Adam Chang
14 May 2019

Bicycles for Whom: Ecobici in Mexico City

Original Research Question

Over the past two decades, "third-wave" bike-sharing systems have grown in popularity worldwide as modes of urban transportation that, amongst other benefits, reduce environmental impacts and encourage physically active lifestyles. In 2010, the Mexico City government[1] launched its own Ecobici system, which has since grown to over 6,000 bicycles and 100,000 registered users. In a city notorious for pollution and congestion, Ecobici's promise centers on transit that is sustainable, convenient, and affordable (annual plans cost $24 USD). In this piece, we investigate how Ecobici lives up to this utopian vision—and for whom.

Using Ecobici's open data[2] alongside a variety of literature that studies transit in Mexico City and bike-sharing systems worldwide, as well as in-person exploration of the system's physical forms and public perceptions of Ecobici, we answer questions like these, and in doing so, locate Ecobici in the complexities and realities that are Mexico City.

Introduction

While I began with a broad interest in the place of Ecobici in everything that is Mexico City, I inevitably narrowed my scope to the socioeconomic dynamics of Ecobici. I seek to compare the promises of Ecobici as a third-wave bike-sharing system with its socioeconomic realities amongst Mexico City's residents and neighborhoods. To do this, I rely on Ecobici's "open data," which includes data on every ride ever done in the system, and combine this encompassing data stream with various geographic and demographic datasets that allow (crude) estimation of characteristics not captured in the open data, particularly socioeconomic dynamics.

[1] In partnership with a private enterprise, with costs borne by the government.
[2] https://www.ecobici.cdmx.gob.mx/en/informacion-del-servicio/open-data

I first examine the temporal-spatial dynamics of Ecobici station locations. Which districts are best served by Ecobici, for example, and might there be some correlation between districts' Ecobici provisioning and their socioeconomic status? While the open data includes a direct data field for "district," it turns out to be highly imperfect, so I construct a measure of my own using public shapefiles from INEGI on colonias in CDMX. Ultimately, I examine capacity of Ecobici in various districts and colonias by the number of (normalized, per population) stations and slots, the latter being an imperfect approximation of bikes.

With this in mind, I then examine ride data from the perspectives of age and gender, and how they differ across socioeconomic characteristics, roughly approximated by the stations in which rides begin and end.

Transportation and Transit in Mexico City

As Rubén Gallo writes in his introduction to *The Mexico City Reader*, Mexico City suffers from a host of issues related to its car-centric transportation infrastructure:

> "…it suffers from the world's most serious pollution problems (on certain days the smog is so thick that one cannot see across the street, and radio broadcasts advise parents against letting their children play outside); its millions of cars and over one hundred thousand taxis cause endless traffic jams (over the years, the city government has unsuccessfully attempted to ease the problem by demolishing entire streets to make room for six-lane expressways)…" (5).

These issues, while widespread in the city, do not impact residents equally, with wildly divergent effects across socioeconomic status; SES largely determines what forms of transport residents have access to and how they interact with the city's transportation infrastructure, whether positively or

negatively (for example, if highways serve them, or add to noise pollution in their neighborhood, or both). As an example, simply consider residents in satellite districts, "built to shelter their residents from the chaos and pollution of downtown life" (Gallo 13), who are only connected to downtown by multi-lane expressways that cut across other neighborhoods. Still, across the city as a whole, major air pollutant levels "routinely exceed [WHO] maximum exposure limits…with ozone levels exceeding WHO standards for 79 percent of all days in 2005" (Davis 38, 40), spurring significant health risks as well as less obvious social costs—Davis writes that "most residents…avoid outdoor activity during periods of low air quality" (39).

For all of the problems related to automobiles, non-vehicular public transit in the city is not much better. Ricardo Garibay, in "Voyage to the Center of the City," describes the metro as "the deep expressway of neurasthenia," writing:

> "You will emerge black and blue, smelling to high heaven, your clothes in rags, your money gone, your umbrella and briefcase gone after it, half-suffocated, furious, frightened by what you've seen down there, and grateful to have resurfaced in more or less one piece: a miracle wrought from sunrise to sunset, seven days a week, for one and a half million riders[3]…" (Gallo 133).

The metro, in its current form, has become an uncomfortable and occasionally unsafe mode of transport largely reserved for the city's poorer classes, with its frequent crowding and occasional instances of sexual offenses and other crimes.

[3] This figure dates back a while, obviously.

Adam Chang
14 May 2019

Third-Wave Bike Sharing Systems Worldwide

Third-wave bike sharing systems, so named because first-generation systems were stationless and second-generation systems incorporated stations but no customer tracking, have been growing worldwide since 1996. Per DeMaio, these systems raise bike mode share 1.0-1.5% in cities with previously low usage, including in Barcelona and Paris, but, importantly, this statistic comes *conflated* with cycle facility improvements like bike lanes. This latter circumstance presents an interesting challenge to Ecobici in Mexico City, since bicycles alone may not do much.

DeMaio suggests that transit use increases in cities with bike-sharing, especially as a first-mile/last-mile modal solution, and personal vehicle trips decrease, offering critical possibilities for Ecobici in Mexico City. Environmental benefits are obvious, eliminating millions of pounds of CO2 from the atmosphere.

The costs of implement bike-sharing systems, on the other hand, are less uniform across settings, and depends significantly on installation, maintenance, and distribution (including re-balancing) costs. Combined, these are around $5000-6000 across the United States and Western Europe, but may be significantly lower in Mexico City based on costs of labor. Still, the challenges of successfully implementing bike-sharing systems may play a significant role in how Ecobici arrives in Mexico City, including which neighborhoods it serves. Re-balancing is costly and polluting, and station installation requires electrical and physical infrastructure in the surrounding area, both factors that may drive Ecobici to favor neighborhoods with certain characteristics.

The Promises of Ecobici in Mexico City

The government of Mexico City has introduced countless reforms aimed at improving transportation in the city, including a selective vehicle ban in 1989 known as *Hoy No Circula* and

deemed ineffective by public health researchers (Davis 38). Amongst other transportation policy alternatives, bike-sharing systems have gained momentum owing to their relative low cost and ease of implementation, as highlighted by DeMaio.

On the Ecobici website, run by the government of Mexico City, the system is touted as "an essential part of the mobility in [Mexico City]," and the website celebrates its remarkable growth, over 400% between its inception in 2010 and six years later. Most of the system's success seems attributed to its large number of users, stations, and coverage area, but these are aggregate measures unrevealing of specific residents and populations. Perhaps Ecobici lives up to its utopian dreams…but for whom? Has this outwardly optimistic program actually impacted transportation in colossal Mexico City? Even the city's metro, as Garibay writes, had been "billed as [its] salvation" (Gallo 134).

Introduction to Ecobici Data

Somewhat shockingly, Ecobici releases to the general public data on *individual rides* dating back to its inception in February 2010. Data is supposedly anonymized, but several quite revealing features are left, as shown below in a sample entry.

Genero_Usuario	M	Hora_Retiro	00:00:54
Edad_Usuario	34	Ciclo_Estacion_Arribo	68
Bici	8861	Fecha_Arribo	01/03/2019
Ciclo_Estacion_Retiro	143	Hora_Arribo	00:13:55
Fecha_Retiro	01/03/2019		

There are significant privacy concerns with this data, since a wide swath of Ecobici users have repetitive and predictable ride patterns (e.g., those commuting to work on a daily basis). Out of ethical considerations, we avoid any analysis that may take advantage of such structure, and only consider user personal information (like gender and age) in highly aggregate settings.

Adam Chang
14 May 2019

Analysis of Ecobici Stations Data

Methodology

We begin this section by analyzing the locations of Ecobici stations, and the temporal-spatial dynamics of Ecobici station locations. Information on Ecobici stations is available publicly from the system website, with a table of data points for each of the 480 (as of May 2019) stations.

A sample station entry is as follows:


```
{"id":"366","district":"EXI","lon":"-99.181973","lat":"19.378345",
bikes":"8","slots":"16","zip":"3740","address":"366 - Porfirio Díaz-Augusto
Rodín ","addressNumber":"S\/N","nearbyStations":"362,367,368",
"status":"OPN","name":"366 PORFIRIO DÍAZ -AUGUSTO RODÍN ",
"stationType":"BIKE"}
```


Within each entry, our primary features are the location (latitude and longitude being the most direct measure) and the number of slots. While the entries include a "district" field, we note several flaws with the measure as it appears in Ecobici open data. For various stations, their "district" is recorded as "1" or "NaN" (null), reflecting an incomplete and imprecise method for assigning districts. These stations are scattered all over the city, so we leave them out of our analysis, unable to decipher what "1" means. Separately but similarly, districts in the open data are provided as two or three-letter codes (e.g., "H-C," "DOC," "GRA"), which are not explained anywhere online or in the accompanying documentation. By brief estimation and tedious checking, some are colonias of Mexico City, but others seem to be random landmarks. "DOC" is obvious, but what about "GRA"[4] and "ESC"[5]?

As an alternative to possibly poor classification by Ecobici open data administrators, we *independently categorize stations' locations* using latitude/longitude data and shape files of CDMX's various colonias. In all results below, we alternate between these two assignment

[4] Granada, which is Nuevo Polanco?
[5] Escandón, which is South of Condesa?

schemes for transparency. The format of the colonia shapefiles are shown immediately below; it's straightforward to locate Ecobici stations given each of the colonia "polygons."

	OBJECTID	POSTALCODE	ST_NAME	MUN_NAME	SETT_NAME	SETT_TYPE	AREA	Shape_Leng	Shape_Area	geometry
0	11065	01000	DISTRITO FEDERAL	ÁLVARO OBREGÓN	SAN ANGEL	COLONIA	711515239.0	0.053887	9.466505e-05	POLYGON Z ((-99.20257000020518 19.345380000187...
1	11066	01010	DISTRITO FEDERAL	ÁLVARO OBREGÓN	LOS ALPES	COLONIA	711515239.0	0.029948	4.433100e-05	POLYGON Z ((-99.19058000036659 19.363460000278...
2	11067	01020	DISTRITO FEDERAL	ÁLVARO OBREGÓN	GUADALUPE INN	COLONIA	711515239.0	0.034212	5.763755e-05	POLYGON Z ((-99.1851499999724 19.3618699998008...
3	11068	01030	DISTRITO FEDERAL	ÁLVARO OBREGÓN	AXOTLA	COLONIA	711515239.0	0.025519	3.213660e-05	POLYGON Z ((-99.17462000001285 19.361860000239...
4	11069	01030	DISTRITO FEDERAL	ÁLVARO OBREGÓN	FLORIDA	COLONIA	711515239.0	0.044882	8.945425e-05	POLYGON Z ((-99.18276000037355 19.362370000375...

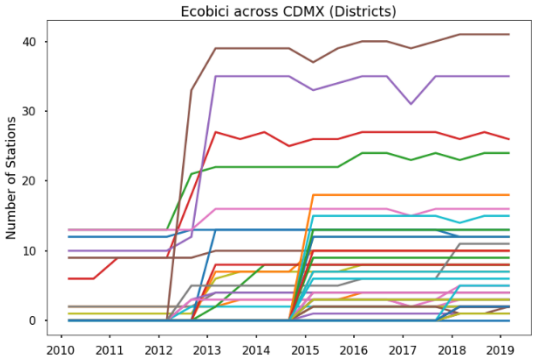
With these features in mind, we begin to compute various aspects of the temporal-spatial distribution of Ecobici. First, we calculate the number of stations in each district in each month, by determining which stations have recorded activity in the open data for each month (with the extremely unlikely caveat that a station may exist yet have no activity). We normalize this metric to district population, reflecting some measure of Ecobici capacity and growth in the districts, and do the same for number of total slots in each district. To be clear, existence of a station and even its number of slots are not perfect measurements of Ecobici capacity in a certain area; with factors like the previously mentioned importance of re-balancing, it's possible that a neighborhood may have many stations and even many slots, but be underserved by the Ecobici system, if there is little re-balancing and slots/stations are left unused. At this stage of an analysis, however, it's impossible to better understand these dynamics without delving deeply into actual ride data, which we do later.

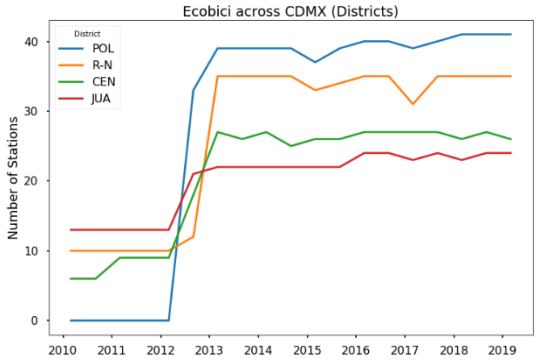
Using INEGI census data, we then aggregate districts across Mexico City, primarily based on residents' socioeconomic status. SES by itself, of course, cannot purport to reflect the actual circumstances of a place (and can often be misleading), so we attempt various other district

groupings. Across aggregated district/colonia groupings, we re-compute the capacity measures described in the previous paragraph.

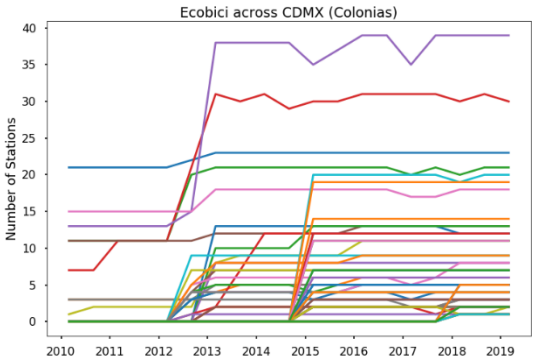
In which districts are Ecobici stations?

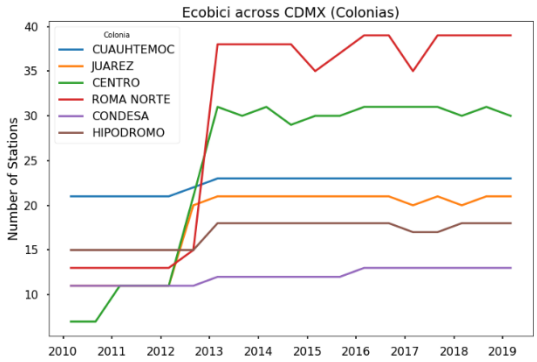
The plots below are of Ecobici capacity per district, measured at six-month intervals.





These first two graphs are across districts, as determined by the creators of the Ecobici open data. We see that several districts (Polanco, Roma Norte, Centro, and Juarez) achieve significant growth in the number of stations, from the labeled graph on the right.

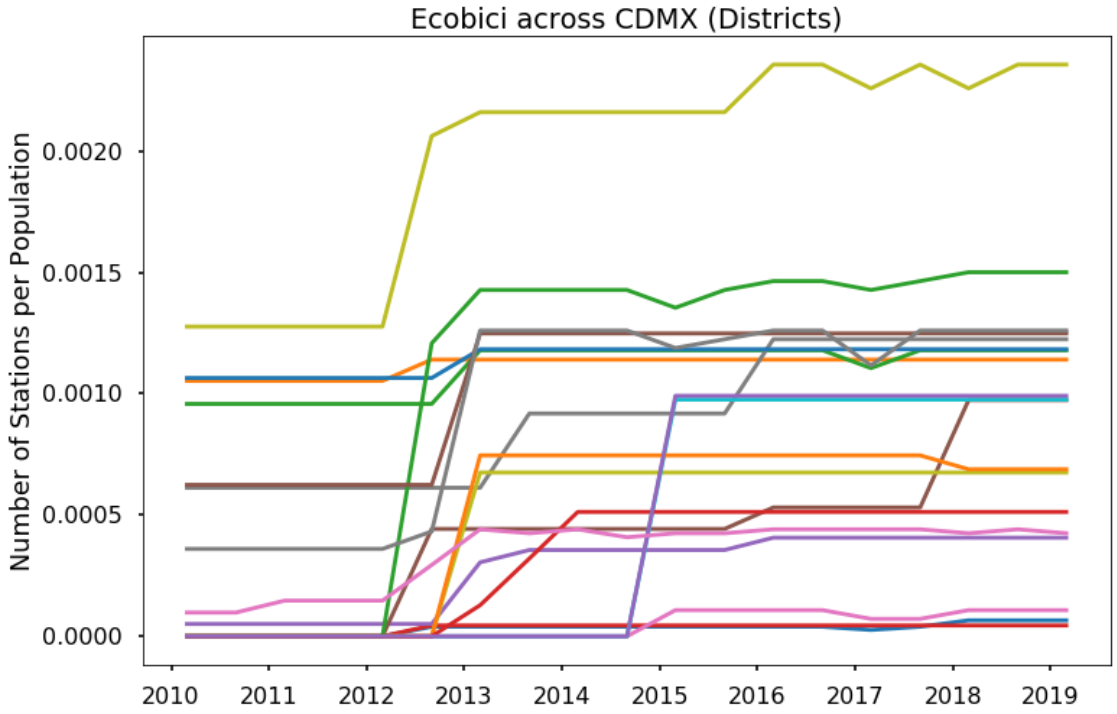




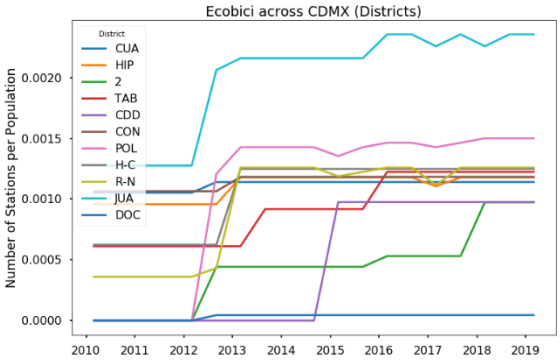
The next two graphs are across colonias (that I independently selected). Almost the exact same takeaway proceeds from this: a few colonias were responsible for adding many stations.

Adam Chang
14 May 2019

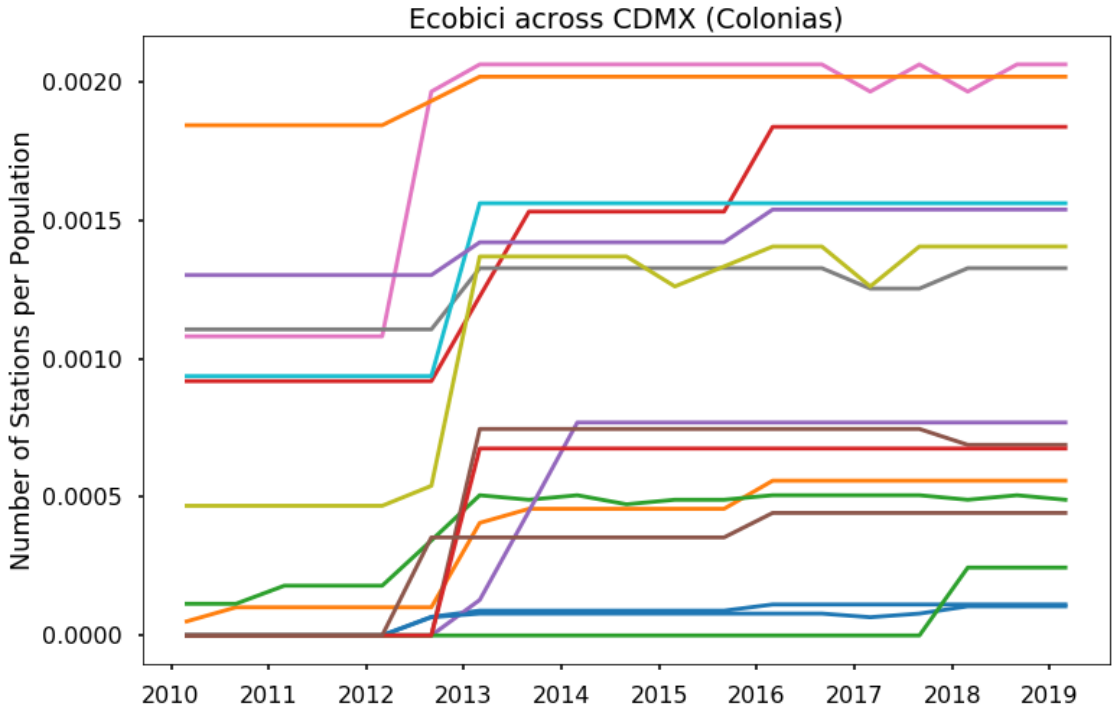
Here's where it gets interesting. Normalizing to population with INEGI census data for each district and colonia, we obtain the following plots.

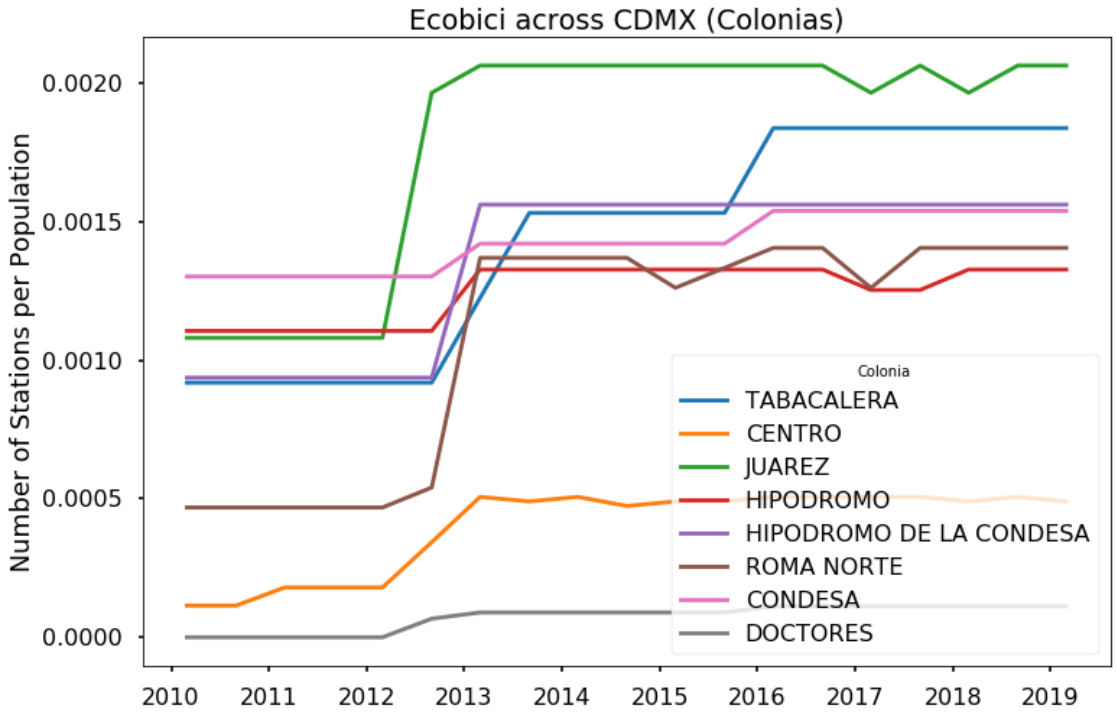


The labeled version is presented at right, and we immediately notice that certain districts are much better served by Ecobici, normalized to population…and that they tend to be more upscale districts. The flat line at the bottom of the graph represents capacity in Doctores, one of the most crime-ridden neighborhoods in CDMX. On the y-axis, the figures represent capacity: 0.0020 stations per population means that there's a station for every 500 people, whereas 0.0005 means that there's a station for every 2000 people. The plots for colonia (independent selection process, based on shapefiles) follow. They speak for themselves.



Adam Chang
14 May 2019



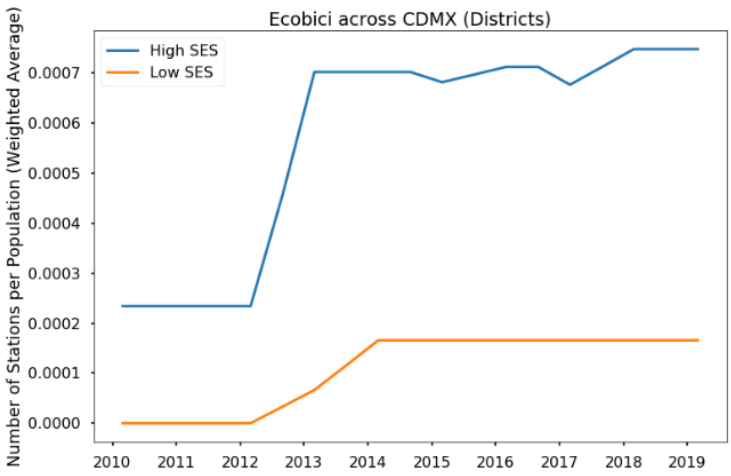


Adam Chang
14 May 2019

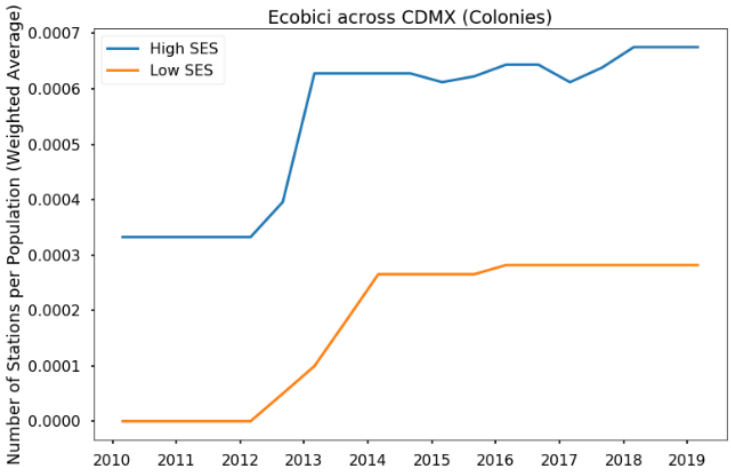
Ecobici across districts by SES-level

We begin this section by emphasizing the large caveat that our initial analysis below does an inherently poor job of examining the socioeconomic dynamics of Ecobici. The SES of an Ecobici station's district is a *highly imprecise* measure of the SES of a station's users, since riders often use stations outside of where they live (indeed, it may be argued that riders *primarily* use stations outside of their residence district, if Ecobici is part of their commute). To achieve precision under such circumstances requires context-specific expertise and understanding, and we proceed to work towards such a goal in later sections. At this time, we limit ourselves to simply examining station/district data, as described in the above section on methodology.

We aggregate districts and colonias by general SES-level, using INEGI and general knowledge. High-income districts are: ["CUA", "R-N", "CON", "HIP", "H-C", "ROM", "POL", "GRA", "2"], with "GRA" representing Nuevo Polanco and "2" representing Anzures. High-income colonias are: ["CUAUHTEMOC", "ROMA NORTE", "CONDESA", "HIPODROMO", "HIPODROMO DE LA CONDESA", "ROMA SUR", "GRANADA", "ANZURES", "LOMAS DE CHAPULTEPEC"]. And low-income districts are: ["DOC", "BUE"]. Low-income colonias are: ["DOCTORES", "BUENAVISTA"].



Adam Chang
14 May 2019

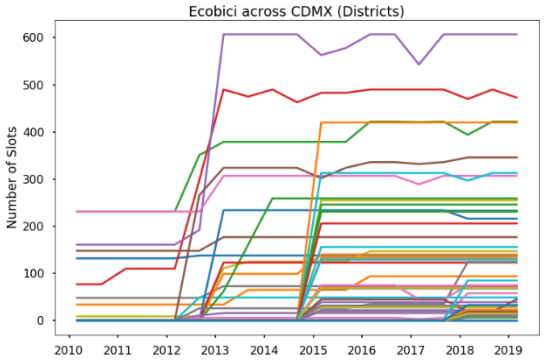


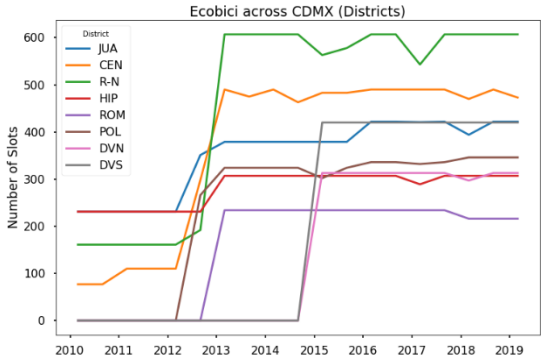
We end our analysis of station data here. Regardless of which districting scheme chosen, whether that which comes with the Ecobici open data or that which I've implemented using shapefiles, the conclusions are the same. In higher SES districts/colonias, there are significantly more stations normalized to population; the effect is diverse for specific districts, but aggregated together, higher SES locales fare much better in Ecobici service delivery than lower SES ones.

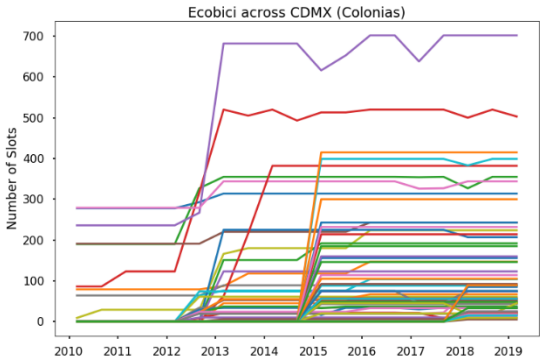
Adam Chang
14 May 2019

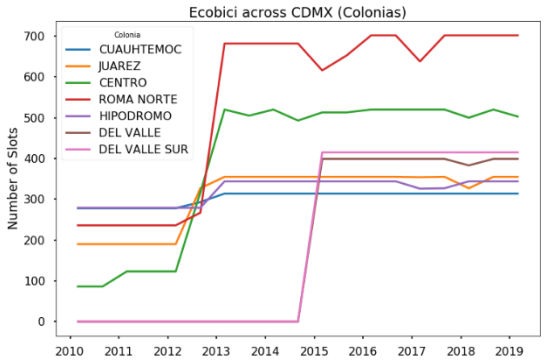
Analysis of Ecobici Slot Data

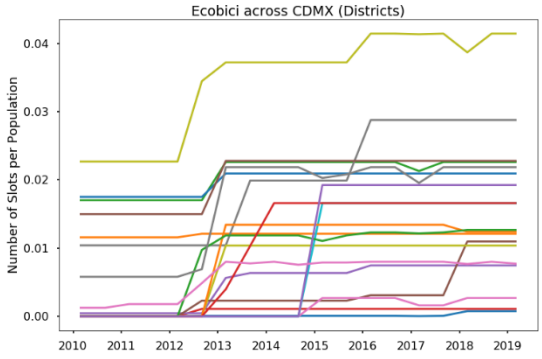
We conduct the above analysis for slots in districts/colonias, instead of just stations. Slots are taken from the stations data table. Most conclusions are the same, though there are some interesting features of the “DVN” and “DVS” neighborhoods (Colonia del Valle), where there were comparatively few stations, but many slots now.

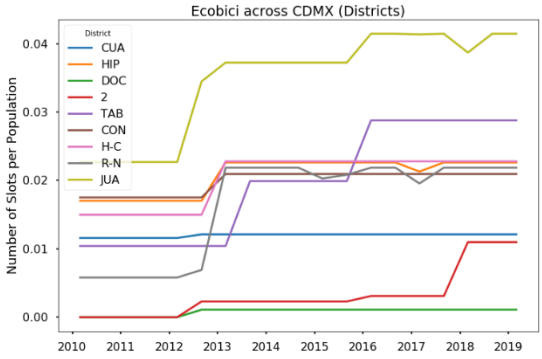




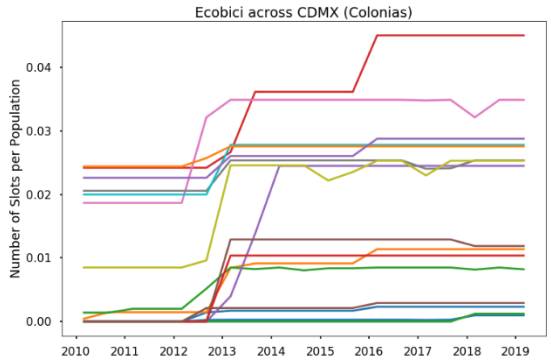


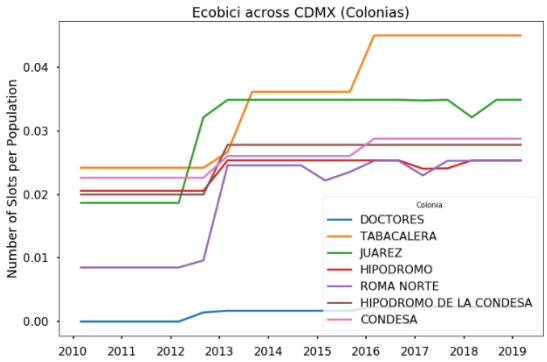


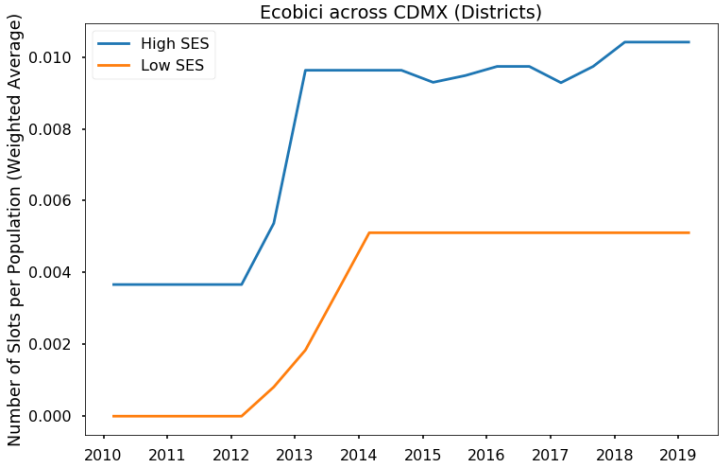


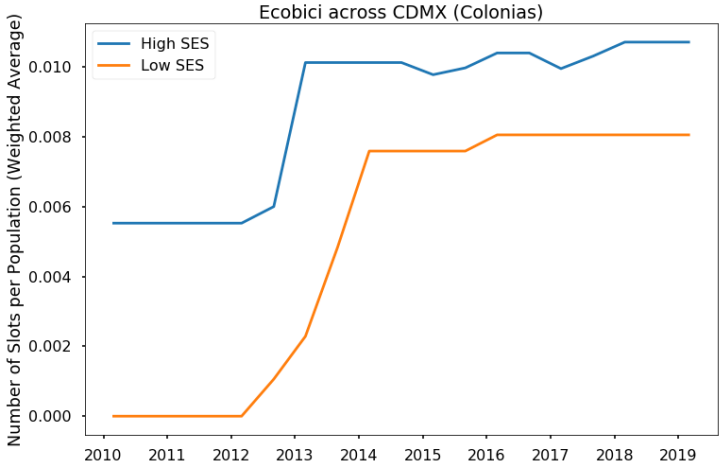


Adam Chang
14 May 2019









Again, high SES districts and colonias are much better served by Ecobici than low SES ones. This holds true not only for stations, but also slots (and essentially bikes), as we now show.

Adam Chang
14 May 2019

Analysis of Ecobici Rides Data

Introduction to ride data

The open data on individual rides offers us significant opportunities in surmounting various caveats based on inferring activity from station/slots presence. Specifically, we now have an actual measurement of activity, and its resolution (i.e., very specific timestamp) allows us to proceed with much greater inference power than we could before—we have some ability now, for example, to separate out "commute" rides.

With great opportunity also comes great challenges, however, and great challenges lie in the sheer quantity of data we have available. Combining monthly data for the past *nine* years results in a data stream of around 4 GB, which may not sound like much but makes even loading the data quite difficult[6]. To overcome the limitations posed by time and computing power on working with such data, we limit our analysis to the single month of March 2019 (the most recent, and one of the largest, dataset available)—even in this single month, there were over 750,000 rides!

We lose out on important aspects of the data in this way, especially those that pertain to historical trends, but make this trade-off for the valuable opportunity of being able to examine ride-specific data, as they relate to gender, timing, and more. We don't claim that our results are generalizable to the history of Ecobici at large, only Ecobici in its current state, which may be significantly different from its previous realities (for example, are older people more comfortable using the system now, since they typically tend to be technologies' "late adopters"?). A worthwhile future endeavor would involve running the analyses that follow on the dataset as large, and extracting conclusions about how the system has evolved over the past nine years.

[6] See https://opendata.stackexchange.com/questions/1256/how-can-i-work-with-a-4gb-csv-file. I did try.

The obvious suggestion here would be simply including another month, from further back, in the analysis, and perhaps deriving some trends that way. But conclusions from two points should mean very little—if the proportion of rides by female users increased significantly from March 2012 to March 2019, does that mean women have become more comfortable using the system? Or is it possible that in March 2012 women just stayed at home more in general, because of a crime spike or because they were more genetically predisposed to having springtime allergies? At this point, it becomes pointless to limit the analysis to just two months, and the argument easily comes for the entire nine years' worth of data.

(We do check that March 2019, the month we use, is not a significant outlier in any of the features we describe below, and it doesn't appear to be, as compared with similar periods.)

Rides and socioeconomic status

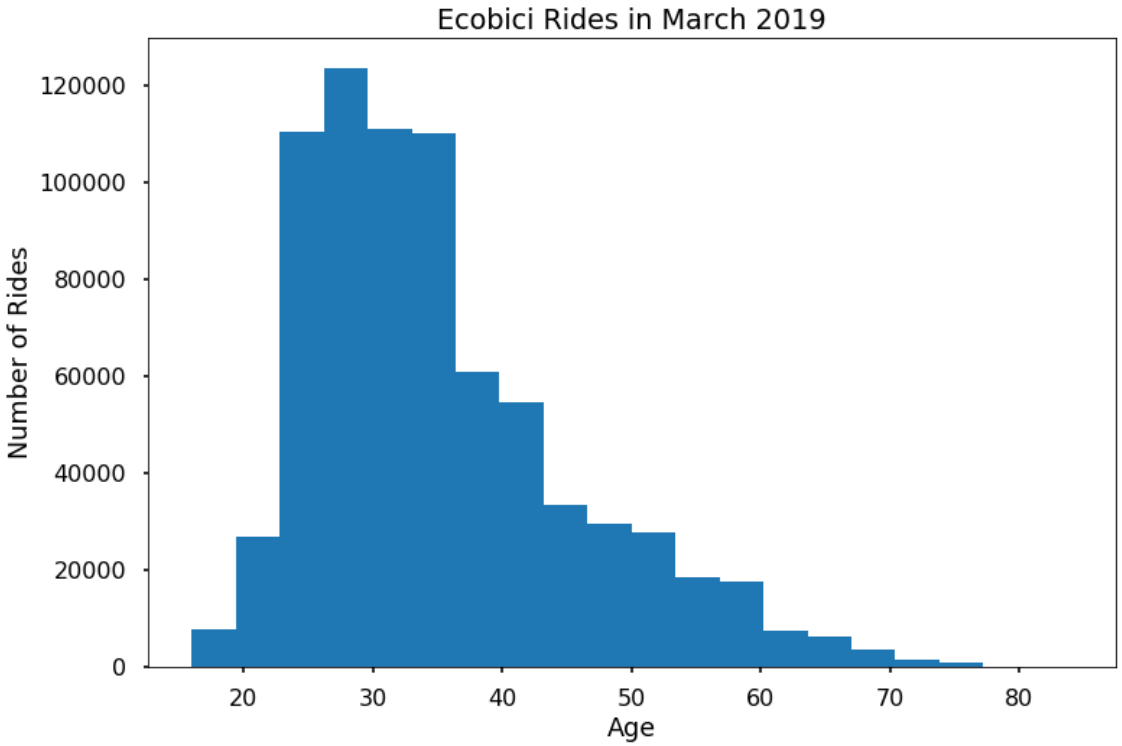
We (crudely) categorize a ride as low SES if its start and end stations are in low SES districts/colonias, and the opposite for high SES.

Of all rides, 1971 (0.26%) are estimated to be low SES, and 217,129 (28.9%) are estimated to be high SES. This disparity highlights the points previously made about Ecobici service provisioning to low and high SES areas.

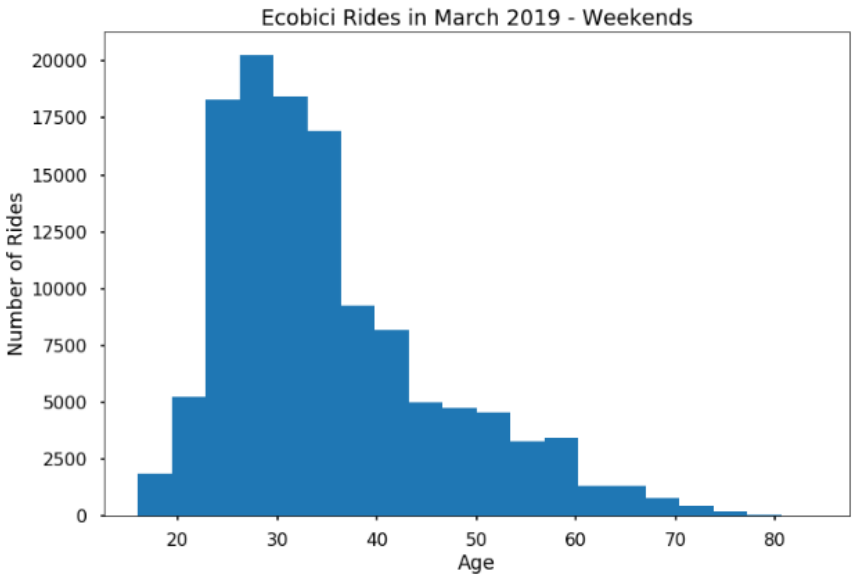
Age of riders

As we would imagine, we find that the majority of rides tend to be by users ages 25-35, with a long tail extending to the upper 70s.

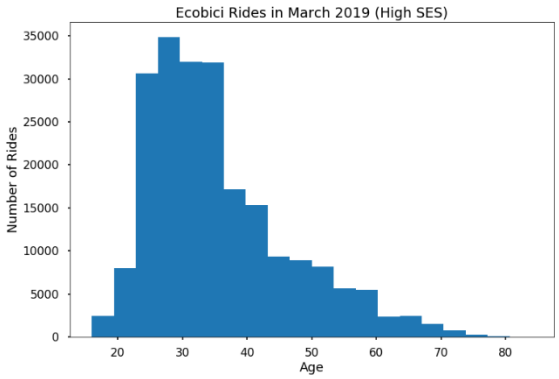
Adam Chang
14 May 2019

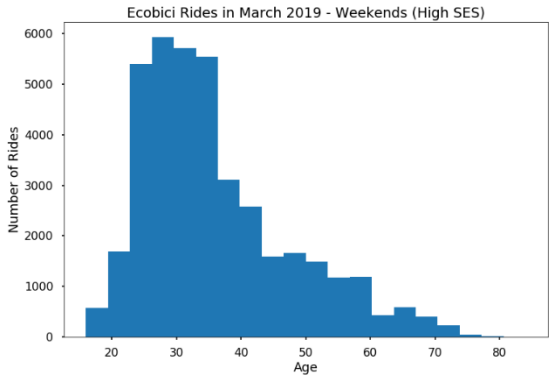


Surprisingly, this distribution of ages-of-rides does not appear to be significantly different on weekends, suggesting that those who use Ecobici on weekdays also use Ecobici on weekends.

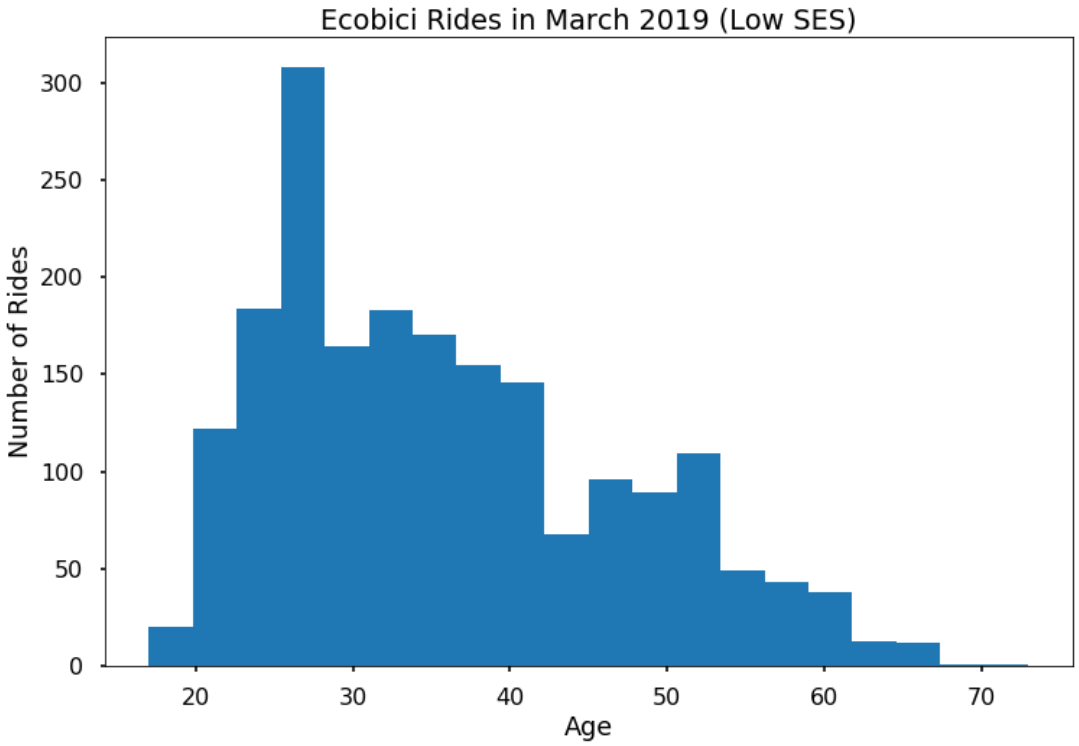


The same patterns held for high SES rides, as shown below.

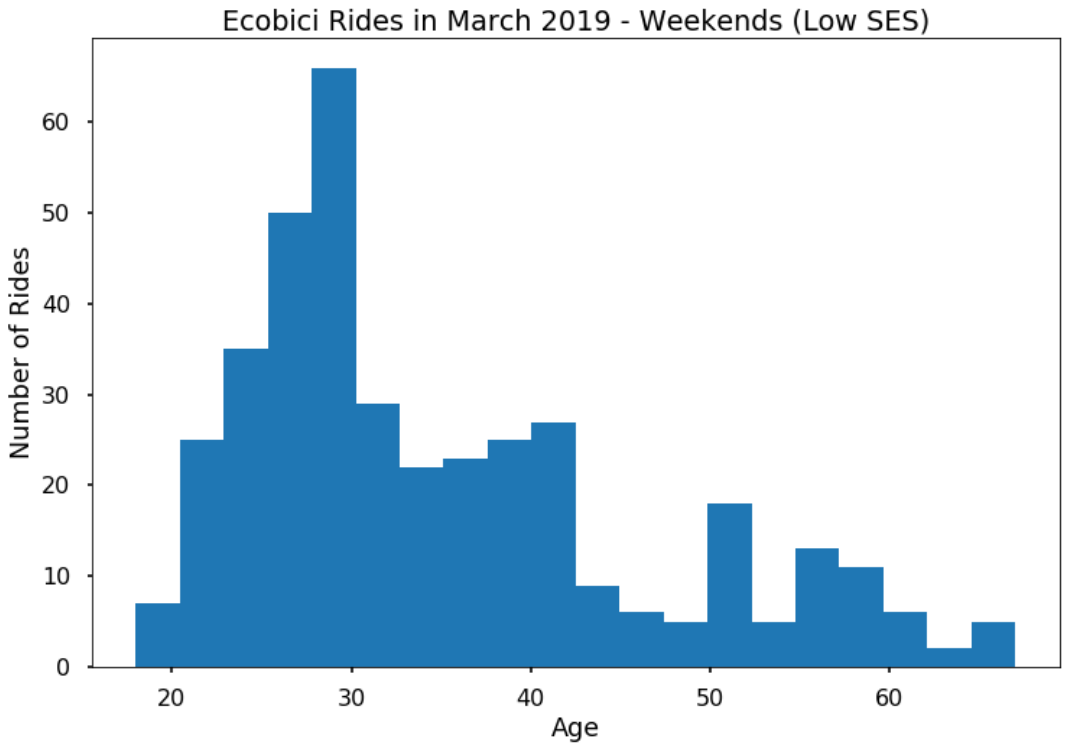




Low SES rides, however, seemed to follow a different pattern of age distribution, not tailing off until age 40 or so, with a high peak at age 25. Because of the very limited data here, however, we simply treat the peak as an outlier.



The pattern is more or less true on weekends as well.

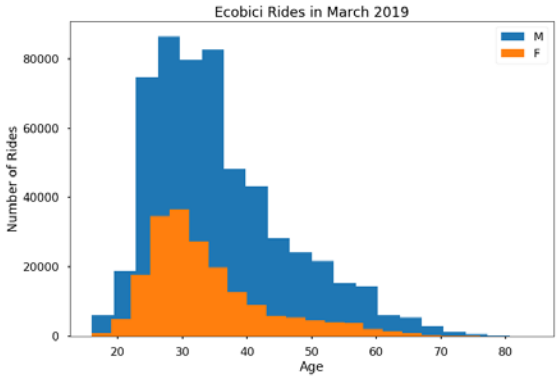


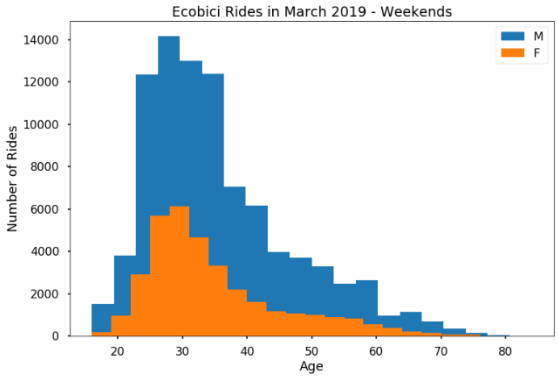
Gender of riders

There was a significant disparity in male and female Ecobici users, with male users accounting for almost 75% of rides (560,305 of 752,299). On weekends, female riders constituted a slightly higher percentage of rides (27.5%, a 10% increase), but their age distribution also remained similar to that of weekdays.

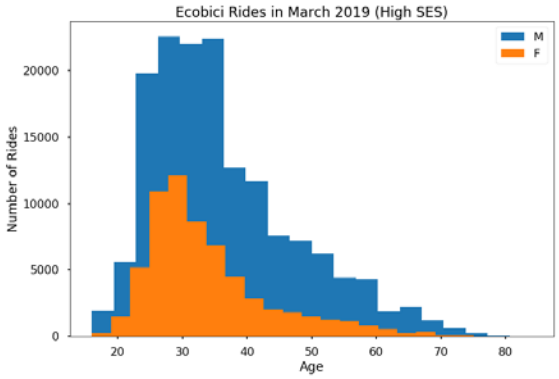
Interestingly, across all high SES rides (not just on weekends), females constituted 62,478 of 217,129 rides, around 28%. This represents a significant increase from their baseline of 25%, suggesting that females are more likely to be users of Ecobici in high SES areas. The opposite is true for low SES areas, where females only made 335 of 1,971 rides, or 17%.

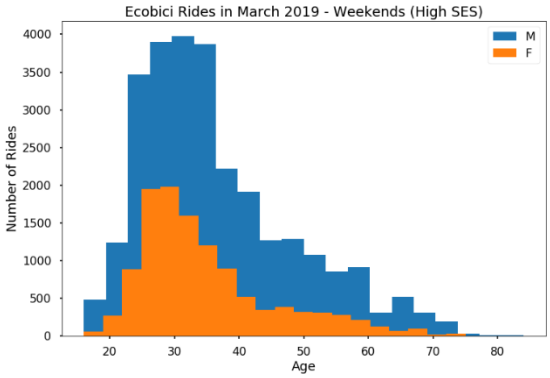
Combining these features, on weekends in high SES areas, females constituted 11,522 of 39,353 rides, or over 29%. In low SES areas, females rode 73 of 389 rides on weekends, around 19%. The clear trend here is that females tend to ride more in high SES areas and on weekends, with the effect magnified once both are combined.





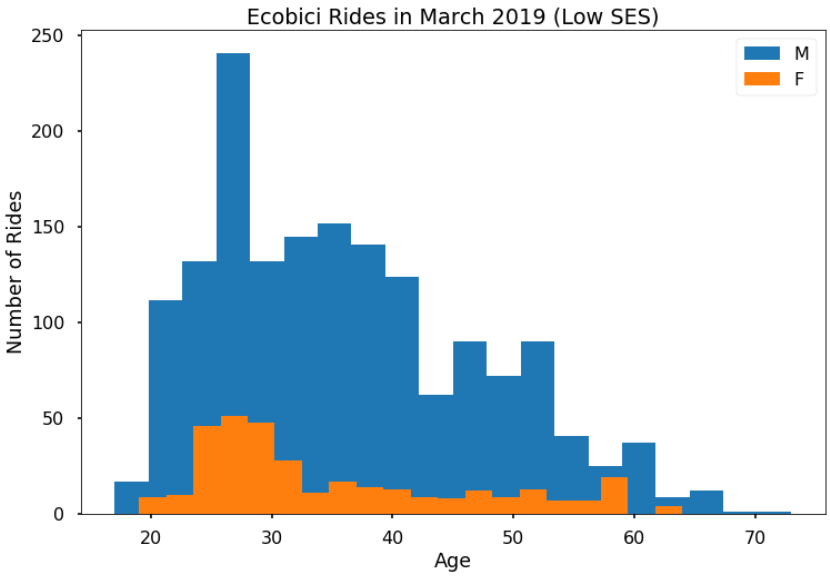
Across the age distribution, female rides tended to be more concentrated amongst younger female users, with the peak quickly tapering off after age 30, compared with males who taper off after their late 30s. This pattern holds for high SES areas and on weekends, and for the combination of the two.

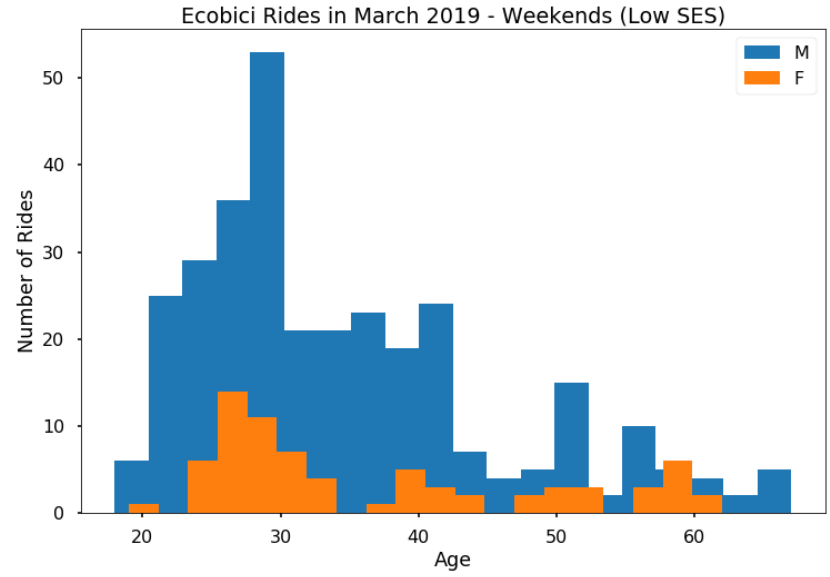




Adam Chang
14 May 2019

In low SES areas, the lack of data again makes analysis difficult, but female riders tended to skew younger, with the peak in the mid-to-late 20s.





Adam Chang
14 May 2019

Conclusions

Mexico City's Ecobici has made headlines as a third-wave bike-share system in a city desperate for alternative transportation policy. Still, its aggregate measures of success—in numbers of stations, riders, and coverage—hide significant disparities in the system's service delivery.

We've shown that the number of stations and number of slots, normalized per population, are distributed highly unequally based on districts' SES. We construct a separate measure of geography, based on INEGI shapefiles, and find that these two measures of capacity remain highly unequal between high SES and low SES areas. With ride data, we conclude that in low SES areas, there tends to be fewer female riders, and these riders skew younger.

An interesting extension to this data problem would be to record the *membership* status of users, even if to differ between yearly plans and short-term riders. There are very few statistics on what proportion of Ecobici users/rides are long-term users, but this could shed light on problems that we attempted to address through estimation—for example, commute rides—and Ecobici does have this data available (and it doesn't seem that much more intrusive than what they currently publish).

With this, Ecobici still has a long way to go toward improving transportation in Mexico City, especially since those facing the most difficult transportation circumstances are likely to be those in lower SES areas, who lack the resources and capacity to mitigate problems like traffic, pollution, and crime on the metro. Only through much better, much more equal integration of these populations can Mexico City build transportation that truly serves the city.

Adam Chang
14 May 2019

Acknowledgment

Simon Bedford had previously done some analysis of Ecobici data and published it online for public use. I adapted some aspects of his analysis, like his decision to use colonia shapefiles, but those adapted parts constitute a *highly insignificant* part of my work. Simon did not have any analysis related to SES, nor did he incorporate the colonia shapefiles in any meaningful way.

Nevertheless, I'm very grateful for his analysis, in that it (1) showed me that the Ecobici open data was out there and usable and (2) jumpstarted my thinking about how to proceed further from the raw data.

Still, all code in my work is original, except for the brief section involved in downloading the open data CSV files at the very beginning, which is copied from him (in Princeton computer science courses, this would be acceptable use).

Source: https://github.com/simonb83/ecobici.

Adam Chang
14 May 2019

Bibliography

Bayón, María Cristina, and Gonzalo A. Saraví. "The cultural dimensions of urban fragmentation: Segregation, sociability, and inequality in Mexico City." *Latin American Perspectives* 40.2 (2013): 35-52.

Davis, Lucas W. "The effect of driving restrictions on air quality in Mexico City." *Journal of Political Economy* 116.1 (2008): 38-81.

DeMaio, Paul. "Bike-sharing: History, impacts, models of provision, and future." *Journal of public transportation* 12.4 (2009): 3.

Gallo, Ruben. *The Mexico City Reader*. University of Wisconsin, 2004.