Histograms and scatter plots

How prices spread, and price vs metascore

Name _____

Superset → **Charts**, dataset `games_clean` (database `Dades classe`, schema `public`): 26,685 Steam games. Keep **SQL → SQL Lab** open in a second tab. Think of ranked mode: a **histogram** counts the players in each rank; a **scatter plot** puts one dot per player. Work alone and write what you see here.

Reference

	Bar chart	Histogram
X-axis	Categories that already exist.	One **numeric** column cut into equal **bins**.
Height	Any metric.	Always a count: rows inside the bin.
It answers	Which category is bigger?	What shape do the values have?

Cloud of dots	Reading
Rises to the right	Positive relationship.
Falls to the right	Negative relationship.
Round blob	No clear relationship.

A relationship never says which of the two causes the other.

Part A · The price histogram (18 min)

Price distribution. Where are most games, and how long is the tail?

Step	Do this	You should see / write
1	New chart on `games_clean`, chart type **Histogram**. It is not in the first list (**Featured**): click **All charts**, or type `histogram` in the search box. **Column**: `price`.	**Column**, **Bins** and **Row limit**.
2	**Row limit**: change `10000` to `50000`. Otherwise the chart reads only 10,000 of your games.	`50000` in **Row limit**.
3	Leave **Bins** at 5. **Create chart**. Hover over the first bar.	Bars you can really see: _____ Range of the first bar: _____

4	In SQL Lab:
	`SELECT game, price FROM games_clean`
	`ORDER BY price DESC LIMIT 6;`
		Most expensive: _____
		Its price: _____ Are they games? _____

5	Back in the chart: **Filters** → `price` → operator **Less or equal (<=)** → type `60` and pick it in the list → **Save**.	A chip `price <= 60` under **Filters**.
6	**Bins**: type `12` and press Enter. **Update chart**.	Bars: _____ Tallest: the _____ one The tail goes to the _____
7	**Customize** → **Y Axis Format**: `,d`. Hover over the first bar.	Exact count, not rounded: _____
8	**Save** → `UT3 · Price distribution`.	**Charts** lists it.
9	Optional. Tick **Cumulative**, **Update chart**: each bar adds all the bars before it. Do not save.	Share under 10: _____

A1 Step 3: bins are equal slices from the cheapest to the most expensive row. Why one giant bar?

A2 Read your final histogram in one sentence with one number in it: “Most games…”.

A3 You left the products above 60 out of the picture. Are they wrong data? What should the chart title say?

Part B · Price vs metascore (16 min)

Price vs metascore. Do more expensive games get better reviews?

Step	Do this	You should see / write
1	In SQL Lab: `SELECT has_metascore, COUNT(*) FROM games_clean GROUP BY has_metascore;`	With a metascore: _____ of 26,685 Roughly, in %: _____
2	New chart on `games_clean`, chart type **Scatter Plot**.	An empty chart in the Explore view.
3	**X-axis**: `price`. **Metrics**: **Simple** tab → column `metascore`, aggregate `AVG`.	Both controls filled in.
4	**Filters** → `has_metascore` → operator **Is true** → **Save**. **Create chart**.	A cloud of dots between 0 and 60 on the X-axis.
5	Hover over three dots: the one at 9.99, the one at 59.99, and the lowest one.	At 9.99: _____ At 59.99: _____ Lowest: _____ at price _____
6	**Save** → `UT3 · Price vs metascore`.	**Charts** lists it.

B1 Many games cost 9.99, but there is only one dot at 9.99. Look at your **Metrics** control: what does one dot stand for?

_____

B2 Does the cloud rise, fall, or neither? Use the reference.

_____

B3 Which dots would you trust less, and why?

_____

B4 A friend reads your chart and says: "On Steam, raising the price of a game improves its reviews." Find the two mistakes. Step 1 helps with one of them.

1. _____

2. _____

Done when Every blank is filled · two charts are saved: `UT3 · Price distribution` and `UT3 · Price vs metascore`.

Stuck?

- **The bars add up to 10,000**: **Row limit** is still at 10,000. Set 50,000.
- **`price` is not offered in** Column: the control lists numeric columns only. Tell the teacher.
- **Still one giant bar after the filter**: the chip `price <= 60` must appear under **Filters**, and you must press **Update chart**.
- **"Must specify a value for filters with comparison operators"**: the chip says `price <=` with no number. Click it, type `60`, pick it in the list, **Save**.
- **Scatter with a single dot**: `price` went to **Metrics** instead of **X-axis**.
- **Dots spread evenly like a bar chart**: **Force categorical** is ticked under the X-axis. Untick it.
- **Superset does not open**: pair up with a neighbour and tell the teacher.

At home Nothing compulsory over the break. Optional, for the final project: on your own open dataset, one histogram of a numeric column. Does it have outliers too? Your saved `UT3 ·` charts are your notes for January.