

---

# IIC2026

## Visualización de Información

— Hernán F. Valdivieso López —  
(2024 - 1 / Clase 23)

---

# Temas de la clase - Fin del *framework*

1. Validación en visualización
  - a. Tipos de resolución.
  - b. Amenazas a la validez
2. Privacidad de datos

# Validación en Visualización

---

# Validación - Tipos de resolución

- Forma de enfrentar un caso de visualización.
- Veremos 2 enfoques donde podemos aplicar el Modelo Anidado de Munzner:
  - Enfoque centrado en problema o necesidad
  - Enfoque centrado en la técnica

# Validación - Tipos de resolución

## Enfoque centrado en problema o necesidad

- Este enfoque consiste en **diseñar y construir una herramienta visual** para solucionar un problema o necesidad.
- Existe una situación y datos específicos para el problema, y hay una o más tareas a ser resueltas por uno o varios usuarios.
- **El enfoque del curso está alineado a este tipo.** Solo se requiere aplicar el Modelo Anidado de Munzner tal como lo hemos visto en clases para desarrollar este enfoque.

# Validación - Tipos de resolución

Enfoque centrado en problema o necesidad



**Domain situation**

Situación de dominio



**Data/task abstraction**

Abstracción de datos y tareas



**Visual encoding/interaction idiom**

Codificación visual e interacciones



**Algorithm**

Algoritmo

# Validación - Tipos de resolución

## Enfoque centrado en la técnica

- Consiste en **inventar nuevos *idioms* o implementaciones** que aporten mejores resultados a dominios y abstracciones específicas.
- Hay que determinar a posteriori los datos y las situaciones posibles para aplicar las nuevas invenciones.
- Un ejemplo de resolución en este enfoque es **Mike Bostock y en todos los contribuidores de D3.js**. Su gran listado de utilidades existen con ese propósito: apoyar y mejorar la creación de *idioms* de visualización. A veces ni sabemos para qué casos aplicarlo.
- Podemos aplicar el modelo anidado de Munzner desde adentro hacia afuera.

# Validación - Tipos de resolución

Enfoque centrado en la técnica



**Domain situation**

Situación de dominio



**Data/task abstraction**

Abstracción de datos y tareas



**Visual encoding/interaction idiom**

Codificación visual e interacciones



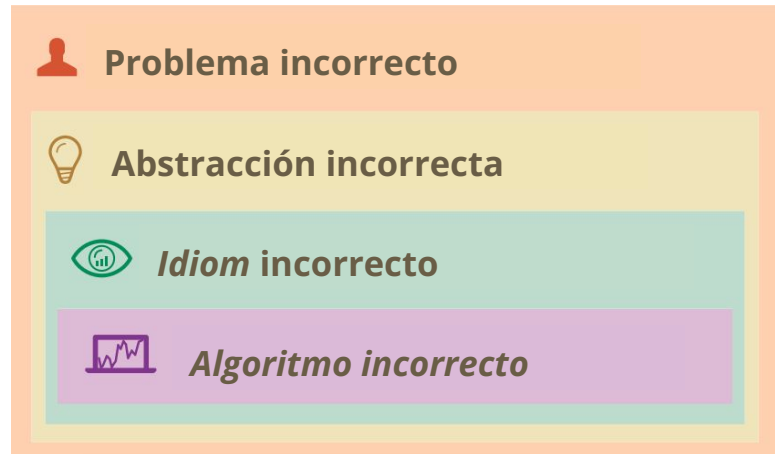
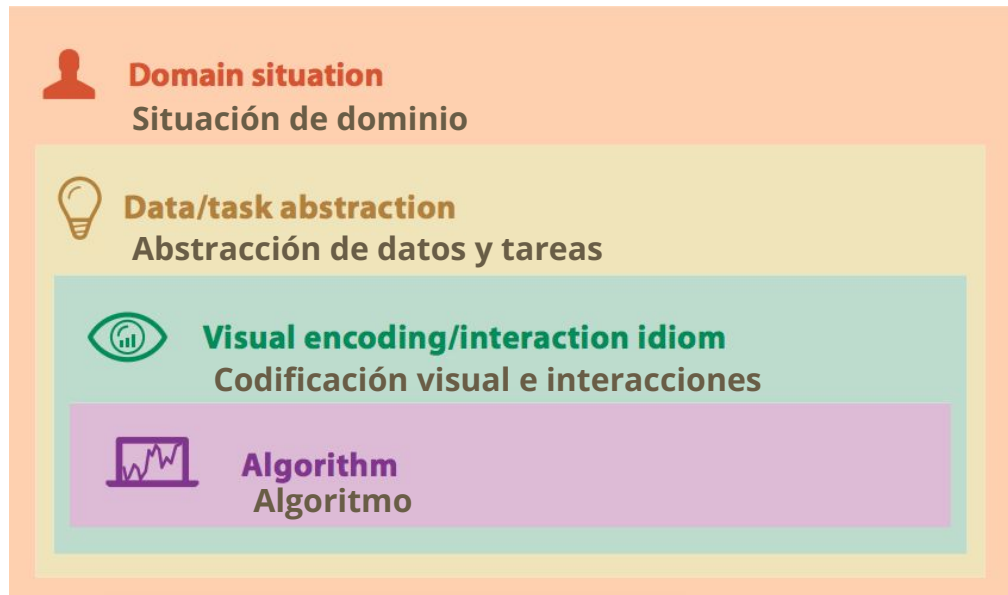
**Algorithm**

Algoritmo



# Validación - Amenazas a la validez


Las validaciones son razones fundamentales de por qué pudieron generarse decisiones erróneas en un proceso de confeccionar una visualización.



# Validación - Amenazas a la validez

Para cada problema, disponemos de diferentes herramientas para mitigarlas. Estas serán nuestras validaciones.

Tenemos 2 tipos de validaciones:





- Inmediatas   → Antes de pasar al siguiente nivel.
- "Río abajo"   → Después de pasar por los niveles interiores.

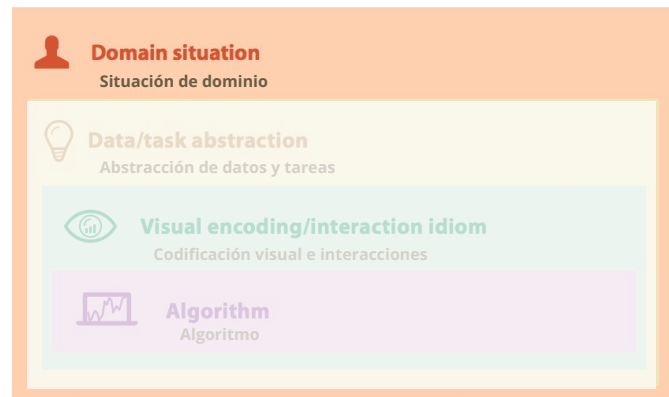
# Validación - Amenazas a la validez

## Situación de dominio

✗ Amenaza ✗ → Problema incorrecto.

La Identificación de necesidades no fue acertada y la audiencia objetivo no tienen los problemas que se buscan solucionar.

- Validación inmediata  
  - Realizar entrevistas a usuarios objetivos, o realizar estudios de campo de observación. En resumen, **conocer al usuario objetivo**
- Validación "Río abajo"  
  - Medir cómo se ha adoptado la herramienta en el contexto que fue concebida después de un periodo de tiempo.



# Validación - Amenazas a la validez

## Abstracción de datos y tareas

✗ Amenaza ✗ → Abstracción incorrecta.

Lo identificado y seleccionado como tareas y datos a usar no resuelven el problema del usuario.

- Validación inmediata ✓ ✓
  - No se definió.
- Validación "Río abajo" ✓ 💧
  - Poner a prueba la herramienta con usuarios objetivos y recolectar anécdotas.
  - Estudio de campo de herramienta en contexto de uso.
  - En resumen, **que los usuarios usen las visualizaciones.**



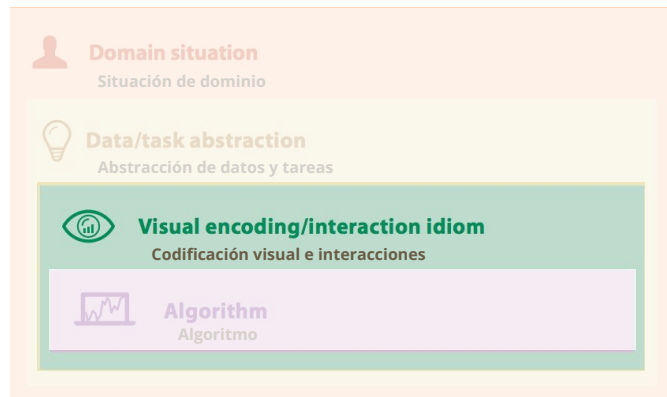
# Validación - Amenazas a la validez

## Codificación visual e interacciones

✗ Amenaza ✗ → Gráfico incorrecto.

El *idiom* seleccionado no funciona para las abstracciones realizadas.

- Validación inmediata ✓ ✓
  - Justificar estas decisiones realizadas en base a principios de percepción y cognición.
- Validación "Río abajo" ✓ 💧
  - Experimentos de laboratorio para medir impacto cuantitativo y cualitativo de decisiones o gráficos aislados.
  - Presentación y discusión cualitativa de imágenes o videos de propuesta de herramienta. En otras palabras, **realizar una exposición de la solución.**







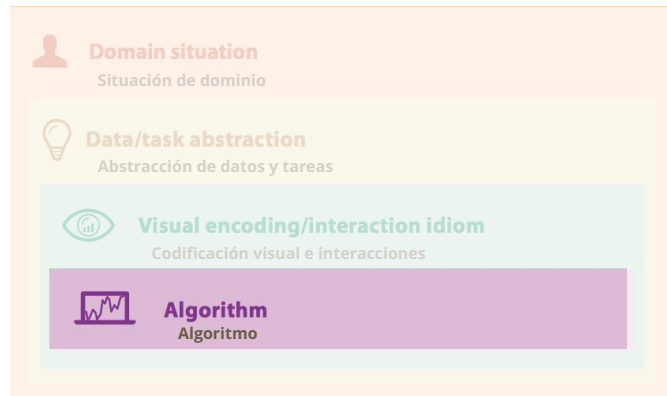
# Validación - Amenazas a la validez

## Algoritmo

✗ Amenaza ✗ → Algoritmo incorrecto.

Código sea muy lento o subóptimo para las decisiones de codificación elegidas.

- Validación inmediata  
  - Análisis de complejidad computacional de implementación.
- Validación "Río abajo"  
  - Medir tiempo de carga y respuesta en la práctica.
  - Asegurarse que la visualización cargue y funcione como corresponde técnicamente.



# Privacidad de datos

---

# Motivación

- A veces nos toca trabajar con datos que tienen información sensible para el usuario
  - Enfermedad.
  - Sueldo.
  - Deudas.
  - Entre otros (dependerá del contexto).
- 🤔 ¿Cómo aseguramos que no puedan identificar a nuestros usuarios?



# Motivación

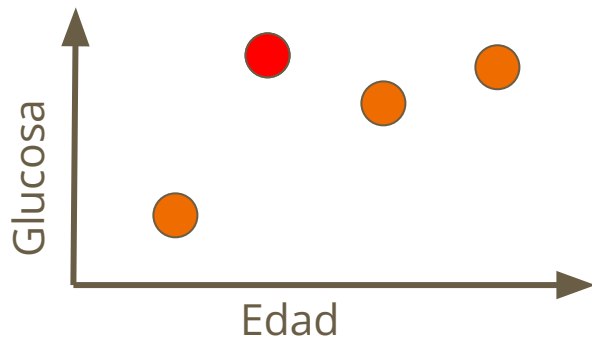
- Los datos se pueden catalogar en diferentes tipos (no excluyentes):
  - **Identificadores.** Información que permite identificar al individuo detrás de los datos como el rut, mail uc, nombre completo.
  - **Potenciales identificadores.** Información general del individuo que eventualmente pueden servir para identificar al individuo como la edad.
  - **No identificadores.** Información que por sí sola no permite identificar al individuo.
  - **Sensibles.** Información privada que se espera desvincular del individuo.

Name	Age	Glucose	Diabetes
Alice Smith	25	237	Yes
Bob Taylor	34	186	Yes
Frank Jones	33	165	Yes
Ivy Smith	42	80	No
Jim Davies	12	190	Yes
...	...	...	...

# Motivación - ¿Y qué pasa en la visualización?

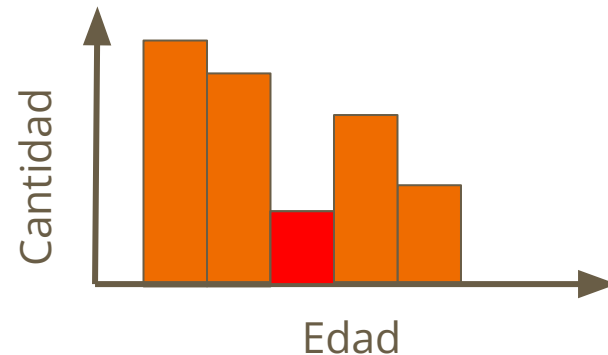
Name	Age	Glucose	Diabetes
Alice Smith	25	237	Yes
Bob Taylor	34	186	Yes
Frank Jones	33	165	Yes
Ivy Smith	42	80	No
Jim Davies	12	190	Yes
...	...	...	...

Scatterplot Glucosa VS Edad



Si conozco la edad de  
Alicia, ahora sé que  
tiene diabetes

Histograma por edad de  
personas diabéticas



# ¿Qué podemos hacer con los datos?

- **Objetivo:** Asegurar que no sea posible vincular un dato sensible al usuario.
- **Acciones:**
  - Se evita mantener identificadores en la visualización.
  - **Intentar generalizar o modificar posibles identificadores.**

# ¿Qué podemos hacer con los datos?

- Se han creado diferentes nociones de privacidad:
  - ***k-anonymity*** → Para cada fila, hay al menos ***K-1*** filas que tienen los mismos datos no identificadores o potenciales identificadores.
  - ***l-diversity*** → Para cada grupo de datos iguales, hay al menos ***L*** datos sensible diferentes
  - ***t-Closeness*** → Para cada grupo, la distancia entre la distribución de datos sensibles no supera el valor ***T***.

**Veamos un ejemplo de estas 2 nociones**

# ¿Qué podemos hacer con los datos?

- Se han creado diferentes nociones de privacidad:
  - ***k-anonymity*** → Para cada fila, hay al menos k-1 filas que tienen los mismos datos no identificadores o potenciales identificadores.

Company	Position	Nationality	Zip	Age	Disease
Alpha	Director	Japanese	10001	32	Galactosemia
Beta	Manager	Indian	11049	53	Cancer
Gamma	Associate	American	10011	38	Galactosemia
Beta	Manager	Russian	10004	43	Fatty Liver
Alpha	Manager	Japanese	10014	48	Hepatitis B
Delta	Consultan	Indian	10017	34	Galactosemia
Gamma	Associate	American	11042	57	Hepatitis B
Delta	Manager	American	10007	42	Hepatitis B
Gamma	Director	Japanese	11043	51	Galactosemia
Beta	Manager	Russian	10009	35	Galactosemia
Delta	Associate	Indian	10019	42	Fatty Liver
Gamma	Manager	Japanese	11047	63	Fatty Liver



Company	Position	Nationality	Zip	Age	Disease
*	*	*	100**	<40	Galactosemia
*	*	*	100**	<40	Galactosemia
*	*	*	100**	<40	Galactosemia
*	*	*	100**	<40	Galactosemia
*	*	*	110**	>=50	Galactosemia
*	*	*	110**	>=50	Cancer
*	*	*	110**	>=50	Hepatitis B
*	*	*	110**	>=50	Fatty Liver
*	*	*	100**	4*	Hepatitis B
*	*	*	100**	4*	Fatty Liver
*	*	*	100**	4*	Fatty Liver
*	*	*	100**	4*	Hepatitis B

*4-anonymity*

# Problemas con *k-anonymity*

- Sí sé que Hernán tiene 38 años y vive en el código postal 10011, puedo descifrar fácilmente que tiene galactosemia (**ataque por homogeneidad**).

Company	Position	Nationality	Zip	Age	Disease
*	*	*	100**	<40	Galactosemia
*	*	*	100**	<40	Galactosemia
*	*	*	100**	<40	Galactosemia
*	*	*	100**	<40	Galactosemia
*	*	*	110**	>=50	Galactosemia
*	*	*	110**	>=50	Cancer
*	*	*	110**	>=50	Hepatitis B
*	*	*	110**	>=50	Fatty Liver
*	*	*	100**	4*	Hepatitis B
*	*	*	100**	4*	Fatty Liver
*	*	*	100**	4*	Fatty Liver
*	*	*	100**	4*	Hepatitis B

# Problemas con *k-anonymity*

- Sí sé que Daniela tiene 49 años, trabaja en la compañía Delta, vive en 10019 y además sé que en dicha compañía están todos vacunados contra la hepatitis B, puedo deducir que Daniela tiene hígado graso (**ataque por información previa**).

Company	Position	Nationality	Zip	Age	Disease
*	*	*	100**	<40	Galactosemia
*	*	*	100**	<40	Galactosemia
*	*	*	100**	<40	Galactosemia
*	*	*	100**	<40	Galactosemia
*	*	*	110**	>=50	Galactosemia
*	*	*	110**	>=50	Cancer
*	*	*	110**	>=50	Hepatitis B
*	*	*	110**	>=50	Fatty Liver
*	*	*	100**	4*	Hepatitis B
*	*	*	100**	4*	Fatty Liver
*	*	*	100**	4*	Fatty Liver
*	*	*	100**	4*	Hepatitis B

Los dos problemas anteriores se abordan con *l-diversity*.

# ¿Qué podemos hacer con los datos?

- Se han creado diferentes nociones de privacidad:
  - l-diversity*** → Para cada grupo de datos iguales, hay al menos L datos sensible diferentes

Company	Position	Nationality	Zip	Age	Disease
Alpha	Director	Japanese	10001	32	Galactosemia
Beta	Manager	Indian	11049	53	Cancer
Gamma	Associate	American	10011	38	Galactosemia
Beta	Manager	Russian	10004	43	Fatty Liver
Alpha	Manager	Japanese	10014	48	Hepatitis B
Delta	Consultan	Indian	10017	34	Galactosemia
Gamma	Associate	American	11042	57	Hepatitis B
Delta	Manager	American	10007	42	Hepatitis B
Gamma	Director	Japanese	11043	51	Galactosemia
Beta	Manager	Russian	10009	35	Galactosemia
Delta	Associate	Indian	10019	42	Fatty Liver
Gamma	Manager	Japanese	11047	63	Fatty Liver



Company	Position	Nationality	Zip	Age	Disease
*	*	*	1000*	<50	Galactosemia
*	*	*	1000*	<50	Fatty Liver
*	*	*	1000*	<50	Hepatitis B
*	*	*	1000*	<50	Galactosemia
*	*	*	1104*	>=50	Hepatitis B
*	*	*	1104*	>=50	Galactosemia
*	*	*	1104*	>=50	Fatty Liver
*	*	*	1104*	>=50	Cancer
*	*	*	1001*	<50	Galactosemia
*	*	*	1001*	<50	Hepatitis B
*	*	*	1001*	<50	Galactosemia
*	*	*	1001*	<50	Fatty Liver

*3-diversity*



# Métodos para intentar asegurar privacidad

- Generalización
- Filtrado
- Aplicar una máscara
- Reemplazar valores
  - Vecinos cercanos
  - Ruido

# Métodos para intentar asegurar privacidad

- **Generalización**
- Filtrado
- Aplicar una máscara
- Reemplazar valores
  - Vecinos cercanos
  - Ruido

Edad	Edad Generalizada
11	< 50
69	>= 50
42	< 50
5	< 50
69	>= 50

# Métodos para intentar asegurar privacidad

- Generalización
- **Filtrado**
- Aplicar una máscara
- Reemplazar valores
  - Vecinos cercanos
  - Ruido

Edad	Edad Filtrada
	69
69	42
42	69
	
69	

# Métodos para intentar asegurar privacidad

- Generalización
- Filtrado
- **Aplicar una máscara**
- Reemplazar valores
  - Vecinos cercanos
  - Ruido

Edad	Edad con máscara
11	**
69	6&
41	4*
5	++
69	6&

# Métodos para intentar asegurar privacidad

- Generalización
- Filtrado
- Aplicar una máscara
- Reemplazar valores
  - **Vecinos cercanos**
  - Ruido

Edad	Reemplazar Valores <i>Promedio Vecinos cercanos</i>
11 (cercano al 5)	<b>8</b>
69 (cercano al 68)	68,5
41	41
5 (cercano al 11)	<b>8</b>
68 (cercano al 69)	68,5

# Métodos para intentar asegurar privacidad

- Generalización
- Filtrado
- Aplicar una máscara
- Reemplazar valores
  - Vecinos cercanos
  - **Ruido**

Edad	Reemplazar Valores Ruido
11	12
69	68
41	40
5	5
69	70

# Decisiones en visualización para facilitar la privacidad

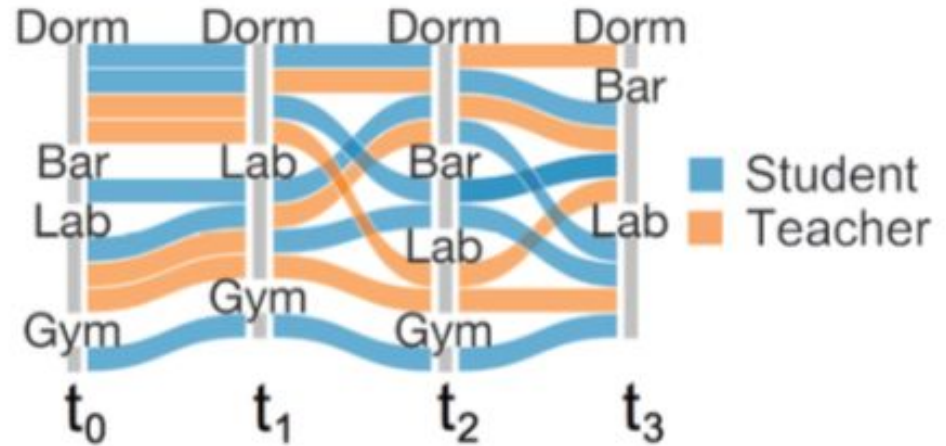
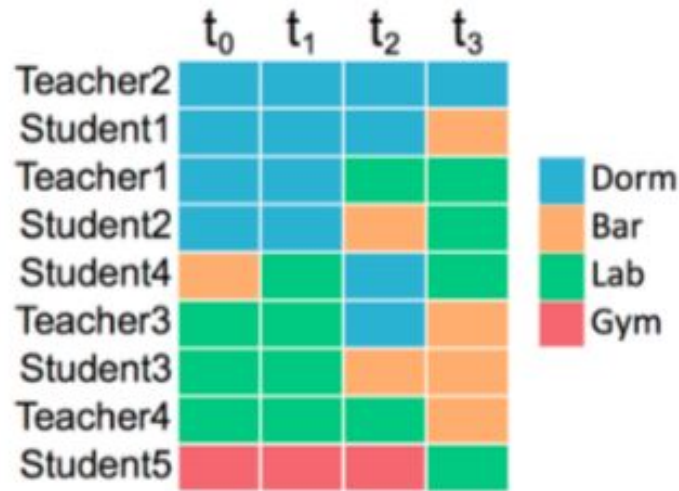
Hay casos donde no se puede modificar el *dataset* y es bueno pensar en formas que la visualización modifique *inline* la información antes de comenzar a dibujar.

Algunas decisiones implican:

1. Utilizar visualizaciones que dificultan la identificación total de un ítem.
2. Eliminación y/o agrupación de datos.
3. Aplicar ruido en función de vecinos cercanos o de forma probabilística.
4. Superposición de enlaces en el caso de grafo.

# Decisiones en visualización para facilitar la privacidad

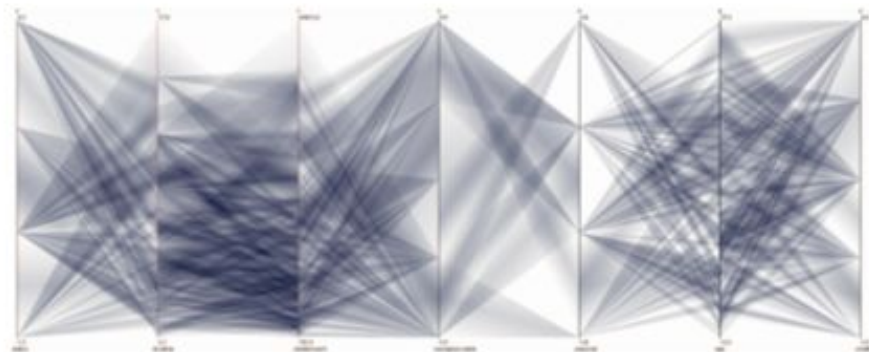
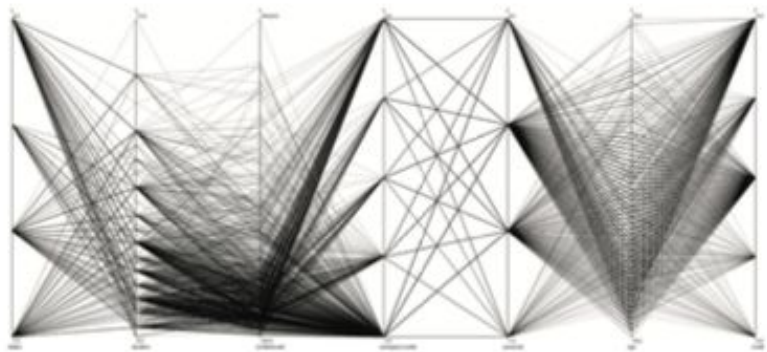
Utilizar visualizaciones que dificultan la identificación total de un ítem.





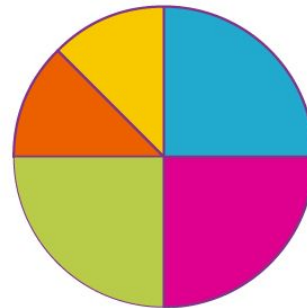
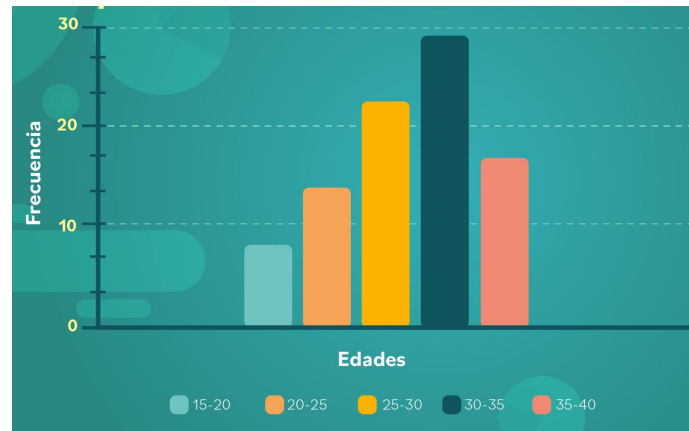
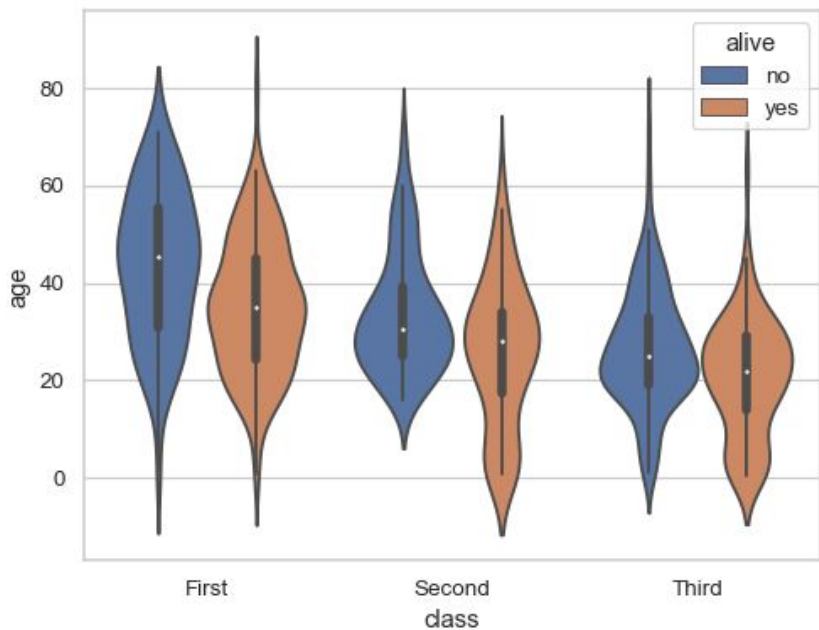
# Decisiones en visualización para facilitar la privacidad

Utilizar visualizaciones que dificultan la identificación total de un ítem.



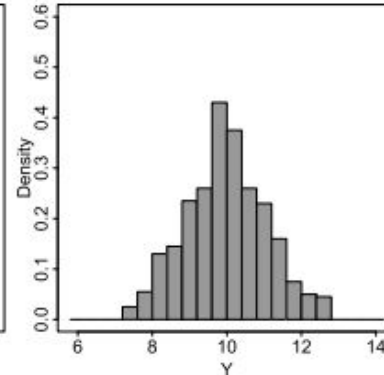
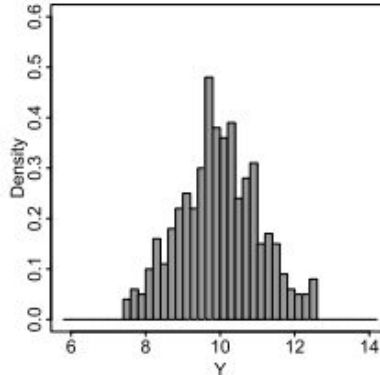
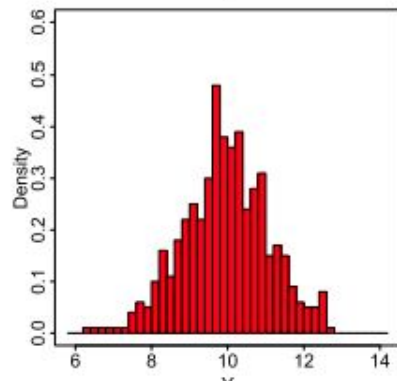
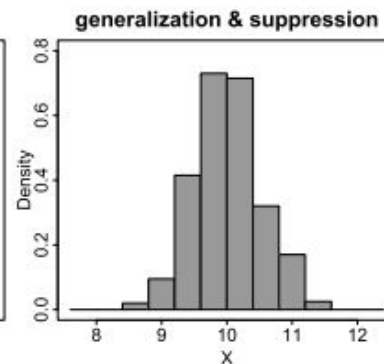
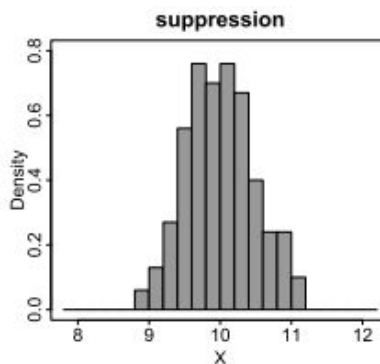
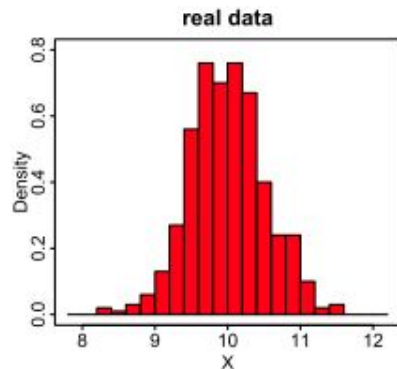
# Decisiones en visualización para facilitar la privacidad

Utilizar visualizaciones que dificultan la identificación total de un ítem.



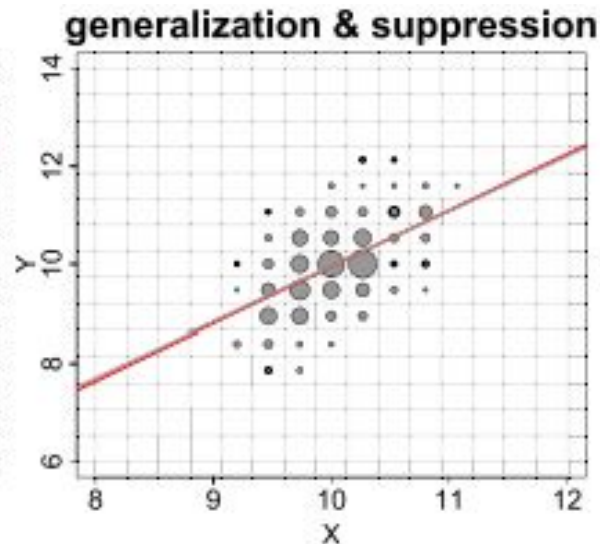
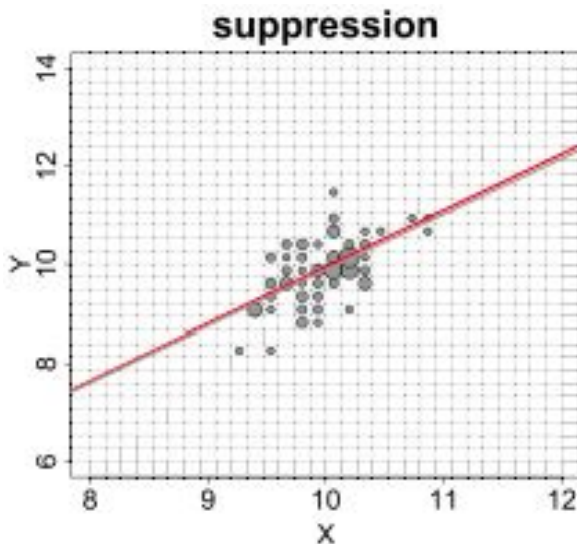
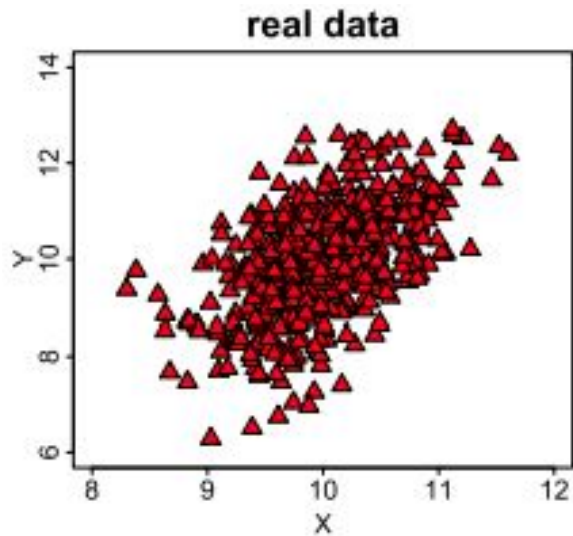
# Decisiones en visualización para facilitar la privacidad

Eliminación y/o agrupación de datos.



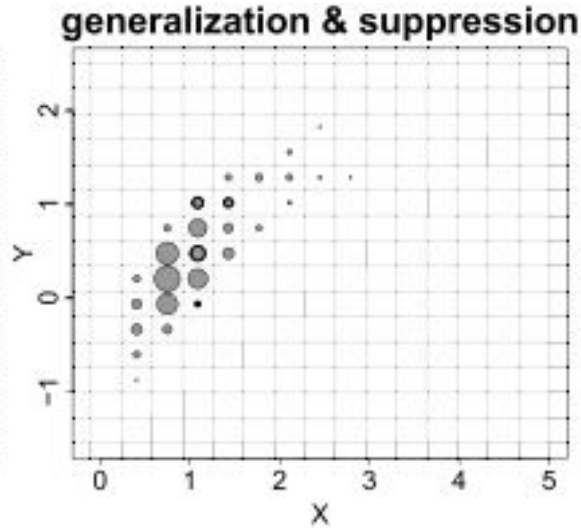
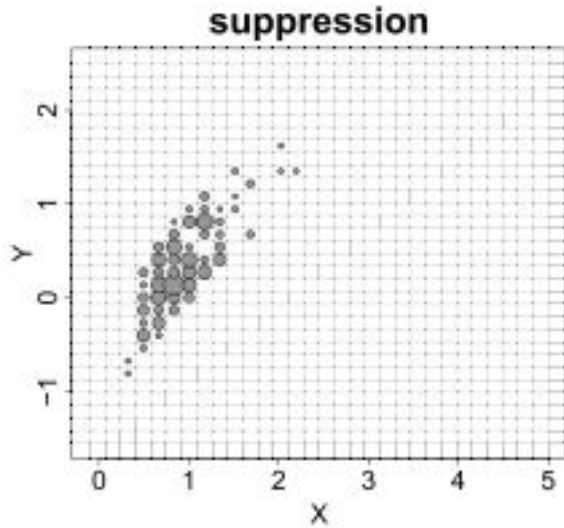
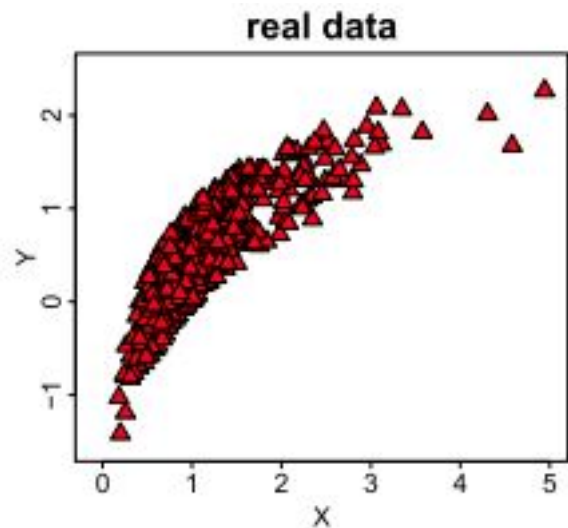
# Decisiones en visualización para facilitar la privacidad

Eliminación y/o agrupación de datos.



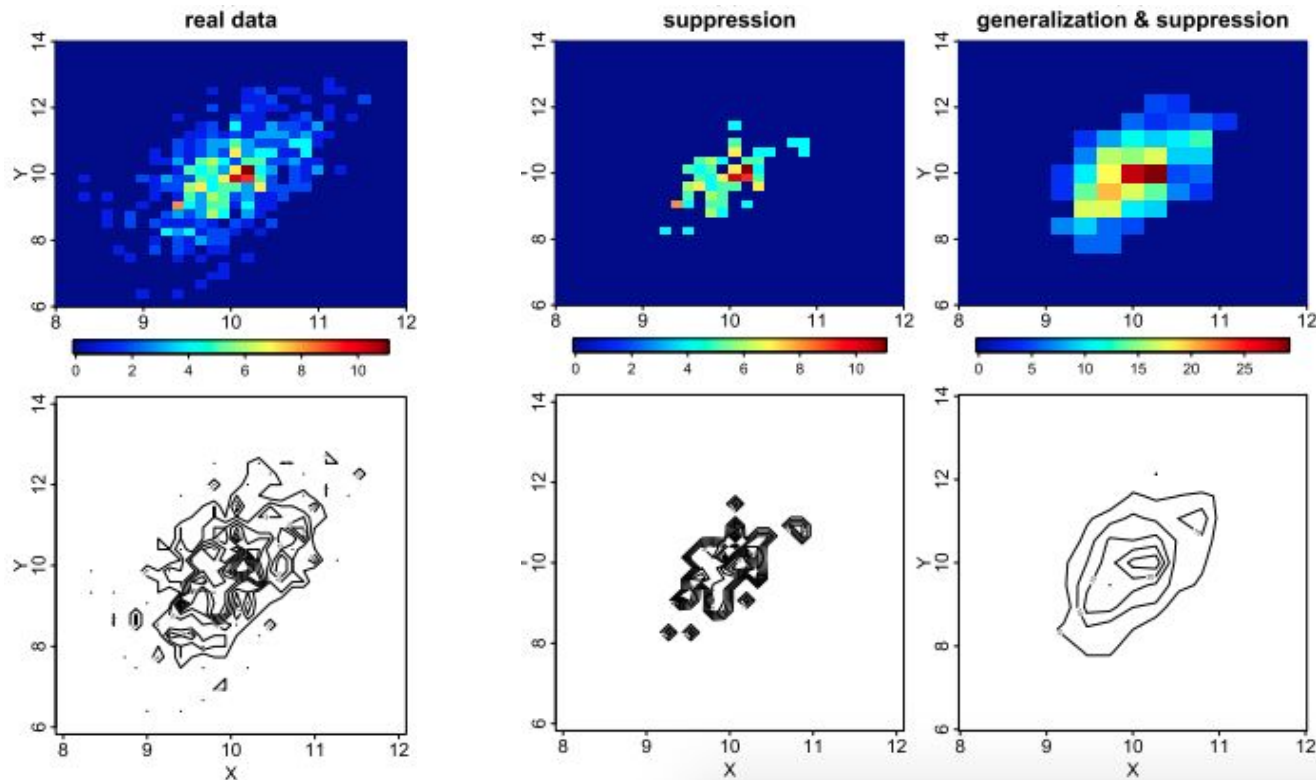
# Decisiones en visualización para facilitar la privacidad

Eliminación y/o agrupación de datos.



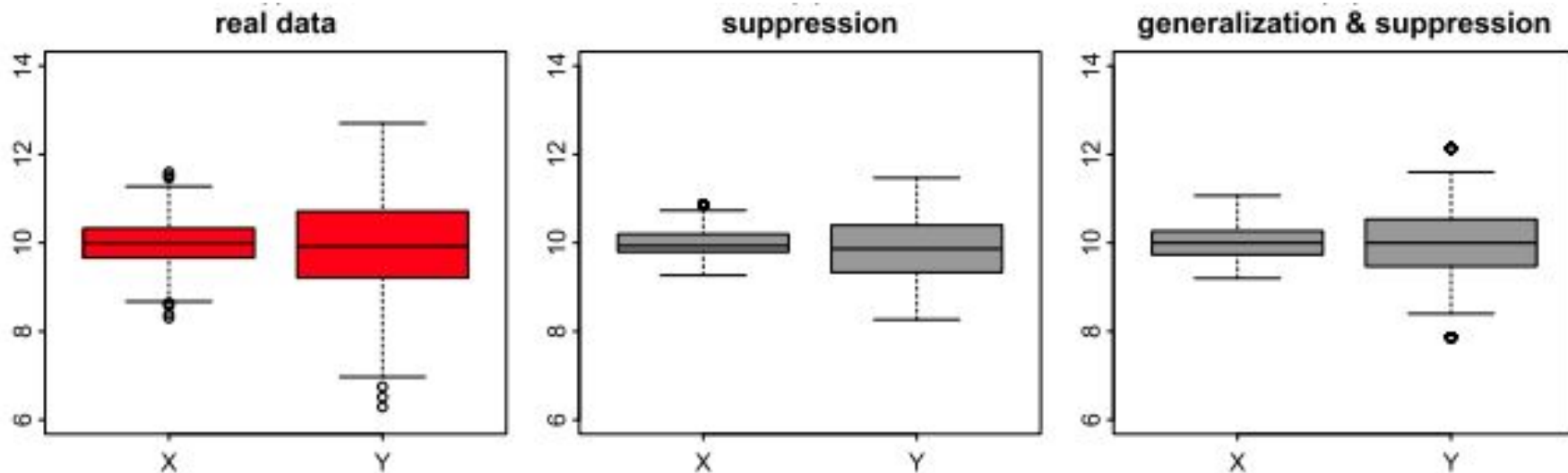
# Decisiones en visualización para facilitar la privacidad

Eliminación y/o agrupación de datos.



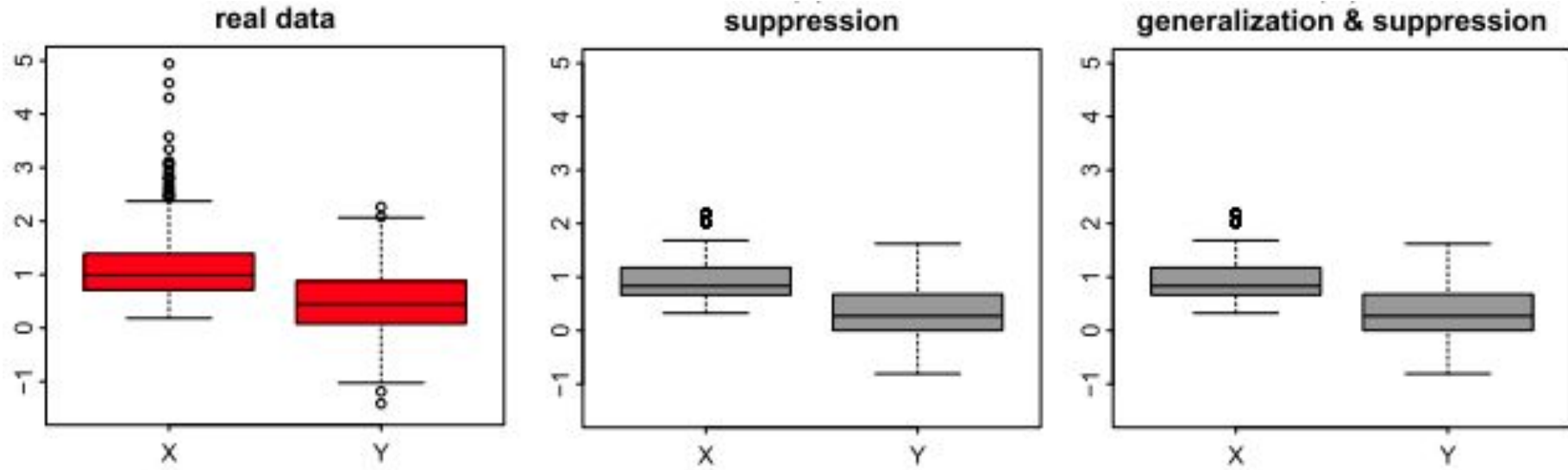
# Decisiones en visualización para facilitar la privacidad

Eliminación y/o agrupación de datos.



# Decisiones en visualización para facilitar la privacidad

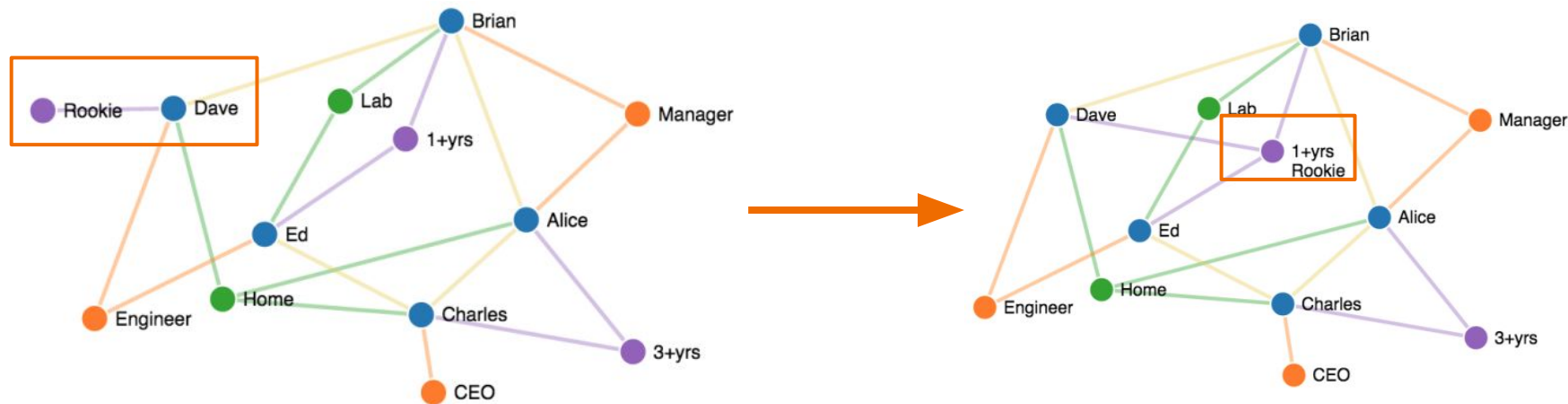
Eliminación y/o agrupación de datos.





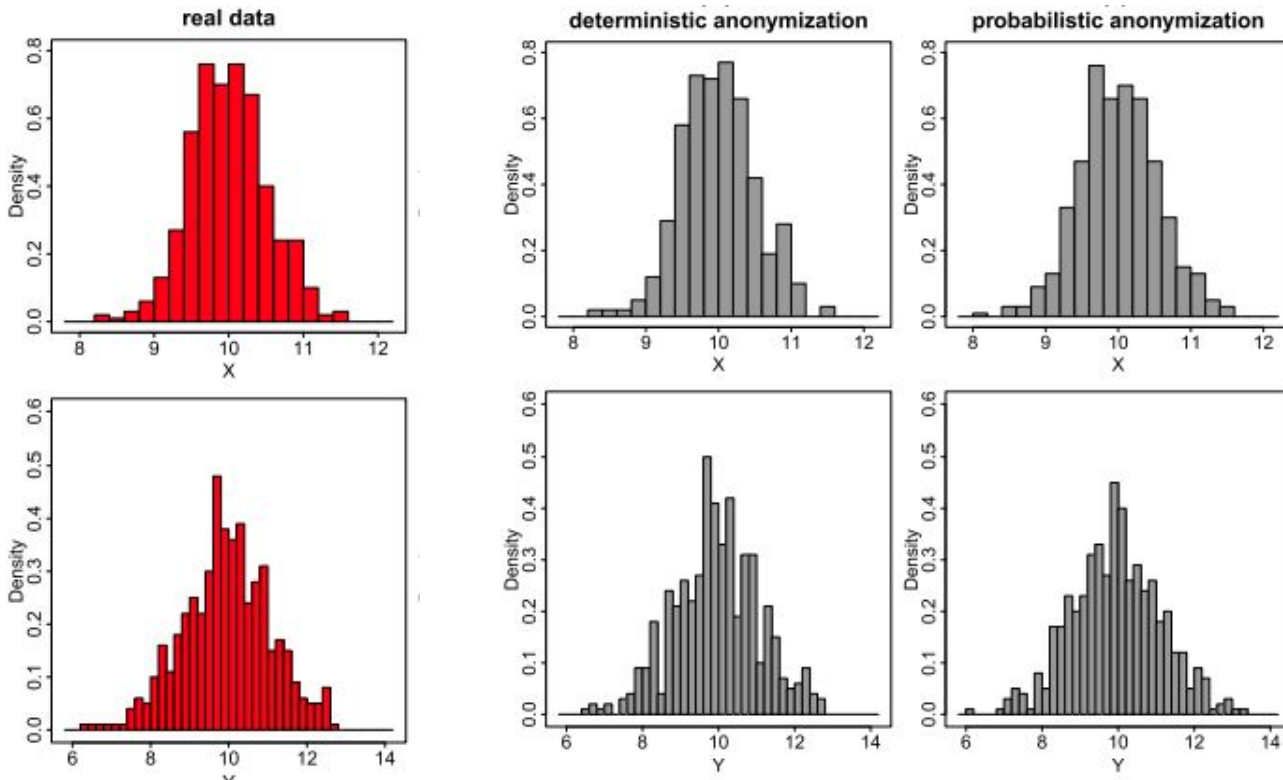
# Decisiones en visualización para facilitar la privacidad

Eliminación y/o agrupación de datos.



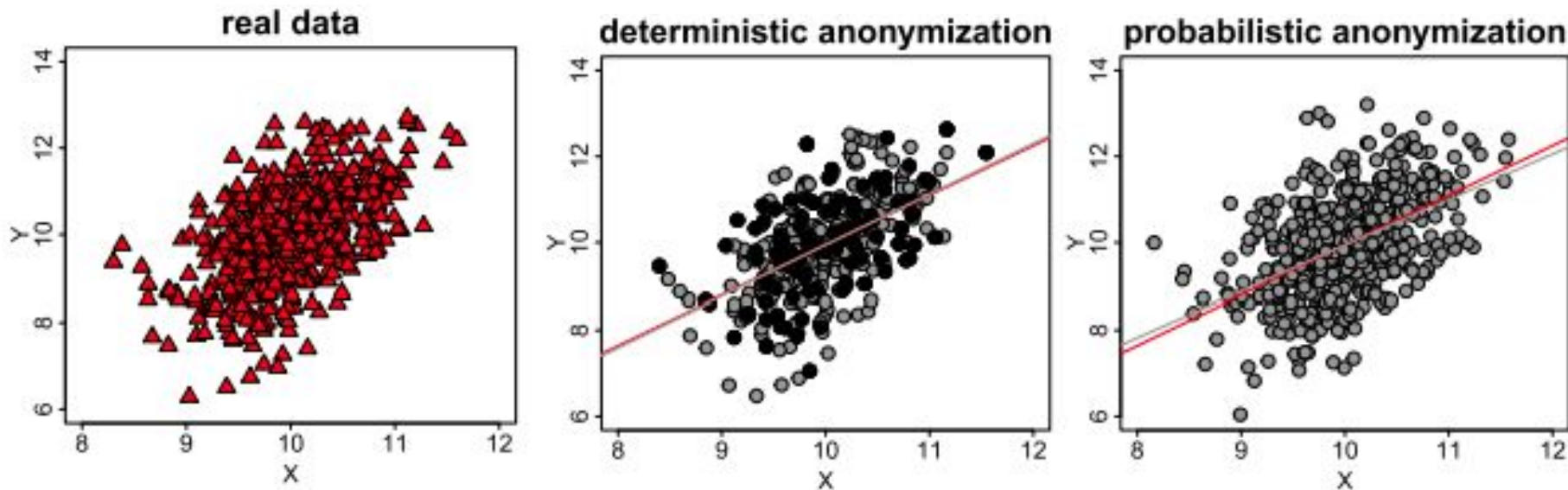
# Decisiones en visualización para facilitar la privacidad

Aplicar ruido en función de vecinos cercanos o de forma probabilística.



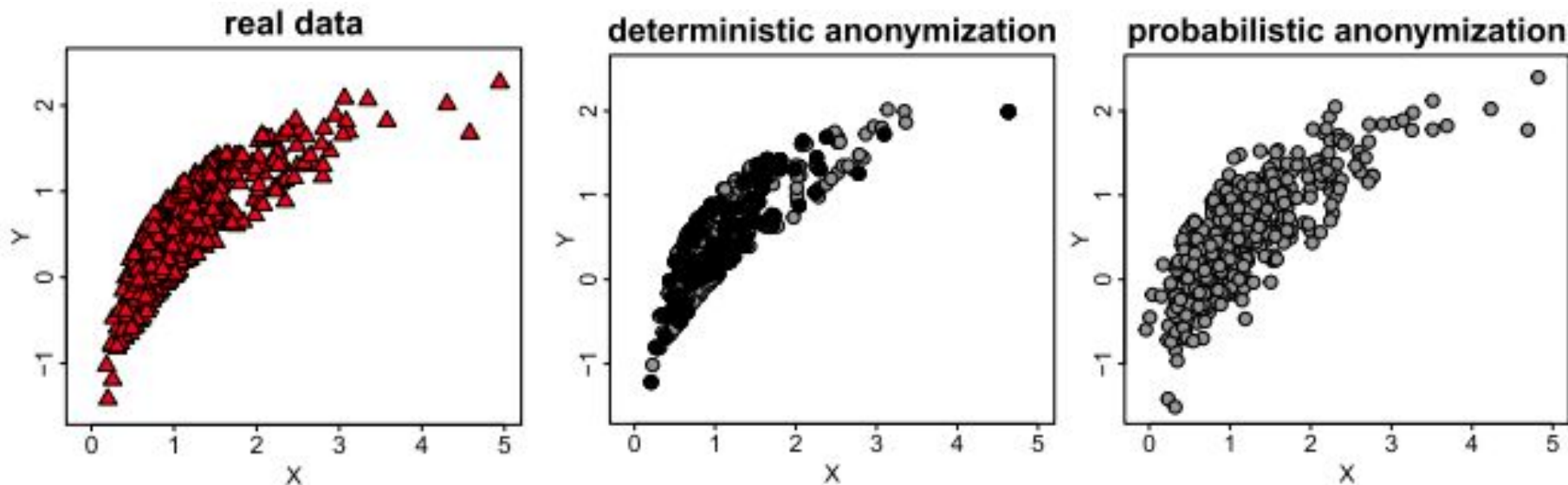
# Decisiones en visualización para facilitar la privacidad

Aplicar ruido en función de vecinos cercanos o de forma probabilística.



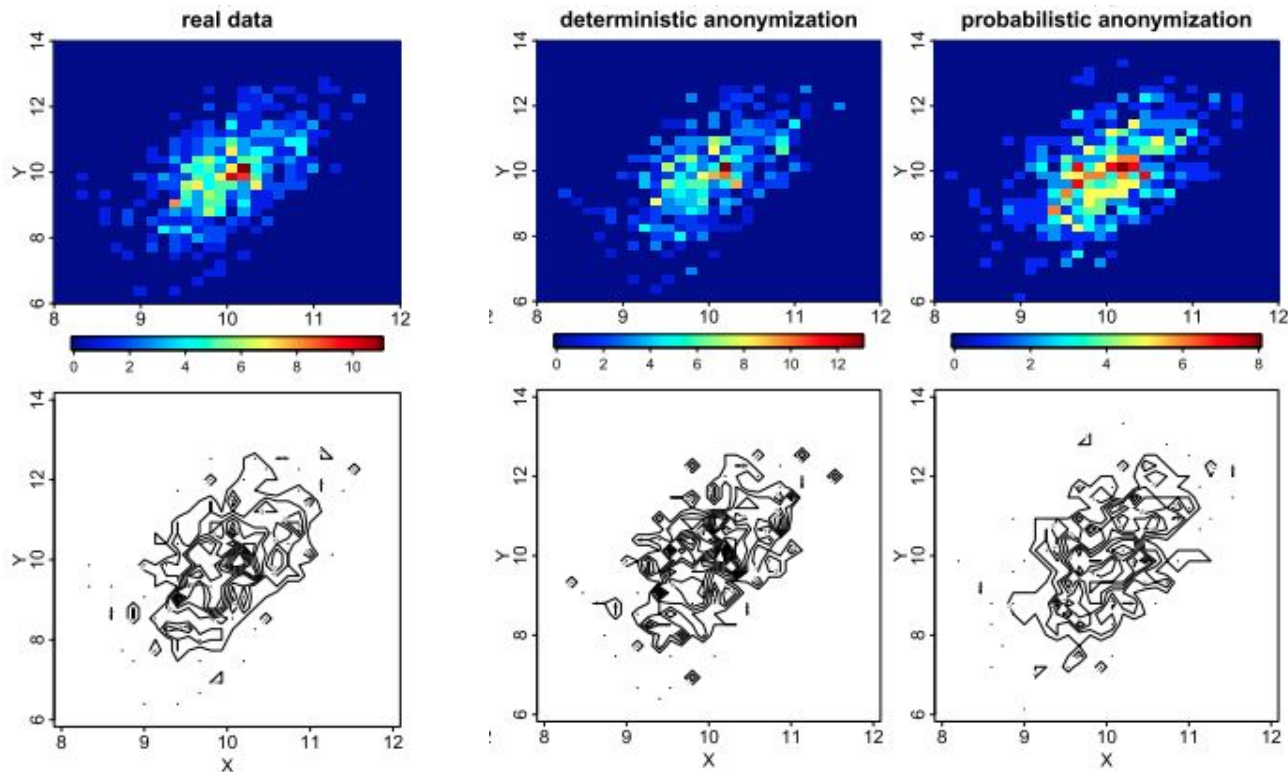
# Decisiones en visualización para facilitar la privacidad

Aplicar ruido en función de vecinos cercanos o de forma probabilística.



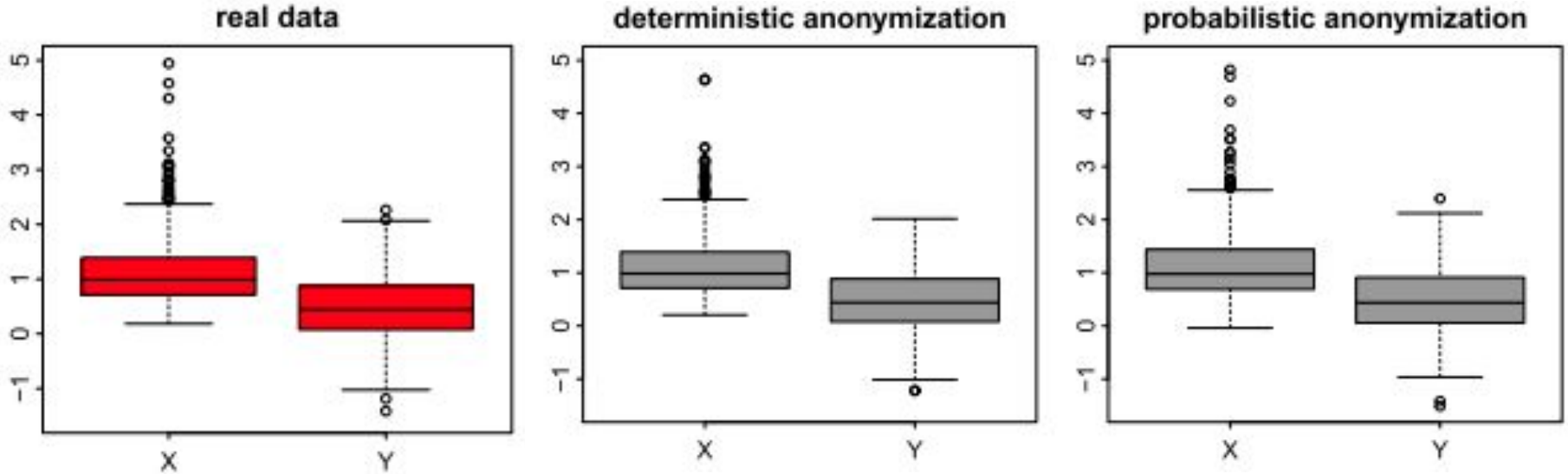
# Decisiones en visualización para facilitar la privacidad

Aplicar ruido en función de vecinos cercanos o de forma probabilística.



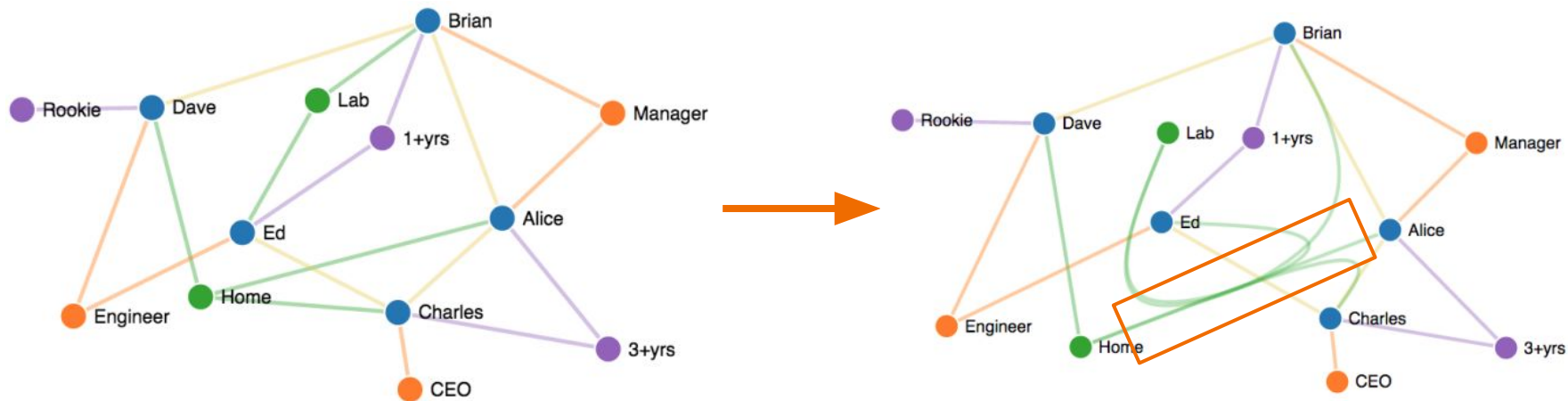
# Decisiones en visualización para facilitar la privacidad

Aplicar ruido en función de vecinos cercanos o de forma probabilística.



# Decisiones en visualización para facilitar la privacidad

Superposición de enlaces en el caso de grafo.



# Decisiones en visualización para facilitar la privacidad

Algunas decisiones implican:

1. Utilizar visualizaciones que dificultan la identificación total de un ítem.
2. Eliminación y/o agrupación de datos.
3. Aplicar ruido en función de vecinos cercanos o de forma probabilística.
4. Superposición de enlaces en el caso de grafo.



# Decisiones en visualización para facilitar la privacidad

Algunas decisiones implican:

1. Utilizar visualizaciones que dificultan la identificación total de un ítem.
2. Eliminación y/o agrupación de datos.
3. Aplicar ruido en función de vecinos cercanos o de forma probabilística.
4. Superposición de enlaces en el caso de grafo.

**¿Cómo aplicamos estas decisiones? ¿cuándo? ¿Cuál es mejor? 🤔**

→ Dependerá del contexto, cada caso es un mundo nuevo.

# Decisiones en visualización para facilitar la privacidad

Se siguen haciendo esfuerzos para apoyar este proceso. Por ejemplo:

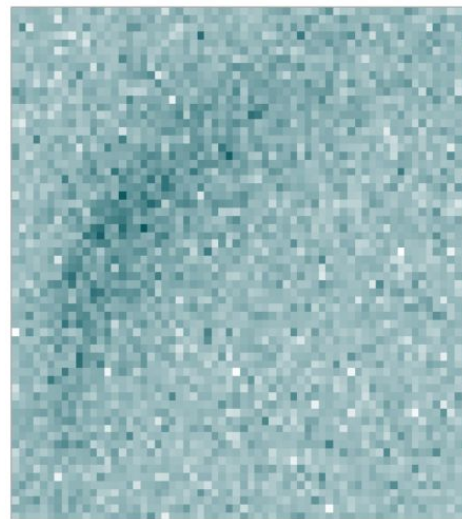
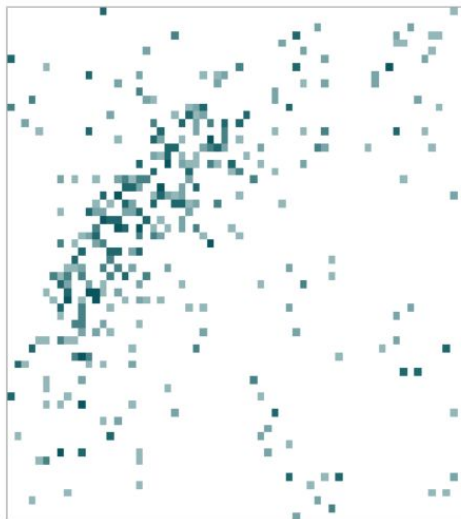
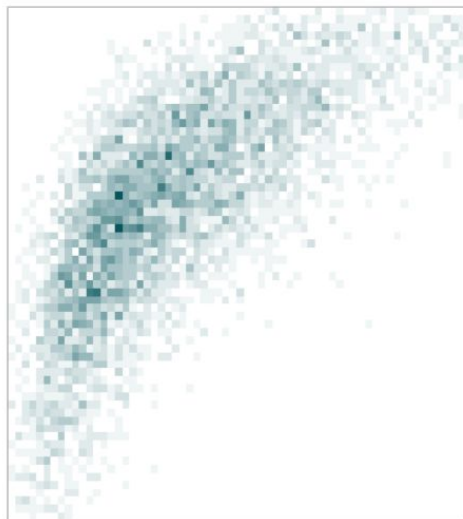
- Investigando la utilidad de diferentes formas de privatizar una visualización (2023)

Original

AHP

DAWA

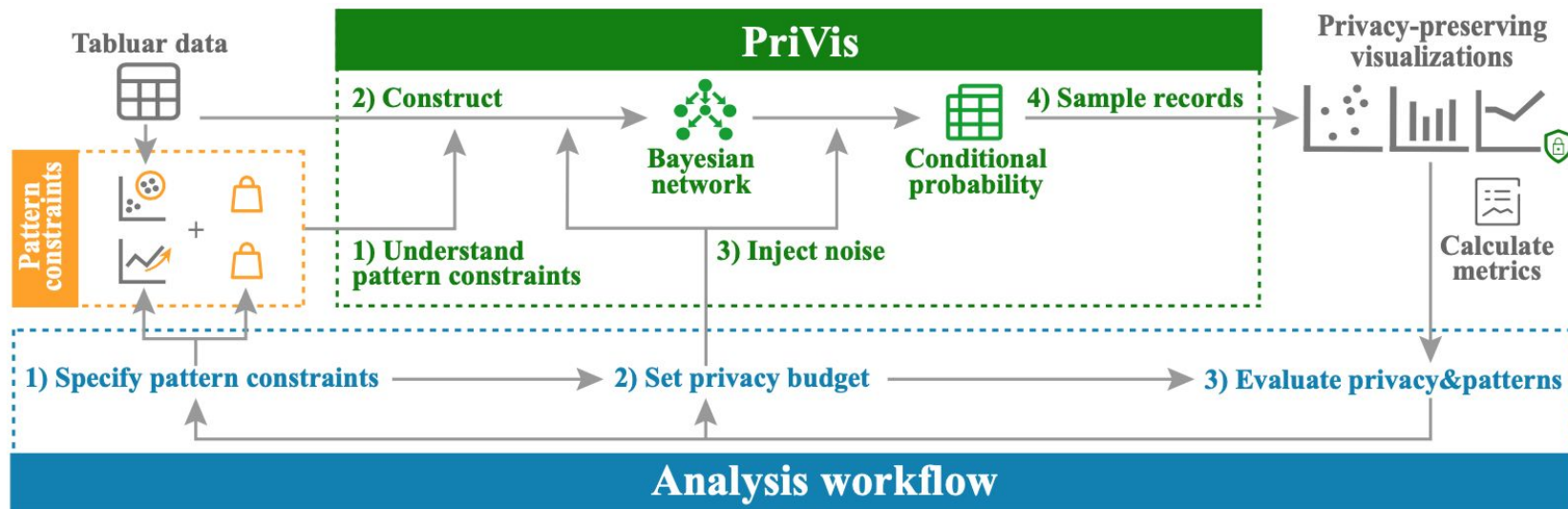
Laplace



# Decisiones en visualización para facilitar la privacidad

Se siguen haciendo esfuerzos para apoyar este proceso. Por ejemplo:

- Creando *software* que sugieren visualizaciones que muestran ciertos patrones (2022)



# Próimos eventos

## Próxima clase

- *Looker Studio*, una aplicación web para desarrollar visualización sin programar.
- Última clase de contenidos

## Próxima ayudantía

- **Esta semana:** no hay, es feriado.
- **La otra semana:** aplicar *brush* y/o zoom en un caso más aplicado.

---

# IIC2026

## Visualización de Información

— Hernán F. Valdivieso López —  
(2024 - 1 / Clase 23)

---