What boosts cycle commuting? Investment, demographics and road safety in England between 2001 and 2011 ==

Introduction

During the first decades of the 21st century the nature of threats to the long-term wellbeing of the planet and its inhabitants have become increasingly clear. We have more data than ever about the problems, from satellites monitoring the impacts of climate change (citep{GRL51730} to detailed global statistics on likely decline rates of unconventional oil and gas supplies (citep{hughes2013reality}. Yet collectively humanity seems unable to act on these cues, raising the spectre of past civilisational collapse and bleak prospects for future generations (citep{Beddoe2009, ehrlich2013can}. Although the problems are well understood, effective solutions have thus far been elusive. In some instances unintended consequences have meant that proposed solutions have only worsened the situation while leaving root causes unaddressed (e.g. (citealp{sheehan2009biofuels}). Various theories have been developed to explain our collective failure (citep{Berners-Lee2013}. Generally these revolve around a failure to understand that long-term problems at the interface between the human economy and the Earth's physical systems are entrenched, complex and interdependent (citep{smil}.

It is within this wider context that questions of transport and health should be considered (citep{Woodcock2009}. The 'global energy challenge' (citep{Nocera2006} underlies this research, even though it is quite specific geographically (focussing solely on England), temporally (analysing decadal shifts) and sectorally, with its focus on transport to work. Yet the subject matter was influenced by and contributes to discussion of humanity's short term sustainability and longer term survival. Many of the benefits of cycling uptake now may only be realised long in the future, yet many papers evaluating bicycle policies focus solely on 'the here and now'. This bias towards the present is present in many areas, not least in economics, and deserves wider discussion in the context transport planning. For now, however, let us press on with the specific matter of the paper: cycling to work.

What is striking about the bicycle is that it is an accessible technology (everyone worldwide can realistically aspire to own one) that tackles a very wide range of issues. Interventions that can simultaneously tackle health, energy and economic problems are the mythical 'golden bullets' that policy makers dream of yet almost never encounter: the sustainability challenge can be seen as an overinflated balloon: squeeze one area of it and a bulge generally appears elsewhere (citep{Berners-Lee2013}. However, a strong argument can be made that policies which increase the rate of cycling for personal transport represent one such 'golden bullet', tackling an array of issues in one simple step. Commuting is an important reason for personal travel, accounting for around a fifth of trips in high income nations and is inflexible compared with other forms of transport:

people have to get to work (citep{Sexton2010}

This paper starts from the premise that cycling is a desirable outcome for health and other reasons. There is much research supporting this premise and on the benefits of active travel more broadly. There is less research, however, about which policies are most likely to boost cycling in different places at the sub national level. There have been international comparative studies on why cycling flourishes in some countries whilst barely accounting for 1% of trips in others (citep{Pucher}. Research seeking to explain variation in bicycle use between smaller administrative zones is described in section x.

There are important social, cultural and spatial-economic differences between nation states, meaning that what works in one country may not always work in another. In addition, we argue that explaining national-level variation is a higher priority from the perspective of local transport planners - those who implement the details of strategic plans and decide precisely how investment in cycling is spent.

What conditions are associated with growth in cycling at the national level? A central motivation of this paper is to shed light on potential answers to this question. The aim is to inform the policy making process by supplying evidence about which important factors are within the scope of local transport planners to influence. (For example, if topography and weather are found to completely explain growth in cycling, there is little transport planners can do compared with if provision of cycle paths is found to explain much of the change.) Within this question, the following hypotheses were generated, based on the literature on the spatial distribution of cycling (see section x). The expectations were, certeris parabis (all things being being equal), that:

- Areas with high incomes would see disprortional increases in cycle commuting.
- Length of cycle paths (a proxy for investment in cycling) will be associate with higher than expected growth rates in cycling.
- A good safety record on cycling (relative to the number of bicycle commuters) will be associated with high growth rates.
- Areas that had received central government support for cycling, through the Cycling Demonstration Towns initiative, would see growth in cycle commuting.

-

The benefits of cycling

Unlike some other types of environmental intervention, there are no major downsides to bicycle uptake: cyclists (often unwittingly) provide many benefits to the wider system whilst simply enjoying a fast and healthy for of tranpsort. To be specific, there is strong evidence of social, environmental and economic benefits of cycling in the following areas:

- improved health of cyclists (Oja, 2011; Rojas-Rueda et al., 2011; Saunders et al., 2013).
- economic benefits at the individual level, for example due to lower trip times (Borjesson and Jonas, 2011), and reduced car use (Semlyen, 2005).
- economic benefits for society at large via reduced public health bills (Rutter et al. 2013; Jarrett et al. 2012) and wider impacts (Cavill et al., 2008; Krizek, 2007; Saelensminde, 2004).
- environmental benefits including lower greenhouse gas emissions and demand for roads and motor vehicles (Lenzen, 1999; Lindsay and Macmillan, 2011; Lovelace et al., 2011).
- reduction in congestion during rush hour - this is a particular benefit of cycle commuting (as opposed to leisure cycling) via improved traffic flow (Arnott et al., 2005; Downs, 2004).
- bicycles pose a lower risk to other road users than do cars, with benefits for social equality (Jacobsen et al., 2009; Furness, 2010).
- 'wide boundary' impacts including heightened sociability of public space and the hope that society may one day be able to operate without burning valuable finite resources (Furness, 2010).

Pro-cycling policies: a UK perspective

Due in part to these benefits, there has been a noticeable increase in political commitment to cycling in many countries in recent years. In the UK, for example, Prime Minister David Cameron announced that "we want to see cycling soar" (Prime Minister's Office, 2013) as well as providing a more specific statement of intent: "This government wants to make it easier and safer for people who already cycle as well as encouraging far more people to take it up" (ibid).

Within this context of widespread political and evidence-based support of policies to promote modal shift to cycling, a major barrier is specific evidence on the effectiveness of different interventions. Clearly, the number of new cyclists resulting from a specific policy intervention cannot precisely be known. However, using an analogy from medicine, 'dose-response' type studies can greatly help predict the impact of planned interventions. Transport planning is a long-term process with even longer-term impacts and such evidence can aid the strategic decision making process (Schweizer and Rupi, 2014) With limited public funds, it is critical to maximise the cost-effectiveness ratio of cycle-related expenditure.

More numerous and rigorous studies

could therefore help increase the rate of cycling in many areas, assuming that funding and political will are abundant.

The purpose of this paper is to help fill this knowledge gap by analysising the change in bicycle commuting across administrative zones across the UK. A geographically weighted regression methodology will be used to estimate how effective different types of intervention - including investment from the Cycling Demonstration Towns (CDT) initiative and number of cycle paths - have been.

Literature review: the spatial distribution of cycling

The Census travel to work statistics have been used in a number of studies to shed light on the spatial distribution of travel patterns overall.

However, the number of studies focussed on active travel, and cycling in particular, is much lower. Goodman (2013) provides a detailed and up-to-date account of the spatial distribution of walking and cycling at the national level across England and Wales and describes how it has changed between 2001 and 2011. The overall rate of walking and cycling to work was found to have changed little, but important important regional differences were identified: at the regional and local authority level cycling growth was largely concentrated in high density urban centres, notably London and Bristol. The paper also provided new insight into the shifting social distribution of travel to work mode: affluence is associated with motorised modes, although this association is weakening. Walking and cycling was found to have grown most in the least deprived areas, with car use tending to fall in the wealthiest areas (Goodman 2013). The underlying reasons for these shifts was not explored: "Future analyses could also explore associations with geographic factors such as hilliness, climate and land use patterns; although outside the scope of this paper, these may play a key role in explaining local and regional variation" (Goodman, 2013, p.9). The present paper follows this suggestion by exploring the factors associated with growth in bicycle commuting.

Multivariate regression models have been used to help explain the spatially uneven distribution of the proportion of commutes made by bicycle. Parkin et al. (2007) used a logistic regression model to identify the most important factors related to cycling at the ward level in England and Wales. 81% of the variability in the proportion of commuters cycling to work could be explained by the model. The proportion of white male commuters (positive), number of cars per household (negative) and income deprivation (negative) were found to be powerful socio-economic explanatory variables. Interestingly, hilliness was found to have a negative impact on cycle commuting in the model. Road traffic (via the proxy 'transport demand intensity') was negatively related to cycling to work whereas the proportion of off-road routes had a positive impact. These latter findings indicating transport infrastructure is an important policy area for encouraging cycling.

Another useful output from this paper was the estimation of a 'saturation point', referred to as 'carrying capacity' in population ecology (Lovelace et al 2011), a theoretical upper limit on the proportion of people cycling to work in any particular area. This was calculated to be 43% of trips, higher than any wards in the UK, but comparable with the proportion cycling to work in some Dutch areas (Parkin et al. 2011)

Data

Census data on commuting provide the highest spatial resolution of transport data in the UK, down to the level of Output Areas, consisting of roughly 100 households each. The decadal Census of population also has the highest response rate and number of participants of any national survey, due to the nationwide coverage of the census and the legal requirement to complete it. Offering many opportunities to transport researchers. The downsides to census data are that it only incorporates one reason for trip (commuting) and its poor temporal resolution. However, recent work shows that the modal split for commuting is highly correlated with modal split for all trips ($r > 0.9$ for private modes and public transport, dropping to $r = 0.77$ for cycling), suggesting that commuting is a reasonable proxy for travel behaviour overall (Goodman2013). In addition, the annual publication of results from the National Travel Survey provide higher temporal resolution to complement the 10 year cycle of the census.

The input data for this study were tables of commuter mode share by administrative zone between the 2001 and 2011 Census of population. In line with the principle of parsimony, the modelling approach was to start simple (with large geographical zones) and move to analyse higher resolution spatial datasets as the analysis progressed.

A data problem that had to be overcome early in the analysis was the conversion of 2001 354 Local and Unitary Authorities (combined with the 'merge' function in QGIS) into 2011 local authority areas (LAs), composed of English Districts, Unitary Authorities and London Boroughs. As shown in Fig. x, there are 8 2011 LAs which encapsulate many (38) 2001 administrative zones. The result of this process of spatial aggregation was all 324 'lower tier' 2011 LAs, for which the travel to work data was directly comparable between 2001 and 2011.

The other scales of analysis used in this study were Medium Super Output Areas and Lower Super Output Areas (MSOAs and LSOAs with average working populations of x and y people, respectively). The smallest administrative geography in England is the Output Area which contain on average around 100 households. These were omitted from the analysis because small values are

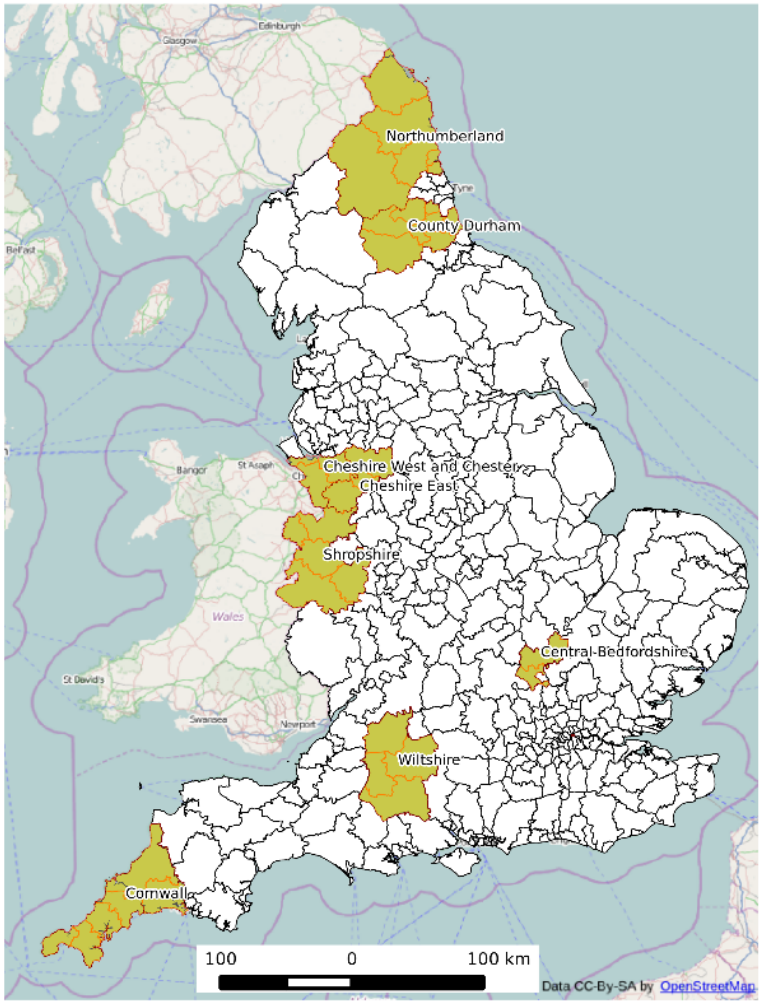

Figure 1: plot of chunk unnamed-chunk-2

"randomly adjusted" in the Census tables for low counts (Lovelace, 2014; see for example http://www.nomisweb.co.uk/livelinks/4652.xls).

The data pre-processing for these areas was simpler as there has been less change between 2001 and 2011 for Output Areas than other types of adminstrative geography. Of the 6,781 2001 MSOAs, 98% (6,640) are unchanged in the 2011 dataset. The remaining 141 zones are contiguous with the more numerous (n = 6791) 2011 zones (fig. x). Because many of 2001 MSOAs were split-up into small 2011 zones, the solution was not as simple as allocating the misfits to the nearest 2011 zones. Instead we took advantage of the unchanging geometry of the OA zones and calculated incomplete records by aggregating 2001 data from OA-zones in each incomplete 2011 zone (fig. x).

The explanatory variables used in the baseline model to test the hypotheses listed in section 1 were:

- *Avinc*, the mean average income for households in each area, downloaded at the MSOA level from the official data repository data.gov.uk.
- *Lpath*, the of high quality cycle paths within each area. This dataset was provided in a series of MySQL databases by CycleStreets.net and processes in R.
- *Bcrash*, a proxy on the safety record on cycling in each area defined as a severity-weighted count of the number of cycle-related road traffic incidents reported in the STATS19 database from the beginning of 2005 until the end of 2012.
- *Invst*, a crude proxy of investment in cycling in each zone, defined as a binary variable: 0 for areas outside the scope of the CDT project and 1 for areas that did receive funding.

In addition to these baseline input datasets, a number of other variables were tested to investigate the relationship between other variables and cycling to work. These are described as and when they appear in the results.

Method

The proportion of people cycling to cycling to work ($pCycle$) was calculated as the total number of bicycle commuters ($Cycle$) divided the number of people commuting by all modes ($Commute$) for each area (a). At the national level, this can be represented as the population-weighted sum of $pCycle$ for all areas:

$$PCycle = \sum_{a=1}^{n} \frac{Commute_a}{\sum_{a=1}^{n} Commute} \frac{Cycle_a}{Commute_a}$$

where $PCycle$ is the proportion cycling overall. This is expressed more concisely as the total number of cyclists divided by the total number of commuters ($PCycle = Cycle/Commute$). However, equation x is useful in demonstrating how a single national value can mask substantial regional variation and highlights the importance of a zone's total population: while the *average* value of $pCycle$ dropped from 3.3 to 3.1% across English LAs from 2001 to 2011, $PCycle$ remained constant. In other words cycling grew in the (urban) LAs with higher than average populations.

The categories of "unemployed" and "work from home" were deliberately excluded from the *Commut* count, to prevent changes in employment structure influencing the result: if a commuter belt shifted away from car driving towards 'teleworking' (working from home), for example, this method could provide an unfairly optimistic impression of the uptake of cycling.

The dependent variable, *change* in the proportion of people cycling to work, can be defined in two ways. First, the *absolute difference* in the percentage of people cycling to work ($\Delta pCycle$) was calculated by subtracting the 2011 results from the 2001 results. Second, *relative change* ($QpCycle$) was defined as proportion of people cycling in 2011 divided by the rate in 2001.

A linear regression model was used to test the impact of the explanatory variables on model fit, and to elucidate the direction of influence. Following the principle of parsimony in model design, a simple model based on the hypotheses presented in the introduction was developed first. Against this baseline the performance of different model runs was compared. To this end the following alterations were made:

- Changes to the number and type of explanatory variables used.
- Subsetting the observations used for the model (e.g. to exclude London)
- Altering the dependent variable itself to explore the relationship between absolute and relative increases in cycling.

Results

The rate of cycling between 2001 and 2011 census was found to have changed very little, being 3.1% in both cases. However, there was substantial variation between the zones in terms of change in cycling. There is a strong positive skew in the distribution the growth rate (fig. x): less than a quarter (74) of LAs saw the modal share of cycling rise, by an average of 30% whereas the majority of LAs (250) saw small declines in the proportion of people cycling.

In terms of absolute growth, the distribution is more symetrical, with the vast majority of zones (264 zones, 81%) seeing less than a 1% change either way

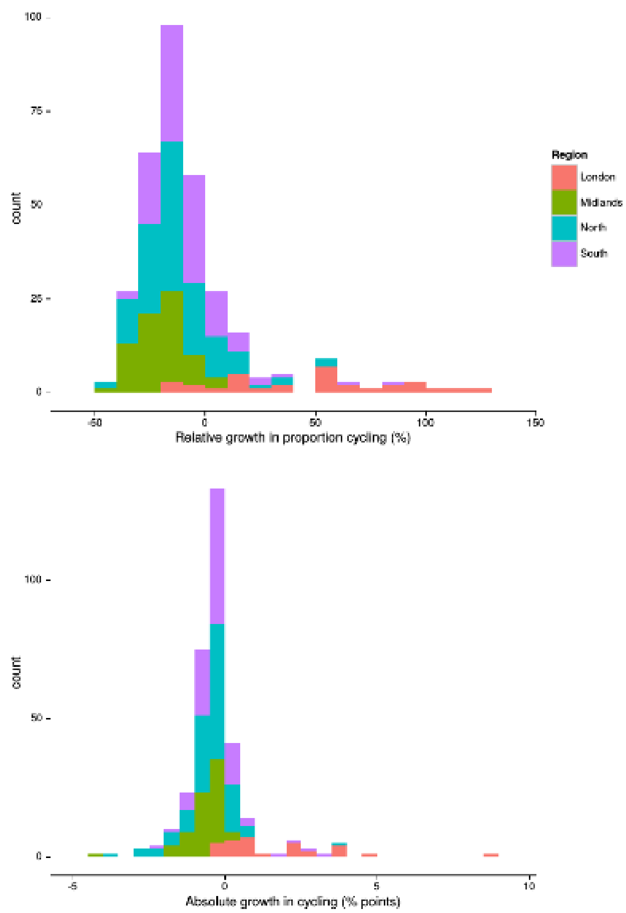

Figure 2: plot of chunk unnamed-chunk-5

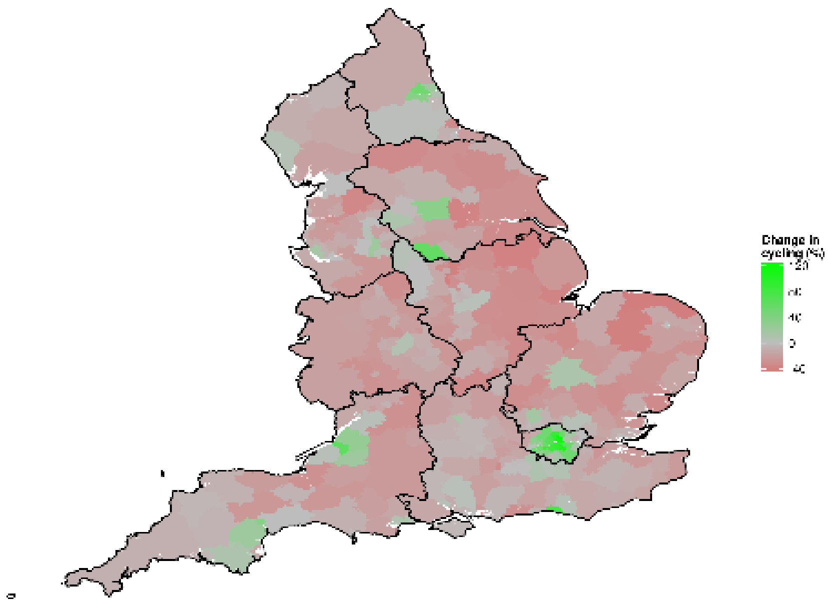

Figure 3: plot of chunk unnamed-chunk-6

in absolute proportion of people cycling to work. These results are plotted geographically in fig. x.

It is interesting to note that the average cycling rate in 2001 was lower for zones where cycling dropped (3.3 %) compared with zones where cycling grew (3.7%). Indeed, the variability in the proportional of people cycling grew between 2001 and 2011, despite the mean remaining the same: the standard deviation increased from 2.5 percentage points in 2001 to 2.7 in 2011. At the MSOA level the standard deviation of the percentage of people cycling to work also grew noticeably, from 2.7 to 2.9.

Far from cycling becoming more accessible to everyone everywhere, these results provide some geographical evidence for a divergence between the cycling 'haves' and 'have nots'. Spatial inequality in cycling as a mode of travel to work in recent years in England, supporting the hypothesis of positive feedback loops in modal shift to cycling through 'strength in numbers' and the normalisation of cycling culture (borjesson2012benefits, goetzke2011bicycle).

The regional distribution of growth in cycling, with its focus on London, is also evident from fig x, which shows the expected high correlation between the percent cycling to work in 2001 and 2011 (r-squared = 0.82). The $x = y$ line in fig x represents the 'break even' point above which cycling has grown and below which it has dropped.Thus there are 74 points above the line and the rest fall below. The further points are located from this break even line the more cycling has changed in that area. The colors illustrate that many of the zones with the greatest growth in cycling are located within the Greater London Authority.

The regional pattern represented by the colors in fig. x is emphasised in in table x, which shows that outside London there were falls in the proportion of people cycling to work, with the greatest declines in the Midlands and the North: the average LA in the Midlands saw the proportion of people cycling to work drop by a fifth.

Table x: regional differences in the change in cycling as a commuter mode in England, 2001 to 2011.

Region	Relative change	Absolute change
London	47.4	1.6
Midlands	-19.7	-0.7
North	-12.3	-0.5
South	-7.8	-0.2

Because of this apparent "London exceptionalism", the analysis was re-run with London removed. It was found that outside London, the proportion of all people cycling to work dropped substantially, from 3.2% in 2001 to 2.9% in 2011. *The average LA outside London saw a 12% drop in the proportion of people cycling.* These results suggest that attempts to 'get Britain cycling' have so far failed outside of the capital, which has seen the country's only large 'congestion charge' scheme encourage active travel (nakamura2014economic).

Some urban areas outside London broke the trend, seeing cycling as a mode of travel to work grow, yet only in 12 zones did the cycling to work increase by more than half a percentage point. (These were, in descending order of growth

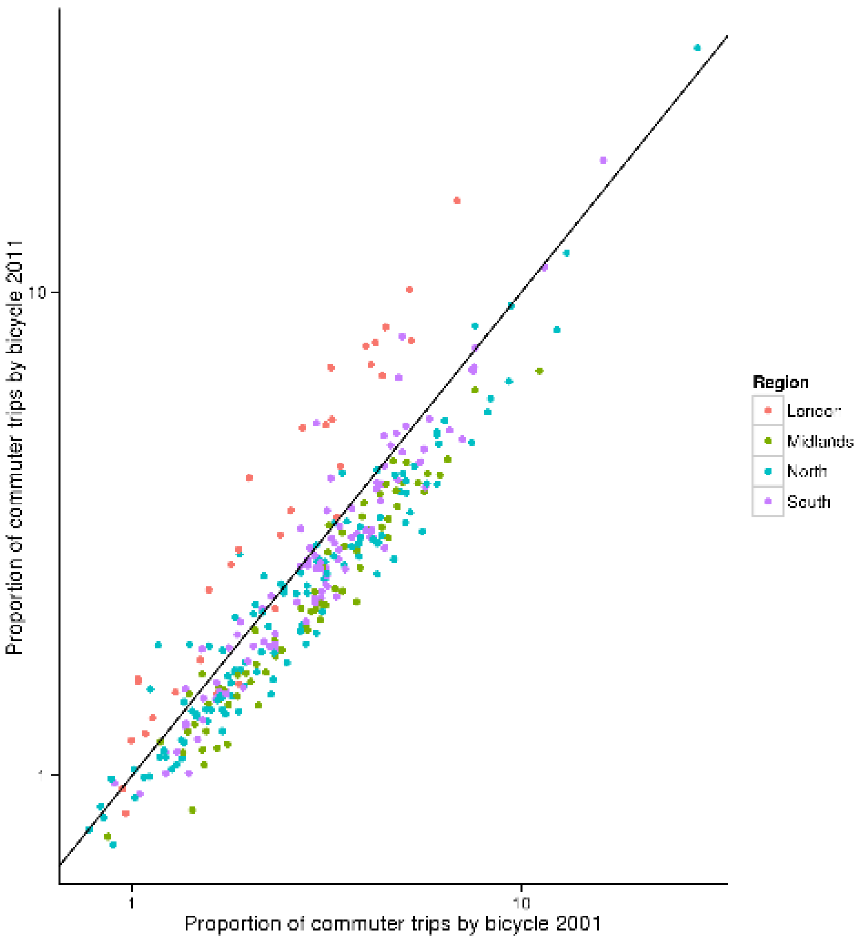

Figure 4: plot of chunk Scatterplot

rate, were Cambridge, Bristol, Oxford, Brighton, Exeter, Newcastle, South Cambridgeshire, South Gloucestershire, Manchester, Sheffield and Bournemouth.) Of these, only 5 had growth rates above 1 percent. In the name of balance, it was decided to focus equally on the areas where cycling has fallen greatly: there is a tendency towards picking 'best case' studies in the cycling literature. This can be seen as analogous to the disproportionate non-publication of medical trials that have negative results: it is important to focus on 'failures' as well as 'success stories' to provide impartial evidence to policy makers (Jones et al., 2013).

Table x presents the growth in cycling in the top 5 and bottom 5 areas in terms of absolute shift in the proportion of people cycling to work. It is noticeable that while 4 of the top 5 received central government funding between 2001 and 2011 from the CDT initiative, none of the bottom 5 did. (Oxford is conspicuously missing from the list of CDT beneficiaries, and some cycle campaigners have accused the local council of failing to properly maintain the city's existing cycle infrastructure (Horne, 2008).) This suggests that, in addition to investment helping to increase the rate of cycling, it can also serve to maintain it and prevent declines in areas that have an already high rate of cycling.

Table x: Statistics on cycle commuting from the top 5 and bottom 5 local authorities in England in terms of the absolute change in the proportion of commuters cycling to work. Note

Local Authority	pCycle 2001	pCycle 2011	Absolute growth	Rel. growth	CDT
Cambridge	28.3	31.9	3.6	12.7	2009
Bristol, City of	4.9	8.1	3.1	63.7	2009
Oxford	16.2	18.7	2.5	15.5	No
Brighton and Hove	3.0	5.3	2.4	80.1	2005
Exeter	4.8	6.6	1.8	37.3	2005
Boston	11.1	6.9	-4.3	-38.4	No
Kingston upon Hull, City of	12.3	8.3	-4.0	-32.2	No
Waveney	9.3	6.5	-2.7	-29.6	No
Fenland	7.4	4.9	-2.6	-34.4	No
North East Lincolnshire	8.2	5.6	-2.6	-31.2	No

Regression model results

The baseline model was able to explain 16% of the variation in $\Delta pCycle$ and 14% of the variability in $QpCycle$ across all zones (Table x). This is a statistically significant by not particularly powerful result. In addition, the most statistically signficant variable, $Bcrash$, had the opposite sign as expected, indicating a cycling may have become proportionally riskier in places where cycling has grown the most. Indeed, the correlation between $Bcrash$ and $\Delta pCycle$ was strongly positive (r = 0.31).

We can speculate why this may be - perhaps car drivers in growth areas were relatively unused to high cyclist volumes or perhaps accident rates are higher amongst new cyclists. Yet correlation does not prove causality and the number of cycle commuters is a poor indication of *exposure*, meaning that the results do not show with any certainty that cycling realy did tend to become less safe in areas where cycling grew.

Avinc and *Invst* had a statistically significant ($p > 5\%$) impact on the model in the direction expected, with average income in 2004/5 having a strong positive correlation with $\Delta pCycle$ ($r = 0.22$) and $QpCycle$ ($r = 0.21$). Despite there being only 17 CTD projects (their impact spread accross 19 LAs), the impact on the model was statistically significant: CDT funding increased the proportion of people cycling in an LA by on average 0.6 % points in the baseline model.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-51.76	7.828	-6.612	1.595e-10
Avinc	0.04622	0.0116	3.986	8.341e-05
Lpath	-1015	638.4	-1.59	0.1129
Bcrash	86.16	16.79	5.131	5.021e-07
Invst	10.81	5.905	1.83	0.06812

Additional variables

Removing London

As can be seen in figure xx, LAs within London had anomolously high growth rates, some of which can be explained by factors exclusive to the capital such as the congestion charge, slow traffic speeds and an influx of young, mobile workers.

Discussion

Conclusions