After the change of firewall settings, the numbers of blocked intrusions during the next 20 days were

$$53, 21, 32, 49, 45, 38, 44, 33, 32, 43, 53, 46, 36, 48, 39, 35, 37, 36, 39, 45.$$

Comparing the number of blocked intrusions before and after the change,

(a) construct side-by-side stem-and-leaf plots;

(b) compute the five-point summaries and construct parallel boxplots;

(c) comment on your findings.

8.2. A network provider investigates the load of its network. The number of concurrent users is recorded at fifty locations (thousands of people),

17.2	22.1	18.5	17.2	18.6	14.8	21.7	15.8	16.3	22.8
24.1	13.3	16.2	17.5	19.0	23.9	14.8	22.2	21.7	20.7
13.5	15.8	13.1	16.1	21.9	23.9	19.3	12.0	19.9	19.4
15.4	16.7	19.5	16.2	16.9	17.1	20.2	13.4	19.8	17.7
19.7	18.7	17.6	15.9	15.2	17.1	15.0	18.8	21.6	11.9

(a) Compute the sample mean, variance, and standard deviation of the number of concurrent users.

(b) Estimate the standard error of the sample mean.

(c) Compute the five-point summary and construct a boxplot.

(d) Compute the interquartile range. Are there any outliers?

(e) It is reported that the number of concurrent users follows approximately Normal distribution. Does the histogram support this claim?

8.3. Verify the use of Chebyshev's inequality in (8.6) of Example 8.16. Show that if the population mean is indeed 48.2333 and the population standard deviation is indeed 26.5170, then at least 8/9 of all tasks require less than 127.78 seconds of CPU time.

8.4. Use Table A4 to compute the probability for *any* Normal random variable to take a value within 1.5 interquartile ranges from population quartiles.

8.5. The following data set shows population of the United States (in million) since 1790,

Year	1790	1800	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900
Population	3.9	5.3	7.2	9.6	12.9	17.1	23.2	31.4	38.6	50.2	63.0	76.2
Year	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010	
Population	92.2	106.0	123.2	132.2	151.3	179.3	203.3	226.5	248.7	281.4	308.7	

Construct a time plot for the U.S. population. What kind of trend do you see? What information can be extracted from this plot?

8.6. Refer to Exercise 8.5. Compute 10-year *increments* of the population growth $x_1 = 5.3 - 3.9$, $x_2 = 7.2 - 5.3$, etc.

(a) Compute sample mean, median, and variance of 10-year increments. Discuss how the U.S. population changes during a decade.

(b) Construct a time plot of 10-year increments and discuss the observed pattern.

8.7. Refer to Exercise 8.5. Compute 10-year *relative population change* $y_1 = (5.3 - 3.9)/3.9$, $y_2 = (7.2 - 5.3)/5.3$, etc.

(a) Compute sample mean, median, and variance of the relative population change.

(b) Construct a time plot of the relative population change. What trend do you see now?

(c) Comparing the time plots in Exercises 8.6 and 8.7, what kind of correlation between x_i and y_i would you expect? Verify by computing the sample correlation coefficient

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})/(n-1)}{s_x s_y}.$$

What can you conclude? How would you explain this phenomenon?

8.8. Consider three data sets.

(1) 19, 24, 12, 19, 18, 24, 8, 5, 9, 20, 13, 11, 1, 12, 11, 10, 22, 21, 7, 16, 15, 15, 26, 16, 1, 13, 21, 21, 20, 19

(2) 17, 24, 21, 22, 26, 22, 19, 21, 23, 11, 19, 14, 23, 25, 26, 15, 17, 26, 21, 18, 19, 21, 24, 18, 16, 20, 21, 20, 23, 33

(3) 56, 52, 13, 34, 33, 18, 44, 41, 48, 75, 24, 19, 35, 27, 46, 62, 71, 24, 66, 94, 40, 18, 15, 39, 53, 23, 41, 78, 15, 35

(a) For each data set, draw a histogram and determine whether the distribution is right-skewed, left-skewed, or symmetric.

(b) Compute sample means and sample medians. Do they support your findings about skewness and symmetry? How?

8.9. The following data set represents the number of new computer accounts registered during ten consecutive days.

$$43, 37, 50, 51, 58, 105, 52, 45, 45, 10.$$

(a) Compute the mean, median, quartiles, and standard deviation.

(b) Check for outliers using the 1.5(IQR) rule.

(c) Delete the detected outliers and compute the mean, median, quartiles, and standard deviation again.

(d) Make a conclusion about the effect of outliers on basic descriptive statistics.