

Freelisting

Eliciting the members of a domain

Free Listing

- Basic idea:
 - Tell me all the <category name> you can think of
 - Typically loosely timed, no questions allowed
 - An example of Spradley’s “grand tour” question
- Contrasts with survey open-ended question
 - Open-end is typically about the respondent:
 - what do you like about this product? what ice-cream flavors do you like? what illnesses have you had?
 - Free list is about the domain:
 - what ice-cream flavors are there? what illnesses exist?

Why we do it

- Analysis of the list itself
 - What makes something a fruit? A bad word?
 - Hypotheses about what will be salient
 - Comparing salience of items for different groups
 - Examining similarities among respondents
 - Who lists the same items
 - Examining similarities among items
 - Which items tend to be mentioned by the same respondents?
- First step in mapping the domain
 - i.e., getting a list of salient items to work with
- Obtaining local terminology
- Tongue loosener

How many respondents?

- Depends on level of consensus – coherence of domain
 - Non-domains like “reasons why organizations fail” need huge Ns, like 200+
- But typically,
 - For developing workable list for further analysis (e.g., doing pilesorts), need 20+
 - For analyzing the domain membership, need about 100
 - For comparing groups, need about 50 in each group

Synonyms, Misspellings, Suffixes

- When list is basis for further research, such as measuring similarity, need to
 - cull synonyms
 - Eliminate items at different levels of contrast
- When it is a linguistic study, you don't cull synonyms
- Spellings should be standardized,
- Plurals, -ing endings, etc should be standardized
 - but careful when you don't know the culture: is "ho" the same as "whore"?

Which items do you keep for further work?

- Most frequent items – as many as you can handle
- Items mentioned by more than 1 person
- Search for natural elbow in frequencies

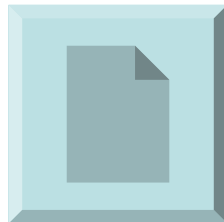
The “Bad Words” Domain

WARNING:
4-Letter words follow!

The squeamish and the moral should go back to work now!

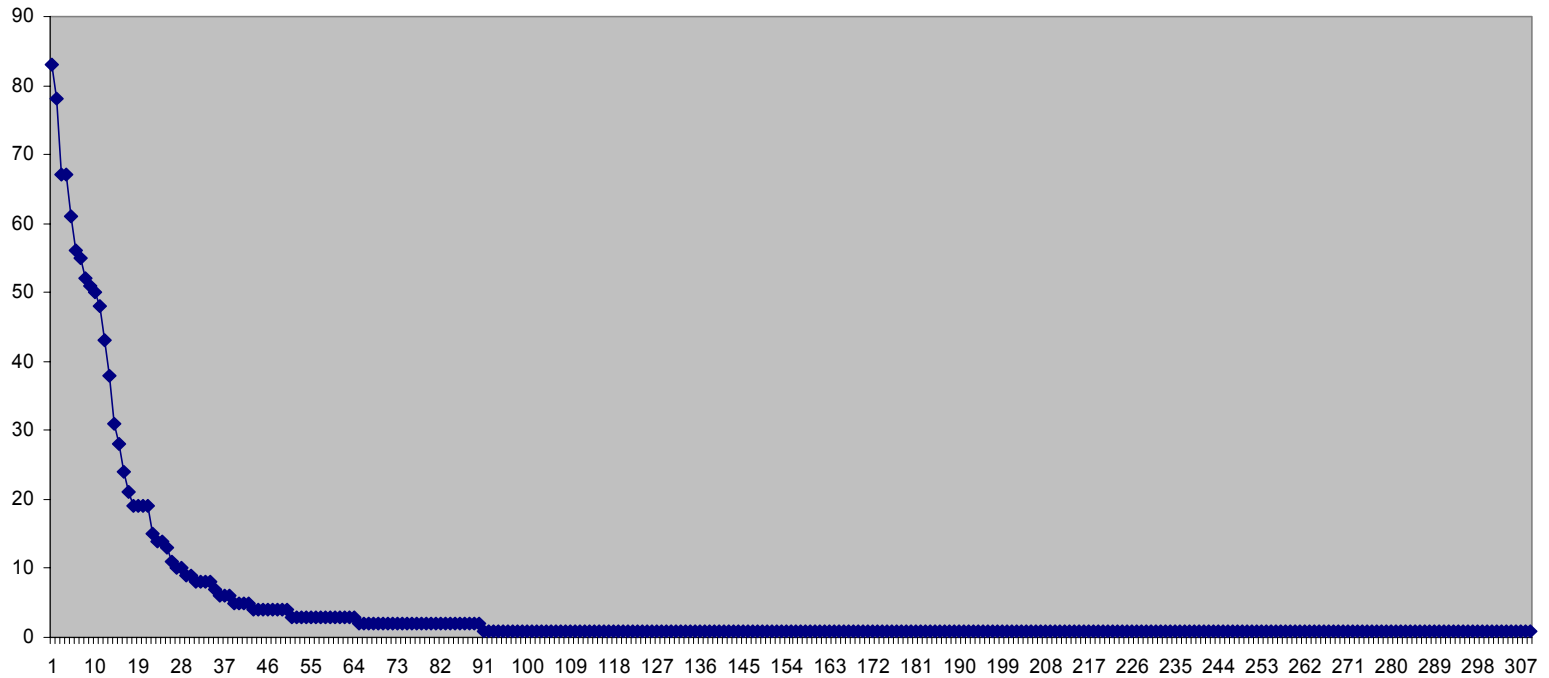
Frequencies

- Sort in descending order
- Tally average position in lists
- Combine frequency and position to create salience measure
- May need editing to standardize spelling
- In some cases, want to collapse synonyms
 - Not in linguistics projects, though



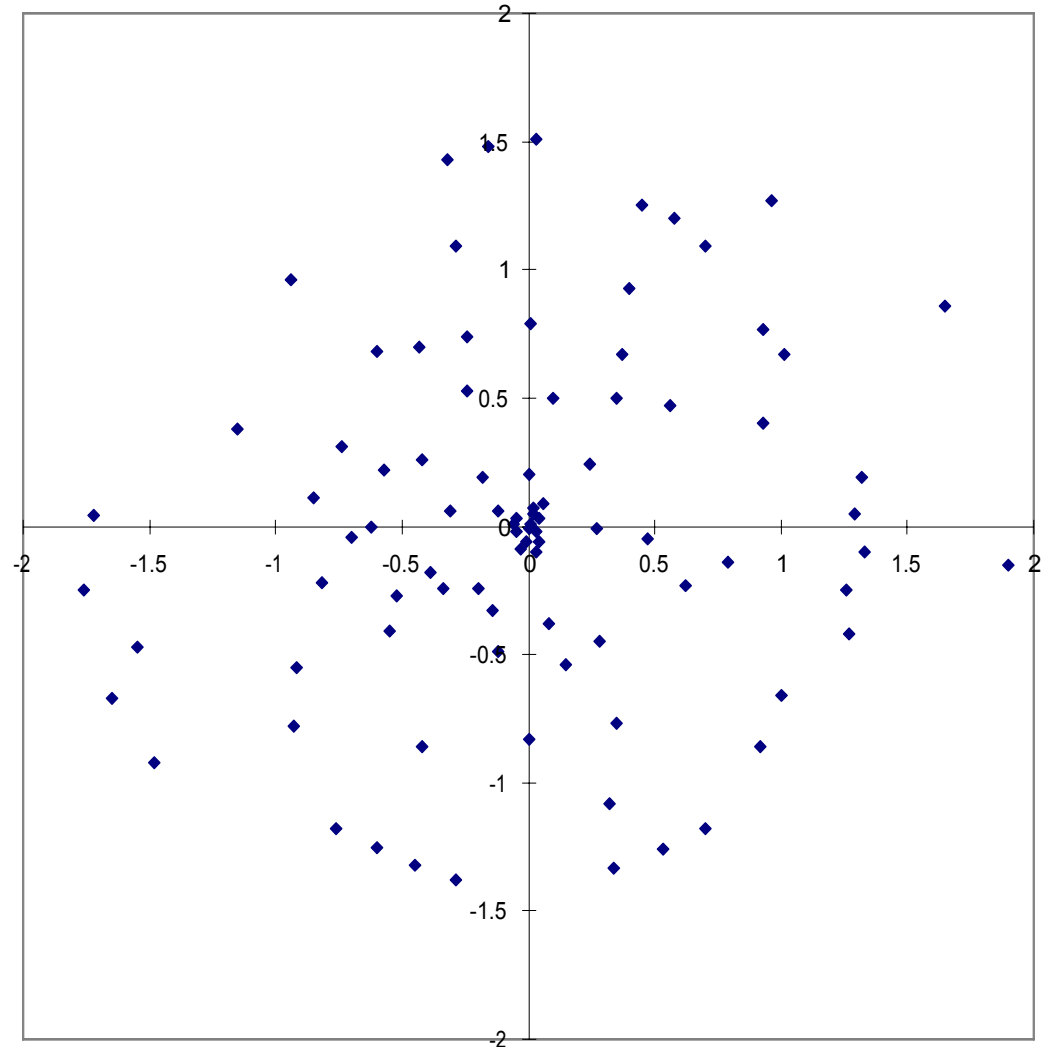
Domain borders are fuzzy

Frequencies of each bad word



Domains have core/periphery structure

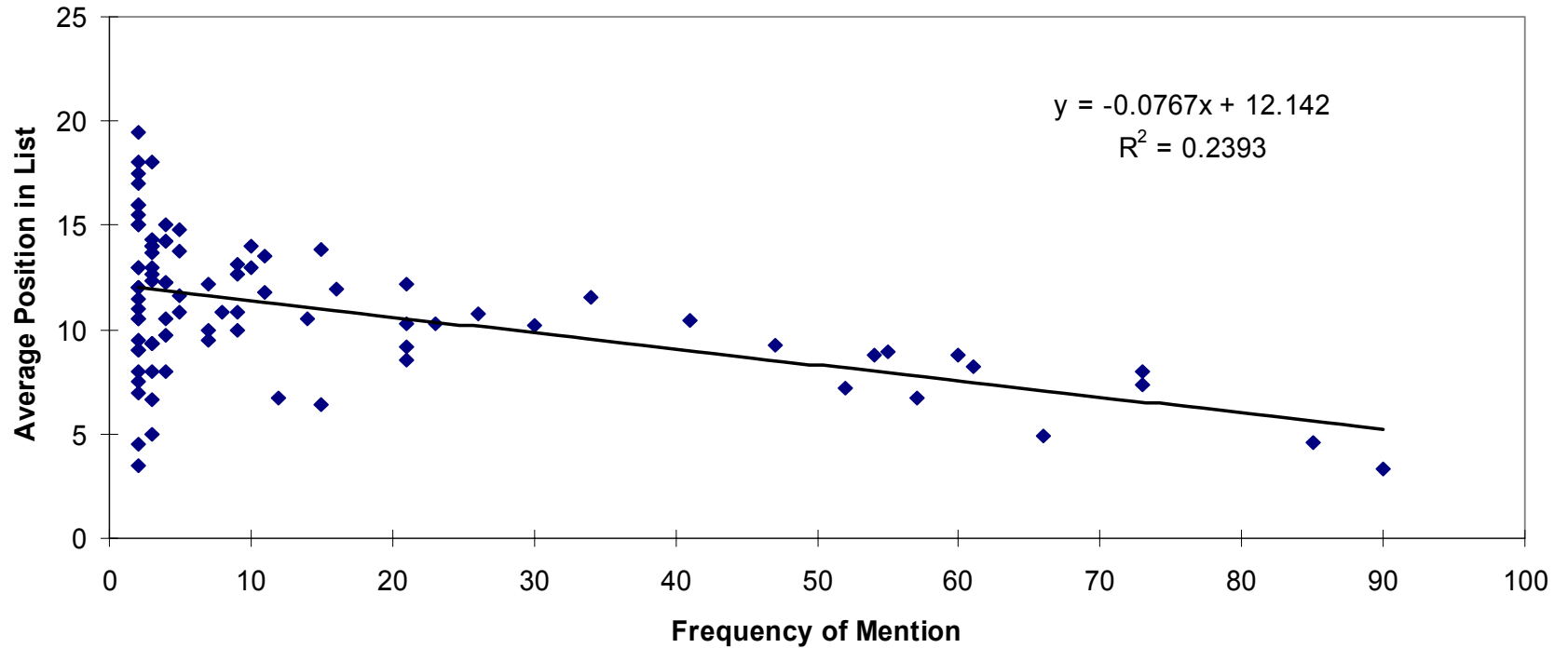
- MDS of item-item co-occurrences
- Each dot is a bad word
- Core items are in the center – in everybody's list – and co-occur with each other



Core items typically mentioned first

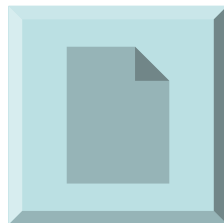
Characteristic negative correlation between avg rank and frequency

Frequency vs Rank



Can analyze respondents as well

- Length of lists
- Conventionalality of their lists (do they tend to list more popular items)
- Correlation between rank (position on list) and sample frequency
- Similarities (overlaps) in people's lists



Things to notice ...

- Boundaries of a domain are fuzzy
 - Not just artifact of aggregation
 - For additional data collection, need inclusion rules
- Simple, established cultural domains have
 - Core/periphery structure
 - Core items recalled first
 - Consensus among respondents:
 - Each list has core items + idiosyncratic
 - We don't see clusters
- Quantitative analysis of qualitative data

Animals Domain

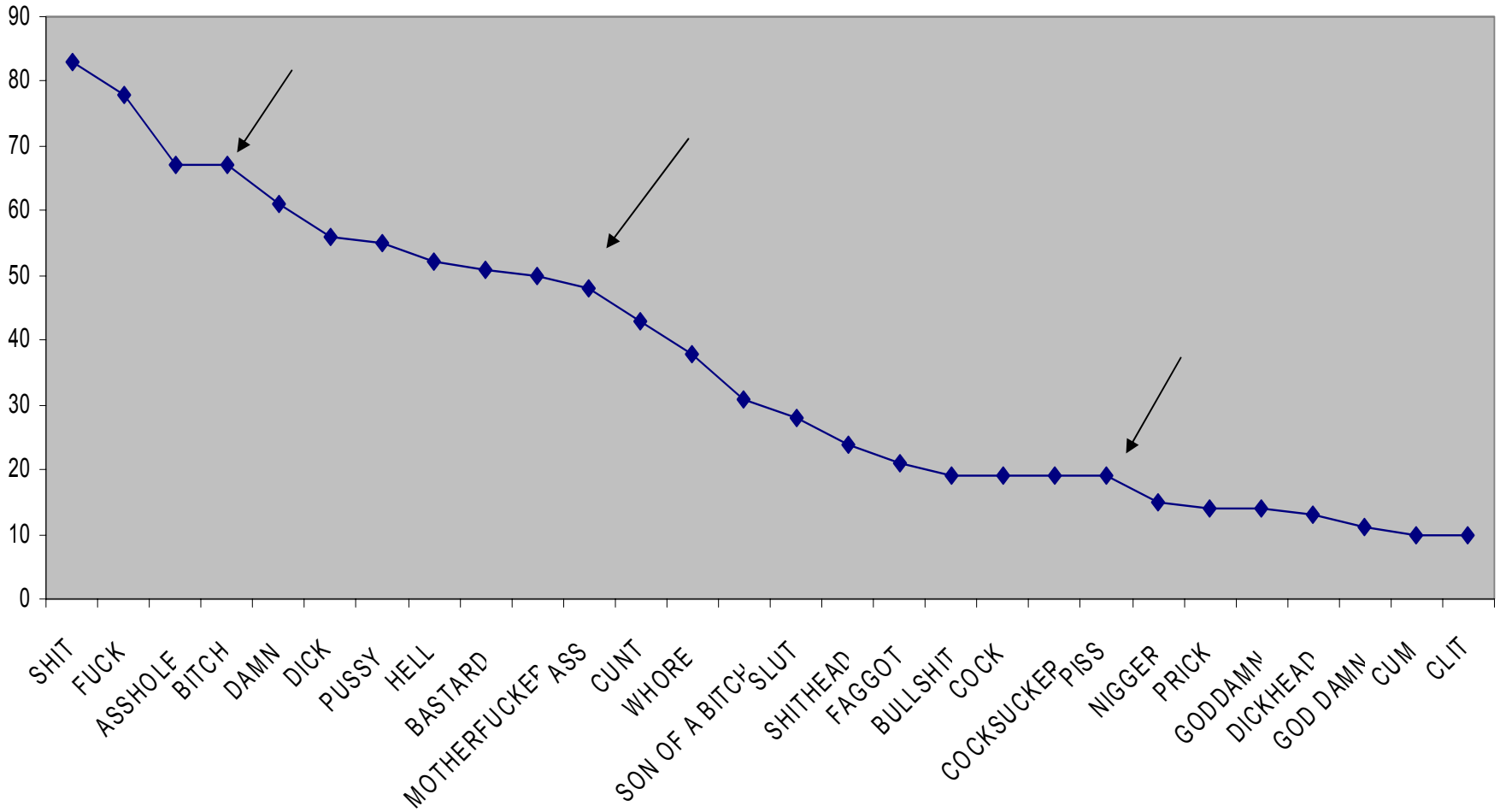
- Please grab a piece of paper and something to write with
- When I say 'go', please write down all the animals you can think of. You will have two minutes

Things to notice ...

- Ordering of items encodes ...
 - sub-category membership
 - Semantic relations such as similarity (lions & tigers)
complementarity (forks & knives)
- Can reproduce map of domain from free lists

Use scree plot to select core

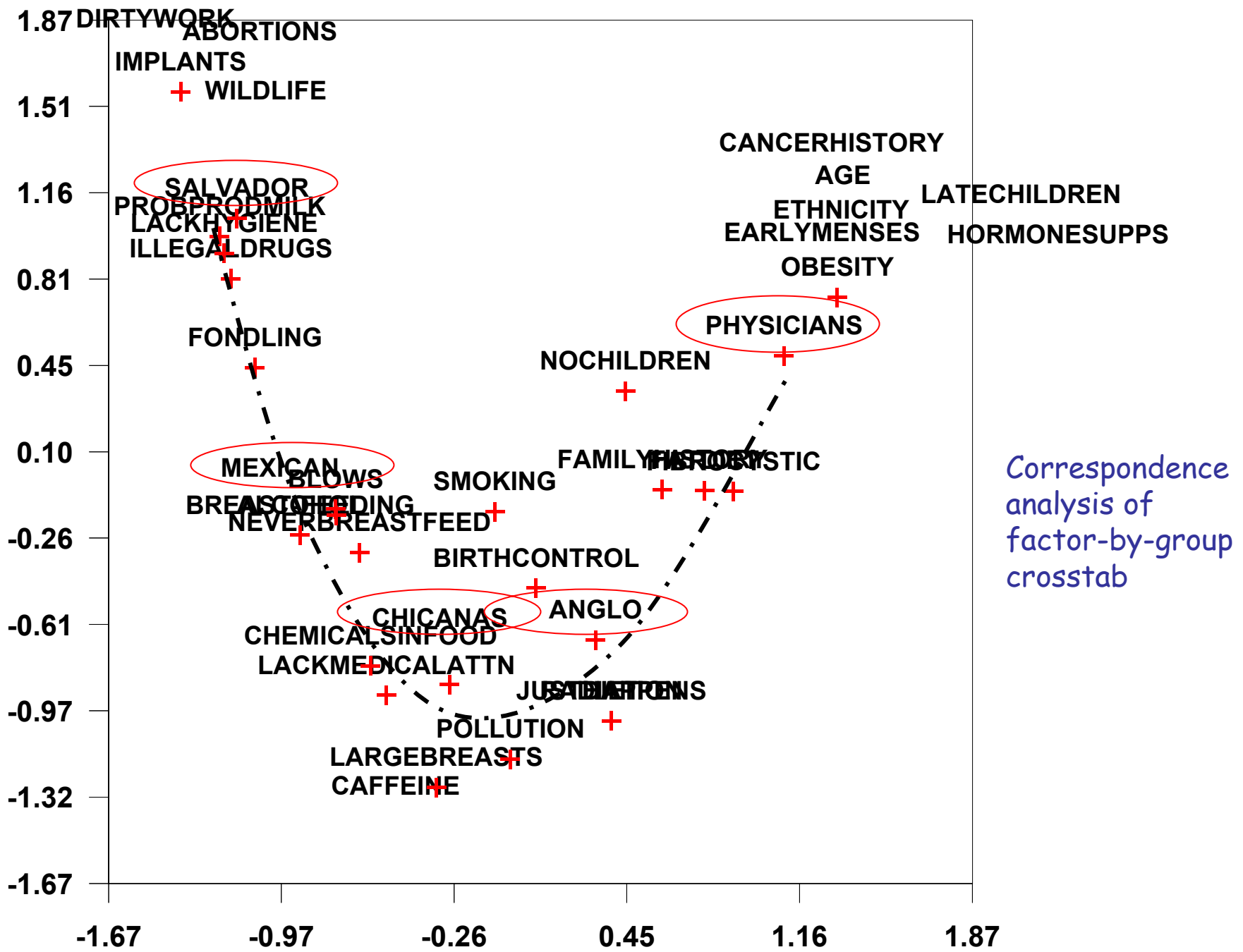
FREQUENCY



Comparative Use of Freelists

Causes of Breast Cancer

Salvadoran women (N = 28)	%^a	Mexican women (N = 39)	%	Chicanas (N = 27)	%	Anglo women (N = 27)	%	Physicians (N = 30)	%
Blows, bruises	29	Blows, bruises	64	Chemicals in food	30	Family history	67	Family history	100
Problems producing milk	29	Never breast-feeding	33	Environmental pollution	26	Radiation	26	Obesity	37
Breast implants	21	Chemicals in food	28	Blows, bruises	26	Unhealthy diet	19	Hormone supplements	33
Disorderly, wild life	16	Excessive fondling	23	Lack of medical atten.	26	Smoking	19	First child after 30	30
Excessive fondling	14	Problem producing milk	23	Family history	26	Birth control pills	19	High fat diet	30
Smoking	14	Birth control pills	18	Never breast-feeding	22	Environmental pollution	19	Prior history of cancer	30
Never breast-feeding	14	Breast-feeding	15	Smoking	19	It just happens	15	Age	27
Lack of hygiene	14	Lack of medical atten.	15	High fat diet	11	Blows, bruises	15	No children	20
Family history	11	Smoking	13	Large breasts	11	Never breast feeding	11	Smoking	17
Abortions	11	Too much alcohol	13	Too much caffeine	11	Fibrocystic breasts	11	Fibrocystic breasts	13
Illegal drugs	11	No children	13	Birth control pills	11	High fat diet	11	Ethnicity	13
Dirty work environment	11	Lack of hygiene	8					Early menses	13
		Illegal drugs	8					Birth control pills	13
		Family history	8						



Correspondence analysis of factor-by-group crosstab

Holiday Destinations

Destination	Girls	Boys	Destination	Girls	Boys	Destination	Girls	Boys
HAWAII	0.68	0.76	NEW YORK CITY	0.16	0.23	EUROPE	0.16	0.13
BAHAMAS	0.45	0.63	LOS ANGELES	0.21	0.19	DC	0.24	0.08
CANCUN	0.53	0.52	MEXICO	0.21	0.18	AMSTERDAM	0.18	0.10
JAMAICA	0.42	0.52	EGYPT	0.11	0.24	BOSTON	0.13	0.13
CALIFORNIA	0.42	0.48	GRAND CANYON	0.13	0.23	ORLANDO	0.13	0.13
FLORIDA	0.45	0.45	LAS VEGAS	0.18	0.18	CHINA	0.11	0.13
PARIS	0.34	0.47	CANADA	0.16	0.18	DISNEYLAND	0.13	0.11
AUSTRALIA	0.39	0.40	CARIBBEAN	0.13	0.19	GERMANY	0.11	0.13
BERMUDA	0.37	0.34	ARUBA	0.13	0.19	SAN DIEGO	0.16	0.10
LONDON	0.39	0.31	COLORADO	0.18	0.16	AFRICA	0.05	0.16
DISNEY WORLD	0.24	0.29	CAPE COD	0.16	0.18	FLORENCE	0.08	0.13
PUERTO RICO	0.16	0.32	NEW ORLEANS	0.18	0.15	NEW ZEALAND	0.16	0.08
ITALY	0.13	0.32	VIRGIN ISLANDS	0.21	0.13	ENGLAND	0.03	0.16
FRANCE	0.18	0.27	MONTREAL	0.16	0.16	VENICE	0.08	0.13
SPAIN	0.13	0.31	CHICAGO	0.18	0.13	CAYMAN ISLANDS	0.13	0.10
MIAMI	0.29	0.21	IRELAND	0.21	0.11	VERMONT	0.05	0.15
NEW YORK	0.26	0.21	ALASKA	0.16	0.15	BRAZIL	0.08	0.13
ROME	0.18	0.26	MAINE	0.16	0.13	HONG KONG	0.16	0.08
SAN FRANCISCO	0.18	0.23	JAPAN	0.13	0.15	ST. THOMAS	0.13	0.08

Statistical comparison: $r = 0.882$, $p (r_{\text{obs}} \leq r_p) = 0.49$

Things to notice ...

- Comparative analysis is particularly powerful
- Correspondence analysis
 - is clearly quantitative
 - Singular value decomposition of frequency matrix adjusted for row and column marginals
 - So we have quantitative analysis of qualitative data
 - On the other hand, the result is a picture – what can be more qualitative than that?

Working with multiple domains

- Domain overlap
- Building a network of domains ...

Domain of Fruits

TABLE 2.1
Frequency of Mention of "Fruits" in Free List Task

Apple	37	Honeydew	9
Orange	35	*Avocado	8
Pear	34	Mango	8
Banana	33	Date	7
Grape	32	Fig	7
Peach	30	Prune	7
Tangerine	27	Gooseberry	6
Cherry	26	Raisin	5
Grapefruit	26	*Pumpkin	4
Pineapple	26	Casaba melon	3
Strawberry	22	Kumquat	3
Watermelon	21	Melon	3
Lemon	20	Breadfruit	2
*Tomato	19	Kiwi	2
Apricot	18	Passionfruit	2
Blueberry	18	Persimmon	2
Plum	18	Cranberry	1
Cantaloupe	17	Crenshaw melon	1
Lime	16	Currant	1
Nectarine	14	Elderberry	1
Papaya	14	Huckleberry	1
Raspberry	14	Loganberry	1
Blackberry	13	Mandarine	1
Boisenberry	12	*Rhubarb	1
Tangelo	11	Salmonberry	1
Guava	10	*Squash	1
Pomegranate	10	Taro	1
Coconut	9	Turnip	1

Domain of Vegetables

TABLE 2.2
Frequency Distribution of "Vegetables" Free Listing Task

Green beans	55	Chinese peas	6
Corn	50	Greens	6
Carrots	49	Okra	6
Peas	41	Summer squash	6
Lima beans	40	Blackeyed peas	5
Lettuce	38	Swiss chard	5
Broccoli	37	Wax beans	5
Califlower	36	Bamboo shoots	4
Brussels sprouts	35	Navy beans	4
*Tomatoes	32	Alfalfa sprouts	3
Onions	30	Chile peppers	3
Spinach	30	Endive	3
Asparagus	29	Kidney beans	3
*Squash	28	Leek	3
Cucumbers	26	Parsnips	3
Celery	25	*Pumpkin	3
Cabbage	24	Redleaf lettuce	3
Zucchini	24	*Rhubarb	3
*Turnips	23	Water chestnuts	3
Potatoes	20	Butterleaf lettuce	2
Artichokes	18	Green onions	2
Bell peppers	18	Kale	2
Radishes	18	Kolari	2
*Avocado	18	Red onions	2
Beets	13	Sauerkraut	2
Rutabaga	11	Butternut squash	1
Bean sprouts	10	Garlic	1
Eggplant	9	Hubbard squash	1
Mushrooms	8	Jicama	1
Parsley	8	Peapods	1
Pinto beans	8	Pickles	1
Yams	7	Soybeans	1

*Indicates items that appear on both "fruit" and "vegetable" lists.