Table of Contents

Abstract

I did an exploratory and explanatory data analysis of the experimental dataset about the ability to absorb energy by sandwich composite panels that have shear-thickening fluids upon ballistic impact.

1. Dataset

The data was obtained from research about a new type of sandwich composite panels made of 3D mats with channels between Kevlar layers, and filled with a high-performance fluid to improve energy dissipation [1]. The panels were tested by shooting them with a 9mm bullet, and it was found that the ones with the fluid could absorb 96.3% of the bullet's energy, which is 67.4% more energy absorbed and 61.26% more energy absorbed per unit weight than the panels without the fluid [1]. Figure 1 shows the data from ballistic tests on sandwich composite panels that have shear-thickening fluids. There are lots of different variables/features measured in the experiment, like the sample mass, areal weight, bullet mass, initial velocity, final velocity, and energy absorbed. The experiment was done 21 times in different conditions. The response variable (desired variable) is the absorbed energy of the composite which is calculated based on the difference between the energy of the bullet before and after the impact.

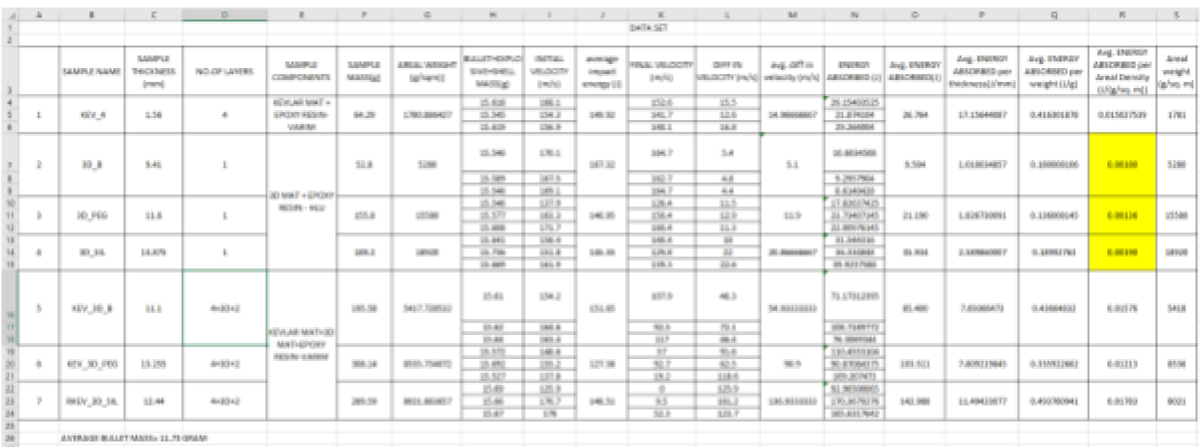

Figure 1. Dataset for ballistic tests on the manufactured sandwich composite panels incorporated with Shear Thickening Fluids

As we see in Figure 2, the variables' datatypes are *num* or *char*.

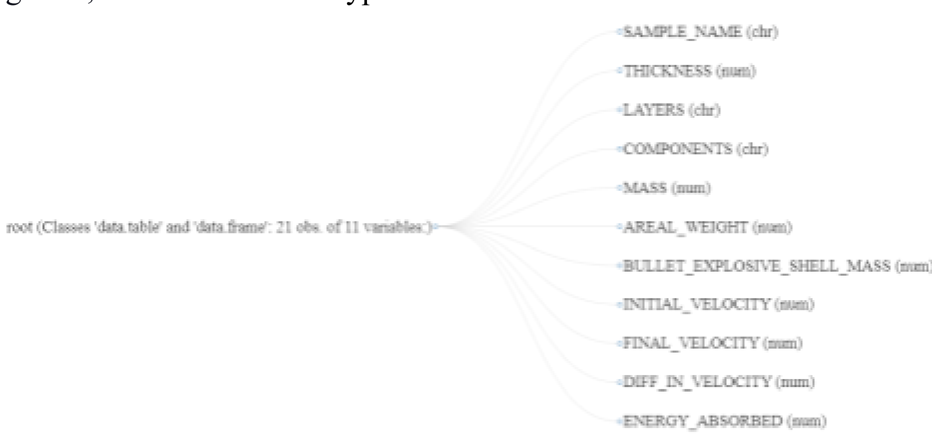

Figure 2. Features and datatypes

2. Cleaning data

As a result of the way datasets are stored, computers have difficulty reading them to do statistical analysis. The following are some of the problems:

- Merging cells
- Long names for variables
- Using color
- Two datatypes in one column like *NO.OF LAYERS*
- Extra and/or empty rows
- Manual index row
- Space between words of column's name

The first thing I did was to preprocess the messy data and eliminate the problems. The clean data is shown Figure 3.

	A	B	C	D	E	F	G	H	I	J	K
1	SAMPLE_NAME	THICKNESS	LAYERS	COMPONENTS	MASS	AREAL_WEIGHT	BULLET_EXPLOSIVE_SHELL_MASS	INITIAL_VELOCITY	FINAL_VELOCITY	DIFF_IN_VELOCITY	ENERGY_ABSORBED
2	KEV_4	1.56	L_4	C1	64.29	1780.886427	15.618	168.1	152.6	15.5	29.15403525
3	KEV_4	1.56	L_4	C1	65.29	1808.587258	15.545	154.3	141.7	12.6	21.874104
4	KEV_4	1.56	L_4	C1	66.29	1836.288089	15.619	156.9	140.1	16.8	29.264004
5	3D_B	9.41	L_1	C2	52.8	5280	15.546	170.1	164.7	5.4	10.6034508
6	3D_B	9.41	L_1	C2	52.8	5280	15.589	167.5	162.7	4.8	9.2957904
7	3D_B	9.41	L_1	C2	52.8	5280	15.546	169.1	164.7	4.4	8.6140428
8	3D_PEG	11.6	L_1	C2	155.8	15580	15.546	137.9	126.4	11.5	17.82637425
9	3D_PEG	11.6	L_1	C2	155.8	15580	15.577	163.3	150.4	12.9	23.73407145
10	3D_PEG	11.6	L_1	C2	155.8	15580	15.608	171.7	160.4	11.3	22.00976145
11	3D_SIL	13.875	L_1	C2	189.2	18920	15.641	158.4	140.4	18	31.544316
12	3D_SIL	13.875	L_1	C2	189.2	18920	15.736	151.8	129.8	22	36.334848
13	3D_SIL	13.875	L_1	C2	189.2	18920	15.669	161.9	139.3	22.6	39.9237588
14	KEV_3D_B	11.1	L_4+3D+2	C3	195.58	5417.728532	15.61	154.2	107.9	46.3	71.17312395
15	KEV_3D_B	11.1	L_4+3D+2	C3	195.58	5417.728532	15.62	164.6	92.5	72.1	108.7189772
16	KEV_3D_B	11.1	L_4+3D+2	C3	195.58	5417.728532	15.64	163.4	117	46.4	76.3069344
17	KEV_3D_PEG	13.255	L_4+3D+2	C3	308.14	8535.734072	15.572	148.6	57	91.6	110.4553104
18	KEV_3D_PEG	13.255	L_4+3D+2	C3	308.14	8535.734072	15.692	155.2	92.7	62.5	90.87084375
19	KEV_3D_PEG	13.255	L_4+3D+2	C3	308.14	8535.734072	15.527	137.8	19.2	118.6	109.207473
20	RKEV_3D_SIL	12.44	L_4+3D+2	C3	289.59	8021.883657	15.69	125.9	0	125.9	92.96500065
21	RKEV_3D_SIL	12.44	L_4+3D+2	C3	289.59	8021.883657	15.66	170.7	9.5	161.2	170.3679276
22	RKEV_3D_SIL	12.44	L_4+3D+2	C3	289.59	8021.883657	15.67	176	52.3	123.7	165.6317642

Figure 3. Preprocessed dataset

3. Percentages and Missing data diagrams

Percentages and missing data diagrams (Figure 4) show how many percentages of the dataset's variables are *discrete* or *continuous variable*. Also how many percentages of the dataset's rows are *complete* or *empty*. Fortunately, we don't have *NAN* or *empty* values in the dataset.

Figure 4. Prenectage diagram

4. Univariate Distribution

4.1. Histogram

Histogram plots (figure 5) show the data distribution for each variable. If we have enough data, the distribution of the variables is usually the *normal distribution*, in which case the algorithms work better.

4.2. QQ plot

In QQ plot, if the distribution of data is normal, the points should fall on the baseline, and the density of points in the middle of the line should be more than the edges. But for our dataset, it is not like that, which indicates that we need more data.

As we can see in Figure 5, for this dataset, the data distribution is not a *normal distribution*, which can complicate the data analysis, but it does not mean that nothing can be done.

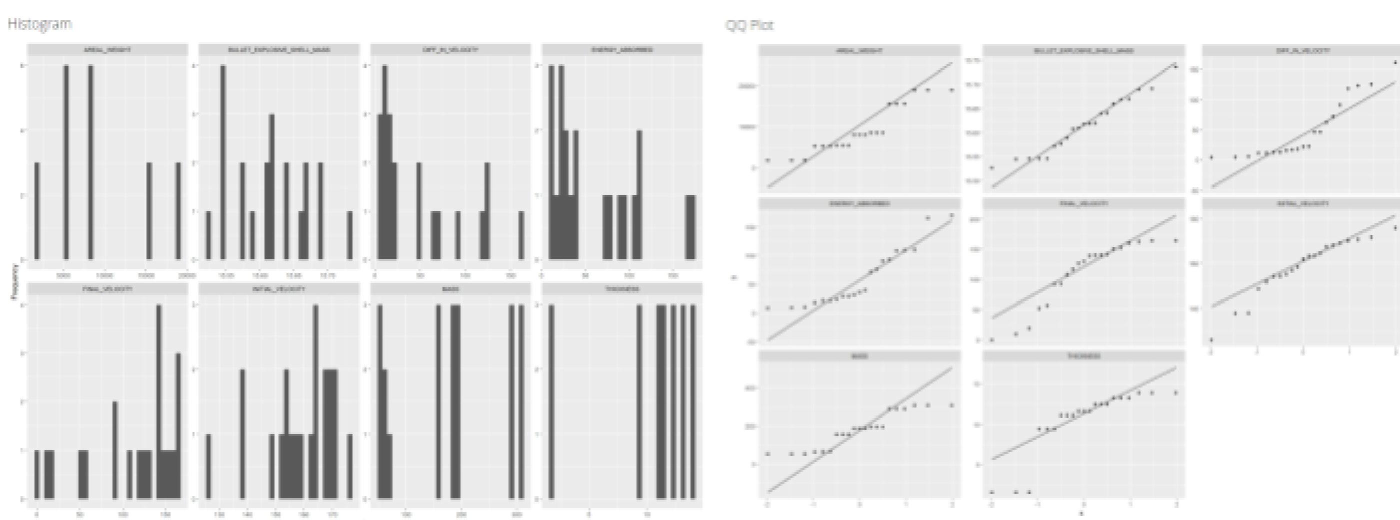

Figure 5. Histogram and QQ plots for variables' distribution.

4.3. Bar Chart (with frequency)

Bar Chart with frequency (figure 6) shows the frequency of components and samples also the thickness of layers.

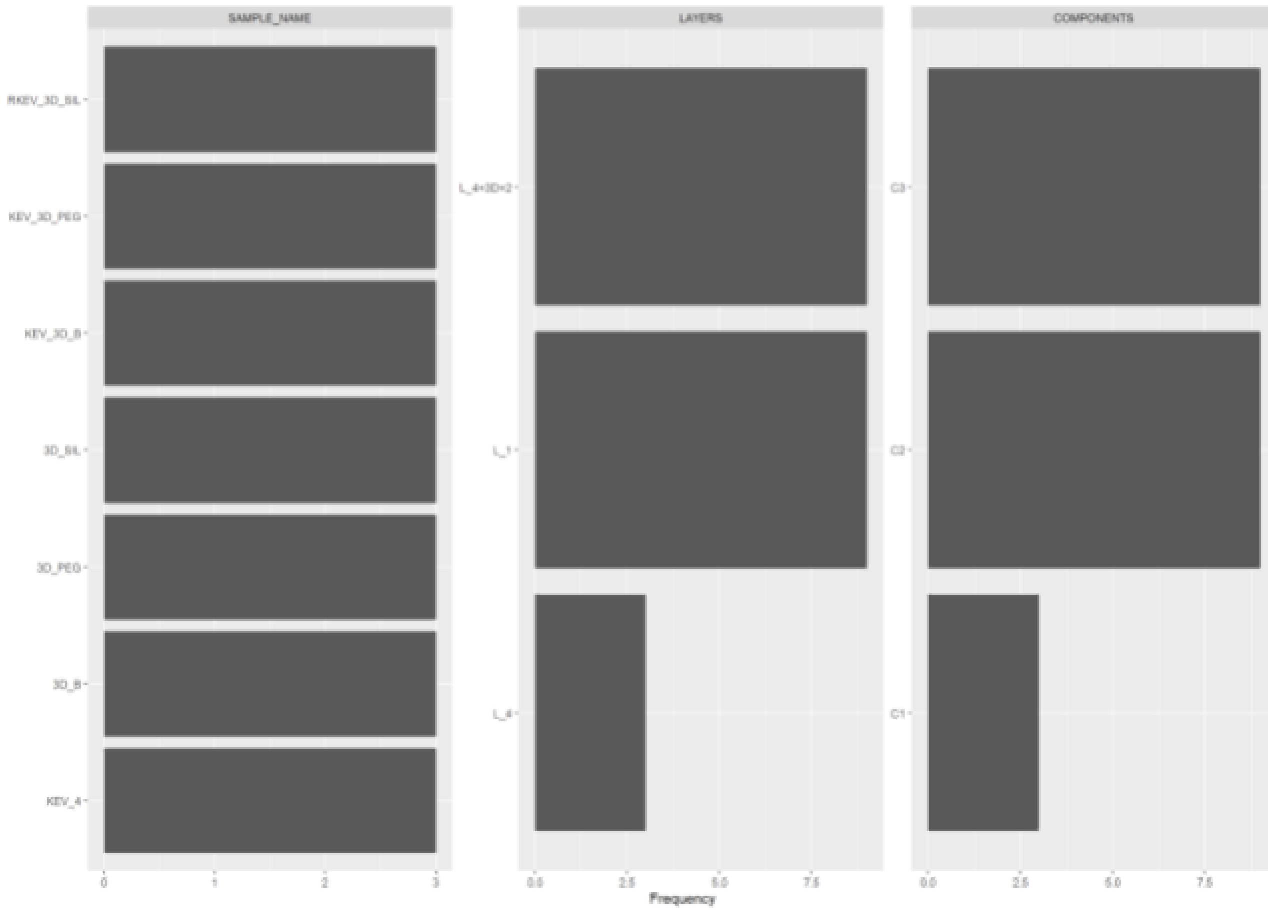

Figure 6. Bar Chart (with frequency)

5. Correlation Analysis

A correlation plot shows the relationship between two or more variables by displaying their correlation coefficient as a color-coded matrix. The desired variable in this research is absorbed energy. Using a correlation plot (figure 7) We can see the intensity and positivity/negativity influence of each variable on absorbed energy and compare the effect of each variable on absorbed energy. As we see the difference between the final and initial velocity of the bullet, sample mass, and sample thickness are the most important variable respectively for increasing the energy absorbed.

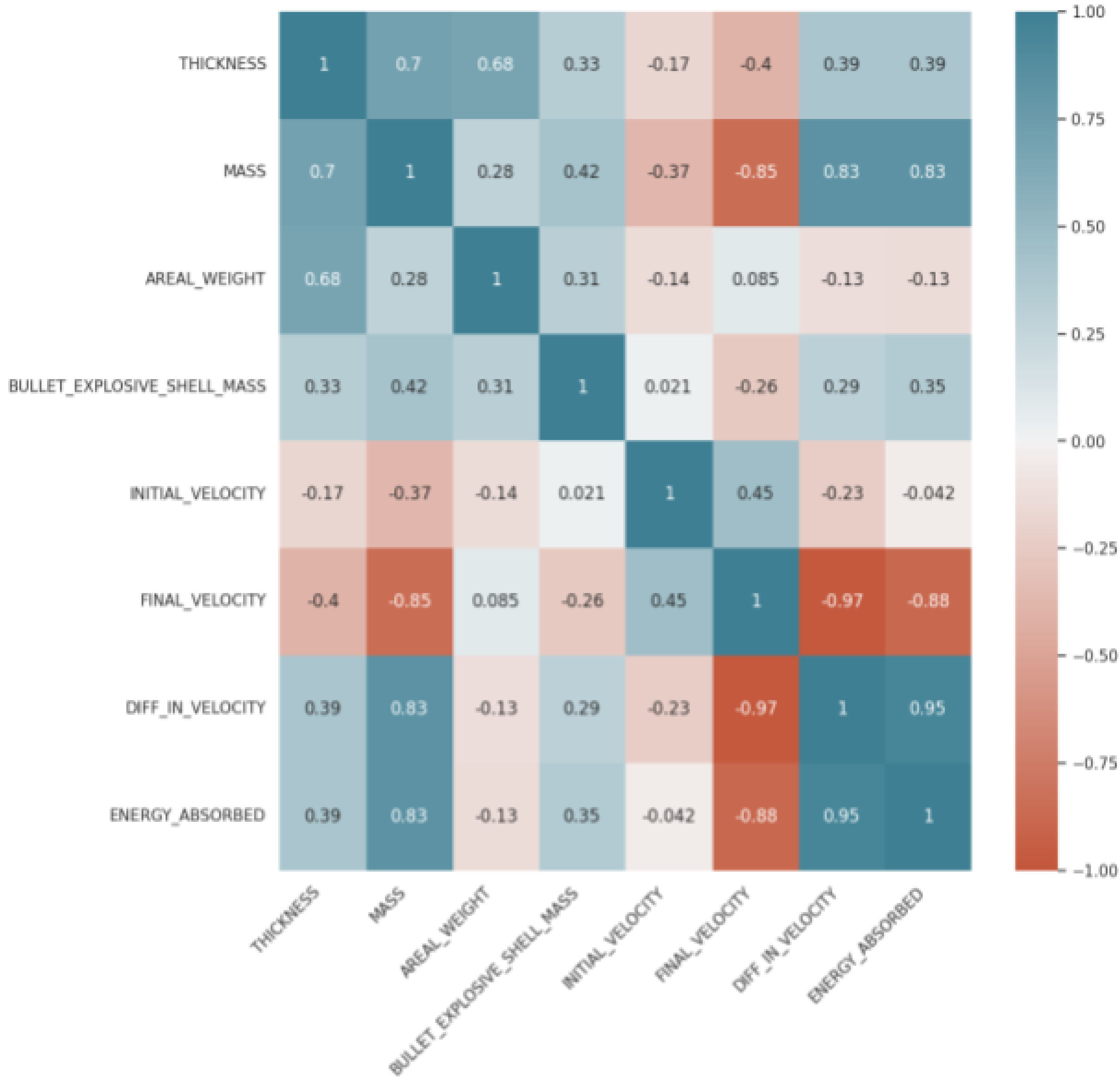

Figure 7. Correlation plot