Journal of Transport & Health ■ (■■■■) ■■■–■■■

Contents lists available at ScienceDirect

Journal of Transport & Health

journal homepage: www.elsevier.com/locate/jth

Tracking global bicycle ownership patterns

Olufolajimi Oke [a,b,*], Kavi Bhalla [c], David C. Love [d,e], Sauleh Siddiqui [a,b,f]

[a] Department of Civil Engineering, The Johns Hopkins University, Baltimore, MD 21218, United States
[b] Systems Institute, The Johns Hopkins University, Baltimore, MD 21218, United States
[c] International Injury Research Unit, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, United States
[d] Center for a Livable Future, The Johns Hopkins Bloomberg School of Public Health, Baltimore MD 21205, United States
[e] Department of Environmental Health Sciences, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, United States
[f] Department of Applied Mathematics and Statistics, The Johns Hopkins University, Baltimore, MD 21218, United States

ARTICLE INFO

Article history:
Received 21 March 2015
Received in revised form
6 August 2015
Accepted 21 August 2015

Keywords:
Bicycle ownership
Sustainable transportation
Data mining

ABSTRACT

Over the past four decades, bicycle ownership has been documented in various countries but not globally analyzed. This paper presents an effort to fill this gap by tracking household bicycle possession. First, we gather survey data from 150 countries and extract percentage household bicycle ownership values. Performing cluster analysis, we determined four groups with the weighted mean percentage ownership ranging from 20% to 81%. Generally, bicycle ownership was highest in Northern Europe and lowest in West, Central and North Africa, and Central Asia. We determine worldwide household ownership patterns and demonstrate a basis for understanding the global impact of cycling as a sustainable transit mode. Furthermore, we find a lower-bound estimate of the number of bicycles available to the world's households. We establish that at the global level 42% of households own at least one bicycle, and thus there are at least 580 million bicycles in private household ownership. Our data are publicly available and amenable for future analyses.

1. Introduction

Bicycles have been an enduring and beloved mode of human-powered travel for over 125 years. Since the late 19th century, bicycles have been seen as a vehicle for social change, for example, empowering women's emancipation (Herlihy, 2006). Historically, significant fractions of populations around the world, particularly in Asia, have used cycling for transportation. Requests for paved roads came first from the cycling community, in some cases predating automobiles (Herlihy, 2006).

In the past century, however, both developed and developing countries have undergone rapid transitions towards motorization, which have disfavored bicycle use (Schäfer et al., 2009; Koglin and Rye, 2014). Simultaneously, there have been transitions in population health away from mostly infectious disease in children to non-communicable diseases (NCDs) and injuries that affect adults (Murray et al., 2013; Lozano et al., 2013). NCDs and injuries now comprise 94% of all deaths in China, 65% in India, and 34% in sub-Saharan Africa (Lozano et al., 2013). Increasing motorization leads to injuries from road traffic crashes, growing vehicular air pollution, and declining physical activity (Bhalla et al., 2014). As countries rapidly urbanize, people living in high-density populations are more vulnerable to vehicular air pollution. There is a growing need to address the rise of NCDs and injuries globally, and the public health community has begun engaging with transportation and urban planning professionals to search for solutions (Pratt et al., 2012; UN, 2011).

In the mid-1970s, Danish population and city planners embraced a novel experiment: a reversal of policies promoting motorization was made in favor of policies promoting cycling as a way to address traffic fatalities, looming energy crises, and environmental concerns (Fietsberaad, 2006). After decades of study, the research community has validated this approach and has shown that cycling creates a virtuous cycle with numerous positive feedback loops (Krizek, 2007). The movement of people and goods by bicycle reduces vehicular air pollution and motor vehicle traffic congestion (ESCAP, 2013). Cycling is a key element to "livable cities" (Geller, 2003), it connects easily to other modes of transit, and it can stimulate local businesses via the addition of new cycling routes (Litman, 2014). From a public health perspective, cycling promotes wellness (Oja et al., 2011), and the benefits of cycling outweigh the risks (de Hartog et al., 2010; Rojas-Rueda

* Corresponding author at: Department of Civil Engineering, The Johns Hopkins University, Baltimore, MD 21218, United States, Tel.: +1 413 329 6183.
E-mail addresses: ooke1@jhu.edu (O. Oke), kbhalla@jhsph.edu (K. Bhalla), dlove@jhsph.edu (D.C. Love), siddiqui@jhu.edu (S. Siddiqui).

http://dx.doi.org/10.1016/j.jth.2015.08.006
2214-1405/

Table 1
Survey sources for bicycle ownership data and the number of countries and country–years available from each.

Survey source	Acronym	Years	Countries	Country-years
Demographic and Health Surveys (DHS, 2014)	DHS	1990–2011	68	169
Enquête Démographique et de Santé et à Indicateurs Multiples (EDSM, 2013)	EDSM	2006	1	1
India National Census (IndiaStat, 2014)	INC	2001, 2011	1	2
Integrated Public Use Microdata Services (IPUMS, 2013)	IPUMS	1990–2006, 2008, 2009	21	26
Integrated Survey on the Welfare of the Population (IBEP, 2013)s	IBEP	2008–2009	1	1
International Crime Victim Surveys (ICVS, 2000)	ICVS	1989–2002	62	130
Malaria Indicator Surveys (MIS, 2014)	MIS	2006–2009	3	3
Multiple Indicator Cluster Surveys 4 (MICS4, 2013)	MICS4	2010–2011	17	17
Multiple Indicator Cluster Surveys 3 (MICS3, 2013)	MICS3	2005–2009	39	39
Study on Global Ageing and Adult Health (SAGE, 2012)	SAGE	2007–2011	5	12
World Health Surveys (WHS, 2013)	WHS	2002	65	65

et al., 2011). As more cyclists use roads, the safer the roads become, following the “safety in numbers” hypothesis (Jacobsen, 2013). On a global scale, cycling as a form of low-carbon transportation combats climate change (Woodcock, 2009; Bannister, 2011; Sheppard, 2011).

Globally, there is ample information about motor vehicles. Nearly every country tracks vehicle registration—some for tax purposes—and global data are gathered by the International Road Federation, the World Bank (via World Development Indicators), the World Health Organization (Global Status Reports on Road Safety), among other agencies. Bicycles, however, have never been systematically counted and presented in the peer-reviewed literature. This study aims to do that by compiling data on household bicycle availability and by grouping nations based on ownership levels.

2. Methods

2.1. Data collection

We obtained data on percentage bicycle ownership (PBO) from national and international surveys conducted at various times from 1971 to 2012 in 150 countries. However, we only consider for analysis the years 1989–2012, as only four countries have data available prior to 1989. Our sources for household bicycle ownership data include the World Health Surveys (WHS, 2013), Demographic and Health Surveys (DHS, 2014), Malaria Indicator Surveys (MIS, 2014), Integrated Public Use Microdata Services (IPUMS, 2013), International Crime Victim Surveys (ICVS, 2000), Multiple Indicator Cluster Surveys (MICS4, 2013; MICS3, 2013), and the India National Census (IndiaStat, 2014). We also had data available from the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ, 2009, 2012), but we did not include these in our analyses, as the respondents were schoolchildren and not representative of the national populations. Table 1 lists the surveys mined for our analyses, indicating the contribution of each source to the dataset. Fig. 1 shows the number of datapoints obtained from each country, showing the density and geographical distribution of the survey data obtained. (Supplementary Table S1 contains a list of all the countries for which data were collected.) Table 2 summarizes the objectives and sampling methodologies of the contributing surveys, most of which were global in scope. All were nationally representative, employing probability sampling of census enumeration areas (EAs) or otherwise determined zones, except for the INC, which involved an actual count. The household bicycle ownership questions vary little. Based on these, the survey data are fairly comparable.

We also collected household population numbers for the country–years in our dataset where available. Our sources primarily included IPUMS (2013) and the United Nations (UN, 2013; UNECE, 2014). In cases where direct household numbers were unavailable for certain country–years, we used a simple heuristic to find an approximation from nearest values or multiply average household sizes and corresponding national population (World Bank, 2015b) totals. (See Supplementary Material Section 1.2.)

2.2. Cluster analysis

The bicycle ownership data obtained[1] were sparse and time series varied considerably in length from one country to another. To find similarities in ownership across geographical regions, clustering presented itself as an effective pattern recognition tool (Jain et al., September 1999). Hierarchical or agglomerative clustering relies on a matrix of pairwise distances between vectors in the dataset, which were nontrivial to compute in our case due to their nonalignment. The number of clusters must also be specified.

First, we used the dynamic time warping algorithm (Sakoe and Chiba, 1978; Giorgino, 2009) to obtain distance alignments between countries. Using the goodness-of-fit test proposed by Mérigot et al. (2001), we then found the best agglomerative clustering method, which produced the minimum separation between the original distance matrix and that obtained from the tree. Of four possibilities, the unweighted pair-group method with arithmetic means (UPGMA) emerged as the best fit. The gap statistic and test (Tibshirani et al., 2001) determined the optimal number of clusters for the data. We performed ordinary least squares regression on the PBO in each cluster but found no significant temporal dependencies on bicycle ownership. Thus, we used the rolling mean with five-year windows to observe possible ownership trends. From household estimates for all the country–years, we were able to determine a lower bound on the number of available bicycles in the world. (See Appendix A for more details on the clustering steps described. The UPGMA dendrogram is shown in Supplementary Fig. S2.)

[1] The data and supporting code are available at http://www.hce.jhu.edu/sauleh/obls-gbu

Please cite this article as: Oke, O., et al., Tracking global bicycle ownership patterns. Journal of Transport & Health (2015), http://dx.doi.org/10.1016/j.jth.2015.08.006

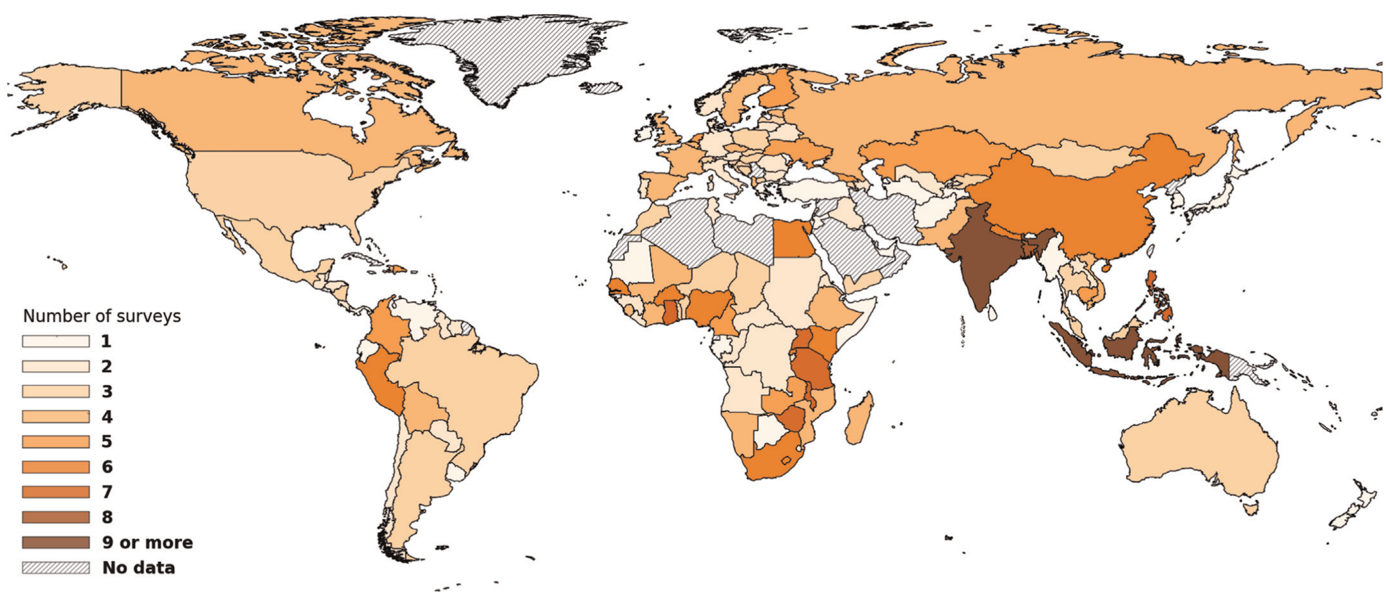

Fig. 1. World map indicating the number of surveys (datapoints) obtained in each of the 150 countries for which household bicycle ownership data were available. (Only one datapoint was available for South Sudan, and that is accounted for by Sudan's tally of 2 datapoints for the purposes of this map.)

3. Results

Household bicycle ownership rates were compiled for 150 countries from survey data. There were wide variations in bicycle ownership across different countries. In 2010, for example, Burkina Faso had a high household percentage bike ownership (PBO) of 84.2%, while Armenia had a low value of 4%. Even within the same region, there can be wide disparities. For instance, Ethiopia had a PBO of 2.3% in 2011, while Uganda (also in East Africa) had a PBO of 37.1%, about 12 times greater. Temporal variations at the country level can also be substantial. China, perhaps, exhibits the most dramatic variation here. In 1992, China had a PBO of 97.2%, indicating that there was at least one bicycle available in almost every household. This statistic had dropped (by nearly half) to 48.7% by the year 2007 but then rose to 63.2% (an increase of about 30%) in 2009. Supplementary Figure S1 shows bicycle ownership trends for each of the countries in this study.

In spite of this variation in bicycle ownership among the 150 countries, we were able to identify four distinct ownership levels using cluster analysis methods. Groups 1, 2, 3, and 4 had an average weighted PBO of 81%, 60%, 40%, and 20%, respectively. Group 1 comprises the countries with the highest PBO values (Fig. 2A). There are only 9 in this group: the Scandinavian countries, the Netherlands, Germany, Austria, Slovenia (all in Europe), and Burkina Faso in West Africa. There are 34 countries in Group 2 (Fig. 2B). They include, notably, USA, Canada, Brazil, Argentina, Uruguay, China, Australia, New Zealand and several European nations (e.g. France, Ireland, Italy, Luxembourg and Poland). Group 3 includes Russia and parts of Eastern Europe, the United Kingdom, four nations in the African Rift Valley system (Malawi, Tanzania, Uganda and Zambia) and five in West Africa (four of which are border states to Burkina Faso), the Indian subcontinent, Maritime Southeast Asia, Mexico, Chile and other South American countries, and Panama and Nicaragua in Central America (Fig. 2C). The lowest PBOs are to be found in most West, Central, and North African nations, as well as the Middle East and Central Asia. These make up Group 4 (Fig. 2D). There are 45 countries in Group 3 and 62 in Group 4.

The world map (Fig. 3) enables us to visually compare countries within their geographical region. We clearly see that Burkina Faso is an outlier in the entire African continent. Peru and the Philippines also stand out as the only Group 4 countries in South America and Southeast Asia, respectively.

From 1989 through 2012, the global household-weighted PBO averaged 42% (Fig. 4A). The median PBO was weighted by the number of households available for each year analyzed. India and China together account for over a third of the world's population (Bosworth and Collins, 2008), and on average, they make up close to a quarter of the household population analyzed in this study. We therefore plot the household-weighted median PBO for these two nations separately (Fig. 4B). The rest of the world accounts for an average PBO of 37% (Fig. 4C).

Finally, we determine a conservative lower bound on the number of bicycles currently available for use globally. An estimated 1.25 billion households (averaged over the years 1989–2012) make up the 150 countries we studied. These account for 80% of the estimated number of households in the world (UN-Habitat, 2007). Using the mean weighted PBO values for each cluster (multiplied by the corresponding household population of each), we estimate that there are at least 580 million bicycles currently in the possession of the world's households.

4. Discussion

In each of the clusters we discovered, the availability of bicycles by household has remained largely unchanged since 1990. Forty percent of the 150 countries considered occupy the group with the lowest PBO. Only 6% of the countries are in the top-ranked group. Household ownership is understandably low in places generally geographically inhospitable to bicycles, such as deserts and mountainous regions, i.e., Central Asia, the Sahara, and so forth. Interestingly, wealth is not always an indicator of bicycle availability. Notably, the United Kingdom with one of the world's highest per capita GDP (World Bank, 2015a) sits in Group 3, which collectively has a PBO of 40%.

Please cite this article as: Oke, O., et al., Tracking global bicycle ownership patterns. Journal of Transport & Health (2015), http://dx.doi.org/10.1016/j.jth.2015.08.006

Table 2
Comparison of survey methodologies. Data collection methods include paper and pen interviews (PAPI), computer-assisted person interviews (CAPI), computer-assisted field editing (CAFE) and face-to-face interviews (F2F). Responses choices include: “yes” (Y), “no” (N), “do not know” (DNK), “unknown” (U) and “not in universe” (NIU). For SAGE and WHS, 50+ and 18+ indicate that only household members at those respective ages and above were sought as respondents.

Survey	Purpose	Scope	Interview methods	Question	Response Choices	Question Type	Sampling
DHS	Monitor/evaluate population, health, nutrition indicators	Global	PAPI, CAPI, CAFE	Does any member of this household own: A bicycle?	Y/N	Household Characteristics	Clusters/census enumeration areas; respondents: women 15–49 years, men 15–59 years
ICVS	Crime and victimization analysis; perceptions of safety and security	EU, Global	CATI, F2F	bicycle ownership	Y/N/DNK	Screening	Random sampling and selection
INC	Census	India	F2F	Bicycle	N/A	Household asset availability	None
IPUMS	National censuses	Global	Various	Various	Y/N/U/NIU	N/A	Probability sampling
MICS3/4	MDG monitoring; wellbeing of women, children	Africa, Asia, South America	F2F	Does any member of your household own: A bicycle?	Y/N	Household Characteristics	Probability sampling of EAs
MIS	Track malaria intervention impact	At-risk malaria populations	F2F	Does any member of this household own: A bicycle?	Y/N	Household Characteristics	Probability sampling of EAs
SAGE	Older adult population health	Global	CAPI, CATI, F2F, PAPI	Does your household or anyone in your household have…? A bicycle?	Y/N	Permanent Income Indicators (Assets)	Used existing national framework; 50+ households targeted
WHS	Monitor critical national health outcomes	Global	F2F, PAPI	Does anyone in your household have: A bicycle?	Y/N	Permanent Income Indicators	Probability sampling by strata; 18+ private households

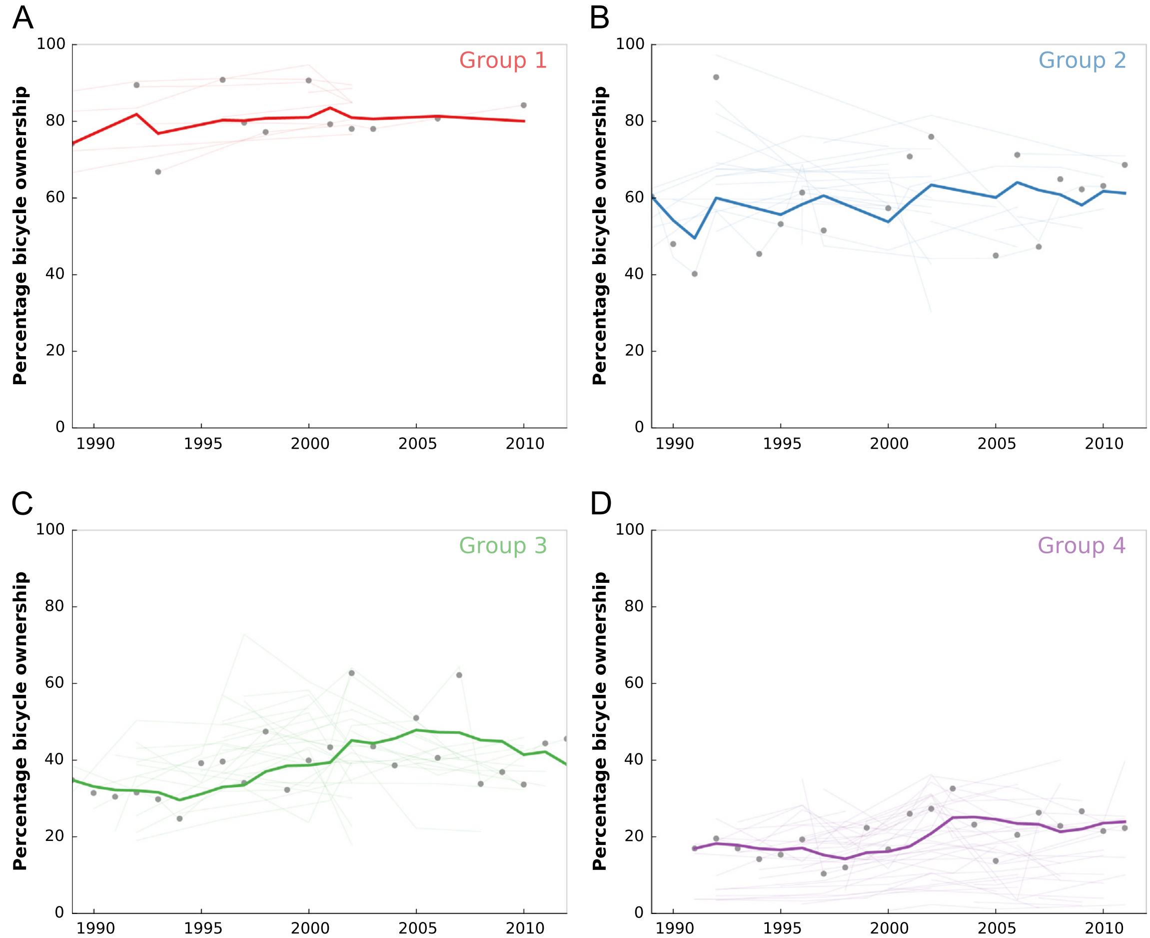

Fig. 2. Bicycle ownership trends (1989–2012) for the four groups determined by clustering analyses. In each plot, the thick line represents the rolling mean (in a five-year window) of the population-weighted annual median bicycle ownership (gray points; the thin lines are the individual time series for the member countries).

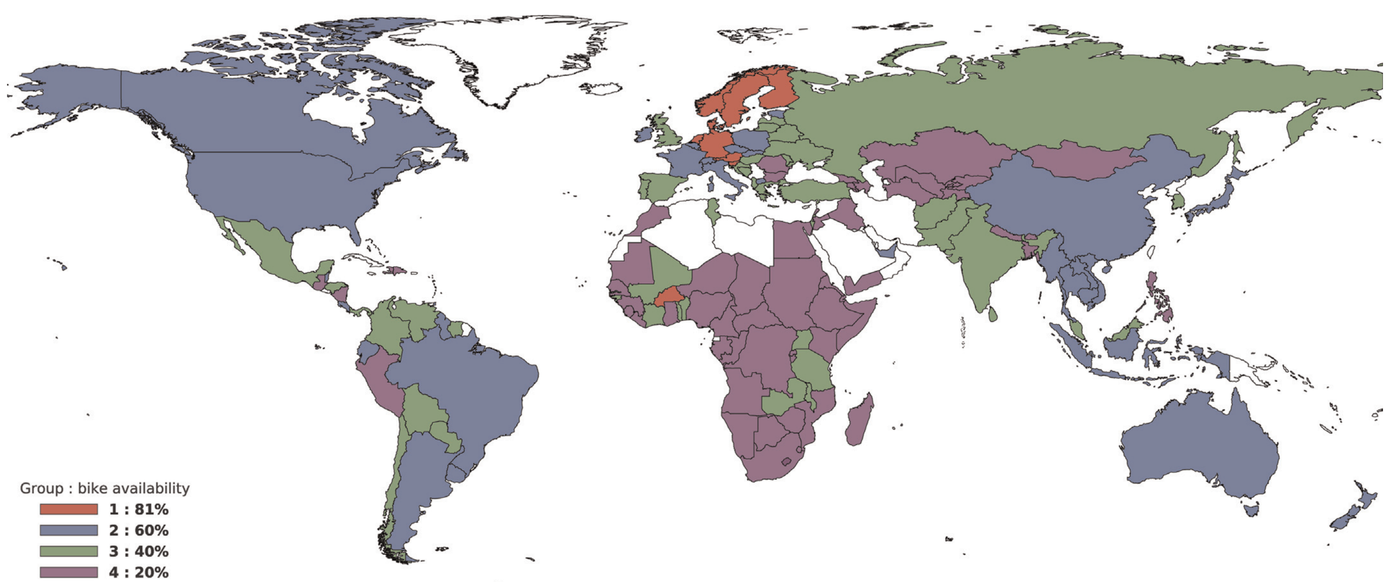

Fig. 3. World map showing countries color-coded by cluster. The weighted mean percentage household bicycle ownership is shown next to each group label. The red countries have the highest ownership numbers. Data were unavailable for the white portions of the map (notably in North Africa and the Middle East). South Sudan is not shown on the map, but it is also in Group 4, as is Sudan. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

Attitudes, safety and poor infrastructure may have contributed to the relatively low level of ownership in the UK (Wardlaw, 2014), compared to its neighbors (see Fig. 3).

Please cite this article as: Oke, O., et al., Tracking global bicycle ownership patterns. Journal of Transport & Health (2015), http://dx.doi.org/10.1016/j.jth.2015.08.006

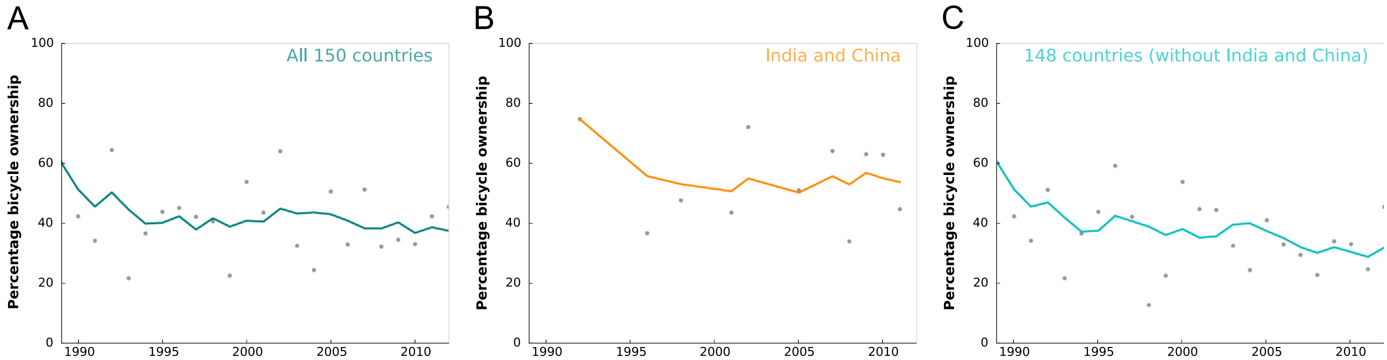

Fig. 4. Global trends in bicycle ownership from 1989 to 2012 for (A) all 150 countries analyzed, (B) India and China, and (C) all countries excluding India and China. The points in gray are median ownership levels weighted by the number of households surveyed in each year. The lines are the rolling means within a 5-year window.

Beyond ownership levels, each cluster shows characteristic behavior (Fig. 2). Group 1 has the most stable trend, and its member nations have had long-existing bicycle prioritization policies. Of note, 8 of the 9 Group 1 countries (those in Europe) are proximate. Group 2 also has a largely static trend. The mean number of households in this group is half a billion, which is about 40% of the households in the dataset. The Group 3 PBO trend has a trough of 30% in 1994 and a peak of 48% in 2006. Similarly, Group 4 pits at 14% in 1998 before rising to a maximum of 25% in 2004. Group 3, however, has a more pronounced post-peak decline (by 10 percentage points in 2011) compared to Group 4 (3 percentage points in 2011). The overall outlook in all four clusters is therefore flat or declining. We note, however, that no statistically significant trend can be observed for any of the ownership plots.

Worldwide, bicycle availability at the household level did not experience any significant increases or decreases (Fig. 4A), as it hovered around a weighted average of 42%. We see from Fig. 4B that China and India together also exhibit a flat trend in ownership, excepting the initial decline in the early 1990s. However, their mean weighted PBO of 54% is 12 percentage points greater than the global average. For the other 148 countries, the mean weighted PBO is 5 percentage points less than the global average. We note, however, that in these countries bicycle availability declined by a half during the period under consideration, from an average PBO of 60% in 1989 to 32% in 2012 (Fig. 4C).

Since India and China are the most influential nations (with regard to household population) in Groups 3 and 2, respectively, it may be instructive to observe their respective outcomes for bicycle ownership. We can also examine the success story of Burkina Faso in West Africa, which has an average PBO of 78%, over three times the unweighted regional average of 26%. Although one of the poorest nations in the world, Burkina Faso has invested substantially in cycling infrastructure (on a scale perhaps not seen in other African nations), and its positive attitudes toward cycling have been well documented (VOA, 2009). Cycling is also popular in Burkina Faso as a sport (for example, Tour du Faso since 1987) and as a tourist activity—further evidence for the widespread acceptance of bicycles in the country. However, the popularity of cycling does not always indicate household bicycle availability, as exemplified by Peru in Group 4. Cycling and mountain biking have been growing as a sport in Peru, and it remains a highly desired destination for bicycle tourists (Jenkins and Guides, 2009), due in part to its sights and scenic routes. Yet, only 20% of Peruvian households own bicycles, compared to the much higher South American average of 52%. Similarly, in Southeast Asia, where the mean ownership is 49%, the Philippines stands out with a low 23% mean ownership level. Congestion, poor roads and inadequate planning have all contributed toward making bicycling unpopular (Gozun and Guillen, 2008). However, like in Peru, mountain biking has gained popularity in recent years both among residents and visitors, and investments are now being made to improve cycling infrastructure in certain areas.

5. Conclusion

While our study was not able to assess bicycle usage, we have found that about four-tenths of households around the world have within arm's reach a powerful tool for low-carbon transportation and healthy physical activity. Governments, with the help of public health experts, health and transport geographers (Davison and Curl, 2014), and other stakeholders, can harness and mobilize cyclists as change agents by developing policies that support bicycle education, infrastructure, and a culture of safety for all road users. Socio-spatial factors, which have been shown to be of significance in characterizing bicycle usage patterns (Harms and Bertolini, 2014), must also be taken into consideration in understanding ownership trends. We do acknowledge, however, that ownership does not necessarily imply usage, and herein lies a limitation of the survey data we have currently compiled. Nevertheless, tracking bicycle ownership and usage should be a priority for countries and cities wanting to increase exercise and urban livability, thus reducing NCDs and the carbon footprint of their populations.

National health surveys are robust multi-year data sources that can have many unintended uses. Here, we have used them to develop global estimates of household bicycle ownership. Through the application of data alignment and clustering techniques, we identified four characteristic ownership groups and trends. It would be useful to investigate how broader ownership trend patterns might depend on variations in climate, development and economic prosperity across the clusters. Further collaboration between the public health and transportation fields on the analyses of nationally available datasets is possible and can be mutually beneficial for advancing priorities within both fields. For example, further research could be undertaken to identify the determining factors of bicycle ownership, motor vehicle ownership, and their interdependencies. These national surveys will also help us to identify countries or regions on new in-depth case studies of bicycle usage can be conducted.

Acknowledgements

This work was partially supported by the Energy, Environment, Sustainability and Health Institute (E^2SHI) at the Johns Hopkins University, Baltimore, Maryland. We would also like to thank the reviewers for their detailed assessments and helpful suggestions, all of which contributed toward making this manuscript a higher quality publication.

Appendix A. The clustering method

This is a three-step process further described below. It involves creating a matrix of distances between the time series, finding the best-fit agglomerative clustering procedure, and then computing the optimal cluster number.

A.1. Dynamic time warping alignment

The dynamic time warping (DTW) algorithm was first introduced by Bellman (1959). Sakoe and Chiba (1978) notably used it as a tool for aligning speech patterns for recognition. In this case, it provides a means of calculating the separation between our nonaligned points for each country, as the years do not all coincide. We use the package developed by Giorgino (2009) to execute the DTW algorithm in our program. A brief explanation is provided below.

Consider two time series A and B (a test and a reference) with P and Q observations, respectively. Elements a_i and b_j reside in series A and B, respectively. DTW computes a warping curve $\phi(k)$ with M elements, each mapped from A and B. Thus,

$$\phi(k) = (\phi_a(k), \phi_b(k)) \tag{A.1}$$

$$\phi_a(k) \in \{1, \ldots, P\} \tag{A.2}$$

$$\phi_b(k) \in \{1, \ldots, Q\} \tag{A.3}$$

The optimal deformation (alignment) minimizes the "average accumulated distortion" between the warped series. The deformation D is given as

$$D(A, B) = \min_{\phi} d_{\phi}(A, B) \tag{A.4}$$

where the distortion d_{ϕ} is

$$d_{\phi}(A, B) = \frac{1}{C_{\phi}} \sum_{k=1}^{M} d(\phi_a(k), \phi_b(k)) c_{\phi}(k), \tag{A.5}$$

with $c_{\phi}(k)$ being the weighting coefficient in each step and C_{ϕ} the normalization constant. The warping functions ϕ_a and ϕ_b are constrained for monotonicity, continuity and endpoint matching.

The output of DTW is robust, but we are only interested in obtaining the minimum cumulative distance for each possible pair of time series in our data. The result, which we call the dissimilarity matrix DM, is of size 150×150. However, since it is symmetric, there are only 11 325 unique entries, including a zero diagonal. We normalize the dissimilarity matrix D by dividing all its elements by the maximum value element.

A.2. Finding the clustering method of best fit

There are several well-defined hierarchical or agglomerative clustering methods, each using a different distance algorithm to determine how new clusters are added to the forest. We perform the clustering procedure using four methods, namely method of complete linkages, method of single linkages, unweighted pair group method with arithmetic means (UPGMA), weighted pair-group method with arithmetic means (WPGMA).

Using the goodness of fit measure defined by Mérigot et al. (2001) (who showed that their measure was superior to the established cophenetic correlation coefficient measure), we find that the UPGMA method gives the best fit clustering. The measure in question is the greatest singular value of difference between the original distance matrix D and the ultrametric matrix U (reordered distance matrix based on the clustering method).

A threshold for λ can also be estimated (Mérigot et al., 2001):

$$\lambda = \| D - U \|_2 \le \theta = 2\sigma\sqrt{N}, \tag{A.6}$$

where σ^2 is the sum of the variances of D and U. A λ value less than the threshold indicates that U is reasonably close to D. For all four

Table A1
Goodness-of-fit test for hierarchical/agglomerative clustering methods.

Method	λ	θ	λ/θ
Single linkage	2723.7	347.8	7.83
Complete linkage	4551.8	759.7	5.99
UPGMA	**883.2**	**423.1**	**2.09**
WPGMA	966.1	434.7	2.22

Table A2
Countries in the four groups determined by clustering with UPGMA.

ISO	Country
Group 1	
AUT	Austria
BFA	Burkina Faso
DNK	Denmark
FIN	Finland
DEU	Germany
NLD	Netherlands
NOR	Norway
SVN	Slovenia
SWE	Sweden
Group 2	
ARG	Argentina
AUS	Australia
BEL	Belgium
BLZ	Belize
BRA	Brazil
KHM	Cambodia
CAN	Canada
CHN	China
CRI	Costa Rica
CZE	Czech Republic
ECU	Ecuador
EST	Estonia
FRA	France
GUY	Guyana
IDN	Indonesia
IRL	Ireland
ITA	Italy
JPN	Japan
LAO	Laos
LUX	Luxembourg
MKD	Macedonia
MUS	Mauritius
MMR	Myanmar
NZL	New Zealand
POL	Poland
ZZZX	Serbia
SVK	Slovakia
CHE	Switzerland
THA	Thailand
TTO	Trinidad and Tobago
ARE	United Arab Emirates
USA	United States
URY	Uruguay
VNM	Vietnam
Group 3	
AFG	Afghanistan
ALB	Albania
BLR	Belarus
BEN	Benin
BOL	Bolivia
BIH	Bosnia and Herzegovina
CHL	Chile
COL	Colombia
CIV	Cote d'Ivoire
HRV	Croatia
GMB	Gambia
GRC	Greece
GNB	Guinea-Bissau
HND	Honduras
HUN	Hungary
IND	India
ISR	Israel
LVA	Latvia
LTU	Lithuania
MWI	Malawi
MYS	Malaysia
MDV	Maldives
MLI	Mali
MLT	Malta
MEX	Mexico
MNE	Montenegro
PAK	Pakistan
PAN	Panama

Please cite this article as: Oke, O., et al., Tracking global bicycle ownership patterns. Journal of Transport & Health (2015), http://dx.doi.org/10.1016/j.jth.2015.08.006

Table A2 (*continued*)

ISO	Country
PRY	Paraguay
PRT	Portugal
KOR	Republic of Korea
MDA	Republic of Moldova
RUS	Russia
ESP	Spain
LKA	Sri Lanka
SUR	Suriname
TZA	Tanzania
TGO	Togo
TUN	Tunisia
TUR	Turkey
UGA	Uganda
UKR	Ukraine
GBR	United Kingdom
VEN	Venezuela
ZMB	Zambia
Group 4	
AGO	Angola
ARM	Armenia
AZE	Azerbaijan
BGD	Bangladesh
BTN	Bhutan
BWA	Botswana
BGR	Bulgaria
BDI	Burundi
CMR	Cameroon
CAF	Central African Republic
TCD	Chad
COM	Comoros
COG	Congo
COD	Congo DRC
DJI	Djibouti
DOM	Dominican Republic
EGY	Egypt
ERI	Eritrea
ETH	Ethiopia
GAB	Gabon
GEO	Georgia
GHA	Ghana
GTM	Guatemala
GIN	Guinea
HTI	Haiti
IRQ	Iraq
JOR	Jordan
KAZ	Kazakhstan
KEN	Kenya
KGZ	Kyrgyzstan
LBN	Lebanon
LSO	Lesotho
LBR	Liberia
MDG	Madagascar
MRT	Mauritania
MNG	Mongolia
MAR	Morocco
MOZ	Mozambique
NAM	Namibia
NPL	Nepal
NIC	Nicaragua
NER	Niger
NGA	Nigeria
PER	Peru
PHL	Philippines
ROM	Romania
RWA	Rwanda
STP	Sao Tome and Principe
SEN	Senegal
SLE	Sierra Leone
SOM	Somalia
ZAF	South Africa
SSD	South Sudan
SDN	Sudan
SWZ	Swaziland
TJK	Tajikistan
TLS	Timor-Leste

Table A2 (*continued*)

ISO	Country
TKM	Turkmenistan
UZB	Uzbekistan
VUT	Vanuatu
YEM	Yemen
ZWE	Zimbabwe

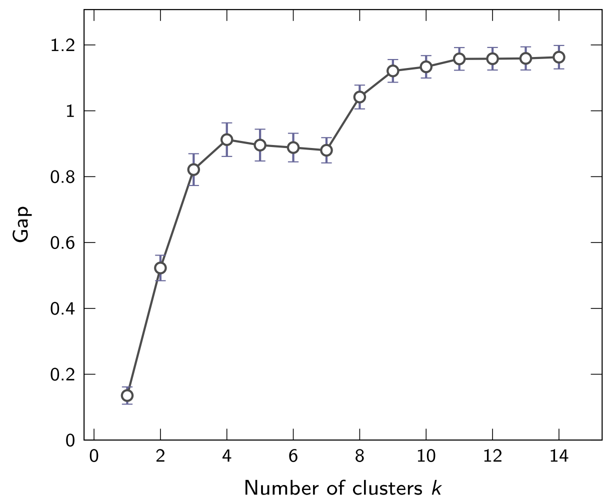

Fig. A1. Gap curve; optimal cluster number $\hat{k} = 4$, chosen as smallest local maximum.

methods, $\lambda > \theta$ (Table A1). This is not surprising considering the sparsity of our data. However, we observe that the method that also minimizes the λ/θ ratio is the best fit, which in this case is UPGMA.

Table A2 lists the countries in each of the four clusters found using the unweighted pair-group method.

A.3. The gap test

While there are rules of thumb (involving root functions) for finding the number of clusters in a dataset, various methods have been developed to find the optimal group number based on the structure of the dataset. We use a quantity referred to as the gap statistic (Tibshirani et al., 2001) to find the optimal number of clusters. The gap statistic is defined as

$$\mathrm{Gap}_n(k) = E_n^*\{\log(W_k)\} - \log(W_k), \tag{A.7}$$

where W_k is the within-cluster sum of pair-wise distances. The expected value $E_n^*\{\log(W_k)\}$ is determined by a Monte Carlo simulation of several samples of the dissimilarity matrix (obtained using a uniform distribution), from which $\log(W_k^*)$ is then calculated. If we let B be the number of Monte Carlo samples generated, then the gap statistic can be redefined as

$$\mathrm{Gap}_n(k) = \frac{1}{B}\sum_b \log(W_{kb}^*) - \log(W_k) \tag{A.8}$$

We define a simulation error term ε_k such that

$$\varepsilon_k = \mathrm{sd}_k \sqrt{\left(1 + \frac{1}{B}\right)}, \tag{A.9}$$

where sd_k is the standard deviation of the reference datasets:

$$\mathrm{sd}_k = \sqrt{\frac{1}{B}\sum_b \left\{\log(W_{kb}^*) - \frac{1}{B}\sum_b \log(W_{kb}^*)\right\}^2} \tag{A.10}$$

The optimal number of clusters $\hat{k}$ is chosen as the smallest k such that

$$\mathrm{Gap}(k) \geq \mathrm{Gap}(k+1) - \varepsilon_{k+1} \tag{A.11}$$

The gap statistic algorithm can be briefly described as follows:

(1) Using any given clustering method (in our case, the best-fit agglomerative method) and varying the number of clusters k from 1 through K, find within-cluster sum of squares W_k for the dissimilarity matrix.

(2) Assume a uniform distribution and produce B instances of the dissimilarity matrix. Again, using k-means clustering, clustering each of these datasets, in each case finding W_{kb} for $b = 1, \ldots, B$ and $k = 1, \ldots, K$.

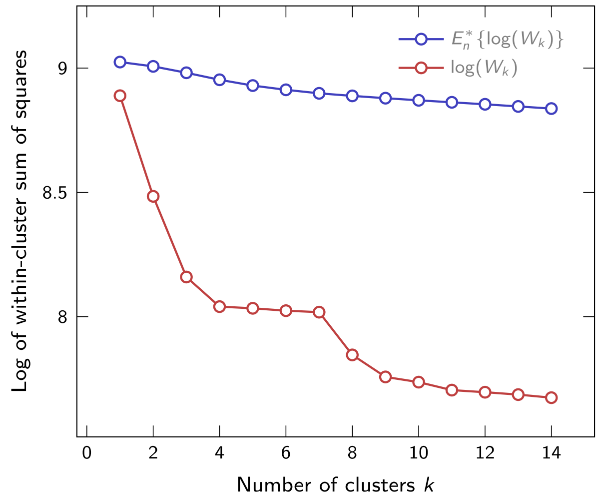

Fig. A2. The observed and expected values ($E_n^*[\,]$) of $\log(W_k)$, where W_k is the within-cluster sum of squares for k clusters. $\log(W_k)$ falls rapidly for $k<\hat{k}$ and less so for $k>\hat{k}$. Given the gap statistic Gap(k) and the test value $\text{Gap}(k+1)-\varepsilon_{k+1}$, we consider cluster numbers at which the gap statistic is greater than the test value. Then we choose the smallest value of k at which this happens. In this case, the k value of interest is 4.

(3) Compute the gap estimate:

$$\text{Gap}_n(k)=\frac{1}{B}\sum_b \log(W_{kb}^*)-\log(W_k) \tag{A.12}$$

(4) Find the simulation error term:

$$\varepsilon_k=\sqrt{\frac{1}{B}\sum_b\left\{\log(W_{kb}^*)-\frac{1}{B}\sum_b \log(W_{kb}^*)\right\}^2\left(1+\frac{1}{B}\right)}, \tag{A.13}$$

(5) Choose the number of clusters $\hat{k}$ such that

$$\hat{k}=\underset{k}{\operatorname{argmin}}\ \text{Gap}(k) \quad \text{subject to} \quad \text{Gap}(k)\geq \text{Gap}(k+1)-\varepsilon_{k+1} \tag{A.14}$$

We choose $B=1000$ for our dataset and obtain the gap curve shown in A1. An “elbow heuristic” suggests $\hat{k}=4$. But the final step in the algorithm formalizes this choice, as shown in Fig. A2.

Appendix B. Supplementary data

Further details on data collection and a table of the post-processed data (Supplementary Table S1) are provided in the Supplementary Material. The document also includes a dendrogram of the UPGMA clustering (Supplementary Fig. S2) and percentage bicycle ownership trends plots for each of the 150 countries analyzed (Supplementary Fig. S2).

Supplementary data associated with this paper can be found in the online version at http://dx.doi.org/10.1016/j.jth.2015.08.006.

References

Bannister, D., 2011. Cities, mobility and climate change. J. Transp. Geogr., 1538–1546.
Bellman, R., Kalaba, R., 1959. On adaptive control processes. IRE Trans. Autom. Control 4 (November (2)), 1–9.
Bhalla, K., Shotten, M., Cohen, A., Brauer, M., Shahraz, S., Burnett, R., Leach-Kemon, K., Freedman, G., Murray, C.J., Dingenen, R.V., Dentener, F., Vos, T., Naghavi, M., Abraham, J., Bartels, D., Weh, P.-H., 2014. Transport for Health: The Global Burden of Disease due to Injuries and Air Pollution from Motorized Road Transport. Report, Global Road Safety Facility, The World Bank; Institute for Health Metrics and Evaluation. IHME, Seattle, WA; The World Bank, Washington, DC.
Bosworth, B., Collins, S.M., 2008. Accounting for growth: comparing China and India. J. Econ. Perspect. 22 (1), 45–66.
Davison, L., Curl, A., 2014. A transport and health geography perspective on walking and cycling. J. Transp. Health 1 (4), 341–345.
de Hartog, J.J., Boogaard, H., Nijland, H., Hoek, G., 2010. Do the health benefits of cycling outweigh the risks? Environ. Health Perspect. 118 (8), 1109–1116.
DHS, 2014. The Demographic and Health Surveys Program, June ⟨http://www.statcompiler.com⟩ (accessed on May 27, 2013).
EDSM, 2013. Enquête Démographique et de Santé et à Indicateurs Multiples. The Demographic and Health Surveys Program, April ⟨http://www.childinfo.org/mics3_surveys.html⟩ (accessed on June 14, 2013).
ESCAP, 2013. Combatting Congestion. Transport and Communications Bulletin for Asia and the Pacific 82, Transport Division (TD) of the United Nations Economic and Social Commission for Asia and the Pacific (ESCAP), December.
Fietsberaad, 2006. Continuous and Integral: The Cycling Policies of Groningen and Other European Cycling Cities. Publication Number 7. Dutch Bicycle Council, Utrecht, The Netherlands.
Geller, A.L., 2003. Smart growth: a prescription for livable cities. Am. J. Public Health 93 (September (9)), 1410–1415.
Giorgino, T., 2009. Computing and visualizing dynamic time warping alignments in R: the dtw package. J. Stat. Softw. 31, 7.

Gozun, B.C., Guillen, M.D.V., 2008. Towards a sustainable transportation environment: The case of "pedicabs" and cycling in the Philippines. In: CODATU XIII. Challenges and Solutions for a Sustainable Development. CODATU, Ho Chi Minh City, Vietnam ⟨http://www.codatu.org/wp-content/uploads/Towards-a-sustainable-transportation-environment.-The-case-of-Pedicabs-and-cycling-in-the-Philippines-Brian-GOZUN-Marie-Danielle-GUILLEN.pdf⟩.
Harms, L., Bertolini, L., te Brömmelstroet, M., 2014. Spatial and social variations in cycling patterns in a mature cycling country exploring differences and trends. J. Transp. Health 1 (4), 232–242.
Herlihy, D.V., 2006. Bicycle: The History. Yale University Press, Taunton, MA.
IBEP, 2013. Inquérito Integrado Sobre o Bem-Estar da Populaç ao. UNICEF Child Info, April ⟨http://www.childinfo.org/mics3_surveys.html⟩ (accessed on June 14, 2013).
ICVS, 2000. International Crime Victimization Survey (ICVS) 1989–2000. ICVS International Working Group, Anna Alvazzi del Frate, Jan J.M. van Dijk, John van Kesteren, Pat Mayhew, Ugi Svekic, United Nations Interregional Crime and Justice Research Institute (UNICRI), Turin, Italy; United Nations Office of Drugs and Crime, Vienna, Austria ⟨http://www.unicri.it/services/library_documentation/publications/icvs/data/⟩.
IndiaStat, 2014. India National Census, June ⟨http://www.indiastat.com⟩ (accessed on June 10, 2013).
IPUMS, 2013. Integrated Public Use Microdata Series, International: Version 6.2 [Machine Readable Database] ⟨https://international.ipums.org/international/⟩ (accessed on June 13, 2013).
Jacobsen, P.L., 2013. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. Inj. Prev. 9 (3), 205–209.
Jain, A.K., Murty, M.N., Flynn, P.J., 1999. Data clustering: a review. ACM Comput. Surv 31 (September (3)), 264–323.
Jenkins, D., Guides, R., 2009. The Rough Guide to Peru. Rough Guides. Penguin, p. 48 ⟨https://books.google.com/books?id=BFZsW8seJ54C⟩.
Koglin, T., Rye, T., 2014. The marginalisation of bicycling in modernist urban transport planning. J. Transp. Health 1 (December (4)), 214–222.
Krizek, K.J., 2007. Estimating the economic benefits of bicycling and bicycle facilities: an interpretive review and proposed methods. In: Coto-Milln, P., Inglada, V. (Eds.), Essays on Transport Economics. Contributions to Economics. Physica-Verlag HD, pp. 219–248. http://dx.doi.org/10.1007/978-3-7908-1765-2_14.
Litman, T.A., 2014. Economic Value of Walkability. Paper, Victoria Transport Policy Institute, presented at the Transportation Research Board, 82nd Annual Meeting.
Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., Abraham, J., Adair, T., et al., 2013. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010. Lancet 380 (9859), 2095–2128 http://www.sciencedirect.com/science/article/pii/S0140673612617280.
Mérigot, B., Durbec, J.-P., Gaertner, J.-C., 2001. On goodness-of-fit measure for dendrogram-based analyses. Ecology 91 (6), 1850–1859.
MICS3, 2013. Multiple Indicator Cluster Surveys. UNICEF Child Info, April ⟨http://www.childinfo.org/mics3_surveys.html⟩(accessed on June 14).
MICS4, 2013. Multiple Indicator Cluster Surveys/MICS4. UNICEF Child Info, April ⟨http://www.childinfo.org/mics4_surveys.html⟩ (accessed on June 14, 2013).
MIS, 2014. Malaria Indicator Surveys. The Demographic and Health Surveys Program, June ⟨http://www.dhsprogram.com/What-We-Do/Survey-Types/MIS.cfm⟩ (accessed on May 27, 2013).
Murray, C.J.L., Vos, T., Lozano, R., Naghavi, M., Flaxman, A.D., et al., 2013. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the global burden of disease study 2010. Lancet 380 (9859), 2197–2223 http://www.sciencedirect.com/science/article/pii/S0140673612616894.
Oja, P., Titze, S., Bauman, A., de Geus, B., Krenn, P., Reger-Nash, B., Kohlberger, T., 2011. Health benefits of cycling: a systematic review. Scand. J. Med. Sci. Sports 21 (4), 496–509. http://dx.doi.org/10.1111/j.1600-0838.2011.01299.x.
Pratt, M., Sarmiento, O., Montes, F., Ogilvie, D., Marcus, B., Perez, L., Brownson, R., Lancet Physical Activity Series Working Group, Andersen, L., 2012. The implications of megatrends in information and communication technology and transportation for changes in global physical activity. Lancet 380 (9838), 282–293.
Rojas-Rueda, D., de Nazelle, A., Tainio, M., Nieuwenhuijsen, M.J., 2011. The health risks and benefits of cycling in urban environments compared with car use: health impact assessment study. Br. Med. J. 343.
SACMEQ, 2009. Data Archive for the SACMEQ I and SACMEQ II Projects (v.6.0). Compact Disc, Southern and Eastern Africa Consortium for Monitoring Educational Quality. Compiled by: Kenneth N. Ross, Mioko Saito, Stephanie Dolata, Miyako Ikeda and Linda Zuze, October.
SACMEQ, 2012. Data Archive for the SACMEQ III Project (v.6.0). Compact Disc, Southern and Eastern Africa Consortium for Monitoring Educational Quality. Compiled by: Demus Makuwa, Njora Hungi, Kenneth N. Ross, Mioko Saito, Stephanie Dolata, Frank van Capelle, Laura Paviot, Jocelyne Vellien, Jan Maarse.
SAGE, 2012. Study on global ageing and adult health. World Health Organization, February ⟨http://apps.who.int/healthinfo/systems/surveydata/index.php/catalog/sage⟩ (accessed on July 1, 2013).
Sakoe, H., Chiba, S., 1978. Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans. Acoust. Speech Signal Process. 26 (February (1)), 43–49.
Schäfer, A., Heywood, J.B., Jacoby, H.D., Waitz, I.A., 2009. Transportation in a Climate-Constrained World. MIT Press, April.
Sheppard, D.C., 2011. Social solutions for climate change mitigation and adaptation: cross cultural lessons from Denmark to the United States. Intersect 4 (1), 67–91.
Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. J. R. Stat. Soc. B 63 (2), 411–423.
UN, 2011. Political Declaration of the High-Level Meeting of the General Assembly on the Prevention and Control of Non-Communicable Diseases. United Nations Item 117, pp. 1–13.
UN, 2013. Demographic Yearbook. United Nations Statistics Division ⟨http://unstats.un.org/unsd/demographic/products/dyb/dyb_Household/dyb_household.htm⟩.
UN-Habitat, 2007. Enhancing Urban Safety and Security: Global Report on Human Settlements 2007. United Nations Human Settlements Programme, Earthscan, London; Sterling, VA.
UNECE, 2014. Private Households by Household Type. ⟨http://w3.unece.org/pxweb/dialog/varval.asp?ma=08_GEFHPrivHouse_r&path=../database/STAT/30-GE/02-Families_households/& lang=1&ti=Private+households+by+household+type⟩, United Nations Economic Commission for Europe Statistical Division Database.
VOA, 2009. Two Wheels a Way of Life in Burkina Faso. Voice of America ⟨http://www.voanews.com/content/a-13-2008-05-19-voa51/406615.html⟩.
Wardlaw, M.J., 2014. History, risk, infrastructure: perspectives on bicycling in the Netherlands and the UK. J. Transp. Health 1 (4), 243–250.
WHS, 2013. World Health Survey. World Health Organization, June ⟨http://apps.who.int/healthinfo/systems/surveydata/index.php/catalog/whs⟩ (accessed on May 27, 2013).
Woodcock, J., 2009. Public health benefits of strategies to reduce greenhouse-gas emissions: urban land transport. Lancet 374 (9705), 1930–1943.
World Bank, 2015a. GDP Per Capita, PPP (Current International $). ⟨http://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD⟩.
World Bank, 2015b. Population, Total. ⟨http://data.worldbank.org/indicator/SP.POP.TOTL⟩.

Please cite this article as: Oke, O., et al., Tracking global bicycle ownership patterns. Journal of Transport & Health (2015), http://dx.doi.org/10.1016/j.jth.2015.08.006