

# Producing Publication Ready Graphs in Stata ©

(or things I wished I had known and been able to do 15 years ago)

## Vernon Gayle

Professor of Sociology & Social Statistics  
University of Edinburgh

vernon.gayle@ed.ac.uk  
@profbigvern

2016

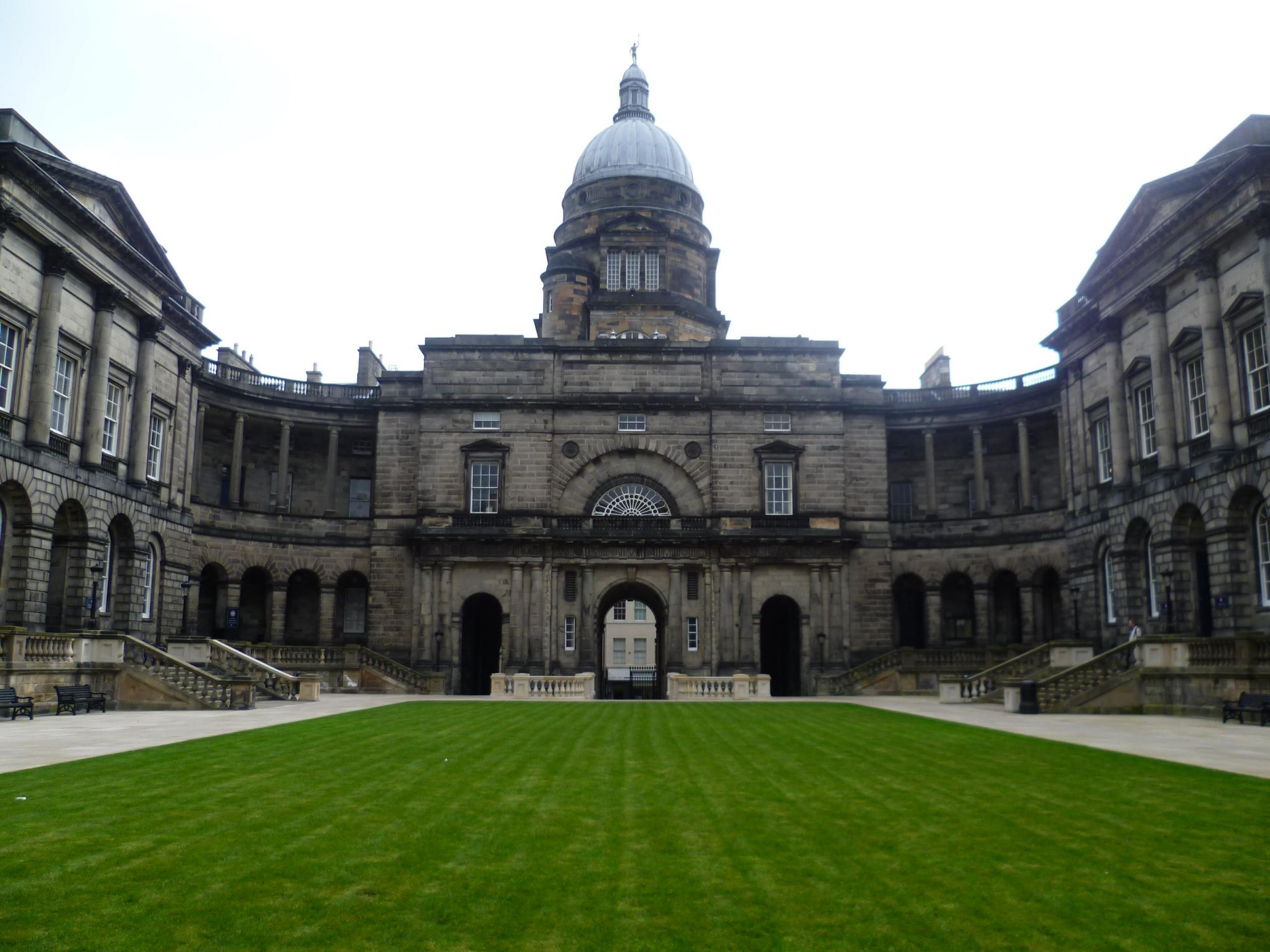
© Vernon Gayle



Administrative Data  
Research Centre  
Scotland



THE UNIVERSITY  
of EDINBURGH





Administrative Data Research Centre Scotland

# Better Knowledge Better Society

- ▶ Supporting researchers to use administrative data for economic and social research
- ▶ Enabling studies that benefit the public
- ▶ Helping inform and improve policy-making

W. adm.ac.uk  
E. adrc-sped.ac.uk





## AQMeN

Applied Quantitative Methods Network



Research Centre developing a dynamic pioneering set of inter-disciplinary methods to improve understanding of social issues in the UK.

conducting an international programme of research using secondary data in three primary areas:

- Crime and Victimisation
- Education and Social Stratification
- Migration, Integration and Inequality

Helping policy makers and researchers with robust independent evidence-based evidence.

Providing high quality statistical analysis and user support.

Facilitating knowledge exchange and collaboration across disciplines and sectors.

Supporting research in quantitative methods in the UK, social science and health.

[www.aqmen.ac.uk](http://www.aqmen.ac.uk)



Administrative Data Research Centre Scotland

An ESRC Data Investment

# Better Knowledge Better Society

- ▶ Supporting researchers to use administrative data for economic and social research
- ▶ Enabling studies that benefit the public
- ▶ Helping inform and improve policy-making



W. adm.ac.uk  
E. adrc-sped.ac.uk

In partnership with the Fair Institute of Scotland and the Scottish Government



## AQMeN

Applied Quantitative Methods Network



A Research Centre pioneering set of interdisciplinary methods to improve understanding of social issues in the UK.

- Crime and Victimisation
- Education and Social Stratification
- Migration, Integration and Inequality

Helping policy makers and researchers with robust independent evidence-based evidence.

Providing high quality statistical analysis and user support.

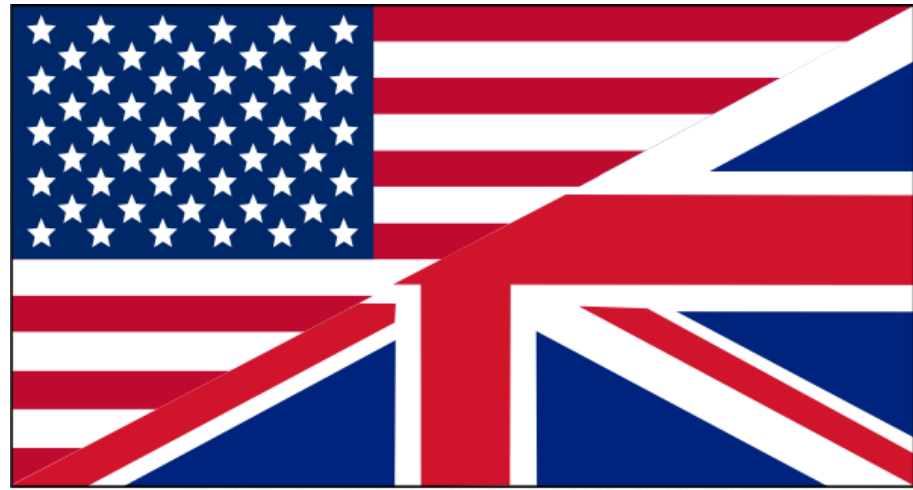
Facilitating knowledge exchange and collaboration across disciplines and sectors.

Supporting research in quantitative methods in the UK, social science and health.

[www.aqmen.ac.uk](http://www.aqmen.ac.uk)







Cookie

Biscuit

Flashlight

Torch

Fries

Chips

Chips

Crisps

Soccer

Football

Candy

Sweets

Restroom

Loo

Gasoline

Petrol

# Topics Today

Introduction (30 mins)

Basic Operations (30 mins)

#delimit; markers; labelling points; axes; legends; line plots etc.

Saving and Exporting (20 mins)

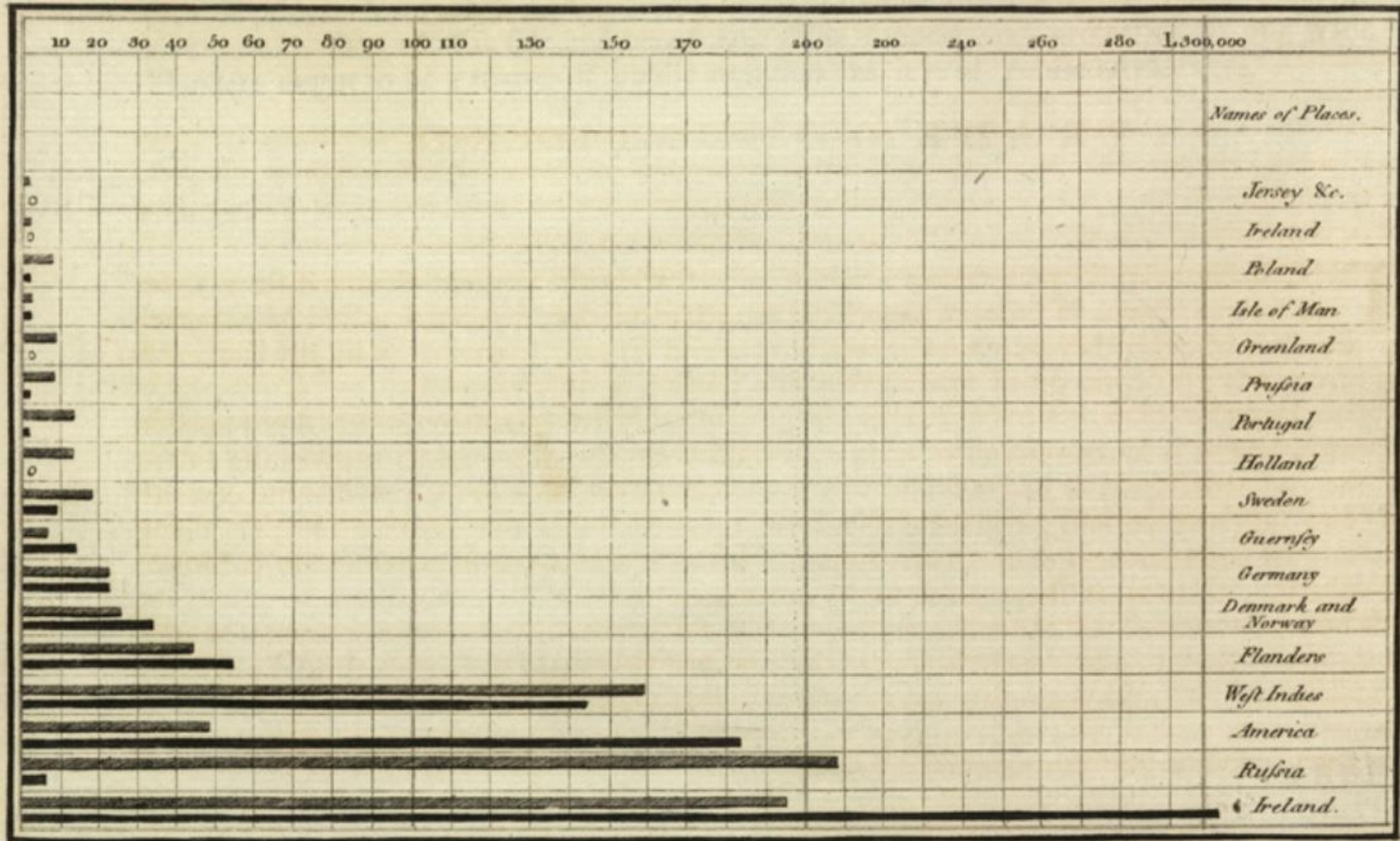
Graphing Results (20 mins)

*The .do file has additional topics*



# A Graph of Raw Data

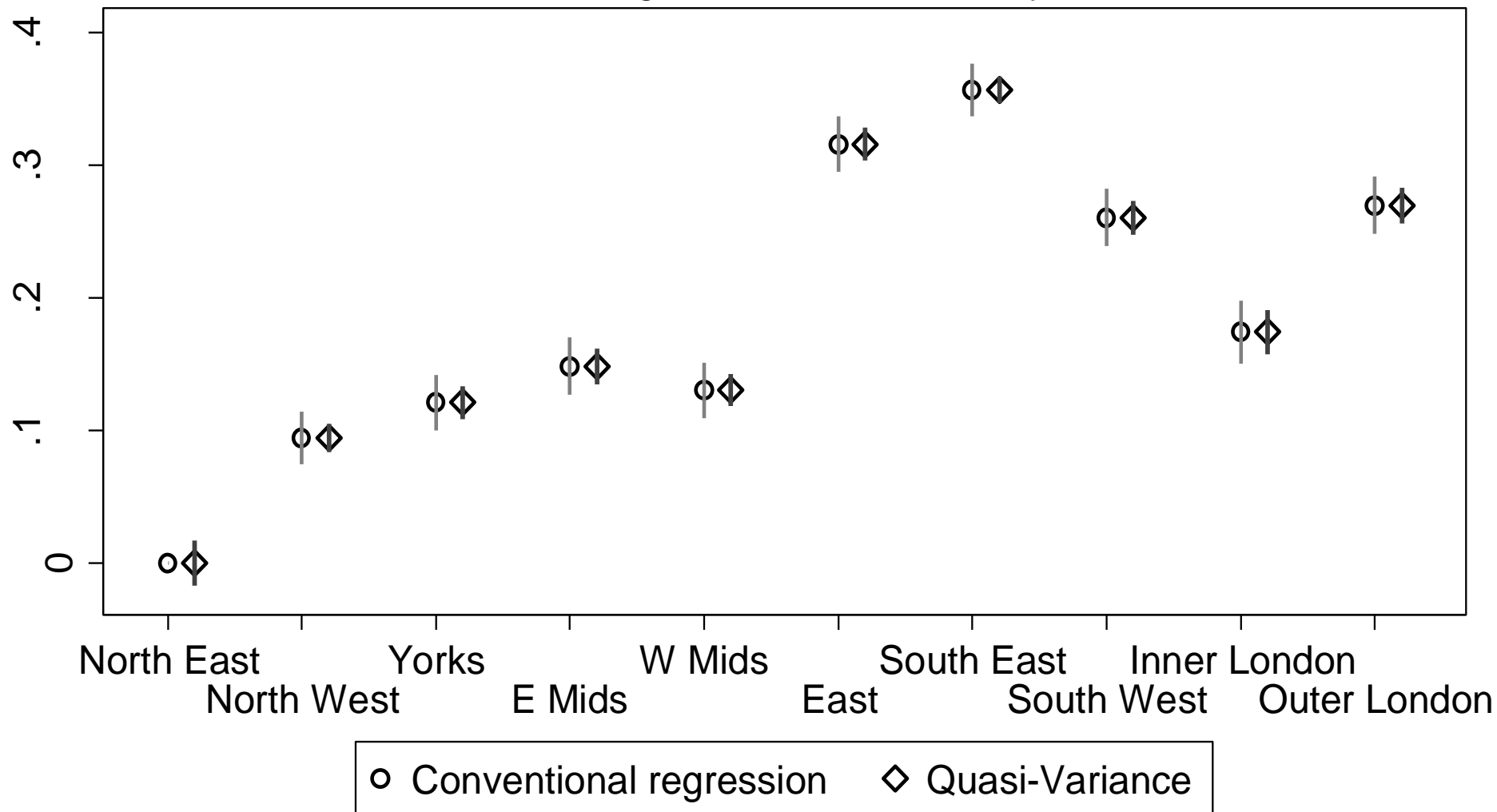
Exports and Imports of SCOTLAND to and from different parts for one Year from Christmas 1780 to Christmas 1781.



The Upright divisions are Ten Thousand Pounds each. The Black Lines are Exports the Ribbed lines Imports.

# Predictions of Good Health, by Government Office Region

Confidence intervals of regression coefficients, by estimation method



Source: UK Census 2001 SARS for England, n=1099294.

Model 1: Logistic regression predicting 'Good Health'. Other controls for education and gender

Gayle and Lambert (2007 p.1195)



```

#delimit ;
graph twoway
    (scatter beta region if source==1, msymbol(circle_hollow) mlcolor(gs0) msize(medium))
    (scatter beta region if source==2, msymbol(diamond_hollow) mlcolor(gs0) msize(medium))
    (rspike upper lower region if source==1, blwidth(medium) )
    (rspike upper lower region if source==2, blwidth(medthick) )
    ,
    ytitle("") xtitle("")
    yscale(range(-0.02 0.39)) xscale(range(-0.3,9.5))
    xlabel(0 1 2 3 4 5 6 7 8 9, valuelabel alternate )
    title("Predictions of Good Health, by Government Office Region",
    size(large) justification(center) )
    subtitle("Confidence intervals of regression coefficients, by estimation method",
    size(medsmall) justification(center) )
    note("Source: UK Census 2001 SARS for England, n=1099294."
    "Model 1: Logistic regression predicting 'Good Health'. Other controls for education and gender"
    "Gayle and Lambert (2007 p.1195)", justification(left) )
    legend( order(1 2)
    label(1 "Conventional regression") label(2 "Quasi-Variance") );
#delimit cr

```

# A brief role-play using Stata...



# The Four Pillars of Wisdom



# A Planned Workflow Has Benefits



Effectiveness

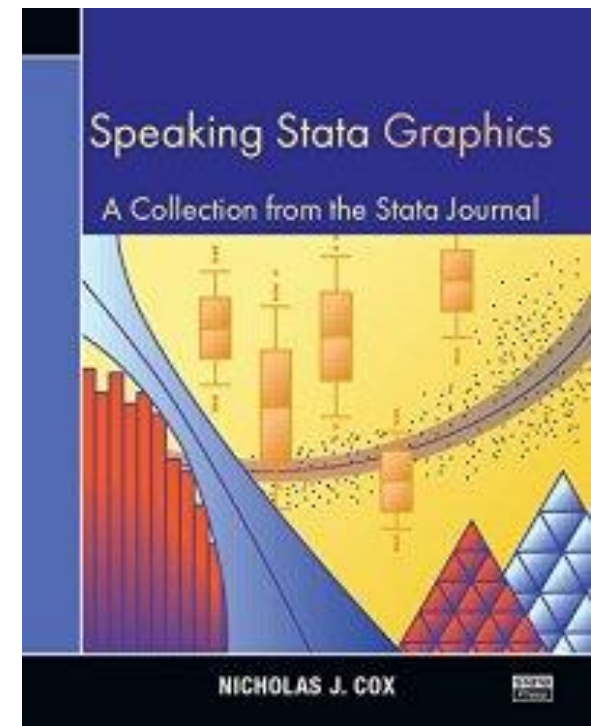
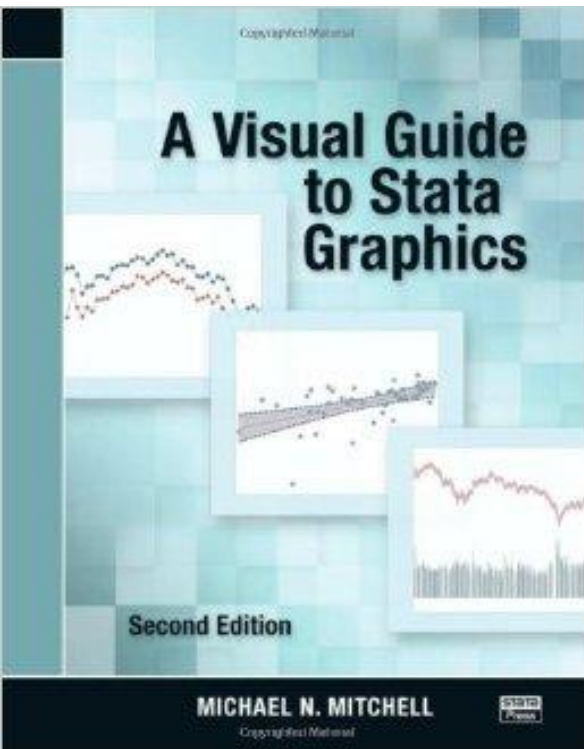
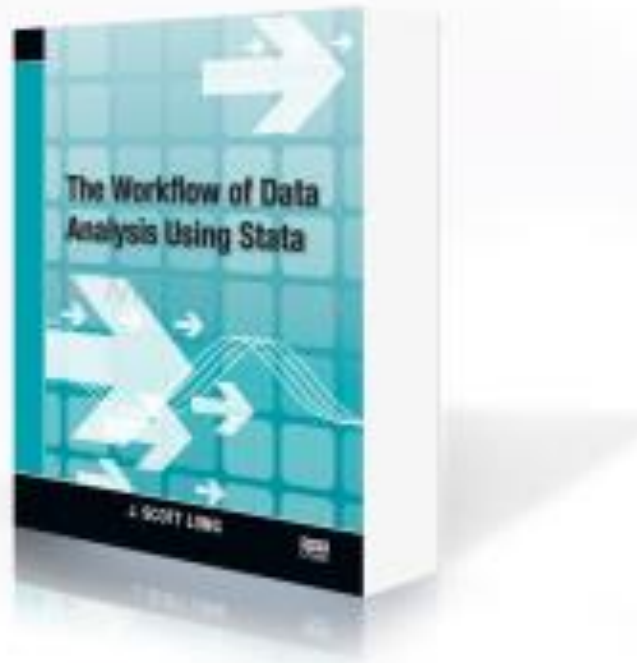
Efficiency

Transparency

Reproducibility

# Four Pillars of Wisdom

- Effectiveness
  - minimising information loss and errors in analyses and output
- Efficiency
  - automation, maximising features in software
- Transparency
  - showing what you did, why, when, how
- Reproducibility
  - same results every time whoever or wherever
  - editing, rewriting dissertation or re-submission of papers



The best habit that you can get into



is to get into good habits!

# The Workflow

Drukker's dictum: Never type anything that you can obtain from a saved result

My dictum (Gayle's dictum):

You can't be too fit or have too many publications

However...



# Long's Law

*It is always easier to document today than it is tomorrow!*

*Corollary 1:*

*Nobody likes to write documentation*

*Corollary 2:*

*Nobody every regrets having written documentation*

# Long's Law

*Has anyone in the history of data analysis ever said*

*“these files are too well documented”*



# Some Guiding Thoughts

- Graphing should be part of the workflow
- Never resort to using Excel
- Always use syntax
- You can never have too many comments

# Some Guiding Thoughts

- Graphs should be 'standalone'
- ABC Always Be Clear
- KIS Keep It Simple
- Black and white often required for journal articles
- Minimise grey
- Colors can be helpful for presentations (but be thoughtful)

# Some Guiding Thoughts

- Be smart! Figure 4 might be Figure 22 in two years time

*Think carefully about what is included in your graph and what can best be handled by your document management system*

# Graphing might even become fun...

