

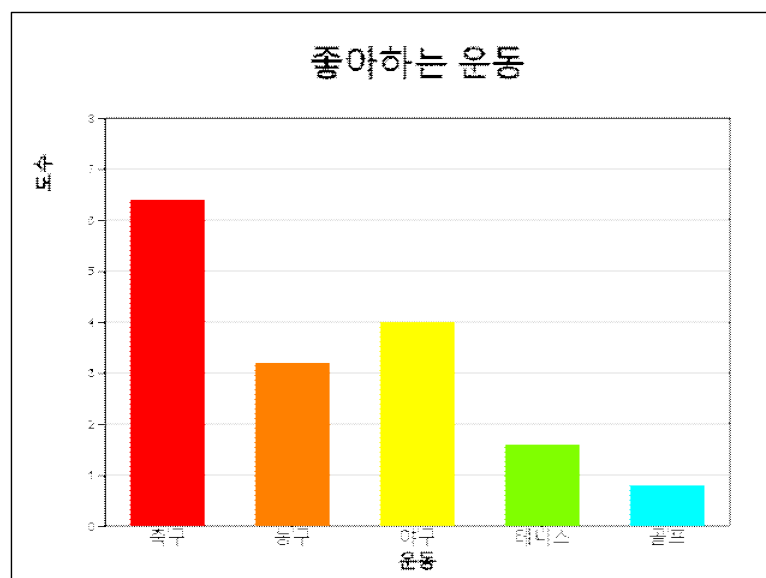
2. 질적 자료의 그래프

- 범주형 자료인 경우에 다음과 같은 그래프를 그려 분석한다
 - 막대그래프
 - 원그래프
 - 띠그래프
- 단어로 이루어진 문자 형태의 자료는 단어구름을 그려 분석한다.

2.1 막대그래프 - 원그래프 - 띠그래프

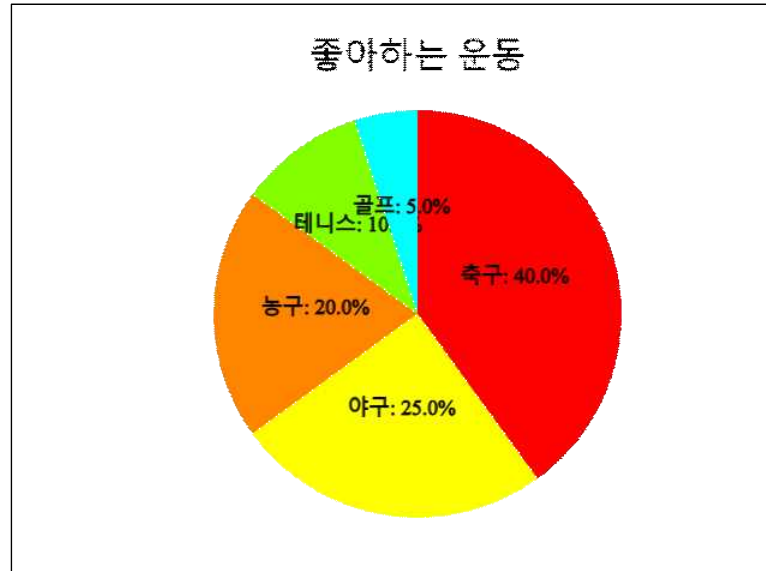
<p>☞ 생각열기</p>	<p>초등학교 한 학급 학생 20명을 대상으로 좋아하는 운동을 조사한 도수분포표가 다음과 같다.</p> <p style="text-align: center;">[표 2.1] 초등학교 학생의 좋아하는 과목의 도수분포표</p> <table border="1" style="margin: auto;"> <thead> <tr> <th>운동</th><th>학생수</th></tr> </thead> <tbody> <tr> <td>축구</td><td>8</td></tr> <tr> <td>농구</td><td>4</td></tr> <tr> <td>야구</td><td>5</td></tr> <tr> <td>테니스</td><td>2</td></tr> <tr> <td>골프</td><td>1</td></tr> </tbody> </table>	운동	학생수	축구	8	농구	4	야구	5	테니스	2	골프	1
운동	학생수												
축구	8												
농구	4												
야구	5												
테니스	2												
골프	1												
<p>탐구</p>	<p>좋아하는 운동이 어떠한 것인지 쉽게 관찰할 수 있는 그래프에는 어떠한 것이 있을까?</p>												

- 학생들에게 좋아하는 운동을 물으면 ‘축구’, ‘농구’, ‘야구’, ... 등의 대답을 얻을 수 있는데 이를 **질적 자료**라 한다. 대개의 질적 자료는 그 값이 범주를 의미하여 범주형 자료로 부르기도 한다. 질적 자료는 [표 2.1]과 같이 좋아하는 운동의 학생수를 세어 도수분포표로 만든 후 **막대그래프**, **원그래프**, **띠그래프** 등을 이용하여 그 특징을 살펴본다.
- **막대그래프**는 각 범주별 자료의 도수를 직사각형 막대의 높이로 나타낸 것이다. 범주형 자료라는 것을 강조하기 위하여 막대사이의 간격을 띄어 놓는다.



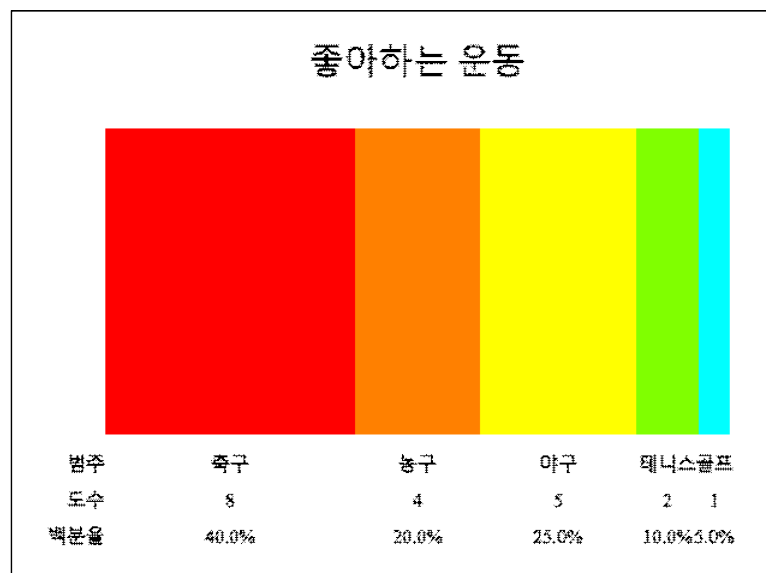
<그림 2.1> 좋아하는 운동의 막대그래프

- **원그래프**는 각 범주의 도수가 전체 자료에서 차지하는 비율을 원조각으로 나누어 표시한 그래프이다. 비율의 비교를 잘 할 수 있도록 시계 방향으로 가장 큰 범주의 원조각부터 순서대로 차례로 그린다.



<그림 2.2> 좋아하는 운동의 원그래프

- **띠그래프**는 원그래프의 변형 형태로 각 범주의 도수가 전체 자료 중에서 차지하는 비율을 사각형 조각으로 나누어 표시한 그래프이다. 범주의 값이 가장 큰 순서부터 정렬할 수도 있으나 원그래프와 구별하기 위하여 자료의 범주 순서대로 사각형 조각을 그린다.



<그림 2.3> 좋아하는 운동의 띠그래프

- 『eStat』 소프트웨어를 이용하여 위의 그래프를 그려보자.

실습 2.1

『eStatH』를 이용하여 좋아하는 운동에 대해 막대그래프, 원그래프, 띠그래프를 그리고 어느 운동을 많이 좋아하는지 관찰하라

풀이

- 왼쪽의 QR을 이용해 『eStatH』 메뉴에서 ‘막대그래프 - 원그래프 - 띠그래프’를 선택하면 <그림 2.4>와 같은 자료 입력창이 나타난다.
- ‘범주’와 ‘도수’를 입력하고 ‘주 제목’, ‘세로축 제목’, ‘가로축 제목’ 원하는 제목을 입력한다.
- [막대그래프] 버튼을 클릭하면 <그림 2.1>과 같은 막대그래프가 나타난다. 원하는 범주의 색을 선택하여 다시 [막대그래프]를 그릴 수도 있다. 제일 오른쪽에 각 범주별 아이콘을 선택하고 그래프 아래의 ‘이모티콘’을 체크하면 <그림 2.5>와 같은 그래프가 나타난다.



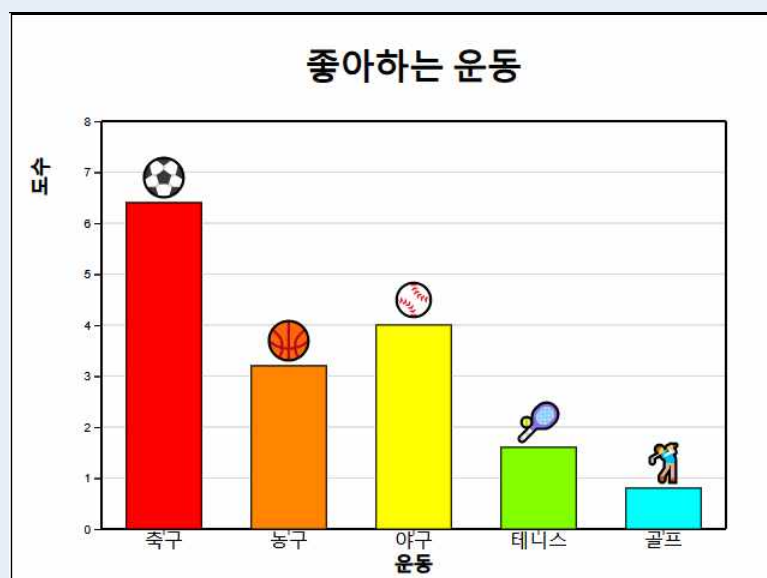
막대그래프 - 원그래프 - 띠그래프 메뉴

주 제목:

세로축 제목: 가로축 제목:

	범주	도수	상대도수	색상	이모티콘
1	축구	8	0.400		
2	농구	4	0.200		
3	야구	5	0.250		
4	테니스	2	0.100		
5	골프	1	0.050		
6	<input type="text"/>	<input type="text"/>	<input type="text"/>		
7	<input type="text"/>	<input type="text"/>	<input type="text"/>		
8	<input type="text"/>	<input type="text"/>	<input type="text"/>		
9	<input type="text"/>	<input type="text"/>	<input type="text"/>		
	합계	20	1.000	<input type="button" value="기본색"/>	

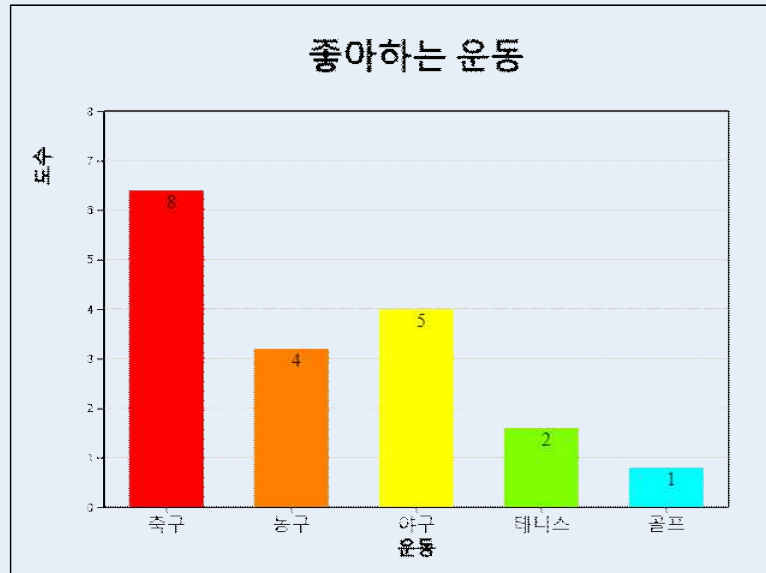
<그림 2.4> 좋아하는 운동의 자료 입력



<그림 2.5> 각 범주의 이모티콘이 표시된 막대그래프

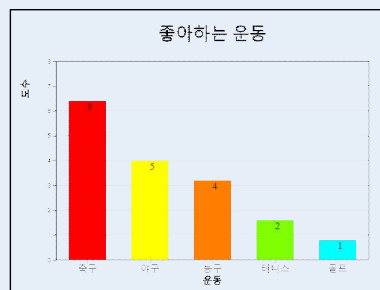
실습 2.1 풀이 계속

- 유사한 방법으로 막대 위에 <그림 2.6>과 같이 도수를 표시할 수도 있다.

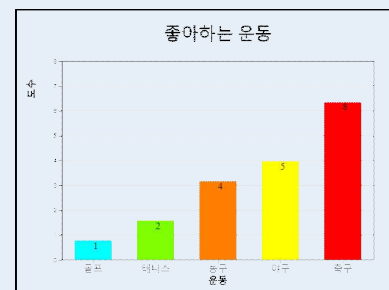


<그림 2.6> 도수가 표시된 막대그래프

- 그래프 아래 (정렬) 선택사항을 이용하여 <그림 2.7> <그림 2.8>처럼 막대를 도수의 내림차순 또는 오름차순으로 정렬할 수 있다.



<그림 2.7> 도수의 내림차순 막대그래프



<그림 2.8> 도수의 올림차순 막대그래프

실습 2.2

초등학생이 좋아하는 과일을 조사하였더니 다음과 같다. 『eStat H』을 이용하여 막대그래프를 그려보자.

[표 2.2] 초등학생들이 좋아하는 과일

과일	학생수
딸기	15
사과	10
바나나	25
포도	20
오렌지	10
토마토	2
수박	13
파인애플	5

실습 2.2 폴이

- 왼쪽의 QR을 이용해 『eStatH』 메뉴에서 ‘막대그래프 - 원그래프 - 띠그래프’를 선택하면 <그림 2.9>와 같은 자료 입력창이 나타난다.
- ‘범주’와 ‘도수’를 입력하고 ‘주 제목’을 입력한 후 제일 오른쪽에 각 과일 아이콘을 선택한다.
- [막대그래프] 버튼을 클릭하고 그래프가 나타나면 그래프 아래의 ‘이모티콘’을 체크한다. 원하는 범주의 색을 선택하여 다시 [막대그래프를] 그릴 수도 있다.



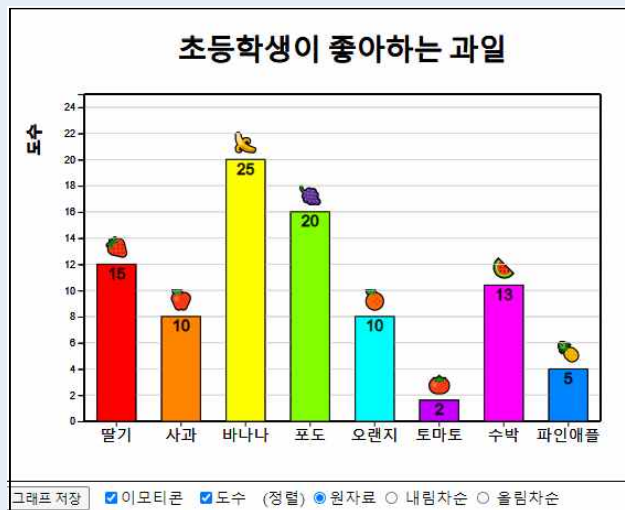
막대그래프 - 원그래프 - 띠그래프 메뉴

주 제목 : 초등학교가 좋아하는 과일
세로축 제목 : 도수 가로축 제목 :

범주	도수	상대도수	색상	이모티콘
1 딸기	15	0.150	Red	🍓
2 사과	10	0.100	Orange	🍏
3 바나나	25	0.250	Yellow	🍌
4 포도	20	0.200	Green	🍇
5 오렌지	10	0.100	Cyan	🍊
6 토마토	2	0.020	Purple	🍅
7 수박	13	0.130	Magenta	🍈
8 파인애플	5	0.050	Blue	🍍
9			기본색	🍌
합계	100	1.000		

막대그래프 원그래프 띠그래프

<그림 2.9> 좋아하는 과일의 자료 입력



<그림 2.10> 좋아하는 과일의 막대그래프

과제 2.1

2015년 우리나라의 8개 특별시 및 직할시에 대한 인구가 [표 2.3]과 같다. 『eStatH』를 이용하여 막대그래프, 원그래프, 띠그래프를 그려 특징을 알아보라.



[표 2.3] 시도별 인구수 (출처: 통계청 단위: 만명)

특별시 및 직할시	인구수
서울	990
부산	345
대구	247
인천	289
광주	150
대전	154
울산	117
세종	20

과제 2.2

어느 초등학교 학생들의 장래 희망직업을 조사한 요약 자료가 다음과 같다. 『eStatH』를 이용하여 막대그래프, 원그래프, 띠그래프를 그려 특징을 알아보라.



[표 2.4] 초등학생들의 장래 희망직업

직업	학생수
연예인	45
의사	30
교사, 교수	60
법조인	24
게이머	10
운동선수	43
경찰, 소방수	26

2.2 단어구름

<p>☞ 생각열기</p>	<p>다음은 4차 산업혁명에 대한 위키백과의 설명이다.</p> <p>[표 2.5] 4차 산업혁명에 대한 위키백과의 설명</p> <p>제4차 산업 혁명은 기술이 사회와 심지어 인간의 신체에도 내장되는 새로운 방식을 대표하는 디지털 혁명 위에 구축되고 있다.[10] 제4차 산업 혁명은 로봇 공학, 인공 지능, 나노 기술, 양자 프로그래밍, 생명 공학, IoT, 3D 인쇄 및 자율주행 차량을 비롯한 여러 분야에서 새로운 기술 혁신이 나타나고 있다.</p> <p>세계 경제 포럼 암시자 슈왈(Klaus Schwab)의 저서 《제4차 산업 혁명》에서 이 4번째 혁명이 기술 발전에 의해 특징지어졌던 이전의 3가지 혁명과 근본적으로 다른 점을 언급하고 있다. 이러한 기술은 수십억 명의 사람들을 계속해서 웹에 연결하고 비즈니스 및 조직의 효율성을 획기적으로 향상시키며 더 나은 자산 관리를 통해 자연 환경을 재생산할 수 있는 커다란 잠재력을 가지고 있다.[11]</p> <p>"제4차 산업 혁명 마스터 하기"는 스위스 Davos-Klosters에서 열린 세계 경제 포럼 연례회의 (World Economic Forum Annual Meeting 2016)의 주제였다.</p>
<p>탐구</p>	<p>과연 4차 산업혁명 설명에는 어떠한 단어가 중요할까?</p>

- 단어구름 (word cloud)은 문자 자료의 정보를 시각적으로 표현한 것이다. 단어구름은 주어진 문장에 나타나는 모든 단어들의 빈도수를 조사하여 많이 나타난 단어의 중요성을 글꼴 크기, 기울기 또는 색상으로 잘 보이도록 표시한다. <그림 2.11>은 위의 4차 산업혁명 설명에 대한 단어 구름이다.





<그림 2.11> 4차 산업혁명 설명에 대한 단어 구름

- 단어구름은 가장 두드러진 단어를 빠르게 인식하여 상대적인 중요도를 결정하는 데 유용하다. 예를 들면 정치 연설의 중요한 주제를 시각화 하든지, 소셜 미디어 소프트웨어에서 한 단어와 관련된 항목의 하이퍼링크를 결정하는 도구로 사용할 수 있다. 단어 구름의 중요 단어는 특정 웹 사이트와 관련된 검색 엔진 마케팅 용어로 사용되기도 한다.
- 단어구름 생성에는 많은 알고리즘이 있는데 eStat는 d3 오픈 소프트웨어의 알

고리즘을 채택했다. d3에는 아직 한글에 대한 알고리즘은 없어 현재 영어 문장에 대한 단어구름만 제공한다. 현재 단어구름은 단어 사이의 연관성을 표시할 수 있는 알고리즘 개발이 진행되고 있다.

<p>실습 2.3</p>	<p>『eStatH』을 이용하여 [표 2.5]의 4차 산업혁명에 대한 단어구름을 그려 관찰하라</p>
<p>풀이</p> <div data-bbox="339 779 523 969">  </div>	<ul style="list-style-type: none"> • 왼쪽의 QR을 이용해 『eStatH』 메뉴에서 ‘단어구름’을 선택하면 <그림 2.12>와 같은 자료 입력창이 나타난다. • 여기에 [표 2.5]의 문장을 복사한 후 [실행] 버튼을 클릭하면 <그림 2.11>과 같은 단어 구름이 나타난다. <div data-bbox="587 734 1369 1070">  </div> <p style="text-align: center;"><그림 2.12> 단어구름을 위한 자료입력</p>

과제 2.3	<p>다음은 통계학에 대한 위키백과의 설명이다. 『eStatH』를 이용하여 단어구름을 그려 관찰하라</p>
	<p>[표 2.6] 4차 통계학에 대한 위키백과의 설명</p> <p>통계학은 관찰 및 조사로 얻을 수 있는 데이터로부터, 응용 수학의 기법을 이용해 수치상의 성질, 규칙성 또는 불규칙성을 찾아낸다. 통계적 기법은, 실험 계획, 데이터의 요약이나 해석을 실시하는데 있어서의 근거를 제공하는 학문이며, 폭넓은 분야에서 응용되어 실생활에 적용되고 있다.[2] 통계학은 실증적인 뿌리를 가지고 있으며 실질적 활용에 초점을 맞추고 있기 때문에, 흔히 순수수학과는 다소 구분되는 응용수학의 일종으로 여겨진다. 통계학의 방법을 통해, 실제의 수치들을 왜곡하여 해석하는 것을 막고 연구를 바탕으로 합리적인 의사결정을 할 수 있다. 통계학은 과학, 산업, 또는 사회의 문제에 적용되며 모집단을 연구하는 과정이 우선시된다. 모집단은 "한나라 안에 사는 모든 사람" 또는 "크리스탈을 구성하는 모든 원자"와 같이 일정한 특성을 지닌 집단이면 어느 것이든 가능하다. 통계학자들은 전체인구(인구조사를 하는 기업)에 대한 데이터를 편집한다. 이것은 정부의 통계 관련 법률요약집같은 조직화된 방법으로 수행될 수도 있다. 기술통계학은 모집단의 데이터를 요약하는데 사용된다. 도수 및 비율(경주 등) 범주 형 데이터를 설명하는 측면에서 더 유용할 동안 수치 기술자는 연속적인 데이터 유형(소득 등)에 대한 평균과 표준 편차를 포함한다. 데이터 분석 방법 엄청난 자료가 연구되는 현대 사회에서 경제 지표연구, 마케팅, 여론조사, 농업, 생명과학, 의료의 임상연구 등 다양한 분야에서 응용되고 있는 통계는 단연 우리 사회에서 가장 필요하고 실용적인 학문이라고 할 수 있다.</p>

과제 2.4	<p>다음은 다음은 존 에프 케네디 미국 대통령의 취임사 중 일부이다. 이 취임사에 대해 단어구름을 작성하여 주요 단어를 분석하라.</p>
	<p>[표 2.7] 존 에프 케네디 미국 대통령의 취임사</p> <p>In the long history of the world, only a few generations have been granted the role of defending freedom in its hour of maximum danger. I do not shrink from this responsibility—I welcome it. I do not believe that any of us would exchange places with any other people or any other generation. The energy, the faith, the devotion which we bring to this endeavor will light our country and all who serve it—and the glow from that fire can truly light the world.</p> <p>And so, my fellow Americans: ask not what your country can do for you—ask what you can do for your country.</p> <p>My fellow citizens of the world: ask not what America will do for you, but what together we can do for the freedom of man.</p>