Part IV: Methodology

This project relies primarily on quantitative methods, although qualitative studies were heavily consulted to inform analytical parameters. More specifically, I first queried Google Maps' shortest-path routing assignment algorithms to calculate the minimum transit-based travel times between (1) a sample of *asentamientos* from across AGBA and (2) a series of ten important activity sites (e.g. central business districts/employment centers, rail stations, schools, and public health centers). Then, I supplemented these values with a battery of statistical tests to see whether travel times from the *asentamientos* were statistically-significantly different from those values associated with some of the conurbation's formal neighborhoods. Given that I used black-boxed, third-party tools, i.e. Google's Distance Matrix API, to determine travel times, the description of my spatial methods is simplistic, especially when compared to the data preparation process outlined below.

In turn, I will first define and describe the Google Distance Matrix API tool and its travel-time calculation process, inputs, parameters, assumptions, outputs, and functional limitations. I will then discuss the steps by which I acquired and formatted the tools' inputs (e.g. origins and destinations) and how I determined modal parameters (e.g. departure dates and times, modes, etc.) selected for my transit-based travel time requests. Lastly, I will briefly discuss formatting the API's travel-time matrix outputs and the differences-of-means tests (ANOVA and independent-sample t-tests) that I implemented.

Part IV(a), Google Distance Matrix API

To calculate transit-based travel times within Greater Buenos Aires, I opted to use Google Maps' Distance Matrix API (Application Programming Interface) web service. In short, this tool allows researchers to determine the shortest-path travel times between hundreds of different origin-destination pairs, utilizing the same transportation network data and routing algorithm that Google Maps employs every day when performing on-demand route-finding requests through its popular web interface (Wang and Xu 2011, p. 200). For reference, shortest-path routing algorithms, like the one employed by Google, try to estimate the route between a given origin and destination pair with the shortest travel impedance (time or cost) through a given network. They are associated with the trip-assignment stage of traditional travel modeling, where planners seek to "determine the trip-maker's likely choice of paths between zones ... along the network of each mode ... and predict the resulting flows ... on the individual links [of that] network", whether highways, transit, or some combination of the two (Papacostas and Prevedouros, 2000, pg. 400).

According to Google, using an API "gives developers several ways ... [for] retrieving data from Google Maps ... [with] simple or extensive customization" that, in the case of the *Distance Matrix API,* "provides travel distance and time from a matrix of origins and destinations ... based on the recommended route between start and end points, as calculated by the Google Maps API (Google 2018)."[1] The API, in turn, produces an output that is a travel time matrix indicating the amount of

---

[1] https://developers.google.com/maps/documentation/distance-matrix/intro

time that Google Maps estimates—based on its (rather secretive) algorithm—would be required to travel between the origins and destinations of interest to the researcher.

Nevertheless, these estimates depend on one of the API's integral functionalities: its parameters regarding travel mode, arrival/departure times, units, language, and day-of-the-week. When it comes to requesting these data, Google has several interfaces (or "client libraries", which "make developing with the Google Maps web service API easier by providing simple, native implementations of common tasks") that allow the user to query their servers via well-known programming languages or software packages. I chose the client library created for the statistical programming language R.[2] Following the advice of Wang and Xu (2011), I selected R because it is good at handling large data requests and would allow me to easily perform statistical tests once the time values were acquired. The specific tool within R for performing these requests is called "gmapsdistance" and requires an input value for each of the parameters listed above.[3] Next, I will describe the steps required to use the Distance Matrix API tool.

The first step is to acquire and prepare the data needed for the tool's three fundamental inputs: origins, destinations, and API key. The origins and destinations are the geographic points representing the starting and ending points for any travel-time analysis. These points can be provided in the form of latitude and longitude coordinates or full addresses. As for the number of origins or destinations that can be entered at, users are permitted (with a standard, free Google account) to enter just twenty-five (of each) at one time. In R, these addresses or latitude/longitude pairs must be entered as a string or a vector of string values, each separated with a "+".

In fact, the monitoring and limitation of user activity explains the importance of the final required input: the API key. An API key is a unique identifier code that a user must enter into the program before making any requests. Each key is linked to a user's Google account. As a result, they can track and, most importantly, limit the number of requests made to their servers at once. For the Distance Matrix API, the daily quota is set to 2,500 free elements (i.e. one origin/destination pair is one element), 25 origins and destinations per request, and 100 requests per second. Once a user reaches this total for the day, the application will stop returning results (in my case, I used one key for each of my three Google accounts to increase my daily return). If someone has a paid account, they can increase their daily quota to 100,000 elements.

After the origins, destinations, and API key are set, the user will then set the remaining parameters. The user first decides the mode of travel (driving via the road network, walking via pedestrian paths and sidewalks, bicycling via paths and preferred streets, or transit via public transit routes), the language and units (metric or imperial) of the output, the arrival and/or departure time (in UTC time), arrival and/or departure date, and the shape of the time table (long or wide). Some parameters are specific to certain modes; if driving is selected, the user can stipulate certain traffic conditions (optimistic, pessimistic, or best guess) or what types of feature to avoid (tolls, highways, or ferries). As for transit-based modeling, the user can stipulate specific modes—bus, subway, train, tram/light rail, or rail (a combination of subway, train, and tram)—as well as routing preferences—

---

[2] https://cran.r-project.org/web/packages/gmapsdistance/gmapsdistance.pdf
[3] https://github.com/rodazuero/gmapsdistance

selecting routes with less walking or fewer transfers (if appropriate). If not stipulated, many of these parameters have a set default value: driving as the mode, English as the language, metric as the unit system, and the present day and time for the departure. In R, these parameters are included within the "gmapsdistance" tool and are all stipulated as text strings. I will walk through the process of selecting my parameters below.

Once the parameters are set, and the tool is run, the final output is a "list with the traveling time(s) and distance(s) between origin(s) and destination(s)." The output table contains a row for each origin and three columns for each destination: one for the distance (in meters), one for time (in seconds), and one that displays a "status" that indicates whether that origin/destination pair calculated correctly ("OK", "NA", "INVALID", and "OVER QUERY LIMIT" are common statuses, should they apply). The first row and column of the table also contains the latitude/longitude coordinate for the corresponding origin or destination, as entered into the API. Serving as unique identifiers for each row, they allowed me to join these data to other tables.

Given how recently Google Maps introduced its Distance Matrix API web tool (launched in 2006), its application within the academic literature is nascent. Nevertheless, researchers have already documented its advantages over alternate methods of estimating network-based travel times. In an article from 2011, Wang and Xu compared driving times estimated by the API with those from a network dataset self-constructed within ESRI ArcMap and its network analyst tools. Looking specifically at access to hospitals, they found Google's estimates to be *longer* than those from ArcGIS, an indication that the former was actually accounting for traffic congestion and other possible delays that are more difficult to represent with ArcMap.

Within their servers, Google Maps maintains and constantly-updates a massive quantity of transportation data on network configurations and characteristics (e.g. real-time traffic data, prior congestion levels, speed limits, restricted turns on urban streets, etc.) that permit estimations of travel time more accurate and up-to-date than most anything a single researcher, or team of researchers, could feasibly pull together within a custom-built network. For instance, most of which, for instance, rely solely on speed limits to calculate travel times along road segments and are unable to account for diurnal variations in traffic, a key feature of the API) (Wu 20170.

In turn, the user does not have to waste time collecting huge sums of transportation data or familiarizing herself with complex road and transit network data structures—all of the "dirty work" of modeling the network is already done. In the words, of Wang and Xu, "modeling is as good as the data get (2001, pg. 202)," and, in the case of Google's data, it is simply the best (especially in Argentina, as I will explain below). Nevertheless, the authors noted some of the API's remaining limitations: the number of requests is limited for users without a paid license, there is little transparency for non-Google users related to data quality or routing algorithms, and the servers are still prone to returning seemingly-random errors for certain requests (pg. 208).

Academic employment of the Distance Matrix API for transit-based analysis, however, appears Sparse, so far. I am not aware of any papers that explicitly look at the methodological advantages of transit-based travel time queries using Google (at least in a manner similar to Wang and Xu's evaluation of drive times), although a bit more has been written about Generalized Transit Feed Specification (GTFS) datasets, the data structure employed and developed by Google for

standardizing transit schedules around the world. GTFS—which was first introduced by Google in 2005—consists of a standardized series of relatable CSV files containing a transit agency's name, routes, schedules, frequencies, stop locations, shapes, transfer points, and fare attributes. Regardless of an agency's size or location, each of these files utilizes the same simple, open-sourced formatting; this helps to "facilitate data sharing and access to information" and to be easily operable with online applications that provide route and schedule information to transit users (Fortin, et al. 2016, pg. 22).

In turn, GTFS has become standard within transportation departments in the United States and, increasingly, the world. Recent innovations have made GTFS datasets interoperable with GIS programs like ArcGIS, permitting users to more accurately study and produce "service area calculations, ... time and distance service calculations, stop location and spacing optimizations, [and] service frequencies (Fayyaz, et al. 2017, pg. 5)." Additionally, when looking specifically at transit, the GTFS schedule estimator is paired with calculations of the walking distances and times required for users to (1) reach the system and (2) make any in-route transfers (Wu 2017).

All the while, two of these authors (Fortin, et al 2016; Wu 2017) acknowledged shortcomings of GTFS and transit travel time estimators derived from it. The data files are not insusceptible to data-entry errors or network "misrepresentations" (e.g. stops or routes with incorrect coordinates) and that the schedules are, *in fact, schedules* and not innately able to account for traffic delays or service disruptions that push transit behind-schedule. Furthermore, since Google does not publicly disclose its routing mechanism or means for estimating travel times, it is difficult to verify their claims (Wu, 2017). There is a need for academic research that validates the accuracy of Google's transit-based travel time estimates, and these limitations are acknowledge vis-à-vis my results.

In terms of applying Google Maps' web tools to questions of transit in a metropolitan area like Buenos Aires, Argentina (or, more generally, cities in Latin America and the Global South), there is actually some precedent. In a piece from 2016, Boisjoly, et al. used the Distance Matrix API to estimate travel times between residences and employment sites in Sao Paulo, Brazil. More specifically, they queried the API for transit-based travel times that left their stipulated origins at 7am (during the metro area's peak travel hour), all while minimizing necessary transfers. All of these were made possible by the conversion of Sao Paulo's transit schedules into GTFS (although not *all* of these have been made public, according to the authors), a process that is similarly ongoing in Buenos Aires. It should be noted, however, that the authors note the risks of using a singular departure date and time for their study (and the greatly different outcomes that could result from small shifts in departure time) but acknowledge that these small variations are not as important when diagnosing access at an aggregated, metro-level scale (pg. 91). Knowing that my technique was used to analyze accessibility in a similarly-sized Latin American city adds confidence to its application in metropolitan Buenos Aires.

In fact, Google is perhaps the most authoritative source for travel data in agglomerated Buenos Aires. According to a 2016 article from *La Nacion*, one of Buenos Aires' primary newspapers, Google worked extensively with the governments of CABA and the departments of AGBA to acquire all the transit schedules needed to allow local travelers to request transit-based directions (with corresponding estimates of travel time, distance, and cost) through Google Maps'

web and mobile applications. According to the article's author, Google collected schedules from an array of government agencies: CABA's municipal government, the federal transport ministry, Ferrovías (a private company operating one of Buenos Aires' commuter rail lines), Metrovías (another private company operating CABA's subway system), and the CNRT (government agency responsible for transport statistics), and other private enterprises (presumed to be the operators of major bus lines). Altogether, they collected data on 800 routes and 34,000 stop locations! Argentine public users gained access to this information—through the Google platform—in early 2016, following in the footsteps of 18,000 other cities in 70 countries from around the world (Tomovose 2016). Combined with existing roads datasets, users could now request directions for driving, walking, and transit (bus, train, and subway) between any two points in the conurbation.

   The incorporation of transit into Google Maps' directions tools in Buenos Aires presupposes that Google converted the collected schedule data (from the different local agencies and companies) into GTFS format. Since these datasets have not been made public (as would be the case for transit agencies in the United States), it is impossible to know for sure (or to manually check their quality). Outside of the City of Buenos Aires (which operates its own travel-time route estimator—known as *ComoLlega*—for requests within its boundaries and publishes GTFS files for its bus and subway routes), there is little to no public official information on transit schedules for routes within the AGBA's departments. The only exceptions are the printed time tables for the individual commuter railway lines (naturally forcing one to ponder how *exactly* Google was able to cajole these data from existing operators, some of which are colloquially said to not even follow exact schedules, operate along informal schedules or headways, or that lack dedicated stops).

   As such, it would be incredibly difficult to self-construct a network dataset—akin to those native to ArcGIS—using publicly-available data in the conurbation. Serious assumptions (with high likelihoods for error) about schedules, transfers, and wait times would needed, likely producing estimates that are significantly biased. In pursuit of accurate travel time-estimations, I am essentially forced to call upon the Google Distance Matrix API. In turn, this immediately places this accessibility study ahead of all previous tabulation attempts within AGBA, most of which, in the absence of reasonably-accurate time estimates, turned to physical distance as their proxy for access. Nevertheless, I acknowledge that has limitations: black-boxed scheduling and routing techniques as well as unverifiable results.

Part IV(b), API Parameters, Origins

Selecting API-compatible "origin" coordinates, in order to represent the *asentamientos*, was an intricate task. I began by visiting the TECHO website and downloading—from their interactive online map—a shapefile of all informal housing neighborhoods in Argentina.[4] Once uploaded into ArcGIS, I then selected, from this file, only those neighborhoods classified as *asentamientos* (TECHO also logs *villas*) that belong to one of the thirty departments of AGBA, a total of 687 units (**Figure 4.1**). While it made initial sense to use these polygons as the "origin" points of my API matrix requests, knowing the travel times from these neighborhoods is useful only for understanding *absolute* accessibility. Without some type of comparison-point, limiting my requests to just the *asentamientos* would provide me with just an illustration of their accessibility levels **relative to each other** and not give any indication as to whether they enjoy greater or lesser transit accessibility relative to people who live **outside** of the *asentamientos*.
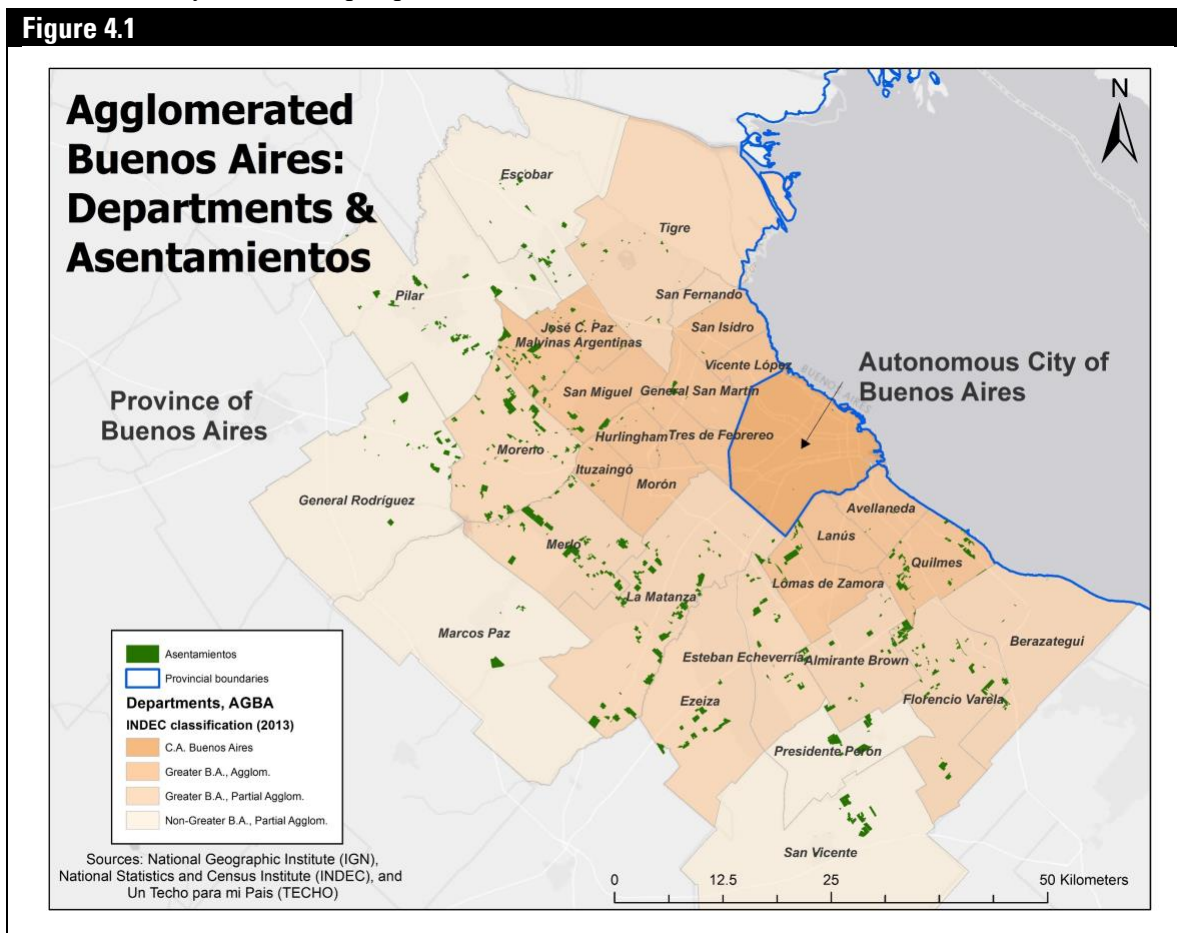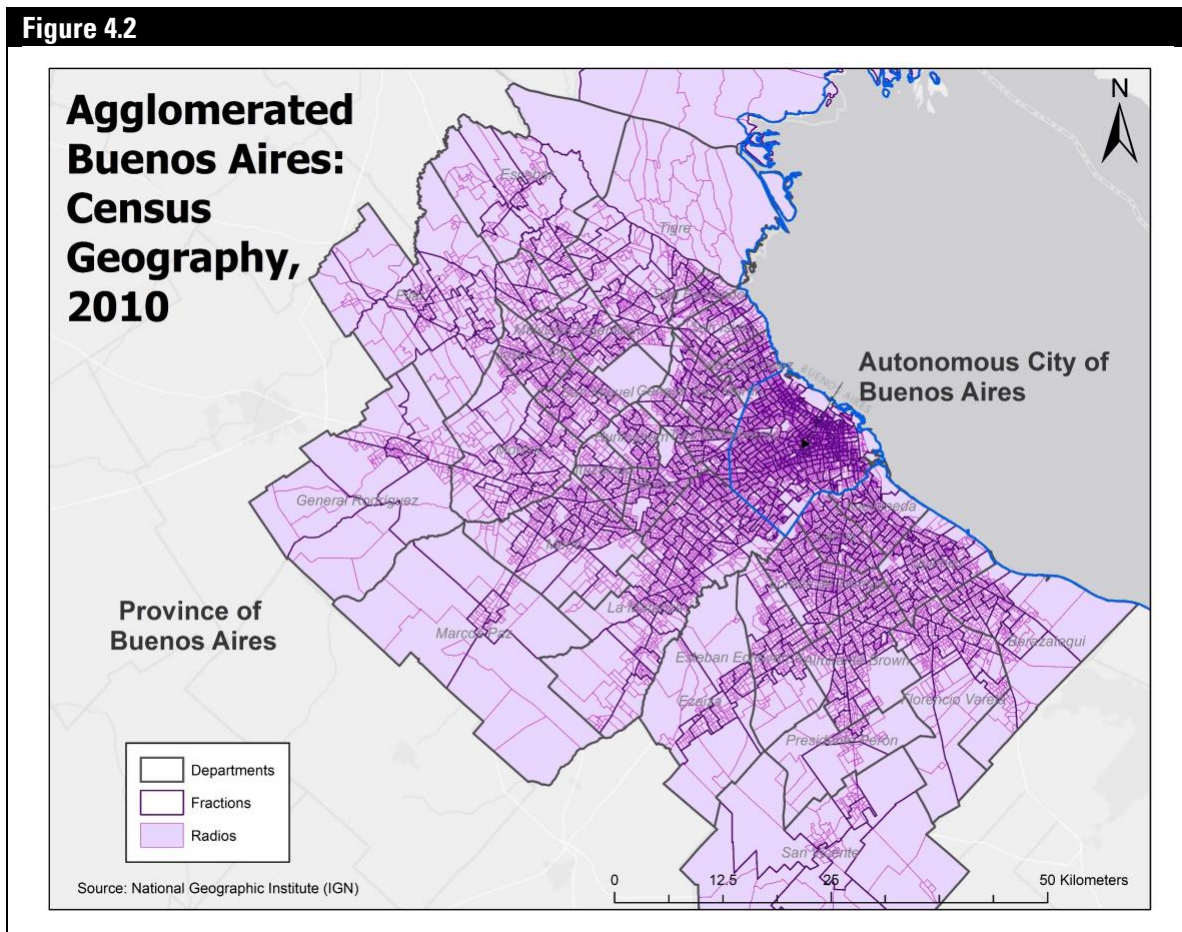

Figure 4.1:

4 http://relevamiento.techo.org.ar/

As a result, I needed a new unit of spatial analysis that permitted the selection of a "control" group of neighborhoods to compare against the *asentamientos*. There is no immediate solution to this problem because the *asentamientos* vary greatly in shape and size and, therefore, do not universally align with any preexisting spatial or geographic administrative units. Nonetheless, I ultimately gravitated towards the census geographies (**Figure 4.2**) created by INDEC as part of the 2010 Argentine national census. The smallest unit of analysis from their survey—known as the *radio*—turned out to be the best option; groups of radios are nested inside of "*fracciones*" (fractions), which are, in turn, further nested inside of the departments.

Figure 4.2

Agglomerated Buenos Aires: Census Geography, 2010

Autonomous City of Buenos Aires

Province of Buenos Aires

Legend:
Departments
Fractions
Radios

Source: National Geographic Institute (IGN)

0   12.5   25   50 Kilometers

Roughly equivalent to the "block" unit within the United States, census *radios* are quite small, often a few city blocks in size. While *asentamientos* rarely overlap perfectly with individual census radios, they are often similar in size (see **Figure 4.3** for an example from La Matanza department), allowing census radios—based on their degree of overlap—to be a nominal stand-in for many of my study sites. Since the entire study area is covered with radios (there are 13,521 within AGBA alone, see **Figure 4.2** above), it is possible to draw a sample of non-overlapping radios to compare against those that contain part of an *asentamiento.* This was the approach I took. Furthermore, INDEC publishes a robust amount of data—through the 2010 census—on demographic and household characteristics at the radio-level, which I can hypothetically compare against my eventual travel time data to see if they produce statistically-significant correlations.[5]

Figure 4.3

La Matanza Department: Asentamiento Centroids

---
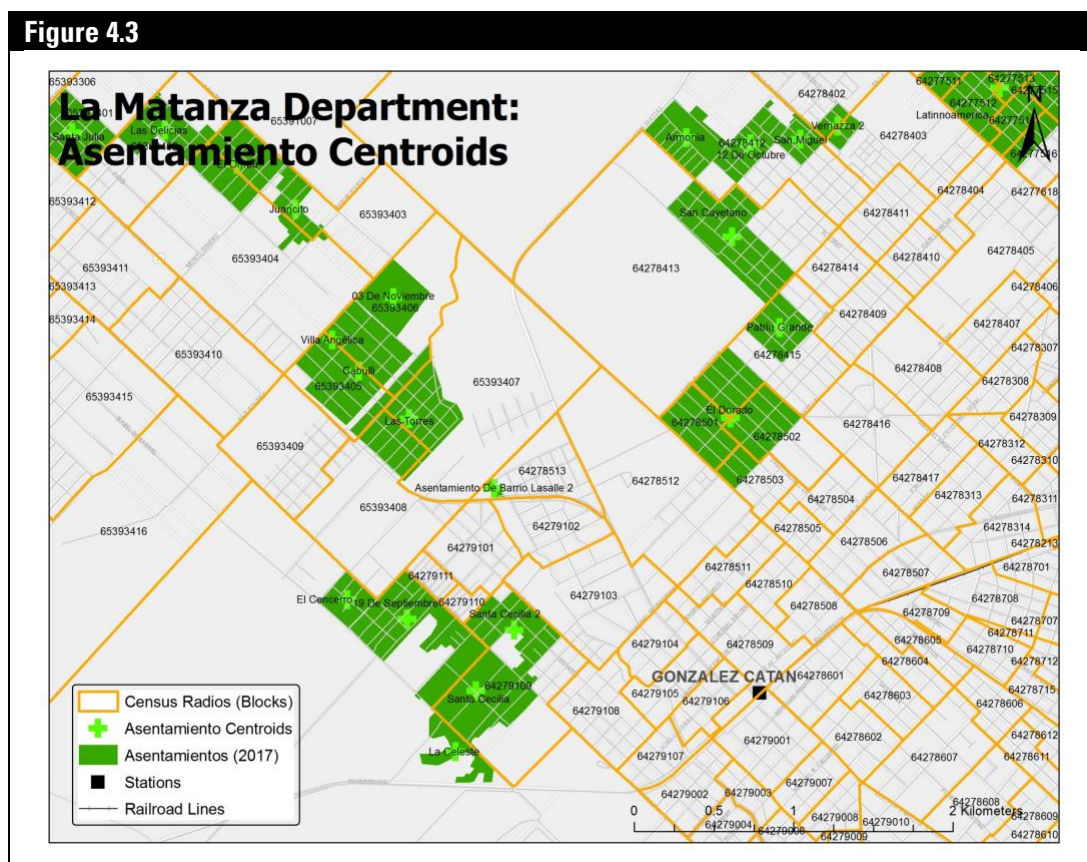
[5] Data include: housing density, land ownership types, roof quality/material, possession of a bathroom, flushing toilets, sewerage, cooking heat type, floor quality/type, clean water availability, computer ownership, freezer ownership, landline ownership, cell phone ownership, literacy, age, employment, education level, age cohorts, gender, national origin, quality of services, construction quality, housing material quality, housing occupation, and housing type (INDEC 2011).

In turn, I used ArcGIS to calculate the degree to which each of those 13,521 radios overlapped with the 687 *asentamiento* polygons. Given the relative dispersion of *asentamientos* across the entirety of the metro area, it is hardly surprising that the average overlap was only 1.98% and that over 94% of all radios registered no overlap at all (12,700 of 13,521). Conversely, nine recorded total-overlap (100%) and 821 saw more than 1% of their territory covered by an *asentamiento.* The map below, **Figure 4.4**, depicts all AGBA census radios, colored by their degree of overlap.

Figure 4.4

Agglomerated Buenos Aires: Radio/Asentamiento Overlap
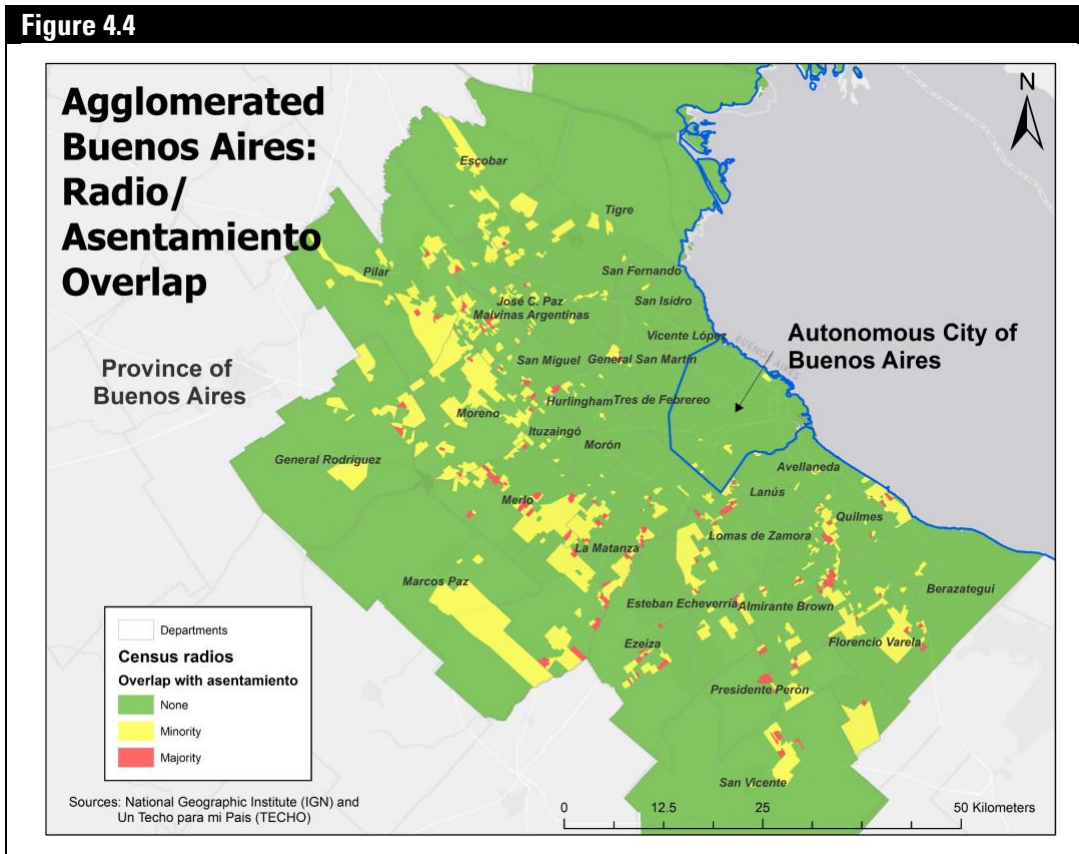
To create "study" and "control" groups for this project, I grouped the population of census radios into (a) those that overlap with an *asentamiento* and (b) those that did not. Furthermore, I also sub-divided the "study" group into "majority" (greater than 50% of territory occupied by an *asentamiento*) and "minority" (between 1% and 50% of territory covered) groups; 206 fell into the former and 615 in the latter. Summary statistics on the overlap calculations, per district and agglomeration level, are shown in **Table 4.5** below; those three case studies that I ultimately selected are highlighted in purple.

With unlimited time, I would have started to make API queries here, requesting travel time information for each of these *radios*. However, as was noted before, standard Google accounts are only allowed 2,500 free origin/destination API requests per day, with only 25 origins or destinations at one given time. Performing travel time requests with all these origins—even with one single destination for each— was not feasible with time and budget constraints, forcing me to narrow down my scope even more.

Department	% of AGBA Population	% of *Asent.* in AGBA.	No. of fam. in *Asent.* (est.)	Radios	Avg. Overlap	Radios, by Overlap			Agglom., Class (2003)
						Maj.	Min.	None	
C.A. Buenos Aires	21.55%	0.00%	0	3,555	0.00%	0	6	3,549	*CABA*
Quilmes	4.33%	7.40%	19,995	557	4.29	16	61	480	
José C. Paz	1.98%	4.48%	6,575	218	6.00	0	35	175	
Malvinas Argentinas	2.40%	3.44%	3,495	281	2.43	2	30	249	
Lomas de Zamora	4.58%	2.93%	11,440	612	2.44	14	19	579	
San Miguel	2.05%	2.93%	7,642	256	4.75	12	16	228	
Ituzaingó	1.25%	2.07%	819	160	0.92	0	9	151	*GABA,*
Hurlingham	1.35%	1.89%	3,011	170	0.48	0	6	164	*Totally*
Avellaneda	2.55%	1.20%	2,305	387	0.24	0	5	382	*Agglom.*
General San Martín	3.08%	1.03%	6,245	435	1.93	8	7	420	
Lanús	3.41%	0.86%	2,200	517	0.78	3	6	508	
Morón	2.39%	0.52%	125	357	0.02	0	2	355	
San Isidro	2.18%	0.52%	320	320	0.15	0	3	317	
Tres de Febrero	2.53%	0.34%	315	377	0.06	0	1	376	
Vicente López	2.00%	0.00%	0	344	0	0	0	344	
Zone 1	57.62%	29.60%	64,487	4,991	1.75	63	200	4,728	
La Matanza	13.20%	11.88%	34,681	1,302	3.09	36	65	1,201	
Moreno	3.36%	11.36%	18,423	389	3.09	4	57	328	
Florencio Varela	3.17%	8.26%	17,925	360	6.74	20	46	294	
Merlo	3.93%	7.92%	19,490	453	6.69	28	43	382	
Almirante Brown	4.11%	3.79%	11,040	484	2.61	10	28	446	*GABA.*
Esteban Echeverría	2.24%	3.27%	13,800	254	2.56	4	16	234	*Partially*
Tigre	2.80%	2.75%	2,920	320	0.71	0	19	320	*Agglom.*
Ezeiza	1.22%	2.07%	10,020	146	8.83	12	14	120	
Berazategui	2.41%	0.52%	460	290	0.37	1	3	286	
San Fernando	1.21%	0.34%	520	164	0.59	1	1	162	
Zone 2	37.64%	52.15%	129,279	4,162	3.44	116	292	3,754	
Greater Buenos Aires	95.26%	81.75%	193,766	12,708	1.82	179	498	12,031	
Pilar	1.73	6.02%	13,170	279	3.45	6	40	233	
Escobar	1.32	3.79%	7,980	194	3.69	3	28	163	
General Rodríguez	0.65	3.61%	5,178	113	4.86	3	29	81	*Non-GABA,*
San Vicente	0.33	2.58%	4,685	83	0.56	4	11	68	*Partially*
Presidente Perón	0.45	1.55%	6,780	73	13.50	11	6	56	*Agglom.*
Marcos Paz	0.32	0.69%	2,340	58	0.62	0	3	55	
Zone 3	4.80%	18.24%	40,133	800	4.64	27	117	656	
Agglom. Buenos Aires	100.0%	100.0%	233,899	13,508	1.98	206	615	12,687	

Table 4.5

While there are *asentamientos* found across AGBA's suburban periphery, I selected just three of these departments to serve as representative case studies, thus reducing the time to complete my API requests. Seeking to produce results that applied to the largest quantity of people possible, I
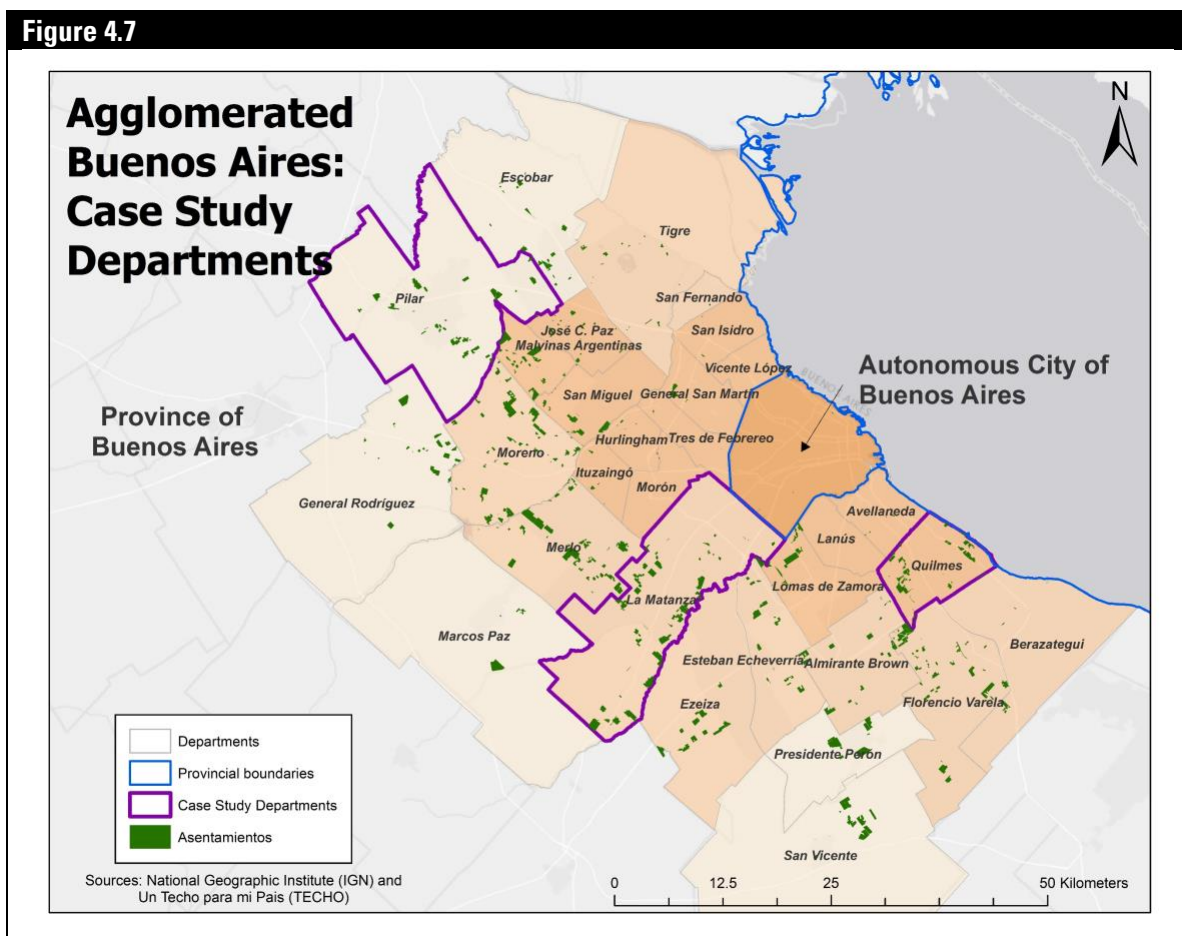
first prioritized departments by the number of families residing in their respective *asentamientos*, eyeing those with the largest values (**Table 4.5**). When these raw data are mapped, however, the districts the most *asentados* are all on AMBA's western side (La Matanza, Moreno, Merlo) and off to the southeast (Quilmes and Florencio Varela)—see **Figure 4.6** below. This distribution coheres to general socio-economic patterns in AGBA and their north/south gradations.

Figure 4.6

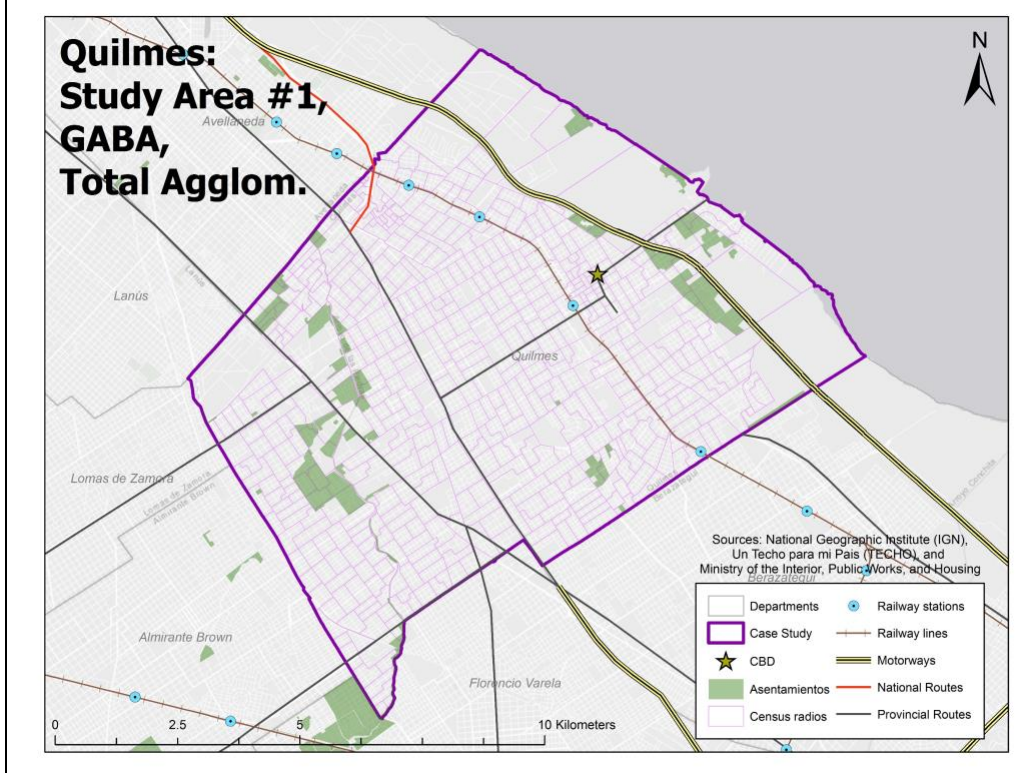Agglomerated Buenos Aires: Asentamiento Families, by Dept.
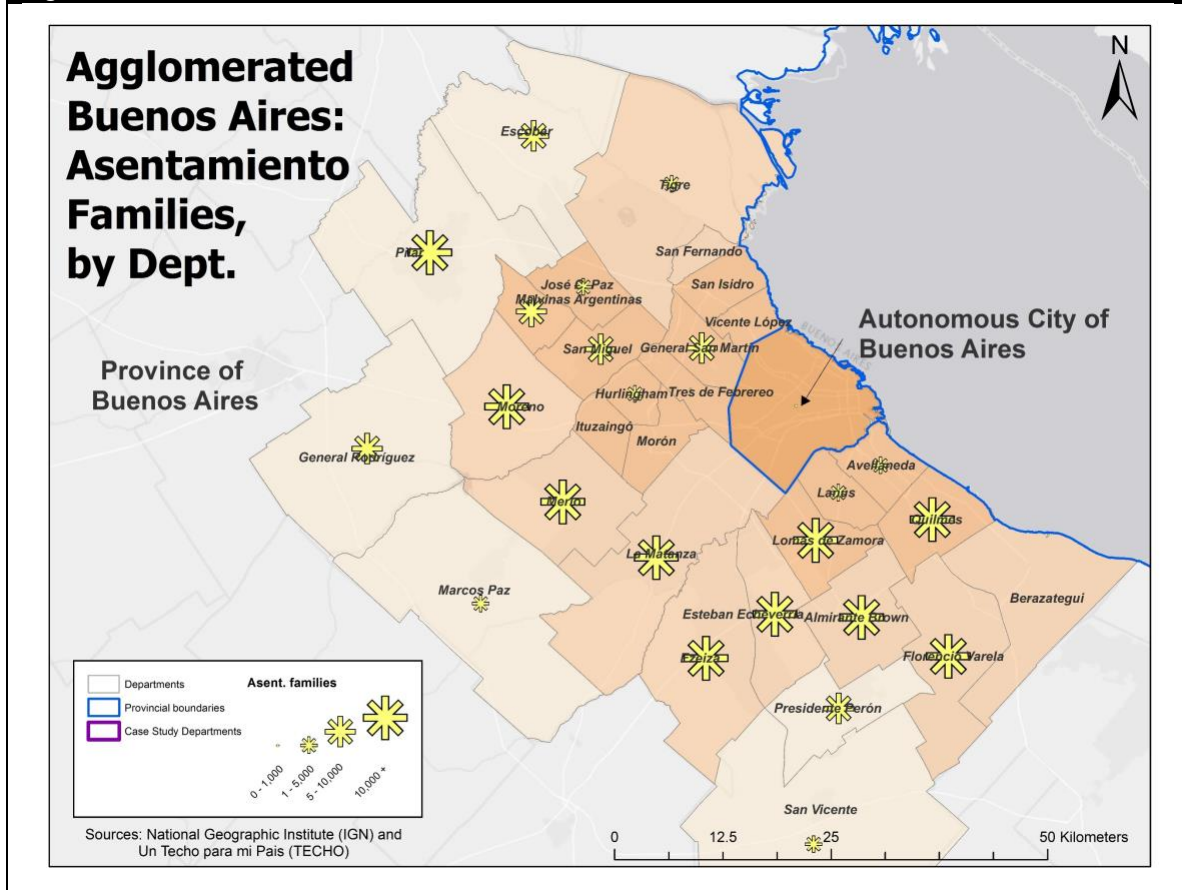
While selecting these five would account for accessibility conditions among a large proportion of the conurbation's total population of *asentados*, it could miss trends in some of the other districts, especially those further removed from CABA's core. As a result, I decided to sub-set the departments once again, using INDEC's urbanization hierarchy (considering history in GABA and degree of agglomeration) mentioned at the beginning of the paper. Considering that each of these categories represents departments of varying degrees of suburbanization—and, therefore, exhibiting a wide range of densities, infill, age of development, and dispersion of roads and transit services—iit would provide a snapshot of accessibility as it relates to each of AGBA's different zones. Chosen primarily to reduce API requests, these differentiations would gone to greatly affect my results.

With this conceptualization in mind, I selected the department from each "agglomeration zone" with *the largest population of asentamiento families* (a statistic tabulated by TECHO during their survey): Quilmes (GABA + total agglomeration; 19,995 families in 43 communities), La Matanza (GABA + partial agglomeration; 34,681 in 69), and Pilar (non-GABA + partial agglomeration; 13,170 in 35). Shown on **Figure 4.7**, these departments are spatially distinct, each representing districts of varying age and development (Quilmes the most established, Pilar the youngest, and La Matanza in between) as well as Buenos Aires' socioeconomically distinct edges: the historically wealthy north (Pilar), the working-class west (La Matanza), and the impoverished south (Quilmes). Maps of these districts, and their asentamientos (shown in green), are provided in **Figures 4.7a-c**. Nonetheless, the number of census radios comprising each of these districts (**Table 4.8**) was *still* massive. I had to perform one last sub-sampling to reduce the number of "origins" to something compatible with the API's strict quotas. Maps
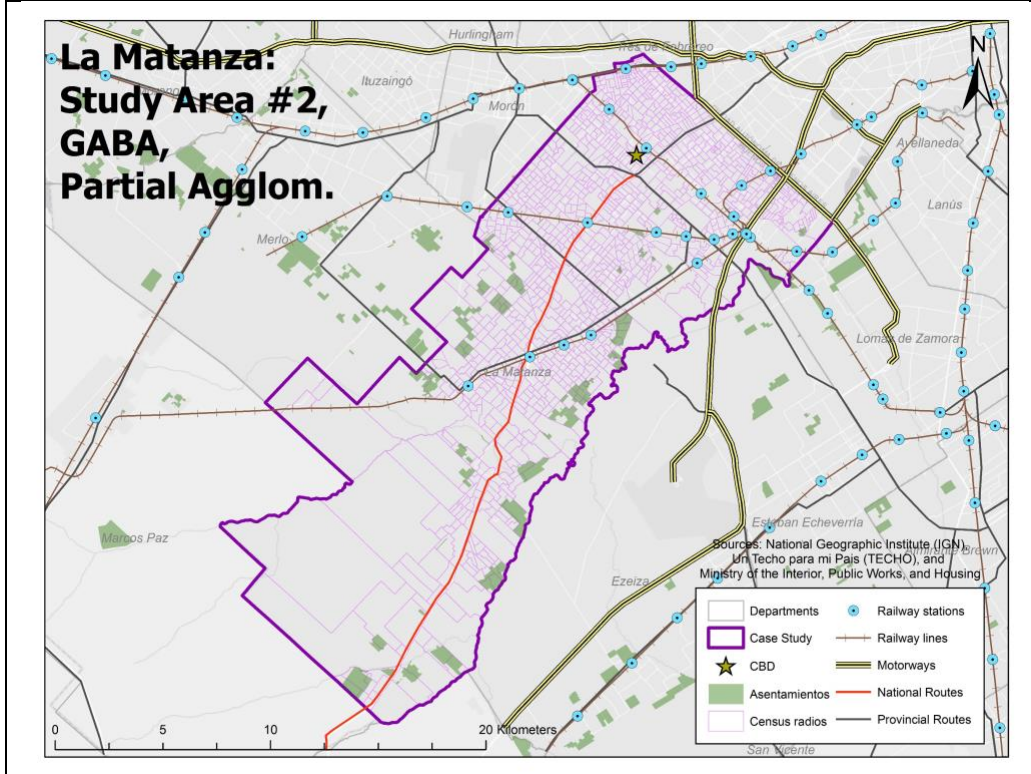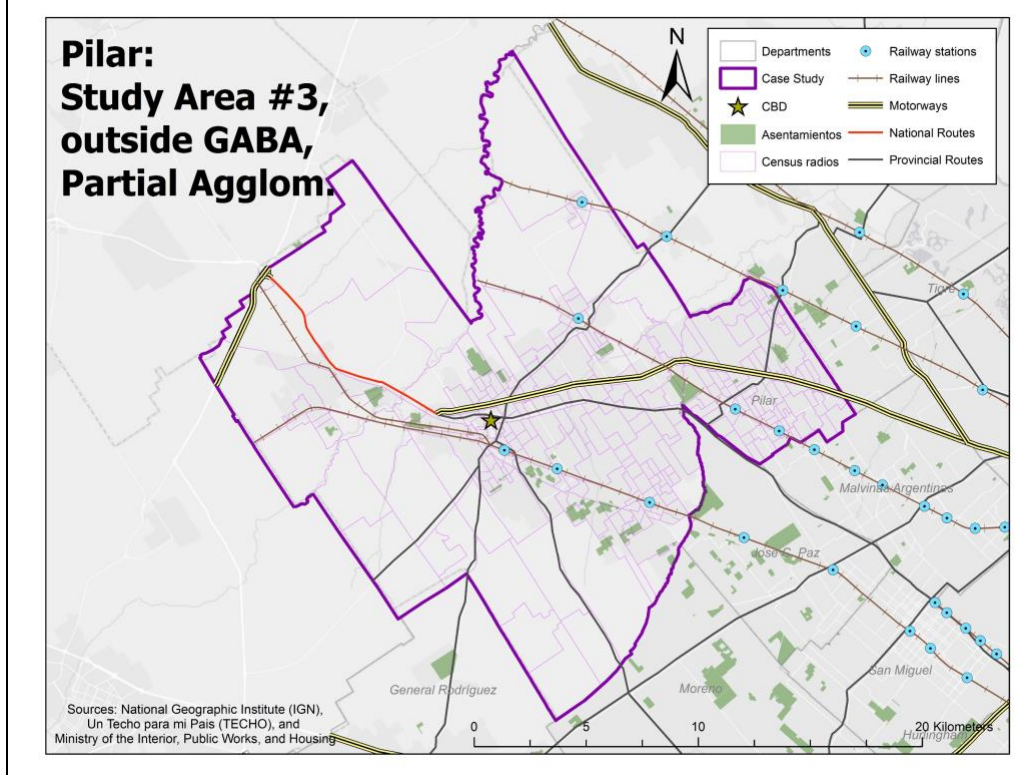

Figure 4.7

Figure 4.7a

Quilmes: Study Area #1, GABA, Total Agglom.

Figure 4.7b

La Matanza: Study Area #2, GABA, Partial Agglom.

Figure 4.7c

Pilar:
Study Area #3,
outside GABA,
Partial Agglom.

Sources: National Geographic Institute (IGN),
Un Techo para mi Pais (TECHO), and
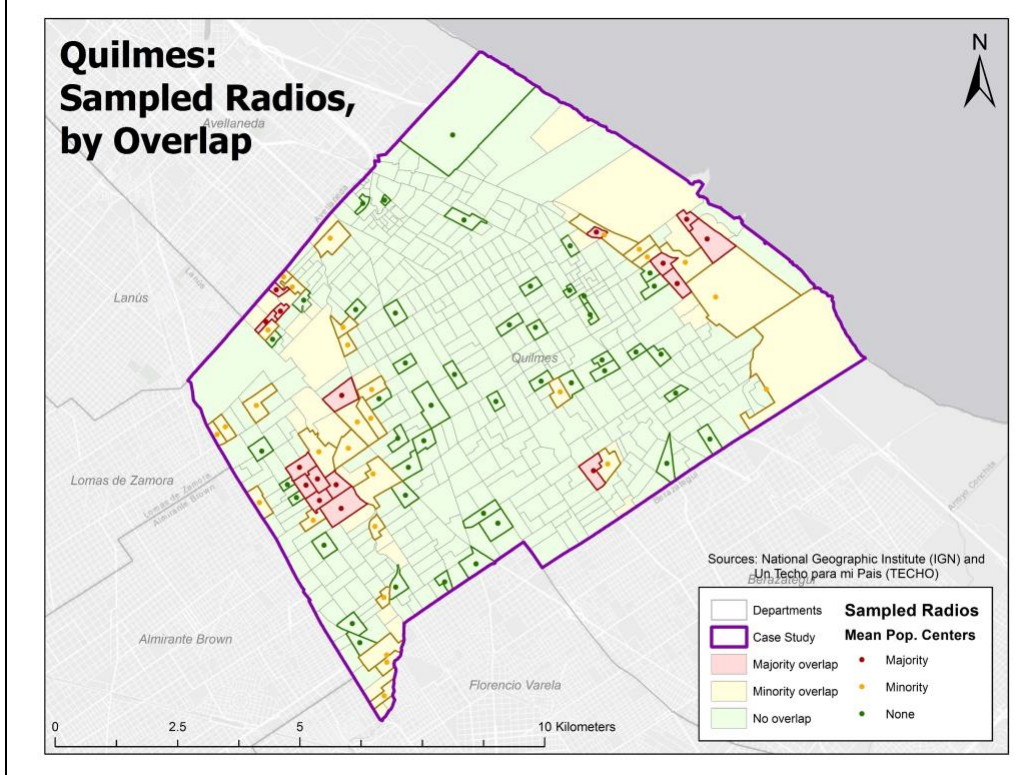Ministry of the Interior, Public Works, and Housing

To complete this task, I performed a series of stratified random samples on the census radios belonging to each case study. For all three districts, I began by dividing their radios into those classified as "majority overlap", "minority overlap", and "no overlap"; **Table 4.8** summarizes the number, and percentage, in each category.

Table 4.8	Radios			Overlap			
District	Radios, Total	Average Overlap	Avg. Pop., per Radio	Majority	Minority	Any	None
Quilmes	557	4.29%	1,046.6	16 (2.9%)	61 (11.0%)	77 (13.8%)	480 (86.2%)
La Matanza	1,302	3.09%	1,363.9	36 (2.8%)	65 (5.0%)	101 (7.8%)	1,201 (2.2%)
Pilar	279	3.45%	1,072.0	6 (2.2%)	40 (14.3%)	46 (16.5%)	233 (83.5%)
Totals	**2,139**	**3.45%**	**1,234.1**	**58 (2.7%)**	**166 (7.8%)**	**224 (10.5%)**	**1,914 (89.5%)**

Serving as "study" groups, I first randomly selected thirty radios from all three districts' populations of "majority overlap" and "minority overlap" radios; in the cases where there were not thirty available, I selected all that were available. Once these thirty were identified, I then added together the number of "majority" and "minority" radios, per case study, into a secondary categorization called "any overlap." I then drew a random sample from the "no overlap" radios belonging to each district, selecting a sample sized in correspondence with the number of radios in each district's "any overlap" category. I did this because the pool of "no overlap" radios was always much larger than those with overlap, and I wanted to ensure I was drawing a proportionally-representative sample of "no overlap" radios to match those drawn for the other two categories. Those numbers are reported in **Table 4.9**, while maps of the sampled radios are depicted by **Figures 4.10a-c**.

Table 4.9		Sample Selection				
District		**Majority**	**Minority**	**Any**	**None**	**Total**
Quilmes	Sample	16 (100.0%)	30 (49.2%)	46 (59.7%)	46 (9.6%)	92 (16.5%)
	Total	16	61	77	480	557
La Matanza	Sample	30 (83.3%)	30 (46.2%)	60 (59.4%)	60 (5.0%)	120 (9.2%)
	Total	36	65	101	1,201	1,302
Pilar	Sample	6 (100.0%)	30 (75.0%)	36 (78.3%)	36 (15.5%)	72 (25.8%)
	Total	6	40	46	233	279
All 3	Sample	52 (89.7%)	90 (54.2%)	142 (63.4%)	142 (7.4%)	284 (13.0%)
	Total	58	166	224	1,914	2,183

Figure 4.10a

Quilmes:
Sampled Radios,
by Overlap

Figure 4.10b

La Matanza:
Sampled Radios,
by Overlap

Figure 4.10c

Pilar: Sampled Radios, by Overlap

Sources: National Geographic Institute (IGN) and Un Techo para mi Pais (TECHO)

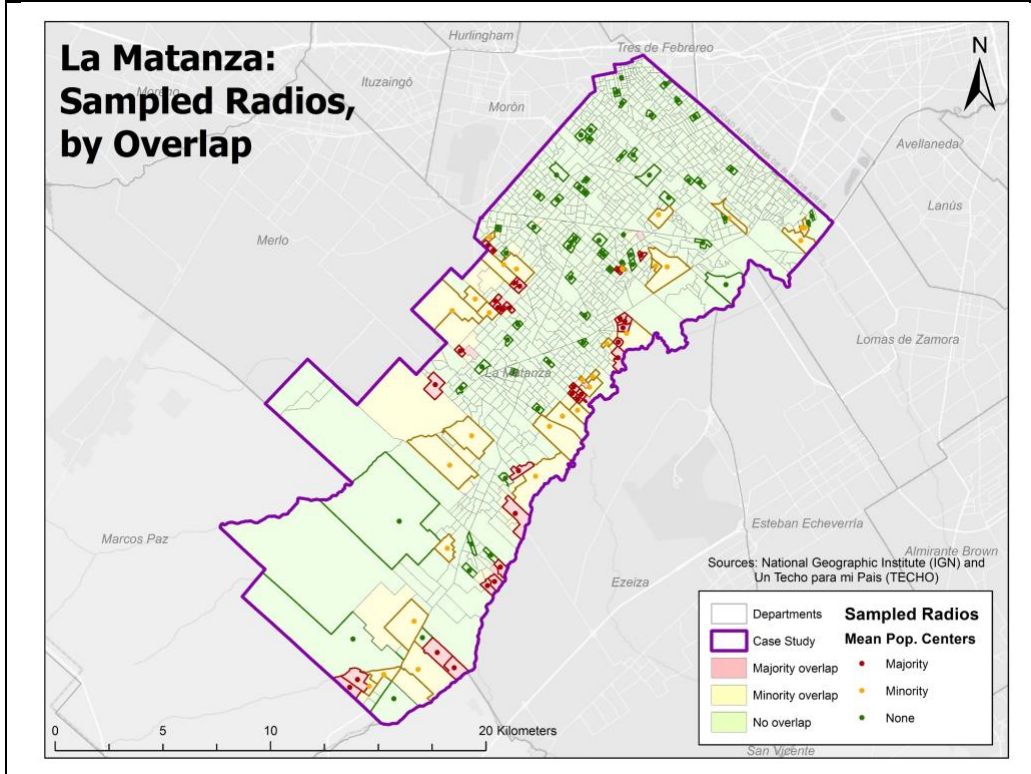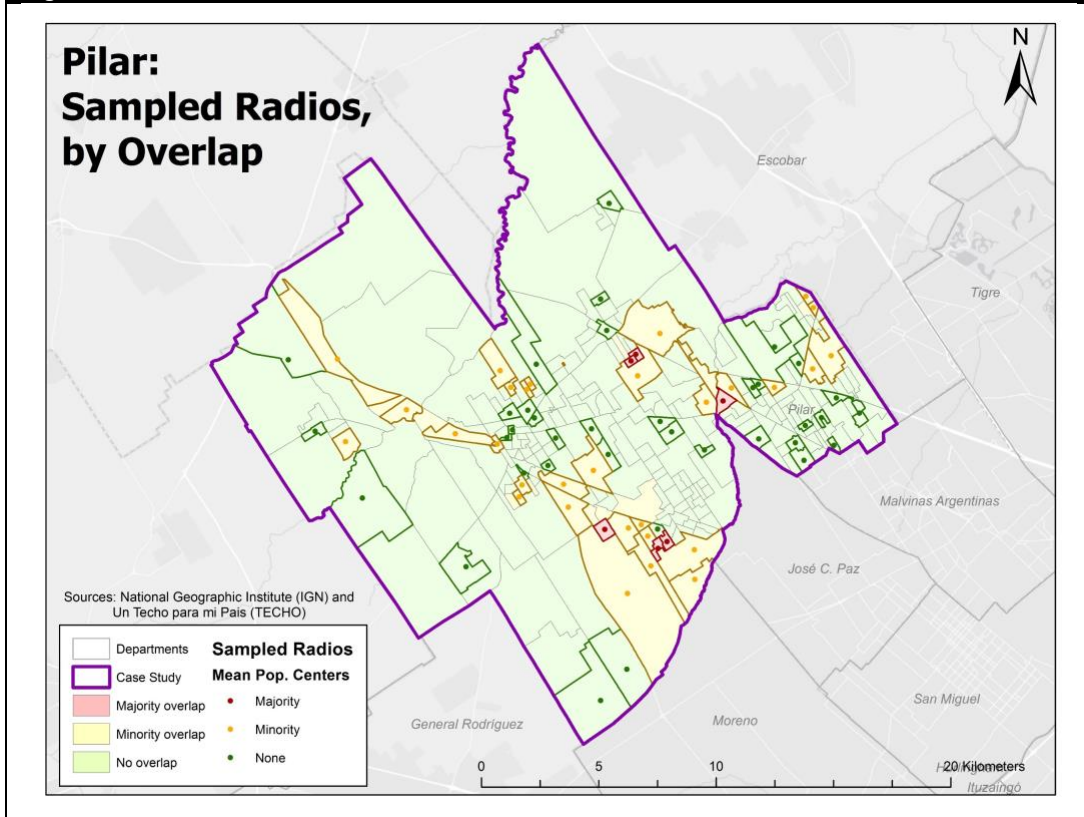| Departments | **Sampled Radios** |
Case Study	**Mean Pop. Centers**
Majority overlap	Majority
Minority overlap	Minority
No overlap	None

Some census statistics and ENMODO results for the three districts are shown in **Table 4.11,** with a focus on variables traditionally associated with the *asentamientos*. La Matanza has the largest overall population and is actually the most populous department in AGBA outside of CABA. Cells highlighted in pink represent the district with the "poorest" score for a given category. Note how Pilar, for instance, scores the worst in many development indicators (piped water, sewerage, basic needs unmet, illiteracy, bus trips, and school attendance) yet also has the highest rates for university attendance, car ownership, trips by auto, and homes in gated communities—inequities are discerned at even an aggregate scale. Also note how, paradoxically, Pilar's *asentamientos* scored the worst is nearly all access metrics *except public transit access.*

17

Table 4.11					
Source	Variable	Quilmes	La Matanza	Pilar	AMBA
INDEC, 2010 census	Total population	520,552 (4.01%)	1,398,891 (10.77%)	266,564 (2.53%)	12,985,885 (100.00%)
	People per radio	1046.576	1363.914	1072.064	1005.957
	Homes per radio	327.14	343.55	335.07	345.28
	% homes with *"basic needs unmet"*	8.82	11.37	11.77	9.10
	% homes, owner does not own land	5.42	4.94	4.42	4.90
	% homes, public network sewage	61.57	47.15	18.58	56.03
	% homes, piped water inside home	92.37	85.47	79.65	90.99
	% people, illiterate	5.45	5.93	6.71	5.05
	% people, never attended school	1.97	2.52	2.66	1.78
	% people, attended university	10.67	6.31	10.02	14.00
	% people, born in Argentina	92.73	90.75	93.11	91.50
	% homes, "insufficient" services	25.71	36.02	65.17	30.38
	% homes, "insufficient" const. quality	14.39	19.66	22.72	12.38
ENMODO respondents, 2009-10	Trip generation rate	1.43	1.48	1.51	1.52
	Average people per home	3.07	3.49	3.77	3.23
	Homes, average income quintile	2.86	2.61	2.45	2.87
	% homes, *villa de emergencia*	4.0	2.3	2.7	2.3
	% homes, gated community/country	0.1	0.1	8.4	0.4
	% homes, with automobile	36.1	35.1	38.3	36.9
	% people, employed as domestic help	2.2	3.0	3.2	2.6
	% people, attending school	9.0	9.3	10.0	9.3
	% students, public school	67.9	73.1	69.0	70.4
	% students, private school	32.1	26.9	31.0	29.6
	% people, actively working	39.6	39.1	37.5	40.9
	% people, working in private sector	36.2	37.0	34.0	37.7
	% people, working in public sector	4.9	3.8	4.0	4.5
	% trips, made 5x weekly	57.8	24.1	58.9	53.3
	Most common hour of trip departure	12:00 PM			
	% trips, on bus	43.4	45.9	23.4	39.4
	% trips, on rail	3.8	2.9	7.9	7.1
	% trips, in private auto	19.1	15.6	27.4	17.9
	% trips, on foot	23.6	25.7	28.9	23.5
TECHO, 2013	% asent., irregular electricity access	69.0	71.6	50.0	62.4
	% asent., water from public network	11.9	16.4	0.0	4.8
	% asent., firewood/coal for heat	9.5	14.9	26.7	12.8
	% asent., ambulance always responds	71.4	25.4	60.0	45.3
	% asent., floods with every rain	42.9	31.3	53.3	35.2
	% asent., hospital beyond 5km	35.7	59.7	63.3	54.0
	% asent., public transit inside neigh.	23.8	13.4	36.7	19.8
	% asent., major prob., service access	33.3	34.3	20.0	29.4
	% asent., public sewerage	7.1	1.5	0.0	3.3

With the number of "origins" down to just 284, I was nearly ready to move on to the next steps collecting destinations and setting my parameters. The only other step was to convert these census radios into points (with latitude and longitude coordinates) that could be fed into the API interface. My initial idea was to use just the centroid of each polygon; however, in the case of the spatially-expansive radios (many of those in Pilar and La Matanza), I ran the risk of having the center point of the polygon located far from the actual population center of the radio.

I crafted my own solution to this problem. I first visited the website of the Socioeconomic Data and Applications Center (SEDAC) at Columbia University.[6] They maintain a repository of geospatial data on world population distributions. There, I downloaded a raster dataset of the estimated distribution of people throughout Argentina; using their in-house model, SEDAC uses interpolation techniques ("random forest classification and regression") to estimate "likely residence locations at a 100-meter scale (Rodriguez, et al. 2017, pp. 36-37)." Their results for AGBA for 2015 are shown on **Figure 4.12,** with the three case studies outlined in purple.

Figure 4.12

___

[6] http://sedac.ciesin.columbia.edu/data/collection/gpw-v3
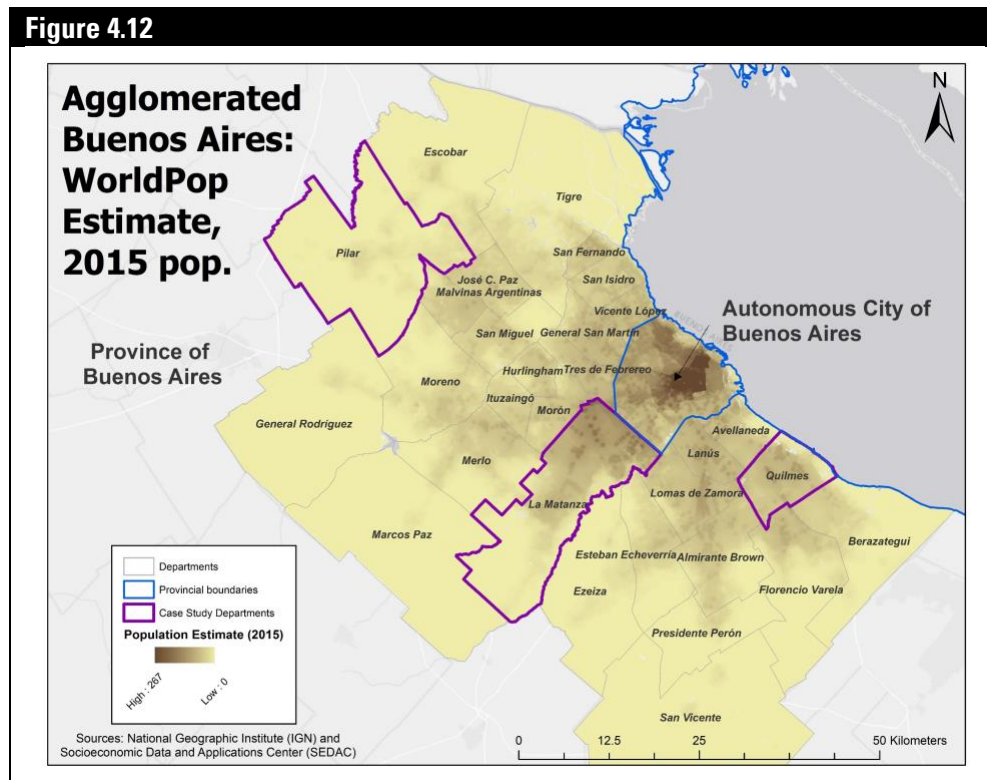
From this raster, I extracted the areas corresponding to each case study and then converted the raster into a point field—one point was placed at the center of each 100-meter raster cell, with that point assigned the value of that raster grid cell. I then used a spatial join to identify the specific census radio to which each point belongs and then calculated the approximate mean center of population for each individual census radio. The result is a series of points corresponding to the weighted mean population centers of each census radio, shown on **Map 4.13** below. These points, I assume, will ensure that the API's "origins" correspond to actual human locations.

The only other step required was to calculate the X/Y coordinate for each point, an easy process in ArcGIS. To make sure these values were compatible with the Distance Matrix API, I created a special column in the attribute table called "YX coordinate" where the coordinate was stored in the proper text-string format ("Longitude+Latitude", ex. "-34.7098979561128+-58.2342871680724"). I then exported, for each district, its list of points as a CSV, which could then be easily imported into R to serve as the "origins" input for "gmapsdistance" API requests.

Figure 4.14
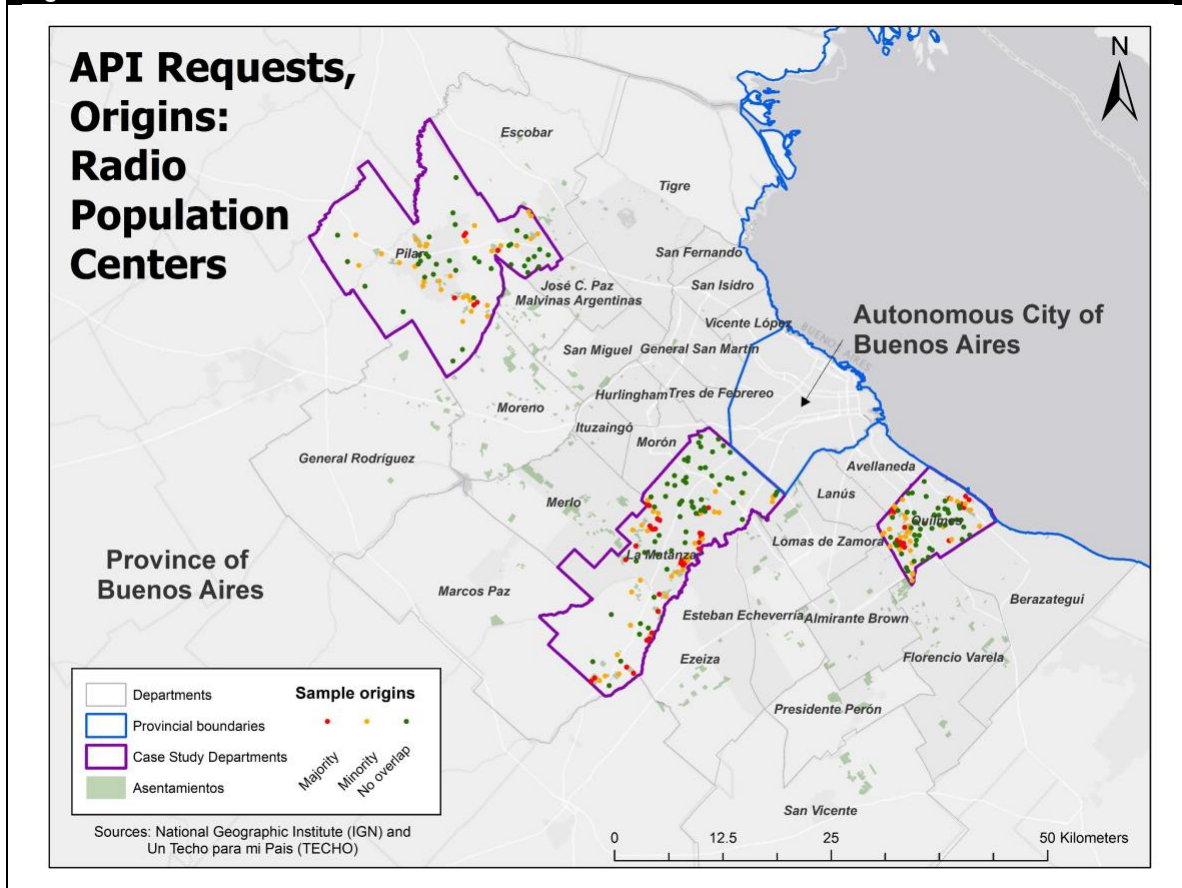
Part IV(c), API Parameters, Destinations

Acquiring and preparing the coordinates of my destinations was more straightforward. Seeking to heed the advice of Handy and Neimeier, I sought datasets representing the activity sites of importance to those people living in Buenos Aires' *asentamientos*. To do this, I returned to the qualitative and quantitative sources discussed earlier: TECHO's survey data, INHABITAT's interviews, and ENMODO's survey statistics. Unfortunately, none of these sources directly asked *asentados* transit-specific questions about travel times, trip types, or common destinations. What we do know, however, is that buses are the most commonly-reported mode of travel, most have transit within a short distance of their settlement, schools, clinics, and jobs seem to be relatively close, and hospitals and recreation centers are further in distance. Whether transit is favored for certain trip-types or destinations, however, remains a mystery, with nothing to prioritize any one feature for inclusion with the API.

Since the purpose of this project is ultimately comparative (looking at differences between the *asentados* and their "legal" neighbors), I used the ENMODO results as an initial guide, since they summarize trips by *all* travelers. The most common trip destination was work (37%), followed by education/accompanying someone to school, (34%), shopping (7%), personal business (5%), health (4%), family (3%), and recreation (2%) (ST 2011). While searching for geospatial datasets depicting these activities, I looked only for those activity sites with a corresponding reliable, region-wide location point dataset. If the location data for a given activity site was only available for a section of the study area or was inconsistently depicted across the whole region, it was discarded. This proved difficult given the little geospatial data made public—for whichever reason—by national, provincial, and municipal governments (the City of Buenos Aires, for instance, has a well-maintained repository of GIS layers but are wholly inapplicable to the suburban scope of this project). Ultimately, I located comprehensive data for only schools and healthcare sites and, in turn, had to create proxies for employment and shopping/commerce. My findings are summarized in **Table 4.15**.

Table 4.15						
Destination Type	**Sub-Type**	**AGBA**	**Quilmes**	**La Matanza**	**Pilar**	**All 3 Cases**
Central Business Districts	C.A. Buenos Aires			1		
	Department C.B.D.	30	1	1	1	**3**
	Proximate C.B.D.	21	6	8	9	**23**
Railroad Stations		240	19	42	23	**84**
Healthcare	Public Urgent Care	15	7	9	6	**22**
	Diagnostic/Treatment	1,076	70	81	51	**202**
	Hospitals	80	11	19	10	**40**
Public Schools	Kindergartens	1,602	120	148	66	**334**
	Primary Schools	1,973	109	203	77	**389**
	Secondary Schools	1,674	78	199	74	**351**

Part IV(c.1), Destinations: Central Business Districts

The Argentine government, lamentably, does not publish geographic information on employment locations. However, as the most common trip type, I sought to include a destination to serve as a proxy for these to ensure their inclusion. After determining that the types of employment associated with low-income, *asentamiento*-dwellers (e.g. carpenters, domestic workers, street vendors, etc.) were not location-specific enough to permit simulation, I elected to instead use central business districts. While the evidence shows that employment in AGBA is more dispersed than ever, I figured CBD's, given their relatively high densities and their general roles as centers of commerce and local government, would be reasonable proxies for some of those job- and shopping-centric trips made by *asentados*. Obviously, this assumption does not account for all jobs, especially those in the service sector that could be more dispersed across the terrain and industrial jobs along AGBA's new outer highways (relocated from southern CABA, where they had been concentrated for much of the city's history).

In fact, I performed three CBD-related queries: to Buenos Aires' central business district (in CABA), to each of the study departments' individual CBD's, and to the nearest departmental CBD (including those departments excluded from my sample) closes. This variation captured those 8.7% of *asentados* (according to Cravino, et al. 2008) who work in the federal district, the 17.7% who worked in separate departments, and the 48.1% who worked in their home department. Including the neighboring departments also accounted for people living in the far periphery of any of the three case study areas who may, in fact, live closer to a neighboring CBD (and its corresponding commercial/employment opportunities) than their own. Understanding the travel times to known employment centers can not only show whether *asentados* have a more difficult journey to work but also—should those times by significantly longer compared to people from formal neighborhoods—why so many of their fellow residents need to work in or nearby their home neighborhood (26% of respondents).

Data on central business districts was downloaded from the geographic data portal of the Ministry of the Interior, Public Works, and Housing.[7] More specifically, I used a layer called "Localities' head of local government" ("Localidades cabecera de gobierno local segun tipo de gobierno") based on locations identified during Argentina's 2010 national census.[8] The data were most recently updated on 19 January 2018. The layer, a point shapefile, was uploaded into ArcGIS, where I used the program's database query tools to select, the points individually pertaining to the three case studies: Quilmes, La Matanza, and Pilar. I then used the program's spatial query tools to select those departments that bordered each these three: nine for Pilar, eight for La Matanza, and six for Quilmes. Once identified, I selected the CBD that corresponded to each. For each of my three case studies, I exported (1) a point file that included its own CBD and (2) a file for those identified as its neighbors'. Lastly, I exported these points as separate files. As for CABA, I created a custom point file with a single feature placed at the *Obelisco,* a major
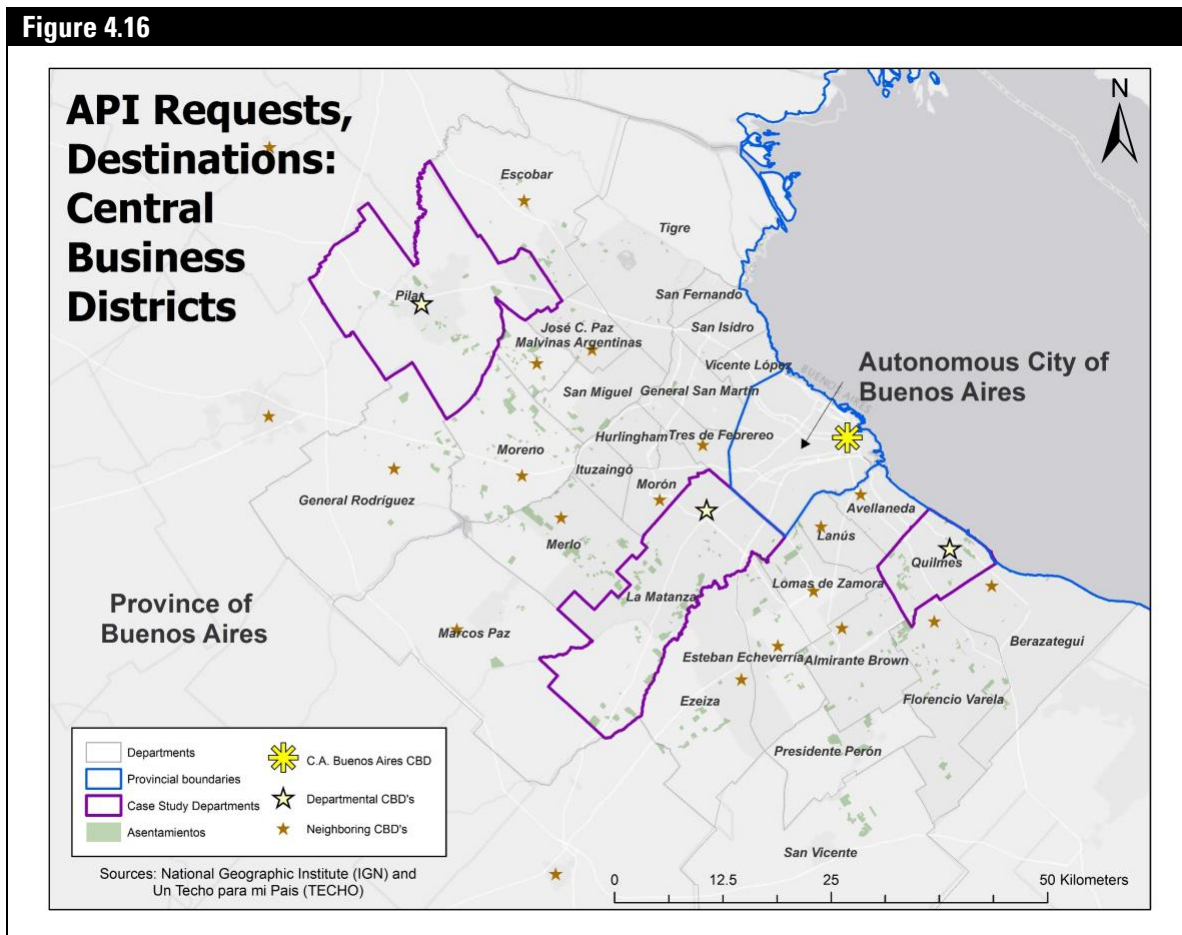
[7] http://sig.planificacion.gob.ar/layers/ultimas
[8] http://sig.planificacion.gob.ar/layers/detalle_capa/awagne_cabecera_gob_local/

landmark along Buenos Aires' central 9 de Julio Avenue widely considered the city's center point. These points are shown on **Figure 4.16**.

Figure 4.16

Part IV(c.2), Destinations: Railroad Stations

The second destination type was the railroad station; while railroad usage has been on the decline for decades in metropolitan Buenos Aires, it still serves as the primary mode of travel for trips between CABA and the departments in the province (56%) and, as a mode, is frequented more by people in the lowest two income quintiles than those at the top (ST 2011). Furthermore, the areas around train stations, given the history of Buenos Aires, are some of the most densely populated pockets outside of CABA and, as a result, serve as local centers of employment and residence (Blanco 2014). While Cravino, et al. (2018) found only 5% of the surveyed *asentados* chose to ride the train, this was because the settlements being studied were all far from their nearest train station anyway. This last characteristic, in fact, speaks to the importance of studying access to rail: longer relative travel times will showcase the *asentamientos* physical isolation relative to existing infrastructure and highlight another way that *asentados* are inherently disadvantaged when looking for employment, especially low-skill positions commonly found in the urban core.

Just as with the central business districts, I acquired my data on railroad station locations from the geographic data portal of the Ministry of the Interior, Public Works, and Housing. The specific layer was called "Railroad Stations" ("Estaciones de Ferrocarril") and was last updated on 28 December 2016.[9] Another point shapefile, I uploaded it to ArcGIS, displaying all 240 stations. Since the goal of this API request, compared with those for the CBDs, is to identify the *nearest* stop to each origin point, the ideal situation would have been to upload all 240 as "destinations", calculate travel time between each origin and all stations, and then select the minimum value. This, however, requires a large quantity of API requests, so I invented an ad-hoc technique to cut down on the number of queries **(a process I also employed for schools and hospitals)**.

I began by assuming that most travelers, by the very nature of accessibility and human travel, would probably only use a station within a reasonable distance of their home; ENMODO, in fact, tell us that most trips—whether for work, school, and health—are within the same or adjoining departments, 80%, 91%, and 85% respectively, suggesting people are not likely to go far out of their way to complete these tasks. As such, I used ArcGIS' spatial proximity tools to determine which train stations, of the 240, were one of the ten closest, using Euclidean distance, to each of the "origins" for each district.[10] I assumed it unlikely that a train station outside of the ten "closest", in straight-line distance, to my study sites was not likely going to be its closest station when considering transit-based network-distances and times.
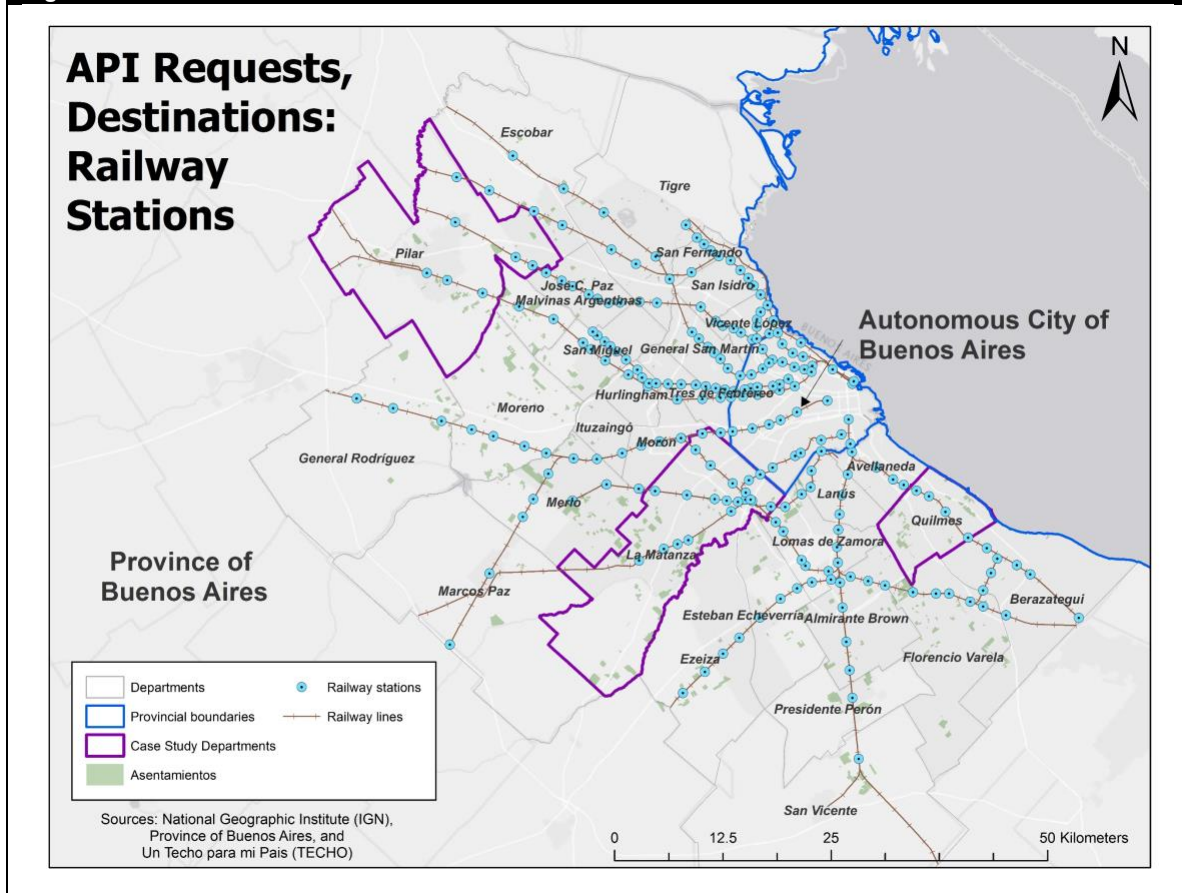
For each of the three study districts, I created a separate point file of just those train stations that qualified under this scheme; it reduced the number of "destinations" to just 19 for Quilmes, 23 for Pilar, and 42 for La Matanza. In each case, some of these stations (if not *most* of them) were geographically outside of the bounds of the home department. Nevertheless, I collected the latitude and longitude coordinates of each station, formatted those coordinates for compatibility with the API engine ("LATITUDE+LONGITUDE"), and exported the values as a "csv" file whose data could be easily read into R. A map of the rail stations are shown below, **Figure 2.17**.

---

[9] http://sig.planificacion.gob.ar/layers/detalle_capa/mrapis_est_ferrocarril_final/
[10] "Near"

Figure 4.16

API Requests, Destinations: Railway Stations

Autonomous City of Buenos Aires

Province of Buenos Aires

Legend:
- Departments
- Provincial boundaries
- Case Study Departments
- Asentamientos
- Railway stations
- Railway lines

Sources: National Geographic Institute (IGN), Province of Buenos Aires, and Un Techo para mi Pais (TECHO)

0   12.5   25   50 Kilometers

Part IV(c.3), Destinations: Public Schools

While prior studies suggest that most *asentamientos* have schools nearby, I included them anyway, primarily because data were easy to find and school-trips a large share of the modal split. The Province of Buenos Aires, within its open data portal, publishes a spreadsheet of all the schools in its territory.[11] Officially titled "Educative Establishments" (Establecimientos educativos), it has attributes for the latitude/longitude coordinates of each school, its level, and whether it is private or public.[12] Downloadable as a comma separated values (csv) file, these data's coordinates were easily displayed as point files within ArcGIS. This file was last updated on 21 February 2018.

A file of 22,290 schools, I had to narrow down the pool of destinations for the API. I first sub-selected the public schools ("estatales"). Aside from colloquial advice from Argentine contacts that *asentados* were more likely to attend to public schools (they are free in Argentina), the results from ENMODO show public schools as not only favored by the population at large (71%) but also by students from all five income quintiles. Furthermore, the lowest quintile (the most likely to

---

[11] http://catalogo.datos.gba.gob.ar/home

[12] http://catalogo.datos.gba.gob.ar/dataviews/245163/establecimientos-educativos/

contain *asentados*) showed a remarkable 84% of its students as public-school attendees (ST 2011). This reduced the pool of schools to just 16,035.

To whittle the total even further, I selected, from the public schools, just three types: kindergartens ("jardin de niños"), primary schools ("primarios"), and secondary schools ("secondarios"). I based this decision on the INHABITAT results showing that most *asentados* had little more than a primary education: 71.9% completed any primary and 23.3% for secondary. Less than 2% had made it beyond (ST 2011). While it is certainly possible that tertiary- or university-level schooling was not pursued for reasons of access, the total number of potential students is certainly smaller than those at other lower levels. This narrowed the number schools to: 2,671 (kindergartens), 4,262 (primary), and 2,729 (secondary).

Lastly, I selected, using the same Euclidean distance technique as before, just those schools that were one of the ten-closest to any one of my "origins"; I replicated this task for all three study areas, reducing the pool significantly. This produced a total of nine files: the list of nearest schools in all three categories for all three districts. Like before, I took the initial step, before exportation, to tabulate the coordinates and re-format them for the API. A map of schools is shown in **Figure 4.17** below.

Figure 4.17
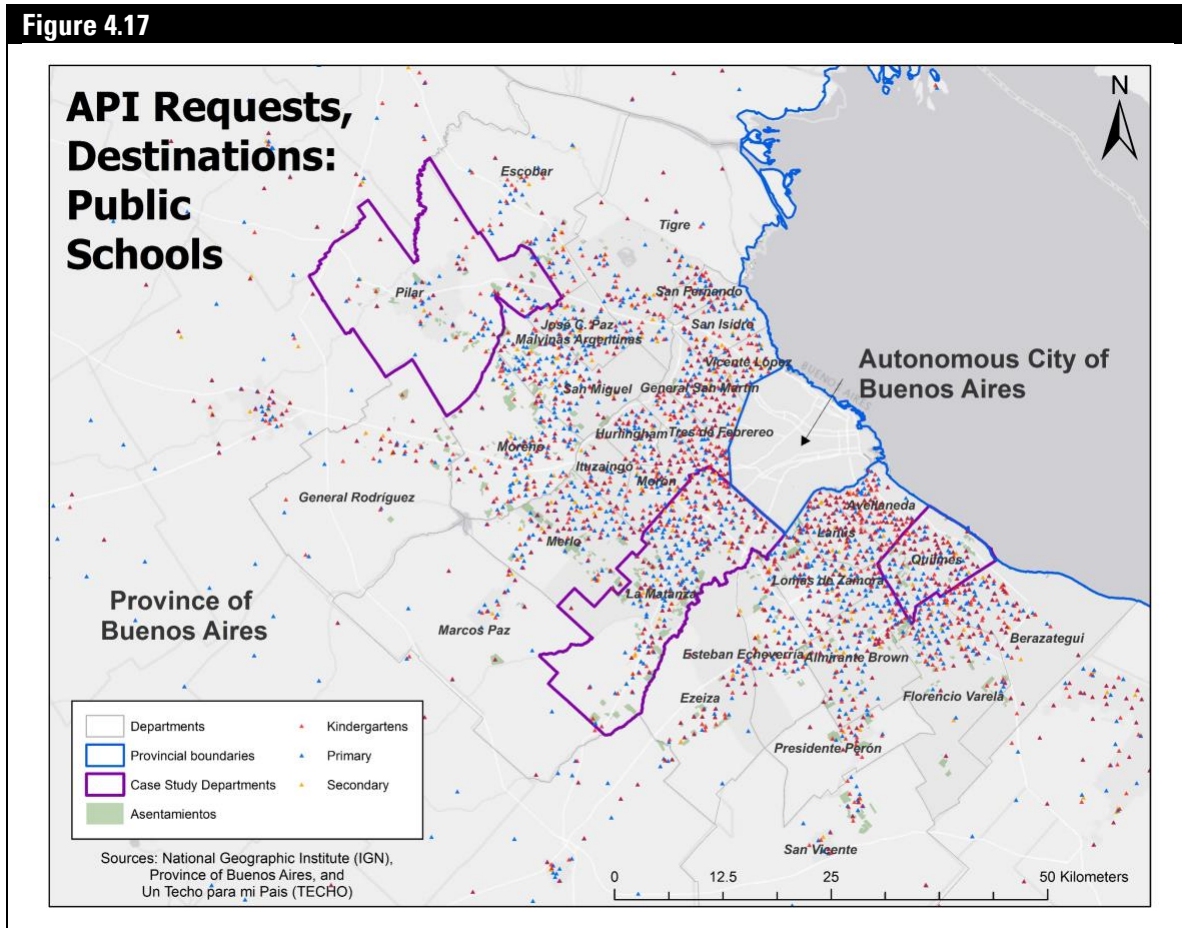
Part IV(c.4), Destinations: Health Care

The last destination category was health care; unlike the other activity classes, these came from multiple sources. My first intention was to use the "public healthcare establishments" ("establecimientos de salud públicos")[13] and "private healthcare establishments" ("establecimientos de salud privados") spreadsheets published in the same data portal from the Province of Buenos Aires. However, neither differentiates sub-categories within each file, making it impossible to know the size or scope of each; small clinics are included with hospitals and other specialists (facilities distinct enough from one another—some higher-order and others low-order—to produce vastly different accessibility patterns). The only exceptions are "urgent care units" ("unidades de pronta atención" or "prompt attention units"), fifteen of which were included in the public establishments sheet. According to local news sources, UPA's (the initialism I adopted) are meant to assist cash-strapped hospitals in the provision of first-aid services to members of the community.[14] I included them because they made sense as possible destinations for low-income *asentados* since, public health facilities are open to all people in Argentina, regardless of health coverage or income (I do not, however, have evidence of *asentados* actually using UPA's, or any other health facilities).

To supplement these data with information on hospitals and healthcare clinics, I returned to the GIS portal from the Ministry of the Interior, Public Works, and Housing. Here, I found two layers: "centers with general hospitalization" ("centros de internecion general")[15] and "centers of diagnosis and treatment without hospitalization" ("centros de diagnóstico y tratamiento sin internación")[16]. Both derived from data from the Ministry of Health and both last updated in May 2017, these are point shapefiles with attributes on the address, geographic coordinates, the name of each facility, and the government (e.g. federal, municipal, or national) responsible for that outlet. Separating out the hospitals from the diagnostic/treatment centers (a full definition of which was not immediately provided) was important to distinguish between their levels of service and, therefore, their likely variation in accessibility thresholds. I assume that hospitals, which allow patients to stay overnight, offer a greater array of services than a diagnostic/treatment center, which does not house patients; in turn, the former likely draws people from further distances than the latter. I selected public establishments under the assumption that low-income *asentados* were more likely to seek services where they are free; this opinion was informed by a conversation with an Argentine university contact.

Given that all three datasets, in their raw form, represented large scales (the whole province for the UPA's and the entire country for the hospitals and diagnostic/treatment centers), I had to narrow the pool of destinations ahead of my API requests. For the UPA's, I first sub-selected them from the provincial dataset of public health establishments (2,586 units); only 19 applied. With this step complete, I then added the nationwide point-shapefiles for hospitals (1,114) and

13 http://catalogo.datos.gba.gob.ar/dataviews/245383/establecimientos-de-salud-publicos/
14 http://www.po.org.ar/prensaObrera/online/politicas/buenos-aires-desmantelamiento-de-las-unidades-de-pronta-atencion-y-crisis-sanitaria
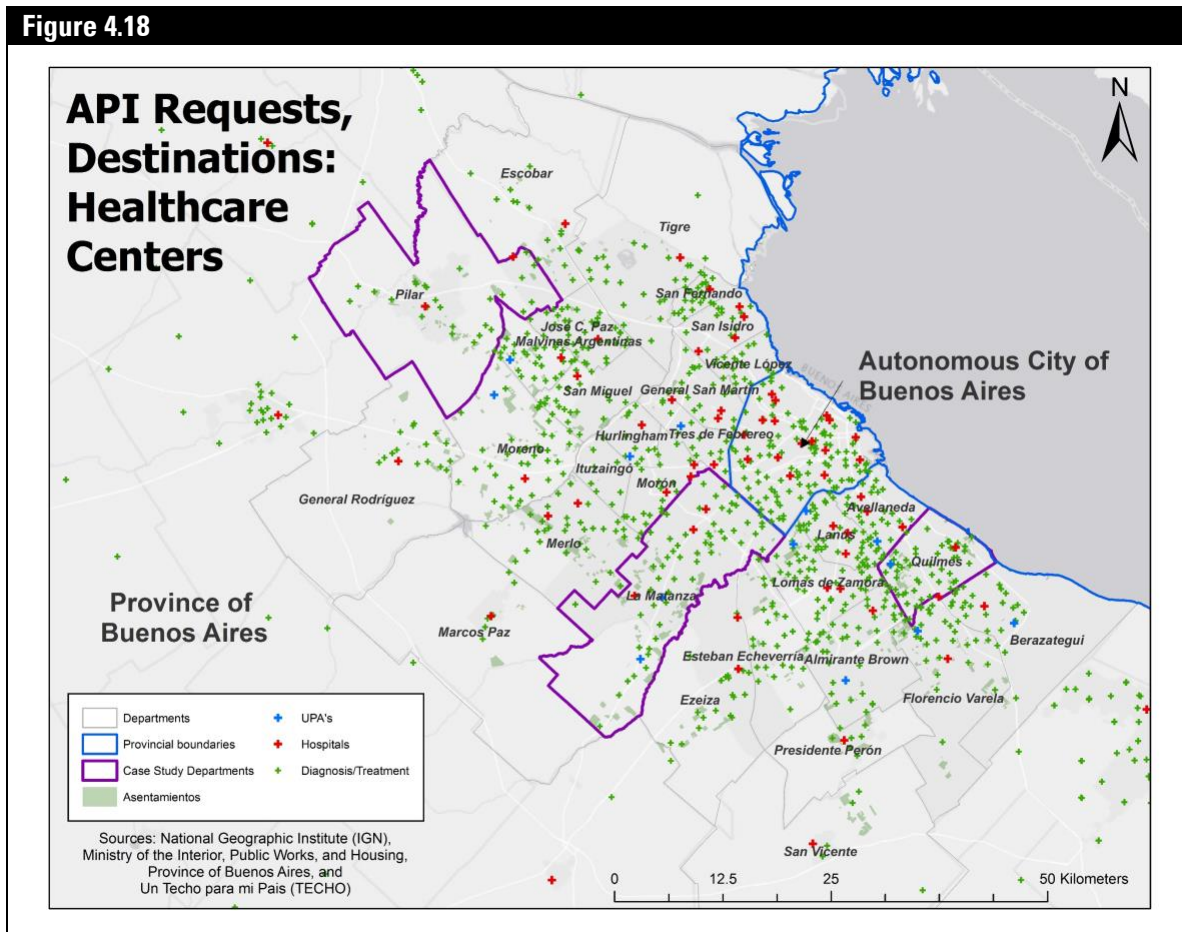15 http://sig.planificacion.gob.ar/layers/detalle_capa/daniela_centros_con_internacion_gral/
16 http://sig.planificacion.gob.ar/layers/detalle_capa/daniela_centros_diag_y_tratam_sin_internacion/

diagnostic/treatment centers (7,985). Next, I then determined, using the same Euclidean distance technique, whether individual establishments were one of the ten closest sites to any of the origins. Performed for each combination of healthcare site-type and department, this produced nine files, each of which also contained the same standardized latitude/longitude information for the corresponding points. The "destinations" falling into each health care category are on **Figure 4.18**.

Figure 4.18

API Requests, Destinations: Healthcare Centers

Sources: National Geographic Institute (IGN), Ministry of the Interior, Public Works, and Housing, Province of Buenos Aires, and Un Techo para mi Pais (TECHO)

Part IV(d), API Parameters: Mode, Departure Day, and Departure Time

   After acquiring the latitude and longitude coordinates of my origins and destinations, I then set values for the remaining travel-time request parameters: the mode of travel, the departure time, and the departure date.

   Given the topic of this paper, the mode was naturally set to "transit", the API's catch-all setting for all public transportation modes in an area: bus, rail, and subway. Considering bus (39%) and rail (5%) together, public transit was easily the most common mode preferred in INHABITAT's survey (subways, only found in CABA, were not relevant). There was no obvious reason to prioritize any of these three modes over the others, so I opted to include them all. This choice was aided by ENMODO's generalized findings that public transit was the favored mode of transit for workers (57% versus 27% for private and 16% for non-motorized) and represents the choice of greater than 60% of domestic workers, industrial workers, and company employees. Income-wise, those from the lowest quintile were also less likely to take private transit to school and more likely to walk or take the bus on health-trips. Students, meanwhile, were also more likely to travel on public transit, although, it should be noted, primary-school students actually preferred non-motorized travel (ST 2011).

   It would be interesting to eventually compare hypothetical travel times across the other modes offered by the API (driving, walking, or biking), although there are reasons for their exclusion from this project (other than reducing API requests). For instance, driving is irrelevant to *asentados* given its price exclusivity (less than 2% of INHABITAT respondents used a car for travel). Knowing the minimum driving times to activity sites from the *asentamientos* might be interesting from a comparative standpoint (showcasing the transit/driving dichotomy in accessibility that is growing in Buenos Aires), it does not represent a feasible choice for typical *asentados*. Walking and biking are necessitated by the absence of reasonable transit, or simple outright poverty (more than 50% of non-motorized trips to school are done by those in the first income quintile but are not viable options on the routes being studied here—*asentados* are more likely to simply not travel than walk or bike to jobs in far-off locations. Additionally, knowing the walking or driving times to activity sites is not of much use without first knowing the quality of transit, which could then contextualize any long walking or biking trips made out of necessity (ST 2011).

   As for departure date, I wanted to capture the transit services that would be available to the citizens of AMBA on an average work or school day, so I selected Wednesday mornings to serve as the model generic weekday. I personally picked Wednesdays since, based on personal experience living in Argentina, they were not likely to overlap with public holidays, which often take place on the days immediately preceding or following the weekend and feature special transit schedules akin to a typical weekend. Justification of a general weekday falls in line with ENMODO statistics showing that 74% of all trips, regardless of type, are made 5-times a week, an indication that trips on Wednesday are probably similar to those made on the other four days.

   When it came to travel time, however, I opted for a more nuanced approach. Using the time-of-travel information from the ENMODO survey, I gleaned the peak hour of different trip types: 7:00am for work trips (9-10% of all work trips—the highest evening peak was 9% at 5pm and 8.5%

at 6pm), 10:30am for health-related travel (peaks at 10am and 11am, with smaller peaks around noon, 9am, and 3-4pm), and 12:00pm for school trips (there were other peaks at 7am and 5pm—corresponding to the start of morning session and culmination of the afternoon session) (ST 2011). While the noontime peak reflects both trips to *and from* schools, there is no clear way to determine what percentage of these trips are going in either direction (it could be that most of the home-to-school trips are also at 7am), I also chose this time simply to capture accessibility trends during a different time of day than the work-trips, already set to the early morning hours.

Part IV(e), Requesting and processing the API output

Once I set all the Distance Matrix API parameters—API keys, origins, destinations, modes, day-of-week, and times-of-departure—I made my requests. With one API key dedicated to each study area, I requested data over the course of five weeks between February-March 2018. In cases where the number of origins or destinations exceeded the API's limit of 25 entries, I sub-divided the latitude and longitude sites into groups of twenty-five and fed them into the API separately. The total number of requests is summarized in **Table 4.19** below.

Table 4.19		Number of Destinations																			
Departments of Origin	Number of Origins	Buenos Aires CBD		Departmental CBD		Train Station		Nearest Dept. CBD		Public Urgent Care Center		Public Hospital		Pub. Diagnostic/Treatment		Public Kindergarten		Public Primary School		Public Secondary School	
		#	Requests	#	Requests	#	Requests	#	Requests	#	Requests	#	Requests	#	Requests	#	Requests	#	Requests	#	Requests
Pilar	72	1	72	1	72	23	1656	9	648	6	432	10	720	51	3,672	66	4,752	77	5,544	74	5,328
Quilmes	92	1	92	1	92	19	1748	6	552	7	644	11	1,012	70	6,440	120	11,040	109	10,028	78	7,176
La Matanza	120	1	120	1	120	42	5040	8	960	9	1,080	19	2,280	81	9,720	148	17,760	203	24,360	199	23,880
Travel Mode	"Transit"																				
Departure Time	7:00 AM							10:30 AM						12:00 PM							
Departure Day	"Wednesday"																				

Post-processing was simple, although a few extra steps were required for destination categories with more than one feature. The output, for each request, was a matrix of origins and destinations, with time values in seconds. The unique identifier for the origins (rows) and destinations (columns) was the "LATITUDE+LONGITUDE" script that was fed into the API. For the two cases where there was only one destination—Buenos Aires' CBD and the local departmental CBD—I only had to convert the times from seconds to minutes, dividing all values by sixty. For the remaining tables, I had to perform an extra function—using a tool within R—to select the minimum time in all the columns for each row. This script, in turn, produced a new matrix with just two columns: the unique identifier for a given origin and the travel time to the closest iteration of that activity class. Lastly, I converted this value to minutes. With the original shapefiles and these new time matrices sharing a common attribute (the latitude and longitude in "LATITUDE+LONGITUDE" format), I could join the R-output times tables to my shapefiles in ArcGIS to map my results.

Part IV(f), Statistical tests: ANOVA

Given that the primary question underlying this project is whether the *asentamientos* enjoy better or worse transit accessibility than the rest of metropolitan Buenos Aires, I naturally opted for statistical procedures that test differences-of-means. More specifically, I was interested in comparing the average travel times between (1) those census radios designated as "majority overlap", "minority overlap", and "no overlap" as well as (2) between those designated "any

overlap" and those with "no overlap". The former, given its three-sample comparison, warranted an ANOVA (analysis of variance) test while the latter required an independent-sample t-test. I will explain both—including their assumptions and test statistics—and how I used them for this study.

ANOVA, or analysis of variance, is a test that compares the means of three-or-more samples of ratio or interval data to see if they are statistically independent. Simultaneously looking at differences in variation within and across the study samples, an ANOVA can test both the variation of each sample around the "total mean" of all samples and the variance within each of these samples, around their respective means. As stated by McGrew, et al. (2014), "if the variability between the group means is relatively large as contrasted with the relatively small amount of variability within each group around its group mean, then the statistical conclusion is likely that the different groups have been drawn from different populations (pg. 175)." In other words, if the means of the independent samples are sufficiently different from one another, they are probably from different populations and, therefore, statistically-significantly different.

The test statistic for an ANOVA test is the F-statistic and is directly derived from the relative difference in variances between and within the different samples. When the F-score is high, indicating that the between-group variance is much larger than the within-group variances (the former is divided by the latter to get the F-score), than it is unlikely that all the samples were drawn from the same population and that *at least one* of those samples is from a separate population. This would reject the null hypothesis for an ANOVA test that all samples are from the *same* population and that their means and variances are all equal. The higher an F-score, the more likely the null hypothesis is not true. Nevertheless, such a conclusion can only be met if the ANOVA's main assumptions are met: that there are three or more independent random samples, that each population is normally distributed, that each population have equal variance, and that the variables are measured on an interval or ratio scale.

In my case, I am comparing travel times between the three samples of census radios, grouped by their overlap with AGBA's *asentamientos*: majority overlap, minority overlap, and none. Performed on the travel time estimates for all ten activity sites, the goal is to see whether the mean times of the "majority" or "minority" radios are statistically-significantly different from those qualified as "none." An ANOVA test on these samples will determine if the travel times for the majority or minority overlap groups are from a different population of travel times than those in the control group. What the test will not do, however, is indicate which specific samples are different from one another (if they are at all)—if the F-statistic is high, this could be because all three samples are different or just between one and the other two.

Part IV(g), Statistical tests: Independent Sample T-Tests

Given the main weakness of the ANOVA test, that it does not specify which, or how many, of the sample means are different from the others, I sought to add robustness by employing a series of independent-sample T-tests. Similar in nature to the ANOVA, they are used to compare the means and variances of two independent samples, instead of three of more. Requiring that the two samples be independent and random, normally distributed, and measured on an interval or ratio scale, the resulting statistic can, if significant, reject a null hypothesis that the two samples' means are equal (and, therefore, from the same population). The higher the t-score, the lower the likelihood that the two samples are from the same population. One caveat with the t-test has to do with whether the population variances for the two samples are known; if they are not, which is the case with my study, then the sample mean will have to stand in for the population mean. If the sample is small, less than 30 cases, than is probable that the sample mean is not a good representation of the population mean and that a t-distribution should be referenced. If the sample is greater than 30, the sample mean is probably more representative of its' population-level equivalent and a Z-distribution can be assumed.

For this study, there are several plausible t-tests, each characterizing the initial ANOVA results. I will look at the differences-in-means between (a) majority- and minority-overlap radios, (b) any-overlap (majority + minority) and no-overlap, (c) minority- and no-overlap, and (d) majority- and (e) non-majority-overlap (minority + none). Alongside the F-test from the initial ANOVA, I can perform a t-test on the travel times in each district (and to each destination) and see if the radios belonging to these respective categorizations display significant differences. A significant ANOVA test, say for the "distance to Buenos Aires CBD" measure, alongside significant t-tests for all the paired combinations of overlap-categorizations (and especially for the majority-none and minority-none tests) would provide strong evidence that the *asentamientos* enjoy worse transit-based accessibility than their neighbors. Contrary results, perhaps a significant F-statistic but insignificant T-tests, would help clarify, if not nullify, the initially-expected conclusion.