Lab 3: More Simple Linear Regression

At this stage you would have developed an intuition behind the concept of regression analysis and I presume you would be eager for further development. Today's lab should provide sufficient content to encourage your fervour and anticipation for what lies ahead, as we will be expanding on the criteria that is used to evaluate the performance of a model.

Let's Begin!

1 Background

After your work with the Housing Agency of Jamaica (HAJ), interest among the business community is being piqued. Sun Island Jamaica has expanded their operations to include an online store that is accessible through their App. Before investing more resources into the app, they would like some insight into the users' time on the app and the amount of money which the user spends annually. They have reached out to you given your reputation of solving problems *simply*.

Below is the information Sun Island has provided you with on their data:

- **Email**: email address of user
- **Address**: address of user
- **Avatar**: user icon
- **Avg. Session Length**: Average session of in-store style advice sessions.
- **Time on App**: Average user time spent on App (min.)
- **Time on Website**: Average user time spent on website (min.)
- **Length of Membership**: number of years the customer has been a user of the App.

Procedure:

1. Upload and store the data in a suitable variable

   Note: Please ensure where the data has been saved, the directory has been set to the same location.

2. Produce some descriptive statistics on the aforementioned variables and evaluate the relationship which exists between them with, but not limited to, a scatterplot.
3. Build a suitable model that can predict for the required variable.
4. Create a scatterplot that is fitted with the model from above.
5. Comment on the significance of the coefficients, the least square estimates.

2 Regression Diagnostics

Here in, we will verify the quality of the model by observing statistics of the residuals as well as several plots by:

- Assessing the extent to which the residuals have satisfied the underlying assumptions. Once we are satisfied with the residual analysis, we can proceed to produce confidence intervals for the coefficients of the model as well as prediction intervals for observed values in our data.
- Using the *confint()* command, produce a confidence interval for coefficients of the model
- Produce a prediction interval when the average time on the app is 30min.