

{ Tidy Data }

(Kyungnam Univ. Bigdata center Kim JiSeong)

1) Tidy Data

행정구역별(읍면동)	2017							2018					
	총인구 (명)	남자 (명)	여자 (명)	내국인-계 (명)	내국인-남자 (명)	내국인-여자 (명)	외국인-계 (명)	총인구 (명)	남자 (명)	여자 (명)	내국인-계 (명)	내국인-남자 (명)	내국인-여자 (명)
전국	51,422,507	#####	#####	49,943,260	24,922,392	25,020,868	1,479,247	51,629,512	#####	#####	49,977,951	24,931,554	25,046,400
읍부	4,794,377	2,456,558	2,337,819	4,625,994	2,347,613	2,278,381	168,383	4,984,023	2,557,807	2,426,216	4,787,535	2,431,315	2,355,220
면부	4,835,090	2,498,157	2,336,933	4,592,308	2,324,414	2,267,894	242,782	4,730,290	2,453,691	2,276,599	4,464,067	2,263,139	2,200,928
동부	41,793,040	#####	#####	40,724,958	20,250,365	20,474,593	1,068,082	41,915,199	#####	#####	40,726,349	20,237,100	20,489,249
서울특별시	9,741,871	4,757,642	4,984,229	9,397,944	4,592,393	4,805,551	343,927	9,673,936	4,718,972	4,954,964	9,299,511	4,536,552	4,763,459
종로구	157,277	76,670	80,607	146,298	71,757	74,541	10,979	157,967	77,147	80,820	145,911	71,611	74,300
중구	127,896	62,195	65,701	117,631	57,391	60,240	10,265	129,797	62,975	66,822	117,809	57,240	60,569
용산구	223,898	108,471	115,427	207,893	99,759	108,134	16,005	226,938	110,685	116,253	207,812	99,806	108,006
성동구	302,367	149,060	153,307	291,931	144,018	147,913	10,436	306,796	150,623	156,173	295,489	145,220	150,269
광진구	363,934	177,156	186,778	345,535	168,833	176,702	18,399	362,304	175,762	186,542	342,575	166,730	176,045
동대문구	357,380	176,460	180,920	339,438	168,977	170,461	17,942	358,141	176,323	181,818	338,293	167,999	170,300
중랑구	396,892	196,346	200,546	390,879	193,871	197,008	6,013	391,668	193,599	198,069	385,082	190,871	194,211
성북구	445,417	215,068	230,349	432,226	209,374	222,852	13,191	438,734	211,204	227,530	424,211	205,031	219,179
강북구	313,698	152,464	161,234	309,255	150,745	158,510	4,443	309,138	150,028	159,110	304,191	148,057	156,134
도봉구	332,586	161,944	170,642	329,957	160,885	169,072	2,629	328,243	159,569	168,674	325,292	158,370	166,922
노원구	543,499	263,655	279,844	538,768	261,563	277,205	4,731	534,096	258,737	275,359	528,268	256,124	272,144
은평구	466,243	225,813	240,430	460,651	223,388	237,263	5,592	462,552	223,452	239,100	456,582	220,818	235,764
서대문구	321,345	151,132	170,213	308,193	146,320	161,873	13,152	318,874	149,438	169,436	304,508	144,208	160,300
마포구	368,841	175,326	193,515	356,881	170,311	186,570	11,960	368,181	174,194	193,987	355,405	168,819	186,586
양천구	452,111	222,503	229,608	446,643	220,035	226,608	5,468	445,591	218,975	226,616	439,910	216,381	223,529

(Tidy Data)

1) Tidy Data

year 변수

행정구역별(읍면동)	2017							2018					
	총인구 (명)	남자 (명)	여자 (명)	내국인-계 (명)	내국인-남자 (명)	내국인-여자 (명)	외국인-계 (명)	총인구 (명)	남자 (명)	여자 (명)	내국인-계 (명)	내국인-남자 (명)	내국인-여자 (명)
전국	51,422,507	#####	#####	49,943,260	24,922,392	25,020,868	1,479,247	51,629,512	#####	#####	49,977,951	24,931,554	25,046,400
읍부	4,794,377	2,456,558	2,337,819	4,625,994	2,347,613	2,278,381	168,383	4,984,023	2,557,807	2,426,216	4,787,535	2,431,315	2,355,220
면부	4,835,090	2,498,157	2,336,933	4,592,308	2,324,414	2,267,894	242,782	4,730,290	2,453,691	2,276,599	4,464,067	2,263,139	2,200,928
동부	41,793,040	#####	#####	40,724,958	20,250,365	20,474,593	1,068,082	41,915,199	#####	#####	40,726,349	20,237,100	20,489,249
서울특별시	9,741,871	4,757,642	4,984,229	9,397,944	4,592,393	4,805,551	343,927	9,673,936	4,718,972	4,954,964	9,299,511	4,536,552	4,763,459
종로구	157,277	76,670	80,607	146,298	71,757	74,541	10,979	157,967	77,147	80,820	145,911	71,611	74,300
중구	127,896	62,195	65,701	117,631	57,391	60,240	10,265	129,797	62,975	66,822	117,809	57,240	60,569
용산구	223,898	108,471	115,427	207,893	99,759	108,134	16,005	226,938	110,685	116,253	207,812	99,806	108,006
성동구	302,367	149,060	153,307	291,931	144,018	147,913	10,436	306,796	150,623	156,173	295,489	145,220	150,269
광진구	363,934	177,156	186,778	345,535	168,833	176,702	18,399	362,304	175,762	186,542	342,575	166,730	176,045
동대문구	357,380	176,460	180,920	339,438	168,977	170,461	17,942	358,141	176,323	181,818	338,293	167,999	170,300
중랑구	396,892	196,346	200,546	390,879	193,871	197,008	6,013	391,668	193,599	198,069	385,082	190,871	194,211
성북구	445,417	215,068	230,349	432,226	209,374	222,852	13,191	438,734	211,204	227,530	424,211	205,031	219,179
강북구	313,698	152,464	161,234	309,255	150,745	158,510	4,443	309,138	150,028	159,110	304,191	148,057	156,134
도봉구	332,586	161,944	170,642	329,957	160,885	169,072	2,629	328,243	159,569	168,674	325,292	158,370	166,922
노원구	543,499	263,655	279,844	538,768	261,563	277,205	4,731	534,096	258,737	275,359	528,268	256,124	272,144
은평구	466,243	225,813	240,430	460,651	223,388	237,263	5,592	462,552	223,452	239,100	456,582	220,818	235,764
서대문구	321,345	151,132	170,213	308,193	146,320	161,873	13,152	318,874	149,438	169,436	304,508	144,208	160,300
마포구	368,841	175,326	193,515	356,881	170,311	186,570	11,960	368,181	174,194	193,987	355,405	168,819	186,586
양천구	452,111	222,503	229,608	446,643	220,035	226,608	5,468	445,591	218,975	226,616	439,910	216,381	223,529

(Tidy Data)

1) Tidy Data

구분할 변수 하나의 변수

행정구역별(읍면동)	2017							2018						
	총인구 (명)	남자 (명)	여자 (명)	내국인-계 (명)	내국인-남자 (명)	내국인-여자 (명)	외국인-계 (명)	총인구 (명)	남자 (명)	여자 (명)	내국인-계 (명)	내국인-남자 (명)	내국인-여자	
전국	51,422,507	#####	#####	49,943,260	24,922,392	25,020,868	1,479,247	51,629,512	#####	#####	49,977,951	24,931,554	25,046,397	
읍부	4,794,377	2,456,558	2,337,819	4,625,994	2,347,613	2,278,381	168,383	4,984,023	2,557,807	2,426,216	4,787,535	2,431,315	2,356,220	
면부	4,835,090	2,498,157	2,336,933	4,592,308	2,324,414	2,267,894	242,782	4,730,290	2,453,691	2,276,599	4,464,067	2,263,139	2,200,928	
동부	41,793,040	#####	#####	40,724,958	20,250,365	20,474,593	1,068,082	41,915,199	#####	#####	40,726,349	20,237,100	20,489,249	
서울특별시	9,741,871	4,757,642	4,984,229	9,397,944	4,592,393	4,805,551	343,927	9,673,936	4,718,972	4,954,964	9,299,511	4,536,552	4,763,459	
종로구	157,277	76,670	80,607	146,298	71,757	74,541	10,979	157,967	77,147	80,820	145,911	71,611	74,300	
중구	127,896	62,195	65,701	117,631	57,391	60,240	10,265	129,797	62,975	66,822	117,809	57,240	60,569	
용산구	223,898	108,471	115,427	207,893	99,759	108,134	16,005	226,938	110,685	116,253	207,812	99,806	108,006	
성동구	302,367	149,060	153,307	291,931	144,018	147,913	10,436	306,796	150,623	156,173	295,489	145,220	150,269	
광진구	363,934	177,156	186,778	345,535	168,833	176,702	18,399	362,304	175,762	186,542	342,575	166,730	175,845	
동대문구	357,380	176,460	180,920	339,438	168,977	170,461	17,942	358,141	176,323	181,818	338,293	167,999	170,294	
중랑구	396,892	196,346	200,546	390,879	193,871	197,008	6,013	391,668	193,599	198,069	385,082	190,871	194,211	
성북구	445,417	215,068	230,349	432,226	209,374	222,852	13,191	438,734	211,204	227,530	424,211	205,031	219,180	
강북구	313,698	152,464	161,234	309,255	150,745	158,510	4,443	309,138	150,028	159,110	304,191	148,057	156,134	
도봉구	332,586	161,944	170,642	329,957	160,885	169,072	2,629	328,243	159,569	168,674	325,292	158,370	166,922	
노원구	543,499	263,655	279,844	538,768	261,563	277,205	4,731	534,096	258,737	275,359	528,268	256,124	272,144	
은평구	466,243	225,813	240,430	460,651	223,388	237,263	5,592	462,552	223,452	239,100	456,582	220,818	235,764	
서대문구	321,345	151,132	170,213	308,193	146,320	161,873	13,152	318,874	149,438	169,436	304,508	144,208	160,300	
마포구	368,841	175,326	193,515	356,881	170,311	186,570	11,960	368,181	174,194	193,987	355,405	168,819	186,586	
양천구	452,111	222,503	229,608	446,643	220,035	226,608	5,468	445,591	218,975	226,616	439,910	216,381	223,529	

(Tidy Data)

2) How to ?

[자료 준비]

https://github.com/KimJiSeong1994/TidyData_lecture

(Tidy Data)

2) How to ?

#----- Step 1) read data -----

```
install.packages(c("tidyverse", "readxl"))  
library(tidyverse) ; library(readxl)
```

```
data = read_xlsx("인구_가구_및_주택_읍면동(2017-2018).xlsx", col_names = F)
```

col_names = F : 데이터를 불러올 때 변수 명을 지정 하지 않음

#----- Step 2) 변수명 변경 -----

```
names(data) <- data[2, 1] # names() 변수 명을 출력 해줌
```

```
names(data) <- paste(data[1, 1], names(data), sep = "_")
```

```
data <- data[-c(1:2), 1] # 변수 명으로 사용한 Obs 제거
```

(Tidy Data)

2) How to ?

#----- Step 3) 전처리 -----

data %>%

**filter(!`행정구역별(읍면동)_행정구역별(읍면동)` %in% c("전국", "읍부", "면부", "동부"),
str_detect(`행정구역별(읍면동)_행정구역별(읍면동)`, "시\$")) %>%**

str_detect(data, 찾을 목록, “찾을 패턴”)

“ 찾을 패턴 ” 에는 정규표현식 or 문자 그래도 넣어도 됨

gather(type, value, `2017_총인구 (명)` : `2018_주택이외의 거처 (호)`) %>%

gather(data, 모든 데이터 변수 이름 , 모든 데이터의 변수 이름, 모을변수 : 모을변수)

<https://github.com/gadenbuie/tidyexplain>

separate(type, into = c("year", "type"), sep = "_") %>%

mutate(type = as.factor(type)) %>%

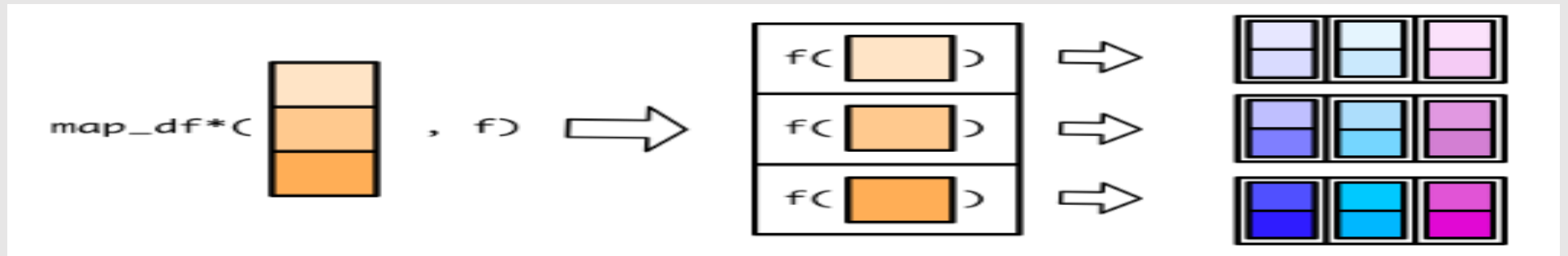
(Tidy Data)

2) How to ?

...

```
mutate(type = as.factor(type)) %>%  
map_dfc(function(x) {  
  if(str_detect(x[[1]], "[0-9]")) {  
    as.numeric(x)  
  } else {  
    as.character(x)  
  }  
})
```

map_dfc() :



(출처 : Functional Programming <https://dcl-prog.stanford.edu/purrr-extras.html>)

(Tidy Data)

2) How to ?

```
...  
mutate(type = as.factor(type)) %>%  
map_dfc(function(x) {  
  if(str_detect(x[[1]], "[0-9]")) {  
    as.numeric(x)  
  } else {  
    as.character(x)  
  }  
}) %>%  
group_split(type) # type 별로 데이터를 분리할 수 있음
```

(Tidy Data)