

2023학년도 2학기 언어데이터과학

제7강 데이터과학 (3)

박수민

서울대학교 인문대학 언어학과

2023년 9월 25일 월요일

지난 시간에 배운 것

- 1 배열을 사용한 데이터 표상
- 2 numpy를 이용한 배열 연산

오늘의 목표

- 1 데이터프레임의 정의를 설명할 수 있다.
- 2 pandas 라이브러리를 이용하여 데이터프레임 형태의 데이터셋에서 기초통계량을 계산할 수 있다.
- 3 데이터의 종류에 따라 적절한 그림을 선택할 수 있다.
- 4 seaborn 라이브러리를 이용하여 데이터를 그림으로 표현할 수 있다.

지금 할 일

- 1 VS Code에서 github LDS2023 레포지토리를 연동한다.
- 2 아래 URL의 notes/ 폴더에서 오늘 날짜에 해당하는 파일들을 다운로드한다.
 - <https://github.com/suparklingmin/LingDataSci2023>
- 3 노트 파일을 username/LDS2023/notes 폴더에 저장한다.

마지막으로 할 일

- 1 Commit & Push를 완료한다.

다음 시간에 배울 것

- 1 다양한 빈도 개념

[숙제04] 국어 정보화의 방향

2023-10-04 13:00까지

- 1 <국어 정보화의 방향> 읽고 요약하기

[숙제05] Word Cloud

2023-10-06 13:00까지

- 1 주어진 코퍼스에서 단어 빈도를 계산하고 시각화하기

[숙제06] 8주차 발표 논문 선정

2023-10-06 13:00까지

- 1 발표를 희망하는 논문 선택(3순위까지)