

데이터 시각화 가이드

제발 이렇게 그리지 마세요

2021.05.24.

한국에너지기술연구원 플랫폼연구실

이제현

지난 발표의 일부 + α

Pega Devlog a space for development <https://jehyunlee.github.io>

Home GIS ImageJ Python About Search

FOLLOW: 2021

RECENTS

- PYTHON > DATA SCIENCE
MATPLOTLIB VIS. FUNCTION
2021-05-18
- PYTHON > DATA SCIENCE
PYTHON을 이용한 연구데이터 시각화 PART 3
2021-06-04
- PYTHON > DATA SCIENCE
PYTHON을 이용한 연구데이터 시각화 PART 2
2021-06-03
- PYTHON > DATA SCIENCE
PYTHON을 이용한 연구데이터 시각화 PART 1
2021-04-30
- PYTHON > DATA SCIENCE
GIBBS SAMPLING IN N-DIMENSION
2021-04-20

CATEGORIES

- GIS (7)
 - Python (5)
 - QGIS (2)
- ImageJ (12)
 - Cookbook (4)
 - Tutorial (8)
- Python (84)
 - Data Science (71)

데이터 시각화 가이드

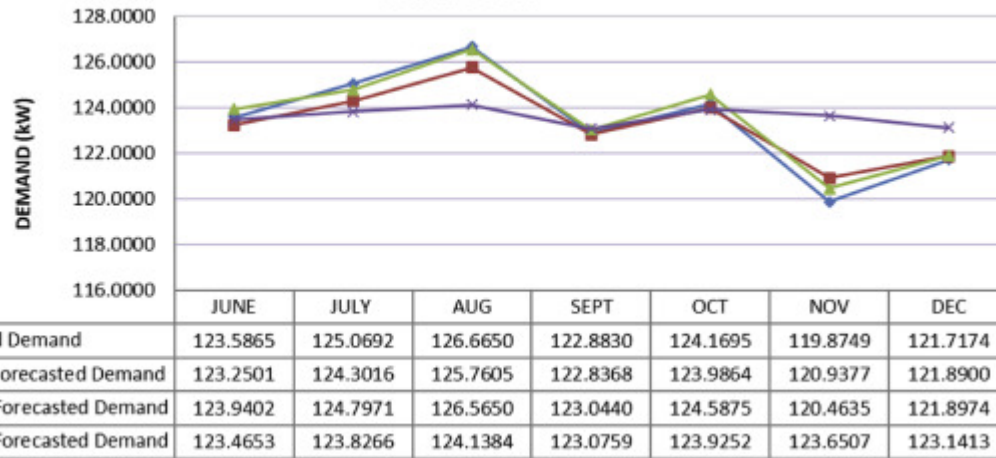
Part 1

Part 2

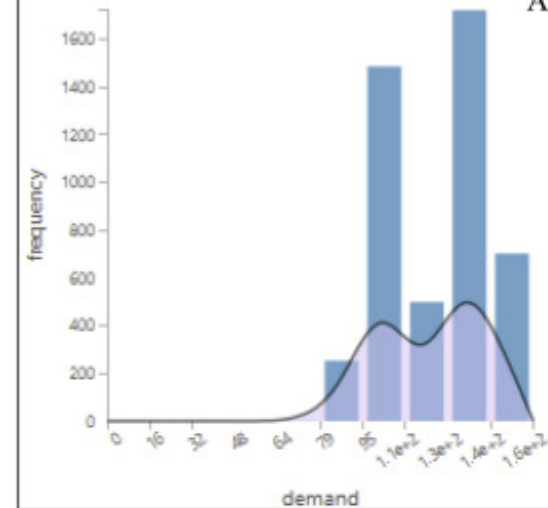
Part 3

연구데이터 시각화

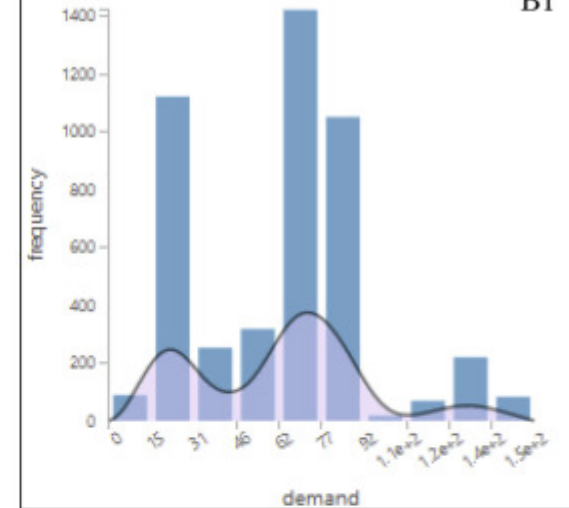
TENANT A1



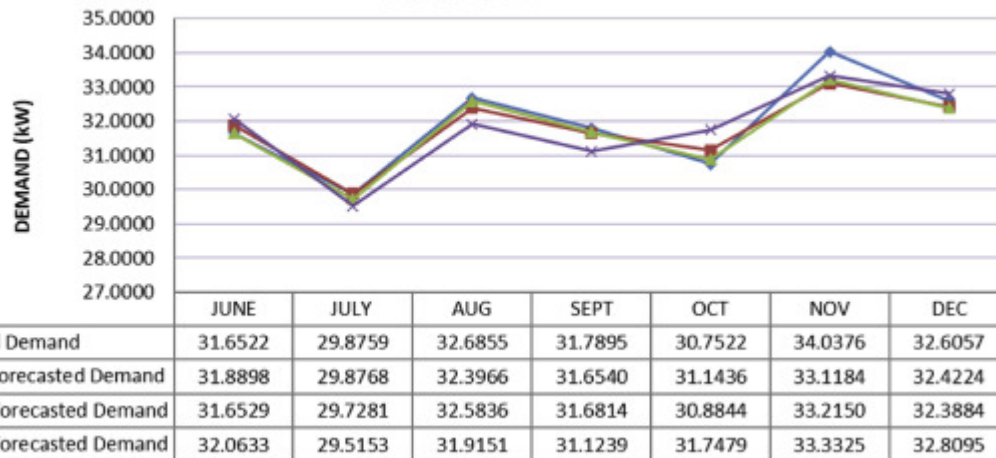
A1



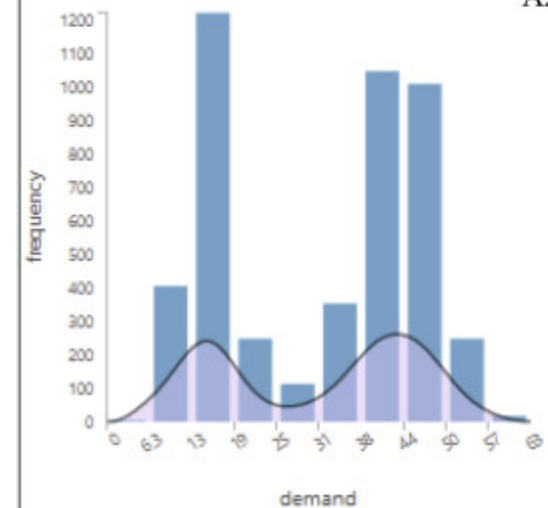
B1



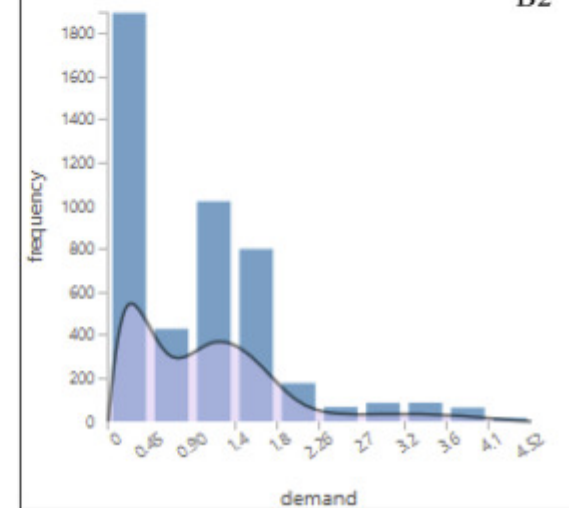
TENANT A2



A2



B2



데이터 시각화 Data Visualization

- 숫자를 그림으로 바꾸는 작업

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1				
1	4.9				
2	4.7				setosa
3	4.6				
4	5.0	3.6	1.4	0.2	setosa
...	...				
145	6.7	3.0	5.2	2.3	virginica
146	6.3				
147	6.5	5.0	5.2	2.0	virginica
148	6.2				
149	5.9				

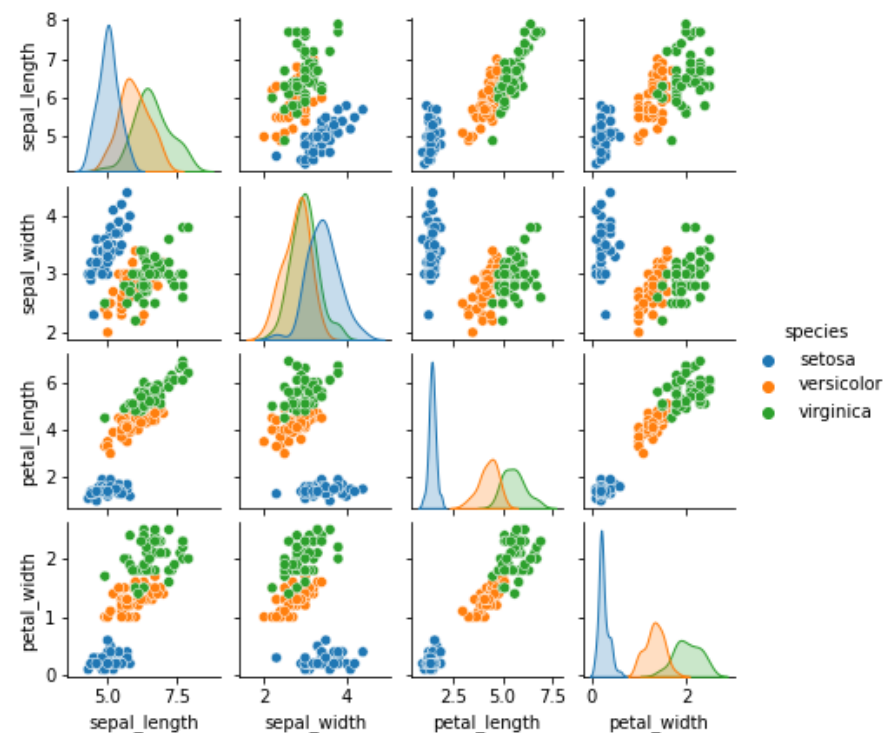
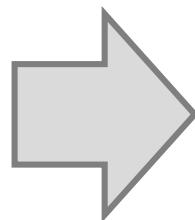
- 데이터 이해도 증진

- 새로운 특징 탐색

- 데이터 문제점 도출

- 전체와 부분 동시 이해

- 가설 수립 & 검증

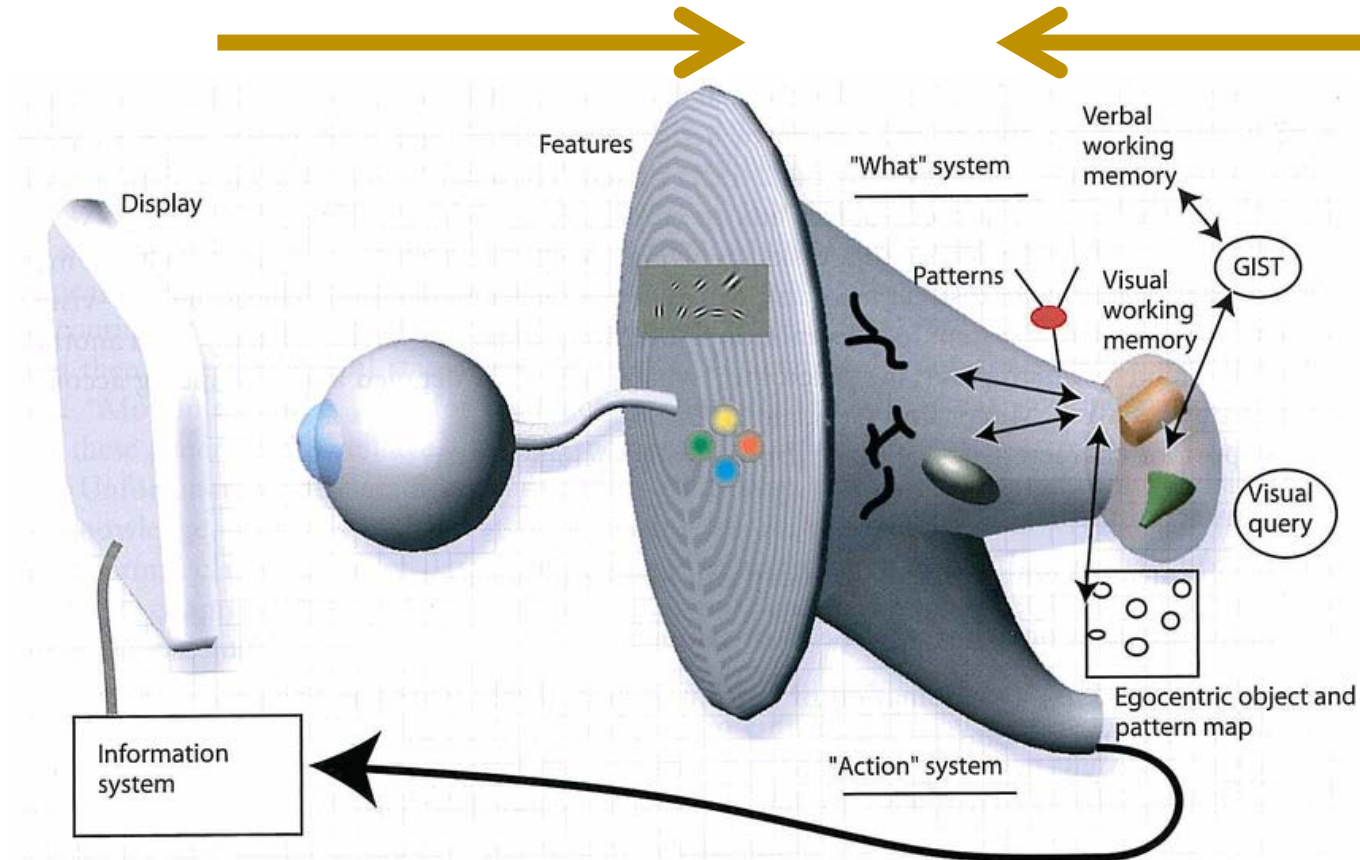


인간 시각 인지 메커니즘

- 상향식 & 하향식

상향식 : 무의식적 정보 수집 (뇌 부담 적음)

하향식 : 시각적 질의에 대한 답 찾기 (뇌 부담 큼)



인간 시각 인지 메커니즘

N

지도 홈

길찾기

버스

지하철

기차

즐거찾기

더보기

대중교통 자동차 도보 자전거

한국에너지기술연구원 대전역 경부선(고속철도)

다시입력

길찾기 >

전체 10 버스 2 지하철 0 버스+지하철 8

오늘 오후 10:58 출발 ON ? 추천순

51분 최적
오후 11:04-오후 11:55 | 13분 | 1,350원

704 한국화학연구원 승차
8분 9정류장

유성온천역 환승 → 대전역 하차

상세보기 >

나중에 출발

45분 ① 운행시간 주의
다음날 오전 5:48-오후 6:33 | 10분 | 1,250원

한국에너지기술연구원 → 대전역 경부선(고속...

전체경로 보기

일반지도 위성지도 지형지도

거리뷰

지적량정보

테마

반경

면적

거리

저장

인쇄

공유

NAVER 1km

인간 시각 인지 메커니즘

Q: 트리플시티 9단지 아파트는 어디?

N

지도 홈
길찾기
버스
지하철
기차
즐거찾기
더보기

대중교통 자동차 도보 자전거

○ 한국에너지기술연구원
○ 대전역 경부선(고속철도)

다시입력 **길찾기 >**

전체 10 버스 2 지하철 0 버스+지하철 8

오늘 오후 10:58 출발 **ON** ? 추천순 ▾

51분 최적
오후 11:04-오후 11:55 | 13분 | 1,350원

704 한국화학연구원 승차
8분 9정류장

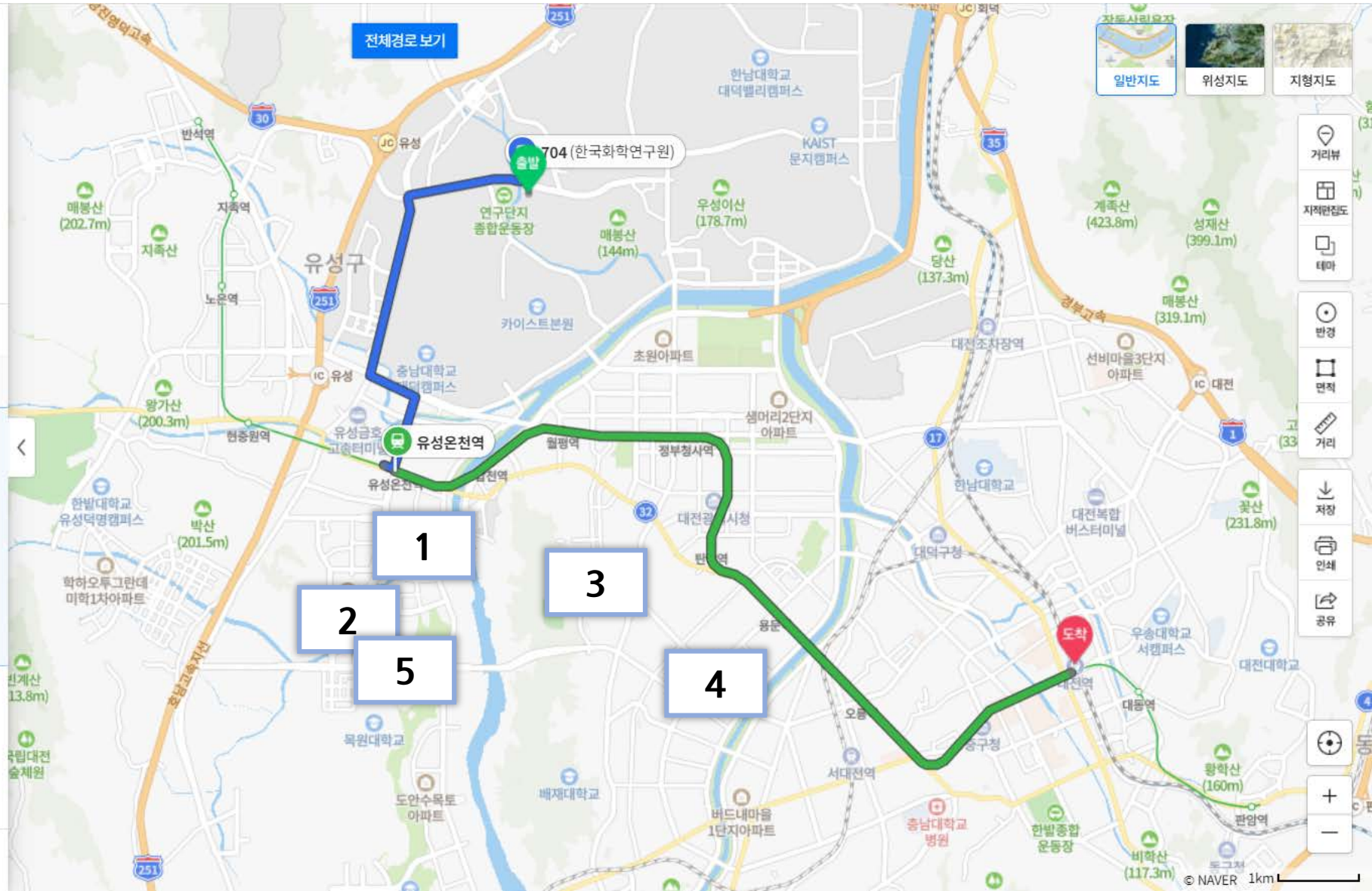
1 유성온천역 환승 → 대전역 하차

상세보기 >

나중에 출발

45분 ① 운행시간 주의
다음날 오전 5:48-오전 6:33 | 10분 | 1,250원

한국에너지기술연구원 → 대전역 경부선(고속...



인간 시각에 대한 연구 : 전주의 처리

- 전주의 처리(1988) - 주의를 기울이지 않아도 처리되는 정보가 있다.
= 뇌에 부담을 더 줄 수도, 덜 줄 수도 있다.

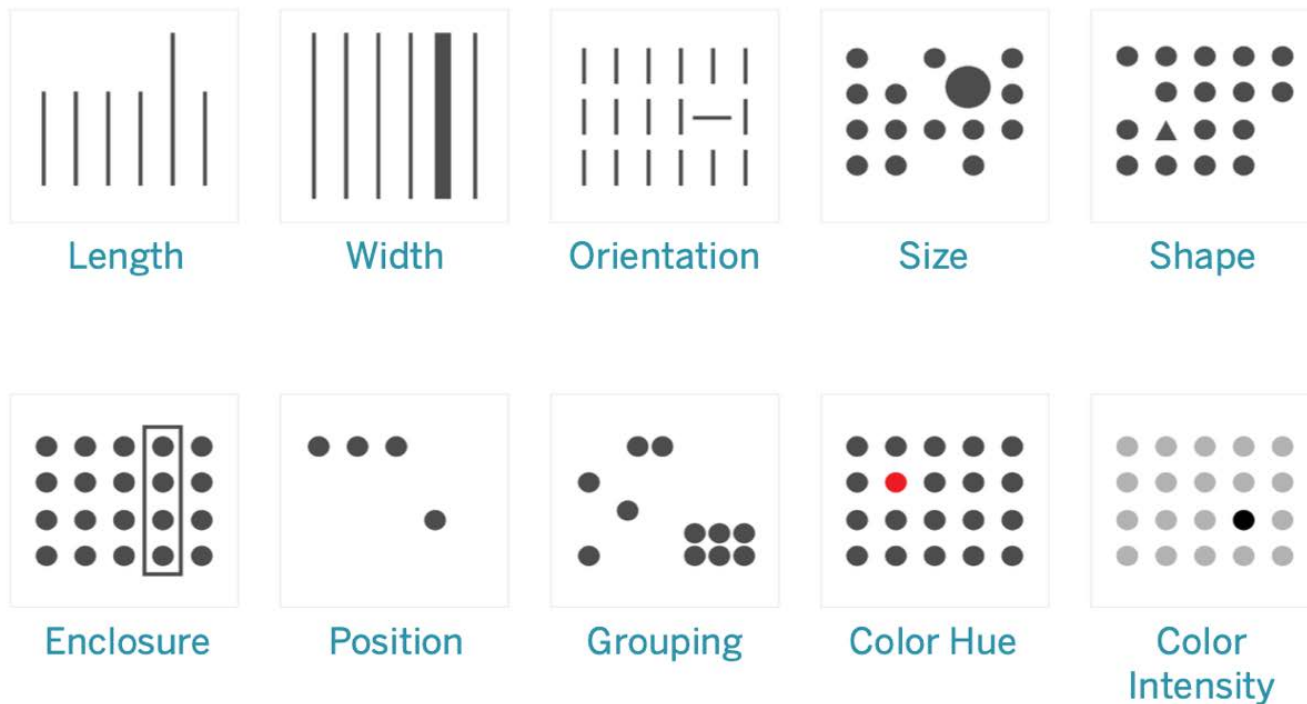
숫자 속 9 개수 세기

6	6	4	5	1	5	0	3	0	5	4	2	6	3	7	5	0	3
1	0	4	3	2	8	9	4	8	8	4	3	0	3	6	6	1	8
8	1	1	8	7	1	3	0	4	9	3	2	6	5	5	7	6	1
9	6	2	3	9	0	9	0	8	6	3	6	1	5	9	6	8	7
3	9	0	3	0	2	2	4	2	2	0	0	9	8	2	1	9	4
0	7	9	1	8	7	3	3	1	5	5	7	1	4	9	7	8	1
7	1	0	6	7	6	7	6	6	8	4	0	0	2	5	6	0	4
3	1	8	6	3	5	9	1	4	7	1	8	1	5	9	0	7	7
1	0	7	3	1	4	8	6	9	4	0	1	8	6	6	1	1	6

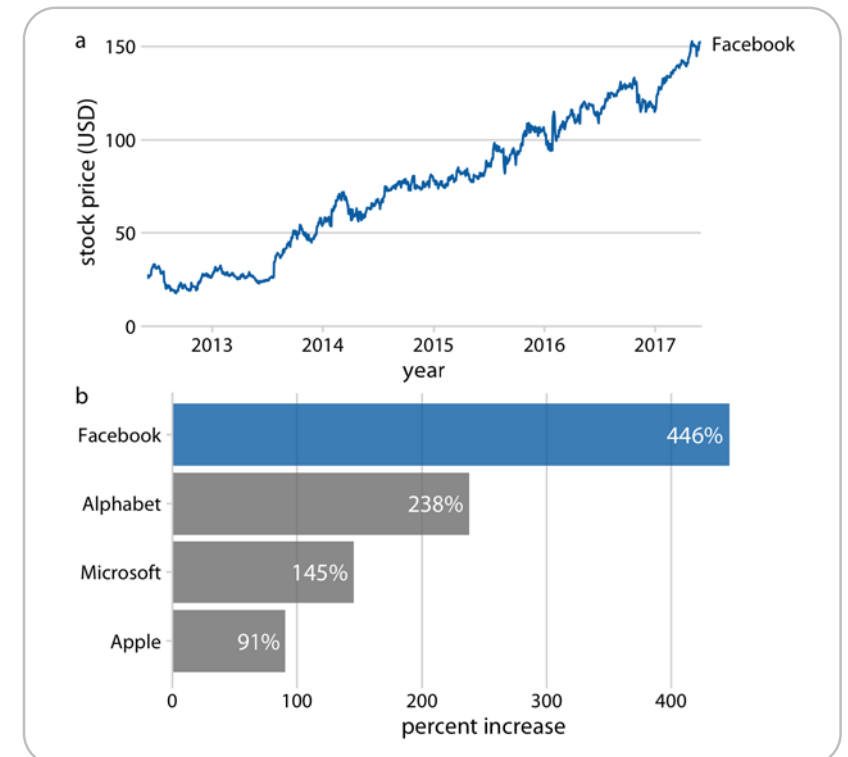
인간 시각에 대한 연구 : 전주의 처리

- 전주의 처리(1988) - 주의를 기울이기 전에 다른 점을 찾는다.
= 뇌에 부담을 더 줄 수도, 덜 줄 수도 있다.

전주의 처리 사례



페이스북 주가 변동 분석

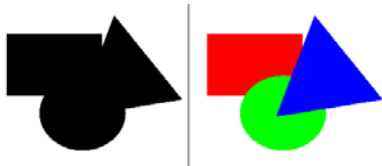


인간 시각에 대한 연구

- 게슈탈트 법칙(1912) - 패턴으로 인식되는 그림이 있다
= 설명 없이 패턴으로 인식시킬 수 있다. 엉뚱한 패턴이 생길 수 있다.

게슈탈트 이론

단순함 : 단순한 형태의 합으로



폐쇄 : 하나로 인식되는 패턴으로



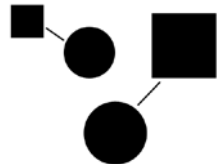
대칭 : 대칭을 기준으로 인식



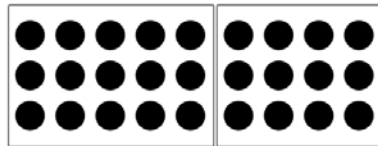
형태와 배경 : 둘 중 하나로 인식



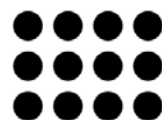
연결 : 연결된 요소간 연관성



공동 영역 : 닫힌 공간의 요소들



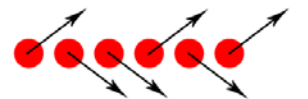
근접 : 가까이 있는 요소들



연속 : 가까이 있는 요소들



공동 운명 : 같이 움직이는 요소



평행 : 평행 요소끼리 연관됨



유사 : 비슷한 요소끼리 연관됨



초점 : 다른 지점은 집중됨



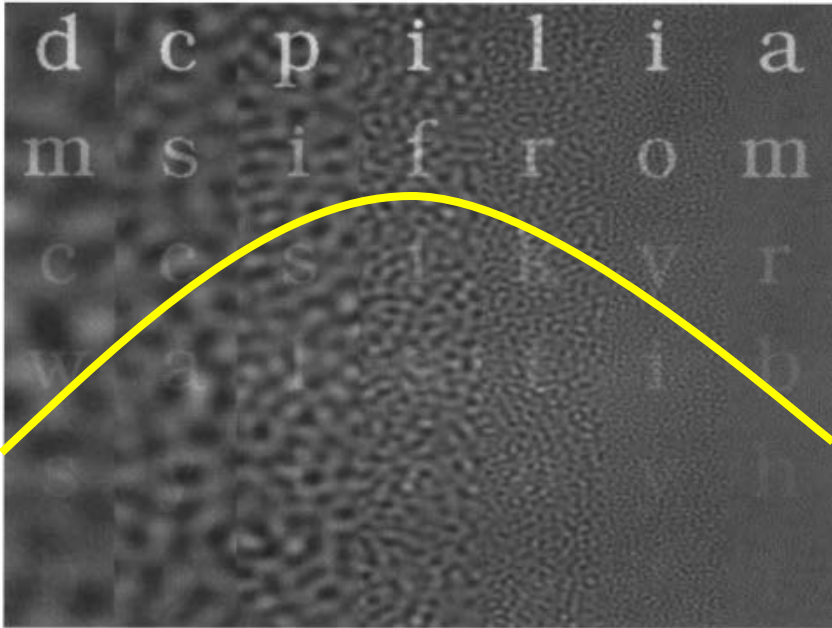
과거의 경험 : 경험에 따른 인식



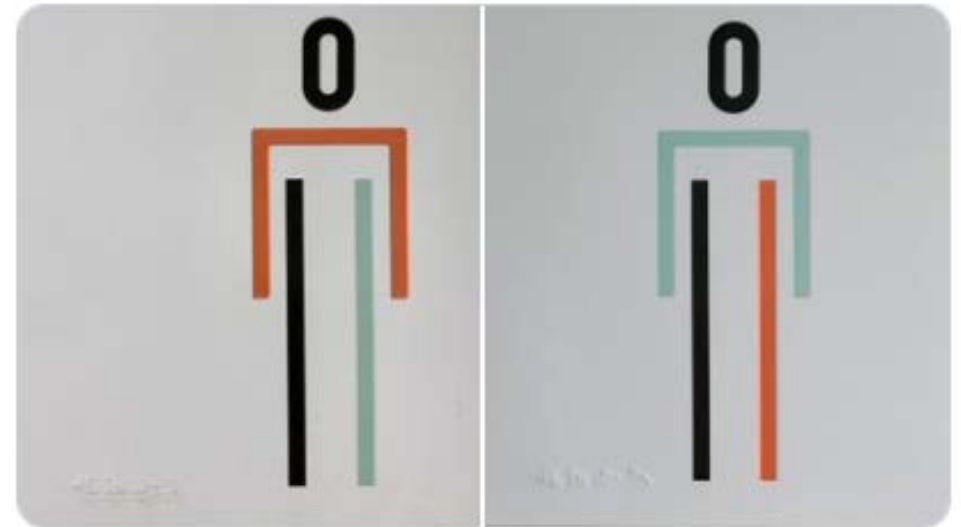
거꾸로, 잘 안보이려고 작성하는 것도 가능

- 전주의처리를 **반대로** 적용한다거나 : 주변이랑 비슷하게 배치
- 게슈탈트 원리를 **거꾸로** 적용한다거나 : 익숙하지 않은 기호 사용

FIG. 1 Letters in noise, demonstrating that, despite the broad $1/f^2$ spectra of letters, noise at 3 cycles per letter (middle column) masks letter identification more effectively than noise at lower or higher frequencies. Read down each column as far as you can; the positions of the faintest identifiable letters trace out the sensitivity of your eye's letter-identification channel as a function of spatial frequency. Note that changes in viewing distance, from 3 to 60 cm, hardly affect the visibility of any given letter, indicating that the channel scales with letter size. (Identification of letters in noise is independent of size for letters subtending 0.5 to 15.8 degrees of visual angle²².) Letter contrast (luminance increment divided by background luminance) decreases by factors of $\times 0.588$ from 0.89 in the top row to 0.063 in the bottom row. The noise contrast c_{rms} is 0.22. Centre frequency (cycles per letter) of the noise's octave-wide pass band is: 0.8, d, m, c, w, s, h; 1.3, c, s, c, a, y, o; 2.0, p, i, s, l, f, t; 3.2, l, f, l, s, r, w; 5.1, l, r, k, t, o, s; 8.1, i, o, y, i, y, f; 13, a, m, r, b, h, t.



이것만 보고 구별 가능한지 여쭙고 싶네요



2

149

145



기술적 한계로 인한 시각 표현 제약

- 스타워즈 - 자바 더 헛 Jabba the Hutt

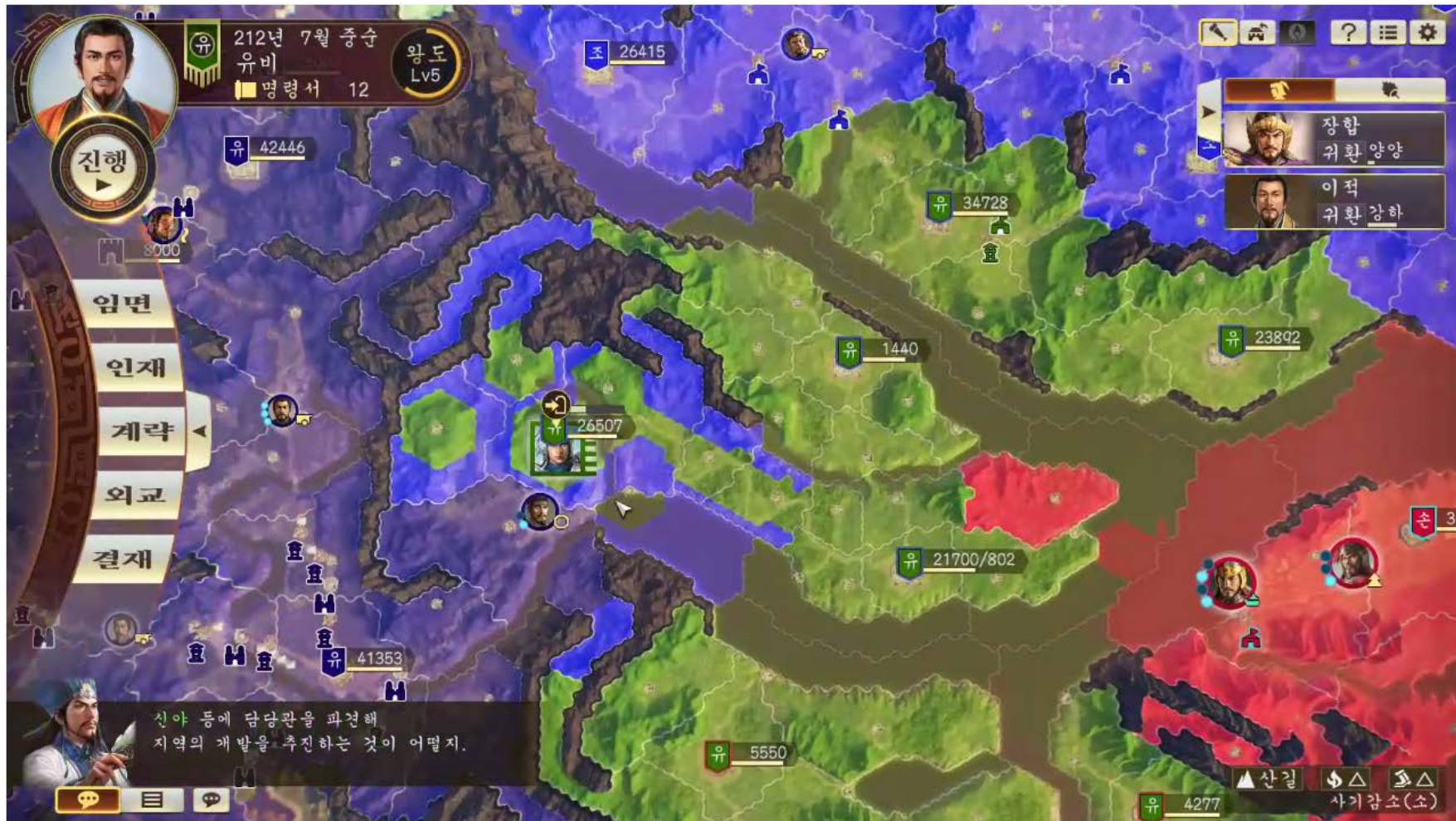
- “fat, slug-like creature with eyes on extended feelers and a huge ugly mouth”
<https://web.archive.org/web/20080501111000/http://www.starwars.com/databank/character/jabbathehutt/?id=bts>



기술적 한계로 인한 시각 표현 제약

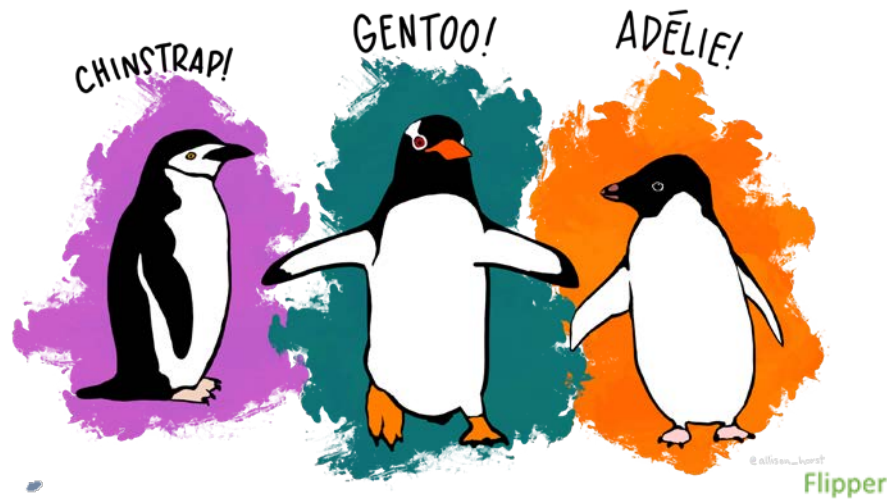
- 삼국지 시리즈 (KOEI)

삼국지 14 (2020)



데이터 분포 시각화

- 펭귄 데이터셋

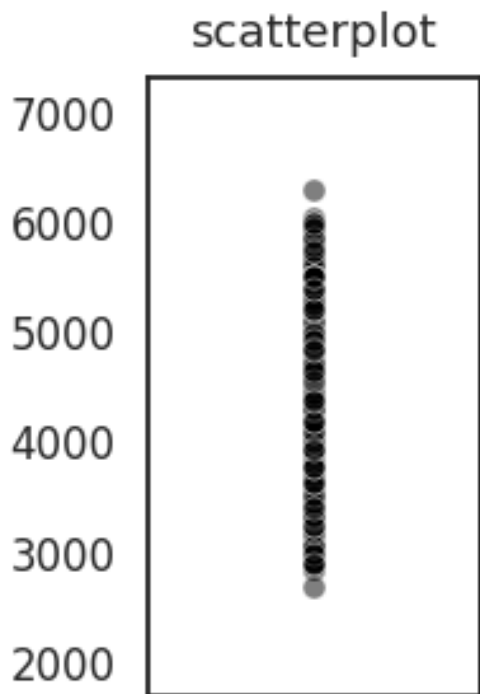


데이터 분포 시각화

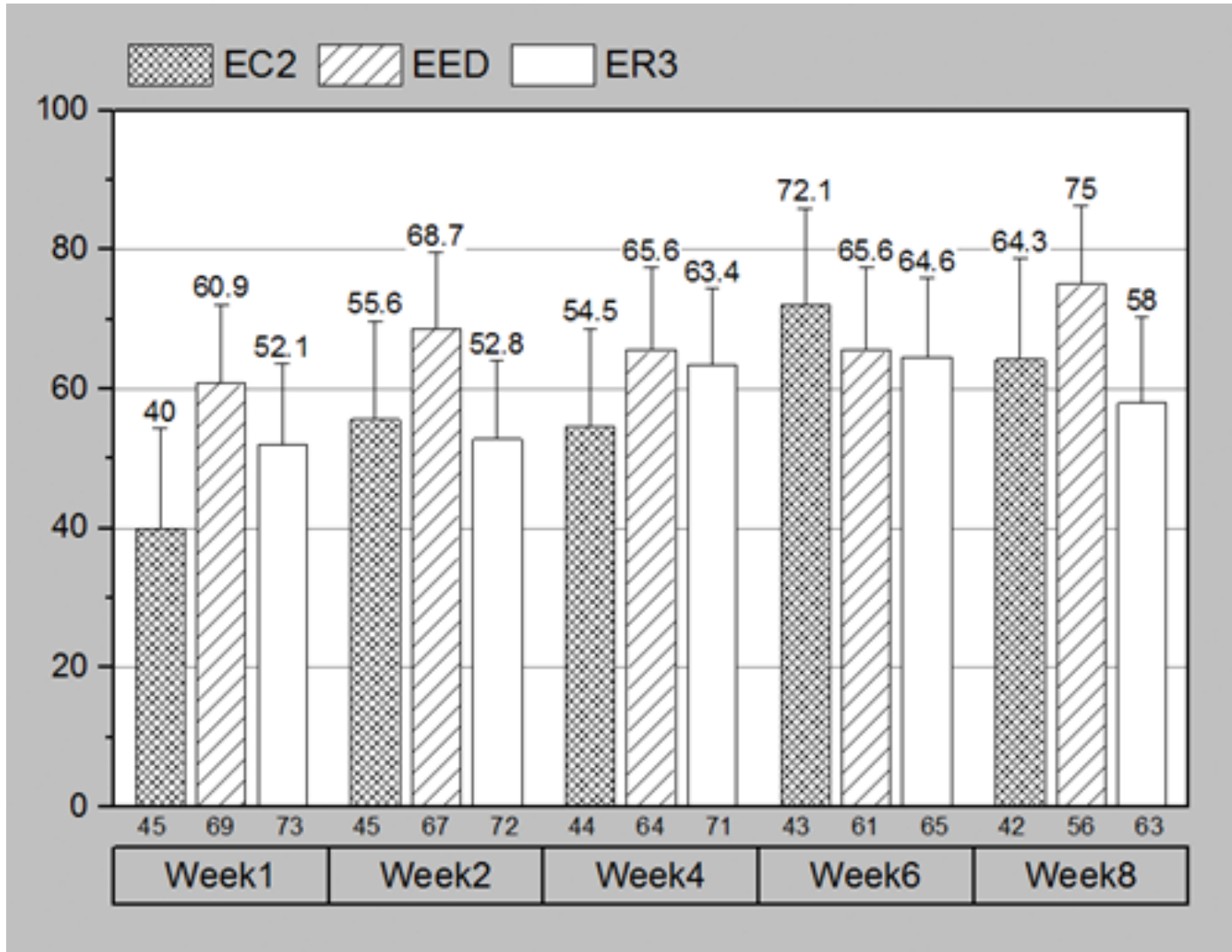
- 펭귄 데이터셋

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	Male
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	Female
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	Female
3	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	Female
5	Adelie	Torgersen	39.3	20.6	190.0	3650.0	Male
6	Adelie	Torgersen	38.9	17.8	181.0	3625.0	Female
7	Adelie	Torgersen	39.2	19.6	195.0	4675.0	Male
8	Adelie	Torgersen	34.1	18.1	193.0	3475.0	NaN

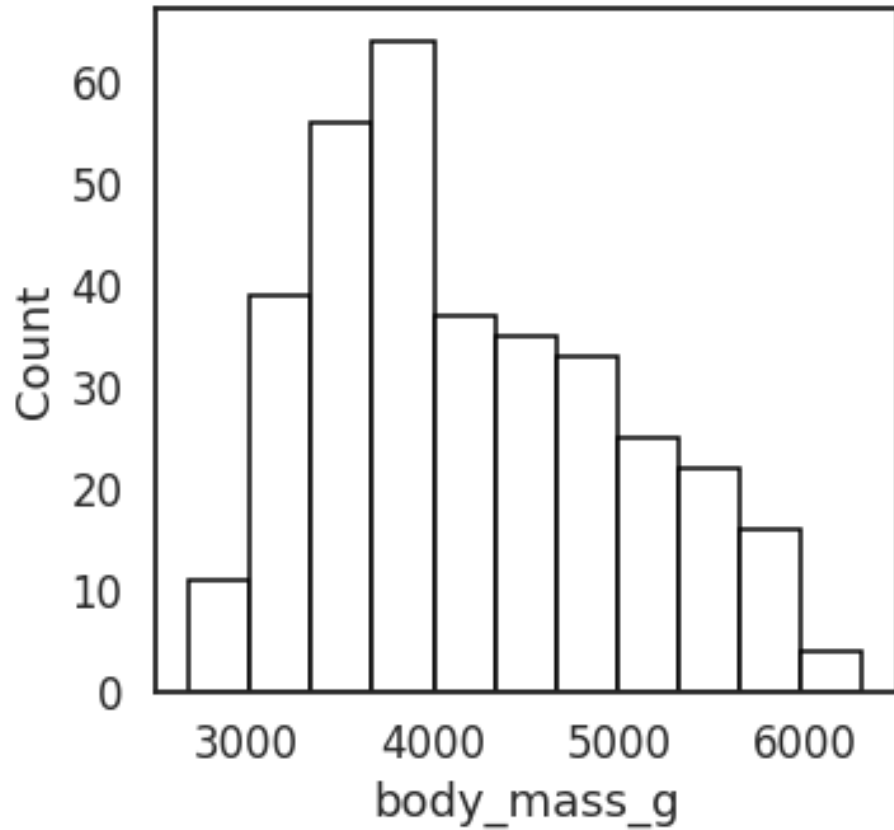
데이터 분포 시각화 방법 : 펭귄 체중



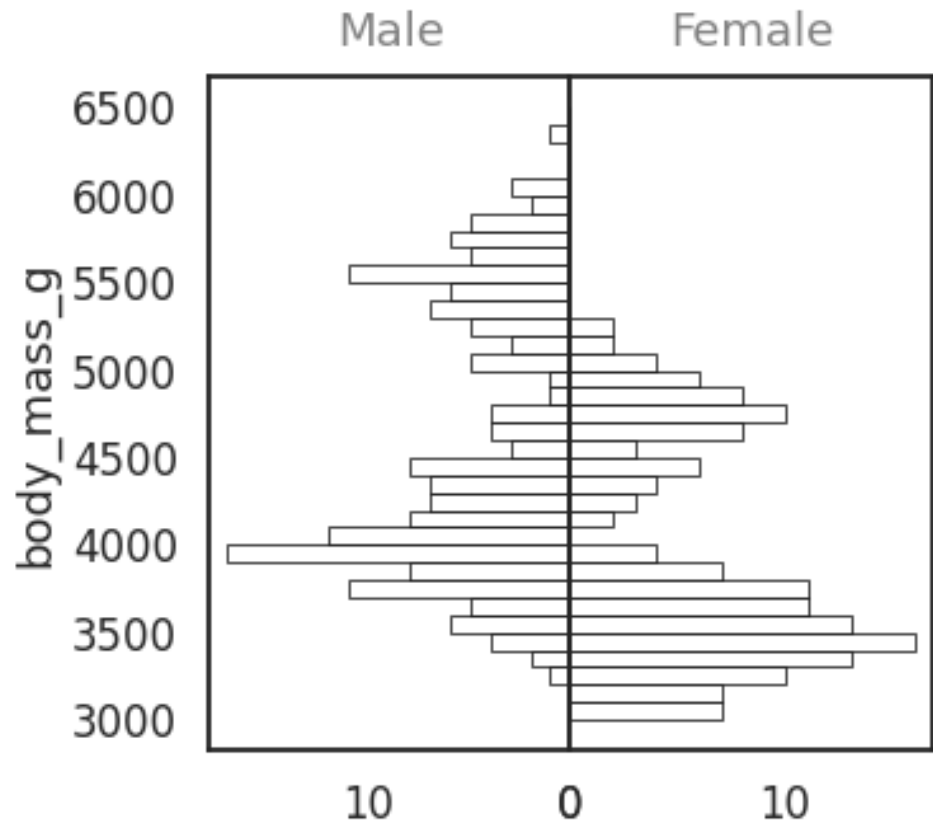
오래된 습관 : Hatching



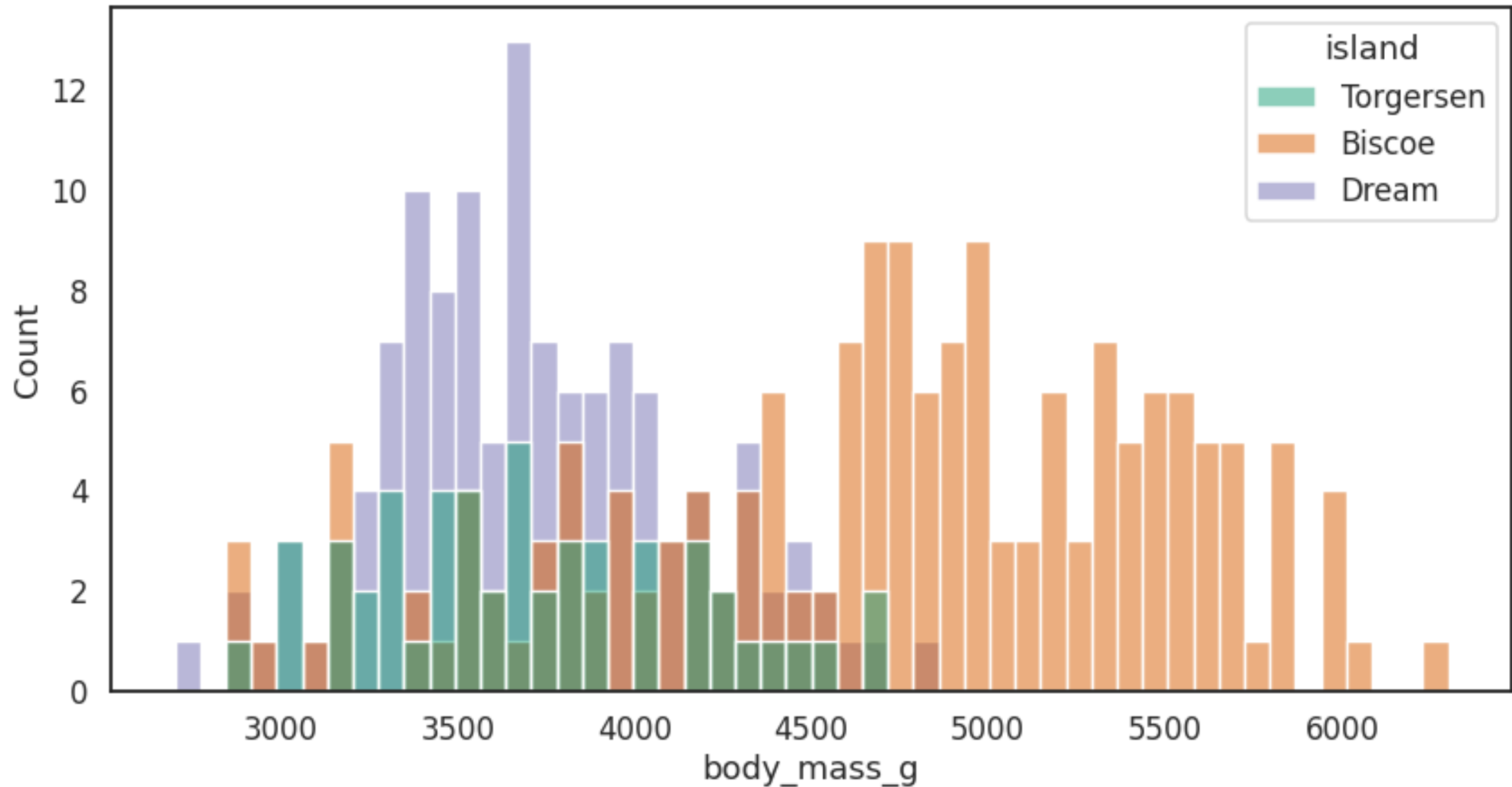
선보다 색을 씹시다



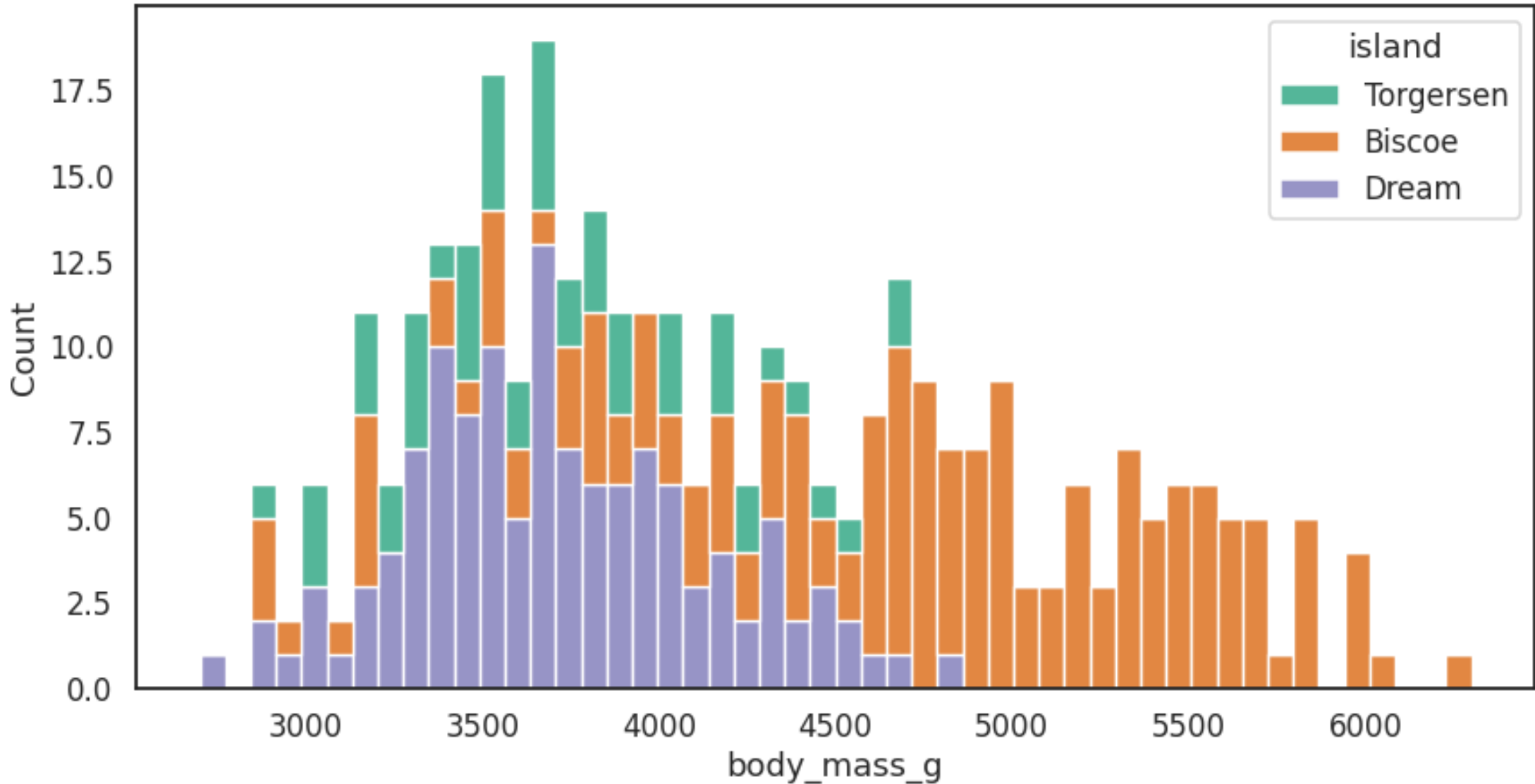
선보다 색을 씹시다



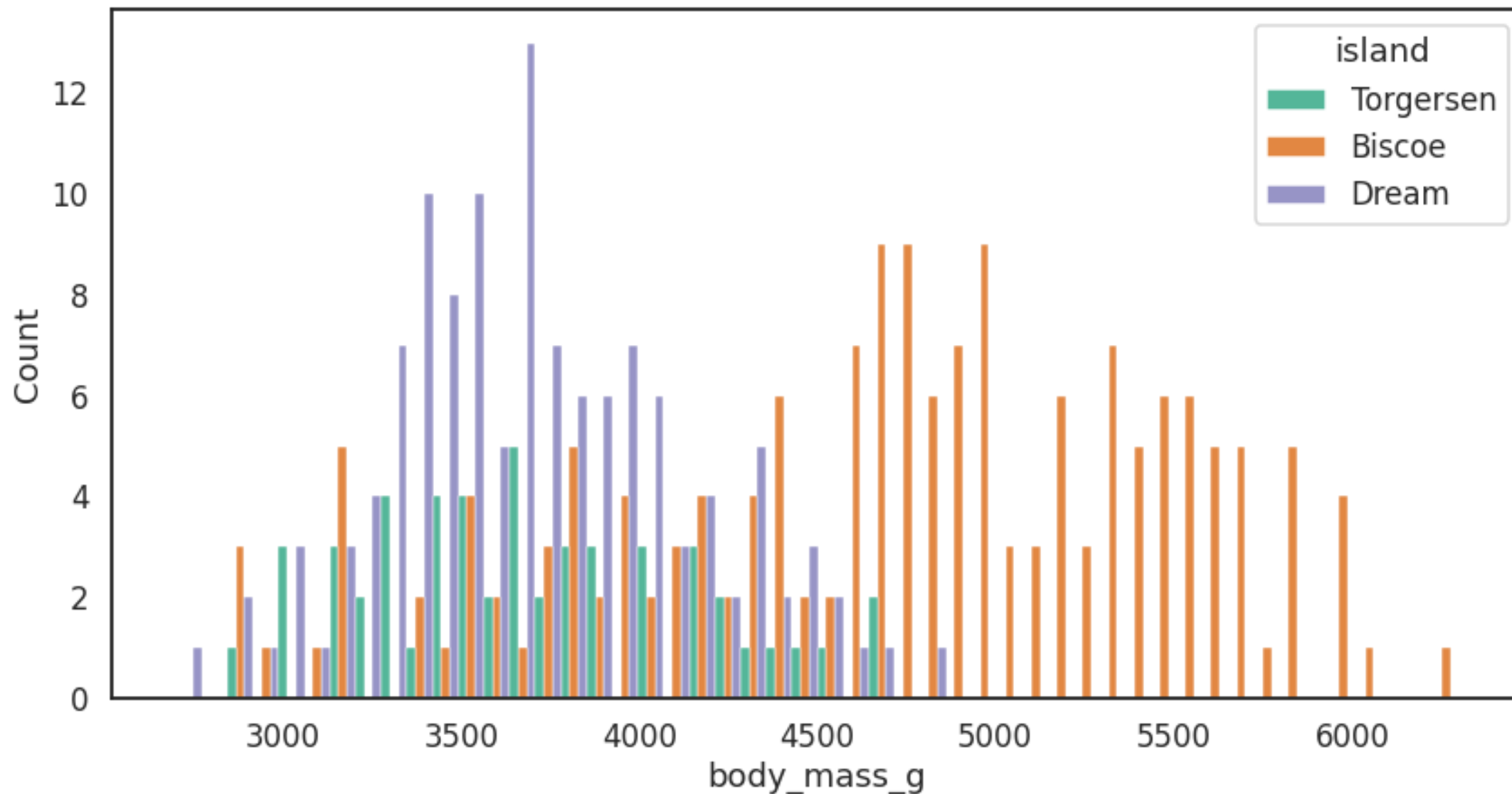
히스토그램을 겹치지 맵시다



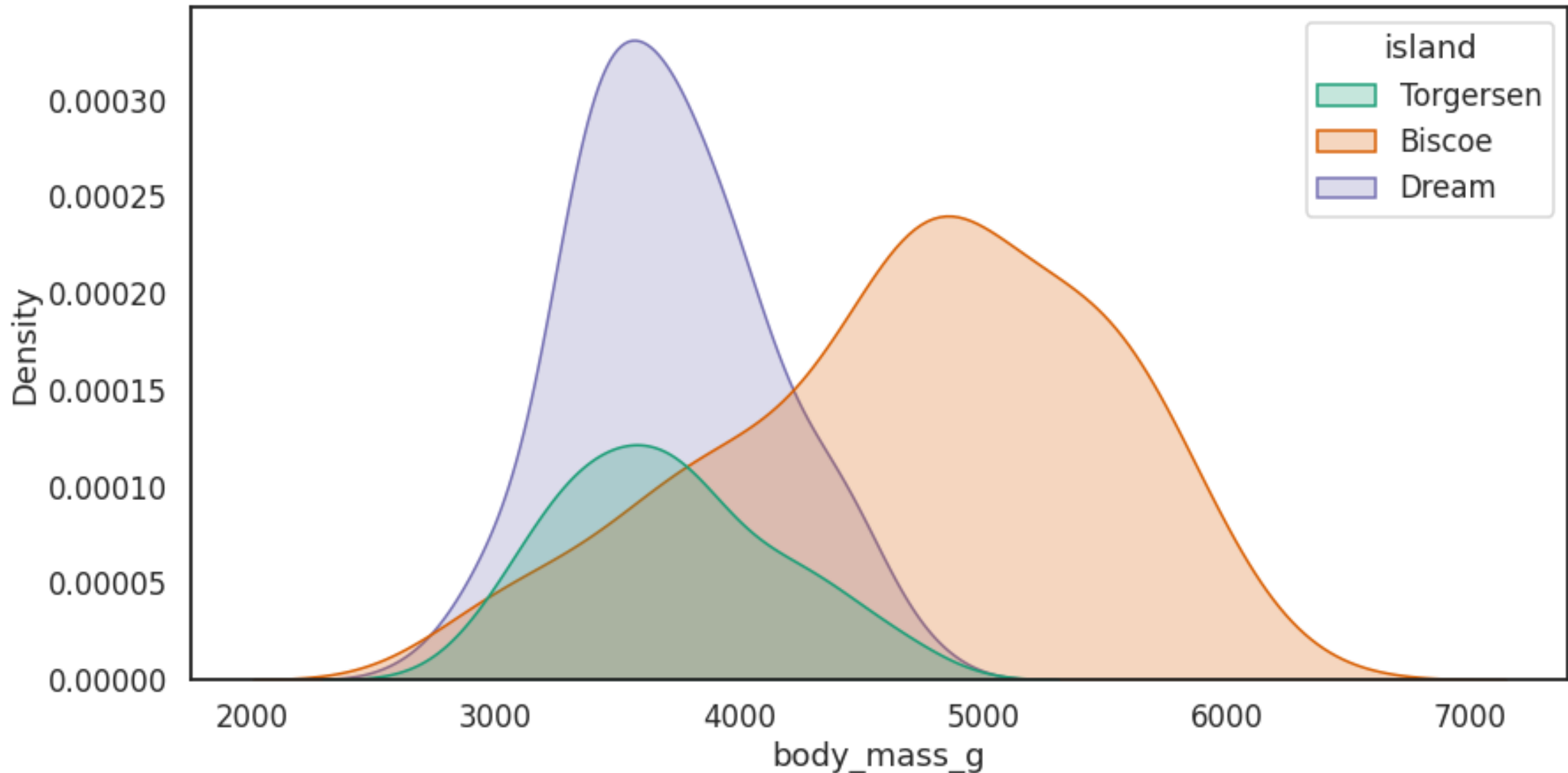
히스토그램 안 겹치기 1: 쌓기^{stack}



히스토그램 안 겹치기 2: 피하기dodge

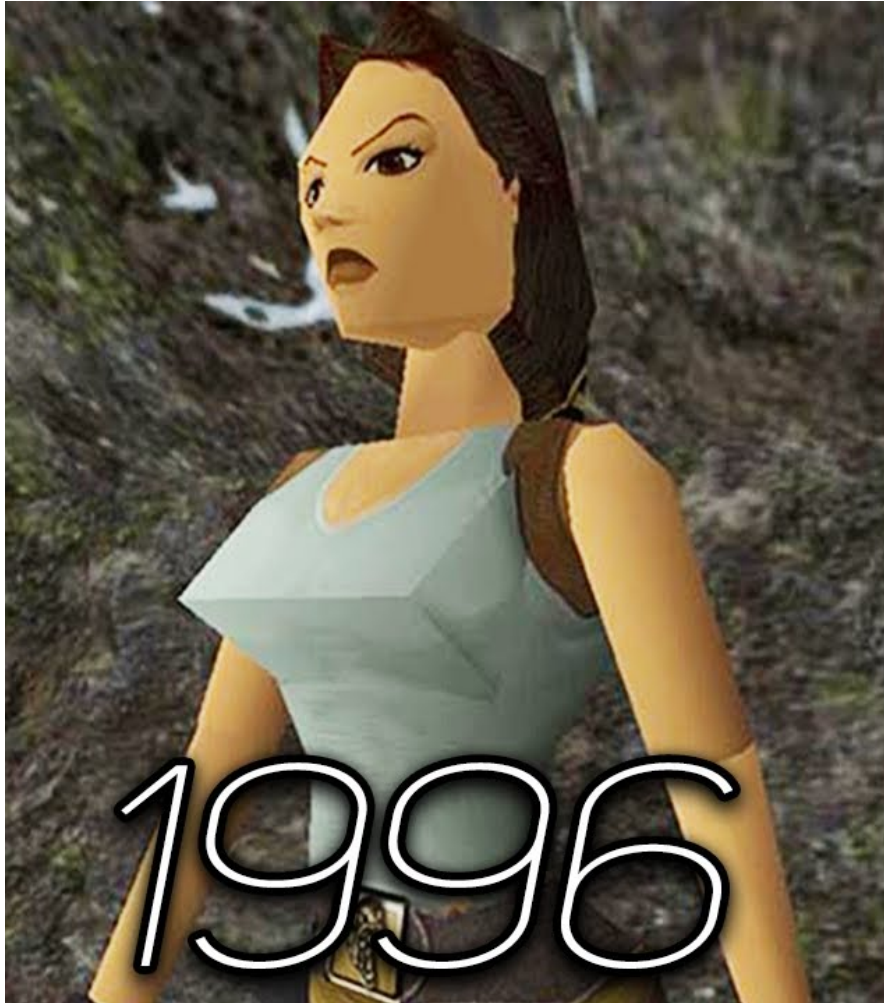


히스토그램 안 겹치기 3: 히스토그램 안 쓰기



시대에 뒤쳐져 보이지 않기를 바랍니다.

- 톰 레이더 (Square Enix)



전달력을 높이려면

- 웬만하면 해칭 안하기
- 히스토그램은 겁치지 말기

약속해!