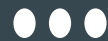


데이터과학



웨어러블 데이터 - Python, R

지난주...

- 제출률 : 31/48
 - 66.7%
- 질문
 - Classroom : 2개
 - E-mail : 8개
- 딜레이 후 과제 제출하면 채점요청 메일보내주세요.
- 과제 결과
 - https://docs.google.com/a/cs-cnu.org/spreadsheets/d/1qR_Alрма9eIqFlvpNW8zMRgoWB7RzoLXtIoRP32ELbI/edit?usp=sharing

목표

- 소쿠리 분석대회 데이터 분석하기
 - 대회 기간동안의 걸음수 TOP 10 구하기
 - 증명하고 싶은 가설을 직접 세우기 (혹은 선택)
 - 그래프로 가설 증명

데이터

- 2016년 소쿠리분석대회
- 68명의 데이터
- 4월1일 ~ 5월20일 데이터
- 일마다 심박수, 걸음수, 수면 데이터

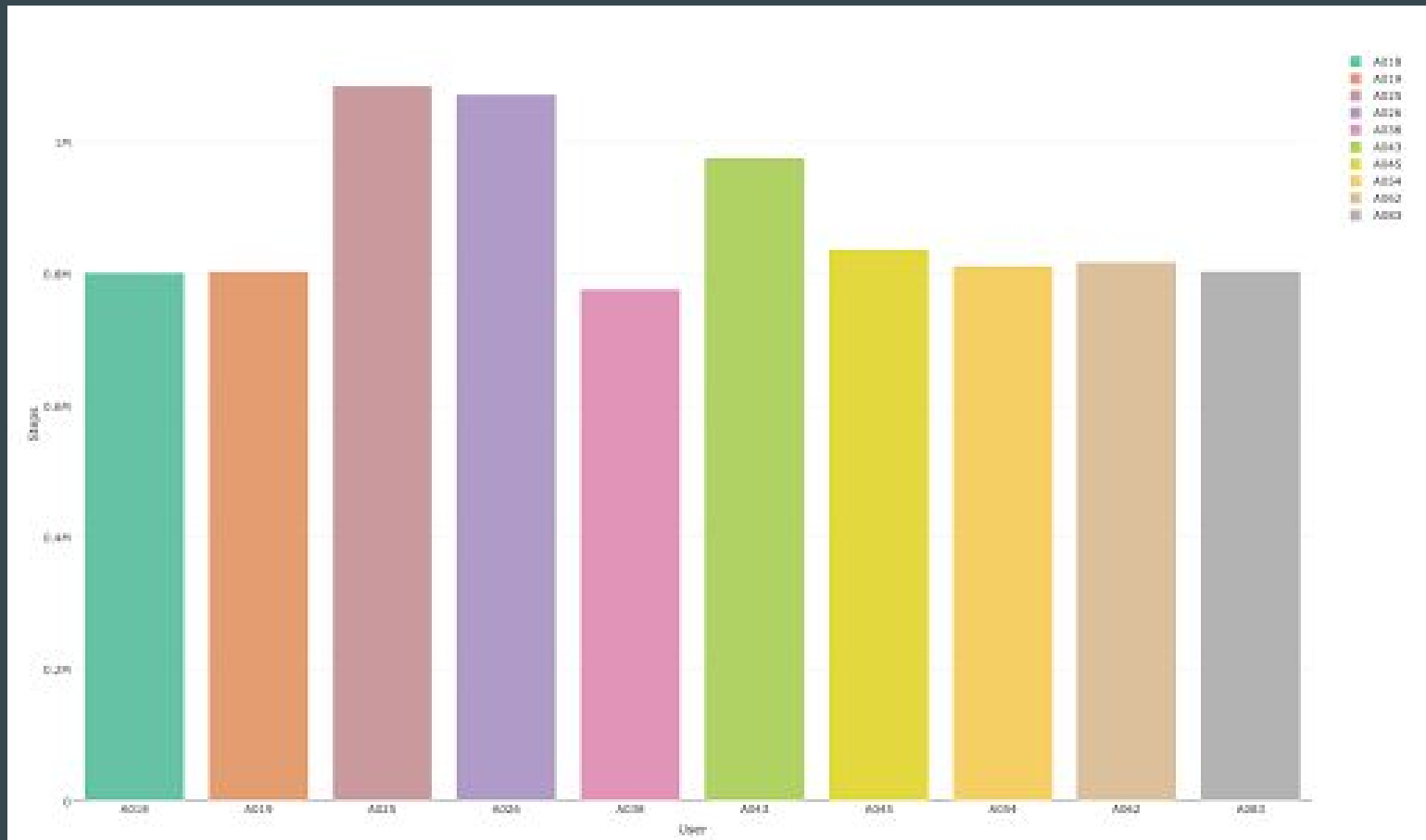
데이터 구조

- sokulee
 - User
 - 20160401_sleep.csv
 - 20160401_steps.csv
 - 20160401_heart.csv
 - ...
 - 20160520_sleep.csv
 - 20160520_steps.csv
 - 20160520_heart.csv

실습 - 걸음수 TOP 10 구하기

- 전체 기간동안의 걸음수 TOP10 구하기
- 데이터 가공과정
 - steps.csv 에서 총걸음수를 어떻게 구할 것인가?
 - 사람별로 , 일별로 나뉘져있는 데이터를 어떻게 합칠 것인가?
- 가시화
 - 그래프?
 - 텍스트?

실습 - 걸음수 TOP 10 구하기



과제

- 해결하고싶은 문제를 직접 설정하고 해결하기
 - a. 데이터 처리 및 가공과정
 - b. 데이터 분석과정
 - c. 가시화 과정

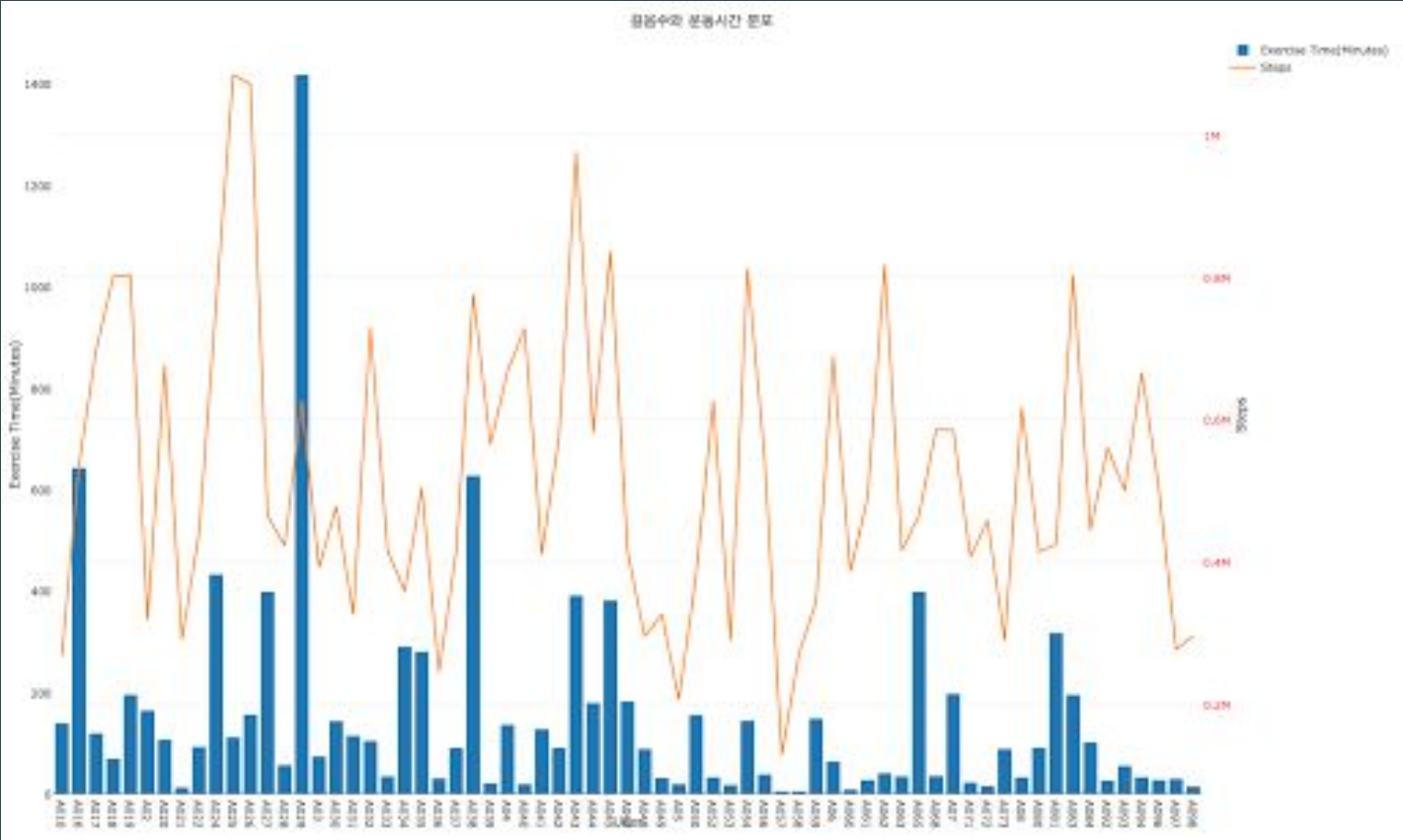
과제 - 가설 예

- 가장 착용 안한 사람찾기
- 아침형 인간과 야행성 인간 구별하기
- 운동을 많이 한사람 찾기
- 사람들이 운동을 많이 하는시간 찾기
- 각 요소별 상관 관계 찾기
 - a. 시간과 걸음수의 관계
 - b. 운동시간과 걸음수의 관계

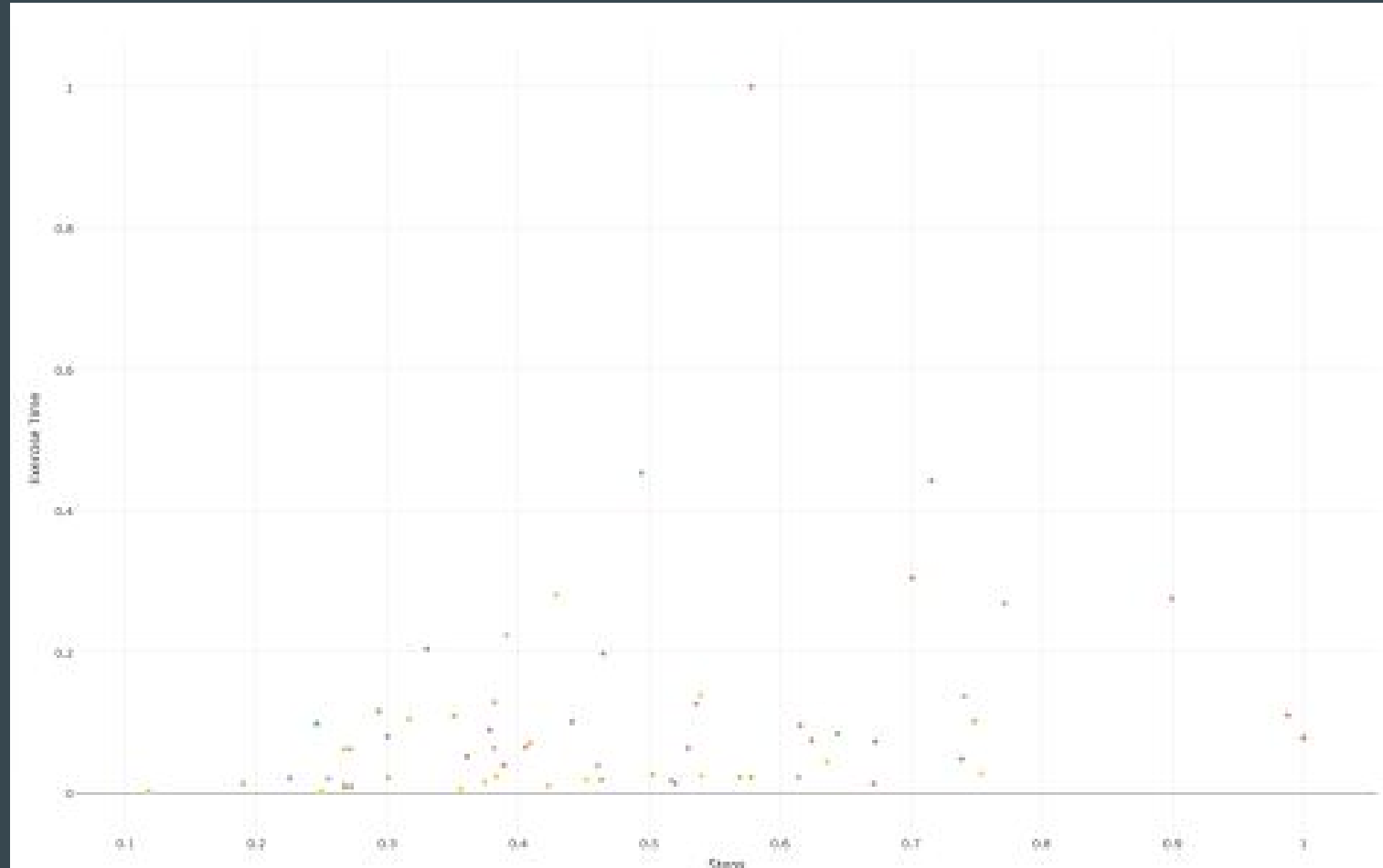
과제 - 예시

- 운동시간과 걸음수의 관계 찾기
- 어떻게?
 - 사용자별 걸음수 분포
 - 사용자별 운동시간 분포
 - 운동시간은 어떻게 계산?
 - heart.csv 의 heartRateZones 필드 사용
 - 두개의 분포를 하나의 그래프로 출력
 - Plotly사용

과제 - 예시



과제-예시



과제 - 예시

- 걸음수가 많다고 운동을 많이 하는것은 아니다.
 - 운동을 많이 하는사람중 걸음수가 적은 사람 존재
 - 걸음수가 많지만 운동시간이 적은 사람 존재
- 운동시간과 걸음수가 정비례 하지 않는다.
 - 대부분은 운동을 적게한다.

과제 정리

- 68명중 걸음수 TOP10 구하기 (텍스트, 가시화)
- 문제 선택 과 해결
 - 데이터 가공과정
 - 데이터 분석과정
 - 가시화 과정

과제 제출 방법

- 과제 제출 기한 : 2017년 4월 5일 **오후 6시까지!**
- Google Classroom에 제출!
 - **24시간 경과시마다 20% 감점**
- 파일 제목 : **DS_학번_이름_주차.pdf, DS_학번_이름_주차.zip**
 - 보고서(PDF형태) : **HWP, DOC**일경우 채점 안함
 - 데이터 가공과정, 데이터 분석 과정, 코드 설명(스크린샷)
 - 결과 그래프
 - 파일 제목 및 형태 틀리면 -1점

질문 사항

- 방문
 - 606호 (데이터네트워크 연구실)
- 메일
 - dbgustlr92@cs-cnu.org
- Google Classroom
 - Good!