R에서 남한지도 그려보기

김병건

2019 9 7

필요한 패키지 설치

```
## -- Attaching packages ---------------------------------------------------------------------

## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   0.8.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

## -- Conflicts ------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

## Linking to GEOS 3.6.1, GDAL 2.2.3, PROJ 4.9.3
```

남한지도를 그려보겠습니다.

동(또는 면/리)단위로 구역이 나뉘어진 지도입니다.

```
## Reading layer `TL_SCCO_EMD' from data source `C:\work_git\drawing_south_korea_map_in_R\ma
## Simple feature collection with 5047 features and 3 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:           xmin: 746110.3 ymin: 1458754 xmax: 1387948 ymax: 2068444
## epsg (SRID):    NA
## proj4string:    +proj=tmerc +lat_0=38 +lon_0=127.5 +k=0.9996 +x_0=1000000 +y_0=2000000 +e
```

지도를 그리기 위해서 알아야할 내용들

지도를 그리기 위해서는 우선, 지도를 구성하고 있는 데이터에 대해서 알아야 합니다.

- 지도 데이터, shape file 에 대한 설명은 아래를 링크를 참고하세요.
- https://endofcap.tistory.com/3
- http://www.snpo.kr/data//file/meetshare_epilogue/2949849334_CGywx7W5_slide3.pdf

shape file만 올바르게 작성되어 있다면, 위에서 보신 것 과 같이 쉽게 지도를 그릴 수 있씁니다.

- shape file 다운 받은 곳
- http://www.gisdeveloper.co.kr/?p=2332&unapproved=28711&moderation-hash=bf738becd6dd0f8b10fd61503ceb93e7#comment-28711

지도 데이터(shape file)의 기본구조

EMD_CD : 법정동 코드의 시/도 + 군/구 + 동 코드입니다. * 법정동 코드는 아래 링크 참고하세요. * https://financedata.github.io/posts/korea-area-code.html

EMD_ENG_NM : 동 영문 표기 EMD_KOR_NM : 동 한글 표기 (cp949로 인코딩 되어 있어서, UTF-8에서 올바르게 표기 되지 않았습니다. 뒤에서 UTF-8으로 변환 하겠습니다.) geometry : 동의 좌표를 가지고 있는 부분입니다.

shape file은 현재 'Simple feature'라른 이름으로 ISO에서 표준화한 포맷이 되었습니다.

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

```
## Simple feature collection with 5047 features and 3 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:           xmin: 746110.3 ymin: 1458754 xmax: 1387948 ymax: 2068444
## epsg (SRID):    NA
## proj4string:    +proj=tmerc +lat_0=38 +lon_0=127.5 +k=0.9996 +x_0=1000000 +y_0=2000000 +e
## First 10 features:
##      EMD_CD      EMD_ENG_NM                                    EMD_KOR_NM
## 1  11110101   Cheongun-dong                               u<U+00BF><U+ED7F>
## 2  11110102     Singyo-dong     <U+00BD><U+0171><U+00B3><U+00B5><U+00BF>
## 3  11110103  Gungjeong-dong         <U+00B1><c3><U+0064><U+00B5><U+00BF>
## 4  11110104      Hyoja-dong                 <U+023F><c0><U+06B5><U+00BF>
## 5  11110105 Changseong-dong            a<U+00BC><U+00BA><U+00B5><U+00BF>
## 6  11110106     Tongui-dong                 <c5><eb><c0><U+01F5><U+00BF>
## 7  11110107   Jeokseon-dong     <c0><fb><U+00BC><U+00B1><U+00B5><U+00BF>
## 8  11110108     Tongin-dong                       <c5><eb><c0>ε<U+00BF>
## 9  11110109     Nusang-dong <U+00B4><U+00A9><U+00BB><f3><U+00B5><U+00BF>
## 10 11110110       Nuha-dong         <U+00B4><U+00A9><c7><U+03F5><U+00BF>
##                          geometry
## 1  MULTIPOLYGON (((953700 1954...
## 2  MULTIPOLYGON (((953233.5 19...
## 3  MULTIPOLYGON (((953560.2 19...
## 4  MULTIPOLYGON (((953519.8 19...
## 5  MULTIPOLYGON (((953516.1 19...
## 6  MULTIPOLYGON (((953546.7 19...
## 7  MULTIPOLYGON (((953569.2 19...
## 8  MULTIPOLYGON (((953274.9 19...
## 9  MULTIPOLYGON (((952981.2 19...
## 10 MULTIPOLYGON (((953098.4 19...
```

지도 데이터 클리닝

EMD_KOR_NM 인코딩을 CP949에서 UTF-8로 변경해서, 깨진 한글을 복구합니다.

```
## Simple feature collection with 6 features and 3 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:           xmin: 952451 ymin: 1953139 xmax: 953700 ymax: 1955073
## epsg (SRID):    NA
## proj4string:    +proj=tmerc +lat_0=38 +lon_0=127.5 +k=0.9996 +x_0=1000000 +y_0=2000000 +e
##     EMD_CD      EMD_ENG_NM EMD_KOR_NM                        geometry
```

```
## 1 11110101   Cheongun-dong     청운동 MULTIPOLYGON (((953700 1954...
## 2 11110102     Singyo-dong     신교동 MULTIPOLYGON (((953233.5 19...
## 3 11110103  Gungjeong-dong     궁정동 MULTIPOLYGON (((953560.2 19...
## 4 11110104      Hyoja-dong     효자동 MULTIPOLYGON (((953519.8 19...
## 5 11110105 Changseong-dong     창성동 MULTIPOLYGON (((953516.1 19...
## 6 11110106     Tongui-dong     통의동 MULTIPOLYGON (((953546.7 19...
```

지도 데이터를, 데이터프레임으로 변환

shape file의 포맷은 주로 사용해왔던 데이터프레임과 함께 다루기가 어렵기 때문에, shape file을 데이터 프레임으로 변환합니다.

shape file에서 데이터 프레임으로

shape file의 데이터 dongs_shp에서 geometry컬럼 부분, 즉 spatial 부분만 뽑아 냅니다.

```
## [1] "SpatialPolygonsDataFrame"
## attr(,"package")
## [1] "sp"
```

dongs_seoul_spatial 의 안을 보면, data 와 polygons 둘로 나뉘어 있습니다.

- data 부분이 데이터 프레임으로 법정동코드, 동이름, 을 담고 있고
- polygons는 동의 모습을 폴리곤으로 담고 있습니다.

두 부분을 각각 아래와 같이 변환 합니다.

```
##     EMD_CD      EMD_ENG_NM EMD_KOR_NM
## 1 11110101   Cheongun-dong     청운동
## 2 11110102     Singyo-dong     신교동
## 3 11110103  Gungjeong-dong     궁정동
## 4 11110104      Hyoja-dong     효자동
## 5 11110105 Changseong-dong     창성동
## 6 11110106     Tongui-dong     통의동
```

```
## Regions defined for each Polygons
```

```
##       long     lat order  hole piece id group
## 1 953700.0 1954605     1 FALSE     1  1   1.1
## 2 953693.9 1954591     2 FALSE     1  1   1.1
## 3 953690.8 1954583     3 FALSE     1  1   1.1
## 4 953689.6 1954577     4 FALSE     1  1   1.1
## 5 953687.9 1954569     5 FALSE     1  1   1.1
## 6 953682.3 1954556     6 FALSE     1  1   1.1
```

2개의 데이터프레임 클리닝

dongs_data_df에 id 컬럼을 추가

- dongs_data_df 와 dongs_geometry_df를 연결하여 사용할 때, 필요한 id 컬럼을 만들겠습니다.
- EMD_CD를 charater 로 변경하겠습니다.
- EMD_ENG_NM charater 로 변경하겠습니다.

```
##     EMD_CD      EMD_ENG_NM EMD_KOR_NM id
## 1 11110101   Cheongun-dong     청운동  1
## 2 11110102     Singyo-dong     신교동  2
## 3 11110103  Gungjeong-dong     궁정동  3
## 4 11110104      Hyoja-dong     효자동  4
## 5 11110105 Changseong-dong     창성동  5
## 6 11110106     Tongui-dong     통의동  6
```

법정동 코드 8자리에서 10자리로 변경

EMD_CD컬럼이 법정동 코드를 의미합니다. 법정동 코드는 총 10자리로 구성된 숫자이며, 아래와 같이 조합합니다.

"시/도" + "시/군/구" + "읍/면/동" + "리"

EMD_CD컬럼은 읍/면/동 코드까지 가진 8자리로 예상됩니다. 8자리가 맞는지 확인해보고, 뒤에 00을 붙여서, 총 10자리인 법정동 코드로 변경하겠습니다.

```
## [1] 8
```

모두 8자리 이기 때문에, 뒤에 00을 붙여 줍니다.

```
##       EMD_CD      EMD_ENG_NM EMD_KOR_NM id
## 1 1111010100   Cheongun-dong     청운동  1
## 2 1111010200     Singyo-dong     신교동  2
## 3 1111010300  Gungjeong-dong     궁정동  3
## 4 1111010400      Hyoja-dong     효자동  4
## 5 1111010500 Changseong-dong     창성동  5
## 6 1111010600     Tongui-dong     통의동  6
```

dongs_data_df에 id 컬럼 타입을 숫자형으로 변환

dongs_geometry_df의 id 컬럼 타입을 숫자형으로 변환합니다.

```
##       long     lat order  hole piece id group
## 1 953700.0 1954605     1 FALSE     1  1   1.1
## 2 953693.9 1954591     2 FALSE     1  1   1.1
## 3 953690.8 1954583     3 FALSE     1  1   1.1
## 4 953689.6 1954577     4 FALSE     1  1   1.1
## 5 953687.9 1954569     5 FALSE     1  1   1.1
## 6 953682.3 1954556     6 FALSE     1  1   1.1
```

2개의 데이터프레임 저장하기

지도 그리기

앞에서 shape file에서 데이터를 추출해서, 클리닝을 한 뒤에 csv 파일로 저장하였습니다. CSV파일에서 데이터를 읽어와서, 지도를 그려보겠습니다.

동이름, 동 지도 데이터 읽기

factor로 되어 있는것을 character로 변환합니다.

```
##   X     EMD_CD      EMD_ENG_NM EMD_KOR_NM id
## 1 1 1111010100   Cheongun-dong     청운동  1
## 2 2 1111010200     Singyo-dong     신교동  2
## 3 3 1111010300  Gungjeong-dong     궁정동  3
## 4 4 1111010400      Hyoja-dong     효자동  4
## 5 5 1111010500 Changseong-dong     창성동  5
## 6 6 1111010600     Tongui-dong     통의동  6
```

```
##   X     long     lat order  hole piece id group
## 1 1 953700.0 1954605     1 FALSE     1  1   1.1
## 2 2 953693.9 1954591     2 FALSE     1  1   1.1
## 3 3 953690.8 1954583     3 FALSE     1  1   1.1
## 4 4 953689.6 1954577     4 FALSE     1  1   1.1
## 5 5 953687.9 1954569     5 FALSE     1  1   1.1
## 6 6 953682.3 1954556     6 FALSE     1  1   1.1
```

첫 지도 그리기

동별로 나뉘어진 첫 지도 그리기 :-)

좀더 예쁜 지도로

- 지도의 삐쭉 나온 선들은 가려지지가 않네요 -_-;
- 동별로 색도 칠해 봅니다. geom_polygon 의 aes(fill = id)
- 가로/세로 비율이 같은 지도를 그리기 coord_fixed(1)

서울/종로구 데이터만 뽑아서 사용하겠습니다. (전국 데이터는 너무 커서)

데이터를 뽑기 위해서, 2개의 데이터프레임을 합치고, 법정동 코드에서, 서울/종로구 데이터만 필터합니다.

```
##   id X.x     EMD_CD    EMD_ENG_NM EMD_KOR_NM X.y     long     lat order
## 1  1   1 1111010100 Cheongun-dong     청운동   1 953700.0 1954605     1
## 2  1   1 1111010100 Cheongun-dong     청운동   2 953693.9 1954591     2
## 3  1   1 1111010100 Cheongun-dong     청운동   3 953690.8 1954583     3
## 4  1   1 1111010100 Cheongun-dong     청운동   4 953689.6 1954577     4
## 5  1   1 1111010100 Cheongun-dong     청운동   5 953687.9 1954569     5
## 6  1   1 1111010100 Cheongun-dong     청운동   6 953682.3 1954556     6
##    hole piece group
## 1 FALSE     1   1.1
## 2 FALSE     1   1.1
## 3 FALSE     1   1.1
## 4 FALSE     1   1.1
## 5 FALSE     1   1.1
## 6 FALSE     1   1.1
```

서울/종로구 동이름도 함께

id,EMD_KOR_NM 로 group_by를 하면, 각 동,동이름별로 데이터가 그룹이 되고, 동이름을 표시할 위도/경도를, 동의 외곽을 구성하는 점들의 위도/경도의 평균으로 정합니다.

1958000
1956000
1954000
1952000
lat
952000
954000
956000
958000
long
id
80
60
40
20

보다 큰 크기로 그리기 (작아서 잘 안보임)

동옆에 바차트 그리기

geom_rect()를 사용해서, 지도위에 위도/경도에 해당하는 사각형을 그려서, 바차트 용도로 사용하였습니다.

- geom_subplot() 이 동작안하는 것 같은대, 잘 되시는 분은 연락 주세요 ^^

마무리 하며

R로 만들어 볼까? 했던 것이 있어서, kormap2014을 사용하면 되겠지... 하면서 시작 했다가, 이런저런 이유로, kormap2014을 사용하기가 어렵게 되면서... 일이 점점 커지다가... 결국은 shape file에서 부터 직접 그리게 되었습니다. :-)

R에서 지도 그리시는 분들에게 도움이 되셨기를 바라며, 이만 마치겠습니다.

김병건 (skysign at_sign gmail.com) 2019년 9월 7일