

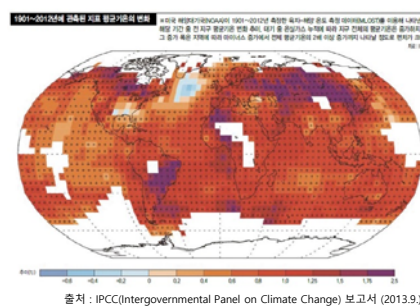
# 융합의 시대를 위한 준비

## INFO-Graphics



### 200년 동안 얼마나 풍족해졌나?

- ✓ 한스 로슬링 박사
- ✓ 동그라미 크기 : 각 국가의 인구
- ✓ 수입 ↑ → 평균수명 ↑



### 지구 온난화

- ✓ 기온측정방법이 제각각
- ✓ 기록이 없는 지역이나 연대가 존재함.
- ✓ IPCC : 100년(1906~2005년)동안 0.74°C 상승 보고

## Data Mining



- ✓ 판매된 물품을 바탕으로 다음에 구입하려는 물품 예측 → 쿠폰 발송



- ✓ 기후분석 시스템 도입 → 허리케인 직전에 팔 수 있는 과자

“가게 주인은 데이터에 감추어진 히트 상품을 찾는다”

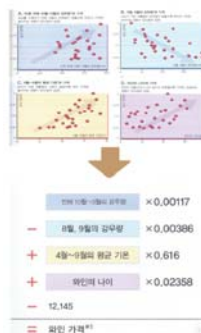
	소스	코코아	레몬	후식	커피	향신료
10대 여성	1					
20대 남성				1	1	
30대 남성	1	1				
40대 남성						
50대 여성	1		1	1		
60대 여성						
70대 여성						
80대 여성	1		1	1	1	1
90대 여성	1					
10대 남성						
합계	15	2	1	2	3	2

3개 이상 판매된 것을 골라 정리

	소스	코코아	레몬	후식	커피	향신료
10대 여성	1					
20대 남성				1	1	
30대 남성						
40대 남성						
50대 여성	1		1	1		
60대 여성						
70대 여성						
80대 여성	1		1	1	1	1
90대 여성	1					
10대 남성						
합계	15	2	1	2	3	2

어느 물품을 구입한 고객이 다른 물품도 구입할 확률은 몇 %인지 예측

## 미래에 대한 예측



### 10년 뒤 얼마짜리 와인이 될까?

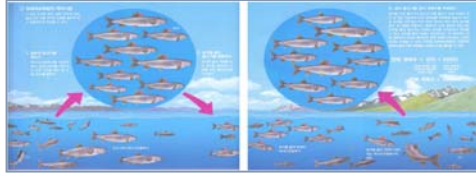
- ✓ 올리 아센펠터 (Orley Ashenfelter)
- ✓ 경제학자, 와인 애호가
- ✓ 통계를 이용해 와인이 장래에 높은 가치를 가질지 여부 예측



### 연간 보험료를 얼마로 책정해야 할까?

- ✓ 에드먼드 핼리 (1656~1742, 영국, 천문학자)
- ✓ 커다란 집단으로 보면, 사람의 사망률은 나이마다 일정하다. → 생명표 작성

## 모집단의 추정



### 미국 Yellowstone Lake의 외래어종 수는?

- ✓ 포획-재포획법
- ✓ 2013년 8월 시작
- ✓ 현재 3000마리의 외래어종에 표지 부착

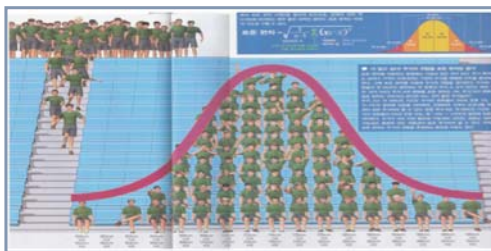


### 지도도 예측

- ✓ 찌개의 간 → 한 숟가락이면 충분
- ✓ 1억 명의 생각을 단 1000명으로 추측
- ✓ 모집단을 잘 닮은 표본으로 전 국민의 의견 추측 가능

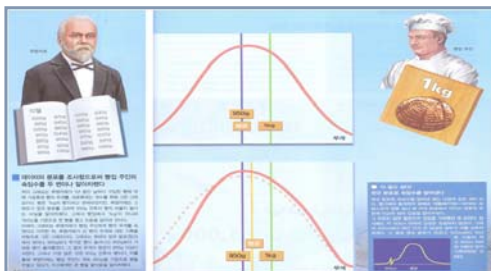
Random → 모든 사람이 선정될 확률을 동일하게 가지도록 !!

## 정규분포의 활용



### Normal Distribution

- ✓ 좌우대칭의 산 모양
- ✓ 평균과 표준편차로 데이터의 전모를 추측



### 빵집의 거짓말

- ✓ 무게를 속이고 있지 않을까?
- ✓ 앙리 푸앵카레 (1854~1912, 프랑스, 수학자)
- ✓ 날마다 빵의 무게를 조사
- ✓ 1kg이 아니라 950g 기준으로 팔고 있음을 밝힘.

## 설문조사의 신뢰성 (불편한 질문에 대한 응답)



### 미성년 음주에 대한 설문 응답

- ✓ 누가 음주 경험이 있는지는 비밀로 하고, 비율만 조사하는 것이 목적
- ✓ 동전 앞면 : 예 (Yes)
- ✓ 동전 뒷면 : 예 (Yes)/아니오 (No)

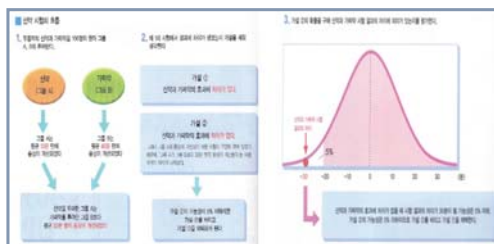


### 오바마 후보에 대한 기부금

- ✓ 댄 시로커 (Dan Siroker)
- ✓ 24가지 웹사이트 디자인 제작
- ✓ 웹사이트 방문자에게 무작위 노출
- ✓ 일정기간 이후 방문자 이메일 등록 비율 조사

➔ 기부금 약 \$60M 증가, 자원봉사자 28만명 증가

## 비교 분석



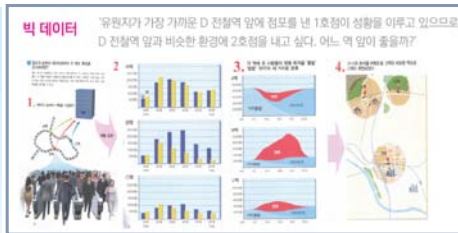
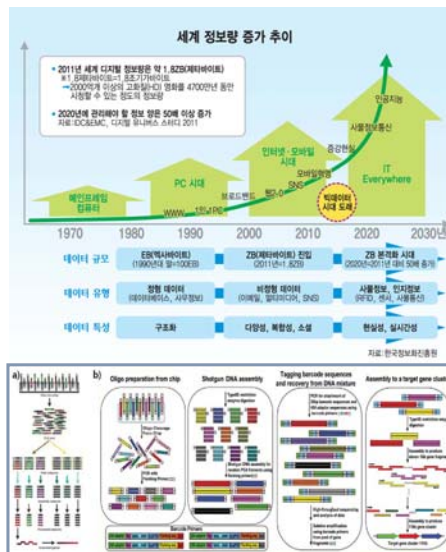
### 새로 개발된 신약의 효과

- ✓ 위약 (placebo)과의 차이 검증
- ✓ 기존 약과의 효과 비교



### SNS를 통한 후보간 강/약점 비교

## Big Data



### 어디에 상점을 내면 좋을까?

- ✓ 히타치 (Hitachi, 일본)
- ✓ 일본 전철 승차카드 4000만장 분석
- ✓ 일본 수도권 약 1800개 역에 대한 특성 파악

### Bio Big Data

- ✓ Next generation sequencing
- ✓ EMR (Electronic Medical Records)

## Big Data Scientist의 핵심 역량

