



# From (lat, lon) to livelihoods: data science, geo data & their ethics

**Dr Marc Cheong**  
Senior Fellow (MLS) & Research Fellow Digital Ethics (CAIDE) &  
Honorary Burnet Institute Senior Fellow

26 May 2021





# About me

CompSci - Did Twitter research ca. 2009-2013

- Was a wonderful world back then, generous support by Twitter, less toxicity online
- Lucky to jump in then, ~3 dozen papers or so worldwide!

Pivoted to philosophical research

- Introducing data science to philosophers (digital humanities/experimental philosophy)
- Digital Ethics (philosophical ethics + data science + practical aspects)
- Contemporary social media usage and social networking trends.



# Overview

Geo data – from GIS's to Tableau

- Applications: geography/agriculture, COVID information and public health...
- Dangers: security/cybersecurity, privacy...

## **Case study: Social media metadata - interactive lecture**

- Twitter and the 2011 London Riots – law enforcement
- Studying the public reaction to the Black Lives Matter – movement for equality & social good
- **The ethics behind this!**

## **Enter: Digital ethics - interactive lecture**

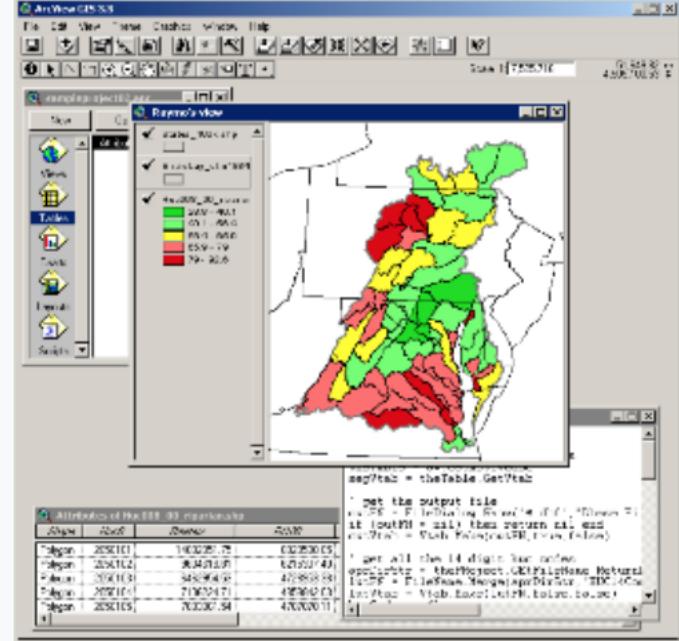
- Mitigating risks – techniques?
- Coghlan et al – the ethical principles
- Further Discussion.



# Geo data - from GIS's to Tableau

# The power of geo data

## ArcView GIS 3.x



**ArcView GIS showing data for the Chesapeake Bay.**

Region	Area	Owner	Attr
Region 1	10000.00	00000000	
Region 2	10000.00	00000000	
Region 3	10000.00	00000000	
Region 4	10000.00	00000000	
Region 5	10000.00	00000000	

**Developer(s)** ESRI

**Initial release** October 5, 1995

**Final release** 3.3 / May 22, 2002

## History and development [edit]

The phrase, "geographic information system", was coined by [Roger Tomlinson](#) in 1968, when he published the scientific paper, "A Geographic Information System for Regional Planning".<sup>[5]</sup> Tomlinson, acknowledged as the "father of GIS",<sup>[6]</sup> is credited with enabling the first computerized-GIS to be created through his work on the [Canada Geographic Information System](#) in 1963. Ultimately, Tomlinson created a framework for a database that was capable of storing and analyzing huge amounts of data; leading to the Canadian government being able to implement its National Land-Use Management Program.<sup>[7][6]</sup>

One of the first known instances in which spatial analysis was used, came from the field of [epidemiology](#) in the, "Rapport sur la marche et les effets du choléra dans Paris et le département de la Seine" (1832).<sup>[8]</sup> French geographer and [cartographer](#), [Charles Picquet](#), created a map outlining the [forty-eight Districts in Paris](#), using [halftone](#) color gradients, to provide a visual representation for the number of reported deaths due to [cholera](#), per every 1,000 inhabitants.

In 1854, [John Snow](#), an epidemiologist and physician, was able to determine the source of a cholera outbreak in [London](#) through the use of spatial analysis. Snow achieved this through plotting the residence of each casualty on a map of the area, as well as the nearby water sources. Once these points were marked, he was able to identify the water source within the cluster that was responsible for the outbreak. This was one of the earliest successful uses of a geographic methodology in pinpointing the source of an outbreak in epidemiology. While the basic elements of [topography](#) and theme existed previously in [cartography](#), Snow's map was unique due to his use of cartographic methods, not only to depict, but also to analyze clusters of geographically dependent phenomena.

The early 20th century saw the development of [photozincography](#), which allowed maps to be split into layers, for example one layer for vegetation and another for water. This was particularly used for printing contours – drawing these was a labour-intensive task but having them on a separate layer meant they could be worked on without the other layers to confuse the [draughtsman](#). This work was originally drawn on glass plates but



E. W. Gilbert's version (1958) of John Snow's 1855 map of the Soho cholera outbreak showing the clusters of cholera cases in the London epidemic of 1854

[https://en.wikipedia.org/wiki/ArcView\\_3.x](https://en.wikipedia.org/wiki/ArcView_3.x)

[https://en.wikipedia.org/wiki/Geographic\\_information\\_system](https://en.wikipedia.org/wiki/Geographic_information_system)

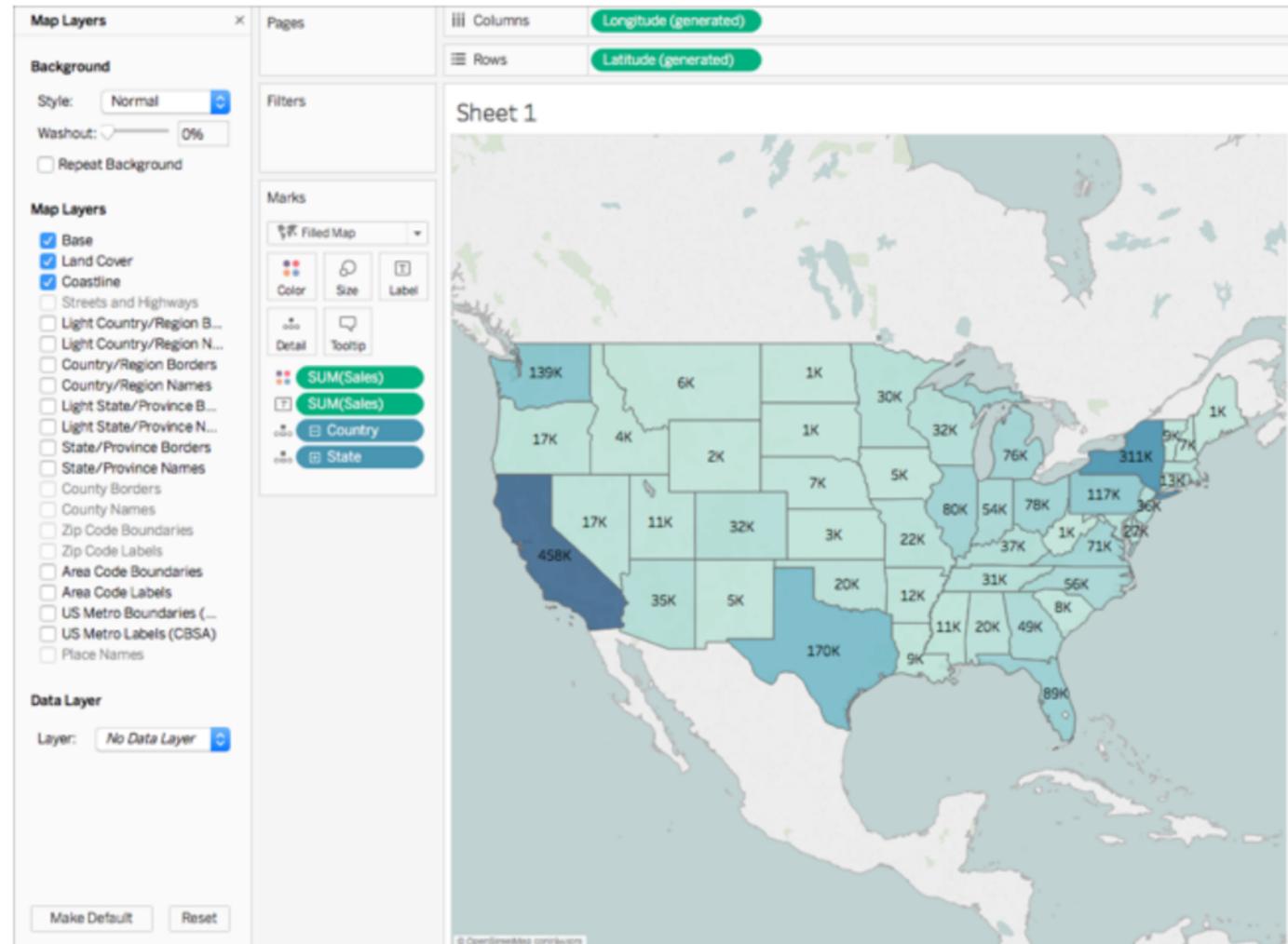
# The power of geo data

Companies, organisations all start to value geo data.

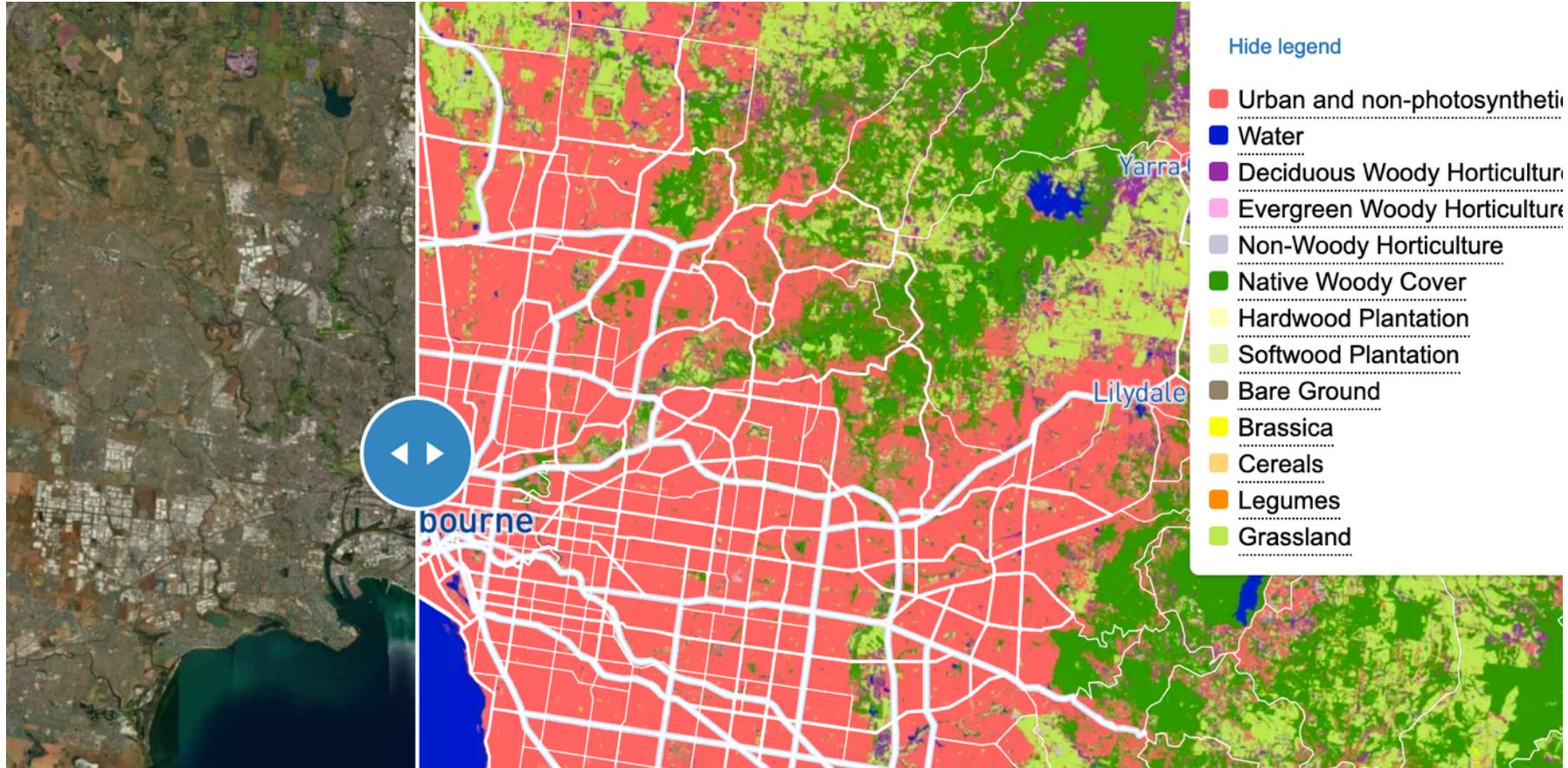
No longer the exclusive domain of cartographers/geographers/Geog. Info Systems (GIS) experts...

Consider Tableau 

[https://help.tableau.com/current/pro/desktop/en-us/buildexamples\\_maps.htm](https://help.tableau.com/current/pro/desktop/en-us/buildexamples_maps.htm)

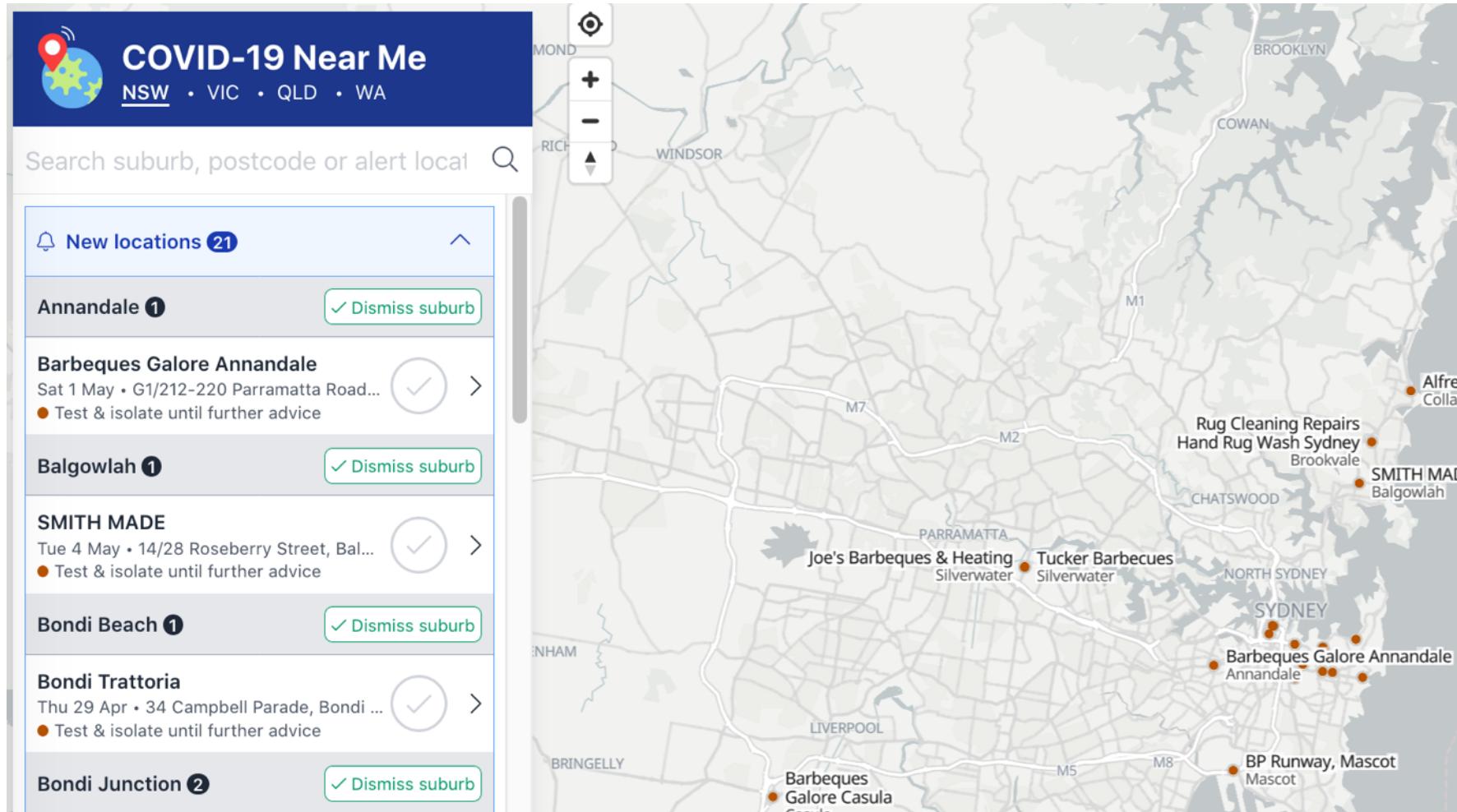


# Applications: from agriculture...



Charlotte Pelletier, Zehui Ji, Olivier Hagolle, Elizabeth Morse-McNabb, Kathryn Sheffield, Geoffrey I. Webb and Francois Petitjean, [Using Sentinel-2 Image Time Series to map the State of Victoria, Australia, MultiTemp, 2019](#).

# Applications: ... to helping us in COVID-19



Aggregating data from official public health websites and visualizes them (including time metadata etc)  
Tsang (2021). <https://covid19nearme.com.au/>

# ...beware danger!

“Sensitive information about the location and staffing of military bases and spy outposts around the world has been revealed by a fitness tracking company...

...data visualisation map that shows all the activity tracked by users of its app, which allows people to record their exercise and share it with others.

**The map, released in November 2017, shows every single activity ever uploaded to Strava - more than 3 trillion individual GPS data points, according to the company..”**

Hern (2018) for The Guardian

<https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases>

----

**Ethical question: Say you're a data scientist at Strava - did you foresee this coming? What should you do?**

GPS

Alex Hern

Mon 29 Jan 2018 08.51 AEDT

This article is more than 3 years old

## Fitness tracking app Strava gives away location of secret US army bases

Data about exercise routes shared online by soldiers can be used to pinpoint overseas facilities

Latest: Strava suggests military users ‘opt out’ of heatmap as row deepens



▲ A military base in Helmand Province, Afghanistan with route taken by joggers highlighted by Strava. Photograph: Strava Heatmap

Sensitive information about the location and staffing of military bases and spy outposts around the world has been revealed by a fitness tracking company.



# Case study: Social media metadata

# Case Study: Twitter metadata London Riots (2011)

The riots took place from 6–11 August 2011 (inclusive), erupting from “a peaceful protest over the police killing of a Tottenham man, Mark Duggan” [May, 2011]. On the 7th of August, the rioting, looting, and arson started in parts of London, which subsequently spread to Birmingham, Manchester and Liverpool [May, 2011]. The riots continued spreading till 10th August, where they started to ebb due in part to bad weather [May, 2011] and police response [May, 2011; Meikle and Jones, 2011]. The riots resulted in over 1,000 arrests [Meikle and Jones, 2011], a rough estimate of 100 million British pounds in damages [Lock, 2011], and a few collateral deaths [May, 2011]. Theories on the cause of the riots range from “police prejudices, a lack of social mobility, unemployment...” [May, 2011], all the way to “welfare dependency... teenage pregnancies... [and] consumerism” [May, 2011].

What is notable in the case of the London Riots is the prominent use of social media technology. BlackBerry Messenger, Facebook, and Twitter were among the oft-mentioned technologies used in the riots for “inciting public disorder” [Meikle and Jones, 2011].

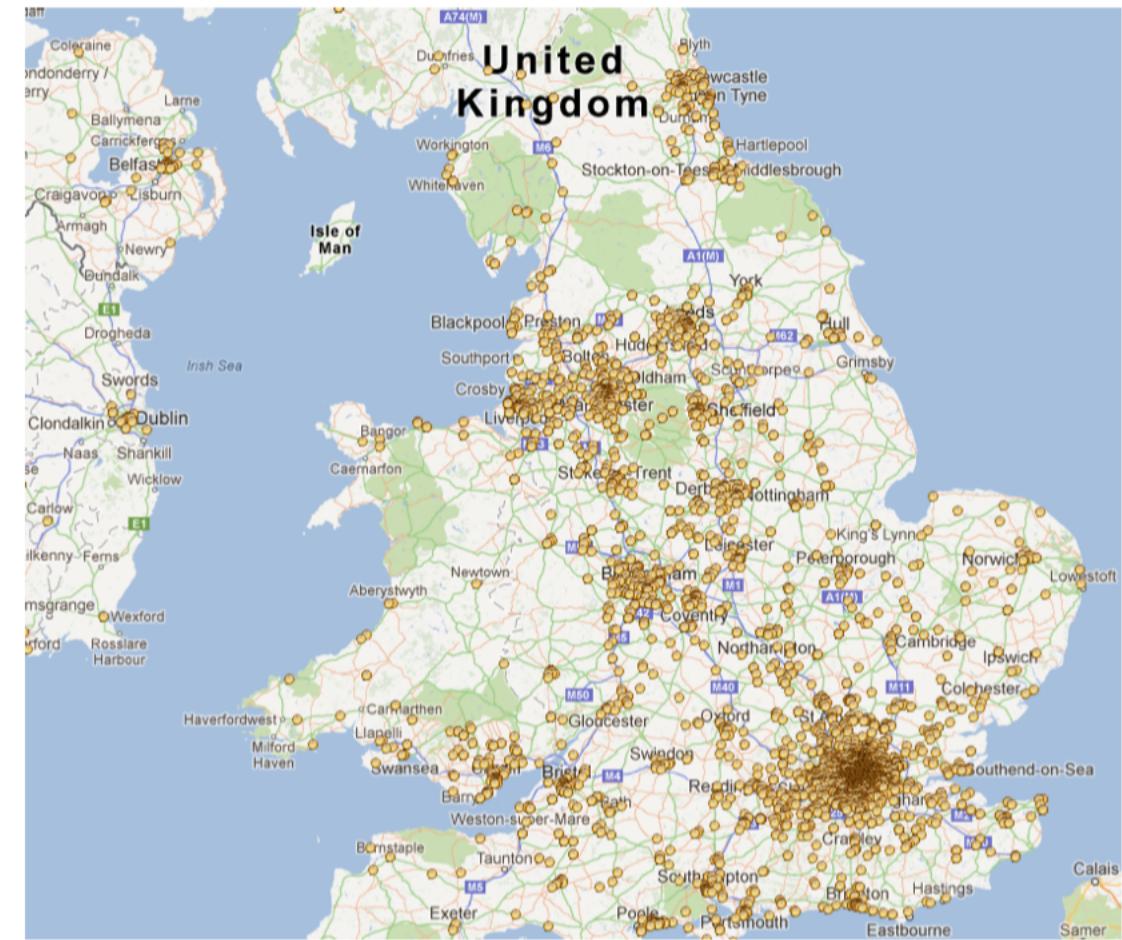


Figure 7.18: Visualization of locations found in Twitter metadata from the *London Riots Dataset*, originating from the United Kingdom. Each yellow dot represents a tweet composed in a particular location. From visual inspection, the spatial distribution of yellow dots roughly correspond to the red dots (representing riot activity) in Figure 7.17.

Cheong (2013) – first published in:

Cheong, M., Ray, S. and Green, D. [2012a]. Interpreting the 2011 London Riots from Twitter Metadata, Proc. SoCPAR 2012.

# Case Study: Twitter metadata London Riots (2011)

Table 7.15: Results of statistical tests on the total tweet count and reported riot outbreaks per unit area.

Parameter	For 1 degree $\times$ 1 degree squares	For 0.5 degree $\times$ 0.5 degree squares
Number of squares evaluated	37	114
Pearson coefficient, $r$	0.9704	0.9370
$t$ -value corresponding to $r$ (d.o.f. = number of squares-2)	23.7727	28.4027
Significant at 5% level?	Yes (exceeds required $t = 1.6896$ )	Yes (exceeds required $t = 1.6586$ )
Significant at 1% level?	Yes (exceeds required $t = 2.4377$ )	Yes (exceeds required $t = 2.3601$ )

From the results in Table 7.15, there is strong enough evidence to refute the null hypothesis that there is no correlation between tweets per unit area and the corresponding number of reported riots. It can be suggested that the correlation — between the number of tweets for a given area with the number of reported riot outbreaks in the same area — is statistically significant at the 1% level, thus accomplishing *Study Goal 3*. The results from Experiments 7.12 and 7.13 corroborates the findings from related studies mentioned earlier [Longueville et al., 2009; Starbird et al., 2010; Guy et al., 2010; Cheong and Lee, 2011] that Twitter metadata can be an accurate source of location information, useful in accurately pinpointing locations of real-world events.

Cheong (2013) – first published in:

Cheong, M., Ray, S. and Green, D. [2012a]. Interpreting the 2011 London Riots from Twitter Metadata, Proc. SoCPAR 2012.

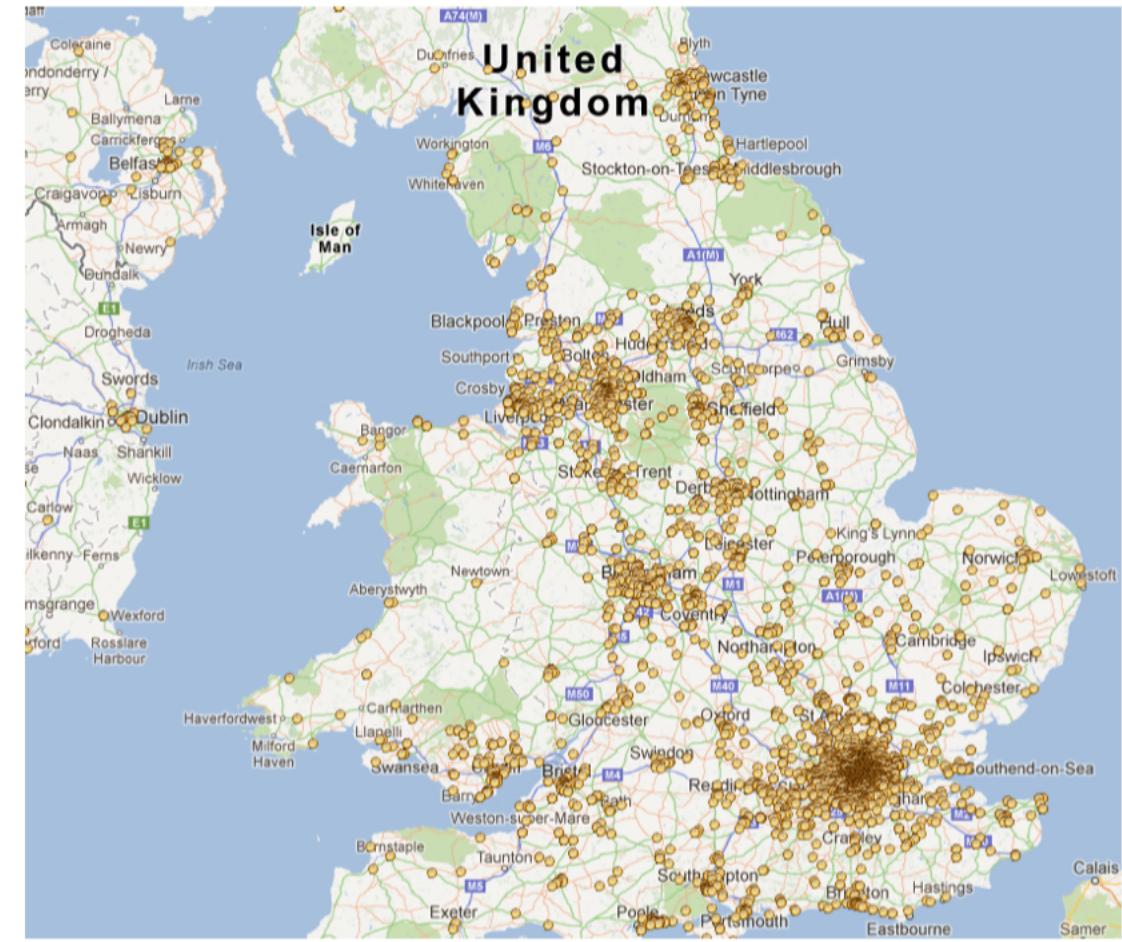


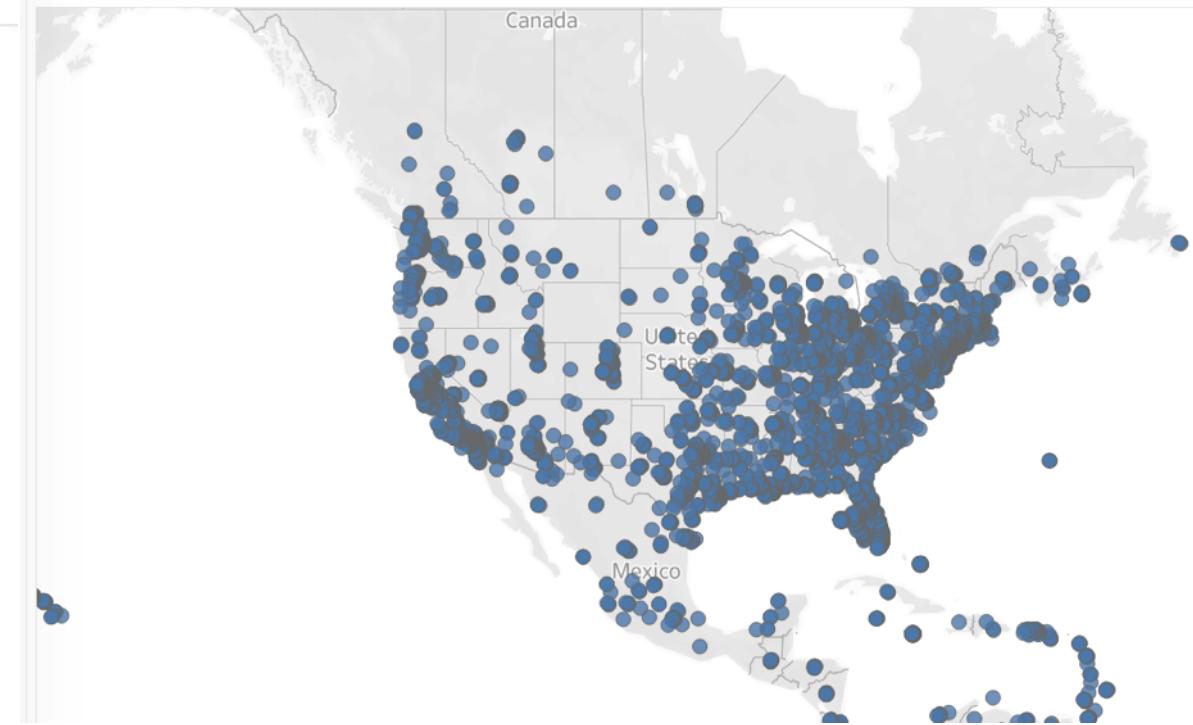
Figure 7.18: Visualization of locations found in Twitter metadata from the *London Riots Dataset*, originating from the United Kingdom. Each yellow dot represents a tweet composed in a particular location. From visual inspection, the spatial distribution of yellow dots roughly correspond to the red dots (representing riot activity) in Figure 7.17.

# Case Study: Twitter metadata BLM peaceful protest movement (2020)

geovis - Week 21



geovis - Week 22



# Ethical discussion I

For riots (London Riots, 2011) –  
the location information can be used for crime prevention.

- Is this ethical?
- Consider stopping of “rioting, looting, and arson”

the 7th of August, the rioting, looting, and arson started in parts of London, which subsequently spread to Birmingham, Manchester and Liverpool [May, 2011]. The riots continued spreading till 10th August, where they started to ebb due in part to bad weather [May, 2011] and police response [May, 2011; Meikle and Jones, 2011]. The riots resulted in

# Ethical discussion II

For peaceful protests (BLM protests, 2020) –  
is it problematic if geotags are used by police or?

- Why?
- Consider Amnesty International (2020)'s report □  
<https://www.amnesty.org/en/latest/news/2020/08/usa-law-enforcement-violated-black-lives-matter-protesters-human-rights/>

## USA: Law enforcement violated Black Lives Matter protesters' human rights, documents acts of police violence and excessive Force

4 August 2020, 10:00 UTC

Amnesty International USA Recorded 125 Separate Incidents of Police Violence Against Protesters, Medics, Journalists and Legal Observers in 40 States and D.C. During May and June Protests

*The Report Chronicles the Stories of More Than 50 People Affected by Police Violence as Black Lives Matter Grows Into the Largest Social Movement in U.S. History*

Today, Amnesty International USA released [a report](#) documenting widespread and egregious human rights violations by police officers against protesters, medics, journalists and legal observers who gathered to protest the unlawful killings of Black people by the police and to call for systemic reform in May and June of 2020. The report, *The World is Watching: Mass Violations by US Police of Black Lives Matter Protesters' Rights*, builds on Amnesty's [interactive mapping](#) of violence against protesters and [new findings](#) on the use of lethal force by the police. It is the most comprehensive human rights analysis of police violence against protesters to date.

Class discussion – as data scientists, discuss:

What determines the ethics of use of social media geo visualization? (5-10 minutes)



# Digital ethics!

## Interactive lecture.

# Risk mitigation: data precision and risk of re-ID

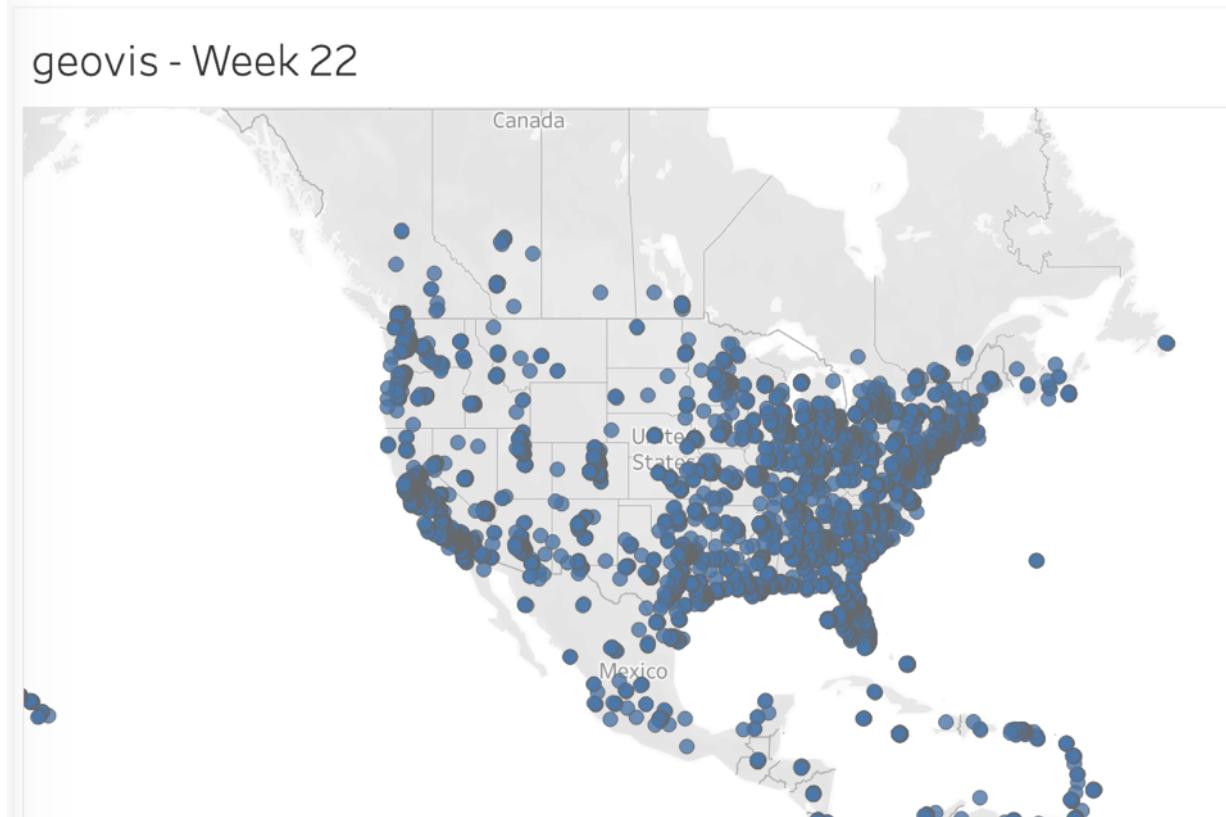
Take, say, the BLM case study –  
this is sensitive data!

Why?

Modern GPS's in phones ↞ accurate geotags!

(What did we do? Added lots of random noise –  
hence some data points in the ocean.)

What else can we do?



# Risk mitigation: another ethical dilemma

Take another relevant issue with user-generated geo data during the times of the pandemic...

**Aggregated location info from mobile apps (Google, Facebook) for public good**

**Is this ethical?**

**What are the risks?**

**What can we do?**

Sources:

<https://spectrum.ieee.org/the-human-os/telecom/wireless/facebook-google-data-publics-movement-covid19>

; Google

## How Facebook and Google Track Public's Movement in Effort to Fight COVID-19

Location data provide rich resource for decision makers, scientists, and the public



COVID-19 Community Mobility Report

Victoria 20 March 2021

Mobility changes

This data set is intended to help remediate the impact of COVID-19. It shouldn't be used for medical diagnostic, prognostic or treatment purposes. Nor is it intended to be used for guidance on personal travel plans.

The data shows how visits to places, such as corner shops and parks, are changing in each geographic region. Learn how you can use this report in your work by visiting [Community Mobility Reports Help](#).

Location accuracy and the understanding of categorised places varies from region to region, so we don't recommend using this data to compare changes between countries, or between regions with different characteristics (e.g. rural versus urban areas).

We'll leave a region out of the report if we don't have statistically significant levels of data. To learn how we calculate these trends and preserve privacy, read [About this data](#).



# Ethical discussion - just a start

Adapted from my talk given at the Victorian Centre for Data Insights (VIC Govt), May 2021.

*How do we put academic ideas of digital ethics into practice*

- As a start: learning basic key tenets (cf Coghlan, Miller, Paterson, 2020)  
<https://arxiv.org/pdf/2011.07647.pdf>
- “**fairness, non-maleficence, transparency, privacy, respect for autonomy, liberty, and trust.**”

Use this as a framework to discuss the COVID tracking scenario...

## How Facebook and Google Track Public's Movement in Effort to Fight COVID-19

Location data provide rich resource for decision makers, scientists, and the public

By Emily Waltz



Thank you  
Any questions?

