Data Science in Transport Planning
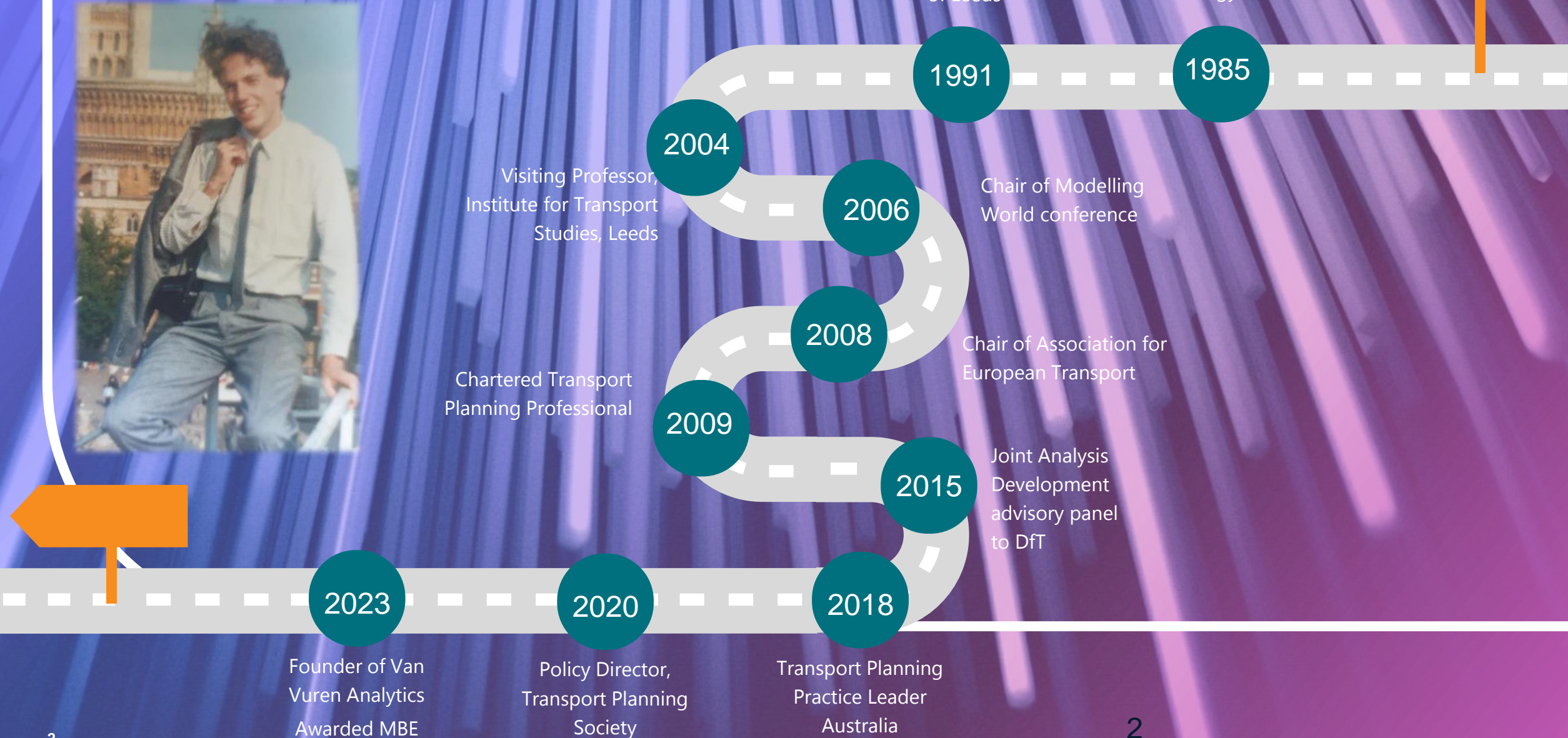
February 2024

Tom van Vuren
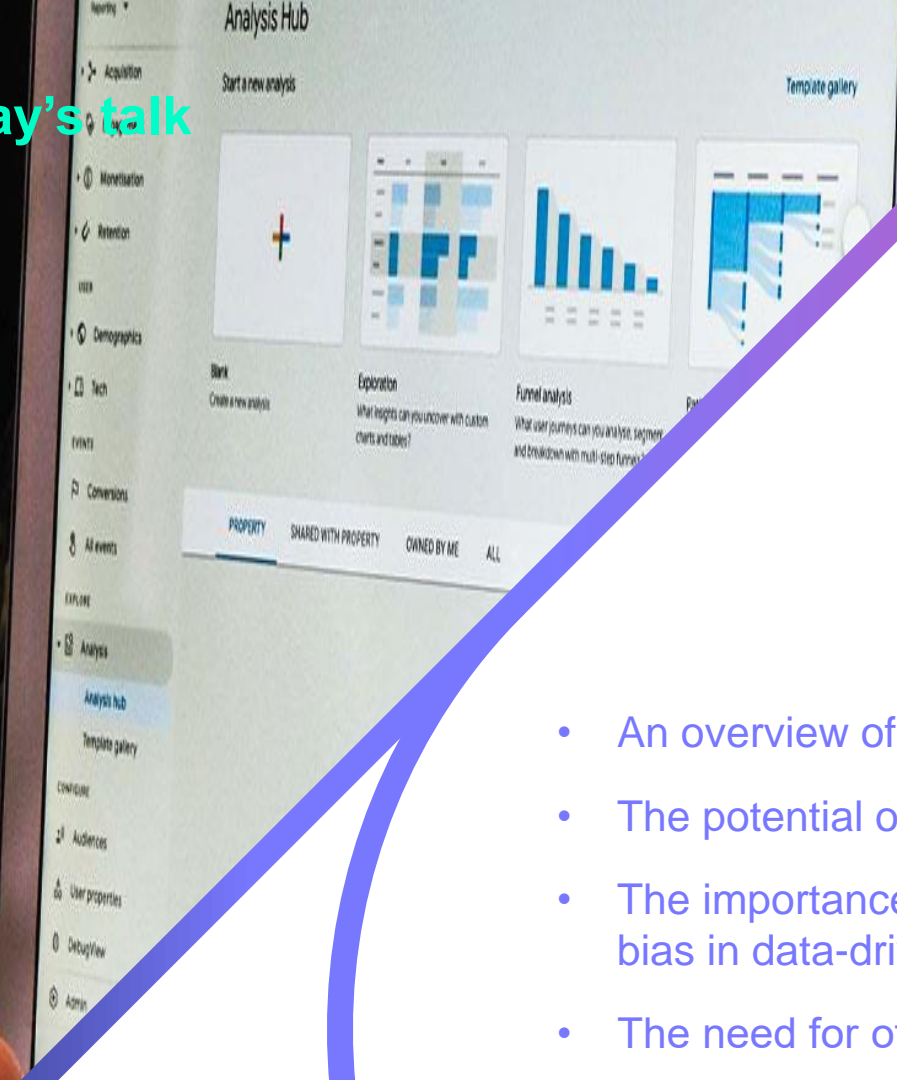
Strategic Business Partner, Amey

Amey

A brief history of ~~Time~~ Tom

MSc Civil Engineering , University of Technology Delft

PhD, University of Leeds

1991

1985

2004

Visiting Professor, Institute for Transport Studies, Leeds

2006

Chair of Modelling World conference

2008

Chair of Association for European Transport

Chartered Transport Planning Professional

2009

2015

Joint Analysis Development advisory panel to DfT

2023

2020

2018

Founder of Van Vuren Analytics Awarded MBE

Policy Director, Transport Planning Society

Transport Planning Practice Leader Australia

2

Contents of today's talk

Amey

- An overview of the use of data in transport planning in the real world

- The potential of new data sources

- The importance of data validation and critiquing and the dangers of bias in data-driven transport planning workflows

- The need for of data science skills in today's industry

Data – your responsibility

Existing Data Sources

Check:

the provenance of the data source

availability and transparency of reports describing the nature of the data, and the collection and processing steps

definitional issues that might impact on compatibility with other sources

suspected or known biases, their size, and how these have been treated

spatial coverage and date of data collection and/or release

any restrictions on access or use of data

Age of data

Newly collected data

Consider:

spatial coverage

data collection period

sample size and rate

the need to represent characteristics of the larger population

consistency of definitions with existing data sources

method of data collection

data privacy

Use of Data

Be aware of:

Errors

Bias

Unconscious bias

National Data

- **ONS Population Data**

- **Employment and Labour Market Data**

- **Valuation Office Agency Floorspace Data**

- **DfE School Census Data**

- **DVLA Vehicle Licensing Statistics**

- **MOSAIC**

- **ACORN**

- **Index of Multiple Deprivation**

- **AddressBase**

- **ONS Boundary Data**

Amey

Amey

National Trip End Model (NTEM) and Trip End Model Presentation Program (TEMPro)

- **Multimodal but…**

- **MSOA level only, therefore…**

- **CTripEnd model**

- **Freight?**

National Road Traffic Projections

- **Based on National Transport Model**

MOIRA (Model Of Inter-Region Activity?) / LENNON (Latest Earnings Networked Nationally Overnight)

- **Rail only**

Network Data

OS Mastermap Topography Layer

OS Mastermap Highways Network

OS Terrain Mapping

OS Open Roads

OpenStreetMap

DfT Major Roads Database

National Cycle Network

National Rail Timetable

CIF Railway Timetable Data

Traveline National Dataset

Realtime Trains

BODS Bus Open Data Service

Google Distance Matrix API

Movement Data

Mobile Network Data

GPS Tracking Data

TomTom

INRIX

(TrafficMaster) TeleTrac Navman

Electronic Ticket Machine Data

LENNON

Rail Users and Drivers Dataset (RUDD)

Census Journey to Work

International Passenger Survey

CAA Survey of Passengers

Usage Data

Office of Road and Rail Station Usage Data

Rail Passenger Numbers and Crowding on Weekdays

DfT Road Traffic Statistics

DfT Congestion Statistics

National Highways Traffic Information System (WebTRIS)

National Travel Survey

Continuous Survey of Road Goods Transport

DfT Bus Statistics

UK Time Use Survey

Land Use Data

NaPTAN National Public Transport Access Nodes

Census Data

**Business Register and Employment Survey (BRES)**

Amey

Household Surveys
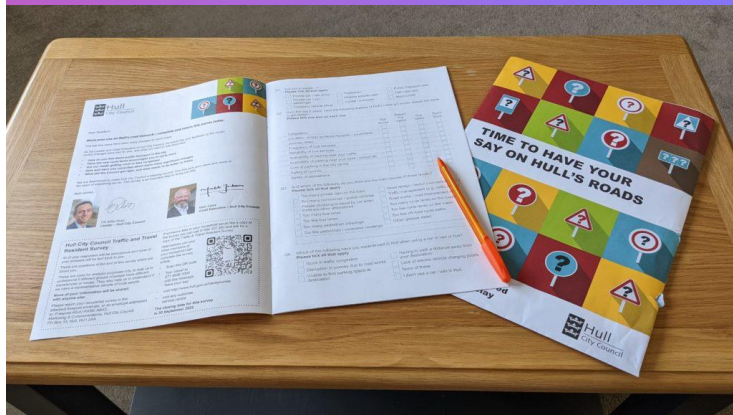
Based on recall

Clarity of terminology

Obtaining representative sample

Intrusive – how long is reasonable

Roadside or Station Interviews

Who to interview

Representativeness

Non-travellers

Intrusive – impact on choices

Mobile Phone Data

Spatial detail

Estimating trip ends – what is a stay

Limited info on purpose and mode

Lack of socio-demographics

Amey
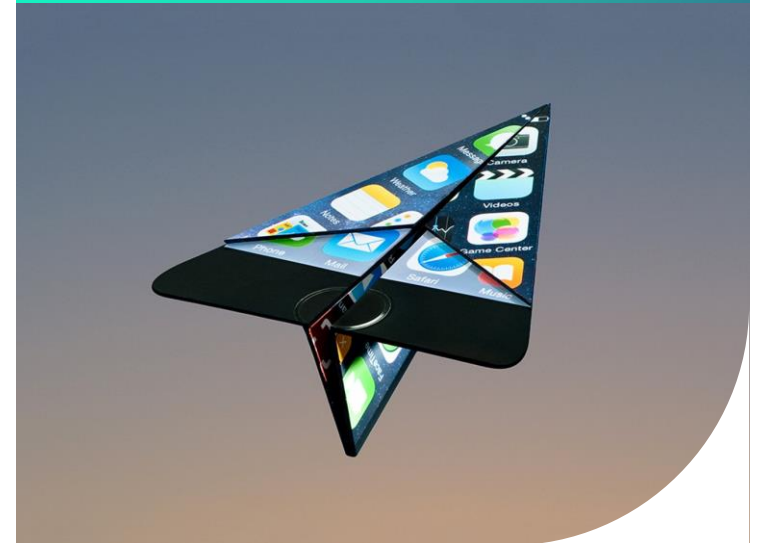
What if a cycling scheme trebles usage on you improved route?

Copenhagen: 90% of newly attracted cyclists have re-routed

DfT study:

11-28% came from cars,

45-58% from public transport,

19-26% from walking,

17-28% induced demand

Amey

How successful is theBristol ULEZ?

- 9.7% overall?

- 4.8% in ULEZ zone?

- what is counterfactual?

"Data for Sites 3 (+15.3μg/m³) and 12 (+10.2 μg/m³) have been impacted by non-CAZ related sources and results excluded from the report analysis. Site 3 is impacted by very localised and site-specific changes to a fast-food restaurant ventilation system. Further investigation is needed for site 12."

Diffusion Tube Monitoring Locations[1]	Average Change in annual NO_2 Concentrations ($\mu g/m^3$)	Average Change in annual NO_2 Concentrations (%)
Sites with >75% Data Collection Rates* - All	-3.2	-9.7%
Sites with >75% Data Collection Rates - Inside CAZ	-4.4	-12.6%
Sites with >75% Data Collection Rates - Outside CAZ	-2.5	-7.8%

Data Science in Transport Planning Practice

Amey

New ways of working

Low entry

Open Source Data

Open Source Tools

Machine Learning

Artificial Intelligence

Agent-based models

Old ways of working

Commercial tools

Barriers to entry

Hierarchy in expertise

Tried and tested

Prescriptive

Opportunity

Disruption

Creating community

New questions

Democratising models and data

Narrative, not numbers

Take-Aways

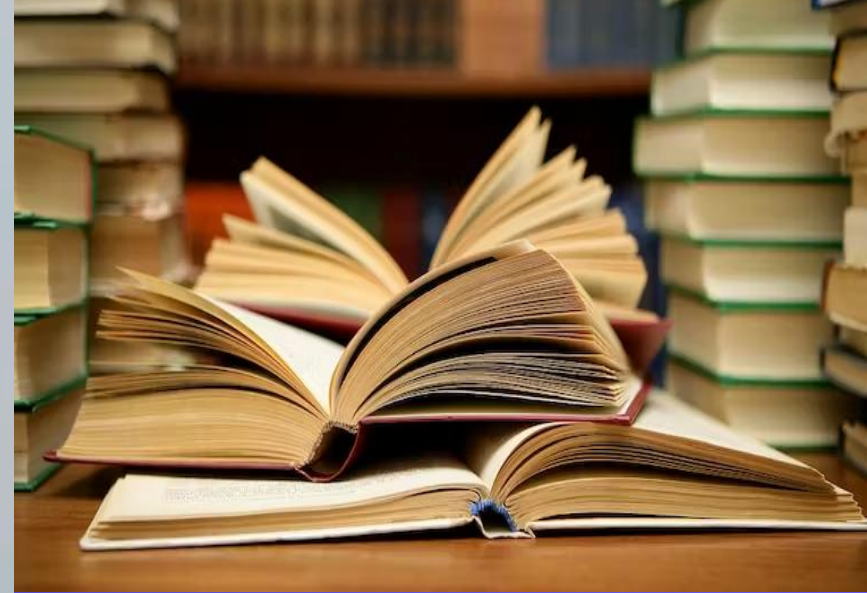Use existing sources where possible

Take responsibility for your data

Be aware of (unconscious) bias

Always challenge

Further Reading

TAG Unit M1.2: Data Sources and Surveys

TAG Data Book

What you were promised

What you got

Thank you

tom-vanvuren@amey.co.uk

tomvanvuren@hotmail.com

t.vanvuren@its.leeds.ac.uk