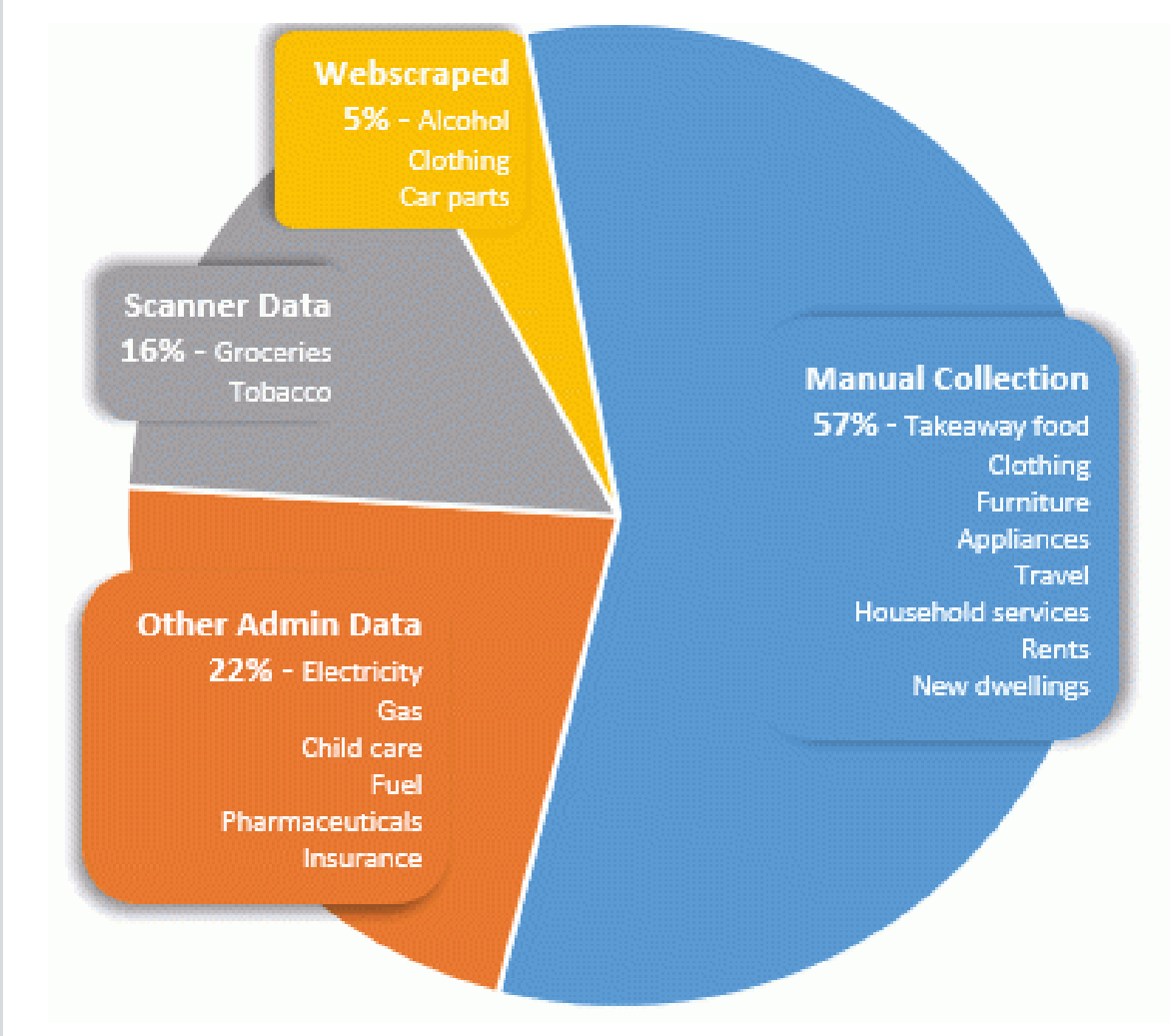
Use of scanner and web-scraped data in Australia's CPI

Leigh Merrington (leigh.merrington@abs.gov.au)
Catherine Smyth (catherine.smyth@abs.gov.au)

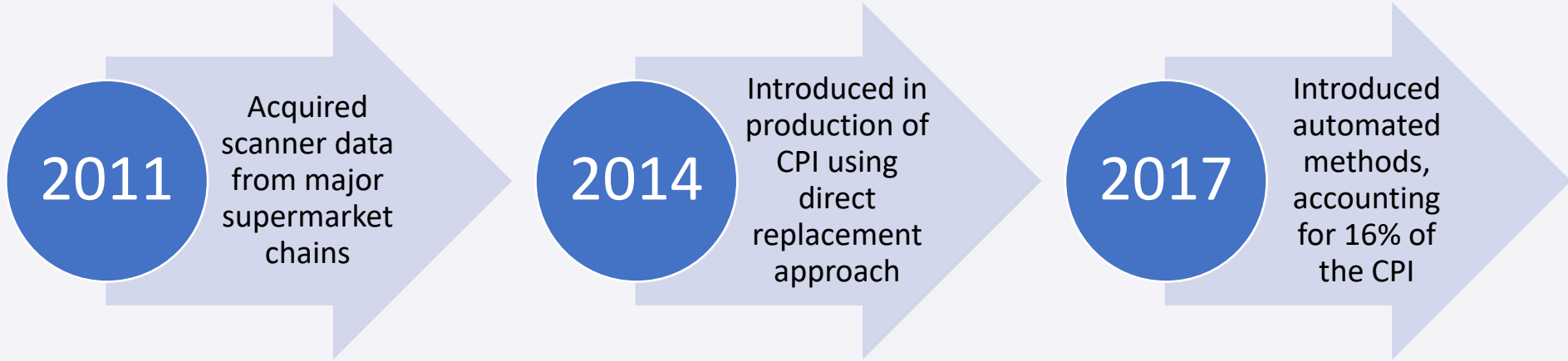
Australian Bureau of Statistics
Informing Australia's important decisions

Data sources used in the CPI



Scanner data in the CPI – a brief history

Australian Bureau of Statistics



Multilateral methods

- Use an entire census of data
- Contribution within a component varies depending on relative expenditure
- Provides insight into consumer behaviour

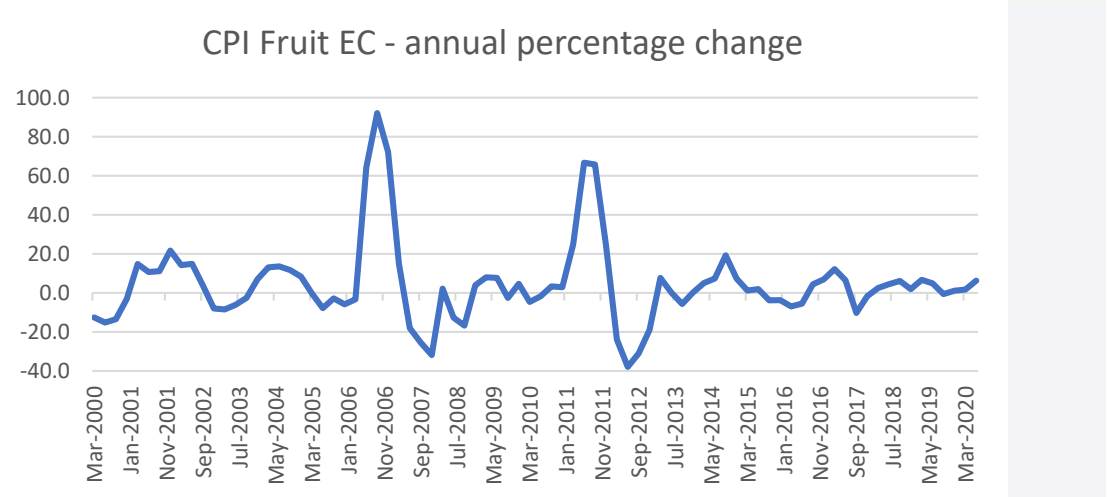
Currently used for the following Expenditure Classes:

- Bread
- Cakes & biscuits
- Breakfast cereals
- Other cereal products
- Beef & veal
- Pork
- Lamb & Goat
- Poultry
- Other meats
- Fish & other seafood
- Milk
- Cheese
- Ice cream & other dairy products
- Fruit
- Vegetables
- Eggs
- Jams, honey & spreads
- Food additives & condiments
- Oils & fats
- Snacks & confectionery
- Coffee. tea & cocoa
- Waters, soft drinks and juices
- Tobacco
- Cleaning and maintenance products
- Personal care products
- Other non-durable household products
- Pets and related products

Engagement/communication

- Regular communication with key users
  - papers, workshops, briefings
- Engagement with international experts and forums
- Parallel runs/proof of concept
- Ambitious change, cautious implementation

Fruit prices gone bananas!



Has the banana's impact on the CPI been overestimated?

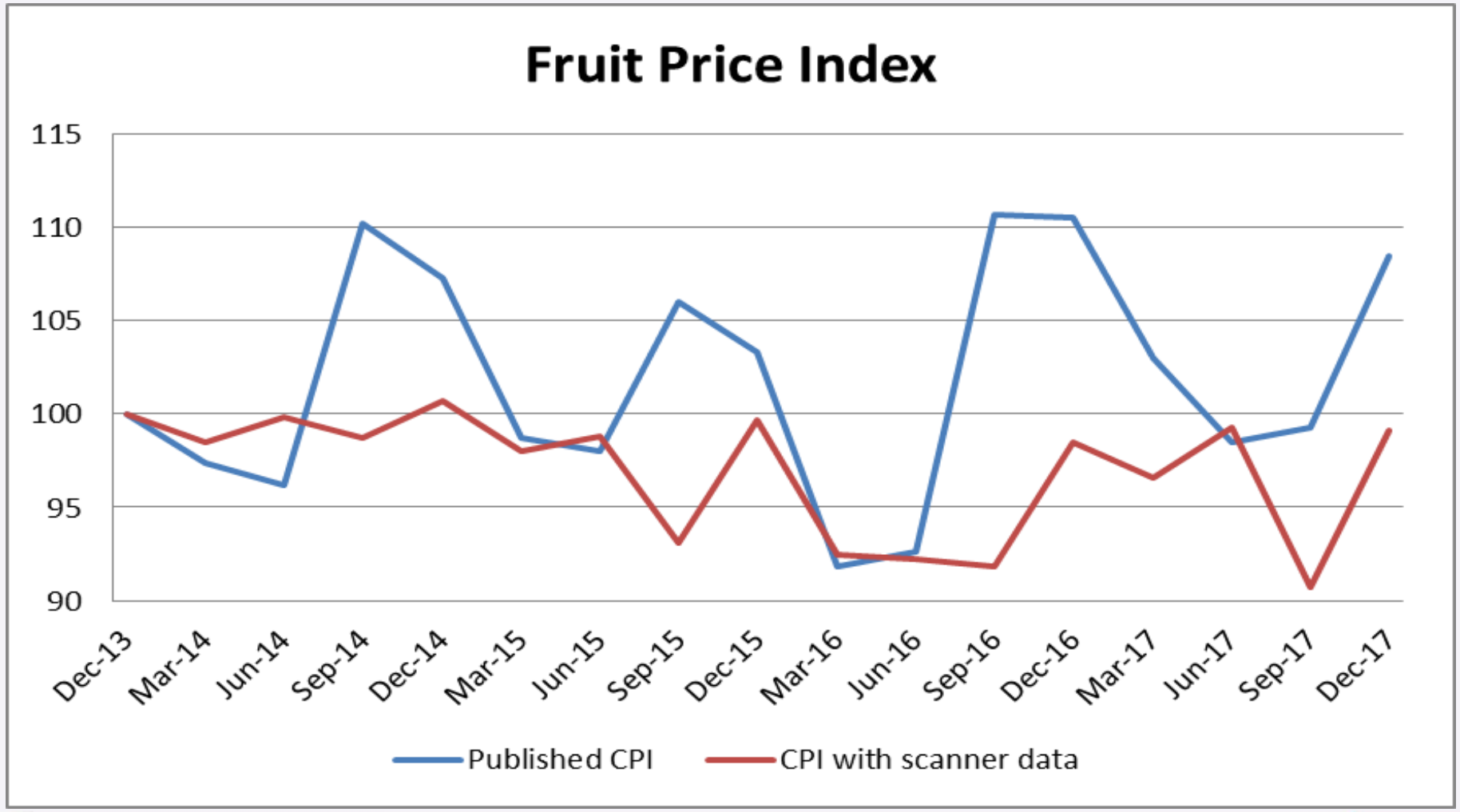
We're told that the banana price blow-out has hit the Consumer Price Index. According to the Australian Bureau of Statistics, fruit contribute 0.6 percentage points to the 4.0% through-the-year Consumer Price Ir increase. And the rise in fruit prices was mainly attributable to an increase of approximately 250% in the price of bananas during the [...]

Inflation stokes rate rise speculation but don't blame the bananas

July 28, 2011 4.00pm AEST

When it comes to bananas, demand outstrips supply. But it won't mean an interest rate rise. AAP

Increased measurement accuracy



Price impact of panic buying – MQ20

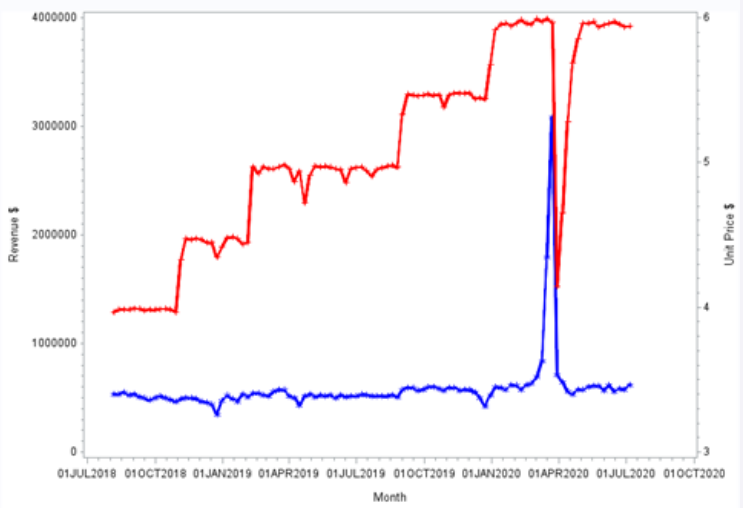
	Movement to mid March (11 weeks of data)	Movement to 31 March (13 weeks of data)	Change (%pts)
FOOD & NON-ALC BEVERAGES GROUP	1.5	1.9	+0.4
Bread and cereal products	0.5	1.2	+0.7
Meat and seafoods	1.4	2.0	+0.6
Dairy and related products	0.8	1.1	+0.3
Fruit and vegetables	4.9	6.0	+1.1
Food products n.e.c	1.6	2.3	+0.7
Non alcoholic beverages	1.2	2.0	+0.8

Select scanner data graphs

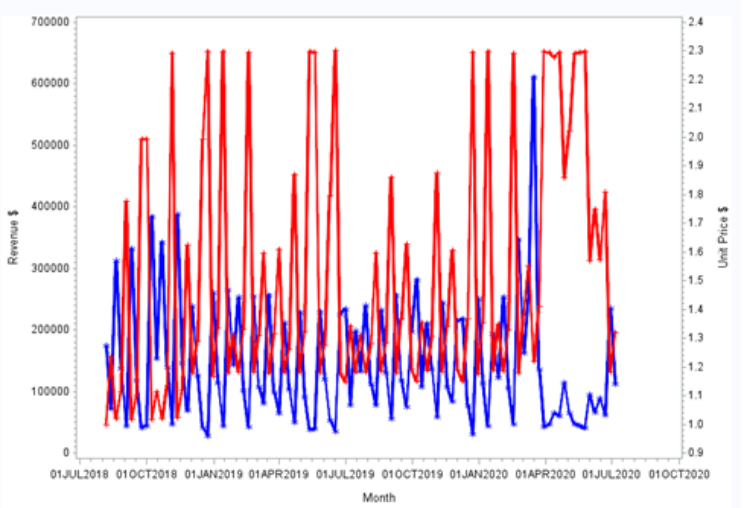
Price
Revenue

Australian Bureau of Statistics

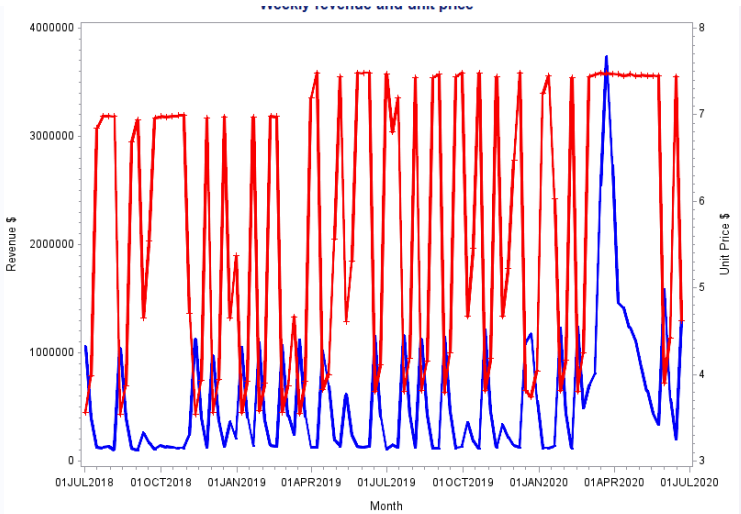
Beef Mince



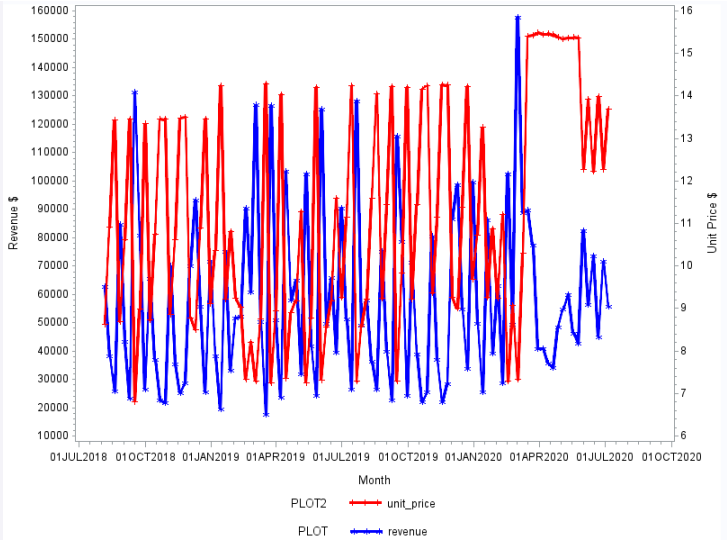
Canned Tuna



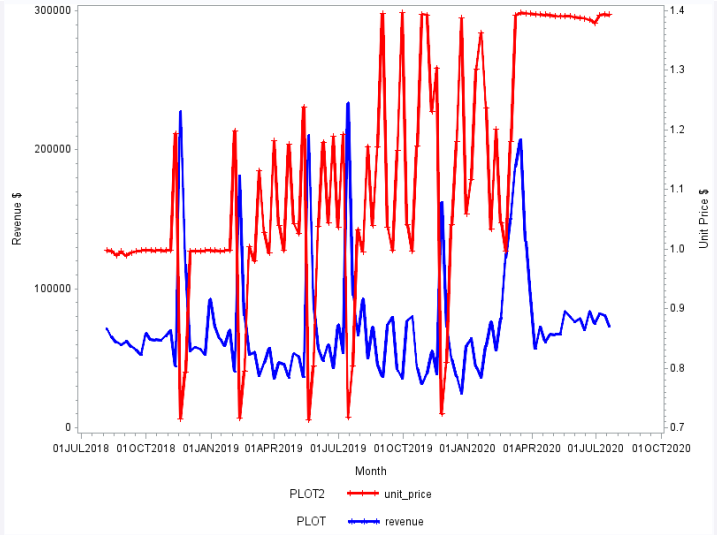
Toilet Paper



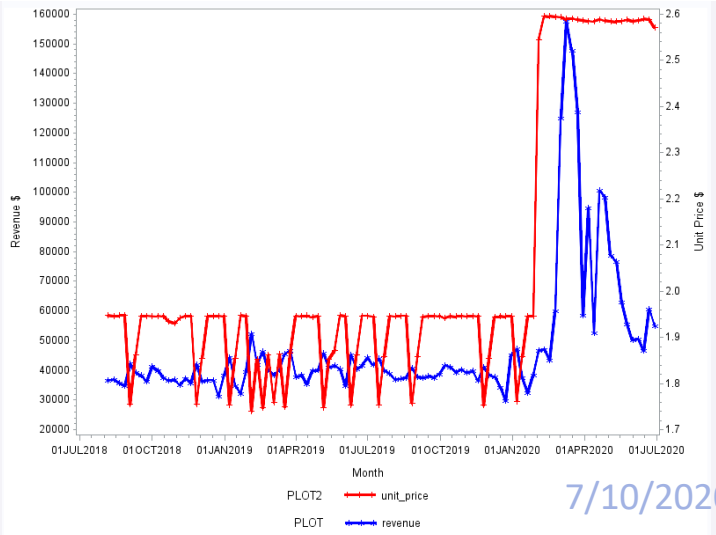
Meat Pies



Canned Tomato



Pasta



Continuous improvement

- Improve analytics
- Reduce time spent running code
- Intelligent coder to standardise classifications and code new products
- Other applications for multilateral methods
- New skills required to process and analyse the data

Web scraping overview

- Web scraping – an automatic collection method which extracts and converts unstructured website data into structured data.
- Web scraped prices progressively incorporated into the CPI since March 2017 – direct replacement strategy currently used.
- CPI Enhancement Team have been investigating methods to betterutilise online price data in the CPI since April 2018.
- Looking into potential for APIs.

Gold standard vs. silver standard

Scanner data	Web-scraped data
• 'Census' of products collected from each retailer • Includes weekly expenditure and quantities for each product • Products defined by stock keeping units	• 'Census' of products collected from each retailer • No expenditure or quantity information provided • Stock keeping units not currently scraped

Web-scraped data example

Date	Retailer	Category	Brand	Type	Characteristics	Description	Price	Count
02-Jan-17	Retailer ABC	Women's Tops	Brand XYZ	T-Shirt	Short Sleeves	Short Sleeve Regular T Shirt "Brand XYZ"	$55.00	1
05-Jan-17	Retailer ABC	Women's Tops	Brand XYZ	T-Shirt	Short Sleeves	S/S Regular Tee Brand XYZ	$55.00	1
05-Jan-17	Retailer ABC	Women's Tops	Brand XYZ	T-Shirt	Short Sleeves	Short Sleeved Oversized T-Shirt "Brand XYZ"	$55.00	1
05-Jan-17	Retailer ABC	Women's Tops	Brand XYZ	T-Shirt	Long Sleeves	Long Sleeve T.S. "Brand XYZ"	$65.00	1
28-Jan-17	Retailer ABC	Women's Tops	Brand XYZ	T-Shirt	Long Sleeves	L.S. Tee Shirt "Brand XYZ"	$65.00	1
28-Jan-17	Retailer ABC	Women's Tops	Brand XYZ	T-Shirt	Short Sleeves	Short-Sleeve Reg T-Shirt "Brand XYZ"	$55.00	1
28-Jan-17	Retailer ABC	Women's Tops	Brand XYZ	T-Shirt	Short Sleeves	Short Sleeved O/S Tee "Brand XYZ"	$55.00	1

Research questions

- How can we define individual products or homogeneous product clusters?
- Can alternative data sources be used to weight products/clusters in the absence of expenditure and quantity information?
- Which index method should be used to aggregate products/clusters to derive elementary aggregate indexes?
  – Bilateral vs multilateral indexes
- Presentation showcasing multilateral results

Questions

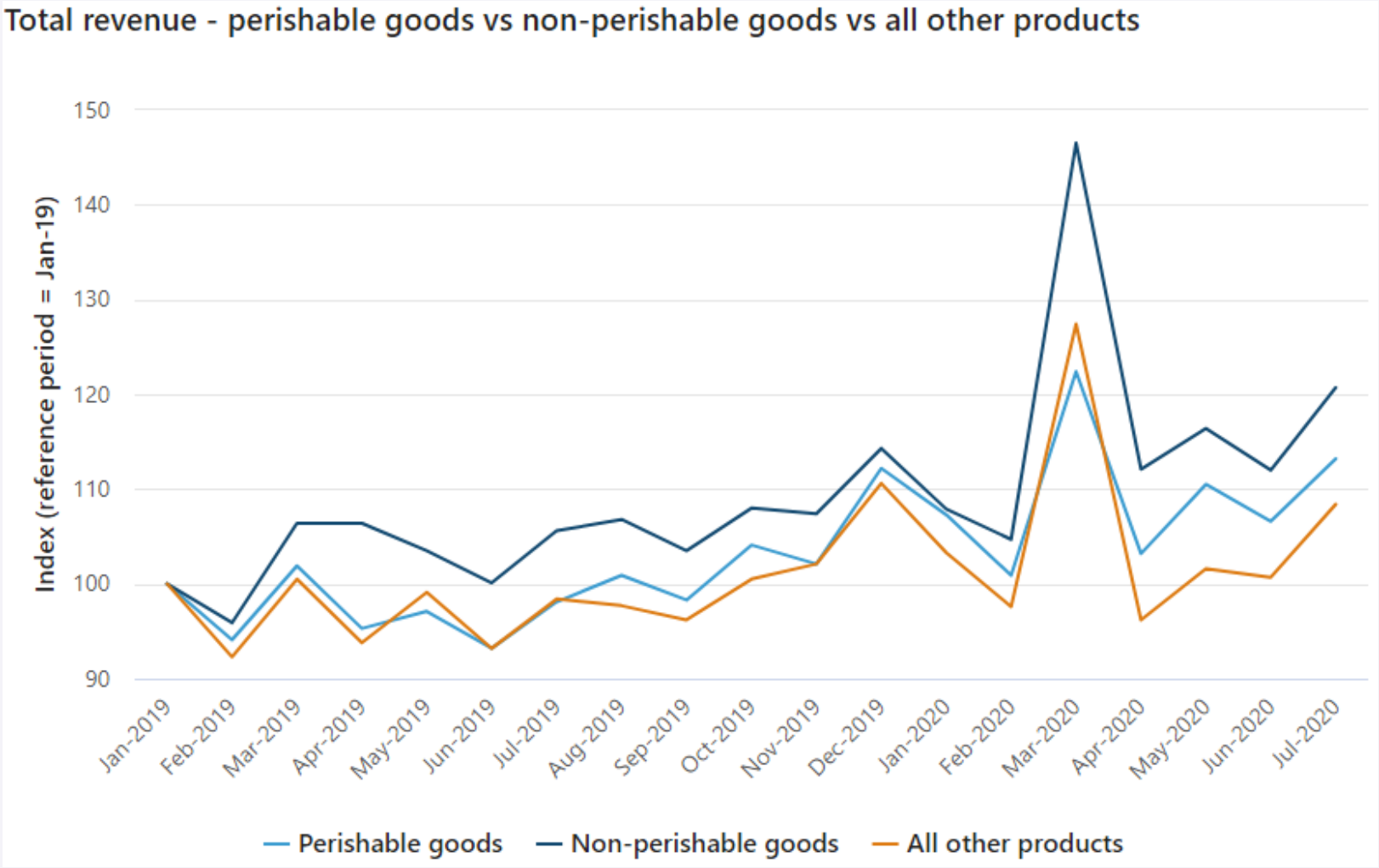
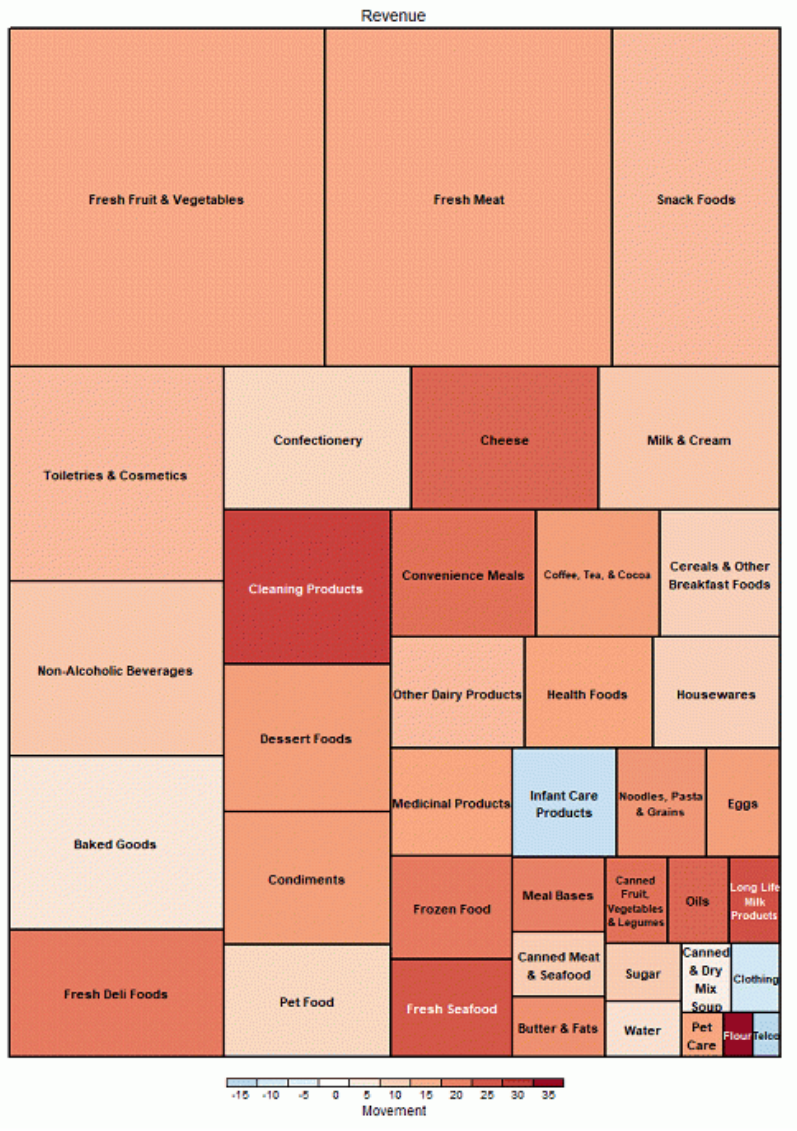
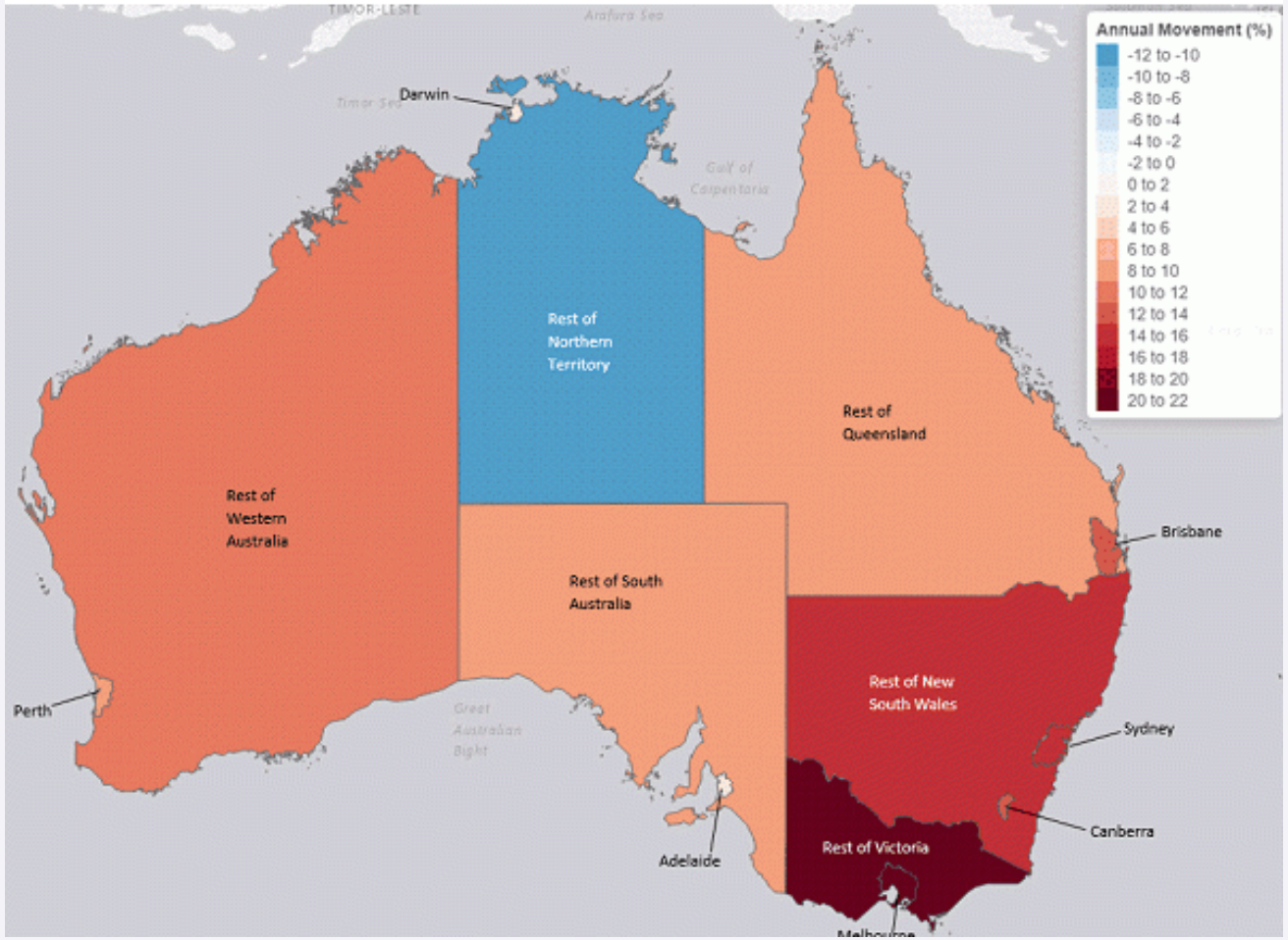
- Thanks for listening
  - Leigh Merrington
  - Catherine Smyth

- There’s a few more slides on scanner data if there’s time.

Other uses of scanner data

- Re-weighting at upper levels.
- Use in National Accounts (HFCE).
- Health statistics – nutrition consumption.
- Near real time spending tracker.....

Spending during COVID-19



Alternative views of spending