Working 9 to 5? Measuring hyperlocal worker productivity with public Wifi network data
DRAFT - NOT FOR CITATION WITHOUT PERMISSION OF THE AUTHORS

Nicholas E. Johnson[1], Bartosz Bonczak[1], Arpit Gupta[2], Constantine E. Kontokosta[1,3*]

1 Marron Institute for Urban Management, New York University, New York City, United States
2 Leonard N. Stern School of Business, New York University, New York City, United States
3 Marron Institute for Urban Management & Center for Urban Science + Progress, New York University, New York City, United States

* ckontokosta@nyu.edu

Abstract

The accurate estimation of workday length is essential to estimate total labor supply, and has a significant bearing on the assessment of labor productivity and worker well-being. Using probe request data from a 53 access-point, publicly-accessible Wi-Fi network in the Lower Manhattan district of New York City, we develop a method to measure localized worker activity patterns. Our Wi-Fi network data consist of over 10,000,000 probe requests per day, accounting for approximately 9.5 million unique devices over the study period from April 2017 to September 2017. We describe worker activity at various spatial and temporal aggregations in order to define baseline workday patterns and compute the workday length. We find a substantial population with characteristic workday lengths (e.g. 9am-5pm) during the workdays, as well as diurnal activity patterns that are consistent with expected worker behavior. These temporal patterns provide sufficient evidence to reinforce our assumptions about the ability to identify worker populations from Wi-Fi data. Finally, we compute the workday length for each identified worker and aggregate these workday lengths to estimate collective workday patterns to understand the uniformity of worker behavior. We find workday lengths of 7 hours and 40 minutes on average, which shorten substantially on Fridays and days surrounding holidays. We also find considerable seasonal variation in total workday hours supplied in our study area. This dynamic pattern of hours-worked suggests that our methodology is able to accurately assess workday lengths at high spatial resolution and temporal frequency. The ability to quantify hyperlocal worker activity patterns has a broad range of applications, including estimates of localized economic output and changes in labor supply.

Introduction

The accurate measurement of workday length is essential to estimate total labor supply, and has a significant bearing on the assessment of labor productivity and worker well-being [1]. Workday lengths are most commonly assessed through interviews and time-diary analysis, such as the American Time Use Survey. Similarly, day-time population density is often estimated using surveys (such as the Longitudinal Employer-Household Dynamics) or payroll data [2–4]. Analysis using such surveys has established considerable heterogeneity in typical workday lengths over time, across countries, and contract types (i.e., full-time, part-time, night shifts); and has been an integral component of a substantial body of work to understand the determinants of labor supply.

However, traditional survey methods have substantial and well-understood limitations. Survey responses are subject to considerable response bias, recall bias, and internal inconsistencies (i.e., total time budgeted to all activities does not add up to 24 hours) [5, 6]. The use of short time lags can mitigate some of these biases, at the expense of restricting time analysis to a narrow window around the interview time. The expense of conducting such interviews is also a substantial constraint to survey-based time use analysis.[1] These constraints limit the extent of time use analysis research based on current datasets.

In this paper, we propose a new method for workday measurement using real-time communication information that occurs between active Wi-Fi enabled devices (cell phones, laptops, tablets, etc.) and a proximate Wi-Fi network. Specific signals between these devices, known as probe requests, occur when a Wi-Fi enabled device is searching for an available network. This standard communication protocol provides the Wi-Fi network with information about the "pinging" device, which can be collected and analyzed for a broad range of applications, including presence detection, determining mobility trajectories, and identifying transportation mode [7–10]. For the purposes of understanding workday patterns, we use Wi-Fi probe request data to identify the frequency and duration of devices as they become visible to the network, using these aggregation measures to estimate workday lengths and labor supply. Relative to other methods of estimating workday length, this approach provides a greater representativeness of the local population, a more expansive time period for analysis, and more precise estimates of workday patterns.

Our objectives are to identify and quantify patterns of workday length using Wi-Fi probe request data and to demonstrate the potential of using these data to understand worker activity at high spatiotemporal resolution. Using probe request data from a 53 access-point, publicly-accessible Wi-Fi network in the Lower Manhattan district of New York City, we develop a method to identify worker activity by capturing first/last seen observations of individual devices in our study region and performing a time series analysis to filter out non-worker related activities. Our Wi-Fi network data consist of over 10,000,000 probe requests per day, accounting for approximately 9.5 million unique devices over the study period from April 2017 to September 2017. In the next stage of our analysis, we describe worker activity discrete and aggregate levels in order to define baseline workday patterns and compute the workday length. We find a substantial worker population with characteristic workday lengths (e.g. 9am-5pm) during the workdays, as well as diurnal activity patterns that are consistent with expected worker behavior. These temporal patterns provide sufficient evidence to reinforce our assumptions about the ability to identify worker populations from Wi-Fi data. Finally, we compute the workday length for each identified worker and aggregate individual workday lengths to estimate collective workday patterns and quantify heterogeneity in worker behavior.

This work fits within a growing body of work using communication technologies and smart devices to understand mobility patterns. This literature exploits the ubiquity of communication networks and their potential to generate data that describe population

[1]The American Time Use Survey, for instance, contacts roughly 1,100 randomly selected individuals (based on their participation in the Current Population Survey) each month.

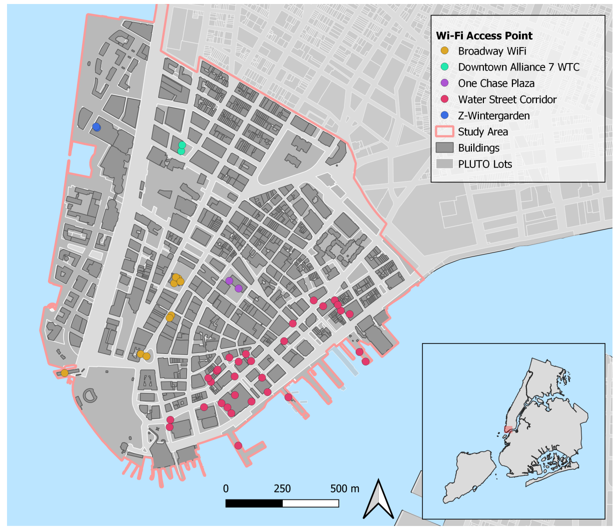

Fig 1. Locations of Wi-Fi access points in Lower Manhattan. Color differences indicate different sub-networks in the study area.

dynamics with unprecedented granularity [11, 12]. Wi-Fi probe request data, in particular, have been used to develop a real-time census and classify population types (i.e. worker, resident, visitor), as well as understand population mobility and trajectories throughout a dense urban area [4, 13, 14]. Similarly, estimating building occupancy and crowd size using Wi-Fi probe request are areas of interest for improving operations and safety [9, 15–17]. While important advances have been made in similar areas, there is a significant gap in the literature on using high resolution spatio-temporal data, such as Wi-Fi probe requests, as a proxy for worker activity and productivity.

In the next section, we describe our source data and processing pipeline, as well as our approach for computing workday patterns. It is followed by the description of our results and a discussion of our findings and potential applications. We conclude with ideas for future research.

Methods

The Wi-Fi probe request data used in this study was obtained from a public Wi-Fi network comprised of 53 access points (APs) located throughout Lower Manhattan [2]. Fig 1 shows a map of Lower Manhattan and the locations of individual APs. As a study area, Lower Manhattan is of considerable interest due to its intensity of business activity, concentration of a range of firm types, and high population density. On average, the network receives approximately 10,000,000 probe request per day. Each probe request observation includes the signal strength based on an RSSI value, the anonymized MAC address of the probing device, the MAC address of the AP which received the probe request, and a timestamp of when the probe request occurred. Data were obtained from April 2017 through Sept 2017, which contained 9.5 million unique device observations. Storage, access, and use of these data are governed by NYU CUSP's data use protocols and a data management plan approved by NYU's Institutional Review Board.

[2] Data were provided by the Alliance for Downtown New York, the organization that owns and maintains the Wi-Fi network

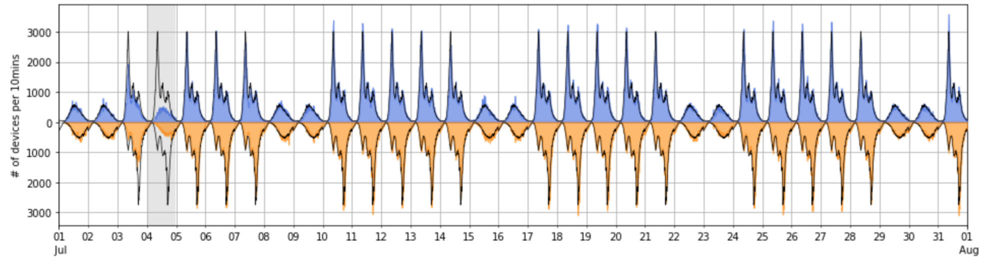

Fig 2. Number of first/last seen events per 10 minute intervals in Lower Manhattan throughout the month of July. The blue area indicates first seen counts while the orange indicates last seen counts. The black lines indicate the average count for the entire study period and the shaded area denotes the July 4th holiday.

Data Processing

The first stage in processing the Wi-Fi dataset is to identify specific observations that represent worker activity. A time series analysis is performed for each device to identify activity patterns based on presence in the study area. We identify the first and last observations ($o_{\text{first/last}}$) of a device from the Wi-Fi dataset for each day and built a binary vector that describes the hourly presence of each device throughout the study period. A device receives a value of 1 if present and a value of 0 if not present for a given hour. The bounding of $o_{\text{first/last}}$ observations to each day is required in order to identify recurring patterns of presence, as well as provide a duration metric from which a viable workday length can be calculated. A Fourier analysis of each discrete time series is used to identify devices with strong weekly patterns, which are likely to represent patterns of worker activity. Devices with non-weekly activity patterns represent activity that may be attributed to populations other than workers in Lower Manhattan (i.e. resident or tourist activity) and were removed.

After filtering the original dataset, we calculate workday length for each worker based on the difference between the first seen observation (o_{first}) and the last seen observation (o_{last}) for each day the device is present in the dataset. The daily median workday length is then computed for each day as the median duration for all workers present during the day.

Results and Discussion

After processing and filtering the original Wi-Fi dataset, the final dataset contains 255,122 worker devices for which we estimate collective worker activity patterns.

Aggregate Patterns of Worker Activity

Aggregate counts of o_{first} (blue) and o_{last} (orange) observations for one month are shown in Fig 2. These observations are aggregated over 10 minute intervals and capture the collective pattern of worker activity in Lower Manhattan. The average number of $o_{\text{first/last}}$ observations are indicated by black lines and the shaded area indicates a major holiday. During the weekday, o_{first} observations reach an average peak of 3,000 observations (per 10 minutes) in the morning at 9 AM and o_{last} observations reach an average of 2,700 observations in the evening at 6 PM. On weekends, however, $o_{\text{first/last}}$ observations reach a maximum of 500 observations at midday with no distinct peaks during the day.

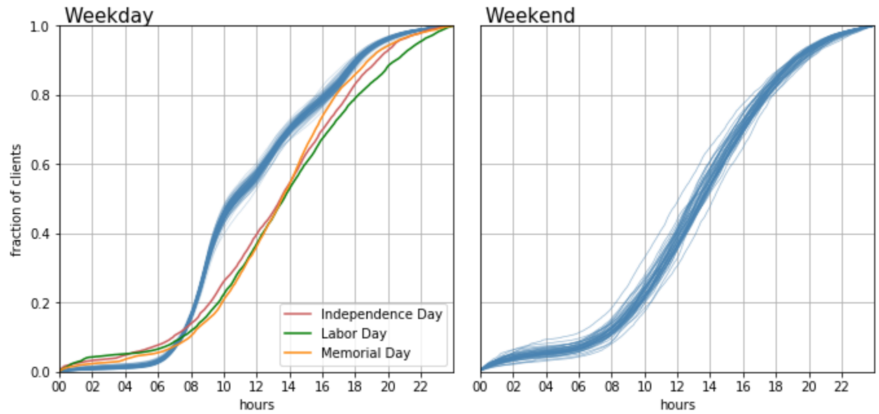

Fig 3. Baseline activity patterns for weekday and weekend in Lower Manhattan based on $o_{\mathbf{first}}$ observations. The blue lines indicate the fraction of clients identified in the network for each day of the study period and differentiated by weekday and weekend. Holidays that occurred during the study period are also noted with the red, green and yellow lines.

A unique diurnal pattern can be identified from aggregate $o_{\text{first/last}}$ observations. In particular, discernible weekday peaks at 9 AM for o_{first} observations and 6 PM for o_{last} observations are consistent with morning and evening rush hour periods and occur at regular daily and weekly frequencies. During weekdays, the only notable exceptions occur on holidays, as shown for July 4th in Fig 2.

The stability of this diurnal activity pattern is shown in Fig 3. The cumulative fraction of o_{first} observations, computed across 10 minute windows, show remarkable day-to-day consistency. Fully 98% of weekday observations are within one standard deviation from the mean and the average time difference from the mean is 20 minutes, reaching a maximum of 30 minutes during periods of peak variability. More concretely, the daily fraction of people arriving in Lower Manhattan varies by 20 minutes, on average, for any given weekday. Fig 3 also captures the activity differences between weekdays, weekends, and holidays. During weekdays, over half of the o_{first} observations occur between 6-10 AM, while weekends and holidays demonstrate a more gradual population increase throughout the day, reflecting the differences in workday patterns for those that work on holidays and/or weekends. Collectively, these observations reveal a consistent pattern of aggregate worker activity in Lower Manhattan.

Patterns of Discrete Worker Activity

While daily worker activity demonstrates remarkable consistency in the aggregate, a second analysis was performed to understand activity for discrete devices. The aim of this process is to identify a 'typical' arrival time for each worker and assess how individual devices deviate from their typical arrival time. The typical arrival time for a device is defined as its median arrival time over the course of the study period. Variability in arrival time is assessed by computing the average deviation (v_j) of o_{first} observations from an individual's mean and median arrival time. The average arrival variability can be formulated as:

$$v_j = \frac{1}{n} \sum_{i=1}^{n} \tilde{x}_j - o_i, \tag{1}$$

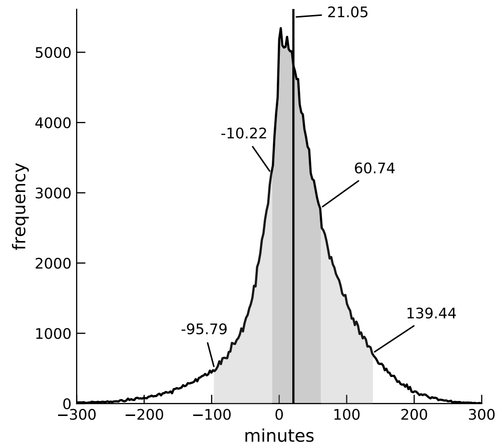

Fig 4. Distribution of deviation from median arrival time. The dark shaded area is the interquartile range and the light grey area is the 95% range. The black vertical line is the median deviation.

where o_i are device o_{first} observations and $\tilde{x}$ is the median arrival time for each device j.

The distribution of average arrival variability for all workers is shown in Fig 4. The average deviation is 21.05 minutes with 50% of observations occurring between -10.22 and 60.74 minutes. Given that a positive deviation indicates a worker arriving prior to their median time, on average, and a negative deviation indicates an arrival after their median arrival time, the calculated range indicates consistent arrival patterns with workers arriving up to an hour prior to their median arrival time on any given day, and a smaller portion of the sample arriving up to 10 minutes later than their median arrival time.

It is also important to note that the tails of this distribution may appear unreasonable for a worker's arrival schedule. Specifically, it seems unlikely that an individual is, on average, arrives one hour later (or earlier) than their typical arrival time. Instead, however, this activity pattern may not represent a worker whose daily work routine is from 9-5, but rather workers who have irregular work hours and work days. For example, it can be assumed a sample of the worker population visit Lower Manhattan only occasionally, as they may regularly work in an office that is not located in Lower Manhattan, travel extensively for work, or have a more flexible work schedule. In such cases, the individual may meet the criteria to be classified as a worker, while also demonstrating large variations in $o_{\text{first/last}}$ observations. Fig 4 shows that 50% of observations fall within a one-hour time window and 95% fall approximately within a two-hour time window, suggesting the presence of these types of workers is minimal in the sample.

Workday Length in Lower Manhattan

Based on the difference between the $o_{\text{first/last}}$ observations for a specific device on a given day, we compute the median duration for all devices present during a single day. The median workday length throughout the study period is shown in Fig 5 and includes the total worker count as defined by our filter method. The median workday length on weekdays is found to be 7 hours and 40 minutes, which remains consistent throughout the study period with exceptions during federal holidays as indicated by the vertical red lines in Fig 5.

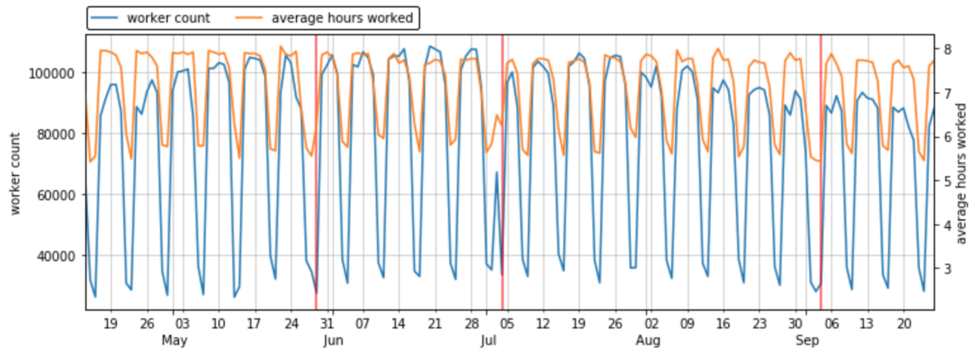

Fig 5. Average workday length (orange) for users in Lower Manhattan and the number of workers (blue). Red vertical lines indicate holidays.

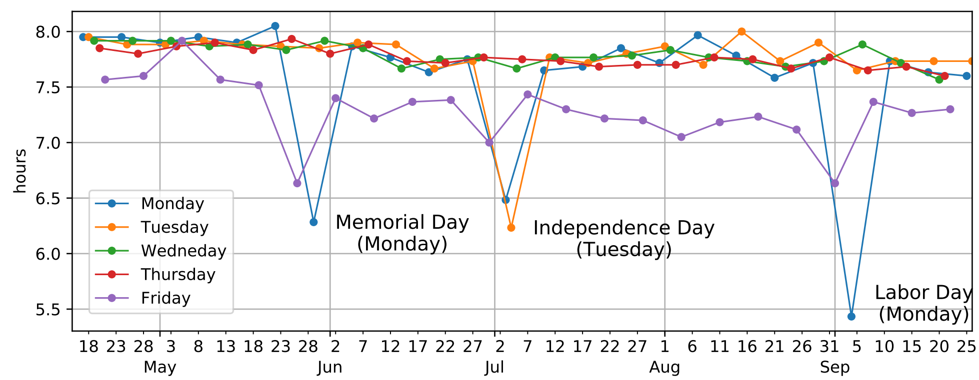

Fig 6. Time series plot of the median workday length for Lower Manhattan differentiated by day of week. Three federal holidays are also labeled for reference.

An important observation in Figure 5 is the relationship between median workday length and the total worker population. The weekday worker population in Lower Manhattan peaks in June with an average population of 110,000 unique devices per day and steadily declines to 90,000 per day in September. However, while a decline in population observed from June to September is likely a result of regular seasonality caused by workers leaving for vacation, the median workday length remains steady. Only 9% of the variance in median workday length during weekdays can be explained by changes in the population, when excluding federal holidays and weekdays that fall directly before a federal holiday.

Workday lengths also vary by day-of-week as the time series shows in Fig 6. As one might expect, the median workday length on Friday is 24 minutes shorter (5.24%) than the median workday length observed Monday through Thursday. Shorter workday lengths are also observed on days preceding federal holidays, which are 59 minutes shorter (12.9%) than Monday through Thursday observations.

Discussion

In this research, we present a new approach for measuring worker activity and quantifying workday length. Unlike traditional survey approaches, using Wi-Fi information provides a robust and comprehensive data source to analyze localized worker activity. We measure worker behavior by extracting distinct patterns in the Wi-Fi probe request data that indicate worker presence, estimating workday lengths, and measuring the regularity of observed patterns over time. Following the classification of workers based on device activity, we are then able to quantify arrivals and departures from the study area, which are further extrapolated to understand the duration of worker presence and the collective workday length in Lower Manhattan.

Indeed, the results from our approach to estimate workday length validate many of the assumptions about worker activity in an urban commercial district. However, our methodology provides new insight into specific characteristics of worker behavior. First, our assessment of aggregate activity patterns shows notable regularity of worker activity in Lower Manhattan. While arrival times may vary day-to-day, the overall fraction of workers arriving in Lower Manhattan at certain times demonstrates a distinct pattern with little variability. Collective $o_{\text{first/last}}$ observations occur during rush hour periods with typical arrival and departure times that correspond to traditional workdays of 9am to 5pm.

In addition to aggregate patterns of activity, we also demonstrate how variability in worker arrival times. By evaluating the difference between typical arrival time and daily variation in o_{first} observation, we show that over 50% of workers on any given day arrive prior to their regular (median) arrival time, suggesting not only are these workers visiting Lower Manhattan regularly, but also highlights consistency in arrivals before a specific time each day. This reinforces the expectation that workers will tend to arrive before the typical start of the workday in order to avoid the negative consequences of being ”late”.

We also demonstrate a new method for quantifying patterns of workday length. Based on the difference between $o_{\text{first/last}}$ observations, the median workday length is determined to be 7 hours and 40 minutes, which is consistent with the traditional eight-hour workday. However, the use of Wi-Fi data to extrapolate this information provides a more robust measurement of workday length, as well as additional insight about how the length of the workday varies over daily, weekly, and seasonal cycles.

Finally, our findings suggest that worker productivity (assuming a constant worker output per hour) fluctuates throughout the work week. While we observe relatively stable workday lengths from Monday through Thursday, there is a measurable decline on Fridays, especially when preceding a holiday. Although we do not explicitly attempt to quantify productivity, our approach can be used to estimate hyperlocal variations that can associated with other worker characteristics to estimate worker output.

Limitations

While this work demonstrates a novel approach to estimating workday length, there are several important limitations that should be considered. One of the main limitations relates to data quality and comprehensiveness. As discussed throughout this work, there are limitations to understanding causes of variations in individual $o_{\text{first/last}}$ observations, which may result directly from device activity or indirectly from the overall data capture process. It is entirely possible that day-to-day variations in an individual's commute pattern and their activity throughout the day may cause a particular device to be visible to the network at times that do not directly represent their workday patterns. While we have tried to reduce this uncertainty as much as possible with our filtering methods, as municipal Wi-Fi networks become more ubiquitous, covering large parts of a city, the issue of network boundaries diminishes.

A similar limitation relates to the representativeness and overall capture rate of the Wi-Fi network. It is likely that a sample population may have more than one Wi-Fi enabled device or no device at all. In addition, devices that have turned off their Wi-Fi transmitter will not be visible to the network. While this does limit our confidence that the Wi-Fi data includes all potential workers throughout the study period, for the purposes of this research, a 100% capture rate is not required to estimate the median workday length.

Furthermore, recent concerns over privacy have led to mobile device manufactures to implement MAC address anonymization, which causes a device to change its MAC address over time preventing identification of the same device across a study period. The data obtained were pre-filtered by the Meraki platform to exclude randomized MAC addresses.

A final limitation in this research is the duration of accessible data. Specifically, there may be seasonality that we do not capture given the time period covered in the dataset. Though it is not clear that this would have any significant implications for our approach to calculating workday length, it is a consideration for future work and should be more thoroughly assessed given data availability.

Conclusion

We have demonstrated a fundamentally new method for quantifying worker activity and workday length using data from a publicly-accessible Wi-Fi network. Our approach is able to produce both individual-level and aggregate estimates about labor supply using the proxy of hours worked. We find workday lengths of 7 hours and 40 minutes on average, which shorten substantially on Fridays and days surrounding holidays. We also aggregate our estimates and find considerable seasonal variation in total workday hours supplied in our study area. This dynamic pattern of hours worked suggests that our methodology is able to accurately assess workday lengths at high spatial resolution and temporal frequency.

The ability to estimate hyperlocal worker behavior has a broad range of applications in numerous fields. From a transportation perspective, information on actual workday patterns and associated commute modalities can assist in more robust planning and timely service provision. Furthermore, understanding the variability in worker arrival times to a specific area unlock new opportunities to evaluate the impact of staggered work shifts on congestion and overcrowding. The analysis of economic statistics can also be assisted through more precise estimates of localized economic output. In particular, our estimates may be valuable in tracking high-frequency changes in labor supplied. Though our study period features fairly constant economic output, sharp changes in labor supply may be potentially detected more quickly using sensor data than through conventional economic statistic collection methods. These data may, therefore, assist in helping to "nowcast" layoffs and regional shocks, providing new tools to forecast and respond to shifts in economic activity.

Supporting information

S2 Fig. 1 **Locations of Wi-Fi access points in Lower Manhattan.** Color differences indicate different sub-networks in the study area.

S3 Fig. 2 **Number of first/last seen events per 10 minute intervals in Lower Manhattan throughout the month of July.** The blue area indicates first seen counts while the orange indicates last seen counts. The black lines indicate the average count for the entire period and the shaded area indicates the July 4th holiday.

S3 Fig. 3 **Baseline activity patterns for weekday and weekend in Lower Manhattan based on $o_{\mathbf{first}}$ observations.** The blue lines indicate the fraction of clients identified in the network for each day of the study period and differentiated by weekday and weekend. Holidays that occurred during the study period are also noted with the red, green and yellow lines.

S3 Fig. 4 **Distribution of individuals' average deviation from their median arrival time.** The dark shaded area is the 50% quartile range and the light grey area is the 95% range. The black vertical line is the median deviation.

S3 Fig. 5 **Average workday length (orange) for users in Lower Manhattan and the number of workers (blue).** Red vertical lines indicate holidays.

S3 Fig. 6 **Time series plot of the median workday length for Lower Manhattan differentiated by day of week.** Three federal holidays are also labeled for reference.

Acknowledgments

We would like to thank the Alliance for Downtown New York for access to the data used in this study. This work was funded, in part, by a grant from NYU C2SMART, a USDOT Tier 1 Transportation Research Center. All errors remain our own.

References

1. Costa DL. The Wage and the Length of the Work Day: From the 1890s to 1991. Journal of Labor Economics. 2000;18(1):156–181.

2. Boeing G. Estimating local daytime population density from census and payroll data. Regional Studies, Regional Science. 2018;5(1):179–182.

3. Schein A, Haruvi N. The Relationship between the Average Workweek Length and Per Capita Gross Domestic Product (GDP). International Studies of Management & Organization. 2017;47(4):311–323.

4. Kontokosta CE, Johnson N. Urban phenology: Toward a real-time census of the city using Wi-Fi data. Computers, Environment and Urban Systems. 2017;64:144–153.

5. Abraham KG, Maitland A, Bianchi SM. Nonresponse in the American time use Survey: Who is missing from the data and how much does it matter? International Journal of Public Opinion Quarterly. 2006;70(5):676–703.

6. Hamermesh DS, Frazis H, Stewart J. Data watch: The American time use survey. Journal of Economic Perspectives. 2005;19(1):221–232.

7. Jiang S, Ferreira J, González MC. Clustering daily patterns of human activities in the city. Data Mining and Knowledge Discovery. 2012;25(3):478–510.

8. Schneider CM, Belik V, Couronné T, Smoreda Z, González MC. Unravelling daily human mobility motifs. Journal of The Royal Society Interface. 2013;10(84):20130246.

9. Gonzalez MC, Hidalgo CA, Barabasi AL. Understanding individual human mobility patterns. nature. 2008;453(7196):779.

10. Sapiezynski P, Stopczynski A, Gatej R, Lehmann S. Tracking human mobility using wifi signals. PloS one. 2015;10(7):e0130824.

11. Calabrese F, Ferrari L, Blondel VD. Urban Sensing using Mobile Phone Network Data: a Survey of Research. Acm computing surveys (csur). 2015;47(2):25.

12. Pappalardo L, Simini F, Rinzivillo S, Pedreschi D, Giannotti F, Barabási AL. Returners and Explorers Dichotomy in Human Mobility. Nature communications. 2015;6.

13. Traunmueller MW, Johnson N, Malik A, Kontokosta CE. Digital footprints: Using WiFi probe and locational data to analyze human mobility trajectories in cities. Computers, Environment and Urban Systems. 2018;72:4–12.

14. Oliveira L, Schneider D, De Souza J, Shen W. Mobile Device Detection Through WiFi Probe Request Analysis. IEEE Access. 2019;7:98579–98588.

15. Li X, Liu X, Qian Z. Towards an Occupancy-Enhanced Building HVAC Control Strategy Using Wi-Fi Probe Request Information. In: Computing in Civil Engineering 2017; 2017. p. 17–24.

16. Hong H, De Silva GD, Chan MC. Crowdprobe: Non-invasive crowd monitoring with wi-fi probe. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2018;2(3):115.

17. Di Luzio A, Mei A, Stefa J. Mind your probes: De-anonymization of large crowds through smartphone WiFi probe requests. In: IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications. IEEE; 2016. p. 1–9.