

Data science with Animal science

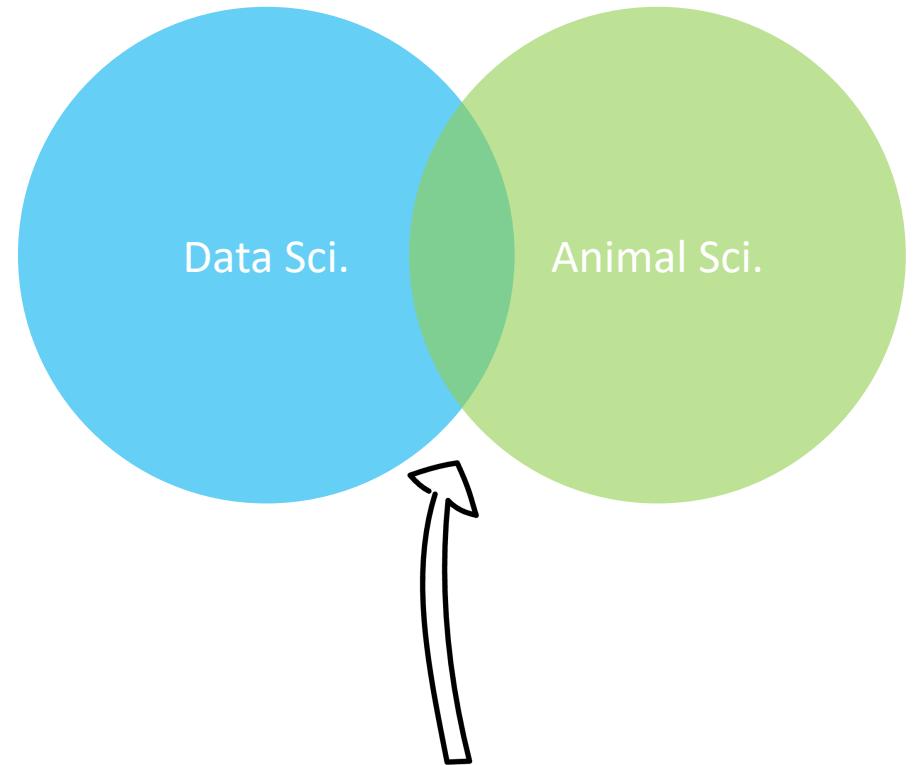
Youngjun Na

Github: <https://github.com/youngjunna>

Email: ruminoreticulum@gmail.com

Introduction:

Data science + Animal science

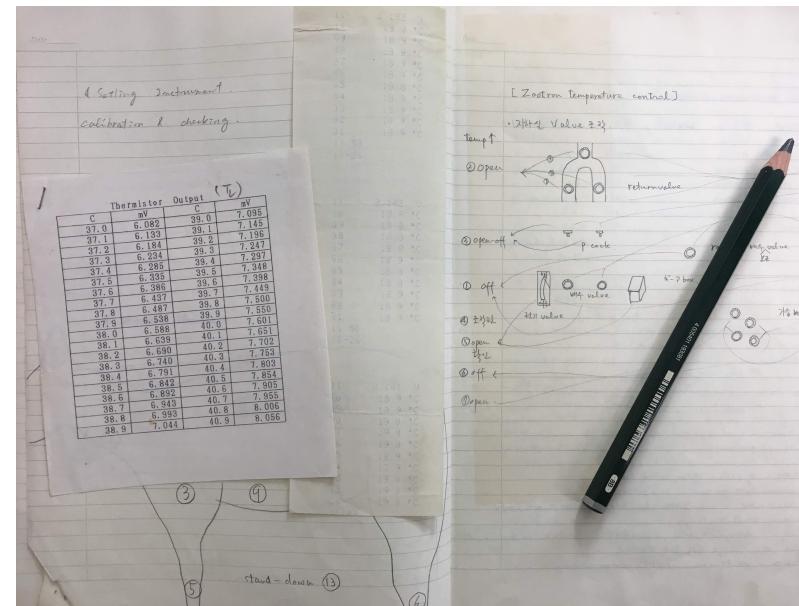


データ?
何?

Data Science + Animal Science

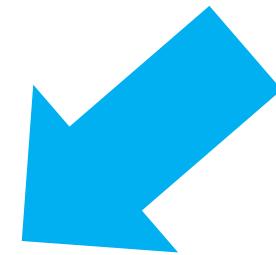
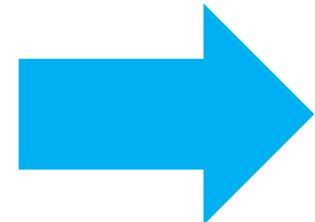
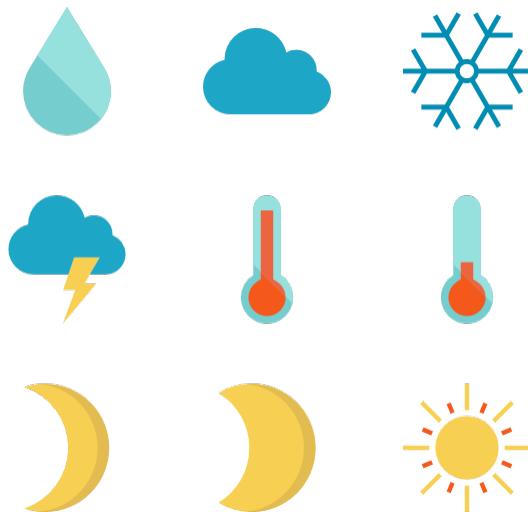
- 이전에는 사람이 손으로 기록하고 관리해야 했던 데이터
- 기록되지 않던 동물들의 정보들이 데이터로 남기 시작함

Insulin RIA	
Sample	
A11 : 12	A101 ~ A112
D2 : 12	A201 ~ A212
D3 : 12	A301 ~ A312
D4 : 12	A401 ~ A412
B-DL : 12	Number : 8
D2 : 12	
D4 : 8	
D5 : 12	D5-1 ~ 12 \leftarrow 8ml \times 2ml \rightarrow 3ml
D6 : 12	D6-1 ~ 12 \leftarrow 6ml \times 3 \leftarrow 1/2 ok
D7 : 12	D7-1 ~ 12 \leftarrow 6ml \times 3 \leftarrow 1/2 X
D8 : 14	D8-1 ~ 14 \leftarrow 8ml \times 2ml \rightarrow 3ml
C-DL : 12	Number : 8
D5 : 12	B5-1 ~ 12 \leftarrow 6ml \times 2ml \rightarrow 3ml
D6 : 12	B6-1 ~ 12 \leftarrow 6ml \times 2ml \rightarrow 3ml
D7 : 12	B7-1 ~ 12 \leftarrow 6ml \times 2ml \rightarrow 3ml
D8 : 14	B8-1 ~ 14 \leftarrow 8ml \times 2ml \rightarrow 3ml
D-DL : 12	Number : 8
D5 : 12	D5-1 ~ 12 \leftarrow 6ml \times 2ml \rightarrow 3ml
D6 : 12	D6-1 ~ 12 \leftarrow 6ml \times 2ml \rightarrow 3ml
D7 : 12	D7-1 ~ 12 \leftarrow 6ml \times 2ml \rightarrow 3ml
D8 : 14	D8-1 ~ 14 \leftarrow 8ml \times 2ml \rightarrow 3ml
E-DL : 12	Number : 8
D5 : 12	D5-1 ~ 12 \leftarrow 6ml \times 2ml \rightarrow 3ml
D6 : 12	D6-1 ~ 12 \leftarrow 6ml \times 2ml \rightarrow 3ml
D7 : 12	D7-1 ~ 12 \leftarrow 6ml \times 2ml \rightarrow 3ml
D8 : 14	D8-1 ~ 14 \leftarrow 8ml \times 2ml \rightarrow 3ml
Total D-DL : 12	Number : 8
Total E-DL : 12	Number : 8



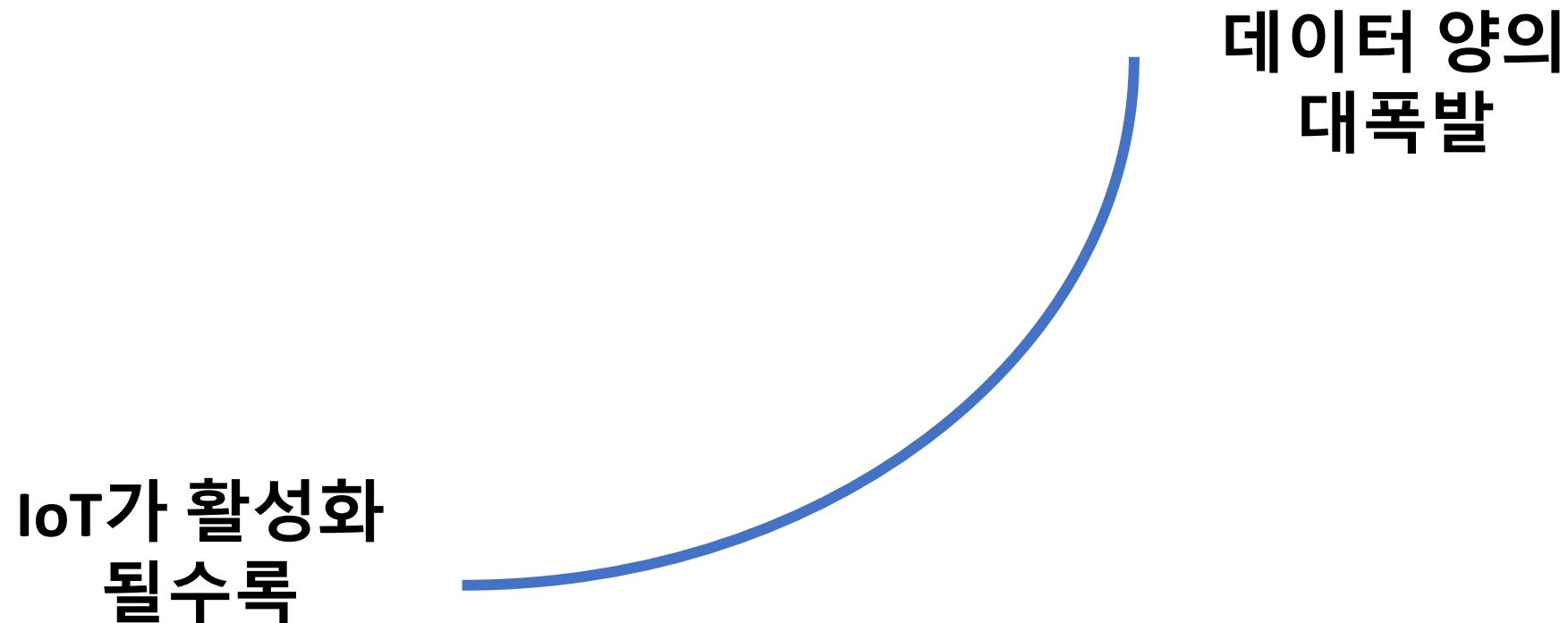
Data Science + Animal Science

- 축산분야에서 IoT의 발달
- IoT의 핵심 == 자동화된 많은 센서(sensor)



Data Science + Animal Science

- IoT (Internet of Things)의 발달로 인한 데이터 양의 대폭발

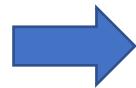


우리나라의 여건 = 수많은 데이터들이 비교적 체계적으로 관리되고 있음



혈통/검정성적

한국종축개량협회
젖소개량사업소
- 혈통정보
- 검정성적
= 유량
= 유성분
= 번식효율
한우개량사업소
- 혈통정보



공공데이터포털



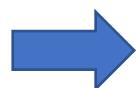
축산물품질평가원 소도체 등급 정보

축산물품질평가원
소도체 정보



분뇨/환경정보

축산환경관리원
가축분뇨 발생 정보
공동자원화 정보
악취 정보

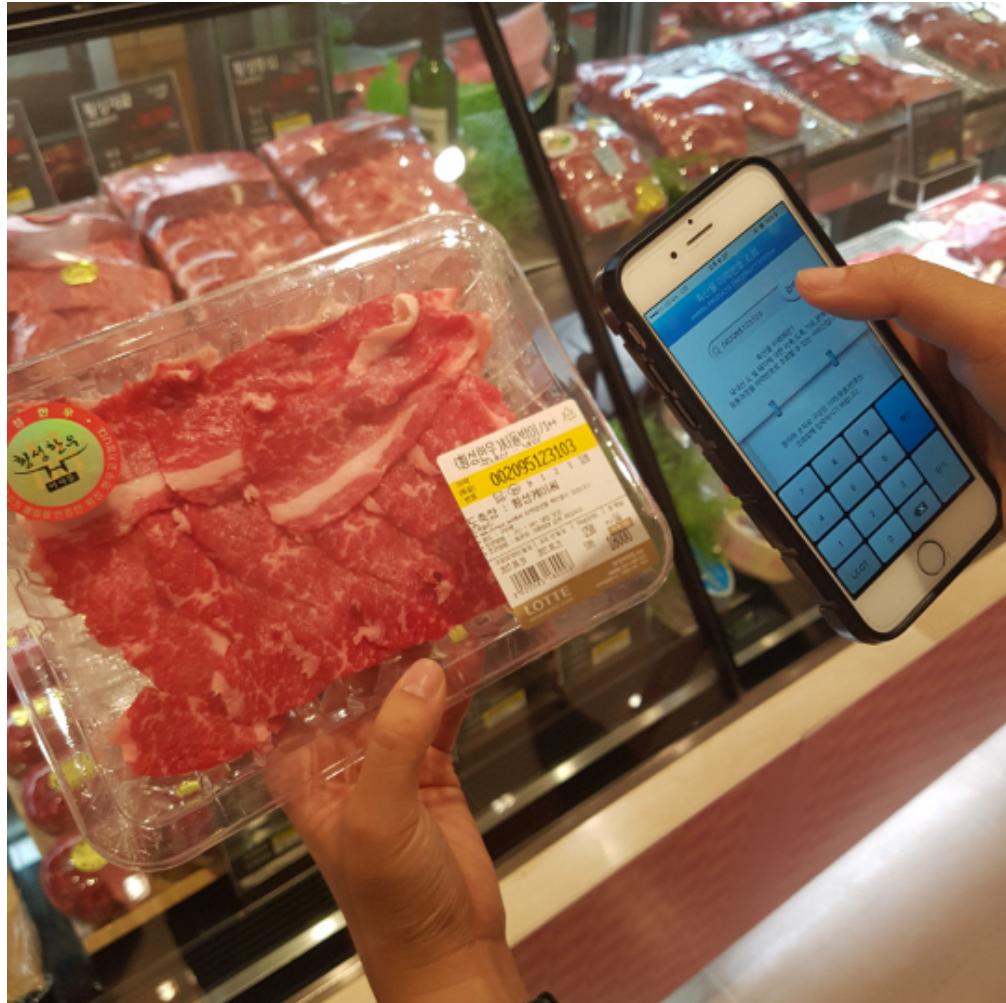


기상정보

기상청
기상 데이터
- 온도
- 습도
- THI
- 풍속



축산물 이력제



혈통정보 조회

HOME > 한우개량 > 개체정보조회

개체정보조회

개체식별번호	002095123103	검색	- 바코드로 조회시 12자리를 사용하세요. 예) 000123456789
			- 등록번호 예 : 221542733

기본정보	공란우여부
------	-------

개체 기본정보	확인서 인쇄	인쇄하기
---------	--------	------

등록번호	등록구분	생년월일	명호(이름)	계대	개체식별번호	성별	등록일
227508918	혈통	2014-11-23	강림14-04-5261	4	002095123103	수	2014-12-16

구분	정액번호 / 명호	생년월일	개체식별번호	등록구분	등록번호
↳ 조부	KPN486	1999-09-13		고등	221418921
부	KPN785	2006-09-20	000195289266	고등	223593901
↳ 조모	21624	1992-03-25	000120551074	혈통	221088540
↳ 외조부	KPN628	2003-03-03	000177170179	고등	221737804
모	강림10-03-0437	2010-12-23	002061553417	혈통	225168855
↳ 외조모	강림08-02-0411	2008-03-22	002009803839	고등	223817948

번식자 성명	도광*	번식자 주소	강원 횡성군 강림면***
소유자 성명	도광*	소유자 주소	강원 횡성군 강림면***

• 특징

모색	면선	미선	배선	뿔	기타
황	중	좌	중		

씨수소 능력조회



[종모우 상세내역 정보]

종모우명(KPN번호)	생년월일	냉도체중(kg)	배최장근 단면적(cm)	등지방두께(mm)	근내지방도(점)	장점	가격
KPN1080	2012-02-13	21.75	7.3	-1.56	1.08	등심형	10,000

[체형유전능력(EPD)]

체고	십자부고	체장	흉심	흉폭	고장	요각폭	gon폭	좌골폭	흉위
0.79	0.74	0.62	0.37	0.24	0.16	0.62	0.72	0.43	0.6

[번식자 정보]

번식자 성명	손종민	번식자 주소	충북 진천군 덕산면 화상리 537
--------	-----	--------	--------------------

[3대혈통정보]

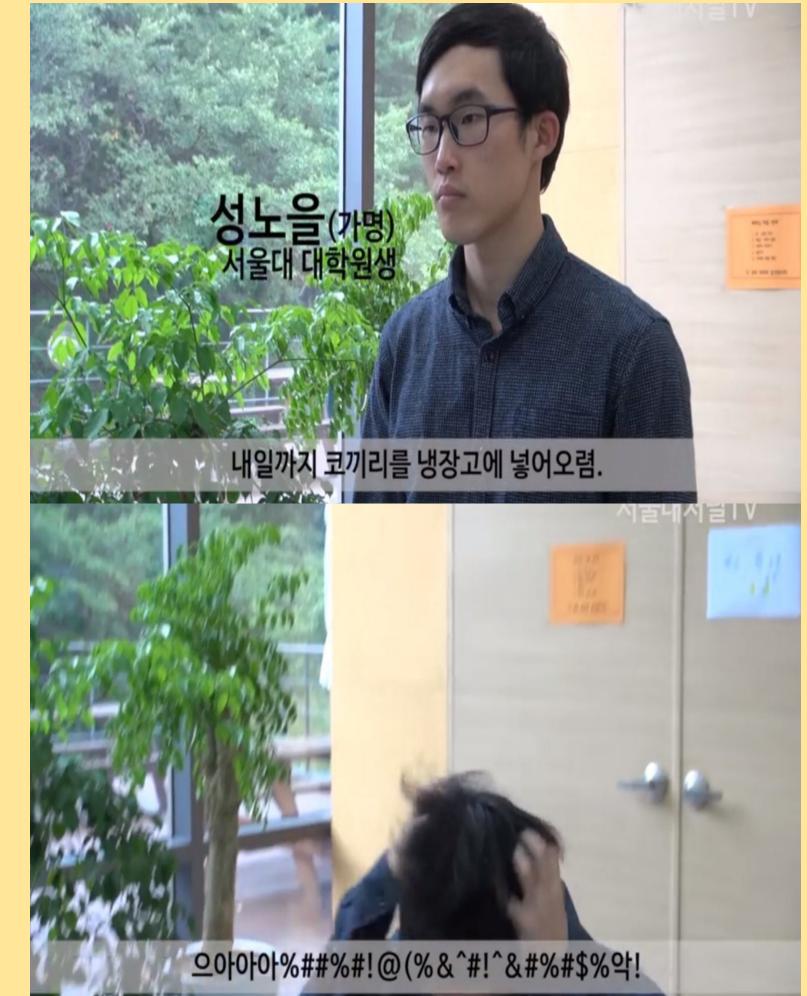
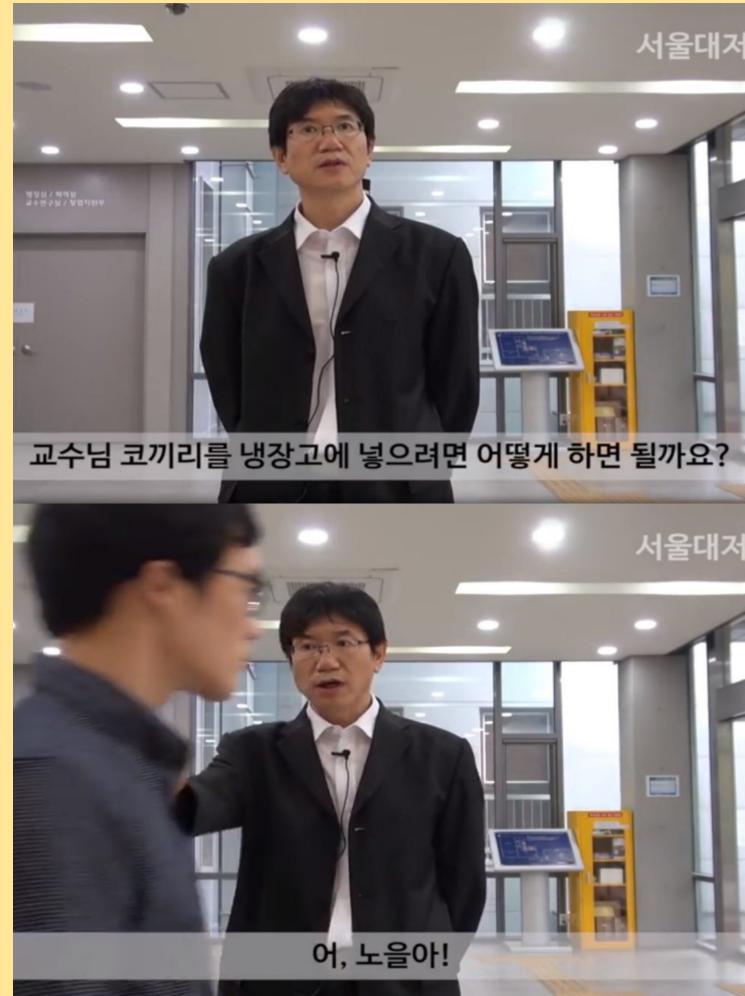
등록번호	부등록번호	모등록번호	조부등록번호	조모등록번호	외조부등록번호	외조모등록번호
226100274	223238314 (KPN768)	223627900 (진천07-03-0607)	221352161 (KPN387)	221861439 (홍천04-02-1054)	221467810 (KPN538)	223149692 (진천05-02-0399)

The background of the slide features a 3D perspective view of a landscape composed entirely of binary digits (0s and 1s). The terrain is rugged, with numerous peaks and valleys, creating a sense of depth and data volume.

넘치는 데이터를 어떻게 감당할 것인가?

넘치는 데이터를 어떻게 감당할 것인가?

1. 내가 분석한다
2. 대학원생에게 시킨다
3. 포닥에게 시킨다



포닥은 넘치는 데이터를 어떻게 감당할 것인가?

1. ~~내가 분석한다~~
2. ~~대학원생에게 시킨다~~
3. 기계에게 시킨다



기계학습(ML, Machine learning)

- 산업혁명 시대
 - = 물건 생산을 위해 육체노동을 기계로 자동화 하자
- 데이터혁명 시대
 - = 데이터 분석을 위해 정신노동을 기계로 자동화 하자

단순하고 반복적인 일들은 기계에게 맡기고

최대한 우리는

1. 어떻게 분석할 것인지 기획하고
2. 서비스의 질을 높일 것인지
3. 좀 더 본질적인 고민을 하자
4. 그리고 사랑하는 사람들과 더 많은 시간을 보내자 😊

ML이 잘 동작할 수 있는 조건

1. 알고리즘

= 기존의 학습 알고리즘 + Deep learning

2. 많은 데이터

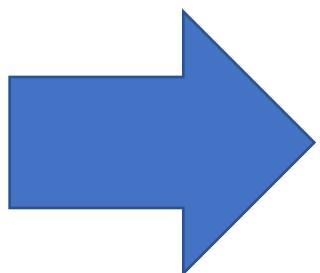
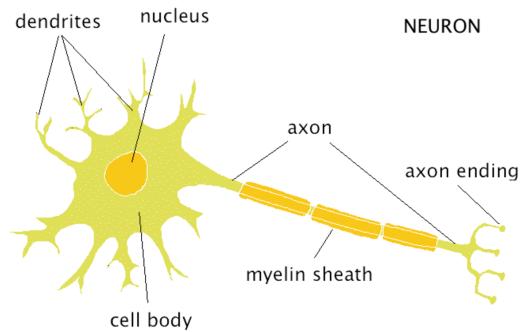
= IoT 발달로 인한 데이터의 축적

3. 컴퓨터 POWER

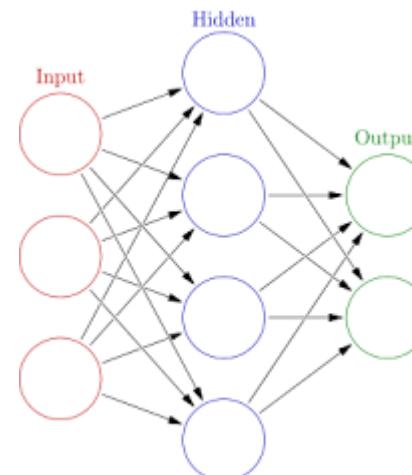
= 컴퓨터의 발달 + 클라우드 컴퓨터(AWS, Google Cloud Platform)

Deep learning

인간 두뇌의 뉴런 구조

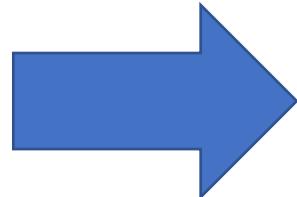


뉴럴 네트워크



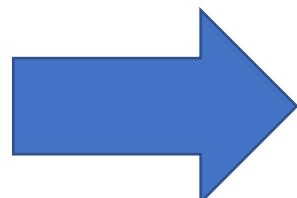
모든 것을 다 알고 시작할 수 없다

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$



=average(A1:A4)

$$SD = \sqrt{\frac{\sum (x - \bar{x})^2}{N - 1}}$$



=stdev(A1:A4)

모든 것을 다 알고 시작할 수 없다

Gradient descent algorithm

repeat until convergence {

$$\theta_0 := \theta_0 - \alpha \left[\frac{1}{m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)}) \right]$$
$$\theta_1 := \theta_1 - \alpha \left[\frac{1}{m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)}) \cdot x^{(i)} \right]$$

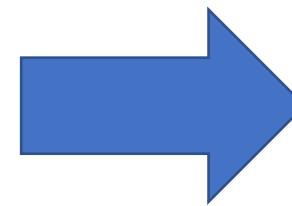
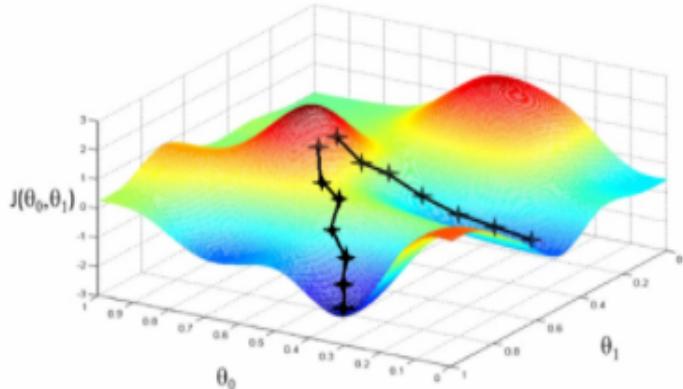
}

$$\frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

update
 θ_0 and θ_1
simultaneously

$$\frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

$$h_\theta(x) = \underline{\theta_0 + \theta_1 x}$$



`tf.train.GradientOptimizer()`

코드 한줄로 끝!

오픈소스 언어 

: 바퀴를 다시 발명하지 마라

오픈소스의 지향점



1. 소스코드 공개
2. 무료로 이용
3. 더 많은 사람들이 사용
4. 버그발견, 기능추가, 문서보강, 새로운 아이디어

GNU 선언문(1985; 1993 개정)



- <https://www.gnu.org/gnu/manifesto.ko.html>
- Once GNU is written, everyone will be able to obtain good system software free, just like air.
- This means much more than just saving everyone the price of a Unix license. **It means that much wasteful duplication of system programming effort will be avoided. This effort can go instead into advancing the state of the art.**
- 일단 GNU가 만들어지면, 모든 사람들은 훌륭한 시스템 소프트웨어를 공기처럼 무료로 얻을 수 있게 될 것이다.
- 이것은 모든 사람이 단지 유닉스 사용에 대한 라이선스 비용을 절약할 수 있다는 것보다 훨씬 더 많은 의미를 갖는다. 이것은 시스템 프로그래밍에 소모되는 불필요한 노력의 중복을 피할 수 있게 된다는 것을 의미하며, 절약된 노력은 기술 수준을 향상시키는데 사용될 수 있을 것이다.

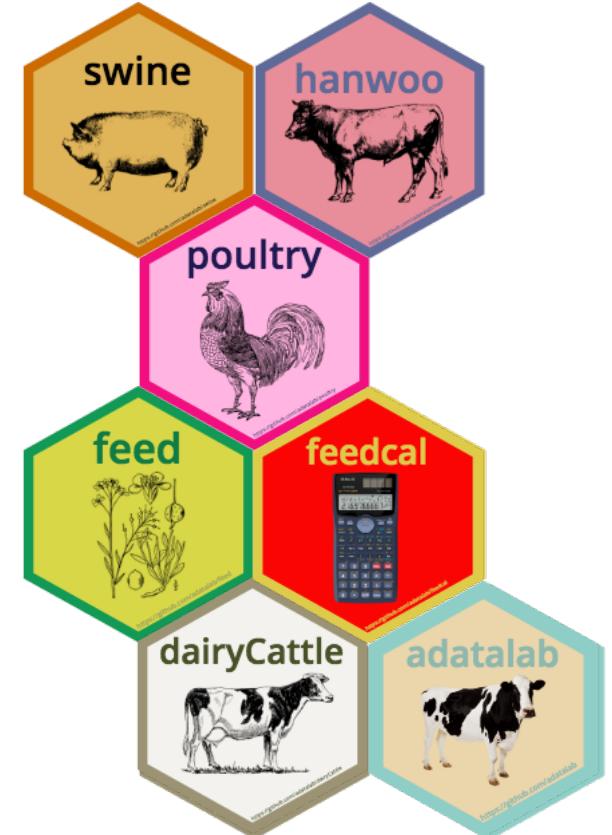
오픈소스 라이센스 R

- R 언어는 GPL (General Public License)를 채택
- 사용자들이 소프트웨어를 자유롭게 공유하고 수정할 수 있음



Animal science를 위한 R 패키지:

Project animal data lab. (adatalab)



animal data lab. (adatalab)



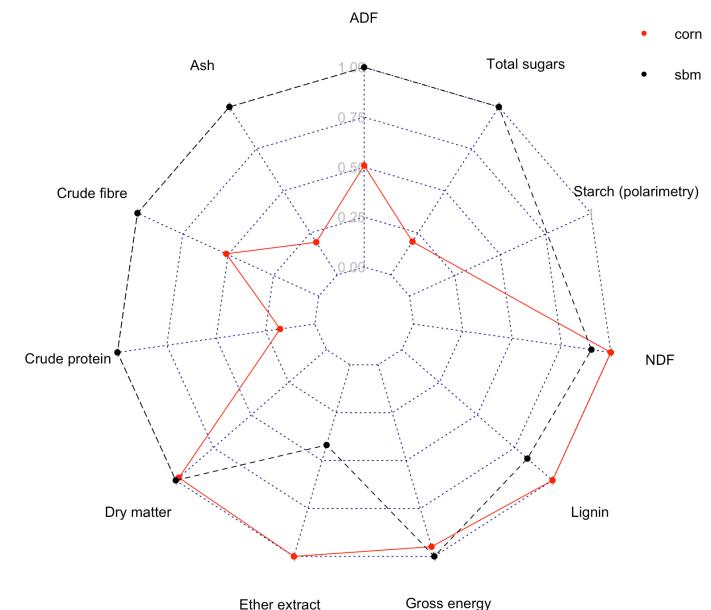
- <https://github.com/adatalab>
- 동물 영양 연구 분석용 R package를 제작

사료	요구량 설정/결과분석	사료배합
The logo for the "feed" package is a yellow hexagon with a green border. Inside, there is a detailed line drawing of a flowering plant, likely a legume, which is a common source of protein in animal feeds. The word "feed" is printed in a bold, sans-serif font at the bottom left of the hexagon.	A cluster of six hexagonal icons arranged in two rows. The top row contains three icons: "swine" (yellow background, pig silhouette), "hanwoo" (pink background, cow silhouette), and "dairyCattle" (tan background, cow silhouette). The bottom row contains three more icons: "dairyCattle" (white background, cow silhouette), "poultry" (pink background, rooster silhouette), and another "dairyCattle" (white background, cow silhouette).	The logo for the "feedcal" package is a red hexagon with a gold border. Inside, there is a black scientific or programmable calculator. The word "feedcal" is printed in a bold, sans-serif font at the top left of the hexagon.

Package: feed



- <https://github.com/adatalab/feed>
- Imports: rvest, ggplot2, dplyr
- `feed_list()`
- `feed_search()`
- `feed_info()`
- `feed_compare()`



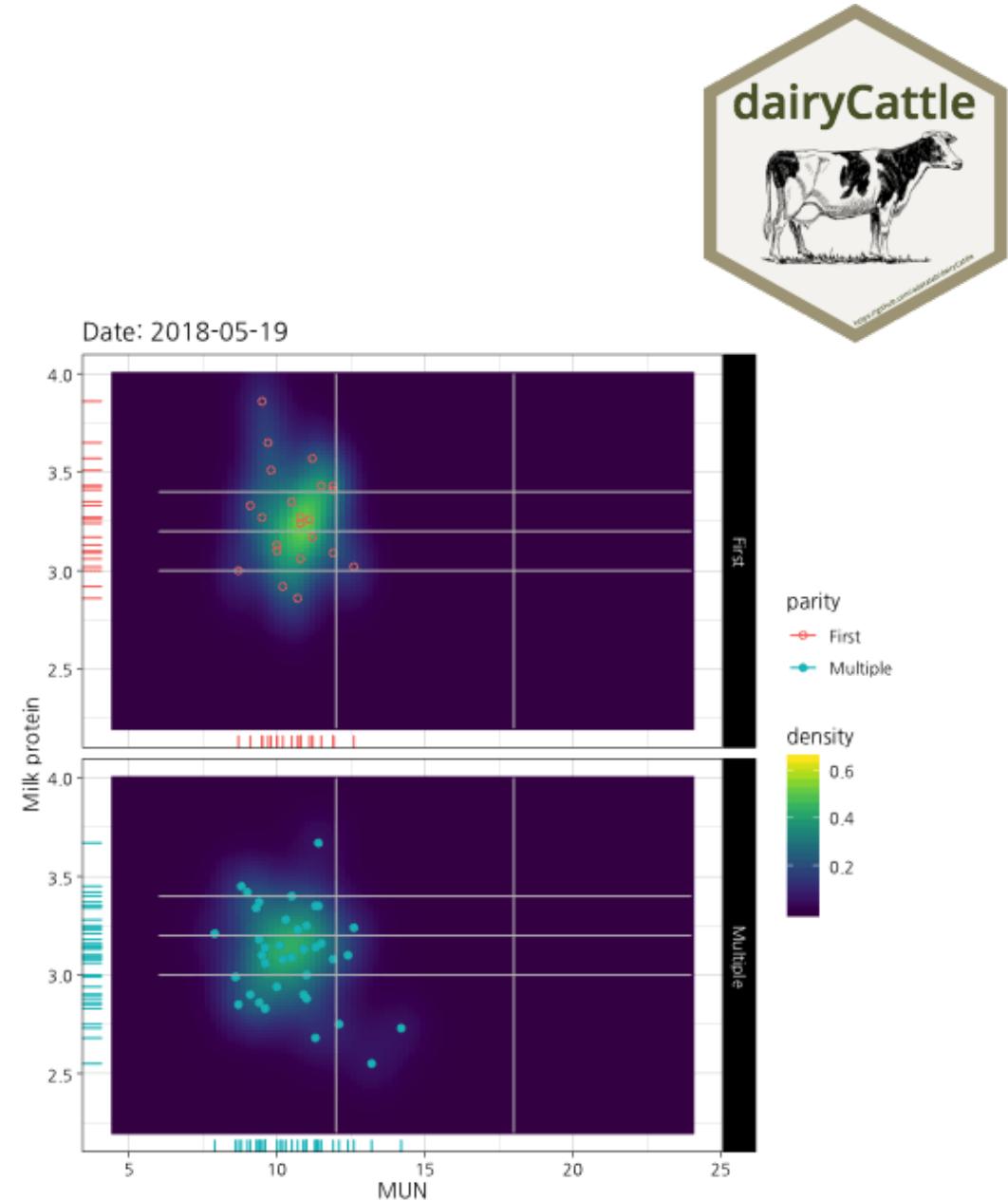
Package: hanwoo



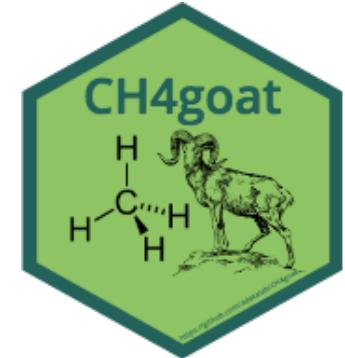
- <https://github.com/adatalab/hanwoo>
- Imports: XML, dplyr
- hanwoo_info()
- hanwoo_bull()
- Functions for import the nutrient requirements of Hanwoo.

Package: dairyCattle

- <https://github.com/adatalab/dairyCattle>
- Imports: `ggplot2`, `dplyr`, `janitor`, `timevis`
- `mun_mp()`
- `parturition()`
<https://youngjunna.github.io/adatalab/parturition>
- Functions for import the nutrient requirements of dairy cattle.

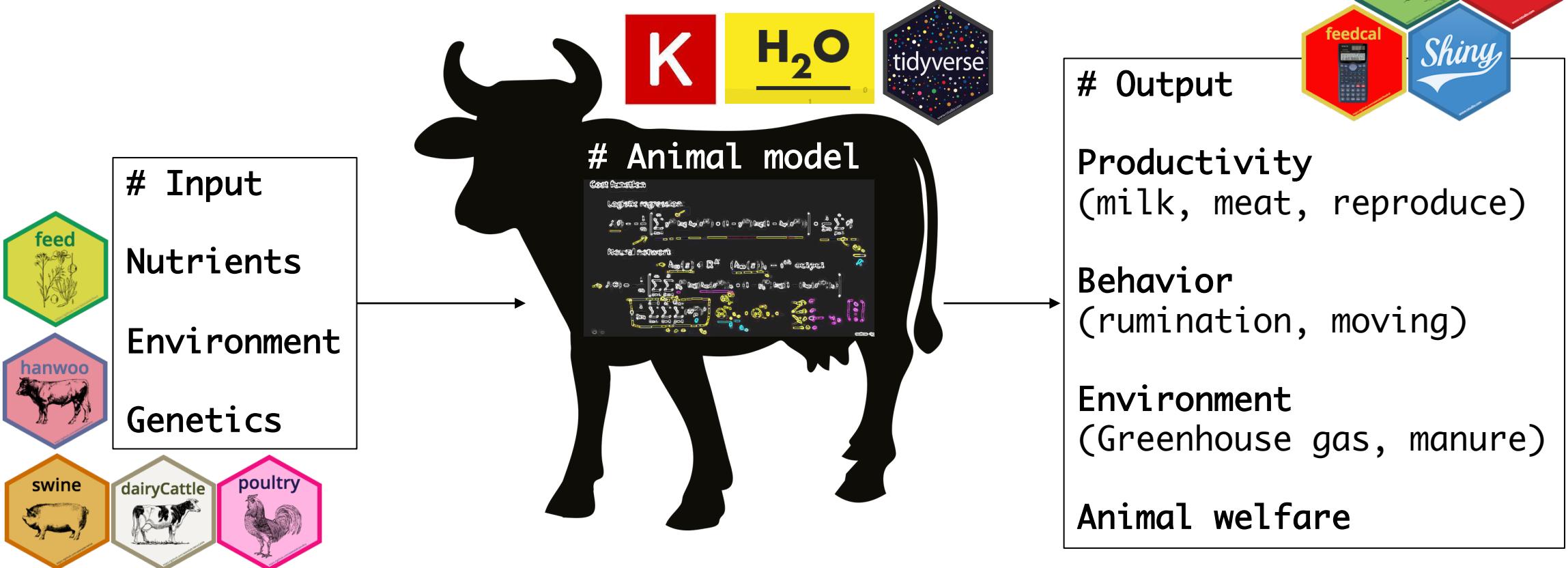


Package: CH4goat



- <https://github.com/adatalab/CH4goat>
- Imports: neuralnet
- The package contains two black-box model derived from the *in vivo* enteric methane dataset using respiration-metabolism chamber system.
- `neuralnet::compute(model1, your_data)`

Workflow for animal modeling





R을 이용한 재현가능한 연구:

R markdown; shiny

재현 가능한 연구(Reproducible Research)

- 논문 또는 연구 문서 내에 실제 연구에 사용한 소스 코드와 데이터를 모두 한꺼번에 배포
- 누구나 쉽게 다시 재현이 가능할 수 있게 배포자가 배포물을 체계적으로 관리하고 배포하는 것

Markdown

마크다운

위키백과, 우리 모두의 백과사전.

마크다운(markdown)은 일반 텍스트 문서의 양식을 편집하는 문법이다^[1]. README 파일이나 온라인 문서, 혹은 일반 텍스트 편집기로 문서 양식을 편집할 때 쓰인다. 마크다운을 이용해 작성된 문서는 쉽게 HTML 등 다른 문서형태로 변환이 가능하다.

목차 [숨기기]

- 1 역사
- 2 문법
- 3 같이 보기
- 4 각주
- 5 외부 링크



역사 [편집]

존 그루버는 2004년에 문법 면에서 에런 스워츠와 중대한 협업을 통해 마크다운 언어를 만들었으며^{[2][3]}, 사람들이 읽기 쉽고 쓰기 쉬운 플레인 텍스트 포맷을 사용하여 쓸 수 있으면서 구조적으로 유효한 XHTML(또는 HTML)로 선택적 변환이 가능하게 하는 것이 목표이다.^[4]

R markdown

- <https://rmarkdown.rstudio.com/>
- <https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf>
- 기본적인 문법은 마크다운과 동일
- 문서 내에 R 분석 결과를 쉽게 삽입할 수 있음

```
```{r}  
plot(iris$Sepal.Length, iris$Sepal.Width)
```
```

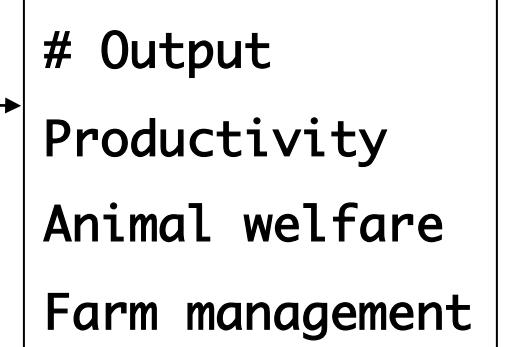
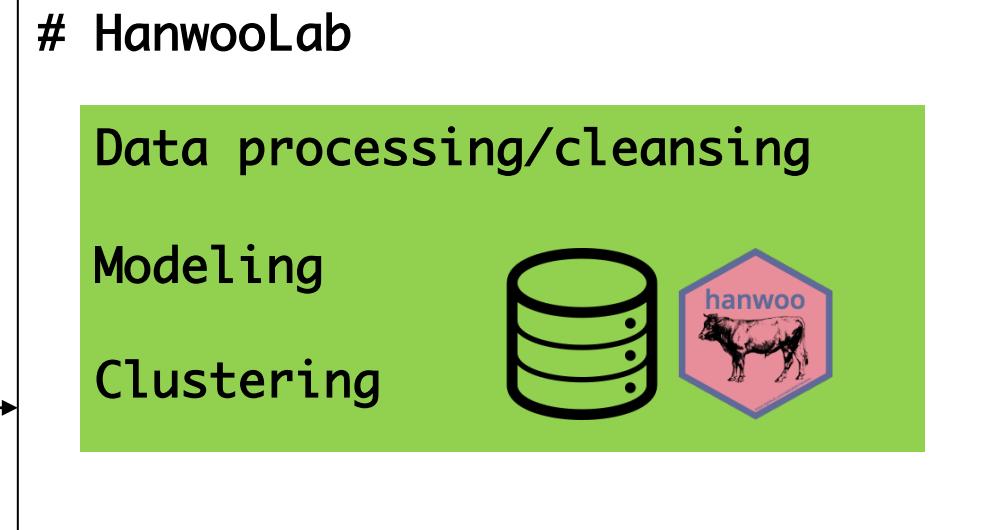
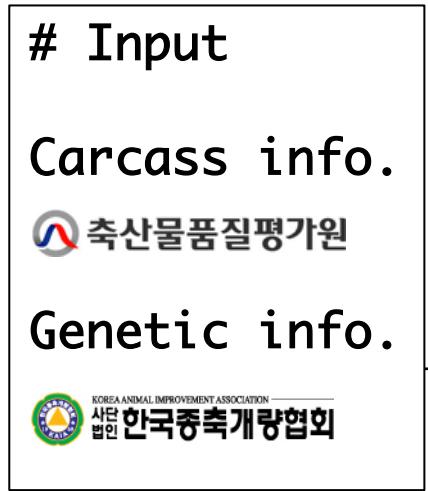
Shiny

- <https://shiny.rstudio.com/>
- <https://shiny.rstudio.com/gallery/>
- <https://youngjunna.github.io/adatalab/kpn-2017-sample.html>
- HTML, CSS, 자바스크립트를 사용하지 않고 순수한 R만으로 Web application을 만들 수 있음
- 반응성(reactivity) 모드를 디폴트로 사용하기 때문에 사용자가 입력한 값에 즉각적으로 반응

HanwooLab

<https://adatalab.github.io/hanwoolab-example/>

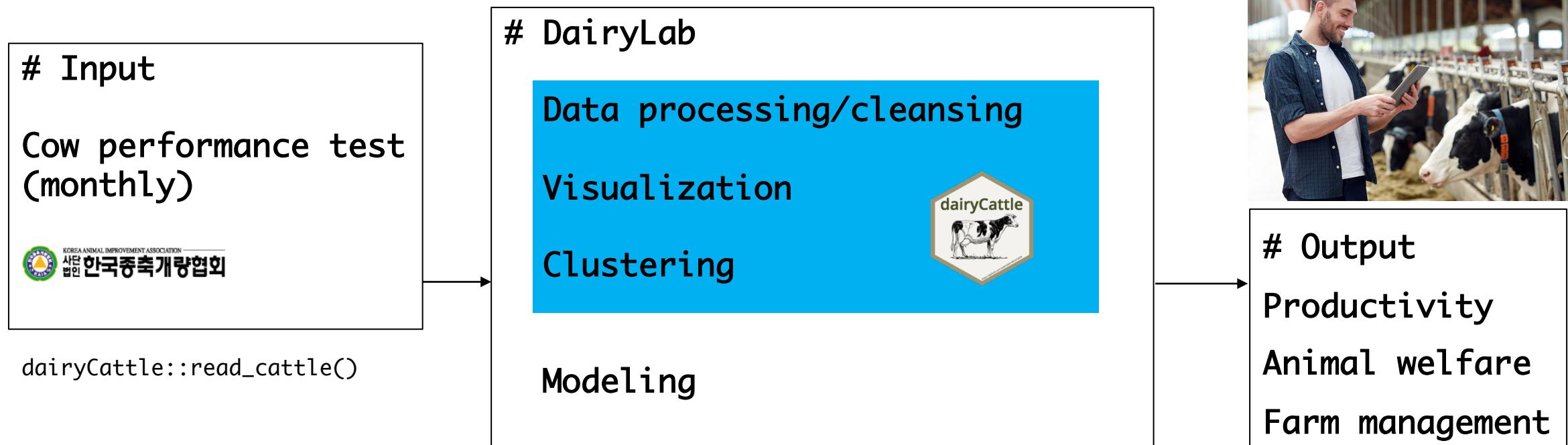
id: user pw: pass



```
hanwoo::hanwoo_info("002123456789")
+
hanwoo::hanwoo_bull("1080")
+
Rselenium + rvest
```

DairyLab

<https://adatalab.github.io/dairylab-pro/>



매번 똑같은 보고서 쓰는게 귀찮다면..



Github을 이용한 협업

- <https://www.youtube.com/watch?v=w3jLJU7DT5E>
- <https://github.com/YoungjunNa>
- <https://github.com/adatalab>

Thank you

Q&A