

gt 패키지로 데이터 테이블 만들기 - 1

표도 데이터 시각화의 방법

데이터 시각화를 한다고 하면 대부분 떠올리는 작업이 그래프를 그리는 작업이다. 데이터 시각화를 시작할 때는 먼저 데이터를 대상으로 할지, 어떤 그래프를 써서 데이터를 예쁘고 직관적으로 표현할지를 고민하게 된다. 하지만 데이터 시각화 결과가 들어갈 보고서에는 대부분 그래프와 함께 작성하는 것이 표이다. 데이터 시각화라고 할 때 많은 사람들은 표까지 생각하지는 않는 듯하다. 인터넷 상의 표의 정의¹를 보면 '시각적 의사소통과 자료의 정렬 양식'으로 나와 있다. 결국 표도 데이터 시각화의 일부인 것이다.

표는 데이터를 직접적으로 표현한다는 점에서 그래프나 플롯과는 다르다. 데이터를 직접 표현하기 때문에 데이터를 정확하게 표현하지만 데이터의 전체적인 흐름이나 분포를 알아보기에는 적절치 않다. 하지만 그래프나 플롯은 정확한 데이터를 알아보기 힘들다는 점에서 표와 그래프, 플롯은 상호 보완적으로 사용되어야 한다.

보통 표는 행과 열로 이루어진 데이터의 모음이다. 행은 각각의 개별 사례²나 개별 사례를 적절히 그룹화한 요약된 데이터가 표현된다. 열은 각각의 사례를 설명하기 위해 필요한 속성이 나열된다. 행과 열은 이 만나는 곳이 사용자가 알고 싶어하는 데이터가 위치하고 이 곳을 칸, 혹은 셀이라고 한다.

표는 그 구성 방법에 따라 1차원 표와 다차원 표로 구성된다. 1차원 표는 단순히 데이터의 나열인 경우를 말한다. 이런 테이블을 '단순 표(simple table)'이라고도 한다. 다차원 표는 1차원 표를 적절한 변환(추상화)하여 표에서 제공하는 몇가지 차원의 정보를 모두 가져야 데이터를 해석할 수 있는 표를 말한다. 다차원 표 중에서 가장 흔하게 보이는 표가 2차원 표, 교차표(cross table)이다. 2차원 표는 행(X축)과 열(Y축)의 정보를 가지고 해당 칸(셀)의 정보를 해석할 수 있다.

아래의 그림은 1차원 표와 2차원 표의 예를 보이고 있다.

¹ <https://ko.wikipedia.org/wiki/%ED%91%9C>

² 데이터베이스에서는 튜플, 레코드 등으로 불린다.

Year	Month	Category	Type	Sales	Cost
2001	January	Fruit	Apples	12	10
2001	January	Fruit	Pears	21	13
2001	January	Fruit	Bananas	29	26
2001	January	Vegetables	Cucumber	9	6
2001	January	Vegetables	Tomatoes	13	11
2001	January	Vegetables	Lettuce	22	20
2001	February	Fruit	Apples	11	9
2001	February	Fruit	Pears	21	14
2001	February	Fruit	Bananas	31	27
2001	February	Vegetables	Cucumber	8	6
2001	February	Vegetables	Tomatoes	12	10
2001	February	Vegetables	Lettuce	20	19
2001	March	Fruit	Apples	9	10
2001	March	Fruit	Pears	19	13
2001	March	Fruit	Bananas	32	27
2001	March	Vegetables	Cucumber	8	6
2001	March	Vegetables	Tomatoes	11	11
2001	March	Vegetables	Lettuce	21	19
2001	April	Fruit	Apples	9	10
2001	April	Fruit	Pears	18	13
2001	April	Fruit	Bananas	32	27
2001	April	Vegetables	Cucumber	10	7
2001	April	Vegetables	Tomatoes	12	11
2001	April	Vegetables	Lettuce	21	19
2001	May	Fruit	Apples	10	9
2001	May	Fruit	Pears	20	13
2001	May	Fruit	Bananas	31	27
2001	May	Vegetables	Cucumber	11	7
2001	May	Vegetables	Tomatoes	13	11
2001	May	Vegetables	Lettuce	21	19
2001	June	Fruit	Apples	9	9
2001	June	Fruit	Pears	18	14
2001	June	Fruit	Bananas	28	26
2001	June	Vegetables	Cucumber	13	8
2001	June	Vegetables	Tomatoes	14	10
2001	June	Vegetables	Lettuce	22	19
2001	July	Fruit	Apples	10	10

Category	Type	2001	2002	2003	Grand total
Fruit	Apples	150	153	162	465
	Bananas	332	336	344	1012
	Pears	267	266	279	812
	Subtotal	749	755	785	2289
Vegetables	Cucumber	140	141	152	433
	Lettuce	246	245	258	749
	Tomatoes	156	161	168	485
	Subtotal	542	547	578	1667
Grand total		1291	1302	1363	3956

위의 예에서 보면 1차원 표의 경우는 데이터를 설명하는 속성이 위쪽에만 설정되어 있다. 하지만 2차원 표의 경우는 데이터를 설명하는 속성이 위쪽과 왼쪽에 모두 설정되어 있다. 특정 칸의 데이터를 해석하기 위해서는 행과 열의 속성 정보를 모두 알아야 한다는 것이다.

1차원 표의 장점은 데이터를 원본 차원에서 확인 할 수 있다는 점이다. 반면 데이터의 행이 길어질 수 있어

보고서에 수록하는 것인 적절치 않을 때가 있다. 하지만 다차원 표의 가장 큰 장점은 데이터를 요약하고 구조화하였기 때문에 대량의 데이터의 특성을 간략히 표현했다는 점이다. 하지만 데이터를 요약하는 방식에 따라 전달되는 정보가 제한적일 수 있다는 단점이 있다.

표의 구성 방법과 요소는 사용자가 표현하고 싶은 데이터와 형태에 따라 표현방식이 매우 다르기 때문에 어느 하나로 정의하기가 어렵다. 다음 그림은 논문의 작성에서 대표적으로 사용되는 미국 심리학회의 표 구성 가이드라인이다.

Table 1
Numbers of Children That Have or Have Not Completed Yearly Home Reading Goals

Grade	Girls		Boys	
	Have	Have Not	Have	Have Not
Wave 1				
K	210 ^a	31 ^b	208	20
1	223	54	230	61
2	254	22	248	44
Total	687	107	686	125
Wave 2				
K	198	49	210	39
1	212	37	221	15
2	227	19	248	26
Total	637	105	679*	80

Note. This table demonstrates the elements of a prototypical table. A *general note* to a table appears first and contains information needed to understand the table, including definitions of abbreviations and the copyright attribution for a reprint or adapted table.

^a A *specific note* appears in a separate paragraph below the general note.

^b Subsequent specific notes follow in the same paragraph.

*A probability *note* (for *p* values) appears as a separate paragraph below any specific notes; subsequent probability notes follow in the same paragraph.

모든 표를 APA 가이드라인에 맞춰 그릴 수는 없겠지만 APA 가이드라인을 보면 표에 꼭 들어가야 할 몇가지 요소들을 알 수 있다.

- 표 제목(Title) : 표에서 표현하고 있는 데이터를 대표하는 제목이 반드시 필요하다.
- 표 헤드(Heading) : 표에서 제시하고 있는 사례들의 속성값들에 대한 이름이 표현된 표 가로축의 맨 위줄을 말한다.
- 표 스텝(Stub) : 표에서 표현되는 다양한 사례를 구분하기 위해 필요한 세로축의 가장 왼쪽 줄을 말한다.
- 표 몸체(Body) : 표의 헤드와 스텝으로 감싸고 있는 데이터가 표현된 셀들이 표현된 부분을 말한다.

실무에서 표를 그릴 때 가장 많이 사용하는 툴은 아마도 MS-Excel일 것이다. Spreadsheet로 거의 유일하게(?) 살아남은 이 툴은 표를 그리는 WYSIWYG(What You See IS Wat You Get) 툴로 거의 모든 Spreadsheet를 없애버렸다고 해도 과언이 아닐 듯 하다. 어쨌던 엑셀은 표를 그리는데 매우 특화된 툴임에는 틀림없고 초보자들도 쉽게 사용할 수 있지만 사용하다보면 몇가지 단점을 만날 수 있다.

가장 큰 단점이 반복된 작업을 하기가 어렵다는 점이다. 이는 WYSIWIG 툴들이 가지는 공통된 단점이다. 표를 쉽게 만들 수 있지만 유사한 표를 다시 만들어야 하는 경우 반복된 작업을 수행해야하고 반복된 작업을 꼼꼼히 기록해두지 않으면 동일한 표를 만들기 어렵다. 두번째 단점이 기초 데이터의 구조가 업데이트 되면 표를 다시 그려야 하는 경우가 발생한다는 점이다. 기초 데이터의 열 이름이 바뀌거나 열이 추가, 삭제되면 이 데이터로부터 파생된 표가 정상적으로 표현되지 않는다는 것이다. 이러한 문제는 엑셀을 사용해 본 사용자라면 한번쯤, 아니 자주 겪는 문제일 것이다.

이제 엑셀에서 탈피해서 R에서 표를 그려보자. R에서도 아주 훌륭한, 그리고 예쁘게 표를 그릴 수 있는 다양한 방법을 제공한다. R로 표를 그려보면 반복된 작업을 할 필요가 없다는 점에서, 기초 데이터의 업데이트가 발생해도 바로 반영할 수 있다는 점, 표를 세세하게 다룰 수 있다는 점에서 엑셀로 다시 돌아갈 수 없을 수도 있다.

gt 패키지

R에서 표를 만들기 위해 사용되는 패키지 중에 하나가 gt 패키지이다. gt 패키지는 표의 세부 구성들을 상세히 구분하고 이들을 구조적으로 조화롭게 구성시켜서 표를 만드는 다양한 함수와 매개 변수들을 제공한다. gt 패키지에서 사용하는 표의 세부 파트는 다음의 그림과 같다.

표의 구성을 보면 앞에서 살펴보았던 APA 가이드라인상의 표와 유사한 형태를 보인다. 이와 같이 표를 세부적인 파트로 구분하고 이들을 각각 설정함으로써 표를 만들 수 있다.

gt 패키지에서 권장하는 표 생성 방식은 다음의 그림과 같다.

우선 표를 만들기 위한 기초 데이터를 생성해야 한다. gt 패키지로 생성되는 표는 tidyverse 방식을 준용하여 생성하는 것이 편하다. 따라서 gt 표를 생성하기 위해 만드는 기초 데이터도 tidyverse의 tibble 이나 R의 기초 데이터 셋인 data.frame 으로 생성한다. 이렇게 생성된 기초 데이터 셋은 gt 객체로 변환되고 gt 패키지에서 제공되는 다양한 함수와 매개변수를 설정하여 사용자가 원하는 형태의 표로 만들어진다. 마지막으로 이 표를 다른 문서에서 활용하기 위해 html, pdf 등의 문서로 변환하여 활용한다.

Data Import

이번 포스트에서는 각각의 학제에 소속된 학과들에 대한 정보를 표현하기 위한 표를 그려보도록 하겠다. 표에서는 각각의 열을 학제로 구분(Stub)하고, 특성으로 표현되는 열은 크게 합계와 평균 으로 묶어주고(span), 학위과정 데이터를 사용하여 행을 그룹핑한 후 그룹별, 전체 합계와 평균 행을 추가하는 표를 그리겠다.

gt 패키지를 사용하여 표를 만들어 보기 위해 사용하는 데이터는 한국교육개발원 교육통계서비스 홈페이지의 학교/학과별 데이터셋 - 대학 - 학과별(상반기) - 2021를 활용하겠다.

```
library(openxlsx)
library(tidyverse)

## `read.xlsx()`
df <- read.xlsx(xlsxFile = './21
                sheet = ' ', ## sheet ' ',
                startRow = 13, ## 13
                na.string = '-', ## NA '-'
                colNames = T) ##

##
df$ <- fct_relevel(df$ , ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ( )', ' ( )', ' ( )', ' ( )', ' (2)'
```

1. 표로 그릴 데이터 셋 만들기 우선 데이터를 정제하기 위해 필요한 데이터만 필터링과 서브세팅하도록 하겠다. 여기서 필요한 데이터는 각각의 데이터 중 '합계'에 해당하는 열이고 이 데이터들을 학제와 학위과정으로 그룹핑하여 합계와 평균을 낸 데이터이다. 다음과 같이 산출한다.

학제	학위과정	학과수_전체_sum	지원자_전체_계_sum	입학자_전체_계_sum	재적생_전체_계_sum
대학교	대학과정	12028	2635154	329306	1938254
교육대학	대학과정	140	15805	3864	15409
산업대학	대학과정	251	22128	2379	14539
사이버대학(대학)	대학과정	357	51840	34279	135155
전문대학(2년제)	전문대학과정	2546	473154	70200	238266
전문대학(3년제)	전문대학과정	3095	603197	87333	304435
전문대학(4년제)	전문대학과정	317	63559	9174	33340
기능대학	전문대학과정	263	19102	7565	23910
일반대학원	대학원과정	10076	121226	69928	161987
특수대학원	대학원과정	4351	84377	45448	124912
전문대학원	대학원과정	1088	39141	15556	40516

앞서 설명한 바와 같이 각각의 행을 구분할 수 있는 구분열(Stub)을 설정해야 한다. 또 목표로 설정한 표는 학위과정별로 소계와 평균을 구해야하기 때문에 학위과정별로 그룹핑해야 한다. 행 구분을 위한 구분 열은 rowname_col로 설정하고 행을 그룹핑하기 위한 열은 groupname_col로 설정한다.

```
gt.table1 <- df.gt |>
  ## ' ', ' ', ' ' `gt()`
  gt(rowname_col = ' ',
     groupname_col = ' ')
```

	학과수_전체_sum	지원자_전체_계_sum	입학자_전체_계_sum	재적생_전체_계_sum	재학생_전체_계_sum
대학과정					
대학교	12028	2635154	329306	1938254	
교육대학	140	15805	3864	15409	
산업대학	251	22128	2379	14539	
사이버대학(대학)	357	51840	34279	135155	
전문대학과정					
전문대학(2년제)	2546	473154	70200	238266	
전문대학(3년제)	3095	603197	87333	304435	
전문대학(4년제)	317	63559	9174	33340	
기능대학	263	19102	7565	23910	
대학원과정					
일반대학원	10076	121226	69928	161987	
특수대학원	4351	84377	45448	124912	
전문대학원	1088	39141	15556	40516	

3. 표 꾸미기

표의 각 부분 설정 먼저 표의 각 부분을 설정한다. 앞서 살펴본 gt 객체의 각각의 부분들을 설정하는 단계로써 tab_*로 시작하는 함수들이 대부분 표의 각 부분을 설정하는 함수이다.

표 각 부분 설정에는 특정한 순서가 정해져 있지 않지만 여기서는 먼저 표의 제목부터 설정하겠다. 표 제목을 설정할 때 사용하는 함수는 `tab_header()`이다. `tab_header()`의 용법과 주요 매개변수는 다음과 같다.

**** note ** :::** infobox `tab_header(data, title, subtitle = NULL)`

- data : 표 그리기에 사용할 gt 객체
- title : 표 제목 설정
- subtitle : 표 부제목 설정 ::: —

```
gt.table2 <- gt.table1 |>
##
tab_header(title = ' ', subtitle = '2021 ')
```

	학과수_전체_sum	지원자_전체_계_sum	입학자_전체_계_sum	재적생_전체_계_sum	재학생_
대학교과정					
대학교	12028	2635154	329306	1938254	
교육대학	140	15805	3864	15409	
산업대학	251	22128	2379	14539	
사이버대학(대학)	357	51840	34279	135155	
전문대학교과정					
전문대학(2년제)	2546	473154	70200	238266	
전문대학(3년제)	3095	603197	87333	304435	
전문대학(4년제)	317	63559	9174	33340	
기능대학	263	19102	7565	23910	
대학원과정					
일반대학원	10076	121226	69928	161987	
특수대학원	4351	84377	45448	124912	
전문대학원	1088	39141	15556	40516	

위의 실행 결과를 보면 표 제목과 표 부제목이 잘 설정되었음을 알 수 있다. 그런데 표의 열이 너무 길고 열 이름에 'sum'과 'mean'이 반복되고 있다. 이 부분을 열 그룹으로 묶어(Scanner)주면 보기가 좋아질 수 있을 것이다. `tab_scanner()`는 이렇게 유사한 속성을 가진 열끼리 묶어주는 기능을 제공한다. 그 용법과 주요 매개변수는 다음과 같다.

tab_scanner(data, label, columns, id = label, gather = TRUE)

- data : 표 그리기에 사용할 gt 객체
- label : 열 그룹에 사용할 라벨
- columns : 열 그룹에 속할 열 이름 벡터

여기서는 3번부터 8번까지의 열을 '합계'로, 9번부터 14번까지 열을 '평균'으로 묶어주었다.

```
gt.table3 <- gt.table2 |>
## 3 8 ' '
tab_scanner(columns = 3:8, label = ' ') |>
```

```
## 9 14 ' '
tab_spanner(columns = 9:14, label = ' ')
```

	합계				
	학과수_전체_sum	지원자_전체_계_sum	입학자_전체_계_sum	재적생_전체_계_sum	재학생_전체_계_sum
대학과정					
대학교	12028	2635154	329306	1938254	
교육대학	140	15805	3864	15409	
산업대학	251	22128	2379	14539	
사이버대학(대학)	357	51840	34279	135155	
전문대학과정					
전문대학(2년제)	2546	473154	70200	238266	
전문대학(3년제)	3095	603197	87333	304435	
전문대학(4년제)	317	63559	9174	33340	
기능대학	263	19102	7565	23910	
대학원과정					
일반대학원	10076	121226	69928	161987	
특수대학원	4351	84377	45448	124912	
전문대학원	1088	39141	15556	40516	

이외에도 표의 각 부분을 설정하는 다음과 같은 함수가 있다.

tab_spanner(data, label, columns, id = label, gather = TRUE)

- tab_spanner_delim() : 특정한 구분자를 통해 열 라벨과 묶음 라벨을 설정
- tab_row_group() : 행 그룹 생성
- tab_stubhead() : 구분 헤더에 사용할 문자열 설정
- tab_footnote() : 각주 부분 생성
- tab_source_note() : 출처 부분 생성
- tab_style() : 표의 각 셀에 대한 스타일 설정
- tab_options() : 전체 표의 옵션 설정

표의 데이터 표기 형태 설정 표에는 많은 데이터들이 표현된다. 데이터의 속성에 따라 데이터의 표현 방식도 달라지게 된다. 따라서 표현되는 데이터의 속성에 따라 그 표기 형태를 정해줄 필요가 있다. gt 패키지에서는 표에 표기되는 데이터의 표기 형태를 설정하는 다양한 함수를 제공하는데 대부분 `fmt_*` 로 시작되는 함수이다. gt 패키지에서 제공하는 주요 `fmt_*`() 함수들은 다음과 같다.

- `fmt_number()` : 수치 형태의 포맷
- `fmt_integer()` : 정수 형태의 포맷
- `fmt_scientific()` : 과학적 표기법의 포맷
- `fmt_engineering()` : 공학적 표기법의 포맷
- `fmt_percent()` : 백분율 형태의 포맷

- `fmt_currency()` : 통화 형태의 포맷
- `fmt_bytes()` : 바이트 형태의 포맷
- `fmt_date()` : 날짜 형태의 포맷
- `fmt_time()` : 시간 형태의 포맷
- `fmt_datetime()` : 날짜, 시간 형태의 포맷
- `fmt_markdown()` : 마크다운 텍스트 포맷
- `fmt_passthrough()` : 문자의 강제 변환을 위한 포맷
- `fmt_missing()` : 누락값에 대한 포맷
- `fmt()` : 사용자 정의 포맷을 설정
- `text_transform()` : 문자열 변환 함수를 적용하는 포맷
- `data_color()` : 데이터의 색을 변경하는 함수

위의 표에서는 모든 데이터들이 수치형 데이터이다. 하지만 합계 그룹의 수치들은 소수점이 필요없고 천단위 구분자가 필요한 형태이고 평균 그룹의 수치들은 소수점 첫자리정도가 의미가 있고 천단위 구분자가 필요하다. 따라서 다음과 같이 설정할 수 있다.

```
gt.table4 <- gt.table3 |>
  ## 3      8              1000
  fmt_number(columns = 3:8, decimals = 0, use_seps = TRUE) |>
  ## 9     14              1000
  fmt_number(columns = 9:14, decimals = 1, use_seps = TRUE)
```

	합계				
	학과수_전체_sum	지원자_전체_계_sum	입학자_전체_계_sum	재적생_전체_계_sum	재학생_
대학과정					
대학교	12,028	2,635,154	329,306	1,938,254	
교육대학	140	15,805	3,864	15,409	
산업대학	251	22,128	2,379	14,539	
사이버대학(대학)	357	51,840	34,279	135,155	
전문대학과정					
전문대학(2년제)	2,546	473,154	70,200	238,266	
전문대학(3년제)	3,095	603,197	87,333	304,435	
전문대학(4년제)	317	63,559	9,174	33,340	
기능대학	263	19,102	7,565	23,910	
대학원과정					
일반대학원	10,076	121,226	69,928	161,987	
특수대학원	4,351	84,377	45,448	124,912	
전문대학원	1,088	39,141	15,556	40,516	

열 수정 이번에는 열에 대한 수정을 위한 함수들을 알아본다. `cols_*`로 시작하는 함수들을 사용하여 전체 혹은 일부 열의 속성을 수정할 수 있다. 열의 정렬, 보이기/감추기, 열 라벨 설정, 열 합치기 등을 할 수 있다. 열의 수정을 위한 함수는 다음과 같다.

- cols_align() : 수치 형태의 포맷
- cols_width() : 정수 형태의 포맷
- cols_label() : 과학적 표기법의 포맷
- cols_move_to_start() : 공학적 표기법의 포맷
- cols_move_to_end() : 백분율 형태의 포맷
- cols_move() : 통화 형태의 포맷
- cols_hide() : 바이트 형태의 포맷
- cols_unhide() : 날짜 형태의 포맷
- cols_merge_range() : 시간 형태의 포맷
- cols_merge_uncert() : 날짜, 시간 형태의 포맷
- cols_merge_n_pct() : 마크다운 텍스트 포맷
- cols_merge() : 문자의 강제 변환을 위한 포맷

위에서 만든 표의 열 이름이 사용자가 알아보기 좀 어렵게 표기되어 있다. 열 이름을 바꾸기 위해 사용하는 함수는 cols_label()이다. cols_label()의 용법과 주요 매개변수는 다음과 같다.

```
cols_label(.data, ..., .list = list2(...))
```

- data : 표 그리기에 사용할 gt 객체
- ... : 열 라벨로 표시할 문자열과 해당 열 이름에 대한 named vector
- .list : 열 이름과 열 라벨의 named vector 대신 사용될 list

```
gt.table5 <- gt.table4 |>
```

```
##
```

```
cols_label( = ' ',
  - _sum = ' ',
  - _sum = ' ',
  - _sum = ' ',
  - _sum = ' ',
  - _sum = ' ',
  - _sum = ' ',
  - _mean = ' ',
  - _mean = ' ',
  - _mean = ' ',
  - _mean = ' ',
  - _mean = ' ',
  - _mean = ' '
)
```

고등교육기관 데이터
2021년 전체 고등교육기관 대상

	합계						학과수	지원자	
	학과수	지원자	입학자	재적생	재학생	휴학생			
대학과정									
대학교	12,028	2,635,154	329,306	1,938,254	1,415,162	504,165	1.5	326.3	
교육대학	140	15,805	3,864	15,409	15,045	364	2.3	263.4	

산업대학	251	22,128	2,379	14,539	11,076	3,374	1.0	90.3
사이버대학(대학)	357	51,840	34,279	135,155	119,995	15,160	0.8	120.8
전문대학과정								
전문대학(2년제)	2,546	473,154	70,200	238,266	162,729	75,386	1.4	257.8
전문대학(3년제)	3,095	603,197	87,333	304,435	213,105	91,162	1.7	326.9
전문대학(4년제)	317	63,559	9,174	33,340	24,242	9,077	1.1	227.8
기능대학	263	19,102	7,565	23,910	14,860	9,050	1.5	107.9
대학원과정								
일반대학원	10,076	121,226	69,928	161,987	143,965	18,022	2.7	31.9
특수대학원	4,351	84,377	45,448	124,912	107,537	17,375	1.7	32.1
전문대학원	1,088	39,141	15,556	40,516	36,108	4,408	1.5	55.5

행 수정 행 수정에 관련된 함수는 `row_group_order()` 뿐이다. 이 함수는 행 그룹의 순서를 지정해주는 역할을 한다.

row_group_order(data, groups)

- data : 표 그리기에 사용할 gt 객체
- groups : 행 그룹의 순서를 지정하는 문자형 벡터

앞선 표 안의 데이터를 잘 살펴보면 행 그룹의 순서가 대학->전문대학->대학원의 순서로 설정되어 있다. 보통 전문대학->대학->대학원으로 표현되는 것이 일반적이다. 다음과 같이 이 순서를 변경시켜줄 수 있다.

```
gt.table6 <- gt.table5 |>
  ## ' ', ' ', ' ', ' '
  row_group_order(groups = c(' ', ' ', ' '))
```

고등교육기관 데이터
2021년 전체 고등교육기관 대상

	합계						학과수	지원자
	학과수	지원자	입학자	재적생	재학생	휴학생		
전문대학과정								
전문대학(2년제)	2,546	473,154	70,200	238,266	162,729	75,386	1.4	257.8
전문대학(3년제)	3,095	603,197	87,333	304,435	213,105	91,162	1.7	326.9
전문대학(4년제)	317	63,559	9,174	33,340	24,242	9,077	1.1	227.8
기능대학	263	19,102	7,565	23,910	14,860	9,050	1.5	107.9
대학과정								
대학교	12,028	2,635,154	329,306	1,938,254	1,415,162	504,165	1.5	326.3
교육대학	140	15,805	3,864	15,409	15,045	364	2.3	263.4
산업대학	251	22,128	2,379	14,539	11,076	3,374	1.0	90.3
사이버대학(대학)	357	51,840	34,279	135,155	119,995	15,160	0.8	120.8
대학원과정								
일반대학원	10,076	121,226	69,928	161,987	143,965	18,022	2.7	31.9
특수대학원	4,351	84,377	45,448	124,912	107,537	17,375	1.7	32.1
전문대학원	1,088	39,141	15,556	40,516	36,108	4,408	1.5	55.5

요약 행 추가 표를 생성할 때 원 데이터에는 없는 소계, 총계가 필요한 경우가 있다. `gt` 패키지에서는 이러한 요약행을 자동으로 생성해 주는 기능을 제공한다. `gt` 패키지에서 제공하는 요약 행 생성 함수는 `summary_rows()`와 `grand_summary_rows()`이다.

- `summary_rows()` : 요약 함수를 사용하여 그룹별 요약 행을 생성
- `grand_summary_rows()` : 요약 함수를 사용하여 전체 요약 행을 생성

앞선 표의 전체적인 표의 형태가 대략 갖춰진 듯 하다. 마지막으로 각 행 그룹의 합계, 평균 행을 넣고 마지막에 전체 합계, 평균 행을 넣어주겠다. 각 행 그룹의 요약 행을 넣는 `summary_rows()`와 `grand_summary_rows()`의 용법과 주요 매개변수는 다음과 같다.

summary_rows(data, groups = NULL, columns = everything(), ...)

- `data` : 표 그리기에 사용할 `gt` 객체
- `groups` : 요약 행을 넣어줄 그룹 설정
- `columns` : 요약 행을 적용할 열을 설정
- `fns` : 요약 행에 적용할 그룹 함수를 설정, `mean`, `sum` 등 R에서 제공하는 함수를 사용할 수도 있고 사용자 생성 함수를 사용할 수도 있다. `list`의 형태로 제공하며 요약 행에 표현될 이름을 같이 지정할 수 있다.
- `missing_text` : 데이터에 NA값이 존재할 경우 대체할 텍스트 설정
- `formatter` : 요약 행의 데이터 포맷 설정

grand_summary_rows(data, columns = everything(), fns, ...)

- `data` : 표 그리기에 사용할 `gt` 객체
- `columns` : 요약 행을 적용할 열을 설정
- `fns` : 요약 행에 적용할 그룹 함수를 설정, `mean`, `sum` 등 R에서 제공하는 함수를 사용할 수도 있고 사용자 생성 함수를 사용할 수도 있다. `list`의 형태로 제공하며 요약 행에 표현될 이름을 같이 지정할 수 있다.
- `missing_text` : 데이터에 NA값이 존재할 경우 대체할 텍스트 설정
- `formatter` : 요약 행의 데이터 포맷 설정

위의 표에 요약 행으로 합계와 평균값, 전체 요약행으로 합계와 평균값을 넣어주는 코드는 다음과 같다. 여기서는 합계 열 그룹과 평균 열 그룹의 값 표기 형태가 다르기 때문에 두 부분으로 나누어서 설정하였다.

```
gt.table7 <- gt.table6 |>
##
summary_rows(
  ##
  groups = T,
  ## 3 8
  columns = 3:8,
  ##      ' ' `sum()` , ' ' `mean()`
  fns = list(
    = ~sum(.),
    = ~mean(.)),
  ##
  formatter = fmt_number,
  ##
  decimals = 0,
  ##
```

```

    use_seps = TRUE
) |>
summary_rows(
  groups = T,
  ## 9 14
  columns = 9:14,
  fns = list(
    = ~sum(.),
    = ~mean(.)),
  formatter = fmt_number,
  ##
  decimals = 1,
  use_seps = TRUE
) |>
##
grand_summary_rows(
  columns = 3:8,
  fns = list(
    = ~sum(.),
    = ~mean(.)),
  decimals = 0,
  use_seps = TRUE
) |>
grand_summary_rows(
  columns = 9:14,
  fns = list(
    = ~sum(.),
    = ~mean(.)),
  decimals = 1,
  use_seps = TRUE
)

```

고등교육기관 데이터
2021년 전체 고등교육기관 대상

	합계							
	학과수	지원자	입학자	재적생	재학생	휴학생	학과수	지원자
전문대학과정								
전문대학(2년제)	2,546	473,154	70,200	238,266	162,729	75,386	1.4	257.8
전문대학(3년제)	3,095	603,197	87,333	304,435	213,105	91,162	1.7	326.9
전문대학(4년제)	317	63,559	9,174	33,340	24,242	9,077	1.1	227.8
기능대학	263	19,102	7,565	23,910	14,860	9,050	1.5	107.9
합계	6,221	1,159,012	174,272	599,951	414,936	184,675	5.7	920.5
평균값	1,555	289,753	43,568	149,988	103,734	46,169	1.4	230.1
대학과정								
대학교	12,028	2,635,154	329,306	1,938,254	1,415,162	504,165	1.5	326.3
교육대학	140	15,805	3,864	15,409	15,045	364	2.3	263.4
산업대학	251	22,128	2,379	14,539	11,076	3,374	1.0	90.3
사이버대학(대학)	357	51,840	34,279	135,155	119,995	15,160	0.8	120.8
합계	12,776	2,724,927	369,828	2,103,357	1,561,278	523,063	5.7	800.9

평균값	3,194	681,232	92,457	525,839	390,320	130,766	1.4	200.2
대학원과정								
일반대학원	10,076	121,226	69,928	161,987	143,965	18,022	2.7	31.9
특수대학원	4,351	84,377	45,448	124,912	107,537	17,375	1.7	32.1
전문대학원	1,088	39,141	15,556	40,516	36,108	4,408	1.5	55.5
합계	15,515	244,744	130,932	327,415	287,610	39,805	5.8	119.5
평균값	5,172	81,581	43,644	109,138	95,870	13,268	1.9	39.8
합계	34,512	4,128,683	675,032	3,030,723	2,263,824	747,543	17.2	1,840.9
평균값	3,137	375,335	61,367	275,520	205,802	67,958	1.6	167.4

다음 포스트에서는 `gt` 표를 예쁘게 꾸며보는 방법을 알아보도록 하겠다..

Coming Soon..