CON 2012
Assignment 3 (120 points)
Due: 11:59pm, April 3 (Sunday)

Write code in `R` to answer the following questions. Clearly label your answers using the comment function and turn in the `R` **script file** only. If there anything to discuss, please do so in your R script file.

The dataset `ecommerce_customers.csv` is having data of customers who buy clothes through phone application. Here, we would like to examine whether average time spent on phone app is associated with yearly amount spent. In your data, average time spent on phone app is coded as `Time.on.App` (x variable; measured in minutes), and yearly amount spent is `Yearly.Amount.Spent` (y variable; measured in US dollar).

1. Estimate a simple linear regression of y on x and report the R results. (**15 pts**)
2. Write down the estimated regression model. (**5 pts**)
3. Create a scatter plot of `Yearly.Amount.Spent` (y) vs. `Time.on.App` (x). Superimpose the estimated regression line on the scatter plot. (**10 pts**)
4. Interpret R^2 of the estimated regression model. (**10 pts**)
5. Conduct hypothesis testing for the association between `Time.on.App` and `Yearly.Amount.Spent`. State the null hypothesis and interpret the test results. (**20 pts**)
6. Calculate the expected annual spending for a customer who spends 10 minutes on phone application. (**10 pts**)
7. Calculate how much additional annual spending will be made by a customer who spends about 3 more minutes than others on the phone application. (**10 pts**)
8. Calculate residual sum of squares of the estimated regression model. (**20 pts**)
9. Demonstrate that an alternative model fit, $\hat{y} = 20 + 38 \cdot x$, has a higher residual sum of squares than the estimated regression model. (**20 pts**)