

# The Data Scientist's skill set

---

Gianluca Campanella

17<sup>th</sup> July 2018

# Types of skills

1. Technical
2. Non-technical

# Technical skills

---

# Technical skills

- Basic probability
- Descriptive statistics
- SQL
- Python and/or R + some libraries

# Technical skills

Scripting language	Python	R
Data analysis and manipulation	pandas	tidyverse
Visualisation	seaborn	ggplot2
Modelling	sklearn	caret, mlr, ...

Standardise and automate  
data collection and analysis

Standardise and automate  
**everything!**

# Coding in academia vs industry

- Reusability
- VCS and code reviews
- Testing
- Integration
- Monitoring and maintenance
- Scalability and distributed systems



# Non-technical skills

---

# Non-technical skills

You need to be able to...

- Identify business value
- Solve problems using dirty data
- Communicate effectively and manage expectations

# How do you identify business value?

- You need domain expertise
- What's interesting might not be what's important
- **Quantify impact** whenever possible

# How do you solve problems using dirty data?

- Be curious... and a bit stubborn
- Understand when to change strategy
- **Focus on quality and execution** rather than quantity

# Certainty vs growth rate of knowledge

## Science

tries to maximise

**certainty**

## Adversarial industries

try to maximise

**growth rate**

# How do you communicate effectively?

- Speak the stakeholders' language
- Explain **why** a certain question is worth answering
- Understand limitations and **don't overstate results**