Mixed Integer Programming (MIP) for Daily Fantasy Sports, Statistics and Marketing

Juan Pablo Vielma

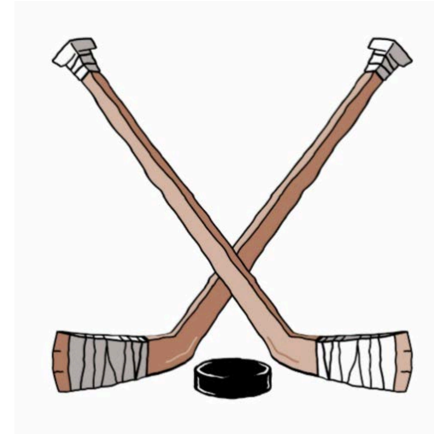
Massachusetts Institute of Technology

AM/ES 121, SEAS, Harvard.

Boston, MA, November, 2016.

MIP & Daily Fantasy Sports

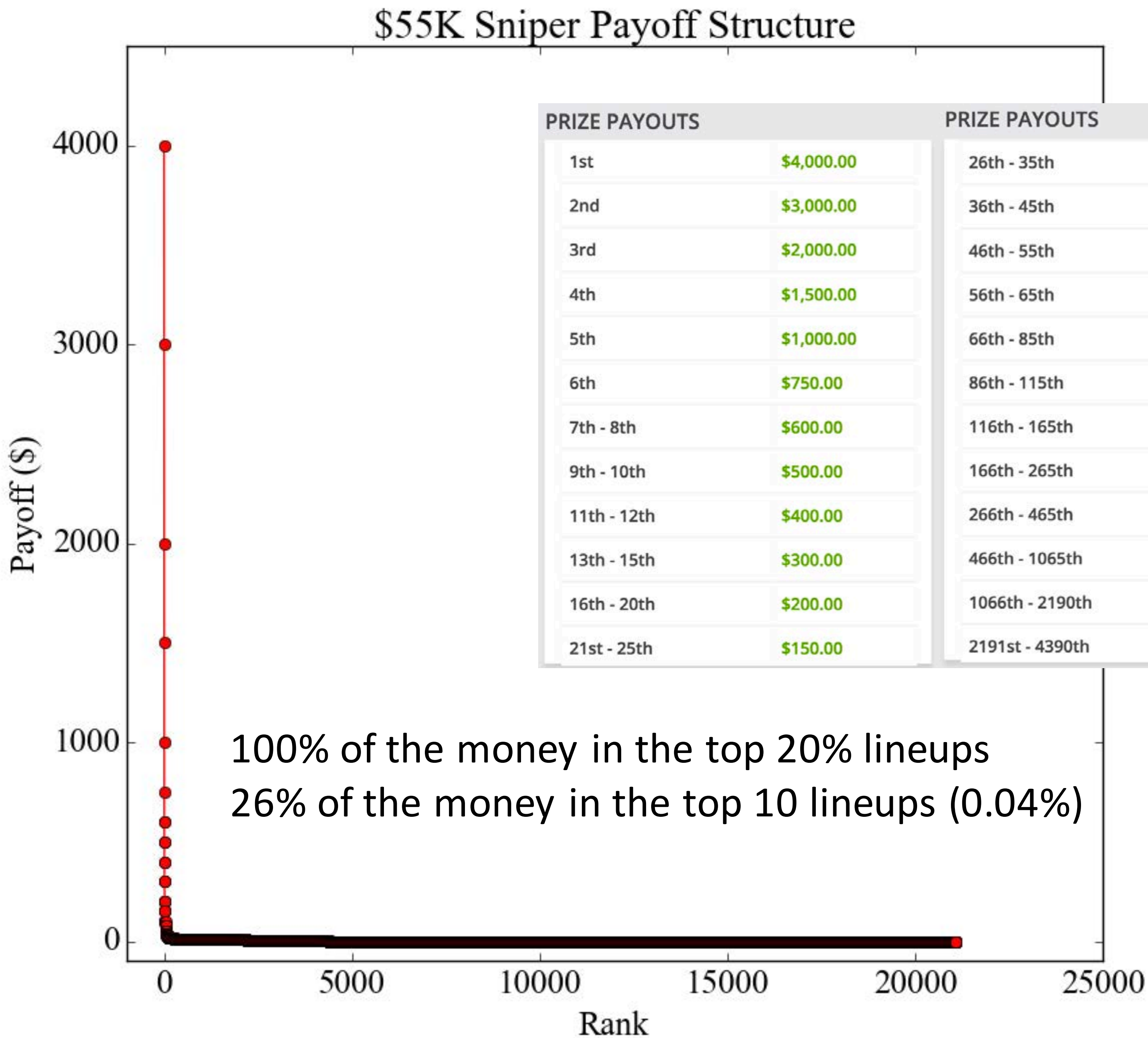
Example Entry



LINEUP

Avg. Rem. / Player: $0
Rem. Salary: $0

POS	PLAYER	OPP	FPPG	SALARY
C	Jussi Jokinen	Fla@Anh	3.1	$5,300
C	Brandon Sutter	Pit@Van	3.0	$4,400
W	Nikolaj Ehlers	Wpg@Tor	3.9	$4,800
W	Daniel Sedin	Pit@Van	3.8	$6,400
W	Radim Vrbata	Pit@Van	3.4	$5,800
D	Brian Campbell	Fla@Anh	2.6	$4,100
D	Morgan Rielly	Wpg@Tor	3.5	$4,200
G	Corey Crawford P	StL@Chi	6.3	$7,800
UTIL	Blake Wheeler	Wpg@Tor	4.8	$7,200



PRIZE PAYOUTS	
1st	$4,000.00
2nd	$3,000.00
3rd	$2,000.00
4th	$1,500.00
5th	$1,000.00
6th	$750.00
7th - 8th	$600.00
9th - 10th	$500.00
11th - 12th	$400.00
13th - 15th	$300.00
16th - 20th	$200.00
21st - 25th	$150.00
26th - 35th	$100.00
36th - 45th	$75.00
46th - 55th	$50.00
56th - 65th	$40.00
66th - 85th	$30.00
86th - 115th	$25.00
116th - 165th	$20.00
166th - 265th	$15.00
266th - 465th	$12.00
466th - 1065th	$10.00
1066th - 2190th	$8.00
2191st - 4390th	$6.00

100% of the money in the top 20% lineups

26% of the money in the top 10 lineups (0.04%)

Building a Lineup

Integer Programming Formulation

- We will make a bunch of lineups consisting of 9 players each
- Use an integer programming approach to find these lineups

Decision variables

$$x_{pl} = \begin{cases} 1, & \text{if player } p \text{ in lineup } l \\ 0, & \text{otherwise} \end{cases}$$

Basic Feasibility

- 9 different players
- Salary less than \$50,000

Basic constraints

$$\sum_{p=1}^{N} c_p x_{pl} \leq \$50,000, \quad \text{(budget constraint)}$$

$$\sum_{p=1}^{N} x_{pl} = 9, \quad \text{(lineup size constraint)}$$

$$x_{pl} \in \{0,1\}, \quad 1 \leq p \leq N.$$

Position Feasibility

- Between 2 and 3 centers
- Between 3 and 4 wingers
- Between 2 and 3 defensemen
- 1 goalie

Position constraints

$$2 \leq \sum_{p \in C} x_{pl} \leq 3, , \quad \text{(center constraint)}$$

$$3 \leq \sum_{u \in W} x_{pl} \leq 4, \quad \text{(winger constraint)}$$

$$2 \leq \sum_{u \in D} x_{pl} \leq 3, \quad \text{(defensemen constraint)}$$

$$\sum_{u \in G} x_{pl} = 1 \quad \text{(goalie constraint)}$$

Team Feasibility

- At least 3 different NHL teams

Team constraints

$$t_i \leq \sum_{p \in T_i} x_{pl}, \quad \forall\, i \in \{1, \ldots, N_T\}$$

$$\sum_{i=1}^{N_T} t_i \geq 3,$$

$$t_i \in \{0,\ 1\}, \quad \forall\, i \in \{1, \ldots, N_T\}.$$

Maximize Points

- Forecasted points for player p: f_p

Score type	Points
Goal	3
Assist	2
Shot on Goal	0.5
Blocked Shot	0.5
Short Handed Point Bonus (Goal/Assist)	1
Shootout Goal	0.2
Hat Trick Bonus	1.5
Win (goalie only)	3
Save (goalie only)	0.2
Goal allowed (goalie only)	-1
Shutout Bonus (goalie only)	2

Table 1 Points system for NHL contests in DraftKings.

Points Objective Function

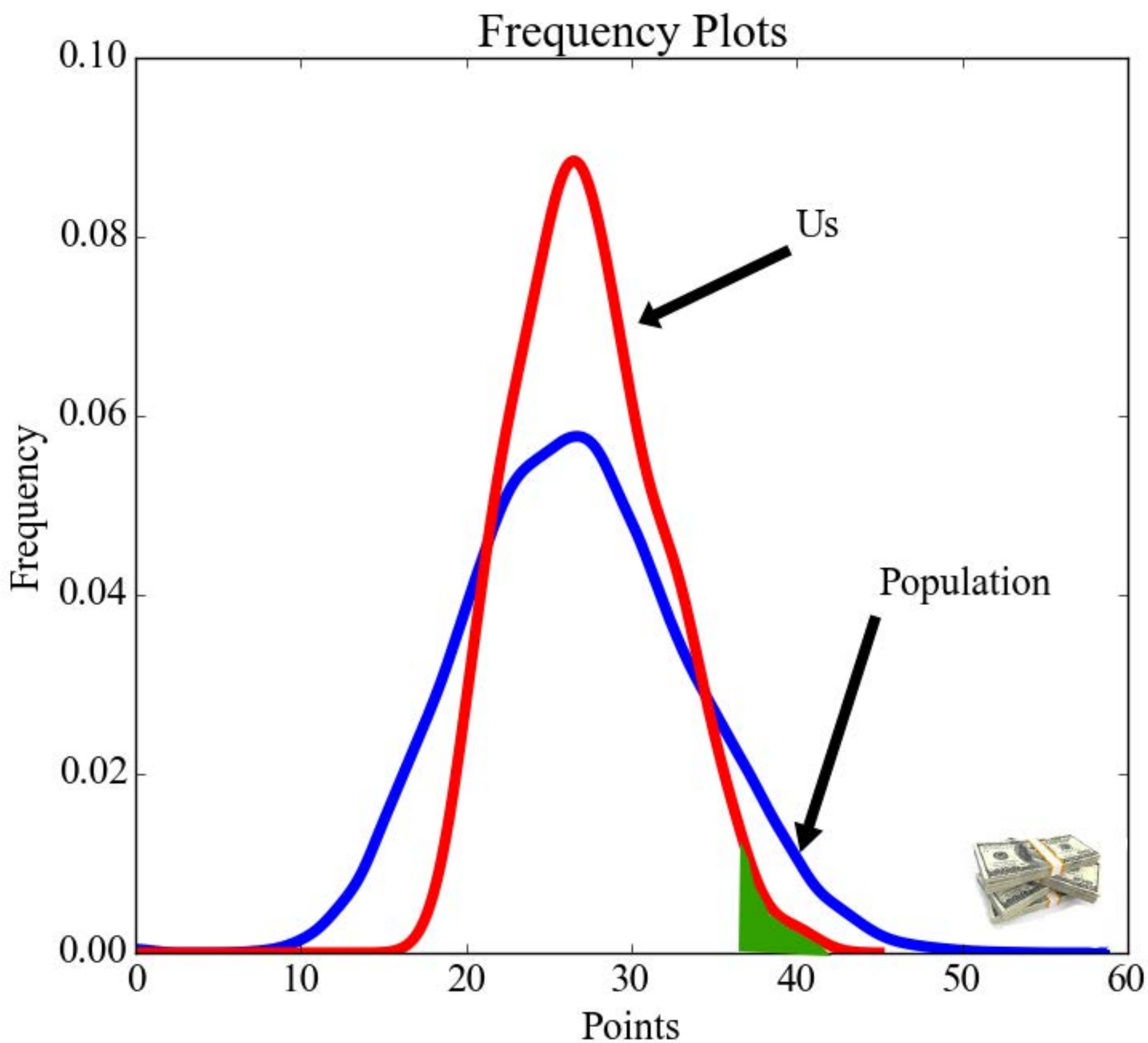
$$\sum_{p=1}^{N} f_p x_{pl}$$

Lineup

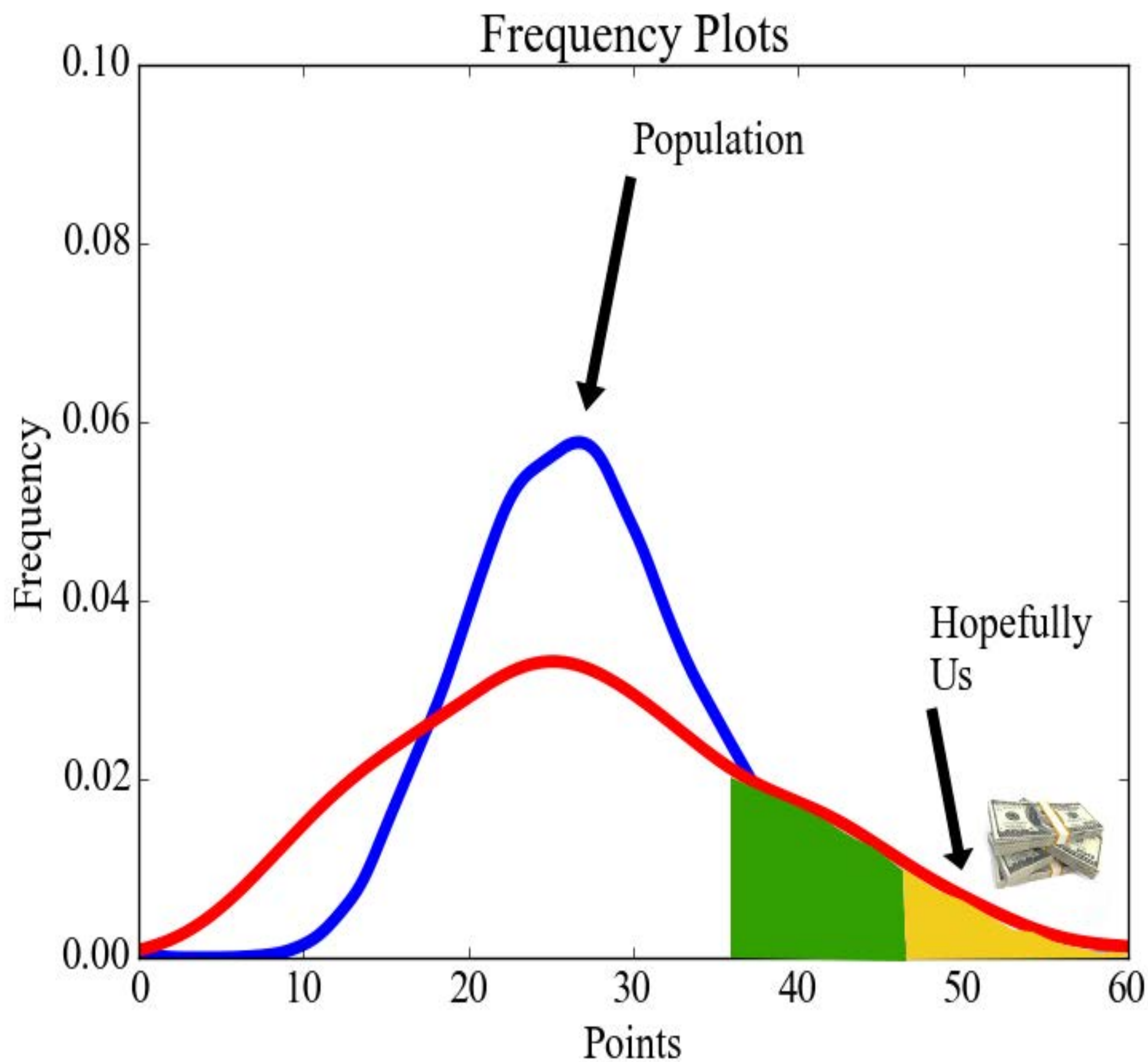
Projections:	5.4	2.5	3.4	3.0	3.2	4.2	3.5	3.4	5.7
	$9500	$2700	$4600	$3800	$4600	$6400	$5200	$5100	$8000
	W	UTIL	D	D	C	C	W	W	G

23 points on average

Need > 38 points for a chance to win



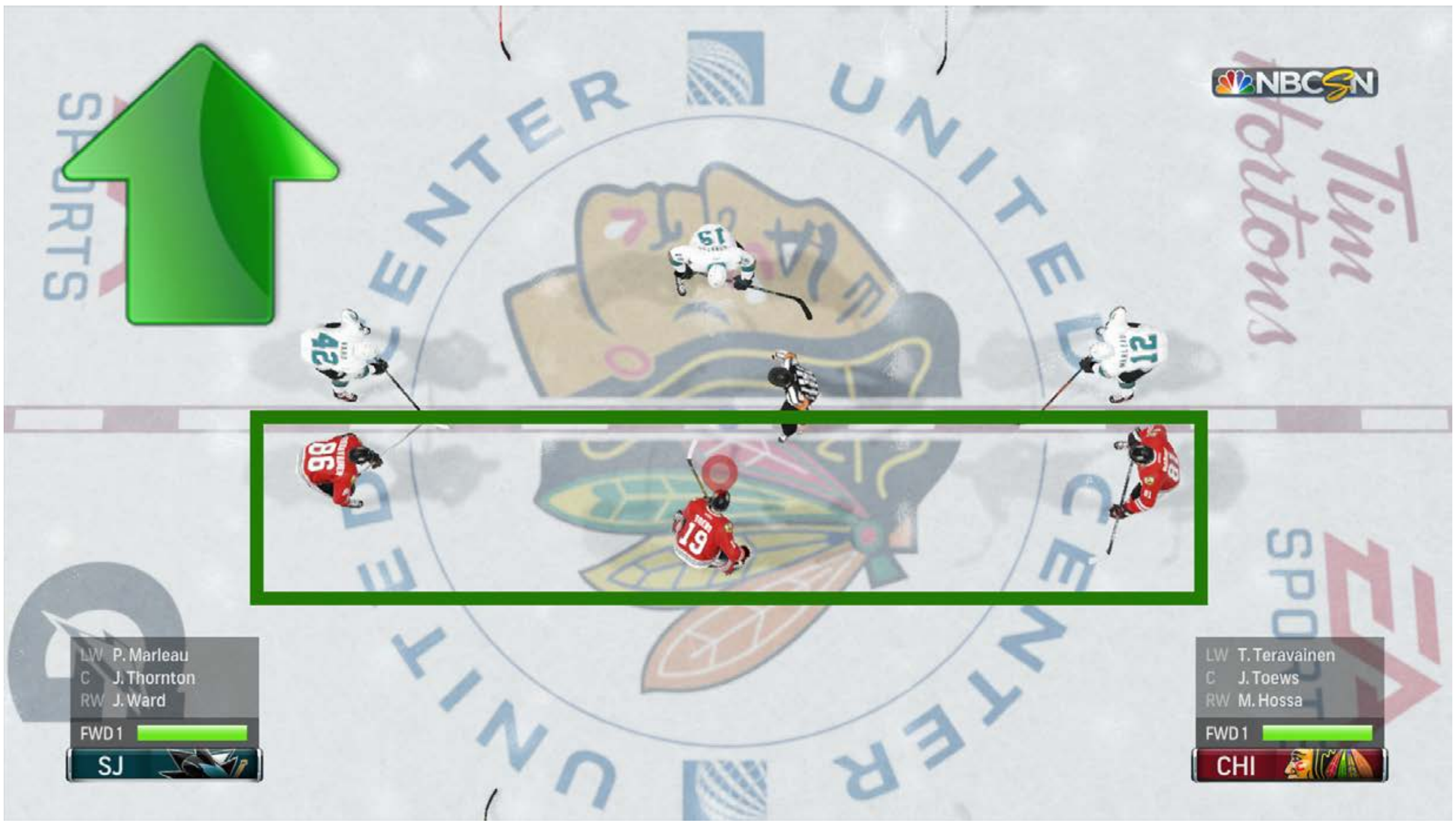
Increase variance to have a chance



Structural Correlations - Teams

Structural Correlations - Lines

- Goal = 3 pt, assist = 2 pt



Structural Correlations – Lines = Stacking

- At least 1 complete line (3 players per line)
- At least 2 partial lines (at least 2 players per line)

1 complete line constraint

$$3v_i \leq \sum_{p \in L_i} x_{pl}, \qquad \forall i \in \{1, \ \ldots \ , N_L\}$$

$$\sum_{i=1}^{N_L} v_i \geq 1$$

$$v_i \in \{0,1\}, \qquad \forall i \in \{1, \ \ldots \ , N_L\}.$$

2 partial lines constraint

$$2w_i \leq \sum_{p \in L_i} x_{pl}, \qquad \forall i \in \{1, \ \ldots \ , N_L\}$$

$$\sum_{i=1}^{N_L} w_i \geq 2$$

$$w_i \in \{0,1\}, \qquad \forall i \in \{1, \ \ldots \ , N_L\}.$$

Structural Correlations – Goalie Against Opposing Players

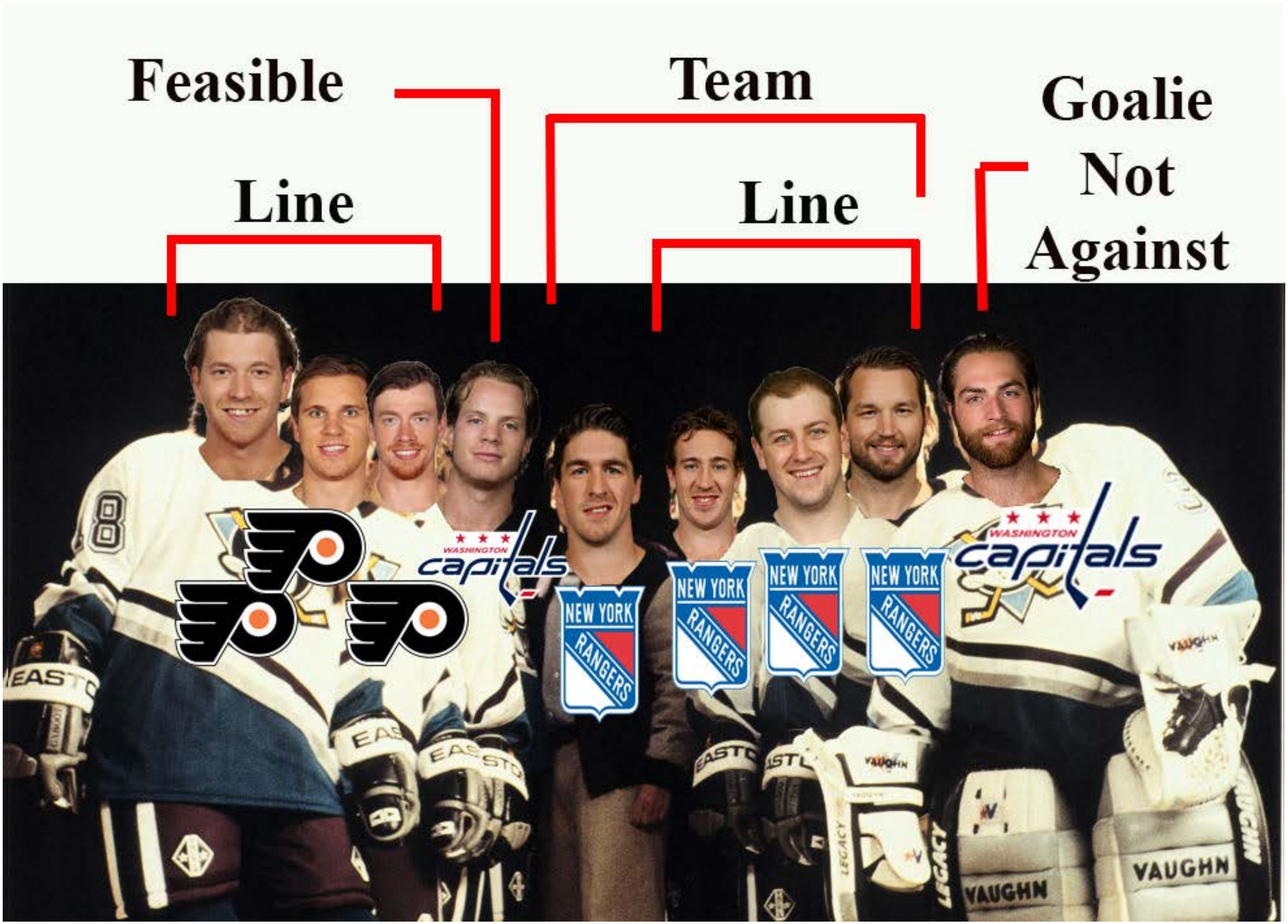
Structural Correlations – Goalie Against Skaters

- No skater against goalie

No skater against goalie constraint

$$6x_{pl} + \sum_{q \in Opponents_p} x_{ql} \leq 6, \quad \forall p \in G$$

Good, but not great chance



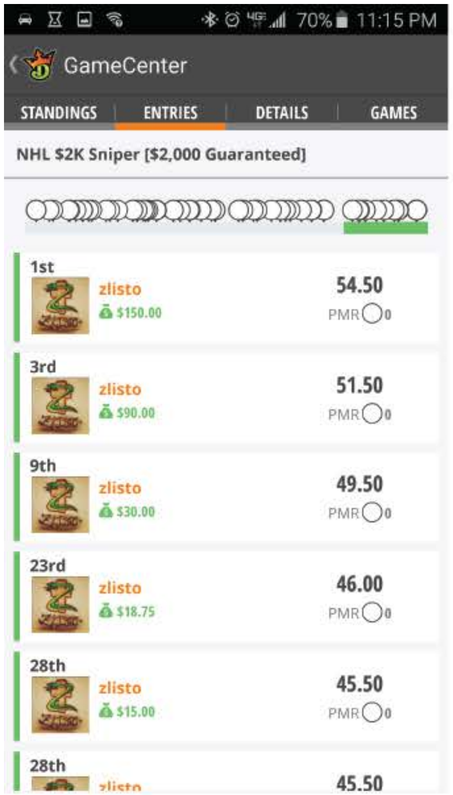
Play many diverse Lineups

- Make sure lineup l has no more than γ players in common with lineups 1 to l-1

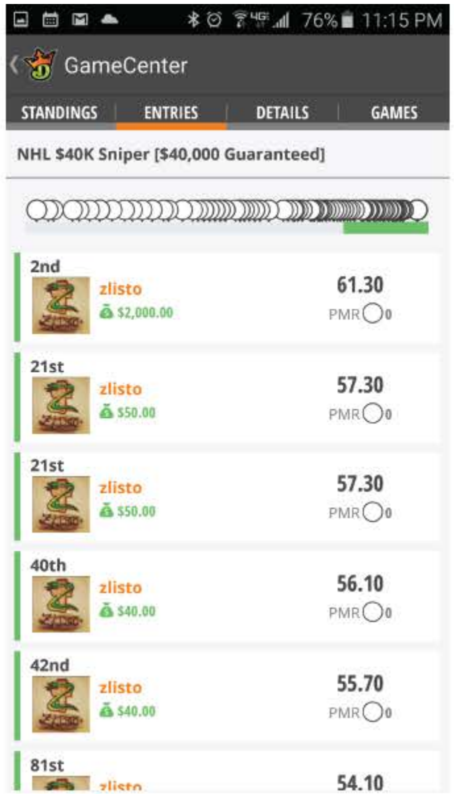
Diversity constraint

$$\sum_{p=1}^{N} x^*_{pk} x_{pl} \leq \gamma, k = 1, \ldots, l-1$$

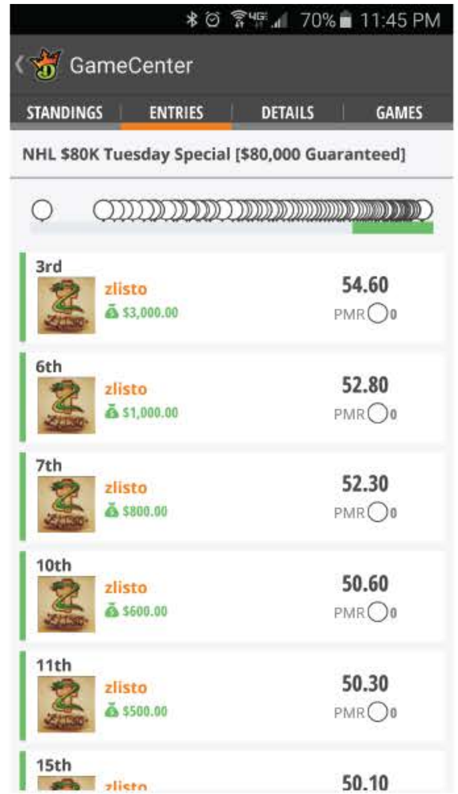
Were we able to do it?



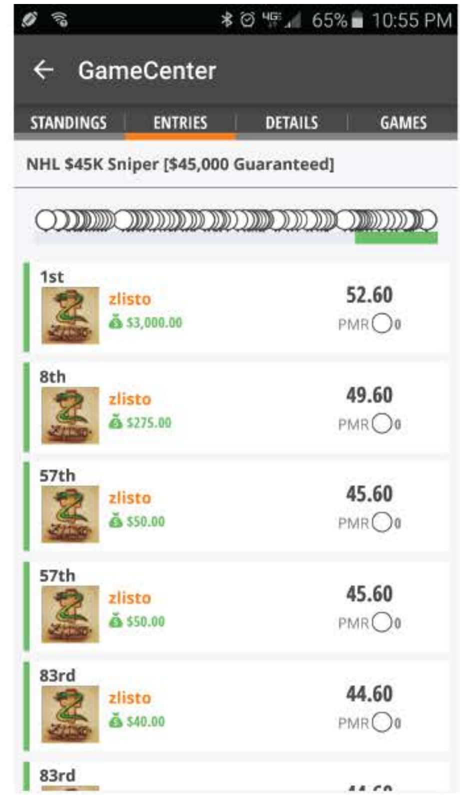
November 15, 2015



November 16, 2015



November 17, 2015



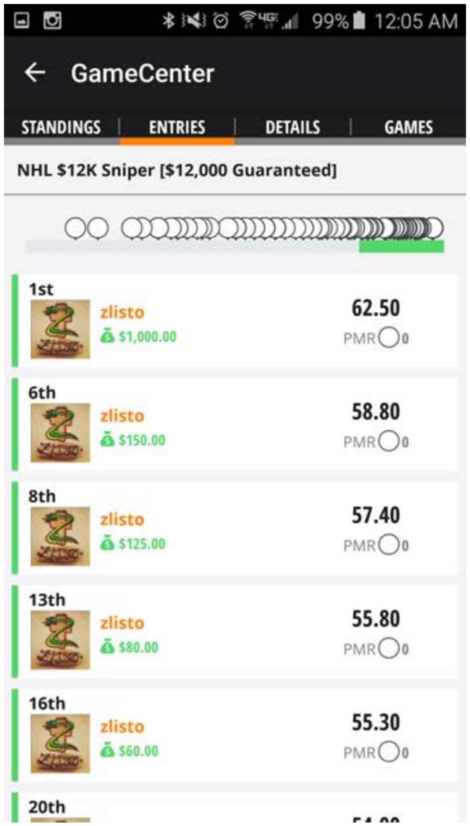
November 23, 2015

200 lineups

Policy Change

200 lineups -> 100 lineups

Were we able to continue it?



December 12, 2015

100 lineups

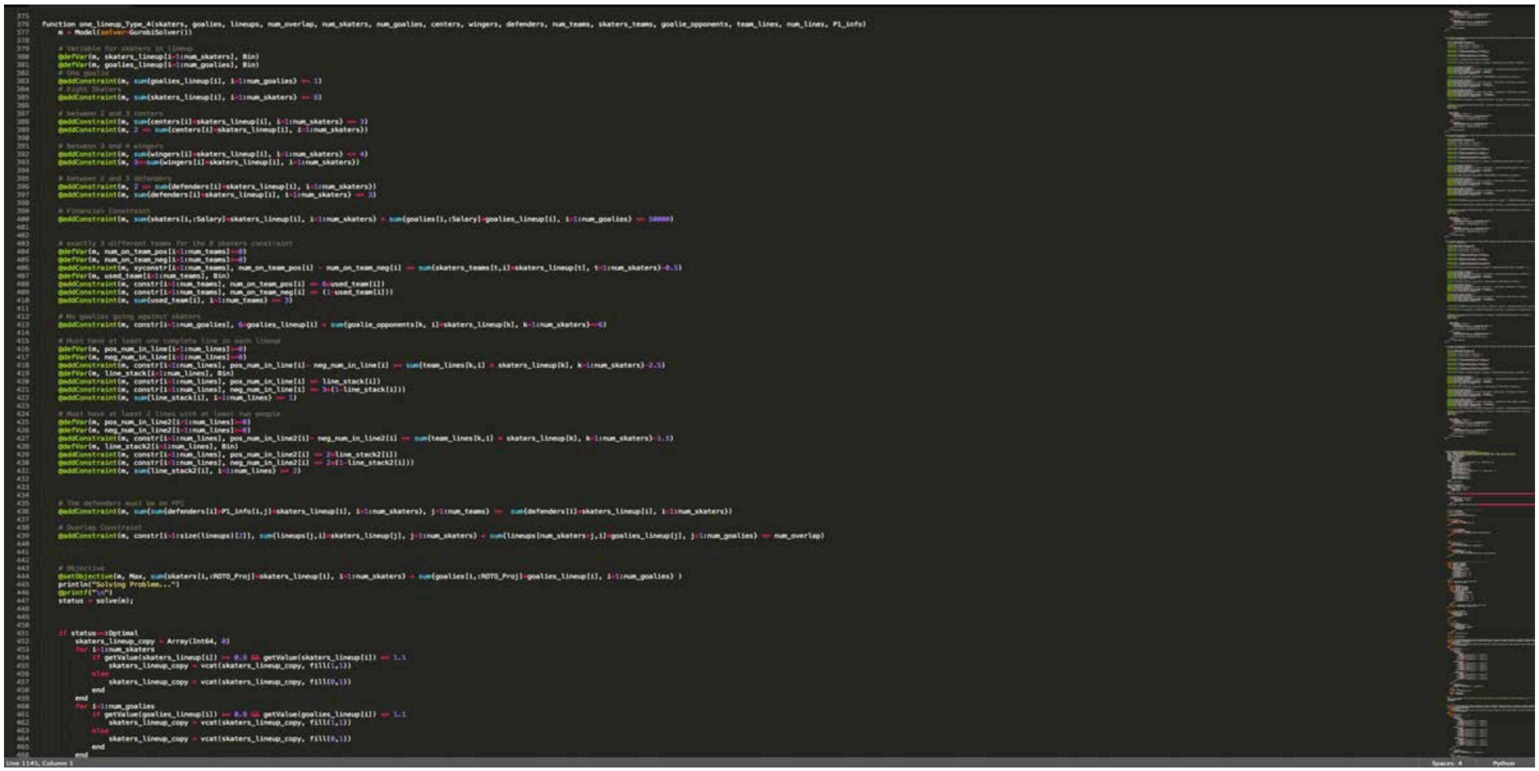
The Greater Boston FOOD BANK®

> $15K

How can you do it?

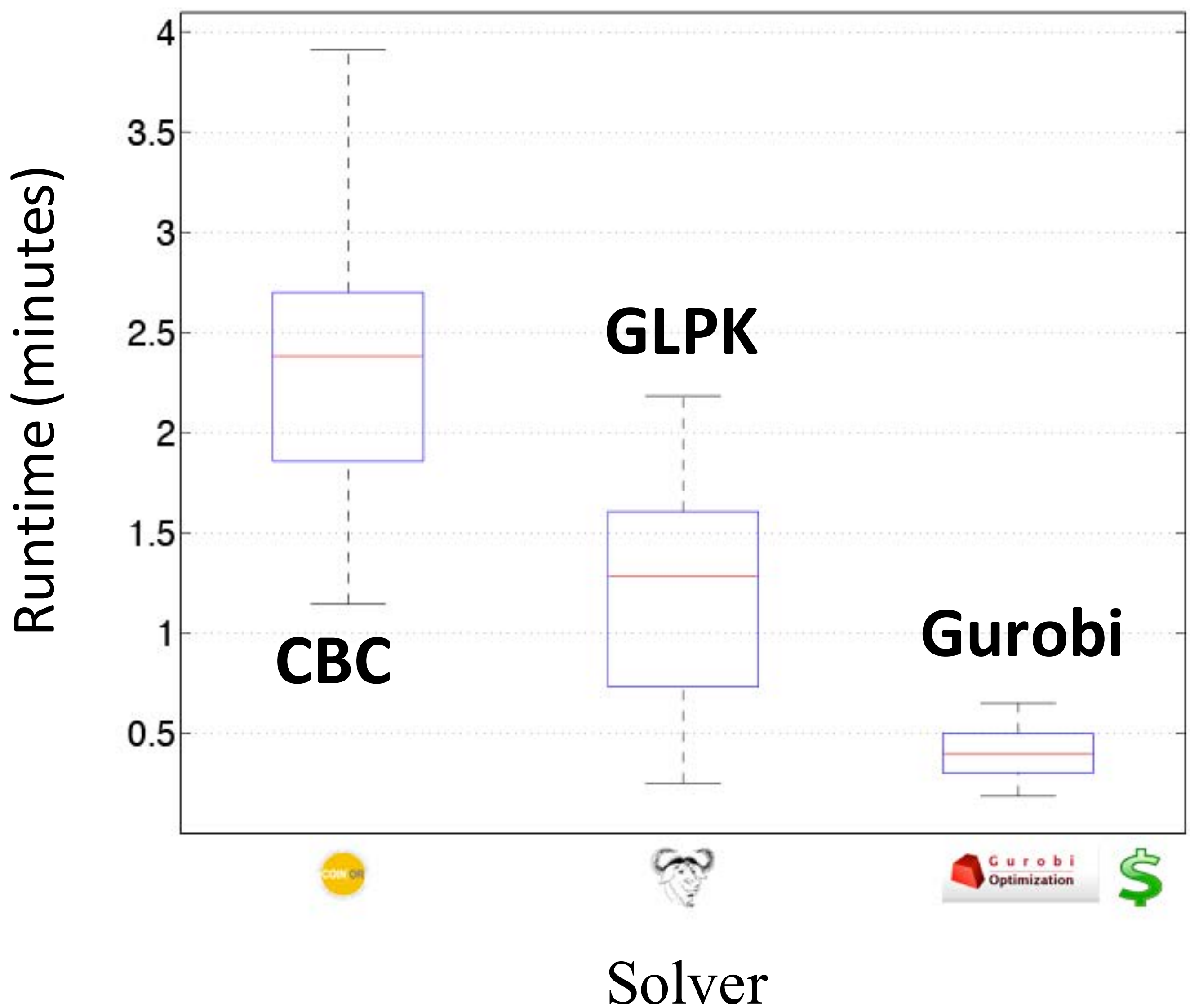
Download Code from Github:

https://github.com/dscotthunter/Fantasy-Hockey-IP-Code



http://arxiv.org/pdf/1604.01455v1.pdf

Performance Time
< 30 Minutes



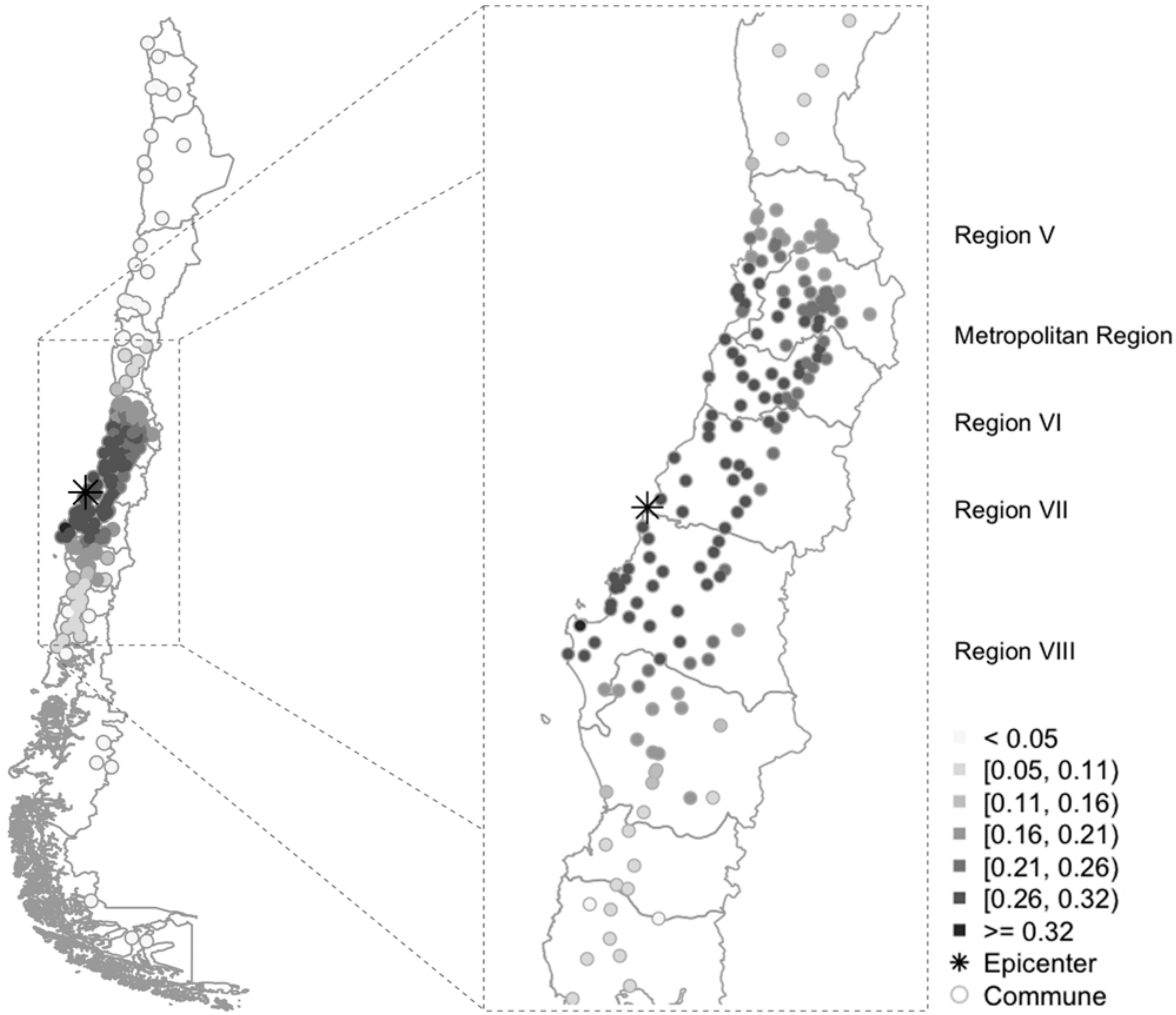
MIP and Statistics: Inference for the Chilean Earthquake

The 2010 Chilean Earthquake

6th Strongest in Recorded History (8.8)

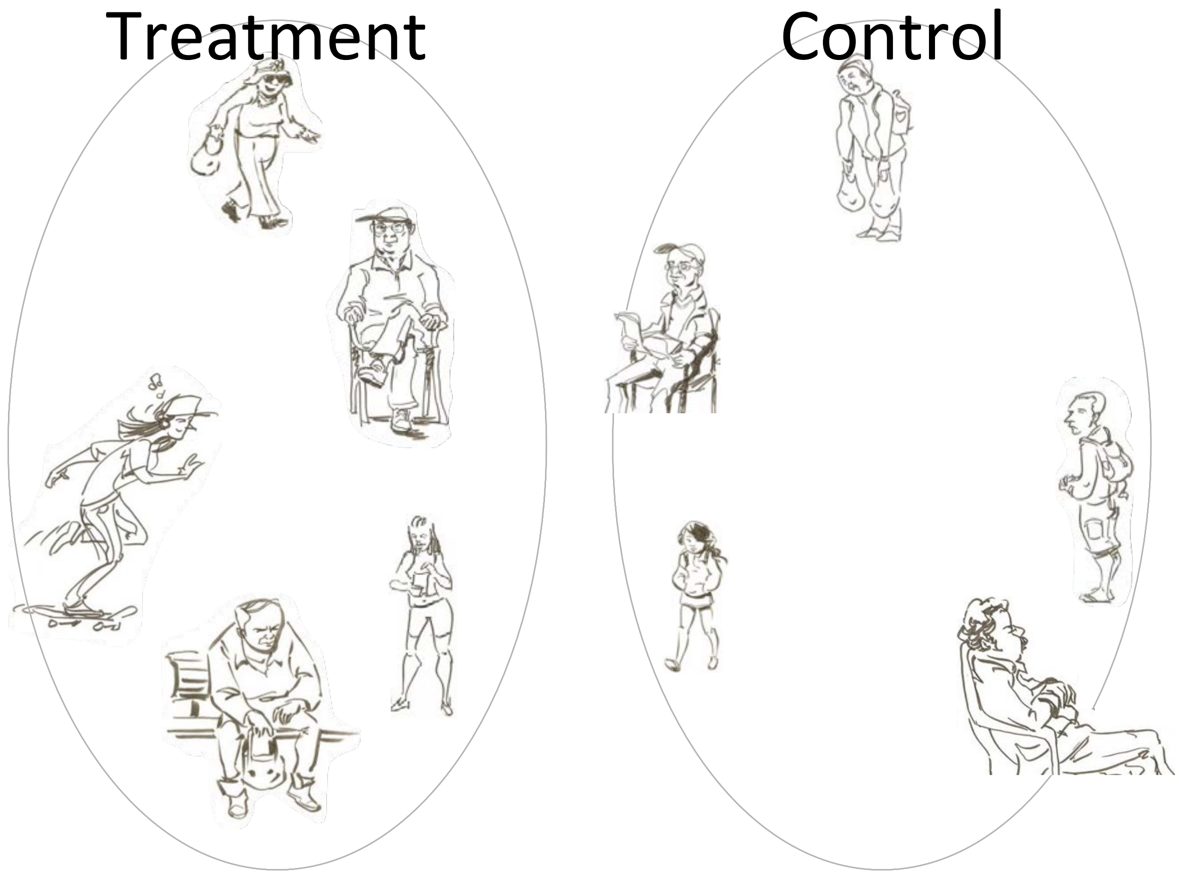
Impact on Educational Achievement? PSU = SAT

Earthquake Intensity + Great Demographic Info



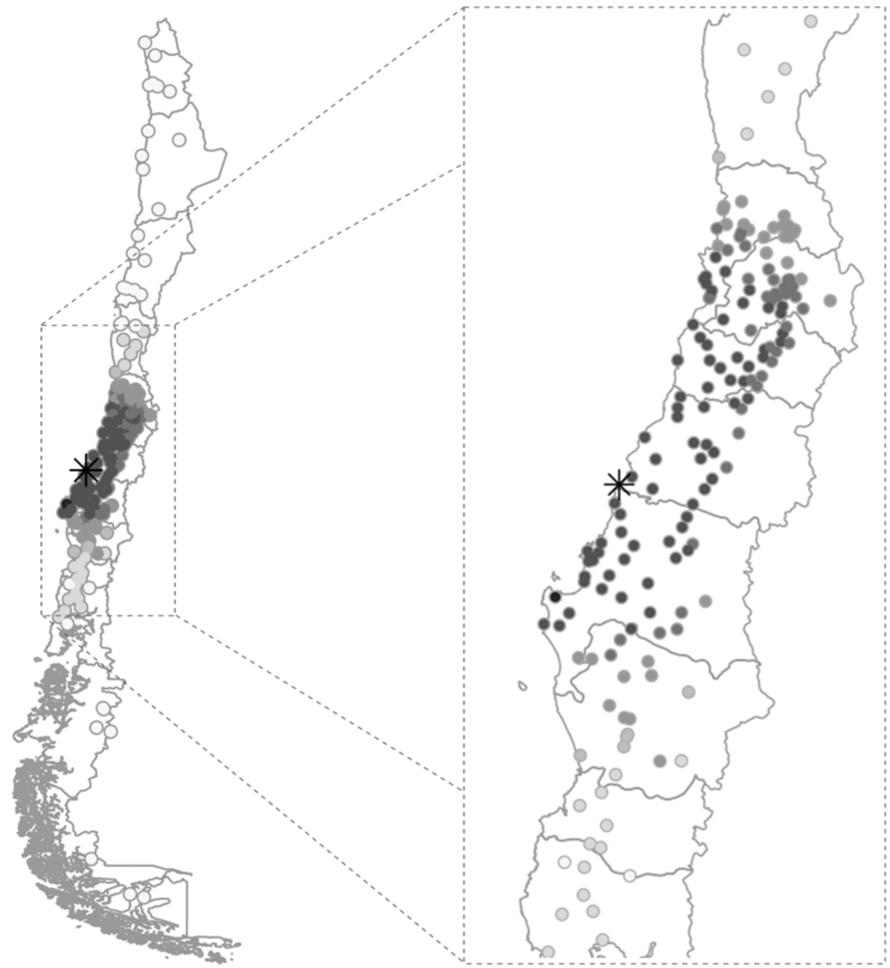
Randomized experiment

- Treatment / control have similar characteristics (covariates).



Covariate Balance Important for Inference

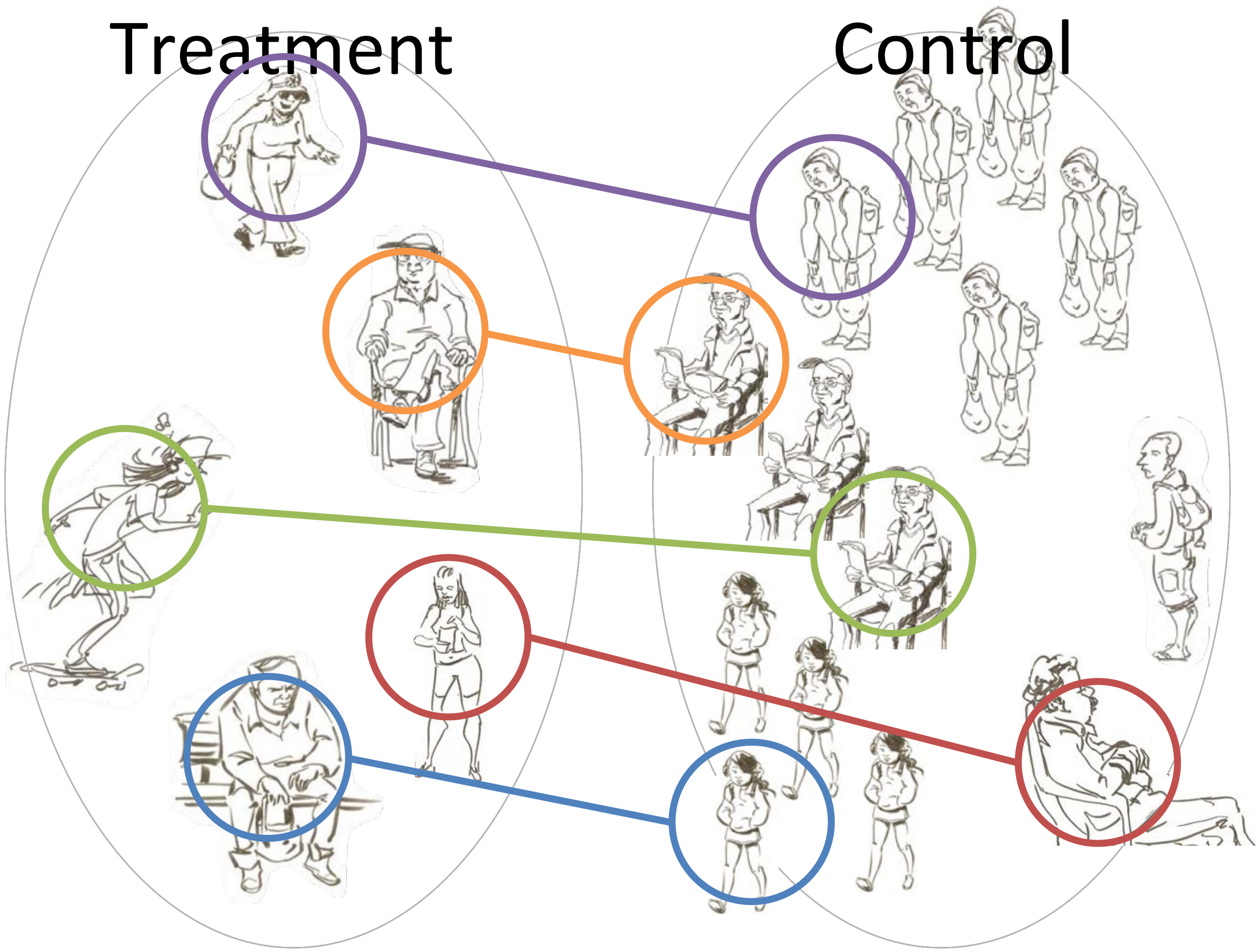
Covariate	Dose 1	Dose 2
Gender		
Male	462	462
Female	538	538
School SES		
Low	75	75
Mid-low	327	327
Medium	294	294
Mid-high	189	189
High	115	115
Mother's education		
Primary	335	335
Secondary	426	426
Technical	114	114
College	114	114
Missing	11	11
⋮		



Observational Study: e.g. After Earthquake

- Treatment / control can have different characteristics.

Solution
=
Matching?



Matching

Treated Units: $\mathcal{T} = \{t_1, \ldots, t_T\}$

Control Units: $\mathcal{C} = \{c_1, \ldots, c_C\}$

Observed Covariates: $\mathcal{P} = \{p_1, \ldots, p_P\}$

Covariate Values:

$$\mathbf{x}^t = \left(x_p^t\right)_{p \in \mathcal{P}}, \quad t \in \mathcal{T}$$

$$\mathbf{x}^c = \left(x_p^c\right)_{p \in \mathcal{P}}, \quad c \in \mathcal{C}$$

Nearest Neighbor Matching

$$\underset{\boldsymbol{m}}{\text{minimize}} \quad \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \delta_{t,c} m_{t,c}$$

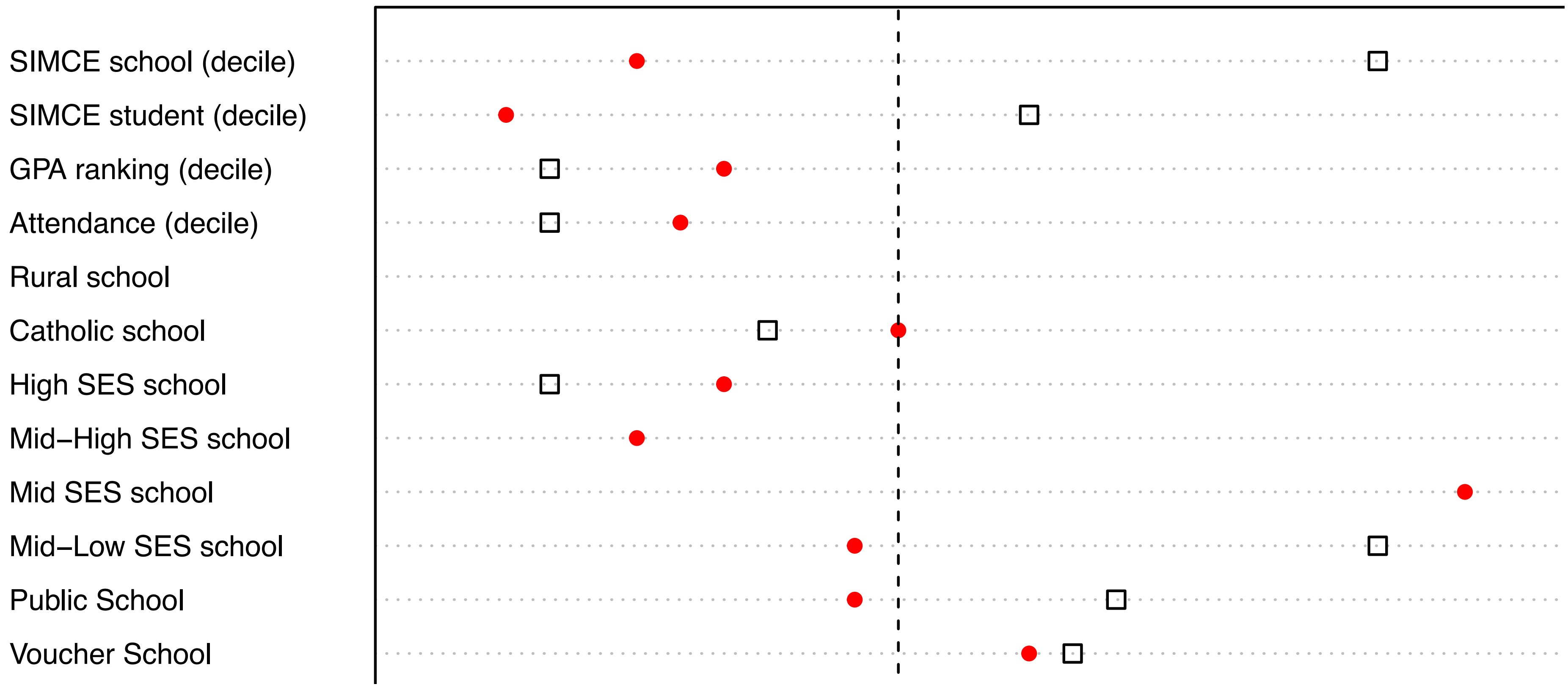
$$\text{subject to} \quad \sum_{c \in \mathcal{C}} m_{t,c} = 1, \ t \in \mathcal{T}$$

$$\sum_{t \in \mathcal{T}} m_{t,c} \leq 1, \ c \in \mathcal{C}$$

$$0 \leq m_{t,c} \leq 1 \quad \sout{m_{t,c}} \in \{0, 1\}, \ t \in \mathcal{T}, c \in \mathcal{C}$$

- e.g. $\delta_{t,c} = \left\| \mathbf{x}^t - \mathbf{x}^c \right\|_2$
- Easy to solve

Balance Before After Matching



Maximum Cardinality Matching

$$\mathcal{K}(p) = \{\mathbf{x}_p^c\}_{c \in \mathcal{P}} \cup \{\mathbf{x}_p^t\}_{t \in \mathcal{T}}$$

$$\mathcal{C}_{p,k} = \{c \in \mathcal{C} \ : \ \mathbf{x}_p^c = k\}$$

$$\mathcal{T}_{p,k} = \{t \in \mathcal{T} \ : \ \mathbf{x}_p^t = k\}$$

$$\max \quad \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} m_{t,c}$$

$$s.t.$$

$$\sum_{t \in \mathcal{T}} m_{t,c} \leq 1, \qquad \forall c \in \mathcal{C}$$

$$\sum_{c \in \mathcal{C}} m_{t,c} \leq 1, \qquad \forall t \in \mathcal{T}$$

$$\sum_{t \in \mathcal{T}_{p,k}} \sum_{c \notin \mathcal{C}_{p,k}} m_{t,c} = \sum_{t \notin \mathcal{T}_{p,k}} \sum_{c \in \mathcal{C}_{p,k}} m_{t,c} \qquad \forall p \in \mathcal{P}, k \in \mathcal{K}(p)$$

$$m_{t,c} \in \{0, 1\} \qquad \forall t \in \mathcal{T}, \quad c \in \mathcal{C}.$$

- Very hard to solve (and very hard to understand!)

Advanced Maximum Cardinality Matching

$$\max \quad \sum_{t \in \mathcal{T}} x_t$$

$$\mathcal{K}(p) = \{\mathbf{x}_p^c\}_{c \in \mathcal{P}} \cup \{\mathbf{x}_p^t\}_{t \in \mathcal{T}}$$

$$\mathcal{C}_{p,k} = \{c \in \mathcal{C} \; : \; \mathbf{x}_p^c = k\}$$

$$\mathcal{T}_{p,k} = \{t \in \mathcal{T} \; : \; \mathbf{x}_p^t = k\}$$

$s.t.$

$$\sum_{t \in \mathcal{T}} x_t = \sum_{c \in \mathcal{C}} y_c,$$

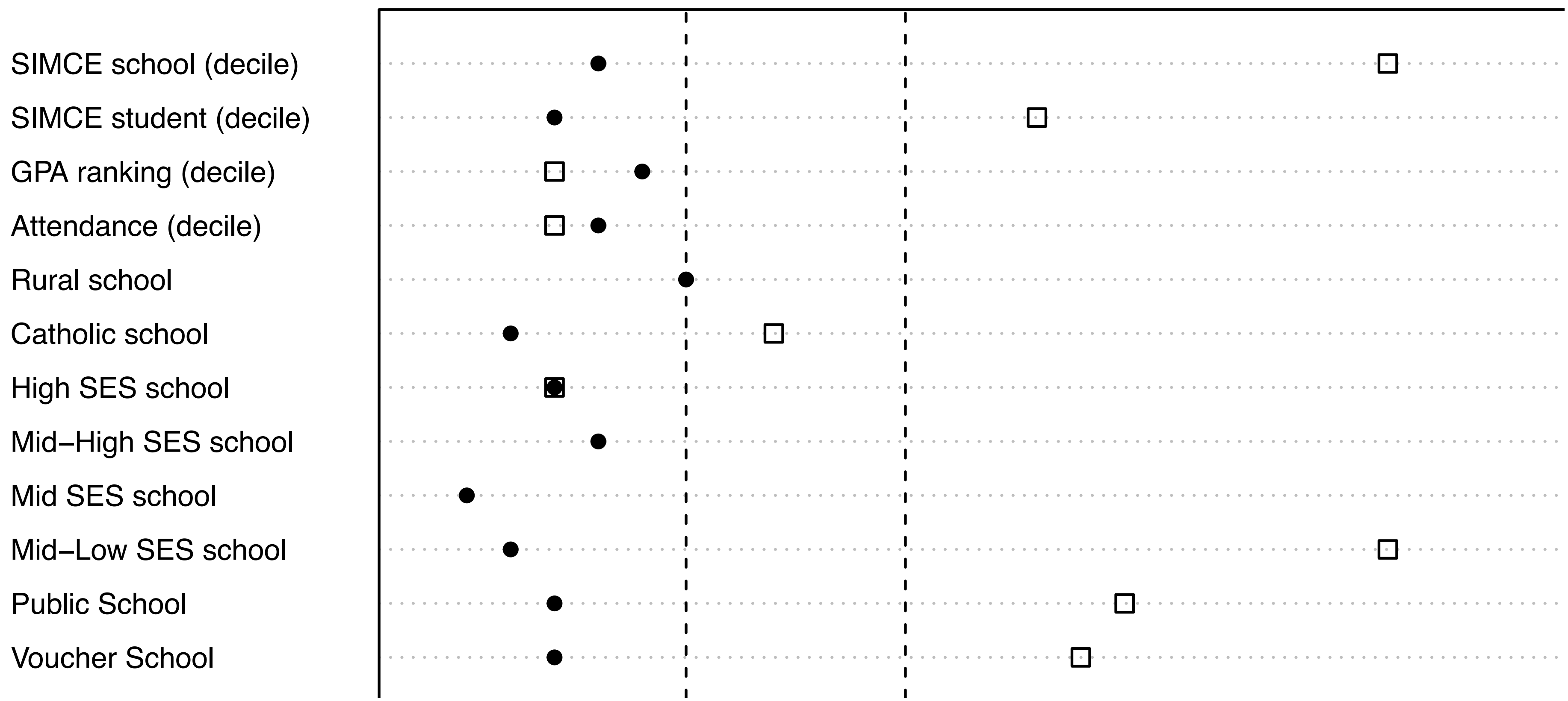
$$\sum_{t \in \mathcal{T}_{p,k}} x_t = \sum_{c \in \mathcal{C}_{p,k}} y_c, \quad \forall p \in \mathcal{P}, \quad k \in \mathcal{K}(p)$$

$$x_t \in \{0,1\} \qquad \forall t \in \mathcal{T}$$

$$y_c \in \{0,1\} \qquad \forall c \in \mathcal{C}.$$

- Matching without matching variables
- Easy to solve: Small, but inherits matching properties

Balance Before After Cardinality Matching

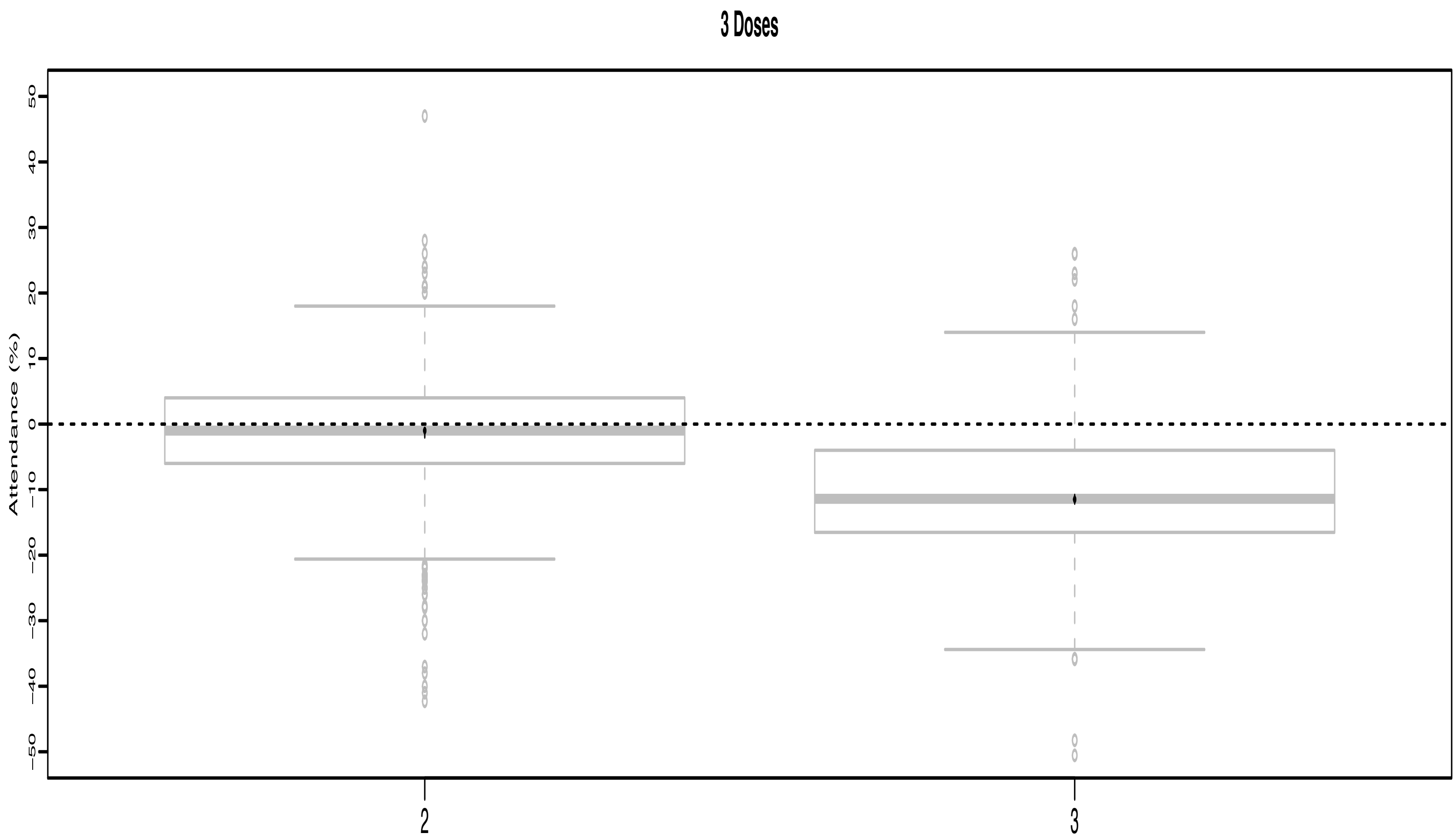


Can Also do Multiple Doses

- Dose
  1. No quake
  2. Medium quake
  3. Bad quake

	Dose		
Covariate	1	2	3
Gender			
Male	462	462	462
Female	538	538	538
School SES			
Low	75	75	75
Mid-low	327	327	327
Medium	294	294	294
Mid-high	189	189	189
High	115	115	115
Mother's education			
Primary	335	335	335
Secondary	426	426	426
Technical	114	114	114
College	114	114	114
Missing	11	11	11
⋮			

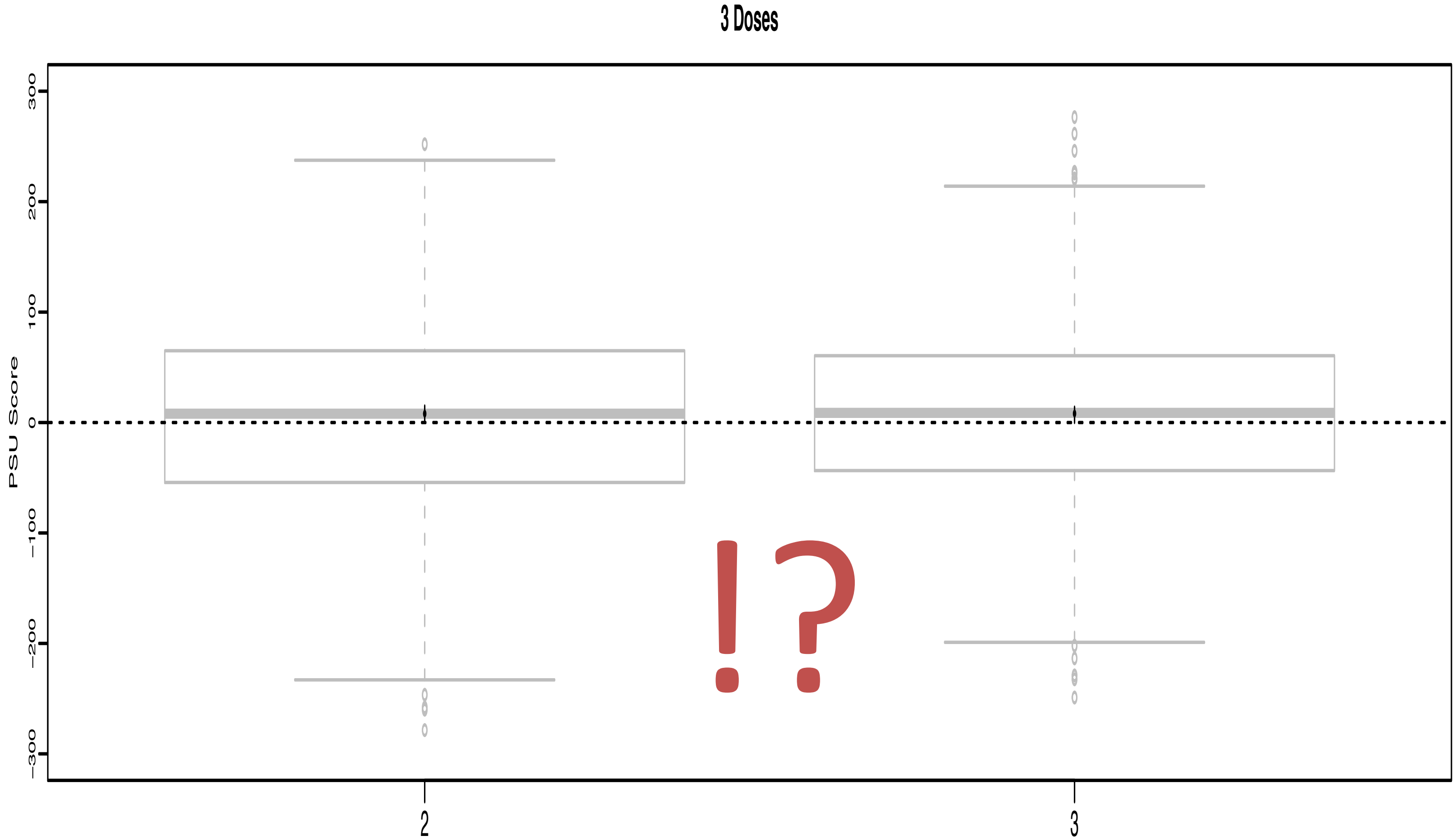
Relative (To no Quake) Attendance Impact



Medium quake

Bad quake

Relative (To no Quake) PSU Score Impact



Medium quake

Bad quake

MIP and Marketing: Chewbacca or BB-8?

Adaptive Preference Questionnaires

Feature	SX530	RX100
Zoom	50x	3.6x
Prize	$249.99	$399.99
Weight	15.68 ounces	7.5 ounces
Prefer	☑	☐


Feature	TG-4	G9
Waterproof	Yes	No
Prize	$249.99	$399.99
Weight	7.36 lb	7.5 lb
Prefer	☐	☑


Feature	TG-4	Galaxy 2
Waterproof	Yes	No
Prize	$249.99	$399.99
Viewfinder	Electronic	Optical
Prefer	☑	☐

We recommend:

Choice-based Conjoint Analysis (CBCA)

Feature	Chewbacca	BB-8
Wookiee	Yes	No
Droid	No	Yes
Blaster	Yes	No
I would buy toy	☑	☐
Product Profile	x^1	x^2

$$\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = x^2$$

Preference Model and Geometric Interpretation

- Utilities for 2 products, d features, logit model

$$U_1 = \beta \cdot x^1 + \epsilon_1 = \sum_{i=1}^{d} \beta_i x_i^1 + \epsilon_1$$

$$U_2 = \beta \cdot x^2 + \epsilon_2 = \sum_{i=1}^{d} \beta_i x_i^2 + \epsilon_2$$

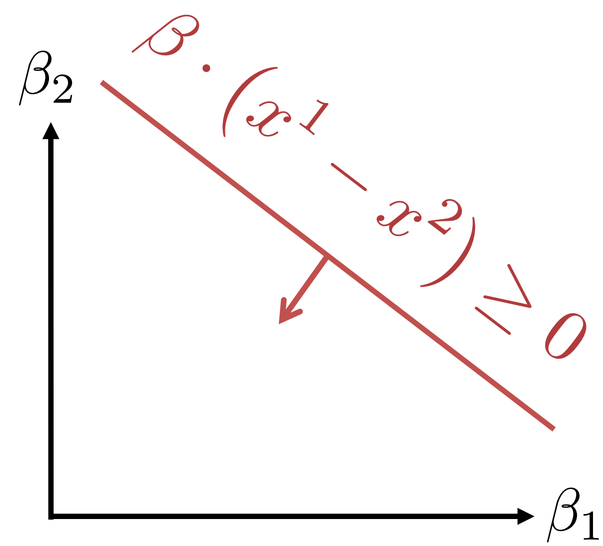
part-worths

product profile

noise (gumbel)

- Utility maximizing customer
  - Geometric interpretation of preference for product 1 without error

$$x^1 \succeq x^2 \Leftrightarrow U_1 \geq U_2$$

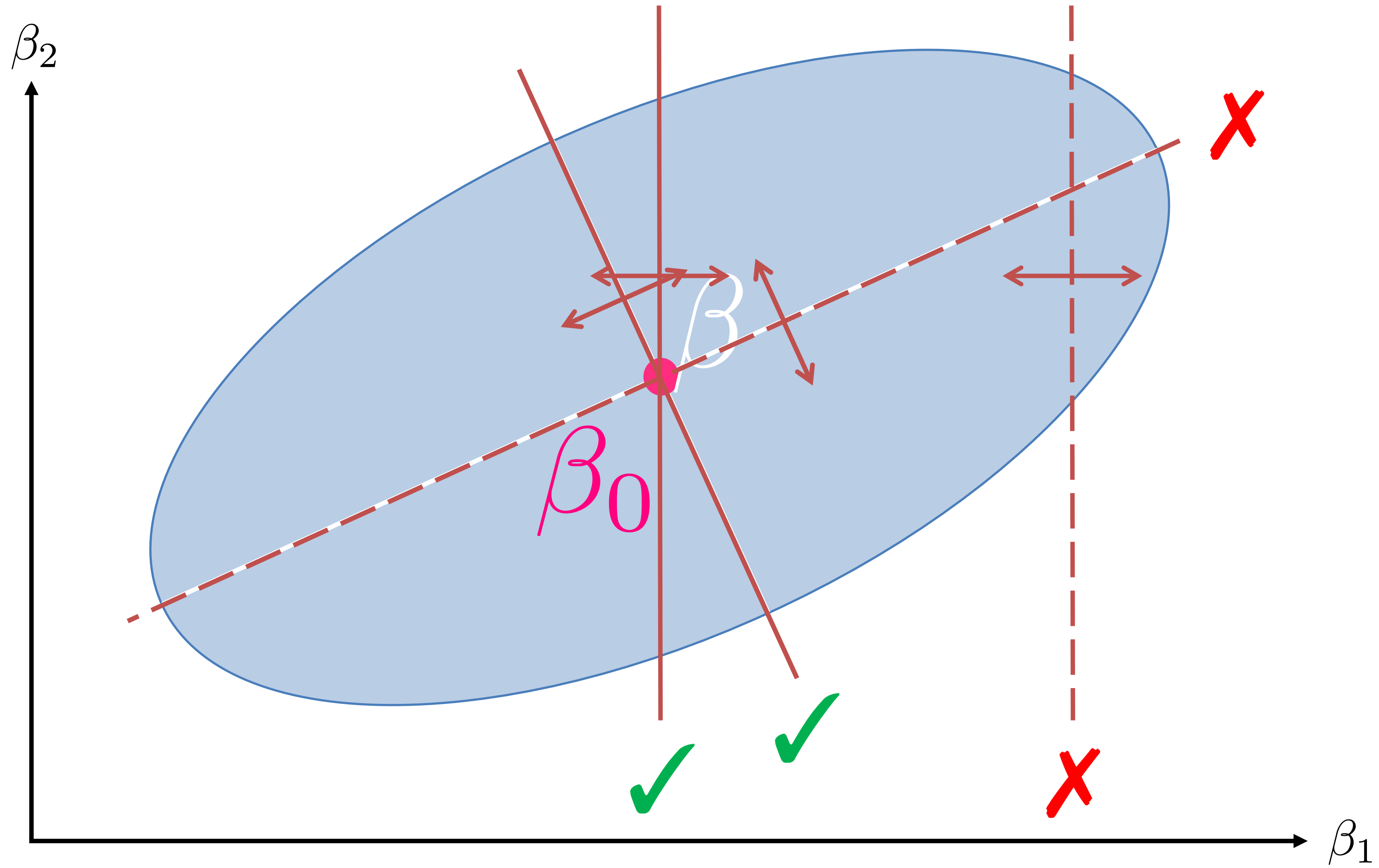


Next Question = Minimize (Expected) Volume

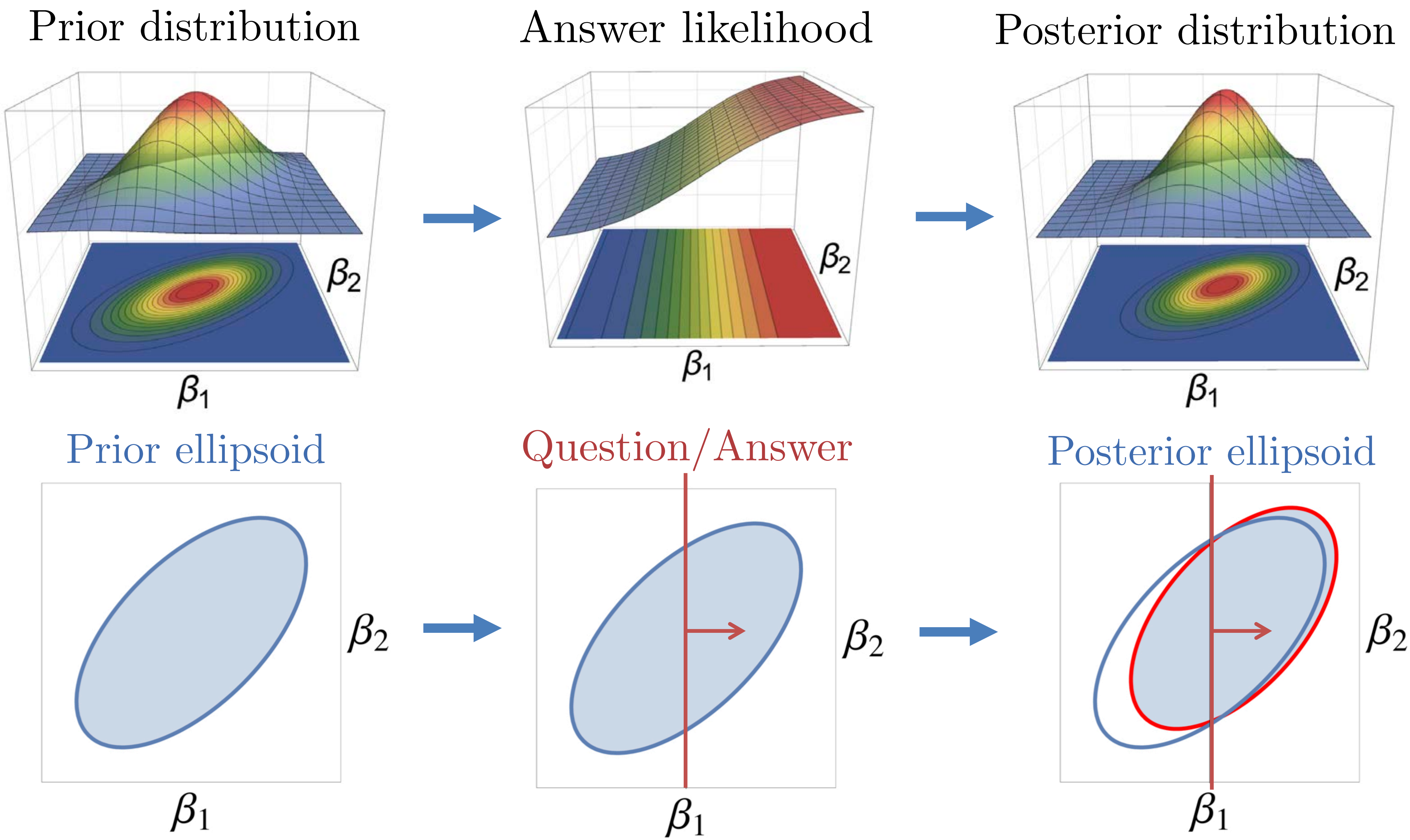
Good Question?

Good estimator for β?

1. Centroid of ellipsoid



With Error = Volume of Ellipsoid $f\left(x^1, x^2\right)$

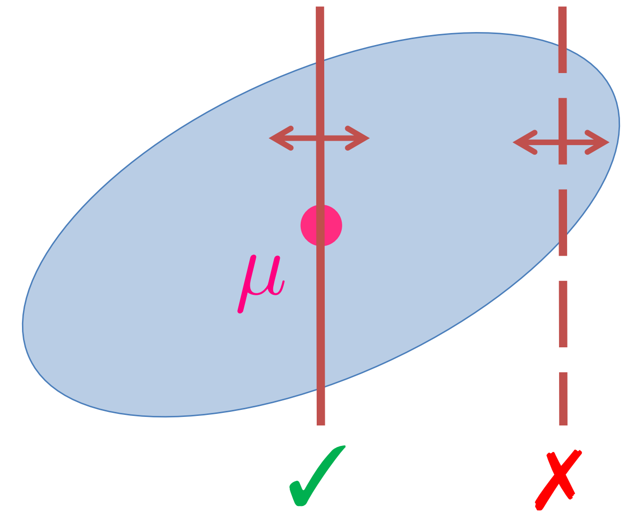


Rules of Thumb Still Good For Ellipsoid Volume

$$(\beta - \mu)' \cdot \Sigma^{-1} \cdot (\beta - \mu) \leq r$$

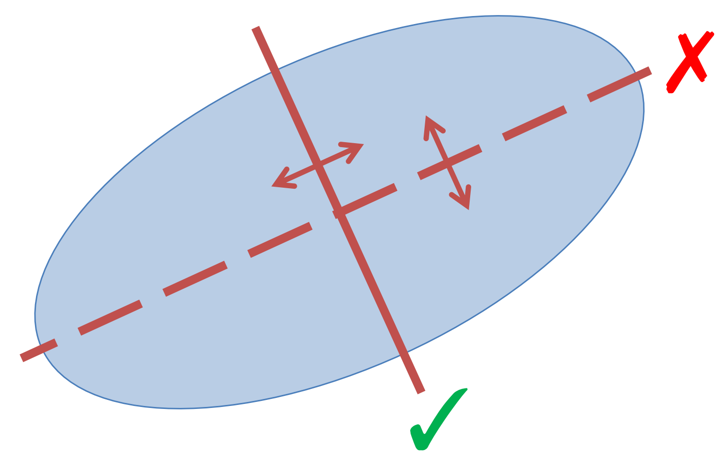
- Choice balance:
  - Minimize distance to center

$$\mu \cdot (x^1 - x^2)$$



- Postchoice symmetry:
  - Maximize variance of question

$$(x^1 - x^2)' \cdot \sum \cdot (x^1 - x^2)$$

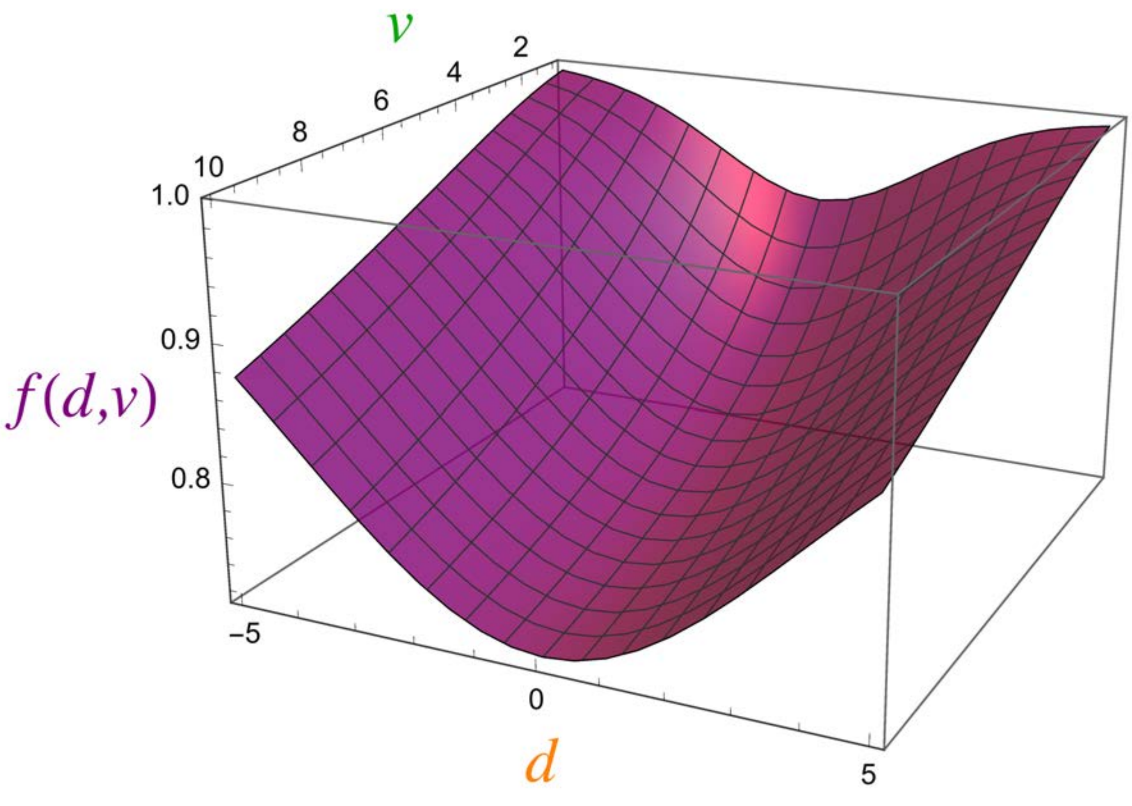


"Simple" Formula for Expected Volume

- Expected Volume = Non-convex function $f(d, v)$ of

  distance: $d := \mu \cdot (x^1 - x^2)$

  variance: $v := (x^1 - x^2)' \cdot \sum \cdot (x^1 - x^2)$



Can evaluate $f(d, v)$ with 1-dim integral 🙂

Optimization Model

$$\min \qquad f(d, v) \qquad ✗$$

$$s.t.$$

$$\mu \cdot (x^1 - x^2) = d \qquad ✓$$

$$(x^1 - x^2)' \cdot \Sigma \cdot (x^1 - x^2) = v \qquad ✗✓$$

$$A^1 x^1 + A^2 x^2 \leq b \qquad ✓$$

Formulation trick:

linearize $x_i^k \cdot x_j^l$

$$x^1 \neq x^2 \qquad ✗✓$$

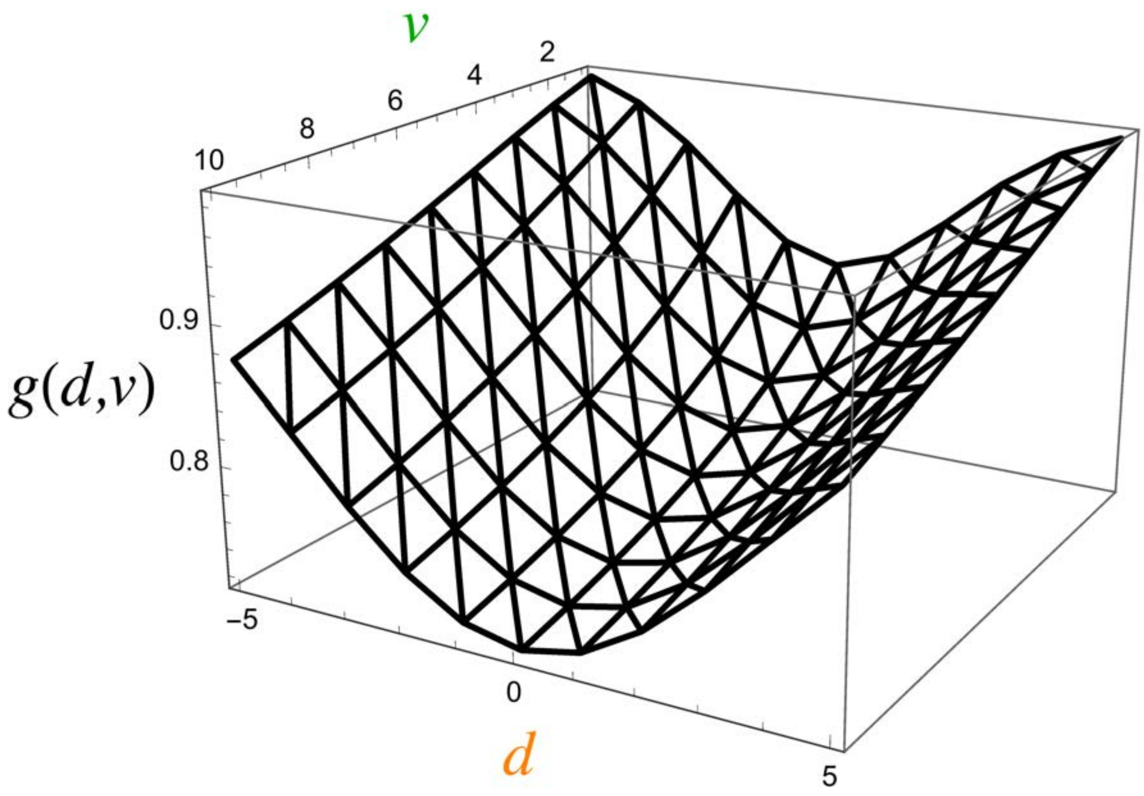
$$x^1, x^2 \in \{0, 1\}^n$$

Technique 2: Piecewise Linear Functions

- D-efficiency = Non-convex function $f(d, v)$ of

  distance: $d := \mu \cdot (x^1 - x^2)$

  variance: $v := (x^1 - x^2)' \cdot \sum \cdot (x^1 - x^2)$



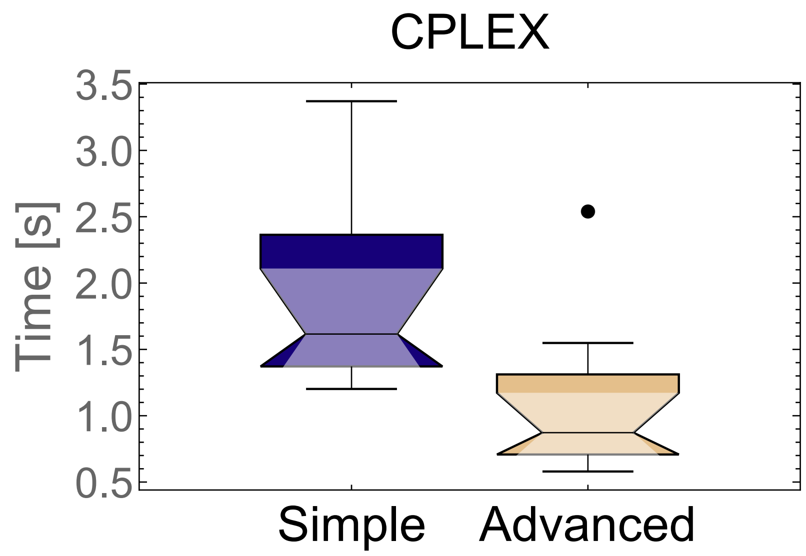
Can evaluate $f(d, v)$ with 1-dim integral 🙂

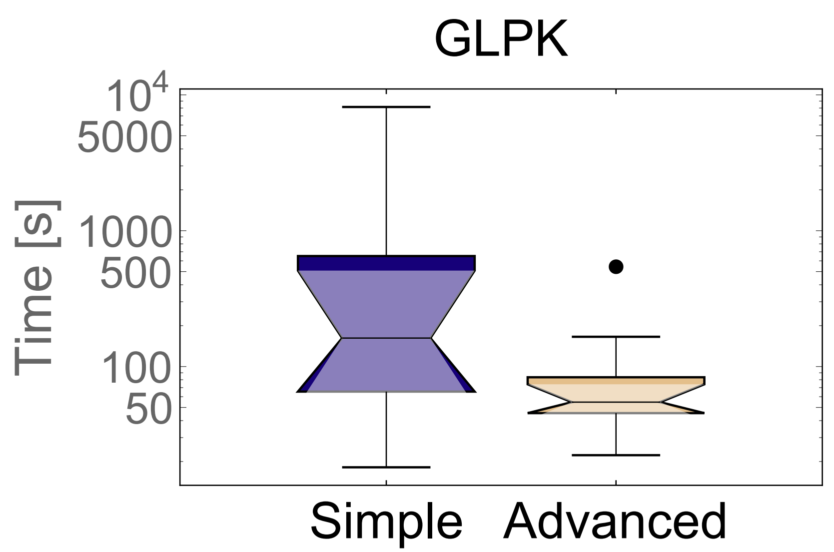
Piecewise Linear Interpolation

MIP formulation

Computational Performance

- Advanced formulations provide an computational advantage
- Advantage is significantly more important for free solvers
- State of the art commercial solvers can be significantly better that free solvers
- Still, free is free!





Summary and Main Messages

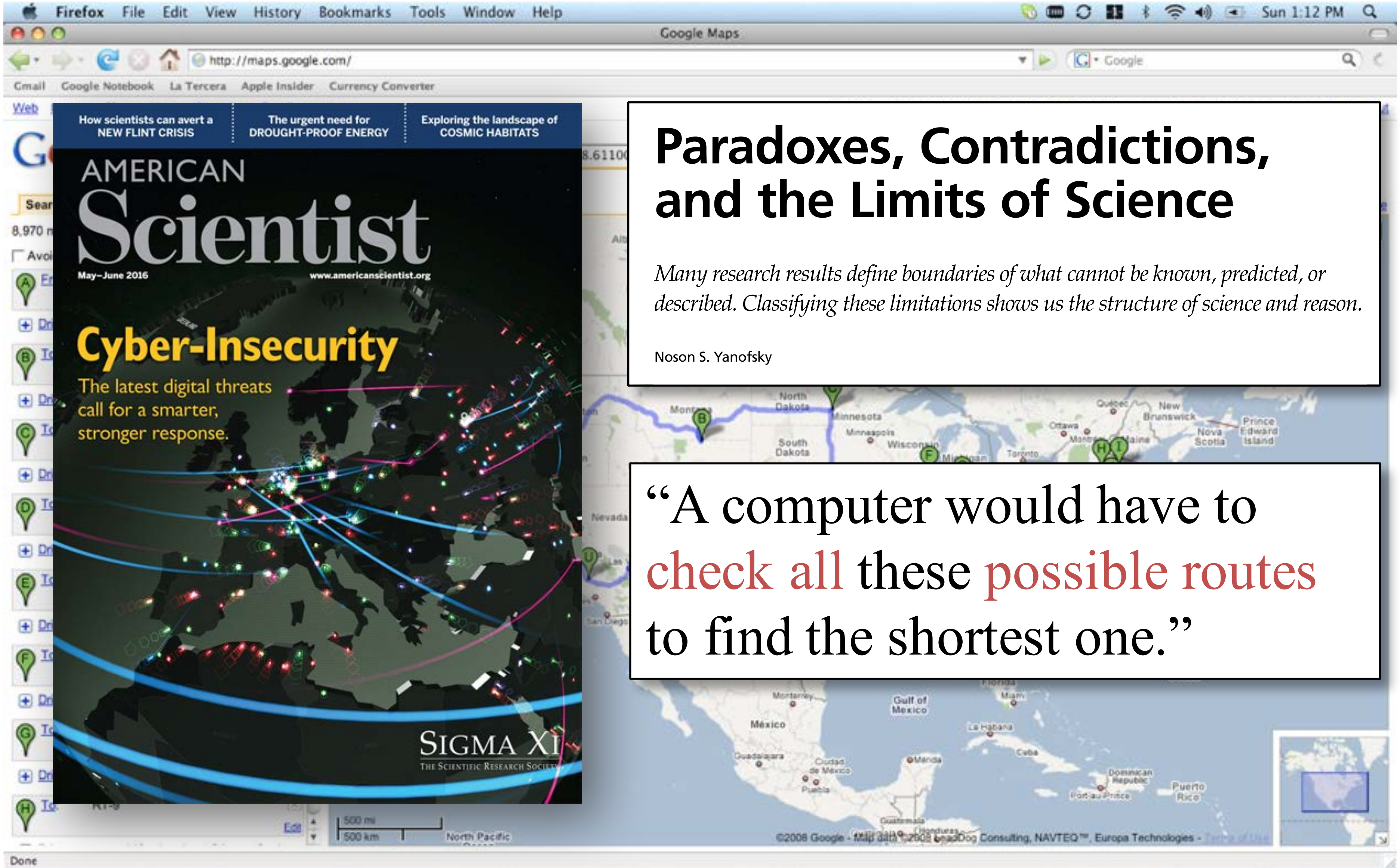
- Always choose Chewbacca!
- How to YOU use MIP / Optimization / OR / Analytics?
  - Study for the 2nd midterm!
  - Use JuMP and Julia Opt.
  - How about grad school down the river?
    - Masters of Business Analytics / OR
    - Ph.D. in Operations Research

OPERATIONS
RESEARCH
CENTER

https://orc.mit.edu

How Hard is MIP?

How hard is MIP: Traveling Salesman Problem ?



"A computer would have to check all these possible routes to find the shortest one."

MIP = Avoid Enumeration

- Number of tours for 49 cities $= 48!/2 \approx 10^{60}$
- Fastest supercomputer $\approx 10^{17}$flops
- Assuming one floating point operation per tour:
  $> 10^{35}$ years $\approx 10^{25}$ times the age of the universe!
- How long does it take on an iphone?
  - Less than a second!
  - 4 iterations of **cutting plane** method!
  - Dantzig, Fulkerson and Johnson 1954 did it by hand!
  - For more info see tutorial in ConcordeTSP app
  - **Cutting planes** are the key for effectively solving (even NP-hard) MIP problems in practice.

50+ Years of MIP = Significant Solver Speedups

- Algorithmic Improvements (Machine Independent):
  - **CPLEX** v1.2 (1991) – v11 (2007): 29,000x speedup
  - Gurobi v1 (2009) – v6.5 (2015): 48.7x speedup
  - Commercial, but free for academic use
- (Reasonably) effective free / open source solvers:
  - GLPK, **COIN-OR (CBC)** and SCIP (only for non-commercial)
- Easy to use, fast and versatile modeling languages
  - Julia based JuMP modelling language
  - http://julialang.org
  - http://www.juliaopt.org

Technique 1: Binary Quadratic $x^1, x^2 \in \{0,1\}^n$

$$(x^1 - x^2)' \cdot \Sigma \cdot (x^1 - x^2) = v$$

$$X^l_{i,j} = x^l_i \cdot x^l_j \quad (l \in \{1,2\}, \quad i,j \in \{1,\dots,n\}):$$

$$X^l_{i,j} \leq x^l_i, \quad X^l_{i,j} \leq x^l_j, \quad X^l_{i,j} \geq x^l_i + x^l_j - 1, \quad X^l_{i,j} \geq 0$$

$$W_{i,j} = x^1_i \cdot x^2_j :$$

$$W_{i,j} \leq x^1_i, \quad W_{i,j} \leq x^2_j, \quad W_{i,j} \geq x^1_i + x^2_j - 1, \quad W_{i,j} \geq 0$$

$$\sum_{i,j=1}^{n} (X^1_{i,j} + X^2_{i,j} - W_{i,j} - W_{j,i}) \Sigma_{i,j} = v$$

Technique 1: Binary Quadratic $x^1, x^2 \in \{0,1\}^n$

$$x^1 \neq x^2 \quad \Leftrightarrow \quad \|x^1 - x^2\|_2^2 \geq 1$$

$$X_{i,j}^l = x_i^l \cdot x_j^l \quad (l \in \{1,2\}, \quad i,j \in \{1,\ldots,n\}) :$$

$$X_{i,j}^l \leq x_i^l, \quad X_{i,j}^l \leq x_j^l, \quad X_{i,j}^l \geq x_i^l + x_j^l - 1, \quad X_{i,j}^l \geq 0$$

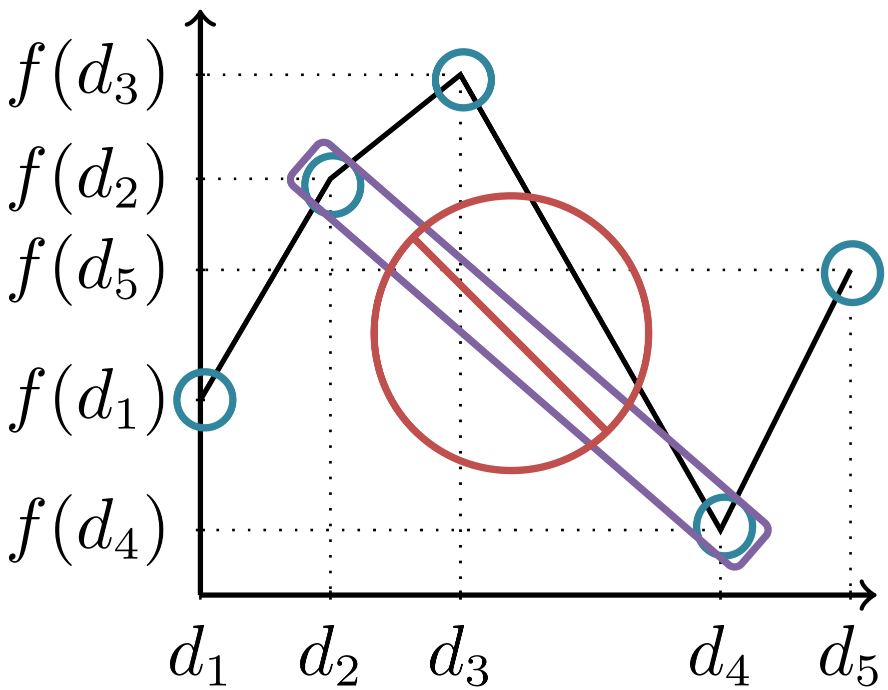
$$W_{i,j} = x_i^1 \cdot x_j^2 :$$

$$W_{i,j} \leq x_i^1, \quad W_{i,j} \leq x_j^2, \quad W_{i,j} \geq x_i^1 + x_j^2 - 1, \quad W_{i,j} \geq 0$$

$$\sum_{i,j=1}^{n} \left(X_{i,j}^1 + X_{i,j}^2 - W_{i,j} - W_{j,i}\right) \geq 1$$

Simple Formulation for Univariate Functions

$$z = f(x)$$



$$\begin{pmatrix} x \\ z \end{pmatrix} = \sum_{j=1}^{5} \begin{pmatrix} d_j \\ f(d_j) \end{pmatrix} \lambda_j$$

$$1 = \sum_{j=1}^{5} \lambda_j, \quad \lambda_j \geq 0$$

$$y \in \{0,1\}^4, \quad \sum_{i=1}^{4} y_i = 1$$

$$0 \leq \lambda_1 \leq y_1$$

$$0 \leq \lambda_2 \leq y_1 + y_2$$

$$0 \leq \lambda_3 \leq y_2 + y_3$$

$$0 \leq \lambda_4 \leq y_3 + y_4$$

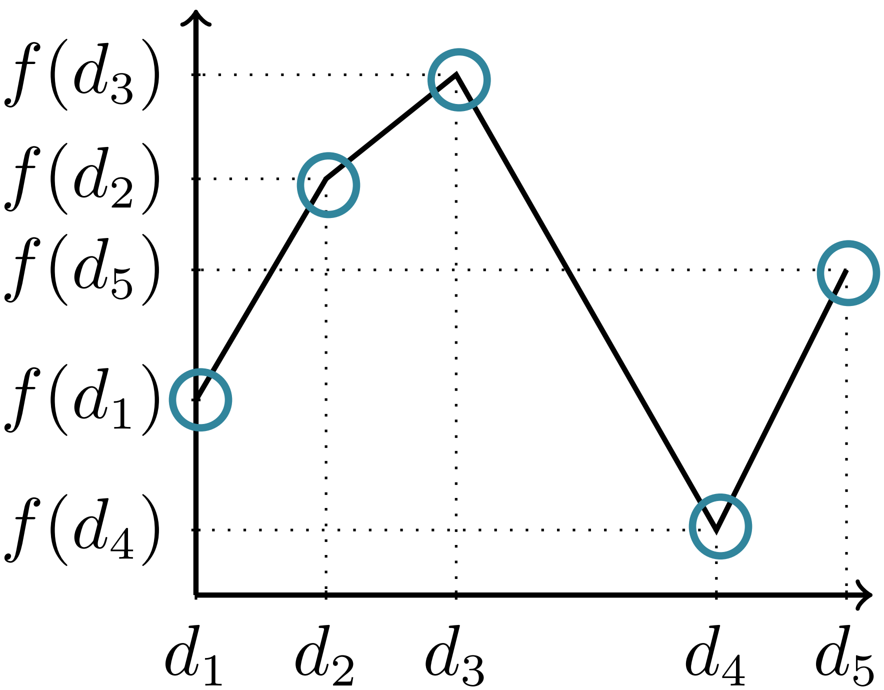
$$0 \leq \lambda_5 \leq y_4$$

$$\text{Size} = O(\#\text{ of segments})$$

Non-Ideal: Fractional Extreme Points

Advanced Formulation for Univariate Functions

$$z = f(x)$$



$$\text{Size} = O\left(\log_2 \# \text{ of segments}\right)$$

Ideal: Integral Extreme Points

$$\begin{pmatrix} x \\ z \end{pmatrix} = \sum_{j=1}^{5} \begin{pmatrix} d_j \\ f(d_j) \end{pmatrix} \lambda_j$$

$$1 = \sum_{j=1}^{5} \lambda_j, \quad \lambda_j \geq 0$$

$$y \in \{0,1\}^2$$

$$0 \leq \lambda_1 + \lambda_5 \leq 1 - y_1$$

$$0 \leq \lambda_3 \leq y_1$$

$$0 \leq \lambda_4 + \lambda_5 \leq 1 - y_2$$

$$0 \leq \lambda_1 + \lambda_2 \leq y_2$$