

지리정보시스템과 프롭테크: 도구의 진화와 미래 방향

이광춘^{1,*}

2023-09-01

요약

오픈소스 소프트웨어와 공간정보 공공데이터가 넘쳐나고 있는 가운데, 시민 데이터 과학자의 중요성이 증가하고 있다. 그러나, 지난 50년 동안 빠르게 발전해 온 지리 정보 시스템(GIS) 분야를 부동산 기술(프롭테크) 관점에서 바라보면 가속되고 있는 도구의 발전이 가져올 미래 방향을 유추할 수 있다. 다른 과학 및 엔지니어링 분야에서와 마찬가지로, GIS 분야 또한 지속적인 도구의 발전을 통해 큰 혁신이 일어나고 있다. 본 연구에서는 GIS 도구의 발전 과정을 살펴보고, 사례로 제시된 코드와 코드를 지탱하는 도구를 코드 뒤에 숨겨진 기술 생태계를 이해하고 프롭테크 미래발전 방향을 함께 가늠하고자 한다.

¹ 한국 R 사용자회, 서울, 대한민국

* Correspondence: 이광춘 <victor@r2bit.com>

1 들어가며

최근 OpenAI사가 공개한 챗GPT는 인공지능 분야 자체도 큰 반향을 일으키고 있지만, 다른 분야에도 커다란 변화를 일으키고 있는데 가장 큰 이유는 쓸만한 도구가 마련되었기 때문이다. 시대에 맞는 적절한 도구 없이 지리정보시스템과 프롭테크를 지향하는 것은 커다란 모순이기 때문에 지리학에 대한 주요 개념을 먼저 살펴본 후에 도구의 진화를 살펴보고 진화된 도구를 통한 사례로 프롭테크의 미래를 함께 조망해보자. [5]

지리학(Geography)은 가장 오래된 과학 중 하나로 꼽히며, 학생때부터 교실 여기저기에서 볼 수 있는 지도가 아마도 처음 접한 데이터 시각화 산출물일 것이다. 지리학은 지구의 곡률을 정확히 파악하여 지구의 크기와 모양, 중력에 대한 이해를 높이는 데 중요한 역할을 수행했을 뿐만 아니라 대항해 시대 지도가 매우 큰 기여를 했고, 현대에 와서 지리학 데이터 응용 범위는 더욱 확장되어, 민간과 국방 범위를 가리지 않고 상업적으로 입점위치 선정, 우주 탐사, 군사 작전, 인구이동 패턴 분석 등 다방면에서 요긴한 역할을 수행하고 있다. 지리정보시스템(GIS)은 공간 위치를 분석하고 정보를 지도와 필요한 경우 3D로 시각화하여 데이터 패턴, 관계, 상황 등에 대한 심도 깊은 이해를 제공하여 의사결정을 지원한다.

먼저 좌표계를 살펴보자. 지도 자체가 불가피하게 왜곡된 현실을 시각적으로 표현하기 때문에 ‘현실의 축소판’(Maps are scale models of reality)이라고 할 수 있다. 예를 들어, 사과 껍질을 벗겨 펼친 것처럼, 평면화시켜 지도를 제작한다. 특히, **지리 좌표계(Geographic Coordinate System)**와 **투영 좌표계(Projected Coordinate System)**는 지도 작성과 공간 분석에서 중요한 개념이다. 간단히 요약하면, 지리 좌표계는 정확한 위치 정보를 제공하는 반면, 투영 좌표계는 거리와 면적 측정에 더 적합하다. 지리 좌표계는 지구상 위치를 설명하며, 지구가 둥글기 때문에 주로 각도 단위로 측정되고, 기준점에 상관없이 위치가 달라지지 않으므로 절대 좌표계다. 투영 좌표계는 지리 좌표계를 평면화한 것으로, 투영 좌표계를 사용할 경우 직선을 이용해 거리를 측정할 수 있어, 토지 이용 계획, 건물 배치, 거리 측정 등에서 유용하다. [1]

다음으로 지도학의 대표적인 파일 형식을 살펴보자. **지도학(Cartography)**은 지도를 작성하는 중요한 실무분야로 과거 벡터 데이터 형식으로 대표적인 **Shapefile** 파일을 통해 점, 다각형, 선으로 지리적 정보를 표현했다면, GIS 분야에 오픈소스 소프트웨어 문화가 확산되면서 **Geopackage**, **GeoJSON**, **KML**과 같은 다른 파일형식도 인기를 얻고 있다. GIS 데이터는 지도 뿐만 아니라 R 혹은 파이썬 프로그래밍 언어를 활용하여 다른 형태의 (비)정형 데이터와 결합한 교차분석이 일반화되고 있다.

전통적인 **벡터(Vector) Shapefile** 파일과 함께 **래스터(Raster)** 데이터는 GIS에서 큰 축을 담당하는 대표적인 데이터 형식이다. 벡터 데이터가 점, 선, 다각형 등의 기하학적 형태로 주로 경계선, 위치 정보 등을 정확하게 표현할 때 사용되는 반면, 래스터 데이터는 픽셀이나 셀 배열로 정보를 표현한다는 점에서 차이가 난다. 각 픽셀은 고유값을 가지며, 색상, 높이, 밀도 등 다양한 형태 정보를 표현하는데 용이하다.

GIS에서 계층(레이어, Layer)별로 나눠 데이터를 표현하고, 이를 결합하여 공간정보 지도형태 산출물을 얻는다. 한 구획의 토지(parcel), 공간 정보를 점, 선, 다각형 같은 벡터 데이터로 표현하고 비공간 정보는 소유권, 토지 이용, 구역 분류, 평가 가치 등 다양한 속성(칼럼) 데이터프레임과 결합하여 작업한다. Figure 1 에 실제 세상이 벡터와 래스터 데이터로 표현하는 사례가 제시되어 있다.

2 도구의 진화

GIS, 좌표계, 파일 형식, 계층(Layer), 교차분석 등 주요개념을 바탕으로 도구 진화과정을 살펴보자. 초기 PC가 보급되면서 1980년대 GIS 시스템은 사일로 형태 고가 시스템이었으며, 1990년대에는 1980년대 개발된 GIS 시스템간 연결이 시작되면서 스파게티 현상이 심해지면서 새로운 애플리케이션과 파일 형식이 도입되는데 필요 이상의 낭비가 심해지는 구조가 되었다.

2000년대 GDAL(Geospatial Data Abstraction Layer)이 등장하여 개발자가 각 파일 형식별로 서로 다른 드라이버를 작성하는 대신 GDAL 클라이언트 드라이버만 개발하면 되기 때문에 일대 혁신이 일어났다. 2010년대는 인공위성을 통한 대량의 데이터가 넘쳐나며 이를 저장하기 위해서 클라우드 서비스가 우후죽순처럼 생겨났고 각기 다른 파편화된 진화와

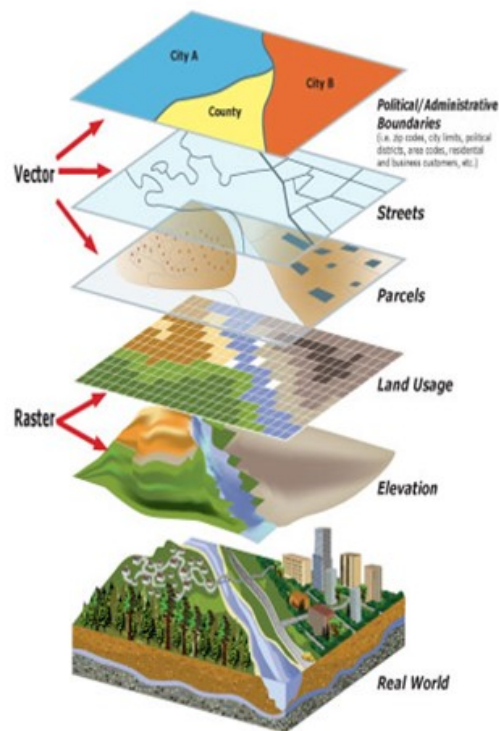


그림 1: GIS 벡터와 래스터 계층, 출처: <http://gis.sbcounty.gov/>

정이 일어났다. 2020년대 넘어서면서 1990년대와 유사한 상황이 재현되었고, 이를 해결하고 새로운 전환점을 만들기 위해 OpenEO가 제시되며 사용자와 데이터 센터 사이에서 새로운 표준으로 자리잡게 되면서 R, 파이썬 등 데이터 과학 프로그래밍 언어를 통한 공간정보 가치 창출이 용이해졌다. [3]

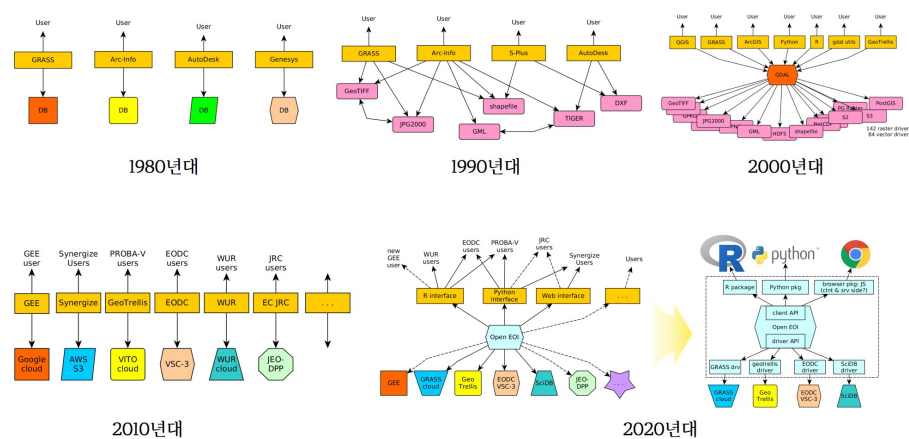


그림 2: GIS 도구 진화과정

3 GIS 생태계

인간은 오랜 역사 동안 다양한 도구를 개발하고 사용해 왔으며, GIS에서도 도구는 단순한 계산부터 복잡한 데이터 분석, 시각화를 통한 의사결정지원, 최근 거대언어모형(LLM) 인공지능까지 다양한 형태로 기여하고 있다.

현재 직면한 문제를 해결하는데 과거 도구를 사용하는 것은 마치 철기시대에 석기시대 도구를 사용하는 것과 다름이 없다. 따라서, 데이터를 통해 가치를 만드는 데이터 과학자 입장에서 도구의 선택은 매우 중요할 수 밖에 없다.

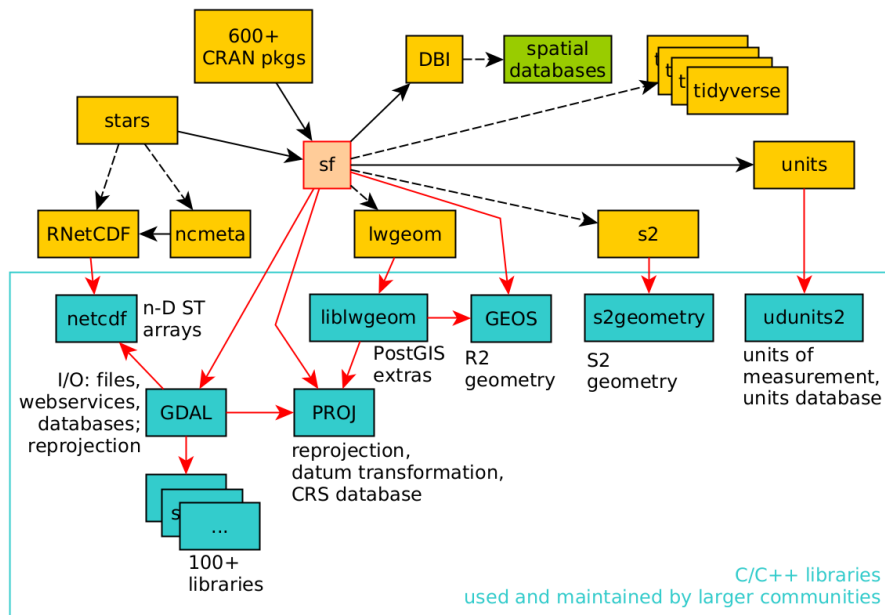


그림 3: **sf** 패키지와 의존성 - 화살표는 강한 의존성, 점선 화살표는 약한 의존성

현재 R 공간정보 생태계는 공간정보 도구 진화과정을 몇차례 경험한 후에 Figure 3 처럼 자리를 잡았다. 결론부터 들어가면 **sf** 패키지가 생태계의 중심으로 자리 잡았으며, 사용자와 개발자는 **sf** 패키지를 통해 상당수 공간정보 데이터 문제를 해결할 수 있게 되었다. 하지만, **sf** 패키지는 C/C++ 라이브러리에 크게 의존하기 때문에 각 라이브러리를 살펴보는 것은 향후 GIS 개발자와 분석가로 현업에서 활약하는데 큰 도움이 될 것으로 보인다.

GDAL, GEOS, PROJ, liblwgeom, s2geometry, NetCDF, udunits2는 C/C++ 라이브러리를 개발, 유지보수, 지원하는 커뮤니티가 존재하는 반면 R, 파이썬, 줄리아(Julia), 자바스크립트는 대화형 인터페이스를 통해 라이브러리를 활용하는 커뮤니티도 존재한다. [2]

- GDAL (Geospatial Data Abstraction Library) 라이브러리는 공간 데이터 처리에 있어 멀티툴의 대명사인 스위스군 칼(SAK, Swiss Army Knife)이라는 별명을 갖고 있고 100개가 넘는 다른 라이브러리와 연결되어 다양한 공간 데이터를 불러오고, 처리하고, 내보내는 기능을 제공한다.

- PROJ는 지도 투영 및 데이터 변환을 위한 라이브러리로, 하나의 좌표계에서 다른 좌표계로 좌표를 변환할 때 유용하다.
- GEOS (Geometry Engine Open Source)와 **s2geometry** 라이브러리는 기하학 연산에 사용하며, 길이, 면적, 거리를 측정하거나 연산작업에 사용되며 R^2 로 표기되며 GEOS는 평평한 2차원 공간에, S^2 로 표기되며 **s2geometry**는 구형 공간에 사용한다.
- NetCDF는 파일 형식이며 C 라이브러리로, 어떤 차원 배열도 정의할 수 있으며 특히 (기후) 모형개발 커뮤니티에서 널리 사용된다. **Udunits2**는 측정 단위의 데이터베이스 및 소프트웨어 라이브러리로, 단위의 변환과 파생 단위, 사용자 정의 단위를 처리한다. **liblwgeom**은 PostGIS 구성 요소로, **GDAL**, **GEOS**에서 누락된 몇 가지 루틴을 포함한다. [3]

4 사례

sf 패키지가 추상화된 함수를 제공하기 때문에 R 공간정보 생태계에서 **GDAL** 라이브러리를 GIS 분석가가 직접적으로 다룰 일은 없다. 하지만, GIS 도구를 활용하여 2023년 7월 기준 시도별 인구수를 대한민국 지도위에 도식화하기 위해서는 R 공간정보 생태계를 구성하는 다양한 도구가 꼭 필요하다. 먼저, 지도 데이터를 다루기 위해 **sf** 패키지를 사용하고 통계청 (KOSIS) 데이터를 처리하기 위해 **tidyverse** 패키지 도구를 활용하고 오픈지도 개발자가 공개한 **geojson** 파일을 결합하여 인구수를 시도 수준에서 시도별 인구수 색상을 달리하여 지도위에 시각화한다. 작성된 코드는 R로 작성되었지만 파이썬 진영에도 공간정보 생태계도 유사한 도구가 준비되어 있어 각자 사용하기 좋은 도구를 가지고 의미있는 결과물을 만들어내고 있다.[4]

```
library(sf)
library(tidyverse)
sf_use_s2(FALSE)

## 지도
korea_map <- read_sf("data/HangJeongDong_ver20230401.geojson")

sido_map <- korea_map |>
  group_by(sidonm) |>
  summarise(geometry = sf::st_union(geometry))

## 23년 7월 인구수 (KOSIS) 행정구역별, 성별 인구수
pop_tbl <- read_csv("data/행정구역_시군구_별__성별_인구수
↪ _20230831223248.csv",
```

```

      locale=locale(encoding="euc-kr"), skip = 1) |>
    set_names(c("sidonm", "인구수")) |>
    mutate(sidonm = if_else(sidonm == "강원특별자치도", "강원도",
      ↪ sidonm))

sigo_gg <- sido_map |>
  left_join(pop_tbl) |>
  ggplot() +
    geom_sf(aes(geometry = geometry, fill = cut(인구수, 10)),
      ↪ show.legend = FALSE) +
    ggrepel::geom_label_repel(aes(label = sidonm, geometry =
      ↪ geometry),
                                size = 3, stat = "sf_coordinates")
                                ↪ +
    theme_void() +
    scale_fill_brewer(palette = "OrRd")

ragg::agg_jpeg("images/GIS_tools.jpeg",
              width = 10, height = 7, units = "in", res = 600)

sigo_gg
dev.off()

```

5 마무리

공간정보 개발자와 사용자가 개발하는 코드는 기계보다 사람 친화적으로 바뀌었으며, 효과적인 디버그와 신속한 개발을 위한 파이프 철학도 도입되어 생산성 향상이 비약적으로 높아졌고 도구의 추상화 수준도 대폭 향상되었다. IoT와 인공위성을 통해 엄청난 공간정보 데이터가 축적되고 있지만, GIS 도구가 꾸준히 발전하면서 이제 누구나 이러한 도구를 활용하여 공간정보 데이터를 통해 의미있는 산출물을 제작하고 도구도 개발할 수 있게 되었고, 다른 한편으로는 도구 없이 도구에 대한 이해없이 프롭테크를 논하는 것조차 의미없는 시대로 접어들고 있다.

참고문헌

- [1] Robin Lovelace, Jakub Nowosad, and Jannes Muenchow. *Geocomputation with R*. CRC Press, 2019.

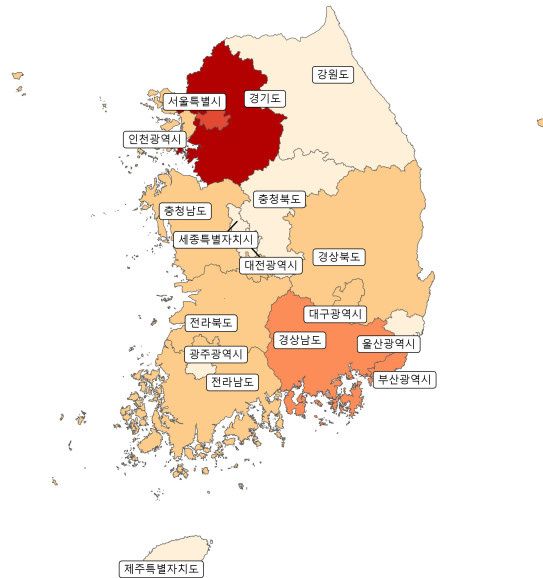


그림 4: GIS 도구 활용 대한민국 시도별 인구수

- [2] Edzer Pebesma and Roger Bivand. *Spatial Data Science: With Applications in R*. Chapman and Hall/CRC, 2023. DOI: [10.1201/9780429459016](https://doi.org/10.1201/9780429459016). URL: <https://doi.org/10.1201/9780429459016>.
- [3] Edzer Pebesma et al. *OpenEO: a GDAL for Earth Observation Analytics*. Accessed: 2023-09-01. 2016. URL: <https://r-spatial.org/2016/11/29/openeo.html>.
- [4] Hadley Wickham et al. “Welcome to the Tidyverse”. In: *Journal of open source software* 4.43 (2019), p. 1686.
- [5] 이광춘 · 신종화. *POD 챗 GPT 유닉스 셸: 데이터 과학 자동화 도구*. 한국 R 사용자회, Aug. 2023.