

Assignment #1

Due date: 4월 5일

- **metro.csv** : 2019년 11월에 대한 일별 시간대별 지하철역 이용인원 데이터
- **metro_coord.csv** : 7호선 지하철역 위치 (위도, 경도) 데이터

위의 데이터파일을 이용하여 아래의 시각화 과제를 수행하세요. 데이터파일을 R에서 읽어오기 전에 미리 수정하지 않기 바랍니다. 즉, R에서 위의 두 데이터파일을 읽은 후, 각 문항에 적합한 형태로 데이터프레임을 생성하거나 수정하도록 합니다.

1. 송실대입구(살피재) 역의 11월 1일의 시간대별 승차 및 하차 인원 수를 하나의 그래프로 시각화해보자.
2. 송실대입구(살피재) 역의 11월 4일(월)부터 11월10일(일)까지 일주일간 각 요일별 시간대별 승차인원과 하차인원의 분포를 각각 heat map으로 시각화해보자.
3. 7호선의 모든 역 중에서 유동인구(월간 승하차 전체인원)가 가장 많은 20개 역에 대한 유동인구 수를 그래프로 시각화해보자.
4. 7호선 지하철역 위치 정보를 활용하여 7호선의 모든 역에 대한 유동인구 분포를 지도 위에 시각화해보자. 크기, 투명도 등을 활용하여 분포를 표현할 수 있다.

최종적으로 아래의 파일들을 스마트캠퍼스에 제출해야 합니다.

- R script 파일 (comment 포함)
- 생성된 그래프 및 그래프에 대한 간략한 해석을 포함하는 파일

Hint.

1. tidyr 패키지의 gather() 함수 활용하여 24개의 변수로 나누어져 있는 시간대별 인원 수를 하나의 변수로 변환시킬 수 있습니다.
2. scale_fill_continuous(type="viridis") 를 ggplot() 함수에 추가하면 heat map에서의 영역별 차이를 더 선명하게 관찰할 수 있습니다.
3. 강의노트 case study의 마지막 그래프를 참고하세요.
4. dplyr 패키지의 left_join() 함수를 활용하여 같은 변수를 가지는 두 데이터프레임을 하나로 합칠 수 있습니다.