

Tableau를 통한 데이터 시각화와 분석

2020/10/30

서울대학교 사회과학대학
컴퓨터이셔널 방법론 부트캠프

소개

서울대학교 언론정보학과 박사과정, 이규호(hci+d lab.)

조사방법론(머신/딥러닝), 사회연결망(그래프이론), 데이터 저널리즘

“사회 현상(미디어)을 컴퓨터에 기반한 연산으로 풀어내는 연구”

(e.g. 이슈에 따른 뉴스 기사의 중요 단어 연결망 분석,

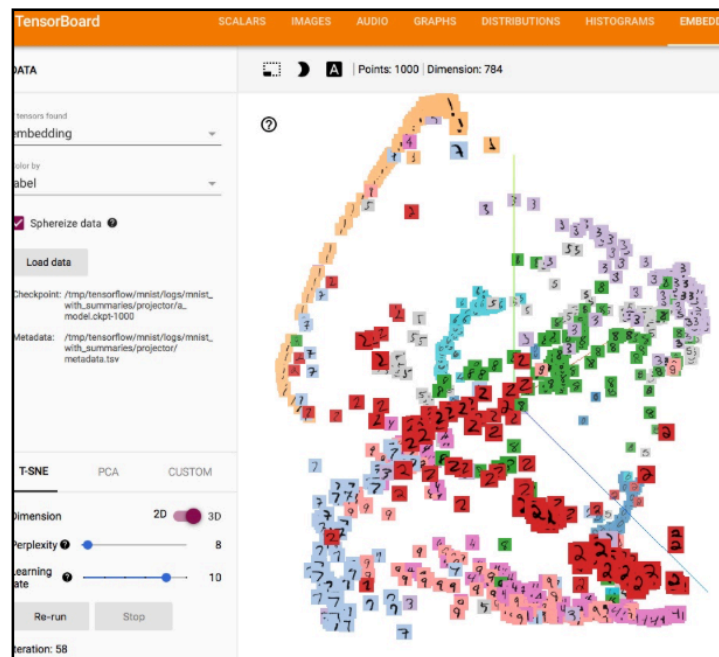
수치 분석을 통한 커뮤니티 성장 가능성 측정,

머신러닝을 통한 학생 성취 예측/Tableau 시각화 등...)

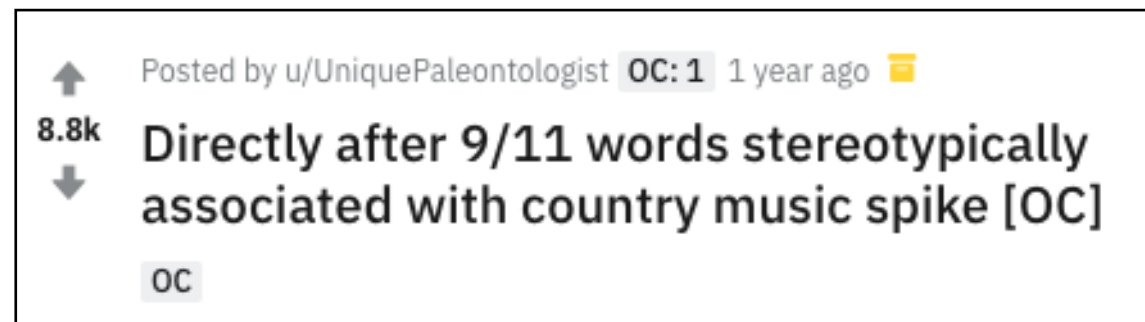
Python, R + Tableau + 기타 여러가지(흥미 있으면 다 씁니다!)

0. 데이터 분석과 tableau

데이터 과학(머신러닝)/저널리즘/사회과학(연구)?



<https://projector.tensorflow.org>



https://www.reddit.com/r/dataisbeautiful/comments/ci2wws/directly_after_911_words_stereotypically/

분야는 다르지만 둘 다 시각화를 적절히 사용합니다

1) 데이터 해석 2) 중요한 메시지 강조

빅데이터의 발전과 시각화의 시대!

“오해에 사로잡힌 사람을 설득할 때는
그의 의견을 데이터와 비교하는 방법이 매우 유용하다”

– 팩트폴니스

(<https://youtu.be/Sm5xF-UYgdg>)

빅데이터/머신러닝을 통한 분석을 어떻게 보여줄 것인가?

Tableau?



Business Intelligence Software

비즈니스 용도로 데이터 분석을 수행하는 소프트웨어
시각화도 가능하지만, 실제로는 더 많은 기능이 존재

왜 Tableau인가?

```
In [9]: df2 = pd.DataFrame({'A': 1.,
...:                        'B': pd.Timestamp('20130102'),
...:                        'C': pd.Series(1, index=list(range(4)), dtype='float32'),
...:                        'D': np.array([3] * 4, dtype='int32'),
...:                        'E': pd.Categorical(["test", "train", "test", "train"]),
...:                        'F': 'foo'})

In [10]: df2
Out[10]:
```

	A	B	C	D	E	F
0	1.0	2013-01-02	1.0	3	test	foo
1	1.0	2013-01-02	1.0	3	train	foo
2	1.0	2013-01-02	1.0	3	test	foo
3	1.0	2013-01-02	1.0	3	train	foo

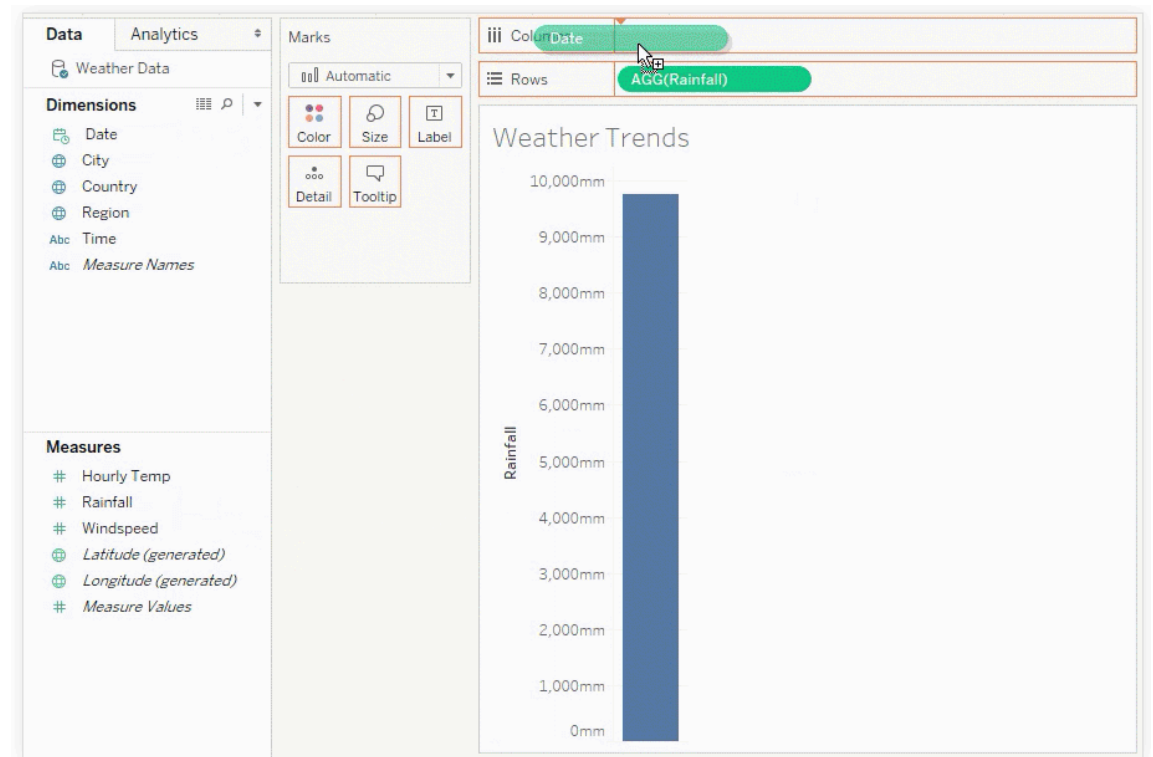
```
def create_app(test_config=None):
    # create and configure the app
    app = Flask(__name__, instance_relative_config=True)
    app.config.from_mapping(
        SECRET_KEY='dev',
        DATABASE=os.path.join(app.instance_path, 'flaskr.sqlite'),
    )

    if test_config is None:
        # load the instance config, if it exists, when not testing
        app.config.from_pyfile('config.py', silent=True)
    else:
        # load the test config if passed in
        app.config.from_mapping(test_config)

    # ensure the instance folder exists
    try:
        os.makedirs(app.instance_path)
    except OSError:
        pass

    # a simple page that says hello
    @app.route('/')
    def hello():
        return 'Hello, World!'

    return app
```



Python(Pandas+Matplotlib/Plotly)와 같은 대체 조합도 가능 하지만
공유까지 생각하면 기술적 요구 사항이 높음

하지만 Tableau는 강력한 인터랙티브 시각화를 손쉽게 구현하고 공유 가능
(코딩에 비해 난이도가 쉽고, 다른 BI에 비해 호환성이 좋은 편)

비교적 고가지만 Public 버전이나, “학생/교직원” 라이선스 활용 가능!

Tableau 학습 준비물

Tableau Desktop(체험판)
Tableau/Public 계정

Google 계정
Google Drive 활성화
(시간이 되면 켜주세요...)

Don't Panic!
(자신감!)

1. 데이터 기초 개념

데이터 종류 확인하기

attribute

item

	A	B	C	D	E	F	G	H
1	번호	발생시각	규모	위치	위도	경도	지역	시기
2	1	1978/08/30 02:29:43	4.5	북한 평안북도 철산 남남서쪽 84km 해역	39.1	124.2	북한	1970~1999
3	2	1978/09/16 02:07:06	5.2	경북 상주시 북서쪽 32km 지역	36.6	127.9	경북	1970~1999
4	3	1978/10/07 18:19:52	5	충남 홍성군 동쪽 3km 지역	36.6	126.7	충남	1970~1999
5	4	1978/11/23 11:06:05	4.6	북한 황해남도 안악 남동쪽 15km 지역	38.4	125.6	북한	1970~1999
6	5	1978/12/04 15:16:16	2.8	강원 정선군 북동쪽 4km 지역	37.4	128.7	강원	1970~1999
7	6	1978/12/12 21:58:38	3.3	전북 군산시 서남서쪽 38km 해역	35.9	126.3	전북	1970~1999
8	7	1979/01/01 00:11:22	2.9	충남 홍성군 동쪽 3km 지역	36.6	126.7	충남	1970~1999
9	8	1979/01/24 00:40:57	3	전북 변산반도 서쪽 약 40km 해역	35.7	126	전북	1970~1999
10	9	1979/01/29 10:06:02	3.3	북한 황해북도 사리원 남동쪽 31km 지역	38.3	126	북한	1970~1999
11	10	1979/01/30 00:33:12	3	경북 구미시 서남서쪽 5km 지역	36.1	128.3	경북	1970~1999
12	11	1979/02/04 09:29:18	3.7	북한 황해북도 사리원 남남서쪽 13km 지역	38.4	125.7	북한	1970~1999
13	12	1979/02/08 08:52:20	4	충남 홍성군 동쪽 3km 지역	36.6	126.7	충남	1970~1999
14	13	1979/02/24 14:44:07	2.7	북한 황해남도 장연 북쪽 6km 지역	38.3	125.1	북한	1970~1999
15	14	1979/02/24 19:00:30	2.9	충남 홍성군 동쪽 3km 지역	36.6	126.7	충남	1970~1999
16	15	1979/03/12 11:09:24	3.8	충남 홍성군 동쪽 3km 지역	36.6	126.7	충남	1970~1999
17	16	1979/04/09 08:49:33	3.3	북한 평안남도 안주 북북동쪽 10km 지역	39.7	125.7	북한	1970~1999
18	17	1979/04/24 05:44:27	3	경북 영주시 서쪽 11km 지역	36.8	128.5	경북	1970~1999
19	18	1979/08/13 12:12:02	3.5	경북 군위군 북서쪽 10km 지역	36.3	128.5	경북	1970~1999
20	19	1979/08/14 10:05:05	2.5	북한 평안남도 안주 남동쪽 17km 지역	39.5	125.8	북한	1970~1999

Tableau에서는 주로 **Tidy Data**를 사용합니다
(데이터 사이언스에서 보편적인 형식)

Tidy data?



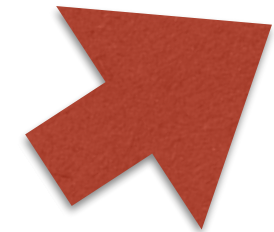
Hadley Wickham
(tidyverse)
(dplyr, reshape2, ggplot2...)

<https://commons.wikimedia.org/wiki/File:Hadley-wickham2016-02-04.jpg>

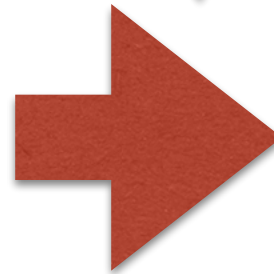
각 변수 = 열(column)
각 관측치 = 행(row)
행과 열은 하나의 기준(실험/관찰)의 결과값

vita.had.co.nz/papers/tidy-data.pdf

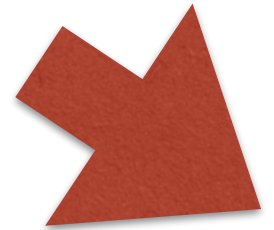
간단히 말해 봅시다



개



흰색



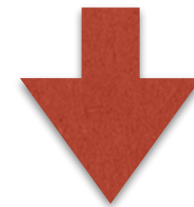
멋짐(?)

...

표로 본다면?

	종류	색	나이
동물1	개	검은색	3
동물2	고양이	흰색	2
동물3	라마	갈색	2
동물4 (...)	개	흰색	3

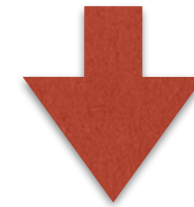
세로에는 서로 다른 특징이 들어갑니다



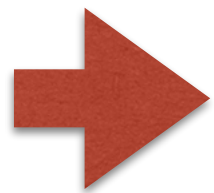
특징 1



특징 2



특징 3



가로에는 각자 다른 조사 대상이 들어갑니다
(e.g. 피험자1, 2, 3, 4... / 국가1,2,3,4...)

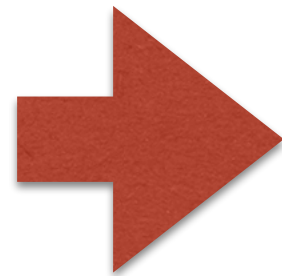
TABLEAU와 TIDY DATA의 관계는?

특징

인원

	특징		
인원			
		측정값	

특징을 섞고, 측정값을 넣으면
시각화가 됩니다 (IV/DV)



세로막대차트

특징 : 전공, 학년
표현값 : 성적

순위차트

특징 : 팀 종류/시간(년)
표현값 : 순위

TABLEAU에 적합하지 않은 데이터?

네트워크 데이터(가로/세로 축이 모두 변수로 사용)
년도별 데이터(년도별로 데이터가 늘어나 있는 형태일때)
텍스트 데이터(가로, 세로축이 아예 존재하지 않음)
...

기타 Tidy Data 형식이 아닌 것

하지만, 이런 데이터도 “전처리”를 잘 하면 쉽게 분석할 수 있습니다
(다음 시간에...)

1. 가볍게 시작하기

시각화의 2가지 방향

데이터 탐색
EDA, Exploratory
Data Analysis

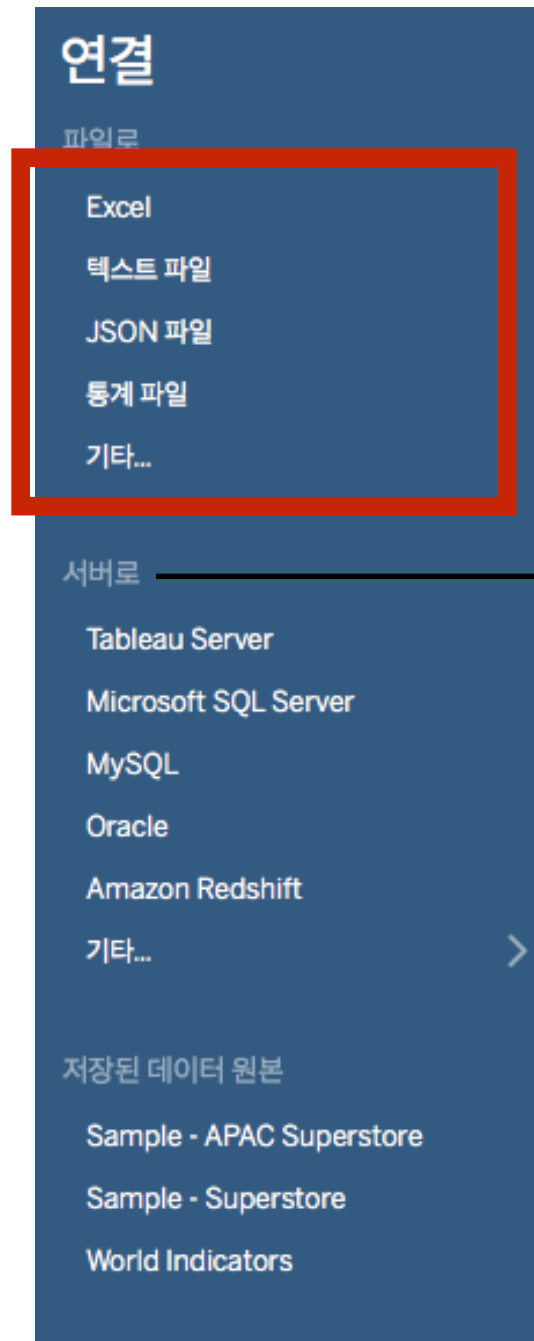


분석 결과 정리
Information/Scientific
Visualization

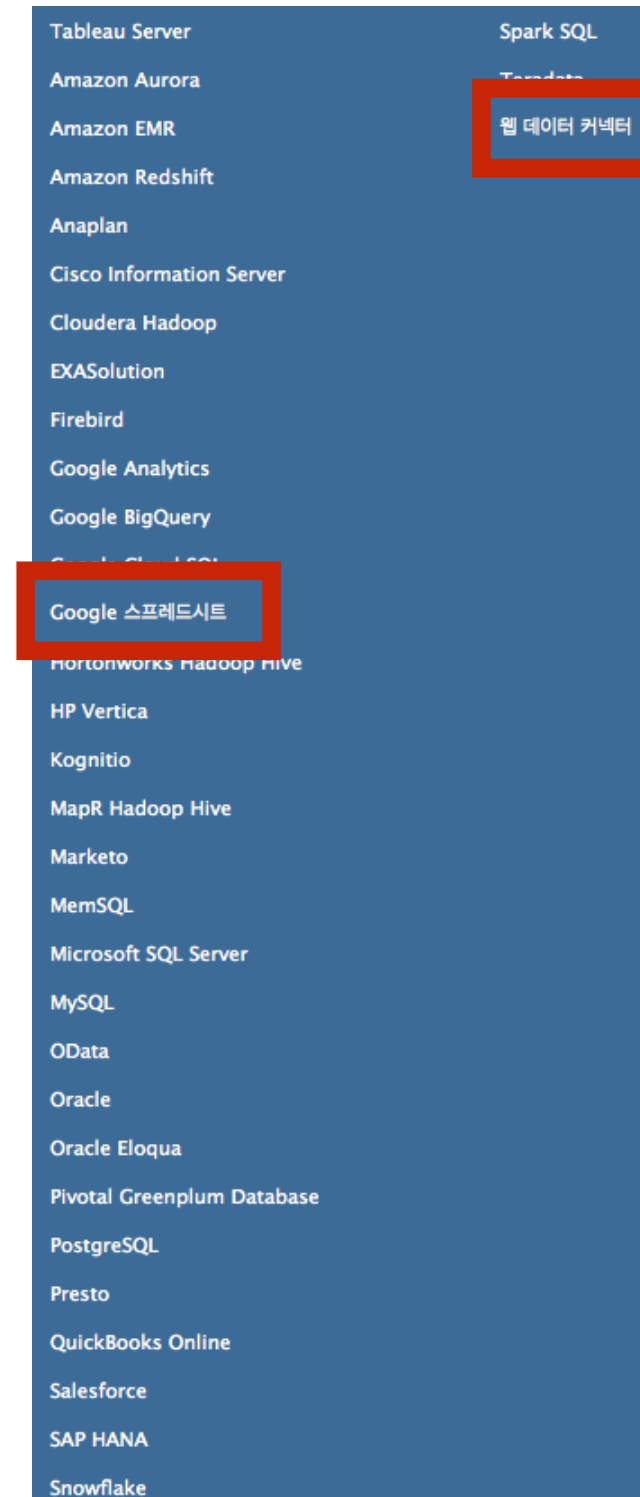
Data Format!

아마도 여러분이 생각할 수 있는 모든 데이터 포맷을
사용 가능합니다...
(CSV, TSV, EXCEL, SPSS, SQL...)

연결(2종류 존재)

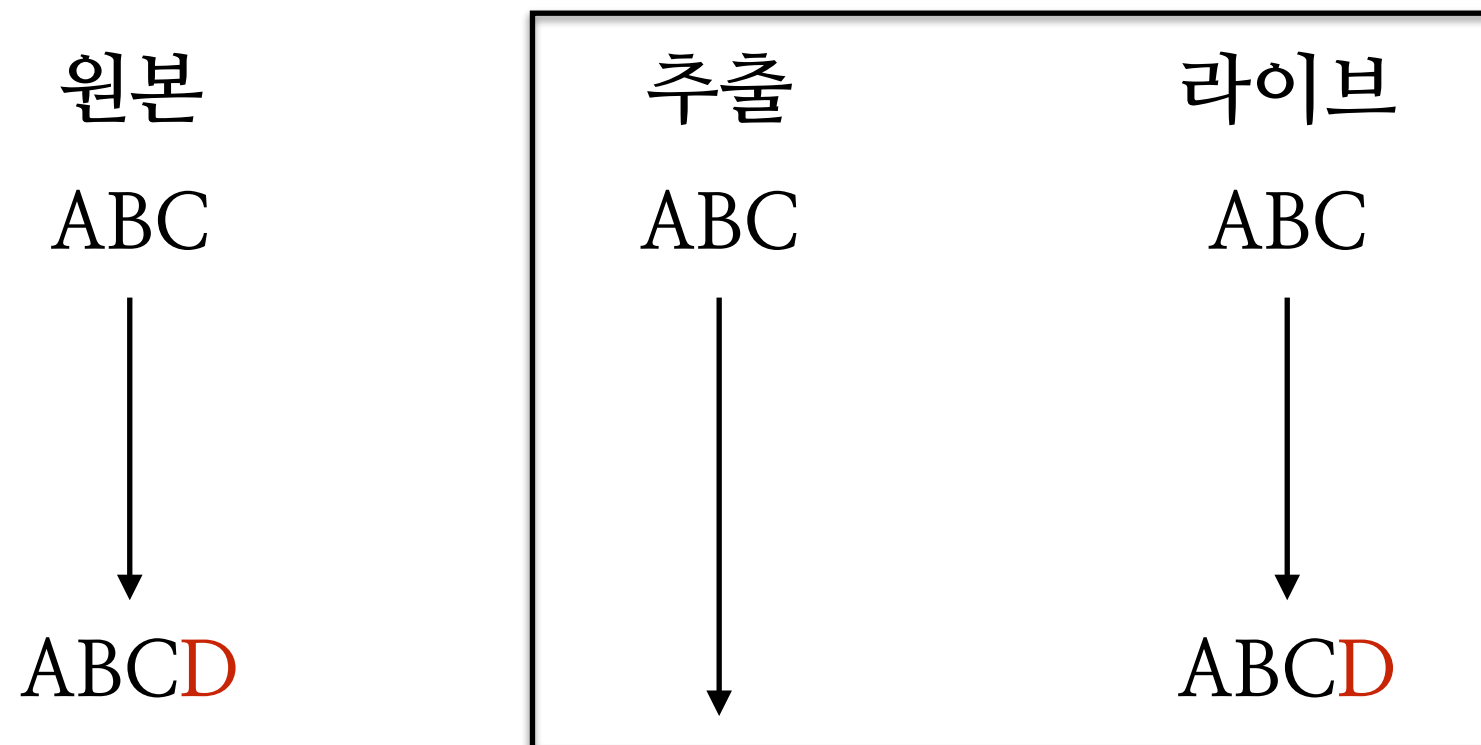


파일에서 가져오기



온라인에서 가져오기
(오늘은 X)

추출? 라이브?



과거 데이터, 혼자 작업 - 추출연결(한번 불러오면 바뀌지 않습니다)
협업, 실시간 데이터 - 라이브 연결(원 파일과 연동됩니다)

data!

data.seoul.go.kr
서울 열린 데이터 광장

서울시 지하철호선별 역별 승하차 인원 정보

Sheet Open API File					
서울시 지하철호선별 역별 승하차 인원 정보에 대한 파일 명세서를 제공합니다. 명세서를 다운로드 하세요.					
No	파일명	파일크기(KB)	마지막수정일	최초공개일	다운로드
1	CARD_SUBWAY_MONTH_201501.csv	1,002	2016.06.10	2016.06.10	Down
2	CARD_SUBWAY_MONTH_201502.csv	1,030	2016.06.10	2016.06.10	Down
3	CARD_SUBWAY_MONTH_201503.csv	1,018	2016.06.10	2016.06.10	Down
4	CARD_SUBWAY_MONTH_201504.csv	1,007	2016.06.10	2016.06.10	Down
5	CARD_SUBWAY_MONTH_201505.csv	1,043	2016.06.10	2016.06.10	Down

데이터 원본창

subway

subway.csv

필드 정렬

수정한 날짜

<div>subway.csv</div> <div>사용일자</div>	<div>subway.csv</div> <div>노선명</div>	<div>subway.csv</div> <div>역ID</div>	<div>subway.csv</div> <div>역명</div>	<div>subway.csv</div> <div>승차총승객수</div>	<div>subway.csv</div> <div>하차총승객수</div>
2017. 1. 1.	경원선	1907	가능	4,194	3,852
2017. 1. 1.	경원선	1908	녹양	2,612	2,468
2017. 1. 1.	경원선	1909	양주	5,036	5,004
2017. 1. 1.	경원선	1910	덕계	1,145	1,123
2017. 1. 1.	경원선	1911	덕정	3,823	4,188



데이터 조합/편집 가능
(오늘은 간단하게만...)

차트 작성하기(1)



요인 선택 후 오른쪽 위 “표현방식” 클릭
Tableau에서는 **자동으로 추천 그래프 작성**
(직접클릭/더블클릭)

차트 작성하기(2)

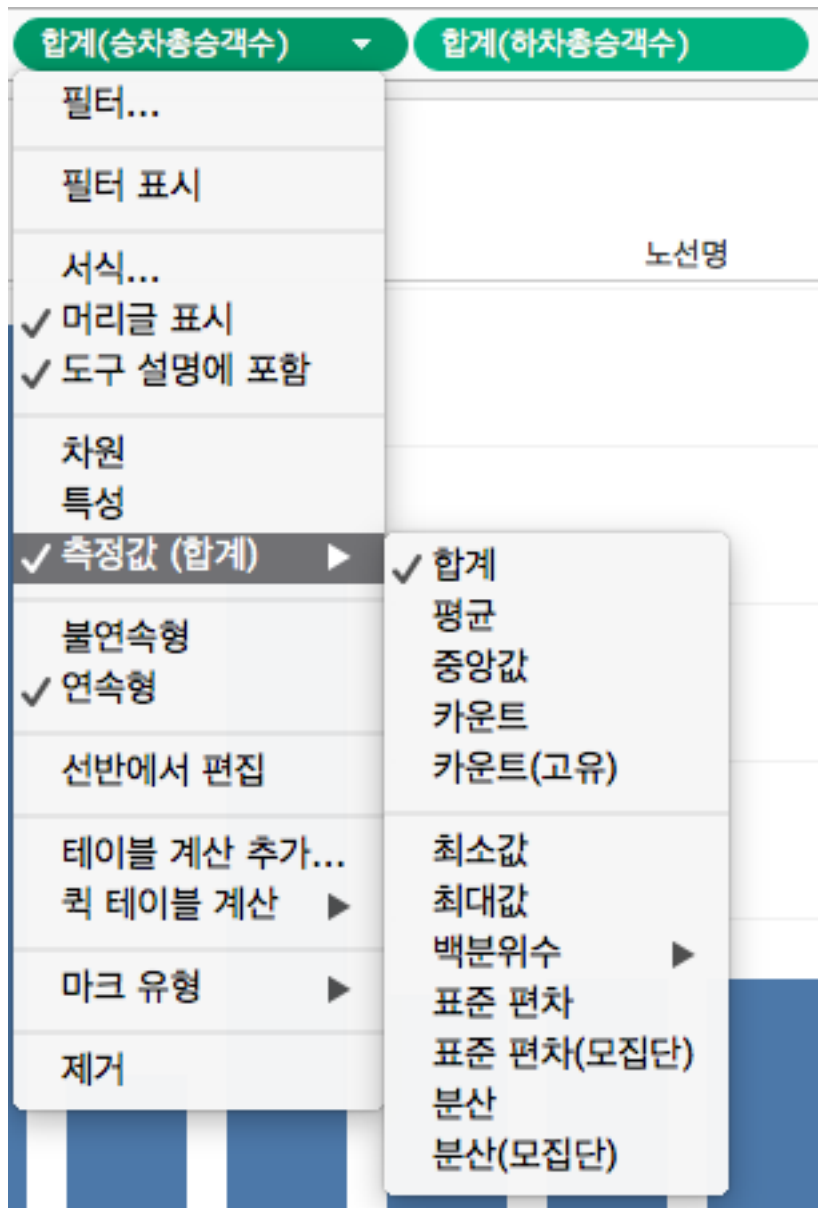


데이터 편집/속성변경후

조합

문자, 색의 차이는 (유형, IV/DV)?
서로 연결된 데이터 형식은? 그룹화/계층화

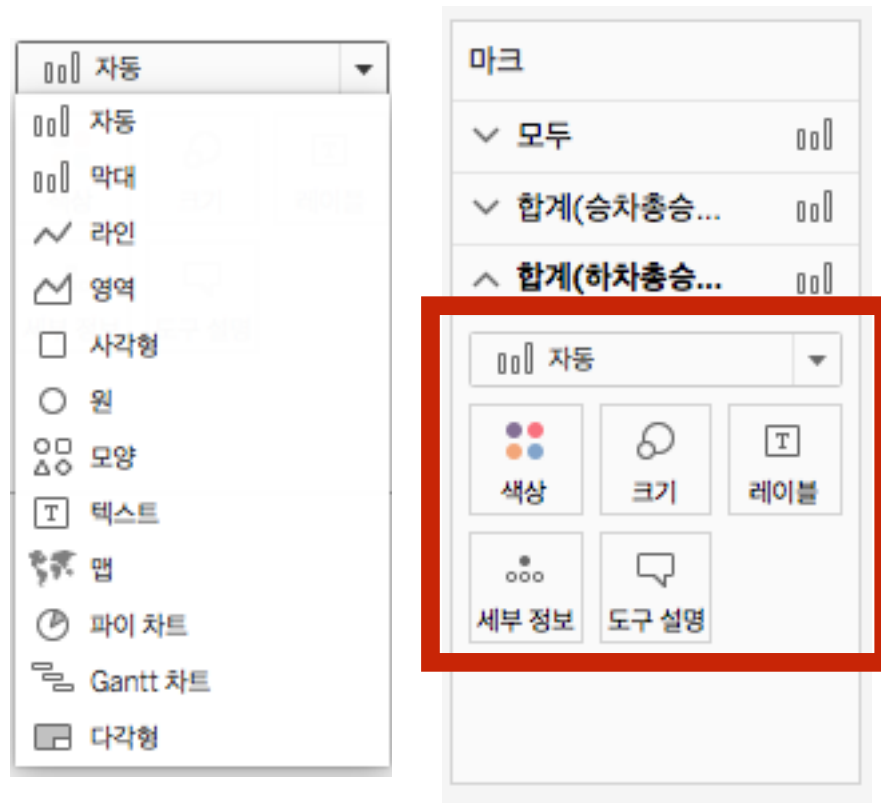
디테일 수정(측정값/표현방식)



내가 원하는 형태의 측정값을 사용 가능

비즈니스 분석에서 빈번하게 쓰는 기능은
(퀵 테이블 계산)에서 사용 가능

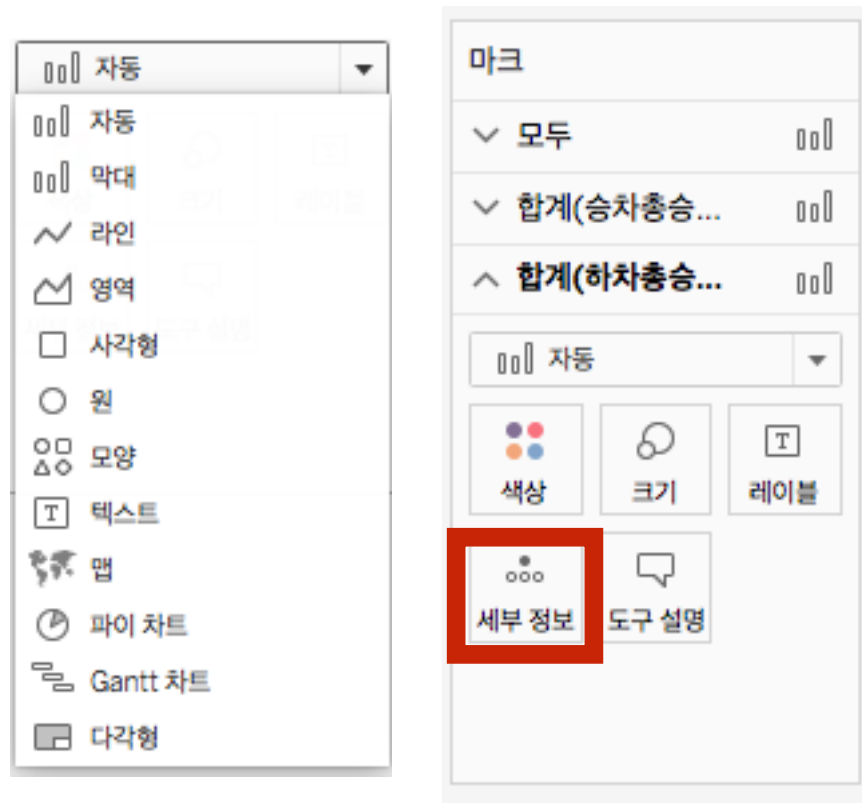
디테일 수정(측정값/표현방식)



차트선택 + 세부옵션

(세밀한 조절은
아래 Box의 조합으로 가능)

세부정보 (데이터를 묶고 풀때 사용)



고급 시각화의 핵심
(원리를 안다면
다양한 형태로 응용 가능)

표현 방식

분산형 차트의 경우 다음 시도:

0개 이상 차원
2~4개 측정값

색상

크기

레이블

세부 정보

도구 설명

열	노선명
행	합계(승차총승객수)

표현 방식

- ✓ 합계
- 평균
- 중앙값
- 카운트
- 카운트(고유)

- 최소값
- 최대값
- 백분위수
- 표준 편차
- 표준 편차(모집단)
- 분산
- 분산(모집단)

“자신감”을 가지고 시도!

2. 데이터 분석

Tableau를 통한 분석 정리

일반 통계
(평균/분포
사분위)

회귀분석
(추세분석)

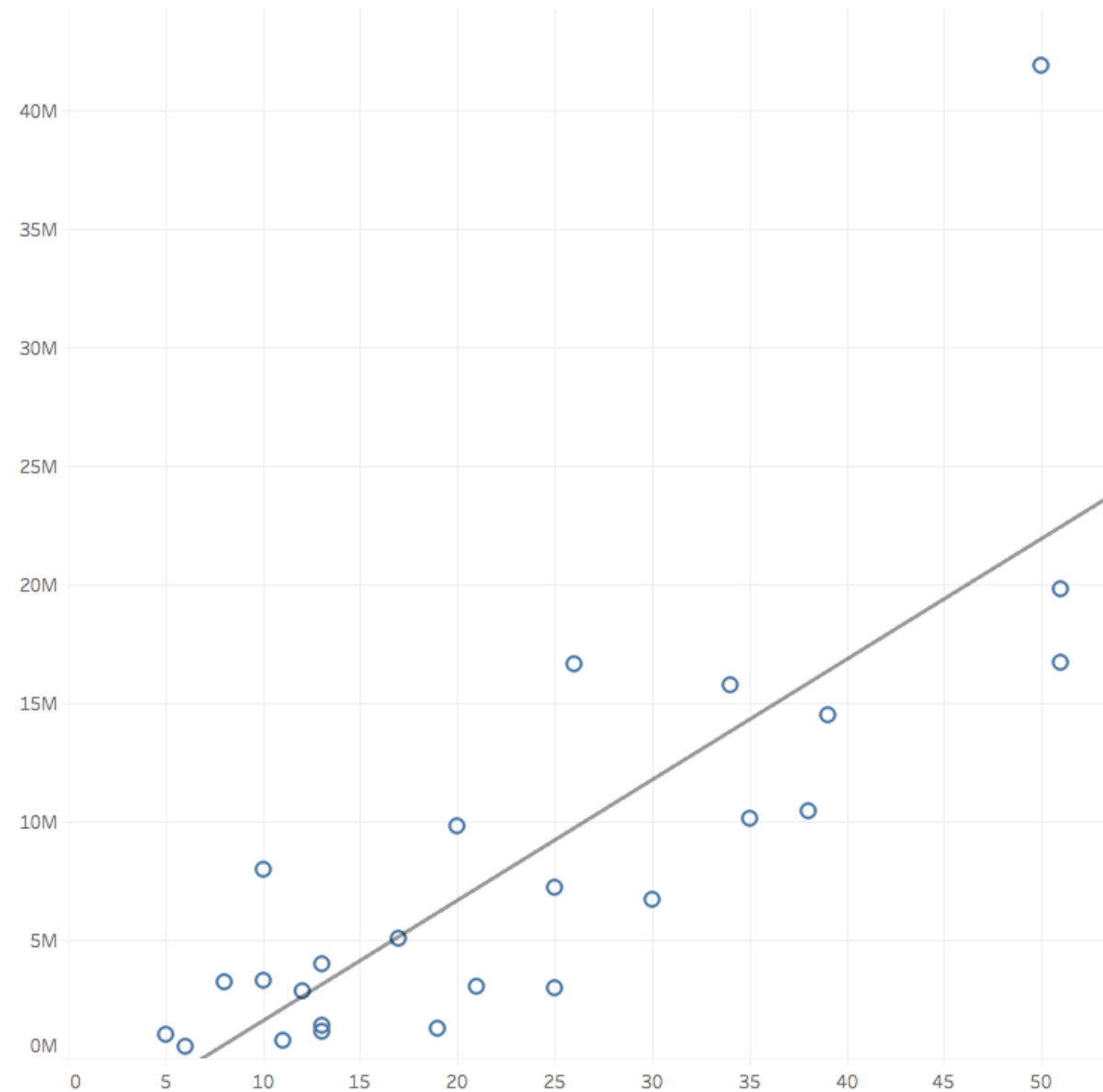
집단분석
(클러스터링)

시나리오
(What-If?)

Tableau에서는 분석과 관련해 크게 4가지 기능을 제공
+ 최근(2019 하반기) : 자연어 검색 기능 추가(Ask Data)
(본 수업에서는 다루지 않음, Desktop에서는 지원불가)

통계와 시각화 1

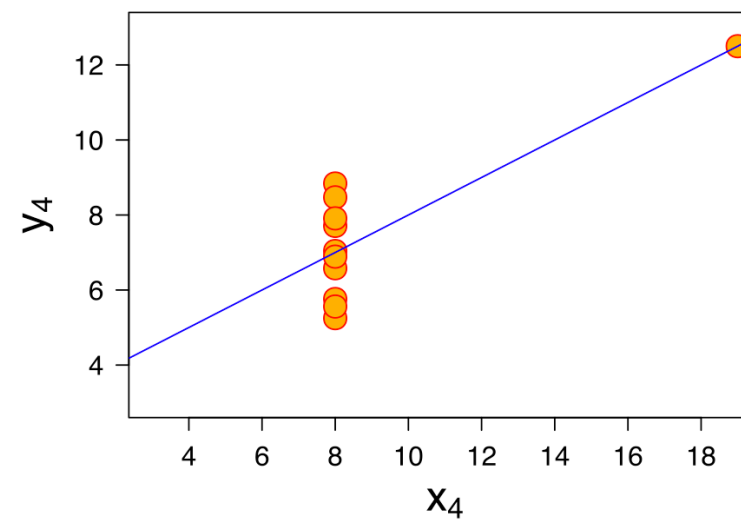
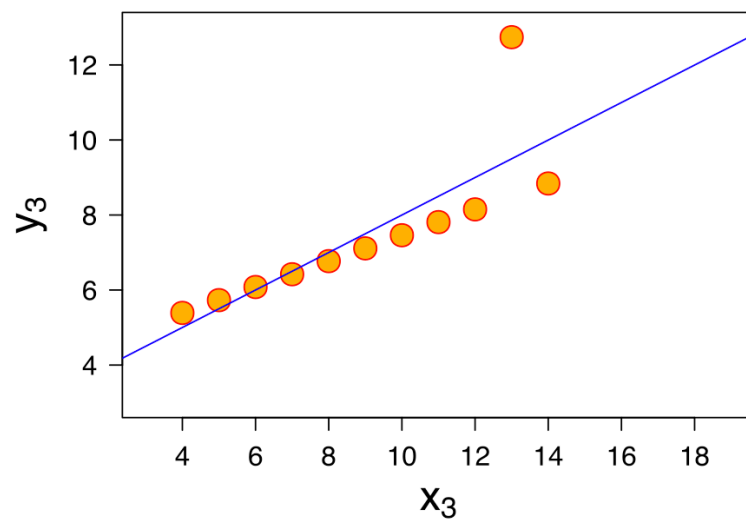
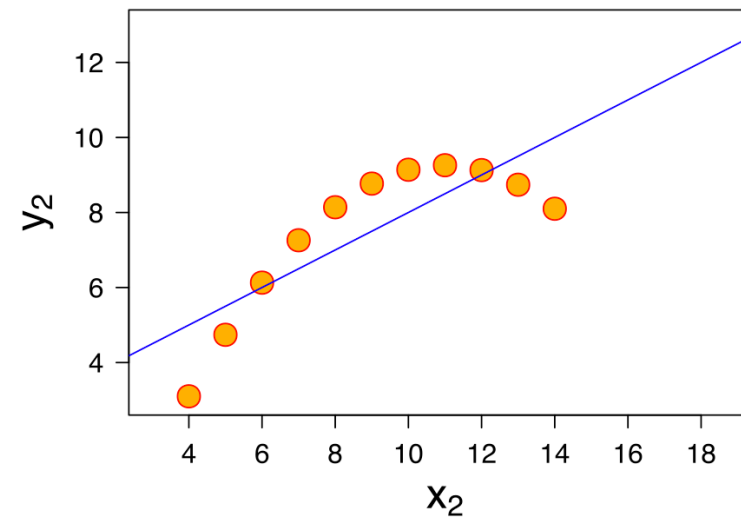
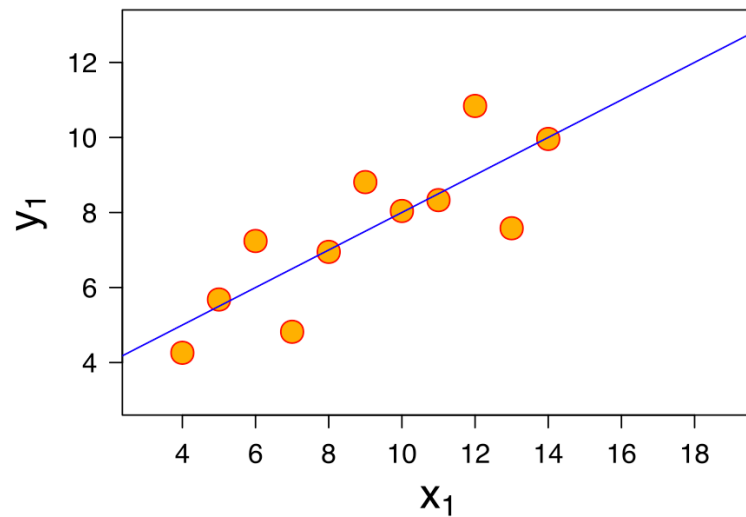
(상어와 아이스크림 - 회귀모형에서 인과와 상관)



노선별 역수는 이용량과 상관관계가 있을까?

통계와 시각화 2

(Anscombe's quartet)



통계수치상 위 데이터는 모두 “동일하다!”

규모에 의한 실수

Covid: how Excel may have caused loss
of 16,000 test results in England

3. 데이터 전처리

왜 중요한가?

CARD_SUBWAY_MONTH_201901.csv
CARD_SUBWAY_MONTH_201902.csv
CARD_SUBWAY_MONTH_201903.csv
CARD_SUBWAY_MONTH_201904.csv
CARD_SUBWAY_MONTH_201905.csv
CARD_SUBWAY_MONTH_201906.csv
CARD_SUBWAY_MONTH_201907.csv
CARD_SUBWAY_MONTH_201908.csv
CARD_SUBWAY_MONTH_201909.csv
CARD_SUBWAY_MONTH_201910.csv
CARD_SUBWAY_MONTH_201911.csv
CARD_SUBWAY_MONTH_201912.csv
CARD_SUBWAY_MONTH_202001.csv
CARD_SUBWAY_MONTH_202002.csv
CARD_SUBWAY_MONTH_202003.csv
CARD_SUBWAY_MONTH_202004.csv
CARD_SUBWAY_MONTH_202005.csv
CARD_SUBWAY_MONTH_202006.csv
CARD_SUBWAY_MONTH_202007.csv

+

신설동	37.575297	127.025087
보문	37.585286	127.019381
성신여대입구(돈암)	37.592624	127.016403
정릉	37.6026712	127.0113053
북한산보국문	37.6154595	127.0034393
솔샘	37.6202672	127.0113983
삼양사거리	37.6213452	127.0183183
삼양	37.6269262	127.0159753
화계	37.6341222	127.0153233
가오리	37.6415632	127.0146213



지하철 데이터셋

데이터 과학(분석) 작업의 80%는 데이터 전처리

실제 사례 (지하철 데이터셋 만들기)

지하철 이용
데이터
(월별로 나눔)

병합..하려는데
파일 형식이 다름
(맥/윈도우)

재인코딩...
컬럼명이 다름

좌표 데이터
없어서 추가
데이터 탐색

기존데이터+
구글 지도 데이터
코딩

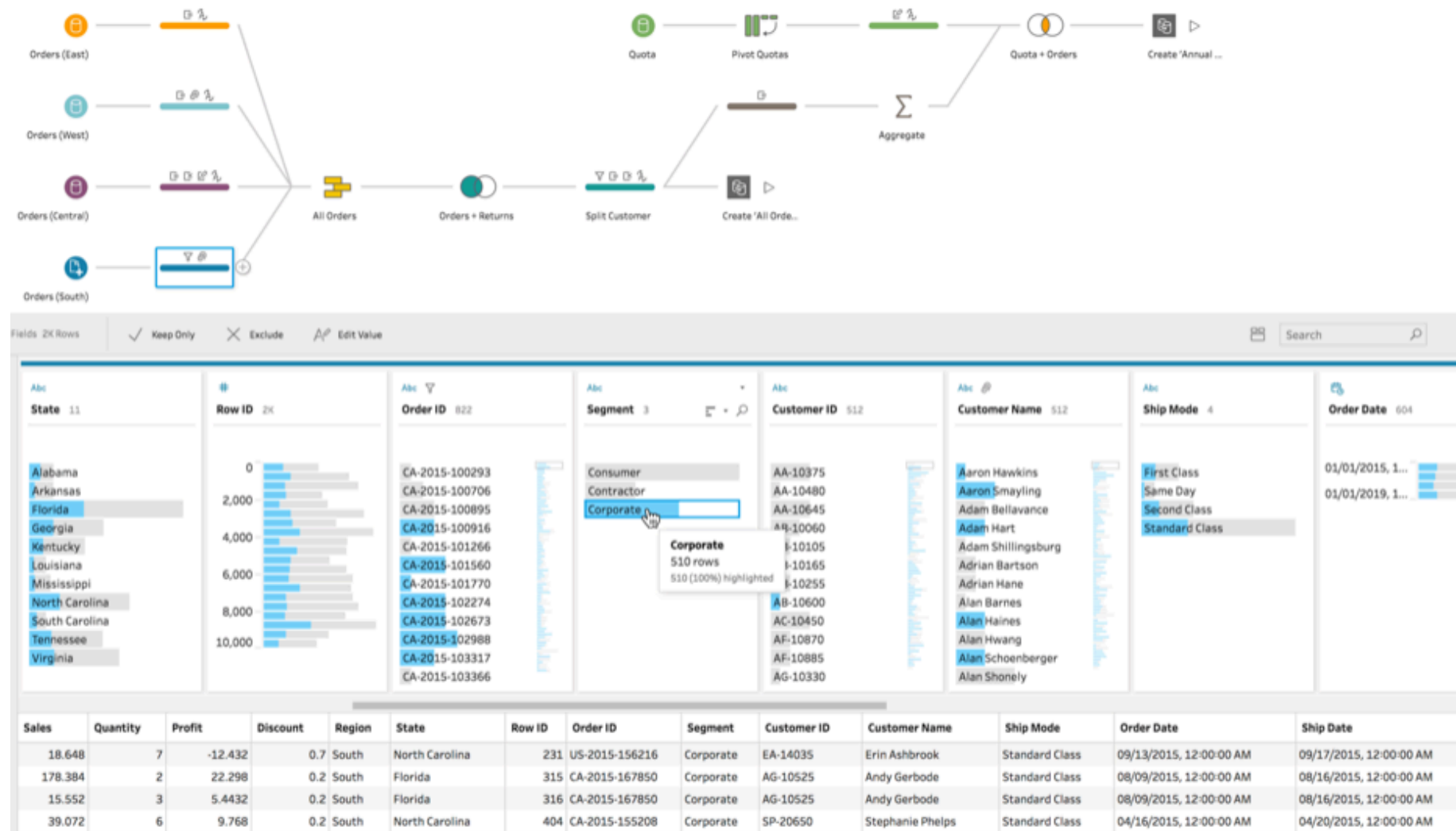
결합
하려는데...
역명이 다름

역명을 수정...
했는데 없는 역
이 존재

이 외의 사례들...

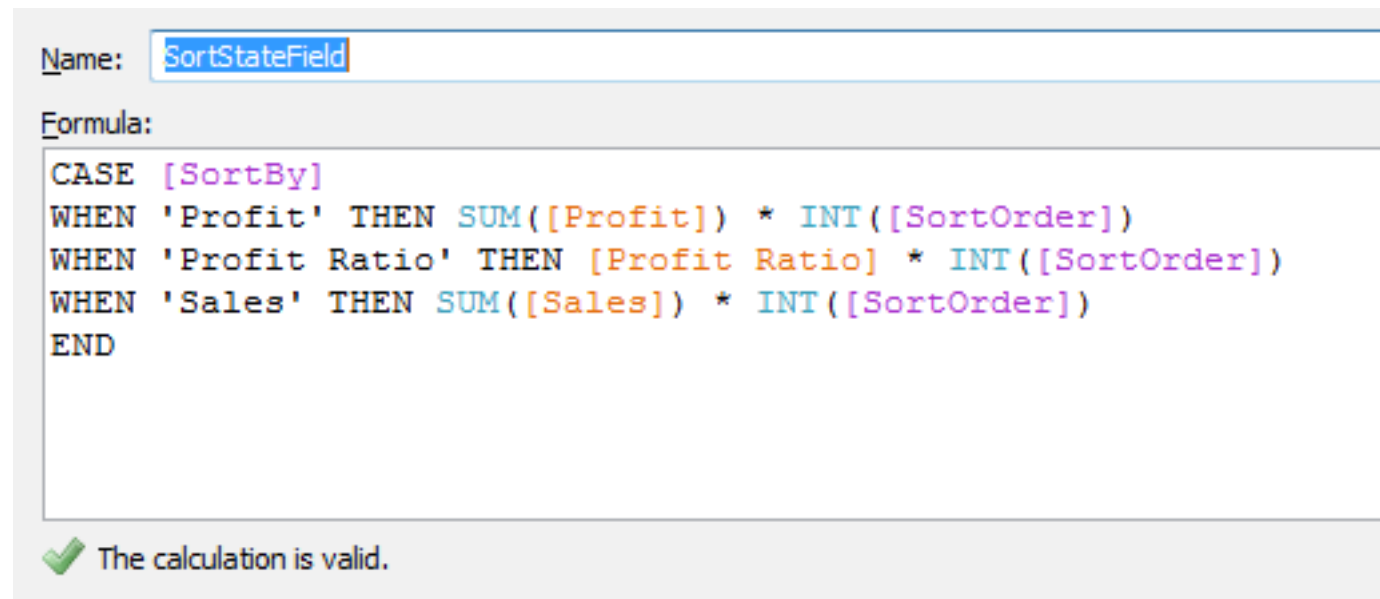
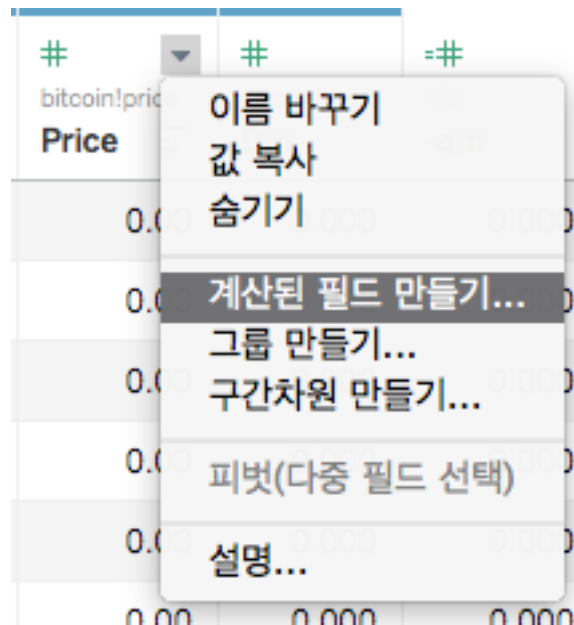
정보를 받았는데 PDF -> 표로 요청 -> 표가 그려진 PDF
주소를 받았는데 도로명 주소와 혼합된 형태
단위 통일이 되지 않은 데이터(인치, cm, 마일, km...)

Tableau Prep?



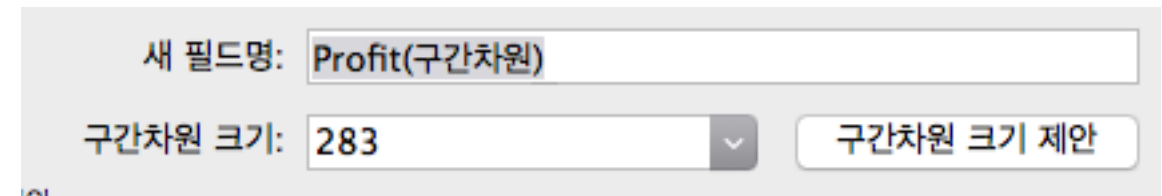
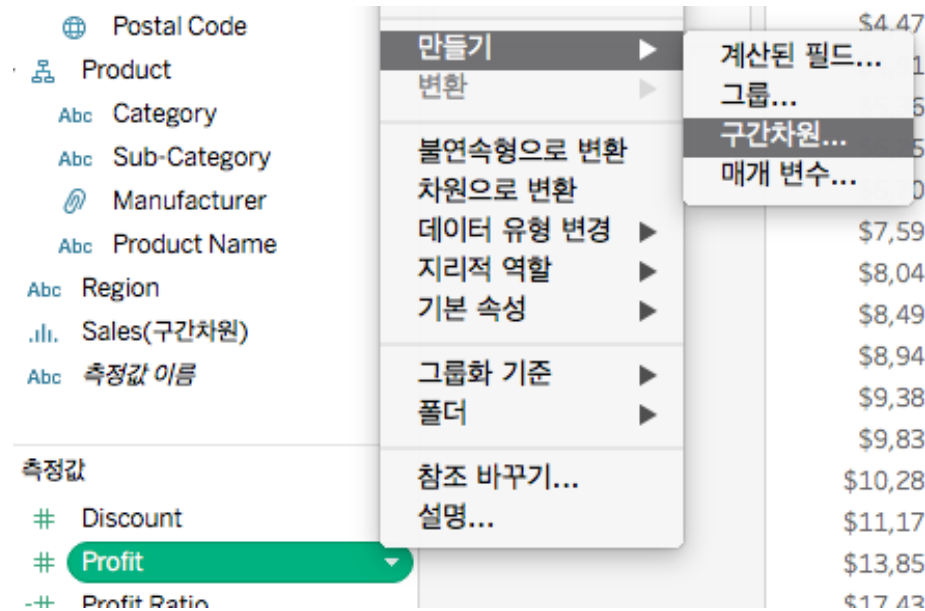
데이터 전처리 용도로 만들어진 Tableau 연계 프로그램
("대용량", "DB작업"에 적합, 본 수업에서는 다루지 않음)

자주 쓰는 전처리 1 - 수식 계산



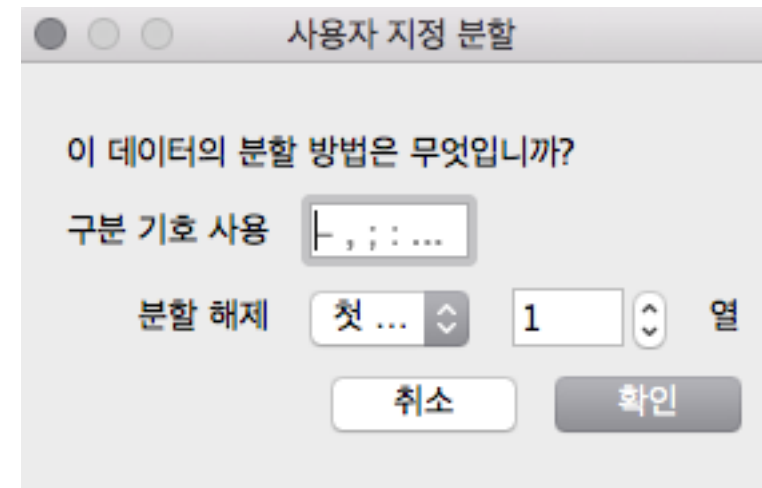
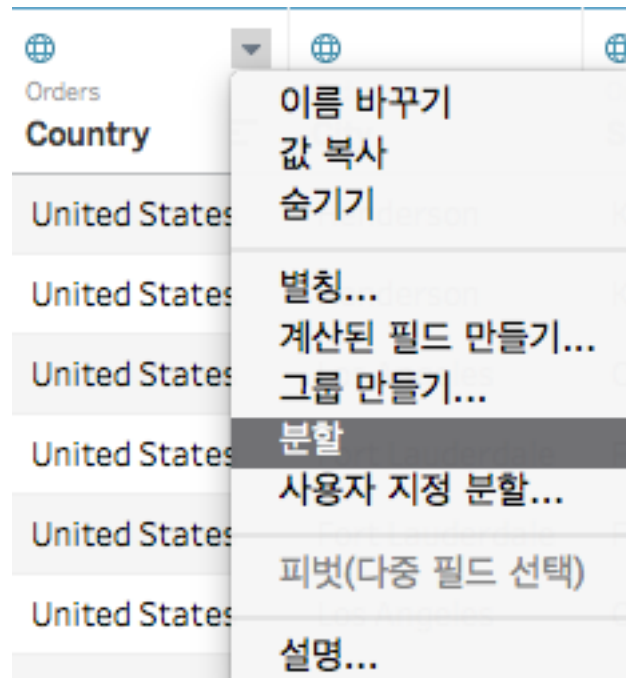
(능력이 된다면) 섬세한 작업도 가능
추천하지는 않음(대부분 Excel, Google Sheets로 처리 가능)
하나라도 “덜” 배우자!

자주 쓰는 전처리 2 - 구간차원 나누기



지정된 범위로 데이터 나눠주기
연속형 데이터를 구간별로 나눠서 확인(e.g. 연령대)
“별칭”을 쓰면 그룹 이름을 내가 만들 수 있음

자주 쓰는 전처리 3 - 셀 나누기



특정 문자열 기준으로 나눠주기
년/월/일, 단위, ID 다양한 형태로 사용 가능

4. Tableau 시각화 심화

data?

홈 > 지진·화산 > 지진 > [국내지진 목록](#)

지진 | 국내지진 목록

? 도움말

인쇄

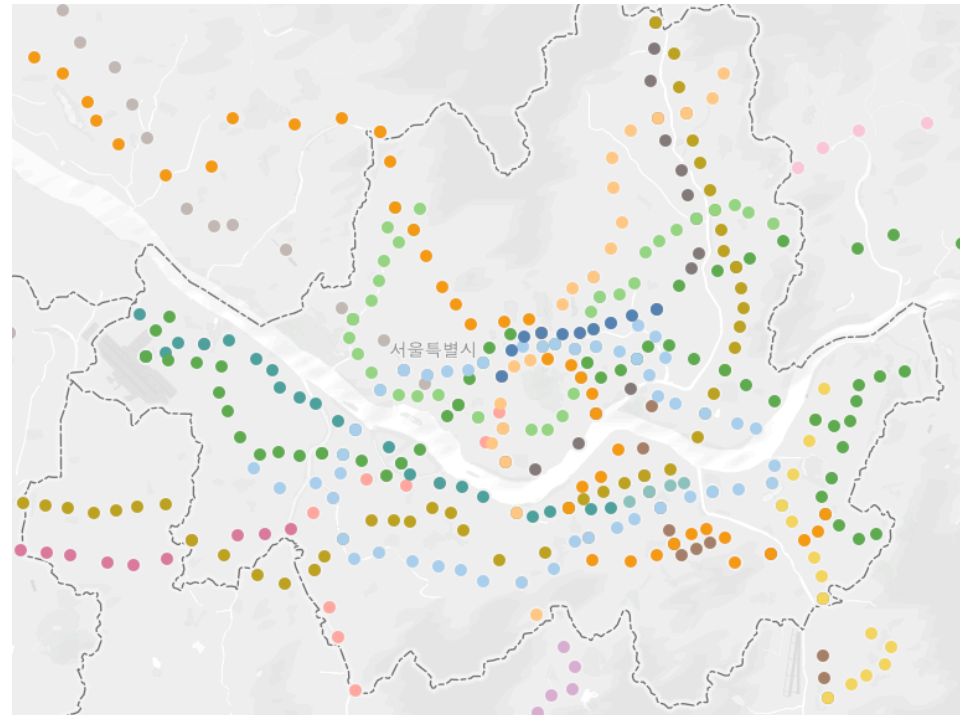
- 지진 재난문자 발송기준 안내 [바로가기](#)
- 규모 2.0이상 국내 지진발생 목록을 알 수 있습니다.
- 포항 여진 발생 현황 : 총 101회
 - 2.0~3.0 미만 : 93회
 - 3.0~4.0 미만 : 6회
 - 4.0~5.0 미만 : 2회
- ※ 최근 발생 여진 : 2020/09/26 20:52:02 경북 포항시 북구 북서쪽 5km 지역 (규모:1.3 / 깊이:6km)

[최근 미소지진 발생 현황(규모 2.0미만)]

- 2020/10/16 09:37:46 경북 청도군 동쪽 24km 지역 (규모:1.6 / 깊이:15km)

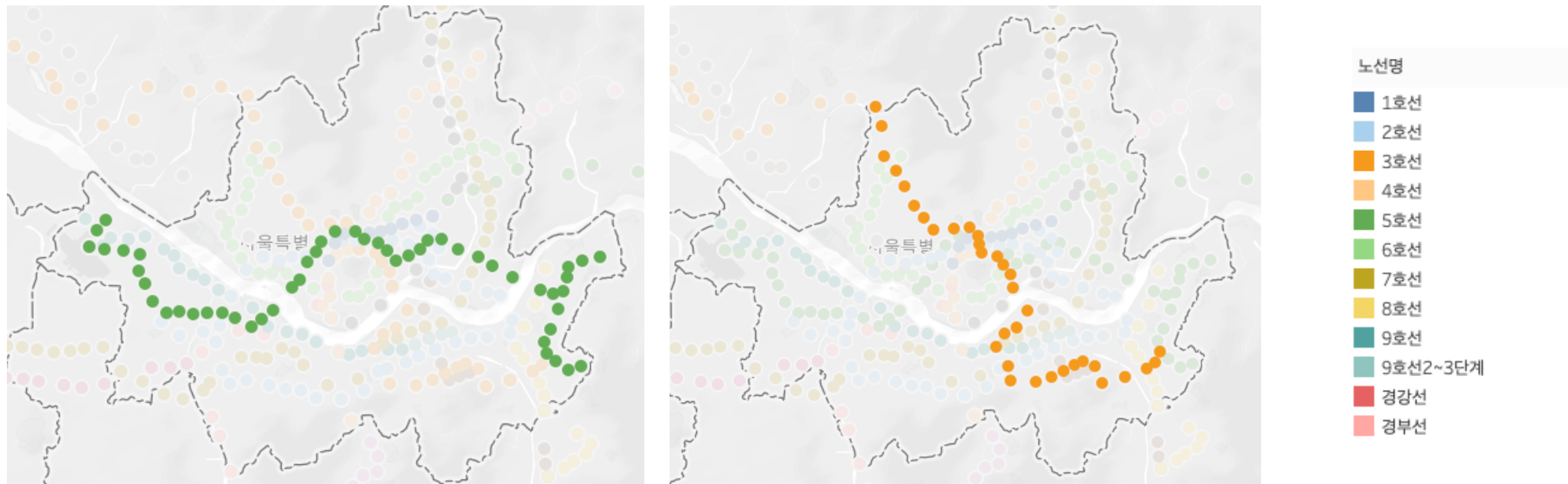
기상청 국내 지진 목록(<http://www.weather.go.kr>)
최대 1200건 조회 가능

지리 시각화



지리 데이터를 표현하는 2가지 방법 존재(지역명, 좌표)
오늘은 좌표 한가지만 배웁니다

필터/페이지

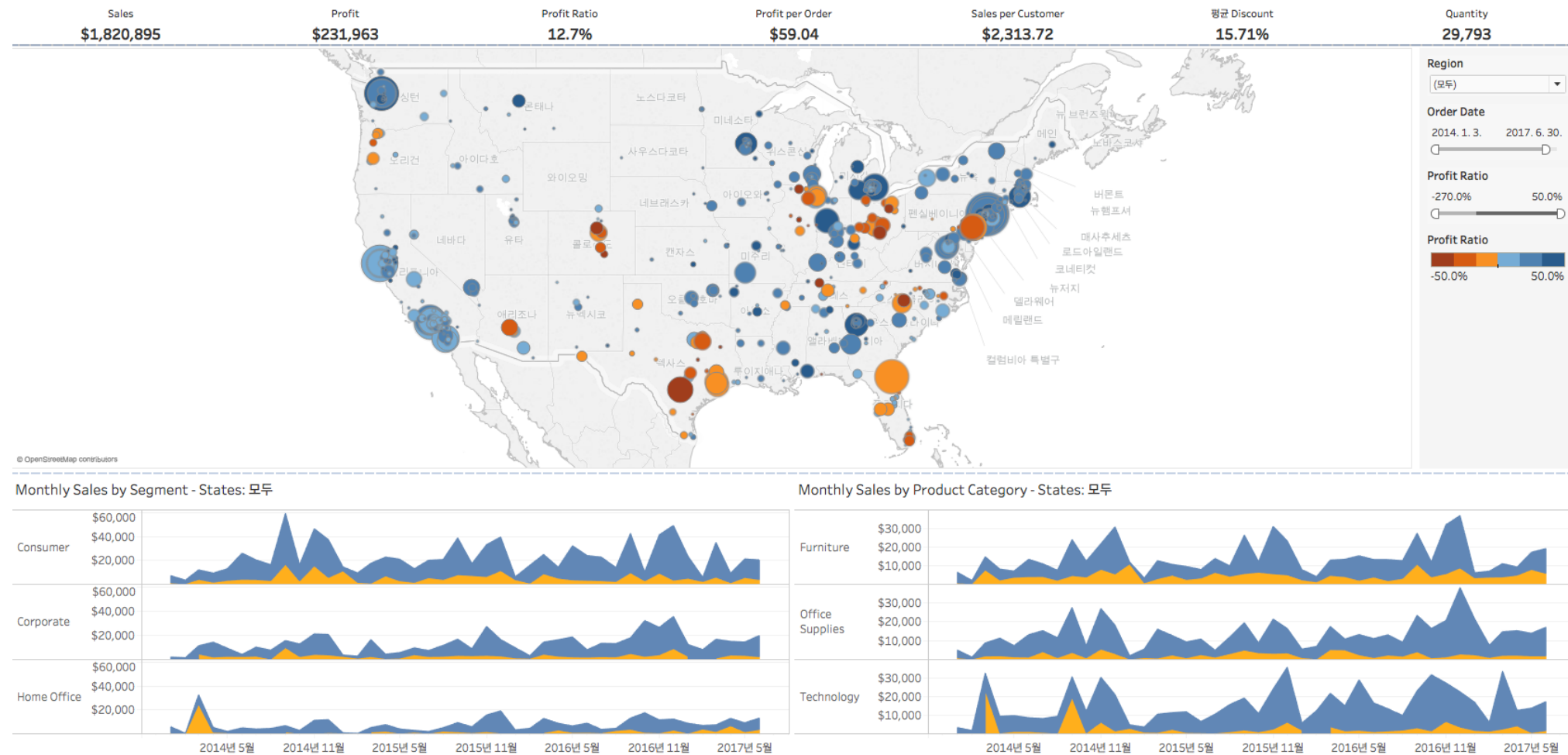


시각화에 인터랙티브 연동 효과 넣기(시나리오 분석)
오늘은 필터만 배웁니다...

5. 대시보드 구성하기

분석 결과 모아보기

Executive Overview - Profitability (모두)



각 그래프는 유기적으로 연동 / 필터 설정 가능


분석 결과 모아보기

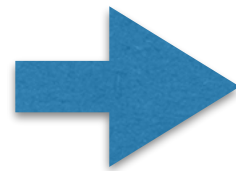


워크시트 -> 대시보드 -> 스토리(데이터연동)

6. 온라인에 게시하기

주의!

 License CC BY-NC-SA 4.0



Attribution-NonCommercial-ShareAlike
4.0 International (CC BY-NC-SA 4.0)

시각화를 공유하기 전에 자료를 검토해야 합니다
→ 연구목적으로만 사용, 공개 불가 등 제한조건 고려
(Tableau 무료 서버 사용시 자동으로 “자료공유”)

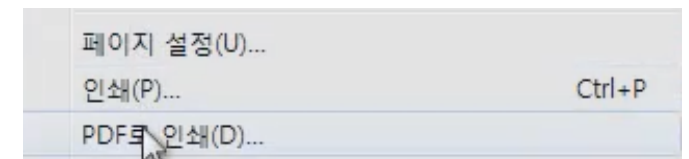
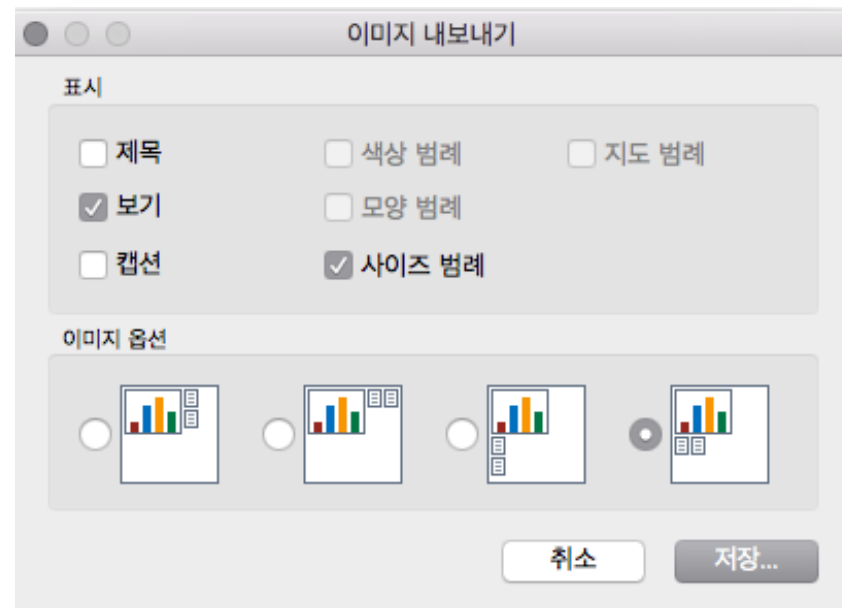
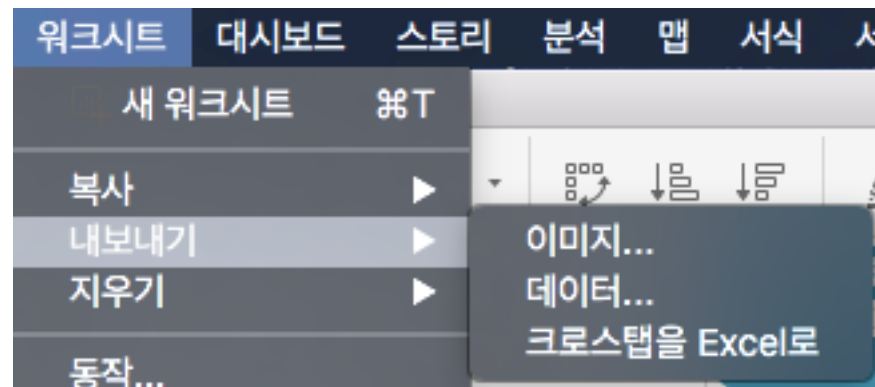
Tableau의 공유방법

스냅샷 - 내보내기 시점 그대로(Pdf, Image 형태, 리더 불필요)

Tableau문서(파일) - 데이터 갱신에 따라 변화(Reader 필요)
데이터 원본을 포함할지 안 할지 선택 가능(그래도 일부 데이터는 공개)

Tableau문서(온라인) - 데이터 갱신에 따라 변화(웹페이지 형태, Reader 불필요)
데이터 원본을 포함할지 안 할지 선택 가능(그래도 일부 데이터는 공개)

스냅샷



워크시트(대시보드) -> 내보내기 -> 옵션 중 선택
파일 -> PDF로 인쇄(통합인쇄) 가능
(+ 2019 PPTX로 내보내기 기능 추가)

tableau문서 공유

파일	데이터	워크시트	대시보드	스토리	분석	맵	서식	서버
새로 만들기								⌘N
열기...								⌘O
닫기								⌘W
저장								⌘S
다른 이름으로 저장...								⇧⌘S
저장된 내용으로 되돌리기								⌘E
패키지 통합 문서 내보내기...								

twb파일 -> 핵심 요소만 추출(간략하게, 추가 수정 어려움)

twbx -> 모든 요소 포함(데이터셋도 포함, 수정가능, 비공개 데이터일 경우 주의!)

tableau 온라인 공유


(Public 계정 or 회사서버 사용)




Tableau Public에서는 저장/내보내기가 무조건 서버 상에서만 가능
온라인에 저장하기 전 데이터 추출해야 함(내 컴퓨터에서 업로드가 이뤄지기 때문)
[데이터 추출은 데이터 -> 원하는 데이터 선택 -> 추출로 가능]

tableau문서




+tableau⁺public



Lee Gyuho  편집

1 Viz

Viz 1

 Favorite  세부 정보 편집  통합 문서 다운로드

제목

2007-2017데이터 모음

다른 사용자가 Tableau Public에서 Viz를 검색할 때 그냥 지나치지 않도록 눈에 띄는

고정 링크

URL 추가

Viz를 내장해 보십시오. 사이트로 트래픽을 유도하는 탁월한 방법입니다.

설명

testing_02

검색 결과에서 상위에 표시되려면 정확한 설명이 필요합니다.

툴바 설정

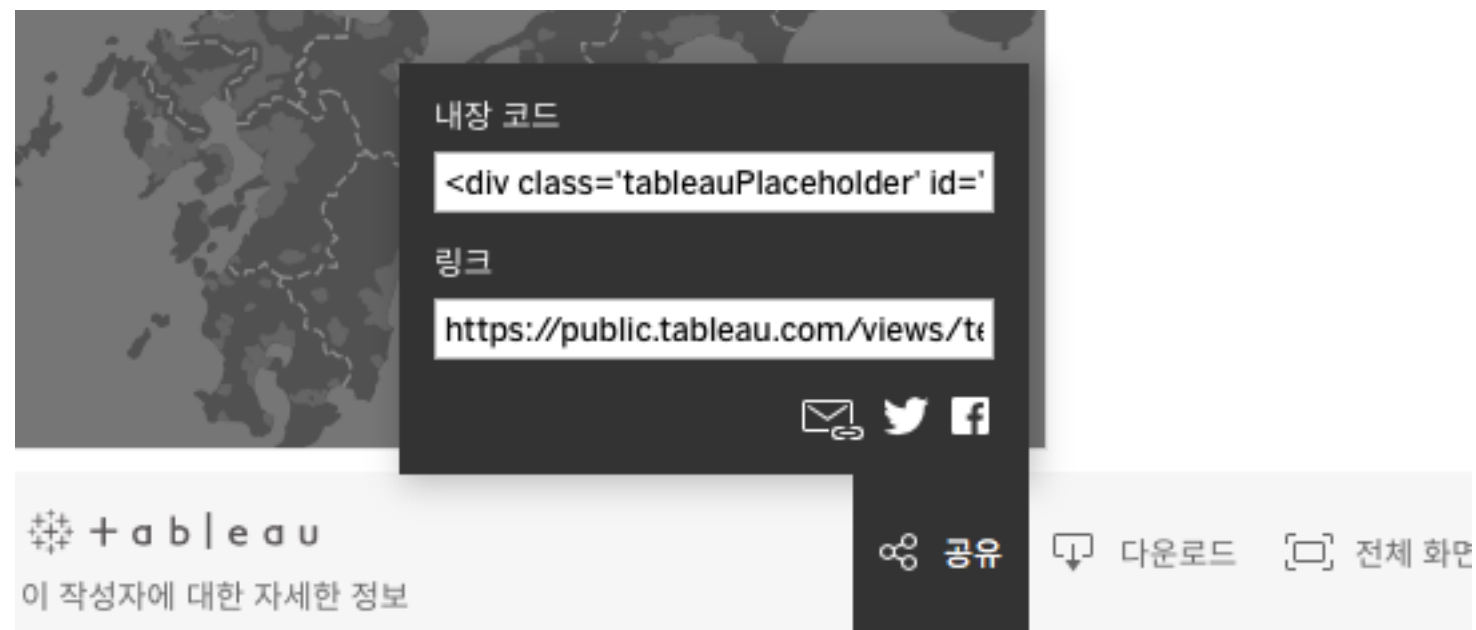
- ☒ 뷰 컨트롤 표시 실행 취소, 다시 실행, 되돌리기
- ☒ 작성자 프로필 링크 표시
- ☐ 통합 문서 및 해당 데이터를 다른 사용자가 다운로드할 수 있도록 허용

기타 설정

- ☐ 탭으로 통합 문서 시트 표시

public.tableau.com에 로그인 하면 게시한 자료 확인, 공개설정 가능

문서 공유하기



public.tableau.com로그인 한 후, 게시한 자료 선택 -> 오른쪽 아래 공유 URL 얻기
(html 문서는 내장 코드 사용)

6. 데이터 연구 사례 분석

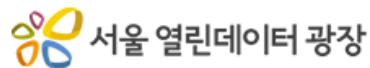
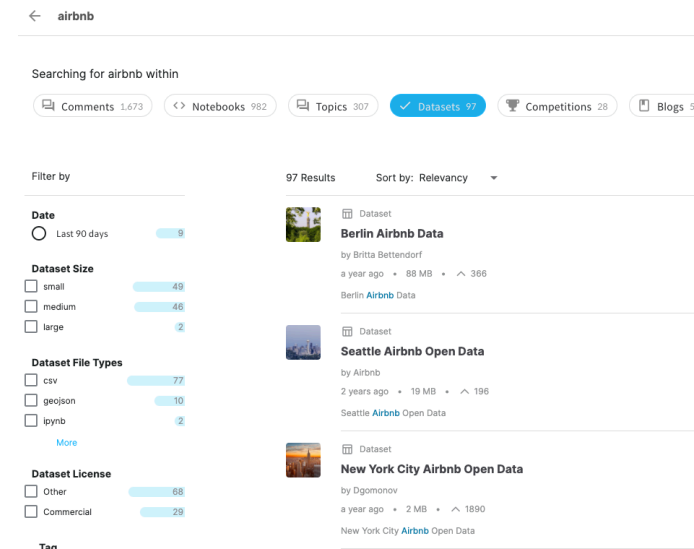
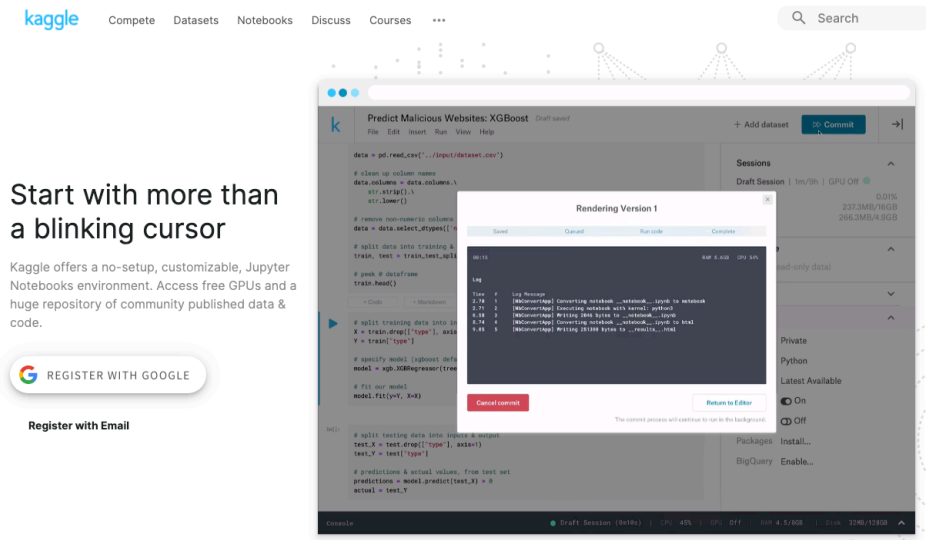
빅데이터 제대로 사용하기

Data is not information,
information is not knowledge,
knowledge is not understanding,
understanding is not wisdom

– Clifford Stoll

빅 데이터 \neq 좋은 데이터

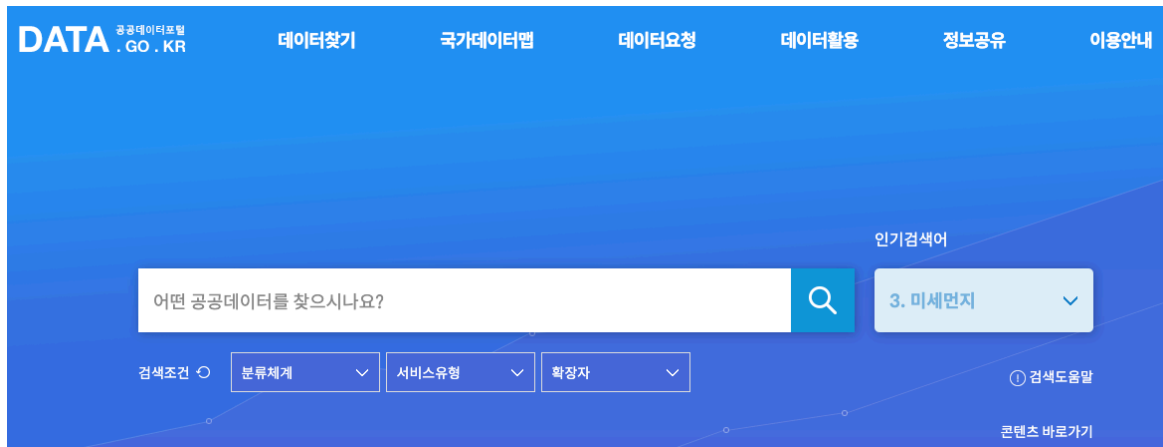
오픈 데이터 활용하기



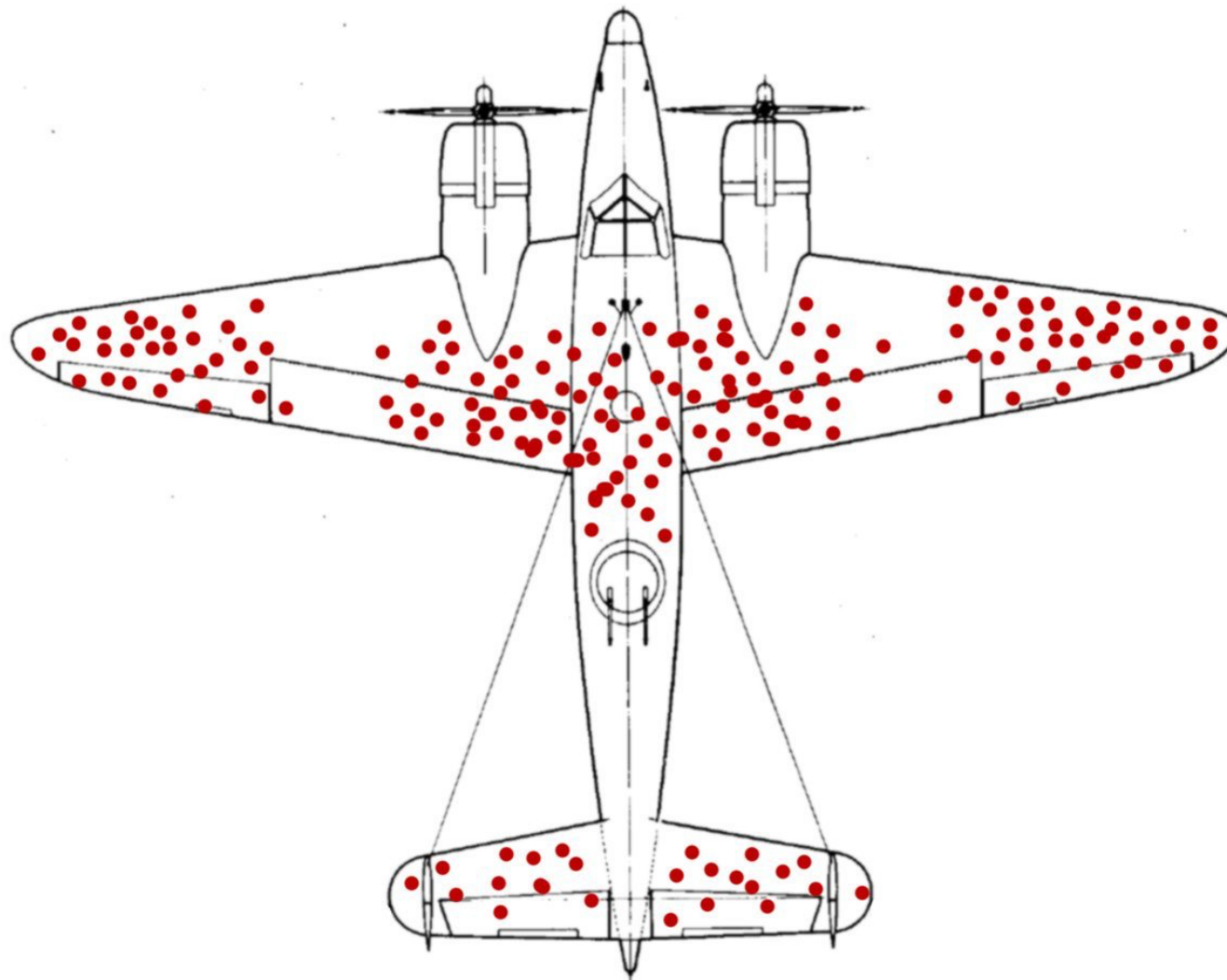
공공데이터 통계

모든 서울시민을 위한 공공데이터

열린데이터광장에서 서울시와 연계 기관이 공개한 공공데이터를 확인하실 수 있습니다.
서울시와 관련된 다양한 공공데이터를 확인해 보세요.



좋은 데이터를 판단하기 어려운 이유 (Survivorship Bias)



https://en.wikipedia.org/wiki/Survivorship_bias
<https://commons.wikimedia.org/wiki/File:Survivorship-bias.png>

일상에서 데이터 발견하기

Would you rather be hot or cold?

[https://old.reddit.com/r/dataisbeautiful/comments/iu1hse/
would_you_rather_be_hot_or_cold_oc/](https://old.reddit.com/r/dataisbeautiful/comments/iu1hse/would_you_rather_be_hot_or_cold_oc/)

Most Popular Colors In Among Us?

[https://old.reddit.com/r/dataisbeautiful/comments/j2pplt/
most_popular_colors_in_among_us_oc/](https://old.reddit.com/r/dataisbeautiful/comments/j2pplt/most_popular_colors_in_among_us_oc/)

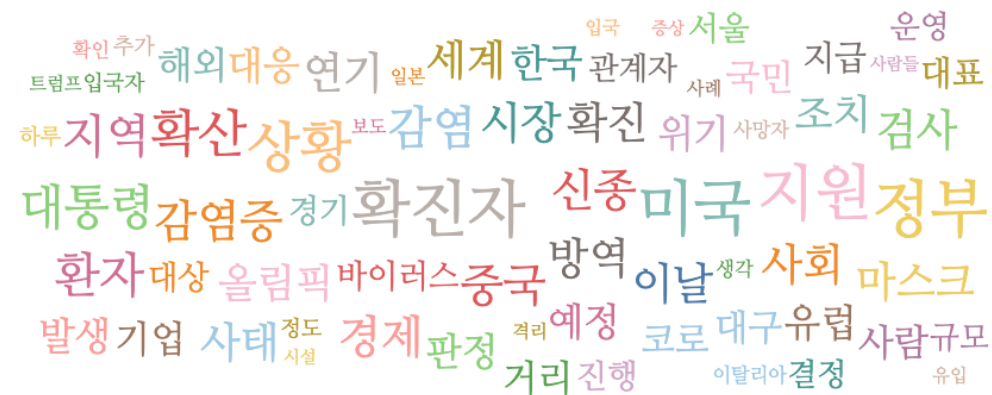
기존 데이터셋 활용하기

2020-03 코로나 보도 키워드 워드 클라우드

2020년 3월 코로나19 보도 기사(빅카인즈 뉴스 데이터 사용)에 등장한 단어(키워드)를 워드 클라우드로 확인합니다.

오른쪽 필터로 주차 및 언론사선택이 가능합니다. 키워드에 마우스를 올리면 빈도 수 확인도 가능합니다.

코로나19, 3월 언론 보도 키워드



시기

- ☐ 3월1주차
- ☐ 3월2주차
- ☐ 3월3주차
- ☒ 3월4주차

언론

- ☒ 방송사
- ☒ 중앙지

Presented by:

Gyuhoo Lee

Seoul National University
hci+d lab.
<http://hci.d.snu.ac.kr/>

자료 출처:
빅카인즈 (<https://www.bigkinds.or.kr/>)

https://eqo295.github.io/news_cloud/

Cause of Death – Reality vs. Google vs. Media

[https://www.reddit.com/r/dataisbeautiful/comments/8cwcbu/
cause_of_death_reality_vs_google_vs_media_oc/](https://www.reddit.com/r/dataisbeautiful/comments/8cwcbu/cause_of_death_reality_vs_google_vs_media_oc/)

지리 데이터 시각화

Inside Airbnb
(Murray Cox)

<http://insideairbnb.com/>

프로젝트 내용
(https://en.wikipedia.org/wiki/Inside_Airbnb)

규모를 통한 시각화

Open Access | Published: 16 May 2017

Inferring personal economic status from social network location

Shaojun Luo, Flaviano Morone, Carlos Sarraute, Matías Travizano & Hernán A. Makse 

Nature Communications **8**, Article number: 15227 (2017) | [Cite this article](#)

2040 Accesses | **25** Citations | **147** Altmetric | [Metrics](#)

(전문)

<https://www.nature.com/articles/ncomms15227>

(시각화)

<https://www.nature.com/articles/ncomms15227/figures/1>

데이터 저널리즘

Tracking Federal Purchases to Fight the Coronavirus

시각화

<https://projects.propublica.org/coronavirus-contracts/>

+Federal Procurement Data System

https://www.fpds.gov/fpdsng_cms/index.php/en/

새로운 기술과 결합하기

Sonic Cities
(MIT Senseable City Lab)

<http://senseable.mit.edu/sonic-cities/>

7. 정리/질문시간

오늘 다루지 않은 내용들

(아마도 다음 기회에?)

데이터 연결/조작

온라인 데이터 연결(+크롤링), 데이터셋 조합(Join, Union 등..), DB다루기

시각화 응용/심화

지리 데이터, 시계열 데이터 심화

페이지, 필터 기능 심화

파라미터 응용(인터랙티브 핵심!)

머신러닝/통계

Python, R + 통계기법 시각화(+상관관계 분석)

기타

온라인 공유 심화 + 팁 + 모범사례(?)

특이한 시각화 기법들(워드 클라우드, 순위 차트, 간트 차트 등...)

...