

Capsule 3 - Préparer les données temporelles

MQT-2101 - Analyse et modélisation des données

Pourquoi ça compte

Un mois absent peut-il être traité comme un mois observé?

Une panne de collecte et une fermeture n'ont pas le même sens. Un nettoyage automatique peut déformer la série.

Objectifs de la capsule

- Auditer le calendrier et les mesures.
- Justifier le traitement des anomalies.
- Limiter le prétraitement à l'information disponible.

Quatre vérifications

Calendrier

doublons et périodes absentes

Mesure

NA, zéro, unités

↓
éviter
manipuler

Contexte

extrêmes et ruptures

Mensuel ne signifie pas trente jours

→ 30, 31, 28 et 29 jan.

```
1 dates_attendues <- seq.Date(min(apprentissage$date),  
2                             max(apprentissage$date), by = "month")  
3 setdiff(dates_attendues, apprentissage$date)  
4 anyDuplicated(apprentissage$date)
```

Valeur manquante ou période absente

DATE PRÉSENTE, NA

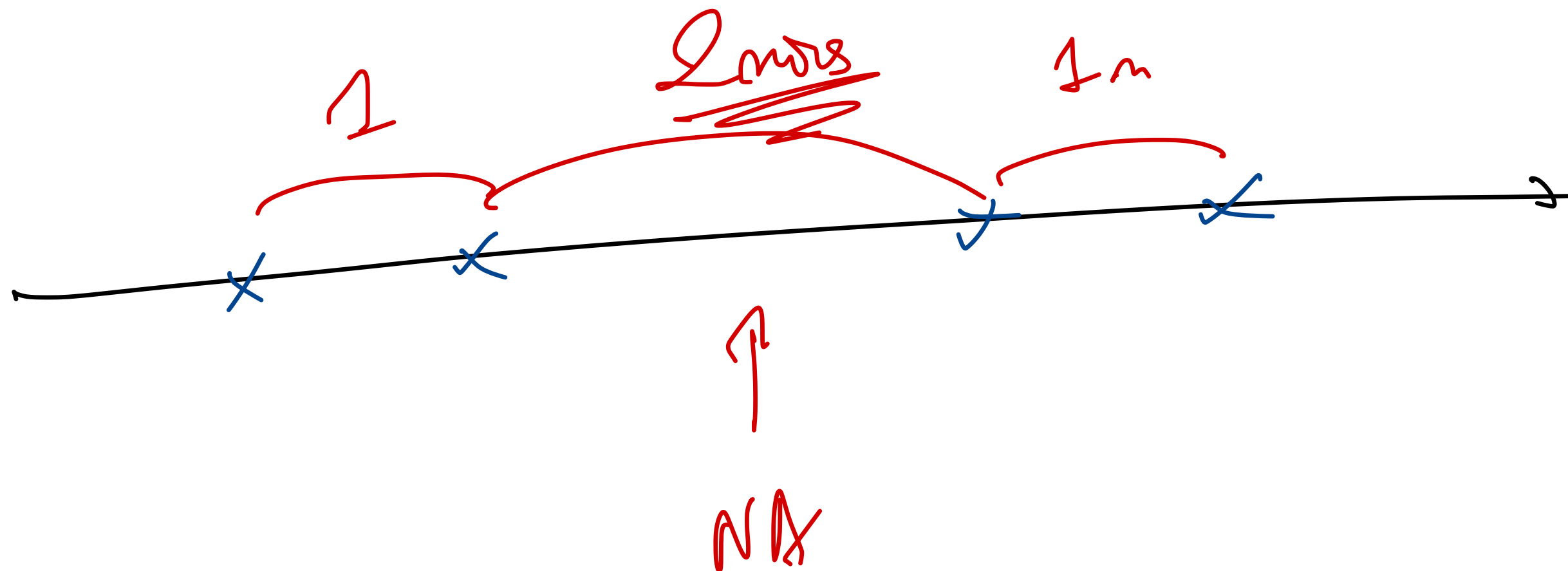
La mesure manque

Feb 2023
y = NA

DATE ABSENTE

Le calendrier a un trou

Retirer une ligne ne rend pas les périodes voisines consécutives.



Imputer exige une règle

HISTORIQUE

estimer avec l'information déjà disponible

SENSIBILITÉ

comparer plusieurs traitements

imputation : processus qui remplace les données manquantes

Espacement irrégulier

ÉVÉNEMENTS

arrivées à des heures variables

observe: arrivé d'un client
à un kiosque.

irrégulier

AGRÉGATION

compte par jour ou semaine

Extrême : erreur ou événement?

ERREUR VÉRIFIÉE

corriger et documenter

ÉVÉNEMENT RÉEL

conserver ou modéliser le contexte

Sans justification, ne pas supprimer automatiquement.



Combien d'années utiliser?

HISTORIQUE LONG

davantage de répétitions

saisonnalité

HISTORIQUE RÉCENT

meilleure pertinence après une rupture

saisonnalité

Les variables externes arrivent-elles à temps?

CONNUES À L'ORIGINE

calendrier, fermeture planifiée

INCONNUES

prix réalisé, demande future

À retenir

Repérer

les anomalies de calendrier

Documenter

chaque traitement

Respecter

l'information à l'origine

Production autonome 6.3

Auditez les dates et les valeurs manquantes sur l'apprentissage. Proposez un traitement documenté pour un mois absent, une erreur de saisie confirmée et un événement réel.

Auto-vérification

- J'ai distingué une valeur manquante, un zéro et une période absente.
- J'ai justifié séparément le traitement d'un mois absent, d'une erreur de saisie et d'un événement réel.
- Mes traitements sont estimés uniquement à partir des données d'apprentissage.