

Introduction to *Urban Data* Science

Introduction

(EPA1316)

Lecture 1

Trivik Verma



NELSON MANDELA 1918–2013

Designed by Sean Perez

How can we move forward together?

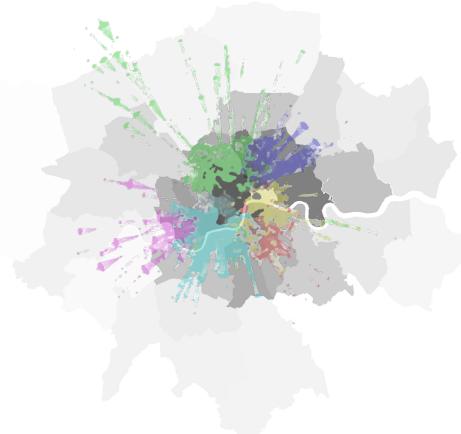
- You do not owe me “productivity” or “efficiency”. I just want your time and presence, so we can all learn something new.
- You do not owe me any information concerning your personal situation or mental or physical health condition.
- You are welcome to talk to me about anything you are going through. Just drop me an email, and we will figure it out from there.
- If you need extra help, or you have to miss class, or you need more time for something, just ask me. I will work with you, I promise.
- There is **no** grade for participation in this class. I hope the material is good enough to keep you interested.
- There is **no** exam that focusses on your ability to regurgitate facts and figures in a 3-hour window. The final project will hopefully provide you an opportunity to practice the real chaos of the real world.

Education + Work



Center for Urban
Science & Policy

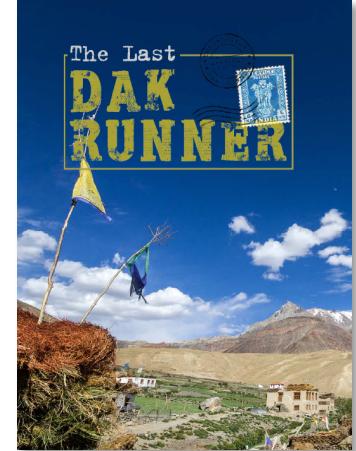
CUSP



Teaching
photography

Startup

The
OUTDOOR
Journal & Voyage



Now also generate
DATA (sensors + citizens)

Draw meaningful
insights from "Accidental"
DATA

Leverage large
DATA

Translate raw DATA
to insights into people
+ cities

My
Research

Real world
Urban problems
→ Inequalities
→ sustainability

Design "open-source"
applications for urban
planning.

Develop neat + novel
algorithms
"reproducible"

Teaching Support



Tess Kim
2nd year EPA



Lotte Lourens
2nd year EPA



Aarthi Sundaram
2nd year EPA



Giulia Reggiani
PhD @ Civil Engineering



Jelle Egbers
2nd year EPA



Bramka Jafino
PhD @ Policy Analysis



Amir Ebrahimi Fard
PhD @ CUSP lab



Talia Kaufmann (Visiting)
PhD @ Network Science
Institute, Northeastern
University

You?

- What is your educational background?
- Which language do you have experience with?

Go to www.menti.com and use the code 348967



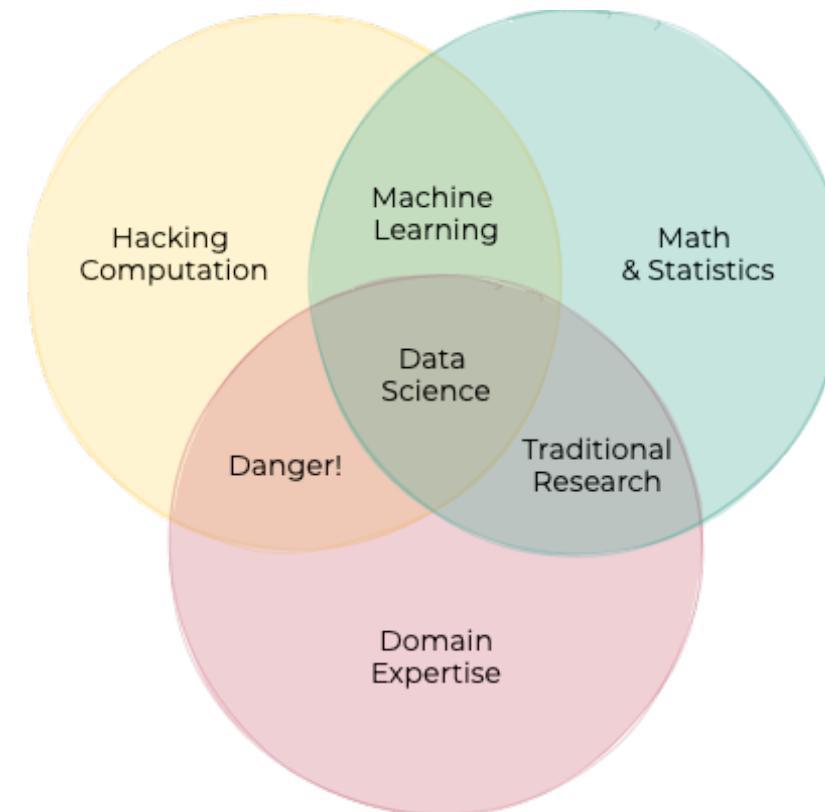
Today

- Introduction to the Course
- The (Geo-)Data Revolution
- (Geo-)Data Science
- Tools - Python and Conda

Introduction to the Course

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

With a few twists!



After this course

You will be able to...

- **Obtain**: Obtaining data from multiple **open** data sources.
- **Scrub**: Data cleaning, munging, sampling to consolidate all information into a dataset that is manageable, informative and relates to your problem.
- **Explore**: Exploratory data analysis to make sense of what your data is trying to say.
- **Model**: Estimation and modelling based on statistical tools such as regression and clustering.
- **Interpret**: Communicating results and reflections through visualisation, storytelling and interpretable summaries.

Philosophy

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

Format – Ways of Working

Eight weeks of:

- **Prep. Materials**: videos, podcasts, articles... 1h. approx. (most recommended!)
- **14x 1h. Lecture**: concepts, methods, examples
- **2h. Computer labs**: hands-on, application of concepts, Python (*highly employable*)
- **1h. Paper Discussions**: reading a paper and debating a set of questions with your peers in small groups. (extremely important if you are interested in applying concepts to real-world problems).
- **Further readings**: how to go beyond the minimum*

*(Very useful for the final project. If we don't read, we cannot write!)

Content

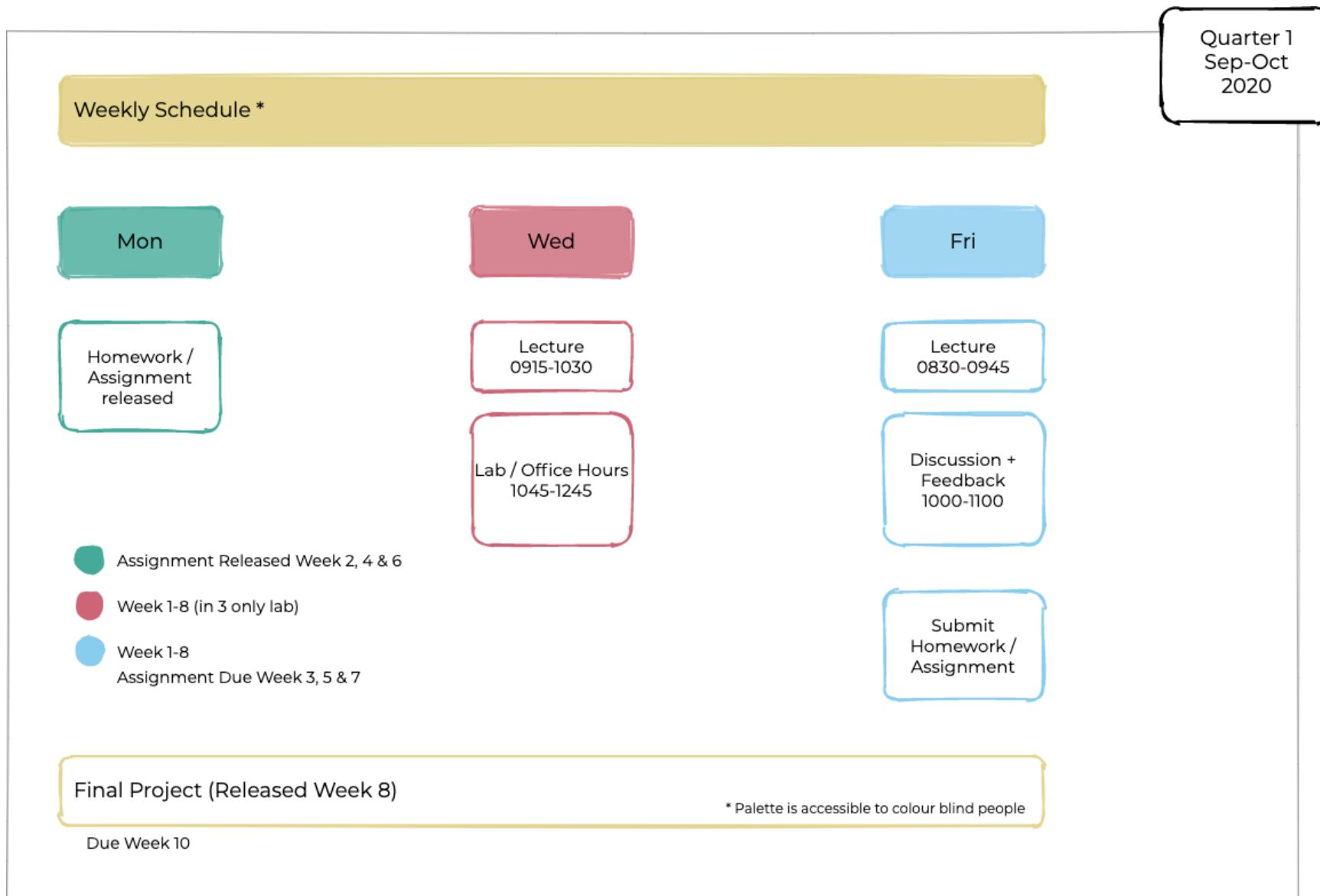
- **Weeks 1-3:** “big picture” lectures + introduction to computational tools (learning curve) + lots and lots of data
- **Weeks 4-7:** lots of spatial, network and machine learning concepts + lots of visualisation
- **Weeks 8-10:** wrap up + prepare an awesome final project in groups (follow up with internships in the ***CUSP*** lab)

An overview of the course

- **Weeks 1-3:** “big picture” lectures + introduction to computational tools (learning curve) + lots and lots of data
- **Weeks 4-7:** lots of spatial, network and machine learning concepts + lots of visualisation
- **Weeks 8-10:** wrap up + prepare an awesome final project in groups (follow up with internships in the ***CUSP*** lab)

Week	Lecture	Topic	Python Libraries	Labs and Homework ¹	Assessment ²
W 1	L1	Introduction to Urban Data Science	Anaconda and Jupyter, Numpy	Lab 1	
W 2	L2	Spatial and Urban Data	Pandas, Seaborn	Lab 2	
	L3	Data Grammar			
W3	L4	Data Engineering	Pandas	Lab 3	Assignment 1
W4	L5	EDA and Visualisation	Geopandas, Matplotlib, Rasterio	Lab 4	
	L6	Geo-Visualisation			
W 5	L7	Networks and Spatial Weights	Networkx, Osmnx, Pysal	Lab 5	Assignment 2
	L8	Exploratory Spatial Data Analysis			
W 6	L9	Machine Learning for Everyone	Sklearn, Scipy, Statsmodels	Lab 6	
	L10	Anatomy of a Learning Algorithm			
W 7	L11	Clustering	Pysal, Sklearn-Cluster	Lab 7	Assignment 3
	L12	Dimensionality Reduction			
W 8	L13	Spatial Density Estimation	More Sklearn	Lab 8	
	L14	Responsible Data Science			
W 10					Final Project ⁴

Schedule



Logistics - Website

- **Course Material:** <https://research.trivikverma.com/courses/epa1316-2020/>
- **Lectures and In-class Labs:** [Virtual Classroom on Brightspace](#)
- **Communication Channels:** [Discussion Forums on Brightspace](#)
- [Announcements on Brightspace](#)
- **Groups, Submissions and Feedback:** Brightspace

The screenshot shows the homepage of the course website. At the top right, there is a navigation bar with links for About, People, Research (with a dropdown arrow), News, Blog, Openings, a Twitter icon, and a search icon. On the far left, there is a sidebar with a search bar and several menu items: Introduction, Overview, Syllabus, Schedule, To Begin With, Lectures, Discussions, Software, Labs, and FAQ. The main content area features a large title "EPA 1316 Introduction to Urban Data Science". Below the title, there is a paragraph of text about the course's history and current name. Underneath the text, there are two sections: "Instructor" (listing Dr. Trivik Verma's contact information) and "Contact" (listing his office details). To the right of the main content, there is a vertical sidebar titled "Contents" which lists various course components.

About People Research ▾ News Blog Openings

Search...

Introduction

[Overview](#)

[Syllabus](#)

[Schedule](#)

[To Begin With](#)

Lectures

[Discussions](#)

[Software](#)

[Labs](#)

[FAQ](#)

EPA 1316 Introduction to Urban Data Science

Welcome to EPA 1316 at Delft University of Technology. The course used to be called Data Analytics and Visualisation and has concurrently been named Introduction to *Urban Data Science*. The entire course will be held online.

Instructor

Dr Trivik Verma - t.verma [at] tudelft.nl

Contact

Assistant Professor in Computational Urban Science & Policy
B1.290, Building 31
Faculty of Technology, Policy and Management
Jafflaan 5
2628 BX Delft
The Netherlands

Contents

[Instructor](#)

[Locations](#)

[Teaching Assistants](#)

[Course Language](#)

[Why Python](#)

[Expected prior knowledge](#)

[Philosophy of the course](#)

[Feedback strategy](#)

[Key texts and learning resources](#)

[Acknowledgement](#)

[License](#)

Self-directed learning

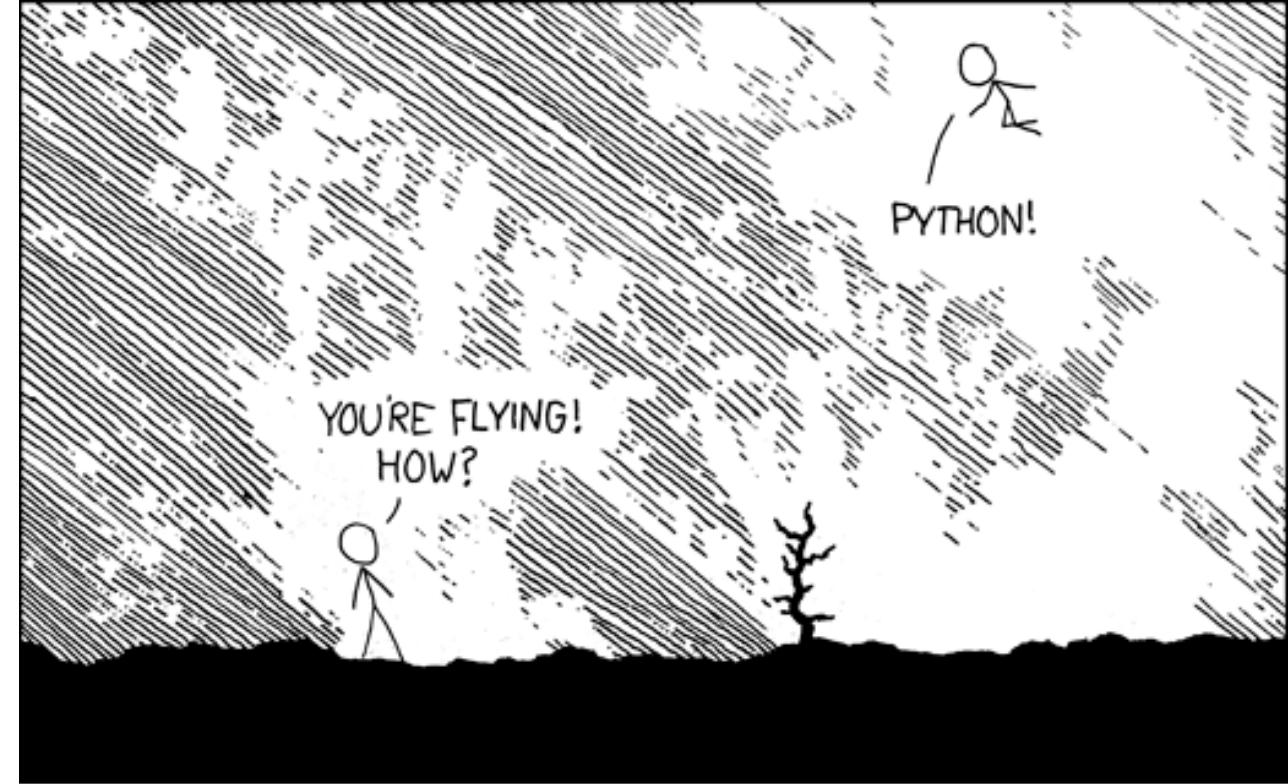
This course is much more about “**learning to learn**” and problem solving rather than acquiring specific programming tricks or stats wizardry

- **Prepare** for the labs
- **I won't** be leading/lecturing at the computer labs. TAs will be present for abundant help and feedback.
- **Go over the notebooks** before the lecture and the computer lab
- If the first time you see a notebook is at the lab, you may find it difficult to follow on. The best thing to do is to prepare a set of questions to ask us.
- **Bring** questions, comments, feedback, (informed) rants to class/labs. The more you bring, the more we all learn.
- **Collaborate** (it's **NOT** a zero-sum win!)

Python

- General purpose programming language
- “Sweet spot” between “*proof-of-concept*” and “*production-ready*”
- Industry standard: **GIS** (Esri, QGIS) and **Data Science** (510, World Bank, OECD, The Atlantic, Gemeente Den Haag...)

How many of you have written a line of computer code before?



I LEARNED IT LAST NIGHT! EVERYTHING IS SO SIMPLE!
/ HELLO WORLD IS JUST
print "Hello, world!"

I DUNNO...
DYNAMIC TYPING?
WHITESPACE?
COME JOIN US!
PROGRAMMING IS FUN AGAIN!
IT'S A WHOLE NEW WORLD UP HERE!
BUT HOW ARE YOU FLYING?

I JUST TYPED
import antigravity
THAT'S IT?
/ ... I ALSO SAMPLED
EVERYTHING IN THE
MEDICINE CABINET
FOR COMPARISON.
/ BUT I THINK THIS
IS THE PYTHON.

Assessments

Formative : These are ungraded

- 8 In-Class Labs
- 8 Peer-Reviewed Homework Questions
- Discussion Sessions

Summative : These are graded components

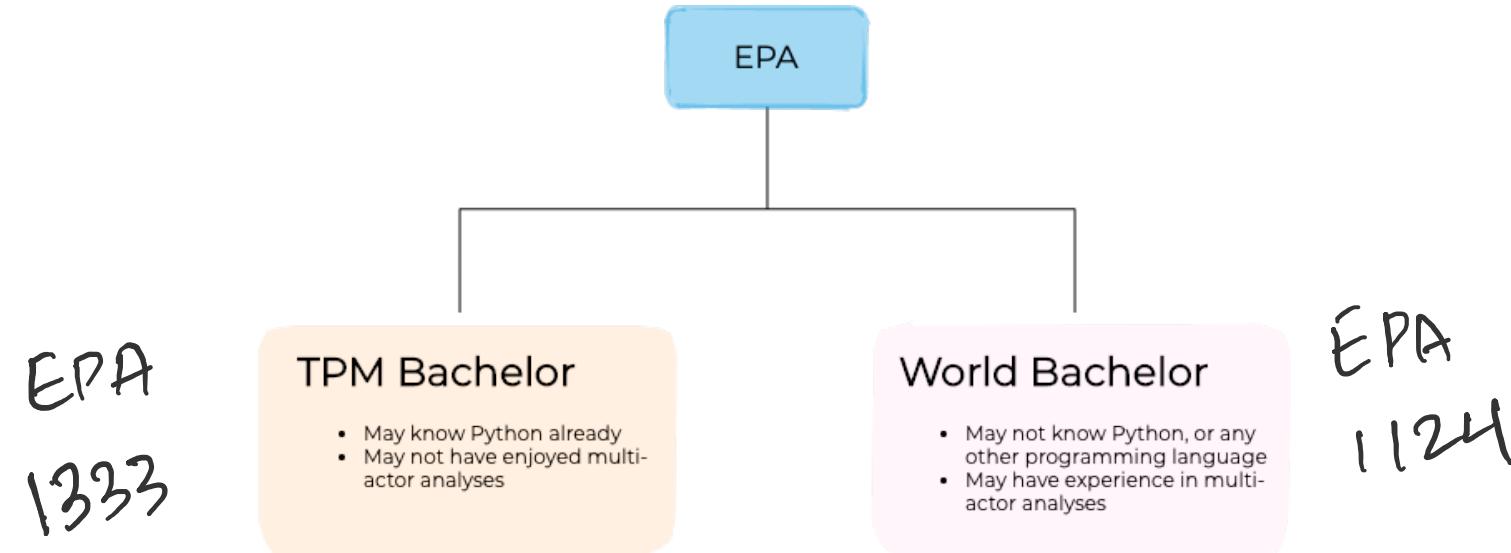
- Assignment 1 (15%)
- Assignment 2 (15%)
- Assignment 3 (10%)
- Final Project (60%)

More Help!!!

- Ask questions
- Help others as much as you can (the best way to learn is to teach)
- Search heavily on Google + Stack Overflow (learn to troubleshoot)
- Bring questions, comments, feedback, (informed) rants to class
- Collaborate with each other
- Use the Discussion Forum **heavily**



EPA Elephant



I am going to give you enough opportunity to
both learn **Python** and gain perspective on
multi-actor systems

Break



WATER



WALK



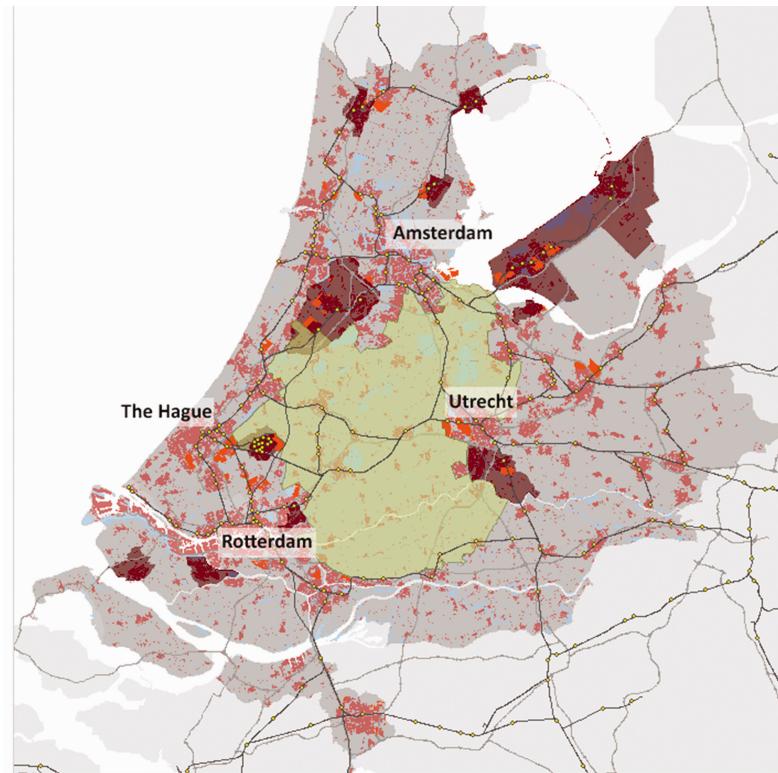
COFFEE OR TEA



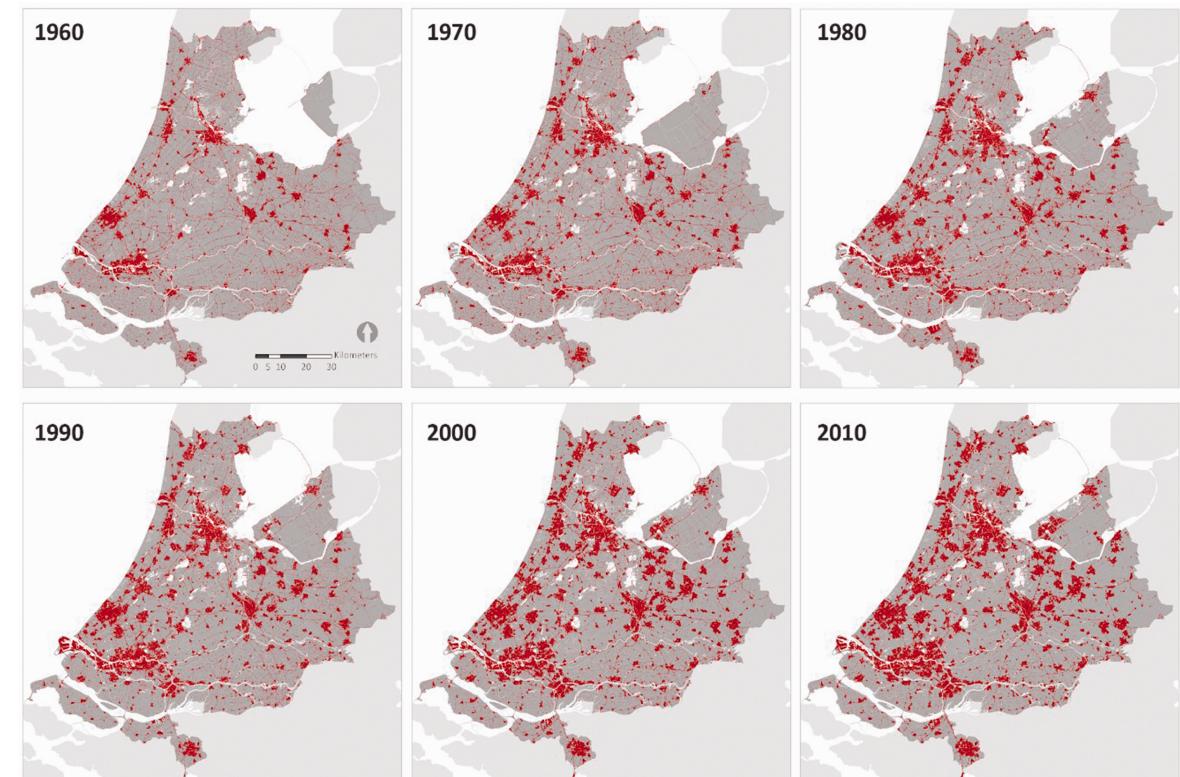
MAKE FRIENDS

The (Geo-)Data Revolution

Example: Randstad



Urbanisation of the Greater Randstad Area 1960-2010



The data revolution

Exciting times to be a:

- Data Scientist
- Urban Planner
- Policymaker

The world is producing a lot of “**data**”...

Massive Data Revolution

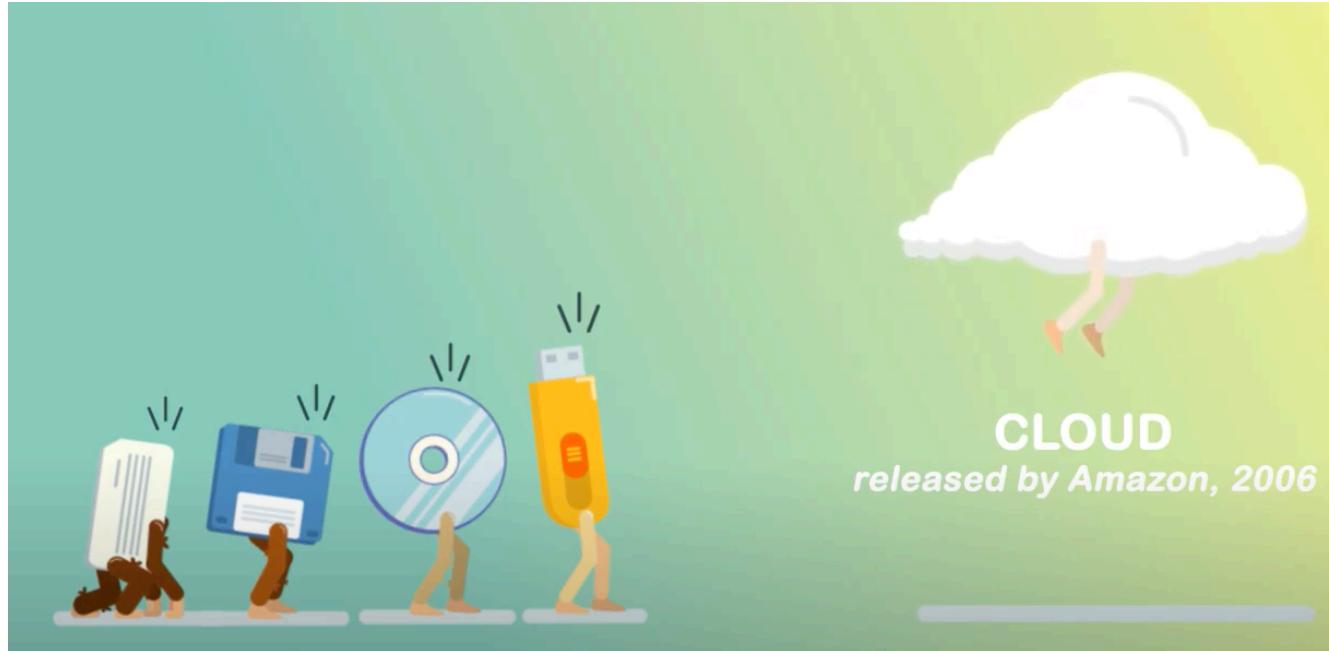
Quantification of phenomena through the systematic recording of data, “taking all aspects of life and turning them into data ([Cukier & Mayer-Schoenberger](#))

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

Implications

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

Why now?



Statistics

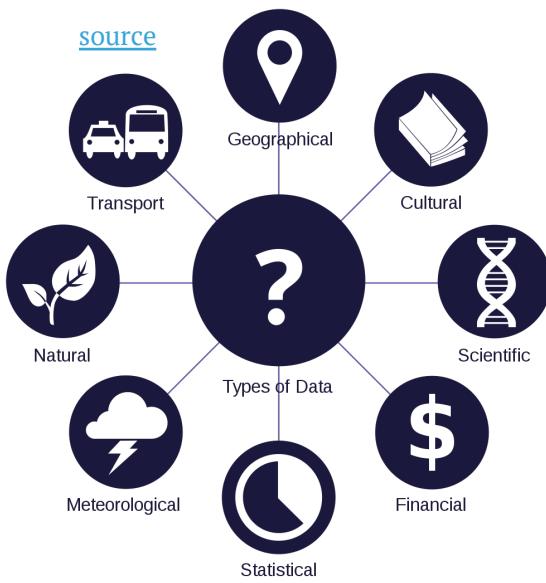
Machine
learning

What's
next?
o

- o Massive Data generation
- o Computing power
- o R + Python
- o Visualisation

Between the dawn of civilization and 2003, we only created five exabytes of information; now we're creating that amount every two days.

Eric Schmidt, Google

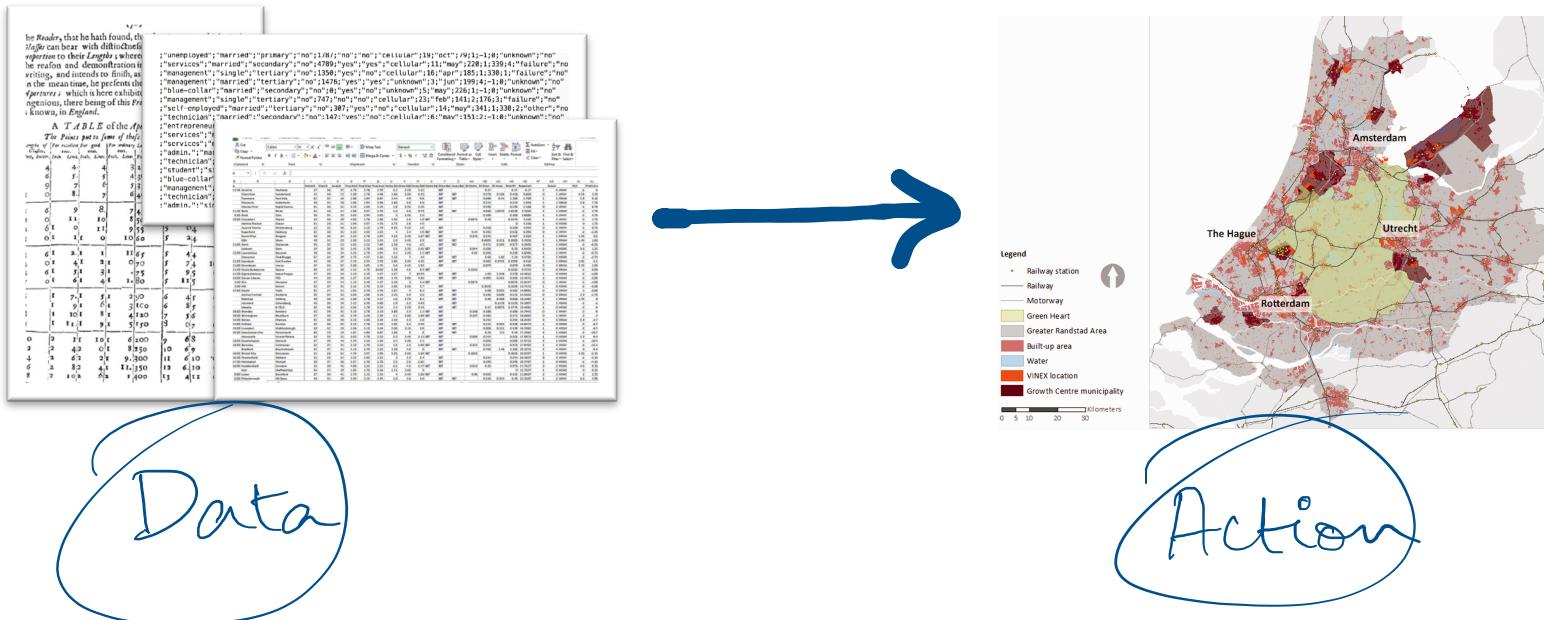


Examples: credit transactions, public transit usage, tweets, census, mobility and migration, etc.

A screenshot of a Microsoft Excel spreadsheet titled "A TABLE of the Apertures". The table has 20 rows and 15 columns. The first column contains numerical values from 1 to 20. The second column contains letters from A to Z. The third column is labeled "Length" and the fourth column is labeled "Width". The data in the table represents a grid of values, likely representing aperture sizes. The table includes a header row and a footer row. The entire table is contained within a single cell, demonstrating the exponential growth of data storage over time.

Formats: CSV, Excel, JSON, Shapefiles

Now, data alone is not very valuable



Methods, tools and techniques to turn data into
actionable knowledge (hence this lesson!)

Class Quiz

Can you think of a real-world context where data and statistics are being used to make a difference?

- Enter a breakout room and discuss for two minutes
- Then I may ask you to summarise your discussion



Data Science

*Methods, tools and techniques to turn data into
actionable knowledge (hence this lesson!)*

Statistics + ...

- **Computational** tools → Programming (hence this course's labs and homework!)
- **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...

Emmy-winning US TV Shows



Police Detective TV Dramas



Critically Acclaimed Witty TV Shows



Free Online Dating | OkCupid - Mozilla Firefox (Private Browsing)

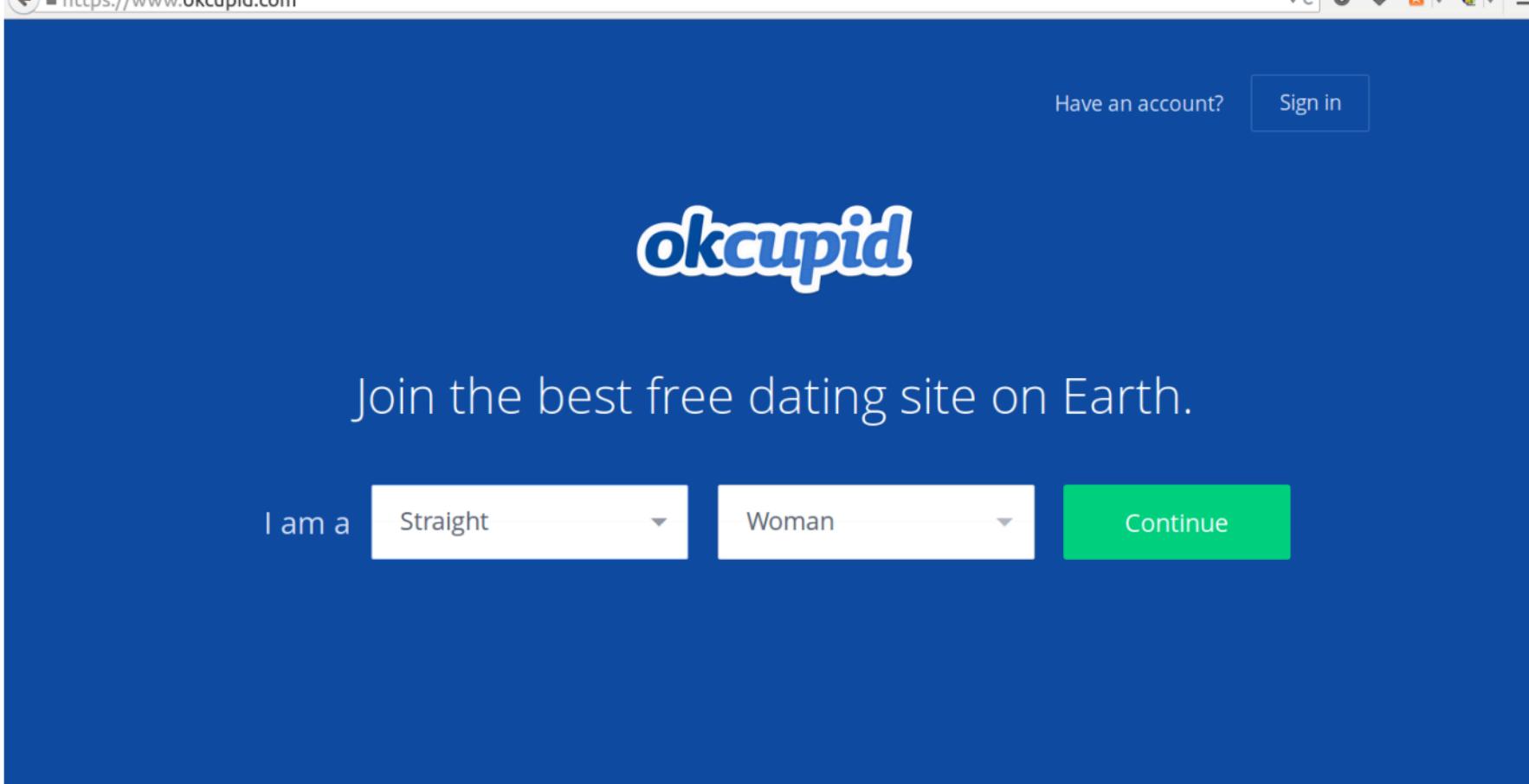
Free Online D... https://www.okcupid.com

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.

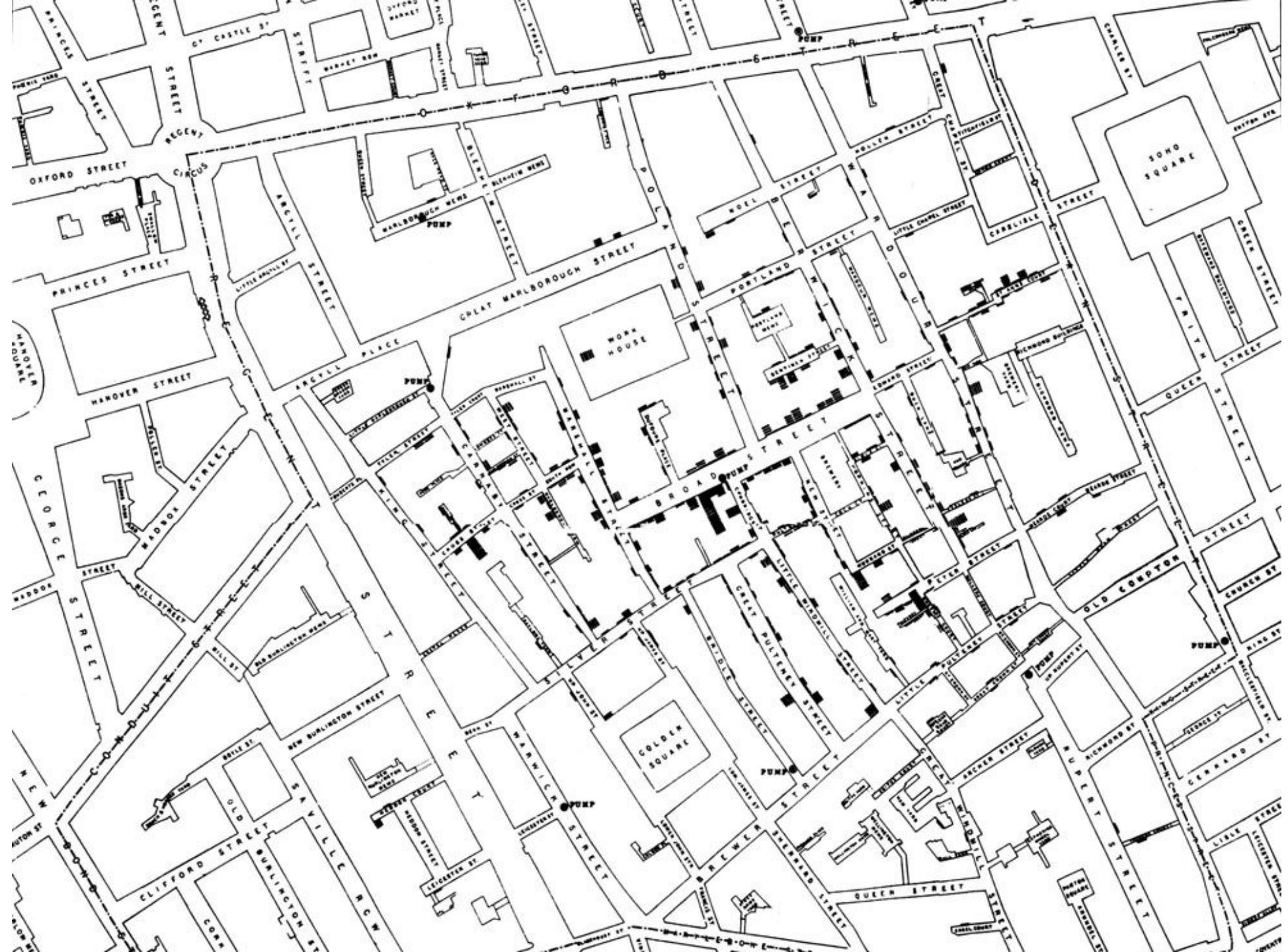


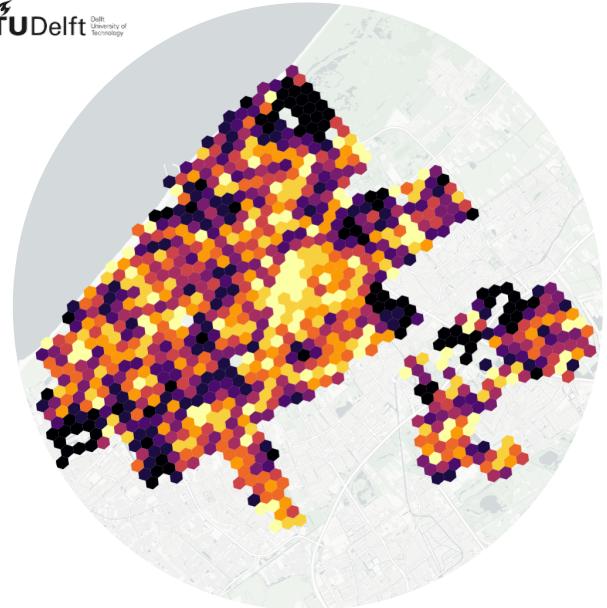
iOS or Android?
You can take us to go.

(Geo-)Data Science

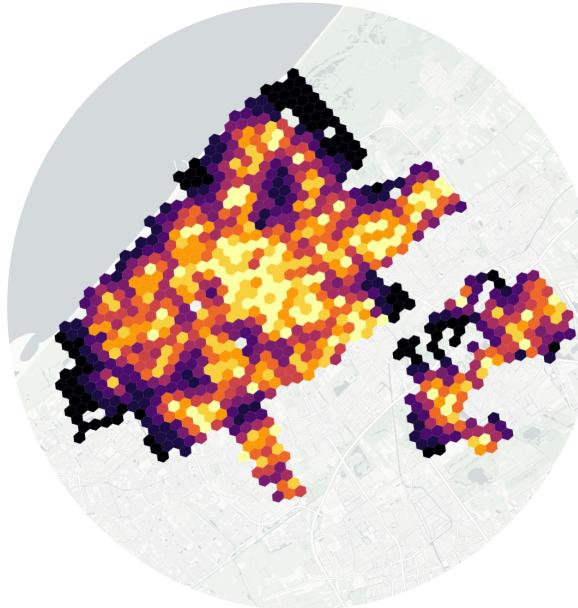
(Geo-)Data Science

- 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...

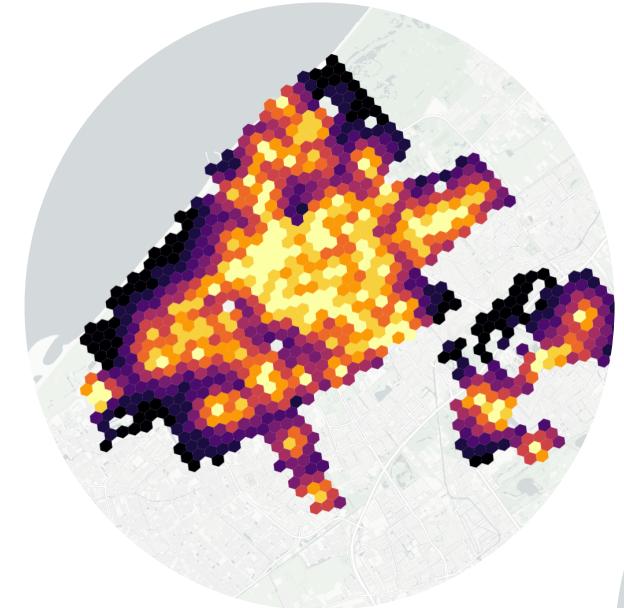




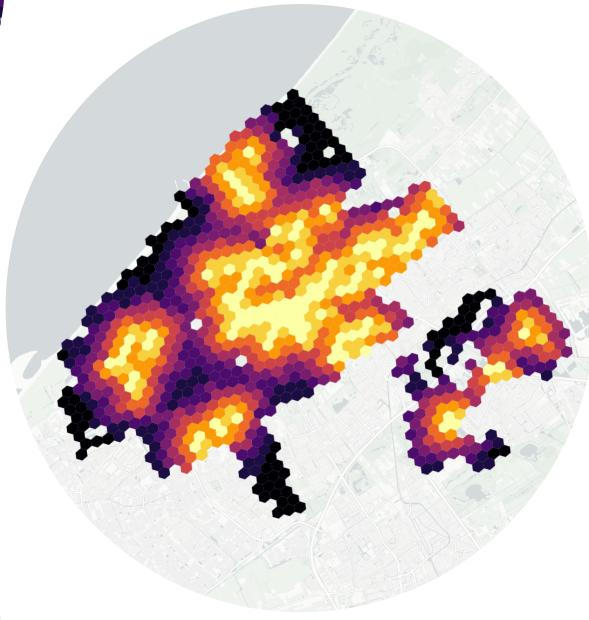
Active Living



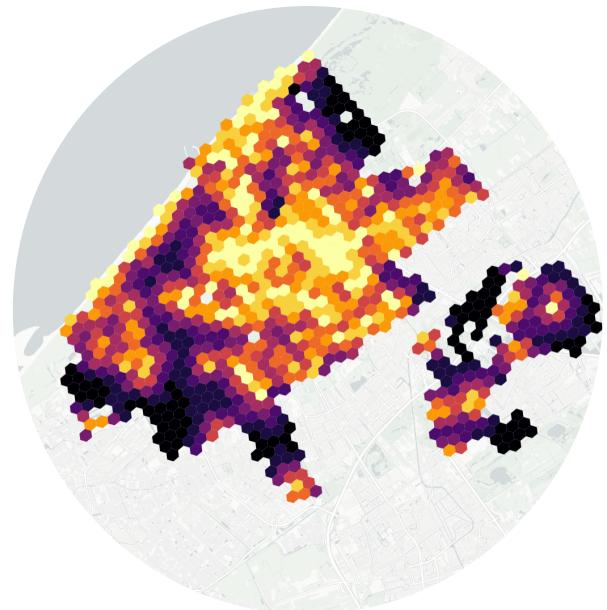
Education



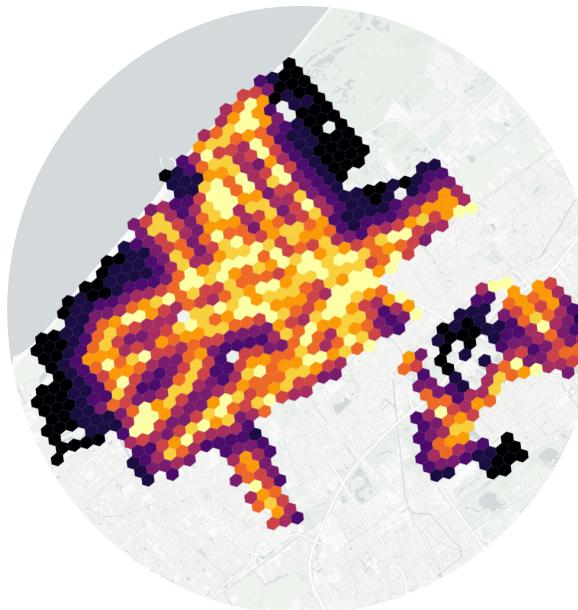
Health and Well Being



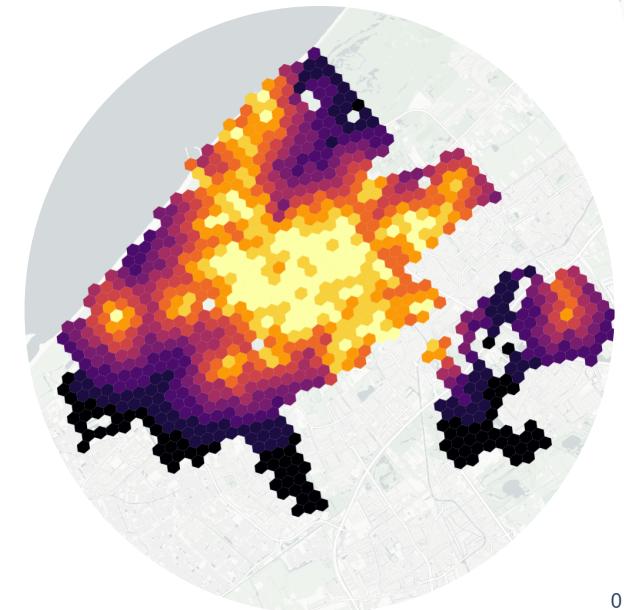
Nightlife



Food Choices



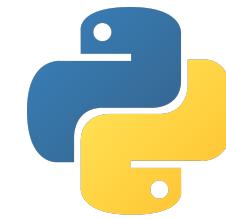
Mobility



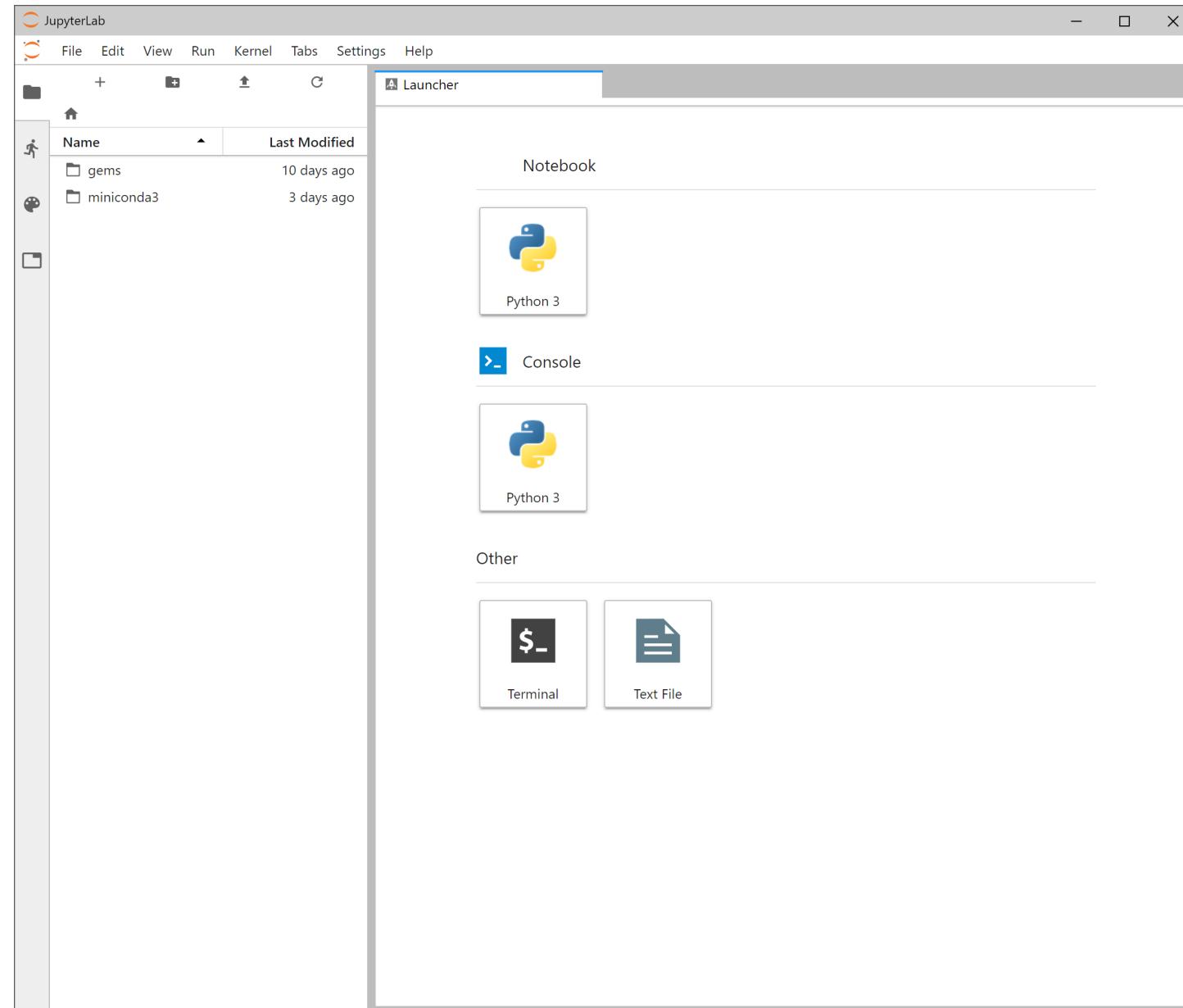
Community Space



Installing Python



<https://research.trivikverma.com/courses/epa1316-2020/software/>



For next class..



Finish Lab 01 to practice programming



Submit Homework 01 for peer review on Brightspace



Check Assignment 1 – due in **Week 3** on Friday at **2330**



See “To do before class” for every lecture (~ 1 hour of self study)



Read paper for **Discussion** session before every Friday



Post questions on the **Discussion** forum on Brightspace (especially on **Anaconda/Jupyter** section for this week)