DSA/ISE 5113

Advanced Analytics and Metaheuristics

Charles Nicholson, Ph.D.
Associate Professor
School of Industrial & Systems Engineering
Data Science and Analytics Institute
Gallogly College of Engineering
University of Oklahoma

Lecture 1

INTRO TO COURSE

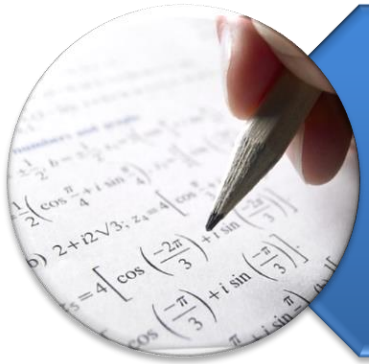
Basic Information

Instructor	**Charles Nicholson, Ph.D.**	
Email	cnicholson@ou.edu	
Office	Zoom	
Office Hours	M:9:30-11a; W:11a-12p	
TAs	Rafia Bushra rafia37@ou.edu	Paxton Leaf jpleaf21@ou.edu
Office hours	TBD	TBD
Software	AMPL www.ampl.com Python www.python.org IDE www.wingware.com	

Textbooks

	Textbooks and links
HTSI	Michalewicz, Z. and D. Fogel. 2004. How to Solve It: Modern Heuristics. Springer http://link.springer.com/book/10.1007%2F978-3-662-07807-5
PO	Chinneck, J. Practical Optimization: A Gentle Introduction. Carleton University: Canada, 2016. http://www.sce.carleton.ca/faculty/chinneck/po.html
AMP	Bradley, S. P., Hax, A.C., Magnanti, T.L. 1977. Applied Mathematical Programming. Addison-Wesley. Out of print, but available here: http://web.mit.edu/15.053/www/AMP.htm)
AMPL	Fourer, R., D. Gay, and B. Kernighan. 2002. AMPL: A Modeling Language for Mathematical Programming. Cengage Learning, 2nd Edition. http://ampl.com/resources/the-ampl-book/
HM	Gendreau, M. and Potvin, J-Y. (Eds.). 2010. Handbook of Metaheuristics. Springer. http://link.springer.com/book/10.1007%2F978-1-4419-1665-5
EAMOP	Coello, C., G. Lamont, and D. Van Veldhuizen. Evolutionary Algorithms for Solving Multi-objective Problems. Second Edition. Barcelona, Spain: Springer Science Business Media, LLC http://link.springer.com/book/10.1007%2F978-0-387-36797-2

Grades



Homework……60%

(6 assignments)

Exams……25%

(1 mid-term, 1 final)

Participation*……15%

(attendance, discussions, quizzes)

Grading Policy

- All assignments must be uploaded to course website by the due date and time indicated.
- Ensure your familiarity with the system in advance and allot sufficient time for assignment submission *and any technical difficulties that may arise*.
- Assignments should be submitted in advance of an anticipated absence.
- ***Late assignments will be penalized.***

Submission is:	Maximum possible grade:
On time + up to 10 min late	100%
Up to 12 hours late	80%
12 to 24 hours late	60%
More than 24 hours late	0%

Slack Day

- Each student will be allocated *one* slack day per semester.
- A slack day is a 24-hour extension for a homework assignment (not applicable to project/quiz/exam/etc.)
- The slack day may be invoked only once per student, but it may be used as the student sees fit. To invoke the slack day provide a comment during the submission of the homework on Canvas.
- Slack days are not retroactive.
- For team assignments, a slack day value = 24 hours / # of students on the team.

Cheating, plagiarism, or any act of dishonesty will NOT be tolerated.

- Incorrect or missing citations is plagiarism...
- Do not copy and paste large swathes of information/code from another source!
- Do not copy work from a previous semester.
- I am interested in *your contribution* and *your success*…
- Know the requirements of the Academic Misconduct Code http://integrity.ou.edu/students_guide.html

python™

GUROBI
OPTIMIZATION

Which of the following are valid ways of contacting the professor?
(select all that apply)

A. **Emailing him at cnicholson@ou.edu**
B. **Zoom virtual office hours**
C. **Quietly communicating with him using only mental telepathy**

Which of the following are valid ways of contacting the professor?
(select all that apply)

- ✓ **Emailing him at cnicholson@ou.edu**
- ✓ **Zoom virtual office hours**
- X ~~**Quietly communicating with him using only mental telepathy**~~

If your assignment was several minutes late due to technical difficulties while submitting your work, this is perfectly okay, everyone understands and appreciates these issues: no one expects you to attempt to submit your assignment more than a few minutes prior to the deadline anyway.

- **True**
- **False**

If your assignment was several minutes late due to technical difficulties while submitting your work, this is perfectly okay, everyone understands and appreciates these issues: no one expects you to attempt to submit your assignment more than a few minutes prior to the deadline anyway.

- ~~True~~
- ✓ **False**

An assignment that is submitted 5 hours late will incur the following:

A. **No penalty – after submitting late, I'll just write an email to the prof and explain why I was late… I'm sure I am special enough as an individual that whatever reason that I give will be something that he has not heard a million times before and therefore, effectively, no penalty.**
B. **100% penalty**
C. **20% penalty**

An assignment that is submitted 5 hours late will incur the following:

- ~~✗ No penalty – after submitting late, I'll just write an email to the prof and explain why I was late… I'm sure I am special enough as an individual that whatever reason that I give will be something that he has not heard a million times before and therefore, effectively, no penalty.~~
- ~~✗ 100% penalty~~
- **✓ 20% penalty**

Discussion Question:
How to invoke a “slack day”?

Participation is graded and accounts for 15% of your total grade.

- **True**
- **False**
- **Well, sort of true and sort of false, I mean, at the end of the course I will ask for (and expect!) all participation points regardless if I actually participated in class, visited the professor or TA during office hours, or contributed at all. So, yes, it's "graded", but you know.**

Participation is graded and accounts for 10% of your total grade.

- **True**
- ~~**False**~~
- ~~**Well, sort of true a… ort c… lse, I mean, at the end of the course I will ask… pect!) all participation points regardless if I a… participated in class, visited the professo… ng office hours, or contributed at all. S… yes, … graded", but you know.**~~

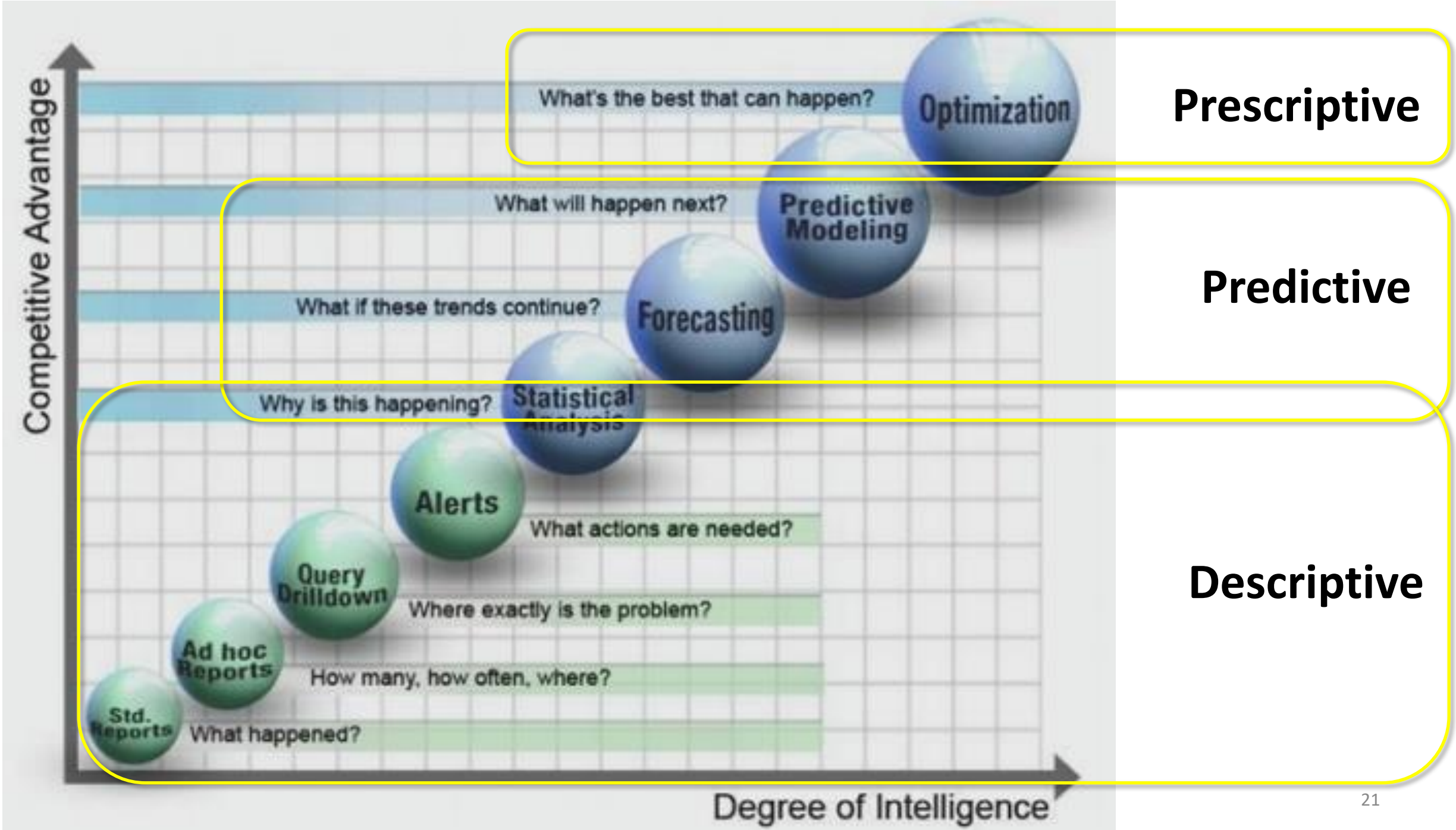
oklahomaAnalytics.com

facebook.com/ou.analytics

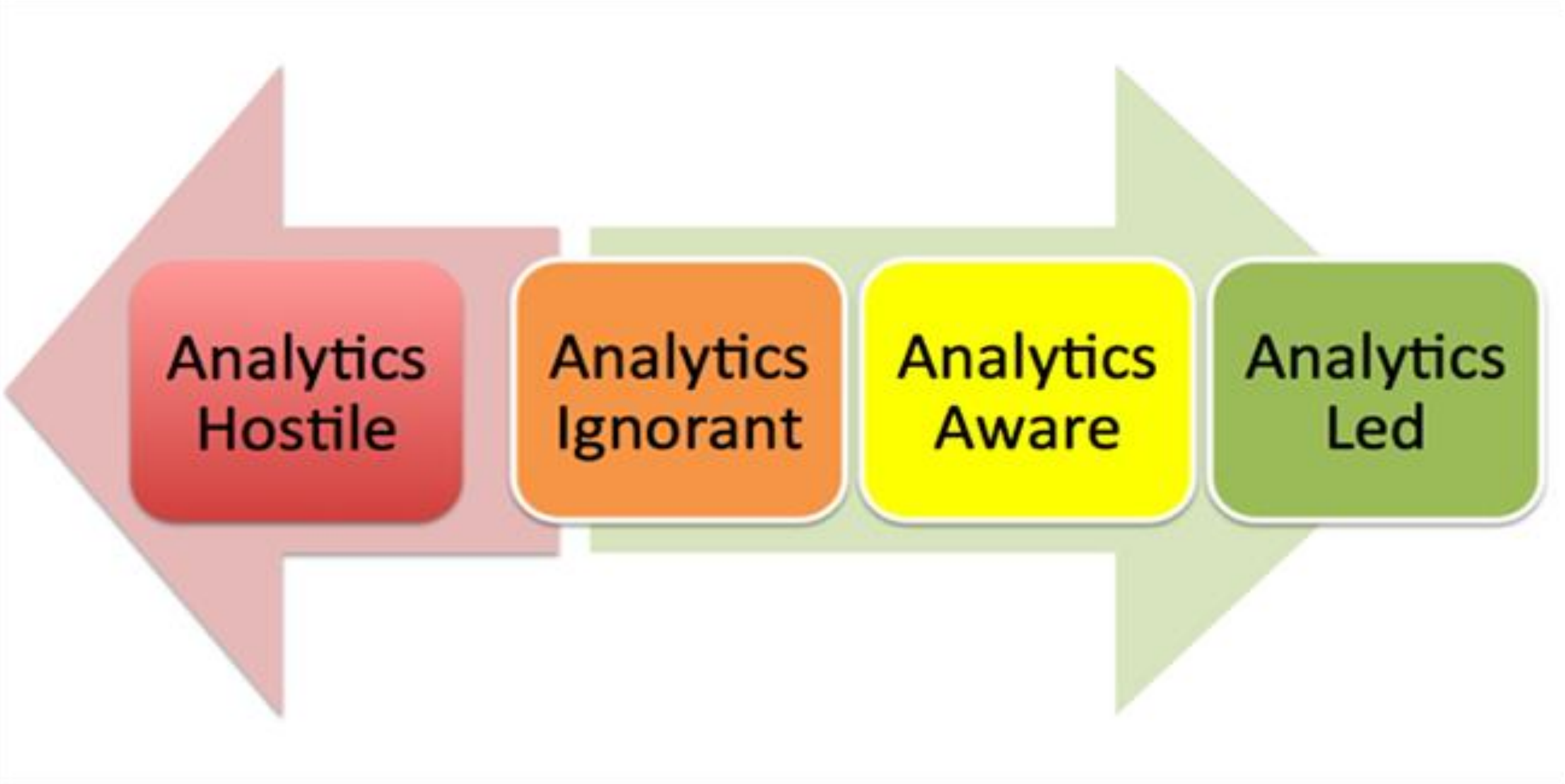
Will "liking" the professors Analytics Lab facebook page and following his blog help my grade?

- ✓ It can't hurt to try!

Advanced Analytics and Metaheuristics


Competitive Advantage
What's the best that can happen?
Optimization
What will happen next?
Predictive Modeling
What if these trends continue?
Forecasting
Why is this happening?
Statistical Analysis
Alerts
What actions are needed?
Query Drilldown
Where exactly is the problem?
Ad hoc Reports
How many, how often, where?
Std. Reports
What happened?
Degree of Intelligence
Prescriptive
Predictive
Descriptive

A word of caution…



Aaron Erickson, January 2014, http://www.thoughtworks.com/insights/blog/analytics-maturity-spectrum

Expand fundamental problem-solving ability and creativity

There are 26 sheep and 10 goats on a ship.

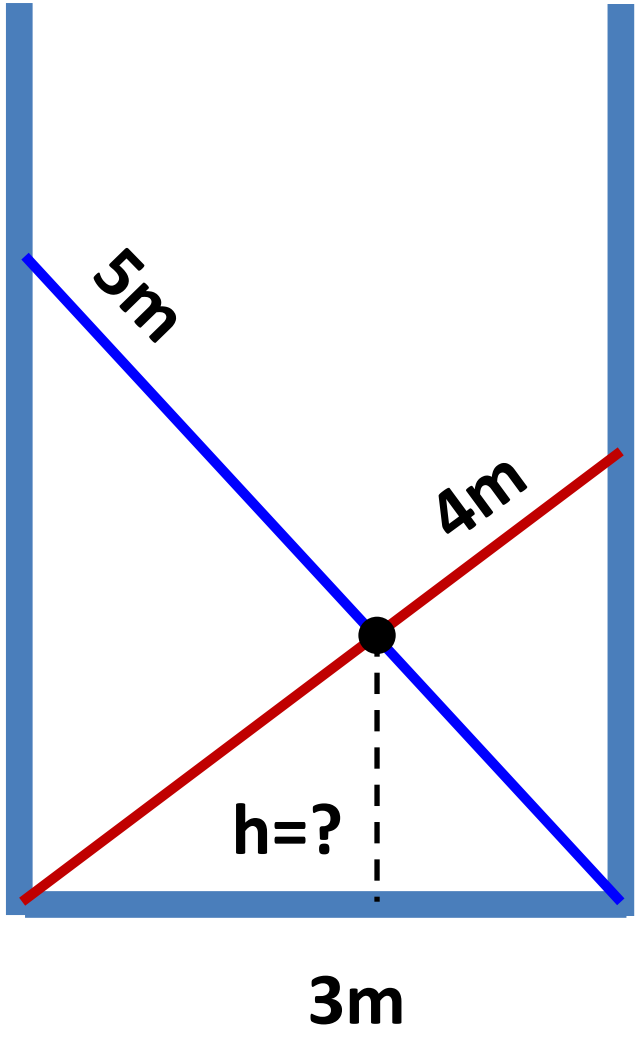
How old is the captain?

An army bus holds 36 soldiers.

If 1,128 soldiers are being bussed to their training site, how many busses are needed?

32	23%
31	18%
31 1/3	29%

Problems are the possession of purpose-driven decision makers


5m
4m
h=?
3m

Why are some problems difficult to solve?

- **large search space**
- **too complex; complications with evaluation**
- **conflicting objectives**
- **ill-defined problem**
- **heavily constrained**
- **dirty data; untrustworthy data**
- **stochastic elements**
- **psychological barrier**

Boolean satisfiability problem (SAT)

Given a boolean expression written using only AND, OR, NOT of variables, determine the true or false variable assignment that will make the entire expression true.

simple SAT problem

$$f(\mathbf{x}) = (x_1 \wedge \neg x_2) \wedge (\neg x_2 \vee x_3)$$

x_1	x_2	x_3	$x_1 \wedge \neg x_2$	$\neg x_2 \vee x_3$	$f(\mathbf{x})$
TRUE	TRUE	TRUE	FALSE	TRUE	FALSE
TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
TRUE	FALSE	TRUE	TRUE	TRUE	TRUE

not so simple SAT problem

What if there were a 100 variable SAT problem?

$$f(\mathrm{x}) = (x_{17} \vee \neg x_3 \vee x_7) \wedge (\neg x_{91} \vee \neg x_6) \wedge \cdots \wedge (x_1 \vee \neg x_{64} \vee x_{11} \vee x_9)$$

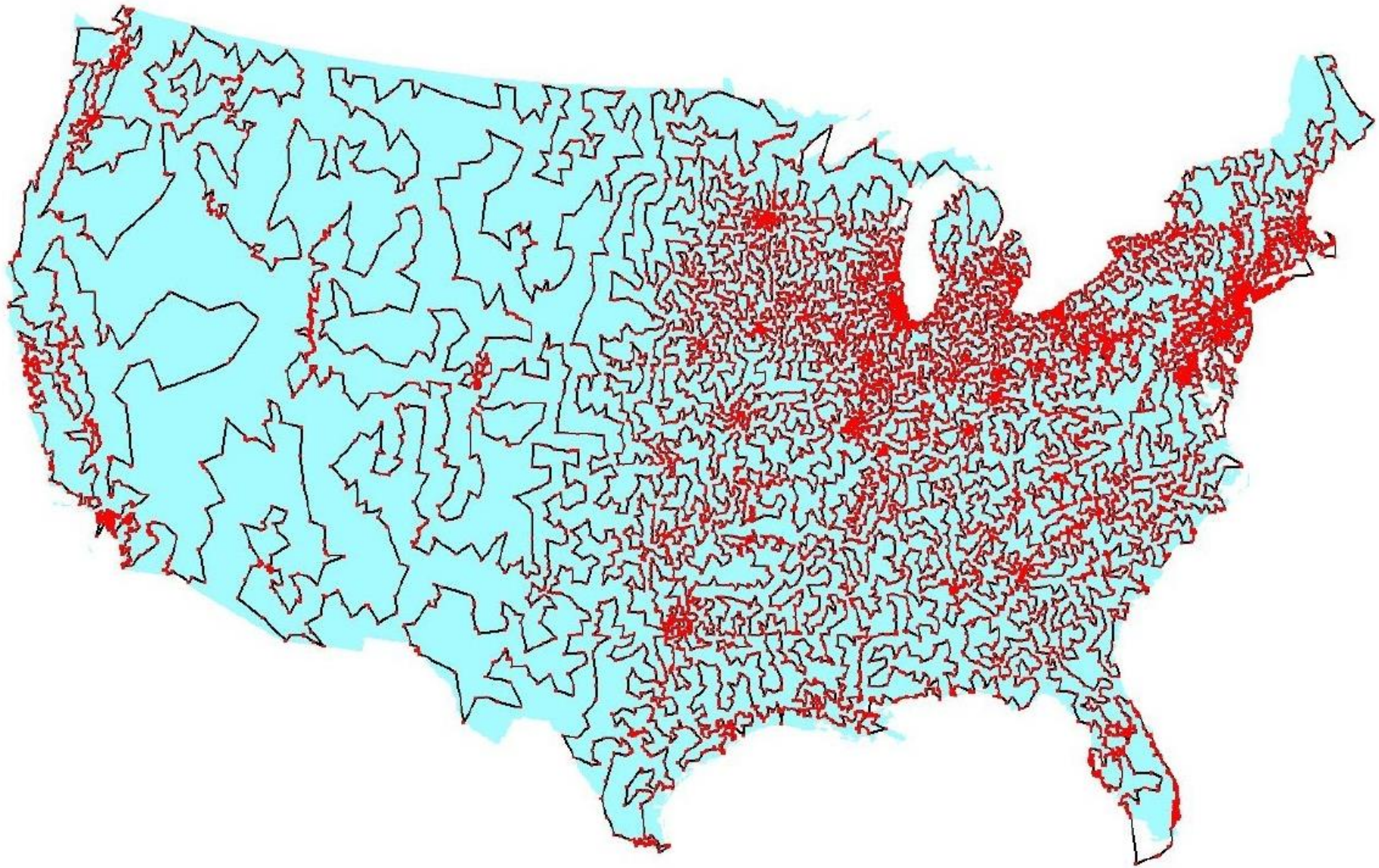
In general, how many possible solutions?

Traveling Salesman problem (TSP)

A traveling salesman must visit every city in his territory exactly once and then return home using a shortest possible route

TSP Solutions

Cities	Possible solutions (approximate)
6	60
10	181,440
20	10,000,000,000,000,000
50	**100,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000**

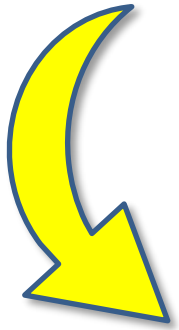


13,509 cities in US TSP solved in 1998

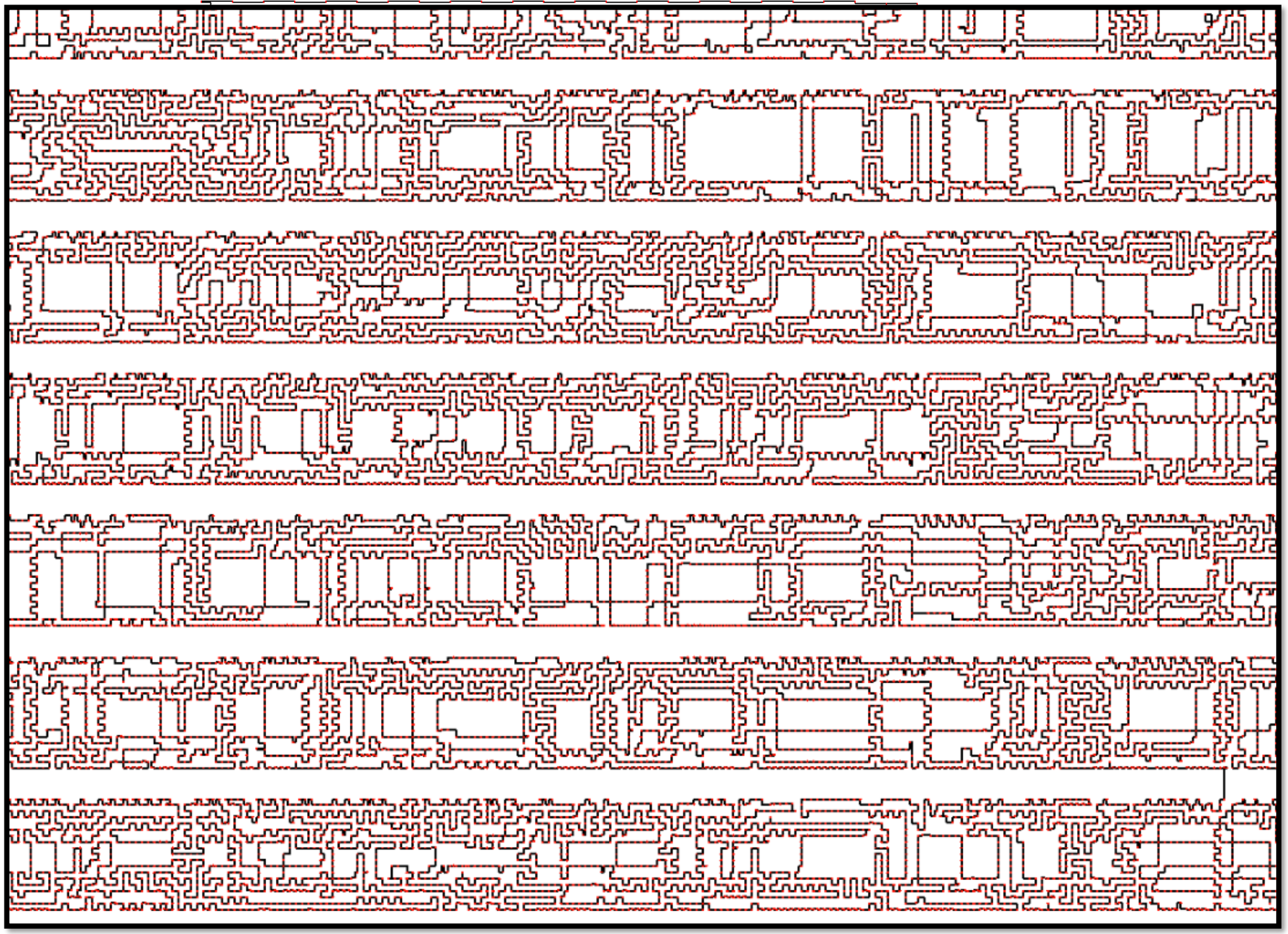
TSP Solutions

- 15,112 cities in Germany solved in 2001
- 24,978 cities in Sweden solved in 2004
- 85,900 locations solved solved in 2006

http://www.math.uwaterloo.ca/tsp/pla85900/



The total amount of computer usage for the computations was approximately **136 CPU years!**



Nonlinear Programming Problem (NLP)

Find $x_1,\ldots,x_n$ so as to

min or *max* $f(x_1,\ldots,x_n)$ (objective function)

subject to $g_i(x_1,\ldots,x_n) \leq b_i$ (functional constraints)

$x_1,\ldots,x_n \in S$ (set constraints)

where at least some of the f and g_i functions are nonlinear.

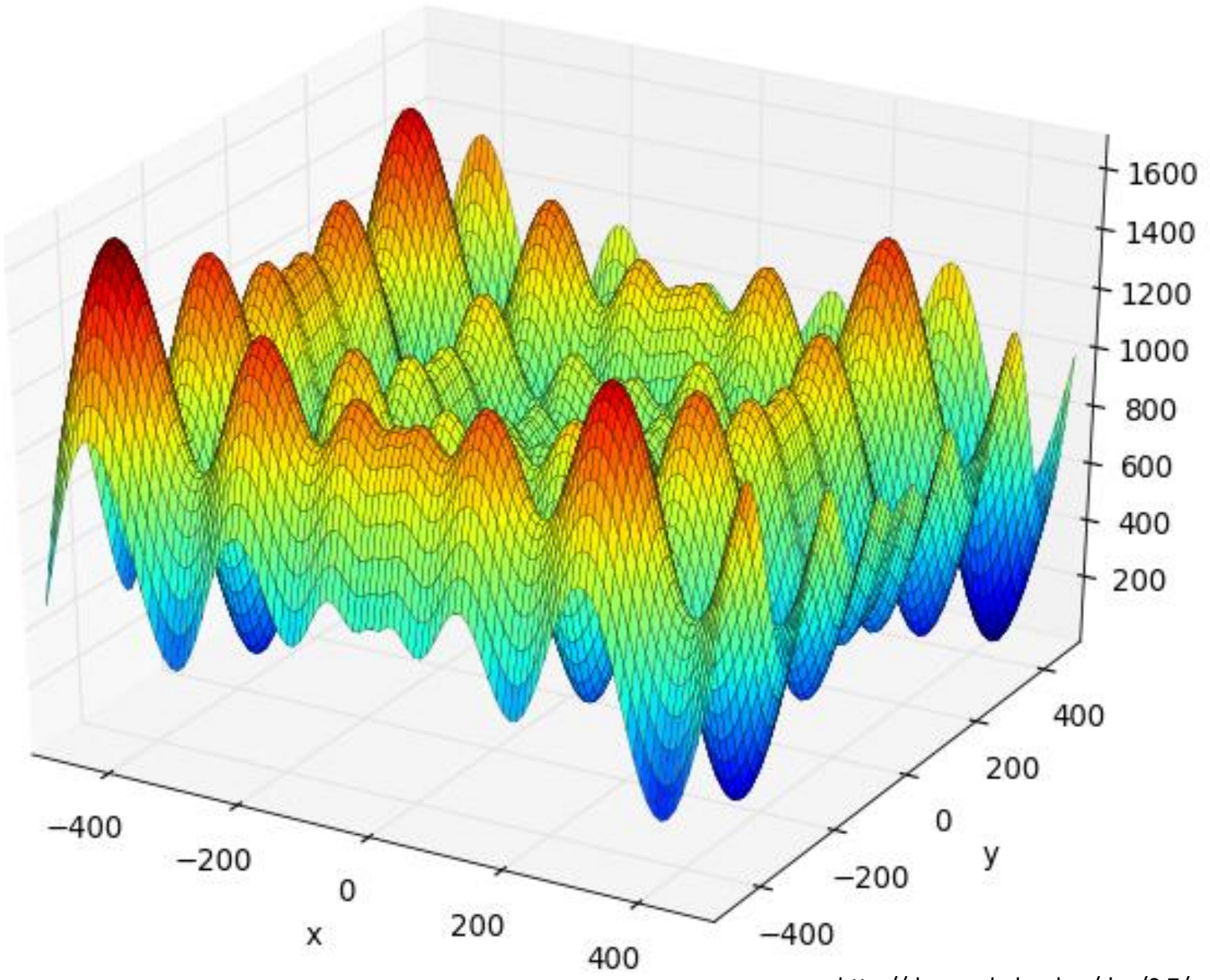
$$\text{maximize } G2(\mathbf{x}) = \left| \frac{\sum_{i=1}^{n} \cos^4(x_i) - 2\prod_{i=1}^{n} \cos^2(x_i)}{\sqrt{\sum_{i=1}^{n} i x_i^2}} \right|$$

$$\prod_{i=1}^{n} x_i \geq 0.75,$$

$$\sum_{i=1}^{n} x_i \leq 7.5n,$$

$$0 \leq x_i \leq 10 \text{ for } 1 \leq i \leq n$$

$$f_{\text{Schwefel}}(\mathbf{x}) = 418.9828872724339 \cdot N - \sum_{i=1}^{N} x_i \sin\left(\sqrt{|x_i|}\right)$$



http://deap.gel.ulaval.ca/doc/0.7/api/benchmarks.html

Acquire information, data, specifications

Visualize, assume, simplify, formulate

Decide on approach, solve

Mr. Smith and his wife invited four other couples over for a party. When everyone arrived, some of the people in the room shook hands with some of the others. Of course, nobody shook hands with their spouse and nobody shook hands with the same person more than once.

After that, Mr. Smith asked everyone how many times they shook someone's hand. He received different answers from everybody.

How many times did Mrs. Smith shake someone's hand?

Any problem worth solving is worth thinking about.