Texas Advancement Analytics Symposium

TEXAS
The University of Texas at Austin

Patterns of Philanthropy: Using Pattern Mining for Predictive Analysis in Advancement and Fundraising

Klaus Mueller, PhD and Eric Papenhausen, PhD

Akai Kaeru LLC and Stony Brook University

Donations & Academia

small donor

University endowment rankings (2019)

- Harvard: $41 Billion
  - annual increase: 1.5 Billion (3%)
  - compare with annual budget: 4.5 Billion (10%)
- Yale: $30 Billion
- Stanford $28 Billion
- Princeton $26 Billion
- ….
- Stony Brook: 380 Million

midsize donor

mega donor

Identifying the Donors

These days a wealth of personal data is collected by universities

- demographics
- family and friends
- geo locations
- academics
- club memberships
- prior donation activities
- ….
- we will call these properties “**features**”

Use these data to shape specific fund raising efforts

.. and evaluate their expected profitability

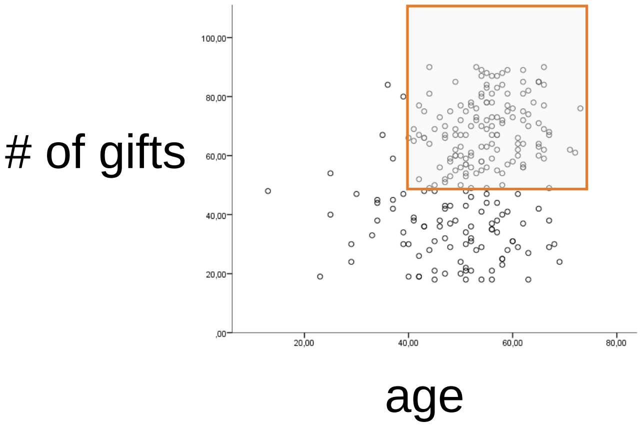
Looking Under the Hood

Our software has two main components

- **Pattern Miner**: searches for groups of donors with similar features and similar donation behavior
- **Pattern Browser:** allows analysts to explore these patterns and extract actionable insights

A pattern is

- a subpopulation of donors that
- fits inside a low-dimensional hypercube that
- has well-defined value ranges of the donor features



Patterns Must Also Be Interesting!

What makes a group of donors interesting?

- right -- when they have a high probability of donating

An interesting pattern is thus a group of similar people where

- their probability of a specific type of donation is significantly higher than the probability of the general population

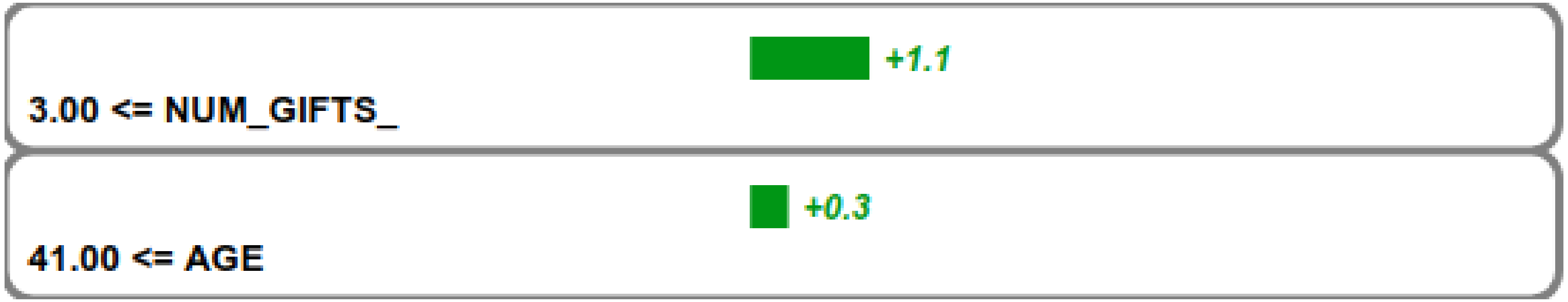
- our Pattern Miner extracts these interesting patterns automatically via statistical hypothesis testing (Mann-Whitney, $\chi 2$ test for independence)

Let’s See an Example (a 2D Pattern)

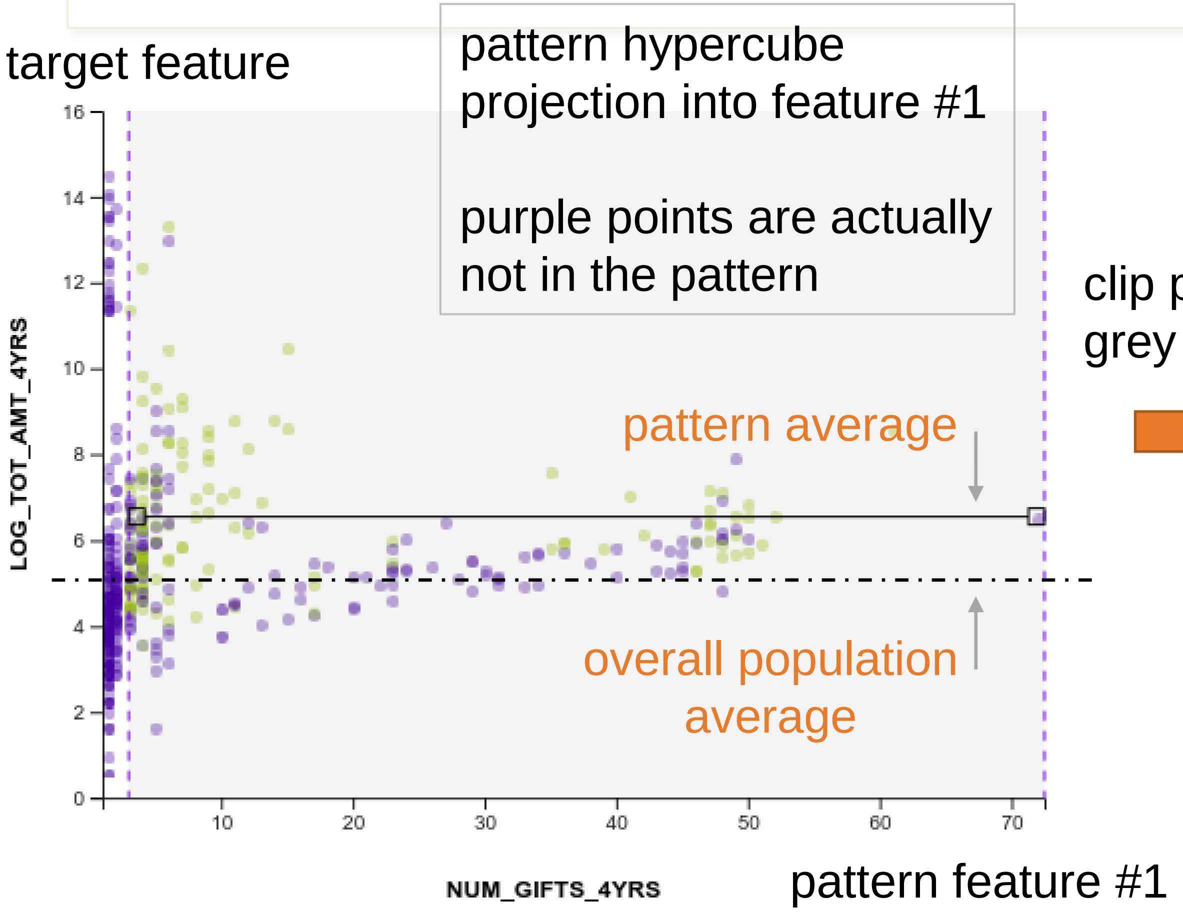
Total of amount of gifts in FY 2015-2018

Description (High NUM_GIFTS_4YRS + High AGE)

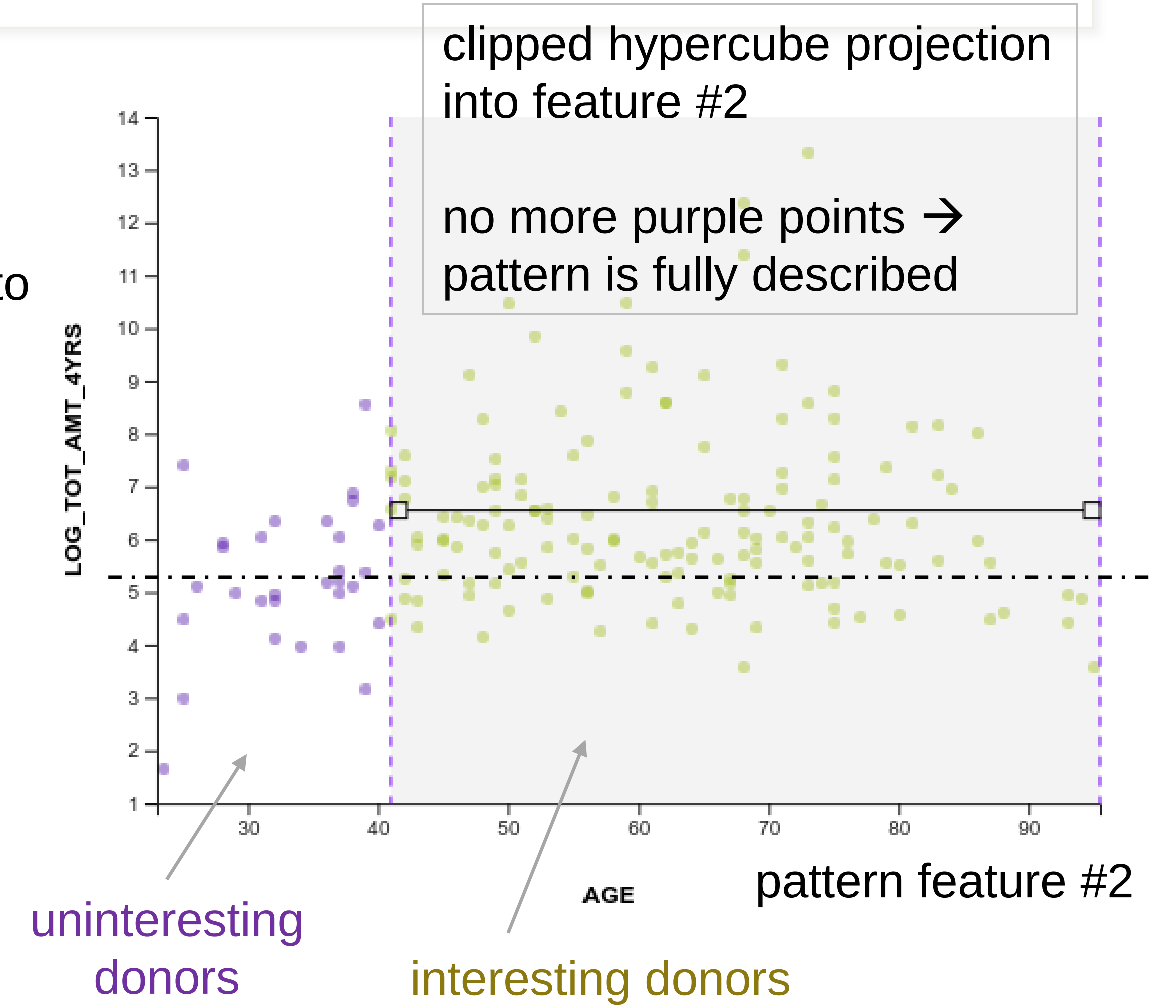
The probability that a randomly selected point within this group has a **higher** **LOG_TOT_AMT_4YRS** than any point outside this group is **0.88**. This finding is statistically highly significant.



Example Continued….



clip points to grey box

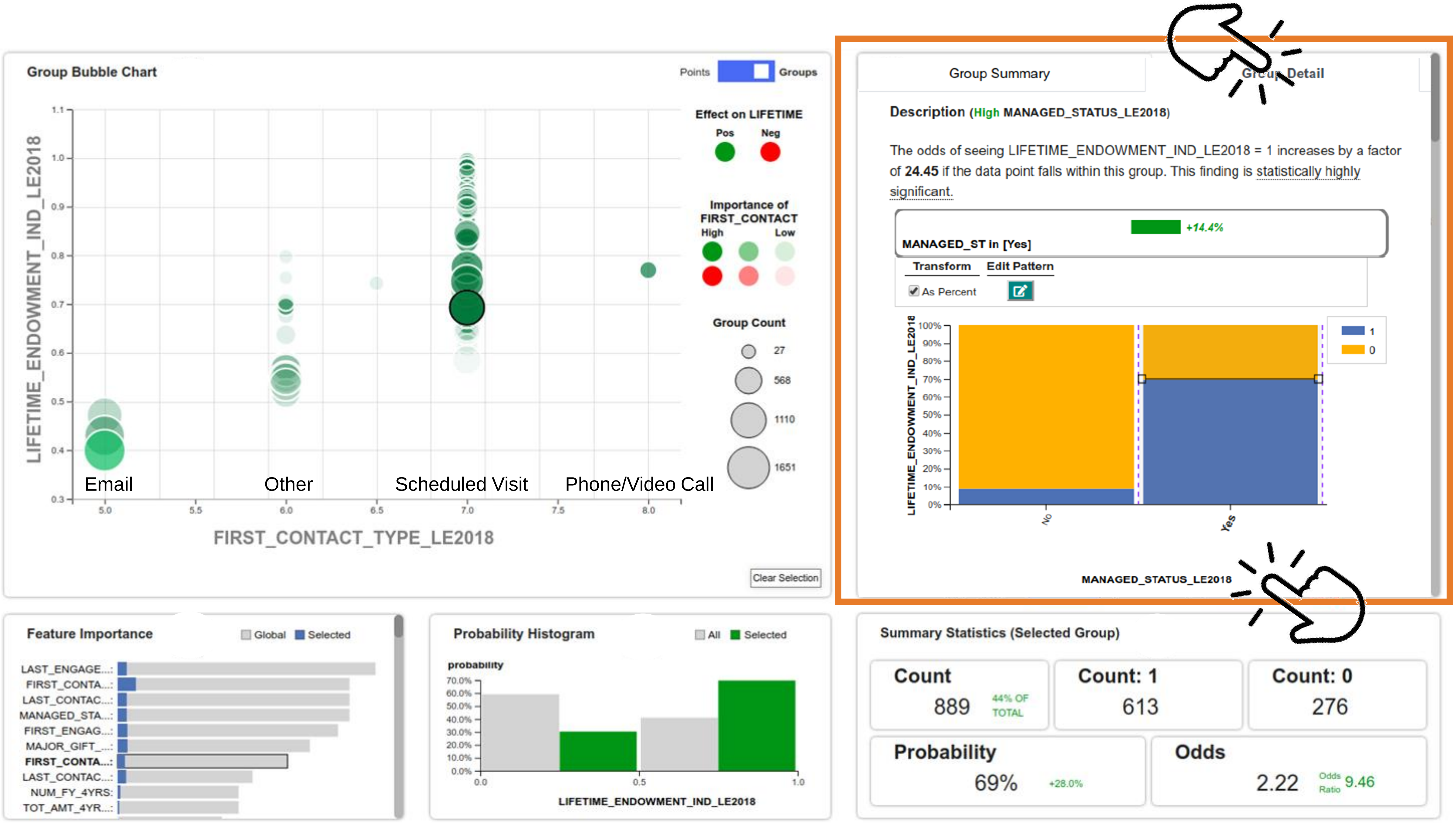


Next: A Tour of the Visual Pattern Browser

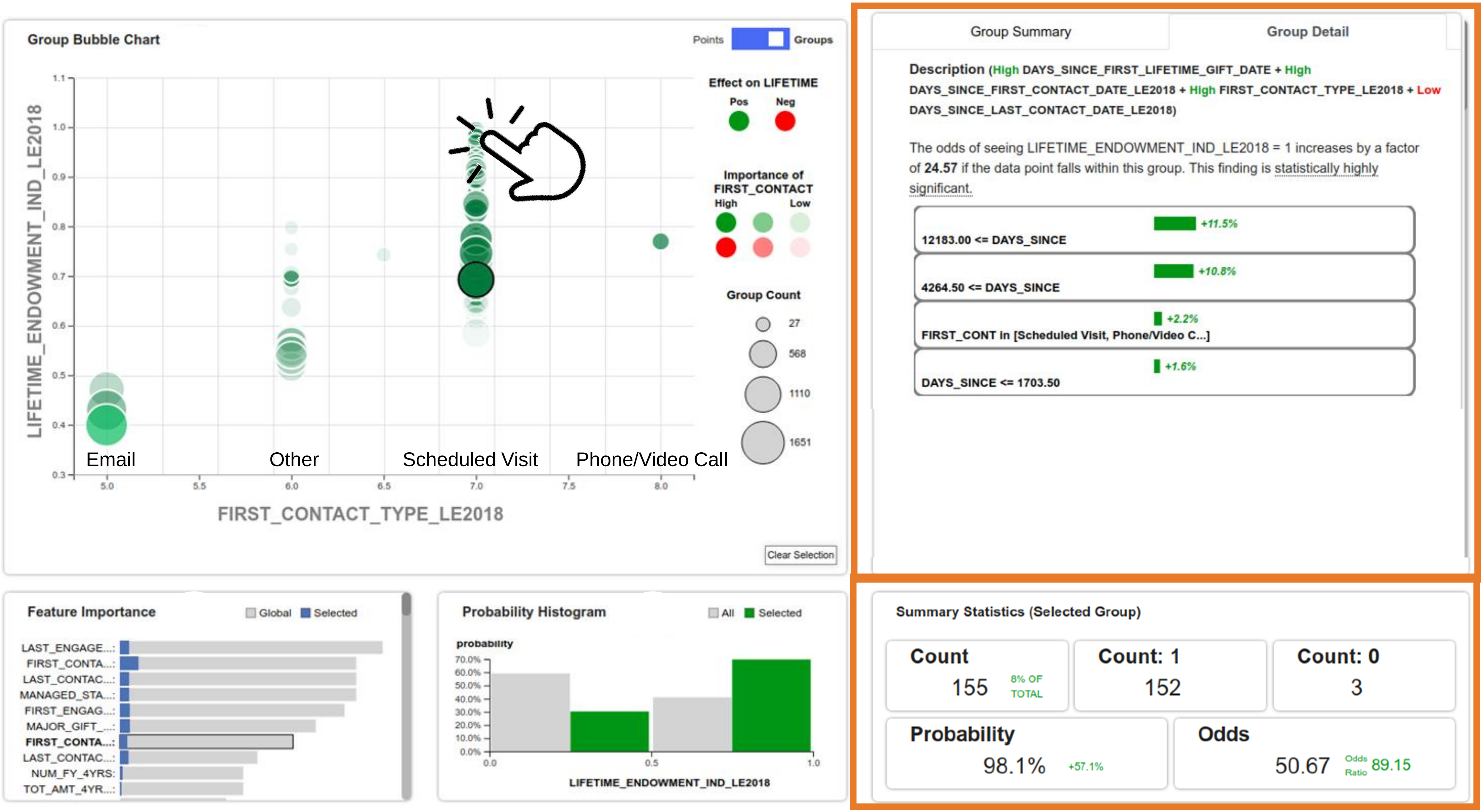
Example:

- what kind of donor is likely to make a Lifetime Endowment and how much
- history is captured by the indicator LIFETIME_ENDOWMENT_IND (0/1)

Group Bubble Chart
Points
Groups
Effect on LIFETIME
Pos
Neg
Importance of FIRST_CONTACT
High
Low
Group Count
27
568
1110
1651
LIFETIME_ENDOWMENT_IND_LE2018
Email
Other
Scheduled Visit
Phone/Video Call
FIRST_CONTACT_TYPE_LE2018
Clear Selection
Group Summary
Group Detail
You selected 1 Group
Points:
Few
Many
Attribute Range
Low
Med
High
Effect on LIFETIME_ENDOWMENT_IN
FIRST_CONTA...
MAJOR_GIFT_...
FIRST_ENGAG...
LAST_CONTAC...
LAST_ENGAGE...
LAST_CONTAC...
MANAGED_STA...
FIRST_CONTA...
UT_FRIEND
NGIFT_4YRS_...
NUM_FY_4YRS
NUM_SIBLING
DB_4YRS
+28.7
+17.6%
+16.0%
+15.4%
+14.4%
+14.4%
+14.4%
+14.0%
+5.8%
+4.7%
+4.7%
-4.0%
-4.0%
Feature Importance
Global
Selected
LAST_ENGAGE...:
FIRST_CONTA...:
LAST_CONTAC...:
MANAGED_STA...:
FIRST_ENGAG...:
MAJOR_GIFT_...:
FIRST_CONTA...:
LAST_CONTAC...:
NUM_FY_4YRS:
TOT_AMT_4YR...:
Probability Histogram
All
Selected
probability
LIFETIME_ENDOWMENT_IND_LE2018
Summary Statistics (Selected Group)
Count
889
44% OF TOTAL
Count: 1
613
Count: 0
276
Probability
69%
+28.0%
Odds
2.22
Odds Ratio
9.46



Group Bubble Chart
Points
Groups
Effect on LIFETIME
Pos
Neg
Importance of FIRST_CONTACT
High
Low
Group Count
27
568
1110
1651
LIFETIME_ENDOWMENT_IND_LE2018
Email
Other
Scheduled Visit
Phone/Video Call
FIRST_CONTACT_TYPE_LE2018
Clear Selection
Group Summary
Group Detail
Description (High MANAGED_STATUS_LE2018)
The odds of seeing LIFETIME_ENDOWMENT_IND_LE2018 = 1 increases by a factor of 24.45 if the data point falls within this group. This finding is statistically highly significant.
MANAGED_ST in [Yes]
+14.4%
Transform
Edit Pattern
As Percent
No
Yes
MANAGED_STATUS_LE2018
Feature Importance
Global
Selected
LAST_ENGAGE...:
FIRST_CONTA...:
LAST_CONTAC...:
MANAGED_STA...:
FIRST_ENGAG...:
MAJOR_GIFT_...:
FIRST_CONTA...:
LAST_CONTAC...:
NUM_FY_4YRS:
TOT_AMT_4YR...:
Probability Histogram
All
Selected
probability
LIFETIME_ENDOWMENT_IND_LE2018
Summary Statistics (Selected Group)
Count
889
44% OF TOTAL
Count: 1
613
Count: 0
276
Probability
69%
+28.0%
Odds
2.22
Odds Ratio
9.46


Group Bubble Chart
Points
Groups
LIFETIME_ENDOWMENT_IND_LE2018
FIRST_CONTACT_TYPE_LE2018
Email
Other
Scheduled Visit
Phone/Video Call
Effect on LIFETIME
Pos
Neg
Importance of FIRST_CONTACT
High
Low
Group Count
27
568
1110
1651
Clear Selection
Group Summary
Group Detail
Description (High DAYS_SINCE_FIRST_LIFETIME_GIFT_DATE + High DAYS_SINCE_FIRST_CONTACT_DATE_LE2018 + High FIRST_CONTACT_TYPE_LE2018 + Low DAYS_SINCE_LAST_CONTACT_DATE_LE2018)
The odds of seeing LIFETIME_ENDOWMENT_IND_LE2018 = 1 increases by a factor of 24.57 if the data point falls within this group. This finding is statistically highly significant.
12183.00 <= DAYS_SINCE
+11.5%
4264.50 <= DAYS_SINCE
+10.8%
FIRST_CONT in [Scheduled Visit, Phone/Video C...]
+2.2%
DAYS_SINCE <= 1703.50
+1.6%
Feature Importance
Global
Selected
LAST_ENGAGE...:
FIRST_CONTA...:
LAST_CONTAC...:
MANAGED_STA...:
FIRST_ENGAG...:
MAJOR_GIFT_...:
FIRST_CONTA...:
LAST_CONTAC...:
NUM_FY_4YRS:
TOT_AMT_4YR...:
Probability Histogram
All
Selected
probability
LIFETIME_ENDOWMENT_IND_LE2018
Summary Statistics (Selected Group)
Count
155
8% OF TOTAL
Count: 1
152
Count: 0
3
Probability
98.1%
+57.1%
Odds
50.67
Odds Ratio
89.15

Takeaways From This First Study

Emailing gives little hope for lifetime endowments

- not much more than not doing anything (40%)

Scheduled visits are much better

- managing donors is the way to go (70%)

Frequent contact pays off for managed donors

- almost 100%

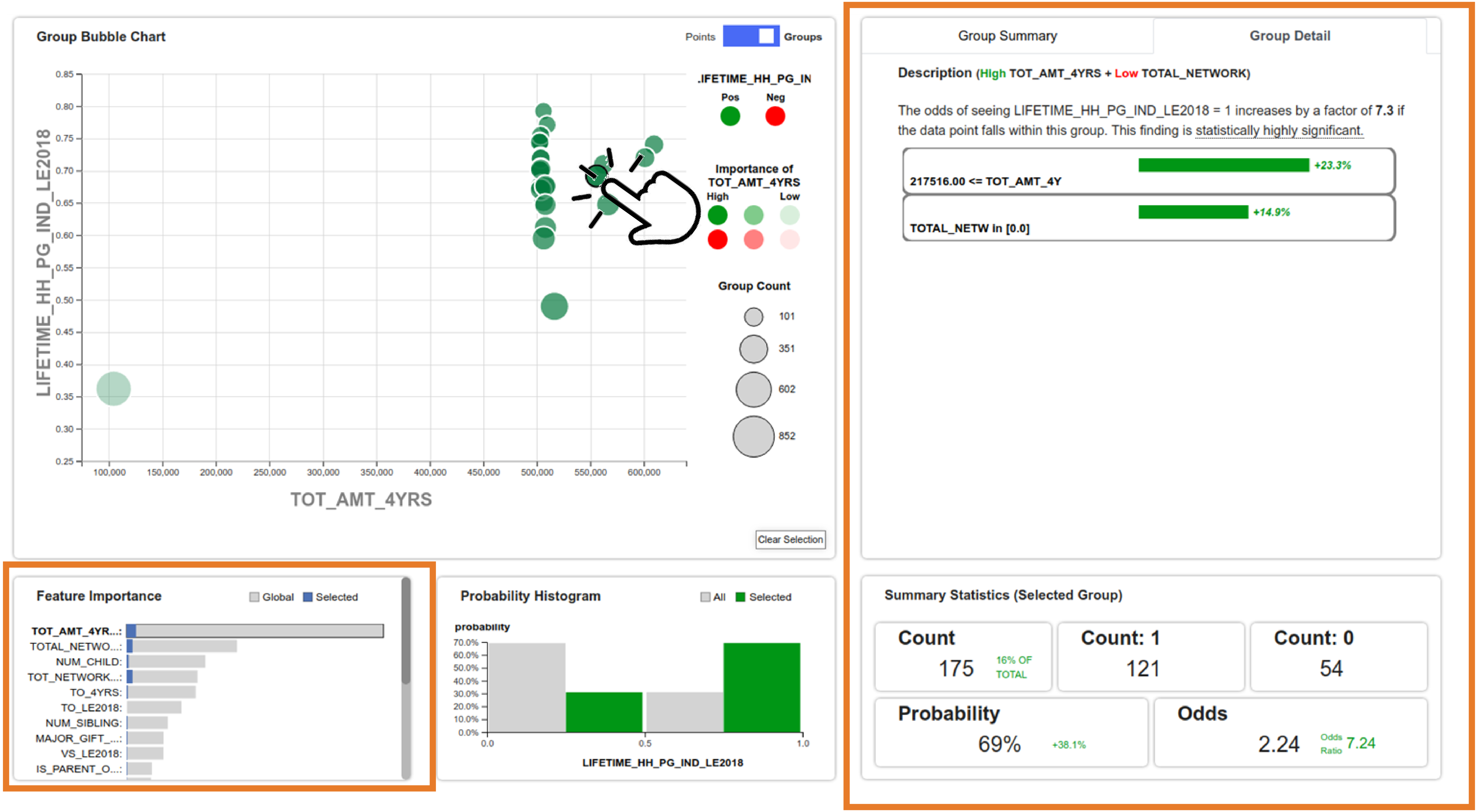
Next: Who Will Make a Planned Gift

Planned gifts are typically difficult to predict

- they often occur in a will, after the donor has passed
- there is rarely a prior announcement
- they are usually considerable sums of money

Predictive analysis based on historical data can give the insight

- find the type of secret donor who will end up making a Planned Gift
- captured by the indicator LIFETIME_HH_PG_IND (0/1)



Group Bubble Chart
Points
Groups
LIFETIME_HH_PG_IND_LE2018
TOT_AMT_4YRS
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
100,000
150,000
200,000
250,000
300,000
350,000
400,000
450,000
500,000
550,000
600,000
.IFETIME_HH_PG_IN
Pos
Neg
Importance of TOT_AMT_4YRS
High
Low
Group Count
101
351
602
852
Clear Selection
Group Summary
Group Detail
Description (High TOT_AMT_4YRS + Low TOTAL_NETWORK)
The odds of seeing LIFETIME_HH_PG_IND_LE2018 = 1 increases by a factor of 7.3 if the data point falls within this group. This finding is statistically highly significant.
217516.00 <= TOT_AMT_4Y
+23.3%
TOTAL_NETW in [0.0]
+14.9%
Feature Importance
Global
Selected
TOT_AMT_4YR...:
TOTAL_NETWO...:
NUM_CHILD:
TOT_NETWORK...:
TO_4YRS:
TO_LE2018:
NUM_SIBLING:
MAJOR_GIFT_...:
VS_LE2018:
IS_PARENT_O...:
Probability Histogram
All
Selected
probability
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
0.0
0.5
1.0
LIFETIME_HH_PG_IND_LE2018
Summary Statistics (Selected Group)
Count
175
16% OF TOTAL
Count: 1
121
Count: 0
54
Probability
69%
+38.1%
Odds
2.24
Odds Ratio
7.24

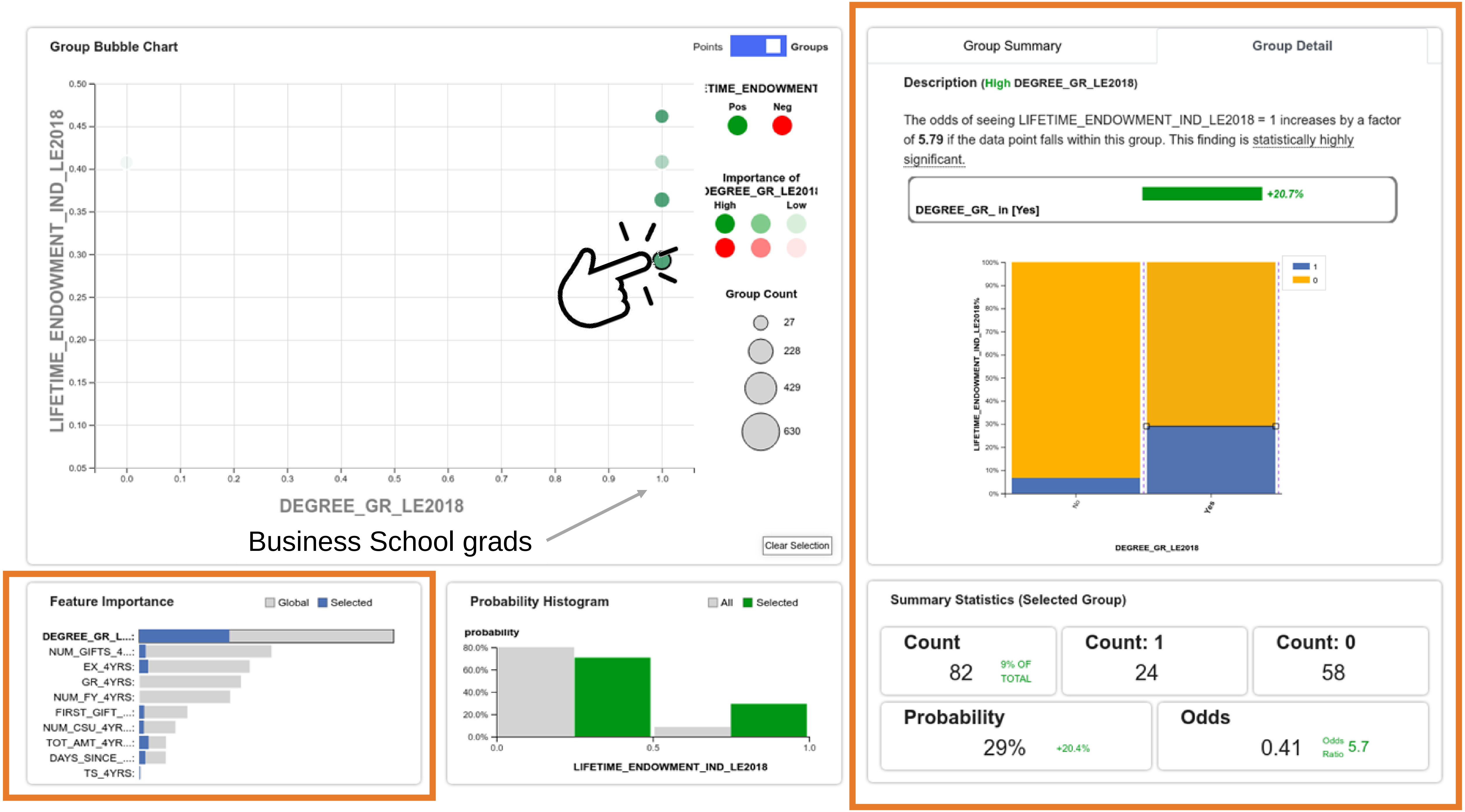

Group Bubble Chart
Points
Groups
LIFETIME_HH_PG_IN
Pos
Neg
Importance of TOT_AMT_4YRS
High
Low
Group Count
101
351
602
852
LIFETIME_HH_PG_IND_LE2018
TOT_AMT_4YRS
Clear Selection
Group Summary
Group Detail
Description (High TOT_AMT_4YRS + Low TOTAL_NETWORK)
The odds of seeing LIFETIME_HH_PG_IND_LE2018 = 1 increases by a factor of 7.3 if the data point falls within this group. This finding is statistically highly significant.
217516.00 <= TOT_AMT_4Y
+23.3%
TOTAL_NETW in [0.0]
+14.9%
LIFETIME_HH_PG_IND_LE2018%
TOTAL_NETWORK
Feature Importance
Global
Selected
TOT_AMT_4YR...:
TOTAL_NETWO...:
NUM_CHILD:
TOT_NETWORK...:
TO_4YRS:
TO_LE2018:
NUM_SIBLING:
MAJOR_GIFT_...:
VS_LE2018:
IS_PARENT_O...:
Probability Histogram
All
Selected
probability
LIFETIME_HH_PG_IND_LE2018
Summary Statistics (Selected Group)
Count
175
16% OF TOTAL
Count: 1
121
Count: 0
54
Probability
69%
+38.1%
Odds
2.24
Odds Ratio
7.24

Identify the Most Charitable Unmanaged Donors

This has been a so-far neglected group

- are there any donors who might be forgotten?
- what kinds of people are they?
- can they be converted to managed donors?

- let’s have a look at Lifetime Endowment


Group Bubble Chart
Points
Groups
LIFETIME_ENDOWMENT_IND_LE2018
DEGREE_GR_LE2018
Pos
Neg
Importance of DEGREE_GR_LE2018
High
Low
Group Count
27
228
429
630
Business School grads
Clear Selection
Group Summary
Group Detail
Description (High DEGREE_GR_LE2018)
The odds of seeing LIFETIME_ENDOWMENT_IND_LE2018 = 1 increases by a factor of 5.79 if the data point falls within this group. This finding is statistically highly significant.
DEGREE_GR_ in [Yes]
+20.7%
No
Yes
Feature Importance
Global
Selected
DEGREE_GR_L...:
NUM_GIFTS_4...:
EX_4YRS:
GR_4YRS:
NUM_FY_4YRS:
FIRST_GIFT_...:
NUM_CSU_4YR...:
TOT_AMT_4YR...:
DAYS_SINCE_...:
TS_4YRS:
Probability Histogram
All
Selected
probability
LIFETIME_ENDOWMENT_IND_LE2018
Summary Statistics (Selected Group)
Count
82
9% OF TOTAL
Count: 1
24
Count: 0
58
Probability
29%
+20.4%
Odds
0.41
Odds Ratio 5.7

Takeaways From This Study

Business school grads are the most valuable prospects for lifetime endowments

- any other grads (College of Fine Arts, School of Engineering, School of Social Work, etc.) not so much
- the probability is not overly high for most (29%)
- but still much higher than for the overall unmanaged population (8.6%)

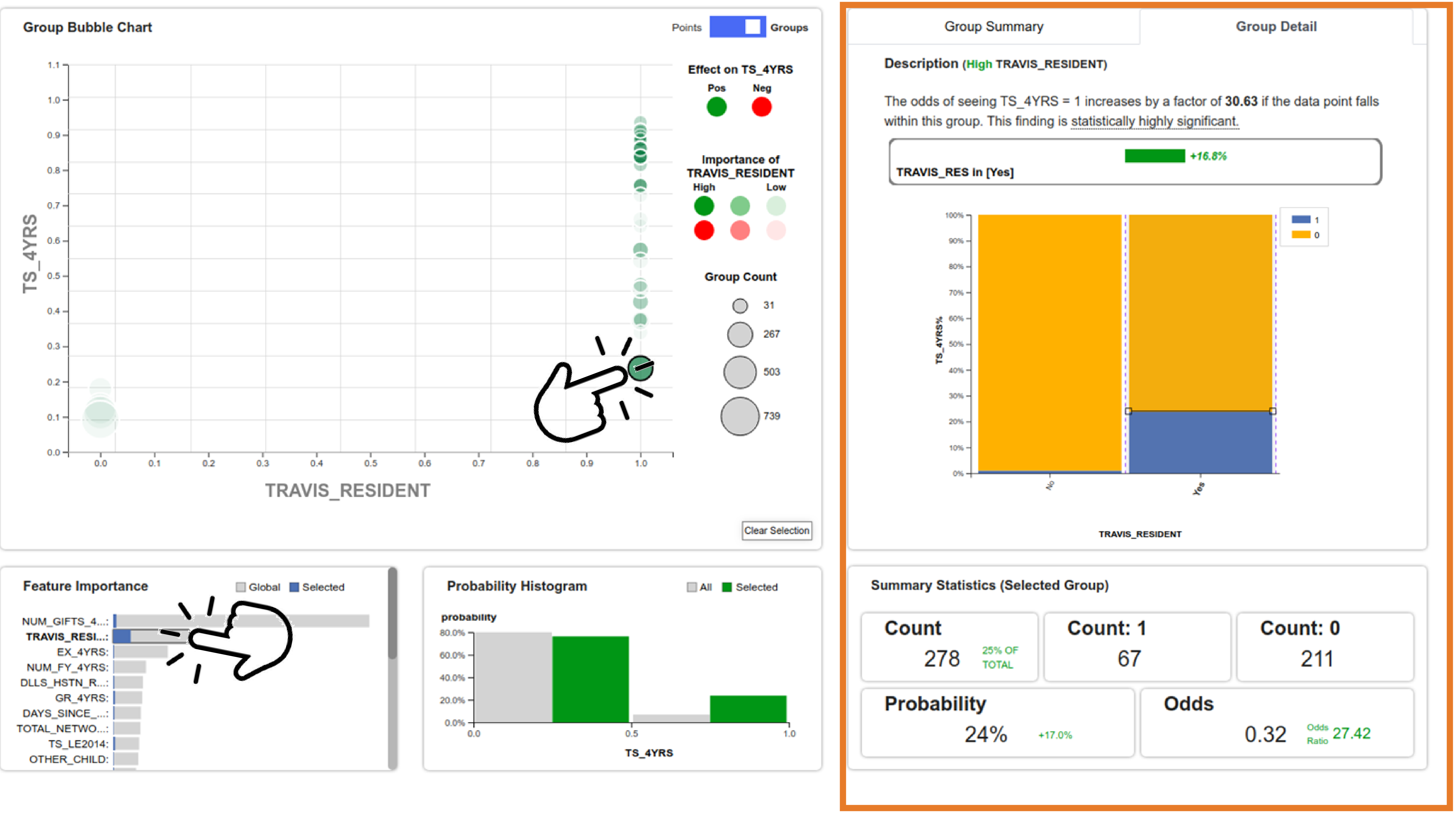
Finally: How About the Radio Station

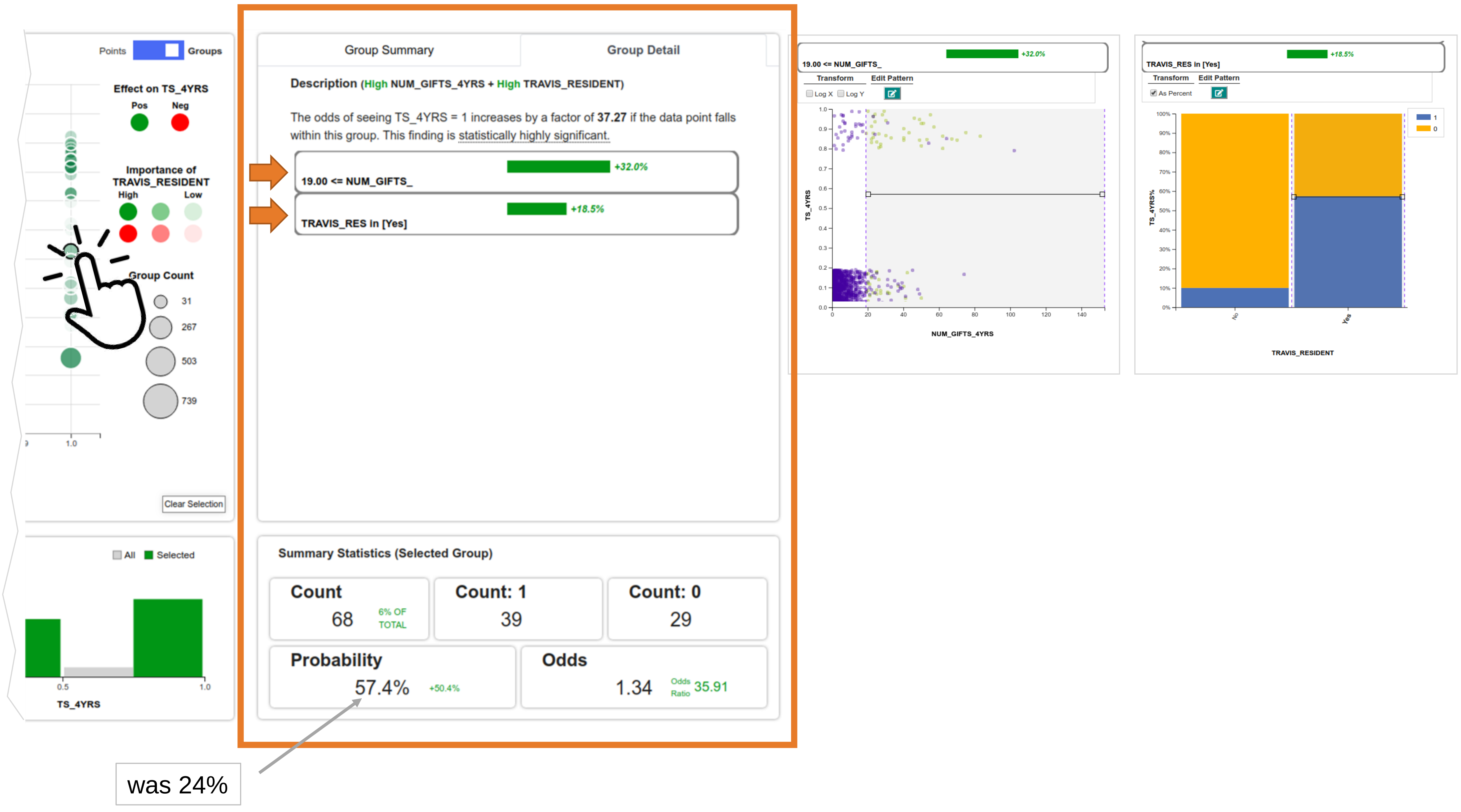
The campus radio station is the pride of many universities

- they depend on donations big time
- where do these funds they come from?
- how to solicit? who?
- knowing it may even help inform (some of the) programming

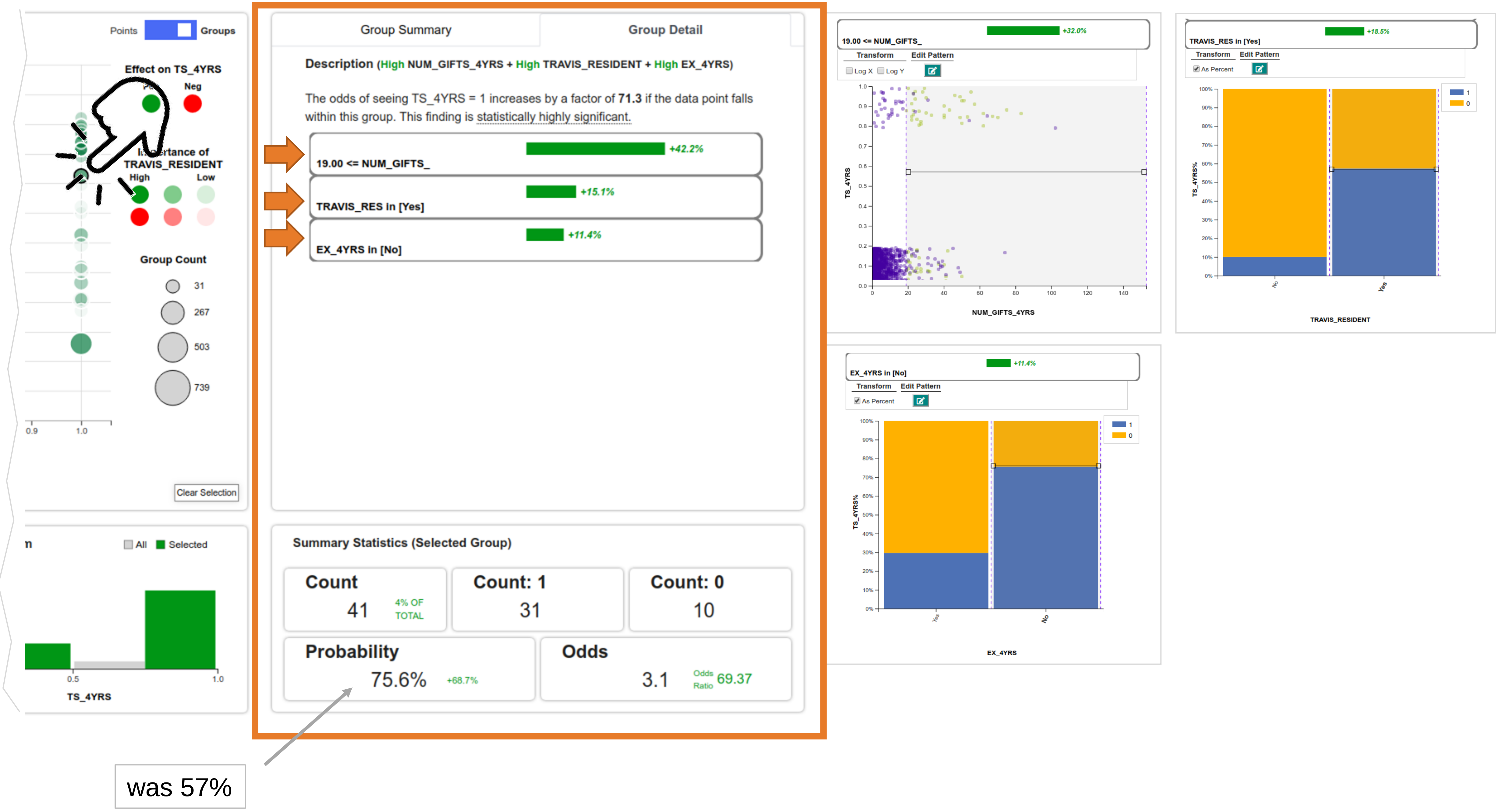
- captured by the indicator feature TS_4YRS, set to 1 if a person has donated to it within the past 4 years


Group Bubble Chart
Points
Groups
Effect on TS_4YRS
Pos
Neg
Importance of NUM_GIFTS_4YRS
High
Low
Group Count
31
267
503
739
TS_4YRS
NUM_GIFTS_4YRS
Group Summary
Group Detail
Select a group in the bubble chart on the left.
Feature Importance
Global
Selected
NUM_GIFTS_4...:
TRAVIS_RESI...:
EX_4YRS:
NUM_FY_4YRS:
DLLS_HSTN_R...:
GR_4YRS:
DAYS_SINCE_...:
TOTAL_NETWO...:
TS_LE2014:
OTHER_CHILD:
Probability Histogram
probability
TS_4YRS
Summary Statistics (All)
Count
1K
Count: 1
76
Count: 0
1K
Probability
7%
Odds
0.1


Group Bubble Chart
Points
Groups
Effect on TS_4YRS
Pos
Neg
Importance of TRAVIS_RESIDENT
High
Low
Group Count
31
267
503
739
TS_4YRS
TRAVIS_RESIDENT
Clear Selection
Group Summary
Group Detail
Description (High TRAVIS_RESIDENT)
The odds of seeing TS_4YRS = 1 increases by a factor of 30.63 if the data point falls within this group. This finding is statistically highly significant.
TRAVIS_RES in [Yes]
+16.8%
TS_4YRS%
1
0
No
Yes
TRAVIS_RESIDENT
Feature Importance
Global
Selected
NUM_GIFTS_4...:
TRAVIS_RESI...:
EX_4YRS:
NUM_FY_4YRS:
DLLS_HSTN_R...:
GR_4YRS:
DAYS_SINCE_...:
TOTAL_NETWO...:
TS_LE2014:
OTHER_CHILD:
Probability Histogram
All
Selected
probability
TS_4YRS
Summary Statistics (Selected Group)
Count
278
25% OF TOTAL
Count: 1
67
Count: 0
211
Probability
24%
+17.0%
Odds
0.32
Odds Ratio
27.42



Points
Groups
Effect on TS_4YRS
Pos
Neg
Importance of TRAVIS_RESIDENT
High
Low
Group Count
31
267
503
739
1.0
Clear Selection
All
Selected
0.5
1.0
TS_4YRS
Group Summary
Group Detail
Description (High NUM_GIFTS_4YRS + High TRAVIS_RESIDENT)
The odds of seeing TS_4YRS = 1 increases by a factor of 37.27 if the data point falls within this group. This finding is statistically highly significant.
+32.0%
19.00 <= NUM_GIFTS_
+18.5%
TRAVIS_RES in [Yes]
Summary Statistics (Selected Group)
Count
68
6% OF TOTAL
Count: 1
39
Count: 0
29
Probability
57.4%
+50.4%
Odds
1.34
Odds Ratio
35.91
was 24%
19.00 <= NUM_GIFTS_
+32.0%
Transform
Edit Pattern
Log X
Log Y
TS_4YRS
NUM_GIFTS_4YRS
TRAVIS_RES in [Yes]
+18.5%
Transform
Edit Pattern
As Percent
TS_4YRS%
TRAVIS_RESIDENT
No
Yes
1
0


Points
Groups
Effect on TS_4YRS
Neg
Importance of TRAVIS_RESIDENT
High
Low
Group Count
31
267
503
739
0.9
1.0
Clear Selection
All
Selected
0.5
1.0
TS_4YRS
Group Summary
Group Detail
Description (High NUM_GIFTS_4YRS + High TRAVIS_RESIDENT + High EX_4YRS)
The odds of seeing TS_4YRS = 1 increases by a factor of 71.3 if the data point falls within this group. This finding is statistically highly significant.
19.00 <= NUM_GIFTS_
+42.2%
TRAVIS_RES in [Yes]
+15.1%
EX_4YRS in [No]
+11.4%
Summary Statistics (Selected Group)
Count
41
4% OF TOTAL
Count: 1
31
Count: 0
10
Probability
75.6%
+68.7%
Odds
3.1
Odds Ratio
69.37
was 57%
19.00 <= NUM_GIFTS_
+32.0%
Transform
Edit Pattern
Log X
Log Y
TS_4YRS
NUM_GIFTS_4YRS
TRAVIS_RES in [Yes]
+18.5%
As Percent
TS_4YRS%
TRAVIS_RESIDENT
No
Yes
EX_4YRS in [No]
+11.4%
EX_4YRS

Takeaways From This Study

Exposed a good strategy on how to use our system

- derive nuanced multi-level fundraising strategies by refining the characteristics of a certain family of groups
- first launch a more general campaign for a broader group
- then address smaller but more specific groups with more targeted campaigns with higher probabilities of success

Now to a Live Demo

Pattern Browser 4 XAI

Pattern Browser allows analysts to

- explore a dataset from multiple perspectives
- quickly follow their instincts via simple mouse-click interactions
- within a single session from one dashboard

Fully embraces the paradigm of explainable machine learning / AI

- shows the results not just as a single number but with visual explanations on how the number was derived and how it relates to the overall data
- explanations are succinct and focus on the important features only

Contrast: Subgroup Analysis

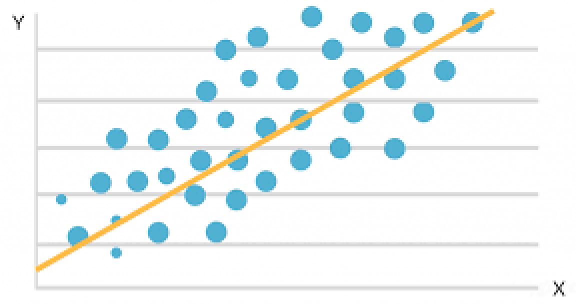
Decomposing large populations into sets of homogenous subgroups is well known in fields like medicine

- seeks to identify a specific patient characteristics that benefit a desired outcome
- typically done using prior knowledge, pre-specification, or stepwise procedures
- not scalable in the number of features

In contrast, we learn these subgroups by automated *discovery*

- robustly via statistical pattern mining
- this can scale to 1,000s and more features/variables

Contrast: Regression Models

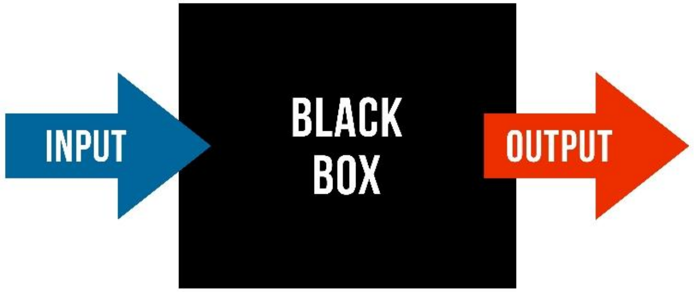


Regression models are a standard approach in data analysis

- intractable to explicitly model all possible interactions between variables
- even with pairwise interactions we would have over 10,000 possible interactions in the study we presented here
- also are restricted to modeling linear relationships -- nonlinear relationships would require additional transformations

In contrast, our system can identify interactions and capture nonlinear relationships automatically

Contrast: Black Box Models



Random forests, neural networks, etc. have become ubiquitous

- lots of libraries are available
- explainable AI tools, such as SHAP, LIME, can help explain a black box model's decision
- no guarantees if the decision is based on a true cause-effect relationship or a spurious correlation

In contrast, our system puts the human in the sense-making loop

- analyst can identify the most likely explanation and choose an action
- e.g. select the most likely explanation why a group is more likely to donate

Acknowledgments

Stony Brook University

Akai Kaeru

The system used for this analysis

- available as a software package called Pattern Browser
- developed by Akai Kaeru LLC http://akaikaeru.com
- development was funded by NSF SBIR grant 192694 (Phase I and II)

Thanks also to

- John Gough from U Texas, Austin for providing the data and his insight interpreting them