

Tableau를 통한 데이터 시각화와 **분석**

2018.10.01 (2.5 hours)

소개

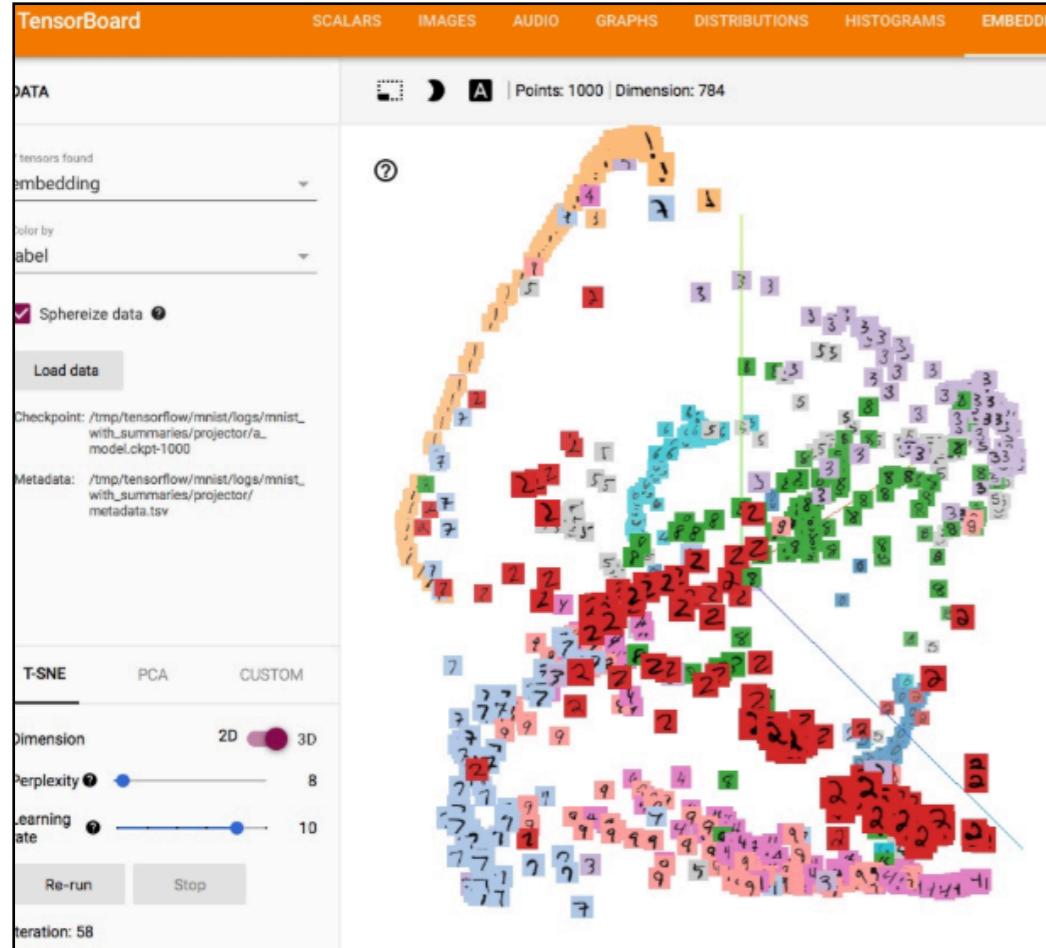
서울대학교 언론정보학과 박사과정, 이규호(HCI+D)

데이터 사이언스(저널리즘), 머신러닝(딥러닝), 조사방법론, 사회연결망

Python, R, Tableau/d3.js(Visualization), AMOS(SEM), Gephi(Network)

0. 데이터 분석과 TABLEAU

데이터 과학(머신러닝)/저널리즘/사회과학(연구)?



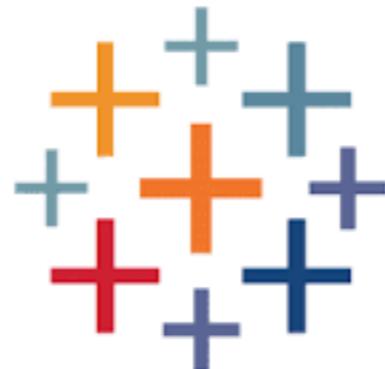
<Google Tensorflow>

<New York Times, The Upshot>

분야는 다르지만 둘 다 시각화를 적절히 사용합니다

1) 데이터 한눈에 파악 2) 중요한 메시지 강조

TABLEAU?



Google Data Studio

Business Intelligence Software

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

TABLEAU FOR FREE?

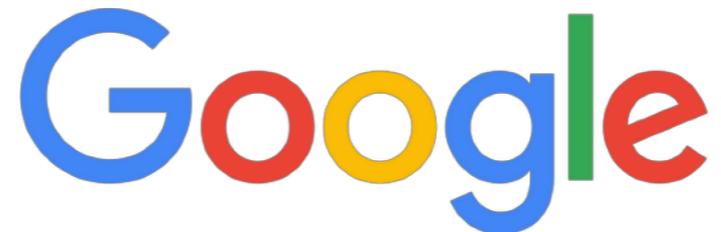


Tableau는 비교적 고가의 소프트웨어
하지만 무료로 Public 버전 사용가능(저장기능제한)
학생/교직원 -> 1년 라이센스 무료(서류 인증 필요, 영어 서류 스캔해서 제출)

TABLEAU학습 준비물



Tableau Desktop(체험판)
Tableau/Public 계정



Google 계정
Google Drive활성화



자신감

1. 데이터 기초 개념

데이터 종류 확인하기

This is what your data should look like

Tidy data = easy analysis

For best success with Tableau, your data should be formatted like a table or spreadsheet as seen here. If your data needs to be prepped before you use it, read on for details on Tableau's built-in tools to help.

A	B	C	D	E	F
Row ID	Order ID	Order Date	Order Priority	Sales	Ship Date
1	CAAB10015140	11/20/16	High	\$221.98	11/22/16
2	ESKM1637548	2/7/16	Critical	\$3,709.40	2/17/16
3	SGRH9495111	11/14/16	Medium	\$5,175.17	10/27/16
4	INJM156557	7/7/16	Medium	\$2,892.51	2/9/16
5	INTS2134092	11/16/14	Critical		
6	INMB1808592	4/24/15	Critical		
7	CAAB10015140	11/20/16	High	\$5,244.84	4/28/15
8	36258	3/16/14	High	\$341.96	11/22/16
9	36259	2/16/14	High	\$48.71	3/17/14
10	28879	11/23/15	High	\$17.94	3/17/14
11	45794	6/15/16	High	\$4,626.15	5/2/15
12	4132	MXVFZ171518	Critical	\$2,616.96	1/7/15
13	27704	INPF1912027	Critical	\$2,221.80	11/23/15
14			Critical	\$3,701.52	6/17/16

Data should be in a spreadsheet format with rows and columns

Each column represents one attribute of the data

Data is row level, one row represents one item

Tableau에서는 주로 Tidy Data를 사용합니다

TIDY DATA?

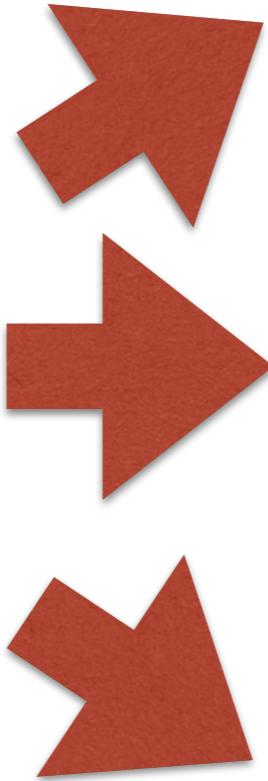


Hadley Wickham
(dplyr, reshape2, ggplot2)

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

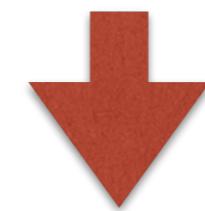
잘생김(?)

...

간단히 말해 봅시다

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

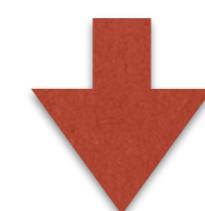
세로에는 서로 다른 특징이 들어갑니다



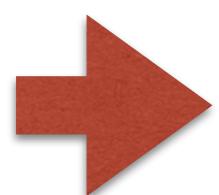
특징 1



특징 2



특징 3



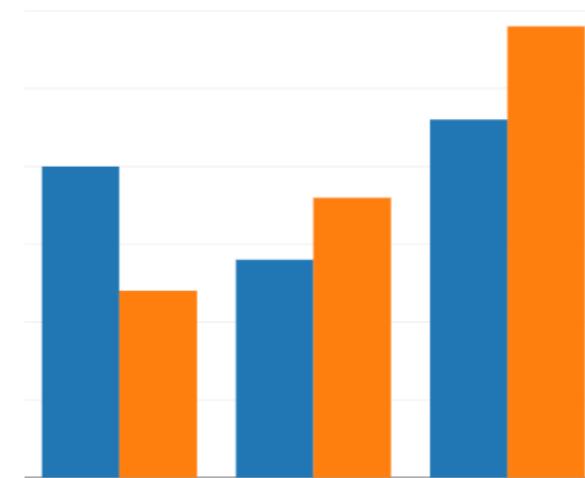
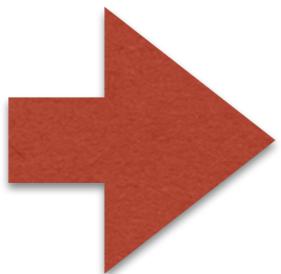
가로에는 각자 다른 조사 대상이 들어갑니다
(e.g. 영희, 민수…/학생1, 학생2/동물1, 동물2…)

TABLEAU와 TIDY DATA의 관계는?

특징

인원		측정값	
1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16
17	18	19	20
21	22	23	24
25	26	27	28
29	30	31	32
33	34	35	36
37	38	39	40
41	42	43	44
45	46	47	48
49	50	51	52
53	54	55	56
57	58	59	60
61	62	63	64
65	66	67	68
69	70	71	72
73	74	75	76
77	78	79	80
81	82	83	84
85	86	87	88
89	90	91	92
93	94	95	96
97	98	99	100
101	102	103	104
105	106	107	108
109	110	111	112
113	114	115	116
117	118	119	120
121	122	123	124
125	126	127	128
129	130	131	132
133	134	135	136
137	138	139	140
141	142	143	144
145	146	147	148
149	150	151	152
153	154	155	156
157	158	159	160
161	162	163	164
165	166	167	168
169	170	171	172
173	174	175	176
177	178	179	180
181	182	183	184
185	186	187	188
189	190	191	192
193	194	195	196
197	198	199	200
201	202	203	204
205	206	207	208
209	210	211	212
213	214	215	216
217	218	219	220
221	222	223	224
225	226	227	228
229	230	231	232
233	234	235	236
237	238	239	240
241	242	243	244
245	246	247	248
249	250	251	252
253	254	255	256
257	258	259	260
261	262	263	264
265	266	267	268
269	270	271	272
273	274	275	276
277	278	279	280
281	282	283	284
285	286	287	288
289	290	291	292
293	294	295	296
297	298	299	300
301	302	303	304
305	306	307	308
309	310	311	312
313	314	315	316
317	318	319	320
321	322	323	324
325	326	327	328
329	330	331	332
333	334	335	336
337	338	339	340
341	342	343	344
345	346	347	348
349	350	351	352
353	354	355	356
357	358	359	360
361	362	363	364
365	366	367	368
369	370	371	372
373	374	375	376
377	378	379	380
381	382	383	384
385	386	387	388
389	390	391	392
393	394	395	396
397	398	399	400
401	402	403	404
405	406	407	408
409	410	411	412
413	414	415	416
417	418	419	420
421	422	423	424
425	426	427	428
429	430	431	432
433	434	435	436
437	438	439	440
441	442	443	444
445	446	447	448
449	450	451	452
453	454	455	456
457	458	459	460
461	462	463	464
465	466	467	468
469	470	471	472
473	474	475	476
477	478	479	480
481	482	483	484
485	486	487	488
489	490	491	492
493	494	495	496
497	498	499	500
501	502	503	504
505	506	507	508
509	510	511	512
513	514	515	516
517	518	519	520
521	522	523	524
525	526	527	528
529	530	531	532
533	534	535	536
537	538	539	540
541	542	543	544
545	546	547	548
549	550	551	552
553	554	555	556
557	558	559	560
561	562	563	564
565	566	567	568
569	570	571	572
573	574	575	576
577	578	579	580
581	582	583	584
585	586	587	588
589	590	591	592
593	594	595	596
597	598	599	600
601	602	603	604
605	606	607	608
609	610	611	612
613	614	615	616
617	618	619	620
621	622	623	624
625	626	627	628
629	630	631	632
633	634	635	636
637	638	639	640
641	642	643	644
645	646	647	648
649	650	651	652
653	654	655	656
657	658	659	660
661	662	663	664
665	666	667	668
669	670	671	672
673	674	675	676
677	678	679	680
681	682	683	684
685	686	687	688
689	690	691	692
693	694	695	696
697	698	699	700
701	702	703	704
705	706	707	708
709	710	711	712
713	714	715	716
717	718	719	720
721	722	723	724
725	726	727	728
729	730	731	732
733	734	735	736
737	738	739	740
741	742	743	744
745	746	747	748
749	750	751	752
753	754	755	756
757	758	759	760
761	762	763	764
765	766	767	768
769	770	771	772
773	774	775	776
777	778	779	780
781	782	783	784
785	786	787	788
789	790	791	792
793	794	795	796
797	798	799	800
801	802	803	804
805	806	807	808
809	810	811	812
813	814	815	816
817	818	819	820
821	822	823	824
825	826	827	828
829	830	831	832
833	834	835	836
837	838	839	840
841	842	843	844
845	846	847	848
849	850	851	852
853	854	855	856
857	858	859	860
861	862	863	864
865	866	867	868
869	870	871	872
873	874	875	876
877	878	879	880
881	882	883	884
885	886	887	888
889	890	891	892
893	894	895	896
897	898	899	900
901	902	903	904
905	906	907	908
909	910	911	912
913	914	915	916
917	918	919	920
921	922	923	924
925	926	927	928
929	930	931	932
933	934	935	936
937	938	939	940
941	942	943	944
945	946	947	948
949	950	951	952
953	954	955	956
957	958	959	960
961	962	963	964
965	966	967	968
969	970	971	972
973	974	975	976
977	978	979	980
981	982	983	984
985	986	987	988
989	990	991	992
993	994	995	996
997	998	999	1000

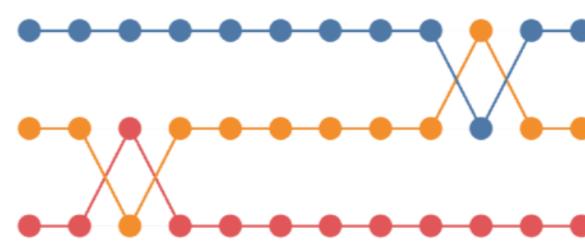
**특징을 섞고, 측정값을 넣으면
시각화가 됩니다**



세로막대차트

특집 : 남녀 고/저학년

표현가 : 성저



순위차트

특집 · 팀 종료/시가(녀)

표현값 : 순위

1. 가볍게 시작하기

DATA!

Tableau에서는 다양한 형태의 데이터 편집 가능
(**CSV**, TSV, **EXCEL**, SPSS, SQL…)

연결(2종류 존재)



파일에서 가져오기

Tableau Server	Spark SQL
Amazon Aurora	Teradata
Amazon EMR	웹 데이터 커넥터
Amazon Redshift	
Anaplan	
Cisco Information Server	
Cloudera Hadoop	
EXASolution	
Firebird	
Google Analytics	
Google BigQuery	
Google Cloud SQL	
Google 스프레드시트	
Hortonworks Hadoop Hive	
FIR Verica	
Kognitio	
MapR Hadoop Hive	
Marketo	
MemSQL	
Microsoft SQL Server	
MySQL	
OData	
Oracle	
Oracle Eloqua	
Pivotal Greenplum Database	
PostgreSQL	
Presto	
QuickBooks Online	
Salesforce	
SAP HANA	
Snowflake	

온라인에서 가져오기

추출? 라이브?

과거 데이터, 혼자 작업 - 추출연결(한번 불러오면 바뀌지 않습니다)
협업, 실시간 데이터 - 라이브 연결(원래 파일이 바뀌면 바뀝니다)

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

연결 추가

subway
텍스트 파일

파일 ⌂

subway.csv

새 Union

필드 정렬 수정한 날짜					
subway.csv 사용일자	subway.csv 노선명	# subway.csv 역ID	Abc subway.csv 역명	# subway.csv 승차총승객수	# subway.csv 하차총승객수
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

데이터 조합?

유니언(행 추가)

조인(열 추가)

교차 DB조인

데이터 블렌딩

Tableau에서는 크게 4가지 조합 지원

오늘은 “조인/유니언”만 배움(나머지는 엑셀/구글 시트로도 충분)

차트 작성하기(1)

데이터 분석

subway

차원

노선명 등록일자 사용일자 역ID 역명 측정값 이름

측정값

-# 격차별이동인구
#+ 승차총승객수
#+ 하차총승객수
#+ 레코드 수
#+ 측정값

→

열	노선명	월(사용일자)
행	합계(승차총승객수)	

데이터 편집/속성변경후

조합

문자, 색의 차이는?

앞의 기호(데이터 형태)/색상(데이터 연속성)
서로 연결된 데이터 형식은? 그룹화/계층화

차트 작성하기[2]



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

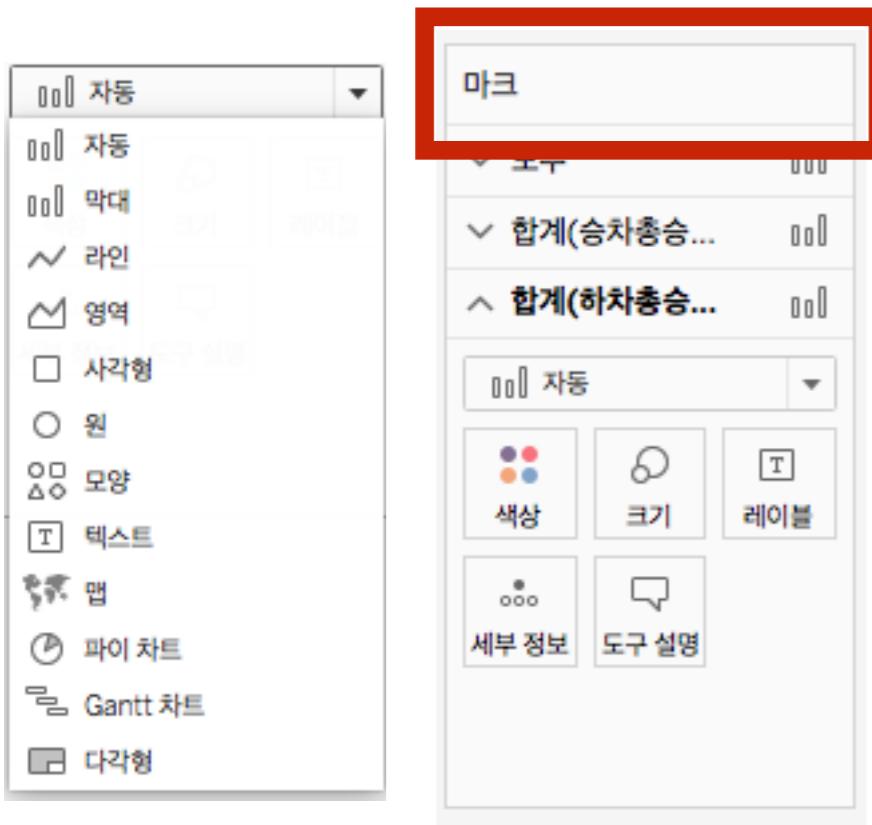
디테일 수정

차원

측정값

표현형식

디테일 수정(측정값/표현방식)

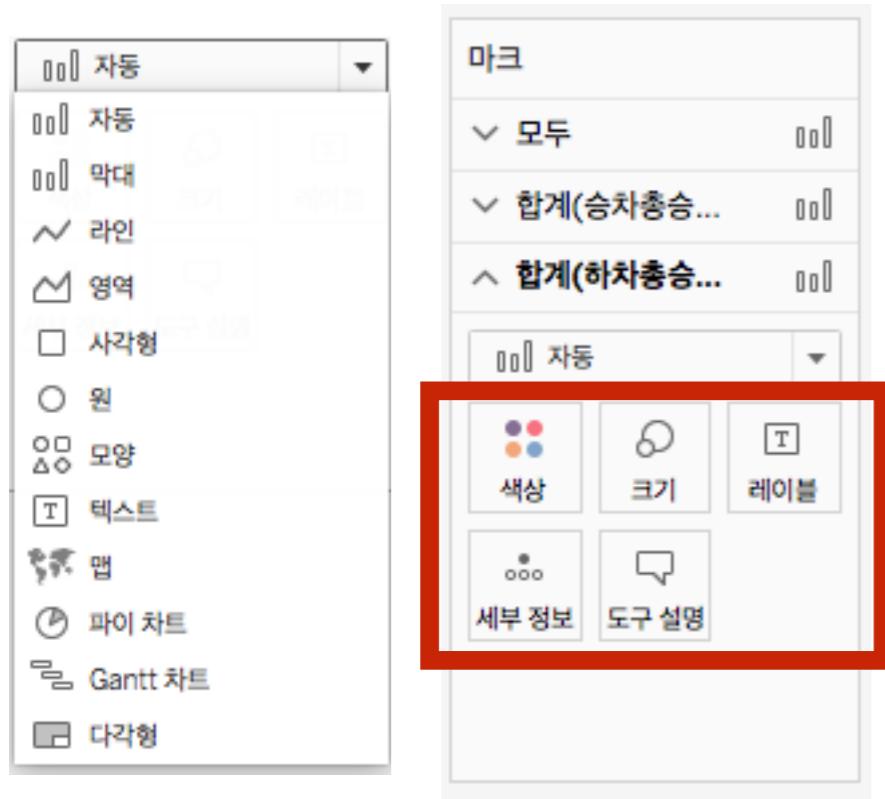


마크(표현 방식 세부 조절 가능)

마크의 조합을 통해 “복잡한”
그래프 생성 가능

새로운 측정값이 추가될때마다 Stack

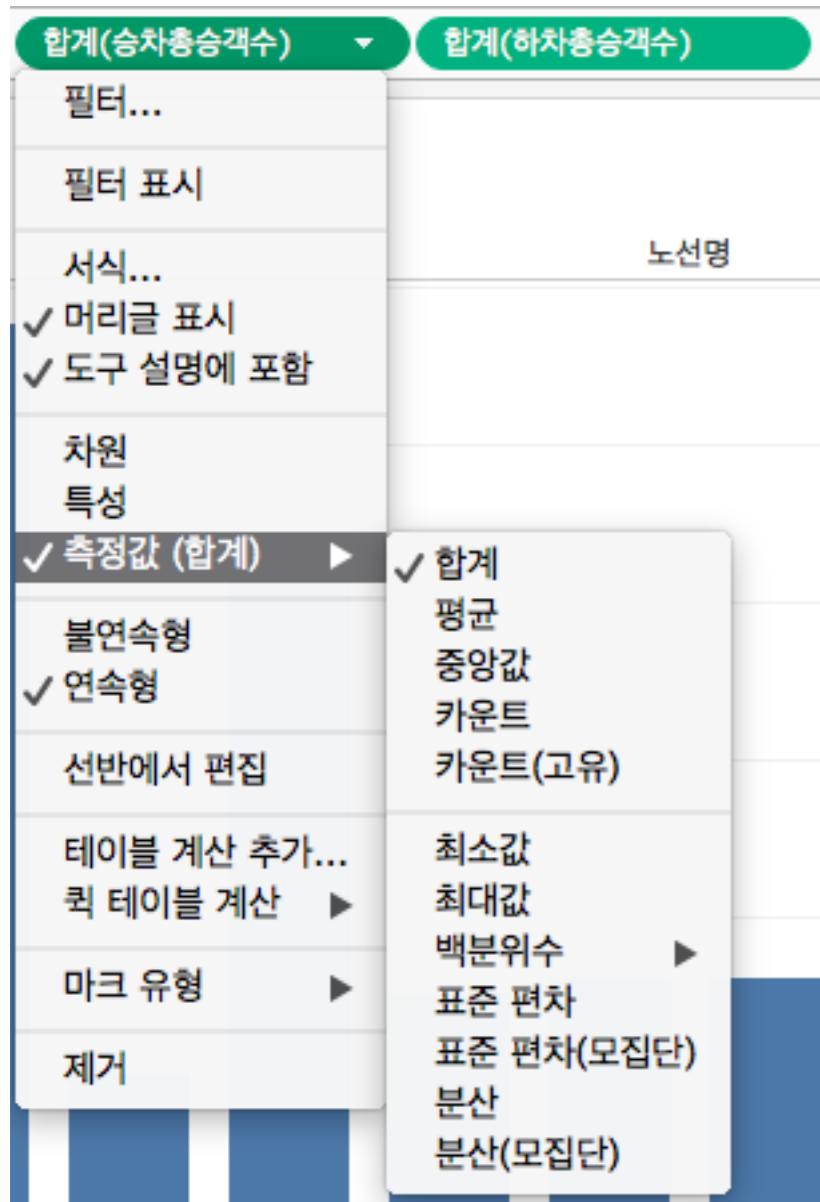
디테일 수정(측정값/표현방식)



세부옵션

(세밀한 조절은
아래 Box의 조합으로 가능)

디테일 수정(측정값/표현방식)



내가 원하는 형태의 측정값을 사용 가능

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

디테일 수정(측정값/표현방식)

카운트 vs 카운트(고유) = “select” vs “select distinct”

“카운트”: 몇개나 있나?, “카운트(고유)”: 몇 종류나 있나?

2. 데이터 분석

시각화의 2가지 방향

데이터 탐색

분석 결과 정리

TABLEAU 에서 지원하는 분석 방법들(그룹분석)

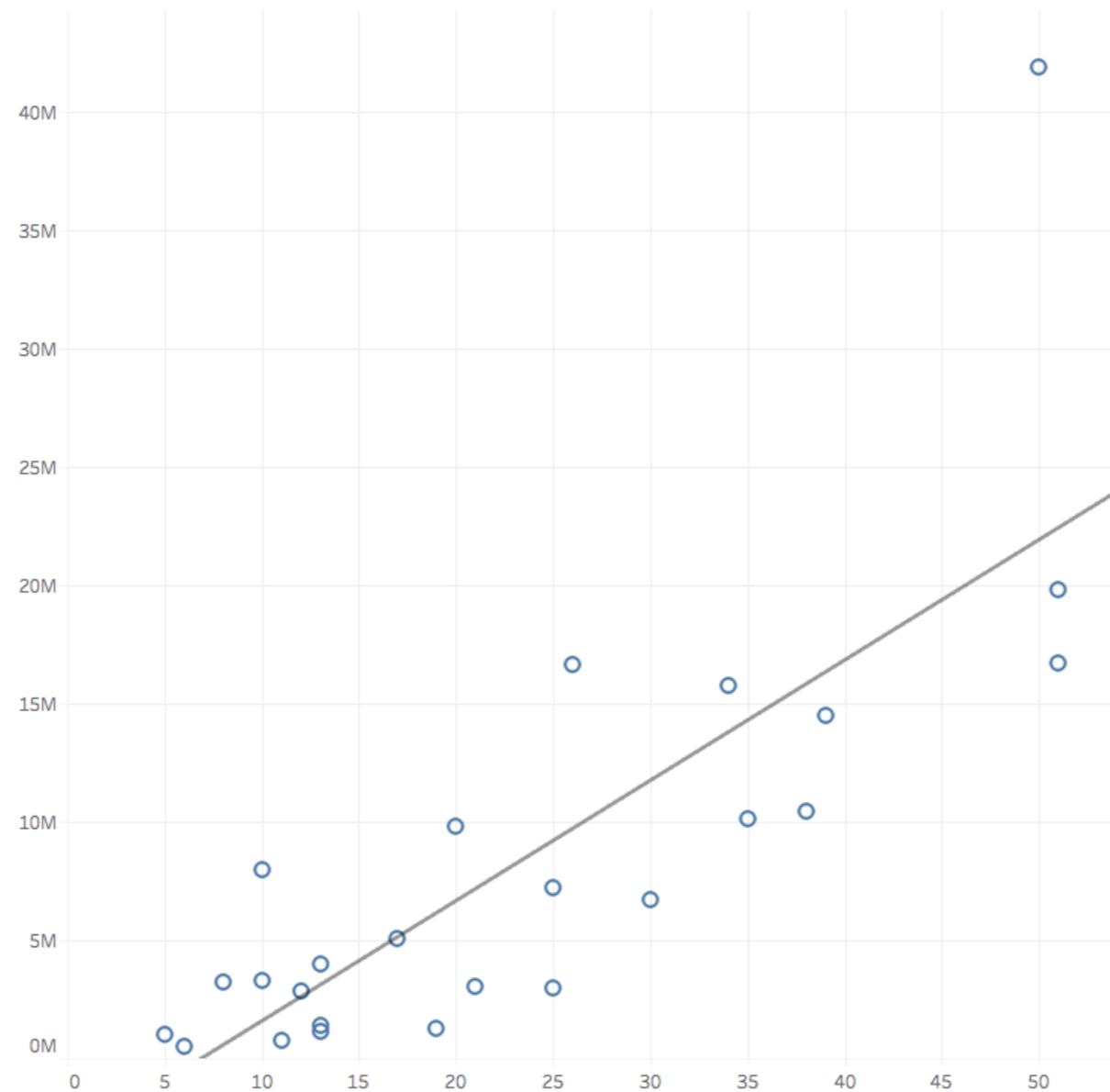
일반 통계
(평균/분포
사분위)

회귀분석
(추세분석)

집단분석
(클러스터링)

시나리오
(What-If?)

노선별 역수는 이용량과 상관관계가 있을까? [황새의 함정 - 인과와 상관의 위험성]



3. 데이터 전처리

데이터 전처리

RAW데이터를 “쓸만한” 데이터로 변환하기

DATA!

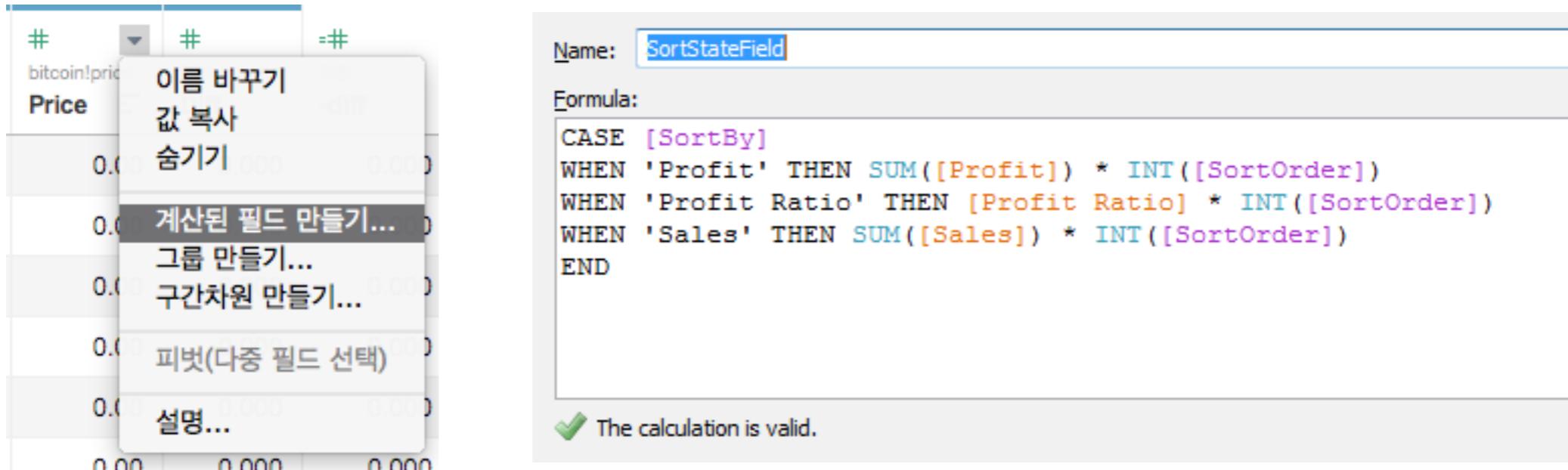
weather.go.kr
(기상청, 날씨누리 종합기상정보)

국내 지진 목록



지진 · 화산	관측자료	기후자료	생활과산업
발표정보			
지진 >		국내지진 목록	
지진해일 >		국외지진 목록	

자주 쓰는 전처리 1- 수식 계산



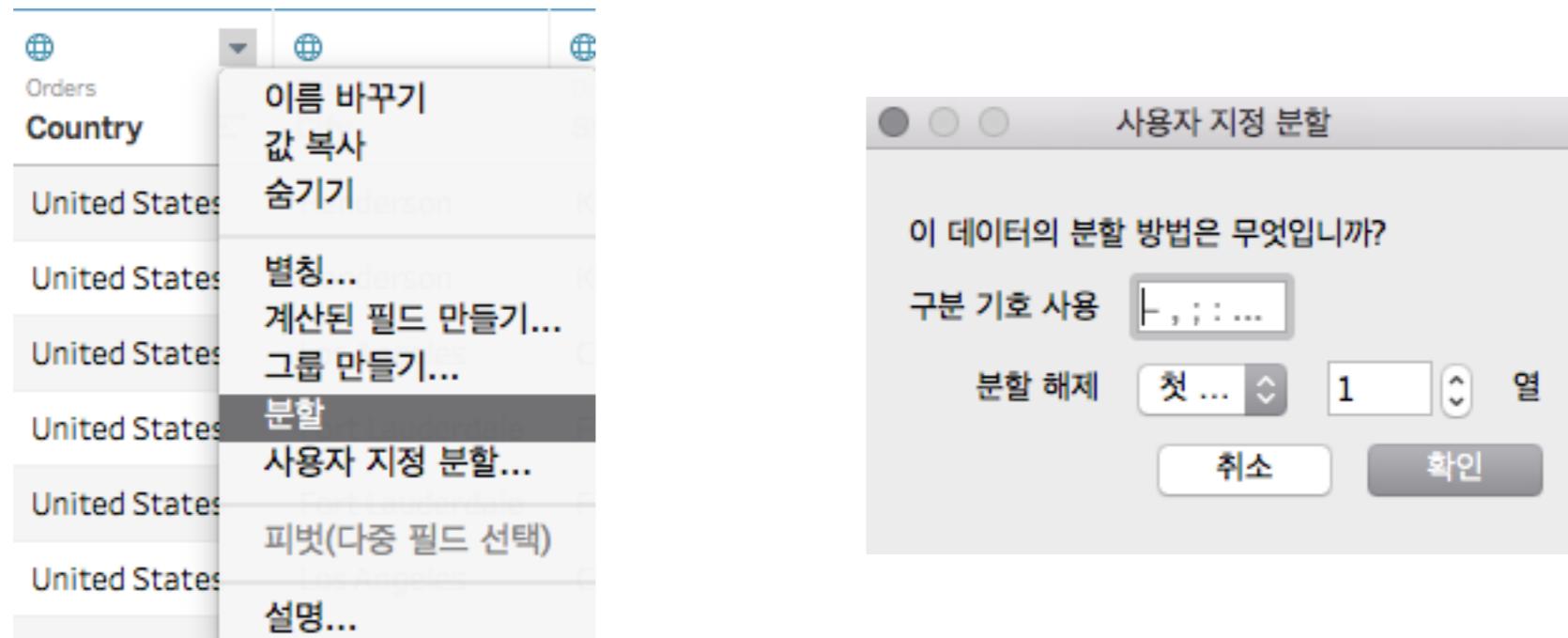
(능력이 된다면) 섬세한 작업도 가능
추천하지는 않음(대부분 Excel, Google Sheets로 처리 가능)
하나라도 “덜” 배우자!

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

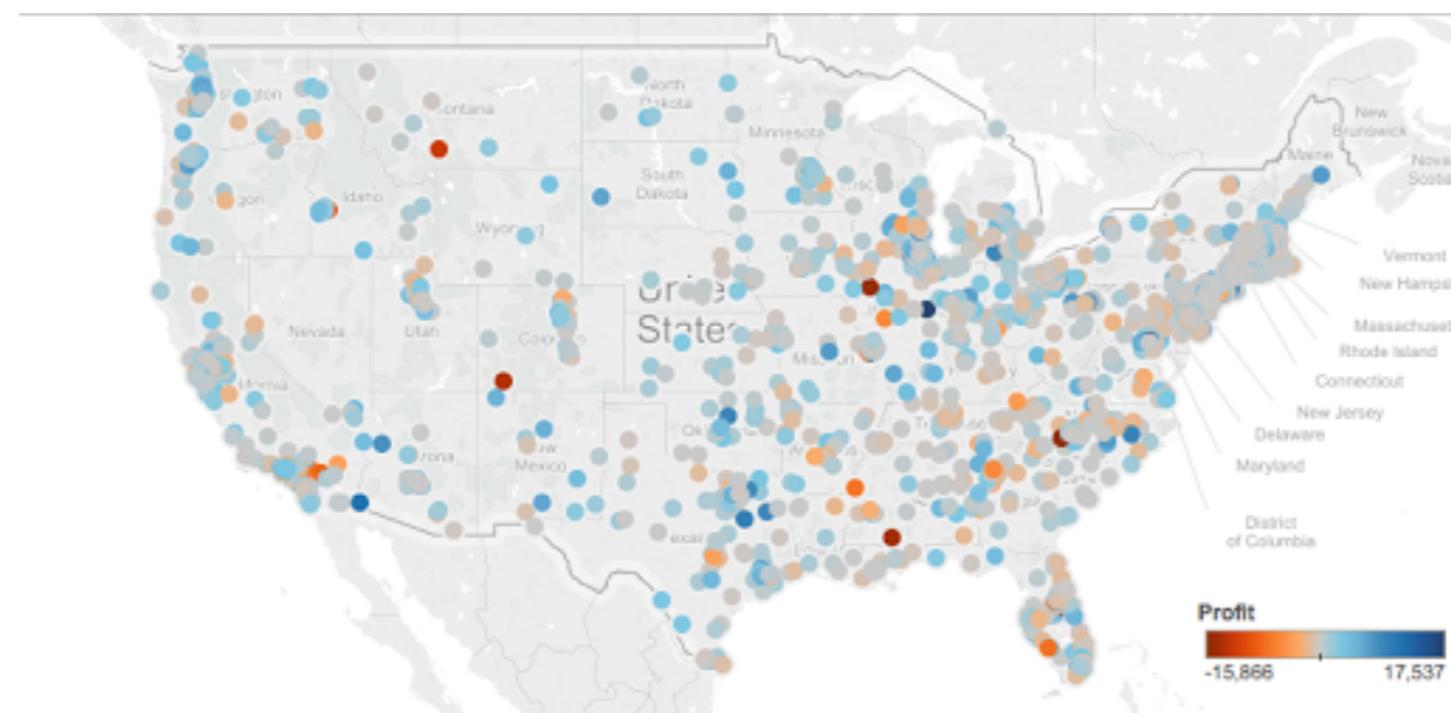
자주 쓰는 전처리 3 - 셀 나누기



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

4. 고급 데이터 시각화 (1) - MAPPING

MAPPING



지리 데이터를 표현하는 2가지 방법

좌표 or 지역명 (적절히 데이터 유형 바꾸기)

5. 고급 데이터 시각화 (2) - 텍스트, 필터/페이지

DATA!

https://ko.wikipedia.org/wiki/대한민국의_영화_흥행_기록

대한민국 영화 흥행 기록

흥행 순위 목록 [편집]

대한민국 내에서의 영화 흥행 기록이다. 표 작성의 출처 및 기준은 대한민국 영화진흥위원회의 영화관 입장권 통합전산망에 기록된 내용을 기반으로 한다. 만약 기록이 없다면 기사 등을 참고한다. 전체 영화 중에서 대한민국 내에서의 흥행 순으로 기록한다. 만약 두 통계에서 모두

순위	제목	감독	한국내 배급사
1	《명량》	김한민	CJ E&M
2	《신과함께: 죄와 벌》	김용화	롯데엔터테인먼트
3	《국제시장》	윤제균	CJ E&M
4	《아바타》	제임스 캐머런	이십세기폭스코리아

데이터 가져오기

Google Sheet 기능으로 최근 데이터 수집하기
(Tableau에서도 가능!)

필터/페이지

시각화에 인터랙티브 연동 효과 넣기(시나리오 분석)

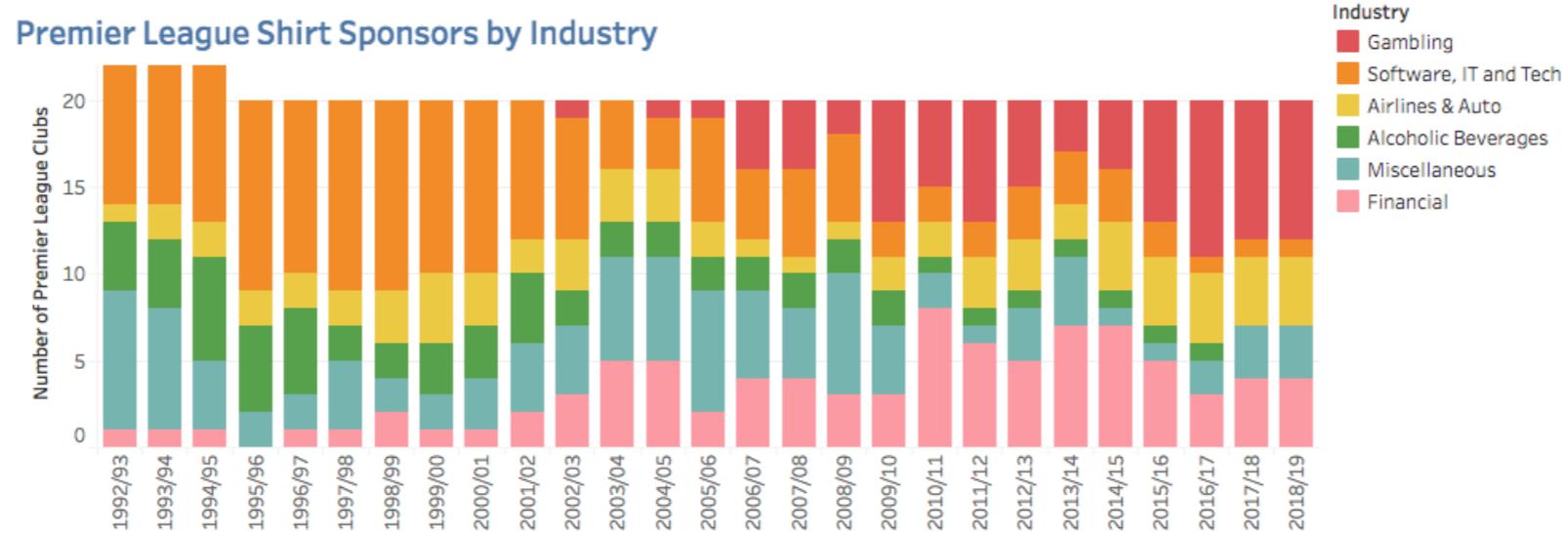
5. 인터랙티브 스토리 텔링(모아보기)

분석 결과 모아보기



워크시트 -> 대시보드 -> 스토리(데이터연동)

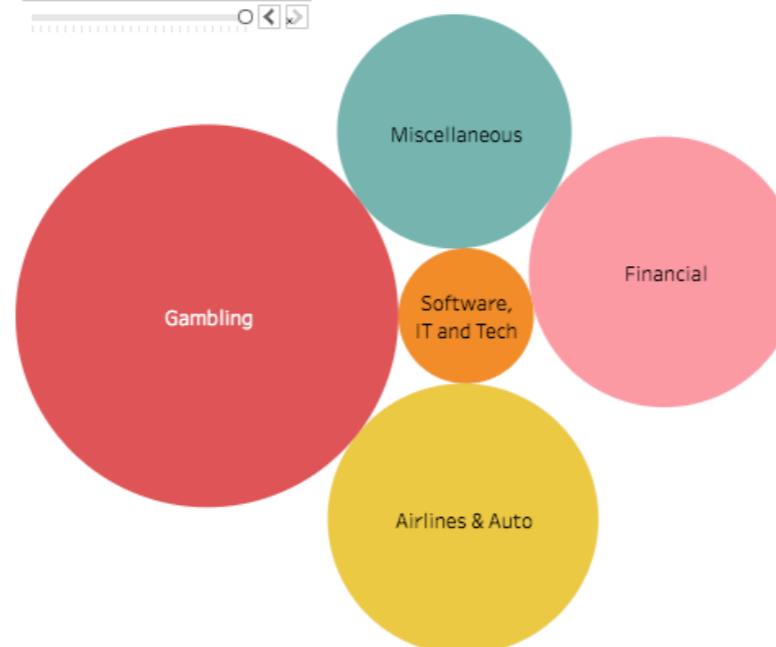
6. 데이터 / 좋은 영감 얻기



Evolution of Shirt Sponsors: 2018/19

(Move the cursor below to see different seasons)

Season
2018/19



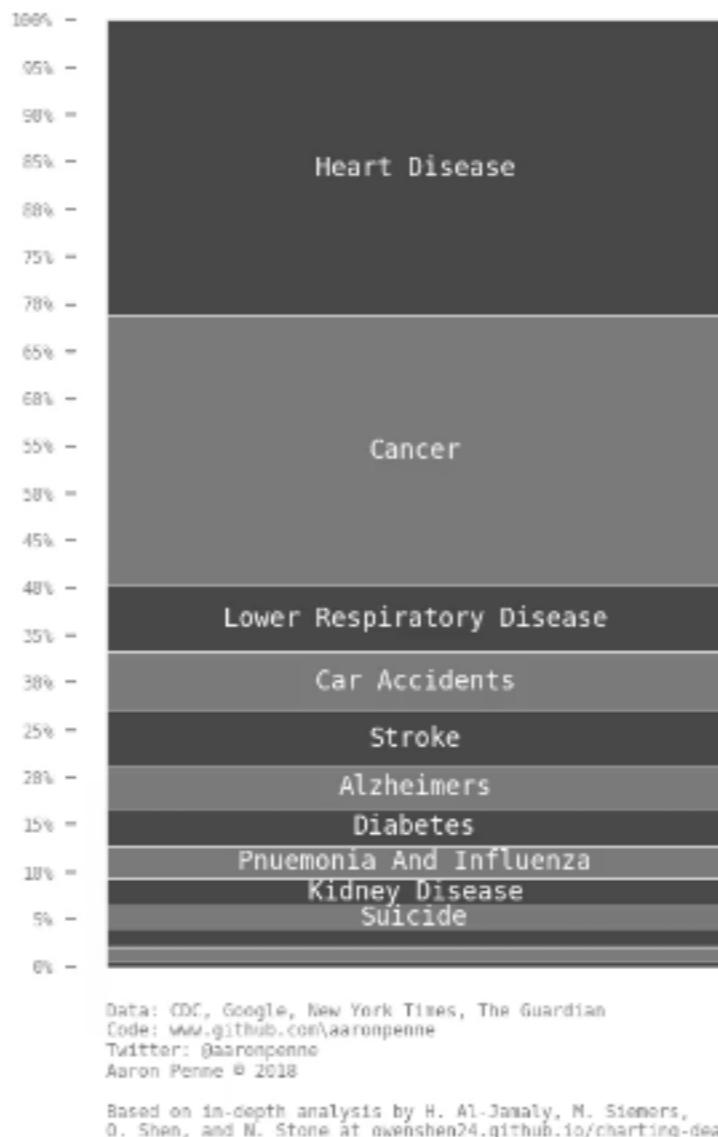
Show me my Club

(Select a club from the drop-down list)

Season	Club	Sponsor
1996/97	Arsenal	Software, IT and Tech
1997/98	Arsenal	Software, IT and Tech
1998/99	Arsenal	Software, IT and Tech
1999/00	Arsenal	Software, IT and Tech
2000/01	Arsenal	Software, IT and Tech
2001/02	Arsenal	Software, IT and Tech
2002/03	Arsenal	Software, IT and Tech
2003/04	Arsenal	Software, IT and Tech
2004/05	Arsenal	Software, IT and Tech
2005/06	Arsenal	Software, IT and Tech
2006/07	Arsenal	Airlines & Auto
2007/08	Arsenal	Airlines & Auto
2008/09	Arsenal	Airlines & Auto
2009/10	Arsenal	Airlines & Auto
2010/11	Arsenal	Airlines & Auto
2011/12	Arsenal	Airlines & Auto
2012/13	Arsenal	Airlines & Auto
2013/14	Arsenal	Airlines & Auto
2014/15	Arsenal	Airlines & Auto
2015/16	Arsenal	Airlines & Auto
2016/17	Arsenal	Airlines & Auto
2017/18	Arsenal	Airlines & Auto
2018/19	Arsenal	Airlines & Auto

https://public.tableau.com/profile/enochchan#!/vizhome/Shirt_Sponsors/ShirtSponsors

CDC Cause of Death in USA
"What actually causes death?"



https://g.redditmedia.com/x1cdV5AkvPOFUAMi1Lar_fQQcpG7VcWhUJvdibJPG0U.gif?fm=mp4&mp4-fragmented=false&s=9627a66446fff1a5dcee66a5cdc17c0b

시각화를 이용한 데이터 탐색 예제

Here's How America Uses Its Land
By Dave Merrill and Lauren Leatherby
July 31, 2018

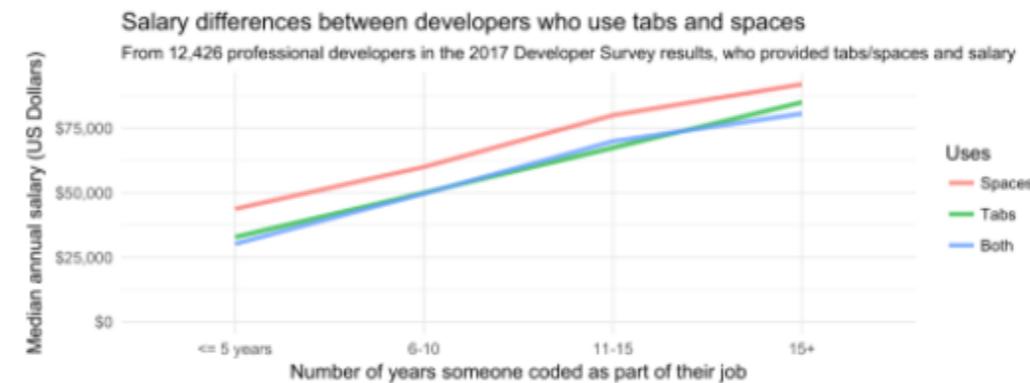
<https://www.bloomberg.com/graphics/2018-us-land-use/>

시각화를 이용한 데이터 탐색 예제

EVELINA GABASOVA'S BLOG

Tabs, spaces and your salary - how is it really?

A couple of days ago, David Robinson published an article on the Stack Overflow blog with a very provocative title: [Developers Who Use Spaces Make More Money Than Those Who Use Tabs](#). It uses the data from Stack Overflow developer survey to show that indeed, using spaces is associated with higher salaries, even when we account for experience level. So, should you start using spaces instead of tabs to increase your salary?



7. 온라인에 게시하기

분석 결과를 공유하기 위해서



Tableau Public(무료버전)

Tableau 계정

공유정신

주의!

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

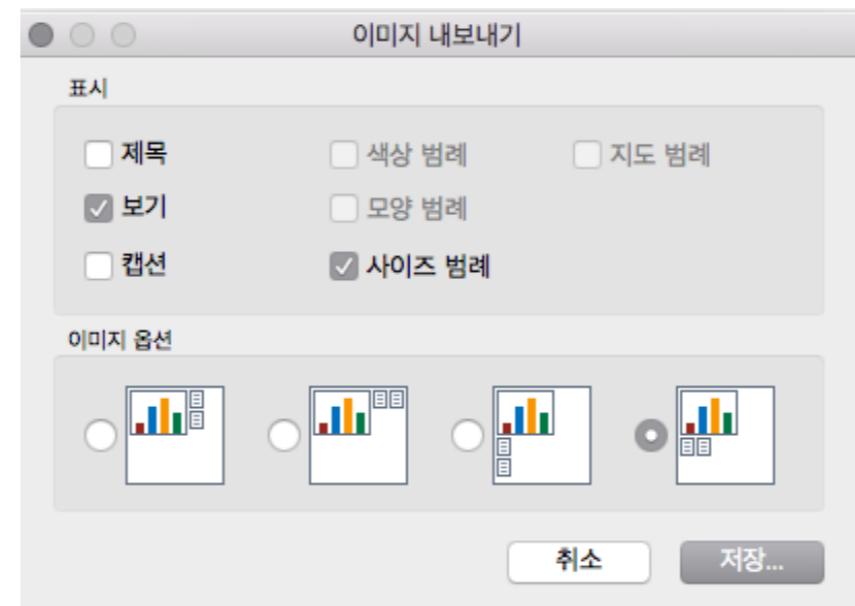
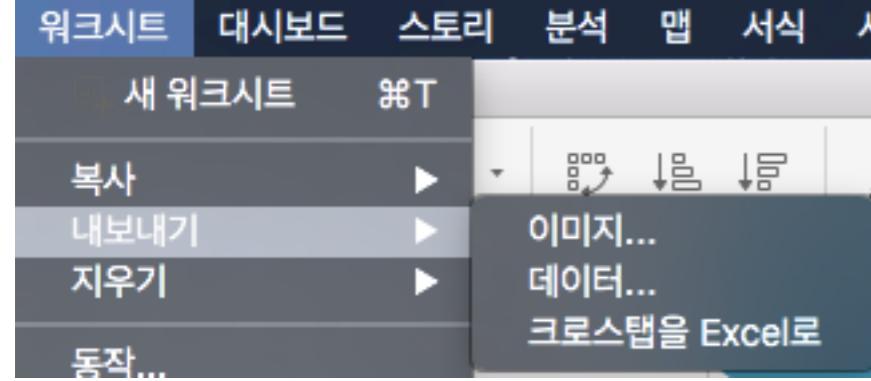
TABLEAU의 공유방법

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

TABLEAU문서 공유



twb파일 -> 핵심 요소만 추출(간략하게, 추가 수정 어려움)

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

TABLEAU 온라인 공유

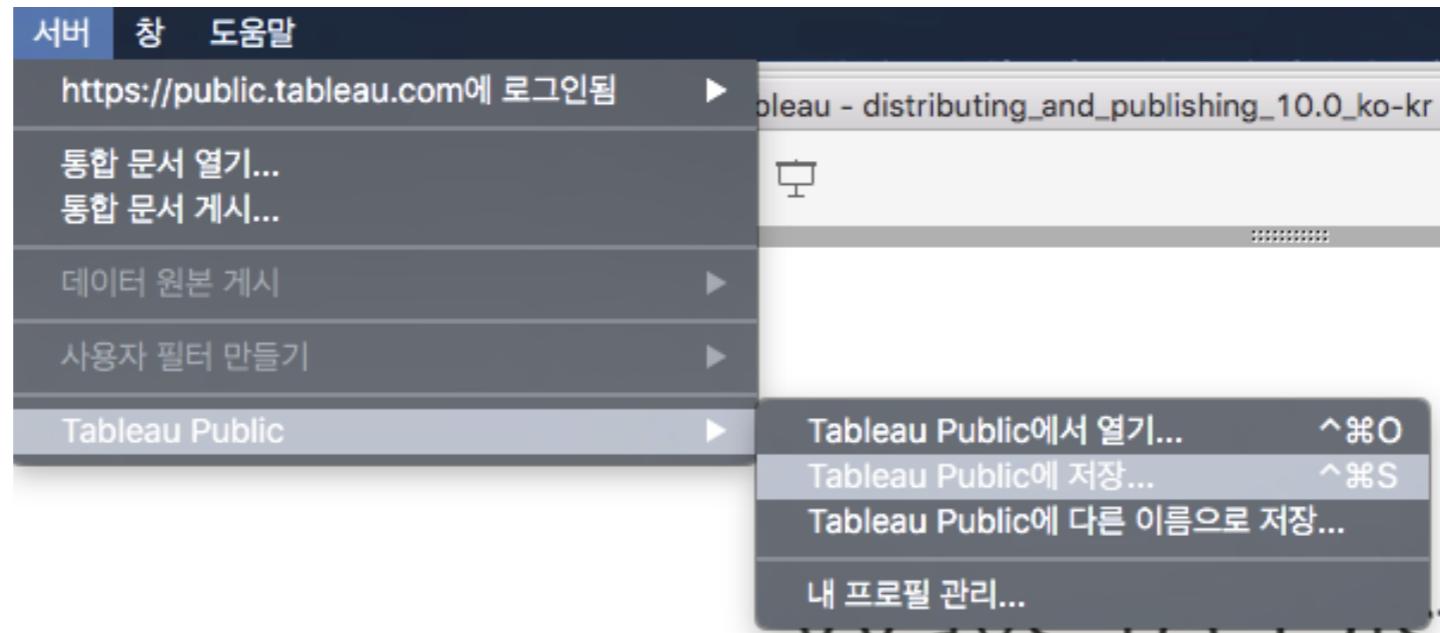


Tableau Public에서는 저장/내보내기가 무조건 서버 상에서만 가능
온라인에 저장하기 전 데이터 추출해야 함(내 컴퓨터에서 업로드가 이뤄지기 때문)
[데이터 추출은 데이터 -> 원하는 데이터 선택 -> 추출로 가능]

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 제목을 설정하세요.

고정 링크

URL 추가

Viz를 내장해 보십시오. 사이트로 트래픽을 유도하는 탁월한 방법입니다.

설명

testing_02

검색 결과에서 상위에 표시되려면 정확한 설명이 필요합니다.

툴바 설정

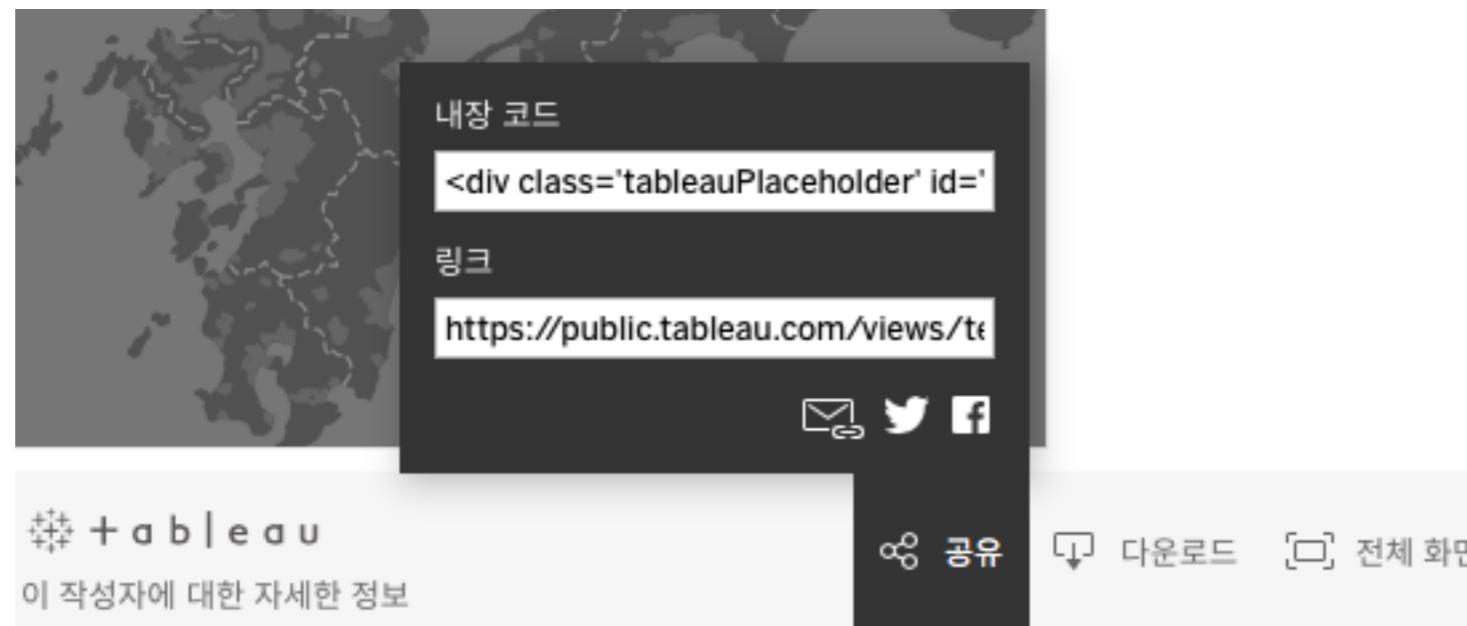
- 뷰 컨트롤 표시 실행 취소, 다시 실행, 되돌리기
- 작성자 프로필 링크 표시
- 통합 문서 및 해당 데이터를 다른 사용자가 다운로드할 수 있도록 허용

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