Monitoring & Improving Quality via Consumer Complaints

A Statistical Engineering Case Study

William A. Brenneman

Research Fellow, Global Statistics Discipline Leader
Data and Modeling Sciences
The Procter & Gamble Company

Adjunct Professor of Practice
Stewart School of Industrial and Systems Engineering
Georgia Tech University

Fall Technical Conference
West Palm Beach, Florida
October 5, 2018

Outline

- Introduction – size and complexity of problem
- Solution Overview
- MGPS Algorithm for Monitoring Complaints
- Improving Data Quality through Machine Learning
- Deployment and Evolution of the Application
- Conclusion

P&G Products

IVORY
Pampers
Baby Dry
Crest
ADVANCED CLEANING
Oral-B
Secret
MVP
Gillette
FUSION
BRAUN
The Gillette Company
Puffs
Pepto-Bismol
Charmin
VICKS 44
VICKS VapoRub
VICKS Vapo Steam
VICKS Sinex
VICKS NyQuil
VICKS DayQuil
NyQuil Cough
Tide
DAWN
original scent
Cascade
OLAY
regenerist
UV defense regenerating lotion
head & shoulders
h&s
Old Spice
Bounty
Designer's Touch

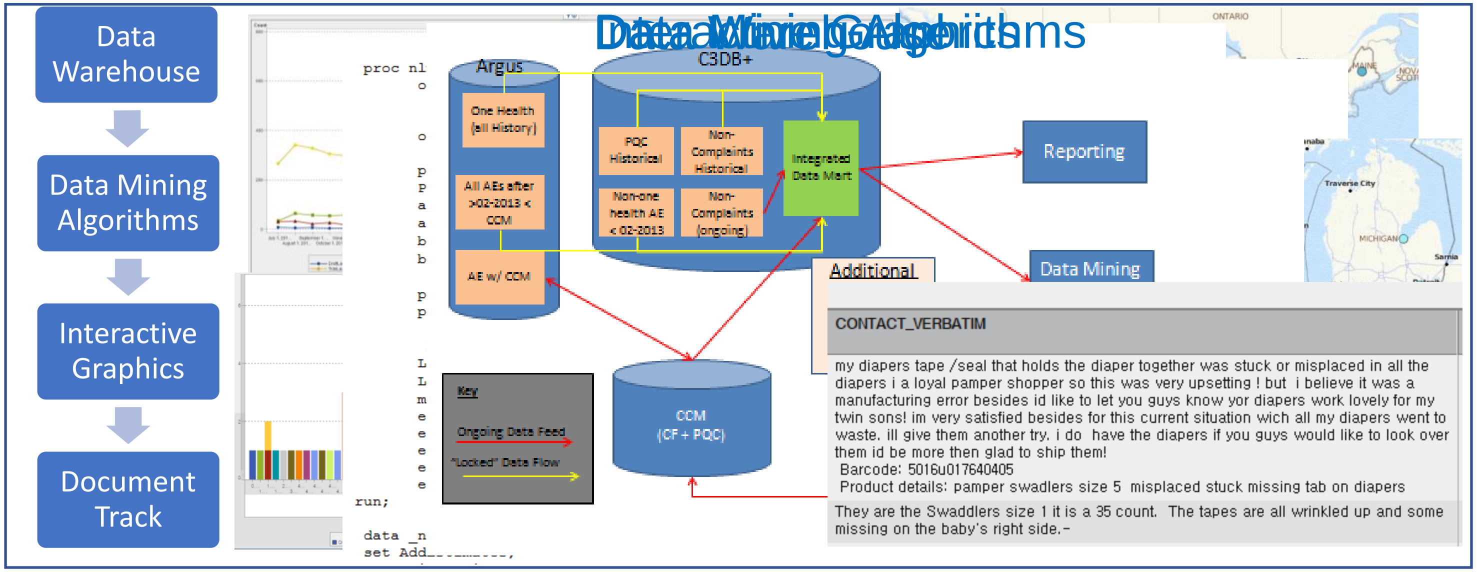
Consumer comments Examples

- Hi. I have a 2 year old daughter who I have only been able to use Pampers diapers for. Every other brand I have tried has caused her to get a rash.
- Do you use Kathon CG, DP, LX, UT, WT; CI+Me-isothiazolinone or Methylchloroisothiazone in your Free and Clear detergent. I am allergic to those chemicals/preservatives and have horrible skin reactions to them. Thanks for your response.
- I bought a huge pack of Ultra strong and I usually buy the Ultra Soft but the strong is causing irritation.

Phone, Email, P&G Websites/Social Media, Letters

Data mining for Signals in Consumer Complaints

- Signal Detection leads to improved products & greater compliance
- P&G receives about 2 million complaints/year globally
- Old System: data silos, different algorithms, lots of data prep
- Large multi-functional team brought together – under Director Leadership
- New System: automated data retrieval, data mining, and document tracking



```
proc nl
    o

    o

    p
    p
    a
    a
    b
    b

    p
    p

    L
    L
    m
    e
    e
    e
    e
    e
run;

 data _n
 set Add__.____,
```

CONTACT_VERBATIM

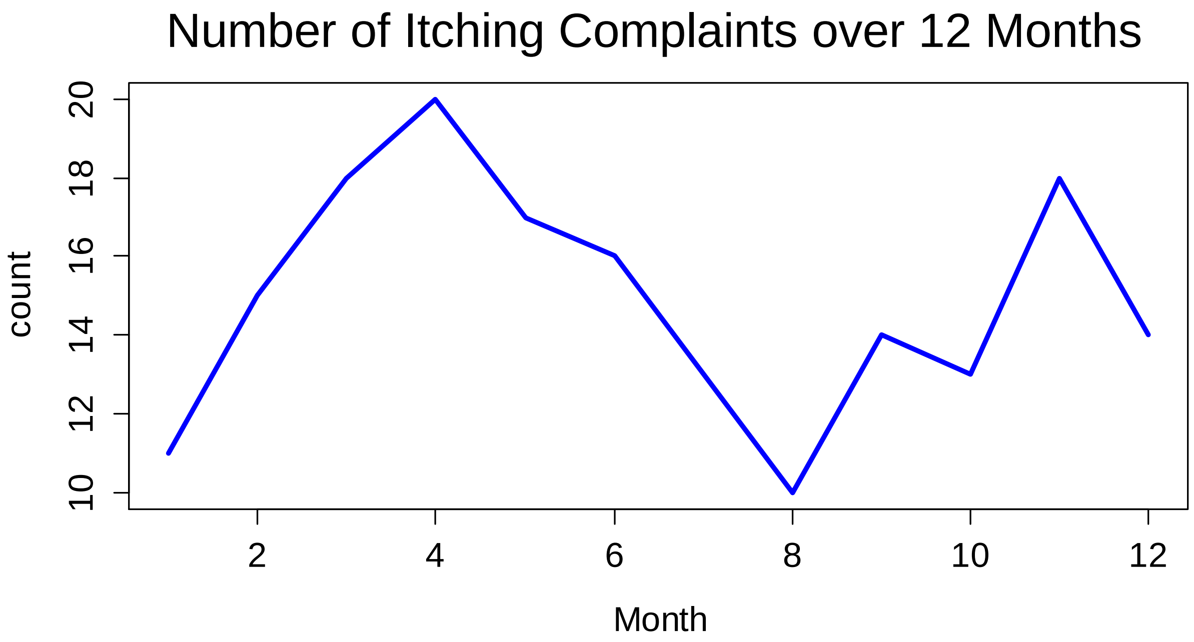
my diapers tape /seal that holds the diaper together was stuck or misplaced in all the diapers i a loyal pamper shopper so this was very upsetting ! but i believe it was a manufacturing error besides id like to let you guys know yor diapers work lovely for my twin sons! im very satisfied besides for this current situation wich all my diapers went to waste. ill give them another try. i do have the diapers if you guys would like to look over them id be more then glad to ship them!
 Barcode: 5016u017640405
 Product details: pamper swadlers size 5 misplaced stuck missing tab on diapers

They are the Swaddlers size 1 it is a 35 count. The tapes are all wrinkled up and some missing on the baby's right side.-

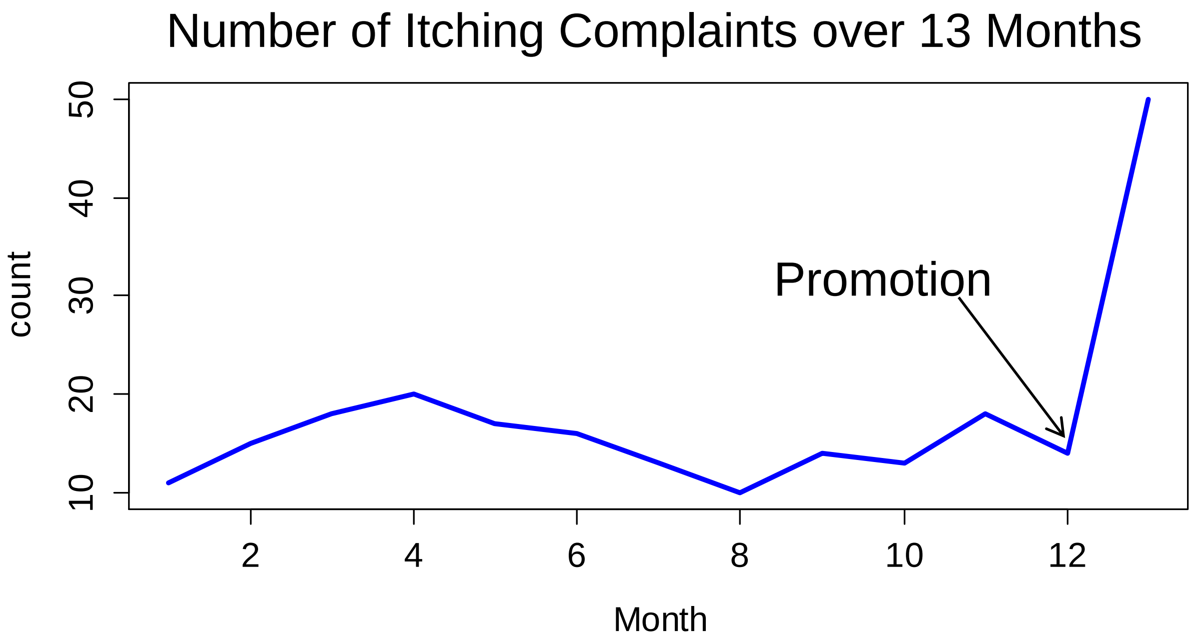
Data Mining Algorithm
Gamma Poisson Shrinker

- W. DuMouchel
- Externally recognized methodology
- Developed to detect adverse drug reactions in prescription drugs

Complaint Data for Lotion



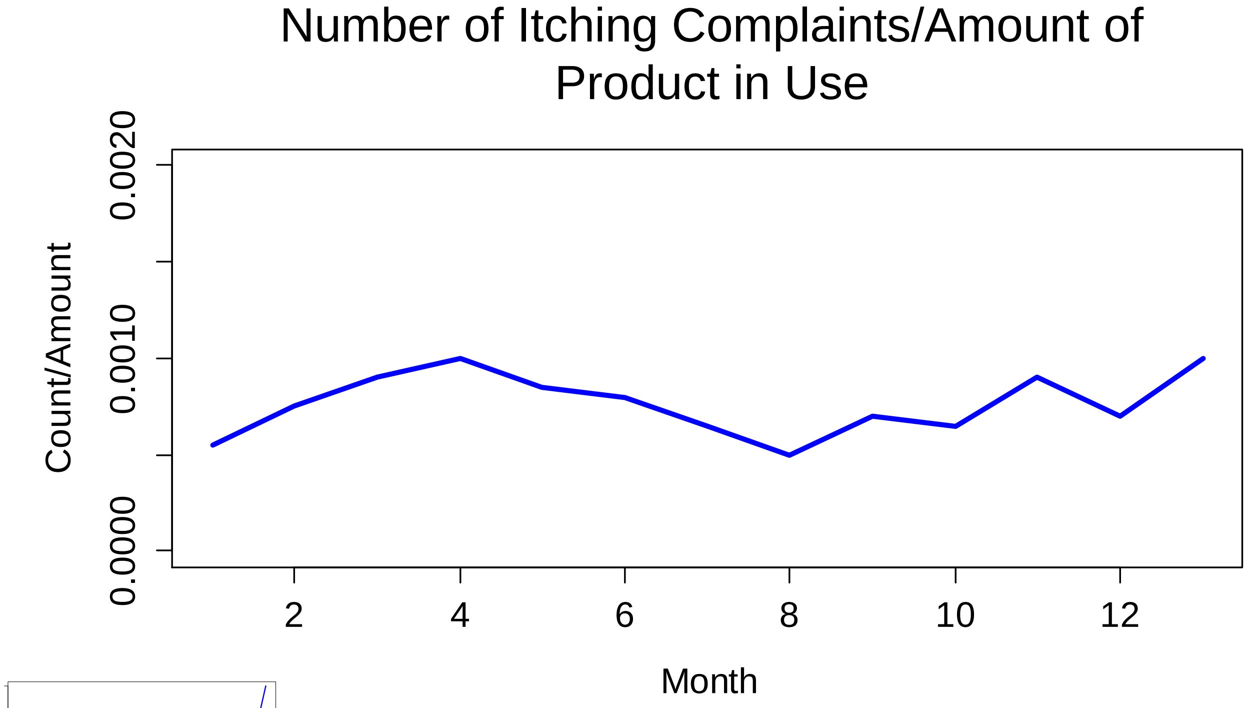
Example Data for Lotion

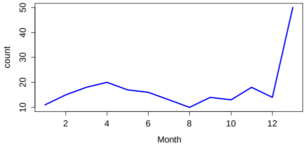


Cause of Spike

- More people using the product
  - Advertising campaign, free samples
- Imagine we knew the amount of product in use:
  - 20000 during the first 12 months
  - 50000 during the last month
- We can normalize the data (complaints/amount)

Known Amount in Use





Problem

- We do not know the amount of product in use at any given time
- Known as the *Denominator Problem*
  - No way to normalize the data
  - (# of complaints)/(?)
- If we know the denominator, signal detection is easy
  - Control Charts work well

Solution

- Look at *all* complaints relating to a product
  - Previous method looks at one complaint category at a time

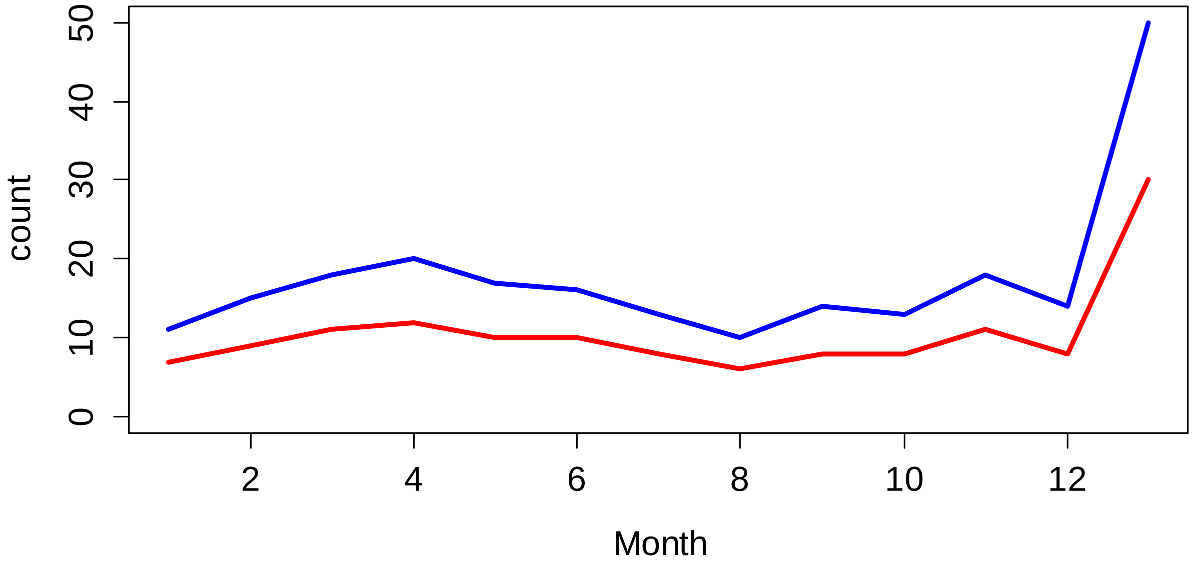
- Are the complaints following the same trend?

- More product in use = more complaints for all categories

Disproportionate Analysis

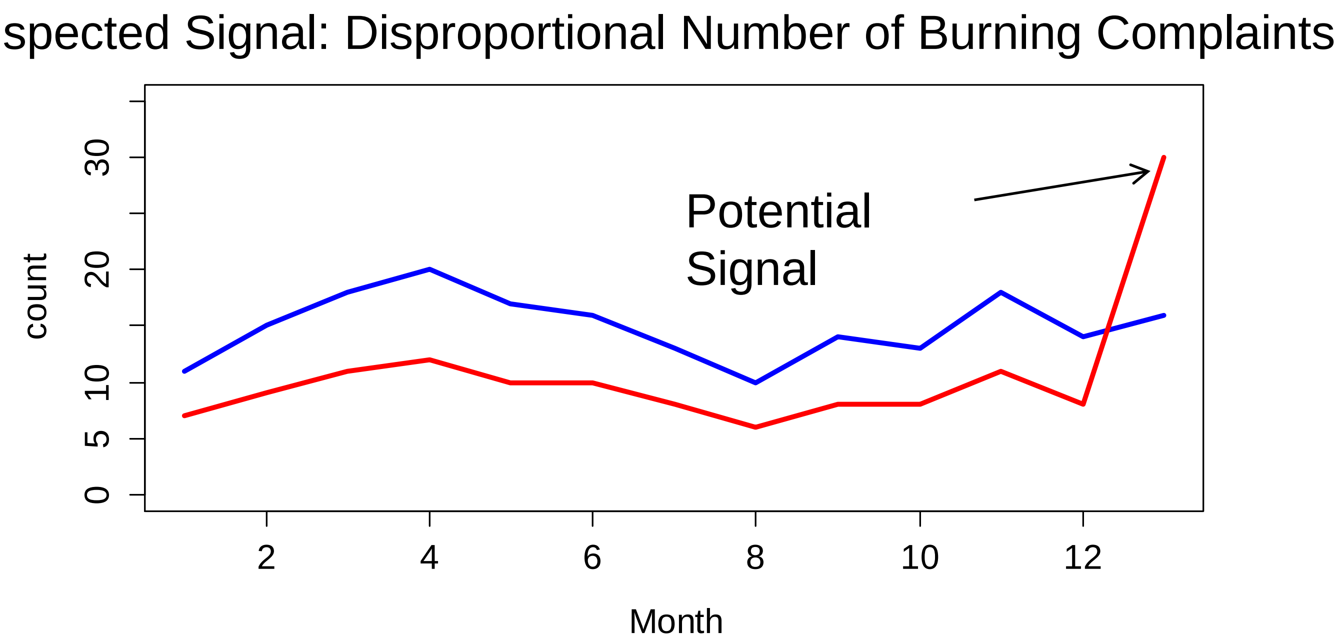
Example Data for Lotion

Number of Itching (blue) and Burning (red) Complaints



- Both complaints follow same trend
- They move proportionally to each other

Example Data for Lotion



- The trend has shifted
- Complaints are now disproportional

Disproportionality Analysis

- No need for outside data
- Detects interesting shifts in the data
    - Reveals shifts in proportionality
- Calculates an *Internal Denominator*
    - (# of complaints)/(# of expected complaints)

Example Data Set

- 24 months of complaint data for Product X

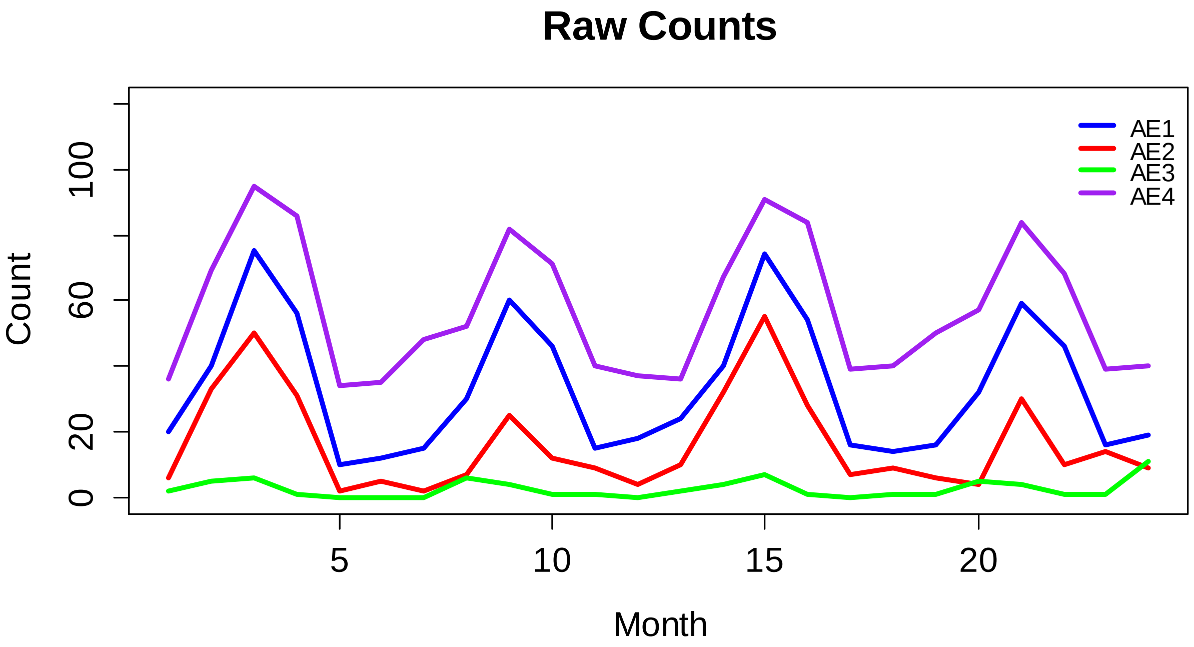
- 4 Adverse Events (AEs)

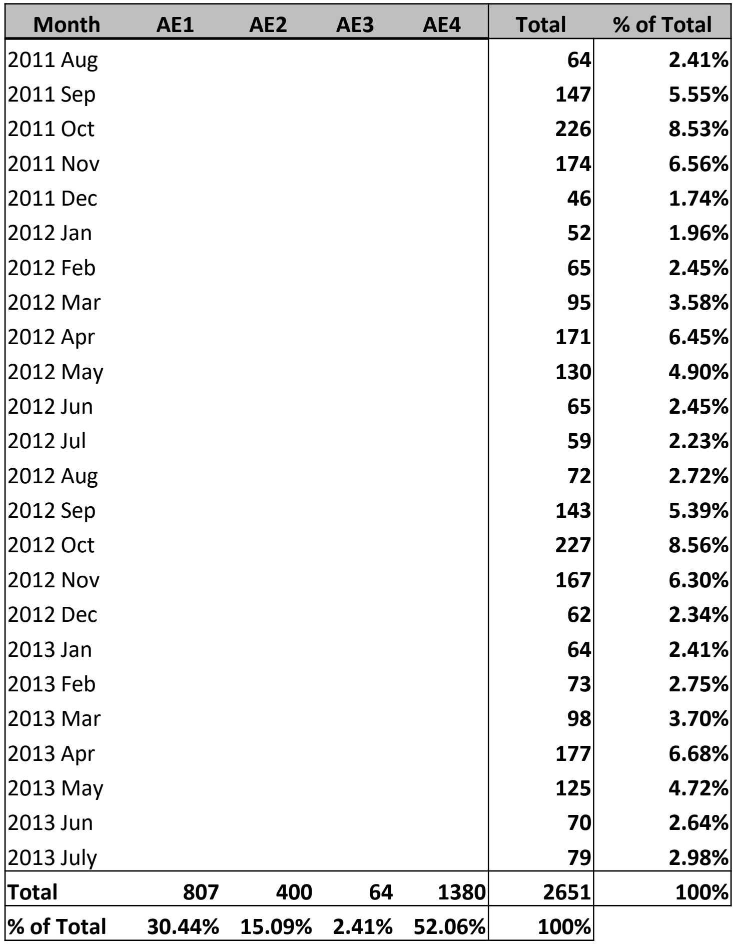
- Goal:
  - Calculate how the data would look if it were perfectly proportional

Table of Complaint Counts

Month	AE1	AE2	AE3	AE4	Total	% of Total
2011 Aug	20	6	2	36	**64**	**2.41%**
2011 Sep	40	33	5	69	**147**	**5.55%**
2011 Oct	75	50	6	95	**226**	**8.53%**
2011 Nov	56	31	1	86	**174**	**6.56%**
2011 Dec	10	2	0	34	**46**	**1.74%**
2012 Jan	12	5	0	35	**52**	**1.96%**
2012 Feb	15	2	0	48	**65**	**2.45%**
2012 Mar	30	7	6	52	**95**	**3.58%**
2012 Apr	60	25	4	82	**171**	**6.45%**
2012 May	46	12	1	71	**130**	**4.90%**
2012 Jun	15	9	1	40	**65**	**2.45%**
2012 Jul	18	4	0	37	**59**	**2.23%**
2012 Aug	24	10	2	36	**72**	**2.72%**
2012 Sep	40	32	4	67	**143**	**5.39%**
2012 Oct	74	55	7	91	**227**	**8.56%**
2012 Nov	54	28	1	84	**167**	**6.30%**
2012 Dec	16	7	0	39	**62**	**2.34%**
2013 Jan	14	9	1	40	**64**	**2.41%**
2013 Feb	16	6	1	50	**73**	**2.75%**
2013 Mar	32	4	5	57	**98**	**3.70%**
2013 Apr	59	30	4	84	**177**	**6.68%**
2013 May	46	10	1	68	**125**	**4.72%**
2013 Jun	16	14	1	39	**70**	**2.64%**
2013 July	19	9	11	40	**79**	**2.98%**
Total	**807**	**400**	**64**	**1380**	**2651**	**100%**
% of Total	**30.44%**	**15.09%**	**2.41%**	**52.06%**	**100%**	

Complaint Counts for Product X





Month	AE1	AE2	AE3	AE4	Total	% of Total
2011 Aug					**64**	**2.41%**
2011 Sep					**147**	**5.55%**
2011 Oct					**226**	**8.53%**
2011 Nov					**174**	**6.56%**
2011 Dec					**46**	**1.74%**
2012 Jan					**52**	**1.96%**
2012 Feb					**65**	**2.45%**
2012 Mar					**95**	**3.58%**
2012 Apr					**171**	**6.45%**
2012 May					**130**	**4.90%**
2012 Jun					**65**	**2.45%**
2012 Jul					**59**	**2.23%**
2012 Aug					**72**	**2.72%**
2012 Sep					**143**	**5.39%**
2012 Oct					**227**	**8.56%**
2012 Nov					**167**	**6.30%**
2012 Dec					**62**	**2.34%**
2013 Jan					**64**	**2.41%**
2013 Feb					**73**	**2.75%**
2013 Mar					**98**	**3.70%**
2013 Apr					**177**	**6.68%**
2013 May					**125**	**4.72%**
2013 Jun					**70**	**2.64%**
2013 July					**79**	**2.98%**
Total	**807**	**400**	**64**	**1380**	**2651**	**100%**
% of Total	**30.44%**	**15.09%**	**2.41%**	**52.06%**	**100%**	

How would the data look if the overall proportions were consistent?

Month	AE1	AE2	AE3	AE4	Total	% of Total
2011 Aug					64	2.41%
2011 Sep					147	5.55%
2011 Oct					226	8.53%
2011 Nov					174	6.56%
2011 Dec					46	1.74%
2012 Jan					52	1.96%
2012 Feb					65	2.45%
2012 Mar					95	3.58%
2012 Apr					171	6.45%
2012 May					130	4.90%
2012 Jun					65	2.45%
2012 Jul					59	2.23%
2012 Aug					72	2.72%
2012 Sep					143	5.39%
2012 Oct					227	8.56%
2012 Nov					167	6.30%
2012 Dec					62	2.34%
2013 Jan					64	2.41%
2013 Feb					73	2.75%
2013 Mar					98	3.70%
2013 Apr					177	6.68%
2013 May					125	4.72%
2013 Jun					70	2.64%
2013 July					79	2.98%
Total	807	400	64	1380	2651	100%
% of Total	30.44%	15.09%	2.41%	52.06%	100%	

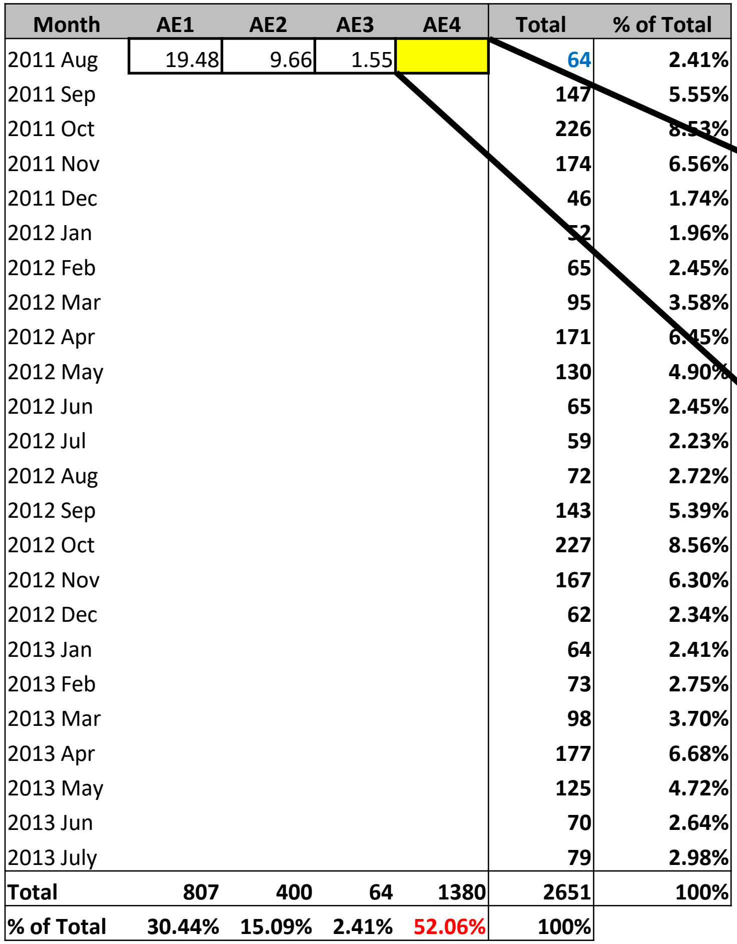
64* 0.3044 =
19.48

Month	AE1	AE2	AE3	AE4	Total	% of Total
2011 Aug	19.48				64	2.41%
2011 Sep					147	5.55%
2011 Oct					226	8.53%
2011 Nov					174	6.56%
2011 Dec					46	1.74%
2012 Jan					52	1.96%
2012 Feb					65	2.45%
2012 Mar					95	3.58%
2012 Apr					171	6.45%
2012 May					130	4.90%
2012 Jun					65	2.45%
2012 Jul					59	2.23%
2012 Aug					72	2.72%
2012 Sep					143	5.39%
2012 Oct					227	8.56%
2012 Nov					167	6.30%
2012 Dec					62	2.34%
2013 Jan					64	2.41%
2013 Feb					73	2.75%
2013 Mar					98	3.70%
2013 Apr					177	6.68%
2013 May					125	4.72%
2013 Jun					70	2.64%
2013 July					79	2.98%
Total	807	400	64	1380	2651	100%
% of Total	30.44%	15.09%	2.41%	52.06%	100%	

**64* 0.1509 =
9.66**

Month	AE1	AE2	AE3	AE4	Total	% of Total
2011 Aug	19.48	9.66			64	2.41%
2011 Sep					147	5.55%
2011 Oct					226	8.53%
2011 Nov					174	6.56%
2011 Dec					46	1.74%
2012 Jan					52	1.96%
2012 Feb					65	2.45%
2012 Mar					95	3.58%
2012 Apr					171	6.45%
2012 May					130	4.90%
2012 Jun					65	2.45%
2012 Jul					59	2.23%
2012 Aug					72	2.72%
2012 Sep					143	5.39%
2012 Oct					227	8.56%
2012 Nov					167	6.30%
2012 Dec					62	2.34%
2013 Jan					64	2.41%
2013 Feb					73	2.75%
2013 Mar					98	3.70%
2013 Apr					177	6.68%
2013 May					125	4.72%
2013 Jun					70	2.64%
2013 July					79	2.98%
Total	807	400	64	1380	2651	100%
% of Total	30.44%	15.09%	2.41%	52.06%	100%	

**64* 0.0241 =
1.55**



Month	AE1	AE2	AE3	AE4	Total	% of Total
2011 Aug	19.48	9.66	1.55		64	2.41%
2011 Sep					147	5.55%
2011 Oct					226	8.53%
2011 Nov					174	6.56%
2011 Dec					46	1.74%
2012 Jan					52	1.96%
2012 Feb					65	2.45%
2012 Mar					95	3.58%
2012 Apr					171	6.45%
2012 May					130	4.90%
2012 Jun					65	2.45%
2012 Jul					59	2.23%
2012 Aug					72	2.72%
2012 Sep					143	5.39%
2012 Oct					227	8.56%
2012 Nov					167	6.30%
2012 Dec					62	2.34%
2013 Jan					64	2.41%
2013 Feb					73	2.75%
2013 Mar					98	3.70%
2013 Apr					177	6.68%
2013 May					125	4.72%
2013 Jun					70	2.64%
2013 July					79	2.98%
Total	**807**	**400**	**64**	**1380**	**2651**	**100%**
% of Total	**30.44%**	**15.09%**	**2.41%**	**52.06%**	**100%**	

64* 0.5206 =
33.32

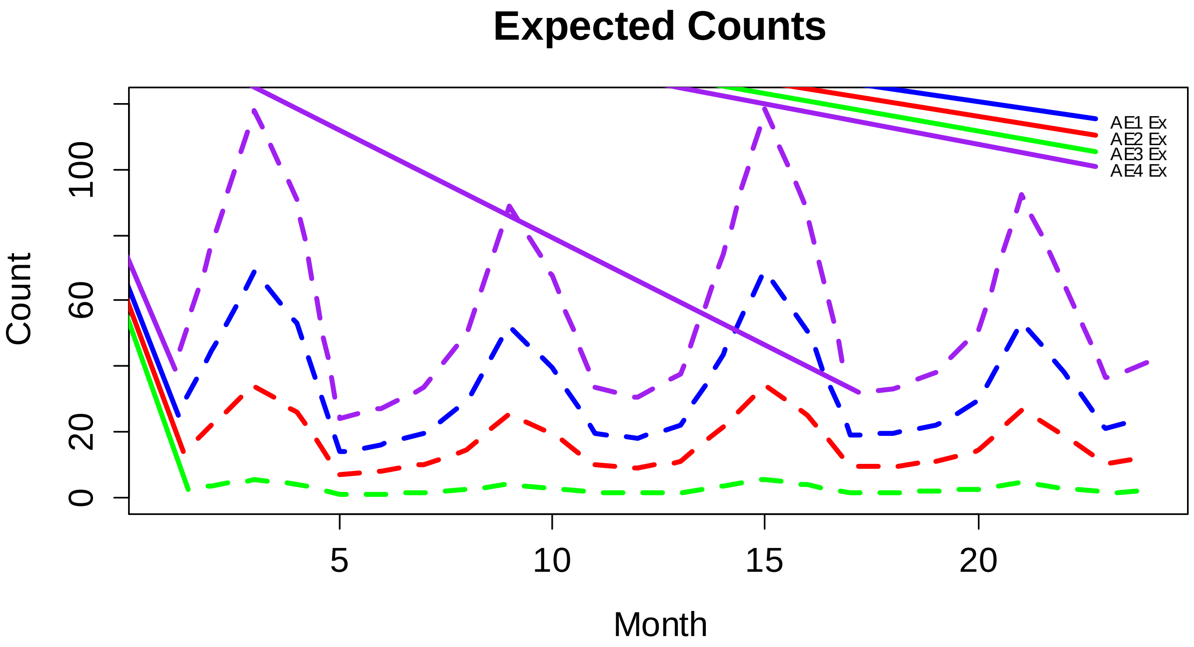
Month	AE1	AE2	AE3	AE4	Total	% of Total
2011 Aug	19.48	9.66	1.55	33.32	64	2.41%
2011 Sep					147	5.55%
2011 Oct					226	8.53%
2011 Nov					174	6.56%
2011 Dec					46	1.74%
2012 Jan					52	1.96%
2012 Feb					65	2.45%
2012 Mar					95	3.58%
2012 Apr					171	6.45%
2012 May					130	4.90%
2012 Jun					65	2.45%
2012 Jul					59	2.23%
2012 Aug					72	2.72%
2012 Sep					143	5.39%
2012 Oct					227	8.56%
2012 Nov					167	6.30%
2012 Dec					62	2.34%
2013 Jan					64	2.41%
2013 Feb					73	2.75%
2013 Mar					98	3.70%
2013 Apr					177	6.68%
2013 May					125	4.72%
2013 Jun					70	2.64%
2013 July					79	2.98%
Total	807	400	64	1380	2651	100%
% of Total	30.44%	15.09%	2.41%	52.06%	100%	

147*.3044 =
44.75

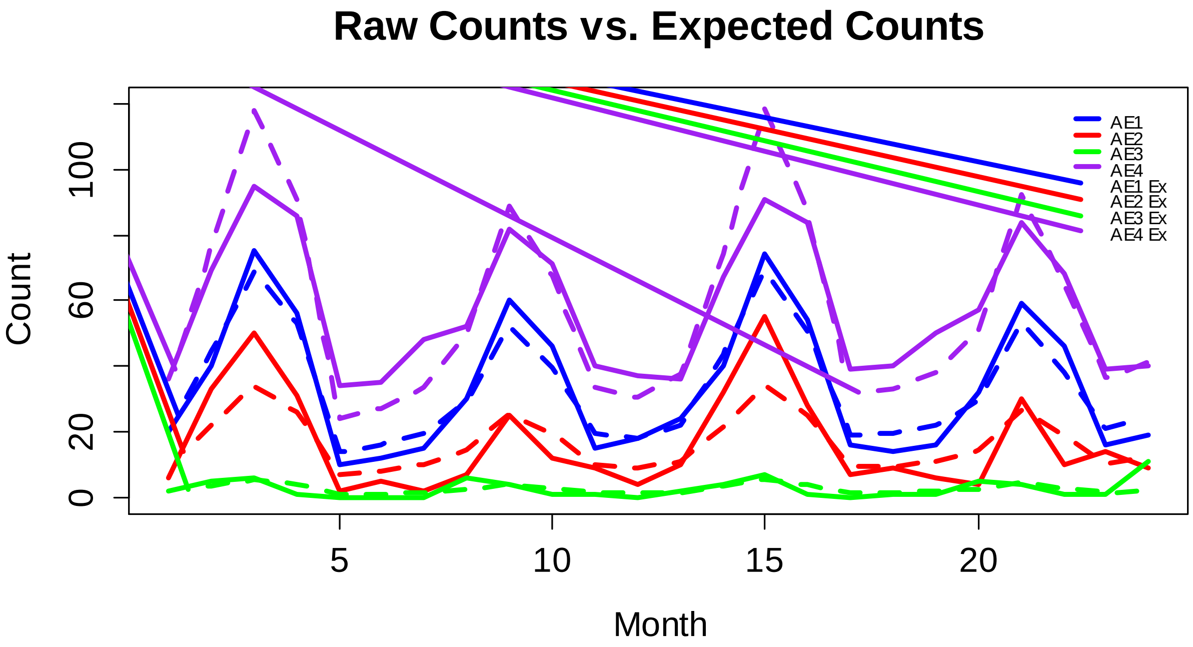
Table of Expected Counts

Month	AE1	AE2	AE3	AE4	Total	% of Total
2011 Aug	19.48	9.66	1.55	33.32	**64**	**2.41%**
2011 Sep	44.75	22.18	3.55	76.52	**147**	**5.55%**
2011 Oct	68.80	34.10	5.46	117.65	**226**	**8.53%**
2011 Nov	52.97	26.25	4.20	90.58	**174**	**6.56%**
2011 Dec	14.00	6.94	1.11	23.95	**46**	**1.74%**
2012 Jan	15.83	7.85	1.26	27.07	**52**	**1.96%**
2012 Feb	19.79	9.81	1.57	33.84	**65**	**2.45%**
2012 Mar	28.92	14.33	2.29	49.45	**95**	**3.58%**
2012 Apr	52.05	25.80	4.13	89.02	**171**	**6.45%**
2012 May	39.57	19.62	3.14	67.67	**130**	**4.90%**
2012 Jun	19.79	9.81	1.57	33.84	**65**	**2.45%**
2012 Jul	17.96	8.90	1.42	30.71	**59**	**2.23%**
2012 Aug	21.92	10.86	1.74	37.48	**72**	**2.72%**
2012 Sep	43.53	21.58	3.45	74.44	**143**	**5.39%**
2012 Oct	69.10	34.25	5.48	118.17	**227**	**8.56%**
2012 Nov	50.84	25.20	4.03	86.93	**167**	**6.30%**
2012 Dec	18.87	9.35	1.50	32.27	**62**	**2.34%**
2013 Jan	19.48	9.66	1.55	33.32	**64**	**2.41%**
2013 Feb	22.22	11.01	1.76	38.00	**73**	**2.75%**
2013 Mar	29.83	14.79	2.37	51.01	**98**	**3.70%**
2013 Apr	53.88	26.71	4.27	92.14	**177**	**6.68%**
2013 May	38.05	18.86	3.02	65.07	**125**	**4.72%**
2013 Jun	21.31	10.56	1.69	36.44	**70**	**2.64%**
2013 July	24.05	11.92	1.91	41.12	**79**	**2.98%**
Total	**807**	**400**	**64**	**1380**	**2651**	**100%**
% of Total	**30.44%**	**15.09%**	**2.41%**	**52.06%**	**100%**	

Expected Complaint Counts



What Did We See vs. What Did We Expect?



Month	AE1	AE2	AE3	AE4
2011 Aug				
2011 Sep				
2011 Oct				
2011 Nov				
2011 Dec				
2012 Jan				
2012 Feb				
2012 Mar				
2012 Apr				
2012 May				
2012 Jun				
2012 Jul				
2012 Aug				
2012 Sep				
2012 Oct				
2012 Nov				
2012 Dec				
2013 Jan				
2013 Feb				
2013 Mar				
2013 Apr				
2013 May				
2013 Jun				
2013 July				

- How different are the actual data and the expected data?
- If the numbers are close, the ratio (# of complaints)/(# of expected complaints) *should be near 1.*
- Let N_{ij} be the number of complaints
- Let E_{ij} be the number of complaints we expect
- $RR_{ij}= N_{ij}/E_{ij}$
- RR is the Relative Reporting Ratio

Month	AE1	AE2	AE3	AE4
2011 Aug				
2011 Sep				
2011 Oct				
2011 Nov				
2011 Dec				
2012 Jan				
2012 Feb				
2012 Mar				
2012 Apr				
2012 May				
2012 Jun				
2012 Jul				
2012 Aug				
2012 Sep				
2012 Oct				
2012 Nov				
2012 Dec				
2013 Jan				
2013 Feb				
2013 Mar				
2013 Apr				
2013 May				
2013 Jun				
2013 July				

20/19.48 = 1.03

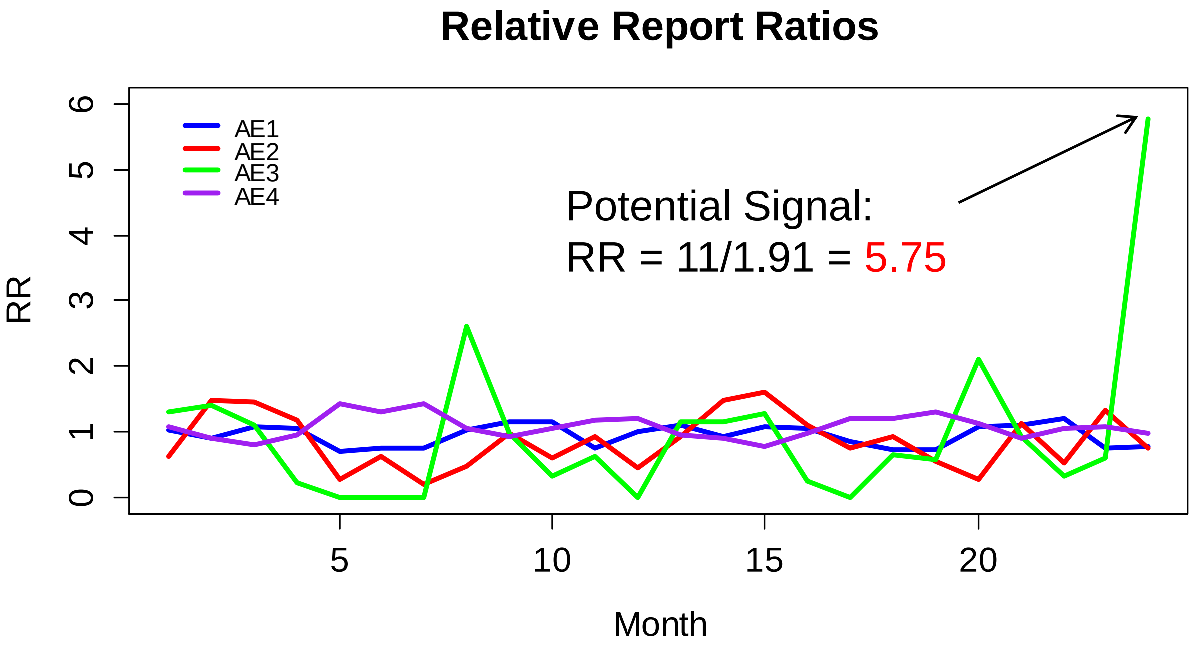
Month	AE1	AE2	AE3	AE4
2011 Aug	1.03			
2011 Sep				
2011 Oct				
2011 Nov				
2011 Dec				
2012 Jan				
2012 Feb				
2012 Mar				
2012 Apr				
2012 May				
2012 Jun				
2012 Jul				
2012 Aug				
2012 Sep				
2012 Oct				
2012 Nov				
2012 Dec				
2013 Jan				
2013 Feb				
2013 Mar				
2013 Apr				
2013 May				
2013 Jun				
2013 July				

6/9.66 = 0.62

Table of Relative Reporting Ratios

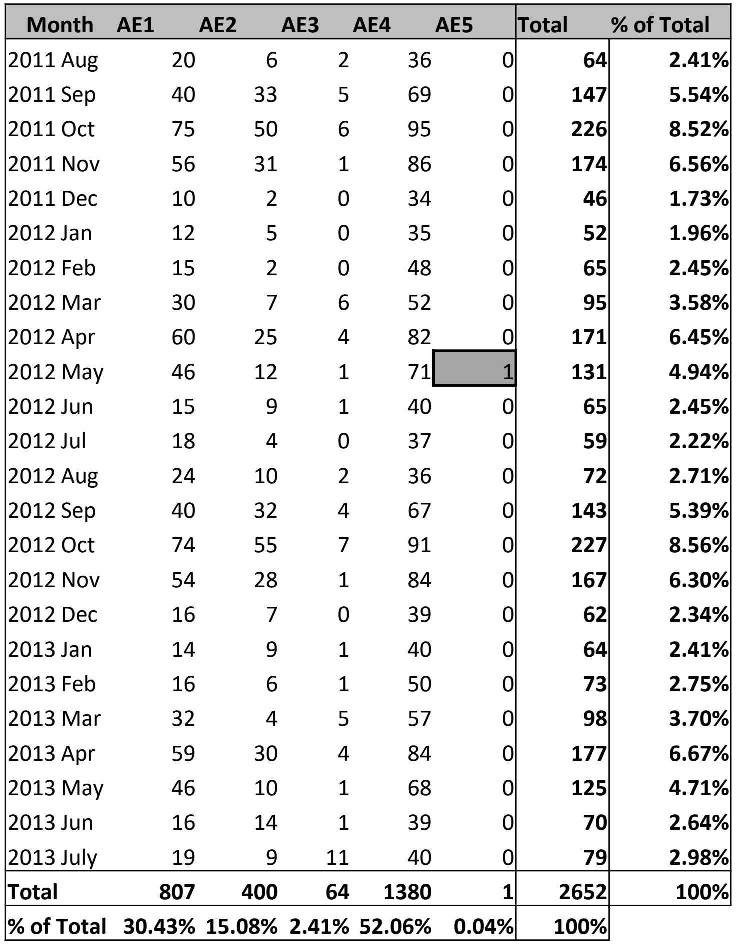
Month	AE1	AE2	AE3	AE4
2011 Aug	1.03	0.62	1.29	1.08
2011 Sep	0.89	1.49	1.41	0.90
2011 Oct	1.09	1.47	1.10	0.81
2011 Nov	1.06	1.18	0.24	0.95
2011 Dec	0.71	0.29	0.00	1.42
2012 Jan	0.76	0.64	0.00	1.29
2012 Feb	0.76	0.20	0.00	1.42
2012 Mar	1.04	0.49	2.62	1.05
2012 Apr	1.15	0.97	0.97	0.92
2012 May	1.16	0.61	0.32	1.05
2012 Jun	0.76	0.92	0.64	1.18
2012 Jul	1.00	0.45	0.00	1.20
2012 Aug	1.10	0.92	1.15	0.96
2012 Sep	0.92	1.48	1.16	0.90
2012 Oct	1.07	1.61	1.28	0.77
2012 Nov	1.06	1.11	0.25	0.97
2012 Dec	0.85	0.75	0.00	1.21
2013 Jan	0.72	0.93	0.65	1.20
2013 Feb	0.72	0.54	0.57	1.32
2013 Mar	1.07	0.27	2.11	1.12
2013 Apr	1.10	1.12	0.94	0.91
2013 May	1.21	0.53	0.33	1.05
2013 Jun	0.75	1.33	0.59	1.07
2013 July	0.79	0.76	5.77	0.97

RRs for Product X



Potential Signal:
RR = 11/1.91 = 5.75

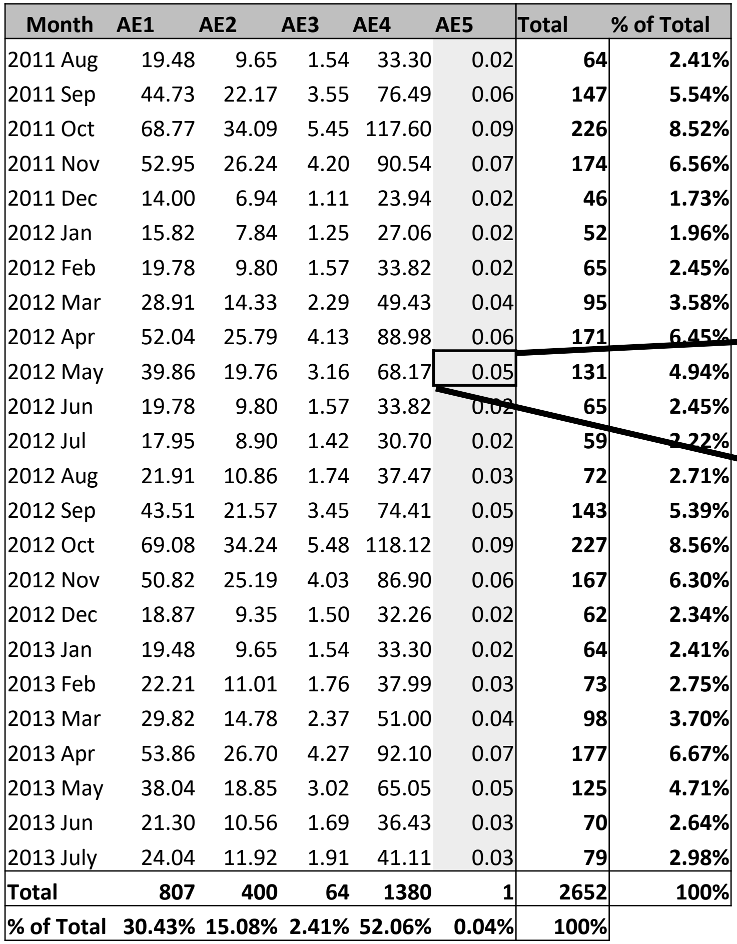
Table of Complaint Counts



Month	AE1	AE2	AE3	AE4	AE5	Total	% of Total
2011 Aug	20	6	2	36	0	**64**	**2.41%**
2011 Sep	40	33	5	69	0	**147**	**5.54%**
2011 Oct	75	50	6	95	0	**226**	**8.52%**
2011 Nov	56	31	1	86	0	**174**	**6.56%**
2011 Dec	10	2	0	34	0	**46**	**1.73%**
2012 Jan	12	5	0	35	0	**52**	**1.96%**
2012 Feb	15	2	0	48	0	**65**	**2.45%**
2012 Mar	30	7	6	52	0	**95**	**3.58%**
2012 Apr	60	25	4	82	0	**171**	**6.45%**
2012 May	46	12	1	71	1	**131**	**4.94%**
2012 Jun	15	9	1	40	0	**65**	**2.45%**
2012 Jul	18	4	0	37	0	**59**	**2.22%**
2012 Aug	24	10	2	36	0	**72**	**2.71%**
2012 Sep	40	32	4	67	0	**143**	**5.39%**
2012 Oct	74	55	7	91	0	**227**	**8.56%**
2012 Nov	54	28	1	84	0	**167**	**6.30%**
2012 Dec	16	7	0	39	0	**62**	**2.34%**
2013 Jan	14	9	1	40	0	**64**	**2.41%**
2013 Feb	16	6	1	50	0	**73**	**2.75%**
2013 Mar	32	4	5	57	0	**98**	**3.70%**
2013 Apr	59	30	4	84	0	**177**	**6.67%**
2013 May	46	10	1	68	0	**125**	**4.71%**
2013 Jun	16	14	1	39	0	**70**	**2.64%**
2013 July	19	9	11	40	0	**79**	**2.98%**
Total	**807**	**400**	**64**	**1380**	**1**	**2652**	**100%**
% of Total	**30.43%**	**15.08%**	**2.41%**	**52.06%**	**0.04%**	**100%**	

- Problem with RR

Table of Expected Counts



Month	AE1	AE2	AE3	AE4	AE5	Total	% of Total
2011 Aug	19.48	9.65	1.54	33.30	0.02	**64**	**2.41%**
2011 Sep	44.73	22.17	3.55	76.49	0.06	**147**	**5.54%**
2011 Oct	68.77	34.09	5.45	117.60	0.09	**226**	**8.52%**
2011 Nov	52.95	26.24	4.20	90.54	0.07	**174**	**6.56%**
2011 Dec	14.00	6.94	1.11	23.94	0.02	**46**	**1.73%**
2012 Jan	15.82	7.84	1.25	27.06	0.02	**52**	**1.96%**
2012 Feb	19.78	9.80	1.57	33.82	0.02	**65**	**2.45%**
2012 Mar	28.91	14.33	2.29	49.43	0.04	**95**	**3.58%**
2012 Apr	52.04	25.79	4.13	88.98	0.06	**171**	**6.45%**
2012 May	39.86	19.76	3.16	68.17	0.05	**131**	**4.94%**
2012 Jun	19.78	9.80	1.57	33.82	0.02	**65**	**2.45%**
2012 Jul	17.95	8.90	1.42	30.70	0.02	**59**	**2.22%**
2012 Aug	21.91	10.86	1.74	37.47	0.03	**72**	**2.71%**
2012 Sep	43.51	21.57	3.45	74.41	0.05	**143**	**5.39%**
2012 Oct	69.08	34.24	5.48	118.12	0.09	**227**	**8.56%**
2012 Nov	50.82	25.19	4.03	86.90	0.06	**167**	**6.30%**
2012 Dec	18.87	9.35	1.50	32.26	0.02	**62**	**2.34%**
2013 Jan	19.48	9.65	1.54	33.30	0.02	**64**	**2.41%**
2013 Feb	22.21	11.01	1.76	37.99	0.03	**73**	**2.75%**
2013 Mar	29.82	14.78	2.37	51.00	0.04	**98**	**3.70%**
2013 Apr	53.86	26.70	4.27	92.10	0.07	**177**	**6.67%**
2013 May	38.04	18.85	3.02	65.05	0.05	**125**	**4.71%**
2013 Jun	21.30	10.56	1.69	36.43	0.03	**70**	**2.64%**
2013 July	24.04	11.92	1.91	41.11	0.03	**79**	**2.98%**
Total	**807**	**400**	**64**	**1380**	**1**	**2652**	**100%**
% of Total	**30.43%**	**15.08%**	**2.41%**	**52.06%**	**0.04%**	**100%**	

RR = 1/0.05 = 20

Problems with RR

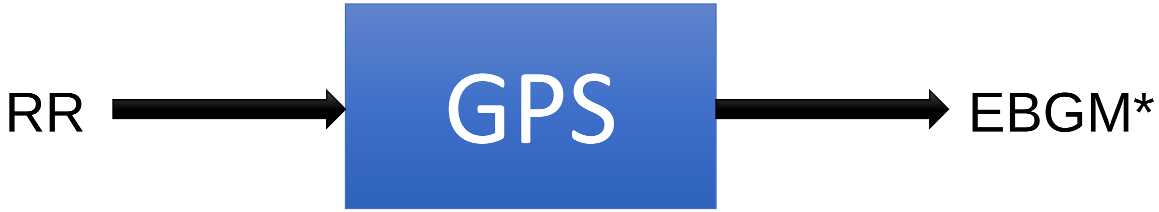
- Small expected counts cause high RR
- No adjustment for magnitude
  - 40 complaints, 2 are expected (RR = 20)
  - 1 complaint, 0.05 are expected (RR = 20)
  - Clearly the 40/2 event is more troubling

- *Multiplicity Problem*
  - Even with 95% C.I.s, we'll have many false alarms with 100,000 events

Solution

- Gamma Poisson Shrinker to the rescue!

- GPS corrects the problems with RR



*Empirical Bayes Geometric Mean

Let’s Take a Step Back

- Accept that complaint counts are *random*
  - We expect 5 and receive between 2-9
  - Think of your daily mail
  - N_{ij} ~ *Poisson*(μ_{ij})
  - 10 random draws from *Poisson*(5) gives (4,4,5,5,3,2,2,4,9,2)

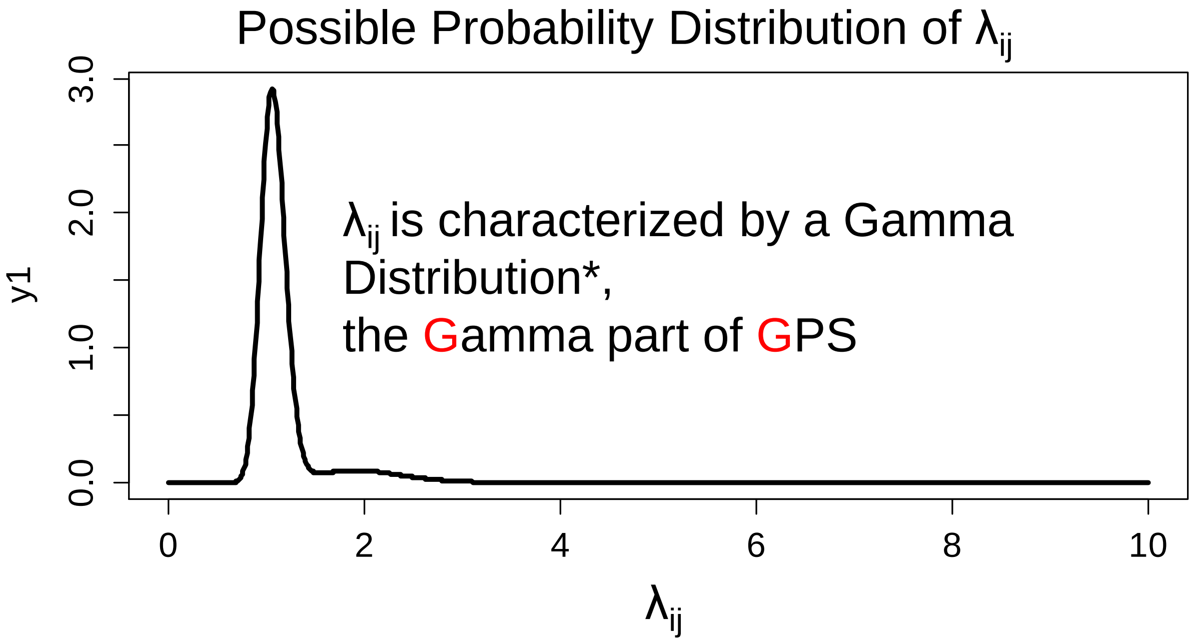
- This is the Poisson piece of GPS

What Do We Know? Forming a prior

- N_{ij} is random, so $RR_{ij} = N_{ij} / E_{ij}$ is random
- RR_{ij} is a random occurrence of the true underlying ratio $\lambda_{ij} = \mu_{ij} / E_{ij}$

- Nature of the world: we'll never know λ_{ij} with 100% certainty
- But we know some things before data is collected…
  - It's greater than 0
  - *It's likely near 1 (signifying a normal event)*
  - It has a small chance of being > 1 (bad signal) and small chance of being < 1 (good signal)

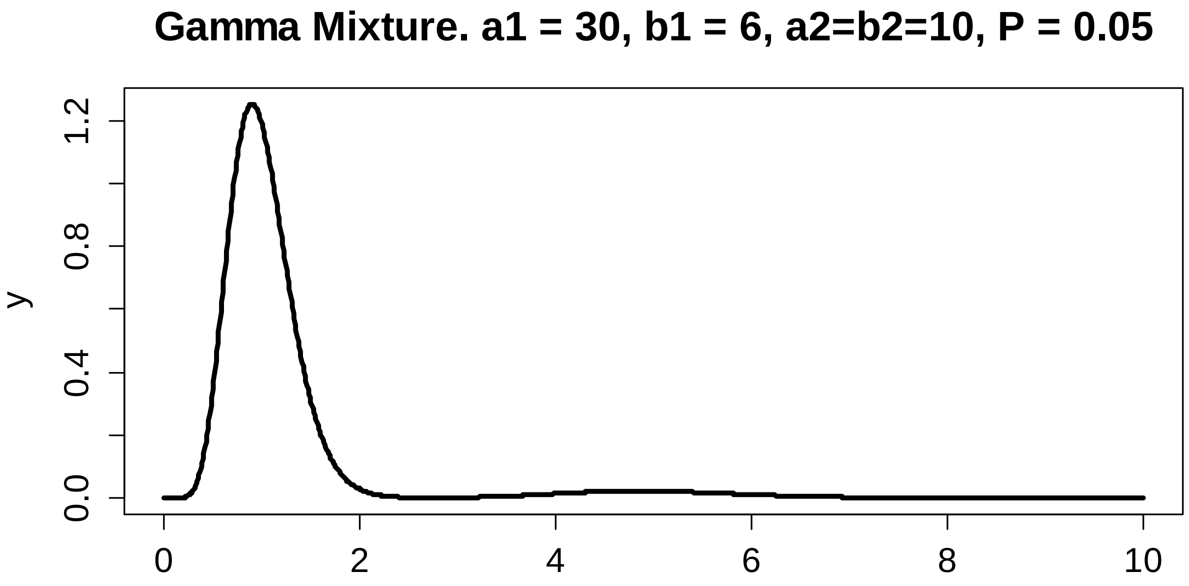
Prior Probability Distribution



Possible Probability Distribution of λ_{ij}

λ_{ij} is characterized by a Gamma Distribution*,
the Gamma part of GPS

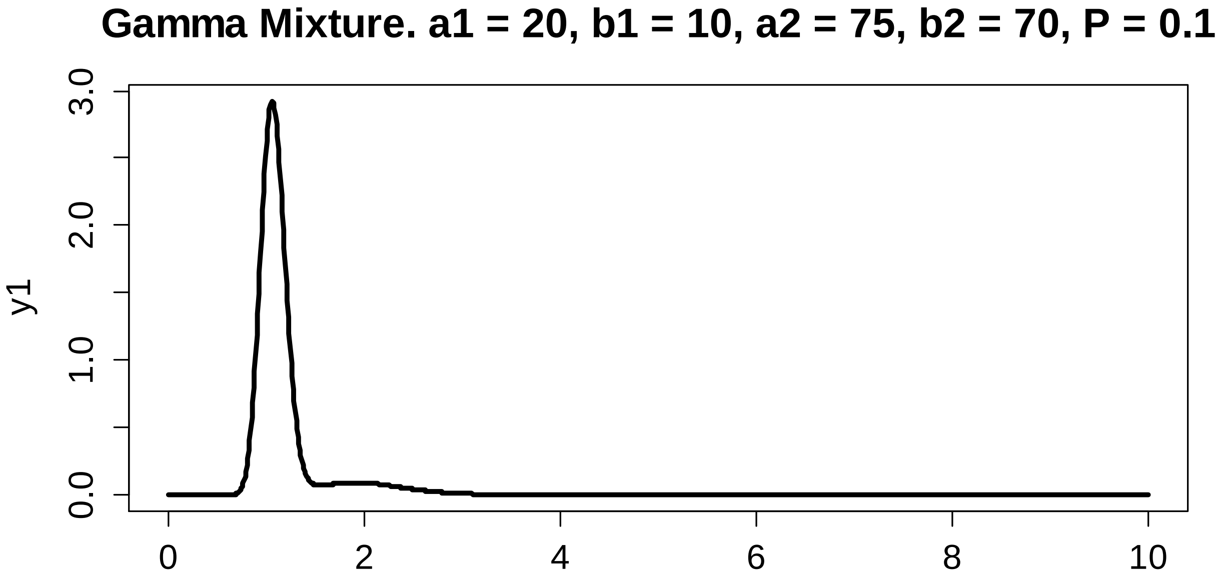
Density of Gamma Mixture (1)



$$\pi(\lambda;\ \alpha_1, \beta_1, \alpha_2, \beta_2, P) =\ Pg(\lambda;\ \alpha_1,\ \beta_1) +\ (1 - P)g(\lambda;\ \alpha_2,\ \beta_2)$$

$$g(\lambda; \alpha, \beta) = \frac{\beta^{\alpha}\,\lambda^{\alpha-1}e^{-\beta\lambda}}{\Gamma(\alpha)}$$

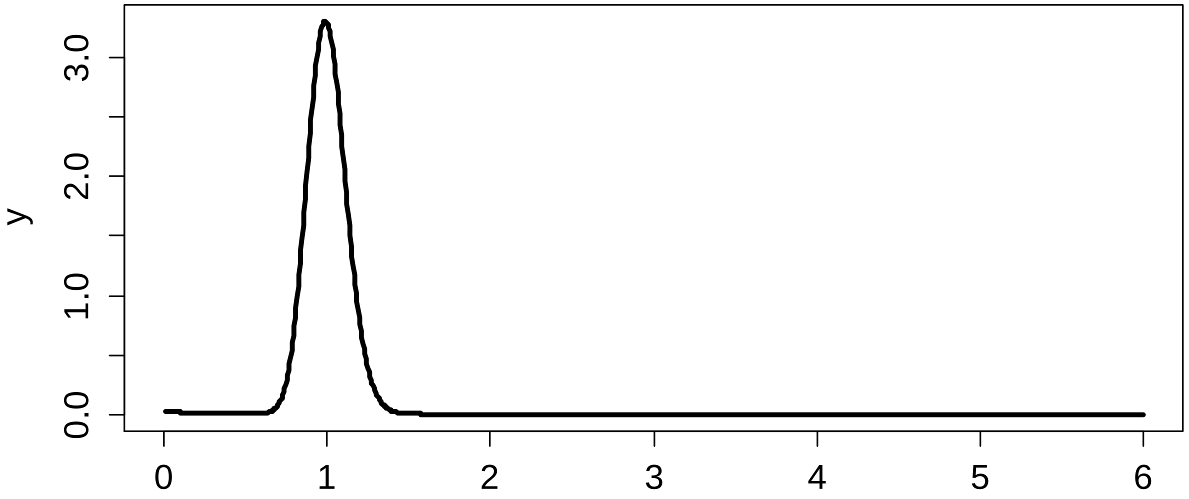
Density of Gamma Mixture (2)



$$\pi(\lambda;\ \alpha_1, \beta_1, \alpha_2, \beta_2, P) = \ Pg(\lambda;\ \alpha_1,\ \beta_1) + \ (1 - P)g(\lambda;\ \alpha_2,\ \beta_2)$$

Prior Information

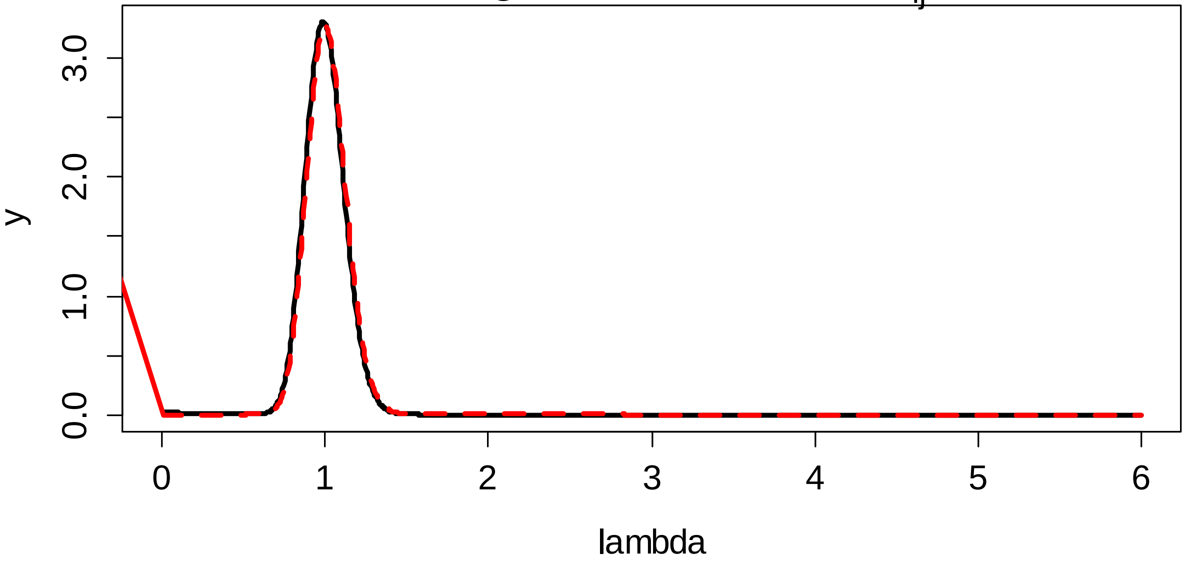
How will our prior knowledge of λ_{ij} be affected by $RR_{ij} = 5$



It depends

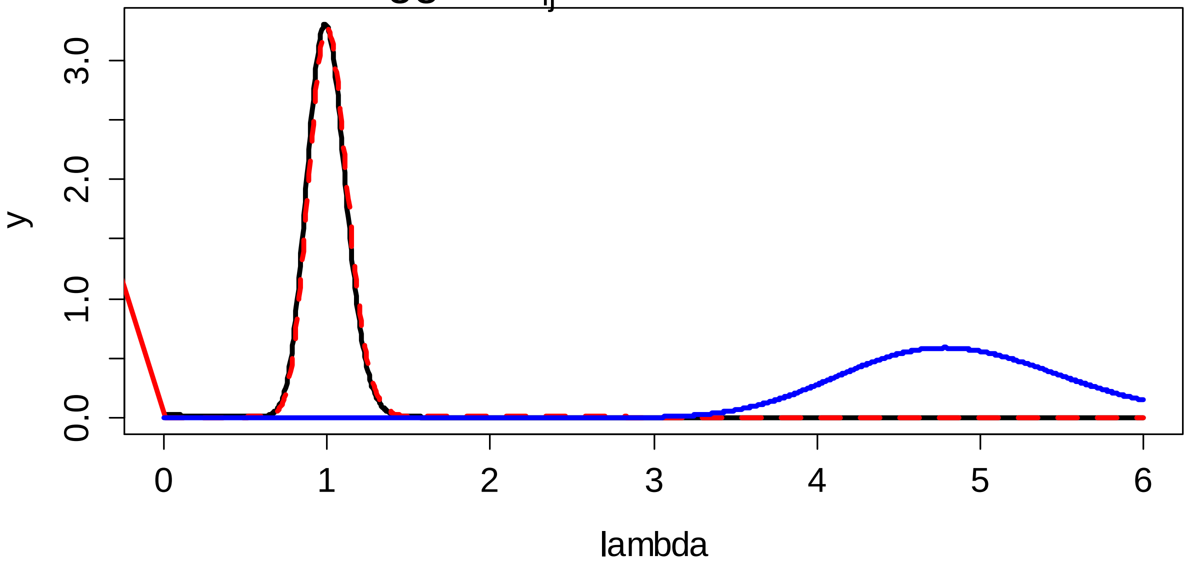
Event with $N_{ij} = 1, E_{ij} = 0.2$

RR_{ij} = 5 but there's not enough information to really change our estimate of λ_{ij}

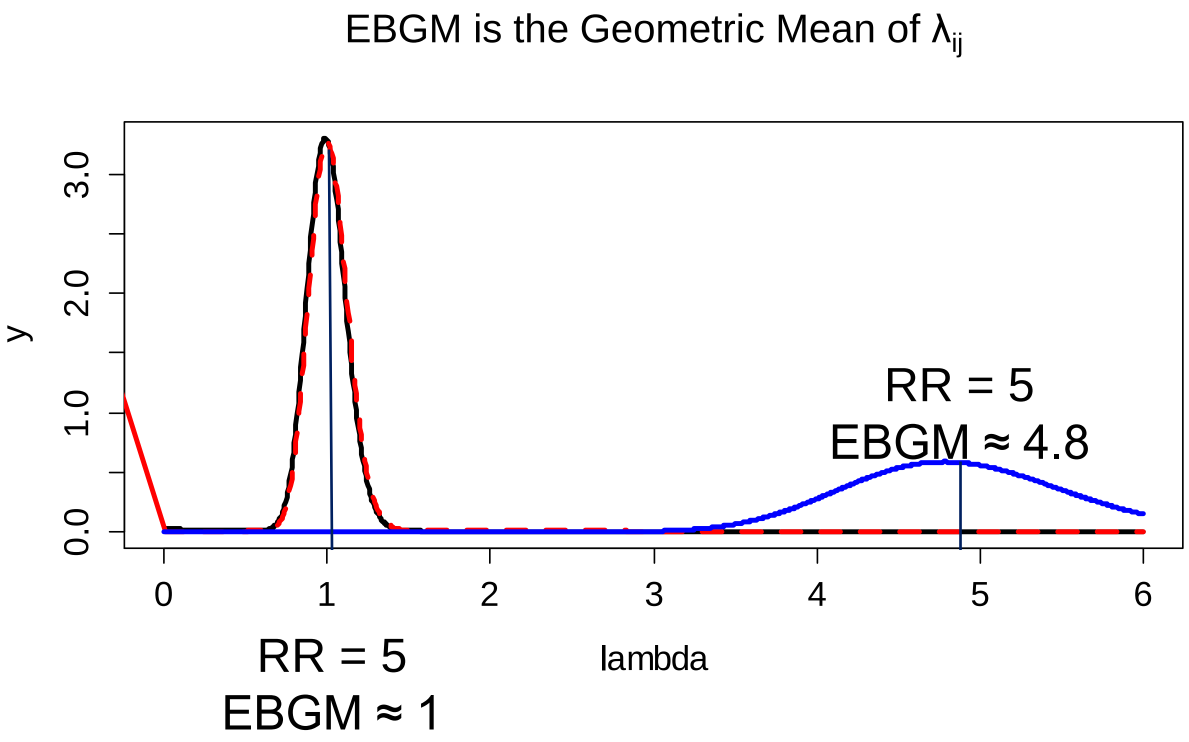


Event with $N_{ij} = 50, E_{ij} = 10$

RR_{ij} = 5 and there's *a lot* enough information to suggest λ_{ij} is between 4 and 6



EBGM = Point Estimate of Lambda

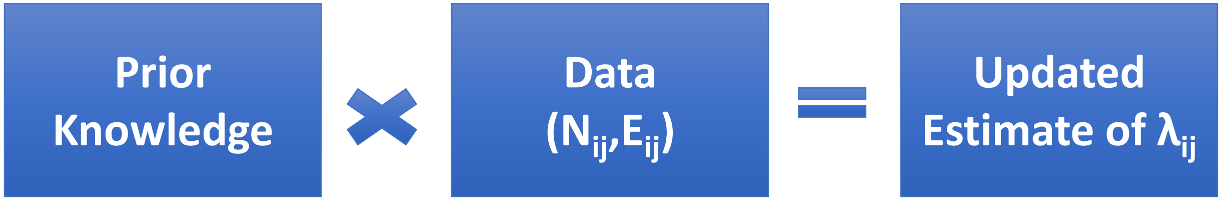


EBGM is the Geometric Mean of λ_{ij}

RR = 5
EBGM ≈ 4.8

RR = 5
EBGM ≈ 1

Summary

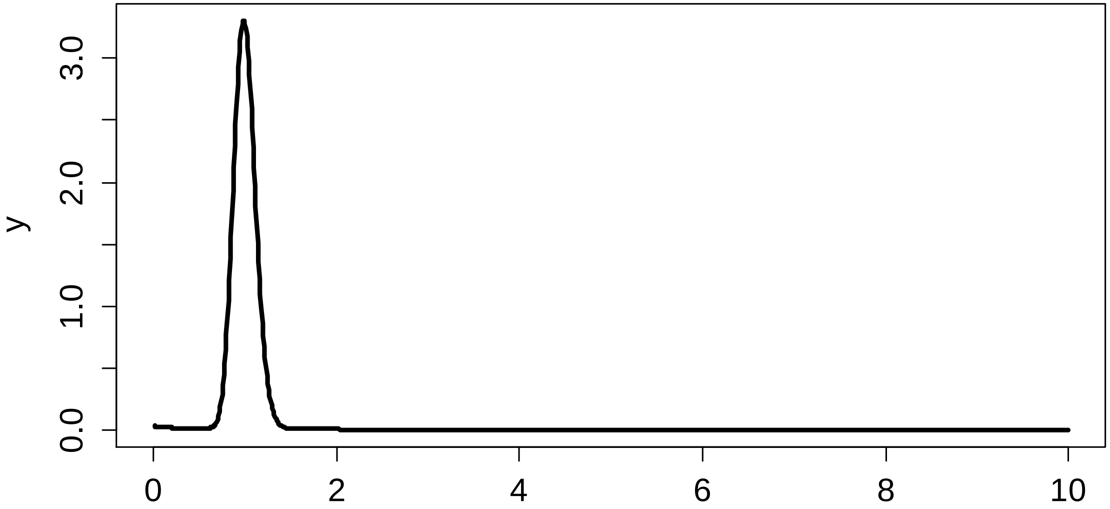


- If the number of complaints is small, there's not enough information in the data to override the prior
  - *Shrink* the influence of data when N_{ij} is small
  - This is the Shrinker piece of GPS

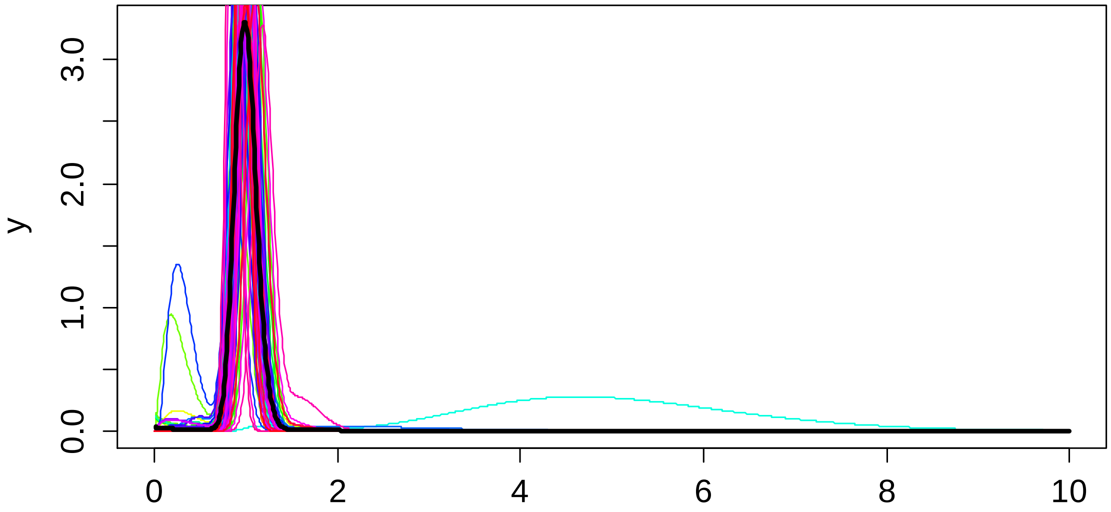
- If (N_{ij},E_{ij}) are different in ratio and value, then the data is telling us something!

Product X Data

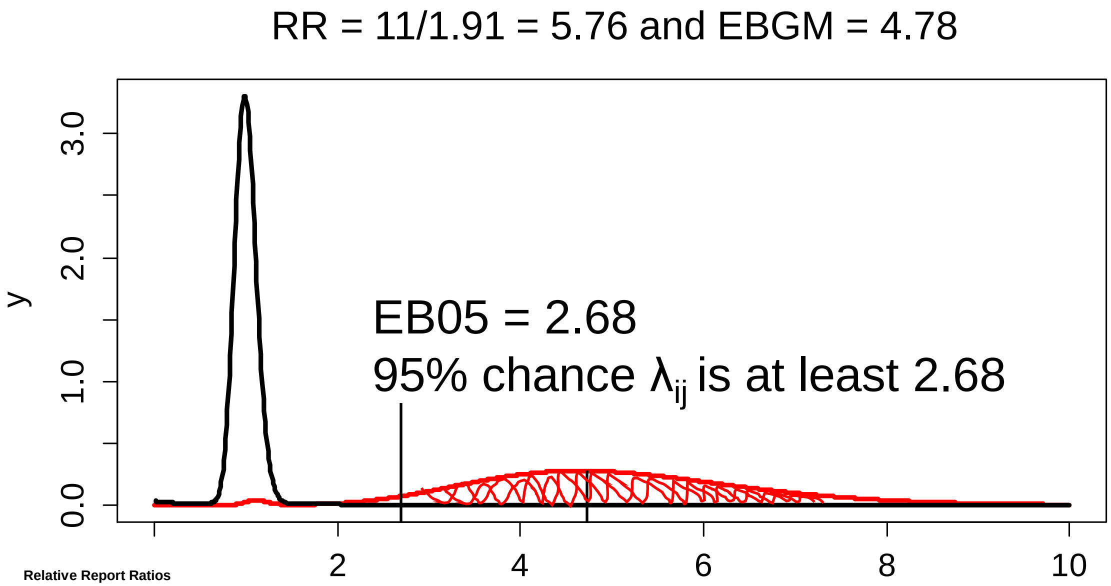
- 96 (24 months x 4 AEs) events
- 96 posterior distributions
- All 96 events have the same initial estimate of λ_{ij}



All 96 Posterior Distributions



Posterior Distribution of Signal

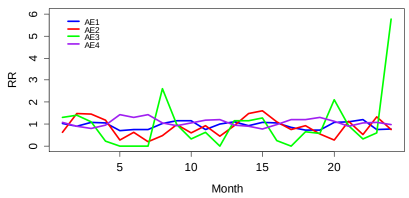


RR = 11/1.91 = 5.76 and EBGM = 4.78

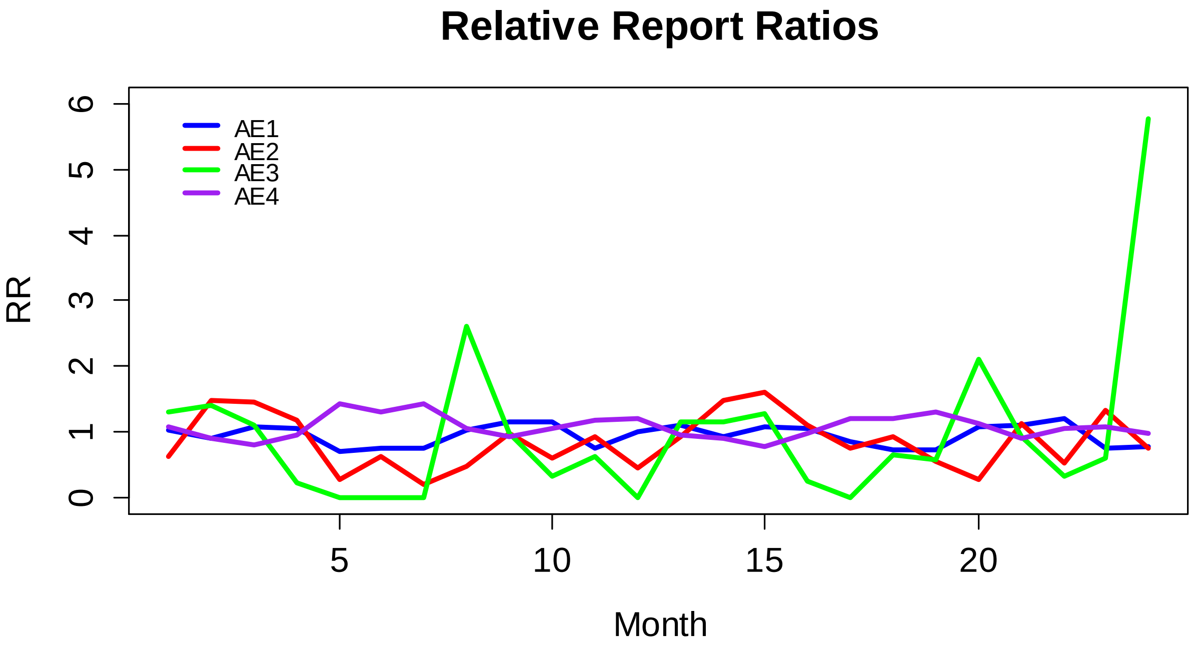
EB05 = 2.68

95% chance λ_{ij} is at least 2.68

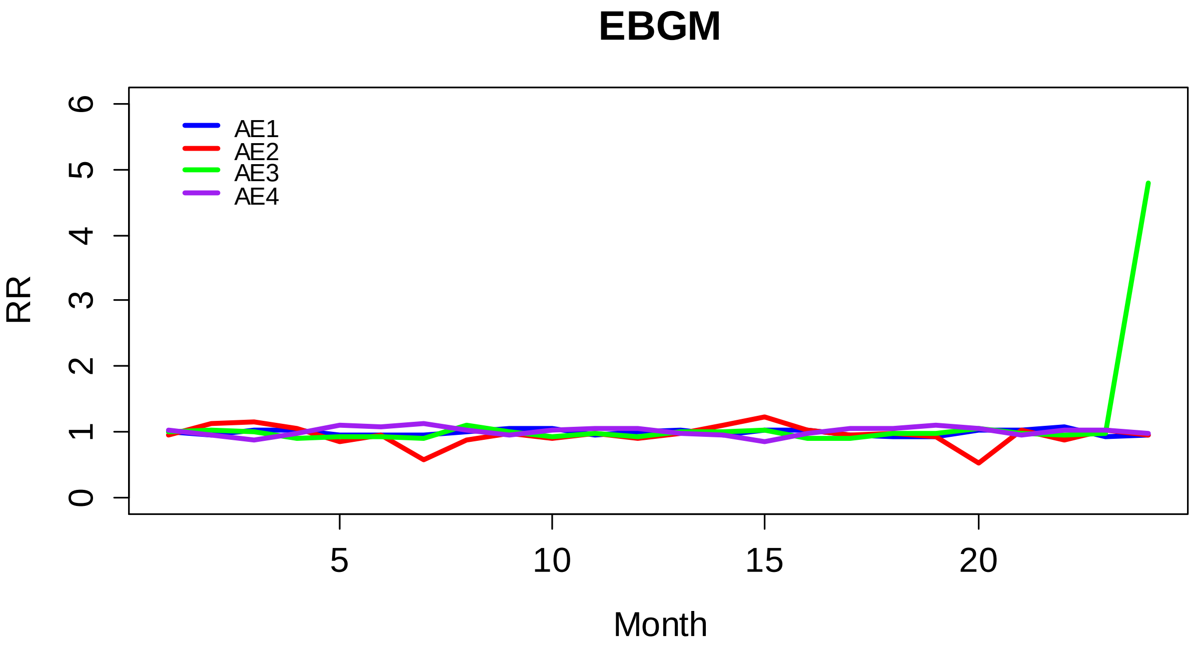
Relative Report Ratios



RRs for Product X vs. EBGMs



RRs for Product X vs. EBGMs

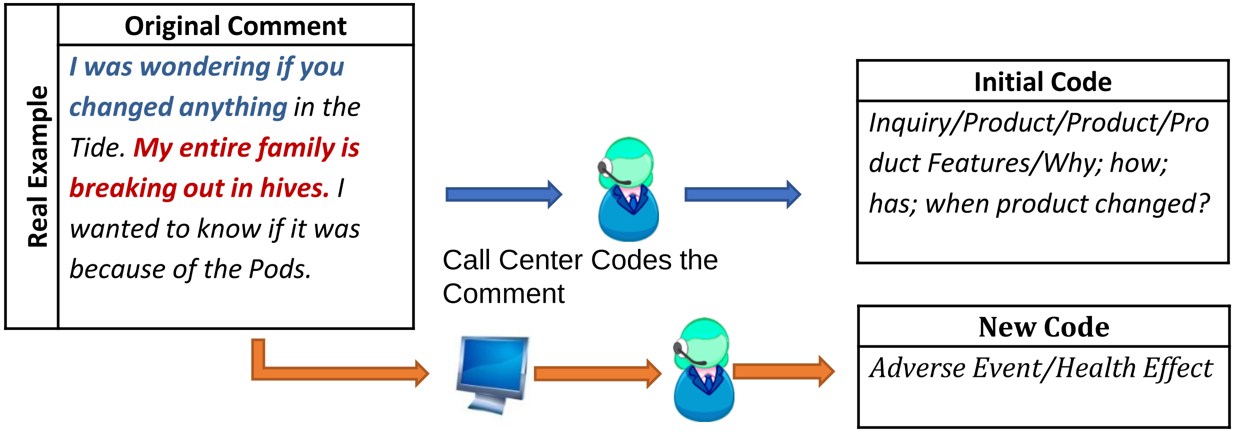


Data Quality using machine learning

Supervised and Unsupervised Techniques

Detecting Misclassified Adverse Events with Machine Learning

- All consumer comments are coded by P&G. 100% of Adverse Events are then reviewed by our Global Safety & Surveillance (GSSA) organization to ensure our products' safety.
- Using Vowpal Wabbit, we can automatically detect consumer comments that may have been incorrectly coded



Real Example

Original Comment

I was wondering if you changed anything *in the Tide.* ***My entire family is breaking out in hives.*** *I wanted to know if it was because of the Pods.*

Call Center Codes the Comment

Initial Code

Inquiry/Product/Product/Product Features/Why; how; has; when product changed?

New Code

Adverse Event/Health Effect

Machine Learning flags the comment, sends back for human review, comment code is corrected

Vowpal Wabbit (http://hunch.net/~vw/)

	w1	w2	w3	w4	w5	w6	w7	w8	w9	w10	w11	w12	w13	w14	w15	w16	w17		
Verbatim	y i	have	a	rash	i have	have a	a rash	tide	and	it	smells	funny	have tide	tide and	and it	it smells	smells funny	...	AE?
I have a rash	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	...	1
I have tide and it smells funny	1	1	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	...	-1
Where can I buy Charmin?	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...	-1
Pampers are the best!	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...	-1
It's hard to open a bottle.	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...	-1

- Linear classifier with a logistic loss function. *y* (label) and *p* (prediction)

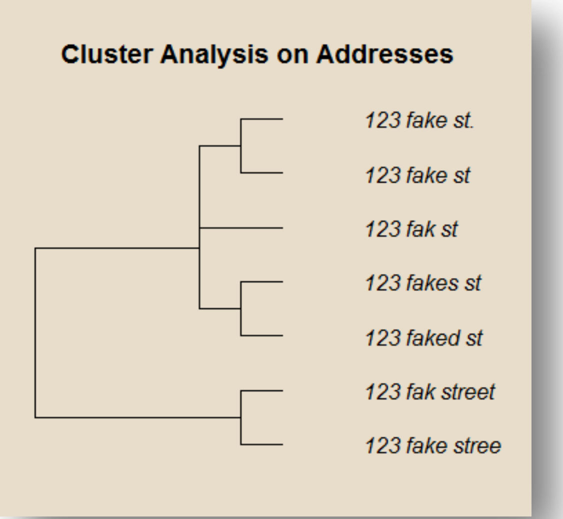
$$\log(1 + \exp(-yp))$$

Can we find Adverse Events on Social Media?

- I got Vicks in my eye last night had me all ****** up
- @Crest Yo crest. I used some pro-health and it burned my gums. How you gonna sell that? Or am i using it wrong?
- Sh*t. I just got Oil of Olay face cream in in my eye. That's a first. Burns like a mother.
- I putted vicks on my face and it's burning me
- Roommate sprayed Febreeze EVERYWHERE on her bed and I instantly got a migraine. It was almost a half hour ago since I took something.

Detecting Potential Fraud with Text Mining and Network Analysis

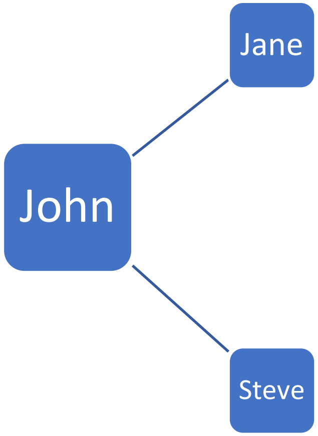
- Repeat complainers (aka *repeaters*) edit their addresses to appear as different households. They register multiple complaints, asking for coupons and gift cards. We use cluster analysis on the edit-distance of address to group similar addresses into unique households.
- Repeat consumers complain and have their coupons sent to a different household or multiple households, creating *repeater networks*. We can isolate these networks with applications from Graph Theory

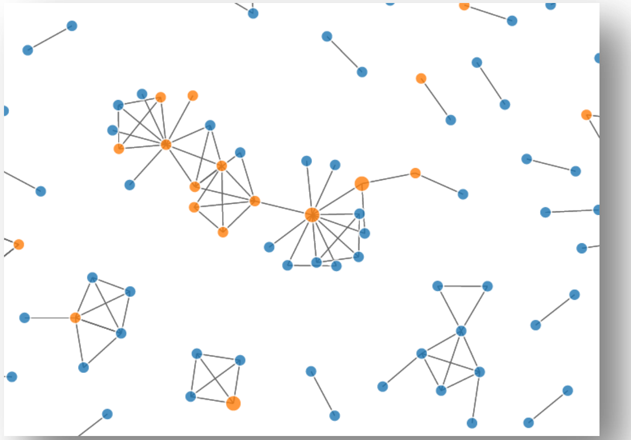


Repeaters

Repeater Networks

Name	Email	Address	City
John Doe	thisisfake@ymail.com	123 Fake St	Mason
Johnny Doe	hello@hotmail.com	123 Fake St	Mason
Jane Doe	hello@hotmail.com	789 Main St	Dayton
Steve Doe	thisisfake@ymail.com	1 Repeater Dr.	Dayton

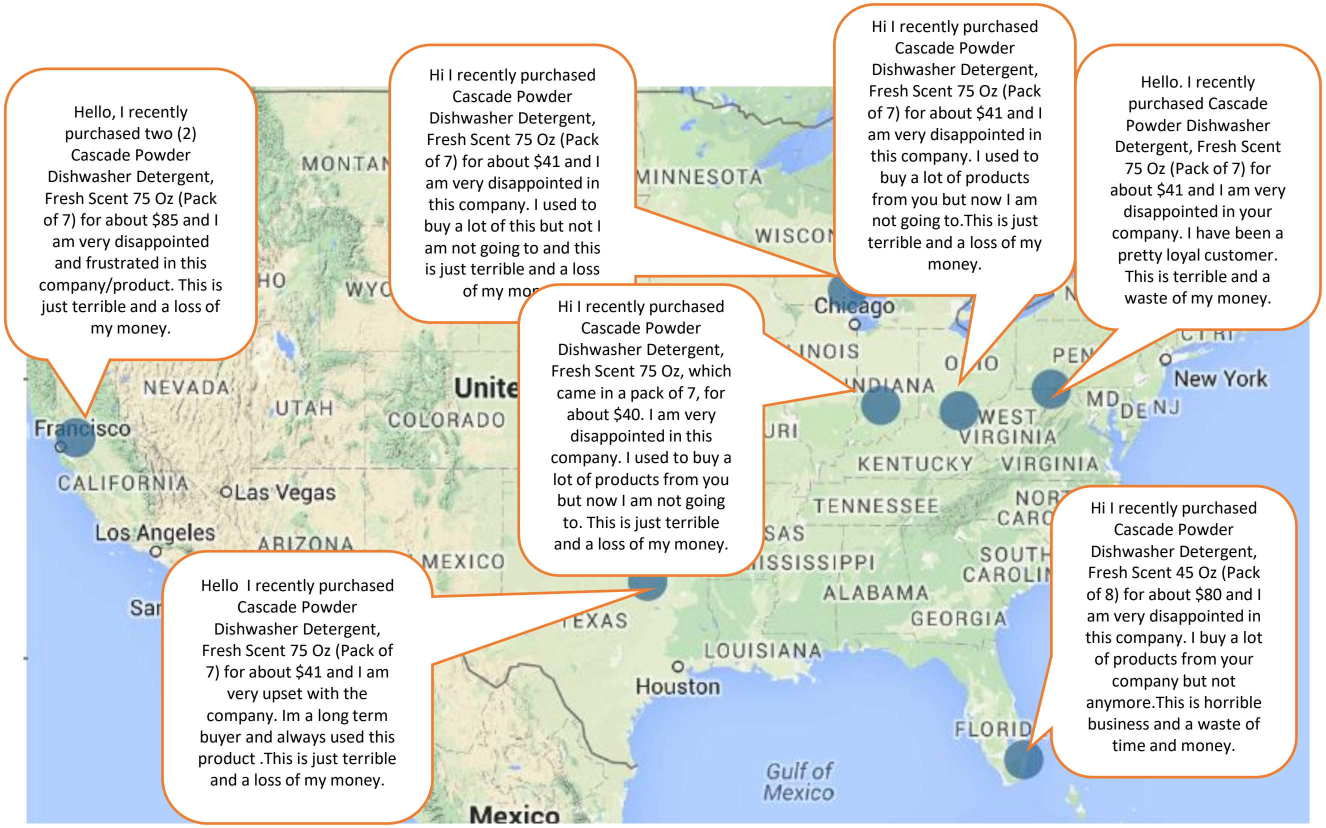




Repeater Networks

Detecting Potential Fraud with Text Mining and Network Analysis

Consumers find complaints online, change a few words, and submit complaints about our products. We call these people "*Copy Cats*" and can detect them with text mining.



Hello, I recently purchased two (2) Cascade Powder Dishwasher Detergent, Fresh Scent 75 Oz (Pack of 7) for about $85 and I am very disappointed and frustrated in this company/product. This is just terrible and a loss of my money.

Hi I recently purchased Cascade Powder Dishwasher Detergent, Fresh Scent 75 Oz (Pack of 7) for about $41 and I am very disappointed in this company. I used to buy a lot of this but not I am not going to and this is just terrible and a loss of my mon

Hi I recently purchased Cascade Powder Dishwasher Detergent, Fresh Scent 75 Oz, which came in a pack of 7, for about $40. I am very disappointed in this company. I used to buy a lot of products from you but now I am not going to. This is just terrible and a loss of my money.

Hello I recently purchased Cascade Powder Dishwasher Detergent, Fresh Scent 75 Oz (Pack of 7) for about $41 and I am very upset with the company. Im a long term buyer and always used this product .This is just terrible and a loss of my money.

Hi I recently purchased Cascade Powder Dishwasher Detergent, Fresh Scent 75 Oz (Pack of 7) for about $41 and I am very disappointed in this company. I used to buy a lot of products from you but now I am not going to.This is just terrible and a loss of my money.

Hello. I recently purchased Cascade Powder Dishwasher Detergent, Fresh Scent 75 Oz (Pack of 7) for about $41 and I am very disappointed in your company. I have been a pretty loyal customer. This is terrible and a waste of my money.

Hi I recently purchased Cascade Powder Dishwasher Detergent, Fresh Scent 45 Oz (Pack of 8) for about $80 and I am very disappointed in this company. I buy a lot of products from your company but not anymore.This is horrible business and a waste of time and money.

More Examples

Pampers

My **daughter** recently had a child who's 6 months old now we bought him 6 packs of the cruiser diapers since there suppose to support movement but my grandson seem to leak from the sides
My **sister** recently had a child who's 6 months old now we bought him 6 packs of the cruiser diapers since there suppose to support movement but my grandson seem to leak from the sides
My **mom** recently had a child who's 6 months old now we bought him 6 packs of the cruiser diapers since there suppose to support movement but my grandson seem to leak from the sides

OLAY

HI I want to say I have been buying **olay soup** for 20 years. I'm in my XXXXXX now and my skin stays dry so I've been buying the age defining bar soap. I'm sorry to say, buy my skin is moisturizing like it did the last 3 years. I bought the olay soap at Wal-Mart and I buy the 8 bars and 3 of the bars when like split and cracked, the other 5 were normal. Just though **I'd let you no.**
I want to say I have been buying **olay soup** for several years. I'm in my getting older now and my skin stays dry so I've been buying the age defining bar soap. I'm sorry to say, but I'm disappointed. My skin is moisturizing like it did the in the past. I bought the olay soap at Wal-Mart. **I'd let you no.**

Conclusion

- Creating a global system for signal detection of consumer complaints satisfied a high-level need within P&G
- Large multi-functional team created to deliver
- Many different statistical and machine learning tools used to deliver overall project goals – innovative use of MGPS algorithm
- Both technical and non-technical skills used to arrive at a meaningful solution
- Solution is embedded in QA and Safety work processes
- Strong Validation Protocols
- Influencing, Deployment & Training - key components of success

Acknowledgements

Alex Gutman, Joel Chaney, John Dunavent, Rob Baker, Joe DiGennaro, Jeff Swartzel, Winter Qi, Beatrice Blum, Russell Jolly, Nelson Webb, Mark Dato

+ many others!

Thank you

brenneman.wa@pg.com