Introductions

Data Sponsor Team

- Laura

- Will

- Amber

- Spencer

WCAI Team

- Ben Adams

- Elea McDonnell Feit

- Melissa Hartz

A Few Operational Notes

This webinar is intended for researchers & scholars. It should not be published or presented without permission from WCAI.

- Questions on program: wcai-research@wharton.upenn.edu

- Press contacts: wcai-press@wharton.upenn.edu

To get access to the data, research teams should submit a proposal to WCAI and the sponsor for approval. More details at the end of this presentation.

This research opportunity is sponsored by a Fortune 500 Specialty Retailer. While awarded teams will be introduced, the sponsor must not be named in publications.

What Defines a Project?

The data sponsor sells products that customers use in complex and/or multi-step "projects". Most of the projects require purchase of multiple products over time.
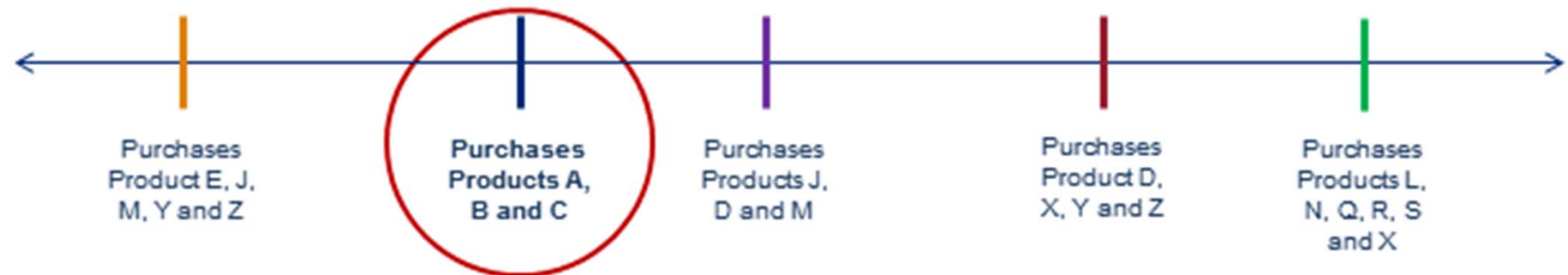
What Defines a Project?

The data sponsor sells products that customers use in complex and/or multi-step "projects". Most of the projects require purchase of multiple products over time.

A project may be identified by certain products being purchased together, such as Products A, B, and C.

What Defines a Project?

The data sponsor sells products that customers use in complex and/or multi-step "projects". Most of the projects require purchase of multiple products over time.
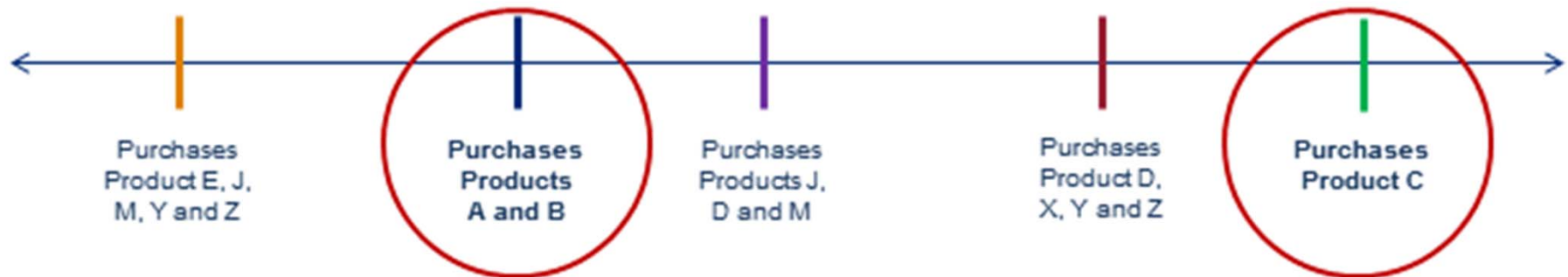
A project may be identified by certain products being purchased together, such as Products A, B, and C.

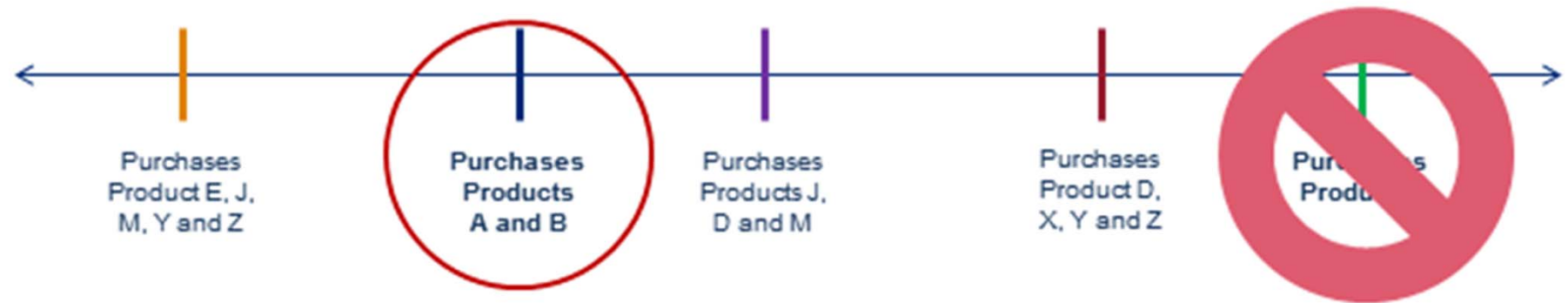| Purchases Product E, J, M, Y and Z | Purchases Products A, B and C | Purchases Products J, D and M | Purchases Product D, X, Y and Z | Purchases Products L, N, Q, R, S and X |

However, detection isn't always so clear.

Business Problems

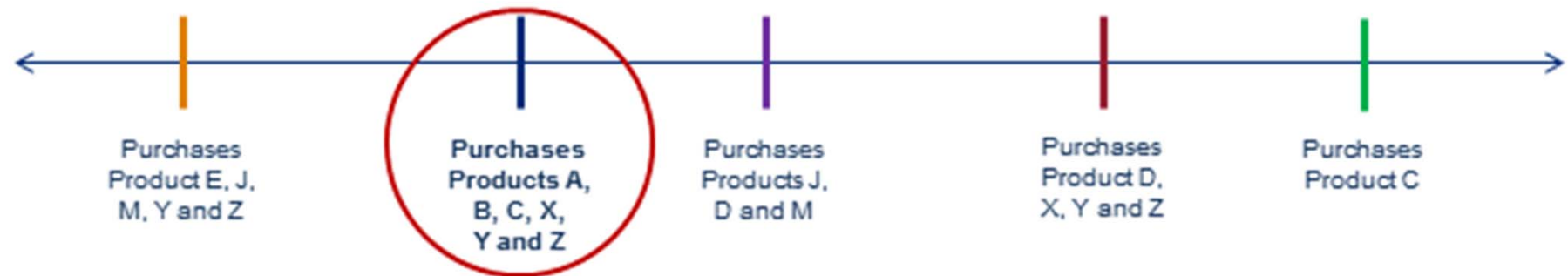**A customer might purchase Products A and B together, but buy Product C several days after.**

| Purchases Product E, J, M, Y and Z | Purchases Products A and B | Purchases Products J, D and M | Purchases Product D, X, Y and Z | Purchases Product C |

Business Problems

Or, the customer may never buy Product C from the data sponsor: possibly purchase it from a competitor, or never.

Purchases Product E, J, M, Y and Z

Purchases Products A and B

Purchases Products J, D and M

Purchases Product D, X, Y and Z

Pur... s Produ...

WHARTON CUSTOMER ANALYTICS INITIATIVE

Wharton UNIVERSITY *of* PENNSYLVANIA

Business Problems

Additionally, the products that comprise a project is latent. We do not know exactly *which* products define a particular project – so Products A, B, C, X, Y, and Z might make up a project, not just A, B, and C.

| Purchases Product E, J, M, Y and Z | Purchases Products A, B, C, X, Y and Z | Purchases Products J, D and M | Purchases Product D, X, Y and Z | Purchases Product C |

Key Research Question

How should retail transaction data be analyzed or conceptualized when customers may be engaged in "projects"?

Wharton
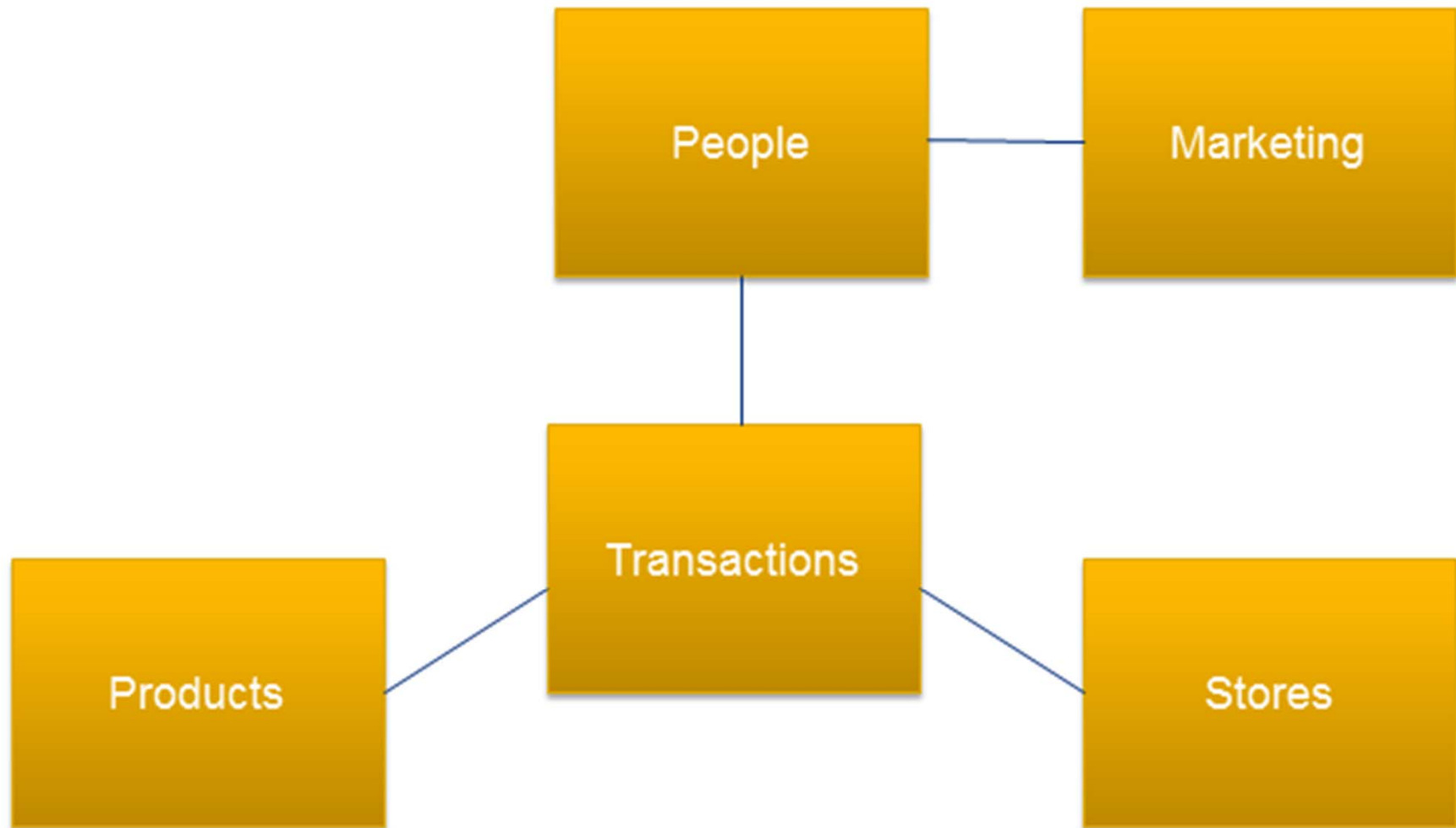UNIVERSITY of PENNSYLVANIA

Key Research Question

How should retail transaction data be analyzed or conceptualized when customers may be engaged in "projects"?

- Can we determine (from the purchase histories) sets of products that seem to be purchased together as part of a project (not just in the same transaction). Can this information be used to construct a "project materials list"?

- Is it possible to detect when a customer is engaged in a project, based on shopping behavior alone?

- Can you predict the customer's next trip or, more importantly, the next big project? Is there a natural order to projects?

- Can customer groups be targeted for offers or promotions based on their prior behavior or current project(s)?

- Can we identify what products or services involved with a project are being purchased elsewhere? Can we determine share of wallet?
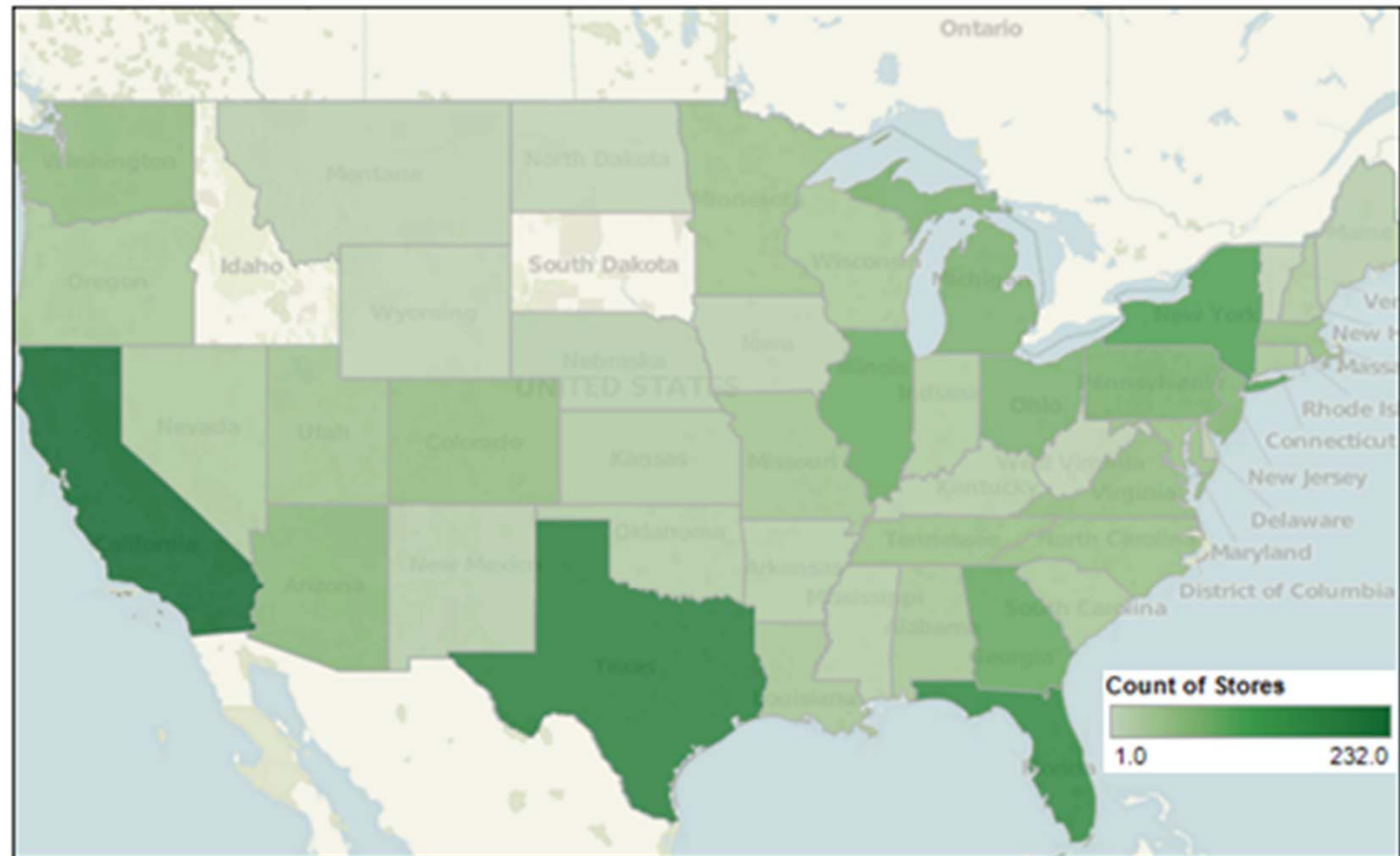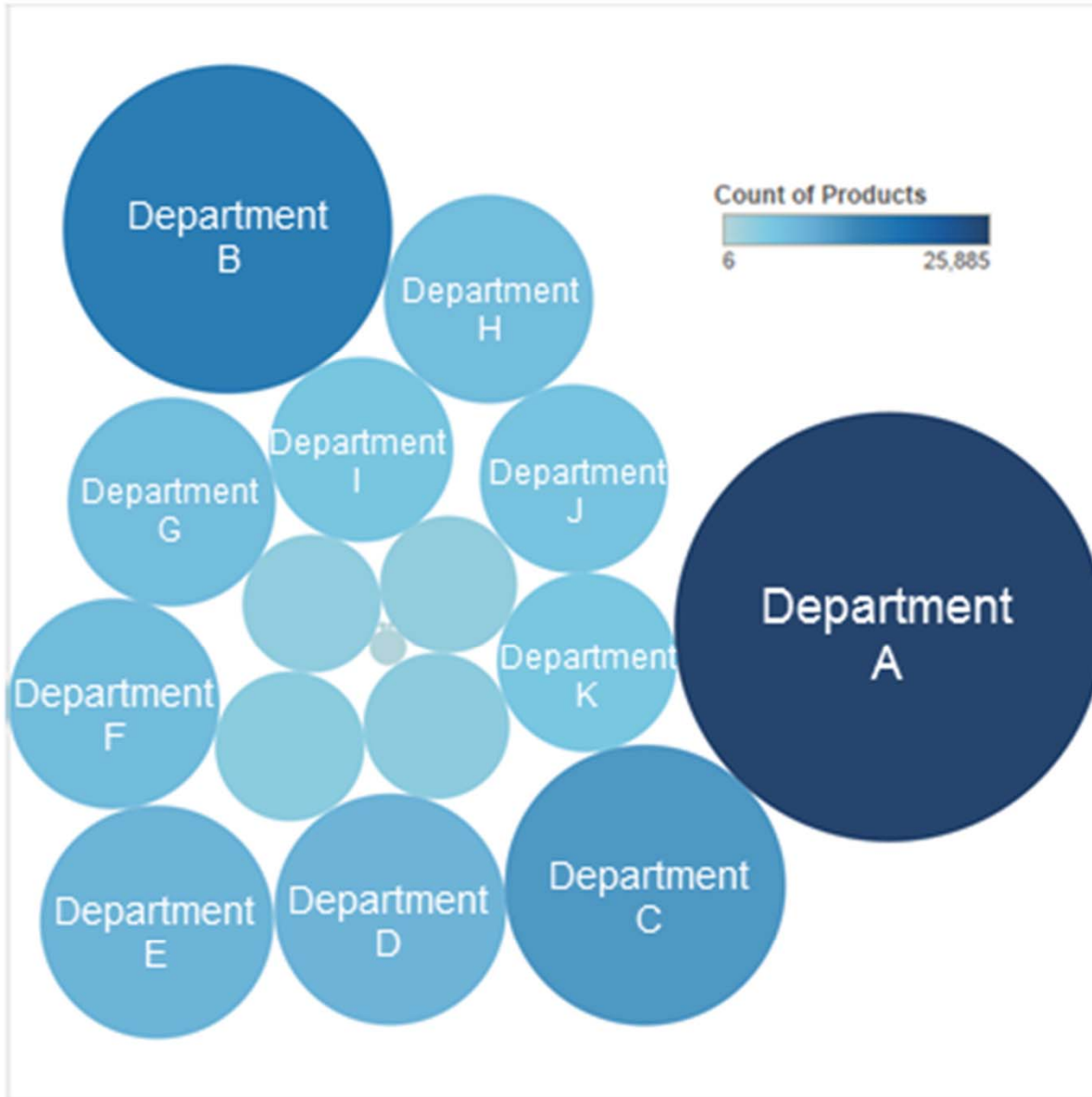
The Data

The Stores

There are 1,982 stores in the U.S.

Data includes
- Store name
- Store open date
- Metro location
- Managerial region

Count of Stores
1.0 — 232.0

The Products

~120,000 Products

Four levels of product hierarchy/organization:
- Department
- Category
- Sub-category
- Product

The dataset contains:
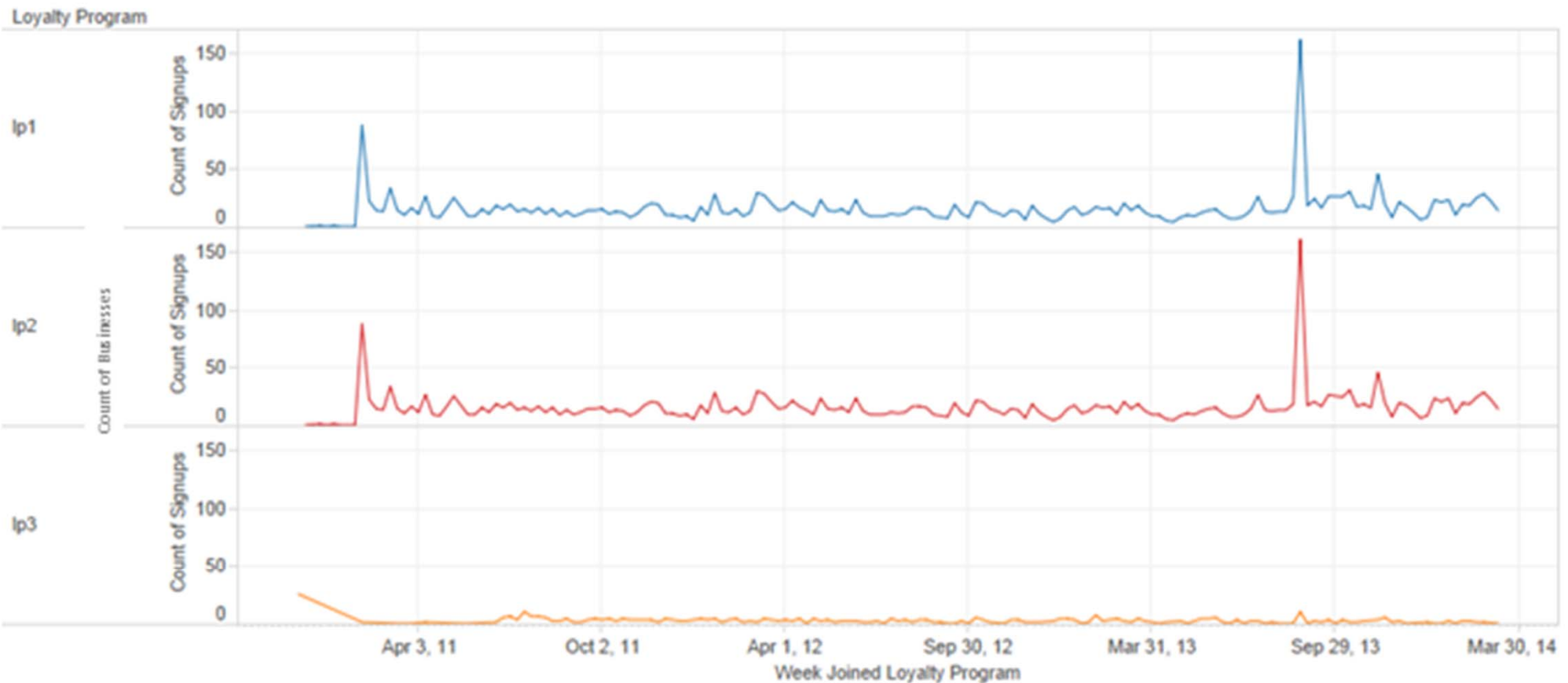Detailed, descriptive names and IDs for each level

The People

3 samples

- Commercial Purchasers
- "Known" Consumer Households
- Randomly Sampled Customers

WHARTON CUSTOMER ANALYTICS INITIATIVE

Wharton
UNIVERSITY of PENNSYLVANIA

The People

3 samples

- **Commercial Purchasers**
- "Known" Consumer Households
- Randomly Sampled Customers

- o **20,000** companies
- o Tend to be high-volume purchasers
- o Dataset includes date of signup in one of three loyalty programs, if applicable

Loyalty Program 3 is a pilot program, offered in limited geographic areas.

The People

3 samples
- Commercial Purchasers
- **"Known" Consumer Households**
- Randomly Sampled Customers

- **20,000** households
- Confirmed demographic data
- Become "known" when customer provides an address

Estimated Household Income
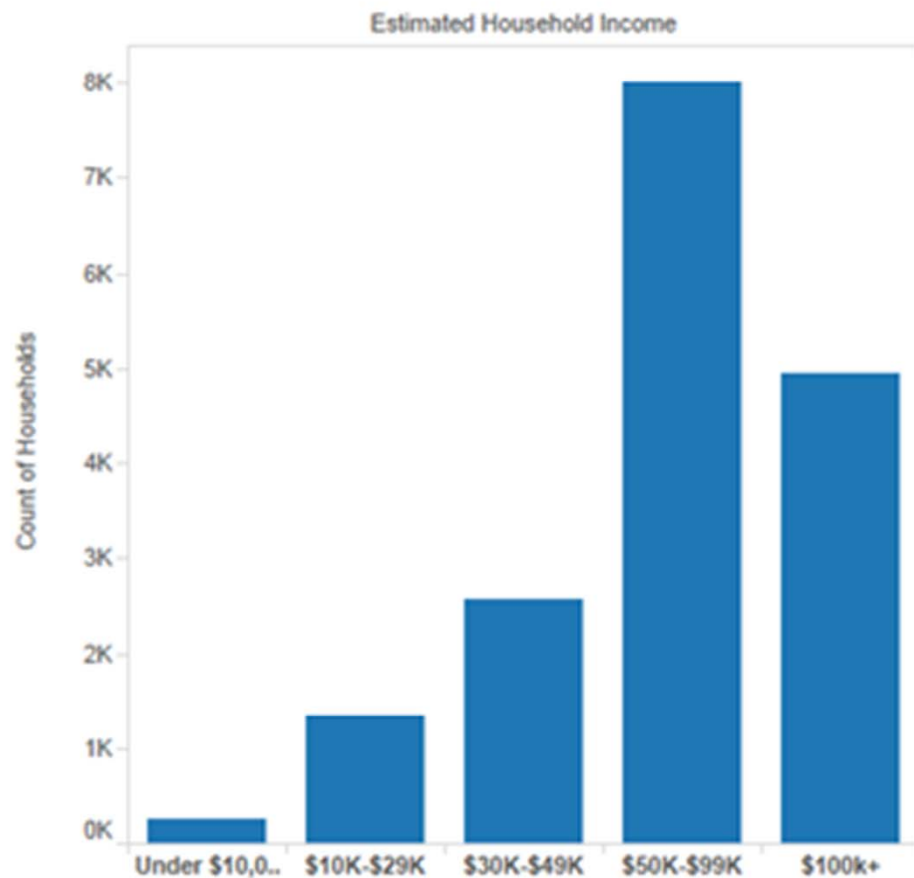
The dataset includes:
- age & gender
- ethnicity
- estimated household income
- estimated net worth
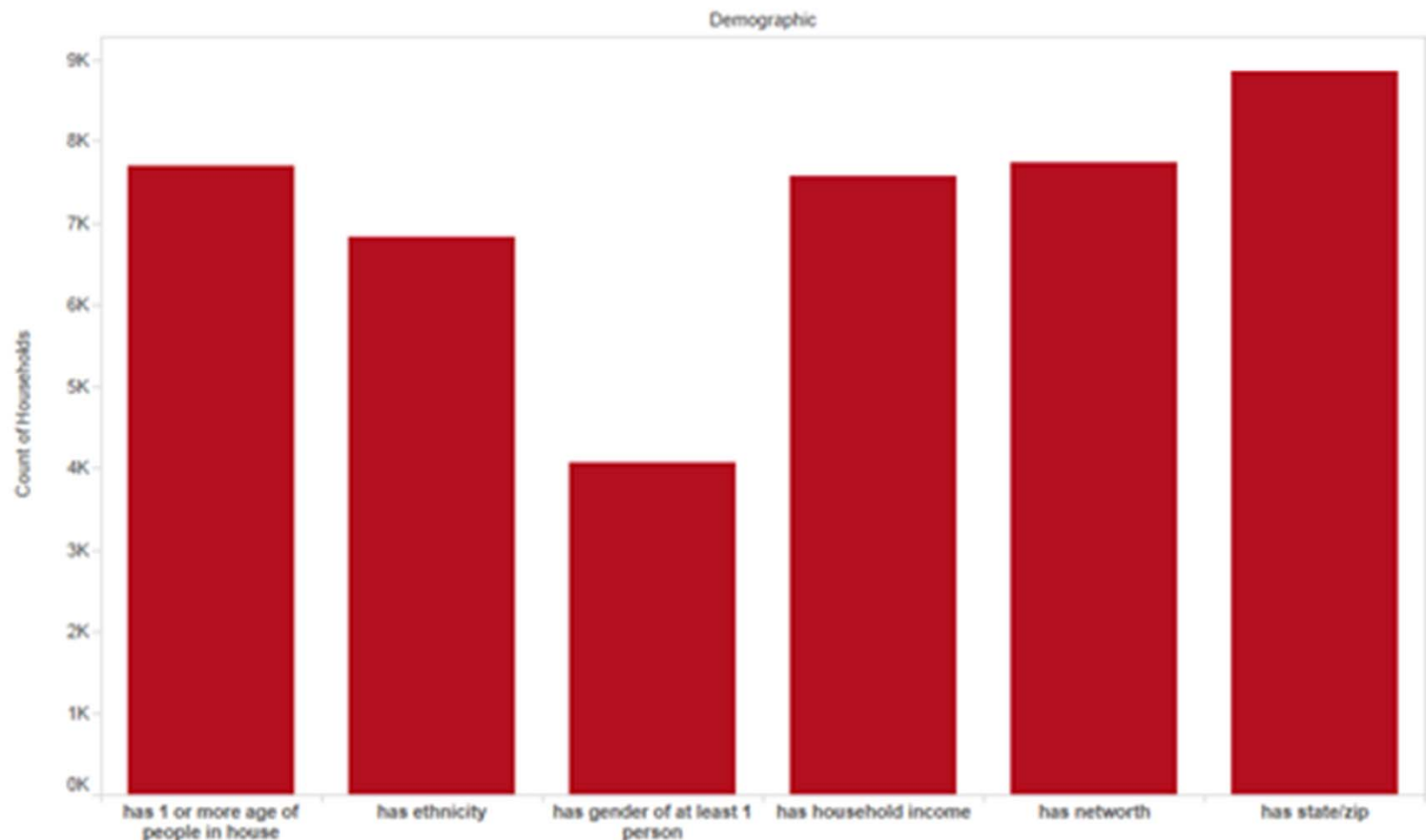- postal code
- additional product-related information

WHARTON CUSTOMER ANALYTICS INITIATIVE

Wharton
UNIVERSITY of PENNSYLVANIA

The People

3 samples
- Commercial Purchasers
- "Known" Consumer Households
- **Randomly Sampled Customers**

- o **20,000** randomly selected customers
- o Could be commercial, known, or unknown

If any data from business or consumer is available, it is recorded with the household/ customer ID

Demographic

Count of Households

has 1 or more age of people in house · has ethnicity · has gender of at least 1 person · has household income · has networth · has state/zip

Marketing

The data contain:

- Individual email marketing for addressable customers in the dataset (about 25%)

- ~3,300 campaigns from March 2012 – March 2014

- Also includes:
  - subject line
  - delivery date
  - campaign name
  - ID of household/ company that received email
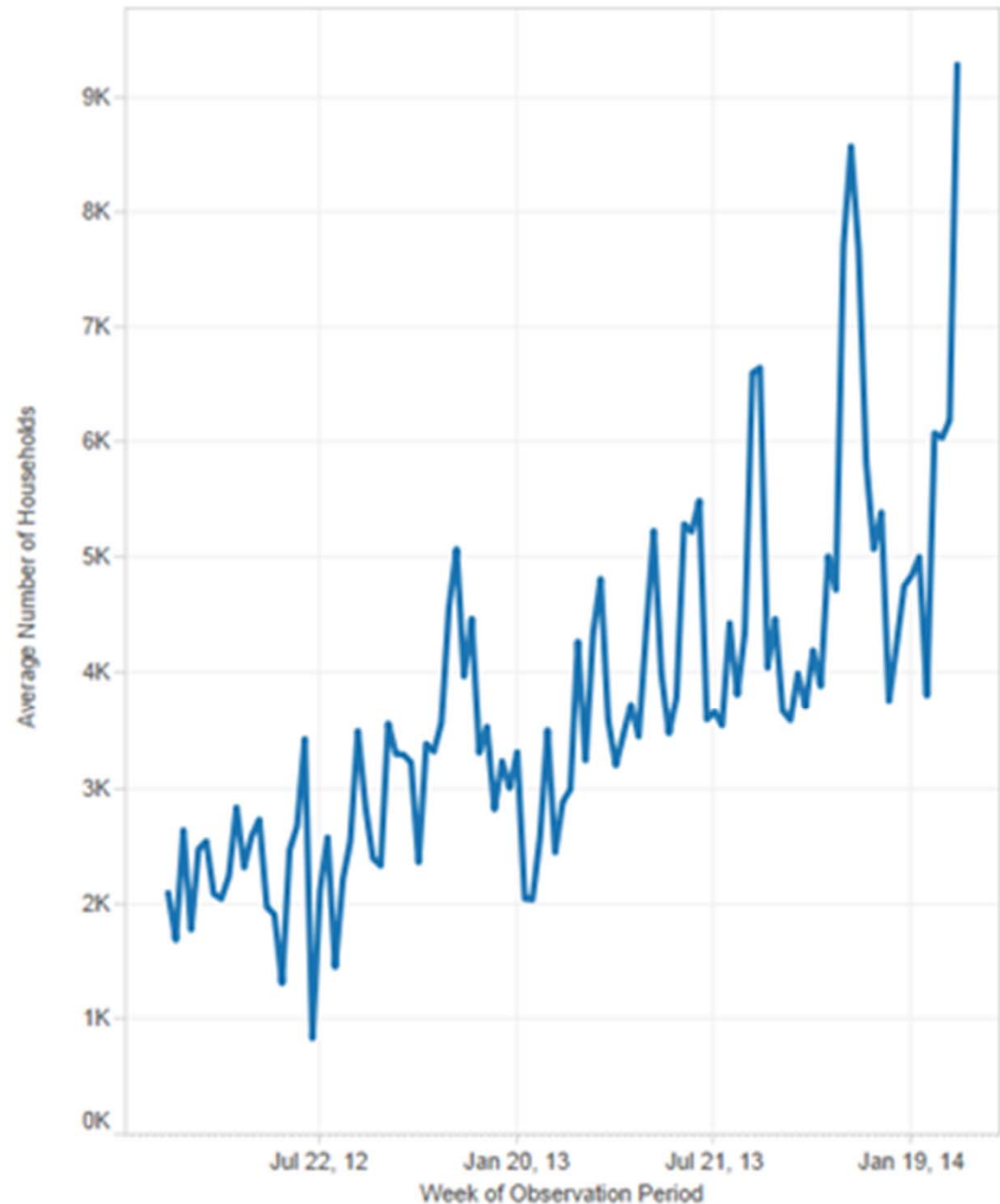
The Data

The Transactions

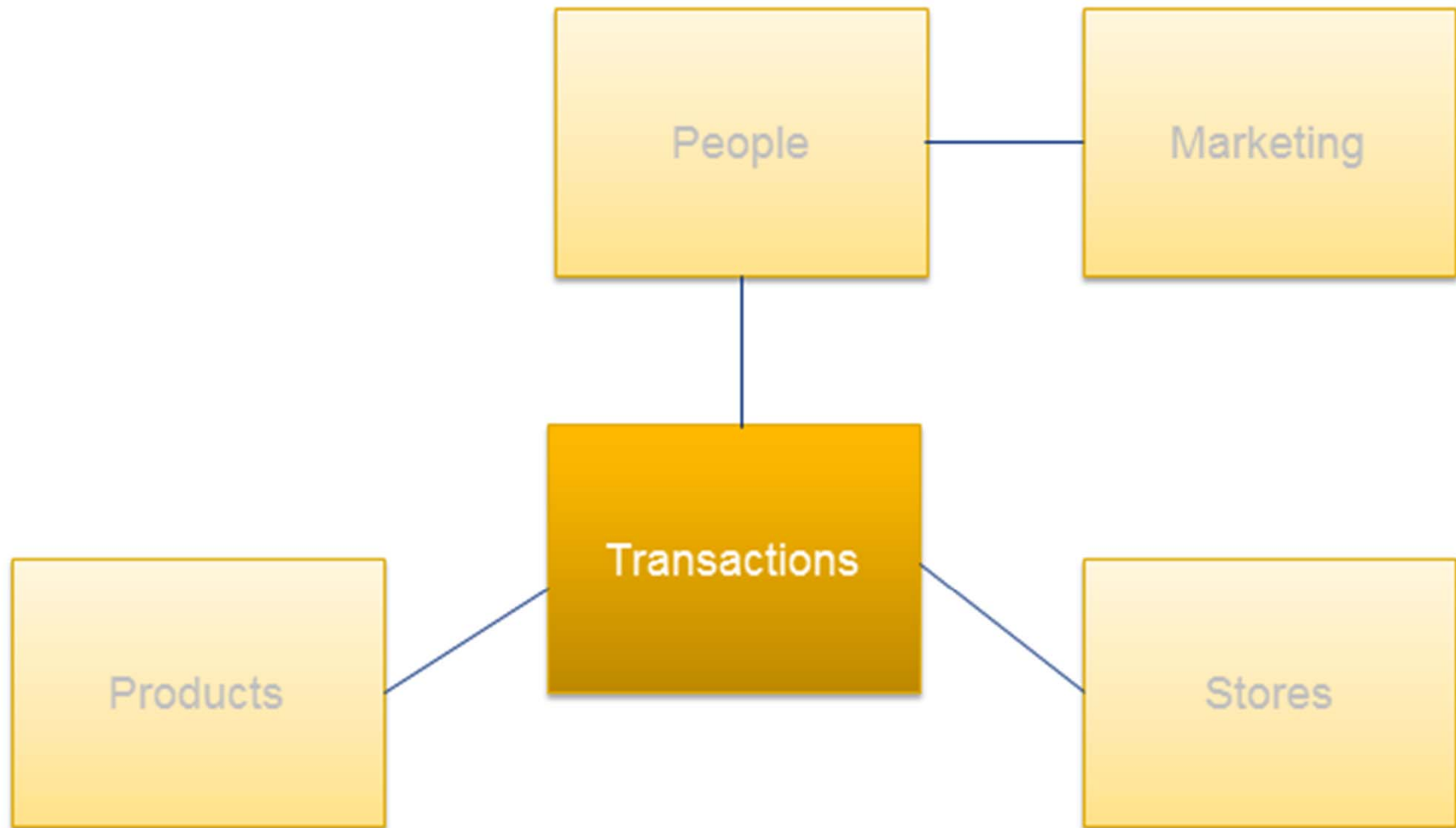**The dataset includes ~2.5 million transactions over 24 months. This data is at the individual item level.**

The dataset includes:
- Item price and quantity
- Total sale amount
- Date and time of purchase
- Purchase channel
- Any discounts, sorted by discount category

This data also includes returns as negative sale amounts.

WHARTON CUSTOMER ANALYTICS INITIATIVE

Wharton
UNIVERSITY of PENNSYLVANIA

The Transactions

The dataset includes ~2.5 million transactions over 24 months. This data is at the individual item level.

- Each transaction has, on average, **4.53 products**
  with StdDev of 4.64

- **73% of the transactions contain 5 or fewer products**

The Transactions

The dataset includes ~2.5 million transactions over 24 months. This data is at the individual item level.
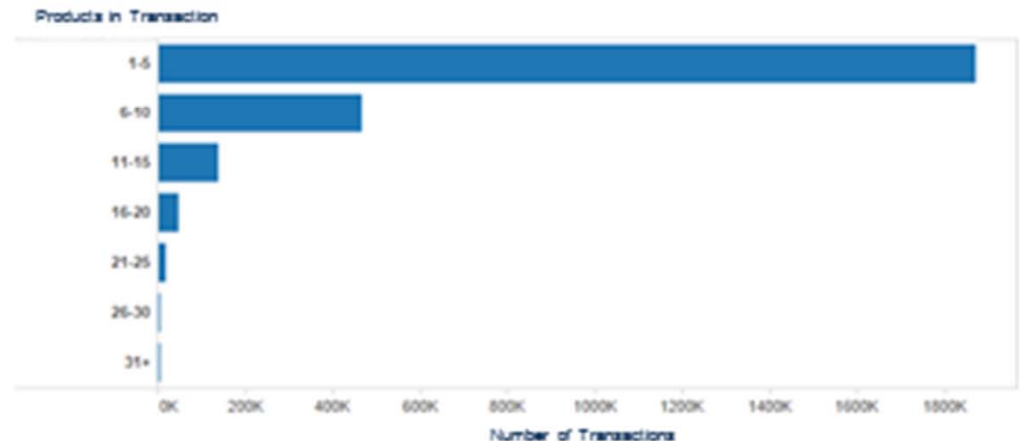
- Each transaction has, on average, **3.65 products**
  with StdDev of 3.81

- 73% of the transactions contain **5 or fewer products**

- Transactions containing more than 2 different products have an **average price of $125.02**
  with StdDev $375.58
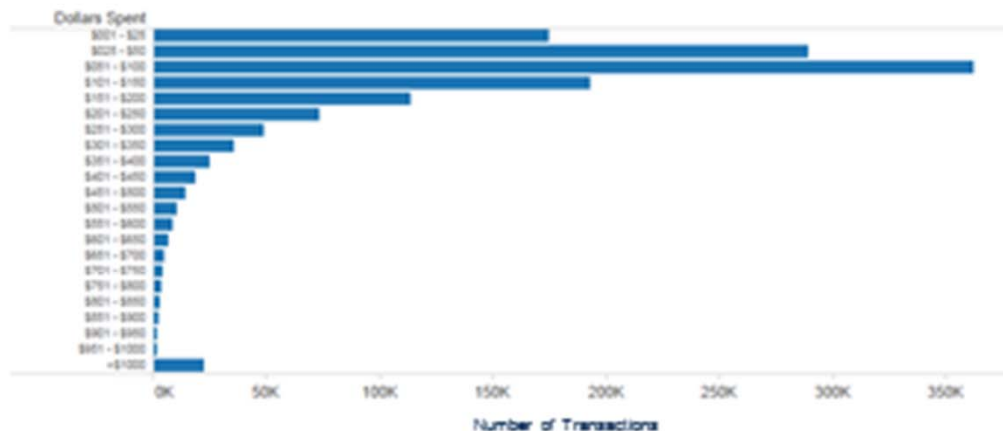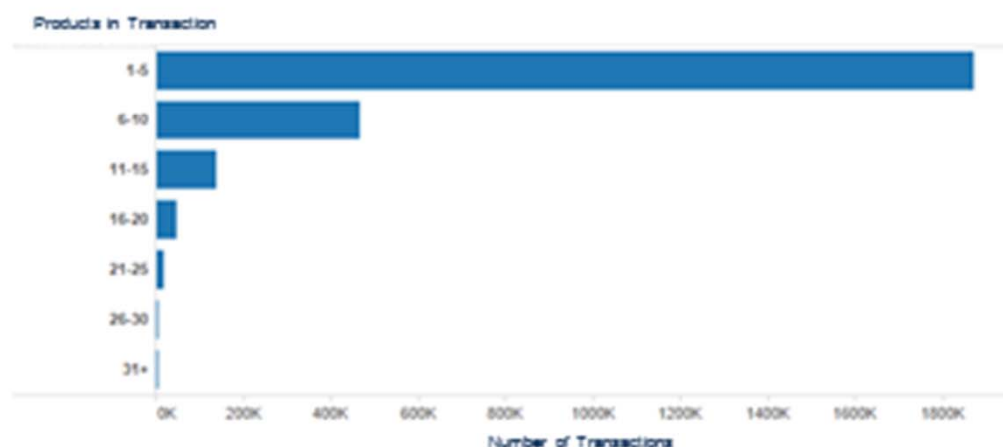
Total Spend per Week by Sample Group
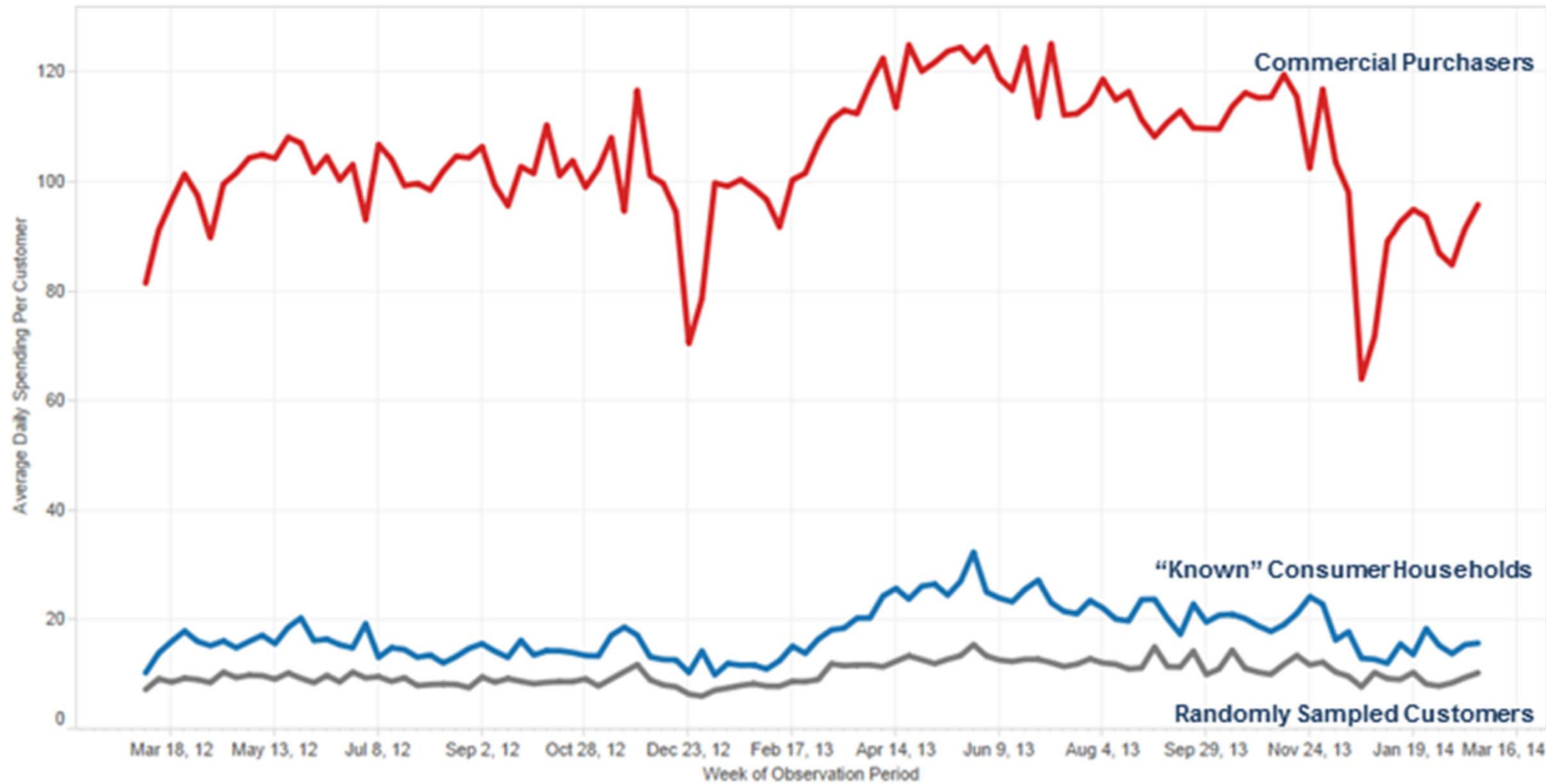
Purchases Over Time – Bob
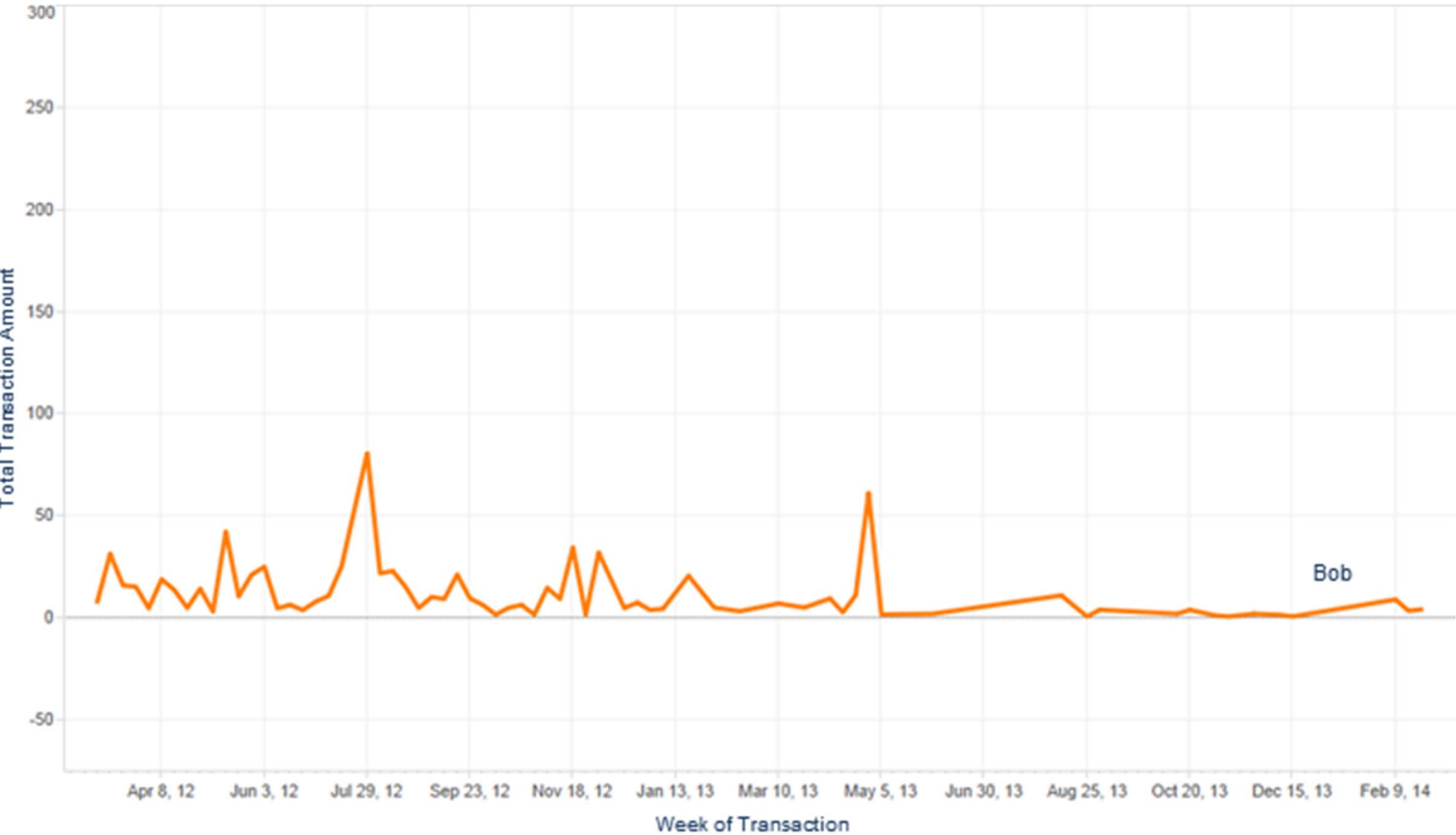
Purchases Over Time – Bob and Joan
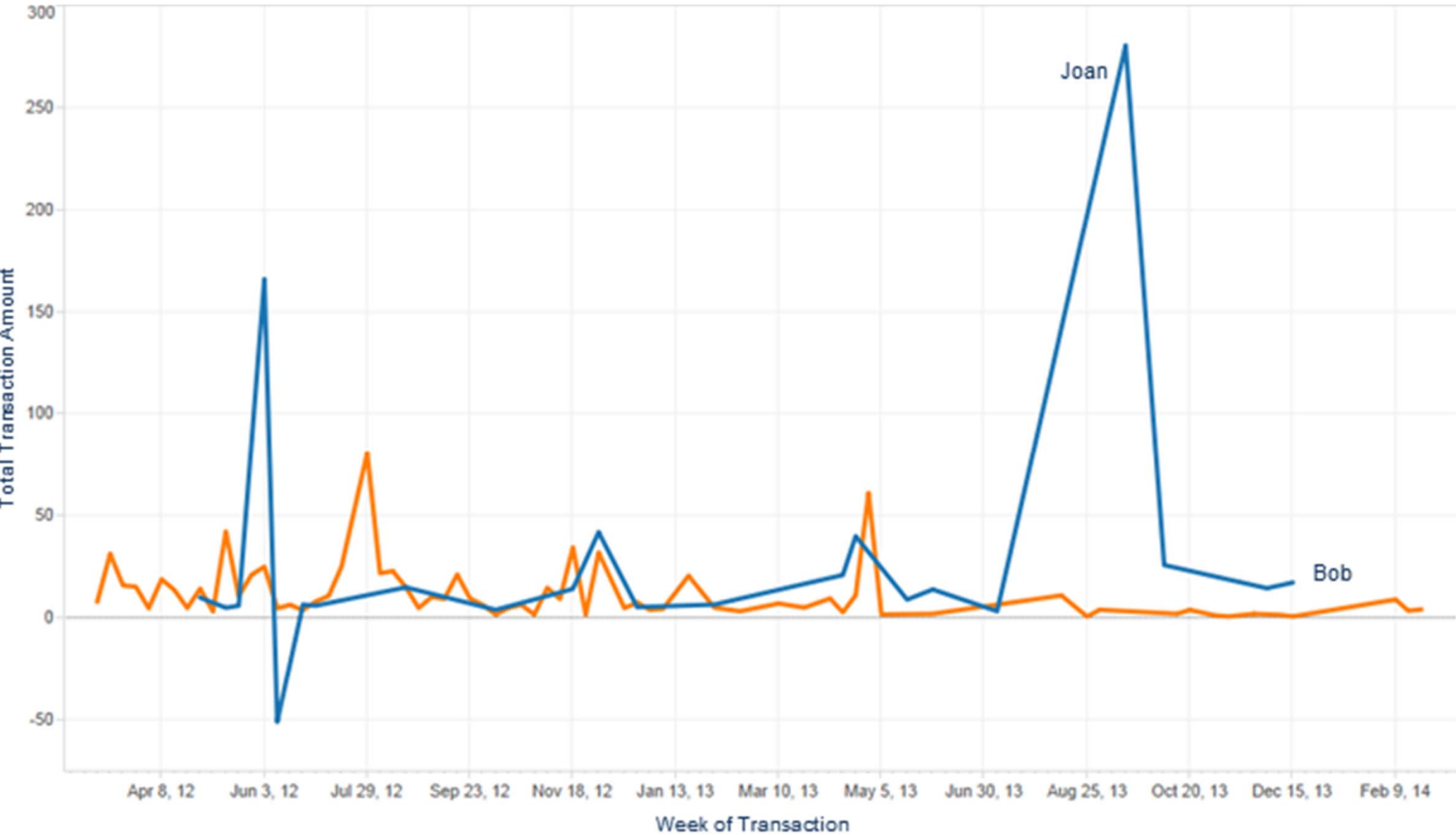
Purchases Over Time – Bob and Joan and Carlos

Purchases Over Time – Bob and Joan and Carlos

Purchases Over Time – Bob and Joan
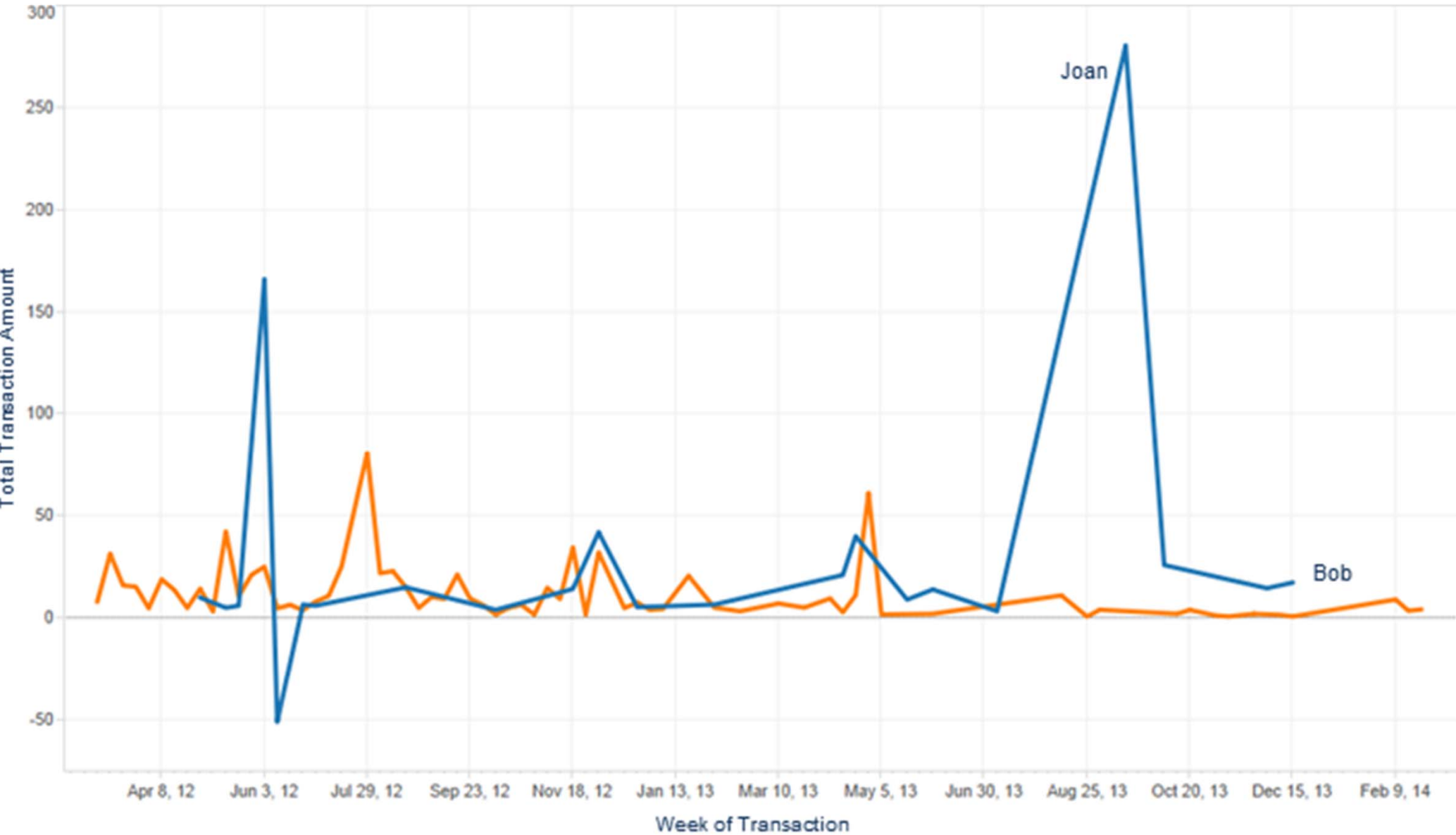
Purchases Over Time – Bob and Joan

The lines in this visualization represent the customers' dollars spent in the given week. The size of the circles represents the quantity of items purchased.
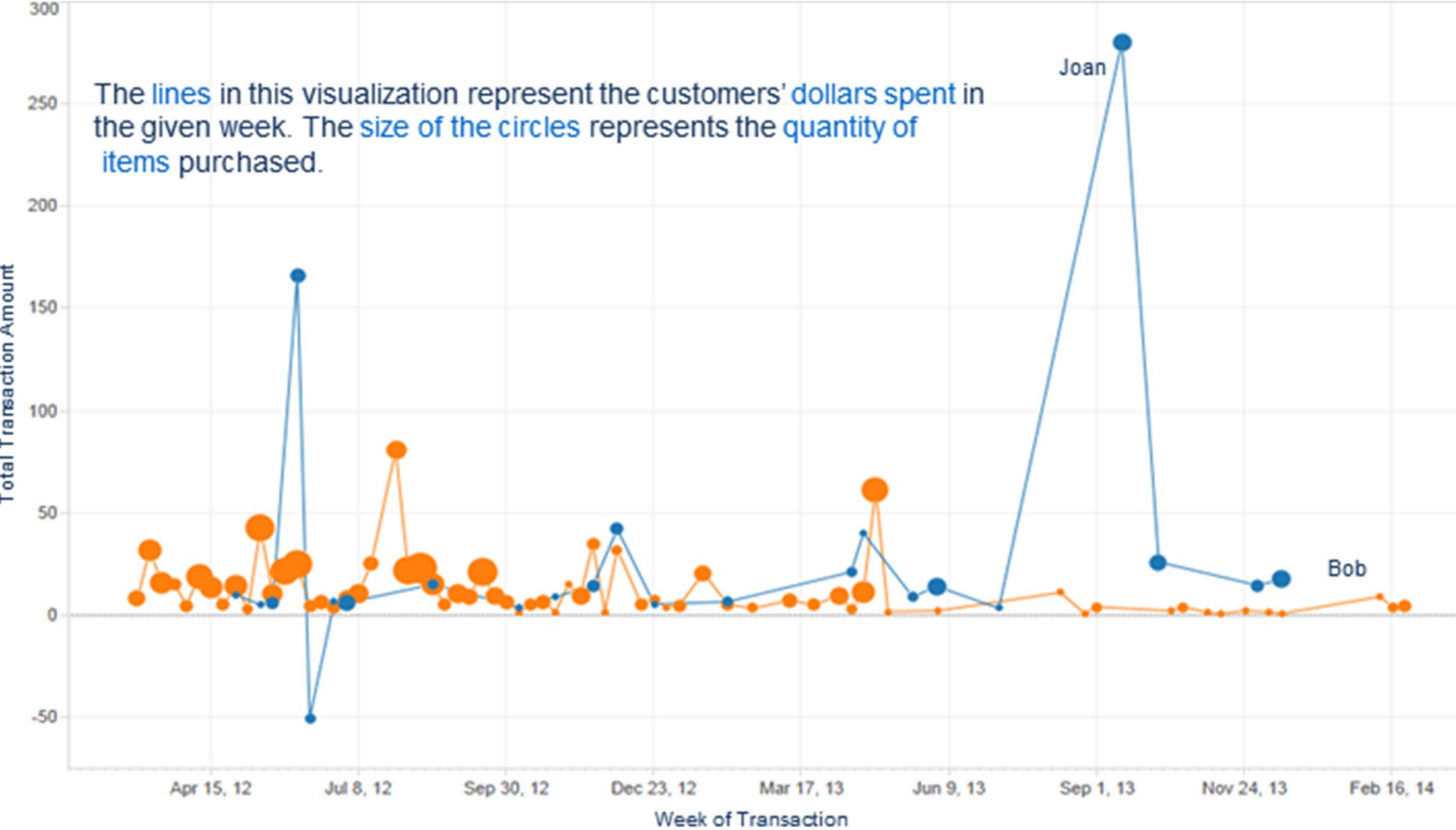
Purchases Over Time – Bob

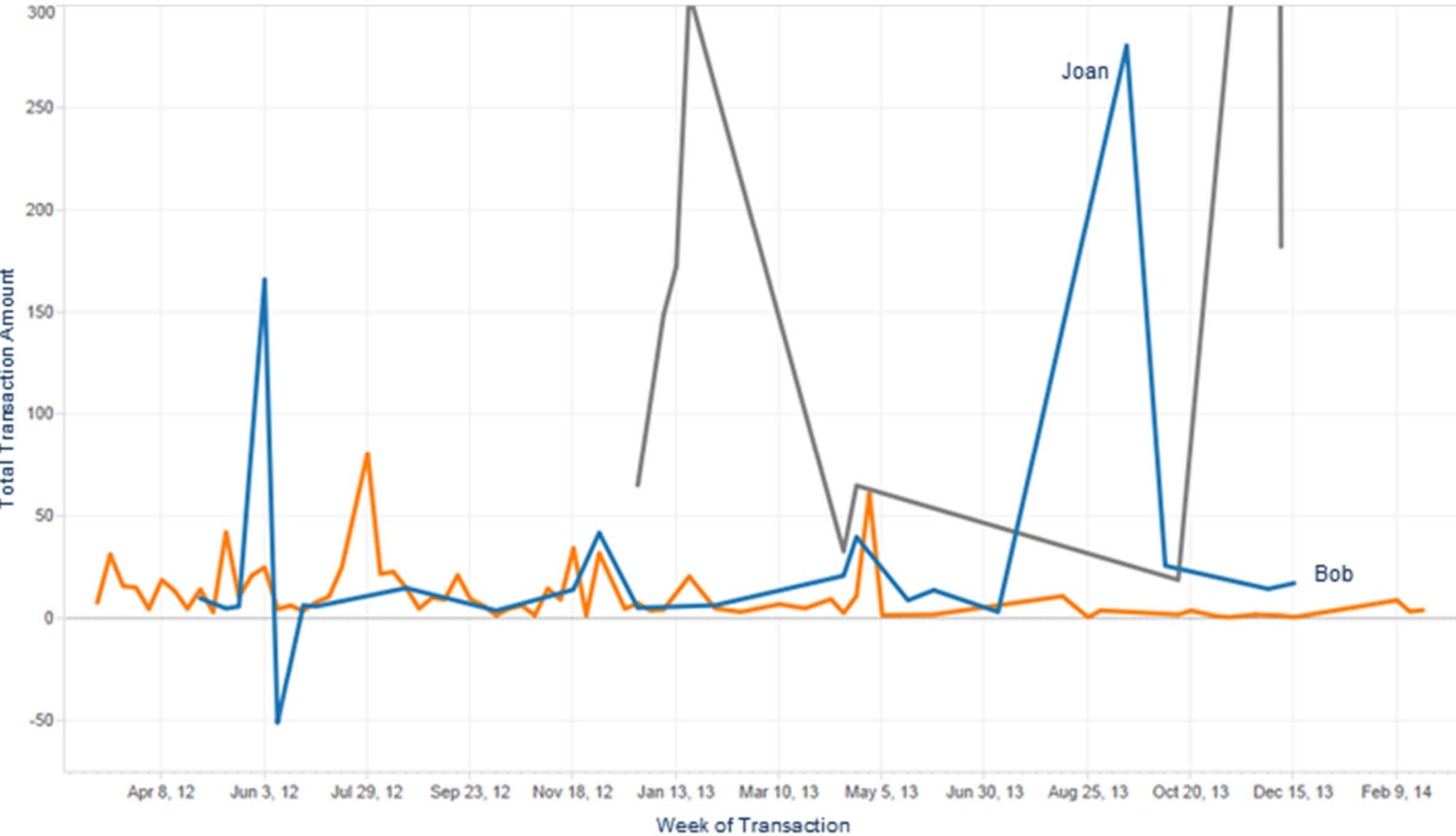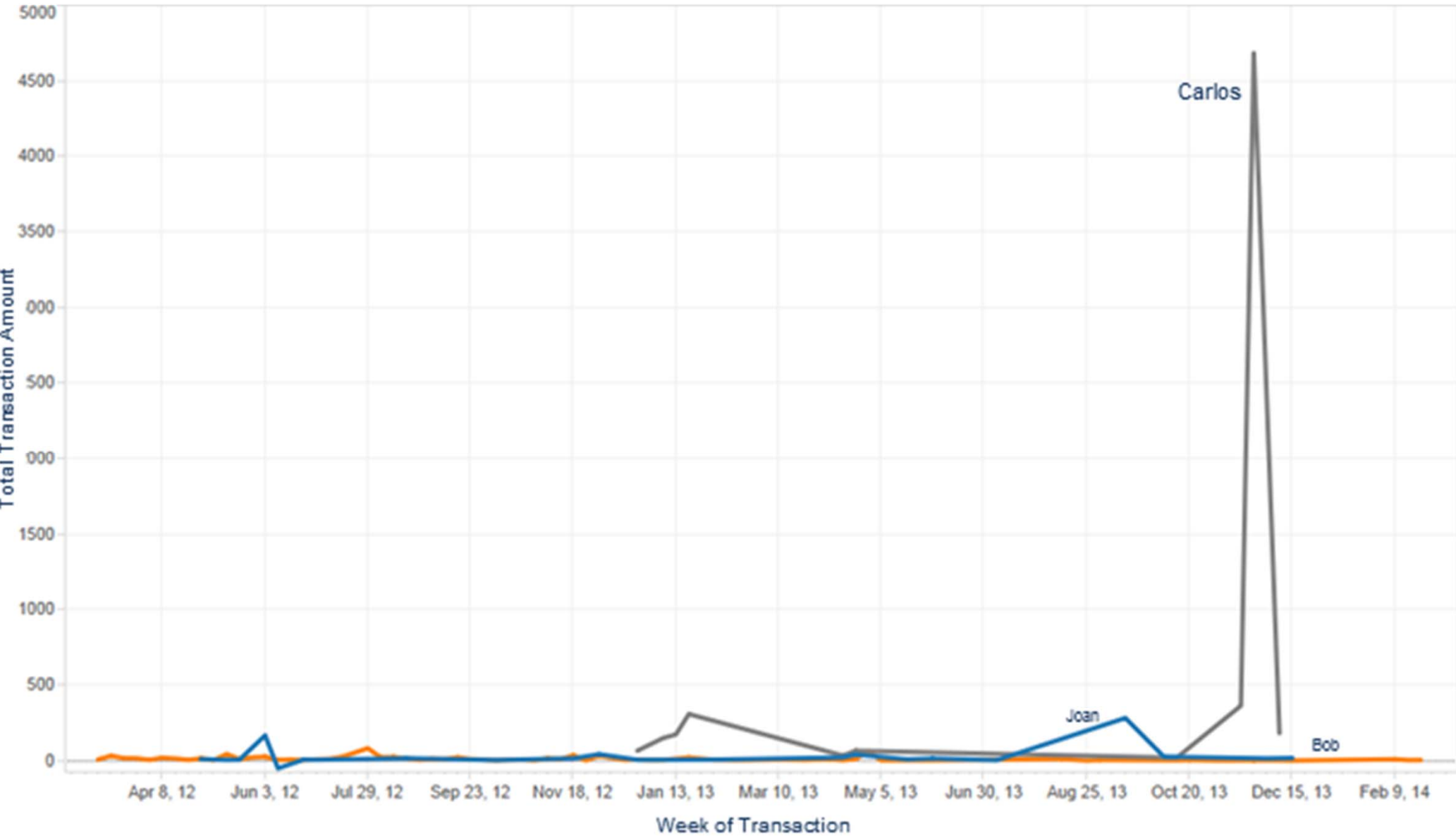| sample_segment	hh_id	channel	loc_id	register_id	transaction_id	transaction_date	unit_qty	unit_price	unit_id
KC	371225153	STORE	557	12	2195	5/25/2013 13:57	1	3.98	232590
KC	371225153	STORE	557	12	2195	5/25/2013 13:57	1	4.97	2315323
KC	371225153	STORE	557	12	3976	6/3/2013 11:42	1	1	197770
KC	371225153	STORE	557	12	3976	6/3/2013 11:42	1	1	199866
KC	371225153	STORE	557	12	3976	6/3/2013 11:42	1	1.35	151229
KC	371225153	STORE	557	12	3976	6/3/2013 11:42	1	2.94	24523
KC	371225153	STORE	557	12	3976	6/3/2013 11:42	1	3.67	206392
KC	371225153	STORE	557	12	3976	6/3/2013 11:42	1	3.93	736134
KC	371225153	STORE	557	10	8411	7/13/2013 13:12	1	3.28	694059
KC	371225153	STORE	557	10	4224	9/20/2013 8:51	8	0.88	529085
KC	371225153	STORE	557	10	4224	9/20/2013 8:51	8	1.36	464582
KC	371225153	STORE	557	13	4171	9/21/2013 15:06	3	11.97	673349
KC	371225153	STORE	557	5	2673	9/21/2013 15:20	1	16.98	24723
KC	371225153	STORE	557	5	2673	9/21/2013 15:20	1	69.97	757694
KC	371225153	STORE	557	5	2673	9/21/2013 15:20	1	179	777521

Purchases Over Time – Joan

| sample_segment	hh_id	channel	loc_id	register_id	transaction_id	transaction_date	unit_qty	unit_price	unit_id
KC	593185567	STORE	2720	1	6699	4/22/2013 8:03	1	0.68	785084
KC	593185567	STORE	2720	1	6699	4/22/2013 8:03	1	1.68	64509
KC	593185567	STORE	2720	1	6699	4/22/2013 8:03	1	1.98	785098
KC	593185567	STORE	2720	16	3883	4/23/2013 7:48	1	0.68	785084
KC	593185567	STORE	2720	16	3883	4/23/2013 7:48	1	0.78	734114
KC	593185567	STORE	2720	16	3883	4/23/2013 7:48	1	1.08	90815
KC	593185567	STORE	2720	16	3883	4/23/2013 7:48	1	2.42	480579
KC	593185567	STORE	2720	1	7457	4/24/2013 8:16	1	0.68	785084
KC	593185567	STORE	2720	1	7457	4/24/2013 8:16	1	1.08	89304
KC	593185567	STORE	2720	13	8865	4/28/2013 13:12	2	2.96	556170
KC	593185567	STORE	2720	13	8865	4/28/2013 13:12	1	5	419403
KC	593185567	STORE	2720	13	8865	4/28/2013 13:12	2	5	295419
KC	593185567	STORE	2720	13	8865	4/28/2013 13:12	1	9.98	59925
KC	593185567	STORE	2720	13	8865	4/28/2013 13:12	2	13.76	498928
KC	593185567	STORE	2720	13	8865	4/28/2013 13:12	2	15.34	2300557
KC	593185567	STORE	2720	13	8870	4/28/2013 13:24	1	2.5	295419
KC	593185567	STORE	2720	16	4078	4/29/2013 8:35	1	0.68	785084
KC	593185567	STORE	2720	16	4078	4/29/2013 8:35	1	1.68	78788
KC	593185567	STORE	2720	57	5951	4/30/2013 10:02	1	0.68	749630
KC	593185567	STORE	2720	57	5951	4/30/2013 10:02	1	1.78	88214

Purchases per Week by Department

Some departments are consistently purchased frequently, and in high volume.
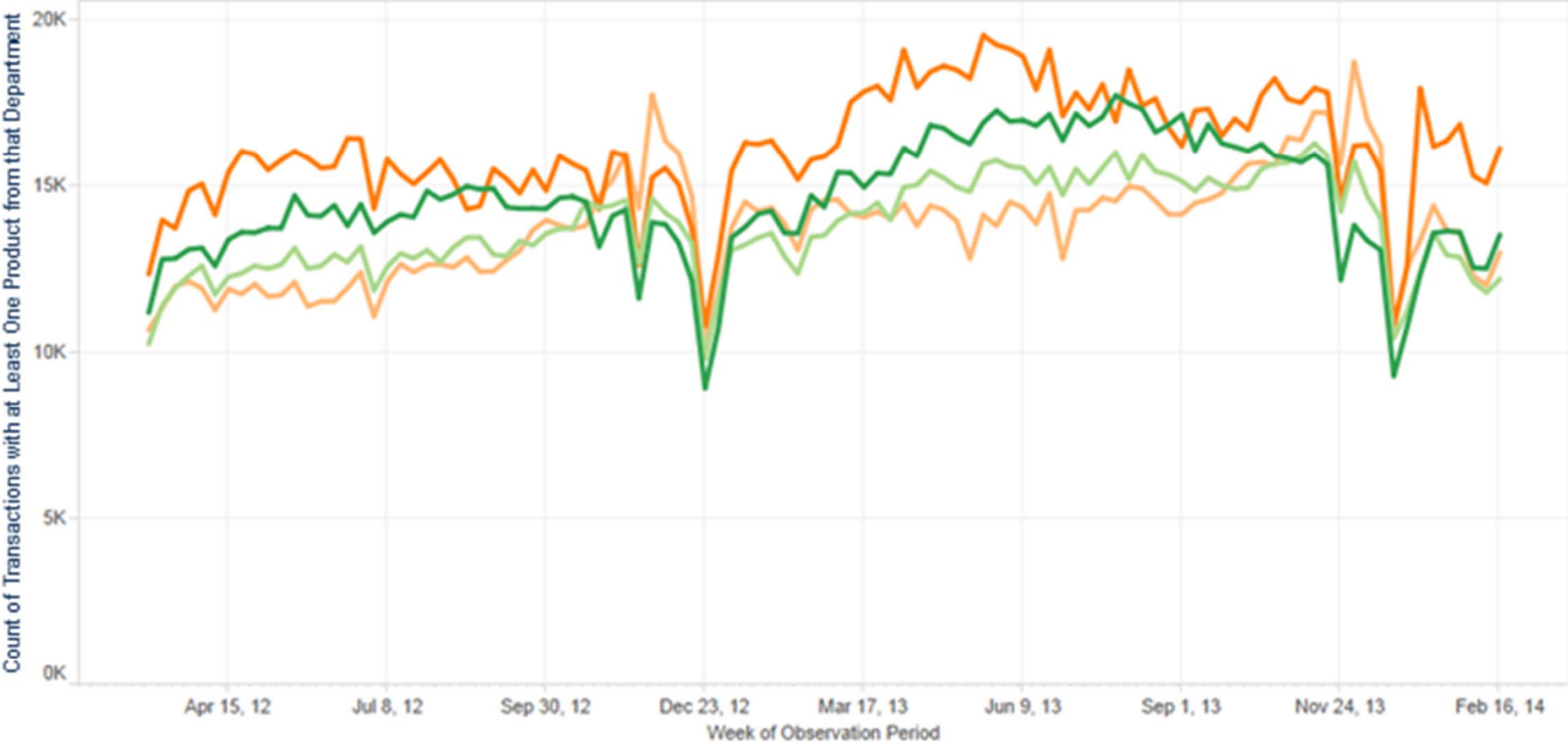
Purchases per Week by Department

While some departments are purchased in lower volume.
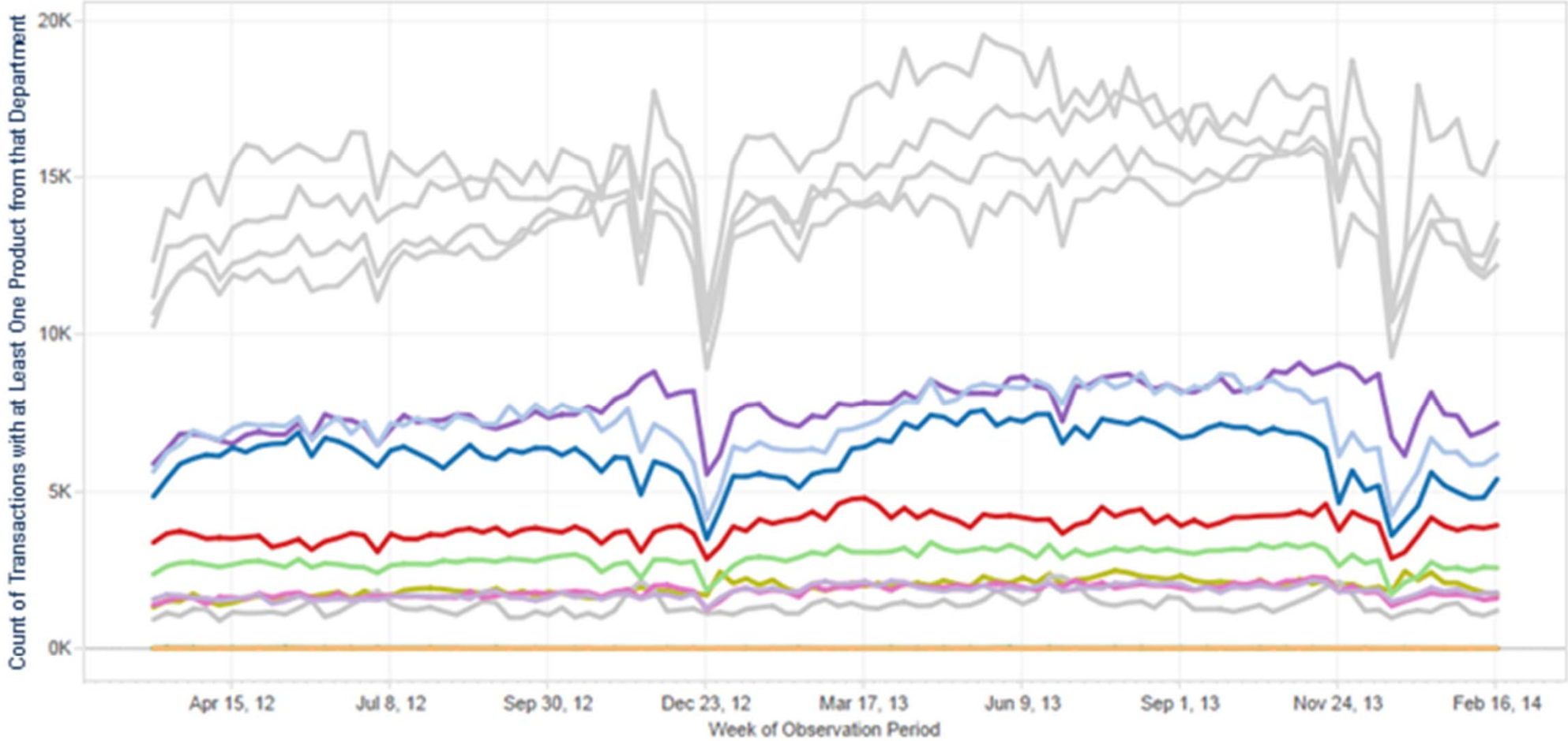
Purchases per Week by Category

There is also an element of seasonality in certain departments.
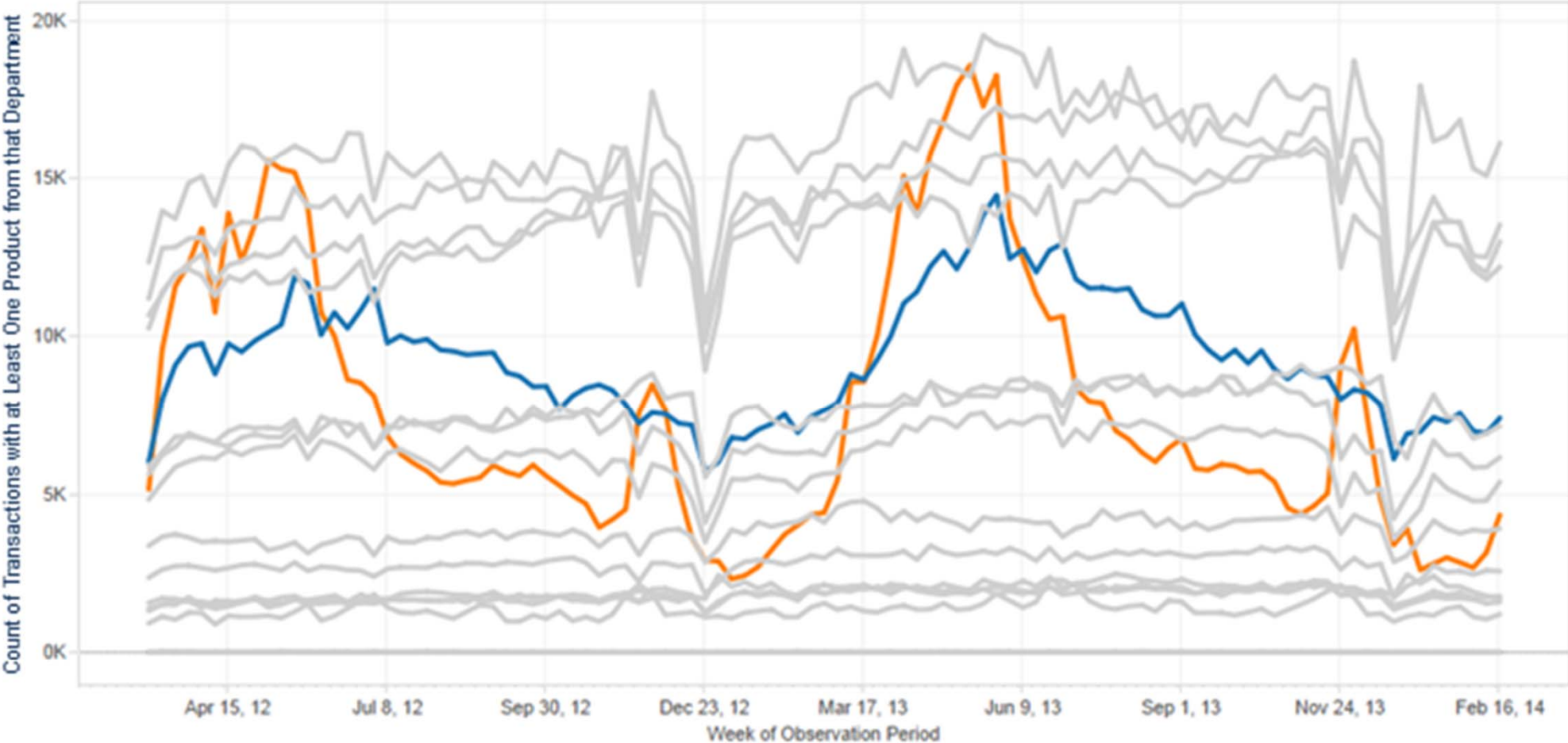
Example Product Associations
(column conditional on row)

Transactions frequently show **multiple subcategories being purchased together.**

This graphic shows 5.83% of all transactions with product from subcategory E have at least one product from subcategory C in them.

Percent In Transactions

0.00% — 5.83%

Subcategory	A	B	C	D	E	F	G	H	I	J
A		0.18%	0.06%	0.25%	0.00%	0.24%	0.26%	0.00%	0.14%	0.15%
B	0.26%		0.22%	2.26%	0.04%	0.89%	1.51%	0.00%	0.31%	0.87%
C	0.10%	0.24%		0.91%	1.41%	1.04%	0.71%	0.02%	0.23%	0.02%
D	0.12%	0.76%	0.28%		0.11%	0.57%	2.46%	0.00%	0.71%	0.90%
E	0.03%	0.19%	5.83%	1.51%		1.79%	1.13%	0.06%	0.72%	1.07%
F	0.20%	0.52%	0.56%	0.99%	0.23%		0.82%	0.00%	0.25%	0.82%
G	0.11%	0.43%	0.18%	2.07%	0.07%	0.40%		0.00%	0.49%	0.49%
H	0.00%	0.00%	0.43%	0.14%	0.29%	0.00%	0.29%		0.14%	0.43%
I	0.15%	0.24%	0.17%	1.67%	0.13%	0.34%	1.37%	0.01%		0.98%
J	0.12%	0.46%	0.01%	1.43%	0.13%	0.75%	0.92%	0.01%	0.66%	

Example Product Associations
(column conditional on row)

0.00% 31.79%

Transactions frequently show **multiple subcategories being purchased together.**
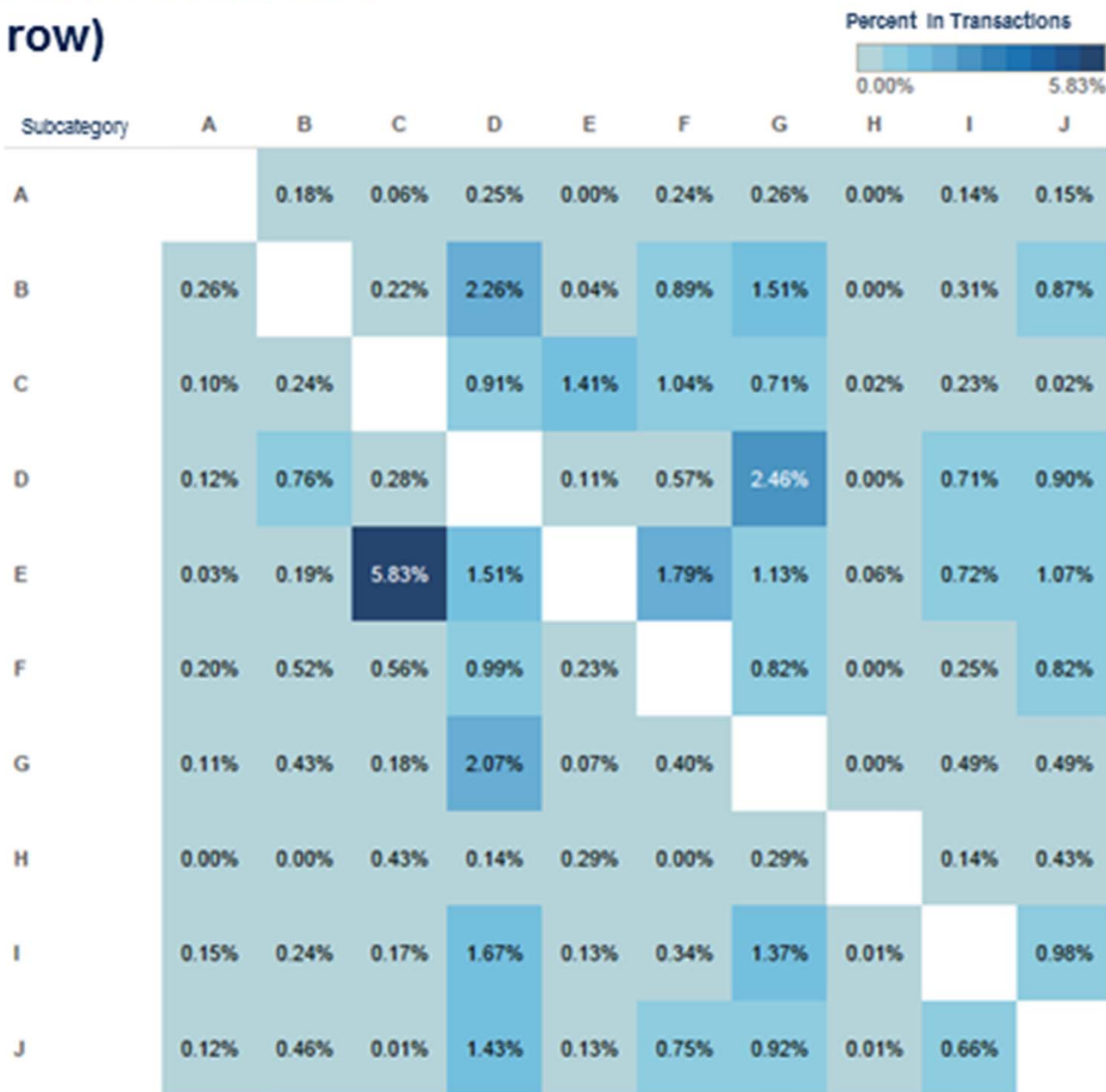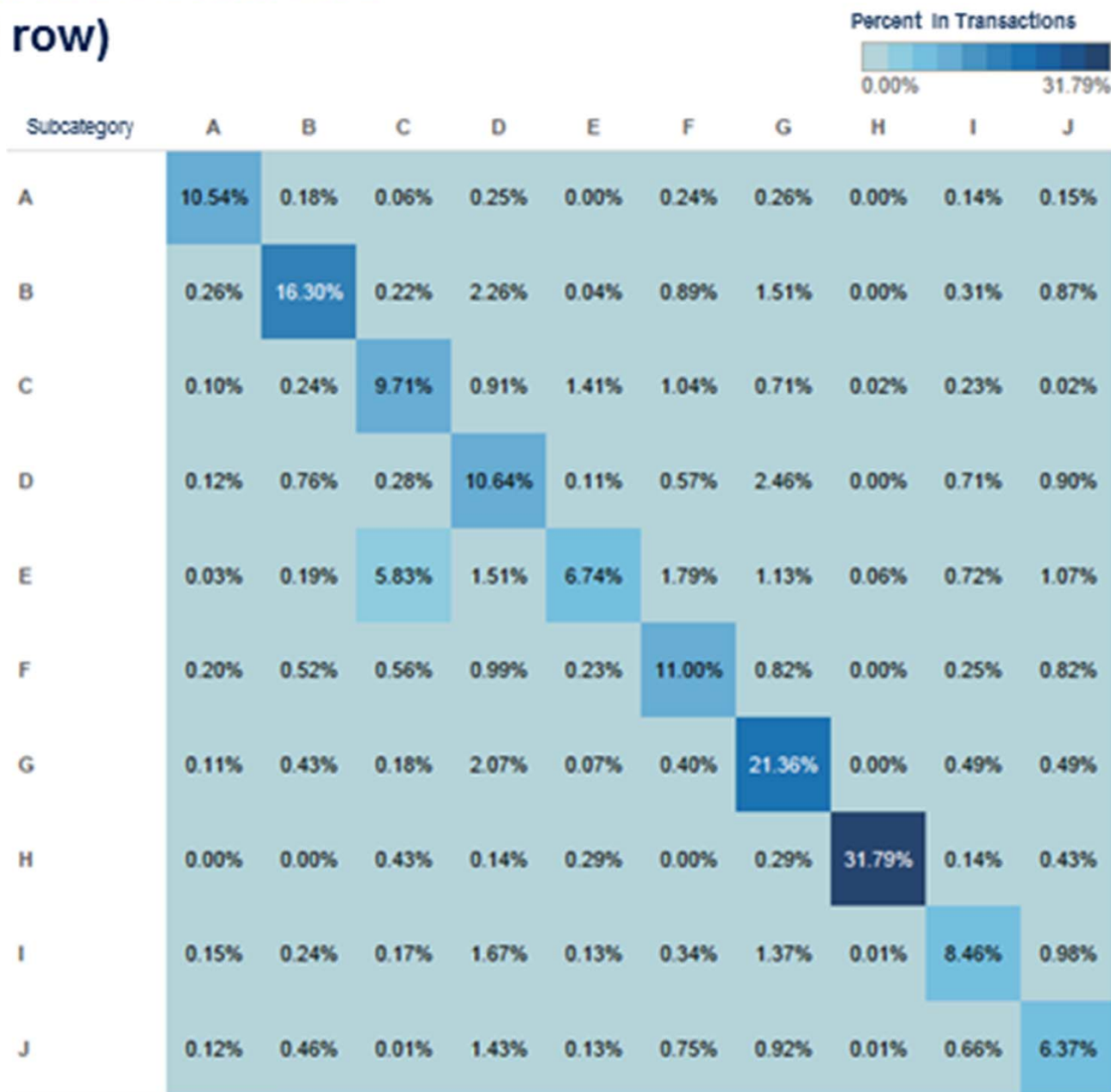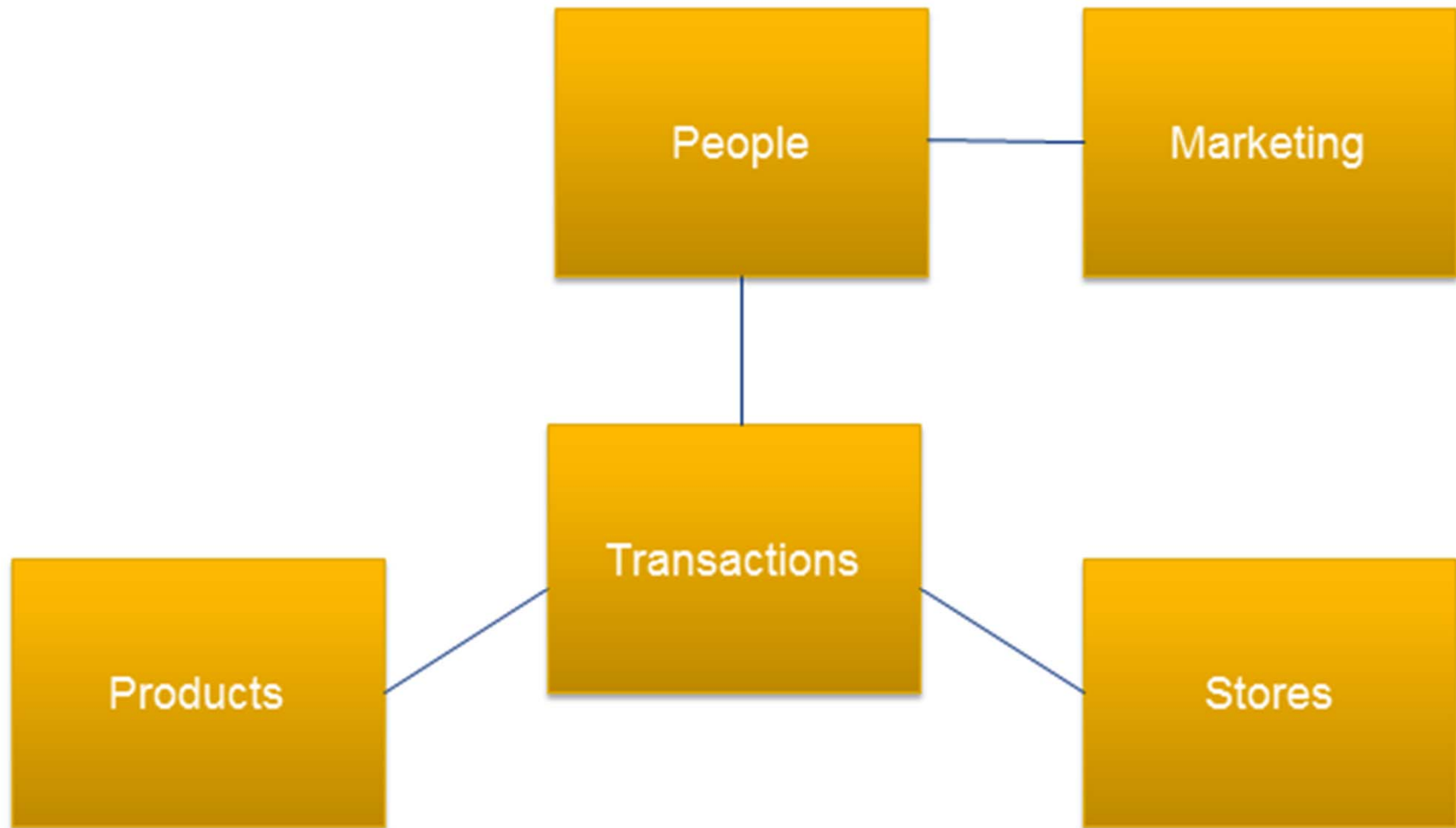
However, it is much more likely that **several products will be purchased within the same subcategory.**

To use a craft store as an example, it is more likely that a customer will purchase **several skeins of yarn** rather than **yarn** and a **hot glue gun.**

Subcategory	A	B	C	D	E	F	G	H	I	J
A	10.54%	0.18%	0.06%	0.25%	0.00%	0.24%	0.26%	0.00%	0.14%	0.15%
B	0.26%	16.30%	0.22%	2.26%	0.04%	0.89%	1.51%	0.00%	0.31%	0.87%
C	0.10%	0.24%	9.71%	0.91%	1.41%	1.04%	0.71%	0.02%	0.23%	0.02%
D	0.12%	0.76%	0.28%	10.64%	0.11%	0.57%	2.46%	0.00%	0.71%	0.90%
E	0.03%	0.19%	5.83%	1.51%	6.74%	1.79%	1.13%	0.06%	0.72%	1.07%
F	0.20%	0.52%	0.56%	0.99%	0.23%	11.00%	0.82%	0.00%	0.25%	0.82%
G	0.11%	0.43%	0.18%	2.07%	0.07%	0.40%	21.36%	0.00%	0.49%	0.49%
H	0.00%	0.00%	0.43%	0.14%	0.29%	0.00%	0.29%	31.79%	0.14%	0.43%
I	0.15%	0.24%	0.17%	1.67%	0.13%	0.34%	1.37%	0.01%	8.46%	0.98%
J	0.12%	0.46%	0.01%	1.43%	0.13%	0.75%	0.92%	0.01%	0.66%	6.37%

The Data

Key Research Question

How should retail transaction data be analyzed or conceptualized when customers may be engaged in "projects"?

Key Research Question

How should retail transaction data be analyzed or conceptualized when customers may be engaged in "projects"?

- Can we determine (from the purchase histories) sets of products that seem to be purchased together as part of a project (not just in the same transaction). Can this information be used to construct a "project materials list"?

- Is it possible to detect when a customer is engaged in a project, based on shopping behavior alone?

- Can you predict the customer's next trip or, more importantly, the next big project? Is there a natural order to projects?

- Can customer groups be targeted for offers or promotions based on their prior behavior or current project(s)?

- Can we identify what products or services involved with a project are being purchased elsewhere? Can we determine share of wallet?

Key Research Question

How should retail transaction data be analyzed or conceptualized when customers may be engaged in "projects"?

- Can we determine (from the purchase histories) sets of products that seem to be purchased together as part of a project (not just in the same transaction). Can this information be used to construct a "project materials list"?

- Is it possible to detect when a customer is engaged in a project, based on shopping behavior alone?

- Can you predict the customer's next trip or, more importantly, the next big project? Is there a natural order to projects?

- Can customer groups be targeted for offers or promotions based on their prior behavior or current project(s)?

- Can we identify what products or services involved with a project are being purchased elsewhere? Can we determine share of wallet?

The data sponsor is also open to unique/novel research areas using this dataset.

Benefits of Participating in a WCAI Research Opportunity

- **Access to unique data**

- **Assistance with data cleaning, preparation, and processing at any point during the Research Opportunity**

- **Teleconference Q&A sessions with the research sponsor**

- **Opportunity to present findings to the Corporate Sponsor at the closing symposium, to be held at Wharton one year from now** (note: attendance by at least one member from each team required)

- **Potential for additional PR for your research**

- **Promotion of your research paper through the WCAI SSRN Research Paper Series**

WHARTON CUSTOMER ANALYTICS INITIATIVE

Wharton
UNIVERSITY of PENNSYLVANIA

Proposal Format

Proposals should be no more than 2,000 words and include the following:

- **Title**

- **Author(s) name, title, affiliation and e-mail address:** Please designate a corresponding author.

- **Summary information:** a single "slide" that visually summarizes team & project.

- **Abstract**

- **Introduction:** Describe expected contribution(s), covering both the academic and practical aspects. Please keep it concise, and cite relevant work as necessary to explain your academic contribution. There is no need to include a lengthy literature review.

- **Detailed project proposal:** Please include enough detail that we can assess the feasibility & merit of the proposed approach. For example, modeling projects should include at least a sketch of the model. In addition, include a rough estimate of how long the project will take. Also include the business relevance of your research and the impact for the sponsor as well.

- **Biographies:** Include up to a paragraph-long biography highlighting what each team member will contribute to the project.

PDFs ONLY PLEASE!

WHARTON CUSTOMER ANALYTICS INITIATIVE

Wharton
UNIVERSITY of PENNSYLVANIA

Proposal Submission Process

- Read more about the proposal format at http://www.wharton.upenn.edu/wcai/Proposals.cfm

- Submit proposals at http://www.wharton.upenn.edu/wcai/proposalform.cfm no later than Wednesday, June 11, 2014, 12 noon US Eastern.

  - Proposals will only be accepted online.

  - A single PDF combining the written portion and the single slide.

- Proposals will be evaluated both on academic contribution and potential to significantly improve the research sponsor's marketing practice.

  - Anand Bodapati (UCLA), Gary Russell (University of Iowa), Eric Bradlow (WCAI), Pete Fader (WCAI), Elea Feit (WCAI), and representatives from the Research Sponsor.

- Contact wcai-research@wharton.upenn.edu, if you have questions prior to submitting your proposal.

WHARTON CUSTOMER ANALYTICS INITIATIVE

Wharton
UNIVERSITY of PENNSYLVANIA

Other WCAI Opportunities for Researchers

If you registered for this webinar, you will receive regular announcements regarding:

- Research Opportunities like this one

  - Desktop Software Use Analysis

  - 2-3 more Research Opportunities in the Fall

Also find us at:

- SSRN Research Paper series:
  http://www.ssrn.com/link/Wharton-Cust-Analytics-Initiative-RES.html

- Announcements: http://wcai.wharton.upenn.edu