# 내가 해온 연구에 데이터 분석 끼얹기

Attaching Data Analysis on the Research I've Working On

KIER 이제현 선임연구원





### N차 산업혁명





- 2016
  - 4차 산업혁명 (Claus Schwab, Davos Forum)

21세기

초반

AlphaGo (DeepMind)



### N차 산업혁명

### • 같은 시대 다른 풍경

#### 사진으로 동영상 제작: Deep Nostalgia





https://www.myheritage.co.kr/

#### 자율주행차 : Tesla



https://www.tesla.com/

#### 문장으로 디자인 제작: "아보카도 닮은 의자" DALL.E



https://openai.com/blog/dall-e/

#### 엑셀 대신 계산기 사용 권유 : 한국의 모 상사



김대리. 내가 감히 조언 하고 싶은것이 있읍니다. 다른것이 아니고, 너무 엑셀 팡션? 사용 하기 마세요. 편리함이 있다면, 위험성은 증대하죠. 소를 잡는데는 그만한 칼날이 있고 닭잡는데는 칼이 필요 한가요? 쉬운것이 정답 일수 있읍니다.

MMS 오후 2:02



Data 취합, 정리, 단순한 방법 있어요. 별, 시간도 필요 없고, 나중 아날로그 방법도 있죠.

오후 2:09



김대리가 전쟁터에 장군이라 가정하죠. 전쟁에서 이겨야 하는것은 당연 한것 아닌가요? 그 상황에 맞는 전략? 기상군으로만 게압한다? 아니죠. 저의 의견은 암산이 빠를수 있고, 물론 사람에 차이는 있지만, 계산기가 좋을수 있죠. 컴퓨터는 소잡는 칼 아닌가 해서 의견 드립니다.

MMS 오후 2:25



메시기를 입력하세요





인터넷 커뮤니티 캡처

### Two Labs

# WET LAB (VS)

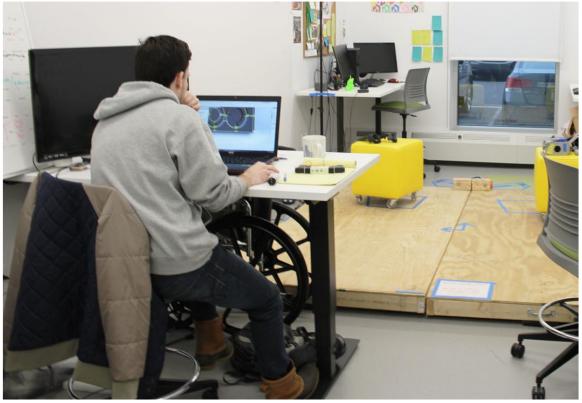
A wet lab is one where drugs chemicals, and other types of biological matter can be analyzed and tested by using various liquids.



# DRY LAB

A dry lab environment focuses more on applied or computational mathematical analyses via the creation of computer-generated models or simulations.





ÜLP

https://www.crbgroup.com/insights/science-technology/lab-design, https://americangardener.net/difference-between-dry-lab-and-wet-lab/, https://southerntierincubator.com/dry-lab/

## **Snapshots of Wet Laboratory**







## Robot Chemist (Univ. of Liverpool, 2020)

- 8일간 688회 실험: 휴식 없이 풀타임 근무. 인간보다 1000배 빠름
- H/W 비용 15만달러 + S/W 3년 〈 정직원 2명 (평균 연봉 95,091천원)



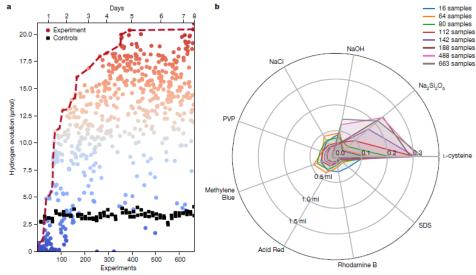


Fig. 3 | Output from the autonomous robotic search. a, Plot showing hydrogen evolution achieved per experiment in an autonomous search that extended over 8 days. Sixteen experiments were performed per batch, along with two baseline controls. The baseline hydrogen evolution was 3.36 ± 0.30 µmol (black squares). The maximum rate attained after 688 experiments was 21.05 µmol h 1. The robot made 430 moves between stations

and travelled a total distance of 2.17 km during this 8-day experiment. **b**, Radar plot showing the evolution of the average sampling of the search space in millillitres; the scale denotes the fraction of maximum solution volume dispensed. The starting conditions (Batch 1) were chosen randomly. The best catalyst formulation found after 43 batches contained P10 (5 mg), NaOH (6 mg), L-cysteine (200 mg) and Na,Si,O<sub>3</sub> (7.5 mg) in water (5 ml).

### 신소재 개발 자율 실험

### • 삼성전자 종합기술원

#### Al set for material breakthroughs

The institute is also working on autonomous R&D, concentrating on autonomous material development (AMD). SAIT is working on an AI algorithm that after learning from experimental data sets, will be capable of driving robot synthesis tools.



Autonomous materials development (AMD) at SAIT. © SAMSUNG

"In about 10 to 20 years, SAIT hopes to reach the point where all the synthesis can be done by machine and human researchers can instead focus on things that require superior creativity," Hwang says.

"It's a promising way to sift through the many possible combinations of elements in different configurations for new materials, a task that would present a huge analysis challenge for humans," says InTaek Han, SAIT senior vice president, who is leading the AMD project.

"Al can utilize enormous datasets in a very short time and search unexplored experimental spaces that humans would have never considered," Han says.

The boost from the AI platform will add to an already impressive output of materials research successes — centred around quantum dots, 2D materials, and other nanostructures — from Han's team.

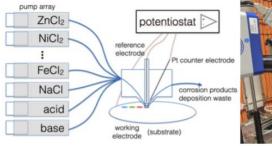
### NIST

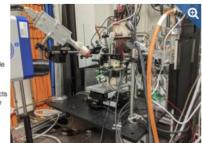
#### **Autonomous Materials Science**

#### Summary

We use autonomous experimentation (the merger of automated synthesis, characterization, Al-driven decision-making) to elucidate the role of composition, processing, and microstructure on the aqueous corrosion of complex metal alloys. To do this we are building a one-of-a-kind combined synthesis-characterization platform that is capable of making and measuring samples on demand. Our autonomous platform is meant to serve as a substrate for answering questions about trust and interpretability of modern Al systems, as well as bias and differences in human interpretation of data in materials science. Through this project we hope to lay the groundwork for autonomous experiments that answer scientific questions, which will be enabled by the future development of scientific Al imbued with materials science knowledge.

#### **DESCRIPTION**

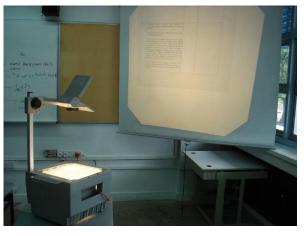




# 2차 (전기) → 3차 (컴퓨터, 인터넷) → 4차 (데이터, AI)

• 1990년대, 2000년대, 그리고.







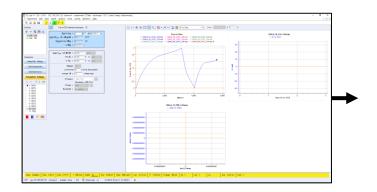






### 2차 (전기) → 3차 (컴퓨터, 인터넷) → 4차 (데이터, AI)

• 1990년대, 2000년대, 그리고.







"우리 때는 이런 걸로 논문 쓰고 연구했어!" 1차 산업혁명 시대 도구

8. Li er el

800 (80) 12Nb fiber
900 (80) 1.5Nb fiber
900 (80) (80) 1.5Nb f

Fig. 3. Street-strain curves of the TI-187z-(10-12)Nb 25n alloy fiber specimens obtained by leading unloading tensile tests: 12Nb (a), 11.5Nb (b), 11Nb (c) and 10Nb (d).

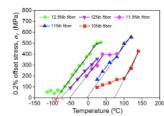


Fig. 4. Temperature dependence of the 0.2% offset stress (σ<sub>Y</sub>) in the TL-16Zr-(10–12.5)Nb-28n alloy fiber specimens.

Fig. 3 shows the stress-strain curves of the 12.50h fiber specimens obtained by rentile tests which were carried out as a wide temperature range from  $-115^\circ$  Cro  $95^\circ$  Cr. Here, it should be montioned that the stress-strain curves obtained  $a = 30^\circ$  Cr.  $-10^\circ$  Cro  $10^\circ$  S Cra F from our provious study [22] and cited in fits article for result completeness purposes. As obtain in Fig.  $200^\circ$ -Cr., the superleasticity, i.e. the strain that is developed through a planeau-like range upon loading it recovered superleastically upon unloading, is observed clearly in a wide

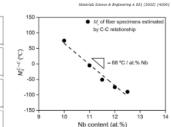


Fig. 5. Nb content dependence of M<sup>2-C</sup> estimated by Clausius Clausius (C-C)
relationship in the Ti-182r-(10-12.5)Nb-25n alloy fiber specimens.

emporature vindove of 100 °C, ranging from 65 °C to 55 °C, Fig. 3 belower the recent strain curves of the 12 1000 biter specimens, the super-laterity is disserved clearly at temperature from 50 °C to 25 °C. In the 11,300 fiber specimen, although the superstandingly is observed at temperature from 50 °C to 60 °C. The superclasedirty starts cappers at 80 °C in the 1108 fiber specimen and 100 °C in the 1000 fiber specimen. To estimate the relationship between the apparent yield roses and temperatures of the fiber specimen, the apparent yield roses (see junificated by an arrow is measured by using the 1,200 fiber methods as thowas in Fig. 400. The cy is measured byte using the 1,200 fiber in the stown in Fig. 80. The cy is no accurate from control curve or the fiber specimens and then plotted as a function of test temperature as shown in Fig. 4.

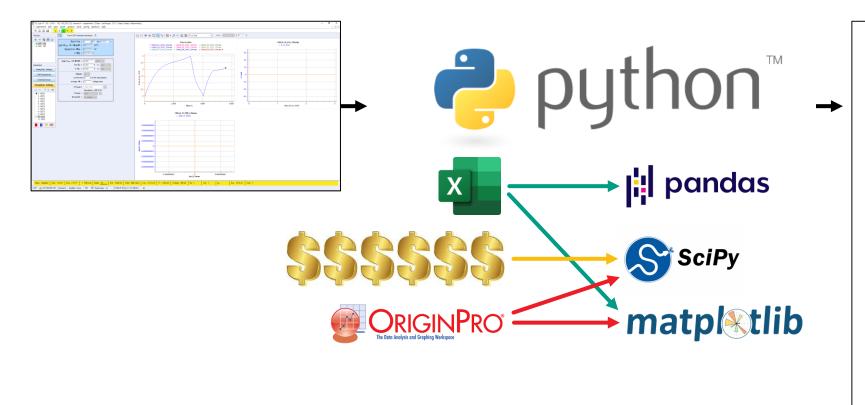
in the 12-350 fiber specimen, we increase illustrative with increasing superature from ~78 (Tel. pt 24, abus > 15 (Tel. fibers a beliatively) deviates, which is caused by the presence of plante obternation severing inclining basiline, the relief attention for individual period of the process of the proces

any LDs. Seals during colouing also beauting processes are no necessive in the rapidity solitised 12.000 half Silbert as reported in our previous study [13], to 1200 half 11.500 hiller aperiment, the C-C relationship is obtained at empirerature range from  $-0.00^\circ$  C = 20°C, and  $-3.00^\circ$  C, respectively.  $M_{\gamma}^{\rm C}$  of the 1200 and [1,300 hiller speciment are measured to be  $-75^\circ$  C and  $-3.0^\circ$  C, respectively. In 1100 and 1000 filber speciments, partially superelastic stress-train curves are domined at 00–120° C in 1100 hiller speciment and 1001–100° C in 1000 hiller specimen and another state of the control of the contro

https://www.originlab.com/origin, https://ko.wikipedia.org/ Lee et al, https://doi.org/10.1016/j.msea.2021.142001

### 2차 (전기) → 3차 (컴퓨터, 인터넷) → 4차 (데이터, AI)

• 20년 전과는 달라져야 하지 않겠습니까?



https://www.originlab.com/origin Lee et al, https://doi.org/10.1016/j.msea.2021.142001

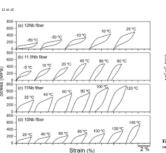


Fig. 3. Stress-strain curves of the TI-182s-(10-12)Nb-29n alloy fiber specimens obtained by loading unloading tensile tests: 12Nb (a), 11.5Nb (b), 11Nb

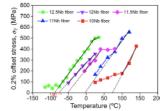


Fig. 4. Temperature dependence of the 0.2% offset stress  $(\sigma_{\rm F})$  in the 11.162r-

rapidly solidified T1-182r (10-12.5)ND-28n fiber specimens. 10ND fiber specimen consists of the minor  $\beta$  and dominant  $\alpha'$  phases as aboven in Fig. (13). The intensity of  $\alpha'$  postal decreases and that of  $\beta$  postal receives an ASI of the content increases from 10 st.Ni to 11.5 st.Ni. When the Nb content increases to 12 at.Ni, the  $\alpha'$  position are observed, which indicates that the  $\beta'$  —  $\alpha'$  matronified transformation start emperature (Ai) in the 12ND fiber specimen decreases to below the toom temperature. (Ai)  $\alpha'$  position are observed in the 12ND fiber specimen decreases to ASI of Position are observed in the 12ND fiber specimen decreases the AI of the 12ND fiber specimen decreases the 12ND fiber specimen decreases the AI of the 12ND fiber specimen dec

Fig. 2 shows the stress-strain curves of the 12.8Nb fiber specimens obtained by remile tests which were extined out as a wide temperature range from −115° Cro 95° C. Hers, it should be mentioned that the stress-strain curves obtained at −20° C, −10° C and 9° C are from our provious study 1201 and cited in this article for result completeness purposes. As shown in Fig. 20° C-p, the superleasticity, is, the tuttin that is developed through a plateau-like range upon loading it recovered superleastically upon unloading, is observed clearly in a wide

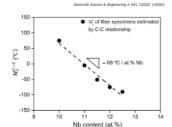


Fig. 8. Nb content dependence of M<sub>1</sub><sup>C</sup> of estimated by Clausius Clapsyron (C-C)

emporature vindovo el 100 °C, ranging from 65 °C to 55 °C. Ple. 9 bowork for terrors strain curves of the 12 1000 blee specimens. The beta 2000 filter specimens, the superstanting is observed elsewly at temperature from 50 °C to 20 °C, the 11.700 blee specimens, although the superstanting is observed at temperature ranges from 5°C to 100 °C, in the compensation range from 5°C to 100 °C. The superclasticity rater to appear at 80 °C in the 13.700 blee in 1500 blee specimen 30 °C to 80 °C. The superclasticity rater to appear at 80 °C in the 1300 blee specimen and 100 °C in the 1000 blee specimens. The stimulate the radiationship intervent the supparent spidel stress small est compensations of the threat specimens of the specimens. The appearant spidel stress small set compensations of the 1000 blee specimens and them gloss of the 1000 blee specimens and them gloss of the 1000 blee specimens should be spinted as a function of less the specimens and them glosted as a function of less the specimens and them glosted as a function of less the specimens and them glosted as a function of less the specimens and them glosted as a function of less the specimens and them glosted as a function of less the specimens and them glosted as a function of less the specimens and them glosted as a function of less the specimens.

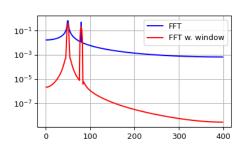
In the 12.5Nb fiber specimen, or increases linearly with increasing

temperature from -75 °C to 15 °C. Above 15 °C, the linear relationship deviates, which is caused by the presence of plastic determation accurring during leading. The critical stress for inducing plastic defor mation is measured to be about 470 MPa, which is higher than that (990 MPa) in the referenced bulk specimen [1]. The linear relationship be tween -75 °C and 15 °C follows the Clausius-Clanevron (C-C) relationship widely applied in SMA systems to analyze the stress induced transformation removaring of the 12 5Nb alloy fibers, an extranolation method is used for the C-C relationship as reported in Ti-Ni-based [29] and nickel-free Ti-Nb-based [30] SMAs. By extrapolating the C-C relationship to zero or (the dashed line), the temperature that is termed "M; or as shown in Fig. 4, is obtained to speculate the actual M, because any DSC peaks during cooling and heating processes are not observed in the rapidly solidified 12.5Nb alloy fibers as reported in our previous study [18]. In 12Nb and 11.5Nb fiber specimens, the C-C relationship is obtained at temperature ranges from  $-50\,^{\circ}\text{C}\sim25\,^{\circ}\text{C}$  and  $-5\,^{\circ}\text{C}\sim40\,^{\circ}\text{C},$ respectively. Mc-C of the 12Nb and 11.5Nb fiber specimens are measured to be -75 °C and -51 °C, respectively. In 11Nb and 10Nb fiber specimens, partially superelastic stress-strain curves are obtained at 50-120 °C in 11Nb fiber specimen and 100-140 °C in 10Nb fiber specimen and the  $\sigma_T$  almost increases linearly with increasing temper ature in both temperature ranges. Thus, the C-C relationship is applied in these two temperature ranges in 11Nb and 10Nb fiber specimens. MC-C of the 11Nb and 10Nb fiber specimens are measured to be -2 °C and 70 °C, respectively. M; C is ploted as a function of Nb content as shown in Fig. 5. It is found that the increase of 1 at.%-Nb content de creases the Mind by 68 °C in the present fiber specimens. This degree of decrement in martensitic transformation temperature is larger than that

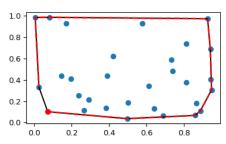
# SciPy로 할 수 있는 일

### • 거의 모든 신호처리, 영상처리

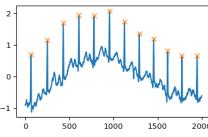
#### **Fast Fourier Transformation**



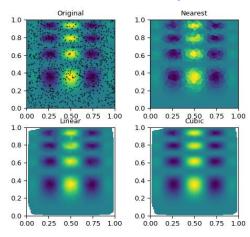
#### Convex Hull



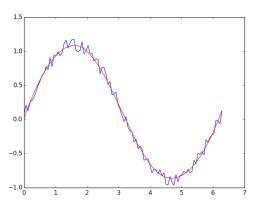
#### **Peak Detection**



#### N-dimensional Interpolation



#### **Smoothing**



https://scipy.org/

### nature methods

**PERSPECTIVE** 

https://doi.org/10.1038/s41592-019-0686-2 There are amendments to this paper

#### OPEN

### SciPy 1.0: fundamental algorithms for scientific computing in Python

Pauli Virtanen<sup>1</sup>, Ralf Gommers <sup>© 2\*</sup>, Travis E. Oliphant <sup>2,3,4,5,6</sup>, Matt Haberland <sup>© 7,8\*</sup>, Tyler Reddy <sup>© 9\*</sup>, David Cournapeau<sup>10</sup>, Evgeni Burovski<sup>11</sup>, Pearu Peterson<sup>12,13</sup>, Warren Weckesser<sup>14</sup>, Jonathan Bright<sup>15</sup>, Stéfan J. van der Walt 914, Matthew Brett 16, Joshua Wilson 17, K. Jarrod Millman 1014, 18 Nikolay Mayorov<sup>19</sup>, Andrew R. J. Nelson<sup>©20</sup>, Eric Jones<sup>5</sup>, Robert Kern<sup>5</sup>, Eric Larson<sup>21</sup>, C J Carey<sup>22</sup>, Ilhan Polat23, Yu Feng24, Eric W. Moore25, Jake VanderPlas26, Denis Laxalde 027, Josef Perktold28 Robert Cimrman<sup>29</sup>, Ian Henriksen<sup>6,30,31</sup>, E. A. Quintero<sup>32</sup>, Charles R. Harris<sup>33,34</sup>, Anne M. Archibald<sup>35</sup>, Antônio H. Ribeiro @36, Fabian Pedregosa37, Paul van Mulbregt @38 and SciPy 1.0 Contributors39

SciPy is an open-source scientific computing library for the Python programming language. Since its initial release in 2001, SciPy has become a de facto standard for leveraging scientific algorithms in Python, with over 600 unique code contributors, thousands of dependent packages, over 100,000 dependent repositories and millions of downloads per year. In this work, we provide an overview of the capabilities and development practices of SciPy 1.0 and highlight some recent technical developments.

ciPy is a library of numerical routines for the Python programming language that provides fundamental building blocks for modeling and solving scientific problems. SciPy includes algorithms for optimization, integration, interpolation, eigenvalue problems, algebraic equations, differential equations and many other classes of problems; it also provides specialized data structures, such as sparse matrices and k-dimensional trees. SciPy is built on top of NumPy1,2, which provides array data structures and related fast numerical routines, and SciPy is itself the foundation upon which higher level scientific libraries, including scikit-learn3 and scikit-image4, are built. Scientists, engineers and others around SciPy, and the M87 black hole imaging project cites SciPy<sup>9</sup>.

Recently, SciPy released version 1.0, a milestone that traditionally signals a library's API (application programming interface) being mature enough to be trusted in production pipelines. This version numbering convention, however, belies the history of a project that

has become the standard others follow and has seen extensive adoption in research and industry.

SciPy's arrival at this point is surprising and somewhat anomalous. When started in 2001, the library had little funding and was written mainly by graduate students-many of them without a computer science education and often without the blessing of their advisors. To even imagine that a small group of 'rogue' student programmers could upend the already well-established ecosystem of research software-backed by millions in funding and many hundreds of highly qualified engineers 10-12—was preposterous.

Yet the philosophical motivations behind a fully open tool the world rely on SciPy. For example, published scripts 5.6 used in stack, combined with an excited, friendly community with a sinthe analysis of gravitational waves import several subpackages of gular focus, have proven auspicious in the long run. They led not only to the library described in this paper, but also to an entire ecosystem of related packages (https://wiki.python.org/moin NumericAndScientific) and a variety of social activities centered around them (https://wiki.python.org/moin/PythonConferences). The packages in the SciPy ecosystem share high standards of

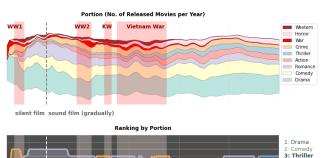
\*University of Jyväskylä, Jyväskylä, Finland. \*Quansight LLC, Austin, TX, USA. \*Ultrasound Imaging, Mayo Clinic, Rochester, MN, USA. \*Electrical Engineering, Brigham Young University, Provo, UT, USA, 5Enthought, Inc., Austin, TX, USA, 6Anaconda Inc., Austin, TX, USA, 7BioResource and Agricultural Engineering Department, California Polytechnic State University, San Luis Obispo, CA, USA. <sup>8</sup>Department of Mathematics, University of California Los Angeles, Los Angeles, CA, USA. ºLos Alamos National Laboratory, Los Alamos, NM, USA. ¹ºIndependent researcher, Tokyo, Japan. ¹¹National Research University Higher School of Economics, Moscow, Russia. 12 Independent researcher, Saue, Estonia. 13 Department of Mechanics and Applied Mathematics, Institute of Cybernetics at Tallinn Technical University, Tallinn, Estonia. 14Berkeley Institute for Data Science, University of California Berkeley, Berkeley, CA, USA. 15Independent researcher, New York, NY, USA. 16School of Psychology, University of Birmingham, Edgbaston, Birmingham, UK. 17Independent researcher, San Francisco, CA, USA, 1th Division of Biostatistics, University of California Berkeley, Berkeley, CA, USA, 1th WayRay LLC, Skolkovo Innovation Center, Moscow, Russia. 20 Australian Nuclear Science and Technology Organisation, Lucas Heights, NSW, Australia. 21 Institute for Learning and Brain Sciences, University of Washington, Seattle, WA, USA. 22 College of Information and Computing Sciences, University of Massachusetts Amherst,

### Matplotlib으로 그릴 수 있는 것

### • 거의 모든 형태의 데이터 시각화

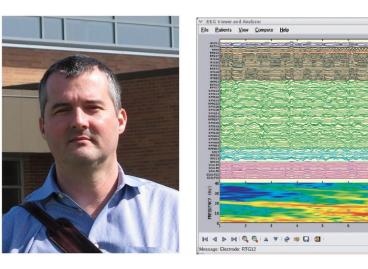






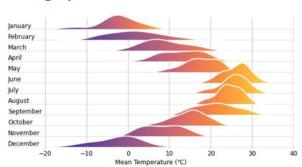
: Action

5: Horror 6: Romance 7: Crime

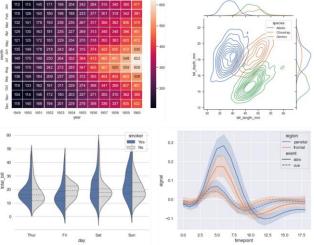




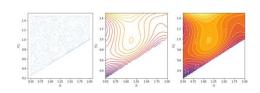








2D Data Distribution



## Python을 쓰기 어려운 분야: 거의 없음











Quantum Computing	Statistical Computing	Signal Processing	Image Processing	Graphs and Networks	Astronomy Processes	Cognitive Psycholog
	~	րկի		M	*	<b>②</b>
QuTiP	Pandas	SciPy	Scikit-image	NetworkX	AstroPy	PsychoPy
PyQuil	statsmodels	PyWavelets	OpenCV	graph-tool	SunPy	
Qiskit	Xarray	python-control	Mahotas	igraph	SpacePy	
	Seaborn			PyGSP		
Bioinformatics	Bayesian Inference	Mathematical Analysis	Chemistry	Geoscience	Geographic Processing	Architecture & Engineering
囡	M	+-	H	<b>(2)</b>	U	<b></b>
BioPython	PyStan	SciPy	Cantera	Pangeo	Shapely	COMPAS
Scikit-Bio	PyMC3	SymPy	MDAnalysis	Simpeg	GeoPandas	City Energy Analys
PyEnsembl	ArviZ	cvxpy	RDKit	ObsPy	Folium	Sverchok
ETE	emcee	FEniCS		Fatiando a Terra		

### Python을 쓰면 가능해지는 일: 머신러닝 포함 Al

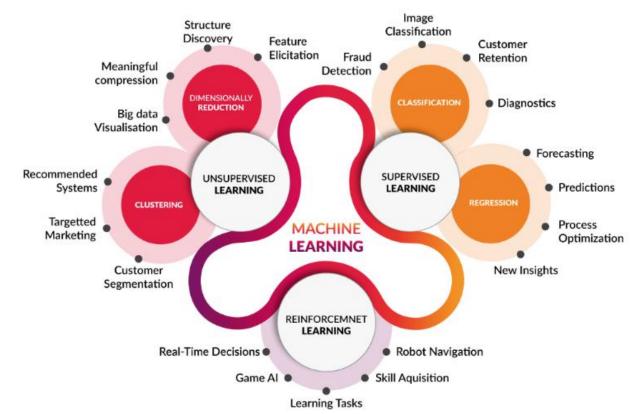








# NumPy <sup>®</sup>







# 감사합니다

내가 해온 연구에 데이터 분석 끼얹기

