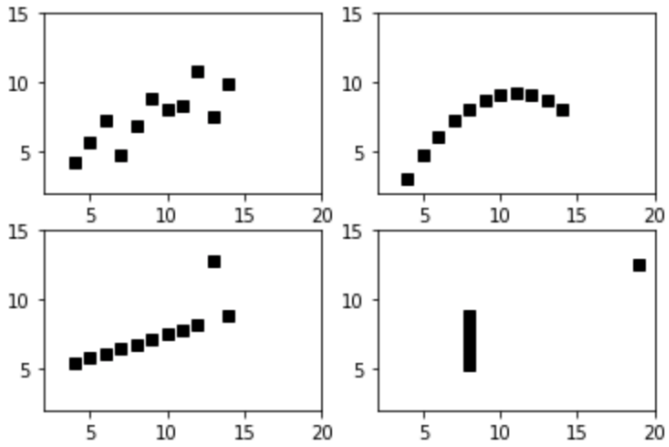
# LAB 3a: 'Anscombe Data'




Consider four datasets (x,y1), (x,y2), (x,y3) and (x4,y4), where

x = [10, 8, 13, 9, 11, 14, 6, 4, 12, 7, 5]
x4 = [8, 8, 8, 8, 8, 8, 8, 19, 8, 8, 8]
y1 = [8.04, 6.95, 7.58, 8.81, 8.33, 9.96, 7.24, 4.26, 10.84, 4.82, 5.68]
y2 = [9.14, 8.14, 8.74, 8.77, 9.26, 8.10, 6.13, 3.10, 9.13, 7.26, 4.74]
y3 = [[7.46, 6.77, 12.74, 7.11, 7.81, 8.84, 6.08, 5.39, 8.15, 6.42, 5.73]
y4 = [6.58, 5.76, 7.71, 8.84, 8.47, 7.04, 5.25, 12.50, 5.56, 7.91, 6.89]

1. Please load and plot the **Anscombe Data**.

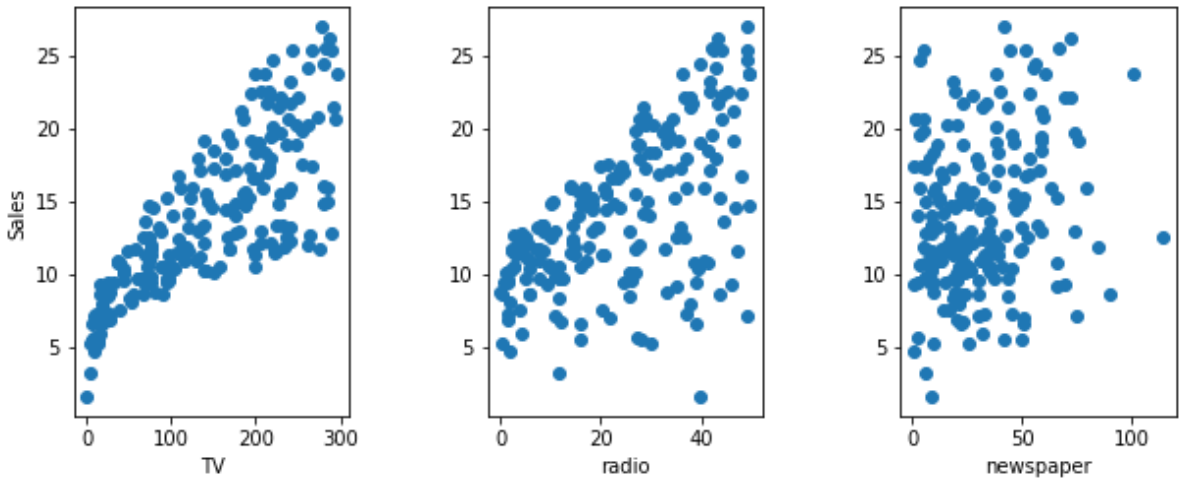
2. For each dataset compute the following basic statistics:
   - mean of y
   - std of y
   - correlation coefficient between x, y

3. For each dataset
   4. compute the regression coefficients
   5. plot the least-squares regression lines

6. What can you conclude looking at the basic statistics and regression lines?

# LAB 3b (Simple Linear Regression): 'Advertising Data'

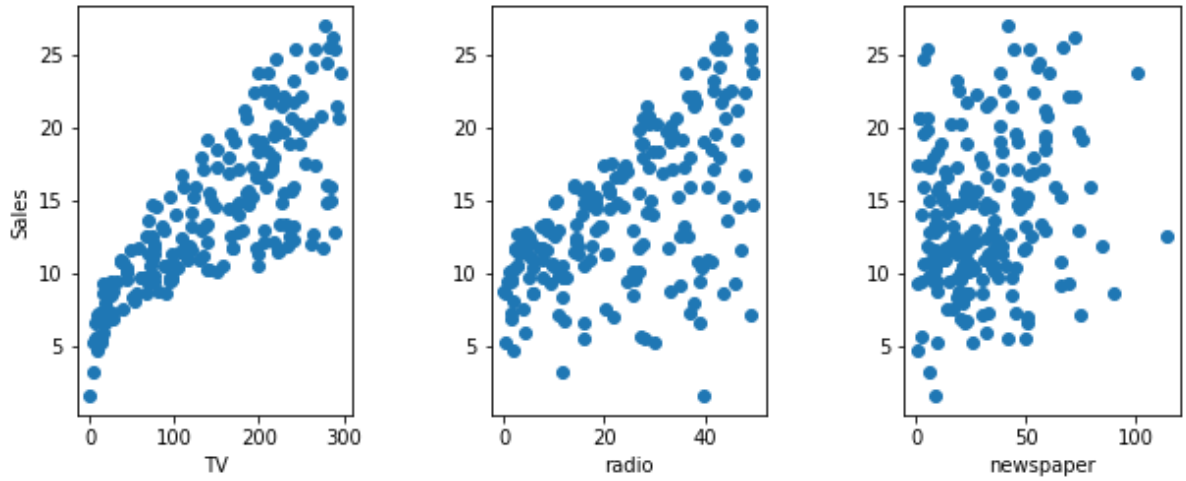
The Advertising data set (http://www-bcf.usc.edu/~gareth/ISL/Advertising.csv) consists of the **sales** of that product in 200 different markets, along with advertising budgets for the product in each of those markets for three different media: **TV**, **radio**, and **newspaper** .




Please load and plot the **Advertising Data** and answer the following questions:

1. Does the relationship between **TV budge** and **sales** seem linear?
2. How well does the least-square regression line fit the data? What is the proportion of the variability in sales that can be explained using TV budget?
3. How big is the coefficient relative to the standard error?
4. What is 95% CI for the slope?
5. Is there a relationship between **TV budget** and **sales**?

# LAB 3c (Multiple Linear Regression): 'Advertising Data'




Please load and plot the **Advertising Data** and answer the following questions:

1. Is there any relationship between **advertising budget** and **sales**? (Hint: F statistics)

2. Which media contribute to sales most? How 'reliable' is this conclusion? (Hint: regression coefficients, CI)

3. How much what we see in the data can we explain using linear regression model? (Hint: RSE, R^2, adj R^2)