Answers at the end

|	

1	This question refers to the assigned R preparation.
	
	In a previous class, we have worked the data frame `country_info`. Here is the top and bottom.
	
	``
	head(country_info)
	  country_name continent population access_to_electricity_percent
	1 Burundi Africa 11890784 7.59
	2 Comoros Africa 869601 77.80
	3 Djibouti Africa 988000 51.80
	4 Eritrea Africa 3546421 46.70
	5 Ethiopia Africa 114963588 42.90
	6 Kenya Africa 53771296 56.00
	
	tail(country_info)
	                 country_name continent population access_to_electricity_percent
	188 Luxembourg Europe 625978 100
	189 Monaco Europe 39242 100
	190 Netherlands Europe 17134872 100
	191 Switzerland Europe 8654622 100
	192 Canada Americas 37742154 100
	193 United States of America Americas 331002651 100
	``
	
	How can we calculate the total population on each continent?
	
	``
	  continent population
	1 Africa 1337666440
	2 Americas 1018121141
	3 Asia 4609091684
	4 Europe 747293775
	5 Oceania 41798096
	``

(a)	`aggregate(population ~ continent, data = country_info, sum)`
(b)	`aggregate(continent ~ population, data = country_info, sum)`
(c)	`aggregate(cbind(population, continent), data = country_info, sum)`
(d)	`aggregate(cbind(continent, population), data = country_info, sum)`

A kite enthusiast took hourly readings of wind speed for an entire year. They then decided to calculate the daily average wind speed and compare the daily averages during summer and winter. To do so, they plotted two histograms as seen below:
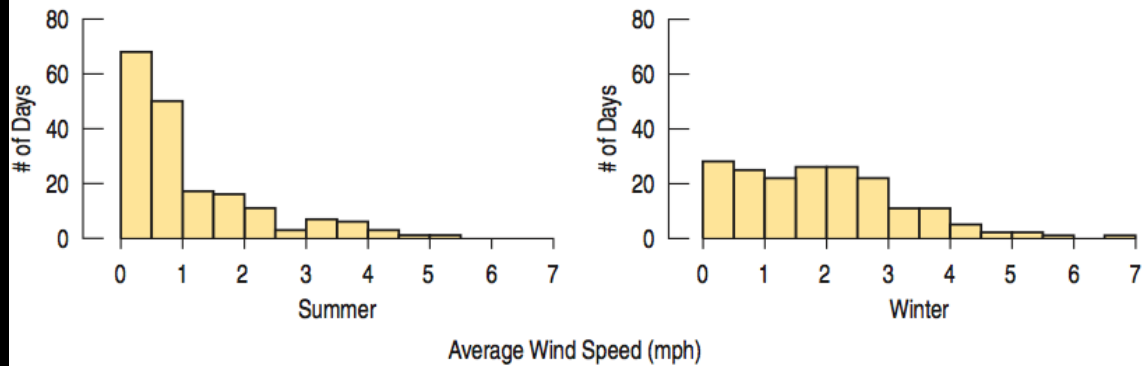

Based upon these two histograms, what can they conclude?

(a)	Wind speed in summer has a larger interquartile range than in winter.
(b)	Wind speed in winter has a larger interquartile range than in summer.
(c)	Wind speed in summer and winter is equally variable.
(d)	More information is required to make such a determination.

Three Statistics classes all took the same test. Histograms and boxplots of the scores for each class are shown below. Match each class with the corresponding boxplot.
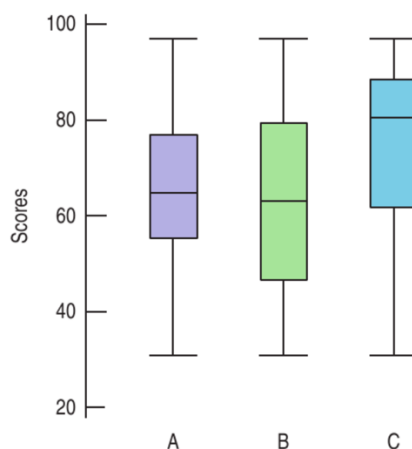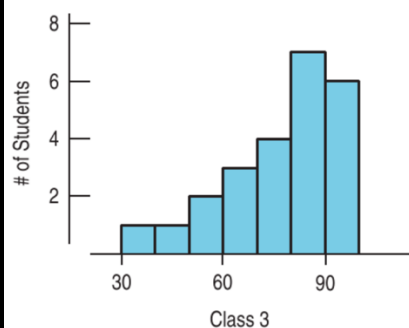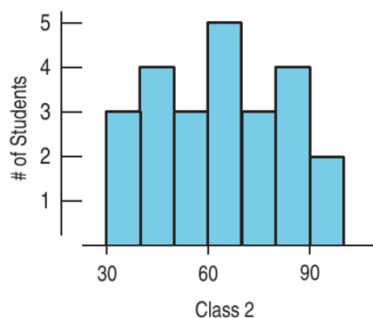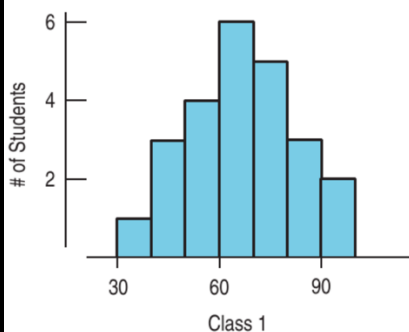
(a)	A2, B1, C3
(b)	A1, B2, C3
(c)	A3, B2, C1
(d)	A2, B3, C1

A class of fourth graders takes a diagnostic reading test, and the scores are reported by reading grade level. The 5-number summaries for the 14 boys and 11 girls are shown below.

| **Boys** | 2.5 | 3.8 | 4.5 | 5.3 | 5.9 |
| **Girls** | 2.8 | 3.8 | 4.6 | 5.4 | 5.7 |

Which statement below is correct?

(a)	The highest score was achieved by a girl.
(b)	The boys' scores have the greater range.
(c)	The boys' scores have the greater interquartile range.
(d)	The boy with the lowest score was an outlier.

Below is a "back-to-back" stem-and-leaf display that shows two data sets at once – one going to the left, one to the right. The display compares the percent change in population for two regions of the United States (based on census figures for 2000 and 2010). The fastest growing state was Nevada at 35%. To show the distributions better, this display breaks each stem into two lines, putting leaves 0–4 on one stem and leaves 5–9 on the other. "2 | 1 | 0" means 12% for an NE/MW state and 10% for an S/W state.

```
NE/MW States |      | S/W States
           1 | -0 |
  4433333220 |  0 | 14
  98777776655 |  0 | 58899
          32 |  1 | 00223344
           5 |  1 | 57889
             |  2 | 114
             |  2 | 5
             |  3 |
             |  3 | 5
```

Which pair of boxplots corresponds to this stem-and-leaf display?

(a)   (b)   (c)   (d)
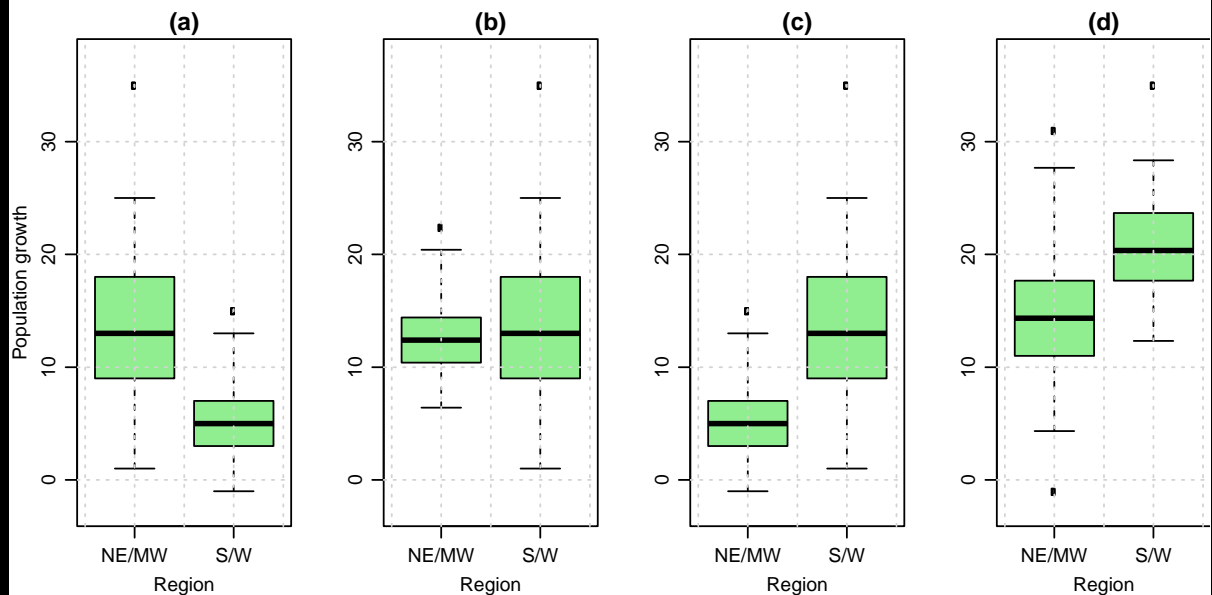
Answers:

1a. The first argument must tell R to split the `population` column by `continent`. So there must be a tilde ~ ("as a function of") in the first argument (ruling out answer options c and d): we must treat the `population` as a function of the `continent`. Viewing the `continent` as a function of the `population` doesn't make sense so that answer option b must be wrong.

2b. See table at the bottom of page 101 in the textbook.

3b. One solution strategy is to compare the skewness and the IQR in histograms and boxplots.

4b.

5c.