# Guidelines for DAG construction exercise

Using the list of variables provided below, we ask you to construct a directed acyclic graph (DAG) for the data generating mechanism behind these variables. Constructing a DAG involves suggesting a number of causal relationships between the variables. In order to decide on which potential causal relationships exist between two variables, we ask you to provide your best educated guess. You do not have to be a subject-matter expert or consult scientific literature; this is merely an exercise! But please make sure you do not propose causal relationships that go against the direction of time (see temporal information on the variable list).

How to add arrows to the DAG template

In order to construct the DAG, you need to add arrows between variables in the attached DAG template. For each pair of variables, we ask you to draw an arrow between them if you believe that one is a potential direct cause of the other. A direct causal effect is a causal effect that is not mediated via other variables. For example, for two variables X and Y, where X is a direct cause of Y, you should draw the following arrow:

$$X \longrightarrow Y$$
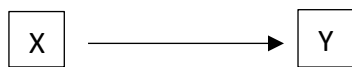
If you draw the arrow in the opposite direction, it means that Y is a potential direct cause of X:

$$X \longleftarrow Y$$
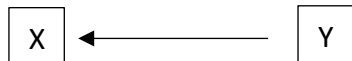
Finally, if you do not draw an arrow between X and Y it means that you do not believe that there is any direct causal relationships between the two variables: X is not a potential direct cause of Y, *and* Y is not a potential direct cause of X:

$$X \qquad Y$$

We ask you to draw at most one arrow between each pair of variables. If you are unsure about the causal relationship between two variables, please choose the option among the three possibilities described above (arrow in one direction, arrow in the other direction, or no arrow) that you have the most confidence in.

The graph is not allowed to contain any cycles. A cycle means that you can follow a directed path of arrows from one variable to another. Here is an example of a graph with a cycle:

$$X \longrightarrow Y, \quad Y \longrightarrow Z, \quad Z \longrightarrow X$$
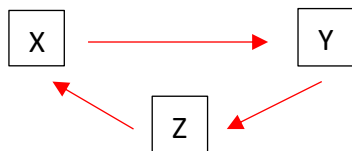
Such a graph will *not* be allowed. We ask you to draw your best suggestion for a DAG under these restrictions.

Data and variables: NLSY97

We will consider data from the The National Longitudinal Survey of Youth 1997 (NLSY97). NLSY97 is a longitudinal study following youth born between 1980 and 1984 with interviews from 1997 to 2022 and a total of 20 rounds (R1 - R20). We consider a subset of the data with measurements in 1997, 2002 and 2008 (R1, R6 and R12), and the variables described in the table below. We consider a total of 6315 complete cases. For details see https://www.nlsinfo.org/content/cohorts/nlsy97.

Variable name	Tier	Type	Description
r1_hhchildren	R1	Numeric	Number of household members younger than 18
r1_urban	R1	Binary	Urban status (Urban / Non-urban)
r1_mcollege	R1	Binary	Did biological mother attend college? (Yes / No)
r6_exercise	R6	Numeric	Days with 30+ minutes exercise /week
r6_depressed	R6	Binary	Having felt depressed during the last month (Sometimes / No)
r6_docvisits	R6	Numeric	Number of doctor's visits during the last year
r12_health	R12	Binary	Perception of own health (Good / Poor)
r12_depressed	R12	Binary	Having felt depressed during the last month (Sometimes / No)
r12_docvisits	R12	Numeric	Number of doctor's visits during the last year

DAG template

hhchildren

exercise

health

urban

docvisits

docvisits

mcollege

depressed

depressed

r1

r6

r12