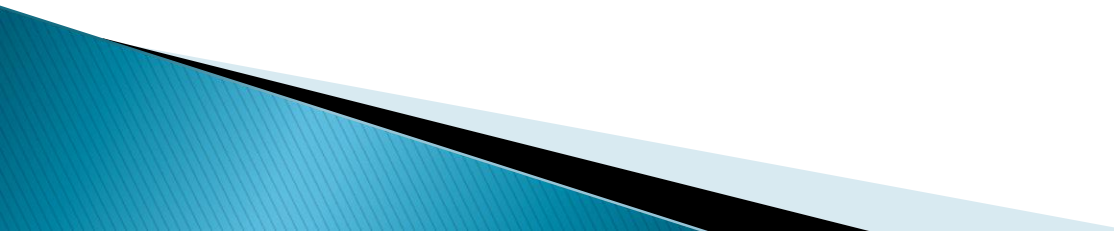


# Turning data into Research at the RTA

Steve Croft  
Research Manager, Regulatory  
Services

SNUG, May 2011

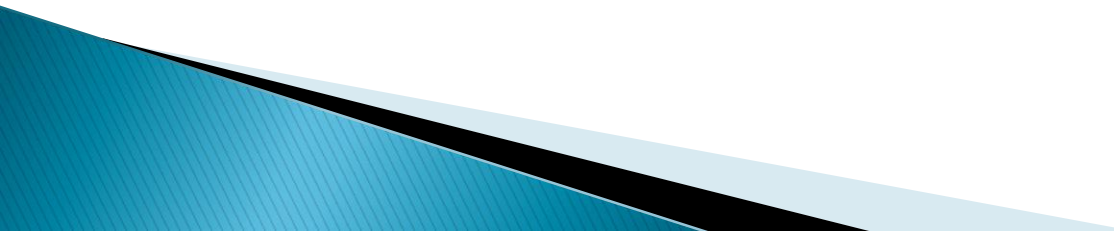
# DATA AT THE RTA

- ▶ Research Unit provides data and “research” internally externally on Licensing and Registration.
  - ▶ A number of databases provide operational data to provide insights into Licensing and Registration
  - ▶ One key objective – making NSW roads safer.
- 

# DATA SOURCES

- ▶ **DRIVES** contains:
  - Vehicles registered in NSW
  - Licence history – drivers and riders
  - Driver testing
  - Driving offences and outcomes
  - Demographic data (age/gender/residential locality)
  - RTA Business transactions
- ▶ **CRASHLINK** contains:
  - Crashes data in NSW for last 20 years

# HOW SAS IS USED

- ▶ 8–10 user licence on stand-alone PCs.
  - ▶ Including SAS Access to Oracle, SAS Graph.
  - ▶ Many RTA's databases are Oracle – based.
  - ▶ Transfer rate from UNIX servers to SAS using Access to Oracle of 1.2Gb per minute
  - ▶ Occasional requests from IT to use SAS's off-the shelf tools rather than SQL / other languages.
- 

# Information or Research

- ▶ 700 requests for information per year
  - suspended drivers
  - vehicle sales
  - heavy vehicle licence holders
  - Mobility Parking Permits
  - Driving Test results
- ▶ Facts to justify a position.

# RTA Menu-driven reports

Home Motor registries

myRTA online services

Live traffic NSW

Centre for Road Safety

GEARED For L and P drivers

- Licensing
- Registration
- Using NSW roads
- Rules and regulations
- Environment
- Construction and maintenance
- Heavy vehicles
- Careers in the RTA
- Publications, statistics & forms**
  - RTA forms
  - Statistics
    - Licence statistics
    - Registration and licensing**
    - Safe-T-Cam offences
    - Performance data
    - Traffic volume maps

## Licensing

Home > Publications, statistics & forms > Statistics > Registration and licensing > Licensing

- Licensing**
- Registration
- Licensing
  - Driver and rider suspensions and cancellations
  - Mobility Parking Scheme
  - NSW Photo Cards

### 2.1. Licences on issue in NSW

1. <u>Quarterly trend in licence class</u>	
2. Licence class by licence type as at (quarter/year)	2010: q4 <input type="button" value="Go"/>
3. Licence class by age group as at (quarter/year)	--Select year and quarter-- <input type="button" value="Go"/>
4. Licence class by LGA as at (quarter/year)	--Select year and quarter-- <input type="button" value="Go"/>
5. Licence class by postcode as at (quarter/year)	--Select year and quarter-- <input type="button" value="Go"/>
6. Licence class by licence type by licence holder's age group as at (quarter/year)	--Select year and quarter-- <input type="button" value="Go"/>

# HTML Reports

**Table 2.1.2 Licence class by licence type as at 31 December 2010**

Licence type	TOTAL	Licence class						
		Class C	Class LR	Class MR	Class HR	Class HC	Class MC	Class R
Learner	288,748	263,552	0	0	0	0	0	25,196
P1	164,970	145,295	0	0	0	0	0	19,675
P2	254,650	251,327	136	1,206	675	0	0	1,306
Unrestricted	4,644,783	3,647,936	84,042	123,369	200,391	112,575	19,523	456,947
<b>TOTAL</b>	<b>5,353,151</b>	<b>4,308,110</b>	<b>84,178</b>	<b>124,575</b>	<b>201,066</b>	<b>112,575</b>	<b>19,523</b>	<b>503,124</b>

*Note: A person may hold a rider licence class as well as a driver licence class.*

# PLDB

- ▶ SAS dataset with 3 dimensions:
  1. Variable of interest
  2. Geography
  3. Time period
- ▶ Statistics held against these dimensions
  - counts of licence holders or
  - average age of vehicle or
  - knowledge test pass rate.
- ▶ Instantaneous access using a SAS AF front end.

# PLDB - AF Front end

SAS 10.51.15.198

File View Tools Solutions Window Help

Results

## Research Unit Processed-Level Database

Select database:  
Registration transfers

Select variable:  
New vehicles by fuel type

Select geography:  
REGION

Select start date:  
31JAN2000

Select end date:  
30NOV2010

Select date format:  
Financial year

Select second row variable:  
Geography

Select column variable:  
Variable

Include totals

Include totals

Select first row variable:  
Date

Include totals

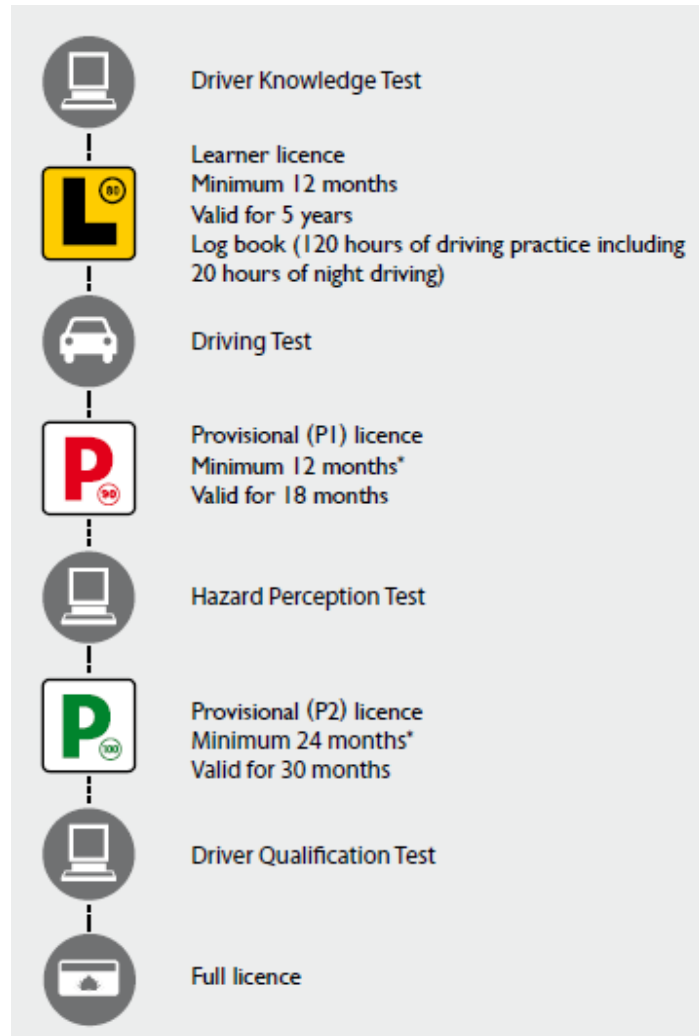
		Licence holder sex female	Licence holder sex male	Licence holder sex unknown	TOTAL
31/03/2007	Region - Sydney	1352264	1489963	1827	2843959
	Region - Rest NSW	828520	867905	1696	1698221
	Region - Other	860	1877	11	2658
	TOTAL	2181664	2358853	3534	4544058
30/06/2007	Region - Sydney	1386604	1498940	1820	2887364
	Region - Rest NSW	834268	871333	1710	1707331
	Region - Other	809	1073	6	1883
	TOTAL	2201701	2371351	3536	4576588
30/09/2007	Region - Sydney	1373094	1504723	1795	2879612
	Region - Rest NSW	836531	873401	1736	1711662
	Region - Other	805	1094	5	1904
	TOTAL	2210830	2378813	3530	4592178
31/12/2007	Region - Sydney	1380169	1526003	1788	2907960
	Region - Rest NSW	839773	875285	1746	1716804
	Region - Other	802	1083	6	1891
	TOTAL	2220744	2389973	3540	4613197
31/03/2008	Region - Sydney	1385862	1516496	1772	2904130
	Region - Rest NSW	842251	876527	1750	1724528
	Region - Other	807	1109	5	1921
	TOTAL	2228920	2394132	3527	4626579

Results Output - (Untitled) Log - (Untitled) pldb\_start.sas

# Research projects

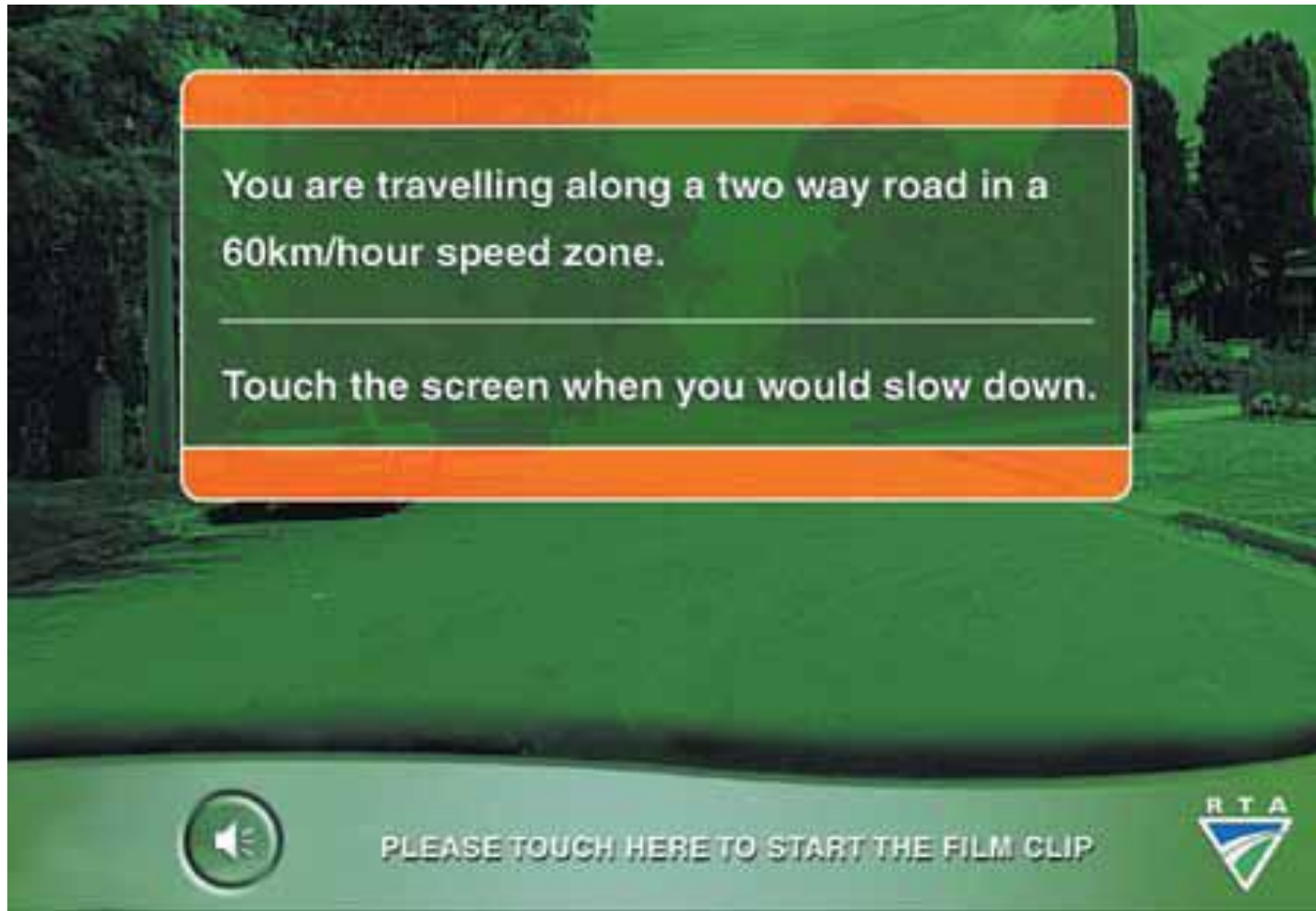
- ▶ Several datasets created combining the Crash data with other datasets.
- ▶ Uses are:
  - Evaluate Driver Education
  - Examine outcomes of driver behaviour
  - Examine crash likelihoods

# HPT and DQT evaluations



- ▶ Hazard Perception Test  
P1 to P2
- ▶ Driver Qualification Test:  
P2 to Unrestricted
- ▶ Crash data joined to test  
performance data.
- ▶ Is a poor score  
associated with a greater  
likelihood of a crash?

# HPT and DQT evaluations



Source: Hazard perception handbook, RTA

# HPT and DQT evaluations

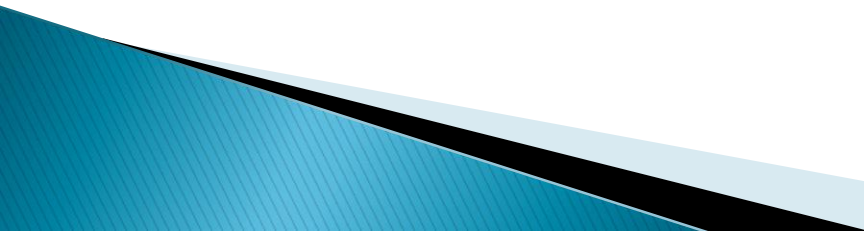


Source: Hazard perception handbook, RTA

# HPT and DQT evaluations

- ▶ Logistic Regression (Propensity modelling) likelihood of crash in following 2 years controlling for
  - Age
  - Gender
  - Socio-economic status
- ▶ First time failure on HPT = 10% more likely
- ▶ First time failure on DQT = 16% more likely
- ▶ Subsequent failures = greater

# HPT and DQT evaluations

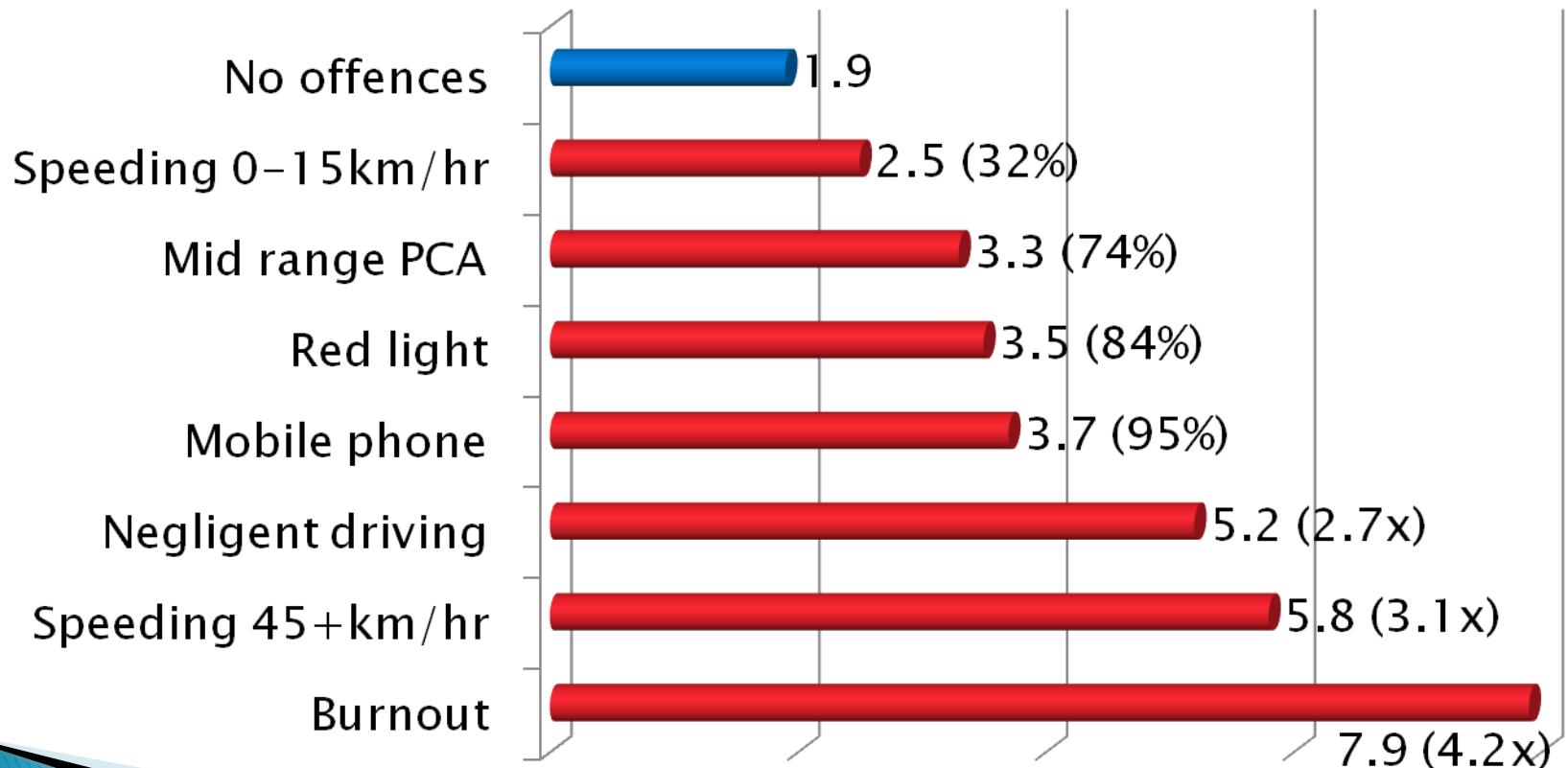
- ▶ Pool of questions examined for predicting types of crashes, such as
    - adjacent direction
    - same direction
    - overtaking
    - alcohol-related
  - ▶ Some questions performed well, others did not.
  - ▶ Customer Education able to vary frequency of presentation of clips.
  - ▶ Crash types not adequately covered.
- 

# Offence history

- ▶ Link between driving offence and sanction history and crash likelihood.
- ▶ Propensity modelling to calculate the probability of a driver having a crash based on :
  - history of driving offences
  - history of sanctions
  - age
  - gender
  - socio-economic and other variables

# Offence history

- ▶ Offences involving greater risk-taking resulted in a much greater likelihood of a subsequent crash.



# Offence history

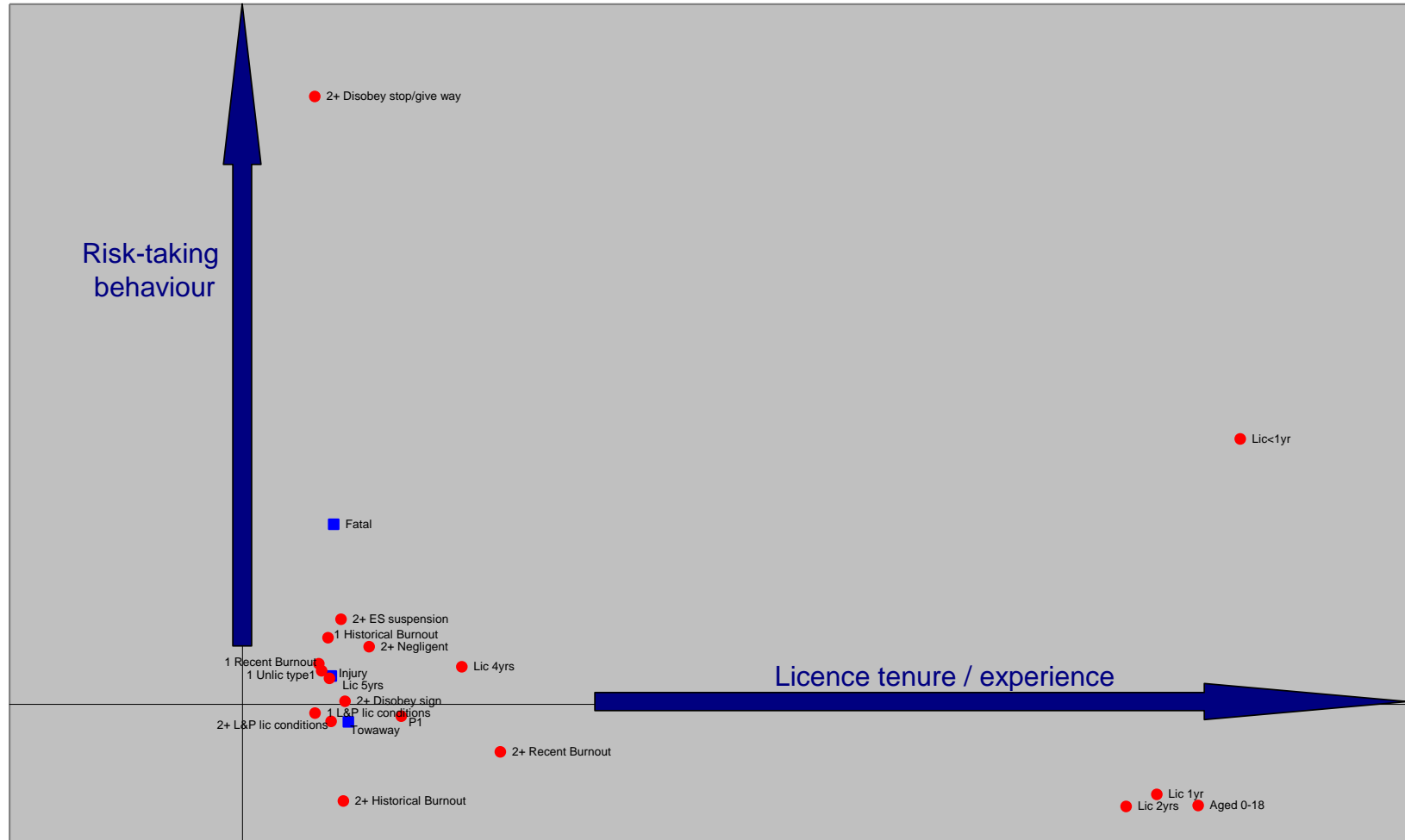
- ▶ Correspondence Analysis portrays hundreds of crosstabs in one scatterplot.
- ▶ Principal Components Analysis derives two “factors” that best predict associations between the variables – in our case, the crash type (Fatal/Injury/Towaway) and numbers of offences by offence group.

# Correspondence Analysis

- ▶ 70 crosstabs :  
 offence x number of offences x crash type  
 = (70) <sup>3</sup> = 3 tables

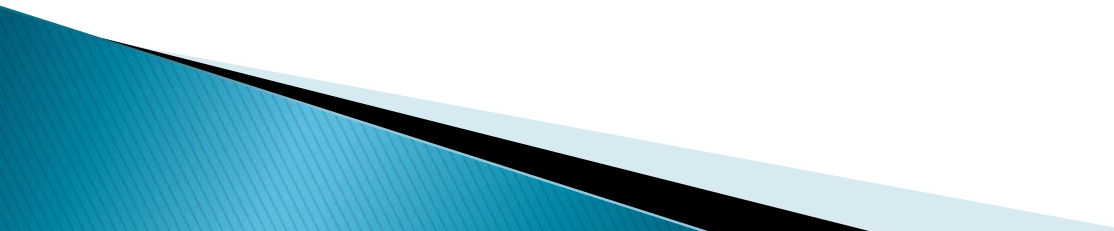
RECENT OFFENCES (prior 4 years)	Crash outcome								Crash Probability
	No Crash		Fatal		Injury		Towaway		
	N	%	N	%	N	%	N	%	
<b>Speeding 0-15km/hr above limit</b>									
0	3,330,101	98.1	496	0	25,465	0.8	38,339	1.1	1.9%
1	705,678	97.4	141	0	7,673	1.1	11,223	1.5	2.5%
2+	284,392	96.9	66	0	3,729	1.3	5,409	1.8	3.1%

# Correspondence analysis





# myRTA penetration

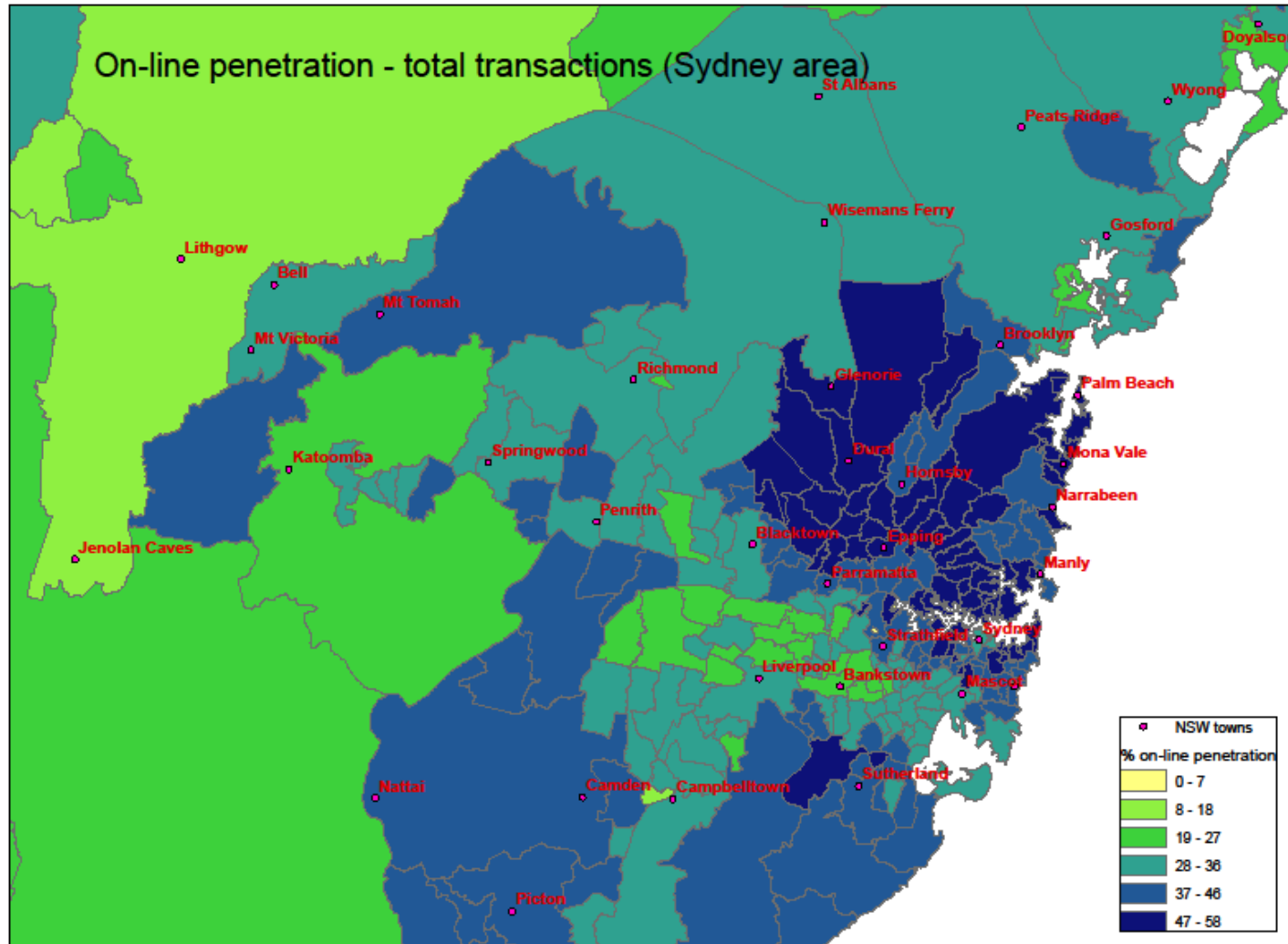
- ▶ the RTA is attempting to improve customer service through the myRTA website.
  - ▶ Data extracted from the log of customer transactions to examine customer usage of the online portal.
- 

# myRTA penetration

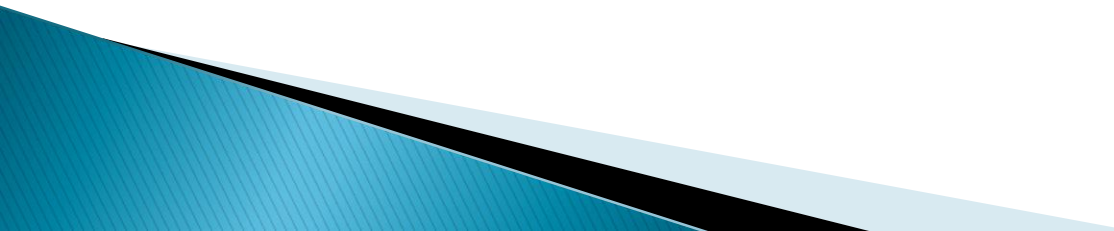
- ▶ *Penetration* = % of customers ever conducted online business
- ▶ *Repeat usage* = % who conduct a second online transaction
- ▶ Used to gauge the success of a new product trial.
- ▶ *Reach* (TV/Radio) or *Circulation* (Print media).



# Penetration by area



# myRTA penetration

- ▶ August 2009:
    - penetration = 33%
    - repeat transaction rate = 34%
  - ▶ Lag time between initial trial and return purchase is very important.
  - ▶ An annual registration renewal versus buying a loaf of bread (a week).
  - ▶ Useful in evaluating products and services, but not commonly used.
- 

# SAS AT THE RTA

- ▶ Base SAS to easily manipulate and analyse data.
  - ▶ Most used = Proc Frequency.
  - ▶ More complex analyses (regressions, factor analyses etc) require a good knowledge of the data.
  - ▶ Good understanding of the data and implications for the business.
  - ▶ SAS used to turn data into better policy.
- 