



Releasenieuws

Datadash 1.9

Elke dag speelt software een rol in de ondersteuning van uw bedrijfskritische processen. Belangrijke applicaties in uw organisatie staan vol met privacygevoelige persoonsgegevens van uw medewerkers, klanten of patiënten. En elke dag komen daar weer nieuwe gegevens bij.

Als u deze applicaties wil testen of de data uit de applicaties wil gebruiken om medewerkers te opleiden of onderzoek te doen, dan doet u dit het liefst met waarheidsgetrouwe gegevens. Maar dit mag niet volgens de Nederlandse en Europese privacywet.

Door de privacygevoelige gegevens te anonimiseren, zijn de gegevens niet meer te herleiden naar een individueel persoon. Zo voldoet u aan de wet- en regelgeving en heeft u een kwalitatief goede da-

taset om te gebruiken voor testen, onderzoek en ontwikkeling. Data volledig anonimiseren kan een flinke klus zijn en daar kan de anonimiseeroplossing Datadash u bij helpen.

Datadash 1.9

In Datadash 1.9 zijn een aantal nieuwe functionaliteiten beschikbaar voor het anonimiseren van persoonsgegevens.

- Preserve variability
- Chained shuffle

Wij wensen u veel plezier met de nieuwe release!

Team Datadash



1 Preserve variability

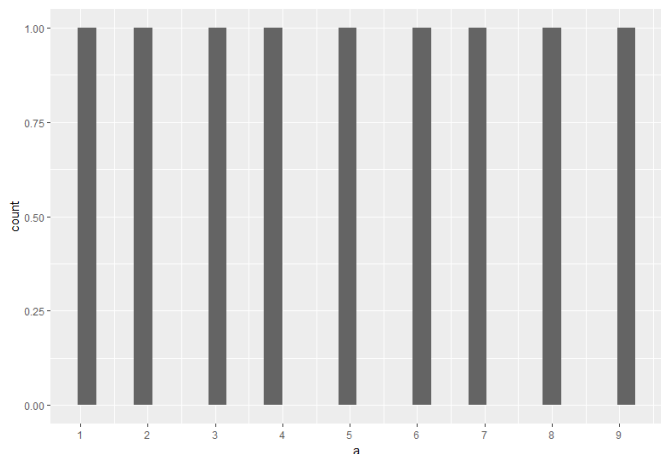
Preserve variability is een aanvullende functie op de shufflemethode in Datadash. Door deze functie in te zetten, blijft de verdeling van waarden voor en na de shuffle gelijk. Dit is met name waardevol als waarden vaker dan één keer terug komen in de kolom, of wanneer het noodzakelijk is dat de distributie van waarden gelijk blijft na de shuffle. Met de inzet van preserve variability is het niet mogelijk dat de oorspronkelijke waarde gelijk blijft na de shuffle.

Wanneer u de shuffle-anonimiseermethode inzet, kiest het onderliggende algoritme voor elke rij een willekeurige andere rijwaarde en vervangt de oorspronkelijke waarde. Deze randomisatie bevordert de anonimisering, maar geeft ook de mogelijkheid dat waarden vaker terugkomen na inzet van de shuffle-methode dan daarvoor.

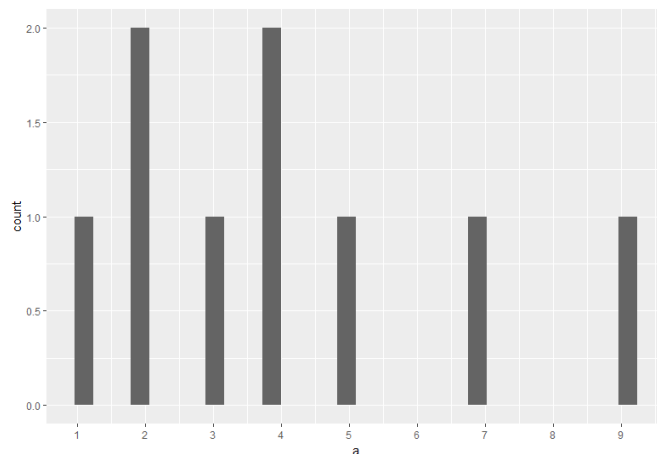
Voorbeeld

Figuur 1 laat de cijfers 1 tot en met 9 zien in een kolom. Dit zijn allemaal unieke waarden en elke waarde komt één keer voor in de originele kolom. We noemen dit uniforme data distributie. Na inzet van de shufflemethode op deze kolom, is het mogelijk dat de verdeling van de data eruit ziet zoals in figuur 2.

Door inzet van de preserve variability-functie blijft de verdeling van de data gelijk na de shuffle. Oftewel, als een waarde één keer voorkomt in de originele dataset, komt deze ook één keer voor in de post-shuffle dataset.



Figuur 1. Uniforme datadistributie voor shuffle



Figuur 2. Mogelijke datadistributie na shuffle



2 Chained shuffle

De chained shuffle is een nieuwe anonimiseer-methode binnen Datadash. Wanneer u de chained shuffle inzet, worden de datarelaties tussen tabellen binnen dezelfde database in stand gehouden. Het is bijvoorbeeld mogelijk om het patiëntnummer te randomiseren of te shuffelen, waardoor deze waar-de verandert, maar de relatie blijft bestaan over verschillende tabellen heen. Op deze manier zijn de gegevens te koppelen over verschillende databases heen, maar niet meer herleidbaar naar een fysiek persoon.

- We raden aan om de chained shuffle alleen te gebruiken op een kolom met unieke waarden, zoals de primary key waarde. Deze primaire sleutel identificeert ieder record in een tabel.
- Relaties tussen tabellen en relaties tussen de hoofdkolom en afhankelijke kolommen worden ondersteund, zolang de waarden in de hoofdkolom uniek zijn. Stel dat in de hoofdkolom één keer de naam 'Jan' voorkomt en in een afhankelijke kolom drie keer de naam 'Jan'. Door de chained shuffle wordt 'Jan' in de hoofdkolom aangepast naar 'Kees'. Vervolgens gebeurt dit in de afhankelijke kolom drie keer, zodat drie keer 'Jan', drie keer 'Kees' wordt.

Meer weten over Datadash?

In dit releasenieuws las u meer over de belangrijkste vernieuwingen in Datadash 1.9.

Datadash 2.0 verwachten wij in mei 2018 op te leveren. In deze omvangrijke release introduceren we een volledig nieuwe grafische interface en breiden we de ondersteuning op verschillende dataplatformen uit.

Heeft u vragen naar aanleiding van dit document? Of wilt u meer informatie over een van deze onderwerpen of de implementatie van Datadash? Neem contact op met Mark Jansen.



Mark Jansen

Productmanager Datadash
mark.jansen@pinkroccade.nl
06 – 4339 2983