

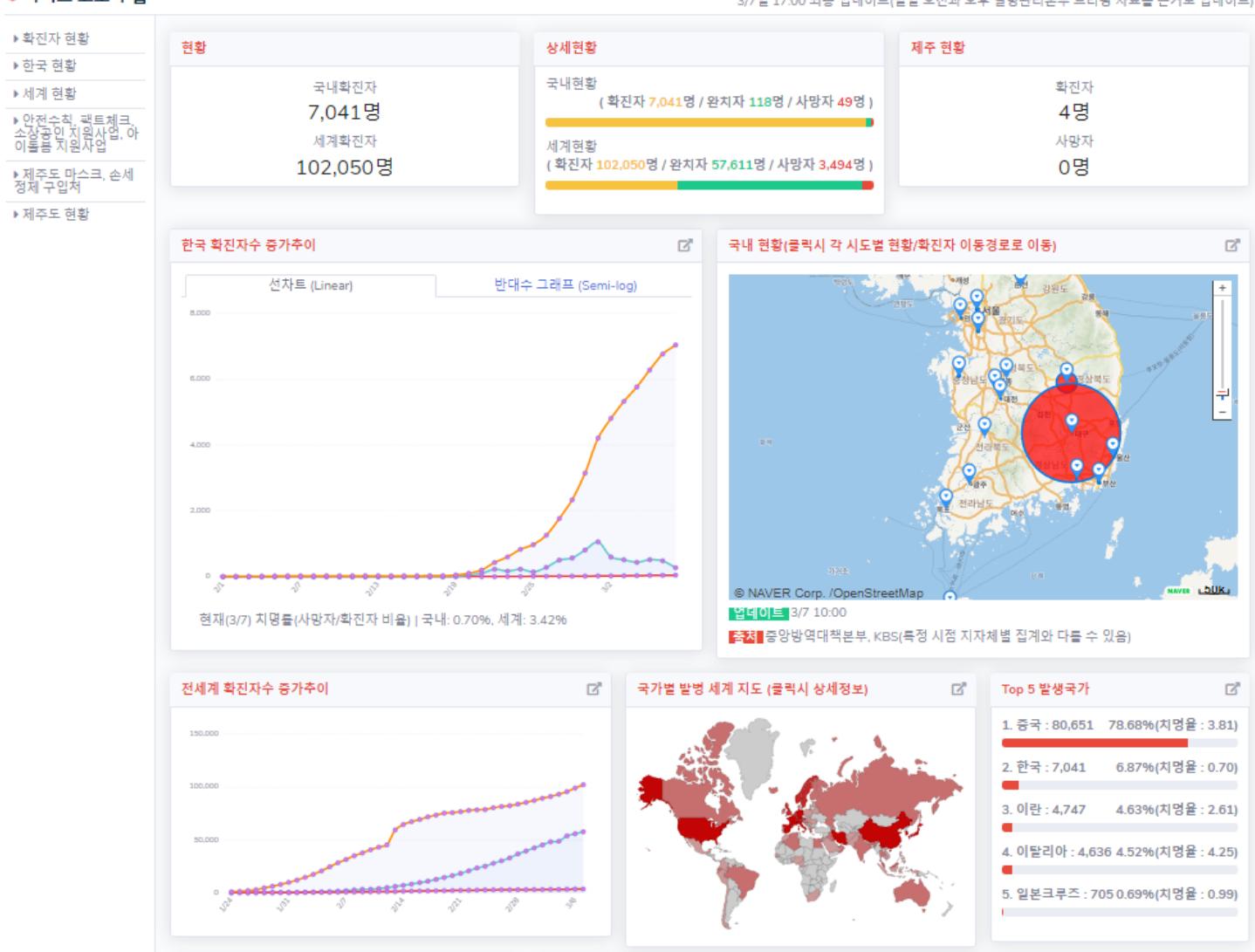
데이터시각화 강좌 소개

- 과목: 데이터시각화
- 교수: 신해웅(logistex@hywoman.ac.kr)

1. 데이터시각화?

1.1 코로나 맵

● 라이브 코로나 맵



1.2 공공데이터

데이터셋

제공신청

활용사례

정보공유

이용안내

/ 활용사례 / 공공데이터 시각화

활용사례

공공데이터 활용사례

기업탐방 인터뷰

공공데이터 시각화 >

국민참여지도

위치정보 시각화

공공데이터 분석서비스

공공데이터 시각화

공공데이터를 활용하여 다양한 시각화 차트를 만들어 공유할 수 있습니다.



교육 (242)



국토관리 (37)



공공행정 (192)



재정금융 (20)



산업고용 (22)



사회복지 (95)



총 822 건

등록일 ▼

~

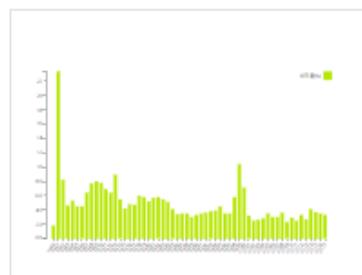
전체



전체

가이드 문서 다운로드

시각화 만들기 +



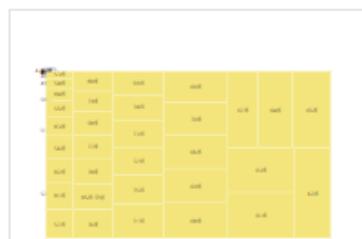
1960 ~ 2017 연도별 공무원 퇴직

제작자 : 공무원연금공단

메타정보

공공행정

2019-12-06



연령별 퇴직연금수급액 비교

제작자 : 공무원연금공단

메타정보

공공행정

2019-12-06



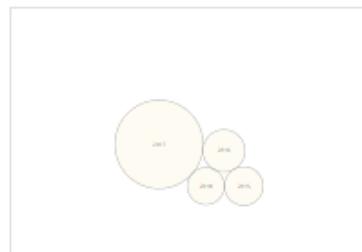
공무원연금 가입자 평균연령 비교

제작자 : 공무원연금공단

메타정보

공공행정

2019-12-06



연도별 공무원연금기금 기금회계

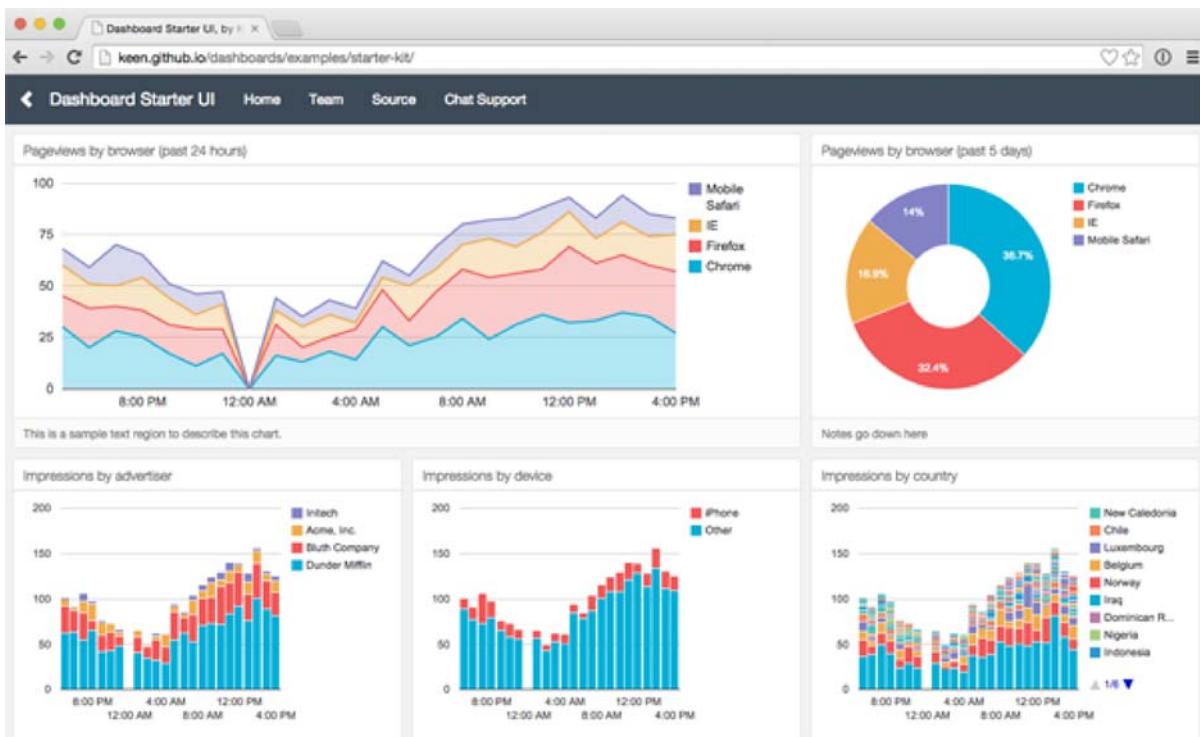


공무원연금 연도별 퇴직일시금 지급액



연도별 공무원연금기금 기금회계

1.3 대시보드



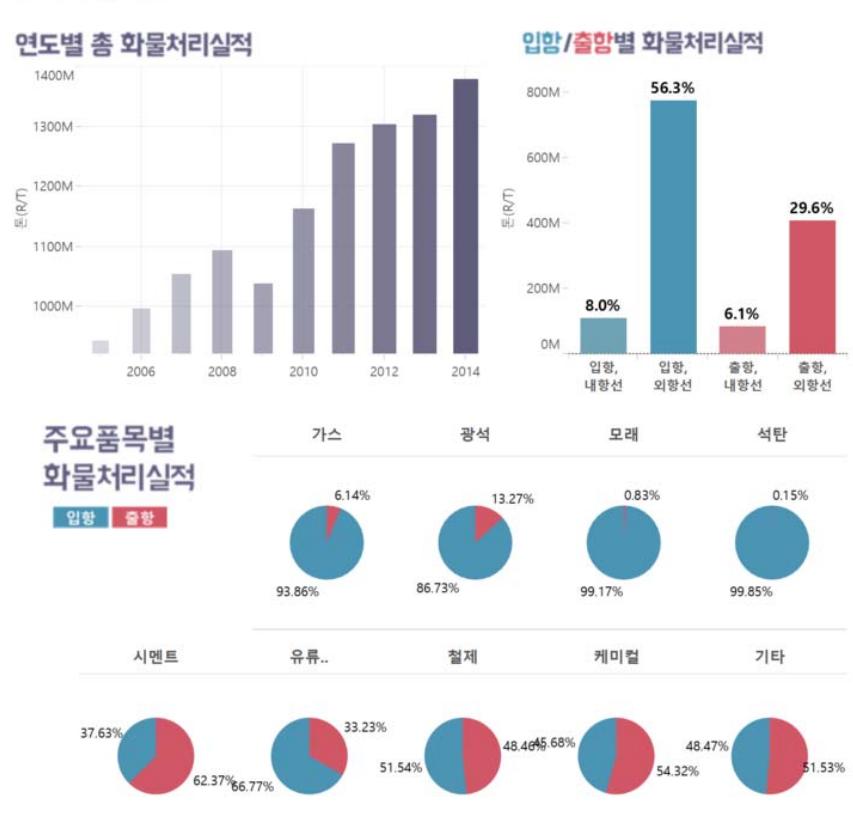
1.4 항만별 물동량

대한민국 항만별 화물처리실적 (2005-2014)

최근 10년 간 항만별 화물처리실적 누적 규모 비교



항만별&연도별 상세 데이터 비교



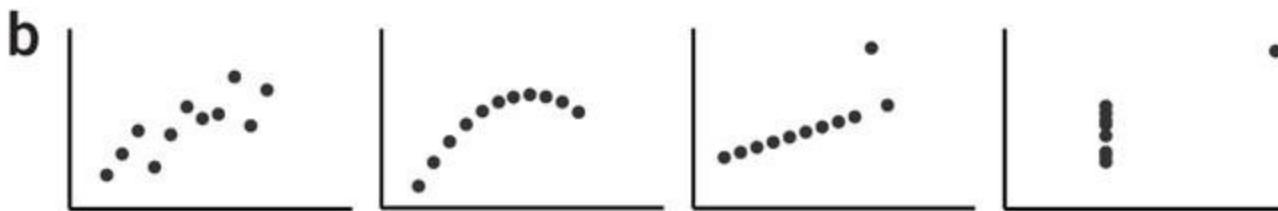
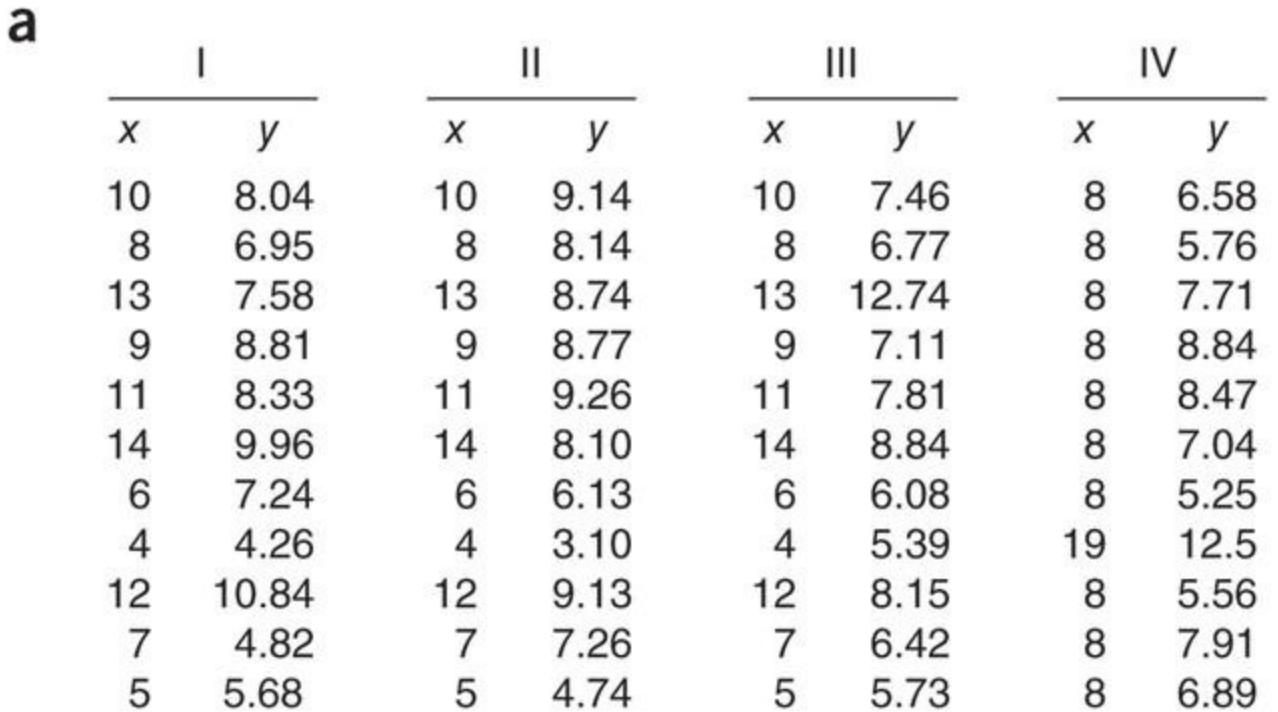
2. 왜 배워야 하나?

- 190122, 뉴스젤리, 당신도 데이터 시각화를 해야 하는 5가지 이유

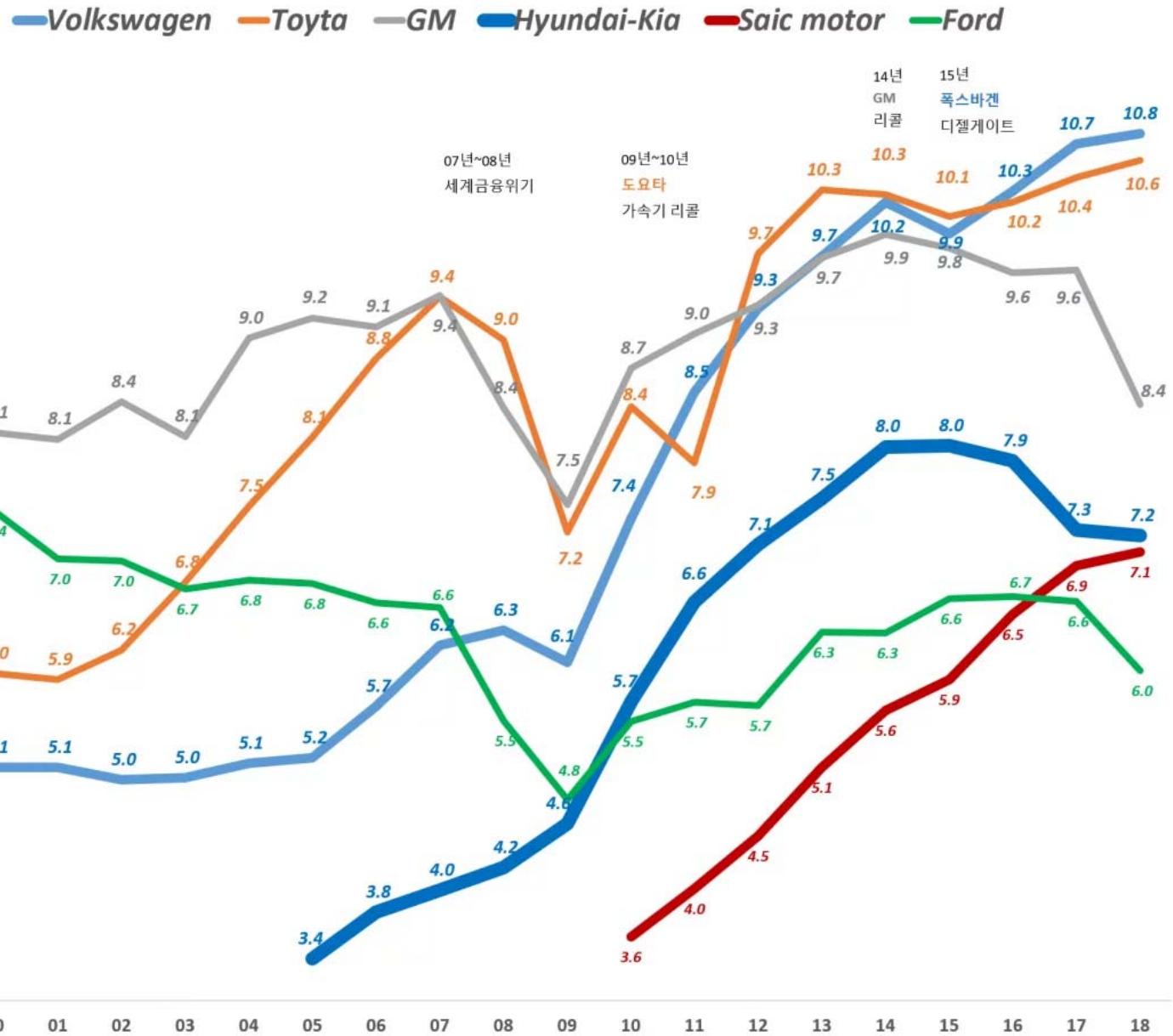
2.1 한눈에 보이는



엑셀 화면을 가득 채우고도 넘치는 양의 데이터를 간단한 선 차트로 요약하여 표현할 수 있습니다. (데이터 출처 : 서울 열린데이터 광장)

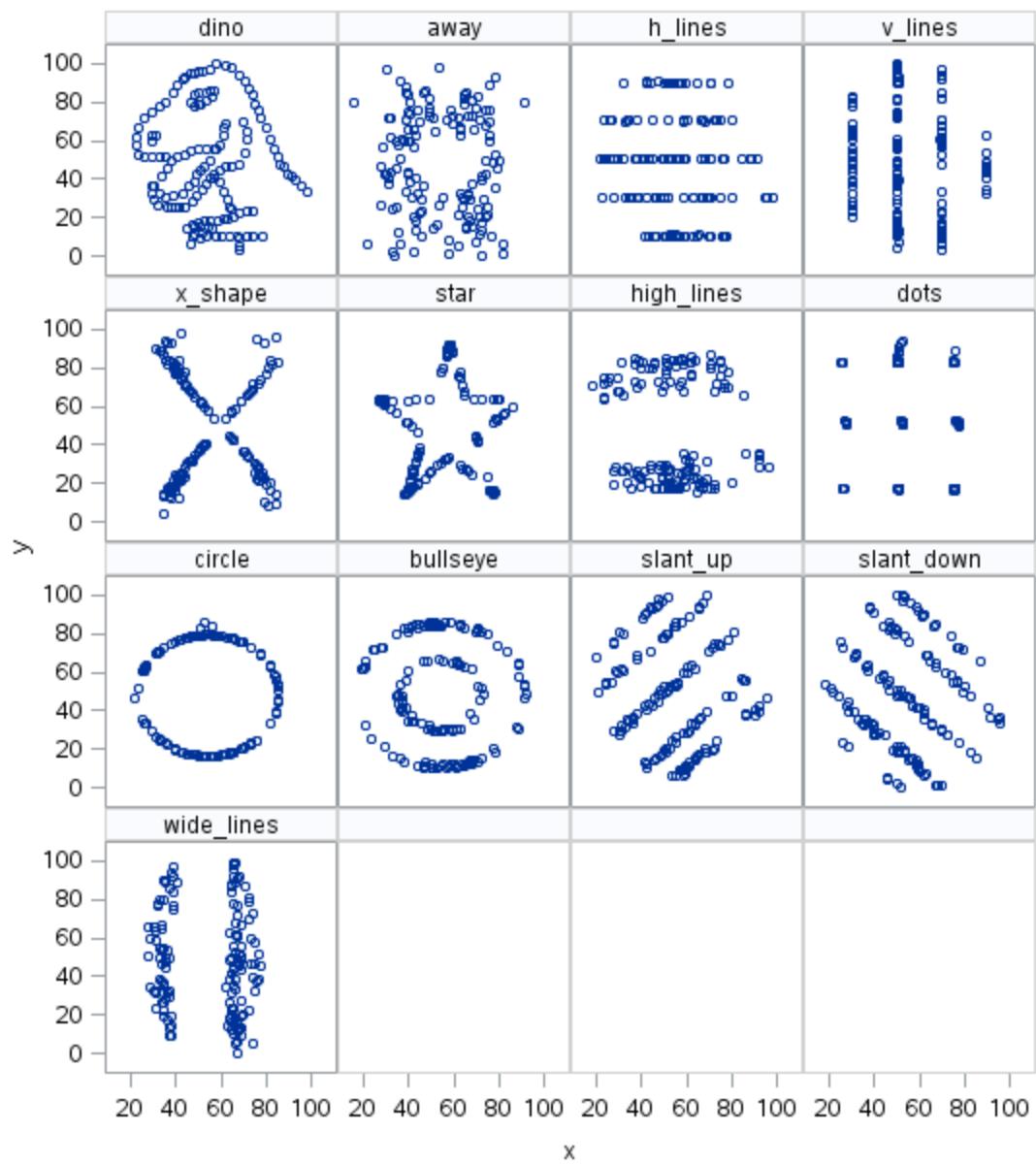


2.2 영감을 주는



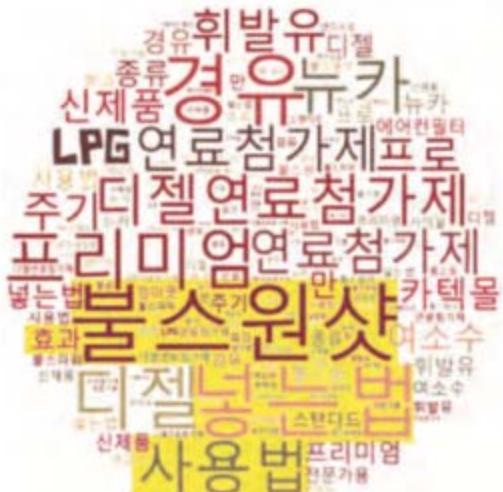
2.3 통계의 합정

THE DATASAURUS DOZEN



2.4 판단을 위해서

빅데이터를 활용해 기업 문제 해결한 사례



谤스원샷 연관 검색어 Top 20			
1	谤스원샷 넣는법	980	11
2	谤스원샷 강유	908	12
3	谤스원샷 70000	905	13
4	谤스원샷 효과	824	14
5	谤스원샷 다큘	717	15
6	谤스원샷 프리미엄	707	16
7	谤스원샷 뉴카	663	17
8	谤스원샷 사용법	536	18
9	다행연료첨가제	514	19
10	검마모	447	20
			카택을
			75

[불스원상 연관 검색어]

불스원샷 연관 검색어와 화제어 분석

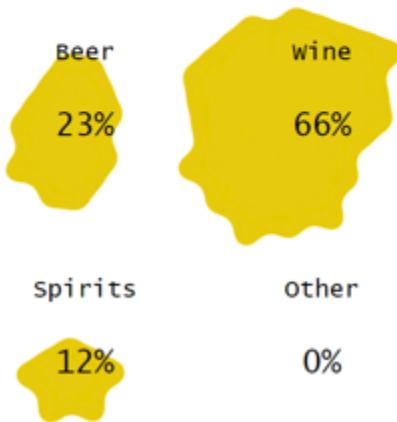
2.5 창의적 활용

What Alcohol People Around the World Drink

[More Info Here](#)

Select a Country

Italy •

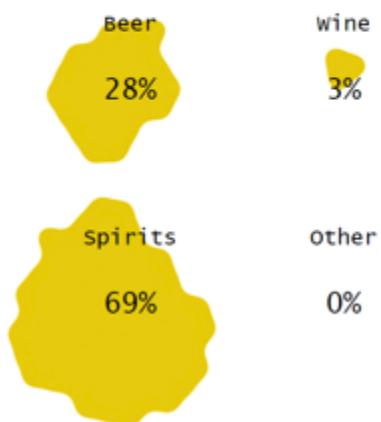


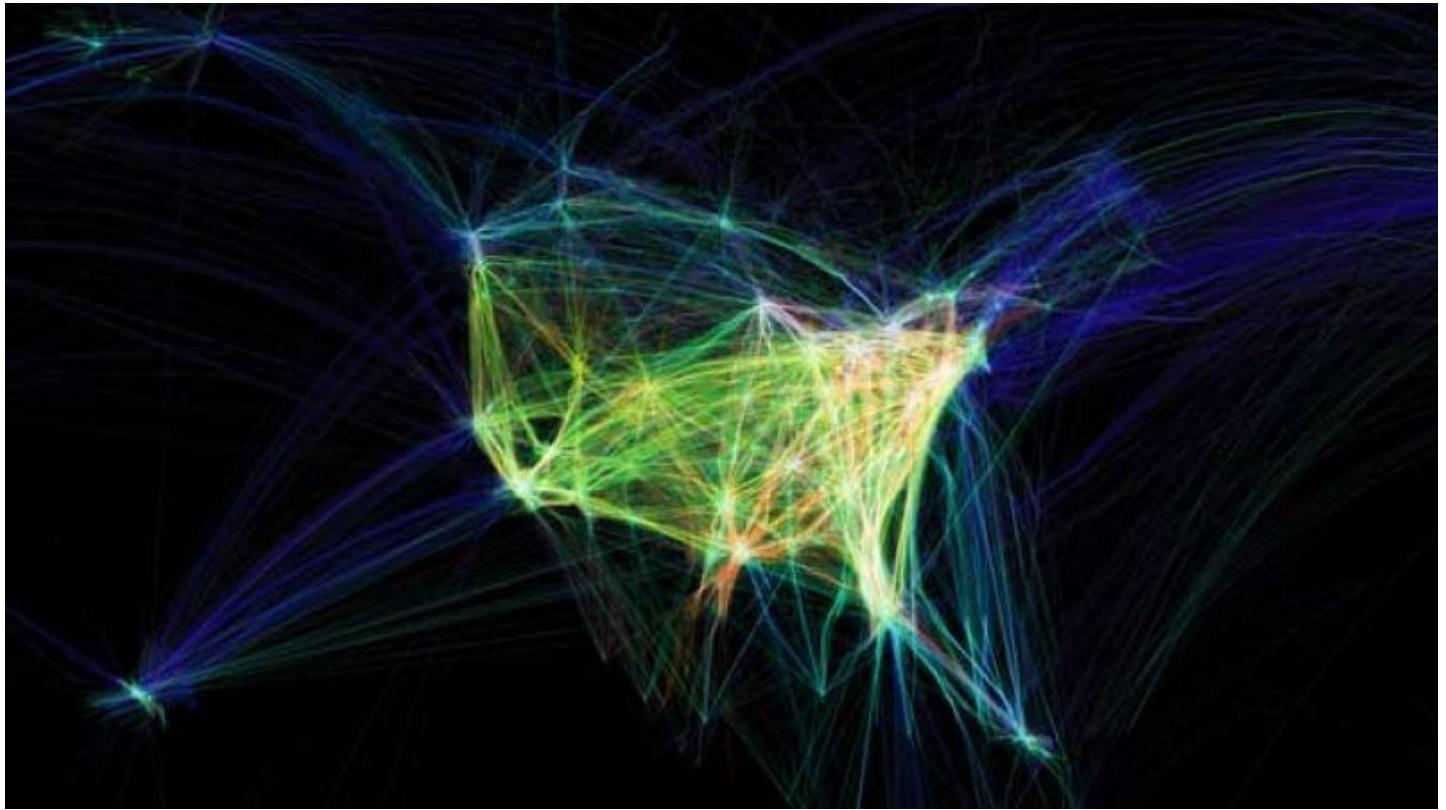
What Alcohol People Around the World Drink

[More Info Here](#)

Select a Country

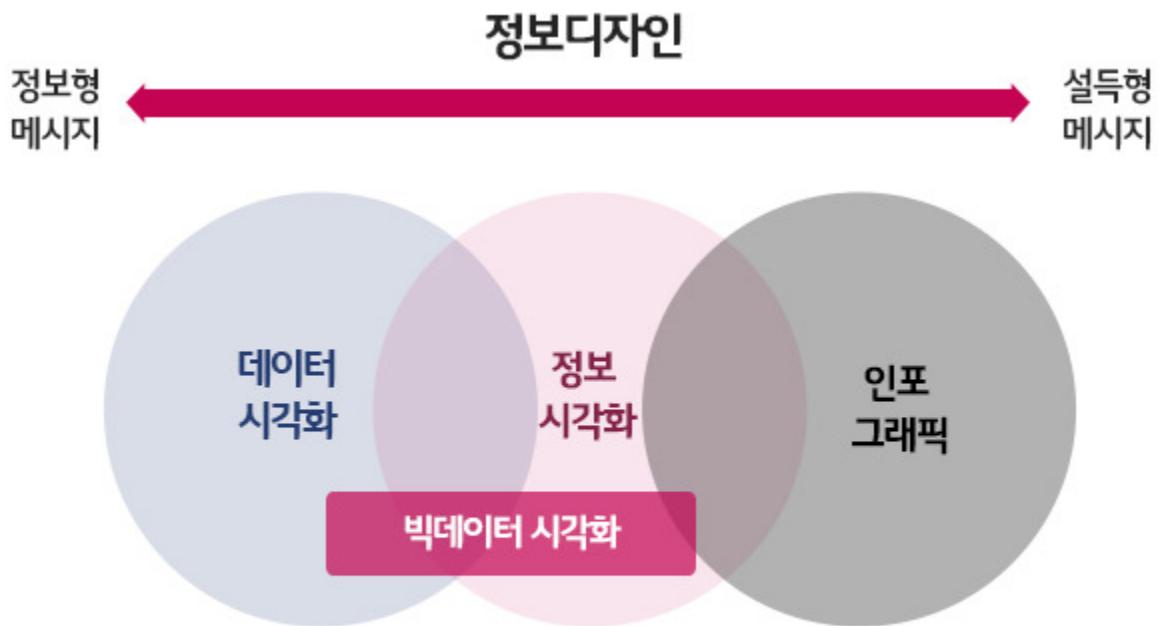
china •





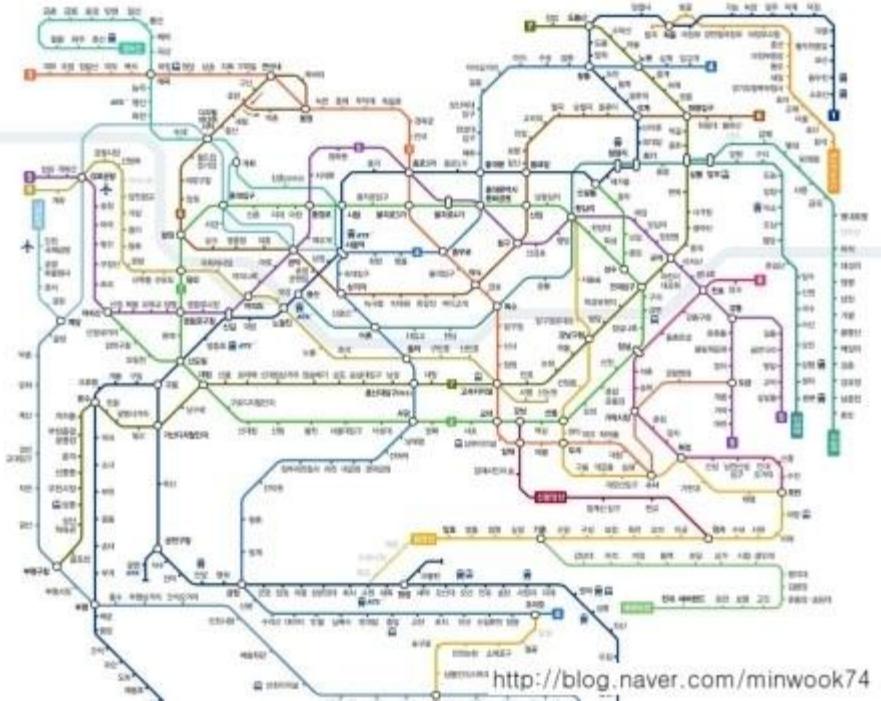
3. 무엇을 배워야 하나?

3.1 범위



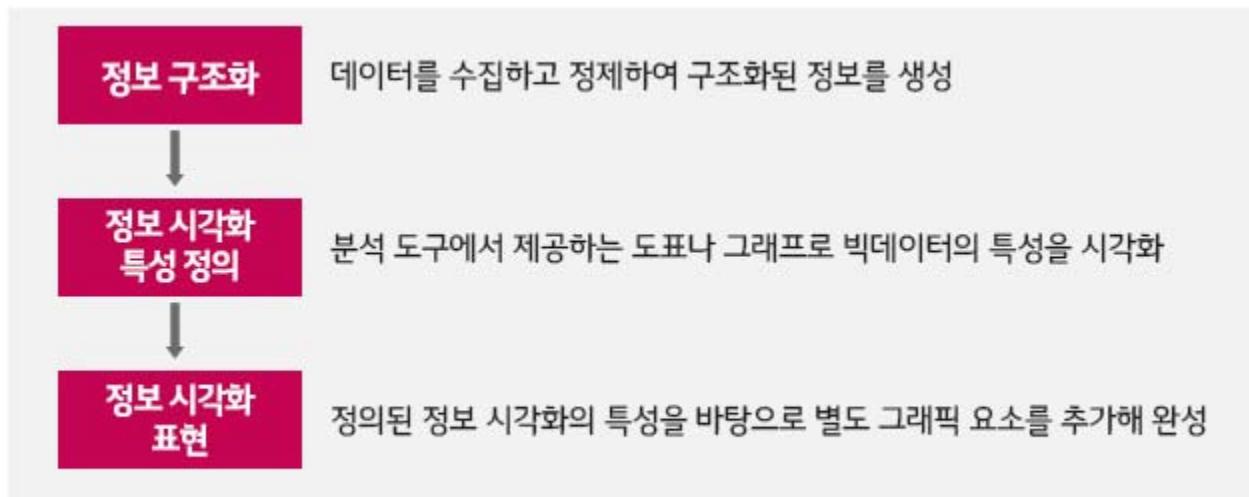
| 정보 디자인에서 빅데이터 시각화 영역
(출처: 한국데이터베이스진흥원의 데이터 분석 전문가 가이드 도서, 재구성)

- 정보 시각화
- 인포그래픽

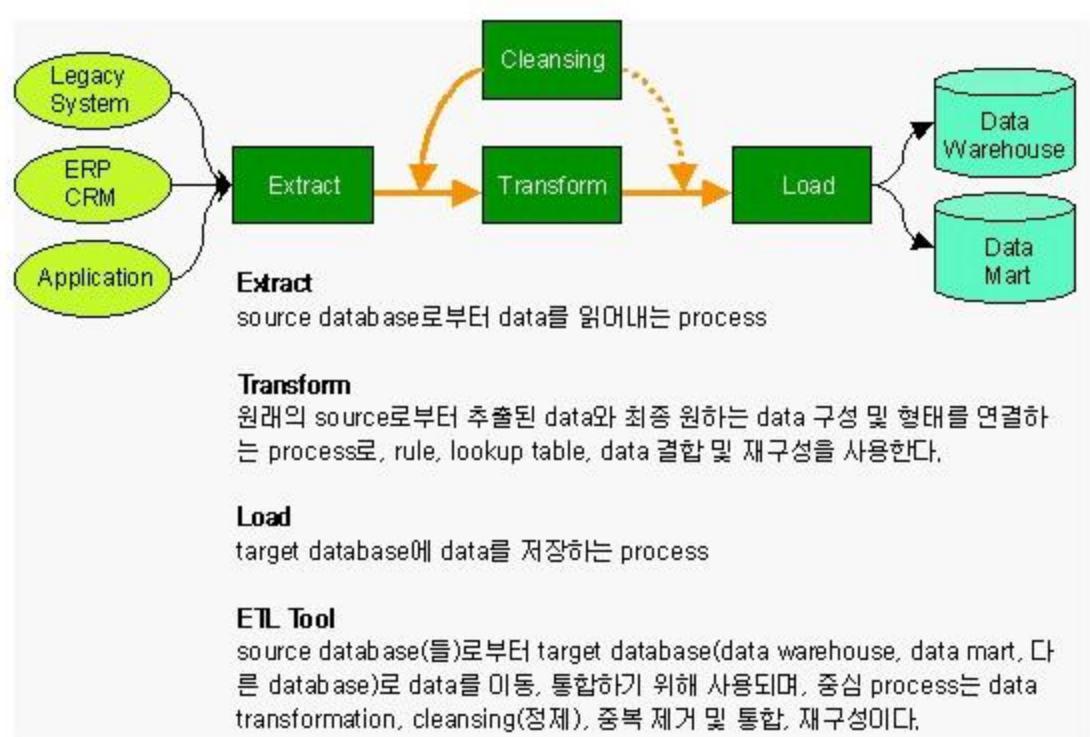


- 데이터 시각화

3.2 절차



| 빅데이터 시각화 프로세스 (출처: 한국데이터베이스진흥원의 데이터 분석 전문가 가이드 도서, 재구성)



3.3 도구

구분	도구
Basic Tools	Excel, CVS/JSON, Google, Chart API, Flot, Raphael, D3(Data-Driven Documents), Visually, infogr.am
Interactive GUI Control	Crossfilter, Tangle
Mapping	Modest Maps, Leaflet, Polymaps, OpenLayers, Kartograph, CartoDB
Professional Tools	Processing, NodeBox, R, Weka, Gephi

| 빅데이터 시각화 도구 (출처: 한국데이터베이스진흥원의 데이터 분석 전문가 가이드 도서, 재구성)

What You Will Learn: [hide]

Top 15 Best Data Visualization Tools In 2020

- #1) Zoho Analytics
- #2) Whatagraph
- #3) Tableau Desktop
- #4) Sisence
- #5) Adaptive Discovery
- #6) TeamMate Analytics
- #7) Jupyter
- #8) Dundas BI
- #9) Infogram
- #10) Google Charts
- #11) Visme
- #12) Domo
- #13) Cluvio
- #14) TapClicks
- #15) Looker
- #16) Microsoft Power BI

Few Additional Data Visualization Software

Conclusion

Recommended Reading

- 데이터 분석, R로 시작할까 파이썬으로 시작할까?

R의 장단점

장점 1 : 데이터 시각화

시각화된 데이터는 Raw 데이터 보다 훨씬 효과적으로 그 의도를 전달할 수 있습니다. R을 처음 만든 목적 자체가 ‘데이터 분석, 통계학 그리고 **시각화**를 하기 위해 더 좋고 친근한 방법을 만들어 보자!’입니다. 그만큼 R은 강력한 데이터 시각화 기능을 자랑하죠. 의미 전달이나 보고가 목적이라면 R을 활용하세요. 대표적인 시각화 패키지에는 ggplot2, ggviz, googleVis, rCharts 등이 있습니다.

장점 2: 생태계

R은 정말 풍부한 최첨단 패키지와 활발한 커뮤니티를 자랑합니다. 말 그대로 ‘R의 생태계’를 이루고 있죠. R 함수와 데이터들의 모음인 패키지, 이를 통해서 언제나 쉽게 최근의 테크닉과 기능에 접근할 수 있습니다. 자료를 찾느라 밤을 지새우지 않아도 되는 거죠!

장점 3: 데이터 사이언스의 통용어

R은 통계학자들에 의해, 통계학을 위해 개발되었습니다. 그래서 통계학자들은 그들의 아이디어와 개념들을 R의 코드와 패키지 등을 통해 공유하죠. 그래서 R을 활용한다면 데이터 사이언스를 시작하기 위해 꼭 컴퓨터 공학을 배워야 할 필요가 없습니다!

단점 1: 느린 속도

말씀드렸듯이 R은 통계학자들이 좀 더 쉽게 연구하려고 만든 언어이기에 컴퓨터를 효율적으로 활용하는 방법은 그리 깊게 고려하지 않은 듯합니다. 다시 말해 R은 읽기 어려운 코드 때문에 좀 느릴 수 있습니다. 하지만 그런 단점을 상쇄해줄 수 있는 pqR, renjin, FastR, Riposte 등의 패키지가 있죠! R의 패키지 생태계는 정말 대단하네요.

단점 2: 어렵다

R을 배우는 것은 그리 쉬운 일이 아닙니다. 풍부한 패키지가 있지만 그걸 사용하는데 익숙하지 않다면 활용하는데 시간이 꽤 오래 걸릴 겁니다. 입문자가 처음부터 R을 활용해 원하는 분석을 할 수는 없습니다. 꾸준한 학습이 필요하죠.

파이썬의 장단점

장점 1: IPython Notebook

IPython Notebook은 파이썬을 가지고 하는 데이터 작업을 훨씬 쉽게 만들어줍니다. 따로 부가적인 설치도 할 필요 없이, 동료들과 메모와 작업을 공유할 수 있습니다. 이를 통해 코드, 결과물, 메모 파일들을 정리하는데 필요한 비용을 획기적으로 줄여주죠. 진짜 해야 할 일에 집중할 수 있습니다.

장점 2: 특정 주제나 용도에 국한되지 않은 배우기 쉬운 언어

파이썬은 특정 분야를 위한 언어가 아니라 비교적 쉽고, 직관적입니다. 데이터 사이언스 말고도 다양한 분야에 많이 쓰이죠. 그만큼 배우기 쉽습니다. 코드를 익히고 쓰는데 드는 시간을 줄일 수 있습니다.

장점 3: 다양한 용도의 언어

파이썬은 다양한 백 그라운드의 사람들이 모두 사용할 수 있는 언어입니다. 이해하기 쉬운 언어이기 때문에 통계 백 그라운드를 가진 사람들도 배울 수 있죠. 워크플로우의 모든 부분을 통합시킬 하나의 도구를 만들 수도 있습니다.

단점 1: 시각화

R에겐 장점이었던 것이 파이썬에겐 단점입니다! 시각화는 데이터 분석 소프트웨어를 고르는데 매우 중요한 판단 기준이지만, 아쉽게도 파이썬은 시각화 기능이 R에 비해 떨어지죠. Seaborn, Bokeh and Pygal과 같은 시각화 라이브러리를 가지고 있긴 하지만, 너무 선택사항이 많은 감이 들기도 합니다. 더욱이 R과 비교하면 파이썬의 시각화는 난해한 모습을 하고 있는 경우가 많습니다. 그래서 시각화된 결과가 썩 만족스럽지 않을 때도 있죠.

단점 2: 파이썬은 도전자다

데이터 분석 툴에 있어서 R이 파이썬보다 선구자입니다. 파이썬은 도전자죠. 그래서 R이 수백 개에 달하는 기본적인 패키지들의 대체자가 되기에는 아직 많이 부족합니다. 파이썬이 최근 핫하게 떠오르고는 있지만, 여전히 사람들이 방대한 패키지를 자랑하는 R을 포기할 만큼인지는 미지수입니다.

- 190619, [Coding World News](#), “파이썬 vs. R” 데이터 사이언스에 더 적합한 언어는?

1. 우아함 (우승자: 파이썬)

우아함에 대한 평가는 매우 주관적이다. 맷로프 교수는 파이썬은 갈호나 교정기 사용을 크게 줄일 수 있기 때문에 파이썬의 명백한 승리이며, 파이썬은 매우 날렵하다고 말했다.

2. 학습 곡선 (우승자: R)

Python을 사용하기 위해선 NumPy, Pandas, matplotlib를 포함한 많은 자료를 배워야 한다. 하지만 R에는 기본적으로 매트릭스 유형과 기본 그래픽이 내장되어 있다.

시스템에 능숙한 경우에도 Python의 라이브러리를 구성하기 까다로울 수 있다. 반면 대부분의 R 패키지는 즉시 실행되며, 초보자도 R을 활용해 몇 분 안에 간단한 데이터 분석을 할 수 있다.

3. 사용 가능한 라이브러리 (무승부)

파이썬 패키지 인덱스(PyPI)는 183,000개 이상의 패키지를 가지고 있으며, CRAN(Compreged Rarchive Network)은 12,000개 이상의 패키지를 가지고 있다. 그러나 데이터 과학에 대한 PyPI는 적다.

예들 들어 맷로프가 주어진 데이터 포인트의 가장 가까운 이웃에 대한 빠른 계산을 하기 위해 코드가 필요할 때, 두 개의 패키지를 즉시 찾을 수 있었다. 반면 파이썬에서는 가장 가까운 이웃의 코드를 찾을 수 없었다. 그는 간단하고 직설적인 구현은 오직 한 개였다고 언급했다.

또한 맷로프는 PyPI에 다음 용어를 검색하면 아무것도 나타나지 않았다고 덧붙였다: 로그 선형 모델, 포아송 회귀 분석, 기약 변수, 공간 데이터, FWE 비율.

4. 머신 러닝 (우승자: Python, 하지만 많이는 아님)

최근 몇 년간 파이썬의 엄청난 성장은 머신러닝과 인공지능의 상승에 일부분 영향을 받았다. Python은 AlexNet과 같이 이미지 인식을 위해 정밀하게 조정된 라이브러리를 많이 제공하지만, R 버전도 쉽게 개발할 수 있다.

맷로프는 파이썬 라이브러리의 힘은 특정한 이미지 평탄화 ops 설정에 의해 생기며, R의 Keras wrapper를 통해 쉽게 구현할 수 있다고 말했다. 순수한 R 버전의 TensorFlow(텐서 플로)가 개발될 수 있으며, R의 랜덤 포레스트와 그래디언트 부스팅의 가능성이 탁월할 수 있다고 주장한다.

5. 통계적 정확성 (우승자: R, 월등히 좋음)

맷로프는 파이썬을 옹호하는 머신러닝 전문가들이 때때로 관련된 통계적 문제에 대해 잘 이해하지 못한다고 지적했다. 반면 R은 통계학자에 의해 작성되었다고 덧붙였다.

6. 병렬 연산 (무승부)

R과 Python의 기본 버전은 멀티 코어 연산에 대한 지원이 약하다. 파이썬의 멀티프로세싱 패키지는 다른 이슈들에 대한 좋은 해결책이 아니며, R의 병렬 패키지도 마찬가지이다.

맷로프는 “클러스터 연산을 지원하는 외부 라이브러리는 두 언어 모두 끌찮다”, “현재 Python은 GPU에 대한 더 나은 인터페이스를 가지고 있다.”라고 말했다.

7. C/C++ 인터페이스 (우승자: R, 하지만 많이는 아님)

R의 Rcpp는 R과 C/C++를 연결하는 강력한 도구이다. 파이썬 swig(스위그)와 같은 도구를 가지고 있지만 그것은 그렇게 강력하지 않으며, Pybind11 패키지는 아직 개발 중이다. R의 새로운 ALTREP 아이디어는 성능과 사용성을 향상시킬 수 있는 잠재력을 가지고 있다. 하지만 Cython과 파이썬의 PyPy 변형은 때때로 C/C++ 인터페이스의 필요성을 없앨 수 있다.

8. 객체 방향, 메타 프로그래밍 (우승자: R, 하지만 많이는 아님)

R과 Python 둘 다에서 기능이 객체임에도 불구하고 R은 이를 더욱 심각하게 받아들이고 있다고 언급했다. 그는 “파이썬을 사용할 때마다 R에서 많이 하는 것처럼 단말기에 기능을 인쇄하지 못한다는 사실에 짜증이 난다”고 말했다.

Python은 단지 하나의 OOP 패러다임을 가지고 있다. 반면 R에서는 몇 가지를 선택할 수 있다. 그는 R의 마법 메타 프로그래밍 기능(코드를 생성하는 코드)을 감안할 때 컴퓨터 과학자들은 R에 대해 주목해야 한다고 주장한다.

9. 언어의 통일성 (우승자: 파이썬, 월등히 좋음)

파이썬이 버전 2.7에서 3.x로 변해도 큰 혼란이 생기지 않을 것이다. 하지만 R은 RStudio의 영향으로 두 개의 다른 언어로 바뀌어가고 있다. (R, Tidyverse)

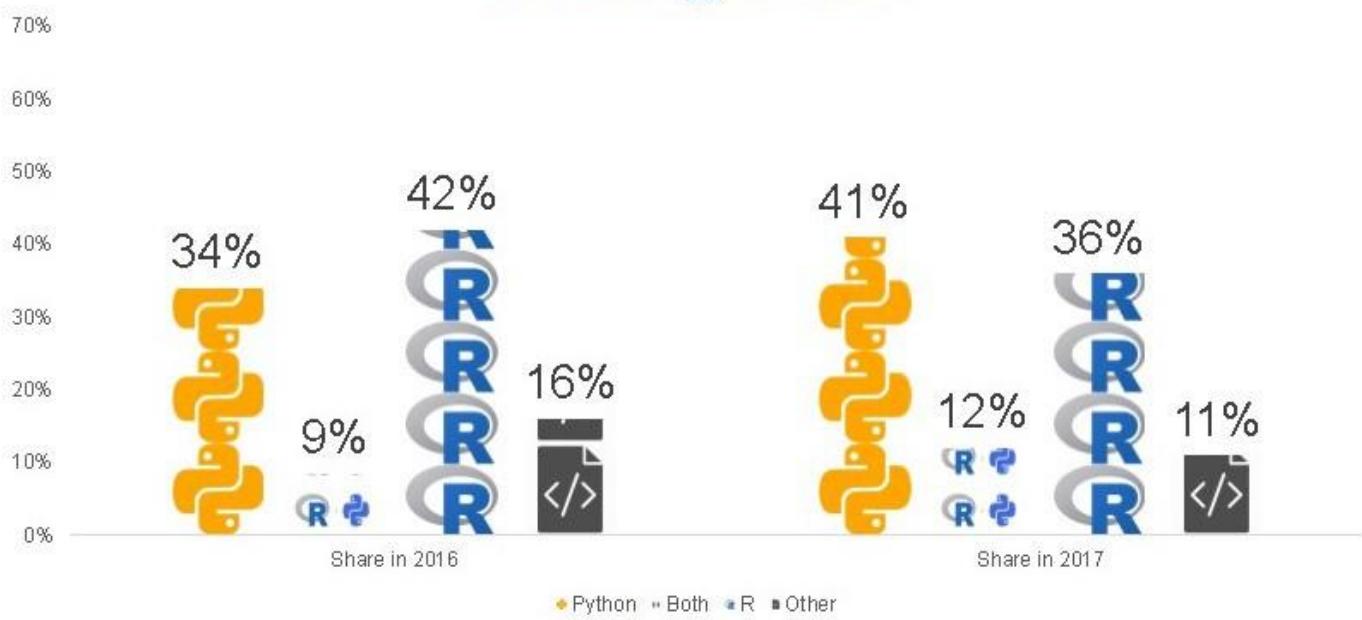
맷로프는 Tidyverse가 일반적인 R보다 우월하다면 받아들일 수 있지만, 개인적으로 그렇게 생각하지 않는다고 언급했다. 또한 이것이 초보자들에게 어렵게 만든다고 말했다.

10. 링크된 데이터 구조 (우승자: 파이썬)

Binary tree 같은 고전적인 컴퓨터 사이언스 데이터 구조는 파이썬에서 구현하기 쉽다. R에서 리스트 클래스를 사용해 구현할 수 있지만, 맷로프는 이것과 느리다고 추측한다.

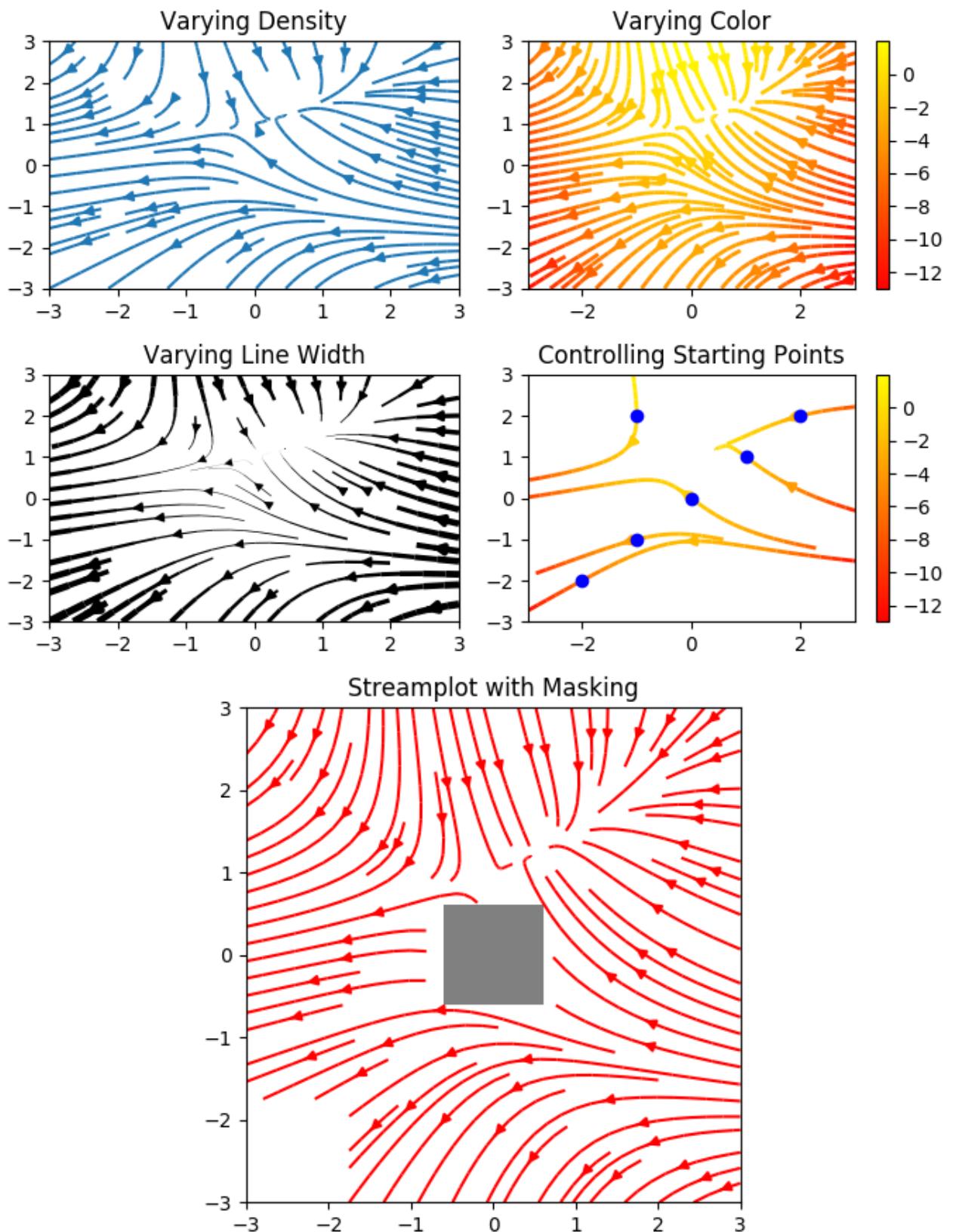
- 180131, Data-Driven Science, Python vs R for Data Science: And the winner is..

Python, R, Both or Other Platforms for Data Science
Source: KDnuggets Poll 2017



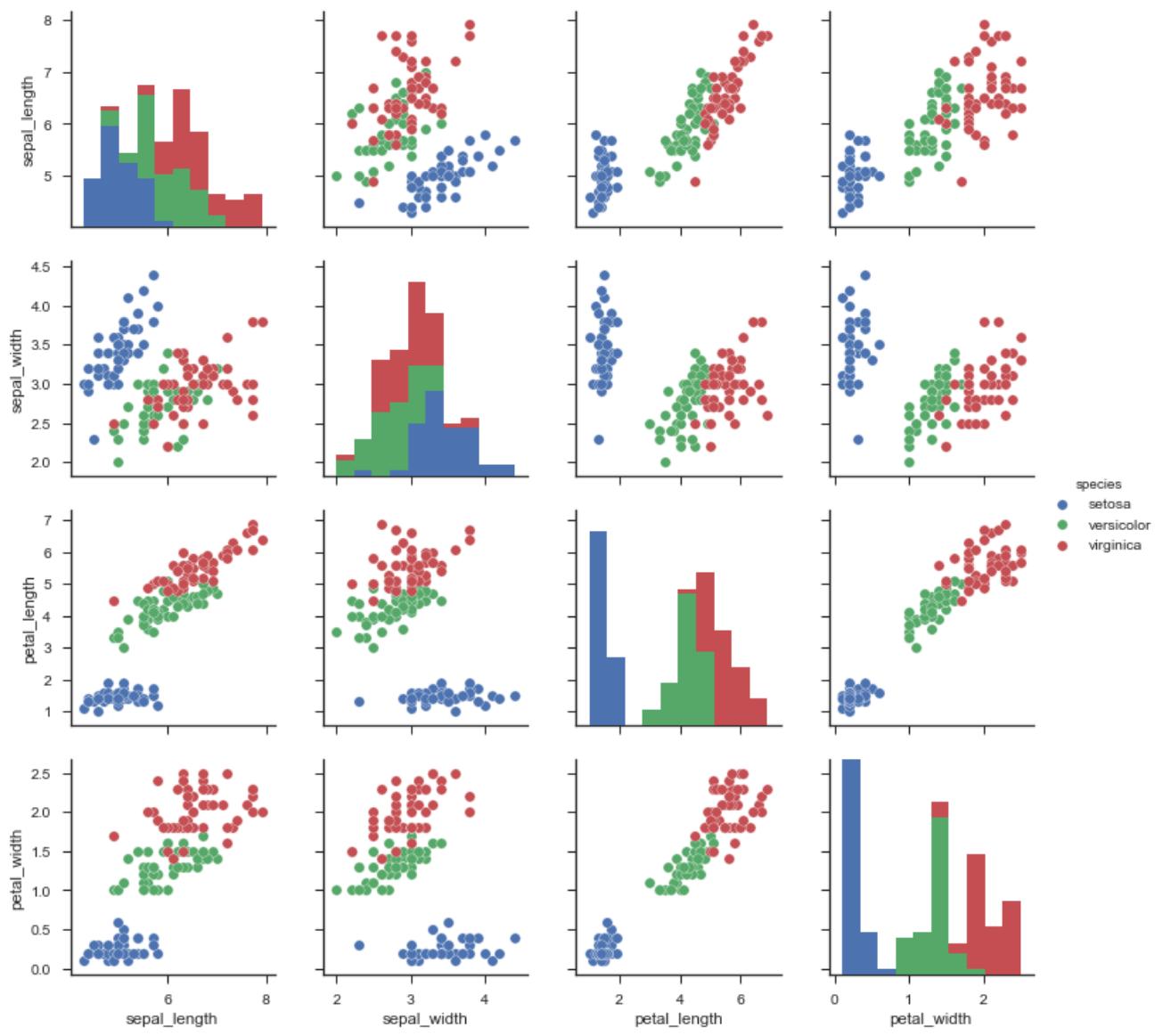
3.4 파이썬

- 파이썬 라이브러리 개괄
 - Matplotlib**
모든 파이썬 라이브러리의 엄마같은 존재,
다른 도구를 활용하기 위한 기반 역할을 하는 라이브러리



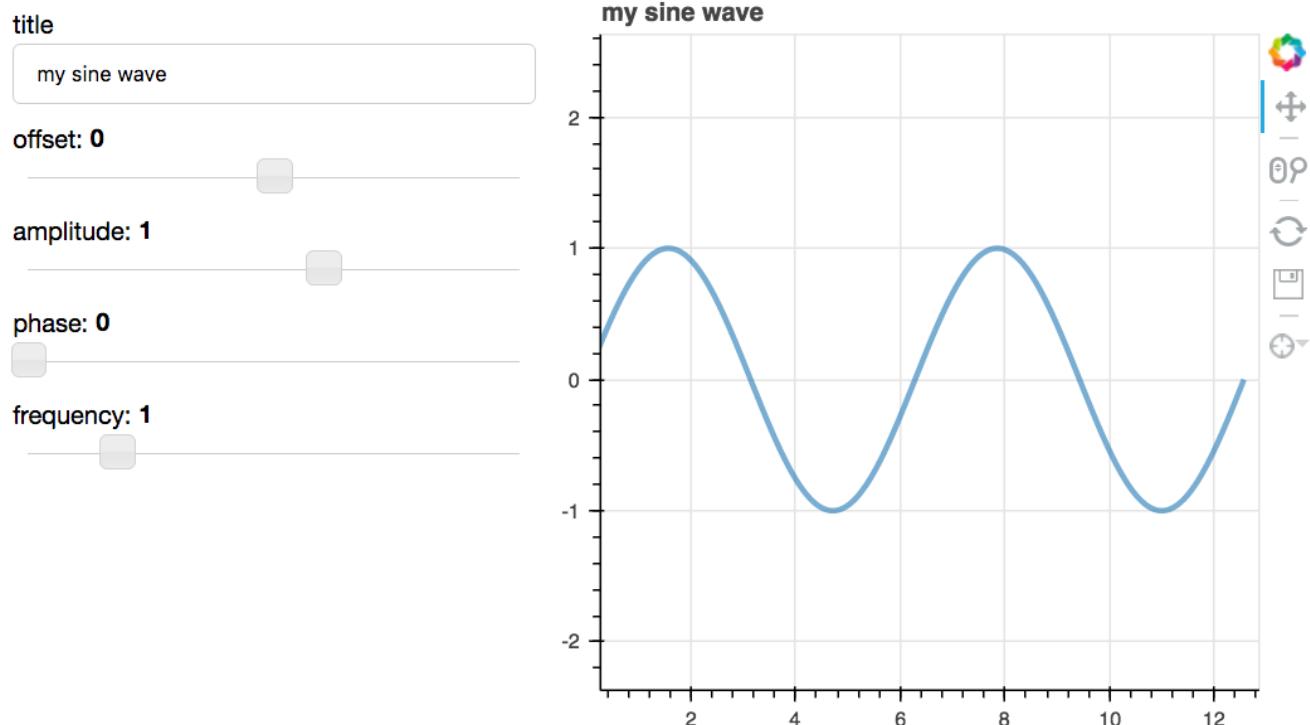
- **Seaborn**

더 복잡한 시각화 접근을 가능하게 해주지만,
섬세하게 다듬기 위해서는 matplotlib에 관한 지식이 여전히 필요함



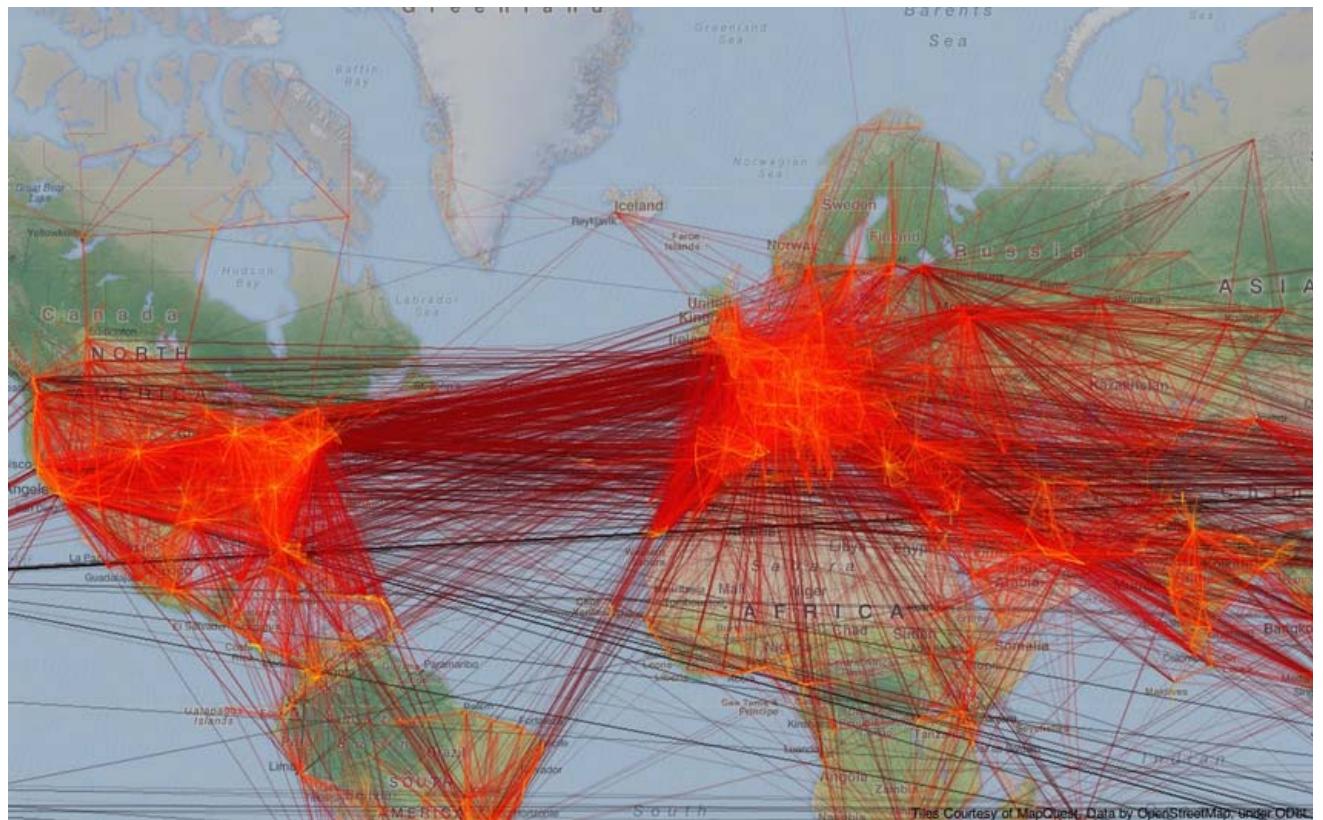
o Bokeh

당신만의 시각화 서버를 구축하기 위한 견고한 도구이지만,
단순한 시나리오를 작성해야 하는 상황이라면 너무 과한 도구



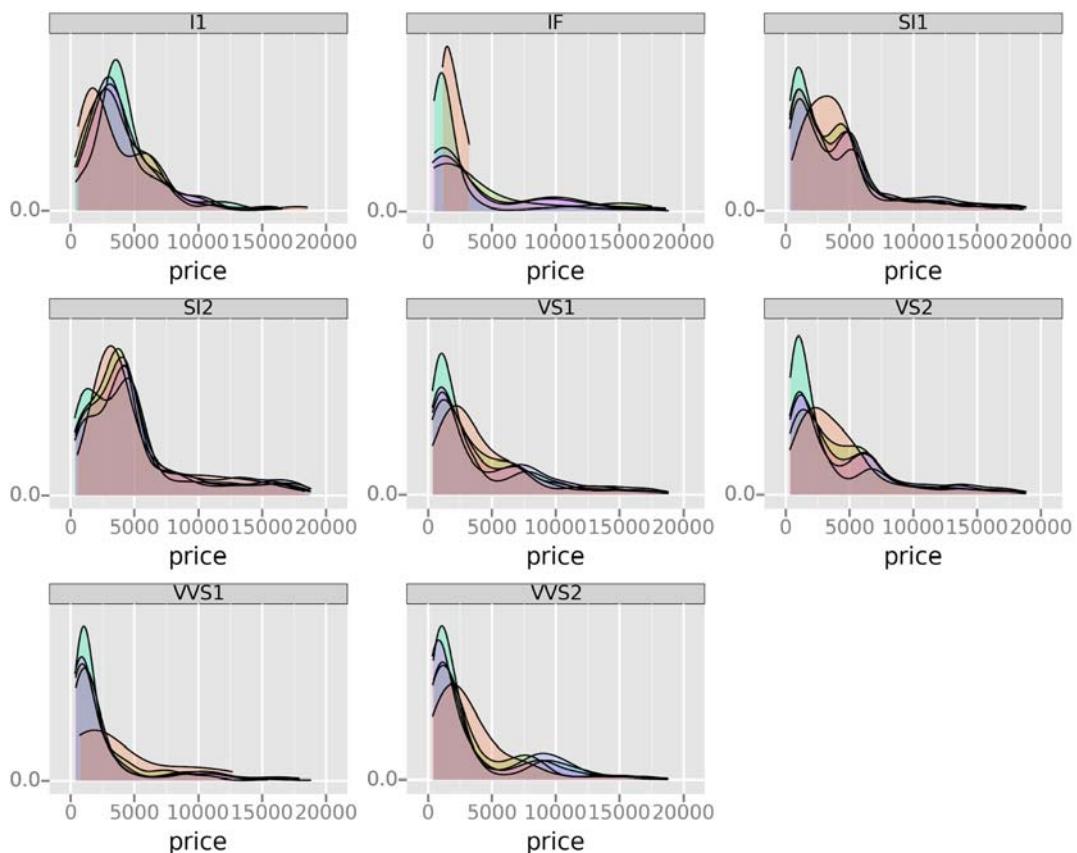
- **Geoplotlib**

지리적 데이터에 관한 시각화 도구



- **ggplot**

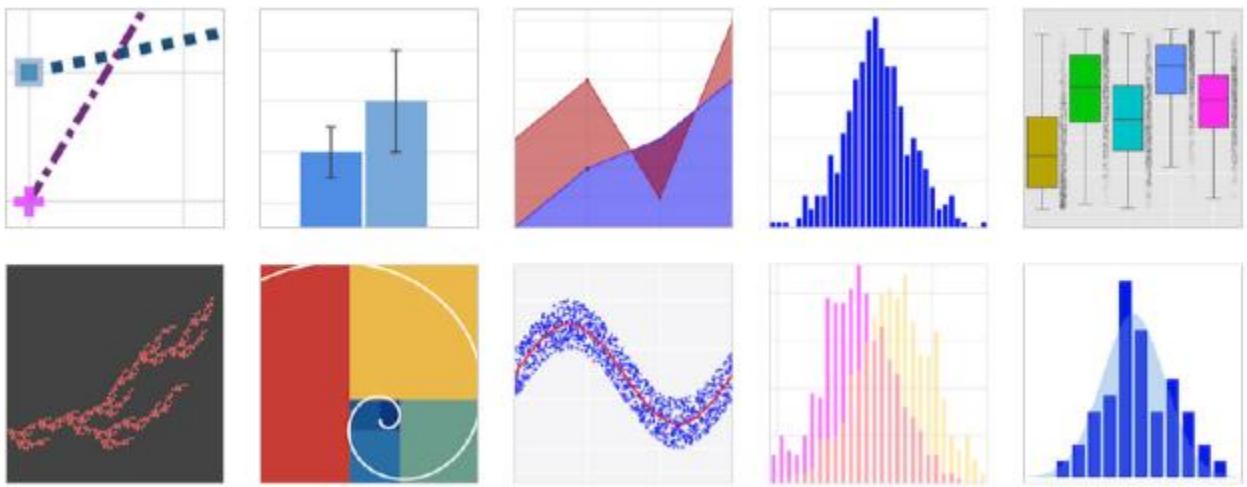
쓸만하지만, 아직도 발전하고 있는 도구



- **Plotly**

상호작용성이 가장 뛰어난 그래프를 생성할 수 있으며,

멋진 웹 기반 시각화를 위하여 결과를 오프라인으로 저장할 수 있음



- 용도에 맞는 프레임워크 설정

- **인공 신경망**

대부분의 튜토리얼이 matplotlib를 사용하고 있으며,
matplotlib라는 범용 시각화 도구는 다른 도구를 적용하기 위한 기반의 역할에 적합함

- **회귀 및 분류 모델**

회귀 또는 분류 작업에서는 주로 Seaborn을 사용하는데,
Seaborn에서 상관 행렬(correlation matrix)을 제공하기 때문임

- **주가 변동 표시기**

데이터 변동을 실시간으로 표시하고,
사용자가 차트 위로 커저를 옮길 때 실시간으로 반응해야 하는 상황,
주로 Plotly 또는 Bokeh를 사용하며, 둘 다 자바스크립트를 지원함

- **지도**

PostgreSQL, MongoDB 및 JSON 형식 파일에 지리 정보를 저장해 두고,
지도와 관련한 데이터를 표시해야 한다면,
Geoplotlib 프레임워크를 추천함

- **통계학 응용 연구**

제약회사의 임상 실험 등의 상황에서는 전통적으로 R 도구가 애용되었으나,
최근에는 Python 도구 활용이 늘어나는 추세이며, ggplot 또한 추천 대상임

- 180824, Seongyun Byeon, [Python에서 데이터 시각화하는 다양한 방법](#)

- 파이썬에서 자주 쓰는 데이터 시각화 라이브러리
 - 다양한 차트 예제
 - 간단한 코드 예제
 - 충실한 원 자료 링크
 - 이 기사에서 링크하고 있는 [누구나 쉽게 이해할 수 있는 Git 입문](#) 공부하기를 강추

- 190325, Zeming Yu, [My top 10 Python packages for data science](#)

- 데이터 처리
 - pandas
 - 데이터 정렬, 병합, 필터링, ...

[Python for Data Analysis](#)

- numpy

pandas 개발의 기반 역할을 한 패키지로서, 기본적 데이터 조작을 위한 패키지
연령에 근거해서 은퇴여부라는 새로운 열을 생성하려면,

```
df['은퇴여부'] = np.where(df['연령']>=65, 'yes', 'no')
```

- qgrid

쥬피터 노트북에서 데이터프레임을 정렬, 필터, 편집하기 위한 멋진 패키지

- 그래프 작성

- matplotlib

모든 종류의 그래프 작성에 유용한 패키지로서,

쥬피터 노트북에서 작성한 그래프가 출력되도록 하려면 다음 코드 라인을 실행해야 함

```
%matplotlib inline
```

- seaborn

matplotlib로 작성한 그래프를 더욱 멋지게 보이도록 해주는 패키지

- plotly

사용자와 상호작용하는 그래프를 쥬피터 노트북에서 작성하도록 해주는 패키지

- 모델링

- statsmodels

보험계리사가 사용하는 Generalized Linear Model(GLM) 작성, 시계열 분석 및
기타 통계적 모델링 기능을 제공하는 패키지

- scikit-learn

기계학습의 주력 패키지로서,

분류, 회귀, 군집화 및 차원 축소 등의 모든 기계 학습 작업을 수행할 수 있음

- lightgbm

Gradient Boost Machine(GBM)을 위한 기계 학습 패키지

- lime

기계학습의 모델 해석 패키지로서, 모델 예측과 예측의 주요 인자 분석에 사용됨

3.5 데이터 과학 코스

	Python Learn the most important language for Data Science
	Machine Learning Machine learning is the hottest field in data science, and this track will get you started quickly.
	Pandas Short hands-on challenges to perfect your data manipulation skills
	Data Visualisation Visualisation is one of the most versatile skills in data science. Make insightful and beautiful graphics to see what's happening in any dataset.
	SQL Learn SQL for working with databases, using Google BigQuery to scale to massive datasets.
	R Learn the language designed for data analysis. This track includes data set-up, machine learning and data visualization.
	Deep Learning Learn TensorFlow to take Machine Learning to the next level. Your new skills will amaze you.

4. 알아야 할 원리는?

4.1 다양한 차트 유형



Arc Diagram



Area Graph



Bar Chart



Box & Whisker Plot



Brainstorm



Bubble Chart



Bubble Map



Bullet Graph



Calendar



Candlestick Chart



Chord Diagram



Choropleth Map



Circle Packing



Connection Map



Density Plot



Donut Chart



Dot Map



Dot Matrix Chart



Error Bars



Flow Chart



Flow Map



Gantt Chart



Heatmap



Histogram



Illustration Diagram



Kagi Chart



Line Graph



Marimekko Chart



Multi-set Bar Chart



Network Diagram



Nightingale Rose Chart



Non-ribbon Chord Diagram



Open-high-low-close Chart



Parallel Coordinates Plot



Parallel Sets



Pictogram Chart



Pie Chart



Point & Figure Chart



Population Pyramid



Proportional Area Chart



Radar Chart



Radial Bar Chart



Radial Column Chart



Sankey Diagram



Scatterplot



Span Chart



Spiral Plot



Stacked Area Graph



Stacked Bar Graph



Stem & Leaf Plot



Stream Graph



Sunburst Diagram



Tally Chart



Timeline



Mind Map



Dendrogram



Venn Diagram



Scatterplot



Tag Cloud



4.2 선택 기준

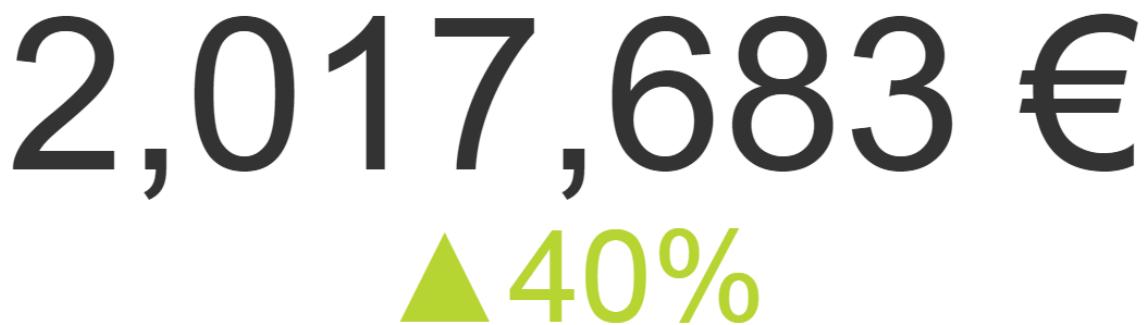
- 190302, Sandra Durcevic, [How to Choose the Right Data Visualization Types](#)
- 어떤 유형의 차트를 사용할 것인가?
 - 스토리 텔링의 관점에서 고민해야 함
 - 누구에게, 어떤 메시지를 전달할 것인가를 고민해야 함
 - 최종적으로 얻어 내고자 하는 성과를 고려해야 함

4.2.1 데이터 시각화 유형 12 종

1. number chart

- 특정 KPI 변동을 보여줄 때
- 많은 종류를 나열하면 곤란
- 등락의 비교 기준이 동일해야 함

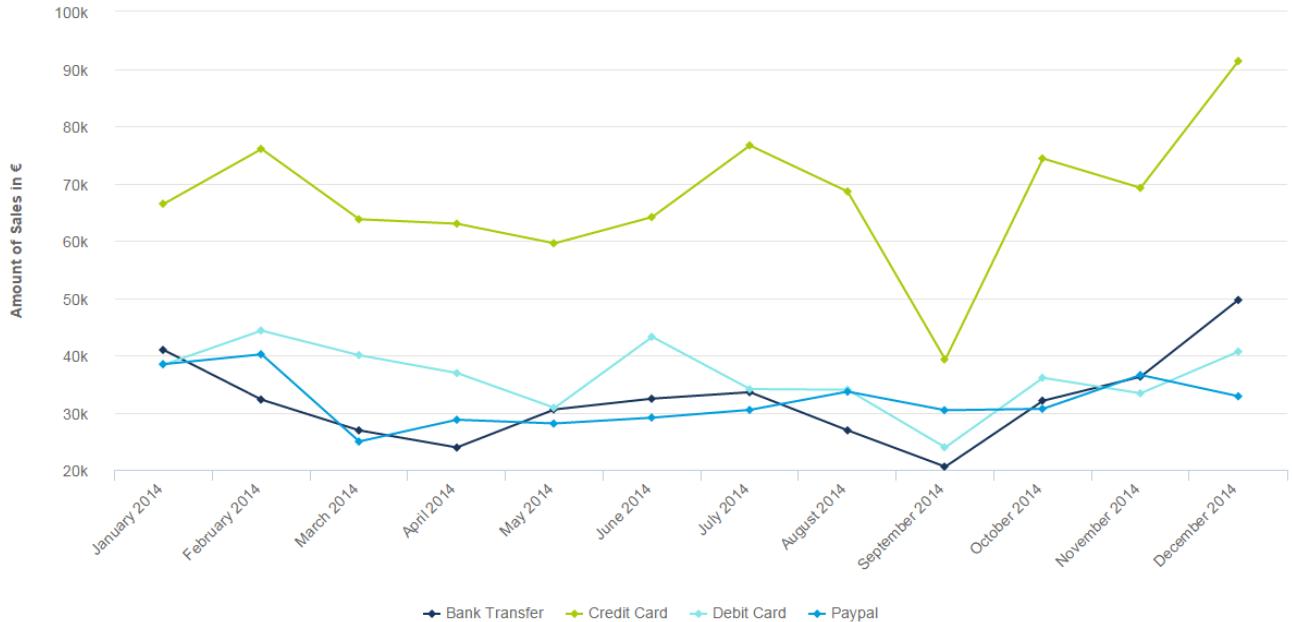
Amount of Sales Year to Date vs Last Period



2. line chart

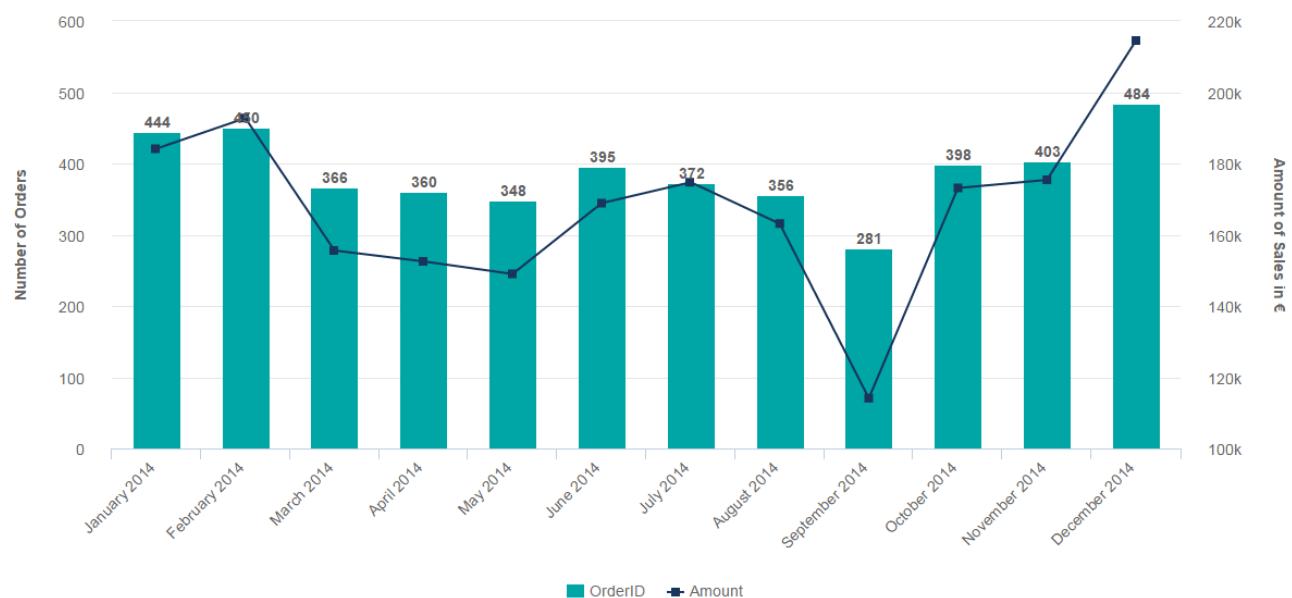
- 시간에 따른 추세성, 경향성, 등락성을 보여줄 때
- 너무 많은 선(변수)은 곤란
- y 축 최대값이 데이터 최대값과 근접하도록

Amount of Sales by Payment Method (last year)



- 이중 y 축 (범례에 주목)

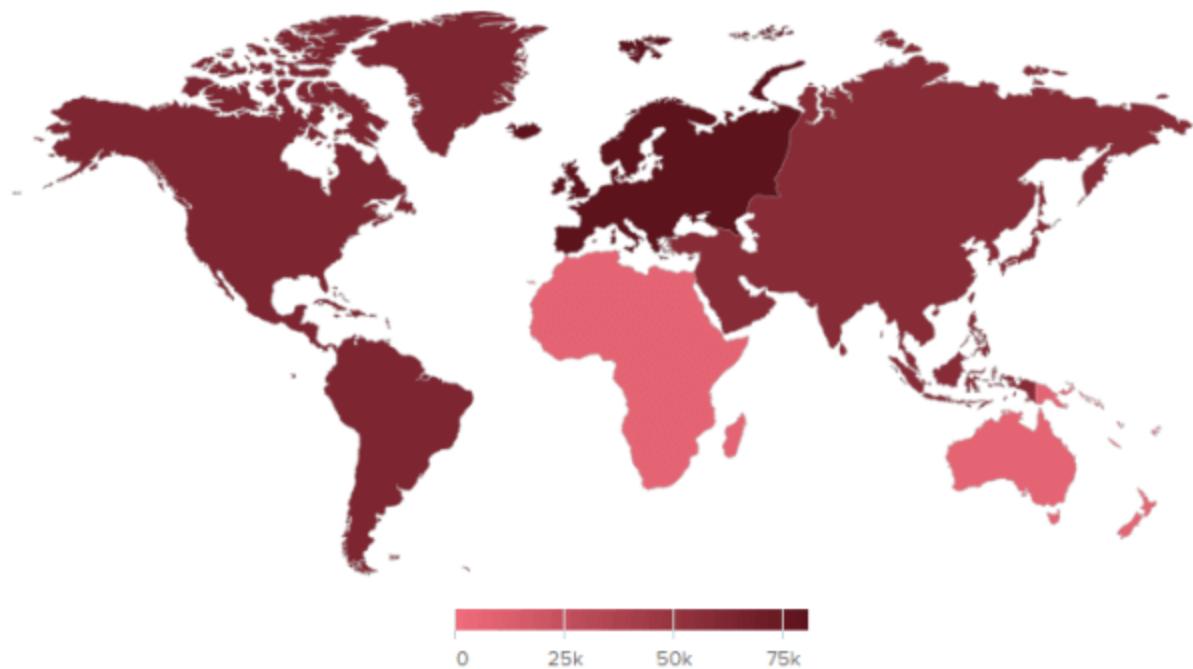
Number of Orders and Amount of Sales (last year)



3. (choropleth) map

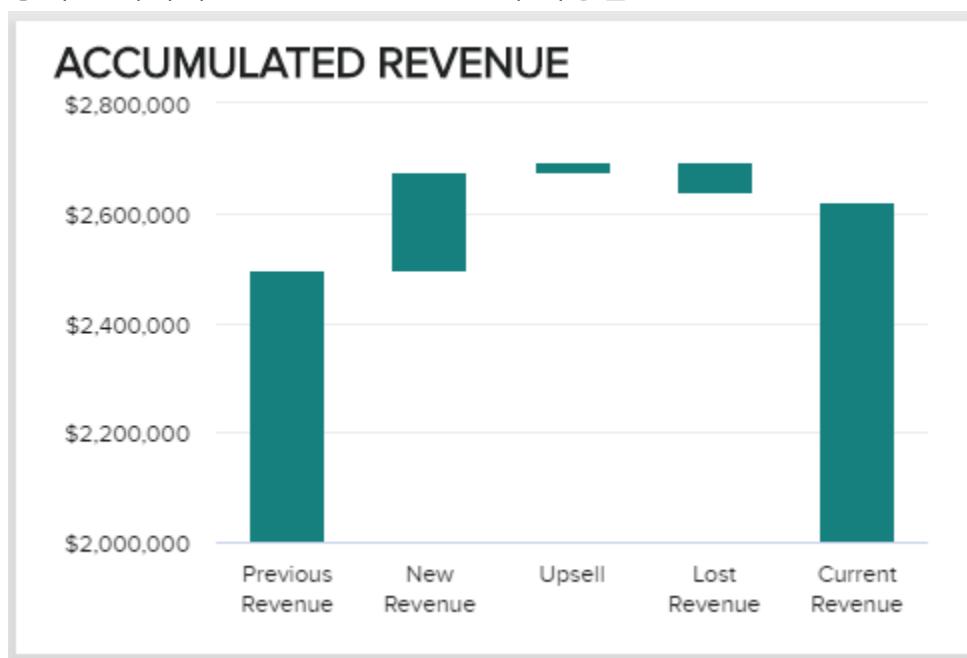
- choropleth map: 지리적 데이터에서 특정 값의 단계를 색상으로 구분하여 표시
- bubble map: 특정 위치에 크기로 구분한 원을 표시
- 지도 위에 매우 많은 점을 찍어 놓은 형식은 비주

Sessions By Continent (Last 6 Months)



4. waterfall chart

- 정적이지만, 시간 흐름이나, 진행 단계에 따른 증감을 구분하여 표시할 때
- 정적인 데이터는 waterfall chart가 적당하지만,
동적인 데이터는 stacked bar chart가 적당함



5. bar graph

- horizontal bar graph
- 순위 비교에 적당함, 데이터 레이블이 긴 경우에도 유용함
항목의 순서에 의미가 있도록

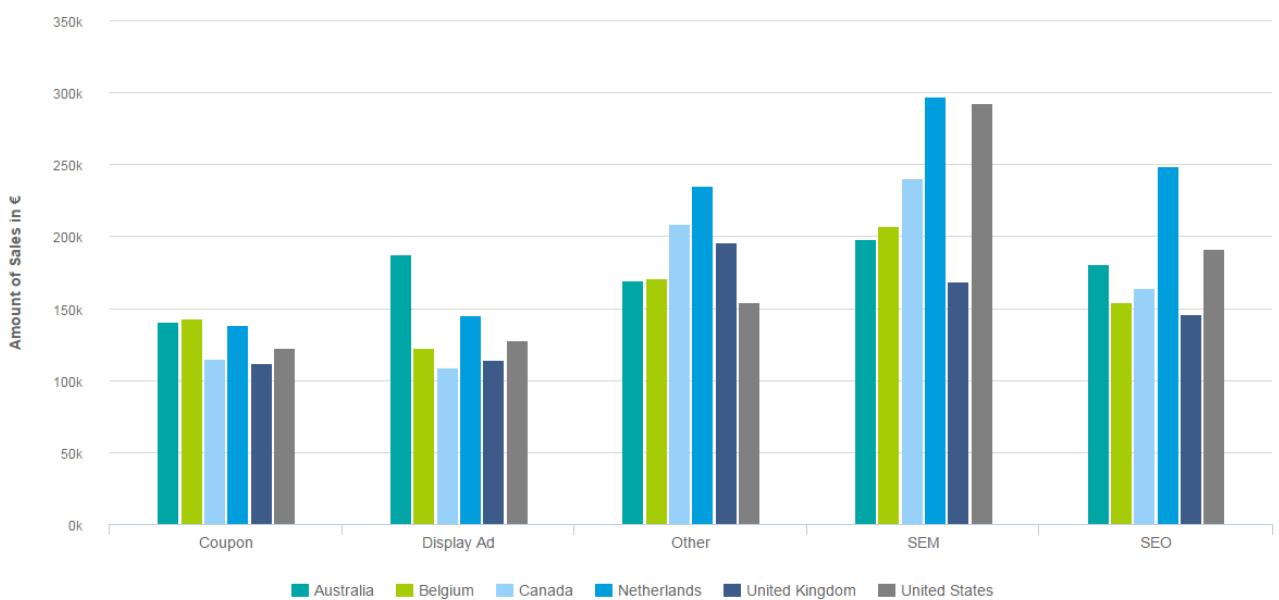
Top 5 Products on Sales (last year)



- column graph

시간 순서 데이터를 표현하는 표준적 방법
범주에 따른 데이터를 비교하는 표준적 방법
국가별, 판매 채널별 판매량

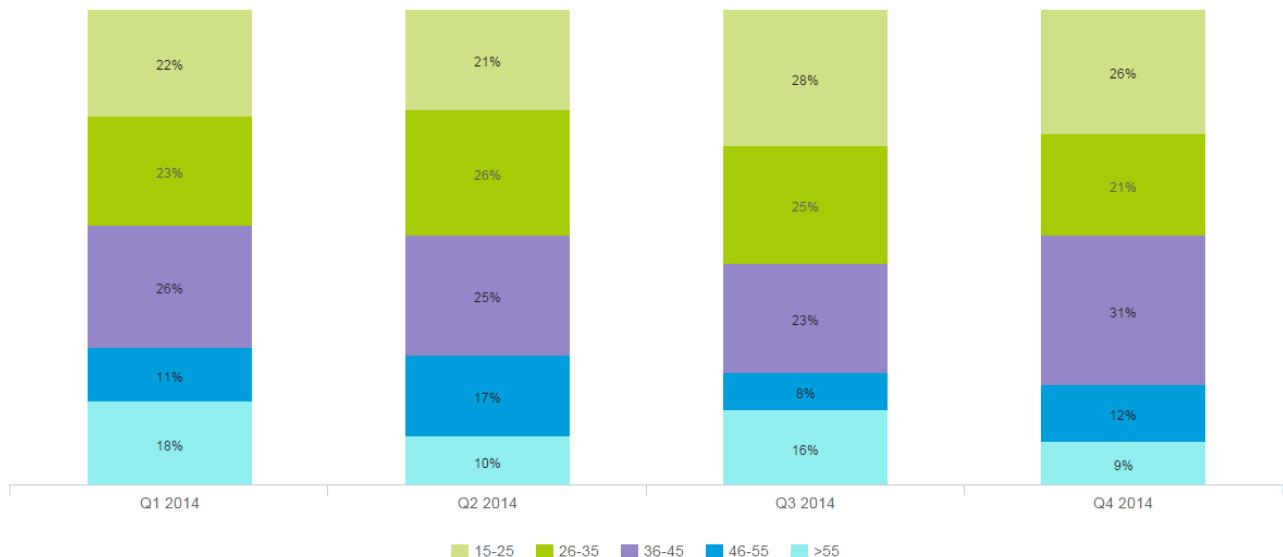
Amount of Sales per Channel and Country (last year)



- stacked column chart

전체-부분 관련성을 표시하는 방법
합계값 자체보다는 상대적 비중(백분율)을 비교하여 표시
단일 항목에 대해서라면 pie chart로 비중을 표시해야 함
항목이 너무 많아지면, 막대가 얇아져서 어색해짐

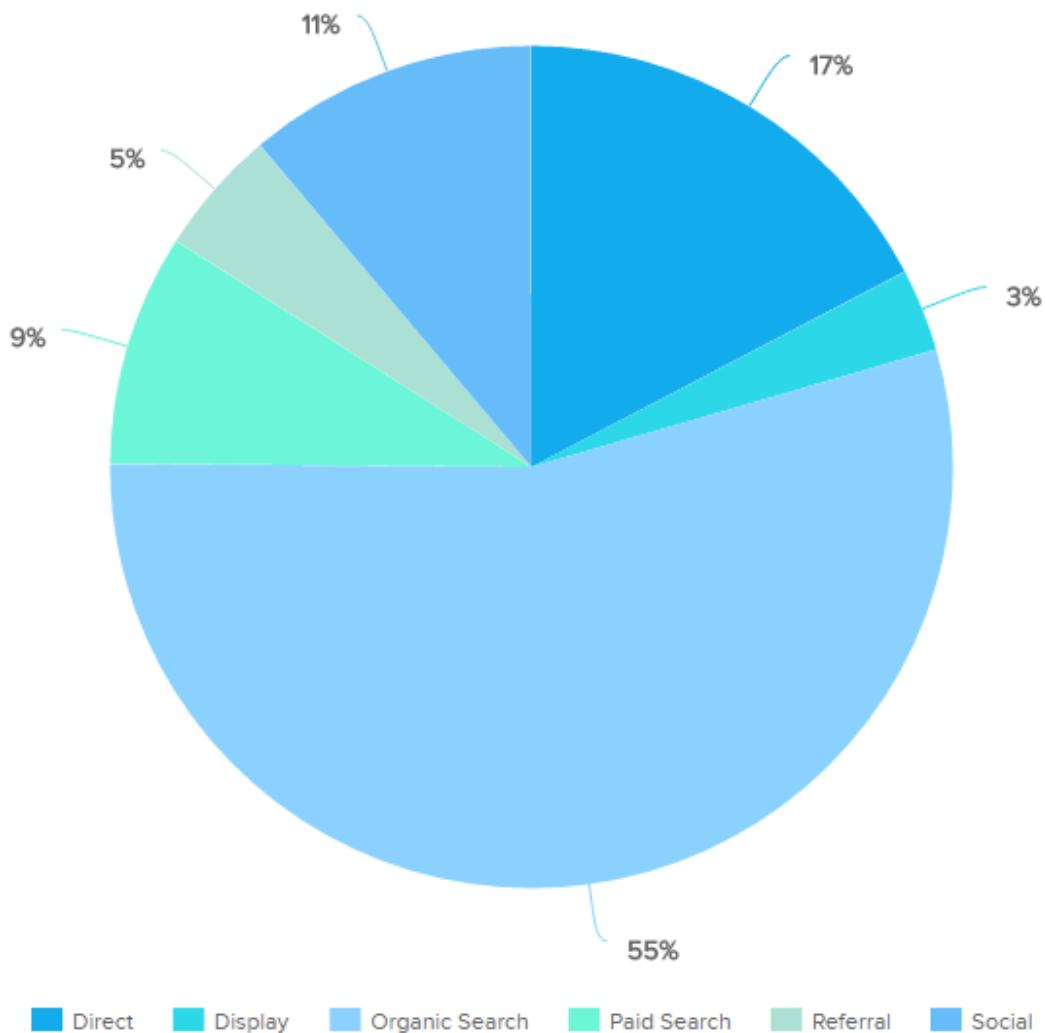
Age of New Customers per Quarter (last year)



6. pie chart

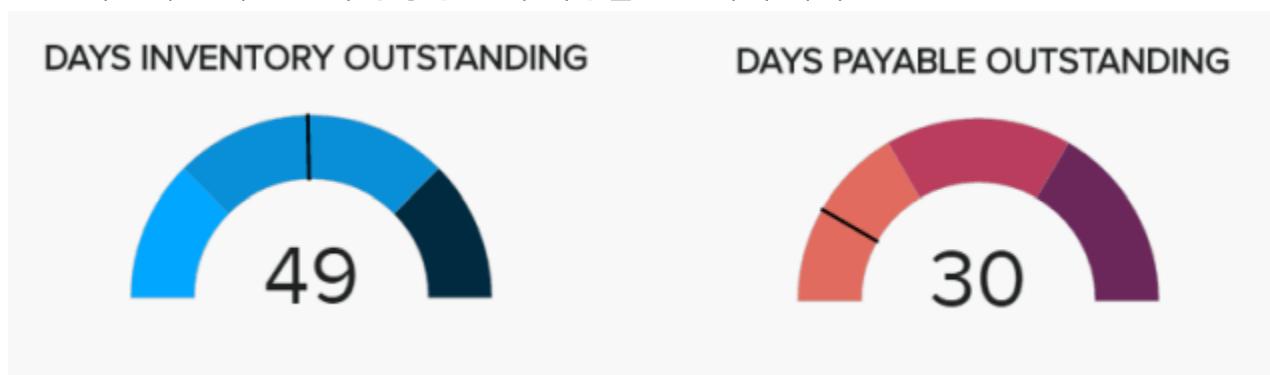
- 조각별 비중을 합한 전체가 100%가 될 때
- 비중 수치의 정밀도가 문제된다면 stacked bar chart가 좋음
- 압도적 비중인 항목을 강조할 때
- 경향성 및 추세성을 강조할 때는 적합하지 않음

Sessions by Channel



7. guage chart

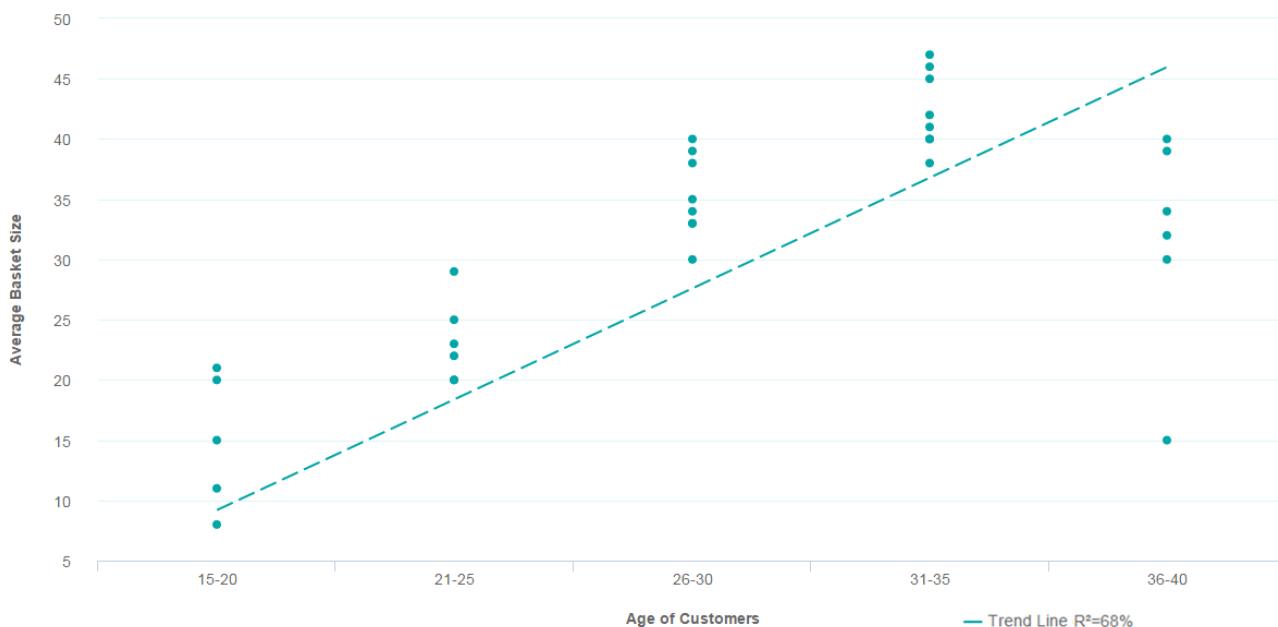
- dial chart, speedometer chart
- 바늘과 색상으로 구분하는 구역
- 단일 수치로 표현되는 핵심 경영 지표를 표시할 때
- 단일 지표의 한계로 인해 경향성 및 추세성을 표현하기 어려움



8. scatter plot

- 대용량 데이터 셋 내부의 상관관계를 표시할 때
- 독립변수를 x 축에, 종속변수를 y 축에 표시하고, 추세선도 함께 표현
- 데이터 셋이 많고, 상관관계가 있는 경우에만 권장

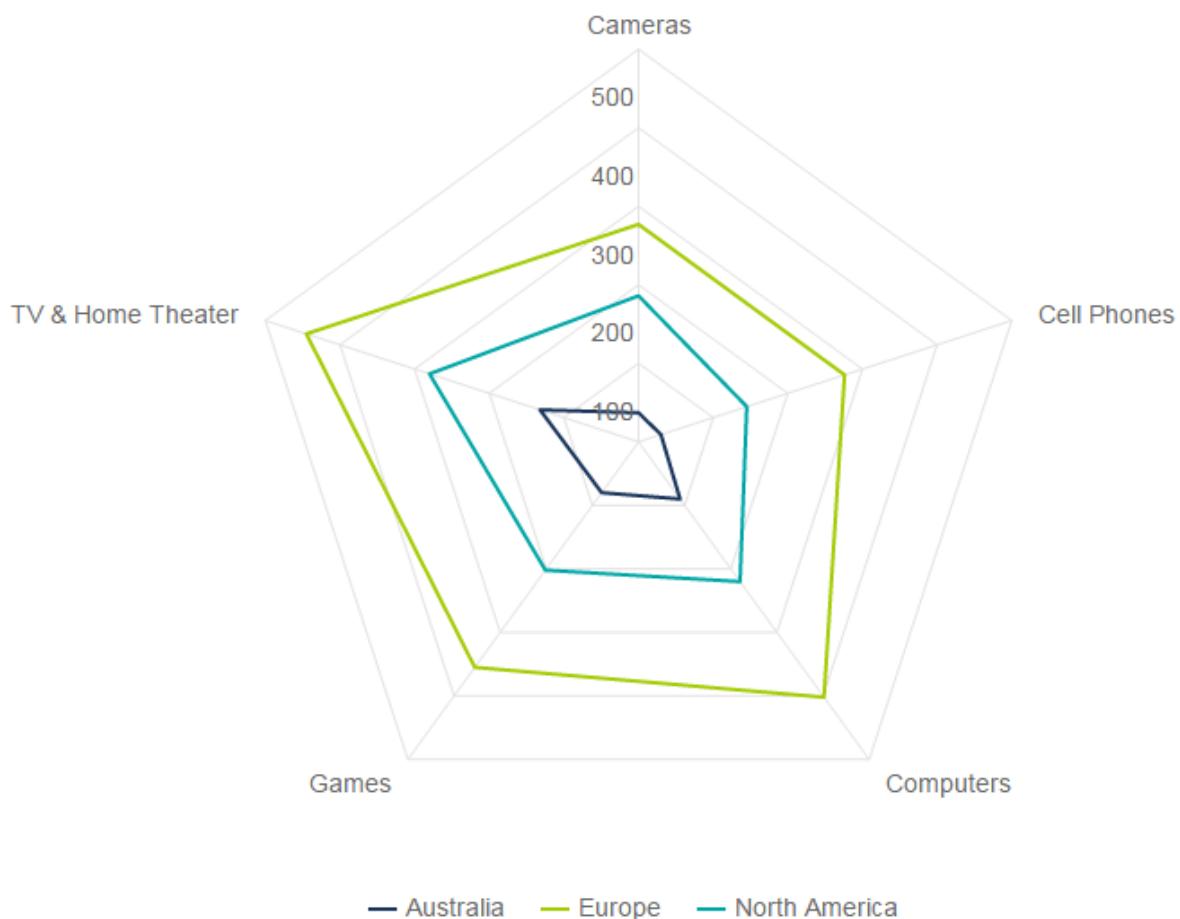
Average Basket Size per Age (last year)



9. spider chart

- radar chart
- 다변량 데이터의 균형감 비교를 위한 차트
- 축 단위가 동일해야 함
- 변수가 5개 초과하여 많아지면 이해하기 어려워짐

Products Sold per Category and Country (last year)



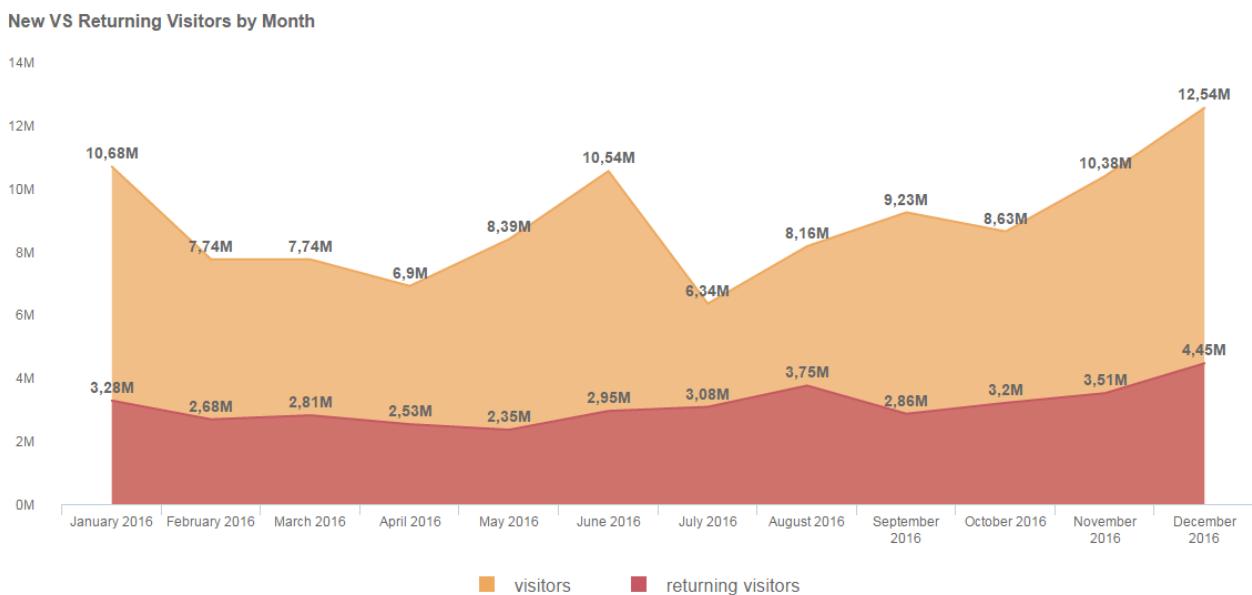
10. tables

- 실질적으로 시각화 도구가 아님
- 여러 개의 정확한 수치와 단위를 표시하기 위하여 사용
- 독자가 자신의 정확한 수치를 확인하고 싶어하는 경우에 사용
- 독자가 읽어야 한다는 상황에 짜증을 낸다면 좋지 않음

Date	Revenue	ProductID
August 2015	€64.468,00	152
September 2015	€172.328,00	402
October 2015	€173.353,00	367
November 2015	€154.950,00	340
December 2015	€116.263,00	287
January 2016	€176.746,00	404
February 2016	€169.579,00	391
March 2016	★ €214.476,00	484
April 2016	★ €239.049,00	551
May 2016	★ €232.288,00	552
June 2016	★ €198.438,00	462
July 2016	★ €199.840,00	470
August 2016	€101.741,00	249
TOTAL	€2.213.519,00	5.111

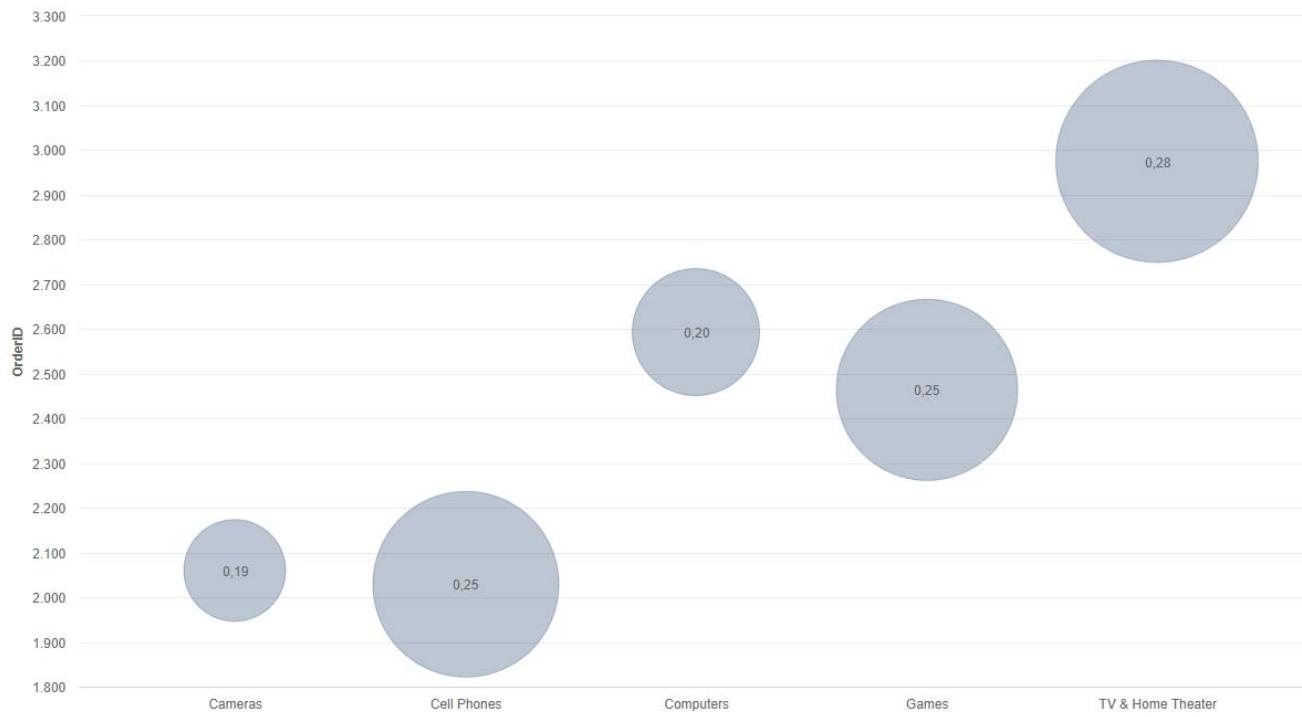
11. area chart

- line chart와 유사하게, 시계열 관계에 기반함
- 개별 수치보다는 추세성 및 경향성을 강조할 때 권장
- 층으로 쌓여서 전체-부분의 비중 관계를 가질 때가 아니라면 권장하지 않음
- 층이 겹치는 경우 곤란

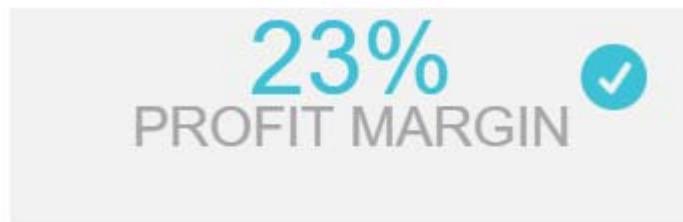


12. bubble chart

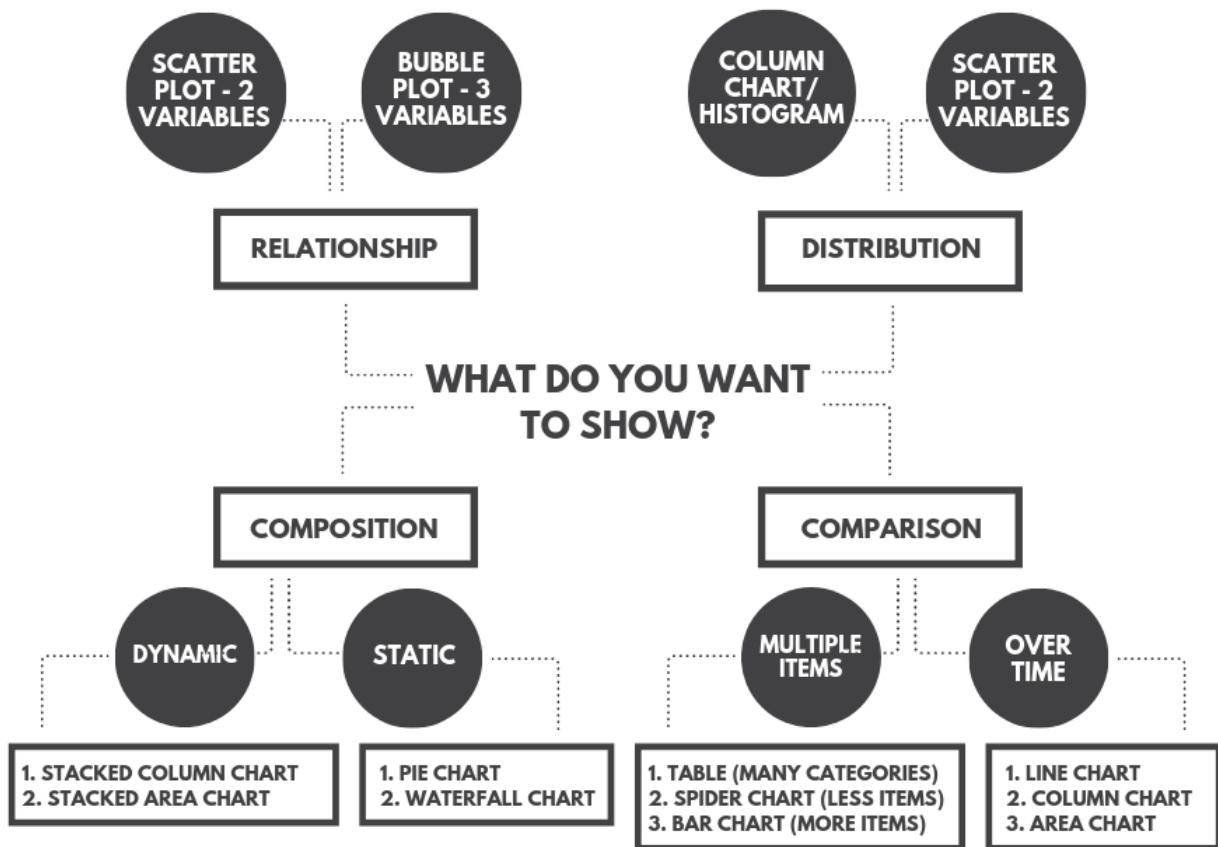
- 다수 값의 대소 관계를 한눈에 비교할 수 있도록 하는 방식
- 아래 차트에서는 가전 제품의 주문량 및 이윤율을 동시에 표시하고 있으며, 높게 위치할수록 주문량이 많고, 원이 클수록 이윤율이 크다는 의미임
- 버블의 면적이 중요하지, 지름이 중요한 것이 아니라는 점에 주의



- 차트에 멋진 텍스트 박스를 추가하여 마무리 하기



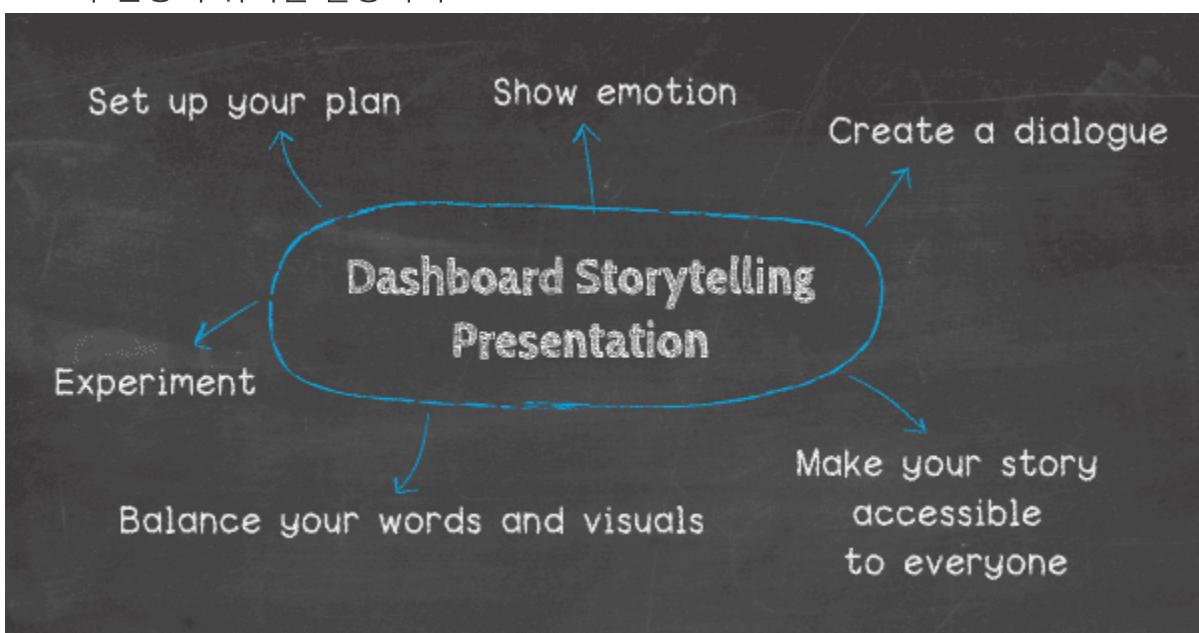
4.2.2 체크 리스트 7종



1. 어떤 스토리를 전달하려고 하는가?

데이터 스토리-텔링에 관해서, [dashboard presentation](#) 지침을 참고

- 데이터 시각화로 시작하라.
- 독자를 최우선으로 존중하라.
- 공허한 말(기술)을 내세우지 말라.
- 스토리-텔링의 위력을 활용하라.



2. 누구를 향해 말하는가?

- 독자(청중)에 대해 연구하라.
- 누가 나의 청중인가?

3. 추세(경향)에 대한 분석인가?

- line chart
- column chart
- area chart

4. 데이터의 구성 비율에 대한 분석인가?

- pie chart
- spider(radar) chart
- bar chart
- columned visualization
- scatter plots

5. 여러 값 집합을 비교하려는가?

- bubble chart
- spider chart
- bar graph
- columned visualization
- scatter plots

6. 시간 축이 필요한가?

- dynamic line chart
- bar chart

7. KPI(핵심 성과 지표)가 문제인가?

- 관계성을 보여주거나 비교하려고 하는가?
- 경향성을 보여주려는가?

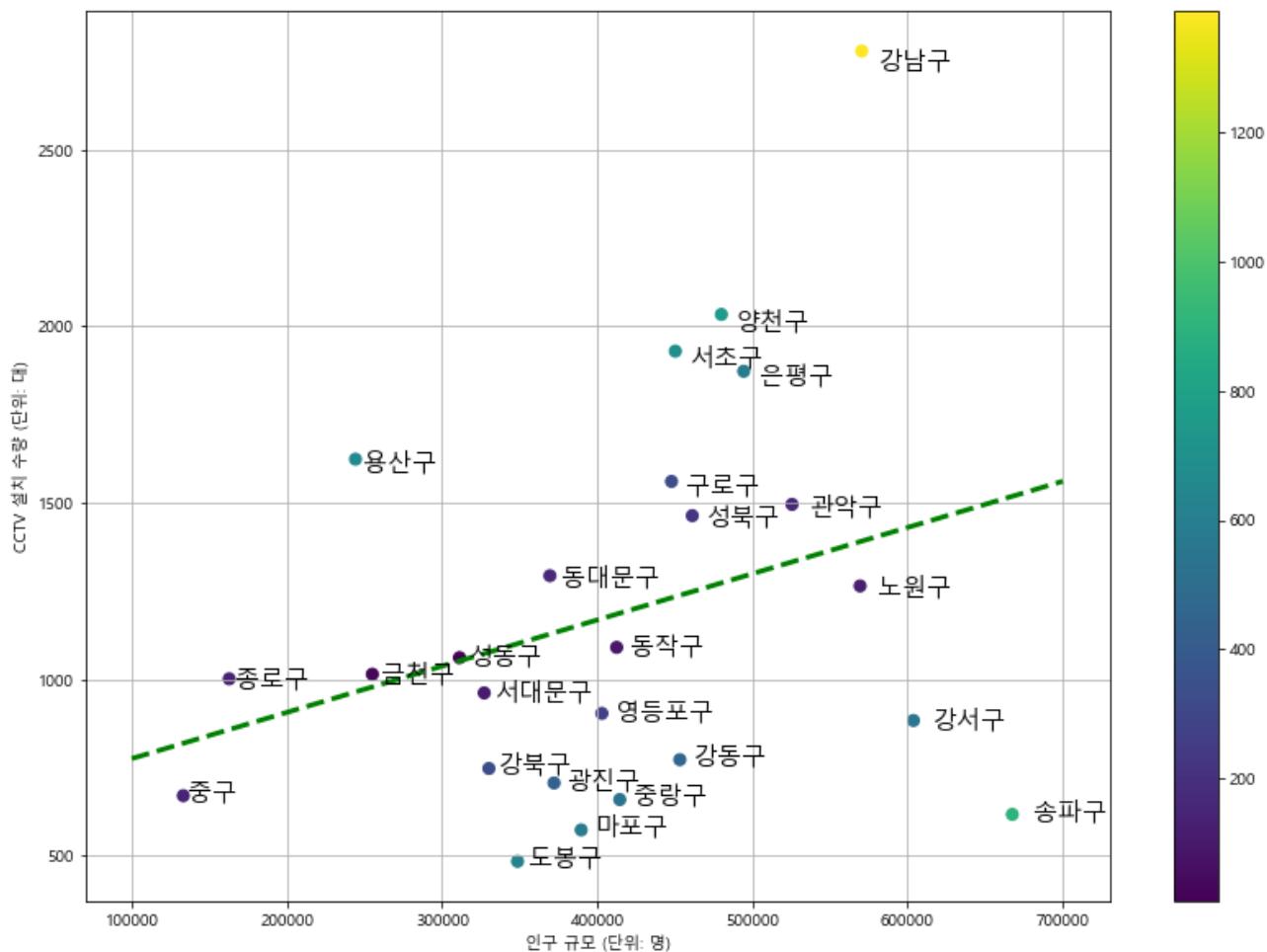
4.3 추천 자료

- 180118, SD아카데미, 빅데이터 시각화 종류 및 도구 정리
- 140428, kingdom99, 02. 데이터 시각화의 기본 원리
- 140602, kingdom99, 03. 다양한 데이터 시각화의 표현 방법
- 170401, 송한나, [The Art of Data - 2회 이럴 땐 이런 그래프: 데이터에 맞는 시각화 방법](#)
- 190725, 임월드, [데이터 시각화 종류 \(기초\) _ matplotlib\(seaborn\)](#)
- ?, Qlik Sense, 목적에 적합한 유형의 시각화 사용
- ?, tableau, [데이터 시각화 현장 가이드: 정의, 예제, 학습 리소스](#)
- [5 Types of Big Data Visualization Categories](#)
- [THE TOP 10 TYPES OF DATA VISUALIZATION MADE SIMPLE](#)
- [Top 16 Types of Chart in Data Visualization
\(상호작용형 차트 예시\)](#)
- [Data Visualization: Visualization Types](#)
- 190302, Sandra Durcevic, [How to Choose the Right Data Visualization Types](#)
- 140314, 도나 M. 웡, [월스트릿저널 인포그래픽 가이드, 인사이트](#)

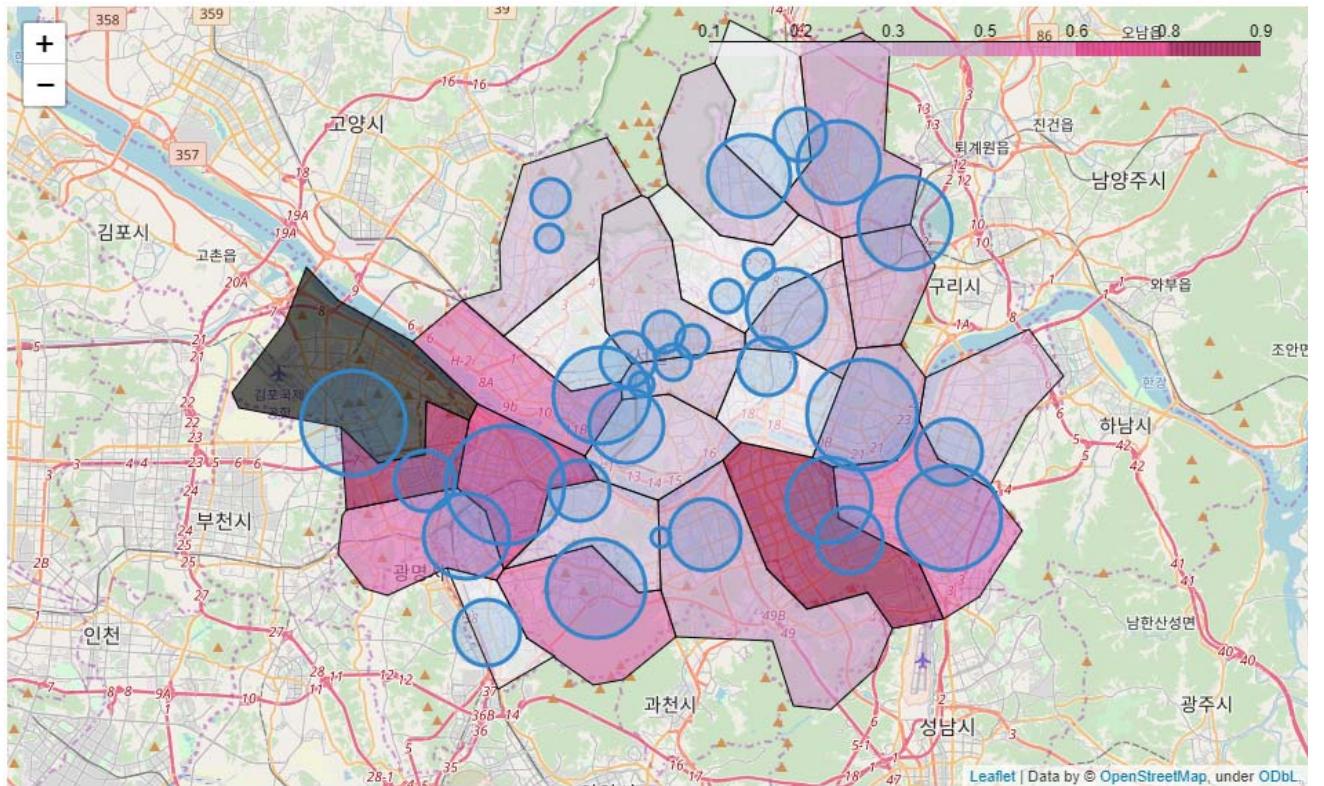
5. 강의 및 평가 계획

5.1 강의 계획

- 공식 교재: 171229, 민형기, [파이썬으로 데이터 주무르기](#), 비제이퍼블릭
- 강의 노트: 200000, 신해웅, <https://github.com/logistex/vd>
- 강좌 내용 구성
 - 강좌 소개 및 준비
 1. 강좌 소개 (이번 강의)
 2. [실습 환경 구축](#)
 - [Anaconda 설치 안내](#)
 - [Anaconda 가상환경 관리](#)
 - 교재 실습
 - 3. [CCTV 설치 규모 분석](#)



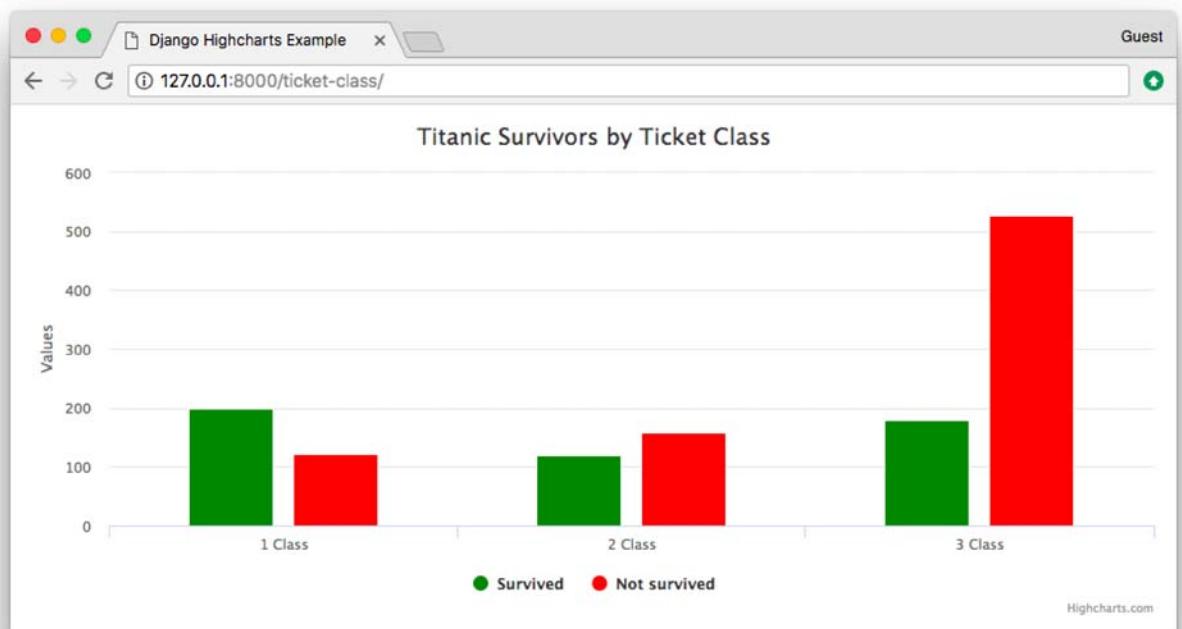
4. 범죄 빈도 분석

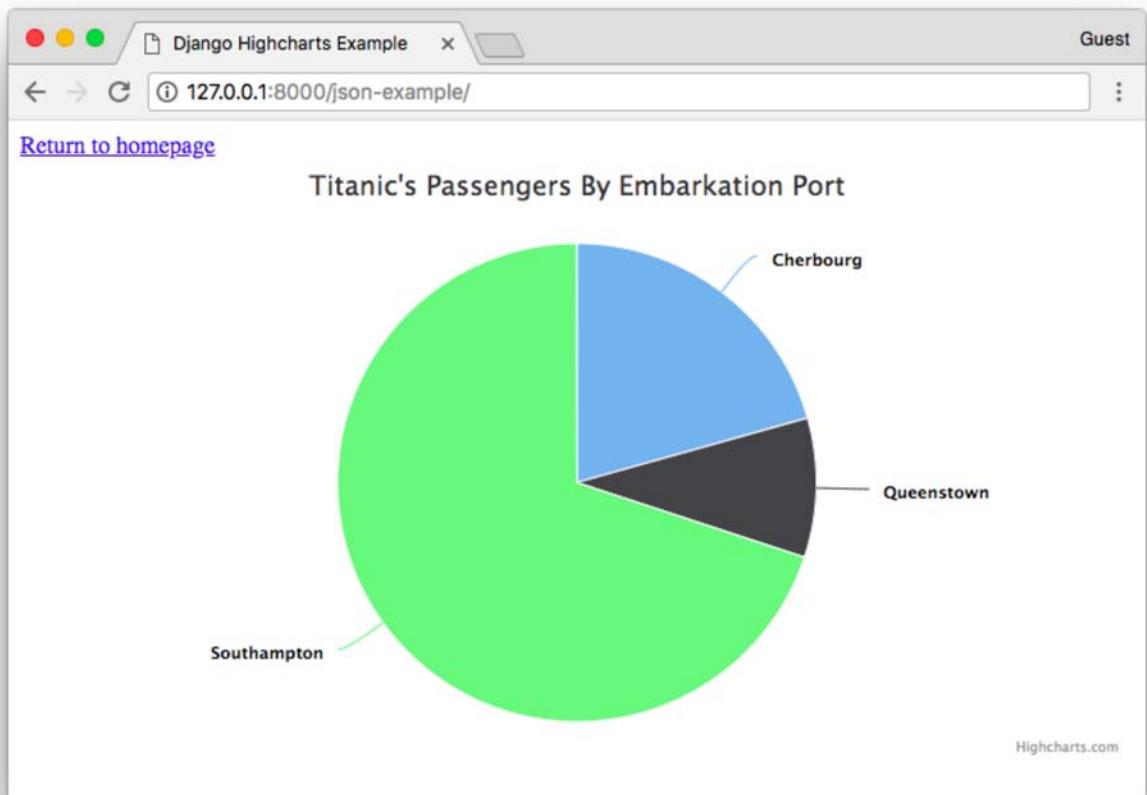


- 인구 대비 범죄 빈도가 높을수록 붉은 색
- 원이 클수록 검거력이 높은 경찰서
- 검거율을 범죄 발생 건수 대비 검거 건수의 비율로 처리해야 하지 않을까?

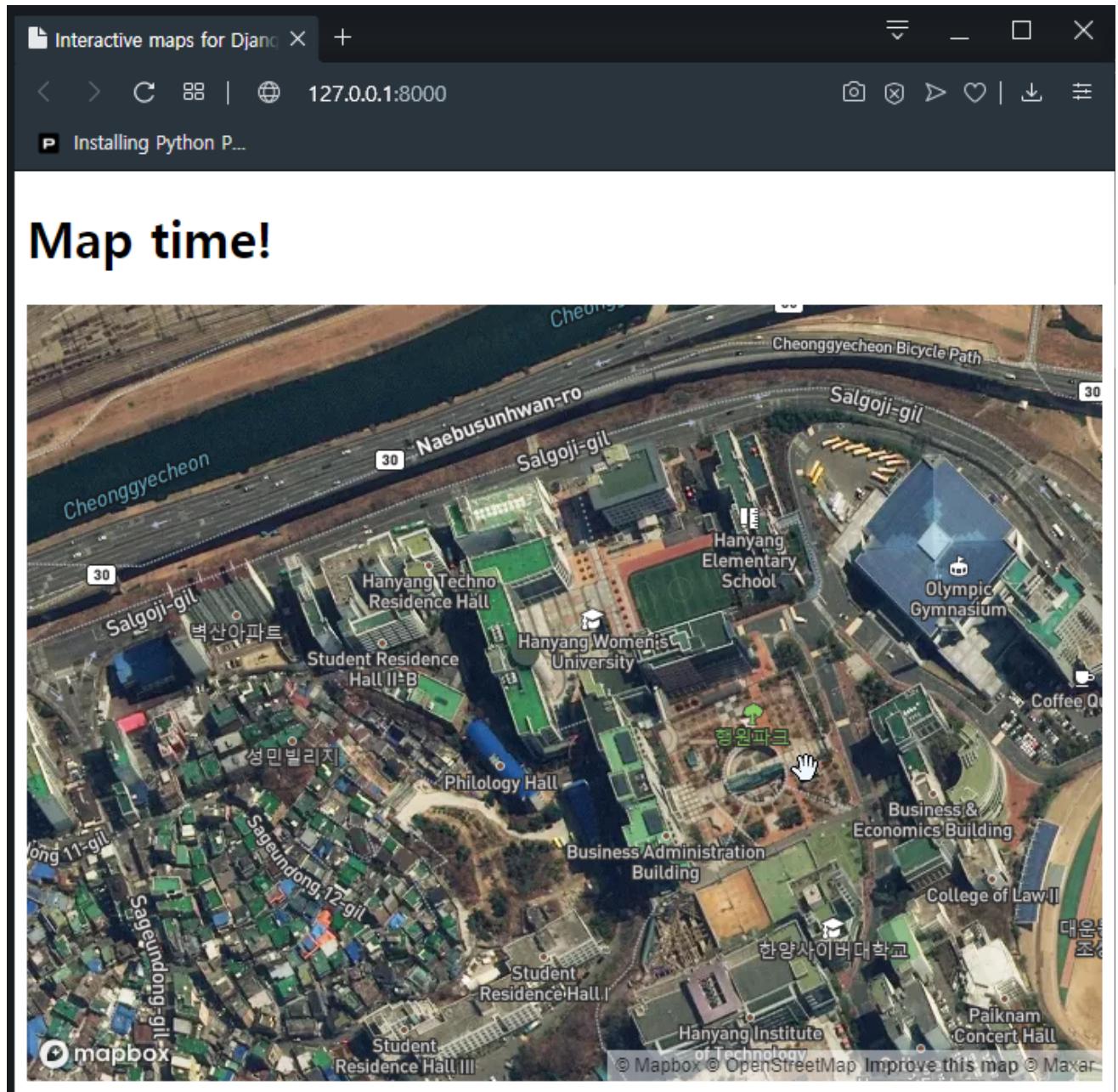
o 장고 데이터 시각화

5. Highcharts With Django





6. Mapbox with Django



Map time!

7. GeoDjango by Ahmed

Add shop

Name:

Location:



Delete all Features

Address:

City:

Nearby Shops

- Brickell City Centre : 624.22206095 m
- Brickell Ace Hardware : 689.90471889 m
- Publix : 835.30059796 m
- Whole Foods Market : 1200.23735559 m
- Whole foods : 1223.04928068 m
- Ross : 1310.74954076 m

8. map with Django by Hakim

Django administration

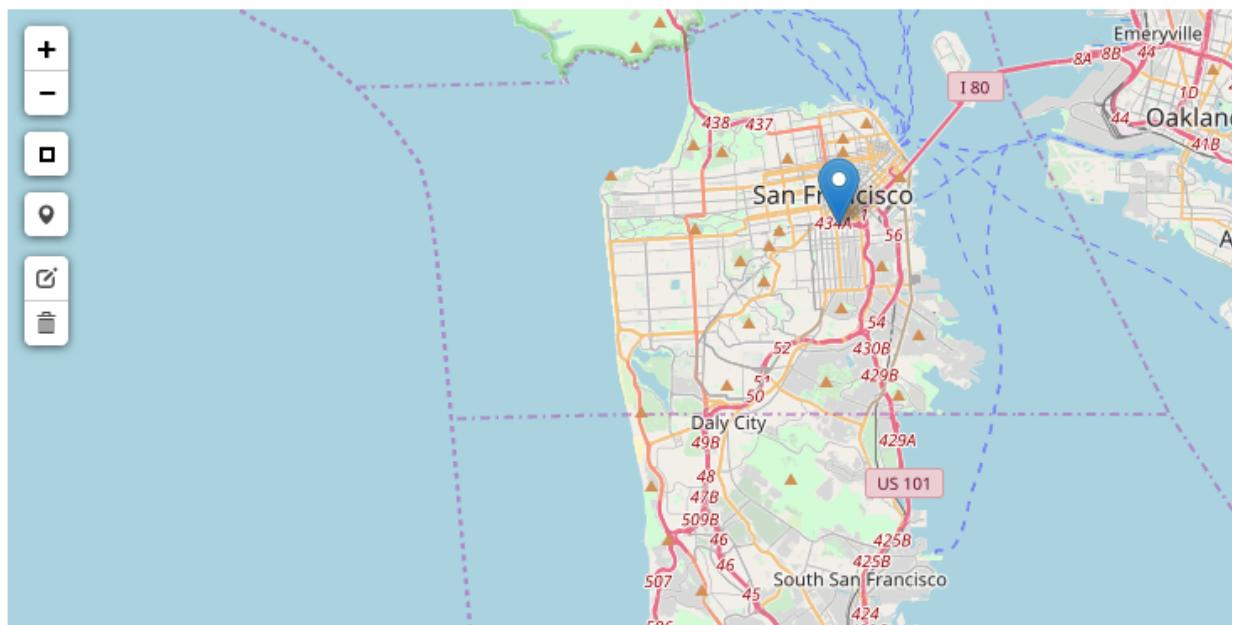
Home > Cities > Cities > City object (7223)

Change city

Name:

San Francisco

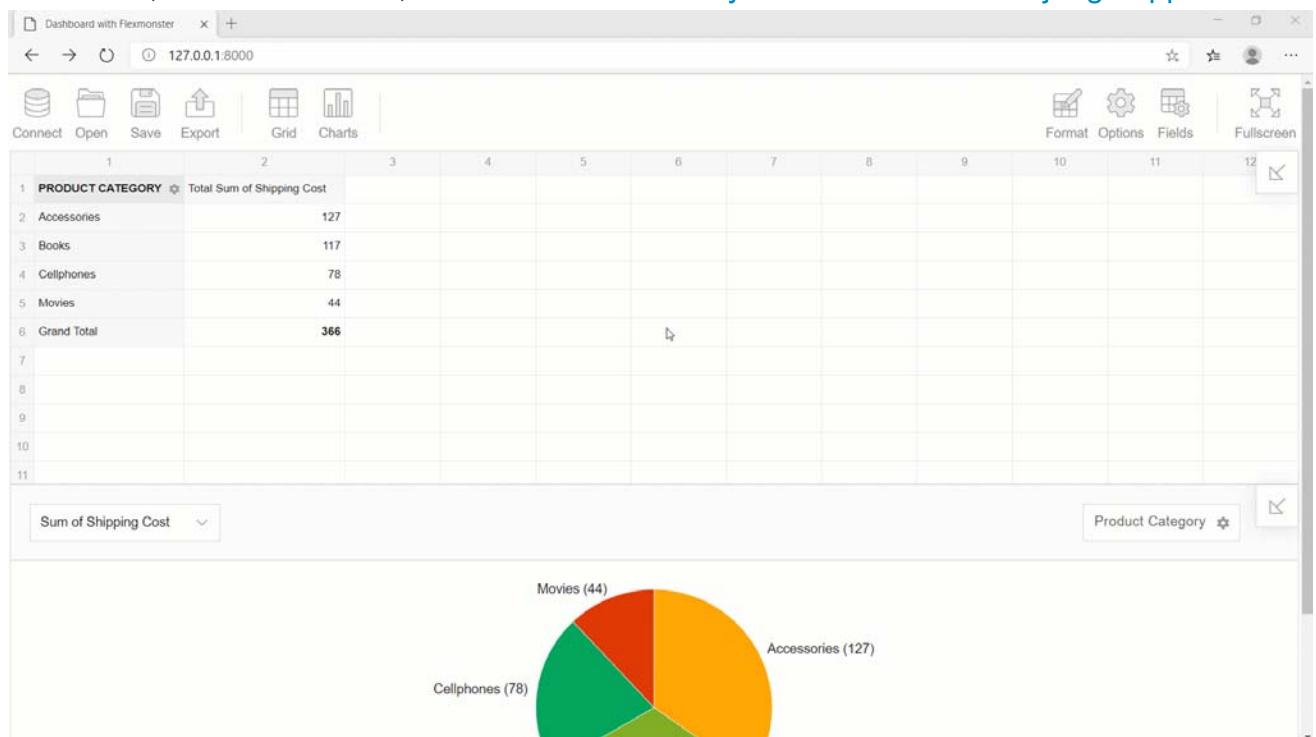
Geometry:



- D3 Map with Django
- 장고 지도 관련 참고 자료 목록

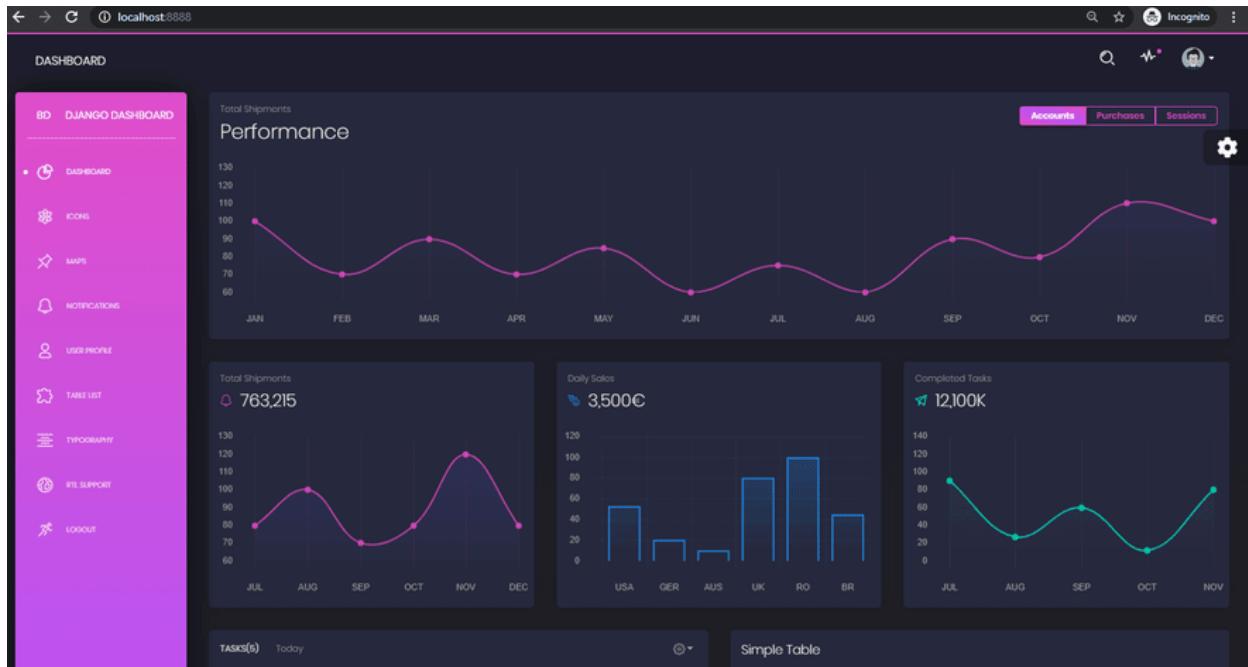
o Dashboard

9. 200212, Veronika Rovnik, [How to create an analytics dashboard in a Django app](#)



- Welcoming Dash 1.0.0
- django-dash

10. 200117, Sm0ke, Django Dashboard - Learn by Coding



11. 190312, David Comfort, [How to Build a Reporting Dashboard using Dash and Plotly](#)

12. 160618, Clinton Dreisbach, [Building dashboards with Django and D3](#)

5.2 평가 계획

구분	배점	평가의 초점
출석	15	성실성
중간	25	기본 개념의 이해 및 전반적 흐름의 파악
기말	30	"
평소	40	문제 해결 능력 및 의지
합계	100	

6. 마무리

- 이번에 공부한 내용
 - 데이터 시각화란 무엇인가?
 - 왜 배워야 하나?
 - 무엇을 배워야 하나?
 - 알아야 할 원리는?
 - 강의/평가 계획
- 다음에 공부할 내용
 - 실습 환경 구축
 - CCTV 설치 규모 분석

