

Chapter 4 - Understanding and Comparing Distributions

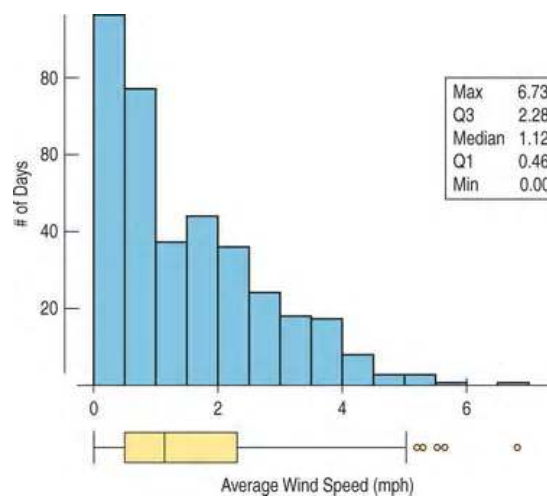
September 5, 2014

In Chapter 3 we learned about graphical representations and numerical summaries for Quantitative Variables. We also learned to describe the distribution of a data set in terms of the Shape, Center, and Spread.

If the distribution was symmetric we used the Mean for the Center and the Standard Deviation for the Spread

If the distribution was skewed or had outliers we used the Median for the Center and the IQR for the Spread.

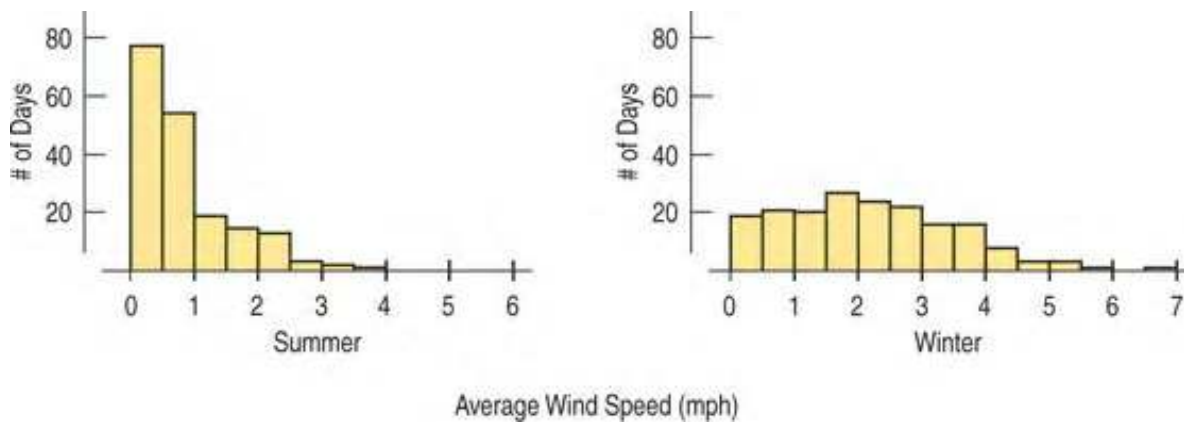
Histogram and Boxplot for daily wind speeds



How does each represent distribution?

4.1 Comparing Groups with Histograms

Axis should be similar. Note the shapes, centers, and spreads of the distributions.



Back to Back Stem and Leaf Plot - Stem plot used to compare two different data sets by putting one set of leaves to the left of the stems and the other set to the right of the stems.

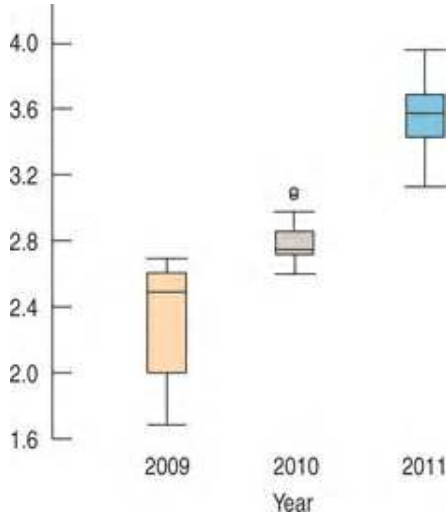
South and West		North and Midwest
5778	8	8
12344	9	03
6667778899	9	67
02334	10	012233334
56	10	6779
	11	122444

4.2 Comparing Groups with Boxplots

Boxplots are useful when comparing different groups of data. Use same axis and plot boxplots for different groups side by side or above each other.

Examples, p. 104:

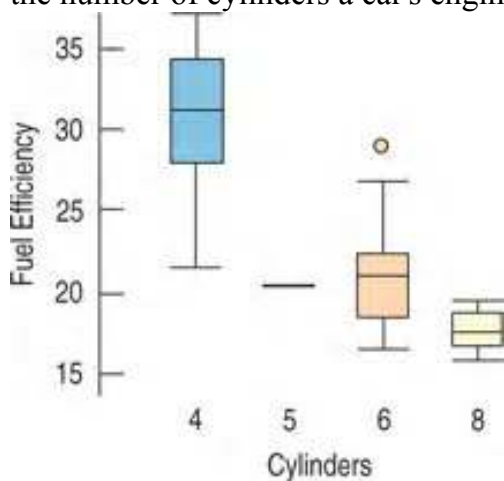
26. Gas prices 2011 Here are boxplots of weekly gas prices for regular gas in the United States as reported by the U.S. Energy Information Administration for 2009, 2010, and 2011.



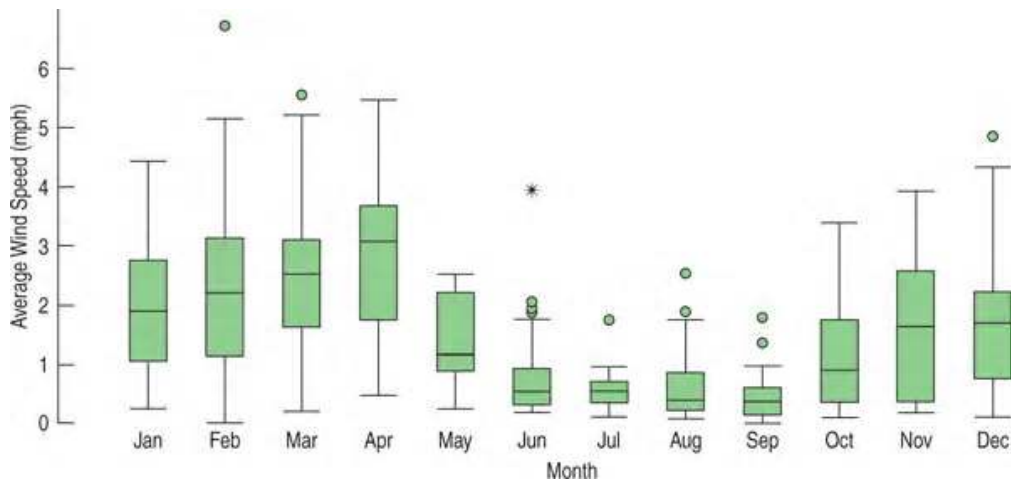
a) Compare the distribution of prices over the three years.

b) In which year were the prices least stable? Explain.

28. Fuel economy Describe what these boxplots tell you about the relationship between the number of cylinders a car's engine has and the car's fuel economy (mpg).



4.3 Outliers



Is an outlier an error? May be able to correct.

How extreme is an outlier? Outlier may not be as extreme in another context.

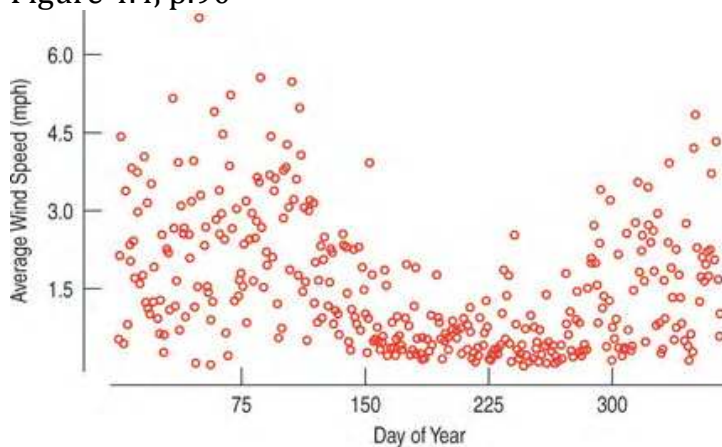
Outliers should not be ignored and treated as normal.

If an outlier is omitted from the data it should be mentioned and justified.

4.4 Timeplots

Timeplots are graphs that plot data values versus time. Used in Stock Market values, Stock prices, Unemployment Rates, Temperature, Global Temperatures, etc.

Figure 4.4, p.90



One should resist using timeplots to predict for the future unless there is strong reasons for doing otherwise. (Path a ball follows when thrown from a certain height at a given speed, Seasonal Temperatures, Interest Rates?)