Dataset generation for M&S

Qc Checklist

**1. Dataset generation strategy**

제공받은 파일은 총 1개(K24027-qPCR AAV DNA result (2-8주차).xlsx)이다. R에서의 data tidying의 편리함을 위해서 xlsx파일을 .csv로 변환한 뒤에 R로 raw상태의 파일을 불러왔다. 파일의 내용은 G1~G11에서 시간에 따른 qPCR의 결과이고 총 7개의 SITE(순서대로 aq, vit, iris, retina, choroid, optic nerve, serum)와 좌/우로 농도값이 나누어져 있었다. ID의 경우 명시된 ID값을 그대로 사용하였고, TIME의 경우 시간(TIME), 날(DAY), 주(WEEK)로 나누었고, 좌우 구분을 위해 LEFT(1: left, 0: right) column을 생성하였고, 편의를 위해서 GROUP, TYPE, BLQ, DRUG(4: Vehicle, 3: 3TF, 2: Aflibercept, 1: AIO), ROW column을 추가하였다.

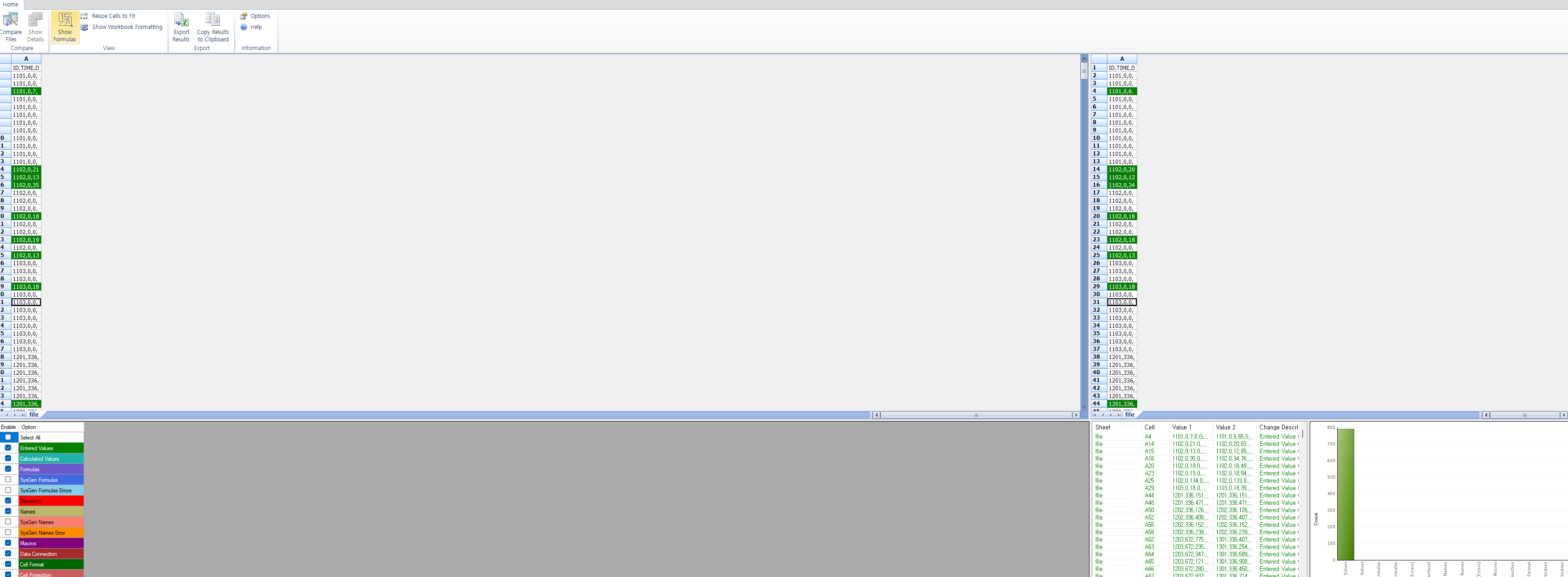
**2. Dataset generation**

Dataset generation은 R (Version 4.3.1), `tidyverse` package (version 2.0.0)을 이용해서 진행하였다. 최종적으로 생성된 dataset은 총 13개의 column, 852의 row가 생성되었다. ID별로 시간에 따른 순서로 정렬하여 최종적으로 dataset를 생성하였다. Dataset을 생성하는데 사용된 R 코드는 아래에 .txt 파일로 첨부하였다.

**2. Dataset evaluation**

Dataset의 비교는 excel에서 inquire 옵션을 사용하였다. Excel inquire는 excel에서 제공하는 추가 기능으로 서로 다른 excel 파일을 비교할 때 사용하는 추가 기능이다. 서로 다른 2개의 dataset를 비교한 결과 column과 동일하게 관찰되었으나 inquire에서 row에서 서로 다른 점을 보고하였다. 농도 값들이 서로 다르다고 보고되었는데 이는 R에서 처리하는 소수점 차이로 인해 서로 같은 값임에도 불구하고 서로 다른 값이라고 인식한 것으로 보인다. R코드 생성한 데이터는 소수점이 모두 1자리 수로 보이는 것으로 보아 raw data의 csv로 변환하는 과정과 R tidying 과정에서 소수점관련 에러가 발생한 것으로 보인다. 그 외에는 특별히 다른 점이 관찰되지 않았다. Inquire 결과는 .xlsx 파일로 저장하였다.

**2. Figures**

****

Excel Inquire과 그 결과 값