RUcausal: An R Package for Representing Uncertainty in Causal Discovery

Ioan Gabriel Bucur

Radboud University Nijmegen

September 29, 2019

Why is causality important?

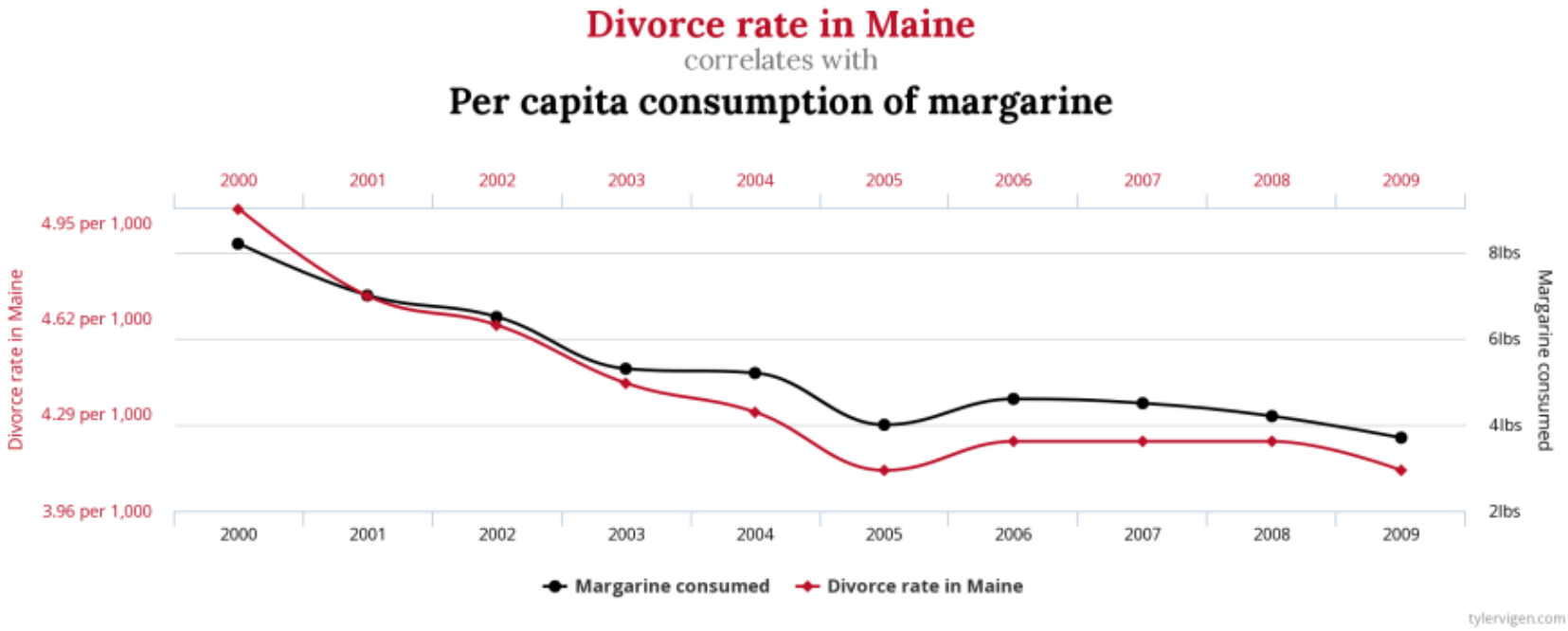
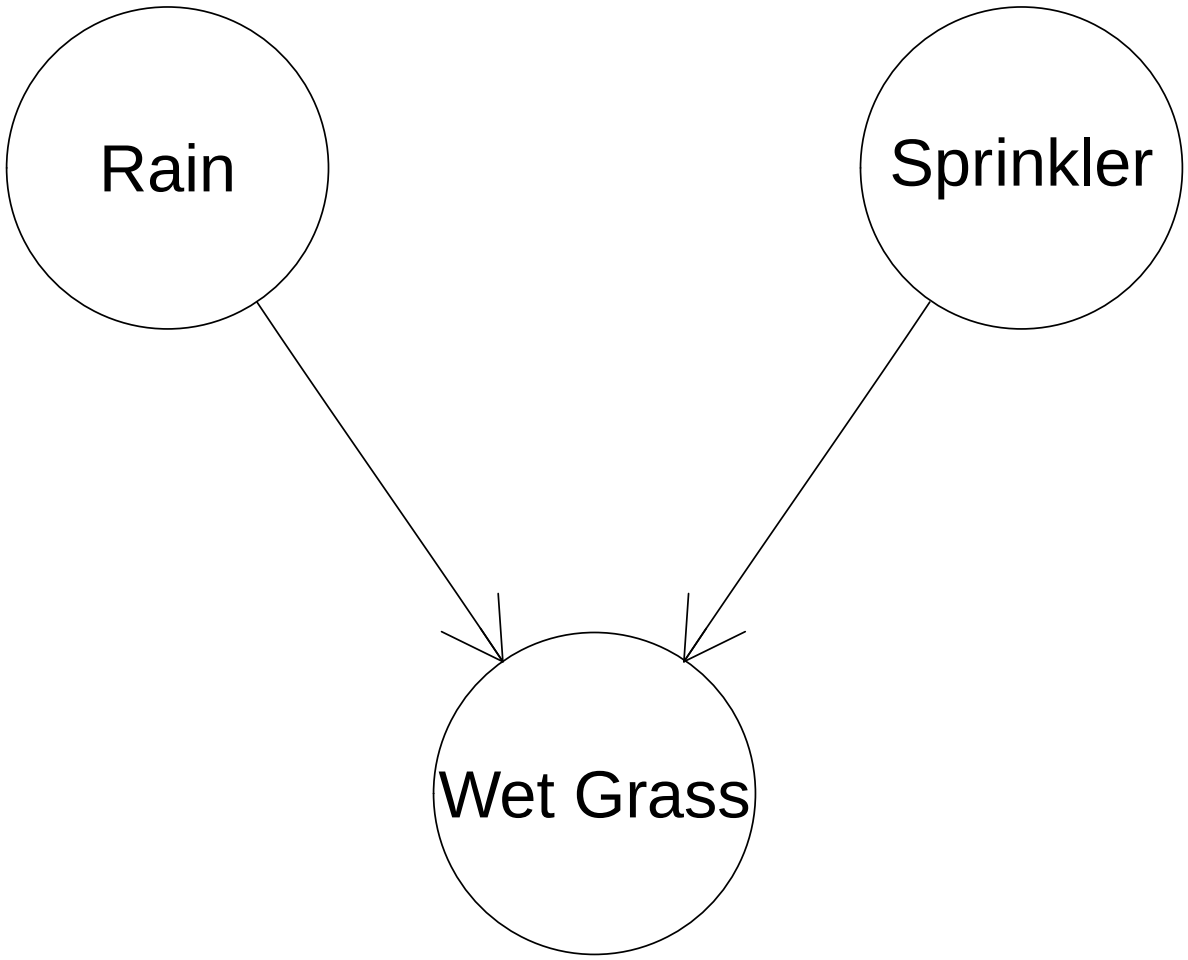


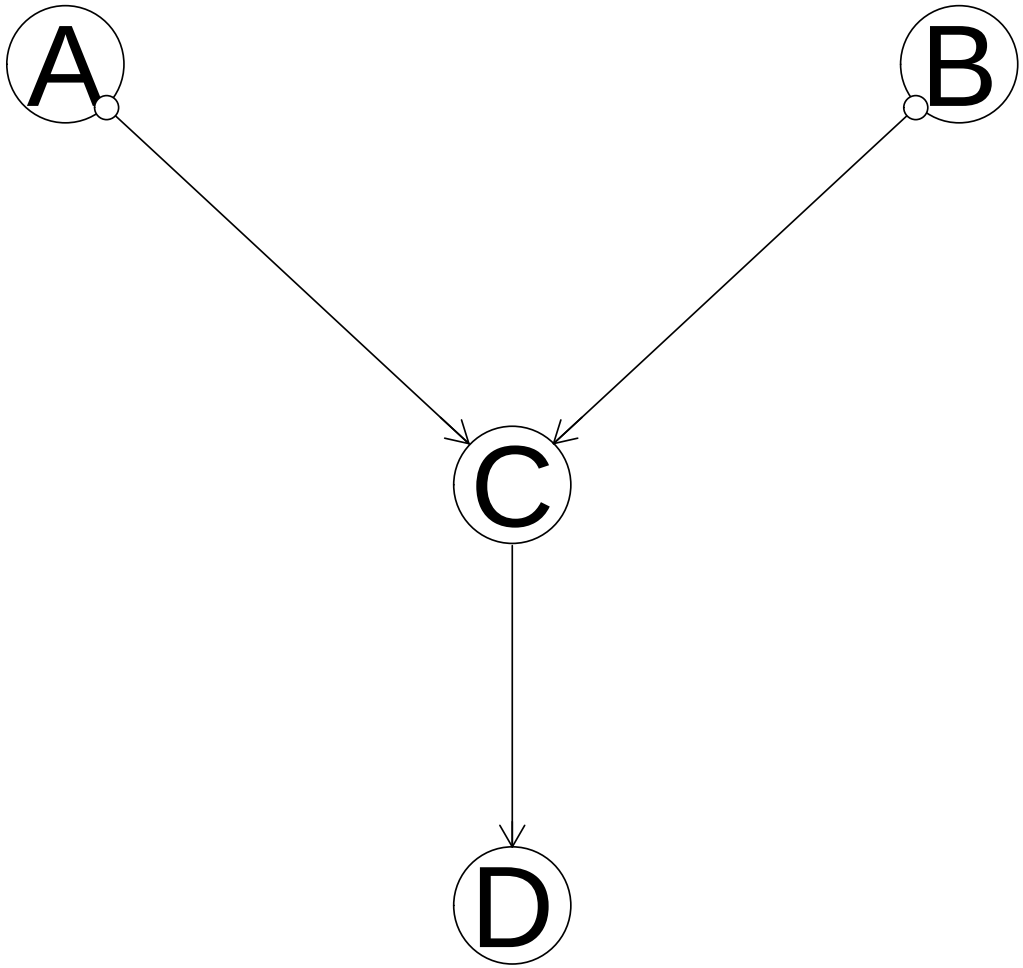
Figure 1: Example of spurious correlation

Causal discovery



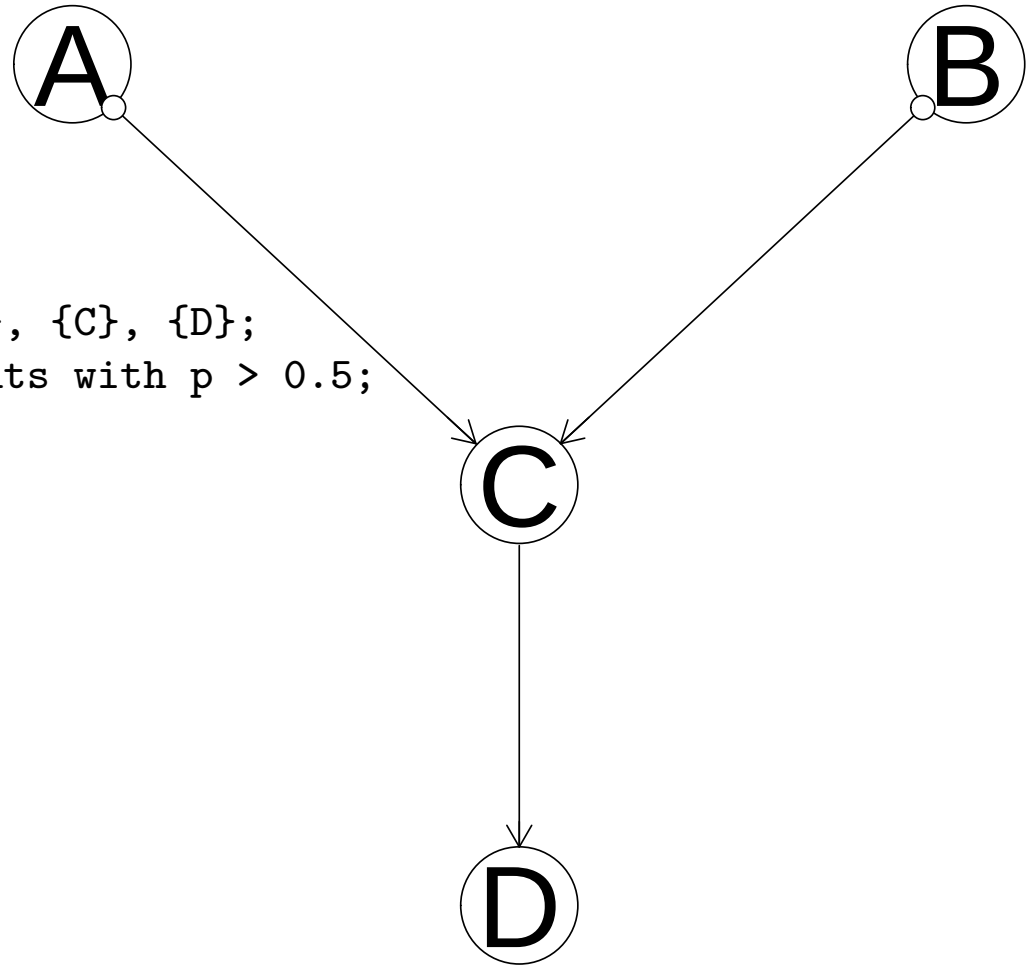
Causal discovery with RUcausal

```r
set.seed(2019)
N <- 1000 # number of observations

# Generate data from Y causal structure
# A ==> C <== B & C ==> D
A <- rnorm(N)
B <- rnorm(N)
C <- A + B + rnorm(N)
D <- C + rnorm(N)
dat <- cbind(A, B, C, D)

# Compute correlation matrix
R <- cor(dat)

# Run the BCCD causal discovery algorithm
capture.output(
  bccd.fit_ystr <- BCCD(R, N),
  file = '/dev/null'
)
```



Causal discovery with RUcausal

```
## Graph on 4 variables: {A}, {B}, {C}, {D};
## Only showing inferred statements with p > 0.5;
## [p = 1.00] {A} === {C};
##   [p = 0.78] {C} =/> {A};
## [p = 1.00] {B} === {C};
##   [p = 0.78] {C} =/> {B};
## [p = 1.00] {C} === {D};
##   [p = 0.73] {C} ==> {D};
##   [p = 0.70] {D} =/> {C};
```



Input background knowledge

Structural information

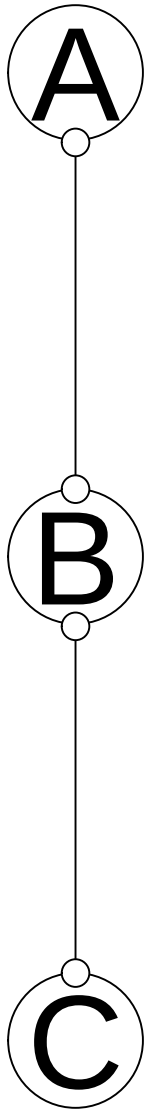
- {A} === {B}
- {A} =/= {B}

Causal information

- {A} ==> {B}
- {A} =/> {B}

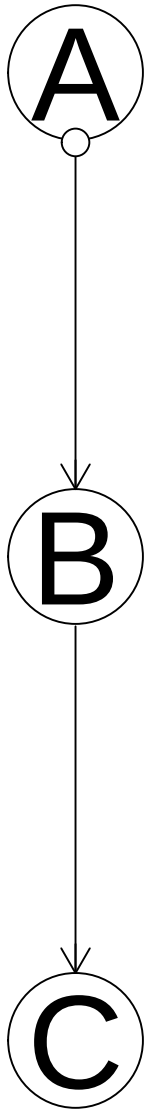
Causal discovery without background knowledge

```r
set.seed(2019)
N <- 1000 # number of observations

# Generate data from LCD causal structure
# A ==> B ==> C
A <- rnorm(N)
B <- A + rnorm(N)
C <- B + rnorm(N)
dat <- cbind(A, B, C)

# Compute correlation matrix
R <- cor(dat)

# Run the BCCD causal discovery algorithm
capture.output(
  bccd.fit_lcd <- BCCD(R, N),
  file = '/dev/null'
)
```



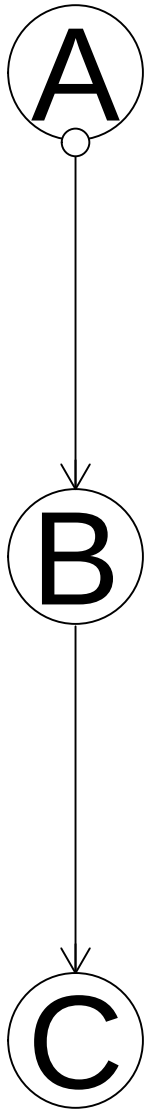
Causal discovery with background knowledge

```
statements <- c(
  "B =/> A   p = 0.95",
  "C =/> A"
)

bkg_info <- express_prior_background_knowledge(
  D = 3, statements, node_names = LETTERS[1:3]
)

# Run the BCCD causal discovery algorithm
capture.output(
  bccd.fit_bkg <- BCCD(
    R, N, bkg_info = bkg_info,
    no_selection_bias = TRUE
  ), file = '/dev/null'
)
```



Causal discovery with background knowledge

```
## Graph on 3 variables: {A}, {B}, {C};
## Only showing inferred statements with p > 0.5;
## [p = 1.00] {A} === {B};
##   [p = 0.95] {B} =/> {A};
## [p = 1.00] {B} === {C};
##   [p = 0.63] {B} ==> {C};
##   [p = 0.57] {C} =/> {B};
```



References

- **RUcausal**: https://gitlab.science.ru.nl/gbucur/RUcausal
- Other packages for causal discovery: *pcalg*, *bnlearn*, *dagitty*
- The **BCCD** algorithm (Claassen and Heskes 2012)

Claassen, Tom, and Tom Heskes. 2012. "A Bayesian Approach to Constraint Based Causal Inference." *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, August. AUAI Press, 207–16.