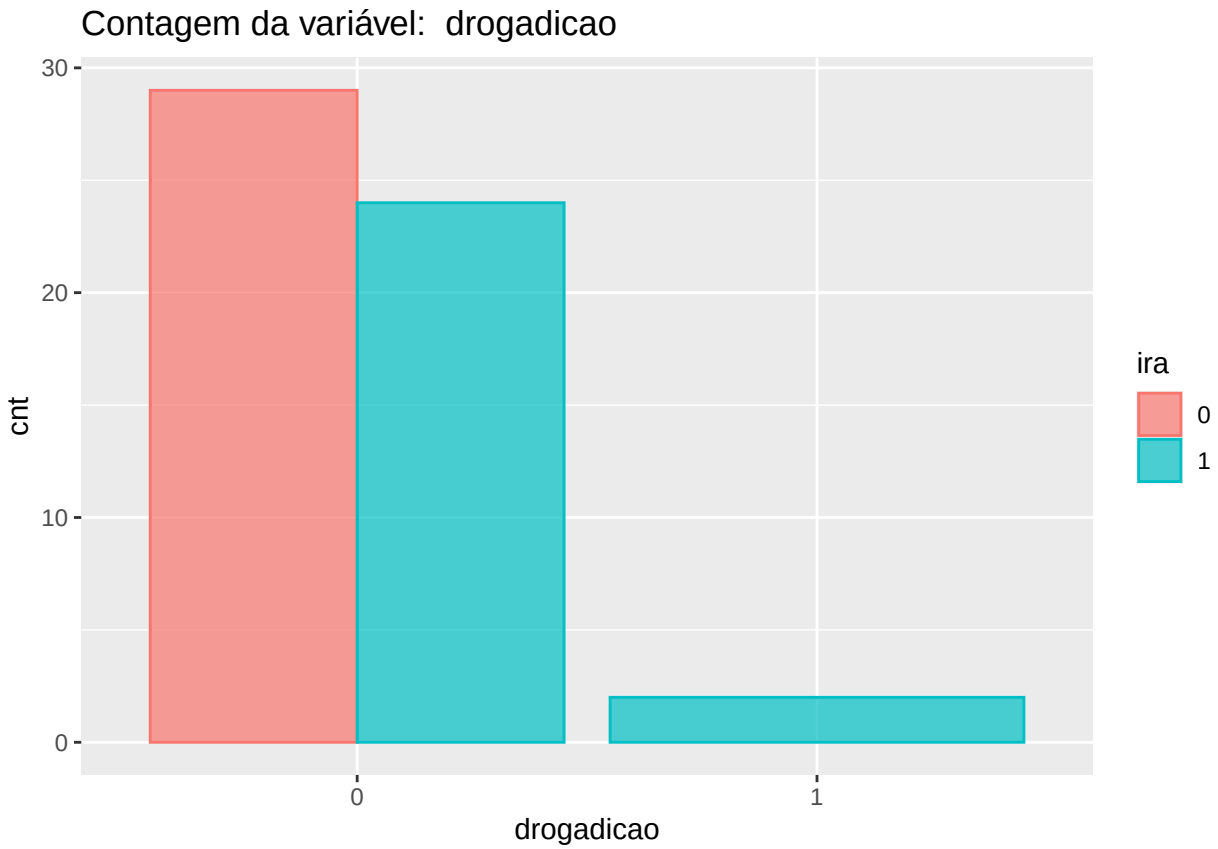

```
## `summarise()` regrouping output by 'dvp' (override with `.groups` argument)
```



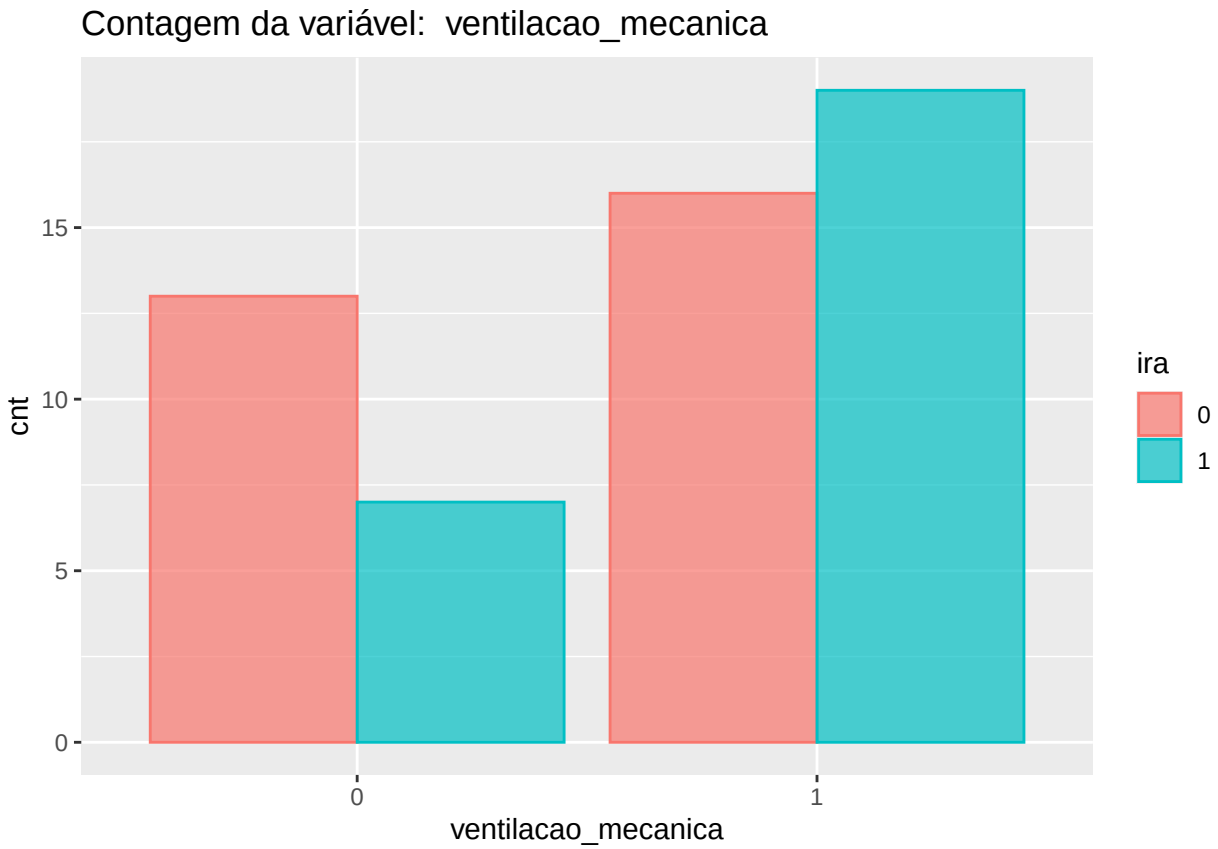

```
## `summarise()` regrouping output by 'drogadicao' (override with `.groups` argument)
```



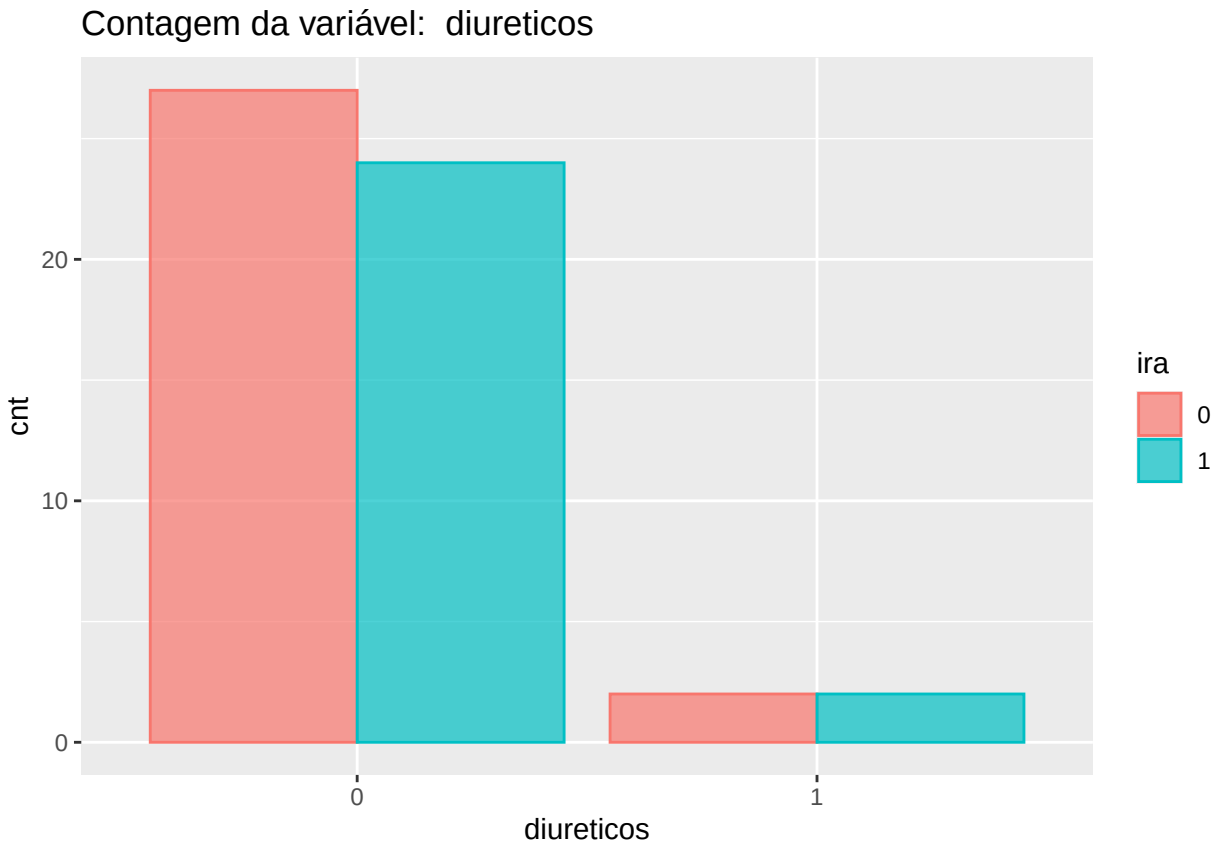

```
## `summarise()` regrouping output by 'ventilacao_mecanica' (override with `.groups` argument)
```



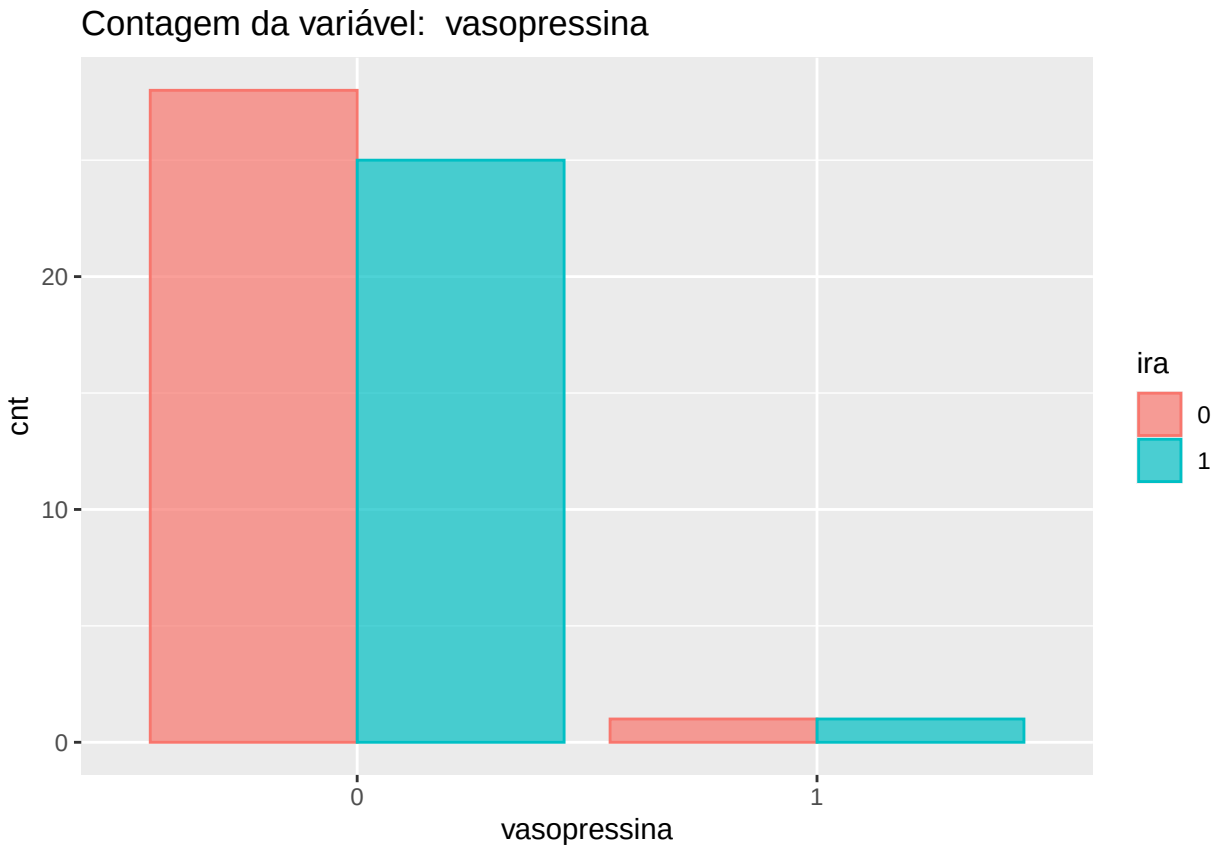

```
## `summarise()` regrouping output by 'diureticos' (override with `.groups` argument)
```



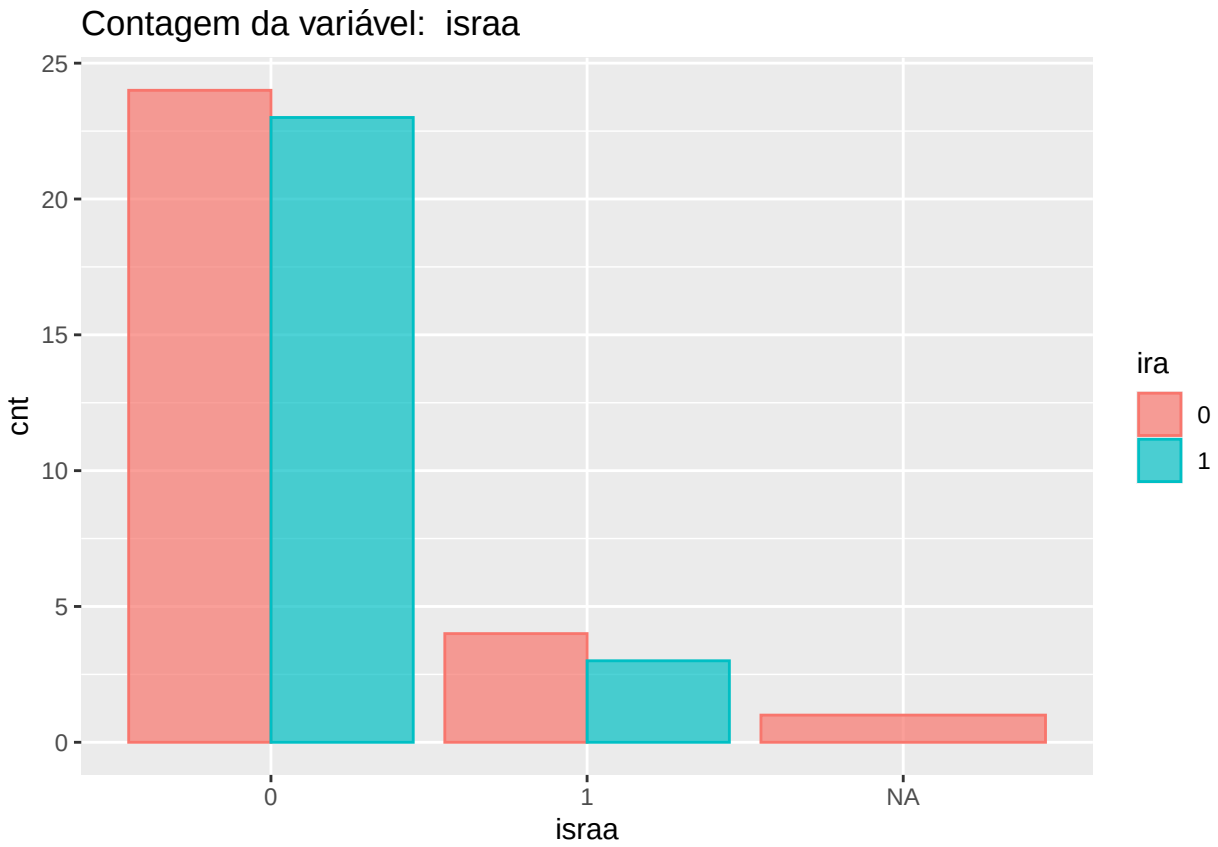

```
## `summarise()` regrouping output by 'vasopressina' (override with `.groups` argument)
```



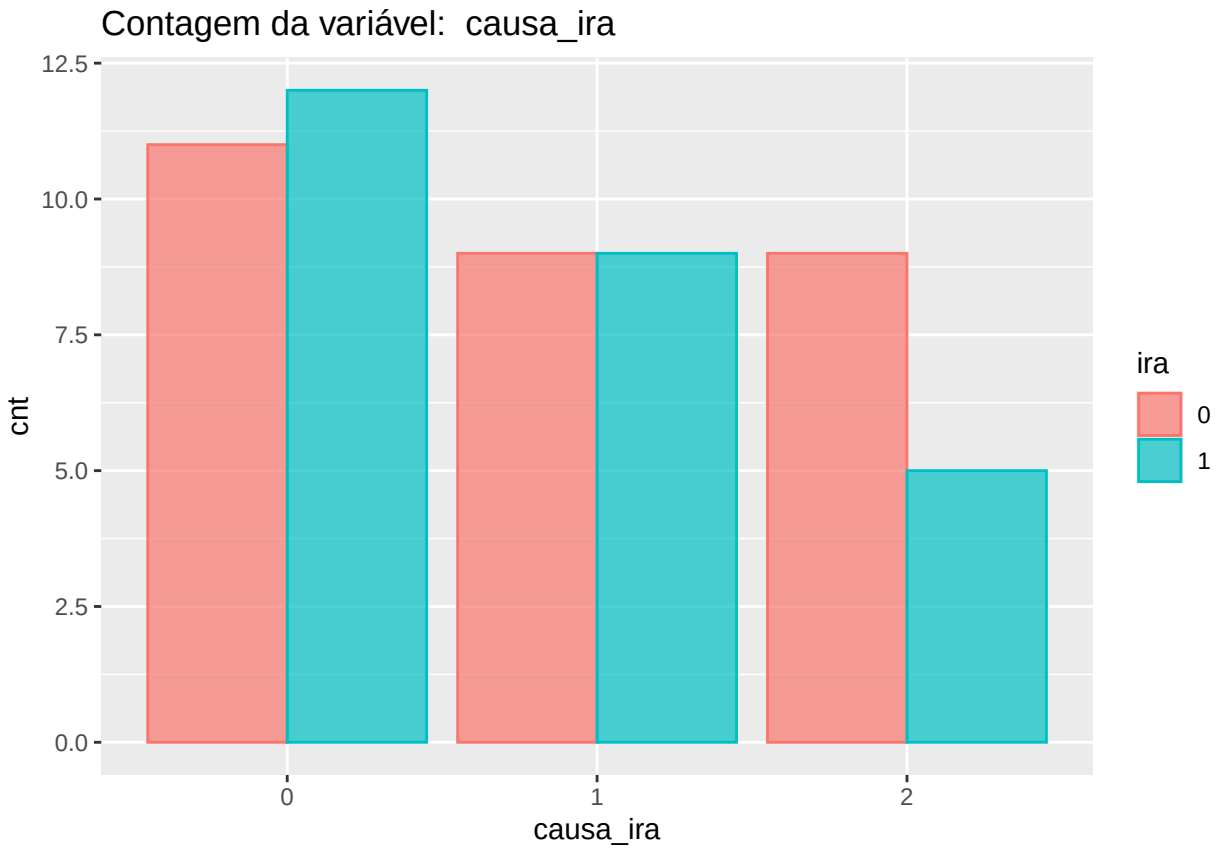

```
## `summarise()` regrouping output by 'israa' (override with `.groups` argument)
```



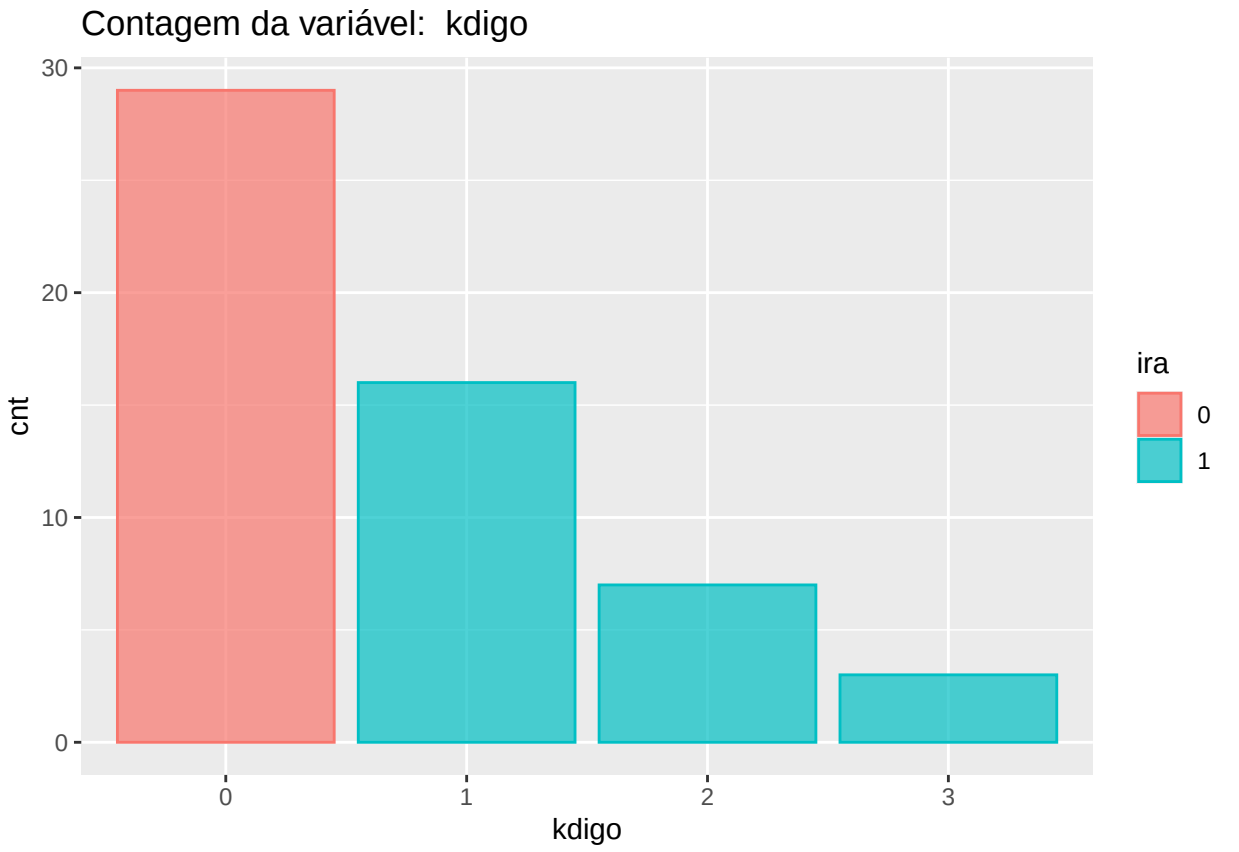

```
## `summarise()` regrouping output by 'causa_ira' (override with `.groups` argument)
```



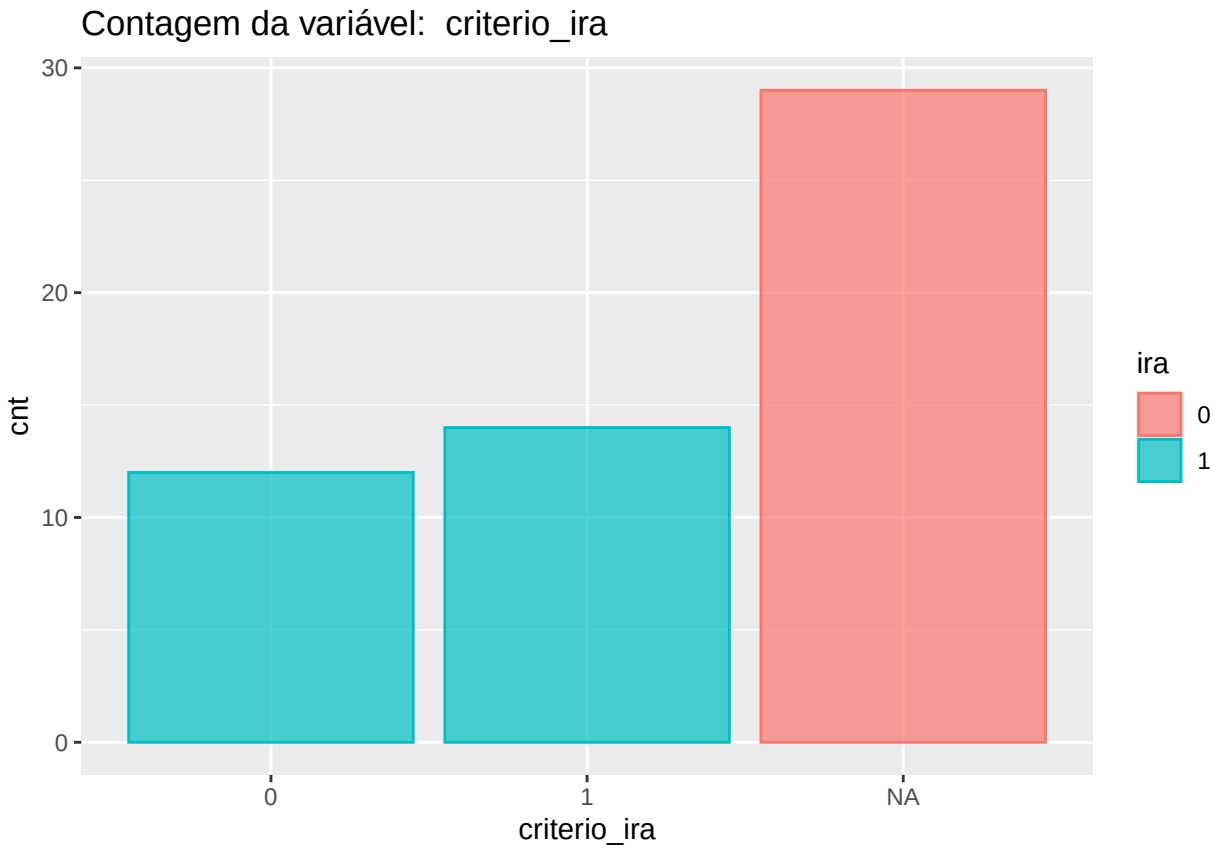

```
## `summarise()` regrouping output by 'kdigo' (override with `.groups` argument)
```



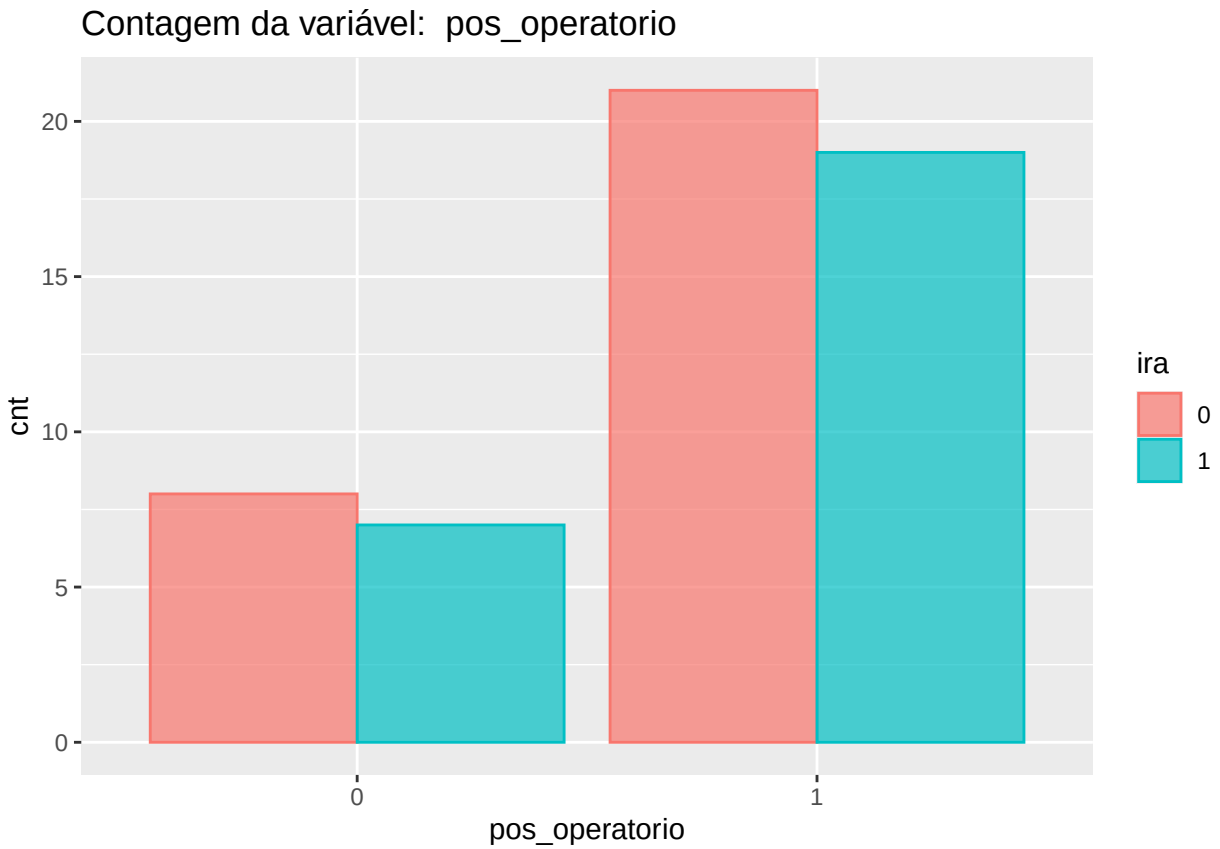

```
## `summarise()` regrouping output by 'criterio_ira' (override with `.groups` argument)
```




```
## `summarise()` regrouping output by 'pos_operatorio' (override with `.groups` argument)
```

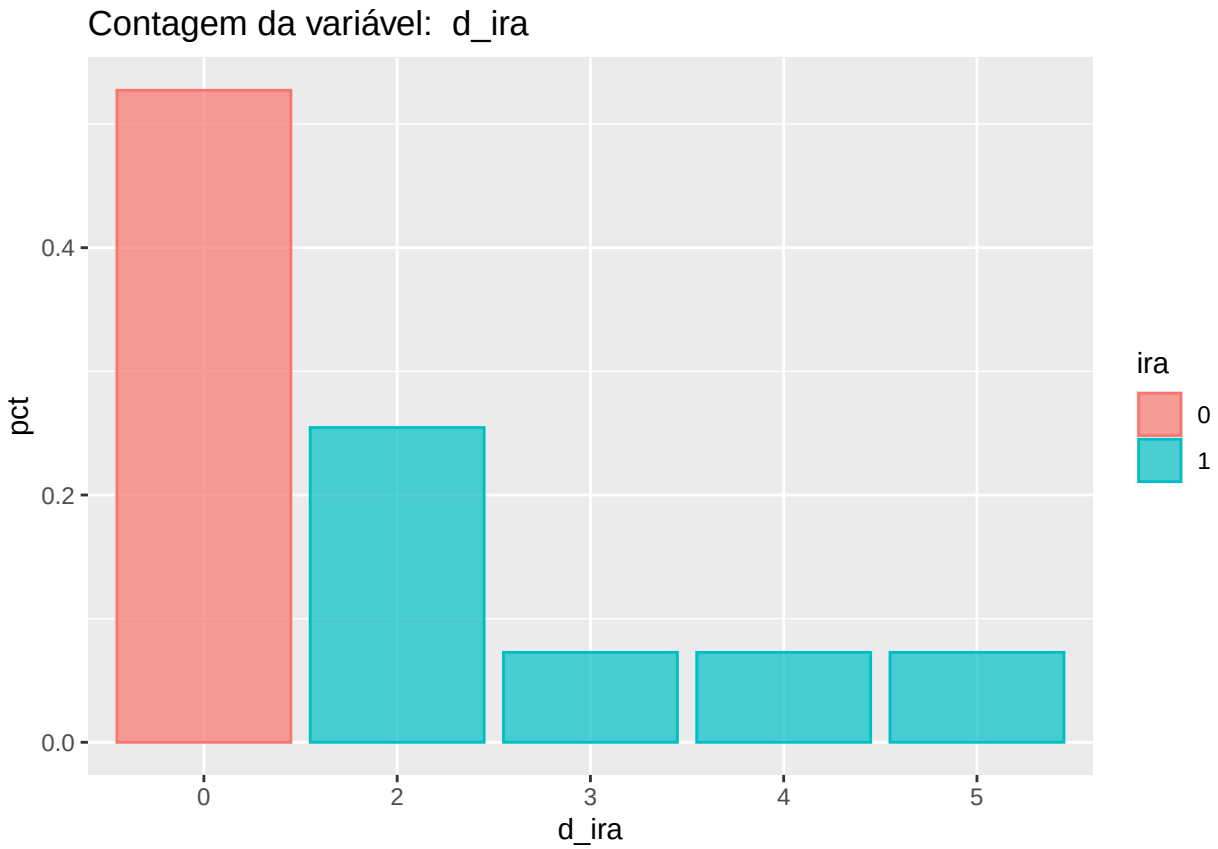



```
## `summarise()` regrouping output by 'mortalidade_intra_hospitalar' (override with `.groups` argument)
```

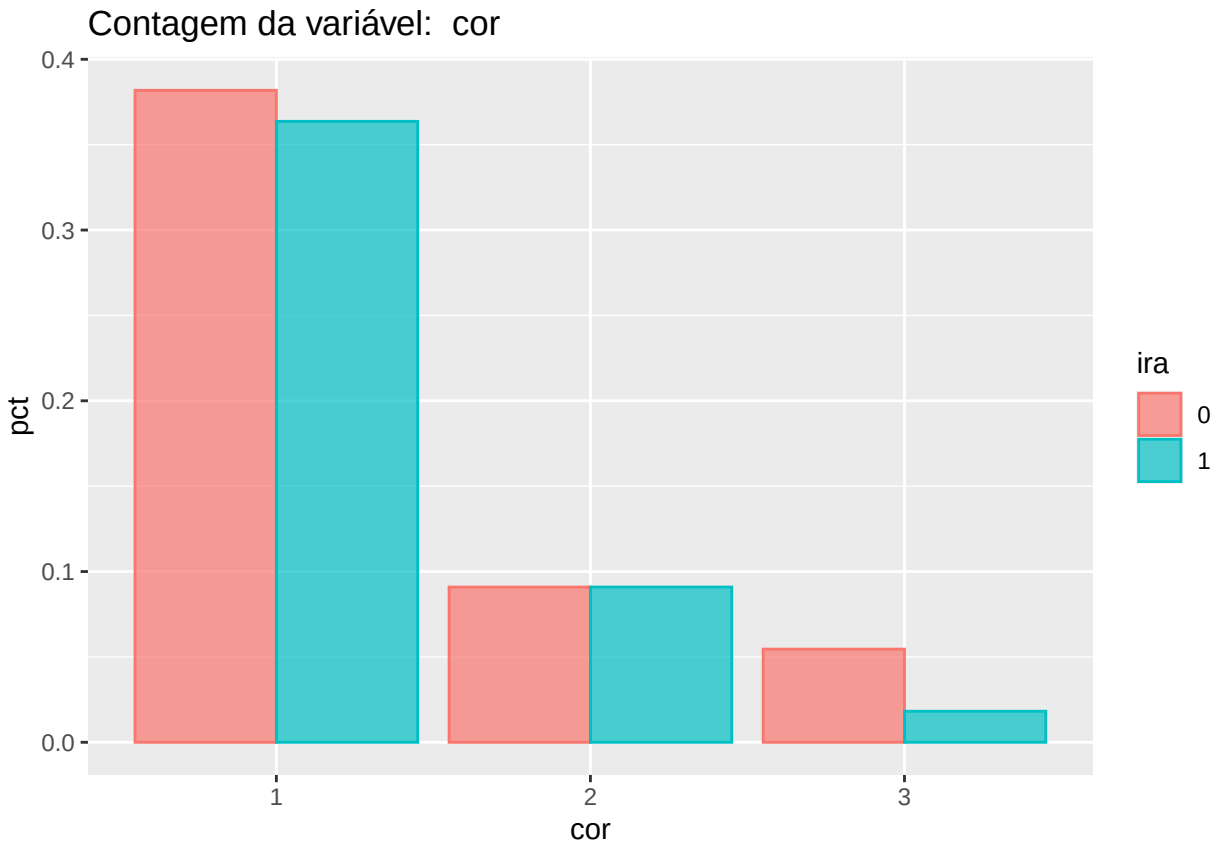

Contagem da variável: pos_operatorio
cnt
20
15
10
5
0
0
1
pos_operatorio
ira
0
1


```
for(column in categorical_columns[categorical_columns != 'D_IRA']){
    p = df %>%
        group_by(!!sym(column), ira) %>%
        summarise(cnt = n()) %>%
        ungroup %>%
        mutate(pct = cnt / sum(cnt)) %>%
        ggplot(aes(x=!!sym(column), y=pct,
                   color=ira, fill=ira)) +
            geom_col(alpha=0.7, position='dodge') +
            labs(title=paste('Contagem da variável: ', column))
    print(p)
}
```

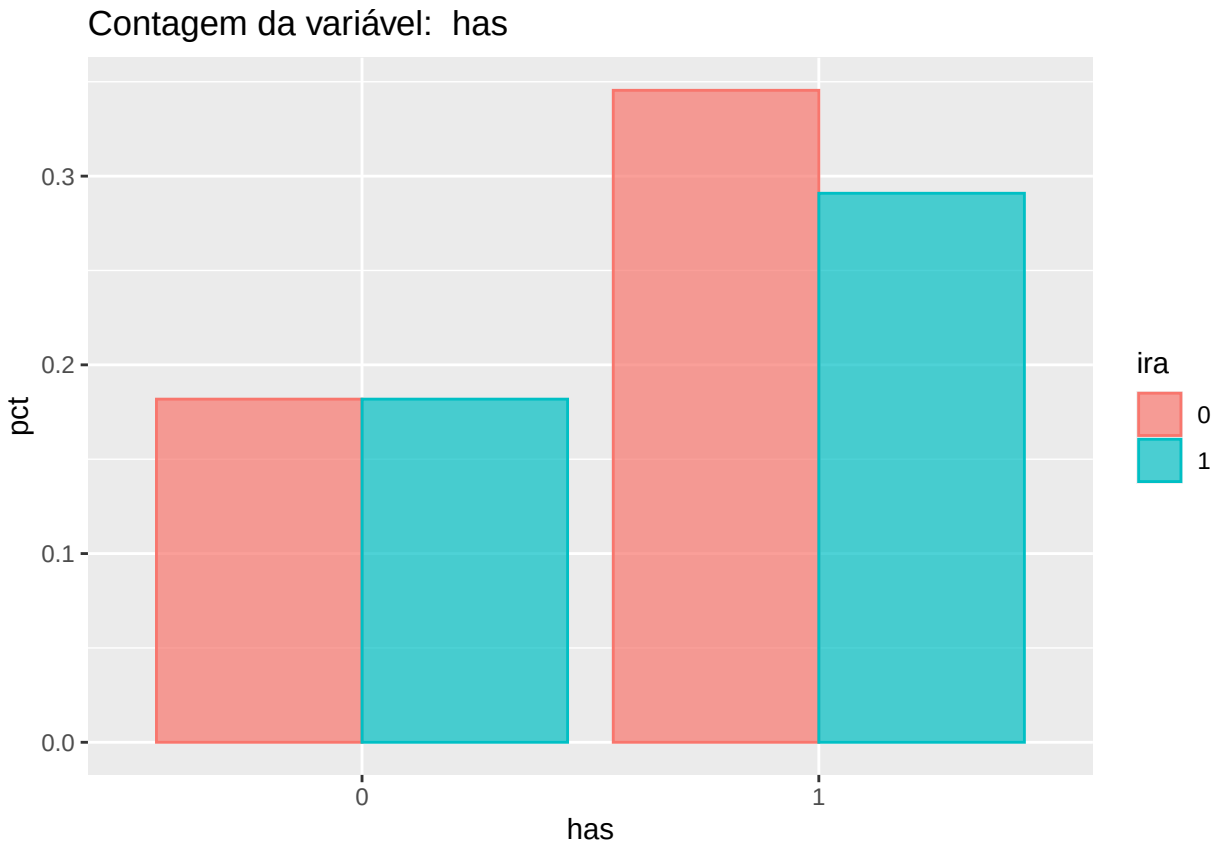
```
## `summarise()` regrouping output by 'd_ira' (override with `.groups` argument)
## `summarise()` regrouping output by 'cor' (override with `.groups` argument)
```



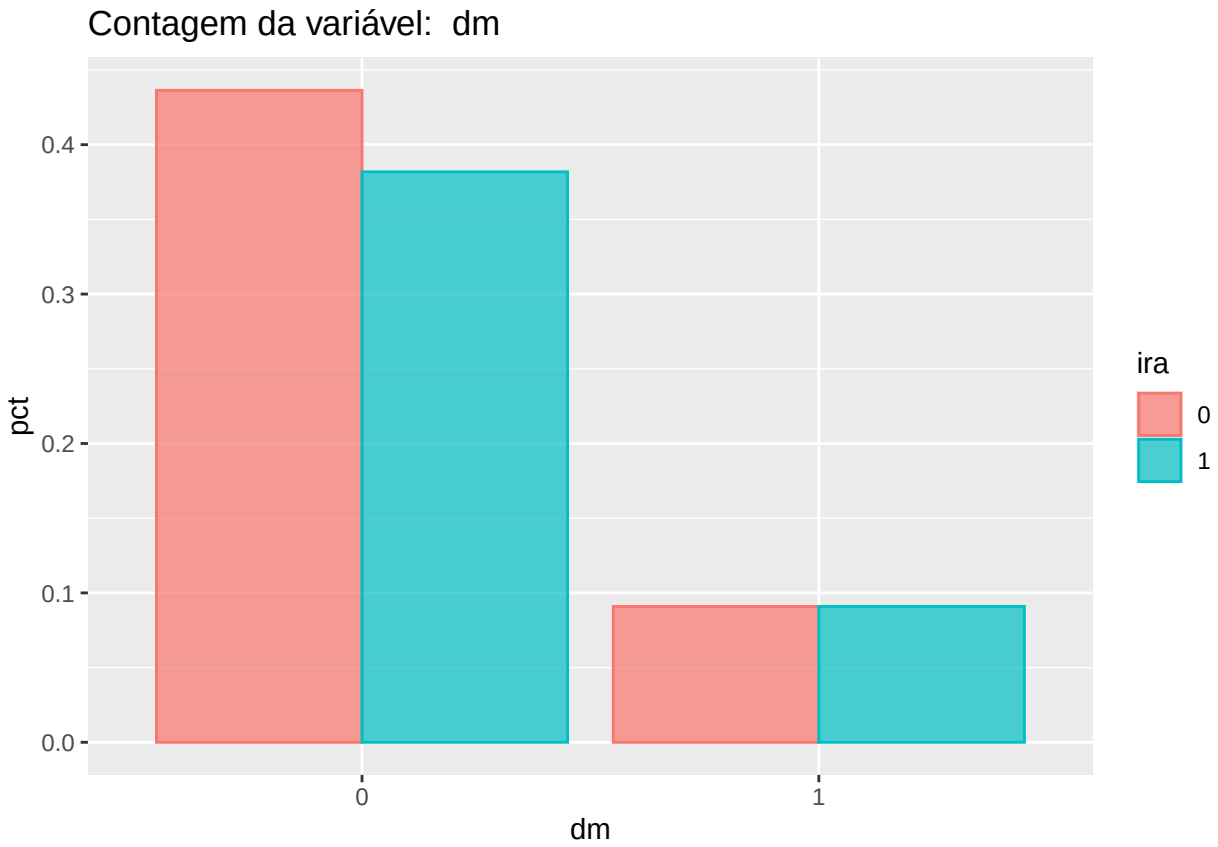

```
## `summarise()` regrouping output by 'has' (override with `.groups` argument)
```



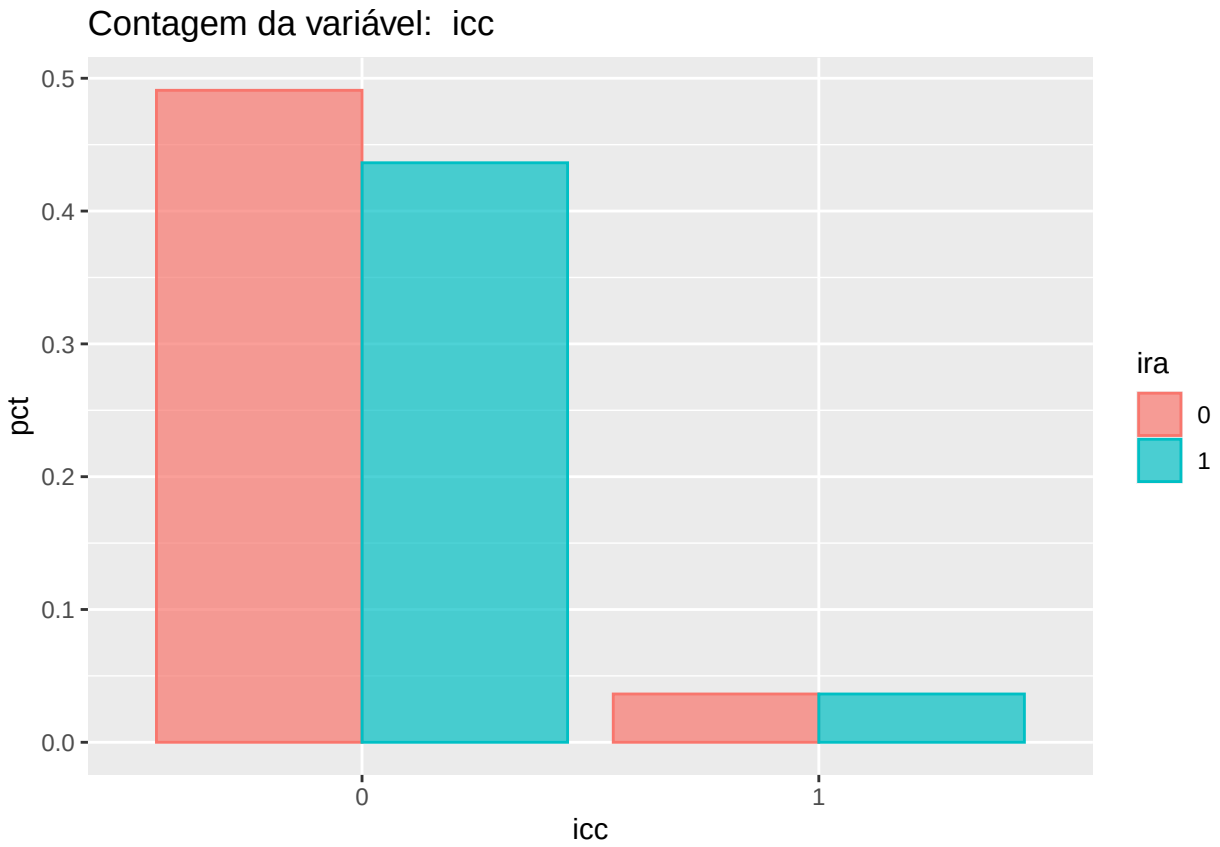

```
## `summarise()` regrouping output by 'dm' (override with `.groups` argument)
```



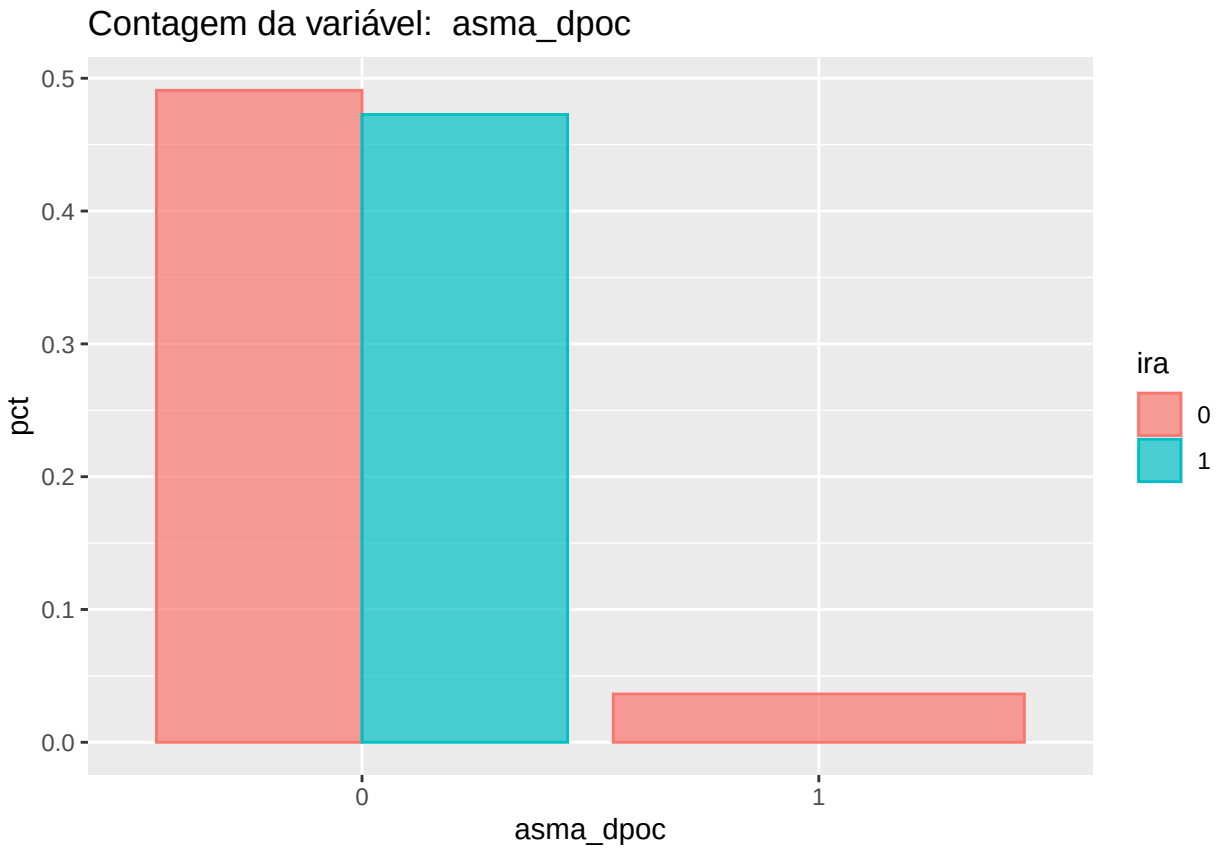

```
## `summarise()` regrouping output by 'icc' (override with `.groups` argument)
```



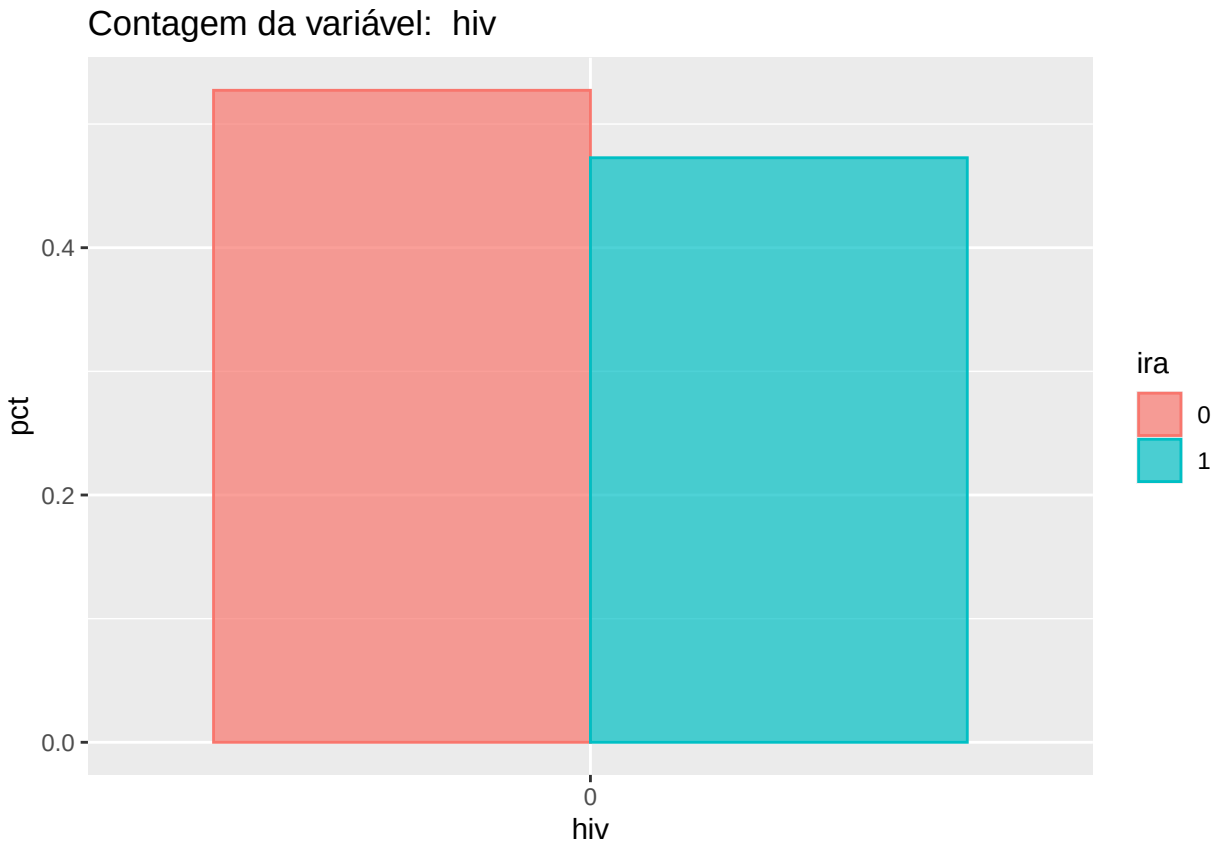

```
## `summarise()` regrouping output by 'asma_dpoc' (override with `.groups` argument)
```



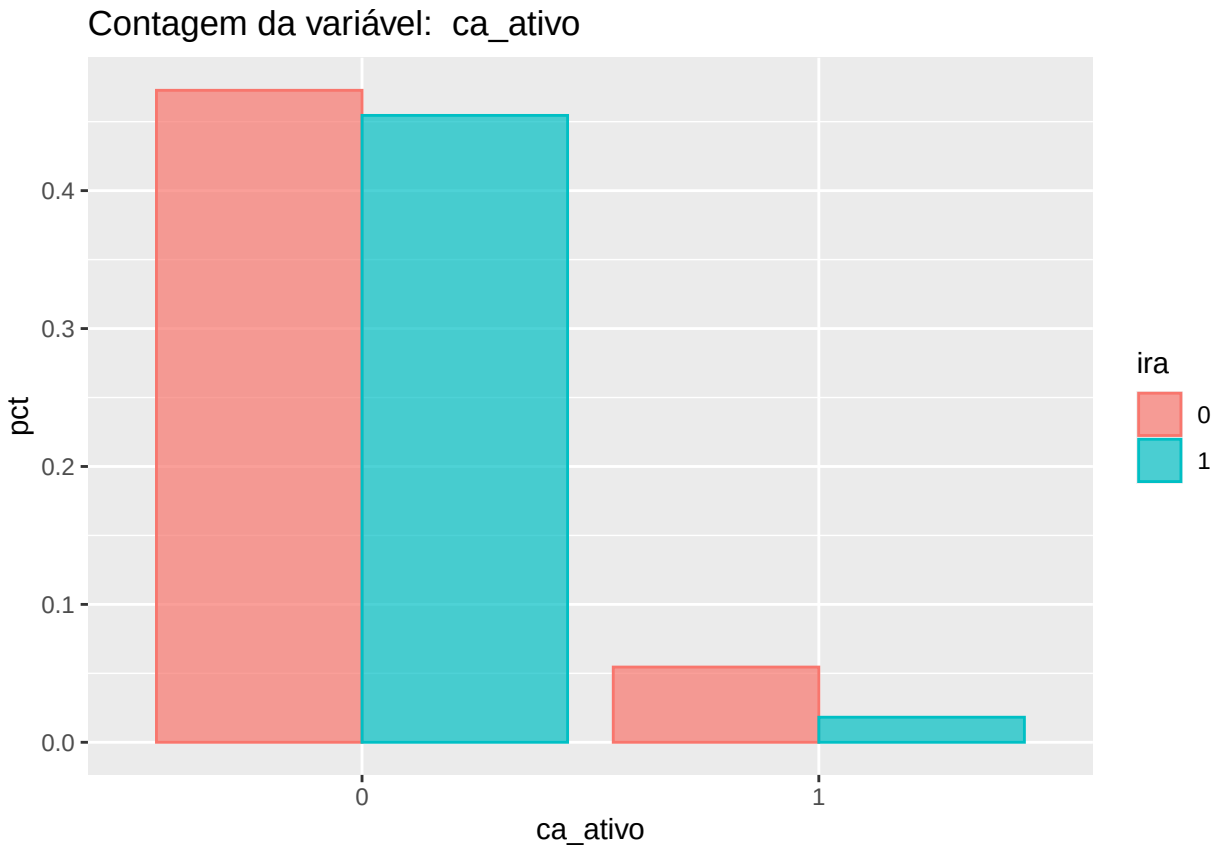

```
## `summarise()` regrouping output by 'hiv' (override with `.groups` argument)
```



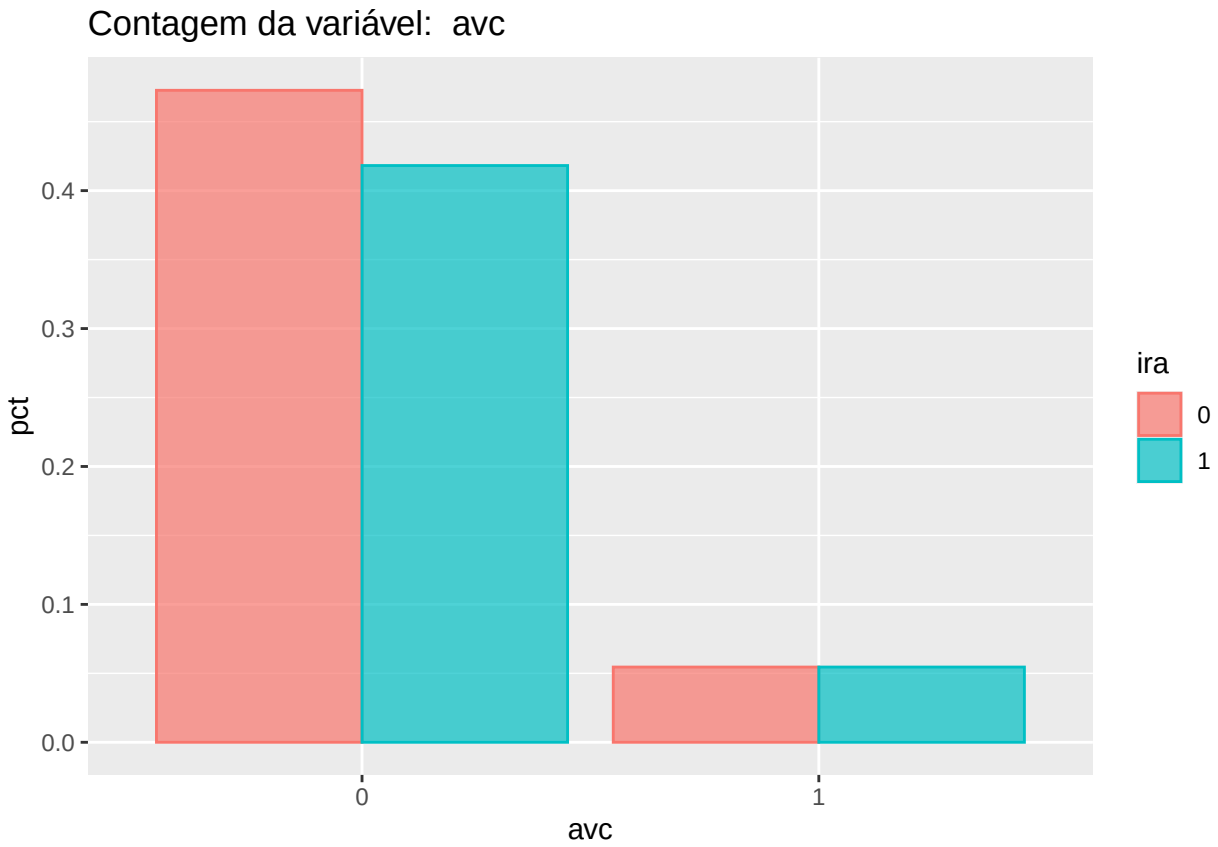

```
## `summarise()` regrouping output by 'ca_ativo' (override with `.groups` argument)
```



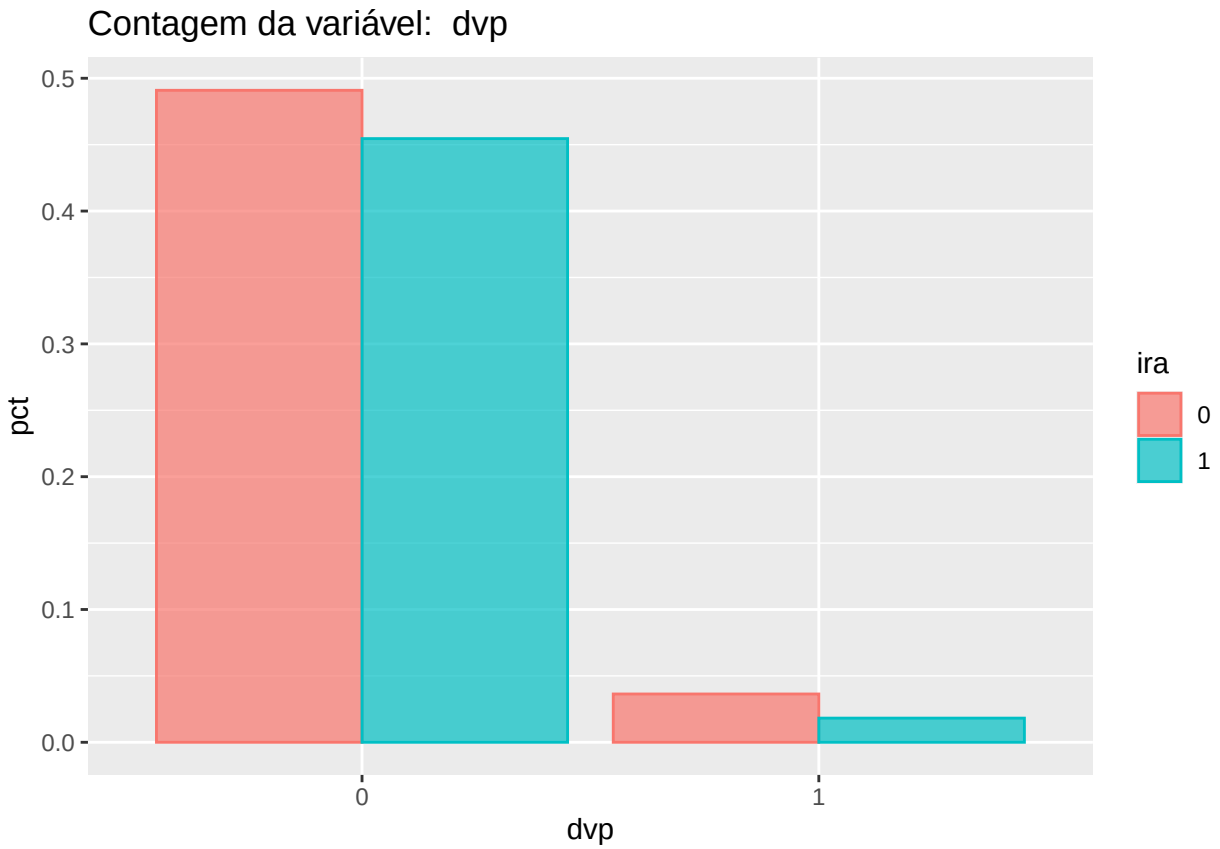

```
## `summarise()` regrouping output by 'avc' (override with `.groups` argument)
```



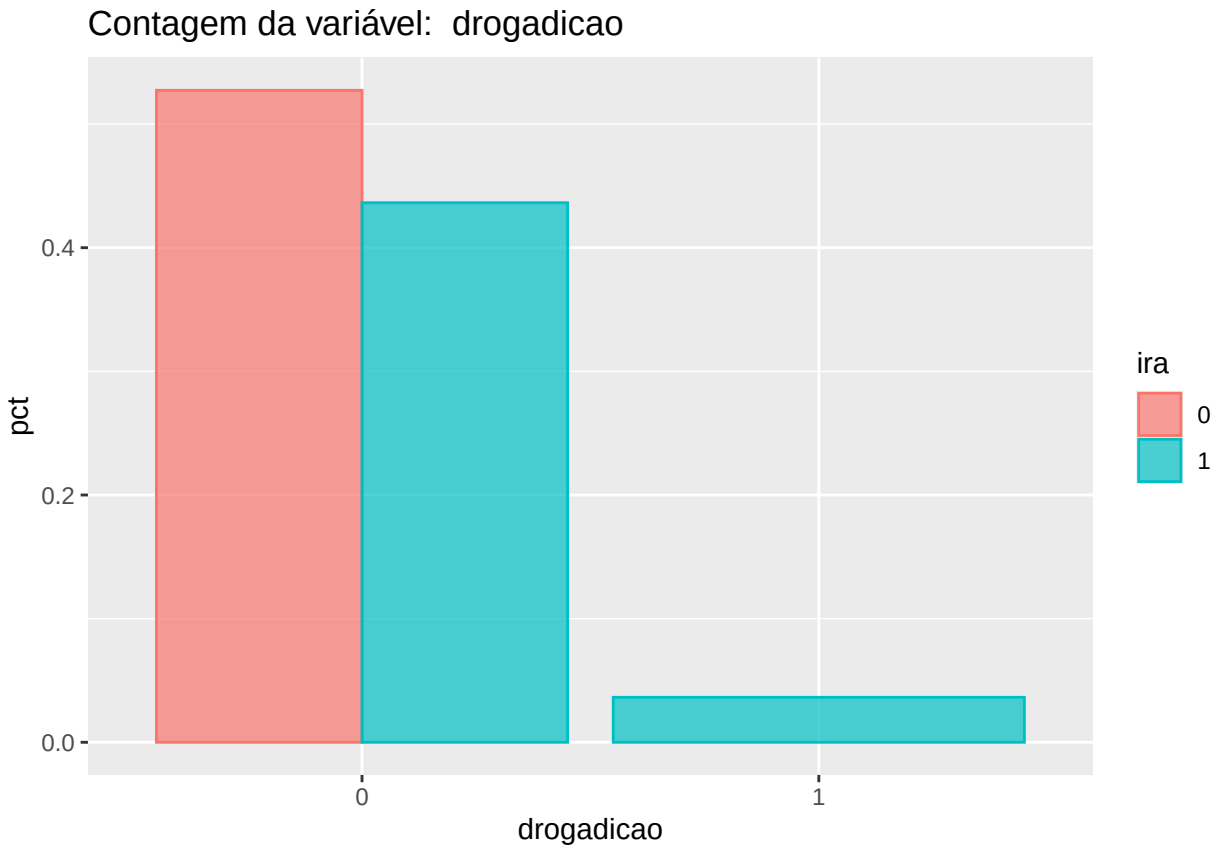

```
## `summarise()` regrouping output by 'dvp' (override with `.groups` argument)
```



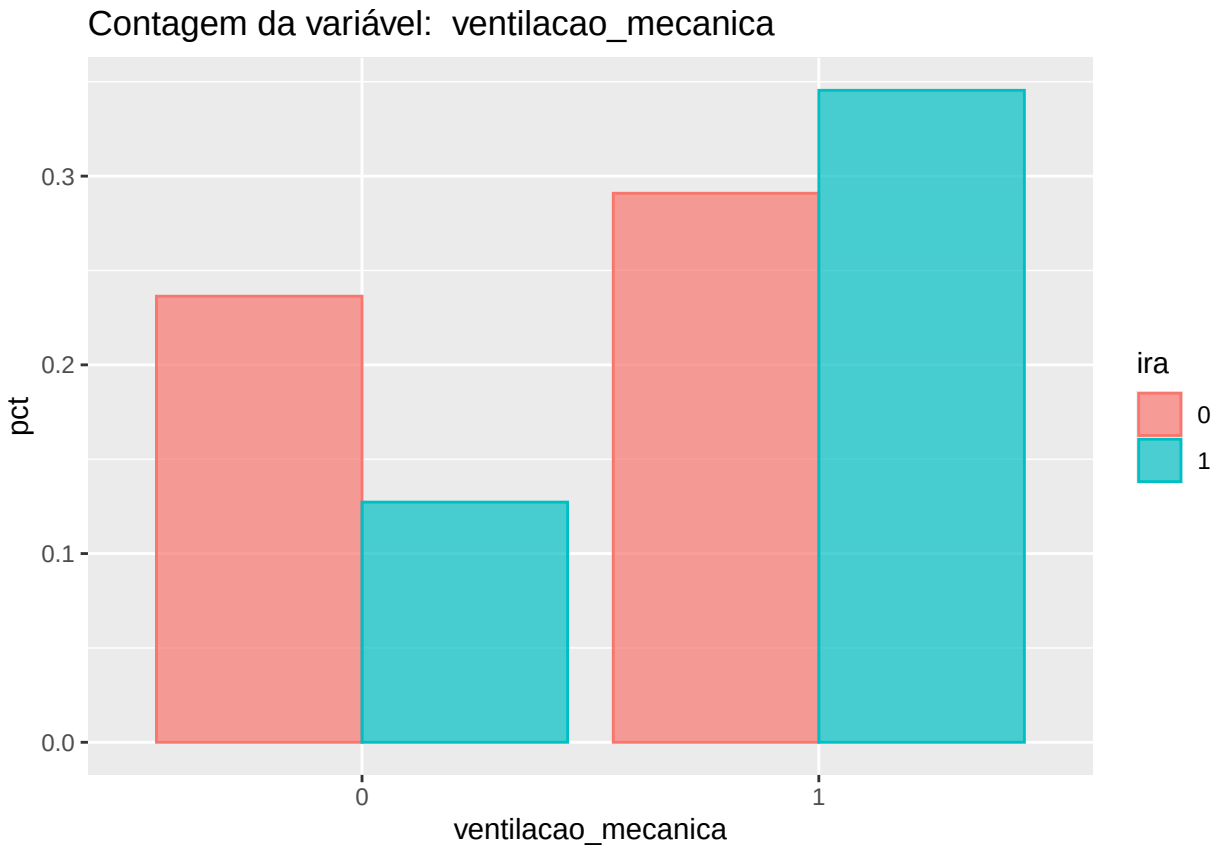

```
## `summarise()` regrouping output by 'drogadicao' (override with `.groups` argument)
```



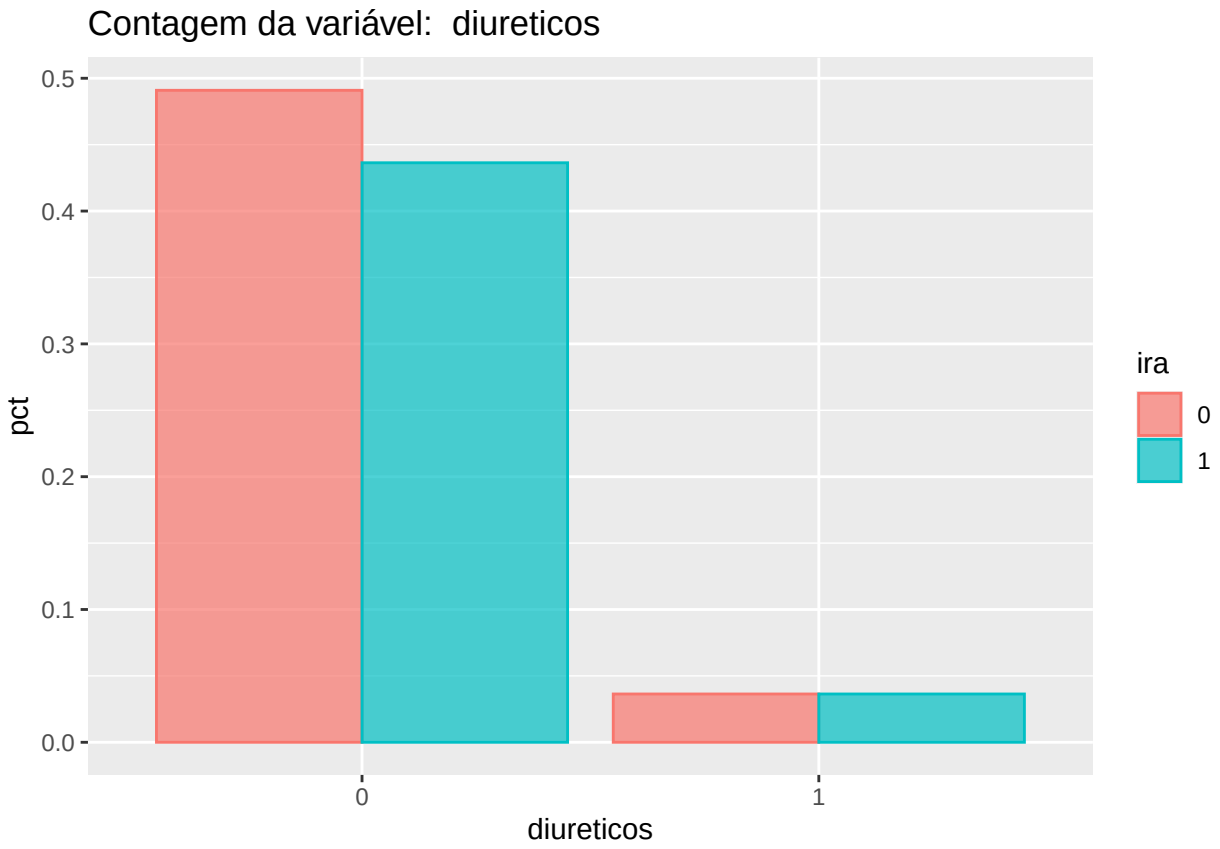

```
## `summarise()` regrouping output by 'ventilacao_mecanica' (override with `.groups` argument)
```



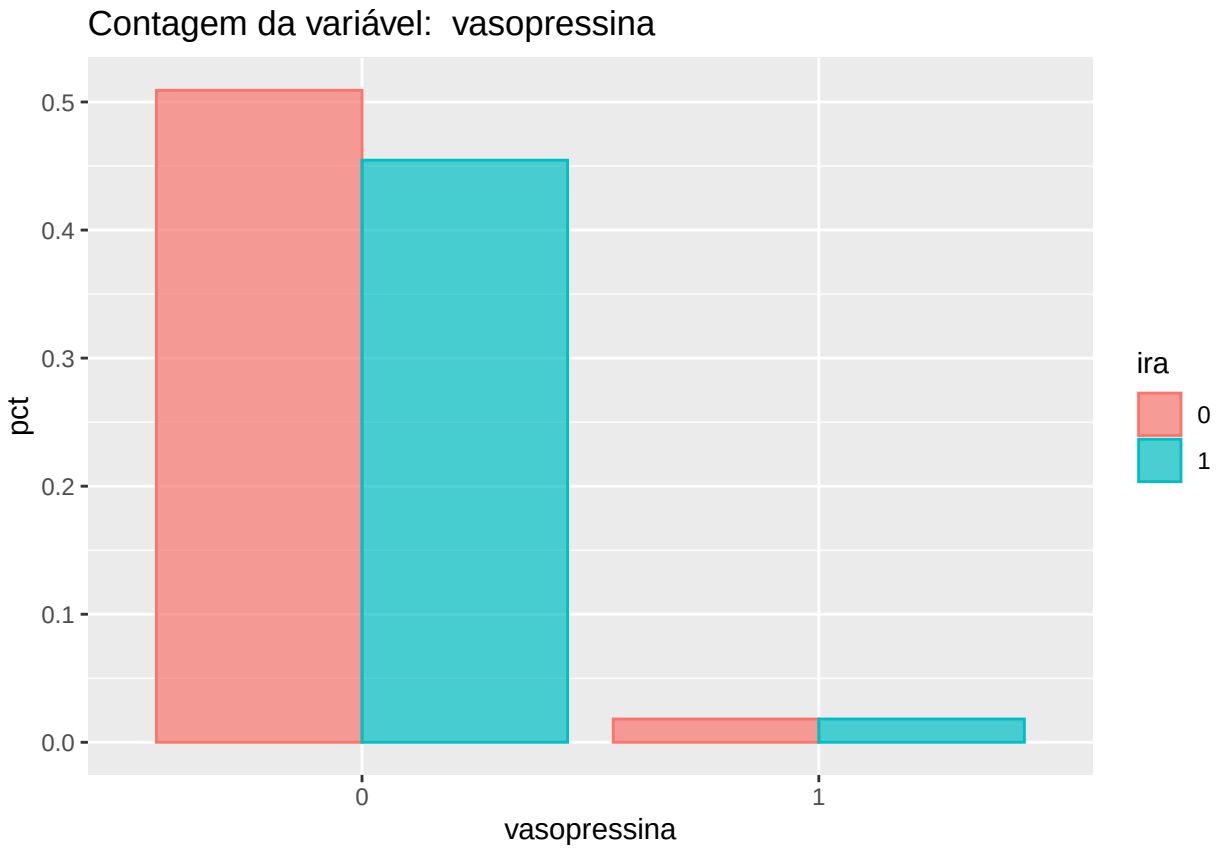

```
## `summarise()` regrouping output by 'diureticos' (override with `.groups` argument)
```



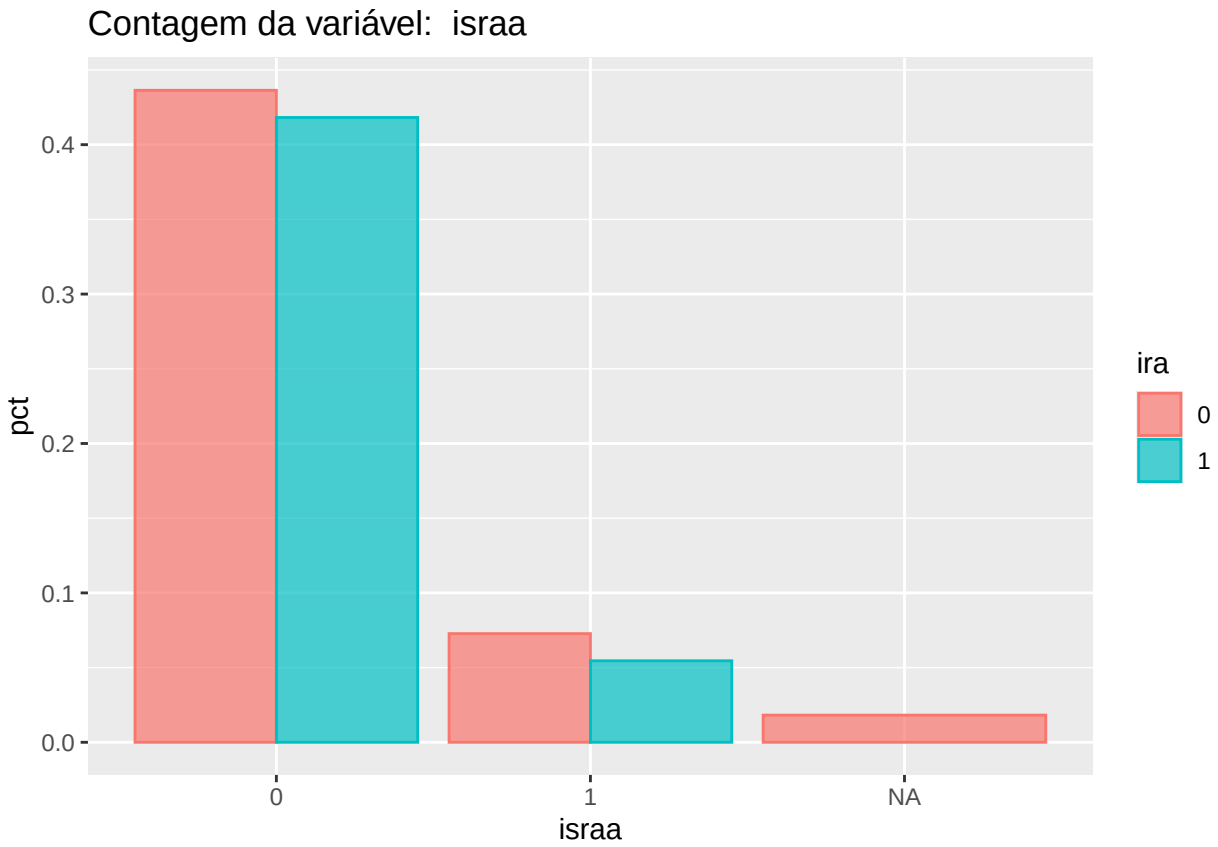

```
## `summarise()` regrouping output by 'vasopressina' (override with `.groups` argument)
```



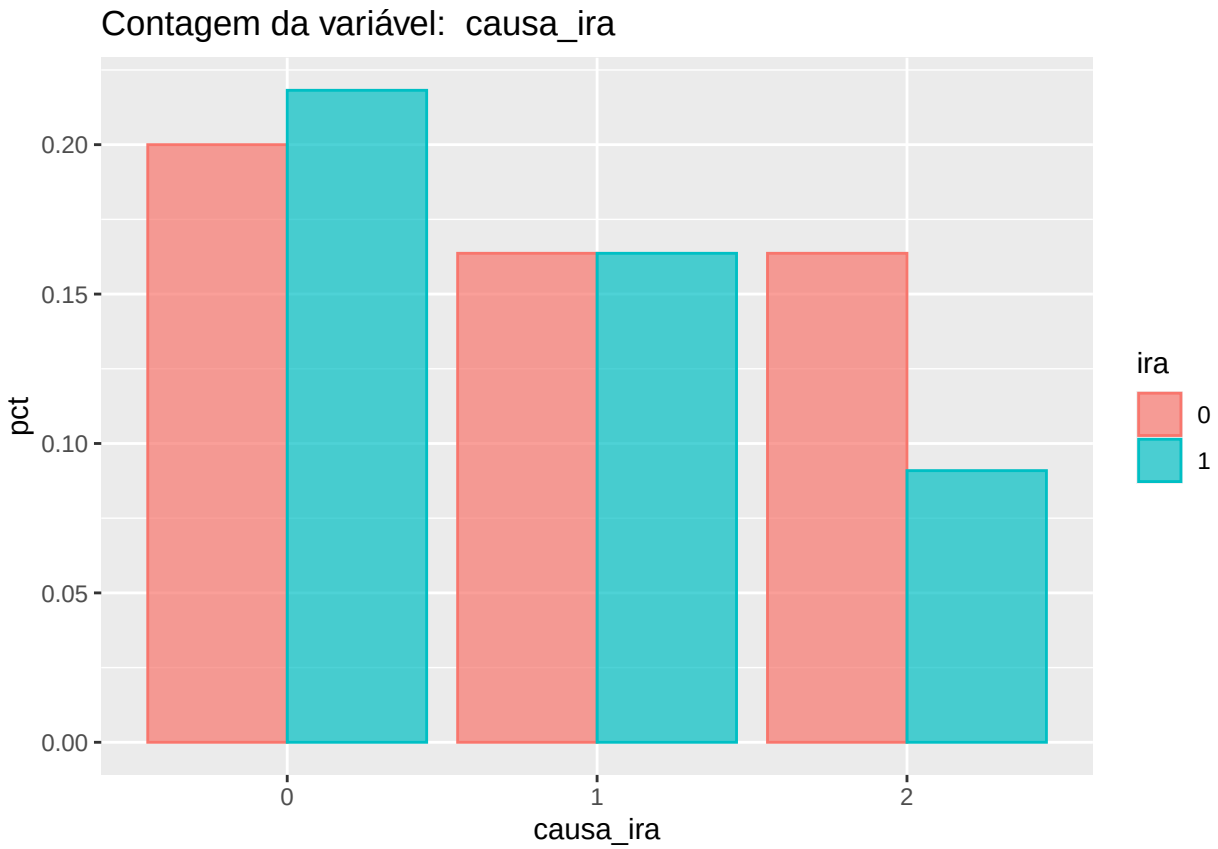

```
## `summarise()` regrouping output by 'israa' (override with `.groups` argument)
```



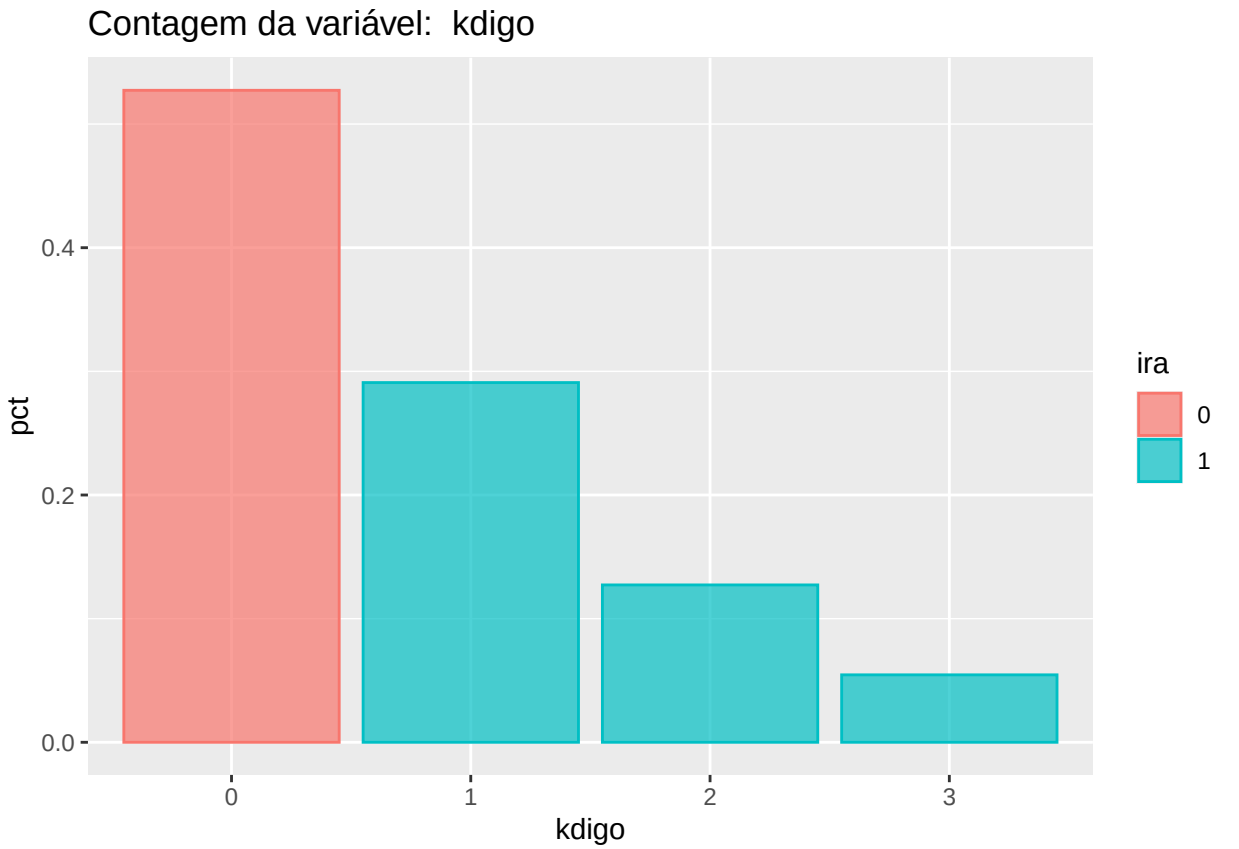

```
## `summarise()` regrouping output by 'causa_ira' (override with `.groups` argument)
```



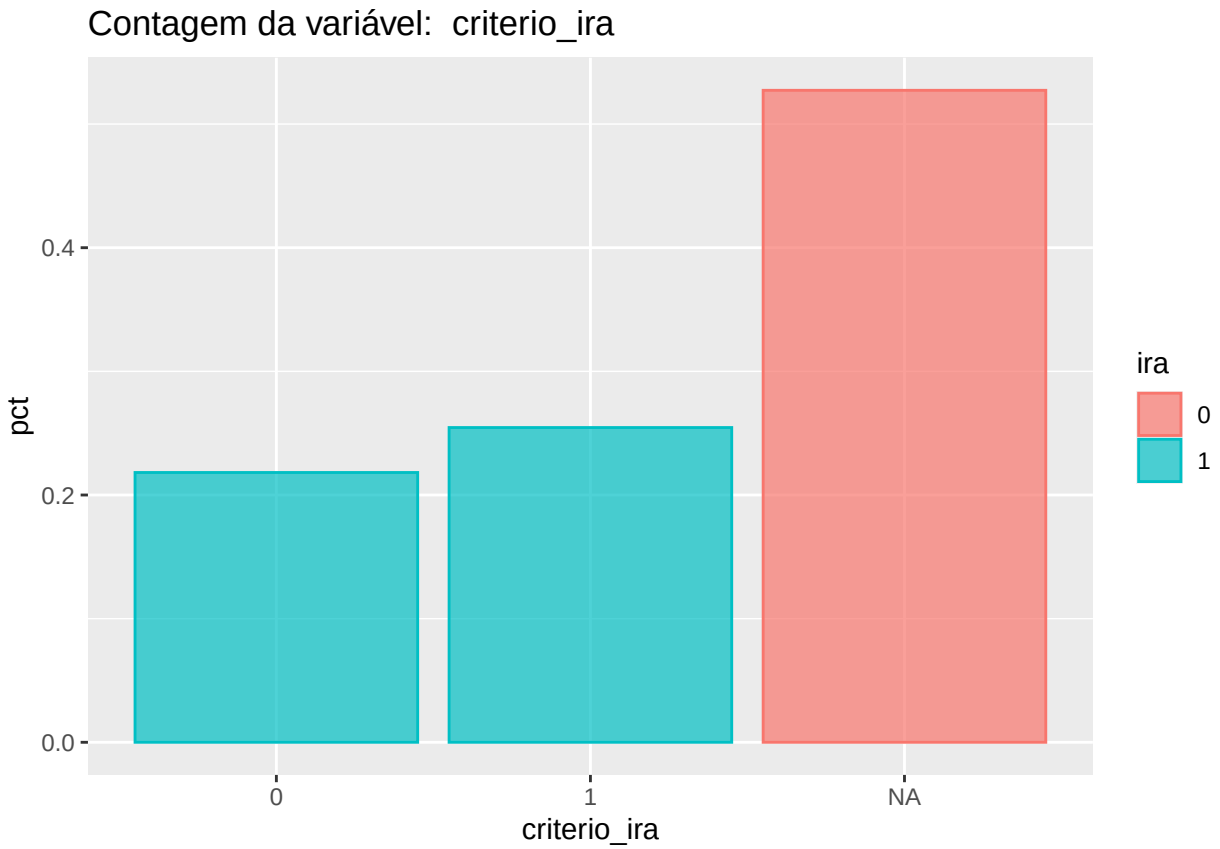

```
## `summarise()` regrouping output by 'kdigo' (override with `.groups` argument)
```



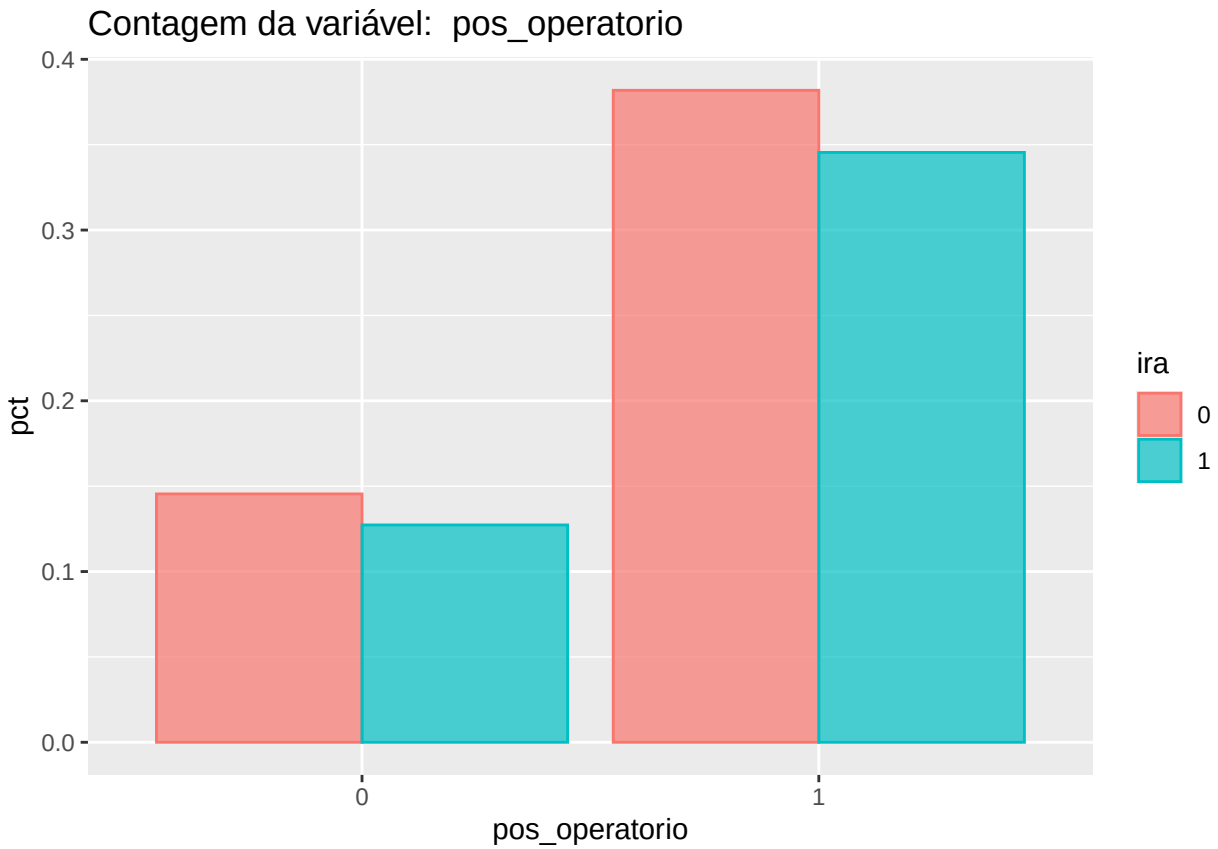

```
## `summarise()` regrouping output by 'criterio_ira' (override with `.groups` argument)
```

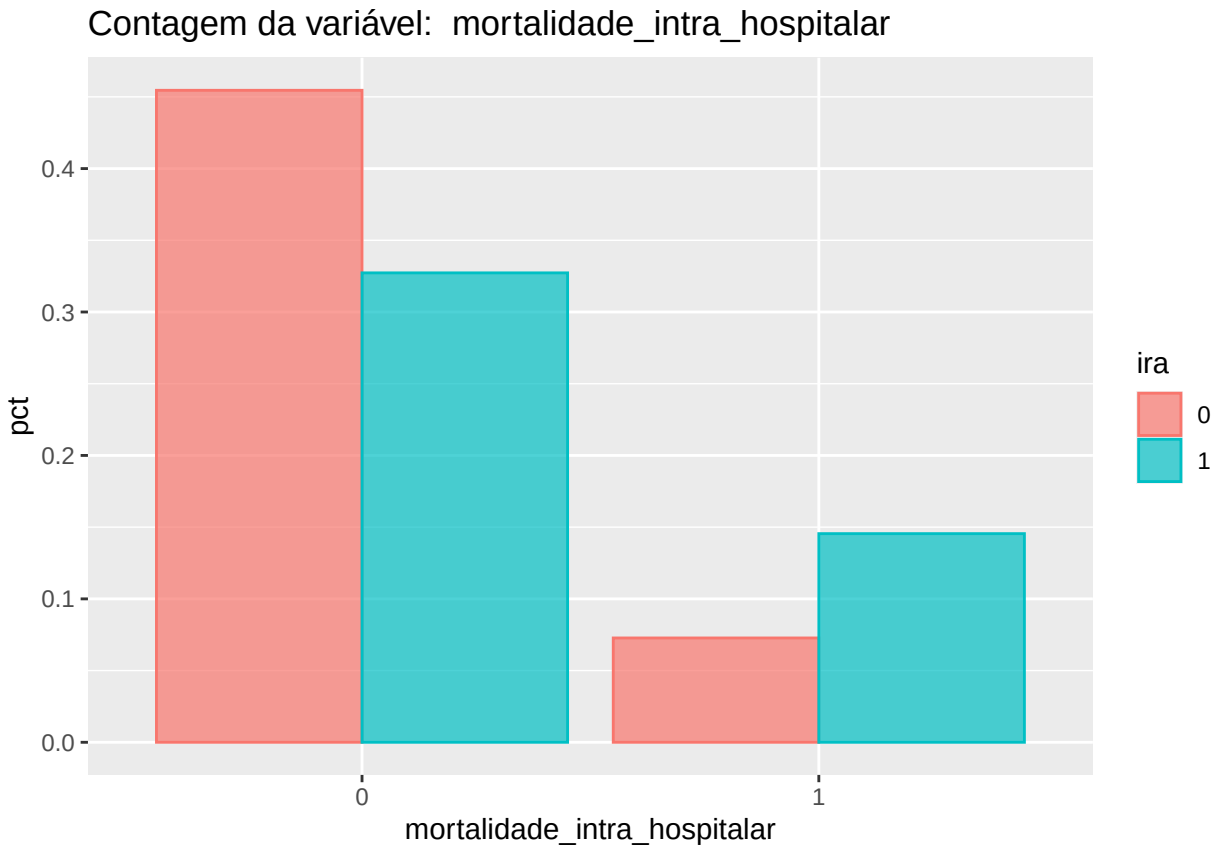



```
## `summarise()` regrouping output by 'pos_operatorio' (override with `.groups` argument)
```




```
## `summarise()` regrouping output by 'mortalidade_intra_hospitalar' (override with `.groups` argument)
```

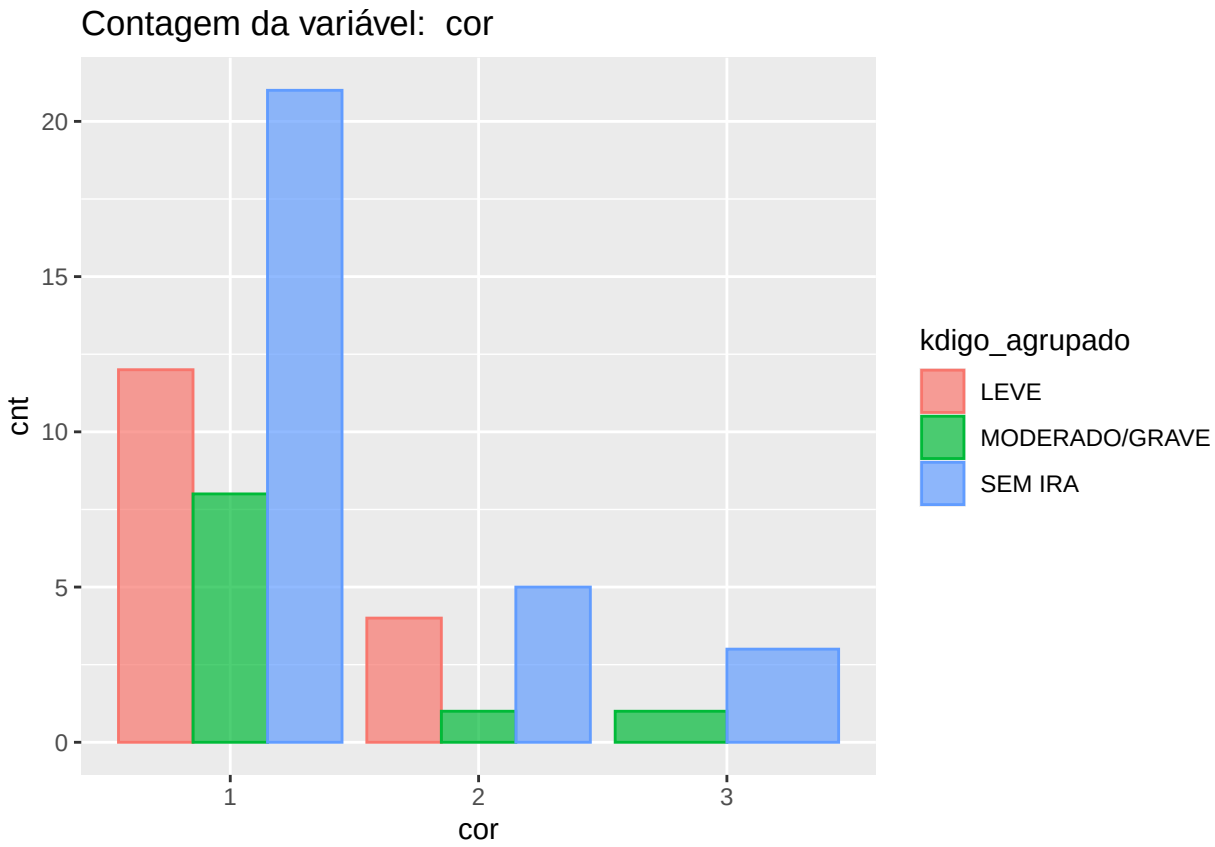

Contagem da variável: pos_operatorio
0.4
0.3
0.2
0.1
0.0
pct
0
1
pos_operatorio
ira
0
1




```
for(column in categorical_columns[categorical_columns != 'd_ira']){
    p = df %>%
        group_by(!!sym(column), kdigo_agrupado) %>%
        summarise(cnt = n()) %>%
        ggplot(aes(x = !!sym(column), y = cnt,
                   color = kdigo_agrupado,
                   fill = kdigo_agrupado)) +
            geom_col(alpha = 0.7, position ='dodge') +
            labs(title = paste('Contagem da variável: ', column))
    print(p)
}
```

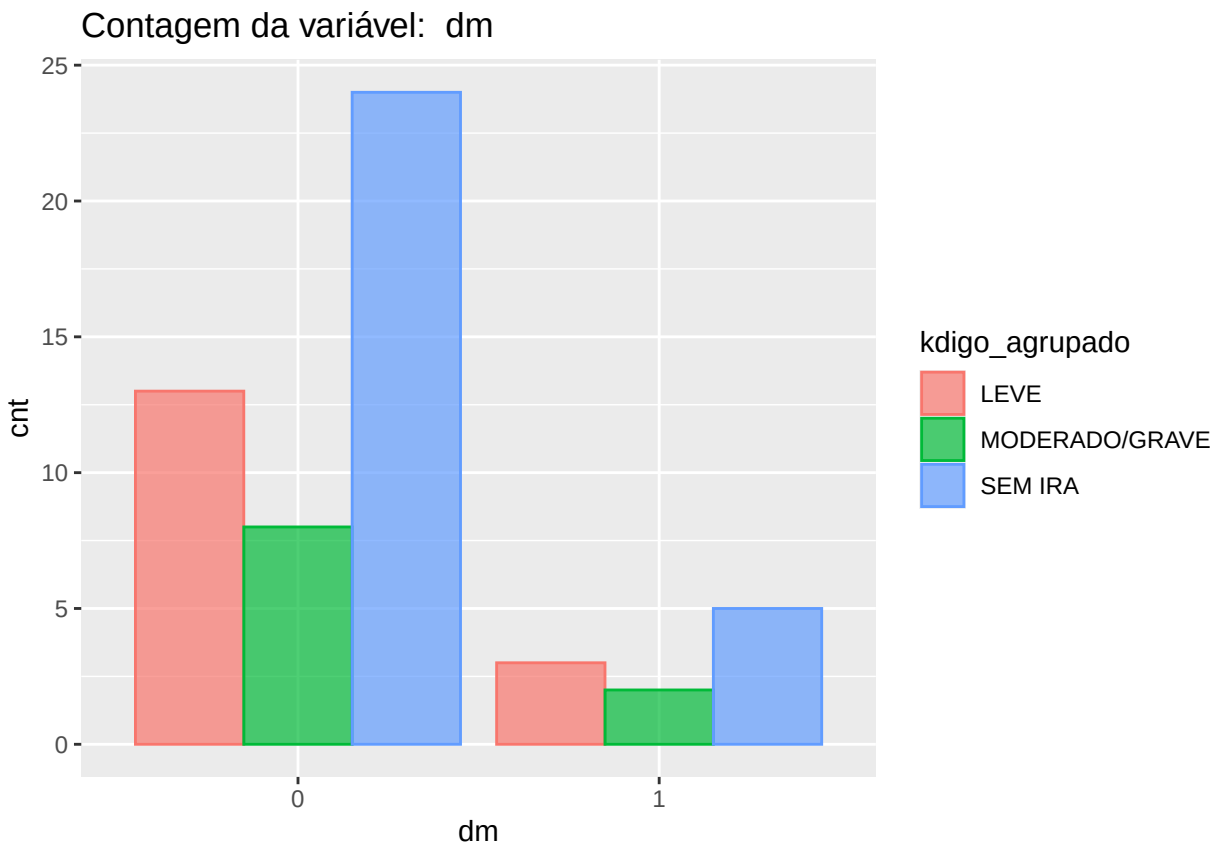
```
## `summarise()` regrouping output by 'cor' (override with `.groups` argument)
```

```
## `summarise()` regrouping output by 'has' (override with `.groups` argument)
```

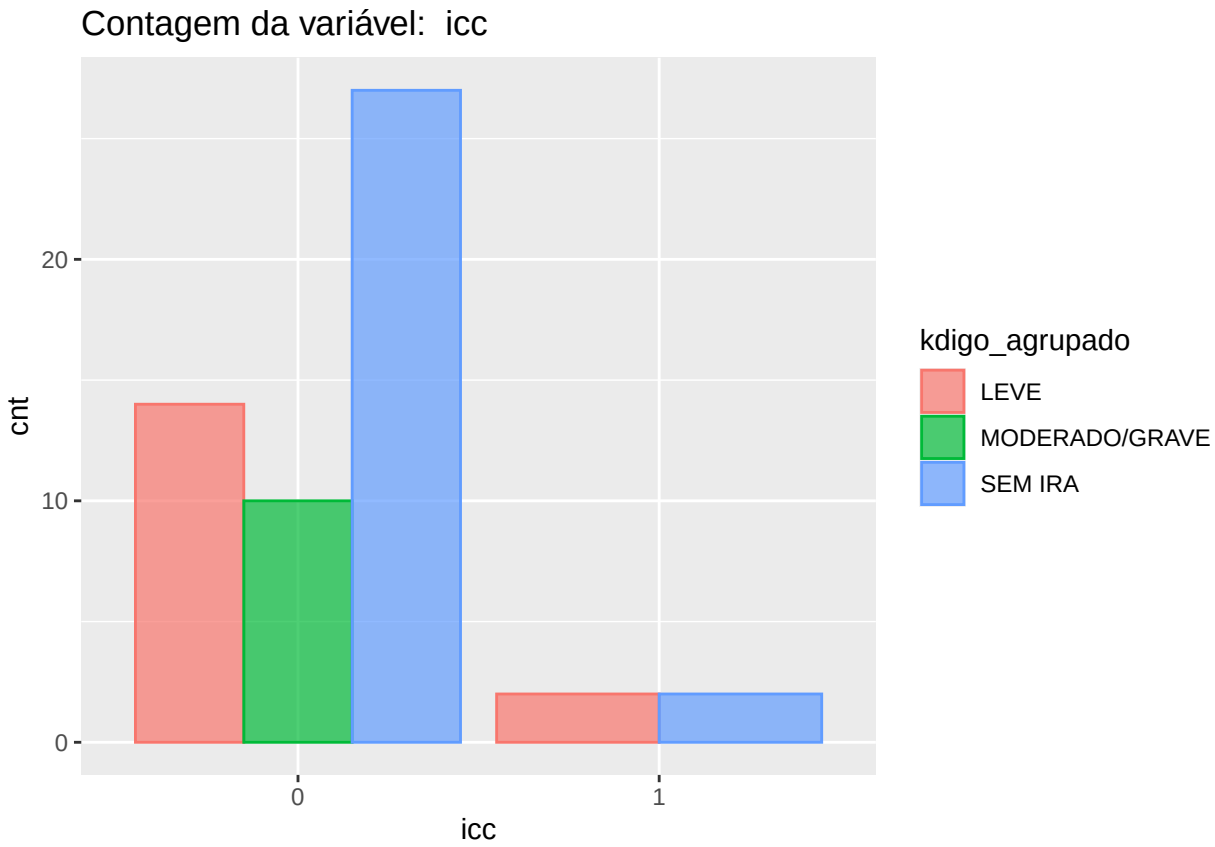



```
## `summarise()` regrouping output by 'dm' (override with `.groups` argument)
```

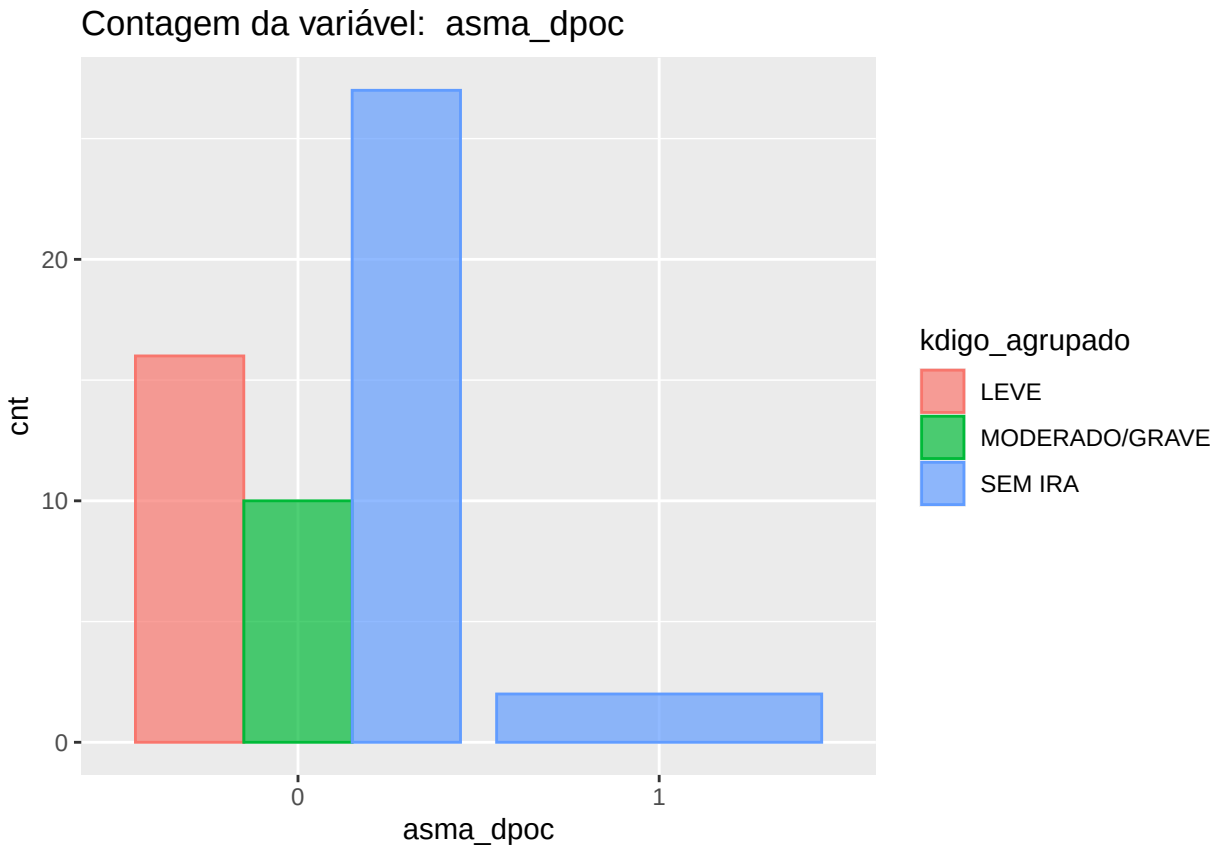

```
## `summarise()` regrouping output by 'icc' (override with `.groups` argument)
```



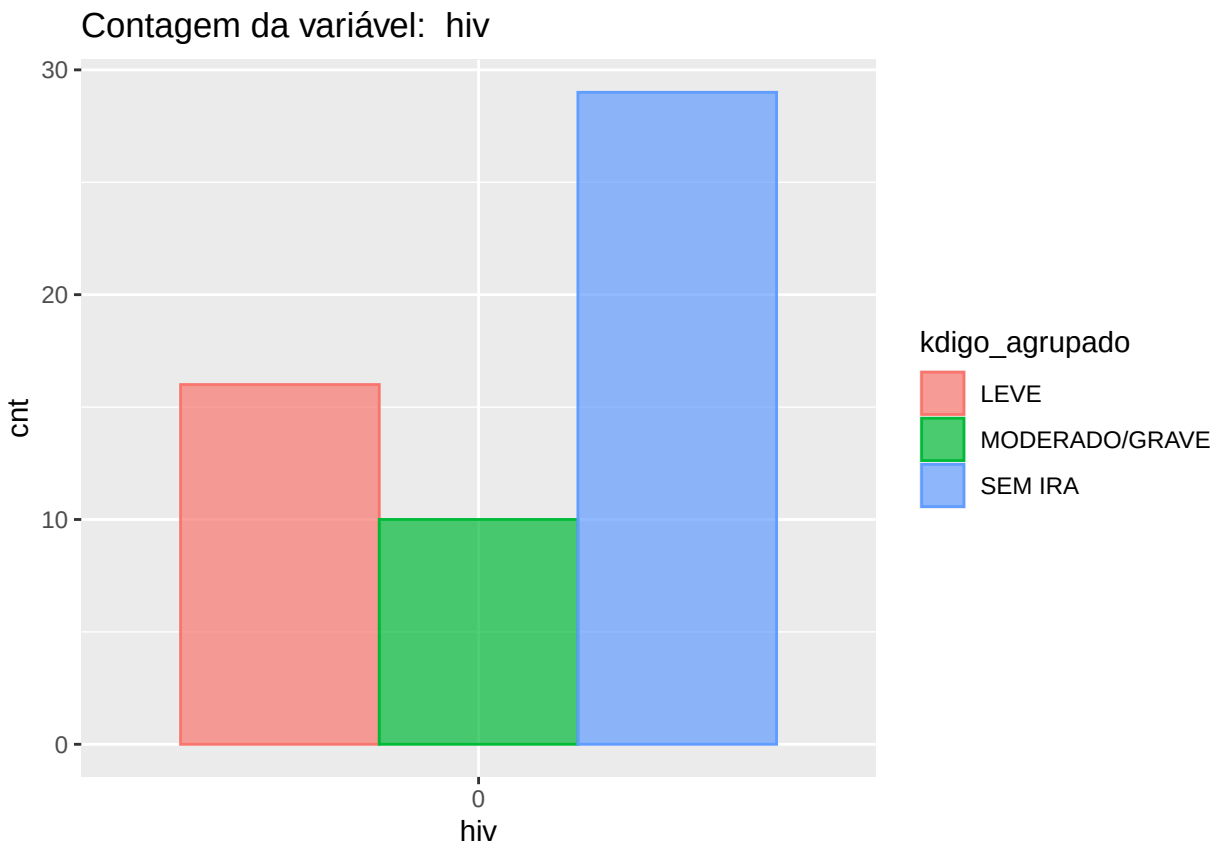

```
## `summarise()` regrouping output by 'asma_dpoc' (override with `.groups` argument)
```



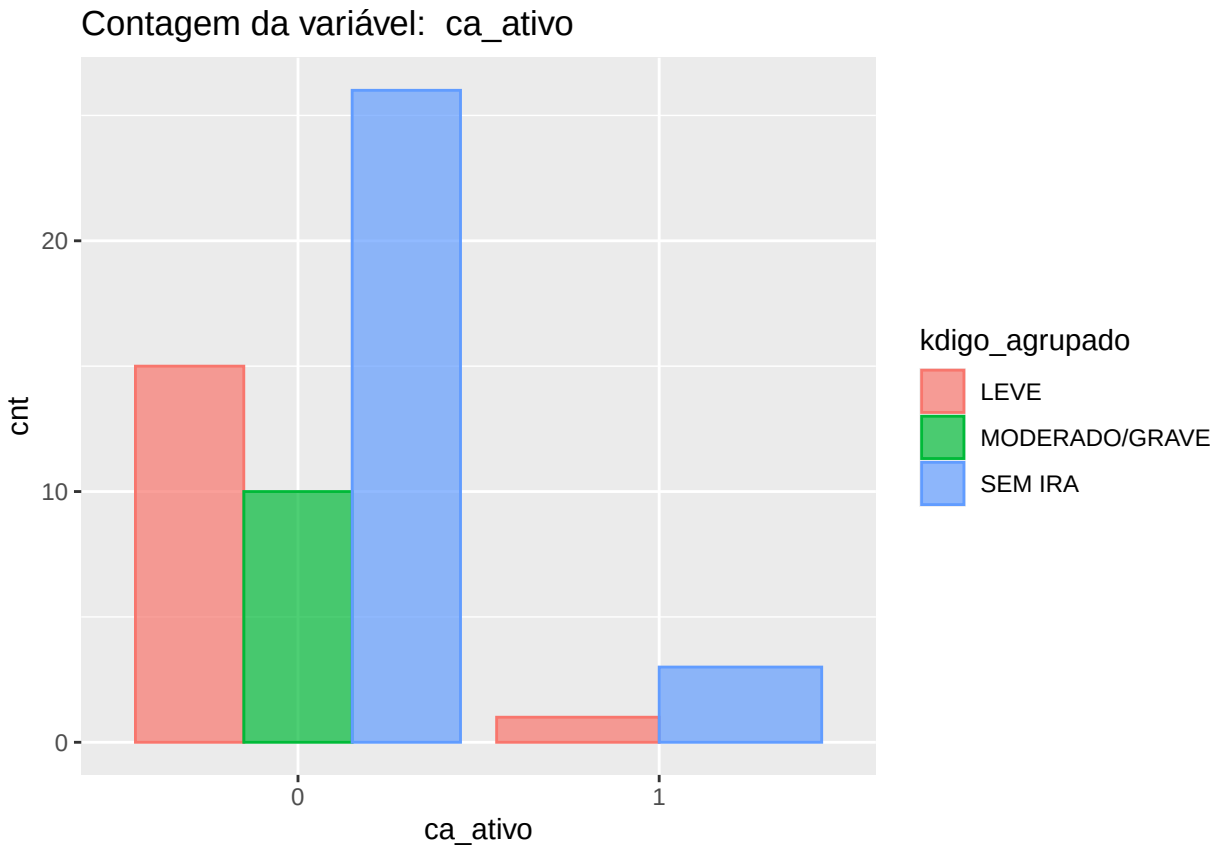

```
## `summarise()` regrouping output by 'hiv' (override with `.groups` argument)
```



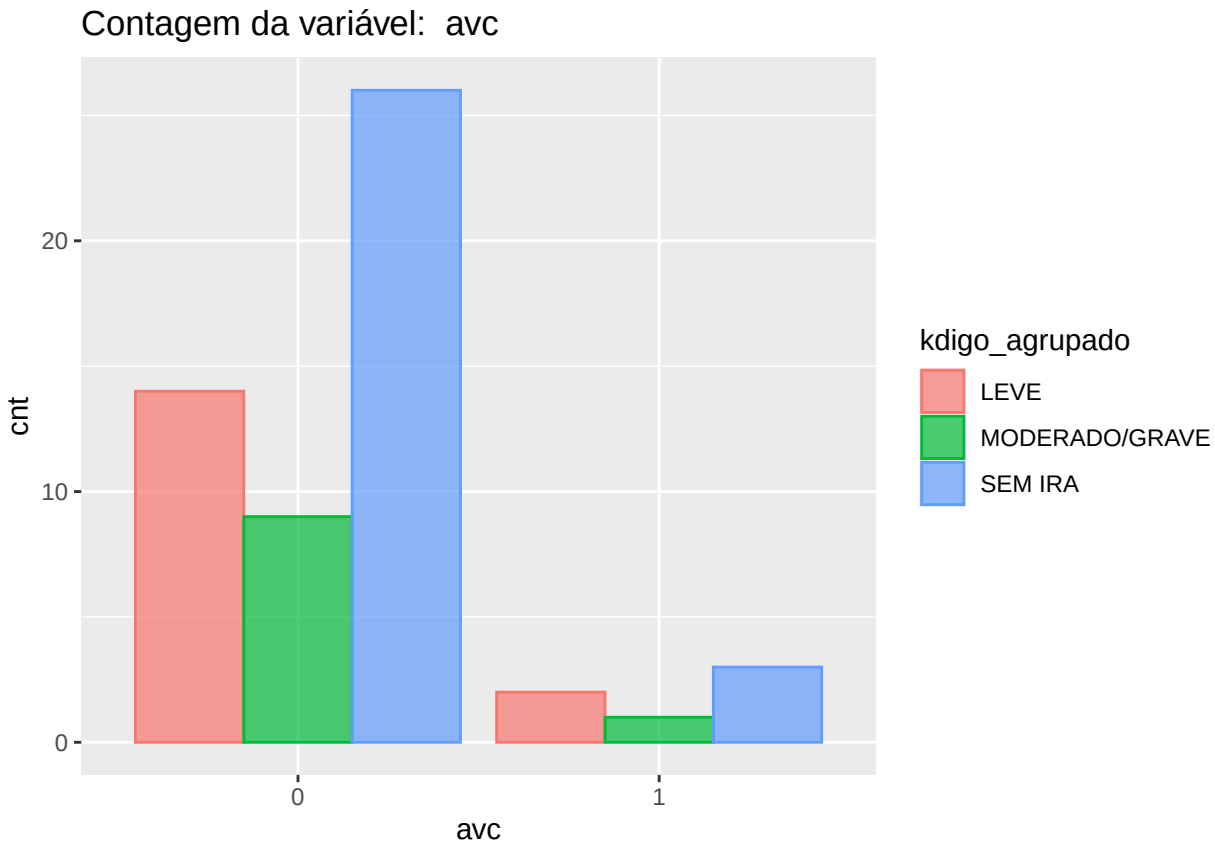

```
## `summarise()` regrouping output by 'ca_ativo' (override with `.groups` argument)
```



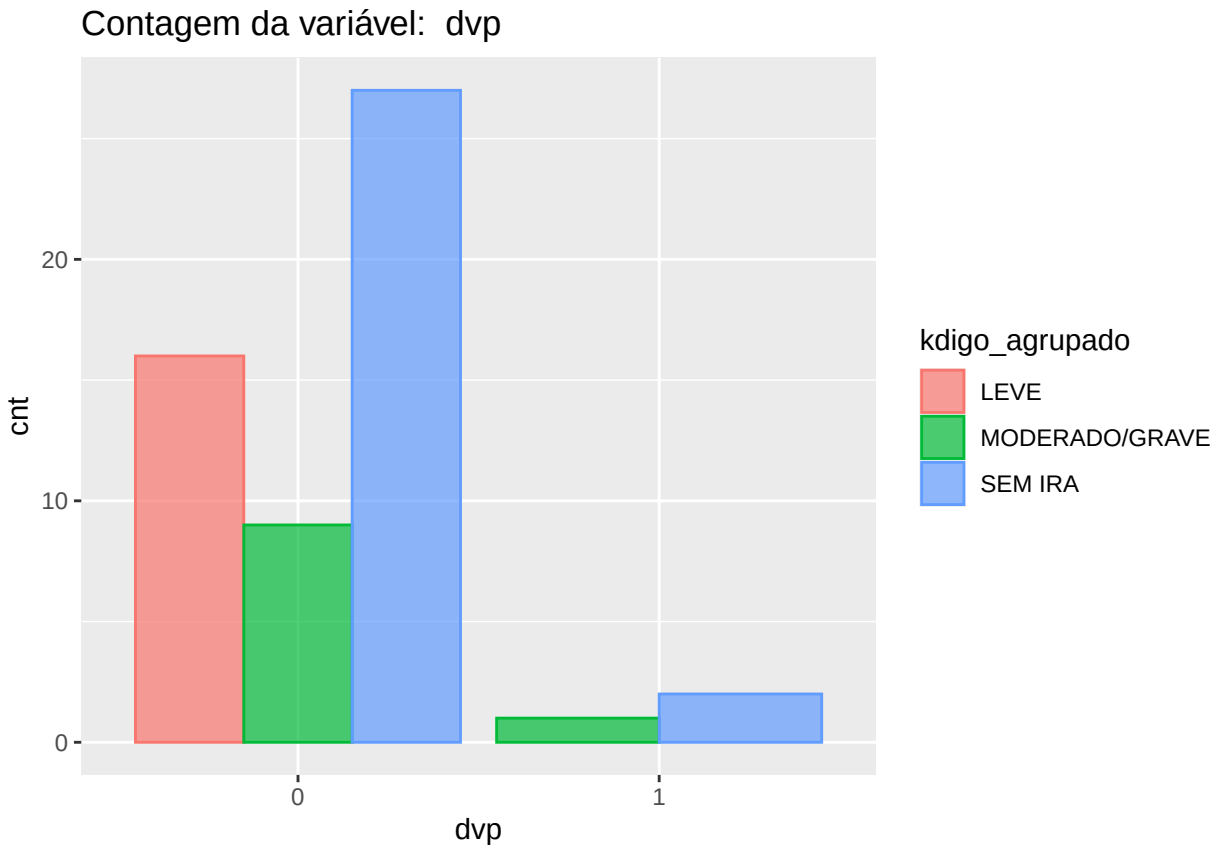

```
## `summarise()` regrouping output by 'avc' (override with `.groups` argument)
```



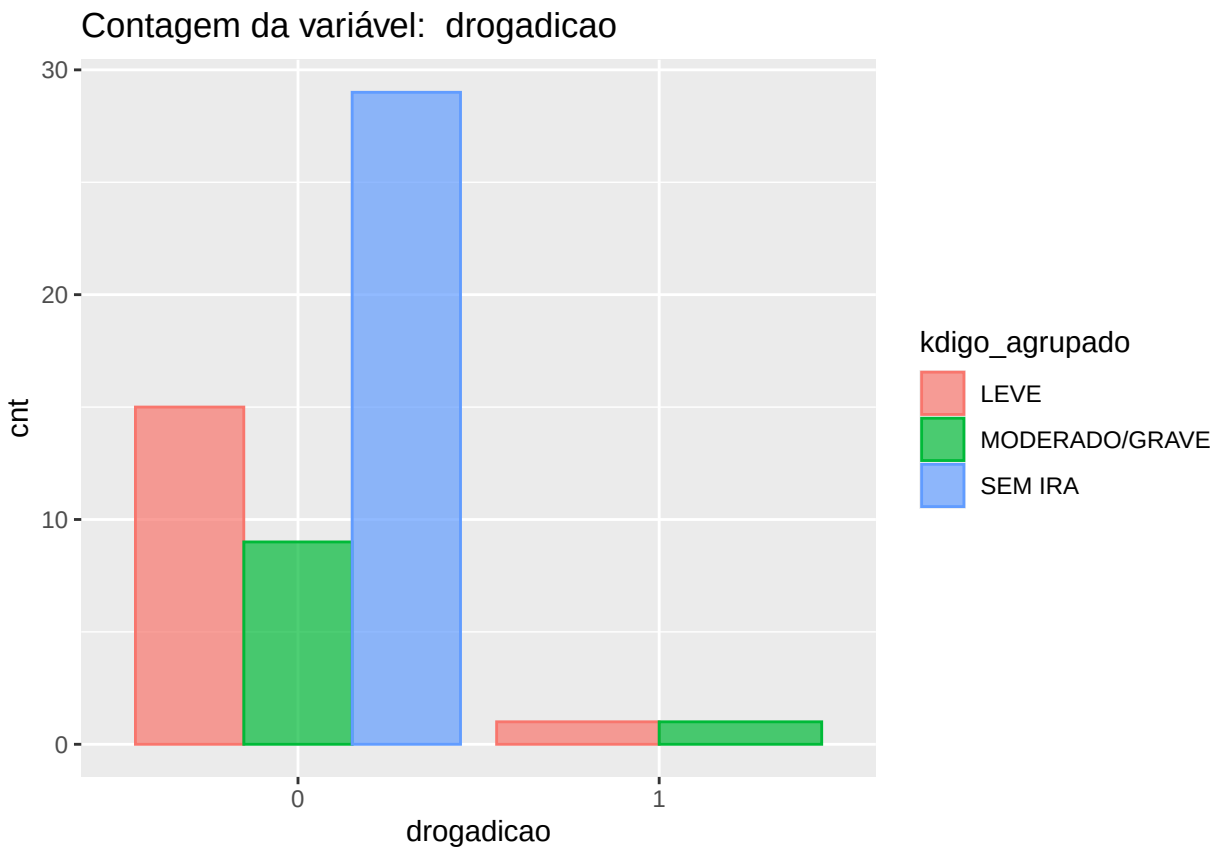

```
## `summarise()` regrouping output by 'dvp' (override with `.groups` argument)
```



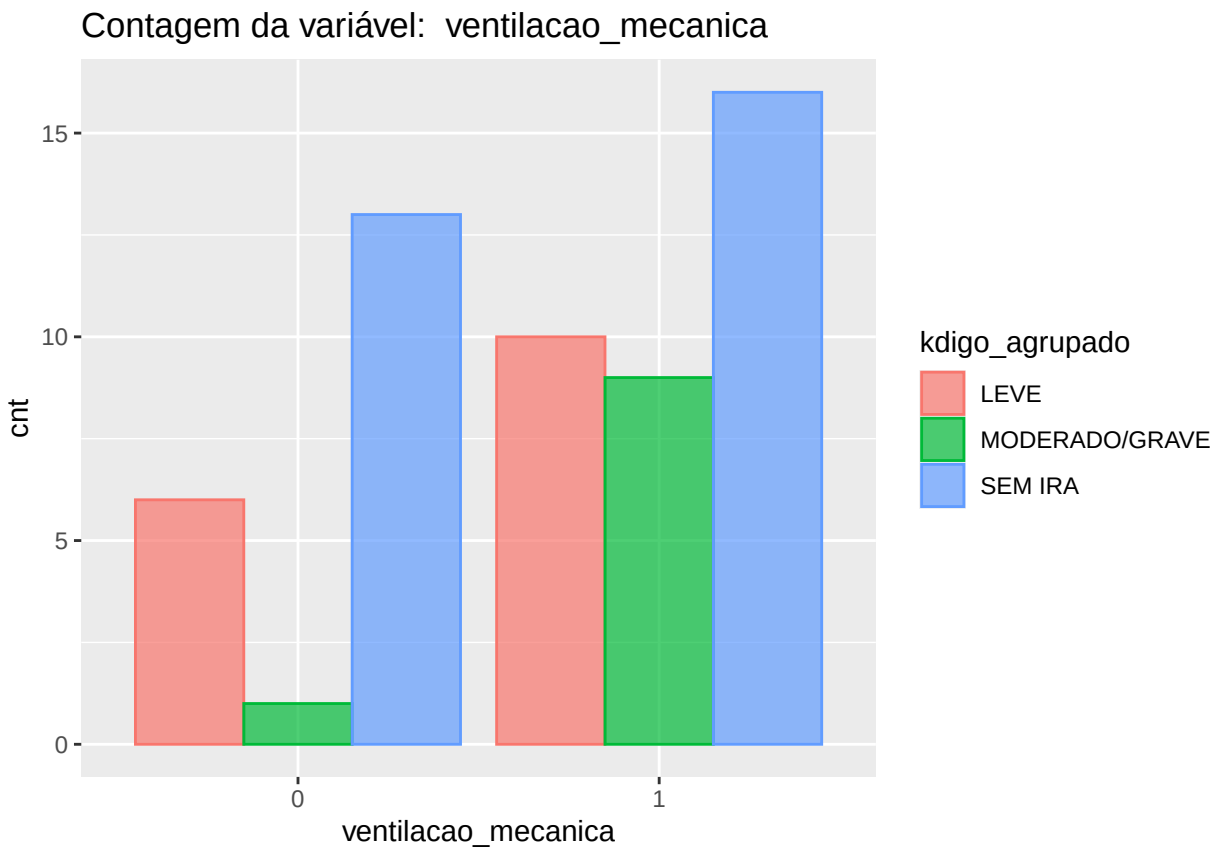

```
## `summarise()` regrouping output by 'drogadicao' (override with `.groups` argument)
```



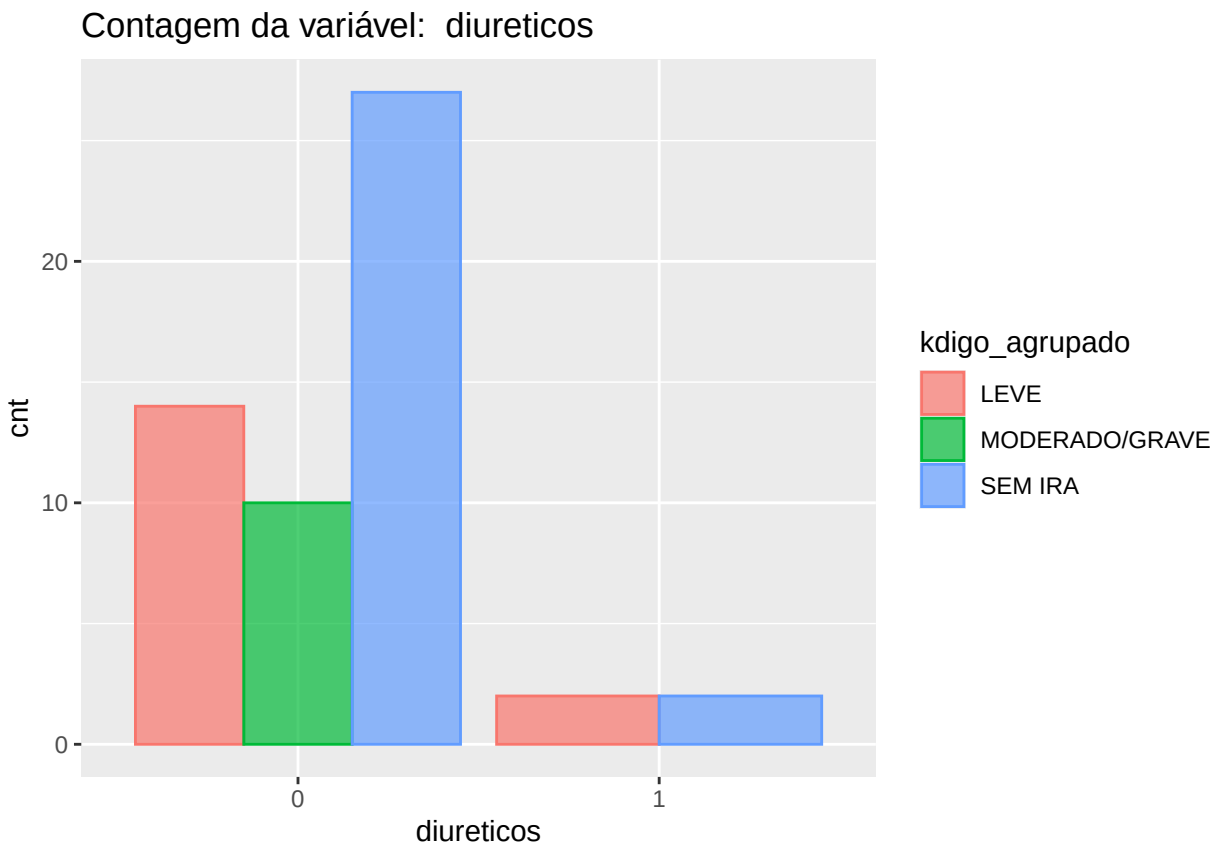

```
## `summarise()` regrouping output by 'ventilacao_mecanica' (override with `.groups` argument)
```



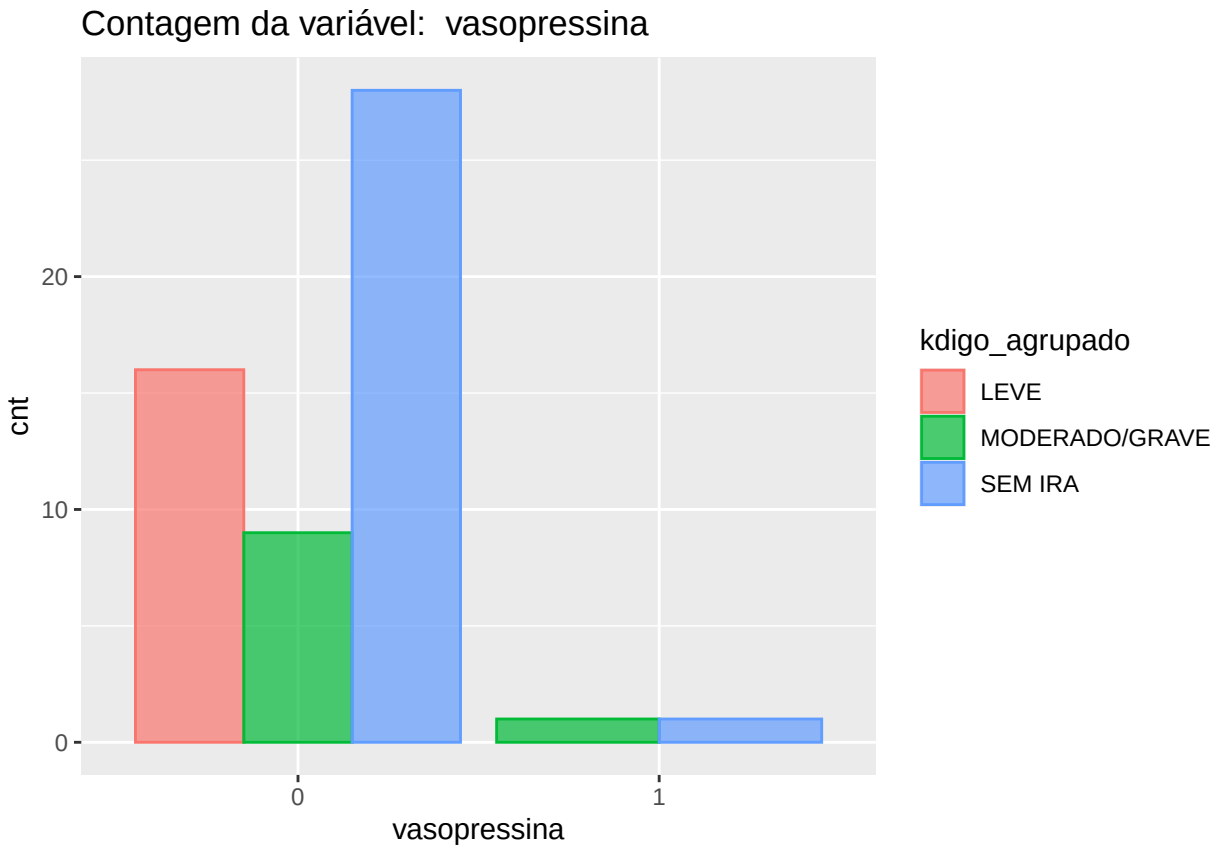

```
## `summarise()` regrouping output by 'diureticos' (override with `.groups` argument)
```



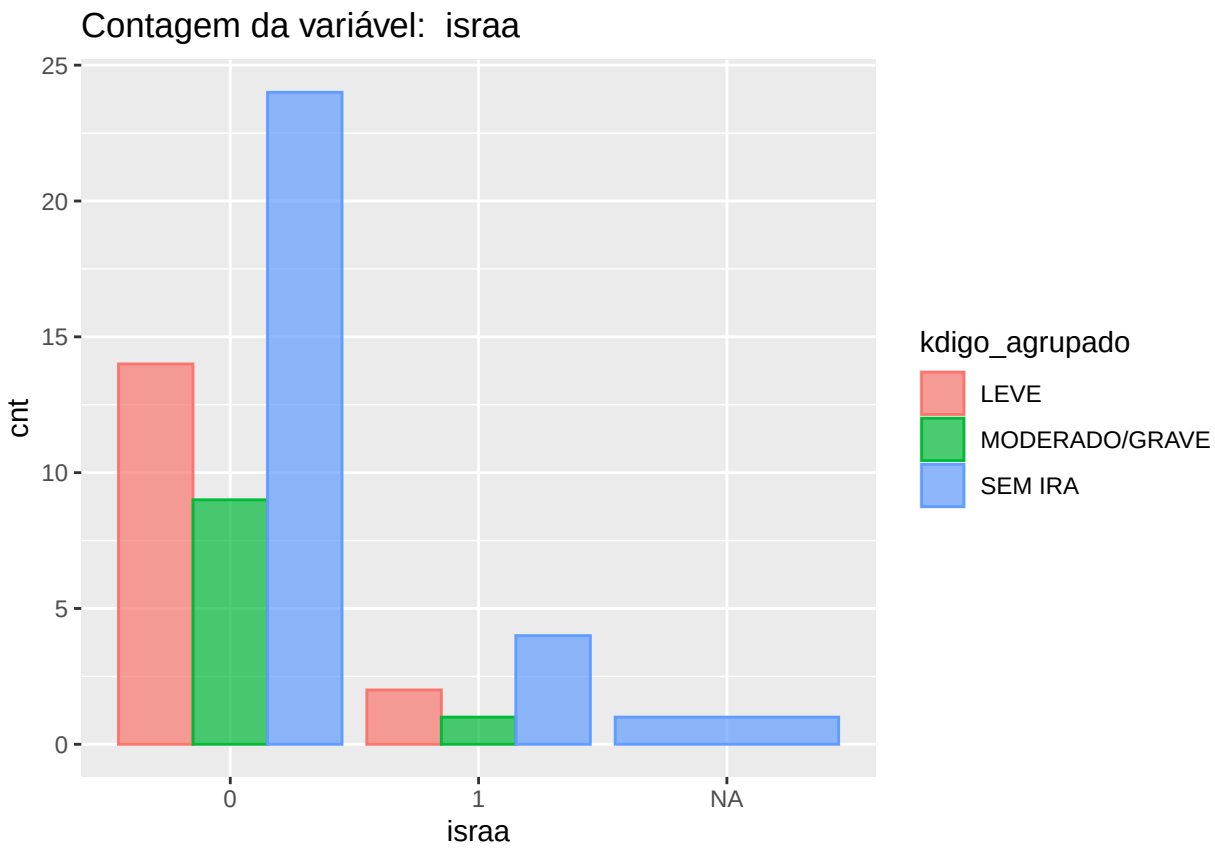

```
## `summarise()` regrouping output by 'vasopressina' (override with `.groups` argument)
```



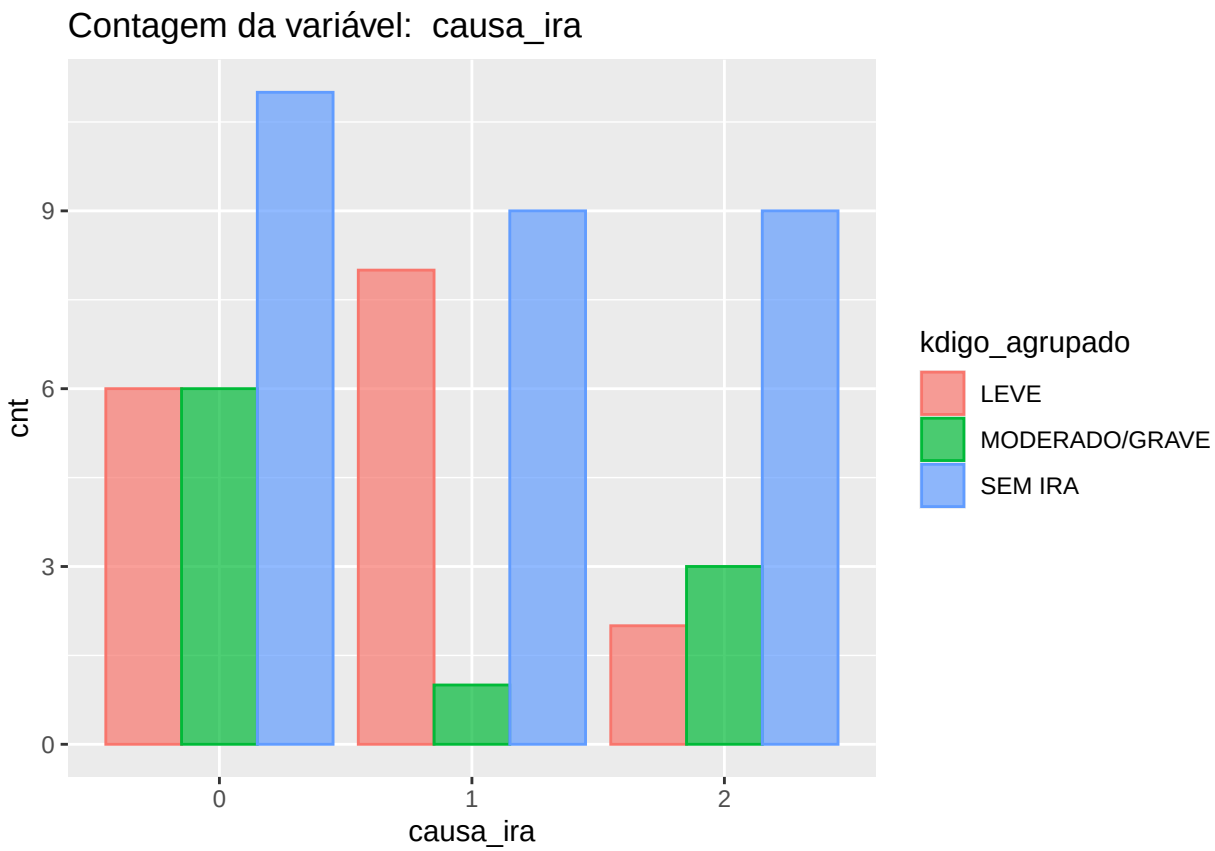

```
## `summarise()` regrouping output by 'israa' (override with `.groups` argument)
```



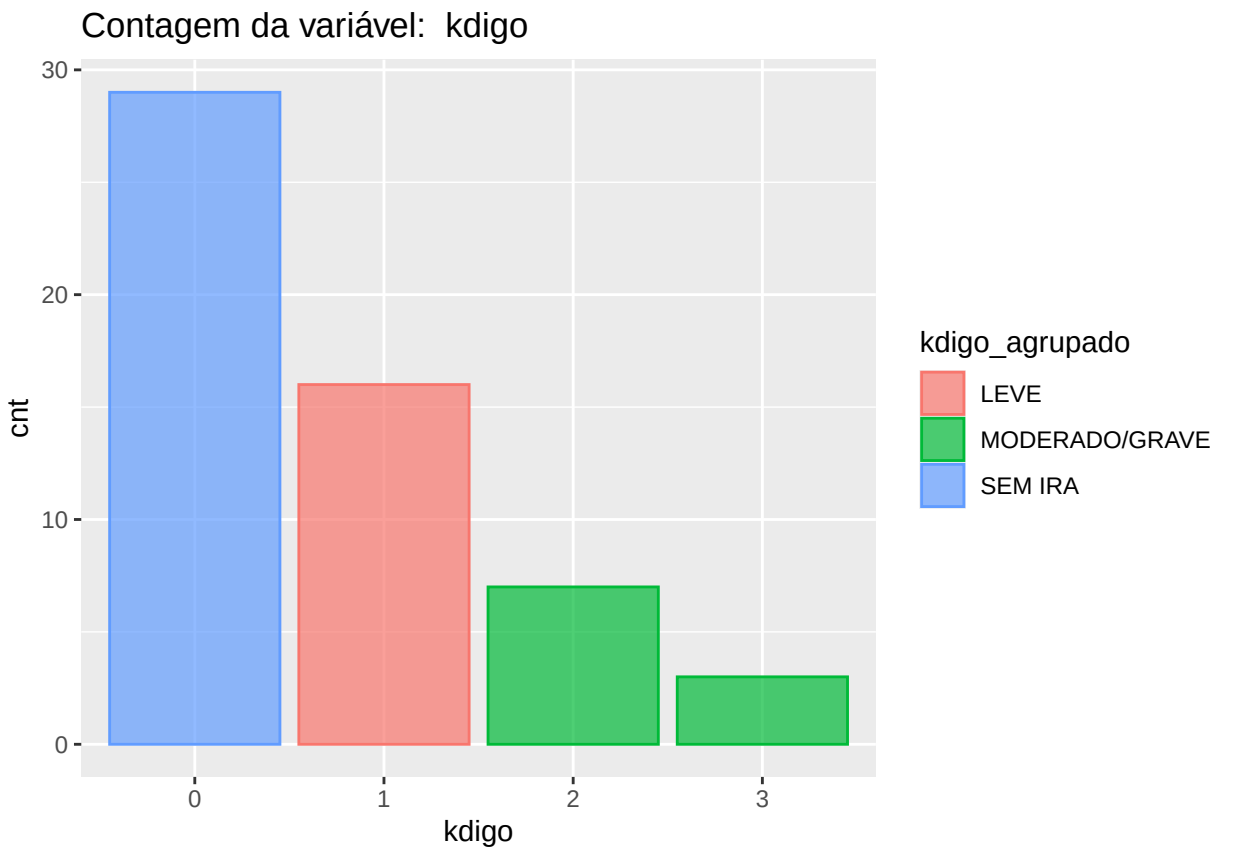

```
## `summarise()` regrouping output by 'causa_ira' (override with `.groups` argument)
```




```
## `summarise()` regrouping output by 'kdigo' (override with `.groups` argument)
```

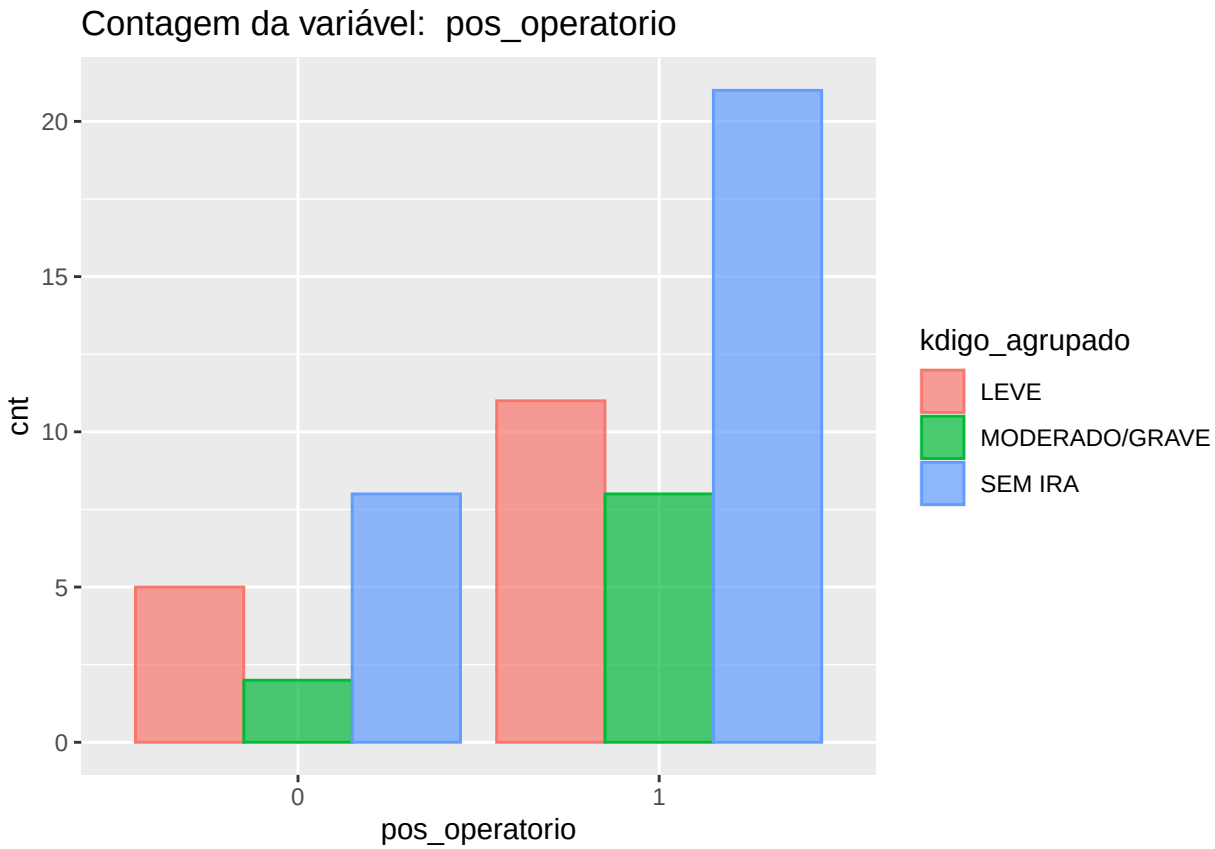


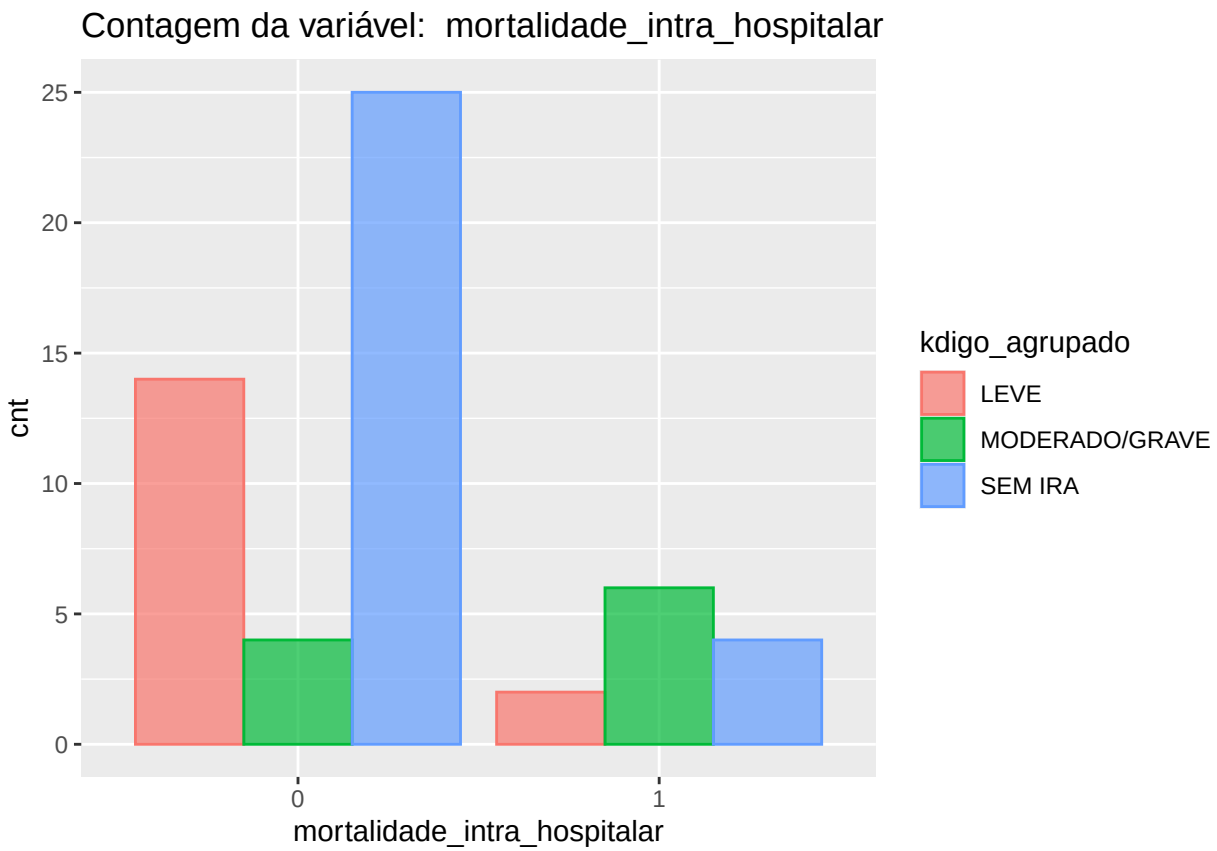

```
## `summarise()` regrouping output by 'criterio_ira' (override with `.groups` argument)
```




```
## `summarise()` regrouping output by 'pos_operatorio' (override with `.groups` argument)
```


```
## `summarise()` regrouping output by 'mortalidade_intra_hospitalar' (override with `.groups` argument)
```


Contagem da variável: pos_operatorio
cnt
20
15
10
5
0
0
1
pos_operatorio
kdigo_agrupado
LEVE
MODERADO/GRAVE
SEM IRA




```
df[rowSums(is.na(df %>% select(-criterio_ira))) > 0,]
```

```
## # A tibble: 5 x 59
##   numero d_ira idade cor    peso altura   imc has   dm    icc   asma_dpoc hiv
##    <dbl> <fct> <dbl> <fct> <dbl>  <dbl> <dbl> <fct> <fct> <fct> <fct>     <fct>
## 1     30 3        71 1      60     1.6   23.4 1     0     0     0         0
## 2     67 5        45 2      75     1.8   23.1 0     0     0     0         0
## 3     76 5        63 2      60     1.65  22.0 1     0     0     0         0
## 4      9 0        81 3      50.7   1.55  21.1 1     1     0     0         0
## 5     34 0        52 1      70     1.7   24.2 1     0     0     0         0
## # ... with 47 more variables: ca_ativo <fct>, avc <fct>, dvp <fct>,
## #   drogadicao <fct>, score_clinico <dbl>, ventilacao_mecanica <fct>,
## #   dva_mcg_kg_min <dbl>, diureticos <fct>, vasopressina <fct>, israa <fct>,
## #   saps3 <dbl>, egfr_basal <dbl>, scr_basal <dbl>, causa_ira <fct>,
## #   kdigo <fct>, criterio_ira <fct>, pos_operatorio <fct>,
## #   mortalidade_intra_hospitalar <fct>, tempo_internacao_hospitalar <dbl>,
## #   ira <fct>, kdigo_agrupado <chr>, diurese <dbl>, bh <dbl>, sofa <dbl>,
## #   su <dbl>, scr <dbl>, sna <dbl>, sk <dbl>, sosm <dbl>, uu <dbl>, ucr <dbl>,
## #   una <dbl>, uk <dbl>, volume_urinario <dbl>, uvu24h <dbl>, feu <dbl>,
## #   uu_ucr <dbl>, una_ucr <dbl>, uvna24h <dbl>, uosm <dbl>,
## #   uosm_estimada <dbl>, `una+uk` <dbl>, uu_su <dbl>, fena <dbl>, fek <dbl>,
## #   su_scr <dbl>, delta_scr <dbl>
```

```
df %>%
    select(uosm, uosm_estimada) %>%
    ggplot(aes(x = uosm_estimada, y = uosm)) +
        geom_point() +
```

```
        geom_abline(intercept = 0, slope = 1,
                    linetype = 2, color = 'red')
```

```
## Warning: Removed 3 rows containing missing values (geom_point).
```