

---

**DATA SCIENCE**

---

**WELCOME TO GA**

## JESSICA BURNS - EDUCATION PROGRAMS PRODUCER

---

- Slack: [gamelbparttimecourses.slack.com](https://gamelbparttimecourses.slack.com)
- Channel: ds2melb
- Direct Message: Jess Burns
- [jessica.burns@generalassemb.ly](mailto:jessica.burns@generalassemb.ly)



## OUR CLASS

---

- Adrian Ciconte
- Daniel Grimsey
- Eamonn McCallum
- Golnoosh Tajadod
- Ian Deutscher
- Jihye Han
- Liz Menheere
- Madhara Udawela
- Mary Hoang
- Mog Van Golstein
- Ollie McNaught-Reynolds
- Oscar Xu
- Pei Wan Wong
- Simon Crean
- Tam Nguyen

## JEREMIAH MANNINGS - ASSOCIATE INSTRUCTOR (AI)

Jeremiah is a Big Data & Analytics Consultant Engineer with Altis Consulting where he is focused on a wide range of skills across the data landscape, primarily in Python. He has previous experience at IBM in the Watson team delivering Cognitive frameworks and technologies; where he had in-depth exposure to machine learning and data science techniques. Utilising SPSS machine learning modelling integrated with Hadoop and Solr, to produce & lead top tier landmark projects.

Jeremiah has a passion for innovation; ideas, structures and producing valuable insights. He believes constantly evolving intelligent computing will change the landscape of data as we know it, with data science on the forefront of this change.

Jeremiah has a particular love for natural language processing, often building bots in his spare time for fun!

Jeremiah has been involved in running Cognitive hackathons as far as South Korea. As well as design, development and implementation of advanced analytics and enterprise search solutions. Jeremiah is always open for a coffee and chat, and working with talented and like-minded people!



## JEREMIAH MANNINGS - ASSOCIATE INSTRUCTOR (AI)

---

- Slack: [gamelbparttimecourses.slack.com](https://gamelbparttimecourses.slack.com)
- Channel: ds2melb
- Direct Message: @jmannings
- [jerry.mannings@gmail.com](mailto:jerry.mannings@gmail.com)
- On-line: 18:00 - 21:00
- Estimate Off-line response time: 24 hours



## ANGELO KLIN - LEAD INSTRUCTOR

Angelo Klin is a Data Scientist based in Melbourne. He has an education background in statistics, information technology, business and management with individual specialisations in data science and machine learning.

Angelo has worked in Australia and internationally with experience in a wide variety of industries and currently is working with Telstra.

Angelo is particularly proud of his involvement in the creation of the very first electronic tally for a Brazilian election in 1989, which involved accurately aggregating 85 million votes

More recently he has analysed federal court rulings to identify patterns using text analytics and topic analysis.

He loves solving complex problems and thoroughly enjoys the challenge of sourcing the best solution for a problem using a wide variety of tools. Angelo is an avid learner, keen photographer, science and space enthusiast, is trying to learn to play the piano and likes chatting in any one of the three languages he speaks.



## ANGELO KLIN - LEAD INSTRUCTOR

---

- Slack: [gamelbparttimecourses.slack.com](https://gamelbparttimecourses.slack.com)
- Channel: ds2melb
- Direct Message: @angeloklin
- [professional@angeloklin.com](mailto:professional@angeloklin.com)
- [angeloklin@katraanalytics.com.au](mailto:angeloklin@katraanalytics.com.au)



# WELCOME TO DATA SCIENCE

**Angelo Klin**  
Katra Analytics

---

## **LEARNING OBJECTIVES**

---

- What is Data Science
- Uses of Data Science
- What is a Data Scientist and What does a Data Scientist do
- Define the Data Science Process / Workflow

---

**DATA SCIENCE**

---

# PRE-WORK

---

## **PRE-WORK REVIEW**

---

- Conclude the on boarding by
  - Creating accounts
    - GitHub, Slack
  - Installing programs
    - Python, conda/anaconda, git, etc.
  - Attending the pre-requisite courses
    - Python, command line, pandas, etc.

---

**DATA SCIENCE**

---

# WHAT IS DATA SCIENCE?

# WHAT IS DATA SCIENCE?

---

- What is your **definition**?
- What is your **interpretation**?
- What is your **understanding**?
- What is your **view**?
- What is your **take**?



# DATA

---

- **data** |'deɪtə|, noun [mass noun]
- **facts and statistics collected together for reference or analysis:** there is very little data available.
  - the quantities, characters or symbols on which operations are performed by a computer.
  - Philosophy things known or assumed as facts, making the basis of reasoning or calculation.
- **ORIGIN:** mid 17th century (as a term in philosophy): from Latin, plural of *datum*.
- **datum** |'deɪtəm|, noun (plural *data*)
- **a piece of information:** the fact is a datum worth taking into account.
  - an assumption or premise from which inferences may be drawn: this is not a permanent and unchangeable datum. See also *sense datum*.
- **a fixed starting point of a scale or operation:** an accurate datum is formed by which other machining operations can be carried out. See also *ordnance datum*.
- **ORIGIN:** mid 18th century: from Latin, literally ‘something given’, neuter past participle of *dare* ‘give’.

---

# SCIENCE

---

- **science** |'sʌɪəns|, noun [mass noun]
- **the intellectual and practical activity encompassing the systematic study of the structure and behaviour of the physical and natural world through observation and experiment:** the world of science and technology.
- a particular area of science: veterinary science | [count noun] : the agricultural sciences.
- a systematically organised body of knowledge on a particular subject: the science of criminology.
- archaic knowledge of any kind: his rare science and his practical skill.
- **ORIGIN:** Middle English (denoting knowledge): from Old French, from Latin *scientia*, from *scire* ‘know’.

# WHAT IS DATA SCIENCE?

---



WIKIPEDIA  
The Free Encyclopedia

## Data science

From Wikipedia, the free encyclopedia

*Not to be confused with [information science](#).*

**Data science** is an interdisciplinary field about processes and systems to extract [knowledge](#) or insights from [data](#) in various forms, either structured or unstructured,<sup>[1][2]</sup> which is a continuation of some of the data analysis fields such as [statistics](#), [machine learning](#), [data mining](#), and [predictive analytics](#),<sup>[3]</sup> similar to [Knowledge Discovery in Databases \(KDD\)](#).

### Contents [hide]

- [1 Overview](#)
- [2 Data scientist](#)
- [3 History](#)
- [4 Criticism](#)
- [5 Software](#)
- [6 References](#)
- [7 Further reading](#)

---

## SIDE DEFINITIONS

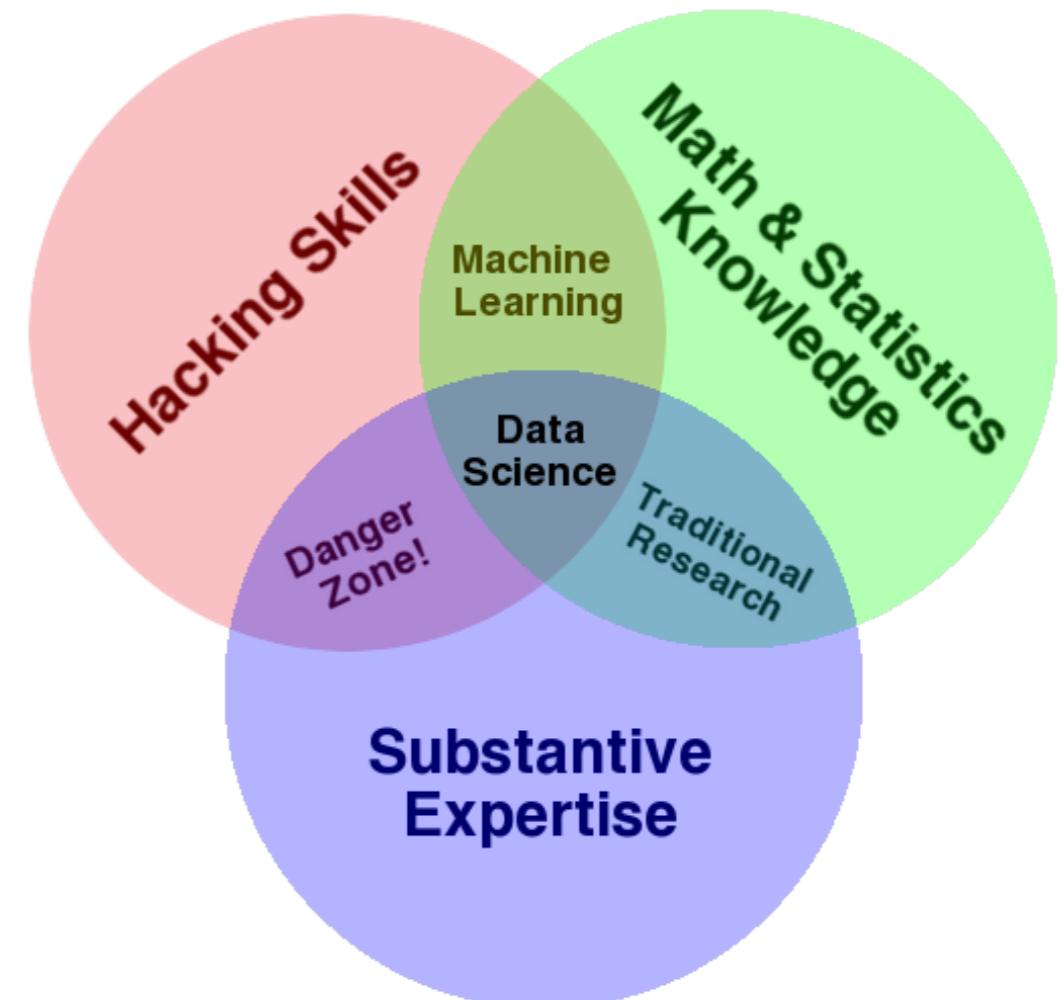
---

- Statistics (discipline)
  - the study of the collection, analysis, interpretation, presentation and organisation of data
- Machine learning
  - subfield of computer science that gives computers the ability to learn without being explicitly programmed. Evolved from pattern recognition and computational learning in artificial intelligence, machine learning explores the study and construction of algorithms that can learn from and make predictions on data.
- Statistic (noun)
  - a fact or piece of data obtained from a study of a large quantity of numerical data

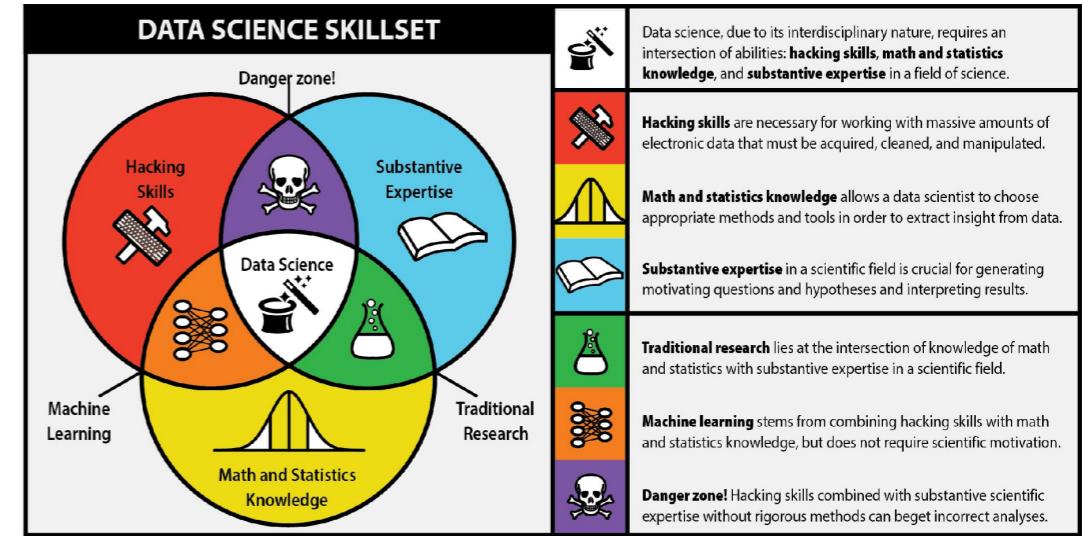
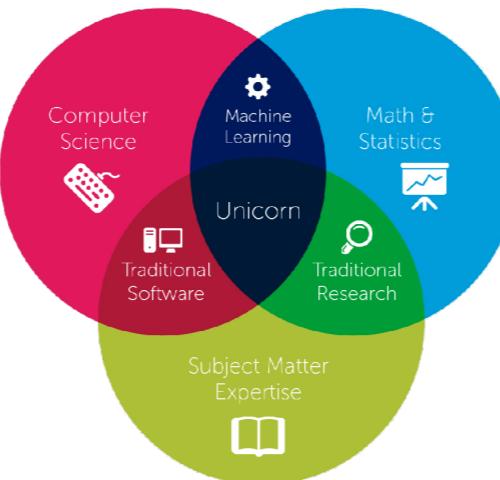
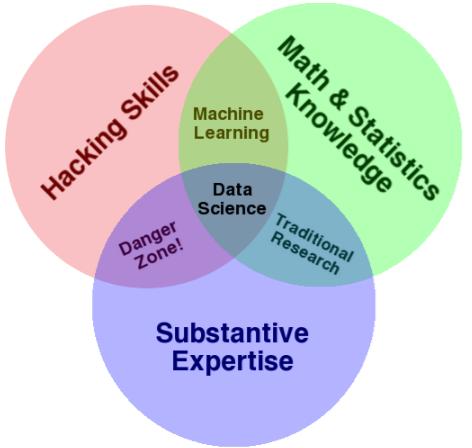
# WHAT IS DATA SCIENCE?

---

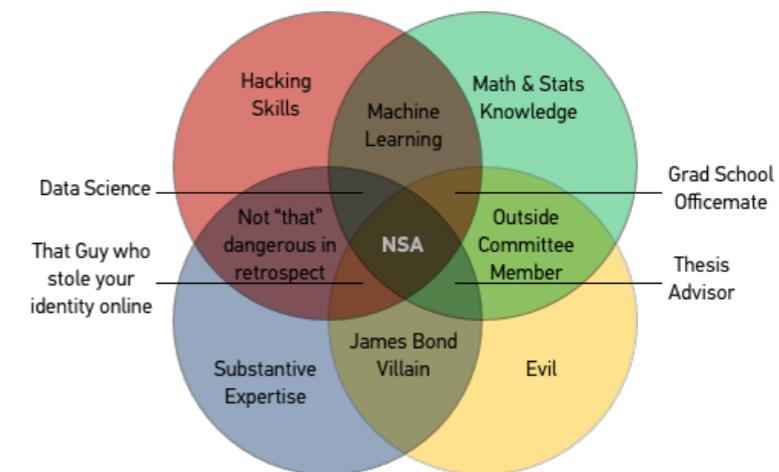
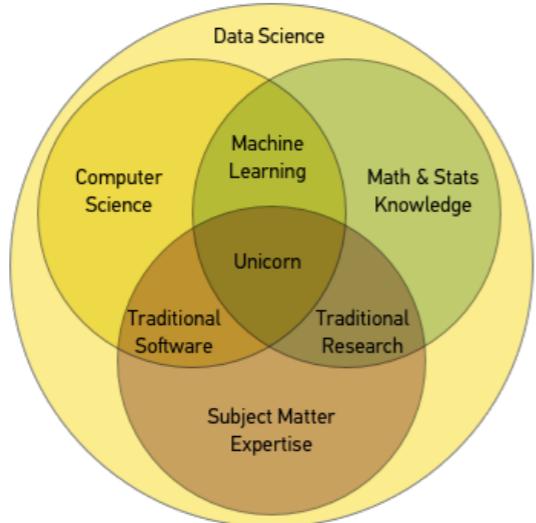
- A set of tools and techniques used to extract useful information from data
- A interdisciplinary, problem-solving oriented subject
- Application of scientific techniques to practical problems



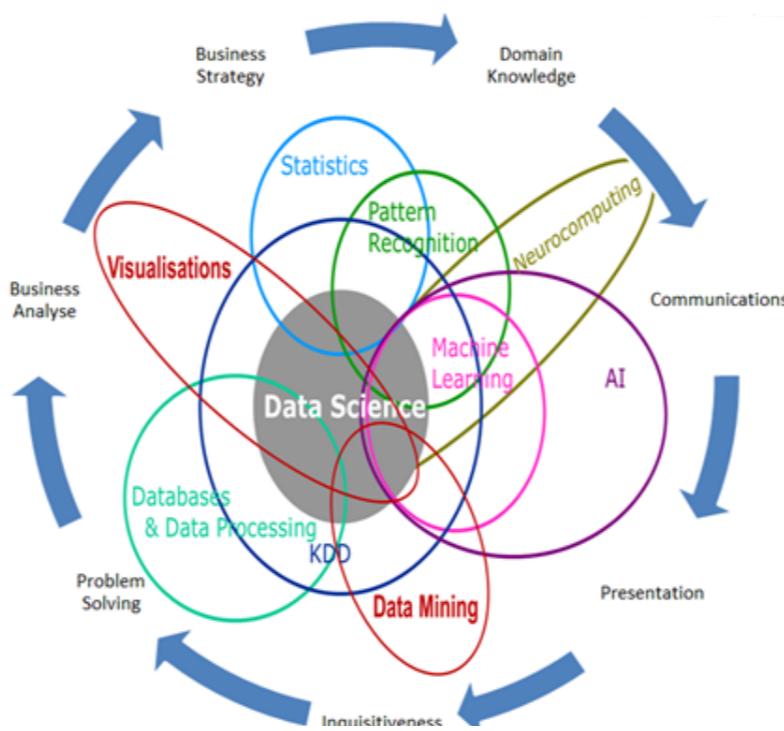
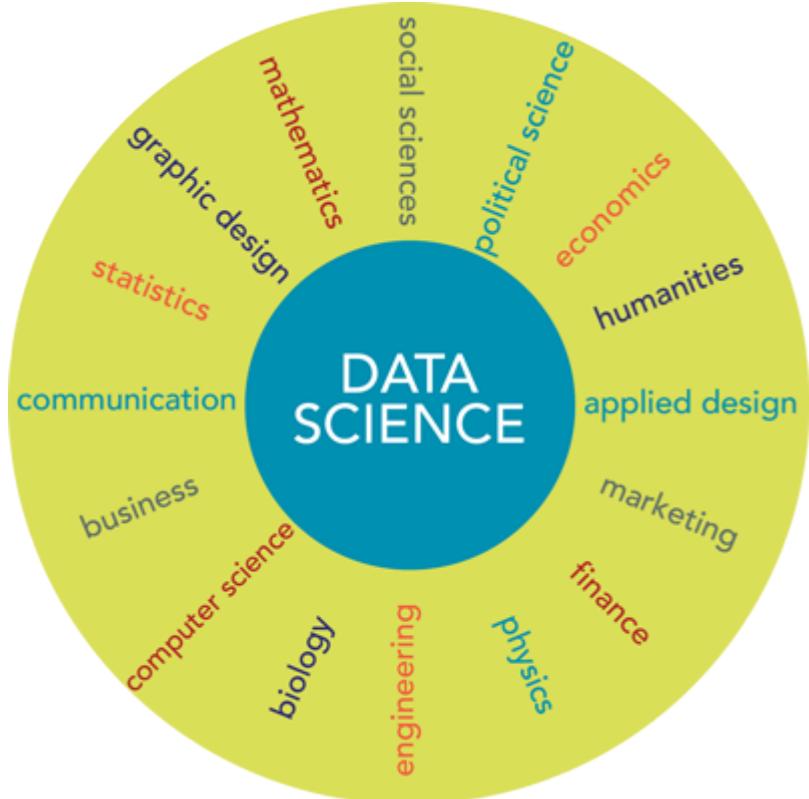
# DATA SCIENCE



Data Science Venn Diagram v2.0



# DATA SCIENCE IS MULTIDISCIPLINARY



---

## WHO USES DATA SCIENCE?

---

**NETFLIX**



🦊 **FiveThirtyEight**

**Google**

**amazon.com®**

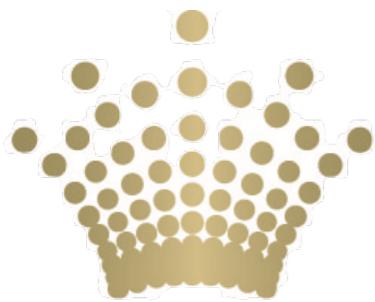


---

## WHO USES DATA SCIENCE? (LET'S BRING IT HOME)

---

coles



CROWN



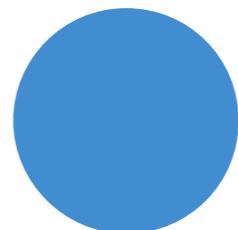
sportsbet  
.com.au

---

## **WHO USES DATA SCIENCE?**

---

- Can you think of others?



---

**DATA SCIENCE**

---

# **USES OF DATA SCIENCE**

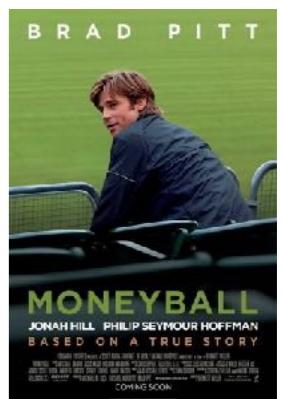
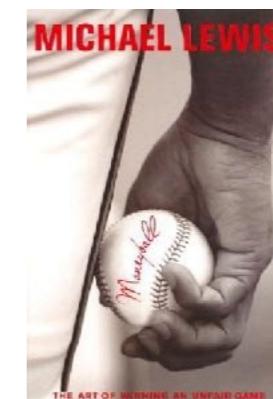
# MONEYBALL

---

## ● Synopsis (from the Book)

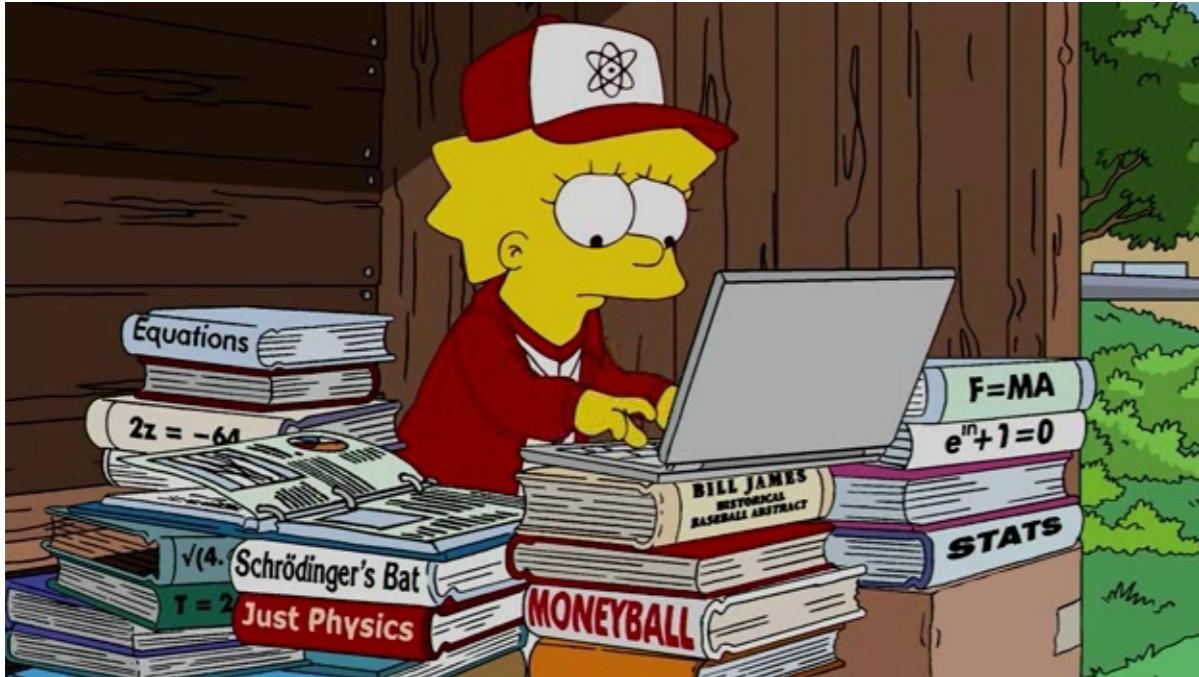
- The central premise of Moneyball is that the collective wisdom of baseball insiders (including players, managers, coaches, scouts, and the front office) over the past century is subjective and often flawed. Statistics such as stolen bases, runs batted in, and batting average, typically used to gauge players, are relics of a 19th-century view of the game and the statistics available at that time. The book argues that the Oakland A's front office took advantage of more analytical gauges of player performance to field a team that could better compete against richer competitors in Major League Baseball (MLB).
- Rigorous statistical analysis had demonstrated that on-base percentage and slugging percentage are better indicators of offensive success, and the A's became convinced that these qualities were cheaper to obtain on the open market than more historically valued qualities such as speed and contact. These observations often flew in the face of conventional baseball wisdom and the beliefs of many baseball scouts and executives.

- <https://en.wikipedia.org/wiki/Moneyball>
- [https://en.wikipedia.org/wiki/Moneyball\\_\(film\)](https://en.wikipedia.org/wiki/Moneyball_(film))



# MONEYBART

---



- <https://en.wikipedia.org/wiki/MoneyBart>
- <http://simpsons.wikia.com/wiki/MoneyBART>

---

## EXAMPLES

---

- Customer/employee churn
- Customer segmentation
- Digital Advertisements (Targeted Advertising)
- Recommendations and Up selling
- Fraud and Risk Detection
- Dynamic pricing, Price Comparison Websites
- Image Recognition, Speech Recognition
- Internet Search
- Gaming
- To come: Self Driving Cars, Robots

---

## **IS THIS A OR B, TRUE OR FALSE, YES OR NO? (BINARY CLASSIFICATION)**

---

- Two-class (or binary) classification for any question that has just two possible answers
  
- Will this customer renew their subscription?
- Is this an image of a cat or a dog?
- Will this customer click on the top link?
- Will this tire fail in the next thousand kilometres?
- Does the \$5 coupon or the 25% off coupon result in more return customers?

---

## **IS THIS A OR B OR C OR ...? (CLASSIFICATION)**

---

- Multi-class classification for questions that have several possible answers
- Which animal is in this image?
- Which aircraft is causing this radar signature?
- What is the topic of this news article?
- What is the mood of this tweet?
- Who is the speaker in this recording?

---

## **HOW MUCH / HOW MANY? (REGRESSION)**

---

- Regression is used for when the answer is a number; can have lots of decimal places or even be negative
- What will the temperature be next Tuesday?
- How many kilowatts will be demanded from a wind farm 30 minutes from now?
- How many new followers will be in the next week?
- Out of a thousand units, how many of this model of bearings will survive 10,000 hours of use?

---

## IS THIS UNUSUAL?

---

- Anomaly detection to identify data points that are not normal.  
Particularly useful when looking rare occurrences.
- Is this pressure reading unusual?
- Is this internet message typical?
- Is this combination of purchases very different from what this customer has made in the past?
- Are these voltages normal for this season and time of day?

---

**DATA SCIENCE**

---

# WHAT IS A DATA SCIENTIST?

# WHAT IS A DATA SCIENTIST?

---



WIKIPEDIA  
The Free Encyclopedia

## Data scientist [\[ edit \]](#)

---

Data scientists use their data and analytical ability to find and interpret rich data sources; manage large amounts of data despite hardware, software, and bandwidth constraints; merge data sources; ensure consistency of datasets; create visualizations to aid in understanding data; build mathematical models using the data; and present and communicate the data insights/findings. They are often expected to produce answers in days rather than months, work by exploratory analysis and rapid iteration, and to produce and present results with dashboards (displays of current values) rather than papers/reports, as statisticians normally do.<sup>[6]</sup>

"Data Scientist" has become a popular occupation with [Harvard Business Review](#) dubbing it "The Sexiest Job of the 21st Century"<sup>[7]</sup> and [McKinsey & Company](#) projecting a global excess demand of 1.5 million new data scientists.<sup>[8]</sup> Universities are offering masters courses in data science.<sup>[9]</sup> Shorter private bootcamps are also offering data science certificates including student-paid programs like [General Assembly](#) to employer-paid programs like [The Data Incubator](#).<sup>[10]</sup>

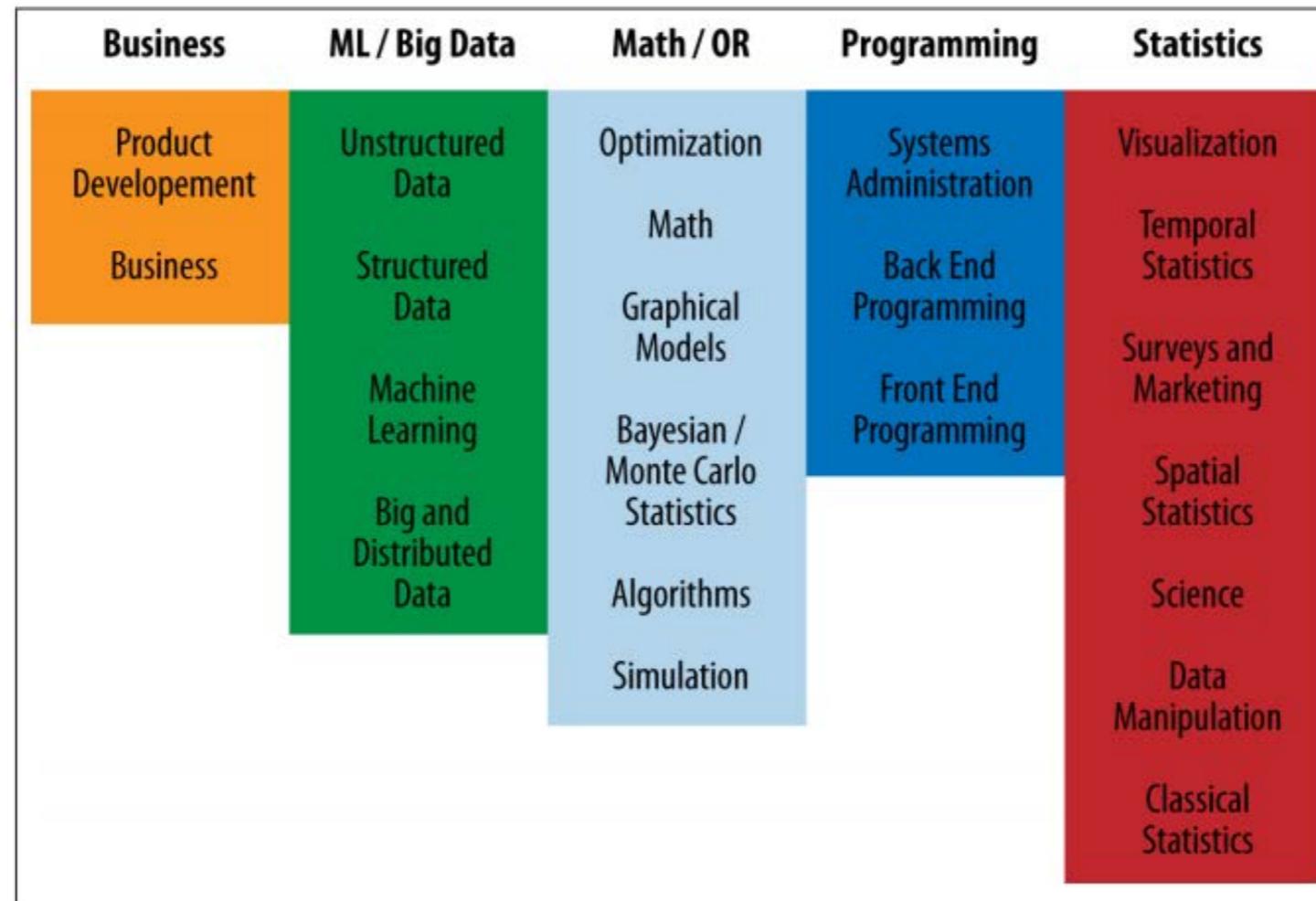
# WHAT ARE THE ROLES IN DATA SCIENCE?

- Data Science involves a variety of **roles**, not just one

Data Developer	Developer	Engineer	
Data Researcher	Researcher	Scientist	Statistician
Data Creative	Jack of All Trades	Artist	Hacker
Data Businessperson	Leader	Businessperson	Entrepreneur

# WHAT ARE THE ROLES IN DATA SCIENCE?

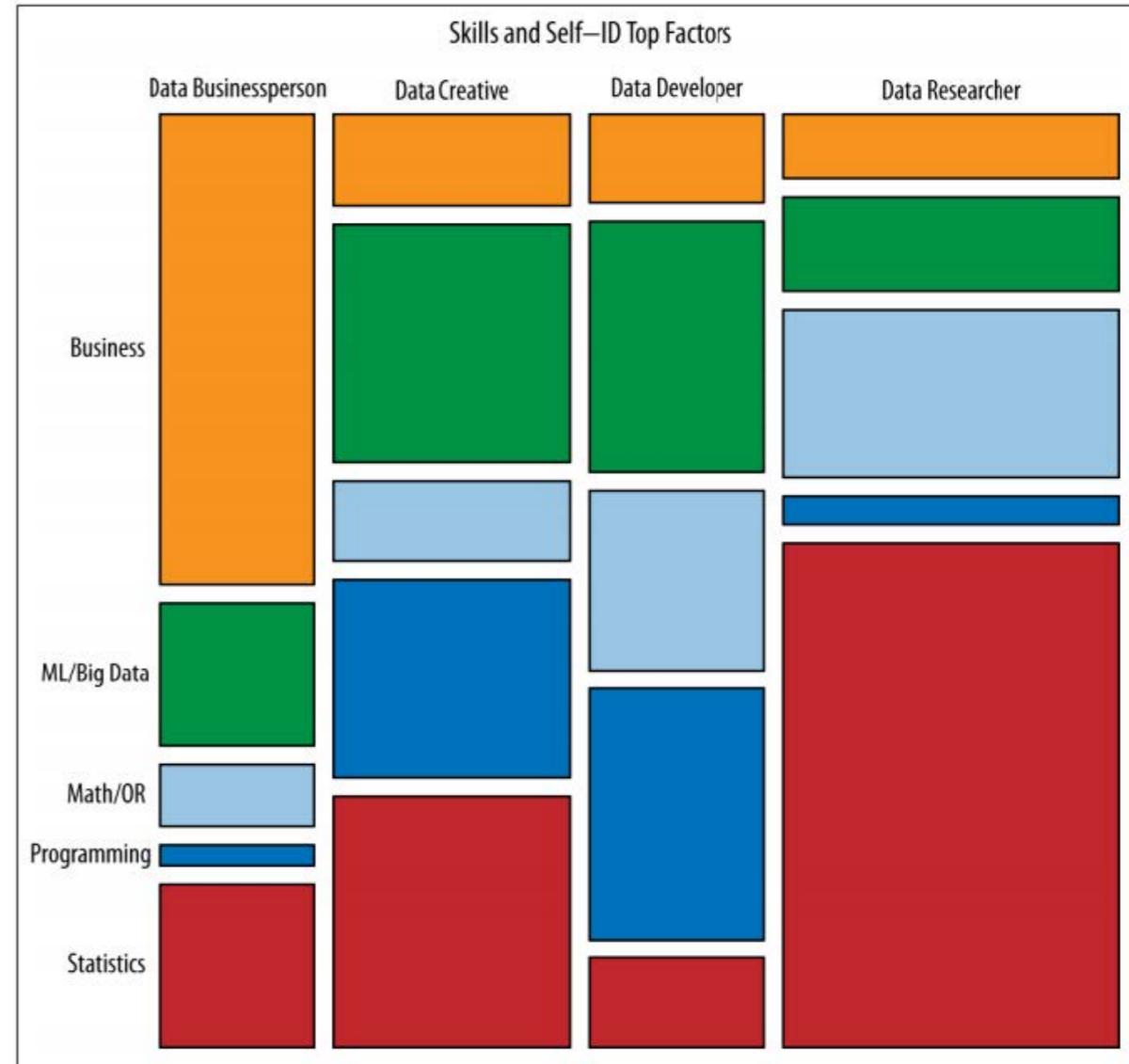
- Data Science involves a variety of **skill sets**, not just one



- Analyzing the Analyzers

# WHAT ARE THE ROLES IN DATA SCIENCE?

- These roles prioritise different skill sets
- However, all roles involve some part of each skill set
- Where are your strengths and weaknesses?



---

**DATA SCIENCE**

---

# **WHAT DOES A DATA SCIENTIST DO?**

---

# WHAT DOES A DATA SCIENTIST DO?

---



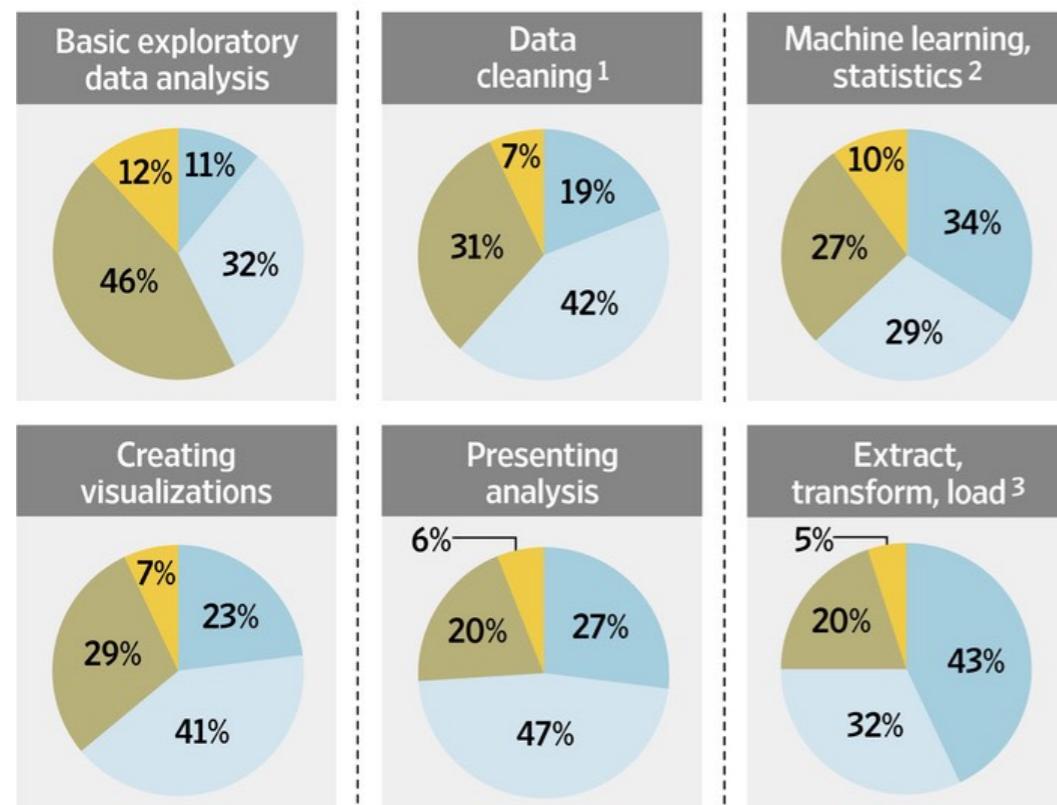
CLIOCRESSWELL

# WHAT DOES A DATA SCIENTIST DO?

## Where Does the Time Go?

The amount of time spent on various tasks by surveyed nonmanagers in data-science positions

Less than 1 hour a week    1 to 3 hours a day  
1 to 4 hours a week    4 or more hours a day



<sup>1</sup> Correcting or removing faulty data    <sup>2</sup> Creating computer models

<sup>3</sup> Also known as ETL — moving information to a data warehouse

Source: O'Reilly Media Inc. online survey of more than 600 datascience professionals, conducted from November 2014 to July 2015

THE WALL STREET JOURNAL.

- ◉ Lot of time spent on data
  - ◉ Exploring
  - ◉ Cleaning
  - ◉ Linking
- ◉ Building models
- ◉ Validating
- ◉ Communicating
- ◉ Engineering solutions

## **WHAT DOES A DATA SCIENTIST DO? INTERPRETABLE INSIGHTS FOR BUSINESS USERS**

---

- Deriving insights from the data so that they can be reviewed and acted upon by business users
- Models or solution frameworks need to be sufficiently transparent
- Business Users should be able to understand and interpret the models at a high level

# **WHAT DOES A DATA SCIENTIST DO? INTERPRETABLE INSIGHTS FOR BUSINESS USERS**

---

- Deriving insights from the data so that they can be reviewed and acted upon by business users
- Models or solution frameworks need to be sufficiently transparent
- Business Users should be able to understand and interpret the models at a high level
- Examples:
  - Understanding key drivers of brand perception
  - Quantifying ROI of different marketing initiatives
  - Quantifying impact of various pricing strategies on sales
  - Forecasting demand under various economic scenarios

---

## **WHAT DOES A DATA SCIENTIST DO? BACKEND INTELLIGENCE FOR SMART SYSTEMS**

---

- Data Science is used to build intelligence into applications
- What matters more is whether the models produce accurate and relevant end results

## **WHAT DOES A DATA SCIENTIST DO? BACKEND INTELLIGENCE FOR SMART SYSTEMS**

---

- Data Science is used to build intelligence into applications
- What matters more is whether the models produce accurate and relevant end results
- Examples:
  - Personalising news articles for a user using recommendation engines
  - Quantifying sentiment of conversations on social media
  - Detecting credit card fraud in real time
  - Digital campaign optimisation and execution
  - Processing unstructured data such as images, audio or videos

---

## IMPORTANT NOTES

---

- **NOTE:** Do not believe any facts and figures until you know how they were produced or came from
  - News and Marketing might use or make unsupported claims
- **NOTE:** Data Scientists are Ethical and Respectful in their activities daily when collecting, using and reporting findings and results

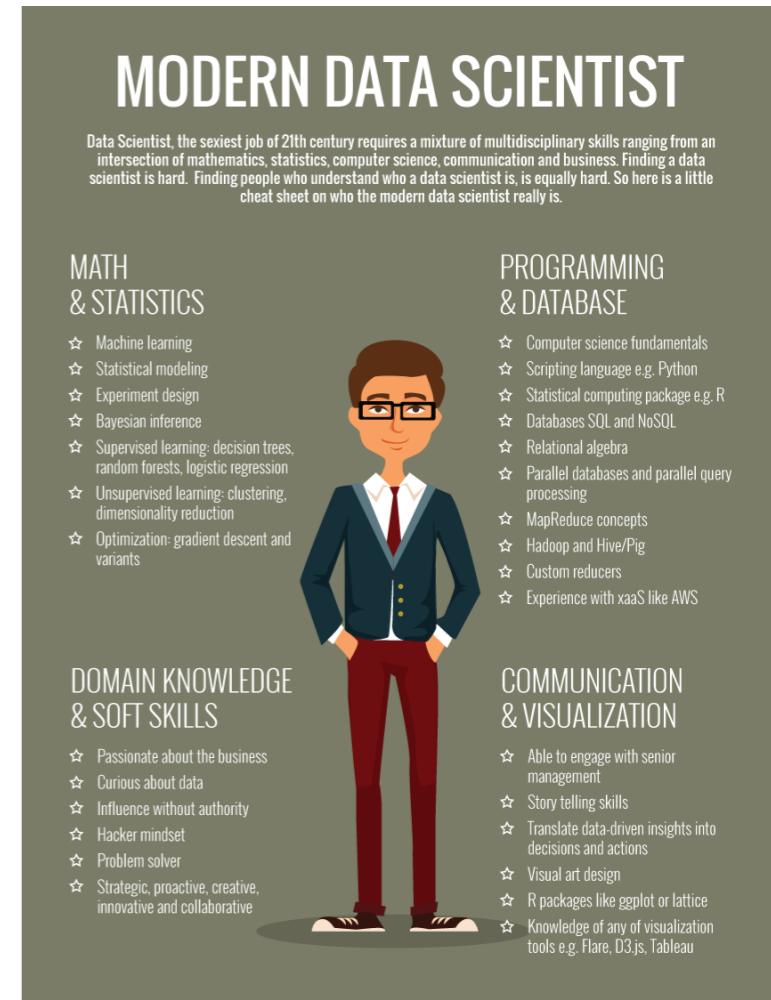
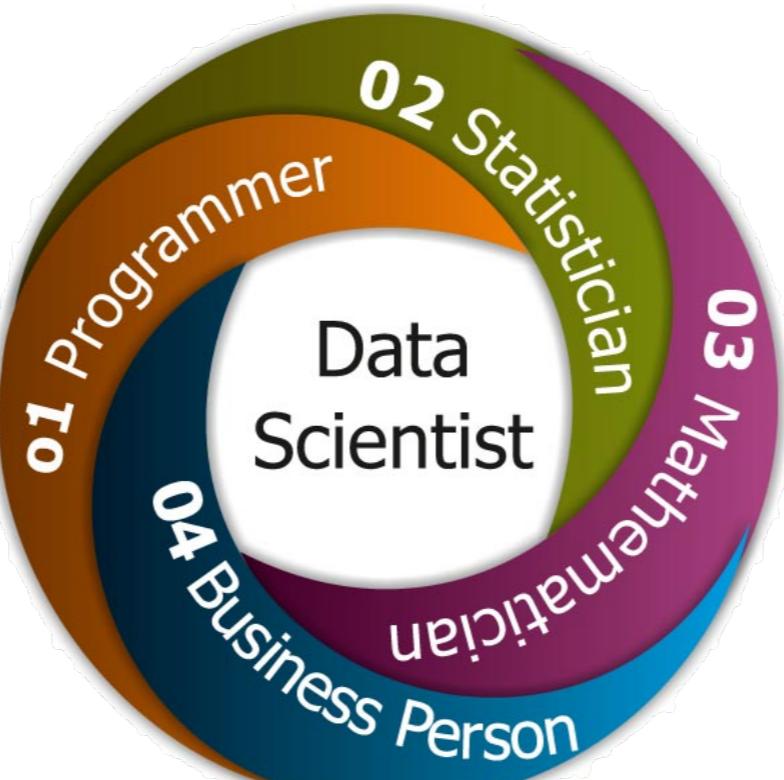
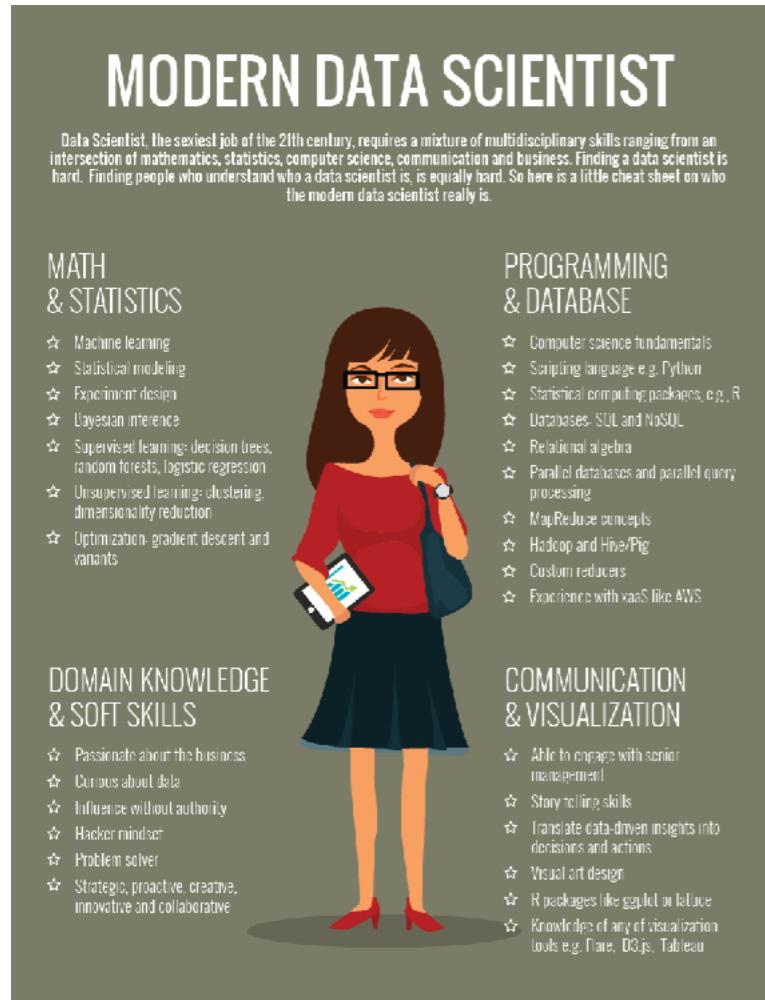
---

**DATA SCIENCE**

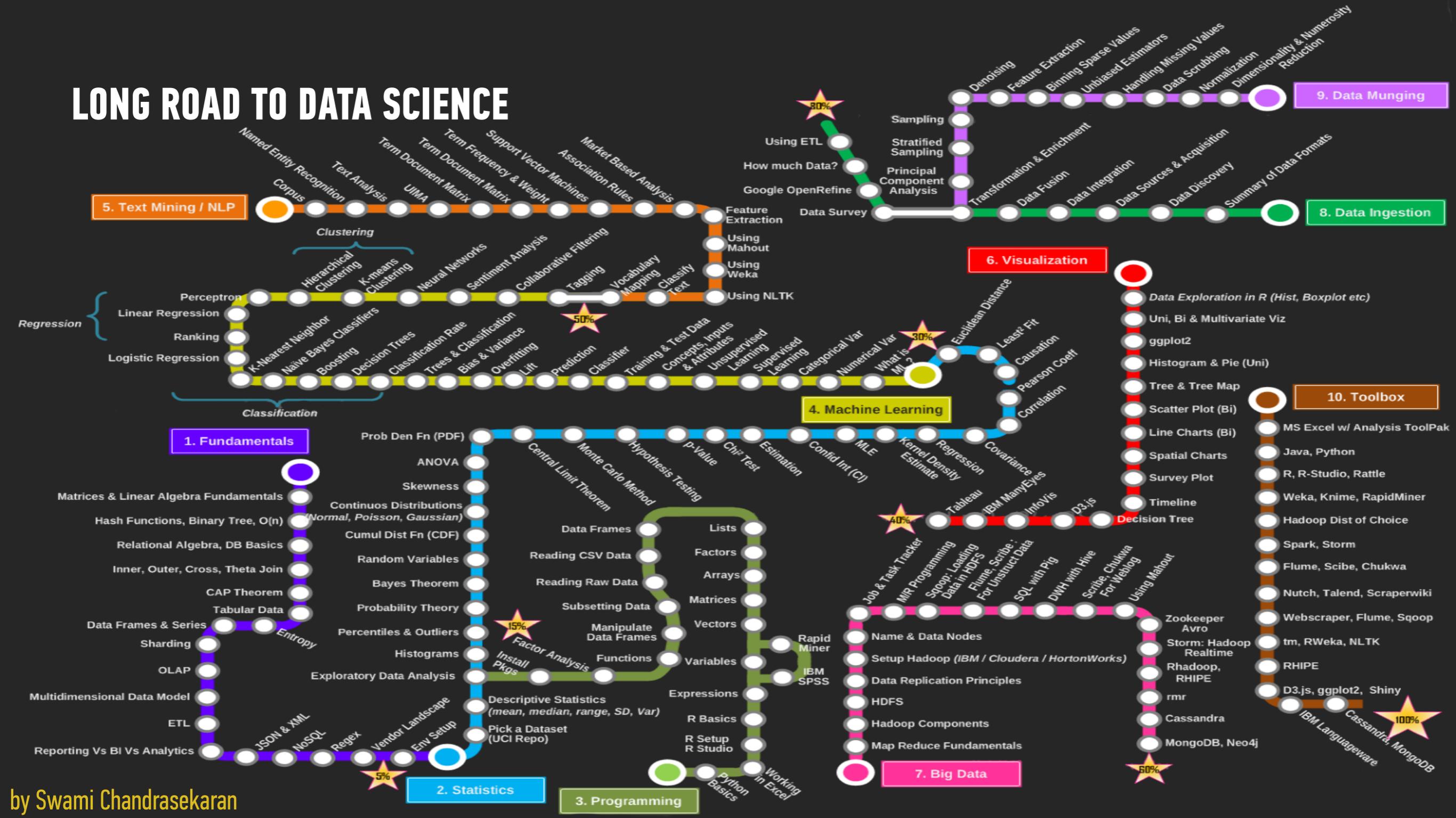
---

# **WHAT DOES A DATA SCIENTIST KNOW AND USE?**

# SKILLS AND TOOLS OF DATA SCIENTISTS



# LONG ROAD TO DATA SCIENCE



---

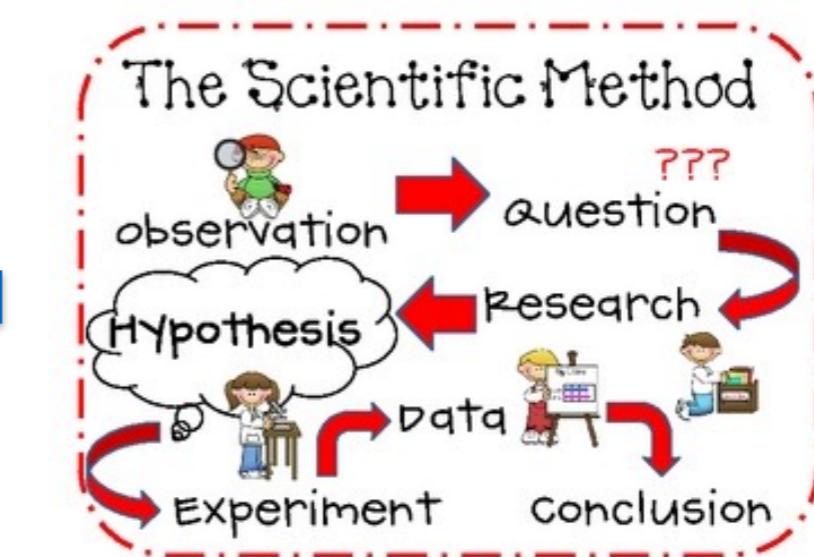
**DATA SCIENCE**

---

# DATA SCIENCE PROCESS

# DATA SCIENCE PROCESS

---

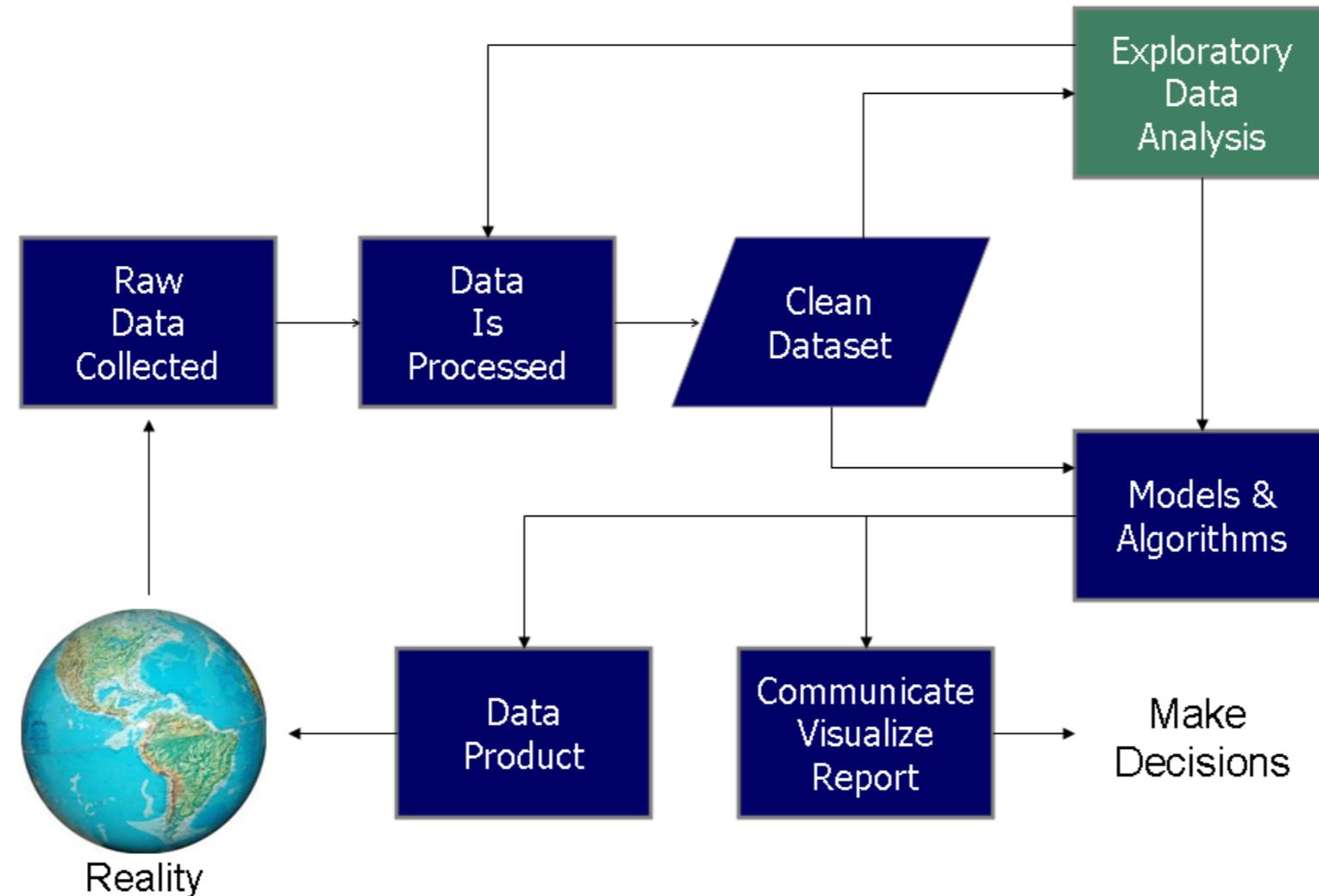


# DATA SCIENCE PROCESS



WIKIPEDIA  
The Free Encyclopedia

## Data Science Process



---

## INTRODUCTION

---

# DATA SCIENCE WORKFLOW

---

## OVERVIEW OF THE DATA SCIENCE WORKFLOW

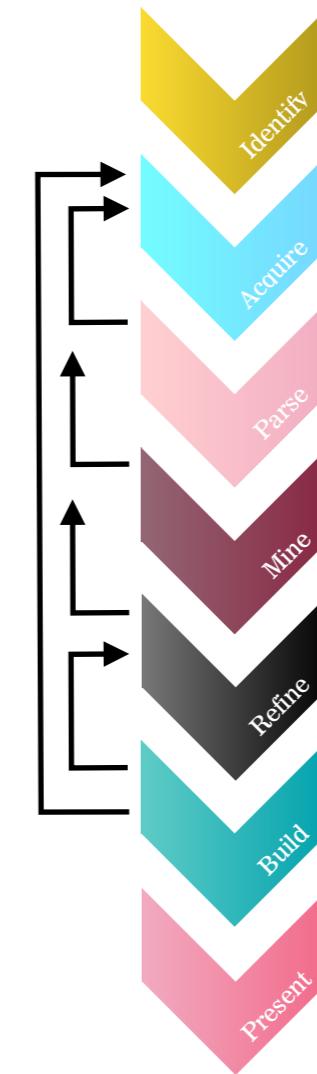
---

- A methodology for doing Data Science
- Similar to the scientific method
- Helps to produce **reliable** and **reproducible** results
  - Reliable: Accurate findings
  - Reproducible: Others can follow your steps and get the same results

# OVERVIEW OF THE DATA SCIENCE WORKFLOW

## ○ The steps

1. Identify the Problem
2. Acquire the Data
3. Parse the Data
4. Mine the Data
5. Refine the Data
6. Build a Data Model
7. Present the Results



# OVERVIEW OF THE DATA SCIENCE WORKFLOW

---

- Identify the Problem
- Identify business / product objectives
- Identify and hypothesise goals and criteria for success
- Create a set of questions for identifying correct dataset



## FUTURAMA EXAMPLE

---

- Problem Statement
- Using Planet Express customer data from January 3001-3005, determine how likely previous customers are to request a repeat delivery using demographic information (profession, company size, location) and previous delivery data (days since last delivery, the total number of deliveries)
- We can use the Data Science Workflow to work through this problem



## FUTURAMA EXAMPLE: IDENTIFY THE PROBLEM

---

- Identify the business/product objectives
  - Are previous customers likely to request a repeat delivery?
- Identify and hypothesise goals and criteria for success
  - What factors are likely to influence a customer's decision to reuse Planet Express for delivery?
- Create a set of questions to help you identify the correct dataset



# OVERVIEW OF THE DATA SCIENCE WORKFLOW

---

- Acquire the Data
- Identify the “right” dataset(s)
- Import data and set up local or remote data structure
- Determine most appropriate tools to work with data



## FUTURAMA EXAMPLE: ACQUIRE THE DATA

---

- Ideal data vs. data that is available
- Often times we start by identifying the ideal data we would want for a project
  
- Learn about limitations of the data
  
- What data is available for this example?
  - demographic information (profession, company size, location)
  - previous delivery data (days since last delivery, total number of deliveries)



---

## FUTURAMA EXAMPLE: ACQUIRE THE DATA

---

- What kind of questions might we want to ask about the data?
- Identifying the “right” dataset(s)
- Is there enough data?
- Does it appropriately align with the question/problem statement?
- Can the dataset be trusted?
- How was it collected?
- Is this dataset aggregated?
- Can we use the aggregation or do we need to get it pre-aggregation?
- Assess resources, requirements, assumptions and constraints

---

## FUTURAMA EXAMPLE: ACQUIRE THE DATA

---

- More questions we might want to ask about the data
  - Import data from the web (Google Analytics, HTML, XML)
  - Import data from a file (CSV, XML, TXT, JSON)
  - Import data from a preexisting database (SQL)
  - Set up local or remote data structure
  - Determine most appropriate tools to work with data
  - Tool follows the format, size of the dataset

## FUTURAMA EXAMPLE: PARSE THE DATA

---

- Secondary data = we did not directly collect it ourselves
- Example data dictionary

Variable	Description	Type of Variable
Profession	Title of the account owner	Categorical / Nominal
Company size	small=1, medium=2, large=3	Categorical, Ordinal
Location	Planet of the company	Categorical / Nominal
Days since last delivery	Integer	Numerical / Discrete
Number of deliveries	Integer	Numerical / Discrete

# OVERVIEW OF THE DATA SCIENCE WORKFLOW

---

- Parse the Data
- Read any documentation provided with the data
- Perform Exploratory Data Analysis
- Verify the quality of the data



## FUTURAMA EXAMPLE: PARSE THE DATA

---

- Questions to ask while parsing
  - Is there documentation for the data?
  - Is there a data dictionary?
  - What kind of filtering, sorting or simple visualisations can help understand the data?
  - What information is contained in the data?
  - What data types are the variables?
  - Are there outliers?
  - Are there trends?



# OVERVIEW OF THE DATA SCIENCE WORKFLOW

---

- Mine the Data
- Determine sampling methodology and sample data
- Format, clean, slice and combine data
- Create necessarily derived columns from the data (Data Engineering)



## FUTURAMA EXAMPLE: MINE THE DATA

---

- Think about sampling
- Get to know the data
- Explore outliers
- Address missing values
- Derive new variables  
(i.e. columns, features)



---

## FUTURAMA EXAMPLE: MINE THE DATA

---

- Common steps while mining the data
- Sample the data with appropriate methodology
- Explore outliers and null values
- Format and clean the data
- Determine how to address missing values
- Format and combine data; aggregate and derive new columns

# OVERVIEW OF THE DATA SCIENCE WORKFLOW

---

- Refine the Data
- Identify trends and outliers
- Apply descriptive and inferential statistics
- Document and transform data



## FUTURAMA EXAMPLE: REFINE THE DATA

---

- Descriptive stats help refine by
  - Identifying trends and outliers
  - Deciding how to deal with outliers
  - Applying descriptive and inferential statistics
  - Document and capture knowledge
  - Determining visualisation techniques for different data types
  - Transforming data



## FUTURAMA EXAMPLE: REFINE THE DATA

---

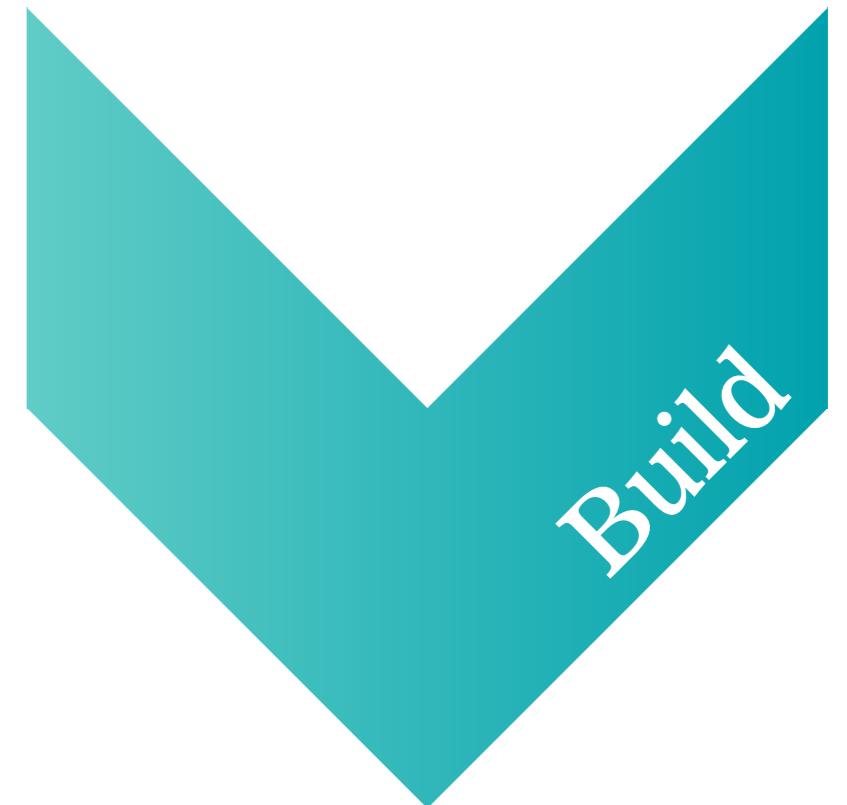
- Use statistics and visualisation to identify trends
- Example of basic statistics

Variable	Mean (Std) or Frequency (%)
Number of Deliveries	50.0 (10)
Earth	50 (10%)
Amphibios 9	100 (20%)
Bogad	100 (20%)
Colgate 8	100 (20%)
Other	150 (30%)

# OVERVIEW OF THE DATA SCIENCE WORKFLOW

---

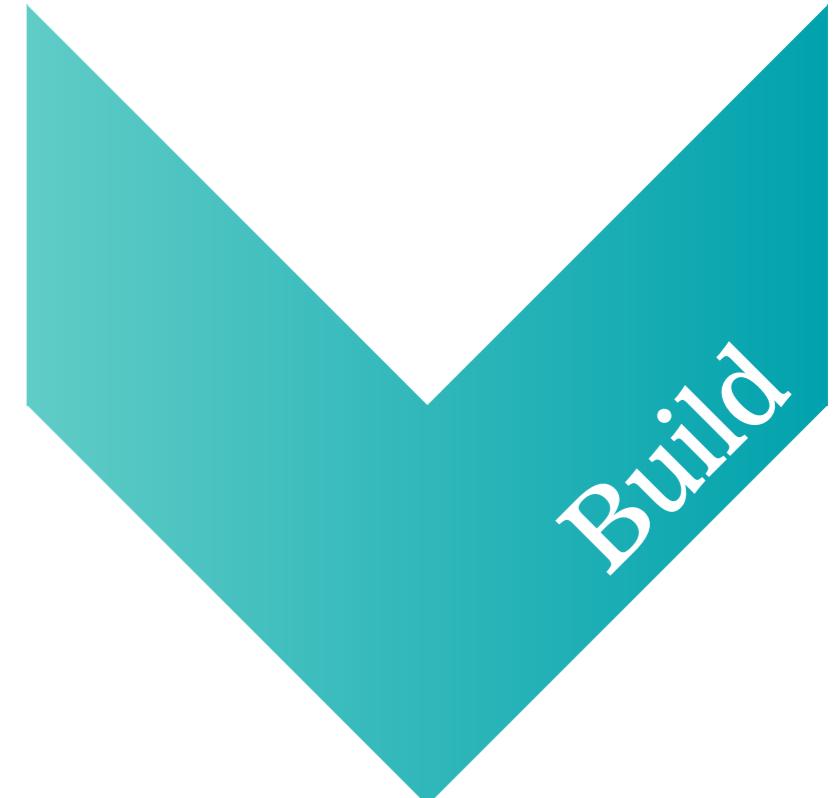
- Build a Data Model
- Select appropriate Modelling
- Build some Models
- Evaluate and refine a Model



## FUTURAMA EXAMPLE: BUILD A DATA MODEL

---

- Select a model based upon the outcome
- Example model statement
  - We completed a logistic regression using Statsmodels v. XX. We calculated the probability of a customer placing another order with Planet Express.
  - Here, we are using a logistic regression model because we are determine the probability that a customer may place a return order, which at its heart is a classification problem



---

## FUTURAMA EXAMPLE: BUILD A DATA MODEL

---

- The steps for model building are
- Select the appropriate model
- Build the model
- Evaluate and refine the model
- Predict outcomes and action items

# OVERVIEW OF THE DATA SCIENCE WORKFLOW

---

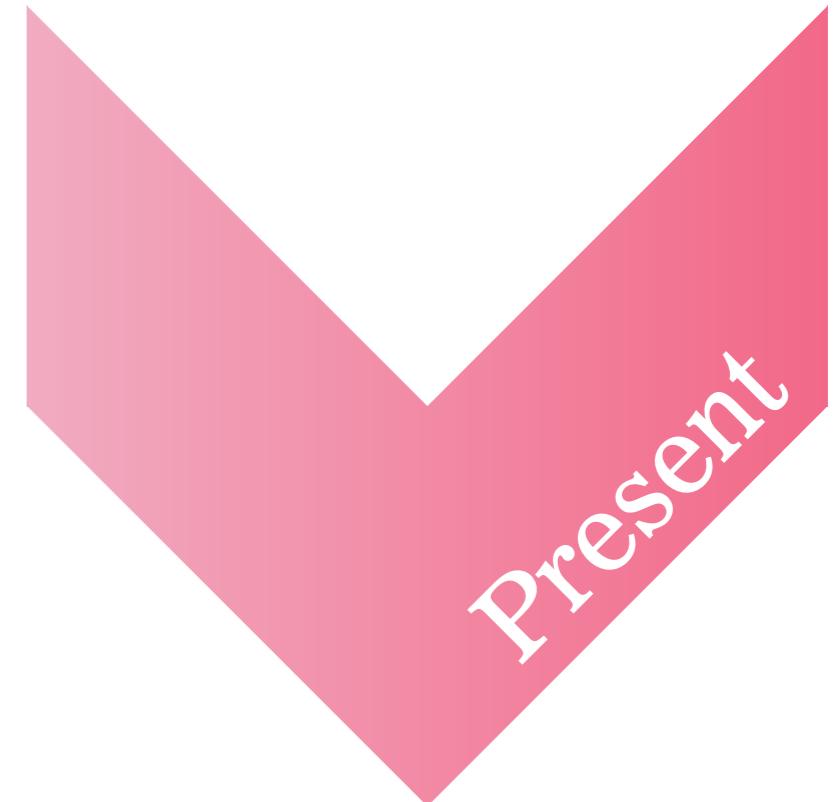
- Present the Results
- Summarise findings with narrative and storytelling techniques
- Present limitations and assumptions of your analysis
- Identify follow-up problems and questions for future analysis



## FUTURAMA EXAMPLE: PRESENT THE RESULTS

---

- You have to effectively communicate your results to matter!
- Ranges from a simple email to a complex web graphic
- Make sure to consider your audience
- A presentation for fellow data scientists will be drastically different from a presentation for an executive



---

## FUTURAMA EXAMPLE: PRESENT THE RESULTS

---

- Key factors of a good presentation include
  - Summarise findings with narrative and storytelling techniques
  - Refine your visualisations for broader comprehension
  - Present both limitations and assumptions
  - Determine the integrity of your analyses
  - Consider the degree of disclosure for various stakeholders
  - Test and evaluate the effectiveness of your presentation beforehand

---

## FUTURAMA EXAMPLE: PRESENT THE RESULTS

---

- Example presentations and infographics

- [Sydney vs Melbourne - Which Is The Best City To Drive In?](#)

- [Ready for a career change? You are not alone](#)

- [Australian Tourism Infographic](#)

---

**GUIDED PRACTICE**

---

# **DATA SCIENCE WORKFLOW**

# ACTIVITY: DATA SCIENCE WORKFLOW



## DIRECTIONS (25 MINUTES)

1. Divide into groups, each located at a whiteboard
2. **Identify:** Each group should develop 1 research question they would like to know about their classmates. Create a hypothesis to your question. Do not share your question yet! (5 minutes)
3. **Acquire:** Rotate from group to group to collect data for your hypothesis. Have other students write or tally their answers on the whiteboard. (10 minutes)
4. **Present:** Communicate the results of your analysis to the class. (10 minutes)
  - a. Create a narrative to summarise your findings
  - b. Provide a basic visualisation for easy comprehension
  - c. Choose one student to present for the group

---

**DEMONSTRATION**

---

# ENVIRONMENT SETUP

---

# DEVELOPMENT ENVIRONMENT SETUP

---

- Brief introduction of tools
- Environment setup
  - Create a GitHub account
  - Install Python 2.7 and Anaconda
  - Practice Python syntax, Terminal commands and Pandas
- Jupyter Notebook test and Python review

---

## DEVELOPMENT ENVIRONMENT SETUP

---

- Test your new setup using the lesson 1 starter code available in the GitHub repository at
  - `~/lessons/lesson-01/code/starter-code/lesson1-starter-code.ipynb`
- Ask your classmates and instructor for help if you have problems!

---

**CONCLUSION**

---

**REVIEW**

---

## CONCLUSION

---

- You should now be able to answer the following questions
- What is Data Science?
- What is the Data Science Process / Workflow?
- How can you have a successful learning experience at GA?

---

**DATA SCIENCE**

---

# PROJECTS

# PROJECTS

---

Unit	Project	Assigned	Deadline
Unit 1	Unit Project 1	Lesson 1	Lesson 3
Unit 1	Unit Project 2	Lesson 3	Lesson 5
Unit 2	Final Project, part 1	Lesson 1	Lesson 8
Unit 2	Unit Project 3	Lesson 5	Lesson 10
Unit 2	Unit Project 4	Lesson 9	Lesson 12
Unit 3	Final Project, part 2	Lesson 8	Lesson 14
Unit 3	Final Project, part 3	Lesson 14	Lesson 16
Unit 3	Final Project, part 4	Lesson 16	Lesson 18
Unit 3	Final Project, part 5	Lesson 18	Lesson 20

**DATA SCIENCE**

---

**BEFORE NEXT CLASS**

---

## **BEFORE NEXT CLASS**

---

## **DUE DATE**

- Projects
  - Begin work on Unit Project 1
  - Begin thinking on Final Project

---

**WELCOME TO DATA SCIENCE**

---

# Q & A

---

WELCOME TO DATA SCIENCE

---

# CREDITS AND REFERENCES

# WELCOME TO DATA SCIENCE

---

- Simon Lehna Singh, MBE is a British popular science author whose works largely contain a strong mathematical element
  - [Simon Singh on Wikipedia](#) and [Simon Singh personal site](#)
- Videos about numbers - it's that simple, by Brady Haran
  - [Numberphile](#)
- Watson is a technology that understands all forms of data
  - [IBM Watson on Wikipedia](#) and [Web site](#)
  - and reasons and learns at scale. (If curious, check the [youTube](#))
- This course gives you easy access to the invaluable learning techniques used by experts
  - [Learning How to Learn](#)