

YOUR LEARNING EXPERIENCE

Christoph Rahmede

DSI Lead Instructor, GA London

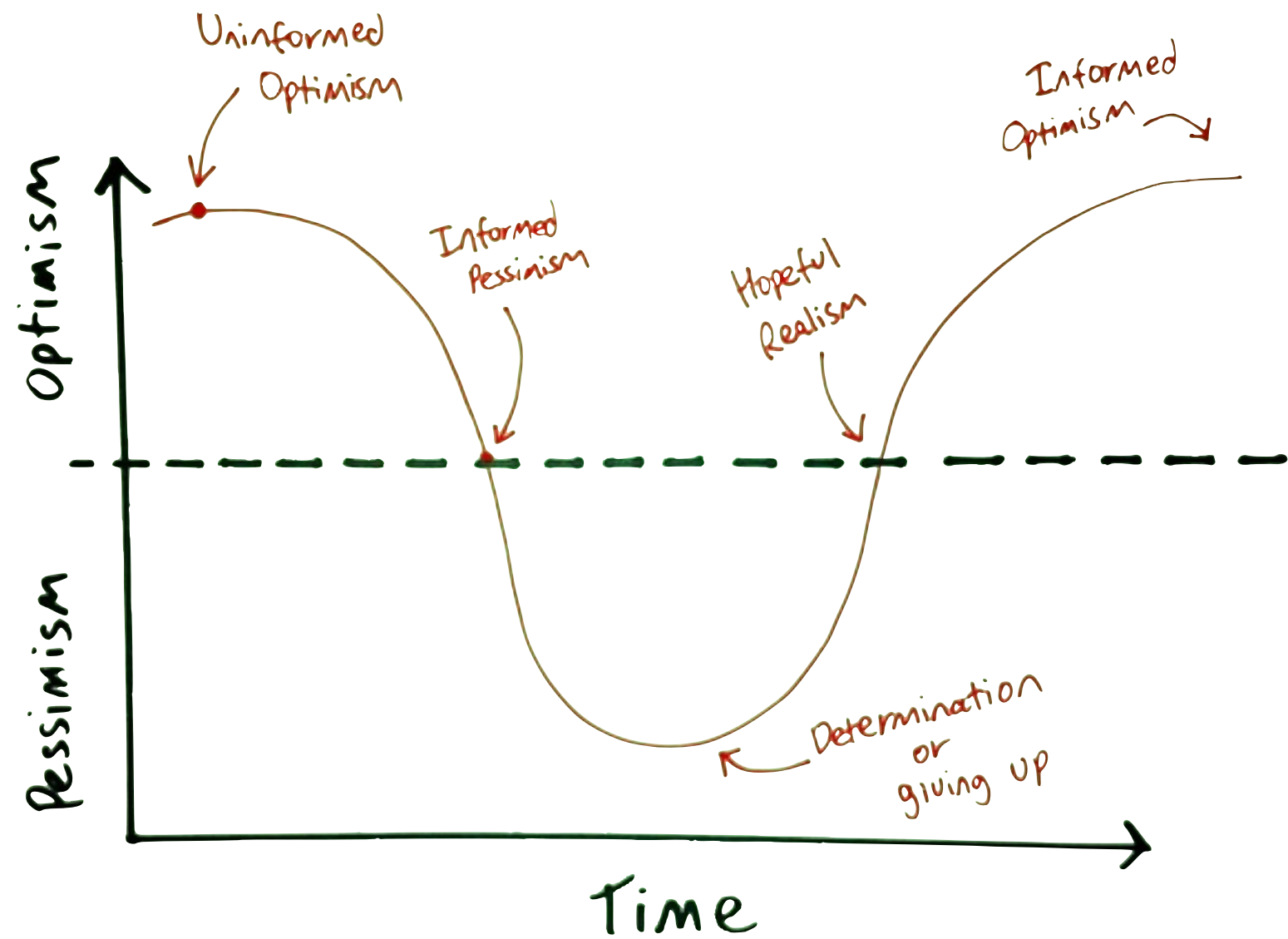
- Formulate course expectations
- Identify the learning environment at GA
- Define ground rules for efficient learning
- Outline course content/projects

EXPECTATIONS

**WHAT ARE YOUR EXPECTATIONS
OF WHAT WE'LL LEARN &
EXPERIENCE DURING THIS
COURSE?**

**HOW MANY OF
YOU HAVE
EVER LEARNED
HOW TO RIDE
A BICYCLE?**





**NO BELL CURVE, WE EXPECT
EVERYONE TO GRADUATE!**



LEARNING AT GA

LEARN
PRACTICE
SHARE
APPLY
REFLECT

**YOU HAVE THE
FREEDOM TO FAIL
EXPERIMENT
MAKE MISTAKES
TRY NEW THINGS**

- Collaborate and help each other
- We are here to help you! Please ask.

GROUND RULES

THINK OF A TIME YOU HAD A **NEGATIVE LEARNING EXPERIENCE.**

LOTS OF DISTRACTIONS. NO ONE PARTICIPATED. IT WASN'T FUN. PEOPLE KEPT SHOWING UP LATE. PEOPLE DIDN'T RESPECT OTHER PEOPLE'S CONTRIBUTIONS. I DIDN'T FEEL COMFORTABLE MAKING A MISTAKE. PEOPLE INTERRUPTED EACH OTHER. NO ONE WAS LISTENING. NOT ENOUGH QUESTIONS.

WHAT GROUND RULES WOULD YOU LIKE TO SET TO MAKE THIS AN **AMAZING LEARNING EXPERIENCE?**

**THINK OF A TIME YOU HAD A GREAT
LEARNING EXPERIENCE.**

EVERYONE PARTICIPATED. WE HAD FUN. PEOPLE ARRIVED ON TIME. EVERYONE RESPECTED EACH OTHER'S CONTRIBUTIONS. I FELT COMFORTABLE MAKING A MISTAKE. PEOPLE GAVE ROOM FOR EACH OTHER TO CONTRIBUTE. WE LISTENED. WE ASKED QUESTIONS.

**WHAT GROUND RULES WOULD YOU LIKE TO SET TO MAKE
THIS AN AMAZING LEARNING EXPERIENCE?**

OUR GROUND RULES













Q&A

THE WORLD OF DATA SCIENCE

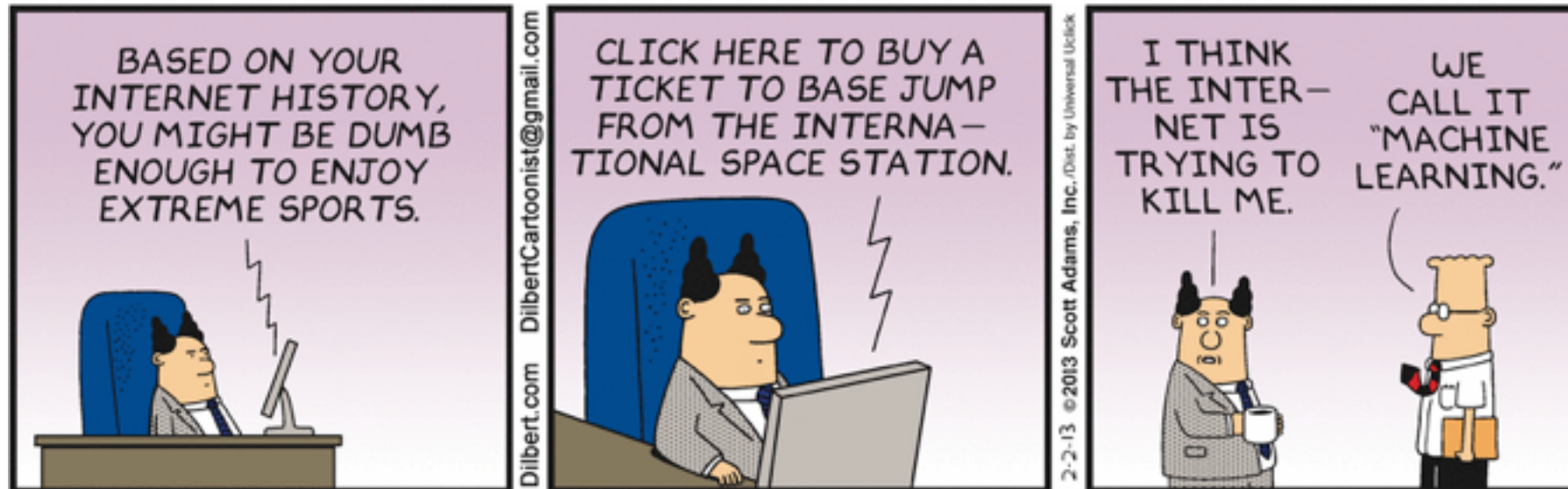
**INFORMATION IS THE OIL OF THE 21ST
CENTURY, AND ANALYTICS IS ITS
COMBUSTION ENGINE.**

PETER SONDERGAARD, SVP, GARDNER RESEARCH

Is exponentially increasing in size and complexity:

- Velocity: 2.5 quintillion bytes of data created daily.
- Variety: 80% is unstructured (can't fit in an excel spreadsheet).
- Volume: 90% of data was created in the last 2 years.

Provides computers the ability to learn without being explicitly programmed.



The study of the generalisable extraction of knowledge from data.



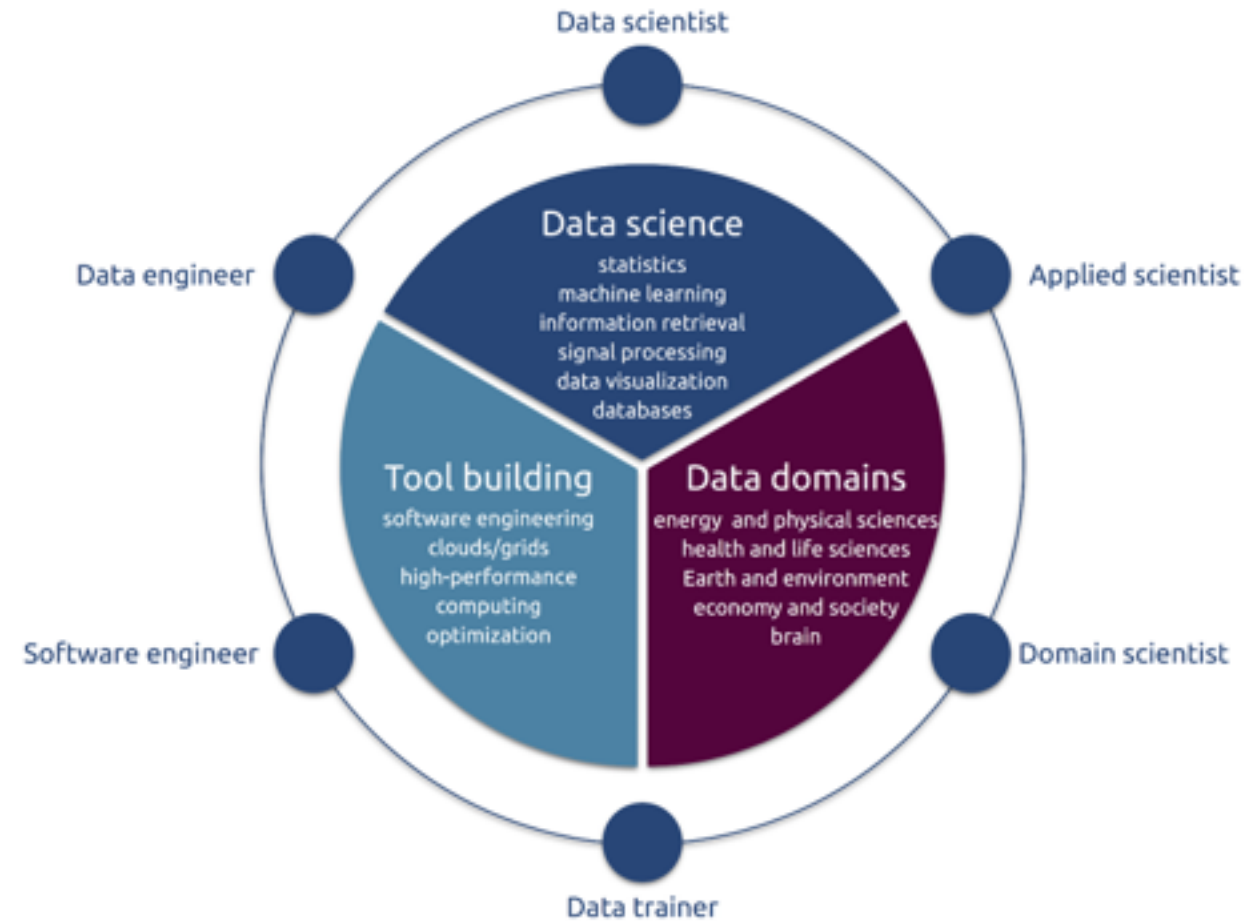
SITS AT THE INTERSECTION OF MANY DISCIPLINES

Data Science



Copyright © 2014 by Steven Geringer Raleigh, NC.
Permission is granted to use, distribute, or modify this image,
provided that this copyright notice remains intact.

- is a hybrid of a statistician and an engineer
- directs business decisions using data
- models data at the most advanced level in a company
- is responsible for clearly communicating data insights
- integrates with nearly every part of a company



COURSE CONTENT



WEEK 1

Python fundamentals, introduction to numpy, matplotlib and pandas

WEEK 2

More pandas, basics of probability and statistics, Exploratory Data Analysis (EDA), working with data, use statistical analysis and visualisation

WEEK 3

Linear regression, regularisation, bias-variance tradeoff, sklearn, statsmodels

WEEK 4

Classification, model evaluation, model tuning with grid search

WEEK 5

SQL, web scraping, apis

WEEK 6

Decision trees, ensemble methods, random forests, boosting, intro to natural language processing

WEEK 7

Unsupervised learning, clustering algorithms, principal component analysis

WEEK 8

Bayesian modelling, Naive Bayes, LDA, sentiment analysis, pymc3

WEEK 9

Time series analysis, ARIMA models

WEEK 10

Big data tools, spark, hadoop, hive, AWS

WEEK 11

Advanced topics, Flask, Networks and network data, finalising the capstone project

WEEK 12

Advanced topics, neural networks, recommender systems, finalising the capstone project

*Subject to adjustments and updates

COURSE PROJECTS

PROJECT #1: WEEK 1

Python fundamentals, working with lists and dictionaries

PROJECT #2: WEEK 2

Exploratory Data Analysis (EDA), working with data, use statistical analysis and visualisation

PROJECT #3: WEEKS 3 & 4

Modelling, develop a regression model on a complex dataset, involves EDA of the dataset, feature engineering and model tuning

PROJECT #4: WEEKS 5 & 6

Web scraping and modelling, obtain data from the web, explore its properties and develop a classification and/or regression model

CAPSTONE PROJECT: WEEKS 7 TO 12

Scope your own data science project, define its goals, obtain the data, explore the data, develop models as appropriate, tune your models, present the final results

IMPORTANT DATES

No classes on:

- Summer bank holiday: Monday, 27 August (week 10)

TWELVE WEEKS

- 12 weeks – will be intense
- Covers breadth and depth of data science concepts
- Multiple projects including capstone project
- Emphasis on application and real world examples over theory
- Rotating instructor lessons

FEEDBACK

HOW WE STRUCTURE FEEDBACK

- You get feedback throughout your entire projects
- Immediately after your presentation
- 1:1 sessions
- Optional drop-in sessions

DROP-IN SESSIONS & ONE-ON-ONES



**MAKE THE MOST OF
THIS EXPERIENCE.**

DATA SCIENCE: THE SEXIEST JOB OF THE 21ST CENTURY

D.J. PATIL, CHIEF DATA SCIENTIST FOR THE US GOV

Q&A