Correlation

The questions below are taken from a textbook produced by Siyavula, which can be accessed [here](https://intl.siyavula.com/read/maths/grade-12/statistics/09-statistics-03).

1. For each of the following data sets, create a scattergraph, calculate the correlation coefficient and comment on the nature of the correlation, e.g., “strong positive”.

Note, the following command will create a list of data:

xa<-c(5,8,13,10,14,15,17,12,18,13)

	1. 

|  |  |
| --- | --- |
| Scattergraph |  |
| Correlation coefficient |  |
| Correlation |  |

* 1. 

|  |  |
| --- | --- |
| Scattergraph |  |
| Correlation coefficient |  |
| Correlation |  |

* 1. 

|  |  |
| --- | --- |
| Scattergraph |  |
| Correlation coefficient |  |
| Correlation |  |

* 1. 

|  |  |
| --- | --- |
| Scattergraph |  |
| Correlation coefficient |  |
| Correlation |  |

* 1. 

|  |  |
| --- | --- |
| Scattergraph |  |
| Correlation coefficient |  |
| Correlation |  |

1. A geography teacher gave the data set below to his class to illustrate the concept that average temperature depends on how far a place is from the equator (known as the latitude). There are 90 degrees between the equator and the North Pole. The equator is defined as 0 degrees.



	1. Create a scatter graph of this data, including a title and axis labels.

|  |
| --- |
|  |

* 1. Determine the value of the correlation coefficient.

|  |
| --- |
|  |

* 1. What can you deduce about the relationship between how far north a city is and its average temperature?

|  |
| --- |
|  |

1. A recording company investigates the relationship between the number of times a CD is played by a national radio station and the national sales of the same CD in the following week. The data below was collected for a random sample of 10 CDs. The sales figures are rounded to the nearest 50.



	1. Create a scatter graph of this data, including a title and axis labels.

|  |
| --- |
|  |

* 1. Determine the value of the correlation coefficient.

|  |
| --- |
|  |

* 1. What can you deduce about the relationship between how many times a CD is played on the radio and the weekly sales of the CD?

|  |
| --- |
|  |

**Useful R commands**

Entering data to R Studio To read in data from an Excel csv file called *excel\_data.csv* to R Studio and name it *mydata*, first use the drop down menus in R Studio Session > Set Working Directory > Choose Directory to indicate the location of *excel\_data.csv* on your computer. The following code will then read the data in to R Studio:

mydata<-read.csv("excel\_data.csv")

attach(mydata) — this adds the variable names

At the end of the analysis remember to use detach(mydata) to disassociate the variable names.

**(a) Graphics**

plot(X,Y,xlab="x-axis label",ylab="y-axis label", main="Scatterplot of Y on X",pch=21,bg="black") — produces a scatterplot of X vs Y with black dots of the size specified by ‘pch’

**(c) Correlation and Regression**

cor.test(X,Y) — computes the correlation between X and Y and performs a test of the null hypothesis of zero correlation