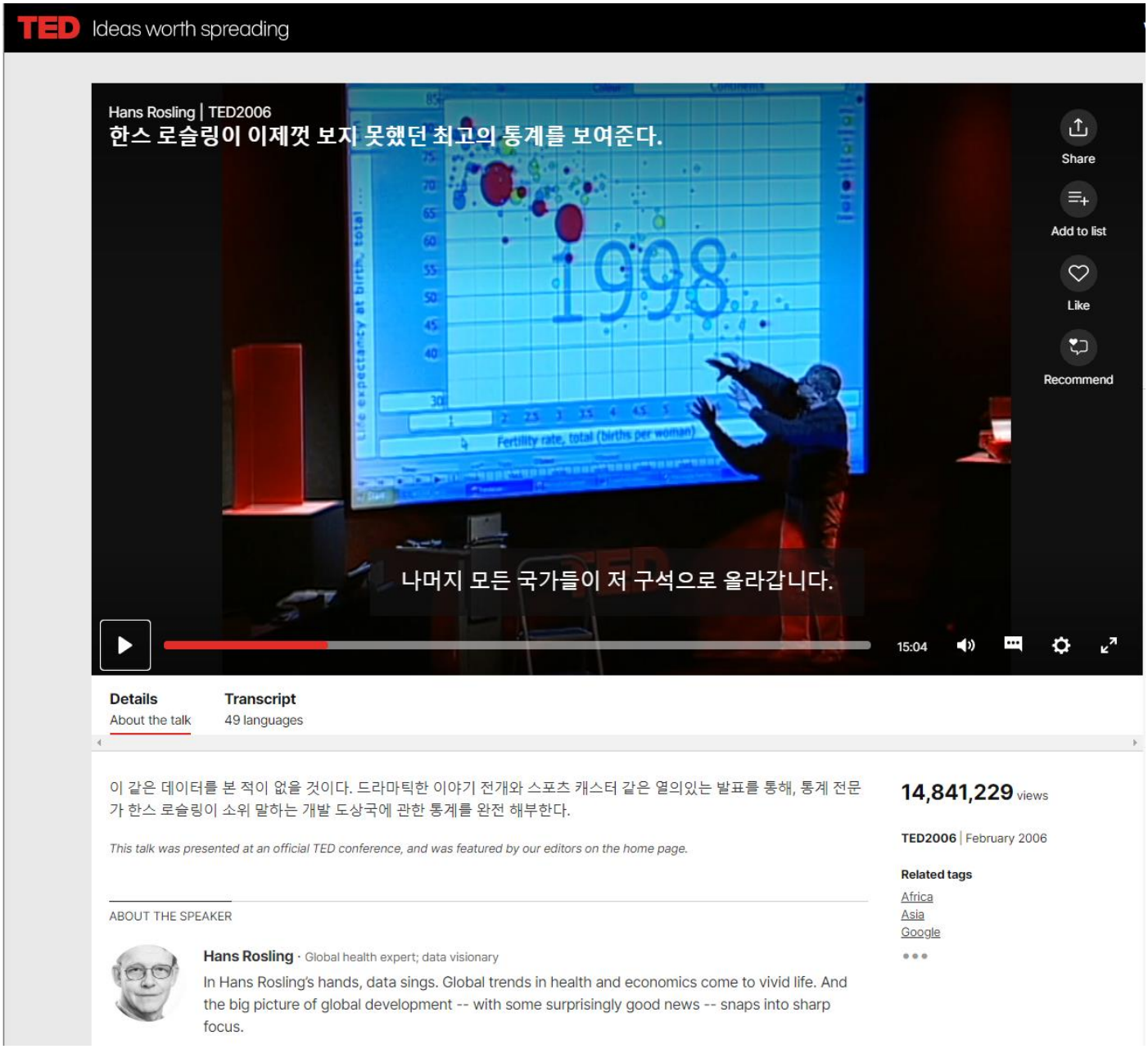
[과제 5] 시각화와 문서화를 통한 탐색적 데이터 분석

최종 결과물인 RMD파일과 이를 변환한 HTML 파일을 모두 제출해야 함

1. 과제의 목적

데이터를 숫자와 문자 그대로만 보는 것이 아니라 이러한 숫자와 문자가 우리에게 주는 메시지를 데이터 분석 및 시각화를 통해 찾는 연습을 하는 것

(데이터 시각화를 통한 메시지 전달의 좋은 예시)



https://www.ted.com/talks/hans_rosling_the_best_stats_you_ve_ever_seen?language=ko#t-270426

2. 주제 선정 및 적합한 데이터셋 다운로드

본인이 본 과제를 수행하면서 하고자 하는 주제를 먼저 설정하고 이를 수행할 수 있는 적절한 데이터셋을 찾아서 다운로드 하시오.

(주제 예시 1) 국가별 COVID-19 발생 동향 분석

(데이터셋이 포함된 웹페이지)

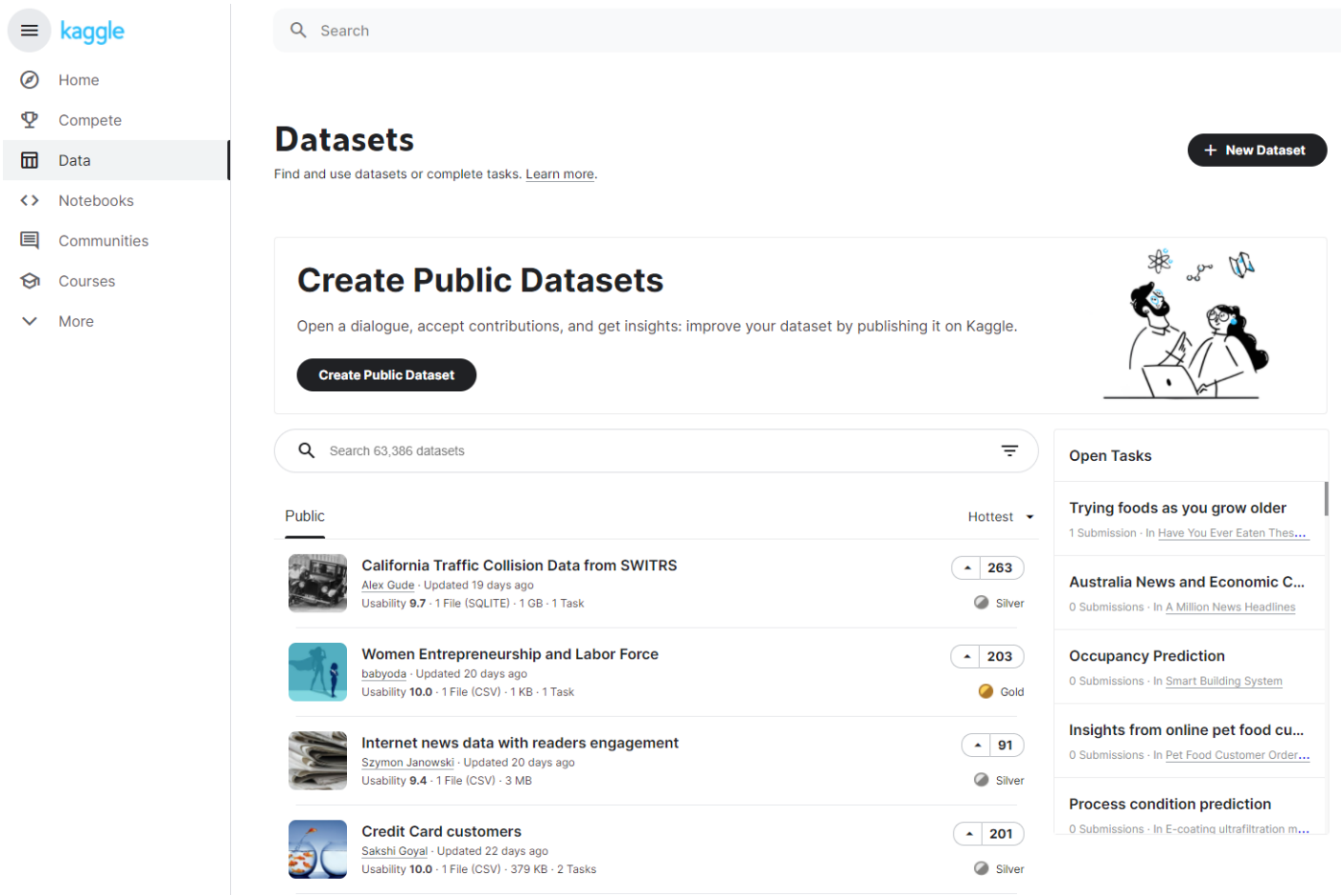
https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide

(주제 예시 2) 우리나라의 산업 성장에 대한 다양한 지표들의 연계 분석

(가용 데이터 출처: 통계청)

http://kostat.go.kr/portal/korea/index.action

주제에 대한 선정이 어려운 학생들은 Kaggle 웹사이트(https://www.kaggle.com/)에서 제시된 데이터셋의 목적을 분석 주제로 선정해도 무방합니다.



3. 데이터에 대한 탐색적 분석 및 시각화

R Notebook과 ggplot2 package에서 제공하는 기능을 사용하여 최소 10개 이상의 다른 형태의 그래프를 그려보고 이를 통해 알 수 있는 사실을 본인만의 언어로 서술하시오.

ggplot2 package로 도시할 수 있는 그래프의 종류는 아래 웹페이지를 포함해서 직접 다양한 사례를 조사하고 이를 적극적으로 활용하세요.

http://r-statistics.co/Top50-Ggplot2-Visualizations-MasterList-R-Code.html

https://www.r-graph-gallery.com/ggplot2-package.html