

---

# SRC-STAT의 시각적 분석 도구 디자인과 개발



## Design and Development of Visual Analytic Tools in SRC-STAT



한구현, GuHyun Han\*, 김영호, Young-Ho Kim\*\*, 신동화, DongHwa Shin\*\*\*,  
이영조, Young Jo Lee\*\*\*\*, 서진욱, Jinwook Seo\*\*\*\*\*



**요약** SRC-STAT은 서울대학교 데이터과학과 지식창출 연구센터에서 개발한 통계패키지로 오픈소스 통계패키지인 R의 엔진을 이용한다. R은 통계 계산, 그래프 표현 기능 모두 강력하지만 차트를 정적인 이미지로 출력하므로 사용자가 직관적인 조작으로 데이터를 탐색하고 분석하기 어렵다. 이 논문에서는 위의 문제를 보완하기 위해 SRC-STAT에 포함된 세 종류의 시각적 분석 기능(산점도 행렬 탐색기, 계층적 군집분석 탐색기, 그래프 시각화 탐색기)을 소개하고 그 설계와 구현에 대해 다룬다. 통계학자들과 협업하였고 반복적 디자인 방법을 따라 사용성을 개선하였다.



**Abstract** SRC-STAT is a statistical package program developed by Data science for knowledge creation research center in Seoul National University. It internally uses an engine of R, an open-source statistical package providing powerful methods for statistical calculations and graph plotting. However, a plot made by R is just a static image so that user cannot explore and analyze data with intuitive interactions. In this paper, we cover the design and implementation of three visual analytic functions in SRC-STAT to relieve the problem. We've co-worked with statisticians in SNU and applied an iterative design process to reflect their feedbacks.



**핵심어:** *Data Visualization: 데이터 시각화, Visual Analytics: 시각적 분석, Iterative Design Process: 반복적 디자인, SRC-STAT, R*

---

본 논문은 2013년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2011-0030813)

\*주저자 : 서울대학교 컴퓨터공학부 석사과정 e-mail: ghhan@hcil.snu.ac.kr

\*\*공동저자 : 서울대학교 컴퓨터공학부 석사과정 e-mail: yhkim@hcil.snu.ac.kr

\*\*\*공동저자 : 서울대학교 컴퓨터공학부 석사과정 e-mail: dhshin@hcil.snu.ac.kr

\*\*\*\*책임저자 : 서울대학교 통계학과 교수 e-mail: youngjo@snu.ac.kr

\*\*\*\*\*교신저자 : 서울대학교 컴퓨터공학부 부교수; e-mail: jseo@snu.ac.kr

## 1. 서론

서울대학교 데이터과학과 지식창출 연구센터에서는 통계 패키지 프로그램 SRC-STAT[1] 개발 프로젝트를 진행 중이며 그 베타 버전을 배포 중이다. SRC-STAT은 내부에 오픈 소스 통계패키지 R의 엔진을 사용하며, 이를 SPSS와 유사한 GUI로 감싸서 통계처리, 분석에 익숙하지 않은 사용자의 편의를 고려하도록 설계되었다.

R은 통계 계산, 그래프 표현 양쪽 모두 강력한 기능을 제공하지만 직관적이지 않은 사용 방법 때문에 진입장벽이 높다는 단점이 있다. 이런 단점을 보완하기 위해 R의 GUI를 제공하려는 시도들이 있었다. RStudio[2] 등이 대표적인 사례이다. 그러나 위의 프로젝트들을 이용한다 해도 R을 사용하여 데이터를 시각적으로 탐색하고 분석하기는 어렵다. R의 그래프 표현 기능은 결과물을 정적인 이미지 파일로 출력하기 때문에 사용자와의 상호작용에 제한이 많기 때문이다.

이를 보완하기 위해 SRC-STAT은 R엔진과 독립적인 세 종류의 시각적 분석 기능을 포함하고 있다. 산점도 행렬을 활용하여 데이터들 간의 상관관계를 탐색하는 산점도 행렬 탐색기, 계층적 군집분석 기능을 사용하여 데이터 분류방법을 탐색하는 계층적 군집분석 탐색기, 그리고 기본 차트들을 연결하여 데이터의 특성을 탐색하는 그래프 시각화 탐색기이다. 위의 기능들은 사용자로 하여금 직관적인 GUI 조작으로 데이터를 탐색할 수 있도록 보조한다.

이 논문에서는 위의 시각적 분석 기능의 디자인과 구현 과정에 대해 다룬다. 이 연구는 통계학자들과의 긴밀한 협업을 통해 진행되었다. 먼저 회의를 통해 베타버전에 반드시 필요한 시각적 분석기능들을 확정하였고, 반복적 디자인 방법을 채택하여 UI를 점진적으로 개선하는 과정에서 통계학자들의 피드백을 반영하도록 애썼다.

## 2. 시각적 분석기능 디자인과 구현

### 2.1 산점도 행렬 탐색기

#### 2.1.1 선정 배경

‘미지의 자료를 처음 받았을 때 가장 먼저 하는 행동이 무엇인가’를 물었을 때 통계학자들은 대부분 산점도(Scatter plot)를 그려본다고 대답했다. 구체적으로, 데이터 시트를 훑어보며 관련성이 있어 보이는 변수들부터 산점도를 하나씩 확인하며 데이터의 특성을 찾겠다고 했다. 데이터 테이블 내의 변수의 수가 많을수록 확인해야하는 변수 쌍의 수도 증가하므로 시간이 많이 소요되고 순서쌍을 놓칠 가능성도 높아지는 단점이 있었다.

#### 2.1.2 디자인

산점도 행렬 형태의 탐색메뉴를 제공하여 산점도를 그릴

수 있는 모든 변수 순서쌍을 한 눈에 확인할 수 있도록 하였다. 행렬의 각 엔트리에는 두 변수의 상관계수 값을 계산하여 기울기 형태로 시각화하였다. 산점도 행렬을 탐색하면서 사용자가 관심을 가진 변수 쌍들의 상세정보를 확인할 수 있도록 하였다.

#### 2.1.3 구현

산점도 행렬과 상관계수 시각화로 변수 간의 관계를 확인할 수 있다. 탐색 메뉴는 공간 활용을 위해 숨김 가능하도록 하였다. 사용자가 선택한 변수 쌍은 산점도 행렬 내에서 하이라이트 되며 새로운 탭으로 추가된다. 탭 내에서는 데이터의 범위 별 값의 분포를 확인할 수 있고, 특정 데이터 범위 내의 산점도를 확인할 수 있다.

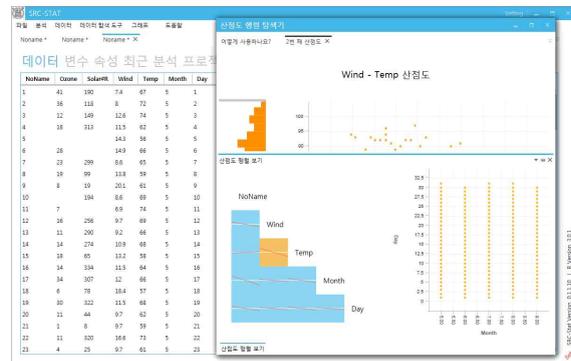


그림 1. SRC-STAT의 산점도 행렬 탐색기

## 2.2 계층적 군집분석 탐색기

### 2.2.1 선정 배경

R 내에도 여러 종류의 군집분석용 라이브러리가 존재하지만 인터랙션이 불가능하므로 직관적인 조작이 불가능하다는 단점이 있었다. 군집 분석의 경우 처음부터 어떤 결과가 나올지 알 수 없는 상태에서 유사성의 수준을 변경해 가면서 군집분석을 다양하게 수행하며 데이터들이 갖고 있는 특성을 확인하고자 하는 욕구가 있었다.

### 2.2.2 디자인

자주 사용되는 군집분석 방법 중 하나인 계층적 군집분석 방법을 사용하기로 했다. HCE(Hierarchical Clustering Explorer)[3]의 시각화 방법과 상호작용 방법을 참고하여 유사도 수준을 Top-down level과 Bottom-up level 양쪽에서 변경할 수 있도록 설계하였다.

### 2.2.3 구현

데이터를 군집분석하여 군집관계를 덴드로그램(Dendrogram)으로, 데이터필드의 값은 히트맵(Heatmap)으로 표현하였다. 두 개의 바를 조작하여 군집분석의 최소 유사도와 세부표현도를 조정할 수 있다. 두 값을 바꾸면 그에 따라 히트맵 또한 변화하므로 사용자는 직관적인 조작으로 군집분석의 변화 양상을 파악할 수 있다. 히트맵에서 관심있

는 부분에 mouse-over하면 데이터의 상세한 정보를 우측 창에서 확인할 수 있다.

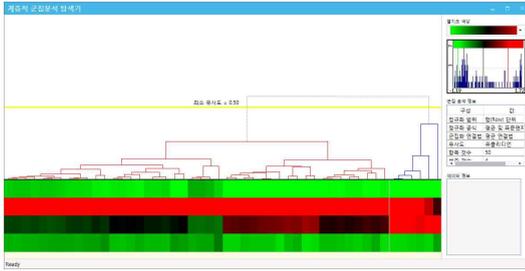


그림 2. SRC-STAT의 계층적 군집분석 탐색기

### 2.3. 그래프 시각화 탐색기

#### 2.3.1 선정배경

같은 데이터를 다른 종류의 차트를 통해 다양한 관점에서 파악할 때 데이터가 가진 특성을 더 잘 파악할 수 있다. 현재 R을 사용하여 만들 수 있는 차트들은 단순한 이미지이므로 이러한 탐색방법을 보조하는 데에 한계가 있다. 따라서 이런 기능을 보조하도록 기본적인 차트(산점도, 꺾은선 그래프, 원 그래프, 막대 그래프, 줄기 잎 그래프, 상자 그림, 히스토그램)를 새로 디자인하고 구현하기로 했다.

#### 2.3.2 디자인

Brushing and Linking을 이용하여 한 차트의 특정한 데이터를 선택하였을 때 다른 차트들도 관련되어 있는 데이터를 하이라이트 하도록 설계하였다. 또한, 여러 종류의 차트를 정돈된 형태로 배치하고 쉽게 확인할 수 있도록 통일된 UI를 제공하기로 했다.

#### 2.3.3. 구현



그림 3. SRC-STAT의 그래프 시각화 탐색기

Avalondock[4] 라이브러리를 사용하여 생성된 차트의 배치를 수평 혹은 수직 방향으로 쉽게 쌓고 떼어낼 수 있도록 설계하였다. Brushing and Linking 기능이 구현되어 있고, 차트의 각 부분에 mouse-over하면 상세 값을 확인할 수 있다. 생성된 차트를 이미지로 저장하는 기능이 필요하다는 의견을 반영하여 마우스 우 클릭 시 차트 이미지를 클립보드에 저장하도록 구현하였다.



### 3. 결론

이 연구에서는 SRC-STAT의 베타버전에 포함된 세 종류의 시각적 탐색기능의 선정배경과 디자인, 그리고 구현에 대해 다루었다. 홈페이지(<http://srcdsc.snu.ac.kr/srcstat/>)에서 다운로드하여 기능들을 체험해볼 수 있다.

향후 개선해야할 점들은 다음과 같다. 먼저, 많은 수의 변수가 포함된 데이터의 경우 조합의 수가 너무 많아져서 산점도 행렬의 각각의 엔트리를 확인할 수 없게 된다. 둘째, 그래프 시각화 탐색기의 Brushing and Linking 기능 중 Brushing 부분에 해당하는 상호작용 방법을 추가할 필요가 있다. 마지막으로 대규모의 데이터를 입력받을 때 처리시간을 단축할 수 있도록 차트 렌더 알고리즘을 최적화할 필요가 있다.

SRC-STAT은 베타버전 릴리즈 후 통계학 기초 수업 등에서 사용되고 있으며, 그로부터 다양한 종류의 피드백을 받을 수 있으리라 기대된다. 시각화 분석기능을 직접 사용한 다양한 사용자들의 경험을 반영한다면 좀 더 사용자에게 친숙한 소프트웨어로 발전시킬 수 있을 것이다.



### 참고문헌

- [1] SRC-STAT([srcdsc.snu.ac.kr/srcstat/](http://srcdsc.snu.ac.kr/srcstat/))
- [2] Rstudio IDE([www.rstudio.com](http://www.rstudio.com))
- [3] Jinwook Seo, Ben Shneiderman, "Interactively Exploring Hierarchical Clustering Results," Computer, vol. 35, no. 7, pp. 80-86, July 2002.
- [4] Project Avalondock([avalondock.codeplex.com](http://avalondock.codeplex.com))