SM-1402: Exercise 1 (Presentation of Data)

1. The following data represents the heights of 20 students in centimetres.

162	177	189	173	151	178	165	179	168	158
168	147	171	168	153	184	173	166	165	158

   (a) What is the minimum and maximum value of the data?

   (b) Tabulate the data frequency by dividing it into 5 classes (i.e. groups) between 140 cm and 190 cm using equal interval widths of 10 cm.

   (c) Draw a histogram for the data.

   (d) Hypothetically, if a new student's height was measured to be 150 cm exactly, which category would this data point be counted in?

2. Consider the annual sales (millions of dollars) of a company by region, as tabulated below.

Region	Africa	America	Asia	Europe
Sales	5.5	6.7	13.2	19.6

   (a) What is the total annual sales of this company?

   (b) Complete the following table

Region	Sales	Percentage	Angle
Africa	5.5		
America	6.7		
Asia	13.2		
Europe	19.6		
Total			

   where the Percentage column represents percentage of total sales, and the angle (in degrees) is calculated using the formula

$$360^\circ \times \text{Percentage}$$

   (c) Draw a bar graph for the amount of sales by region of the company.

   (d) Draw a pie chart for the amount of sales by region of the company.

3. The data file `gsseduc_2021.csv` contains data for 1,495 respondents in the 2021 round of the U.S. General Social Survey (GSS). The GSS is a survey of the characteristics and attitudes of the general U.S. population. We will look at a selection of the data set relating to education levels.

   - `educ`: the number of years of education that the survey respondent has completed
   - `speduc`: years of education completed by the respondent's spouse

Here it is of interest to examine whether there seems to be any association between the two, i.e. whether, for example, people with high levels of education tend to have spouses who also have high levels of education. This would be evidence of *homogamy*, i.e. people tending to marry people who have similar characteristics to themselves.

(a) Using MS Excel, import the data and produce an appropriate graph to answer the research question. What can you observe from the plot?

(b) Suppose that you had access to the GSS data for the years 1972–2021, with the interest of observing the mean `educ` of respondents over time. What kind of graph would you plot?