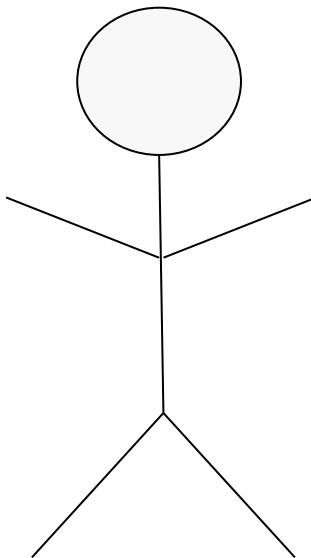




Looking at Associations with Multivariate Quantitative Data

Julie Deke and Tim NeCamp

Gathering Multivariate Quantitative Data



What is your age?

Let's measure your:

- Body mass index (BMI)
- Blood pressure
- Cholesterol level

What is Multivariate Quantitative Data?

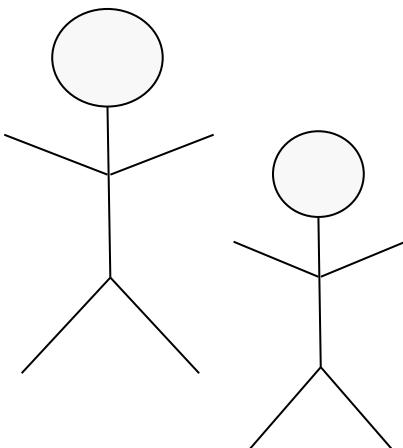
Multivariate

more than one trait measured
per unit

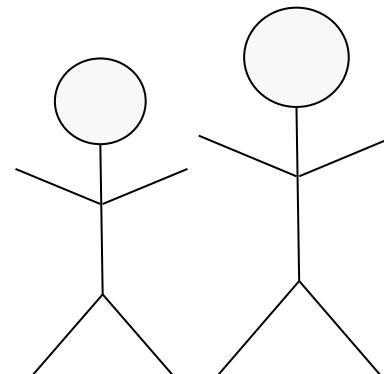
Quantitative

takes on numeric value

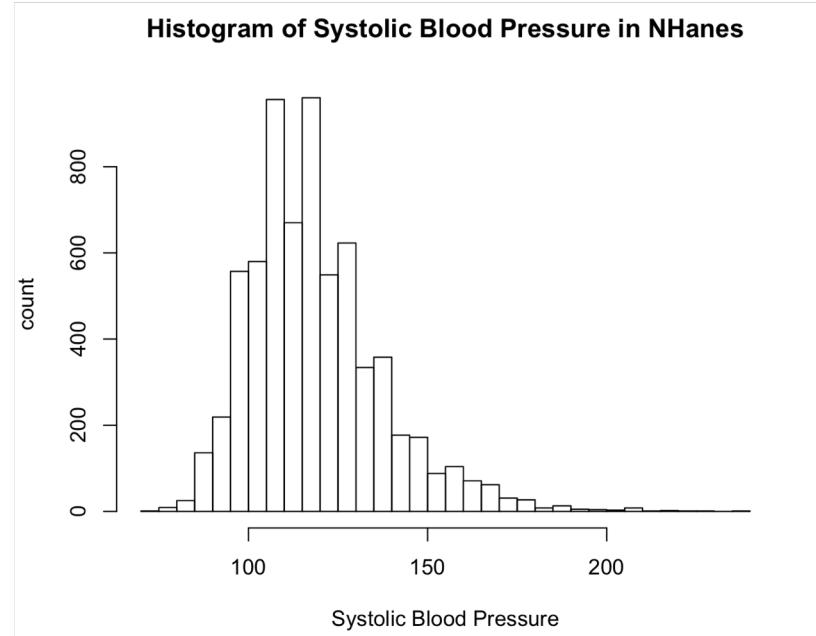
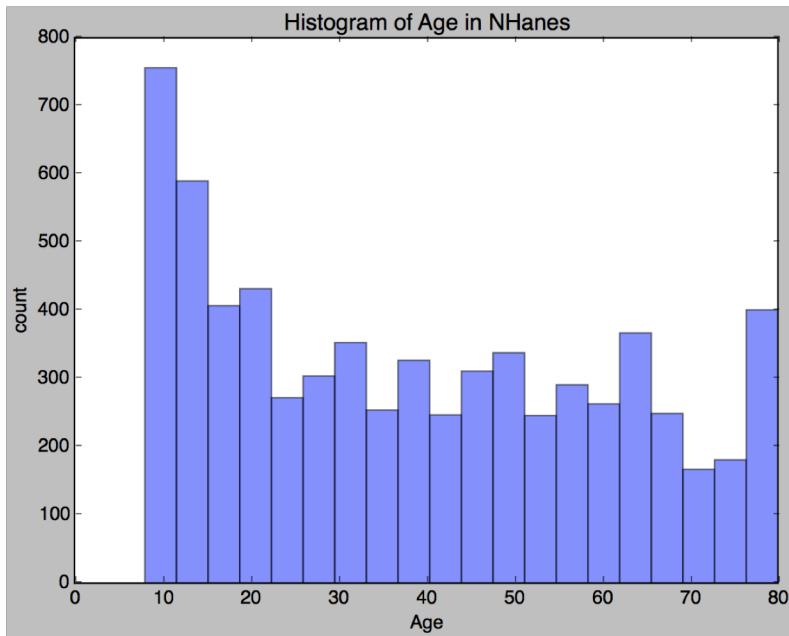
Recording Multivariate Quantitative Data



age	systolic.blood.pressure	bmi	hdl.cholesterol
22	110	23.3	41
14	112	17.3	44
44	116	23.2	28
14	110	27.2	63
21	124	20.1	43
15	124	18.2	61
14	112	19.9	42
43	100	33.3	73
51	152	20.1	43
80	124	28.5	47
55	126	27.6	54
35	108	27.9	33
26	120	22.1	61
17	108	22.9	54
30	94	22.4	48
15	110	17.0	63
11	108	26.7	41
17	136	28.5	42
9	106	14.7	71



Displaying with Univariate Histograms



Displaying with a Scatterplot

Correlation:

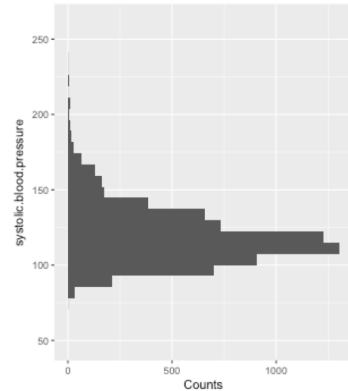
R: 0.58

Correlation:
R: 0.58

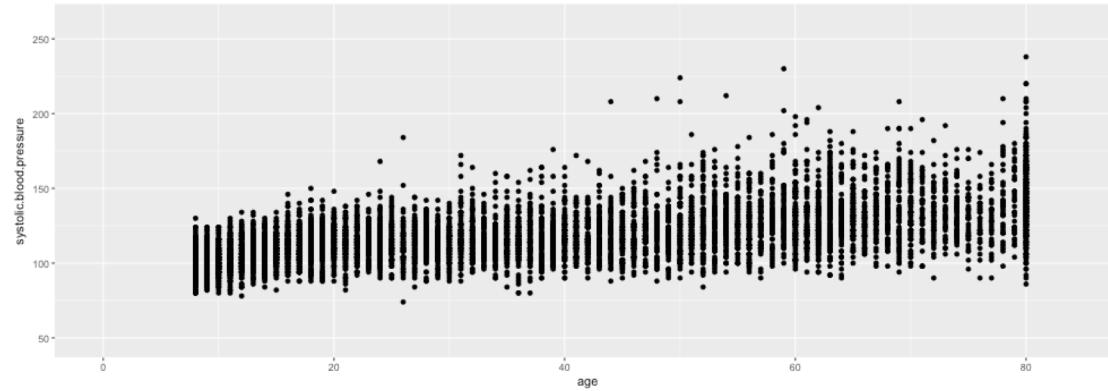
R²: 0.34

R²: 0.34

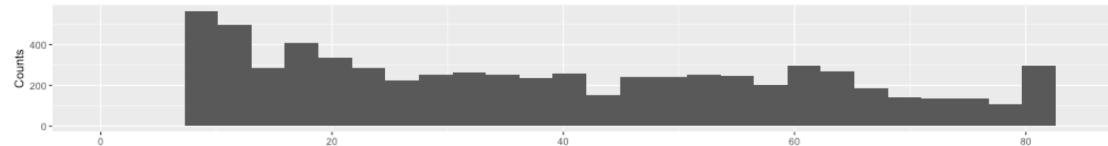
Marginal Distribution of Y



Joint Distribution

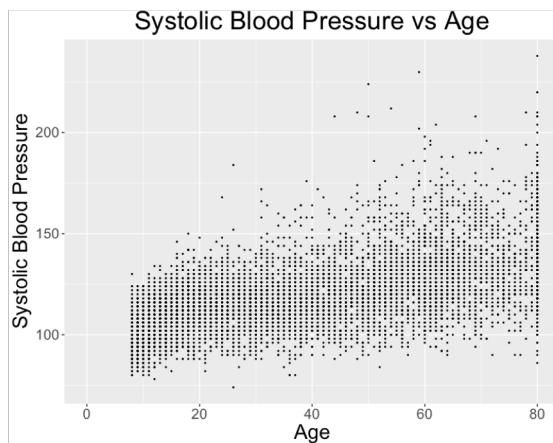


Marginal Distribution of X

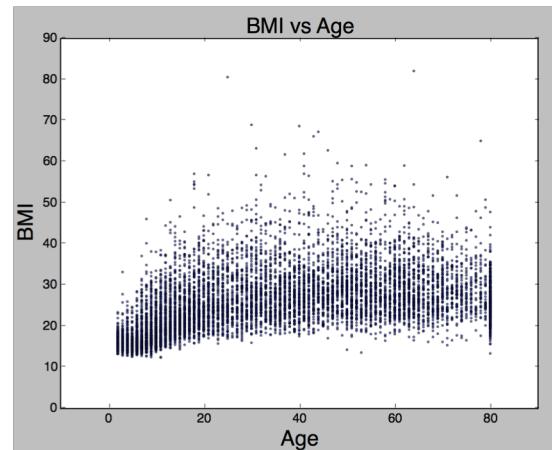


Association- Type

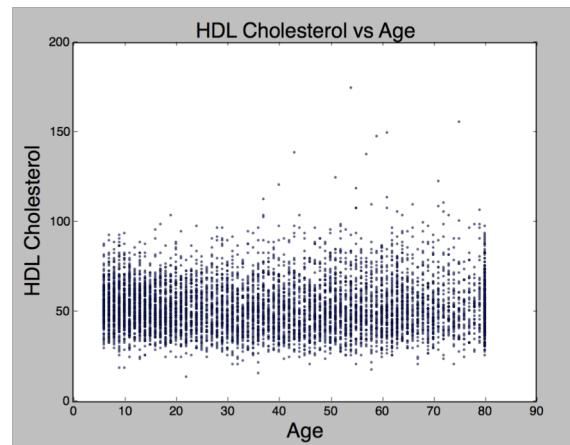
Linear association-
the pattern is a line



Quadratic association-
the pattern is parabolic

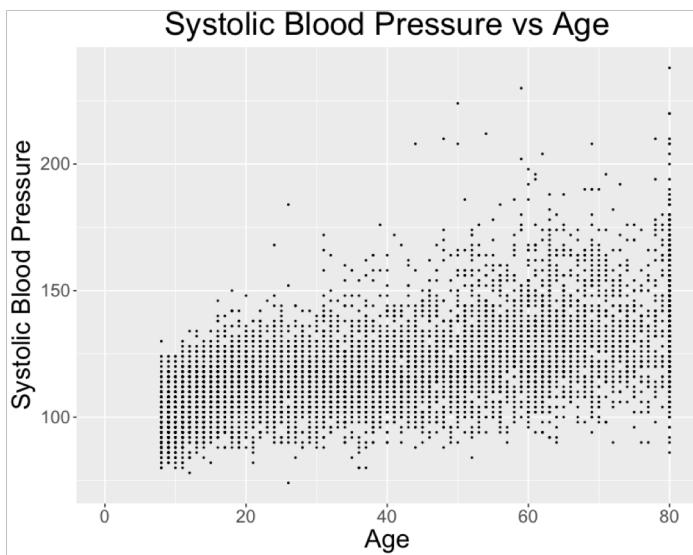


No association-
there is no pattern

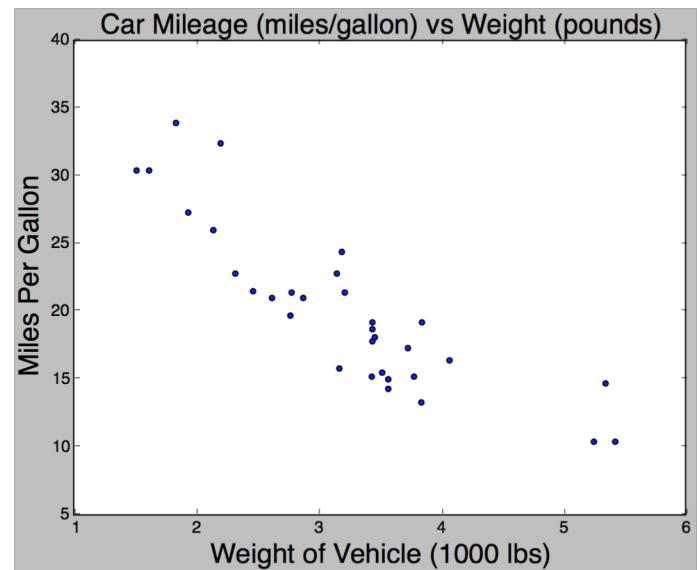


Association- Direction

Positive linear association - pattern has a positive slope, when x increases, y increases



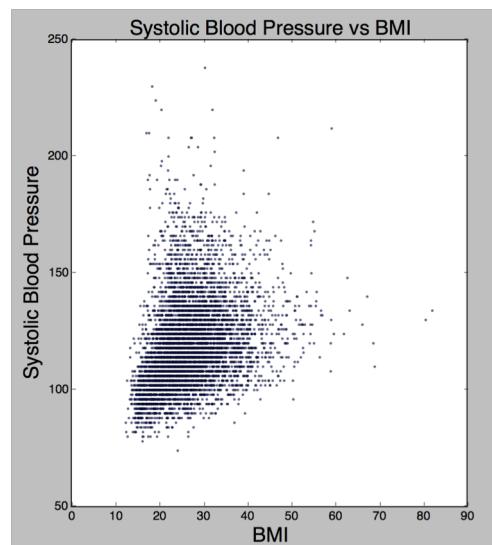
Negative linear association - pattern has a negative slope, when x increases, y decreases



Association- Strength

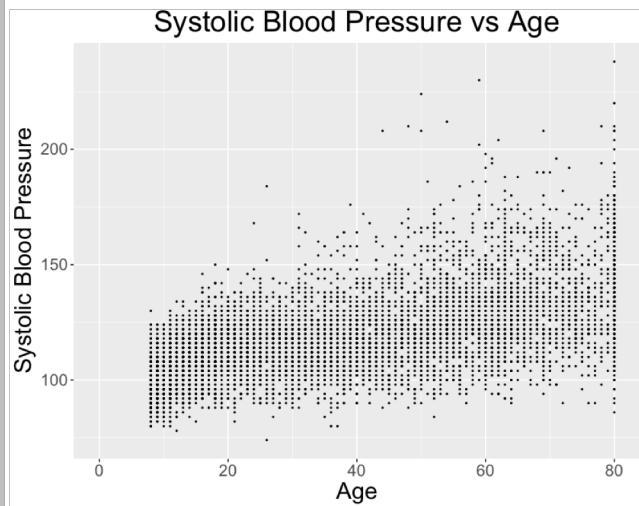
Weak linear association-

points are largely scattered along a line



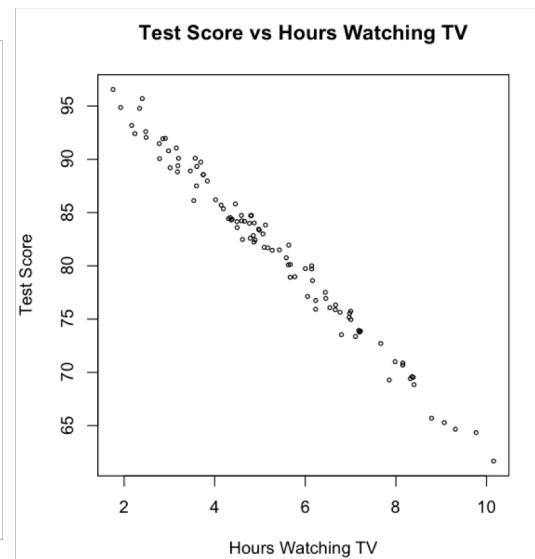
Moderate linear association-

points are partially scattered along a line



Strong linear association-

points are minimally scattered along a line

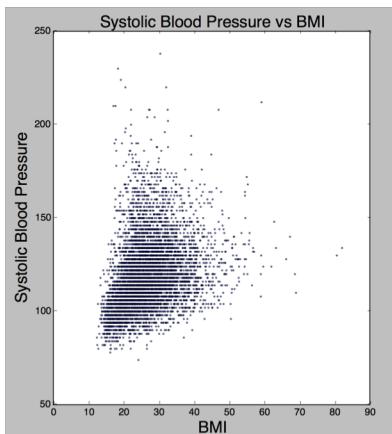


Correlation

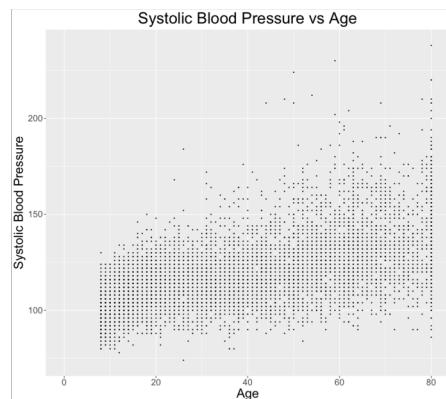
Pearson correlation (R or ρ): number between -1 and 1 indicating the strength and sign of association between 2 variables

The sign of the correlation is the sign of the association

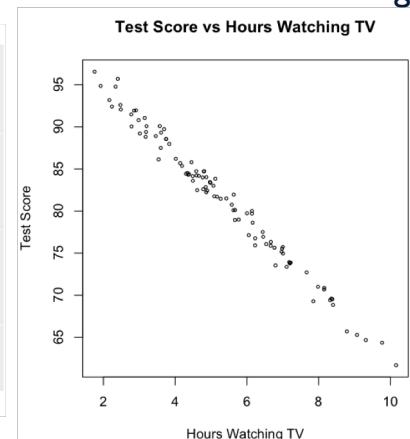
The closer the number is to 1 or -1, the stronger the association



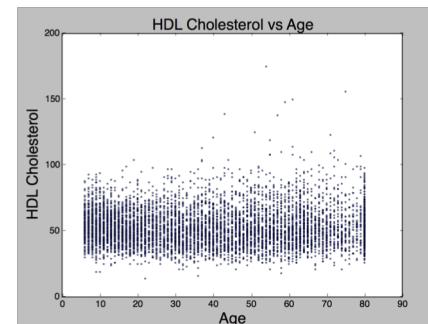
R = 0.30



R = 0.58

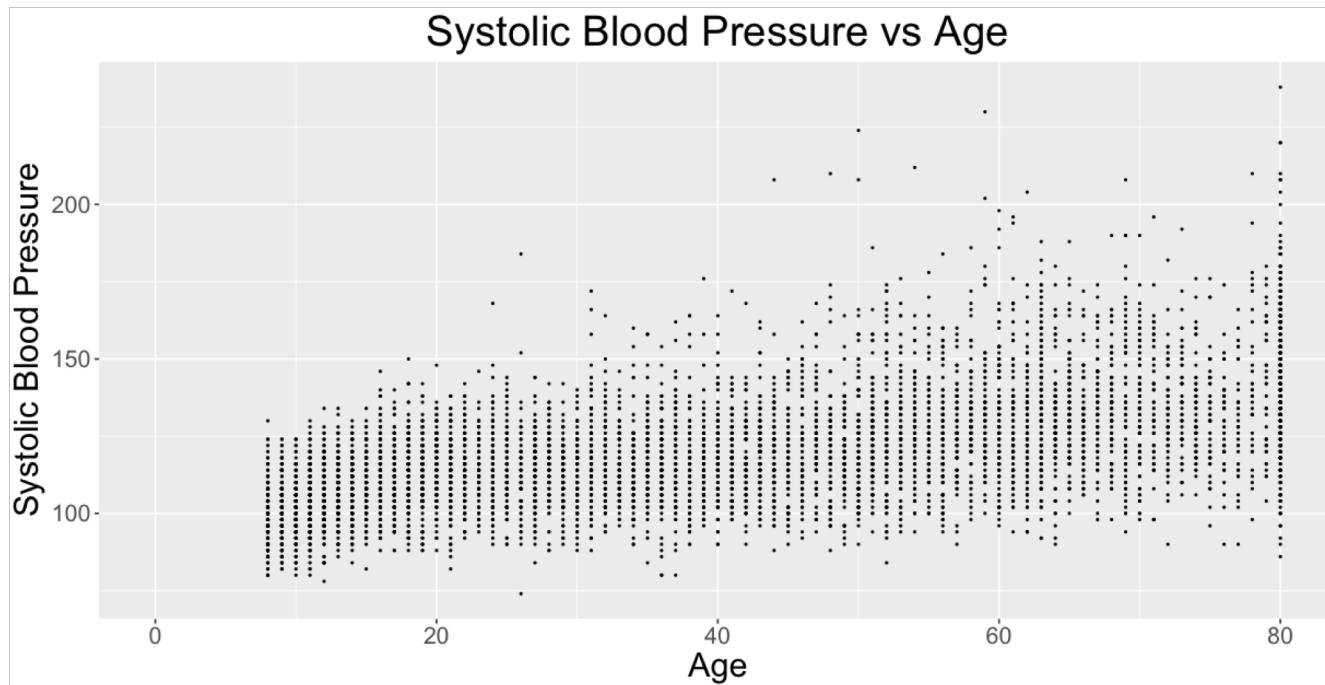


R = 0.99

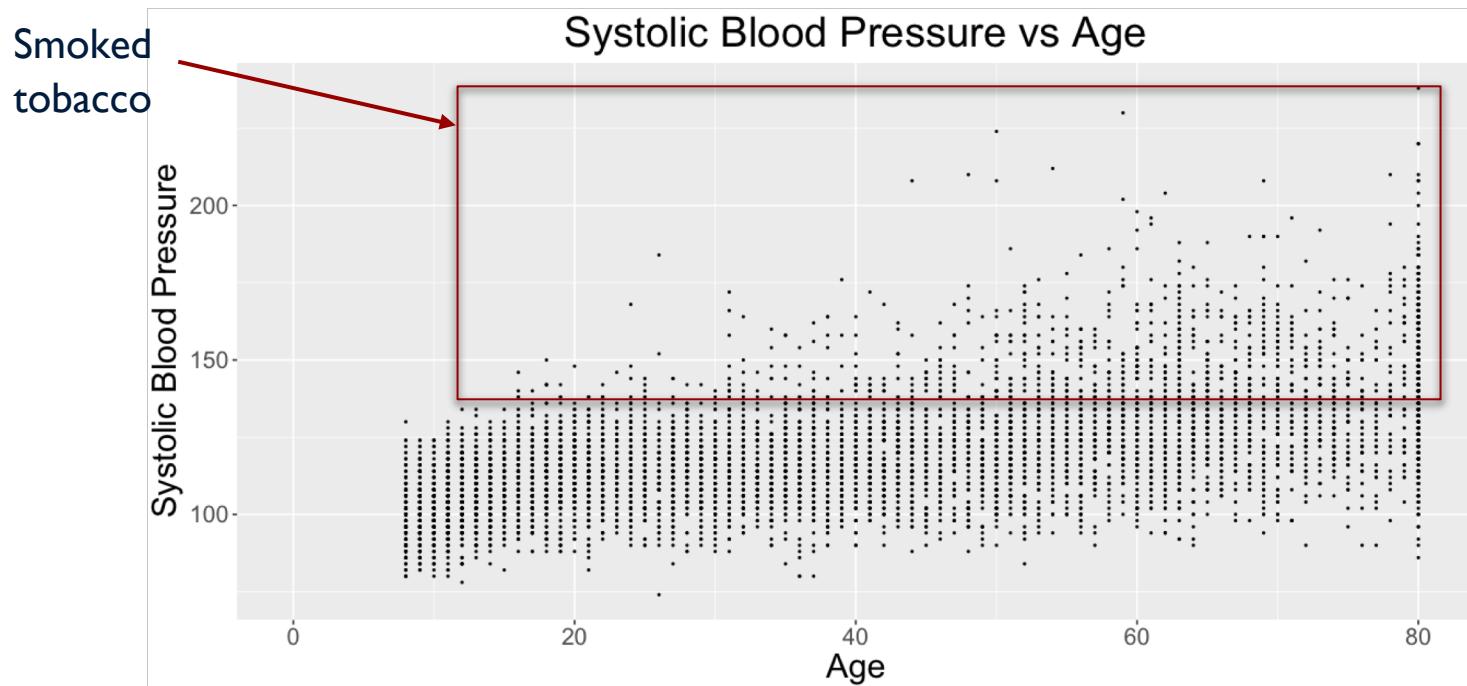


R = -.01

Correlation Does Not Imply Causation

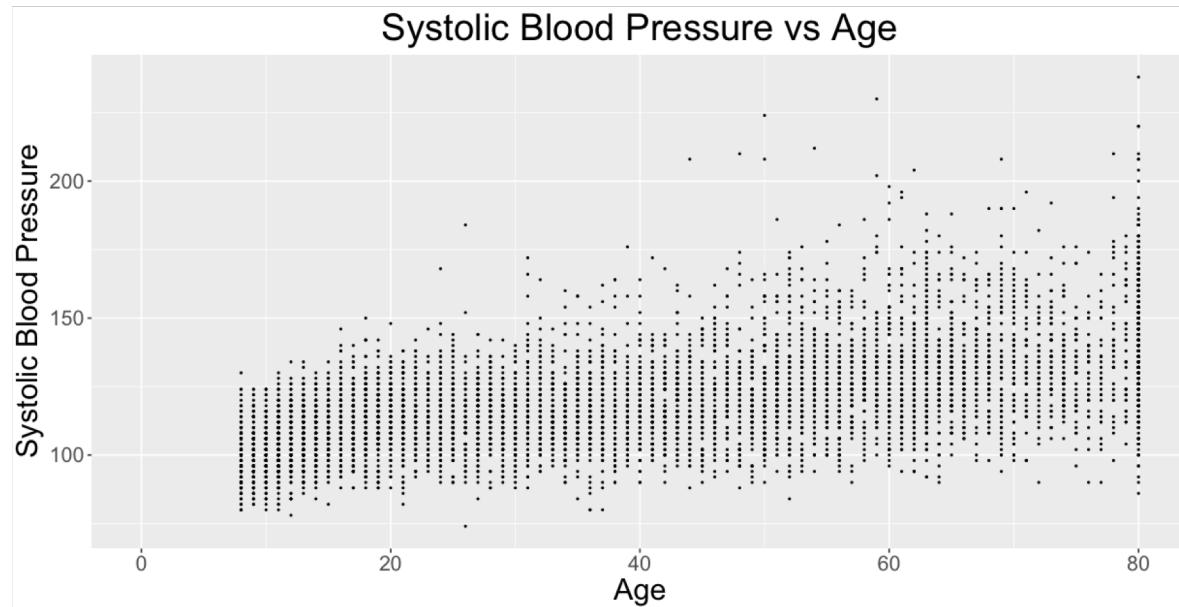


Correlation Does Not Imply Causation



Outliers in Multivariate Quantitative Data

Outliers - extreme data points that deviate from patterns in the rest of the data



Displaying Quantitative and Categorical Data

