در این مقایسه ی سال تولد و infected_by به بیماری با توجه به شکل کشیده شد در حالت scatter ضمیمه شده در قسمت G ، اکثر داده ها در بازه ی 0 تا 50 قرار دارند و داده های outlier در بازه ی بالای 300 می باشند. در نتیجه دارای داده ی outlier می باشیم.

این داده ها لزوما داده ها بدی نیستند که حتما حذف شوند و بستگی به کاربرد و کاری که میخواهیم بکنیم دارد برای مثال ممکن است ما دنبال داده ها استثنایی می باشیم (مثلا مثل نمره های خارج العاده ی مدارس یا نخبه های یک کار خاص) که در این حالت ما دقیقا دنبال همین داده های outlier می باشیم ولی ممکن است کاربرد دیگری مد نظرمان باشد که بخواهیم به روی کل داده ها کاری بکنیم(مثل شیف دادن به نمره ی یک کلاس یا محاسبه میانگین درامد اکثر افراد جامعه) که در نتیجه باید این داده ها را نباید در نظر گرفت.