

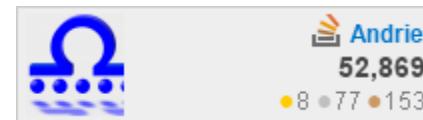
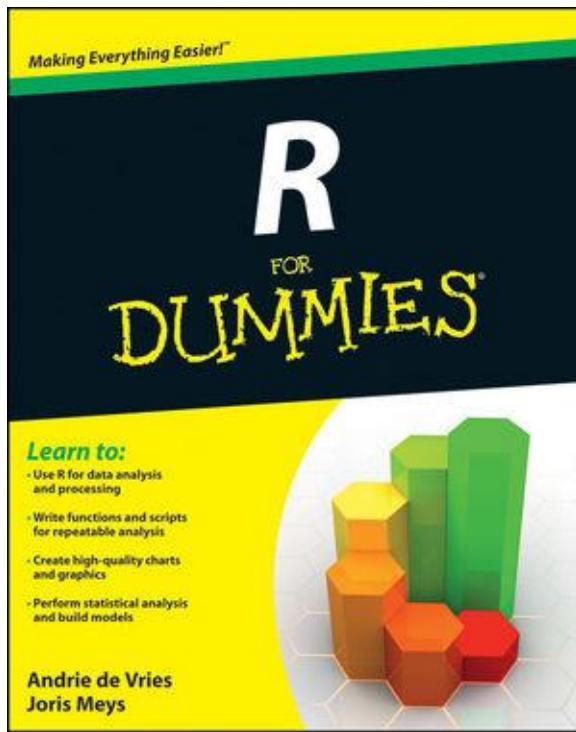
Survival Analysis for Marketing Attribution

Webinar
July 2013

Andrie de Vries
Business Services Director – Europe
@RevoAndrie

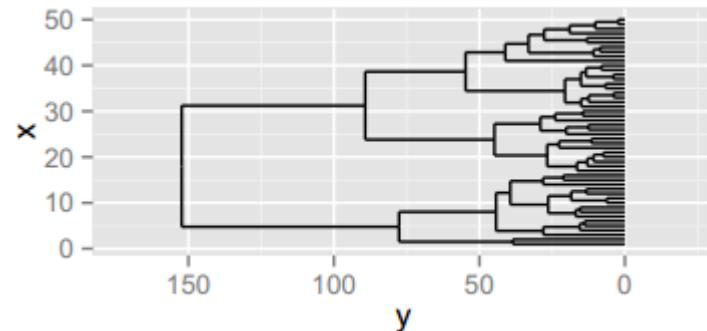
Revolution Analytics
@RevolutionR

Who am I?



StackOverflow

CRAN package ggdendro



From Toledo to Albacete



Rent a car?

Google car rental toledo spain andrie@revolution-computing.com ▾

Web Images Maps Shopping More ▾ Search tools ⚙

About 664,000 results (0.25 seconds)

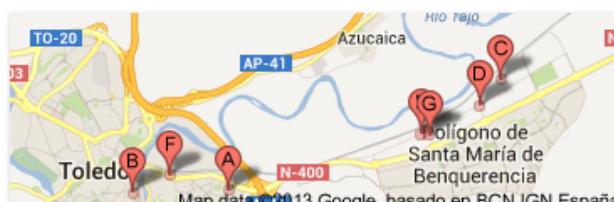
Cookies help us deliver our services. By using our services, you agree to our use of cookies.
OK Learn more

Ads related to car rental toledo spain ⓘ

Toledo Car Hire - Quick & Easy: Book in 3 Steps - Europcar.co.uk
www.europcar.co.uk/Toledo
Car Hire Toledo up to 25% off!
Europcar UK has 152 followers on Google+
Airport Pick Up Valencia
Barcelona

Spain Car Rental - Up to 35% Discount - Book Today
www.auto-europe.co.uk/Spain
★★★★★ 1,624 reviews for auto-europe.co.uk
Guaranteed Best Prices Online!
Auto Europe UK has 627 followers on Google+
Car Hire Spain - Discounted Rates - No Credit Card Fees - Car Hire Alicante

Car Rental Toledo Airport - Best Price Car Rental - Guaranteed
www.rentalcars.com/Toledo-Car-Rental
★★★★★ 9,703 reviews for rentalcars.com
Call Us Free Or Book Online Today!
43 people in London, UK +1'd this
Best Car Rental Prices - Guaranteed - Free 24 Hour UK Based Call Centre



A map showing the Toledo area with several red location markers labeled A through G. Key landmarks include the city of Toledo, the AP-41 highway, the N-400 road, and the Rio Tajo. Other labels on the map include 'Azucaica', 'Hio Tajo', 'Polígono de Santa María de Benquerencia', and 'N'. A copyright notice at the bottom right states 'Map data ©2013 Google, basado en BCN IGN España'.

Map for car rental toledo spain

Ads ⓘ

Spain Car Rental
www.arguscarhire.com/Spain
★★★★★ 682 seller reviews
Guaranteed Best Prices Online.
Spain Car Rental Book Here And Save

Cheapest Spain Car Rental
www.spain-car-rental.com/
Nº1 Spain Car Rental. From 32€/wk
Top Companies at Bargain Rates!

Spain Car Hire Fr £5/Day
www.dealchecker.co.uk/Spain_Car_Rental
★★★★★ 104 seller reviews
Compare Cheapest Car Hire Prices

... or take a bus?

Noticias de este mes...

- [el agua a través del ingenio, es la hora de castilla-la mancha](#)
- [los mayos en motilleja](#)
- [mujer y hombre en igualdad, concurso pintura](#)
- [Ver Más...](#)

Albacete - Toledo

Lunes salida: 06:00 llegada: 08:45
Martes a viernes* salida: 07:00 llegada: 09:45
Lunes a viernes* salida: 16:00 llegada: 18:45
Sábados salida: 08:00 llegada: 10:45
Sábados salida: 16:00 llegada: 18:45
Domingos salida: 18:30 llegada: 21:15

Toledo - Albacete

Lunes salida: 06:00 llegada: 08:45
Martes a viernes* salida: 07:00 llegada: 09:45
Lunes a viernes* salida: 16:00 llegada: 18:45
Sábados salida: 09:00 llegada: 11:45
Sábados salida: 17:00 llegada: 19:45
Domingos salida: 18:30 llegada: 21:15
*Servicios establecidos para días laborables.

Precio billete Albacete - Toledo

Billete sencillo 14,20 euros
Billete de ida y vuelta 25,25 euros
Abono de 10 viajes 110,65 euros

Puertollano - Ciudad Real - Toledo

Lunes a viernes* salida: 06:15 07:00 llegada: 08:30
Lunes a viernes* salida: 15:30 16:15 llegada: 17:45
Sábados salida: 08:00 08:45 llegada: 10:15
Sábados salida: 15:00 15:45 llegada: 17:15
Domingos salida: 17:00 17:45 llegada: 19:15

Toledo - Ciudad Real - Puertollano

Lunes a viernes* salida: 07:00 08:30 llegada: 09:15
Lunes a viernes* salida: 16:00 17:30 llegada: 18:15
Sábados salida: 10:30 12:00 llegada: 12:45
Sábados salida: 17:30 19:00 llegada: 19:45
Domingos salida: 20:00 21:30 llegada: 22:15

Precio billete Puertollano - Toledo

Billete sencillo 10,80 euros
Billete de ida y vuelta 19,20 euros
Abono de 10 viajes 84,25 euros

Cine en Albacete

Programación Filmoteca de Albacete
Cartelera de Cine de Albacete

AdChoices ▶

Auto Europe - Car Hire

[www.auto-euro...](#)
From £7 per day
with Auto Europe
Guaranteed Best
Prices Online!



Conciertos Albacete

suspendido. Jero Romero en concierto en Albacete.
"Albacete Dance". Fiesta aniversario Máxima FM
Conciertos en Albacete organizados por AMIGA 2013

Exposiciones

Ouka Leele inédita

... but what's happening here?

Noticias de este mes...

- [el agua a través del ingenio, es la hora de castilla-la mancha](#)
- [los mayos en motilleja](#)
- [mujer y hombre en igualdad, concurso pintura](#)
- [Ver Más...](#)

Albacete - Toledo

Lunes salida: 06:00 llegada: 08:45
Martes a viernes* salida: 07:00 llegada: 09:45
Lunes a viernes* salida: 16:00 llegada: 18:45
Sábados salida: 08:00 llegada: 10:45
Sábados salida: 16:00 llegada: 18:45
Domingos salida: 18:30 llegada: 21:15

Toledo - Albacete

Lunes salida: 06:00 llegada: 08:45
Martes a viernes* salida: 07:00 llegada: 09:45
Lunes a viernes* salida: 16:00 llegada: 18:45
Sábados salida: 09:00 llegada: 11:45
Sábados salida: 17:00 llegada: 19:45
Domingos salida: 18:30 llegada: 21:15
*Servicios establecidos para días laborables.

Precio billete Albacete - Toledo

Billete sencillo 14,20 euros
Billete de ida y vuelta 25,25 euros
Abono de 10 viajes 110,65 euros

Puertollano - Ciudad Real - Toledo

Lunes a viernes* salida: 06:15 07:00 llegada: 08:30
Lunes a viernes* salida: 15:30 16:15 llegada: 17:45
Sábados salida: 08:00 08:45 llegada: 10:15
Sábados salida: 15:00 15:45 llegada: 17:15
Domingos salida: 17:00 17:45 llegada: 19:15

Toledo - Ciudad Real - Puertollano

Lunes a viernes* salida: 07:00 08:30 llegada: 09:15
Lunes a viernes* salida: 16:00 17:30 llegada: 18:15
Sábados salida: 10:30 12:00 llegada: 12:45
Sábados salida: 17:30 19:00 llegada: 19:45
Domingos salida: 20:00 21:30 llegada: 22:15

Precio billete Puertollano - Toledo

Billete sencillo 10,80 euros
Billete de ida y vuelta 19,20 euros
Abono de 10 viajes 84,25 euros

Cine en Albacete

Programación Filmoteca de Albacete
Cartelera de Cine de Albacete

AdChoices ▶

Auto Europe - Car Hire
www.auto-euro.com
From £7 per day
with Auto Europe
Guaranteed Best Prices Online!

Conciertos Albacete

suspendido. Jero Romero en concierto en Albacete.
"Albacete Dance". Fiesta aniversario Máxima FM
Conciertos en Albacete organizados por AMIGA 2013

Exposiciones

Ouka Leele inédita

Marketing attribution: the Question

How to **attribute** conversion success to marketing spend?

Where to spend the next marketing dollar?

Agenda

Digital marketing attribution

Using Survival models

At scale, on big data

Agenda

Digital marketing attribution

Using Survival models

At scale, on big data

Two-click conversion journey



Click1:
Open landing page

A screenshot of an O2 website landing page. The main headline says "10% off when you top up". Below it, a sub-headline says "Get rewards when you top up your free sim card. A little thank you from us every three months." There are several promotional offers listed with "Order now" buttons: Pay & Go Go Go, O2 Unlimited, BlackBerry Text & Web, International Sim, iPad sim, and Simplicity for Tablets.

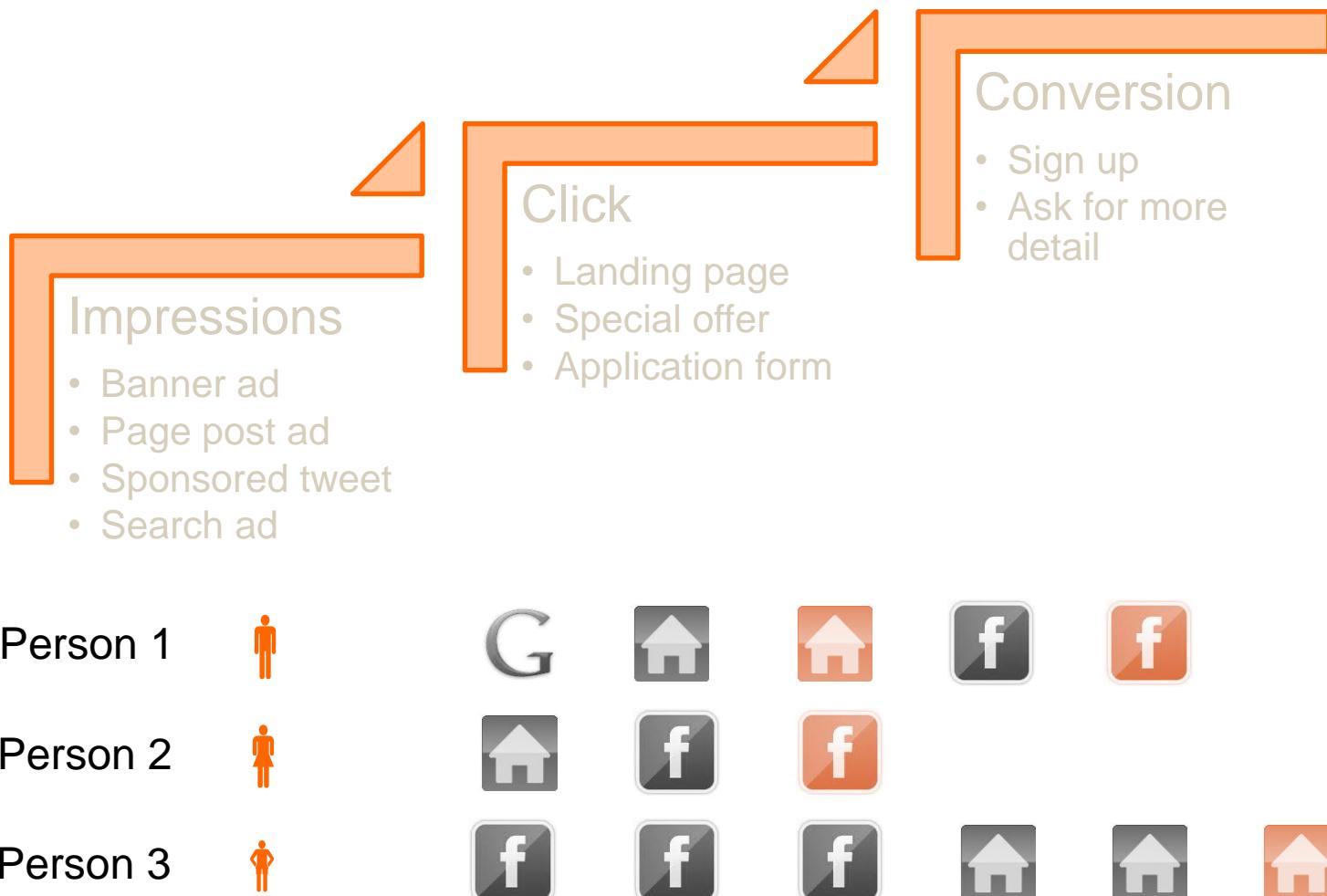
Click 2:
Sign up to offer

A screenshot of a sign-up form for the "Simplicity for Tablets" offer. The form includes fields for "Title", "First Name", "Last Name", "Email", "Mobile", "House Name", "Postcode", and "Address". There is also a note about existing users and a "Find Address" button.

Typical conversion journey...



...but no two journeys are the same...



...so how to attribute the value?

Attribution models

Last click only
All events even
Rule based
Statistical modelling



Attribution with statistical modelling

- Regression
 - In many cases, log data is available **only for conversions**
 - And when non-conversion data is available, these people **may convert in near future**



Attribution with statistical modelling

- Regression

- In many cases, log data is available **only for conversions**
 - And when non-conversion data is available, these people **may convert in near future**



- Survival analysis

- Use **time to conversion** as dependent variable
 - Can **use each interaction** (view or click) as an observation
 - Can include **censored** (incomplete) data
 - No need to flatten the data



Agenda

Digital marketing attribution

Using Survival models

At scale, on big data

Survival models

- Kaplan-Meier survivor function

$$\hat{S}_{km} = \prod_{t_i < t} \frac{r(t_i) - d(t_i)}{r(t_i)}$$

> library(survival)
> Surv(...)

- Cox proportional hazards model

$$L(\beta) = \prod L_i(\beta)$$

Partial likelihood

$$L_i(\beta) = \frac{r_i(t^*)}{\sum_j Y_j(t^*) r_j(t^*)}$$

Likelihood that individual i dies

> library(survival)
> coxph(Surv(...) ~ ...)

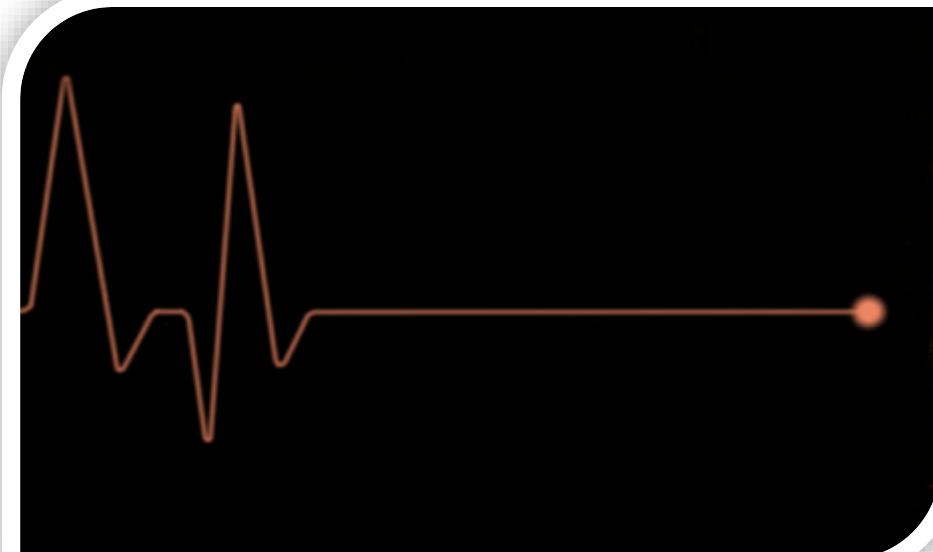
$$\lambda(t; Z_i) = \lambda_0(t) r_i(t)$$

Hazard function

$$r_i(t) = e^{\beta Z_i(t)}$$

Risk score

What is death?

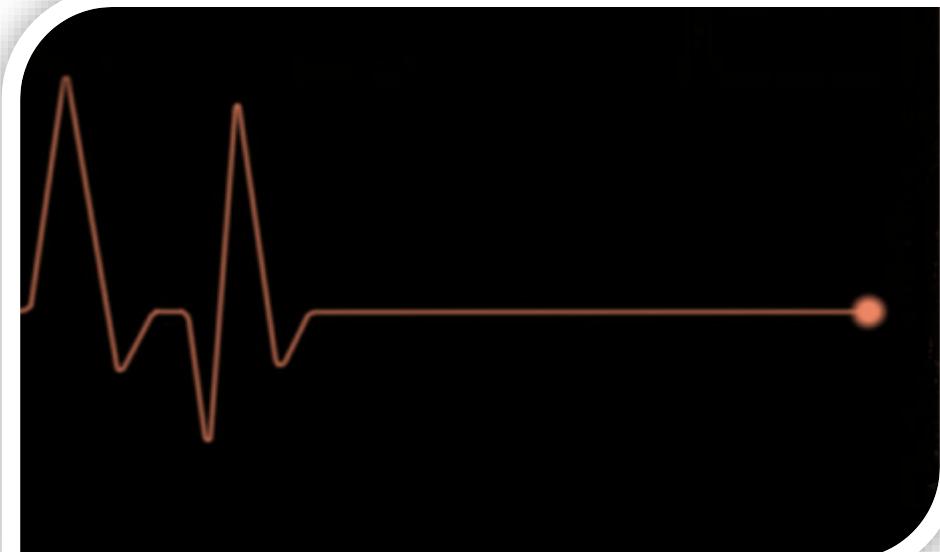


Medicine:

Engineering:

actual death of patient
failure of component

What is death?



*For attribution: **cookie conversion***

Worked example

Attribution of digital media for
telecoms client

Read the data

```
> rdsFile <- "survival_data.rds"
> xd <- readRDS(rdsFile)

> class(xd)
[1] "data.table" "data.frame"

> nrow(xd)
[1] 775782

> ncol(xd)
[1] 31
```

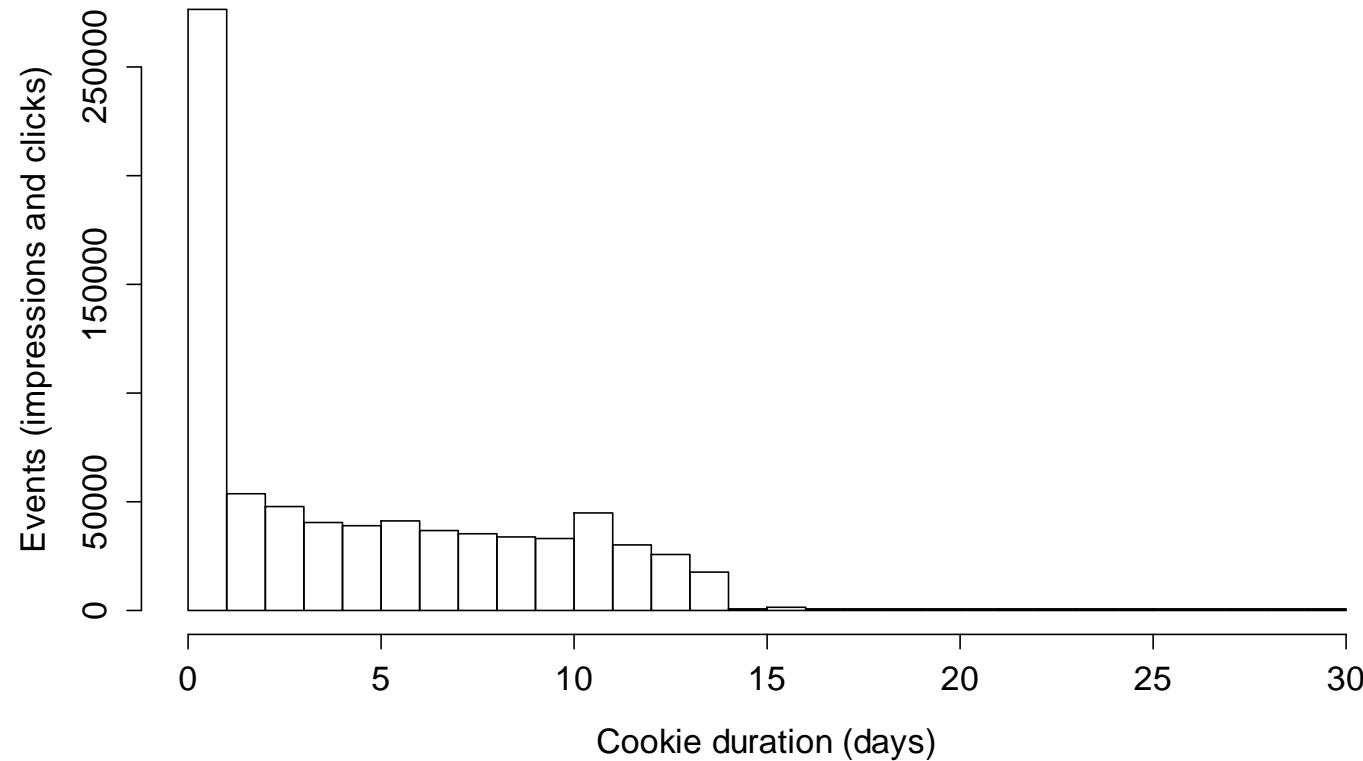
What does the data look like?

```
> xd[1:25, 1:6, with=FALSE]
```

	id	Conversion.Time	Event.Number	Event.Time	Event.Type	Campaign
1:	10101:49721794	01/10/2012 00:05		1 01/10/2012 00:02	Click	Free Sims
2:	10101:49721801	01/10/2012 00:05		1 29/09/2012 16:25	View	BAU High Media
6:	10101:49721854	01/10/2012 00:07		3 17/09/2012 18:32	View	BAU High Media
7:	10101:49721854	01/10/2012 00:07		4 17/09/2012 19:13	View	BAU High Media
8:	10101:49721854	01/10/2012 00:07		5 17/09/2012 19:17	View	BAU High Media
9:	10101:49721854	01/10/2012 00:07		6 17/09/2012 19:20	View	BAU High Media
10:	10101:49721854	01/10/2012 00:07		7 17/09/2012 19:21	View	BAU High Media
11:	10101:49721854	01/10/2012 00:07		8 17/09/2012 19:47	View	BAU High Media
12:	10101:49721854	01/10/2012 00:07		9 17/09/2012 19:49	View	BAU High Media
13:	10101:49721854	01/10/2012 00:07		10 17/09/2012 19:53	View	BAU High Media
14:	10101:49721854	01/10/2012 00:07		11 17/09/2012 20:04	View	BAU High Media
15:	10101:49721854	01/10/2012 00:07		12 18/09/2012 10:02	View	BAU High Media
16:	10101:49721854	01/10/2012 00:07		13 18/09/2012 10:03	View	BAU High Media
17:	10101:49721854	01/10/2012 00:07		14 18/09/2012 10:03	View	BAU High Media
18:	10101:49721854	01/10/2012 00:07		15 18/09/2012 20:06	View	BAU High Media
19:	10101:49721854	01/10/2012 00:07		16 18/09/2012 20:10	View	BAU High Media
20:	10101:49721854	01/10/2012 00:07		17 19/09/2012 18:14	View	BAU High Media
21:	10101:49721854	01/10/2012 00:07		18 19/09/2012 20:23	View	BAU High Media
22:	10101:49721854	01/10/2012 00:07		19 20/09/2012 20:22	View	BAU High Media
23:	10101:49721854	01/10/2012 00:07		20 22/09/2012 14:57	View	BAU High Media
24:	10101:49721854	01/10/2012 00:07		21 22/09/2012 22:18	View	BAU High Media
25:	10101:49721854	01/10/2012 00:07		22 23/09/2012 21:06	View	BAU High Media

Histogram of cookie lifetime

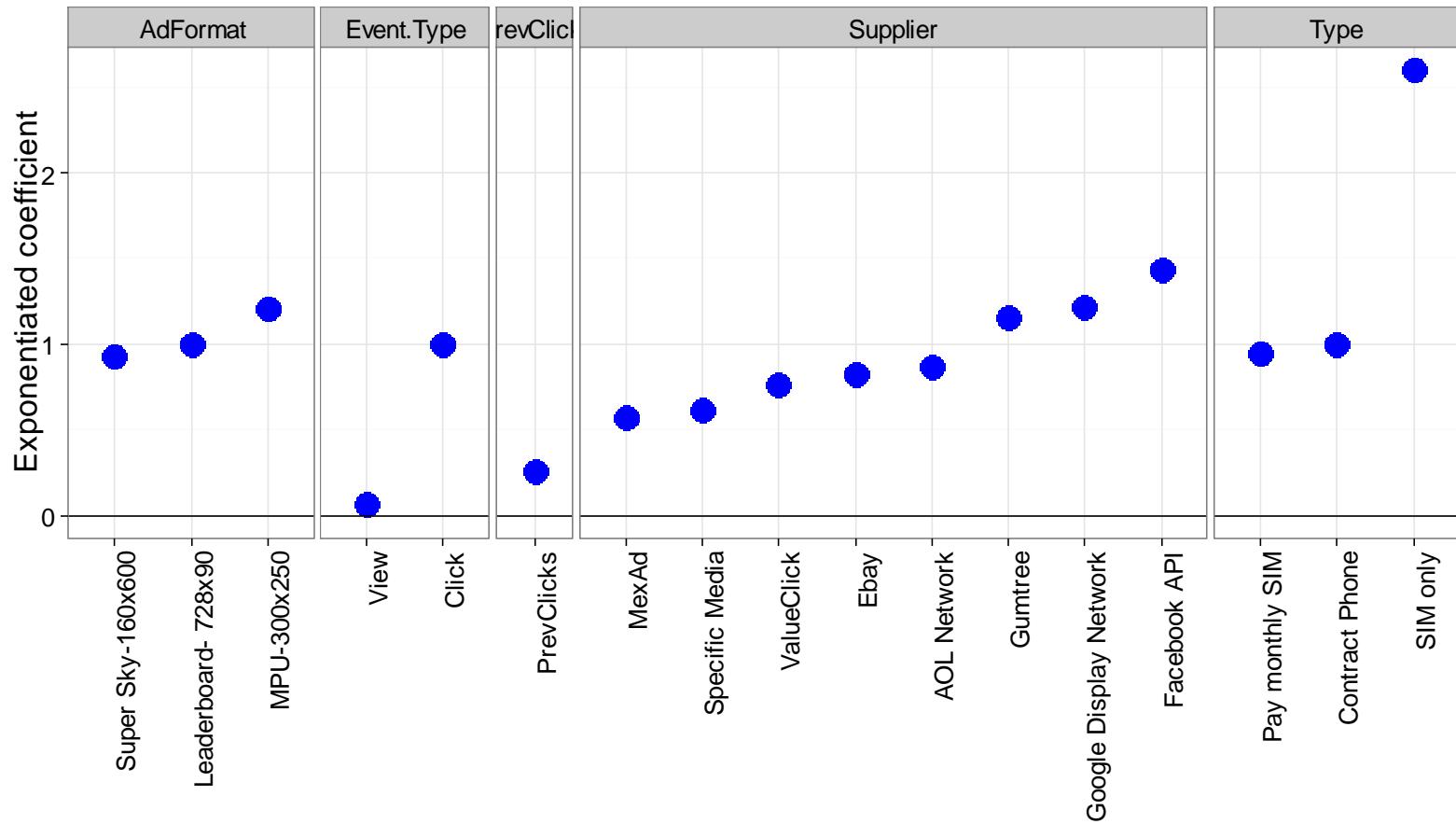
Impressions and clicks in customer journey



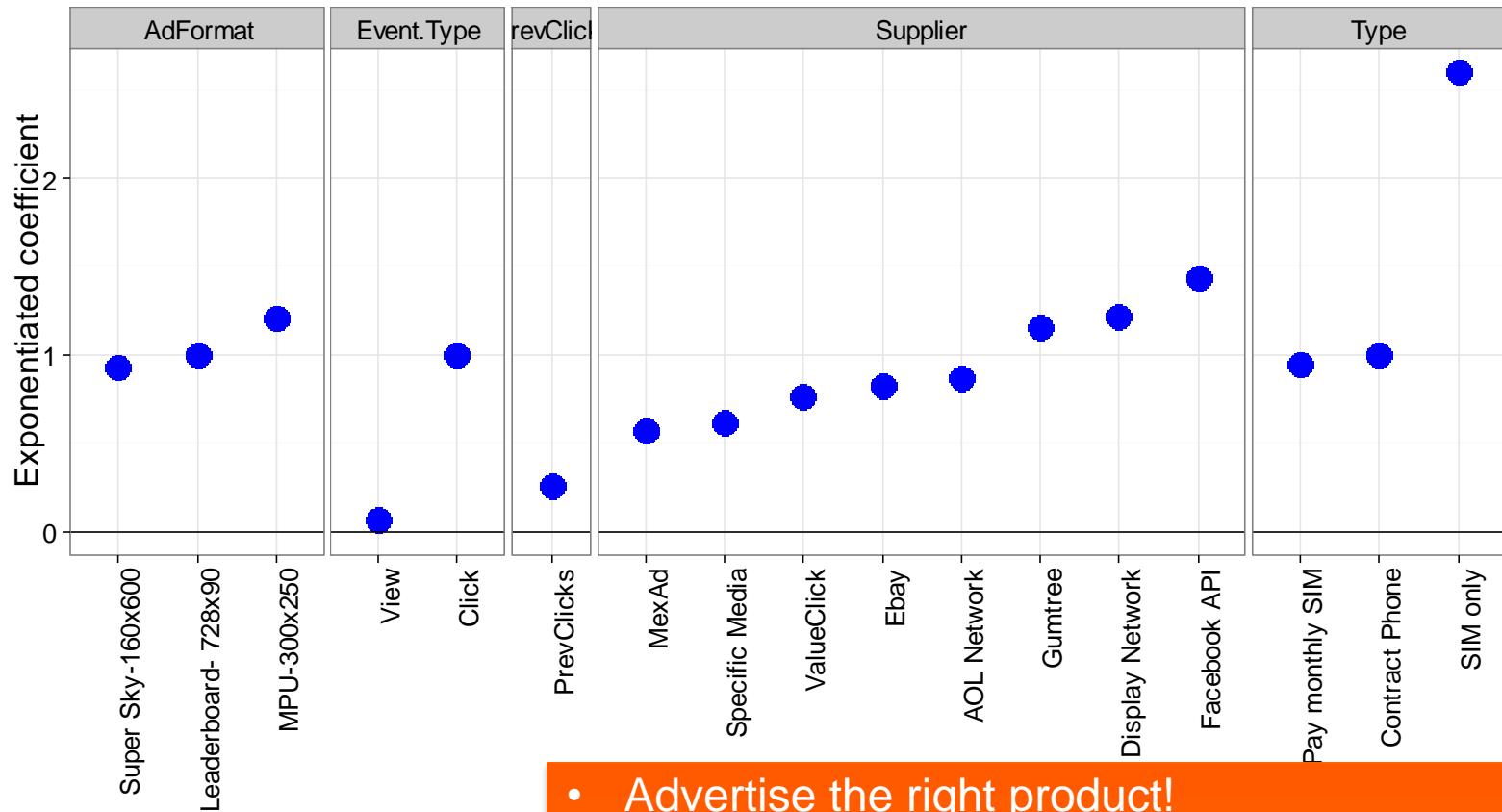
Fitting the model

```
> library(survival)
> fitp <- coxph(
  Surv(times, event=Converted) ~ Type +
    Event.Type +
    Supplier +
    PrevClicks +
    AdFormat,
  data=xd)
```

What does the data say?



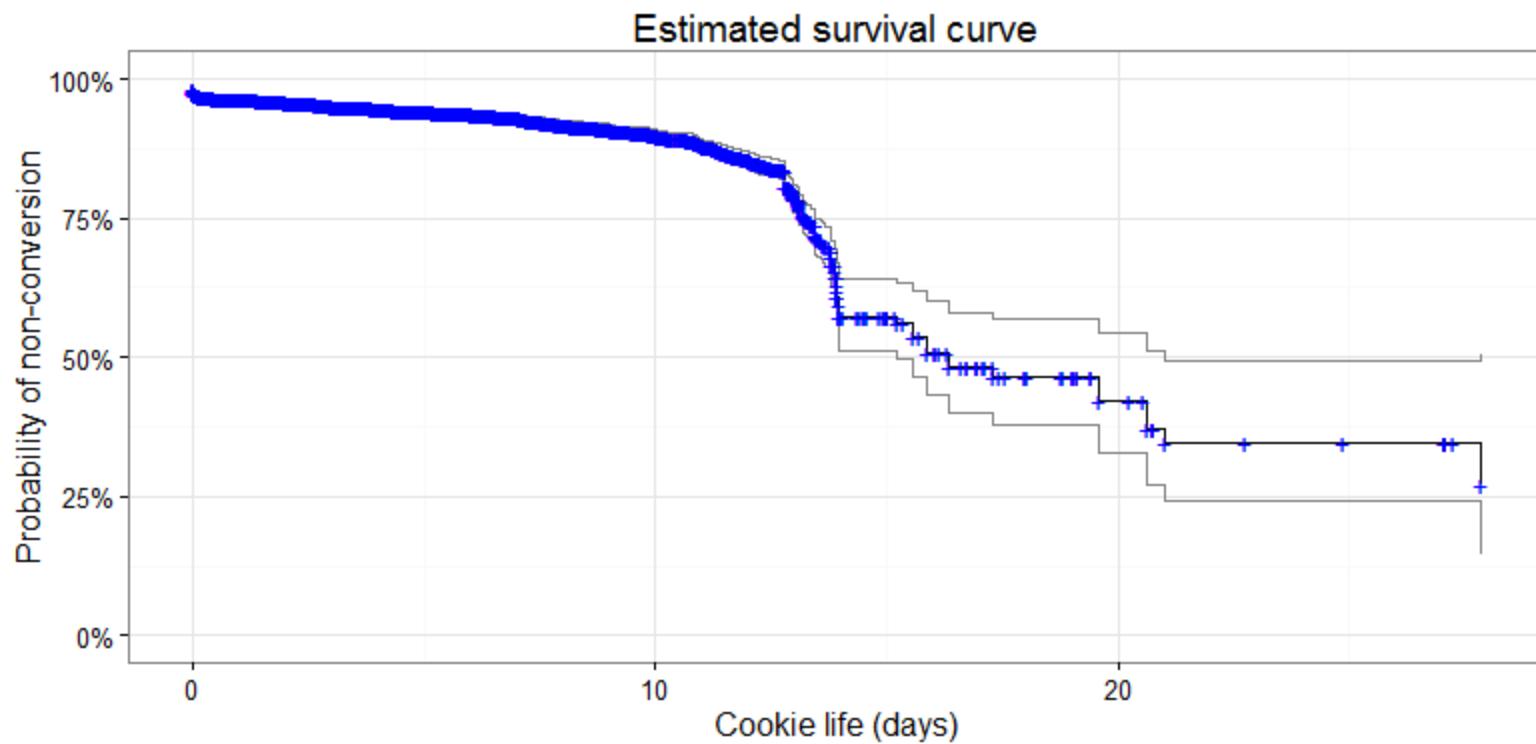
What does the data say?



- Advertise the right product!
- Some suppliers are better at generating conversion
- But note the data wasn't an unbiased experiment!

Estimated hazard function

```
> x <- survfit(fitp)
> xx <- with(x, data.frame(time, surv, upper, lower))
> ggplot(xx, aes(time, surv)) + geom_step() ...
```



Agenda

Digital marketing attribution

Using Survival models

At scale, on big data

Where Revolution helps

Revolution R Enterprise
Parallel external memory algorithms (PEMAs)

Import

- Text formats
- SAS
- High-speed database
- Hadoop

Pre-process

- DataStep
- Clean
- Refactor
- Sort
- Merge

Analyse

- Cube
- Summarise
- Parallelise (rxExec)

Model

- Regression
- GLM
- Tweedie
- Clustering
- Decision trees

Score

- Predict

Deploy

- Web services



Case study: Datasong

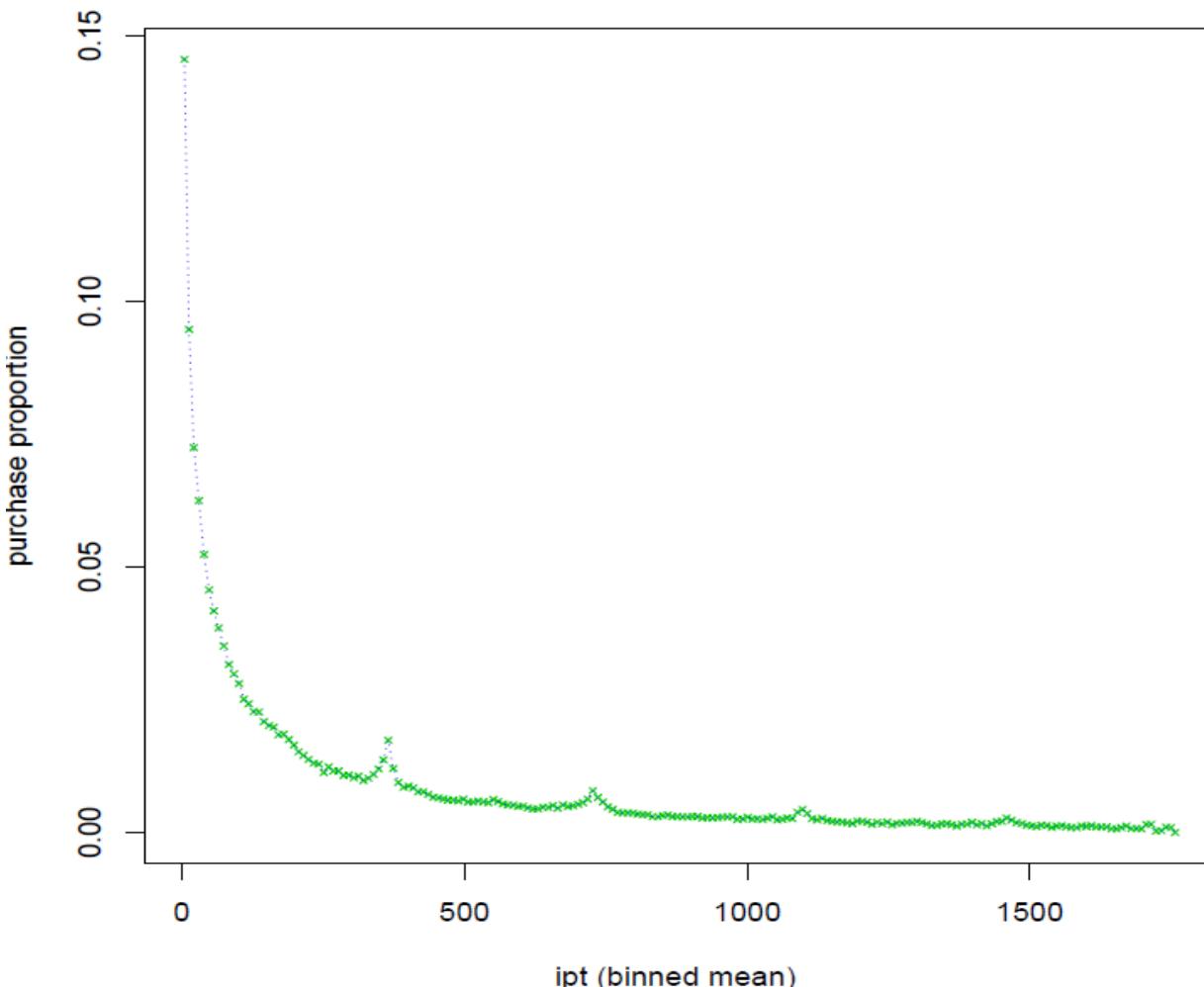
- Profile:
 - Multi-channel marketing analytics
 - Software developer and service provider
 - Growing, innovative, cost-conscious

- Technology:



Modeling the Baseline Hazard

Purchase Hazard by Binned Interpurchase Time



Capture nonlinear trends in baseline, while overlaying marketing treatment variables as well as other customer attributes

Revolution R package used:

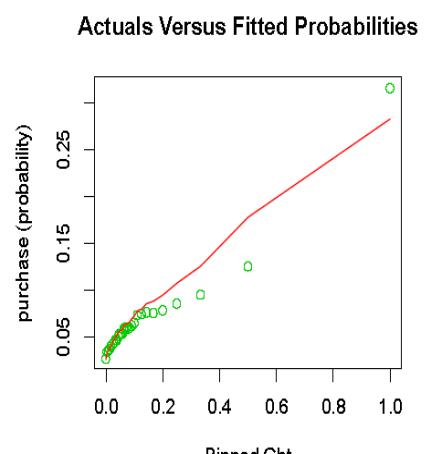
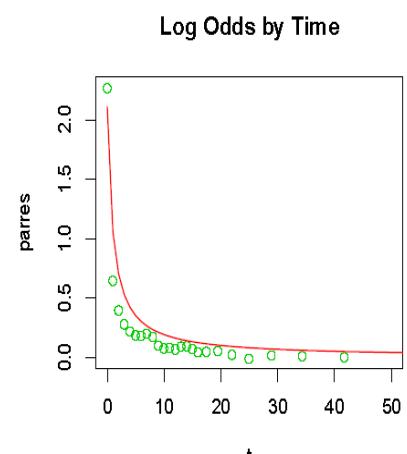
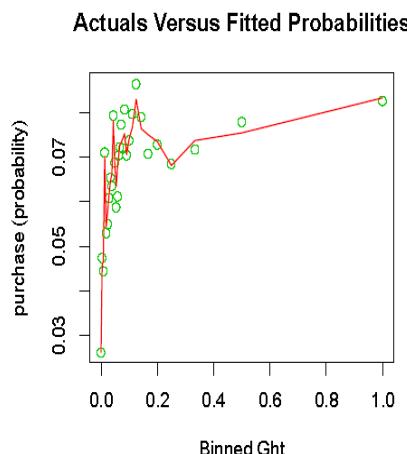
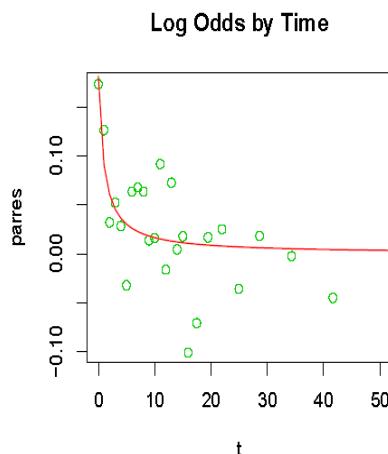
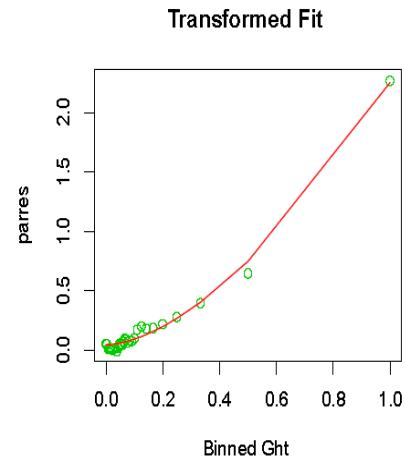
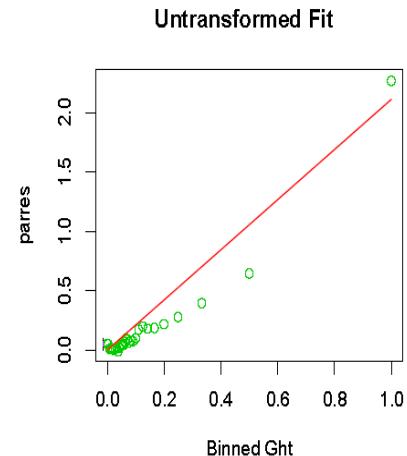
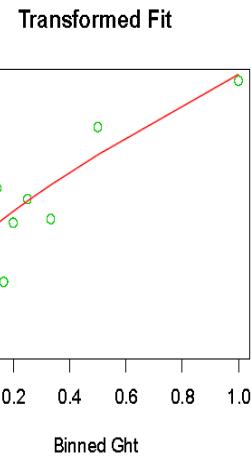
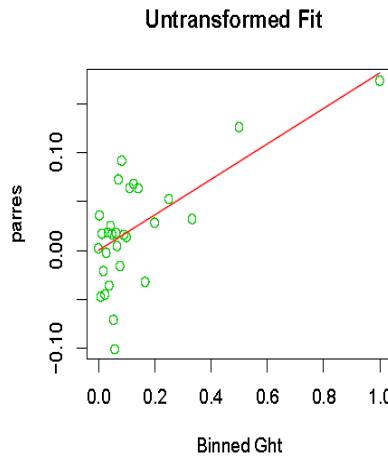
- RevoScaleR

Revolution R functions used:

- rxImport()
- rxSummary()
- rxCube()
- rxLogit()
- rxPredict()
- rxRoc()

Transformations

Catalog



Outcome

- Massively scalable infrastructure
 - Attribution and optimization at individual customer level for clients such as Williams-Sonoma
 - Client **saved \$250K** in one campaign
- Rapid deployment of customer-specific models
 - Innovative techniques, e.g. GAM Survival models
- Performance improvement
 - *Experienced 4x **performance** improvement on 50 million records*



The leading commercial provider of software and support for the popular open source R statistics language.

www.revolutionanalytics.com

Twitter: @RevolutionR