1. Consider the data set given in the table below:

x	5	8	9	7	14
y	3	1	6	7	19

   (a) Use the `lm` function in R to find the equation of the least-squares regression line $y = a + bx$

   (b) Plot the data and regression line together

   (c) Construct a density plot of the residuals. Comment on the overall fit of the model.

2. Consider the following data relating GPA to SAT score:

SAT	500	530	590	660	610	700	570	640
GPA	2.3	3.1	2.6	3.0	2.4	3.3	2.6	3.5

   (a) Assume a linear model $SAT = a + b \cdot GPA$, and compute maximum likelihood estimates for the parameters a and b.

   (b) Based on your model, what SAT score would you predict for someone with a GPA of 3.2?

   (c) Construct 95% confidence intervals for the parameters a and b.

   (d) Based on your 95% CIs, can you conclude that GPA is a significant predictor of SAT? Explain.

3. This exercise illustrates the importance of looking at your data before assuming that it is linear. Consider the following data:

x	1	2	3	4	5
y	0.74	2.22	6.04	16.20	44.55

   (a) Construct a scatter plot of the data. Does it look linear?

   (b) Use `lm` to fit a linear model $y = a + bx$.

   (c) Construct a new scatter plot of x versus $\log y$. Describe the shape. (Hint: just type `plot(x,log(y))` in the R console). What do you notice?

   (d) Use `lm` to fit a linear model for $\log(y) = a + bx$

   (e) Construct residual plots for both models. Based on these plots, which do you think is the better model? Explain.

4. This exercise illustrates a different method of linear model fit called "least absolute value" regression. Consider the following data:

x	1	2	3	4	5
y	5.25	10.12	15.40	18.55	202.12

   (a) Construct a scatter plot of the data. What do you notice?

   (b) Use `lm` to fit a linear model $y = a + bx$.

   (c) Instead of minimizing the squared residuals (as `lm` does), lets try minimizing the *absolute value* of the residuals. Modify the code on line 39 of `week7.R` (the code from the lecture) to compute the *absolute value* of the residuals instead of the square of the residuals.

   (d) Using `optim`, find the parameters a and b that minimize the absolute value of the residuals.

   (e) Using `abline`, add both regression lines to your scatter plot. Which is the better fit?

   (f) Based on this exercise, when do you think using "least absolute value" regression might be most useful?