Causality and biases

Rafał Urbaniak, Nikodem Lewandowski
(LoPSE research group, University of Gdansk)

Some hypotheses to consider

Newsworthiness

"It seems the most newsworthy scientific studies the least trustworthy." What could explain this?

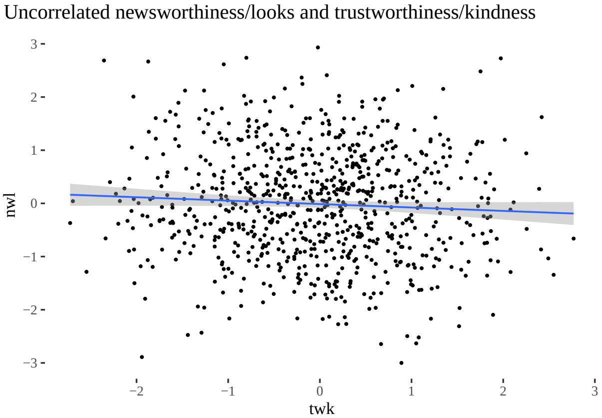
Jerks

"It seems good-looking men are jerks." What could explain this?

Selection-distortion effect

```r
N <- 800 #proposals/candidates
p <- .5  #proportion to select
# uncorrelated newsworthiness/looks and trustworthiness/kindness
nwl <- rnorm(N)
twk <- rnorm(N)
cor(nwl,twk)
```

```
## [1] -0.06152328
```



Selection-distortion effect

```
s <- nwl + twk # total score
q <- quantile( s , 1-p ) # top 10% threshold
selected <- ifelse( s >= q , TRUE , FALSE )
cor( twk[selected] , nwl[selected] )
```

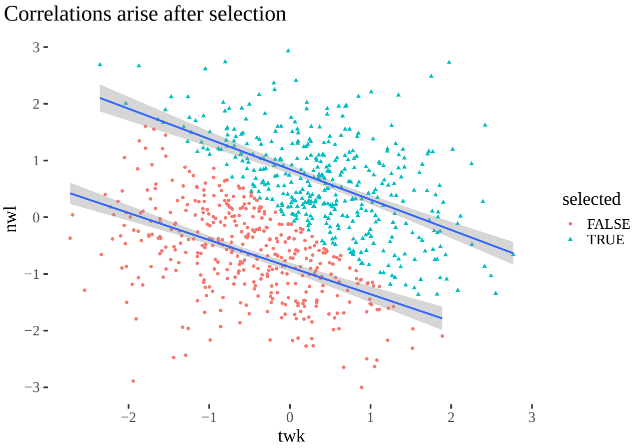
```
## [1] -0.5433415
```

```
cor( twk[!selected] , nwl[!selected] )
```

```
## [1] -0.5040041
```

Selection-distortion effect

```
ggplot() + geom_point(aes(
  x = twk, y = nwl, color = selected, shape = selected))+
geom_smooth(aes(
  x = twk, y = nwl, group = selected), method = "lm")+th+
  ggtitle("Correlations arise after selection")
```



Multiple regression will not save you

```r
news <- list(nwl = nwl, twk = twk,
             sel = as.integer(selected+1))

newsTWK <- quap(
  alist(
    nwl ~ dnorm( mu , sigma ) ,
    mu <- a + t * twk ,
    a ~ dnorm( 0 , 2 ) ,
    t ~ dnorm( 0 , .2 ) ,
    sigma ~ dexp( 1 )
  ) ,
  data= news )

precis(newsTWK)
```

```
##              mean         sd        5.5%        94.5%
## a     -0.01252908 0.03613758 -0.07028391  0.045225751
## t     -0.06223462 0.03627592 -0.12021054 -0.004258699
## sigma  1.02131620 0.02550846  0.98054875  1.062083648
```

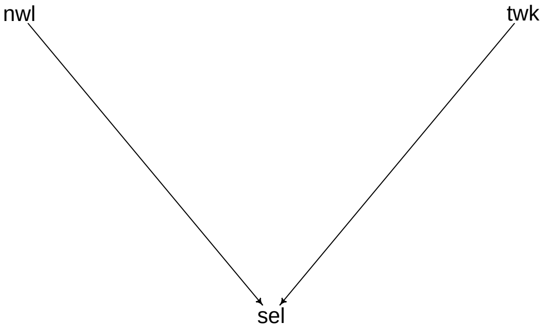
Multiple regression will not save you

```r
newsTWKselected <- quap(
  alist(
    nwl ~ dnorm( mu , sigma ) ,
    mu <- a[sel] + t[sel] * twk ,
    t[sel] ~ dnorm( 0 , .2 ) ,
    a[sel] ~ dnorm(0, 2),
    sigma ~ dexp( 1 )
  ) ,
  data= news )

precis(newsTWKselected, depth  = 2)
```

```
##              mean         sd       5.5%      94.5%
## t[1]   -0.4583950 0.04047583 -0.5230832 -0.3937068
## t[2]   -0.5129769 0.04012847 -0.5771100 -0.4488439
## a[1]   -0.8693598 0.04099812 -0.9348828 -0.8038369
## a[2]    0.8342353 0.03918357  0.7716123  0.8968582
## sigma   0.6967769 0.01741858  0.6689387  0.7246152
```

Collider bias

```r
newsDAG <- dagitty (
  "dag{
  nwl -> sel <- twk
  }"
)
coordinates(newsDAG) <- list(
  x=c(nwl=0,sel=1,twk=2) , y=c(nwl=0,sel=1,twk=0) )
drawdag(newsDAG, cex = 2,
        radius = 3, goodarrow = TRUE, xlim = c(-.2,2.2), ylim = c(-1.2,.2))
```



Post-treatment bias

Blindly tossing in predictors is never a good idea

```
set.seed(21)
# number of subjects
N <- 100
# simulate initial aggression levels
aggression0 <- rnorm(N,1,.4)
#simulate vaccine
vaccine <- rep( 0:1 , each=N/2)
#simulate fungus
cordyceps <- rbinom( N , size=1 , prob=0.95 - vaccine * 0.5 )
# assign vaccines and simulate cordyceps and aggression
aggression1 <- aggression0 + rnorm(N, 1.2 + 2 *cordyceps, .4)
# compose a clean data frame
d <- data.frame( aggression0=aggression0 , aggression1=aggression1 ,
                 vaccine=vaccine , cordyceps=cordyceps )
precis(d)[,-5]
```

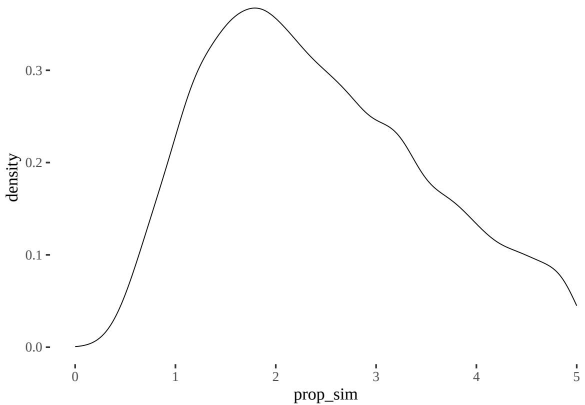
```
##                 mean        sd      5.5%    94.5%
## aggression0 1.028994 0.4108600 0.3452403 1.702803
## aggression1 3.703373 1.0376606 1.8749605 5.038115
## vaccine     0.500000 0.5025189 0.0000000 1.000000
## cordyceps   0.740000 0.4408440 0.0000000 1.000000
```

Choosing prior for proportional change

```
set.seed(23)
prop_sim <- rlnorm( 1e4 , .9 , .6 )
precis( data.frame(prop_sim) )[,-5]
```

```
##                mean       sd      5.5%    94.5%
## prop_sim 2.942808 1.890235 0.9431349 6.351591
```

```
ggplot()+geom_density(aes(x = prop_sim))+xlim(0,5)+th
```



Modeling aggression change

```r
set.seed(12)
aggressionModel <- quap(
  alist(
    aggression1 ~ dnorm( mu , sigma ),
    mu <- aggression0*p,
    p ~ dlnorm( .9 , 0.6 ),
    sigma ~ dexp( 1 )
  ), data=d )

precis(aggressionModel)[,-5]
```

```
##           mean         sd     5.5%    94.5%
## p     3.230631 0.12417941 3.032169 3.429094
## sigma 1.375921 0.09630769 1.222003 1.529840
```

Throwing in all predictors

```r
aggressionModelAll <- quap(
  alist(
    aggression1 ~ dnorm( mu , sigma ),
    mu <- aggression0 * p,
    p <-  a + bv*vaccine + bc*cordyceps,
    a ~ dnorm(1,2),
    bv ~ dnorm( 1 , 1 ),
    bc ~ dlnorm( .9 , 0.6 ),
    sigma ~ dexp( 1 )
  ), data=d )

precis(aggressionModelAll)[,-5]
```

```
##            mean        sd       5.5%     94.5%
## a     1.7361277 0.2838563  1.2824704 2.1897849
## bv    0.2785707 0.2290837 -0.0875492 0.6446906
## bc    1.8454706 0.2607080  1.4288088 2.2621323
## sigma 1.1118171 0.0779973  0.9871624 1.2364719
```

Throwing in all predictors

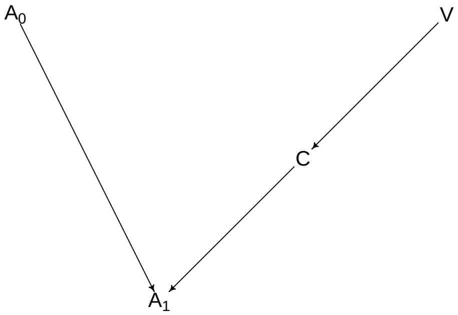
```r
aggressionModelVaccine <- quap(
  alist(
    aggression1 ~ dnorm( mu , sigma ),
    mu <- aggression0 * p,
    p <- a +   bv*vaccine,
    a ~ dnorm(1,2),
    bv ~ dnorm( 1 , 1 ),
    sigma ~ dexp( 1 )
  ), data=d )


precis(aggressionModelVaccine)[,-5]
```

```
##              mean         sd       5.5%      94.5%
## a       3.4774336 0.16402146  3.2152956  3.7395716
## bv     -0.5186644 0.23454769 -0.8935169 -0.1438119
## sigma   1.3340238 0.09356404  1.1844904  1.4835572
```

```r
cordyceps <- rbinom( N , size=1 , prob=0.95 - vaccine * 0.5 )
```

C d-separates V from A1

```
aggressionDAG <- dagitty( "dag {
A_0 -> A_1
C -> A_1
V -> C
}")
coordinates( aggressionDAG ) <- list( x=c(A_0=0,V=1.5,C=1,A_1=.5) ,
                                      y=c(A_0=0,V=0,C=.5,A_1=1) )
drawdag( aggressionDAG,  cex = 2, radius = 3, goodarrow = TRUE,
         xlim = c(-.3,1.7), ylim = c(-1.2,.2))
```



C d-separates V from A1

```
impliedConditionalIndependencies(aggressionDAG)
```

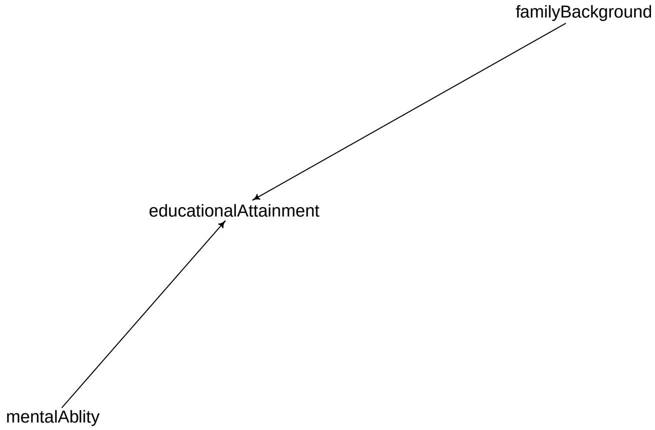
```
## A_0 _||_ C
## A_0 _||_ V
## A_1 _||_ V | C
```

The counterfactual model of data analysis

- individual: i
- potential outcomes: Y_i^0, Y_i^1 (only one observable)
- look at groups, with defendable assumptions estimate the **average effect**

Causality creeps

Example: status attainment tradition



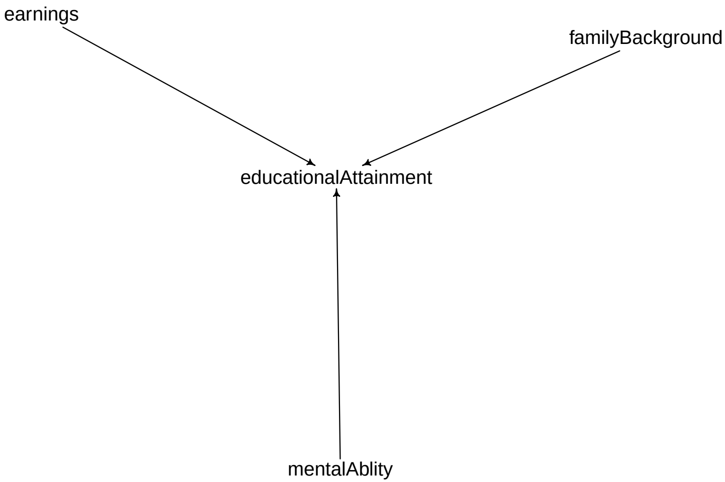
Implicit Wisconsin model

Students follow their own aspirations.

Causality creeps

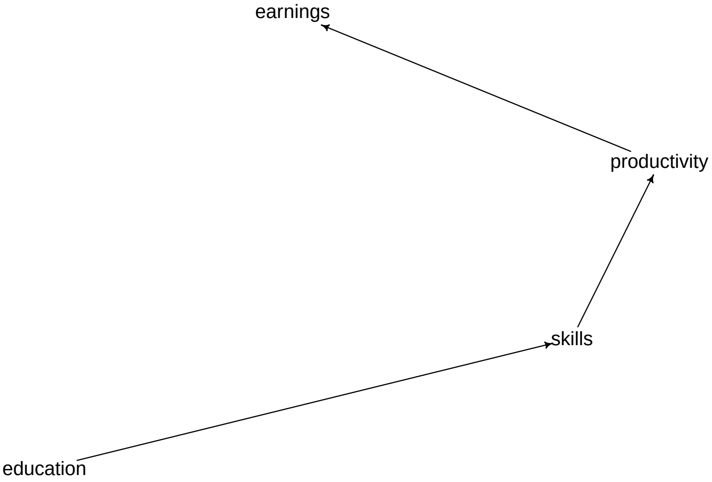
Critics

You can have all aspirations you want, resources will limit you



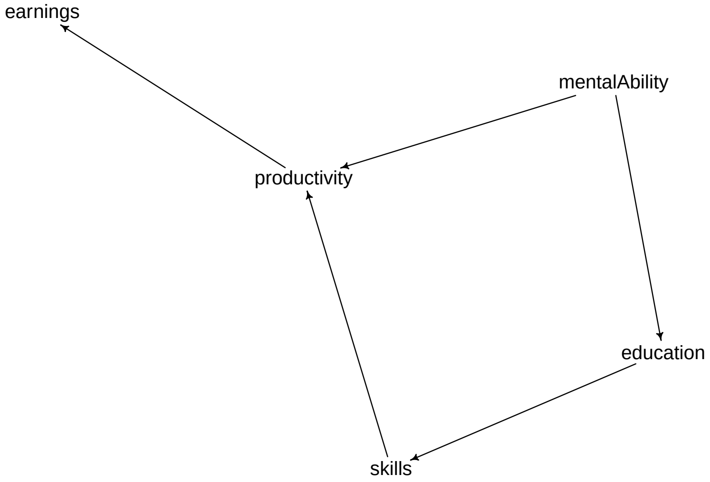
Causality creeps

Example: economic theory of human capital



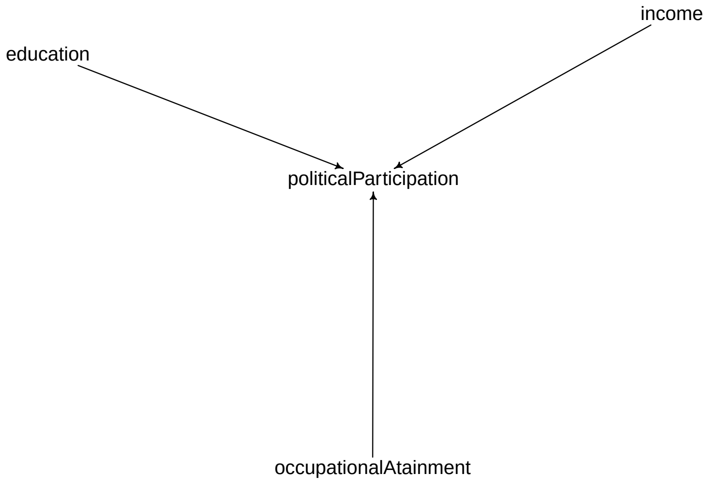
Causality creeps

Example: economic theory of human capital



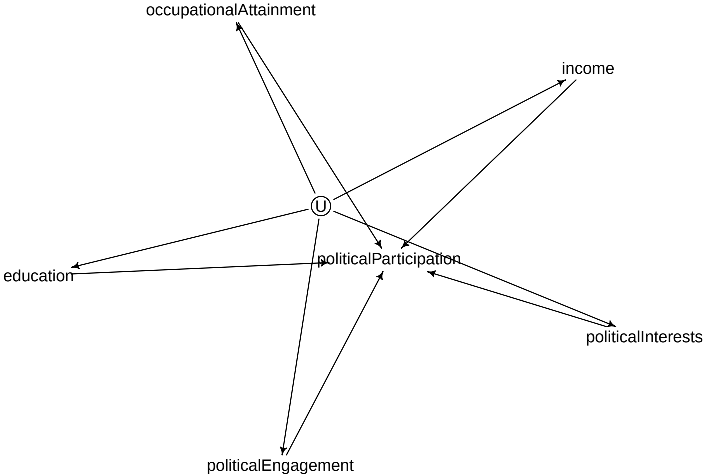
Causality creeps

Example: political participation



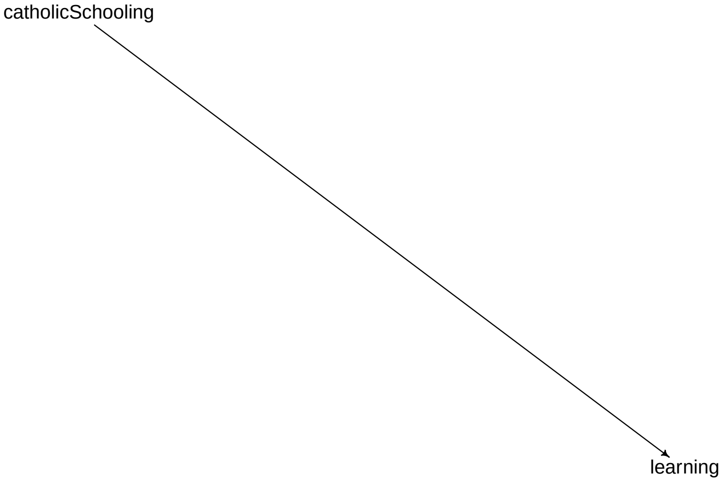
Causality creeps

Example: political participation



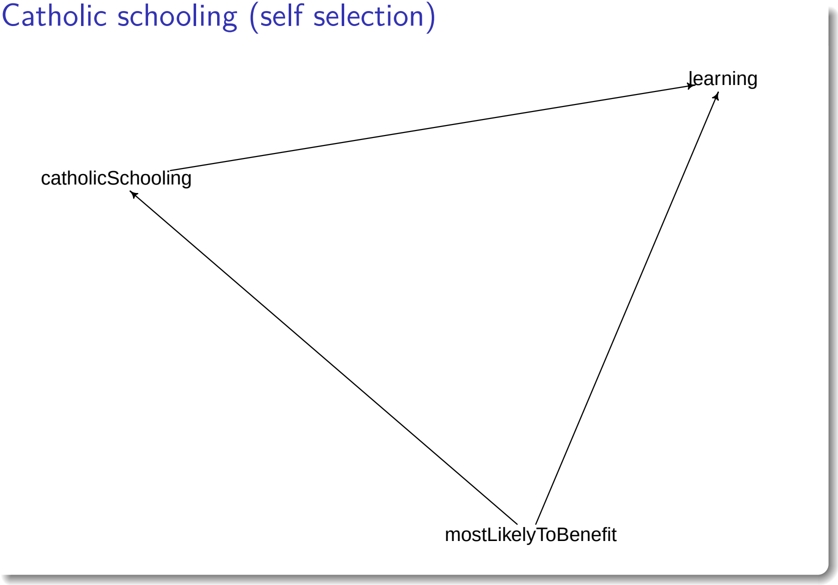
Causality creeps

Example: Catholic schooling



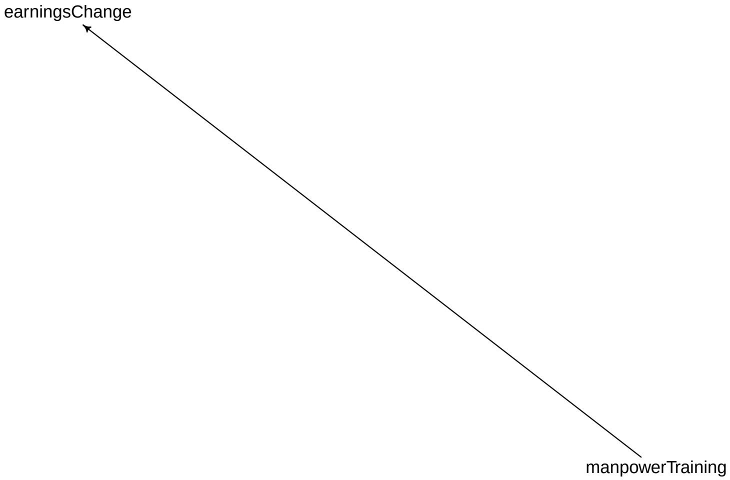
Causality creeps

Catholic schooling (self selection)



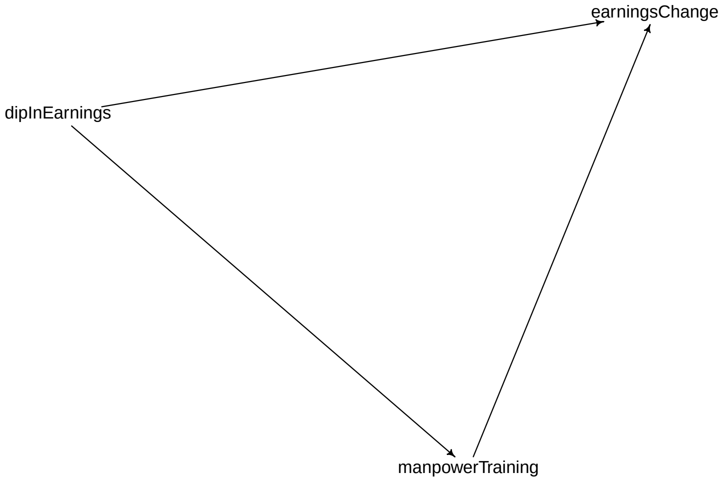
Causality creeps

Example: manpower training



Causality creeps

Manpower training (Ashenfelter's dip)



The straightforward solution

Randomize

Cut the arrows coming into the predictors.

The straightforward solution

Randomize

Cut the arrows coming into the predictors.

The problem

- most data are observational
- randomization is often impossible, impractical, or unethical