

Window of Statistics

2020
WINTER
Vol.26

| ISSUE |

통계로 바라보는 코로나 이후 세상

코로나 이후 사회생활의 변화와 통계 업무의 진화 방향
코로나가 던져준 과제, 신뢰성 있는 정보를 확보하라
코로나19 이후, 세계 경제는 어떻게 변해갈 것인가

| 통계광장 |

가까운 미래를 알려주는
나우캐스트가 세상을 바꾸고 있다
셀프서비스로 데이터 분석하기
기후 변화, 지구는 어디로 가고 있을까

| 통계탐방 |

업무 자동화 경진대회
다양한 통계업무 자동화 사례를 만나보자

| PEOPLE |

식당 데이터 경영자 박노진
알면 다르게 보이고, 보이면 길이 생긴다



CONTENTS

ISSUE

코로나 이후 사회생활의 변화와 통계 업무의 진화 방향 002

박성현 · 서울대학교 통계학과 명예교수, 한국과학기술한림원 전임 원장

코로나가 던져준 과제, 신뢰성 있는 정보를 확보하라 010

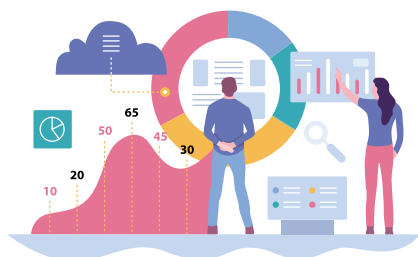
황명진 · 고려대학교 사회학 교수

코로나19 이후, 세계 경제는 어떻게 변해갈 것인가 016

김광석 · 한국경제산업연구원 경제연구실장



통계광장



가까운 미래를 알려주는 나우캐스트가 세상을 바꾸고 있다 024

윤지숙 · 통계청 빅데이터과장

데이터 리터러시: 셀프서비스로 데이터 분석하기 030

구자룡 · 밸류바인 대표

기후 변화, 지구는 어디로 가고 있을까 040

김준래 · 통계의 창 객원기자

통계집중탐구

범죄자를 추적하는 데이터: 사회연결망분석을 활용한 범죄수사 050

한민경 · 경찰대학교 교수



통계탐방

업무 자동화 경진대회 다양한 통계업무 자동화 사례를 만나보자 058

양병렬 · 통계청 기획조정관실 사무관

교육



R에 도전하자... 따라가다 보면, 나도 R유저 ⑧

심송용 · 한림대학교 데이터과학스쿨 교수

064

데이터 인포그래픽 강좌 series 11

정책홍보에 필요한 비주얼데이터 시각화 방법

이수동 · 한국인포그래픽협회 대표

076

데이터로 세상을 만나는 공간, 통계데이터센터

안향숙 · 통계청 통계데이터기획과 사무관

082

PEOPLE

식당 데이터 경영자 박노진

알면 다르게 보이고, 보이면 길이 생긴다

090



참가의 여유

이것만 알면 나도 유튜브!

스마트폰 속 사진을 동영상으로 만들기

정영국 · 디자이너

096

우리의 면역력을 좋게 만드는 힐링 여행

김여환 · 가정의학과전문의, 김앤권 대표원장

102

간추린 통계 소식

통계로 바라보는 세상 이야기

신동헌 · 통계의 창 객원기자

106

발행일 | 2020년 11월 30일

발행인 | 은순현

발행처 | 통계교육원

기 획 | 김정란, 최희정

주 소 | 대전광역시 서구 한밭대로 713(월평동) 통계센터 통계교육원

전 화 | 042-366-6151

팩 스 | 042-366-6498

이메일 | ccn420@korea.kr

디자인 | ALL contents group

인 쇄 | 올인피앤비(031-955-9497)

ISSN 2005-1379

©2020. 통계교육원

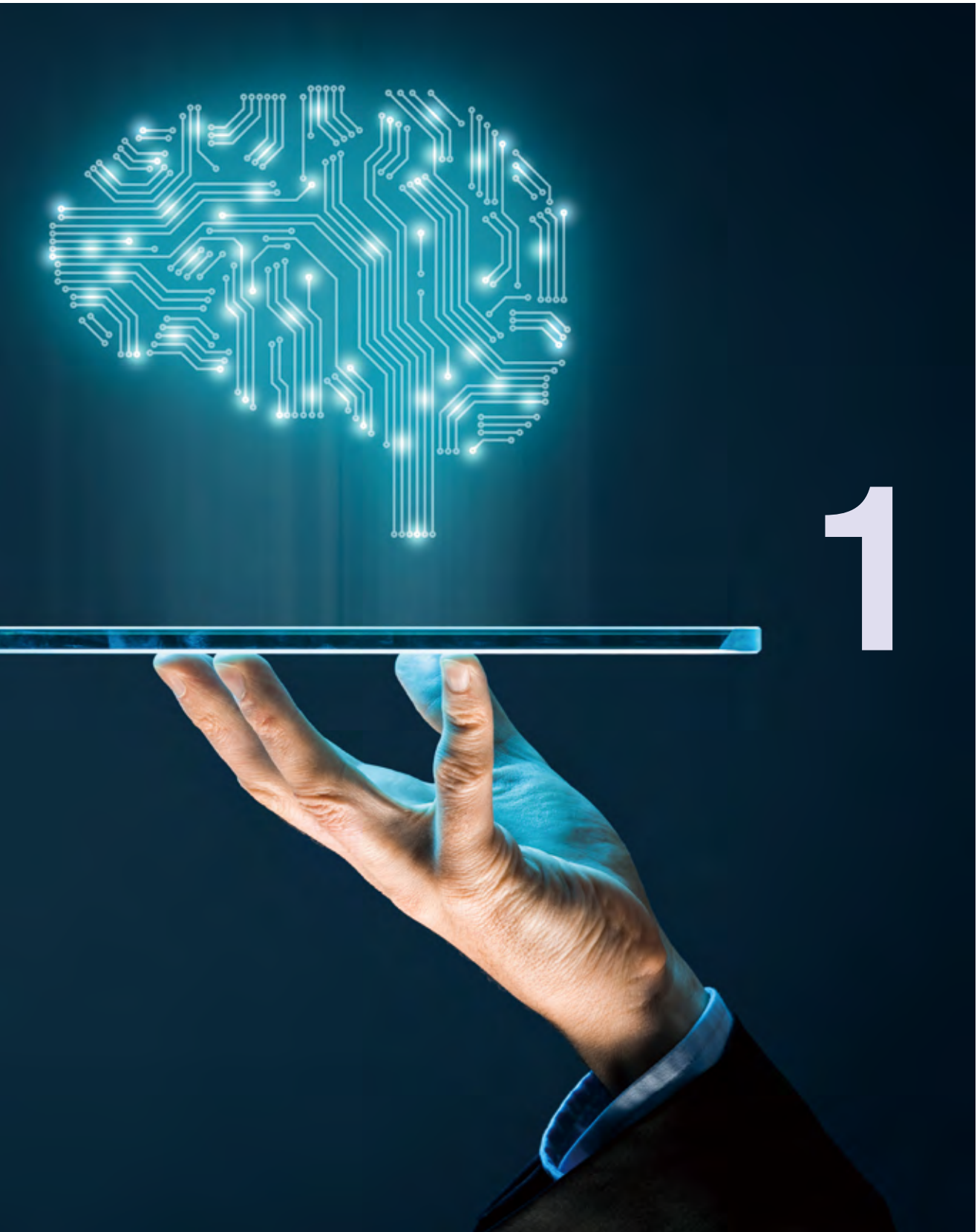
※ 『통계의 창』에 실린 내용은 필자 개인의 의견이므로 필자의 소속기관이나 본지의 공식적인 견해를 대변하는 것은 아닙니다.

코로나 이후 사회생활의 변화와 통계 업무의 진화 방향

지난해 12월 말 중국 우한(武漢)에서 시작된 ‘신종 코로나바이러스 감염증(코로나19)’ 전염 사태가 전 세계적인 팬데믹(대유행 전염병)으로 발전되면서 약 9개월 후인 10월 21일 현재 216개 나라에서 4,100만 명 이상이 감염되고 113만 명 이상이 사망했다. 이 엄청난 사건은 아직 그 끝을 모르고 진행되고 있다. 미국이 확진자 850만 명 이상, 사망자도 22만 명 이상으로 가장 피해가 크고, 다음으로 인도(확진자 760만 명 이상, 사망자 11만 명 이상)와 브라질(확진자 520만 명 이상, 사망자 15만 명 이상) 등에서 맹위를 떨치고 있다. 이 전염병을 막을 백신이 나오기까지, 앞으로 얼마나 더 지속될지 예측할 수 없다. 1918년에 발생한 스페인 독감(사망자 5,000만 명 이상)에 이어 거의 100년 만에 발생한 코로나19 팬데믹으로 인해 전 세계가 엄청난 충격을 받고 있다. 코로나19는 전 세계적으로 의료 보건의 위기를 가져왔을 뿐 아니라 우리 삶에 100년 만의 최대 충격을 안겨줄 것으로 예상된다. 이번 위기의 범위는 전 세계를 망라한다. 따라서 사람과 물자의 이동이 제한되면서 글로벌 공급망이 붕괴되고 있다. 역사적으로 고찰할 때 이러한 큰 충격은 인류의 삶의 방식을 바꾸어놓는다. 그러면 코로나19는 우리 사회를 어떻게 바꾸어놓고 있는가?

ISSUE

1



코로나가 바꾸는 삶의 방식들

우선 재택근무 방식이 급증하고, 사회의 디지털화가 가속화되며, 데이터의 수집·분석·활용 활동이 증가하고 있다. 코로나로 인한 사회적 거리두기, 비대면 업무, 화상회의 등의 증가로 노트북과 스마트폰으로 집에서 업무를 처리하는 재택근무가 급증한다. 이로 인하여 디지털 기술(빅데이터, 인공지능(AI), 사물인터넷 등)의 도입이 빨라지고 사회가 디지털화로 치닫고 있다. 디지털 기술에는 데이터의 활용이 필수적이며, 데이터의 수집, 분석, 활용 등에서 앞서가는 나라가 결국 디지털 기술에서 앞서가게 될 것이다.

두 번째로, 사람들이 외부 활동을 자제하고 집에서 머무는 시간이 늘면서 '인터넷 온라인 쇼핑'이 증가하고, '스트리밍 서비스 콘텐츠'를 이용하는 사람이 증가하고 있다. 또한 많은 대중을 상대로 하는 비즈니스 플랫폼 기업들(아마존, 구글, 알리바바, 네이버, 카카오 등)의 가치가 올라갈 것이며, 데이터에서 새로운 가치를 창출하는 새로운 비즈니스가 생겨나는 등 산업 생태계에도 변화가 있을 것이다.

세 번째로, 원격 교육이 활성화되어 교육 방식에 변화가 생기고 있다. 인터넷에 의한 온라인 원격수업이 확대되고 있다. 이로 인해 지방과 수도권 대학 구분이 무의미해지고, 교수 수요가 감소하며, 지역사회에 대한 대학의 사회교육이 활발해지는 등 대학 구조조정을 촉진할 가능성이 커지고 있다. 일반 사회인을 대상으로 하는 각종의 오프라인 교육도 온라인 교육으로 그 비중이 재편될 것이며, 일반인들이 집에서 각종의 원하는 교육을 받을 수 있는 형태로 바뀌어가고 있다.

마지막으로 산업에서는 스마트공장 전환이 빠르게 진행돼 4차 산업혁명이 가속화될 것이다. 산업현장에서는 사람이 적게 일하는 디지털 워크플레이스(digital workplace) 개념이 도입되면서 AI 기술, 챗봇(채팅 기술), 빅데이터, 태그정보기술, 5G, 가상·증강현실 기술 등이 발전한다. 이에 필요한 데이터와 정보를 순각적으로 활용하는 스마트공장이 빠르게 도입되고, 모든 기업 업무에서 4차 산업혁명의 기술들이 급속하게 접목될 것이다.

결국 코로나 이후의 사회변화상에서 중요한 것은 데이터의 활용 기술이 모든 분야에서 급속히 발전하게 될 것이라는 사실이다.

데이터 · AI 경제 시대의 도래에 따른 데이터 컨트롤타워 설치 필요

우리는 '데이터 경제' 시대에 살고 있다. '데이터 경제'란 개념은 유럽연합(EU) 집행위원회가 발표한 '유럽 데이터 경제 육성책(Building a European Data Economy, 2017)'에서 제시한 것으로, 데이터의 활용이 모든 산업의 발전과 새로운 가치 창출에 촉매 역할을 담당하는 시대의 경제라는 뜻이다. 이 육성책의 골자를 보면, EU 내에서의 데이터의 자유로운 공개와 사용을 보장하고, 데이터 분석 역량을 강화해 새로운 비즈니스 모델을 발굴하는 데 각국이 역량을 강화하도록 권장하고 있다.

또한 지금은 데이터 시대이면서 동시에 AI의 시대이다. 모든 분야(과학, 산업, 의료, 문화, 게임 등)에서 AI와 접목된 기술이 발전하면서 AI 기술의 발전은 한 국가의 역동적인 발전에 가장 중요

한 요소가 되었다. 따라서 우리는 ‘데이터·AI 경제 시대’에 살고 있다고 볼 수 있다.

우리 정부에서도 이를 인식하고 2019년 1월 16일 ‘데이터·AI 경제 활성화 계획; 혁신성장 전략투자’를 발표하면서 대처하고 있다. 또한 정부는 올해 7월 14일 ‘한국판 뉴딜 국민보고 대회’를 열었다. ‘한국판 뉴딜’은 ‘디지털 뉴딜, 그린 뉴딜, 안전망 강화’로 구성되며, 이 중에서 디지털 뉴딜에서는 2025년까지 데이터댐을 구축하여 공공데이터 14.2만 개를 개방하고, AI 학습데이터 1,300종을 마련하며, 정부도 AI 정부로 가겠다는 의지를 밝혔다.

이런 정부의 움직임에서 가장 중요한 핵심은 데이터의 축적과 활용이다. 행정안전부에서 운영하는 공공데이터포털(<https://www.data.go.kr>)이 있는데, 여기에는 3.2만 개 정도(2020년 6월 9일 현재)의 파일 데이터가 공개되어 있다. 또한 통계청의 통계데이터센터(<https://data.kostat.go.kr>)에서도 통계청 생산 데이터와 일부 민간자료를 모아 이용자들에게 서비스하고 있다.

우리나라는 현재 통계법에 의하여 통계청이 국가통계의 컨트롤타워 기능을 하고 있으나 분산형 통계 시스템을 운영하고 있어 수없이 많은 통계작성기관(승인통계 작성기관만 425개)을 가지고 있다. 따라서 각 기관에서 생산되는 통계데이터의 종합적인 관리, 표준화 및 통합, 개방, 활용 등에서 미흡한 것도 사실이다. 데이터·AI 경제 시대에 특히 중요한 민간 부분에서 생산되는 데이터의 관리의 중진국 수준을 못 벗어나고 있다.

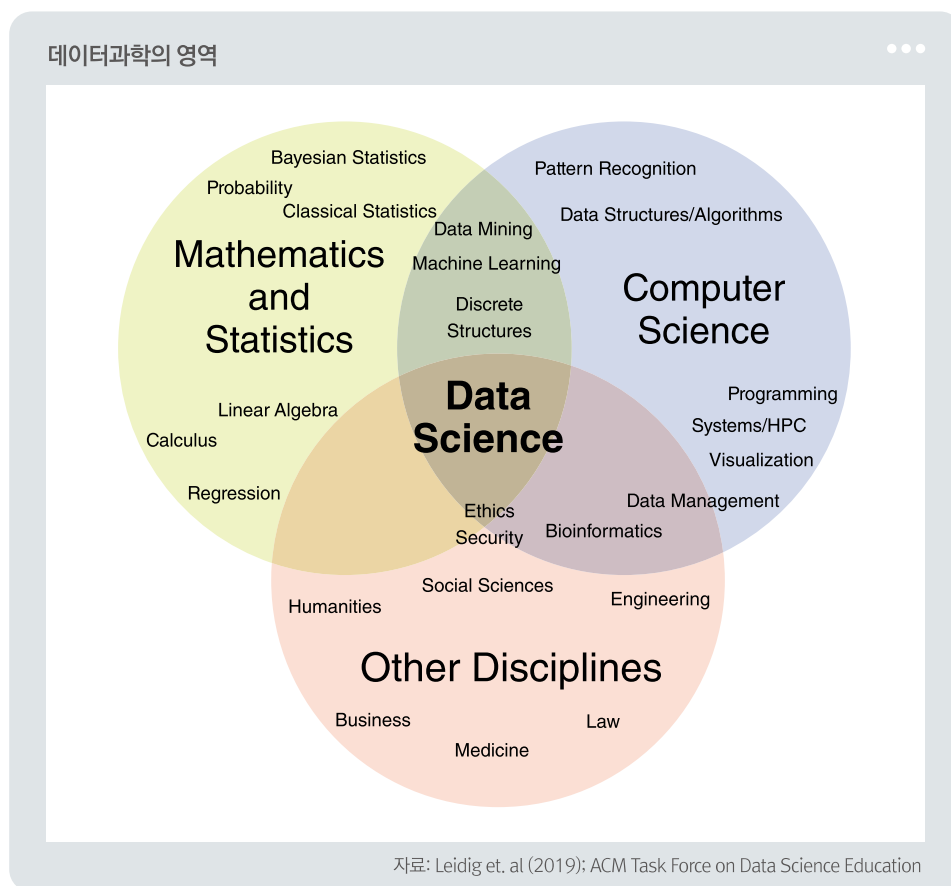


따라서 우리 사회의 모든 분야(공공, 민간, 개인 등)와 관련된 데이터의 생산·관리·개방·유통·활용 등을 종합적으로 기획하고 운영하며, 데이터 거래를 위한 인적·물적·제도적 기반 조성을 전담하면서 데이터·AI 경제 활성화의 효율적인 중심 역할을 담당할 조직으로서의 데이터 컨트롤타워 설립이 시급하다.

데이터 · AI 경제 시대에 가장 각광을 받고 있는 학문은 데이터과학

데이터·AI 경제 시대에 가장 각광을 받고 있는 학문은 데이터과학(data science)이다. 데이터과학은 데이터의 수집과 저장에 필요한 데이터 프로세싱 기술과 데이터 분석에 관한 지식을 기반으로 다량의 데이터로부터 패턴을 찾아내고, 통계적 추정, 예측 모델링 등을 통하여 필요한 정보를 창출하며, 이를 실제로 활용하는 것을 연구하는 융합과학으로, 빅데이터가 하나의 중요한 응용기술이다. 데이터과학에 대해서 좋은 참고문헌은 박성현 외(2018)와 한림연구보고서(2019)를 참조하기 바란다.

넓은 의미의 데이터과학은 데이터 입력으로부터 통찰력 있는 정보를 담은 결과물을 산출하기까지의 전 과정을 연구하는 학문으로 인공지능도 여기에 포함되며, 아래 그림은 그 연구영역을 보여주고 있다.





기존 학문의 영역에서는 수학, 통계학, 컴퓨터과학 및 기타 학문분야(사회과학, 공학, 경영학, 의학, 인문학 등)이며, 데이터과학은 이들을 아우르는 다학제 간 연구(interdisciplinary studies)를 하는 학문이다. 수학·통계학과 컴퓨터과학이 합쳐지는 부분에 있는 데이터 마이닝(data mining), 기계학습(machine learning), 이산구조(discrete structures) 등은 인공지능의 핵심적인 방법론이기도 하다.

즉, 데이터과학은 데이터로부터 수학·통계학과 컴퓨터과학의 이론을 기반으로 응용 분야인 기타 학문 분야의 지식(이를 도메인 지식(domain knowledge)라 칭함)을 활용하여 정보를 창출해내는 새로운 과학인 것이다.

사회생활상 변화에 따른 새로운 통계작성 방식의 대두

코로나로 사회생활 양상이 변화하고, 데이터과학의 발전으로 우리는 차츰 새로운 데이터·AI 경제 시대에 접어들고 있다. 이 새로운 시대에 통계청에서 주로 하고 있는 국가통계 작성 방법에도 어떤 변화가 생길 것인가?

중장기로 볼 때에는 반드시 변화가 있을 것이 확실하며, 이런 변화와 관련하여 김정민 외(2020)과 전영일 외(2020)는 이미 귀중한 연구보고서를 발표하였다. 또한 박재현(2020)은 국가통계 작성에 AI를 활용하는 두 개의 통계청 프로젝트를 진행 중에 있다고 보고하였다.

하나는 국가통계포털(KOSIS)에 AI 챗봇을 도입하여 이용자를 위한 서비스 혁신 방안을 강구 중이며, 또 하나는 산업/직업 분류 자동코딩 시스템의 한계(정확도 하락으로 인한 수동 코딩 비율 증가)를 극복하기 위해 기계학습 기법을 활용한 고도화를 추진 중이다. 새로운 통계 작성 방식의 대두에 대하여 다음과 같은 변화가 예상됨을 지적하고자 한다.

① 비대면 통계조사의 확대

코로나로 인하여 사람이 직접 만나지 않고도 일을 처리하는 언택트(untact) 비즈니스가 확장 되면서, 통계작성에서도 통계조사원이 직접 사람을 방문하여 대면 작성하는 조사통계의 비중이 줄어들고, 비대면 통계조사의 비중이 확대될 것이다. 예를 들면, 인구총조사에서 이미 일부 온라인 조사가 시행된 바 있으나, 이러한 비대면 온라인 조사의 비중이 증가할 것이다. 데이터 3법(개인정보보호법, 정보통신망법, 신용정보법)이 개정됨에 따라 개인의 가명정보는 통계작성이나 빅데이터 분석에 사용 가능하게 되었고, 이는 비대면 통계조사의 확대에도 도움을 줄 것이다.

② AI·빅데이터 기술을 활용하여 기존의 통계 생산 프로세스의 혁신

AI와 빅데이터 기술의 통계 작성 접목은 먼저 기존의 조사통계 생산 과정에 적용될 것이다. 통계 생산 과정에서 그 효율성과 품질 향상, 시간 단축 등을 위하여 AI와 빅데이터 기술이 활용될 것이다. 예를 들면, 데이터의 품질을 높이기 위해 이상치(outlier) 검색, 데이터 정제(cleaning), 무응답 대체(imputation), 설문 응답에 대한 자동 코드 분류 작성 등에 데이터 마이닝 기법이나 기계학습·심층학습 기법을 통한 AI가 적용되는 것이다. 이런 활동을 ‘AI·빅데이터를 활용한 통계 생산 프로세스의 혁신’이라고 말할 수 있다.

③ AI·빅데이터 기술을 활용한 신규 통계의 작성

다음 단계로는 기존의 통계 작성 틀에서 벗어나 ‘AI·빅데이터 기술을 활용한 신규 통계’를 만들어보는 것이다. AI나 빅데이터는 과거의 데이터로부터 현재나 미래를 예측할 수 있으며, 이러한 예측된 추정값을 실제 통계값으로 사용하여 통계를 작성하는 방법이 사용될 수 있다. 이는 아직 AI·빅데이터 기반의 분석이나 예측 결과(생산된 통계)에 관한 신뢰가 보편적으로 형성되지 않았으므로 국가통계로 승인하기에는 갈 길이 멀다. 그러나 정부에서 인정한 ‘AI 표준수립기관’에서 ‘AI 통계 품질 기준 및 심사 가이드라인’을 만들고, ‘AI 기술 표준화’에 성공한다면 결국 이러한 신규 통계도 승인 통계로서 그 역할을 할 수 있는 날이 올 것으로 예상된다.



외국에서는 AI·빅데이터 기술을 활용한 통계 작성이 차츰 늘어나고 있다. 영국 통계청은 3년 전에 ‘데이터과학국’을 출범시키고, 국가통계 작성에서 데이터과학을 활용하여 GDP 조기지수 추정, 위성 이미지 패턴 분석에 기초한 도시지역 산림 측정, 다양한 경제·사회적 변화 예측 등을 하고 있다.



독일 통계청은 구인·구직 데이터 사이트에 노출된 빅데이터를 기계학습을 활용하여 일자리를 추정하고 있다. 미국 인구 통계국은 온라인 데이터를 연계하여 기계 학습을 활용한 물가통계를 추정하고 있다. AI·빅데이터를 활용한 이러한 선진국의 통계 작성 사례들은 우리에게 귀중한 벤치마킹 대상이 되고 있다.

미래는 데이터과학자가 되어야 제대로 일할 수 있는 시대

현재 코로나가 극성을 부리고 있지만 이도 언젠가는 종식될 것이다. 그러나 코로나가 우리 사회에 준 충격은 크며 한국을 포함하여 인류의 사회생활에 큰 변화가 예상된다. 확실한 것은 사회가 데이터 기반의 디지털화로 빠르게 변화할 것이라는 점이며, 4차 산업혁명의 대표적인 기술인 빅데이터, AI 등이 빠르게 이 사회에서 사용될 것이다. 통계의 중요성은 차츰 더욱 증대될 것이며, 이 사회는 데이터·AI 경제 시대로 접어들게 될 것이다. 이에 따라 학문으로서의 데이터과학이 대학에서 중요한 학문으로 자리 잡을 것이며, 이 사회에서 데이터과학자들의 활동 영역이 넓어질 것이다.

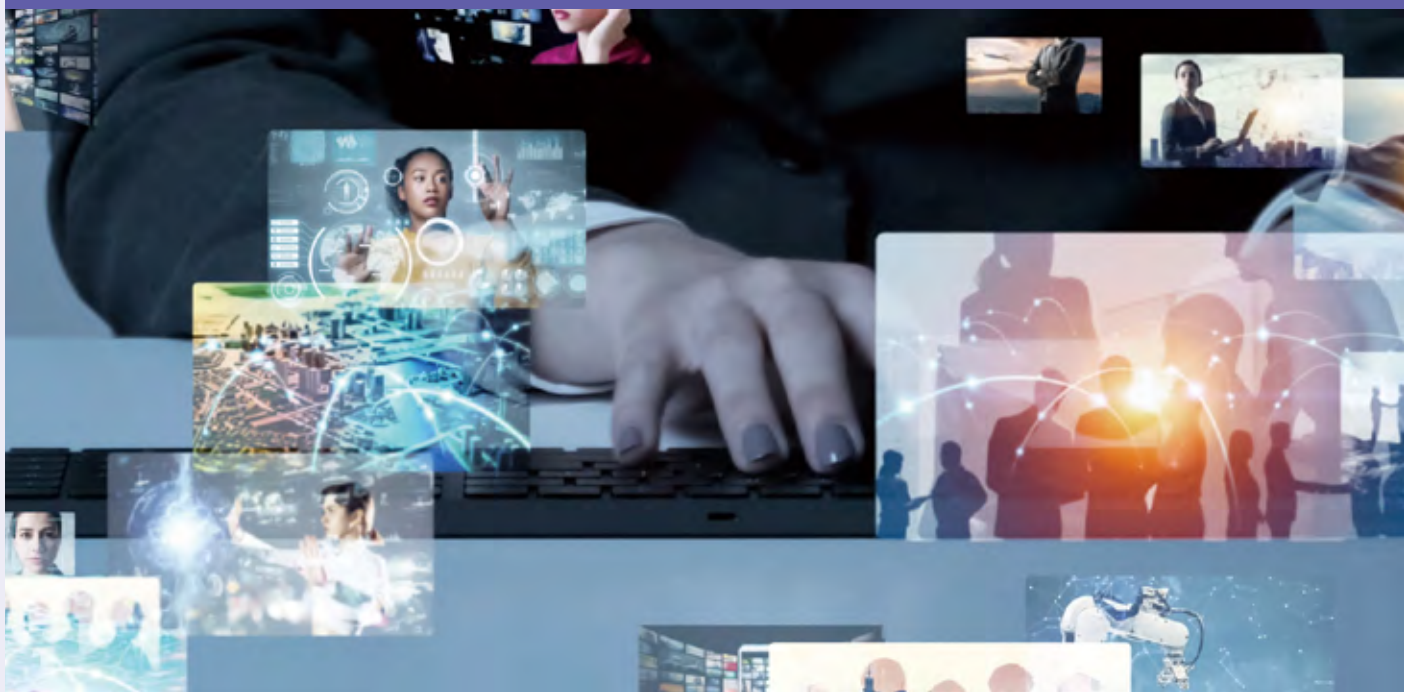
통계청과 같은 통계 작성기관에서도 빅데이터와 AI 기법(기계학습, 심층학습, 챗봇 등)들이 차츰 그 활용도를 증가시킬 것이며, 통계 전문가들도 앞으로는 데이터과학자가 되어야 제대로 일할 수 있는 시대가 올 것이다. 통계 업무에 종사하는 모든 분이 이러한 변화 추세를 읽고 미래를 대비해주기를 당부하고 싶다.

참고 문헌

1. 김정민, 전이슬, 황유림 (2020); AI 기술의 국가통계 활용 사례 및 국내 도입 촉진 방안, 소프트웨어정책연구소 ISSUE REPORT I, 2020. 6. 26.
2. 박성현, 박태성, 이영조 (2018); 빅데이터와 데이터 과학, 한국과학기술한림원 「석학, 과학 기술을 말한다」 시리즈 30, 자유아카데미 발행, 2018년 11월 15일.
3. 박재현 (2020); 머신러닝 기반 국가통계 적용 사례, 제10회 국가통계방법론 온라인 심포지엄 발표, 2020년 9월 24일, 통계청 통계센터 국제회의실.
4. 전영일, 김혜란, 임경민 (2020); 인공지능(AI)과 데이터 사이언스, 국가통계의 도약, 계간 KISTAT 통계플러스 여름호, 통계청 통계개발원 발행.
5. 한림연구보고서 (2019); 빅데이터·인공지능 산업 진흥을 위한 데이터 과학의 발전 전략 연구, 한림연구보고서 130, 한국과학기술한림원 발행, 2019년 12월.
6. Leidig, P., Danyluk, A., Cassel, L. and Servin C. (2019); ACM Task Force on Data Science Education: Draft Report and Opportunity for Feedback, Conference Paper of the 50th ACM Technical Symposium, February 2019.



2 코로나가 던져준 과제, 신뢰성 있는 정보를 확보하라





지구촌의 위기를 몰고 온, 코로나

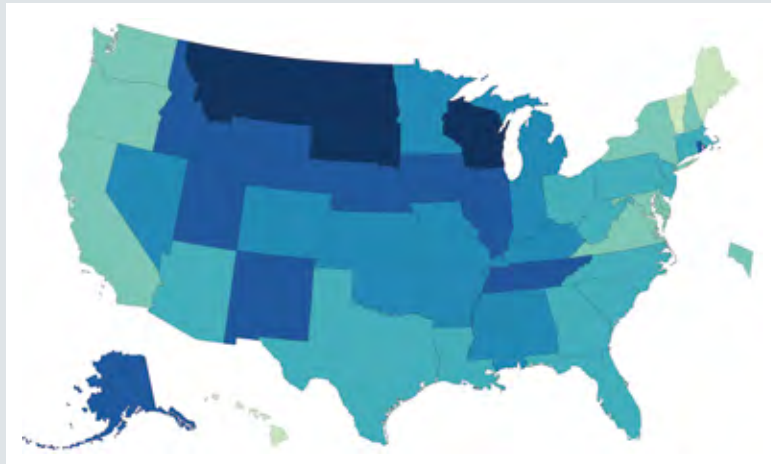
코로나의 시대다. 작년 말, 중국에서 폐렴환자들이 연이어 발생하였고, 의료진은 환자들의 호흡기에서 이전과는 다른 바이러스를 발견했는데, 이 바이러스는 아주 빠른 속도로 중국 전역을 거쳐 이웃 한국, 일본, 이탈리아 그리고 전 세계로 전파되었다.

국제보건기구(WHO)에서는 이를 코로나바이러스감염증으로 명명하였고, 2020년 3월 11일에는 1968년도 홍콩독감, 2009년 신종플루에 이어, 사상 3번째로 전염병 정보 6단계 중 최고의 위험 수준인 팬더믹(pandemic, 대유행)을 선언하였다.

2020년 10월 30일까지 약 113개국에서 약 44백만 명이 감염되었고, 29백만 명이 완치되었으며, 117만 명이 사망하여 치명률 2.65%를 기록하였다. 지금까지는 미국의 피해가 가장 커서 8,752,794명의 환자가 발생했고, 그중 225,985명의 사망자가 발생하였으며, 10월 말을 기준으로 하루 74,410명의 신규확진자가 발생하고 983명이 사망하고 있다.

지역적으로는(아래 그림), 초기에 급속도로 변졌던 뉴욕이나 대도시보다는 오히려 뉴멕시코, 노스다코타, 사우스다코타 주와 같이 소도시나 농촌지역 혹은 인구밀집이 덜한 지역에서 강하게 전파되고 있다.

미국 질병관리국 코로나19 발병보고서(10만 명당)



US COVID-19 Case Rate Reported to the CDC in the Last 7 Days, by State/Territory (cases per 100K)



코로나에 효과적으로 대응할 수 있었던 비결, 신속한 정보



한국은 이미 뉴스를 통해 널리 알려진 바와 같이, 초기 감염지역이었음에도 발병 케이스나 사망자가 극히 적은 편이다. 총감염자는 10월말 기준 26,146명이며, 사망자 463명으로 치명률이 1.6% 수준에 그쳤다.

초기에 국경봉쇄와 같은 강력한 대응에 실패한 비난이 있었지만 이후 전 세계에 소개되었던 드라이브스루(Drive-Thru) 검역소와 같은 체계적인 K-방역과 함께, 중증감염자나 기저질환이 있는 감염자를 중심으로 집중적인 의료보전 시스템을 가동한 의료조치로 치명률(Case Fatality Rate: CFR)을 지속적으로 낮추는 성과를 거두었다.

이 같은 성과에는 국제검역과 치료에 참여했던 보건복지부, 질병관리청, 의사, 간호사, 복지사를 비롯한 의료진의 헌신적인 공로가 있었음을 우리는 잘 알고 있다. 방역과 치료에 가장 필수적인 요소가 정확한 정보다. 우리나라를 비롯한 각국의 중앙질병방역사무소는 질병에 대한 연구와 함께 정확한 통계자료를 수집해서 국제기구(WHO)에 보내고, 동시에 자료를 통계적으로 작성하여서 지역에 배포하였다. 지역이나 국가는 정확하고 신속한 정보를 바탕으로 방역 및 보건 의료를 하였고, 더 나가서 정책적인 대응을 하였다.

한국의 성공적인 방역사례가 단순히 우리끼리 선부르게 자화자찬하는 것으로 볼 게 아닌 것이, 우리나라 질병통계의 작성과 분석 및 배포의 수준이 매우 체계적이고 국제적인 수준에 이르렀다는 것은 지난 사스, 메르스, 에볼라와 같은 전염병 전파 시기에도 알려진 사실이다.



주요 언론 매체들이 가짜 뉴스 생산으로 혼란 키워

재난 시기의 미디어는 평소보다 영향력이 커진다. 사재기 현상은 어느 나라에서든 예외 없이 SNS나 기존의 뉴스 매체를 통해 전파되고 있다. 정보를 제공하는 것이 미디어의 임무인데, 정확한 정보는 그것이 어떤 가치를 지향하고 있느냐에 상관없이 긍정적인 영향을 갖는다.

가짜뉴스는 당연히 그렇지 않다. 흔히 SNS의 발달로 가짜뉴스의 전달력이 커졌다고 이야기들 한다. 하지만 최신 연구에서는 그렇지만도 않다고 한다. 오히려 정부나 기존의 레거시 언론 매체도 가짜뉴스를 만들고 있다. 제대로 된 보건정책연구소의 보고서도 과학적인 실증적 검증을 거치지 않고, 마구잡이로 미디어와 SNS를 통해서 전파되고 있다. 문제는 당장은 어떤 게 가짜뉴스인지 명확치 않다는 것이다.



모든 나라가 소위 '보건의학' 통계에 있어 높은 수준의 신뢰할 수 있는 정보를 배포하는 것은 아니다. 이번 코로나 사태에서 WHO와 같은 국제보건기구의 역할에 실망한 이유는 정치적인 역학, 특히 중국을 옹호하는 듯한 태도만이 아니라 국제적인 전염병 통계수집과 정보의 전달에 있어서 구글과 같은 신종 미디어와의 경쟁에서 밀렸기 때문으로 본다. 각국의 보건당국은 질병관리본부를 통해 수집된 자료를 공표하는데, 구글이 이를 수집하고 지역적인 전파나 분포와 같은 정보를 제공하는 데 탁월한 효과를 발휘했다.

Facebook과 같은 사적·공적인 정보가 망라되는 SNS에서는 가짜뉴스나 검증되지 않은 연구결과가 과학학술지에 등재된 논문이라 하더라도 검열되었다. 이처럼 정부나 보건당국 혹은 국제기구가 갖는 권위는, 비록 레거시 언론의 지원을 받는다 하더라도, SNS와 구글에 상당히 밀리고 있는 실정이다. 실제로 최근의 한 연구에서는 신문이나 방송매체와 같은 레거시 언론이 SNS보다 더 부정적이고 자극적인 기사를 다루었다고 한다.

미디어의 역할과 관련해서 최근 주목받는 단어가 있는데 그것은 인포데믹(Infodemic)이라는 용어이다. 영어로 인포메이션(information)과 질병(epidemic)의 합성어인데, 그 의미는 코로나와 같은 질병에 대한 정보가 매우 빠르게 전파되는 것을 의미한다. 뉴스나 SNS를 통해 퍼지는 질병에 대한 정보는 정확하기도 하고 부정확하기도 한데, 진실을 파악하기 어려운 상황에서 가짜뉴스, 소문, 그리고 공포가 혼재되어서 퍼져나가기 때문에 정작 필요로 하는 사람들이 좋은 정보와 가짜뉴스를 선별하기가 매우 어려운 상황이 조장된다.

참고로, 옥스퍼드 영어 사전에서는 이를 'An excessive amount of information about a problem that is typically unreliable, spreads rapidly, and makes a solution more difficult to achieve'로 정의하여 부정확하고 신뢰할 수 없는 정보가 신속하게 전파되어 문제해결이 힘들어지는 상황을 의미하고 있다.

잘못된 정보 중에는 "세계의사협회에서 코로나19가 단순 감기라고 선언하였다"라든지 "94%의 확진자가 오진이었다"라는 내용이 포함되어 있다. 보다 심한 가짜뉴스 중에는 특정 국가가 코로나바이러스를 인공적으로 배양했고 고의로 퍼뜨렸다, 혹은 한국인이 코로나바이러스에 선천적인 내성이 있어서 치명률이 매우 낮다"라는 내용도 있다.





부정확한 정보의 세 가지 유형

언론전문가들은 부정확한 정보 소위 가짜뉴스를 크게 세 가지로 구분한다: mis-information, disinformation, 그리고 mal-information이다.

특별한 악의 없이 사진이나 지도와 같은 정보를 잘못 인용하는, 오인에 의한 정보의 왜곡을 mis-information이라고 하고, 특정 국가나 인종, 종교집단에 대한 위해를 목적으로, 다시 말해 악의를 가지고 고의로 가짜뉴스를 만들거나 배포하는 것을 dis-information이라고 한다.

mal-information은 비록 사실에 기반한 정보이지만, 디스인포메이션과 마찬가지로 특정 집단에 위해를 가할 수 있는 정보다.

가장 위험한 정보는 악한 목적으로 고의로 만들어내는 디스인포메이션이라 할 수 있다.

흔히 가짜뉴스는 신문이나 공중파 방송과 같은 정통언론, 혹 레거시 뉴스매체 보다는 개인의 사적인 정보교환이 많은 SNS를 통해 많이 전파된다고 생각할 수 있다. 하지만 해외의 연구에 의하면, 코로나 초기 페이스북, 트위터와 같은 SNS에 비해 레거시 언론은 상대적으로 파괴적이고, 자극적이면서 동시에 공포심을 조장하는 뉴스를 집중적으로 생산, 전파하였고 특정 종교집단이나 지역에 대한 거짓되고 왜곡된 정보를 제공하는 데 기여했다고 한다.

위기 상황에서의 질병에 대한 정확한 정보는 비단 방역과 치료에만 도움이 되는 것이 아니다. 사실에 근거한 믿을 만한 정보는 공포심과 불안감을 제거하고 생존을 위한 이기적인 행동을 선택하지 않는 데 상당한 효과가 있다. 서양의 식품이나 휴지 사재기가 한국에서 발생하지 않았던 이유는 바로 정확하고 믿을 수 있는 정보가 개방적으로 제공되었기 때문이다.





코로나를 해결하기 위해 신뢰성 있는 정보유통 필요

언론 매체의 중요성은 새삼 강조할 필요가 없을지 모른다. 기존의 뉴스 미디어뿐 아니라 인터넷 뉴스포털, SNS가 매체와 미디어의 역할을 경쟁적으로 수용하고 있는 상황에서 정작 중요해지는 것은 진실된 팩트, 즉 뉴스의 내용이다.

가짜뉴스는 애초에 신뢰하지 않는 사람들에게는 다가가지 않는다. 거짓되고 왜곡된 정보를 믿으려는 의지와 관심이 있는 사람들에게, 특별히 불신과 의심의

고리를 풀고자 하는 노력 없이 쉽게 수용되고 전파된다. 결국

가짜뉴스는 하나의 정치이념으로 뭉친 진영이나 문화

종교적인 쿨트 집단을 중심으로 만들어지고 퍼져

나가기 때문에, 반대편 진영에서 제아무리 부정

해봤자 소용없는 것이다.

해결책은 Re-information이다. 비록 시

간이 지났다고 하더라도 정확한 정보를 다시

수요자에게 제공해야만 한다. 성급한 결론을

도출하거나 돌이킬 수 없는 판단을 유보하는

것이 당장은 최선이다.

개개인의 역할도 중요하다. 검증되지 않거나 사

회에 위해가 된다고 판단되는 정보가 전달되면 강력

하게 이의를 제기해야만 한다. 재차 동일한 뉴스에 대한 수

용을 거부하거나 전달자에게 항의해야만 한다. 과편적이고 부정확

한 정보로 중요한 정책의사결정을 내려서는 안 된다. 특히나 팬데믹 초기의 위기감으로 인해 완결되고 신뢰성 있는 정보가 부재한 상황에서 선택적으로 취득한 정보를 기반으로 정형화된 인식적인 틀(cognitive schema)을 만들었다면 개인과 집단 그리고 정부는 이를 특별히 조심해야 한다.

우리나라 중앙방역대책본부와 질병관리처는 코로사19 사태를 기회로 방역과 지역보건 체계의 정비와 함께 정확한 통계정보의 작성과 배포뿐 아니라 가짜뉴스에 대한 검색에도 상당한 노력을 해오고 있는 것으로 알고 있다. 중앙통계작성 기관이 통계청 역시 기존의 통계작성과 부가된 활동을 전개하면서 방역당국을 돕고 있는 것으로 알고 있다.

하지만 방역과 치료가 어느 한 정책부서의 역할이 아니듯 재난 시 정확한 통계 작성과 발표 역시 범정부적인 협력이 필요하다. 그 중심에 전문가들이 모여 있는 통계청이 있었으면 하는 바람이다.



3

코로나19 이후, 세계 경제는 어떻게 변해갈 것인가

비 온 뒤에 땅이 굳는 법이다. 소 잃고 외양간을 고친다고 비판하지만, 소를 잃었기 때문에 외양간을 고쳐야 할 필요성을 갖게 되는 것이다. IMF 외환위기를 겪은 후, 외환 건전성을 확보하고 금융시장에 대한 감독 기능을 강화해오고 있다. 일본의 수출규제라는 뼈아픈 경험 후에 소재·부품·장비 국산화를 정책적으로 추진하게 되었다. 위기에 처해본 자가 위기 극복 방법을 알지 않겠는가?

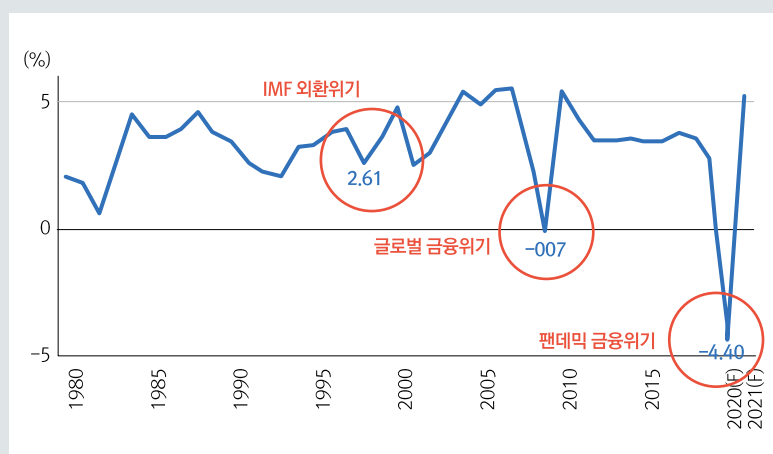


ISSUE

2020년 코로나19와 2021년 뉴노멀(New Normal)

2020년을 살아가는 사람들은 1930대 대공황 이후 가장 흉악한 경제위기를 경험했다. IMF의 2020년 세계 경제성장률 전망치는 -4.4%다. 1930년대 대공황 이후 가장 낮은 세계 경제성장률이다. IMF 외환위기 당시 세계 경제는 2.6% 플러스 성장했고, 글로벌 금융위기 당시 세계 경제성장률은 -0.07%로 거의 0%에 가까운 마이너스 수준이었다. 2020년 팬데믹 경제위기는 그만큼 가혹한 위기였다.

주요 경제위기와 세계경제성장률 추이



자료: IMF

주: 2020년 세계 경제성장률은 2020년 10월 기준 전망치임



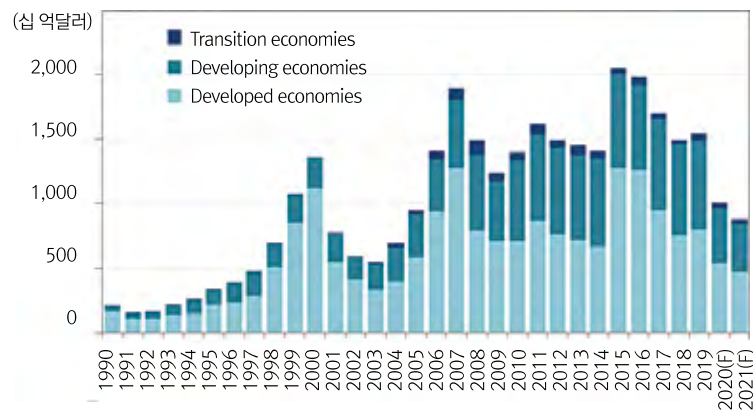
충격이 지나가고 나면 전과 후는 달라지곤 한다. 일시적 변화가 아니라 구조적 변화가 나타나는 것이다. 필자는 신간 『포스트 코로나 2021년 경제전망』을 통해 2021년의 구조적 변화를 조망했다. 예를 들어, 보건에 관한 국민적 필요와 정책적인 강조가 한층 증가할 것이다. 코로나19가 종식될지라도 다른 전염병이 다시 찾아올 우려를 불식시킬 수 없을 것이다. 사회·문화적으로도 낯선 군중들과 어울리는 활동들에 의식적·무의식적으로 부정적 감정을 갖게 될 것이다. 그밖에도 비대면 서비스나 디지털 플랫폼에 대한 수요가 지속해서 늘어나다가 코로나19의 영향으로 수요가 폭발적으로 증가한 이후, 다시 이전 수준으로 돌아가는 것이 아니라 이전 수준보다 한층 높은 수준으로 자리하게 될 것으로 전망된다. 즉, 코로나19는 일시적 변화만이 아니라 구조적 변화를 가져올 것으로 보인다.

코로나19가 가져올 구조적 변화

첫째, 글로벌 벨류 체인(Global Value Chain, 이하 'GVC') 상의 변화가 가속화될 것으로 보인다.

보호무역주의가 팽배해지고, 미국을 비롯한 주요국들이 리쇼어링 정책을 추진함에 따라 이미 제조업 회귀현상이 전개되어왔다. 세계적으로 해외직접투자 유입액(Foreign Direct Investment Inflow)은 2015년 20,418억 달러 규모를 기록한 이후 급속도로 감소하면서 2018년에는 14,952억 달러를 기록했다. 2020년들어 해외직접 투자는 큰 폭으로 감소하고, 2021년까지 이런 추세가 두드러지게 나타날 것으로 전망된다.

세계 해외직접투자(FDI Inflow) 추이



자료: UNCTAD(2020) 『World Investment Report 2020』

주: Transition economies는 UN에서 분류하고 있는 기준으로, 선진국(Developed economies)과 개발도상국(Developing economies)을 제외한 나라들이고, 경제시스템이 근본적으로 상이한 시스템으로 이행되는 과정에서 있는 경제권을 뜻함. 보통 사회주의 경제체제에서 시장경제로 이행하는 동유럽 및 CIS 지역들을 의미함.

2020년 세계 주요국들이 코로나19 사태를 경험하면서 지역 간의 봉쇄(lock down)와 시설들의 폐쇄(shut down)가 본격화되었고, 주요 산업들은 주요 부품의 공급에 차질을 빚은 바 있다. 한국도 자동차, 전자, 철강, 석유화학 등의 주요 산업들이 중국, 미국, 유럽 등으로부터 원자재와 부품의 조달에 문제를 겪은 바 있다. 특히 많은 선진국은 그동안 마스크를 비롯한 위생용품들의 제조기지를 신흥국에 두었기 때문에 전염병 대응에 상당한 차질이 빚어지기도 했다.

코로나19 사태는 GVC상의 일부 부문을 해외에 의존하기보다 자국에 집중하는 현상을 강화시킬 것으로 보인다. 한국 정부는 주요 정책들 중 하나로 리쇼어링 정책(U턴 기업 지원정책)을 강화하고 있다. 해외 현지 법인들을 본국으로 회귀시켜 경제성장과 고용창출에도 기여할 수 있도록 하는 자국 우선주의가 확대되는 것이다. GVC가 약화되고 LVC(Local Value Chain)이 강조되는 구조적 변화가 나타나는 모습이다.

둘째, 디지털 트랜스포메이션(Digital Transformation)이 가속화될 전망이다.

코로나19 이전부터 아날로그에서 디지털 환경으로의 변화가 가파르게 진전되고 있다. 코로나19는 이러한 변화를 앞당겨 놓는 것이다. 글로벌 기업들을 보면, 디지털 트랜스포메이션을 이끄는 기업들을 중심으로 축의 이동이 일어나고 있다. 세계 경제의 핵심축이 '전통산업 → ICT 제조 → ICT 서비스'로 이동하고 있는 모습이다. 시가총액을 기준으로 글로벌 Top 10 기업들의 분포를 확인해 보면, 1990년대 ICT 기업은 IBM, General Electric, BT Group 등 3개에 불과했다.



2000년대에는 4개의 ICT 기업이 Top 10 기업에 들어섰고, 2020년에는 Apple, Amazon, Microsoft, Alphabet, Facebook, Alibaba 등의 ICT 기업들이 주를 이루게 되었다. 특히 ICT 산업 중에서도 ICT 제조에서 ICT 서비스로의 변화가 눈에 띈다. 모든 변화가 코로나19로 인한 것은 아니었지만, 코로나19가 경제주체들의 디지털 서비스에 대한 요구를 폭발시켰고, 이에 적극적으로 대응한 기업들을 중심으로 주요 산업들이 재편되고 있는 것이다.

글로벌 Top 10 기업(시가총액 기준)

여타산업		IT 제조업	IT 서비스업	
	1990년	2000년	2016년	2020년
1	IBM	Itautec	Apple	Apple
2	Exxon Mobil	General Electric	Alphabet	Amazon
3	General Electric	Exxon Mobil	Microsoft	Microsoft
4	Altria Group	Pfizer	Berkshire Hathaway	Alphabet
5	Toyota	Cisco	Exxon Mobil	Facebook
6	Bristo-Myers Sqibb	Citi	Amazon	Alibaba
7	Merck&Co.	Walmart	Facebook	Bershire Hathaway
8	British Petroleum	Vodafone	Johnson&Johnson	Visa
9	Walmart	Microsoft	JPMorgan Chase	Taiwan Semiconductor
10	BT Group	AIG	General Electric	Johnson&Johnson

자료: Bloomberg

주: 시가총액은 매년 12월 31일 기준, 2020년은 8월 13일 작성일 기준

셋째, 비대면 서비스(Untact service)에 대한 의존도가 높아질 전망이다.

‘사회적 거리두기’가 보편화됨에 따라 온라인쇼핑과 게임 서비스 수요가 급증했다. 과거 비대면 서비스는 짧은 층의 전유물로 여겨졌지만, ‘반강제적으로’ 소비자 전체로 확산했다. 온라인쇼핑이 급증하면서 지급결제서비스가 고도화되고, 온라인 교육 및 화상회의가 도입되면서 ZOOM과 같은 플랫폼 사용자가 급증했다.

오프라인 환경하에서 조차도 키오스크(Kiosk)가 빠른 속도로 보급됨에 따라 판매원 등을 만나지 않고 서비스를 이용하는 비대면 환경으로 전환되고 있다. 홈트(홈트레이닝)와 커넥티드 게임의 수요가 증가하고, 앱으로 배달음식을 즐기면서 ‘홈노미’라는 용어가 등장하기도 했다. 디지털 플랫폼을 경험해본 사용자들은 편리성과 유용성을 인식해 코로나19 이후 더욱 의존적이 될 것으로 판단된다.



키오스크를 활용한 비대면 주문과 지급결제



자료: Taco Bell

넷째, 사회·문화도 상당 부분 바뀔 것으로 보인다.

테니스를 즐기는 저자는 동호인 테니스 풍경이 바뀌었음을 실감한다. 테니스 경기는 통상적으로 네트 앞에서 상대방과 악수를 하며 시작하고, 경기 중 파트너와 손바닥을 마주치며 응원한다. 코로나19는 이러한 접촉을 최소화한 채 라켓을 부딪치며 인사하고, 응원하는 방식으로 전환되게 만들었다. 엘리베이터에서 대화를 나누지 않는 에티켓도 자리 잡게 되고, 기침할 때 팔꿈치를 이용하는 교육도 확대되었다.

한국을 상징하던 집단주의(collectivism) 문화도 쇠퇴하며, 개인주의(individualism)로의 전환이 가팔라질 것으로 보인다. 대가족에서 핵가족 제도로 변화하고, 1인가구가 주된 가구 유형으로 등장하며, 아파트형 주거공간이 확대되면서 개인주의적 문화로 전환되고 있었다. 코로나19는 이러한 문화적 변화를 가속화시킬 것으로 전망된다.

마지막으로 조직문화도 마찬가지다. 대면보고와 대면회의가 최소화될 것으로 보인다.

재택근무도 효율성을 잃지 않는 범위 내에서, 더욱 적극적으로 활용될 것으로 전망된다. 비대면 보고와 재택근무가 원활히 이루어질 수 있도록 하는 ERP(Enterprise Resource Planning, 전사적자원관리) 시스템 보급도 확대되고 고도화될 것이다. 비대면 회의의 편리성을 경험한 기업들은 이러한 환경에 적합한 시스템과 소프트웨어를 구축하고 이를 더욱 활용해 나갈 것으로 전망된다.

회식문화가 상당 비중 줄어들고, 유연근무제도가 안착할 것으로 보인다. 이미 밀레니얼 세대가 기존 세대들과 함께 일하기 시작한 직장환경에서 서로의 라이프스타일을 존중하고 2차-3차로 이어지는 회식문화를 지양해왔던 터에 코로나19는 이러한 변화를 증폭시켰다. 최근 G7 재정부 장관 회담, G20 정상회담, IMF 정기 미팅 등과 같은 국제 주요 회담들이 화상으로 진행되었고, 이는 향후 공공 및 민간 조직 내에서 실시간 화상 회의 및 교육의 활용이 늘어날 가능성을 알리는 역할을 했다.



거스를 수 없는 변화와 대응


외양간을 어떻게 고칠지를 고민해야 할 때다. 또다시 소를 잃지 않도록 말이다. 기업들은 코로나 이후 펼쳐질 변화를 그려보고, 그 변화를 선도할 수 있는 제품과 서비스를 제공해야 하겠다. 제조기업들은 GVC의 변화를 인식하고, 생산기지 이동 및 다변화를

검토하며, 정책적인 지원들을 활용해야 한다. 디지털 플랫폼을 확보하고, 아날로그 방식으로 전달하던 서비스를 디지털·비대면 방식으로 전환해 나가야 한다.

사회문화 및 조직문화의 변화에 걸맞은 경영 시스템을 구축하는 것도 중요하다. 개인들도 그런 변화에 기초해 유망한 영역으로의 진로를 설정하고, 요구되는 역량을 개발하기 위해 노력해야 한다. 투자 의사결정 면에서 이러한 변화를 반영하는 것은 너무나 당연한 일일 것이다. 정부는 코로나 19의 경제적 충격을 복구할 수 있도록 하는 정책적 노력을 집중하고, 기업과 가계가 발 빠르게 변화에 대응해 나갈 수 있는 환경을 구축해야 한다.







가까운 미래를 알려주는 나우캐스트가 세상을 바꾸고 있다

2006년 영국의 수학자이자 기업가인 클라이브 험비(Clive Humby)가
“데이터는 새로운 석유다(Data is the new oil)”라고 선언한 지 10여 년이 지나고 있다.
과연 클라이브의 말처럼 오늘날 우리는 데이터가 원유인 시대에 살고 있는 것일까.



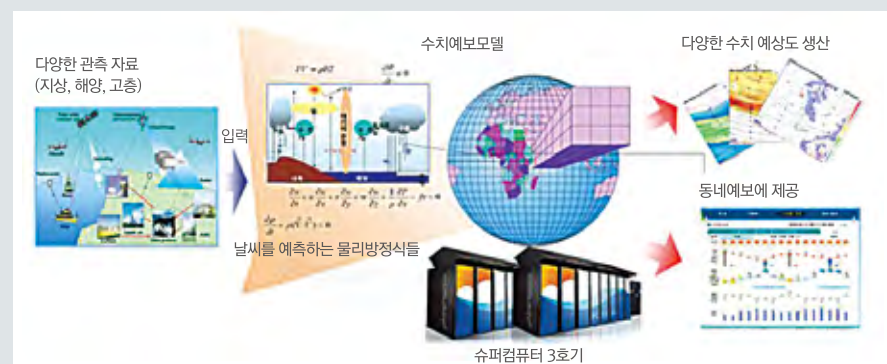
우리 삶 속에 스며든 데이터 활용



요즘 같은 환절기에는 아침에 일어나서 가장 먼저 확인하는 것 중에 하나가 바로 날씨앱일 것이다. 낮 기온은 얼마가 오를지, 혹시 퇴근 무렵에 비가 오지나 않을지, 주말 날씨는 어떻게 바뀔지 살펴보고 우산이나 여분의 옷을 준비한다. 그런데 일기예보만 믿고 외출 준비를 했다가 당황스러운 일을 겪은 경험이 누구나 한번쯤 있을 것이다. 지구온난화로 지구촌 곳곳에서 발생하는 이상기후로 인해 기상예측이 더욱 어려워지고 있다.

날씨를 정확히 예측하는 것은 인류의 오랜 꿈이었다. 30여 년 전만 해도 기상예측은 예보자의 경험을 기반으로 한 주관적 예보방식을 사용했었으나, 1990년대에 들어서면서 보다 객관적인 수치예보의 중요성이 대두되었다. 즉, 기온, 바람, 기압과 같은 기상변수들을 시공간적으로 수집하고 이를 슈퍼컴퓨터 기반의 수치예보모형에 적용하는 날씨 예측방식을 이용하게 되었다.

수치예보 생산과정(기상청 제공)



예보정확도를 결정하는 세 가지 주요 요소는 수치예보모델, 데이터와 인적 역량이다. 기상청 연구에 따르면 각각의 요소별 기여도는 수치예보모델성능 40%, 관측자료의 양과 질 32%, 예보관역량 28% 순이라고 한다. 즉, 모델과 데이터가 차지하는 비중이 전체 72%나 된다.

최근에는 빅데이터와 인공지능을 적용해서 예측력을 높이기 위한 연구가 진행되고 있어 이 수치는 점점 더 높아질 것이다. 올해 초 구글은 동(洞) 단위보다 작은 가로세로 각 1km 내 국지기상현상을 최대 3시간 전에 정확히 예측하는 AI 예보기술을 개발했다고 한다. 바로 기상예측모델 나우캐스트(Nowcast)이다.

구글의 나우캐스트는 U-Net 머신러닝 아키텍처(신경망(Convolutional Neural Networks, CNN) 아키텍처)를 이용하고 있으며, 159개 관측소에 설치된 넥스라드(NEXRAD) 레이더 안테나를 통해 데이터를 수집하고, 고스(GOES)-16, GOES-17의 미국 기상 위성 데이터를 사용하고 있다고 한다. 이러한 시스템을 통해 예측소요 시간이 5~10분 수준으로 매우 짧아, 뇌우나 태풍 등 급박한 기상 재난에 대한 위기 대응에 큰 도움이 될 것으로 내다보고 있다.



약 하루 동안의 예보를 시각화한 영상(구글 제공)



매시간 정각마다 수행한 1시간 동안의 미국 기상청 예보(1시간은 미국 기상청 예보 주기의 최대값임)



구글이 예보하기 위해 시도하고 있는 지상 실측 결과



구글 모델이 수행한 예보 (구글의 예보는 2분마다 이뤄지며 공간 해상도는 미국 기상청의 약 10배 수준)

이렇듯 우리 삶에 크고 작은 영향을 주는 기상예측 분야에서 빅데이터가 몰고 올 변화를 우리는 몸소 체험하고 있다고 할 수 있다.

가까운 미래를 예보하는 나우캐스트(Nowcast)로 발전

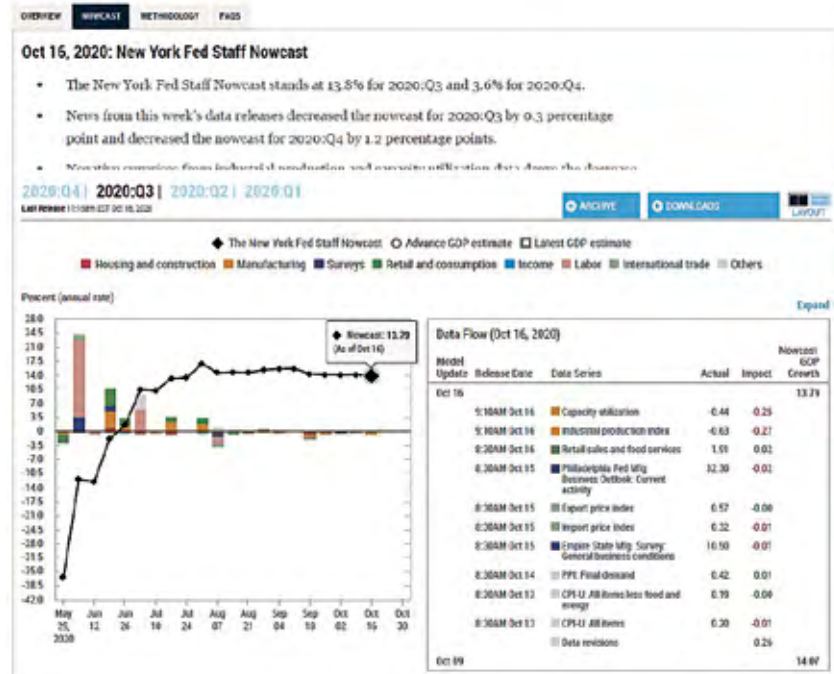


여기서 Nowcast란 단어에 주목해보자. ‘현재’를 뜻하는 Now와 ‘예측’을 뜻하는 Forecast의 Cast를 결합하여 만든 단어로, 우리말로는 ‘실태예보’ 혹은 ‘현황예측’ 정도로 해석할 수 있을 것이다. 다시 말해, Nowcast의 개념은 현재 혹은 아주 가까운 미래를 예측하는 것이다.

Nowcast는 국가통계 분야에서도 활발하게 사용하고 있는데 대표적으로 GDP 추정에 활용되고 있다. GDP(국민소득통계)는 실물 경제활동의 성과를 종합적으로 보여주는 대표적인 거시 경제지표로, 정책당국뿐 아니라 기업, 가계 등의 경제 주체들도 경제적 의사결정을 하는 데 빈번하게 사용하고 있다. 통상 분기 단위 경제지표를 사용해서 분기말 이후 30일에서 두 달 이상의 시차를 두고 발표하던 기존 방법과는 달리 Nowcasting을 통해 월별, 주별 혹은 실시간 변수를 활용한 GDP 성장률을 추정할 수 있다.

미국 뉴욕연방준비은행의 Nowcasting Report

Nowcasting Report



미국 뉴욕연방준비은행은 입수 가능한 모든 경제지표를 수집하여, 예측변수를 실시간으로 업데이트하고, 이를 기반으로 GDP성장률 추정치도 실시간으로 보정하여, 시장 참여자에게 홈페이지 통해 공유하고 있다.

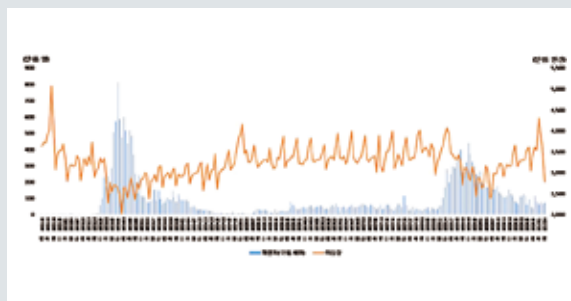
나우캐스트(Nowcast) 정보 공급자로 거듭나고 있는 통계청

우리나라 통계청에서도 궁극적으로 사회·경제 분야에서의 Nowcast를 구현하기 위해 다양한 빅데이터를 활용해서 국가 통계를 생산하는 시도를 하고 있다. 통계청 빅데이터통계과가 2015년 10월에 신설한 이래로, 온라인 물가작성, 임금근로자 및 개인사업자 가계부채 통계, 모바일 인구이동 분석을 위해 웹 스크래핑 데이터, 모바일 통신 데이터, 금융거래 데이터 등을 이용하고 있다.

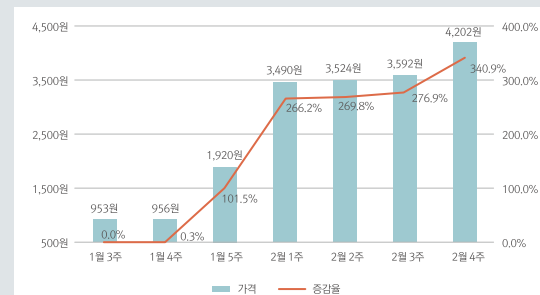
특히 올초 국내에 코로나19 확진자 최초 발생 이후 전국적인 확산의 위기에 놓여 있던 시기에, 코로나19 발생 전·후 인구이동 패턴을 모바일 빅데이터로 분석하여 중앙재난안전대책본부를 비롯한 관련 부처와 언론에 제공한 바 있었다. 그 밖에도 코로나 발생 초기, 마스크의 품귀 현상과 가격 급등이 사회적 이슈가 되었던 시기에는 마스크 등 코로나 19 예방품목의 온라인 판매가격 추이를 웹 스크래핑 기술을 이용해서 일별로 분석하여 관련 부서와 공유하기도 하였다.



코로나19 확진자 수와 모바일 인구이동량



KF94 마스크 온라인 판매가격



이를 통해, 국가 긴급 재난 상황에서 정부의 선제적 대응 지원을 위한 속보성 통계지표의 역할과 가치를 입증할 수 있었다. 통계청은 이에 머물지 않고 한 단계 발전한 '성장기'로의 발돋움을 위해 센서 데이터, 로그 데이터 등 이용 빅데이터 유형을 다양하게 하고 사회·경제의 여러 분야로 활용 범위를 확대하기 위해 지속적으로 노력할 것이다.

데이터 리터러시: 셀프서비스로 데이터 분석하기



데이터가 풍부한 세상이 되었다. 내가 원하든 원하지 않든 우리 앞에는 이미 수많은 데이터가 있다. “앞으로는 석유나 금이 아니라 데이터가 지구상에서 가장 값진 자산이 될 것이다”라고 트렌드 분석가인 마크 펜은 『마이크로트렌드X』에서 주장한 바 있다. 우리 속담에 ‘구슬이 서 말이라도 꿰어야 보배’라는 말도 있다. 아무리 좋은 재료가 많이 있어도 다듬고 정리하지 않으면 가치가 없다는 것을 비유적으로 한 말이다. 널려 있는 데이터를 쓸모 있도록 만들어 보배가 되도록 하려면 무엇을 어떻게 해야 할까? 먼저 데이터를 수집하고 정리하고 분석하여 해석하면 된다. 구슬을 꿰듯이 데이터 분석을 통해 새로운 가치를 만들면 가장 값진 자산이 된다. 이 일을 누가 해야 할까?

현업 실무자가 스스로 분석하는 시대에 접어들고 있다

2015년 가트너(Gartner, Inc.)에서 발표한 「신기술 하이프 사이클(Hype Cycle)」에서 ‘시민 데이터 과학(Citizen Data Science)’이 처음 등장한 후 데이터 분석에 대한 현실적인 관심이 높아지기 시작했다.

전문 데이터 과학은 과학적 방법, 프로세스, 알고리즘 및 시스템을 이용하여 데이터에서 가치를 도출한다. 데이터 과학자는 데이터 분석을 위한 전략 개발, 데이터 준비, 데이터 탐색, 분석 및 시각화, R 및 파이썬(Python)과 같은 프로그래밍 언어를 사용하여 모델을 구축한다. 이런 전문 데이터 과학은 전문적인 학습과 숙련이 되지 않은 현업 실무자가 할 수 있는 업무가 아니다.

하지만 현업에서는 수시로 데이터를 활용해야 하는 이슈가 등장한다. 이럴 때마다 전문 데이터 과학자에게 요청하고 결과를 기다리는 데 많은 한계가 노출되면서 ‘시민 데이터 과학자(Citizen Data Scientist)’에 대한 관심이 높아지고 있다. 이는 수학 또는 통계에 대한 깊은 지식은 없지만 내가 속한 업(domain, field, 산업)에 대한 지식을 바탕으로 데이터 과학의 원리를 일부 적용하여 업무를 하는 사용자를 지칭한다. 가트너는 수년간 시민 데이터 과학자 수가 정규 데이터 과학자보다 5배가량 빠르게 증가할 것으로 전망했다.

전문가가 분석하던 업무 중 일부를 현업 실무자가 스스로 분석하는 시대에 접어든 것이다. 현업에 있는 우리는 모두 데이터 과학자까지는 아니더라도 적어도 데이터 분석가는 되어야 한다. 국내 상황에 맞추어 표현하면, 현업에서 구슬을 꿰어 보배로 만드는 ‘현업 데이터 분석가’가 절실히 요구되는 시점이다. 현업에서 스스로 데이터 분석을 통해 통찰을 얻고 바로 실행에 옮기는 실무자들이다. 이것이 가능하게 된 것은 셀프서비스 분석 도구들이 많이 등장했기 때문이다.

데이터를 읽고 쓰는 리터러시 능력, 즉 데이터를 다루는 힘을 기르는 가장 빠른 방법은 데이터 분석을 통해 스스로 경험을 쌓는 것이다. 운전 연습을 많이 하면 할수록 운전 실력이 향상되는 이치와 같다. 현업 실무자의 셀프서비스 데이터 분석이 필요한 이유다. 현업 실무자 모두가 데이터 분석가가 될 수 있다. 운전면허증만 있으면 모두가 운전을 할 수 있듯이, 컴퓨터만 사용할 수 있다면 모두가 데이터를 분석할 수 있는 것이다.



목적에 맞는 분석 도구를 선택해보자

빅 데이터의 시대가 되면서 데이터 분석은 셀프서비스로 진화하고 있다. 어느 순간 물이 셀프 서비스가 되었듯이 데이터 분석도 셀프서비스가 되고 있다.

어떤 지역으로 이동하려면 자동차가 필요하듯이 데이터 분석을 하려면 분석 도구가 필요하다. 데이터 분석을 전문적으로 해보지 않았던 현업 실무자는 목적에 맞는 현명한 분석 도구를 선택하기 어려울 수 있다. 아무거나 선택한다면 서울에서 대전을 가는 데 승용차가 아니라 트럭을 잘못 선택하여 고생하는 꼴이 될 수 있다.

만약 서버를 통해 수집한 데이터를 단순 집계한다면 엑셀로도 충분하다. 굳이 R을 사용할 이유가 없다. 그러나 빅 데이터를 수집하고 처리하고 모델링을 해야 한다면 R이나 파이썬을 다룰 수 있어야 한다.

목적에 맞는 분석 도구의 선택은 셀프서비스 데이터 분석에서 특히 더 중요하다. 업무를 효율적으로 해야 하는데 잘못 선택하면 오히려 비효율이 될 수 있기 때문이다. 가장 대표적인 도구 몇 가지를 살펴보자.

첫째, 검색 플랫폼에서 제공하는 데이터 분석 도구를 이용하면 손쉽게 데이터 분석을 할 수 있다. 구글의 구글트렌드, 네이버의 데이터랩, 바이브(다음소프트)의 썬트렌드, 그리고 빅카인즈 등이 대표적이다.

구글트렌드와 데이터랩은 특정 검색어가 얼마나 많이 검색되었는지 확인할 때 유용한 도구다. 글로벌 트렌드를 알고 싶다면 구글트렌드를, 국내 트렌드를 알고 싶다면 데이터랩이 더 유용할 수 있다.



썬트렌드는 온라인 데이터러를 통해 키워드분석, 평판분석, 비교분석을 제공한다. 무료로 이용할 수 있으며 고급 분석은 유료 서비스를 받아야 한다.

빅카인즈는 신문, 방송 등 국내 54개 주요 언론사의 6천 5백만 건 뉴스를 축적하고 있는 국내 최대의 공공 뉴스 아카이브다. 빅카인즈 내에서 뉴스 검색, 연관어 등 다양한 빅 데이터 분석을 쉽고 간단하게 할 수 있다. 무료 회원가입으로 분석한 내용을 다운로드해서 다른 분석 도구에서 추가 분석을 할 수도 있다. 특히 텍스트 마이닝으로 뉴스에 언급된 키워드의 빈도와 연관어를 분석할 수 있다.

썬트렌드의 역세권과 슬세권 연관어 비교 분석



빅카인즈의 슬세권 연관어 분석



둘째, 마이크로소프트의 엑셀은 가장 보편적으로 사용하는 데이터 분석 도구다. 엑셀은 데이터가 테이블 형태로 되어 있어서 직관적으로 데이터를 관리할 수 있다. 여기에 파워 쿼리나 피벗 기능을 이용하면 쉽게 데이터를 정리하고 시각화할 수 있다.

파워 쿼리는 분석 요구에 부합하도록 데이터 원본을 검색, 연결, 결합 또는 구체화할 수 있게 해주는 데이터 연결 기술이다. 파워 피벗은 데이터 모델을 만들고, 관계를 설정하고, 계산을 만들 수 있는 데이터 모델링 기술이다. 그리고 파워 BI와 연동이 되기 때문에 비즈니스 인텔리전스(BI)까지 구현할 수 있다.

정품 사용자라면 추가적인 비용 부담 없이 이런 모든 기능을 사용할 수 있다. 다만 처리용량이나 고급 통계분석기능, 그리고 자동화 측면에서 한계를 가지고 있다. 엑셀은 데이터 과학을 처음 접하는 현업 데이터 분석가에게 유용한 도구로 모자람이 없다.

셋째, 전문적인 통계 패키지로 SPSS와 SAS가 있다. SPSS는 어느 정도의 진입장벽이 있긴 하지만 그래픽 사용자 인터페이스(GUI)로 되어 있어서 초보자도 쉽게 사용할 수 있다. 엑셀과 같은 메뉴 구조로 되어 있다. 깊이 있는 혹은 추가적인 분석을 위해 프로그래밍이 필요한 경우도 있지만, 대다수는 기본 기능으로 고급통계분석, 데이터 마이닝까지 가능하다.

SAS는 프로그래밍해야 되기 때문에 어느 정도 전문성이 요구되는 분석 도구다. 강력한 통계분석 기능을 제공하고 있으며, 시각화 및 비즈니스 인텔리전스 기능이 추가되고 있다. 두 패키지 모



두 유료이고 시험판을 일정 기간 사용할 수 있다. 만약 통계와 프로그래밍에 대한 이해가 약하다면 SPSS를 사용하는 편이 훨씬 스트레스를 줄일 수 있다.

넷째, R과 파이썬과 같은 오픈소스 프로그래밍 언어가 있다. R은 통계 분석과 그래픽을 위한 프로그래밍 언어이자 다양한 패키지의 집합으로 구성된 데이터 분석 도구다. 워크시트 대신에 입력창만이 존재하기 때문에 통계와 프로그래밍에 대한 기본적인 이해가 있어야 사용할 수 있다.

R을 포함한 모든 프로그래밍 언어에서는 변수라는 개념을 사용하는데 이런 변수는 작업공간에 존재하지만 고정된 위치를 갖지는 않으며, 벡터, 행렬, 테이블 등 고유의 자료형을 갖는다.

파이썬은 보다 빠르게 작업하고 시스템을 보다 효과적으로 통합할 수 있는 프로그래밍 언어이며 개발도구다. 두 언어 모두 통계분석이 가능하지만, R은 통계분석에 특화된 언어이고, 파이썬은 개발에 특화되어 있다. 그 쓰임새에 따라 선택적으로 사용할 필요가 있다. 무료라는 장점이 있는 반면에 언어를 익혀야 하는 어려움이 있다.

다섯째, 래피드마이너(RapidMiner)와 나임(Knime) 같은 오픈소스 플랫폼도 있다. 드래그 앤 드롭만으로 프로그래밍 없이 분류와 패턴 발견 등의 복잡한 분석을 할 수 있다. 현업 실무자인 초보 데이터 분석가부터 전문 분석가까지 이용할 수 있는 데이터 분석 플랫폼이다.

래피드마이너는 GUI 방식으로 데이터 마이닝을 할 수 있는 분석 도구다. 나임은 Work-Flow 기반으로 데이터 분석을 수행할 수 있으며, 코딩 기반의 다른 데이터 분석 프로그램보다 코딩에 대한 기능 요구도가 낮고, 분석 과정을 시각화하여 볼 수 있다.

마지막으로 비즈니스 인텔리전스 도구로 파워 BI와 태블로 등도 있다. 모든 데이터 분석은 분

석 결과를 해석하여 통찰을 얻어야 하는데 그 중심에 시각화 분석이 있다. 데이터 시각화는 데이터를 수집 및 처리하여 시각적으로 도식화하는 기법으로 아이디어를 탐구하고 정보를 전달할 수 있는 빠르고 효과적인 방법이며, 비즈니스 인텔리전스를 위한 대시보드 역할을 한다.

대시보드를 만들게 되면 한눈에 동태적으로 데이터 분석의 결과를 실시간으로 볼 수 있고 결과물을 공유할 수 있기 때문에 매우 유용한 통찰을 제공할 수 있다.

파워 BI는 엑셀 데이터를 그대로 사용할 수 있고 무료로 이용할 수 있어서 초보자에게 유용한 도구다. 태블로는 강력한 시각화 도구이지만 유료이기 때문에 전사적으로 조직에서 도입할 때 유용하다. 현업 실무자로서 대시보드에서 데이터를 자유롭게 분석할 수 있어야 한다.

이 외에도 다양한 데이터 분석 도구들이 있다. 중요한 것은 도구 자체가 아니다. 어떤 도구를 사용하느냐는 해결하고자 하는 문제 및 데이터의 특성에 따라 달라질 수 있다. 목적에 맞는 도구를 선택하는 것 역시 현업 실무자들이 고민해야 한다. 현업 실무자로서 데이터 분석을 하고자 한다면 먼저 엑셀과 검색플랫폼에서 필요한 분석을 하면서 데이터 분석을 하는 습관을 기를 필요가 있다.

좀 더 깊이 있는 분석을 하고자 한다면 SPSS와 같은 통계 패키지를 사용하고, 전문적인 데이터 분석 및 시각화 분석을 하고자 한다면 R과 같은 프로그래밍 언어를 익혀야 한다. 코딩에 자신이 없다면 래피드마이너와 같은 분석 플랫폼으로 데이터 마이닝 및 예측분석을 할 수 있다.

현업 실무자는 모든 분석 도구와 방법을 익히는 것이 아니라 나의 업무에 필요한 분석 도구와 방법만 집중적으로 학습할 것을 권장한다. 무료이면서도 초보자가 사용할 수 있는 분석 도구로 텍스트 마이닝은 검색플랫폼을, 집계 및 요약 등 기초적인 기술 통계와 가설검정 등은 엑셀을, 그리고 예측 분석 및 데이터 마이닝은 래피드마이너를 분석 도구로 추천한다. R이나 파이썬 같은 프로그래밍 언어를 사용하지 않아도 셀프서비스로 데이터 분석을 할 수 있다.



데이터 분석의 복병, 통계의 벽을 넘어라

데이터 분석은 통계를 기반으로 하고 있다. 통계학을 몰라도 데이터 분석을 할 수 있지만, 통계를 이해하면 분석 결과의 의미를 깊이 있게 파악할 수 있다. 과거에는 주로 표본을 추출하여 서베이를 했기 때문에 통계분석을 통해 구한 통계량으로 모수를 추정하여 모집단의 특성을 파악했었다.

빅 데이터 시대가 도래하면서 표본이 아닌 모집단의 데이터 수집이 가능해졌다. 추정을 할 필요가 없어지면서 일부 통계 무용론이 제기되기도 했지만, 이는 오해라고 할 수 있다. 아직 대다수의 데이터 분석은 전체가 아닌 일부의 데이터를 사용하고 있으며, 예측하거나 모델링을 하는 데 통계량을 기반으로 하고 있다. 분석의 결과를 통계적으로 유의미한지 살펴보고 해석하기 위해서는 통계의 벽을 넘어야 한다.

첫째, 데이터 분석을 위한 기초 통계 및 기술 통계로 최소한 빈도 분포에 대해 이해하고 있어야 한다. 빈도 분포는 모든 분석에서 가장 기본적이면서도 매우 중요한 개념이다.

빈도 분포의 통계량은 위치와 변동성, 왜도와 첨도가 대표적이다. 집중하는 경향을 측정하는 위치 통계량은 평균이다. 일반적으로 평균값(mean)이라고 하면 산술평균을 말한다. 그러나 산술평균은 극단치(outlier)가 있는 경우에 변수의 특성을 왜곡할 가능성이 있다.

예를 들어, 국회의원의 평균재산이 22억 원이라고 할 때 특정의 1인이 2,311억 원의 재산을 가지고 있다면 평균의 의미가 무의미해진다. 이때 최빈값(mode)과 중앙값(median)을 함께 살펴봐야 한다. 소득분포를 볼 때 특히 평균값에 유의해야 한다. 변동성 측정 방법으로 범위, 4분위 범위, 분산과 표준편차, 변동계수 등이 있다. 분산은 어떤 변수의 평균을 중심으로 흩어진 정도를 나타낸다.

분산이 크다는 것은 데이터들이 많이 퍼져 있다는 것을 의미한다. 형태를 측정하는 첨도는 분포의 모양이 중심점에서 뾰족한가를 나타내는 통계량이고, 왜도는 분포의 모양이 얼마나 좌우대칭인지를 나타내는 통계량이다. 이런 통계량으로 변수의 특성을 이해하게 된다.

둘째, 가설검정은 모집단의 특성에 대한 통계적 가설을 모집단으로부터 추출한 표본을 사용하여 검토하는 통계적인 추론 방법이다. 가설을 설정하고, 유의수준을 설정하고, 검정 통계량을 산출한 다음, 가설에 대해 기각 및 채택을 판단하게 된다. 가설은 연구자가 관심을 두고 있는 현상이나 요인에 관한 증명되지 않은 진술이나 제안





으로 차이가 없거나 영향이 없다는 것을 의미하는 귀무가설과 어떤 견해나 행동에 변화를 유발할 수 있음을 의미하는 대립가설을 설정한다.

예를 들어, '성별에 따라 소득수준에 차이가 없다'라는 귀무가설을 설정하고, 유의수준을 0.05(신뢰수준 95%)로 설정했다면, P값이 유의수준 값보다 작게 되면($p < 0.05$) 귀무가설을 기각하고 대립가설을 채택하게 된다. 즉, 유의수준 5%(신뢰수준 95%)에서 성별에 따른 추천 의향에 차이가 있다고 해석한다.

셋째, 변수와 변수 간의 관계에 대한 인과관계와 상관관계를 이해해야 한다. 인과관계(causation)는 원인과 결과의 관계로 원인이 선행되고 그에 따라 결과로 나타나는 관계다. 실험 및 시계열 데이터를 수집하여 회귀분석을 통해 파악할 수 있다. 상관관계(correlation)는 두 변수 중에서 어떤 변수가 원인이고 어떤 변수가 결과인지 알 수 없지만 두 변수 간에는 관계가 있는 것을 말한다. 설문조사나 소셜데이터를 수집하여 상관분석으로 상관계수를 구해서 파악한다.

예를 들어, 무더위로 인해 아이스크림 판매량이 증가했다면 인과관계가 있다. 무더위로 인해 익사자 수가 증가했다면 역시 인과관계가 있다. 하지만 아이스크림 판매량과 익사자 수의 증가 간에는 인과관계를 알 수 없으며 단지 상관관계(?)가 있을 따름이다. 단, 이 상관관계는 통계적으로는 의미가 있을지언정 현실에서는 무의미한 상관이다. 어떤 변수 간에 상관관계가 강하다고 해서 그사이에 인과관계가 있다고 단정하면 오류나 왜곡이 일어날 가능성이 매우 높다는 점을 유의해야 한다. 제대로 된 데이터 분석을 하려면 조사방법론에 대한 이해와 분석적 사고가 필요하다.

셀프서비스 데이터 분석 기술을 갖추자

데이터 분석은 요리와 비슷하다. 신선한 식재료를 구하고 적절한 조리도구를 사용해서 조리할 때 맛있는 음식이 된다. 마찬가지로 좋은 분석 결과물을 얻기 위해서는 재료인 데이터를 수집하고 적절한 분석 도구를 선택하여 분석해야 한다. 그리고 데이터 분석을 제대로 하려면 데이터 수집이 잘되어야 하고, 데이터를 잘 수집하려면 분석의 목적이 분명해야 한다. 즉, 데이터 분석의 목적과 수집과 분석은 일련의 연결된 프로세스로 인식하고 분석 업무를 해야 한다.

데이터에는 정형 데이터(고객 수, 판매량 등)와 비정형 데이터(텍스트, 위치 정보, 사진, 동영상 등)가 있다. 데이터의 유형에 따라 데이터 분석 방법도 정형 데이터 분석(빈도분석, 회귀분석, 데이터 마이닝 등)과 비정형 데이터 분석(텍스트 마이닝, 소셜 분석 등)이 있다.

분석 과정에서 통찰을 얻고 통찰의 결과를 바탕으로 실행에 옮길 때 새로운 가치를 만들 수 있다. 데이터 분석을 통해 매출 증가, 수요 예측, 리스크 경감, 평판 개선, 신제품 개발, 이직률 개선 등 다양한 분야에서 가치를 창출할 수 있다.

데이터 분석을 통한 가치창출 프로세스



데이터 분석에도 수준이 있다.

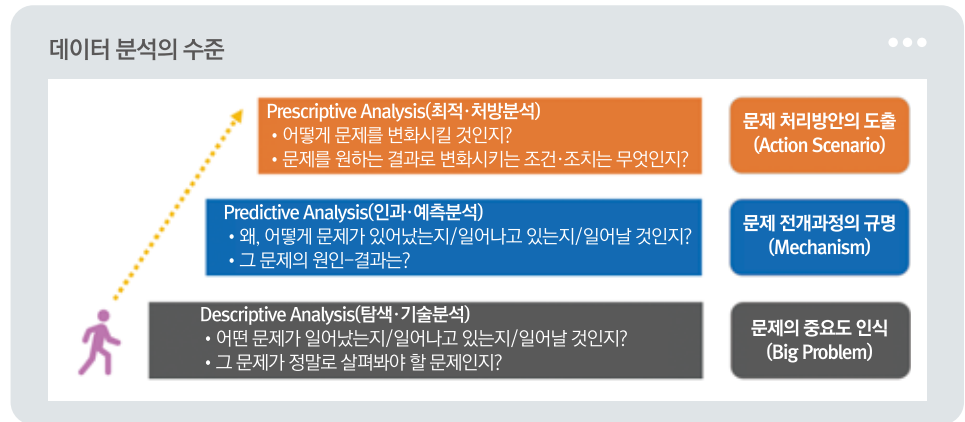
첫 번째 수준은 문제의 중요도 인식을 위한 분석으로 탐색 혹은 기술 분석(Descriptive Analysis)이다. 어떤 문제가 일어났는지, 일어나고 있는지, 또는 일어날 것인지, 그리고 그 문제가 정말로 살펴봐야 할 문제인지 등을 탐구하는 분석 단계다.

두 번째 수준은 문제 전개 과정의 규명으로 인과 혹은 예측 분석(Predictive Analysis)이다. 왜, 어떻게 문제가 일어났는지, 일어나고 있는지, 일어날 것인지, 그리고 그 문제의 원인-결과논 어떻게 되는지를 파악하는 분석 단계다.

세 번째 수준은 문제 처리 방안의 도출을 위한 분석으로 최적 혹은 처방 분석(Prescriptive Analysis)이다. 어떻게 문제를 변화시킬 것인지, 문제를 원하는 결과로 변화시키는 조건 혹은 조

치는 무엇인지를 밝히는 분석 단계다.

현업 실무자는 셀프서비스 데이터 분석으로 최소한 첫 번째 수준인 탐색 혹은 기술 분석까지는 기본적으로 분석할 수 있어야 한다. 그리고 두 번째 수준인 인과 혹은 예측 분석을 할 수 있도록 노력해야 한다. 세 번째 수준인 최적 혹은 처방 분석에 도전하기 위해서는 전문적인 학습과 연습이 필요하다.



셀프서비스로 데이터 분석의 기술을 익히려면,
 첫째, 많은 숫자 속에서 원인과 결과를 찾아내는 것에 관심을 가져야 한다.
 둘째, 변수와 변수 간의 연관성을 찾아내야 한다.
 셋째, 데이터 홍수 속에서 진짜 의미를 찾아내기 위한 과감한 도전이 있어야 한다.
 넷째, 현명한 의사결정 정보를 얻기 위해 항상 문제를 먼저 정의하고 분석적 사고를 해야 하며
 나의 가설을 지지할 데이터를 갖고 있어야 한다.
 마지막으로 데이터를 현명하게 활용하기 위해서는 생각하고, 생각하고, 또 생각해야 한다. 그
 리고 연습하고, 연습하고, 또 연습해야 한다.



기후 변화, 지구는 어디로 가고 있을까



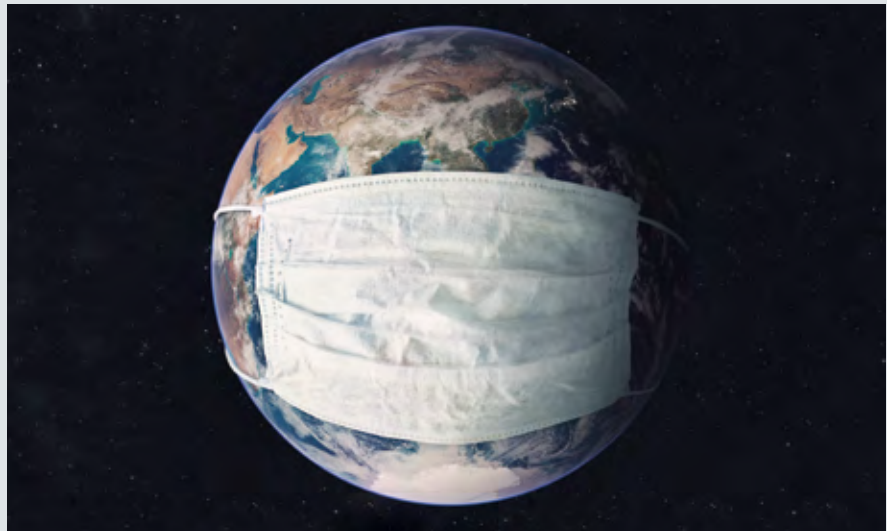


그야말로 앞친 데 덮친 격이다. 전 세계가 코로나19 사태로 초유의 고통을 겪고 있는 가운데 산불과 폭우, 태풍 등 숨돌릴 틈도 없이 연이어 발생하고 있는 자연재해로 지구가 신음하고 있다.

국내의 경우 유래가 없는 긴 장마를 경험했고, 미 서부 지역은 산불로 인해 남한의 5분의 1에 해당하는 면적이 잿더미로 변한 상황이다. 아시아는 또 어떤가. 끊임없이 이어지는 태풍들로 인해 수많은 인명피해와 천문학적인 손실을 입었다.

이처럼 전 세계에서 동시다발적으로 일어나고 있는 자연재해의 원인에 대해 대다수 전문가들은 ‘기후변화’를 주범으로 보고 있다. 기후변화의 속도가 최근 들어 더욱 빨라지면서 기상과 관련된 각종 재난을 일으키는 원인이 되고 있다는 것이다.

지구가 각종 재해로 신음하고 있는 이유는 기후변화 때문이다



출처: unctad.org

물론 기후변화의 심각성이 거론된 것은 어제오늘의 일이 아니다. 또한 기후변화에 맞서 전 세계가 동참 의지를 밝힌 것도 사실이다. 지난 2015년, 200여 개에 달하는 국가들이 파리기후변화협약을 맺고 온실가스 배출량을 단계적으로 감축하자고 약속한 것은 그 같은 의지를 보여준 사례다.

하지만 자국 이익만을 생각한 일부 국가들의 이기적인 행동과 코로나19 사태 등으로 인해 기후변화에 대응하고자 하는 인류의 노력은 별다른 결실을 얻지 못한 채 답보 상태에 머무르고 있다. 그리고 그런 상황을 틈타 기후변화의 속도는 더욱 가속화되고 있는 중이다.

그렇다면 앞으로 기후변화의 미래는 어떻게 진행될까? 지금으로서는 비관도 낙관도 시기상조라는 것이 대다수 전문가의 의견이다. 이들은 무엇보다도 과학적 통계 시스템을 기반으로 하여 기후변화가 어떤 영향을 미치고 있는지를 조사하는 것이 필요하다고 입을 모은다. 그런 과학적 조사 결과를 바탕으로 해야만 미래를 보다 정확하게 예측할 수 있기 때문이다.



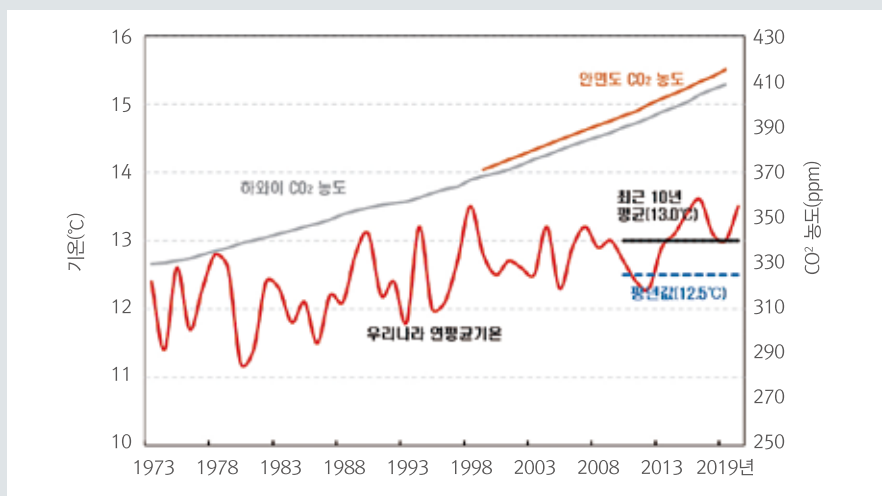
환경 분야: 국내 온실가스 배출량 1990년 이후 약 2.37배 증가

환경부의 발표에 따르면 국내 온실가스 배출량은 지난 1990년 이후 26년 동안 약 2.37배가 증가한 것으로 나타났다. 그중 이산화탄소 배출량이 차지하는 비율은 1990년 86.1%에서 2016년 91.9%로 증가했다.

또한 2018년에 발간된 미국의 연례기후보고서를 살펴보면 2017년을 기준으로 세계 온실가스 배출량이 현대 대기 관측 사상 최고치를 기록한 것으로 파악됐다.

이는 화석에너지의 연소 등 빠르게 진행되는 산업화로 인해 대기 중의 이산화탄소 농도가 급속히 증가했기 때문이다. 이런 이유로 세계기상기구(WMO)와 국제연합환경계획(UNEP)은 이산화탄소를 지구 온난화의 주범이라고 공식적으로 선언한 상황이다.

연도별 이산화탄소 농도와 기온 변화



출처: 환경부

온실가스 배출량이 증가함에 따라 북극의 지표면 온도도 1981년부터 2010년까지의 평균 온도보다 1.6°C가 상승했다. 또한 빙하 크기는 38년 동안 계속해서 감소하는 추세를 보여주었으며, 이에 따라 해수면 높이도 6년 연속 최고치를 달성한 상황이다.

국내 평균 기온이 높아지는 것도 어찌 보면 이 같은 추세에 기인하고 있다. 일기예보 전문업체인 케이웨더(KWeather)는 한반도가 아열대 기후로 변하고 있으며, 유엔 기후변화협의체(IPCC)에서 예상하는 것보다 더 빠르게 아열대로 진입할 가능성이 크다고 예측한 바 있다.





농업 분야: 배 재배는 줄고 아열대 과일은 매년 증가

기후변화로 인해 가장 타격을 받고 있는 분야는 농업이다. 4계절이 뚜렷했던 한반도의 기후는 어느새 열대성 기후로 변하고 있고, 이에 따라 곡물 및 과일 등의 재배지역도 과거와는 판이한 결과를 보여주고 있다.

달고 물이 많은 것으로 유명한 우리나라의 배가 대표적인 사례다. 농림축산식품부가 공개한 자료를 보면 2008년부터 2019년까지의 배 재배면적은 1만 8277ha에서 9616ha로 47.4%가 줄어든 것으로 나타났다.

또한 같은 기간 생산량도 47만 743톤에서 20만 732톤으로 무려 57.4%가 급감한 것으로 드러났다. 특히 국산 배의 주산지인 전남에서조차 생산량이 12만 7188톤에서 5만 582톤으로 60.2% 줄어든 것은 기후변화의 영향을 극명하게 보여주는 수치다.

반면에 참다래와 무화과, 그리고 망고 및 백향과 등 아열대 과일의 경우는 관련 통계가 작성된 2017년부터 지난해까지 3년간 매년 늘고 있는 추세다. 전체 재배면적은 50% 증가했고, 재배 농가수도 45.9% 급증했는데, 그중 망고의 재배면적이 가장 많이 증가했다.







이 같은 현상에 대해 농림축산식품부는 이상 기온으로 인해 지난 20년간 배의 개화 반응이 약 2일 정도 빨라진 점을 주목하고 있다. 또한 봄철 서리 등 냉해 피해 위험이 증가하면서 재배 적합지가 변동하고 있는 점도 배의 생산량이 감소한 이유로 판단하고 있다.

농촌진흥청 역시 '신농업 기후변화 기획보고서'를 통해 현재 기온보다 1.5℃ 상승하게 되면 2040년대의 한반도 기후에서는 고품질 배의 재배 적지가 급감하게 되고, 고랭지 배추 역시 90% 이상의 재배지가 감소할 것으로 예상하고 있다.

따라서 농업정책이 기후변화와 소비자 수요에 세밀하게 대응하지 못한다면, 농업의 미래는 불투명해질 수밖에 없다는 것이 대다수 전문가의 의견이다.

주요 작물의 재배면적 변화 추이

(단위: ha, %)

구분		 사과	 배	 감귤류	 단감	 포도	 복숭아
2019년	재배면적①	32,954	9,615	20,021	8,639	12,676	20,636
	전망면적②	32,976	10,215	20,038	8,959	12,960	21,288
	차이(①-②)	△22	△600	△17	△320	△284	△652
2018년 재배면적		33,234	10,303	20,119	9,108	12,795	21,087
2018년 대비 증감률	재배면적③	△0.84	△6.68	△0.49	△5.15	△0.93	△2.14
	전망면적④	△0.78	△0.85	△0.40	△1.64	1.29	0.95
	차이(③-④)	△0.06	△5.82	△0.09	△3.51	△2.22	△3.09

출처: 농촌경제연구원



질병 분야: 코로나도 기후변화와 밀접한 관계가 있다

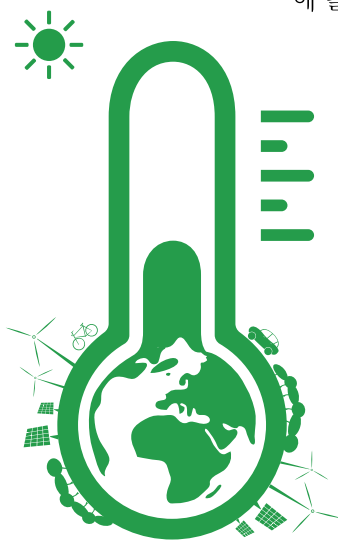
기후변화는 농업만큼이나 사람의 건강에 많은 영향을 미치는 것으로 알려져 있다. 그중에서도 지구온난화로 인해 자주 발생하는 폭염은 건강을 해치는 요인 중 가장 큰 요인이다.

기상청 발표에 따르면 올해 6월의 평균 일 최고기온은 28.0℃였고, 일 평균기온은 22.8℃인 것으로 파악됐다. 이는 지난 1973년 이래 최고 기록을 달성한 것으로서, 폭염일수도 2.0일에 달했다. 6월의 폭염일수가 2.0일인 것은 그동안의 6월 평균 폭염일수보다 1.4일이 나 늘어난 것이다.

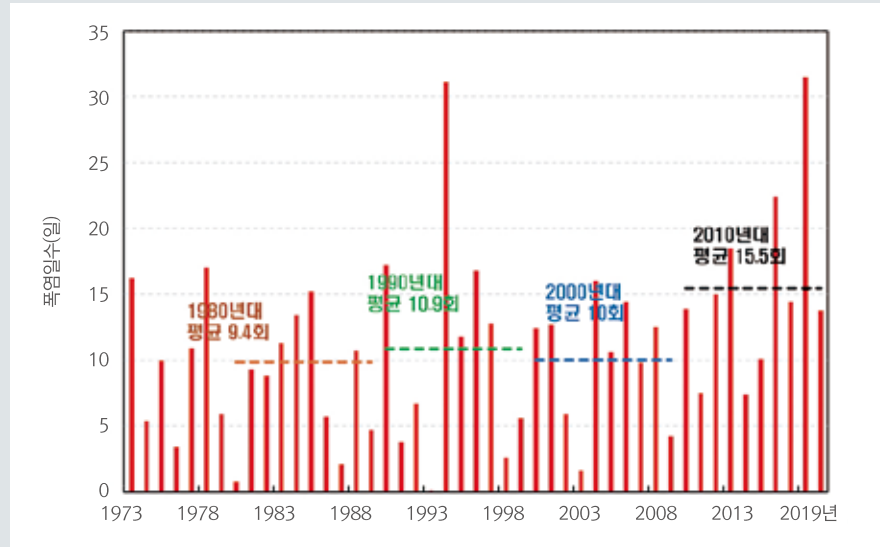
폭염은 열사병뿐 아니라 탈수 관련 합병증 등 광범위한 질병을 초래한다. 또 폭염은 만성 폐질환과 함께 심장이나 신장에 문제가 있는 환자들에게 악영향을 미친다. 기존 질환의 증상을 악화시키고 심지어는 정신질환에도 영향을 미치는 것으로 나타났다.

질병관리본부의 통계에 따르면 지난 2018년 한국에서 폭염과 관련한 발병자는 4526명이었고, 사망자도 48명에 이르는 것으로 조사됐다. 그러나 폭염으로 인한 간접적인 건강 영향이 반영되지 않은 점을 고려하였을 때 실제 사망자는 훨씬 더 많을 것으로 추정되고 있다.

기후변화의 직접적 요인인 폭염이 질병의 원인 중 하나라면 어린이는 기후변화에 따라 발생하는 질병의 직접적인 대상이라 할 수 있다. 면역력이나 체력이 성인에 비해 많이 떨어지는 만큼 기후변화에 적응하기 어렵기 때문이다.



연도별 폭염일수 추이



출처: 환경부

세계보건기구(WHO)에 따르면, 기후변화로 인해 발생하는 질병의 88%는 5세 미만의 어린이에게 발생하고, 미국소아과학회에서는 2015년에 기후변화가 어린이들의 건강과 직결된다고 경고한 바 있다.

그 이유로 여름철 이상고온이 1세 미만 어린이들의 건강에 치명적인 위협을 줄 수 있고, 적절





한 영양 공급을 막을 수 있다는 점을 들었다. 영양 공급의 경우 대기 중 이산화탄소의 농도 변화가 곡물의 단백질 함량을 낮출 수 있다는 것이 세계보건기구 측의 설명이다.

이 외에도 소두증을 유발하는 지카바이러스나 5세 미만 아동 사망에서 큰 비중을 차지하는 말라리아, 그리고 현재 전 세계를 공포의 도가니로 몰아넣고 있는 코로나19 등이 모두 기후변화가 미치는 영향 때문에 발생한 것으로 파악되고 있다.

일찍 생각하면 코로나19 같은 전염병과 기후변화는 별다른 상관관계가 없는 것처럼 보이는 것이 사실이다. 그러나 조금만 더 그 속을 들여다보면 전염병과 기후변화는 상당히 밀접한 관계가 있음을 알 수 있다.

이는 온도가 상승하면서 질병이 퍼지는 데 적합한 환경을 만들었기 때문이다. 코로나바이러스만 해도 과거에는 깊은 밀림 속의 박쥐에서나 발견되었지만, 개발로 인해 밀림이 사라지고 온도는 따뜻해지면서 인간의 영역에까지 침범한 것으로 추정하고 있다.



에너지 분야: 여름철 열대야와 겨울철 한파로 에너지 소비 증가

에너지 부문에서는 폭염 및 열대야로 7월~8월 건물 부문 전력 소비량이 증가하여 아파트 정전 횟수가 증가했으며 1월~2월 사이에는 겨울철 한파로 인한 에너지 소비량이 증가한 것으로 나타났다.

전력 소비의 경우 지구온난화에 따라 여름철 냉방 전력 소비는 증가하고 있고, 겨울철 난방 전력 소비는 감소 추세에 있는 등 계절적 대비가 뚜렷하게 발생하고 있다. 따라서 2020년대 중반

기후변화 대응기술을 활용한 우리나라의 미래 모습



출처: 환경부



무렵부터는 여름철 냉방에 의한 전력소비가 겨울철 난방에 의한 소비를 넘어설 것으로 추정되고 있다.

한편 원자력과 석탄 및 가스 발전, 그리고 수력 및 일부 태양력 발전의 경우 냉각 과정이 필수적이다. 그런데 기후변화에 의해 가뭄 발생이 증가하여 냉각수 부족으로 인한 전력 생산 차질이 우려되고 있다. 또한 높은 기온으로 인하여 냉각수 온도가 높아지게 되면서 전력 생산의 효율성은 떨어질 것으로 예상되고 있다.

시야를 해외로 돌리면 기후변화와 에너지 소비의 상관관계가 보다 명확해진다. 지난 2018년에 국제에너지기구(IEA)가 조사한 에너지통계연보에 따르면 석유의 비중이 32%로서 가장 크고 그다음인 27%를 차지한 석탄과 22%를 점유한 천연가스인 것으로 나타났다.

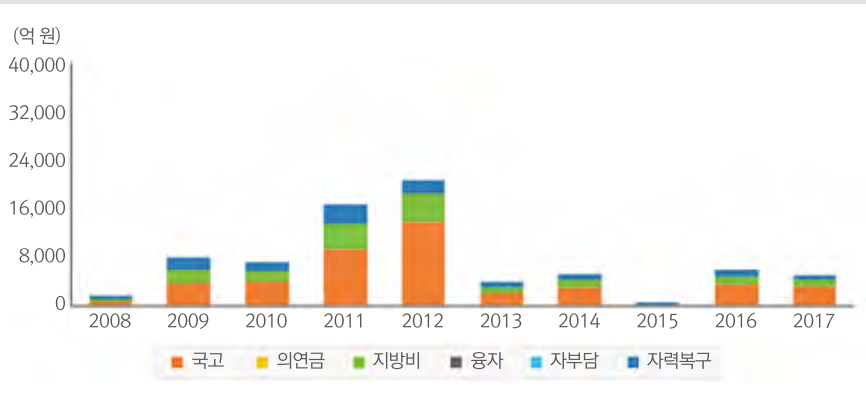
소비되는 연료의 대부분이 화석연료이며 발전용으로 사용되는 석탄의 소비가 증가했다는 것이 눈에 띈다. 이 같은 추세를 고려할 때 재생에너지 사용 비중이 아직은 부족한 상황에서 사용하는 화석연료는 지구온난화를 앞당기는 촉매제로 사용되고 있는 것으로 파악되고 있다.



재난분야: 태풍, 대설, 지진, 풍랑, 강풍 등 자연재해 증가

행정안전부가 발간한 「재해연보」에 따르면, 지난 2004년부터 2018년까지 15년 동안 연평균

최근 10년간 재원별 자연재난 복구비 추이



출처: 행정안전부

자연재난 피해액은 약 5430억 원이며, 복구액은 1조 320억 원에 달하는 것으로 나타났다.

이 중에서 최근 10년(2009년~2018년) 사이로 기간을 좁히면 원인별 자연재난 피해액인 총 3조 6281억 원을 기준으로 호우와 태풍이 88.5%, 그리고 대설 6.6%, 지진 2.7%, 풍랑 1.2%, 강풍이 1.1%를 차지하는 것으로 드러났다.

또한 총 7조 7095억 원에 달하는 복구액은 호우·태풍 94.5%, 지진 2.5%, 대설 1.7%, 풍랑 0.5%, 강풍 0.2%인 것으로 파악됐다. 이에 따라 국회예산정책처는 기후변화의 영향을 고려한 향후 자연재난 피해액을 연간 11.5조 원으로 추정하고 있다.

향후 2020~2060년 동안 발생 가능한 연간 자연재난 피해액은 2002년에 발생한 최대 피해액의 1.4배인 11조 4,794억 원으로 추정하였다.

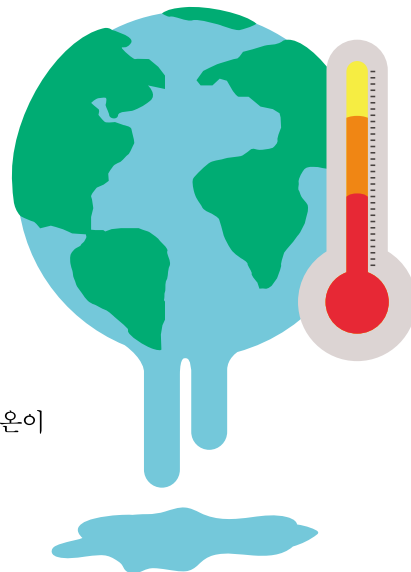
기후변화와 결부되어 시설물의 노후화 및 성능 저하가 빠르게 진행되고 있다. 붕괴 또는 기능 상실시 상당한 인명·재산 피해 발생이 우려되는 댐, 하천 관리시설, 상·하수도, 절토사면 등 주요 시설물들의 노후화가 빠르게 진행되고 있다.



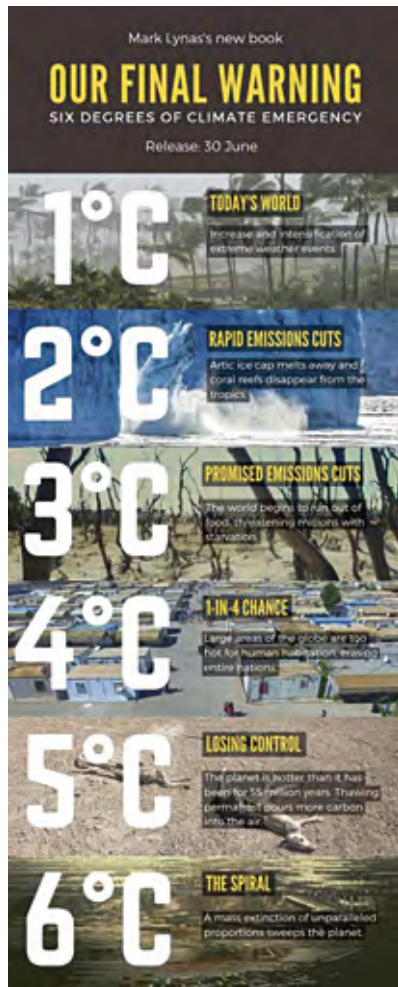
기후변화와 미래 전망 시나리오

기후변화와 재난의 상관관계를 묘사한 저서인 『6도의 악몽(Six Degrees)』은 영국의 환경운동가이자 저널리스트인 '마크 라이너스(Mark Lynas)'가 집필했다. 라이너스는 이 저서를 통해 지구의 평균기온이 1℃에서 6℃까지 상승할 때 나타날 수 있는 기후변화 시나리오를 알기 쉬우면서도 섬뜩하게 묘사했다.

우선 지구평균 기온이 1℃ 상승하면 만년빙이 사라지고 사막화가



지구 평균기온이 6℃ 상승할 때의 과정을 경고한 저서 ...



출처: MarkLynas.org

심화되면서 기상이변 현상이 더욱 빈번하게 나타날 것이라고 예측했다. 현재 우리가 겪고 있는 기상이변 현상과 많이 닮아 있음을 알 수 있다.

또한 지구 평균기온이 2℃ 상승하면 대가뭄과 대홍수가 닥치고, 북극의 빙하가 대부분 녹으면서 해수면이 상승하는 것으로 전망했다. 해수면이 상승하면 지구 평균기온이 3℃ 상승하면서 최악의 상황으로 가게 된다. 뉴욕시 등 해안가 주변에 형성된 도시들이 침수 사태를 겪으면서 수많은 이재민이 발생할 것이다.

이어서 4℃가 상승하게 되면 시베리아의 영구 동토층이 본격적으로 녹으면서 남극의 빙하는 하루가 다르게 사라져 버린다. 빙하가 사라진다는 것은 빙하 속에 들어있던 물질이 방출된다는 것을 의미한다. 예를 들어 영구동토층에 갇혀 있던 온실가스인 메탄이 분출되면 지구온도는 더 상승하게 된다. 빙하가 사라질수록 지구온난화가 가속될 수밖에 없는 것이다.

5℃ 상승부터는 지구가 지옥처럼 변하는 것을 목격하게 된다. 살아남은 사람들은 식량과 물을 확보하기 위해 닥치는 대로 강탈하고 살인을 저지를 것이다. 해저에 갇혀있던 메탄 하이드레이트가 분출하면서 지구의 기온은 6℃까지 올라가는데, 기온이 6℃까지 상승하게 되면 인류를 포함한 모든 동식물이 멸종하게 된다.

이 같은 비극적인 상황이 닥치기 전에 전 세계는 절박하게 힘을 합쳐야만 하는 상황이다. 물론 코로나19라는 변수가 발생하면서 인류의 생존에 위협을 주고 있는 상황임은 분명하지만, 모든 인류를 위협에 빠뜨릴 정도의 규모는 아니다. 게다가 바이러스 발생은 위

생이나 날씨 등의 영향에 의해 간헐적으로 나타나는 경우가 대부분이다.

하지만 기후변화는 다르다. 전 세계 인류를 멸종시킬 수 있는 힘을 갖고 있고, 지역과 시간에 상관없이 거의 동일한 현상이 일어나게 된다. 사람이 없는 오지에 들어가 산다고 한들 시간의 차이만 있을 뿐 생명을 잃는 것은 매한가지다.

따라서 그 어떤 방법을 써서라도 석탄이나 석유 같은 화석연료가 내뿜는 온실가스를 과격적으로 감축해야만 인류가 조금이라도 더 생존할 가능성이 높아진다는 것을 과학적 통계 시스템이 보여주고 있다.

범죄자를 추적하는 데이터: 사회연결망분석을 활용한 범죄수사





경찰 등 형사사법기관이 범죄 사건을 해결하기 위해 데이터를 수집하고 분석한다는 것은 이제 일반 시민들도 한번쯤 들어보았을 정도로 널리 알려진 이야기가 되었다. ‘데이터 중심 경찰활동(data-driven policing)’과 ‘증거 기반 경찰활동(evidence-based policing)’이 강조됨에 따라 데이터를 체계적으로 수집하고 분석하는 것이 한층 중요해졌다.

통상 범죄 관련 데이터를 수집·분석하는 기준은 크게 ① 범죄가 발생한 시·공간, ② 범죄사건, ③ 범죄자의 세 가지로 나누어볼 수 있다. 우리나라 경찰의 경우 범죄가 발생한 시·공간에 대한 정보는 지리적 프로파일링 시스템(GeoPros; Geographic Profiling System), 범죄사건에 대한 정보는 과학적 범죄분석 시스템(SCAS; Scientific Crime Analysis System), 범죄자에 대한 정보는 전자수사자료표 시스템(E-CRIS; Electronic Criminal Record Identification System)을 통해 관리하고 있다.

과학치안을 지향하는 경찰은 여기서 한 단계 더 나아가 범죄자들의 연결 관계 데이터를 수집, 분석하는 시스템을 운용 중이다. 범죄수사에서 범죄자의 역할을 특정하고 여죄를 파악하는 데 데이터가 긴요하게 활용되고 있는 것이다. 이때 범죄자들의 연결 관계 데이터를 분석하는 틀은 사회연결망분석(Social Network Analysis; SNA)에 기반하고 있다.





9·11 테러 사건을 계기로 범죄수사 방법으로 본격 사용된 사회연결망분석

사회적 관계에서 개인이 차지하는 위치를 사회적 지위라고 정의할 때, 범죄단체 조직원도 그 단체의 맥락에서는 하나의 사회적 지위에 해당한다. 사회연결망분석은 사회학의 전통에 따라 사회적 관계에 접근하는 방식을 다양하게 고안하였다. 역할분담과 위계구조가 명확하고 강한 결속력을 보이는 범죄단체들은 조직원들 간의 관계 유형과 상호작용을 모두 살펴볼 수 있는 전형적인 연결망이다.¹

사회연결망분석이 발전되어온 초기부터 집단에 의한 범죄현상을 설명하는 데 사회연결망분석이 활용될 수 있음을 포함한 연구들이 발표되었다. 갱의 일원으로 활동하는 청소년들에 대한 연구(Baron & Tindall, 1993)나 성범죄자들의 동료 지지(peer support)에 대한 연구(Hanson & Scott, 1996), 교도소 수감자들을 대상으로 마약 거래 연결망을 파악한 연구(Lichtenstein, 1997) 등이 그 대표적인 예이다.

2001년 발생한 9·11 테러는 당시까지 학문적인 '연구' 방법에 그치고 있던 사회연결망분석이 실무적인 범죄 '수사' 방법으로 전환되는 계기가 된 사건이었다. 사회연결망분석은 테러범들이 어

1 사회연결망분석에서 전체 연결망에서의 위치와 효과를 측정하는 것에 중점을 두는 것을 위치적 접근(positional approach), 연결망의 직접적인 관계에 중점을 두는 것을 관계적 접근(relational approach)라고 한다. 위치적 접근은 행위자가 연결망 내 다른 행위자들과 갖는 관계의 유형(patterns of relations)에, 관계적 접근은 행위자(actor)들 간의 직접적인 상호작용에 관심을 둔다.

떻게 역할을 분담하고 동시다발적으로 활동을 전개할 수 있었는지를 파악하는 데 활용되었다.

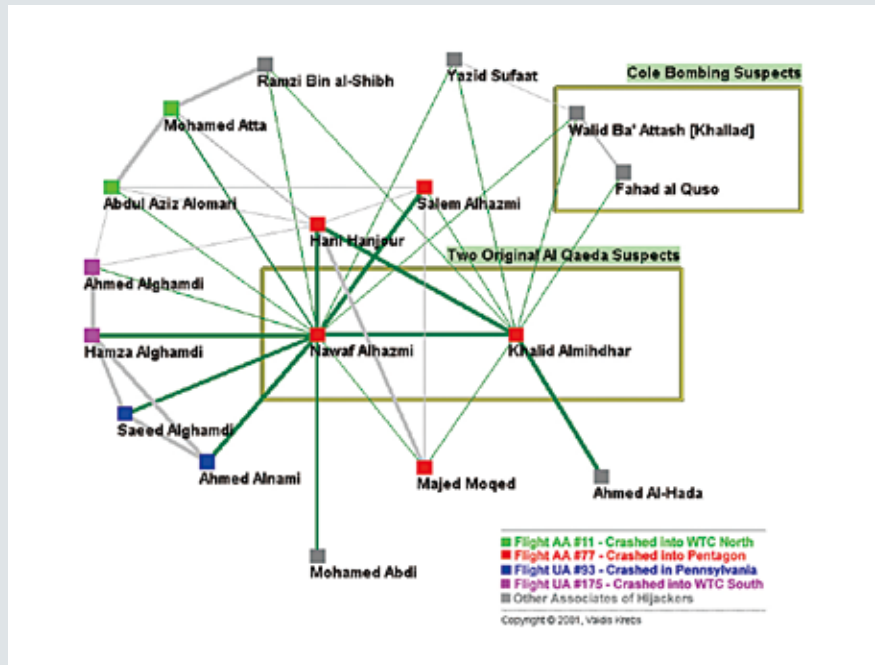
4대의 항공기를 납치한 테러범 19명은 모두 알-카에다 조직원으로, 이들의 관계는 Nawaf Alhazmi(이하 '알-하즈미')와 Khalid Almihdhar(이하 '알-미드하')를 중심으로 수집된 정보에 근거해서 파악되었다. 사건 발생 전인 2000년 초부터 CIA는 알-하즈미와 알-미드하가 테러를 모의하고 있다는 첩보를 입수하고 감시해왔는데, 이 둘이 테러범 19명 중에 포함되어 있었던 것이다.

테러범 19명의 직접적인 연결 관계를 시각적으로 나타내면 다음 그림에서 보는 바와 같이 나머지 테러범 모두 알-하즈미 또는 알-미드하와 직접적인 교류를 맺고 있었음을 확인할 수 있다. 연결망 중앙에 알-하즈미와 알-미드하가 위치하고 있는 것은 CIA가 이 두 조직원을 감시하여 수집한 자료를 바탕으로 파악된 연결망이라는 점에 기인한다.



9·11 테러범 19명의 '직접' 연결망

CIA가 데이터를 수집한 대상이었던 알-하즈미와 알-미드하가 공범들의 '직접' 연결망 중심에 위치하고 있다.



출처: <http://orgnet.com/tnet.html>, 2020년 10월 21일 접속.

사회연결망분석으로 살펴보기 전에는 공범 중 한 명이라는 것 외에는 크게 두드러져 보이지 않았던 아티는 실제로는 9·11테러에 직접 가담했을 뿐만 아니라, 알카에다 조직의 자금거래를 지원하고 은신처를 제공하는 등 핵심적인 역할을 담당했다. 사회연결망분석이 아니었다면 9·11테러를 간접적으로 지원한 연결망이 배후에 존재하였다는 사실이 자칫 간과될 수 있었다.

■ Flight AA #11 - Crashed into WTC North
 ■ Flight AA #77 - Crashed into Pentagon
 ■ Flight UA #93 - Crashed in Pennsylvania
 ■ Flight UA #175 - Crashed into WTC South
 ■ Other Associates of Hijackers

Copyright © 2001, Vaidis Krebs

2 사회연결망분석은 연결망 내에서 행위자가 갖는 위치를 중심성(centrality)이라는 개념으로 설명한다. 중심성은 크게 연결중심성(degree centrality), 근접중심성(closeness centrality), 매개중심성(betweenness centrality)로 구분할 수 있다. 연결망 내 행위자들과 많은 연결관계를 가지고 있는 행위자일수록 높은 연결중심성을 가진다. 근접중심성은 연결망 안에서의 간접적인 연결까지 고려하여 전체 연결망에서의 행위자와 다른 모든 행위자 간의 거리에 주목한다. 높은 근접중심성을 갖는 행위자일수록 연결망 안에서 교환되는 정보를 신속하게 접할 수 있다. 매개중심성은 행위자가 연결망 내 어디에 위치하고 있는지를 중요한 요소로 고려한다. 높은 매개중심성을 갖는 행위자일수록 연결망 전체에서의 의사소통의 흐름에 큰 영향을 미치게 된다. 사회연결망분석의 용어로 이야기하면, 이타는 '간접' 연결망 내 조직원들과 가장 많은 연결관계를 갖고 있다는 점에서 높은 연결중심성을 보이고, 조직원들 사이를 가장 짧은 거리로 연결한다는 점에서 근접중심성이 높으며, 직접적인 연결이 없는 조직원들의 관계도 매개하고 있어 매개중심성 또한 높다.

데이터 간의 관계에서 범죄수사의 실마리를 찾는다

범죄를 수사하는 경찰이 데이터 간 관계 분석에 관심을 기울이게 된 데에는 범죄 수법이 끊임 없이 변화하는 가운데 조직화·지능화되고 있는 경향과 밀접히 관련되어 있다. 예컨대, 과거와는 달리 소매치기나 날치기 사건은 크게 감소한 데 비해, 전화로 수사기관 등을 사칭하면서 송금을 요구하는 전기통신금융사기, 소위 보이스피싱 범죄는 2000년대에 최초 등장한 이래 좀처럼 근절되지 않고 있다.

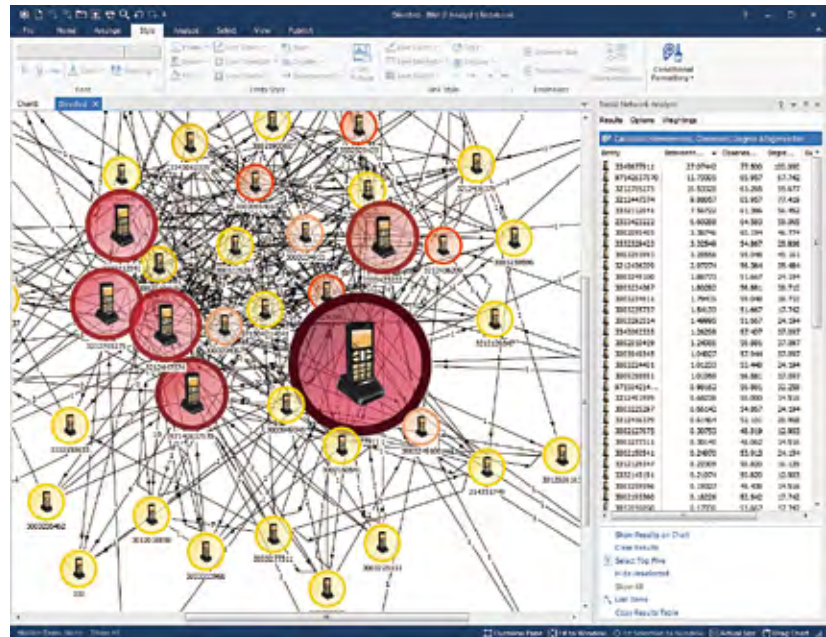
보이스피싱 범죄는 통상 ‘총책’을 중심으로 자금관리 또는 교육을 담당하는 ‘간부’급 조직원, ‘상담원’, 피해금을 인출해서 전달해주는 ‘인출책’ 내지 ‘전달책’으로 구성된 범죄단체에 의해 이루어진다. IT 기술이 발달하면서 고도로 분업화·익명화된 보이스피싱 사기 조직을 검거하는 것은 한층 어려워졌다.

그러나 역설적이게도 IT 기술은 범행을 뒤쫓을 수 있는 가능성을 제공한다는 양면성을 가지고 있다. 범행과정에서 무심코 노출된 정보, 예컨대 통화내역·IP 주소 등과 같은 통신자료 내지 통신사실확인자료³, 계좌거래내역 등을 발견한다면 범죄수사를 개시하기 위한 단서로 활용해볼 수 있다. 사회연결망분석은 막연히 서로 관련되어 있을 것으로 추정되는 단서들이 실제 긴밀하게 관련되어 있음을 시각적으로 보여준다.



3 수사기관이 통신자료(이용자의 성명, 주민등록번호, 주소, 전화번호, 아이디, 가입 또는 해지일자) 내지 통신사실확인자료(이용자의 IP 주소 등)를 요청하는 것은 전기통신사업법 제83조(통신비밀의 보호)에 근거한다. 이때 수사기관은 법원이 발부한 영장을 제시하며 통신자료 또는 통신사실확인자료 제공을 요청한다.

i2를 통한 전화통화 연결망 분석결과 예시



출처: 한국 IBM 홈페이지 <https://www.ibm.com/kr-ko/marketplace/analysts-notebook>, 2020년 10월 21일 접속.



4 IBM은 2011년 범죄정보분석을 전문으로 하는 회사인 i2를 인수하면서 이 회사가 개발한 동명의 프로그램명을 그대로 유지했다.

경찰이 수사 현장에서 범죄자의 인적 관계를 분석하기 위해 사용하고 있는 프로그램은 IBM의 i2⁴이다. 경찰관들은 범죄를 수사하는 과정에서 수시로 대량의 데이터를 살펴봐야 하는 상황에 처하는데, 전문적인 데이터 분석기법을 익힌 경우가 드물다 보니 며칠 밤낮을 고심한다 하더라도 데이터에서 유의미한 결과를 도출하는 데 한계가 있었다. 수집된 데이터의 형태가 일관되지 않은 것 또한 데이터 속에서 빠르게 범죄의 단서를 찾아야 하는 경찰관들에게는 곤혹스러운 문제였다.

이에 경찰은 2015년 데이터 간 관계를 체계적이면서도 신속하게 분석하기 위한 목적에서 i2를 도입하였다. 그와 동시에 경찰은 수집된 데이터들을 사회연결망분석에 활용할 수 있도록 클리닝 및 표준화하는 프로그램(i1) 및 관찰 경찰관서에서 인지된 범죄사건에 대하여 다른 경찰관서에 공조수사 및 정보공유를 요청하는 시스템(i3)을 함께 구축했다.

모든 범죄는 흔적을 남긴다

사회연결망분석에 기반한 i2는 위 그림에서 보는 바와 같이 분산 내지 분절된 형태로 수집된 데이터들이 서로 연결되어 있음을 간명하게 보여줌으로써 경찰이 범죄단체 구성원들의 역할을 파악하고 범죄수익의 흐름을 추적하는 데 도움을 주고 있다.

경찰은 i2로 분석된 데이터들 간의 연관관계를 통해 숨은 공범들을 찾아내기도 하고, 의심되는 자금의 흐름을 추적하며, 수사를 보강해야 할 부분을 발견하기도 한다. 또한 시각화된 피의자들의 연결 관계는 이들의 범행 관련성을 입증하는 증거로도 활용될 수 있다.

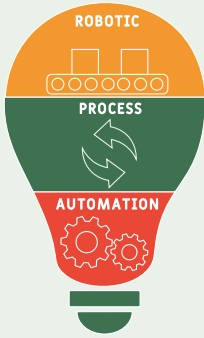
‘모든 범죄는 흔적을 남긴다’는 말이 있다. 아주 작은 단서가 범죄사건을 해결하는 결정적인 실마리가 될 수 있음을 의미한다. 오프라인에서의 범죄자는 지문이나 족적을 남긴다면, 온라인에서의 범죄자는 데이터를 남긴다는 차이가 존재할 뿐이다. 조직화·지능화된 범죄는 아주 작은 데이터 조각도 좀처럼 발견하기 어렵게 하지만, 경찰은 범행과정에서 생성된 사소한 데이터라도 무시코 넘기지 않고 그 연관관계를 파악하기 위한 노력을 기울이고 있다.



업무 자동화 경진대회 다양한 통계업무 자동화 사례를 만나보자

RPA





지금은 4차산업혁명 시대다. 민간에서는 상품 제조과정에서 실시간으로 데이터를 연계한 업무 자동화를 통해 생산 효율성을 극대화하고 있다. 행정도 예외일 수 없다.

2020년 범정부 정부혁신 종합추진계획(행정안전부, 관계부처 합동)에 따르면, 정부혁신의 새로운 역점추진분야로 '디지털기반 행정업무 효율화'가 선정되었다. 단순반복적인 업무를 자동화하여 국민을 위한 정책과 서비스를 확산하는, 일하는 방식의 혁신이 핵심 내용이다.

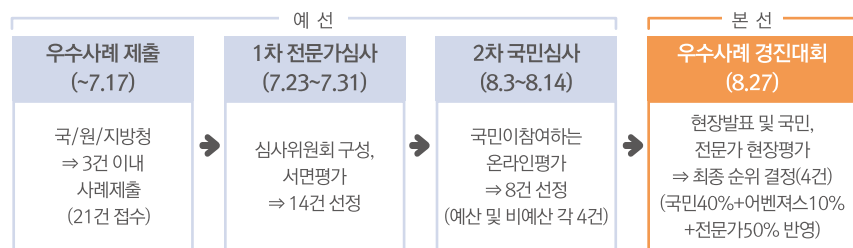
통계청은 인구, 주택, 산업, 고용 등 국민의 삶과 연결되는 방대한 데이터를 활용하여 정책에 필요한 통계와 지표를 만들어내는 기능을 수행하고 있다. 기초자료의 수집, 정제, 연계, 분석, 생산, 서비스의 통계업무 전반에서 다양한 업무자동화 사례들이 나타나고 있고, 이를 더욱 체계화할 필요가 있다.

통계청은 데이터뿐만 아니라 데이터 분석 역량을 갖춘 충분한 인적자원도 보유하고 있다. 이런 환경과 여건에 따라 통계청은 중앙부처 중에서 처음으로 업무자동화 경진대회를 개최하게 되었다.

「제1회 통계청 업무자동화 경진대회」는 언제, 어떻게 진행되었을까?

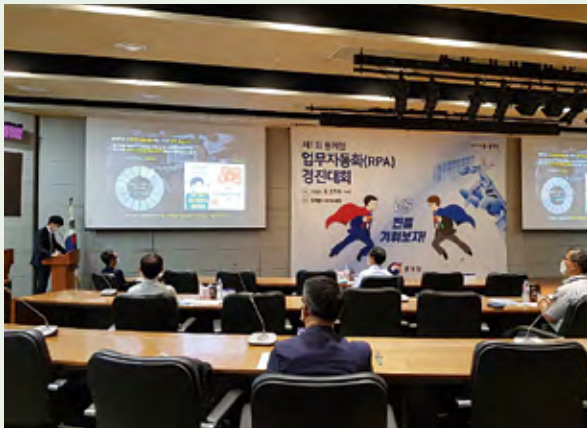
통계청은 지난 7월 17일까지 총 21건의 업무자동화 혁신사례를 접수하였다. 이 중 1차 전문가 심사(7.23~7.31)를 통해 14건으로 압축한 후, 전 국민을 대상으로 2차 온라인 국민심사(8.3~8.14)를 진행하여 최종 8건을 선정하였다.

경진대회 전체 추진 일정



이렇게 선정된 8건에 대하여 지난 8월 27일 통계센터 국제회의실에서 ‘제1회 통계청 업무자동화 경진대회’ 발표심사를 개최하였다. 이 행사는 코로나19 방역수칙을 지켜가며 전국에서 모인 국민평가단이 지켜보는 가운데 2시간 동안 조심스럽게 진행되었다.

심사는 전문가, 국민평가단, 청내 정부혁신 서포터즈 역할을 수행하고 있는 정부혁신 어벤저스 등 3개 그룹으로 나누어 진행하였고 전체적으로 가장 우수한 평가를 받은 4건이 시상(통계청장상)을 받게 되었다.



통계청은 심사에 참여한 국민평가단의 이해를 높이기 위해 인트로 동영상을 제작하고, 사례내용을 요약한 리플릿과 자료집을 배부하였다. 또한 통로에는 배너를 설치하여 궁금해하는 국민의 질문에 자세한 설명을 덧붙여 이해를 높였다.

전문심사단으로 참가한 정부혁신 주관부처인 행정안전부(협업정책과)와 국내 RPA 대표기업인 UiPath, 포스코ICT 대표는 심사평에서 “통계청이 데이터 활용기관의 특성을 잘 살려 디지털 정부를 선도하고 있다”고 대회성적을 긍정적으로 평가하였다.





엄격한 심사를 거쳐 선정된 우수사례는 무엇일까?

이번 대회 평가과정에서 가장 중시한 부분은 국민 참여와 국민평가였다. 2차 온라인 심사에서도 국민평가 결과를 100% 반영하여 본선 진출 사례를 선정하였고, 현장 발표심사에서도 국민평가단을 심사그룹에 포함하여 국민의 시선에 부합하는 사례를 발굴하기 위해 노력하였다. 국민이 직접 우리나라 통계행정의 변화와 혁신과정을 지켜보며 공감하는 것이 국가통계 품질과 신뢰성 제고에 가장 중요하다고 판단한 것이다.

제1회 통계청 업무자동화 경진대회 개최결과

- ① 최우수 국가통계서비스 챗봇서비스 '국민통계비서관'(통계서비스기획과)
- ② 우수 행정자료처리 자동화 프로그램 '데이터원클릭'(충청지방통계청 지역통계과)
- ③ 장려 행정자료 산업분류 코딩 자동화(행정자료관리과),
ICT를 활용한 버 날알 수 측정 자동화(동북지방통계청 농어업조사과)



통계청 업무자동화는 어떻게 이루어지고 있을까?

RPA(Robotic Process Automation)는 '각종 데이터의 수집·입력·비교 등 단순하고 반복적으로 처리해야 하는 업무를 기계가 자동으로 처리할 수 있도록 프로세스를 자동화하여 비용과 시간을 줄이고 생산성을 높이는 업무 혁신'이다.

이 대회는 통계청에서도 올해 처음으로 개최하는 행사이기에, 시행 초기에 업무자동화의 개념과 사례 정의에 대해 각 부서에서 문의하는 경우가 많았다. 그러나 업무자동화(RPA)의 의미를 살펴보자면 통계생산과 서비스 업무수행과정에서 나타나는 일상적인 업무수행 사례들이 많다. 그만큼 업무자동화는 통계행정 내부에 깊숙이 들어와 있다.

이번에 접수된 사례를 살펴보면, 통계기초자료 수집과정에서 통계 대상처로부터 자동으로 자료를 수집하거나 수집된 자료의 검증에 활용한 사례가 많았고, 조사과정에서 일일이 수작업으로 수행해야 하는 업무를 ICT 기술을 활용하여 자동화한 사례들도 선보였다.

통계 작성과정에서도 국세청의 업종분류코드를 한국표준산업분류로 자동연계하는 분류코딩 자동화, 행정자료 우선활용제도 시행을 위한 행정 및 조사자료의 항목검색 자동화 등 행정자료를 통계자료와 접목하는 기능에도 활용되고 있었다.

통계서비스 과정에서는 챗봇시스템을 도입하여 이용자가 복잡한 과정을 거쳐 확인할 수 있는 통계자료를 간단한 질문과 대답으로 바로 확인할 수 있는 국민통계비서관이 돋보였다. (참고: 통계청 업무자동화 경진대회 본선 진출 우수사례 8건)

「제1회 업무자동화 경진대회」 개최가 갖는 의미

업무자동화가 범정부 정부혁신 중점과제로 부각되는 상황에서 통계청이 그에 부응하여 중앙부처 최초로 자체 경진대회를 개최한 것에 자부심이 크다.

업무자동화에는 여러 단계가 있다. 초보적인 매크로에서부터 첨단 AI 활용에 이르기까지 자동화의 수준과 정도는 차이가 난다. 행정도 각기 처한 상황과 목표가 다르므로 업무효율화를 위한 적합한 자동화 기술을 이용하면 된다. 관심을 갖고 업무자동화 사례들을 차근차근 모으고 공유해 나간다면 더욱 수준 높고 좋은 사례들이 나타날 수 있을 것으로 기대한다.

벌써 몇몇 기관에서는 통계청의 경진대회 소식을 듣고 관련 자료를 요청하는 등 벤치마킹을 시도하고 있어서, 기관 간 사례를 공유하고 확산하는 계기가 마련되었으면 한다.





두 마리 토끼가 한 손에... 업무 효율은 높이고 비용은 줄이고

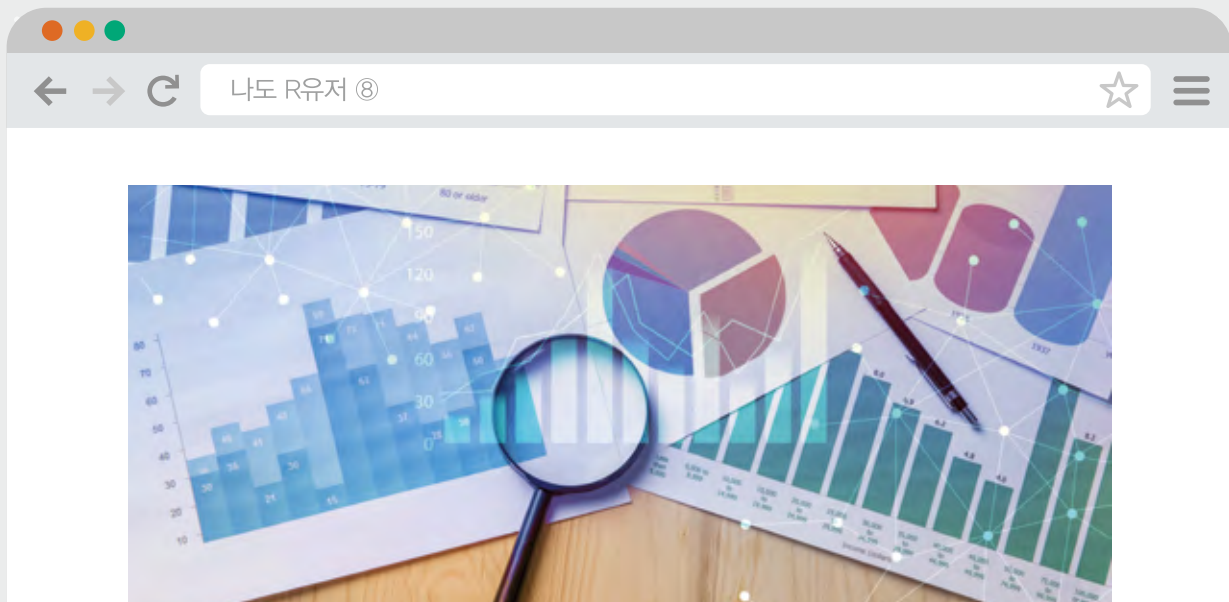
통계청은 중앙통계기관으로서 데이터와 떼려야 뗄 수 없는 기관이다. 데이터의 수집과 연계, 통계의 생산 및 서비스에 이르기까지 모두 데이터와 관련되어 있기에 업무자동화를 통해 비용과 시간을 줄이고 시의성 있고 정확한 통계를 생산하는 미션을 수행해야 한다.

올해 개최된 ‘제1회 통계청 업무자동화 경진대회’에서 나타난 사례들을 공유하며, 더 다양한 사례가 발굴·확산될 수 있도록 통계청 정부혁신의 중점과제로 관리를 지속해 나갈 예정이다.

업무자동화 경진대회를 통해 통계청이 데이터 수집·연계·활용 등 국가데이터 허브로 자리매김하기를 기대해본다.

참고 | 통계청 업무자동화 경진대회 본선진출 우수사례(요약)

제목	주요 내용
중앙부처 통계예산 One-stop 검토시스템 구축	<ul style="list-style-type: none"> • (자동화 내용) 통계예산 정보시스템 구축 및 기존 시스템과 연계, One-Stop 웹 시스템 구축을 통해 예산편성안을 일괄처리하고 진행 상황에 대한 실시간 집계 및 정보 공유 • (기대 효과) One-stop 시스템 기반으로 표준화 양식을 사용하여 편성-제출-검토 단계 수행으로 오류 발생 가능성을 사전에 차단, 기존정보 연계 및 자동계산 기능 지원으로 문서량, 업무 소요일 획기적 감축 달성
행정 및 조사자료의 항목 검색 자동화	<ul style="list-style-type: none"> • (자동화 내용) 행정자료 우선활용 제도 시행('18.2)에 따른 행정자료 우선활용 시스템을 통해 조사항목DB 및 메타DB구축, 검색 기능 구축 • (기대 효과) 행정자료 우선활용 시스템을 통한 행정자료 우선활용 판단 및 통계작성에 필요한 행정·조사자료의 항목검색(12천여 개 법령서식, 24만여 개 항목) 자동화
행정자료 산업분류 코딩 자동화	<ul style="list-style-type: none"> • (자동화 내용) 국세청 업종분류코드 개편을 통한 한국표준산업분류로 연계 및 분류코딩 자동화 • (기대 효과) 국세청과의 협업을 통해 자동화 절차를 정교화하여 자동화 비율 제고 국세청 업종코드와 산업분류코드 1:1 매칭비율은 64%에서 98%로 높아지고 업무처리 투입 인원도 70.5명('09)에서 11명('20)으로 대폭 감소
국가통계서비스 챗봇서비스 '국민통계비서관'	<ul style="list-style-type: none"> • (자동화 내용) 변화하는 ICT 환경에 맞추어 국가통계포털에 차세대 지능화 기술인 챗봇시스템 도입, 지능형 통계비서관은 누구에게나 대화형 AI 기반으로 원하는 통계정보를 추천 및 안내해주는 통합서비스 개발 • (기대 효과) 4차혁명 인공지능 서비스기반 조성으로 국가통계포털 이용건 수 10% 이상 증가 및 통계표 검색시간은 10초대로 가능
AI를 활용한 건설경기동향조사 민간자료 수집 자동화	<ul style="list-style-type: none"> • (자동화 내용) 건설 민간 발주내역 수집 및 AI 기법을 적용한 자료 필터링 시스템 개발 • (기대 효과) 자료수집 시간이 16시간에서 30분으로 단축, 자료정제 시간이 30시간에서 5분 내로 단축, 또한 폭넓은 자료수집으로 누락 방지
ICT를 활용한 벼 낱알수 측정 자동화	<ul style="list-style-type: none"> • (자동화 내용) 짧은 조사기간 내에 작은 벼 낱알을 일일이 수작업으로 생산량 조사를 하여 조사자별 업무피로도가 상당하고 계수 착오를 방지하고자, Image J* SW를 이용한 낱알 수 계수 실험을 통해 낱알 수 측정 자동화 • (기대 효과) 수작업 대비 ICT 계수시간은 1,197.46분(평균 19.96분), 90.52% 감소, 통계청 혁신과제로 채택되어 현장 친화형 방법론 개발을 목표로 '20년 시험조사 실시
예산집행관리 선진화를 위한 '클린UP' 예산 내검 자동화	<ul style="list-style-type: none"> • (자동화 내용) 예산집행 내용검토를 위한 반복적 업무를 '클린Up' 예산내검 프로그램을 개발하여 회계업무의 체계적 관리 • (기대 효과) 프로그램 시뮬레이션 결과 당초 내검결과와 91.3%의 정합성, 자료 검토 시간이 8시간에서 5분으로 감소
행정자료처리 자동화 프로그램 '데이터 원클릭'	<ul style="list-style-type: none"> • (자동화 내용) 통계역량이 부족한 지자체에 행정통계 자체생산을 지원하기 위하여 행정자료 자동화 프로그램「데이터 원클릭」 개발·지원 • (기대 효과) 행정자료처리 및 분석기법 등 지자체 기술전수로 통계역량강화 및 고품질 행정통계 확충에 기여



R에 도전하자... 따라가다 보면, 나도 R유저 ⑧

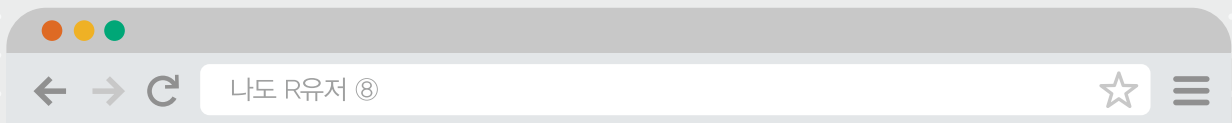
‘R에 도전하자 ①’부터 ⑦에서 R의 설치부터 시작하여 몇 가지 기능을 따라 해보고 R에서의 색깔 처리, R의 기초적인 통계함수 및 R의 벡터 및 데이터 프레임을 사용한 자료처리, 함수작성, 그래픽 장치 등에 대해 알아보았다. 이번 호에서는 지난 호의 기본 그래픽에 ggplot2 패키지를 사용한 R 그래픽에 대해서 알아보기로 하자.

1 | ggplot2 패키지

ggplot2 패키지는 R의 기본 그래픽을 확장하여 데이터 시각화에서 기본 그래픽보다 더 나은 형태의 그림을 만들 수 있고 사용법이 상당히 직관적이다. ggplot2는 R base 설치 시에 설치되지 않는 패키지이기 때문에 R의 기본 설치만으로는 사용할 수 없고 이 패키지를 설치하여야 한다. 패키지의 설치에는 R에서

```
> install.packages("ggplot2")
```

라고 명령하면 설치할 패키지를 내려받을 사이트의 목록이 나오고 이 목록에서 아무 곳이나 선택하면 인터넷을 통해 패키지를 내려받아 로컬 컴퓨터에 설치가 된다. ggplot2를 설치할 때 ggplot2가 필



요로 하는 패키지들도 함께 설치되므로 여러 개의 관련 패키지가 자동 설치된다. 패키지 설치가 완료 되었으면 이 패키지를 탑재하여야 한다. 패키지 탑재는 `library` 함수를

```
> library(ggplot2)
```

로 사용하여 `ggplot2`를 탑재하며 설치된 R의 버전이 `ggplot2`의 버전보다 오래된 것이면 경고 메시지가 출력되기도 하는데(이는 모든 R 패키지에서 동일) 예를 들면 R 4.0.2에서 `ggplot2`의 R 4.0.3 버전이 탑재하면 아래의 경고 메시지가 나온다.

경고메시지(들):

패키지 'ggplot2'는 R 버전 4.0.3에서 작성

R 버전보다 더 최신의 패키지 버전을 사용하는 경우 대부분 큰 문제없이 사용 가능하지만 이런 경우 R을 최신 버전으로 업그레이드 하는 것을 권장한다.

위의 `install packages` 명령은 한 번만 실행하면 관련 프로그램이 내 컴퓨터에 설치되므로 추가로 실행하지 않아도 되지만 `library` 함수는 R을 켜다가 켜 때 새로 호출하여야 한다.

2 | qplot 함수



`ggplot2` 패키지를 사용할 때 간단히 시각화를 하고자 할 때 사용할 수 있는 함수로 다음과 같이 사용한다. `qplot` 함수는 `quickplot` 함수의 줄임말로 뒤에서 소개할 `ggplot` 함수를 사용하지 않고 빠르게 `ggplot`과 해당 `geometry`를 적용하고자 할 때 사용한다. 사용법은

```
qplot(x, y = NULL, ..., data, facets = NULL, geom = "auto",  
      xlim = c(NA, NA), ylim = c(NA, NA), log = "", main = NULL,  
      xlab = deparse(substitute(x)), ylab = deparse(substitute(y)), ...)
```

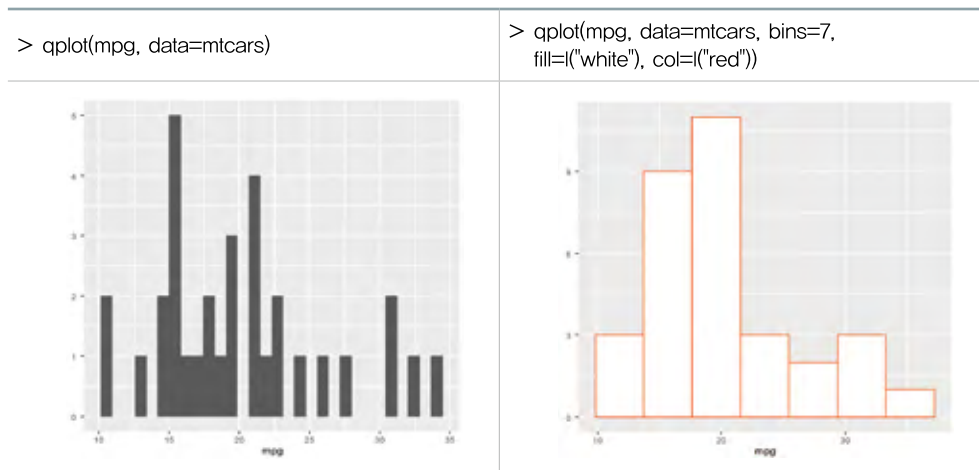
이며 각 매개변수는

- `x, y`: 그림에 사용할 속성들. 산점도 등에서는 `x, y`가 모두 필요하나 히스토그램 등에서는 `x`만 있어도 된다.
- `data`: 사용할 데이터 프레임의 이름



- **facets**: 그룹별로 그림을 그리고자 할 때 설정한다. 주로 행렬산점도를 그릴 때 행과 열에 사용할 변수를 row ~ col 형식으로 설정한다.
- **geom**: 무엇을 그릴지 설정한다. 기본값은 x, y가 주어졌으면 산점도("point"), x만 설정되었으면 히스토그램("histogram")이 그려진다. 이 둘외에 사용할 수 있는 값은 "path", "boxplot", "jitter", "dotplot", "smoothe" 등이 있다.
- **asp**: y축과 x축의 비율 y/x. (aspect ratio).
- **xlim, ylim, main, xlab, ylab**: plot 함수와 같은 설정. 지난 호 'R에 도전하자 ⑦' 참고.
- ...: 각 geom에서 사용할 추가 매개변수들. ggplot2를 구글링하면 ggplot2 패키지의 설명서가 284 페이지의 PDF로 제공되므로 이를 참고하기로 한다.

보기 1: R에 내장된 데이터 프레임 mtcars는 Motor Trend에서 조사한 자동차의 연비, 마력,=배기량 등의 자료가 저장되어 있다. help(mtcars)로 자세한 정보를 확인할 수 있다. 이 자료에서 연비(변수 mpg, 1 갤런당 주행거리 mile)의 히스토그램을 그려보자.

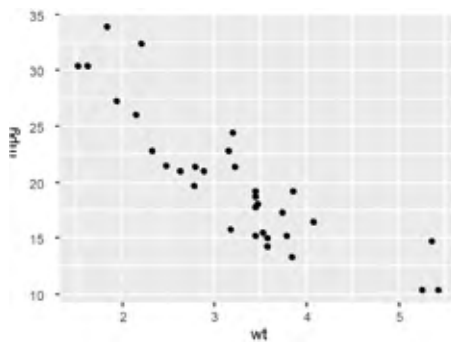


왼쪽 그림은 아무 옵션 설정 없이 mtcars의 변수 mpg에 대한 히스토그램을 그린 것이며 오른쪽 그림은 bins에서 기둥의 수를 7개로, 히스토그램의 기둥에 대해서 내부는 흰색, 기둥 테두리는 빨간색으로 설정한 것이다.

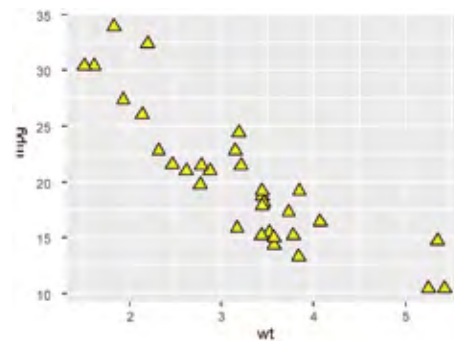
보기 2: (x, y를 함께 설정하여 산점도 얻기) 자동차의 무게(wt)와 연비(mpg)의 산점도를 다음과 같이 그려보자.



```
> qplot(wt, mpg, data=mtcars)
```

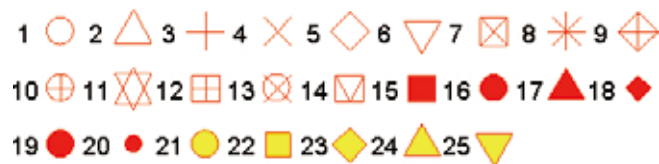


```
> qplot(wt, mpg, data=mtcars, size=(3),  
shape=(24), fill=(\"yellow\"))
```



왼쪽 그림은 기본값으로 x축은 자동차의 무게(wt), y축은 연비로 하여 산점도를 그린 것이며 오른쪽은 점의 모양은 24번, 점의 크기는 3, 내부의 색은 노란색으로 추가 설정한 것이다. 산점도에서 숫자에 따른 점의 모양은 다음과 같으며 21-25번은 점 내부색 추가로 설정할 수 있는데 아래 그림은 노란색으로 설정한 결과이다.

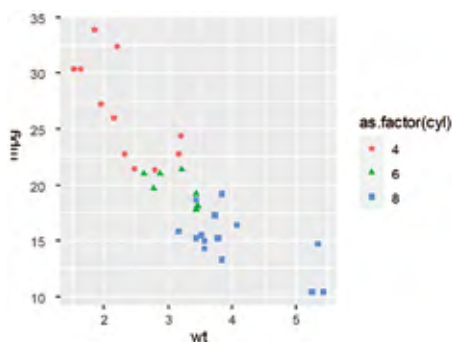
Symbols for Point CHaracter pch = 숫자



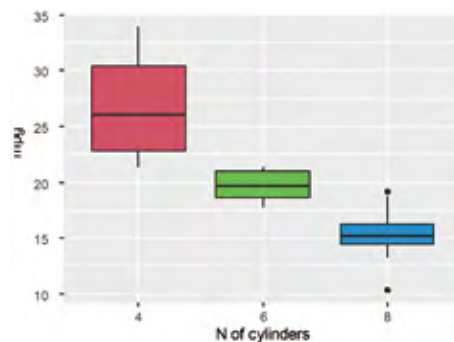


보기 3: (그룹별 산점도 및 상자그림 그리기) mtcars에의 cyl 변수는 해당 자동차 엔진의 실린더 수가 저장되어 있다. 이 숫자가 4이면 4기통, 6이나 8이면 각각 6기통 및 8기통이란 뜻이다. 실린더 개수에 따른 wt와 mpg의 산점도와 mpg의 상자그림을 얻어보자. 왼쪽 그림은 x축에 wt, y축에 mpg를 사용한 산점도를 그리는 데 shape과 color에 cyl을 사용함으로써 cyl의 값에 따라 점의 모양(shape)과 점의 색(color)을 다르게 하였다. 범례는 자동으로 생성된다. 오른쪽 그림은 상자그림을 그리되 x축의 값을 cyl로 둬으로써 cyl의 값에 따른 상자그림을 그린다(당연히 이 경우 geom에 "boxplot"으로 설정). 색을 달리하기 위해 fill에 색의 번호 2, 3, 4(=2:4)를 설정하였다.

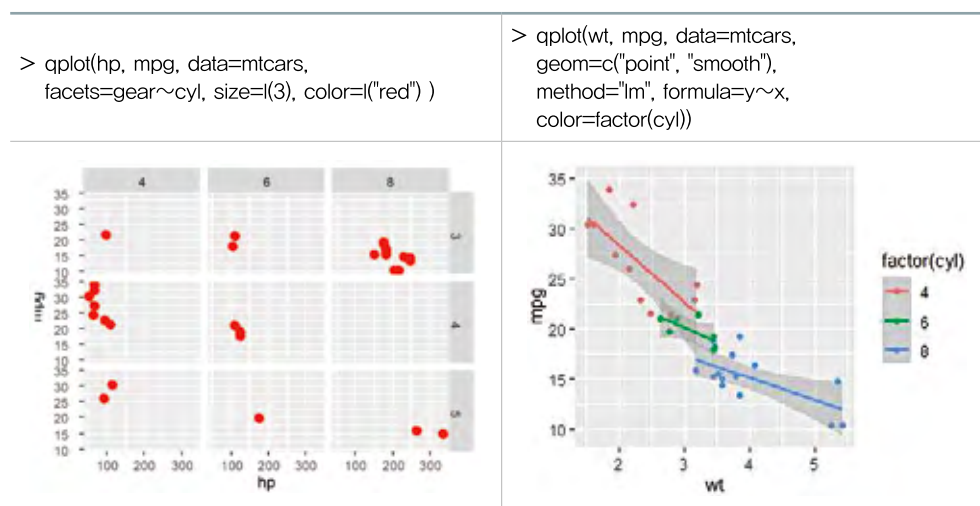
```
> qplot(wt, mpg, data=mtcars,
  shape=as.factor(cyl),
  color=as.factor(cyl))
```



```
> qplot(factor(cyl), mpg, data=mtcars,
  geom="boxplot", fill=(2:4),
  xlab="N of cylinders")
```



보기 4: (행렬 산점도와 회귀선 그리기) 행렬 산점도 `facets`에 규칙을 설정한다. `facets`에 설정된 범주에 해당되는 자료에 대해 `x, y`에 설정한 산점도를 그린다. 왼쪽 그림은 `facets`에 `gear~cyl`로 설정되어 `gear`(변속기의 전진 기어의 단수, 4단 기어면 4)의 값은 `y`축, `cyl`의 값은 `x`축에 사용하여 이 조합에 속한 자료에 대해서 `hp`에 따른 `mpg`의 산점도를 그린다. 예를 들어 아래의 왼쪽 그림에서 왼쪽 위의 산점도는 기어가 3단이고 4기통인 자동차의 `hp`와 `mpg`의 산점도이다.



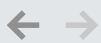
오른쪽 그림은 산점도에 추가하여 산점도를 잘 설명하는 회귀직선(곡선 포함)과 신뢰구간을 그리는 보기이다. 이 경우 산점도와 적합한 회귀선을 함께 그리기 위해 `geom`에 "point"와 "smooth"를 모두 설정하였다. `point`에 의해서 산점도가 그려지고 `smooth`에 의해서 자료를 적합한 회귀선이 얻어진다. `smooth`에 사용할 방법(`method`)과 식(`formula`)은 각각 "lm", `y~x`로 설정되어 선형모형(회귀분석)과 $y = \beta_0 + \beta_1 x$ 를 적합한다. 적합할 모형으로는 "lm", "glm", "gam", "loess" 등이 사용 가능하다. 일부 버전의 R에서 `method`와 `formula`를 인식하지 못한다는 경고가 나오는데 실제로는 인식이 된다.

3 | ggplot 함수



앞에서 본 `qplot` 함수는 `ggplot` 함수를 사용하여 필요한 그래픽 변수를 각각 설정하지 않고 간단히 그릴 때 사용할 수 있는 함수이다. 좀더 세밀한 설정이 필요할 때는 `ggplot` 함수를 사용하는 것이 권장된다. `ggplot` 함수는

```
ggplot(data = NULL, mapping = aes(), ..., environment = parent.frame())
```



로 사용하며

- **data**: 사용할 데이터 프레임 이름
- **mapping**: 그래픽에서 사용할 기본 mapping을 설정하며 aes로 설정한다. aes는 aesthetic의 줄임이다. 설정하지 않은 경우 뒤따르는 각 레이어에서 aes를 설정하여야 한다.
- **environment**: 만일 ggplot 함수에서 사용한 변수 이름이 없을 경우 이곳에서 설정한 environment에서 찾는다.



ggplot 함수는 시각화할 기본 영역을 설정한다고 보아야 하며 구체적인 시각화는 +를 사용하여 geom을 설정한다. 사용할 수 있는 그림은 geom_point, geom_line, geom_path, geom_smooth, geom_histogram 등이며 각각 점, 선(함수 그리기에 유용), 경로(다각형), 산점도에 대한 모형적합, 히스토그램 등을 그리게 된다. 예를 들어 산점도를 그리려면

```
> ggplot(ggplot 설정) + geom_point(geom_point의 설정)
```

으로, 히스토그램을 그리려면

```
> ggplot(ggplot 설정) + geom_histogram(geom_point의 설정)
```

으로 ggplot 함수 +를 사용하여 무엇을 그릴지 추가로 설정한다. +는 필요한 만큼 추가할 수 있다. 즉, 점을 그리고 선도 그리고 싶으면

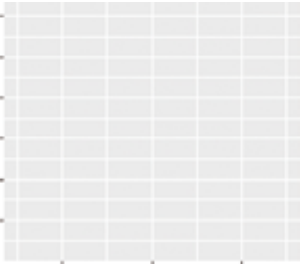
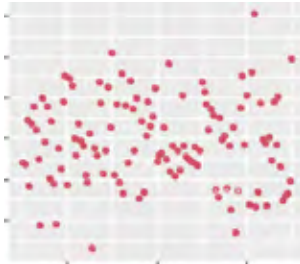
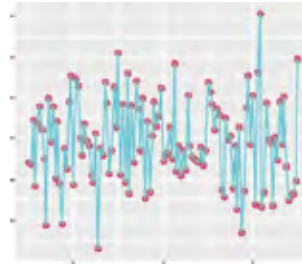
```
> ggplot(ggplot 설정) + geom_point(geom_point의 설정) + geom_line(geom_line 설정)
```

으로 하면 된다. 예를 들어 x와 y1을 다음과 같이 -3부터 3사이의 값과 정규분포의 난수로 구성된 변수로 하고 이 둘을 변수로 갖는 데이터 프레임이 있다면

```
> x <- seq(-3, 3, by=0.05); y1 <- rnorm(length(x))
```

```
> df <- data.frame(x, y1)
```

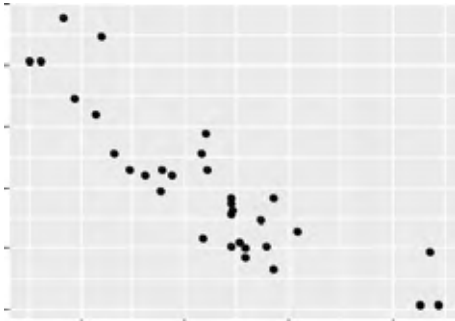
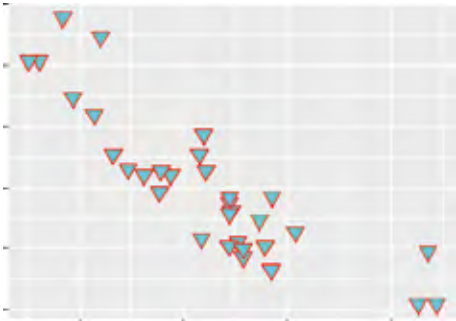
아래의 그림과 같이 ggplot 함수만 사용하면 산점도의 영역만 확보하고, geom_point를 +할 때 점이 추가된다. 여기에 geom_line을 추가하면 점들을 연결한 선이 그려진다.

<pre>> ggplot(df, aes(x, y1))</pre>	<pre>> ggplot(df, aes(x, y1)) + geom_point(col=2)</pre>	<pre>> ggplot(df, aes(x, y1)) + geom_point(col=2) + geom_line(col="cyan")</pre>
		

이 그림에서 보는 것처럼 `ggplot` 함수만 사용하면 실제 그림영역에 점이든 선이든 아무 것도 생기지 않고, `geom_point`를 +하면 비로서 점이 그려지며, 여기에 `geom_line`을 +하면 선이 추가되는 방식이다. 따라서 한 번의 `ggplot` 함수에 +를 사용하여 필요한 만큼의 그림 요소를 그릴 수 있다.

보기 5: ggplot을 사용한 기본 산점도

산점도는 `ggplot`이라는 이름이 의미하는 것처럼 `ggplot`의 기본적인 기능으로 `aes`에 x축, y축에 사용할 변수를 설정하면 된다. 점을 그리기 위해서는 `geom_point`를 추가한다. 왼쪽 그림은 기본 산점도를 그린 것이고, 오른쪽 그림은 점의 모양, 크기, 색 등을 설정하였다. 점의 색은 "red", 크기는 3, 모양은 25번, 점내부의 색은 "cyan"이다. 산점도에서 사용하는 점들 중 21번부터 25번은 내부의 색을 추가로 설정할 수 있다.

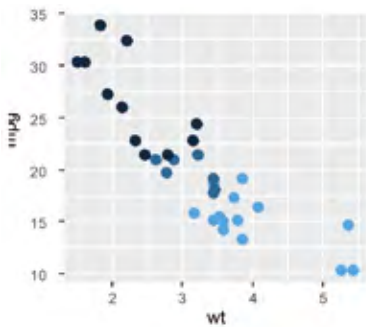
<pre>> ggplot(mtcars, aes(wt, mpg)) + geom_point()</pre>	<pre>> ggplot(mtcars, aes(wt, mpg)) + geom_point(color="red", size = 3, shape=25, fill="cyan")</pre>
	



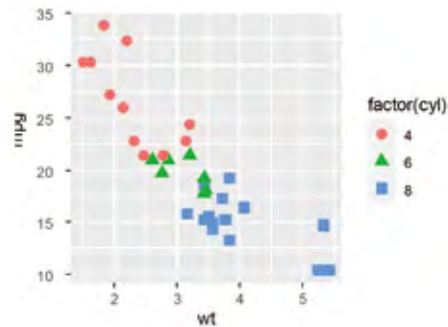
보기 6: (그룹별 산점도: 그룹변수의 값이 연속형인 경우와 이산형인 경우) 산점도를 그릴 때 x축 변수와 y축 변수 외에 제3의 변수의 값에 따라 산점도의 색이나 점의 형태를 다르게 할 수 있다. mtcars에서 cyl의 값에 따라 산점도를 그리되 cyl의 값을 연속으로 간주할 때와 범주형으로 간주할 때의 산점도는 아래와 같다. 왼쪽 그림은 cyl의 값이 연속으로 간주했을 때의 그림으로 color에 cyl을 설정하여 cyl의 값에 따라 색의 농도를 다르게 하였다. 범례는 자동으로 생성된다.

오른쪽 그림은 cyl이 범주형일 때(factor 함수로 cyl의 값을 명목형으로 변경)의 그림으로 cyl의 값에 따라 점의 색깔과 모양을 다르게 하도록 color와 shape을 설정하였다. 이 경우도 범례는 자동으로 만들어진다.

```
> ggplot(mtcars, aes(wt, mpg, color=cyl  
  shape=cyl)) + geom_point(size = 3, shape=19)
```



```
> ggplot(mtcars, aes(wt, mpg,  
  color=factor(cyl), shape=factor(cyl))) +  
  geom_point(size = 3)
```



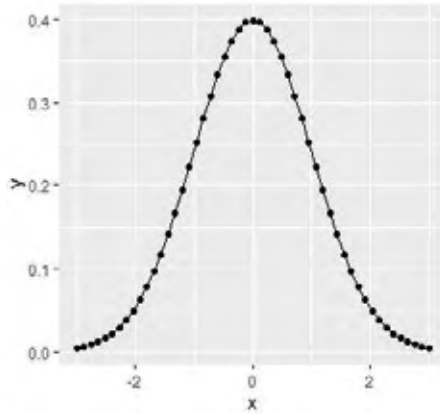
보기 7: (함수 그리기) 함수 $y=f(x)$ 를 그리는 경우를 생각해보자. 이 경우 x축에는 x의 값, y축에는 $f(x)$ 를 사용하게 된다. 예를 들어 $y=f(x)=x(x-2)(x+2)$ 인 함수를 (-3,3) 사이에 그리려면

```
> x <- seq(-3, 3, length=51)  
> y <- x*(x-2)*(x+2)# dnorm(x)  
> df <- data.frame(x=x, y=y)
```

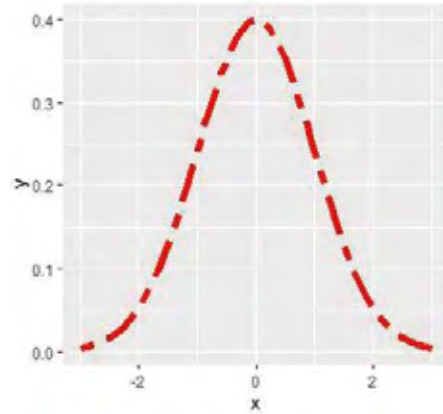
로 df 라는 데이터 프레임 생성 후 geom_line을 사용하는 것이 기본이다. 다음 그림의 왼쪽은 선과 점을 동시에 그릴 경우이며, 오른쪽은 점은 그리지 않고, 선의 형태에 대한 설정을 추가하여 선의 색(color)은 빨강, 선의 굵기(size)는 2, 선의 종류(linetype)는 12로 설정한 결과이다.



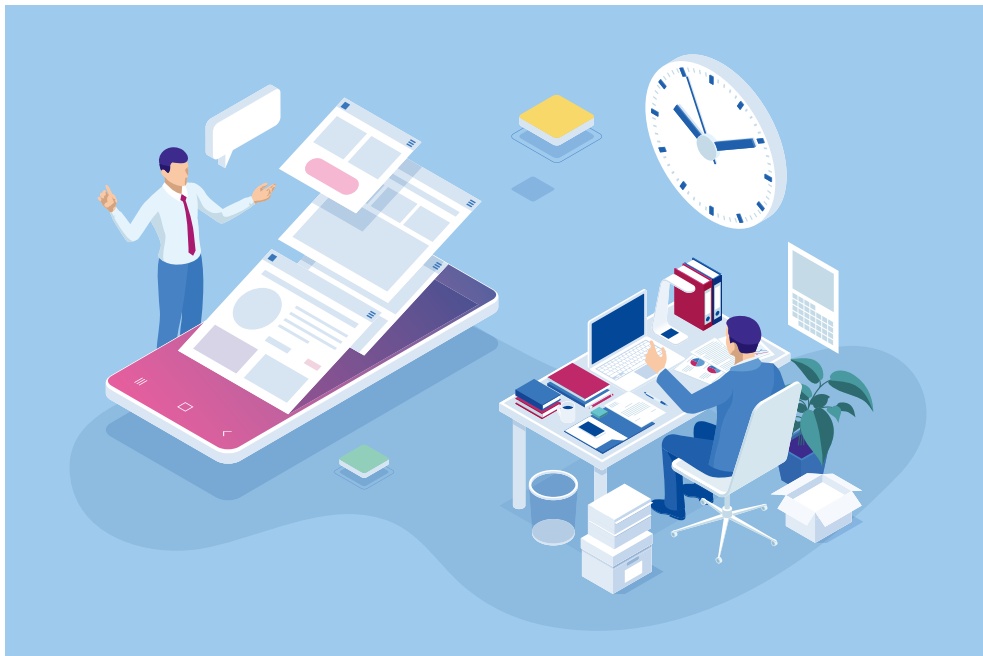
```
> ggplot(df, aes(x=x, y=y)) +  
  geom_point() + geom_line()
```

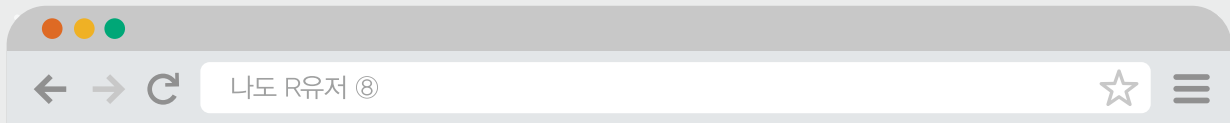


```
> ggplot(data=df, aes(x=x, y=y)) +  
  geom_line(color="red", size = 2,  
  linetype=12)
```



R의 기본적인 선의 형태는 0부터 6까지 설정 가능한데 0은 선이 없는 것이므로 6가지의 선 형태를 제공한다. ggplot의 경우 12개의 선의 형태를 사용할 수 있으며 각 번호에 따른 선의 형태는 다음과 같다.





ggplot의 종류



보기 8: (산점도에서 직선 또는 곡선으로 모형적합 추가) 산점도에 선형회귀분석 등의 모형적합 결과로 얻는 방정식을 추가하려면 `geom_smooth`를 사용한다. `geom_smooth`는

```
geom_smooth(method=NULL, formula=NULL, se=TRUE, level=0.095, ...)
```

로 사용하며

- `method`에는 모형 적합 방법을 설정하며 이 값은 "loess"가 기본값이며, "lm", "glm", "gam", "loess"을 사용하거나 `package::function`과 같이 패키지 이름 `package`와 해당 패키지 함수 이름을 설정하여 모형을 적합한다. 예를 들어 선형회귀분석함수를 사용할 때 `method="lm"`이라고 하거나 `method=stats::lm`으로 한 것은 같은 결과를 얻는다(lm 함수는 R base 설치에서 자동설치되는 stats 패키지의 함수이다).
- `formula`에는 적합할 모형을 설정한다. 기본값은 $y \sim x$ 이며, $y \sim \log(x)$, $y \sim \text{poly}(x, 2)$ 등으로 설정한다. $y \sim \text{poly}(x, 2)$ 는 2차 다항식 $y = \beta_0 + \beta_1 x + \beta_2 x^2$ 을 적합한다.
- `se`에는 적합된 방정식의 신뢰구간을 출력할지 설정한다.
- `level`에는 `se`로 신뢰구간을 그릴 때 신뢰구간의 신뢰도를 설정한다.

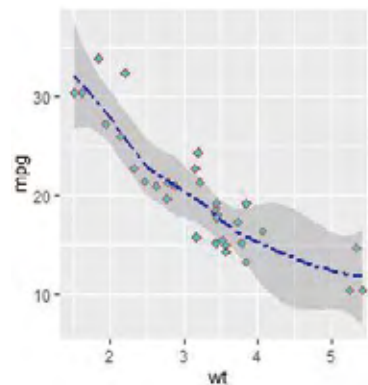
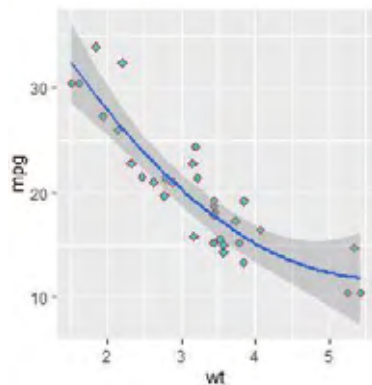


아래의 실행에는

```
p1 <- ggplot(mtcars, aes(wt, mpg)) +  
  geom_point(color="red", fill="cyan", shape=23)
```

```
p1 + geom_smooth(method=stats::lm,  
formula=y~poly(x,2), level=.99)
```

```
p1 + geom_smooth(color="blue",  
linetype=12, lwd=1, level=.99)
```



먼저 p1에 ggplot 함수로 영역을 확보하고 geom_point 함수로 산점도를 그린다. 이때 결과를 p1에 저장하였으므로 그림이 만들어지지 않는다. 이 상태의 그림을 보려면 p1을 입력하면 된다. 이 p1에 추가하여 geom_smooth 함수를 사용하여 왼쪽 그림은 자동차무게의 2차 함수로 연비를 추정한 곡선과 평균의 99% 신뢰구간을 그린 것이며, 오른쪽 그림은 method와 formula를 설정하지 않아서 기본값 "loess"와 y~x가 각각 적용된 결과이다.

ggplot2 패키지의 ggplot 함수를 사용하여 상자그림, 원그래프, 히스토그램 등 많은 시각화를 할 수 있는데 이 부분은 기회가 닿는 대로 소개하기로 한다.



데이터 인포그래픽 강좌 series 11

포스트코로나 시대는 비주얼커뮤니케이션 시대! 정책홍보에 필요한 비주얼 데이터시각화 방법

-농림축산식품부, 경기도 사례 中心-



Infographics

비대면 시대다. 영구 재택근무를 도입하는 기업은 증가하고, 일상에서 디지털 트랜스포메이션, 인디펜던트 워커가 되기 위한 능력 보유는 필수다. 정책홍보 또한 디지털을 통해 국민들로 하여금 공감과 설득을 유도하여 정보를 전달하는 노력도 적극 기울이고 있다. 이번 호에서는 초기 통계 데이터를 분석하여 표와 그래프로 변환하는 과정에서 배워야 할 유용한 시각화 표현법을 알아본다.



가. 강조할 메시지에 따라 달라지는 표 제작 방법

① 업체 수를 강조하는 경우

사업 기간, 설립연도, 업체 수, 투자유치금액 등 4개의 항목으로 이루어진 표다. 통상 표는 1열에 들어가는 것이 '독립변수'이므로, 해당 표는 사업기간(독립변수)으로 살피본 업체 수, 투자유치금액을 말한다. 제작자는 메시지에서 어떤 항목을 강조할지 먼저 결정해야 표 디자인을 할 수 있다.

▶ 개선 전

단위: 백만 원

사업기간	설립연도	업체 수	투자유치금액
1년 미만	2016	2	3,000
1년	2015	20	13,000
2년	2014	18	43,500
3년	2013	5	80,000
4년	2012	15	130,000
5년	2011	10	5,000
6년	2010	2	6,500
7년	2009	3	3,000
8년	2008	1	4,500
9년	2007	3	13,500
계		79	302,000

제시된 표의 열은 사업기간 1열, 설립연도 2열, 업체 수 3열, 투자유치금액 4열이며, 총 10년간의 데이터를 10행으로 구성했다. 표는 제작자가 어떤 메시지(주제)를 전달하려는지에 따라 '표 구성'이 달라진다. 예를 들어 '등록 업체 수 상위 5개'를 강조하고 싶다면 1열은 '업체 수'를 순차 배열로 재구성한다. 업체 수 상위 5개만 남겨두고 나머지 데이터는 제거한다. 투자유치금액 역시 원하는 항

목이 아니므로 제거한다. 업체 수 상위 5개만 남겨두고, 다시 업체 수를 순서대로 정리한다.

▶ 표 요약 과정

단위: 백만 원

사업기간	설립연도	업체 수	투자유치금액
1년 미만	2016	2	3,000
1년	2015	20	13,000
2년	2014	18	43,500
3년	2013	5	80,000
4년	2012	15	130,000
5년	2011	10	5,000
6년	2010	2	6,500
7년	2009	3	3,000
8년	2008	1	4,500
9년	2007	3	13,500
계		79	302,000

▶ 표 요약 후

순위	업체 수	사업기간	설립연도
1	20	1년	2015년
2	18	2년	2014년
3	15	4년	2012년
4	10	5년	2011년
5	5	3년	2013년
계	68		

합계 칸을 만들면서 순위 항목을 1열에 넣었으나 사실 상 1열은 '업체 수'라 할 수 있다. 또 '설립연도'보다는 '사업기간' 항목이 중요하기 때문에 2열은 '사업기간', 3열은 '설립연도'로 위치를 바꾸었다.



▶ 표 시각화 모습

경기도 사업기간별 스타트업 기업 업체 수

사업기간 1년차 업체 수 20개로 가장 많다

순위	업체 수	사업기간	설립연도
1	20	1년	2015년
2	18	2년	2014년
3	15	4년	2012년
4	10	5년	2011년
5	5	3년	2013년
계	68		

출처: 경기도

▲ 표 제목, 출처 표시, 항목 강조, 연관 아이콘 등 시각화 요소를 포함한 모습

키노트(Key Note): 표와 그래프 사용법

표와 그래프는 다른 목적으로 사용한다. '그래프'는 숫자 간 관계를 묘사하기 위해 사용하는 시각적 표현 방법으로 복수 이상의 데이터의 상관관계를 확인할 때 사용하며 자료가 무엇을 말하는지 이해시키는 데 유리하다. 이와 달리 '표'는 정확한 숫자를 전달하고자 할 때 사용한다. 구체적인 양(숫자)을 기억하는 데는 표가 유리하며, 데이터를 열과 행으로 어떻게 분류하느냐에 따라 해석이 달라질 수 있다.



② 투자유치금액을 강조하는 경우

변화율을 나타내는 경우 일단 시간 순으로 자료를 나열하는 것이 중요하다. 특히 변화율에서 큰 차이를 나타낼 때를 강조할 때는 '100% 2중 막대그래프' 또는 '선 그래프'로 표현하는 것을 우선 생각해볼 수 있다. 두 개의 그래프는 연도 차이를 동시에 살펴볼 수 있다는 장점이 있다. 다만 수직으로 그릴 때와 수평으로 그릴 때 연도순서는 다르다. 특히 수평으로 표현하는 경우 최근연도, 즉 2008년은 맨 위에 그리고 강조해야 하는 막대는 명도를 낮게 처리한다.

- '사업기간별'과 '투자유치금액'의 상관관계가 핵심이다. 2개 항목은 1, 2열에 위치한다.
- 투자유치금액 상위 5개만 남기고 나머지 하위 행은 모두 제거한다.
- 설립연도는 사업기간과 밀접한 항목이므로 남겨놓는다. '업체 수' 항목은 '투자유치금액' 총액에 영향을 주기는 하지만 전체금액을 알리는 것이 목적이므로 제거한다.
- 투자유치금액이 중요한 변수다. 투자유치금액 합계 및 단위는 반드시 표기한다.

▶ 표 요약 과정

단위: 백만 원

사업기간 (2열)	설립연도 (3열)	업체 수	투자유치금액 (1열)
1년-미만	2016	2	3,000
1년(5위)	2015	20	13,000
2년(3위)	2014	18	43,500
3년(2위)	2013	5	80,000
4년(1위)	2012	15	130,000
5년	2011	10	5,000
6년	2010	2	6,500
7년	2009	3	3,000
8년	2008	1	4,500
9년(4위)	2007	3	13,500
계		79	302,000



▶ 표 요약 후

단위: 백만 원

순위	투자유치금액	사업기간	설립연도
1	130,000	4년	2012
2	80,000	3년	2013
3	43,500	2년	2014
4	13,500	9년	2007
5	13,000	1년	2015
계	280,000		

▶ 표 시각화 모습

경기도 스타트업 기업 사업기간별 투자 유치금액 Top 5

4년 차 기업 투자유치금액 1천 3백 억 원으로 1위

단위: 백만 원

순위	투자유치금액	사업기간	설립연도
1	130,000	4년	2012
2	80,000	3년	2013
3	43,500	2년	2014
4	13,500	9년	2007
5	13,000	1년	2015
계	280,000		

출처: 경기도

▲ 투자유치금액을 내림차순으로 정렬하여 표로 나타낸 모습

키노트(Key Note)

1. '표'에서는 '독립변수'가 1열이며, 그래프로 나타내면 X축과 같다.
2. 제목은 구체적인 메시지를 담고 있어야 해독시간을 줄일 수 있다. '그래프와 표 제목'은 데이터를 해독할 때 중요한 '나침반' 역할을 한다.
3. 열과 행은 제목에 맞추어 재편집하는 것이 중요하다.
4. 시각표로 만들 경우 어떤 항목(열 또는 행)의 텍스트를 그림으로 동시에 보여줄지 결정한다.
5. 표는 숫자 및 글자 크기, 컬러 등으로 강조가 가능하다.
6. 표는 그래프로 가기 위한 준비 과정이다. 표를 그릴 수 있어야 그래프를 제대로 그릴 수 있다.





나. 신기술 보급 목표와 실적을 강조한 실적현황자료 만들기

2개 이상의 복합적인 데이터를 동시에 표현해야 하는 경우 시각적 차이를 보다 강하게 주는 것이 중요하다. 제목과 그래프 및 표는 물론, 성격이 다른 그래프를 그릴 때 확실한 차이를 알 수 있도록 컬러에 세심한 신경을 써야 한다.

이공계 연구원들이 자주 작성하는 연구실적 데이터 사례를 통해 표현 방법을 살펴본다.

▶ 농림축산부 보급실적 데이터 분석

농림축산식품부는 농식품 분야에서 생산성 향상과 노동력 절감 등으로 경쟁력을 높이기 위해 2014년부터 원예 및 축산 분야에 ICT를 접목한 ‘스마트팜’ 확산 사업을 추진 중에 있다. 매년 시설원예 부분의 데이터 단위는 (ha)를 사용하고, 축산 부분의 데이터 단위는 (호)로 집계된다. 단위가 다른 데이터는 그래프로 나타내는 방법과 ‘도해’로 배열하는 방법이 있다. 두 가지 방법의 차이를 알아보자.

▶ 스마트팜 보급실적 데이터

시설현대화 사업과 연계한 스마트팜 확산 가속화로 도입 첫해인 2014년 대비 2016년 보급 실적이 8~19배 확대되는 등 스마트팜이 선도 농가에서 인근 농가로 확산되면서 주산지 중심으로 규모화, 집적화하는 단계로 진입하고 있다. 스마트팜 보급현황을 살펴보면 2014년 시설원예 60ha, 축산농가수 30호에서 출발해 2015년 시설원예 364ha, 축산농가수는 156호로 늘어났다. 2016년 시설원예 1,143ha, 축산농가수 234호, 2017년은 시설원예 4,000ha, 축산농가수 730호를 목표로 한다.

▶ 전체 분석

총 4년간(2014년, 2015년, 2016년, 2017년)의 스마트팜 보급 현황을 시설원예, 축산 등 두 가지 항목으로 분석한 데이터다. 해당 자료는 먼저 표 분류부터 해야 그래프나 도해 처리가 가능하다. 문장의 앞부분에 있는 ‘2014년 대비 2016년 보급 실적이 8~19배’는 요약 메시지로 별도 표현이 가능하다. 출처는 농림축산식품부이다.

▶ 데이터 요약

• 대표 제목

- 제목1: 스마트팜 2014년 대비 2016년 ‘8~19배’로 증가

- 제목2: 연도별 스마트팜 보급 추이

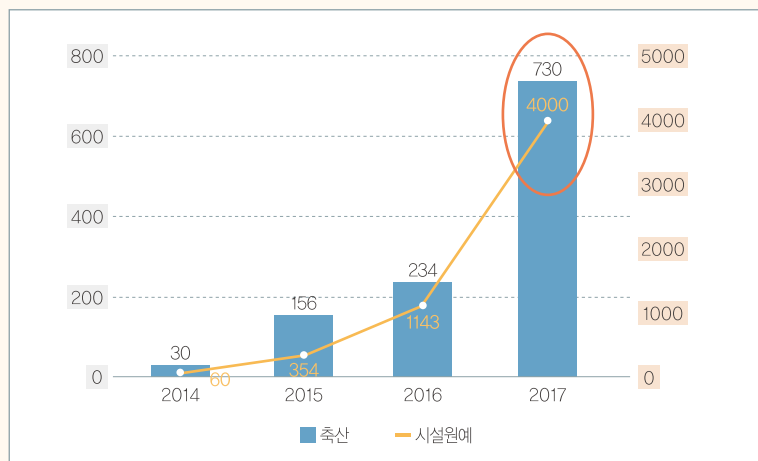
• 데이터 요약: 먼저 ‘표’로 내용을 분류한다. 단위는 ha, 호이다.

항목	시설원예(ha)	축산(호)
2014년	60ha	30호
2015년	364ha	156호
2016년	1,143ha	234호
2017년(목표)	4,000ha	730호



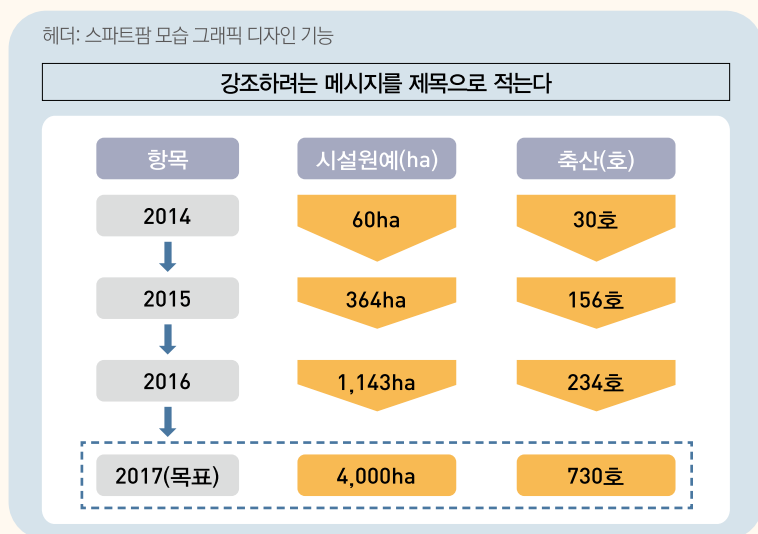
시각화(디자인) 방법 1

독립변수가 연도이므로 추이를 나타내는 동시에 대소의 높낮이를 나타낼 수 있는 복합 그래프(선그래프(시설원예)+막대그래프(축산))를 사용할 수 있다. 특히 최댓값이 4,000ha로 가장 높은 '시설원예' 데이터는 선그래프로 처리한다. 복합 그래프는 동시에 2개의 그래프를 확인해야 하므로 분명한 메시지 포인트를 어디에 주는지가 핵심이다. 제시된 데이터에서는 2017년을 강조하기 위해 다음 그래프처럼 표현하였다.



시각화(디자인) 방법 2

전체 데이터를 시간순으로 전개하는 도해로 나타내는 방식이다. 표보다는 발전된 형태로, 데이터의 흐름을 메시지 전달자가 결정한다. 왼쪽에서 오른쪽으로, 위에서 아래로 그리는 방법 등이 있으며, 통상적으로 시간순의 데이터는 왼쪽 → 오른쪽 도해를 선택하나 해당 데이터는 4년간의 데이터를 나타내야 하므로 가로 공간이 부족할 수 있다. 따라서 위에서 아래(↓)로 흐르는 플로어(flow)형 도해를 사용할 수 있다.



전문가 T·I·P

도해에서는 화살표 도형의 역할이 매우 중요하다. 도해는 읽는 방법, 보는 흐름을 제작자가 결정하는 것으로 한정된 공간에 함축된 정보를 전달할 수 있는 장점을 가지고 있다. 문제는 '스마트팜'이란 단어는 일반인들이 쉽게 받아들이기 어려운 신조어 및 기술용어라는 점이다. 따라서 개념을 글자로만 해독하기엔 매우 어렵고 분석 시간이 걸리므로 중요한 키워드는 이미지에이션(Imagination)해 뇌에서 쉽게 받아들일도록 표현해야 한다. 어떤 의미를 다른 사람에게 전달하는 방법은 글(텍스트), 말, 그림 등이 있다. 그림과 단어(스마트 팜)를 일치시키는 나만의 기호를 만드는 연습이 필요하다.

출처: 데이터인포그래픽 디자인제작실무(이수동, 김선주 예문사)



이수동 samyicamper@gmail.com

(주)인포그래픽웍스 대표 겸

(사)한국인포그래픽협회 대표(현)

전 언론사 근무, 국내 인포그래픽 및 데이터

시각화 분야 개척자다. 국내 첫 인포그래픽

실전서 인포그래픽기획과실전략(베스트셀러),

인포그래픽인사이트57(중국 인민체신출판사

독점계약), 데이터인포그래픽 디자인제작실무

(베스트셀러)를 포함해 여러 권을 집필했다.



데이터로 세상을 만나는 공간, 통계데이터센터

통계와 정책 간 연계 강화, 과학적 정책수립 체계 확립을 위해 통계기반정책평가제도가 도입되고 이는 증거기반정책(Evidence-based policy)이라는 정책학의 흐름이 반영된 결과라고 볼 수 있다.

이렇듯 합리적인 정책 결정에 대한 요구가 높아지고, 국가체계가 선진화될수록 합리적인 정책 결정에 대한 요구가 높아지고 있다.

이런 설득을 위해서는 객관적이고 합리적인 증거 자료, 즉 통계가 필요하다.

통계청은 정확하고 신속한 통계, 그리고 더 다양하고 유용한 통계를 요구하는 각 경제주체가 만족할 수 있는 통계정보를 제공하기 위해 노력하고 있다. 자료의 수집부터 생산·보급·이용까지 통계 행정 전 영역에서 정보화 기술을 적용하여 더욱 정확하고 효율적으로 통계를 작성하고, 이용자가 쉽고 편리하게 활용할 수 있도록 뒷받침하고 있다. 최근에는 품질 높은 통계수치를 생산·제공하는 역할에서 더 나아가 수집된 다양한 데이터를

국민이 유용하게 연계·분석할 수 있는 기반 조성에도 주력하고 있다.

통계청은 이러한 요구에 부응하기 위하여 ‘통계데이터센터’ 서비스를 개발·제공하고 있다.



통계데이터센터를 통해 다양한 데이터 연계·집계 서비스 제공

통계데이터센터는 통계작성을 목적으로 수집한 다양한 데이터를 이용자 관점에서 직접 연계·집계하거나 분석할 수 있도록 구축한 분석플랫폼이다. 행정자료 등을 활용해서 통계청에서 구축한 각종 데이터베이스를 이용자가 목적에 맞게 직접 분석하여 부가가치 높은 데이터를 생산할 수 있도록 지원하고 있다. 또한 엄격한 보안환경을 구축하여 데이터 활용 과정에서 개인·기업의 개별정보가 절대 유출되는 일이 없도록 관리하고 있다.

2016년 10월부터 시범서비스를 시작하여 대전, 서울, 부산, 광주 지역에 통계데이터센터를 정식 개소하였다. 이용자의 분석 편의를 위해 SAS, SPSS, STATA, R서버 등을 갖추었고, 온라인서비스시스템, 자료연계시스템, SGIS-pro 시스템을 추가하여 다양한 분석·활용을 지원하고 있다.



통계데이터센터가 제공하는 서비스

통계데이터센터에서는 크게 분석센터 이용 서비스와 주문형 분석 서비스라는 두 가지 서비스를 제공하고 있다.



분석센터 이용 서비스는 통계데이터센터에서 제공하는 자료뿐만 아니라 이러한 자료들을 분석하고 이를 통해 보고서 작성까지 할 수 있는 플랫폼을 제공하는 서비스를 말한다. 주문형 분석 서비스는 말 그대로 물리적, 환경적으로 직접 방문·이용하기 어려운 이용자들을 위해 전문가가 분석한 결과물을 제공해주는 서비스를 의미한다.

통계데이터센터는 대전, 서울, 부산, 광주 총 4곳만 존재하기 때문에 일부 이용자들은 거리상의 문제로 직접 방문이 어려울 수가 있다. 또 분석의 방법에 따라 접근하기 어려운 영역이 있을 수 있다. 이러한 경우에 '주문형 분석 서비스'를 활용하여 총계표, 기술통계와 같은 다양한 형태의 자료를 얻을 수 있다.





통계데이터센터가 제공하는 자료

데이터센터를 이용하는 가장 큰 목적은 바로 원하는 데이터가 원하는 형태로 이용 가능한가에 있다. 통계데이터센터에서 어떤 자료를 제공하는지 다음에 나오는 표를 통해 자세히 확인할 수 있다.

통계데이터센터에서는 기업등록부, 인구DB, 가구DB 등을 포함한 통계등록부 14종을 서비스하고 있으며, 향후 다양한 데이터를 지속적으로 추가할 예정이다. 또한, 유동인구나 신용정보, 건축물 정보와 같은 민간데이터들도 33종을 서비스하고 있어 다양한 데이터 간 연계와 융합이 가능하다.

인구·가구에서는 등록센서스의 인구 DB, 가구 DB, 주택 DB 등이 있으며, 농림어업 분야에서는 농업 DB, 임업 DB, 어업 DB 등 다양한 자료들을 제공하고 있다.

통계청 자료들 외에 다양한 민간자료들도 제공하고 있다.

인구·가구 분야에서는 가구의 소득과 성별 나이별(유입인구, 주거인구, 직장인구) 그리고 유동 인구들을 알아볼 수 있는 자료들이 있고, 경제·사회 분야에서는 신용정보(대출, 소득, 카드 소비), 카드 매출, 온라인 가격정보와 주택 정보, 차량 통행량 정보를 이용할 수 있다.



1. 통계등록부

부문	자료명	시계열	주기
경제 · 사회	기업등록부DB(사업자등록기준)	2010~2018년	년
	기업등록부DB(대표자기준)	2010~2018년	년
	기업등록부DB_분기(사업자등록기준)	2018년 4분기~ 2020년 1분기	분기
	기업등록부DB_분기(대표자기준)	2018년 4분기~ 2020년 1분기	분기
	사업장 기초DB(4대보험)	2009~2016년 (17,18년 중지)	년
	사업장 기초DB(법인)	2008~2018년	년
	종사자-기업체 연계DB	2015~2018년	년
	육아휴직 사용현황DB	2015~2018년	년
인구 · 가구	등록센서스 인구DB	2015~2018년	년
	등록센서스 가구DB	2015~2018년	년
	등록센서스 주택DB	2015~2018년	년
농림어업	농업DB	2015~2018년	년
	임업DB	2015~2018년	년
	어업DB	2015~2018년	년



2. 민간데이터

부문	자료명	범위	시계열	주기
인구 · 가구	유동인구정보(성연령)	전국	2015.01.~2020.12.	월
	유동인구정보(시간)	전국	2015.01.~2020.12.	월
	유동인구정보(요일)	전국	2015.01.~2020.12.	월
	유입인구정보(성연령)	부산	2017.01.~2018.06.	월
	주거인구정보(성연령)	전국	2016~2019	년
	직장인구정보(성연령)	전국	2016~2019	년
	건물별 가구평균소득	서울, 부산	2016~2017	년
	(코로나19) 인구이동정보	전국	2020.01~2020.05	월
	방문인구정보	전국		월
경제 · 사회	성 · 연령별 카드매출	전국	2015.01.~2018.06.	월
	업종별 매출액	전국	2015.01.~2018.06.	월
	성 · 연령별 카드매출	전국	2018.10~2020.12	월
	신용융합통계정보_대출	전국	2018.10.~2020.12	월
	신용융합통계정보_소득	전국	2018.10.~2020.12	월
	신용융합통계정보_소비	전국	2018.10.~2020.12	월
	전입인구의 평균소득소비	전국	2019.10.~2020.12	분기
	전출인구의 평균소득소비	전국	2019.10.~2020.12	분기
	통근인구의 평균소득소비	전국	2019.10.~2020.12	분기
	유통고객 구매상품 정보	전국	2019.12.~2020.11.	월
	1000대 상권 정보	전국	2018	년
	영업 중 업소	부산	2017~2018	월
	표준공시지가	서울, 부산	2016~2017	년
	주택(단독/다세대/기타)	서울, 부산	2016~2017	년
건물 · 지가	공동주택정보	서울, 부산	2016~2017	년
	건축물정보	전국	2017~2018	월
	인구집중유발시설	서울, 부산	2016~2018	분기
차량	도로별 차량통행량	서울, 부산	2017	년
	시 · 군 · 구별 차량등록대수	전국	2017	년
기초 정보	행정동코드 정보	전국	2018	년
	법정동코드 정보	전국	2018	년
기타	SNS 지역축제정보	전국	2017	년
	온라인가격 정보	전국	2014.01~2020.09	월
	모바일상품권 거래 정보	전국	2018	년



통계데이터센터 홈페이지 활용 팁

○ 빅데이터 활용

빅데이터 활용 메뉴 안에는 뉴스 기반 통계검색, 모바일 유동 인구지도, 생활경제지표, 온라인 가격정보, 지도로 본 빅데이터, 국내·외 동향 등 다양한 기능이 있다. 한 번의 클릭으로 유용한 정보를 손쉽게 얻을 수 있다. 그리고 '시각화분석 서비스' 기능이 있어 다양한 도표 및 그래프를 통해 원하는 정보를 명료하게 확인할 수 있다.



○ 소통 마당

소통 마당에서는 궁금한 사항에 대하여 질문할 수 있는 묻고 답하기, 참고할 수 있는 자료들로 채워져 있다. 특히 분석·활용 사례를 통해 통계데이터센터를 이용하기 전에 어떤 연구들이 가능한지 미리 볼 수 있다.



○ 센터 소개

데이터 분석이나 활용 등 접근하기 어려운 경우 ‘교육지원’을 활용할 수 있다. 별도의 교육 신청을 통해서 직접 통계데이터센터에 방문해 분석 전문가와 함께 데이터를 분석해보고 시각화까지 해볼 수 있는 서비스를 제공하고 있다.

이런 서비스들은 특히 데이터 분야에 관심이 있는 대학생들에게는 좋은 기회가 될 수 있다.

○ 통계데이터 분석·활용대회

통계데이터 분석·활용대회 같은 경우 개인이나 팀으로 누구든지 참가할 수 있다. 데이터 분석 보고서 분야뿐만 아니라 분석 초보자들을 위한 데이터 분석 포스터 분야가 있어 활용 초보자도 도전해볼 수 있다. 우수 보고서나 포스터로 선발된 참가자에게는 통계청장상을 표창하고 상금도 있기에 대학생이나 대학원생들에게도 좋은 도전이 될 수 있다.



손가락 반상 미·실



INTERVIEW

식당 데이터 경영자 박노진

알면 다르게 보이고, 보이면 길이 생긴다



요즘 신문 방송에 데이터라는 말이 많이 나옵니다. 하지만 데이터에 접근해서 친해지기란 쉽지 않습니다. 숫자로 된 표를 잠시 바라보는 것만으로도 머리가 아파지고 하품이 나옵니다.

『식당 공부』란 책을 보는 순간 반가운 마음이 들었습니다. 아, 식당에서도 이런 접근을 하고 있었구나.

천안에 있는 한 식당 앞마당에 부부가 겨울을 준비하며 나무에 작은 전구를 달고 있었습니다. 인사를 하자 작업을 멈추고 먼지 묻은 손을 털며 반갑게 맞아주었습니다. 싱긋 웃으시는 모습이 연구실에서 잠시 산책을 나온 학자 같다는 생각이 들었습니다.

책 얘기를 꺼내자 쑥스러워하시다가 불쑥 이런 말을 던집니다.

“이 책 안 팔릴 거라고 출판사 여러 군데서 거절당했어요”

사실 나조차 식당 운영에 데이터 활용을 생각해보지 않았습니다. 우리는 예로부터 손맛을 중요하게 여겼습니다. 손맛이라 것은 참 묘한 것이죠. 할머니가 똑딱 만들어내는 음식에 누구도 흉내 낼 수 없는 맛이 스며 있듯이, 오랜 세월 동안 내공을 닦은 사람만이 가지는 능력 같기 때문입니다.

“식당경영은 만성질환인 당뇨병 관리와 같이 일정 손익프레임에서 관리해야 하는 거예요. 무엇이 문제인지를 알아야 해결 방법도 찾을 수 있죠. 이를 통해 작은 시도를 해나가야 해요. 그렇게 실패하고 성공하는 과정을 통해 나만의 것을 가질 수 있어요.”

○ 직장생활을 하다 보면 매일 방문하는 곳이 식당일 겁니다. 식당을 방문할 때마다 유심히 살펴보는 것이 왜 이 집은 사람이 많고 왜 저 집은 사람이 없을까입니다. 보통 음식 맛이나, 상권, 서비스 특성에 따라 식당의 성공과 실패를 많이 얘기합니다. 데이터를 살펴보게 된 어떤 계기가 있으신지요.

식당을 운영하다 보면 밤에 잠자리에 들 때 수많은 생각이 떠오릅니다.

오늘 많이 판 걸까? 얼마 남은 거지? 부가세는 왜 이렇게 많이 나온 거야? 재료비는 얼마나 들어가는 거야? TV에 나온 식당은 어떻게 저렇게 만들었을까? 우리 집 메뉴 구성은 괜찮은 걸까? 지난달과 이번 달의 매출은 차이가 있는데 이익은 왜 500만 원이나 줄어든 걸까? 손님은 몇 달에 한 번씩 우리 가게를 찾아주는 걸까? 손님들이 어떻게 하면 자주 오게 할 수 있을까?

이런 고민을 하지만 답이 찾기가 참 힘들어요. 어떻게 하면 여기에 대한 답을 찾을 수 있을까 고민하다가 포스 매출 자료를 들여다보기 시작했어요. 그 데이터를 들여다볼수록 그 안에 해답이 있다는 것을 알게 되었죠. 15년 동안 숫자들을 보며 문제를 해결하는 방법을 찾다보니 여기까지 오게 된 것 같습니다.

○ 다른 식당을 방문하실 때도 이런 데이터 경영적 시각으로 살펴보실 것 같아요. 식당을 방문하실 때 무엇 무엇을 살펴보시는지요.

제일 먼저 식당을 가서 보는 것은 청결한 이미지를 주는가입니다. 홀이나 주방이 지저분하거나 무질서한 느낌을 받으면 아무래도 기분이 좋지 않아서 맛에도 나쁜 선입견을 주게 되더라고요.

그리고 이 집의 손익프레임은 어떻게 될까 생각해봅니다. 직원들이 몇 명이니까 매출이 얼마인지 추정해보고 메뉴의 원가를 분석해서 이 둘의 합이 매출의 몇 %를 차지하는지 어렵잡아봅니다. 그러면 자연스럽게 테이블이 몇 석이고 방문한 시간대에 따라 회전율도 대략 짐작되니까 그 음식점의 대략적인 손익계산이 머릿속에 들어옵니다. 그것에 따라 직원들과 주인의 움직임과 얼굴을 봅니다. 허허, 하지만 음식이 나오면 다 잊고 맛있게 먹습니다.

○ 식당을 보면 대략 손익계산이 나오시는군요. 잘되는 식당과 안되는 식당의 차이가 바로 보이시겠어요.

잘되는 식당은 직원들의 분위기가 밝아요. 인사를 잘하거나 잘 웃거나 대답을 잘해요. 음식도 15분 이내로 빨리 나오고요. 신기하게 두 명 손님이 많아요. 재방문율이 높은 거라 유추할 수 있죠. 보통 단체 손님은 돈이 되니 잘해주고 두 명 손님을 신경을 덜 쓰는 편이거든요. 그런데 이렇게 만족한 손님들이 단체 손님들을 데리고 오는 거예요.

반면에 안되는 식당은 주인이 카운터에서만 있고 손님한테 잘 오지 않아요. 주로 종업원을 시키기만 하죠. 또 직원이 손님들이 있어도 핸드폰을 보거나 관심을 기울이지 않아요. 그리고 정리 정돈이 안 되어 있는 곳입니다. 이런 곳은 음식에도 정성이 들어가지 않을 확률이 높아요.



○ 식당 경영하시는 분이 이 글을 보고 계실 수도 있는데, 경영 상황에 따라 어떻게 데이터 활용이 가능할까요? 예를 들어 망해가는 곳, 어정쩡한 매출의 식당, 잘되는 식당에 따라 데이터 활용 방법이 다를 것 같아요.

망해가는 식당에서 왜 망해가는지 분석할 수 있다면 그래서 망하는 요인을 제거할 수 있다면 다시 한번 회생의 기회를 얻을 수 있겠죠. 그 원인이 어떻게 해볼 수 없는 요인이라면 빠른 폐업 결정을 내릴 수 있어요.

어정쩡한 매출이 일어나는 식당은 선택과 집중을 하게끔 도와줄 수 있습니다. 이것도 맛있고 저것도 맛있는 음식점은 성공하기 쉽지 않습니다. 청국장이면 청국장, 김치찌개가 맛있는 식당은 김치찌개에 집중해야 합니다. 다 잘한다고 해도 손님들은 어느 하나의 음식만 기억하는 법이거든요.

천안에 있는 어떤 고깃집은 입지도 별로인데다 메뉴도 여러 가지라 손님들한테 잘 기억되지 않았던 식당이었는데 점심때 김치찌개에 고기를 듬뿍 넣어주어 대박 난 사례가 있습니다. 김치찌개로 대박 나고 저녁에 고기로 다시 손님을 유치하는 전략으로 성공한 사례입니다.

잘되는 식당은 굳이 데이터를 활용하지 않아도 잘 됩니다. 이런 식당이 기업형 외식 업소로 성장하고 싶다면 데이터 경영을 도입하기를 추천합니다.

○ 김치찌개 사례를 말씀해 주셨는데요. 아마 여러 식당에 데이터 경영 방법을 전수해주셨을 듯합니다. 그 사례를 좀 더 듣고 싶습니다.

청주에 어떤 보리밥은 장사가 엄청나게 잘되는데 데이터 분석을 해보니까 인건비 비율이 생각보다 낮다는 것을 알고 직원을 더 충원했습니다. 그랬더니 손님들한테 음식이 제공되는 시간도 줄고 테이블을 치우는 시간도 줄어들어 오히려 매출이 더 늘어난 윈윈 효과를 경험했습니다. 시



간을 줄여 회전율을 높일 수 있었던 것이죠. 만약 데이터를 보지 않으면 직원 보장을 생각하기 어려워요. 당장 나가는 인건비만 머릿속에 떠오르거든요. 데이터를 통해 전체적인 구조를 파악해야만 가능한 결정이죠.

냉면집 사재인데요. 여름엔 장사가 괜찮은데 나머지 계절은 힘들다고 해서 데이터 분석을 한 적이 있었습니다. 데이터를 분석해보니 겨울에 불고기 주문이 많다는 결과가 나왔습니다. 냉면집이다 보니 냉면을 겨울에도 팔 방법이 필요했어요. 그래서 불고기와 냉면을 세트 메뉴화했습니다. 추운 겨울에 불고기와 냉면은 끌리는 조합이었어요. 이 메뉴를 통해 겨울철에도 많이 사람들이 찾아오게 되었지요.

그 외도 원가분석을 통한 메뉴를 개선한 초밥집, 수익성을 개선한 연어전문집 등 여러 사례가 있습니다.



○ 책을 보면 가슴에 와닿는 문구가 많이 있습니다. ‘느낌을 숫자로 설명할 수 있는가’, 이 문장은 많은 것을 생각나게 해주는데요. 느낌으로 아는 것과 그것을 숫자로 설명할 때의 차이점은 무엇인지요.

장사가 잘되거나 무난할 때는 느낌을 구체적인 숫자로 표현하기 어렵지 않습니다. 어떤 이유를 가져다 붙여도 다 맞거든요. 이래서 잘되는 것 같고 저래서 잘되는 것 같고요.

그렇지만 안되거나 어려울 때, 특히 이유도 없이 장사가 안되면 정말 미칠 것 같습니다. 마치 등이 가려운데 엉뚱한 데를 계속 긁고 있는 그런 느낌이에요. 알 것 같은데 그것이 아닌 것 같은 그런.

그럴 때 데이터를 보면 그 느낌의 실체를 발견할 수가 있습니다. 무엇 때문에 이런 현상이 일어나는지 설명할 수 있는 거예요. 그 원인을 알게 되면 다르게 보입니다. 다르게 보이면 해결할 방법을 찾을 가능성이 커요. 방법을 찾으면 그 위기를 넘기게 돼요. 그러면 한 단계 올라서게 되는 겁니다.

우리 가게의 하루, 한 달, 그렇게 1년을 온전히 숫자로만 표현해놓으세요. 그러면 문제가 닥쳤을 경우 그 안에서 막연한 느낌의 정체를 찾을 수 있어요.

○ 또 책에서 인상 깊은 구절이 ‘모든 매출의 움직임에는 다 이유가 있다. 매출이 떨어지는데 그 원인을 모르면 끌어올릴 방법이 없다. 그 시행착오에서 무엇을 얻고 무엇을 잃었는지 알 수만 있다면 그것이 자산이 된다’입니다. 코로나로 많은 자영업자가 어려움을 겪고 있습니다. 데이터 경영이라는 말에 공감은 하면서도 쉽게 다가오지는 못할 것 같습니다. 그들에게 어떤 말씀을 해주시고 싶으신지요.

생각보다 자신의 가게에 대해서 모르는 분이 많아요. 지금 운영하시는 가게에 대해 어떤 진단을 하고 있는지, 매출이 어떻게 흘러가고 있는지, 메뉴 구성은 적절하고 객단가 회전율은 어떤지, 고객은 누구인지, 직원의 생각에 대해서 알고 있는지, 원가 구성은 어떻게 되는지, 마케팅 방법은 어떻게 가야 하는지에 대해서 구체적으로 답할 수 있어야 합니다. 이런 객관적인 분석을 통해야만 집중과 선택의 포인트가 보입니다. 즉 우리 가게만의 적정 손익프레임을 맞추고 그에 기초해 가성비 메뉴를 개발하고 고객의 만족을 높이기 위해서 무엇을 해야 하는지 그 길이 보이는 거죠.

작은 식당의 경우 대표의 역량이 80%로 좌우한다고 생각해요.
만약 과거로 되돌아간다면 저한테 구사일행(九思一行)을 얘기하고 싶어요.
아홉 번 생각하고 한번 실행한다는 의미입니다.

○ 시간대별, 요일별, 홀과 배달, 점심 저녁, 식사와 술 메뉴 등으로 매출을 쪼개어 본 것이 인상적이었습니다. 결국 이렇게 세부적으로 판단하고 이에 대한 전략을 세우려면 다양한 자료가 필요할 것 같은데요. 이런 부분에 대한 데이터화가 필요하다 또는 이런 데이터가 있었으면 좋겠다고 생각되시는 게 있으신지요.

요즘 포스는 매출에 대한 분석은 잘되어 있습니다. 이 자료를 활용하면 매출 위주의 분석자료는 확보할 수 있습니다. 반면에 매입 관련 데이터는 잘 정리되어 있지 않습니다. 매입과 원가를 구성하는 요소들이 가게마다 다양하므로 이에 대한 취합데이터를 모으기 어렵죠. 여기에 대한 데이터 수집이 쉬운 시스템이 나오면 좋겠어요. 마음 같아서는 인공지능형 데이터 경영시스템을 만들어보고 싶네요.

인공지능이라는 말이 나와서 말인데, 요즘 많이 얘기되고 있는 빅데이터는 소상공인에게는 크게 도움이 안 돼요. 배추나 삼겹살 가격 동향 같은 이런 자료들이 있다면 식당을 운영하는 데 많은 도움이 될 것 같아요. 이런 가격의 움직임을 예측할 수 있다면 그에 따라 효율적인 원가 계산이 가능할 거라 여겨집니다.

○ 사업자 등록증이 14개라고 하셨는데 13번의 시련을 거쳐오신 것 같습니다. 그때로 되돌아가서 한마디 해줄 수 있는 타임머신이 있다면 그 시절의 본인한테 무슨 말을 해주시고 싶습니까.

작은 식당의 경우 대표의 역량이 80%로 좌우한다고 생각해요. 만약 과거로 되돌아간다면 저한테 구사일행(九思一行)을 얘기하고 싶어요. 아홉 번 생각하고 한번 실행한다는 의미입니다.

2002년 첫 번째 식당으로 고깃집을 시작했습니다. 천안에서 규모가 꽤 큰 식당이었는데 2년 차에 광우병이 터져 쫓딱 말아먹었습니다. 광우병이 터지기 전에는 월1억 이상 팔았는데 80% 이상 매출이 급감해 어떻게 손써볼 도리도 없이 문을 닫았죠.

올해도 코로나라는 전혀 생각지도 못한 외부환경에 많은 사람들이 어려움을 겪고 있어요. 이런 외부환경의 변화나 위협에 충분히 대응할 힘을 길러야 해요. 그게 무엇이냐 하면 바로 내 가게에 대해 숫자로 파악하고 있느냐예요.

제가 데이터 경영 수업을 하면 80%가 포기를 해요. 장기간 데이터를 모으고 숫자를 들여다보면서 이런 생각 저런 생각을 해보지 않으면 작은 외부 변화에도 당황하게 됩니다. 충분한 준비와 구체적인 고민을 통해 행동으로 옮겨야 합니다.



이것만 알면 나도 유튜버! 스마트폰 속 사진을 동영상으로 만들기

어느덧 아침저녁으로 쌀쌀한 가을이 깊어지고 있다.

이렇게 한 해가 마무리 되어가는 때가 오면 스마트폰에는 많은 추억을 담은 사진이 간직되어 있다.

‘아 맞다 이때 그랬지’, 사진을 보면 신기하게도 평소 잊고 있었던 많은 추억이 다시 생생하게 되살아난다.

그래서 역시 남는 것은 사진밖에 없다고 말하나 보다. 이 사진들을 이용하여 영상 앨범을 만들면 어떨까?

그래서 준비해보았다. 내 스마트폰 속에 잠들어 있는 사진으로 영상 만들기!

지난 호에서 학습했던 곰믹스를 이용해 폰 속에 잠들어 있는 사진을 영상 앨범으로 손쉽게 만들어보자.

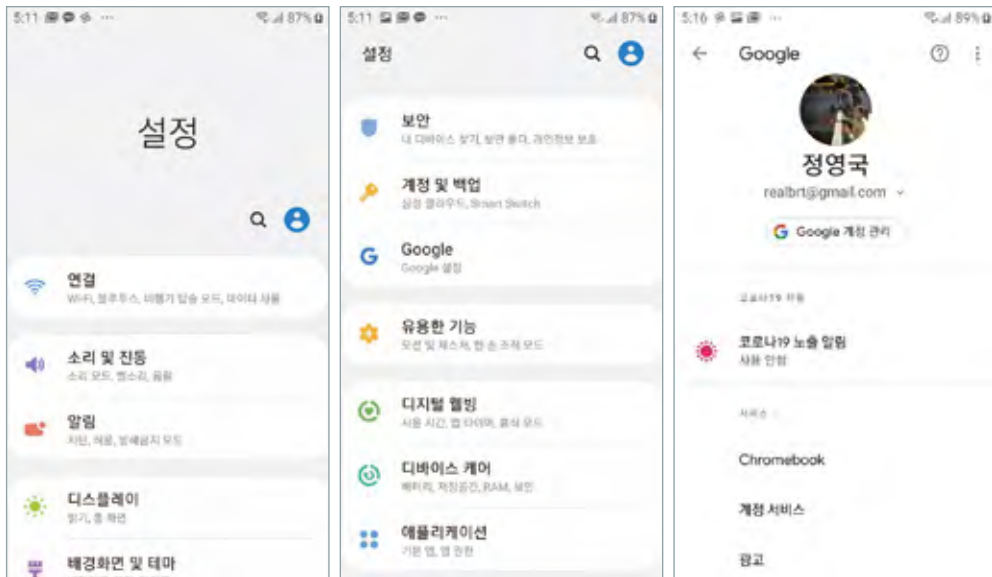
지난 호에서 다루었던 영상 만들기의 기본 순서를 다시 한번 살펴보자.

영상 클립을 가져오고, 음악 넣고, 컷 편집하고, 장면전환 하면 끝. 자, 오늘도 이 순서로 시작해보자.

스마트폰에서 PC로 가져오기

안드로이드 안드로이드 OS는 USB를 이용해 쉽게 사진을 PC로 복사할 수 있다. 그러나 파일이 많은 경우 구글포토를 이용한 방법을 추천한다.

- 1 여러분의 휴대전화의 OS가 안드로이드라면 휴대전화에 등록된 지메일 계정이 있다. (설정 > Google 메뉴에서 확인) 먼저, 나의 지메일 계정을 확인해야 한다.



- 2 PC로 이동 후 인터넷 브라우저에서 구글포토(<https://photos.google.com>)에 접속해보자.
※인터넷 브라우저는 크롬(chrome)을 추천합니다.



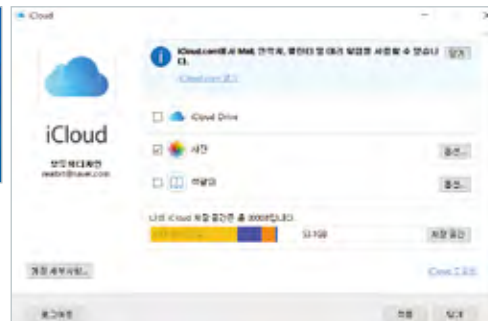
3 스마트폰과 동일한 계정으로 로그인해보자.
그러면 스마트폰에서 그동안 담았던 추억들이 ‘앨범’카테고리 안에 인물, 장소, 사물 등으로 인공지능(AI)에 의해 잘 정리되어 있는 것을 볼 수 있다! 영상을 만들 사진들을 선택해서 다운로드(shift+D)를 받으면 된다.



iOS iOS의 경우 Mac 사용자는 기본 사진 앱에서 쉽게 사진들을 복사할 수 있다. 그러나 Windows 사용자는 iCloud를 설치해야 한다는 것을 알아두자.

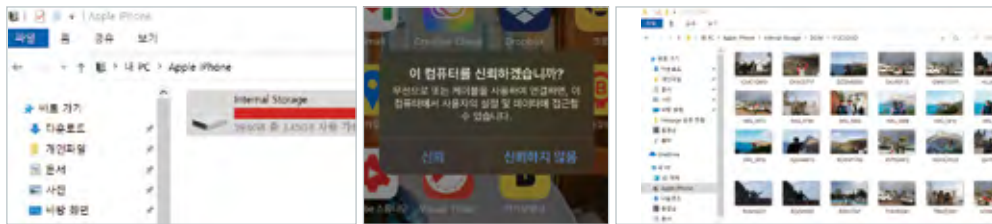


〈iCloud 다운로드 및 설치〉



〈iCloud 연결〉

포털사이트 또는 Windows store에서 'Windows 용 iCloud'를 검색한 후 다운받아 설치한다. iCloud 사용이 불편하다면 USB를 이용하면 된다.



〈USB를 이용한 연결〉

〈신뢰하기〉

〈파일 탐색기에서 확인〉

PC에 USB를 연결하고 '이 컴퓨터를 신뢰하겠습니까?'라는 안내 팝업에 '신뢰'를 선택한다. 그러나 사진의 정리가 되어있지 않기 때문에 불편한 면이 있어서 iCloud 설치를 추천한다.

음악 선곡 및 준비하기

사진을 선택했다면 그다음 음악을 준비해보자. 곰믹스에서 제공하는 기본 음악을 사용하면 된다.

무료 음원 다운로드		
파일명	재생시간	미러듣기
♪ Eternal Sunshine.mp3	00:01:51	▶
♪ Fantasy Piano.mp3	00:02:29	▶
♪ Iceberg.mp3	00:01:04	▶
♪ Last Travel.mp3	00:02:13	▶
♪ Planet Travel.mp3	00:02:43	▶
♪ Red Guitar.mp3	00:01:31	▶

그런데 만약 완성된 영상을 유튜브 채널에 올려야 하는데 곰믹스의 음악이 마음에 들지 않아 다른 음악을 선곡하고 싶다면 저작권에 문제없는 음악을 구해야 한다. 하지만 일반적으로 어떤 음악이 저작권 사용이 가능한지 아닌지 구분하기 어렵다. 그런 경우 유튜브에서 제공하는 음악 라이브러리에서 음악을 다운받아서 사용하면 가장 안전하다. '무료 음원 다운로드'에서 유튜브를 선택하면 유튜브 스튜디오로 페이지 이동이 된다.

유튜브에서 저작권 문제없는 음악다운 받기

유튜브에 가입하여 개인 채널이 있다는 가정하에 로그인하면 우측상단 내 프로필 사진을 클릭하면 유튜브 스튜디오 메뉴 항목이 있다. 유튜브 스튜디오에서 왼쪽 하단 메뉴 카테고리에서 오디오 보관함을 선택하면 그곳에서 다양한 음원들을 저작권 걱정 없이 무료로 사용할 수 있다.

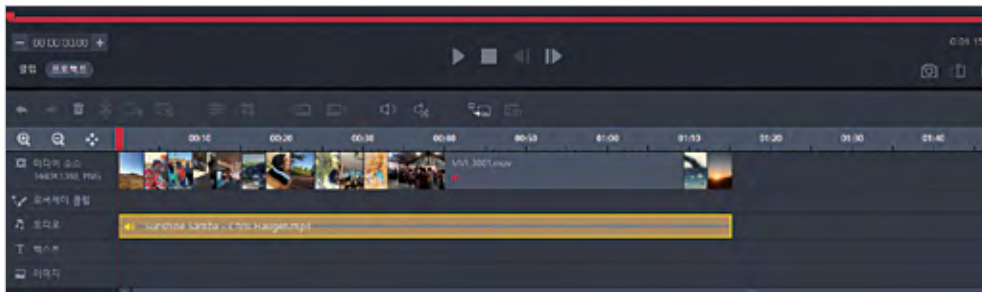
곰믹스에서 음악 불러오기

마음에 드는 음악을 찾았다면 [미디어소스] 탭에서 [파일추가]를 선택한 후 음악을 불러온다. 오디오 타임라인에 음악을 드래그하여 옮겨 넣어보자.

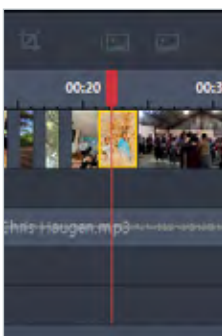


영상(사진) 컷 편집하기

자, 이제 모든 준비가 되었다. 준비한 음악 위에 사진들을 올려놓으면 된다. 이때 감각 있게 편집하는 비결은 음악의 박자에 맞춰 사진들의 시간 조절을 하며 영상 자르기를 하는 것이다. 그러면 영상이 음악과 함께 굉장히 자연스러운 느낌을 줄 수 있다.

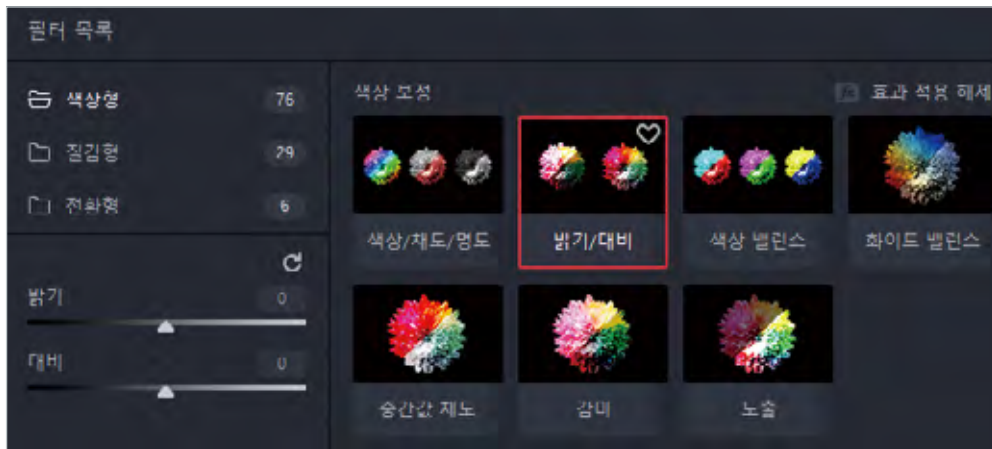


장면전환과 기타효과



장면전환과 색 보정은 개인의 취향에 따른 선택사항이다. 음악의 템포가 빠르다면 장면전환도 따로 넣을 필요 없이 바로 바로 넘어가는 게 어울린다.

하지만 음악의 템포가 느리거나 서정적이라면 부드럽게 장면전환을 하는 것이 어울린다.



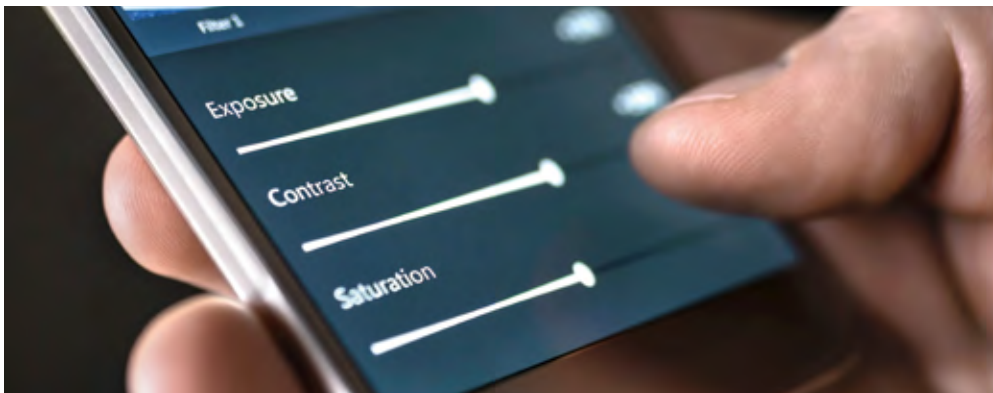
색 보정은 어두운 사진이나 밝게 나온 사진이 있다면 적정 노출을 맞추고 화이트밸런스를 맞추어준다. 그리고 영상의 제목과 자막을 넣어주면 멋진 여행 영상이 완성된다! 이 부분의 더 자세한 설명과 인코딩은 지난 호를 참조해보자.

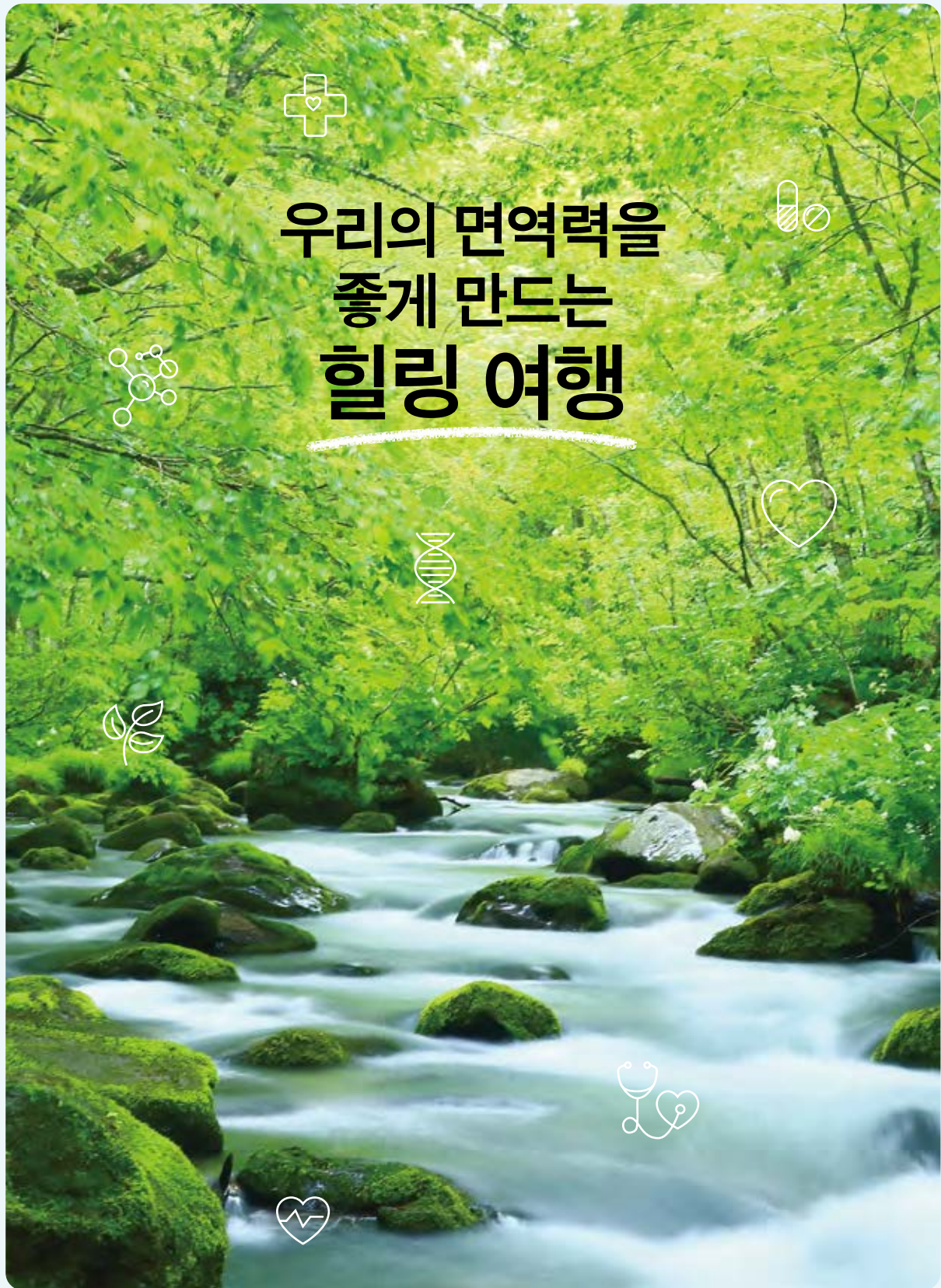
마치며

지난 시간 알아본 곰믹스를 활용해서 폰에 잠들어있는 오래된 사진을 PC에 불러와서 간단한 스케치 영상을 만들어보았다. 주변에 영상 초보자분들을 보면 휴대전화로 찍은 사진 또는 영상들을 PC로 옮기지 못해 시작조차 하지 못하는 것을 자주 본다. 오늘 내용을 보고 따라 하면 아마 쉽게 PC로 옮겨서 영상 앨범을 만들 수 있을 것이다.

영상편집 방법을 다시 한번 정리해보자.

촬영한 영상 클립을 1. 가져오고 2. 음악 넣고 3. 컷 편집 4. 후보정(장면 전환과 색 보정, 자막 등) 순서로 진행한다는 것만 기억하자. 이것만 기억하시면 어떠한 도구를 사용해도 영상 제작을 할 수 있다. 자, 그럼 여러분의 휴대전화 속에 잠든 추억들을 꺼내서 영상으로 만들어보자!







“코로나19, 결국 면역력 싸움입니다!”

- 『면역의 혁명』, 정신과전문의 이시형 박사

코로나19 이전으로 돌아갈 수 있을까? 1980년대 후반에 등장해 우리를 공포에 떨게 만들었던 에이즈를 지금은 보건소에서 검사하고 관리하는 일상이 된 것처럼, 코로나19도 언젠가는 그렇게 될 것이다. 인간에게 해롭지 않은 비병원균성 바이러스가 백만 명 이상의 사망자를 내면서 순식간에 전 세계를 마비시킬 수 있다는 사실을 뼈저리게 체험했고, 코로나19가 무증상으로 쉽게 넘어가는 사람이 있는가 하면, 면역력이 약하면 독하게 앓다가 사망에 이르는 극단적인 예후가 있다는 것도 알게 됐다. 그래서 요즘은 무조건 ‘면역’이다.

한 방송국에서 「면역력이 좋아지는 힐링 여행!」을 촬영하자고 했다. 의사와 환자가 공기 좋고 경치 좋은 곳을 여행하면서 건강의 핵심을 찾아가는 주제다. 모든 병이 그렇겠지만 개인의 면역 상태도 라이프스타일과 밀접하다. 달랑 하루 여행을 간다고 해서 잃었던 건강을 되찾을 수는 없겠지만 건강가이드로서 내가 할 일은 환자의 나쁜 라이프스타일을 알아내고 좋은 쪽으로 안내하는 역할이었다. 진료실이라는 한정된 공간과 짧은 진료 시간으로는 라이프스타일을 세세하게 알아낼 수 없다. 이야기하면서 건고 건강한 음식을 함께 만들어 먹는 힐링 여행은 잘못된 라이프스타일을 알아내고 바꿀 수 있는 절호의 기회가 될 것 같았다.

잘못된 라이프스타일은 수십 년에 걸쳐 사람을 병들게 만든다. 서서히 나를 파먹기 때문에 몸을 망쳐놓은 범인인 줄도 모른다. 설사 안다고 해도 그 습관에 익숙해져 있어서 쉽게 못 바꾼다. 오죽하면 ‘습관이 쌓이면 운명이 된다’고 하겠는가.

나도 건강을 해치는 몇 가지 나쁜 습관이 있었고 수년에 걸쳐서 바꿨다. 운동신경이 둔해서 움직이는 것을 싫어했지만 헬스트레이너 자격증에 도전했고, 계란말이 하나도 깔끔하게 하지 못할 정도로 요리에 소질이 없었지만 항암 건강 요리책을 썼다. 부모님을 비롯한 가족 모두가 인슐린을 써야 할 정도의 당뇨병에 시달리고 있지만, 나만은 56세인 나이에 아직 당뇨가 없는 것은 이런 노력 덕분일 것이다. 여행에서 만나는 환자에게도 내가 학교에서 배운 의학적 지식이 아니라, 내가 실천해 온 라이프스타일을 알려주고 싶었다.

의사가운 대신 등산복을 입고 진료실을 벗어나 충청남도 청양군에 위치한 칠갑산으로 떠났다. 매운맛의 대명사, 청양고추로 유명한 바로 그 청양이다. 고추박물관을 비롯해서 빨간 고추 조형물이 버스 정류장 곳곳에 보였다. 아침 9시, 칠갑산 출렁다리 앞에서 애경(가명, 46세) 씨가 기다리고 있었다. 그녀는 겉보기에는 건강해 보였지만 벌써 7년째 당뇨에 시달리고 있었다. 최근

에 측정한 당화혈색소가 10.3이라고 하니 평소 혈당이 250 이상인 환자였다.(당뇨병: 공복혈당 125 이상, 식후 2시간 혈당 200 이상.) 내과에서 지속적으로 치료를 받고는 있고 식단도 열심히 하는데 당이 떨어지지 않아서 걱정이라고 했다.

그녀를 제일 괴롭히는 것은 자꾸 재발하는 허벅지 뒤쪽의 수포성 대상포진과 칸디다성 질염이었다. 그야말로 면역이 떨어져서 정상인은 잘 걸리지 않는 감염이 생기고 있었다. 이런 경우 혈당을 적극적으로 조절하는 것이 급선무다. 대상포진이 느껴지면 바로 병원으로 가서 항바이러스제를 처방받아야 하는 것도 알려졌다. 그녀는 이미 심각하다는 것을 잘 알고 있었고, 이번 여행을 마치면 입원을 해서라도 혈당을 조절해하겠다고 했다.

칠갑산은 평지가 많은 산이라서 무릎이나 허리가 약한 사람도 가볍게 등산하기 좋은 곳이다. 피톤치드가 많다는 침엽수 아래에서 애경 씨와 나는 마스크를 잠시 벗었다.(코로나19 때문에 2미터 이상의 거리를 유지했다.) 칠갑산 입구에 넓게 자리 잡은 천장호 호수에는 구름 속 솜털가지 선명하게 비쳤다. 미세먼지 한 톨 없는 깨끗한 칠갑산의 산소를 온몸에 담아서 가져가고 싶었다. 애경 씨에게 복식호흡을 가르쳐주었다.

복식호흡법

- 1 가을 하늘을 품고 있는 호수를 지국이 바라보면서 호흡이 자연스럽게 나가고 들어오는 것을 느낀다.
(시선을 15°유지하는 것이 좋다. 고개를 숙이면 목에 있는 중요한 호흡근인 사각근이 긴장하므로 고개를 살짝 들어야 한다.)
- 2 완전히 '후~' 하고 입으로 공기를 내뿜는다.
- 3 한 번 코로 공기를 들이마시고, 2번에 걸쳐 나누어서 입으로 '후~ 후~' 내쉴다.
- 4 두 번 코로 공기를 들이마시고, 4번에 걸쳐 나누어서 입으로 '후~ 후~ 후~ 후~' 내쉴다.
- 5 세 번 코로 공기를 들이마시고, 6번에 걸쳐 나누어서 입으로 '후~ 후~ 후~ 후~ 후~ 후~' 내쉴다.
- 6 들이마신 횟수의 두 배로 호흡을 천천히 뱉는 연습이다. 열 번까지 들이마시고 스무 번으로 나누어 내쉴다.
이 동작을 3회 반복한다.

호흡에는 들숨과 날숨이 있다. 보통은 들어 마시는 호흡을 잘해야 한다고 생각하지만, 오히려 그 반대다. 날숨을 천천히 잘 내뿜어야 새로운 공기를 들이마시는 들숨도 잘할 수 있다. 들이마신 횟수의 두 배로 천천히 내쉬는 것이 복식호흡의 매력이다.

“비워야 채워지고, 버려야 얻는다”라는 의미심장한 말이 인체에서도 통한다.

산행 때문에 혹시 애경 씨가 저혈당이 되지 않을까 싶어서 당검사를 했더니 136이 나왔다. 최근 측정한 혈당 중에는 제일 낮다고 했다. 코로나 때문에 집에만 있었는데 청양으로 오니까 몸과 마음이 날아갈 것 같다고도 했다. 역시 스트레스는 만병의 근원인가 보다.

청양이 고추 말고 구기자도 유명하다는 것을 청양에 가서야 알았다. 촬영 팀은 우리를 2020년 청양 구기자왕으로 선정된 오용재 씨 농장으로 안내했다. 마침 수확철이라서 아이들처럼 생과를 직접 따보는 체험학습도 할 수 있었다. 한 줄기에 30개 정도 조롱조롱 열린 새빨간 구기자는 보기만큼 맛도 달았다.





구기자왕은 구기자가 치매, 당뇨, 면역 등 모든 병에 좋다고 설명했지만, 의사인 내 눈에는 토마토처럼 빨간 색소에 많은 라이코펜이 풍부해서 좋아 보였다. 그래도 당분이 꽤 있어서 당뇨병자는 혈당을 체크하면서 먹어야 할 것 같았다. 오동통하고 빨간 구기자를 한 소쿠리 얻어서 마지막 여행코스인 글램핑 캠핑장으로 향했다.

면역력이 떨어진 당뇨병 환자를 위해서 버섯불고기와 채소 쌈을 준비했다. 불고기는 단백질 공급원이고 버섯은 섬유소와 베타글루칸이 풍부해서 당 조절과 면역세포 활성화에 좋다.

늘 해 먹는 불고기지만 맛은 같고 조리법은 달랐다. 단 맛에는 저칼로리 감미료인 스테비아를 사용했고, 비타민 D가 풍부한 목이버섯을 많이 넣었다. 스테비아만 넣으면 인위적인 단맛을 느낄까봐 농장에서 얻어온 구기자도 약간 넣었다. 구기자가 없으면 올리고당이나 설탕으로 해도 된다.

파, 마늘과 양파도 평소보다 많이 넣었다. 우리가 양념이라고 부르는 파, 마늘 같은 식재료에는 천연방부제인 항산화물질이 아주 많다. 파이토케미컬이라고 부르는 성분은

암이나 병균과 싸우는 면역세포 활성화에 좋다.

그러나 오늘의 가장 중요한 요리의 포인트는 탄수화물을 최소로 한 불고기를 만드는 것이었다. 애경 씨에게 가장 중요한 면역력은 혈당조절이기 때문이다. 생애주기에 따라 질병이 달라지듯 요리법도 나이에 따라 달라져야 한다. 다르게 먹어야 내가 먹는 음식이 나를 보호한다. 그렇지 않으면 음식 때문에 병에 걸리고 음식 때문에 병이 악화된다. 같은 소고기로 요리를 하더라도 물엿과 설탕을 듬뿍 넣고 거기다 참기름까지 고소하게 끼어엮은 불고기에 길들여 있다면 서서히 바꾸어야 한다.

꼬박 12시간을 촬영했다. 산에서는 해가 일찍 저서 저녁에는 추웠다. 애경 씨는 지쳐 보였다. 그러나 카메라가 돌아가면 언제 그랬느냐는 듯이 말을 잘했다. 나는 그 반대였다. 재미있게 이야기하다가도 카메라만 비추면 영 어색했다. 애경 씨는 취미가 연극이라고 했다. 정식으로 연기를 배운 적은 없지만 가끔 엑스트라 섭외가 들어오면 영화도 찍는단다. “수입이 많은 것은 아니지만 제가 참여한 영화가 개봉하면 주인공도 아니면서 괜히 뿌듯해요”라고 말했다.

촬영 팀은 정면에서 한번 찍고 드론을 띄워서 똑같은 장면을 한 번 더 촬영했다. 그러니까 시간이 두 배가 걸렸다. 우리도 같은 장면을 두 번씩 반복했다. 그녀는 드라마 촬영을 할 때는 똑같은 신을 정면, 좌우, 후면으로 계속 반복해 찍는다고 했다. 그러다 보면 밤을 꼬박 새고 혈당이 높아서인지 매우 지친다고 했다. “혈당이 조절되면 지금보다 덜 피곤하게 그 일을 하실 겁니다”라고 말하면서 나의 꿈도 곱곰이 생각해봤다.

여행은 떠나는 즐거움도 있지만 일상으로 행복하게 돌아오기 위한 기쁨도 있다. 애경 씨가 ‘힐링 여행’을 통해서 자신의 꿈을 건강하게 완성하는 멋진 일상이 펼쳐지길 바란다.

통계로 바라보는 세상 이야기...

뉴노멀 시대, 대한민국의 변화와 이슈

01

2020 여성 경력단절 원인 1위 '육아'

통계청은 1997년 이후, 매년 양성평등주간마다 여성의 모습을 부문별로 조명하는 '통계로 보는 여성의 삶'을 발표하고 있습니다. 2020년 우리나라에서 여성 인구는 총 인구의 49.9%(2,583만 5천 명)입니다. 2019년 여성 고용률은 51.6%로 10년 전(2009년, 47.8%) 대비 3.8%p 상승했습니다. 지난 4월에 있었던 제21대 국회의원 선거에서 당선된 의원 총 300명 중 여성이 57명으로 역대 최대 규모 및 비율(19%)을 기록했고, 이 중 여성 장관은 6명으로 전체(18명)의 33.3%에 달합니다. 2019년 15~54세 기혼여성 중 경력단절여성 비율은 19.2%(169만 9천 명)인데, '육아(38.2%)', '결혼(34.4%)', '임신·출산(22.6%)', '가족돌봄(4.4%)', '자녀교육(4.1%)' 때문에 경력단절이 일어났다고 합니다.

02

2020 Census 인구주택총조사 실시

인구주택총조사는 우리나라에 사는 모든 내·외국인과 주택의 규모 및 특징을 알 수 있는 기본적인 국가 통계조사입니다. 인구와 가구, 주택 등 일반적인 특성은 등록센서스로 매년 전수 조사하고 있으며, 지역의 복지와 경제, 교통 등 정책수립에 필요한 실질적인 자료를 수집하기 위해서 5년마다 우리나라 전체 가구의 20% 표본을 선정해 현장조사를 실시하고 있습니다. 인구주택총조사는 역사적이고 범세계적 통계조사로 고대 바빌로니아(B.C. 3600년경) 시대에 기원을 두며, 근대적인 인구조사는 1790년 미국에서 최초로 실시하였습니다. UN 통계청에 따르면, 2010 라운드 센서스(2005~2014년)에서는 214개국에서 인구총조사를 실시하여 전 세계 인구의 93%가 조사되었다고 합니다.

03

기후변화 대응과 경제성장, 그린뉴딜로!

기상청이 발간한 2019년 이상기후 보고서에 따르면 장마뿐만 아니라 폭염, 태풍, 가뭄 등 한반도에 미치는 이상기후 현상이 갈수록 뚜렷해지고 있습니다. 기후위기로 부터 지구를 구하기 위해 우리나라도 최근 한국형 그린뉴딜 정책을 발표했는데요. '그린뉴딜'이란 기후변화에 대응하는 동시에 고용도 촉진하는 정책으로, 환경과 사람이 중심이 되는 지속 가능한 발전을 뜻합니다. 환경부의 2020년 8월 5일 보도자료에 따르면, 목표관리제를 실시한 기관의 온실가스 배출량은 2019년 398만 톤 CO₂eq으로 기준배출량인 521만 톤 CO₂eq 대비 23.6% 감소했다고 합니다. 이러한 결과는 제도를 처음 시행한 2011년 배출량인 473만 톤 CO₂eq과 비교하면 15.8%를 감축한 성과라고 하네요!

04

2020년 대한민국 국민의 삶의 질은?

통계청은 2014년부터 '국민 삶의 질 지표'를 작성하고 있는데, 지난 제6회 국민의 삶의 질 측정 포럼의 '한국의 삶의 질 현황'에서 심수진 사무관은 2020년 3월 기준 건강, 여가, 안전, 등 총 11개 영역 내 71개 지표 중에서 건강수명, 1인당 국민총소득 등 개선된 지표가 45개(63.4%), 비만율, 범죄피해율 등 악화된 지표가 21개(29.6%)로 확인됐습니다. 한국청소년정책연구원 변금선 부연구위원은 청년들의 삶의 질 지표를 분석했는데, 경제지표에서 '채무 있음'으로 응답한 응답자는 전체의 16.81%, 건강지표에서 '정신건강이 나쁨'의 응답자는 28.10%, 행복지표의 행복만족도가 낮다고 답한 응답자는 26.61%에 달하는 등 청년 4명 중 1명 이상은 자신이 행복하지 않다고 답했습니다.

05

10월 5일 세계 한인의 날, 재외동포는 한국인인가요?

10월 5일은 세계 한인의 날입니다. 세계 한인의 날은 재외동포가 거주하는 국가 내에서 권익을 증가시키고 한민족으로서 정체성과 자긍심을 높이고자 마련된 날입니다. 외교부의 '외교부의 2019년 재외동포 정의 및 현황'에 따르면, 2019년에 발표된 전 세계 재외동포 수는 749만 4천 명으로 동북아시아(43.9%)가 가장 많았고 북미(37.2%), 유럽(9.2%), 남아시아태평양(7.9%), 중남미(1.4%), 중동(0.3%), 아프리카(0.2%) 순으로 나타났습니다. 재외동포가 한국 발전에 기여한다고 생각하는 사람이 2019년 49.9%로 2017년(46.0%)보다 3.9%p 증가한 모습을 볼 수 있고, 부정적으로 응답한 비율은 2017년 11.7%에서 2019년 9.0%로 떨어지면서 인식이 나쁘지 않음을 알 수 있죠.

06

너도 나도 주식 열풍, 시작해도 괜찮을까?

금융투자협회에서 2020년 5월 발표한 보도자료에 따르면 올해 초 2,935만 개였던 주식 활동계좌 수는 4월 말 3,125만 개까지 늘어났다고 합니다. '주린이(주식+어린이)' '영끌(영혼까지 끌어모으다)' 등의 용어가 유행하고, 최근 들어 주식을 시작하는 사람이 많아진 이유는 무엇일까요? 주식을 시작한 이유로는 '실수익 기대'가 17.2%로 1위를 차지했고, '예적금으로는 수익이 없어서(14.3%)', '단순 호기심(13.7%)' 순으로 나타났습니다. 트렌드모니터, 2020 주식투자 관련 인식 및 행태 조사를 보면 '개인투자자의 증가는 부의 상승이 어려운 현실과 관련 있다(71.5%)', '마땅한 재테크수단이 없기 때문이다(69.9%)' 등 주식투자에 대한 부정적인 인식을 엿볼 수 있습니다.

07

'라떼'는 가라, 이제는 리버스 멘토링의 시대!

한국행정연구원, 2019년 「사회통합실태조사」에 의하면 61.1%가 세대 간 소통이 이루어지지 않는다고 답했습니다. 최근 세대 간 소통을 원활히 하고, 이해를 높이기 위한 노력으로 급부상하고 있는 것이 바로 역멘토링, 리버스 멘토링입니다. 1999년 제너럴일렉트릭 회장이던 잭 웰치가 CEO 시절 리버스 멘토링을 통해 젊은 소비자들이 원하는 제품을 만들 수 있는 감각을 갖출 수 있다는 취지로 실시했는데, 이는 기업들 사이에서 꾸준히 도입되었다고 합니다. 성공적인 멘토링을 위해서는 첫째, 도입 목적과 영역을 분명하게 정의해야 하고, 둘째, 목적에 맞는 방식을 단계적으로 시행해야 하며, 셋째, 젊은 세대의 목소리를 존중하고 이를 경영에 적극 반영해야 한다고 합니다.

08

대한민국 청년을 통계로 파헤친다!

법으로 정해진 청년의 범위는 만 19세 이상부터 34세 이하입니다. 통계청, 2020년 8월 고용동향에 따르면 15~29세 청년층의 고용률은 42.9%, 실업률은 7.7%로 나타났습니다. 한국청소년정책연구원이 만 15~39세 이하 청년을 대상으로 실시한 청년사회·경제실태조사에서 2019년 '창업을 생각해 보았고 실제로 창업을 했다' 비율은 12.2%, '창업을 생각해 보았으나 실제로 창업을 해본 적은 없다' 비율은 25.3%로 나타났습니다. 잡코리아와 알바몬에서 직장인을 대상으로 2019년 저축 현황을 조사한 결과, 지난해 저축을 한 직장인 비율은 20대 75.3%, 30대 62.8%였는데, 이때 평균 저축액은 20대 직장인 776만 9천 원, 30대 직장인 906만 8천 원으로 30대가 금액이 높았습니다.

09

트렌드 중 트렌드는? 통계로 보는 MZ세대 트렌드

지금은 개인의 파워가 강해지고 소수의 트렌드 형성자들이 다수의 트렌드를 만들어 서로 즐기는 추세입니다. MZ세대란 1980~2000년생을 일컫는 밀레니얼 세대와 1995~2004년생을 뜻하는 Z세대를 합친 말로, 만 16세부터 40세까지가 포함됩니다. 통계청의 2019년 인구총조사에 따르면, MZ세대에 해당하는 인구는 총 1,797만 4천 명으로 전체 인구의 34.7%를 차지합니다. 트렌드모니터의 2020년 트렌드 관련 인식 조사에서, 요즘은 트렌드를 잘 알아야 하는 시대라고 73.7%가 답했고, 앞으로 트렌드를 공부하는 사람이 늘어날 것이라고 77%가 예측했죠. 국립국어원은 트렌드를 '사람들의 사고, 사상, 활동이나 일의 형세 따위가 움직여 가는 방향이나 추세'로 정의하고 있습니다.

10

취미와 재능 다잡은 하비프러너, 수입은 얼마?

얼마 전 직장인 커뮤니티 내에서 회자되고 있던 우스갯소리는 '퇴사하기'와 '유튜버되기'라고 합니다. 얼마나 많은 사람이 어떠한 이유로 N잡을 하고 있을까요? 예상했다시피 가장 큰 이유는 '수입을 높이기 위해서(86.8%)'입니다. 최근 취미로 수익을 창출하는 '하비프러너'가 늘고 있는데요, 잡코리아의 2020년 '하비프러너-취미' 관련 조사를 살펴보면 직장인 중 19.6%가 취미생활을 통해 추가 수익을 창출하고 있다고 답했는데요. '유튜브 등 SNS'가 44.5%로 1위를 차지했습니다. 유튜버를 부업으로 하는 사람들의 월 소득 중위값은 85만 원입니다. 하비프러너들에게 '향후 현재의 취미생활을 본업으로 삼을 계획이 있는지'에 대한 질문에 73.1%가 '그렇다'라고 답했습니다.

11

포스트 코로나, 통계 생산 패러다임 변화 필요!

지난 6월 24일 진행된 '제17차 OECD 통계정책위원회', 37개 회원국 통계기관장들이 참여한 올해 회의는 '코로나19 특별세션'을 통해 각국의 코로나19 관련 긴급 통계 작성 사례를 공유하였습니다. 회의에 참석한 강신욱 통계청장은 통계 작성과 빅데이터 분석을 통해 K-방역으로 이어진 사례를 발표했고, 코로나19 발생 초기에 마스크와 손소독제 등 예방품목이 품귀 현상을 보이자 신속하게 일일가격조사를 실시하여 해당 품목의 가격 및 수급 안정화 대책을 수립하는 기초자료로 활용될 수 있도록 하였습니다. 또한 모바일 빅데이터를 기반으로 코로나19 발생 전후 국내 인구 이동 패턴을 분석해 사회적 거리 두기 시기와 강도를 조절하는 정책 결정을 뒷받침하였다고 합니다.

12

스마트폰 없이 아이 키울 수 있을까요?

급격하게 확산되고 있는 영유아의 스마트폰 사용, 요즘 아이들은 몇 살 때부터 스마트미디어를 접할까요? 육아정책연구소의 '2019년 영유아의 스마트미디어 사용실태 및 부모인식 분석'에서 12개월 이상 만 6세 이하 자녀를 둔 부모를 대상으로 조사한 결과, 스마트미디어를 처음 접하는 시기는 만 1세가 45.1%로 가장 많았으며, 만 2세(20.2%), 만 3세(15.1%) 순으로 높았습니다. 특히 영유아에게 스마트미디어를 보여주는 가장 큰 이유는 '아이에게 방해받지 않고 다른 일을 하기 위해서'(31.1%)였습니다. 영유아 스마트미디어 중독을 치유하기 위해서는 부모와 자녀가 함께 스마트폰 이용규칙을 정하고, 부모와 아이가 상호작용할 수 있는 과정을 마련해주는 게 좋다고 합니다.

13

뉴노멀이 된 재택근무, 코로나19 잠잠해도 계속?

지난 7월, 한국경제연구원의 국내 매출액 500대 기업을 대상으로 한 ‘코로나19 이후 근로형태 변화 및 노동환경 전망 조사’ 결과, 국내 대기업의 75.0%가 코로나19에 대응하여 유연근무제를 활용하고 있었습니다. 유연근무제를 운영하는 기업의 51.1%가 코로나19가 진정된 이후에도 유연근무제를 지속, 확대할 계획이라고 응답했습니다. 또한 트렌드모니터의 ‘2020 직장인 재택근무 관련 인식 조사’에 따르면, 재택근무로 일해 본 직장인의 84.4%가 재택근무 경험에 만족스러웠다고 평가했습니다. 재택근무의 활성화를 위해 가장 필요한 조건으로 ‘자율적 재택근무를 위한 기업문화의 조성’(58.9%)을 1위로 꼽았습니다. 기술적 문제보다도 다른 부분에 대한 보완을 언급한 것이 인상적입니다.

14

냉동고에 방치된 아이스팩, 재사용 쉬워진다!

환경부의 2020년 7월 29일 보도자료에 따르면, 2019년 아이스팩 사용량은 2억 1천만 개로 추정되며, 이는 2015년 대비 약 2배입니다. 이 중 80%는 종량제 봉투에 버려져 소각되거나 매립되고 15%는 하수구에 버려져 미세 플라스틱의 원인이 되고 있다고 합니다. 사실 아이스팩 재사용 의사는 있는데, 재사용 비용이 높고 제조사별 규격이 달라 그동안 재사용이 어려웠습니다. 이에 정부는 아이스팩 크기와 표시사항을 표준화해서 아이스팩 재사용을 활성화하고, 일상생활 속에서 헛갈리는 분리수거를 돕고자 분리배출법 표시도 바꾼다고 하는데, 기존의 표시는 ‘종이팩’, ‘플라스틱’처럼 재질만 표기되어 있었지만, 앞으로 ‘깨끗이 접어서’, ‘라벨을 떼서’처럼 방법까지 함께 기재합니다.

15

9월 21일은 치매극복의 날, 잊지 말아야 할 치매!

2019년 9월 보건복지부 보도자료에 따르면 치매국가책임제 시행으로 치매 원인과 진단, 예방, 치료기술 개발을 위해 2020년부터 2028년까지 9년간 예산 2,000억 원을 투입한다고 했는데요. 특히 건강보험 제도 개선으로 중증치매질환자의 의료비 부담비율이 최대 60%에서 10%로 대폭 낮아졌습니다. 중앙치매센터의 치매 예방수칙 3·3·3에 따르면 3가지를 즐기고, 3가지를 참으며, 3가지를 챙길 것을 권장했습니다. 우선 운동과 식사, 독서를 즐기며 건강한 생활 습관을 만들고 금주, 금연 등 몸에 좋지 않은 것은 멀리합니다. 또 운동할 땀 보호장구를 착용해 뇌손상을 예방해야 하죠. 마지막으로 가족과의 소통, 정기적인 건강검진, 매년 보건소에서의 치매 조기 검진을 안내했습니다.

16

통계를 보면 안다고? 유튜브 유입률 높이는 꿀팁!

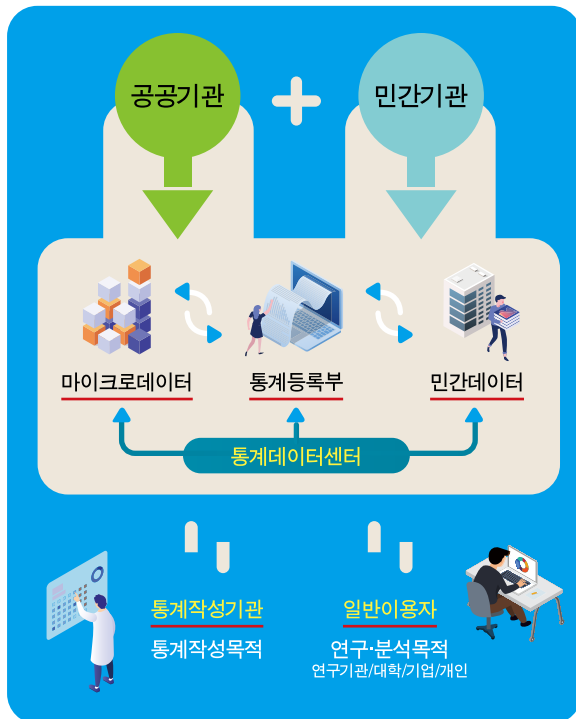
교육부의 2019년 초·중등 진로교육 현황조사를 보면, ‘크리에이터’가 초등학교 희망직업 3위(5.7%)를 차지했습니다. 통계청 제12기 기자단 이나영 기자는 유튜브 유입을 높이는 꿀팁으로 첫째, 구글 트렌드를 활용해서 유튜브 검색 키워드량을 분석하라고 합니다. 둘째, 유튜브 크리에이티브 대시보드, 혹은 스튜디오 활용하기입니다. 본인 채널에 대한 기본적인 개요와 영상에 대한 통계들을 분석해볼 수 있는 아주 유용한 기능입니다. 셋째, 트래픽 소스 확인하기입니다. 쉽게 말해서 내 채널의 영상을 보는 사람들은 어떤 경로로 유입이 되는지 확인하는 겁니다. 마지막은 섬네일입니다. 섬네일은 가독성이 높고 영상의 특징을 잘 보여주는 문구나 사진이 들어가면 매우 좋습니다.

행정통계자료와 민간자료를 한곳에!

통계데이터 서비스

통계데이터센터가 새로운 서비스로
정보화 사회를 선도합니다.

행정자료를 수집하여 가공한 행정통계자료(통계등록부),
통계청이 제공하는 승인된 통계기초자료(마이크로데이터) 등
통계자료뿐만 아니라 민간자료까지 한 곳에서 분석이 가능한 통계데이터센터(SDC)



제공서비스

분석센터 이용서비스

- 데이터분석 플랫폼 제공 (분석시스템, 통계패키지)
- 통계자료 및 민간자료, 이용자 반입자료 등 연계·분석

주문형 분석서비스

- 시간 및 거리상 센터 방문이 어려운 이용자를 위한 서비스
- 이용자가 원하는 형태로 결과 제공(비식별화처리된 형태)

교육지원

- 대학생을 대상으로 통계데이터 센터의 실제 데이터를 통한 다양한 분석 체험 기회 제공
- 통계학과 등 관련학과와 정규 전공 수업 실습 분석 지원 및 이용 지원

전문가분석 지원

- 분석 경험이 없는 이용자를 위한 분석 지원 및 상담

빅데이터를 활용한 서비스 (홈페이지 내 제공)



제공자료

통계등록부

부문	자료명	시계열	주기
경제 · 사회	기업등록부DB(사업자등록기준)	2010~2018년	년
	기업등록부DB(대표자기준)	2010~2018년	년
	기업등록부DB_분기(사업자등록기준)	2018년 4분기~2020년 1분기	분기
	기업등록부DB_분기(대표자기준)	2018년 4분기~2020년 1분기	분기
	사업장 기초DB(4대보험)	2009~2016년(17,18년 중지)	년
	사업장 기초DB(법인)	2008~2018년	년
	중사자-기업체 연계DB	2015~2018년	년
	육아휴직 사용현황DB	2015~2018년	년
인구 · 가구	등록센서스 인구DB	2015~2018년	년
	등록센서스 가구DB	2015~2018년	년
	등록센서스 주택DB	2015~2018년	년
농림어업	농업DB	2015~2018년	년
	임업DB	2015~2018년	년
	어업DB	2015~2018년	년

민간데이터

부문	자료명	범위	시계열	주기	부문	자료명	범위	시계열	주기
인구 · 가구	유동인구정보(성연령)	전국	2015.01.~ 2020.12.	월	경제 · 사회	통근인구의 평균소득소비	전국	2019.10.~ 2020.12	분기
	유동인구정보(시간)	전국	2015.01.~ 2020.12.	월		유통고객 구매상품 정보	전국	2019.12.~ 2020.11.	월
	유동인구정보(요일)	전국	2015.01.~ 2020.12.	월		1000대 상권 정보	전국	2018	년
	유입인구정보(성연령)	부산	2017.01.~ 2018.06.	월		영업 중 업소	부산	2017~2018	월
	주거인구정보(성연령)	전국	2016~2019	년	건물 · 지가	표준공시지가	서울, 부산	2016~2017	년
	직장인구정보(성연령)	전국	2016~2019	년		주택(단독/다세대/기타)	서울, 부산	2016~2017	년
	건물별 가구평균소득	서울, 부산	2016~2017	년		공동주택정보	서울, 부산	2016~2017	년
	(코로나19) 인구이동정보	전국	2020.01~ 2020.05	월		건축물정보	전국	2017~2018	월
	방문인구정보	전국		월	차량	인구집중유발시설	서울, 부산	2016~2018	분기
경제 · 사회	성 · 연령별 카드매출	전국	2015.01.~ 2018.06.	월		도로별 차량통행량	서울, 부산	2017	년
	업종별 매출액	전국	2015.01.~ 2018.06.	월		시 · 군 · 구별 차량등록대수	전국	2017	년
	성 · 연령별 카드매출	전국	2018.10~ 2020.12	월		행정동코드 정보	전국	2018	년
	신용융합통계정보_대출	전국	2018.10.~ 2020.12	월		법정동코드 정보	전국	2018	년
	신용융합통계정보_소득	전국	2018.10.~ 2020.12	월	기타	SNS 지역축제정보	전국	2017	년
	신용융합통계정보_소비	전국	2018.10.~ 2020.12	월		온라인가격 정보	전국	2014.01~ 2020.09	월
	전입인구의 평균소득소비	전국	2019.10.~ 2020.12	분기		모바일상품권 거래 정보	전국	2018	년
	전출인구의 평균소득소비	전국	2019.10.~ 2020.12	분기					

통계청, 정부부처, 지방자치단체, 연구기관 등
모든 기관의 마이크로데이터를 한 곳으로...

보다 심도 있고 다양한 분석을 원한다면
지금 바로 **MDIS**를 클릭해보세요

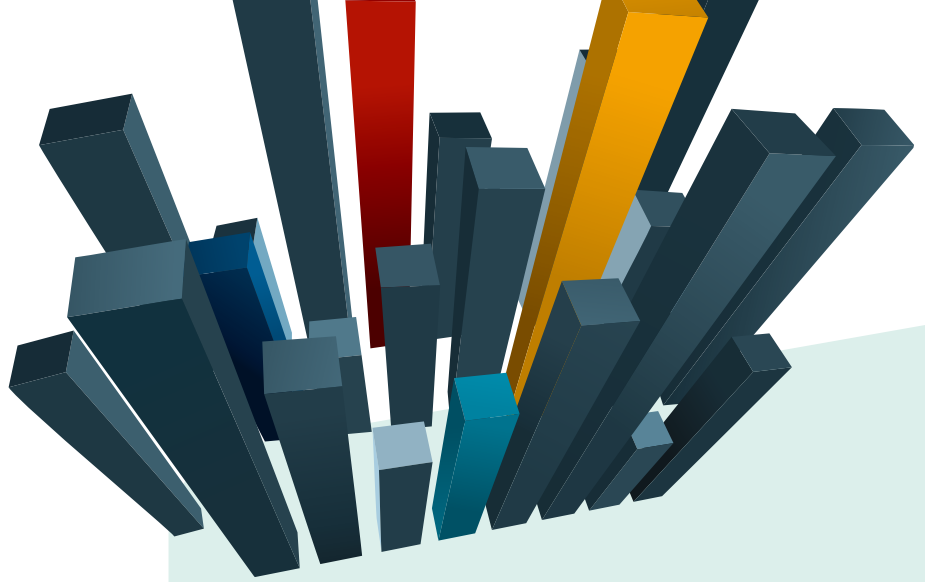
■ 서비스 소개 (2020년 11월 기준)

가. 서비스명 : 마이크로데이터통합서비스(MDIS, mdis.kostat.go.kr)

나. 제공통계수 : 총 265종(통계청 48종 및 통계작성기관 217종)

다. 제공형태 : 마이크로데이터(통계에 따라 사람, 사업체, 가구 기반 자료)

기준		주요 통계
통계청	인구 · 가구	경제활동인구조사, 가계동향조사, 국내인구이동통계, 사망원인통계, 가계금융복지조사, 지역별고용조사, 인구주택총조사, 인구동향조사, 생활시간조사, 사회조사 외 5종
	사업체 · 농어가	전국사업체조사, 광업·제조업조사, 농가경제조사, 기업활동조사, 농업총조사, 농산물생산비조사, 경제총조사, 어가경제조사, 운수업조사 외 16종
	행정통계 및 기타	귀농귀촌인통계, 영리법인기업체행정통계, 신혼부부통계, 주택소유통계, 중장년층행정통계, 퇴직연금통계, 일자리행정통계, 기업생멸행정통계
통계작성기관		전국다문화가족실태조사, 가족실태조사, 자동차주행거리통계, 직종별사업체노동력조사, 보육실태조사, 기상관측통계, 국민여가활동조사, 외래관광객실태조사, 한부모가족실태조사, 청소년종합실태조사 외 207종



■ 서비스 내용

가. 구분: 자료의 민감성 정도에 따라
공공용, 인가용, 특수목적용으로 구분 운영

나. 수수료

- 무료: 공공용 자료
- 인가용/특수목적용: 선택제 수수료 부과

다. 서비스 방법

- 추출·다운로드: MDIS 포털에서
직접 무료 다운로드
- 원격접근서비스: 승인 후 이용자가
집·사무실 등에서 통계청 서버 접속 후 활용
- 이용센터: 승인 후 지정된 장소를 방문·활용

■ 문의

- 연락처: 재단법인 한국통계진흥원
- 전화: (02) 512-0167 FAX: (02) 515-0240
- 주소: (우) 06097 서울특별시 강남구 선릉로 612, 6층
- E-mail: MDIS@stat.or.kr

공공자료의 개방 및 공유 확대

MDIS

[mdis.kostat.go.kr]

**통계청 46종 및 통계작성기관 230종의
통계자료로 제공확대**

MDIS 소개 제공자료 자료이용 이용자지원 마이페이지 LOGIN

마이크로데이터 통합서비스

마이크로데이터는 일상의 의사결정에서부터 국가 주요정책결정에 이르기까지 다양한 분야에서 활용할 수 있습니다.

**RDC 심층분석을 위한
제공확대 확대**
통계청은 이용자의 요청을 적극적으로 검토하여 제공정책을 확대하여 서비스중입니다.
[자세히보기 >](#)

MDIS 주요 서비스

마이크로데이터 이야기

이용자가 원하는 공공용 마이크로데이터를 20만 개에 달하는 마이크로데이터 분석 분석물을 할 수 있는 방법입니다.

다운로드 서비스

공공용 자료에 대해 상세한 수준에 자료를 통계청과 서비스나 마이크로데이터 이용안내에서 이용할 수 있는 서비스입니다.

인가용 서비스

공공용 자료를 가공하여 정부 서비스, 민간기업 등에 활용 가능하도록, 통계청에 특화된 서비스입니다.

공지사항

18
2016

'2011~2019년 경제활동인구조사(4...
'2011~2019년 경제활동인구조사(4분기), 제1권 MDIS 서비스를 다음과 같이

태마존

공공자료

인가자료

POPUPZONE

등록된 팝업이 없습니다.

통계청에서 국가통계를 활용하세요!

통계청은 통계개발·활용에 필요한 모든 정보와 도움을 제공합니다.
다양한 국가통계정보 제공사이트를 활용하세요.

원하는 자료를 직접 분석 및 요청

MDIS
[mdis.kostat.go.kr]

온라인으로 추출/다운로드 선택 시
공공용 마이크로데이터를
무료로 분석 활용 가능



국가통계 쉽게 찾기

KOSIS
[kosis.kr]

국내, 국제, 북한의 주요 통계를
한 곳에 모아 알기 쉽게 분류해 제공



행정통계자료와 민간자료를 한곳에

통계데이터센터
[data.kostat.go.kr]

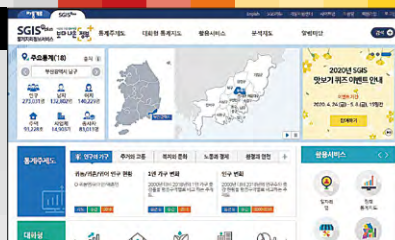
행정통계자료(통계등록부), 민간자료의
연계·융합이 가능한 데이터 플랫폼



지도 위 통계정보 살펴보기

SGIS
[sgis.kostat.go.kr]

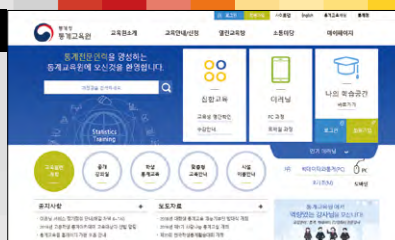
인구, 가구, 주택, 사업체 통계 등 각종
통계를 지도(GIS) 위에서 한눈에 파악



국내 유일의 국가통계교육 전문기관

통계교육원
[sti.kostat.go.kr]

통계작성 및 활용 전문통계과정,
기관맞춤형과정, e-러닝 과정



통계청
통계교육원