MATH 1342

Section 5.1

Bivariate Data

Bivariate data is data for two different variables (usually related in some way).

Variables are classified as response variables and explanatory variables.

 A **response variable** (dependent) measures the outcome of a study. An **explanatory variable** (independent) attempts to explain the observed outcomes. Algebraically speaking, your explanatory variable is your "*x*" and the response variable is your "*y*".

Scatterplots

Once the explanatory and response variables are identified, we can display the association between the two using a **scatterplot**.

Example:

Suppose we want to know if there is an association between the number of spaces a property is from GO and the cost of the property in a monopoly game. The data is below:
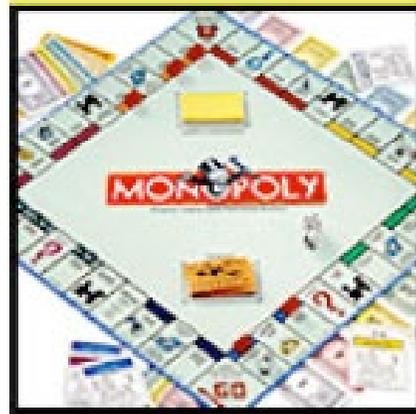
Property	Spaces from GO	Cost
Mediterranean Avenue	1	60
Baltic Avenue	3	60
Reading Railroad	5	200
Oriental Avenue	6	100
Vermont Avenue	8	100
Connecticut Avenue	9	120
St. Charles Place	11	140
Electric Company	12	150
States Avenue	13	140
Virginia Avenue	14	160
Penn Railroad	15	200
St. James Place	16	180
Tennessee Avenue	18	180
New York Avenue	19	200
Kentucky Avenue	21	220
Indiana Avenue	23	220
Illinois Avenue	24	240
B & O Railroad	25	200
Atlantic Avenue	26	260
Ventnor Avenue	27	260
Water Works	28	150
Marvin Gardens	29	280
Pacific Avenue	31	300
North Carolina Avenue	32	300
Pennsylvania Avenue	34	320
Short Line Railroad	35	200
Park Place	37	350
Boardwalk	39	400

Assigning Variables

First we must decide which variable is explanatory and which is response.

Which variable "causes" the change in the other variable?

The one that causes the change is your x (explanatory). The one is changed because of the other one is your y (response).

Creating a scatterplot in R-Studio

Assign each set of data a variable name:

Choose variable names. I will use spaces and cost.

Enter the lists in R:

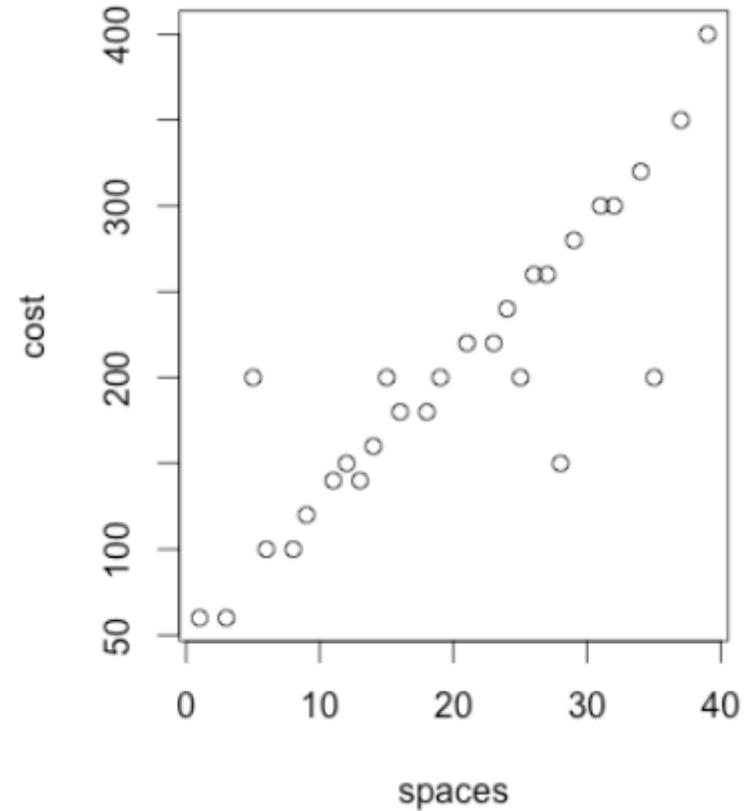assign("spaces",c(1,3,5,6,8,9,11,12,13,14,15,16,18,19,21,23,24,25,26,27,28,29,31,32,34,35,37,39))

assign("cost",c(60,60,200,100,100,120,140,150,140,160,200,180,180,200,220,220,240,200,260,260,150,280,300,300,320,200,350,400))

Now use the plot command:

plot(spaces, cost)

• Note plot command is plot(explanatory,response)

Creating a Scatterplot in R-Studio

Trends in a scatterplot:

To interpret a scatter plot we will look at the direction, form and strength.

Positively related:

Negatively related:

Linear relationship:

Let's discuss the Monopoly problem above.

What is the relationship of the data?

What is the strength of this relationship?

Is the relationship linear?