

Discussion of Descriptive Exercises

By now you've probably realized that these data are *not* actually data on police expenditures and crime from Massachusetts, but rather a set of datasets that have been constructed to illustrate how easily summary statistics can lead you wrong.

Indeed, *all* of the "counties" in this dataset are actually different datasets with the same means, standard deviations, and correlation between the two variables. And yet...

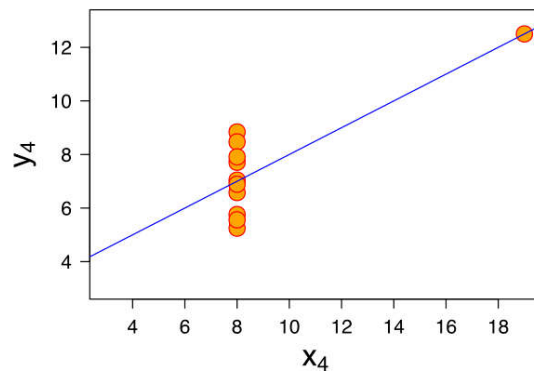
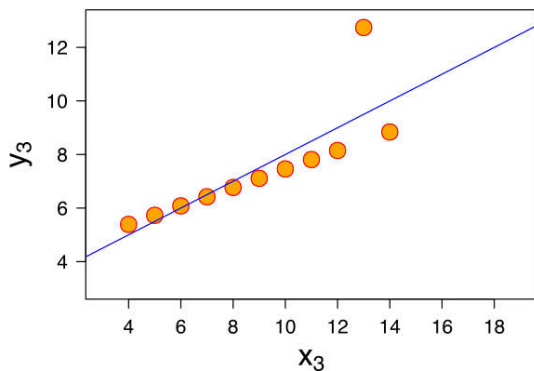
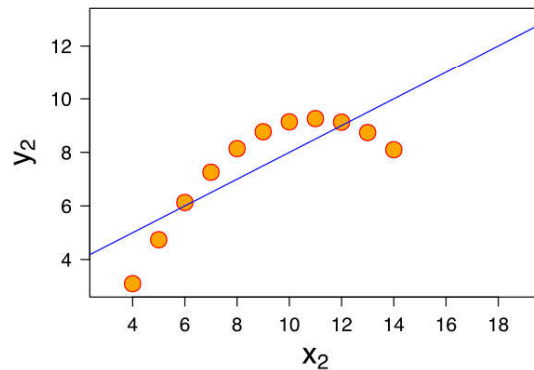
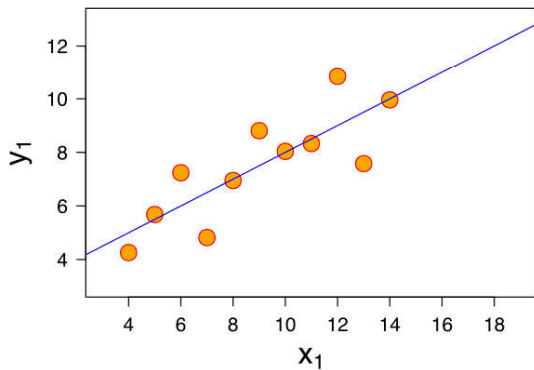
Exercise 8

Write a loop that plots the relationship between `policeexpenditures` and `crimeindex` for all 13 counties in the dataset separately!

Takeaways

Yup. All remarkably different.

While these are especially cute, they are in fact just fancy versions of Anscombe's Quartet [https://en.wikipedia.org/wiki/Anscombe%27s_quartet] – a set of four ways in which standard analyses of simple, bi-variate regressions can go wrong. Each of these datasets have the same mean, same variance, same correlation, and same regression line. And yet they are obviously very different – the first is reasonable, the second doesn't have a linear functional form, the third has an outlier in y , and the fourth has a constant x , save a single outlier.



And these are just some of the ways that your summary statistics can betray you when you're working with *only two variables*! Now think about some of the real datasets you've worked with where you may have a dozen explanatory variables.

So what should you take away from this?

- First, never be too trusting of your data or your summary statistics. It takes forever to get a feel for real datasets, so explore, explore, explore.
- Second, plot your data. Plotting is an incredibly information-rich, and is one of the best ways to diagnose problems in your data.

Absolutely positively need the solutions?

Don't use this link until you've really, really spent time struggling with your code! Doing

so only results in you cheating yourself.

Link [[../solutions_warning.html](#)]