

엔트리를 활용한 초등 데이터 과학 교육 사례 연구

허경

경인교육대학교 컴퓨터교육과

요약

데이터과학은 스몰데이터 분석에서 출발하여, 빅데이터 분석을 위한 머신러닝, 딥러닝까지 포함하고 있다. 데이터과학은 인공지능 기술의 핵심 영역이고, 학교 교육과정에 체계적으로 반영해야 할 내용이다. 데이터과학 교육을 위해, 엔트리에서도 초등교육용 데이터 분석 도구를 제공하고 있다. 빅데이터 분석에서는 데이터 표본을 추출하여, 통계학적인 추측과 판단을 통해 분석결과를 해석한다. 본 논문에서는 통계학적인 지식을 필요로 하는 빅데이터 분석 영역을 초등영역에서 제외하기로 하고, 초등영역에 초점을 맞춘 데이터과학 교육 사례를 제안하였다. 이를 위해서, 일반적인 데이터과학 교육 단계를 먼저 설명하고, 초등 데이터과학 교육 단계를 새롭게 제안하였다. 그리고 엔트리에서 제공하는 공공 스몰 데이터를 사용한 데이터 변수 값 비교 사례와 데이터 변수 간 상관관계 분석 사례를 초등 데이터과학 교육 단계에 따라 제안하였다. 본 논문에서 제안된 엔트리 데이터분석 사례들을 활용하면, 여러 교과에서 발생하는 데이터를 사용한 초등 데이터과학 융합 교육이 가능하다. 또한, 엔트리를 사용하여 텍스트, 음성 및 영상인식 AI 도구와 결합한 데이터과학 교육 자료도 개발 가능하다.

키워드 : 인공지능, 데이터과학, 엔트리, 스몰데이터, 소프트웨어 교육

A Study on Elementary Education Examples for Data Science using Entry

Kyeong Hur

Dept. of Computer Education, Gyeongin National University of Education

ABSTRACT

Data science starts with small data analysis and includes machine learning and deep learning for big data analysis. Data science is a core area of artificial intelligence technology and should be systematically reflected in the school curriculum. For data science education, The Entry also provides a data analysis tool for elementary education. In a big data analysis, data samples are extracted and analysis results are interpreted through statistical guesses and judgments. In this paper, the big data analysis area that requires statistical knowledge is excluded from the elementary area, and data science education examples focusing on the elementary area are proposed. To this end, the general data science education stage was explained first, and the elementary data science education stage was newly proposed. After that, an example of comparing values of data variables and an example of analyzing correlations between data variables were proposed with public small data provided by Entry, according to the elementary data science education stage. By using these Entry data-analysis examples proposed in this paper, it is possible to provide data science convergence education in elementary school, with given data generated from various subjects. In addition, data science educational materials combined with text, audio and video recognition AI tools can be developed by using the Entry.

Keywords : Artificial Intelligence, Data Science, Entry, Small Data, Software Education

논문투고 : 2020-09-14

논문심사 : 2020-09-20

심사완료 : 2020-10-05

1. 서론

‘정보’가 중요한 가치를 갖게 된 현재 시대에서 데이터를 이용하여 다양한 문제를 해결하는 능력은 SW역량과 더불어 누구에게나 필요한 것이 되었다. 최근, 데이터과학이라고 하면 AI, 빅데이터 그리고 머신러닝 등을 연상하고, 컴퓨터가 대량의 데이터를 처리해서 다양한 문제를 자동으로 해결하는 기술을 데이터과학 또는 인공지능이라고 간주하고 있다[1-5].

그러나, 데이터의 특징이나 속성을 이해하기 위해서는 그 데이터가 어떠한 과정을 거쳐서 얻어진 것인지 이해해야 한다. 현대 사회에는 대량의 데이터가 발생 및 유통되고 있고, 누구라도 그 데이터를 구할 수 있다. 데이터가 많으면 많을수록 데이터의 특징이나 속성을 적절히 가려내어, 가공하고 활용하는 능력이 개개인에게 필요하다[1-5]. 또한, 사회에서는 많은 데이터를 축적하고 관리하여 다양한 분석 결과를 통해, 인사이트를 도출할 수 있도록 데이터 기반의 의사결정 체계를 원하고 구축하고 있다. 이러한 의사결정 체계를 구축하기 위해서는, 데이터를 통해 트렌드 및 인사이트를 찾을 수 있는 데이터 분석가, 데이터 과학자 등의 역할이 더욱 중요해지고 있다[1-5].

우리나라에서는 2019년 12월 국무회의를 통해 과학기술정보통신부를 비롯한 전 부처가 참여하여 ‘인공지능국가전략’을 발표하였으며 초·중·고 학교 급 별로 교육해야 할 인공지능 핵심 주제가 제시되었다[6-7]. 이러한 국가 차원의 전략을 바탕으로 교육부에서도 AI교육 및 AI융합교육의 필요성을 강조하고 있다. 데이터를 과학 교육을 위해, 엔트리에서도 초·중·교육용 데이터 분석 도구를 제공하고 있다. 빅데이터 분석에서는 데이터 표본을 추출하여, 통계학적인 추측과 판단을 통해 분석결과를 해석한다. 본 논문에서는 통계학적인 지식을 필요로 하는 빅데이터 분석 영역을 초·중·교육에서 제외하기로 하고, 초·중·교육에 초점을 맞춘 데이터과학 교육 사례를 제안하였다. 이를 위해, 일반적인 데이터과학 교육 단계를 우선 설명하고, 초·중·데이터과학 교육 단계를 새롭게 제안하였다. 그리고 엔트리에서 제공하는 공공 스몰 데이터를 사용한 데이터 변수 값 비교 사례와 데이터 변수 간 상관관계 분석 사례를 초·중·데이터과학 교육 단계에 따라 제안하였다. 본 논문에서 제안된 엔트리 데이터분석

사례들을 활용하면, 여러 교과에서 발생하는 데이터를 사용한 초·중·데이터과학 융합 교육이 가능하다.

2. 이론적 배경

2.1. 데이터과학 교육에 대한 선행연구 분석

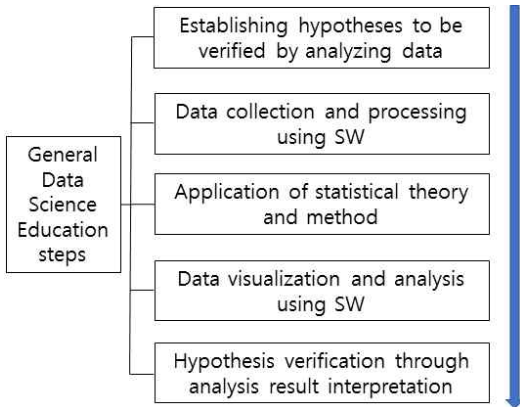
데이터 과학 교육에 대한 선행 연구를 조사한 결과, 인공지능 교육과는 달리, 데이터과학 교육에 대한 선행 연구는 많이 이루어지지 않은 상태이다. 데이터과학 교육에 대한 선행 연구를 살펴보면, 장영재(2017)은 ‘4차 산업혁명 시대 데이터과학 교육 방향성 모색’ 연구를 통해, 데이터과학 교육의 필요성을 주장하였다[8]. 이를 위해, 데이터과학 교육을 통해 갖추어야 할 5가지 소양을 제시하였다. 박윤수 및 이수진(2020)은 보편적 빅데이터를 교육하기 위해서, 중점적으로 교육해야 할 지식 영역을 산출하였다[9]. 이를 위해, 빅데이터 전문가 설문 조사를 실시한 결과로, 빅데이터 가공 과정에 반드시 빅데이터 처리 프레임워크 또는 고성능 컴퓨터가 필요한 것은 아니라고 하였다. 다시 설명하면, 빅데이터를 교육하기 위해서는, 컴퓨터 과학적 지식과 스킬보다는 빅데이터 분석 방법과 응용 방법 중심으로 교육해야 한다고 하였다[9].

김용민(2018)은 컴퓨팅 도구를 활용한 데이터 과학 교육 프로그램을 개발 및 적용한 결과를 제시하였고, 이를 통해, 컴퓨팅 사고력과 창의성을 향상할 수 있다고 하였다[10]. 그리고, 김용민 및 김중훈(2017)에서는 엑셀 스프레드시트를 이용한 42차시 데이터 과학 교육 프로그램의 내용으로, 엑셀의 기본 기능 익히기와 스스로 데이터를 수집하여 안전한 생활을 위해 노력해야 할 점을 실천하기, 바람직한 학습 방법에 대해 생각해보기, 독서가 창의력 신장에 도움을 주는지 알아보기, 개선하고자 하는 문제 주제를 정하고 스스로 데이터를 수집하여 이를 해결해 보기’ 등을 제안하였다[11]. 또한, 제안한 엑셀 스프레드시트 데이터 과학 교육 프로그램을 수업에 적용하면, 컴퓨팅 사고력 향상에 효과적이라고 분석하였다[11]. 이러한 선행 연구와 달리, 본 논문에서는 스몰데이터 크기의 한국 공공데이터가 제공되고, 데이터 수집, 가공 및 시각화에 사용되는 교육용 프로그래밍 활동이 가능한 엔트리를 선택하여, 초·중·데이터과학 교육 단계

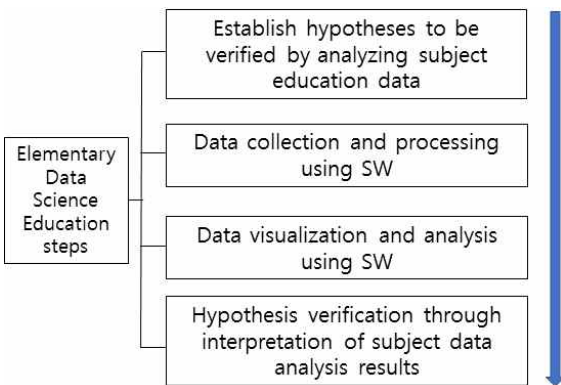
를 새롭게 제안하였다. 그리고 엔트리 데이터 분석 프로그램 교육 사례들을 초등 데이터과학 교육 단계에 따라 제시하였다.

2.2. 초등 데이터과학 교육 단계

데이터과학에서는 데이터를 이용하여 다양한 문제를 구체적으로 해결한다. 이를 위해서 해당 문제를 해결하기 위해 필요한 가설을 설정하여, 가설이 타당한지 여부를 결정하고 이에 따라 해당 문제에 대한 해결방안을 결정한다. (Fig. 1)에 나타난 일반적인 데이터과학 교육 단계서는 빅데이터를 분석하기 위한 통계적 방법을 결정하고 적용한다. 특히, 빅데이터에서 분석에 필요한 데이터 표본을 추출 및 집계할 때, SW를 개발하고 적용한다[5].



(Fig. 1) General Data Science Education steps



(Fig. 2) General Data Science Education steps

이에 대해, (Fig. 2)에서는 초등 데이터과학 교육 방법에 있어서 필요한 단계들을 정의하였다. 본 논문에서는 통계학적인 지식을 필요로 하는 빅데이터 분석 영역을 제외하고, 초등 교과에서 발생하는 데이터로부터 검증하고자 하는 가설을 설정하는 단계부터 정의하였다. 여기서 사용되는 SW는 엔트리로 설정하였고, 엔트리에서 제공하는 초등교육용 데이터 수집, 가공 및 분석 도구를 사용한다.

초등 데이터과학 교육에서는 엔트리에서 수집된 교과 관련 공공데이터로부터 가설 검증에 필요한 변수들을 새롭게 생성한다. 이 신규 변수 하나는 수집한 데이터 변수들 중 속성이 동일한 변수 값들만 선택하여, 평균과 합계 등 기초 통계처리를 통해 생성된다.

한편, 공공데이터가 아닌, 교과별 수업에서 발생하는 데이터를 행과 열로 엑셀 파일에 축적하여 기록 및 저장해 두고, 이를 엔트리에 넣어 원하는 데이터로 가공할 수 있다. 예를 들어, 일주일간 특정 장소에서 특정 시간 간격으로 측정된 온도, 습도, 풍향, 풍력 및 강수량 정보를 축적하여, 과학 수업과 연계한 데이터과학을 실습할 수 있다. 다른 교과를 추가로 예를 들면, 한 달 동안 경제, 정치, 사회, 문화, IT 분야별 신규 뉴스 주제 개수 등을 1일 간격으로 기록 및 축적하여, 사회 수업과 연계한 데이터과학을 실습할 수 있다.

3. 엔트리와 스몰 데이터를 활용한 데이터 변수 값 비교 사례

2.2 절에서 제안한 초등 데이터과학 교육 단계에 따라, 본 장에서는 데이터 변수 값 비교 사례를 제안하였다. <Table 1>에 제시된 가설에 대해, <Table 2>에서 엔트리에서 검증하고자 ‘월평균 미세먼지농도’ 공공데이터를 테이블에 추가하였다. 그리고 <Table 3>에서 데이터 시각화를 위한 분석을 위해, ‘5년단위미세먼지농도’라는 이름의 테이블을 엔트리에서 직접 만들어 추가하고, 차트 생성을 위한 가로축과 세로축, 표현값을 미리 설정하였다. 여기서 만들어진 신규 변수는 ‘5년간전국평균미세먼지농도’이다. 이 신규 변수의 2010년~2014년 1행의 값과 2015년~2019년 2행의 값을 가공하여 생성하기 위한 엔트리 프로그램 작성과정이 <Table 4>에서 데이터 가공 및 수집 단계로 제시되었다.

<Table 1> Data Variable Value Comparison Step 1 Example

step	Elementary data science education content
Step 1	<ul style="list-style-type: none"> Hypothesis: The national fine dust concentration has been higher (deteriorated) in the last five years from 2015 to 2019 compared to the previous five years from 2010 to 2014.
hypothesis setup	<ul style="list-style-type: none"> The reason for setting the hypothesis: The last five years will represent the most serious situation due to air pollution such as global warming and industrialization.

‘월평균 미세먼지농도’ 공공데이터 테이블에서 각 연도 내 월별 전국 평균 미세먼지농도 값들을 60개월, 즉, 2010년도부터 2014년까지 5년 12개월에 걸쳐 합산한 후, 평균값을 취한 값을 ‘5년단위미세먼지농도’ 테이블의 해당위치에 입력하였다. 2015년도부터 2019년도 데이터 중 2019년도 12월 데이터는 존재하지 않아, 해당 기간은 59개월에 걸친 반복 연산이 SW에 설정되었다.

<Table 2> Data variable value comparison step 2-1

Elementary data science education content, step 2-data collection

- Import ‘Monthly Average Fine Dust Concentration’ public data by using Entry SW

No.	월	전국	강원도	경기도
1	2010.01	57	62	68
2	2010.02	49	55	56
3	2010.03	64	70	67
4	2010.04	51	58	57
5	2010.05	59	60	66
6	2010.06	52	48	62
7	2010.07	36	33	41
8	2010.08	36	32	39
9	2010.09	32	24	34
10	2010.1	45	38	51
11	2010.11	73	68	83
12	2010.12	63	63	71

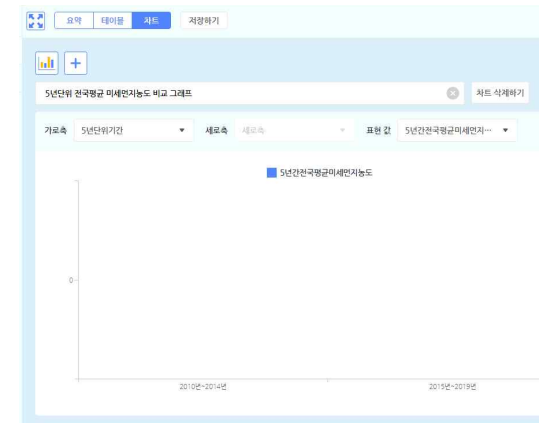
<Table 3> Data variable value comparison step 2-2

Elementary data science education content, step 2-data processing and collection

- Directly make and add “Fine Dust Concentration by Five Years” table by using Entry SW

No.	5년단위기간	5년간전국평균미세먼지
1	2010년~2014년	0
2	2015년~2019년	0

- Preset the horizontal / vertical axis and expression values of the chart to be created



<Table 4> Data variable value comparison step 2-3

Elementary data science education content, step 2-data processing and collection

- Existing data processing using entry SW coding, collecting new data through it

```

for (let i = 1; i <= 60; i++) {
    // ... (code for data processing) ...
}
    
```

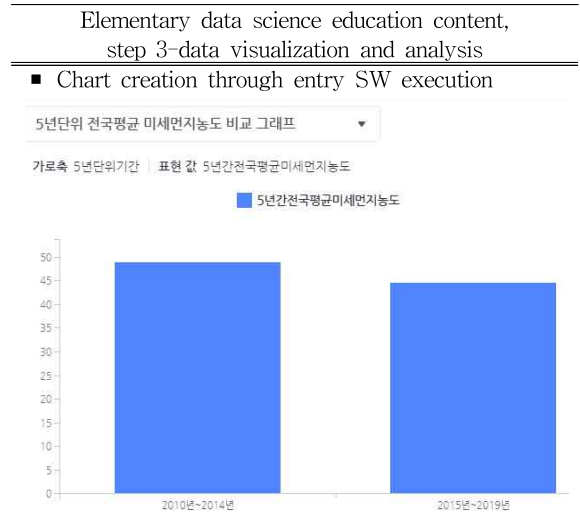
<Table 4>에 제시한 엔트리 프로그램을 구현하는 과정을 설명한다. ‘5년단위미세먼지농도’ 테이블의 ‘5년간

전국평균미세먼지농도' 열 데이터의 1행의 값을 계산하기 위해, 10-14년미세먼지 변수를 생성한다. 그리고, 엔트리에서 불러온 '월평균 미세먼지농도' 공공데이터 테이블의 '전국' 열의 데이터를 1행, 즉, 2010년 1월 평균 값부터 읽어와 더하기 위해, 시작행번호 변수를 생성한다. 이 시작행 변수값을 1부터 60까지 변화시키면서, 2010년 1월부터 2014년 12월까지 60개 월간 전국평균미세먼지 값들을 10-14년미세먼지 변수에 더하고 저장한다. 그리고 60으로 나누어 평균값을 구한 것을 10-14년미세먼지 변수 공간에 다시 저장한다. 이 저장된 10-14년미세먼지 변수 값을 '5년단위미세먼지농도' 테이블의 '5년간전국평균미세먼지농도' 열 데이터의 1행에 넣고 저장한다. 같은 방식으로, 시작행 번호가 61번부터 119번까지 변화하면서, 2015년 1월부터 2019년 11월 전국평균미세먼지 값들을 이번에는 15-19년미세먼지 변수 공간을 만들어 더하고, 59로 나눈 평균값으로 최종 저장한다. 그리고 저장된 최종 15-19년미세먼지 변수 값을 '5년간전국평균미세먼지농도' 열 데이터의 2행에 넣고 저장한다. 그 후 엔트리에서 차트창을 열면, <Table 5>의 차트가 생성된다,

여기서 제시된 엔트리 프로그램은 함수가 등장하지 않는 등 컴퓨팅 사고 절차에 따라 작성된 SW가 아니다. 여기서 처리한 데이터의 크기보다 더 큰 데이터를 통계 처리하여 데이터를 수집할 경우는 문제분할 함수들과 더 많은 변수들을 사용한 자동화 SW가 필요하다. 이 부분이 데이터과학 교육과 컴퓨팅 사고기반 SW교육이 결합될 수 있는 부분으로 판단된다[12-13].

<Table 5>의 3단계 데이터 시각화 및 분석에서는 데이터 가공 및 수집 단계에서 작성된 엔트리가 실행된 모습이 나타나 있다. 실행 결과를 보면, 1단계에서 설정된 가설과 달리, 2010년~2014년 5년간 전국평균미세먼지농도값이 48.73, 2015년~2019년 5년간 전국평균미세먼지농도값이 44.57로 가설이 기각된 결과가 나타났다. 이 결과로부터, 4단계 '교과데이터 분석결과 해석을 통한 가설 검증'이 실시되어야 한다. 이는 해당 교과 지식을 바탕으로 해석하고, 본 결과가 발생한 원인을 분석해야 한다.

<Table 5> Data variable value comparison step 3



4. 엔트리와 스몰 데이터를 활용한 데이터 변수 간 상관관계 분석 사례

2.2 절에서 제안한 초등 데이터과학 교육 단계에 따라, 본 장에서는 데이터 변수 간 상관관계 분석 사례를 제안하였다. 데이터 변수 간 상관관계 분석사례는, 해당 데이터 변수들에 대한 데이터들을 가공하여 수집한 후, 차트 생성을 통한 데이터 시각화를 통해, 일반적인 비례 및 반비례 개념에 대한 해석이 가능한 예제로서, 통계학적 지식이 필요 없는 기초적인 내용으로 판단하였다. 이에 본 논문에서 제안하는 사례로 추가하였다.

<Table 6>에 제시된 가설에 대해, 엔트리에서 검증하고자 '월평균 미세먼지농도' 외에 <Table 7>에서 '계절별 기온' 공공데이터를 테이블에 추가하였다. 그리고 <Table 8>에서 데이터 시각화를 위한 분석을 위해, '기온과 미세먼지관계'라는 이름의 테이블을 엔트리에서 직접 만들어 추가하고, 막대그래프와 분산형 차트 생성을 위한 가로축과 세로축, 표현값을 미리 설정하였다. 여기서 만들어진 신규 변수는 연도별 '전국평균미세먼지농도'이다. 이 신규 변수의 2010년 1행부터 2018년 9행까지 값들을 가공하여 생성하기 위한 엔트리 프로그램 코드 내용이 <Table 9> 데이터 가공 및 수집 단계에 제시되었다.

<Table 6> Correlation analysis between data variables
Step 1 Example

step	Elementary data science education content
Step 1	<ul style="list-style-type: none"> Hypothesis: The annual national average temperature and the national average fine dust concentration measured from 2010 to 2018 have a proportional correlation.
hypothesis setup	<ul style="list-style-type: none"> The reason for setting the hypothesis: The increased annual average fine dust concentration would have caused an increase in air pollution, leading to an increase in the annual average temperature.

<Table 7> Data Correlation analysis step 2-1

Elementary data science education content,
step 2-data collection

- Import "Seasonal temperature" public data by using Entry SW

No.	연도	연평균	봄	여름
1	1973	12.4	11.6	24.5
2	1974	11.4	10.8	22.4
3	1975	12.6	11.2	23.9
4	1976	11.7	10.9	22.6
5	1977	12.3	11.8	23.5
6	1978	12.8	11.7	24.7
7	1979	12.6	11.2	23.5
8	1980	11.2	11	22.1
9	1981	11.4	11.5	23.6
10	1982	12.4	12	23.4
11	1983	12.3	12.2	23.4
12	1984	11.8	10.7	24.3

1가지 통계 데이터의 특성을 해석하려면, 원그래프, 막대그래프, 꺾은선 그래프가 편리하지만, 동시에 2가지 통계 데이터의 특성을 분석할 때에는 산포도가 효과적이다. 산포도는 엔트리에서 생성된 테이블에서 차트를 추가할 때, 분산형을 선택하면 산포도 결과를 생성할 수 있다[5,14-15]. 여기서는 전국평균기온과 전국평균미세먼지농도 간의 상관관계를, 2010년부터 2018년까지 연도별 산포도로 시각화하여 분석하는 것이다.

<Table 8> Data Correlation analysis step 2-2

Elementary data science education content,
step 2-data processing and collection

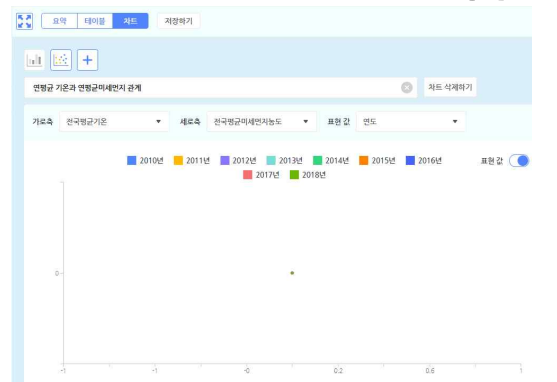
- Directly make and add 'Relationship between temperature and fine dust' table by using Entry SW

No.	연도	전국평균기온	전국평균미세먼지농도
1	2010년	0	0
2	2011년	0	0
3	2012년	0	0
4	2013년	0	0
5	2014년	0	0
6	2015년	0	0
7	2016년	0	0
8	2017년	0	0
9	2018년	0	0

- Preset the horizontal / vertical axis and expression values of the chart to be created-Bar graph



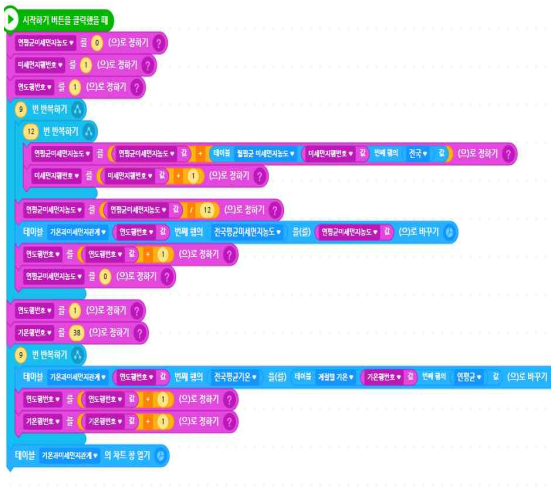
- Preset the horizontal / vertical axis and expression values of the chart to be created-Scatter graph



<Table 9> Data Correlation analysis step 2-3

Elementary data science education content,
step 2-data processing and collection

- Existing data processing using entry SW coding, collecting new data through it



<Table 2>의 ‘월평균 미세먼지농도’ 공공데이터 테이블의 2010년도부터 2018년까지 각 연도 내 월별 전국 평균 미세먼지농도 값들을 12개월에 걸쳐 합산한 후 평균값을 취하였다. 그 9개 평균값들을 ‘기온과 미세먼지관계’ 테이블의 ‘전국평균미세먼지농도’ 1행부터 9행까지 해당위치에 입력하였다. 여기서 제시된 엔트리 프로그램에서도 문제분할과 패턴인식에 의한 함수 구조가 등장하지 않는 등 컴퓨팅 사고 절차에 따라 작성된 SW가 아닌 것으로 보완이 가능하다[12-13].

<Table 9>에 제시한 엔트리 프로그램을 구현하는 과정을 설명한다. ‘기온과 미세먼지관계’ 테이블의 ‘전국평균미세먼지농도’ 열 데이터의 1행부터 9행까지 값들을 계산하기 위해, 연평균미세먼지농도 변수를 생성한다. 한편, ‘기온과 미세먼지관계’ 테이블의 ‘전국평균미세먼지농도’ 열 데이터의 1행부터 9행까지 해당 위치로 이동하기 위해 ‘연도행번호’ 변수를 생성했다. 연도행번호가 1인 상태로 출발하면, 그것은 ‘기온과 미세먼지관계’ 테이블의 ‘전국평균미세먼지농도’의 2010년도 값을 구하는 것을 시작한다는 것이다.

그리고, 엔트리에서 불러온 ‘월평균 미세먼지농도’ 공공데이터 테이블의 ‘전국’ 열의 데이터를 1행, 즉, 2010

년 1월 평균 값부터 읽어와 더하기 위해, 미세먼지행번호 변수를 생성한다. 이 미세먼지행 변수 값을 1부터 12까지 변화시키면서, 2010년 1월부터 2010년 12월까지 12개 월간 전국평균미세먼지 값들을 연평균미세먼지농도 변수에 더하고 저장한다. 그리고 12로 나누어 평균값을 구한 것을 연평균미세먼지농도 변수 공간에 다시 저장한다. 이 값이 2010년도 연평균 전국미세먼지농도값이 된다. 연도행번호 변수 값이 1이므로, 처음에 2010년도 연평균미세먼지농도 변수 값이 ‘기온과 미세먼지관계’ 테이블의 ‘전국평균미세먼지농도’ 열 데이터의 1행 위치에 저장된다, 같은 방식으로, 연도행번호가 1씩 늘어나고, 연평균미세먼지농도 변수 값이 0으로 초기화되고 미세먼지행 변수 값이 13부터 12씩 증가하면서 다음연도 연평균 전국미세먼지농도 값이 계산되어, ‘기온과 미세먼지관계’ 테이블의 ‘전국평균미세먼지농도’ 열 데이터의 다음 행 위치에 저장된다.

이렇게 연도행번호가 1부터 9까지 변화하면서 2010년부터 2018년까지 연평균 전국미세먼지농도 9개 값들이 ‘기온과 미세먼지관계’ 테이블의 ‘전국평균미세먼지농도’ 열 데이터의 1행부터 9행 위치에 채워진다. 그 후, 연도행번호 변수값을 다시 1로 초기화한다. 이번에는 ‘기온과 미세먼지관계’ 테이블의 ‘전국평균기온’ 열 데이터의 1행부터 9행까지 값들을 채우기 위한 코드가 작성되었다. 연도행번호 변수 값은 여기에서 ‘기온과 미세먼지관계’ 테이블의 ‘전국평균기온’ 열 데이터의 1행부터 9행까지 이동하기 위해 사용된다. 그리고 ‘계절별기온’ 테이블의 ‘연평균’ 기온 값을 읽어오기 위해, 기온행번호 변수가 새롭게 생성되었다. 2010년도 연평균 기온 값은 ‘계절별기온’ 테이블의 38번째 행에 있다. 그래서 기온행번호 변수 값이 38로 설정된다. 연도행번호가 1부터 9까지 변화하고, 기온행번호 변수 값이 38부터 46까지 변화하면서, ‘계절별기온’ 테이블에 있는 2010년부터 2018년 연평균 기온 값들이 ‘기온과 미세먼지관계’ 테이블의 ‘전국평균기온’ 열 데이터의 1행부터 9행까지 값들로 채워지게 된다. 최종적으로 <Table 8>에 제시한 ‘기온과 미세먼지관계’ 테이블의 내용이 모두 채워지게 된다. 그 후 엔트리에서 차트창을 열면, <Table 10>의 차트가 생성된다,

<Table 10>의 3단계 데이터 시각화 및 분석에서는 데이터 가공 및 수집 단계에서 작성된 엔트리가 실행된

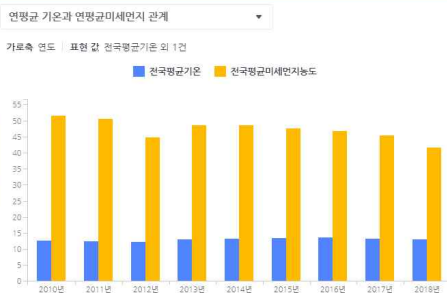
모습이 나타나있다. 실행 결과를 보면, 막대그래프와 분산형 산포도 차트 두 가지를 볼 수 있다. 1단계에서 설정한 가설은 ‘2010년도부터 2018년도까지 측정된 연도별 전국평균기온과 전국평균미세먼지농도는 서로 비례하는 상관관계를 갖는다.’이었다. 이러한 상관관계는 막대그래프에서 해석해 내기 어렵다. 반면에 분산형 산포도에서는 해석하기가 비교적 쉽다[3-5].

분산형 산포도에는 9개의 점들이 나타나 있다. 각 점은 2010년도부터 2018년도까지 연도를 나타낸다. 즉, 각 점은 가로축은 전국 연평균기온 값, 세로축은 전국 연평균 미세먼지농도 값으로 2차원 점이 도시된 것이다. 이 9개의 점들이 Y=X 그래프 형태로 배치되었다면, 서로 비례하는 상관관계를 갖는 것이며, 가설은 타당한 것으로 검증된다. 이와 반대로 Y=A-X 형태로 반대방향으로 배치되면, 이는 서로 반비례하는 상관관계를 나타내는 것이다. <Table 10>의 경우처럼 흩어진 경우는 상관관계가 불분명한 경우에 해당된다[3-5].

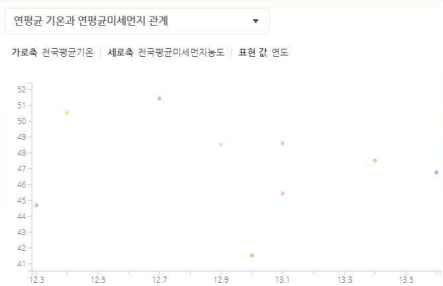
<Table 10> Data Correlation analysis step 3

Elementary data science education content, step 3-data visualization and analysis

■ Chart creation through entry SW execution-Bar graph



■ Chart creation through entry SW execution-Scatter graph



1단계에서 설정된 가설과 달리, ‘2010년도부터 2018년도까지 측정된 연도별 전국평균기온과 전국평균미세먼지농도는 상관관계가 없다.’라는 결과로 분석되어, 가설은 기각되었다. 이 결과로부터, 4단계 ‘교과데이터 분석 결과 해석을 통한 가설 검증’이 실시되어야 한다. 이 또한 해당 교과 지식을 바탕으로 해석하고, 본 결과가 발생한 원인을 분석하여, 또 다른 가설을 설정하고 검증하는 과정을 반복하여 실시해야 한다.

5. 결론 및 향후 연구과제

본 논문에서는 초등 데이터과학 교육 단계를 새롭게 제안하였다. 그리고 엔트리에서 제공하는 스몰 데이터 크기의 공공 데이터를 사용한 데이터 변수 값 비교 사례와 데이터 변수 간 상관관계 분석 사례를 초등 데이터과학 교육 단계에 따라 제안하였다. 이를 통해, 초등 데이터과학 교육 단계가 실제로 실행 가능하다는 것을 나타내었다.

본 논문에서 제안된 데이터분석 예제들을 확장하면, 여러 교과 교육 강좌에서 활용할 수 있는 데이터를 바탕으로 데이터과학 융합 교육이 가능하다. 본 논문에서 제시한 두 가지 예제도 과학 또는 환경 교육과 융합 가능한 소재로 판단된다. 더 나아가, 여러 교과 수업 내에서 발생하는 스몰 데이터들을 축적하고 엔트리 프로그램에 직접 추가하여, 데이터분석 활동을 실행할 수 있다. 분석하고자 하는 데이터속성들을 정의하고, 정의한 데이터를 SW를 통해 가공하여 수집하는 것은 진정한 SW교육이라고 할 수 있다.

교과 수업에서 생성한 데이터 또는 엔트리 및 국가통계포털(KOSIS)에서 수집하고자 하는 공공데이터의 규모가 커지고, 가설 분석에서 요구하는 데이터 속성(열)의 개수가 증가하면, 단순한 SW코딩에서 컴퓨팅 사고 기반 SW 프로그래밍으로 확장되게 된다. 문제분할, 패턴인식 및 자료구조를 생각한 프로그래밍으로 사고가 확장되게 된다. 즉, 교과 및 공공 데이터를 활용한 데이터과학 융합교육을 통해 컴퓨팅 사고 기반 SW교육이 활성화될 수 있다. 이에 첫 번째 후속 연구과제로 컴퓨팅 사고 기반 SW교육과 데이터과학 융합교육이 결합된 연구를 진행할 수 있다. 두 번째 후속 연구과제로는 엔트리 또는 다양한 AI플랫폼에서 제공하는 텍스트, 음성

및 영상인식 AI 도구를 사용하여 발생하는 데이터들을 수집하고 분석하는 데이터과학 교육 프로그램 개발 및 효과분석 연구도 가능하다고 판단된다.

참고문헌

[1] Ministry of Science and ICT(2017), The 4th Industrial Revolution in History, R&D KIOSK, 40.

[2] Ministry of Science and ICT(2017), The Various Aspects of the Fourth Industrial Revolution, the Realized Future, R&D KIOSK, 41.

[3] Kim, J.Y.(2016). Hello Data Science, Seoul : Hanbit Media.

[4] Kwon, J.K.(2020). Learning data science, Seoul : Jpub.

[5] Ichiro, U., Hiroaki, N., Masami, A. and Eichi, M.(2020). Learning Data science with Excel, Seoul : Hanbit Media.

[6] Ministry of Science and ICT(2019). Beyond an IT powerhouse to an AI powerhouse. Report Material, 2019. Dec.17.

[7] Joint ministries(2019). Artificial intelligence national strategy, Report Material, 2019.

[8] Jang, Y.J.(2017). Searching for the direction of data science education in the era of the 4th industrial revolution. *Integrated Humanities Research*, 9(10), 155-180.

[9] Park Y.S. and Lee, S.J.(2020). Study on the Direction of Universal Big Data and Big Data Education-Based on the Survey of Big Data Experts, *Journal of The Korean Association of Information Education*, 24(2), 201-214.

[10] Kim, Y.M.(2018). *A data science education program to improve computing thinking and creativity*, Jeju National University of Education, Ph.D thesis.

[11] Kim, Y.M. and Kim, J.H.(2017). Effect of data science education program using spreadsheet on improvement of elementary school computational thinking, *Journal of The Korean Association of*

Information Education, 21(2), 219-230.

[12] Wing, J. M.(2006). Computational thinking, *Communications of the ACM*, 49(3), 33-35.

[13] Wing, J. M.(2008). Computational thinking and thinking about computing, *Philosophical transactions of the royal society of London A: mathematical, physical and engineering sciences*, 366(1881), 3717-3725.

[14] Lee, J.S.(2013). A Study on Visualization Methods and Expressions of Information Design for Big Data, *Basic Formulation Studies*, 14(3) 261-269.

[15] Park, J.W. and Kim, H.Y.(2011). Artistic Data Visualization Review. *Digital Design Study*, 11(3), 194-202.

저자소개



허 경

1998년 고려대 전자공학과 학사
 2000년 고려대 전자공학과 석사
 2004년 8월 고려대 전자공학과 통신공학박사
 2004년 8월 ~ 2005년 8월 삼성종합기술원(SAIT) 전문연구원
 2005년 9월 ~ 현재 경인교대 컴퓨터교육과 교수
 관심분야: 피지컬컴퓨팅 교육, AI 교육, SW 및 AI 융합교육
 e-mail: khur@ginue.ac.kr