



## M2-06: Anscombe's Quartet

Part of the "Exploratory Data Analysis" Learning Badge

Video Walkthrough: <https://discovery.cs.illinois.edu/m2-06/>

### Descriptive Descriptive Statistics

Visualization is particularly useful even when you begin to understand your dataset through descriptive statistics. **Anscombe's Quartet** is a famous set of four data points that has a some deceptively interesting properties:

	Set 1		Set 2		Set 3		Set 4	
	x	y	x	y	x	y	x	y
0	10	8.04	10	9.14	10	7.46	8	6.58
1	8	6.95	8	8.14	8	6.77	8	5.76
2	13	7.58	13	8.74	13	12.74	8	7.71
3	9	8.81	9	8.77	9	7.11	8	8.84
4	11	8.33	11	9.26	11	7.81	8	8.47
5	14	9.96	14	8.1	14	8.84	8	7.04
6	6	7.24	6	6.13	6	6.08	8	5.25
7	4	4.26	4	3.1	4	5.39	19	12.5
8	12	10.84	12	9.13	12	8.15	8	5.56
9	7	4.82	7	7.26	7	6.42	8	7.91
10	5	5.68	5	4.74	5	5.73	8	6.89
Mean:								
SD:								
Corr:								

Dataset URL: <https://waf.cs.illinois.edu/discovery/anscombe.csv>

The Anscombe's Quartet has several interesting properties:

1. The x values in Sets 1, 2, and 3 are **identical**.
2. The mean and SD of the x values between each set are \_\_\_\_\_.
3. The mean and SD of the y values between each set are \_\_\_\_\_.
4. The correlation and line of best fit (regression line) are \_\_\_\_\_.

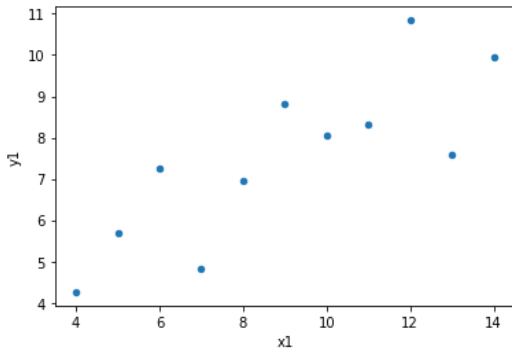


## M2-06: Anscombe's Quartet

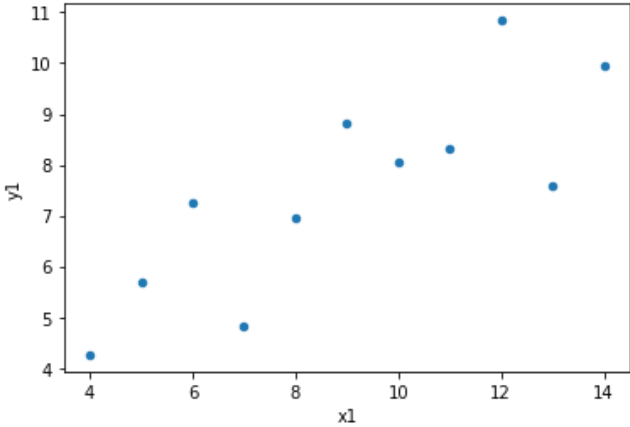
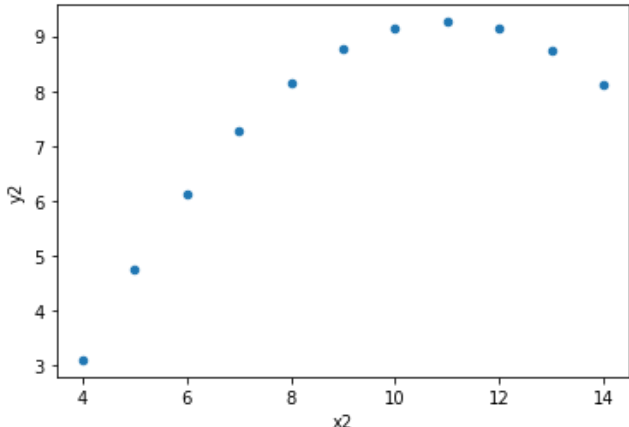
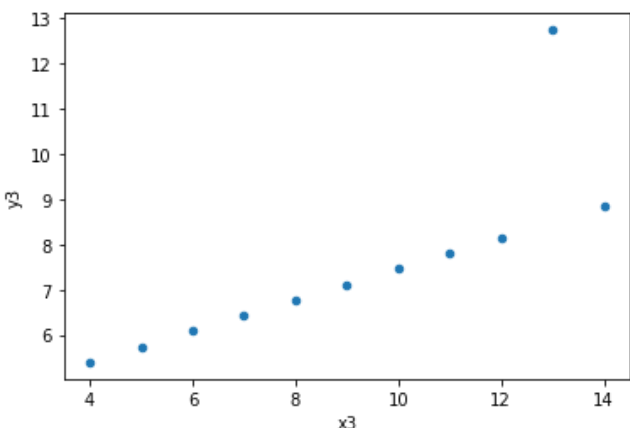
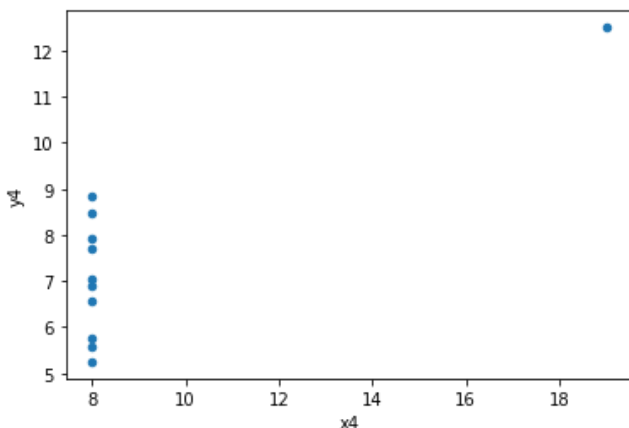
Part of the "Exploratory Data Analysis" Learning Badge

Video Walkthrough: <https://discovery.cs.illinois.edu/m2-06/>

### Creating a Scatter Plot

<b>Python:</b>	<b>General Syntax:</b> <code>df.plot.scatter(x="x3", y="y3")</code>  <b>Require Options:</b> <code>x = "column", numeric column</code> <code>y = "column", numeric column</code>	
<b>Example Shown:</b>	<pre># df contains Anscombe's Quartet df.plot.scatter(x="x1", y="y1")</pre>	

### Anscombe's Quartet Visualized

<p style="text-align: center;"><b>Set 1</b></p>  <p style="text-align: center;"><code>df.plot.scatter(x="x1", y="y1")</code></p>	<p style="text-align: center;"><b>Set 2</b></p>  <p style="text-align: center;"><code>df.plot.scatter(x="x2", y="y2")</code></p>
<p style="text-align: center;"><b>Set 3</b></p>  <p style="text-align: center;"><code>df.plot.scatter(x="x3", y="y3")</code></p>	<p style="text-align: center;"><b>Set 4</b></p>  <p style="text-align: center;"><code>df.plot.scatter(x="x4", y="y4")</code></p>