

AT!-TCH4H

Fundraising Analytics
Challenge

Feb 8- May 3, 2017



TTWYGTTT (i.e., Agenda)

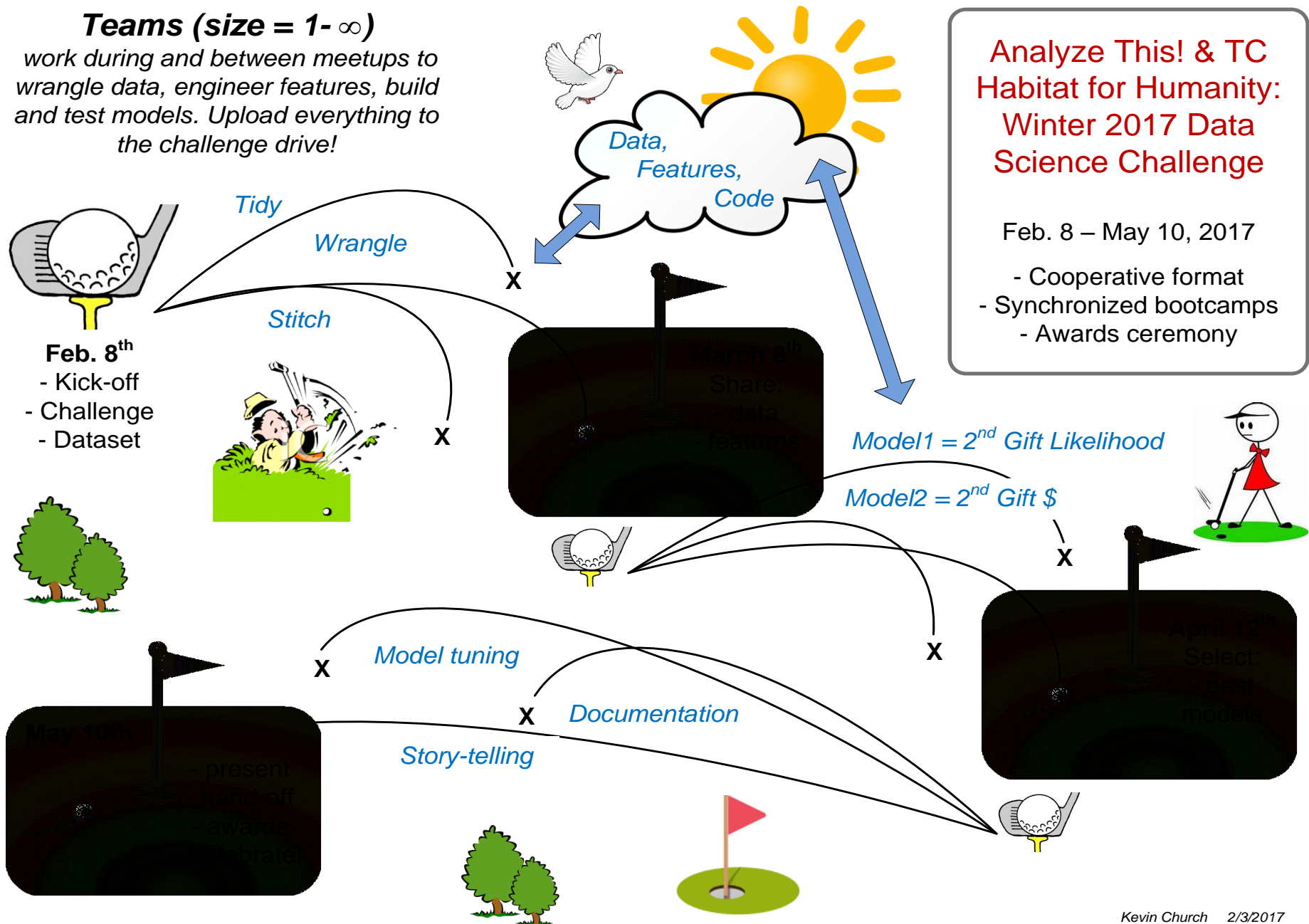
- *Cooperation vs. Competition*
- *The Cooperative framework – visual*
- *The Cooperative framework – in words*
- *The Challenge Question & proposed approach*
- *How to Apply the Model(s) x 2*
- *The Analytic Dataset*
- *Model 1 & 2*

Why Cooperate a.o.t. Compete?

- *Produce the “best” solution rather than the “winning” solution by combining the best data, features, models, tuning and presentation.*
- *Easier for members looking to learn because anyone can contribute (i.e., no prior expertise required to participate).*
- *Enable hack-a-thon like experiment at meetups.*
- *Redirect prize-money to a celebration*

Teams (size = 1- ∞)

work during and between meetups to wrangle data, engineer features, build and test models. Upload everything to the challenge drive!



Analyze This! & TC Habitat for Humanity: Winter 2017 Data Science Challenge

Feb. 8 – May 10, 2017

- Cooperative format
- Synchronized bootcamps
- Awards ceremony

A bit more about the challenge format

- Pre-work: formalize the question, prepare the data, generate baseline model
- Month1 (February 8 – March 8)
 - Feb8 Kickoff. Introduce the challenge, data & baseline models. Brainstorm features & 3rd party data. Distribute the tasks.
 - Feb11 Training No.1 (8A-12N, \$40). How to prepare an analytics dataset.
 - Feb9-Mar7. Upload data, features, code. Test various modeling approaches.

Challenge format (*cont.*)

- Month2 (March 8 – April 12)
 - Mar8 Meetup. Compare model(s) versions. Live demonstrations of model “tweaking”. Reach decision on best modeling approach. Agree on and distribute tuning tasks.
 - Mar11 Training No.2 (8A-5P, \$40/\$40).
 - AM – Binary Logistic Regression, PM – XGBoost.
 - Mar9-Apr12. Tune models. Continue to upload data, features, code.

Challenge format (*cont.*)

- Month3 (April 12 – May 3,10,17?)
 - Apr12 Meetup. Select final model(s) version. Live demonstrations of model “tweaking”. Reach decision on best modeling approach. Outline the presentation. Distribute presentation tasks.
 - Apr15 Training No.3 (8A-12N, \$40). How to communicate model results to a non-technical audience.
 - Mar9-Apr12. Validate the model with new, 1st-time donors from 10/1/13-21/31/13. Build and finalize the presentation

Challenge format (*cont.*)

- Grand Finale (May 3,10,17?) at TCH4H in St. Paul
 - Present the final models & code to TCH4H
 - Create a “transition” team
 - Awards ceremony & recognition
 - Celebrate!
 - Announce the next challenge

The Challenge Question

- How do we leverage data in Raiser's Edge and HubSpot to improve the efficiency of Renewals for 1st time donors? Or put another way, can we increase net revenue?
- *A proposed approach ...*
 - *Model 1. Predict the likelihood that a 1st-time donor gives a 2nd gift within 3.25 years*
 - *Model 2. Predict the expected 2nd gift amount*
 - *2nd Gift \$core = Likelihood * Gift Amount, \$*

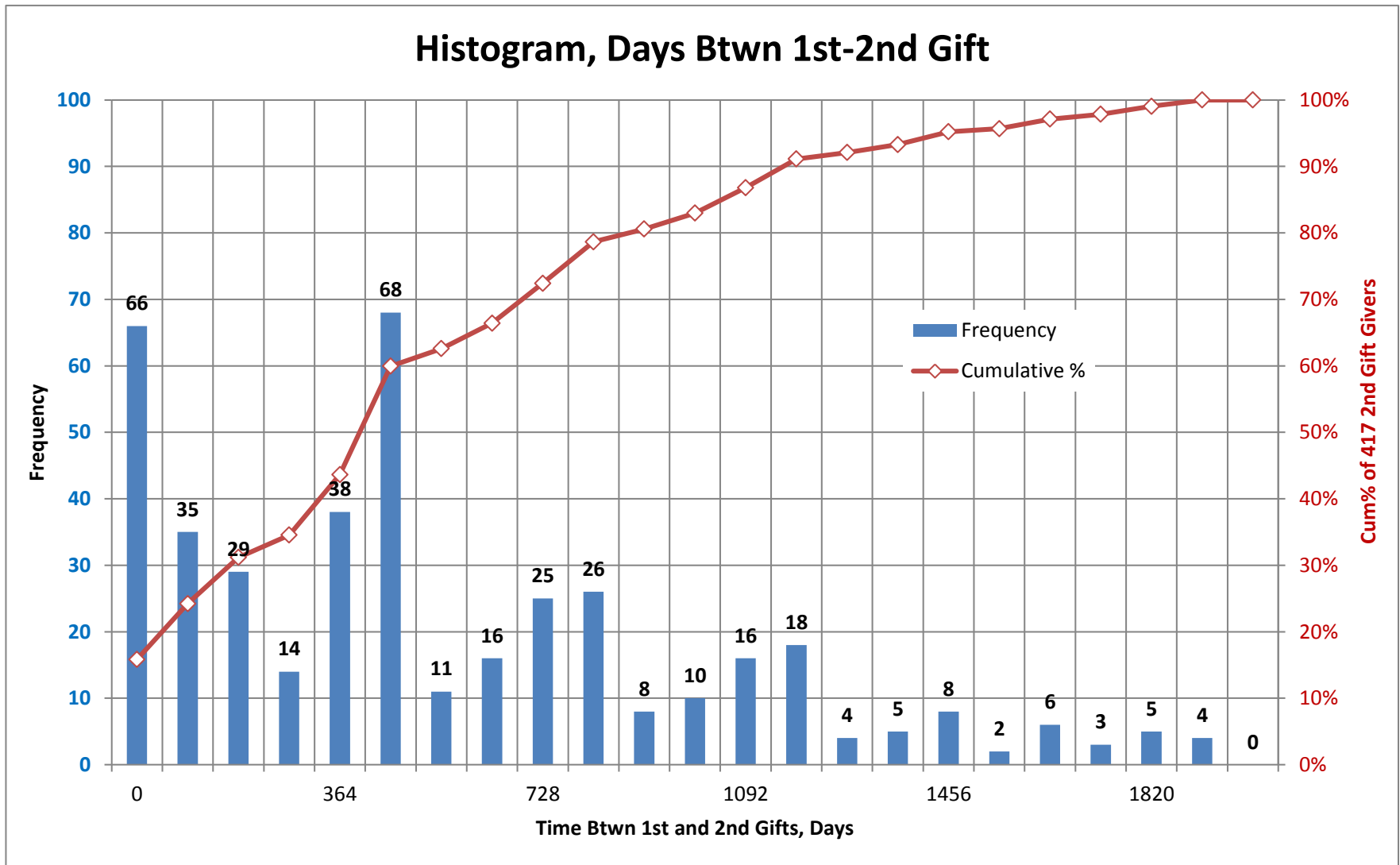
The Challenge Data Timeframe

1. First time donors were selected starting in 2011 to allow incorporation of HubSpot data into the analytic dataset
2. First time donors were stopped in September of 2013 to allow 3.25 years for a possible second gift (95% chance)
3. If a first time donor did not give a 2nd gift within 3.25 years of the first gift, there were labeled as 2ndGift=No

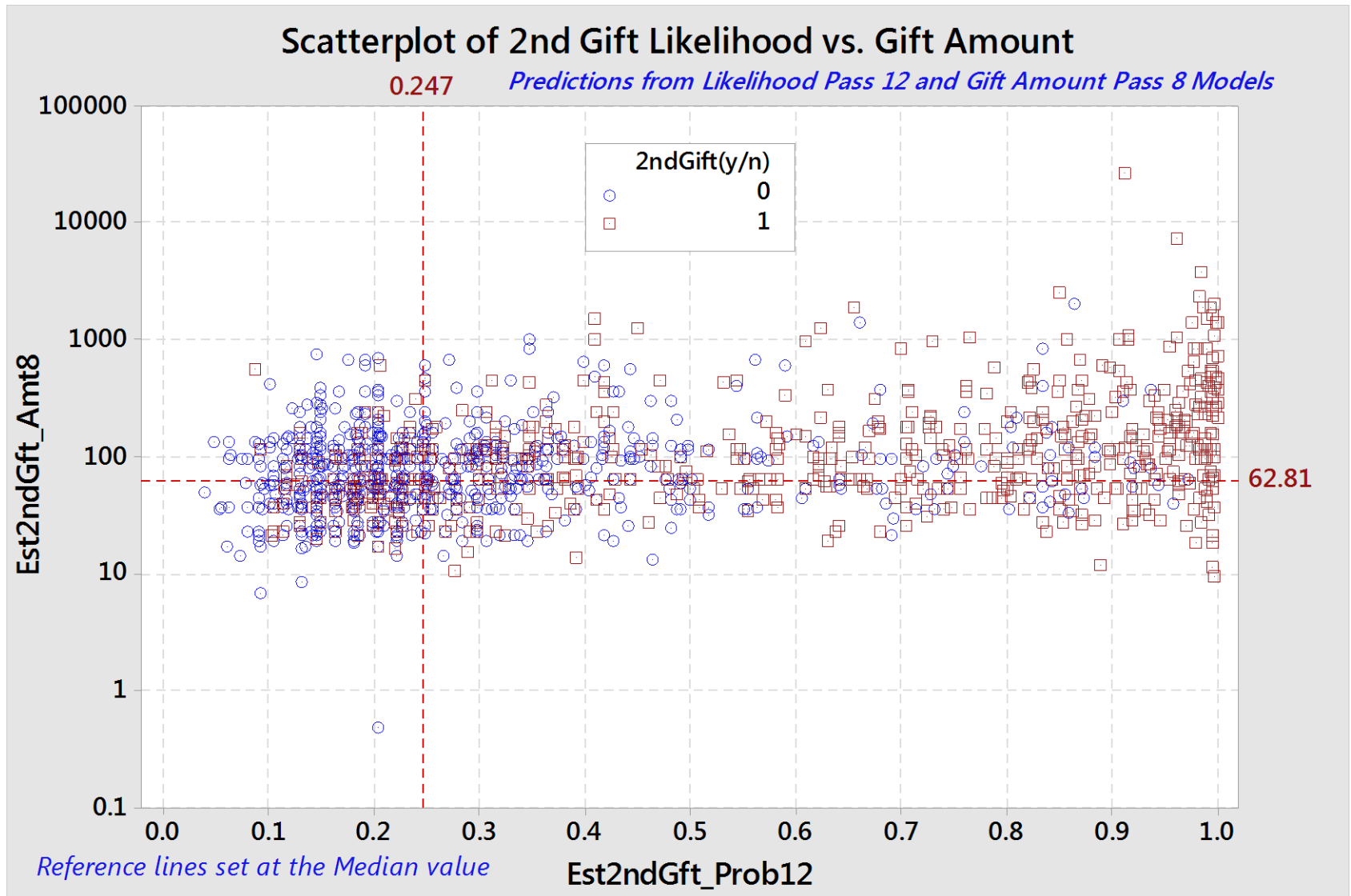
Target Timeframe for 1st-time donors												Validation?																																					
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec												
2011												2012	2013												2014												2015	2016											
1st Gift -----												2nd?																																					
1st Gift -----												2nd?																																					
1st Gift -----												2nd?																																					
												1st Gift -----												2nd?																									
												1st Gift -----												2nd?																									
												1st Gift -----												2nd?																									

Why wait 3.25 years for that 2nd gift?

1,294 1st-time donors from 2011, 417 (32.2%) gave a 2nd gift



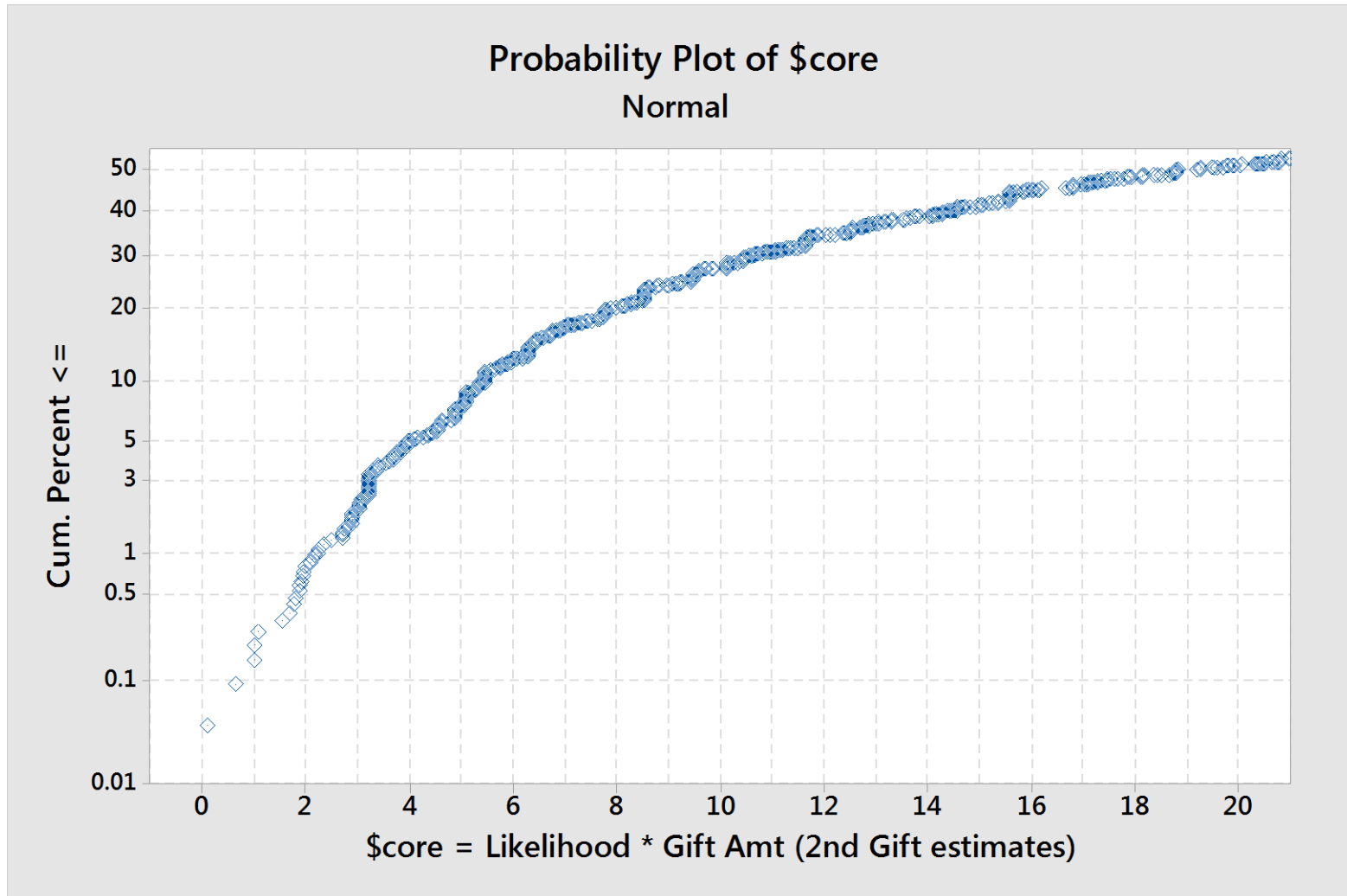
How to apply the models, part 1?



Based on 1,851 1st-time donors from 1/1/11-9/30/13, 37.8% of which gave a 2nd gift

How to apply the models, part 2?

What is the cost of a telemarketing phone call?



Insert overview of the dataset(s)

Analytic Dataset: Baseline model(s)

- 1,851 1st-time donors from Jan. 2011 to Sept. 2013.
- 699 gave a 2nd gift within 3.25 years of the first gift.
- Y1=Gave 2nd Gift (Y=1, N=0); Y2=2nd Gift Amount (\$).
- Demographics: ID, Gender, Marital Status, Age, City, State, ZIP, Email domain.
- Habitat Sites. Distance to nearest, ???
- 1st Gift: Date, Amount, Unsolicited (y/n).
- Email Activity: [(1st Send date + 1st Open date) = NotSent, Sent>3.25Yrs, NotOpened, Opened], Delivered, Clicked, Bounced, Sends since last engagement, Time first seen, HubSpot Scores, Original Source type, Lead date, Subscriber/Lead date, Blog Subscriber, Lifestyle stage, Volunteer date, Opt-Outs.
- Web Activity: Visits, Pageviews, Form Submissions, Last Send/Open dates.
- Engagement: Is Organization contact (y/n), Event participation, Meetings & Phone calls, First Conversion date.
- Social Media: Twitter Followers, Clicks (Twitter, Facebook, LinkedIn, Google+), Klout Score.
- Appeals: 1st Appeal date.
- 3rd Party: Median Income by ZIP code from 2011 US Census
- *Chicken vs. Egg: Total gifts (count, \$), min/max/avg gift amount*

Model 1: Likelihood of 2nd Gift

Binary Logistic Regression: 2ndGift(y/n) versus ...					
P12, Final Model					
Link function Logit					
Rows used	1851				
Response Information					
Variable	Value	Count			
2ndGift(y/n)	1	699	(Event, 37.76%)		
	0	1152			
	Total	1851			
Deviance Table					
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	22	656.87	29.858	656.87	0.000
Error	1828	1797.17	0.983		
Total	1850	2454.03			
Model Summary					Goodne
					Test
	R-Sq	R-Sq(adj)	AIC		Devianc
	26.77%	25.87%	1843.17		Pearson

Goodness-of-Fit Tests				
<u>Test</u>	<u>DF</u>	<u>Chi-Square</u>	<u>P-Value</u>	
Deviance	1828	1797.17	0.692	
Pearson	1828	1800.86	0.670	
Hosmer-Lemeshow	7	12.83	0.076	
Measures of Association				
<u>Pairs</u>	<u>Number</u>	<u>Percent</u>		
Concordant	650,469	80.8		
Discordant	145,339	18.0		
Ties	9,440	1.2		
Total	805,248	100.0		
* Association is between the response variable and predicted probabilities				

Model 1: Likelihood of 2nd Gift

	Coefficients							
	Term	Coef	SE Coef	T-Value	P-Value	VIF	Odds Ratio	95% CI
	Constant	2.6820	0.5260					
1	HbSptScrBin	0.0608	0.0191	3.18	0.001	1.26	1.063	(1.0237, 1.1032)
2	HbSptScrBin_10	-1.1900	0.2670	-4.46	0.000	1.82	0.304	(0.1803, 0.5131)
3	HbSptScrBin_15	0.7480	0.4000	1.87	0.057	1.54	2.112	(0.9642, 4.6268)
4	TotEvnts_0	-2.6370	0.2410	-10.94	0.000	3.08	0.072	(0.0446, 0.1149)
5	TotEvnts_1	-2.3420	0.2710	-8.64	0.000	3.01	0.096	(0.0565, 0.1633)
6	TotMtngFone_0	-0.7400	0.2390	-3.10	0.002	1.06	0.477	(0.2985, 0.7632)
7	AgeInd_NA	0.4020	0.1310	3.07	0.002	1.20	1.494	(1.1565, 1.9310)
8	AgeInd_44	-0.6720	0.3840	-1.75	0.073	1.08	0.511	(0.2405, 1.0851)
9	AgeInd_79	0.9900	0.5480	1.81	0.074	1.03	2.691	(0.9198, 7.8723)
10	AgeInd_80	2.0610	0.6980	2.95	0.002	1.03	7.852	(1.9998, 30.8282)
11	Pgvus_0	-0.8160	0.1660	-4.92	0.000	1.66	0.442	(0.3195, 0.6123)
12	FrmSubmits_0	-0.6190	0.1660	-3.73	0.000	1.29	0.538	(0.3886, 0.7455)
13	BrdcstClckBin_0	0.9380	0.2330	4.03	0.000	1.42	2.554	(1.6163, 4.0345)
14	1stEmail_Sent>3.25Yrs	1.5140	0.4700	3.22	0.001	1.12	4.545	(1.8108, 11.4085)
15	1stGftInd_150	0.6240	0.2380	2.62	0.010	1.03	1.866	(1.1710, 2.9730)
16	1stGftInd_50	0.2640	0.1480	1.78	0.077	1.04	1.302	(0.9738, 1.7396)
17	1stGftMo_Jul	-0.5420	0.1740	-3.11	0.001	1.05	0.582	(0.4140, 0.8176)
18	1stGftYr_2013	-0.3740	0.1340	-2.79	0.005	1.05	0.688	(0.5294, 0.8945)
19	OrigSrcTyp_OrgSrch	-0.8610	0.4740	-1.82	0.067	1.09	0.423	(0.1669, 1.0711)
20	OrigSrcTyp_Rfrs	-1.3160	0.7870	-1.67	0.075	1.04	0.268	(0.0574, 1.2537)
21	ZlPtxt_551	0.5120	0.1500	3.41	0.001	1.21	1.668	(1.2425, 2.2397)
22	ZlPtxt_554	0.2570	0.1380	1.86	0.062	1.19	1.293	(0.9875, 1.6940)

Model 2: Amount of 2nd Gift

Regression Analysis: Log10(2ndGift) versus ...					
P8, Final Model					
Rows unused 1155 (1152 no 2nd gift, 3 outliers)					
Analysis of Variance					
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	12	133.077	11.0898	84.45	0.000
Error	683	89.693	0.1313		
Total	695	222.77			
Model Summary					
Sy.x	R-sq	R-sq(adj)			
0.362384	59.74%	59.03%			

Coefficients						
Term	Coef	SE Coef	T-Value	P-Value	VIF	
Constant	0.8571	0.07480	11.5	0		
1 log10(1stGft\$)	0.5742	0.02930	19.62	0.000	1.45	
2 1stGftInd_20	-0.1479	0.06360	-2.32	0.020	1.13	
3 1stGftInd_100	0.0828	0.03530	2.34	0.019	1.06	
4 TotMtngFone_2-24	0.4400	0.05440	8.09	0.000	1.34	
5 Visits	-0.0013	0.00046	-2.77	0.006	1.27	
6 HbSptScrBin_15	0.1798	0.04370	4.12	0.000	1.05	
7 EmailClckd_10-212	-0.2130	0.06720	-3.17	0.002	1.39	
8 OrigSrcTyp_OffInSrc	-0.1144	0.04660	-2.45	0.014	1.42	
9 OrigSrcTyp_OrgSrch	-0.1948	0.08760	-2.22	0.026	1.30	
10 AgeInd_NA	0.0797	0.02950	2.70	0.007	1.14	
11 AgeInd_64	0.1769	0.06310	2.80	0.005	1.09	
12 ZIPtxt_SE	0.4560	0.21300	2.14	0.032	1.03	