

Development of the HotProducts software system using R and gWidgets

A scanning tool for trends in property crime

Dr Lisa Wainer

Department of Security and Crime Science

University College London

Jill Dando Institute

l.wainer@ucl.ac.uk

LondonR 7th September 2011

Overview

- Our aims and some background
- Developing the software system
- Screen shots
- Software benefits
- Conclusions and further work

Our aims were to

1. Identify hot products
2. Discover emerging hot products
3. Present this information in a timely way

The HotProducts software was developed, in association with police analysts, that goes part way to fulfilling these aims.

What is a hot product?

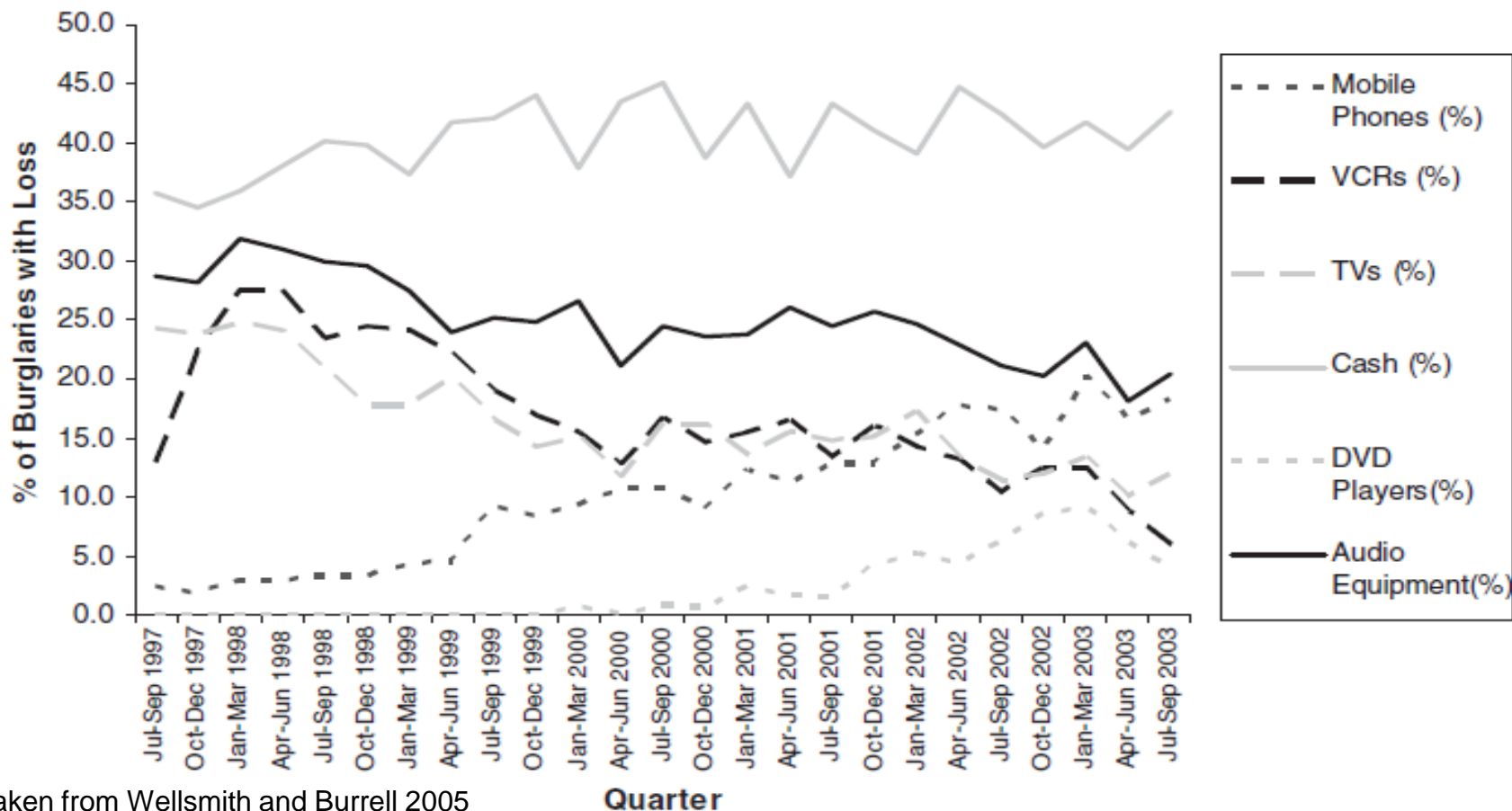
- Crime concentrates in time and space and on individuals
- Some items are stolen more frequently than others e.g. cars, laptops, mobile phones
- These items are referred to as Hot Products
- Hot products are product items that are desirable to thieves
- Hot products can be categorised as having certain attributes: C.R.A.V.E.D.

Why are police interested?

- Crime science is interested in understanding patterns of crime with a view to crime prevention
- Prevention research – specific interventions more effective than general ones
- Current police analysis focuses crime type
- Want to view acquisitive crimes by product type
 - Moving from the general to the specific
- Target crime prevention activity better
 - Design against crime

Trends over 6 years from 1997-2003

Percentage of Burglaries with Specified Items Stolen, Quarterly



Taken from Wellsmith and Burrell 2005

Overview

- Our aims and some background
- Developing the software system
- Screen shots
- Software benefits
- Conclusions and further work

Issues

1. Identify hot products
 - What data exists? Where is it? What format?
 - How is data gathered and stored?
 - How can we extract it?
2. Discover potential new hot products
 - Text mining
3. Presenting the data
 - HotProducts software

Property Item data

- Property stolen in crime was recorded via the NICHE crime recording system
- The police database contains records of each property item stolen in a crime
- Multiple records for each crime
- These records contain two main sources of property data information
 - Drop down menu NICHE categories
 - Free text entry

The Hot Products software

R used for analysis of data to show

- Summaries : *which* products are stolen most often?
- Time series analysis : *when* are products stolen?
- Kernel density estimation : *where* are products stolen?

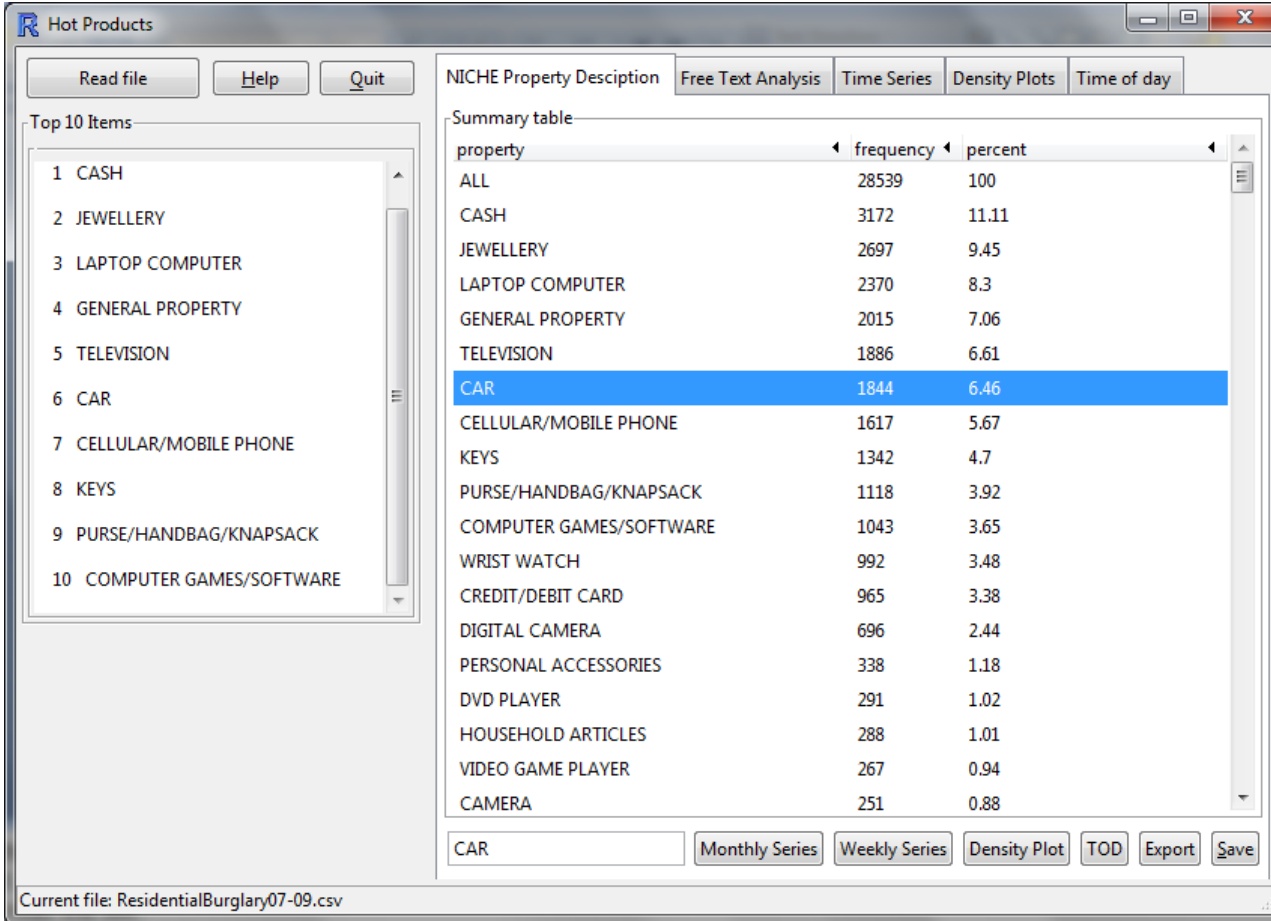
gWidgets used to develop the GUI

- gWidgets is a cross-toolkit API for working with GUI objects
- gWidgetsRGtk2 provides a link between gWidgets and the GTK libraries through the RGtk2 package

Overview

- Our aims and some background
- Developing the software system
- Screen shots
- Software benefits
- Conclusions and further work

Summary of residential burglary



The screenshot shows the 'Hot Products' software interface. The window title is 'Hot Products'. The main area is divided into two panes. The left pane, titled 'Top 10 Items', lists the following items in descending order of frequency:

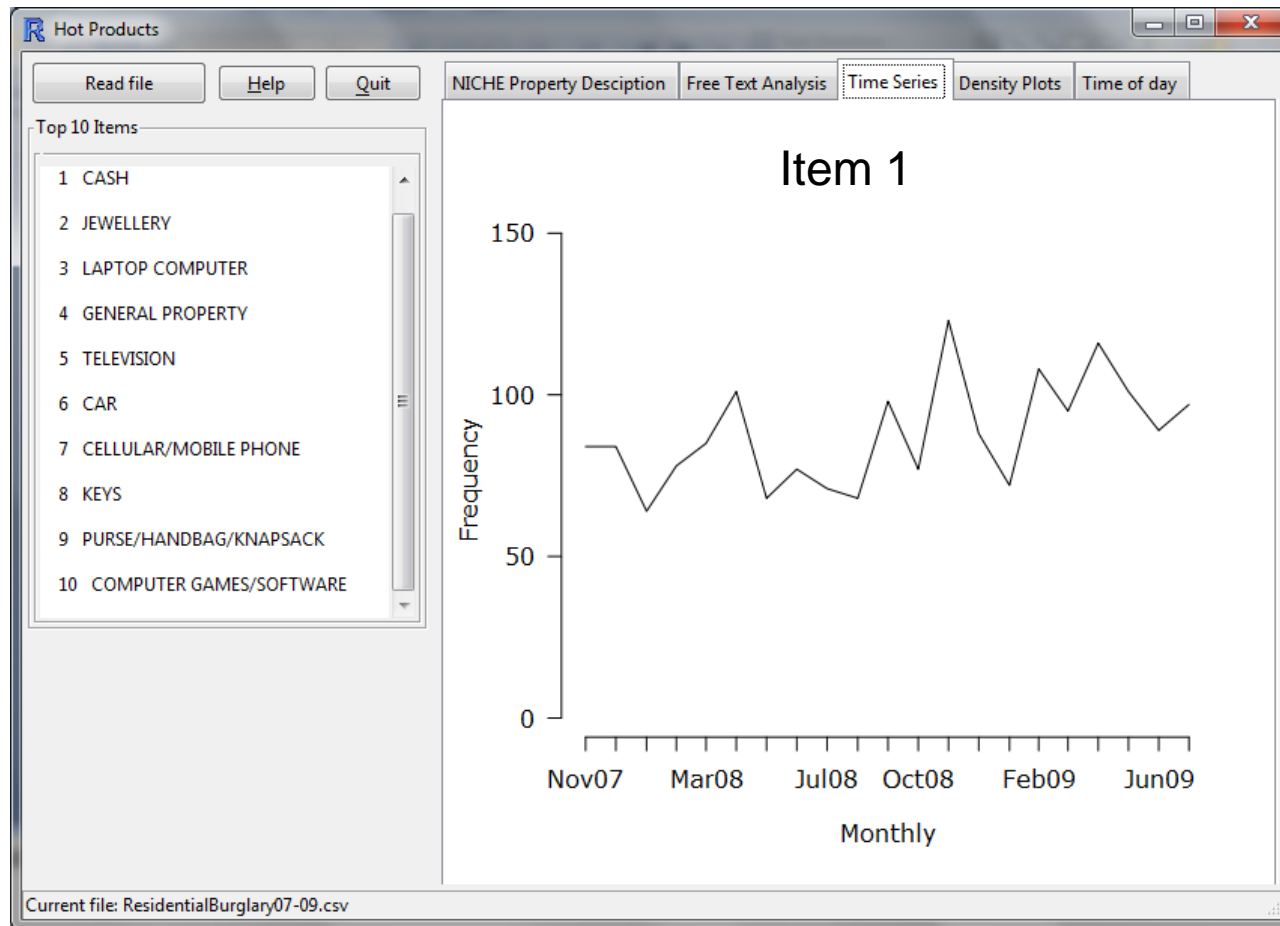
- 1 CASH
- 2 JEWELLERY
- 3 LAPTOP COMPUTER
- 4 GENERAL PROPERTY
- 5 TELEVISION
- 6 CAR
- 7 CELLULAR/MOBILE PHONE
- 8 KEYS
- 9 PURSE/HANDBAG/KNAPSACK
- 10 COMPUTER GAMES/SOFTWARE

The right pane, titled 'NICHE Property Description', shows a 'Summary table' with columns for 'property', 'frequency', and 'percent'. The 'CAR' row is highlighted in blue. The data in the table is as follows:

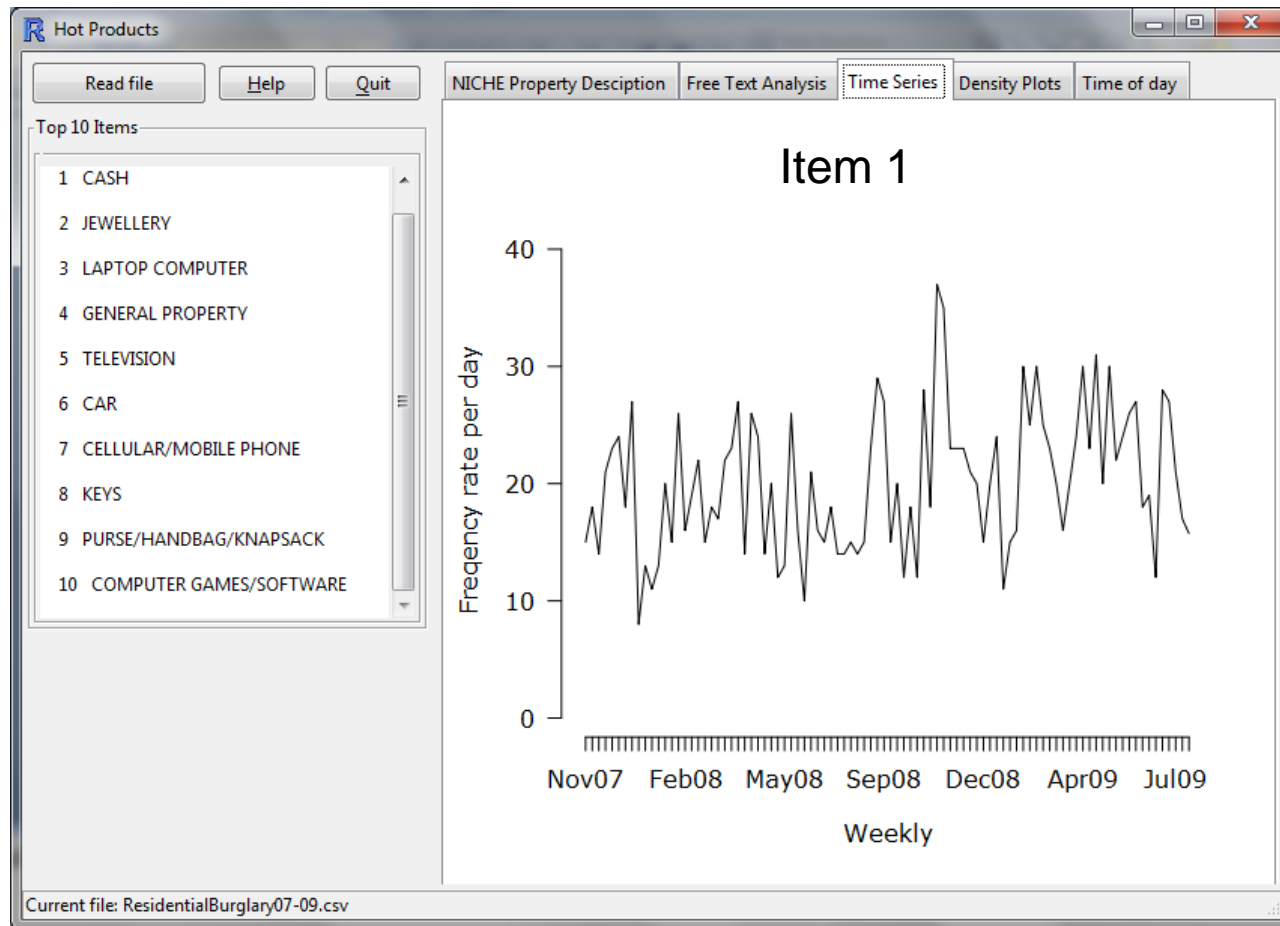
property	frequency	percent
ALL	28539	100
CASH	3172	11.11
JEWELLERY	2697	9.45
LAPTOP COMPUTER	2370	8.3
GENERAL PROPERTY	2015	7.06
TELEVISION	1886	6.61
CAR	1844	6.46
CELLULAR/MOBILE PHONE	1617	5.67
KEYS	1342	4.7
PURSE/HANDBAG/KNAPSACK	1118	3.92
COMPUTER GAMES/SOFTWARE	1043	3.65
WRIST WATCH	992	3.48
CREDIT/DEBIT CARD	965	3.38
DIGITAL CAMERA	696	2.44
PERSONAL ACCESSORIES	338	1.18
DVD PLAYER	291	1.02
HOUSEHOLD ARTICLES	288	1.01
VIDEO GAME PLAYER	267	0.94
CAMERA	251	0.88

At the bottom of the right pane, there is a search box containing 'CAR' and several buttons: 'Monthly Series', 'Weekly Series', 'Density Plot', 'TOD', 'Export', and 'Save'. The status bar at the bottom of the window indicates 'Current file: ResidentialBurglary07-09.csv'.

Item 1 burglary time series



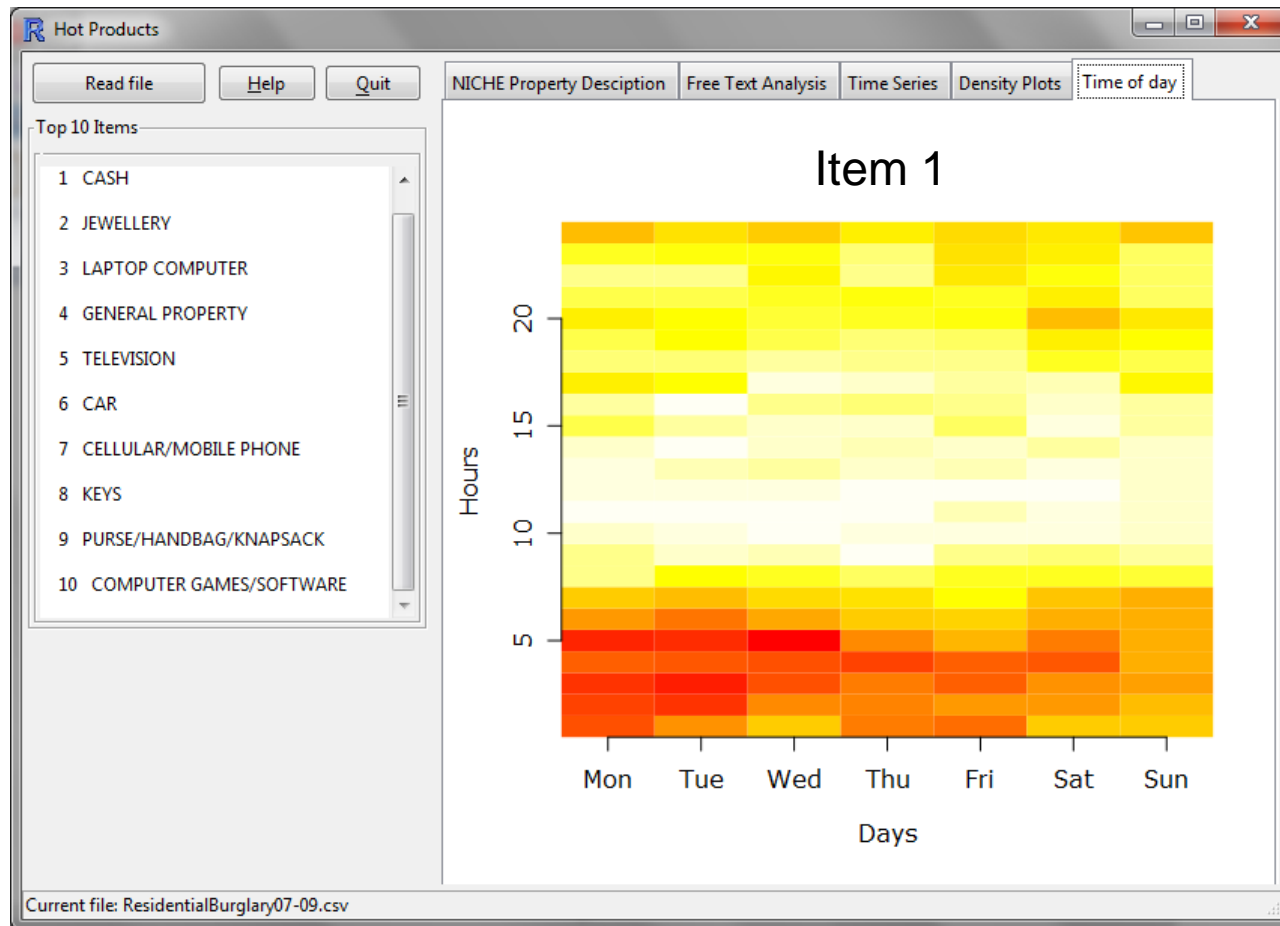
Item 1 burglary weekly time series



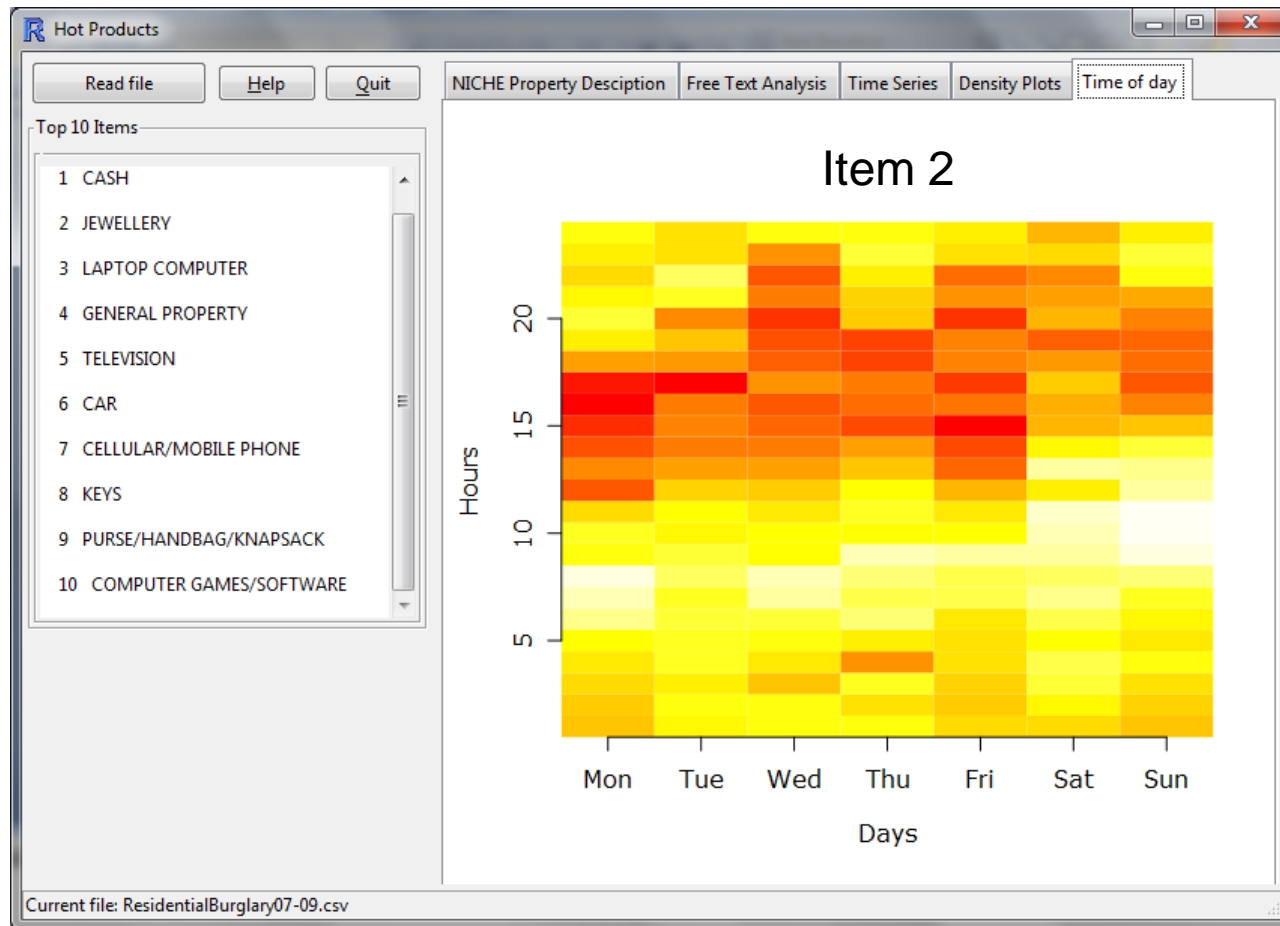
Comparing items

- The next two slides show the time of day and day of week analysis for two different items
- The screen shot images clearly show differing patterns for when item 1 is stolen as compared to when item 2 is stolen
- Item 1 is stolen predominantly late at night and
- Item 2 in the mid afternoon
- The differences are easily spotted from this type of visual representation

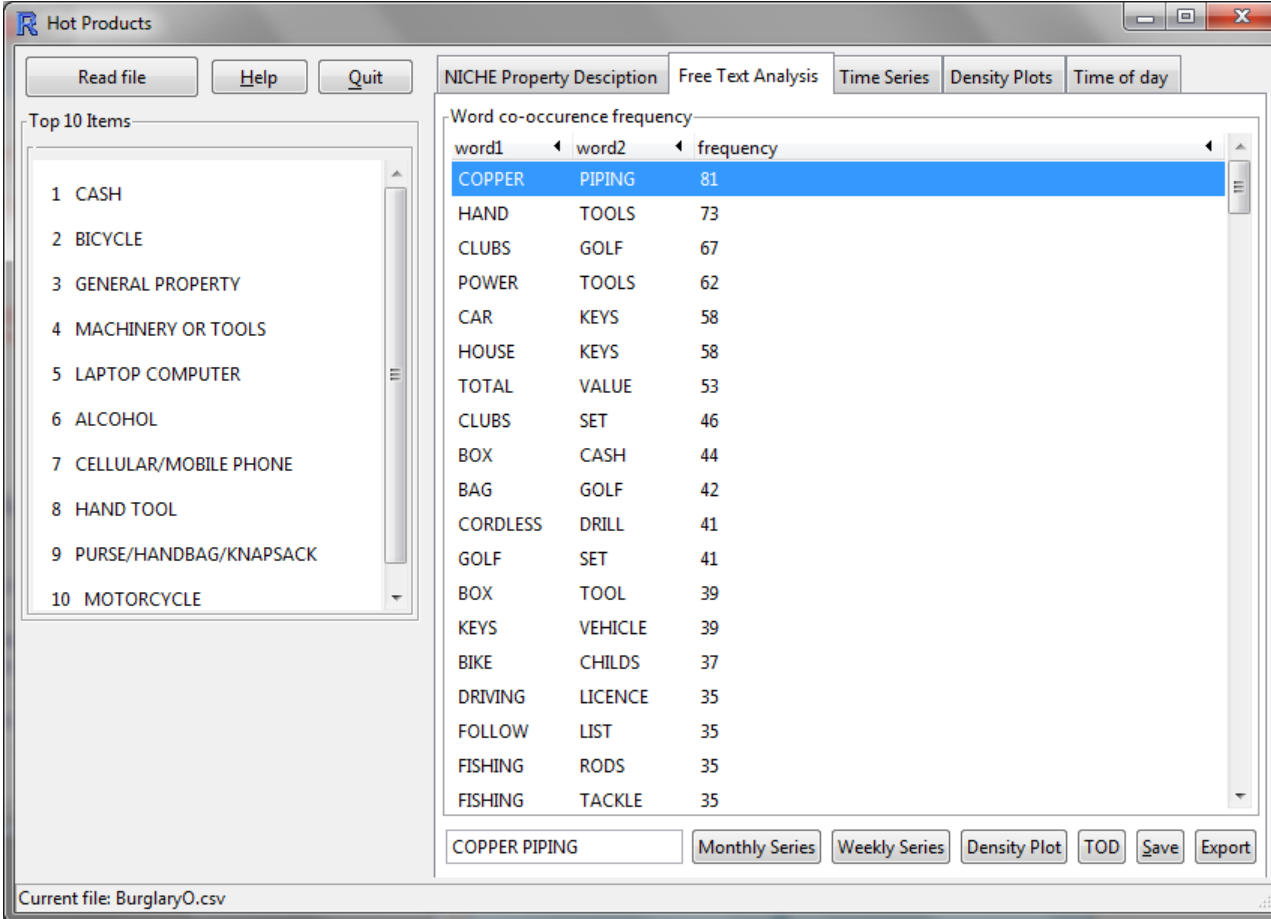
Item 1 burglary time of day and day of week



Item 2 burglary



Free text analysis non-residential burglary



The screenshot displays the 'Hot Products' software interface, which is used for analyzing text data. The window title is 'Hot Products' and it has standard Windows window controls. The interface is divided into several sections:

- Top 10 Items:** A list of the most frequent items found in the text, ranked from 1 to 10.
- Word co-occurrence frequency:** A table showing pairs of words that frequently appear together and their frequency.
- Buttons:** A row of buttons at the bottom right for further analysis, including 'Monthly Series', 'Weekly Series', 'Density Plot', 'TOD', 'Save', and 'Export'.
- Current file:** The bottom left corner shows the current file being analyzed is 'BurglaryO.csv'.

The 'Top 10 Items' list is as follows:

- 1 CASH
- 2 BICYCLE
- 3 GENERAL PROPERTY
- 4 MACHINERY OR TOOLS
- 5 LAPTOP COMPUTER
- 6 ALCOHOL
- 7 CELLULAR/MOBILE PHONE
- 8 HAND TOOL
- 9 PURSE/HANDBAG/KNAPSACK
- 10 MOTORCYCLE

The 'Word co-occurrence frequency' table is as follows:

word1	word2	frequency
COPPER	PIPING	81
HAND	TOOLS	73
CLUBS	GOLF	67
POWER	TOOLS	62
CAR	KEYS	58
HOUSE	KEYS	58
TOTAL	VALUE	53
CLUBS	SET	46
BOX	CASH	44
BAG	GOLF	42
CORDLESS	DRILL	41
GOLF	SET	41
BOX	TOOL	39
KEYS	VEHICLE	39
BIKE	CHILDS	37
DRIVING	LICENCE	35
FOLLOW	LIST	35
FISHING	RODS	35
FISHING	TACKLE	35

The 'Free Text Analysis' tab is selected, and the 'Word co-occurrence frequency' table is the active view. The 'COPPER PIPING' pair is highlighted in blue. Below the table, the text 'COPPER PIPING' is entered into a text box, and the 'Monthly Series' button is highlighted.

Overview

- Our aims and some background
- Developing the software system
- Screen shots
- Software benefits
- Conclusions and further work

How does HotProducts help?

- Scanning tool – hypothesis generation
- Time saving – months of data coding and analysis drops to seconds
- Exploration of free text
 - Successfully identifies objects that are ‘items of property’
 - Discovers potential new target property items e.g. ‘copper piping’ or trends ‘car keys’
- Outputs can now be filtered by product type (and used for reports etc)

Overview

- Our aims and some background
- Developing the software system
- Screen shots
- Software benefits
- Conclusions and further work

Summary

- Hot Product discovery can be automated to a significant degree given the right tools
- The software provides an overview for what is being stolen, when and where
 - Estimates the current state of the hot product landscape
- The software provides clues and ideas for further investigation
 - Hypothesis generation
- Operational advantage of speed
 - The software has massively reduced processing time

Further work

- Porting to other forces
 - Those that use NICHE most promising
- Easier installation process
- Spectral clustering of words in free text
- Generating prettier density plots

Acknowledgements

John Verzani

Author and maintainer of gWidgets API

<http://cran.r-project.org/web/packages/gWidgets/vignettes/gWidgets.pdf>

Professor Shane Johnson

UCL Department of Security and Crime Science

Merseyside Police

Strategic Development Department, Research and Analysis Unit

This project was funded by the Home Office and Merseyside Police

Many thanks

Dr Lisa Wainer

Department of Security and Crime Science

University College London

Jill Dando Institute

l.wainer@ucl.ac.uk

LondonR 7th September 2011