Lecture 0: Course Introduction

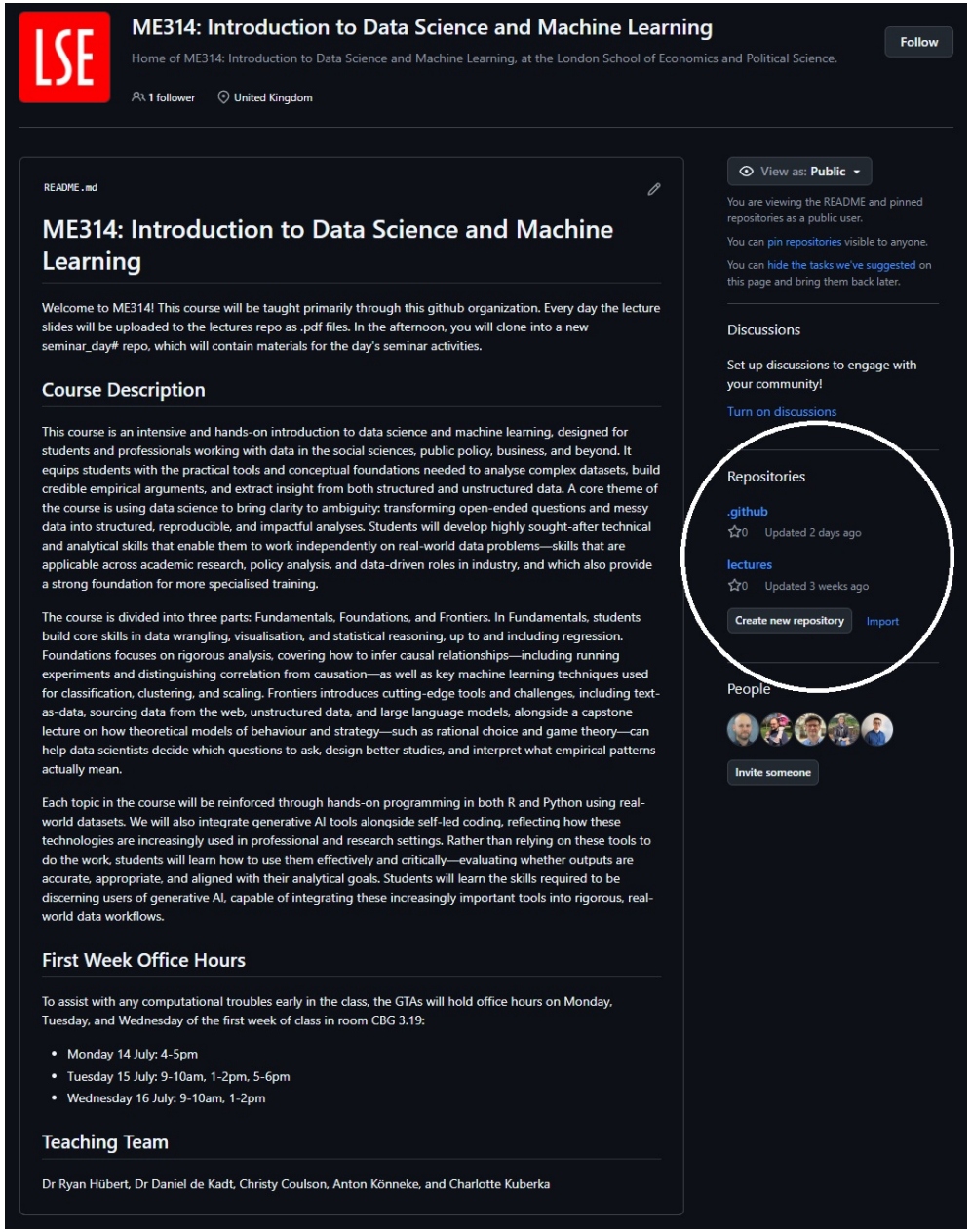
LSE ME314: Introduction to Data Science and Machine Learning (https://github.com/me314-lse)

2025-07-14

Daniel de Kadt and Ryan Hübert

Slides last updated: 13 July 2025

Course homepage: https://github.com/me314-lse



ME314: Introduction to Data Science and Machine Learning

Home of ME314: Introduction to Data Science and Machine Learning, at the London School of Economics and Political Science.

Follow

1 follower · United Kingdom

README.md

ME314: Introduction to Data Science and Machine Learning

Welcome to ME314! This course will be taught primarily through this github organization. Every day the lecture slides will be uploaded to the lectures repo as .pdf files. In the afternoon, you will clone into a new seminar_day# repo, which will contain materials for the day's seminar activities.

Course Description

This course is an intensive and hands-on introduction to data science and machine learning, designed for students and professionals working with data in the social sciences, public policy, business, and beyond. It equips students with the practical tools and conceptual foundations needed to analyse complex datasets, build credible empirical arguments, and extract insight from both structured and unstructured data. A core theme of the course is using data science to bring clarity to ambiguity: transforming open-ended questions and messy data into structured, reproducible, and impactful analyses. Students will develop highly sought-after technical and analytical skills that enable them to work independently on real-world data problems—skills that are applicable across academic research, policy analysis, and data-driven roles in industry, and which also provide a strong foundation for more specialised training.

The course is divided into three parts: Fundamentals, Foundations, and Frontiers. In Fundamentals, students build core skills in data wrangling, visualisation, and statistical reasoning, up to and including regression. Foundations focuses on rigorous analysis, covering how to infer causal relationships—including running experiments and distinguishing correlation from causation—as well as key machine learning techniques used for classification, clustering, and scaling. Frontiers introduces cutting-edge tools and challenges, including text-as-data, sourcing data from the web, unstructured data, and large language models, alongside a capstone lecture on how theoretical models of behaviour and strategy—such as rational choice and game theory—can help data scientists decide which questions to ask, design better studies, and interpret what empirical patterns actually mean.

Each topic in the course will be reinforced through hands-on programming in both R and Python using real-world datasets. We will also integrate generative AI tools alongside self-led coding, reflecting how these technologies are increasingly used in professional and research settings. Rather than relying on these tools to do the work, students will learn how to use them effectively and critically—evaluating whether outputs are accurate, appropriate, and aligned with their analytical goals. Students will learn the skills required to be discerning users of generative AI, capable of integrating these increasingly important tools into rigorous, real-world data workflows.

First Week Office Hours

To assist with any computational troubles early in the class, the GTAs will hold office hours on Monday, Tuesday, and Wednesday of the first week of class in room CBG 3.19:

- Monday 14 July: 4-5pm
- Tuesday 15 July: 9-10am, 1-2pm, 5-6pm
- Wednesday 16 July: 9-10am, 1-2pm

Teaching Team

Dr Ryan Hübert, Dr Daniel de Kadt, Christy Coulson, Anton Könneke, and Charlotte Kuberka

View as: Public

You are viewing the README and pinned repositories as a public user.

You can pin repositories visible to anyone.

You can hide the tasks we've suggested on this page and bring them back later.

Discussions

Set up discussions to engage with your community!

Turn on discussions

Repositories

.github
☆0 Updated 2 days ago

lectures
☆0 Updated 3 weeks ago

Create new repository · Import

People

Invite someone


오징어 게임 2
456억, 동심의 게임은 끝나지 않았다
ONLY ON NETFLIX | 12월 26일 공개

Welcome

Welcome to ME314: Introduction to Data Science and Machine Learning!

Together we will take a journey through the landscape of modern data science and machine learning.

Welcome

Welcome to ME314: Introduction to Data Science and Machine Learning!

Together we will take a journey through the landscape of modern data science and machine learning.

Hopefully it won't feel like Squid Game!

Welcome

Welcome to ME314: Introduction to Data Science and Machine Learning!

Together we will take a journey through the landscape of modern data science and machine learning.

Hopefully it won't feel like Squid Game!

But the material is challenging, and fast paced, so be prepared.

Instructors

Dr. Daniel de Kadt
Assistant Professor in Quantitative Research Methods

Dr. Ryan Hübert
Associate Professor in Computational Social Science

Christy Coulson
PhD Student in Methodology

Anton Könneke
PhD Student in Government

Charlotte Kuberka
PhD Student in Government

Course Outline

Fundamentals:

- ➜ Computational tools (Day 1, Day 2)
- ➜ Data, manipulation, and visualization (Day 2, Day 3)
- ➜ Probability, statistics, and regression (Day 3, Day 4)

Foundations:

- ➜ Causal inference (Day 5, Day 6)
- ➜ Machine learning (Day 7, Day 8, Day 9)

Frontiers:

- ➜ Non-standard data (audio, visual, and maps) (Day 10)
- ➜ Text-as-data (+ brief overview of LLMs) (Day 11)
- ➜ Theory $\longleftrightarrow$ data science (Day 12)

How it Works: Lectures and Seminars/Classes

Each day you will attend:

- 3 hour interactive **lecture**
- 1.5 hour **seminar/class** where you apply what you learned using *your own computer*

It's important that you have your own computer, and that it is set up for success. We'll spend some of the first week working on this.

There are dedicated **office hours** for computational troubles you encounter early on (room CBG 3.19):

- Monday 14 July: 4-5pm
- Tuesday 15 July: 9-10am, 1-2pm, 5-6pm
- Wednesday 16 July: 9-10am, 1-2pm

How it Works: Evaluation

1. Take-home problem set on Day 5 (Monday 21 July):
   - ➜ 24 hours, open book, internet allowed
   - ➜ Covers material from the first week of the course
   - ➜ Please describe how you used GenAI, if you did

How it Works: Evaluation

1. Take-home problem set on Day 5 (Monday 21 July):
   - ➜ 24 hours, open book, internet allowed
   - ➜ Covers material from the first week of the course
   - ➜ Please describe how you used GenAI, if you did

2. Final exam (1 August):
   - ➜ 2 hours, pen and paper, closed book
   - ➜ Covers all material from the course
   - ➜ No make-up exams: You must attend the exam at the date/time/location provided by the SSO
   - ➜ If you have an "exceptional circumstance" (very urgent or pressing difficulty) please contact the SSO to discuss

What's On Offer

By the end of ME314 we hope you have acquired...

What's On Offer

By the end of ME314 we hope you have acquired...

1. Familiarity with the data science landscape:
   - ➜ What data scientists do (and how that is changing)
   - ➜ The types of questions data scientists ask
   - ➜ The types of solutions data scientists develop

What's On Offer

By the end of ME314 we hope you have acquired...

1. Familiarity with the data science landscape:
   - ➜ What data scientists do (and how that is changing)
   - ➜ The types of questions data scientists ask
   - ➜ The types of solutions data scientists develop

2. Practical knowledge about key data science and ML skills:
   - ➜ How to use various computational tools
   - ➜ How to use various designs, methods, and models

What's On Offer

By the end of ME314 we hope you have acquired...

1. Familiarity with the data science landscape:
   - ➜ What data scientists do (and how that is changing)
   - ➜ The types of questions data scientists ask
   - ➜ The types of solutions data scientists develop

2. Practical knowledge about key data science and ML skills:
   - ➜ How to use various computational tools
   - ➜ How to use various designs, methods, and models

3. An appreciation of the scientific ethic:
   - ➜ Ethics, honesty, transparency, reproducibility

What We Ask of You

Some informal pre-requisites:

1. Some familiarity with some math and notation (the slides will have math)
2. Open-mindedness and willingness to learn to write code (you will do a lot of programming in `R` and some in `python`)

Seminars/classes:

- ➜ You cannot change class/seminar group
- ➜ Attendance in seminars/classes is mandatory, and attendance in lecture is strongly encouraged — we will be keeping track of all attendance

Academic integrity:

- ➜ Please review LSE's policies on academic misconduct — we will refer suspected academic misconduct to the school

Data Science: "Seems Good"

Analytics And Data Science

Data Scientist: The Sexiest Job of the 21st Century

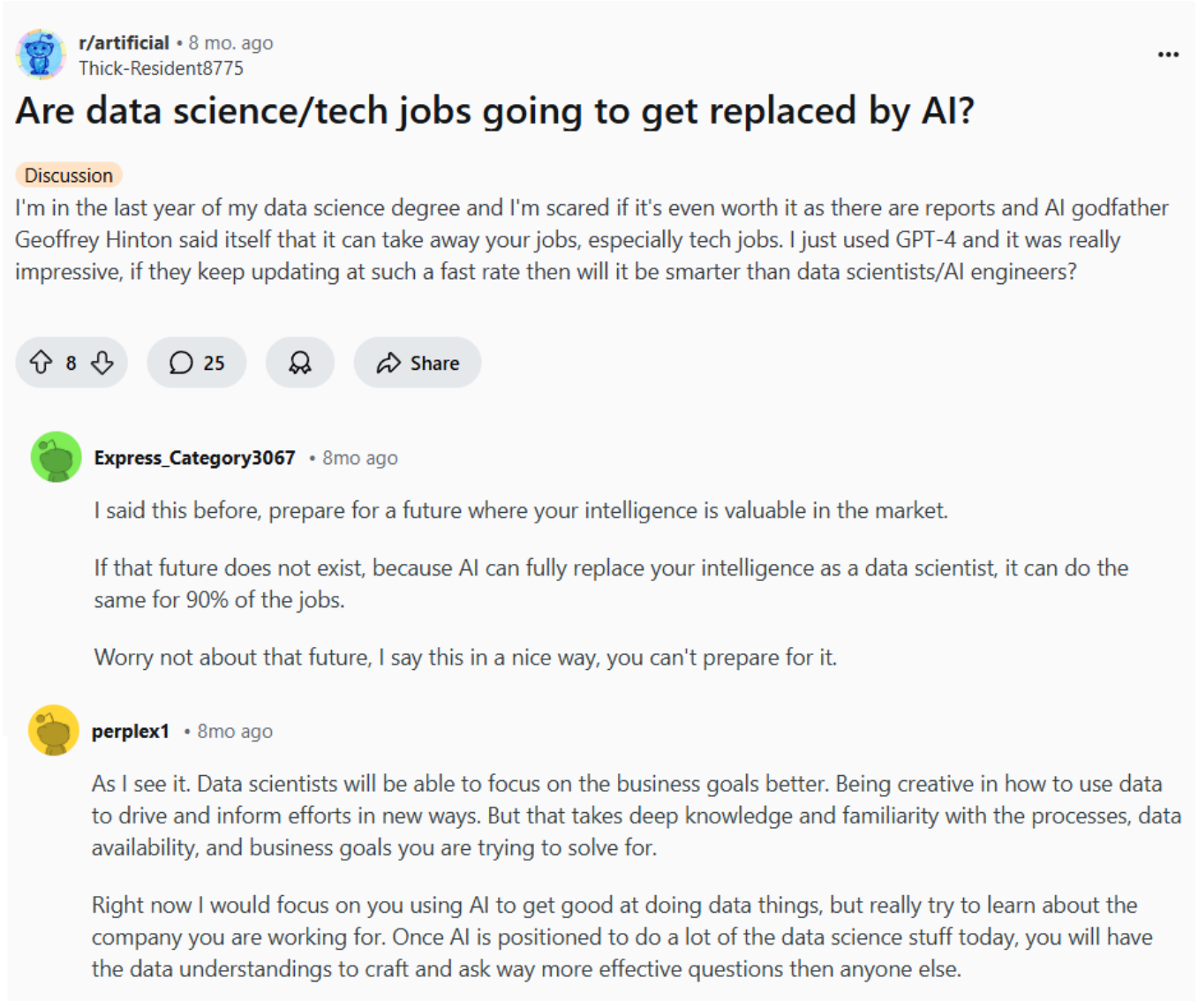
by Thomas H. Davenport and DJ Patil

From the Magazine (October 2012)

Andrew J Buboltz, silk screen on a page from a high school yearbook, 8.5" x 12", 2011 Tamar Cohen

Summary. Back in the 1990s, computer engineer and Wall Street "quant" were the hot occupations in business. Today data scientists are the hires firms are competing to make. As companies wrestle with unprecedented volumes and types of information, demand for... **more**

Data Science: "Seems Bad"



r/artificial • 8 mo. ago
Thick-Resident8775

Are data science/tech jobs going to get replaced by AI?

Discussion

I'm in the last year of my data science degree and I'm scared if it's even worth it as there are reports and AI godfather Geoffrey Hinton said itself that it can take away your jobs, especially tech jobs. I just used GPT-4 and it was really impressive, if they keep updating at such a fast rate then will it be smarter than data scientists/AI engineers?

8 25 Share

Express_Category3067 • 8mo ago

I said this before, prepare for a future where your intelligence is valuable in the market.

If that future does not exist, because AI can fully replace your intelligence as a data scientist, it can do the same for 90% of the jobs.

Worry not about that future, I say this in a nice way, you can't prepare for it.

perplex1 • 8mo ago

As I see it. Data scientists will be able to focus on the business goals better. Being creative in how to use data to drive and inform efforts in new ways. But that takes deep knowledge and familiarity with the processes, data availability, and business goals you are trying to solve for.

Right now I would focus on you using AI to get good at doing data things, but really try to learn about the company you are working for. Once AI is positioned to do a lot of the data science stuff today, you will have the data understandings to craft and ask way more effective questions then anyone else.

LLMs and Generative AI

LLMs and GenAI are an increasingly important part of data scientists' toolkit.

A bunch of the demo code I will show you has been co-written and augmented with GenAI.

These are powerful tools, and anyone who says otherwise is in denial.

LLMs and Generative AI

LLMs and GenAI are an increasingly important part of data scientists' toolkit.

A bunch of the demo code I will show you has been co-written and augmented with GenAI.

These are powerful tools, and anyone who says otherwise is in denial.

But they are **not** a substitute for understanding.

LLMs and Generative AI

We ask you **not** to do the following:

- ➜ Install/activate Copilot in VSCode, or Cursor on your computer
- ➜ Copy and paste questions we ask to GenAI tools, then copy and paste the answers back into your assignments
- ➜ Prompt GenAI to write code, then copy and paste into your assignments

It's **okay** for you to use GenAI tools to:

- ➜ Help you debug code you've written
- ➜ Help you understand some code or a concept
- ➜ Refactor, lint, or improve code

LLMs and Generative AI

We ask you **not** to do the following:

- ➜ Install/activate Copilot in VSCode, or Cursor on your computer
- ➜ Copy and paste questions we ask to GenAI tools, then copy and paste the answers back into your assignments
- ➜ Prompt GenAI to write code, then copy and paste into your assignments

It's **okay** for you to use GenAI tools to:

- ➜ Help you debug code you've written
- ➜ Help you understand some code or a concept
- ➜ Refactor, lint, or improve code

Again, these tools are **not** a substitute for understanding.

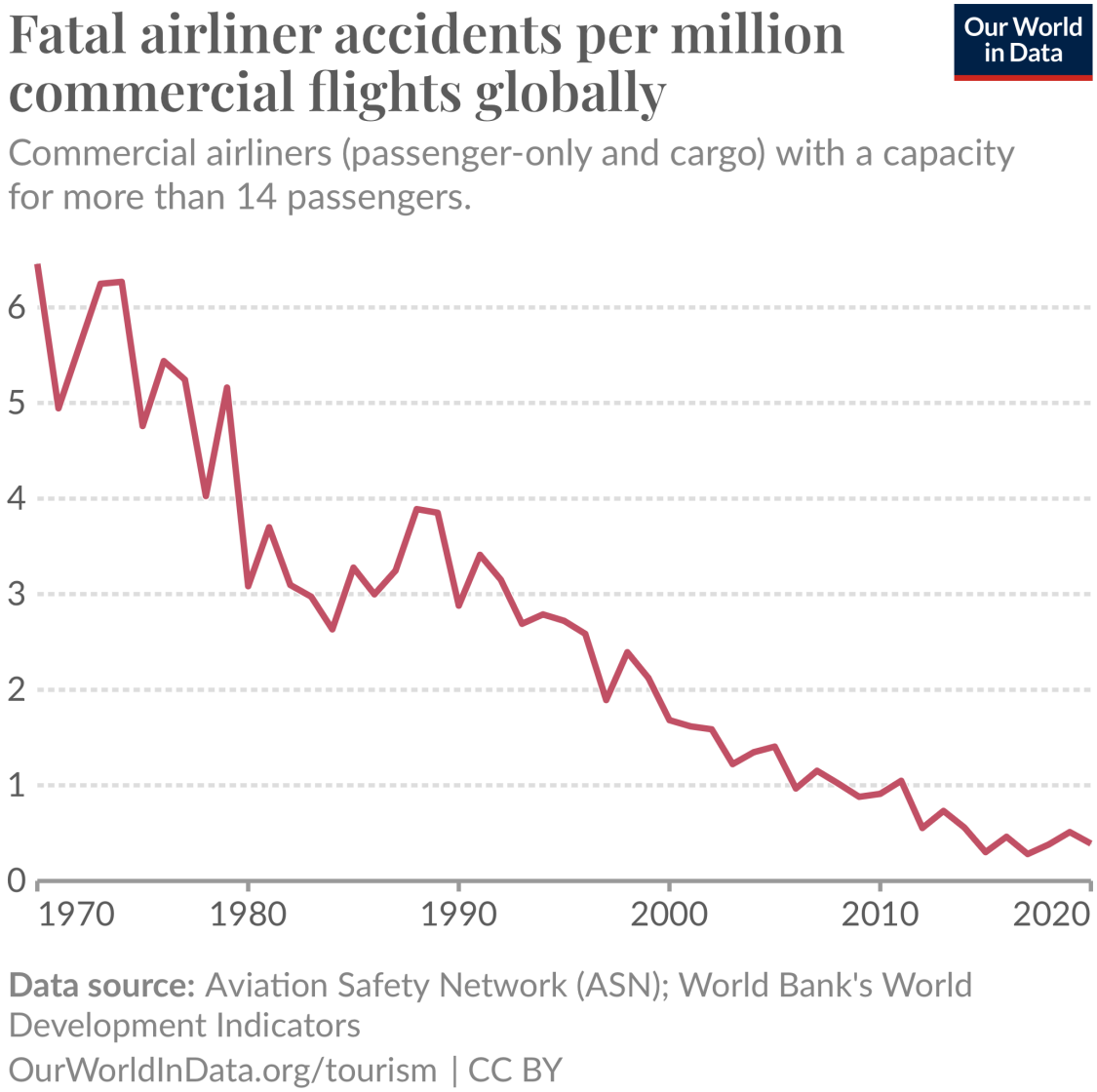
Automation: Upsides and Downsides

Source: Wikipedia, REDACTED

Automation: Upsides and Downsides

Source: British Airways, REDACTED

Automation: Upsides and Downsides



Source: Our World in Data, Commercial Flight Safety

Automation: Learn to Fly

Yes, 'automation' (or computer delegation) is powerful and the future.

Be like pilots $\longrightarrow$ learn how to fly the plane, and how to delegate to the computer.

Automation: Learn to Fly

Yes, 'automation' (or computer delegation) is powerful and the future.

Be like pilots $\longrightarrow$ learn how to fly the plane, and how to delegate to the computer.

If you can fly and you can delegate, you will be equipped to do **responsible**, **high quality** data science at **scale**.

Automation: Learn to Fly

Yes, 'automation' (or computer delegation) is powerful and the future.

Be like pilots $\longrightarrow$ learn how to fly the plane, and how to delegate to the computer.

If you can fly and you can delegate, you will be equipped to do **responsible**, **high quality** data science at **scale**.

Punchline: You ***need to know how to fly***!