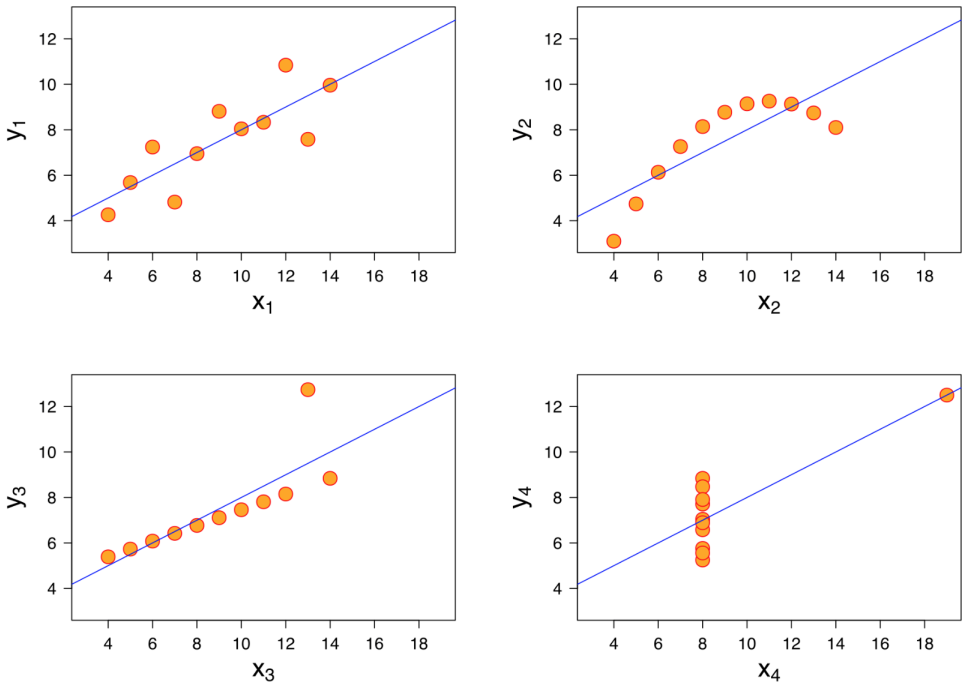
HGDS 200 - Foundations of Data-Driven Analysis 2018/2019 - Handout 4
1/14/2019 - 1/21/2019 Name:

The following should be completed after reading Chapter 5 (Measuring Performance in Regression Models) of *Applied Predictive Modeling*.

Consider the four data sets graphed below known as "Anscombe's Quartet." These four data sets have the same simple descriptive statistics (written below the figure) and are modeled by the same linear regression line with the same R^2 value.



Property	Value	Accuracy
Mean of x	9	exact
Sample variance of x	11	exact
Mean of y	7.50	to 2 decimal places
Sample variance of y	4.125	±0.003
Correlation between x and y	0.816	to 3 decimal places
Linear regression line	$y = 3.00 + 0.500x$	to 2 and 3 decimal places, respectively
Coefficient of determination of the linear regression	0.67	to 2 decimal places

(1) Which model(s) seem(s) to exhibit a bias issue? How might you fix this issue?
(2) Which model(s) seem(s) to exhibit a variance issue? How might you fix this issue?
(3) Is it obvious which model has the smallest MSE?
(4) How can you describe the MSE of these models geometrically?