# Constructing Kirq, software for set-theoretic social research: A software development travelogue

Claude Rubinson
University of Houston—Downtown
rubinsonc@uhd.edu
cjr@grundrisse.org
http://grundrisse.org

PyHou
Houston, Texas
March 19, 2013

# Overview

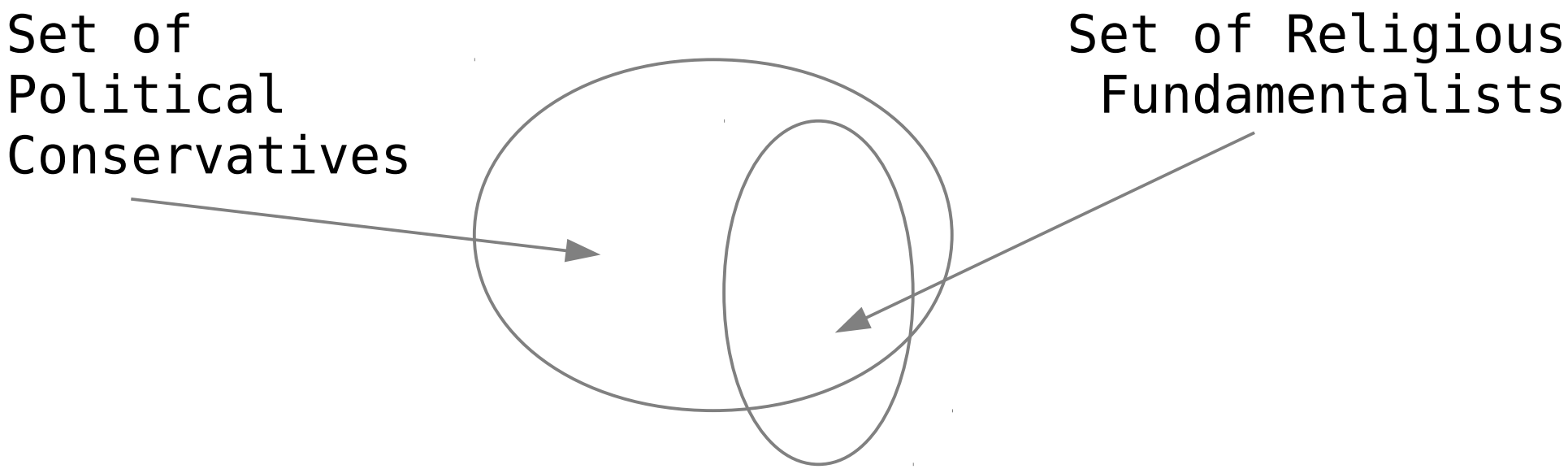
- Introduction to QCA
- History of QCA software
- First (mis-)steps in developing Kirq: the fsQCA package for R
- Use cases and design goals for acq and Kirq
- Python's role in meeting these design goals
- Lessons learned: Using Python for academic software projects

# What is Qualitative Comparative Analysis?

- A method of conducting social research by analyzing subset relationships, using Boolean algebra

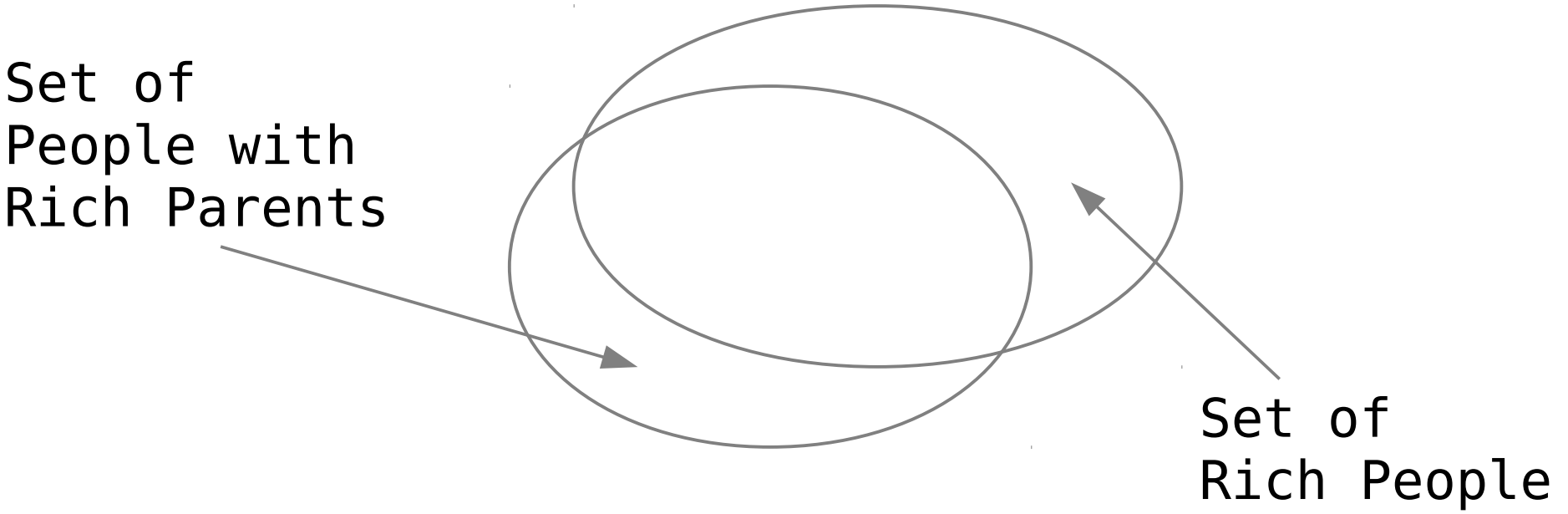
# What is Qualitative Comparative Analysis?

- A method of conducting social research by analyzing subset relationships, using Boolean algebra
- Example: Religious fundamentalists tend to be politically conservative.



# What is Qualitative Comparative Analysis?

- A method of conducting social research by analyzing subset relationships, using Boolean algebra
- Example: Wealthy individuals tend to come from privileged families.

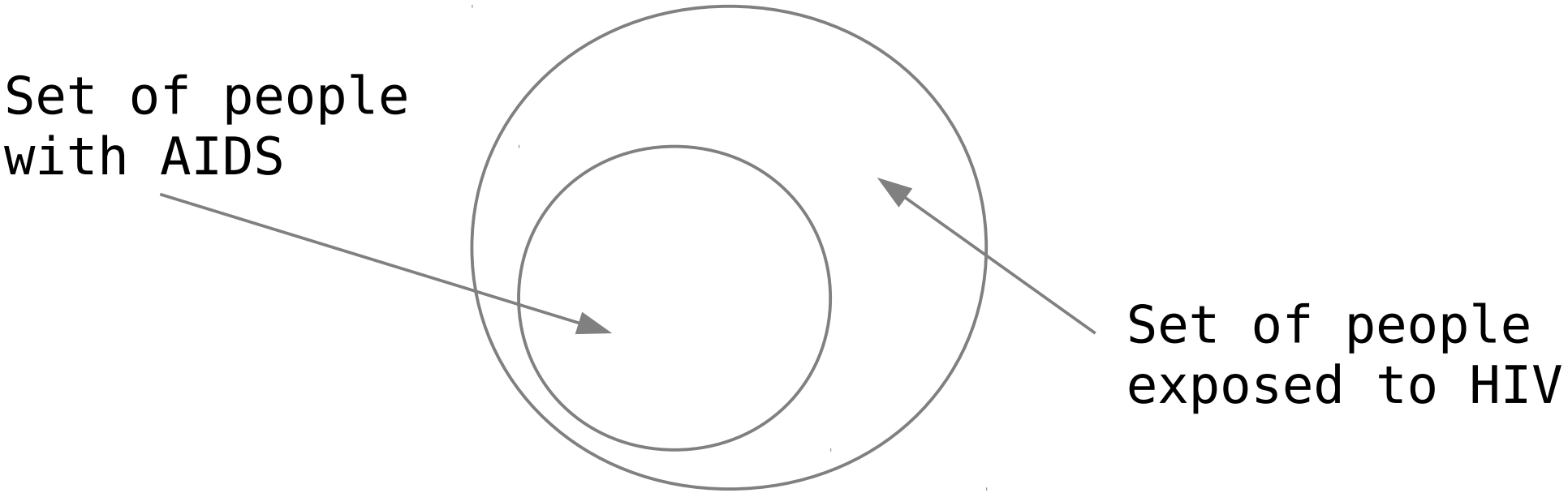


# What is Qualitative Comparative Analysis?

- Particularly concerned with two types of causal relationships: necessary conditions and sufficient conditions

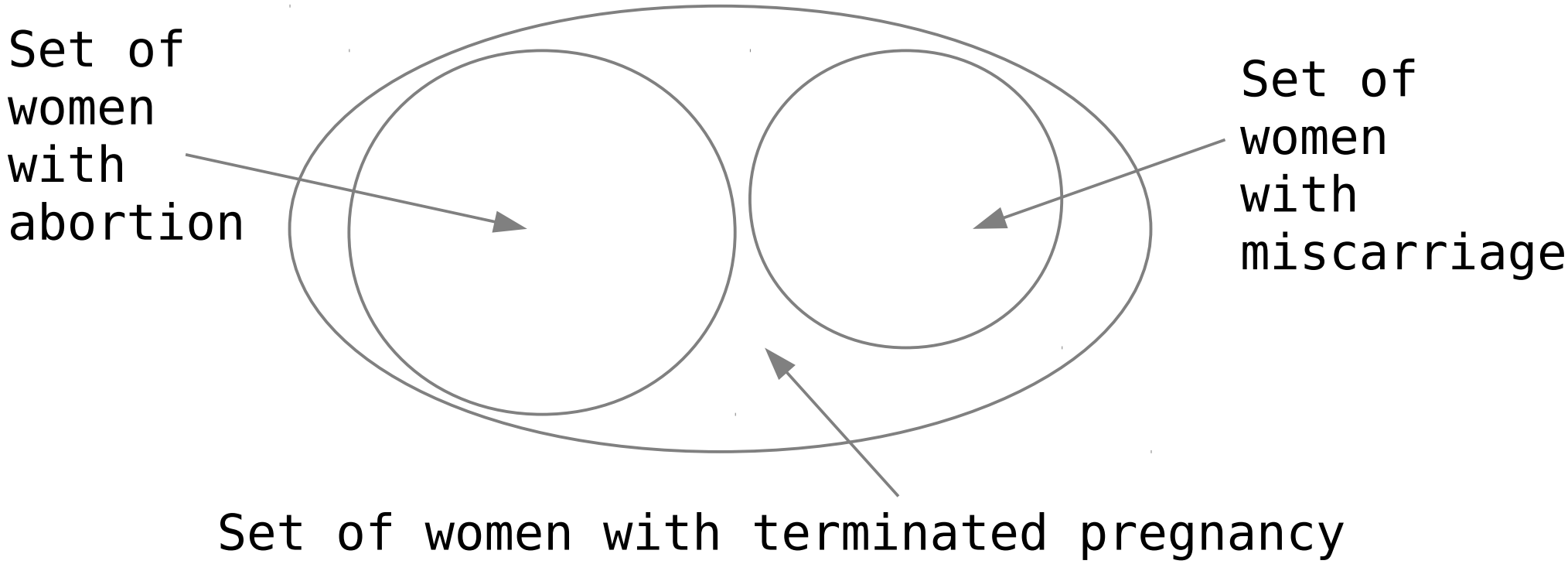
# What is Qualitative Comparative Analysis?

- Necessary condition: cause must be present for outcome to occur
- Example: Must be exposed to HIV to contract AIDS



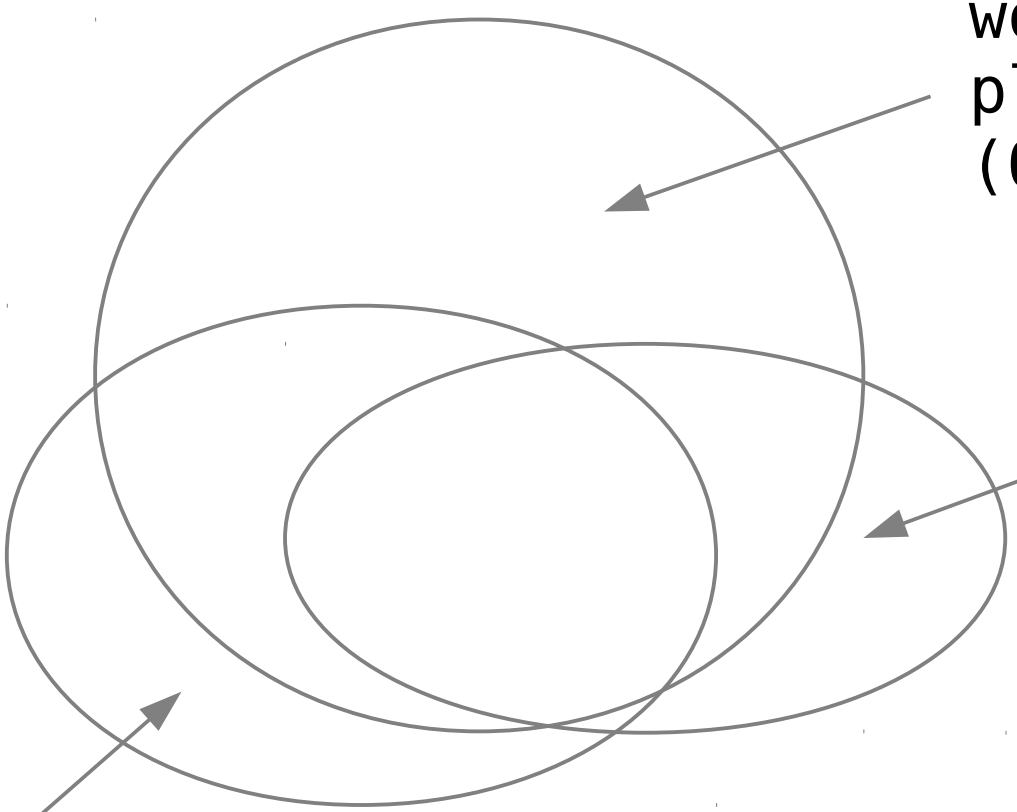
# What is Qualitative Comparative Analysis?

- Sufficient condition: if cause occurs, outcome will occur
- Example: Abortion *or* miscarriage will terminate pregnancy



# What is Qualitative Comparative Analysis?

- Sufficient condition: if cause occurs, outcome will occur



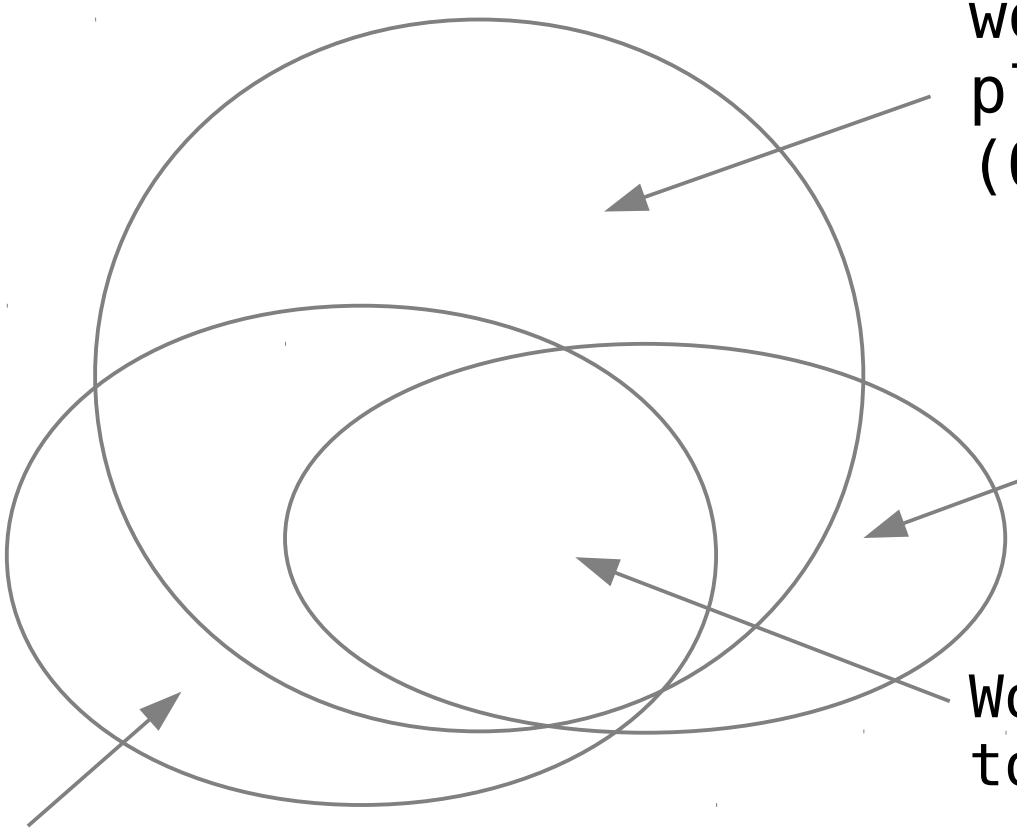
Recently deported women who do not plan to cross again (Outcome)

High SES women who haven't lived in the U.S. and aren't traveling with family ($X_1$)

High SES women who haven't lived in the U.S., have only attempted cross a few times and felt that their last crossing experience was very dangerous ($X_2$)

# What is Qualitative Comparative Analysis?

- Sufficient condition: if cause occurs, outcome will occur



Recently deported women who do not plan to cross again (Outcome)

High SES women who haven't lived in the U.S. and aren't traveling with family ($X_1$)

Women belonging to sets $X_1$ and $X_2$

High SES women who haven't lived in the U.S., have only attempted cross a few times and felt that their last crossing experience was very dangerous ($X_2$)

# What is Qualitative Comparative Analysis?

- Challenges conventional statistical analysis, which is based upon a linear-additive model
- Complements other set-theoretic research methods (e.g., SNA and QNA)
- Does not depend upon degrees of freedom, so is useful for small-, medium-, and large-N studies
- Encourages research process that is “retroductive” and “case-oriented”

# What is Qualitative Comparative Analysis?

## Example: Brown and Boswell (1995)

Truth Table with Contradiction (from Table 4 of Brown and Boswell 1995)

| Recent Black Migrants | Weak Union | Black Strikebreaking | Observations |
|---|---|---|---|
| T | T | T | East Chicago, Pittsburgh, Youngstown |
| T | F | Con | Buffalo, Chicago, Gary, Johnstown, [Cleveland] |
| F | T | F | Bethlehem, Joliet, McKeesport, Milwaukee, New Castle, Reading |
| F | F | F | Decatur, Wheeling |

# What is Qualitative Comparative Analysis?

## Example: Brown and Boswell (1995)

Revised Truth Table without Contradiction (from Table 5 of Brown and Boswell 1995)

| Recent Black Migration | Weak Union | Local Govt Repression | Black Strikebreaking | Observations |
|---|---|---|---|---|
| T | T | T | T | East Chicago, Pittsburgh, Youngstown |
| T | T | F | — | |
| T | F | T | T | Buffalo, Chicago, Gary, Johnstown |
| T | F | F | F | Cleveland |
| F | T | T | F | Bethlehem, Joliet, McKeesport, New Castle, Reading |
| F | T | F | F | Milwaukee |
| F | F | T | F | Decatur |
| F | F | F | F | Wheeling |

# What is Qualitative Comparative Analysis?

## Example: Brown and Boswell (1995)

Revised Truth Table without Contradiction (from Table 5 of Brown and Boswell 1995)

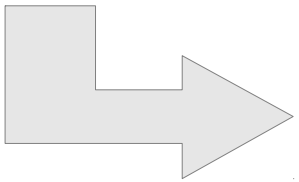
| Recent Black Migration | Weak Union | Local Govt Repression | Black Strikebreaking | Observations |
|---|---|---|---|---|
| T | T | T | T | East Chicago, Pittsburgh, Youngstown |
| T | T | F | — | |
| T | F | T | T | Buffalo, Chicago, Gary, Johnstown |
| T | F | F | F | Cleveland |
| F | T | T | F | Bethlehem, Joliet, McKeesport, New Castle, Reading |
| F | T | F | F | Milwaukee |
| F | F | T | F | Decatur |
| F | F | F | F | Wheeling |

# What is Qualitative Comparative Analysis?

## Example: Brown and Boswell (1995)

Revised Truth Table without Contradiction (from Table 5 of Brown and Boswell 1995)

| Recent Black Migration | Weak Union | Local Govt Repression | Black Strikebreaking | Observations |
|---|---|---|---|---|
| T | T | T | T | East Chicago, Pittsburgh, Youngstown |
| T | F | T | T | Buffalo, Chicago, Gary, Johnstown |



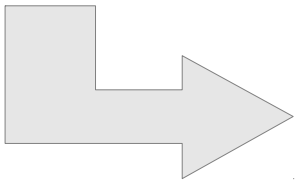
```
RBM * WU * LGR +
RBM * ~WU * LGR
        = Black Strikebreaking
```

# What is Qualitative Comparative Analysis?

## Example: Brown and Boswell (1995)

Revised Truth Table without Contradiction (from Table 5 of Brown and Boswell 1995)

| Recent Black Migration | Weak Union | Local Govt Repression | Black Strikebreaking | Observations |
|---|---|---|---|---|
| T | T | T | T | East Chicago, Pittsburgh, Youngstown |
| T | F | T | T | Buffalo, Chicago, Gary, Johnstown |

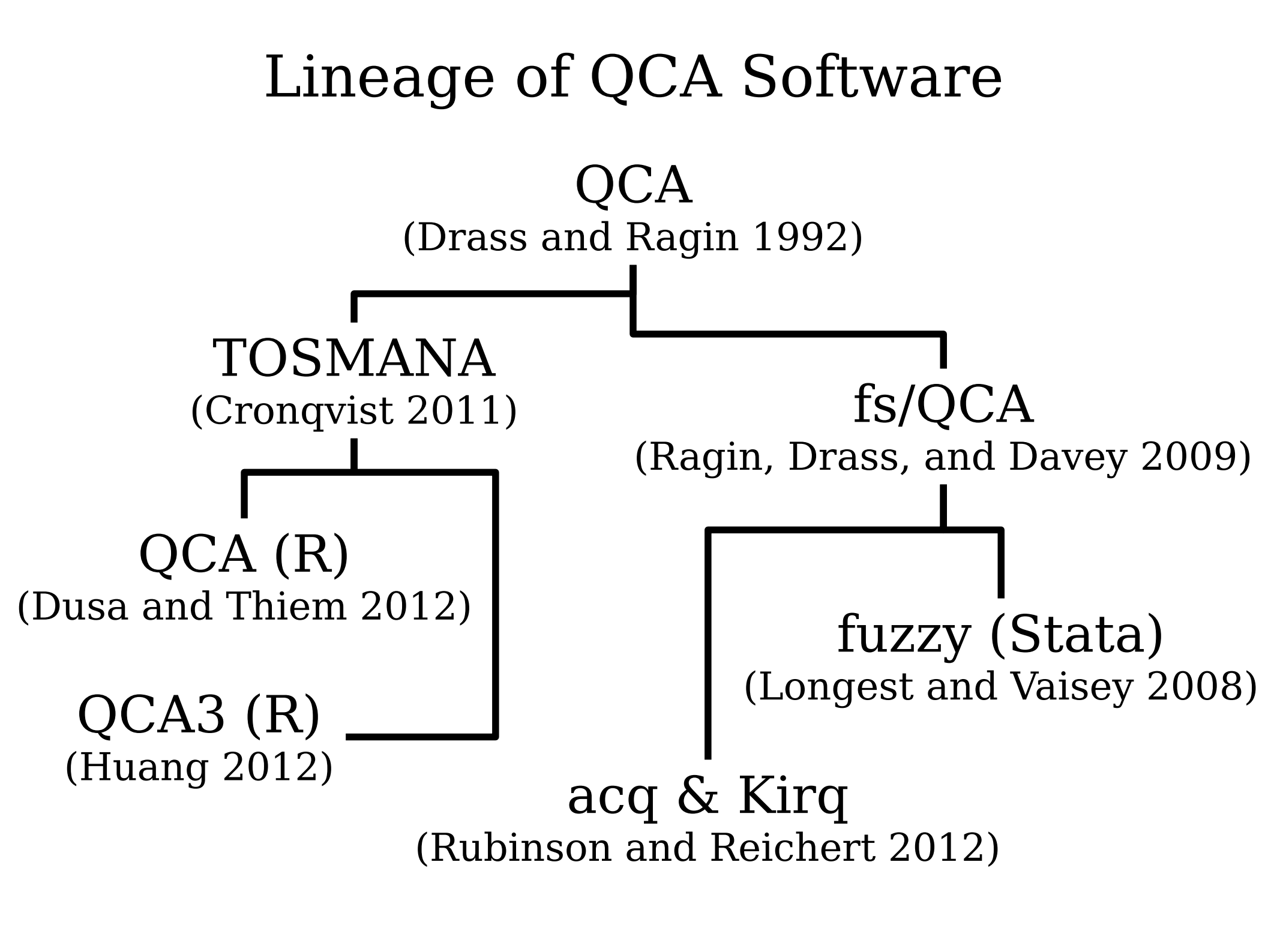


```
RBM *  WU * LGR +
RBM * ~WU * LGR
                = Black Strikebreaking

RBM * LGR       = Black Strikebreaking
```

# Technical and Usability Challenges

- QCA algorithms are:
  - NP-hard (no exact algebraic solution)
  - $O(2^N)$ complexity, where $N$ indicates the number of variables (not observations) in the data set
- Because data sets tend to be small and matrix algebra isn't used, no use for NumPy
- How to maintain and encourage retroductive, case-oriented research process?
- How to make software that's efficient, useful, and usable?


Lineage of QCA Software
QCA
(Drass and Ragin 1992)
TOSMANA
(Cronqvist 2011)
fs/QCA
(Ragin, Drass, and Davey 2009)
QCA (R)
(Dusa and Thiem 2012)
QCA3 (R)
(Huang 2012)
fuzzy (Stata)
(Longest and Vaisey 2008)
acq & Kirq
(Rubinson and Reichert 2012)

# “Plan to throw one away; you will, anyhow”

- fsQCA module for R
  - Cross-platform, but requires R
  - Not user-friendly
  - Too slow
  - R programming “considered harmful”
  - But: allowed me to realize that the user interface should be task-oriented

# “Plan to throw one away; you will, anyhow”

- fsQCA module for R
  - **Cross-platform, but requires R**
  - **Not user-friendly**
  - Too slow
  - R programming “considered harmful”
  - But: allowed me to realize that the user interface should be task-oriented

# “Plan to throw one away; you will, anyhow”

- fsQCA module for R
  - Cross-platform, but requires R
  - Not user-friendly
  - **Too slow**
  - R programming “considered harmful”
  - But: allowed me to realize that the user interface should be task-oriented

# “Plan to throw one away; you will, anyhow”

- fsQCA module for R
  - Cross-platform, but requires R
  - Not user-friendly
  - Too slow
  - **R programming “considered harmful”**
  - But: allowed me to realize that the user interface should be task-oriented

# “Plan to throw one away; you will, anyhow”

- fsQCA module for R
  - Cross-platform, but requires R
  - Not user-friendly
  - Too slow
  - R programming “considered harmful”
  - **But: allowed me to realize that the user interface should be task-oriented**

# Use Cases for acq and Kirq

- acq: QCA at the Unix commandline
  - a “scratch my own itch” project
- Kirq: QCA for everybody else
  - a user-friendly, crossplatform GUI program

# Design Goals for acq and Kirq

- Software that is efficient, useful, and usable:
  - Conform to the Unix philosophy
  - Good “out of the box” performance, plus ability to optimize performance
  - Support and encourage good QCA research practices
  - Kirq should be crossplatform and user-friendly
- Also important: Avoid sucking up all of my time

# Why Python?

- The surrounding ecosystem
  - Ability to hire others
  - Confidence that the supporting environments is stable and will continue to be maintained
  - Python is *lingua franca* in academia
  - Rich environment for GUI toolkits, installers, etc.
  - Chose Qt for GUI toolkit and PyInstaller for installer

# Design Goal: Avoiding a time sink

- Relatively easy to recruit and hire good programmers
- Easy to mix procedural and OOP programming
- Official online documentation remains top notch
- The core Python language remains relatively compact
  - but not the standard library, and certainly not the surrounding environment (PyPI, etc)

# Design Goal: Conform to the Unix Philosophy

- Build a prototypes as soon as possible
- Small is beautiful/do one thing well
  - acq's GUI scripts: gtt and concov
  - have resisted adding a data editor to Kirq
  - decided to offload Kirq's visualizations
- Make every program a filter
  - Because Kirq can read data from the commandline, it's easy for other programs to call out to it

# Design Goal: Good “out of the box” performance and ability to optimize

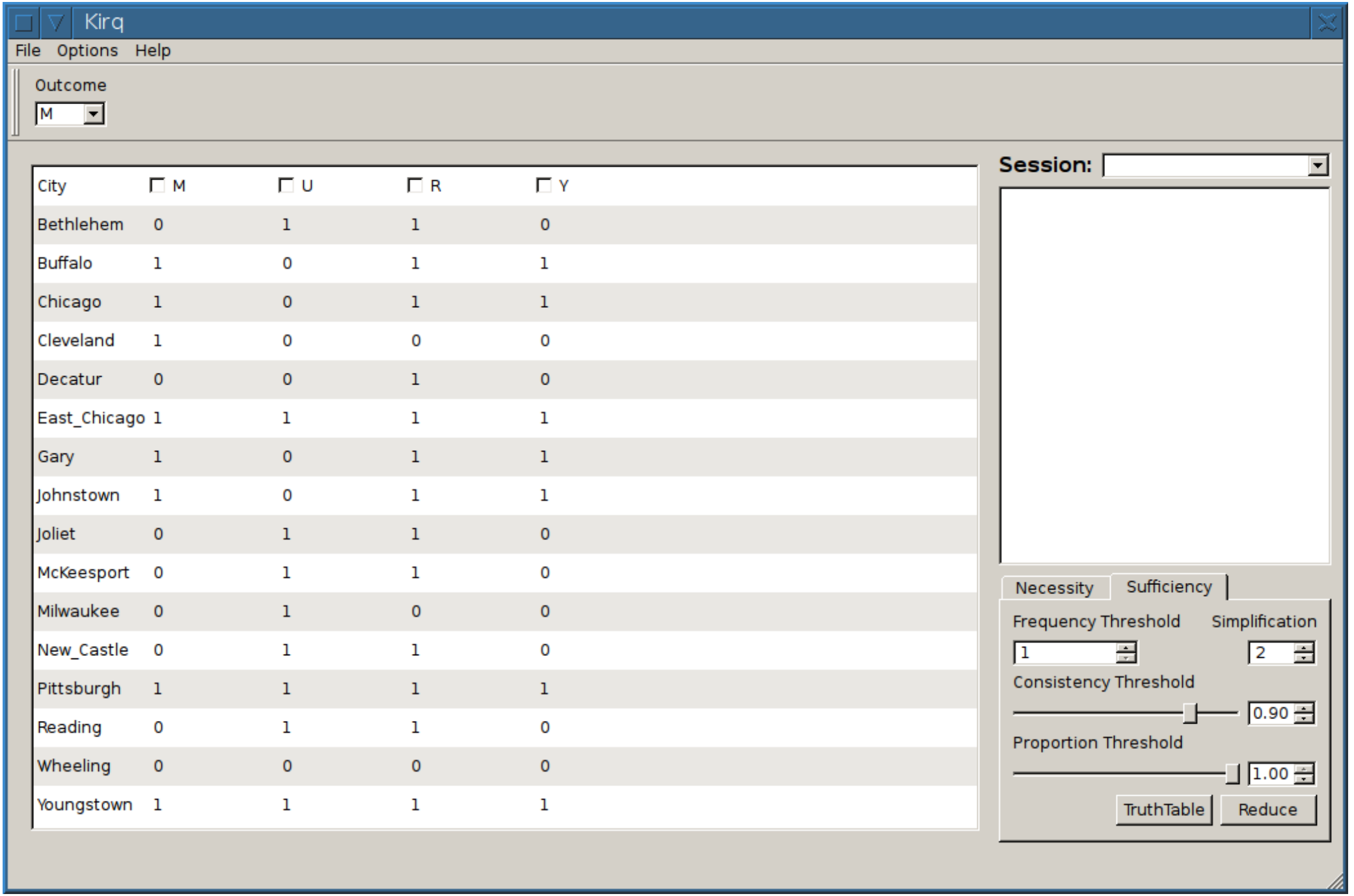
- acq had fewer lines of code than fsQCA for R module, and was faster
  - compare to QCA module for R
- Good tools for profiling
- Some standard, well documented practices for improving performance, although Python optimization often requires expertise
- Promise of projects such as PyPy

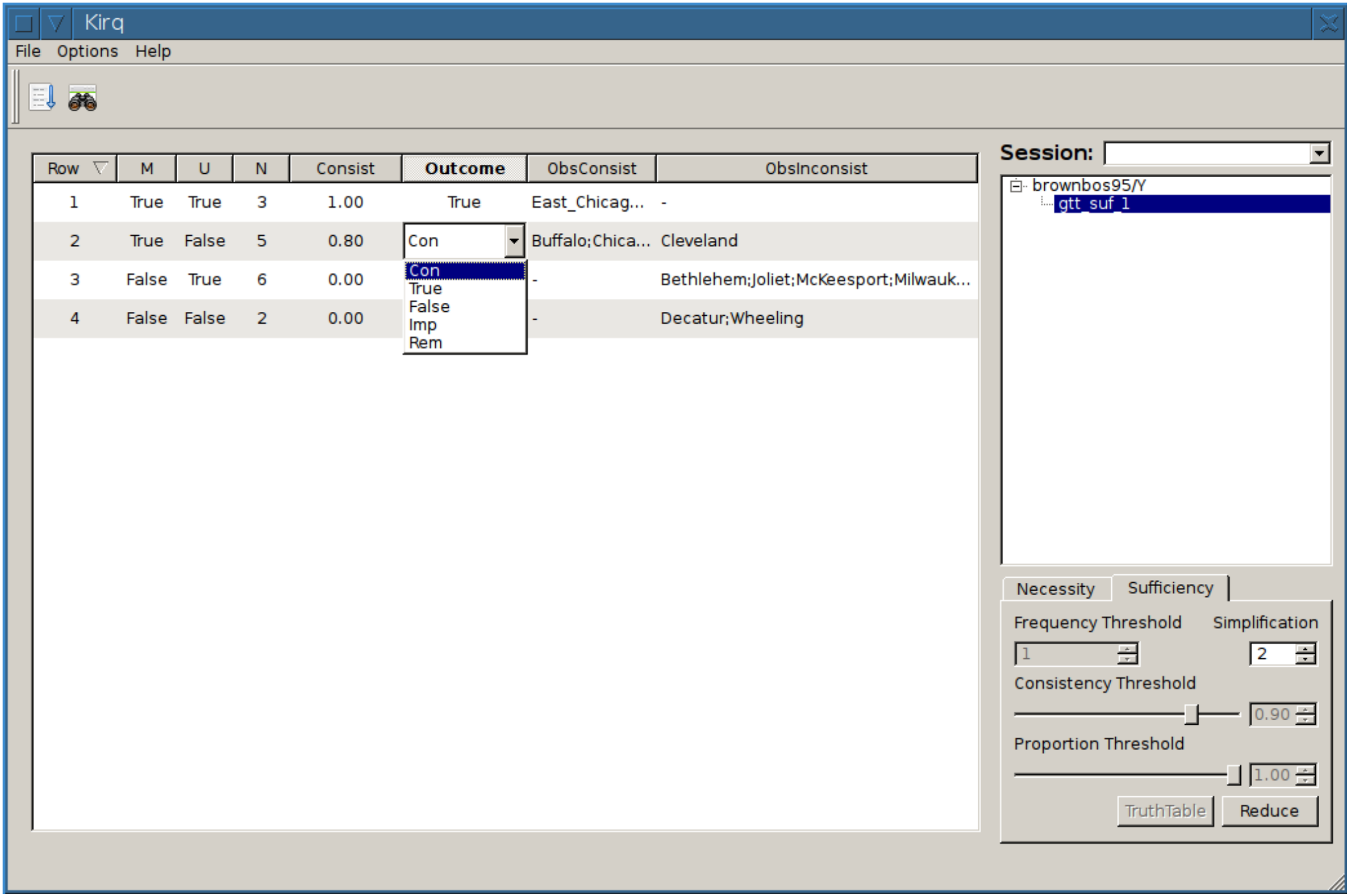
# Design Goal: Support and encourage good QCA research practices

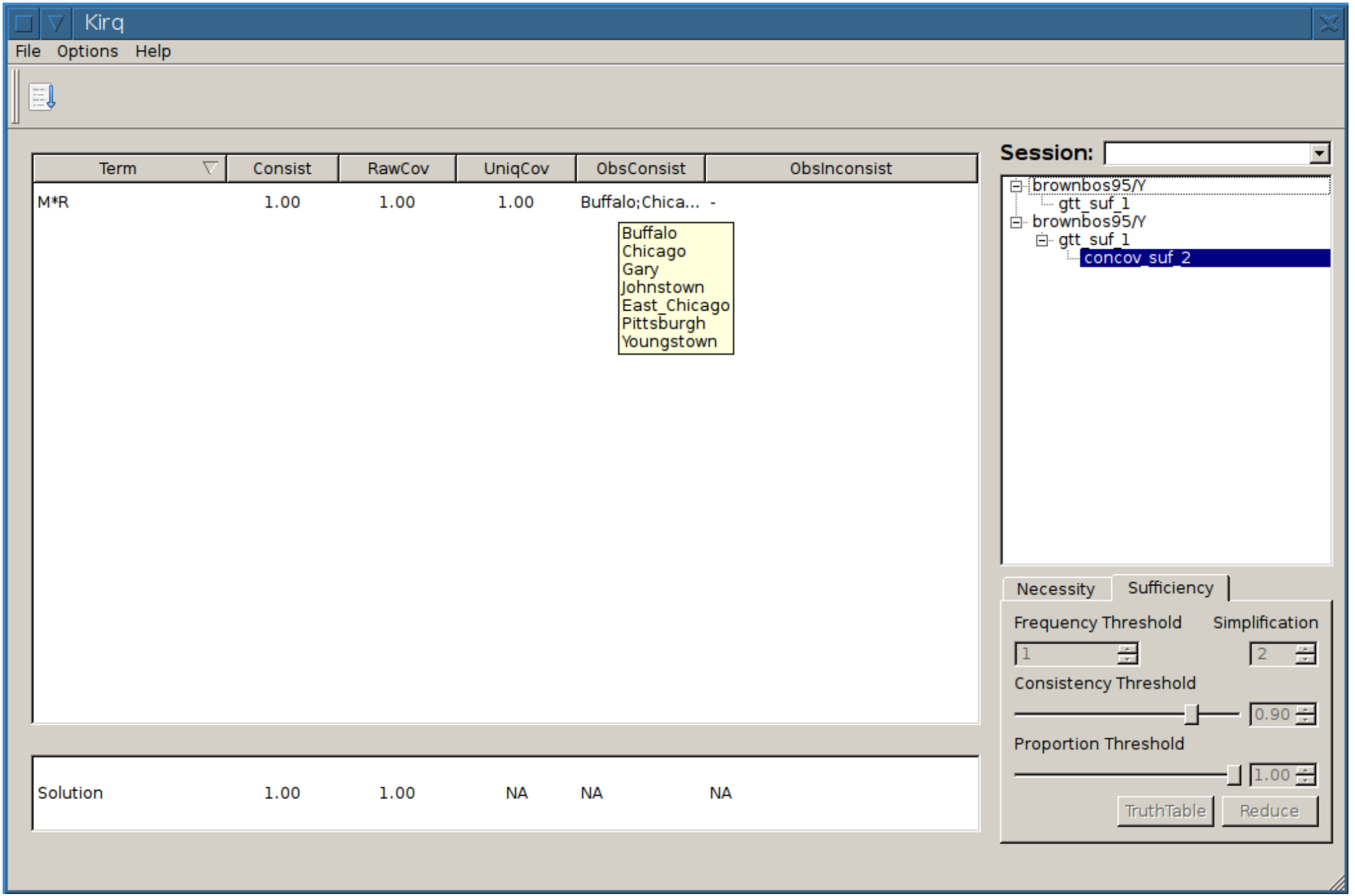
- Less concern for performance means more attention to user-interface issues
- Writing acq as Unix shell scripts helped me streamline the QCA analysis; both acq and Kirq make it easy to modify and rerun analyses
- Have designed Kirq to facilitate interrogation and comparisons of solutions
- Lots of GUI niceties, such as tooltips and pop-out windows
- Importance of “eating your own dogfood”

# Design Goal: Kirq should be cross-platform and user-friendly

- Standardized on Python 2.7, for PyQt
- Only minor compatibility issues with PyQt bindings and OSX, and none with Windows – Kirq always feels native
- Could never build Qt for OSX; used MacPorts instead (slow, but works well)
- PyInstaller works really well, but needed to use development branch for OSX (dev branch is now stable)
- Session history is Kirq's killer feature


Kirq
File
Options
Help
Outcome
M
City
M
U
R
Y
Bethlehem 0 1 1 0
Buffalo 1 0 1 1
Chicago 1 0 1 1
Cleveland 1 0 0 0
Decatur 0 0 1 0
East_Chicago 1 1 1 1
Gary 1 0 1 1
Johnstown 1 0 1 1
Joliet 0 1 1 0
McKeesport 0 1 1 0
Milwaukee 0 1 0 0
New_Castle 0 1 1 0
Pittsburgh 1 1 1 1
Reading 0 1 1 0
Wheeling 0 0 0 0
Youngstown 1 1 1 1
Session:
Necessity
Sufficiency
Frequency Threshold
Simplification
1
2
Consistency Threshold
0.90
Proportion Threshold
1.00
TruthTable
Reduce


Kirq
File
Options
Help
Row
M
U
N
Consist
Outcome
ObsConsist
ObsInconsist
1
True
True
3
1.00
True
East_Chicag...
-
2
True
False
5
0.80
Con
Buffalo;Chica...
Cleveland
3
False
True
6
0.00
-
Bethlehem;Joliet;McKeesport;Milwauk...
4
False
False
2
0.00
-
Decatur;Wheeling
Con
True
False
Imp
Rem
Session:
brownbos95/Y
gtt_suf_1
Necessity
Sufficiency
Frequency Threshold
Simplification
1
2
Consistency Threshold
0.90
Proportion Threshold
1.00
TruthTable
Reduce


Kirq
File
Options
Help
Term
Consist
RawCov
UniqCov
ObsConsist
ObsInconsist
M*R
1.00
1.00
1.00
Buffalo;Chica...
-
Buffalo
Chicago
Gary
Johnstown
East_Chicago
Pittsburgh
Youngstown
Solution
1.00
1.00
NA
NA
NA
Session:
brownbos95/Y
gtt_suf_1
brownbos95/Y
gtt_suf_1
concov_suf_2
Necessity
Sufficiency
Frequency Threshold
Simplification
1
2
Consistency Threshold
0.90
Proportion Threshold
1.00
TruthTable
Reduce

# Lessons Learned – Python's Advantages (for Academic Projects)

- Core language is relatively compact, with excellent documentation
- Relatively easy to find developers
- Strong, well-developed environment of GUI toolkits, installers, etc.
- Good performance out of the box, with ability to optimize if necessary

# Lessons Learned – Python's Disadvantages (for Academic Projects)

- Package distribution is a mess, as is associated documentation
- Churn in the standard library is too rapid to keep up with for a part-time developer
- Dead tree documentation is lousy
- Online signal-to-noise ratio is low
- Python community online is too insular; overly concerned with “idiomatic Python”