

# Geographic Data Science - Lecture I

## Introduction

Dani Arribas-Bel

# Today

- This course
- The (geo-)data revolution
- (Geo-)Data Science

This course

# Quiz

- Can you think of a real-world context where data and statistics are being used to make a difference?
- Have you ever heard the term "**Big Data**"?
- Have you ever heard the term "**Data Science**"?
- Have you ever written a line of computer code?

*More stats than a GIS course, more GIS than a stats course...*

*More stats than a GIS course, more GIS than a stats course...*

*...but in a fun way!*

# Philosophy

- (Lots of) **methods** and techniques
  - General overview
  - Intuition
  - Very little math
  - Lots of ways to continue on your own
- Emphasis on the **application** and **use**
- Close connection to "real world" applications

# Logistics - Website

<http://darribas.org/gds16>

# Logistics - Format

11 weeks of:

- **Prep. materials:** videos, podcasts, articles... 1h.  
approx. (most recommended!)
- **1h. Lecture:** concepts, methods, examples
- **2h. Computer practical:** hands-on, application of  
concepts, Python (highly *employable*)
- **Further readings:** how to go beyond the minimum

# Logistics - Content

# Logistics - Content

- **Weeks 1-3:** "big picture" lectures + introduction to computational tools (learning curve)

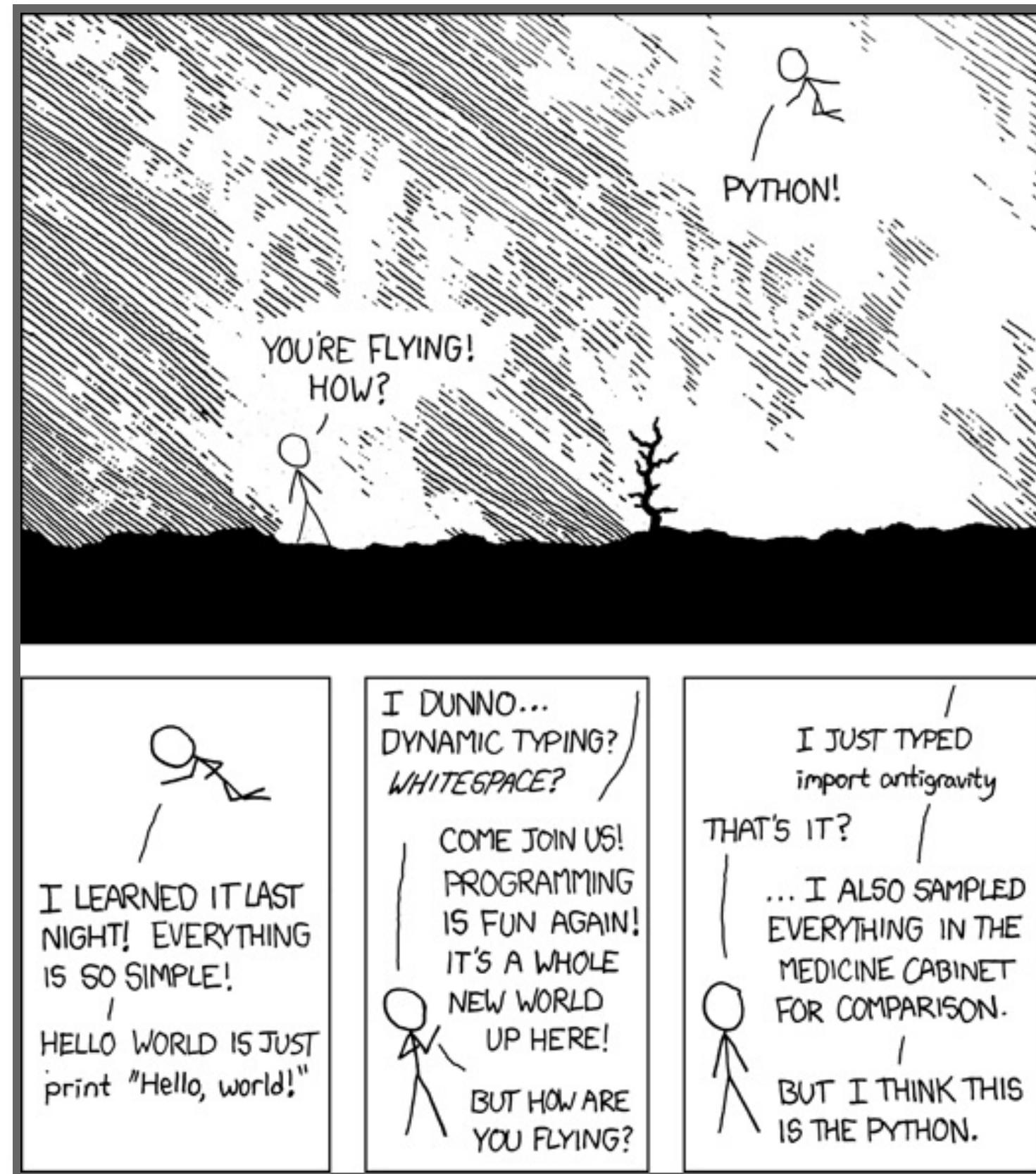
# Logistics - Content

- **Weeks 1-3:** "big picture" lectures + introduction to computational tools (learning curve)
- **Weeks 4-8:** "meat" of the course (lots of concepts packed)

# Logistics - Content

- **Weeks 1-3:** "big picture" lectures + introduction to computational tools (learning curve)
- **Weeks 4-8:** "meat" of the course (lots of concepts packed)
- **Weeks 9-11:** catch up + prepare an awesome Assignment II

# Python



# Python

- General purpose programming language
- Sweet spot between "*proof-of-concept*" and "*production-ready*"
- Industry standard: **GIS** (Esri, QGIS) and **Data Science** (Google, Facebook, Amazon, Netflix, The New York Times, NASA...)

# Self-directed learning (or "*it's all about YOU*")

**Prepare** for the labs

# Self-directed learning (or "*it's all about YOU*")

## **Prepare** for the labs

- I won't be leading/lecturing at the computer labs

# Self-directed learning (or "*it's all about YOU*")

## **Prepare** for the labs

- I won't be leading/lecturing at the computer labs
- Go over the notebooks between the lecture and the computer lab --> If the first time you see a notebook is at the lab, you won't be able to follow on

# Self-directed learning (or "*it's all about YOU*")

## **Prepare** for the labs

- I won't be leading/lecturing at the computer labs
- Go over the notebooks between the lecture and the computer lab --> If the first time you see a notebook is at the lab, you won't be able to follow on
- Bring questions, comments, feedback, (informed) rants to class/labs

# Self-directed learning (or "*it's all about YOU*")

## **Prepare** for the labs

- I won't be leading/lecturing at the computer labs
- Go over the notebooks between the lecture and the computer lab --> If the first time you see a notebook is at the lab, you won't be able to follow on
- Bring questions, comments, feedback, (informed) rants to class/labs
- Use the forum (link on VITAL)

# Self-directed learning (or "*it's all about YOU*")

## Prepare for the labs

- I won't be leading/lecturing at the computer labs
- Go over the notebooks between the lecture and the computer lab --> If the first time you see a notebook is at the lab, you won't be able to follow on
- Bring questions, comments, feedback, (informed) rants to class/labs
- Use the forum (link on VITAL)
- Collaborate with your peers (it's NOT a zero-sum win!!!)

# More help!!!

This course is much more about "**learning to learn**" and **problem solving** rather than acquiring specific programming tricks or stats wizardry

# More help!!!

This course is much more about "**learning to learn**" and **problem solving** rather than acquiring specific programming tricks or stats wizardry

- Learn to ask questions

# More help!!!

This course is much more about "**learning to learn**" and **problem solving** rather than acquiring specific programming tricks or stats wizardry

- Learn to ask questions
- Help others as much as you can (the best way to learn is to teach)

# More help!!!

This course is much more about "**learning to learn**" and **problem solving** rather than acquiring specific programming tricks or stats wizardry

- Learn to ask questions
- Help others as much as you can (the best way to learn is to teach)
- Use heavily Google + Stack Overflow

# Assignments

- Mark (mostly) based on two assignments, due:
  1. Week 7 (47.5%)
  2. Week 12 (47.5%)
- Coursework
- **Equivalent** to 2,500: report (*notebook*) with **code**, **figures** (e.g. maps), and **text**

# The (geo-)data revolution

# The (geo-)data revolution

Exciting times to be a:

- Geographer
- Map fan
- Data fan

The world is being "**datafied**"...

# "Datafication"

*Quantification of phenomena through the systematic recording of data*

*“taking all aspects of life and turning them into data”*  
Cukier & (Mayer-Schoenberg)

Examples: credit transactions, public transit, tweets, facebook likes, spotify songs, etc.

# "Datafication"

Many implications:

- Opportunities for optimization of systems  
(Industrial IoT, planning systems...)
- **Window** into human **behaviour** (this course)
- Issues with **intentionality** and **privacy**
- ...

*Why now?*

# *Why now?*

Advances in:

- Computing power and storage
- Connectivity
- Geospatial technology

# The (geo-)data revolution

The confluence of the three (computing, communication and geospatial) is creating large amounts of data.

**Now, data in itself is not very valuable:**

- Data --> Information --> Knowledge --> Action

# Data Science

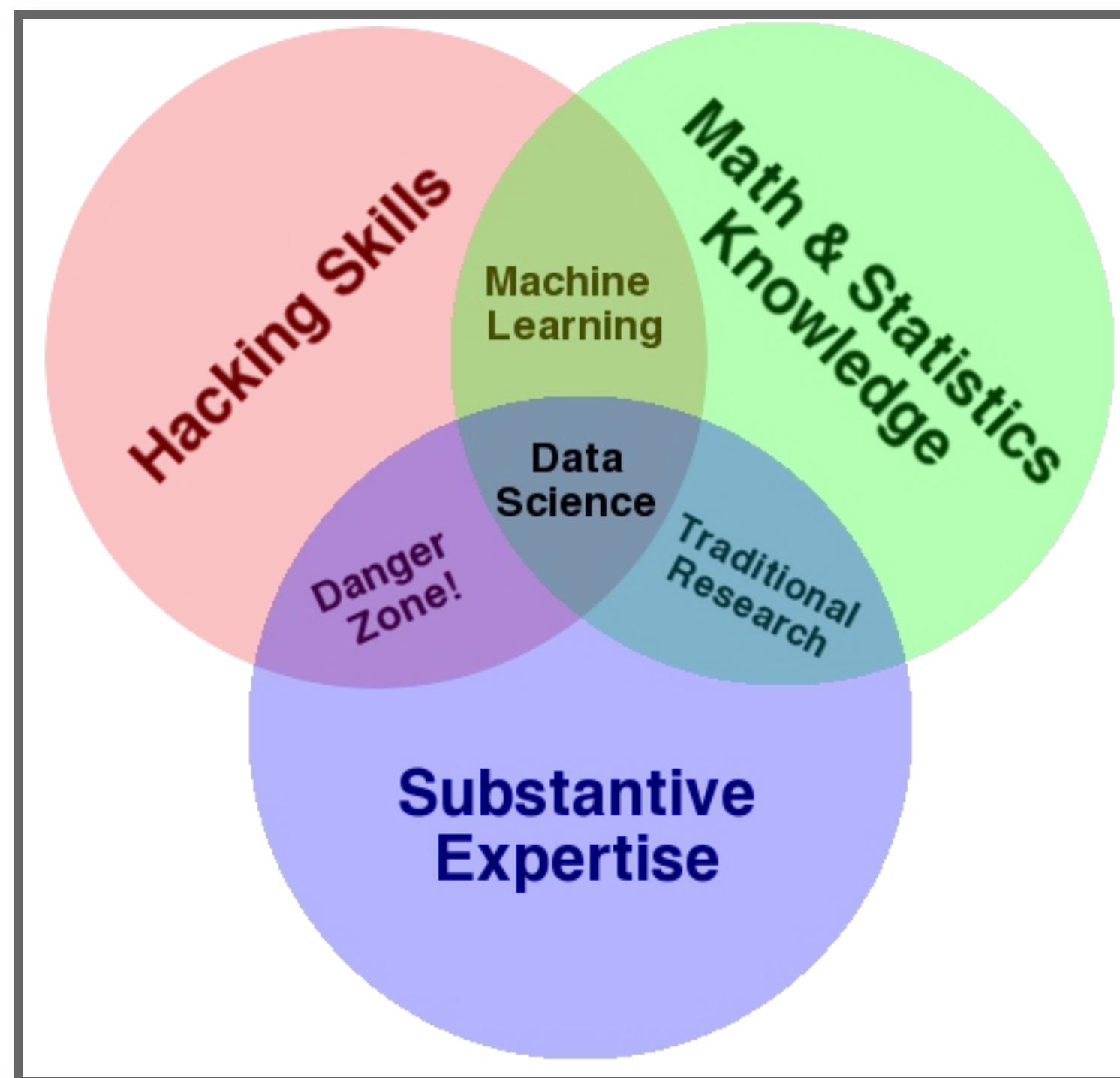
*Methods, tools and techniques to turn data into  
actionable knowledge*

But wait, isn't statistics just that?

Not only...

# Data Science

Source: Drew Conway



# Data Science

Statistics is a *very* important part of DS...

... but not the only one:

- **Computational tools** --> Programming (hence this course's tutorials!)
- **Communication skills** --> "Story telling" (hence this course's assignments)
- **Domain expertise** --> Theories about why the data are the way they are (hence the rest of your degree)

Some examples...

## Frequently Bought Together



+



+



Total price: £32.97

Add all three to Basket

*i* These items are dispatched from and sold by different sellers. [Show details](#)

- This item:** Green and Black's Organic Dark Chocolate 85 Percent Cocoa 100 g (Pack of 5) £11.62 (£2.32 / 100 g)
- [Green and Black's Organic Ginger Dark 100 g \(Pack of 5\)](#) £10.40 (£2.08 / 100 g)
- [Green and Black's Organic Dark Chocolate Maya Gold 100 g \(Pack of 5\)](#) £10.95 (£2.19 / 100 g)

## Customers Who Bought This Item Also Bought



Green and Black's Organic  
Ginger Dark 100 g (Pack of  
5)  
 15  
£10.40



Green and Black's Organic  
Dark Chocolate Maya Gold  
100 g (Pack of 5)  
 5  
£10.95



Green and Black's Organic  
Dark Chocolate 100 g  
(Pack of 5)  
 22  
£8.20



Vivani Organic Dark  
Chocolate with 85% Coco  
100 g (Pack of 5)  
 25  
£11.95

Have an account?

[Sign in](#)

# okcupid

Join the best free dating site on Earth.

I am a

Straight

Woman

[Continue](#)



Signing up takes two minutes and is totally free.



Our matching algorithm helps you find the right people.



iOS or Android?  
You can take us to go.

# Geo-Data Science

# Geo-Data Science

- A (very) large portion of all these new data are inherently **geographic** or can be traced back to some location over space.
- Spatial is special.
- Some of the methods require an explicitly spatial treatment --> (Geo-)Data Science

Some examples...

Where are you going?

Browse ▾

Sign Up Log In Help ▾

Saved

# Find a Neighbourhood in London

What kind of neighbourhood are you looking for?

Dining X

Artsy 7

Nightlife 9

Shopping 15

Loved by Londoners 11

Touristy 14

Great Transit 13

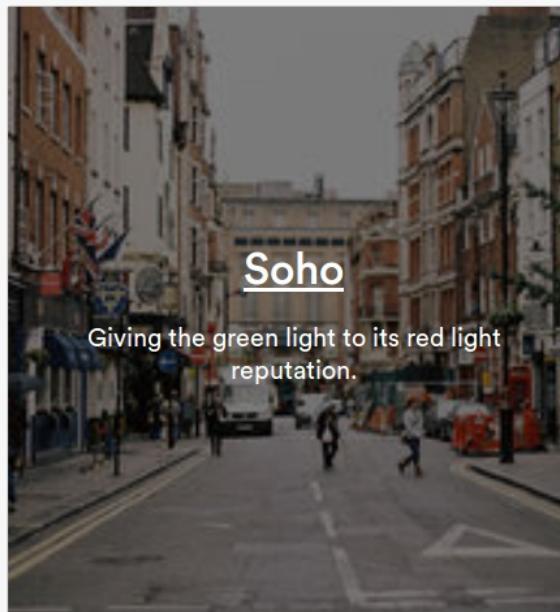
Museums &amp; the Arts 10

Peace &amp; Quiet 7

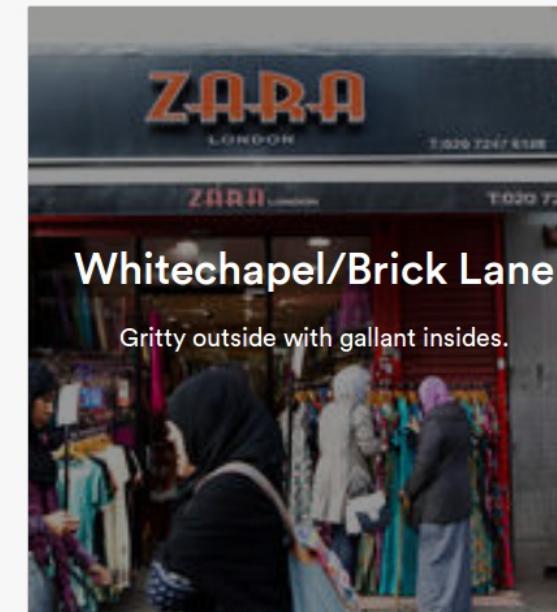
Along the River Thames 5

West End 7

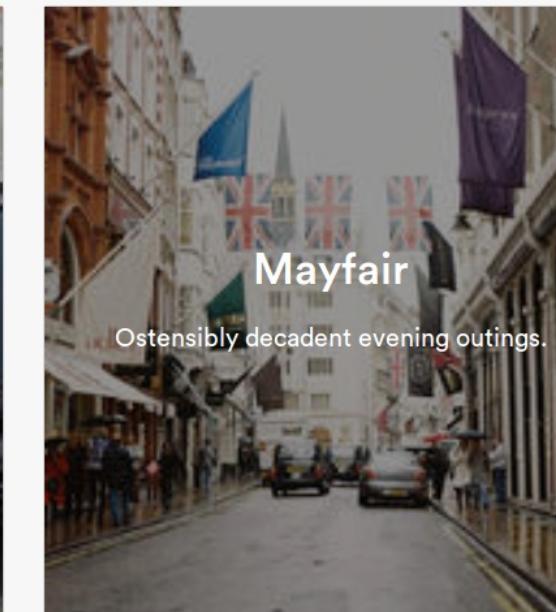
23 neighbourhoods match Dining. [See all listings](#)



Eclectic • Open-minded • China Town • Lgbt Friendly



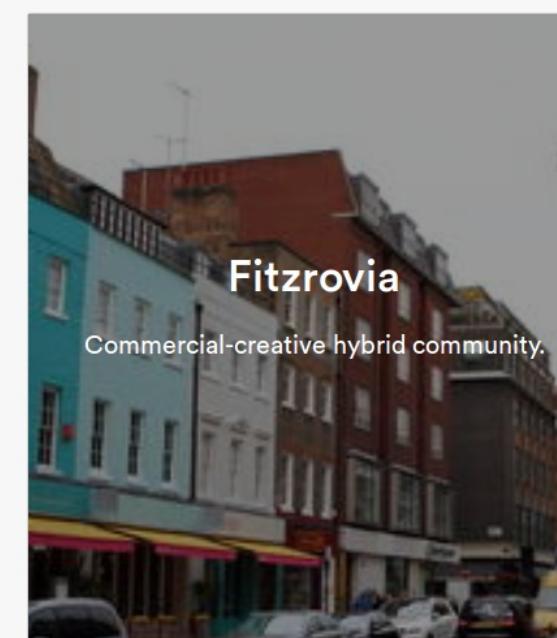
Buzzing • Curry • Arty • Graffiti



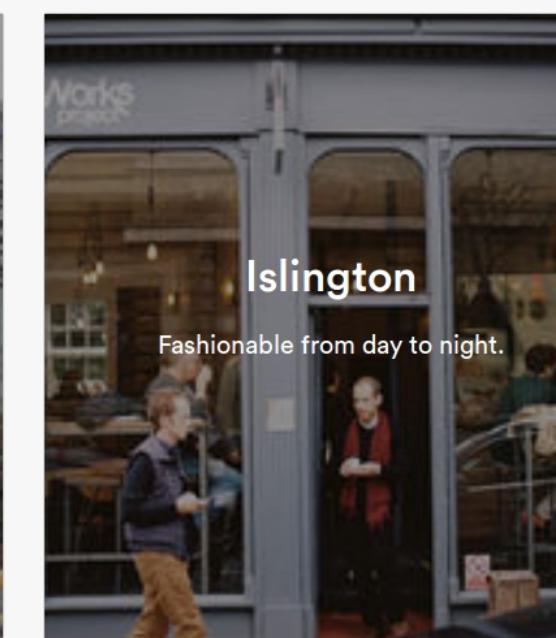
Luxury Shopping • Ballrooms • Heart Of London • Fashionable



Bohemians turned businessmen.



Commercial-creative hybrid community.



Fashionable from day to night.



Penny Ln, Live... x +

<https://www.google.co.uk/maps/dir/Penny+Ln,+Liverpool+L18+1DE,+UK/University+of+Liverpool,+Liverpool/@53.3974201,-2.9788596,13z>

## Walking directions



via A562

49 min

2.4 miles

caution - may involve errors or sections not suited for walking

## Penny Ln

Liverpool L18 1DE, UK

Head southwest on Penny Ln toward Russell Rd

79 ft

Turn right onto Russell Rd

0.2 mi

Turn left onto Smithdown Rd/A562

1.3 mi

Slight right to stay on Smithdown Rd/A562

i Continue to follow A562

364 ft

Turn right toward Smithdown Ln

325 ft

Continue onto Smithdown Ln

0.2 mi

Turn left to stay on Smithdown Ln

0.5 mi

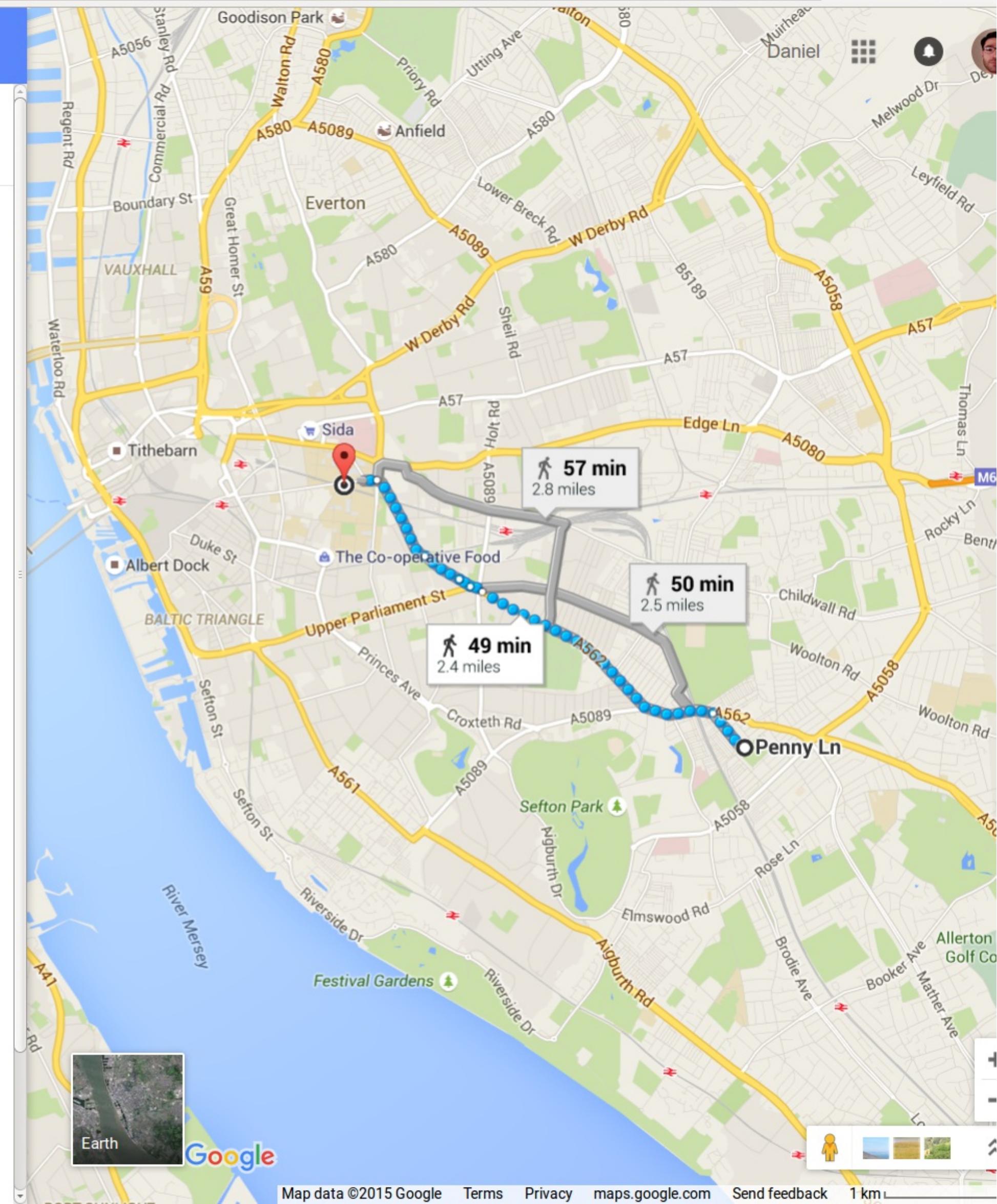
Continue straight onto Brownlow Hill/A5048

i Continue to follow Brownlow Hill

325 ft

## University of Liverpool

Liverpool, Merseyside L69 3BX, United Kingdom



Google

Map data ©2015 Google

Terms

Privacy

maps.google.com

Send feedback

1 km





Geographic Data Science'15 - Lecture 1 by [Dani Arribas-Bel](#) is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](#).