

Temas de Ciencia de Datos



Paradoja de Simpson



Paradoja de Simpson



La gente que incursiona en la ciencia de datos redescubre temas que otras disciplinas ya habían descubierto hace un tiempo, tal es el caso de la paradoja de Simpson. Dicho fenómeno (que no es una paradoja formalmente) ya fue estudiado entre quienes hacían encuestas en la década del 60, era un fenómeno bastante usual.

Ahora la descubrimos nosotros.

Para entenderlo vamos con un ejemplo en forma de cuentito.

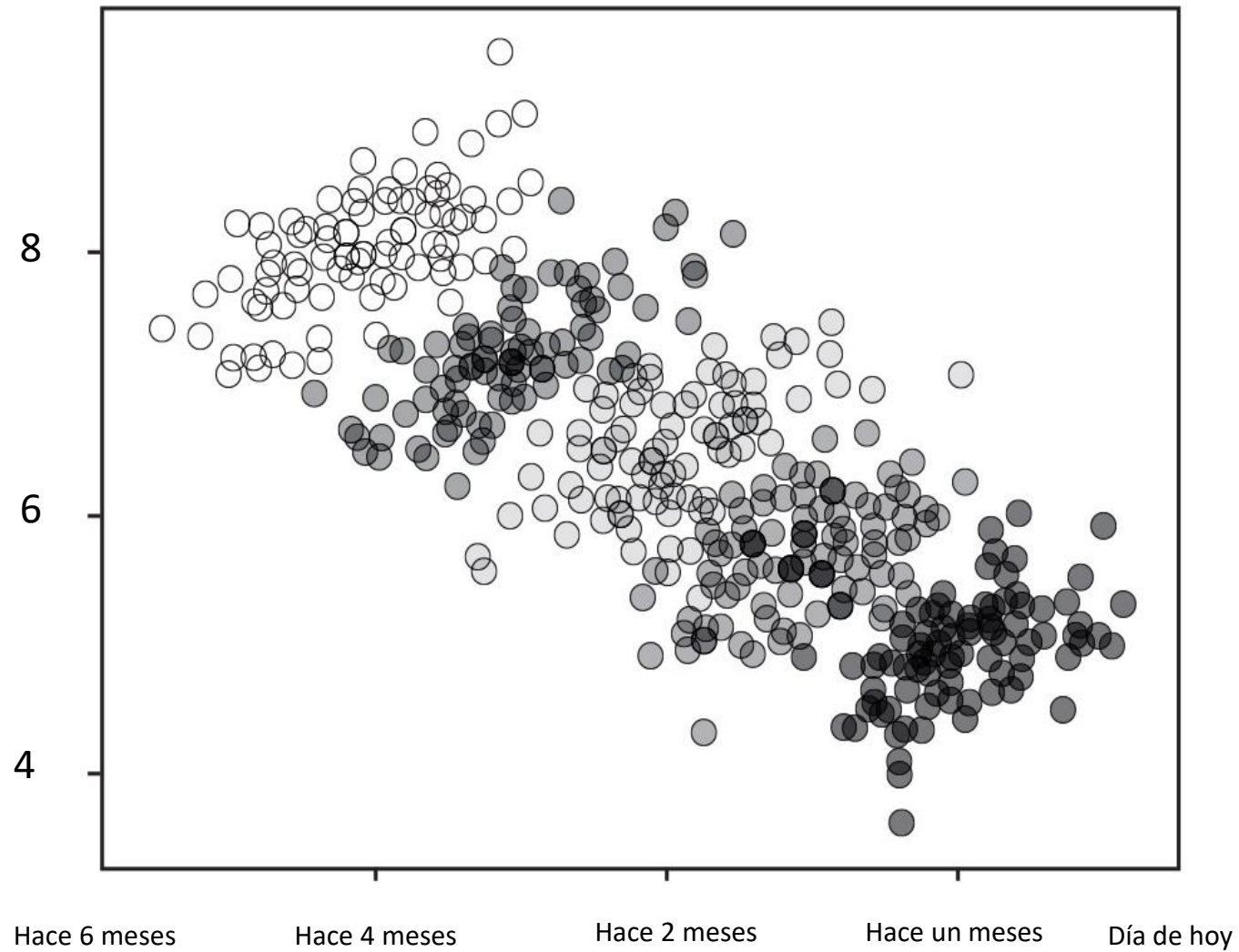
La historia podría contarse así (*)

Un director de colegio pide que le informen la evolución de las notas de los alumnos durante el ultimo semestre. Le pide a la gente de Sistemas que le armen un reporte que en el eje “X” figure el tiempo (ultimo semestre) y sobre el eje “Y” las notas de los alumnos. Luego recibe un cuadro como el que sigue:

(*) De todos los ejemplos de la Paradoja de Simpson este es el que mas me gusta, me lo conto Daniel Yankelevich hace un tiempo y desde entonces lo tome prestado



Paradoja de Simpson





Paradoja de Simpson



El director se lamenta, mirando la nube de puntos es fácil notar que las notas están bajando en el tiempo. A continuación decide hablar con cada uno de los Profesores para ver como abordar el tema.

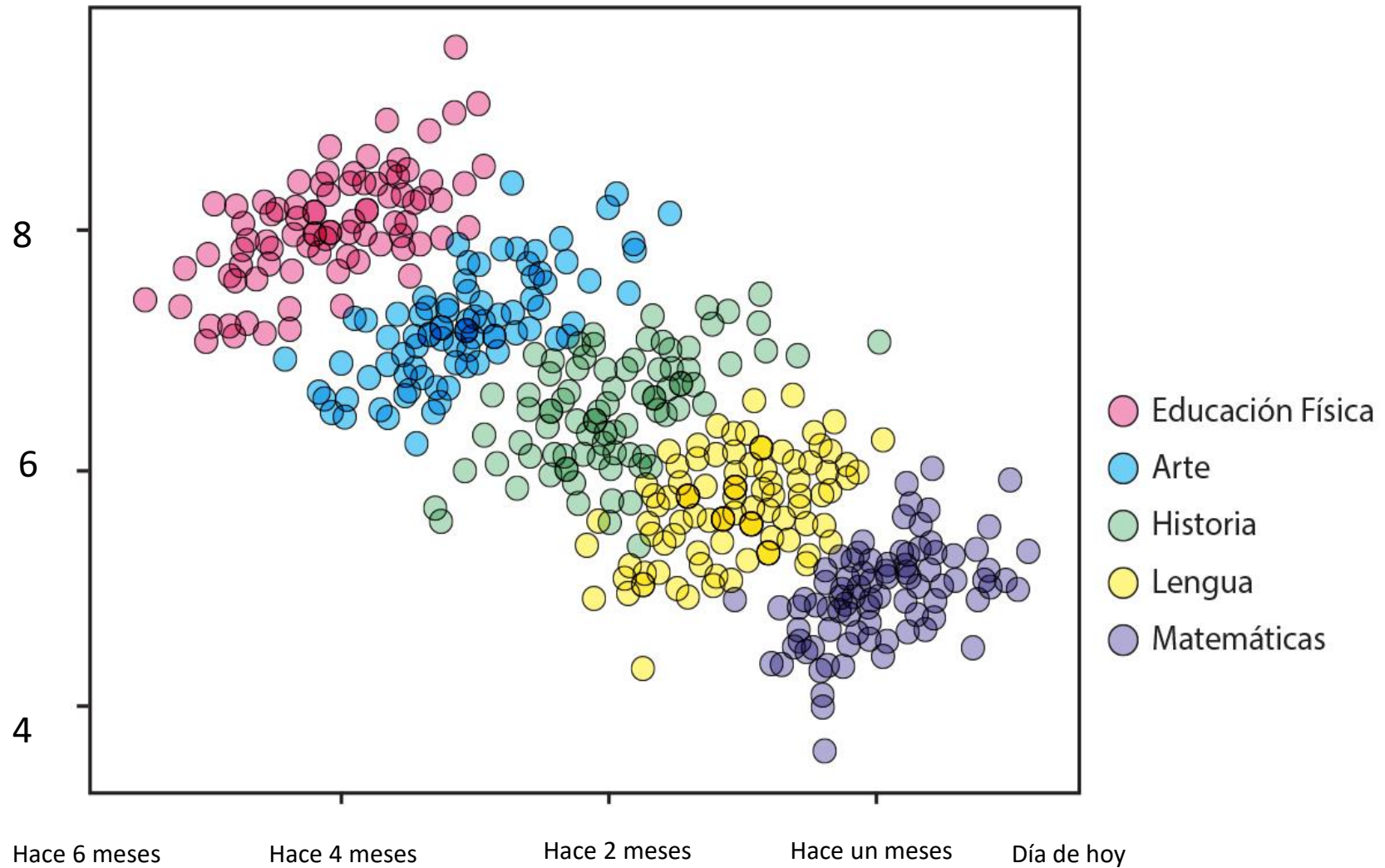
Primero habla con el profesor de matemáticas y le consulta “que podemos hacer? dado que las notas de los alumnos han bajando en el ultimo semestre”. El profesor lo mira sorprendido y le contesta “En matemáticas las notas están subiendo”, el director se queda pensando, decide hablar con otro profesor, el de historia y le hace el mismo planteo. Sorpresivamente obtiene la misma respuesta “las notas están subiendo”. Lo mismo le ocurre con todos los profesores a los que consulta.

Totalmente confundido pide que le hagan el mismo cuadro que antes pero que le marquen que notas son de que materia.

El cuadro que recibe el director es el siguiente:



Paradoja de Simpson





Paradoja de Simpson



Claramente era cierto lo que le decían los profesores. Todas las notas habían subido en el ultimo semestre.

Los dos reportes estaban bien hechos y los dos fenómenos son ciertos. Las notas estaban “bajando” en el ultimo semestre globalmente aunque para cada materia subían.

La moraleja es clara. Siempre que medimos un fenómeno nos puede quedar alguna variable “importante” sin medir. Quizás sea esa variable la que nos explica bien el fenómeno que nuestras primeras conclusiones en realidad no lo vieron.