Research & Development

Authored by:
Michael W. Kruger
Daniel Pagni
Analytics Research & Development
*** Confidential Distribution[1] ***

[1] Distribution limited to those who have signed the appropriate contract for this dataset, agreeing to abide by the terms of use.

Table of Contents

1. Preliminary information

1.1 Revision History

Date	Version	Description	Author
September 13, 2006	0.5	First version for Mela, Bronnenberg, Johnson review	Dan Pagni, Mike Kruger
October 4, 2006	0.51	Minor notes	Mike Kruger
March 7' 2007	0.6	Extending to internal files	Mike Kruger
July 29, 2007	0.9	Aligning to first 3 years of data pull	Mike Kruger
August 28, 2007	1.0		Mike Kruger
November 13, 2007	1.0MSci	As sent to Mela for Marketing Science	Mike Kruger
March 26, 2008 – April 20, 2008	1.1, 1.11	Minor edits to prepare for distribution	Mike Kruger
May 8, 2008	1.12	Added IRI contract/nondisclosure and TNS Terms of Use as appendices	Mike Kruger
June 10, 2008	1.22	Added consolidated trip file information. Consolidated panel trip file replaces individual trip files in 3 different formats across time.	Mike Kruger
July 22, 2008	1.3	Trip files corrected to Jul08 versions, downloadable from website. Documentation updated to describe them	Mike Kruger
September 2, 2008	1.311	Updated ***Marketing Science*** citation with specific page numbers. Updated URL's.	Mike Kruger
October 21, 2008 – November 11, 2008	1.312, 1.313, 1.314	Added clarifications on panel type and chain assignment, Added description of Pacesetters, distributed with ADS_34 and higher, Added description of size attributes for carbonated soft drinks and beer	Mike Kruger
November 26, 2008 – February 5, 2009	1.4-1.404	Start to add changes for year 6. Note that some files were extended to year 7 (static file, chain xref) for convenience, but year 7 will not be released until 2010.	Mike Kruger
February 22, 2011	1.5	Updated with release of 2007 data. Note stub section. Note IRI has been renamed SymphonyIRI Group, but this has not been changed in this document.	Mike Kruger
	1.51	Minor examples added; in process	Mike Kruger
October 5, 2011	1.52	FAQ from Google Sites added as appendix	Mike Kruger

November 29, 2012	2.00	Update with 2008-2011 data	Mike Kruger
January-April 2013	2.01-2.03	Minor: Panel static information, clarification of SY, GE and VEND relationship in UPC code.	Mike Kruger
May 8, 2013	2.1	Documentation of issues with year1 Pepsi and Years 8-11 demographics. (Section 4.1, errata) Rebranding from SymphonyIRI back to IRI.	Mike Kruger

1.2 Acknowledgments and thanks

With this update of the data and the documentation in November, 2012, this data set will have 11 years of data, from January 1, 2001 through the end of December, 2011. We have added data from 2008 through 2011, although the basic principle of not releasing data for the most recent two years will continue. This exception is due to my attempt to complete this update prior to my retirement.

This data set was assembled by Lori DeHerrera working with Dan Pagni. It has been a great help having the same team do these updates. I've appreciated the support of Rob Holston, who now heads the SymphonyIRI analytics business.

We have kept the basic structure of the data the same, in order facilitate long term studies. In each section below, we have noted changes in the 2008-2011 data structure from the previous data structure in 2001-2005 and the 2006 and 2007 updates. This will be a bit clumsier for new users of the data, but probably clearer for those who are updating the data set.

Mike Kruger
EVP, R&D, Consumer and Shopper Marketing, SymphonyIRI Group
August 20, 2012

Earlier acknowledgment

Today, June 10, 2008, we received notice that ***Marketing Science*** has officially accepted the paper describing this data set.

There are a lot of people to thank here. The idea originated with Carl Mela (Duke), who spent years getting IRI interested at all, and who formally proposed this in August, 2005 to the IRI Analytics Advisory Board (AAB). Carl and Bart Bronnenberg (UCLA, Tilberg) did the high level database design. Carl got the cooperation of TNS for some advertising data.

The Analytics Advisory Board provided further advice, and the development of this data set was formally approved by Sunil Garga, president of IRI's Business and Consumer Insights Group, February 14, 2006 (Valentine's Day).

Dan Pagni organized assembly of the dataset and did the early version of this documentation paper. Much of the work on the dataset itself was done by Michael Schlemp (years 1,2,3) and Lori Mudrak DeHerrera (years 4 onward), with help from a variety of others.

During this period, Arvid Johnson headed the AAB and twice headed the Analytics Research and Development department. Without his support, leadership and nagging the data set would never have been released.

We are proud to be the first database paper in ***Marketing Science***. Eric Bradlow wrote in his editorial: "As ***Marketing Science*** publishes its first database paper .... our hope at ***Marketing Science*** is that the IRI marketing database paper … can have similar impact [to the Dominick's Finer Foods Data[2], which] has led to the empirical validation of many of our important theories, to the creation of new theories, and it continues to do so almost 20 years later. Furthermore, it has had impact beyond our own field as … used in economics and other related fields to answer important questions."[3]

Mike Kruger
EVP, Strategic Initiatives, IRI
June 10, 2008

[2] The Dominicks Finer Foods data can be found on the Kilts Center for Marketing web site, http://research.chicagogsb.edu/marketing/databases/dominicks/index.aspx

[3] Bradlow, Eric T. (2008). Editorial: Maximizing impact via database submissions. ***Marketing Science***, **27**(4), 541.

1.3 Citation and Update Information:

#1: Publications using this database should include the following reference:

> Bronnenberg, Bart J., Michael W. Kruger, Carl F. Mela. 2008. Database paper: The IRI marketing data set. ***Marketing Science***, **27**(4) 745-748.

If you need to cite this document containing database descriptions, the appropriate reference is:

> Kruger, Michael W. and Daniel Pagni, **IRI Academic Data Set Description**, version *x.x*, Chicago: Information Resources Incorporated, 2008.

This document will be updated as needed. The current version of this document can be found on the Google Sites support group for this data set, https://sites.google.com/site/irimarketingdataset/

IRI Marketing Data Set web page at IRI: http://us.infores.com/academic

#2: In addition, papers using this data set should include the following footnote:

We would like to thank IRI. for making the data available. All estimates and analysis in this paper, based on data provided by IRI. are by the authors and not by IRI.

Purpose

The purpose of this document is to describe the IRI academic data set in order to enable valid usage of this data set for academic research.

The usage conditions of this data are described in the legal agreement covering this data set[4]. ***It is important that this agreement be adhered to.*** These agreements are enclosed in the appendix.

A description of the purposes of the data set and a broad description of the data scope can be found in Bronnenberg, Kruger and Mela (2008)[5].

Exclusions to the scope of this document:

Daily data, and the description of daily data, is not included here.
Non-US data, and the description of non-US data, is not included here.
TNS advertising data is not described here.
Files which are not part of the academic data set (e.g. files which describe the conversion of actual chain names to chain aliases) are not included here.

2. General description of files

In this section we describe the files included, and provide descriptive information.

2.1 Overall organization

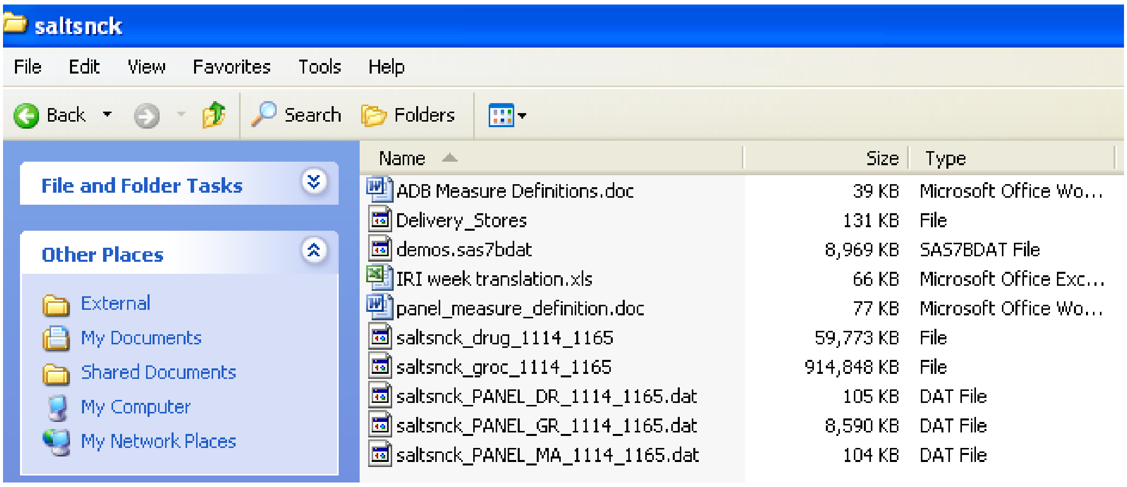
Originally, each year and each category is a separate folder, or separate DVD. This is to allow the researcher to combine the data into the form they need for analysis, without the need to subset very large files.

This distribution method assumed the academic users would want to get a few categories for a few years, so each DVD would be freestanding. However, it became clear academics preferred to get the entire dataset, and preparing and shipping 150 DVD's per academic user would have been clumsy. We shifted to sending out USB hard drives with the entire data set, but some vestiges of the original plan remain in the directory structure.

Example: The salty snacks year 1 DVD contained a directory called "saltsnck", which now is in the directory structure D:\Academic Dataset External\Year1\External\saltsnck
This directory contains the following files:

[4] See appendix 3.

[5] Bronnenberg, Bart J., Michael W. Kruger, Carl F. Mela. 2008. Database paper: The IRI marketing data set. ***Marketing Science***, **27**(4) 745-748.

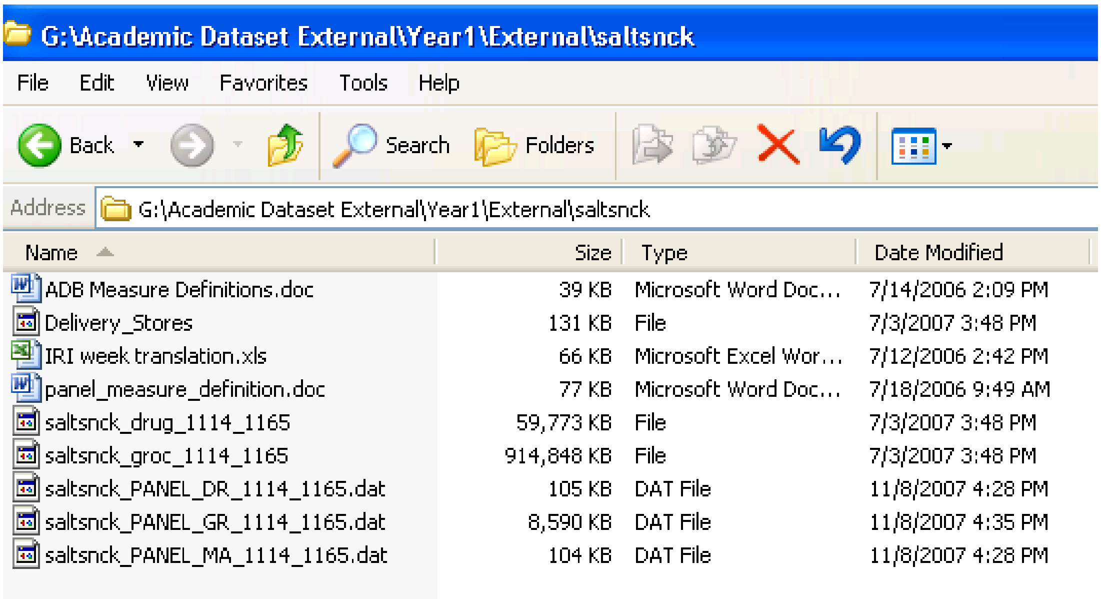


Name	Size	Type
ADB Measure Definitions.doc	39 KB	Microsoft Office Wo...
Delivery_Stores	131 KB	File
demos.sas7bdat	8,969 KB	SAS7BDAT File
IRI week translation.xls	66 KB	Microsoft Office Exc...
panel_measure_definition.doc	77 KB	Microsoft Office Wo...
saltsnck_drug_1114_1165	59,773 KB	File
saltsnck_groc_1114_1165	914,848 KB	File
saltsnck_PANEL_DR_1114_1165.dat	105 KB	DAT File
saltsnck_PANEL_GR_1114_1165.dat	8,590 KB	DAT File
saltsnck_PANEL_MA_1114_1165.dat	104 KB	DAT File

On an external hard drive, the directory structure is likely to be similar to this:

G:\Academic Dataset External\Year**1**\External**saltsnck**

Similarly, the salted snacks year 2 DVD contains a directory called “saltsnck” with these files:


G:\Academic Dataset External\Year1\External\saltsnck
File
Edit
View
Favorites
Tools
Help
Back
Search
Folders
Address
G:\Academic Dataset External\Year1\External\saltsnck

Name	Size	Type	Date Modified
ADB Measure Definitions.doc	39 KB	Microsoft Word Doc...	7/14/2006 2:09 PM
Delivery_Stores	131 KB	File	7/3/2007 3:48 PM
IRI week translation.xls	66 KB	Microsoft Excel Wor...	7/12/2006 2:42 PM
panel_measure_definition.doc	77 KB	Microsoft Word Doc...	7/18/2006 9:49 AM
saltsnck_drug_1114_1165	59,773 KB	File	7/3/2007 3:48 PM
saltsnck_groc_1114_1165	914,848 KB	File	7/3/2007 3:48 PM
saltsnck_PANEL_DR_1114_1165.dat	105 KB	DAT File	11/8/2007 4:28 PM
saltsnck_PANEL_GR_1114_1165.dat	8,590 KB	DAT File	11/8/2007 4:35 PM
saltsnck_PANEL_MA_1114_1165.dat	104 KB	DAT File	11/8/2007 4:28 PM

On an external hard drive, the year will be different:

G:\Academic Dataset External\Year**2**\External**saltsnck**

Brief descriptions of these files are in the following table.

Example of name	General name	Description

Example of name	**General name**	**Description**
ADB Measure Definitions.doc[6]	Same	Definitions for store measures
Delivery_Stores	Same	Information about the stores included in this year's files.
demos.csv Note: • the standardized files for years 2001-2007 are in the directory "demo trips external", • the demo.csv files in the individual category directories contain the same data as the standardized files; either can be used. • in year 6 and 7, these files are ONLY in the directory "demo trips external". • The demos were current as of the data of the update. So, for 2008-2011 the demographic files represent panelists active during that year, but represent the demographics which were most current as of August, 2012. They do NOT allow you to do a time path of demographic change from 2008-2011. (similar situation for 2001-2005, 2006-2007).	Same	Demographics for the panelists
IRI week translation.xls[7]	Same	IRI week numbers converted to standard calendar
panel_measure_definition.doc[8]	Same	Definitions for panel measures. Note 2008-2011 are slightly different as discussed below.
Saltsnck_drug_1166_1217 and Saltsnck_groc_1166_1217	*Category_outlet_startweek_endweek*	Store data file at store week upc level
Saltsnck_PANEL_DR_1166_1217.dat and Saltsnck_PANEL_GR_1166_1217.dat and	*Category*_PANEL_*outlet_startweek_endweek*.dat **Note**: for 2001-2007 the outlet codes	Panel data file at transaction level[9]

[6] In year 6, this was moved to the "Demos Trips External" folder, so there is one copy overall rather than one in each category. Since this file is the same for each year, and is unlikely to be part of any data scripts, this was done to save space on the year 6 update disks. In years 2008-2011 it is in both places.

[7] In year 6, this was moved to the "Demos Trips External" folder, so there is one copy overall rather than one in each category. In years 2008-2011 it is in both places.

[8] In year 6, this was moved to the "Demos Trips External" folder, so there is one copy overall rather than one in each category. In years 2008-2011 it is in both places.

[9] If there is no data for a particular outlet (e.g. beer is not sold in drug stores in this market), there will either be an empty file or no file at all.

Example of name	General name	Description
Saltsnck_PANEL_MA_1166_1217.dat	are DR (drug), GR (grocery) and MA (mass). For 2008-2011 the outlet codes are DK (drug), GK (grocery) and MK (mass).	
Saltsnck_prod_attr (obsolete in 2001-2007 because we replaced this with prod_category.xls) In 2008-2011 this contains expanded product information data.	*Category*_prod_attr	Product attributes for upcs in this category (replaced by prod_Category.xls)
Prod_saltsnck.xls (found in "parsed stub files" directory)[10]	Prod_*Category*.xls	Product attributes for upcs in this category (improved format) See section on "Product Attributes" for information about these files.

The store data files are the largest files.
Both the store data and panel data files are keyed to the dimensional information (store, week, UPC fields, [panelist]).

3. Detailed file descriptions

3.1 Store data sets: category_outlet_startweek_endweek

Naming convention: The naming convention for these is category name then outlet then start week and then end week, all separated by underscores, with no extension, so salted snacks drug data for the earliest year would be **saltsnck_drug_1114_1165**.

Files vary by: category, outlet, and time.
Records within a file represent a store / week / upc.

This file can be read in via a flat file or directly into excel (not, obviously, the entire file).

```
IRI_KEY WEEK SY GE VEND  ITEM  UNITS DOLLARS  F     D PR
 681530 1373  0  1 28400  4874     2    1.98 NONE  0 0
 681530 1373  0  1 28400  4853     7    6.93 NONE  0 0
 681530 1373  0  1 28400  4361    20   40.00 A     0 1
 681530 1373  0  1 28400  4852     1    0.99 NONE  0 0
 681530 1373  0  1 28400  4363     5   10.00 A     0 1
 681530 1373  0  1 28400  4854     3    2.97 NONE  0 0
 681530 1373  0  1 28400  4855     1    0.99 NONE  0 0
 681530 1373  0  1 28400  4365     8   16.00 A     0 1
```

[10] For years 1-6, the same product stub was used. By stub, we mean the hierarchical assignment of UPCs to brands to vendors to types to categories. This reflects the assignment was current as of early 2007. This stub information was updated for year 7. The year 7 stubs drop UPCs that have not moved in the past few years, add UPCs that were introduced in year 7 (2007), and update the hierarchical relationships that may have changed as of 2008 when this data was pulled (e.g. vendor and parent could change due to merger and acquisition activity, brand name could have changed from 2006 to 2007, etc. The years 8-11 have the same stub, but it is different than 2001-2006 and 2007.

```
681530 1373  0  1 28400  4861     3     2.97 NONE 0 0
```

The movement data is not sorted. Field description follows (and is also in the file **ADB measure definitions.doc**).

Header	**Definition**
IRI_KEY	Masked Store number, keyed to **delivery_stores** file.
WEEK	IRI Week: see **IRI Week Translation.xls** file for calendar week translation
SY	UPC - System
GE	UPC – Generation
VEND	UPC - Vendor
ITEM	UPC - Item
UNITS	Total Unit sales
DOLLARS	Total Dollar sales
F	Feature: see table below
D	Display: (0=NO, 1=MINOR, 2=MAJOR. MAJOR includes codes lobby and end-aisle)
PR	Price Reduction flag: (1 if TPR[11] is 5% or greater, 0 otherwise)

Possible Values for Feature (F)	**Definition**
NONE	No feature
FS-C	FSP C (for frequent shopper program members only)
C	C - small ad, usually 1 line of text
FS-B	FSP B
B	B – medium size ad
FS-A	FSP A
A	A – large size ad
FSA+	FSP A+
A+	A+ ad – also known as “Q” or “R” – retailer coupon or rebate

.

WEEK is the IRI week.
SY, GE, VEND, ITEM are the UPC code fields.

SY is the system code.
VEND is the vendor code.
ITEM is the item code.
The check digit is not supplied.

GE is the generation number of the UPC. All UPC’s begin with generation 1, but as product attributes change will have higher generation numbers applied. For example, a UPC that was used for a floor wax in 1984 (generation 1) may be used for a dessert topping in 2006 (generation 2).

DOLLARS reflects the retail price paid, on average. This includes retail features, displays, and retailer coupons. It does not include manufacturer coupons or any discount that might be applied

[11] TPR = temporary price reduction.

by the retailer that is not applicable to the item. For example, if a retailer gave $5 off if you purchased more than $200, that discount is not applied. Sales taxes are not included.

3.1.1 Loyalty program pricing

Loyalty programs (also called Frequent Shopper Card programs) are reflected in the following manner in the dollar data across time.

The following table summarizes the **Reported dollars** across time.

Store Type	Pre-January 2002	January 2002 and beyond
Participating FSP Store	LRD	MD
Non-Participating FSP Store	LRD	LRD
Non-FSP Store	LRD (=MD by definition)	MD (=LRD by definition)

Definitions:

A **Non-FSP Store** is a store that does not have a frequent shopper program.

An **FSP Store** is a store that does have a frequent shopper program.

A **Non-Participating FSP Store** is an FSP store that does not send IRI movement data that reflects frequent shopper discounts.

A **Participating FSP Store** is an FSP store that does send IRI movement data that reflects frequent shopper discounts.

Movement Dollars (MD) are the movement dollars sent to IRI by the retailer. If a non-FSP feature exists then this field is calculated as the MINIMUM(non-FSP feature price, movement price) X Unit sales. This calculation is commonly referred to by IRI as the Feature Price Override (FPO).

Lowest Reported Dollars (LRD) are calculated during the data load process as follows: MINIMUM(available feature prices, movement price) X Unit sales.

Reported Dollars are the most accurate (or best estimate of) dollar sales used in the calculation of all dollar-based measures and any other client deliverable application. This field is the result of the UPCSelect data extraction program.

3.2 Delivery Stores

Naming convention: "**Delivery Stores**".

Varies by: time. Only stores which are active in the particular year are included in that year's file. Does not vary by category.

This is a flat file with information about the stores. The first record is field names. This is a fixed column width file. This file can be read in as a flat file or directly into excel. The file contains each store "masked" using the sequence key as it's identifier across the various tables. This file also contains outlet, estimated acv, the market name so data can be aggregated by market, an open and close week, and finally a "chain" number representing a particular retailer. All the stores belonging to Chain8 are part of the same retailer that year. Note divisions of a large retailer are likely to have different chain numbers.[12]

[12] *Just a reminder: please note that we have masked the retailer identities in this file, and we have masked the private label information in the product definitions, in order to emphasize that the purpose of this data set is not retail consulting, and not a comparison of two retailers (e.g., Albertsons versus American Stores) on a named basis. Retailers are data suppliers to IRI (and Nielsen and others) IRI has contractual restrictions placed on it by retailers as a result of retailers providing this data to IRI, which is why this data is provided without retailer identification. It is not in the best interest of the marketing research industry, consumer packaged goods manufacturers, or academics seeking data to create a situation in which the retailers feel vulnerable for having supplied this data. While the letter of this is spelled out in the data contract signed as part of getting access, we also ask you to respect the spirit of this dataset as well.*

```
 1   IRI_KEY OU EST_ACV  Market_Name               Open Clsd MskdName
 2    200161 GR 11.16299 DETROIT                   1366 9998 Chain8
 3    200171 GR   23.631 MILWAUKEE                  522 9998 Chain87
 4    200197 GR 12.27599 PEORIA/SPRINGFLD.          903 9998 Chain51
 5    200272 GR   12.256 LOS ANGELES                873 9998 Chain113
 6    200287 GR 7.714996 SAN FRANCISCO              795 9998 Chain83
 7    200297 GR 21.76999 PORTLAND,OR                999 9998 Chain69
 8    200334 GR   18.963 PORTLAND,OR                922 1329 Chain125
 9    200341 GR 21.35199 SAN DIEGO                 1197 9998 Chain113
10    200372 GR 7.810997 HOUSTON                   1389 9998 Chain16
```

Open and closed weeks are from the point of view of IRI data, with a value of 9998 meaning the store is currently open and providing data. Record 8 indicates the store provided data to IRI from week 922 to week 1329. It cannot not be determined from this file whether this store closed, or stopped providing data to IRI.

The estimated ACV reflects an estimate of annualized sales in millions for the store (not the actual). $11.16299 reflects estimated sales in the store of $11,162,990 across all categories (including bakery, meat, produce, etc.) in grocery and non-prescription sales in drug.

The masked names are different in each year. So, what is chain13 in one year may be called chain12 in a second year. A cross-reference is provided in appendix 2. (UPDATE TBD)

3.2.1 Multiple records for same key

In some cases of merger and acquisition activity, there may be more than one record for a store in a year (in other words, more than one IRI_KEY with the same number).

IRI_KEY	OU	EST_ACV	Market_Name	Open	Clsd	MskdName
230501	GR	19.21899	BIRMINGHAM/MONTG.	807	1114	Chain28
230501	GR	10.55099	BIRMINGHAM/MONTG.	1120	9998	Chain20

The merger and acquisition patterns at retail can be complex and do not always occur neatly at the end of one year and the beginning of the next[13]. Note that there is a gap in the data. There is no data for weeks 1115 through 1119. The store may have been closed, may not have been providing data, or may not have passed QC at IRI due to the circumstances of the change.

3.2.2 Stores by market

A count of stores by market for year 5 is given below. There are 50 IRI markets included: 48 standard markets and 2 BehaviorScan markets with panel data.

Count of IRI_KEY	**Column Labels**

[13] This particular store’s activity was part of this larger set of events:

ChainA was purchased by ChainB. Some were closed, some kept the ChainA name, some took the ChainB name. However, the purchase and ChainB's takeover by an investment firm proved ill-advised, and 2 years later the combined chains went into bankruptcy. The next year some stores were sold off to ChainC; most of those locations have since closed in the wake of ChainC's own troubles. Some other locations were sold to ChainD and converted to their brands. At least two locations were later converted to ChainE.

Row Labels	DR	GR	Grand Total
ATLANTA	13	42	55
BIRMINGHAM/MONTG.	6	38	44
BOSTON	15	47	62
BUFFALO/ROCHESTER	11	21	32
CHARLOTTE	5	41	46
CHICAGO	39	54	93
CLEVELAND	6	17	23
DALLAS, TX	11	56	67
DES MOINES	2	8	10
DETROIT	20	32	52
EAU CLAIRE	2	7	9
GRAND RAPIDS	2	14	16
GREEN BAY	1	10	11
HARRISBURG/SCRANT	13	29	42
HARTFORD	7	35	42
HOUSTON	12	42	54
INDIANAPOLIS	7	22	29
KANSAS CITY	7	20	27
KNOXVILLE	2	21	23
LOS ANGELES	45	92	137
MILWAUKEE	6	24	30
MINNEAPOLIS/ST. PAUL	10	17	27
MISSISSIPPI	6	25	31
NEW ENGLAND	7	34	41
NEW ORLEANS, LA	8	31	39
NEW YORK	55	97	152
OKLAHOMA CITY	1	11	12
OMAHA	3	15	18
PEORIA/SPRINGFLD.	7	20	27
PHILADELPHIA	22	44	66
PHOENIX, AZ	13	45	58
PITTSFIELD	7	7	14
PORTLAND,OR	3	38	41
PROVIDENCE,RI	5	13	18
RALEIGH/DURHAM	8	45	53
RICHMOND/NORFOLK	7	35	42
ROANOKE	6	32	38
SACRAMENTO	5	32	37
SALT LAKE CITY	1	14	15
SAN DIEGO	15	30	45
SAN FRANCISCO	14	44	58
SEATTLE/TACOMA	6	47	53
SOUTH CAROLINA	15	76	91
SPOKANE	2	10	12
ST. LOUIS	6	27	33
SYRACUSE	5	25	30
TOLEDO	5	15	20

TULSA,OK	4	11	15
WASHINGTON, DC	22	60	82
WEST TEX/NEW MEX	5	16	21
Grand Total	**505**	**1588**	**2093**

Note the stores in the Pittsfield BehaviorScan market are also in the Hartford Infoscan market.

3.3 Panel data sets: Category_PANEL_outlet_startweek_endweek.dat

Panel data is provided for two BehaviorScan markets, Eau Claire, Wisconsin and Pittsfield, Massachusetts.

The naming convention for these is category name then "panel" then outlet then start week and then end week, all separated by underscores, with the extension DAT, so salted snacks drug data for the earliest year would be saltsnck**_PANEL_DR_1114_1165.**

2001-2007: This file can be read in via a flat file or directly into Excel (the entire file may not fit). The fields in this file are delimited by one or more spaces. It is not a fixed width file.

```
PANID      WEEK      UNITS     OUTLET     DOLLARS   IRI_KEY    COLUPC
 1197178       1175          2  DR              1   8003059              11600012250
 1197178       1175          6  DR              3   8003059              11600012250
 1227785       1174          1  DR           1.99   8000583              11600012530
 1137612       1200          1  DR           0.99    642166              11600012606
 1137612       1214          2  DR           1.98    642166              11600012606
 1401877       1166          2  DR           1.98   8003042              11600012606
 1401877       1175          1  DR           0.99   8003042              11600012606
 1401877       1182          1  DR           0.99   8003042              11600012606
 1401877       1183          2  DR           1.98   8003042              11600012606
```

2008-11: The file is a comma delimited file, and also includes the transaction minute, which is useful to matching the trip records.

Definitions of these fields are below (and in **panel_measure_definition.doc**) for 2001-2007

Measure	Definition	Calculation
PANID	panelist number within a market	
UNITS	Total number of units purchased by the Buying households.	The sum of total units purchased by the households buying the Product.

OUTLET	Channel to which the store/chain belongs
	MA=Mass
	GR=Grocery
	DR=drug	

DOLLARS	Total Paid dollars	This is drawn from the store data, not entered by the panelist, in cases where IRI has store data. In cases where IRI does not receive store data, some panelists do

		record price and this price is extended to other panelists.
IRI_KEY	Masked store number	
WEEK	IRI WEEK	
COLUPC	(Collapsed UPC). This is the UPC which matches the internal form (e.g. private label collapsed). The information in COLUPC is the same as in the combination of SY, GE, VEND, ITE.	This is the combination of a upc's system (2 digits), generation (1 digit), vendor (5 digits) and item (5 digits) fields. See product description section for an explanation of these fields. No leading zeroes are shown.

2008-11: The file is a comma delimited file, and also includes the transaction minute, which is useful to matching the trip records.

Measure	Definition	Calculation
PANID	panelist number within a market	
WEEK	IRI WEEK	
MINUTE	Minute of the week the transaction occurred (or, for key panelist, was scanned). Note for some key panelists, the equipment can separate trips, but does NOT provide a true time stamp; these trips are moved to the middle of the night. These types of key panelists should not be in these two markets.	
UNITS	Total number of units purchased by the Buying households.	The sum of total units purchased by the households buying the Product.

OUTLET	Channel to which the store/chain belongs
	MA=Mass
	GR=Grocery
	DR=drug	

DOLLARS	Total Paid dollars	This is drawn from the store data, not entered by the panelist, in cases where IRI has store data. In cases where IRI does not receive store data, some panelists do record price and this price is extended to other panelists.
IRI_KEY	Masked store number	
COLUPC	(Collapsed UPC). This is the UPC which matches the internal form (e.g. private label collapsed). The information in COLUPC is the same as in the combination of SY, GE, VEND, ITE.	This is the combination of a upc's system (2 digits), generation (1 digit), vendor (5 digits) and item (5 digits) fields. See product description section for an explanation of these fields. No leading zeroes are shown.

3.4 Panel trips

These files represent the trips made by panelists who purchased at least one item.

These files have been standardized in format from the way they were originally constructed, and placed in the directory “parsed stub files”. The naming convention is **trips*N* jul08.csv**, where ***N*** is the year[14]. Fields are listed below. These files contain the following fields:

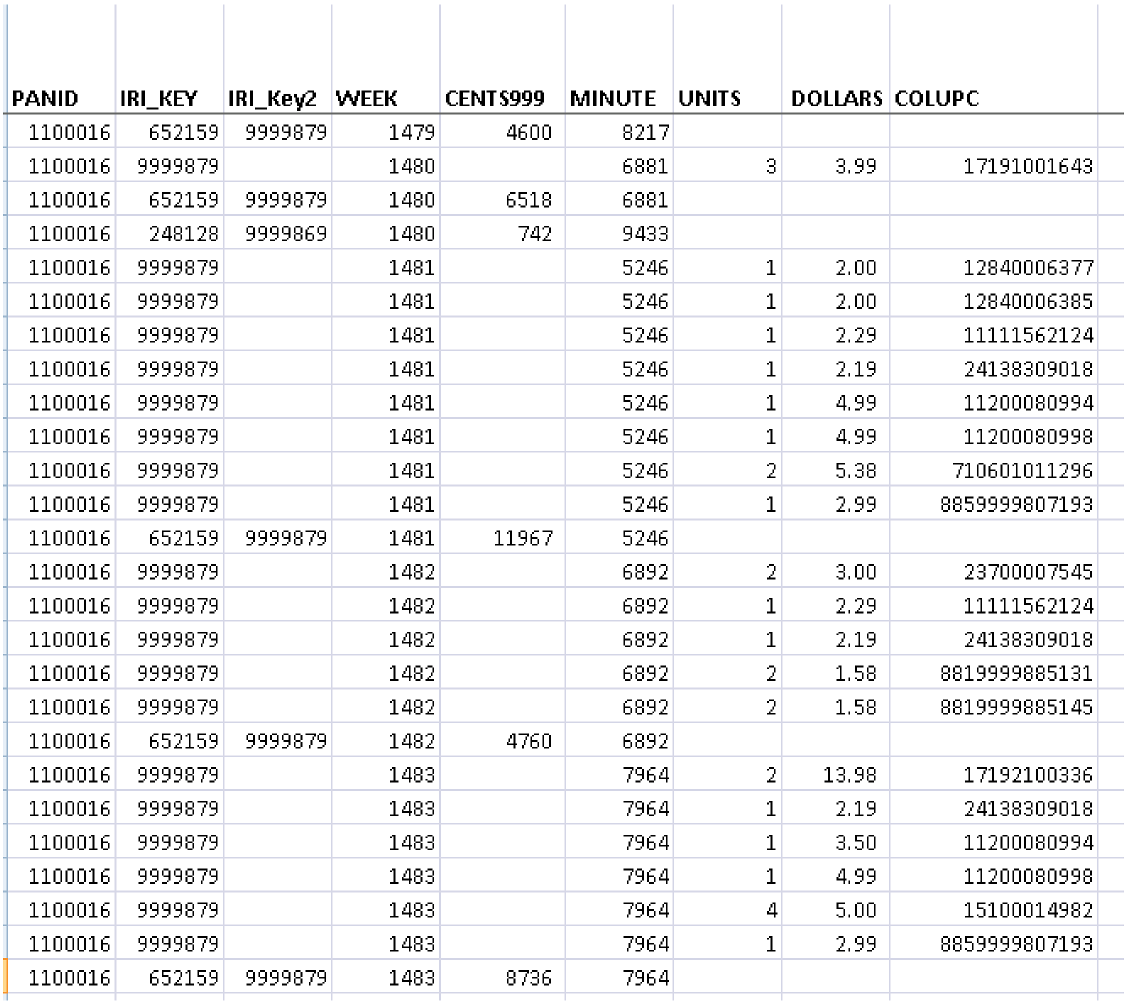
PANID (Panelist ID)	Panelist ID number
Week	IRI defined week; for an explanation of these codes see the section on “IRI week translation”.
IRI_KEY	Store
MINUTE	Minute within the week the transaction occurred (or the scankey was used to record the purchase). 0 - 1439 is Monday, 1440 - 2879 is Tuesday and so on. For example: *1 is 1201 am Monday* *8 is 1208 am Monday* *1438 is 1158 pm Monday* *1441 is 1201 am Tuesday* *And so on.*
KRYSCENTS	Generally the same as CENTS999, it’s scrubbed a bit and is probably a better field.
CENTS998	The cents on the overall register tape, as entered by the panelist. This is missing for the card panelists because they do not enter their total register tape. (Processing procedure changed and this is present in 2008-2011).
CENTS999	The trip total obtained by adding up the individual scanned items. For key panelists, this total will generally be <= the 998 record because of items that were not scanned (non CPG). For card panelists, the 999 record will be similar to the 998 record because the card records all purchases, even ones (such as random weight) that are not used in the panel. For research purposes, the “trip total” might be best considered to be the 999 record for card panelists, and the 998 record for key panelists.

2008-2011: The trip files which are in the product directories have the format above BUT they contain only the trips with a product purchase. So the file in the beer directory only has trips that had a purchase of beer.

[14] Trip files produced earlier than July 2008 have errors and should not be used.

The overall trip files (trip8.csv, trip9.csv, trip10.csv, trip11.csv) contain all trips by the panelists and are likely more useful to researchers. *Note because of different scan equipment the same store may be indicated in two different ways in some cases, and so on the overall trip records there are TWO values for IRI_Key, the store number. The actual transactions for a panelist will match ONE of these.*

Example: We have taken the trips for one panelist and merged them with the transaction records. (This is NOT a file you have in the data set). Store 652159 and store 9999879 are the same store; Both keys are on the trip record (the rows without a UPC). Only one key is on the transaction record (the rows with a UPC).



PANID	IRI_KEY	IRI_Key2	WEEK	CENTS999	MINUTE	UNITS	DOLLARS	COLUPC
1100016	652159	9999879	1479	4600	8217			
1100016	9999879		1480		6881	3	3.99	17191001643
1100016	652159	9999879	1480	6518	6881			
1100016	248128	9999869	1480	742	9433			
1100016	9999879		1481		5246	1	2.00	12840006377
1100016	9999879		1481		5246	1	2.00	12840006385
1100016	9999879		1481		5246	1	2.29	11111562124
1100016	9999879		1481		5246	1	2.19	24138309018
1100016	9999879		1481		5246	1	4.99	11200080994
1100016	9999879		1481		5246	1	4.99	11200080998
1100016	9999879		1481		5246	2	5.38	710601011296
1100016	9999879		1481		5246	1	2.99	8859999807193
1100016	652159	9999879	1481	11967	5246			
1100016	9999879		1482		6892	2	3.00	23700007545
1100016	9999879		1482		6892	1	2.29	11111562124
1100016	9999879		1482		6892	1	2.19	24138309018
1100016	9999879		1482		6892	2	1.58	8819999885131
1100016	9999879		1482		6892	2	1.58	8819999885145
1100016	652159	9999879	1482	4760	6892			
1100016	9999879		1483		7964	2	13.98	17192100336
1100016	9999879		1483		7964	1	2.19	24138309018
1100016	9999879		1483		7964	1	3.50	11200080994
1100016	9999879		1483		7964	1	4.99	11200080998
1100016	9999879		1483		7964	4	5.00	15100014982
1100016	9999879		1483		7964	1	2.99	8859999807193
1100016	652159	9999879	1483	8736	7964			

Note also that the trip in the first row (week 1479, minute 8217) did not have any items in these categories purchased.

3.5 Panel static file (static1_*n*.csv)

This file lists panelists who made the standard IRI static during the year (satisfied minimal requirements for reporting). This evaluation is done for each panelist each year. It was

IRI’s intention to include only the trip and purchase data for panelists who made static. However, due to trip processing problems there are both trips and transactions for panelists who did not made static, particularly for years 1 and 2.

The file **static1_5.csv** provides static information for years 1 through 5. The file **static1_7.csv** provides static information for years 1 through 7. The file **static1_11.csv** provides static information for years 1 through 11.The information for years 1 through 5 is the same in all files.

PANID	Panelist ID number
Trip Count	Number of trips made by this panelist in this year
Make_static	This is always “yes”. Panelist/years which do not satisfy static requirements are omitted from this file.
Year	Data year. All years are included in a single file.

The static is a standard 1 of 4 static. There are 13 four week periods in a 52 week period. The respondent must make at least one transaction in each of these to make static.

Note that for years 3-7 panelists who did not make static were generally excluded from the data. In years 1,2,8-11 they are generally included. In order to get consistent results, you should always use the static file.

3.5.1 Managing transactions, trips and static

In merging the data to produce a valid analysis, the following tips may be helpful.

First of all, determine how long the analysis is to run, and whether you need the same households to report for all of that time. Thus, if you need to have a 3 year analysis and households have to be good reporters for the entire time, you should use households who make the static in 3 successive years (years 1,2,3 or years 2,3,4, for example). Alternatively, you may want to run a long-term analysis, but NOT require the same person to be in the sample the entire time. If you want to filter by particular characteristics of panelists (e.g. card versus key, dog owners) this would be a good spot to merge in the demographic information.

Files needed:

static1_5.csv or **static1_7.csv or static1_11.csv**
ads demo*N*.csv

Second, filter out the data for only those households who make the static in the required time period. The should be the transaction data for the category (e.g. the purchases of cereal) for the outlets needed (of grocery, drug, mass). If trips are needed, filter the trips data files as well.

Files needed:

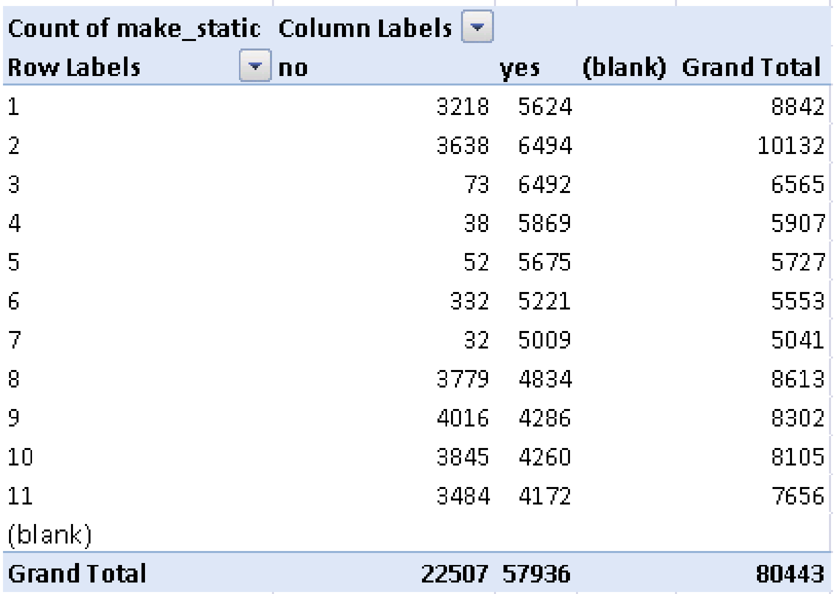
Category_PANEL_outlet_startweek_endweek.dat
trips*N* jul08.csv

Third, if you need to match purchases to trips you must do so at the panelist – week – store level. The trips are coded by minute within the week, but the transaction data in this data set is not for years 1-7.

3.5.2 Some numbers

It may be helpful to provide some numbers here.

How many static panelists?



Count of make_static	Column Labels			
Row Labels	no	yes	(blank)	Grand Total
1	3218	5624		8842
2	3638	6494		10132
3	73	6492		6565
4	38	5869		5907
5	52	5675		5727
6	332	5221		5553
7	32	5009		5041
8	3779	4834		8613
9	4016	4286		8302
10	3845	4260		8105
11	3484	4172		7656
(blank)				
Grand Total	**22507**	**57936**		**80443**

This doesn't reflect change in the panel philosophy; it reflects the fact that the trip and transaction files in year 1,2,8-11 were not adjusted for static and include panelist churn / poor reporters.

How many transactions?

Let's look at carbonated soft drinks and cold cereal together. We get the following transaction counts by year across all 3 outlets. In other words, we have combined 6 files each year in the tabulations below. This is before applying any static.

Year	Count of transactions
1	384363
2	490947
3	399816
4	336652
5	325478

How many transactions, among static panelists?

Year	Count of transactions among static panelists
1	309018
2	405267
3	397666
4	333360
5	322116

How many transactions can be matched to trips?

Note when we are matching to trips, we are first summing trips by unique panelist_store_week combinations. This may include more than one trip in a week.

The number of transactions which are not matched to a trip total record in this manner is small (0.07%). This is improved in year 8-11.

Year	Can be matched	Cannot be matched
1	308580	438
2	404693	574
3	397662	4
4	333086	274
5	322029	87

How many panelists are in how many years?

Panelists keep the same panel ID across years. The static file can enable you to see how many panelists will qualify across more than one year. A count of how many panelists are in how many years is below. There are 1348 panelists who are in all 11 years. There are 873 who are in 3 years. These years may not be consecutive.

Years making static	Number of panelists
1	1527
2	1216
3	873
4	726
5	684
6	635
7	628
8	703
9	794
10	923
11	1348
Grand Total	10057

3.6 Panel stores

Most panel transactions can be referenced using the store information as described in the section on “Delivery Stores”.

There are some stores which do not have store data included. An example would be Wal-mart, a small independent drug store, or other case in which the store data is not available. There are other cases in which the panel scankey includes a general retailer (e.g. “CVS”) but we cannot be specific as to the specific store. These instances are referenced in the file **manual store entry external.csv**. This file applies to earlier years. The file **manual store entry external 8_11.csv** applied to years 8-11.

Field name	Description
IRI_KEY	Store number
Outlet	GR grocery DR drug MA mass NA or n/a other outlets
Year1Chain	Chain number in year1. See discussion of **masked_chain_xref.csv** for explanation
Year2Chain	Chain number in year2
Year3Chain	Chain number in year3
Year4Chain	Chain number in year4
Year5Chain	Chain number in year5

3.7 Panel demographics

Panel demographic files have been standardized and are called **ads demo*N*.csv**, where ***N*** is the year number: ads demo1.csv, ads demo2.csv … ads demo5.csv.

The panelists included are those who satisfied IRI’s standard 52 week reporting static. This means that (1) the panelists included reported all year, and (2) the panelists are different between years.

For the initial set of data provided, the panelist demos reflect data current at that time. So, for the year 1, 2, and 3 (2001-2003) data, the panelist demos are from early 2007, not 2001. For this reason, there may be panelist records without demographics. For years 4 and 5 (2004-2005) the panelist demos are from later in 2007 and may be slightly different due to the demographic updates. Similarly for years 8-11: the demos reflect information pulled in summer, 2012.

The field names and the first two panelist values are shown below. Due to the demographic updates, there are minor differences in the values for the two panelists. For example, the male head in household in 1100180 is now listed as “some college” rather than post-secondary “technical school”, and the male head occupation from laborer to machine operator.

In these files, a missing value may appear as an empty field, a blank, a period, or a zero.

Panelist ID	**…032**	**…180**
Panelist Type	0	6
Combined Pre-Tax Income of HH	5	11
Family Size	2	2
HH_RACE	1	1
Type of Residential Possession	2	2
COUNTY	C	C
HH_AGE	5	5
HH_EDU	7	5
HH_OCC	6	1
Age Group Applied to Male HH	7	5
Education Level Reached by Male HH	9	5
Occupation Code of Male HH	11	7
Male Working Hour Code	7	3
MALE_SMOKE	.	1
Age Group Applied to Female HH	5	5
Education Level Reached by Female HH	7	5
Occupation Code of Female HH	6	1
Female Working Hour Code	3	3
FEM_SMOKE	0	0
Number of Dogs	0	1
Number of Cats	2	1
Children Group Code	3	8
Marital Status	1	2
Language	.	.
Number of TVs Used by HH	1	2
Number of TVs Hooked to Cable	1	2
HISP_FLAG	0	0
HISP_CAT	.	.
HH Head Race (RACE2)	1	1
HH Head Race (RACE3)	1	1
Microwave Owned by HH	1	1
ZIPCODE	1201	1201
FIPSCODE	25003	25003
market based upon zipcode	1	1
IRI Geography Number	1	1
EXT_FACT	1	1

Panelist ID	**…032**	**…180**
Panelist Type	0	6
Combined Pre-Tax Income of HH	6	11
Family Size	2	2
HH_RACE	1	1
Type of Residential Possession	2	2
COUNTY	C	C
HH_AGE	5	5
HH_EDU	7	5
HH_OCC	6	1
Age Group Applied to Male HH	7	5
Education Level Reached by Male HH	9	6
Occupation Code of Male HH	11	6
Male Working Hour Code	7	3
MALE_SMOKE	.	1
Age Group Applied to Female HH	5	5
Education Level Reached by Female HH	7	5
Occupation Code of Female HH	6	1
Female Working Hour Code	3	3
FEM_SMOKE	0	0
Number of Dogs	0	1
Number of Cats	1	1
Children Group Code	3	8
Marital Status	1	2
Language	.	.
Number of TVs Used by HH	2	2
Number of TVs Hooked to Cable	1	2
HISP_FLAG	0	0
HISP_CAT	.	.
HH Head Race (RACE2)	1	1
HH Head Race (RACE3)	1	1
Microwave Owned by HH	1	1
ZIPCODE	1201	1201
FIPSCODE	25003	25003
market based upon zipcode	1	1
IRI Geography Number	1	1
EXT_FACT	1	1

Field definitions are shown below and are in **panel_measure_definition.doc**. “Plan to drop” fields should not be used. Zipcode does not have leading zero (01201 is shown as 1201)

Panel demographics definitions follow (and are also in the file “panel_measure_definition.doc”)

Panelist type: Panelist type determines the data scope for a panelist. A card panelist shows a card, similar to a loyalty card, at participating retailers. Not all retailers participate (notably, Wal-mart does not). A key panelist has a key and wands their purchases at all retailers, but due to the heavier burden has a lower compliance rate. A card+key panelist uses a card in participating retailers, and a key to wand their purchases at non-particpating retailers[15]. A “card switch from key” is a panelist who was recruited as a key panelist, but is now a card panelist (possibly because they found the key too burdensome). A count of the panelists by type from the demos.csv file is listed below:

Panelist Type	Count of Panelist Type
0	2793
5	675
6	2439
9	4
Grand Total	**5911**

Measure	**Definition**
Panelist ID	panelist number within a market
Panelist Type	0=Card Only 5= Card + key 6= Card switch from key 7 = Key only 8= Canceled panelist (only found in a few year 1 and year 2 records; ignore panelist) 9 = Key switch from card
Combined Pre-Tax Income of HH	combined pre-tax income of the heads of household 0 = 'N/A'; 1 = '$00,000 to $ 9,999 per yr' 2 = '$10,000 to $11,999 per yr' 3 = '$12,000 to $14,999 per yr' 4 = '$15,000 to $19,999 per yr' 5 = '$20,000 to $24,999 per yr' 6 = '$25,000 to $34,999 per yr' 7 = '$35,000 to $44,999 per yr' 8 = '$45,000 to $54,999 per yr' 9 = '$55,000 to $64,999 per yr' 10 = '$65,000 to $74,999 per yr' 11 = '$75,000 to $99,999 per yr'

[15] On the trip data, a key transaction will have both CENTS98 and CENTS99 values. A card transaction will only have CENTS99.

	12 = '$100,000 and greater per year'
Family Size	family size 0 = 'N/A' 1 = 'One person' 2 = 'Two people' 3 = 'Three people' 4 = 'Four people' 5 = 'Five people' 6 = 'Six or more people'
HH_RACE	3 = 'Hispanic' Everything else=’non Hispanic’
Type of Residential Possession	The type of residential possession 0 = 'N/A' 1 = 'Renter' 2 = 'Owner'
COUNTY	County sizes
HH_AGE	0 = ‘N/A’ 1 = '18 - 24' 2 = '25 - 34' 3 = '35 - 44' 4 = '45 - 54' 5 = '55 - 64' 6 = '65 + ' 7 = 'No such person'
HH_EDU	0 = ‘N/A’ 1 = 'Some grade school or less' 2 = 'Completed grade school' 3 = 'Some high school' 4 = 'Graduated high school' 5 = 'Technical school' 6 = 'Some college' 7 = 'Graduated from college' 8 = 'Post graduate work' 9 = 'No such head of household'
HH_OCC	0 = 'Other' 1 = 'Professional or technical' 2 = 'Manager or administrator' 3 = 'Sales' 4 = 'Clerical' 5 = 'Craftsman' 6 = 'Operative (machine operator)' 7 = 'Laborer' 8 = 'Cleaning, food, health service worker' 9 = 'Private household worker'

	10 = 'Retired' 11 = 'No such head of household' 13 = 'Not employed'
Age Group Applied to Male HH	age group applied to the male head of household
Education Level Reached by Male HH	the education level reached by the male head of household
Occupation Code of Male HH	the occupation code of the male head of household
Male Working Hour Code	male work hours 1 = 'Not employed' 2 = 'Part time, < 35 hrs./wk.' 3 = 'Full time, > 35 hrs./wk.' 4 = 'Retired' 5 = 'Homemaker' 6 = 'Student' 7 = 'N/A'
MALE_SMOKE	Plan to drop
Age Group Applied to Female HH	age group applied to the female head of HH
Education Level Reached by Female HH	the education level reached by the female head of household
Occupation Code of Female HH	the occupation code of the female head of household
Female Working Hour Code	female work hours 1 = 'Not employed' 2 = 'Part time, < 35 hrs./wk.' 3 = 'Full time, > 35 hrs./wk.' 4 = 'Retired' 5 = 'Homemaker' 6 = 'Student' 7 = 'N/A'
FEM_SMOKE	
Number of Dogs	number of dogs 0 = 'None' 1 = 'One' 2 = 'Two' 3 = 'Three' 4 = 'Four' 5 = 'Five +'
Number of Cats	number of cats 0 = 'None' 1 = 'One' 2 = 'Two' 3 = 'Three' 4 = 'Four' 5 = 'Five +'
Children Group Code	children group 0 = 'N/A'

	1 = 'Child in [0-5)' 2 = 'Child in [6-11)' 3 = 'Child in [12-17)' 4 = 'Children in [0-5) & [6-11)' 5 = 'Children in [0-5) & [12-17)' 6 = 'Children in [6-11) & [1217)' 7 = 'Children in [0-5),[6-11) & [12-17)' 8 = 'Family size>0 yet no children'
Marital Status	marital status code 0 = 'N/A' 1 = 'Single' 2 = 'Married' 3 = 'Divorced' 4 = 'Widowed' 5 = 'Separated'
Language	
Number of TVs Used by HH	Actual number
Number of TVs Hooked to Cable	Actual number
HISP_FLAG	N/A- planning to drop
HISP_CAT	N/A- planning to drop
HH Head Race (RACE2)	N/A- planning to drop
HH Head Race (RACE3)	Ethnicity 0 = 'N/A' 1 = 'White' 2 = 'Black-African American' 3 = 'Hispanic' 4 = 'Asian' 5 = 'Other' 6 = 'American Indian-Alaska Native' 7 = 'Native Hawaiian-Pacific Islands'
Microwave Owned by HH	N/A- planning to drop
ZIPCODE	As is
FIPSCODE	As is
market based upon zip code	Plan to drop
IRI Geography Number	1=Pittsfield 3=eau Claire 7=grand junction 10=cedar rapids-Iowa
EXT_FACT	equal market/demo weight (multi-outlet (4M) weights) [This has no meaning for this data set.]

The field values for 2008-2011 may be slightly different and some fields are in slightly different columns. Please check.

Measure 2008-2011	**Definition 2008-2011**

Panelist ID	panelist number within a market
Panelist Type	Plan to drop
Combined Pre-Tax Income of HH	combined pre-tax income of the heads of household 0 = 'N/A'; 1 = '$00,000 to $ 9,999 per yr' 2 = '$10,000 to $11,999 per yr' 3 = '$12,000 to $14,999 per yr' 4 = '$15,000 to $19,999 per yr' 5 = '$20,000 to $24,999 per yr' 6 = '$25,000 to $34,999 per yr' 7 = '$35,000 to $44,999 per yr' 8 = '$45,000 to $54,999 per yr' 9 = '$55,000 to $64,999 per yr' 10 = '$65,000 to $74,999 per yr' 11 = '$75,000 to $99,999 per yr' 12 = '$100,000 and greater per year'
Family Size	family size 0 = 'N/A' 1 = 'One person' 2 = 'Two people' 3 = 'Three people' 4 = 'Four people' 5 = 'Five people' 6 = 'Six or more people'
HH_RACE	3 = 'Hispanic' Everything else='non Hispanic'
Type of Residential Possession	The type of residential possession 0 = 'N/A' 1 = 'Renter' 2 = 'Owner'
COUNTY	County sizes
HH_AGE	0 = 'N/A' 1 = '18 - 24' 2 = '25 - 34' 3 = '35 - 44' 4 = '45 - 54' 5 = '55 - 64' 6 = '65 + ' 7 = 'No such person'
HH_EDU	0 = 'N/A' 1 = 'Some grade school or less' 2 = 'Completed grade school' 3 = 'Some high school' 4 = 'Graduated high school' 5 = 'Technical school' 6 = 'Some college'

	7 = 'Graduated from college' 8 = 'Post graduate work' 9 = 'No such head of household'
HH_OCC	0 = 'Other' 1 = 'Professional or technical' 2 = 'Manager or administrator' 3 = 'Sales' 4 = 'Clerical' 5 = 'Craftsman' 6 = 'Operative (machine operator)' 7 = 'Laborer' 8 = 'Cleaning, food, health service worker' 9 = 'Private household worker' 10 = 'Retired' 11 = 'No such head of household' 13 = 'Not employed'
Age Group Applied to Male HH	age group applied to the male head of household
Education Level Reached by Male HH	the education level reached by the male head of household
Occupation Code of Male HH	the occupation code of the male head of household
Male Working Hour Code	male work hours 1 = 'Not employed' 2 = 'Part time, < 35 hrs./wk.' 3 = 'Full time, > 35 hrs./wk.' 4 = 'Retired' 5 = 'Homemaker' 6 = 'Student' 7 = 'N/A'
MALE_SMOKE	Plan to drop
Age Group Applied to Female HH	age group applied to the female head of HH
Education Level Reached by Female HH	the education level reached by the female head of household
Occupation Code of Female HH	the occupation code of the female head of household
Female Working Hour Code	female work hours 1 = 'Not employed' 2 = 'Part time, < 35 hrs./wk.' 3 = 'Full time, > 35 hrs./wk.' 4 = 'Retired' 5 = 'Homemaker' 6 = 'Student' 7 = 'N/A'
FEM_SMOKE	
Number of Dogs	number of dogs 0 = 'None'

	1 = 'One' 2 = 'Two' 3 = 'Three' 4 = 'Four' 5 = 'Five +'
Number of Cats	number of cats 0 = 'None' 1 = 'One' 2 = 'Two' 3 = 'Three' 4 = 'Four' 5 = 'Five +'
Children Group Code	children group 0 = 'N/A' 1 = 'Child in [0-5)' 2 = 'Child in [6-11)' 3 = 'Child in [12-17)' 4 = 'Children in [0-5) & [6-11)' 5 = 'Children in [0-5) & [12-17)' 6 = 'Children in [6-11) & [1217)' 7 = 'Children in [0-5),[6-11) & [12-17)' 8 = 'Family size>0 yet no children'
Marital Status	marital status code 0 = 'N/A' 1 = 'Single' 2 = 'Married' 3 = 'Divorced' 4 = 'Widowed' 5 = 'Separated'
Language	
Number of T Vs Used by HH	Actual number
Number of TVs Hooked to Cable	Actual number
HISP_FLAG	N/A- planning to drop
HISP_CAT	N/A- planning to drop
HH Head Race (RACE2)	N/A- planning to drop
HH Head Race (RACE3)	Ethnicity 0 = 'N/A' 1 = 'White' 2 = 'Black-African American' 3 = 'Hispanic' 4 = 'Asian' 5 = 'Other' 6 = 'American Indian-Alaska Native' 7 = 'Native Hawaiian-Pacific Islands'

Microwave Owned by HH	N/A- planning to drop
ZIPCODE	As is
FIPSCODE	As is
market based upon zip code	Plan to drop
IRI Geography Number	1=Pittsfield 3=eau Claire 7=grand junction 10=cedar rapids-Iowa
EXT_FACT	equal market/demo weight (multi-outlet (4M) weights)

3.8 Product attributes

The improved file format, which incorporates further information, is **prod_*category*.xls**, for example, **prod_saltsnck.xls**[16]**.**

There are three sets of files.

The first set of files are applicable to years 1-6 and are provided in a directory called “parsed stub files”.

The second set of files are applicable to year 7 and are provided in a directory called “parsed stub files 2007”.

The third set of files are applicable to years 8-11 are are provided in a directory called “parsed stub files 2008-2011”

Why are there multiple sets of stub files?

(a) Roughly every year, IRI reworks the Infoscan Review product stubs which are the basis for the product definitions used in this database. When that happens, we restate the Infoscan Reviews and make them available to subscribing clients. Through careful timing, we were able to get 2001 through 2006 all on the same stub. So, as you have undoubtedly noticed and appreciated, each UPC maps upward to the same brand, same vendor, same parent, same type and same category for the entire 2001-2006 time period.

(b) It's not a static world. The 2001-2006 stub stopped being maintained in early 2007, which means no new items were added. So, the 2007 data were pulled with a newer stub. What does this mean?

(b1) It means UPCs may map up the hierarchy differently -- most notably, to a different parent corporation.

(b2) Is means private label UPCs are mapped to different system 88's.

[16] The original file format used the following naming convention: category name then an underscore, then “prod_attr” with no extension. For example, the salty snack product attributes are in **saltsnck_prod_attr.** These files should not be on the dataset, and should be ignored if they are.

Depending on your research, it may be trivial to link 2001-2006 to 2007, or it may be very difficult.

Simularly, the 2008-2011 data uses different product stubs.

File structure

This is an excel file. The first line of the file contains the attribute labels. The specific product attributes 1-7 included will vary by category. For some categories, less than 7 additional attributes are provided.

Column name	Description
L1	Level 1 value (Large category)
L2	Level 2 value (Small category)
L3	Level 3 value (Parent Company)
L4	Level 4 value (Vendor)
L5	Level 5 value (Brand)
L9	level 9 value (UPC description)
Level	This field is always 9 because these are UPCs
UPC	UPC number (2 digit system,2 digit generation[17], 5 digit vendor, 5 digit item, separated by dashes)
SY	UPC system code; note that private label UPCs are collapsed across retailers and have a system code of 88. See "Delivery Stores" section for more information.
GE	UPC generation code. This is IRI's version number for the UPC; not a formal part of the UPC. All UPC's begin with generation 1, but as product attributes change will have higher generation numbers applied. For example, a UPC that was a floor wax in 1984 (generation 1) may be a dessert topping in 2006 (generation 2).
VEND	UPC vendor code -- 5 digits
ITEM	UPC item code -- 5 digits
*STUBSPEC 1527RC 00004 (name varies by category)	UPC recipe description.
VOL_EQ	Volume equivalent
PRODUCT TYPE (name varies by category)	Attribute 1 for this category
SUGAR CONTENT (name varies by category)	Attribute 2 for this category
PROCESS (name varies by category)	Attribute 3 for this category
TEXTURE (name varies by category)	Attribute 4 for this category
FORM (name varies by category)	Attribute 5 for this category
TYPE OF COMBINATION (name varies by category)	Attribute 6 for this category (if provided)
STYLE (name varies by category)	Attribute 7 for this category (if provided)

The check digit is not supplied.

[17] Note this is a different format than the COLUPC in the panel data (dashes, and use of a 2 digit generation code, and explicit "00" for system 0).

GE is the generation number of the UPC. All UPC's begin with generation 1, but as product attributes change will have higher generation numbers applied. For example, a UPC that was a floor wax in 1984 (generation 1) may be a dessert topping in 2006 (generation 2).

The STUBSPEC column provides an economical description of the item. Each category has a volume equivalent (VOL_EQ) that provides a way to compare units of different sizes[18]. Let's look at the first few items from the carbonated soft drink stub:

```
*STUBSPEC 1428RC                                         00004                    VOL_EQ
    +ATRET BCHR SODA REG CAN    72OZ 7 1 29228    11 1  1 0.3750RP   00017          0.375
    +ATRET BCHR SODA REG NPB    20OZ 7 1 29228   181 1  1 0.1042RP   00018         0.1042
    +ATRET BCHR SODA REG      67.6OZ 7 1 29228   150 1  1 0.3521RP   00019         0.3521
    +ATRET BCHR SODA REG CAFFR  12OZ27 1     7   678 1  1 0.0625RP   00020         0.0625
```

The STUBSPEC uses a recipe that varies. In this case,

A-TREAT BOTTLING CO.
Black Cherry Soda
Regular (not diet)
Can
72 ounces in the total package (from the size attribute file we can tell this is 6 12-ounce cans)
7-1-29228-11 is the UPC code in system-generation-vendor-item format.
The next fields "1 1" are for internal use
The next field indicates this is 0.3750 volume equivalents (see next paragraph) of regular pack product.
The last field (00017, 00018) is for internal use.

The volume equivalent (VOL_EQ) used in carbonated soft drinks is the 192 ounce case, which was originally 24 8-ounce bottles.

The first item is 72 ounces, and the VOL_EQ is 72/192 = 0.375
The second item is 20 ounces, and the VOL_EQ is 20/192 = 0.1042
The third item is 67.6 ounces (2 liters), and the VOL_EQ is 67.6/192 = 0.3521
The fourth item is 12 ounces, and the VOL_EQ is 12/192 = 0.0625

3.8.1 *Additional size attributes*

For beer and carbonated soft drinks, additional size attribute information is provided. These files are **prod_beer_sz.xls, prod_yogurt_sz2001-2006.xls, prod_mustketc.xls** and **prod_carbbev_sz.xls**[19]**.**

The text fields (VALTEXTn) are extremely repetitive. These are included because these attributes are defined by IRI's internal categories (keycats) and vary by these keycats. For example, "total ounces" is a good volume equivalent for beer, but "tablets" is a more likely one to use for aspirin. The user should check that all these VALTEXTn values are the same. They are all included for beer (4 size fields, 4 label fields), but for carbonated beverages there are 5 size fields, so only 3 label fields could be accommodated.

Column name	**Description**

[18] Trickiest is the photo supplies category.Tthe volume equivalent in the data should reflect the number of rolls. The volume equivalent field in the parsed stub files is a compound field: the first number is the number of rolls and the second one is the total number of exposures (see also the "exposure" field). There are a few records which follow a more obscure convention.

[19] These are applicable for 2001-2006. As of version 1.5, they are not provided for 2007 onward.

L1	(These fields are defined as in the previous table.)
L2	
L3	
L4	
L5	
L9	
Level	
UPC	
SY	
GE	
VEND	
ITEM	
*STUBSPEC 1527RC	
00004 (name varies by category)	
The remaining columns are different. The ones for beer are:	
VALNUM5	Total ounces. If this is 1200, there are 12 ounces in the total unit. If this is 72000, there are 72 ounces in the total unit
VALNUM6	Total count. If this is 1000, there is 1 subunit. If this is 6000, there are 6 subunits.
VALNUM8	Base ounces. These are ounces in the “base “ product. This will differ from total ounces if this is a bonus pack (e.g. 3 extra ounces free).
VALNUM9	Per unit ounces. If this is 12000, there are 12 ounces in each unit.
VALTEXT5	LOUNTOTAL OUNCES – the name of the measure above, total ounces in the sales unit
VALTEXT6	LCOUTOTAL COUNT – the count in the sales unit (subunits)
VALTEXT8	OUNSBASE OUNCES – base ounces in the sales unit (will be different than total ounces for a bonus pack)
VALTEXT9	TOUNPER UNIT OUNCES – per unit ounces, the ounces in each can or bottle that makes up a subunit.
So, if VALNUM5,6,8,9 are 12000, 1000, 12000, 12000 this is a single 12 ounce can/bottle of beer.	**So, if VALNUM5,6,8,9 are 72000, 6000, 72000, 12000 this is a six pack of 12 cans/bottles of beer.**
The remaining columns are different. The ones for carbonated soft drinks are:	
VALNUM5	Total ounces. If this is 1200, there are 12 ounces in the total unit. If this is 72000, there are 72 ounces in the total unit
VALNUM6	Total count. If this is 1000, there is 1 subunit. If this is 6000, there are 6 subunits.
VALNUM7	Total pack count. This field is seldom used (21 out of 10,000+ records), and these entries may not be of value. The idea of this field is that if you are selling a box with 16 packs of 2 aspirins each, this would be “16” and the total count would be “32”.
VALNUM8	Base ounces. These are ounces in the “base “ product. This will differ from total ounces if this is a bonus pack (e.g. 3 extra ounces free).

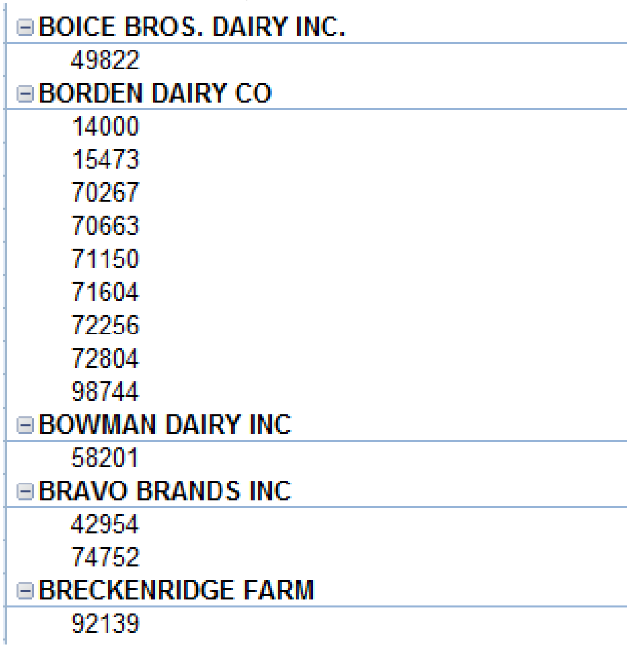
VALNUM9	Per unit ounces. If this is 12000, there are 12 ounces in each unit.
VALTEXT5	LOUNTOTAL OUNCES – the name of the measure above, total ounces in the sales unit
VALTEXT6	LCOUTOTAL COUNT – the count in the sales unit (subunits)
VALTEXT7	KCT TOTAL PACK COUNT
So, if VALNUM5,6,7,8,9 are 12000, 1000, 1000, 12000, 12000 this is a single 12 ounce can/bottle of carbonated beverage.	**So, if VALNUM5,6,7,8,9 are 72000, 6000, 1000, 72000, 12000 this is a six pack of 12 cans/bottles of carbonated beverage.**

3.8.2 VEND and vendor

Question: Could you please tell me why there are more vendor id (VEND) than vendor names (vendor) in the stub files. After I consolidate the stub file across years, I find that there are 5723 vendor ids (VEND) but only 4949 vendor names.

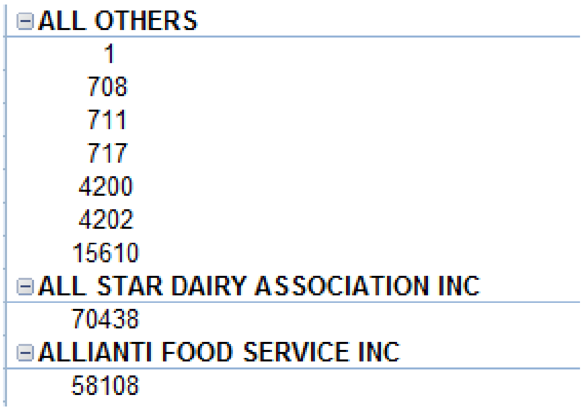
This question allows us to address two important points.

#1. It's normal for there to be more vendor IDs than vendor names for at least two reasons. The main reason is merger and acquisition activity. This is why, when we look at milk, we see multiple vendor numbers under one company.



⊟BOICE BROS. DAIRY INC.
49822
⊟BORDEN DAIRY CO
14000
15473
70267
70663
71150
71604
72256
72804
98744
⊟BOWMAN DAIRY INC
58201
⊟BRAVO BRANDS INC
42954
74752
⊟BRECKENRIDGE FARM
92139

Second, there are sometimes minor brands combined into 'all other brands', although this should not happen much. Looking at milk again, we see



⊟ALL OTHERS

1
708
711
717
4200
4202
15610

⊟ALL STAR DAIRY ASSOCIATION INC

70438

⊟ALLIANTI FOOD SERVICE INC

58108

#2. Note that to get a company, you need the fields System and Vendor. The company that's 07-30000 for system and vendor may be a different company than the one that's 03-30000 or 00-30000.

3.9 IRI week translation

This provides a translation from the IRI week number used in the files to the standard calendar.

The following conversion formulas will also work in an excel context, assuming the week is in cell A9

(equation 1) End date = (A9-400)*7+31900

(equation 2) Start date = (A9-400)*7+31900-6

So, IRI week 1369 evaluates to a start date of 11/21/2005 and an end date of 11/27/2005.

In reverse, a week can be provided for a date. Assuming the date is in cell E9, The exact week is given in equation 3, and the fractional week in equation 4.

(equation 3) Exact week = TRUNC(((E9-31894)/7)+400)

(equation 4) Fractional week = ((E9-31894)/7)+400

So, November 27, 2005 evaluates to IRI week 1369 in equation 3 and 1369.857 as the last day of 1369 in equation 4. These formulas assume Microsoft Excel date logic in which 1/1/2001 is day 36892. Date logic used in other software may vary.

3.9.1 Weeks in each year

References here to year 1, year 2, and so forth refer to the following weeks. Conveniently enough, year 1 starts 01/01/01 regardless of whether month/day/year, day/month/year, or year/month/day ordering is used.

Academic Data Set

Year	Start Week	Start day	End Week	End day
1	1114	January 1, 2001	1165	December 30, 2001
2	1166	December 31, 2001	1217	December 29, 2002
3	1218	December 30, 2002	1269	December 28, 2003
4	1270	December 29, 2003	1321	December 26, 2004
5	1322	December 27, 2004	1373	December 25, 2005
6	1374	December 26, 2005	1426	December 31, 2006
7	1427	January 1, 2007	1478	December 30, 2007
8	1479	December 31, 2007	1530	December 28, 2008
9	1531	December 29, 2008	1582	December 27, 2009
10	1583	December 28, 2009	1634	December 26, 2010
11	1635	December 27, 2010	1686	December 25, 2011

Year 6 has 53 weeks.

3.10 Counties (FIPS) in IRI markets[20]

The file **"fips by IRI market.xls"** is in the folder "**demos trips external**". This contains the list of counties in the United States, and the IRI market they belong to. All counties are listed; those which are not in an IRI standard market are labeled "white space".

Key	FIPS code (assigned by US Census Bureau)
Name	This is a combination of the IRI market name, the state, and the county name, separated by "|".
NAME_COUNTY	County name
NAME_STATE	State name
MARKET NAME	IRI market name, version 1. Note counties which are not in an IRI market are called White Space.
REGION NAME	IRI standard region name
IRI_MARKET_NUMBER	IRI market number
MARKET NAME2	IRI market name, version 2

[20] This file added April 1, 2009 (drives ADS_80 and higher). It is also available on the Google support site.

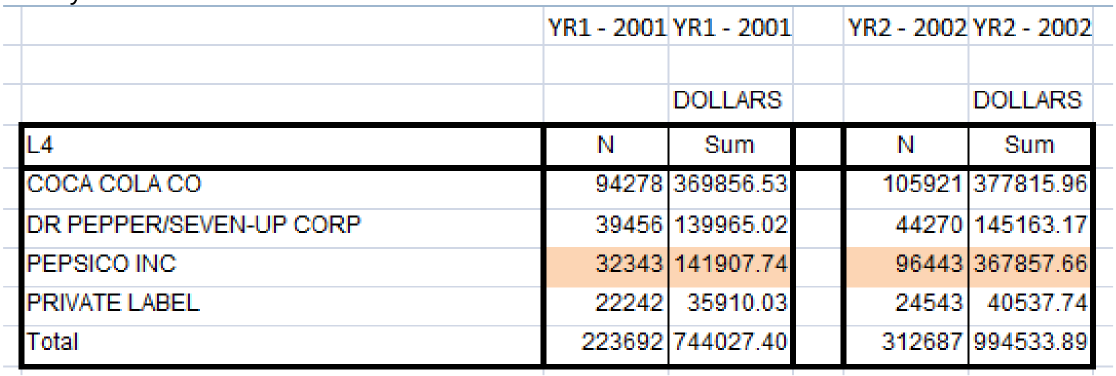
4. Appendices

4.1 Errata

This is a listing of known issues with the data set.

1. Year 1 Pepsi data(May 8, 2013) 1

. The 2001 Pepsi panel data is too low. The corporate level numbers are below. There is no listed testing activity related to this. The effect is particularly strong in regular Pepsi 8, 12 and 24 pack,suggesting this may be related to missing PLU activity.



	YR1 - 2001	YR1 - 2001		YR2 - 2002	YR2 - 2002
		DOLLARS			DOLLARS
L4	N	Sum		N	Sum
COCA COLA CO	94278	369856.53		105921	377815.96
DR PEPPER/SEVEN-UP CORP	39456	139965.02		44270	145163.17
PEPSICO INC	32343	141907.74		96443	367857.66
PRIVATE LABEL	22242	35910.03		24543	40537.74
Total	223692	744027.40		312687	994533.89

With data this old, that's been off the system for years, there's not much to do except recommend you don't use it.

2. Year 8-11 demographics of single/married (May 8, 2013)

The married/single panelist household demographic information for years 8-11 is "flipped". We are seeing whether this is a general problem with the demographic file that goes beyond this field and plan to re-issue the file at that time.

4.2 Chain cross-reference

Chain information is masked by year.

Chain1 only occurs in year1.

There is a chain which is Chain13 in year1, chain12 in year2, chain 11 in year3, and chain 10 in years 4 and 5.

The chain “NONE” is listed as “999” here to distinguish it from blank, but in the data this is just called “NONE”.

This table is contained in the files

- **masked_chain_xref.csv** for years 1-5 and
- **masked_chain_xref1_7.csv** for years 1-7
- **masked_chain_xref1_11.csv** for years 1-11

in the **demos trips external** folder. Only part of the table is shown here.

Year1	Year2	Year3	Year4	Year5

1				
2	1			
3	2	1	1	1
4	3	2	2	2
5	4	3	3	3
6	5	4	4	4
7	6	5	5	5
8	7	6	6	6
9	8	7	7	7
10	9	8	8	8
11	10	9	9	9
12	11	10		
13	12	11	10	10
…				
999	999	999	999	999

4.3 Pacesetters new product information

In October 2008, the Pacesetter files were added to the Google Groups site. This information has been added to distribution disks of the IRI Marketing Data Set effective with drive ADS_34[21].

The IRI New Product Pacesetter set of reports uses a new product screen in which a brand must achieve a certain national distribution (30%), and then have 52 week sales by December of $7,500,000 or more. Top new products from these years are listed.

Because the naming conventions have changed over the years, the index file *"_readme Pacesetters files ADS.xls*" is provided for background. An earlier version is listed below – later files have been added.

File name	**Scope**	**Notes**
IRITTNovDec00NewProducts2.pdf	Overall	Pacesetters 2000
SepOctCPGNewProducts 2000.pdf	Overall	Pacesetters 2000
IRIPacesetters1ExSum 2001 part1.pdf	Overall	Pacesetters 2001
NPP2001ExecBrief_Food.pdf	Food and Beverage	Pacesetters 2001
NPP2001ExecBrief_NonFood.pdf	Nonfood	Pacesetters 2001
NPP2002.pdf	Overall	Pacesetters 2001
npp_benefittrends_foodbev0503.pdf	Food and Beverage	Pacesetters 2002
npp_benefittrends_nonfood0603.pdf	Nonfood	Pacesetters 2002
nppfoodbevfeb04.ppt	Food & Beverages	Pacesetters 2003
nppnonfoodmar04.ppt	Nonfood	Pacesetters 2003
thought_times0504.ppt	Food & Beverages benefits	Pacesetters 2003
thought_times0604.ppt	Nonfood	Pacesetters 2003
tt_issue032005.pdf	Overall	Pacesetters 2004
tt_pacesetters_0306.pdf	Overall	Pacesetters 2005
tt_pacesetters_0307.pdf	Overall	Pacesetters 2006
TT_March_2008_NPP_Final.pdf	Overall	Pacesetters 2007

[21] Lower numbers, such as ADS_33, do not have this information, and it must be downloaded from the Google Groups site if desired. The Google Groups site may have updated information.

4.4 IRI contract/nondisclosure agreement

This version is dated June 5, 2008. The version you signed may be slightly different. The version you signed is the relevant one; this is included for quick reference.

[Date]
[Researcher Name]
[Address/Zip]

RE: Proposal and Confidentiality Agreement regarding InfoScan® Services

Dear _______________:

On behalf of Information Resources, Inc. ("IRI"), I am pleased to make the following proposal to provide the InfoScan Services outlined on Attachment 1 to this letter (the "InfoScan Services") to ___________________________________ (referred to herein as "Researcher" or "you") in accordance with the following terms:

1. IRI will provide the InfoScan Services to you as described in Attachment 1 to this letter agreement. Unless otherwise agreed in writing, additional extra cost InfoScan-related services subscribed to by you during the term of this Agreement will be subject to the same terms and conditions as the InfoScan Services provided hereunder.

2. If you accept this proposal, you agree to pay the following price for the InfoScan Services in accordance with the following payment terms:

Price: _______________

An invoice will be generated upon contract signature and the invoiced amount must be paid in full prior to IRI delivering the data. The prices in this proposal are valid only if you sign and return this agreement by no later than _________.

The prices quoted herein are exclusive of taxes, if any are applicable. You agree promptly to reimburse IRI (or pay directly if so requested by IRI) all taxes, charges and fees imposed by any governmental body or agency upon or in connection with the transaction contemplated by this Agreement excluding all taxes measured by net income. Upon request, you agree to provide IRI with proof of such payment.

3. (a) The reports, data and/or related analysis (if any) provided by IRI under this Agreement (collectively, "IRI Data") is provided to you solely for your own use as one of many sources for your academic research for the project(s) described in Exhibit A attached (the "Research Projects"). You agree to take all reasonable precautions not to disclose or allow to be

disclosed any of the IRI Data or other IRI proprietary or confidential to which you have been provided access hereunder to any other person, firm or other entity without the prior written consent of IRI, except as otherwise explicitly permitted by the provisions in this Paragraph 3. You may not use the IRI Data for any other purpose, including without limitation, any commercial purpose.

(b) You may publish results of your academic research relating to the Research Projects based on the IRI Data in academic publications or academic working papers relating to such Research Projects or at academic conferences ("Academic Publications") Academic Publications provided that:

Any category level, brand level, market or regional level, and/or retailer or store level data included as part of the deliverable set forth in Exhibit 1 is provided to you solely for your internal purposes in connection with a Research Project. Unless otherwise expressly permitted in Exhibit 1 with respect to category or sub-category (or type) level data, in no event shall you publish, report or otherwise disclose (or allow any third party to whom you are permitted to disclose data hereunder (if any), to publish, report or disclose) IRI Data below the category, sub-category (or type) level or in any manner that could allow a third party to identify, derive or otherwise infer the identify of a specific retailer. In publications in which a lower level of detail than category or type or market totals is required, this should be published in a way that respects the fact that that this data is provided for the advancement of general marketing science, rather than the evaluation of the strategies used by particular manufacturers or retailers. Without specific written permission, retailer names should be obscured by referring to code names such as "retailer X" in such a manner that third parties could not easily infer the identity of a specific retailer.

i. IRI is referenced substantially as follows: "University of __________ estimate or analysis based [in part] on Information Resources, Inc. data as analyzed by University of ___________". The reference section should contain the following reference to the dataset:

   Bronnenberg, Bart J., Kruger, Michael W, and Mela, Carl F. The IRI Marketing Dataset. submitted to **Marketing Science**, 2008. *(specific page reference not yet available)*

ii. The scope of the disclosure of IRI Data is limited to the extent necessary to support the specific results of the Research Project, and without limiting the generality of the foregoing you may not under any circumstances post IRI Data to the internet or in any electronic format in any form or manner without IRI's express prior written consent;

iii. The IRI Data is disclosed in a way that is not misleading, and accurately identifies the four components of time period, category, geography and measures;

iv. IRI is provided an opportunity to review the final results based on or incorporating IRI Data prior to any submission for publication, solely to confirm that IRI Data is being represented in a non-misleading fashion and otherwise in accordance with the requirements of this Agreement. It is not IRI's interest or intent to provide any editorial control of the outcome of the Research Project, rather IRI's review shall be limited solely to the presentation of IRI Data. IRI shall use all reasonable efforts to respond in writing to the Researcher within ten (10) business days of receiving a copy of the research. IRI's review may require shorter or longer timeframe depending on the scope, length and nature of the publication based on IRI Data. To expedite that necessary review, you agree to send the request to IRI Legal

Department, Attn: General Counsel, Information Resources, Inc. 150 North Clinton Street, Chicago, Illinois 60661).

v. Publication includes publication in printed or on-line journals, and in media which are generally available to the public or academic community. This would include working papers available on an archive, such as a web site (unless restricted to the research team). This would not include conference presentations or colloquia which may include publication of a brief abstract, but for which no proceedings are published.

vi. The final results described in subparagraph (b) above may be distributed internally within IRI. IRI will make reasonable efforts to avoid external distribution but assumes no legal liability for unauthorized distribution. IRI may publish references to academic publications which have used this data set.

(c) Requests for use of IRI Data provided hereunder to Researcher for academic research projects other than the Research Projects described in Exhibit A will require IRI's prior written consent. All such requests describing the research project scope should be directed to the IRI Legal Department, Attn: General Counsel, Information Resources, Inc., 150 North Clinton Street, Chicago, IL 60661. IRI shall review such request and use reasonable efforts to respond in writing within ten (10) business days from receipt of such request. Any consent granted by IRI shall be in writing referencing this Agreement and shall be subject to the limited use and nondisclosure provisions of this Paragraph 3 as well as all other applicable provisions of this Agreement and any other terms that IRI may specify. In the event that IRI provides its consent, there will be no additional charge to you for using the IRI Data previously provided and paid for under this Agreement.

(d) Researcher agrees that neither Researcher nor any third party academic researcher working with you will use, or attempt to use, or permit or allow the use of, any IRI Data provided by IRI hereunder, in any legal proceedings (including, but not limited to, any use in litigation and/or use with any governmental, investigatory, regulatory or other body or authority) except (i) if and to the extent compelled by service of legal process or in response to an official governmental demand; and (ii) in those cases only if Researcher (a) gives IRI prompt advance notice thereof; and (b) make reasonable efforts to obtain appropriate confidentiality agreements and/or protective orders in form and substance reasonably acceptable to IRI.

(e) You also agree to provide IRI a copy or summary of your research based on the IRI data.

4. The data to which you are provided access hereunder shall belong to IRI. IRI reserves the right to resell the data in any form to third parties. You may not resell data, reports or portions of reports in any form.

5. **THE IRI DATA IS PROVIDED AS IS, WITHOUT WARRANTY. IRI MAKES NO REPRESENTATION OR WARRANTY AS TO THE VALUE, MERCHANTABILITY, DESIGN OR FITNESS FOR USE FOR A PARTICULAR PURPOSE OF THE DATA TO BE PROVIDED HEREUNDER.** In the event that IRI is unable to perform hereunder for any reason, IRI's liability shall be limited to a refund or credit of the amount paid for that portion of this Agreement that IRI has not fulfilled and in no event shall IRI be liable for lost profits, good will or other special or consequential damages of any kind.

6. Neither party shall be liable to the other party for any loss, injury, delay, damages or casualty suffered by the other party due to strikes, governmental action, unusually severe weather, acts of God or public enemy, or any other cause which is beyond the reasonable control of either party, and any failure or delay by either party in the performance of its obligations under this Agreement due to one or more of the foregoing causes will not be considered a breach of this Agreement.

7. No waiver, alteration or modification of any provision herein shall be binding upon either party unless made in writing and agreed to by a duly authorized officer of the party sought to be bound. This Agreement may not be assigned by you. Waiver by either party of any default hereunder shall not be deemed a waiver by such party of any default by either party which may thereafter occur. This Agreement shall be governed by and construed under the laws of the State of Illinois. This Agreement sets forth the entire agreement between the parties and takes the place of all prior verbal or written communications concerning the subject of this Agreement.

Please acknowledge your acceptance of these proposal terms and your agreement to the foregoing by countersigning where indicated below.

Sincerely,

INFORMATION RESOURCES, INC.

By: ______________________________

Title: ____________________________

ACCEPTED AND AGREED TO BY [INSERT NAME OF RESEARCHER]

Signature: ___________________________

Name/Title:___________________________

Date: ___________________________

EXHIBIT A

RESEARCH PROJECT(S)

Overview of Projects To Be Undertaken by Researcher Utilizing IRI Data

PROVIDE BRIEF SUMMARY

ATTACHMENT 1

(this attachment will contain a list of the database files and the information contained)

This is the dataset called "Academic Dataset External". This database is described in

Bronnenberg, Bart J., Kruger, Michael W, and Mela, Carl F. The IRI Marketing Dataset. submitted to **Marketing Science**, 2008.

4.5 TNS Terms of Use

TNS has generously made available some advertising data sets. These data are described in a separate document.

Operationally, the terms of use should be signed at the same time that the IRI contract/nondisclosure agreement is signed and returned to IRI, which will then distribute the datasets. The agreement on the terms of use of the TNS data is between TNS and the academic researcher only.

The TNS Attorney would like the Academics using the data to sign the attached Terms Of Use Sheet.

Gaurav
Gaurav Bhalla
Global Innovation Director
TNS
8605 Westwood Center Drive Suite 207
Vienna VA 22182

Marc Levin, Esq.
SVP & General Counsel, North America
TNS
100 Park Avenue, 4th Floor
New York, NY 10017

TERMS OF USE

USE OF THE DATA PROVIDED BY TNS CUSTOM RESEARCH, INC. (“TNS”) TO YOU AND/OR YOUR ORGANIZATION (THE “DATA”) IS SUBJECT TO THESE TERMS. YOU AND YOUR ORGANIZATION ACCEPTS THESE TERMS BY USING THE DATA. IF YOU AND/OR YOUR ORGANIZATION DO NOT AGREE TO ALL OF THESE TERMS, DO NOT USE THE DATA.

OWNERSHIP AND COPYRIGHT; USE RESTRICTIONS. The Data are owned by TNS and are being licensed and not sold to you and/or your organization; and thus, other than the license, you and/or your organization shall not receive any right, title or interest in the Data. You and/or your organization may use the Data only for academic purposes. You and/or your organization may not use the Data for any other purpose, including, without limitation, for consultative purposes. Any dissemination of the Data must identify TNS as the source of the Data. You and/or your organization may not sell, license, sublicense or otherwise commercially transfer the Data. Any breach or attempted breach by you and/or your organization of the provisions hereof may cause TNS irreparable injury, for which TNS may seek, in addition to any and all other remedies available to TNS, temporary and permanent injunctive relief.

LIMITED WARRANTY; LIMITED LIABILITY. THE DATA ARE PROVIDED “AS IS” AND WITHOUT WARRANTY OF ANY KIND. WITHOUT LIMITING THE GENERALITY OF THE FOREGOING, TNS DOES NOT WARRANT, GUARANTEE, OR MAKE ANY REPRESENTATIONS REGARDING THE USE OR THE RESULTS OF THE USE OF THE DATA IN TERMS OF CORRECTNESS, ACCURACY, RELIABILITY OR OTHERWISE. TNS MAKES NO WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. TNS WILL NOT BE LIABLE FOR ANY DAMAGES, WHETHER INCIDENTAL OR DIRECT, IN CONNECTION WITH THE USE OF THE DATA BY YOU AND/OR YOUR ORGANIZATION, EVEN IF TNS HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

5. Appendix: How to Update Year 6 and 7, and years 8-11

This appendix was written for the year 6 update, and has been edited for the year 7 update. The procedure is the same EXCEPT for the additional product stubs applicable to year 7.

5.1 Purpose

This document describes how to update the IRI Marketing Data Set for academic use, and add a 6th year [7th year] [8th through 11th year] to that data set.

.***By using additional years, you agree to the same NDA / Terms of Use which applied to years 1-5***[22]

Note that drives which are numbered ADS_55 and higher have year 6 on them already.
Note that drives which are numbered ADS_144 and higher have year 6 and 7 on them already.
Note that drives numbered ADS_215 and higher have years 8-11 on them already.

The 6th year [7th year] is in the same format as years 1-5, with minor exceptions. These exceptions are certain definitional files which were identical across all categories. In order to save space, these were included once, in the "demos trips" folder.

5.2 Do you really want to do this?

The answer is almost certainly yes for year 6.

- This process does not overwrite any files.
- This process makes the data set on disk bigger, but doesn't otherwise change it.
- You almost certainly aren't working with the original files, but with a copy of them formatted for your needs and stored in some other set of folders, so if you are following this basic hygiene this process shouldn't interfere with anything you have done[23].

As to whether you want to incorporate this data in any project which is already in progress – that's a more difficult question that's up to you. It does provide a convenient holdout sample.

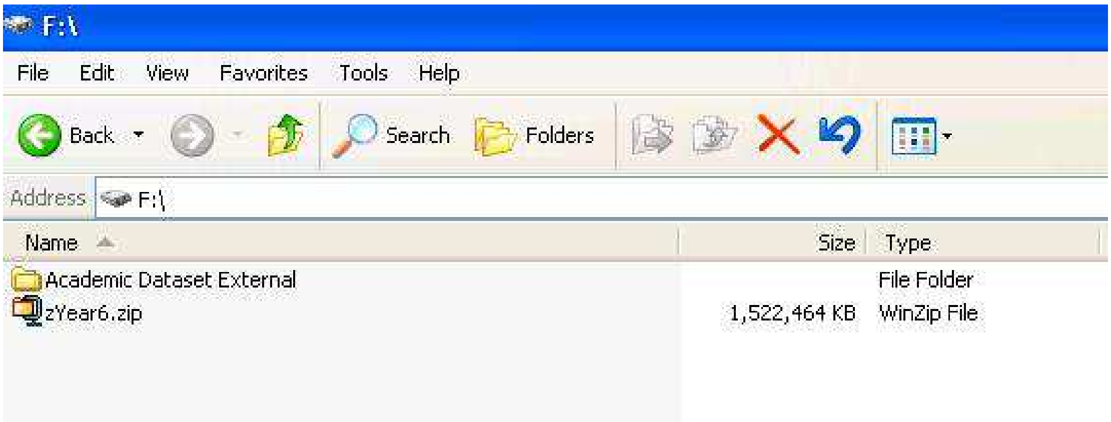
For year 7, the answer is less clear due to the stub update. See the product attribute section for a discussion

5.3 Copy the zipped data file from the DVD onto the hard drive.

The resulting drive is likely to look like this. It may have another drive letter other than **F:**, and there may be other files on that drive, but only these two are of interest right now.

[22] IRI does not promise to continue this program. Do not make the completion of any project depend on IRI providing additional data.

[23] Nevertheless, note that no warranty is expressed or implied, and neither Michael Kruger nor IRI is responsible for any problems that may occur.

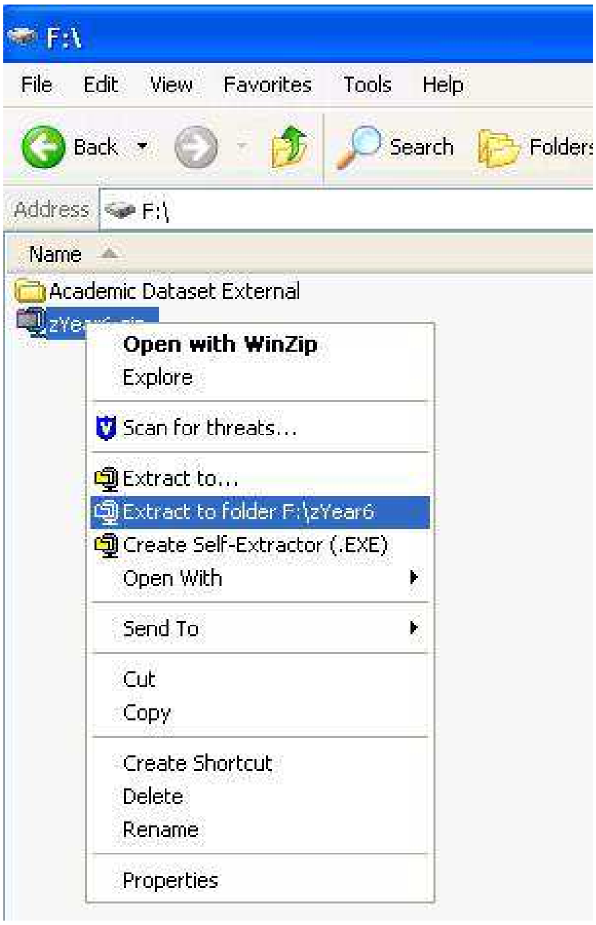


Name	Size	Type
Academic Dataset External		File Folder
zYear6.zip	1,522,464 KB	WinZip File

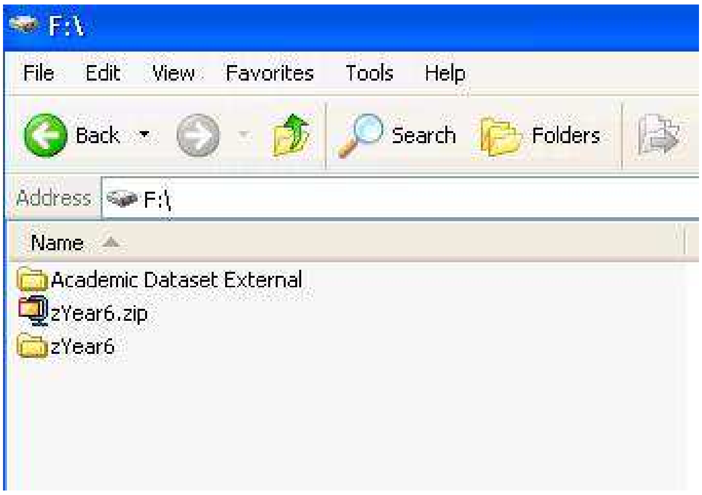
5.4 Unzip zYear6.zip file [zYear7.zip file, Year8.zip, Year9.zip, Year10.zip, Year11.zip]

It is most likely that you will be able to right click on the file and indicate that you want to unzip it, so that's what I've shown here.

If you have some other way to get to WinZip, be sure you extract these files in a way that preserves the directory structure.

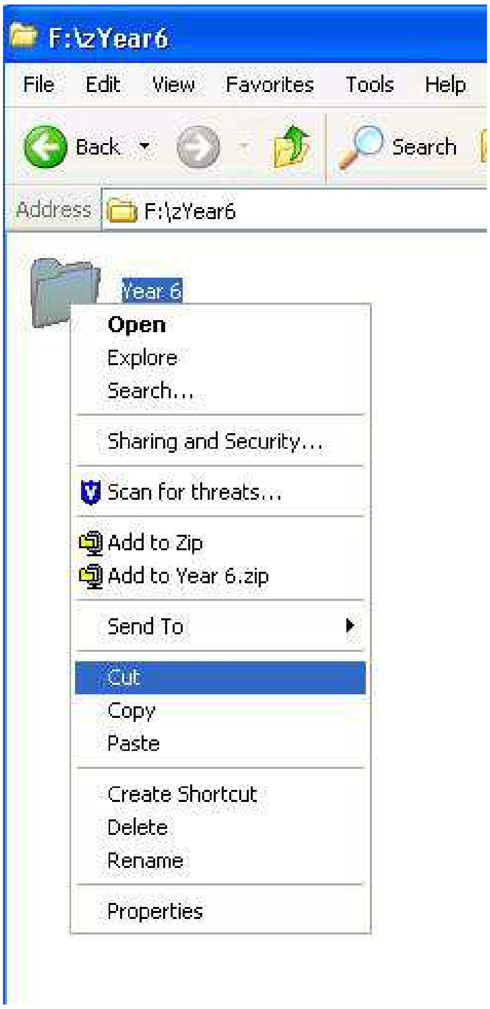


At the end of this process, which will take several minutes, your folder will look like this:



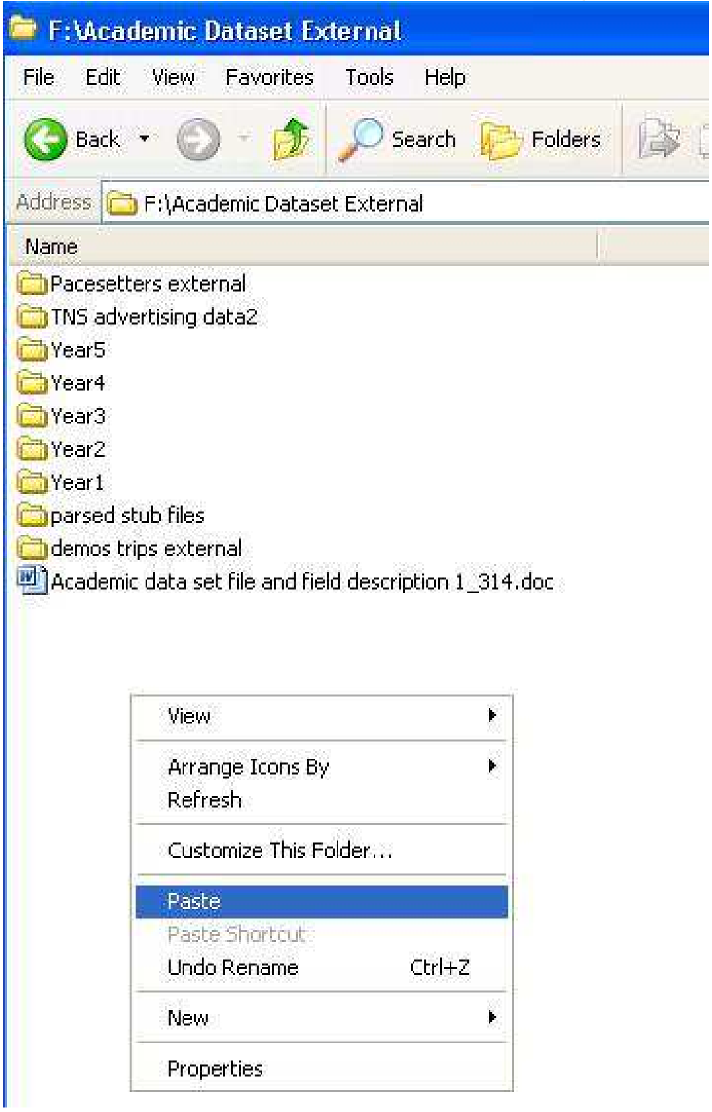
Academic Dataset External
zYear6.zip
zYear6

Select the **zYear6** folder or **Year6** folder. Inside it you will see another folder called **Year6**.
[***OR,*** you may just see a folder called “External”. If that’s what you see, just rename the folder from zYear6 to Year6 if necessary and skip the cutting and pasting.]
Cut it.

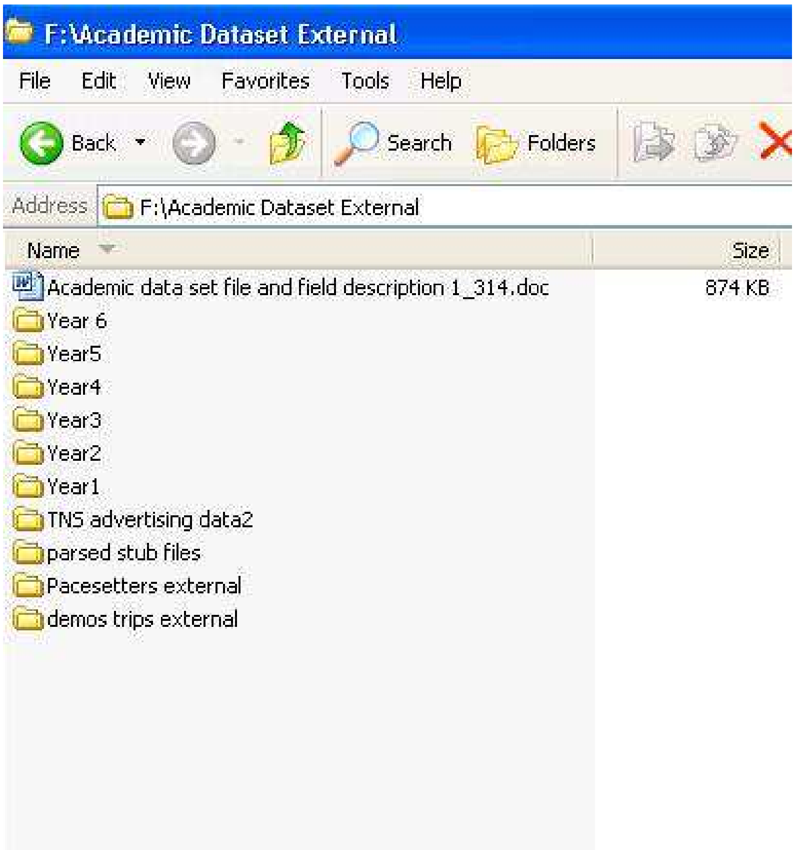


F:\zYear6
File
Edit
View
Favorites
Tools
Help
Back
Search
Address
F:\zYear6
Year 6
Open
Explore
Search...
Sharing and Security...
Scan for threats...
Add to Zip
Add to Year 6.zip
Send To
Cut
Copy
Paste
Create Shortcut
Delete
Rename
Properties

Go to the **Academic Dataset External** folder and paste:



After you do, the folder will look like this:



Name	Size
Academic data set file and field description 1_314.doc	874 KB
Year 6	
Year5	
Year4	
Year3	
Year2	
Year1	
TNS advertising data2	
parsed stub files	
Pacesetters external	
demos trips external	

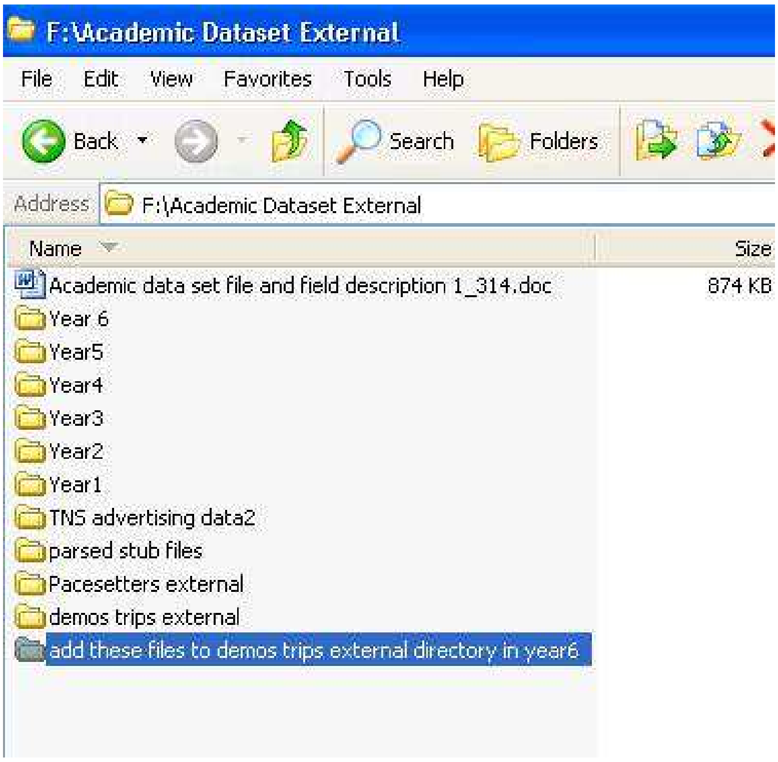
If you now open the Year 6 folder, you should see the same familiar structure as was provided in year 1 through year 5.

[follow the same routine for year 7]

5.5 Updating demos trips external folder

The newer demo files, trip files, and some of the files which were previously duplicated in all the category folders are in a folder on the CD called **add these files to demos trips external directory in year6**. [**add these files to demos trips external directory in year7 – this is a smaller set of files because some were already included with year 6]**

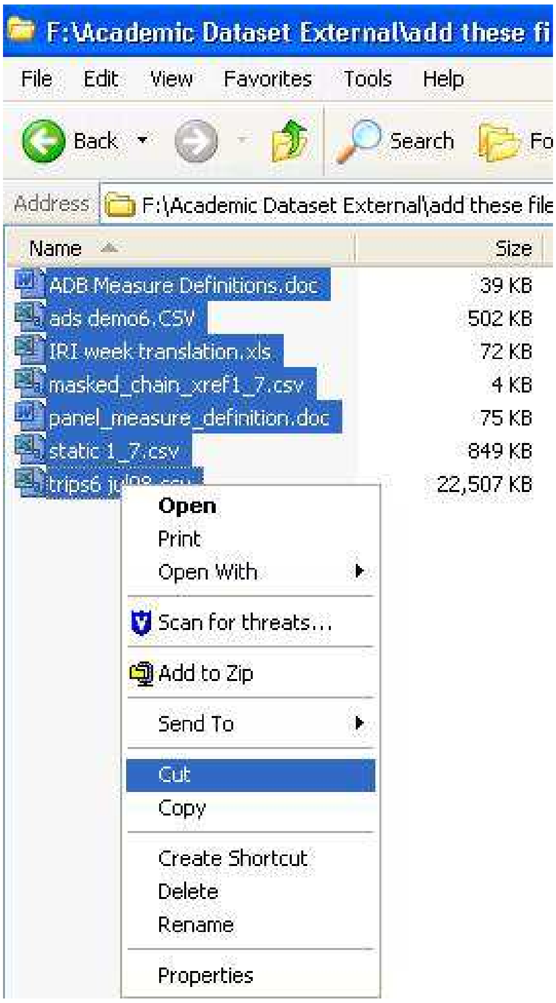
Copy this folder from the CD to the drive, in the **Academic Dataset External** folder. This folder will look like this:



F:\Academic Dataset External

Name	Size
Academic data set file and field description 1_314.doc	874 KB
Year 6	
Year5	
Year4	
Year3	
Year2	
Year1	
TNS advertising data2	
parsed stub files	
Pacesetters external	
demos trips external	
add these files to demos trips external directory in year6	

Open the **add these files to demos trips external directory in year6** folder, and cut all the files there:



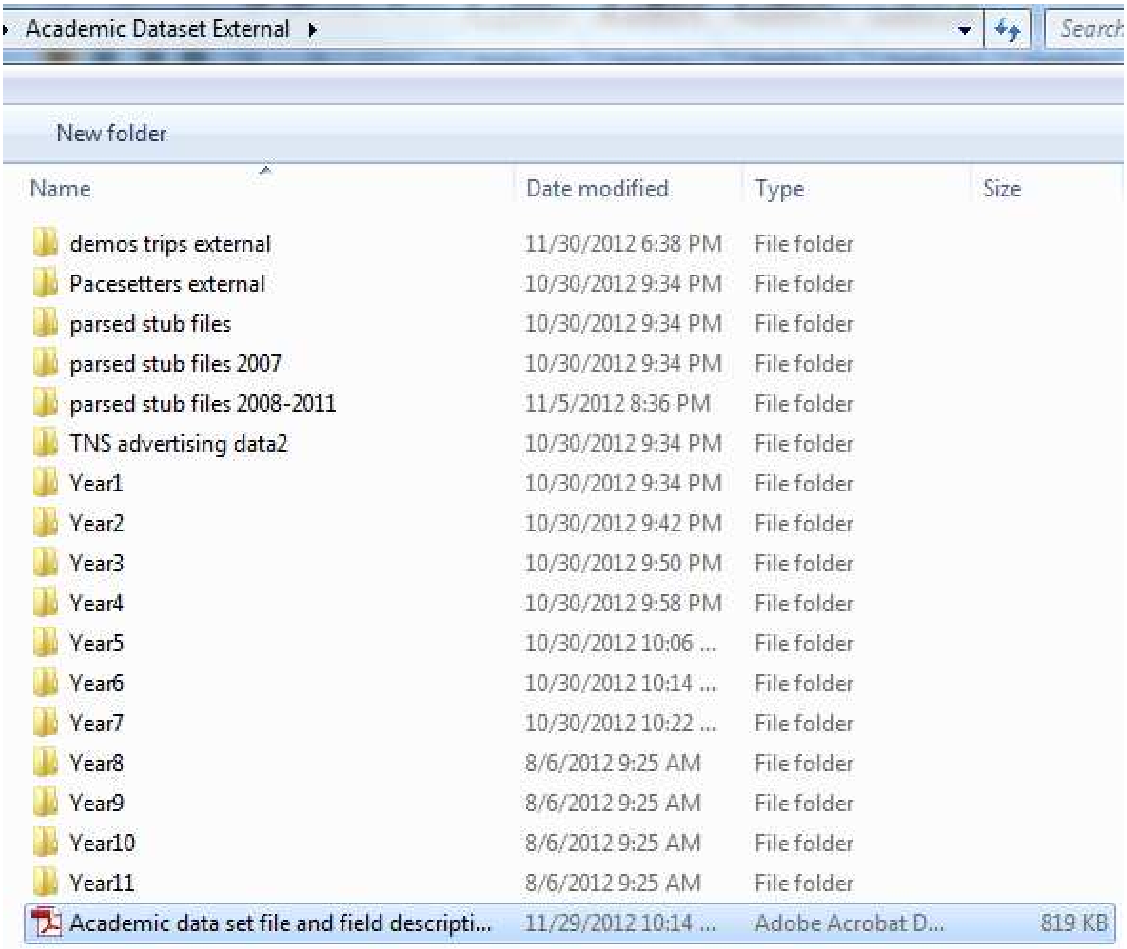
F:\Academic Dataset External\add these fi

Address: F:\Academic Dataset External\add these file

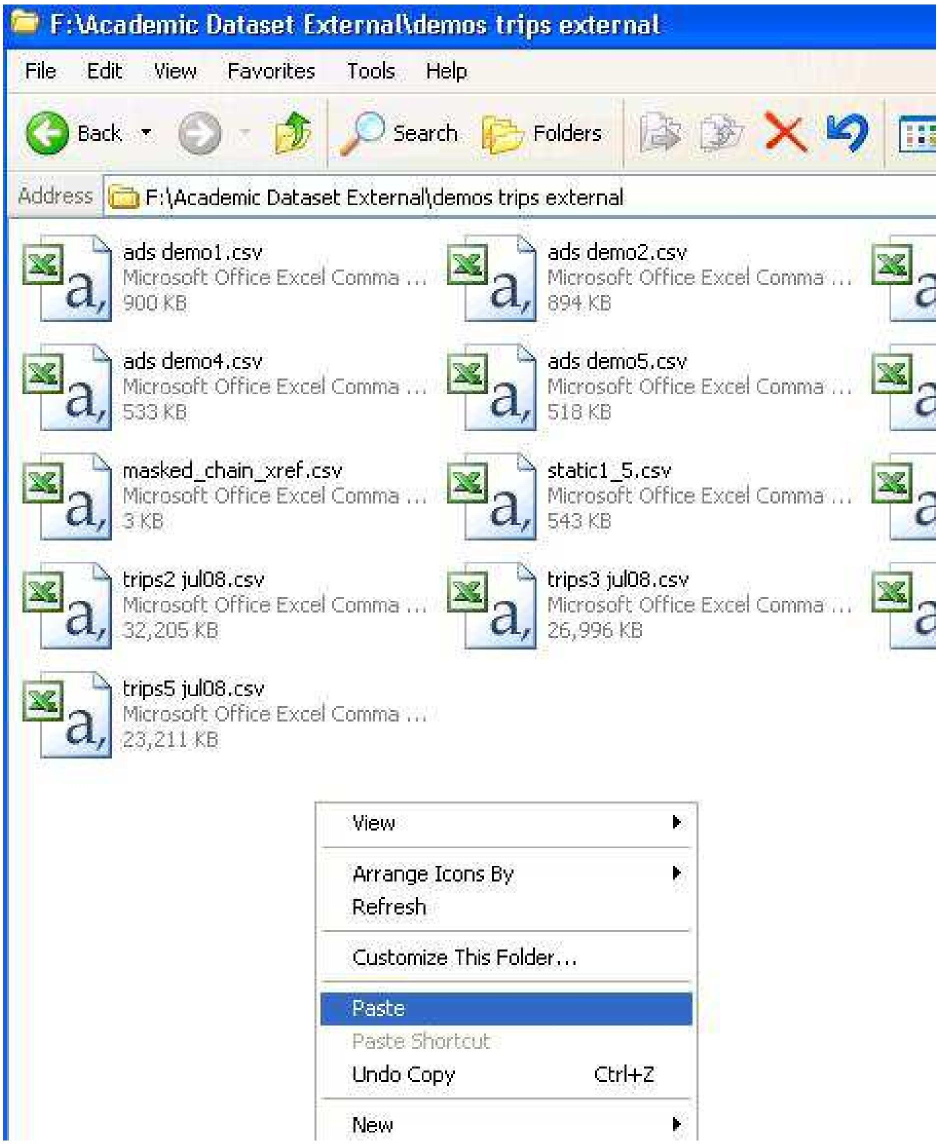
Name	Size
ADB Measure Definitions.doc	39 KB
ads demo6.CSV	502 KB
IRI week translation.xls	72 KB
masked_chain_xref1_7.csv	4 KB
panel_measure_definition.doc	75 KB
static 1_7.csv	849 KB
trips6 ju	22,507 KB

Go to the **Academic Dataset External\demos trips external** folder and past those files there.

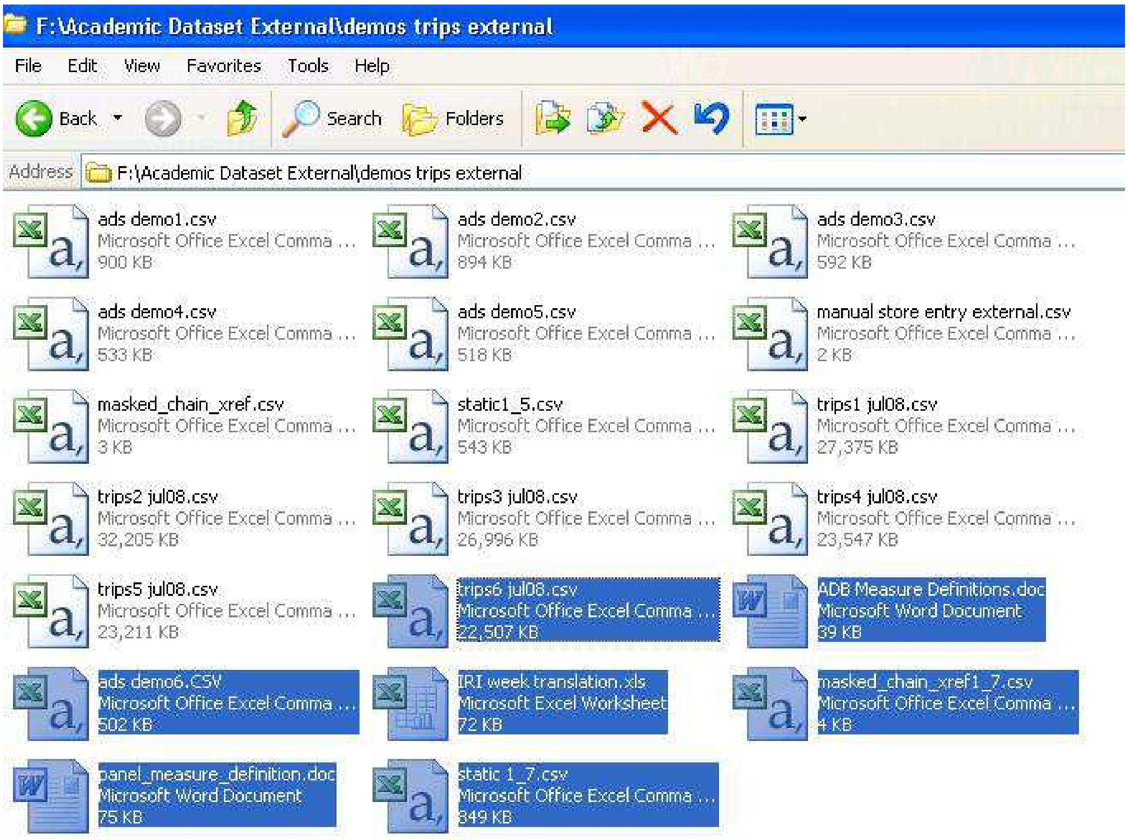
How it should look at the end:



Name	Date modified	Type	Size
demos trips external	11/30/2012 6:38 PM	File folder	
Pacesetters external	10/30/2012 9:34 PM	File folder	
parsed stub files	10/30/2012 9:34 PM	File folder	
parsed stub files 2007	10/30/2012 9:34 PM	File folder	
parsed stub files 2008-2011	11/5/2012 8:36 PM	File folder	
TNS advertising data2	10/30/2012 9:34 PM	File folder	
Year1	10/30/2012 9:34 PM	File folder	
Year2	10/30/2012 9:42 PM	File folder	
Year3	10/30/2012 9:50 PM	File folder	
Year4	10/30/2012 9:58 PM	File folder	
Year5	10/30/2012 10:06 ...	File folder	
Year6	10/30/2012 10:14 ...	File folder	
Year7	10/30/2012 10:22 ...	File folder	
Year8	8/6/2012 9:25 AM	File folder	
Year9	8/6/2012 9:25 AM	File folder	
Year10	8/6/2012 9:25 AM	File folder	
Year11	8/6/2012 9:25 AM	File folder	
Academic data set file and field descripti...	11/29/2012 10:14 ...	Adobe Acrobat D...	819 KB


F:\Academic Dataset External\demos trips external
File Edit View Favorites Tools Help
Back
Search
Folders
Address F:\Academic Dataset External\demos trips external
ads demo1.csv
Microsoft Office Excel Comma ...
900 KB
ads demo2.csv
Microsoft Office Excel Comma ...
894 KB
ads demo4.csv
Microsoft Office Excel Comma ...
533 KB
ads demo5.csv
Microsoft Office Excel Comma ...
518 KB
masked_chain_xref.csv
Microsoft Office Excel Comma ...
3 KB
static1_5.csv
Microsoft Office Excel Comma ...
543 KB
trips2 jul08.csv
Microsoft Office Excel Comma ...
32,205 KB
trips3 jul08.csv
Microsoft Office Excel Comma ...
26,996 KB
trips5 jul08.csv
Microsoft Office Excel Comma ...
23,211 KB
View
Arrange Icons By
Refresh
Customize This Folder...
Paste
Paste Shortcut
Undo Copy Ctrl+Z
New

This will leave this folder with these files in it:



F:\Academic Dataset External\demos trips external

F:\Academic Dataset External\demos trips external

ads demo1.csv
Microsoft Office Excel Comma ...
900 KB

ads demo2.csv
Microsoft Office Excel Comma ...
894 KB

ads demo3.csv
Microsoft Office Excel Comma ...
592 KB

ads demo4.csv
Microsoft Office Excel Comma ...
533 KB

ads demo5.csv
Microsoft Office Excel Comma ...
518 KB

manual store entry external.csv
Microsoft Office Excel Comma ...
2 KB

masked_chain_xref.csv
Microsoft Office Excel Comma ...
3 KB

static1_5.csv
Microsoft Office Excel Comma ...
543 KB

trips1 jul08.csv
Microsoft Office Excel Comma ...
27,375 KB

trips2 jul08.csv
Microsoft Office Excel Comma ...
32,205 KB

trips3 jul08.csv
Microsoft Office Excel Comma ...
26,996 KB

trips4 jul08.csv
Microsoft Office Excel Comma ...
23,547 KB

trips5 jul08.csv
Microsoft Office Excel Comma ...
23,211 KB

trips6 jul08.csv
Microsoft Office Excel Comma ...
22,507 KB

ADB Measure Definitions.doc
Microsoft Word Document
39 KB

ads demo6.CSV
Microsoft Office Excel Comma ...
502 KB

IRI week translation.xls
Microsoft Excel Worksheet
72 KB

masked_chain_xref1_7.csv
Microsoft Office Excel Comma ...
4 KB

panel_measure_definition.doc
Microsoft Word Document
75 KB

static 1_7.csv
Microsoft Office Excel Comma ...
849 KB

5.6 Updating the stub files (product attributes) for year 7

This section is only applicable to year 7.

Unzip the file **zParsed stub files 2007.zip** and place these files in the directory **Parsed stub files 2007.**

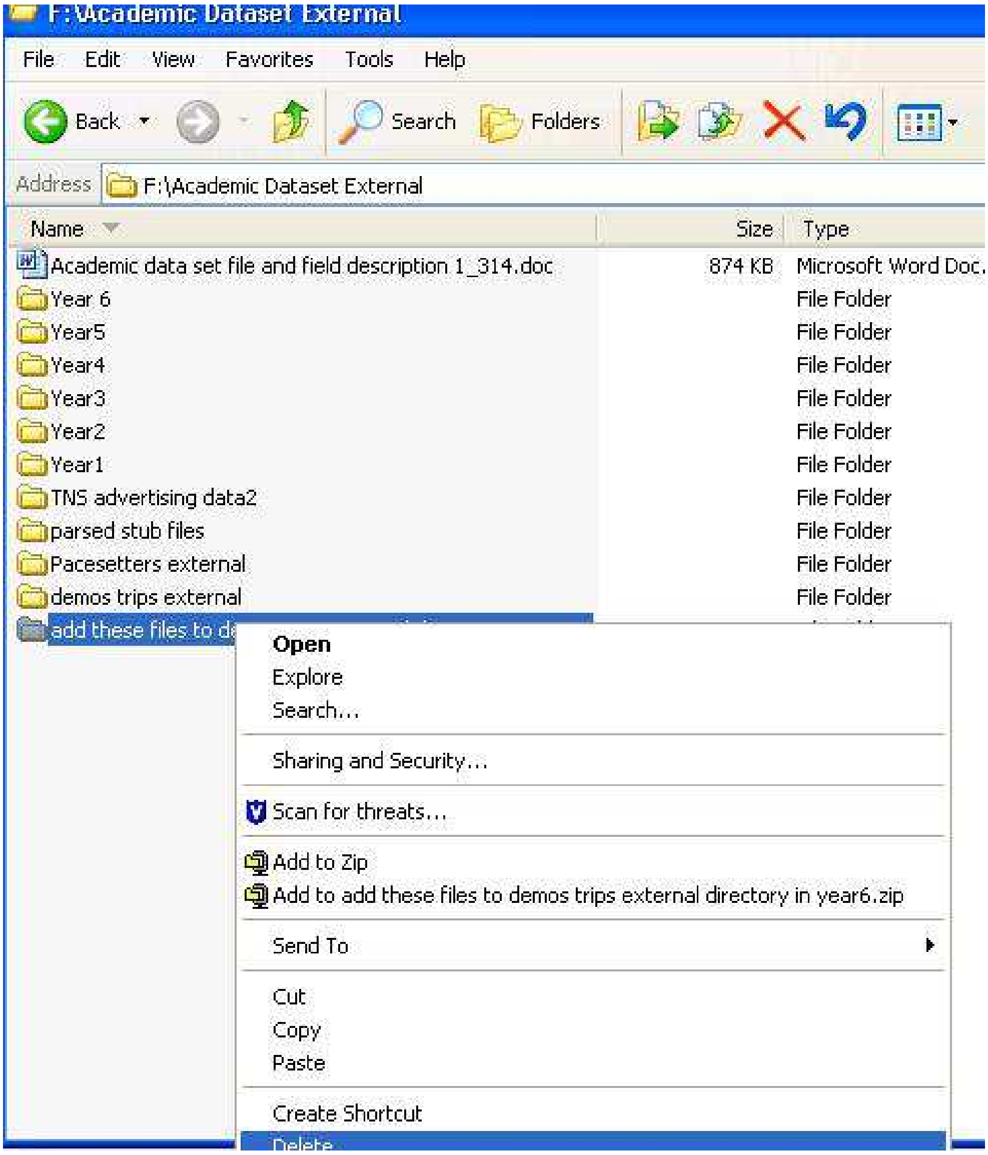
Because of the amount of manual handling the file names in 2001-2006 may be different in 2007, although the fields within the files are the same.

5.7 Cleanup

Now we want to get rid of the folders and files we don't need any more.

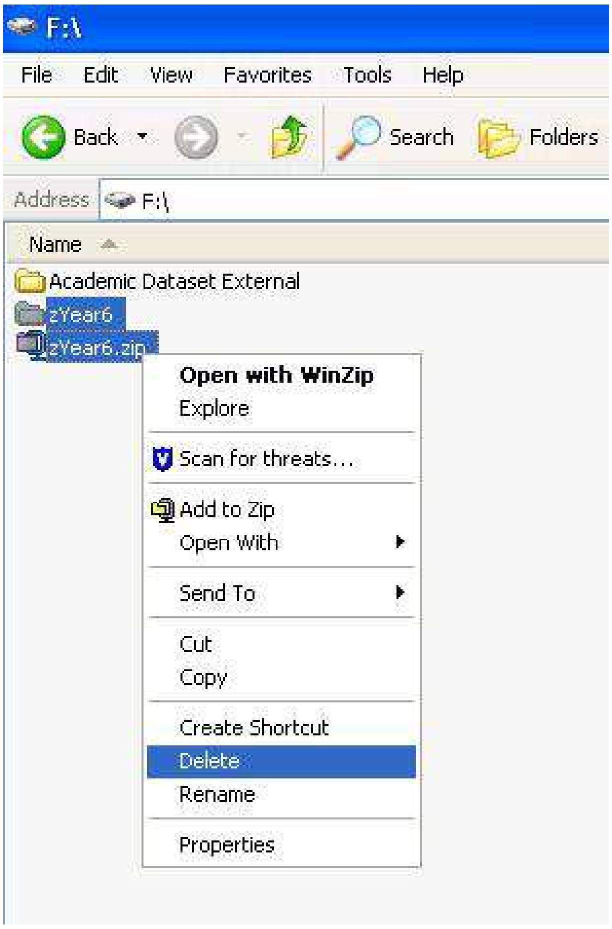
Delete the folder **add these files to demos trips external directory in year6**. **[add these files to demos trips external directory in year7].**

Delete any folders which are now empty.



Name	Size	Type
Academic data set file and field description 1_314.doc	874 KB	Microsoft Word Doc.
Year 6		File Folder
Year5		File Folder
Year4		File Folder
Year3		File Folder
Year2		File Folder
Year1		File Folder
TNS advertising data2		File Folder
parsed stub files		File Folder
Pacesetters external		File Folder
demos trips external		File Folder
add these files to d		

Delete the folder **zYear6** and the file zYear6.zip. [equivalents in year7].



5.8 That’s it. You’re done.

Put the DVD with year6 [year7, etc.] in a safe place.

There is no additional New Product Pacesetters data at this time. This will likely be posted on the Google Groups site when it is available.

There is no update of the TNS advertising data available.

FAQ - Frequently Asked Questions

These are the FAQ accumulated on the Google Groups and Google Sites site. The newest version is accessible here: https://sites.google.com/site/irimarketingdataset/home/faq . Because some of these questions have been moved twice on the way to this document, some links may be broken.

Data Definition Questions

Q. If I have a question for IRI about the data, who should I ask?

A. Send an e-mail to the Google Groups support list iri_marketing_data_set@googlegroups.com .Answers to questions of general interest may be posted here.

Note: Mike Kruger will be retiring at the end of 2012.

Q. "I have attached two data files I had from Homescan. The problem I have is I cannot map the store numbers in ?charstore? data set to the TDLink number in char_store_detail1 data set. I wonder if you could give me some hints on that?"

A. [*testily from MK*] Please don't send me random undocumented data files you have acquired from IRI at some point in IRI's existence (*or, in this case, from Nielsen's Homescan and Nielsen's TDLinx services*) and expect me to figure out what's in them and how they relate. When you received the files, you received them from someone and they had some documentation.

Data Handling Questions

Q. There's a lot of files here. Are the any utilities to aid in assembling the data I need?

A. Bart Bronnenberg has posted some SAS scripts here; see http://groups.google.com/group/IRI_Marketing_Data_Set/web/sas-scripts

Q. I need something simpler that Bart's scripts to start -- for example, how do I even take a look at a file this big?

A. See Starting out hints.zip .

Data Set Policy Questions

Q. Does IRI plan to continue this program?

A. IRI plans to continue this program, although this is not guaranteed. Per our plan, we added a 6th year of data (2006) in early 2009, and plan to add more. 2007 was delayed, but will be available before the end of 2010. Long term data sets should be of great value to the field.

Terms of Use Questions

Q. Why are there restrictions on the data?

A. IRI is a commercial business. We have been non-profit in the past, but not intentionally.

IRI's aim is to stimulate research in the marketing of consumer packaged goods, and therefore perhaps expand its market. IRI is in the business of selling data. Therefore, the data are not to be used for commercial purposes and the most recent two years are not provided. We also purchase the data we sell from retailers, and this results in further restrictions. For this reason, the retailers are not named (e.g. they are identified as "chain 1" or "chain 3"). IRI wants to "review the final results... solely to confirm that IRI Data is being represented in a non-misleading fashion and otherwise in accordance with the requirements of this Agreement. It is not IRI’s interest or intent to provide any editorial control of the outcome of the Research Project".

If IRI feels that expanding this data set into future years would hurt, rather than help, its long-term interests, then release of subsequent years will be less likely. The restrictions are there to protect IRI's interest, and to make it more likely that this project will continue for a long time.

Q. Why is IRI charging for this data?

A. Note we're not selling the data, we're making the data available for a nominal fee for academic research purposes. We're also trying not to ship free USB hard drives to everybody who asks for one. IRI also has certain costs related to the update of the database and we'd like to recover some of these.

ISMS has a grant program to help. As of October 2010, IRI is not aware that the costs of this data set have been any impediment. We know of no case in which ISMS has issued a grant.

Q. What if there are two of us working on a project?

A. Both of you should sign the NDA. We will charge you one fee and ship one drive.

Data Definition Questions and Answers

Q. What counties are in which IRI markets?

A. The file "fips by IRI market.xls" has been added in the "Files" area to provide this information. It is included in the standard data shipment beginning with ADS_80.

Q. I see unit sales and dollar sales, but I need volume sales!

The detailed data files just have units and dollars. The stub files contain the equivalent volume which is applicable to each UPC. See VOL_EQ in section 4.8 of the documentation.

A sample from the ready to eat cereal product listing is below. Note Kix 12.75 ounce box has a volume equivalent of .7969 pounds.

L5	L9	Level	UPC	VOL_EQ
GENERAL MILLS KIX	+GMKIX BR_BR CRN BX 12.75OZ	9	00-01-16000-62680	0.7969

GENERAL MILLS KIX	+GMKIX BR_BR CRN BX FJC 15.3OZ	9	00-01-16000-86410	0.9563
GENERAL MILLS KIX	+GMKIX BR_BR CRN BX FJC 18OZ	9	00-01-16000-62640	1.125
GENERAL MILLS KIX	+GMKIX BRY CRN BX SWTND 11.5OZ	9	00-01-16000-83360	0.7188
GENERAL MILLS KIX	+GMKIX REG CRN BX 9OZ	9	27-01-00710-72125	0.5625
GENERAL MILLS KIX	+GMKIX REG CRN BX 9OZ	9	00-01-16000-66310	0.5625
GENERAL MILLS KIX	+GMKIX REG CRN BX 10.8OZ	9	00-02-16000-86400	0.675
GENERAL MILLS KIX	+GMKIX REG CRN BX 18OZ	9	00-01-16000-62570	1.125
GENERAL MILLS KIX	+GMKIX REG CRN BX 31.5OZ	9	27-01-04200-67890	1.9688
GENERAL MILLS KIX	+GMKIX REG CRN BX L/SGR 36OZ	9	00-01-16000-88490	2.25
GENERAL MILLS KIX	+GMKIX REG CRN BX SWTND 9OZ	9	00-01-16000-68279	0.5625
GENERAL MILLS KIX	+GMKIX REG CRN BX SWTND 13OZ	9	27-01-00710-71837	0.8125
GENERAL MILLS KIX	+GMKIX REG CRN BX SWTND 13OZ	9	00-01-16000-66370	0.8125
GENERAL MILLS KIX	+GMKIX REG CRN BX SWTND 15.5OZ	9	00-01-16000-86730	0.9688
GENERAL MILLS KIX	+GMKIX REG CRN BX SWTND 27.5OZ	9	00-01-16000-60440	1.7188

Q. Does the IRI Academic Data Set capture information on out-of-stock brands or stockouts?

A. The data are sales data at store/week/UPC level. There is a record when there is movement. There is no direct measure of out of stock situations.

Q. **is there any chance that you have the information about order quantity or inventory level?**
Could you provide me the name of the companies so that I could check their financial statement?

A. IRI has no data on the order quantity (how much the retailer orders from the manufacturer for replenishment) or inventory level.
The names of the product manufacturers for each product are included in the product stub files.
Our contracts with retailers do not allow us to name the retailers in the data set.

Q. I'm interested in the issue of multipack. It seems to me that the soda dataset does not contain information on whether a soda item contains, say, 6 or 24 cans. Did I miss something?

A. The files **prod_beer_sz.xls** and **prod_carbbev_sz.xls** in the parsed stub files directory contain this additional information. It is described in the documentation beginning with version 1.314. These files are included in the data sent out by IRI beginning with drive #39, and are available in zipped form in the files area for those who have earlier drive numbers.

Similar information for mustard and ketchup and yogurt is now available on this web site and will be put in the drives beginning with drive #100. http://groups.google.com/group/iri_marketing_data_set/web/prod_mustketc_sz.zip

Q. In IRI data set, there are only sales (units & $) data. Can I know the regular price for products?

A. We receive units and dollars from retailers, which determine the price that week. The regular price is not sent.

IRI has proprietary algorithms to determine regular price (some versions known as base price or baseline price), These are IRI's intellectual property and are not being shared.

An earlier version of this algorithm was published in **Marketing Science** in 1993: "An Implemented System for Improving Promotion Productivity Using Store Scanner Data**",** Magid M. Abraham, Leonard M. Lodish volume 12, #3, pages 248-269.

Q. I am curious about the mechanism of "missing" field in the attribute information for UPCs. Is it missing because for some UPCs there are no value for that attribute (in the case of beer, where some UPCs have flavor_scent (lemon, citrus etc) while others by nature don't) or because there should be a value, it is just not observed (in case of coffee->form, where the product has to take a shape of either bean, ground, pod, etc, but it is recorded missing because it is not observed)? I guess what would help is to know how IRI identifies SKUs from a list of UPCs.

A. The value of "missing" means exactly that -- that there is no known value.

This most commonly arises as you've indicated:

1. The item has the most common value in the category, the one that everyone assumes when they see the category unless they see something different and which is therefore not specified on the package.

2. The item did not hang around the market long enough to be fully identified. Identification is a triage; when I last dealt with this problem about ten years ago we were getting in 35,000 new UPCs per week (US only).

3. The attribute is fairly new. Missing is the default when an attribute is added.

4. The attribute can't be observed from outside the package. In these cases, only leading items are likely to be coded.

There's no magic bullet on the SKU definition. In general, each UPC active at a particular time is a different SKU. The problem can then be seen as determining whether there are products which should be combined -- there may be multiple UPCs due to multiple plants, attempts to track some promotional campaign, UPCs being changed due to merger and acquisition, etc.

Some of the worst issues arise in packaging: depending on how the product was identified, we might know that it is a bottle, a glass bottle, a returnable glass bottle, a nonreturnable glass bottle, a brown glass bottle, and so forth. Obviously, there's a hierarchy of information here ("returnable glass bottle" is a subset of "bottle"), but this may not be known from the information available.

Q. I was just wondering about the figures in cent998 and cent999 fields in the trips files -- many are shown as fraction of cents. I'm not sure why this is, unless the products are priced in fraction of cents. Could you let me know the reason for this?

A. There are a couple of reasons, although the net advice is to round to the nearest cent and ignore this.

First, there will be a bit of this if the price is 3 for a dollar -- best description of the individual unit price is $0.3333333 (as far out as we carry the decimal).

Second, there's floating point conversions. We looked at a sample record and in dollars it shows up as $22.95 but in cents is 2294.64. There is floating point storage in the IRI system for fields that might turn out to be big under some circumstances (such as total records). Note you are seeing the panelist totals for a trip, but this same field is used for the store totals (for a week, across all items in a store), a number that might be over a million dollars if the store had sales of $40,000,000 in a single year.

Third, years 1 and 2 will differ slightly from years 3 onward because years 1 and 2 were "too old" to pull using the standard process when we pulled them. There was an additional SAS step involved. A lot of decimals kept hanging around in that process (again, due to floating point conversions, but in this case due to additional conversions across operating systems). This was visually unfortunate, but doesn't affect the data.

If you are unfamiliar with "floating point conversion issues" ask the oldest person in your computer lab.

Q. **Two quick questions on stubs.**

(1) What is the "PLU SOFT DRINKS" in the L2(small category) field of carbev category?

(2) What does "all ~ products" in L5 field mean? (for example in carbev there is "all coke products" on top of just plain Coke.) Is it any different than just plain Coke? Does it include different type of products in one UPC?

A. Some retailers sell some UPC coded products under PLUs (Product Look-Up codes).

Most produce is sold this way (that's the number on the little sticker on apples -- a 4015 means these are medium red delicious apples).

In some cases, products that have a UPC code will be sold this way for the retailer's convenience. This is particularly common in carbonated soft drinks. The following two cases illustrate why:

Case 1: Single bottles or cans are sold based on splitting larger sizes, such as 6 packs. In these cases, the single sold at $.75 may have a PLU sticker placed over the UPC code to distinguish it from the 6 pack, sold

6 cans for $2.29. In the worst case, all 12 oz cans of soda may have the same PLU, regardless of manufacturer.

Case 2: 24 packs of cans may be heavy to lift, they may reflect in-and-out product that's not in permanent distribution, and all Pepsi products [or all Coke products] may have the same pricing. In these cases, the retailer may designate a PLU number to be used for "All Pepsi 24 pack cans" this week at $5.99.

Note in case 2 the PLU made be used for some weeks, and in other weeks the regular UPC may be used.

Where it is possible to determine what the PLU is, IRI does so. In some cases it is not. That's why you see "ALL COKE PRODUCTS" in L5 -- this may include Coke, Diet Coke, Mello Yello, Sprite, etc.

Each retailer will use a different PLU code (and this may vary by week), IRI converts these to a single arbitrary number.

PLUs are not a plus in this data. PLUs on UPC coded items are a nuisance. They slow down IRI's delivery, they are expensive, and they make the data harder to use.

Q. What's the relationship between the UPC I see on a package and the UPC IRI is providing?

A. In the US, most packages will show a 12 digit UPC that is S-VVVVV-IIIII-C where S is the system, V is the vendor, I is the item and C is a check digit to verify the code is read properly.

[The vendor is really determined by the combination of S and V; the S field is really two digits in the international version, the 10 digits of VVVVVIIIII are not always 5 and 5, the full code has been expanded to 14 digits to include such things as pallet codes -- but what I'm providing here is sufficient.]

The IRI UPC fields provided are System, Vendor and Item. The IRI version number, called generation, allows for re-use of UPC numbers over time, so what was a floor wax in 1990 might be a dessert topping now. We do not provide the check digit, but it can be algorithmically calculated, as shown in http://en.wikipedia.org/wiki/Universal_Product_Code#Check_digit_calculation .

On smaller packages there may be only a 6 digit code, called zero-suppressed UPC-E. IRI expands this to its 12 digit equivalent. For how this is done, see http://en.wikipedia.org/wiki/Universal_Product_Code#Zero-compressed_UPC-E

Q. What are the UPC codes not seen in nature in the data, such as those with system code 27 and system code 88?

A. These are UPCs generated by IRI to replace the actual UPC. There are a variety of reasons for this. First, PLUs (see question above). Second, to allow reporting of private label, private label UPCs for the same type of product, are pooled into a single UPC, such as 2 liter diet cola in a plastic bottle. Third, for some retailers in some cases we are provided SKU information rather than UPC information. These SKUs are converted to an IRI generated UPC. Fourth, system 4's are converted, because they are only unique within a retailer.

Q. In my analysis, I want to exclude purchases that were wanded. To do this, I need to identify those stores which card+key panelists have to wand their purchases at home after the trip. Is it safe to say that all IRI_keys in "manual store entry external.csv" are nonparticipating stores?

A. Yes. In addition, note that wand transactions will have both a "98" and "99" entry in the trips file.

Q. The panel data seems to include a number of stores (iri_key) that I can't find in the store-level files or the Delivery_Stores files. Here are a few, which all seem to be 7-digit iri_key's that begin with "99": 9918685 9979660 9979663 9979673 9979872. Any suggestions on this?

A. See Section 4.6 of the documentation, describing the file "manual store entry.csv".

Q. I'm trying to calculate a price per unit, but a number of observations have zeros as an entry for the units variable. How shoudl I interpret these observations?

A. These are hiconed observations. Hicones can be a problem, particularly in beer and soft drinks. They arise because multipacks (e.g. 6 packs) can be sold/scanned as 6-12 ounce units for $.50 each ($3 total) or as 1-72 ounce unit for $3.

Usually, multipacks which are broken up as singles carry a PLU code (e.g. a sticker over the UPC) so they can be scanned as a single. But this does not always occur, which can result in one-sixth of a unit being sold. The units field in the academic data set is carried as an integer and these parts were lost.

An examination of the 10,165,789 store records for the beer category in 2004 showed 4663 records (.00046) with zero units. Over half of there were from 15 stores, which is as expected because this is a store policy issue.

These packs are called hicones because the plastic rings that link the cans were first provided by the Hi-Cone Corporation.

Q. In the panel data there are purchase records in which units are non-integer. Does this happen when customers break the bulk, for example, buy one can of beer out of 6-pack which is registered as one UPC?

A. Yes, see explanation of hicones in previous question.

Q. I saw somewhere in the documentation that the chain code changes from year to year and this got me concerned that perhaps the store code (IRI_key) also changes from year to year. Can you confirm whether or not this is the case? Also, I am correct that market number does not change from year to year, right?

Store code (IRI_Key) does not change. Market number does not change. Panelist numbers do not change. The chain code is discussed in section 5.2, on chain cross-reference.

Q. Chain code=NONE?

I came across a problem when merging the "delivery store" file with the "Y4_panel trip" file. Specifically, there are some iri_key in the "delivery store" file with "mskdname" listed as "NONE". Yet, a large amount of panelist in Eau Claire and Pittsfield visited those stores with the "NONE" channels. I checked the "manual store entry" file but could not find the corresponding information. For example,

```
  iri_key   ou   est_acv   market_n~e   open   clsd   mskdname
 ------------------------------------------------------------
228037   GR   9.877998   EAU CLAIRE    435   9998      NONE
```

In the BehaviorScan markets of Pittsfield and Marion we have included independent stores which do not belong to a chain. Store 228037 is an independent ***GR***ocery store.

Q. What's I found out a store (e.g. 238522) that is in market HARTFORD and market PITTSFIELD. Is it the same store?

A. Same store. Not an error. Pittsfield is an original BehaviorScan market from 1979. The Hartford market was added later as an InfoScan market (1988) and includes this county. This county is assigned to the Pittsfield market in the BehaviorScan service and the Hartford market in the InfoScan service.

Q. What'sthe easiest way to work with the UPC code? For example, how do I compute COLUPC in the panel files from SYstem, GEneration, VENDor and ITEM fields in the stub file?

A. One trick is to convert the entire set of fields into a number. For example, the collapsed UPC (COLUPC) can be computed as

*LET colupc = sy*100000000000+ge*10000000000+vend*100000+item*

This assumes that your software will handle this large of an integer, or a float with this many significant digits.

Examples:

sy	ge	vend	item	Collapsed_UPC
0	1	12000	230	11200000230
0	1	70470	309	17047000309

Q. The format of 2 of the panel files, margbutr_PANEL_GR_1114_1165.dat and margbutr_PANEL_GR_1114_1165.dat, changes at the end.

A. This is a glitch that was fixed beginning with drive #87. For drives below this number, there's replacement files in the files area, compressed as margbutr_PANEL_GR_1114_1165.zip and margbutr_PANEL_GR_1166_1217.zip. For more information see http://groups.google.com/group/IRI_Marketing_Data_Set/browse_thread/thread/96183d80c78c60c1/9b5909c30235c8ce?hl=en#9b5909c30235c8ce

Q. We can see that during the week 1280-1313, the total number of trips in store 213290 and store 648764 were incredibly small. Would this be due to transmission losses ?

A. These are two stores of one chain. The card data was not transmitted and was not recovered later as backdata.

Q. If a product is defined as private label, is the manufacture company masked or the retailer company masked? For example, a Kroger ketchup is produced by company A. Does the private label in column parent company and vendor in file prod_mustketc.xls mean company A or Kroger?

A. The retailer (Kroger in this example) is masked. IRI doesn't track who manufactures the private label. I don’t know anyone who does. I’m sure private label manufacturers keep some track of who has the current contracts for what (in order to aid their sales efforts), but I don’t know of any regular source for this on a current basis, let alone a back-data basis. If you find one, please let me know.