Saint-
PetersburgStateUniversityMathe
maticsandMechanicsFaculty

Makarov Alexandr

Dynamic Market Analysis

Termpaper

Scientificadvisor:
Korf V.

Saint-
Petersburg201
4

1 Introduction

The point of the research was to analyze consumer purchase patterns. The main points were analyzing sales effect on tendency of purchasing, renewing subscription patterns and consequences after transition to subscription system instead of upgrade system. In addition, Basket analysis has been done.

2 Input data

The data for the research was a list of consumers' transactions with following information:

- Customer.Code – a unique customer ID
- Item.Product.Name – name of the product
- Subscription.Type – name of the subscription type of license used in transaction
- Purchase.License.Type – name of the license type
- Offer.Name – name of the offer used during transaction
- Discount.in.USD – discount summary is US dollars
- Order.Placed.Date – the date when order was placed
- License.Quantity –the number of licenses used in transaction
- Purchase.Amount.in.USD –total purchase summary in US dollars

The data was collected from several countries:

Finland, Canada, Israel, France, Germany, Greece, Turkey, New Zealand, Korea, India, Italy, Spain, Czech, Austria, Belgium, Denmark, Brazil, Ukraine, Russia, China, Norway, Switzerland, Japan

The product names were replaced with English letters(A-J) to preserve commercial secrecy.

3 Salesanalysis

During the specified period, there were three main large sales which were worldwide:

1. Earth Day Offer 2013

2. Mayan Doomsday Offer
3. Back to school 2012
4. Easter Discounts

Full sales info:

Sale	Start	End	days	A	B	C	D	E	F	G	H	I	J
Earth Day Offer 2013	Monday 15 April 13	Monday 22 April 13	7	34	50	40	40	41	40	0	40	40	35
Mayan Doomsday Offer	Thursday 20 December 12	Friday 21 December 12	1	75	75	75	75	75	75	75	75	75	75
Back to school 2012	Monday 3 September 12	Sunday 16 September 12	13	50	50	50	50	50	50	50	50	50	50
Easter Discounts	Monday 2 April 12	Monday 16 April 12	14	0	30	30	30	0	30	0	30	0	0
10% OFF IDEA personal licenses	Thursday 1 August 13	Saturday 31 August 13	30	0	10	0	0	0	0	0	0	0	0

In column “r” the discount percent for product “r” is given during a particular sale. 0 indicates that the product was not discounted during the sale.

The most important difference is length. It was challenge to find a way to compare sales with different lengths.

To predict neutral sales during sale period, HoltWinters[1] forecasting model was used with seasonal and trend components.

The statistic we used to analyze the sales:

$$y = \frac{\frac{sum\_profit * (100\% - discount\_percent)}{period}}{Average\_neutral\_sales}$$

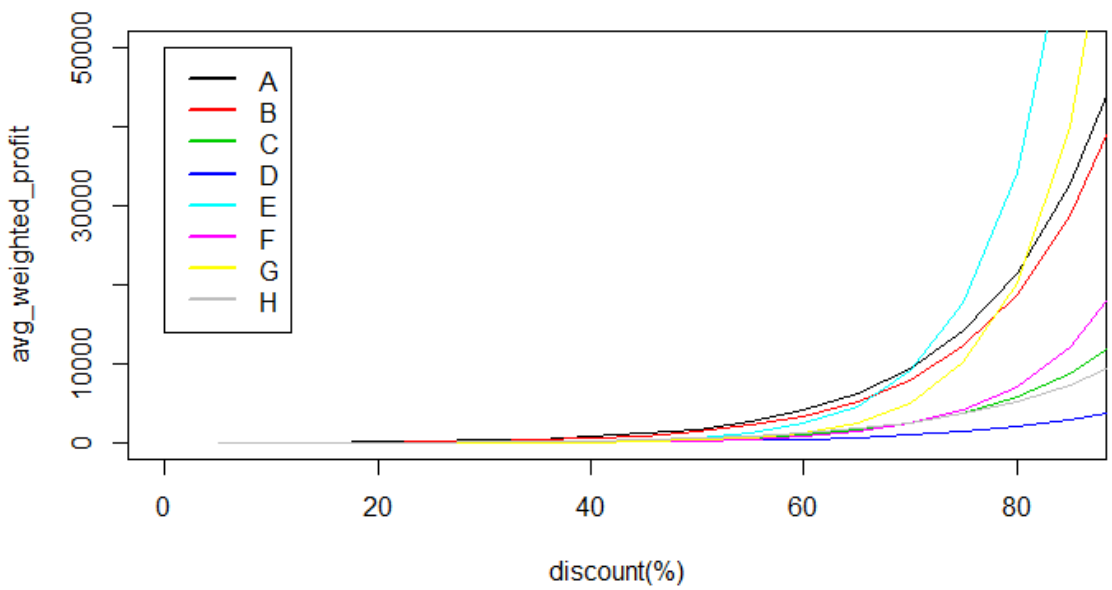
sum_profit – the number of additional licenses purchased during the sale (is calculated as difference between real value and predicted using forecast model).
$discount\_percent$ – the discount percent of the sale for the specific product.

$period$ – the sale length in days
$average\_neutral\_sales$ – the average daily number of licenses purchases during post-sale period of the same length.

For all the products, graphs were drawn withdiscount percent on the X-axis and variable value on the Y-axis.
We assumed that this function rises exponentially, as it is believed to be in standard discount-demand scenario. Consequently, we introduced a parametric function of exponential form to approximate the unknown function.

$$y = a(e^{bx} - 1)$$

The formula has (0, 0) point, because there is no additional sales during 0% sale.
Here, parameter b is the main one and represents the speed of interest growth for a certain product with increasing discount percent. Parameter a was used for normalization.
Optimization process was HillClimbing[2], minimizing Mean Square Error[1] on data points.
All the parameters were written into a table and were used to draw a collaborative graph of consumer reaction on product sales:

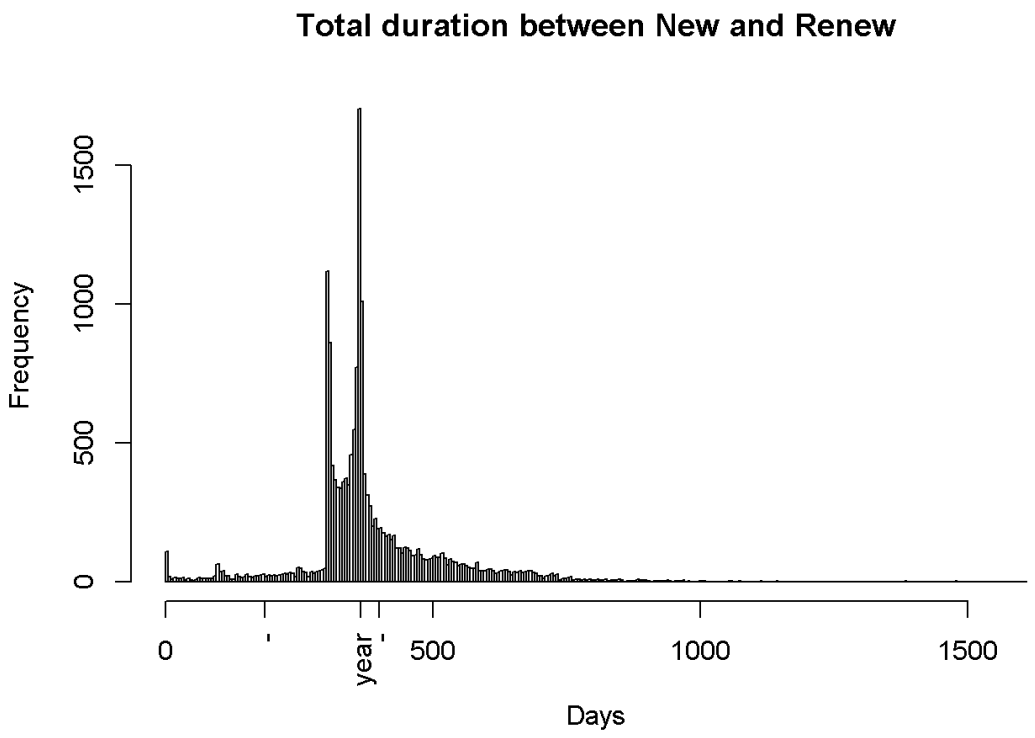


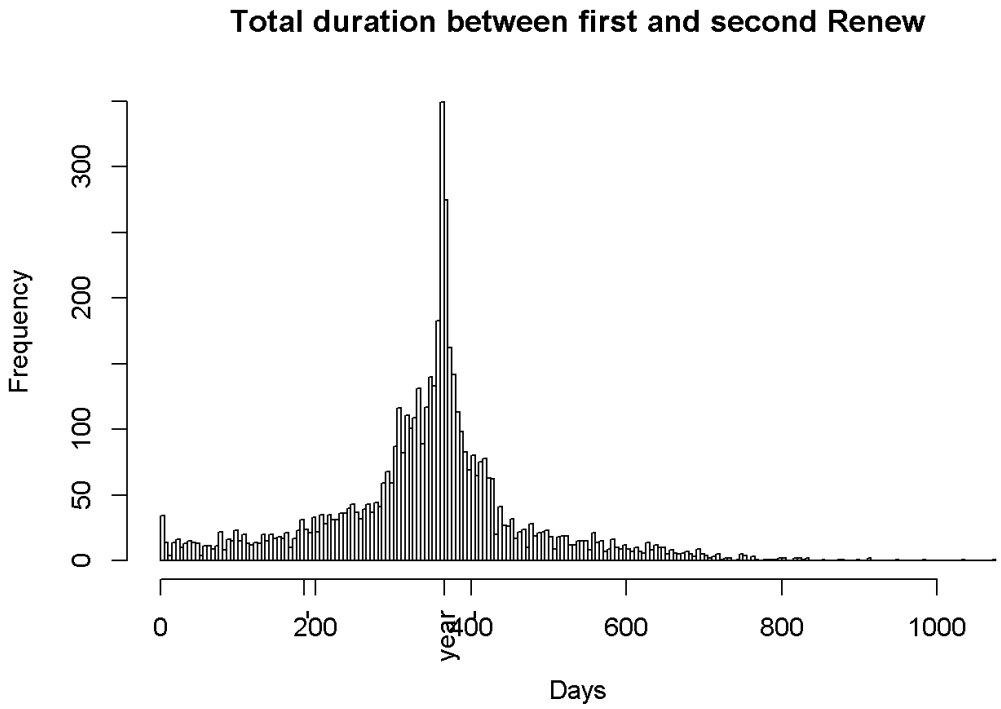
This graph can be used to compare sale effect on different products. However, to find optimal discount for a specific product, the b parameter should solely be looked at.

4 Loyal customers

Our marketing definition for loyal customer is “a customer, behavingin an expected and beneficial way for the company”. Examples would be that they do not react on small price changes, use and update their products frequently and purchase additional products on sales. Exactly these customers are supposed to be the main source of the revenue as the company provides subscription-based software.

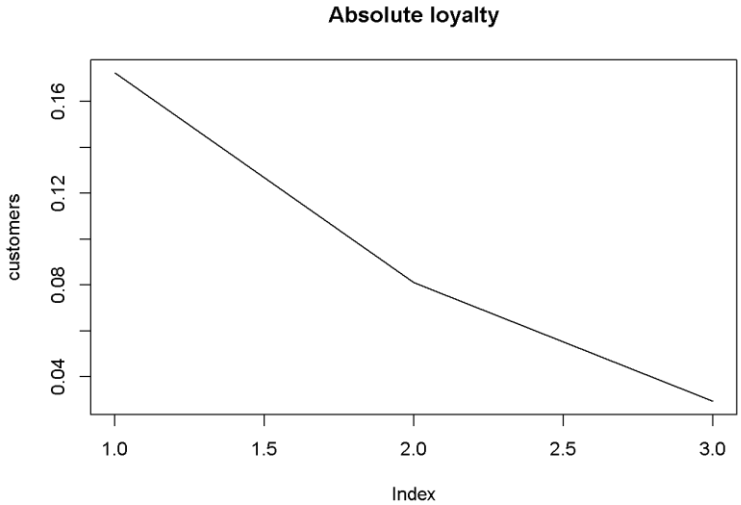
Having the definition in mind, we have to find a criterion that we can calculate to distinguish the loyal customers from the rest. We agreed to use this one: a customer is loyal, if he extends the subscription as it ends. Formally, it occurs between half a year before the end of the subscription and exactly 35 days after. If a customer is extending his subscription before is has ended, this is a positive sign, although if it happens too early, it is unexpected and should not be considered loyal as it is probably some kind of noise in customer’s behavior. Examples of the histogram are shown below:





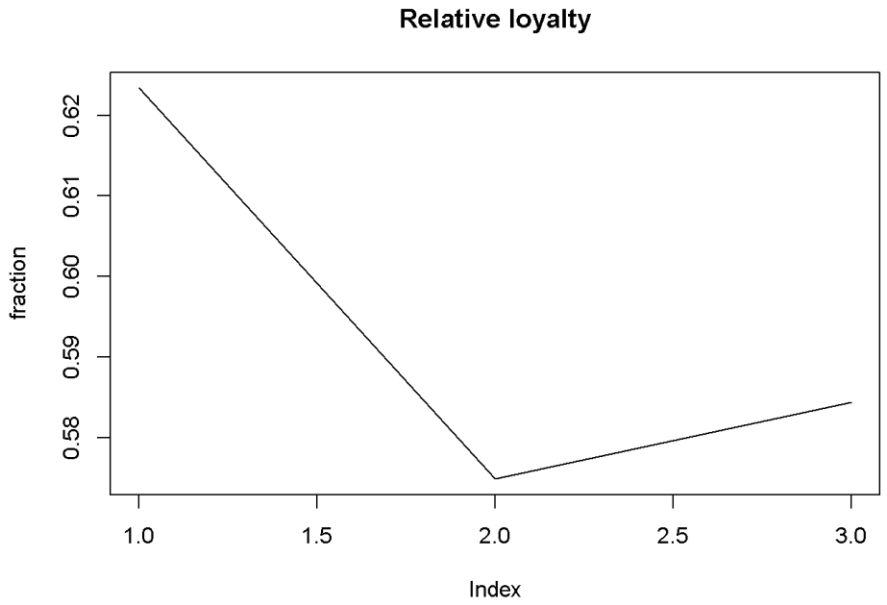
“Year” mark indicates the one year period on the “days” axis and there are small segments, constituting the loyalty criterion period of renew.

For each product, we have drawn a “Relative loyalty” graph, showing a fraction of loyal customers for each consequent Renew among all the customers, who have done that Renew and an “Absolute loyalty” graph, showing a fraction of loyal customers for each consequent Renew among all the customers of the product.
“Absolute loyalty” has shown an expected tendency for all the products:

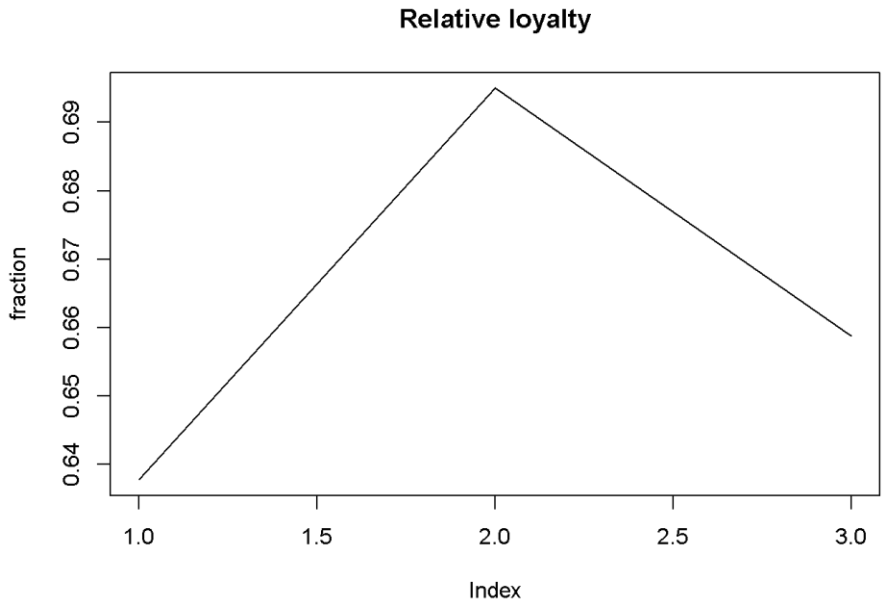


However, based on “Relative loyalty”, the products were divided into 2 classes, showing similar customers’ behavior.

Class A:



Class B:



Actually, we have expected “Relative loyalty” graph to be increasing. The reason for that being that the longer a customer is using the company’s product, the more loyal and committed he/she should be.

Therefore, we suggest setting up a sale for the Renew with dumped loyalty fraction value as it seems to be the critical point between an indifference to a product and an attachment to it. Assuming that Renew sales affect them the same way as sales for new licenses do, we can calculate the required discount percent, using the exponential formula from 3 part.

5 Basket analysis

We looked at products do customers buy. Preliminarily, we have excluded the purchases, occurred during sales, considering them as noise. We've created a table of Booleans, where i,j bit shows whether i-th customer has bought j-th product. Then we launched "associative rules"[3] search. Nearly all the found rules were about 0 values.
Example: {B = 0, C = 0} => {D = 0}.
The conclusion to be made is the most products are purchased independently. Therefore, we divided all the customers into 5 clusters, which had sizes of the same order(порядок). The result are shown in the table:

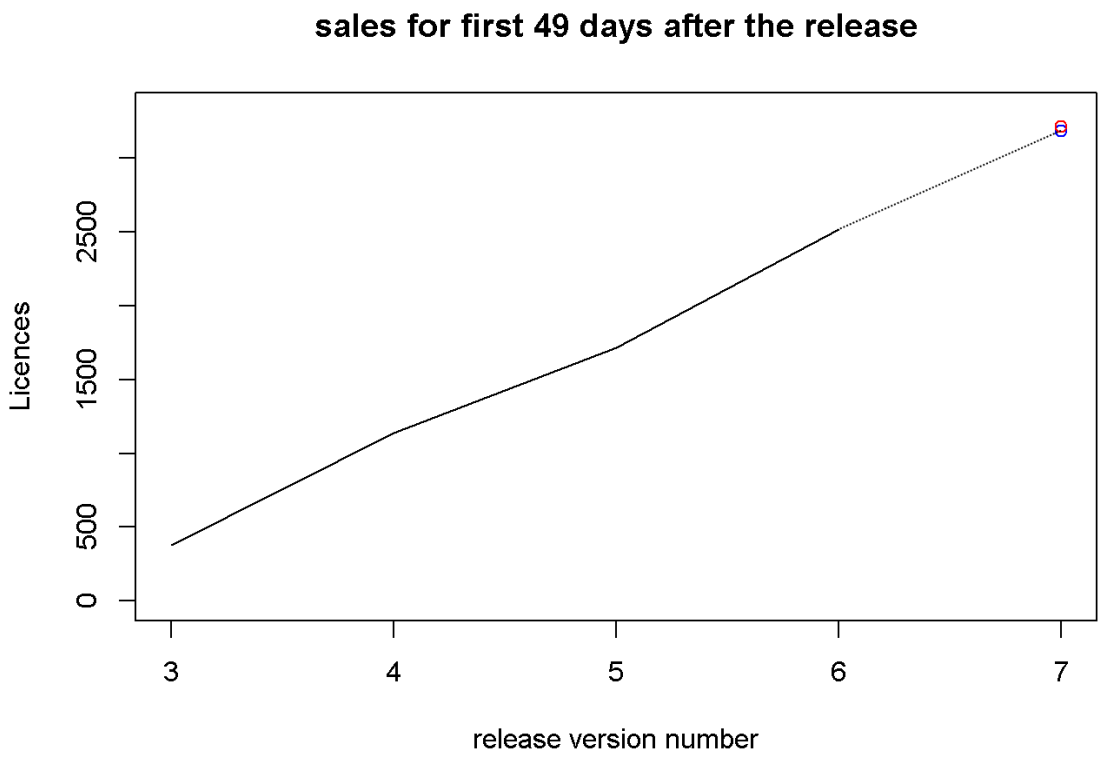
Product combination	Number	Fraction
Others	10405	0.16
A	12698	0.20
B	15611	0.24
B + A	171	<0.01
C	8940	0.14
C + A	695	0.01
C + B	563	0.01
C + B + A	31	<0.01
D	17363	0.27
D + A	392	0.01
D + B	648	0.01
D + B + A	44	<0.01
D + C	746	0.01
D + C + A	64	<0.01
D + C + B	110	<0.01
D + C + B + A	22	<0.01

It is easily noticeable, that any combination of 2 or more products has <0.01 fraction. So the 5 clusters are "A","B","C","D","Others". "Others" means that the customer hasn't bought any products from "A,B,C,D" list, although has purchased some others.

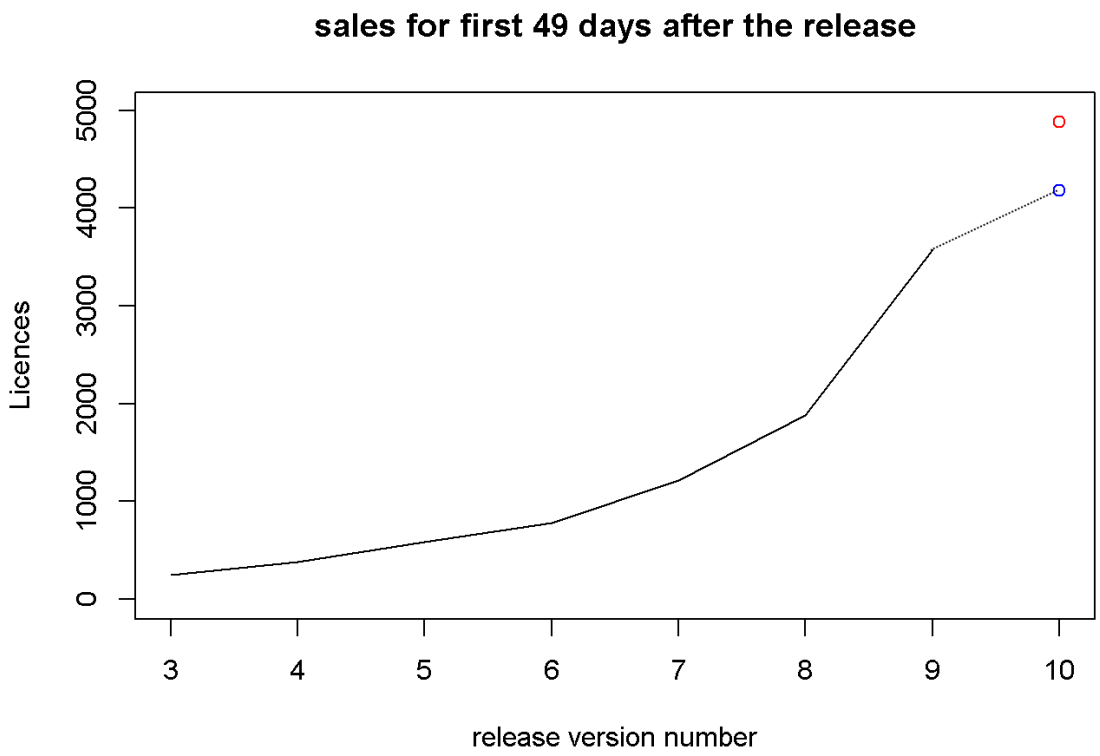
6 Upgrade regression

Recently, the company moved from Upgrade system to Renew system for updating the products. Upgrade is a payment to update one's product to the latest version, while

Renew is 1-year subscription on all major updates. Renew is on average 1.6 time more expensive than upgrade. We decided to research how customers' behavior had changed after the company removed the upgrade system.
After each major release, there is an enormous spike of upgrades for first several days. Releases are made usually uniformly, although not ideally, so calculating the number of upgrades between the releases could be misinforming. Thus, we decided to look at upgrades for the first 7 weeks (49 days) after the release. After each new release, the value grows, as the client base grows. We have built a general regression model[4] from i – version number to predict the number of upgrades at first 7 weeks after release. For the latest release, when upgrades were removed, we compared the number of predicted upgrades using the model and the number of purchased renews. The results are on the graphs:
<u>Product A</u>



Product B



The blue point shows the prediction, the red point shows the summary of upgrades and renews. We discovered that renew system worked out well, as the client base seems to have grown larger than we expected. Considering that renew costs more, there is a consistent increase of income.

7 References

[1] The Holt-Winters Forecasting Procedure *C. Chatfield* Journal of the Royal Statistical Society. Series C (Applied Statistics)

[2] Comparative Analysis of Hill Climbing Mapping Algorithms *David L. Smitley* Supercomputing Research Center

[3] Visualizing Association Rules: Introduction to the R-extension Package arulesViz *Michael Hahsler* Southern Methodist University,*Sudheer Chelluboina* Southern Methodist University

[4] Effect Displays in R for Generalised Linear Models*John Fox*Hamilton, Ontario, Canada