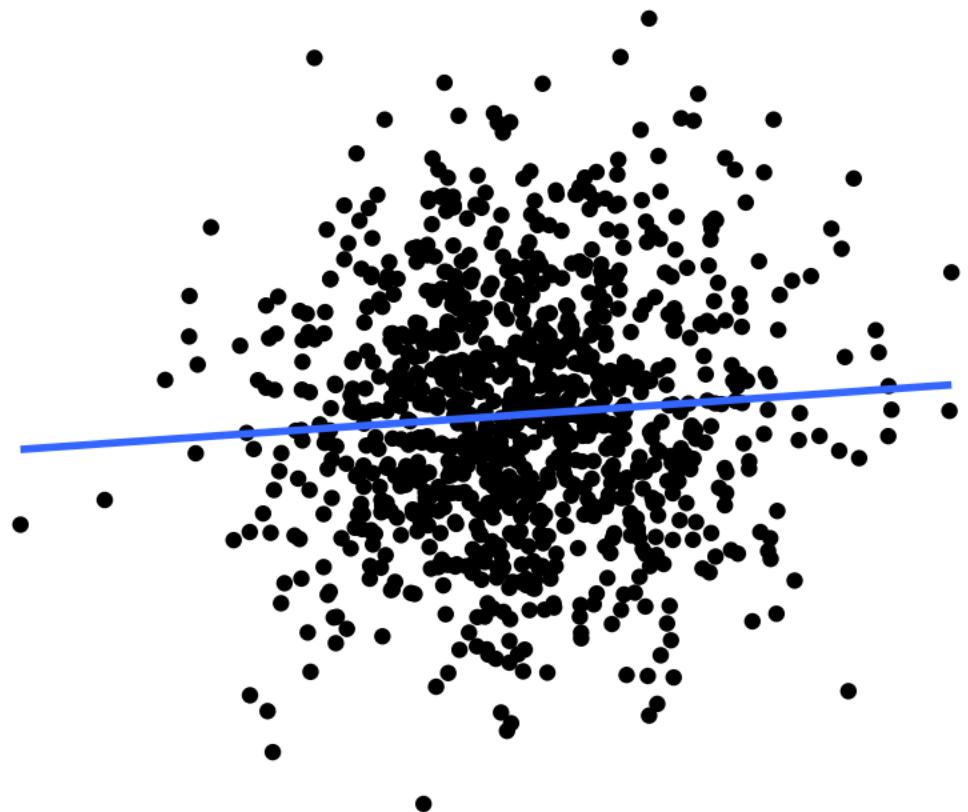
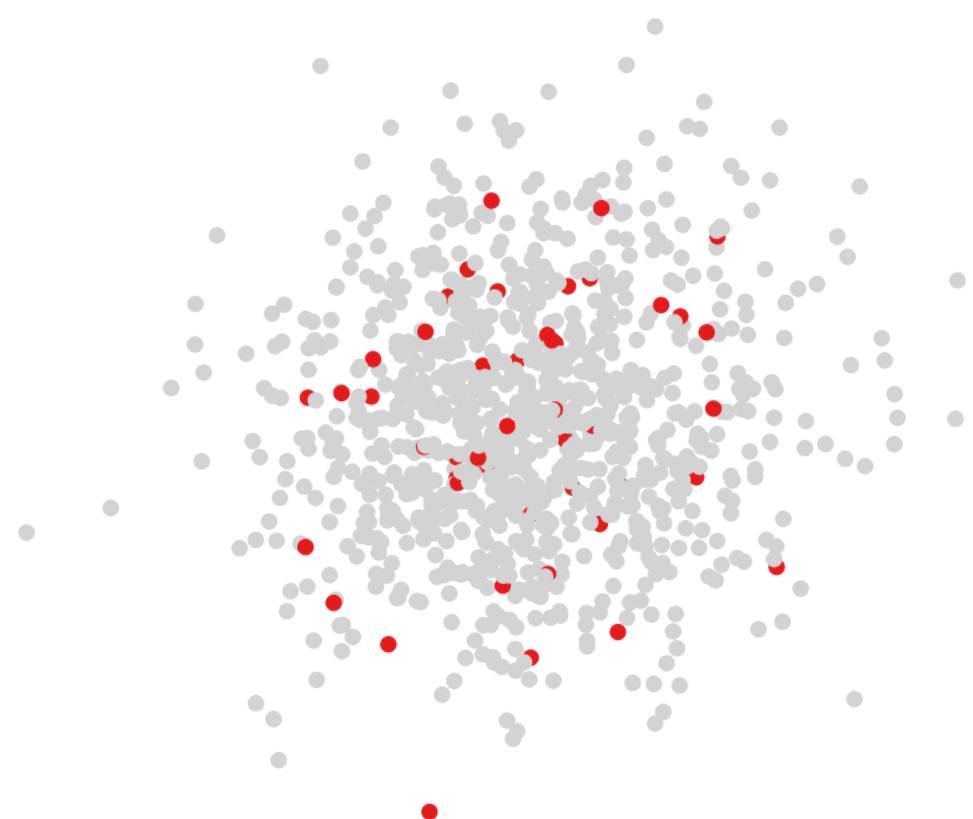


Ezeket a megfigyeléseket x és y értékek definiálják

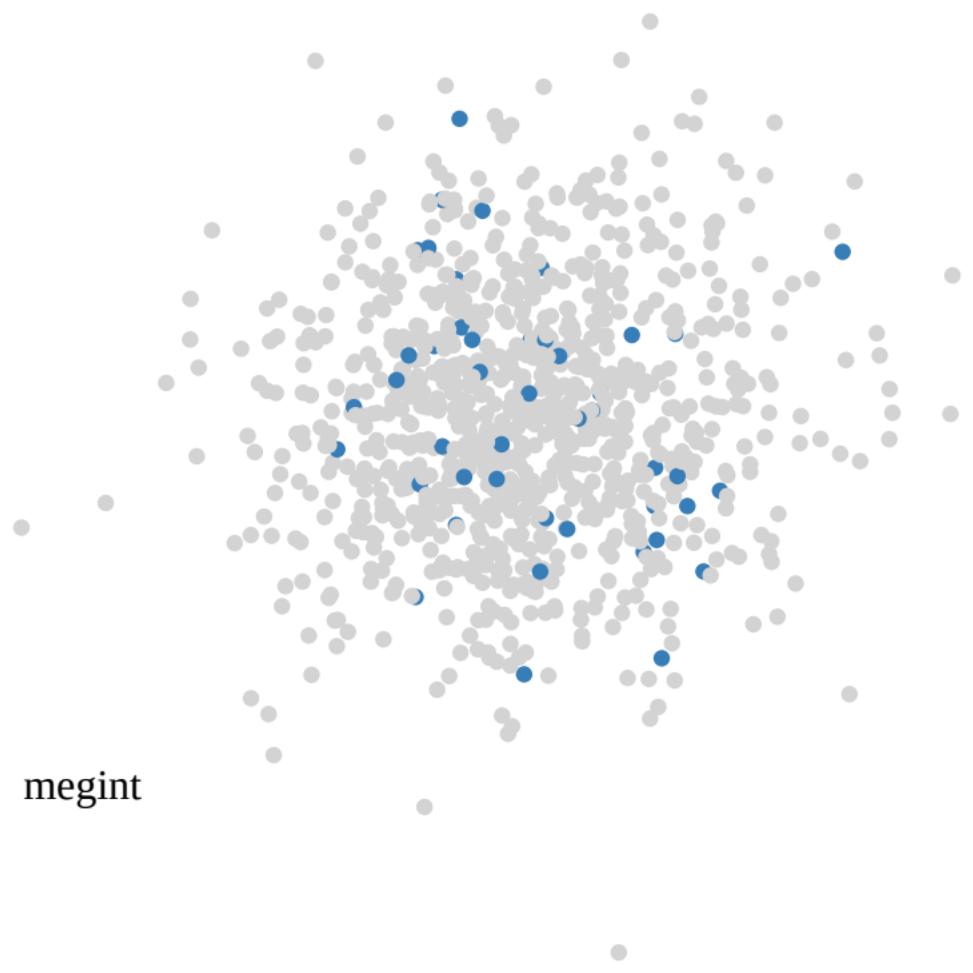


Kiszámolhatjuk, mennyivel nő az y , ha az x 1-el nő:
ez a szokásos vonal.

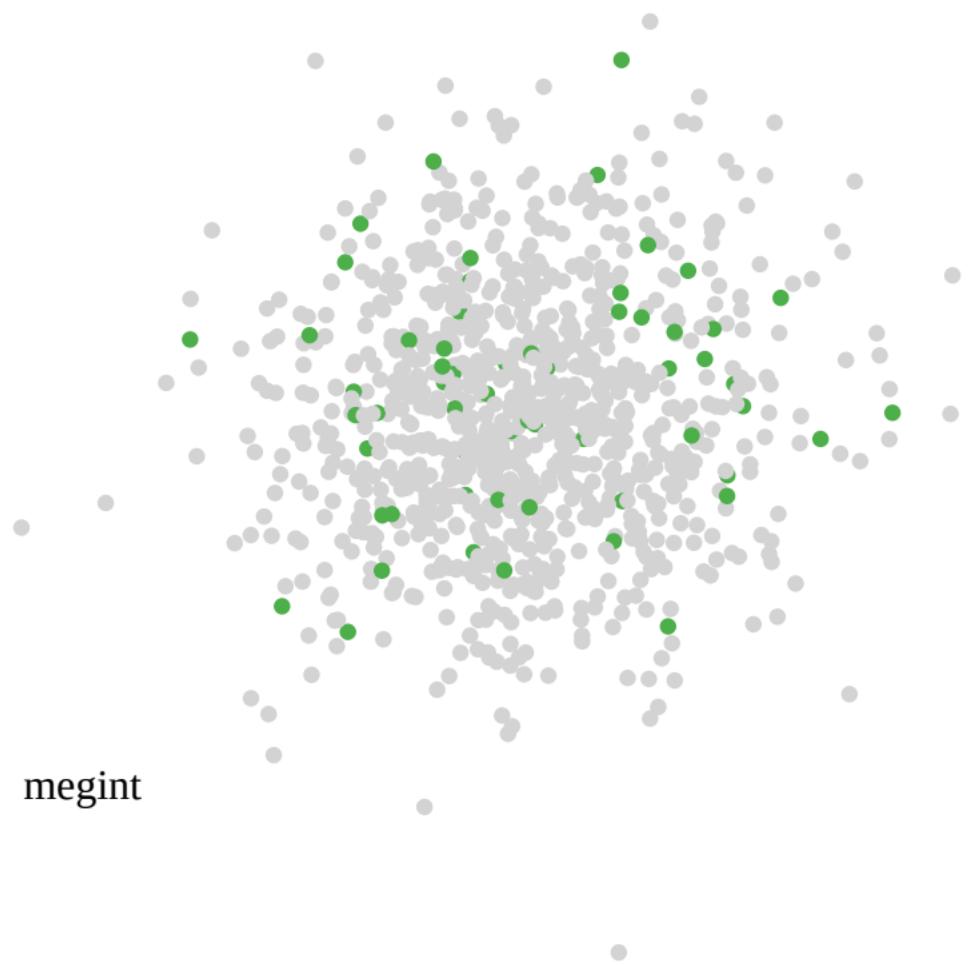


Vegyünk ki a megfigyelések közül ötvenet véletlenszerűen

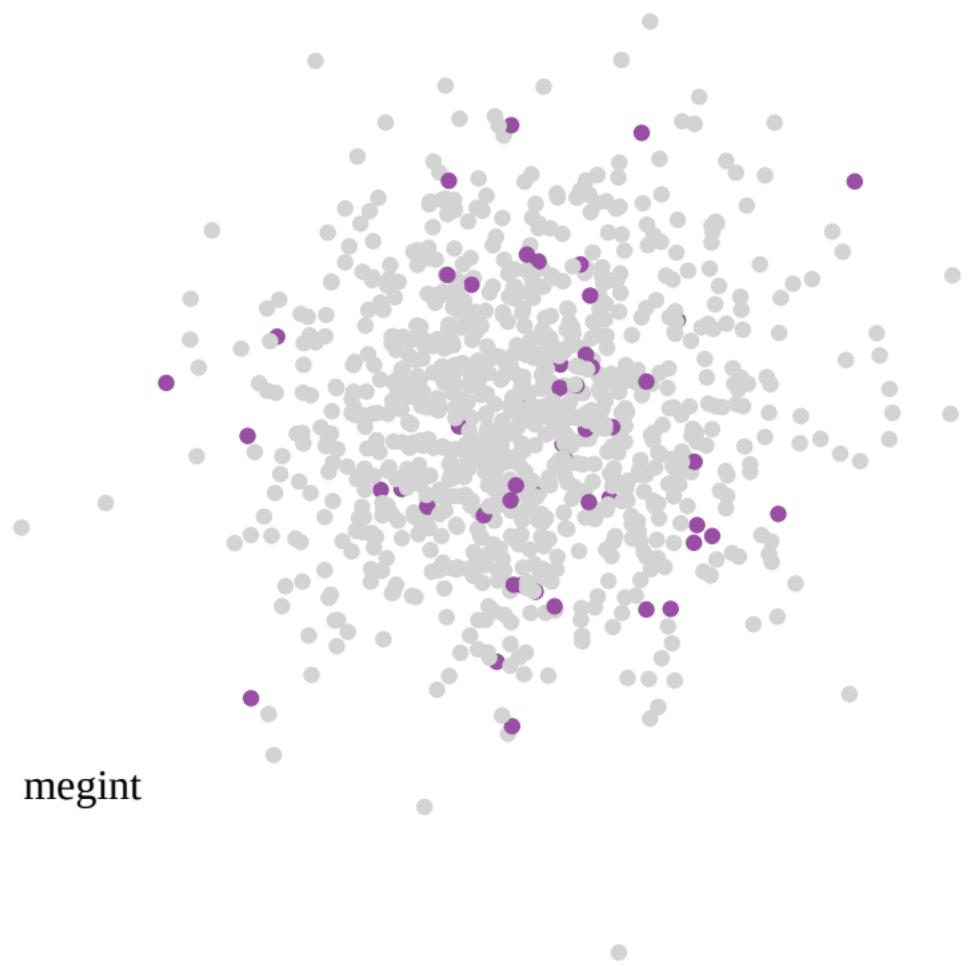
megint

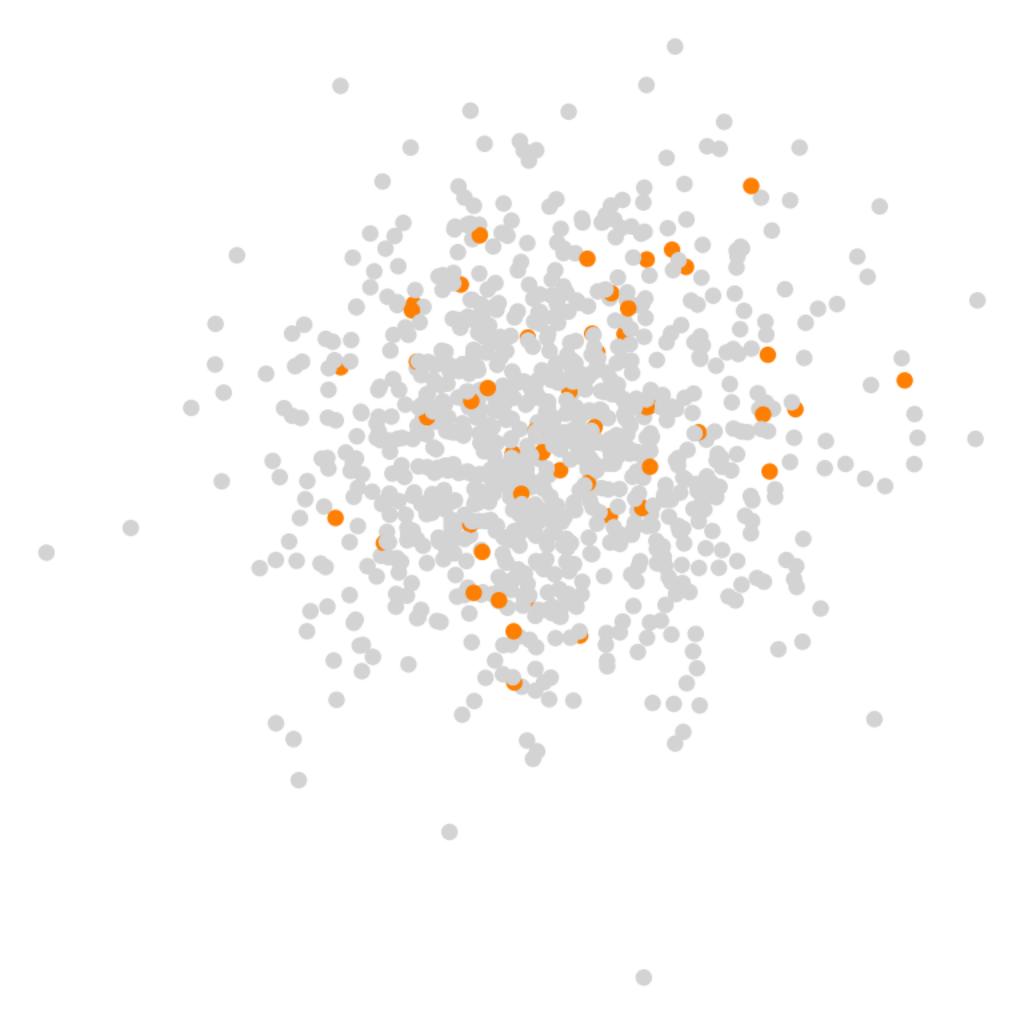


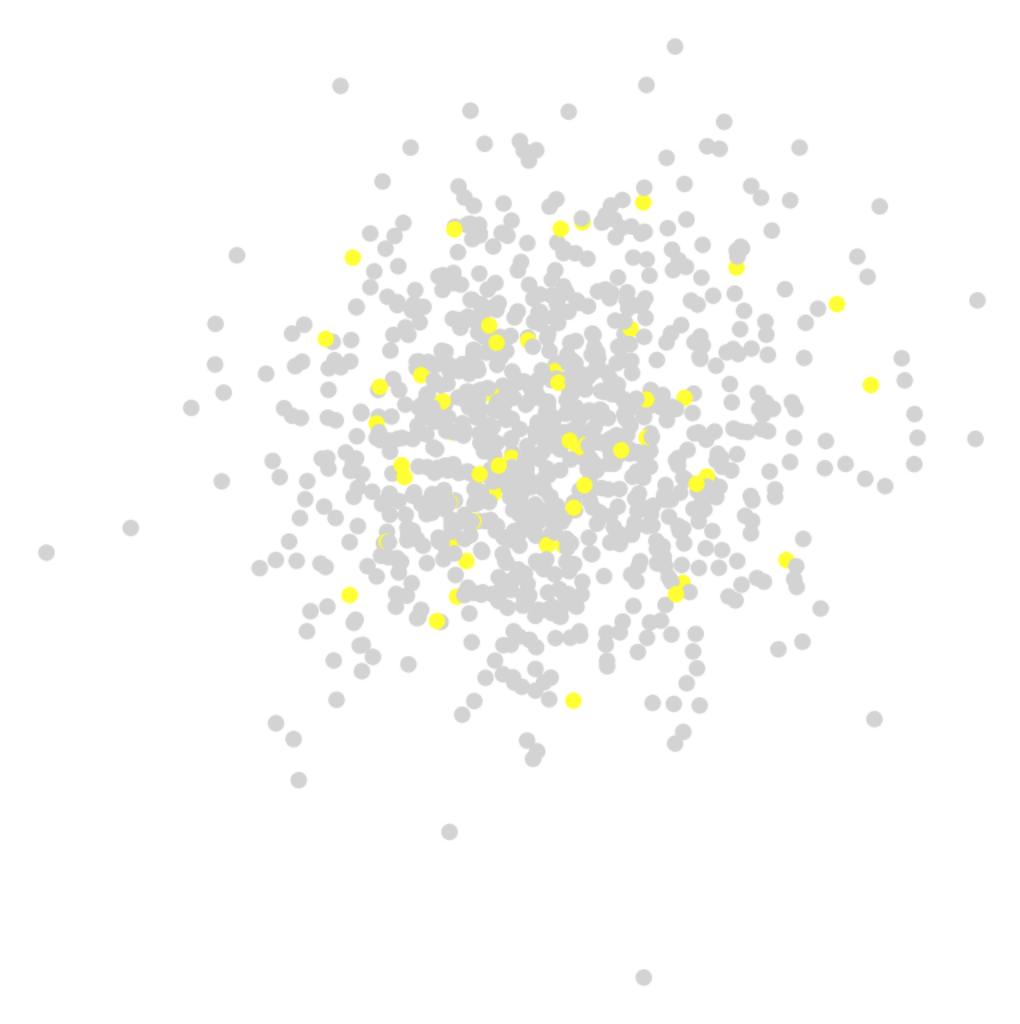
megint

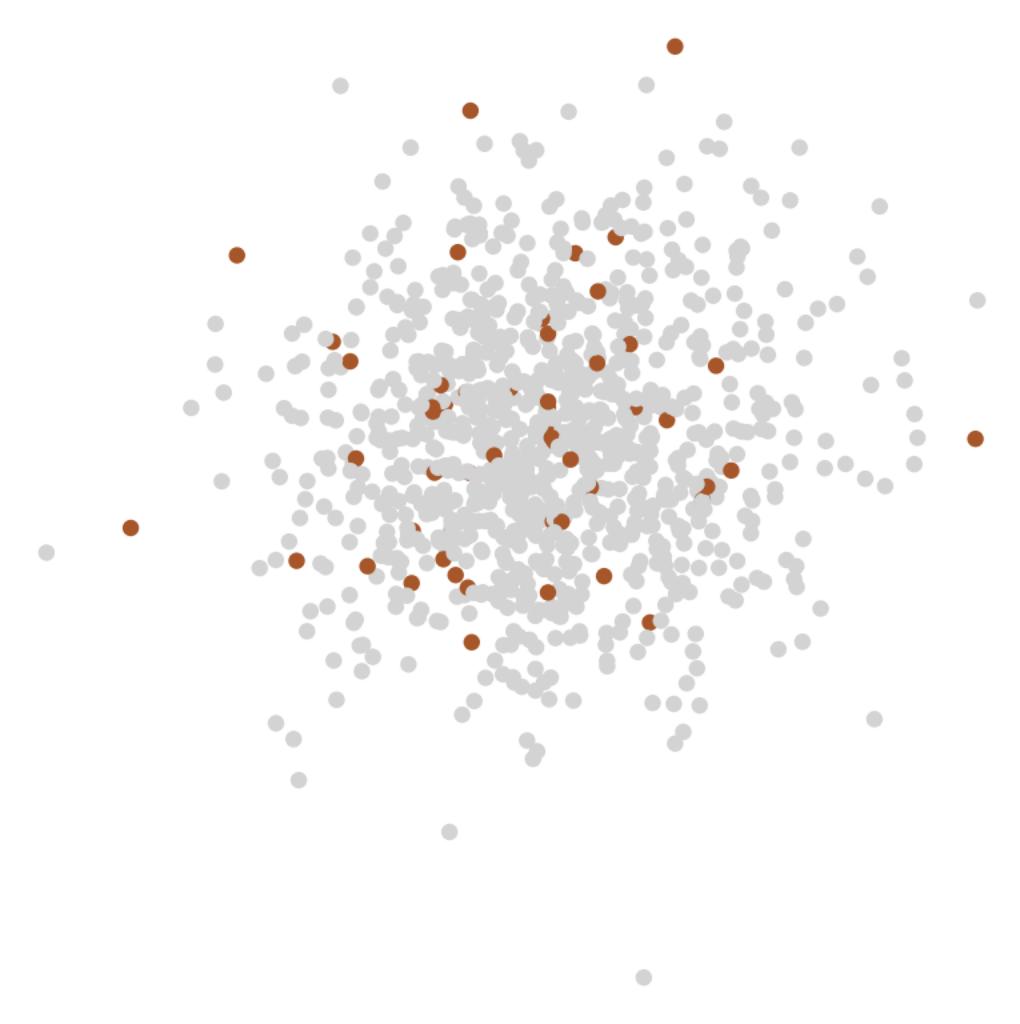


megint

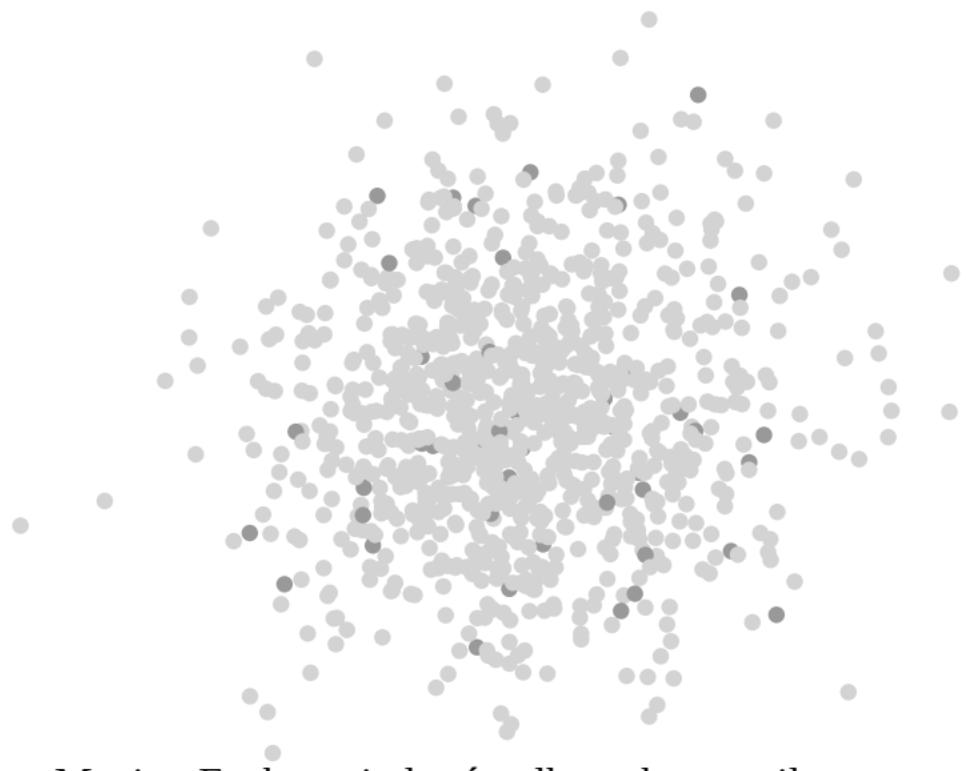




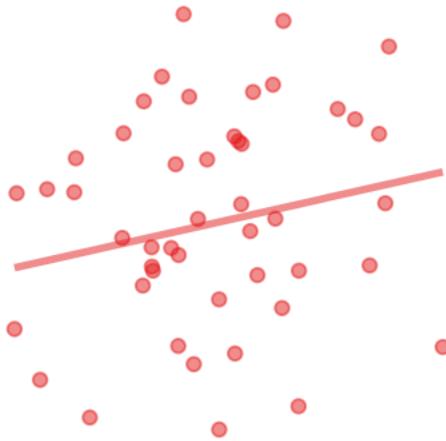




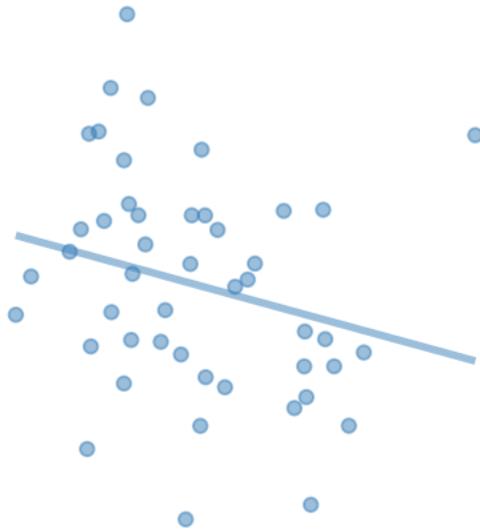




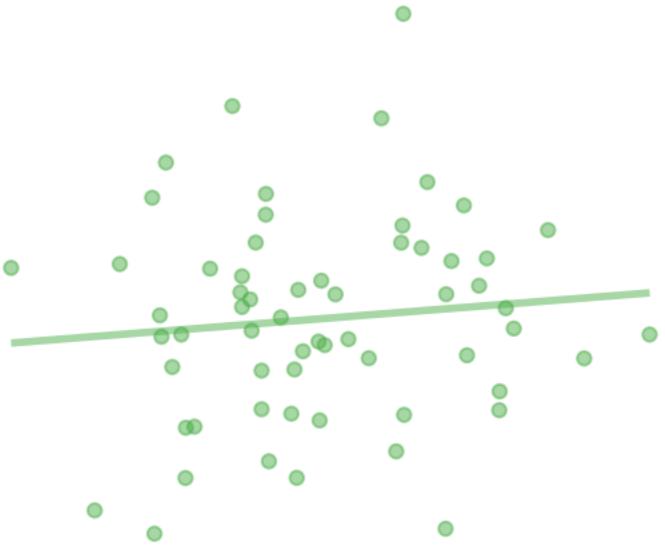
...Megint. Ezekre mind számolhatunk ugyanilyen regressziós vonalat.



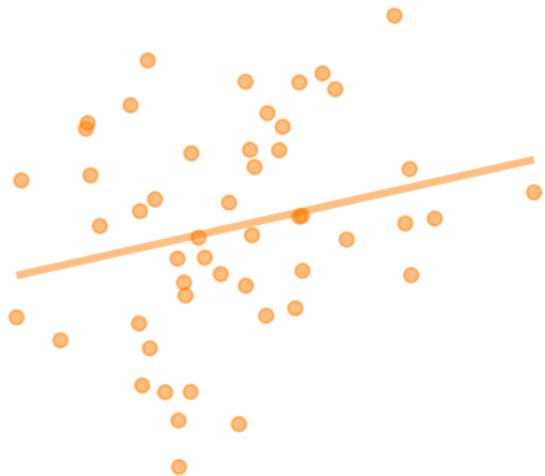
A kis mintára számolt vonal általában hasonlítani fog az összes adatra számolt vonalra



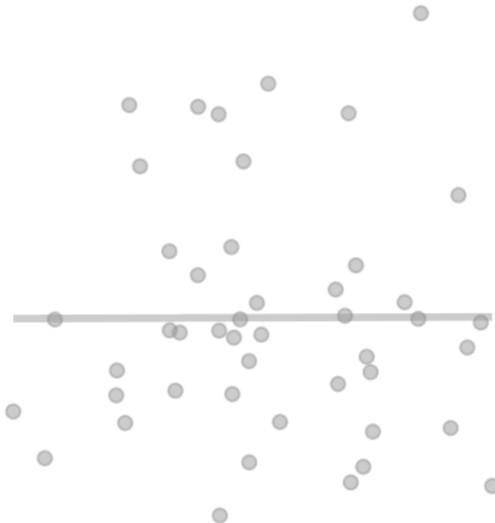
De nem mindig: véletlenszerűen húzgáltunk ki megfigyeléseket, nem mindig kapjuk vissza az eredeti összefüggést



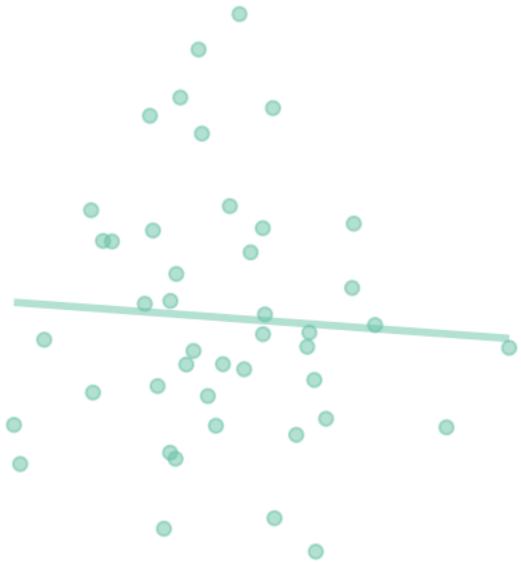
Néha gyengébbet kapunk



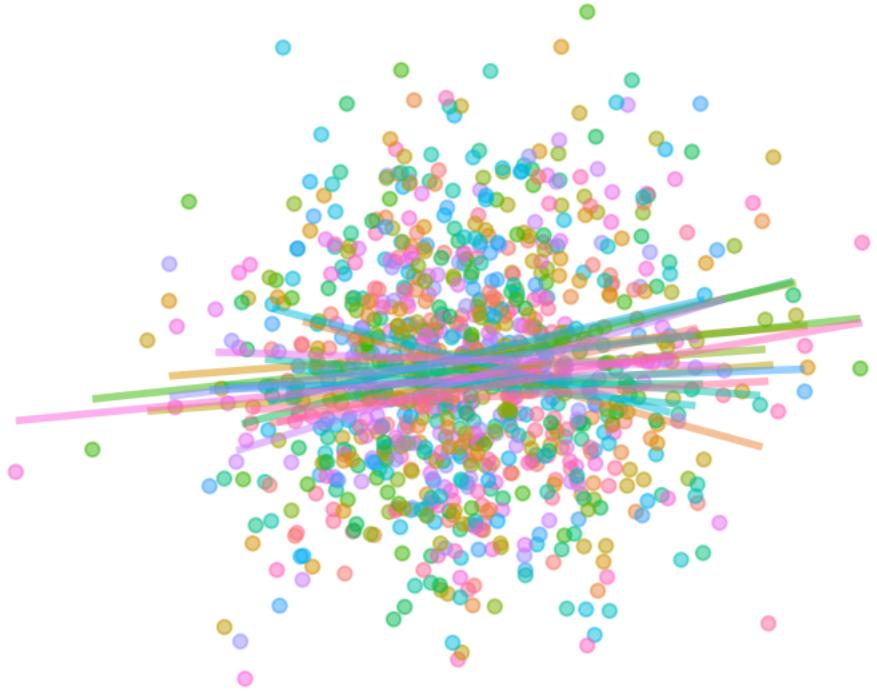
Néha erősebbet



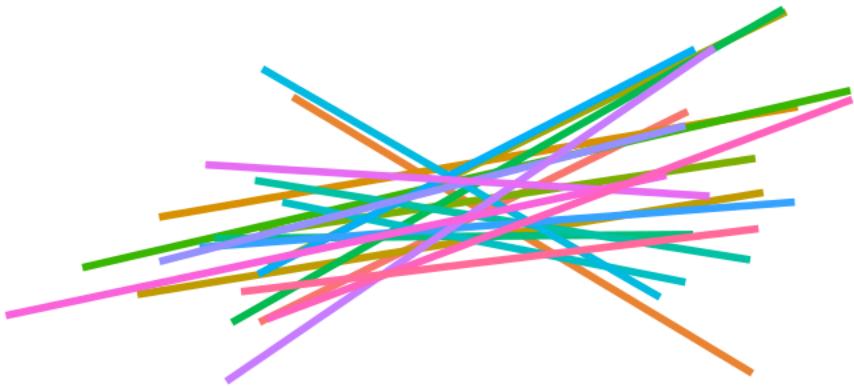
Néha nem találunk összefüggést a mintán belül



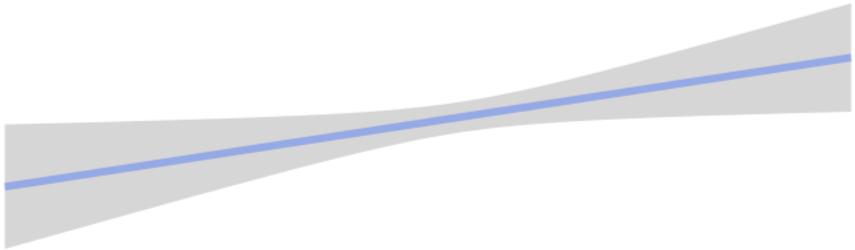
De a mintában negatív is lehet az összefüggés, hiába
tudjuk, hogy a nagy halmazban pozitív.
VÉLETLENEK



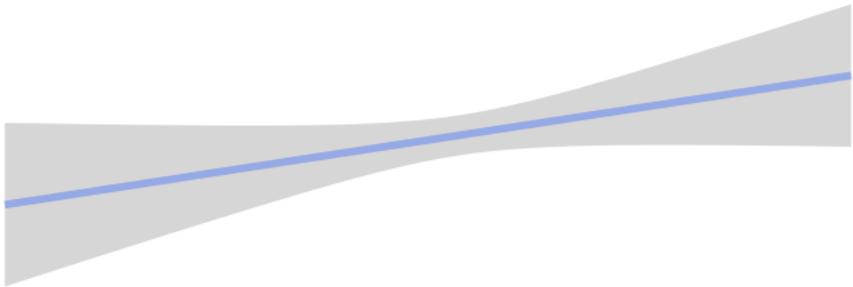
Itt van az összes kis vonal, amit húsz kis részmintára rajzoltunk. Ilyenek, olyanok, amolyanok.



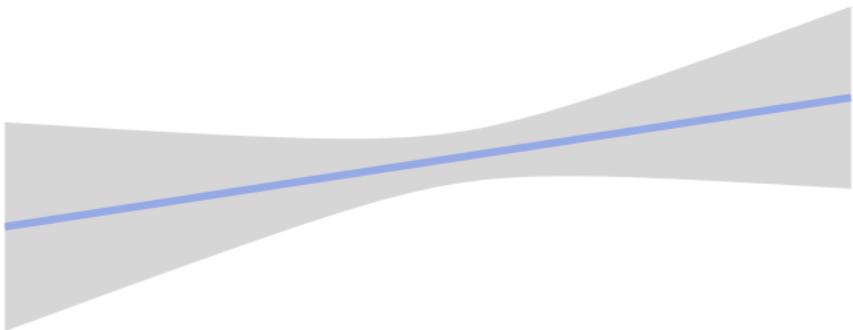
Itt vannak csak a vonalak, pontok nélkül. A többség hasonlít a nagy vonalra (enyhe pozitív viszony), de nem mindegyik. A 95% **konfidencia intervallum** az, ahová a húsz vonalból 19 esik.



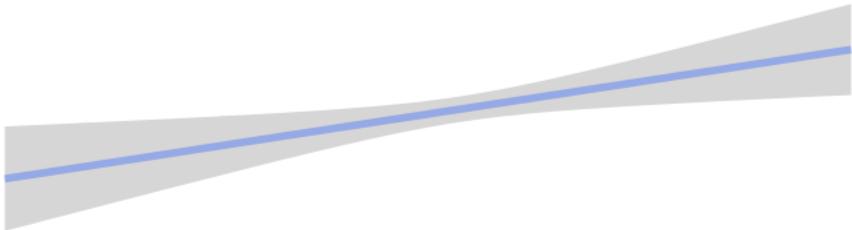
Ezt a zónát mutatja meg a szürke sáv, amit a ggplot rajzol.



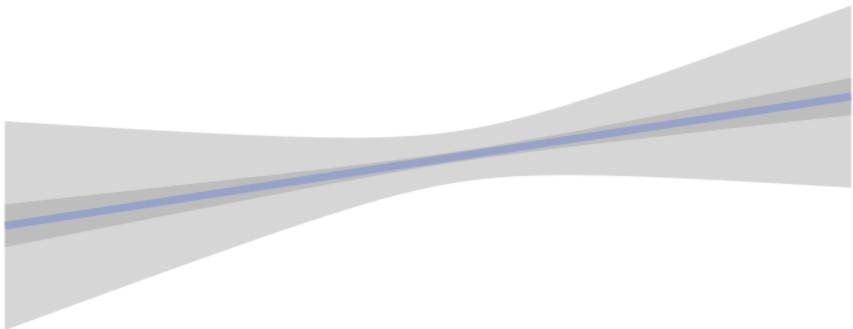
A konfidencia intervallumot mi definiáljuk. Lehet az, hogy szélesebb: a vonalak 95%-a helyett a vonalak 99%-a van benne.



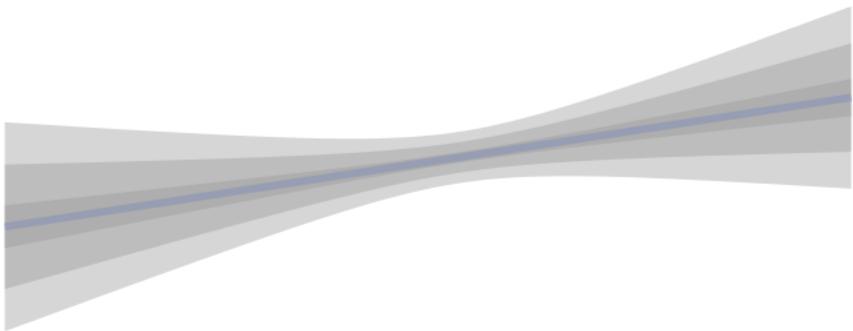
Vagy annál is szélesebb: a vonalak 99.9%-a benne van! Jó, most nem volt ennyi vonal, de lehetne.



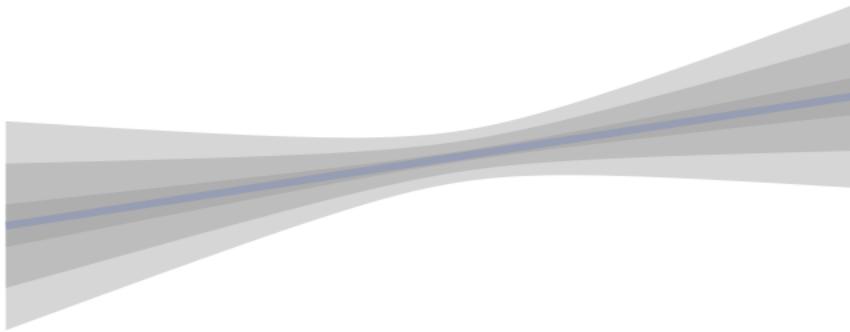
Vagy azt mondjuk, hogy nem kell ennyi vonal: elég,
ha a vonalak 90%-a benne van.



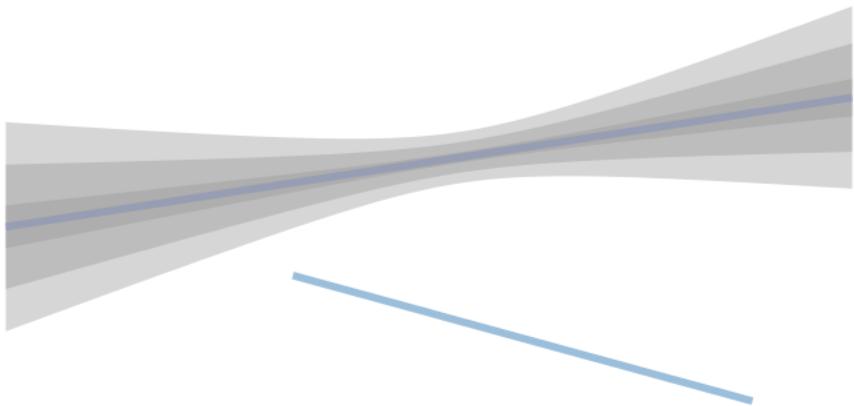
Ha 99.9%-ban szigorú vagy, és a zaj nagy része érdekel, akkor azt fogod mondani, hogy ez a mérés túl zajos. Ha csak 50%-ban, akkor azt fogod mondani, hogy egyértelmű pozitív viszony van.



A viselkedéstudományokban 95%-ot szoktunk használni. Ez egy szokás. Nincs értelme.



Ezt jelenti az, hogy $p < 0.05$. Az esetek 95%-ában azt várjuk, hogy pozitív viszonyt kapunk, tehát 95%-ban vagyunk biztosak abban, hogy a viszony tényleg pozitív.



De ez a **tehát** sok munkát végez. Lehet, hogy valójában teljesen máshol van a valódi viszony, és nem találtuk el egyáltalán. A valóságot nem lehet megismerni soha. Mérni lehet.