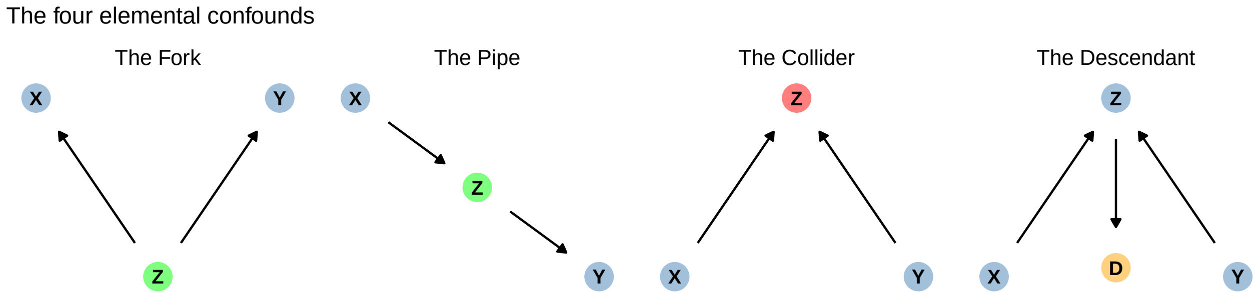
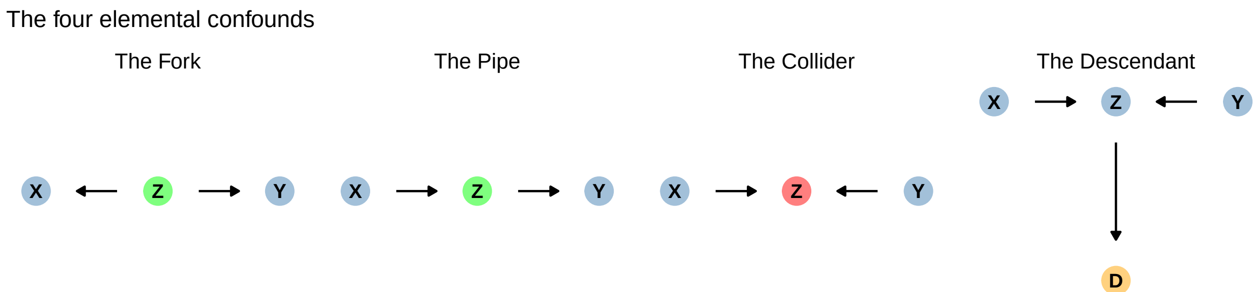
The Four Elemental Confounds

Here are the representations for our four types of variable relations: the fork, pipe, collider, and descendant.



Since we will be thinking about these *along a path*, we might prefer to look at them like this.



Each node on a path is either a fork, a pipe, or a collider. (Note: this status depends on the path; the same node may play different roles on different paths.)

Opening and closing paths

Our goal is to have all backdoor paths closed.

1. **Fork**
   - Example: Growth ← Moisture → Fungus
   - This is the "common cause" confound.
   - $X \perp\!\!\!\perp Y \mid Z$
   - Conditioning on Z blocks the path (of information) between X and Y.
2. **Pipe**
   - Example: Treatment → Fungus → Growth
   - This is the "mediated effect" confound.
   - $X \perp\!\!\!\perp Y \mid Z$
   - Conditioning on Z blocks the path (of information) between X and Y.
3. **Collider**
   - Example: Trustworthy → Selection ← Newsworthy
   - This is the "common effect" confound.
   - $X \not\!\perp\!\!\!\perp Y \mid Z$
   - Conditioning on Z opens the path (of information) between X and Y.
4. **Descendant**
   - Conditioning on a descendant is like a weak version of conditioning on its parent.
   - D can be used as a proxy for Z.

The recipe (page 185)

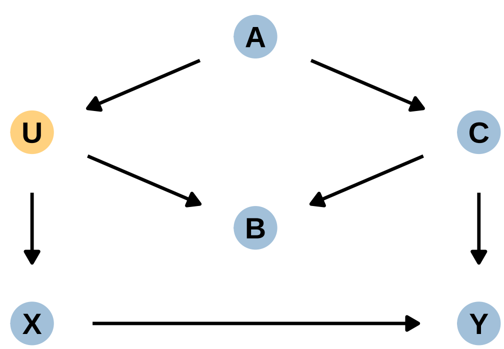
The recipe given in *Statistical Rethinking* is a little bit imprecise. Here's a modified version:

1. **List all paths** connecting X (the potential cause of interest – the eXposure) and Y (the outcome).
2. Classify each path as **causal** or **backdoor** (non-causal)
   - A backdoor (or non-causal) path = at least one arrow followed "backwards"
   - Causal path = a path that follows all the arrows "forwards"
3. Classify each backdoor path by whether it is **open or closed**.
   - open = no collider on path
   - closed = collider on path
4. **Close any open backdoor paths** (if possible) by conditioning on one or more variables **without closing any causal paths**.
   - Rule 1: Conditioning on any non-collider blocks/closes a path. [green]
   - Rule 2: Not conditioning on any collider blocks/closes a path. [red]
   - Rule 3: Conditioning on all colliders and on no non-colliders opens a path.
   - Rule 4: Conditioning on a descendant of a collider (partially) conditions on the collider. [orange]

     So Rules 2 and 3 need a little updating to be completely correct. We need to avoid conditioning on colliders and all of their descendants to close a path.

Example: Two roads

"The DAG below contains an exposure of interest X, an outcome of interest Y, an unobserved variable U, and three observed covariates (A, B, and C)" (p. 186).



In this DAG, there are two backdoor paths from X to Y

- $X \leftarrow U \leftarrow A \rightarrow C \rightarrow Y$, which is open; and
- $X \leftarrow U \rightarrow B \leftarrow C \rightarrow Y$, which is closed.

Conditioning on either C or A will close the open backdoor.

```
dag_6.1 <-
  dagitty("dag { U [unobserved]
  X -> Y;  X <- U <- A -> C -> Y;  U -> B <- C }" )

adjustmentSets(dag_6.1, exposure = "X", outcome = "Y")
```

```
## { C }
## { A }
```

More Practice

In each example below a possible causal influence in indicated. Determine which variables in the DAG should be included in your model to estimate this causal influence. Do this by following our recipe:

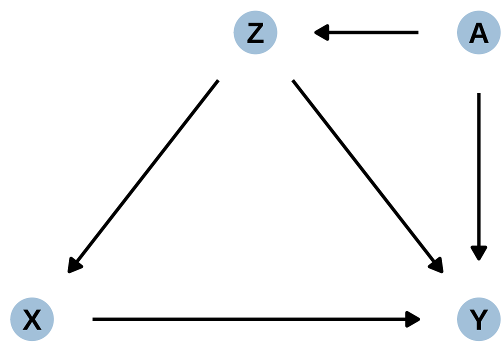
- List all backdoor paths between the indicated variables;
- For each backdoor path, determine whether it is open or closed;
- Choose variables to condition on that close all backdoor paths without closing any causal paths.

Any nodes in orange are unobserved. If possible, avoid conditioning on unobserved variables.

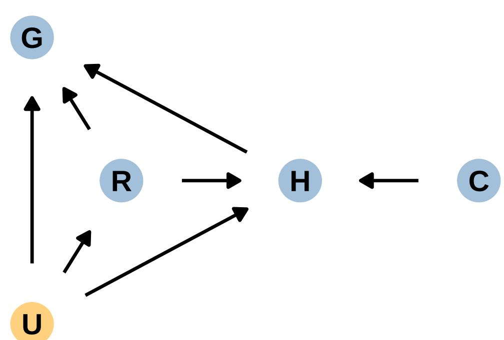
Note: You can check your work by creating the DAG with `dagitty()` or `dagify()` and using `adjustmentSets()`.

1.

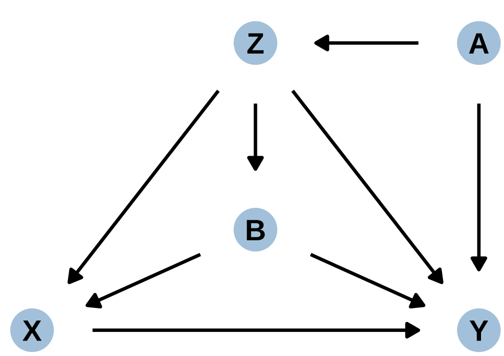
a. $X \to Y$
b. $Z \to Y$



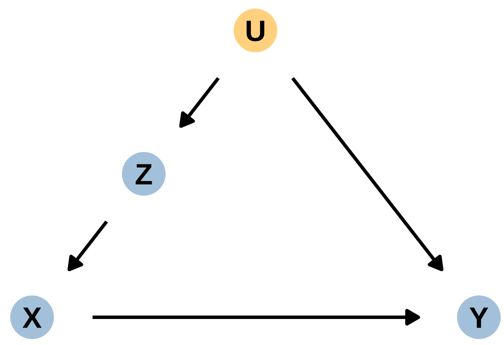
2. $R \to G$



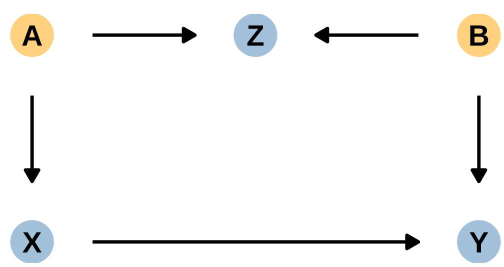
3. $X \to Y$



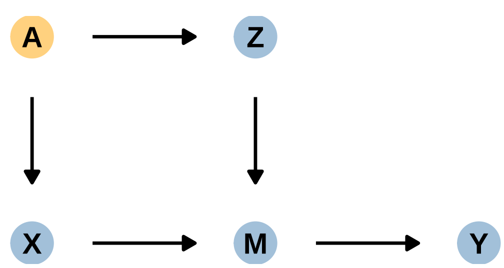
4. $X \to Y$



5. $X \to Y$



6. $X \to Y$



7. $X \to Y$