Predicting the Unpredictable: Cognitive Bias in Elite Figure Skating

Slicing the data behind the often misaligned judgement system

Why figure skating?

- I know next to nothing about figure skating
- The +/- 3 GOE ISU judging system is designed to be both objective and subjective (starting this year the ISU judging system has switched to a +/-5 GOE system)
- Figure skating is unique in the sense that there are artistic and technical sides that are awarded scores based on mathematical formulas

Why the ISU judging system was reformed

"I voted with my conscience on February 11, 2002, and today I would make the same choice without hesitation."

- At the 2002 Olympics in Salt Lake City two pairs teams were awarded the gold medal because of a controversy
- The Canadian team fell during the final pose of their performance but since it wasn't during an element they weren't given a deduction. The Russian team had a minor technical error when one of the skaters stepped out of a double axel
- The Russians were originally declared the winners, but the French judge said she was pressured into voting for the Russian
- After a review of the competition by the ISU, both teams were awarded gold and a new ISU judging system created

New ISU Judging System

- Started in 2004, also called the Code of Points System
- Judgement is split into 2 categories:
    1. Technical elements
    2. Components
- Technical elements: scores are calculated by awarding a grade of execution (GOE) factor for each element, which is then translated into a value using a scale of value (SOV). The GOE values are then averaged using a trimmed mean process, and finally the averaged value is added to the base value for the element.
- Components: scores are based on a 1-10 scale, are not averaged nor dependent on a GOE, but components can be marked for deductions and are based on a factor depending on type category of competition (mens, ladies, pairs). The components are largely considered the more artistic, or subjective, of the two judged aspects and therefore considered the most open to interpretation.

SOV - http://www.usfsa.org/content/2018-19%20SP%20Scale%20of%20Values.pdf

Examples of Components and Elements

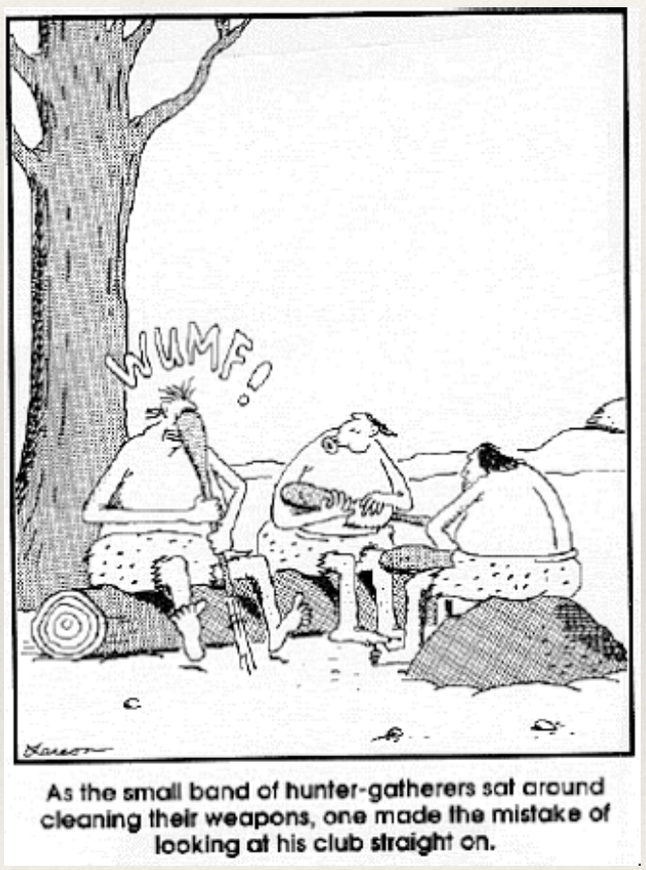
Components:

Alena KOSTORNAIA (RUS)| Ladies Short Program | Linz 2018
https://www.youtube.com/watch?v=qNqkzHtJQF4&t=150s

Elements:

Mirai NAGASU (USA) | Ladies Short Program | Winter Olympics 2018
https://youtu.be/GBWLP-iDUqg?t=9

Data Cleaning/Inspection



As the small band of hunter-gatherers sat around cleaning their weapons, one made the mistake of looking at his club straight on.

- Original data from Buzzfeed's article on national bias in figure skating judging - https://github.com/BuzzFeedNews/2018-02-figure-skating-analysis
- Used SQL to combine performance data with competitor data
- Created data frames for ladies competitions, ladies elements and components

Ladies Long Program

- **Maximum of seven jump elements**
  - One must be an axel type jump
  - Maximum of three jump combinations/sequences; one jump combination may contain three jumps
  - Double jumps cannot be included more than twice
  - Of all the triple and quadruple jumps only two can be executed twice; of the two, only one can be a quadruple jump

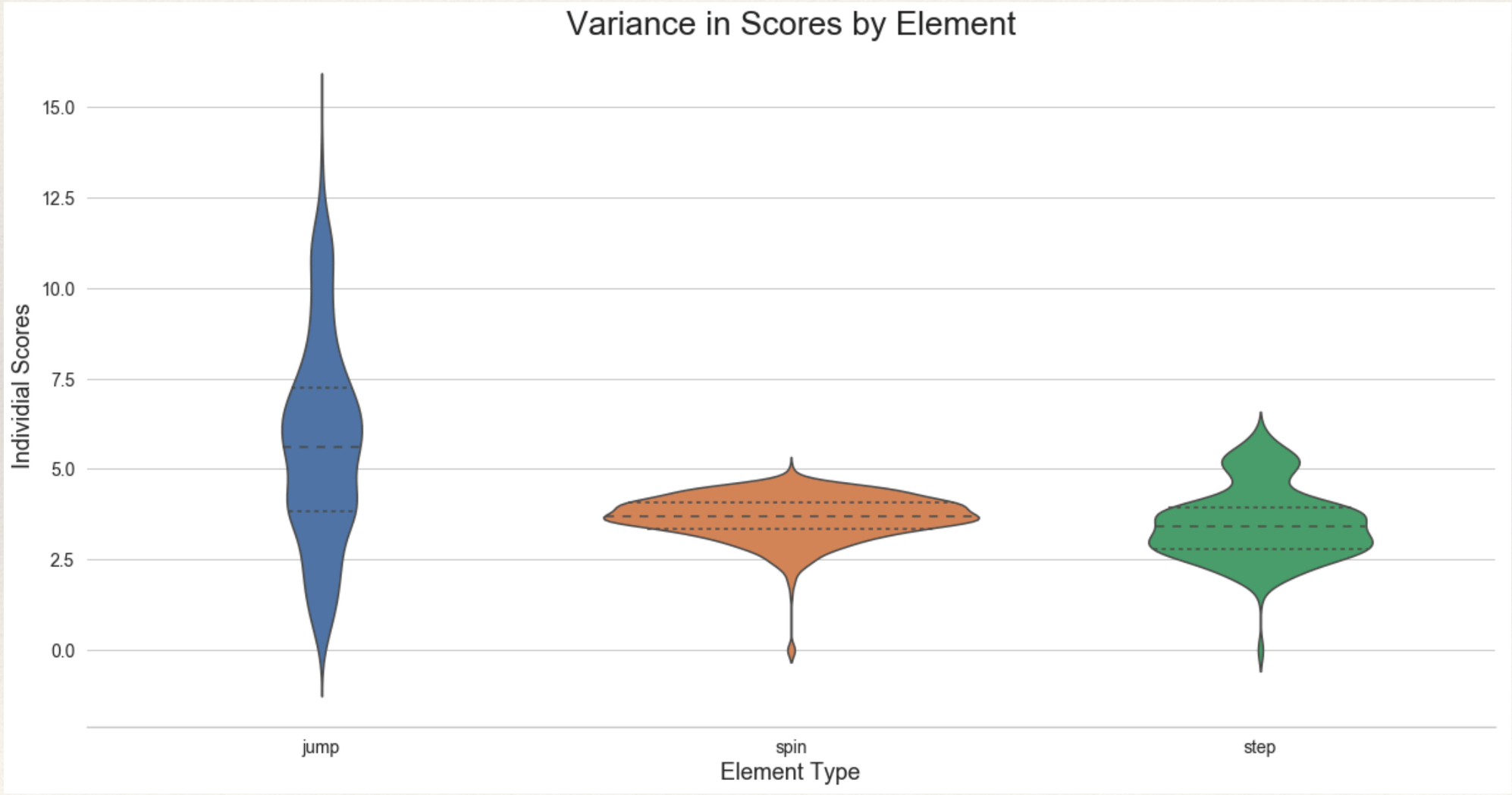
- **Maximum of three spins**
  - One combination spin; minimum of 10 revolutions
  - One flying spin or spin with a flying entrance; minimum of six revolutions
  - One spin with only one position; minimum of six revolutions
  - A change of foot optional in all spins

- **Maximum of one step sequence**

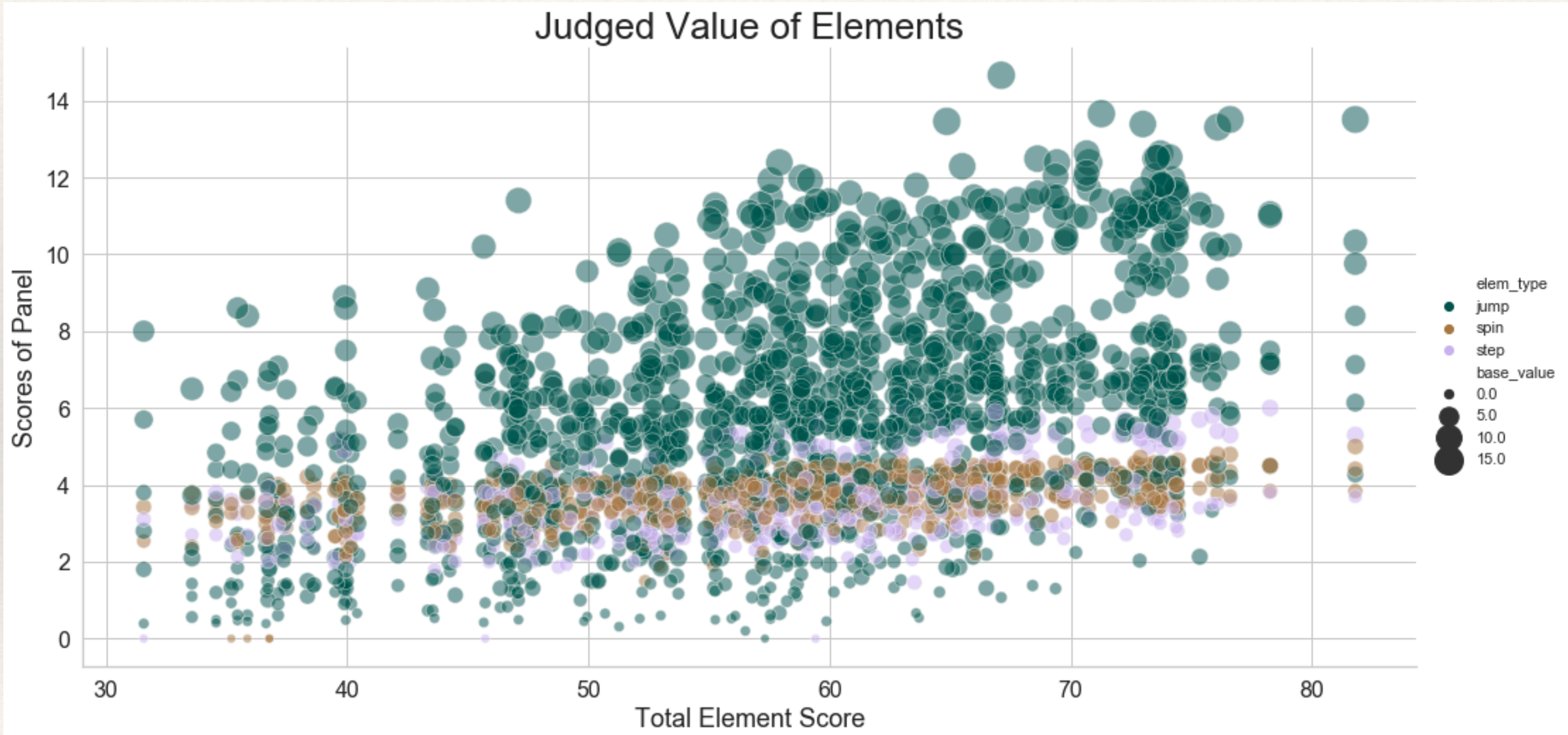
- **Maximum of one choreographic sequence**

Elements

- The technical elements of ladies competitions are organized into the following categories:
  - Jumps
  - Spins
  - Steps



Elements

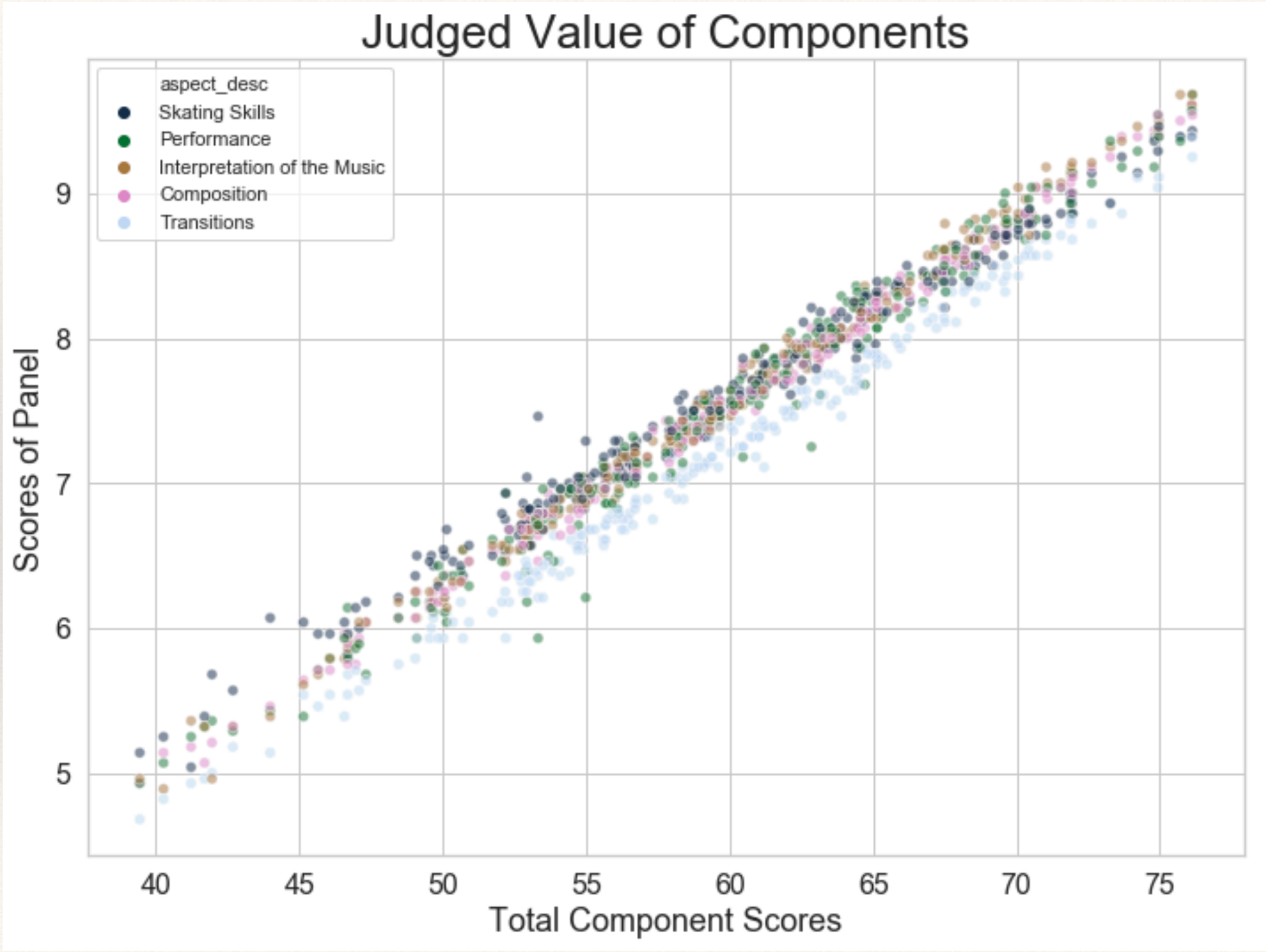

Judged Value of Elements
Scores of Panel
Total Element Score
0
2
4
6
8
10
12
14
30
40
50
60
70
80
elem_type
jump
spin
step
base_value
0.0
5.0
10.0
15.0

Components

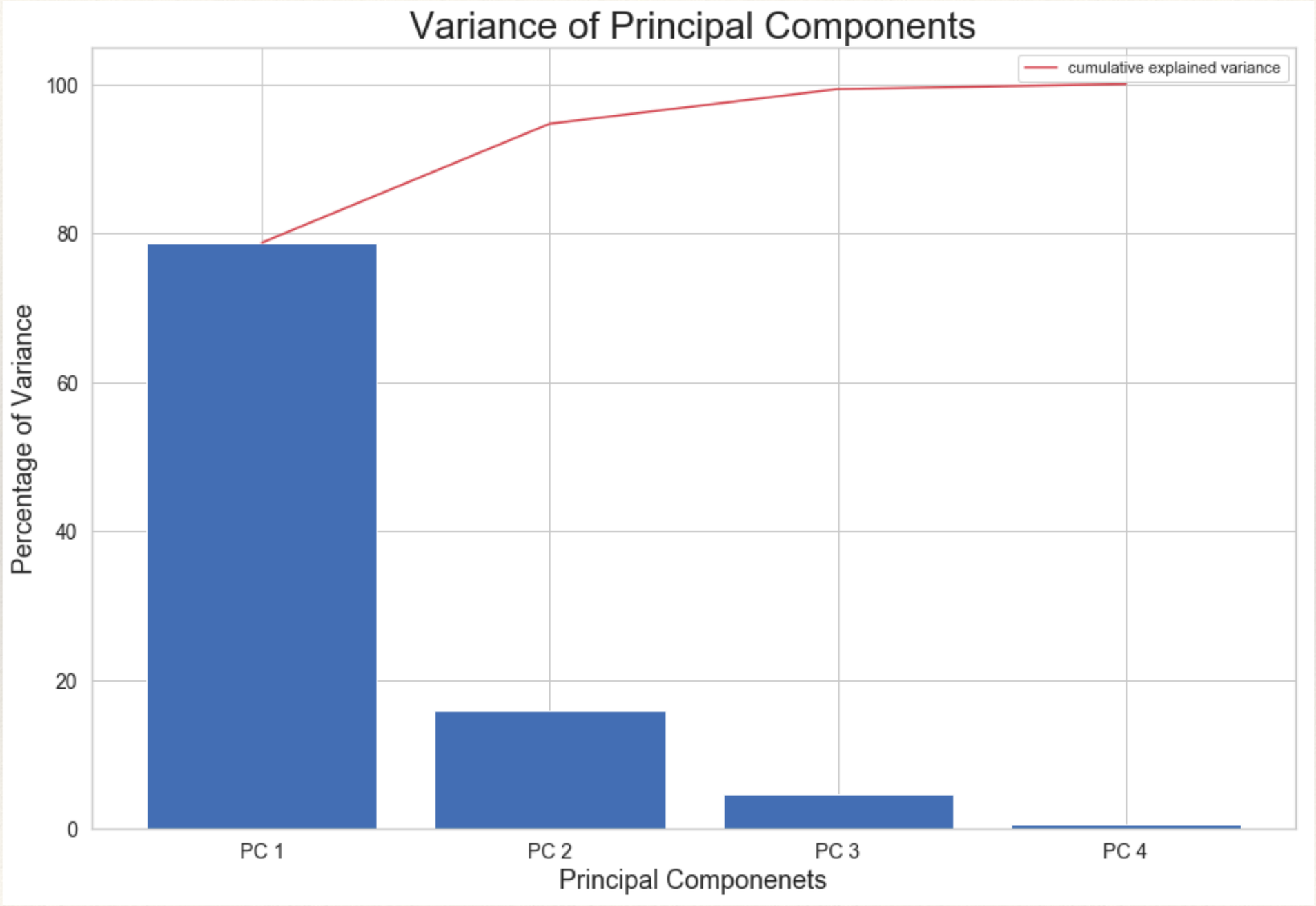
- Program Component Scores (PCS) fall into five categories:
  - Skating Skills
  - Transitions
  - Performance
  - Composition
  - Interpretation of the Music/Timing

The component scores are awarded at the end of the performance

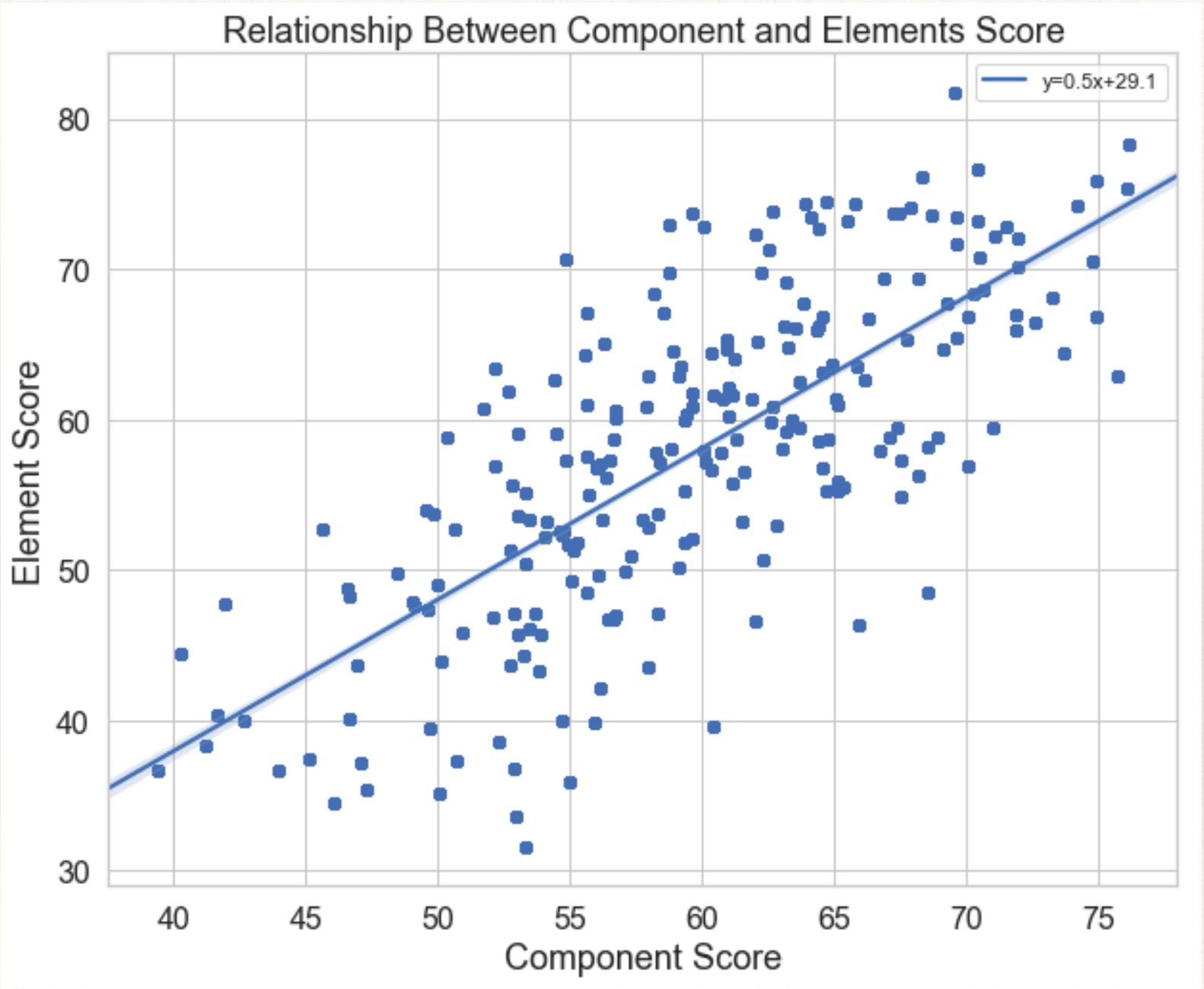
Components


Judged Value of Components
aspect_desc
Skating Skills
Performance
Interpretation of the Music
Composition
Transitions
Scores of Panel
9
8
7
6
5
40
45
50
55
60
65
70
75
Total Component Scores

PCA of Component Scores



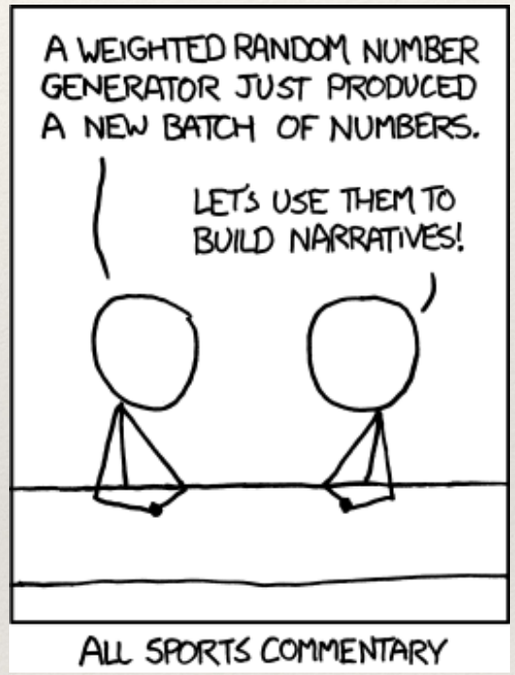
Question



Can component scores be predicted from the attributes of technical elements?

Data Engineering

- GOE in the dataset is the point value of the factor, divided this value by base value to get actual weighted GOE value



- In order to properly create a model showing the effect of element scores on component scores, the element scores of each performance must be engineered to show proper format/logic
- Decided to rank the each element in a performance from highest base value to lowest

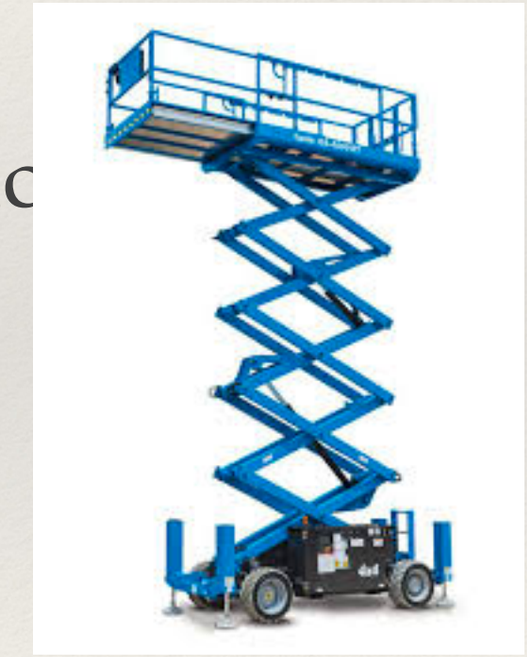
Data Engineering

	index	performance_id	total_deductions	base_value	goe	goe_actual	rank
2416	21048	059039bcf6	0.0	11.22	0.6	0.053476	1.0
1063	9525	c351343d7d	0.0	8.30	-0.4	-0.048193	1.0
747	6647	7d9640cf7e	1.0	9.60	1.3	0.135417	1.0
296	2667	8712edfbc0	0.0	8.20	0.8	0.097561	1.0
294	2658	265a603120	0.0	11.33	-0.5	-0.044131	1.0

- Process:
  - Add rank column to each element, grouped by performance id
  - There is only one deduction per performance, added this back to the data frame after engineering
  - Sorted all values by the new rank column, then created pivot table of data ready for the model
  - For target data, showed each component score for each performance, then added the sum as the final target column

Gradient Boosting Machine

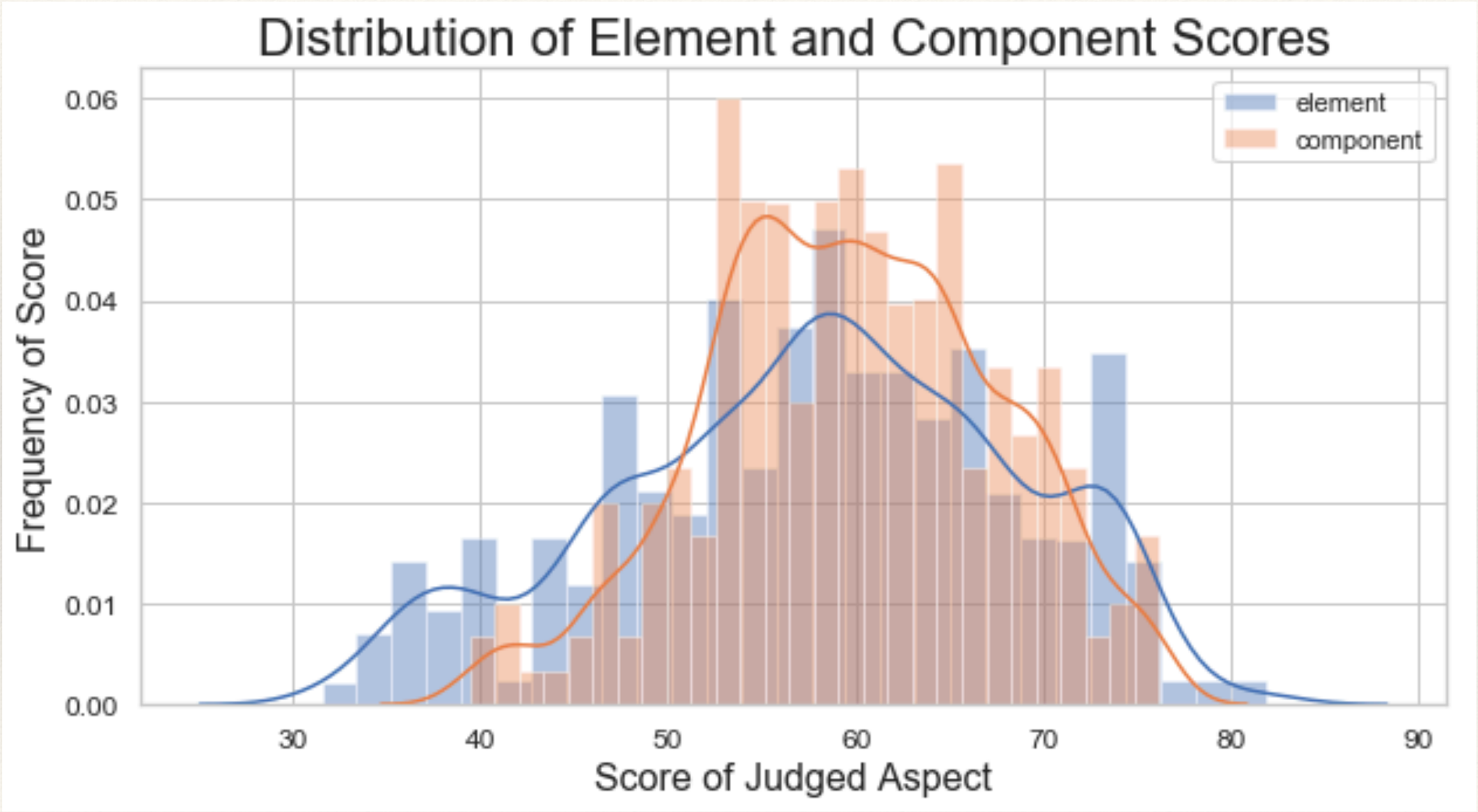
- Wanted to use machine learning to help explain the variance in the component scores
- GBM performs well on a variety of tasks
- GBM provides statistical analysis and metric for further analysis
- GBM provides a ranking of feature importance
- Used H2O Flow GBM to gain experience with this tool/work with something new



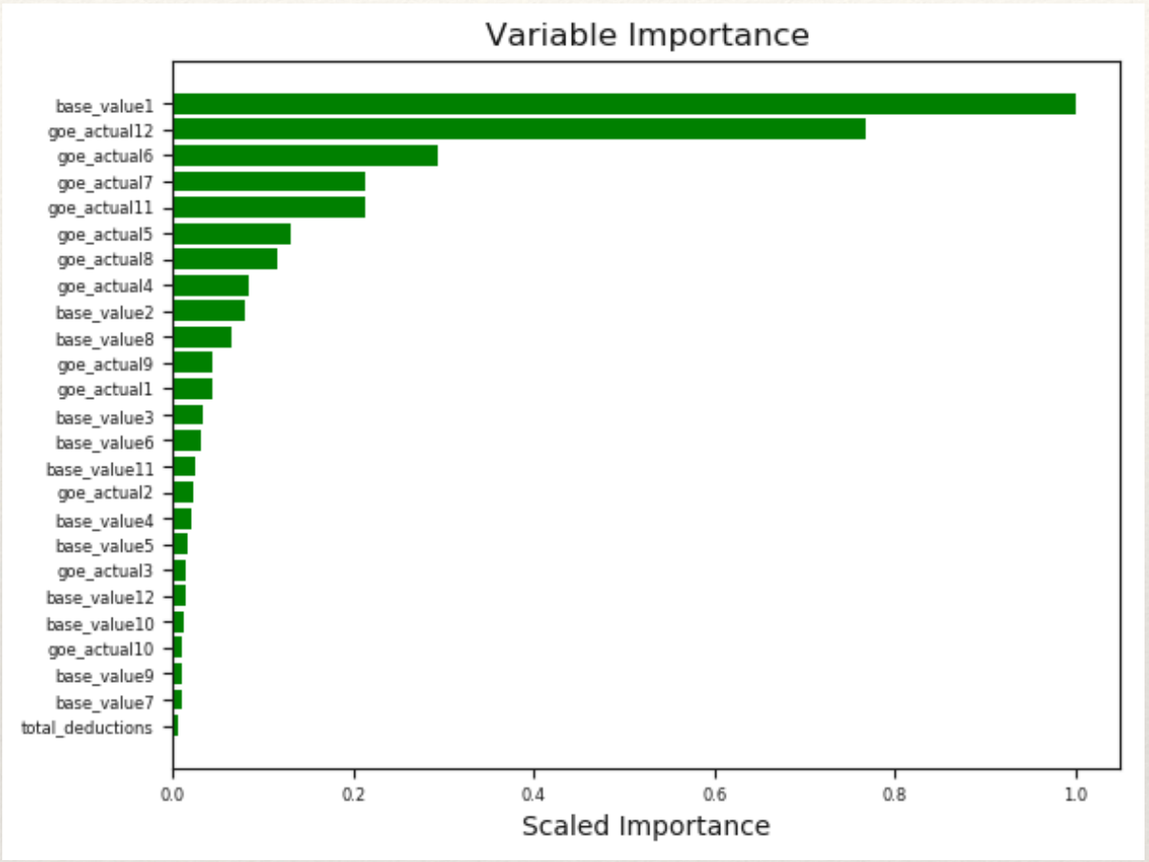
Model Parameters

- Distribution: Gaussian
- Train/Test/Valid Ratios: 0.7/0.15
- ntrees: 100
- Learn Rate: 0.1
- Sample Rate: 0.75
- Max Depth: 4
- N-folds: 5
- Stopping metric: Deviance
- Seed generator: 1234

Distribution of Scores



Predicting Component Scores



Mean R-squared: 72% +/- 0.05%

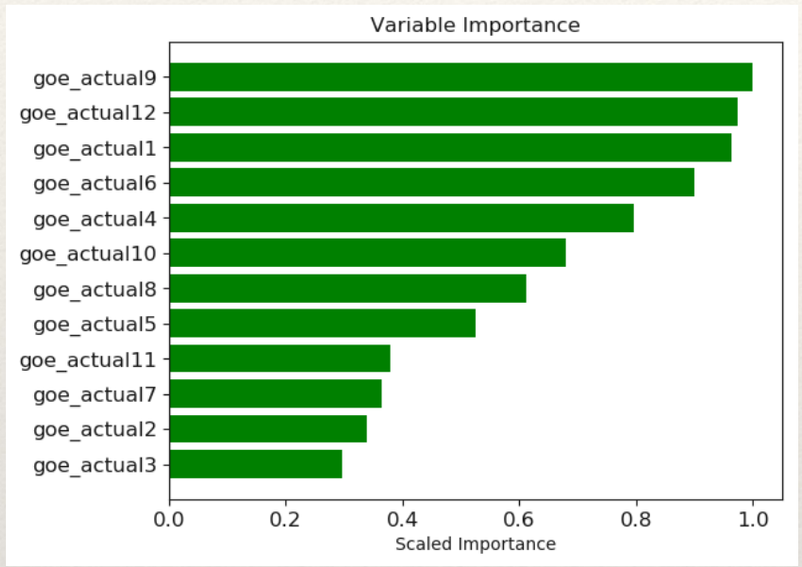
MSE: 0.772

- The base value of the highest scoring element has the greatest weight on the model. The fact that the goe_actual12 has the second most importance among the GOE features shows that the goe of the lowest scoring element has a large effect on the component score.

- Total deductions are the least important feature to predicting component scores

Further Testing

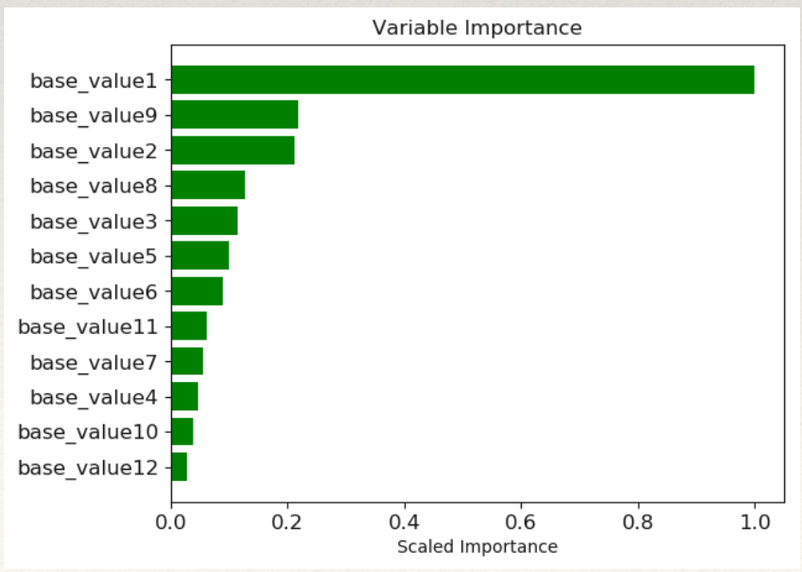
GOE



Mean R-squared: 71% +/- 0.06%

MSE: 1.41

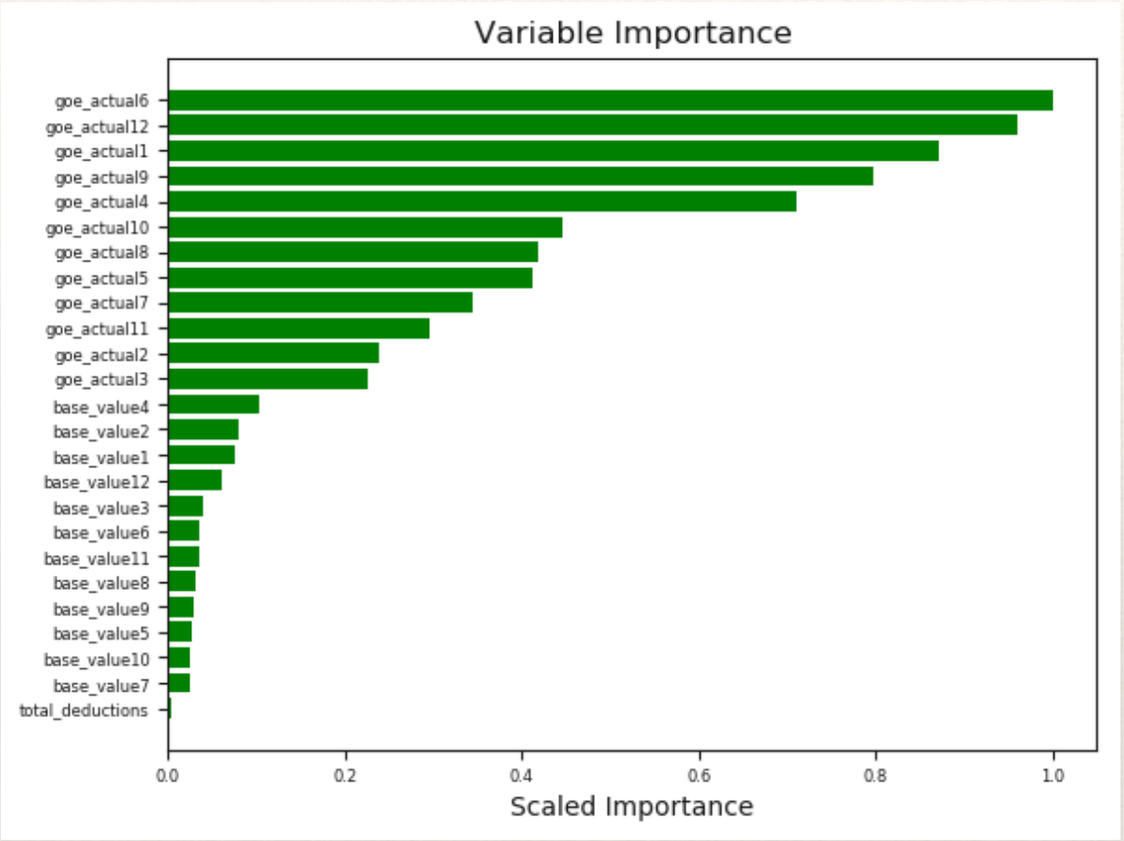
Base Value



Mean R-squared: 32% +/- 0.06%

MSE: 11.92

Chronologically Formatted Data



- **Mean R-squared: 56% +/- 0.05%**
- MSE: 0.791

- While this model is less accurate, it shows that component scores can still be predicted from chronologically ordered elements
- The fact that GOEs are the most important features show that the type of element does not have an effect on component scores, but rather its difficulty
- Again, deductions are the least important feature

Test on Olympic Data

	Component Scores	Predicted
0	96.84	94.498499
1	70.58	71.268268
2	78.66	82.720768
3	92.84	83.329291
4	89.06	88.035280
5	67.92	73.308292
6	62.58	66.665805
7	94.56	88.282318
8	83.00	77.749108
9	93.78	90.991074
10	77.20	68.915513
11	85.14	80.885272
12	69.98	78.203245
13	63.80	58.367511
14	84.14	75.155064
15	93.78	88.851385
16	80.20	77.683993
17	73.50	70.690732
18	63.00	66.856125
19	85.34	80.599615
20	67.44	70.484974
21	77.56	72.821638
22	64.26	72.416549
23	94.56	79.460213
24	65.16	63.576605
25	78.44	74.619251
26	80.12	79.279675
27	71.00	71.450746
28	70.74	70.486764

The correlation between the score sets is 0.8790248156611701%.

- The predictive capability of the model will be tested on Olympic data not included in the original analysis
- The only difference between the Olympic competition and others is deductions are given a negative value

Summary

- Figure skating judges are tasked with judging at least 12 elements for numerous competitors (30 for the ladies long program in the Olympics for example).
- It seems natural that judges would take mental 'shortcuts' to process this information, and using the element score as a basis for the component score could be one of these shortcuts.
- There is statistical evidence that component scores are greatly influenced by the technical elements of the ladies long program, showing one example of such a “shortcut”