Publication quality tables in Stata using `tabout`

Ian Watson

Macquarie University & SPRC UNSW

Stata User Group Meeting
Sydney
29 September 2016

Overview

- What is `tabout`: quick tour
- Background to `tabout`
- Who `tabout` is for
- What makes for a good table
- Reproducible research & single source publishing
- `tabout` in practice
- New features in `tabout`
- Extending `tabout` with simple programming
- User feedback and requests

Quick tour

- Illustrates:
  - aesthetics
  - ease of use
  - design principles
  - reproducibility
  - **new feature**: integration with Word and Excel
  - **new feature**: easier use with LaTeX

Aesthetics I

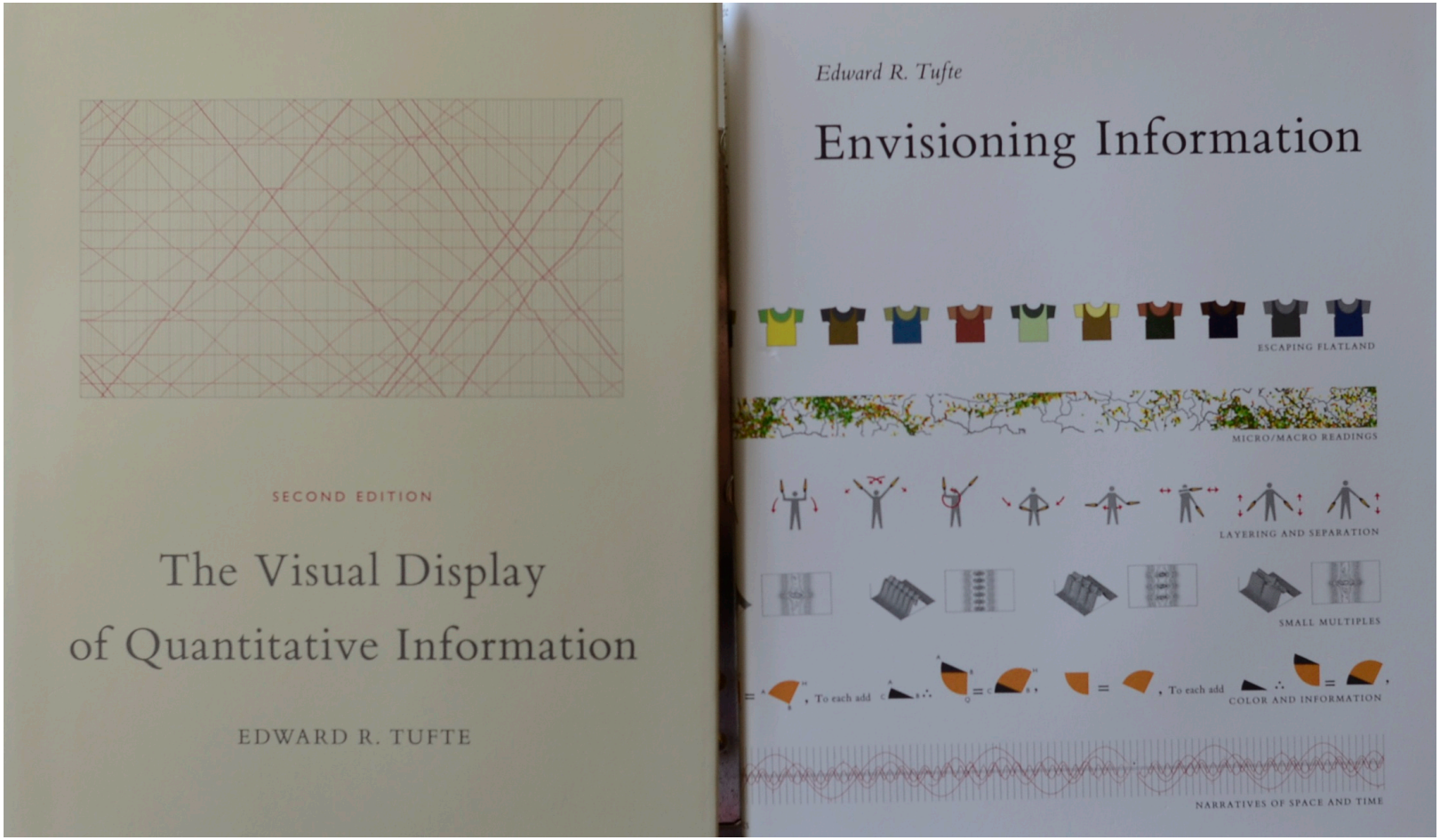
- More than beauty: encoding data and decoding information
- Theory most developed for graphics, but applicable to tables

- William Cleveland, **Visualizing Data** (Hobart Press, 1993)
- Website: http://www.stat.purdue.edu/ wsc/

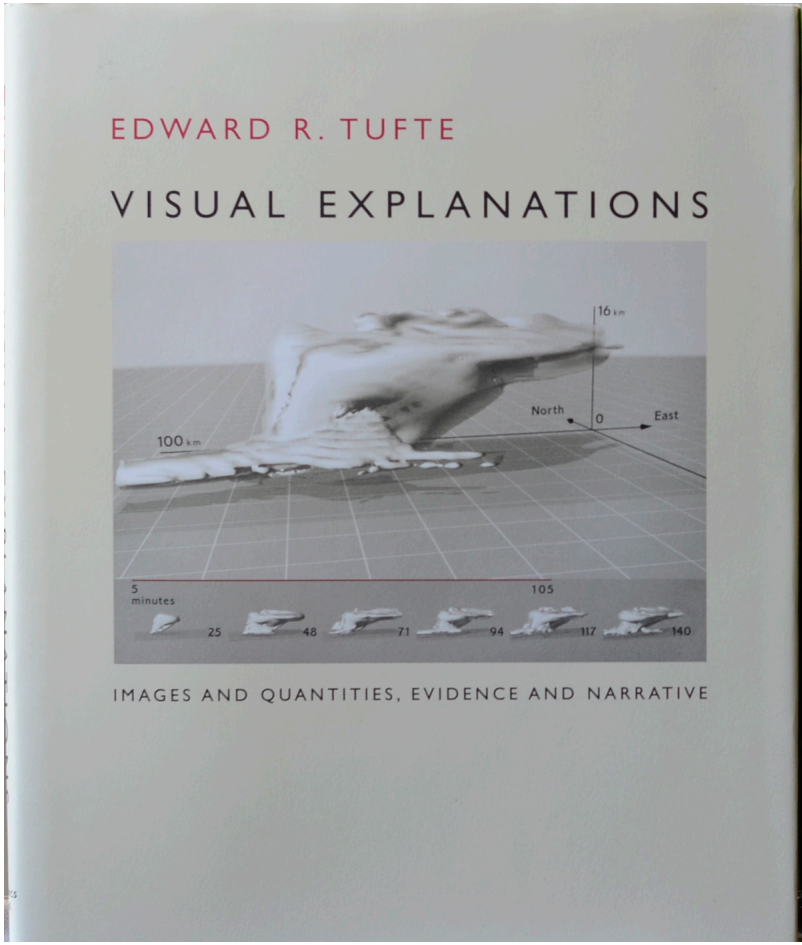
Aesthetics II

- Concept of "mapping from data to aesthetic attributes"
- Based on Leland Wilkinson, **The Grammar of Graphics**, (Springer 2005) and implemented in Hadley Wickham's `ggplot2` in **R**.
- Exemplified in work of Edward Tufte (`http://www.edwardtufte.com/`), especially **The Visual Display of Quantitative Information**, (Cheshire 2001)

Edward Tufte's books



Edward Tufte's books



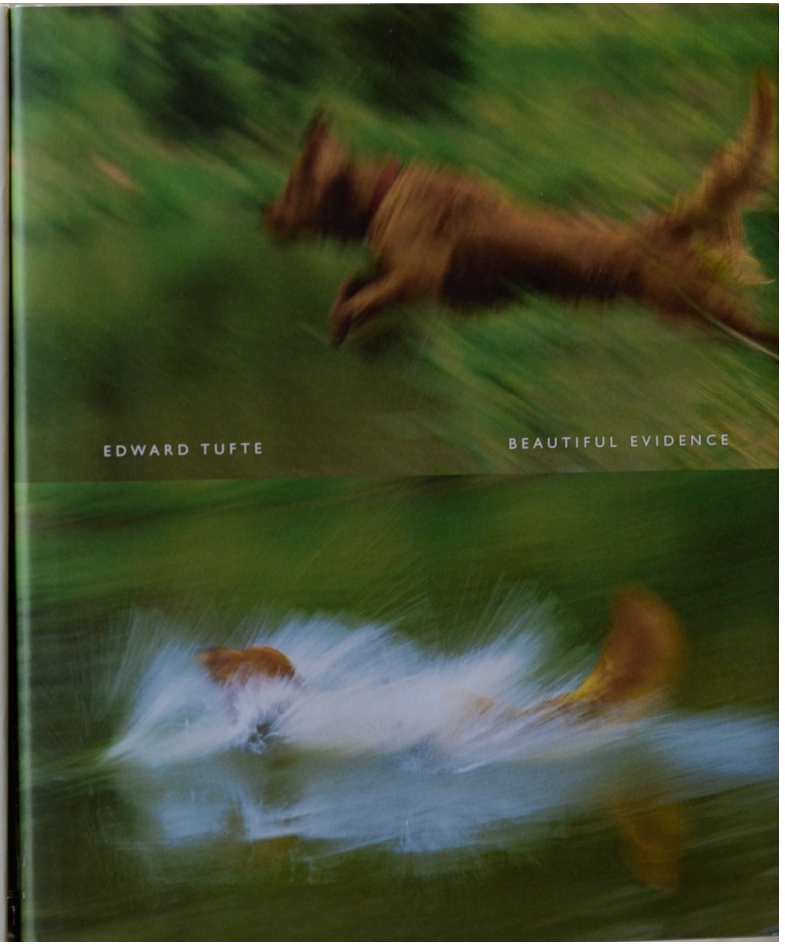


Table aesthetics I

- Tufte's "principles of graphical excellence" apply equally to tables.
- Goal: the well-designed presentation of interesting data—a matter of *substance*, of *statistics*, and of *design*.
- Consists of: complex ideas communicated with clarity, precision and efficiency.
- Gives the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.
- Nearly always multivariate.

Table aesthetics II

- Tufte scorns 'chart junk': we should maximise data component, minimise decorative junk - hence minimalist approach to extraneous ink
- Simon Fear (author of LaTeX package, `booktabs`) advocates: 'never use vertical rules' and 'never use double rules'.
- Importance of the readership:
  - *Generalist*: graphs in chapters, tables in appendix
  - *Specialist*: graphs and key tables in chapter, detailed tables in appendix

Implications for tables

- Key principles:
  - present many numbers in a small space;
  - encourage the eye to compare different pieces of data;
  - make the process of decoding efficient for the reader.
- Contrast with stats package output:
  - separate individual tables;
  - unnecessary additional information (DKs or the NO when only YES really relevant)
- Contrast with "lazy tables":
  - missing bits of information which make the reader undertake tedious mental calculations (eg. no 100%)
  - missing notes at base of table

Key elements in a table I

Table A.23: Household financial stress—C10 ‡

	Household comparisons					
	Adult low paid		Other		All households	
	'000s	%	'000s	%	'000s	%
Family finances: optimists						
Poor or very poor	20	1.6	44	1.2	64	1.3
Just getting along	285	23.8	720	19.0	1,005	20.1
Reasonably comfortable	645	53.9	2,039	53.7	2,684	53.8
Prosperous or v comfort	246	20.6	991	26.1	1,237	24.8
Total	1,196	100.0	3,793	100.0	4,990	100.0
Family finances: pessimists						
Poor or very poor	46	3.8	104	2.8	150	3.0
Just getting along	401	33.5	1,054	27.8	1,454	29.1
Reasonably comfortable	645	53.9	2,097	55.3	2,742	55.0
Prosperous or v comfort	105	8.8	539	14.2	644	12.9
Total	1,196	100.0	3,793	100.0	4,990	100.0
Episodes of financial hardship						
Three or more	135	11.3	295	7.8	430	8.7
Two	115	9.7	282	7.5	397	8.0
One	160	13.4	509	13.5	668	13.5
None	781	65.6	2,691	71.3	3,472	69.9
Total	1,191	100.0	3,776	100.0	4,967	100.0
How easily raise $2000 in one week						
Could not raise it	244	20.4	481	12.7	725	14.6
Have to do something drastic	194	16.2	399	10.5	593	11.9
Raise it, but some sacrifices	321	26.8	949	25.1	1,270	25.5
Easily raise it	436	36.5	1,956	51.7	2,393	48.0
Total	1,196	100.0	3,785	100.0	4,981	100.0
Ownership of credit card						
No credit card	453	34.1	999	23.7	1,452	26.2
Owns credit card	876	65.9	3,210	76.3	4,086	73.8
Total	1,330	100.0	4,209	100.0	5,538	100.0
Sample size	1,200		3,849		5,049	

Notes: First two panels: self-perceptions of financial prosperity. Optimists and pessimists result from differing evaluations by first two members of household. Counts are lower in this table because of missing observations. Third panel: episodes of financial hardship. Since beginning of year have any of following happened (due to lack of money): not pay utility bills on time; not pay rent or mortgage on time; pawned or sold something; went without meals; unable to heat home; asked for financial help from family or friends; asked for help from welfare organisation. Fourth panel: worst situation reported by at least one person in household. Fifth panel: no credit card = no one in household had a credit or charge card or store account; credit card = at least one person had one. Weighted by cross-sectional household population weights. Definition of low pay: earning at or below $15.94 per hour.
Population: Adult = Households with at least one adult low paid employee; Other = Households with at least one employed person (excluding Adult etc); All = Households with at least one employed person. Data from Wave 5 (2005).
Source: HILDA Release 5. †Responding person survey form; ‡Responding person self-completion survey form; §Household survey form.

- Shows population estimates and percentages
- Population estimates give readers a feel for the numbers involved

Key elements in a table II

Table A.23: Household financial stress—C10 ‡

	Household comparisons					
	Adult low paid		Other		All households	
	'000s	%	'000s	%	'000s	%
Family finances: optimists						
Poor or very poor	20	1.6	44	1.2	64	1.3
Just getting along	285	23.8	720	19.0	1,005	20.1
Reasonably comfortable	645	53.9	2,039	53.7	2,684	53.8
Prosperous or v comfort	246	20.6	991	26.1	1,237	24.8
Total	1,196	100.0	3,793	100.0	4,990	100.0
Family finances: pessimists						
Poor or very poor	46	3.8	104	2.8	150	3.0
Just getting along	401	33.5	1,054	27.8	1,454	29.1
Reasonably comfortable	645	53.9	2,097	55.3	2,742	55.0
Prosperous or v comfort	105	8.8	539	14.2	644	12.9
Total	1,196	100.0	3,793	100.0	4,990	100.0
Episodes of financial hardship						
Three or more	135	11.3	295	7.8	430	8.7
Two	115	9.7	282	7.5	397	8.0
One	160	13.4	509	13.5	668	13.5
None	781	65.6	2,691	71.3	3,472	69.9
Total	1,191	100.0	3,776	100.0	4,967	100.0
How easily raise $2000 in one week						
Could not raise it	244	20.4	481	12.7	725	14.6
Have to do something drastic	194	16.2	399	10.5	593	11.9
Raise it, but some sacrifices	321	26.8	949	25.1	1,270	25.5
Easily raise it	436	36.5	1,956	51.7	2,393	48.0
Total	1,196	100.0	3,785	100.0	4,981	100.0
Ownership of credit card						
No credit card	453	34.1	999	23.7	1,452	26.2
Owns credit card	876	65.9	3,210	76.3	4,086	73.8
Total	1,330	100.0	4,209	100.0	5,538	100.0
Sample size	1,200		3,849		5,049	

Notes: First two panels: self-perceptions of financial prosperity. Optimists and pessimists result from differing evaluations by first two members of household. Counts are lower in this table because of missing observations. Third panel: episodes of financial hardship. Since beginning of year have any of following happened (due to lack of money): not pay utility bills on time; not pay rent or mortgage on time; pawned or sold something; went without meals; unable to heat home; asked for financial help from family or friends; asked for help from welfare organisation. Fourth panel: worst situation reported by at least one person in household. Fifth panel: no credit card = no one in household had a credit or charge card or store account; credit card = at least one person had one. Weighted by cross-sectional household population weights. Definition of low pay: earning at or below $15.94 per hour.
Population: Adult = Households with at least one adult low paid employee; Other = Households with at least one employed person (excluding Adult etc); All = Households with at least one employed person. Data from Wave 5 (2005).
Source: HILDA Release 5. †Responding person survey form; ‡Responding person self-completion survey form; §Household survey form.

- Always show 100s, so instant awareness that dealing with column percentages

Key elements in a table III

Table A.23: Household financial stress—C10 ‡

	Household comparisons					
	Adult low paid		Other		All households	
	'000s	%	'000s	%	'000s	%
Family finances: optimists						
Poor or very poor	20	1.6	44	1.2	64	1.3
Just getting along	285	23.8	720	19.0	1,005	20.1
Reasonably comfortable	645	53.9	2,039	53.7	2,684	53.8
Prosperous or v comfort	246	20.6	991	26.1	1,237	24.8
Total	1,196	100.0	3,793	100.0	4,990	100.0
Family finances: pessimists						
Poor or very poor	46	3.8	104	2.8	150	3.0
Just getting along	401	33.5	1,054	27.8	1,454	29.1
Reasonably comfortable	645	53.9	2,097	55.3	2,742	55.0
Prosperous or v comfort	105	8.8	539	14.2	644	12.9
Total	1,196	100.0	3,793	100.0	4,990	100.0
Episodes of financial hardship						
Three or more	135	11.3	295	7.8	430	8.7
Two	115	9.7	282	7.5	397	8.0
One	160	13.4	509	13.5	668	13.5
None	781	65.6	2,691	71.3	3,472	69.9
Total	1,191	100.0	3,776	100.0	4,967	100.0
How easily raise $2000 in one week						
Could not raise it	244	20.4	481	12.7	725	14.6
Have to do something drastic	194	16.2	399	10.5	593	11.9
Raise it, but some sacrifices	321	26.8	949	25.1	1,270	25.5
Easily raise it	436	36.5	1,956	51.7	2,393	48.0
Total	1,196	100.0	3,785	100.0	4,981	100.0
Ownership of credit card						
No credit card	453	34.1	999	23.7	1,452	26.2
Owns credit card	876	65.9	3,210	76.3	4,086	73.8
Total	1,330	100.0	4,209	100.0	5,538	100.0
Sample size	1,200		3,849		5,049	

Notes: First two panels: self-perceptions of financial prosperity. Optimists / pessimists result from differing evaluations by first two members of household. Counts are lower in this table because of missing observations. Third panel: episodes of financial hardship. Since beginning of year have any of following happened (due to lack of money): not pay utility bills on time; not pay rent or mortgage on time; pawned or sold something; went without meals; unable to heat home; asked for financial help from family or friends; asked for help from welfare organisation. Fourth panel: worst situation reported by at least one person in household. Fifth panel: no credit card = no one in household had a credit or charge card or store account; credit card = at least one person had one. Weighted by cross-sectional household population weights. Definition of low pay: earning at or below $15.94 per hour.
Population: Adult = Households with at least one adult low paid employee; Other = Households with at least one employed person (excluding Adult etc); All = Households with at least one employed person. Data from Wave 5 (2005).
Source: HILDA Release 5. †Responding person survey form; ‡Responding person self-completion survey form; §Household survey form.

- Show sample sizes, so that cell counts can be calculated and reader can sense the precision of the estimates

Key elements in a table IV

Table A.23: Household financial stress—C10 ‡

	Household comparisons					
	Adult low paid		Other		All households	
	'000s	%	'000s	%	'000s	%
Family finances: optimists						
Poor or very poor	20	1.6	44	1.2	64	1.3
Just getting along	285	23.8	720	19.0	1,005	20.1
Reasonably comfortable	645	53.9	2,039	53.7	2,684	53.8
Prosperous or v comfort	246	20.6	991	26.1	1,237	24.8
Total	1,196	100.0	3,793	100.0	4,990	100.0
Family finances: pessimists						
Poor or very poor	46	3.8	104	2.8	150	3.0
Just getting along	401	33.5	1,054	27.8	1,454	29.1
Reasonably comfortable	645	53.9	2,097	55.3	2,742	55.0
Prosperous or v comfort	105	8.8	539	14.2	644	12.9
Total	1,196	100.0	3,793	100.0	4,990	100.0
Episodes of financial hardship						
Three or more	135	11.3	295	7.8	430	8.7
Two	115	9.7	282	7.5	397	8.0
One	160	13.4	509	13.5	668	13.5
None	781	65.6	2,691	71.3	3,472	69.9
Total	1,191	100.0	3,776	100.0	4,967	100.0
How easily raise $2000 in one week						
Could not raise it	244	20.4	481	12.7	725	14.6
Have to do something drastic	194	16.2	399	10.5	593	11.9
Raise it, but some sacrifices	321	26.8	949	25.1	1,270	25.5
Easily raise it	436	36.5	1,956	51.7	2,393	48.0
Total	1,196	100.0	3,785	100.0	4,981	100.0
Ownership of credit card						
No credit card	453	34.1	999	23.7	1,452	26.2
Owns credit card	876	65.9	3,210	76.3	4,086	73.8
Total	1,330	100.0	4,209	100.0	5,538	100.0
Sample size	1,200		3,849		5,049	

Notes: First two panels: self-perceptions of financial prosperity. Optimists and pessimists result from diverging evaluations by first two members of household. Counts are lower in this table because of missing observations. Third panel: episodes of financial hardship. Since beginning of year have any of following happened (due to lack of money): not pay utility bills on time; not pay rent or mortgage on time; pawned or sold something; went without meals; unable to heat home; asked for financial help from family or friends; asked for help from welfare organisation. Fourth panel: worst situation reported by at least one person in household. Fifth panel: no credit card = no one in household had a credit or charge card or store account; credit card = at least one person had one. Weighted by cross-sectional household population weights. Definition of low pay: earning at or below $15.94 per hour.
Population: Adult = Households with at least one adult low paid employee; Other = Households with at least one employed person (excluding Adult etc); All = Households with at least one employed person. Data from Wave 5 (2005).
Source: HILDA Release 5. †Responding person survey form; ‡Responding person self-completion survey form; §Household survey form.

- Notes may consist of: **notes**, **population** and **source**
- Notes may explain decision rules, definitions and weighting
- Source may explain where data items came from

Reproducible research I

- Principles of **efficiency** and **accuracy**
- Provides an audit trail
- Example of revisiting results a year later
- Re-running analysis with different data or methods
- Dynamic report writing with data still coming in
- Slogan: "copy and paste" is your enemy: instead aim for "files talking to files"
- Encourages single source publishing

Reproducible research II

- Example in Stata of nested do file structure:
  - master.do → final tables and/or final report
  - master.do made up of:
    - raw.dta → clean.do → clean.dta
    - clean.dta → recode.do → final.dat
    - final.dta → tables.do → actual table files
  - Tables then inserted (with link) in Word document or referenced in LaTeX file
  - Contrast with large single data file which becomes "precious" (eg. in SPSS) and unreproducible

Single source publishing

- Multiple audiences:
  - PDF report for printing
  - Excel file for data provision
  - HTML report for the web and for conversion to ebook formats
- DRY ("don't repeat yourself") applicable to report generation - change something only in one location
- Notion of "chained files" - *text* files invoking other *text* files in *sequential time* (Unix principle) versus *binary* behemoths (eg. word processors) which try to achieve everything in *real time*.

Example master file

```
* master file 21 for project XYZ 16jun2016
* purpose is to ...
cd [your working directory]
do clean
do recode
do tables
shell pdflatex xyz.tex
shell open xyz.pdf
```

Example clean file

```
* clean file 21 for project XYZ 16jun2016
* data provided by ...
cd [your working directory]
use raw.dta, clear
```

Various coding to eliminate duplicates, check integrity etc. Use of regular expressions.

May use edit mode, but capture the commands and include in the file eg.

```
replace abcd = 10 in 13
```

echoed by **Stata** becomes:

```
replace abcd = 10 if id == 1416
```

Why? Observation numbers can change!

Example of LaTeX report file

LaTeX example for composing report. Different to MS Word (with linked files) or Sweave in **R**.

```
\documentclass[a4paper, 11pt, oneside]{memoir}

\begin{document}
\section{Introduction}

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod
tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam

As Table \ref{t_part_timers} shows, Lorem ipsum dolor sit amet ...

\input ./tables/t_part_timers

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod
tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim
veniam, ...

\end{document}
```

Example of LaTeX table file

```
\begin{table}[H]
  \begin{center}
  \footnotesize
    \begin{minipage}{13cm}
    {\caption{Full-time and part-time employees, Australia 2013}{\label{t_part_timers}}}
      \vspace{1ex}
      \begin{tabularx}{13cm}{ l  Y  Y  Y  Y }
        \toprule

\emph{Industry} & \emph{Full-time} & \emph{Part-time} & \emph{Total} & \emph{Part-time as \%} \\
        \midrule
\lt Agric, forestry, fishing & 79,397 & 21,356 & 100,753 & 21.2 \\
\dk Mining & 234,305 & 13,591 & 247,896 & 5.5 \\
\lt Manufacturing & 653,036 & 127,606 & 780,642 & 16.3 \\
\dk Elect, gas, water, waste & 90,600 & 9,084 & 99,683 & 9.1 \\
...
\lt Arts and recreation services & 93,561 & 78,111 & 171,673 & 45.5 \\
\dk Other services & 205,181 & 93,238 & 298,419 & 31.2 \\
\lt Total & 6,485,837 & 3,193,333 & 9,679,169 & 33.0 \\

        \bottomrule
        \addlinespace
      \end{tabularx}
     {\scriptsize Source: Unpublished HILDA data. Population: Employees (excluding owner managers or incorporate
        \vspace*{-3ex}
    \end{minipage}
  \end{center}
\end{table}
```

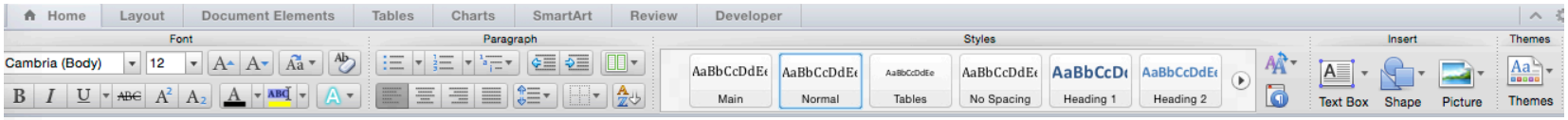
Example of PDF table file

TABLE 2.3: FULL-TIME AND PART-TIME EMPLOYEES, AUSTRALIA 2013

Industry	*Full-time*	*Part-time*	*Total*	*Part-time as %*
Agric, forestry, fishing	79,397	21,356	100,753	21.2
Mining	234,305	13,591	247,896	5.5
Manufacturing	653,036	127,606	780,642	16.3
Elect, gas, water, waste	90,600	9,084	99,683	9.1
Construction	522,625	61,391	584,016	10.5
Wholesale trade	301,722	45,630	347,352	13.1
RETAIL	317,356	585,151	902,508	64.8
OTHER DIVISION G	84,517	37,740	122,257	30.9
Accomm and food services	247,600	521,527	769,127	67.8
Trans, postal, warehousing	387,364	97,473	484,837	20.1
Information media, telecomm	141,136	44,074	185,209	23.8
Finance and insurance	341,447	55,244	396,691	13.9
Rental, hiring, real estate	92,057	27,806	119,863	23.2
Profess, scientific tech	573,960	148,740	722,700	20.6
Admin and support services	150,605	87,364	237,969	36.7
Public admin and safety	576,233	85,731	661,964	13.0
Education and training	593,107	403,361	996,468	40.5
Health and social assistance	800,028	649,115	1,449,143	44.8
Arts and recreation services	93,561	78,111	171,673	45.5
Other services	205,181	93,238	298,419	31.2
Total	6,485,837	3,193,333	9,679,169	33.0

Source: Unpublished HILDA data. Population: Employees (excluding owner managers or incorporated enterprises) in main job.

MS Word example I

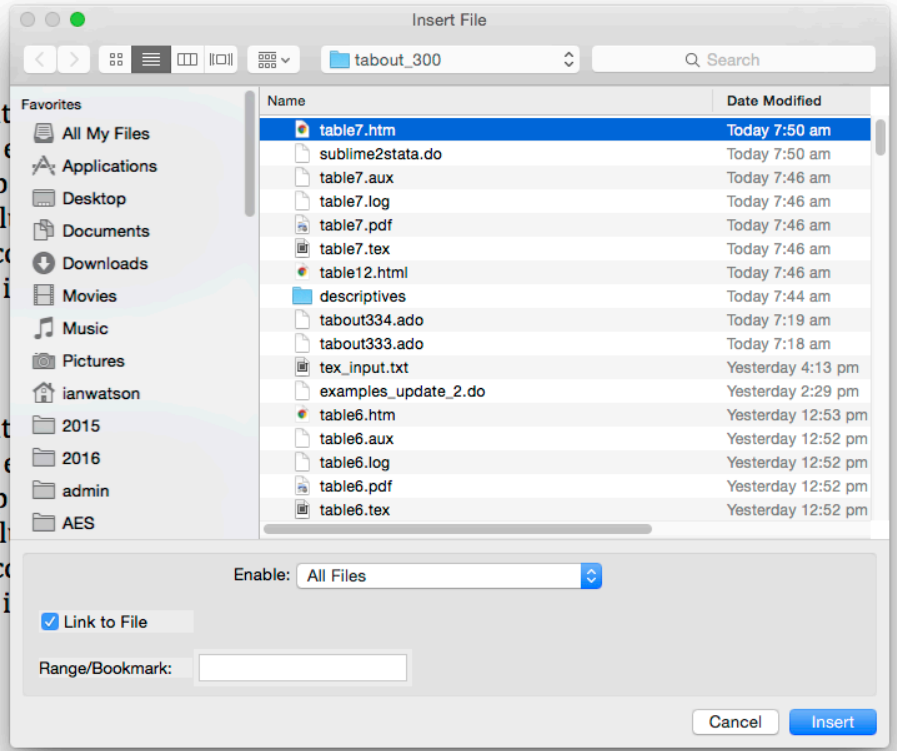


My report

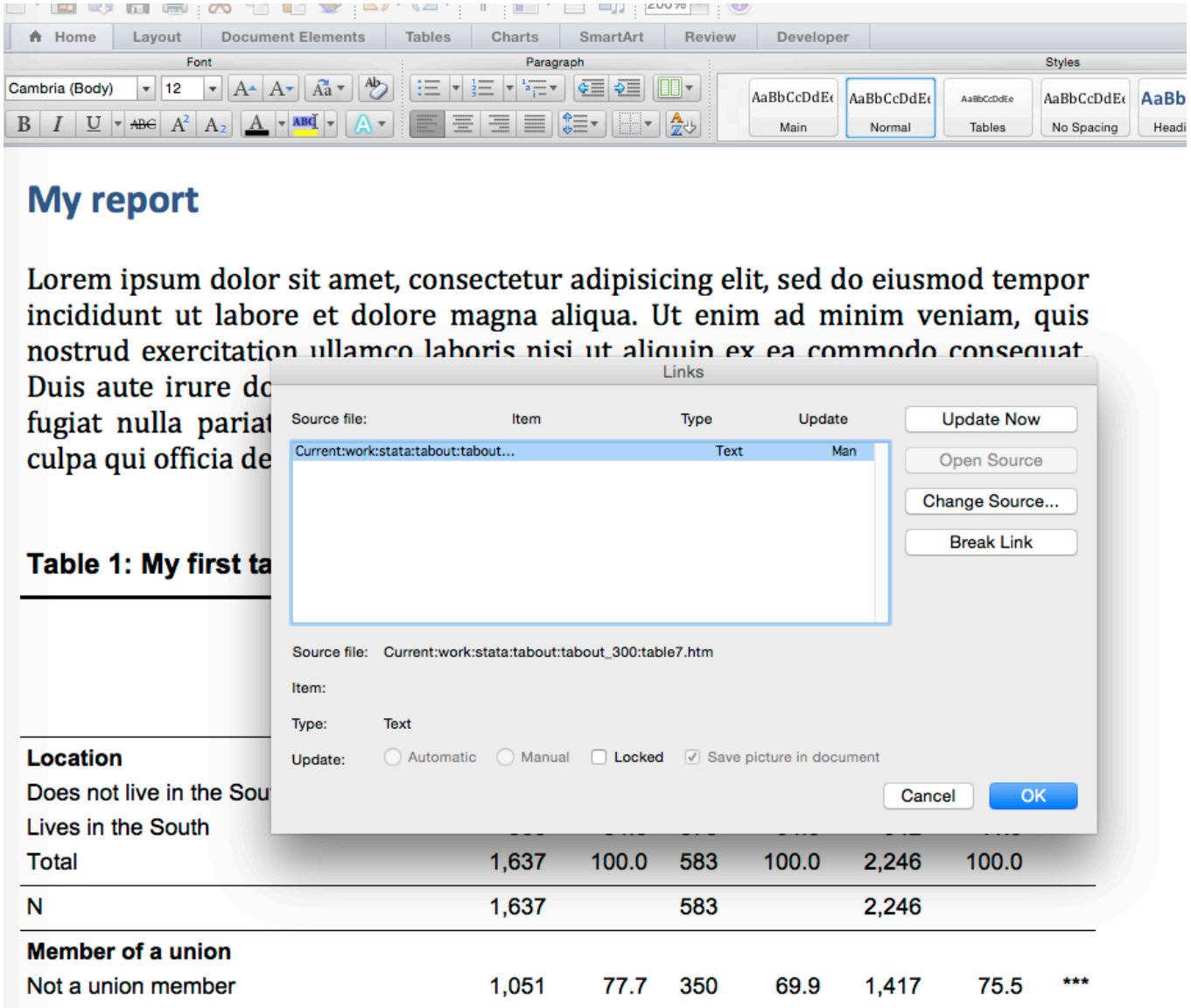
Lorem ipsum dolor sit amet, consectetur adipisicing elit
tempor incididunt ut labore et dolore magna aliqua. Ut
quis nostrud exercitation ullamco laboris nisi ut aliquip
consequat. Duis aute irure dolor in reprehenderit in vol
cillum dolore eu fugiat nulla pariatur. Excepteur sint oc
proident, sunt in culpa qui officia deserunt mollit anim i

Lorem ipsum dolor sit amet, consectetur adipisicing elit
tempor incididunt ut labore et dolore magna aliqua. Ut
quis nostrud exercitation ullamco laboris nisi ut aliquip
consequat. Duis aute irure dolor in reprehenderit in vol
cillum dolore eu fugiat nulla pariatur. Excepteur sint oc
proident, sunt in culpa qui officia deserunt mollit anim i

—



MS Word example II



MS Word example III

My report

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat nonproident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Table 1: My first table

	Race						Sig
	White		Black		Total		
	No.	%	No.	%	No.	%	
Location							
Does not live in the South	5,785	64.6	1,187	36.9	7,107	57.6	***
Lives in the South	3,169	35.4	2,034	63.1	5,227	42.4	
Total	8,954	100.0	3,221	100.0	12,334	100.0	
N	1,637		583		2,246		
Member of a union							
Not a union member	5,723	78.1	1,879	68.5	7,717	75.5	***
Union member	1,601	21.9	863	31.5	2,504	24.5	
Total	7,324	100.0	2,742	100.0	10,221	100.0	
N	1,353		501		1,878		
Sex							
Male	2,351	26.5	808	25.3	3,180	26.0	***
Female	6,513	73.5	2,389	74.7	9,038	74.0	
Total	8,864	100.0	3,197	100.0	12,218	100.0	
N	1,547		559		2,130		

This is a small footnote of various size and enterprises and even has a line break in the middle of it just for testing the text box which has been put in there.

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu

How `tabout` fits in

- Reproducible research:
  - `tabout` → final version of table - no further editing needed
  - lends itself to "chained files" concept
  - **new feature:** expanded file writing capacities
  - **new feature:** compiling and previewing tables
- Single source publishing:
  - `tabout` → various outputs eg. HTML, PDF, MS Word, MS Excel
  - **new feature:** native docx and xlsx file formats
  - **new feature:** configuration files for minimal effort for multiple outputs

`tabout` design principles I

- Concept of panels: "horizontal" variable and many "vertical variables" - Tufte's principles
- Integration of diverse **Stata** commands under one hood: `tabulate`, `summarize`, various `svy` commands.
- Table should need no further editing: "camera ready" appearance.
- Building new table structures with judicious use of `replace` and `append` and various user-defined input (`h1 h2 h3` etc)
- Flexibility increased with **new features**: `topbody` and `botbody`

tabout design principles II

- Flexibility in layout: columns, rows, column blocks or rowblocks

Type of table	Allowable cell contents	Available layout
Basic	freq cell row col cum **any number of above, in any order** *for example: cells(freq col)*	col row cb rb
Basic with SE or CI (turn on *svy* option)	freq cell row col se ci lb ub **only one of:** freq cell row col *(must come first in the cell)* **and any number of:** se ci lb ub *for example: cells(col se lb ub)*	col row cb rb
Summary -as a oneway table (turn on *sum* option; also may need to turn on *oneway* option)	**any number of:** N mean var sd skewness kurtosis sum uwsum min max count median iqr r9010 r9050 r7525 r1050 p1 p5 p10 p25 p50 p75 p90 p95 p99 **with each followed by variable name** *for example: cells(min wage mean age)*	no options (fixed)
Summary -as a twoway table (turn on *sum* option)	**only one of:** N mean var sd skewness kurtosis sum uwsum min max count median iqr r9010 r9050 r7525 r1050 p1 p5 p10 p25 p50 p75 p90 p95 p99 **followed by one variable name** *for example: cells(sum income)*	no options (fixed)
Summary with SE or CI (turn on *sum* option and *svy* option)	mean **followed by one variable name** **and any number of:** se ci lb ub *for example: cells(mean weight se ci)*	col row cb rb

`tabout` design principles III

- Trade-off between complexity and flexibility
  - large number of options, but no sub-options (**Stata** graphics counter-example)
  - inspiration of `estout` but also complexity of sub-options
  - thus preference for switches eg. the N family of switches: `npos nlab nwt nnoc noffset‘`
  - only use switch if needed, otherwise default setting used
- **new feature:** configuration files:
  - remove clutter and reliance on memory
  - share with colleagues or learners

`tabout` design principles IV

- Combines **Stata** and `mata` (**Stata** Version 9+)
- Programming advantages:
  - matrix processing & file writing more efficient
  - pointers for run-time user choices
  - structs for passing complex parameters
- User advantages:
  - faster experience
  - flexibility: column dropping & adding
  - docx output (**Stata** Version 13+)
- Programming disadvantages:
  - frustrating inconsistencies in using two languages simultaneously
  - frustrating passing parameters back and forth between **Stata** and `mata`

tabout: new features I

- Long overdue user requests:
  - dropping columns eg Totals
  - plugging gaps eg. missing categories
- Enhanced output for non-LaTeX users:
  - write to multiple sheets in Excel files using native xls/xlsx formats and place multiple tables on sheets
  - write to Word files in native docx format
  - improved HTML output including CSS (cascading style sheets) support
  - specify font sizes and font families for HTML, Word and Excel outputs

`tabout`: new features II

- Configuration files
- Provision of table `title` and `footnote` options—no longer necessary to use `topf` and `botf` for simple material
- Makes it easier for novice LaTeXusers
- Enhanced handling of table statistics (eg. chi2):
  - test statistics in columns or rows
  - choice of statistic and/or p-value
  - choice of p-values or stars
  - user-defined labels

`tabulate` in practice

```
tab south race, col row
```

Key
frequency *row percentage* *column percentage*

Location	Race White	Black	Other	Total
Does not live in the	1,071 82.13 65.42	210 16.10 36.02	23 1.76 88.46	1,304 100.00 58.06
Lives in the South	566 60.08 34.58	373 39.60 63.98	3 0.32 11.54	942 100.00 41.94
Total	1,637 72.89 100.00	583 25.96 100.00	26 1.16 100.00	2,246 100.00 100.00

tabout in practice I

```
tabout south union using table1.htm, c(freq col row) ///
 f(0c 1) style(htm) font(bold)
```

Location	Member of a union								
	Not a union member			Union member			Total		
	No.	%	%	No.	%	%	No.	%	%
Does not live in the South	754	53.2	69.9	325	70.5	30.1	1,079	57.5	100.0
Lives in the South	663	46.8	83.0	136	29.5	17.0	799	42.5	100.0
Total	1,417	100.0	75.5	461	100.0	24.5	1,878	100.0	100.0

tabout in practice II

```
tabout south union using table1.tex, c(freq col row) ///
f(0c 1) style(tex) font(bold) twidth(14) body ///
title(Table 1: My first table) ///
fn(Some useful additional information)
```

Table 1: My first table

Location	Member of a union								
	Not a union member			Union member			Total		
	No.	%	%	No.	%	%	No.	%	%
Location									
Does not live in the South	754	53.2	69.9	325	70.5	30.1	1,079	57.5	100.0
Lives in the South	663	46.8	83.0	136	29.5	17.0	799	42.5	100.0
Total	1,417	100.0	75.5	461	100.0	24.5	1,878	100.0	100.0

Some useful additional information

Stata with survey data

Stata output for two separate tables:

```
svyset psuid [pweight=finalwgt], strata(stratid)
svy: tabulate diabetes race, row ci format(%7.3f)
svy: tabulate diabetes sex, row ci format(%7.3f)
```

Diabetes	Race			
	White	Black	Other	Total
No	0.881	0.093	0.026	1.000
	[0.844,0.911]	[0.070,0.122]	[0.011,0.059]	
Yes	0.820	0.165	0.015	1.000
	[0.748,0.876]	[0.120,0.221]	[0.003,0.067]	
Total	0.879	0.096	0.025	1.000
	[0.841,0.909]	[0.072,0.125]	[0.011,0.058]	

Key: row proportion
[95% confidence interval for row proportion]

Pearson:
Uncorrected chi2(2) = 21.3483
Design-based F(1.52, 47.26) = 15.0056 P = 0.0000

Diabetes	Sex		
	Male	Female	Total
No	0.482	0.518	1.000
	[0.470,0.494]	[0.506,0.530]	
Yes	0.408	0.592	1.000
	[0.351,0.468]	[0.532,0.649]	
Total	0.479	0.521	1.000
	[0.468,0.491]	[0.509,0.532]	

Key: row proportion
[95% confidence interval for row proportion]

Pearson:
Uncorrected chi2(1) = 7.4897
Design-based F(1, 31) = 6.2012 P = 0.0183

tabout with survey data

tabout combines output into panels in a single table, removes unwanted column and includes sample size. Also sets font, adds title and footnote.

```
tabout race sex diabetes using table2.htm, c(row ci) svy f(3) ///
style(htm) stats(chi2) body font(bold) npos(col) cisep(-) ///
family(Arial) dropc(6) title(Table 2: My second table) ///
fn(Some more useful information, perhaps about the sample design)
```

Table 2: My second table

	Diabetes				
	No		Yes		N
	Prop.	CI	Prop.	CI	
Race					
White	0.968	[0.964-0.972]	0.032	[0.028-0.036]	9,063
Black	0.941	[0.927-0.952]	0.059	[0.048-0.073]	1,086
Other	0.980	[0.957-0.991]	0.020	[0.009-0.043]	200
Total	0.966	[0.962-0.969]	0.034	[0.031-0.038]	10,349
Pearson: Uncorrected chi2(2)= 21.3483					
Design-based F(1.52, 47.26)= 15.0056					
P-value= 0.000					
Sex					
Male	0.971	[0.965-0.976]	0.029	[0.024-0.035]	4,915
Female	0.961	[0.955-0.966]	0.039	[0.034-0.045]	5,434
Total	0.966	[0.962-0.969]	0.034	[0.031-0.038]	10,349
Pearson: Uncorrected chi2(1)= 7.4897					
Design-based F(1.00, 31.00)= 6.2012					
P-value= 0.018					

Some more useful information, perhaps about the sample design

tabout with configuration file I

tabout can remove the clutter and "memory load" for detailed options with new configuration option cfg.

```
tabout race sex diabetes using table2.htm, cfg(svytabs.txt) ///
title(Table 2: My second table) fn(Some more useful information, ///
perhaps about the sample design) ///
```

Configuration file (svytabs.txt) holds generic information:

```
c(row ci) svy f(3) style(htm) stats(chi2) body font(bold) npos(col)
dropc(6) family(Arial) cisep(-)
```

and each table's syntax just addes the unique elements, eg. variable names and table title. Another cfg file (eg. appendix.txt) could hold options to produce more detailed information:

```
tabout race sex diabetes using appendix2.htm, ///
cfg(svyapps.txt) ///
title(Table 2A: Detailed breakdown of ...) ///
fn(Other detailed information, required in an appendix)
```

tabout with configuration file II

Also switch between different types of outputs:

```
tabout race sex diabetes using table2.tex, cfg(texsvy.txt) ///
title(Table 2: My second table) fn(Some more useful information, ///
perhaps about the sample design) ///
```

Configuration file (texsvy.txt) might hold:

```
c(row ci) svy f(3) style(tex) stats(chi2) body font(bold)
dropc(6) cisep(-) twidth(12) fsize(11) stpos(col) ppos(only)
plab(Sig) stars
```

Table 2: My second table

	Diabetes				
	No		Yes		Sig
	Prop.	CI	Prop.	CI	
Race					
White	0.968	[0.964-0.972]	0.032	[0.028-0.036]	***
Black	0.941	[0.927-0.952]	0.059	[0.048-0.073]	
Other	0.980	[0.957-0.991]	0.020	[0.009-0.043]	
Total	0.966	[0.962-0.969]	0.034	[0.031-0.038]	
Sex					
Male	0.971	[0.965-0.976]	0.029	[0.024-0.035]	*
Female	0.961	[0.955-0.966]	0.039	[0.034-0.045]	
Total	0.966	[0.962-0.969]	0.034	[0.031-0.038]	

Some more useful information, perhaps about the sample design

Extending `tabout`: three way tables I

Industry	Member of a union		
	Not in union	In union	Total
Race: white			
Ag/Forestry/Fisheries	8	1	9
Mining	2	0	2
Construction	14	2	16
Manufacturing	155	40	195
Transport/Comm/Utility	31	28	59
Wholesale/Retail Trade	188	18	206
Finance/Ins/Real Estate	127	6	133
Business/Repair Svc	41	3	44
Personal Services	33	2	35
Entertainment/Rec Svc	9	2	11
Professional Services	372	160	532
Public Administration	65	37	102
Total	1,045	299	1,344
Race: black			
Ag/Forestry/Fisheries	2	1	3
Construction	2	1	3
Manufacturing	76	43	119
Transport/Comm/Utility	8	19	27
Wholesale/Retail Trade	51	3	54
Finance/Ins/Real Estate	16	3	19
Business/Repair Svc	9	4	13
Personal Services	25	3	28
Entertainment/Rec Svc	3	0	3
Professional Services	119	58	177
Public Administration	35	16	51
Total	346	151	497
Race: other			
Construction	1	0	1
Manufacturing	3	1	4
Transport/Comm/Utility	0	1	1
Finance/Ins/Real Estate	1	0	1
Business/Repair Svc	1	1	2
Professional Services	9	2	11
Public Administration	1	3	4
Total	16	8	24

Source: nlsw88.dta

- Creative use of replace/append and other `tabout` options
- Exploiting some **Stata** programming tricks inside loops

Extending `tabout`: three way tables I

- Some simple programming:
  - learn how to use `macros`; and
  - become familiar with **Stata's** `levelsof` command:

```
sysuse nlsw88, clear

* normal bys approach
bys race: tabulate industry union

* pseudo bys approach
levelsof race, local(levels)
foreach l of local levels {
    tabulate industry union if race == `l'
}
```

Extending tabout: three way tables I

Then, incorporate tabout features h1 h2 h3 and file options replace and append:

```
* setup macros for loops
levelsof race, local(levels)
local racelabels : value label race
local counter = 0
local filemethod = "replace"
local heading = ""

* begin looping through the values of the by category
foreach l of local levels {
    if `counter' > 0 {
        local filemethod = "append"
        local heading = "h1(nil) h2(nil)"
    }
    local vlabel : label `racelabels' `l'
   tabout industry union if race == `l' using "table.txt", `filemethod' ///
        `heading' h3("Race: `vlabel'") f(0c)
    local counter = `counter' + 1
}
```

Future of `tabout`

- Version 3 currently being developed:
  - Most new features working
  - docx output under development
  - video tutorials also under development
  - beta version ready in next month or so with feedback sought
  - aim to have final version ready at end of 2016
- User requests and feedback?