Crowd Cycling:
Understanding cyclist behaviour using the mobile tracking app Strava

Mark Dunleavy

A research paper submitted to the University of Dublin, in partial fulfilment of the requirements for the degree of Master of Science Interactive Digital Media

2015

Declaration

I declare that the work described in this research paper is, except where otherwise stated, entirely my own work and has not been submitted as an exercise for a degree at this or any other university.

Signed: _____________________________

Mark Dunleavy

14 May 2015

Permission to lend and/or copy

I agree that Trinity College Library may lend or copy this research paper upon request.

Signed: ______________________________

Mark Dunleavy

14 May 2015

Acknowledgement

I would like to thank my research supervisor Professor Glenn Strong for his guidance and encouragement during this research.

I would also like to thank my wonderful wife Jennifer for her support and amazing MS Word skills.

Summary

Concerns over traffic congestion, climate change and increased morbidity and mortality due to the rising rate of obesity have led to the promotion of cycling as a means to address all these issues simultaneously.

Governments and city planners have created policies, and reserved significant funds, to promote cycling as a mode of transport. Many cities, however, have evolved for motor vehicles with little consideration for cyclists. In order to ensure that the cycle-friendly policies are implemented as efficiently and effectively as possible, planners need to understand how cyclists behave, both as individuals and as a group. Current information on cyclist behaviour is derived mainly from surveys and bicycle counting studies, providing a very limited snapshot of cyclist behaviour in cities.

With the rise of the *Smart City* new opportunities exist to monitor every aspect of city life. The smartphone offers a whole new platform to record dynamic data while moving through a city. Furthermore, the ubiquity of smartphones also offers the opportunity to turn people into sensors, and create extensive databases of crowd-generated data. In the current study we use GPS data recorded in Dublin, by the users of the exercise-tracking app Strava, as proof-of-principle that crowdsourced location data can be a valuable resource in understanding cyclist behaviour.

The results of this study show that it is possible to gather information on cyclist behaviour using Strava. In particular, this study found data on the activity of cyclists with respect to dedicated cycle lanes, restricted turns and movement contrary to traffic flow on one-way streets. Our research also demonstrated that change in cycle traffic flow over time could also be observed using this data.

Overall, this study successfully shows that crowdsourced location data can give an insight into cyclist behaviour, and can help governments and planners make better informed planning and policy decisions in the future.

Table of Contents i

“If it’s not on Strava it didn’t happen”

- Unofficial motto of Strava users

1. INTRODUCTION

Two of the biggest issues facing society today are climate change and the ever-increasing rates of obesity. The search for solutions to these problems has highlighted the potential benefits of cycling in addressing these problems. Cycling is being seen as a multiple satisfiers that can address the need of many aspects of society at the same time. Using cycling as a mode of transport has been shown to reduce harmful carbon emissions and improved health outcomes. As a result governments have begun to promote cycling as an alternative mode of transport. Numerous *cycle-friendly* policies have been drawn up and, working in conjunction with city planners, attempts are being made to increase the number of people choosing cycling as a mode of transport.

Most modern cities, however, have developed during the time of motorized vehicles resulting in a road infrastructure that is not necessarily suitable for cyclists. In order for planners to create a cycling an effective cycling network, planners need to understand the behaviour of cyclists in their cities. In Dublin city the data, on which they make their assumptions, is limited. Restricted to models of traffic based on poor quality data, destination surveys, and bicycle counts over limited time periods, the current information does not provide a full picture of the dynamics of cycling traffic.

The rise of the Smart City, where large networks of sensors are distributed through a city, provide an opportunity to gain an insight into city dynamics. While most sensors are static, smartphones offer an array of sensors that may *sense* the city as they move through its streets. Recent studies have shown that by using diverse crowds of people to generate data solutions to difficult problems can be found. As a result, the current study intends to use GPS location data, generated by users of the exercise-tracking app Strava, as a proof-of-principle that crowdsourced location data can provide high-resolution insights into cyclist behaviour in Dublin city centre.

1.1 Crowdsourcing

In 2006 Jeff Howe coined the phrase “crowdsourcing”, stating that it “represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call. This can take the form of peer-production (when the job is performed collaboratively), but is also often undertaken by sole individuals. The crucial prerequisite is the use of the open call format and the large network of potential laborers" (Howe 2006). Of course, this was not a modern phenomenon. As far back as the 1700’s Queen Anne of England created the Longitude Act of 1711, offering a monetary prize for any solution, which would allow sailors to accurately calculate their longitude at sea, and thus prevent shipwrecks that were responsible for a significant loss of life and money at the time (Quill 1966). The competition was open to all and attracted entries from all professions and walks of life. Ultimately it was not the famous physicists like Isaac Newton, or the renowned astronomer Edmund Halley that found the solution, but the self-educated carpenter John Harrison who created a watch capable of accurately telling the time at sea, and hence calculating the longitude.

While Queen Anne’s competition shared many of the characteristics that Howe associated with crowdsourcing (outsourcing a problem to a diverse range of people) it is separate from the modern interpretation of the concept by several magnitudes of scale. The development of the Internet has created an infrastructure where groups, diverse in terms of geographic location, expertise, social status, etc. can be connected in milliseconds. While it took the best part of the century for Harrison to create his final timepiece, solutions to complex problems can now be found in a seconds, and large databases of information can be generated with relative ease. Early iteration of web-based crowdsourcing concentrated mainly on the exchange of opinion and expertise to create value in the space (Brabham 2008). Projects like InnoCentive allowed large pharmaceutical and chemical companies access a broad range of outside users to solve problems which had eluded their in-house research teams (ibid). By gathering the opinions of tourists from all over the world, Trip

Advisor was able to create a business from content provided voluntarily by visitors to hotels and tourist locations. The t-shirt makers Threadless, similarly created a space where anyone could submit a t-shirt design to their website, other users voted on their favourite, with the winner being printed and sold by Threadless. In this way the website managed to create an interested community and create a product without any of the design overhead. As with many of the earlier crowdsourcing products, these companies simply took advantage of a diverse disparate group of people, made available to them by the increased connectivity of the World Wide Web (Brabham 2008; Howe 2006).

Key to the success of these projects was the continued participation of users. This was achieved by utilizing user-generated data to create products that the user themselves could benefit from. Since then, successful crowdsourcing projects have shared the principle that the users generating the content are getting some form of reward, be it individual, local or on the global scale.

The crowdsourcing movement has always adapted to reflect the technological environment in which it exists. *The Wisdom of the Crowds* (Surowiecki 2005) discusses the enhanced potential of crowd thinking. The aggregate opinion of a crowd can often out perform the smartest person within the crowd. It might be argued that many of the early crowdsourcing applications were performing this same function, while taking advantage of the increased connectivity afforded by the Internet. With the development of information and communications technology (ICT) came the opportunity to bring crowdsourcing to the next level. *Mobile Crowdsourcing* took advantage of the shift from Desktop to Smartphone. By exploiting the location sensors present in mobile devices, users could use applications like Four Square and Trip Advisor to locate nearby places of interest (POI); the status of POI often being bestowed on a location by review content generated by other users of the application.

Apps such as these characterised the first generation of mobile crowdsourcing apps, where users actively provided geolocation information by “checking-in” at a location.

In many ways these apps were preforming the same function of the pre-mobile apps (sharing expertise and opinion) but with the addition of a geolocation tag. Smartphones are no longer just to device to make a phone call, or indeed surf the Web. In addition to the location sensors (GPS, Wi-Fi, Bluetooth) phones now carry an array of sensors that can track both the user and the environment, from gyroscopes, barometers, accelerometers, humidity sensors to heart-rate monitors (Ali et al. 2014). The new generation of mobile crowdsourcing apps have begun to embrace the potential of these sensors to create data; moving beyond the place where users needed to actively generate data, to a space where a smartphone can act as a single sensor in a larger network of sensors. Constantly recording data about the user and the environment in which they exist. Many applications of this principle already exist. A research group in California has created an application which monitors a phone's GPS sensor for disturbances which may indicate the presence of an earthquake (Minson et al. 2015). By monitoring multiple phones in disparate locations simultaneously, the developers are able to distinguish a real seismic event from a truck passing by or a phone shaking in a pocket. This is an apparently simple solution which could provide early warning of impending earthquakes. Pressurenet[1] goes beyond simple geolocation sensors to create a real-time local weather. The app works in the background to gather information on atmospheric pressure from the phone's barometer. This allows the app to gather all the information and provide a local weather report to the users.

1.2 Smart Cities

In the current study we want to look at the potential of crowdsourcing data to gain a greater understanding of the behaviour of cyclists in Dublin city centre. As Salim and Haque (2015) have said, cities are "complex, dynamic and messy". Almost like a living organism cities are in a constant state of flux, characterized by the flow of people, goods, information and capital, with these flows directed by the cities' infrastructure network (Williams & Dourish 2006). The complexity of cities, both

[1] http://www.pressurenet.io/

spatially and temporally, have made them a difficult subject to study. However the rise of ubiquitous computing has shed new light on the area and give cause for great optimism. The rise of the “Smart City” has come at a time where, as predicted by Mark Weiser (1999), computers would become so commonplace as to disappear from the consciousness of everyday life. By 2020 it is predicted that there will be 30 billion devices will be connected to the Internet, sensing, analysing and actuating in the environment around them (Gartner 2013).

The global urban population is set to increase to almost 5 billion by 2030 (United Nations 2013). As urban populations increase so do the pressures on a city’s resources. Increased population brings with it elevated emissions, greater demands on utilities like electricity and water, increased infrastructural development and increased traffic flows. In order to be able to efficiently deal with these challenges cities across the world are taking advantage of the developments in ICT, and adopting the “Smart City” model. Smart Cities incorporate ubiquitous sensing within the city’s infrastructure to help it function in an intelligent and sustainable way (Giffinger et al. 2006). This is achieved by placing sensors (up to millions) throughout a city to monitor everything from environmental factors such as temperature, humidity, air and water quality, to monitoring the structural integrity of bridges and other large structures to sensors that track the consumption of electricity by households. As pointed out by Salim & Haque (2015), “Urban computing research should not focus only on solving the technical problems and challenges, but also on wider challenges requiring transdisciplinary inputs from computer science, engineering and technology, health science and medicine, social science and anthropology, architecture and design, urban design and planning, arts disciplines, and policy makers.”

Projects of various scales are underway all over the world. A number of projects have involved he construction of dedicates Smart Cities from scratch, such as PlanIT Valley in Portugal (PlanIT 2013) and Songdo in South Korea (Strickland 2011). In Songdo, systems have been put in place to monitor almost every aspect of life. Consumption of electricity and water is monitored at home, cars are tagged to optimize traffic flow, streetlights respond to pedestrian activity and water flowing through the city's canals is regulated from a central source. While it might seems easy to incorporate these features into a purpose built city, Smart City features are being incorporated into existing cities with great success; creating more efficient, liveable cities as a result of linitiatives like systems to create more energy efficient buildings and monitoring systems to improve traffic flow resulting in reduced emissions and a better quality of life for commuters (Avelar et al. 2014).

The systems incorporated by Smart Cities falls under the umbrella of the Internet of Things (IoT) (Mitton et al. 2012), a concept where everyday objects are embedded with sensors (smart objects) and form part of a network which facilities the exchange of information with a central processor. These systems present significant challenges to researchers trying to optimize these systems.

With million of data points being created everyday (Gartner, 2013), the technology to store and analyse the data is constantly advancing, with Cloud computing playing a key role in meeting the demand (Hancke et al. 2012; Mitton et al. 2012). While Cloud services may be hosted on servers at optimum locations chosen by the provider, sensors, by their ubiquitous nature, must be located in or nearby the object they are sensing. Recent advances in semi-conductor physics and nano-technology have created small, inexpensive devices that can be placed almost anywhere (Girbau et al. 2012; Ueno et al. 2011). However, practical reasons prevent each sensor being connected by wires, giving rise to issues of power supply and information transfer. Wireless Sensor Networks are one way in which the data connectivity issue is being resolved (see Hancke et al. 2012 p 396-99 for review) with long range (Dash 7, Zig Bee, 3G networks) and short range (radio frequency ID (RFID) and near field communication (NFC)) wireless technologies. The energy constraints on a sensor network are significant; due to the scale of installation it is infeasible to create a system that requires replaceable or rechargeable batteries. Accordingly researchers have turned their attention to alternative energy sources. The energy requirements, of course, vary based on the type of sensor but researcher have attempted to exploit wind, radio frequency, solar, thermal, vibrations and movement (Hancke et al. 2012). Advances have been constant but this still remains a stumbling block for the creation of a full scale Smart City.

1.3 Crowdsourcing with Smartphones

With these limitations in mind, it is interesting to consider the potential role of crowdsourcing to provide sensor nodes for a Smart City. By exploiting the sensors present on the smartphones that people carry every day throughout the city, we can move beyond Smart Objects to Smart People. Tapping into the multiple sensors that mobile phones contain we have the potential to generate a significant data sets with relatively low cost. In contrast to the majority of sensors discussed above, these data points would be dynamically collected from different location, and analysis of these data points may prove extremely valuable to understanding many aspects of city life. In the current study we are looking specifically at how such an imitative could gather data on transport patterns. Campbell et al. (2006) proposed the idea of “people centric urban sensing”, where very large scale “opportunistic sensor networks” could be set up to take advantage of the relationship between mobile and static sensors. This approach has been adopted in a number of studies, e.g. looking at crowdsourced mobile phone data to explore urban motility with a view to optimizing public transport (Berlingerio et al. 2013).

There is some precedent to suggest that gathering data in such a way is effective and can improved city living through improved traffic management. Waze[2] is a mobile navigation app for iPhone and Android that provides motorists with the fastest route to their destination. Unlike other apps it does not rely on inputs from local authorities, but passively tracks the speed and location of every active user to create a profile of traffic flow. The app’s 20 million+ users are creating millions of data points every day while allowing users to find the best routes in real time. The value of the application does not stop there however. Although the main reason for interacting with the app is to get the best route, by running the app, users are

---

[2] http://www.waze.com

allowing Waze to passively crowdsource location data which is also being used by cities to improve traffic management systems (Ravindranath 2014).

2. BACKGROUND

2.1 Cycling and Strava

Although the primary function of Waze is to provide users with directions to their desired location, it is clear that the location data that is collected through passive crowdsourcing could play a vital role in creating smart solution to the problems that motorised traffic face. In the current study we intend to investigate if such a model could also be applied to other modes of transport within a city. Cycling is being seen as a mode of transport which may answer many of the challenges of modern urban living. Cycling is being viewed as a “multiple satisfier”; an activity that is “able to help fulfil many valuable human needs simultaneously” (Horton et al. 2012). Significant evidence exists that cycling can be of benefit in addressing many modern social issues.

2.1.1 Health

Over one third of adults do not perform enough physical activity every day (Hallal et al. 2012), and this lack of activity has been associated with increased morbidity and mortality due to cancer, cardiovascular disease, type II diabetes, and impaired mental health (Mueller et al. 2015). Cycling has been proposed as strategy to combat the spread of these diseases. Some studies have suggested that the increase in physical activity associated with cycling can act as an important tool in weight control (Frank et al. 2004), and reducing risk of cardiovascular events. In a systematic review of 30 studies that met the eligibility criteria, Mueller et al. (2015) found that the benefits of moving to an active mode of transport were almost exclusively positive; with increased physical activity, reduced visits to the general practitioner, reduced air pollution and a reduction in obesity cited by many studies. Conversely the main negative outcomes reported by the study were increased exposure (of the traveller) to air pollutants and an increase in incidents with traffic. An investigation into the implication of transport mode in an urban environment, suggested that a modal shift away from cars to active transport modes would improve individual health, the environment, and society in general, while at the same time decreasing the negative aspects of transport (Mindell et al. 2011). A Danish longitudinal study on a population of regular cycle commuters aged 20-93, found that, in any given year, cyclist had a 39% lower risk of mortality after multivariate adjustment (Andersen et al. 2002).

2.1.2 Urban congestion

The European Commission-funded WALCYNG (walking & cycling) project (Hydén 1997) investigated conditions that may replaces shorter car trips with walking and cycling. The study found that issues such as social climate (how walking and cycling were viewed by society), health (a basal level of fitness was required), comfort (availability of shelters, public toilets, etc.), subjective safety (separation of walkers from both cyclists and cars) mobility (cycling infrastructure should be continuous and of good quality) and financial advantage all influenced a persons decision to cycle.

2.1.3 Environmental concerns

A reduction in CO_2 emissions has now become a priority for governments for both an environmental and financial standpoint. There is now unequivocal evidence that there has been a warming of the climate system in the second half of the 20th and early decades of the 21st century. The report from the Intergovernmental Panel on Climate Change (IPCC 2013), highlights the recorded changes and implicates an increase in the release of greenhouse gases, such as CO_2, as significant contributors to these changes. Consequently governments have been examining areas where the production of these gases may be reduced or eliminated. Studies have shown that substituting a car trip with a form of active transport, such as cycling can result in decreased emission and air pollution (Maibach et al. 2009; Federation 2011)

2.1.4 Socio-economic

The cost of car transport in cities is high, both economically and socially. The majority of trips taken by car in urban areas are short (< 6km), habitual (usually on the same route) on a congested network. Consequently, there is are significant ecologic, economic and social costs to expanding road capacity to meet peak traffic demand (Harrison 2012; Coffin 2007).

Governments all over the world have implemented policies to promote cycling as a mode of transport and increase modal share (the proportion of trips taken by bicycle) (Lanzendorf & Busch-Geertsema 2014). This trend had been reflected in Ireland with the implementation of the National Cycle Policy Framework 2009-2020 (Department of Transport, Tourism and Sport 2009). This policy outlines a number of objectives with the goal of 10% of all trips being undertaken by bike by 2020. The framework outlines 19 wide ranging objectives (ibid p9) to help achieve this vision. These objectives range from, at the low level, the provision of fiscal incentives to cycle and the availability of secure bike parking facilities, to higher-level considerations, such as ensuring that any new development of towns and cities is undertaken in a cycle-friendly way. Of particular interest to the current study are:

> **Objective 2:** Ensure that the urban road infrastructure (with the exception of motorways) is designed /retrofitted so as to be cyclist-friendly and that traffic management measures are also cyclist friendly.
>
> **Objective 11.** Improve cyclists' cycling standards and behaviour on the roads
>
> **Objective 16:** Improve enforcement of traffic laws to enhance cyclist safety and respect for cyclists.

As will be outlined below, these objectives are of particular interest in Dublin city centre, given the increasing number of cyclists competing for space on a road network with limited dedicated cycle network infrastructure (Figure 1).

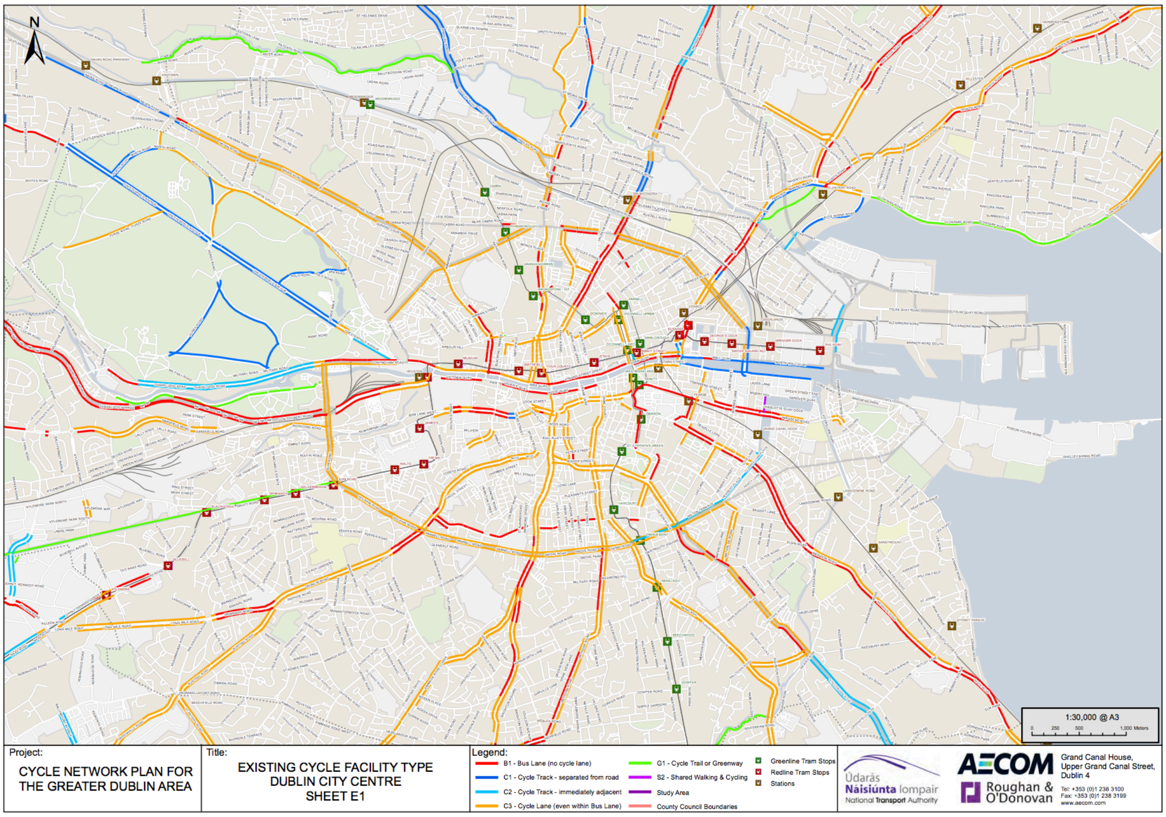

Figure 1. *Map of the existing cycling infrastructure in Dublin (reproduced from the Greater Dublin Area Cycle Network Plan 2013).*

To effectively achieve these objectives it is vital to understand how cyclists behave both as a group and as individuals. However, the current state-of-the-art methods for gathering data on cyclists are limited. These techniques range from surveys, manual cycle counts and pneumatic tube counts. Unfortunately none of these techniques provide a full picture of the flow of cycling traffic. Surveys generally only provide information on the origin and destination of a cycling journey but not the route taken (Milward-Brown 2013; CSO 2012), while counting only looks at the number of cyclists passing certain points over a limited period (NTA & DCC 2014) provided a dataset constrained by time and space. In recent years studies have begun to examine the potential of crowdsourcing information to improve cycling infrastructure. A study in Zurich, Switzerland used GPS in attempt to understand cyclist route choices (Menghini et al. 2010). While Hood et al (2013) used data generated by the mobile-tracking app *CycleTracks* to examine cyclist behaviour in San Francisco. Despite these studies large-scale studies using GPS-data seem to be rare.

Exercise tracking applications are growing increasingly popular and may offer an alternative way to monitor cycling traffic flow and cyclist behaviour. One of the most popular apps among cyclists is Strava[3], a mobile app that uses GPS to allow users to record their riding activity. Set up in California in 2009, Strava provides users with maps of their rides and information on ride distance, continuous and average speed, elevation, heart rate, cadence, temperature, and calories burned among many more (Williams 2012). The application proved instantly popular among "data geeks" who wished to look into their performance in greater detail but the popularity of the app continued to grow into the wider cycling community, with Strava estimating that 2.5 million GPS-tracked activities uploaded to the site every week, contributing to their database of over 300 billion data points across the globe (Strava n.d.). The success of Strava has been attributed to the *segment* feature of the application, which allows users to virtually race other users over sections of road, with the user with the fastest time being crowned *King or Queen of the Mountains.* Tapping into the concept of gamification Strava have created a application that appeals to the competitiveness associated with social network gaming (Shin & Shin 2011). With the Strava dataset available to any developer via a public application program interface (Strava V3 API), the current study intends to investigate Strava user activity in Dublin city centre, to determine if this crowdsourced dataset may provide greater insight into cyclist behaviour in Dublin city centre.

Models of cyclist behaviour play a key role in predicting cyclist behaviour within a city and in turn inform future planning decisions relating to the cycling network. As part of the Greater Dublin Area Cycle Network Plan (NTA & DCC 2013), consultant engineers Roughan O'Donovan developed a model of bicycle traffic in Dublin up until 2021 (Figure 2). The model made assumptions on trip demand based on a number of the data sources discussed above (Census, 2011; The National Household Travel Survey; Canal Condon Counts). Using this model it is possible to forecast the demand assessment of current and future patterns in the Greater Dublin Area. While these

[3] http://www.strava.com

models can be valuable tools in understanding future demand, it is important to note that the current model has a significant number of limitations. When creating a trip demand matrix, the designers chose to use the existing road network, supplemented by some additional cycle network links. Furthermore, route choice was assigned on the shortest distance between points only. These decisions did not take into account the route choice decisions that cyclists make based on Quality of Service (i.e the quality of the cycle lane) or the fact that cyclists do not necessarily follow the flow of traffic on the road network (Aherne and MacGearailt 2013). This is backed up by evidence-based studies using GPS tracking (Broach et al. 2012).

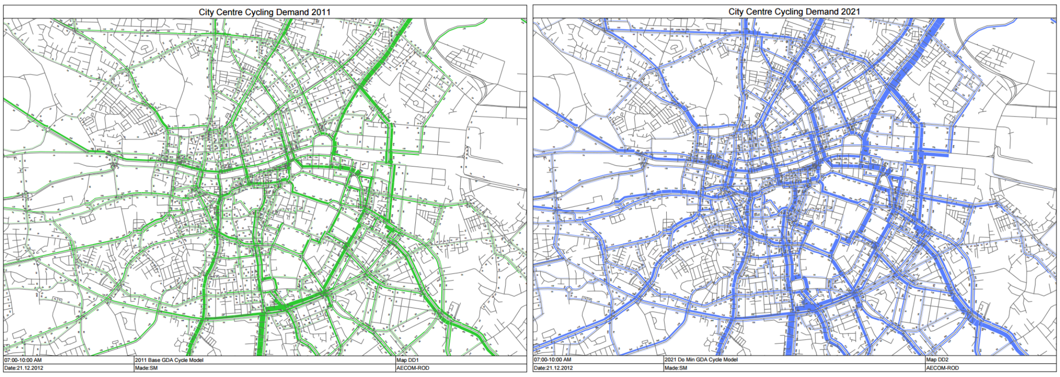

Figure 2. *A graphical representation of cycling demand in 2011(left) and predicted for 2021 (right). The predictions are based on the cycle traffic model developed for the Greater Dublin Area Cycle Network Plan.*

The current study will use data generated in Dublin city centre by users of Strava, as proof-of-principle that crowdsourced location data can provide improve our understanding if cyclist behaviour within a city, and help inform better policy and planning choices going forward.

3. METHODOLOGY

The current chapter describes the specific methods that were employed to achieve the aims of the project. The principle aims were to investigate cyclist behaviour using three case studies that represented the most common areas of the transport network where cyclist may deviate from the idealised model of cycling traffic flow. The three case studies represented:

1. Restricted turns
2. One-way streets
3. Cycle-lane alternative

Furthermore, the overall cycle traffic flow was examined in case study 3 to see if it gave any insight to the daily and season variation in demand on that route.

3.1 Strava Segments

As outlined in the Introduction, one of the key distinguishing features of Strava is the facility for users to create segments, where they can track their own performances and compare them with others. This *create segment* functionality was exploited in the current study to create test segments in each of the case studies. A segment can be created on any existing activity simply by dragging start (green) and end (red) markers, which are represented on a map of the activity (Figure 3). Once created, the segment will be saved in the Strava database and populated by all activity that has taken place on that particular segment.

3.2 Strava API

To access this data we used the *Strava API*, a publically available interface allowing registered developers to access to user-generated data. API's are an increasingly important element of interactions on the Web. It was estimated that there were over 10,000 publically accessible API's published in 2013 (Berlind 2013), a four-fold increase on the amount published in 2010 (DuVander 2011). This rate of growth continues to rise with API's such as Google Maps API, Google AJAX Search API, Twitter API and Facebook API being integral to many online interactions. The Strava V3 API allows access to over 300 billion user-generated data points[4]. Data is stored in the JavaScript Object Notation (JSON) format, allowing easy access to details on athletes, activities and segments.

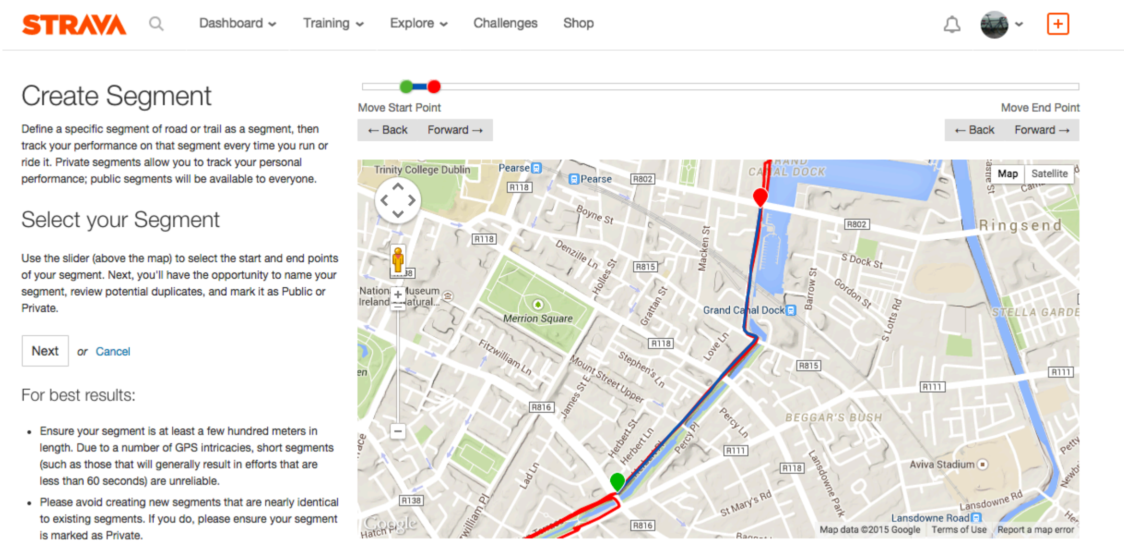

Figure 3. *Screenshot of the "create segment" screen on Strava. The segment can be selected by simply sliding cursors that display start and end point on a map.*

3.3 Authorization

Once the application was created on the Strava Developer site, we were granted a public access_token, which when included with a request, allowed access to *segment* and *segment leader board* data. By default, only cycling activities were downloaded.

It should be noted that applications may also access detailed information on every activity undertaken by an athlete including data on route, instantaneous speed/heart rate/power output, etc., however, this requires each athlete to individually authorize the application to access their data. Due to ethical considerations and time contraints this was beyond the scope of the current study, however it may be worth considering for future, more in depth, analyses of cyclist behaviour.

3.4 Python

The Strava API was accessed using a computer programme written in Python (Python Software Foundation, Python Language Reference, version 3.4.1[5]). Appendix I contains

[4] http://strava.github.io/api/

the Python code used by to access the data through the Strava API. Data for each query was written to a comma separated value file(.csv) and saved for later analysis using MS Excel.

3.5 Case Studies of Strava User Behaviour

3.5.1 Restricted turns

Restricted right turn from St Stephen's Green East to St Stephen's Green South

The activity of cyclists when exiting the junction at the south east corner of St Stephen's Green was examined using Strava data. Three segments were created to represent the three potential exits for a cyclist arriving at the junction from the cycle lane on St Stephen's Green East (Figure 4).

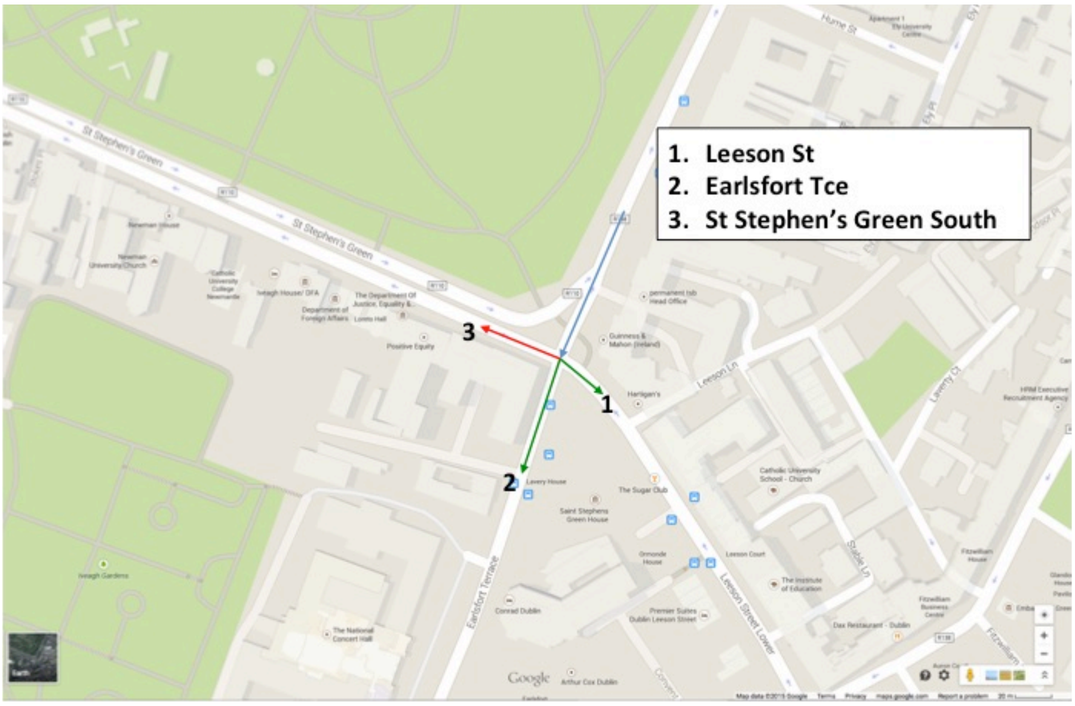

Figure 4. *Map showing the exits from St Stephen's Green East. Permitted turns are indicated in green and restricted turns marked in red.*

The first two exits were legal, while the third represented a restricted right-hand turn in the context of the current road layout.

- Left turn onto Leeson Street Lower - Permitted
- Straight onto Earlsfort Terrace - Permitted
- Right turn onto St Stephen's Green South - Restricted

[5] Available at www.python.org

Using the Strava API, data for each segment was collected on:

- Total number of athletes that have completed the segment
- Total number of times the segment has been completed
- The average speed of each athlete's fastest attempt
- The date and time of each athlete's fastest attempt

The data for the restricted right-hand turn was then compared with the other segments to determine any differences in cyclist behaviour.

3.5.2 One-way Street

Route along St Stephen's Green West

The second case study examined the path followed by cyclists when travelling from the junction of St Stephen's Green and Grafton St in the North, to the junction of St Stephen's Green and Harcourt St in the South (Figure 5). The study looked at two segments which represent two routes that cyclists can use to travel between the two endpoints.

- The **Direct Route** travels along the side of St Stephen's Green and represents the shortest distance between the points. By using this route cyclists are travelling contrary to the flow of traffic in a one-way system for approximately 50% of the journey.
- The **Model Route** represents the shortest distance between the two points if the cyclist uses the proper traffic layout for the area (as calculated using the Google Maps Cycling Route Planner).

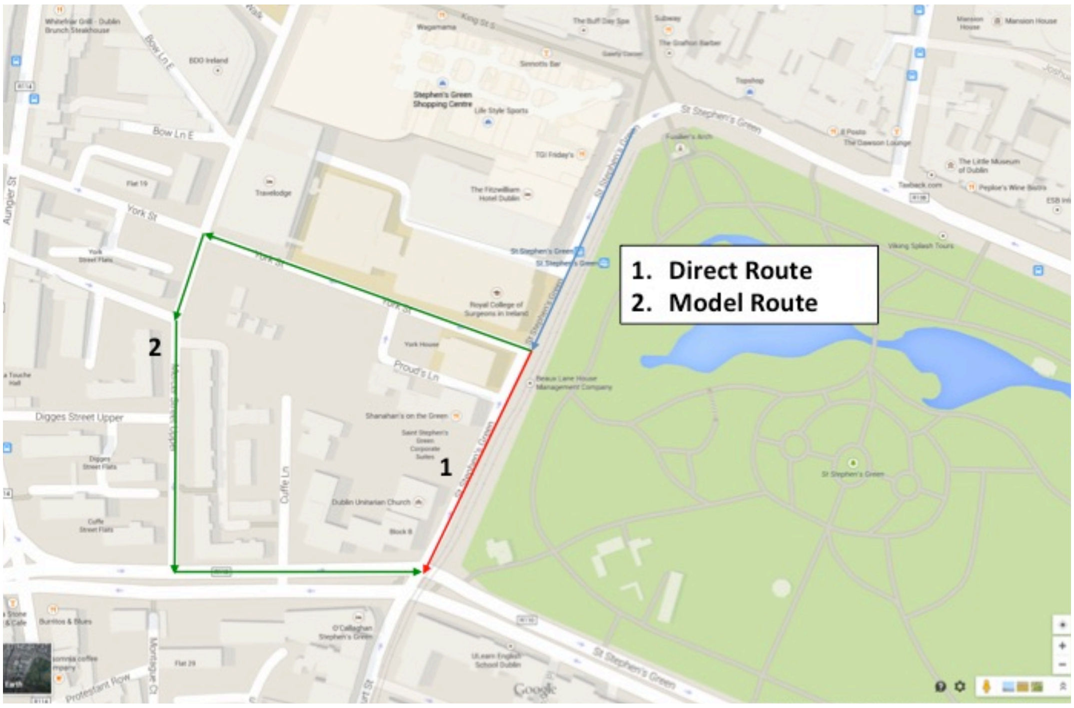

Figure 5. *Map representing the Direct (red) and Model (green) routes between St Stephen's Green and Harcourt St.*

3.5.3 Alternative cycle lane

Baggot St Bridge to Leeson St Bridge on cycle lane and road

The final case study looked at the route choice of cyclists when presented with two alternative routes running in parallel, with the same start and end points. For this study we chose to look at the route choice of cyclists travelling from east to west on the orbital route to the south of the city centre, along the Grand Canal. In 2011 the Grand Canal Cycle Lane was opened, providing cyclists with a bidirectional lane exclusively for bicycles. The lane is positioned on the north bank of the canal running through a quiet tree-lined area completely separated from traffic. In contrast, the existing cycle lane runs on the south bank of the canal sharing space with vehicular traffic. Our segments both began at Baggot St Bridge and continued to Leeson St Bridge (Figure 6).

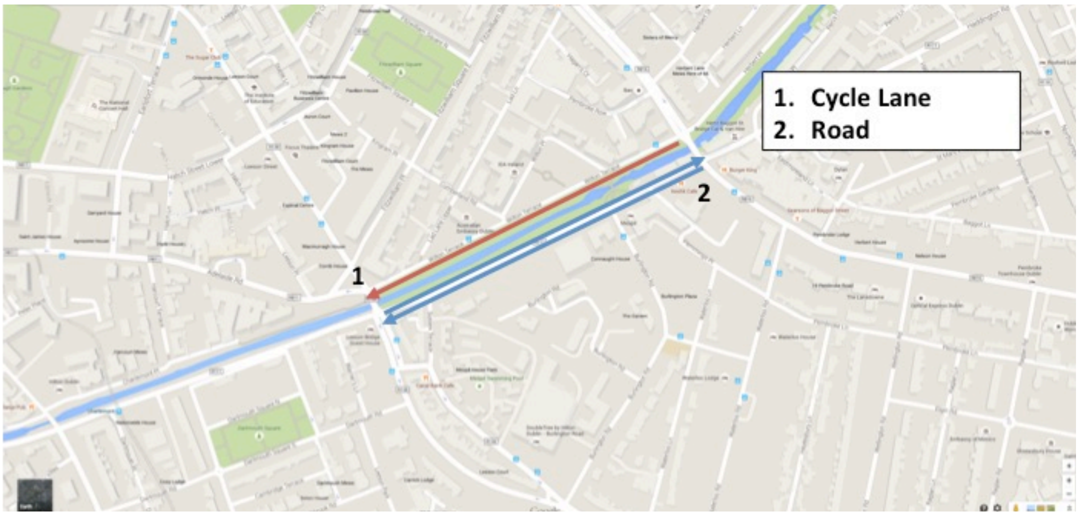

Figure 6. *Map representing the cycle lane (red) and road (blue) routes between Baggot St Bridge and Leeson St Bridge. The road route to the south of the canal includes segments going in both directions.*

3.6 Daily and Seasonal Trip Demand

An analysis of the time distribution of the trips going in both directions on the road section was also performed to see if the data gathered by Strava could provide any insight into the daily and seasonal variations in traffic**.** This was achieved by comparing the time and date distribution of all activities in the leader board.

3.7 Google Maps Cycle Maps

In the absence of access to traffic modelling software such as MapInfo[6] the idealised route for case study 2 was calculated using the route planner feature of Google Maps. The route calculated was based on the fastest time between two points using the selected mode of transport.

[6] http://www.mapinfo.com

4. RESULTS

This chapter will present the main findings from the current study.

4.1 Adoption of Strava

Initial analysis looked at the trend of Strava adoption over time to see if sufficient numbers were using the app to make it a viable strategy for population analysis. By aggregating all the data available on the leader boards for the three case studies analysed, we found that there was a progressive increase in the number of fastest segments being set every year since Strava was launched. Figure 7 shows the increase from 6 users setting their fastest times in 2010 to 779 setting their fastest times in 2014. Data was only available for the first 4 months of 2015 (336) but the trend of increased user activity seems to be continuing. While this data does not provide absolute user numbers it is suggestive of a trend that Strava usage is increasing steadily.

Analysis of the gender distribution revealed that 85% of Strava users in the current study are male; almost 10% are female, with no data available on approximately 5% of users.

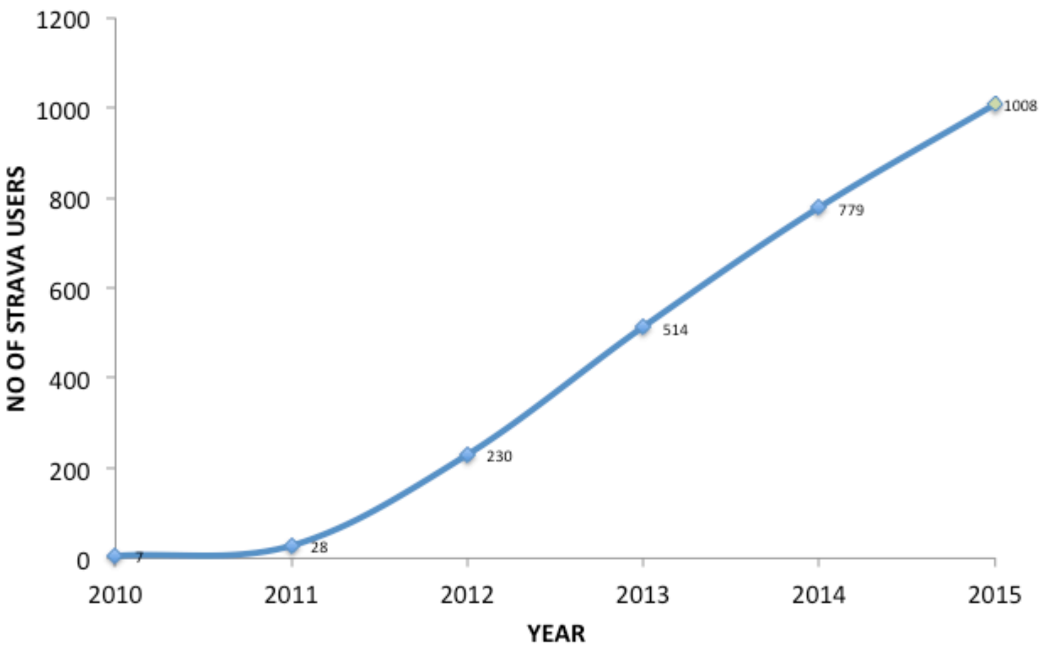

Figure 7. *Number of "fastest times" set in the selected segments since 2010. Data from 2010 to 2014 is based on observed data while the 2015 is extrapolated from data for the first 4 months of 2015.*

The next part of the study investigated if Strava could provide any information on the behaviour of cyclists at a number of different parts of the cycle network.

4.2 Case Study 1 - Restricted Turns

The junction of St Stephen's Green, Leeson St and Earlsfort Tce was selected as a case study of an area where cyclists may choose a route where turning in normally restricted. Three segments were created to represent the three potential exists from the junction and the results obtained from the Strava data set. Overall 5907 trips where analysed at the junction. Table 1 shows the main findings of the study.

Table 1. *Summary of user behaviour in the Restricted Turn case study.*

Route	Users	Trips	Trips per User	Avg Speed (km/h)
Leeson St	390	2374	6.1	15.2
Earlsfort Tce	334	3143	9.4	14.3
SSG South	108	399	3.7	17.7

While the vast majority of Strava users exited the junction using one of the permitted ways, a small but sizable minority (7% of all trips) continued onto St Stephen's Green South via a restricted turn (Figure 8).

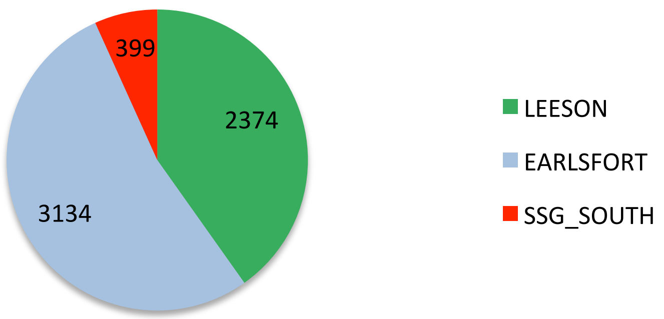

Figure 8. *Route choice of Strava users entering the Leeson St, Earlsfort Tce and St Stephen's Green Junction.*

4.3 Case Study 2 - One-way streets

The second case study examined the behaviour of Strava users when presented with the option to travel against traffic on a one-way street. This study was undertaken on St Stephen’s Green West in Dublin city centre. Analysis of the two segments; one representing the shortest recommended route for cycle traffic (recommended by Google Maps) and the other represented a straight line between the start and endpoints, which was made up of stretch where cyclists travelled against the flow of oncoming traffic (Figure 5)

Overall 610 trips were recorded over these two segments. Analysis of the data revealed that over 98% of the trips chose the direct route which involved going contrary to traffic on a one-way street, in the absence of any contra-flow bike lane. A summary of the key findings can be seen in Table 2. Interestingly, users that attempted the *correct* route only took that route once (1 trip per user).

Route	Users	Trips	Trips per User	Avg Speed (km/h)	Distance (m)
SSG- Harcourt Direct	232	599	2.4	15.4	250
SSG-Harcourt Correct	11	11	1	17.3	612

Table 2. *Summary of user behaviour in the One-way Street case study.*

4.4 Case Study 3 – Alternative cycle lane

In 2011 a dedicated cycle lane was opened along the Grand Canal marking the orbital route around the south of the city centre, with the intention of removing bicycles from the main flow of traffic. This study compared cycle traffic on a section of the cycle lane with cycle traffic on a parallel section of road. Both segments started at Baggot St Bridge and ended at Leeson St Bridge. There were a total of 12993 trips on both routes with no significant difference in route choice observed

(53% of trips were on the cycle track and 47% on the road). Table 3 shows in key findings in case study 3.

The gender distribution was also examined to see if route choice was subject to a gender bias. The results revealed no significant difference between the routes. 12% of the trips on the canal were by female users, compared with 11% trips on the road.

Route	Users	Trips	Trips per User	Avg Speed (km/h)	Distance (m)
Grand Canal Road	1038	6094	5.9	24.4	500
Grand Canal Cycle Lane	1183	6895	5.8	25.4	476

Table 3. *Summary of user behaviour in the Alternative Cycle Lane case study.*

4.4.1 Daily and seasonal trends

We also used these sections to examine if Strava may be capable of providing an insight into the changing patterns of bicycle traffic over time. We first looked at the monthly distribution of trips between 2013 and 2014 to see if any seasonal trend could be detected. The results reveal a clear peak in the summer months dropping off in the winter months (Figure 9).

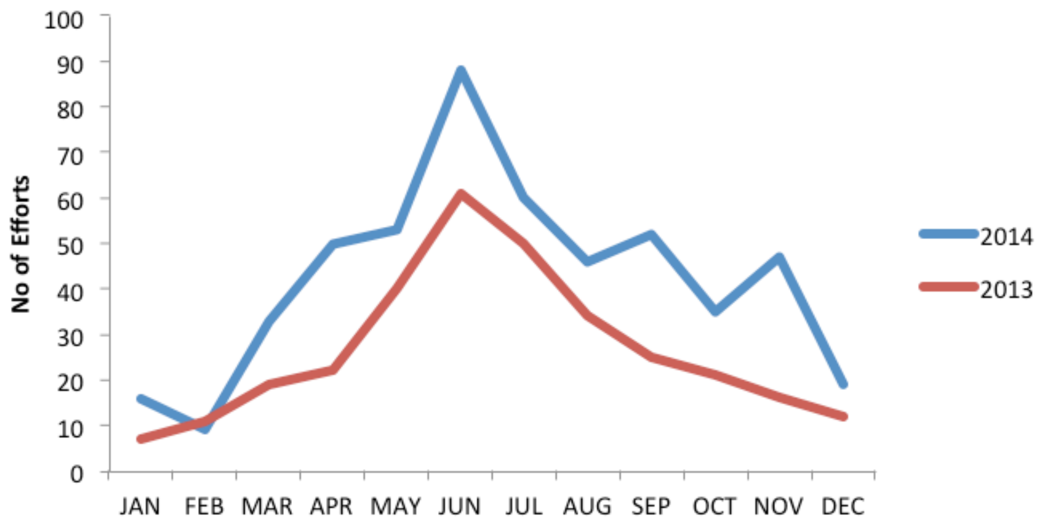

Figure 9. *Seasonal variations in the number of trips completed per month on the Grand Canal Orbital cycle path between 2013 – 2014.*

Finally, we investigated if Strava could reveal any insight into the flow of cycling traffic. To do this we created an additional segment on the road between Leeson St Bridge and Baggot St Bridge (west to east) and compared it to the segment going in the opposite direction. Figure 10 shows a scatter plot of the time of each effort from the leader board. While there is no clear correlation between the time of the effort and the rank in either set of data, there is a slight clustering of the fastest segments going from east to west in the late afternoon, corresponding with peak commuting time.

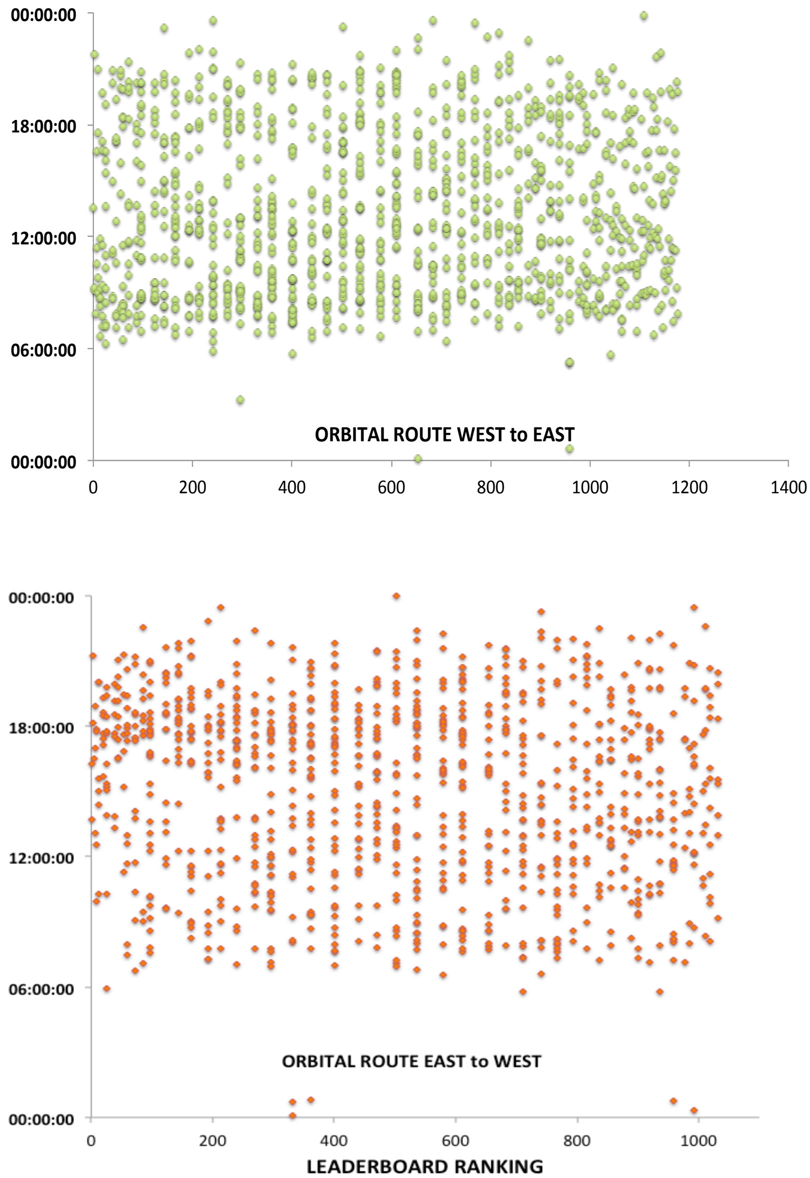

Figure 10. *Scatter plots displaying the time distribution of leader board segments on the Grand Canal Orbital route going from west to east (top) and from east to west (bottom).*

5. DISCUSSION

The current study examined the potential of user-generated data to provide information on cyclist behaviour in Dublin city centre. This was achieved by examining data logged to the exercise-tracking app Strava, in three different case studies. The first case study examined the potential of the app to detect route choice involving restricted turns. Secondly we investigated if the app could be used to understand how users behaved on one-way streets. The third case study investigated the use of Strava as a means to evaluate the efficacy of measures that have already been introduced to improve the cycle network. Finally we examined if this user-generated data could provide insight into the temporal variations in cycle traffic.

Briefly, our study demonstrated great potential for exercise tracking apps, such as Strava, to provide information on cyclist behaviour, on both an individual and population level. The study demonstrated that Strava data was able to detect the proportion of cyclists that made restricted right hand turns or choose the shorter straight route up a one-way street, rather than follow the flow of traffic through a longer route to their destination. The study also showed that measures to improve the cycling network in the city are being adopted by the cycling community, with significant numbers choosing to use the dedicated cycle lane running in parallel to the shared space of the road. At a population level the study also demonstrated the potential for Strava data to help understand the daily and seasonal variations in trip demand and highlight nodes in the network model where cyclists do not behave as predicted.

5.1 Limitations of the current data set

It must be noted that the current study is based on *segment* and *segment leader board* information made available by the Strava V3 API. The *segment* resource provides information on the total number of users and total number of trips on a given segment. The *segment leader board* containing more detailed information on each user's fastest time on that segment, including leader board rank, average speed, time and date of the trip. However, they do not provide information on any other trip that the cyclist may have taken on that particular segment. Therefore, when we are discussing details of individual trips (e.g. time and date of trip or average speed) we are talking only about the users "fastest time". We are making the assumption that the pattern of activity in these data sets is representative of all trips taken by users on these segments.

While this clearly limits the conclusions that may be drawn from the current study it may be seen as "proof-of-principle" that Strava can be a source of valuable data. Due to ethical considerations and time constraints we were unable to canvas Strava users and ultimately chose to concentrate the current study on these publically accessible data sets. However, it is important to note that developers may access a much more comprehensive data set if they receive authorization from individual athletes. The *activity* data set[7] contains information on every ride the user has undertaken, including activity type (commute/leisure/race), a map of the activity, and numerous averaged performance data. Detailed GPS data for each activity is also available through the *streams*[8] data set. This resource is particularly powerful, in that it allows developers to analyse user's location over time and, in addition to given just route information, can provide details on the movement pattern of the cyclist: if the user stops (e.g. at traffic lights or junctions) or changes in speed over different sections of their trip. Future studies accessing this data will be able to gain a much deeper insight into cyclist behaviours and traffic flows beyond the scope of the current study.

5.2 How is this data improving on the State-of-Art?

Until recently most information regarding cyclist route choice came from stated preference surveys and computer models of traffic flow (Vasic & Ruskin 2012; Caulfield et al. 2012). Verification of these models was problematic due to the cost

[7] http://strava.github.io/api/v3/activities/
[8] http://strava.github.io/api/v3/streams/

of data collection and computational complexity associated with developing high-resolution computer models of traffic flow. Indeed a critical review of the current research on stated preference surveys revealed serious flaws in the experimental design of many of the studies (Sener et al. 2009). Furthermore experiments that used GPS in parallel with a travel survey, found significant deviation between recorded activity and reported activity (Bricka et al. 2015). The same study raised the issues of poor response rate, bias in reported data and failure of memory as complicating factors in survey studies.

The National Cycling Framework Policy is basing many of its future plans for Ireland, and Dublin city centre in particular, on similarly limited information. Data on cyclist activity comes from survey-style collections in the form of the National Census (CSO 2012) and the National Household Travel Survey (Milward-Brown 2013). While these surveys give a general overview of the numbers of people cycling and the general commute which they undertake (i.e. moving from one town to another), they lack sufficient detail to create high-resolution models of cyclist movement.

Similarly, the studies that record actual cyclist behaviour in Dublin are limited to the Canal Cordon Count (NTA & DCC 2014), which looked at the mode of transport employed by people entering the Canal Cordon, an area broadly representing the borders of the city centre (Figure 11), during the morning peak (7-10 am) over two days in November between 2006 and 2013. These studies (ibid p4) showed a total number of cyclists crossing the canal of 90,479 in 2013.

The Canal Cordon Count (2013) clearly shows a growth in cycling as a mode of transport in the city centre (4.74% of all trips taken into the canal cordon in 2013 were by bicycle compared with 2.33% in 2006). The data however, is limited, both in terms of timeframe and scope. The surveys were carried over two days in November each year. It has been previously demonstrated that the variability in cycling demand is strongly dependent on weather conditions (Brandenburg et al. 2007) with cycling

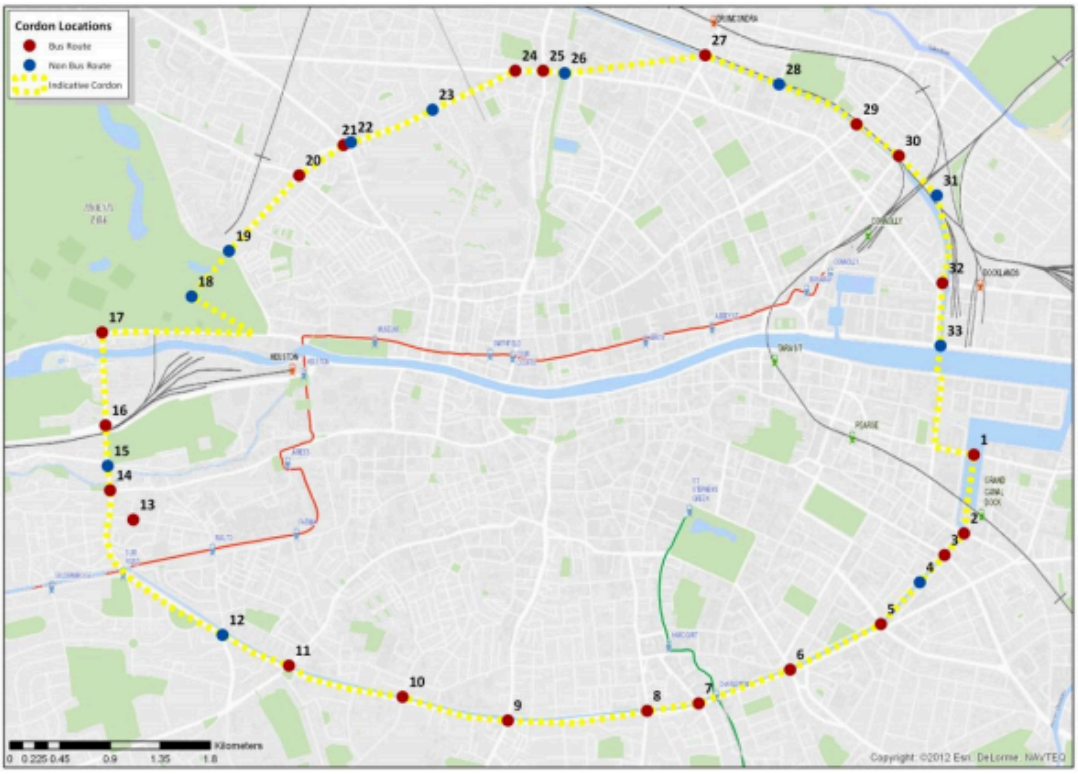

Figure 11. The *Canal Cordon* outlining the points at which traffic/cyclists were surveyed entering Dublin city centre in the 2014 study by the NTA and DCC (*image reproduced from the report*).

primarily a fine weather activity. However these studies were undertaken over two days in the month of November. With an average temperature of 6-8 °C and 26 days of precipitation (Met Eireann 2015), it may be argued that, although valuable, data collected in November is not representative of the real cycling trend, and is severely underestimating the peak cycling figures. Thomas et al (2009) reported that the weather conditions had a much greater effect on recreational cycling, further supporting the idea that the current data underestimates the peak cycling volume. The trend observed in the current study seems to support this hypothesis. The number of efforts recorded along a section of road on the orbital route in 2012 and 2013 demonstrated peak activity between May and July and a severe drop-off in the winter months. Due to the small population of cyclists on the road in November, relative to the summer peak, the Canal Cordon count may be detecting changes that are not full representative of the overall cycling trend. If the seasonal trend in these data is confirmed by more in-depth studies using the Strava activities and streams datasets, it may be worthwhile reconsidering the time of year that the counting studies are performed.

5.3 Is the Strava dataset representative of the general cycling population?

Over the three case studies in the current research, data was recorded for 27,729 trips on selected segments, created by approximately 3000 users. With the total amount of cyclists entering the city centre every day estimated to be around 9000 in 2013 (NTA & DCC 2014), it may be argued that Strava users represent a significant section of the cycling community. However the number of trips per user is quite low with the one-way street from St Stephen's Green having the highest value at almost 10 trips per user, although spread over 5 years that the data is being collected from, it equates to only 2 trips per user per year. It may be reasonable to assume that the number of trips per active user is much higher that this, and that the vast majority of users participate much less frequently. This has been seen throughout numerous crowdsourcing studies (Nielsen 2006). From the data currently available to us it appears that, although a large number of Strava users travel through the city by bicycle, they do not always have Strava activated.

People motivated enough to use a tracking app like Strava are generally more experienced riders, with Strava estimating that only 22% of the rides recorded on Strava were commute (personal communication form Brian Riordan to (Griffin & Jiao n.d.)). It is reasonable to assume therefore, that most of the people using Strava are more experienced cyclists. The research tells us that more experienced cyclists have a slight preference for on-road cycle lanes over separate cycle paths (Larsen & El-Geneidy 2011; Akar & Clifton 2015), however other studies report the opposite (Broach et al. 2012). It must be noted that almost all the data in these studies comes from stated preference surveys, no evidence-based study have addressed the issue. It is, therefore, difficult to conclude if the user profile has biased the results on separate cycle lane usage in the current study. There is a similar dearth of evidence-based data on the risk taking behaviour of cyclists such as making illegal turns and travelling against traffic on a one-way street. A number of qualitative studies suggest that experienced cyclist assess their risk of accident as lower than non-cyclists

(Martha & Delhomme 2009), and cyclist might, therefore be more willing to make risk manoeuvres. Johnson et al. (2013), also report that cyclists that cycle over 100 km per week were slightly more likely in infringe at red lights. In the context of the current study, the rate of infringement observed may over represent the rate seen by the general public, however much more study need to be performed before any conclusion may be made.

Males made up 85% of all users in the current study, with females making up less than 10%. Even allowing for the 5% of users where no gender information was available, it is clear that there is a overrepresentation of males compared with the general population where there are approximately 98 men per 100 women (CSO 2013). Even when considering the proportion of females that use cycling as a mode of transport to get to work (3 male: 1 female) (CSO 2014) it still does not account for the 8.5 fold difference in the current study.

The gender imbalance in cycling has been well established in the research, with numerous studies reporting small numbers of females using bicycles as a mode of transport (Garrard et al. 2008; Bell et al. 2006). Some studies suggest that the perceived risk of cycling is a deterrent for females and that when they do cycle are more likely to use routes separated from traffic (Garrard et al. 2008; Beecham & Wood 2013). In the current study we found no significant difference in the proportion of females choosing the dedicated Grand Canal cycle lane instead of the road. However, as previously pointed out, these data only pertain to the user's fastest trip and not every trip that a user has taken. Any conclusions about the risk-taking behaviour of cyclists in the current study should also consider the greater propensity of males to perform risky manoeuvres on a bicycle such as making illegal turns and breaking red lights (Johnson et al. 2013), therefore given the over-representation of males in the study population, the results may not be typical of the cycling community at large.

In a presentation at FutureStack14, Strava co-founder Mark Shaw spoke about the initial appeal of Strava to "data geeks" (FutureStack14 2014). The user base has increased massively since the early days of the app, with over one millions downloads of the app in Google Pay alone (Google Play n.d.), but how representative of the general cycling population are Strava users, and can any conclusions be drawn about the general cycling public from Strava activity data? A review of tracking apps by Navarro et al (2013), placed Strava as an app aimed more at motivated cyclists than beginners.

5.4 Potential of Strava as a Smart City sensor

Early studies using GPS for activity tracking used external GPS units and were subject to many limitations. One of the major limitations was the large amount of data loss associated with them. In their systematic review of the literature Krenn et al. (2011) examined the main reasons for this loss, finding a wide range of loss across the studies (2-92%). This was attributed to many factors such as "signal drop-outs, loss of device battery power, and poor adherence of participants to measurement protocols". Many of the studies included in the review involved the use of hand-held GPS units that required constant charging and for the participant to remember to carry them. In more recent times however, an alternative to independent GPS units has emerged. One of the most important features of modern mobile phone technology is the almost ubiquitous presence of GPS receivers. A study by Schaefer and Woodyer (2015) revealed that the accuracy of a range of recreational-grade GPS devices on popular smartphone models (Samsung S3mini, S4; iphone4, 5, 5c, and Sony Xperia E, P, Z) exceeded the manufacturer's specifications giving an accuracy to within 3.63 ± 1.41 m. This is in contrast to the relative inaccuracy of the early iPhone 3G, the first phone to fully integrate GPS technology (Zandbergen 2009). Taken together these results show the potential of the modern smartphone as a source of GPS data. It overcomes the problems of inaccuracy and the reluctance of users to carry and charge GPS units, since the majority of the population now habitually carries phones.

The implications of this are two-fold for the current study. Although data may be uploaded to Strava from external GPS units, the vast majority of activities are uploaded to the site from the Strava apps on Android and iPhone. The accuracy of the GPS data recorded on Strava, therefore, is reliable, meaning that any analysis done on the data sets is based on real activity recording and not subject to significant artefact. The presence of the app on smartphones also means that many of the issues surrounding charging and forgetting to carry external GPS devices it are reduced.

Furthermore, Strava also promotes engagement through the creation of social networks between cyclist, allowing users to “give kudos” or comment on other rider’s activities. As with other social media this model had been shown to be highly effective at establishing and maintain engagement with an application or website (Shin & Shin 2011). By creation of the “King of the Mountain” prize for the fastest time on any given segment, Strava also take advantage of social network gaming (ibid) by addressing many of the motivations (social interaction, self-presentation, escapism, entertainment and challenge/competition) that people have for playing online games such as *Texas HoldEm Poker* and *Farmville* (Lee et al. 2012). Turning what was originally an app to allow users to track their activity, into an app where users could compete with their peers, and even professional athletes for the fastest time.

Even allowing for the limited data set available, the most striking result of the current study was the almost complete disregard for the one-way street on St Stephen’s Green West. 98% of users took the direct route instead of following the advised flow of traffic to their destination on Harcourt St. This is keeping with other studies where cyclists will tolerate only small deviations from the minimum distance path (Menghini et al. 2010). While beyond the scope of the current study other environmental factors such as road quality, elevation and traffic lights may also influence a cyclist’s choice of route (ibid; Hood et al. 2013). Analysis of cyclist activity at the junction containing a restricted right turn in Case Study 1 revealed that 7% of all trips entering the junction exited via the restricted turn. This again maybe explained by cyclists not wishing to be diverted from the shortest path.

There are some interesting implications for models of cycling traffic when we consider the findings on the restricted turns and contrary use of one-way streets. Although the values are representative, the data clearly shows that cyclists do take routes that are not laid out in the traffic network. Since computer modelling plays a vital part in the forecast of future cycling traffic demands (Aherne and MacGerailt 2013), identifying areas of the network where cyclists deviate from the predefined matrix are vitally important for accurate predictions, particularly in areas where almost all users are not moving as predicted (i.e. One-way street). Futures studies may take advantage of user-generated GPS data to identify nodes and routes such as this where the reality of cycling behaviour differs significantly from the model.

In 2011 the Grand Canal cycle lane was opened along the orbital route to the south of Dublin city centre. The cycle lane is completely separate from any motorized traffic. In the current study we examined the use of this lane in comparison to the road, which runs parallel to the cycle lane but is a shared space with cars. Our results revealed that cycle traffic was split evenly between the two routes. Since the cycle lane was brought into to reduce the amount of traffic on the main road, the data suggests that this was a successful initiative – moving over half the traffic from shared space to the low-risk cycle lane. This study shows how user-generated data may play a role in evaluating the efficacy of changes to the cycle network.

5.5 The future of crowdsourced cycling data

In the coming years the IoE will create huge volumes of data about every of our lives and environment. These vast amounts of data are referred to as “Big Data”, and are expected to double annually, reaching 44ZB (44 trillion gigabites) by 2020 (EMC & IDC 2012). The development of API’s has meant that a lot of this data will be easily accessible to third parties (Berlind 2013), bringing with it the potential to determine if cyclist behaviour if correlated with other factors. For example correlating behaviour with weather conditions accessible through an API[9].

[9] http://www.wunderground.com/weather/api/

In addition, data from all collisions reported on Irish roads are recorded by the Road Safety Authority (RSA n.d.). A large proportion of these accidents occur in urban areas (There have been 344 collisions involving cyclists in Co. Dublin since 2009 (ibid)). Analysis of the collision map reveals that clusters of collisions may be seen in certain areas of the city. In cases like this, data generated by Strava users may be able to provide an insight into cyclist behaviour in these areas and help planners develop effective solutions.

Exercise tracking apps are becoming increasingly popular (Kahalaf 2014) and Strava has been joined by other extremely successful apps, such as *Nike+, Mapmyride*, *Endomundo*, *FitBit* and *RunKeeper* as top sellers in Google Play and Apple App Store. Although the current study focused on data generated by Strava users, many of the other exercise tracking applications provide API's to access their data sets. There is no reason why future studies could not investigate these data sets, in addition to Strava, to create an even more complete picture of cycling. Indeed, apps aimed at the wider cycling public like *Endomondo* and *Mapmyride* could provide information on the sections of the cycling population that may not be represented by data generated by Strava's more "motivated users.

All of the results in the current study come from data collected and analysed by the researchers on a relatively small scale, through actively querying the Strava API. It is interesting to note that Strava themselves have recognised that the data the have collected may have potential in this field and have recently launched the *Strava Metro*[10] service. This is a subscription service aimed at helping departments of transport, advocacy groups and city planners "make informed and effective decisions when planning, maintaining, and upgrading cycling and pedestrian corridors." Strava Metro provides data sets that are compatible with classic geographical information systems environments. All the data is anonymised preventing any privacy issues with regard to location data (It is also worth noting that in the raw Strava data, a privacy area extending for 1km around a persons home

[10] http://metro.strava.com/

prevents their home location being seen by other users). This product has already been used by a number of cities from Portland, Oregon, to London and Glasgow in the UK (road.cc 2014) This development may be seen as reinforcing the proof-of-principle explored in the current study.

6. CONCLUSION

Overall, this study has successfully demonstrated the proof-of-principle that location data, generated by users of the exercise-tracking app Strava, can provide insight into cyclist behaviour in Dublin city centre. While some caution must be taken in interpreting the quantitative data, given the limited data set available, the current study has demonstrated the potential value of mobile tracking applications in understanding cycling activity which may help governments and planners make better informed planning and policy decisions in the future.

LIST OF FIGURES

BIBLIOGRAPHY

Aherne, M.M.S., 2013. Draft Greater Dublin Area Cycle Network Plan. In *Engineers Ireland*.

Akar, G. & Clifton, K., 2015. Influence of Individual Perceptions and Bicycle Infrastructure on Decision to Bike. *Transportation Research Record: Journal of the Transportation Research Board*. Available at: http://trrjournalonline.trb.org/doi/abs/10.3141/2140-18 [Accessed May 15, 2015].

Ali, S. et al., 2014. Sensors and Mobile Phones: Evolution and State-of-the-Art.: Discovery Service for Trinity College Dublin (University). *Pakastani Journal of Science*, pp.385–400. Available at: http://eds.a.ebscohost.com.elib.tcd.ie/eds/pdfviewer/pdfviewer?sid=95a2ab36-1b37-44e1-b934-65627275bd7d%40sessionmgr4005&vid=0&hid=4211 [Accessed May 15, 2015].

Andersen, L.B. et al., 2002. [Mortality associated with physical activity in leisure time, at work, in sports and cycling to work]. *Ugeskrift for laeger*, 164(11), pp.1501–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11924475 [Accessed April 30, 2015].

Avelar, E. et al., 2014. Interoperability issues on heterogeneous wireless communication for smart cities. *Computer Communications*, 58, pp.4–15. Available at: http://www.sciencedirect.com/science/article/pii/S0140366414002485 [Accessed January 23, 2015].

Beecham, R. & Wood, J., 2013. Exploring gendered cycling behaviours within a large-scale behavioural data-set. *Transportation Planning and Technology*, 37(1), pp.83–97. Available at: http://www.tandfonline.com/doi/abs/10.1080/03081060.2013.844903#.VVO7DJNVhBc [Accessed January 3, 2015].

Bell, A.C., Garrard, J. & Swinburn, B.A., 2006. Active transport to work in Australia: is it all downhill from here? *Asia-Pacific journal of public health / Asia-Pacific Academic Consortium for Public Health*, 18(1), pp.62–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16629440 [Accessed May 13, 2015].

Berlind, D., 2013. ProgrammableWeb's Directory Hits 10,000 APIs. And Counting. *ProgrammableWeb*. Available at: http://www.programmableweb.com/news/programmablewebs-directory-hits-10000-apis.-and-counting./analysis/2013/09/23.

Berlingerio, M. et al., 2013. *Machine Learning and Knowledge Discovery in Databases* H. Blockeel et al., eds., Berlin, Heidelberg: Springer Berlin Heidelberg. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-84886497592&partnerID=tZOtx3y1 [Accessed May 14, 2015].

Brabham, D.C.., 2008. Crowdsourcing as a Model for Problem Solving. *Convergence*, 14(1), pp.75–90.

Bricka, S. et al., 2015. Household Travel Surveys with GPS. *Transportation Research Record: Journal of the Transportation Research Board*. Available at: http://trrjournalonline.trb.org/doi/abs/10.3141/2105-07?journalCode=trr [Accessed May 13, 2015].

Broach, J., Dill, J. & Gliebe, J., 2012. Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*, 46(10), pp.1730–1740. Available at: http://www.sciencedirect.com/science/article/pii/S0965856412001164 [Accessed October 15, 2014].

Campbell, A.T. et al., 2006. People-centric urban sensing. In *Proceedings of the 2nd annual international workshop on Wireless internet - WICON '06*. New York, New York, USA: ACM Press, p. 18–es. Available at: http://dl.acm.org/citation.cfm?id=1234161.1234179 [Accessed May 14, 2015].

Caulfield, B., Brick, E. & McCarthy, O.T., 2012. Determining bicycle infrastructure preferences – A case study of Dublin. *Transportation Research Part D: Transport and Environment*, 17(5), pp.413–417. Available at: http://www.sciencedirect.com/science/article/pii/S1361920912000363 [Accessed February 10, 2015].

Coffin, A.W., 2007. From roadkill to road ecology: A review of the ecological effects of roads. *Journal of Transport Geography*, 15(5), pp.396–406. Available at: http://www.sciencedirect.com/science/article/pii/S0966692306001177 [Accessed July 11, 2014].

CSO, 2012. Census 2011. Profile 10: Door to Door. *Census 2011*.

CSO, 2013. *Men and Women in Ireland 2013*,

CSO, 2014. *National Travel Survey: How we travelled*, Available at: http://www.cso.ie/en/releasesandpublications/ep/p-nts/nationaltravelsurvey2012and2013/overview/#.VVNa65NVhBc.

Department of Transport, T.& S., 2009. *National Cycle Policy Framework 2009-2020*, Available at: http://www.smartertravel.ie/sites/default/files/uploads/0902 02 EnglishNS1274 Dept. of Transport_National_Cycle_Policy_v4%5B1%5D.pdf.

DuVander, A., 2011. API Growth Doubles in 2010, Social and Mobile are Trends. *ProgrammableWeb*. Available at:

http://www.programmableweb.com/news/api-growth-doubles-2010-social-and-mobile-are-trends/2011/01/03.

Eireann, M., 2015. Monthly Averages. Available at: http://www.met.ie/climate/monthly-data.asp?Num=532.

EMC & IDC, 2012. *Digital Universe Invaded By Sensors*, Available at: http://www.emc.com/about/news/press/2014/20140409-01.htm.

Federation, E.C., 2011. *Cycle more often 2 cool down the planet! – quantifying CO2 savings of cycling.*,

Frank, L.D., Andresen, M.A. & Schmid, T.L., 2004. Obesity relationships with community design, physical activity, and time spent in cars. *American journal of preventive medicine*, 27(2), pp.87–96. Available at: http://www.sciencedirect.com/science/article/pii/S074937970400087X [Accessed February 3, 2015].

FutureStack14, 2014. The Story Behind Strava Metro. Available at: https://www.youtube.com/watch?v=0zAG3j5JAAc.

Garrard, J., Rose, G. & Lo, S.K., 2008. Promoting transportation cycling for women: the role of bicycle infrastructure. *Preventive medicine*, 46(1), pp.55–9. Available at: http://www.sciencedirect.com/science/article/pii/S0091743507003039 [Accessed March 1, 2015].

Gartner, 2013. *Personal Worlds and the Internet of Everything Are Colliding to Create New Markets*, Available at: http://www.gartner.com/newsroom/id/2621015 [Accessed May 12, 2015].

Giffinger, R. et al., 2006. *Smart Cities: City-ranking of European Medium-Sized Cities*,

Girbau, D. et al., 2012. Passive Wireless Temperature Sensor Based on Time-Coded UWB Chipless RFID Tags. *IEEE Transactions on Microwave Theory and Techniques*, 60(11), pp.3623–3632. Available at: http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=6296730 [Accessed May 1, 2015].

Griffin, G.P. & Jiao, J., Where does Bicycling for Health Happen? Analysing Volunteered Geographic Information through Place and Plexus. *Journal of Transport & Health*. Available at: https://soa.utexas.edu/sites/default/disk/Where does Bicycling for Health Happen, Journal of Transport and Health accepted manuscript.pdf.

Hallal, P.C. et al., 2012. Global physical activity levels: surveillance progress, pitfalls, and prospects. *Lancet*, 380(9838), pp.247–57. Available at: http://www.sciencedirect.com/science/article/pii/S0140673612606461 [Accessed July 10, 2014].

Hancke, G., Silva, B. & Hancke, Jr., G., 2012. The Role of Advanced Sensing in Smart Cities. *Sensors*, 13(1), pp.393–425. Available at: /pmc/articles/PMC3574682/?report=abstract [Accessed March 11, 2015].

Harrison, M., 2012. The Problem of Road Congestion: The Futility of "Avoidable Cost" Estimates: Discovery Service for Trinity College Dublin (University). *Agenda: A Journal of Policy Analysis & Reform*, 19(2), pp.77–86. Available at: http://eds.b.ebscohost.com/eds/pdfviewer/pdfviewer?sid=7e11c6ac-805a-4395-a7b8-7c8d1f3edc73%40sessionmgr110&vid=0&hid=114 [Accessed April 30, 2015].

Hood, J., Sall, E. & Charlton, B., 2013. A GPS-based bicycle route choice model for San Francisco, California. *Transportation Letters*. Available at: http://www.maneyonline.com/doi/abs/10.3328/TL.2011.03.01.63-75 [Accessed May 13, 2015].

Horton, D., Cox, P. & Rosen, P., 2012. Cycling and Society. In *Cycling and Society*. Ashgate Publishers Ltd.

Howe, J., 2006. The Rise of Crowdsourcing. *Wired Magazine*, pp.1–5. Available at: http://www.wired.com/wired/archive/14.06/crowds_pr.html.

Hydén, C., 1997. *WALCYNG - Walking & Cycling instead of short car trips.*,

IPCC, 2013. *Summary for Policymakers. In: Climate Change 2013: The Physical Science Basis.* Stocker & V. B. and P. M. M. T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, eds., Cambridge, United Kingdom and New York, NY, USA.: Cambridge University Press.

Johnson, M. et al., 2013. Why do cyclists infringe at red lights? An investigation of Australian cyclists' reasons for red light infringement. *Accident; analysis and prevention*, 50, pp.840–7. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-84870298778&partnerID=tZOtx3y1 [Accessed February 26, 2015].

Kahalaf, S., 2014. *Health and Fitness Apps Finally Take Off, Fueled... | Flurry Insights Blog*, Available at: http://flurrymobile.tumblr.com/post/115192181465/health-and-fitness-apps-finally-take-off-fueled [Accessed May 14, 2015].

Krenn, P.J. et al., 2011. Use of global positioning systems to study physical activity and the environment: a systematic review. *American journal of preventive medicine*, 41(5), pp.508–15. Available at: http://www.sciencedirect.com/science/article/pii/S0749379711005460 [Accessed December 1, 2014].

Lanzendorf, M. & Busch-Geertsema, A., 2014. The cycling boom in large German cities—Empirical evidence for successful cycling campaigns. *Transport Policy*, 36, pp.26–33. Available at:

http://www.sciencedirect.com/science/article/pii/S0967070X14001498 [Accessed May 12, 2015].

Larsen, J. & El-Geneidy, A., 2011. A travel behavior analysis of urban cycling facilities in Montréal, Canada. *Transportation Research Part D: Transport and Environment*, 16(2), pp.172–177. Available at: http://www.sciencedirect.com/science/article/pii/S1361920910001100 [Accessed January 17, 2015].

Lee, J., Lee, M. & Choi, I.H., 2012. Social network games uncovered: motivations and their attitudinal and behavioral outcomes. *Cyberpsychology, behavior and social networking*, 15(12), pp.643–8. Available at: http://online.liebertpub.com/doi/abs/10.1089/cyber.2012.0093 [Accessed April 26, 2015].

Maibach, E., Steg, L. & Anable, J., 2009. Promoting physical activity and reducing climate change: opportunities to replace short car trips with active transportation. *Preventive medicine*, 49(4), pp.326–7. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-70349997049&partnerID=tZOtx3y1 [Accessed May 15, 2015].

Martha, C. & Delhomme, P., 2009. Risk comparative judgments while driving a car among competitive road cyclists and non-cyclists. *Transportation Research Part F: Traffic Psychology and Behaviour*, 12(3), pp.256–263. Available at: http://www.sciencedirect.com/science/article/pii/S1369847808000995 [Accessed May 15, 2015].

Menghini, G. et al., 2010. Route choice of cyclists in Zurich. *Transportation Research Part A: Policy and Practice*, 44(9), pp.754–765. Available at: http://www.sciencedirect.com/science/article/pii/S0965856410001187 [Accessed November 14, 2014].

Milward-Brown, 2013. National Transport Authority National Household Travel Survey 2012.

Mindell, J, Rutter, H, W.S., 2011. Urban Transportation and Human Health. In *In Encyclopedia of Environmental Health*. Elsevier B.V., pp. 578–589.

Minson, S.E. et al., 2015. Crowdsourced earthquake early warning. *Science Advances*, 1(3), pp.e1500036–e1500036. Available at: http://advances.sciencemag.org/content/1/3/e1500036.abstract [Accessed April 13, 2015].

Mitton, N. et al., 2012. Combining Cloud and sensors in a smart city environment. *EURASIP Journal on Wireless Communications and Networking*, 2012(1), p.247. Available at: http://jis.eurasipjournals.com/content/2012/1/247 [Accessed March 5, 2015].

Mueller, N. et al., 2015. Health impact assessment of active transportation: A systematic review. *Preventive medicine*. Available at: http://www.sciencedirect.com/science/article/pii/S0091743515001164 [Accessed April 27, 2015].

Nations, U., 2013. *World Urbanization Prospectus 2012*, New York. Available at: http://esa.un.org/unpd/wup/index.htm.

Navarro, K.F., Gay, V. & Golliard, L., 2013. SocialCycle what can a mobile app do to encourage cycling? In *Local Computer Networks Workshops (LCN Workshops), 2013 IEEE 38th Conference*. Syndey, NSW, pp. 24–30. Available at: http://ieeexplore.ieee.org/xpls/icp.jsp?arnumber=6758494 [Accessed May 14, 2015].

Nielsen, J., 2006. Participation Inequality: The 90-9-1 Rule for Social Features. Available at: http://www.nngroup.com/articles/participation-inequality/ [Accessed May 13, 2015].

NTA & DCC, 2014. *Report on trends in mode share of vehicles and people crossing the Canal Cordon: 2006-2013*,

PlanIT, 2013. PlanIT Valley. Available at: http://planitvalley.org/.

Play, G., Strava App in Google Play. Available at: https://play.google.com/store/apps/details?id=com.strava&hl=en.

Quill, H., 1966. John Harrison: the man who found longitude. *London*.

Ravindranath, M., 2014. Crowdsourced map app Waze plans government data exchange for road conditions - The Washington Post. *The Washington Post*. Available at: http://www.washingtonpost.com/business/on-it/crowdsourced-map-app-waze-plans-government-data-exchange-for-road-conditions/2014/10/02/bec51eea-4a5b-11e4-a046-120a8a855cca_story.html [Accessed May 13, 2015].

Road.cc, 2014. Strava moves into “big data” - London & Glasgow already signed up to find out where cyclists ride | road.cc. *road.cc*. Available at: http://road.cc/content/news/118098-strava-moves-big-data-london-glasgow-already-signed-find-out-where-cyclists-ride [Accessed May 14, 2015].

RSA, Ireland Road Collisions Statistics. Available at: http://rsa.ie/RSA/Road-Safety/Our-Research/Collision-Statistics/Ireland-Road-Collisions/.

Salim, F. & Haque, U., 2015. Urban computing in the wild: A survey on large scale participation and citizen engagement with ubiquitous computing, cyber physical Systems, and internet of Things. *International Journal of Human-Computer Studies*. Available at:

http://www.sciencedirect.com/science/article/pii/S1071581915000488 [Accessed March 31, 2015].

Schaefer, M. & Woodyer, T., 2015. Assessing absolute and relative accuracy of recreation-grade and mobile phone GNSS devices: a method for informing device choice. *Area*, 47(2), pp.185–196. Available at: http://doi.wiley.com/10.1111/area.12172 [Accessed May 6, 2015].

Sener, I.N., Eluru, N. & Bhat, C.R., 2009. An analysis of bicycle route choice preferences in Texas, US. *Transportation*, 36(5), pp.511–539. Available at: http://link.springer.com/10.1007/s11116-009-9201-4 [Accessed May 13, 2015].

Shin, D.-H. & Shin, Y.-J., 2011. Why do people play social network games? *Computers in Human Behavior*, 27(2), pp.852–861. Available at: http://www.sciencedirect.com/science/article/pii/S0747563210003535 [Accessed May 11, 2015].

Strava, Frequently Asked Questions | Strava. *2014*. Available at: http://metro.strava.com/faq/ [Accessed May 12, 2015].

Strickland, E., 2011. Cisco bets on South Korean smart city. *IEEE Spectrum*, 48(8), pp.11–12. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5960147 [Accessed May 11, 2015].

Surowiecki, J., 2005. The wisdom of crowds : why the many are smarter than the few / James Surowiecki. Available at: http://stella.catalogue.tcd.ie/iii/encore/record/C__Rb12396045__Sjames surowiecki the wisdom of crowds__Orightresult__U__X2?lang=eng&suite=cobalt [Accessed May 13, 2015].

Ueno, K., Asai, T. & Amemiya, Y., 2011. Low-power temperature-to-frequency converter consisting of subthreshold CMOS circuits for integrated smart temperature sensors. *Sensors and Actuators A: Physical*, 165(1), pp.132–137. Available at: http://www.sciencedirect.com/science/article/pii/S0924424710001354 [Accessed May 12, 2015].

Vasic, J. & Ruskin, H.J., 2012. Agent-based Space-time Discrete Simulation of Urban Traffic Including Bicycles. *Procedia Computer Science*, 10, pp.860–865. Available at: http://www.sciencedirect.com/science/article/pii/S1877050912004693 [Accessed May 13, 2015].

Weiser, M., 1999. The computer for the 21 st century. *ACM SIGMOBILE Mobile Computing and Communications Review*, 3(3), pp.3–11. Available at: http://dl.acm.org/citation.cfm?id=329124.329126 [Accessed August 13, 2014].

Williams, A. & Dourish, P., 2006. Imagining rhe City: The Cultural Dimensions of Urban Computing. *Computer*, 39(38-43).

Williams, A.M., 2012. *King of the Mountain: A Rapid Ethnography of Strava Cycling*. University College London.

Zandbergen, P.A., 2009. Accuracy of iPhone Locations: A Comparison of Assisted GPS, WiFi and Cellular Positioning. *Transactions in GIS*, 13, pp.5–25. Available at: http://doi.wiley.com/10.1111/j.1467-9671.2009.01152.x [Accessed May 6, 2015].

APPENDIX

Sample Python code used to obtain leader board and segment information from the Strava V3 API.

```python
'''
PROGRAM TO QUERY THE STRAVA V3 API
Mark Dunleavy APRIL 2015

'''


import json, requests
import urllib.request
import urllib

#variables obtained when developer registered as user of the Strava API
client_id = "4787"
client_secret = "dc24c30db29e1d07de160065c846cd4e7816050e"
access_token = "7756a413b9bb37b0cd5f83c7cfdfcd51b714da1e"


#Segement id's
#the approrpiate id in inserted into the request url

SGSC_HARCOURT_DIRECT = "9383716"
SGSC_HARCOURT_CORRECT = "9379983"

LEESON = "9346620"
EARLSFORT = "9346611"
SSG_SOUTH = "9346593"

CANAL_CYCLELANE = "9380597"
CANAL_ALTERNATIVE = "9416013"

BAGGOT_LEESON_CANAL = "4703736"
BAGGOT_LEESON_ROAD = "1829289"
LEESON_BAGGOT_ROAD = "2006269"


#making a request to get leaderboard information for certain segments
url = "https://www.strava.com/api/v3/segments/"+LEESON_BAGGOT_ROAD+"/leaderboard?access_token=7756a413b9bb37b0cd5f83c7cfdfcd51b714da1e&per_page=200&page=7"
print(url)
response = requests.get(url)
data = response.json()
entries = data['entries']
total_entries = len(entries)

#making a request to get segment information
segment_url = "https://www.strava.com/api/v3/segments/"+LEESON_BAGGOT_ROAD+"?access_token=7756a413b9bb37b0cd5f83c7cfdfcd51b714da1e&page=1"
poly_response = requests.get(segment_url)
poly_data = poly_response.json()
polyline = poly_data['map']['polyline']

print (polyline)


print("Total activities retrieved: {total}".format(total=total_entries))

#turn read information to a string and print it to a .csv file
def xstr(s):
        if s is None:
                return ''
        else:
                return str(s)

print ('Opening file to write...')
outputfile = open('leeson.csv', 'a') # Append the entry for each run to this file

schema = '"ID","Gender","distance", "Moving time", "average Speed (km/h)","rank", "date_time"\n'
print ('Writing schema...')

outputfile.write('"segment_id",' + xstr(poly_data["id"])+ '\n')
outputfile.write('"segment_name",' + xstr(poly_data["name"])+ '\n')
outputfile.write('"effort count",' + xstr(poly_data["effort_count"])+ '\n')
outputfile.write('"no. of times segment used",'+ xstr(data['entry_count'])+ '\n')

outputfile.write('"polyline",' + polyline + '\n')
outputfile.write(schema)

runs = 0
print ('Writing activities...')
print (str(total_entries))
for x in range(total_entries-1,-1,-1):
    curr_activity = entries[x]
    avg_speed = (curr_activity['distance'])/(curr_activity['moving_time'])*3.6

    print("Writing segment {i} Athlete_ID: {act_id}".format(i=x, act_id=curr_activity['athlete_id']))
    record = ''
    record = record + '"' + xstr(curr_activity['athlete_id']) + '",'
    record = record + '"' + xstr(curr_activity['athlete_gender']) + '",'
    record = record + '"' + xstr(curr_activity['distance']) + '",'
    record = record + '"' + xstr(curr_activity['moving_time']) + '",'
    record = record + '"' + xstr(avg_speed) + '",'
    record = record + '"' + xstr(curr_activity['rank']) + '",'
    record = record + '"' + xstr(curr_activity['start_date_local']) + '"\n'
    outputfile.write(record)
    runs+=1
print("Printing to .csv complete")
outputfile.close()
```