

# Week4 Selecting, Filtering, and Arranging Variables, Creating New Variables, Renaming, Relocating, Summarizing Variables and Grouping Variables

2024年2月3日 16:34

## select()

选择需要的列

## filter()

filter能筛选出符合条件的行

## arrange()

排序，多个变量时后续变量用于排列并列的部分，默认是升序。desc()用于降序

```
starwars %>% arrange(desc(height))
```

## slice()

根据行号选择行

slice\_head(n=)选择头n行

slice\_tail(n=)选择末尾n行

slice\_sample(n=)随机选择n行

## mutate()

添加用函数处理后的新行

```
starwars %>% mutate(height_m = height / 100)
```

.keep参数选择是否保留旧行

```
starwars %>%  
  mutate(  
    height_m = height / 100,  
    BMI = mass / (height_m^2),  
    .keep = "none"  
  )
```

```
#> # A tibble: 87 × 2
```

```
#>   height_m   BMI
```

```
#>   <dbl> <dbl>
```

```
#> 1     1.72  26.0
```

```
#> 2     1.67  26.9
```

```
#> 3     0.96  34.7
```

```
#> 4     2.02  33.3
```

```
#> # i 83 more rows
```

## rename()

改变列名

## relocate()

调整列的顺序

```
starwars %>% relocate(sex:homeworld, .before  
= height)
```

## summarise()

分组汇总，通常结合group\_by()函数使用

```
starwars %>% summarise(height = mean(height,  
na.rm = TRUE))
```

```
#> # A tibble: 1 × 1
#>   height
#>   <dbl>
#> 1    175.
```

## group\_by()

按类别变量进行分组

```
starwars %>% group_by(sex)
```

## left\_join()

```
result<-df1%>%left_join(df2,by='variable1')
```

以df1为母体，将df2中variable1变量下与df1的variable1一致的部分的其它df1中没有的变量数据添加到df2中，组合成一个新的数据框

例子：

```
df1
  Sepal.Length Species
1         5.1   setosa
2         4.9   setosa
3         4.7   setosa
```

```
df2
  Sepal.Length Petal.Length
1         5.1           1.4
2         4.9           1.4
3         4.7           1.3
4         6.0           4.0
5         6.2           4.5
```

```
result
  Sepal.Length Species Petal.Length
1         5.1   setosa           1.4
2         4.9   setosa           1.4
3         4.7   setosa           1.3
```