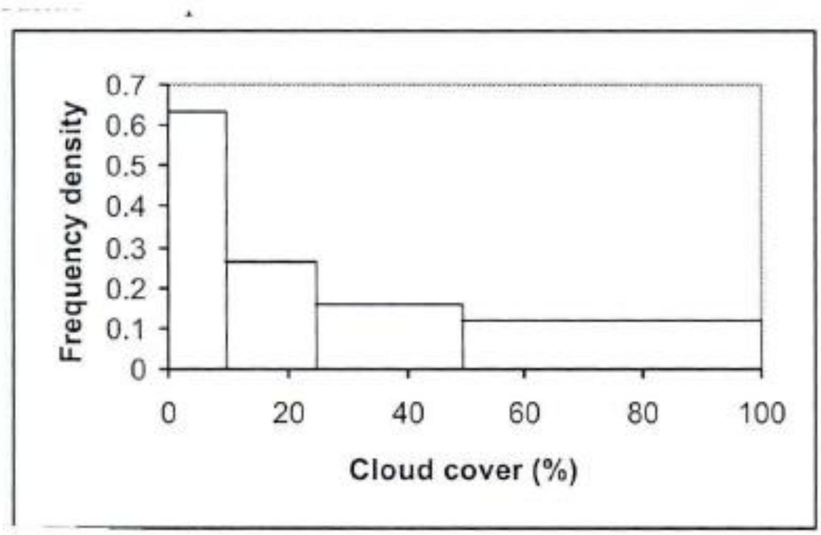
Probability and Statistics 1 - Surgery Hours class (Andres Villegas)
Exercise Sheet 1 Solutions: Exploratory Data Analysis

1. a) Use histogram. The fact that each bar has height equal to frequency divided by interval width is suitable compensation for the unequal interval widths.



b) The student has tried to equalise the widths of the intervals by combining two short intervals and splitting one long one. Combining two intervals is permissible (though seldom useful), but splitting an interval and arbitrarily assigning half the observations to each sub-interval is not allowed. The fact that the intervals now have equal width means that the use of frequency rather than frequency density for the vertical axis is not too serious an error.

2. a)

a.m.		p.m.
3 2 1 0	0	0 0 0 0 0 1 2 4
7 7 6 5	0	6 7
4	1	
	1	
3	2	

b) **For morning**: median=5.5 mins, LQ=2 mins, UQ=7 mins; IQR=5 mins. Therefore anything above 22 mins is a definite outlier (one observation: 23 mins); anything else above 14.5 mins a possible outlier (none).

For afternoon: median=0.5 mins, LQ=0 mins, UQ=4 mins; IQR=4 mins, So the cut-off point for possible outliers is 10: there are none.

c)
Location: the lateness is much greater in the morning; all quartiles and max value are higher.
Spread: The IQRs are about the same. Because of the outlier the morning observations have a much wider range.
Skewness: both data sets are positively (right) skewed (mean greater than median).

3. a)

i)

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}(a + bx_i) = \frac{1}{n}\sum_{i=1}^{n} a + \frac{1}{n}\sum_{i=1}^{n} bx_i = \frac{na}{n} + \frac{b}{n}\sum_{i=1}^{n} x_i = a + b\bar{x}$$

ii)

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(a + bx_i - a - b\bar{x})^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n} b^2(x_i - \bar{x})^2 = \frac{b^2}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = b^2 s_y^2$$

b) Denote by x_i the temperatures in °C and by y_i the temperatures in °F

i)

$$\bar{x} = \frac{280.2}{12} = 23.4$$

$$s_x^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) = \frac{1}{11}(6583.22 - 12(23.4)^2) = 3.69$$

$$s_x = 1.92$$

ii) Since $y_i = \frac{9}{5}x_i + 32$, then $\bar{y} = \frac{9}{5}\bar{x} + 32 = \frac{9}{5}23.4 + 32 = 74.0$, and $s_y = \frac{9}{5}s_x = \frac{9}{5}1.92 = 3.46$