Data Preparation for EventFlow:  
A Case Study of over 12000 Daily Reports from Women in Abusive Relationships  
by Seth Powsner & Tami Sullivan  
Dept of Psychiatry, Emergency Medicine & Center for Medical Informatics  
Yale University School of Medicine

THANKS are due:  
Megan Monroe, Rongjian Lan, Catherine Plaisant, Ben Shneiderman and the rest of HCIL who have been very friendly and helpful during my sabbatical and since.

PHOTOS are from Flickr Creative Commons (credits at end).

Overall Data Transform

Data Set – Stat Pkg Format
262 Columns
14004 Rows + Header
IPV - Intimate Partner Violence
aka Domestic Violence, Wife Beating, Spousal Abuse
Includes Physical, Sexual, & Psychological Violence
CDC NIPSVS 2010 revealed that
36% of women, 29% of men victims
including severe physical violence
against 24% women, 14% men

IPV - a Public Health Problem
Fellow travelers: Drug Abuse, HIV
Collateral damage: Children
Treatment? Prevention? Etiology?
What’s really happening?
Arguments-> Fights
Drugs / Drink -> Fights
Fights -> Drugs / Drink

Experience Sampling Method
Beeper study too dangerous, so--
Diary vs Call-in vs Monthly Review
Sullivan, Khondkaryan, et al 2011
147 women, 90 days -> 13230
+ some confirmatory reviews
- report failures
14004 data records in raw data set
-------> 62% days, no IPV

Easy Right?
Just some basic Data Prep
Little Format Conversion
--> [ sub# | Fight | start time | end time ]

(first page of form)
Basic IPV Form – 2 pages

- Page 1 -

<table>
<thead>
<tr>
<th>Date</th>
<th>Fight details</th>
<th>Together</th>
<th>Saw</th>
<th>Arrest</th>
<th>Treatment</th>
</tr>
</thead>
</table>

- Page 2 -

<table>
<thead>
<tr>
<th>Arugments</th>
<th>Drinking</th>
<th>Drugs</th>
<th>Unwanted Sex</th>
<th>Other</th>
</tr>
</thead>
</table>

Basics: 130 fields (81 date time)

Events: $5 \times (4 \times 4 + 1) \rightarrow 85 + 20$

- Fights, Arguments, Sex, Drugs, Drink up to 4 times / day
  - start, am/pm, stop, am/pm
  - + confirming count
  - + $4 \times (4$ fight modifiers, 1 drink cnt
- Upset-by x3, Psych abuse x4, Drugs x7
- Together, In-treatment, In-jail ->6
- Date, Subject-id, Survey misc ->5

Data Prep Work
First pass - Quick & dirty

*Clojure* (Java based *Lisp* variant)
*Python*, others would be fine

Positives - able to display small sample
revealed basic translation difficulty
revealed time scale problems

Negatives - code soon unmanageable
didn’t catch data field errors

Time Matters

Human time is not Computer time
Study day starts at 6am
- study day spans 2 calendar days

Calendrical calc is hard, often wrong
Feb 28 + 1d -> Feb 29 or Mar 1

**Time Is Not** (around 2a, United States)
  *Monotonic*, Nov 2, 1, 7

Time Matters – A Lot

Happy data records are mostly alike
Every unhappy data record is
unhappy in its own way
  [after Tolstoy]
Over time, all records diverge
  EventFlow - ->
  Confetti
  Out of Memory

Time Matters - Unsolved

Survey time scales of
  90 day: Hopeless at 20 events/day
  Month: Doubtful, but interesting
  Week: Not good so far
  Day: Yes (need a uniform day)
What Time is “we’re still Together”?
Together as Event? Parameter?
Got Schema?
Consider Database over Unit Records

Disadvantages
- Initial investment
- Memory or Database requirements

Easier to
- Test data integrity
- Try different time scales
- Segregate various subpopulations
- Make data handles for EventFlow

What Time is “we’re still Together”?
Together as Event? Parameter?

Currently
Well along in rewrite that
- Keeps records in memory
- easy data review
- easy event regeneration
  ie, 2-20 min to code, <5 sec run

Uses Joda Time package
Hoping to clarify of drugs <-> fights
Photo Credits & References

Collin Campbell "the clutterbells" on Flickr
Warrior in Kitchen
Creative Commons Attribution-ShareAlike 2.0 Generic
Accessed 2014.05.24 13:00 UTC
https://www.flickr.com/photos/clutterbells/5911845428

Josh DiMauro
Moleskine Concept Diagram 1
CC Attribution-NonCommercial-NoDerivs 2.0 Generic
Accessed 2014.05.24 13:30 UTC
https://www.flickr.com/photos/jazzmasterson/3038597

kitchener.lord
Eterna Les Historiques Calendar Chronograph
CC Attribution, Non-Comercial, No Derivative
Accessed 2014.05.24 12:20 UTC
https://www.flickr.com/photos/27862259@N02/5838852411/

REFERENCES

NISVS 2010 Summary Report
Accessed 2014.05.24 23:30 UTC

Sullivan TP, Khondkaryan E, Dos Santos NP, Peters E.
Applying Experience Sampling Methods to Partner Violence Research:
Safety and Feasibility in a 90-Day Study of Community Women
Violence Against Women 17(2) 251–266 2011

Sullivan TP, McPartland TS, Armeli S, Jaquier V, Tenne H.
Is it the Exception or the Rule? Daily Co-Occurrence of Physical, Sexual, and
Psychological Partner Violence in a 90-Day Study of Substance-Using,
Community Women

Calendrical Calculations, 3rd ed;
Dershowitz & Reingold, Cambridge Univ Pres 2008
(preface, page XX, lists a number of publicly known serious and expensive calendrical miscalculations)

Joda Time
Accessed 2014.05.26 17:30 UTC
http://www.joda.org/joda-time/
(see also Java SE 8 Date and Time
http://www.oracle.com/technetwork/articles/java/jf14-date-time-2125367.html )