

A Study on the Development of a Problem Bank in an Automated Assessment Module for Data Visualization Based on Public Data

HakNeung Go*, Sangsu Jeong**, Youngjun Lee***

*Teacher, Songjeong Jungang Elementary School, Gwangju, Korea

**Teacher, Geyonggi Science High School for the Gifted, Suwon, Korea

***Professor, Dept. of Computer Education, Korea National University of Education, Cheongju, Korea

[Abstract]

Utilizing programming languages for data visualization can enhance the efficiency and effectiveness in handling data volume, processing time, and flexibility. However, practice is required to become proficient in programming. Therefore public data-based the problem bank was developed to practice data visualization in a programming automatic assessment system. Public data were collected based on topics suggested in the curriculum and were preprocessed to make it suitable for users to visualize. The problem bank was associated with the mathematics curriculum to learn various data visualization methods. The developed problems were reviewed to expert and pilot testing, which validated the level of the questions and the potential of integrating data visualization in math education. However, feedback indicated a lack of student interest in the topics, leading us to develop additional questions using student-center data. The developed problem bank is expected to be used when students who have learned Python in primary school information gifted or middle school or higher learn data visualization.

▶ **Key words:** Data Visualization, Python, Online Judge, Graph, Public data

[요 약]

프로그래밍 언어를 활용한 데이터 시각화는 처리하는 데이터 양, 처리 시간, 유연성에서 효율성과 효과성을 향상시킬 수 있으나 프로그래밍에 익숙해지기 위해 연습이 필요하다. 이에 본 연구에서는 프로그래밍 자동 평가 시스템에서 데이터 시각화를 연습하기 위한 공공데이터 기반 문제 은행을 개발하였다. 공공데이터는 교육과정에서 제시한 주제로 수집하였으며 학습자가 데이터 시각화하기에 적절한 형태로 가공하였다. 문제는 다양한 데이터 시각화 방법을 학습하기 위해 수학 교육과정과 연계하여 개발하였다. 개발한 문제는 전문가 검토 및 파일럿 테스트를 실시하였으며 문항의 수준, 데이터 시각화를 통한 수학 교육의 가능성을 확인하였다. 하지만 학생에게 흥미가 떨어지는 주제라는 의견을 받았으며 이를 보완하기 위해 학생이 중심이 되는 데이터를 활용하여 추가로 문항을 개발하였다. 개발한 문제 은행은 초등학교 정보영재 또는 중학교 이상에서 파이썬을 학습한 경험이 있는 학생이 데이터 시각화를 배울 때 활용될 수 있을 것으로 기대된다.

▶ **주제어:** 데이터 시각화, 파이썬, 온라인 저지, 그래프, 공공 데이터

- First Author: HakNeung Go, Corresponding Author: Youngjun Lee
- *HakNeung Go (snddl3@outlook.kr), Songjeong Jungang Elementary School
- **Sangsu Jeong (whiteedu84@gmail.com), Geyonggi Science High School for the Gifted
- ***Youngjun Lee (yjlee@knue.ac.kr), Dept. of Computer Education, Korea National University of Education
- Received: 2024. 01. 25, Revised: 2024. 04. 18, Accepted: 2024. 04. 18.

I. Introduction

정보 기반 사회에서 데이터는 의사 결정, 마케팅, 연구 개발 등 다양한 분야에서 활용되고 있다. 특히 정보 기술의 발달과 데이터의 디지털화로 대용량의 데이터를 수집할 수 있게 되었고 데이터를 처리하고 분석할 수 있는 기술도 급격히 발전함에 따라 데이터의 가치가 점점 증가하고 있다[1][2]. 특히, 수집한 데이터를 이용하여 분석한 결과를 전달하는 것은 중요하며, 데이터 시각화는 분석된 결과를 그래프, 이미지 등으로 나타내어 핵심 내용을 탐색하는데 효과적인 방법이다[6].

데이터 시각화는 자료에 내포된 양적 정보를 쉽게 이해할 수 있도록 시각적으로 표현하고 전달되는 과정으로 그래프를 통해 정보를 명확하고 정보를 전달하는 것을 목적으로 한다[5]. 데이터 시각화는 패턴, 추세 및 관계를 식별할 수 있게 해주며 이상치, 결측치 등 데이터의 문제점을 발견하여 직관적이며 효과적으로 정보를 전달할 수 있다. 많은 양의 데이터를 시각화하기 위해서는 공학도구가 필요하며 사용법이 간단한 상용 프로그램과 코드를 작성하여 사용하는 프로그래밍 언어가 있다.

엑셀, 구글 스프레드시트와 같은 상용 프로그램을 활용한 데이터 시각화는 사용법이 간단하여 쉽게 시각화를 할 수 있다. 하지만 분석에 사용할 수 있는 데이터 양에 제한이 있고 프로그램에서 제공하는 형태로 시각화를 할 수 있다. 반면 프로그래밍 언어를 활용한 시각화는 코드를 작성해야 하므로 사용법이 어렵다. 하지만 대규모 데이터를 처리할 수 있으며 사용자 필요에 맞게 그래프를 작성할 수 있는 유연성이 있다[7][15]. 프로그래밍 언어를 활용한 데이터 시각화는 시각화 과정에서 프로그래밍 언어를 사용하므로 프로그래밍 언어를 학습할 수 있고 이는 프로그래밍과 데이터 시각화가 중요한 시대에 필요한 교육이라고 할 수 있다[8].

프로그래밍 언어를 활용한 데이터 시각화 교육으로 파이썬을 활용한 연구가 있다. 대부분의 연구에서 데이터 분석 전 과정인 데이터 수집, 데이터 전처리, 데이터 시각화, 평가의 과정을 학습한다. 데이터 분석에 활용한 데이터는 공공데이터 포털이나 설문 조사를 통해 수집하며 데이터 시각화 학습 방법은 예제 코드를 따라하면서 방법을 익히고 프로젝트에서 데이터 시각화를 활용한다[8-12]. 데이터 분석 전 과정을 다루는 연구는 주를 이루고 있지만 데이터 시각화에 초점을 맞춘 연구는 아직이다.

프로그래밍 언어를 활용한 데이터 시각화는 프로그래밍을 연습할 기회가 필요하다. 이를 지원하기 위한 학습 도구로 문제를 해결하고 실시간으로 평가하고 피드백을 해주는 주피터 노트북 기반 자동 평가 모듈이 개발되어 있으나 문항이 개발되어 있지 않다[13].

이에 본 연구에서는 주피터 노트북 기반 데이터 시각화 자동 평가 모듈에서 활용하기 위한 문항을 개발하고자 한다. 특히 실제 데이터를 가지고 학습할 수 있도록 공공데이터를 활용하며 데이터 시각화 학습 요소는 수학 교육과정의 자료와 가능성 영역에서 제시한 내용 체계와 연계하고자 한다.

II. Preliminaries

1. Data Visualization

통계 영역에서 그래프는 어떠한 표현 목적을 가지고 표현하려는 바가 잘 드러나도록 간결하게 만든 시각적 도구로 데이터 시각화는 데이터를 쉽게 이해할 수 있는 형태로 변환하는 과정을 의미한다. 데이터는 '시각화'의 도움으로 정보로서의 가치를 가지고, 정보 또한 시각화의 과정을 통해 누군가의 지식으로서의 의사 결정이나 지적 발전의 역할을 수행한다[4][14].

데이터 시각화가 효과적인 이유는 인간이 시각적 패턴을 인지하는 능력이 발달하여 글이나 말과 같은 수단보다 그래프를 활용할 때 문제 상황을 보다 쉽게 이해하고 의사 결정을 내리고 결과를 분석할 때 효과적이다[2]. 또한 데이터 시각화는 작은 공간에 다량의 정보를 효율적으로 제시할 수 있다[3].

데이터를 시각화하는 방법으로 스프레드 시트와 같은 상용 프로그램을 사용하는 방법과 파이썬, R과 같은 프로그래밍 언어를 사용하는 방법이 있다. 상용 프로그램은 쉽게 사용할 수 있는 장점이 있다. 하지만 데이터를 시각화하여 표현하는데 있어 유연성이 제한되며 대용량의 데이터를 처리하는데 한계가 있고 일부 프로그램은 사용하는 데 비용을 지불해야 한다. 반면 프로그래밍 언어를 사용하는 방법은 프로그래밍 언어를 익혀야 하는 어려움이 있다. 하지만 시각화하는데 유연성이 뛰어나며 대용량의 데이터를 빠르게 처리할 수 있다. 또한 프로그래밍 언어 및 라이브러리는 대부분 무료이며, 몇 줄의 코드로 데이터를 시각화할 수 있고 이를 활용해 쉽게 재사용할 수 있으며 웹 등 다른 프로그램에서 활용할 수 있다[7][15].

2. Data Visualization education using programming language

프로그래밍 언어를 활용한 데이터 시각화 교육은 파이썬을 활용하여 교육 프로그램을 개발한 연구와 학생들에게 적용한 연구가 있다. 교육 프로그램 개발 연구에서는 데이터 시각화 단계를 바탕으로 프로젝트를 진행할 수 있도록 개발하였다. 심희원은 시각화 단계를 '문제 정의', '데이터 전처리', '시각화 설계', '시각화 구현', '결과 분석 및 공유'로 정의하였다. 교육 프로그램은 이를 바탕으로 '데이터 분석과 시각화', '파이썬 시각화 기초', '문제 정의', '데이터 전처리', '다양한 분석 유형의 시각화 실습', '데이터 분석 프로젝트 수행 및 발표'를 개발하여 FGI, 타당화 검토를 실시한 후 최종 개발하였다[9]. 김민규는 '데이터 필요성과 수집', '데이터 정리 및 표현', '데이터 해석'의 절차에 따라 '파이썬 및 데이터 시각화 기초', '데이터 수집', '구글 설문지와 데이터 시각화', '데이터 시각화에서 그래프 선정', '프로젝트'로 개발하였다[12].

프로그래밍 언어를 활용한 데이터 시각화 교육 적용 연구는 정보영재를 대상으로 한 연구와, 초등학교 일반학급을 대상으로 한 연구가 있다. 배연진은 초등학교 6학년 학생을 대상으로 파이썬을 활용한 데이터 시각화 교육을 개발하고 적용하여 학생들의 데이터 리터러시가 향상됨을 보고하였다[10]. 이진영은 정보영재 학생을 대상으로 파이썬을 활용한 데이터 시각화 교육을 하여 컴퓨팅 사고력이 향상된 것을 보고하였다[11]. 김정아는 초등학교 6학년을 대상으로 파이썬을 활용한 데이터 시각화 교육 프로그램 개발 및 적용하여 컴퓨팅 사고력과 창의성 중 유창성, 정교성, 독창성이 향상됨을 보고하였다[12].

3. Data Visualization Automatic Assessment System

데이터 시각화 자동 평가 시스템은 주피터 노트북을 기반으로 작동하며, matplotlib 라이브러리를 MATLAB 방식으로 데이터 시각화하는 것을 자동 평가 및 자가 평가할 수 있도록 개발되었다. 개발된 모듈은 GitHub에 배포하여 무료로 사용할 수 있다.

데이터 시각화 자동 평가 모듈은 문제 제시, 평가할 코드 작성, 평가 기능으로 구현되었다. 문제 제시는 문제 출력 함수를 이용하여 Code cell output에 문제를 출력한다. 문제를 서식 있는 형태로 제공하기 위하여 문제는 HTML Tag가 포함된 문자열로 저장하였으며 문제 출력 함수는 HTML 태그를 렌더링하여 출력한다.

코드 평가는 개인의 주피터 노트북 커널에서 평가가 실시된다. 따라서 평가할 코드 작성은 Code cell을 커널에 파일로 저장하는 매직 커맨드(%%writefile filename)을 이용한다. 평가 기능은 파일로 저장한 평가할 파일을 평가 함수의 매개변수로 이용하며 Code cell output에 평가 결과를 출력한다. 평가는 자가 평가와 자동 평가가 이루어진다. 자가 평가는 한 행에 학생 코드로 작성한 그래프와 모범 답안으로 작성한 그래프를 동시에 출력하고 자가 평가 요소를 출력한다. 자동 평가는 학생이 작성한 그래프와 모범 답안으로 작성한 그래프를 각각 배열로 변환하여 일치 여부에 따라 평가 결과를 출력한다. 수학 교육과정에서 제시한 막대그래프, 꺾은선그래프, 원그래프, 수평선, 수직선, 히스토그램, 상자그림, 산점도를 자기평가와 자동평가로 피드백을 제공한다[13].

4. Education using Public Data

공공데이터는 공공기관에서 보유하거나 관리하고 있는 데이터로 국민들이 쉽게 접근하고 활용할 수 있도록 공공데이터 포털을 통해 데이터에 접근할 수 있도록 하였다[21]. 학생들이 실제 데이터인 공공데이터를 분석하고 해석하는 일련의 활동을 경험함으로써 과제에 대한 흥미를 높이고 몰입할 수 있도록 하며 과제에 대한 진입 장벽이 낮추어 내재적 동기가 높아지며 고차원 사고를 요구하는 활동에 더 적극적으로 참여할 수 있다[16][19]. 김석전과 김태영은 인구 절벽 문제를 주제로 인구 데이터를 분석하는 활동을 하였고 학생들의 내재적 동기와 자기효능감, 학업 성취도가 유의미하게 향상됨을 보고하였다[20].

하지만 공공데이터는 처음부터 학습 목적으로 만들어진 것이 아니기 때문에 인공지능 모델을 만드는 것과 같은 학습 활동을 위해서는 데이터를 바로 사용하기 전 데이터를 정제하는 과정이 필요하다[23].

III. The Proposed Scheme

본 연구에서는 데이터 분석 전 과정을 다루는 것이 아닌 정제된 데이터를 바탕으로 데이터 시각화에 한정하여 문제를 개발하였다.

데이터 시각화에 활용하는 데이터는 교육과정에서 제시하고 학생의 관심을 고려한 공공데이터를 활용하였다. 또한 수학 교육과정에서 제시한 그래프와 관련된 내용요소를 반영하여 다양한 데이터 시각화 방법을 익히고 데이터 특징에 맞는 그래프를 선택할 수 있도록 하였다.

1. Structure of Question Bank

공공데이터 기반 데이터 시각화 문제은행은 Fig.1과 같이 설계하였다. 공공데이터별로 수학 교육과정에서 제시한 그래프와 관련된 학습 요소를 반영한 2~4개의 문제와 모범 답안을 개발하였다. 공공데이터, 문항, 모범 답안은 데이터 시각화 자동 평가 모듈에서 문제출력함수, 평가함수에서 활용할 수 있도록 문제은행으로 작성하였다.

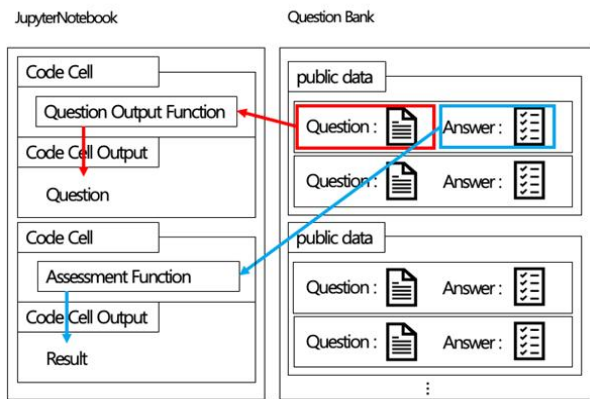


Fig. 1. Question Bank Structure

2. Implementation

2.1 Selection of Public Data

문제를 개발하는 데 활용한 공공데이터는 수학 교육과정에서 제시한 ‘여러 가지 사회 문제’, ‘환경 문제’, ‘지속 가능한 발전’인 범교과 주제를 선정하였다. 선정한 데이터는 음식물 쓰레기, 전기 발전 원료, 육아 휴직, 사교육비, 교통사고, 영화, 지하철 혼잡도, 공공자전거, 고령화 데이터이다. Table 1.은 선정한 공공데이터와 이를 표현하기에 적절한 그래프 관련 학습 요소이다.

Table 1. Data Topics and Graph

Order	Data Topic	Graph
1	Food Waste	Bar
2	Power Generation	Bar, Line
3	Parental Leave	Line
4	Private Education	Line, Pie
5	Car Accident	Pie, Histogram
6	Movie	Histogram, Bar
7	Subway of congestion	Histogram, Bar, Vertical, Box,
8	Public bicycle	Box, Histogram Vertical, Scatter
9	Aging population	Vertical, Scatter Histogram

2.2. Question

문제 틀은 ‘문제 설명’, ‘데이터’, ‘그래프’, ‘데이터 예시’, ‘그래프 예시’로 구성하였다. ‘문제 설명’은 데이터를 사용하여 시각화해야 할 그래프를 설명한다. 프로그램 자동 평가 시스템과 데이터 시각화 자동 평가 모듈은 모두 입력된 데이터를 코드나 알고리즘을 통해 처리하고, 그 결과를 토대로 평가를 수행하는 유사한 작동 원리를 가지고 있다. 이러한 유사성을 고려하여, 데이터 시각화 자동 평가 모듈에서는 ‘입력’과 ‘출력’의 개념을 데이터 시각화의 맥락에 맞게 ‘데이터’와 ‘그래프’로 수정하였다. ‘데이터’는 데이터와 columns에 대한 설명을 제공하였으며 ‘데이터 예시’에는 데이터 시각화에 초점을 맞추기 위해 데이터를 불러오는 코드로 제공하였다. ‘그래프’는 시각화해야 할 그래프에 대한 설명을 제공하며, ‘그래프 예시’에는 모범 답안 그래프를 이미지로 제공하였다. 문제 구성 틀에 대한 설명은 Table 2.와 같다.

Table 2. Draft_Framework of the question

Data Visualization	Note
Description	Description of the Question
Data	provide a description of the public data
Graph	Description of the graph that need to be visualized
Example of Data	Provide code to use data
Example of Graph	Provide graph image

개발한 문항을 체계적으로 관리하기 위하여 8500으로 시작하는 4자리 번호를 부여하였다. 십의 자리는 공공 데이터의 종류를 나타내며 일의 자리는 문항 번호이다. 십의 자리가 0은 음식물 쓰레기량, 1은 원료별 발전량, 2는 성별에 따른 육아 휴직 수, 3은 학교급별 사교육비, 4는 교통 사고, 5는 영화 과점, 6은 지하철 혼잡도, 7은 공유자전거와 날씨, 8은 고령화와 소득 데이터이다.

공공데이터는 전처리하여 csv_file 폴더에 저장하였다. 개발한 문제는 주피터 노트북에서 Code cell Output에서 출력하기 위해 HTML 태그를 포함한 문자열로 저장하여 problem.py 모듈에 저장하였다. Fig 2.는 주피터 노트북에서 문제를 출력한 결과이다.

Question(quest ion_8501)

문제 설명

경상남도 시에서 1년간 발생한 음식물 쓰레기량(톤)과 경상남도 시군구별 인구 데이터입니다. 경상남도 시에서 발생한 음식물 쓰레기량을 막대그래프로 그려봅시다.

데이터

경상남도 시에서 발생하는 음식물 쓰레기량, 인구와 인구당 발생하는 음식물 쓰레기량으로 구성된 데이터
 열 : city_province, emission, population, emission_per_person

그래프

경상남도 시에서 제공하는 시군구별 음식물 쓰레기 및 인구와 관련하여 막대 그래프로 그려봅시다.
 그래프 제목 : '경상남도 시에서 발생하는 음식물 쓰레기량'
 가로축 제목 : '경상남도 시'
 세로축 제목 : '음식물 쓰레기량(톤)'

데이터 예시

```
data =
pd.read_csv('/content/jupyter_judge/csv_file/food_w:
'city_province', 'emission' 열에 있는 데이터를 사용해
봅시다.
```

그래프 예시

Fig. 2. Output of a Problem

2.3. Answer

평가를 위해 필요한 모범 답안, 자가 평가 요소는 problem.py 모듈에 저장하였다. 자동 평가에 활용되는 그래프 이미지는 graph 폴더에 저장하였다. Fig 3은 모범답안을 바탕으로 평가한 결과로 수직선이 일치하지 않아 자동 평가 결과 오답을 출력하고 수직선 내용이 틀림을 피드백으로 제공한 경우이다.

plot_check('._8561.py')

왼쪽 그래프는 여러분이 작성한 그래프입니다. 오른쪽 그래프는 예시 답안입니다.

아래 요소들이 그래프에 반영되었는지 확인해봅시다.
 히스토그램 vlines vlines 범위 가로축 이름 세로축 이름 label과 legend

오답입니다.
 수직선의 x, ymin, ymax값이 틀렸습니다.

Fig. 3. Results of the assessment when answer is incorrect

출력된 2개의 그래프 중 오른쪽에 해당하는 모범 답안 그래프와 하단에 체크박스과 자기 평가 요소가 정상적으로 출력된 것을 확인할 수 있다. Fig 4는 모범 답안을 바탕으로 평가한 결과로 정답인 경우이다.

plot_check('._8561.py')

왼쪽 그래프는 여러분이 작성한 그래프입니다. 오른쪽 그래프는 예시 답안입니다.

아래 요소들이 그래프에 반영되었는지 확인해봅시다.
 히스토그램 vlines vlines 범위 가로축 이름 세로축 이름 label과 legend

정답입니다.
 그래프를 보고 알 수 있는 내용을 적어봅시다. 혼잡도가 매우 높은 경우가 있다.

Fig. 4. Results of the assessment when answer is correct

3. Expert Review Result

데이터 시각화 자동 평가 시스템에서 활용할 공공 데이터 기반한 문제는 타당도를 확인하기 위해 전문가 검토를 실시하였다. 내용 타당도 검토 문항은 선행연구를 토대로 컴퓨터 교육 석사학위 이상의 교사 3인과 협의로 구성했으며 문항의 상세 내용은 Table 3과 같다.

Table 3. Questionnaire of content validity review

Area	Question
Appropriateness of Learning Elements	This question is developed according to the 2022 Curriculum.
	This question reflects the graph and graph elements well.
	The learning elements presented in this question are systematic.
	This question is likely to be applied in the classroom.
	Using this question will help students learn using engineering tools.
	This question is suitable for utilizing the online judge.
Appropriateness of Data	Food waste-related data and graph learning elements are appropriate.
	Electricity generation data and graph learning elements are appropriate.
	Parental leave data and graph learning elements are appropriate.
	Private education expenses data and graph learning elements are appropriate.
	Traffic accident-related data and graph learning elements are appropriate.
	Movie audiences data and graph learning elements are appropriate.
	Subway congestion data and graph learning elements related are appropriate.
	Public bicycles & weather data and graph learning elements are appropriate.
The average monthly income & population composition by region data and graph learning elements are appropriate	

Compete in Data Visualization	Using developed questions based on public data will improve students' data visualization ability.
	Using developed questions based on public data will help students cultivate data programming· statistical convergence competencies.
	By using developed questions based on public data, programming· statistical convergence education can be promoted.
Student's Need and Interest	Using developed questions based on public data will satisfy data programming and statistical training.
	Using developed questions based on public data are at an appropriate level for learners learning data visualization.
	Using developed questions based on public data will arouse interest and interest in programming and statistical education.

문항의 타당도 검토에 참여한 전문가는 컴퓨터 교육 박사 수료이며 영재학교에서 근무하는 현직 정보 교사 2인, 자동 평가 시스템을 활용한 경험이 있고 컴퓨터 교육 박사 학위를 소지한 초등 교사 2인, 컴퓨터 교육 박사과정 중인 교사 3인, 초등수학교육 석사학위를 소지하고 초등학교에서 근무하는 교감 1명, 수학 중등교사 자격증을 소지하고 교육 경력 10년 이상의 교사 2인이다. 전문가의 전공, 학위, 교육경력은 다음과 Table 4과 같다.

Table 4. Experts involved in the review

	Major	Education	Career(year)
A	Computer	Master	18
B	Computer	Master	21
C	Computer	Master	10
D	Computer	Master	8
E	Computer	PhD	9
F	Computer	PhD	10
G	Computer	Master	15
H	Math	bachelor	13
I	Math	bachelor	13
J	Math	Master	22

각 문항은 ‘매우 부적절하다’, ‘부적절하다’, ‘적절하다’, ‘매우 적절하다’의 4점 리커트 척도로 구성하였고 영역별로 전문가 의견을 서술할 수 있도록 제시하였다. 전문가들의 검토 결과를 수집하여 I-CVI(Item Content Validity Index) 산출식을 활용하여 내용 타당도를 분석하였다. I-CVI 계산식은 다음과 같다.

$$I-CVI = \frac{n_r}{N}$$

N = 전체 전문가인원, n_r = ‘적절하다’, ‘매우 적절하다’를 준 응답자

내용 타당도 검토에 참여한 전문가 인원이 10명이어서 절단값은 0.78로 설정하였다[19]. 문항별 통계 결과는 Table 5와 같다.

Table 5. Result of I-CVI statics analysis

area	Item	Mean	SD	I-CVI
Appropriate-ness of Learning Elements	1	3.8	0.42	1
	2	3.9	0.32	1
	3	3.9	0.32	1
	4	3.3	0.48	1
	5	3.7	0.48	1
	6	3.8	0.42	1
Appropriate-ness of Data	7	4	0	1
	8	3.7	0.48	1
	9	3.3	0.67	0.9
	10	3.1	0.74	0.8
	11	3	0.67	0.8
	12	3.7	0.48	1
	13	3.4	0.52	1
	14	3.8	0.42	1
	15	3.3	0.67	0.9
Compete in Data Visualization	16	3.9	0.32	1
	17	3.7	0.48	1
	18	3.7	0.48	1
Student's Need and Interest	19	3.7	0.48	1
	20	3.4	0.52	1
	21	3.7	0.48	1

개발한 문항에 대한 내용 타당성 검사를 위한 I-CVI 값은 모두 0.78이상이므로 내용 타당도가 갖춰졌다고 할 수 있다. 이는 개발한 문항이 교육과정에 부합하여 개발되었고 개발한 문항으로 프로그래밍 및 통계 역량을 높이기에 타당하며 학습자를 고려하여 개발했음을 확인할 수 있다. 4개의 영역 중 데이터의 적절성 영역에서 데이터의 신뢰성과 데이터 소재와 관련하여 전문가 의견이 제시되었다. ‘공공 데이터의 분명한 출처, 객관성을 바탕으로 해야 할 것 같다.’, ‘소재가 좀 더 학생에게 가까웠으면 좋겠습니다 :) 사교육비는 학생들과는 다소 거리가 있어 보이기도 합니다.’, ‘육아휴직자 및 학교별 사교육비 총액은 학생 교육에 다른 자료들 보다 적절성이 떨어지는 것 같습니다.’, ‘학생 경험과 관련있는 데이터를 선정하는 것이 더 좋겠습니다.’와 같은 의견을 제시하였다.

4. Reflect Expert Review

전문가 내용 타당도 검토 결과를 바탕으로 다음과 같이 수정하였다.

첫째, 학생들이 데이터에 친숙하기 위하여 데이터와 관련된 뉴스를 제공하였다. 선정한 데이터는 교육과정에서 제시한 여러 가지 사회 문제, 환경 문제를 다룬 것으로 학생들에

게 친숙하지 않을 수 있다. 따라서 학생들이 데이터에 친숙하고 이해하기 위하여 문제에 관련 뉴스를 추가하였다.

둘째, 데이터를 제공하는 웹 사이트를 링크로 연결하였다. 공공 데이터이기는 하지만 출처, 제공하는 공공기관을 제시하지 않아 실제 데이터로 인식하지 않을 수 있다. 따라서 데이터를 설명하는 곳에 데이터를 제공하는 웹사이트를 링크로 연결하여 필요시 제공하는 웹사이트를 방문하여 확인할 수 있도록 하였다. Table 6.은 수정한 문제 틀이며, Fig. 5는 문제 틀을 ‘관련 뉴스, 데이터, 그래프, 데이터 예시, 그래프 예시’로 수정하여 출력한 결과이다. 관련 뉴스 및 데이터에서 뉴스와 데이터 제공처를 제공하기 위하여 하이퍼링크를 이용하였다.

Table 6. Revised_Framework of the question

Data Visualization	Note
Related article	To arouse student’s interest
Data	provide a description of the public data and Website that provide the data
Graph	Description of the graph that need to be visualized
Example of Data	Provide code to use data
Example of Graph	Provide graph image

Question(question_8501)

문제 설명

관련 뉴스
음식을 쓰레기는 환경적인 문제와 경제적인 문제를 모두 가지고 있습니다. 다음 뉴스를 보고 음식물쓰레기의 문제점을 생각해보세요. [음식물 쓰레기 관련 뉴스](#)

데이터
문제를 제공하는 데이터는 경상남도에서 발생한 1년간 음식물 쓰레기량(톤)과 인구 데이터를 바탕으로 경상남도 시별 음식물 쓰레기량과 1인당 음식물 쓰레기량으로 구성되어 있습니다.
예 : city_province(시), emission(배출량), population(인구), emission_per_person(1인당 음식물 쓰레기 배출량)

그래프
경상남도 시에서 제공하는 시군구별 음식물 쓰레기 및 인구에 관련하여 막대 그래프로 그려봅니다.
그래프 제목 : '경상남도 시에서 발생하는 음식물 쓰레기량'
가로축 제목 : '경상남도 시'
세로축 제목 : '음식물 쓰레기량(톤)'

데이터 예시 그래프 예시

```
data =
pd.read_csv('content/jupyter_judge/csv_file/food_waste.csv')
'city_province', 'emission' 열에 있는 데이터를 사용해봅시다.
```

Fig. 5. Output of a Problem

5. Pilot Text Result

개발한 문항을 초등학생 2명(6학년 1명 학생A, 6학년 1명 학생B)에게 파일럿 테스트를 하였다. 초등학생 2명은

2022년, 2023년 파이썬 기초 프로그래밍을 프로그래밍 자동 평가 시스템을 이용하여 학습했으며 2023년 10월 3주부터 11월 1주까지 방과후 수업을 통해 데이터 시각화를 학습한 학생들이다. 학생들은 데이터 시각화에 앞서 그래프에 대한 설명을 제공했으며 이후 공공 데이터를 기반으로 한 데이터 시각화 자동 평가 모듈을 이용하여 학습하였다. 공공데이터 기반 데이터 시각화 학습 이후 각각 학생들과 인터뷰를 통해 문제의 적절성, 데이터의 적절성, 문제에 대한 학생의 의견을 검토하였다.

A학생과 B학생 모두 그래프로 나타내는 코드는 비교적 단순하고 참고할 수 있는 자료가 있어서 문항의 난이도는 어렵지 않다고 대답하였다. 데이터를 바탕으로 다양한 그래프를 나타낼 수 있고, 많은 데이터를 손으로 그래프를 그리는 것이 아니라 파이썬을 활용하여 나타내는 것이 수학(그래프) 공부에 도움이 된다고 하였다. 또한 문항 정답시 ‘그래프를 통해 알 수 있는 것을 적어봅시다.’를 통해서 데이터를 통해 새롭게 알게 된 점이 있었다고 대답하였다.

하지만 A학생의 경우 뉴스가 제공되어 데이터를 이해하는데 도움이 되지만 데이터가 다루는 주제가 학생들과 공감되지 않은 자료들(육아휴직, 교통사고 등)이어서 흥미가 떨어진다고 대답하였다. B학생은 데이터와 데이터를 불러오는 코드를 제공해 주는 것이 배우기에 쉽지만 한편으로는 문제를 해결하는데 힌트를 많이 제공하는 것 같다고 대답하였다.

6. Final Development Result

전문가 의견과 학생 의견 중 데이터가 학생에게 공감되지 않는 의견을 반영하여 학생을 주제 한 학교 내 안전사고, 학생 건강검진 결과 및 표본 데이터를 바탕으로 문항을 추가로 개발하였다. Table 7.은 추가한 공공데이터와 이를 표현하기에 적절한 그래프 관련 학습 요소이다.

Table 7. Data Topics and Graph

Order	Data Topic	Graph
10	Safety Accident	Bar
11	Medical Checkup	Line, Hist, Vlines

문항 번호는 기존 8500에서 학교 내 안전사고 관련 문제는 십의 자리가 0, 학생 건강검진 결과 데이터 관련 문제는 십의 자리가 2, 학생 건강검진 표본 데이터 관련 문제는 십의 자리가 4로 설정하였다. Table 8.은 개발한 문항 수이다.

Table 8. Number of Questions developed

Data Topic	bar	line	pie	hist	vertical	scatter	box
Food Waste	2						
Power Generation	1	2					
Parental Leave		3					
Private Education		1	3				
Car Accident	1			1			
Movie	1			1			
Subway of congestion	2			1	1		1
Public bicycle				1	1	2	1
Aging population				1	1	2	
Safety Accident	2						
Medical Checkup		3		3	1		
Sum	9	9	3	8	4	4	2

IV. Conclusion and Proposal

정보 기반 사회에서 데이터의 중요성이 강조되고 있으며 이와 더불어 데이터의 수집과 이를 활용할 수 있는 기술이 발달하고 있다. 데이터를 시각화하는 것은 데이터를 직관적으로 파악할 수 있으며 전달하는 효과를 높일 수 있다.

이에 본 연구에서는 프로그래밍 언어를 활용한 데이터 시각화 학습에서 데이터 시각화 자동 평가 모듈에서 활용할 수 있는 공공 데이터 기반의 문제를 개발하였다.

본 연구에서 개발한 문제는 기존의 연구에서 제시한 데이터 분석 전 과정을 다루는 것이 아닌 데이터 시각화에 초점을 맞추어 연습할 수 있도록 하였다. 이를 위해 인위적인 데이터가 아닌 실제 데이터인 공공데이터를 수집하고 전처리하여 제공하였다. 또한 사용자가 다양한 데이터 시각화 방법을 익히고 데이터에 어울리는 시각화 방법을 익힐 수 있도록 수학 교육과정과 연계하여 체계적으로 제시하였다.

이를 위해 교육과정에서 제시하는 ‘지속 가능 발전’, ‘여러 가지 사회 문제’, ‘환경’을 주제로 9개의 데이터를 선정하였다. 또한 문항을 바탕으로 그래프를 체계적으로 그래프 학습 요소와 학습 순서를 작성하였다. 공공데이터와 그래프 학습 요소를 바탕으로 문제 설명, 데이터·그래프 설명, 데이터·그래프 예시를 문제 구성요소로 하여 문제와 모범답안을 개발한 후 데이터 시각화 자동 평가 모듈에 적용하였다. 개발한 문항은 10인의 프로그래밍·수학 전문가

로부터 내용타당도 검사를 실시했으며 모든 문항에 대해서 타당도가 있음을 확인하였다. 하지만 전문가 의견을 반영하여 문제 구성요소를 ‘관련 뉴스’, ‘데이터·그래프 설명’, ‘데이터·그래프 예시’로 수정하고, 데이터 설명에서 공공 데이터를 제공한 기관의 링크를 제공하여 데이터의 신뢰성을 높였다.

파이썬을 학습한 경험이 있는 2인을 대상으로 데이터 시각화 자동 평가 모듈에서 개발한 문항을 토대로 학습하며 파일럿 테스트를 실시하였다. 학생들은 문항의 난이도는 적절하며, 파이썬으로 시각화 하는 것이 수학(그래프) 공부에 도움이 된다고 대답하였다. 하지만 데이터가 학생에게 친숙하지 않아 흥미가 떨어진다고 대답하였다. 이에 학생들에게 친숙할 수 있는 자료(학교 내 안전 사고, 건강 검진)를 바탕으로 추가로 문항을 최종 개발하였다.

본 연구에서 개발한 문항은 전문가 검토를 통해 전체적인 내용 타당도를 확보하였고 전문가 의견을 바탕으로 수정·보완했지만 효과성 검증은 이루어지지 않았다. 추후 연구에서 학습자를 대상으로 데이터 리터러시 등의 검사지를 이용하여 사전 사후 점사를 통해 본 문항의 효과성을 통계적으로 검증하고자 해야 할 것이다.

REFERENCES

- [1] Fry, B. "Visualizing Data: Exploring and Explaining Data with the Processing Environment," O'Reilly Media, Inc. 2007.
- [2] Iakang, Jylee, and Mskim, "Improving visual literacy through data visualization-based art class: A case study of high school," Art Education Review, Vol. 49, pp.1-28. 2014, UCI : G704-000621. 2014.49.001
- [3] Wyner, Y., "A case study: Using authentic scientific data for teaching and learning of ecology," Journal of College Science Teaching, Vol. 42(5), pp. 54-60. 2013
- [4] Bkgye, Yjpark, "Data Visualization Trends in Education: Focusing on Cases and Issues," KERIS, 2017.
- [5] Jslee, "A Study for Cultivating Data Literacy in the Social Studies Education - Focusing on the Analysis of the Data Visualization Contents of the Social Studies Textbook in the 2015 Curriculum," Vol. 58(4), pp. 501-512, 2019, DOI : 10.15812/ter.58.4.201912.501
- [6] Kirk, A. "Data Visualization : A successful design process," Packt publishing LTD, 2012.
- [7] Mark C., "DATA VISUALIZATION AS PART OF HIGH SCHOOL PROGRAMMING," EDULEARN21 Proceedings, 7.2021, DOI:10.21125/edulearn.2021.1916
- [8] Jakim, Mskim, Ymkim, Hjkim amd Jhkim, "Effect of data visualization education with using Python on dcomputational

- thinking of six grade in elementary,” JOURNAL OF The Korean Association of information Education, Vol. 23(3). pp. 197-206. 2019, DOI : 10.14352/jkaie.2019.23.3.197
- [9] Hwsim, “Development of data visualization education program using python for data analysis,” Master’s Dissertation, Sungkyunkwan University, 2022.
- [10] Yjbae, “Effect of data visualization education with using Python on data literacy of elementary school students,” Master’s Dissertation, Daegu National University of Education, 2023.
- [11] Jylee, “Effects of Data Visualization Education Using Python on Improvement of Computational Thinking Ability in Information Gifted Students of Elementary school,” Master’s Dissertation, Korea National University of Education, 2021.
- [12] Mngkim, “Development and application of data visualization education using python and their effects : for sixth grade in elementary school,” Master’s Dissertation, Jeju University, 2020.
- [13] Hngo, Yjlee, “Design and Implementation of a Data Visualization Assessment Module in Jupyter Notebook,” Journal of the Korea society of computer and information, Vol. 28(9), pp. 167-176, 09.2023. DOI : 10.9708/jksci.2023.28.09.167
- [14] Bjtak, “A Study on the Frames of Statistical Graphs in the Elementary School Mathematics Textbooks,” Education of primary school mathematics, Vol. 23(4), 2020, DOI : 10.7468/jksmec.2020.23.4.157.
- [15] Lo, L.Y., Ming, Y., and Qu, H., “Learning Vis Tools: Teaching Data Visualization Tutorials,” 2019 IEEE Visualization Conference (VIS), pp. 11-15., 2019. DOI: 10.1109/VISUAL.2019.8933751
- [16] Erwin Jr, R. W., “Data literacy: Real-world learning through problem-solving with data sets,” American Secondary Education, Vol. 43(2) pp. 18-26, 04.2015.
- [17] Hngo, Swkim, Yjlee, “Design and Implementation of a Programming Automatic Assessment System in Jupyter Notebook,” International Journal on Advanced Science, Engineering and Information Technology, Vol. 13, No. 3, pp. 1080-1086, 2023. DOI: <http://dx.doi.org/10.18517/ijaseit.13.3.18457>
- [18] Wychang and Sskim, “Development and application of algorithm judging system : analysis of effects on programming learning.” The Journal of Korean Association of Computer Education, Vol. 17, No. 4, pp. 45-57, 2014. DOI : 10.32431/kace.2014.17.4.005
- [19] Tjseong, “Understanding and Application of Modern Basic Statistics,” Hakgisa, 2019.
- [20] Sjkim and Tykim, “The Effect of Data Literacy on Students’ Learning Motivation and Academic Achievement in Secondary Problem Solving Class Based on Computational Thinking.” Korean Journal of Teacher Education, Vol. 34(4) ,2018, DOI : 10.14333/KJTE.2018.34.4.125
- [21] Ministry of Interior and Safety, Public Data portal, <https://www.data.go.kr>
- [22] Ministry of Education, “Math Curriculum,” December 2022
- [23] SyPark, Dylee and Jakim, “Cases of Public Data Purification Needed to Build Learning Data,” Asia-pacific Journal or Convergent Research Interchange, Vol. 9(5), 2023, DOI : 10.47116/apjcri.2023.05.10.

Authors



HakNeung Go received the B.S. degree in Science education in Gwangju National University of Education, Korea in 2015. He completed Ph.D. degree in Computer education from Korea National University of

Education, Korea, in 2023. He is currently a Teacher in the Songjeong Jungang Elementary School, Korea. His research interests include programming, informatics education, online judge, AI education, Data Visualization, Data Science.



Sangsu Jeong received the B.S. degree in Computer Education from Gyeongsang National University in 2009. He is currently a doctoral candidate in Computer education from Korea National University of Education,

Korea. He is currently working as an informatics teacher at Gyeonggi Science High School for the gifted. His research interests include informatics education, programming education, and artificial intelligence education.



Youngjun Lee received the B.S. degree in Computer Science from Korea University, Korea in 1988. He received the Ph.D. degree in Computer Science from the University of Minnesota, Minneapolis, in 1994.

He is currently a Professor in the Department of Computer Education, Korea National University of Education. His research interests include intelligent system, learning science, informatics education, technology & engineering education and AI.