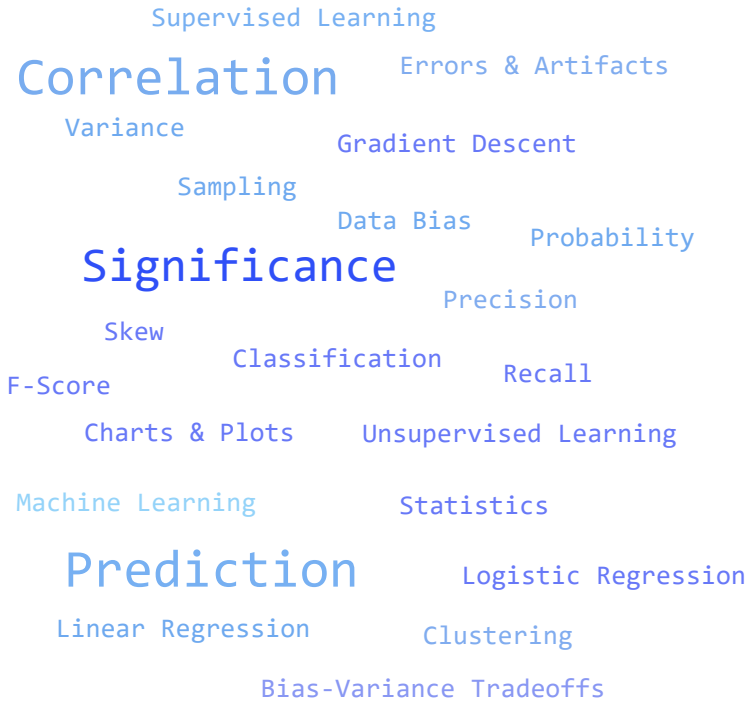


Data Science 1:
Introduction to Data Science

Scores & Rankings

Winter 2025

Wolfram Wingerath, Jannik Schröder

Department for Computing Science
Data Science / Information Systems

Lecture slides based on content from “The Data Science Design Manual” (Steven Skiena, 2017) and associated course materials generously made available online by the author at https://www3.cs.stonybrook.edu/~skiena/data-manual/.

Special thanks to Professor Skiena for sharing these valuable teaching resources!

Semester Schedule

CW 42	14. Oct	Lecture	1	Orga & Intro	1-26
CW 43	21. / 23. Oct	Lecture + Exercises	2	Probability, Statistics & Correlation	27-56
CW 44	28. Oct	Lecture	3	Data Munging, Cleaning & Bias	57-94 / “Invisible Women”
CW 45	04. / 06. Nov	Lecture + Exercises	4	Scores & Rankings	95-120
CW 46	11. Nov	Lecture	5	Statistical Distributions & Significance	121-154
CW 47	18. / 20. Nov	Lecture + Exercises	6	Building & Evaluating Models	201-236
CW 48	25. Nov	Guest Lecture	7	Data Visualization	155-200
CW 49	02. / 04. Dec	Lecture + Exercises	8	Intro to Machine Learning	351-390
CW 50	09. Dec	Lecture	9	Linear Algebra	237-266
CW 51	16. / 18. Dec	Lecture + Exercises	10	Linear Regression & Gradient Descent	267-288
CW 02	06. Jan	Lecture	11	Logistic Regression & Classification	289-302
CW 03	13. / 15. Jan	Lecture + Exercises	12	Nearest Neighbor Methods & Clustering	303-350
CW 04	20. Jan	Lecture	13	Data Science in the Wild	391-426
CW 05	27. / 29. Jan	Lecture + Exercises	14	Q&A / Feedback	
CW 06	03. / 04. Feb	Oral Exams (Block 1)		Preparation in our last session („Oral Exam Briefing“)	
CW 13	24. / 25. Mar	Oral Exams (Block 2)			

Scores and Rankings

Scoring functions are measures that reduce multi-dimensional data to a single value, to highlight some particular property.

Rankings order items, usually by sorting scores.

BEST COLLEGES
U.S.News & WORLD REPORT
RANKINGS

Coca-Cola
MEN'S WORLD RANKING
FIFA

BTI CLIENTOPIA® 24 2016

1	Skadden, Arps, Slate, Meagher & Flom
2	Kirkland & Ellis
3	Hogan Lovells
4	McGuireWoods
5	Jones Day
6	Alston & Bird
7	Bryan Cave
8	Sidley Austin
9	Mayer Brown
10	Ogletree, Deakins, Nash, Smoak & Stewart
11	Orrick, Herrington & Sutcliffe
12	Faegre Baker Daniels

Click Here for Firms 13 – 24

THE WORLD UNIVERSITY RANKINGS
ACADEMIC RANKING OF WORLD UNIVERSITIES
QS WORLD UNIVERSITY RANKINGS
BEST GLOBAL UNIVERSITIES U.S.News RANKINGS
THOMSON REUTERS
multirank

ITA ORACLE
COLLEGE TENNIS RANKINGS

POWER RANKINGS

SPEEDRUN.COM

Assigning Grades

Course grades get assigned by scoring functions. Observe that grading systems have:

- Degrees of arbitrariness: each teacher differs.
- Lack of validation data: there is no *right* grade.
- General robustness: students tend to get similar grades in all their classes anyway.

Calling scores *statistics* lends them more dignity.

Student Transcript Performance?

Propose a statistic to map a student's grade transcript at UOL to a single numerical score such that better performance gets a higher score!

Scoring vs. Regression

The critical issue in designing scoring functions is that there is no gold standard/right answer.

Machine learning techniques like linear regression can learn a scoring function from features if you had training data, which generally you don't.

Google's ranking algorithms train on click data.

The Body-Mass Index (BMI)

BMI is a score designed to capture whether your weight is under control:

$$BMI = \frac{mass}{height^2}$$

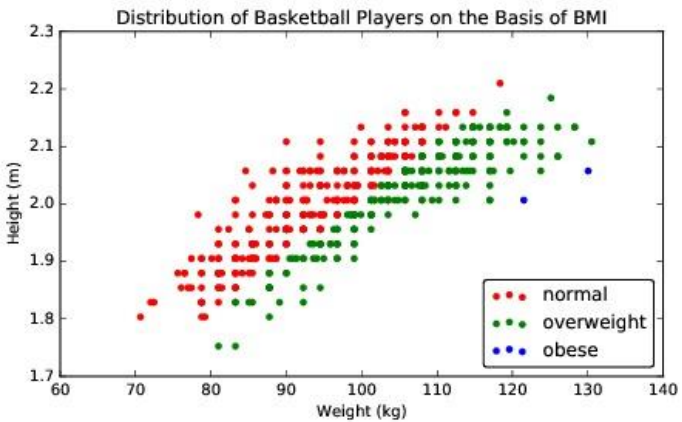
Mass is in kg and height in meters

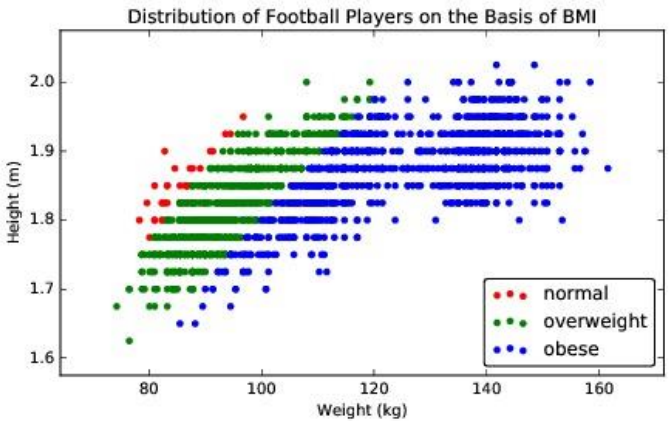
- Underweight: below 18.5
- Normal: 18.5 to 25
- Overweight: 25 to 30
- Obese: over 30

$$BMI_{Wolle} = \frac{40}{1.5^2} \approx 17.78$$

BMI: Pro Basketball and US Football

BMI is easy to interpret, and correlates with body fat. Mass should scale with the square (or cube) of height.

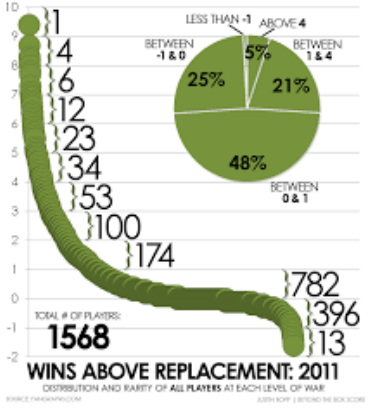




Wins Above Replacement (WAR)

- Map the raw components each player's record to the number of games it "wins".
- Subtract the expected "wins" of a garbage player given as many chances.
- Correct for parks that are easy/hard to hit in, or pitch in.

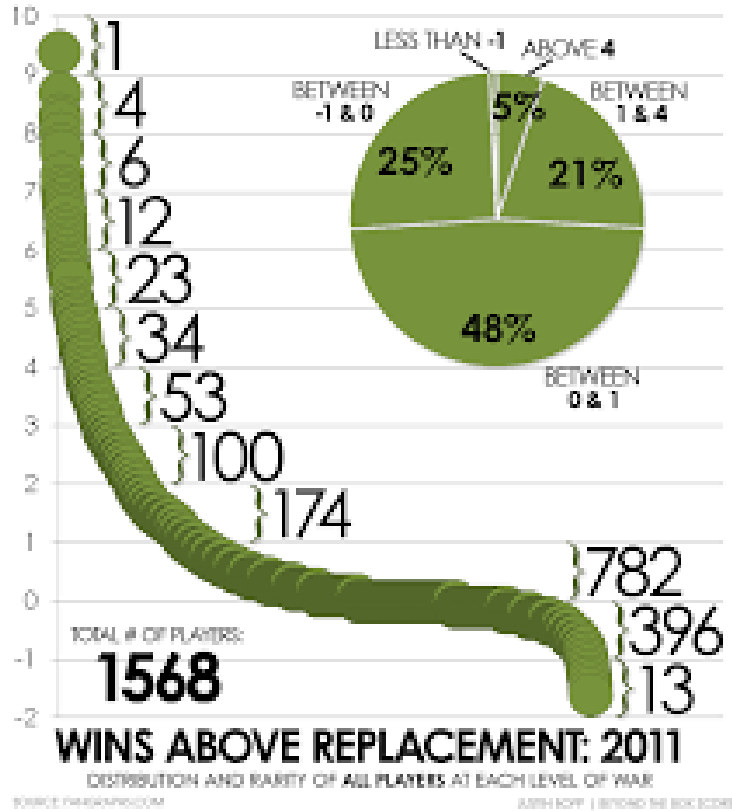
This stat lets you meaningfully compare players across teams, positions, and historical eras.



WAR	Ability
0-1	Replacement level
1-2	Utility player
2-3	Average starter
3-4	Above average starter
4-5	All-Star level player
5-6	Superstar, very well know players
6+	MVP level

Wins Above Replacement (WAR)

WAR	Ability
0-1	Replacement level
1-2	Utility player
2-3	Average starter
3-4	Above average starter
4-5	All-Star level player
5-6	Superstar, very well know players
6+	MVP level



Gold Standards and Proxies

Gold standards are labels or answers we trust to be correct, reflecting the scoring goal.

Proxies are available quantities correlated with what we want to measure.

Your grading average is a proxy for how you should do in my class.

Scores vs. Rankings

Which is more interpretable depends on:

- Will the numbers be presented in isolation?

Simply achieved 1h37m06 in SM64 (10^{th} position)

- What is the distribution of scores?

How much better is #1 than #10?

- Do you care about the middle or extremes?

Small changes in score can cause big rank diffs

Scores vs. Rankings

Last year's slide: Note how Simply's ranking has gotten worse, even though he improved!

Which is more interpretable depends on:

- Will the numbers be presented in isolation?

Simply achieved 1h37m35 in SM64 (7th position)

- What is the distribution of scores?

How much better is #1 than #7?

- Do you care about the middle or extremes?

Small changes in score can cause big rank diffs

Recognizing Good Scoring Functions

- Easily computable
- Easily understandable
- Monotonic interpretation of variables
- Produces satisfying results on outliers
- Use systematically normalized variables

Normalization and Z-scores

It is critical to normalize different variables to make their range/distribution comparable.

Z-scores are computed: $Z_i = \frac{X_i - \bar{X}}{\sigma}$

Z-scores of height measured in inches is the same as height measured in miles.

Your biggest analysis sins will come in using unnormalized variables for analysis!

Z-score Examples

Z-scores have mean 0 and sigma=1.

Thus Z-scores of different variables are of comparable magnitude.

The sign identifies if it is above/below the mean.

$$\mu(B) = 21.9 \quad \sigma(B) = 1.92$$
$$\mu(Z) = 0 \quad \sigma(Z) = 1$$

B	19	22	24	20	23	19	21	24	24	23
Z	-1.51	0.05	1.09	-0.98	0.57	-1.51	-0.46	1.09	1.09	0.57

Advanced Ranking Techniques

Linear combinations of normalized values generally yield reasonable scores, but other techniques include:

- Elo rankings
- Merging rank orderings
- Directed graph orderings
- PageRank

Binary Comparisons

Rankings are often formed by analyzing series of binary comparisons:

- Team A beats team B
- Expert votes for A instead of B
- Student choses university A over B

Vote counts fail to pick the best when different teams face different levels of competition (e.g. 1st vs. 2nd league).

Elo Rankings

After starting equally ranked, scores are then adjusted to reflect the surprise of each match.

$$r'(A) = r(A) + k(S_A - \mu_A)$$

S is the actual score (-1,1) for *A*, with *mu* the expected score from the previous *r(A)* and *r(B)*.

Parameter *k* modulates the maximum possible swing in any one match.

What is the Expected Match Score?

If P(A>B) estimates the probability that A beats B, then:

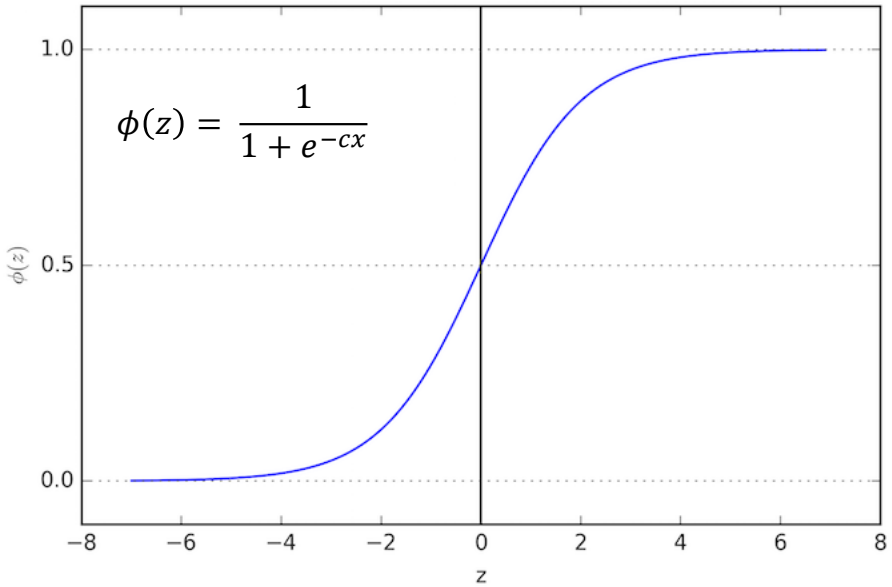
$$\mu_A = 1 \cdot P_{A>B} + (-1) \cdot (1 - P_{A>B})$$

If the ranking system is meaningful, this probability should be a function of the difference between the scores *r(A)* and *r(B).*

The Logistic Function

We need a function f(x) that takes x and yields a probability:

- f(0) = ½
- f(infty) = 1
- f(-infty) = 0



Logistic vs. Logit Function

Error in the Book: The *Data Science Design Manual* does not separate these concepts clearly!

Logistic & logit function are inverse to each other:

Logistic function

(score → probability)

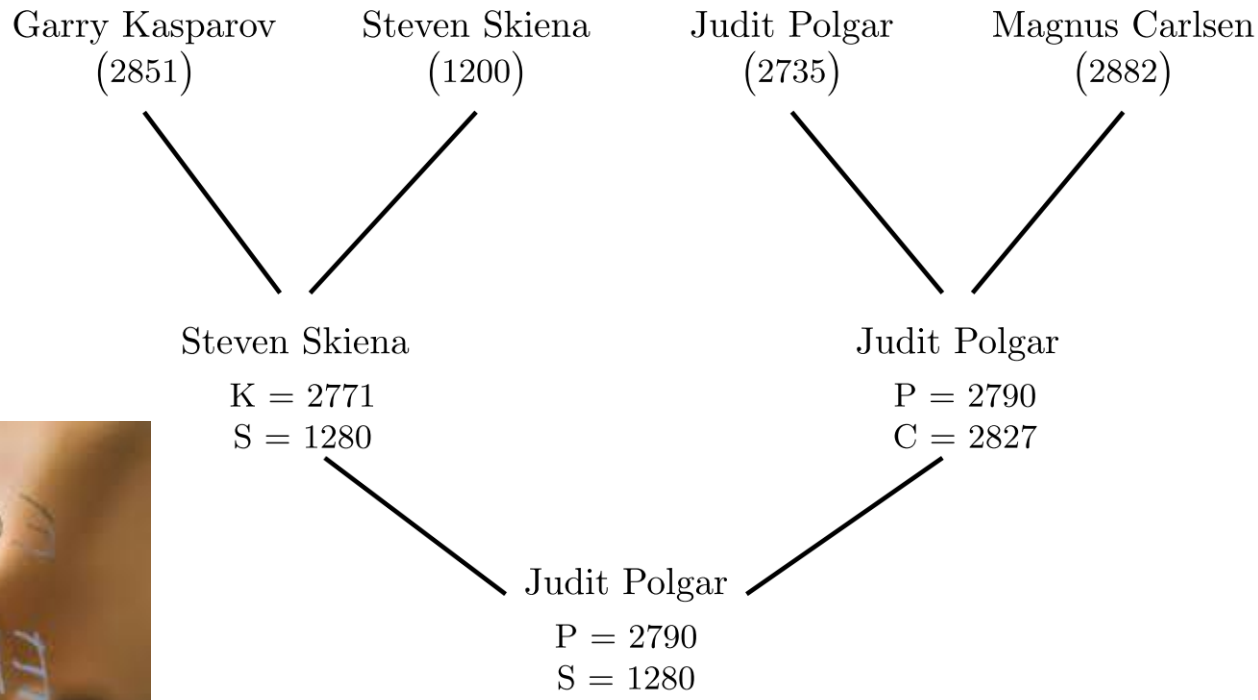
$$logistic(x) = \frac{1}{1 + e^{-x}}$$

Logit function

(probability → score)

$$logit(p) = log\left(\frac{p}{1 - p}\right)$$

Elo Chess Ranking Example



Merging Rankings / Votes

Consider determining the winner of a multiparty election where each voter ranks the candidates in order of preference.

1. UOL 2. TUM 3. UHH 4. ….

Equivalently, consider merging rankings independently drawn on different features.

Borda's Method

By assigning an increasing score per position, the resulting point total ranks the items:

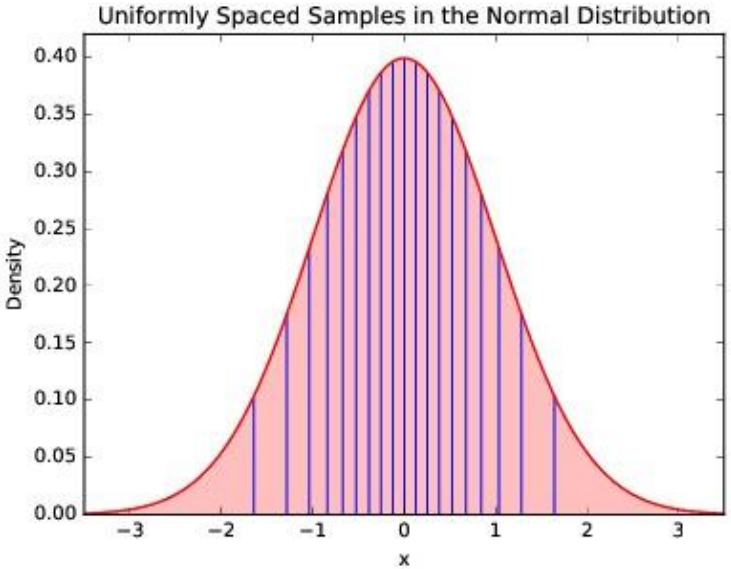
1	A	B	A	A	A:5
2	C	A	B	B	B:8
3	B	C	C	D	→ C:12
4	D	D	E	C	D:16
5	E	E	D	E	E:19

Four voters, each ranking five items.

Weights for Borda's Method

Linear position weights make sense when we have equal confidence across all positions.

But we presumably trust our distinctions among the best/worst more than the middle elements, suggesting normally distributed weights.

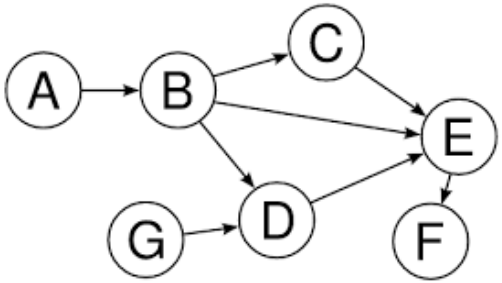


Directed Graph Orderings

Treating the vote (A>B) as an edge (A,B) yields a directed graph.

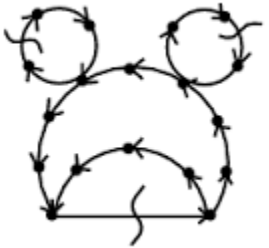
If there are no inconsistencies, we get a directed acyclic graph (DAG).

Topologically sorting this DAG gives a reasonable order, like ABCGDEF or GABCDEF



Ranking General Digraphs

For general directed graphs, we seek the order minimizing the number of “wrong way’’ edges.



Cutting the minimum number of edges to leave a DAG is NP-complete.

But reasonable heuristics start by sorting by the difference between in/out degree.

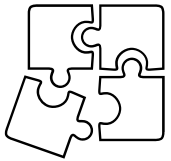
Arrow's Impossibility Theorem

There is no ranking system that satisfies all desirable properties:

- The system should be complete: given A and B it must pick one or say equal preference.
- The system must be transitive, meaning that if A>B and B>C then A>C.
- If every voter prefers A to B, then A wins over B.
- Preferences cannot depend only on one dictator.
- The preference of A to B should be independent of preferences for all other candidates.

Voter	Red	Green	Blue
X	1	2	3
Y	2	3	1
Z	3	1	2

Example: Red beats Green, Green beat Blue, Blue beats Red



Wrapup: Scores & Rankings

- **Scores** reduce the dimensionality of data
- **Rankings** represent orderings of items
- Scores and rankings are not fail-safe, but generally useful
- Examples include the Body-Mass Index (score) and PageRank (ranking)