

Jock D. Mackinlay 및 Kevin Winslow

---

# 뛰어난 비주얼리제이션 디자인

# 목차

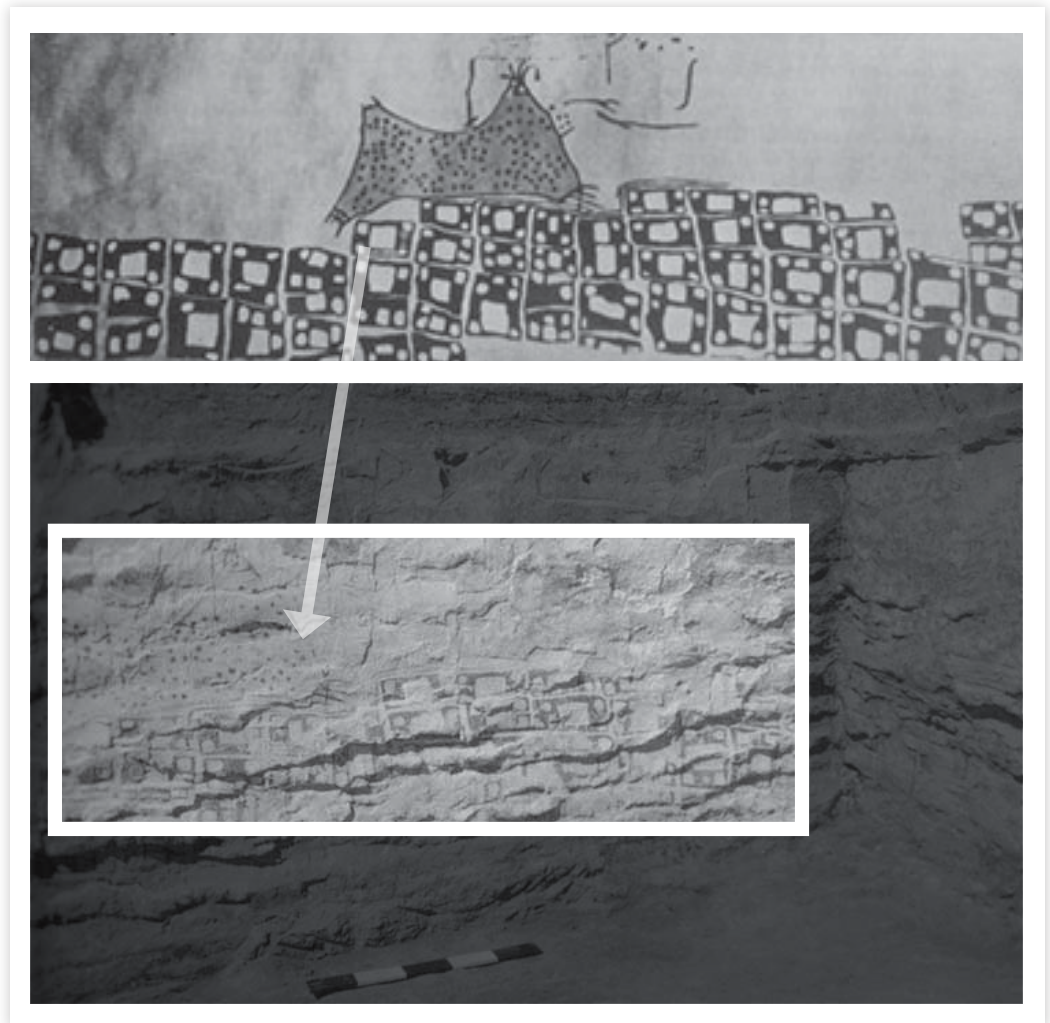
과거의 시각화. ....	3
필요에 따라 다른 데이터 시각화 사용. ....	4
1700-1900: 비주얼리제이션의 전환. ....	5
발전과 오용이 함께 등장한 20세기. ....	6
컴퓨터 기반 비주얼리제이션: 개인의 역사. ....	7
현재: 인간의 인지 능력. ....	9
데이터 유형에 따라 다른 뷰 사용. ....	10
인간의 인지: 능력 및 한계. ....	11
구성 및 상호 작용 활용. ....	12
데이터를 활용한 스토리텔링. ....	13
점검하기: 배운 교훈. ....	16

이 책에서는 초기 동굴 벽화에서부터 컴퓨터 혁명과 Tableau 출시에 이르기까지 시각적 표현의 역사를 추적합니다. 그리고 데이터 연구의 선구자에 대해 알아보며 선구자들의 업적으로 인해 어떻게 시각적 커뮤니케이션 기법이 혁신되었는지 설명합니다. 또한, 다양한 스타일의 데이터 시각화를 검토하고, 효과적인 시각화를 구현하는 데 방해가 되는 장애 요소를 살펴보며, 이러한 장애 요소를 극복하기 위한 방법에 대해 설명합니다. 마지막에는 인간의 인지 능력(및 한계)과 초기 동굴 벽화와 매우 유사하게 데이터를 활용하여 스토리를 전달할 수 있는 방법에 대해 살펴봅니다.



## 과거의 시각화

시각적 표현은 인간의 학습 및 이해에 기본적인 구성 요소입니다. 데이터 시각화의 영향과 진화에 대해 알아보려면 우선 과거를 확인하고 시간이 지남에 따라 데이터 구성과 표현이 변화해 온 방법을 살펴봐야 합니다. 인류가 시작된 이래로 인류는 비주얼리제이션을 활용하여 무언가를 지시하거나, 의미와 스토리를 전달하고 있습니다. 특히, 지도는 오랜 역사를 가지고 있습니다. 아래 그림은 BC 6200년 터키에 지어진 9피트 크기의 벽화입니다. 전경에서는 Hasan Dag 화산이 뒤에 보이는 도시인 Catal Hyük를 볼 수 있습니다[1, 2]. 이 벽화의 의도는 확실하지 않으며, 예술 작품일 수도 또는 데이터 맵일 수도 있습니다. 의도가 무엇이든 간에 이 벽화는 인류가 그림을 활용하여 개념과 아이디어를 소통하는 초기 방법의 예가 됩니다.



## 필요에 따라 다른 데이터 시각화 사용

데이터의 일반적인 시각화 표현 유형은 두 가지입니다. 두 유형 모두 매우 중요하지만, 뛰어난 비주얼리제이션을 디자인할 때의 요구 사항은 각각 다릅니다.

첫 번째 유형은 제시(Presentation)로 데이터 시각화를 활용하여 소통합니다. 이러한 유형의 시각화 표현에는 발표자와 청중이라는 두 가지 역할이 있습니다. 기본적으로 수행하는 기능은 소통 및 설득입니다. 예를 들어, 다음 점토 명판은 도시의 방어 체계를 복원하는 방법을 설명하는 데 사용된 것으로 보이는 바빌론 니푸르의 메소포타미아 도시 계획으로 추정됩니다[3].

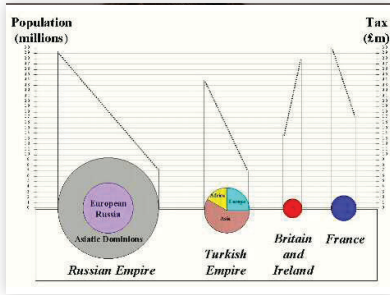


두 번째 유형은 비주얼리제이션입니다. 비주얼리제이션은 상당히 새로운 용어로 시각화를 활용하여 사고하는 개념입니다. 여기에서는, 사람들이 질문에 대한 답을 찾으려고 하는 것과 같은 능동적인 체험이 일어납니다. 지도 비유로 돌아가, 1959년부터 바다 여행과 탐색에 활용된 유명한 Mercator Projection 지도에 대해 생각해 보겠습니다[4]. 이러한 유형의 지도는 작업 수행을 위해 활용될 수 있습니다. 지점 사이를 선으로 이음으로써 지도는 거리가 먼 이동 경로를 보여줄 수 있습니다. 사람들은 지도를 활용하여 경로를 결정할 수 있습니다. 이러한 유형의 '질문 응답'은 비주얼리제이션 데이터 표현의 예입니다.



## 1700-1900: 비주얼리제이션의 전환

비주얼리제이션 역사에서 중요한 시점은 통계적 제시의 아버지라고 불리는 스코틀랜드 엔지니어인 William Playfair와 관련이 있습니다. Playfair는 1786년 영국의 무역 수지를 설명하기 위해 데이터를 도표로 표현한 책인 경제와 정치의 지도(Commercial and Political Atlas)를 출간했습니다. Playfair의 혁신적인 여러 데이터 시각화는 현재에도 활용되고 있습니다. Playfair의 발명품 중 하나인 파이 차트는 왼쪽 도표에서 확인할 수 있습니다[5].

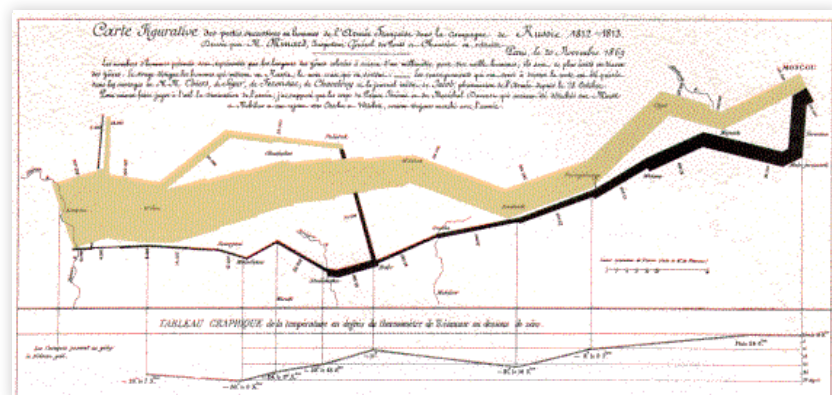


여기를 클릭하여 파이 차트 모범 사례에 대해 알아보십시오.

제시를 위해 통계 도표가 사용되기 시작한 것은 그리 오래되지 않았습니다. 통계 도표를 활용하여 런던의 1855년 콜레라 전염병을 표시한 영국의 의사인 Dr. John Snow의 사례도 유명합니다. Snow는 런던의 지도에 개별 콜레라 발병을 점으로 표시했습니다. 이러한 점들은 대부분의 발병이 Broad Street의 워터 펌프로 추적될 수 있음을 보여주었습니다. 특이 사례를 조사한 결과도 Broad Street 펌프와 연관이 있음을 보여주었습니다. Snow는 오염된 펌프의 핸들을 제거했고 콜레라 전염이 소강상태가 되었습니다. 이는 비주얼리제이션이 질문 응답에서 얼마나 큰 힘을 발휘하며 이 사례에서도 볼 수 있듯이 공공의 이익을 위해서도 효과가 있음을 보여줍니다. 또한, Snow의 지도는 제시 스타일의 효과적인 예시가 되었습니다. Snow의 데이터는 도시 공무원들이 오염된 핸들을 제거하여 발병을 진압하도록 설득하기에 충분히 강력했습니다.

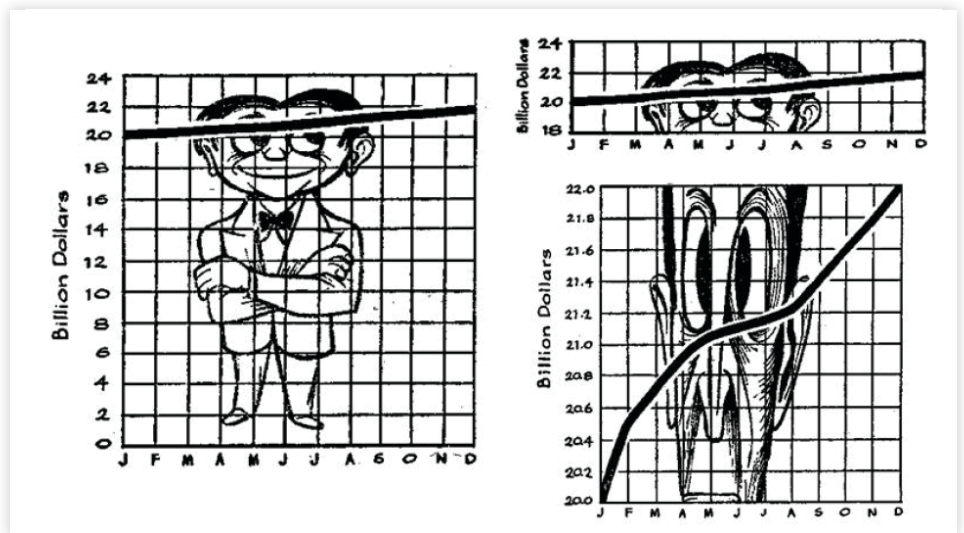
데이터 제시의 가장 유명한 예는 나폴레옹의 모스크바 행진(및 이후의 퇴각)과 관련한 스토리를 설명하기 위해 비주얼리제이션을 활용한 프랑스 토목 엔지니어 Charles Minard의 사례입니다. 아래 그림에서 볼 수 있듯이, Minard의 그림은 밝게 음영 처리된 막대를 사용하여 전진하는 군대의 규모를 나타냅니다[6]. 군대가 모스크바를 향해 나아감에 따라 막대의 두께가 지속적으로 감소합니다. 아래 검정색 막대는 모스크바에서 퇴각 시 군대의 힘이 감소함을 보여줍니다.

그림의 하단에는 외부 온도를 추적하는 선이 있고, 외부 온도로 인해 군대의 규모가 크게 손실되었음을 나타냅니다. 도표 중간의 검정색 막대는 앞서 갈라진 부대가 주력군에 재합류할 때 일시적으로 커집니다. 그러나 차트에 그려진 강을 지나 이동하는 부분에서 막대가 훨씬 적게 증가하며 이는 차가운 물의 영향을 보여줍니다. 음영 처리된 검정색 막대는 행진 시작 시 마지막으로 다시 만나고 이를 통해 보는 사람들은 이전에는 대규모였던 군대가 소수의 병사로 줄어든 방법을 명확하게 살펴볼 수 있습니다. 이러한 방식으로 단순 차트는 매우 강력한 스토리를 전달할 수 있습니다.



## 발전과 오용이 함께 등장한 20세기.

1900년대 중반에는 통계 도표의 인기와 오용이 모두 증가했습니다. 이러한 오용으로 1955년 미국의 작가 Darrell Huff는 새빨간 거짓말, 통계(How to Lie with Statistics)라는 책을 출간했습니다. Huff가 작성한 조작 유형의 예는 아래 도표에서 확인할 수 있습니다. 왼쪽 차트는 데이터를 정직하게 나타내는 반면 오른쪽 상단 차트는 극적인 소비 증가를 표시하기 위해 0인 지점을 제거하였습니다. 오른쪽 하단의 차트는 경고성 수치를 보여주기 위해 한번 더 조작되었습니다. Huff의 책은 상당히 다른 내용의 스토리를 전달하기 위해 어떻게 시각화가 조작될 수 있는지를 보여줄 뿐만 아니라 효율적인 통계를 설계하는 방법을 알려주기 때문에 가치가 있고 권장됩니다.

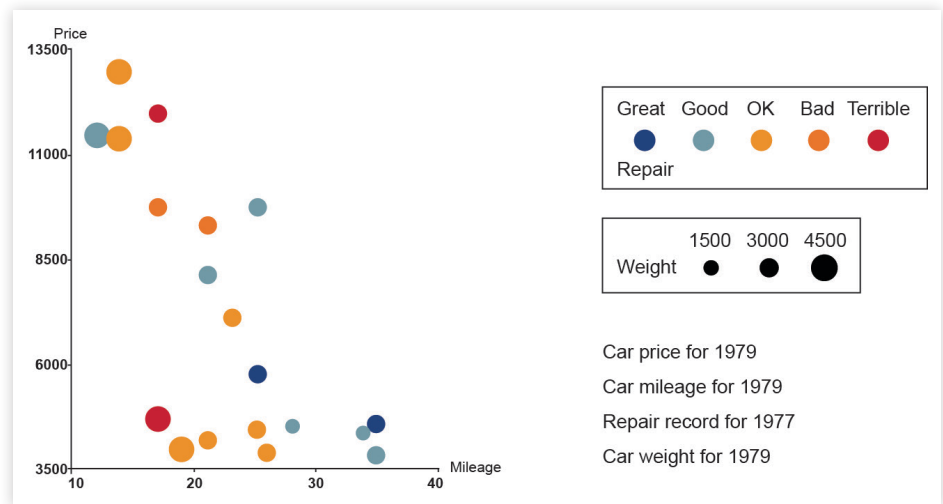


십수 년 후, 프랑스의 이론가 Jacques Bertin은 그래픽의 기호학(Semiology of Graphics) [8]라는 책을 출간했습니다. Bertin은 특히 통계적 매핑에 관심을 두고 데이터 뷰에 다음과 같이 3가지 종류의 마크 (점, 선 및 영역)가 포함됨을 발견했습니다. 이러한 마크는 특정 속성을 가지며 그 중 가장 중요한 속성은 위치(Position)입니다. 위치 (Position) 외에 다른 6가지 속성은 색상(Color), 크기(Size), 모양(Shape), 명도(Gray), 방향(Orientation) 및 질감(Texture)입니다. Bertin은 이러한 마크와 속성을 활용하여 그래픽적 데이터 디자인 방법에 관한 지침을 마련했습니다.

또한, Bertin은 테이블 정렬 등 시각적 분석에서 매우 중요한 기법을 개발했습니다. 치환 행렬이라는 이 기법은 테이블의 행과 열을 조작하여 이전에 정렬되지 않은 데이터에서 패턴을 발견하고 값 사이의 상관 관계를 보여줍니다. 혁신가였던 Bertin은 1977년 컴퓨터 기반 비주얼리제이션과 관련한 초기 활동에 대해 설명한 그래픽 및 그래픽 정보 처리(Graphics and Graphic Information Processing)라는 다른 도서를 출간했습니다. Bertin의 컴퓨터는 그의 작업을 수행하기에는 너무 원시적이었지만 기술의 폭발로 이 분야가 곧 크게 발전되었습니다.

## 컴퓨터 기반 비주얼리제이션: 개인의 역사.

비주얼리제이션 및 제시의 역사는 컴퓨터의 발명으로 빨라졌습니다. 여기서는 저의 개인 역사에 대해 중점적으로 설명하겠습니다. 1986년 필자는 APT(A Presentation Tool)라는 박사학위 논문을 작성했습니다. Bertin의 기호학을 확장 및 자동화함으로써 필자는 아래 그림과 같이 그래픽을 통한 제시를 자동으로 디자인하는 제시 도구를 개발했습니다. 도표는 전통적인 분산형(또는 버블) 차트를 4차원으로 보여줍니다. 디자인 작업에서 컴퓨터를 활용할 수 있는 효과적인 데이터 뷰를 디자인하는 방법에 대해 우리가 충분히 알고 있음을 보여주었기 때문에 이것은 혁신적이었습니다.

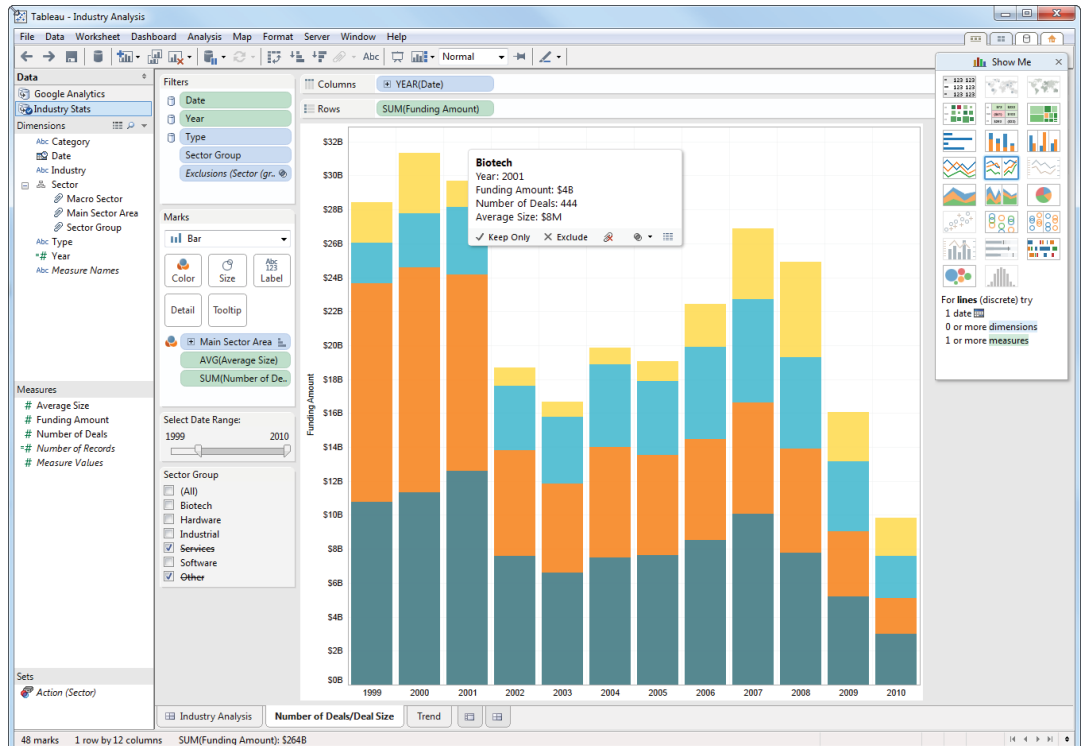
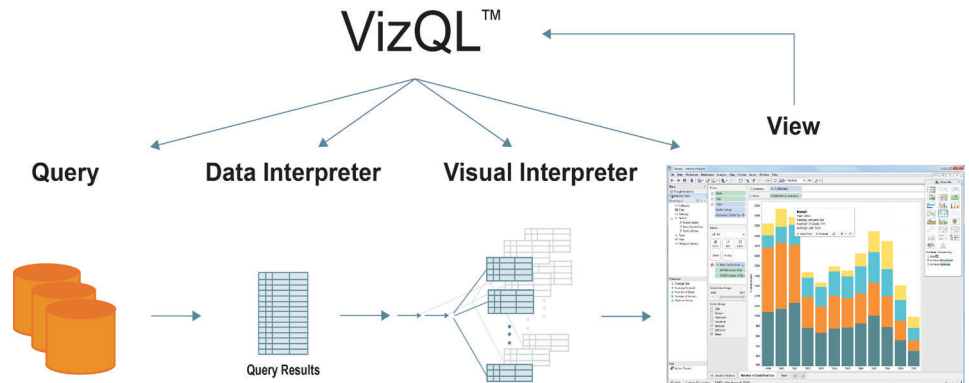


같은 해, 미국 국립과학재단(NSF)의 패널은 과학적 비주얼리제이션 및 특히 컴퓨터 그래픽을 활용한 과학적 문제 해결에 관한 보고서를 작성했습니다. 이 보고서에서 활용된 컴퓨터로 설계된 뇌우 이미지는 매우 효율적이었으며 비주얼리제이션의 지원을 위해 컴퓨터를 활용하는 것과 관련하여 의회로부터 상당한 추가 연구 지원을 받을 수 있었습니다.

1987년 Richard Becker와 William Cleveland는 현재 대부분 문제의 핵심이 되는 다차원 데이터를 다루었습니다. 상호 작용 브러싱(interactive brushing)이라는 기법을 사용하면 한 세트의 점을 선택한 후 모든 다른 뷰에 이 점들을 표시할 수 있습니다. 이러한 상호 작용 요소는 다차원 공간에 대한 이해도를 더욱 확장해 주었습니다.

90년대에 필자는 Xerox Palo Alto Research Center에서 근무하던 Stuart Card 및 George Robertson과 팀을 구성했습니다. 우리는 컴퓨터를 활용하여 Perspective Wall 및 Cone Tree 등 2차원 및 3차원 뷰를 개발했습니다. 이 연구를 통해 필자는 Card 및 Ben Shneiderman과 함께 정보 비주얼리제이션 읽기(Readings in Information Visualization)라는 책을 출간했습니다. 이 책은 컴퓨터 기반 비주얼리제이션에 대한 연구는 그래픽 뷰에 데이터를 인코딩하기 위한 기본 구조 및 프로세스를 가지며 여기에는 데이터와 뷰 모두가 포함됨을 보였습니다. 원시 데이터에서 시작하여 데이터를 변환하고 시각적 구조로 인코딩한 다음 뷰 변환을 사용하여 모두를 컨텍스트 안에서 표시합니다. 이러한 프로세스는 개인 작업에 적용되었으며 해당 작업자가 모든 변환을 제어했습니다. 본질적으로 이러한 프로세스는 대화형 데이터 시각화를 설명합니다. 언급되지 않은 사항은 원시 데이터 원본입니다.

그때 Tableau가 등장했습니다. 2003년 Chris Stolte는 스탠포드 박사학위 논문에서 효율적으로 원시 데이터에 액세스하는 방법을 설명했습니다. Stolte와 지도 교수 Pat Hanrahan은 데이터 그래픽 및 데이터베이스로의 연결 방법을 설명하는 VizQL 라는 명세 언어를 개발했습니다. 이 작업은 Bertin의 기호학 및 필자의 박사학위 논문을 기반으로 수행되었습니다. 또한, VizQL은 직관적인 끌어 놓기 시스템을 지원했으며 이것이 현재의 Tableau 인터페이스로 진화되었습니다. 본질적으로 VizQL 를 사용하면 사용자는 질의를 수행한 다음 결과를 뷰로 전환할 수 있습니다. Stolte는 데이터를 시각화하는 것에 성공하였고 실제 응용 프로그램으로 개발했습니다.



여기를 클릭하여 VizQL™ 기술에 대해 자세히 알아보십시오.



## 현재: 인간의 인지 능력

이러한 모든 역사가 우리에게 알려준 한 가지는 인간의 인지 능력이 강력하다는 것입니다. 그러나 인간의 인지는 시각적 유도를 통해 도움을 받고 증강될 수 있습니다.

예를 들어, 아래 그림을 보고 왼쪽 테이블에 집중해 보십시오. 이제 숫자 9가 한 눈에 몇 개나 보이는지 시험해 보십시오. 이제 오른쪽의 테이블을 보십시오. 이제 9가 빨간색으로 표시되고 이러한 시각적 유도는 숫자 9의 개수를 세는 시간을 몇 초로 줄여줍니다. 이는 '팝아웃'이라는 매우 오래된 기법이며 데이터 비주얼리제이션으로 쉽게 대규모 데이터 집합을 파악하고 발견 사항을 이해할 수 있도록 해주는 방법 중 하나입니다.

3 3 0 3 0 1 8 7 6 8 2 1 4 0 3 8 3 7 7 2 0 6 2 3 2 7 0 2 0	3 3 0 3 0 1 8 7 6 8 2 1 4 0 3 8 3 7 7 2 0 6 2 3 2 7 0 2 0
7 1 4 4 6 0 2 1 3 2 7 6 0 2 5 6 3 2 5 7 6 3 0 2 0 3 0 7 2	7 1 4 4 6 0 2 1 3 2 7 6 0 2 5 6 3 2 5 7 6 3 0 2 0 3 0 7 2
8 7 5 7 2 8 3 8 7 7 8 2 0 7 7 5 2 3 1 1 5 6 3 8 4 7 8 2 0	8 7 5 7 2 8 3 8 7 7 8 2 0 7 7 5 2 3 1 1 5 6 3 8 4 7 8 2 0
0 5 0 5 1 6 1 7 5 6 8 0 4 4 6 7 4 7 1 4 0 0 8 4 4 3 0 3 2	0 5 0 5 1 6 1 7 5 6 8 0 4 4 6 7 4 7 1 4 0 0 8 4 4 3 0 3 2
2 4 3 1 3 5 4 9 5 0 7 6 0 7 4 3 1 8 2 7 3 4 6 0 2 4 8 2 3	2 4 3 1 3 5 4 9 5 0 7 6 0 7 4 3 1 8 2 7 3 4 6 0 2 4 8 2 3
8 6 2 2 6 5 4 6 7 0 7 6 0 0 3 9 1 8 2 7 1 7 2 3 3 5 8 7 0	8 6 2 2 6 5 4 6 7 0 7 6 0 0 3 9 1 8 2 7 1 7 2 3 3 5 8 7 0
0 8 4 5 1 3 1 7 6 4 5 4 1 2 4 5 3 3 5 4 9 6 7 7 6 3 4 2 5	0 8 4 5 1 3 1 7 6 4 5 4 1 2 4 5 3 3 5 4 9 6 7 7 6 3 4 2 5
4 7 7 0 2 2 0 1 1 7 7 7 0 2 6 6 4 7 5 8 6 1 4 3 7 8 5 4 6	4 7 7 0 2 2 0 1 1 7 7 7 0 2 6 6 4 7 5 8 6 1 4 3 7 8 5 4 6
4 3 6 6 4 6 6 2 8 4 8 5 3 7 8 8 1 3 8 5 4 5 7 4 0 3 2 8 4	4 3 6 6 4 6 6 2 8 4 8 5 3 7 8 8 1 3 8 5 4 5 7 4 0 3 2 8 4
5 5 0 3 5 3 5 3 8 3 2 3 8 2 3 1 6 2 7 2 4 6 3 6 4 4 3 2 5	5 5 0 3 5 3 5 3 8 3 2 3 8 2 3 1 6 2 7 2 4 6 3 6 4 4 3 2 5
4 4 0 2 1 7 2 4 4 7 4 1 9 2 4 5 2 5 0 4 0 0 5 3 6 3 3 6 7	4 4 0 2 1 7 2 4 4 7 4 1 9 2 4 5 2 5 0 4 0 0 5 3 6 3 3 6 7
7 4 6 6 8 7 5 7 9 2 0 2 8 8 8 8 3 2 4 2 6 4 0 4 6 3 7 2 1	7 4 6 6 8 7 5 7 9 2 0 2 8 8 8 8 3 2 4 2 6 4 0 4 6 3 7 2 1
0 1 7 1 5 9 1 4 2 8 7 3 7 1 4 5 1 8 7 8 0 5 1 7 0 5 8 8 1	0 1 7 1 5 9 1 4 2 8 7 3 7 1 4 5 1 8 7 8 0 5 1 7 0 5 8 8 1
2 8 5 2 1 2 8 7 7 6 2 5 6 2 6 4 1 5 1 6 1 2 1 1 0 5 6 4 0	2 8 5 2 1 2 8 7 7 6 2 5 6 2 6 4 1 5 1 6 1 2 1 1 0 5 6 4 0
2 1 1 7 7 2 0 0 1 8 7 0 2 9 0 2 8 5 7 8 4 6 0 6 5 0 7 1 2	2 1 1 7 7 2 0 0 1 8 7 0 2 9 0 2 8 5 7 8 4 6 0 6 5 0 7 1 2
0 5 2 4 1 5 3 3 1 6 5 1 4 0 1 6 4 3 3 9 8 8 3 4 6 8 4 8 6	0 5 2 4 1 5 3 3 1 6 5 1 4 0 1 6 4 3 3 9 8 8 3 4 6 8 4 8 6
7 3 7 5 2 4 0 2 7 6 3 8 5 5 4 5 8 8 7 5 5 6 5 6 7 9 7 7 4	7 3 7 5 2 4 0 2 7 6 3 8 5 5 4 5 8 8 7 5 5 6 5 6 7 9 7 7 4
0 3 2 8 1 4 4 6 0 8 2 3 0 1 3 4 6 2 0 5 7 7 3 6 1 8 7 3 5	0 3 2 8 1 4 4 6 0 8 2 3 0 1 3 4 6 2 0 5 7 7 3 6 1 8 7 3 5
4 4 8 3 3 3 5 0 1 0 3 8 6 3 2 0 5 0 6 1 3 3 4 3 6 1 5 8 6	4 4 8 3 3 3 5 0 1 0 3 8 6 3 2 0 5 0 6 1 3 3 4 3 6 1 5 8 6
1 0 2 2 7 6 3 3 0 8 8 0 3 1 8 8 1 2 1 7 5 2 9 3 5 8 3 2 5	1 0 2 2 7 6 3 3 0 8 8 0 3 1 8 8 1 2 1 7 5 2 9 3 5 8 3 2 5

팝아웃은 시각적 유도의 유형(초보적) 중 하나이자, 인간의 비주얼리제이션 시스템의 힘을 활용하는 시작일 뿐입니다. 작가 Colin Ware는 이러한 주제를 다루고 지각적 처리에 대한 기본 3단계를 알아보는 일련의 책을 출간했습니다. 첫 번째 단계는 시각적인 특징을 찾는 저수준 단계입니다. 이러한 프로세스는 눈에서 뇌의 시각 피질로 다시 이동합니다. 두 번째 단계는 패턴을 인식하며 세 번째 주목 단계에서 실제적인 계산이 수행됩니다.

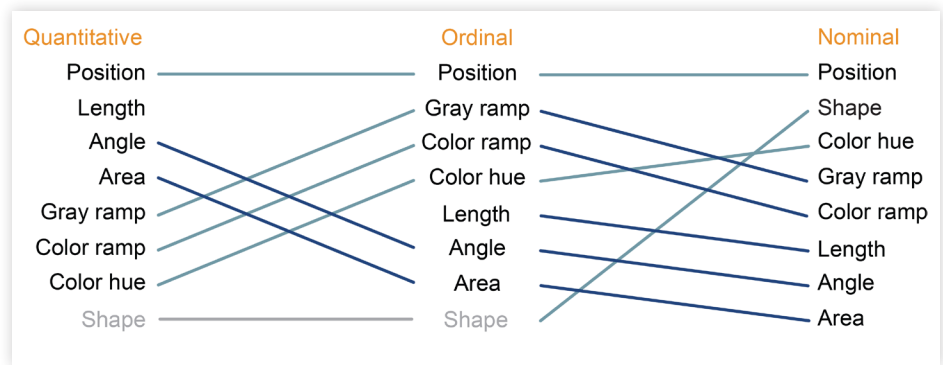
AT&T Bell Laboratories의 통계 과학자 William S. Cleveland와 Robert McGill이 수행한 연구는 Ware의 분석을 지원합니다. 이 연구에서는 정량적 가치를 평가하는 데에서는 위치(Position)가 가장 효율적인 방법이라는 것이 밝혀졌습니다. 위치(Position) 다음으로는 길이(Length), 영역(Area), 부피(Volume), 각도(Angle) 및 기울기(Slope), 색상(Color) 및 밀도(Density) 등의 순서로 정량적 가치 평가에 대해 효율적인 것으로 나타났습니다. Cleveland와 McGill은 위치(Position)를 시작점으로 하여 뒤쪽으로 갈수록 인간의 정량적 인지 정확성이 감소함을 발견했습니다. 이러한 방식에 따라, 스프레드시트에서 막대 그래프 뷰로 데이터 집합을 이동하고, 그러한 데이터를 여러 라인 길이 및 색상으로 변환함으로써, 더 빠르고 효율적인 분석을 수행할 수 있습니다. 물론, 이러한 모든 것은 진행 중인 작업 및 궁극적으로 알아내길 원하는 데이터가 무엇인지에 따라 다를 수 있습니다.

이 시점에서 독해에는 3가지 다른 수준이 있다는 Bertin의 설명을 다시 살펴볼 수도 하겠습니까. 기본 수준은 단일 값을 식별할 때 사용됩니다. 이 경우에는 단순한 스프레드시트가 적합합니다. 그러나 값 사이의 관계를 살펴보는 경우에는 중간 수준으로 이동해야 합니다. 이러한 뷰의 예는 막대 차트입니다. 막대는 모두 정렬되므로 막대 사이의 차이를 쉽게 파악하는 것이 가능합니다. 마지막으로, 글로벌(또는 계슈탈트) 뷰는 전체의 관계를 살펴봅니다. 이 방법의 경우 분산형

차트가 적합합니다. 이러한 유형의 차트에는 점 컬렉션이 포함되며 점의 위치는 수직축 및 수평축에 표시됩니다. 이러한 뷰는 매우 효율적이고 '전역적인' 데이터 뷰를 제공합니다.

## 데이터 유형에 따라 다른 뷰 사용

우리는 데이터 인코딩의 효율성은 데이터 유형에 따라 달라진다는 것을 알게 되었습니다. 데이터에는 3가지 유형(명목, 순서 및 정량적)이 있습니다. 명목(Nominal) 데이터 유형의 예는 서로 다른 종류의 새에 이름을 부여하는 것과 같은 명명법입니다. 순서(Ordinal) 데이터 유형은 요일 등과 같은 순서를 따릅니다. 정량적(Quantitative) 데이터는 숫자 및 측정이 가능한 항목을 다룹니다. 정량적 데이터의 예로는 길이, 시간, 온도 및 속도 등이 있습니다.



이러한 데이터 유형을 배치하는 그래픽 옵션을 살펴봄으로써 여러가지를 파악할 수 있습니다. 예를 들어 어떻게 일부의 경우 영역(Area)이 효과적이고 다른 나머지는 효과적이지 않은지 확인할 수 있습니다. 영역 뷰에 명목 데이터를 배치하면 적용할 수 없는 순서를 나타내게 될 수도 있습니다. 영역은 순서 데이터에 약간 더 효과적일 수 있지만 이러한 값들은 구분이 가능하므로 다른 크기를 구분하는 일부 영역만 가질 수 있습니다. 영역은 순서에 어떤 정량적 값이 있는지를 정확하게 제공한다는 점에서 정량적 데이터에 가장 적합합니다.

그래픽 표현을 색상(Color)으로 변경할 경우 이러한 발견 사항이 반대가 됩니다. 색상으로는 정량적 데이터 집합 전체를 정확하게 표현하기가 매우 어렵기 때문에 색상은 정량적 데이터에서 가장 효과가 낮습니다. 색상은 각 데이터 요소가 구별되는 고유 색상으로 할당될 수 있다는 점에서 명목 데이터에 가장 적합합니다.

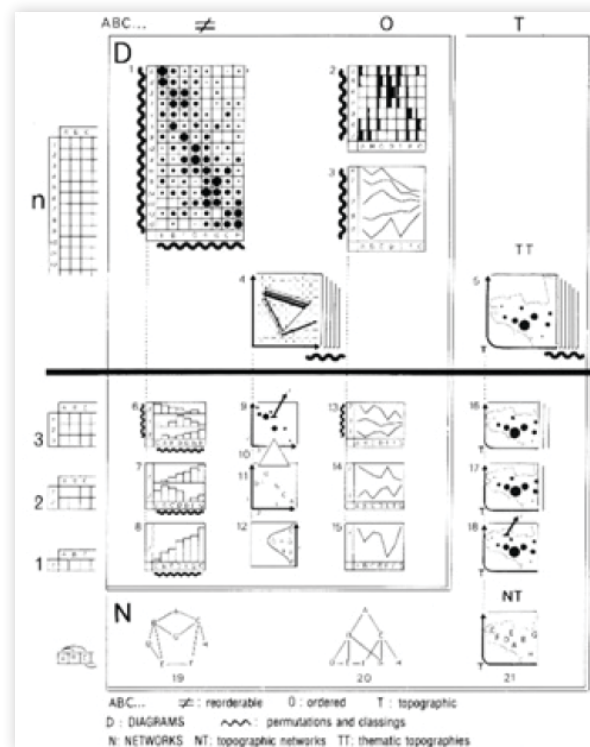
위의 그림은 효과적인 비주얼리제이션을 디자인하기 위해 어떻게 Tableau가 이러한 순위를 활용하는지 보여줍니다. 또다시, 이러한 3가지 유형의 데이터 모두를 표현하는 가장 효과적인 방법으로 위치(Position)가 최상위를 차지했습니다. 그러나 차트를 살펴보면 데이터의 유형에 따라 효과적인 방법에 차이가 있음을 이해할 수 있습니다. 데이터 유형에 따라 일부 뷰는 위로 이동하고, 일부 뷰는 아래로 이동합니다. 모양(Shape)은 정량적 또는 순서 데이터를 배치하는 데에는 적용할 수 없지만 명목 데이터에서는 상위에 위치해 있다는 점이 흥미롭습니다.

## 인간의 인지: 능력 및 한계

인간의 인지 능력은 강력하며 대량의 복잡한 데이터 집합을 처리할 수 있는 반면 특정 한계에 의해 제약됩니다. 여기서 우리는 다양한 유형의 데이터 뷰를 요약하는 Bertin의 기호학으로 돌아가보도록 하겠습니다. 한 가지 주요 요인은 포함된 차원 수이며, 1차원, 2차원, 3차원 또는 "n"차원이 될 수 있습니다. 두 번째 요인은 데이터의 고정 여부 또는 순서 재정렬 가능 여부에 따라 다릅니다. 마지막 구성 요소는 데이터가 토폴로지 형식인지의 여부입니다. Bertin은 이러한 요인을 통해 가능한 모든 다른 유형의 뷰를 보여주었습니다. 현재 Tableau는 이러한 각 뷰를 지원할 뿐만 아니라 최근에는 맵도 도입했습니다. 현재 노드-링크 뷰는 표시할 수 없지만 가까운 미래에 제공될 예정입니다.

아래 그래픽은 이러한 뷰를 보여줍니다. 주황색 선 아래의 뷰는 전체 또는 계슈탈트 형식으로 볼 수 있다는 점에 유의하십시오. 주황색 선 위의 뷰는 즉각적인 인지 즉 계슈탈트 형식으로 볼 수 없습니다. 이는 인간의 인지 및 데이터 분석에는 근본적인 한계가 있음을 나타냅니다. 인간이 데이터에 인코딩하고 짧은 순간에 파악할 수 있는 차원의 수에는 한계가 있습니다.

3D 그래픽은 Bertin의 장애 요인을 해결할 수 없다는 점에 유의해야 합니다. 그러나 세 번째 차원은 또 다른 레이어를 추가하여 데이터를 인코딩하는 동시에 폐색(occlusions)도 추가됩니다. 이러한 폐색으로 인해 모든 데이터를 살펴보는 것이 어려워져 특정 그래픽 기술이 필요하게 됩니다. 3D 추가와 관련한 다른 문제는 방향 문제도 더해진다는 것입니다. 공간에서의 이동이 가능하다는 점으로 인해 방향을 잃을 수 있습니다.



### 구성 및 상호 작용 활용

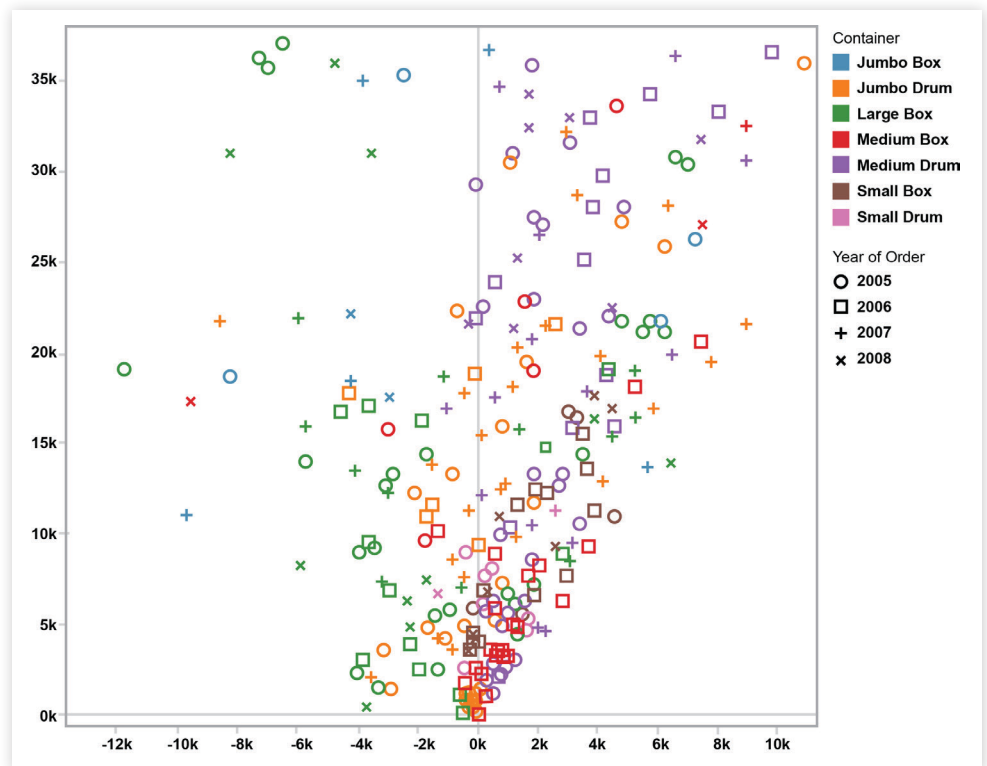
구성이 뛰어난 데이터 시각화를 개발하기 위한 핵심이라는 것은 분명한 사실입니다. 다차원 케이스에 대한 설명을 위해 나폴레옹의 모스크바 행진에 대한 Minard의 지도로 돌아가 보도록 하겠습니다. 이 지도는 군대의 행진과 모스크바로부터의 후퇴를 설명하는 상부의 두 라인과 온도를 보여주는 아래의 다른 라인 등 2개의 이미지로 구성됩니다. 이 지도는 효과적인 그래픽이지만 모든 데이터를 보고 이해하려면 시선을 앞뒤로 이동해야 하기 때문에 계슈탈트 뷰가 아닙니다.

효과적인 계슈탈트 뷰의 예는 Edward Tufte가 작성한 뷰(여러 개의 작은 뷰)입니다. 부동산 데이터를 보여주는 분산형 차트를 생각해 보도록 하겠습니다. 이 그래프는 계슈탈트 뷰와 같이 여러 이웃을 보여줄 수 있습니다. 정렬되어 있고 공통 축을 공유하므로 눈이 그래프에서 데이터가 가장 전달하고자 하는 요소를 빠르게 파악할 수 있습니다.

또한, 구성은 대시보드 유형의 뷰에서도 중요한 역할을 합니다. 여러 그래프를 결합하고 색상을 유의하여 사용함으로써 시선이 빠르게 이동하여 값을 비교하며 대량의 데이터를 이해할 수 있습니다.

Bertin은 상호 작용이 다차원 문제를 다루는 데에서 매우 효과적인 방법이라고 설명했습니다. 정렬 외에도 다차원 데이터에 대한 이해를 향상시키기 위한 여러가지 다른 기법이 있습니다.

아래와 같은 표준 분산형 차트를 생각해 보십시오. 이 차트는 매출과 수익으로 세분화되어 표시되며, 다시 컨테이너 및 연도별로 구분할 수 있습니다. 보시다시피, 너무 많은 데이터 집합을 겹쳐서 표시할 경우 발견 사항을 이해할 수 없으며 진정한 계슈탈트 뷰가 될 수 없습니다.



이러한 문제를 해결하기 위한 한 가지 방법은 집계를 사용하는 것입니다. 예를 들어, 단순히 데이터를 가져와 연도별로 구분하면 데이터의 양과 차트에 결과적으로 나타난 겹침이 크게 감소하게 됩니다. 이러한 방식으로, 집계는 다차원 데이터를 처리하는 데 강력한 도구가 될 수 있습니다.

다른 기법은 대화형 필터링을 사용하는 것입니다. 위의 차트에서, 관심이 있는 데이터 집합만 볼 수 있도록 해주는 필터 위젯을 생성하는 것을 상상해 보십시오. 이러한 위젯을 사용하면 필요에 따라 데이터를 더하거나 빼는 것이 가능해 전반적인 발견 사항을 명확하게 살펴볼 수 있게 됩니다. 필터링은 다차원 데이터를 다루는 데 매우 강력한 도구이지만 약점도 있습니다. 예를 들어, 특정 데이터 집합을 분리하면 데이터 집합 사이의 관계를 확인하는 것이 어렵거나 불가능해질 수 있습니다.

브러싱 프로세스를 활용하면 필터링과 관련된 제한 사항을 극복할 수 있습니다. 이 기법을 사용하여 하나 이상의 데이터 집합에 색상을 지정하면 데이터 집합 사이의 관계와 대조를 더 잘 살펴볼 수 있습니다.

상호 작용을 활용하는 다른 강력한 방법은 링크를 사용하는 것입니다. 예로 Google Maps를 살펴보겠습니다. 단순한 쿼리를 수행하면 영역이 강조 표시된 지도가 나타나고 이 영역을 클릭하면 자세한 데이터가 제공됩니다. 페이지에서 드롭다운 또는 위젯을 통해 데이터를 필터링하면 보다 정교한 뷰를 얻을 수 있습니다. 따라서 대화형 링크는 다차원 데이터 집합을 다루는 매우 강력한 방법입니다.

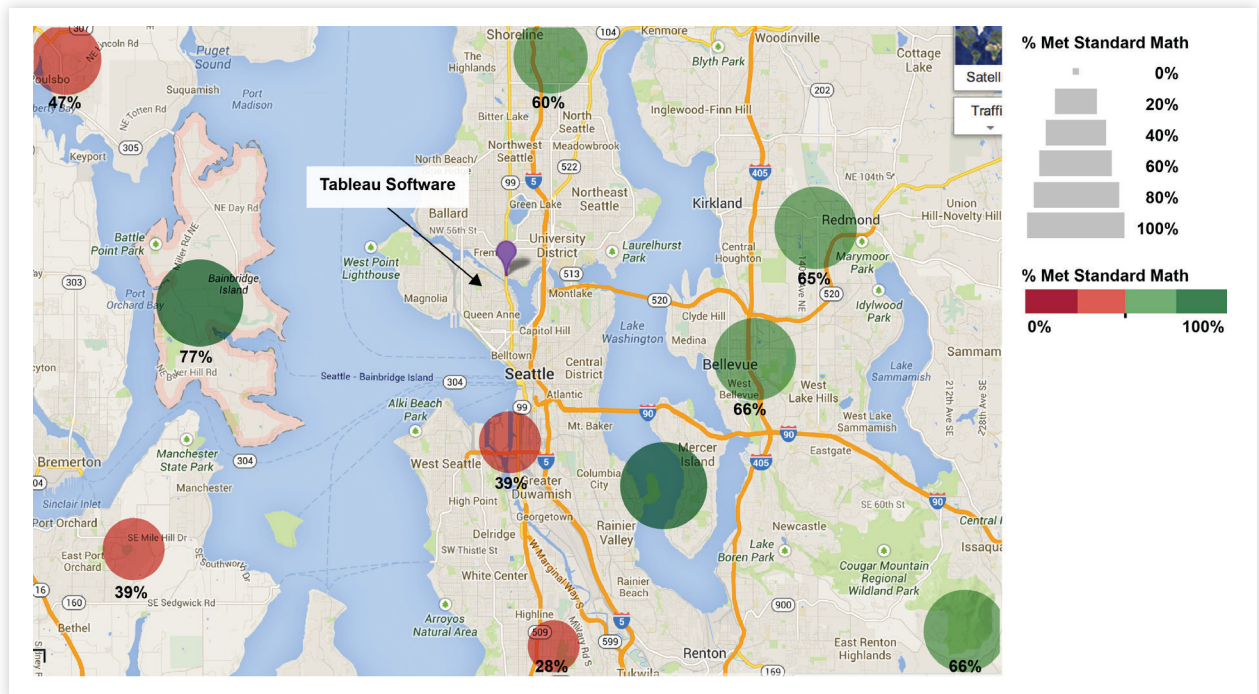
## 데이터를 활용한 스토리텔링

데이터를 활용한 스토리텔링에 대한 아이디어를 설명하기 위해 2004년 Xerox의 팔로알토 사무소에서 시애틀의 Tableau로 이직한 제 스토리를 들려 드리겠습니다. 저의 첫 번째 우선 순위 중 하나는 시애틀에서 최상의 학군을 찾는 것이었습니다. 저는 Seattle Times 웹 사이트에서 정리된 온라인 가이드를 빠르게 확인했습니다. 이 가이드는 모든 지역 학군의 Washington Assessment of Student Learning(WASL) 점수 및 각 과목에서 점수 표준을 충족한 학생의 비율을 제공합니다. 이를 통해 점수를 알게 되었지만 지리적 학군은 볼 수 없었습니다. 시애틀은 처음이라 지역을 알지 못했기 때문에 지도 즉 통계 지도가 필요했습니다.

하지만 지도가 제공되지 않아 아래와 같은 지도를 하나 작성했습니다. 이 지도는 외곽의 학군과 관련하여 Tableau의 위치를 보여줍니다. 색상이 지정된 원은 수학 점수의 측면에서 각 학군의 WASL 성취도를 나타냅니다. 그러므로 예를 들어 시애틀 학군은 사무실에서 가장 가깝지만 WASL 성취도가 낮았습니다. 이 지도를 보고 저는 Tableau에서 그래도 가깝고 WASL 성취도도 높은 Bellevue 학군으로 이사하기로 결정했습니다.

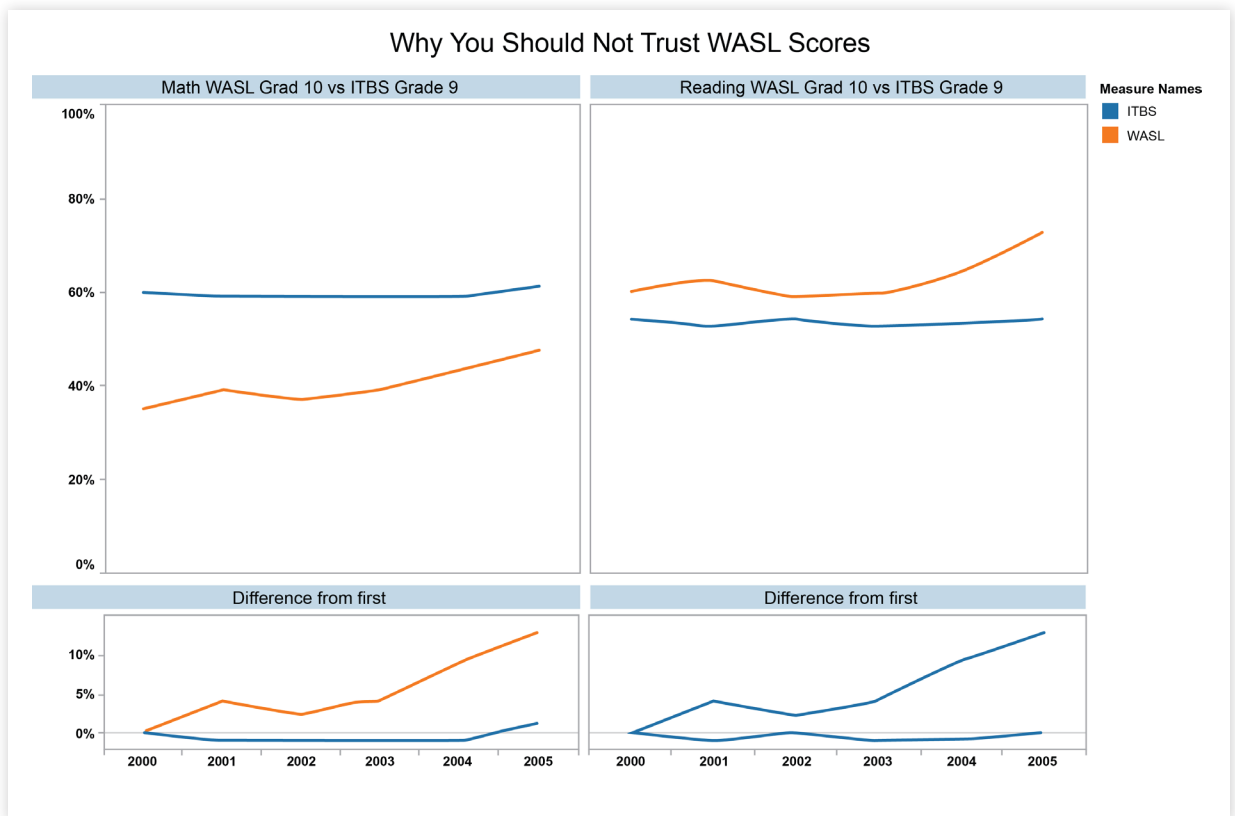
얼마 후 저는 학교의 수학 점수에서 불규칙성을 발견했습니다. 잠시 조사한 후 WASL의 수학 및 읽기 영역 모두에 대한 그래프를 보여주는 자세한 웹 사이트를 발견했습니다. 두 가지 뷰 모두 시간 경과에 따른 데이터 추세를 보여주었습니다. 이들 추세는 모두 증가 추세였습니다. 그리고 Bellevue 학군이 수학 및 읽기 영역 모두에서 주 평균을 능가하는 것을 보여주었습니다. 수학과 읽기 점수가 서로 어떤 관련성이 있는지에 대해서는 명확하지 않았습니다. 좀 더 조사한 후 저는 실제 데이터 원본을 발견했습니다.

이러한 데이터를 정리한 후 Tableau에 입력했습니다. 그 결과 수학 점수가 읽기 점수보다 훨씬 낫다는 보다 흥미로운 점이 발견되었습니다. 저는 수학에 관심이 있고 배경 지식도 가지고 있기 때문에 데이터를 계속해서 모니터링한 후 아래와 같은 차트를 얻게 되었습니다.



발견 사항은 지난 5년 동안 워싱턴 주는 WASL 및 ITBS(Iowa Testing Program)의 두 가지 다른 시험을 활용했다는 것입니다. 아래 차트는 두 시험을 비교하여 놀라운 결과를 보여주었습니다. WASL 결과는 시간이 지남에 따라 상승하지만 ITBS 결과는 거의 변동이 없다는 것입니다. ITBS 점수는 미국 전체 학생의 결과를 나타내므로 저는 WASL 점수를 신뢰할 수 없다는 결론을 내렸습니다.

이 스토리의 교훈은 “항상 데이터에 대해 질문하라”는 것입니다. 데이터 출처가 어디인지에 관해 질문하고, 시각적 시스템을 활용하여 데이터에서 얻어지는 정보를 신뢰하고 이해할 수 있는지 확인해야 합니다.



## 점검하기: 배운 교훈

효과적인 스토리 전달에 도움을 주는 몇 가지 간단한 교훈에 대해 점검해 보도록 하겠습니다.

### 교훈 1:

신뢰가 디자인의 핵심적인 사안입니다. 뷰를 적절하게 디자인하지 못한다면 해당 뷰를 보는 숙달된 사용자로부터 신뢰를 얻을 수 없습니다.

### 교훈 2:

비주얼리제이션이 표현적이며 데이터를 정확하게 전달하는지 확인하십시오.

### 교훈 3:

인간의 인지 능력을 이용하여 시각화를 효율적으로 활용하십시오. 다음 규칙은 시각화를 효율적으로 활용하는 데 도움이 됩니다.

- 적절한 그래픽 언어 활용.
- 공백 활용.
- 불필요한 소재 및 군더더기 제거.

### 교훈 4:

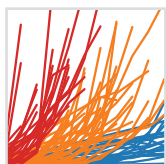
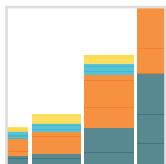
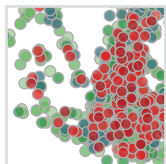
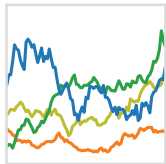
뷰에 컨텍스트를 포함하십시오. 제목, 캡션, 단위, 설명과 같은 항목들은 대상 사용자가 데이터 뷰를 더 잘 이해하는 데 도움이 됩니다. 항상 데이터 및 시각화를 활용하여 스토리를 전달하도록 노력하십시오. 좋은 스토리는 단순히 데이터를 포함하는 것이 아니라 그 이상의 것이 필요하다는 이해를 바탕으로 다음을 고려하십시오.

- 미적인 부분에 유의할 뿐 아니라 종종 감성적인 것들이 효과적이라는 것을 이해하십시오. 즉, 효과적인 뷰는 대상 사용자들로부터 감정적 반응을 이끌어내고 그들과 진솔한 소통을 할 수 있게 해줍니다.
- 스타일 또한 중요합니다. 뷰가 일관적이고 보기에 좋은지 확인하십시오. 뷰는 본인 및 본인의 관심사를 대표합니다.
- 뷰는 즐거움을 줄 수 있습니다. 사람들이 즐길 수 있는 대화형 뷰는 참여도가 매우 높습니다. 대화형 요소를 사용하면 사용자는 스스로 데이터를 조작하고, 질문을 묻고 답하여 발견에 이를 수 있습니다. 이는 데이터의 신뢰성을 높이는 데 도움이 됩니다.
- 생생하고 기억에 남는 뷰를 작성하십시오. 구조와 컨텍스트에 유의하십시오.



## Tableau 정보

Tableau Software를 통해 사람들은 데이터를 보고 이해할 수 있습니다. Tableau는 사용자가 정보를 신속하게 분석하고 시각화하고 공유할 수 있도록 도와드립니다. 1만 5천개 이상의 고객 계정에서 사무실 및 이동 중에 Tableau를 통해 빠른 결과를 받아보고 있습니다. 또한, 수만 명의 사람들이 Tableau Public을 이용해 블로그와 웹 사이트에서 데이터를 공유하고 있습니다. [www.tableausoftware.com/ko-kr/trial](http://www.tableausoftware.com/ko-kr/trial)에서 무료 평가판을 다운로드하여 Tableau를 통해 얻을 수 있는 이점이 무엇인지 확인하십시오.



### 추가 리소스

무료 평가판 다운로드

### 추가 리소스 탐색

- 제품 데모
- 교육 및 자습서
- 커뮤니티 및 지원
- 고객 스토리
- 솔루션

### 참고 자료

- [1] <http://www.henry-davis.com/MAPS/AncientWebPages/100B.html>
- [2] <http://www.math.yorku.ca/SCS/Gallery/milestone/sec2.html>
- [3] <http://www.henry-davis.com/MAPS/AncientWebPages/101.html>
- [4] <http://www.henry-davis.com/MAPS/Ren/Ren1/406B.htm>
- [5] <http://www.math.yorku.ca/SCS/Gallery/images/playfair2.jpg>
- [6] <http://www.math.yorku.ca/SCS/Gallery/images/minard.gif>
- [7] Darrell Huff. How to Lie With Statistics. Norton, New York. 1955.
- [8] Jacques Bertin. Semiology of Graphics. University of Wisconsin Press. 1983 (Semiologie Graphique Gauthier-Villars, Paris 1967).
- [9] Stuart K. Card, Jock D. Mackinlay, & Ben Shneiderman. Readings in Information Visualization: Using Vision to Think. Morgan Kaufman, San Francisco. 1999.