Pedestrian Demand Forecasting Methods Guidance

Technical Report 1: Direct demand models

Prepared for Department of Transport and Main Roads

CDM RESEARCH

Contents

Document history and status

Revision	Date issued	Author	Revision type
1	21/9/2020	C. Munro	Issue-1

Distribution of Copies

Revision	Media	Issued to
1	PDF	Department of Transport and Main Roads

Printed:	21 September 2020
Last saved:	17 September 2020 10:03 AM
File name:	0171 Pedestrian Demand Forecasting - Direct Demand Models - Technical Report (Issue-1).docx
Project manager:	C. Munro
Name of organisation:	Department of Transport and Main Roads
Name of project:	Pedestrian Demand Forecasting Methods Guidance
Project number:	0171

Executive Summary

This technical note describes an effort to develop a regression model to predict pedestrian demand depending on:

- the type of facility (e.g. signalised crossing, zebra crossing or shared path),
- nearby land uses (e.g. population, employment, retail facilities and parks),
- resident demographic characteristics (e.g. age, car ownership and income), and
- transport network characteristics (road connectivity).

In this report these regression models are referred to as direct demand models.

Models were estimated using the following data:

- Average weekday pedestrian demand (6 am – 6 pm) determined from 445 counts provided by the Department of Transport and Main Roads and local authorities in Queensland that were obtained between 2009 and 2020.
- Resident population, employment, car ownership, median household income, median age, household size, method of travel to work and social deprivation were obtained from the ABS 2016 Census of Population and Housing.
- Residential, commercial, and industrial land uses were obtained from the Queensland Land Use Mapping Program (QLUMP). This dataset provides only an indication of the likely use of the land parcel; land uses may be adjusted by local authorities or development may not have occurred.
- The location of parks, hospitals, retail facilities, supermarkets, hotels, motels, existing signals and road crossings were obtained from OpenStreetMap.
- School location and enrolment data was obtained from the Queensland Department of Education.
- Public transport (train, light rail and bus) stop locations and boarding data was obtained from Translink for South East Queensland.
- The State Digital Road Network maintained by the Department of Natural Resources, Mines and Energy was used as a proxy indicator for the footpath network.

Multivariate regression models were estimated using numerous model specifications and error structures. The preferred model is a negative binomial generalised linear model to account for overdispersion in the pedestrian counts. The model is implemented online at http://178.128.61.75:443/PedTool and is sensitive to the following predictors:

- Increasing walking and public transport mode shares for commuting increase the pedestrian forecast, as does increasing median age, the proximity of a school, presence of parkland and retail facilities.
- Increasing household income is associated with lower pedestrian demand.

- Footpaths located alongside roads, roundabouts and sign-controlled intersections are associated with lower demand than paths located outside of road-corridors, signalised intersections and zebra crossings.

The model is insensitive to proximity to population and employment, and cannot differentiate between, for example, zebra crossings and signalised crossings. These shortcomings are related to the quality of the underlying data. An extensive data collection activity is warranted, particularly where before and after-construction of new pedestrian facilities. This data collection should involve, at a minimum, three sequential days of collection to reduce the effect of interday variation.

1 Introduction

This report provides technical documentation on the development of a direct demand model for forecasting pedestrian demand in Queensland. The model is implemented online at http://178.128.61.75:443/PedTool. The direct demand model is a multiple regression model relating a number of *predictor* variables (or *independent* variables) with pedestrian demand (the *outcome* or *dependent* variable). The intention with this model is to provide practitioners with a rapid means of assessing the likely *magnitude* of demand for pedestrian facilities and in so doing assist in the development of business cases and prioritisation.

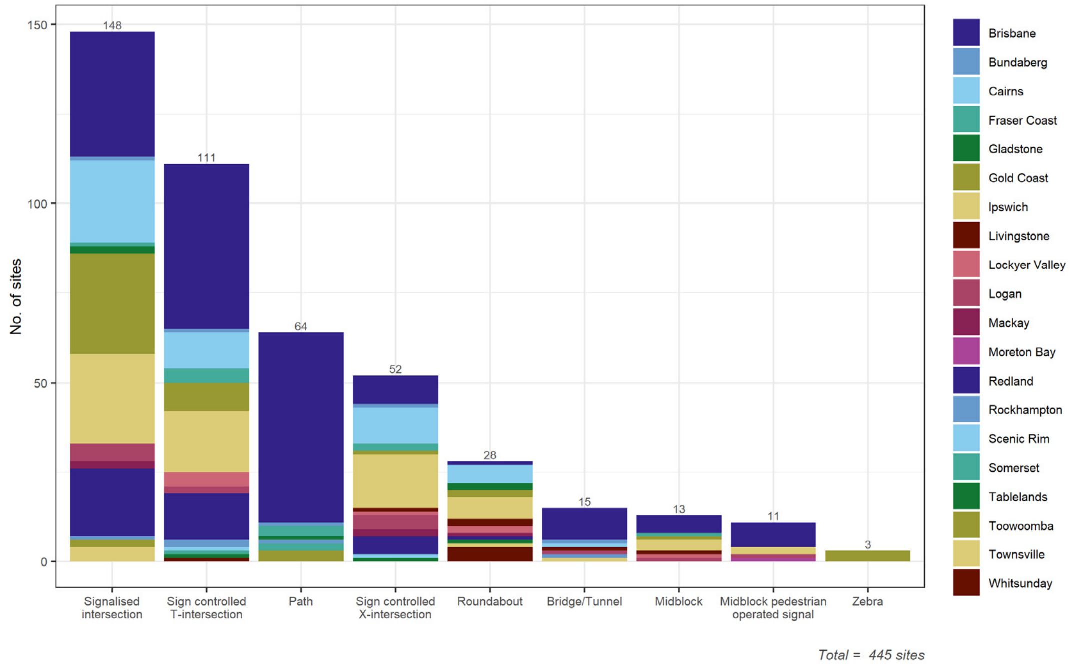
The data available for this model estimation is based on convenience; almost all pedestrian counts have been collated based on data provided by TMR and local authorities for other purposes, and have been collected over a wide period of time (2009 to 2020), at different times of year, using different methodologies (mainly video-based manual counts, but including some automatic counters) and in many cases extending only over one weekday from 6 am to 6 pm. Similarly, the predictor data such as population, land uses and demographics have shortcomings and do not necessarily reflect the land uses at the time the count was undertaken. Given these limitations the models developed herein are subject to uncertainty and should be considered as a guide to the plausible level of demand only; the practitioner may have a compelling argument as to why demand at a subject site will be significantly less or greater than predicted by these models.

2 Datasets

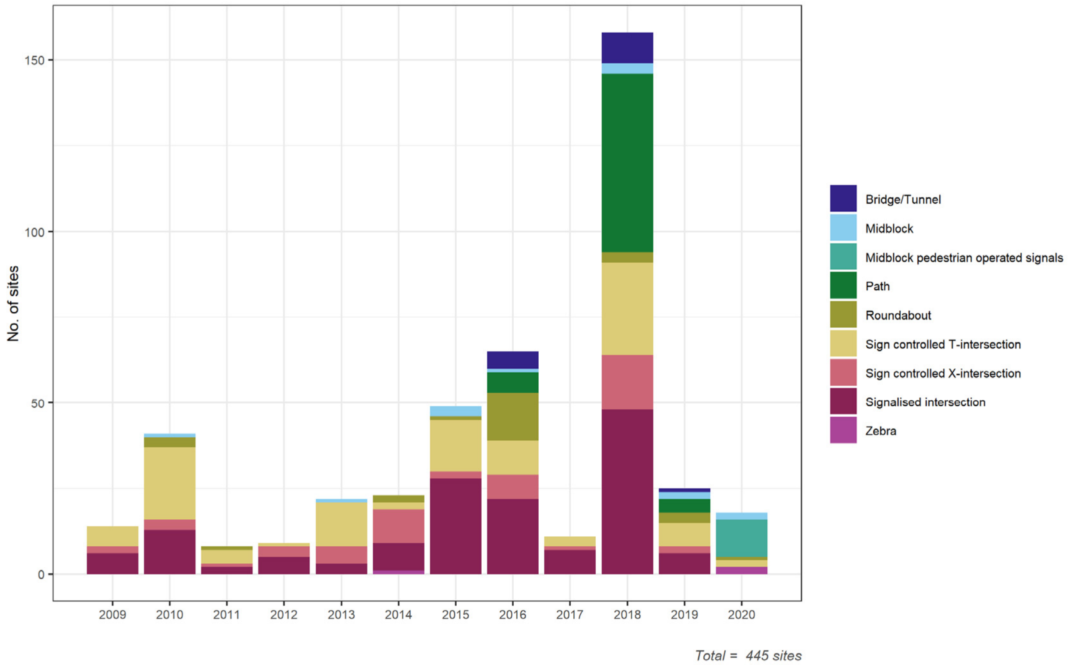
2.1 Pedestrian demand

Pedestrian demand was obtained from counts obtained at 445 locations across Queensland between 17 February 2009 and 29 March 2020. In most instances the demand was obtained from one-day counts between 6 am and 6 pm, although there were instances where counts were obtained on multiple days or continuously using automatic counters. The models in Section 4 are based on the 425 sites for which one or more weekday counts were available; for the remaining 20 sites only weekend counts were available.

A third of sites were at signalised intersections, followed by sign-controlled T-intersections (25%) and paths located away from roads (14%). There were few midblock pedestrian operated signals or zebra crossings in the dataset. Moreover, most paths located away from roads were in Brisbane (Figure 2.1). The distribution of counts by year is shown in Figure 2.2; many counts from 2018 are attributable to a tranche from Brisbane City Council.



■ **Figure 2.1: Sites by type and LGA**



- **Figure 2.2: Sites by year and type**

The demand at sites with multiple legs, such as signalised intersections and roundabouts, was aggregated to represent *total* crossing demand at this location.

Demand is reported both as a 12-hour count (6 am – 6 pm) and as the peak hour using the maximum rolling hourly count from the 15-minute bins. Where multiple days of data were available for a site the simple average was calculated for the daily and peak hour count.

2.2 Predictor variables

Predictor variables were selected on that basis that (a) they could plausibly be expected to affect pedestrian demand, and (b) data was readily available for Queensland.

2.2.1 Population

Population data was obtained from the 2016 Census of Population and Housing and mesh block (MB) geographic areas. MBs are the smallest geography available and generally contain between 30 and 60 dwellings. Persons are coded to mesh blocks based on the place of usual residence.

■ **Figure 2.3: Mesh blocks sample**

2.2.2 Employment

Counts of employed persons were obtained from the 2016 Census of Population and Housing for destination zones (DZ) using origin-destination tables obtained using the ABS TableBuilder Pro application. DZs align to Statistical Area Level 2 (SA2) zones. Employments coded to no geographic area (9499 Place of Work No Fixed Address, 9799 Migratory – Offshore – Shipping, 9099 Place of Work Capital City undefined, 9899 Place of Work State/Territory undefined) were excluded.

All full- and part-time employments were included (and given equal weight) where an origin and destination were stated.

■ **Figure 2.4: Destination zones sample**

2.2.3 Car ownership

Household car ownership was obtained from the 2016 Census of Population and Housing at the Statistical Area Level 1 (SA1) geography. SA1s have a population between 200 and 800 persons with an average of around 400. Place of enumeration data was used (table P29) and the no-car owning ratio was obtained by dividing the number of dwellings with no car by the total number of dwellings within each SA1.

2.2.4 Household income

Household median weekly income was obtained at the SA1 geographic level using place of enumeration table P02 from the 2016 Census of Population and Housing.

2.2.5 Age

Median age was obtained at the SA1 geographic level using place of enumeration table P02 from the 2016 Census of Population and Housing.

2.2.6 Household size

Average household age was obtained at the SA1 geographic level using place of enumeration table P02 from the 2016 Census of Population and Housing. This statistic is the number of persons usually resident in occupied private dwellings; it includes partners, children and co-tenants but excludes visitors. “Usual” residency is defined as the place the person was most commonly resident over the past 90 days.

2.2.7 Method of travel to work: place of enumeration

The method of travel to work on census day was obtained from the 2016 Census of Population and Housing using SA1s for the origin (usually home) location. Place of enumeration table P44 was used to obtain counts of employed persons travelling to work by one method (walked only) and using public transport (train, bus, ferry or tram) either as a sole mode or in conjunction with other modes. Public transport-involved trips were used as a predictor on the assumption that walking is likely to be required at the access and/or egress end of most public transport trips.

The counts were converted to mode shares by dividing by the total number of employed persons in the SA1 excluding those who (a) did not state a method of travel to work, (b) did not go to work on the census day or (c) worked from home.

2.2.8 Method of travel to work: destination zone

The method of travel to work on census day was obtained from the 2016 Census of Population and Housing using DZs for the destination (usually workplace) location using origin-destination tables obtained using the ABS TableBuilder Pro application. Identically to place of enumeration, walking trips as a sole mode and trips involving public transport were selected. These counts were converted to mode shares by dividing by the total number of employed persons in the SA1 excluding those who (a) did not state a method of travel to work, (b) did not go to work on the census day or (c) worked from home.

2.2.9 Social deprivation

Social deprivation was estimated at the SA1 geographic level using the Index of Relative Socio-economic Disadvantage (IRSD) derived as part of the Socio-Economic Indexes for Areas (SEIFA) by the ABS. The index is derived from the 2016 Census of Population and Housing and is based on income, education, employment, occupation, housing and other variables[1]. The index is generally between 700 and 1,100 with lower values indicating higher levels of deprivation.

2.2.10 Land use

2.2.10.1 Commercial

Commercial land uses were obtained from the Queensland Land Use Mapping Program (QLUMP) updated in June 2019. QLUMP provides planning cadastres for the entire state. It is noted that the planning designation does not necessarily accord with the as-built land use for a location. Those cadastres with the tertiary designation of “Commercial services” were considered in this category. These commercial land uses are generally retail or office-type establishments.

[1] The derivation of the IRSD is described by the ABS: https://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/756EE3DBEFA869EFCA258259000BA746/$File/SEIFA%202016%20Technical%20Paper.pdf.

2.2.10.2 Water

Proximity to water is likely to encourage walking activity, such as paths along lakes, rivers or shorelines. Inland watercourses were obtained from QLUMP using tertiary designations including river, lake, reservoir, wetland and estuary. These spatial objects were combined with the Queensland coastline obtained from OpenStreetMap[2].

2.2.10.3 Parks

Paths adjacent or running through parks and gardens may be associated with higher walking demand. Features with key = “leisure” and value = “park” were obtained from OpenStreetMap for Queensland.

2.2.10.4 Signals and road crossings

The proximity to traffic signals may affect pedestrian demand. Features with key = highway and value = “traffic_signals” were obtained from OpenStreetMap for Queensland.

Pedestrian crossings were obtained from OpenStreetMap using the following features:

- Crossing: feature = “highway” and value = “crossing” (these are mainly signalised pedestrian crossings but also include some zebra crossings)
- Zebra: feature = “crossing” and value = “zebra”
- Refuge: feature = “crossing” and value = “island”

2.2.10.5 Hospitals

It is likely that proximity to a hospital will influence pedestrian demand. Features with key = “building” and value = “hospital” were obtained from OpenStreetMap for Queensland.

2.2.10.6 Retail

Retail land uses are likely to be significant pedestrian traffic generators and attractors. Features with key = “building” and value = “retail” were obtained from OpenStreetMap for Queensland.

2.2.10.7 Supermarkets

Supermarkets are likely to be significant pedestrian traffic generators and attractors. Features with key = “shop” and value = “supermarket” were obtained from OpenStreetMap for Queensland.

2.2.10.8 Hotels and motels

Hotels and motels will generate pedestrian movements that would not otherwise be captured using predictors such as residential population or employment alone. This is likely to be particularly true in areas with high tourism demand. Features with key = “tourism” and value = “hotel” or “motel” were obtained from OpenStreetMap for Queensland.

[2] A full list of OpenStreetMap feature attributes is at https://wiki.openstreetmap.org/wiki/Map_Features.

2.2.11 Schools

Full-time equivalent student enrolments for public schools were obtained from the Queensland Department of Education based on day 8 of the 2020 school year. Equivalent data for private schools was available only for the last day of February 2019.

2.2.12 Transport network

2.2.12.1 Public transport

Train, bus and ferry boarding data was obtained from TransLink origin-destination data for 2019[3]. Stop locations were extracted from the Translink GTFS API specification[4]. Boardings data was only available for South East Queensland (Gold Coast to Gympie and as far west as Toowoomba). The origin-destination matrix was aggregated to determine the annual boardings per stop and were then divided by 365 to represent an average day.

2.2.12.2 Road network

Street centrelines were obtained from the State Digital Road Network maintained by the Department of Natural Resources, Mines and Energy[5]. These streets were divided into intersections (points of intersecting streets) and links (lines between intersections).

The implicit assumption is that all streets have footpaths. This will not be correct, and there will be areas where there will be no footpaths alongside streets and areas where there will be footpaths but no streets (e.g. through parks). Footpath data was only available for Gold Coast, Logan, Rockhampton and Ipswich City Council areas.

[3] https://www.data.qld.gov.au/dataset/go-card-transaction-data

[4] https://www.data.qld.gov.au/dataset/general-transit-feed-specification-gtfs-qconnect

[5] https://www.data.qld.gov.au/dataset/baseline-roads-and-tracks-queensland

3 Analytical procedures

3.1 Catchments

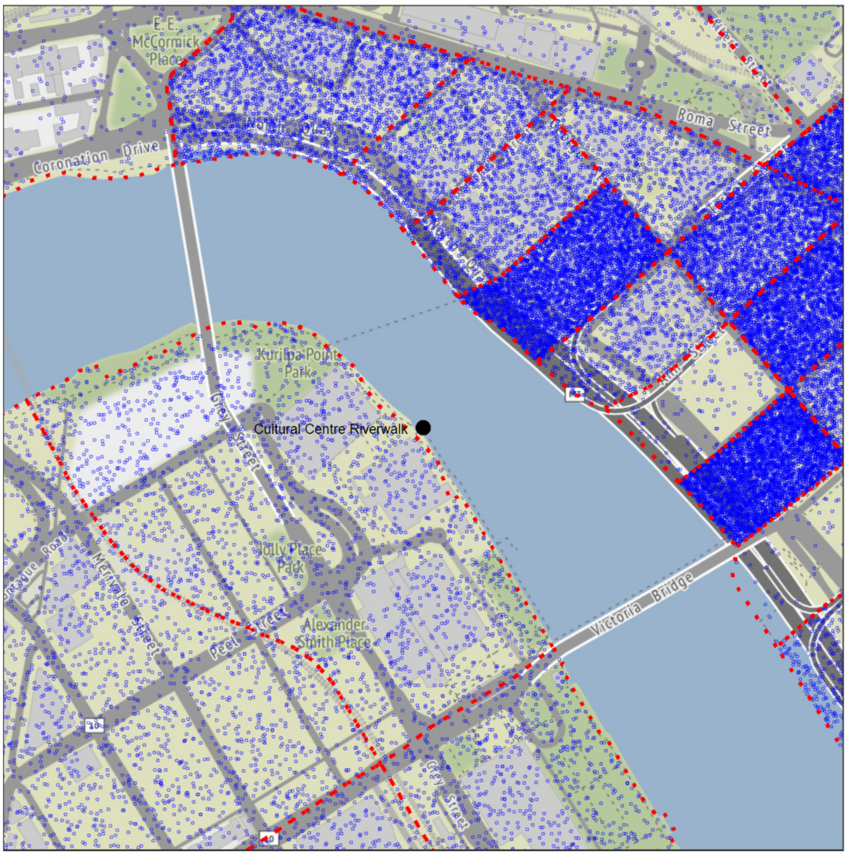
Two main approaches were tested to calculate predictor variables from the input data:

1. **Simple catchment:** under this approach variables were summed within a defined buffer of the count site (e.g. population within 2,000 m radius)
2. **Weighted sum:** this alternative approach applied a decay function to each spatial unit, thereby allocating higher weighting to units closest to the count site and lower values to those further away (e.g. a school directly adjacent to the count site would be weighted more highly than only located some distance away).

3.1.1 Apportioning partial polygons

Under either the simple catchment or weighted sum approach it was necessary to convert predictors described as spatial polygons (e.g. population and employment) to point objects. This was done by filling each polygon with points randomly distributed in proportion to the predictor variable. For example, if a mesh block had a resident population of 120 persons the polygon would be filled with 120 points. This is illustrated in Figure 3.1. This process leads to more plausible sensitivity to increasing the catchment radius and also better facilitates the application of a distance decay function in the weighted sum approach[6].

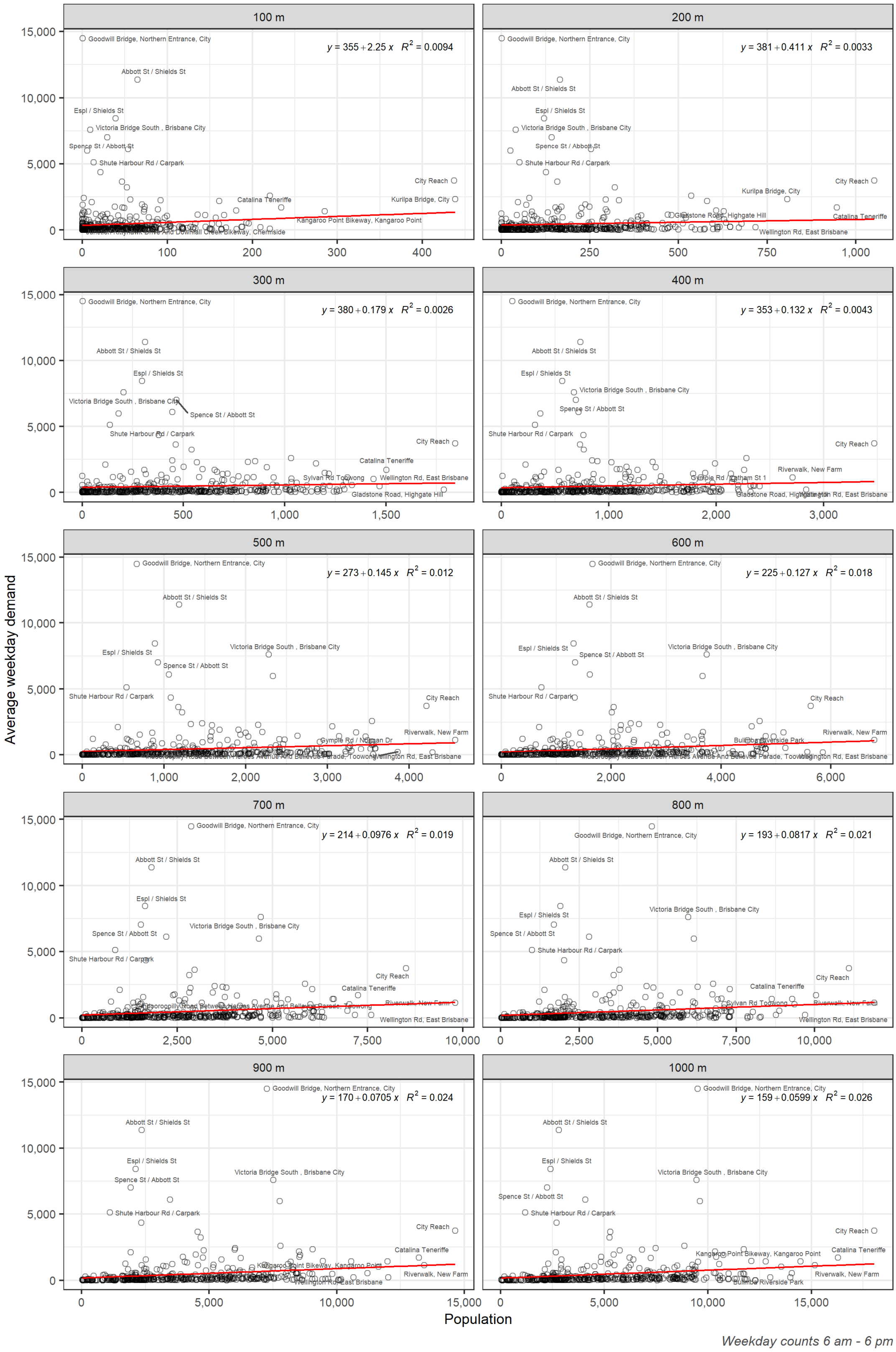
[6] A simpler procedure would be to apportion values based on the area overlap. However, such an approach does not allow for the subsequent application of a distance decay function.



■ **Figure 3.1: Destination zone employment data represented as points; destination zones with higher employment densities contain more points**

3.2 Simple catchment

Catchments for each predictor were calculated for distances 100 – 2,000 m from the count site in 100 m increments. For example, the relationship between population and pedestrian demand for radii from the count site varying from 100 to 1,000 m is shown in Figure 3.2. The relationship between population and demand is generally weak, as indicated both by visual inspection and the coefficient of determination (R^2) for a simple linear fit of the data. Similarly weak correlations were observed for many of the other predictors, suggesting few of the predictors are strongly correlated with demand *and* that the selection of a catchment radius is not critical to the model development. As such, in the modelling a catchment of 2 km is assumed where required on the basis that this catchment is likely to catch the vast majority of walking activity.



■ **Figure 3.2: Relationship between population catchment and pedestrian demand by catchment radius**

3.3 Weighted sum

This second approach weights predictors that are closer to the count site higher than more distant sites. It seems reasonable to expect that, for example, a school would have the greatest influence on pedestrian demand immediately adjacent to the count site and that this impact would decay as the distance between the school and count site increases.

For the purpose of calculating a weighted sum all predictors within 2 km of the count site were selected. A score was then calculated based on the sum of the predictors *i* multiplied by a distance decay:

$$Score = \sum_{i}^{n} A_i \cdot e^{\beta d_i}$$

where:

i is each observation of the predictor within the 2 km buffer

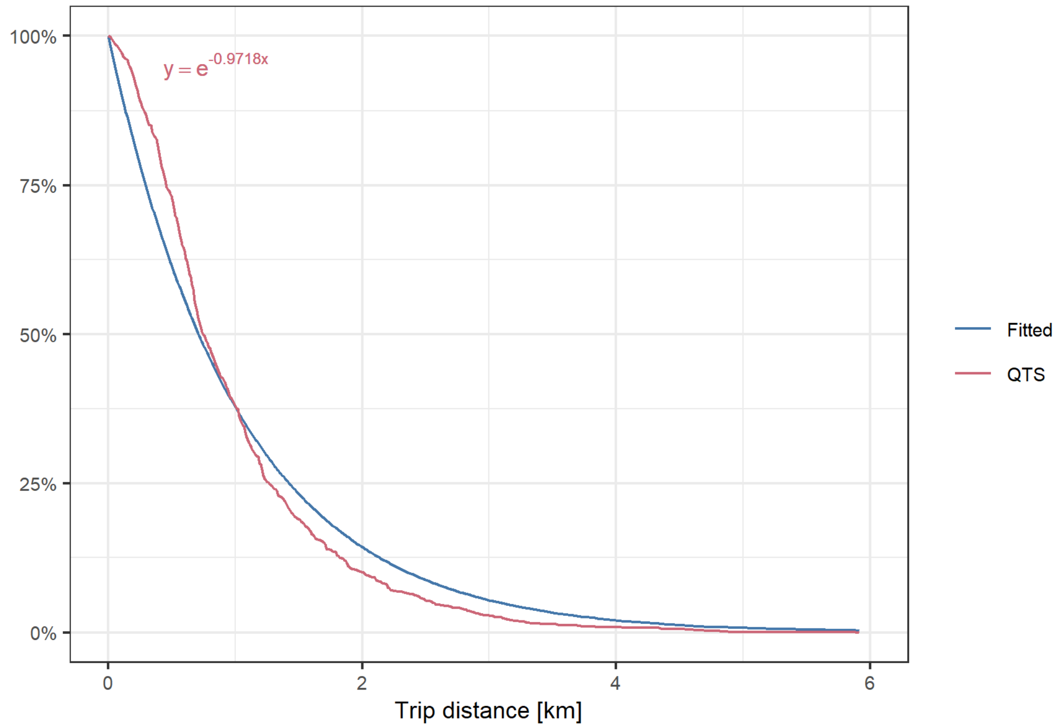
A is a size variable for each observation

d is the crow fly distance between the count site and the observation.

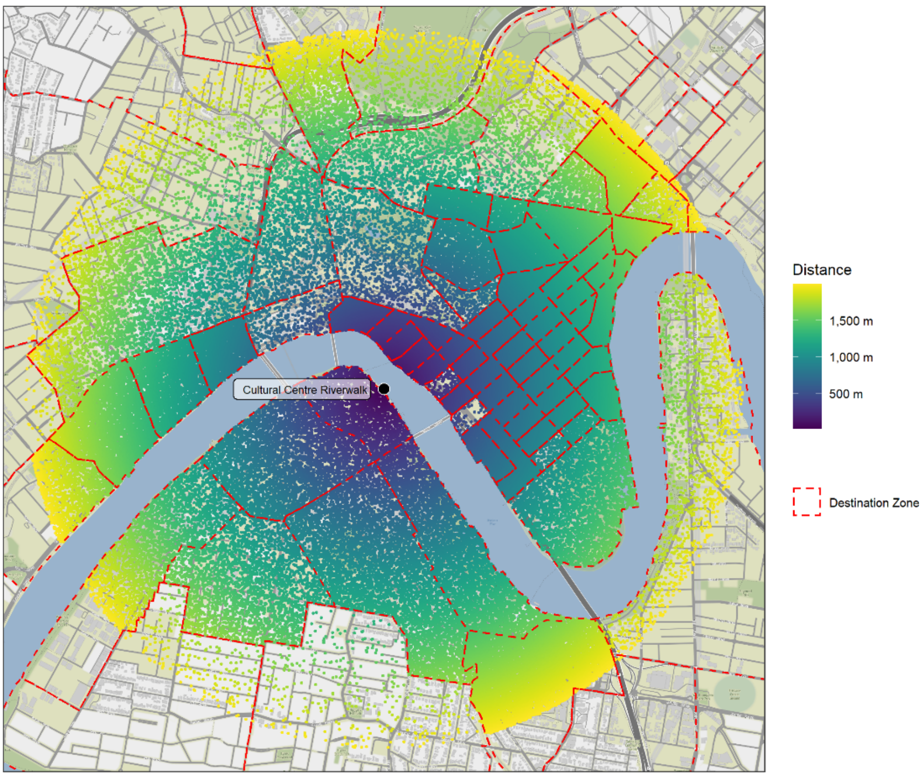
The size variable differed depending on the predictor:

- For population, employment, mode share, household income, household size and median age, the size variable was one (i.e. each point represented one individual or household)
- For schools the size variable was the student enrolment
- For public transport the size variable was the average number of boardings per day.
- For land uses (commercial, retail, supermarkets, parks) the size variable was the projected area; for multistorey facilities this will severely understate the floor area.

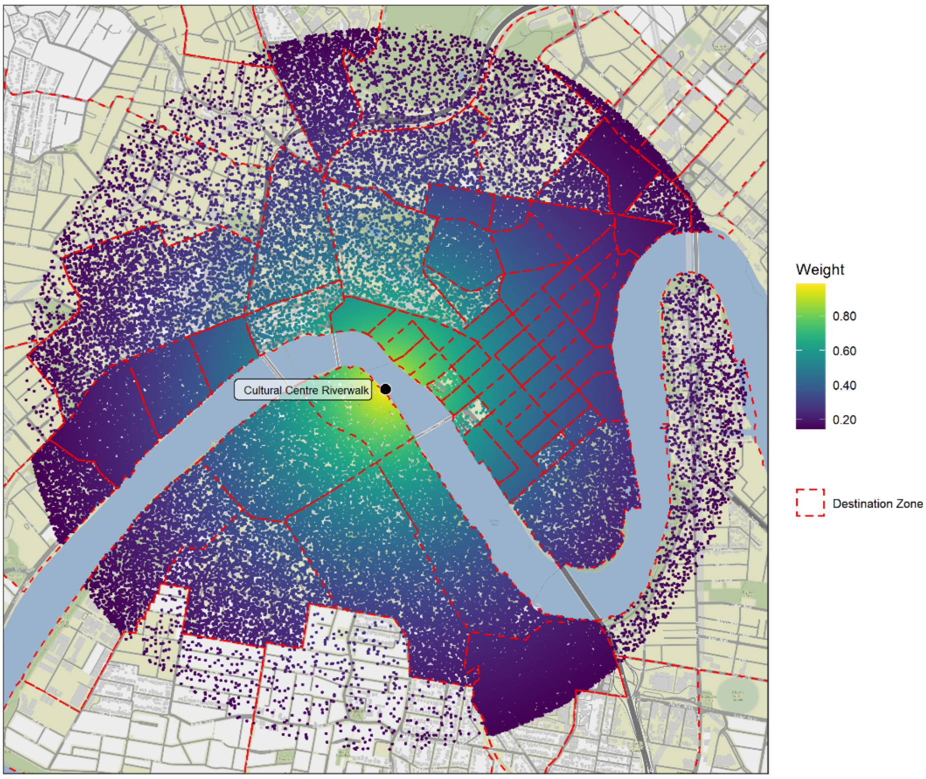
The distance decay weighting was determined from the inverse cumulative distribution function for walking trips (as a sole mode) derived from QTS 2017-19. This distribution is shown in Figure 3.3; an exponential decay exponent of -0.9718 was fit to this data and implies a weighting of 0.38 to features 1 km from the count site and 0.14 to features 2 km from the count site. The application of this decay function to sample site in Brisbane (Southbank) is shown in Figure 3.4 and Figure 3.5. In Figure 3.4 the distance from the count site is illustrated along with the employment density. Figure 3.5 then applies the distance decay function to each of these points, illustrating the much higher weight (and hence influence) assigned to employment close to the count site.



■ **Figure 3.3: Distance decay function (QTS 2017-19 walking trips)**



■ **Figure 3.4: Example of distribution of employment and distance from count site**



■ **Figure 3.5: Weighting of employment using the decay function**

3.4 Correlations

An initial analysis of the relationship between the predictors and pedestrian demand was undertaken by visual inspection. The relationship between the main predictors and weekday pedestrian demand is provided in Appendix B. These graphs generally indicate a weak relationship. Moreover, some sites are clearly outliers – particularly a number of sites in the Cairns CBD.

3.5 Outliers

Influential outliers were identified and removed by applying a linear regression to the population catchment within 100 – 1,000 m in 100 m increments against the daily demand. Those sites with Cook's D > 4 / N = 4 / 445 = 0.0058 at each of the ten catchment radii were flagged and then inspected manually to assess whether their removal is warranted. This manual process checked the sites against not only population but also employment, school students and proximity to parks and retail. This process resulted in the removal of ten sites:

- Abbott St / Shields St, Cairns
- City Reach, Brisbane
- Esplanade / Shields St, Cairns
- Goodwill Bridge, Brisbane
- Sheridan St / Shields St, Cairns
- Shite Harbour Rd / Carpark, Airlie Beach
- Spence St / Abbott St, Cairns
- Spence St / Grafton St, Cairns
- Victoria Bridge, Brisbane.

It is speculated that the very high pedestrian counts at the Cairns and Airlie Beach Site (accounting for eight of the ten outliers) is attributable to tourist traffic which is not adequately accounted for in the predictors. The exclusion of the three Brisbane sites may be due to the significance of these three sites (City Reach, Goodwill Bridge and Victoria Bridge) in the local walking network which is not well captured by the predictors.

4 Model estimation

4.1 Models

Two categories of regression models were tested:

1. Ordinary Least Squares (OLS)
    a. Untransformed (linear) outcome variable
    b. Log-transformed outcome variable
2. Generalised Linear Models (GLM)
    a. Quasi-Poisson (QP)
    b. Negative Binomial (NB)

Numerous model specifications were tested, including linear and logarithmic transformation of the pedestrian demand[7].

The daily and peak hour walking demands were severely over-dispersed, with a residual dispersion of 226 for the daily and 49 for the peak hour data. To redress this overdispersion, and to constrain the counts to positive values, the quasi-Poisson and negative binomial models were preferred. Model parameter selection was undertaken in the first instance by forward and backward stepwise variable selection using minimum Akaike Information Criterion (AIC) as the selection criteria[8]. Manual observation of the model parameters was then used from the optimum model to verify parameter plausibility and, in some cases, to add back in variables that add additional model sensitivity.

4.2 Results

The results for naïve predictors within a 2 km catchment are shown in Table 4.1. In these models the predictor values are the unweighted totals within 2 km of the count site. Dummy variables are used for facility type with off-road pathways as the base. Continuous predictors have been standardised[9].

The following are the main insights from these models:

- Population with 2 km is positively associated with higher pedestrian demand; this finding is consistent across all four models.
- Employment within 2 km is <u>negatively</u> associated with pedestrian demand; this finding is consistent across all models and is counterintuitive (all else being equal, it would be expected that an increase in employment near the site would be associated with an increase in pedestrian demand).

[7] Wherever the pedestrian demand was log transformed a value of one was added to the count prior to transformation to account for sites with zero weekday demand.

[8] The process used the stepAIC function within the MASS package in the R statistical programming language. For the quasi-Poisson model for which no AIC is defined the variable selection process was entirely manual and based on the p-values.

[9] Standardisation was achieved by subtracting the mean and dividing by one standard deviation.

- Public transport boardings are positively associated with pedestrian demand; again, this finding is consistent across all models.
- The presence of schools students within 2 km is statistically insignificant for most models aside from the linear OLS model, where students are negatively associated with pedestrian demand.
- In the log-transformed OLS model and negative binomial GLM higher median household income is associated with lower pedestrian demand.
- Higher median age is associated with higher pedestrian demand in all models.
- Walk (as a sole mode) mode shares for journeys to work are associated with higher pedestrian demand across all models based on the home location, and for the log-transformed OLS model and negative binomial GLM when based on the workplace (i.e. destination).
- Higher public transport mode shares for journeys to work are associate with increased pedestrian demand for the negative binomial GLM model.
- The area of hotel, park and retail facilities within 2 km is positively associated with pedestrian demand for some models, although the association is strongest for retail area.
- The proximity to the nearest hospital and school is associated with higher pedestrian demand (i.e. shorter distances are associated with higher demand) across all four models.
- The proximity to supermarkets is only significant for the negative binomial model.
- The proximity to universities is associated with a decrease in pedestrian demand for the log-transformed OLS model and negative binomial GLM; this result is counterintuitive.
- All facilities, such as sign-controlled and signalised intersections and roundabouts, are associated with lower pedestrian demand than for paths located outside of road corridors.
- The intercept of the linear OLS model is negative, which is nonsensical for count data.

Table 4.1: Simple catchment model coefficients

	OLS: LM	OLS: Log	GLM: QP	GLM: NB
(Intercept)	269.285***	5.043***	5.615***	5.587***
	(0.000)	(0.000)	(0.000)	(0.000)
pop	127.620***	0.801***	0.415***	0.423***
	(0.000)	(0.000)	(0.000)	(0.000)
emp	-224.804***	-0.709***	-0.560***	-0.476***
	(0.000)	(0.000)	(0.000)	(0.000)
pt	94.610***	0.177**	0.125***	0.177***
	(0.000)	(0.015)	(0.003)	(0.010)

	OLS: LM	OLS: Log	GLM: QP	GLM: NB
students	-66.030**			
	(0.038)			
dist.university		0.099		
		(0.118)		
dist.hospital	-26.692			
	(0.157)			
dist.school	-33.024*	-0.335***	-0.213***	-0.360***
	(0.083)	(0.000)	(0.004)	(0.000)
hh.income.wgt		-0.185***		-0.260***
		(0.008)		(0.000)
median.age.wgt		0.217***	0.237***	0.258***
		(0.002)	(0.002)	(0.000)
walk.share.wgt	396.134***	0.932***	0.768***	0.736***
	(0.000)	(0.000)	(0.000)	(0.000)
pt.share.wgt				0.283***
				(0.007)
dz.walk.share.wgt		-0.159**		-0.110*
		(0.030)		(0.084)
retail.km2	30.318	0.112*	0.188***	0.129**
	(0.158)	(0.088)	(0.001)	(0.025)
park.km2		0.133**	0.132**	
		(0.032)	(0.016)	
LNR_500				0.092
				(0.100)
dummyFootpath		-1.084***	-1.144**	-1.270***
		(0.001)	(0.021)	(0.000)
dummyRbt		-0.359	-0.551**	-0.374*
		(0.135)	(0.014)	(0.068)
dummySign	-78.925*	-0.793***	-0.841***	-0.844***
	(0.093)	(0.000)	(0.000)	(0.000)
dummySignal	186.337***			
	(0.000)			
dummyBridge				-0.485
				(0.118)
Num.Obs.	425	425	425	425
R^2	0.468	0.569		
R^2 Adj.	0.455	0.554		
AIC	6220.8	1284.6		5303.2

	OLS: LM	OLS: Log	GLM: QP	GLM: NB
BIC	6269.4	1349.4		5372.1
Log.Lik.	-3098.383	-626.300		-2634.599

Values are standardised
Values in brackets are p-values
* p < 0.1, ** p < 0.05, *** p < 0.01

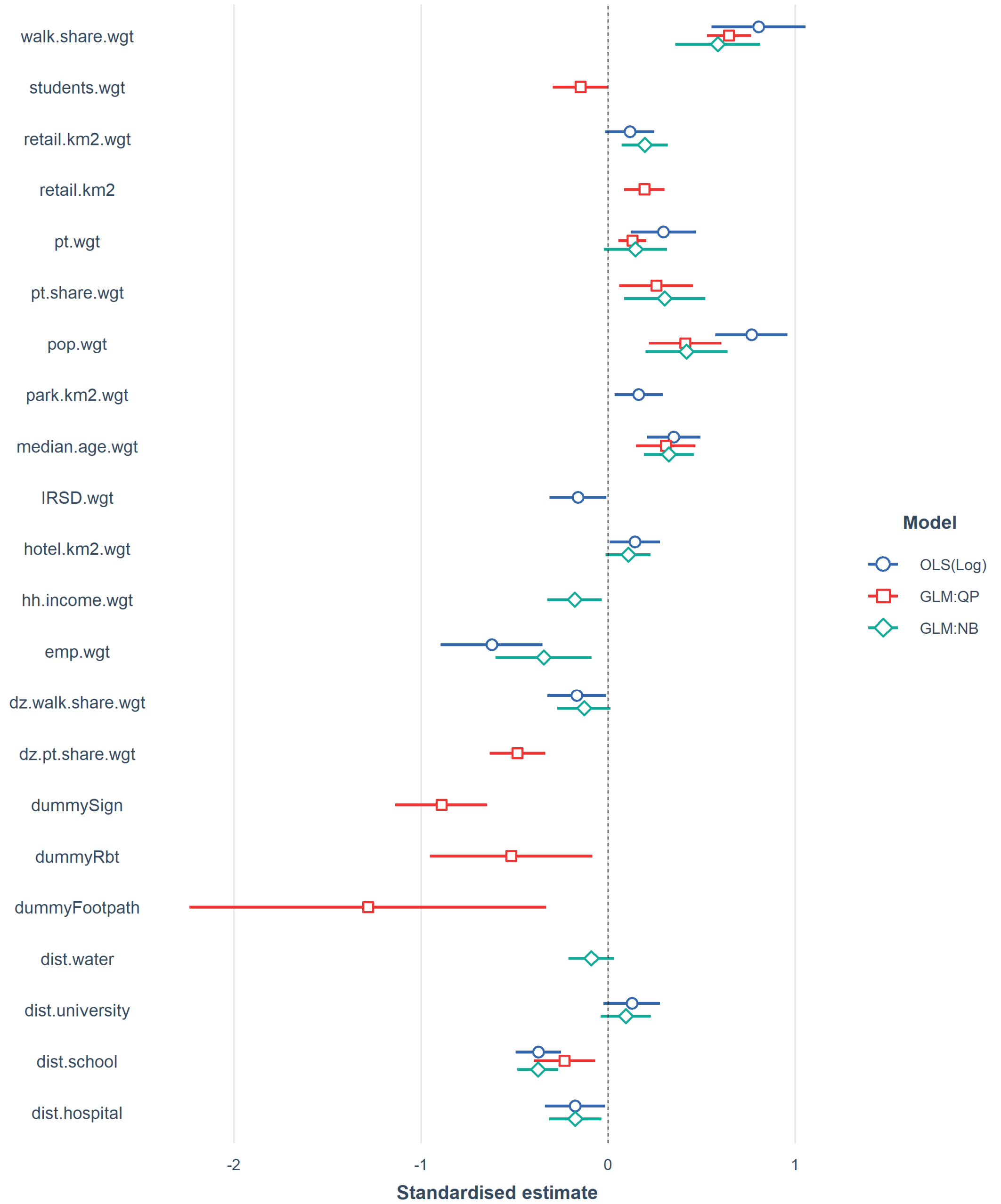
The models using the decay weighted population, employment, public transport and school student predictors are shown in Table 4.2. The model insights are similar to the simple catchment model. The coefficients for the three models are illustrated in Figure 4.1; in general the coefficients are similar between the models.

- **Table 4.2: Weighted catchment model coefficients**

	OLS: LM	OLS: Log	GLM: QP	GLM: NB
(Intercept)	300.515***	4.690***	5.619***	5.250***
	(0.000)	(0.000)	(0.000)	(0.000)
pop.wgt	142.422***	0.766***	0.432***	0.419***
	(0.001)	(0.000)	(0.000)	(0.000)
emp.wgt		-0.623***		-0.346***
		(0.000)		(0.008)
pt.wgt	105.112***	0.295***	0.107***	0.146*
	(0.000)	(0.001)	(0.006)	(0.092)
students.wgt			-0.088	
			(0.230)	
dist.hospital	-50.806**	-0.177**		-0.176**
	(0.019)	(0.032)		(0.014)
dist.school	-39.955*	-0.374***	-0.233***	-0.376***
	(0.051)	(0.000)	(0.005)	(0.000)
dist.water	-42.409*			-0.090
	(0.067)			(0.147)
dist.university		0.126		0.094
		(0.102)		(0.172)
walk.share.wgt	359.118***	0.805***	0.602***	0.586***
	(0.000)	(0.000)	(0.000)	(0.000)
pt.share.wgt			0.209**	0.302***
			(0.040)	(0.007)
dz.walk.share.wgt		-0.168**		-0.129*
		(0.036)		(0.074)
dz.pt.share.wgt	-182.945***		-0.448***	
	(0.000)		(0.000)	

	OLS: LM	OLS: Log	GLM: QP	GLM: NB
hh.income.wgt	95.037*			-0.179**
	(0.093)			(0.016)
hh.no.car.wgt	-104.591**			
	(0.045)			
IRSD.wgt	-97.904*	-0.162**		
	(0.099)	(0.036)		
median.age.wgt	74.969***	0.350***	0.319***	0.324***
	(0.006)	(0.000)	(0.000)	(0.000)
retail.km2.wgt	77.593***	0.115*	0.221***	0.195***
	(0.001)	(0.084)	(0.000)	(0.002)
hotel.km2.wgt		0.142**		0.106*
		(0.039)		(0.082)
park.km2.wgt		0.164**		
		(0.013)		
dummyFootpath			-1.297***	
			(0.007)	
dummyRbt			-0.478**	
			(0.029)	
dummySign			-0.877***	
			(0.000)	
Num.Obs.	425	425	425	425
R^2	0.432	0.524		
R^2 Adj.	0.416	0.509		
AIC	6252.3	1324.9		5348.0
BIC	6309.1	1385.7		5412.9
Log.Lik.	-3112.174	-647.465		-2658.014

Values are standardised
Values in brackets are p-values
* p < 0.1, ** p < 0.05, *** p < 0.01



■ **Figure 4.1: Standardised model estimates for weighted predictors**

While these models generally have plausible signs on the coefficients (with the most apparent exception of employment) the elasticities are implausibly low to key changes in population, public transport and school students. For example, using the weighted negative binomial GLM and assuming median values for all predictors:

- Increasing the distance-weighted population in the catchment from 10,000 to 20,000 person-km increases pedestrian demand by 8 pedestrians/day; increasing from 10,000 to 50,000 person-km increases pedestrian demand by 43 pedestrians/day.

- Increasing the distance-weighted employment from 10,000 person-km to 20,000 person-km reduces pedestrian demand by 18 pedestrians/day.
- Increasing the distance-weighted student enrolment from 100 person-km to 10,000 person-km reduces pedestrian demand by less than 1 pedestrians/day.
- Increasing public transport boardings from 1,000 to 10,000 person-km increases pedestrian demand by 117 pedestrians/day.

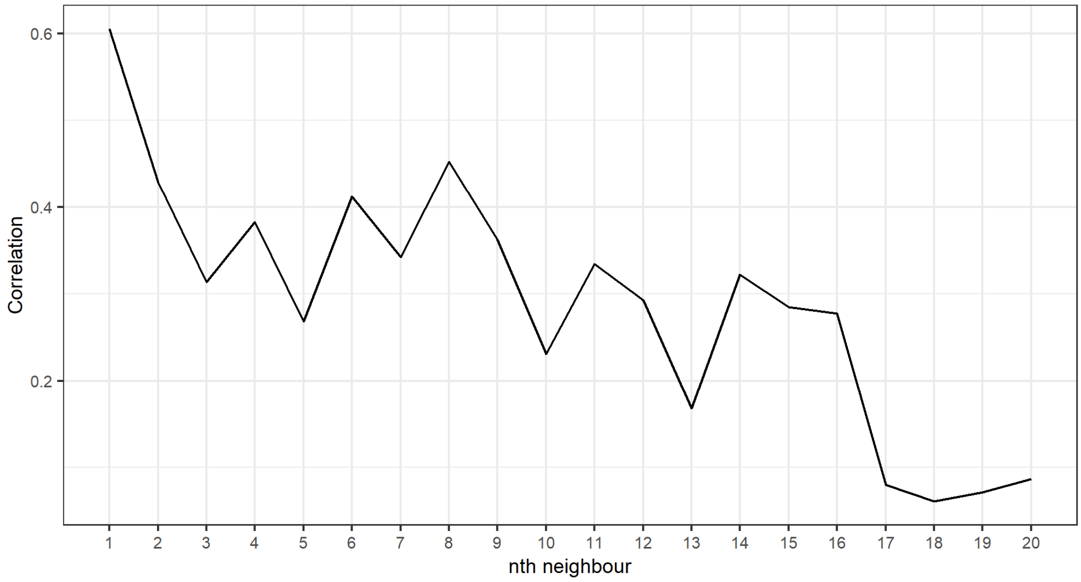
In considering these model sensitivities the distance weighting person-km unit can be thought of as this number of additional users immediately adjacent to the count site (that is, no distance decay applies). Under this interpretation an additional 10,000 residents adjacent to a count site would be expected to introduce far more than 8 additional pedestrian trips per day.

4.2.1 Spatial autocorrelation

Spatial interdependency was examined in two ways:

- dummy variables for LGAs
- spatial autocorrelation models.

Dummy variables for LGAs in the OLS and GLM models did not improve the model fit and so were excluded from the models described in Section 4.2. Spatial autocorrelation was examined by identifying the correlation between nearest neighbours; the correlation to the nearest neighbour was 0.61 and decreased to rapidly to 0.43 for the second closest neighbour (Figure 4.2). Moran's I test for spatial autocorrelation was statistically significant ($I = 0.1402$, $p < 0.00$).



■ **Figure 4.2: Nearest neighbour correlation**

Spatial autocorrelation models with error (SEM) and lag (SAR) terms were estimated using inverse distance weights and both the nearest two and nearest 20 points on the weighted negative binomial GLM. These models improved the AIC and log likelihood but did not significantly alter the model responsiveness to changes in key land use characteristics such as population, employment, public transport boardings and school students.

4.3 Conclusion

The models are limited by the weak relationship between the key land use characteristics of population, employment, public transport boardings and school students and pedestrian demand. Clearly, all else being equal, an increase in any one of these predictors would increase pedestrian activity and therefore increase demand at a facility. However, this effect cannot be drawn out from the data. Possible explanations for this shortcoming are:

- The pedestrian counts are too "noisy": most are single day counts obtained over a wide range or years and months such that there are likely many other factors that are influencing these short-period counts that are not included in the models.
- The land use attributes are spatially too coarse to adequately capture the very *local* influences on pedestrian demand.

While these direct land use attributes are insignificant there are indirect proxies which are both significant and of plausible sign and magnitude, for which the most important is the commuting mode share for walking trips. The final model is shown in

Table 4.3 and uses a negative binomial GLM error structure and predictors optimised to AIC using stepwise regression. The key characteristics of this model are:

- Higher commuting walking to work and public transport mode shares are associated with higher pedestrian demand
- Higher household income is associated with slightly lower pedestrian demand
- Higher median age is associated with higher pedestrian demand
- Proximity to a school is associated with higher pedestrian demand
- Proximity to, and higher areas of, parks or retail facilities are associated with higher pedestrian demand
- Facilities such as footpaths (at mid-blocks), roundabouts and sign-controlled intersections are associated with lower pedestrian demand than paths outside of road corridors, signalised intersections and zebra crossings.

The marginal effects for this model are shown in Appendix C and generally appear to be of a plausible magnitude. For example, for a shared path and average parameters:

- Increasing the walk mode share to work from 10% to 20% will increase weekday pedestrian demand by 433 pedestrian per day
- Increasing public transport mode share to work from 10% to 20% will increase weekday pedestrian demand by 190 pedestrians per day
- Increasing the distance of the nearest school from 300 m to 1,000 m will reduce weekday pedestrian demand by just under 100 pedestrians per day
- Areas with a median weekly household income of $1,000 will have pedestrian demand around 117 higher than areas with median weekly demand of $2,000.
- Areas with a median age of 45 will have pedestrian demand around 120 higher than areas with a median age of 35.

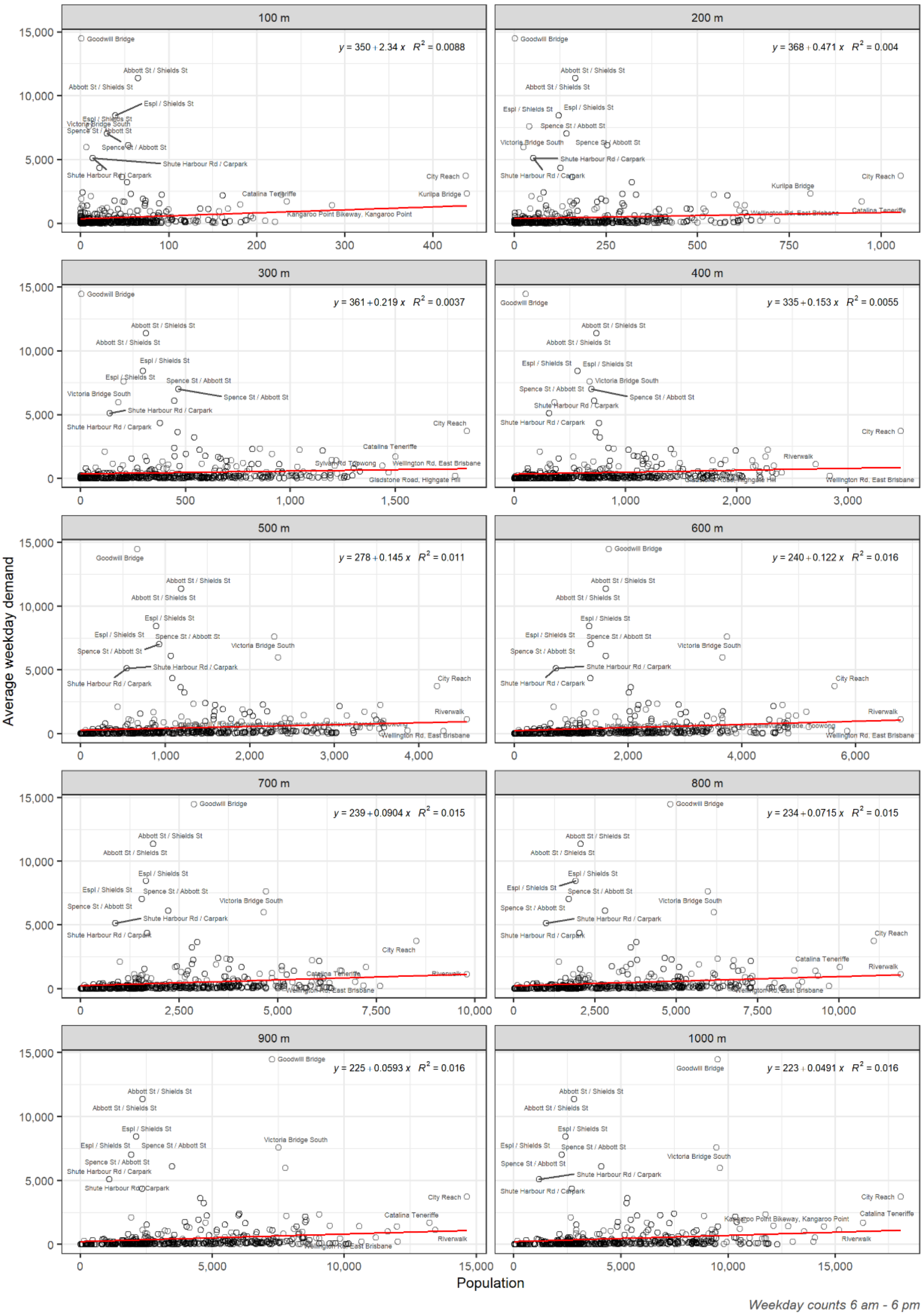
Table 4.3: Final model

	GLM: NB
(Intercept)	3.649***
	(0.000)
walk.share.wgt	9.157***
	(0.000)
pt.share.wgt	6.421***
	(0.000)
hh.income.wgt	-0.001***
	(0.001)
median.age.wgt	0.056***
	(0.000)
dist.school	-0.703***
	(0.000)
park.km2.wgt	2.7E-7
	(0.117)
retail.km2.wgt	6.69E-6***
	(0.001)
dummyFootpath	-1.155***
	(0.000)
dummyRbt	-0.432**
	(0.033)
dummySign	-0.753***
	(0.000)
Num.Obs.	425
AIC	5310.4
BIC	5359.0
Log.Lik.	-2643.203

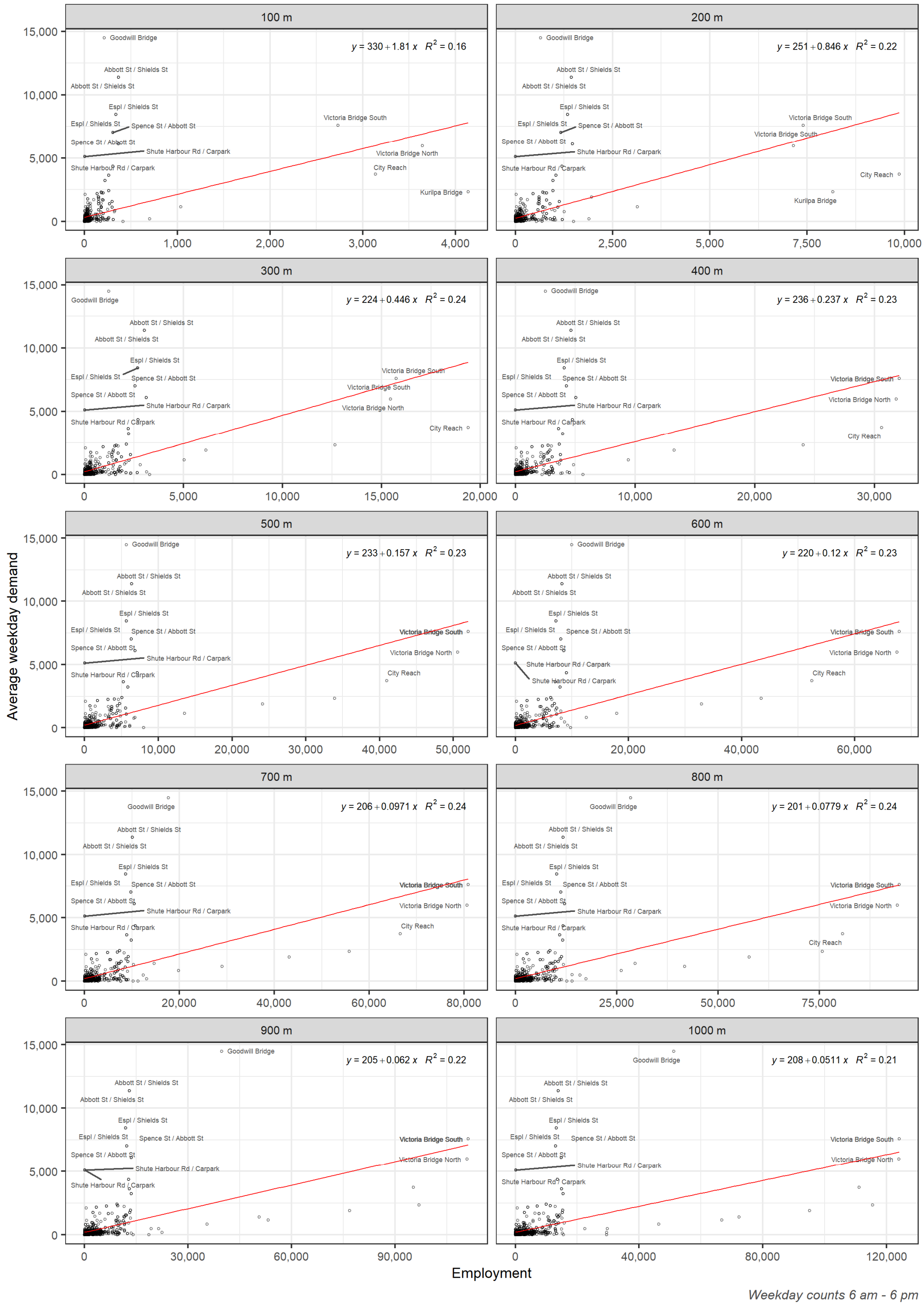
Values in brackets are p-values
* p < 0.1, ** p < 0.05, *** p < 0.01

Appendix A: Outliers

Population

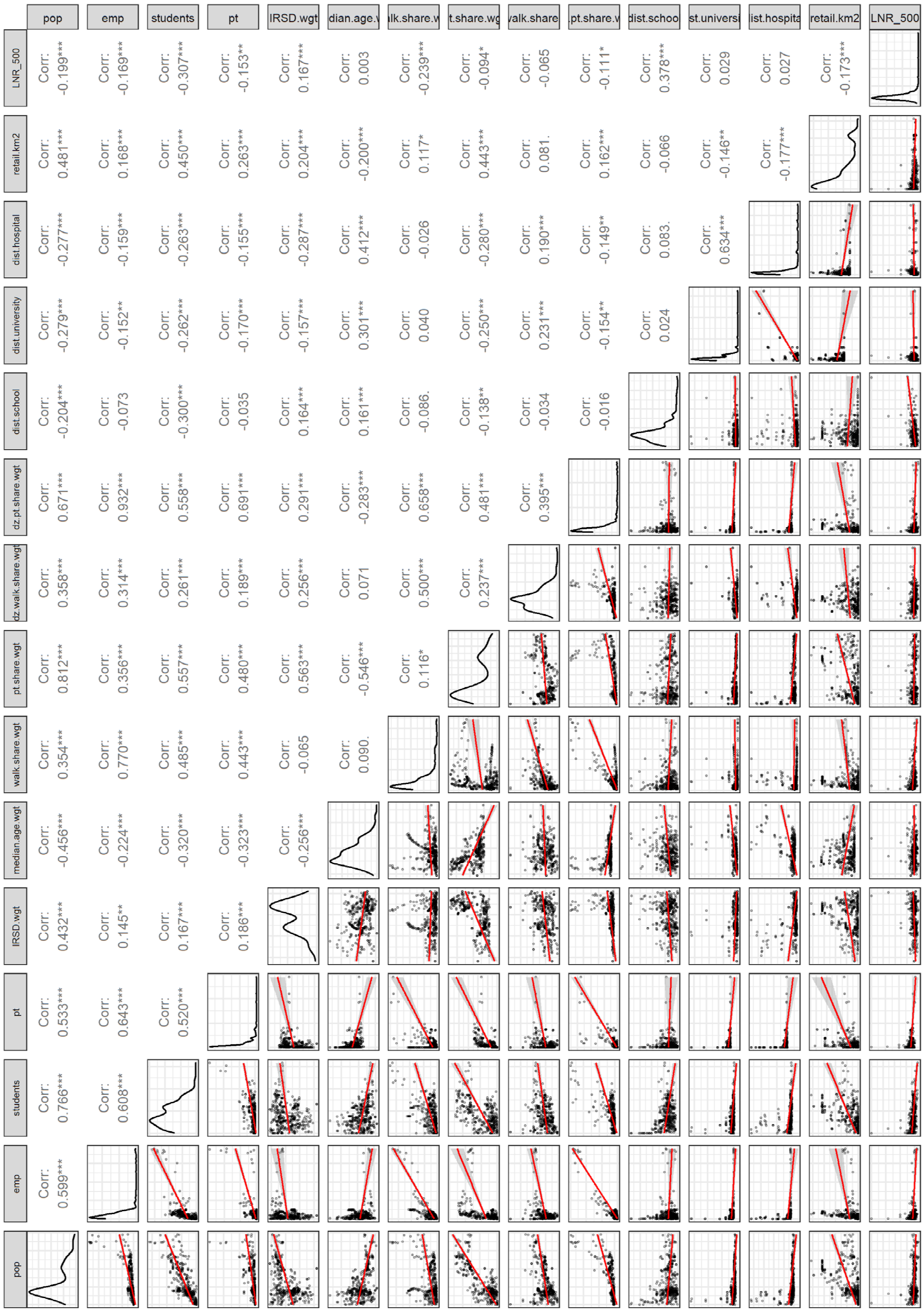


Employment


100 m
y = 330 + 1.81 x R^2 = 0.16
200 m
y = 251 + 0.846 x R^2 = 0.22
300 m
y = 224 + 0.446 x R^2 = 0.24
400 m
y = 236 + 0.237 x R^2 = 0.23
500 m
y = 233 + 0.157 x R^2 = 0.23
600 m
y = 220 + 0.12 x R^2 = 0.23
700 m
y = 206 + 0.0971 x R^2 = 0.24
800 m
y = 201 + 0.0779 x R^2 = 0.24
900 m
y = 205 + 0.062 x R^2 = 0.22
1000 m
y = 208 + 0.0511 x R^2 = 0.21
Goodwill Bridge
Abbott St / Shields St
Espl / Shields St
Spence St / Abbott St
Shute Harbour Rd / Carpark
Victoria Bridge South
Victoria Bridge North
City Reach
Kurilpa Bridge
Average weekday demand
Employment
Weekday counts 6 am - 6 pm

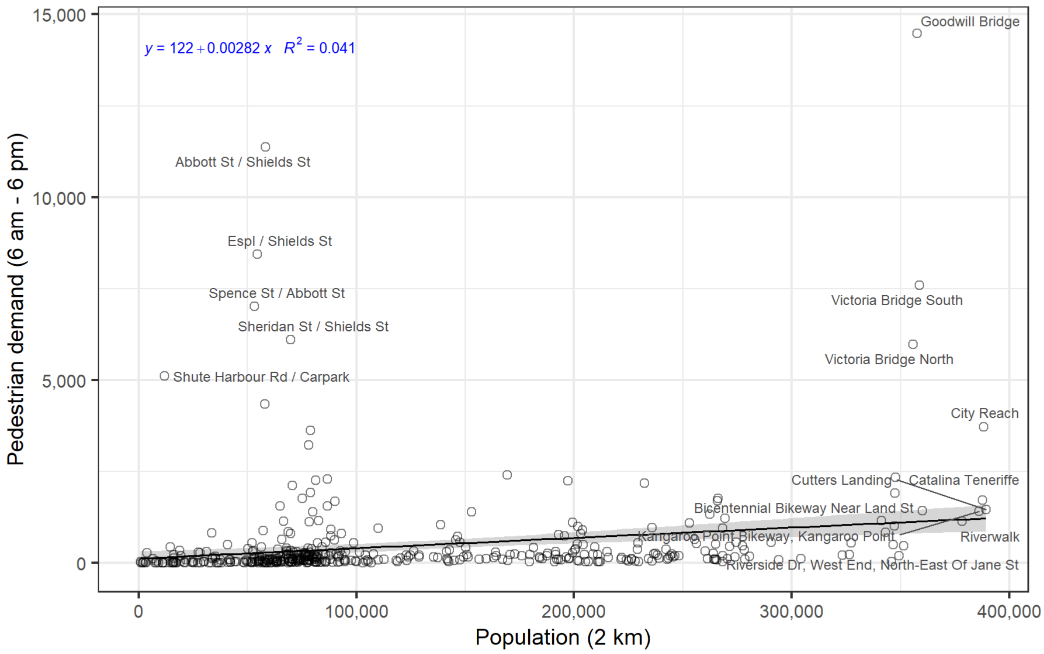
Appendix B: Correlations

B.1 Pairs plot



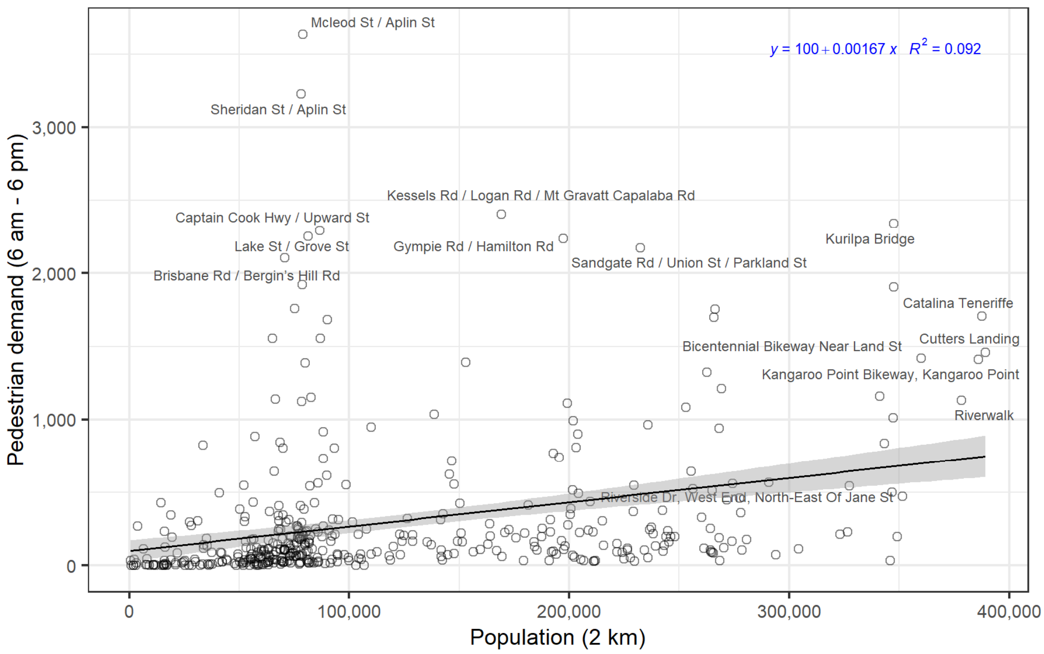
B.2 Population

Full dataset



Line is linear regression and shaded area is 95% confidence interval

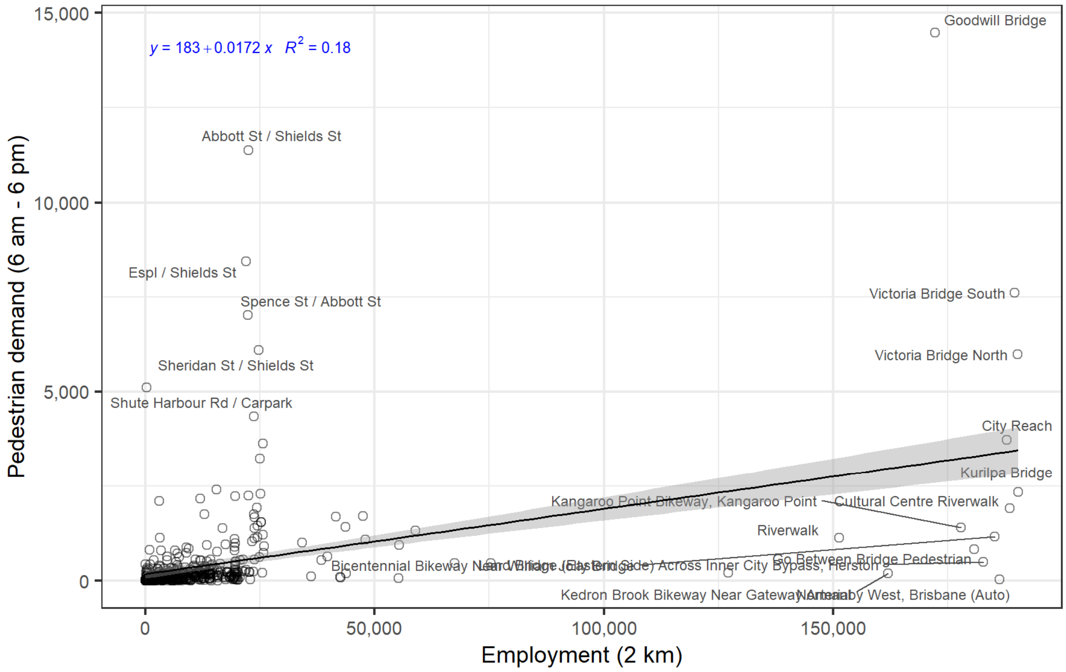
Removal of influential outliers



Line is linear regression and shaded area is 95% confidence interval

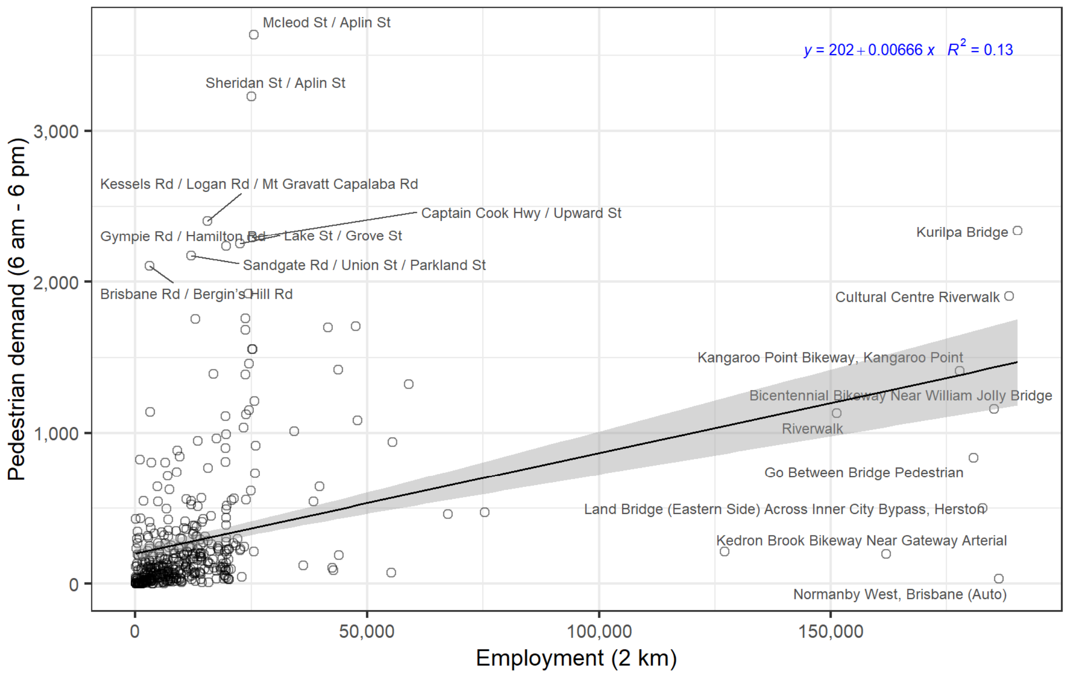
B.3 Employment

Full dataset



Line is linear regression and shaded area is 95% confidence interval

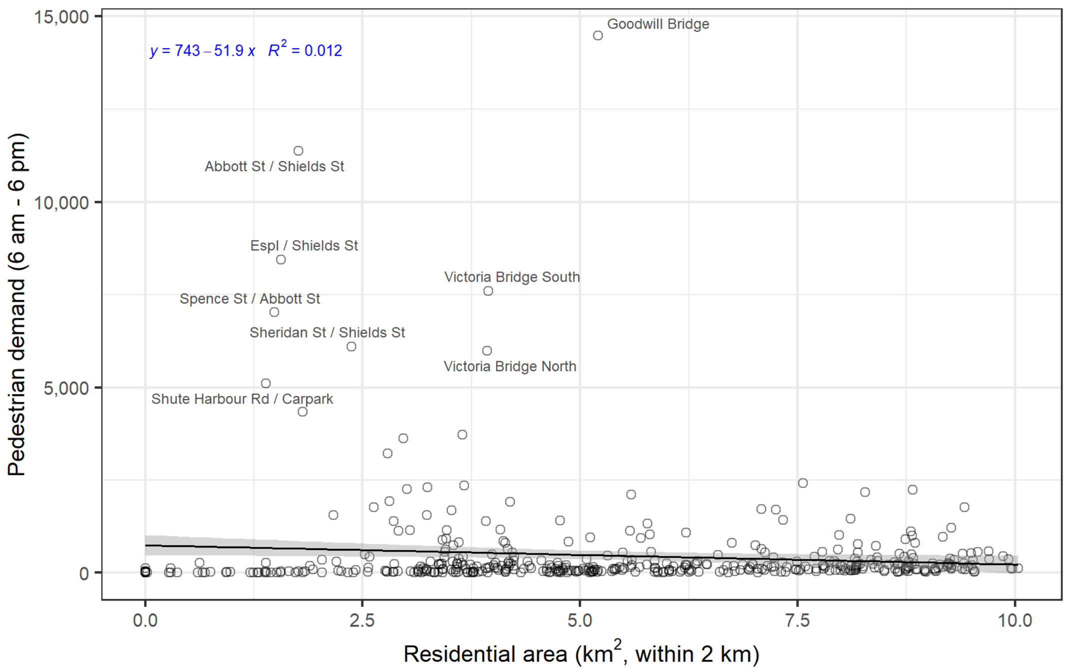
Removal of influential outliers



Line is linear regression and shaded area is 95% confidence interval

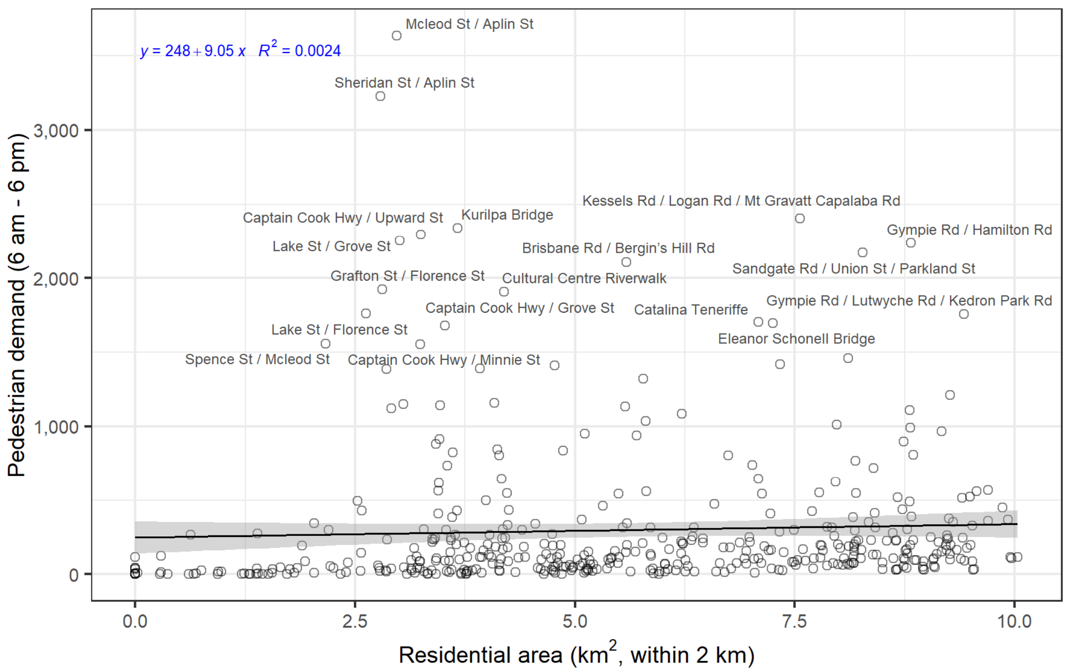
B.4 Residential zoning

Full dataset



Line is linear regression and shaded area is 95% confidence interval

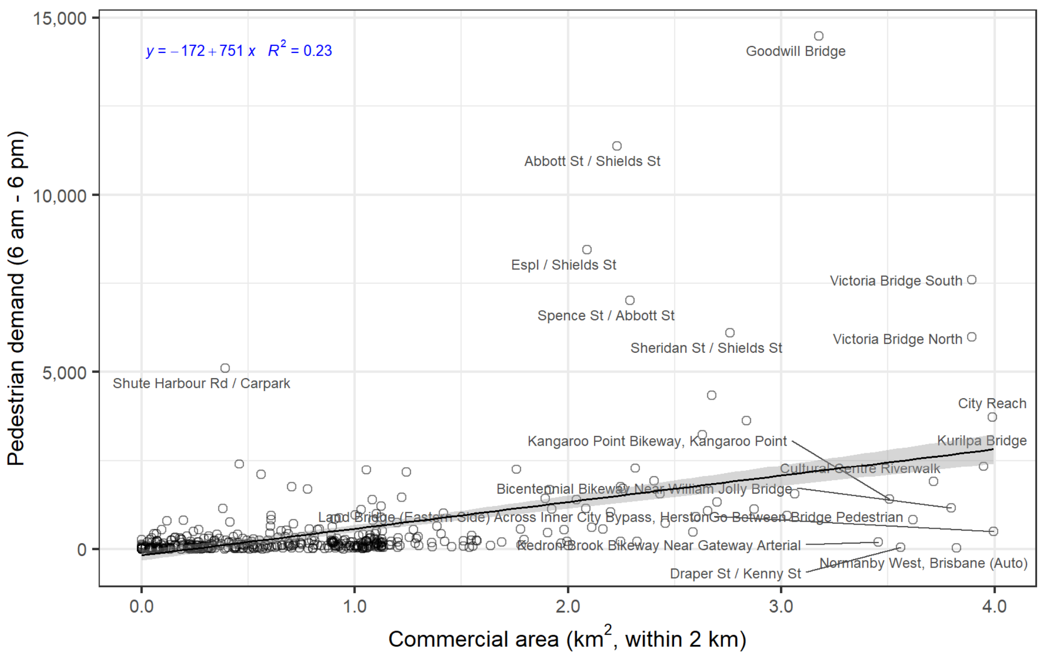
Removal of influential outliers



Line is linear regression and shaded area is 95% confidence interval

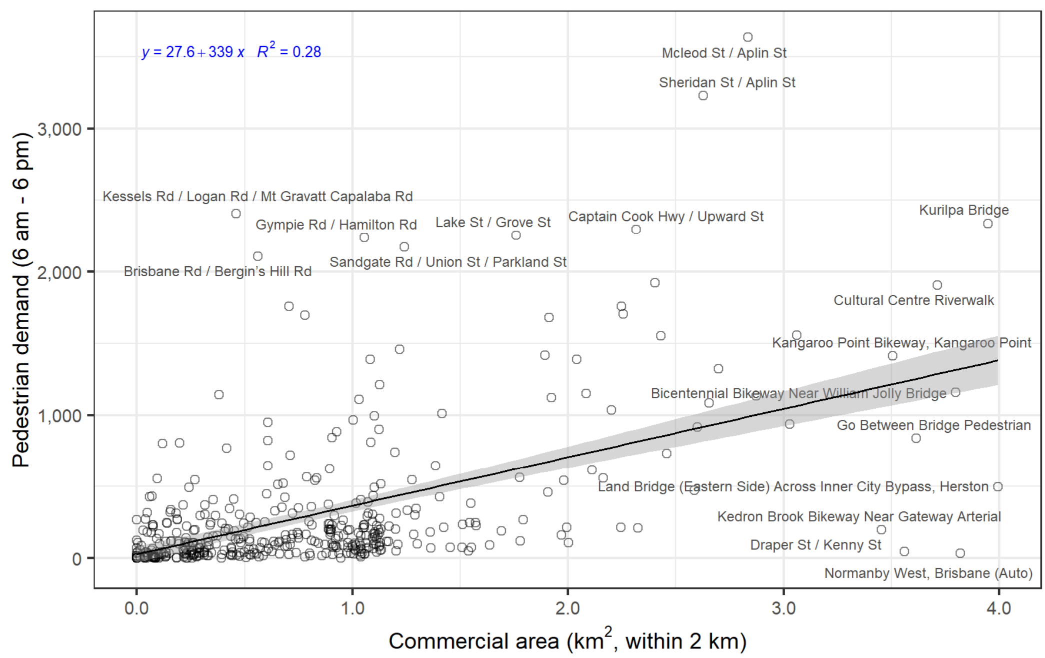
B.5 Commercial zoning

Full dataset



Line is linear regression and shaded area is 95% confidence interval

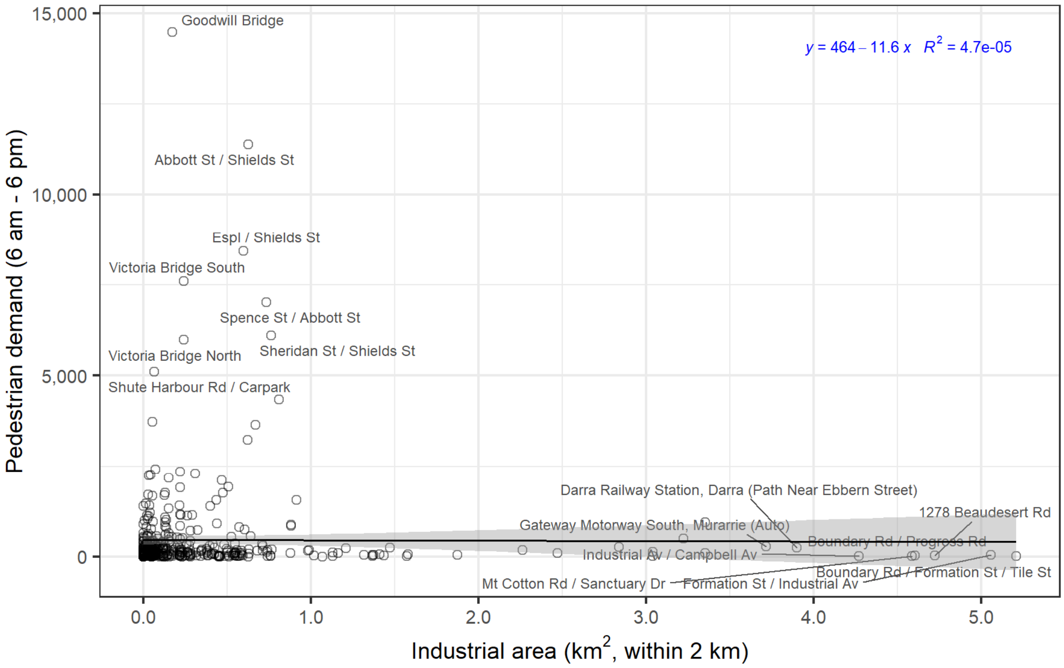
Removal of influential outliers



Line is linear regression and shaded area is 95% confidence interval

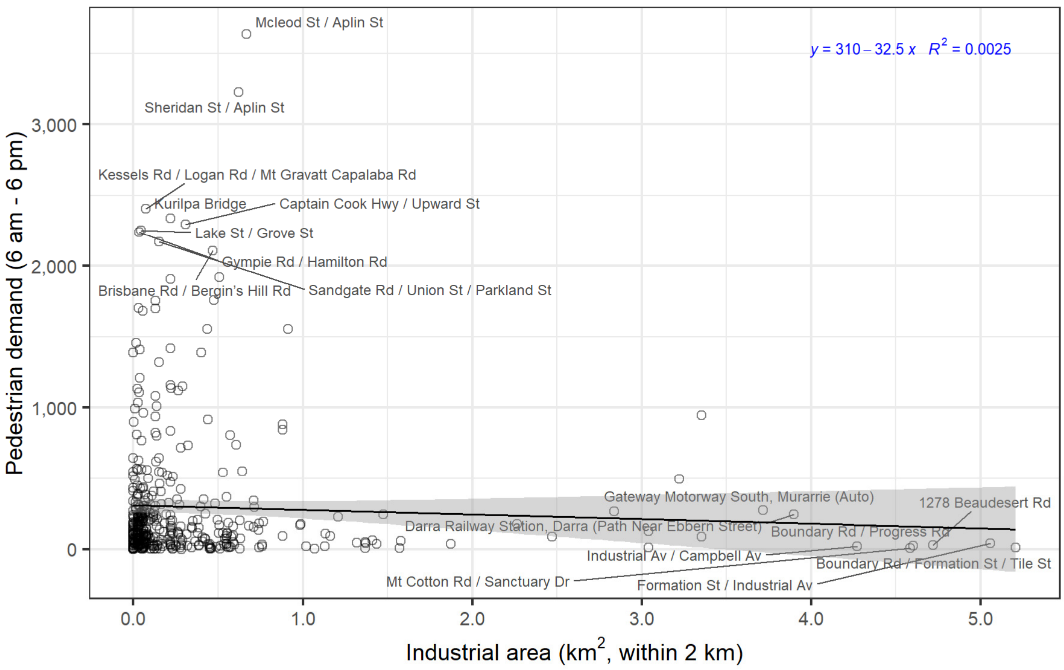
B.6 Industrial zoning

Full dataset



Line is linear regression and shaded area is 95% confidence interval

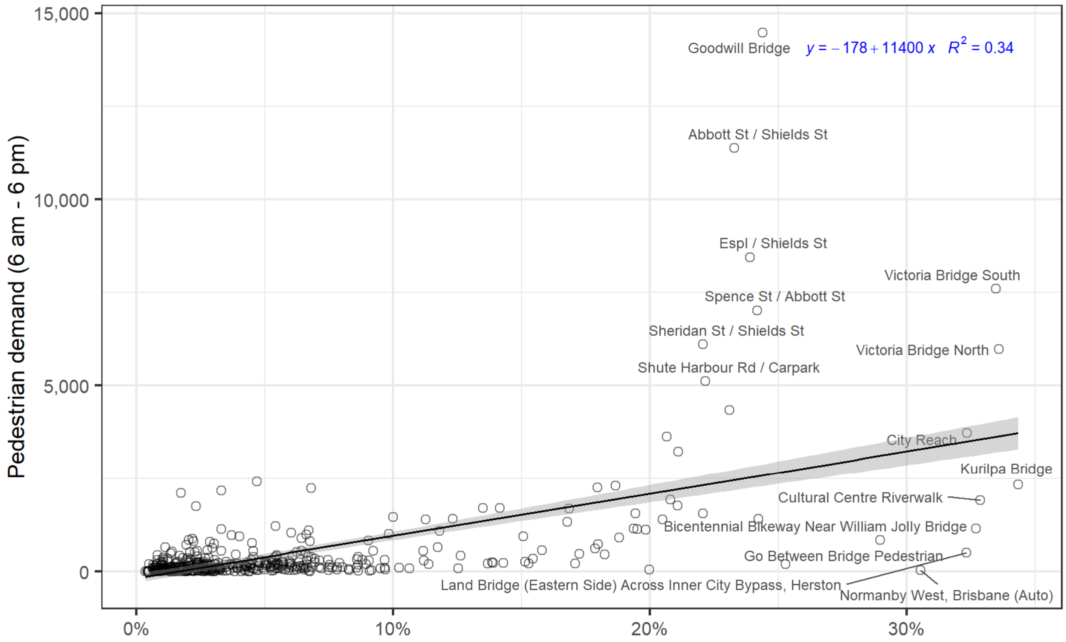
Removal of influential outliers



Line is linear regression and shaded area is 95% confidence interval

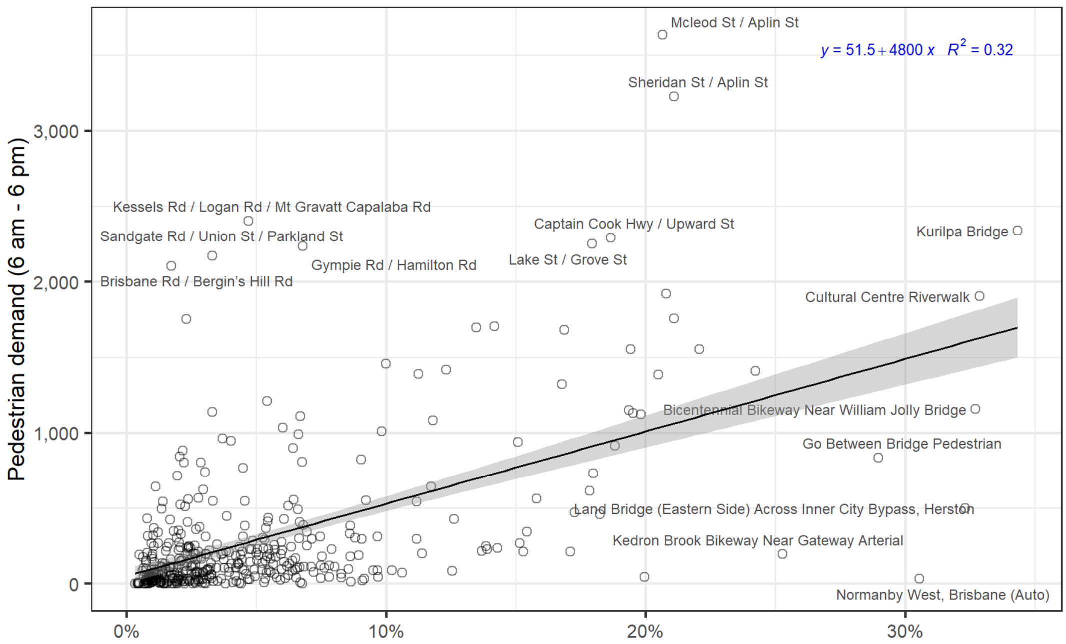
B.7 Walk to work mode shares (based on home location)

Full dataset



Line is linear regression and shaded area is 95% confidence interval

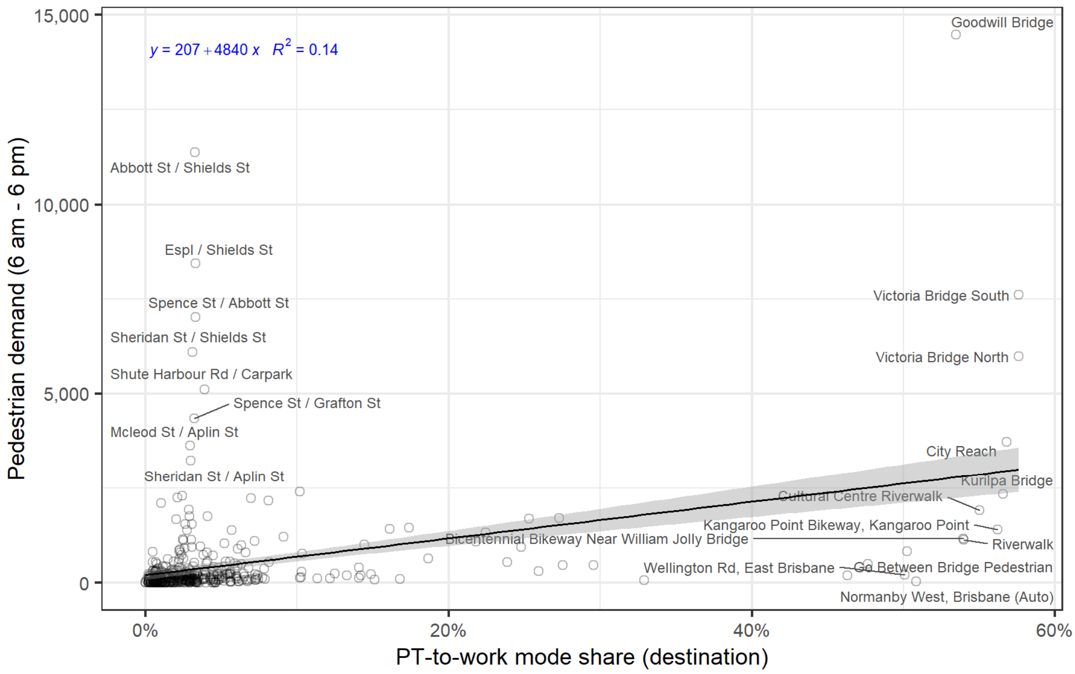
Removal of influential outliers



Line is linear regression and shaded area is 95% confidence interval

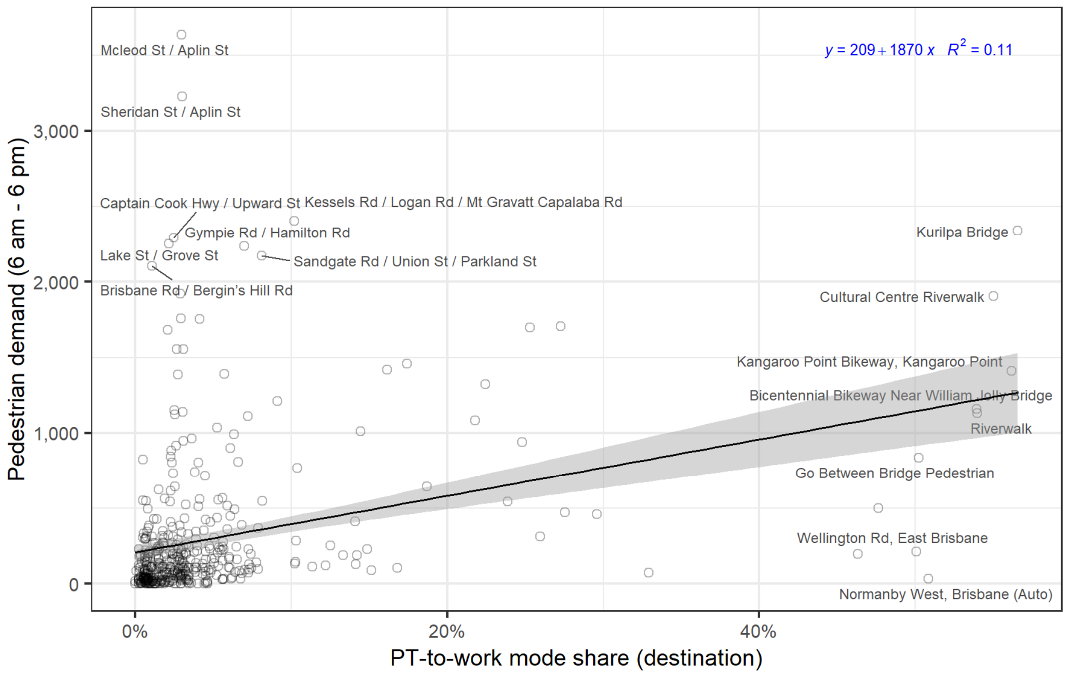
B.8 Public transport to work mode shares (based on home location)

Full dataset



Line is linear regression and shaded area is 95% confidence interval

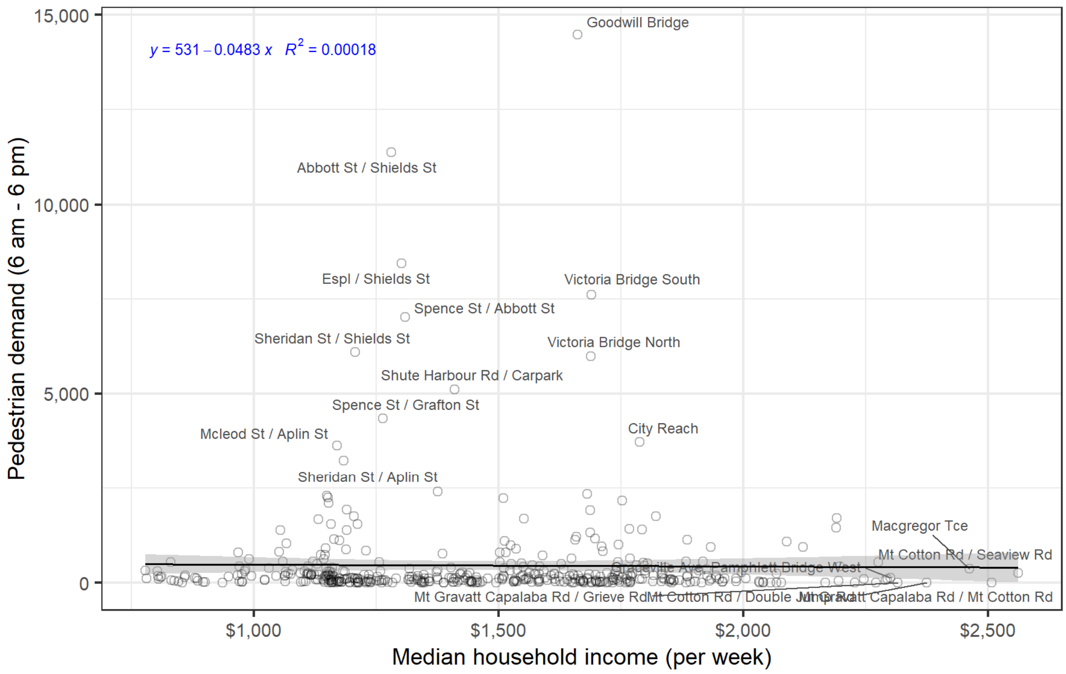
Removal of influential outliers



Line is linear regression and shaded area is 95% confidence interval

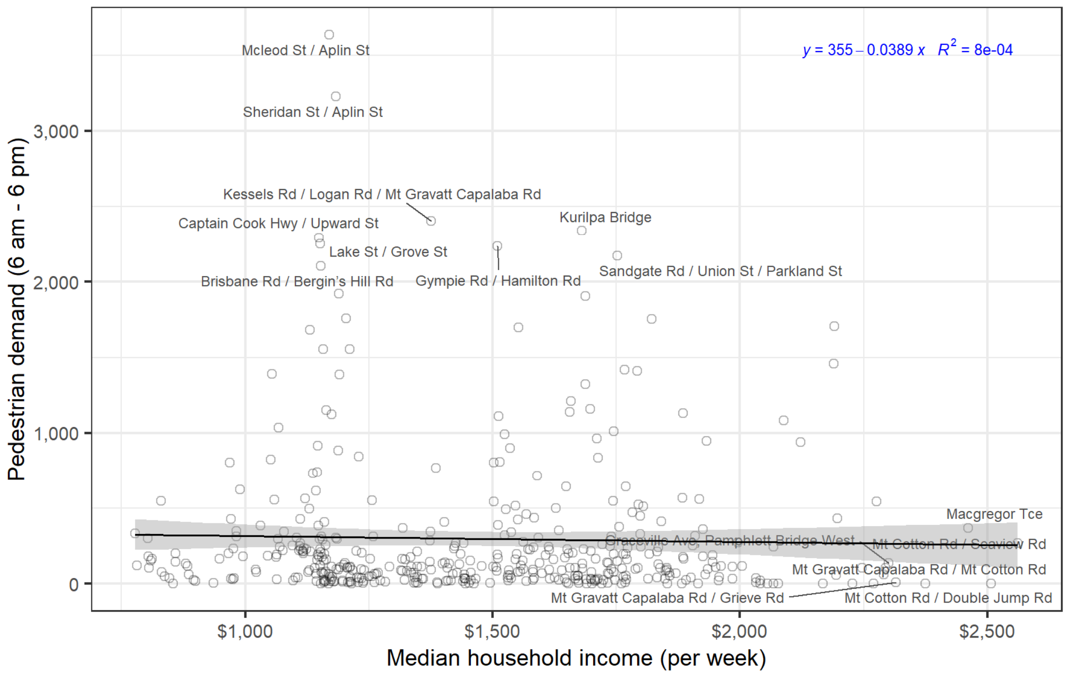
B.9 Household income

Full dataset



Line is linear regression and shaded area is 95% confidence interval

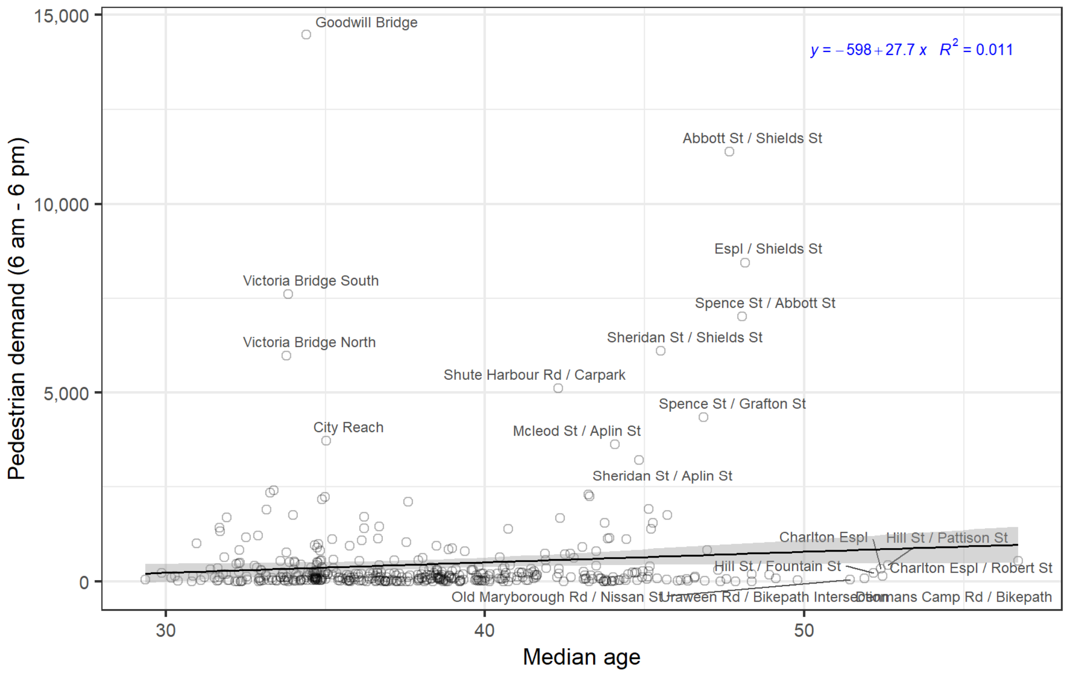
Removal of influential outliers



Line is linear regression and shaded area is 95% confidence interval

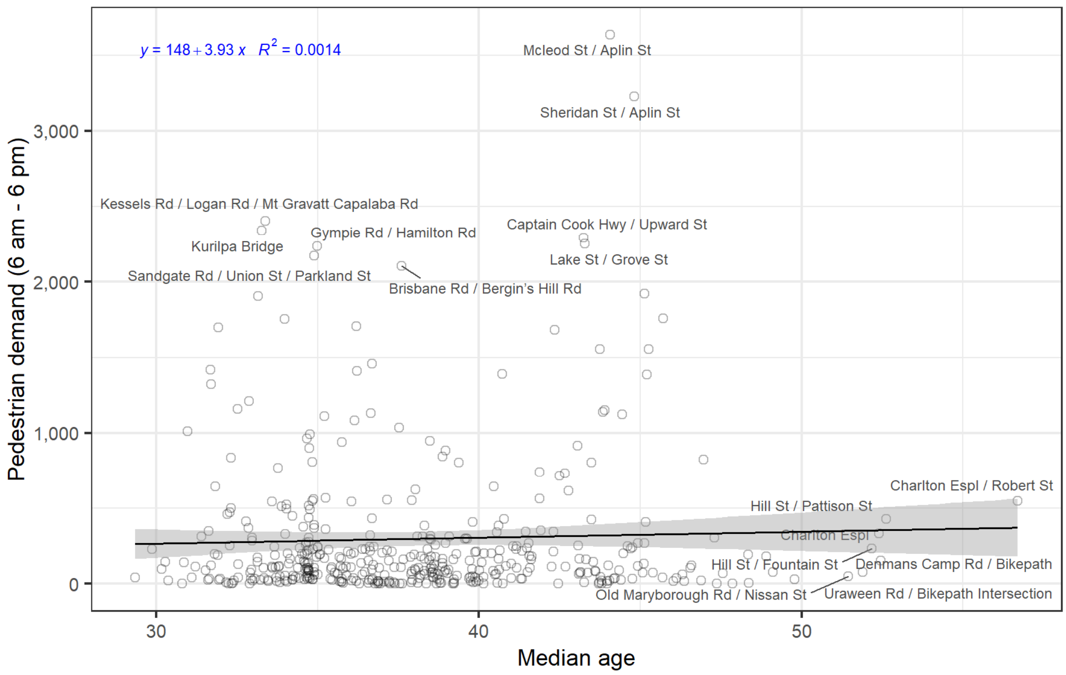
B.10 Median age

Full dataset



Line is linear regression and shaded area is 95% confidence interval

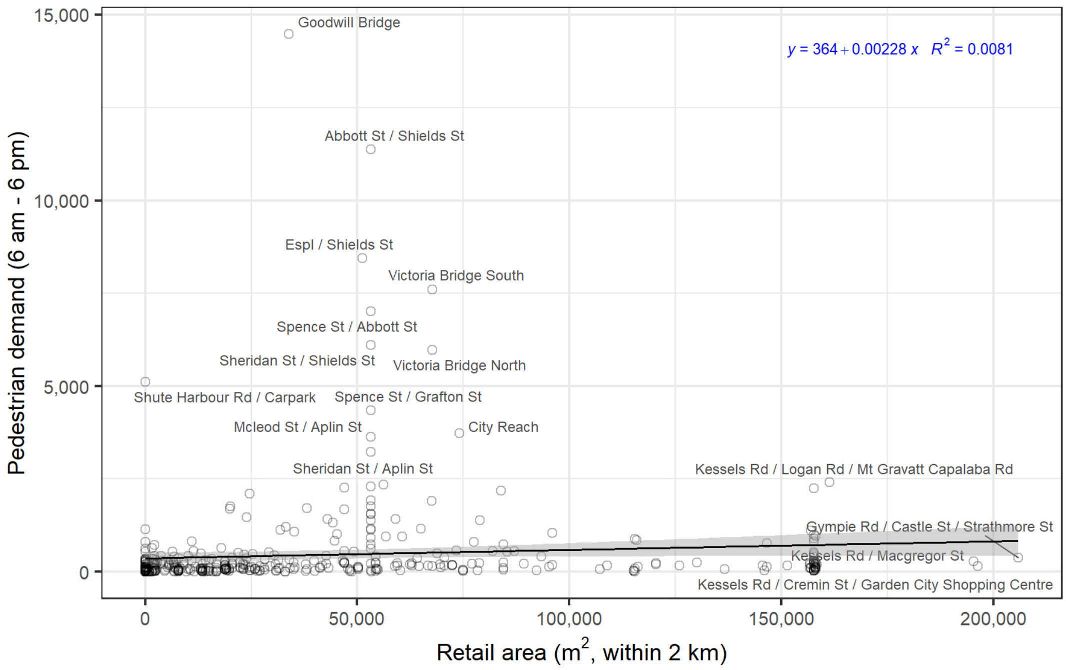
Removal of influential outliers



Line is linear regression and shaded area is 95% confidence interval

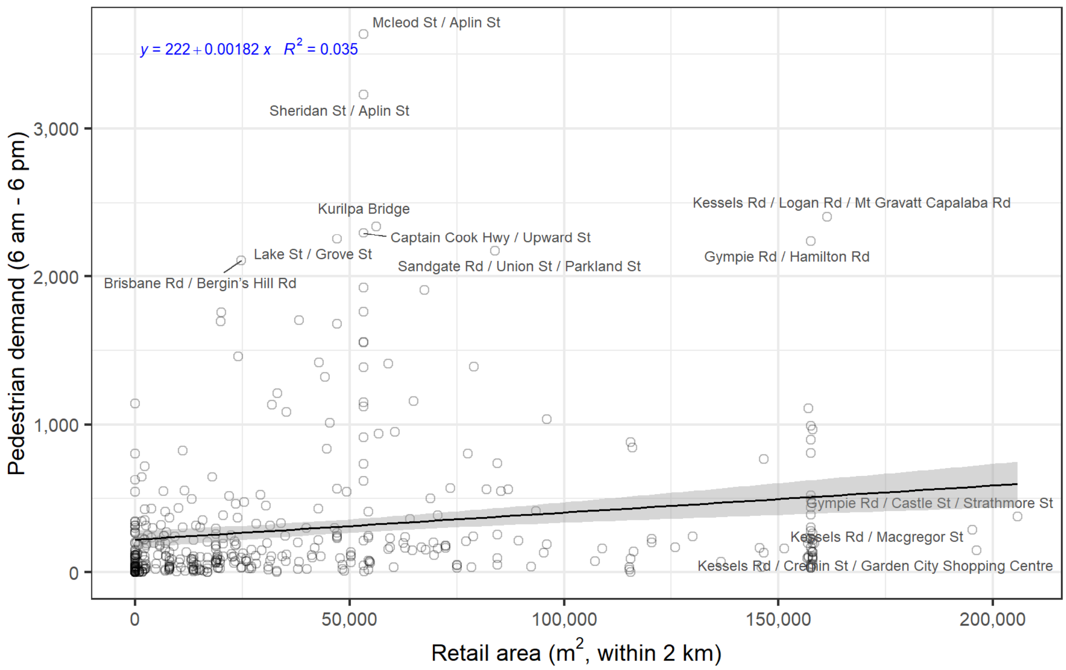
B.11 Retail floorspace

Full dataset



Line is linear regression and shaded area is 95% confidence interval

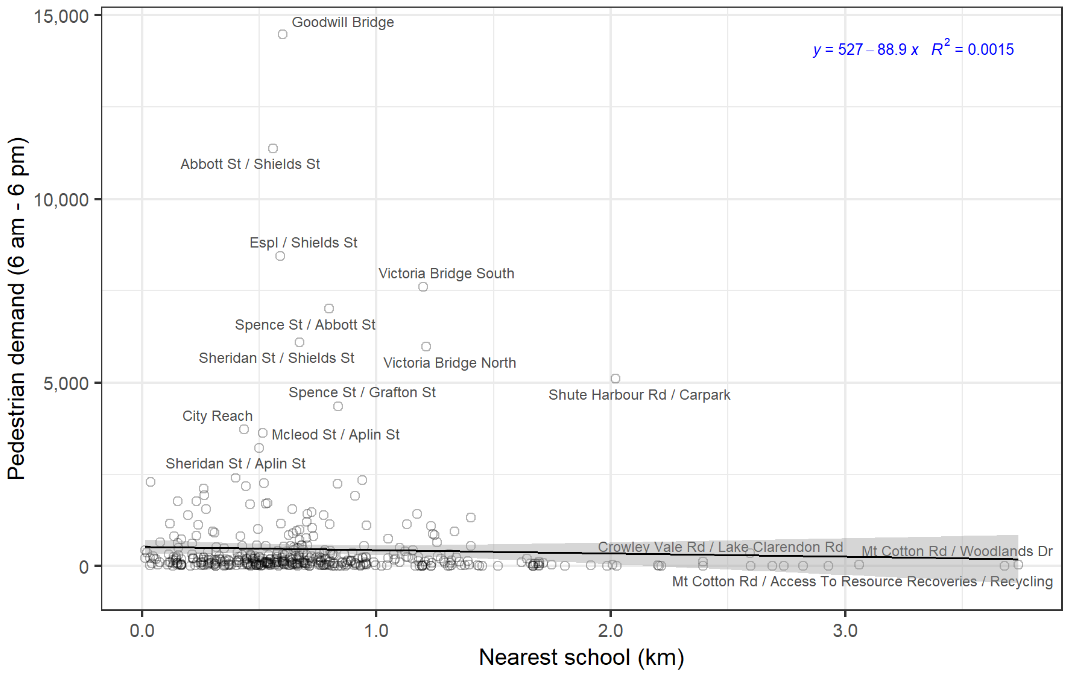
Removal of influential outliers



Line is linear regression and shaded area is 95% confidence interval

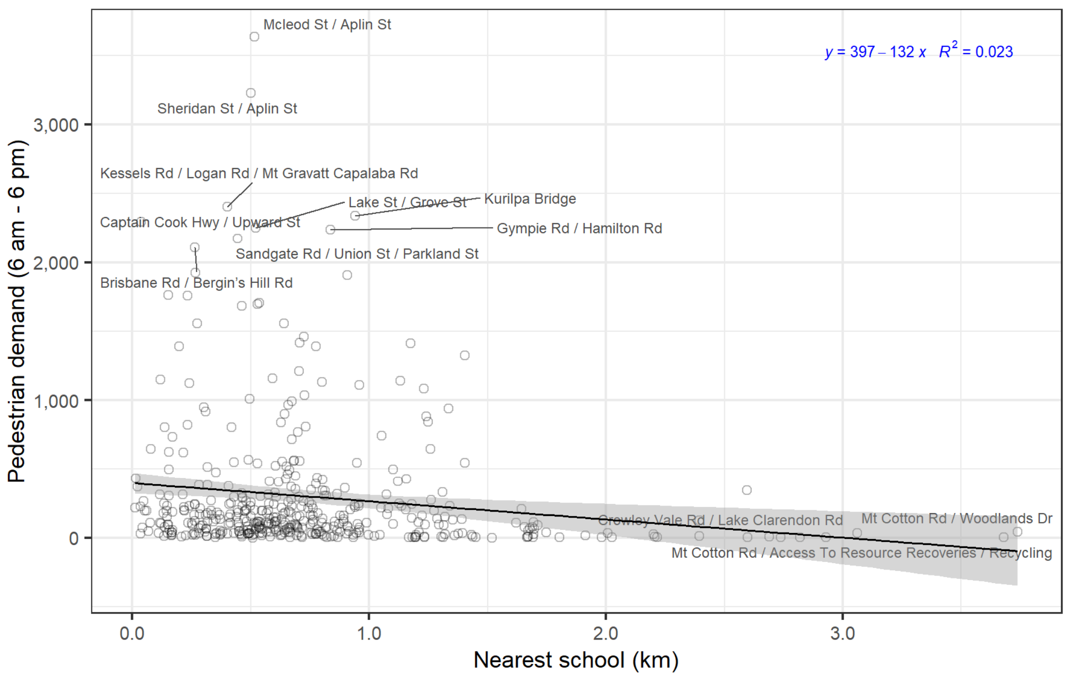
B.12 Nearest school

Full dataset



Line is linear regression and shaded area is 95% confidence interval

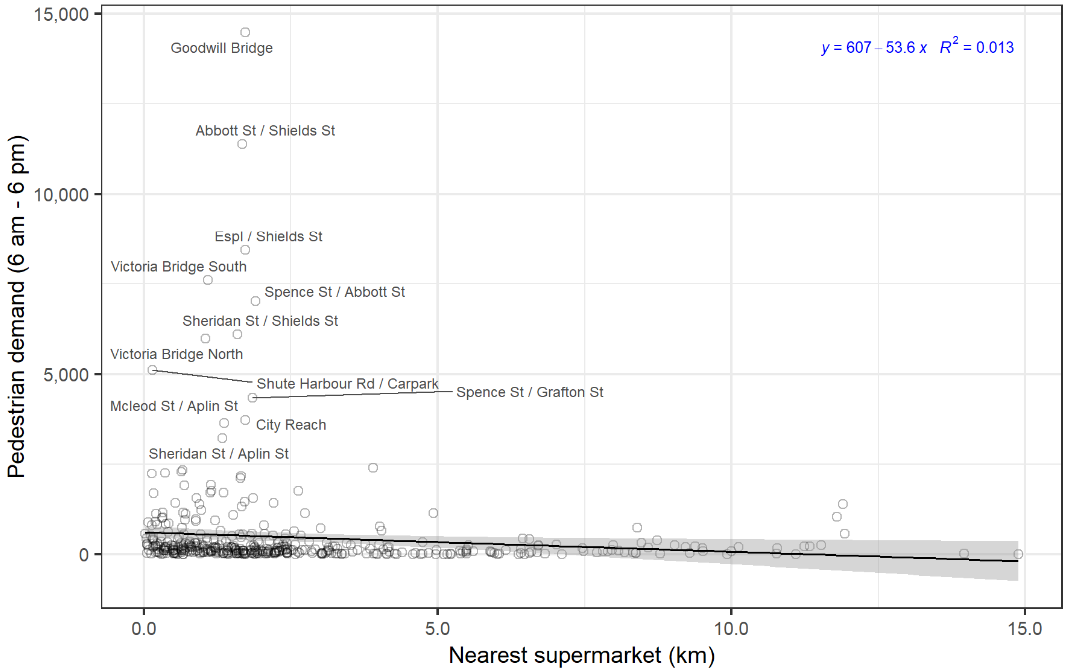
Removal of influential outliers



Line is linear regression and shaded area is 95% confidence interval

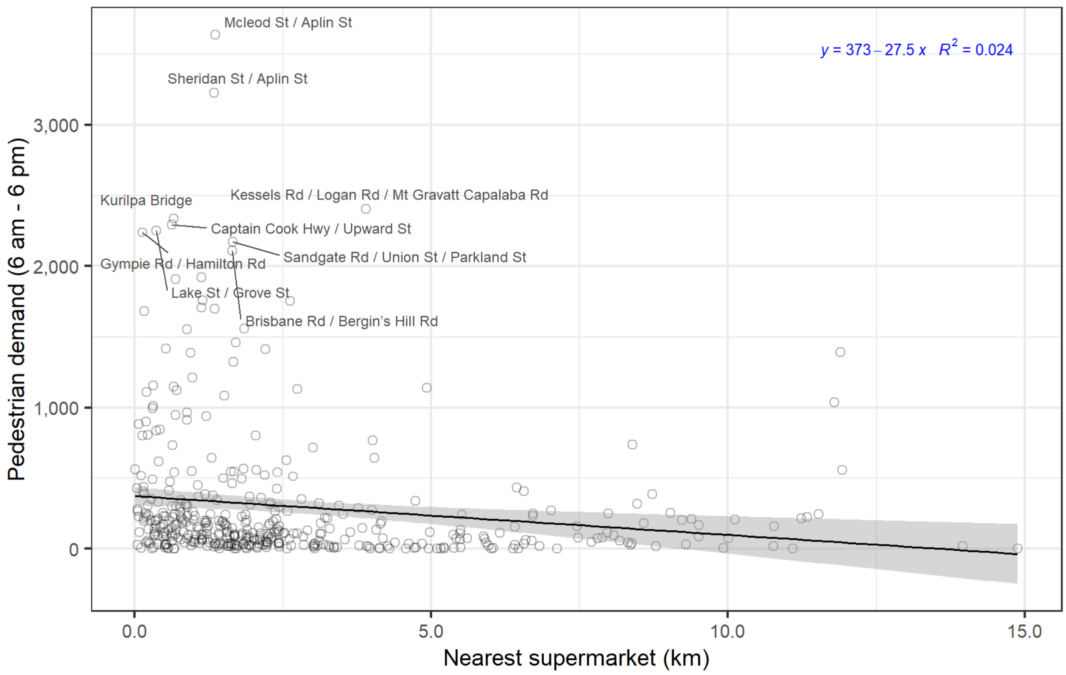
B.13 Nearest supermarket

Full dataset



Line is linear regression and shaded area is 95% confidence interval

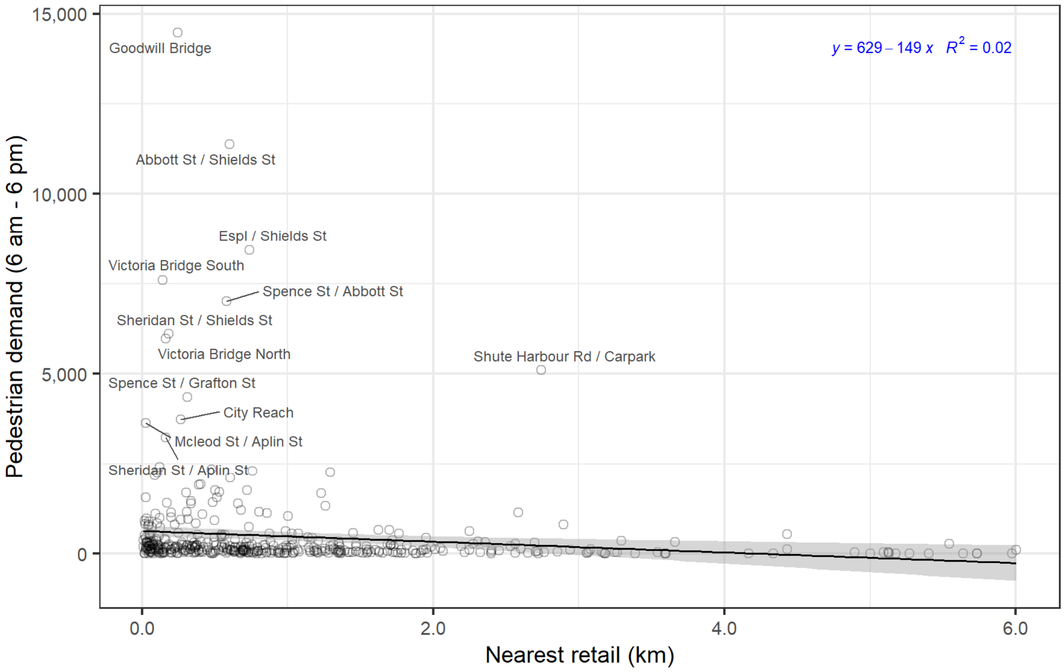
Removal of influential outliers



Line is linear regression and shaded area is 95% confidence interval

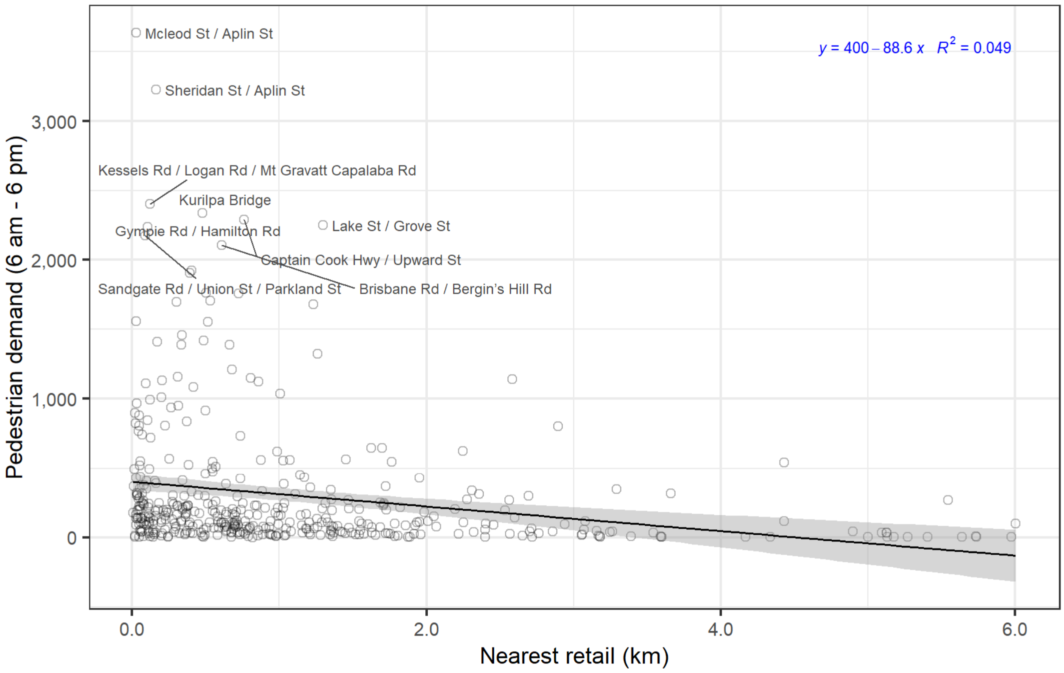
B.14 Nearest retail

Full dataset



Line is linear regression and shaded area is 95% confidence interval

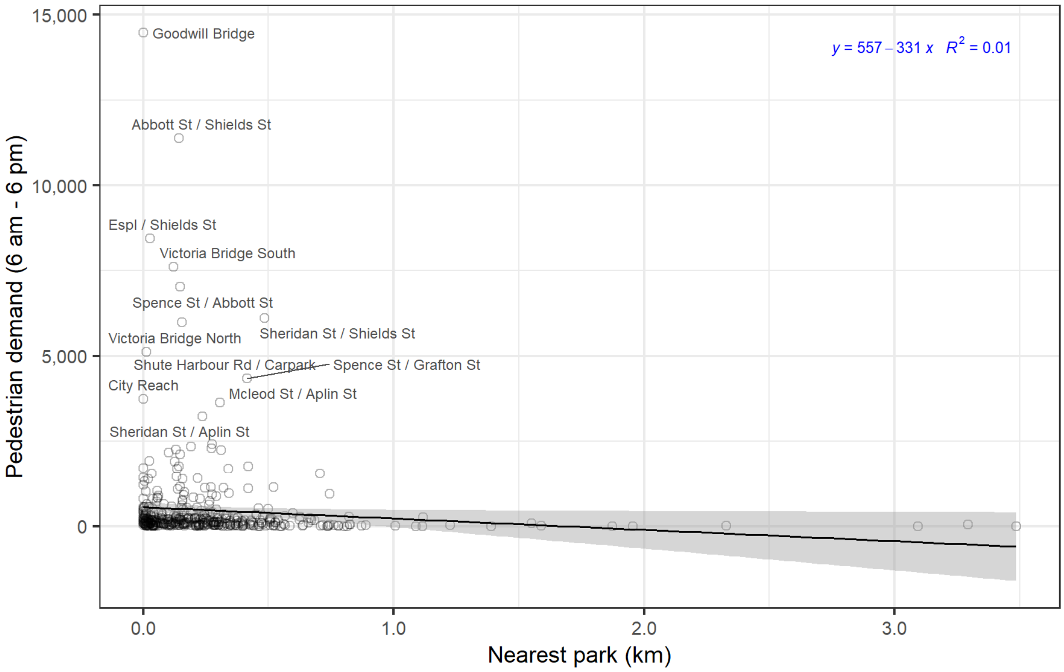
Removal of influential outliers



Line is linear regression and shaded area is 95% confidence interval

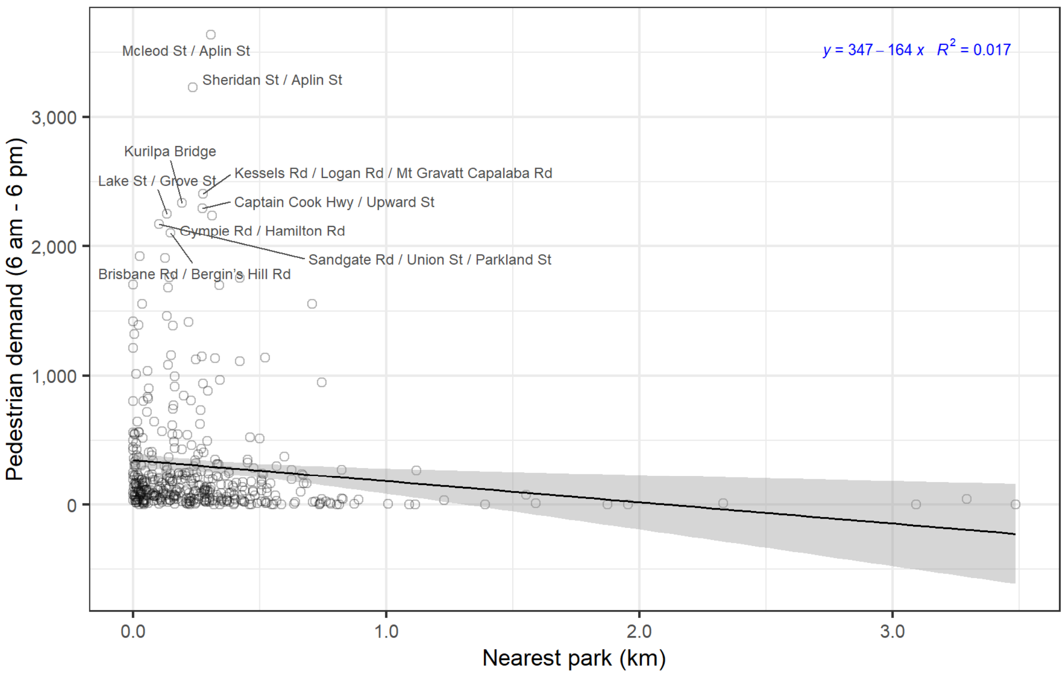
B.15 Nearest park

Full dataset



Line is linear regression and shaded area is 95% confidence interval

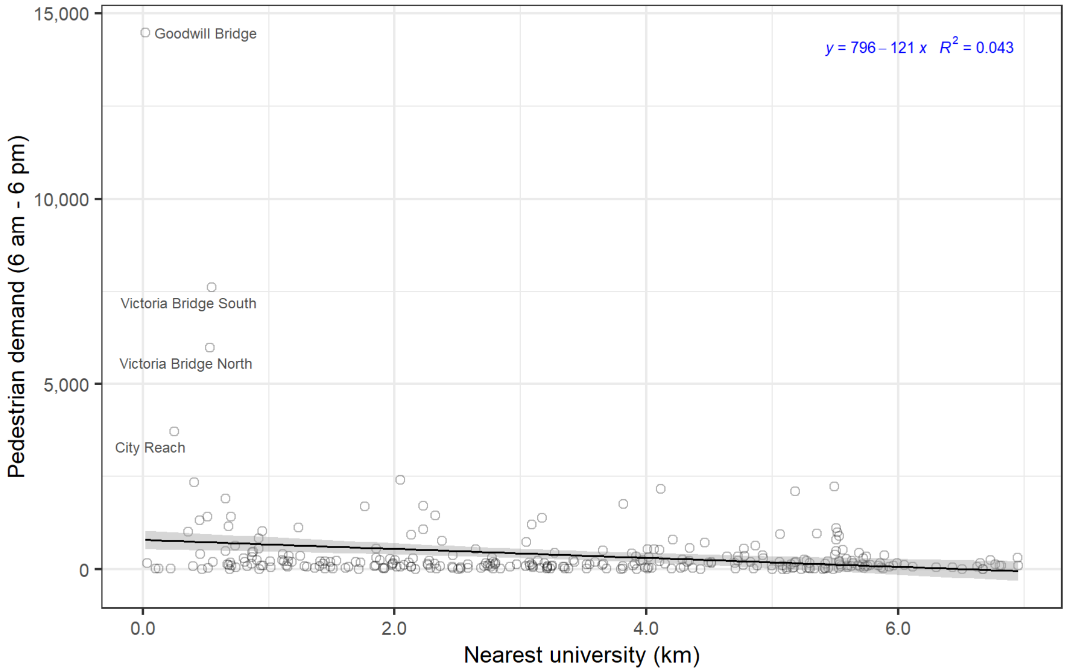
Removal of influential outliers



Line is linear regression and shaded area is 95% confidence interval

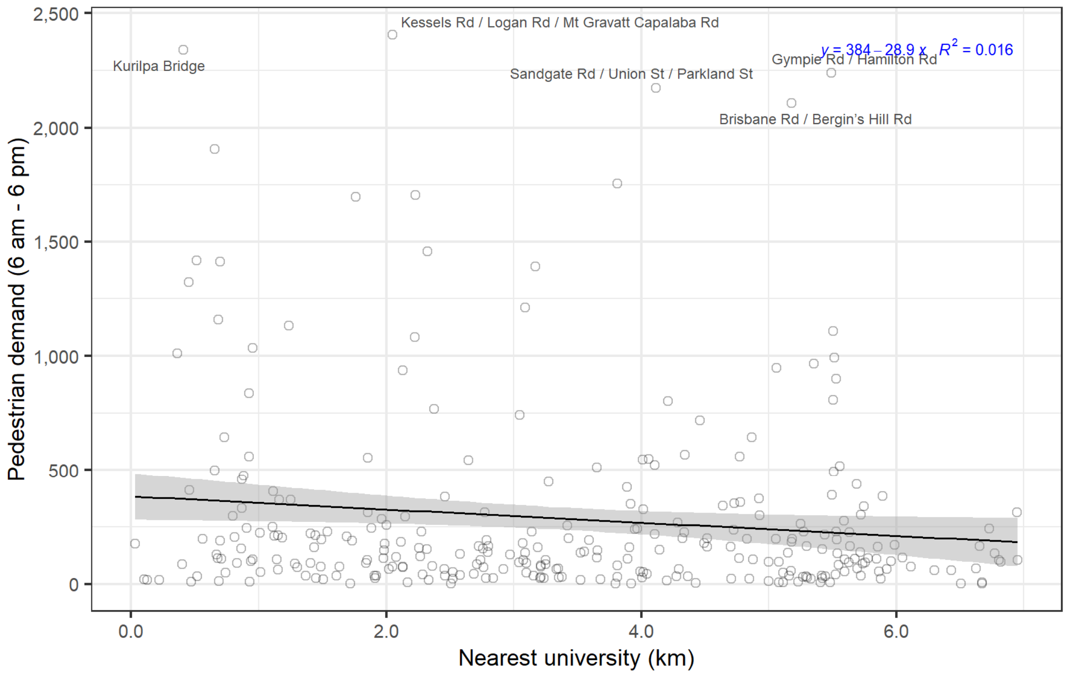
B.16 Nearest university

Full dataset



Line is linear regression and shaded area is 95% confidence interval

Removal of influential outliers



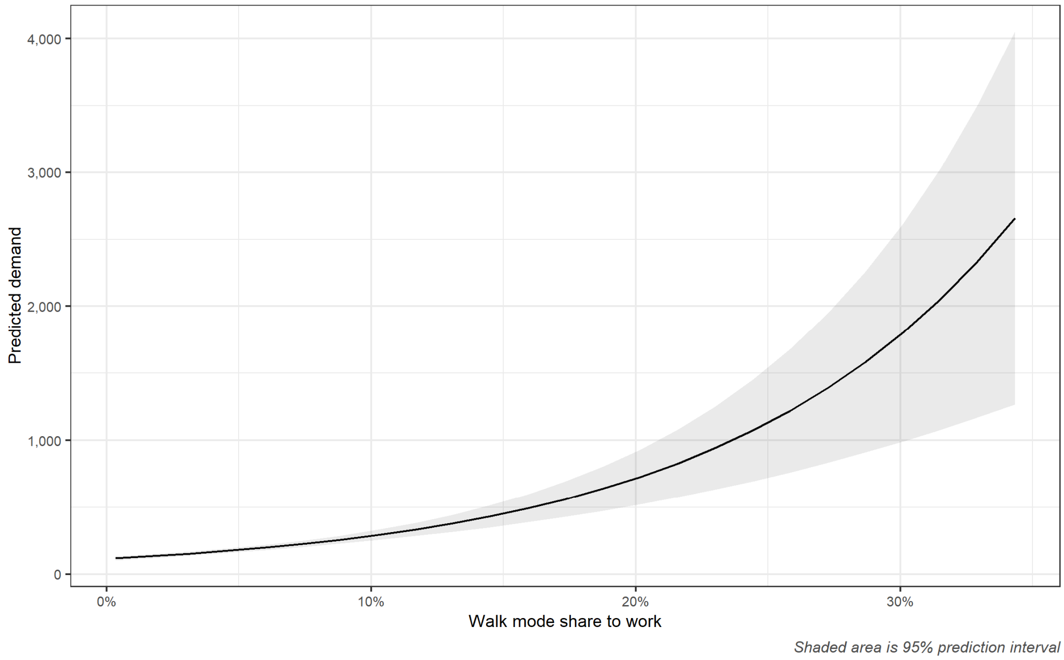
Line is linear regression and shaded area is 95% confidence interval

Appendix C: Marginal effects

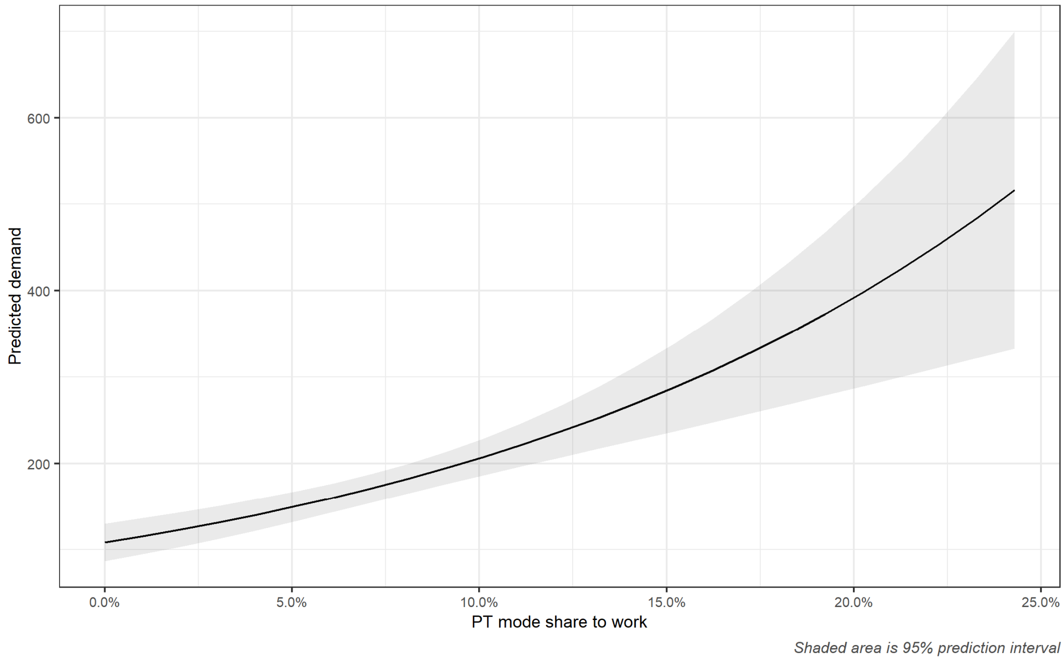
The marginal effects in these graphs are for the final model in

Table 4.3.

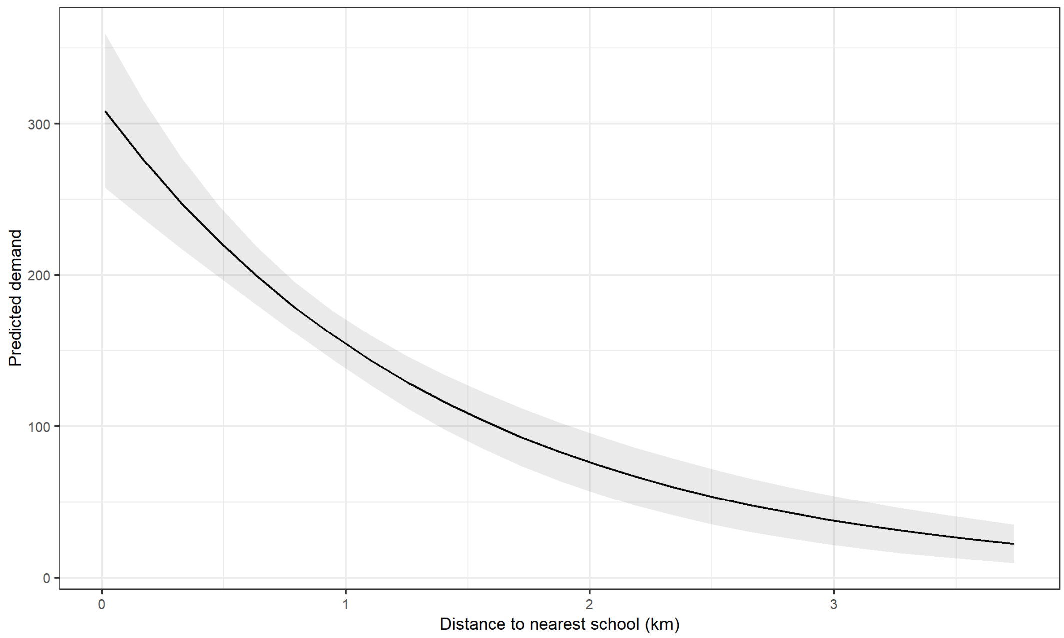
Walk (as a sole mode) mode share for commuting



Public transport mode share for commuting

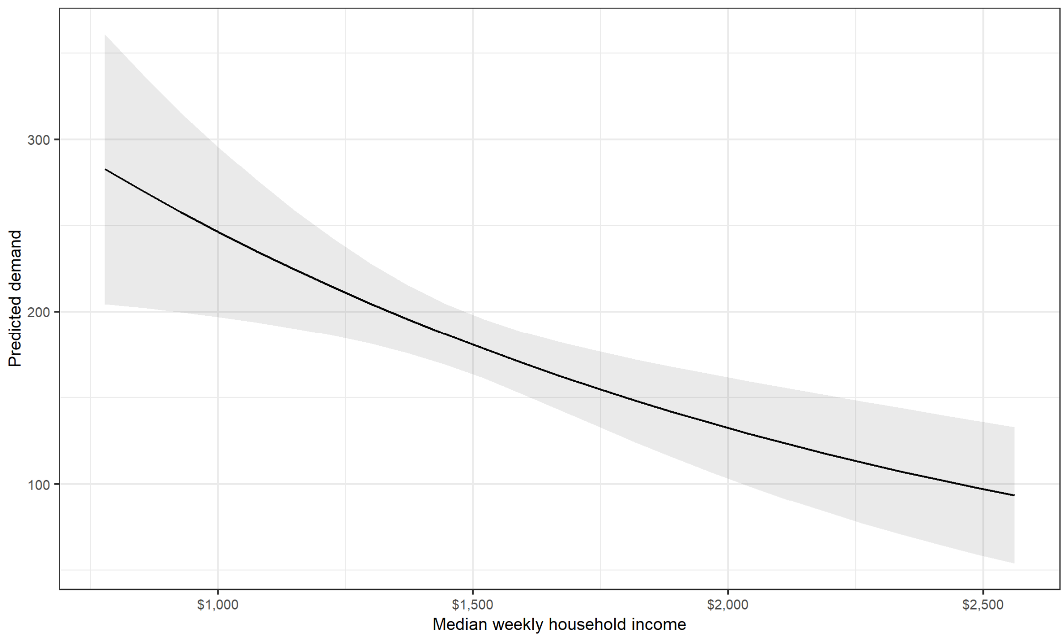


Distance to nearest school



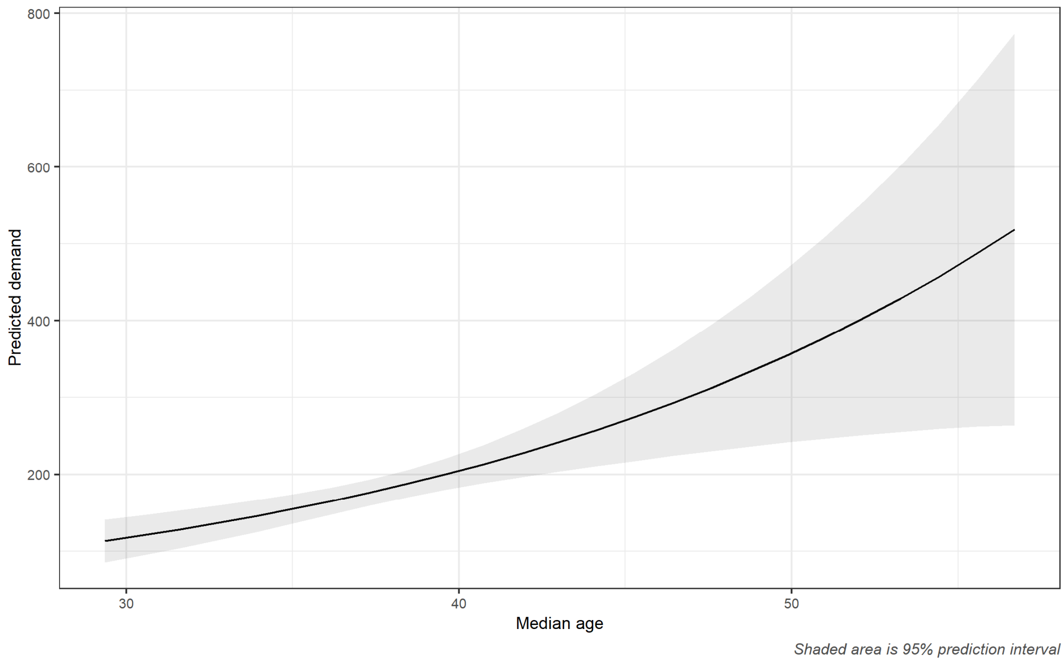
Shaded area is 95% prediction interval

Household income

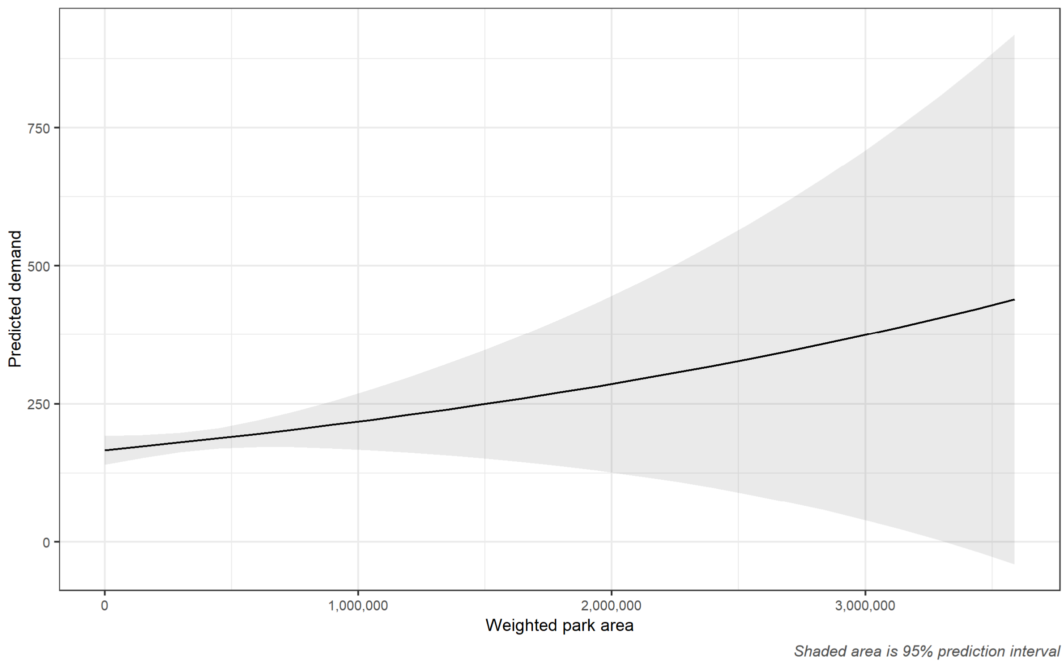


Shaded area is 95% prediction interval

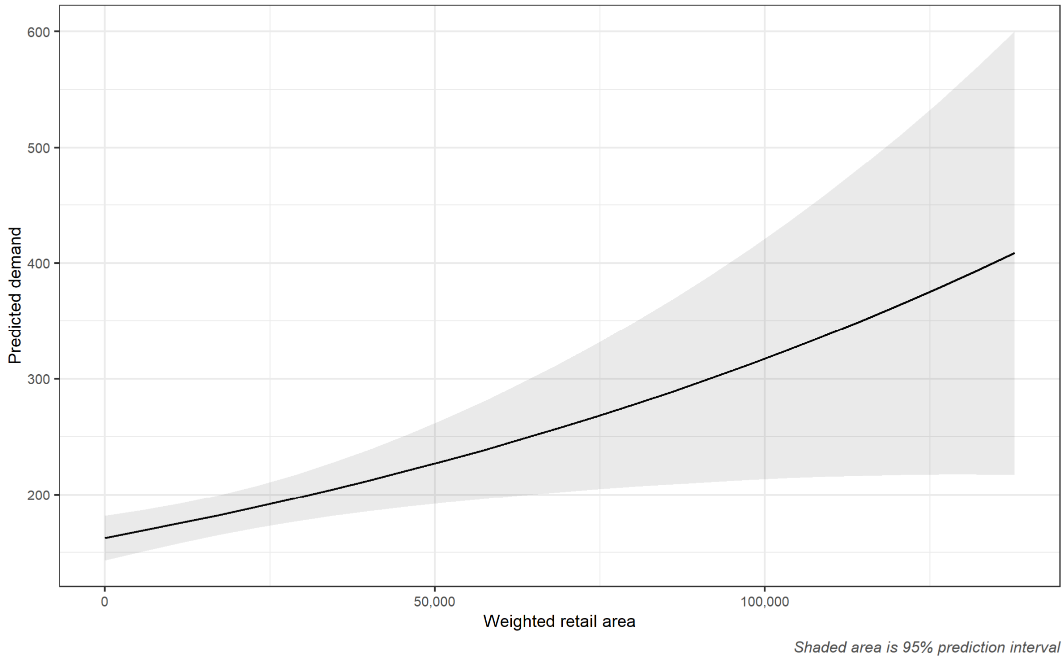
Median age



Park area



Retail area


600
500
400
300
200
Predicted demand
0
50,000
100,000
Weighted retail area
Shaded area is 95% prediction interval