

I 발표 자료 작성 Guide

1. 15~20분 분량에 맞게 세부 목차 및 페이지 수 자율 조정
2. 목차는 전체 3개 Section으로 구성
 - 1) 팀 소개
 - 2) 분석 보고서/아이디어 제안서 요약 발표
 - 3) Lessons Learned (경연 수행하면서 느낀점 등)
3. 페이지 별로 큰 글씨 사용(폰트14이상), 이미지/그래픽 등 적극 활용
4. 전체 참가자들이 참여해 듣는 발표이므로,
최대한 이해하기 쉽고 재밌게 설명

AI Challenge for Biodiversity

「경연」 Jam Study 발표

맹꽁이2

2022. 12. 28

목 차

1. 팀 소개
2. 주제 선정 배경
3. 데이터 전처리
4. 시각화
5. 결론
6. Q&A

1. 팀 소개

맹꽁이2

: 기후변화 생물지표 100종에 포함된 **맹꽁이**



[멤버 구성]

기지원 (숙명여자대학교 생명시스템학부)

김윤진 (숙명여자대학교 통계학과)

박지원 (숙명여자대학교 통계학과)

유영서 (숙명여자대학교 영어영문학부)

이민지 (숙명여자대학교 통계학과)

2. 주제 선정 배경

설명적/진단 분석의 일환으로 시각화 분석, 상관/인과 분석, 패턴/주기 분석을 수행

- 시민과학자 관찰 데이터에서 자료 개수가 **3번째로 풍부한** 매미 데이터를 활용하는 것이 분석의 **신뢰도**를 향상시킬 수 있다고 판단
- **매미 서식지 특성과 각 매미 종별 개체수의 상관관계**를 시각화하여 매미의 종다양성에 영향을 미치는 요인을 분석 및 진단
- 생물 다양성을 감소시키는 주요 **원인**과 함께 다양한 지역 이해관계자와 데이터를 기반으로 생태계를 분석·예측

3. 데이터 전처리

- POSTING_MASTER 데이터 중에 'CATEGORY' 명이 **매미**인 것만 추출

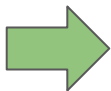
분석에 필요 없는 변수와 모든 행이 결측치인 컬럼명 제거

- 제거한 컬럼명:

'TITLE','READ_CNT','SUBMIT_YN','DEL_YN','CATEGORY','PRESSURE','Column21','Column22','Column2 3','Column24','Column25',
'Column26','Column27','Column28'

- Object 변수

➤ IDX
➤ TEMP
➤ HUMIDITY
➤ WIND_SPEED



- Int/Float형 변수

➤ IDX
➤ TEMP
➤ HUMIDITY
➤ WIND_SPEED

3. 데이터 전처리

POSTING_MASTER

IDX : int(11) - PK
EXPEDITION_NAME : varchar(30)
TITLE : varchar(255)
CATEGORY : varchar(255)
CONTENTS : mediumtext
HOW_MANY : varchar(10)
READ_CNT : int(11)
POS_ADDR : varchar(255)
LATITUDE : varchar(11)
LONGITUDE : varchar(11)
WEATHER : varchar(50)
TEMP : varchar(10)
WIND_SPEED : varchar(10)
HUMIDITY : varchar(10)
PRESSURE : varchar(10)
WRITER_ID : varchar(20) - FK (Refer to MEMBER.ID)
WRITE_TIME : timestamp
DEL_YN : varchar(1)
IS_Q : enum('Y','N')
SUBMIT_YN : varchar(1)



POSTING_EARTH_HISTORY

IDX : int(11) - PK
POSTING_IDX : int(11) - FK (Refer to POSTING_MASTER.IDX)
RESCH_1_Q : varchar(100)
RESCH_2_Q : varchar(100)
RESCH_3_Q : varchar(100)
RESCH_4_Q : varchar(100)
RESCH_5_Q : varchar(100)
RESCH_6_Q : varchar(100)
RESCH_7_Q : varchar(100)
RESCH_8_Q : varchar(100)
RESCH_9_Q : varchar(100)
RESCH_10_Q : varchar(100)
RESCH_11_Q : varchar(100)
RESCH_12_Q : varchar(100)
RESCH_1 : varchar(50)
RESCH_2 : text
RESCH_3 : text
RESCH_4 : text
RESCH_5 : text
RESCH_6 : varchar(25)
RESCH_7 : varchar(25)
RESCH_8 : text
RESCH_9 : varchar(25)
RESCH_10 : varchar(25)
RESCH_11 : varchar(25)
RESCH_12 : text



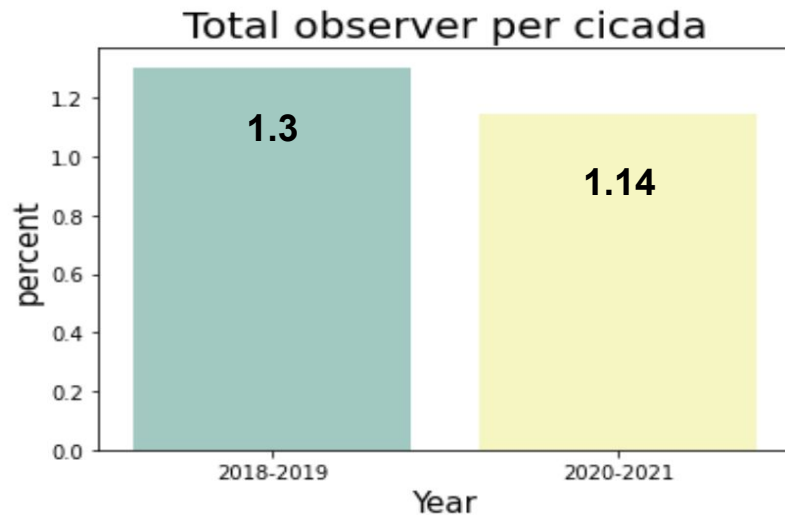
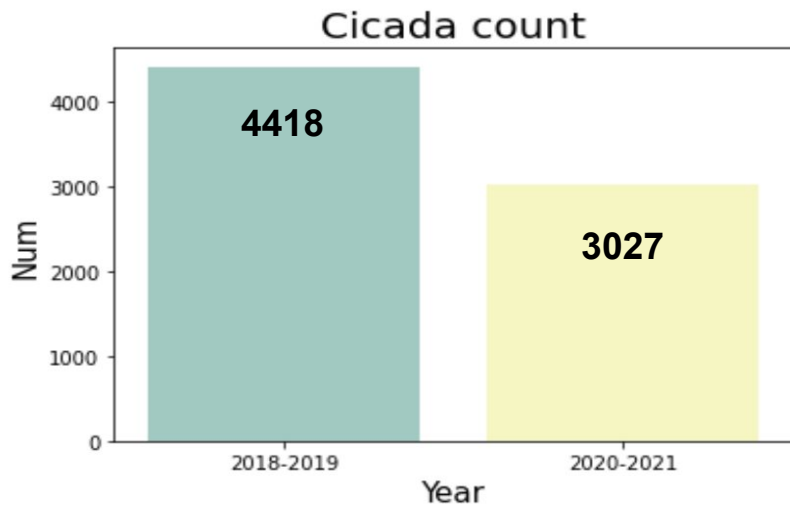
df_merge로 통합

3. 데이터 전처리

- 관찰 연도를 기준으로, 질문이 2갈래로 나뉨
- 데이터 프레임을 2018~2019, 2020~2021로 나눔
- 연도별 데이터 프레임의 경우, 질문이 다른 경우가 있어 새로운 열 생성
→ cicada_type, place_park, place_downtown, place_forest, place_farmland, place_marshland
- 개체수가 적은 소요산매미, 늦털매미 제거
- 2018~2019 데이터의 경우, 도심, 숲 공원, 농경지 등 중복 답변 존재.
- 2020~2021 데이터의 경우, 주거단지(아파트, 빌라), 공원
→ 매미 지도 분석을 위한 데이터 프레임 cicada_final에 place 칼럼 추가
- 매미 종별 지도를 위해 각 종별로 df_매미종 데이터프레임 생성

4. 시각화

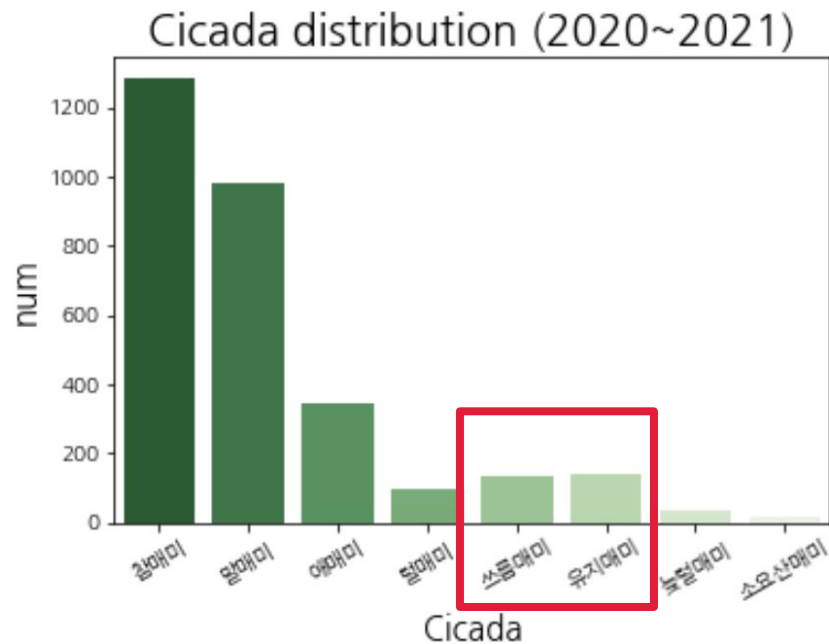
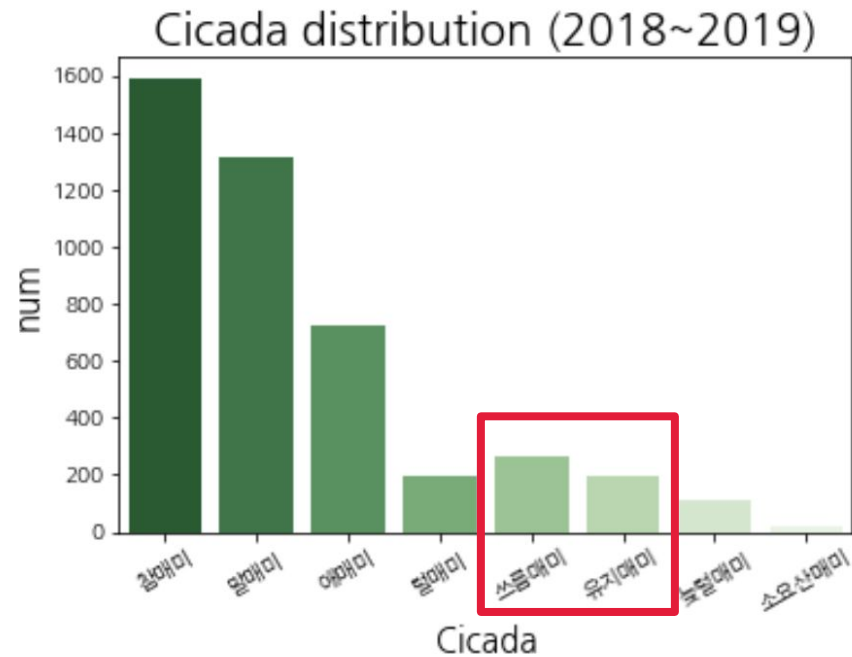
(1) 전체 매미 수 시각화



연도	매미 수	관찰자 수	관찰자 당 매미 수
2018~2019	4418	3387	1.3
2020~2021	3027	2648	1.14

4. 시각화

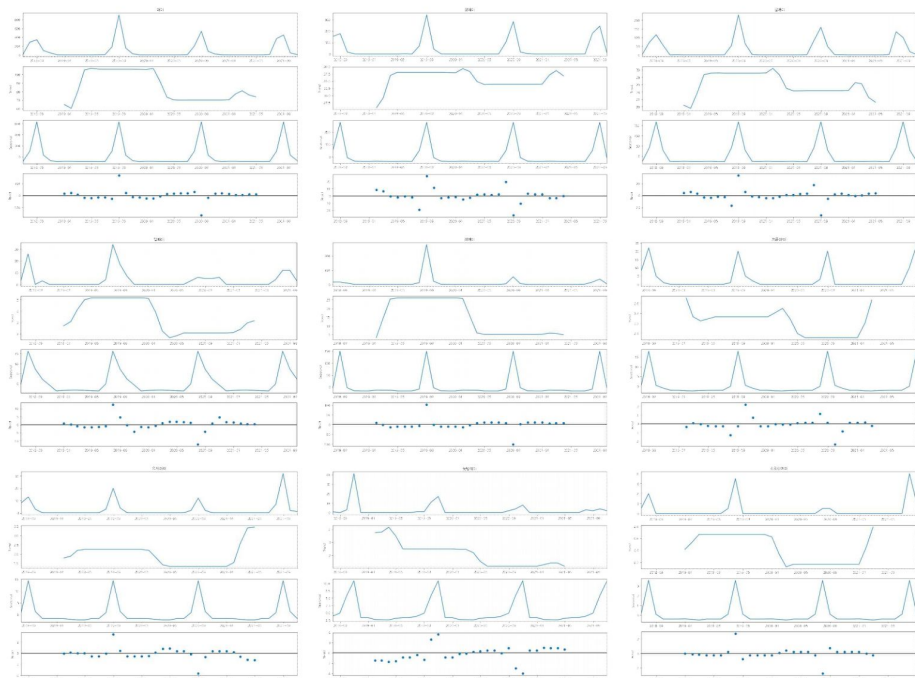
(1) 전체 매미 수 시각화



4. 시각화

(2) 매미 관측 시계열 그래프

매미의 전반적인 출현 기간이 늦어지면서 활동 기간이 감소한 것을 확인할 수 있다.



털매미, 소요산매미(여름 초)

활동 시기 늦어짐

참매미,말매미,쓰름매미(여름 중기)

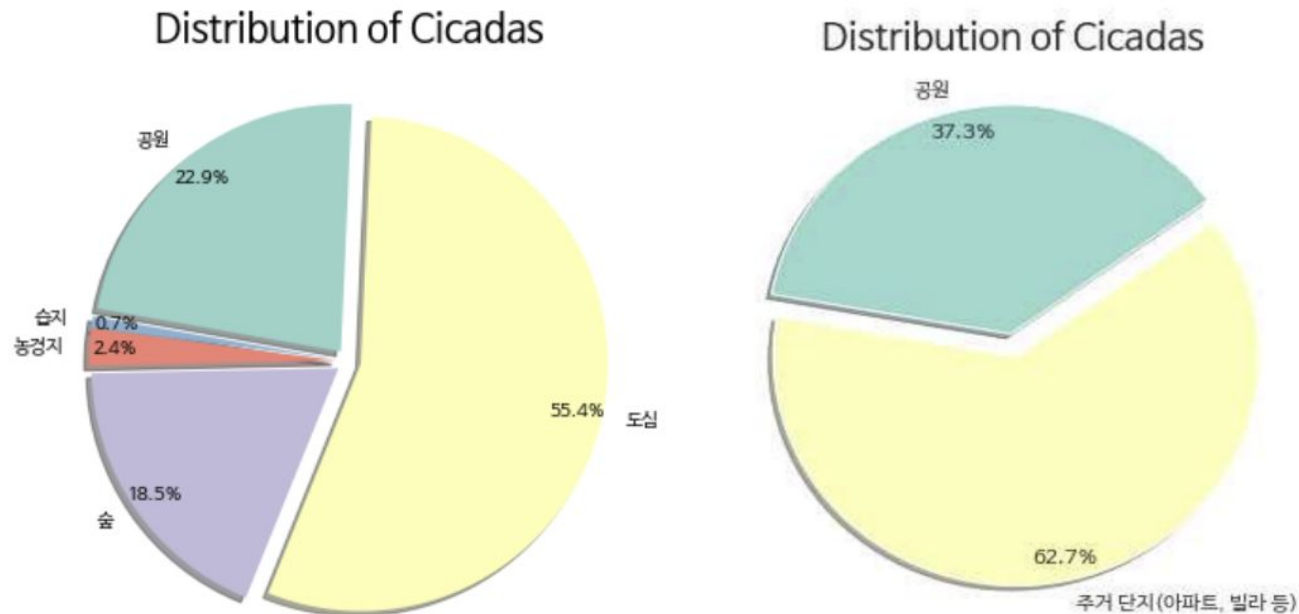
활동 시기 빨라짐

늦털매미(여름 말)

활동 시기 빨라짐

4. 시각화

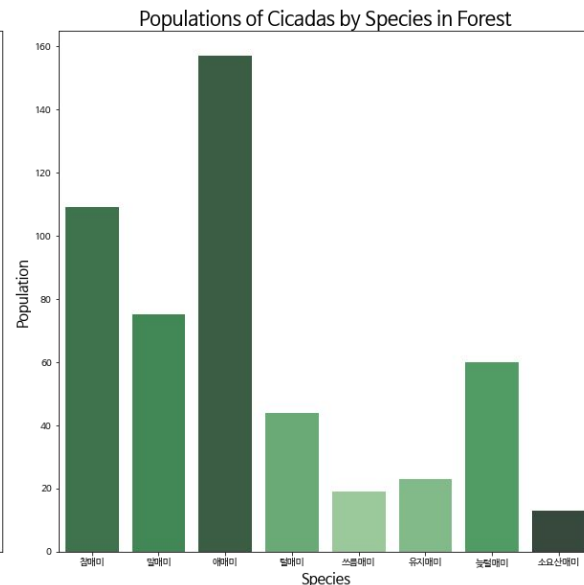
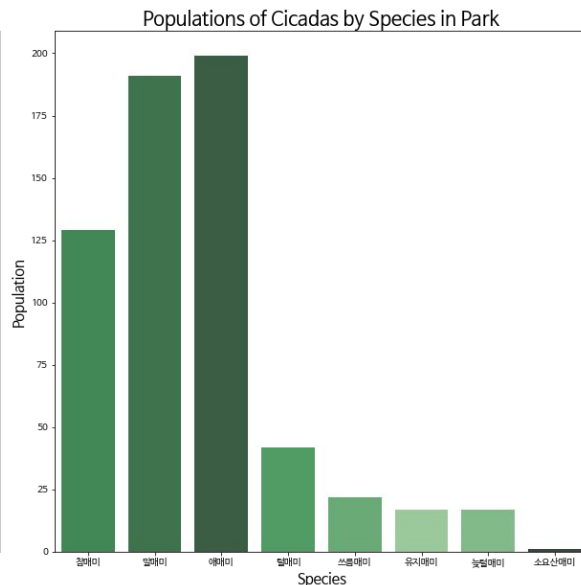
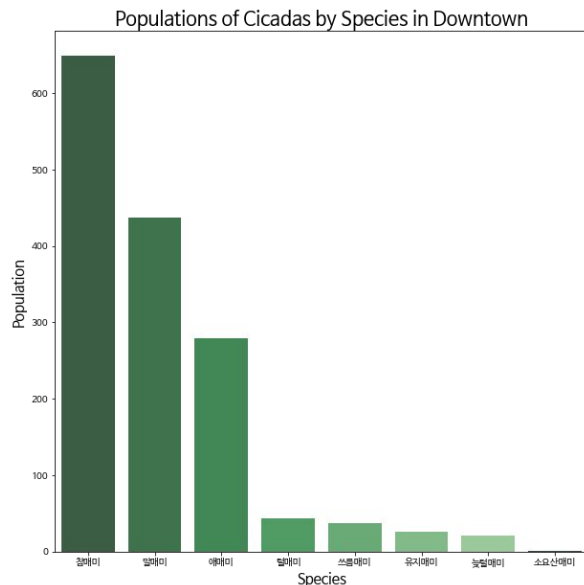
(3) 서식지에 따른 매미 종별 분포



- 2018-2019년도 : 도시 > 공원 > 숲 > 농경지 > 습지
- 2020-2021년도 : 주거 단지가 62.7%로 공원보다 많은 비율을 차지
➢ 시민과학자들의 주 활동 장소와 연관된 것으로 보임.

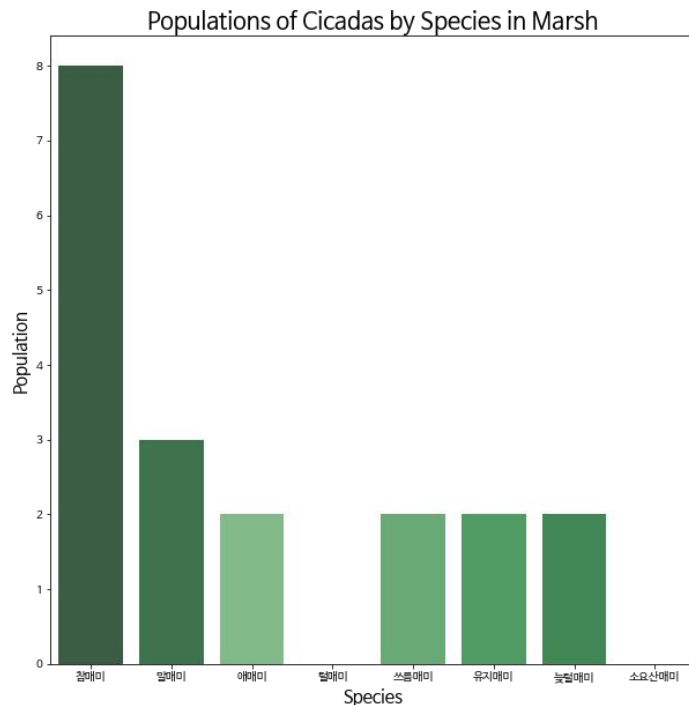
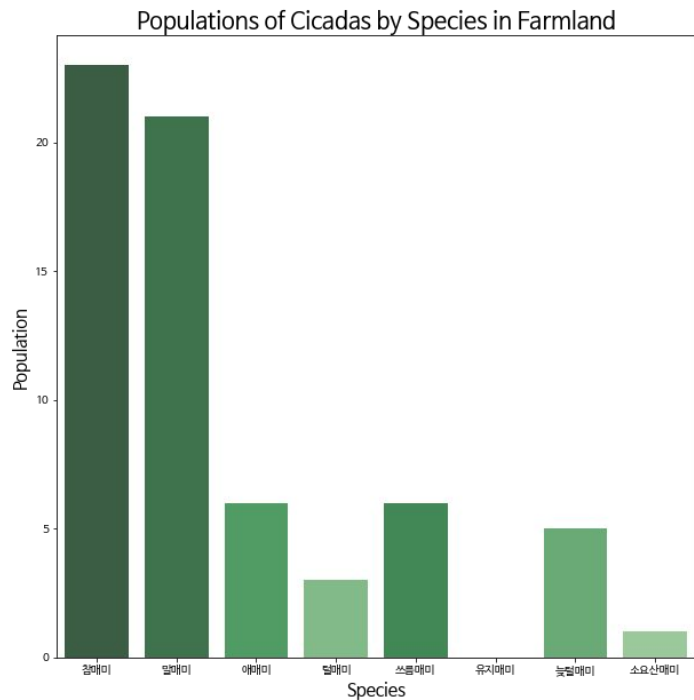
4. 시각화

(3) 서식지에 따른 매미 종별 분포



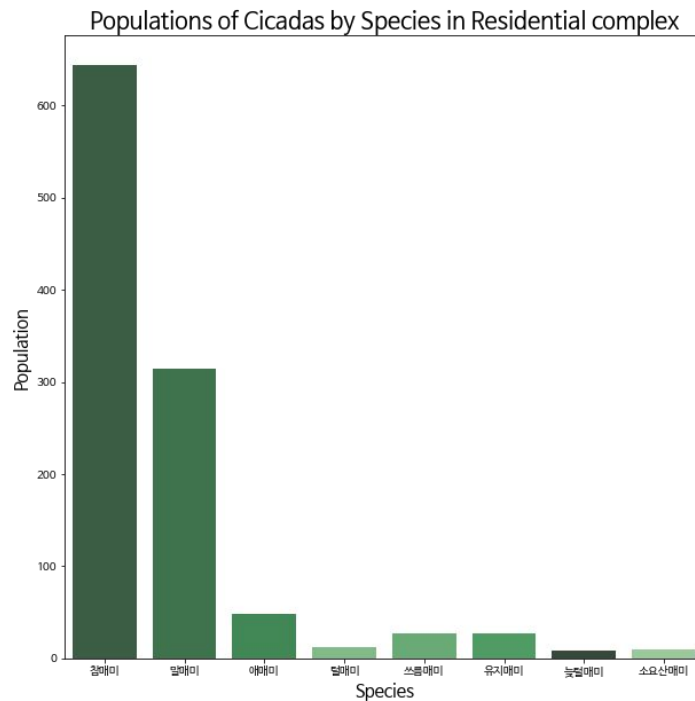
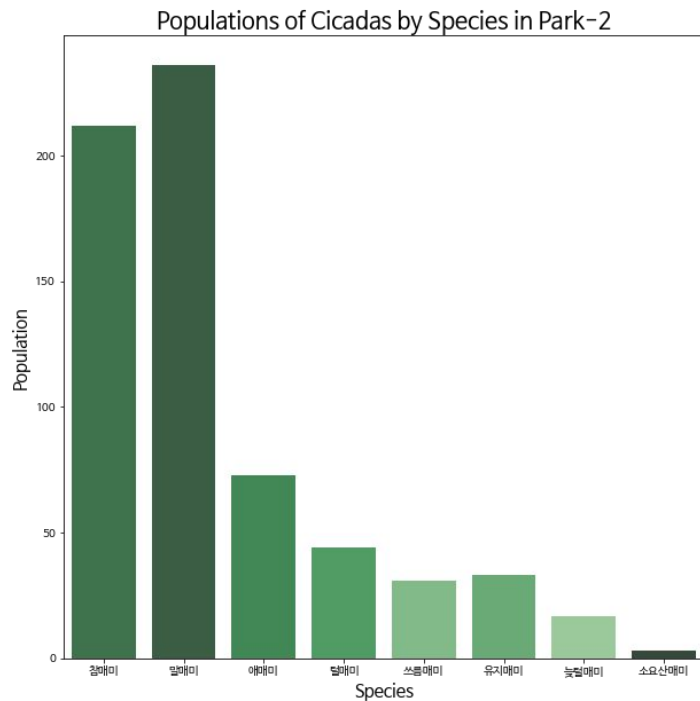
4. 시각화

(3) 서식지에 따른 매미 종별 분포



4. 시각화

(3) 서식지에 따른 매미 종별 분포



- 관측 장소에 따라 매미 종별 분포가 다르고 **도심지로 갈수록 참매미, 말매미의 비율이 우세**
- 종별 분포가 균등하지 못한 것을 확인 가능

4. 시각화

(3) 서식지에 따른 매미 종별 분포

Simpson 지수

분포하는 종의 수가 얼마나 많은지를 나타내는 지수, 풍성함에 민감

Shannon 지수

해당하는 개체 수들이 얼마나 균등하게 분포하는지를 나타내는 지수, 균등함에 민감

- Haines-Young, R. and Chopping, M., 1996, Quantifying landscape structure: a review of landscape indices and their application to forested landscapes, Progress in Physical Geography, 20, 418-445.

⇒ 시민과학자 관찰 기록 특성 상, 표본 수가 도심지에 치우쳐져 있어 표본 크기 변화에 상대적으로 영향을 덜 받는 Simpson 지수를 기준으로 분석을 진행

- Tuomisto, H., 2010, A consistent terminology for quantifying species diversity? Yes, it does exist, Oecologia, 164, 853-860.

4. 시각화

(3) 서식지에 따른 매미 종별 분포

Simpson 지수

한 군집으로부터 두 개체를 랜덤하게 추출하였을 때 두 개체가 같은 종에 포함될 확률인 우점도로부터 계산되는데, 이 수치는 종 풍부도와 개체들의 균등도를 포함한 개념

$$C \text{ (Simpson의 우점도지수)} = \frac{\sum n_i(n_i-1)}{N(N-1)}, \quad (N = \text{total number}, n_i: \text{the number of each species})$$

한 군집의 우점도 지수가 높으면 다양도 지수가 낮게 된다.

=> 다양도 지수는 전체 값 1에서 우점도지수를 뺀 값으로 다음과 같이 나타낼 수 있다.

$$D_s \text{ (Simpson의 다양도 지수)} = 1 - C$$

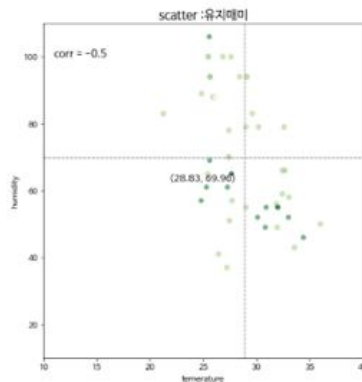
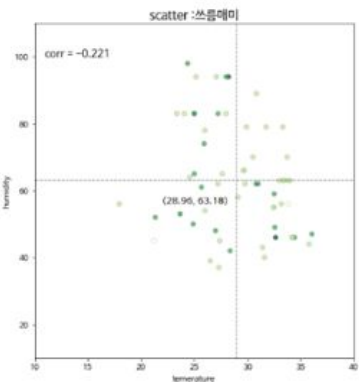
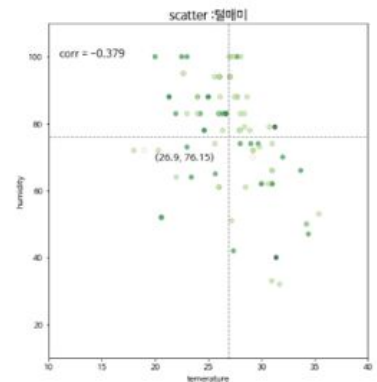
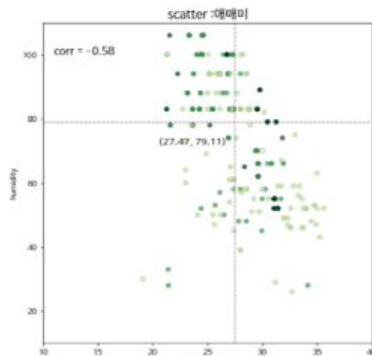
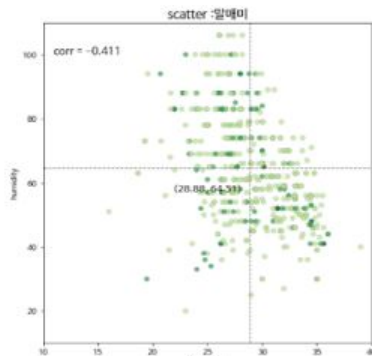
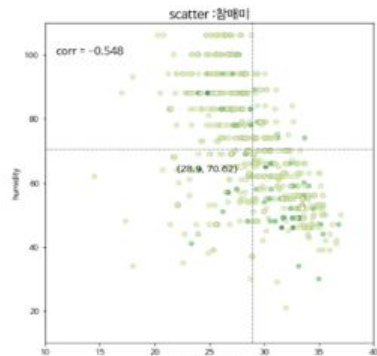
4. 시각화

(3) 서식지에 따른 매미 종별 분포

2018-2019					
관측 장소	도심	공원	숲	농경지	습지
Simpson 지수	0.69	0.75	0.81	0.76	0.8
2020-2021					
관측 장소	주거 단지		공원		
Simpson 지수	0.57		0.74		

4. 시각화

(4) 기온, 습도, 날씨 - 2018~2019



상관계수:

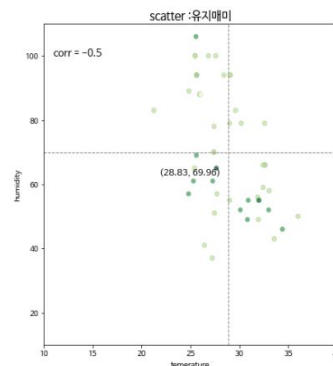
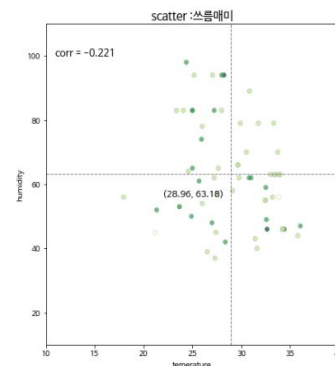
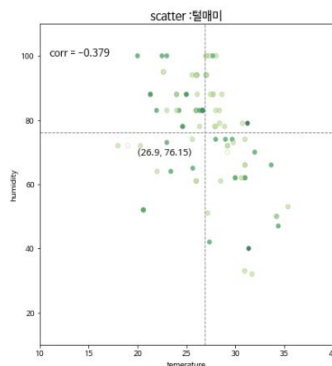
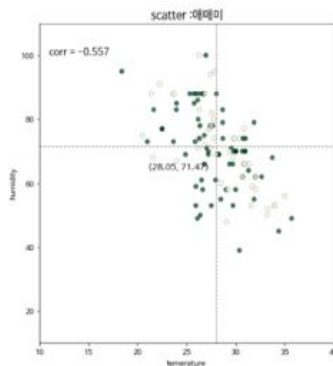
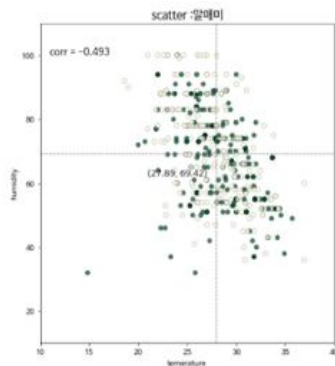
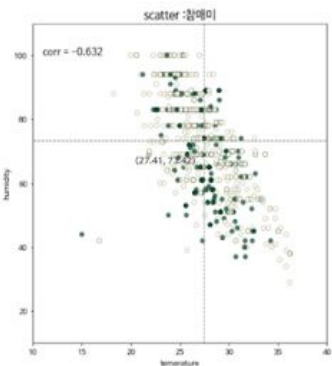
애매미 (0.58) > 참매미 (0.548) >

유지매미 (0.5) > 말매미 (0.411) >

털매미 (0.379) > 쓰름매미 (0.221) 순

4. 시각화

(4) 기온, 습도, 날씨 - 2020~2021



상관계수:
유지매미 (0.697) > 쓰름매미 (0.669) >

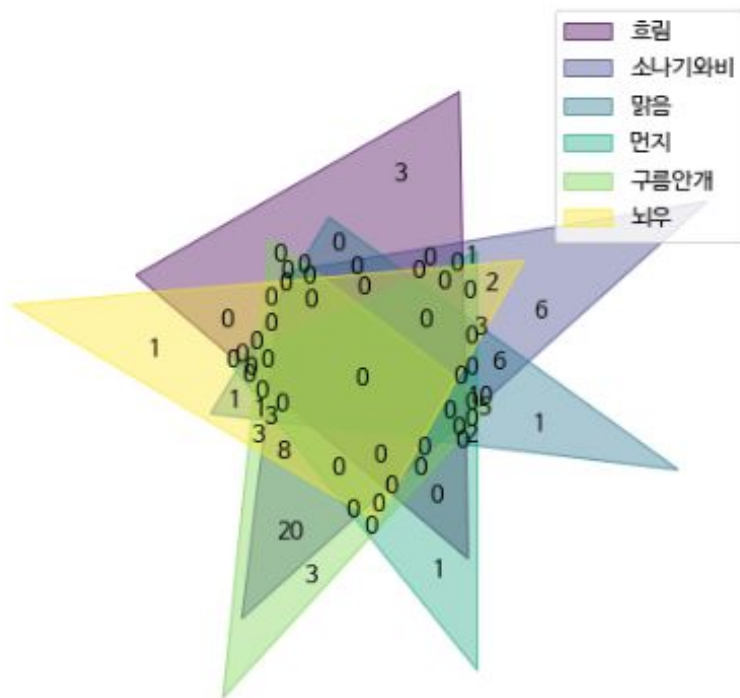
참매미 (0.632) > 애매미 (0.557) >

말매미 (0.493) > 털매미 (0.149) 순

기온과 습도의 상관관계를 통해,
두 요소의 **영향력은 비슷하나**
모든 종의 개체 수에 대해서는 반비례 한다.

4. 시각화

(4) 기온, 습도, 날씨에 따른 매미 분포 - 날씨에 따른 종별 분포 (복수 응답)



‘소나기와 비(‘소나기’, ‘비’가 들어간 응답)’ 와 ‘구름안개(‘구름’, ‘안개’가 들어간 응답)’가 가장 많음을 확인

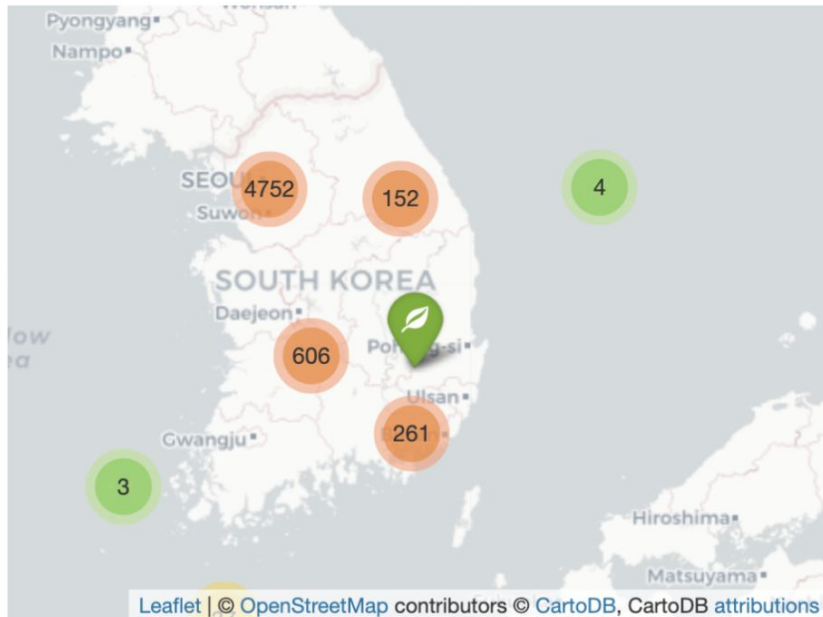
4. 시각화

(5) 맵미 지도

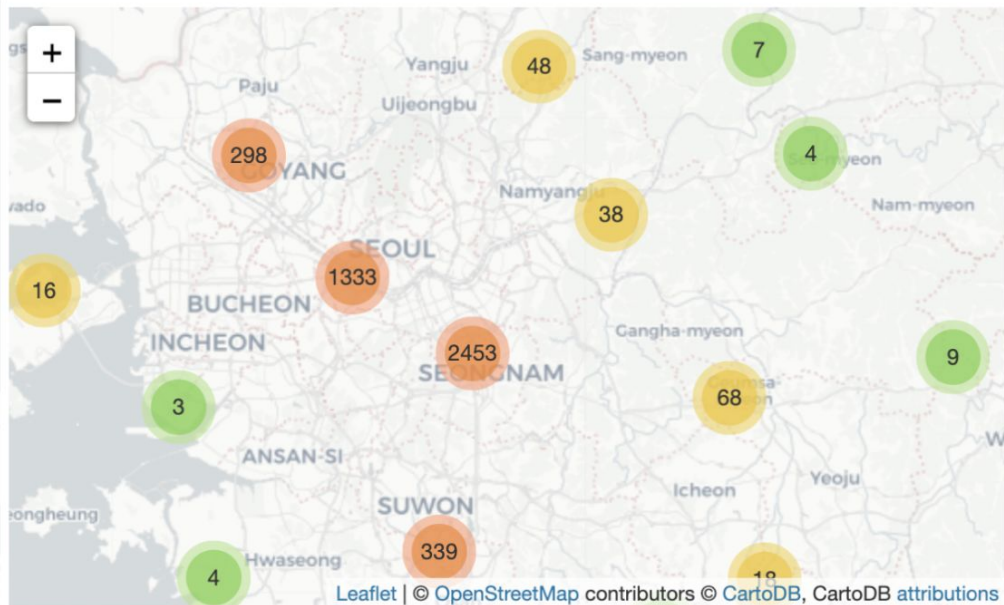
- 지도 시각화를 위해 Python의 **folium** 라이브러리 사용.
- MarkerCluster Map을 활용해 전체 맵미 지도 시각화
- 사용자가 계속해서 확대하는 경우, 숫자가 쪼개지는 형태로 각 맵미의 구체적인 종과 서식지를 확인할 수 있음.
- popup으로 맵미종과 관찰 장소를 표기

4. 시각화

(5) 매미 지도



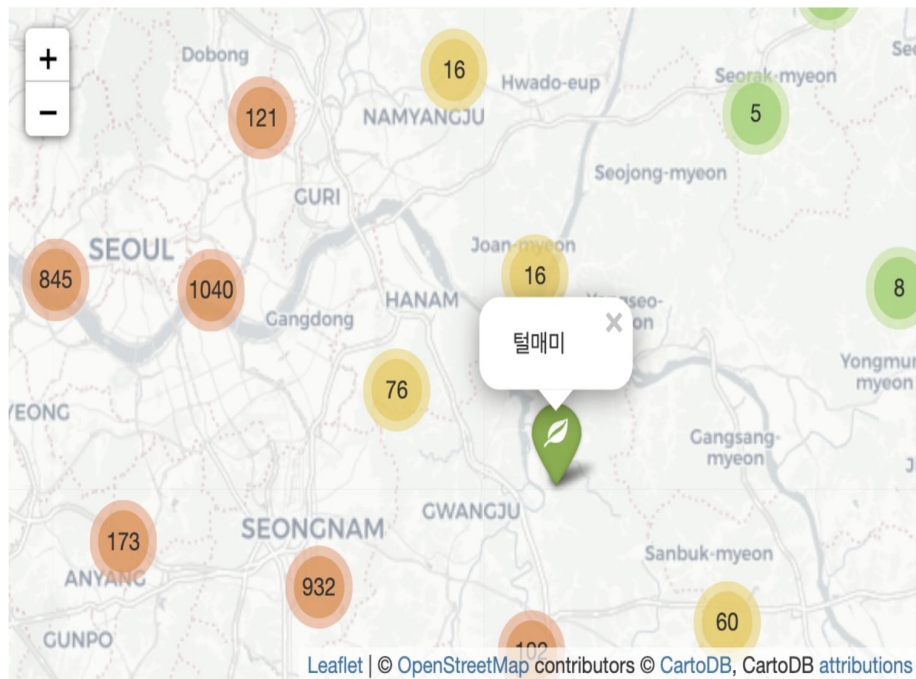
<전체적인 지도>



<확대 했을 때 숫자가 쪼개지는 모습>

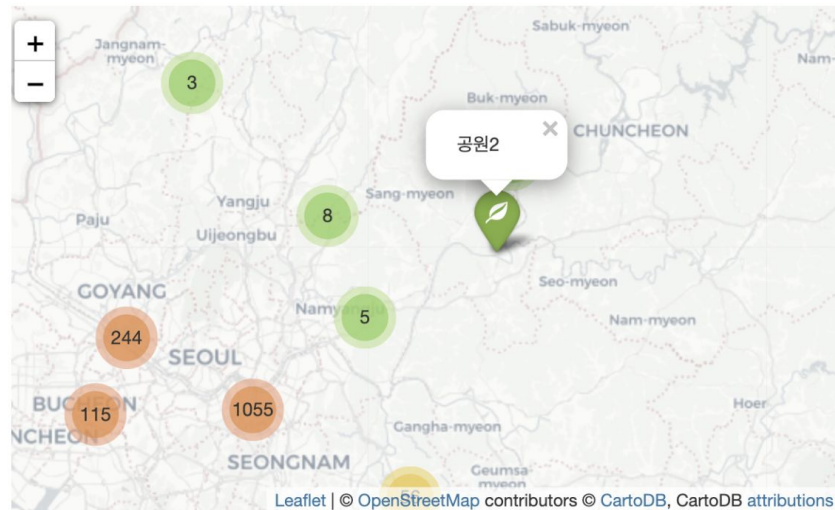
4. 시각화

(5) 매미 지도



<popup시에 매미 종>

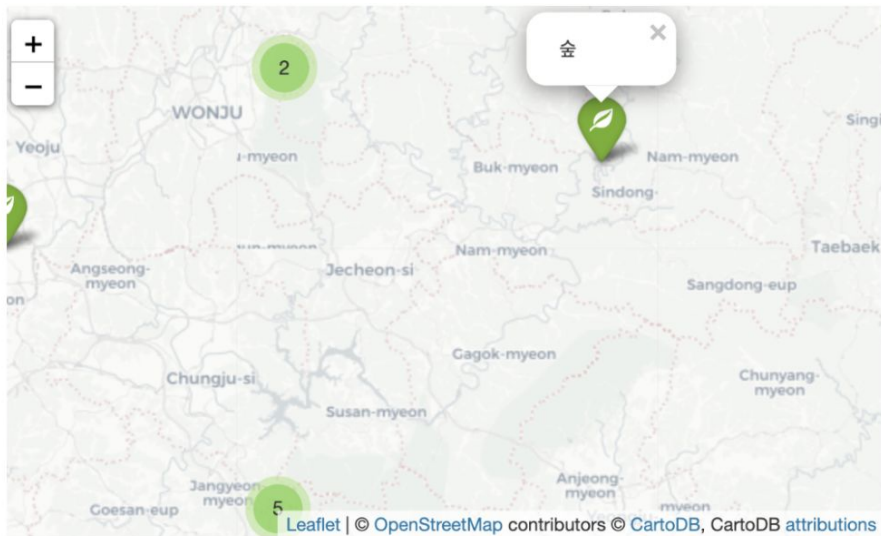
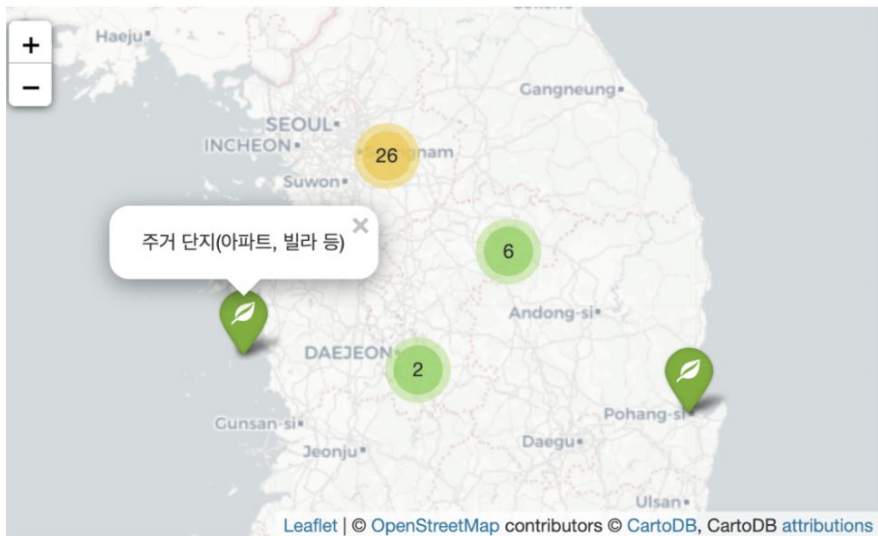
<말매미 지도>



4. 시각화

(5) 매미 지도

<소요산매미 지도>



5. 결론

- 매미는 기후에 따라 개체 수, 번식 율음 여부, 활동 기간 등이 모두 영향을 받고, 종별로 영향을 받는 요인이 각기 다른 곤충이다.
- 이에, 각 매미 종의 관측 시간,기간, 분포를 통해 해당 관측지역/전 지역의 기후 변화를 예측할 수 있다고 판단하였다.
- 참매미, 말매미는 도심에 주로 서식하는 만큼 도심일수록 그 분포가 우세하였다. 이때, **종 다양성 지수가 도시화 정도와 반비례**하는 경향을 보였기에, 도시화 정도를 나타내는 지표로도 활용이 가능해 보인다.
- 시민 과학자들의 특성 상 관측 범위가 넓고 다양하기에, 매미 분포의 환경적 요인을 보다 깊이 파악할 수 있는 자료라고 생각된다.

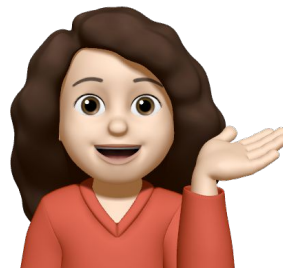
5. Lessons Learned



- 이번 대회를 통해서 단순히 분석을 하는 것보다는, “왜” 이 분석을 진행했는지, 분석 결과를 통해서 “우리는 무엇을 할 수 있을까?” 등이 더 중요하는 것을 배웠습니다.
- 심사위원분들의 평가를 통해 부족한 점을 깨닫고 한층 성장할 수 있는 좋은 기회였습니다.

기지원

- 생물다양성이라는 주제가 광범위한 만큼 접근하기가 어려웠는데, 이번 공모전을 통해서 매미라는 개체를 통해 주제에 대한 접근성을 높인 것 같습니다.
- 또한 생물 종에 대한 관심을 넓히는 계기가 되었습니다.



박지원

5. Lessons Learned



- 초점을 두었던 매미 종에 대해 데이터 분석적 사고를 가지고 상관관계를 파악하는 것은 매우 흥미로웠습니다.
- 뿐만 아니라 시민과학자의 데이터에서 이제까지 몰랐던 다양한 생물종들까지 접할 수 있어 생물다양성을 필히 보존해야 한다는 생각을 되새길 수 있었습니다.

유영서

- Raw data로 주제 선정부터 데이터 가공, 분석까지 진행하는 것은 처음이었는데 생각보다 좋은 정보들이 도출되어 신기했고 다양한 분석을 적용해볼 수 있어 좋았습니다.
- 구체적인 피드백을 통해 보완점을 생각해볼 수 있는 기회가 되었습니다.



김윤진

Q & A