서울대학교 4차 산업혁명 아카데미 - 인공지능 에이전트

Information Visualization and HCI

Tableau 실습

Wonjae Kim 김원재 Human-Computer Interaction Laboratory Seoul National University





Tableau

1. 태블로 개요



태블로란?



Analytics for everyone, everywhere, anytime

비즈니스 환경에 최적화 된 강력한 시각화 소프트웨어



태블로의 강점

• 강력한 시각화: 엑셀에서 지원하는 기본적인 차트뿐만 아니라 복잡한 시나리오에 적합한 다양한 차트 지원

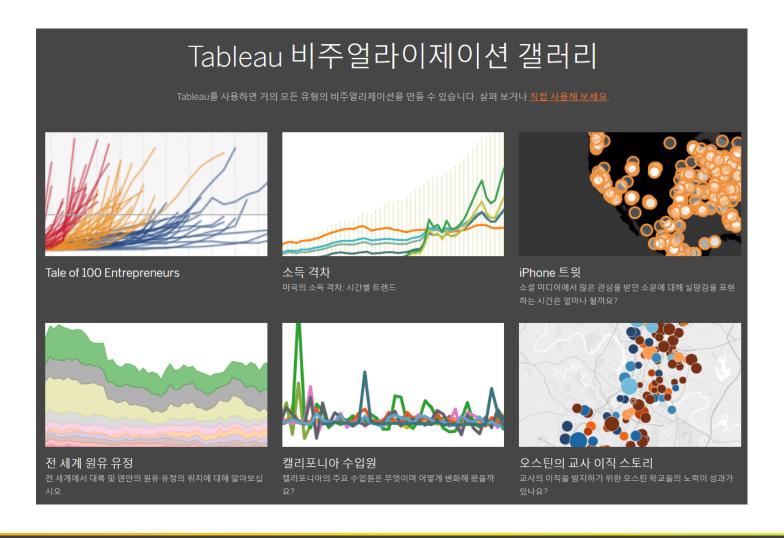
• 사용하기 쉬움: 쉽게 데이터를 불러와 몇 번의 클릭으로 인터랙티브하고 공유 가능한 차트를 만들 수 있음

• 뛰어난 성능 : 큰 데이터도 효과적으로 처리하여 빠르게 차트를 그려볼 수 있음



태블로 시각화 갤러리

https://www.tableau.com/ko-kr/stories/gallery





예제

- 수도권 톨게이트 교통 정보 데이터
 - 날짜 별로 어떤 영업소에 어떤 차종이 몇 대 지나갔나?

• 데이터를 Tableau에서 불러와 탐색하며 어떤 특이한 점들을 발견할 수 있는지 봅시다.

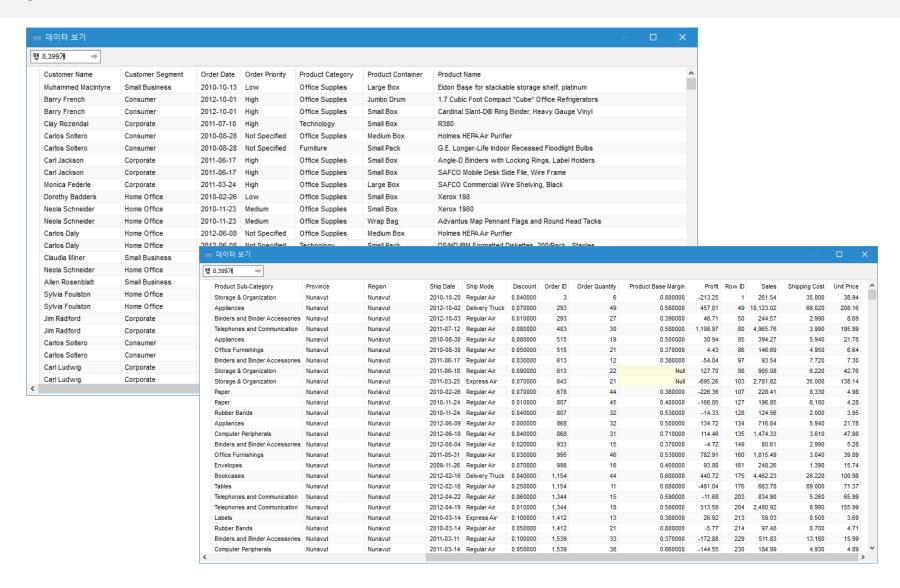


오늘 배울 것

- 1. 데이터 연결
- 2. 기본 뷰 만들기
- 3. 데이터 다루기와 계산된 필드 만들기
- 4. 태블로 대시보드
- 5. 데이터 다루기+



Superstore 데이터





목표는 데이터를 이해하는 것

- Superstore 데이터는 아래와 같은 많은 정보를 담고 있음
 - 판매량의 전년 대비 성장률이 가장 높았던 때는 2011년 8월
 - 판매량의 전년 대비 성장률이 가장 낮았던 때는 2010년 3월
 - 2011년만 1월 수익이 2월 수익보다 낮음
 - 2010년 3월에는 수익이 아예 없었음
 - 등등...

• 예제 데이터를 태블로를 이용해서 탐색해 봅시다.



준비물

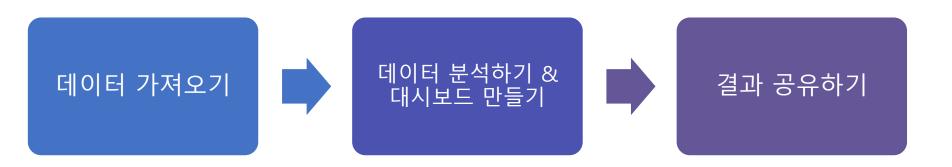
• 태블로 10.4 평가버전 (http://www.tableau.com/products/trial)

• 보고 싶은 데이터 파일

• 그 데이터를 어떻게 시각화 할 것인지에 대한 아이디어



태블로의 세 단계



태블로 개요

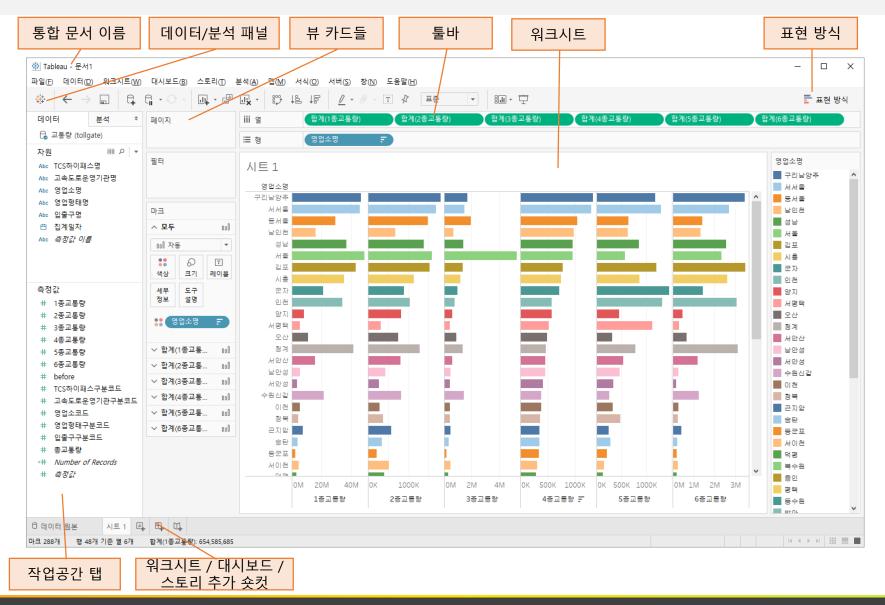


시작 화면





태블로 작업공간



13



기본 메뉴

- 시작페이지 / 작업공간 전환
 - 태블로 아이콘 클릭



• 혹은 Ctrl+2

- 데이터 원본 탭으로 이동
 - 작업공간에서 아래에 있는 데이터 원본 탭 선택





통합 문서와 워크시트

• 통합 문서: 데이터 + 시각화



- 데이터 원본: 데이터를 불러올 수 있음
- 시트: 하나의 차트가 보이는 공간
- 대시보드: 여러 시트를 조합해서 여러 개의 차트를 한 화면에서 볼 수 있는 공간

• 데이터 원본, 시트, 대시보드는 작업공간 아래에 탭으로 배치되어 이동할 수 있음

태블로 개요



실습

16

- 1. 태블로 실행
- 2. 새로운 통합 문서 만들기
- 3. 새로운 시트 추가하기
- 4. 데이터 원본 탭 들어가보기
- 5. 새로운 대시보드 추가하기



Tableau

1. 데이터 연결

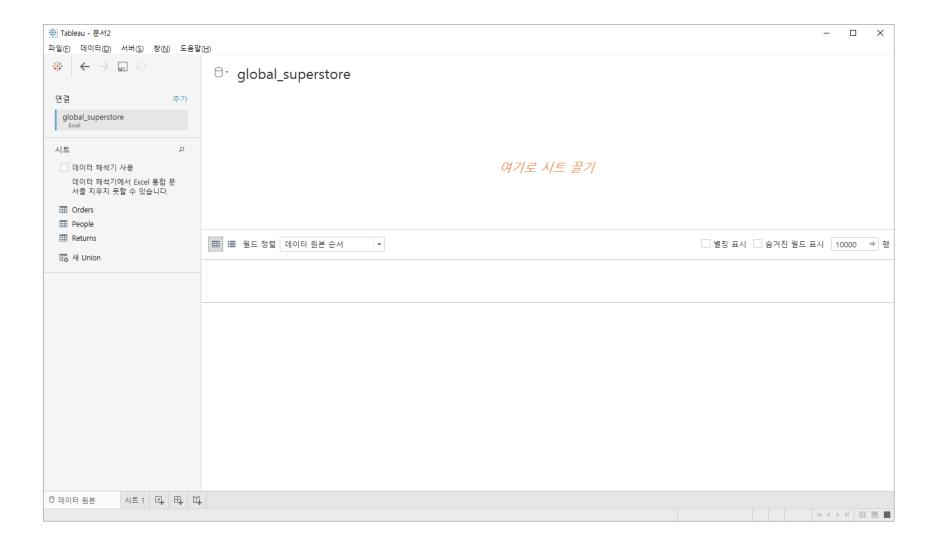


태블로에서 불러올 수 있는 데이터

- 거의 모든 형태의 데이터를 불러올 수 있음
 - Excel, csv, tsv, sas, spss, rdata
 - 혹은 데이터베이스 서버에서 직접 연결도 가능
 - MySQL, Oracle, Redshift
- 가장 간단한 방법: 데이터 (엑셀 파일)를 드래그-앤-드랍으로 태블로 영역에 끌어서 넣기



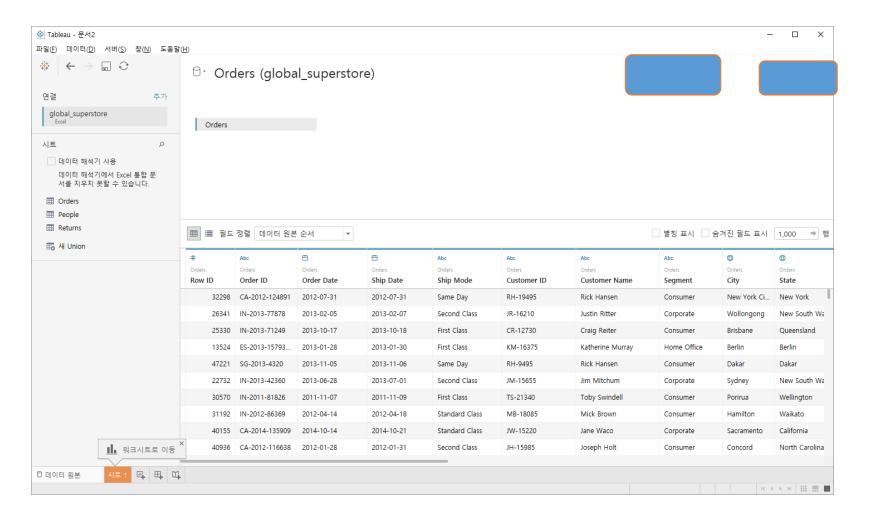
데이터 불러오기





데이터 불러오기

• 불러온 엑셀 파일에 있는 시트 중 분석에 사용할 시트들을 끌어다 놓기





라이브 연결 / 추출 연결과 필터

• 연결의 종류

• 라이브 연결과 추출 연결이 있는데 기본값인 라이브 연결을 사용하시면 됩니다.

• 필터

- 큰 데이터 중 일부만 불러오고 싶을 때 조건을 지정하여 원하는 데이터만 불러올 수 있습니다.
- 엑셀의 필터 기능과 유사



데이터 열의 종류

- 차원 (Dimension)
 - 질적인, 범주형 정보
 - 성별, 국적, 결혼상태
 - 날짜
- 측정값 (Measure)
 - 양적인, 수치형 정보, 연속형과 불연속형이 있음
 - 나이, 신용등급 (불연속형)
 - 수입 (연속형)

#	Abc	ä	=	Abc	Abc	Abc	Abc	⊕	#
Orders	Orders	Orders	Orders	Orders	Orders	Orders	Orders	Orders	Orders
Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State
32298	CA-2012-124891	2012-07-31	2012-07-31	Same Day	RH-19495	Rick Hansen	Consumer	New York Ci	New York
262/11	INI_2012_77979	2012-02-05	2012-02-07	Second Class	IR-16210	luctin Ritter	Corporate	Wollongong	New South



데이터 형식

- 태블로가 자동으로 값들을 보고 열에 알맞은 데이터 형식을 지정함
- 만약 이것이 틀린 경우 사용자가 직접 데이터 형식을 정할 수 있음
 - 데이터 원본 탭 > 열 이름 위에 있는 데이터 아이콘 클릭 > 원하는 데이터 형식 설정

아이콘	데이터 형식
Abc	텍스트(문자열) 값
Ė	날짜 값
超	날짜 및 시간 값
#	숫자 값
T F	부울 값(관계형에만 해당)
#	지리적 값(맵에 사용)



실습

- 1. global_superstore.xls 불러오기
- 2. 데이터 원본 탭에서 Order 시트를 드래그 앤 드랍을 통해 연결해보기
- 3. Order 시트에 어떤 열과 값들이 있는지 확인하기
- 4. 시트 1로 들어가기
- 5. 시각 마크 확인해보기
- 6. 표현 방식 확인해보기
 - 태블로는 어떤 시각화들을 프리셋으로 지원할까?



Tableau

2. 기본 뷰 만들기



기본 아이디어

- 데이터에 있는 차원 및 측정값들을 차트의 어떤 역할에 매핑 할지 결정
 - 드래그 & 드랍
- 가능한 역할: 열, 행, 페이지, 필터, 색상, 크기, 레이블 등

• 열 하나에 특정 역할을 지정했을 때 어떤 시각화가 보일 지 이해하는

것이 중요







기본 아이디어

• barley.csv 데이터는 종/재배 위치/년도에 따른 보리의 수확량을 담고 있는 데이터입니다.

- 데이터 원본에서 이 데이터는 어떻게 표시될까요?
- 차원과 측정값에는 어떤 것이 있을까요?
- 태블로 작업공간의 각 카드는 어떤 역할을 할까요?
- 직접 열을 옮겨보면서 확인해 봅시다!

기본 뷰 만들기



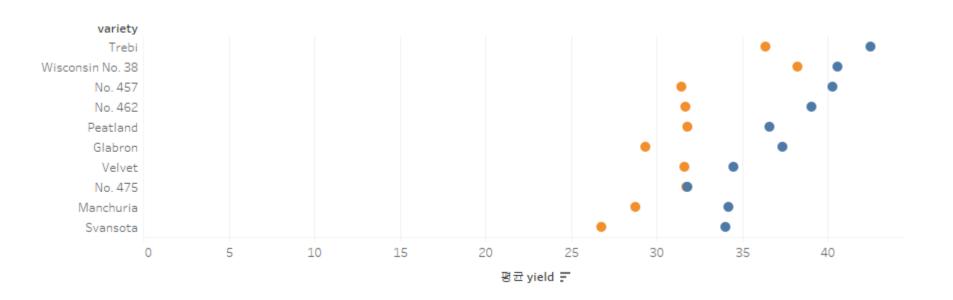
실습: 기본적인 차트 그리기

- 1. barley.csv 데이터 연결하기
- 2. variety에 따른 yield의 평균을 바 차트로 나타내기
- 3. 파이 차트로 변환하기
- 4. 파이 차트 소팅하기
- 5. 바 차트를 통해 1931년에 어떤 variety가 가장 많은 yield를 냈는지 알아보기



실습: 기본적인 차트 그리기

• 아래 같은 차트는 어떻게 만들 수 있을까요?



기본 뷰 만들기



실습: 트랜드 차트 그리기

- 1. global_superstore.xls 데이터 연결하기
- 2. Order Date에 따른 Profit 시계열 라인 차트 그려보기
 - 1. 수익이 무슨 요일에 가장 높고, 가장 낮은지 알아보기
 - 2. 수익이 몇월에 가장 높고, 가장 낮은지 알아보기
 - 3. 월에 대한 Profit 시계열 라인 차트에서 추세선 그려보기

기본 뷰 만들기



실습: 마크와 맵

- 1. global_superstore.xls 데이터 연결하기
- 2. 채워진 맵 차트와 색상 마크를 통해 높은 Profit을 얻고있는 나라를 판단해보기
- 3. 페이지 기능을 이용하여 2014년 Profit의 전년 대비 성장률이 낮아진곳과 높아진 곳을 확인해보기



Tableau

3. 데이터 다루기와 계산된 필드 만들기

데이터 다루기와 계산된 필드 만들기



실습:계산된 필드

- 1. global_superstore.xls 데이터 연결하기
- Discount를 한 것에 비해 수익이 얼마나 나는지를 알기 위하여 ratio
 계산된 필드를 생성
 - ratio = Profit / Discount
- 3. Ratio 필드를 통해서 멕시코와 브라질은 discount에 비해 많은 profit을 올리고 있음을 알 수 있음
- 4. 색상표에서 빨간색-녹색 다중을 선택하고 전체 색상 범위 사용에 체크하면 터키와 나이지리아는 discount는 많이 했지만 profit은 낮음을 알 수 있음



실습: 날짜와 관련된 함수와 필터

- 1. global_superstore.xls 데이터 연결하기
- 2. 배송 딜레이가 얼마나 일어났는지 알아보기 위해서 Delay 필드를 생성
 - delay = Ship Date Order Date
 - 또는 delay = DATEDIFF('day', [Order Date], [Ship Date])
- 3. 색상표에서 빨간색-녹색 다중을 선택하고 반전에 체크해서 딜레이가 짧을수록 녹색을 띄게 함
- 4. Delay 필드의 평균을 통해서 Chad가 delay가 매우 낮은 걸 알 수 있지만, Quantity의 합계를 필터로 걸어서 관찰하면 표본의 수가 적어서 나타난 현상임을 알 수 있다.
- 5. Quantity 합계가 100 이상인 나라만 보면, 수단과 아프가니스탄 등에서 딜레이가 많이 일어났고, 카자흐스탄, 러시아, 캐나다 등이 딜레이가 적다는 것을 알 수 있다.



Tableau

4. 태블로 대시보드



실습: 비쥬얼 필터

- 1. global_superstore.xls 데이터 연결하기
- 2. 시트 1에 월(Order Date) x Profit 합계, 그리고 연도를 컬러 마크에 지정, Profit 합계는 전년 대비 성장률로 하여 라인 차트를 생성
- 3. 워크 시트를 추가하여 시트 2에 Country x Profit 합계의 채워진 맵 생성
- 4. 대시보드를 추가하여 크기를 자동으로 지정 후 시트 1과 시트 2를 상하로 배치
- 5. 시트 1과 시트 2의 **필터로 사용**을 클릭
- 6. 라인 차트의 점과 지도를 브러싱해보면서 필터가 어떻게 적용되는지 살펴보기



Tableau

5. 데이터 다루기+



Join과 Union

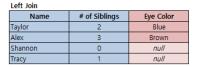
• 두 개 이상의 시트를 가진 엑셀파일이나 다른 데이터 형식에 대해서 이 시트들을 필드명 기준으로 합쳐서 하나의 데이터로 만드는 방법

Name	# of Siblings
Taylor	2
Alex	3
Shannon	0
Tracy	1

Name	Eye Color
Taylor	Blue
Alex	Brown
Morgan	Brown

Inner Join			
Name	# of Siblings	Eye Color	
Taylor	2	Blue	
Alex	3	Brown	





Right Join			
Name	Eye Color	# of Siblings	
Taylor	Blue	2	
Alex	Brown	3	
Morgan	Brown	null	

	~	
(



Union			
Name	# of presentations	Expenses	Expense report
Kai	10	\$3,761	null
Piper	3	\$287	null
Lien	4	\$1,008	null
Kai	10	null	\$4,024
Tori	7	null	\$2,930
Clair	4	null	\$987

Expenses

\$3,761

\$287

\$1,008

Name # of presentations

Outer Join			
Name	# of Siblings	Eye Color	
Taylor	2	Blue	
Alex	3	Brown	
Shannon	0	null	
Tracy	1	null	
Morgan	null	Brown	



Join

Union

Name # of presentations

Expense report

\$4,024

\$2,930

\$987

데이터 다루기+



계층 구조

- 필드가 계층 구조를 이루고 있을때(e.g., global_superstore 데이터의 Category 필드와 Sub-Category 필드) 이를 묶는 기능
- 계층 구조를 형성했을 때는 차원이 여러 개인 마크(e.g., 색상(Hue, Saturation, Value))에 대응시킬 수 있다.

데이터 다루기+



실습: Join과 계층 구조

- 1. global_superstore.xls 데이터 연결하기
- 2. Orders와 People의 Inner Join 형성하기
- 3. Sub-Category를 Category에 끌어서 계층 구조 형성하기
- 4. Category와 Sub-Category를 색상 마크에 매핑하고 Sales 합계를 각도에 맵핑하여 어떤 파이 차트가 형성되는지 보기
- 5. Person을 세부정보 마크에 매핑하고, Category와 Person을 Sales 필드에 대해 내림차순 정렬하여 파이 차트의 변화를 관찰하기



자동 생성되는 필드들

- 자동 생성되는 필드들은 *이탤릭체*로 표시된다
 - 측정값 이름, 측정값, 레코드 수, 경도, 위도
- 측정값 이름 : 모든 측정값들의 이름이 한 필드로 모아진 것
- 측정값 : 모든 측정값이 한 필드로 모아진 것
 - 측정값 이름 x 측정값 바 차트를 그려보면 쉽게 이해 가능
- 레코드 수 : 지정된 기준으로 묶인 레코드의 수를 보여줌
- 경도, 위도 : 지도에 매핑시키고자 할 때 지리적 역할 필드의 열과 행 값을 나타냄



Tableau

6. 태블로 10에서 추가된 점

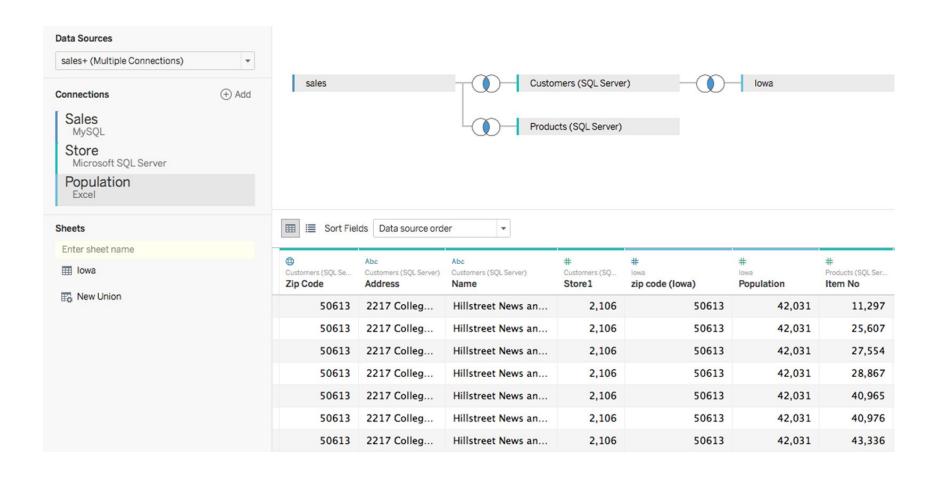


태블로 10

- 2016년 9월에 업데이트 된 태블로의 최신 메이저 버전
- 현재 버전은 10.4

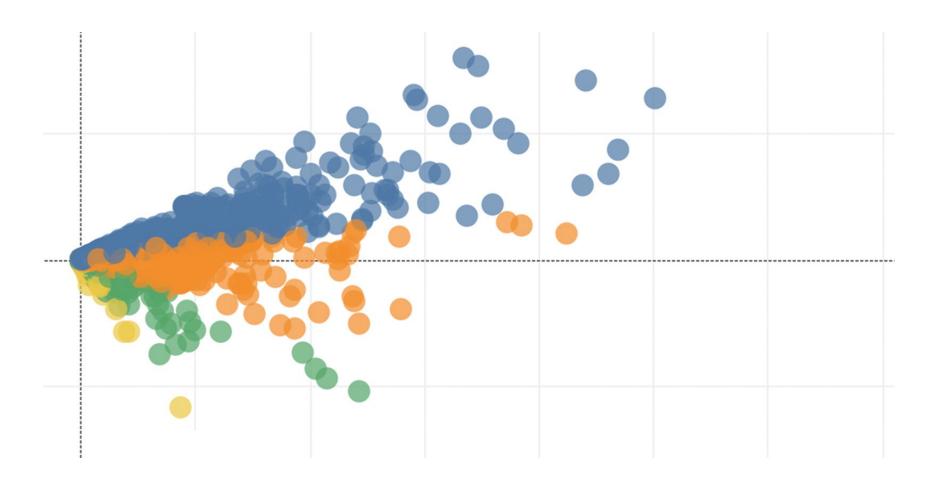


기능 – 크로스 데이터베이스 조인





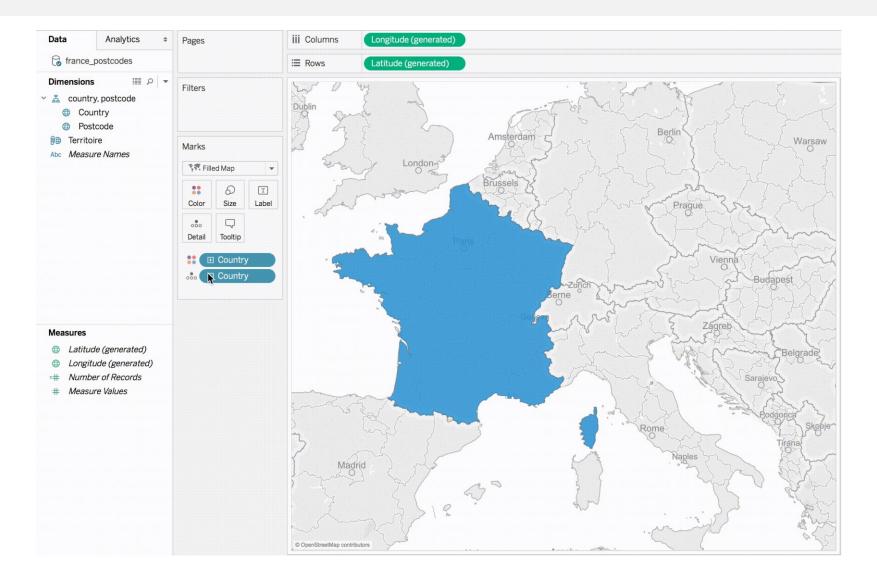
기능 - 클러스터링



태블로 10에서 추가된 점

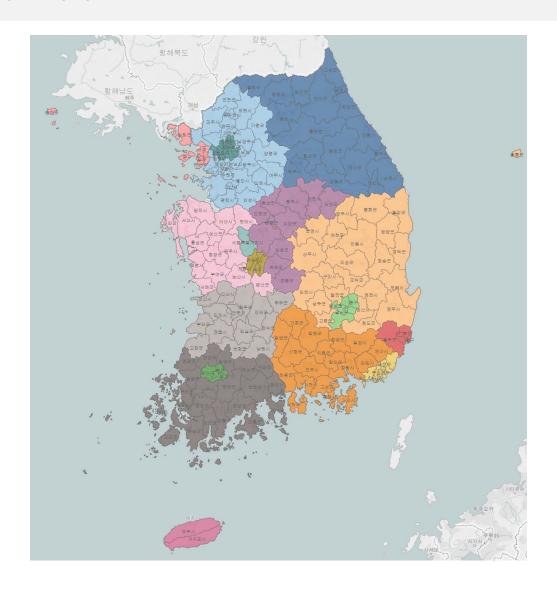


기능 – 사용자 지정 지역



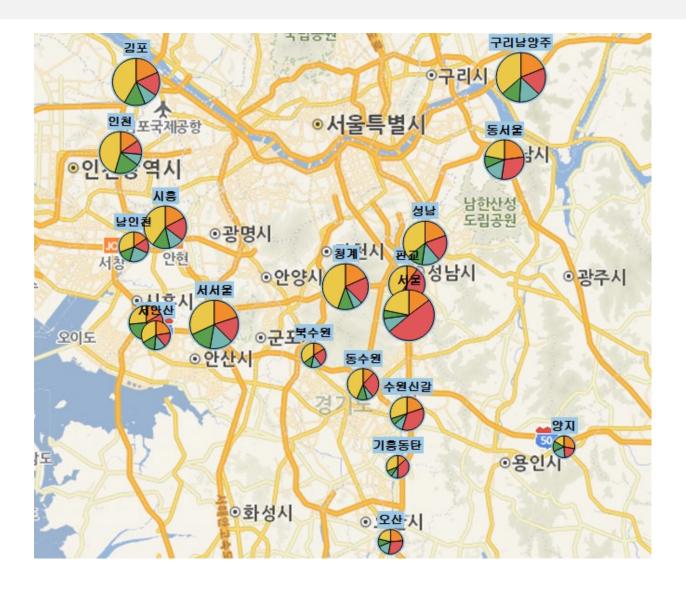


기능 – 사용자 지정 지역





국토부 v-world 지도





Thank You