STA 674

Regression Analysis And Design Of Experiments

Measuring Association between Two Variables – Lecture 4

STA 674, RADOE:
Measuring Association between Two Variables

- Interpretation of Correlation
  - Last time – sign
  - This time – magnitude

STA 674: Measuring Association between Two Variables Correlation

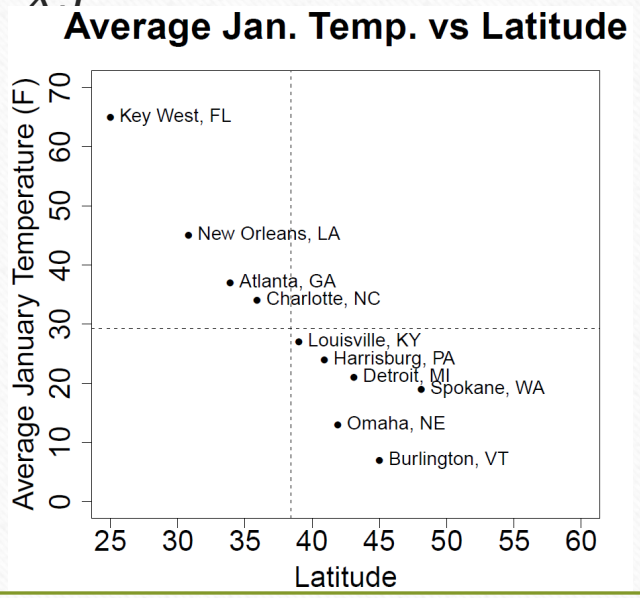
- The correlation, $R_{X,Y} = \frac{\sum_{i=1}^{n}[(x_i - \bar{x})(y_i - \bar{y})]}{(n-1)s_X s_Y}$, always lies between -1 and 1, inclusive.
- Suggested language:

If $R_{X,Y}$ is between	then we may say* X and Y have …
$0.5 \leq R_{X,Y} \leq 1$	a strong positive correlation
$0 < R_{X,Y} < 0.5$	a weak positive correlation
$-0.5 < R_{X,Y} < 0$	a weak negative correlation
$-1 \leq R_{X,Y} \leq -0.5$	a strong negative correlation

STA 674: Measuring Association between Two Variables Correlation

- Example: Average January temperature in US cities

Previously, we found $R_{XY} = -0.94$

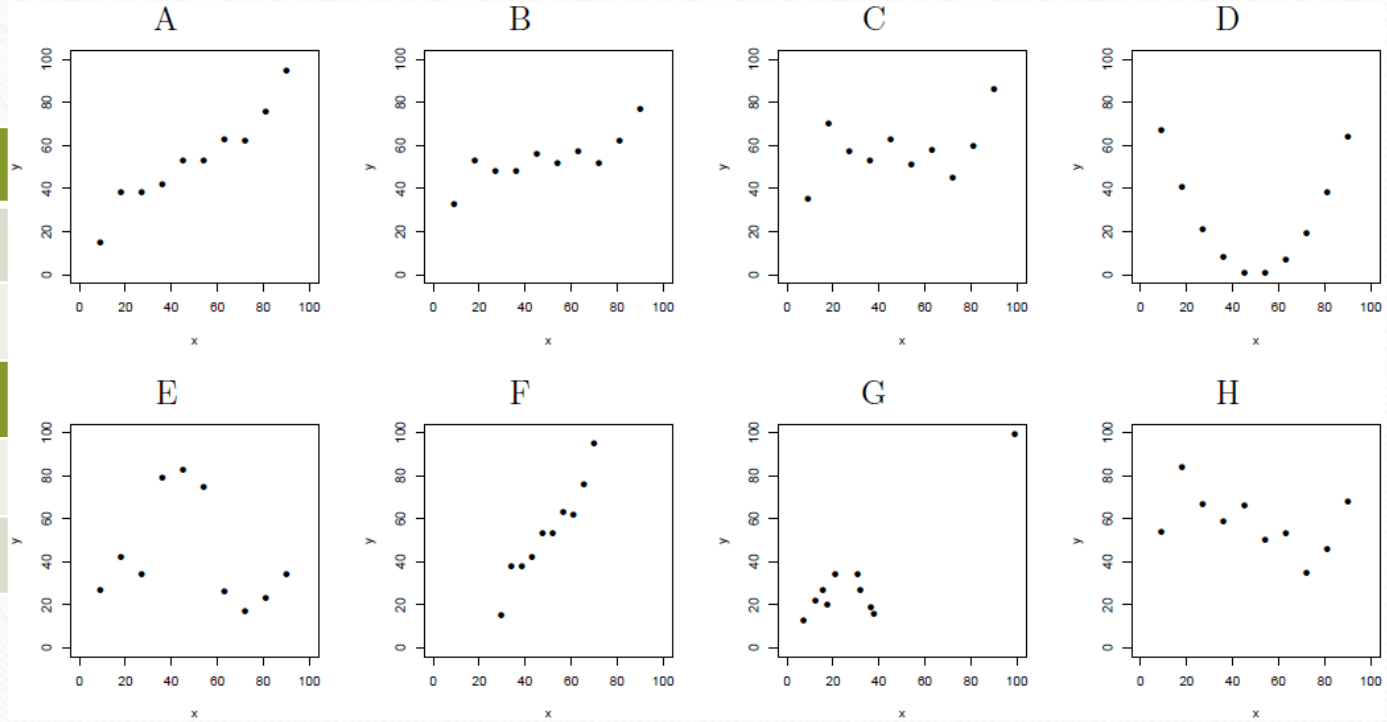


City		Latitude	Average Jan Temp
Louisville,	KY	39	27
Key West,	FL	25	65
New Orleans,	LA	30.8	45
Atlanta,	GA	33.9	37
Charlotte,	NC	35.9	34
Harrisburg,	PA	40.9	24
Omaha,	NE	41.9	13
Detroit,	MI	43.1	21
Burlington,	VT	45	7
Spokane,	WA	48.1	19
Means		38.4	29.2
Standard deviations		7	16.9

STA 674: Measuring Association between Two Variables Correlation

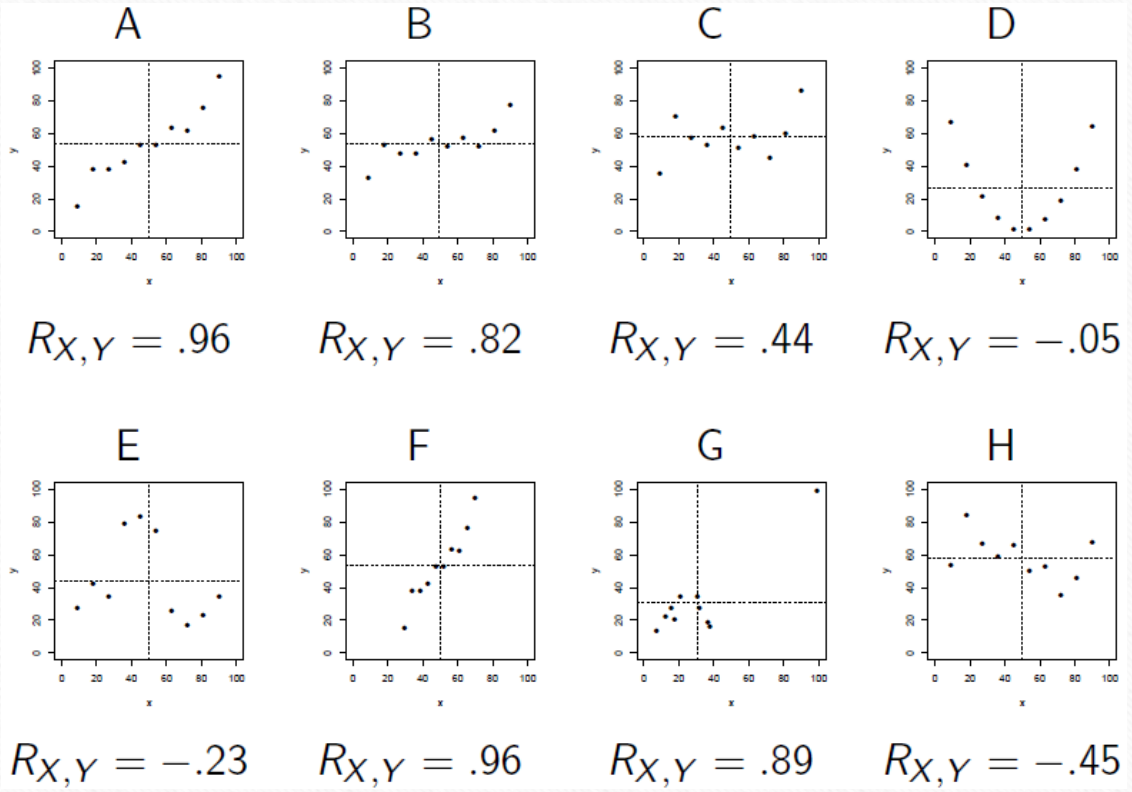
- Practice (?)

Graph	A	B	C	D
$\bar{x}$	50	50	50	50
$\bar{y}$	54	54	58	27
Graph	E	F	G	H
$\bar{x}$	50	50	31	50
$\bar{y}$	44	54	31	58



STA 674: Measuring Association between Two Variables Correlation

- True values:



STA 674: Measuring Association between Two Variables Correlation

- Next time – *cautions* about interpretations of the correlation coefficient.