

Chapter 5 - LSRL Cautions

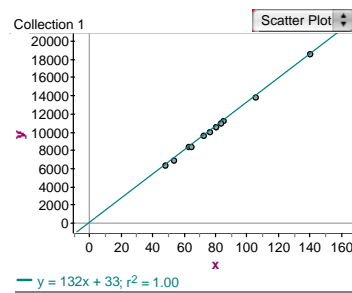
- Correlation and Regression are NOT RESISTANT to outliers and Influential Points!
- Correlations based on “averaged data” tend to be higher than correlations based on all raw data.
- Extrapolating beyond the observed data can result in predictions that are unreliable.

Correlation vs. Causation

- Consider the following historical data – Ministers vs. Imported Rum

- x = # Methodist Ministers in New England
- y = # of Barrels of Rum Imported to Boston

	Year	Ministers	Rum
1	1860	63	8376
2	1865	48	6406
3	1870	53	7005
4	1875	64	8486
5	1880	72	9595
6	1885	80	10643
7	1890	85	11265
8	1895	76	10071
9	1900	80	10547
10	1905	83	11008
11	1910	105	13885
12	1915	140	18559



- $r=0.999997$
- There is an almost perfect linear relationship between Ministers (x) and Imported Rum (y)

■ CORRELATION DOES NOT IMPLY CAUSATION!