Week 5 Homework

---

1. **Using matrix operations, perform linear regression with the least squares method (see link for more info), where:**

$$Y = \beta 0 + X.\,\beta 1 + \varepsilon$$

Step 1: Import the *Plots* and *LinearAlgebra* packages.

Change the plotting backend to plotly, i.e. plotly()

Step 2: Create a vector of random numbers called X.

Step 3: Create a vector of random numbers called Y, where

```
Y = X .+ rand(100)/3.
```

Step 4: Plot X vs Y as a scatter plot.

Step 5: Create a 100x2 element array called X1 by horizontally concatenating a vector of 100 ones with X. [Hint: hcat(), ones()]

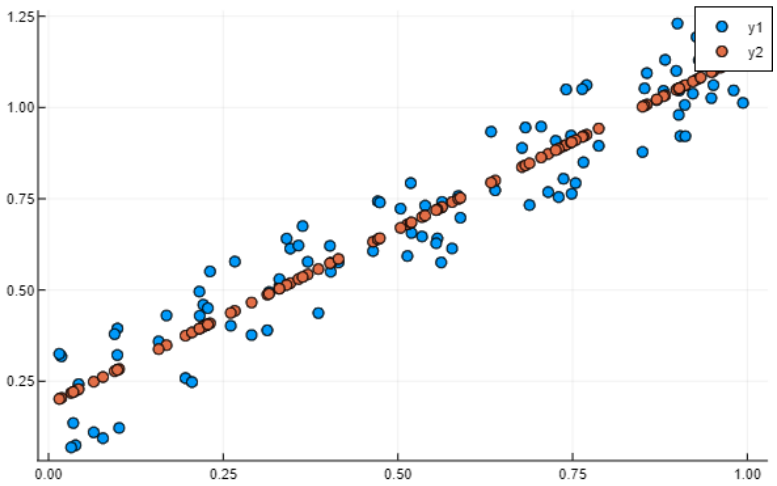
Step 6: Calculate the values of β0 and β1 such that:

$$\beta = \begin{bmatrix}\beta 0\\ \beta 1\end{bmatrix} = (X1'.X1)^{-1}.X1'.Y$$

Step 7: Overlay the previous plot with the linearly fitted data.

[Hint: `scatter!(X, β[1] .+ β[2].*X)`]

The final plot should look something like this:



2. Import data from the file *Homework_Data.txt*.
[Hint: DelimitedFiles > readdlm()]

Calculate the mean, median, standard deviation, variance, skewness, kurtosis, standard error of the mean, and quantiles of each column. [Hint: Statistics, StatsBase]

Plot the histogram of each column in a single plot, adjusting the transparency of the bar colours. [Hint: alpha]
Overlay the plot with the histogram of the entire dataset.

Round each element of the data matrix and convert values to integers.
Use the counts function to see how often the values in each column occur.

Create a set for each column of the integer data matrix in the previous step.
Find the union and intersection of all 3 sets.