

¹ **My First Four-Step Model: a Simple and Accessible Tool to Teach Travel Demand Modeling**

²

³ **Matthew Bhagat-Conway**

⁴ Department of City and Regional Planning

⁵ University of North Carolina at Chapel Hill

⁶

⁷ Email: mwbc@unc.edu

⁸ ORCiD: [0000-0002-1210-2982](https://orcid.org/0000-0002-1210-2982)

⁹

¹⁰

¹¹

¹²

¹³ **Abstract**

¹⁴ Travel demand modeling is taught in most transportation engineering and planning programs. It is generally taught
¹⁵ using a ‘bottom-up’ approach where students first learn about the theory and mathematics, and then work with indi-
¹⁶ vidual model components. They may never run a complete travel demand model. I propose a ‘top-down’ approach
¹⁷ where students first run a full demand model, focusing on outputs and the “big picture” of how the model works. This
¹⁸ introduction provides students the tools to be critically engaged consumers of model output. Most students in planning
¹⁹ and engineering will not become modelers, but many will work with model output.

²⁰ After a single lecture on demand modeling, I introduce a simple R package for running a four-step model on student’s
²¹ computers, and provide instructions for how to use it. Students have a homework assignment where they apply the
²² model and interpret the outputs. Students do well on the assignment, demonstrating an understanding of the basic
²³ structure of travel demand modeling. Most students take no additional modeling courses, but those that do have a
²⁴ strong foundation to build upon. Having students run a travel demand model soon after being introduced to the topic
²⁵ can help students focus on the big picture of modeling, and be informed consumers of model outputs, even if they take
²⁶ no further modeling classes.

²⁷ **Keywords:** travel demand modeling, teaching, education, open-source

²⁸ **Practical applications**

²⁹ This article introduces a very simple travel demand modeling package, designed for use in educational settings. It runs
³⁰ on any modern consumer-grade computer, uses only free software, and provides forecasts for any region of the US.
³¹ The goal of the tool is to give planners and engineers who do not work professionally with models enough experience
³² to collaborate effectively with those who do. I describe how I use the tool in an introductory planning class, to give
³³ future planning practitioners experience with travel forecasting.

34 Introduction

35 Most transportation planning and engineering programs teach travel demand modeling to some extent (Zhou &
36 Schweitzer, 2009). In my experience, it is almost invariably taught with a ‘bottom-up’ approach. Students take classes
37 on transportation planning, econometrics, and choice modeling. Students then work with individual components
38 of demand models, such as trip generation or mode choice. Often, students never “put it all together” to run a
39 regional model from start to finish. The vast majority of learning time is spent on theory and mathematics, rather than
40 applications. In this article, I introduce a ‘top-down’ approach where running a complete model is one of the first
41 steps, and discuss how I implement this in the classroom.

42 To facilitate teaching in this way, I introduce the open-source “My First Four-Step Model” R package, a simple four-step
43 model that can install easily and run quickly on consumer-grade computers. It can be estimated for any US metropolitan
44 area using publicly-available data. Complete code for running the model is presented in the results section.

45 Theory and mathematics are paramount for those who will build and run travel demand models themselves. This is
46 a very small group of students, however. At most metropolitan planning organizations and DOT’s, demand models
47 are estimated and run by consultants or a small in-house team. Consumers of model output are a larger group: trans-
48 portation planners and engineers, land use planners, developers, advocates, and so on. For this larger group, only a
49 cursory understanding of the mathematics is required; the general mechanisms and assumptions the model relies on
50 are far more important.

51 Most students in transportation planning and engineering programs will fall into the latter group. A better understanding
52 of modeling among this group will help promote better communications between modelers and model consumers.
53 Consumers will be more aware of what the model can and can’t do, and more able to come up with situations where
54 the model may be helpful. Understanding will also promote a “healthy skepticism” of the model, enabling feedback
55 from users on the model and ultimately leading to better models and decision support.

56 Students interested in modeling may take further classes, but all students will have some first-hand experience with
57 demand models—something many students do not get at all today, even after taking many classes on modeling. I use
58 this approach in my introductory Planning Methods course, where we spend only a week discussing transportation
59 modeling and engineering, and at the conclusion the students run a simple demand model for the Research Triangle
60 region of North Carolina.

61 This article discusses the model structure and implementation and how I use it in the classroom, and includes R code
62 to estimate and run a complete four-step model for anywhere in the US.

63 I am a planning faculty member, and draw on significant academic experience with modeling, theory, and mathematics,

64 as well as a previous career as a software developer. I am a white male, a member of a group that is overrepresented
65 in math and software development. I strive to make mathematics and modeling accessible to my students, who come
66 from diverse backgrounds both socioeconomically and academically.

67 Literature review

68 There is little literature on pedagogical practices surrounding travel demand modeling. Most transportation planning
69 and engineering programs cover the topic (Zhou & Schweitzer, 2009), but it was not even included on a semi-regular
70 survey of transportation faculty regarding what they consider the most important topics in introductory courses (Tur-
71 ochy, 2013).

72 There is, however, a long history of pedagogy around teaching through simulating real-world activities undertaken by
73 practitioners, rather than through one-to-many classroom instruction. This is most well established in the medical field,
74 with positive outcomes for student learning (McGaghie, Issenberg, Petrusa, & Scalese, 2010). Simulation activities
75 are widely used in transportation engineering instruction (Hurwitz, Bernhardt, Turochy, & Young, 2015), and research
76 on active learning techniques in transportation engineering goes back decades (Weir, 2004). Simulation activities in
77 planning have a similarly long history (e.g., Meier & Duke, 1966).

78 Effective simulations as an educational tool often take the form of a game. Solving transportation challenges is one of
79 the recurrent examples in the foundational book on gamification in education, Clark Abt's *Serious Games* (Abt, 1970).
80 More recently, physical board games have been used to teach transportation planning using both popular-press games
81 (Huang & Levinson, 2012) and purpose-built educational games (Paget-Seekins, 2021).

82 Computer-based simulations have rapidly become ubiquitous in transportation engineering education (Hurwitz et al.,
83 2015). Liao, Liu, and Levinson (2009) built a web-based traffic simulation tool to help students experiment with
84 signal timing practices. The interactive A/B Street traffic-simulation software has likewise been used in undergraduate
85 courses at Arizona State University (Carlino, Li, & Kirk, 2024). An economic simulation of airline operations has also
86 been applied to help budding engineers understand airline operations (Luken, Hotle, Alemdar, & Garrow, 2011).

87 Computer-based simulations have also been applied in planning, although perhaps less frequently. Simulations in
88 planning classrooms often take the form of commercial planning games, such as SimCity or Cities: Skylines (Gaber,
89 2007; Khan & Zhao, 2021), likely due to less funding for purpose-built simulations in planning as opposed to engi-
90 neering. A significant challenge with commercial games is that they are intended primarily for entertainment, and thus
91 may oversimplify or even modify system dynamics to support enjoyable gameplay rather than educational outcomes
92 (Gaber, 2007; Walker, 2009). The advantage is that commercial games are more likely to receive significant upfront
93 investment as well as continued support, a significant problem with games developed for educational purposes (Söbke,

94 Harder, & Planck-Wiedenbeck, 2018).

95 Public education and communication are another arena of planning where gamification and simulation have been
96 deployed. The *Future Energy Chicago* exhibit at Chicago's Museum of Science and Industry engaged participants
97 in a several-hour, facilitated game to improve energy outcomes. Survey data suggests that the game improved some
98 aspects of willingness to conserve energy (Applebaum, Price, & Foster, 2021). The CityScope platform provides
99 a hands-on physical environment wherein members of the public can make land-use changes to a Lego model of a
100 neighborhood and see computer simulation output regarding transport and energy consumption in real time (Alonso
101 et al., 2018). The CoAXs platform allows meeting participants to see how proposed Bus Rapid Transit routes would
102 affect their ability and the ability of other citizens to reach key destinations, and was found to support improved learning
103 and discussion outcomes among participants (Stewart, 2017; Stewart & Zegras, 2016). All of these simulations are
104 perhaps somewhat simpler than might be used in a classroom environment, since they target the general public rather
105 than future practitioners.

106 Teaching travel demand modeling differs from other places where simulations have been deployed in transportation
107 education. Travel demand models are themselves simulations of complex urban systems. Applying them in a classroom
108 environment does not demand developing a new simulation. Rather, it means simplifying the existing structure of
109 demand models to create one suitable for students with only a rudimentary understanding of the theory and mathematics
110 involved.

111 The only travel demand modeling software designed specifically for education I am aware of is the now-defunct Agent-
112 Based Demand and Assignment Model (ADAM) software (Zhu, Xie, & Levinson, 2011). ADAM implemented a sim-
113 ple agent-based model for transportation education. This model focused on a network assignment for simple networks;
114 it started with production (workers) and attractions (jobs), and students modified the network to reduce congestion.
115 Likely due to computational limits in place at the time, it worked with a very simple network of only 24 nodes and 68
116 links.

117 I focus on the ubiquitous four-step model in my introductory courses. The four-step model was one of the earliest
118 travel demand models developed (Federal Highway Administration, 1977; Weiner, 2013). While it has come under
119 significant criticism lately (Mladenovic & Trifunovic, 2014), it remains in common use. Many large regions have
120 transitioned to more modern activity-based models, but many smaller regions and even some large ones continue to
121 use the four-step model.

¹²² Pedagogical goals

¹²³ My First Four-Step Model is a software package that allows students with minimal experience and consumer-grade
¹²⁴ computer hardware to run a simple four-step travel demand model. Specifically, it is designed to address these student
¹²⁵ learning outcomes:

¹²⁶

- ¹²⁷ 1. Have a basic understanding of the structure and mathematics of travel demand model,
- ¹²⁸ 2. Understand the types of scenarios travel demand models are appropriate for,
- ¹²⁹ 3. Understand the limitations and uncertainty of travel demand modeling, and
- ¹³⁰ 4. Be able to have constructive conversations with travel modelers.

¹³¹

¹³² It is implemented as an R package (R Core Team, 2024), which has several advantages. R is a free, open-source,
¹³³ and cross-platform statistical programming language, allowing students to run it on their own computers regardless
¹³⁴ of configuration. Furthermore, R is becoming the *lingua franca* of quantitative urban planning. Using the My First
¹³⁵ Four-Step Model package in an assignment gives students a gentle introduction to the language and potentially piques
¹³⁶ their interest in learning more. The package has several key design goals:

¹³⁷

- ¹³⁸ 1. The four steps of the model map directly onto four functions in the package;
- ¹³⁹ 2. Any place where there is a tradeoff between simplicity and predictive accuracy, simplicity is chosen;
- ¹⁴⁰ 3. It can be estimated for any location in the United States using only publicly-available data;
- ¹⁴¹ 4. There are intuitive tools to visualize model inputs, outputs, and parameters, so students can interpret and understand the model;
- ¹⁴³ 5. Preparing land-use and transportation scenarios is simple;
- ¹⁴⁴ 6. It runs on any computer a student is likely to have (including Windows, Macs, and Chromebooks); and
- ¹⁴⁵ 7. It depends only on R itself and common R packages that are easily installed from the Comprehensive R Archive Network (CRAN).

¹⁴⁷

¹⁴⁸ To meet these goals, the model is highly simplified, and this certainly affects its predictive accuracy, but predictive
¹⁴⁹ accuracy is not one of the goals. Epstein (2008) lists 16 reasons to build models other than prediction. One of them,
¹⁵⁰ train practitioners, is the primary goal of this model. This goal does not depend on high predictive accuracy.

¹⁵¹ The open-source package is available on Github at <https://github.com/mattwigway/MyFirstFourStepModel>.

¹⁵² Results: implementation in the classroom

¹⁵³ I use the package in my Planning Methods course in the Department of City and Regional Planning at the University
¹⁵⁴ of North Carolina at Chapel Hill. This is an introductory master's level course which all planning students (not only
¹⁵⁵ those in transportation) take, generally during their first semester. In the course, I spend a week discussing transporta-
¹⁵⁶ tion planning and engineering. I give a single hour-and-fifteen-minute lecture on transportation modeling, covering
¹⁵⁷ primarily the four-step model, but with a nod to activity-based models. We discuss the general structure of the model.
¹⁵⁸ We cover some basic mathematical underpinnings of discrete-choice models. We discuss travel surveys and how they
¹⁵⁹ are used in model development.

¹⁶⁰ I then have an assignment where all students run and interpret the output of a four-step model. Specifically, students first
¹⁶¹ run a baseline model. They then change land use inputs to reflect Chatham Park, a large proposed housing development
¹⁶² in Pittsboro, NC. Pittsboro is a town that is quickly becoming an exurb of the Research Triangle. I give students
¹⁶³ guidance on the number of households to add based on development plans, but give them discretion in determining the
¹⁶⁴ demographics of those households. They then run the model again. This portion of the assignment is worth five points
¹⁶⁵ plus 0.5 extra credit points for calculating statistical significance from a coefficient and standard error. For each step
¹⁶⁶ of the model, I have 1–2 homework questions. These questions cover interpretation the coefficients of the constituent
¹⁶⁷ models, and what the outputs indicate about the transportation system and differences between scenarios.

¹⁶⁸ There is also an extra credit section worth two points where students run their scenario again with a network where 15-
¹⁶⁹ 501, the main highway between Pittsboro and Chapel Hill, is widened from two to three lanes. I have them interpret
¹⁷⁰ how congestion changes in this scenario. Modeled results of widening projects are often overstated due to induced
¹⁷¹ demand, the phenomenon where building more roadway capacity induces more people to take trips on that road. Like
¹⁷² many four-step models, My First Four-Step Model does not account for induced demand. I have students discuss how
¹⁷³ this biases the results.

¹⁷⁴ I provide the students with an R file, all of the code of which is included inline in this section to run the model. The
¹⁷⁵ full homework assignment is in the appendix.

¹⁷⁶ Students do well on this assignment. In Spring 2025, the mean score was 5.96 (n=42, s.d.=1.51, median=5.5, 25th
¹⁷⁷ percentile=5, 75th percentile=7.45). 20 students (48%) attempted the two-point extra credit section looking at highway
¹⁷⁸ widening. Unfortunately, this class did not cover travel demand modeling prior to the introduction of My First Four-
¹⁷⁹ Step Model, so it is not possible to evaluate how the tool has changed learning outcomes.¹

¹This use of student grade data was reviewed and approved by the University of North Carolina at Chapel Hill Registrar and Institutional Review

180 The goal is to enable students in all specializations, not just transportation, to understand basic demand model struc-
181 ture and have productive conversations with modelers. For students who want to work more closely with models in
182 their careers, I also teach a semester-long travel demand modeling course using a more complex model developed in
183 TransCAD. Most transportation specialization students take this course later in their program.

184 **Installation**

185 All my students already have R and RStudio installed on their machines from a previous exercise on linear regression.
186 Installing My First Four-Step Model is simple; students run the following R code at the command prompt to install the
187 package:

```
install.packages('MyFirstFourStepModel',  
  repos = c('https://mattwigway.r-universe.dev', 'https://cloud.r-project.org'))
```

188 This installs a pre-compiled R package containing the model. It also automatically installs the R packages
189 `MyFirstFourStepModel` depends on—notably `tidyverse` (Wickham et al., 2019), `sf` (Pebesma & Bivand, 2023),
190 `igraph` (Csardi & Nepusz, 2006), `readxl` and `writexl` (Ooms, 2024; Wickham & Bryan, 2023), `tidycensus` (K.
191 Walker & Herman, 2024), `tigris` (K. Walker, 2024), and `nnet` (Venables & Ripley, 2002).

192 **Loading the package and the model**

193 Next, students load the `MyFirstFourStepModel` package, and an already-estimated model. Most model users will
194 never estimate a model themselves, so I provide an already-estimated model for the Research Triangle region of North
195 Carolina. The model can be read either from a local file or directly from an `https` URL. Reading directly from a URL
196 requires the instructor to have access to a server, but avoids needing to troubleshoot local file paths. The URL included
197 below is a live URL with the model for the Research Triangle.

```
library(MyFirstFourStepModel)  
model = load_model("https://files.indicatrix.org/rdu_chatham.model")
```

198 **Trip generation**

199 The first step of the four-step process is trip generation. Trip generation is done at the household level using linear
200 regression. I use linear regression rather than traditional cross-classification or more complex regression methods
201 because of its ease of interpretation, and because I teach demand modeling shortly after teaching linear regression.

Board (approval 24-2069). The requirement for consent was waived due to the aggregate nature of the data.

Table 1: Trip production regression for AM Peak home-based work trips

Coefficient	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.00	0.02	-0.17	0.86
Number of vehicles	0.03	0.01	2.97	0.00
Household size	-0.03	0.01	-4.68	0.00
Income 35,000-74,999	0.02	0.02	1.48	0.14
Income 75,000-99,999	0.08	0.02	3.76	0.00
Income > 100,000	0.07	0.02	3.69	0.00
Housing unit density in home tract (units/square mile)	0.00	0.00	-0.36	0.72
Number of workers	0.40	0.01	48.76	0.00

- 202 Before running the trip generation model, I have students view the trip generation model for AM Peak home-based
 203 work trips, using the code below, and interpret the coefficients (Table 1).

```
summary(model$production_functions`AM Peak`$HBW)
```

- 204 I similarly have them interpret one of the regressions for attraction functions (coefficients not shown for brevity):

```
summary(model$attraction_functions`AM Peak`$HBW)
```

- 205 Once students have interpreted the regressions, I have them run the trip generation step. True to the design goals, this
 206 requires only a single function.

```
productions_attractions = trip_generation(model, model$scenarios$baseline)
```

- 207 Students can then map the number of trips produced and attracted in each Census tract in the region using the
 208 `map_trip_generation` function as shown below (Figure 1).

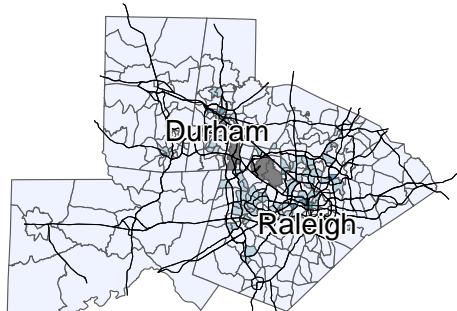
```
map_trip_generation(  
  model,  
  productions_attractions,  
  "Productions",  
  "AM Peak",  
  "HBW"  
)  
  
map_trip_generation(  
  model,  
  productions_attractions,
```

```

    "Attractions",
    "AM_Peak",
    "HBW"
)

```

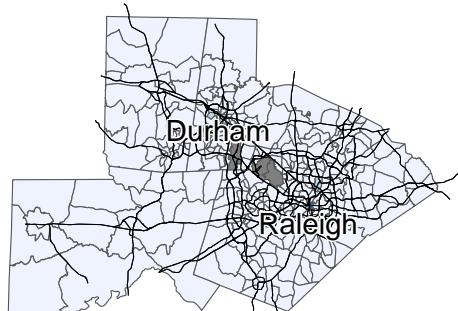
HBW productions, AM Peak



Trips per square kilometer

250	500	750	1000
-----	-----	-----	------

HBW attractions, AM Peak



Trips per square kilometer

2000	4000	6000
------	------	------

Figure 1: Home-based work trip productions and attractions for the Research Triangle region, AM Peak

209 Trip distribution

210 Trip distribution uses a simple gravity model, with different parameters estimated for home-based work, home-based
 211 other, and non-home-based trips. I first have students print the parameters using the code below, and interpret them
 212 (the estimated values for the Research Triangle region are -1.21 for home-based work trips, -1.87 for home-based other
 213 trips, and -1.71 for non-home-based trips).

```
model$distribution_betas
```

214 Then, they can run the trip distribution step with the following code

```
flows = trip_distribution(
  model,
  model$scenarios$baseline,
  productions_attractions
```

215 Students can then map the trip distribution results for any origin Census tracts, and interpret them. Figure 2 shows the
216 results for AM Peak home-based work trips originating from a tract in suburban Durham; results show that many trips
217 stay local, but there are also pockets of activity in further-flung large employment centers (e.g., Raleigh).

```
map_trip_distribution(  
    model,  
    flows,  
    "AM Peak",  
    "HBW",  
    origin_tract="37063002025"  
)
```

HBW trips, AM Peak from tract 37063002025

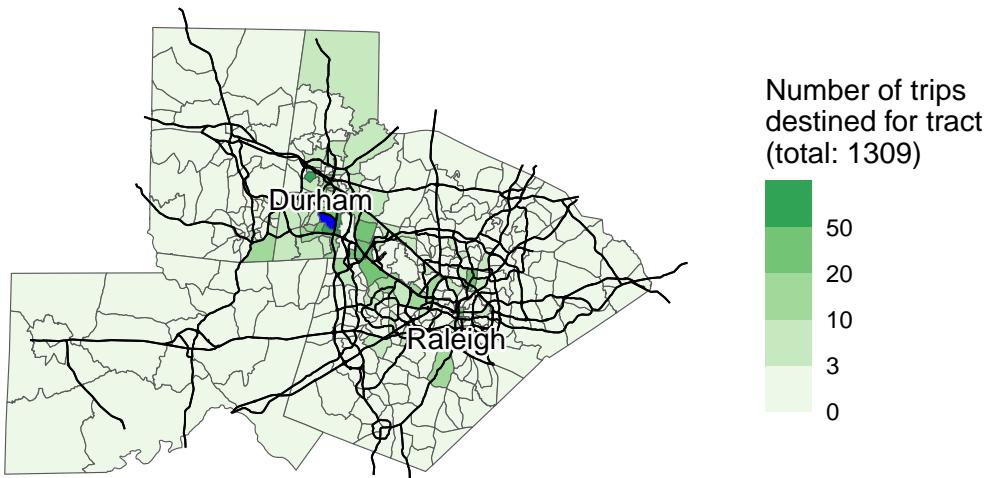


Figure 2: AM Peak trip distribution, from a selected tract in suburban Durham, NC

218 Mode choice

219 The mode choice model is a multinomial logit model with four modes: walk, bike, transit, and drive. Since the model
220 is estimated from public data, there are minimal attributes of each individual trip available, so the model is very simple
221 and primarily based on Euclidean distance between the origin and destination. I give a very basic explanation of the
222 multinomial logit model, highlighting commonalities with linear regression, and have students print the mode choice

Table 2: Coefficients from the multinomial logit mode choice model, for home-based trips

Mode	(Intercept)	Density [†]	Trip length (km)	Midday	PM Peak	Overnight	Work trip
Transit	-2.6375***	0.0003***	-0.0204***	-0.4486***	-1.7884***	-2.1183***	-1.3445***
Walk	-0.6018***	0.0001***	-0.4696***	-0.3797***	-0.2467***	-0.0537	-0.9591***
Bike	-4.6994***	0.0004***	-0.0609***	-0.0888	-0.0585	-0.2398***	-0.1505***

[†]Homes/sq. mi. in home Census tract

* = p < 0.05, ** = p < 0.01, *** = p < 0.001

223 model using the code below, and interpret a few coefficients (Table 2):

```
summary(model$mode_choice_models$HB)
```

224 I then have students run the mode choice step and calculate mode shares, using the code below; for the Triangle region,
225 they are 91% car, 5% walk, 3% transit, and 1% bike in the baseline.

```
flows_by_mode = mode_choice(model, model$scenarios$baseline, flows)
get_mode_shares(flows_by_mode)
```

226 Network assignment

227 The final step of the model is network assignment. A simple Frank-Wolfe static traffic assignment algorithm is
228 used to assign trips to the network. Since network assignment is time consuming, a relatively simple network is
229 used. The network used in the Research Triangle example model has 7529 nodes and 10497 edges. The code be-
230 low performs assignment for the PM Peak and displays congestion (Figure 3). To maximize network assignment
231 performance, a small amount of code written in the Rust language (Klabnick, Nichols, Krycho, & Rust Community,
232 2025) is used to efficiently process routing results; this code is pre-compiled and will be installed automatically when
233 MyFirstFourStepModel is installed.

```
pm_network_flows = network_assignment(
  model,
  model$scenarios$baseline,
  model$networks$baseline,
  flows_by_mode,
  "PM Peak"
)
map_congestion(model, model$networks$baseline, pm_network_flows)
```

234 Agencies are increasingly interested in vehicle miles traveled. The assignment step can also estimate VMT by period
235 using the code below. For the PM Peak in the Research Triangle, this is estimated to be 7 million miles/day. The

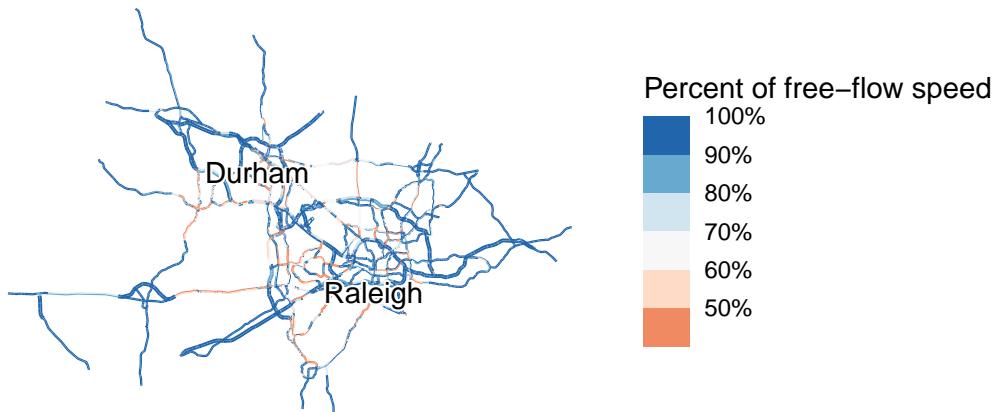


Figure 3: Forecast PM Peak congestion

236 2017 Local Area Transportation Characteristics for Households (LATCH) statistics estimate total VMT in the Triangle
 237 region to be 26 million miles/day (author calculations from Bureau of Transportation Statistics, 2024), so 7 million in
 238 the PM Peak is reasonable.

```
estimate_vmt(model, model$networks$baseline, pm_network_flows, "PM Peak")
```

239 [1] 7174647

240 Land-use scenarios

241 The code above runs a complete, simplified, four-step model for baseline conditions. However, we are generally in-
 242 terested in forecasts for future years under different scenarios. Above, I requested baseline land use by specifying
 243 `model$scenario$baseline`. The baseline is automatically created during model estimation based on Census popu-
 244 lation and employment data.

245 There are two ways to create additional scenarios. The simpler method is to use the `add_households` function. This
 246 function modifies a scenario by adding households with particular characteristics to a particular area. The code below
 247 adds 20,000 households to Census tract 37037020801, which is the location of a large new residential development.
 248 The households to add are specified in a tabular format. In this case, are all four-person, two-worker households; half
 249 have three cars and income of \$150,000/year, and half have two cars and income of \$75,000/year.

```

model$scenarios$future = model$scenarios$baseline |>
  add_households(
    "37037020801",
    tribble(
      ~hhszie, ~workers, ~vehicles, ~income, ~n,
      4,        2,        3,        150000,  10000,
      4,        2,        2,        75000,   10000
    )
  )

```

250 The model can then be re-run using the same steps as above, substituting `model$scenarios$future` for
 251 `model$scenarios$baseline`. This results in the forecast congestion shown in Figure 4. The new development is
 252 highlighted with a red dot.

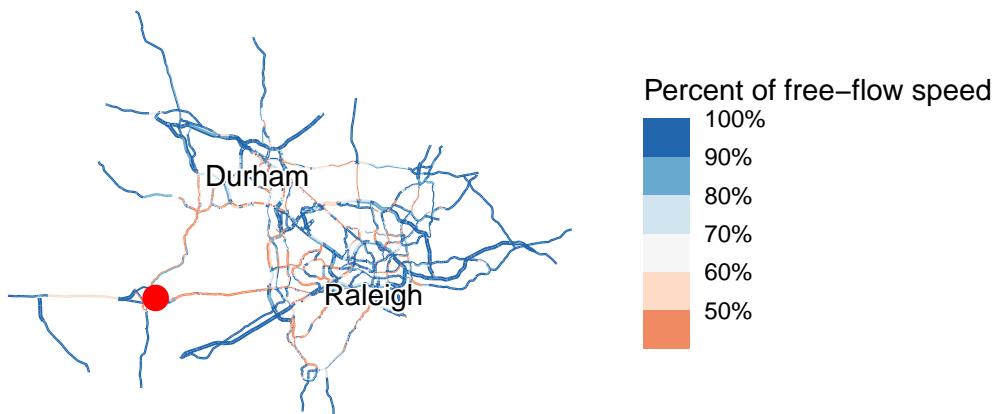


Figure 4: Forecast congestion levels after adding 20,000 households, PM Peak

253 If more control is needed, scenarios can be created in a tabular format by hand or using external tools. This format is
 254 shown in Table 3, specifying the number of households in different household size, income, vehicle ownership, and
 255 number of worker categories. A similar format is used to specify employment numbers (used as a proxy for out-of-
 256 home destinations).

Table 3: Specification of a demographic scenario

geoid	marginal	value	count
37183053411	hhszie	1	514
37183053411	hhszie	2	711
37183053411	hhszie	3	940
37183053411	hhszie	4	1907
37183053411	income	0	358
37183053411	income	35000	595
37183053411	income	75000	183
37183053411	income	100000	2936
37183053411	vehicles	0	110
37183053411	vehicles	1	921
37183053411	vehicles	2	2089
37183053411	vehicles	3	952
37183053411	workers	0	288
37183053411	workers	1	1784
37183053411	workers	2	1711
37183053411	workers	3	289

- 257 In the assignment I give to my students, I use this format to create a scenario based on a demographic forecast we work
 258 earlier in the semester. To do this, I first export the baseline scenario into Excel format:

```
save_landuse_scenario(model$scenarios$baseline, "baseline.xlsx")
```

- 259 After modifying the scenario to be consistent with the forecast growth using external tools, I re-load it:

```
model$scenarios$projected = load_landuse_scenario("projected.xlsx")
```

260 **Network scenarios**

- 261 Changes to the network are just as important as changes to land use. Currently, My First Four-Step Model only supports
 262 changes to existing links.

- 263 Two link attributes can be changed: the lane count and the roadway type. The roadway type uses OpenStreetMap
 264 highway tag taxonomy; notable values are ‘motorway’, ‘trunk’, and ‘primary’. Changing the attributes of a link requires

265 determining its OpenStreetMap ‘way ID’, which can be done either by looking at the data on openstreetmap.org, or
266 exporting the network to GIS format and investigating in GIS:

```
network_to_gis(model$networks$baseline, "baseline.gpkg")
```

267 Link attributes can then be modified using the `modify_links` function. For example, the code below widens the road
268 going north from the development, from two to three lanes in each direction, and upgrades it to a motorway. Some
269 way IDs are suppressed for brevity (see Supplemental Materials for full code).

```
model$networks$widen = model$networks$baseline |>  
  modify_ways(  
    # US 15-501 between Pittsboro and Chapel Hill  
    c(  
      "16468788", "133051274", "16471803", "285898984",  
      . . .  
      "712336821", "712336826", "712336827", "998595932"  
    ),  
    lanes_per_direction=3,  
    highway_type="motorway"  
  )
```

270 Students can then repeat the network assignment step with the new network to see how congestion changes—in this
271 case, that there is no longer forecast congestion along the road we modified (Figure 5). This also provides a jumping-off
272 point to discuss the (lack of) modeling of induced demand.

273 Methods: estimation of a new model

274 Estimating a new model requires only a few lines of code, however it does require the 2017 National Household Travel
275 Survey (NHTS, Federal Highway Administration, 2017) and an OpenStreetMap PBF file for the region modeled.² The
276 code to estimate a model for the Research Triangle region is below. First, it loads the relevant libraries, and then the
277 NHTS (NHTS_PATH should be replaced with a directory containing the NHTS CSV files). I filter the NHTS to only
278 North Carolina households with a weekday travel day ($n = 7,146$). The final line estimates the model. It requires
279 the (possibly filtered) NHTS, the path to the OpenStreetMap data (written as OSM_PATH below but should be replaced
280 with the actual path), the state and a vector of counties to define the region under study, and a year. Currently 2021 is

²OpenStreetMap PBF files for any region are easily obtained from <https://slice.openstreetmap.us>

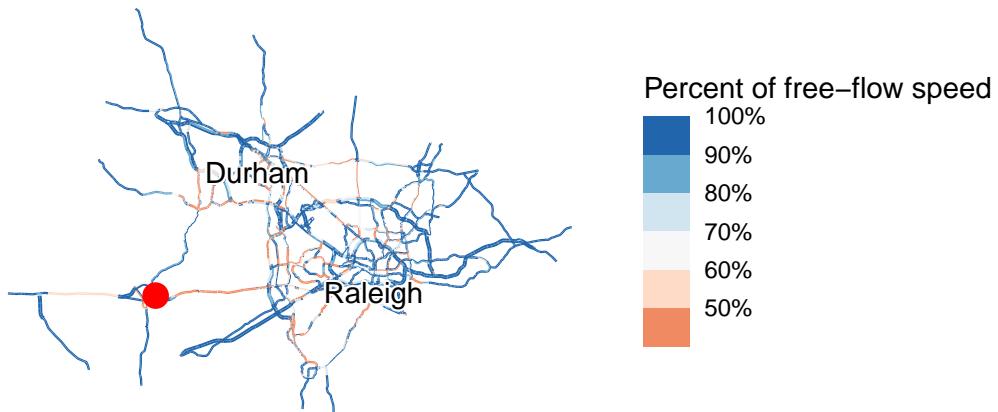


Figure 5: Forecast congestion, PM Peak, with widened 15-501.

most recent year available, as this is based on American Community Survey and Longitudinal Employer-Household Dynamics data availability.
 Parsing the OpenStreetMap data uses Julia (Bezanson, Edelman, Karpinski, & Shah, 2017) for performance, which can be installed if it is not already by running `JuliaCall::install_julia()` from within R. Julia is only required for estimation; students do not need to install Julia.

```

library(MyFirstFourStepModel)
library(tidyverse)

# Load NHTS and filter to North Carolina weekday data
nhts = load_nhts(NHTS_PATH)
nhts$households = filter(
  nhts$households,
  HHSTATE == "NC" & TRAVDAY %in% c(2, 3, 4, 5, 6)
)

# Estimate the model using 2021 Census/LODES data for the Triangle
model = estimate(nhts, OSM_PATH, "NC", c("Durham", "Orange", "Wake", "Chatham"), 2021)
  
```

286 Lastly, the model can be saved to a file for distribution to students.

```
save_model(model, "rdu.model")
```

287 This can be loaded by the `load_model` function described above, either from a file or a URL. If any land-use or network
288 scenarios are created or loaded prior to saving the model, they will be included in the saved file.

289 **Model architecture and input data**

290 **Trip generation**

291 Trip generation is based on the NHTS. The model uses four time periods: overnight (7:00 pm–5:59 am), AM Peak
292 (6:00 am–9:59 am), midday (10:00 am–3:59 pm), and PM Peak (4:00 pm–6:59 pm). The model divides trips into three
293 purposes: home-based work (HBW), home-based other (HBO), and non-home-based (NHB). While this is fewer trip
294 types and time periods than might be included in production travel models, it is consistent with the general practice of
295 dividing trips by time of day and count.

296 Household-level trip counts are estimated for each time period and trip type using a simple linear regression with
297 the trip count as the dependent variable and independent variables for number of vehicles, household size, household
298 income, Census tract residential density, and number of workers. This results in 12 regression equations, for each time
299 period and trip purpose. Household income is represented by dummy variables for less than \$35,000, \$35,000–\$74,999,
300 \$75,000–\$99,999, and \$100,000 or more.

301 These regression models are disaggregate, household-level models, whereas the four-step model is an aggregate model
302 using marginal data at the TAZ level (for simplicity, TAZs correspond directly to Census tracts). Specifically, the
303 model uses household size (topcoded at 4), number of workers (topcoded at 3), number of vehicles (topcoded at 3),
304 and income (in the categories used in the regression). To apply the household-level model to this aggregate data, I
305 disaggregate the data to household-level records (i.e. create a synthetic population) using iterative proportional fitting
306 with a seed matrix derived from the Integrated Public Use Microdata Sample 2021 Five-Year American Community
307 Survey data for the entire US (Ruggles et al., 2024). This seed matrix ships with the software. The regressions predict
308 household-level tripmaking, which is re-aggregated to the tract level.

309 The NHTS does not provide sufficient spatial detail to estimate trip attractions. Instead, I use the Puget Sound House-
310 hold Travel Survey, which includes origin and destination Census tract in the public-use dataset (Puget Sound Regional
311 Council, 2024). I calculate the number of HBW and HBO trips in each time period that have the non-home end in each
312 Census tract in the Puget Sound region. For NHB trips, the production and attraction ends are not clearly defined. I
313 therefore assume that NHB attractions and productions are half of the total number of NHB trips originating from or

314 destined to each tract.

315 To extrapolate this data to tracts outside the region, I build linear regression models for each trip type and time period
316 based on total employment and employment in retail, education, and accomodation/food services from the US Cen-
317 sus Bureau Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics. I balance total
318 attractions by trip type in each time period to match estimated productions.

319 **Trip distribution**

320 The trip distribution step uses a singly-constrained (at the production end) gravity model (Travel Forecasting Resource,
321 2020). The exponent for the gravity model is calibrated for each trip type using the method introduced by Merlin (2020)
322 based on median trip length. The method observes that half of the weighted destinations should be closer to the origin
323 than the median trip, and half should be further away. I make two slight changes to the function presented in Merlin
324 (2020); see appendix.

325 The median trip distance comes from the NHTS. I approximate the crow-flies distance for each NHTS trip by dividing
326 the network distance by 1.3, a factor determined by Wang, Bette, Schreckenberg, & Guhr (2024).

327 For intrazonal trips, I assume a travel distance of $0.52\sqrt{s}$, where s is the area of the TAZ. This is based on a Monte
328 Carlo simulation of the average distance between random points in a square. There are two opposing factors that bias
329 this. TAZ's are not square, increasing average travel distance. However, development within a TAZ is concentrated in
330 certain areas, decreasing average travel distance. I assume these roughly cancel out.

331 **Mode choice**

332 The mode choice model is a multinomial logit model based on the NHTS. Because there is not detailed information
333 about each trip and the alternatives available in the NHTS, the model is based solely on trip type, time period, travel
334 distance, and housing unit density in the home tract. For NHB trips, a separate model is estimated excluding density.
335 Goodness of fit is poor, but the point of the model is to demonstrate a simple demand model, not to produce accurate
336 forecasts.

337 **Network assignment**

338 The network assignment uses a Frank-Wolfe traffic assignment algorithm (Boyles, Lownes, & Unnikrishnan, 2023).
339 First, I convert production-attraction format home-based trips into origin-destination format, using directionality factors
340 estimated from the NHTS. I calculate peak hourly vehicle flows during each period using average vehicle occupancy
341 for each period and trip type, and an assumed “peaking factor” that accounts for the proportion of traffic during the

342 time period that occurs in the busiest hour—for example, I assume 45% of the traffic in the three-hour PM Peak period
343 occurs during the busiest hour. Then, I run the assignment algorithm, with impedances based on a Bureau of Public
344 Roads-style function:

$$t_{\text{congested}} = \left(1 + 0.6 \left[\frac{f}{c} \right]^5 \right) t_{\text{freeflow}}$$

345 where f is the predicted flow, c is the link capacity, and $t_{\text{congested}}$ and t_{freeflow} are congested and free-flow link travel
346 times. The factors 0.6 and 5 are from the Southern California Association of Governments travel demand model
347 (Southern California Association of Governments, 2012).

348 The assignment algorithm is written in pure R. This is quite slow, but will run anywhere R does without any installation
349 complexity. To keep runtimes reasonable (on the order of minutes or tens of minutes, depending on the computer),
350 networks need to be simple, and regions small. In the example, I use the central four counties of the Research Triangle
351 region.

352 I derive the network from OpenStreetMap. By default, I retain only the most major roads—motorways, trunk, and
353 primary roads (and associated ramps). Furthermore, I run the assignment algorithm only until the “relative gap”—a
354 measure of the error in the estimate—is 1%, rather than the typically recommended 0.01% (Boyce, Ralevic-Dekic, &
355 Bar-Gera, 2004), to improve performance.

356 Discussion and conclusion

357 This article introduced a new way of teaching travel demand modeling in introductory courses. The primary change
358 is to move the step of working with an actual demand model *much* earlier in the process, after only a few lectures. To
359 facilitate this, I introduced the My First Four-Step Model R package, which implements a highly simplified demand
360 model that can be run on students’ machines and estimated anywhere in the US with only publicly-available data.
361 With only a few lines of code, which I provide to students as a runnable R file, students can run a four-step model,
362 and interact with and interpret the outputs. This gives them firsthand experience with a demand model, which will
363 promote improved communications between model output consumers and model developers once the students enter
364 the workforce.

365 Insufficient accounting for induced demand—the phenomenon of roadway expansion leading to additional demand
366 (Downs, 2004)—is a common criticism of four-step models. This is a particular concern among planning students. My
367 First Four-Step Model is worse even than most production travel models; it does not account for induced demand at all.
368 While most models would use estimates of network travel time and travel cost in the distribution and mode choice, and
369 thus be at least somewhat sensitive to changes in the network, My First Four-Step Model relies entirely on crow-flies

370 distances.

371 My First Four-Step Model will never be appropriate for production travel demand modeling. It is also not appropriate
372 as a sole teaching tool for students who will ultimately become modelers. However, it is useful as a first exercise even
373 in courses that focus only on demand modeling, where students can have a chance to work with a simple model before
374 diving into the more complex theories and software that are necessary for a detailed education in this area.

375 **Data availability statement**

376 Some or all data, models, or code generated or used during the study are available in a repository online.

- 377 • Models and code are available on Github: <https://github.com/mattwigway/MyFirstFourStepModel>
- 378 • Street network data are available from OpenStreetMap: <https://slice.openstreetmap.us>
- 379 • Demographic data are available from the US Census Bureau, at several locations:
 - 380 – Residential locations: through the `tidycensus` package (K. Walker & Herman, 2024)
 - 381 – Job locations: LEHD LODES program (US Census Bureau, n.d.)
 - 382 – Seed matrix: Integrated Public Use Microdata Sample (Ruggles et al., 2024)
- 383 • NHTS survey data (Federal Highway Administration, 2017)
- 384 • Puget Sound Survey Data (Puget Sound Regional Council, 2024)
- 385 • Student grade data are not publicly available to protect privacy

386 **Acknowledgements**

387 Road network data in Figure 1–Figure 5 © OpenStreetMap contributors. An earlier version of this work was presented
388 at the Transportation Research Board Annual Meeting (Bhagat-Conway, 2025).

389 **Statement on artificial intelligence**

390 No generative AI tools were used in this research.

391 **References**

- 392 Abt, C. C. (1970). *Serious games*. New York, Viking Press. Retrieved from <http://archive.org/details/seriousgames0000abtc>

- 394 Alonso, L., Zhang, Y. R., Grignard, A., Noyman, A., Sakai, Y., ElKatsha, M., ... Larson, K. (2018). CityScope: A
395 Data-Driven Interactive Simulation Tool for Urban Design. Use Case Volpe. In A. J. Morales, C. Gershenson,
396 D. Braha, A. A. Minai, & Y. Bar-Yam (Eds.), *Unifying Themes in Complex Systems IX* (pp. 253–261). Cham:
397 Springer International Publishing. https://doi.org/10.1007/978-3-319-96661-8_27
- 398 Applebaum, L. R., Price, C. A., & Foster, A. Y. (2021). Collaboration and Competition in Exhibit Facilitation Around
399 Energy Literacy. *Journal of Museum Education*, 46(1), 113–126. <https://doi.org/10.1080/10598650.2020.1858268>
- 400 Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing.
401 *SIAM Review*, 59(1), 65–98. <https://doi.org/10.1137/141000671>
- 402 Bhagat-Conway, M. W. (2025). *My First Four-Step Model: A Simple and Accessible Tool to Teach Travel Demand
403 Modeling*. Presented at the Transportation Research Board Annual Meeting. Retrieved from <https://annualmeeting.mytrb.org/FileUpload/FullPaper?ID=58574&SessionID=22907&ConferenceID=13>
- 405 Boyce, D., Ralevic-Dekic, B., & Bar-Gera, H. (2004). Convergence of Traffic Assignments: How Much is Enough?
406 *Journal of Transportation Engineering*, 130(1), 49–55. [https://doi.org/10.1061/\(ASCE\)0733-947X\(2004\)130:1\(49\)](https://doi.org/10.1061/(ASCE)0733-947X(2004)130:1(49))
- 408 Boyles, S. D., Lownes, N. E., & Unnikrishnan, A. (2023). *Transportation network analysis* (Vol. 1).
- 409 Bureau of Transportation Statistics. (2024). *2017 Local Area Transportation Characteristics for Households Data
410 [Data set]*. Retrieved from <https://www.bts.gov/latch/latch-data>
- 411 Carlino, D., Li, Y., & Kirk, M. (2024). A/B Street. Retrieved July 28, 2024, from <https://a-b-street.github.io/docs/index.html>
- 413 Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex
414 Systems*, 1695. Retrieved from <https://igraph.org>
- 415 Downs, A. (2004). *Still stuck in traffic: Coping with peak-hour traffic congestion*. Washington, D.C: Brookings
416 Institution Press.
- 417 Epstein, J. M. (2008). Why model? *Journal of Artificial Societies and Social Simulation*. Retrieved from <http://jasss.soc.surrey.ac.uk/11/4/12.html>
- 419 Federal Highway Administration. (1977). *PLANPAC/BACKPAC general information manual : Computer programs
420 for urban transportation planning*. Retrieved from <https://hdl.handle.net/2027/ien.35556021025184>
- 421 Federal Highway Administration. (2017). *2017 national household travel survey*. Retrieved from <https://nhts.onl.gov/downloads>
- 423 Gaber, J. (2007). Simulating Planning: SimCity as a Pedagogical Tool. *Journal of Planning Education and Research*,
424 27(2), 113–121. <https://doi.org/10.1177/0739456X07305791>
- 425 Huang, A., & Levinson, D. (2012). To Game or Not to Game: Teaching Transportation Planning with Board Games.
426 *Transportation Research Record*, 2307(1), 141–149. <https://doi.org/10.3141/2307-15>

- 427 Hurwitz, D. S., Bernhardt, K. L. S., Turochy, R. E., & Young, R. K. (2015). Transportation Engineering Instructional
428 Practices: Analytic Review of the Literature. *Transportation Research Record*, 2480(1), 45–54. <https://doi.org/10.3141/2480-06>
- 430 Khan, T. A., & Zhao, X. (2021). Perceptions of Students for a Gamification Approach: Cities Skylines as a Pedagogical Tool in Urban Planning Education. In D. Dennehy, A. Griva, N. Pouloudi, Y. K. Dwivedi, I. Pappas, & M. Mäntymäki (Eds.), *Responsible AI and Analytics for an Ethical and Inclusive Digitized Society* (pp. 763–773). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-85447-8_64
- 431 Klabnick, S., Nichols, C., Krycho, C., & Rust Community. (2025). *The Rust programming language* (2nd ed.).
- 432 Liao, C.-F., Liu, H. X., & Levinson, D. M. (2009). Simulating Transportation for Realistic Engineering Education and Training: Engaging Undergraduate Students in Transportation Studies. *Transportation Research Record*, 2109(1), 12–21. <https://doi.org/10.3141/2109-02>
- 433 Luken, B., Hotle, S., Alemdar, M., & Garrow, L. (2011). A Case Study: Educating Transportation Engineers with Simulation Software. *2011 ASEE Annual Conference & Exposition Proceedings*, 22.19.1–22.19.10. Vancouver, BC: ASEE Conferences. <https://doi.org/10.18260/1-2--17301>
- 434 McGaghie, W. C., Issenberg, S. B., Petrusa, E. R., & Scalese, R. J. (2010). A critical review of simulation-based medical education research: 2003–2009. *Medical Education*, 44(1), 50–63. <https://doi.org/10.1111/j.1365-2923.2009.03547.x>
- 435 Meier, R. L., & Duke, R. D. (1966). Gaming Simulation for Urban Planning. *Journal of the American Institute of Planners*, 32(1), 3–17. <https://doi.org/10.1080/01944366608978484>
- 436 Merlin, L. A. (2020). A new method using medians to calibrate single-parameter spatial interaction models. *Journal of Transport and Land Use*, 13(1, 1), 49–70. <https://doi.org/10.5198/jtlu.2020.1614>
- 437 Mladenovic, M., & Trifunovic, A. (2014). The shortcomings of the conventional four step travel demand forecasting process. *Journal of Road and Traffic Engineering*.
- 438 Ooms, J. (2024). *Writexl: Export data frames to excel 'xlsx' format*. Retrieved from <https://CRAN.R-project.org/package=writexl>
- 439 Paget-Seekins, L. (2021, October 20). Transform Transit. Retrieved July 28, 2024, from <https://laurelintransit.com/transform-transit/>
- 440 Pebesma, E., & Bivand, R. (2023). *Spatial Data Science: With applications in R*. Chapman and Hall/CRC. <https://doi.org/10.1201/9780429459016>
- 441 Puget Sound Regional Council. (2024). *Household travel survey trips*. Retrieved from <https://psrc-psregenc.hub.arcgis.com/datasets/household-travel-survey-trips/explore>
- 442 R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

- 460 Ruggles, S., Flood, S., Sobek, M., Backman, D., Chen, A., Cooper, G., ... Schouweiler, M. (2024). *IPUMS USA: Version 15.0* (Version 15.0) [Data set]. <https://doi.org/10.18128/D010.V15.0>
- 461
- 462 Söbke, H., Harder, R., & Planck-Wiedenbeck, U. (2018). Two Decades of Traffic System Education Using the Simulation Game MOBILITY. In S. Göbel, A. Garcia-Agundez, T. Tregel, M. Ma, J. Baalsrud Hauge, M. Oliveira, ... P. Caserman (Eds.), *Serious Games* (pp. 43–53). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-02762-9_6
- 463
- 464
- 465
- 466 Southern California Association of Governments. (2012). *SCAG Regional Travel Demand Model and 2012 Model Validation* (p. 224).
- 467
- 468 Stewart, A. F. (2017). Mapping transit accessibility: Possibilities for public participation. *Transportation Research Part A*, 104, 150–166. <https://doi.org/10.1016/j.tra.2017.03.015>
- 469
- 470 Stewart, A. F., & Zegras, P. C. (2016). CoAXs: A Collaborative Accessibility-based Stakeholder Engagement System for communicating transport impacts. *Research in Transportation Economics*, 1–11. <https://doi.org/10.1016/j.retrec.2016.07.016>
- 471
- 472
- 473 Travel Forecasting Resource. (2020). Destination Choice: Theoretical Foundations. Retrieved July 31, 2024, from <https://tfresource.org>
- 474
- 475 Turochy, R. E. (2013). Structuring the Content of the First Course in Transportation Engineering: Perspectives of Engineers and Educators. *Journal of Professional Issues in Engineering Education & Practice*, 139(3), 206–210. [https://doi.org/10.1061/\(ASCE\)EI.1943-5541.0000149](https://doi.org/10.1061/(ASCE)EI.1943-5541.0000149)
- 476
- 477
- 478 US Census Bureau. (n.d.). *Longitudinal Employer-Household Dynamic Origin-Destination Employment Survey* [Data set]. Retrieved from <https://lehd.ces.census.gov/lodes/>
- 479
- 480 Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (Fourth). New York: Springer. Retrieved from <https://www.stats.ox.ac.uk/pub/MASS4/>
- 481
- 482 Walker. (2009, June 8). Did Sim City Make Us Stupid? Retrieved July 29, 2024, from <https://humantransit.org/2009/06/did-sim-city-make-us-stupid.html>
- 483
- 484 Walker, K. (2024). *Tigris: Load census TIGER/line shapefiles*. Retrieved from <https://CRAN.R-project.org/package=tigris>
- 485
- 486 Walker, K., & Herman, M. (2024). *Tidycensus: Load US census boundary and attribute data as 'tidyverse' and 'sf'-ready data frames*. Retrieved from <https://CRAN.R-project.org/package=tidycensus>
- 487
- 488 Wang, S., Bette, H. M., Schreckenberg, M., & Guhr, T. (2024, June 10). How much longer do you have to drive than the crow has to fly? <https://doi.org/10.48550/arXiv.2406.06490>
- 489
- 490 Weiner, E. (2013). *Urban transportation planning in the United States: History, policy, and practice*. New York, NY: Springer New York. <https://doi.org/10.1007/978-1-4614-5407-6>
- 491
- 492 Weir, J. A. (2004). *Active learning in transportation engineering education* (PhD thesis, Worcester Polytechnic Insti-

- 493 tute). Worcester Polytechnic Institute, United States – Massachusetts. Retrieved from <https://www.proquest.com/docview/305366413/abstract/FD41D00F66B9407EPQ/1>
- 494
- 495 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ... Yutani, H. (2019). Welcome to
496 the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- 497 Wickham, H., & Bryan, J. (2023). *Readxl: Read excel files*. Retrieved from <https://CRAN.R-project.org/package=readxl>
- 498
- 499 Zhou, J., & Schweitzer, L. (2009). Transportation Planning Education in the United States: Literature Review, Course
500 Survey, and Findings. *Transportation Research Record*, 2109(1), 1–11. <https://doi.org/10.3141/2109-01>
- 501 Zhu, S., Xie, F., & Levinson, D. (2011). Enhancing Transportation Education through Online Simulation Using an
502 Agent-Based Demand and Assignment Model. *Journal of Professional Issues in Engineering Education & Practice*,
503 137(1), 38–45. [https://doi.org/10.1061/\(ASCE\)EI.1943-5541.0000038](https://doi.org/10.1061/(ASCE)EI.1943-5541.0000038)