

WHY DO WE VISUALISE DATA?



AGENDA.

- Table With Four Groups Of Numbers
- Why Do We Visualise Data
- Preattentive Attributes
- Aggregation
- Summary

Agile Hitman

TABLE WITH FOUR GROUPS OF NUMBERS.

- In the following **table** there are **four groups of numbers**, each with **11 pairs**.
- In a moment, we will **create a chart** from them, but before we do, **take a look** at the numbers.
 - **What can you see ?**
 - Are there any **discernible differences** in the **patterns** or **trends** among them ?
- You **don't see anything clearly** – it's too **hard**.

Group A		Group B		Group C		Group D	
x	y	x	y	x	y	x	y
10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89

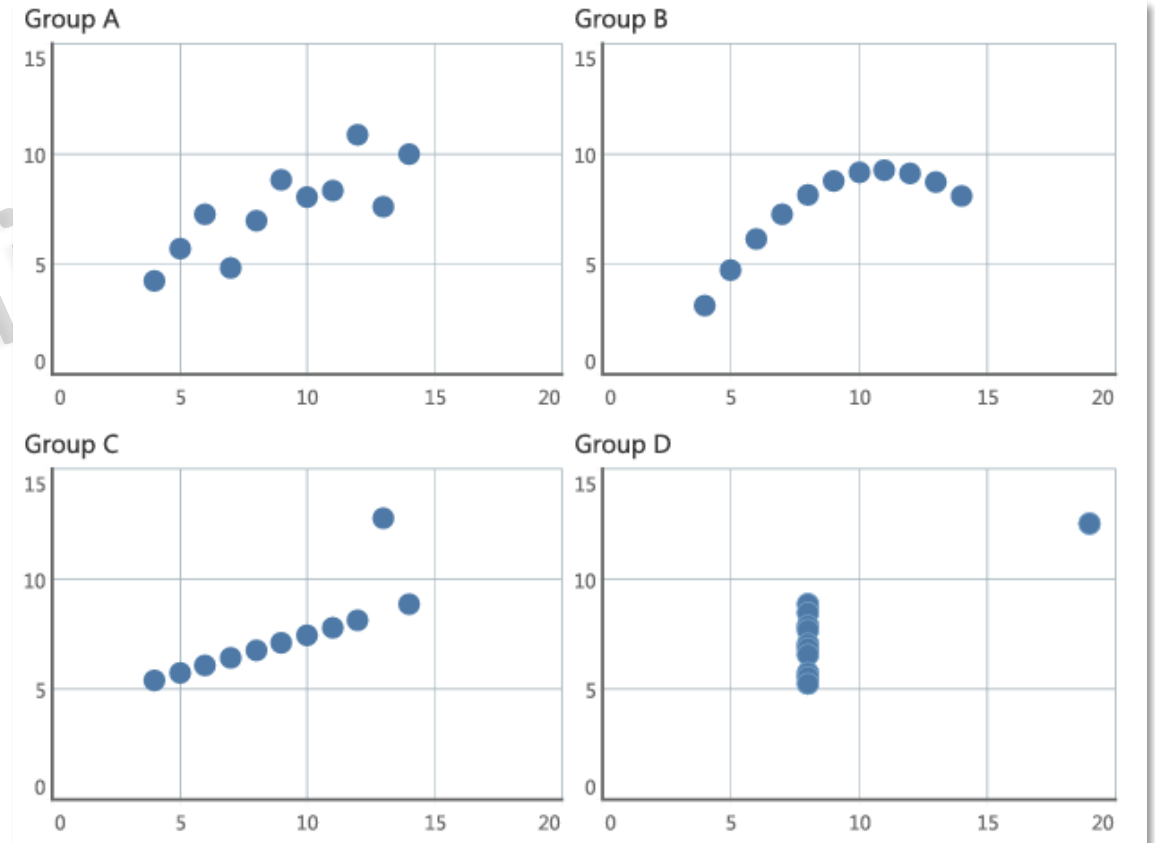
WHY DO WE VISUALISE DATA ?

- Before we put the **numbers in a chart**, we might consider their **statistical properties**.
- Were we to do that, we'd find that the **statistical properties** of each group of numbers are **very similar**.
- If the **table doesn't show anything** and **statistics doesn't reveal much**, what happens when we **plot the numbers** ?

Agile Hitman

WHY DO WE VISUALISE DATA ?

- Now do you see the **differences** ?
 - Seeing the numbers in **a chart** shows you something that **tables** and some **statistical measures cannot**.
 - We **visualise** data to harness the incredible **power of our visual system** to spot **relationships** and **trends**.
- This brilliant example is the creation of **Frank Anscombe**, a **British statistician**.
 - He created this set of numbers—called “**Anscombe’s Quartet**”—in his paper “**Graphs in Statistical Analysis**” in 1973.
 - In the paper, he fought against the notion that “**numerical calculations are exact, but graphs are rough.**”



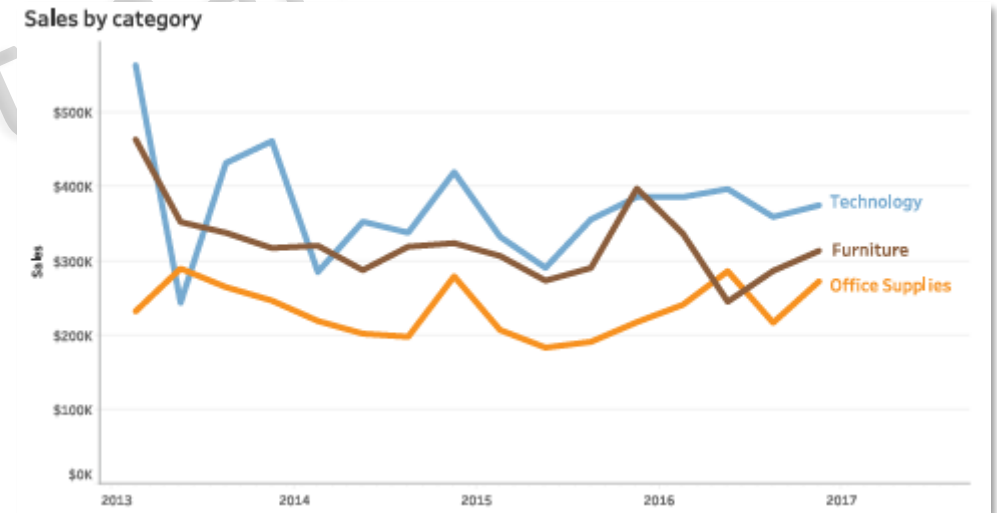
WHY DO WE VISUALISE DATA ?

- Another reason to **visualise** numbers is to help our **memory**.
 - The table shows **sales numbers** for three categories, by quarter, over a four-year period.
 - What **trends** can you see ?
- **Identifying trends** is as **hard** as it was with Anscombe's Quartet.
 - To read the table, we need to **look up every value**, one at a time.
 - Unfortunately, our **short-term memories** aren't designed to store many pieces of information.
 - By the time we've reached the fourth or fifth number, we will have **forgotten** the first one we looked at.

Category	2013 Q1	2013 Q2	2013 Q3	2013 Q4	2014 Q1	2014 Q2	2014 Q3	2014 Q4
Furniture	\$463,988	\$352,779	\$338,169	\$317,735	\$320,875	\$287,934	\$319,537	\$324,319
Office Supplies	\$232,558	\$290,055	\$265,083	\$246,946	\$219,514	\$202,412	\$198,268	\$279,679
Technology	\$563,866	\$244,045	\$432,299	\$461,616	\$285,527	\$353,237	\$338,360	\$420,018
Category	2015 Q1	2015 Q2	2015 Q3	2015 Q4	2016 Q1	2016 Q2	2016 Q3	2016 Q4
Furniture	\$307,028	\$273,836	\$290,886	\$397,912	\$337,299	\$245,445	\$286,972	\$313,878
Office Supplies	\$207,363	\$183,631	\$191,405	\$217,950	\$241,281	\$286,548	\$217,198	\$272,870
Technology	\$333,002	\$291,116	\$356,243	\$386,445	\$386,387	\$397,201	\$359,656	\$375,229

WHY DO WE VISUALISE DATA ?

- Office supplies has been the **lowest-selling** product category in all but two quarters.
- Furniture trends have been **dropping slowly** over the time period, except for a bump in sales in 2015 Q4 and a **rise** in the last two quarters.
- Technology sales have mostly been the **highest** but were particularly **volatile** at the start of the time period.



PREATTENTIVE ATTRIBUTES.

- These are things that our brain **processes** in **milliseconds**.
 - **How many 9s are there ?**
 - How did you do ?
- It's **easy** to answer the question— you just look at all the values and **count the 9s**—but it takes a **long time**.
- We can make **one change** to the grid and make it very **easy** for you.

2	2	5	6	7	1	1	6	9	1
9	1	7	5	5	5	6	2	5	9
4	5	2	9	6	9	7	6	4	6
8	1	5	7	8	5	6	6	6	7
7	2	3	6	8	9	1	7	9	1
3	8	6	8	4	5	6	9	4	5
4	9	9	2	3	7	1	9	1	2
3	7	8	1	6	1	5	6	1	6
5	6	6	8	6	6	9	1	2	6
3	2	4	2	6	9	4	2	7	1

PREATTENTIVE ATTRIBUTES.

- **Colour differences pop out.** It's as easy to find one **red 9** on a table of hundreds of digits as it is on a 10-by-10 grid.
 - Think about that for a moment:
 - Your brain **registers** the **red 9s** before you **consciously** addressed the grid to count them.

2	2	5	6	7	1	1	6	9	1
9	1	7	5	5	5	6	2	5	9
4	5	2	9	6	9	7	6	4	6
8	1	5	7	8	5	6	6	6	7
7	2	3	6	8	9	1	7	9	1
3	8	6	8	4	5	6	9	4	5
4	9	9	2	3	7	1	9	1	2
3	7	8	1	6	1	5	6	1	6
5	6	6	8	6	6	9	1	2	6
3	2	4	2	6	9	4	2	7	1

Agile Hitman

PREATTENTIVE ATTRIBUTES.

- **Difference** in **size** is easy to see too.
- **Size** and **hue**: That's all very well when **counting** the 9s.
- What if our task is to **count the frequency** of each digit?
 - That's a slightly more realistic task, but we **can't** just use a **different colour** or **size** for each digit.

2	2	5	6	7	1	1	6	9	1
9	1	7	5	5	5	6	2	5	9
4	5	2	9	6	9	7	6	4	6
8	1	5	7	8	5	6	6	6	7
7	2	3	6	8	9	1	7	9	1
3	8	6	8	4	5	6	9	4	5
4	9	9	2	3	7	1	9	1	2
3	7	8	1	6	1	5	6	1	6
5	6	6	8	6	6	9	1	2	6
3	2	4	2	6	9	4	2	7	1

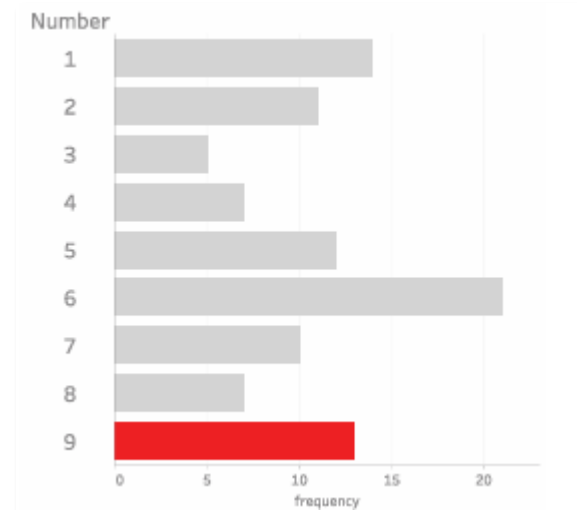
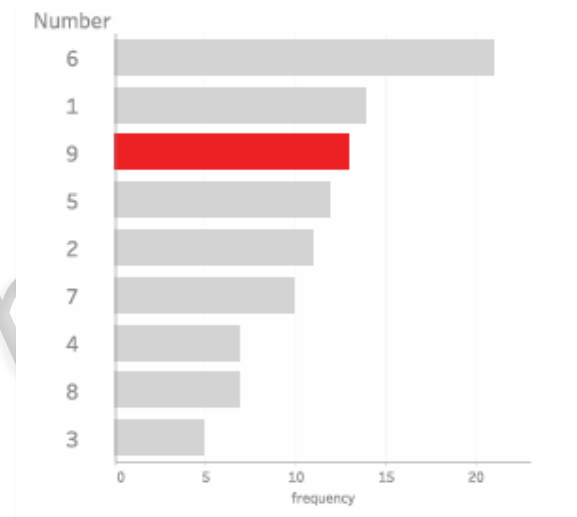
PREATTENTIVE ATTRIBUTES.

- **Colouring** every digit is nearly as bad as having no colour.
 - It's not a complete disaster - if you're looking for the 6s, you just need to work out that they are **red** and then **scan quickly** for those.
 - Using **one colour** on a visualisation is highly effective to make **one category** stand out.
 - Once you're up to around eight to ten categories, however, there are **too many colours** to easily **distinguish** one from another.
- To count each digit, we need to **aggregate**.

2	2	5	6	7	1	1	6	9	1
9	1	7	5	5	5	6	2	5	9
4	5	2	9	6	9	7	6	4	6
8	1	5	7	8	5	6	6	6	7
7	2	3	6	8	9	1	7	9	1
3	8	6	8	4	5	6	9	4	5
4	9	9	2	3	7	1	9	1	2
3	7	8	1	6	1	5	6	1	6
5	6	6	8	6	6	9	1	2	6
3	2	4	2	6	9	4	2	7	1

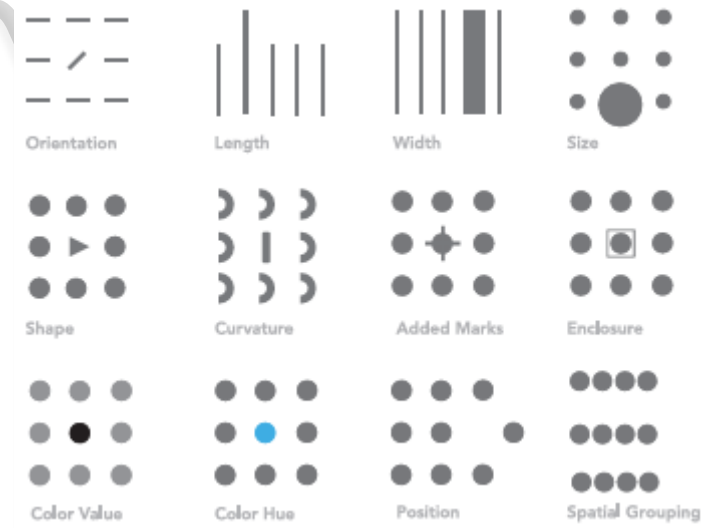
AGGREGATION.

- To **count** each digit, we need to **aggregate**.
 - **Visualisation** is, at its core, about **encoding aggregations**, such as **frequency**, in order to gain **insight**.
 - We need to move **away** from the **table entirely** and **encode the frequency** of each digit.
 - The most **effective** way is to use **length**, which we can do in a bar chart.
- Since the task is to **count** the 9s in the data source, the **bar chart** is one of the **best ways** to see the results.
 - This is because **length** and **position** are **best** for **quantitative comparisons**.



SUMMARY.

- To recap, we've seen how **powerful** the **visual system** is and looked at some **visual features** we can use to display data effectively.
- In the next course we will look at **Types of Data**.



Preattentive features