

Garbage In Garbage Out : DAG to understand?



Guillaume AVENIN
Réunion du CISP Club 2023

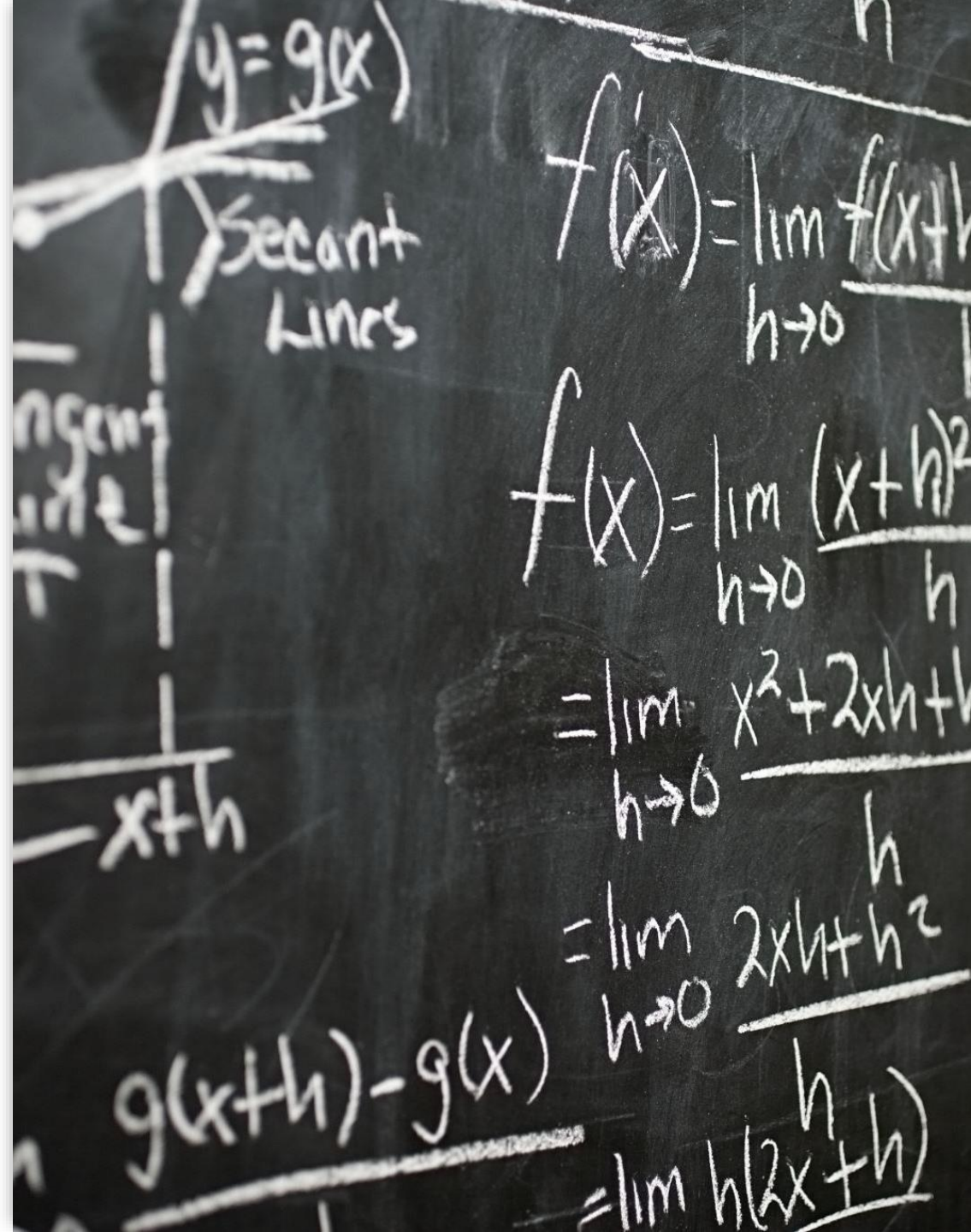


Comprendre

- Médecine rationnelle : Pour affronter un problème en déterminer la cause est une idée clef
- Médecine scientifique : une approche expérimentale pas toujours facile
- Passer des données à la compréhension = inférence causale

Inférence causale

- Comprendre ce qui pourrait se passer en changeant un paramètre
 - Individu A reçoit traitement Y et meurt alors qu'en l'absence de traitement il survit
 - Individu B survit qu'il reçoive ou non le traitement Y
- Inférence individuelle difficile
- Approximation par Inférence populationnelle
- Notion de population contre factuelle



DAG



Acronyme

Directed : il y aura des flèches

Acyclic : on ne tourne pas en rond

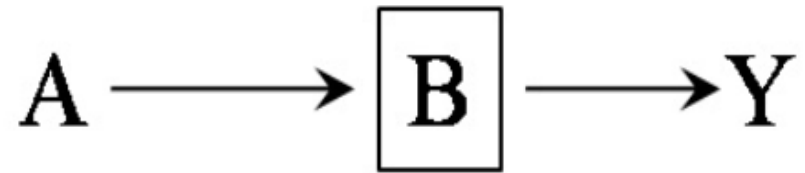
Diagram : un petit dessin



Il s'agit de représenter la chaîne de causalité suspectée lors de l'étude d'un phénomène donné : Inférence causale

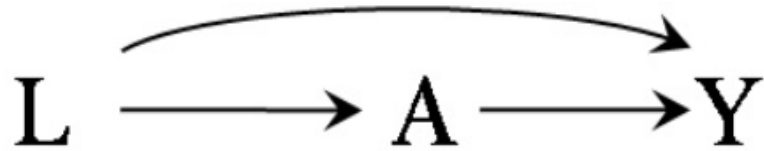


« Draw your assumption before your conclusions »



Principes et outils

- Distinguer Association et Causalité
- Des items
- Des flèches
- Des carrés (pour selection ou ajustement)



Notion de D separation

- 4 règles de dessin :

1. Si aucune variable n'est conditionnée, un chemin est bloqué si et seulement si deux flèches se rencontrent sur ce chemin au niveau d'une variable.
2. Tout chemin qui contient un "non-collider" sur lequel on a conditionné est bloqué.
3. Un "collider" sur lequel on a conditionné ne bloque pas un chemin.
4. Un "collider" qui a un descendant sur lequel on a conditionné ne bloque pas un chemin

$$A \rightarrow Y \leftarrow L$$

$$A \longrightarrow \boxed{B} \longrightarrow Y$$

$$A \xrightarrow{\quad\quad\quad} Y \longrightarrow \boxed{L}$$

$$A \xrightarrow{\quad\quad\quad} Y \longrightarrow L \longrightarrow \boxed{C}$$

Facteur de confusion

- Une cause commune perturbe l'évaluation de A sur Y

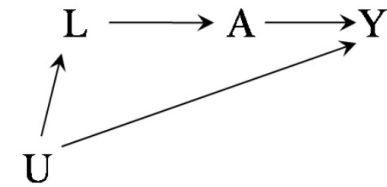
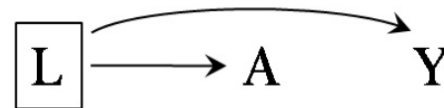
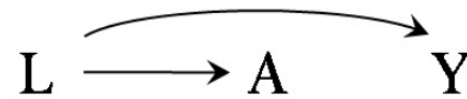
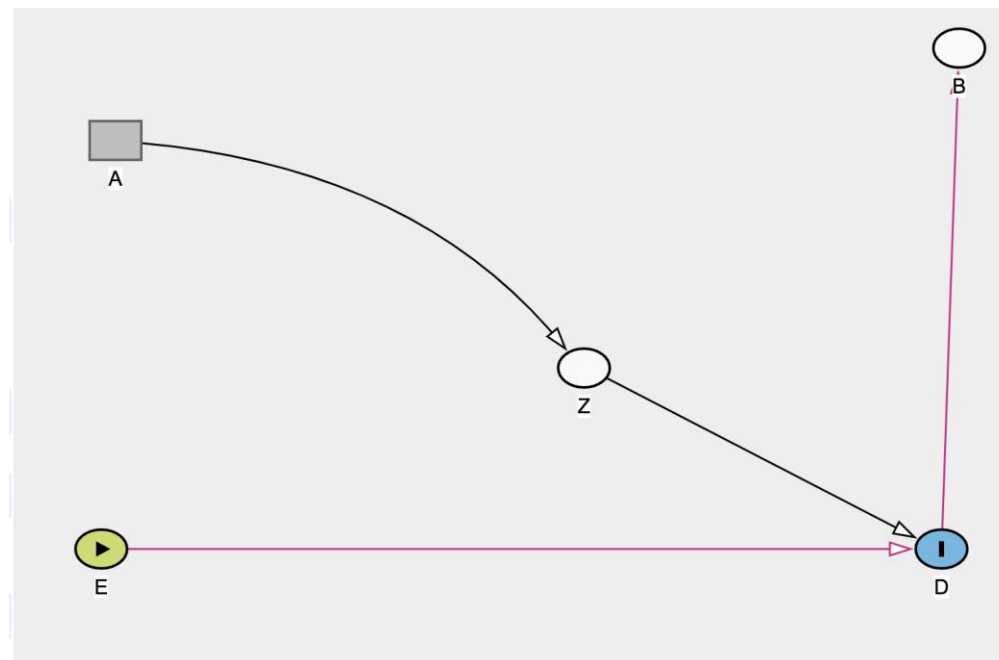


Figure 7.2

Some authors prefer to replace the unmeasured common cause U (and the two arrows leaving it) by a bidirectional edge between the measured variables that U causes.

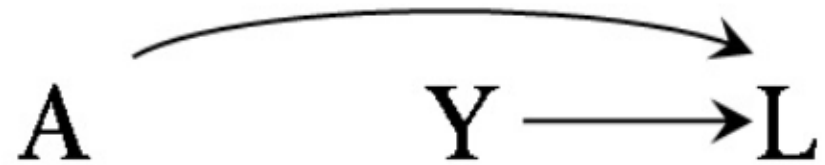
Biais de selection

- Healthy worker effect
 - Biais dans le choix des controles
 - Perdu de vus ...
-
- Le biais ne pose de souci que si le biais affecte l'outcome

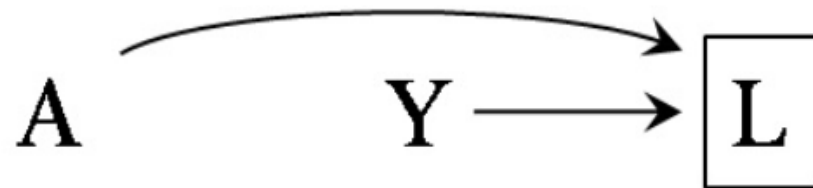


Biais de sélection

Un effet Commun



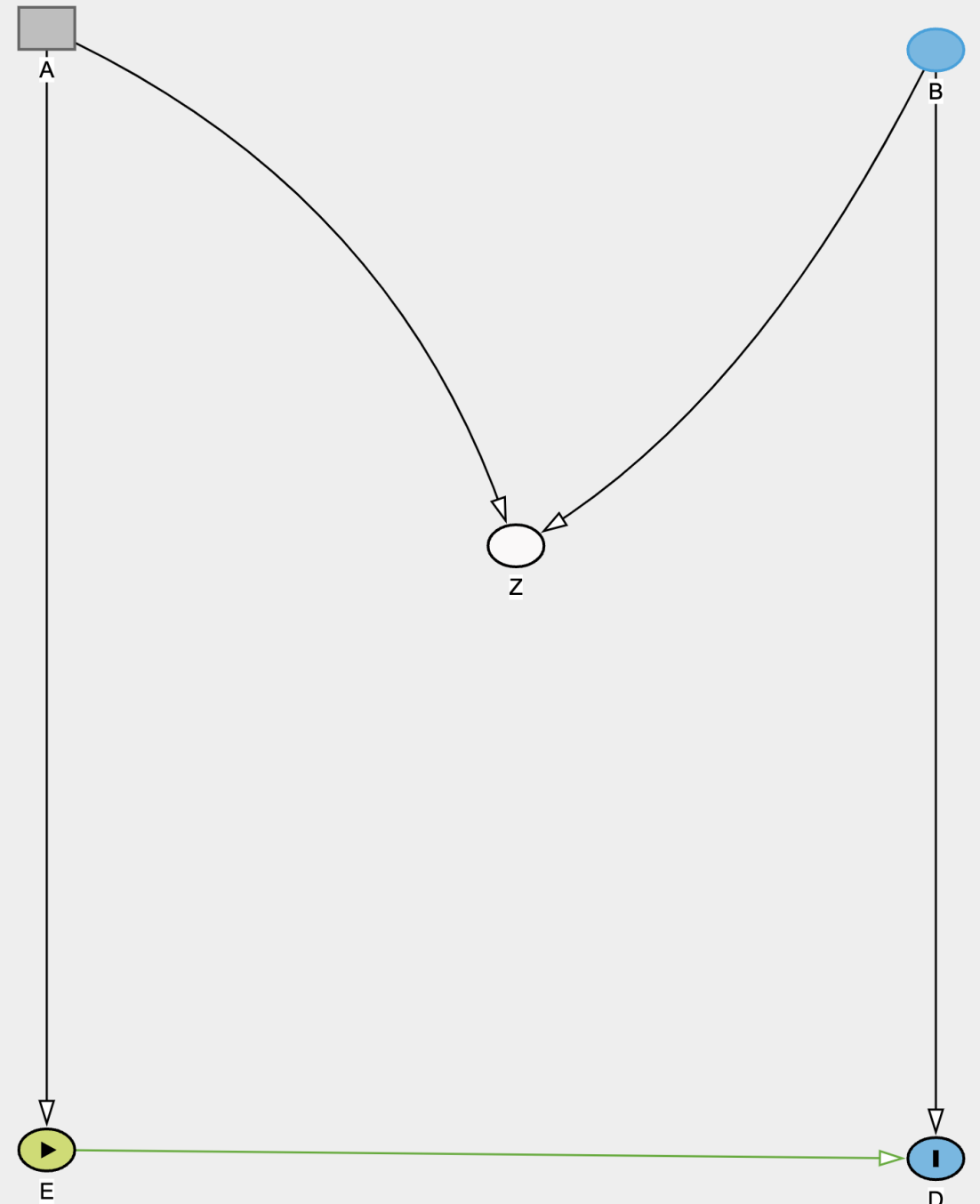
Perturbe l'évaluation de l'effet de Y sur L



Exemple et mise en pratique

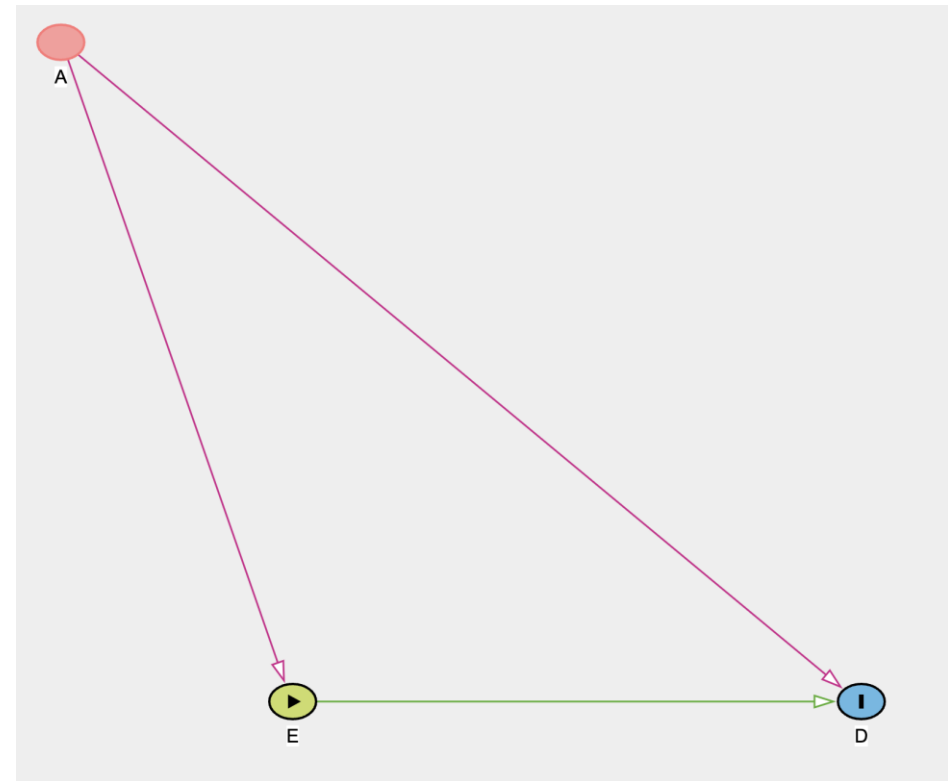
- Avec le logiciel dagitty
- Package R
- <https://dagitty.net/>

s
effects
paths
paths
ral structure
sure
me
stor of exposure
stor of outcome
stor of exposure and
ed variable
erved (latent)
ariable
ath
ath



Comment bloquer :

- Design d'étude : l'essai clinique randomisé vise à rendre l'exposition A indépendante du recrutement
- Stratification/ sélectionner les populations (les fumeurs, les CSP..)
- Matching/appariement
- G techniques
 - Standardization
 - IP weighting
 - ...



Adjustement

many variables are collected in a 1-year baseline period before the index date to adjust

4 dimensions previously identified:

- Conditions health
- Social economics status
- Therapy related factors
- Health system ambulatory

Propensity score

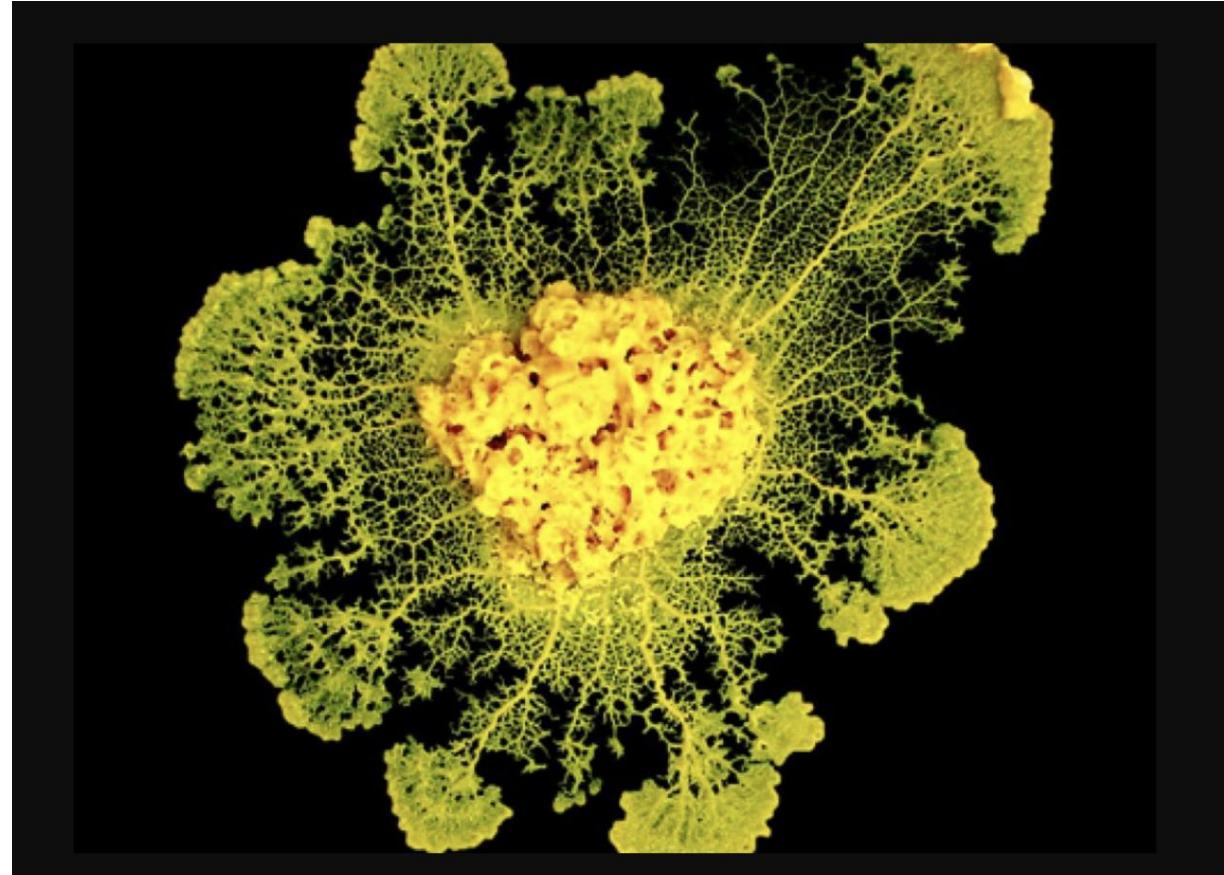
- calculate a probability to exposure related to the outcome
- variable related
 - ALD, Hospitalization code
 - Insurance status and other information
 - other Drug consumption
 - Health care consumption

High density

- No such variable but hundreds of them available: semi-automated approach
- Selection on prevalence
 - more frequent ICD code
 - (depends on the data available in each country)
 - more frequent ATC code
 - more frequent consumption of referral laboratory ...

L'apport des modèles computationnelles des grandes bases de données

- Intelligence Artificielle et la prédiction en dehors de toute signification
- Apprendre à renoncer à comprendre ce qui détermine le résultat



Conclusion : une introduction à l'analyse des causalités



- Cette approche permet de représenter l'avis d'expert et la simplicité apparente de la méthode tient davantage de l'élégance que de la trivialité
- Cette représentation présente plusieurs limitations et une évolution est proposée avec les SWIG pour Single World Intervention Graph développées par Richardson and Robins (2013)

Intelligence naturelle et artificielle



LUC JULIA : IL FAUT AVOIR VU 2
CHATS POUR QU'UN ENFANT
SACHE RECONNAITRE UN CHAT



IL EN A FALLU 100 000 POUR
QU'UN ORDINATEUR PUISSE LE
FAIRE (SAUF DANS LE NOIR)



COMPILER UNE CONNAISSANCE
D'EXPERT AVEC UNE PUISSANCE
DE CALCUL

Sources et logiciels

- Nombreuses références sur Internet
- Miguel Hernan / James Robins
 - What If :
https://www.hsph.harvard.edu/miguel-hernan/wp-content/uploads/sites/1268/2023/10/hernanrobins_WhatIf_30sep23.pdf
 - <https://www.edx.org/course/causal-diagrams-draw-assumptions-harvardx-ph559x>
- <https://dagitty.net/>

