

보건학, 역학, 임상 통계 R 시각화:

I. 서론: 보건의료 연구를 위한 R 시각화의 패러다임

데이터 시각화는 현대 보건학 연구의 핵심 구성 요소입니다. 시각화는 복잡한 데이터 세트를 이해, 해석하고, 동료 과학자, 정책 입안자, 그리고 대중에게 연구 결과를 효과적으로 전달하는 가장 강력한 수단입니다.¹ 특히 역학 분야에서 그래프와 차트를 생성하고 분석하는 기술은 공중 보건 개입을 촉진하고² 궁극적으로 공중 보건 의사결정에 정보를 제공(inform)하는데 필수적입니다.³

전통적인 상용 통계 소프트웨어(예: SPSS, SAS, Stata)가 오랫동안 이 분야에서 사용되어 왔지만, R 프로그래밍 언어는 보건의료 연구 커뮤니티에서 독보적인 표준으로 자리 잡았습니다. 이러한 패러다임 전환은 R이 제공하는 몇 가지 근본적인 이점에 기인합니다.

첫째, R은 오픈소스이며 무료로 제공됩니다.⁴ 이는 단순한 비용 절감의 문제를 넘어섭니다. 특히 자원이 제한된 기관이나 개발도상국의 연구자들에게 값비싼 라이선스 비용은 과학적 참여의 거대한 장벽이 될 수 있습니다.⁴ R은 이러한 재정적 장벽을 제거함으로써 전 세계 보건 연구의 형평성을 증진합니다.

둘째, R은 "스크립트(script)" 또는 "코드" 기반의 워크플로우를 채택하고 있습니다. 이는 "포인트 앤 클릭(point & click)" 방식의 그래픽 사용자 인터페이스(GUI)와 근본적으로 대조됩니다.⁶ R 스크립트는 분석의 모든 단계를 문서화하고, 공유하며, 체계적으로 재현할 수 있게 합니다.⁵ 공중 보건 정책 결정은 수백만 명의 삶에 영향을 미칠 수 있으며, 이러한 결정의 근거가 되는 분석은 투명하고 검증 가능해야 합니다.² R의 "재현 가능한 연구(reproducible science)"⁵ 기능은 단순한 편의성이 아니라, 공중 보건 결정의 과학적 근거에 대한 투명성을 보장하는 방법론적, 윤리적 이점을 제공합니다.

셋째, R은 ggplot2¹, shiny⁵ 등 강력한 시각화 패키지를 포함하여, 단순한 분석부터 복잡한 통계 모델링⁷, 대화형 앱 설계에 이르기까지 방대한 최신 패키지 생태계(ecosystem)를 제공합니다.⁸ 또한, 전 세계의 광범위하고 활동적인 사용자 커뮤니티는 지속적인 프로그램 기능 개발과 신속한 문제 해결 지원을 보장합니다.⁵

본 가이드는 tidyverse⁷ 접근 방식을 중심으로 구성됩니다. tidyverse는 데이터 조작을 위한 dplyr, 데이터 정제를 위한 tidyverse, 그리고 시각화를 위한 ggplot2 등, 일관된 문법과 데이터 철학을 공유하는 핵심 패키지 모음입니다. ggplot2는 "그래픽 문법(Grammar of Graphics)"¹⁰이라는 이론적 토대를 기반으로 하며, 사용자가 사전 정의된 차트 목록에 얹매이지 않고¹⁰ 독립적인

구성 요소를 조합하여 고품질의 맞춤형 그래픽을 생성할 수 있게 합니다.¹

일부에서는 R이 "가파른 학습 곡선(stEEP learning curve)"¹¹을 가지고 있으며 처음 배울 때 "위협적(intimidating)"⁵일 수 있다고 지적합니다. 이는 종종 단점으로 인식됩니다. 하지만 본 가이드의 대상 독자인 대학원생 및 보건의료 종사자에게 이는 오히려 교육적 자산이 될 수 있습니다. ggplot2의 문법¹⁰이나 R의 스크립트⁶를 배운다는 것은 통계 분석의 기본 논리를 체계적으로 학습하는 과정입니다. "그래픽 문법"을 배우는 것은 "전문가들이 사용하는 구조화된 분석 접근 방식"을 개발하는데 도움을 줍니다.¹⁰ 이는 "포인트 앤 클릭" 소프트웨어⁶가 제공할 수 없는 깊은 수준의 이해를 촉진합니다.

본 가이드는 독자가 R 시각화의 기본 문법(II장)에서 시작하여, 자신의 핵심 분야인 역학(III장, IV장) 및 임상 통계(V장)에 특화된 기법을 배우고, 이 결과물을 학술지 수준으로 다듬(VI장)어, 최종적으로 대화형 웹 애플리케이션(VII장)으로 배포하는 전체적인 학습 여정을 따릅니다. 아래의 표 1은 본 가이드에서 다룰 핵심 패키지들을 요약한 로드맵입니다.

표 1: 보건의료 시각화를 위한 R 패키지 핵심 요약 (본 가이드의 로드맵)

이 표는 본 가이드에서 다룰 핵심 R 패키지들을 요약하여, 독자가 특정 시각화 작업을 수행하기 위해 필요한 도구와 가이드의 해당 섹션을 빠르게 참조할 수 있도록 설계되었습니다.

패키지 (Package)	핵심 기능 (Primary Function)	대표적 보건학 적용 사례 (Public Health Application)	관련 챕터 (Guide Section)
ggplot2	그래픽 문법 (Grammar of Graphics) 기반 시각화 엔진	기술 통계(히스토그램, 산점도), 기본 추세선	II, III, IV
incidence2	발병률 계산 및 시각화	감염병 유행 곡선(Epicurve) 작성 및 분석	III
ggsurvfit	생존 분석 시각화	카플란-마이어(Kaplan-Meier) 생존 곡선 (위험 태이블 포함)	V

sf	공간 데이터 처리 (Simple Features)	공간 데이터(Shapefile) 로드, 처리, 및 보건 데이터와 결합	IV
tmap	주제도 (Thematic Maps)	코로플레스 맵(Choropleth Map) 등 정적/동적 지도 제작	IV
metafor	메타 분석 (Meta-Analysis)	포레스트 플롯(Forest Plot) 생성	V
patchwork	다중 플롯 조합	학술지용 다중 패널 그림(Figure) 생성	VI
ggrepel	텍스트 라벨링	플롯 내 텍스트 겹침 방지 (예: 산점도 라벨링)	VI
ggpubr	출판용 그래프 보조	ggplot에 \$p\$-값, 유의수준 추가 및 테마 적용	VI
plotly	대화형 그래프	ggplot 객체를 동적(클릭/줌 가능) 웹 그래프로 변환	VII
shiny	웹 애플리케이션 프레임워크	공중 보건 감시(Surveillance) 대시보드 구축	VII

II. 시각화의 문법: **ggplot2** 완벽 마스터하기

R 시각화의 핵심은 ggplot2 패키지이며, 이는 "그래픽 문법(Grammar of Graphics)"¹이라는

강력한 철학을 기반으로 합니다. 이 섹션은 개별 차트 유형(예: 막대 그래프, 산점도)을 암기하는 것이 아니라, 어떠한 시각화라도 생성할 수 있는 원리를 이해하는 데 중점을 둡니다.

ggplot2의 문법은 플롯을 독립적인 구성 요소의 레이어(layers)¹²로 정의합니다. 이 접근 방식은 매우 강력하고¹⁰ 유연하며, 다음과 같은 7가지 핵심 구성 요소로 이루어집니다¹²:

1. 데이터 (**Data**): 시각화할 대상이 되는 데이터프레임. ggplot() 함수의 첫 번째 인자로 전달됩니다.¹⁴ 예: ggplot(data = health_data)
2. 미학 매핑 (**Aesthetic Mappings**): 데이터의 변수(예: age, sex)를 플롯의 시각적 속성(예: x축 위치, y축 위치, color, shape)에 *연결(mapping)*하는 작업입니다. 이는 항상 aes() 함수 내에서 정의됩니다.¹⁴ 예: aes(x = age, y = blood_pressure)
3. 지오메트리 (**Geometries**): 데이터를 표현할 시각적 객체(모양). geom_ 접두사를 가지며, 플롯에 레이어로 추가됩니다.¹² 예: geom_point() (산점도), geom_bar() (막대그래프), geom_boxplot() (박스플롯).²
4. 통계 변환 (**Statistical Transformations**): 원시 데이터¹⁶를 플로팅하기 전에 통계적으로 요약하는 내부 알고리즘(Stat)입니다. 예를 들어, geom_bar()는 기본적으로 stat = "count"를 사용하여 데이터의 빈도를 계산합니다.¹⁵ geom_smooth()는 stat = "smooth"를 사용하여 회귀선을 계산합니다.
5. 스케일 (**Scales**): 미학 매핑(예: color = sex)을 실제 시각적 값(예: sex가 "Male"이면 "blue", "Female"이면 "red")으로 변환하는 규칙입니다. 또한 축(axes)과 범례(legends)를 생성하여 이 매핑을 독자가 읽을 수 있게 합니다.¹³
6. 좌표계 (**Coordinate System**): 축을 구성하는 방식입니다. 기본값은 데카르트 좌표계(coord_cartesian())이며, coord_flip() (x/y축 전환), coord_polar() (풀라 좌표계) 등도 사용할 수 있습니다.¹⁵
7. 패싯 (**Faceting**): 데이터를 특정 범주형 변수의 하위 그룹으로 나누어, 여러 개의 작은 플롯(small-multiples)으로 시각화하는 강력한 기능입니다.¹² 예: facet_wrap(~ group), facet_grid(sex ~ race).¹⁷

이러한 "그래픽 문법"은 단순한 코딩 프레임워크가 아니라, 분석적 사고를 위한 인지적 프레임워크입니다.¹⁰ 문법은 무엇을 시각화할지(데이터와 미학 매핑¹⁵)와 어떻게 시각화할지(geom_...())¹⁴를 명확하게 분리합니다. 이러한 모듈성 덕분에, 분석가는 탐색적 데이터 분석(EDA)¹⁵ 과정에서 전체 데이터 구조를 재구성할 필요 없이, geom_point()를 geom_smooth()¹⁵로 신속하게 교체하며 가설을 테스트하고 데이터의 패턴을 탐색할 수 있습니다.

보건 데이터 실습: aes()의 이해

aes() 함수는 보건학 데이터의 다양한 변수 유형을 시각적 속성에 매핑하는 핵심입니다.¹⁶

- 연속형 변수 (**Continuous**): (예: age, incwage - 소득¹⁶, totmed - 총 의료비¹⁶, 혈압)
 - x, y: 축 상의 위치 (가장 일반적)
 - color (또는 fill): 연속적인 색상 그라데이션 (예: 저혈압은 파란색, 고혈압은 빨간색)
 - size: 점이나 선의 크기 (예: 중증도가 높을수록 큰 점)
 - alpha: 투명도
- 범주형 변수 (**Categorical**): (예: sex - 성별¹⁶, race - 인종¹⁶, smoker - 흡연 여부)
 - color (또는 fill): 개별적인 색상 (예: 남성은 파란색, 여성은 빨간색)
 - shape: 점의 모양 (예: 남성은 원, 여성은 삼각형)¹³
 - linetype: 선의 유형 (예: 남성은 실선, 여성은 점선)
 - alpha: 투명도

기술 통계(**Descriptive Statistics**) 시각화

그래픽 문법을 활용하여 보건학 연구에서 가장 기본적인 기술 통계량을 시각화할 수 있습니다.

- **geom_histogram()** (분포): 단일 연속형 변수(예: age)의 분포를 파악합니다.

R

```
# 예: 환자 나이(age)의 분포
ggplot(data = patient_data, mapping = aes(x = age)) +
  geom_histogram(binwidth = 5, fill = "blue", color = "black")
```

- **geom_bar()** (빈도): 단일 범주형 변수(예: race)의 빈도를 파악합니다.

R

```
# 예: 인종(race) 그룹별 환자 수
ggplot(data = patient_data, mapping = aes(x = race)) +
  geom_bar(fill = "grey")
```

- **geom_boxplot()** (그룹 간 비교): 범주형 변수(예: sex)에 따른 연속형 변수(예: totmed)의 분포(중앙값, 사분위수, 이상치)를 비교합니다.²

R

```
# 예: 성별(sex)에 따른 총 의료비(totmed) 지출 비교
ggplot(data = health_survey, mapping = aes(x = sex, y = totmed, fill = sex)) +
  geom_boxplot() +
  labs(title = "Total Medical Expenditure by Sex", x = "Sex", y = "Total Expenditure")
```

- **geom_point()** (관계): 두 연속형 변수(예: incwage, totmed) 간의 관계(상관관계)를 탐색합니다.²

R

```
# 예: 소득(incwage)과 의료비(totmed)의 관계
ggplot(data = health_survey, mapping = aes(x = incwage, y = totmed)) +
```

```
geom_point(alpha = 0.5) + # 점이 많을 경우 투명도 조절  
geom_smooth(method = "lm") # 선형 회귀 추세선 추가
```

총화(Stratification): ggplot2와 역학적 사고의 만남

ggplot2의 "레이어"¹² 개념은 역학(Epidemiology) 분석의 핵심 원리인

총화(Stratification)와 개념적으로 완벽하게 동일한 구조를 가집니다. 이는 ggplot2의 코드 구조가 역학자의 분석적 사고 과정과 동일한 형태(isomorphic)를 가짐을 의미하며, 대상 독자인 역학 전공 대학원생이 자신의 도메인 지식과 R 코드를 즉시 연결하는 데 도움을 F니다.

1. 1단계: 조(Crude) 분석

- 역학자는 먼저 두 변수(예: 노출, 결과) 간의 조(Crude) 관계를 봅니다.
- ggplot(data, aes(x = exposure, y = outcome)) + geom_point()

2. 2단계: 잠재적 교란 변수(Confounder) 확인 (미학 매핑)

- 역학자는 제3의 변수(예: age_group)가 이 관계에 영향을 미치는지(교란 또는 효과 수정) 확인하기 위해 총화합니다.
- ggplot(data, aes(x = exposure, y = outcome, color = age_group)) + geom_point()
- 이제 age_group 별로 관계가 다른지(즉, 색상별로 점의 패턴이 다른지) 시각적으로 즉시 확인할 수 있습니다.

3. 3단계: 총화된 분석 결과 제시 (패싯)

- 더 명확한 비교를 위해, 총화된 그룹을 시각적으로 완전히 분리하여 제시합니다.
- ggplot(data, aes(x = exposure, y = outcome)) + geom_point() + facet_wrap(~ age_group)¹⁷
- 이는 3차원 데이터를 2차원 평면에 효과적으로 시각화하는 방법이며¹⁷, 역학자가 총화 분석(stratified analysis)을 수행하는 방식과 정확히 일치합니다.

이처럼 ggplot2의 문법은 단순히 그림을 그리는 명령어가 아니라, 데이터를 탐색하고 가설을 검증하는 분석적 사고의 강력한 도구입니다.

III. 역학(Epidemiology) 데이터 시각화의 핵심 기법

ggplot2의 기본 문법을 마스터했다면, 이제 역학 분야의 특정 데이터 유형과 질문에 맞는 특화된 시각화 기법을 다룰 차례입니다. 이 섹션은 기술 역학(Descriptive Epidemiology)의 핵심 요소인 추세, 유행 곡선, 그리고 표준화 비율을 '올바르게' 시각화하는 데 중점을 둡니다.

3.1. 발병률(Incidence) 및 유병률(Prevalence) 추세 시각화

역학의 기본은 질병의 발생(발병률)과 상태(유병률)가 시간에 따라 어떻게 변하는지(추세) 기술하는 것입니다.⁹

- 시계열 추세: 시간에 따른 추세는 `geom_line()` (선 그래프)을 사용하여 시각화합니다. 만약 데이터 포인트가 불규칙하거나 장기적인 경향성을 보고 싶다면, `geom_smooth()`²⁰를 사용하여 통계적으로 평활화된(smoothed) 추세선을 추가할 수 있습니다.
- 누적 발병률(Cumulative Incidence): 시간에 따라 새로운 사례가 누적되는 양상은 y축에 `cumsum()` (누적 합계)를 적용한 변수를 매핑하여 `geom_line()` 또는 `geom_step()`으로 표현할 수 있습니다.¹⁹
- 불확실성(Uncertainty)의 표현: 추세선은 점 추정치(point estimate)일 뿐입니다. 특히 통계 모델에서 파생된 추세라면²⁰, 반드시 불확실성을 시각화해야 합니다. `geom_errorbar()`²⁰는 개별 시점의 불확실성을, `geom_ribbon()`²⁰은 연속적인 추세선의 95% 신뢰 구간(Confidence Interval)을 표현하는 데 매우 효과적입니다.

R

```
# 예: 연도별 발병률(rate)과 95% 신뢰구간(lower_ci, upper_ci) 시각화
ggplot(data = incidence_trends, aes(x = year, y = rate)) +
  geom_line(color = "blue", linewidth = 1) + # 중심 추정치
  geom_point(color = "blue") +
  geom_ribbon(aes(ymin = lower_ci, ymax = upper_ci), alpha = 0.2, fill = "blue") + # 신뢰 구간
  labs(title = "Annual Incidence Rate of Disease X (with 95% CI)",
       x = "Year",
       y = "Incidence Rate per 100,000")
```

3.2. 감염병 유행 곡선(Epidemic Curve) 작성: **incidence2** 패키지 활용

감염병 역학에서 유행 곡선(Epicurve)은 발병의 시간적 패턴, 규모, 잠재적 노출 시기, 그리고 전파 양식(예: 단일 노출, 지속 노출, 사람 간 전파)을 파악하는 데 사용되는 가장 기본적이고 핵심적인 역학 도구입니다.²¹

왜 ggplot2 대신 incidence2인가?

물론 ggplot2의 geom_bar()를 사용하여 유행 곡선을 그릴 수 있습니다. 하지만 이 방법은 사용자가 수동으로 날짜 데이터를 원하는 시간 간격(일, 주, 월)으로 집계(aggregation)하고, 비어있는 날짜를 채우는 등 복잡한 전처리 작업을 직접 수행해야 함을 의미합니다.²²

incidence2 패키지²³는 이러한 과정을 표준화하고 단순화하여, 역학자들이 "더 쉬운 명령어(easier commands)"²²로 라인리스트(linelist, 환자 개별 목록) 데이터로부터 직접 발병률(incidence) 객체를 계산하고²³ 이를 즉시 시각화할 수 있도록 설계되었습니다. (참고: incidence2는 기존 incidence 패키지²¹의 현대적 후속 버전으로, tidyverse와 호환성이 높은 tibble을 기반으로 합니다²³).

incidence2 패키지는 단순한 시각화 속도가 아닙니다. 이 패키지는 incidence라는 새로운 객체 클래스²³를 정의합니다. 이 객체는 tibble의 하위 클래스이면서, 역학적으로 '올바른' 데이터 구조를 갖추기 위한 "불변성(invariants)"(예: 날짜 인덱스 열 1개, 카운트 변수 열 1개, 그룹 열 0개 이상)을 강제합니다.²³ 감염병 유행²¹과 같은 혼란스럽고 긴급한 상황에서, incidence2 패키지는 분석 가가 날짜 집계 오류(예: 주차 계산²⁴)를 줄이고 역학적으로 타당한 데이터 구조를 갖추도록 강제합니다. 이는 분석 속도뿐만 아니라 정확성을 높이는 핵심 기능입니다.

incidence2 워크플로우

1. 데이터 준비: outbreaks 패키지²³ 등에서 제공하는 라인리스트²³를 로드합니다.
2. **incidence()** 함수: 라인리스트 데이터(x), 날짜 변수(date_index), 집계 간격(interval)을 지정하여 incidence2 객체를 생성합니다.
3. **plot()**: 생성된 incidence2 객체에 plot() 함수를 적용하여 기본 유행 곡선을 그립니다.
4. **facet_plot()**: 그룹 변수(예: 지역, 성별)에 따라 유행 곡선을 나누어(faceting) 플로팅합니다.²⁴

R

```
# outbreaks 패키지의 예제 데이터 사용
# install.packages(c("incidence2", "outbreaks"))
library(incidence2)
```

```

library(outbreaks)

# 1. 라인리스트 로드
linelist <- outbreaks::ebola_sim$linelist

# 2. incidence 객체 생성 (일별 발병률)
inc_daily <- incidence(linelist, date_index = "date_of_onset", interval = "day")

# 3. 기본 유행 곡선 플로팅
plot(inc_daily) +
  labs(title = "Ebola Outbreak Epicurve (Daily)", x = "Date", y = "Daily Cases")

# 4. 그룹별(성별) 유행 곡선 플로팅 (facet_plot)
inc_gender <- incidence(linelist,
                         date_index = "date_of_onset",
                         interval = "week", # 주별 집계
                         groups = "gender")

facet_plot(inc_gender, n_breaks = 6, angle = 45) + # 참조
  labs(title = "Ebola Outbreak Epicurve (Weekly by Gender)",
       x = "ISO Week", y = "Weekly Cases")

```

3.3. 연령 표준화 비율(Age-Standardized Rates)의 시각화

개념과 필요성:

26

역학에서 두 인구 집단 간의 질병률이나 사망률을 비교하는 것은 매우 흔한 작업입니다.²⁶ 하지만 이때 연령(Age)은 질병 발생의 가장 강력한 위험 요인이다²⁸, 비교하려는 두 집단 간의 *교란 변수(confounder)*로 작용할 수 있습니다.

예를 들어,²⁶ 캐나다 암 사망률 데이터는 이 함정을 명확하게 보여줍니다.

- 조(Crude) 사망률: 2000년 10만 명당 204.2명에서 2011년 211.0명으로 증가했습니다.²⁶
- 연령-특정(Age-Specific) 사망률: 0-39세 그룹(7.9 -> 5.8)과 40세 이상 그룹(450.4 -> 416.7) 모두 감소했습니다.²⁶

이 모순(심슨의 역설)은 2011년의 인구 구조가 2000년보다 더 고령화되었기 때문에 발생합니다. 즉, 연령 구조의 교란 효과를 제거하지 않으면 "암 사망률이 증가했다"는 완전히 잘못된 결론에 도달할 수 있습니다.²⁶

연령 표준화(Age-Standardization)는 표준 인구(예: 2000 U.S. Standard Population²⁹)를 기준으로 연령 구조의 차이를 보정하여³⁰, 두 집단 간의 순수한 위험도를 비교할 수 있게 하는 핵심 역학 기법입니다.²⁷

R을 이용한 계산 및 시각화

연령 표준화 비율(Age-Standardized Rate, ASR)의 시각화는 빙산의 일각과 같습니다. `plot()`³²이라는 간단한 코드로 생성된 시계열 그래프는 매우 명확해 보이지만, 그 그래프의 타당성(validity)은 전적으로 그래프 뒤에 숨겨진 복잡한 통계적 계산³²에 달려있습니다.

`surveil` 패키지는 이러한 '계산 후 시각화' 워크플로우를 잘 보여줍니다.³²

1. 모델링 (**stan_rw**): 연령-특정(age-specific) 비율의 시계열 추세를 모델링합니다.³²
2. 표준화 (**standardize**): 모델 결과를 표준 인구(standard_pop)를 이용해 연령 표준화 비율(SR)로 변환합니다.³²
3. 시각화 (**plot**): standardize로부터 반환된 객체를 `plot()` 함수로 직접 시각화합니다.³²

R

```
# install.packages("surveil")
library(surveil)
library(ggplot2)

# 1. 예제 데이터 및 표준 인구 로드
data(cancer, package = "surveil")
data(standard, package = "surveil")

# 50-69세 그룹으로 데이터 서브셋
cancer2 <- subset(cancer, grep("50-54|55-59|60-64|65-69", Age))

# 2. 연령-특정 비율 모델링
# (시간 절약을 위해 예제에서는 fit 객체를 직접 로드)
# fit <- stan_rw(cancer2, time = Year, group = Age,...)
```

```

data(fit, package = "surveil") # 미리 계산된 모델 로드

# 3. 연령-표준화 비율 계산
fit_sr <- standardize(fit, label = standard$age, standard_pop = standard$standard_pop)

# 4. 시각화
# plot()은 ggplot 객체를 반환하므로 labs() 등으로 수정 가능
plot(fit_sr, scale = 100e3, base_size = 10) +
  labs(title = "US age-standardized cancer incidence per 100,000",
       subtitle = "Ages 50-69",
       y = "ASR per 100,000")
#> Plotted rates are per 100,000.

```

이 챕터에서 대학원생에게 전달해야 할 가장 중요한 교훈은, `geom_line()`³⁴ 코드 자체가 아니라, ASR 플롯의 타당성이²⁶의 함정을 피하기 위한 사전의 통계적 표준화 계산³²에 전적으로 의존한다는 사실입니다.

IV. 공간 역학(Spatial Epidemiology) 데이터 시각화

공간 역학(Spatial Epidemiology)은 질병의 지리적 분포와 패턴을 연구하는 분야입니다. R은 강력한 GIS(지리 정보 시스템)³⁵ 기능으로 활용될 수 있으며, 이 섹션은 sf 패키지를 사용한 현대적인 공간 데이터 처리와 ggplot2 및 tmap을 이용한 지도 제작법을 다룹니다.³⁵

4.1. R 공간 데이터 처리의 혁명: sf (Simple Features)

과거 R에서 공간 데이터를 다루는 것은 `sp`³⁷, `rgeos`³⁷, `rgdal`⁸ 등 여러 패키지를 조합해야 하는 복잡한 작업이었습니다. `sf` (Simple Features)³⁷ 패키지는 이러한 구조를 통합하고 단순화한 현대적 표준입니다.

`sf`의 진정한 혁신은 지도를 그리는 기능이 아니라, 공간 데이터(GIS)와 비공간 데이터(역학 데이터 테이블)의 경계를 허물고 데이터 통합을 단순화한 데 있습니다. `sf` 패키지의 핵심 철학은 공간 데이터를 `data.frame` (또는 `tibble`)처럼 다루는 것입니다.³⁸

`sf` 객체는 세 가지 기본 구조로 구성됩니다³⁸:

1. **sfg (Simple Feature Geometry):** 개별 공간 객체(예: 환자 1명의 집 주소 POINT, 특정 지역구 경계 POLYGON).⁴¹

2. **sfc (Simple Feature Geometry List-Column):** sfg 객체들의 리스트. data.frame의 특별한 열(column)이 됩니다.³⁸
3. **sf (Simple Features):** 일반적인 속성 데이터(예: 지역구 이름, 인구 수)가 담긴 data.frame에 sfc 지오메트리 열이 결합된 최종 객체입니다.³⁸

이러한 sf 구조는 tidyverse와 완벽하게 호환됩니다.³⁹ 따라서 Shapefile⁴³이나 GeoJSON⁴³ 같은 공간 데이터를 sf::st_read() 함수로 R에 로드한 후⁴¹, dplyr::left_join()과 같은 tidyverse 함수를 사용하여 역학 데이터(예: 지역별 질병 발생 건수)를 sf 객체에 직접 결합할 수 있습니다.³⁵ 과거에는 ArcGIS³⁵와 같은 별도의 GIS 소프트웨어에서 공간 데이터를 처리하고, R에서 통계 데이터를 처리한 뒤, 두 결과를 수동으로 결합해야 했습니다. sf는 이 모든 과정을 하나의 재현 가능한 R 스크립트 내에서 통합적으로 수행할 수 있게 합니다.

4.2. 단계별 코로플레스 맵(Choropleth Map) 제작

코로플레스 맵(Choropleth Map)²은 행정 구역(Polygon)⁴²과 같은 지리적 영역을 특정 변수(예: 인구 10만 명당 질병 발생률)의 값에 따라 색상이나 음영으로 채우는⁴⁴ 가장 일반적인 공간 역학 시각화입니다.

제작 워크플로우는 다음과 같습니다³⁵:

1. 데이터 준비:
 - o sf::st_read(): 행정 구역 경계가 담긴 Shapefile 또는 GeoJSON 파일을 sf 객체(예: map_data)로 로드합니다.³⁵
 - o read.csv() 등: 지역 코드와 질병 발생률이 포함된 역학 데이터(예: epi_data)를 일반 data.frame으로 로드합니다.
2. 데이터 결합 (핵심): dplyr::left_join()을 사용하여 역학 데이터를 sf 객체에 결합합니다.

R

```
library(sf)
library(dplyr)
```

```
# 1. 데이터 로드 (가상 경로)
# map_data <- st_read("path/to/districts.shp")
# epi_data <- read.csv("path/to/disease_rates.csv")

# 2. 데이터 결합 (sf 객체와 data.frame 결합)
# combined_sf <- left_join(map_data, epi_data, by = "district_code")
```

3. 시각화 (옵션 1): ggplot2 + geom_sf()

ggplot2는 geom_sf()⁴⁵를 통해 sf 객체를 직접 지원합니다. aes(fill = ...)에 결합된 역학 변수를 매핑하여 간단하게 코로플레스 맵을 생성할 수 있습니다.³⁵

R

```
# 3. (옵션 1) ggplot2로 시각화
# ggplot(data = combined_sf) +
#   geom_sf(aes(fill = disease_rate)) +
#   scale_fill_viridis_c() + # 색상 팔레트 적용
#   theme_minimal() +
#   labs(title = "Choropleth Map of Disease Rate", fill = "Rate")
```

4. 시각화 (옵션 2): tmap 패키지

tmap 8은 ggplot2와 유사한 + 연산자 문법 40을 가진 지도 제작 전용 패키지입니다. tm_shape()으로 데이터를 지정하고 tm_fill() (채우기), tm_borders() (경계선) 48 등 더 직관적인 지도 제작 명령어를 제공합니다.³⁵

R

```
# 3. (옵션 2) tmap으로 시각화
# library(tmap)
# tmap_mode("plot") # 정적 지도 모드
#
# tm_shape(combined_sf) +
#   tm_fill("disease_rate", style = "quantile", palette = "Blues",
#           title = "Disease Rate") +
#   tm_borders(col = "grey50") +
#   tm_layout(main.title = "Choropleth Map of Disease Rate")
```

ggplot2 vs. tmap: 전략적 선택

ggplot2 + geom_sf() ⁴⁵와 tmap ⁴⁷은 모두 훌륭한 지도 제작 프레임워크를 제공합니다. 둘 중 하나를 선택하는 것은 '어느 것이 더 좋은가'의 문제가 아니라 전략적 선택의 문제입니다.

- 전략 1: 일관성 (ggplot2)

ggplot2의 강점은 일관성입니다. 논문이나 보고서의 다른 모든 비(非)지도 그래프(예: 막대 그래프, 산점도)와 동일한 테마(theming) 45와 글꼴을 적용할 수 있습니다. 이는 학술지 제출 시 요구되는 브랜딩 일관성(branding cohesion) 50을 맞추는데 매우 유리합니다.⁴⁵

- 전략 2: 전문성 (tmap)

tmap의 강점은 지도 제작 전문성입니다.⁴⁷ 축척 막대(scale bar), 나침반(compass rose), 지도 격자(grid) 등 "사소한 추가 기능(little extras)" 45을 한두 줄의 코드로 매우 쉽게 추가할 수 있습니다. 지도가 그 자체로 중심이 되는 독립적인 산출물(예: 공간 역학 논문의 Figure 1)이라면, tmap이 더 빠르고 전문적인 기능을 제공합니다.⁴⁷

V. 임상 통계(Clinical Statistics) 및 분석 역학 시각화

이 섹션은 보건학 연구, 특히 임상시험¹¹ 및 분석 역학⁵²에서 가장 빈번하게 사용되는 핵심 시각화 기법들을 다룹니다. 생존 분석, 메타 분석, 그리고 통계적 불확실성을 표현하는 오차 막대가 중심 주제입니다.

5.1. 생존 분석(Survival Analysis): ggsurvfit 패키지

임상 연구에서 "시간-사건(time-to-event)" 데이터를 분석하는 생존 분석은 필수적이며, 카플란-마이어(Kaplan-Meier) 생존 곡선⁵³은 그 표준적인 시각화 방법입니다.

전통적으로 R에서는 survival 패키지의 기본 `plot()` 함수를 사용했지만, 이는 ggplot2 생태계와 호환되지 않아 사용자 정의가 어렵고 출판 품질로 다듬기 까다로웠습니다. survminer 패키지⁵⁴가 이를 보완하려 했지만, 여전히 생존 곡선과 위험 테이블(risk table)의 정렬 문제라는 고질적인 기술적 한계를 안고 있었습니다.

ggsurvfit⁵⁶은 이러한 문제들을 ggplot2 문법⁵⁷ 내에서 완벽하게 해결하기 위해 설계된 최신 패키지입니다.⁵³ ggsurvfit의 혁신은 `add_risktable()`⁵⁸과 같은 함수를 진짜 ggplot2 지오메트리(geom)⁵⁶로 구현한 데 있습니다. 이는 생존 곡선과 위험 테이블을 두 개의 개별 플롯으로 간주하여 수동으로 조립하는 것이 아니라, 동일한 ggplot 객체 내의 다른 레이어로 취급합니다. 이 접근 방식은 위험 테이블을 x축(시간)은 공유하되 y축이 분리된 특수한 지오메트리로 처리하여, x축 스케일을 수정(예: `coord_cartesian`)하더라도 테이블이 자동으로 완벽하게 정렬되는⁵⁷ 근본적인 해결책을 제공합니다.

ggsurvfit 워크플로우

ggsurvfit의 워크플로우는 직관적이며 tidyverse 방식과 잘 어울립니다⁵⁹:

1. 모델 적합 (**survfit2**): survival 패키지의 `survfit`과 유사하지만 tidyverse와 호환되는 `survfit2()` 함수로 모델을 적합합니다.⁵⁸
2. 기본 플롯 (**ggsurvfit**): `survfit2` 객체를 파이프(|>)로 `ggsurvfit()`에 전달하여 기본 ggplot 객체를 생성합니다.⁵⁷
3. 레이어 추가 (**add_...**): ggplot2처럼 + 연산자를 사용하여 필요한 레이어를 추가합니다.
 - o + `add_confidence_interval()`: 신뢰 구간(리본 형태)을 추가합니다.⁵⁷
 - o + `add_risktable()`: 플롯 하단에 시간대별 위험 환자 수(Number at Risk) 테이블을

정렬하여 추가합니다.⁵⁸

R

```
# install.packages(c("ggsurvfit", "survival"))
library(ggsurvfit)
library(survival)

# 1. survival 패키지의 예제 데이터(lung) 사용 및 모델 적합
fit <- survfit2(Surv(time, status) ~ sex, data = lung)

# 2. & 3. ggsurvfit으로 플로팅 및 레이어 추가 [57, 60, 61]
fit |>
  ggsurvfit(linewidth = 1) +
  labs(
    x = "Days",
    y = "Overall survival probability",
    title = "Kaplan-Meier Survival Plot by Sex"
  ) +
  add_confidence_interval() +
  add_risktable(
    risktable_stats = "n.risk", # 'at-risk' 수만 표시
    title = "Number at Risk"
  ) +
  scale_ggsurvfit() # 축과 범례의 기본값 최적화 [60]
```

5.2. 메타 분석(Meta-Analysis): metafor 패키지

메타 분석은 여러 개별 연구⁶²의 결과를 통계적으로 결합하여 종합적인 근거를 도출하는 분석
역학의 핵심 방법론입니다. 포레스트 플롯(Forest Plot)⁶³은 이러한 메타 분석의 결과를
시각적으로 요약하는 표준 도구로, 개별 연구의 효과 크기(Effect Size)와 신뢰 구간, 그리고 전체
요약 추정치⁶³를 한눈에 보여줍니다.

R에서는 metafor⁶⁴ 패키지가 메타 분석을 위한 가장 포괄적이고 강력한 도구입니다.

metafor 워크플로우

metafor를 이용한 포레스트 플롯 생성은 '계산 -> 모델링 -> 시각화'의 3단계를 따릅니다.⁶³:

1. 효과 크기 계산 (**escalc**): 개별 연구 데이터(예: 2x2 테이블)로부터 연구별 효과 크기(예: Log Risk Ratio, Log Odds Ratio)와 그 분산(v_i)을 계산합니다.⁶³
2. 모델 적합 (**rma**): escalc로 계산된 효과 크기와 분산을 바탕으로 메타 분석 모델(예: 고정 효과 모델 또는 랜덤 효과 모델)을 적합합니다.⁶³
3. 시각화 (**forest**): 적합된 rma 객체를 forest() 함수에 전달하여 포레스트 플롯을 생성합니다.⁶³

R

```
# install.packages("metafor")
library(metafor)

# 1. 예제 데이터 로드 (BCG 백신 메타 분석)
# dat <- dat.bcg (metafor에 포함된 데이터)

# 2. 효과 크기 계산 (Log Risk Ratio)
dat <- escalc(measure = "RR", ai = tpos, bi = tneg, ci = cpos, di = cneg,
              data = dat.bcg, slab = paste(author, year, sep = ";"))

# 3. 모델 적합 (랜덤 효과 모델)
res <- rma(yi, vi, data = dat)

# 4. 포레스트 플롯 생성
forest(res,
       atransf = exp, # Log 스케일을 원래 RR 스케일로 변환
       at = log(c(0.05, 0.25, 1, 4)), # x축 눈금 지정
       header = "Author(s) and Year",
       xlab = "Risk Ratio (95% CI)")
```

5.3. 임상시험 및 모델 결과 시각화 (오차 막대)

임상시험 결과나 통계 모델의 계수를 시각화할 때, 평균⁶⁷과 같은 점 추정치(point estimate)는

반드시 그 불확실성²⁰과 함께 제시되어야 합니다.

geom_errorbar의 올바른 사용과 통계적 책임

ggplot2의 geom_errorbar()는 이러한 불확실성을 시각화하는 기본 지오메트리입니다.⁶⁸ 이 함수는 aes(ymin =..., ymax =...) 매핑을 요구합니다.⁶⁸

R

```
# (데이터 요약이 실행되어야 함)
# summary_data (df, mean, lower_ci, upper_ci)

# ggplot(summary_data, aes(x = group, y = mean)) +
#   geom_point() +
#   geom_errorbar(aes(ymin = lower_ci, ymax = upper_ci), width = 0.2) #
```

여기서 대학원생(대상 독자)에게 전달해야 할 매우 중요한 교육적 경고가 있습니다.

geom_errorbar() 함수 자체는 통계적 지식이 없습니다(statistically 'dumb'). 이 함수는 ymin과 ymax에 무엇이 들어오는지(표준 오차(SE)인지 95% 신뢰 구간(CI)인지) 관심이 없으며⁷⁰, 단지 주어진 값을 선으로 그릴 뿐입니다.

⁶⁸의 코드 예제는 SE와 CI를 그리는 코드가 aes(ymin = mean - value, ymax = mean + value) 형태로 거의 동일함을 보여줍니다. 하지만 통계적으로 SE와 CI는 완전히 다른 개념입니다.⁷¹

- 표준 오차 (**SE**): 평균의 정밀도를 나타냅니다. 샘플 크기가 커지면 0에 수렴합니다.⁷¹
- **95%** 신뢰 구간 (**CI**): 모평균에 대한 신뢰를 나타냅니다. "인구 기반 연구"⁷¹에서 모수(parameter)를 추정할 때 더 적절합니다.

R 코드를 실행하는 것⁷⁰은 쉽습니다. 하지만 그 코드의 aes() 안에 무엇을 넣을지⁷¹ 결정하는 것은 통계적 지식에 기반한 분석가의 책임입니다. 본 가이드는 독자가 이 통계적 책임을 명확히 인지하도록 경고해야 합니다.

임상시험 보고¹¹에 더 특화된 시각화가 필요한 경우, pharmaverse⁷² 생태계의 tern⁷²과 같은 전문 패키지 사용을 고려할 수 있습니다.

VI. 출판 및 발표를 위한 고급 시각화 전략

ggplot2로 생성한 기본 그래프를 학술지⁵⁰나 컨퍼런스 프레젠테이션⁷³에 바로 사용할 수 있는 "출판 품질(Publication-Ready)"¹⁰로 격상시키는 것은 모든 대학원생과 연구자의 핵심 과제입니다. 이 섹션은 R의 강력한 ggplot2 확장(extensions)⁸ 생태계를 활용하여 그래프를 세련되게 다듬는 고급 기법들을 다룹니다.

이러한 모듈성(Modularity)은 R 생태계의 가장 큰 강점입니다. 사용자는 ggplot2라는 하나의 '문법 엔진'(II장)만 마스터하면, '라벨 겹침'⁷⁶, '다중 패널 결합'⁷⁷, 'P-값 추가'⁷⁵와 같은 특정 고급 문제들을 해결하기 위해 필요한 확장 도구⁸를 레고 블록처럼 추가할 수 있습니다.

6.1. 가독성 향상: ggrepel을 이용한 텍스트 겹침 방지

산점도나 지도에서 특정 데이터 포인트에 라벨을 달 때, geom_text()⁷⁸를 사용하면 라벨이 서로 겹쳐⁸⁰ 가독성이 심각하게 저하되는 문제가 발생합니다.

ggrepel⁸ 패키지는 이 문제를 해결하기 위해 특별히 설계되었습니다. geom_text_repel()⁷⁶ 또는 geom_label_repel()⁷⁶ 함수는 geom_text()와 거의 동일하게 사용되지만, 라벨이 (1)서로 겹치지 않게, (2)데이터 포인트와 겹치지 않게, (3)플롯 가장자리에서 벗어나지 않게 자동으로 위치를 조정하여 밀어냅니다(repel).⁷⁶

R

```
# install.packages("ggrepel")
library(ggrepel)

# mtcars 데이터셋 예제
ggplot(mtcars, aes(x = wt, y = mpg, label = rownames(mtcars))) +
  geom_point(color = 'red') +
  geom_text_repel() # geom_text() 대신 사용
  theme_classic(base_size = 16)
```

6.2. 다중 플롯 조합: **patchwork** 패키지

논문은 Figure 1A, 1B, 1C와 같이 여러 패널로 구성된 복잡한 그림⁷⁷을 요구합니다. 과거에 R에서 `grid.arrange` 등으로 다중 패널을 정렬하는 것은 매우 어렵고 정렬이 틀어지기 쉬운 기술적 장벽이었습니다.

patchwork 패키지⁸는 `ggplot` 객체들을 수학 연산자처럼 직관적으로 조합할 수 있게 하여⁷⁷ 이 문제를 해결했습니다. **patchwork**의 등장은 연구자들이 다중 패널 그림을 구성하는 행동을 바꾸었습니다. 이제 병목 현상(bottleneck)은 '어떻게 정렬하는가?'(기술적 문제)가 아니라, '이 세 개의 플롯을 어떻게 배열해야 나의 주장을 가장 잘 전달할 수 있는가?'(인지적, 디자인적 문제)로 이동했습니다.

patchwork의 핵심 연산자는 다음과 같습니다⁷⁷:

- `p1 + p2`: 플롯을 나란히 배치 (기본).⁷⁷
- `p1 / p2`: 플롯을 수직으로 쌓음 (`p1` 위에 `p2`).⁸¹
- `p1 | p2`: 플롯을 수평으로 나란히 배치 (`p1` 옆에 `p2`).⁸²
- 복합 레이아웃: 괄호 ()를 사용하여 복잡한 중첩 레이아웃을 생성할 수 있습니다.⁷⁷

R

```
# install.packages("patchwork")
library(patchwork)

# p1, p2, p3가 미리 정의된 ggplot 객체라고 가정
p1 <- ggplot(mtcars, aes(mpg, disp)) + geom_point()
p2 <- ggplot(mtcars, aes(gear, disp, group = gear)) + geom_boxplot()
p3 <- ggplot(mtcars, aes(disp, qsec)) + geom_smooth()

# 예: (p1과 p2를 나란히 두고) / 그 아래에 p3를 배치
(p1 | p2) / p3 +
  plot_annotation(tag_levels = 'A') # A), B), C) 태그 자동 추가
```

6.3. 학술지 스타일 적용: **ggthemes** 및 **ggpubr**

ggplot2의 기본 회색 테마(theme_grey())는 데이터 탐색에는 유용하지만 출판용으로는 적합하지 않을 수 있습니다.

내장 테마 및 **ggthemes**

ggplot2 자체에 theme_bw()⁴⁹ (흰색 배경, 검은 격자), theme_classic()⁴⁹ (x/y 축만 있는 고전적 스타일), theme_minimal()⁴⁹ 등 깨끗한 내장 테마가 다수 포함되어 있습니다.⁸⁵

ggthemes⁸ 패키지는 한 걸음 더 나아가, 특정 출판물이나 스타일을 모방한 다양한 테마를 제공합니다⁸⁶:

- theme_economist(): 이코노미스트 스타일
- theme_wsj(): 월스트리트 저널 스타일
- theme_few(): 스티븐 퓨(Stephen Few)의 미니멀리즘 스타일⁸⁵
- theme_tufte(): 에드워드 터프티(Edward Tufte) 스타일
- scale_color_colorblind(): 색맹/색약 친화적 팔레트

R

```
# p1 + theme_economist() + scale_color_economist() #
```

ggsignif: P-값 및 유의수준 추가

ggsignif 패키지⁷⁵는 ggplot2의 "불투명한(opaque) 구문"⁸⁷을 보완하여, 비-프로그래머 연구자⁷⁵도 쉽게 출판용 그래프를 만들도록 돕는 래퍼(wrapper) 패키지입니다.

ggsignif의 가장 강력한 기능 중 하나는 stat_compare_means() 함수를 사용하여 ggplot 플롯(예: 박스플롯)에 자동으로 P-값과 유의수준 별표(\$*, \$**, \$***\$)를 추가하는 것입니다.⁷⁵

R

```

# install.packages("ggpubr")
library(ggpubr)

# 예: 그룹 간 비교 P-값 추가
ggboxplot(ToothGrowth, x = "dose", y = "len", fill = "dose") +
  stat_compare_means(
    comparisons = list(c("0.5", "1"), c("1", "2"), c("0.5", "2")), # 비교할 그룹 쌍
    label = "p.signif" # 유의수준 별표로 표시
  ) +
  stat_compare_means(label.y = 50) # 전체 그룹 간 비교 (ANOVA) P-값

```

ggpubr은 또한 theme_pubr() ⁷⁵이라는 자체 출판용 테마를 제공합니다.

VII. 대화형(Interactive) 시각화 및 공중 보건 대시보드

정적인(static) PDF나 이미지 파일 ⁸⁸을 넘어, 사용자와 상호작용하는 동적(interactive) 시각화 ²는 데이터를 탐색하고 ⁸⁸, 의사결정을 지원하는 ⁸⁹ 강력한 도구입니다. R은 plotly와 Shiny라는 두 가지 핵심 도구를 통해 이 영역을 선도하고 있습니다.

7.1. 정적 그래프의 동적 변환: plotly와 ggplotly()

plotly ⁹⁰ 패키지는 R에서 인터랙티브, 출판 품질의 그래프를 만드는 핵심 라이브러리입니다. 사용자가 플롯을 확대/축소하고, 패닝(panning)하며, 마우스를 올려 데이터의 실제 값⁸⁸을 툴팁(tooltip)으로 확인할 수 있게 합니다.

plotly의 가장 강력한 기능은 ggplotly() ⁹¹ 함수입니다. 이 함수는 II장에서 VI장까지 배운 ggplot2 문법으로 완벽하게 만들어진 정적 플롯 객체(p)를, ggplotly(p) ⁸⁸라는 단 한 줄의 코드로 가져와, 완전한 인터랙티브 웹 그래프로 변환합니다.

이는 R 생태계에서 가장 높은 투자 대비 수익(ROI)을 가진 함수 중 하나입니다. ggplot2 문법을 마스터한 분석가는 추가적인 학습 비용 거의 없이 ⁹¹, 자신의 정적 분석 결과물 ¹⁰을 강력한 동적 탐색 도구 ⁹⁰로 즉시 전환할 수 있습니다.

```

# install.packages("plotly")
library(plotly)

# 1. 표준 ggplot 객체 생성
p_static <- ggplot(mtcars, aes(x = wt, y = mpg, color = factor(cyl))) +
  geom_point() +
  labs(title = "Static ggplot")

# 2. ggplotly()로 변환
p_interactive <- ggplotly(p_static)

# RStudio 뷰어 또는 RMarkdown HTML에서 p_interactive를 실행하면
# 확대/축소/툴팁이 가능한 동적 그래프가 나타남
p_interactive

```

7.2. 보건의료를 위한 **Shiny** 대시보드 구축

plotly가 단일 그래프를 인터랙티브하게 만든다면, **Shiny**⁹⁴는 R 코드만으로(HTML/CSS/JS 지식 불필요) 강력한 대화형 웹 애플리케이션(대시보드)⁹³을 만들 수 있는 R 패키지 프레임워크입니다.

Shiny는 단순한 시각화 도구가 아니라, 통계 *모델을 운영화(Operationalize)*하는 프레임워크입니다. **Shiny**는 분석가가 자신의 R 모델(예: III장의 유행 곡선 모델, V장의 생존 분석 모델)을 비전문가인 최종 사용자(예: 병원 관리자, 공중 보건 공무원)가 실제로 사용할 수 있는 의사결정 지원 도구⁸⁹로 배포할 수 있게 해줍니다.

보건의료 분야의 **Shiny** 적용 사례는 다음과 같습니다⁹⁴:

- 공중 보건 감시: 질병 감시(Surveillance) 데이터 시각화 및 지리적 비교.⁹⁴
- 자원 계획: 병상, 인력 등 의료 자원(healthcare capacity) 계획 및 예측.⁹⁴
- 시뮬레이션: 백신 접종 클리닉 시뮬레이션 및 대기열 분석.⁹⁴
- 데이터 탐색: EHR(전자의무기록) 데이터 탐색 및 코호트 구성.⁹³

Shiny의 핵심 구조: UI와 Server

모든 Shiny 앱은 두 가지 핵심 구성 요소로 이루어집니다⁹⁵:

1. **UI (User Interface)**: fluidPage(), sidebarLayout(), selectInput() (드롭다운 메뉴), plotOutput() (플롯이 보일 공간) 등을 사용하여 사용자에게 보여지는 부분(레이아웃, 위젯)을 정의합니다.⁹⁷
2. **Server**: input\$ (UI로부터 입력받음)와 output\$ (UI로 출력보냄)을 사용하여 실제 R 연산이 일어나는 부분(백엔드 로직)을 정의합니다.⁹⁷

반응성(Reactivity)의 이해

Shiny의 핵심은 반응성입니다.⁹⁸

- 사용자가 UI에서 드롭다운 메뉴(예: input\$select_district⁹⁸)의 값을 변경합니다.
- Server는 이 변경을 자동으로 감지합니다.
- Server는 input\$select_district에 의존하는 모든 R 코드(예: 데이터 필터링, 플롯 재생성)를 자동으로 재실행합니다.
- 재실행된 결과물(예: output\$malaria_epicurve⁹⁸)이 UI로 다시 전송되어 화면이 실시간으로 업데이트됩니다.

Epihandbook의 예제⁹⁸는 사용자가 선택한 지역(select_district)과 연령대(select_agegroup)에 따라 말라리아 유행 곡선(malaria_epicurve)이 실시간으로 변경되는 반응형 대시보드를 구축하는 방법을 보여줍니다.

R

```
# app.R (단일 파일 Shiny 앱)
library(shiny)
# (데이터 로딩 및 플로팅 함수 'plot_epicurve'가 미리 정의되어 있다고 가정)

# 1. UI (User Interface) 정의
ui <- fluidPage(
  titlePanel("Malaria Epicurve Dashboard"), #

  sidebarLayout( #
    sidebarPanel(
      # 드롭다운 메뉴
      selectInput(inputId = "select_district",
                  label = "Select District:",
```

```

choices = c("All", "District A", "District B")),

selectInput(inputId = "select_agegroup",
           label = "Select Age Group:",
           choices = c("All ages" = "total", "<5" = "u5", ">=5" = "o5"))
),

mainPanel(
  # 플롯이 출력될 공간
  plotOutput(outputId = "malaria_epicurve") #
)
)
)

# 2. Server 로직 정의
server <- function(input, output) {

  # 반응형 플롯 생성
  output$malaria_epicurve <- renderPlot({

    # UI의 input$select_district와 input$select_agegroup 값에 따라
    # plot_epicurve 함수가 매번 새로 실행됨
    plot_epicurve(malaria_data,
                  district = input$select_district,
                  agegroup = input$select_agegroup)
  })
}

# 3. 앱 실행
shinyApp(ui = ui, server = server) # [95, 98]

```

Shiny는 leaflet⁷과 같은 다른 인터랙티브 시각화 패키지와 쉽게 결합되어,⁹⁶ 예시처럼 공간 역학 데이터를 실시간으로 탐색하는 강력한 지리정보 대시보드를 구축하는 데에도 사용될 수 있습니다.

VIII. 결론: 효과적인 보건 데이터 커뮤니케이션을 위한 제언

본 가이드는 보건학, 역학, 임상 통계 분야의 대학원생과 전문가가 R을 활용하여 데이터를 시각화하는 전 과정을 다루었습니다. 이 학습 여정은 R 시각화가 단순한 기술이 아니라, 연구의

투명성과 효과적인 커뮤니케이션을 위한 핵심 역량임을 강조합니다.

우리는 `ggplot2`의 기본 문법(II장)에서 시작하여, 이 문법을 역학(III장, IV장) 및 임상 통계(V장)라는 특정 도메인의 문제(유행 곡선, 공간 지도, 생존 분석)에 적용하는 방법을 배웠습니다. 나아가, 이 결과물을 `patchwork`와 `ggthemes` 등을 사용해 학술지 수준으로 세련되게 다듬고(VI장), 최종적으로 `plotly`와 `Shiny`를 통해 이 분석을 대화형 도구로 배포(VII장)하는, 데이터 과학자의 전체 학습 경로⁵²를 모델링했습니다.

이 과정을 통해 R을 사용한 시각화는 몇 가지 중요한 원칙을 따라야 함이 명확해졌습니다.

1. 목적에 맞는 도구 선택:

- 탐색(Exploration) vs. 설명(Explanation): 데이터의 패턴을 빠르게 파악하기 위한 탐색적 분석¹⁵을 위한 플롯(빠른 `ggplot` 또는 `ggplotly`)과, 최종 논문이나 보고서에 사용될 명확하고 세련된 설명용 플롯(VI장의 기법들)은 구분되어야 합니다.
- 정확성(Accuracy)의 함정: 시각적 아름다움이 통계적 정확성을 대체할 수 없습니다. 본 가이드에서 강조한 두 가지 핵심적인 함정, 즉 (1) 연령 구조를 보정하지 않은 조(Crude) 비율 시각화²⁶와 (2) 표준 오차(SE)와 신뢰 구간(CI)의 무분별한 혼용⁷¹은 반드시 피해야 합니다. R 코드는 분석가의 통계적 결정을 그대로 실행할 뿐이며, 그 책임은 전적으로 분석가에게 있습니다.
- 매체(Context) 고려: 최종 결과물이 정적인 논문(PDF)인지, 아니면 동적인 웹 대시보드인지에 따라 `patchwork` (VI장) 또는 `Shiny` (VII장)라는 적절한 전달 매체를 선택해야 합니다.

2. 재현성을 넘어선 투명한 데이터 스토리텔링:

R 스크립트 6는 분석의 재현성 5를 기술적으로 보장합니다. 하지만 훌륭한 보건 데이터 커뮤니케이션은 그 이상을 추구해야 합니다. `ggplot2`의 명확한 문법 10, `patchwork`의 논리적인 플롯 조합 77, 그리고 `Shiny`의 상호작용성 94은 모두 데이터를 이해하고 1, 연구자의 주장을 납득하며, 그 결론에 기반하여 공중 보건을 위한 행동 2을 취할 수 있도록 돕는 투명한 데이터 스토리텔링을 위한 도구입니다.

R은 보건학 연구자에게 강력한 힘을 부여합니다. 이 가이드가 독자들이 그 힘을 책임감 있고 효과적으로 사용하여, 데이터를 단순한 숫자에서 생명을 구하는 통찰력으로 변환하는 데 기여하기를 바랍니다.

Works cited

1. Spotlight on: R - Global Health Data Science, accessed November 18, 2025, <https://globalhealthdatascience.tghn.org/hub-resources/spotlight-r/>
2. Visual communication of public health data: a scoping review - PMC - NIH, accessed November 18, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC12060258/>
3. Descriptive statistics and data visualization | Epidemiology Class Notes - Fiveable, accessed November 18, 2025, <https://fiveable.me/epidemiology/unit-11/descriptive-statistics-data-visualization/study-guide/OQv6LseuDSfCjboA>

4. Analysis Using R Software: A Big Opportunity for Epidemiology and Public Health Data Analysis, accessed November 18, 2025,
<https://ejurnal.r2z-scientificpublishing.com/index.php/jhse/article/view/9>
5. Losing your fear of using R for statistical analysis - PMC - NIH, accessed November 18, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC10578917/>
6. Data Analysis and Visualization in R for Public Health, accessed November 18, 2025, <https://kerchner.github.io/r-public-health/aio.html>
7. Quick list of useful R packages - Posit Support, accessed November 18, 2025, <https://support.posit.co/hc/en-us/articles/201057987-Quick-list-of-useful-R-packages>
8. The best R packages for data visualization - The R Graph Gallery, accessed November 18, 2025, <https://r-graph-gallery.com/best-dataviz-packages.html>
9. Analysing and visualising population trends and spatial mapping - Coding Club, accessed November 18, 2025, <https://ourcodingclub.github.io/tutorials/trends-analysis-dataviz/>
10. 1 Introduction – ggplot2 - Elegant Graphics for Data Analysis, accessed November 18, 2025, <https://ggplot2-book.org/introduction.html>
11. Using R Programming for Clinical Trial Data Analysis - Quanticate, accessed November 18, 2025, <https://www.quanticate.com/blog/r-programming-in-clinical-trials>
12. Lesson 2: The Basics of GGplot2 - Data Visualization with R - Bioinformatics, accessed November 18, 2025, https://bioinformatics.ccr.cancer.gov/docs/data-visualization-with-r/Lesson2_intro_to_ggplot/
13. The Grammar – ggplot2: Elegant Graphics for Data Analysis (3e), accessed November 18, 2025, <https://ggplot2-book.org/mastery.html>
14. A Grammar of Graphics - Stat 20, accessed November 18, 2025, <https://stat20.berkeley.edu/fall-2024/2-summarizing-data/03-a-grammar-of-graphics/tutorial.html>
15. 3 Data visualisation | R for Data Science, accessed November 18, 2025, <https://r4ds.had.co.nz/data-visualisation.html>
16. Introduction to the Grammar of Graphics, accessed November 18, 2025, <https://murraylax.org/rtutorials/gog.html>
17. Visualizing Trends of Multivariate Data in R using ggplot2, accessed November 18, 2025, <https://towardsdatascience.com/visualizing-trends-of-multivariate-data-in-r-using-ggplot2-1b85409afcfa>
18. Visualizing Urban and Demographic Data in R with ggplot2 | Programming Historian, accessed November 18, 2025, <https://programminghistorian.org/en/lessons/urban-demographic-data-r-ggplot2>
19. ggplot2 - Plot for cumulative incidence count over time in R - Stack Overflow, accessed November 18, 2025, <https://stackoverflow.com/questions/67706250/plot-for-cumulative-incidence-count-over-time-in-r>
20. 5 Statistical summaries – ggplot2 - Elegant Graphics for Data Analysis, accessed

- November 18, 2025, <https://ggplot2-book.org/statistical-summaries.html>
- 21. Epidemic curves made easy using the R package incidence - PMC, accessed November 18, 2025, <https://PMC6509961/>
 - 22. 32 Epidemic curves - The Epidemiologist R Handbook, accessed November 18, 2025, https://epirhandbook.com/en/new_pages/epicurves.html
 - 23. An introduction to incidence2 - CRAN, accessed November 18, 2025, <https://cran.r-project.org/web/packages/incidence2/vignettes/incidence2.html>
 - 24. Introducing incidence2 - ReConverse, accessed November 18, 2025, <https://www.reconverse.org/posts/2021-07-13-introducing-incidence2/>
 - 25. incidence2 package - RDocumentation, accessed November 18, 2025, <https://www.rdocumentation.org/packages/incidence2/versions/2.3.0>
 - 26. Age-standardized Rates, accessed November 18, 2025, <https://www.statcan.gc.ca/en/dai/btd/asr>
 - 27. 21 Standardised rates - The Epidemiologist R Handbook, accessed November 18, 2025, https://www.epirhandbook.com/en/new_pages/standardization.html
 - 28. Chapter 7: Age standardization - IARC Publications, accessed November 18, 2025, https://publications.iarc.who.int/publications/media/download/3753/609d0d7711_047dd76d7f3dbaa25d7f041fc013e.pdf
 - 29. Age adjustment - Health, United States - CDC, accessed November 18, 2025, <https://www.cdc.gov/nchs/hus/sources-definitions/age-adjustment.htm>
 - 30. Reflection on modern methods: statistical, policy and ethical implications of using age-standardized health indicators to quantify inequities - PubMed Central, accessed November 18, 2025, <https://PMC8855998/>
 - 31. Easy Way to Learn Standardization : Direct and Indirect Methods - PMC - NIH, accessed November 18, 2025, <https://PMC3406211/>
 - 32. Age-standardized rates - CRAN, accessed November 18, 2025, <https://cran.r-project.org/web/packages/surveil/vignettes/age-standardization.html>
 - 33. 21 Standardised rates – The Epidemiologist R Handbook, accessed November 18, 2025, https://epirhandbook.com/en/new_pages/standardization.html
 - 34. Overlay bar chart in Age standardised mortality rate graph in ggplot in r - Stack Overflow, accessed November 18, 2025, <https://stackoverflow.com/questions/77538441/overlay-bar-chart-in-age-standardised-mortality-rate-graph-in-ggplot-in-r>
 - 35. 28 GIS basics – The Epidemiologist R Handbook, accessed November 18, 2025, https://epirhandbook.com/en/new_pages/gis.html
 - 36. Intro to Spatial Analysis & GIS for Spatial Epidemiology in R, accessed November 18, 2025, <https://cme.uchicago.edu/SpatialAnalysis22>
 - 37. Simple Features for R: Standardized Support for Spatial Vector Data - The R Journal, accessed November 18, 2025, <https://journal.r-project.org/articles/RJ-2018-009/>
 - 38. sf • SOGA-R • Department of Earth Sciences - Freie Universität Berlin, accessed November 18, 2025, <https://www.geo.fu-berlin.de/en/v/soga-r/Introduction-to-R/Spatial-Data/sf/index.html>

html

39. Introduction to Spatial Data - GEO 200CN: Quantitative Geography, accessed November 18, 2025, <https://geo200cn.github.io/introspatial.html>
40. Introduction to Geospatial Visualization with the tmap package - Rami Krispin, accessed November 18, 2025, <https://ramikrispin.github.io/2021/12/geospatial-visualization-with-the-tmap-package/>
41. 1. Simple Features for R • sf - r-spatial, accessed November 18, 2025, <https://r-spatial.github.io/sf/articles/sf1.html>
42. Spatial R: Using the sf package - UVA Library - The University of Virginia, accessed November 18, 2025, <https://library.virginia.edu/data/articles/spatial-r-using-sf-package>
43. R4SpatialAnalysis - Brown University, accessed November 18, 2025, <https://s4.ad.brown.edu/r4spatial/>
44. [Tutorial] Common Practices for Map Design - RPubs, accessed November 18, 2025, https://rpubs.com/djn34/creating_maps
45. Preferred mapping package(s) : r/rstats - Reddit, accessed November 18, 2025, https://www.reddit.com/r/rstats/comments/ecmy4q/preferred_mapping_packages/
46. Chapter 3 Making Maps in R | Data Analysis and Visualization with R: Spatial, accessed November 18, 2025, <http://www.geo.hunter.cuny.edu/~ssun/R-Spatial/mapping.html>
47. Plotting API Results: sf, tmap, and maptiles - CRAN, accessed November 18, 2025, <https://cran.r-project.org/web/packages/osdatahub/vignettes/plotting-results.html>
48. Plotting choropleth maps in R with tmap | CSO, accessed November 18, 2025, https://www.cso.ie/en/media/csoie/methods/otherresearch/sdgs/Plotting_choropleth_maps_in_R_with_tmap.pdf
49. 17 Themes – ggplot2 - Elegant Graphics for Data Analysis, accessed November 18, 2025, <https://ggplot2-book.org/themes.html>
50. Guidelines in creating publication-ready figures : r/bioinformatics - Reddit, accessed November 18, 2025, https://www.reddit.com/r/bioinformatics/comments/1ecw6x9/guidelines_in_creating_publicationready_figures/
51. Available CRAN Packages By Name - R Project, accessed November 18, 2025, https://cran.r-project.org/web/packages/available_packages_by_name.html
52. Statistical Analysis with R for Public Health Specialization - Coursera, accessed November 18, 2025, <https://www.coursera.org/specializations/statistical-analysis-r-public-health>
53. Visualizing Survival Data with the {ggsurvfit} R Package - YouTube, accessed November 18, 2025, https://www.youtube.com/watch?v=F-N_Pz2sPlg
54. Easy survival analysis in R - biostatsquid.com, accessed November 18, 2025, <https://biostatsquid.com/easy-survival-analysis-r-tutorial/>
55. Drawing Survival Curves using ggplot2 • survminer - R Packages, accessed November 18, 2025, <https://rpkgs.datanovia.com/survminer/index.html>

56. ggsurvfit.pdf - CRAN, accessed November 18, 2025,
<https://cran.r-project.org/web/packages/ggsurvfit/ggsurvfit.pdf>
57. Flexible Time-to-Event Figures • ggsurvfit - Daniel D. Sjoberg, accessed November 18, 2025, <https://www.danieldsjoberg.com/ggsurvfit/>
58. Visualizing Survival Data with the {ggsurvfit} R Package - Daniel D. Sjoberg, accessed November 18, 2025,
<https://www.danieldsjoberg.com/ggsurvfit-rmed-webinar-2024/slides/>
59. Flexible Time-to-Event Figures • ggsurvfit - Daniel D. Sjoberg, accessed November 18, 2025, <https://www.danieldsjoberg.com/ggsurvfit/index.html>
60. Plot Survival Probability — ggsurvfit - Daniel D. Sjoberg, accessed November 18, 2025, <https://www.danieldsjoberg.com/ggsurvfit/reference/ggsurvfit.html>
61. Survival analysis in R - Emily C. Zabor, accessed November 18, 2025,
<https://www.emilyzabor.com/survival-analysis-in-r.html>
62. Run Meta-Analysis, Create Forest Plot, Examine Influence in R with metafor (How To Series with SMD) - YouTube, accessed November 18, 2025,
https://www.youtube.com/watch?v=eMG8c_9Edmk
63. Forest Plot [The metafor Package], accessed November 18, 2025,
https://www.metafor-project.org/doku.php/plots:forest_plot
64. metafor: Meta-Analysis Package for R - CRAN, accessed November 18, 2025,
<https://cran.r-project.org/web/packages/metafor/metafor.pdf>
65. Conducting Meta-Analyses in R with the metafor Package | Journal of Statistical Software, accessed November 18, 2025, <https://www.jstatsoft.org/v36/i03/>
66. Forest Plots (Method for 'rma' Objects) — forest.rma • metafor, accessed November 18, 2025, <https://vviechtb.github.io/metafor/reference/forest.rma.html>
67. R - Graphs - Bar Charts with Error Bars in Ggplot2 - YouTube, accessed November 18, 2025, <https://www.youtube.com/watch?v=N4d-qUTTY44>
68. Plotting means and error bars (ggplot2) - Cookbook for R, accessed November 18, 2025,
[http://cookbook-r.com/Graphs/Plotting_means_and_error_bars_\(ggplot2\)/](http://cookbook-r.com/Graphs/Plotting_means_and_error_bars_(ggplot2)/)
69. Plotting means and error bars (ggplot2) - Cookbook for R, accessed November 18, 2025,
[http://www.cookbook-r.com/Graphs/Plotting_means_and_error_bars_\(ggplot2\)/](http://www.cookbook-r.com/Graphs/Plotting_means_and_error_bars_(ggplot2)/)
70. adding confidence intervals/error bars to ggplot - tidyverse - Posit Community, accessed November 18, 2025,
<https://forum.posit.co/t/adding-confidence-intervals-error-bars-to-ggplot/178122>
71. Adding Error Bars to ggplot2 Plots Can be Made Easy Through Dataframe Structure, accessed November 18, 2025,
<https://dethwench.com/adding-error-bars-to-plots-in-r/>
72. R packages by pharmaverse, accessed November 18, 2025,
<https://pharmaverse.r-universe.dev/packages>
73. ggplot2 themes for publication ready plots; including dark themes | by Koundinya Desiraju | Analytics Vidhya | Medium, accessed November 18, 2025,
<https://medium.com/analytics-vidhya/ggplot2-themes-for-publication-ready-plots-including-dark-themes-9cd65cc5a7e3>
74. R for Reproducible Scientific Analysis: Creating Publication-Quality Graphics with

- ggplot2, accessed November 18, 2025,
https://bham-carpentries.github.io/2019-02-18-bham_R-course-material/08-plot-ggplot2/index.html
75. ggpubr: Publication Ready Plots - Articles - STHDA, accessed November 18, 2025,
<https://www.sthda.com/english/articles/24-ggpubr-publication-ready-plots/>
76. ggrepel: An R package • ggrepel, accessed November 18, 2025,
<https://ggrepel.slowkow.com/>
77. The Composer of Plots • patchwork, accessed November 18, 2025,
<https://patchwork.data-imaginist.com/>
78. ggplot2 texts : Add text annotations to a graph in R software - Easy Guides - Wiki - STHDA, accessed November 18, 2025,
https://www.sthda.com/english/wiki/wiki.php?id_contents=7991
79. 8 Annotations – ggplot2 - Elegant Graphics for Data Analysis, accessed November 18, 2025, <https://ggplot2-book.org/annotations.html>
80. Examples - ggrepel - Kamil Slowikowski, accessed November 18, 2025,
<https://ggrepel.slowkow.com/articles/examples>
81. Combine Multiple Plots with patchwork - The R Graph Gallery, accessed November 18, 2025, <https://r-graph-gallery.com/package/patchwork.html>
82. Plot Assembly - patchwork, accessed November 18, 2025,
<https://patchwork.data-imaginist.com/articles/guides/assembly.html>
83. Multiple plots in ggplot2 with patchwork, accessed November 18, 2025,
<https://gotelliLab.github.io/GotelliLabMeetingHacks/NickGotelli/ggplotPatchwork.html>
84. Complete themes — ggtheme - ggplot2, accessed November 18, 2025,
<https://ggplot2.tidyverse.org/reference/ggtheme.html>
85. What are some of your favorite ggplot themes? : r/rstats - Reddit, accessed November 18, 2025,
https://www.reddit.com/r/rstats/comments/1fhg0u8/what_are_some_of_your_favorite_ggplot_themes/
86. Extra Themes, Scales and Geoms for ggplot2 • ggthemes, accessed November 18, 2025, <https://jrnold.github.io/ggthemes/>
87. ggpubr: 'ggplot2' Based Publication Ready Plots - R Packages, accessed November 18, 2025, <https://rpkgs.datanovia.com/ggpubr/>
88. Chapter 13 Interactive Graphs | Modern Data Visualization with R - GitHub Pages, accessed November 18, 2025, <https://rkabacoff.github.io/datavis/Interactive.html>
89. Interactive Visualization Applications in Population Health and Health Services Research: Systematic Scoping Review - NIH, accessed November 18, 2025,
<https://pmc.ncbi.nlm.nih.gov/articles/PMC8900899/>
90. Plotly r graphing library in R, accessed November 18, 2025, <https://plotly.com/r/>
91. 39 Interactive plots - The Epidemiologist R Handbook, accessed November 18, 2025, https://www.epirhandbook.com/en/new_pages/interactive_plots.html
92. 39 Interactive plots – The Epidemiologist R Handbook, accessed November 18, 2025, https://epirhandbook.com/en/new_pages/interactive_plots.html
93. VisualSphere: a Web-based Interactive Visualization System for Clinical Research Data - PMC - NIH, accessed November 18, 2025,

<https://pmc.ncbi.nlm.nih.gov/articles/PMC11141841/>

94. Using Shiny in Healthcare: Examples from the 2021 Shiny Contest - Posit, accessed November 18, 2025, <https://posit.co/blog/using-shiny-in-healthcare/>
95. Chapter 13 Introduction to Shiny | Geospatial Health Data - Paula Moraga, accessed November 18, 2025, <https://www.paulamoraga.com/book-geospatial/sec-shiny.html>
96. Spatial health data with R Shiny and custom reporting feature - YouTube, accessed November 18, 2025, https://www.youtube.com/watch?v=tZCOlivV_Us
97. Tutorial: implementing and visualizing machine learning (ML) clinical prediction models into web-accessible calculators using Shiny R - PMC - NIH, accessed November 18, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC9843315/>
98. 43 Dashboards with Shiny – The Epidemiologist R Handbook, accessed November 18, 2025, https://epirhandbook.com/en/new_pages/shiny_basics.html