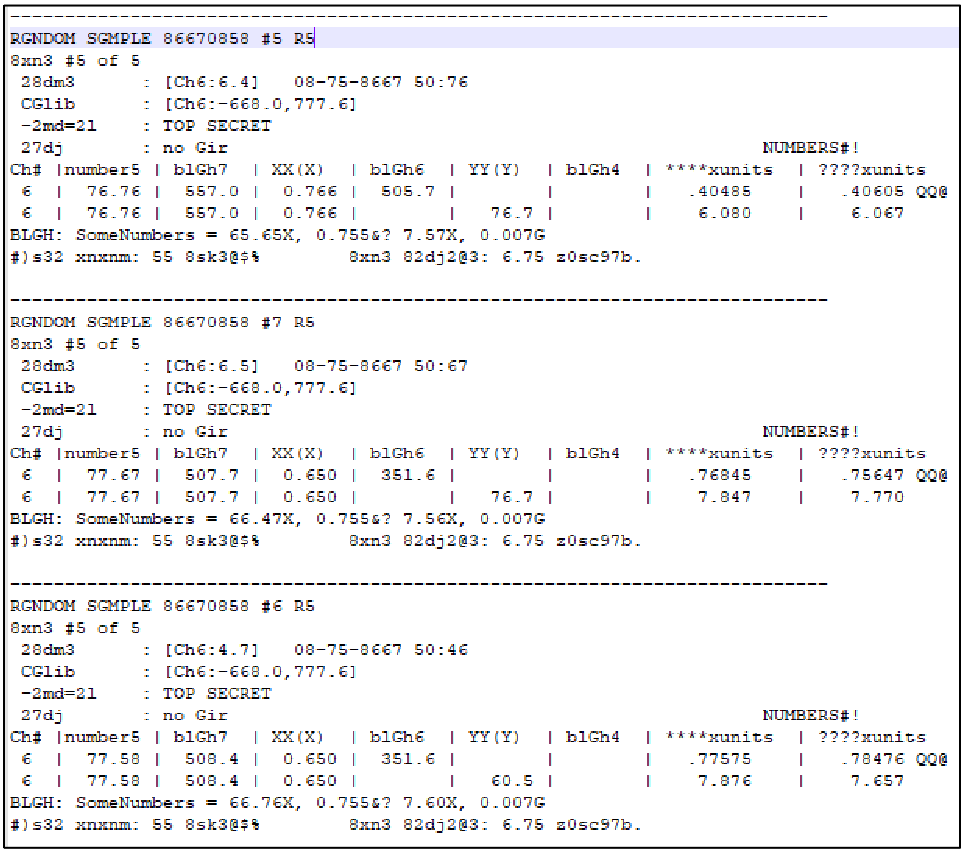
Problem: A .txt file of data was generated by the software for a commonly used lab test. The data was presented in the format seen in Fig. 1. The majority of the data and test parameters were of no concern, important values were visually difficult to discern, and subsequent data analysis was cumbersome.

Solution: A script and web application was written in R and R Shiny, respectively, that utilized parsing and text mining principles (via the *grep()* and *gsub()* functions in the *stringr* package) to extract the useful data, and then produce the desired calculations and statistical analysis on the data set. The user interface (UI) is seen in Figs. 2 and 3.

Example Usage: The user first uploads a .txt file containing the desired data, and selects which machine from which the data file was generated ("Left" or "Right"); each machine generates data files which differ slightly in their format. Once a data file is uploaded and the appropriate machine is selected, a preview of parsed data is automatically displayed under the UI. The user is allowed to use the slider to adjust the number of replicates per sample treatment group, which facilitates statistical calculations. Once the user is satisfied with the preview, they are able to download the results by clicking "Download as .csv".

Benefit: A .csv file was produced as output. The .csv file, shown as Fig. 4, represents the important data and calculations for all samples in a functional and readable format. The resulting output eliminated the need for manual data entry by lab technicians, significantly reducing both the time spent and risk of errors in data preparation for lab reports.

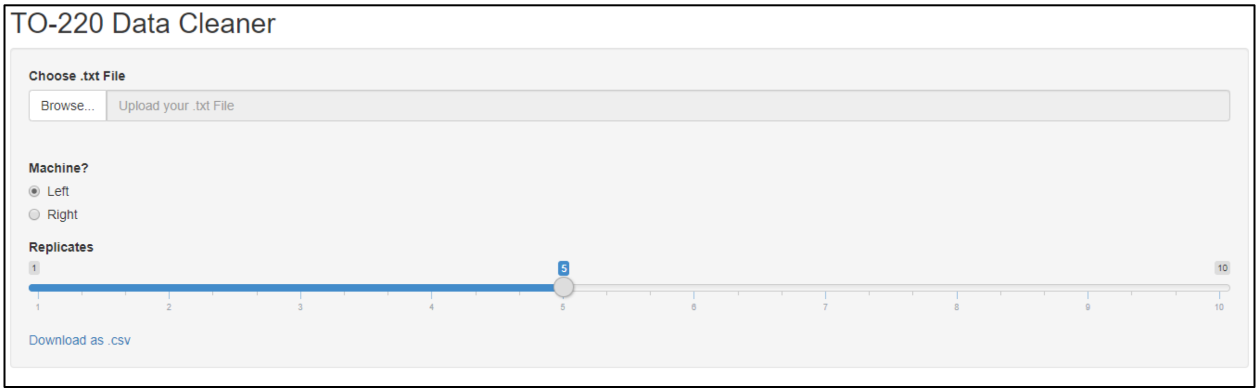


```
---------------------------------------------------------------------------
RGNDOM SGMPLE 86670858 #5 R5
8xn3 #5 of 5
 28dm3      : [Ch6:6.4]   08-75-8667 50:76
 CGlib      : [Ch6:-668.0,777.6]
 -2md=21    : TOP SECRET
 27dj       : no Gir                                                 NUMBERS#!
Ch# |number5 | blGh7  | XX(X)  | blGh6  | YY(Y)  | blGh4  | ****xunits  | ????xunits
 6  |  76.76 |  557.0 |  0.766 |  505.7 |        |        |   .40485    |   .40605 QQ@
 6  |  76.76 |  557.0 |  0.766 |        |   76.7 |        |    6.080    |    6.067
BLGH: SomeNumbers = 65.65X, 0.755&? 7.57X, 0.007G
#)s32 xnxnm: 55 8sk3@$%        8xn3 82dj2@3: 6.75 z0sc97b.

---------------------------------------------------------------------------
RGNDOM SGMPLE 86670858 #7 R5
8xn3 #5 of 5
 28dm3      : [Ch6:6.5]   08-75-8667 50:67
 CGlib      : [Ch6:-668.0,777.6]
 -2md=21    : TOP SECRET
 27dj       : no Gir                                                 NUMBERS#!
Ch# |number5 | blGh7  | XX(X)  | blGh6  | YY(Y)  | blGh4  | ****xunits  | ????xunits
 6  |  77.67 |  507.7 |  0.650 |  351.6 |        |        |   .76845    |   .75647 QQ@
 6  |  77.67 |  507.7 |  0.650 |        |   76.7 |        |    7.847    |    7.770
BLGH: SomeNumbers = 66.47X, 0.755&? 7.56X, 0.007G
#)s32 xnxnm: 55 8sk3@$%        8xn3 82dj2@3: 6.75 z0sc97b.

---------------------------------------------------------------------------
RGNDOM SGMPLE 86670858 #6 R5
8xn3 #5 of 5
 28dm3      : [Ch6:4.7]   08-75-8667 50:46
 CGlib      : [Ch6:-668.0,777.6]
 -2md=21    : TOP SECRET
 27dj       : no Gir                                                 NUMBERS#!
Ch# |number5 | blGh7  | XX(X)  | blGh6  | YY(Y)  | blGh4  | ****xunits  | ????xunits
 6  |  77.58 |  508.4 |  0.650 |  351.6 |        |        |   .77575    |   .78476 QQ@
 6  |  77.58 |  508.4 |  0.650 |        |   60.5 |        |    7.876    |    7.657
BLGH: SomeNumbers = 66.76X, 0.755&? 7.60X, 0.007G
#)s32 xnxnm: 55 8sk3@$%        8xn3 82dj2@3: 6.75 z0sc97b.
```

Figure 1: Original form of test data in .txt file, from which data extraction was quite unenjoyable.



TO-220 Data Cleaner

Choose .txt File

Browse... | Upload your .txt File

Machine?

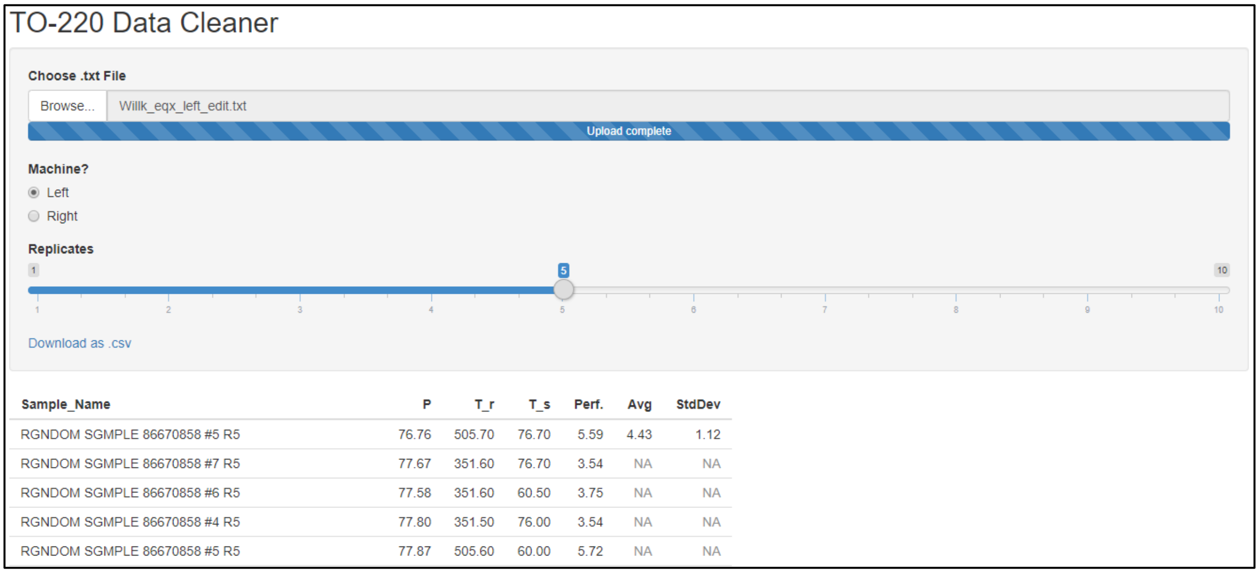
● Left

○ Right

Replicates

Download as .csv

Figure 2: User interface of application. User uploads her .txt file, selects the machine from which the data was generated, and selects the appropriate number of replicates via slider.



TO-220 Data Cleaner

Choose .txt File

Browse... | Willk_eqx_left_edit.txt

Upload complete

Machine?

● Left

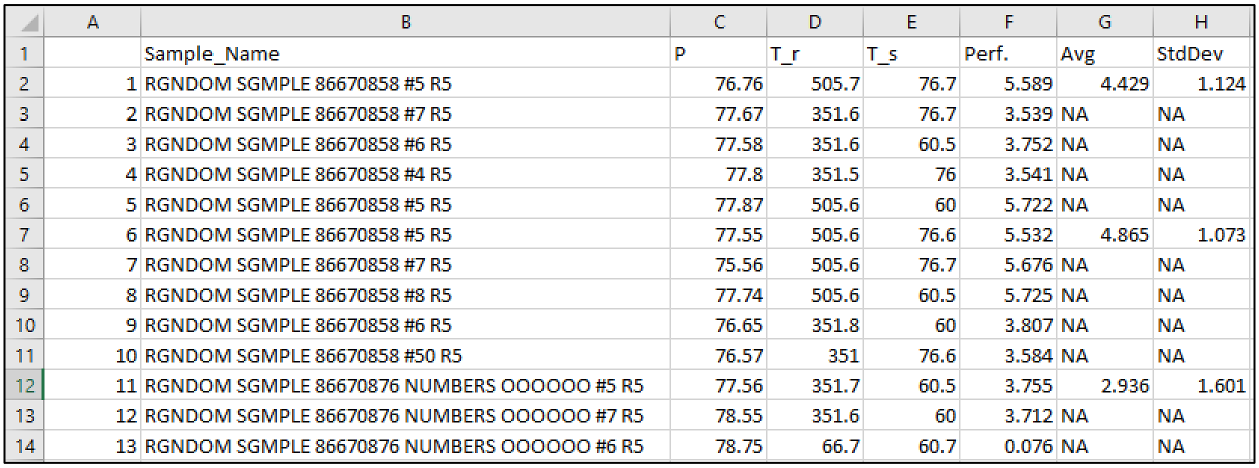
○ Right

Replicates

Download as .csv

Sample_Name	P	T_r	T_s	Perf.	Avg	StdDev
RGNDOM SGMPLE 86670858 #5 R5	76.76	505.70	76.70	5.59	4.43	1.12
RGNDOM SGMPLE 86670858 #7 R5	77.67	351.60	76.70	3.54	NA	NA
RGNDOM SGMPLE 86670858 #6 R5	77.58	351.60	60.50	3.75	NA	NA
RGNDOM SGMPLE 86670858 #4 R5	77.80	351.50	76.00	3.54	NA	NA
RGNDOM SGMPLE 86670858 #5 R5	77.87	505.60	60.00	5.72	NA	NA

Figure 3: Once file is uploaded, a preview of the parsed data is shown below the user interface. The preview adjusts itself as the number of replicates varies. When satisfied, the user is able to download a .csv of the clean data.



	A	B	C	D	E	F	G	H
1		Sample_Name	P	T_r	T_s	Perf.	Avg	StdDev
2	1	RGNDOM SGMPLE 86670858 #5 R5	76.76	505.7	76.7	5.589	4.429	1.124
3	2	RGNDOM SGMPLE 86670858 #7 R5	77.67	351.6	76.7	3.539	NA	NA
4	3	RGNDOM SGMPLE 86670858 #6 R5	77.58	351.6	60.5	3.752	NA	NA
5	4	RGNDOM SGMPLE 86670858 #4 R5	77.8	351.5	76	3.541	NA	NA
6	5	RGNDOM SGMPLE 86670858 #5 R5	77.87	505.6	60	5.722	NA	NA
7	6	RGNDOM SGMPLE 86670858 #5 R5	77.55	505.6	76.6	5.532	4.865	1.073
8	7	RGNDOM SGMPLE 86670858 #7 R5	75.56	505.6	76.7	5.676	NA	NA
9	8	RGNDOM SGMPLE 86670858 #8 R5	77.74	505.6	60.5	5.725	NA	NA
10	9	RGNDOM SGMPLE 86670858 #6 R5	76.65	351.8	60	3.807	NA	NA
11	10	RGNDOM SGMPLE 86670858 #50 R5	76.57	351	76.6	3.584	NA	NA
12	11	RGNDOM SGMPLE 86670876 NUMBERS OOOOOO #5 R5	77.56	351.7	60.5	3.755	2.936	1.601
13	12	RGNDOM SGMPLE 86670876 NUMBERS OOOOOO #7 R5	78.55	351.6	60	3.712	NA	NA
14	13	RGNDOM SGMPLE 86670876 NUMBERS OOOOOO #6 R5	78.75	66.7	60.7	0.076	NA	NA

Figure 4: Output .csv in Excel which allowed readable record-keeping. Data includes calculations of interest and basic descriptive statistics from which Excel graphs were easily generated.