Election Cycling:
Quantifying the Impact of Bike Infrastructure on Local Elections and Democracy

Luke Kolar

Presented to the Department of Statistics
and the Department of Government
in partial fulfillment of the requirements
for a Bachelor of Arts degree

Harvard College
Cambridge, Massachusetts
April 1, 2024

Abstract

Abstract goes here

Acknowledgements

Acknowledgements go here

Contents

Chapter 1

Introduction and Overview

Chapter 2

Literature Review and Contributions

The contributions this thesis makes to the literature are twofold. First, literature on the interaction between American cities' ever-increasing bike infrastructure and its municipal elections is absent. Political punditry, on the other hand, follows phenomena like bikelash closely. This means our understanding of this connection is painted by a qualitative, anecdotal brush. These stories are informative, but they may exaggerate or extrapolate effects of bike infrastructure on elections. I bring a quantitative lens to the subject: big data on bike infrastructure, though decentralized and difficult, offers strong research opportunities.

The second contribution is more general. Literature on local elections and infrastructure policy – or infrastructure itself – generally lacks electoral analysis, instead frequently relying on survey or qualitative research. Feedback loops between public goods distribution and incumbent electoral performance or turnout are under-explored.

To begin, I define "bike infrastructure" and outline its growing role in contemporary U.S. cities. Then, I build three qualitative hypotheses for quantitative testing, exploring literature related to each of the analyses I perform and identifying existing gaps. Together, these three hypotheses and the distinct aspects of electoral politics they touch provide exciting potential for understanding how bike infrastructure shapes elections. These also relate to larger questions on the mechanics of urban political opinion, expression, and lifestyle.

The rise of bike infrastructure

Bike infrastructure constitutes anything ranging from bike lanes painted on an existing street to a new bike share station as part of a larger system. These are like most other public goods, part of a universe of budgetary and distributive political decisions. Yet the recent rise of bike infrastructure has been precipitous. The first painted bike lane was installed in Davis, California, in 1967[1]; today, most cities regularly update Bicycle Plans, which imagine vast networks of pristine, grade-separated cycleways making intracity trips convenient and safe on two wheels. Some cities, including the ones I analyze, are closer to that vision than others. Overall, though, United States government spending on bicycling infrastructure increased from a mere $22.9 million in 1992 to $970 million in 2017 (Nunno, 2021) as planners increasingly recognize cycling's importance to goals of health, climate, and mobility.

Those who study urban public goods provision are wary that bike infrastructure's *placement* is solely based on these goals, however, and distributive analyses constitute most research on bike infrastructure. The most holistic American analysis comes from Braun et al. (2019), who found across 22 large U.S. cities that in the mid-2010s, concentration of bike lanes was lower in more Hispanic, lower income Census block groups. But this analysis treated all lanes, from cheap paint to high–quality parkways, as the same treatment. For my analysis, quality is a crucial metric, since higher-quality infrastructure is likely more noticeable to voters. Research that distinguishes by quality is lacking in the United States, though international studies from Santiago de Chile (Mora et al., 2021) to London (Tait et al., 2022) find that wealthier areas generally receive higher-quality (safer) bike infrastructure.

Unequal distribution of bike infrastructure is not my primary focus, and including confounders in an electoral analysis helps control for disparities. But these studies highlight the relationship between bike infrastructure and urban demography. This in turn raises three hypotheses about the role of bike infrastructure in urban political participation.

[1]This is widely cited and reinforced by the City of Davis itself: `https://www.cityofdavis.org/about-davis/history-symbols/first-bicycle-lanes-in-davis`

Hypothesis 1: Bikelash and incumbency

Opposition to bike lanes is generally known as *bikelash*, and its study focuses on protests during planning stages. Scholars of this phenomenon find that bike lanes increasingly clash with two demographically-unique urban interest groups domestically.

First, *anti-gentrification activists* characterize bike lanes as "white lanes of gentrification" (Lubitow, 2016, 250); however, the mechanism by which this infrastructure has become associated with urban displacement is unclear. For instance, in the Shaw / U Street neighborhood in Washington D.C., a historically Black neighborhood that gentrified quickly in the 2000s, Hyra (2015, 1767) describes how new white, wealthy gentrifiers successfully fought for "limited parking, the removal of go-go clubs, bikes lanes and dog parks, which cater to their tastes and preferences." On the other hand, Betancur (2002, 806) analyzes gentrification in Chicago's West Town and concludes that the government "steered gentrification" to the neighborhood via "major infrastructural improvements." Whether gentrifiers spur infrastructure or planners use infrastructure to attract the so-called "creative class" of young professionals who accelerated downtown revival in the 2000s is unclear. The takeaway is that gentrification and bike amenities are closely linked in the minds of urban residents.

Second, typically older *motorists* see bike infrastructure as a threat to cars' unique cultural dominance in American cities. As the primary trade-off of high-quality bike projects that require expanded right-of-way, parking operates as a convenient microcosm of this culture. For instance, in *Paved Paradise*, Henry Grabar (2023) describes how parking scarcity intensely shapes urban behavior; competition for parking spaces culminates in murder surprisingly often in the United States. More broadly, Wild et al. (2018, 510) include motorist opposition to bike infrastructure under the banner of "conservatism," arguing that "conservative value commitments, in particular, the centrality of car travel to notions of family and economic responsibility; [and] the commitment to sub-urbanism [...] all play out in conservative resistance to progressive planning projects." Since bike projects in the 2010s and beyond are motivated by pedestrian safety and climate concerns, it is no surprise that conservatives resist these under the car-dominant regime of post-urban renewal U.S. cities.

Journalism on bikelash reinforces the primary opponents outlined by the literature: consider recent headlines like "In Mattapan, bike lanes divide the community: 'They're just trying to push us out'" from *The Boston Globe*[2], or "Bikes or cars? The battle over one Bay Area bridge lane is heating up" from the *San Francisco Chronicle*[3].

Unlike academia, journalism has also focused on bikelash in an electoral setting. Many cities have lobbying and advocacy groups on both sides of the bike debate, whose endorsements candidates seek and advertise. Take Cambridge, MA, as an example: in its 2023 City Council elections, the city's "Cycling Safety Ordinance" requiring a massive bike network expansion emerged as a key issue (Arshad, 2023). Pro-cycling advocacy organization Cambridge Cycling Safety distributed leaflets at bike share stations and knocked on doors; opponents of the ordinance touted their stance in interviews with local newspapers. Similar debates over cycling infrastructure happened in urban contests across over the past year: in Minneapolis, the *Star Tribune*[4] rhetorically asked City Council candidates whether a marginal budgetary dollar should go to "vehicles or [. . .] a bike lane," and in an unexpected Los Angeles culture war, firefighters launched a campaign against a bike and bus lane referendum[5]. The lack of research on electoral bikelash represents a significant gap contemporary urban political studies, in addition to a general lack of quantitative local electoral analysis.

Local political contests are unique; unlike those at higher levels, the vast majority are nonpartisan contests, and turnout is generally lower. As a result, J. Eric Oliver (2012, 179) demonstrates via survey research in *Local Elections and the Politics of Small-Scale Democracy* that the local electorate is "far more interested, engaged, and informed about

[2] https://www.bostonglobe.com/2022/08/01/metro/mattapan-bike-lanes-divide-community-theyre-just-trying-push-us-out/

[3] https://www.sfchronicle.com/bayarea/article/richmond-san-rafael-bridge-bike-lane-future-18444983.php

[4] https://www.startribune.com/2023-city-election-minneapolis-city-council-mayor-jacob-frey-ward-mpls-candidates/600305882/

[5] https://www.latimes.com/california/story/2024-02-14/firefighters-launch-campaign-against-measure-hla

local politics." Across localities of different sizes and features, disapproval of local government performance is correlated with votes for challengers, and relatedly, issue-sharing between a voter and candidate is associated with votes for that candidate. These suggest that engaged anti-gentrification activists and car culture warriors are more likely to vote against an incumbent pushing a cycling agenda.

The question of *geography* remains, and few studies have looked at the place-based impacts of infrastructure policy. On similar issues, however, issue-based activists overwhelmingly emerge from near the site of a controversy. In studying the composition of "NIMBYs," or those who routinely block high-density housing projects, Einstein (2020, 102) found in *Neighborhood Defenders* that "eighty-two percent of commenters live in the same census tract [...] as the housing development" they intend to block with requests for delays or more red tape reviews. NIMBYism is a useful analogy for bikelash, given that the older, home-owning demographic composition of housing opponents parallels the profile of conservative cycling opponents (Wild et al., 2018). While housing developments themselves are largely led by private developers, research focusing on public infrastructure supports the idea that voters punish incumbents for unpopular distribution. Burnett and Kogan (2017), for instance, find that pothole complaints are negatively correlated with mayoral and city council member incumbent vote share.

Thus, a reasonable theory of bikelash suggests that we should see the most pronounced incumbent vote share decrease in areas with higher concentrations of older automobile users and areas at risk of or actively gentrifying. Though literature is not helpful in explaining effects beyond these, reporting on bikelash in elections suggests that substantively, bikelash will be less pronounced or absent elsewhere. Formally:

Hypothesis 1 (H1): Bike lanes are unpopular among (a) older, car-commuting and (b) gentrifying neighborhoods; so they will punish incumbents upon receiving new lanes.

Hypothesis 2: Bike lanes and local government visibility

A defining feature of local elections is their low turnout. In Houston, the nation's fourth-largest city with almost 2.5 million residents, only about 17% of eligible voters cast a ballot to select their mayor in December 2023 (Sledge, 2023). Compare this to the 2020 U.S. presidential election, where 60% of Harris County, TX, voters cast ballots (Kanik, 2022).

Studies on the impacts of extraordinarily low local participation detail many associated effects, including lower minority representation in local government (Hajnal and Trounstine, 2005) and unequal municipal spending incentives (Trounstine, 2013). Oliver (2012), however, argues that low turnout is not an existential problem for localities, as citizen preferences are more homogenous at the local level, and the number of divisive issues are self-evidently fewer at a less powerful level of government. Furthermore, homeownership is strongly correlated with mayoral election turnout (Jiang, 2018), leading (Oliver, 2012, 84) to argue these voters have "strong emotional and material connections to their communities [...] exemplifying all that we expect in a classical notion of the ideal democratic citizen."

Some may find this problematic beyond class undertones; most large U.S. cities are densifying in large part due to growth in renters. Progressive cities encourage this growth for reasons economic, cultural, or environmental, and they may consider higher renter turnout healthier for local democracy as this population segment grows.

As such, many researchers have turned attention to interventions that increase participation in municipal elections. Perhaps the most common is moving these "on-cycle" with federal elections, as most local contests take place in odd-years (Hajnal et al., 2022). However, local races, particularly in larger cities, often involve early primaries, runoff elections, or ranked-choice ballots; this intervention involves more system upheaval than one might expect. Others note that increasing campaign spending, introducing partisanship, or higher contestation in mayoral elections may increase turnout (Holbrook and Weinschenk, 2014; Marschall and Lappie, 2018). But these changes themselves may bring negative political externalities, and an outside treatment like encouraging higher spending may be difficult.

While literature discusses the above structural reforms or vote choice mechanisms that treat turnout as constant, few have explored how increased local government action on display, what I call *visibility*, impacts turnout. Even local politicians are unsure of the relationship between their distributive spending and its mobilization effects (Dynes, 2020). Just as literal local election reminders increased turnout (Hopkins et al., 2023), local government action on display may serve as a constant reminder to vote. In other words, any local government action resulting in visible change – perhaps a street reconstruction — may increase the salience of future local elections for nearby residents. A potential issue may be that citizens are bad at assigning responsibility for government services in other areas like train delays (de Benedictis-Kessner, 2018). But early evidence supports the theory that at least some of the electorate may respond to higher local government visibility: Hajnal and Lewis (2003), for example, demonstrate that cities providing more in-house services like trash collection or libraries, as opposed to outsourcing, see higher turnout.

Bike lanes offer an opportunity to test whether more local government action coincides with greater electoral participation by way of increasing its visibility to voters. In the past two decades, many cities have changed their approach to a core service: street maintenance. Mill and overlay projects increasingly incorporate a painted bike lane; full reconstructions more frequently integrate grade-separated bikeways. Streets are perhaps the most visible, ubiquitous element of a city, and one might imagine significant changes to its layout, such as removing parking or lanes of traffic for cyclists, are among the most salient *everyday* evidence of local government in action. In voters' minds, this may up the stakes, or simply serve as a reminder that local government and its elections exist at all. Finally, in a city with stagnant bike ridership across a period of bike lane network growth, one can more confidently isolate the impact of visibility versus changes in urban mobility. Formally:

Hypothesis 2 (H2): Bike lanes amplify local government visibility, increasing the salience of municipal elections and thus increasing local turnout.

Hypothesis 3: Bike share stations and going-to-the-polls

The previous hypothesis focuses on how local government infrastructure may increase the salience of a local election, but since urban transportation infrastructure is usually meant to ease movement, another theory suggests bike infrastructure may make the *act of voting* easier in dense cities. The popular rational choice model of turnout portrays potential voters' choice as a cost-benefit analysis (Riker and Ordeshook, 1968). Aldrich (1993, 248) notes that costs include those "of obtaining information, processing it, and deciding what to do and the direct costs of registering and going to the polls." But little is known about the "*going*" in "going to the polls," and for those interested in boosting voter turnout, lowering transportation barriers to voting is important.

Despite the traditional American view of voting as being in person, on the first Tuesday following the first Monday of November, it is true that absentee voting or voting by mail has increased significantly in recent cycles. Yet the in-person picture is still largely accurate: during the 2020 COVID-19 pandemic, 54% of voters cast their ballots in-person, and over half of these were on Election Day[6]. In the 2022 midterm elections, levels of in-person Election Day voting rebounded to nearly half of all votes[7], down from 60% in 2018, but still demonstrating a regression to pre-pandemic patterns of voting.

Literature on geography and costs to participation has focused largely on the placement of polling places. Findings agree that turnout and distance to a polling place are related, but models differ: these range from findings that distances exhibit the largest negative effect at close changes in distance (Haspel and Knotts, 2005), further distances (Burke et al., 2019), or the middle ground of suburbs (Gimpel and Schuknecht, 2003). Survey research finds that transportation is indeed a barrier to voting; nearly 30% of nonvoting respondents in a 2016 survey listed lack of transportation as a reason for skipping the election, though

[6] https://www.pewresearch.org/politics/2020/11/20/the-voting-experience-in-2020/

[7] https://www.eac.gov/sites/default/files/2023-06/2022_EAVS_Report_508c.pdf

the geography of these nonvoters and their mode of transportation is unknown (Stewart, 2017). Notably, studies overlook explicitly urban contexts, where urban form, non-motorized mobility, and public transportation may have a greater effect.

Despite a lack of concrete evidence, interventions that lower transportation barriers are common in recent elections. The FTA notes that ten large urban transit systems in America, from Los Angeles to Dallas, launched free-fare services on Election Day in 2020[8]. Some private companies like Lyft continued initiatives offering discounted rides to the polls in 2020[9]. Bike share systems, including Citi Bike in NYC[10], Divvy in Chicago[11], and Blue Bikes in Boston[12], have offered free or lower-cost rides to users for the past three federal elections.

These bike share systems present an intriguing opportunity for studying transportation patterns and turnout on Election Day, especially in the nation's densest neighborhoods. Expanding bike share systems in cities like New York City and Boston have seen corresponding massive increases in ridership, and patterns of traveling to polling places may change when these are located near stations. Furthermore, in dense neighborhoods, higher concentration of bike share stations in election precincts may make transportation to polling places easier, marginally reducing the costs of voting. Bike share ridership data allows for distinguishing this mobility effect from the previous hypothesis on visibility. Formally:

Hypothesis 3 (H3): Bike share infrastructure facilitates transportation to polling places on Election Day, and in dense neighborhoods, thus increasing total electoral turnout.

[8] https://www.transit.dot.gov/funding/grants/transit-programs-increase-access-voting

[9] https://www.lyft.com/blog/posts/helping-more-people-get-to-the-polls-on-election-day

[10] https://www.thrillist.com/news/new-york/lyft-offering-off-rides-citi-bikes-election-day

[11] https://www.nbcchicago.com/news/local/chicagoland-divvy-bikes-offers-free-rides-on-election-day/170157/

[12] https://blog.bluebikes.com/blog/bike-to-vote-2020

Conclusion and next steps

In summary, I hypothesize three ways bike infrastructure impacts our elections. First, some groups – older car commuters or anti-gentrification activists – may oppose bike lanes, affecting incumbent electoral performance (H1). Second, bike lanes may serve as a reminder to residents of local government in action, increasing turnout (H2). Third, bike share systems may make traveling to polling places easier in urban areas, increasing turnout (H3). Each of these hypotheses requires an independent test, given the diversity of bike infrastructure data, electoral systems, and urban dynamics across different cities.

First, in **Chapter 3** on Portland, Oregon, a national leader in urban cycling, I test H1 by comparing mayoral incumbent reelection performance across neighborhoods both altered and untouched by the city's expanding high-quality bike network. **TAKEAWAY:**

Second, in **Chapter 4** on Seattle, Washington, another cycling leader with more elected turnover and significantly more precincts, I test H2 by looking at the association between local electoral turnout and the distribution of bike infrastructure between cycles. **TAKEAWAY:**

Third, and finally, in **Chapter 5** on Manhattan, New York City, I explore H3 by studying election turnout at several levels, the expansion of bike share system Citi Bike, and ample data on Election Day system ridership. **TAKEAWAY:**

Together, these amount to first steps in the study of how bike infrastructure interacts with electoral politics – and the dynamics of electoral politics in the first place.

Chapter 3

Electoral bikelash in Portland

This chapter aims to quantify the relationship between the geographic distribution of new, protected bike lanes and the reelection performance of an incumbent overseeing it. Specifically, I test:

Hypothesis 1 (H1): Bike lanes are unpopular among (a) older, car-commuting and (b) gentrifying neighborhoods; so they will punish incumbents upon receiving new lanes.

I examine Portland, OR, Mayor Ted Wheeler's vote share change by precinct between the city's 2016 and 2020 primary elections as a case study in electoral bikelash [Figure 3.4]. I first cleaned precinct and election data, then projected electoral factors including neighborhood gentrification stage, age, proportion car-commuting, and population density onto precincts. To test the impact of bike infrastructure, I categorized bike lanes by quality and year using several sources and constructed two treatment options. Treatment 1 uses the proportion of new bike lanes to a precinct's total street grid, since lanes in most cases follow existing roads [Figure 3.2]. Treatment 2 uses precinct percent overlap with 15-minute "walksheds" for points imputed on bike lanes, or shapes representing areas that are a realistic 15-minute walk away from this infrastructure [Figure 3.3].

My analysis consists of linear regression models with confounders, treatment effects distinguishing by infrastructure implementation and quality, and interactions between these.

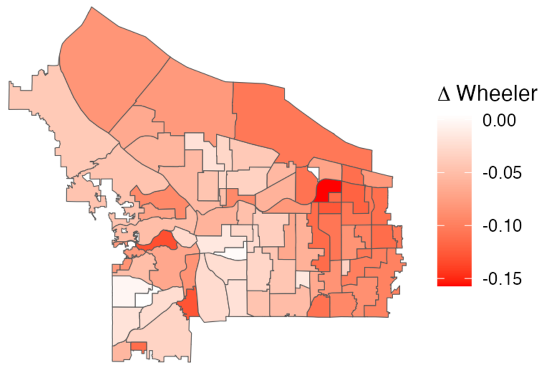
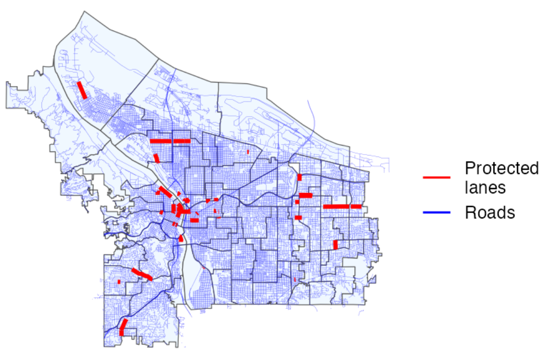


Figure 3.1: Wheeler vote share change, 2016-2020



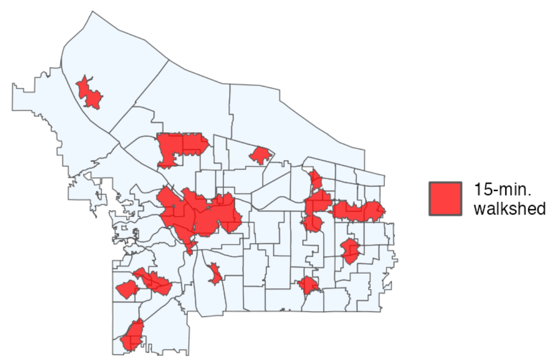


Figure 3.2: Protected bike lane aggregated 15-minute walksheds, whose overlap with precinct areas constitutes *Treatment 1*

Figure 3.3: Protected bike lanes overlaid on the street network, whose proportion by precinct constitutes *Treatment 2*

My results somewhat contradict *H1*, indicating that the there is no statistically significant relationship between the *geographical placement* of bike infrastructure and incumbent backlash when controlling for confounders. However, variables associated with groups more likely to exhibit *bikelash* – older, car-commuters and those in gentrifying neighborhoods – did punish Wheeler significantly more in his second election. This suggests while *electoral bikelash* is not place-based, it may exist citywide.

I end by discussing the implications of these findings, including that bike-forward politicians concerned with electoral performance need not worry about *where* lanes are distributed; residents far from these projects may punish a pro-cycling agenda regardless.

Why Portland, Oregon?

Portland, Oregon, is among a select few American cities lauded for its Scandinavian-style bike network and corresponding bike-forward political culture. The city reports having "over 400 miles of bikeways, including 100 miles of low-stress neighborhood greenways," thanks to an early start in drafting its first Bicycle Master Plan in 1973[1].

Mayor Ted Wheeler, elected in 2016, bolstered Portland's reputation as a top American biking city in his first term, overseeing a massive expansion of the city's protected cycling network. Urbanist analytics website Walk Score ranked Portland second on its "Bike Score" rankings in 2021[2], and Portland has consistently held its place among national cycling advocate People for Bikes' top five large U.S. biking cities since Wheeler entered office[3].

From a research perspective, three features make Portland the ideal city for studying *electoral bikelash*. First, among top biking cities, Portland is one of few which has seen an incumbent run in a close reelection race. To measure incumbency punishment, it is necessary to have an incumbent's original and reelection years coincide with a period of significant bike network expansion; in Portland, like many cities, this took place in the late 2010s.

Second, Ted Wheeler's elections both occurred at the latest possible four-year window in this decade, 2016 and 2020. Both elections taking place within the same decennial U.S. Census period means that precinct boundaries are largely unchanged in Portland. Furthermore, since precincts are constructed from Census blocks, the latter provide a common, comparable source of demographic information across the election cycle.

Third, and finally, Portland is one of very few cities with accurate, extensive data on a bike lane's most recent upgrade. Other top cycling cities considered for this analysis release only partial or inaccurate data on timing or lanes themselves.

[1] https://www.portland.gov/transportation/walking-biking-transit-safety/biking-portland

[2] https://www.walkscore.com/cities-and-neighborhoods/

[3] Various, from: https://www.peopleforbikes.org

Outcome: Incumbent vote share change

Portland holds nonpartisan mayoral elections in presidential years: a first round primary takes place in May, and a runoff occurs in November if no candidate wins a majority. Ted Wheeler won his first election outright in 2016, so my analysis uses May primary results for 2016 and 2020.

Using May primary results may be the only longitudinally-comparable option in Portland, but it also has some advantages over a higher-turnout November mayoral election in testing *H1*. Both mayoral primaries saw similar turnout around 50%, which is similar to off-year mayoral races common in most large U.S. cities. The May primary also allows the most options for voting against an incumbent, whereas binary-choice November runoffs may force anti-cycling voters to vote for an incumbent against a more aggressively pro-cycling challenger. There is no perfectly generalizable American local election, since U.S. cities hold mayoral elections by a variety of systems taking place throughout the year[4], but this set of elections is a reasonable case study when keeping caveats in mind.

Election precinct shapefiles were crucial for calculating treatment effects and aggregating confounders. These shapefiles for Portland come from GIS analysts Kevin Rancik[5] and Ian R. McDonald[6] for 2016 and 2020, respectively. Both used statewide data to reconstruct these for analysis of presidential results. One cluster of precincts were aggregated in 2020 but not 2016; I combined these and their statistics for both cycles. Precincts visually matched city-released overview maps, and precinct shapes overlapped across years within 1% error, suggesting technical differences and not boundary changes themselves. Chapters on Seattle and Manhattan elaborate on the tedious process of aligning precincts across years; thankfully for Portland, algorithmically merging precinct shapes and data was unnecessary.

[4] For example: New York City has partisan elections, and given its Democratic bent, the winner is essentially crowned when the Democratic primary race is called in June. Chicago's elections kick off with a first round in February and a runoff in April. Etc.

[5] `https://www.kevinrancik.com/home`

[6] `https://www.ianrmcdonald.com`

Election returns come from Multnomah[7], Clackamas[8], and Washington Counties[9] in Oregon[10]; the vast majority of Portland and its precincts fall into the former. These were released in PDF document form; some were cleanable using R packages like `tabulizer`, others required manual entry. I extensively verified the accuracy of these via official summary statistics and randomly-selected precincts.

I removed all precincts with fewer than 150 votes cast in 2016. These were largely on the perimeter, with small land areas, whose outcome variable was skewed by small changes in net Wheeler change that resulted in larger percentage swings. Removing these also alleviated issues in collecting confounders using block group overlap as outlined later. I selected the cutoff of 150 as it conveniently resolved the several above issues at once, and a steep increase in votes cast begins after this threshold. For i in 78 resulting precincts or combined precincts [Figure 3.4], our outcome variable is defined as[11]:

$$\frac{(\text{2020 Wheeler votes})_i}{(\text{2020 total votes})_i} - \frac{(\text{2016 Wheeler votes})_i}{(\text{2016 total votes})_i} = (\Delta\text{Wheeler})_i$$

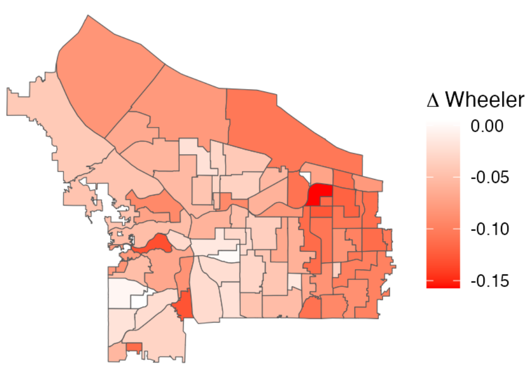


Figure 3.4: Wheeler vote share difference by precinct, 2016-2020

[7]https://www.multco.us/elections

[8]https://www.clackamas.us/elections

[9]https://www.washingtoncountyor.gov/elections

[10]Portland is a rare large U.S. city with a handful of residents who vote in a separate county, seemingly outside of the city limits. Most of these precincts were small and filtered out in the end.

[11]Note that the result is Δ% Wheeler, not %Δ Wheeler, which is to say I use a simple difference in percents and not the percent change in percent.

Treatment: Protected bike lanes

Since *H1* revolves around bikelash fueled by older motorist conservatism and anti-gentrification sentiment, I chose bike lanes as a treatment. These are disruptive for automobile users and the dominant symbol of bike infrastructure as gentrification. While a later chapter treats bike share stations as infrastructure, Portland's own system, BIKETOWN, is less prominent than others nationally and relatively uniformly dispersed across the city[12]; its effect as a treatment is likely minimal across a mere 78 precincts. Additionally, public system data stretches only as far back as September 2020, after elections measured here.

There are many methodological challenges to analyzing bike infrastructure, but three general questions stand out: how to conceptualize the *timing* of an introduction, how to capture *quality*, and how to measure primary *recipients* of a project based on its placement. My analysis explores recency, quality, and geography, so all are important variables. I make several assumptions which simplify the above.

First, on the question of *timing*, I consider a bike lane's "introduction" to be when its construction is completed. It is reasonable to believe that most residents of a given election precinct see bike infrastructure projects as discrete change represented by a "before" and an "after." In the long run, voters, especially long-term residents, will remember a change for these two longer-term states. To be sure, construction is an annoyance, but it happens constantly all around any given city; this analysis seeks to look at bike infrastructure as a treatment, not its implementation process. Additionally, in inspecting all rows in the data individually, I found that most projects are on a relatively quick construction schedule. These projects typically start and end within the six month "warm weather" window and all are complete by Election Day.

Second, I propose categorizing the *quality* of bike infrastructure by how easy it is to ignore, not by the comfort of its use. Many more Portland voters drive cars than bike; the

[12] `https://account.biketownpdx.com/map`

question of quality is partly how disruptive it is to car users. Additionally, higher quality infrastructure is more visible as a symbol of gentrification to some groups. I divide bike infrastructure into two categories of quality: *painted* and *protected*.

Painted lanes are self-explanatory: these follow a street network and require essentially no construction but technically allocate dedicated space to cyclists [Figure 3.5]. Standard bike lanes typically sit on the shoulder of a road as painted white lines, sometimes between traffic and parking, and provide little protection from drivers. Others, such as those running through downtown Portland, are bright green strips of the road, more visible to non-cyclists yet just as easy to drive over. Buffered bike lanes, which place thicker double-striped lines or patterned neutral space between bike and traffic lanes, legally occupy more of the street but are, again, just as easy to drive over. Installing a painted bike lane changes little about the topography of a street, meaning disgruntled motorists can simply refuse to change bad habits, like refusing to look over one's shoulder or speeding through a turn. It is for this reason that some cyclists complain painted bike lanes are merely symbolic. Painted bike lanes are less important for my analysis, and I treat both variations as one.

Protected bike lanes are more complicated as they vary in convenience to motorists more widely. These are like bike lanes in that they follow a street network, but some form of barrier exists physically preventing collision between cyclists and motorists. In Portland, many forms are categorized under the banner of "protected bike lane." The most common barrier are "flex posts," a type of bollard that emergency vehicles can easily drive over but strongly disincentivizes driving in a bike lane [Figure 3.6]. Others involve small raised curbs, usually laid out in a dashed line with open intervals where driveways exist. Many involve a combination of flex posts and curb sections, particularly when driveways are present along a route [Figure 3.7]. Finally, the highest-quality projects are grade-separated, serving as more of an extension to the sidewalk than something that looks like part of the street [Figure 3.8]. All of these are highly visible and significantly more disruptive to drivers than paint, though construction schedules vary.

Figure 3.5: *Painted* lanes on SE Belmont Street

Figure 3.6: *Flex post / bollard* lanes on SW 49th Avenue

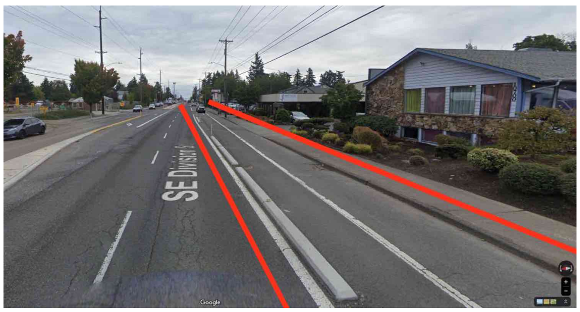


Figure 3.7: *Flex/curb mixed* lanes on SE Division Street

Figure 3.8: *Fully protected* lanes on SW Capitol Highway

Figures above display the four categories of data I used[13]. Bike lane data comes from the City of Portland[14], which regularly updates shapefiles for all existing bike routes.

I split this data into two subsets. The first encapsulates the *painted* category, or all standard and buffered bike lanes. The second includes all *protected* lanes. Some segments were named, and most routes had an associated year of construction. To verify the segments included in the protected lanes subset across all years, I individually "visited" each one on Google Street View, using historical imagery to find pre- and post-construction years. I saved hyperlinks to images for each of these [see Appendix # for more details]. This not only allowed me to confirm all construction years were correct, but it also helped me ensure no recent (post-2020) construction projects were built as lanes of a lower quality in between the 2016 and 2020 elections.

[13] All examples are taken from the final data set of lanes, with images from Google Street View (red lines added to delineate the lane): `https://www.google.com/streetview/`

[14] `https://gis-pdx.opendata.arcgis.com/datasets/PDX::bicycle-network/about`

Finally, this allowed me to categorize each protected segment into two types based on protection: *type 1* includes flex post only lanes [Figure 3.6] or those with a combination of flex posts and raised curbs [Figure 3.7], and *type 2* includes grade-separated or fully curb protected lanes [Figure 3.8].[15]

Filtering for projects constructed from 2016 to 2020[16], and converting all to the `sf` "linestring" type, the final protected data set includes 61 shapefile segments, some of which were part of the same project. At the segment level, there were 48 of *type 1* and 13 of *type 2* [see Appendix # for maps of these by *type*, year, and more]. I took painted lane installation dates as given upon confirming dates for protected lanes and filtered in the same way.

Having defined the *timing* and *quality* components, the final question revolves around *recipients* of bike infrastructure. I propose two treatment methods to quantify geographical proximity to bike lanes, since these should be more salient to those nearer to them.

<u>*Treatment 1*</u> (T1) is defined as new bike lane length as a proportion of a precinct's total street network. To gather the Portland street network, I used the `tinytiger` package[17], which downloads the U.S. Census Bureau's TIGER/Line shapefiles on roads for selected counties. I calculated the total length of roads within each precinct, as well as the length of all paint and protected projects in that precinct. For protected lanes, I distinguished between construction years to analyze project recency and treatment. For each precinct, I calculated the total proportion of painted project lengths to total street grid length, and the total proportions of protected project lengths by year and type (type 1 versus type 2, as defined above) to total street grid length. For i in 78 precincts, and for category $= c \in$ $\{paint$; *protected*: *all*; *protected*: *type* 1 by *year*; *protected*: *type* 2 by *year*$\}$, T1 is:

$$\frac{(\text{Total bike lane length})_{c,i}}{(\text{Street grid length})_i} = (\text{T1})_{c,i}$$

[15]Note that I combined the flex post and flex post / curb combination categories as the latter only had six examples, and these were similarly "disruptive" to the streetscape visually and in construction.

[16]I included 2020 lanes since for all 11 in this subset, installation had occurred. The COVID-19 pandemic and civil unrest that year may have slowed the infrastructure pipeline temporarily.

[17]`https://alarm-redist.org/tinytiger/index.html`

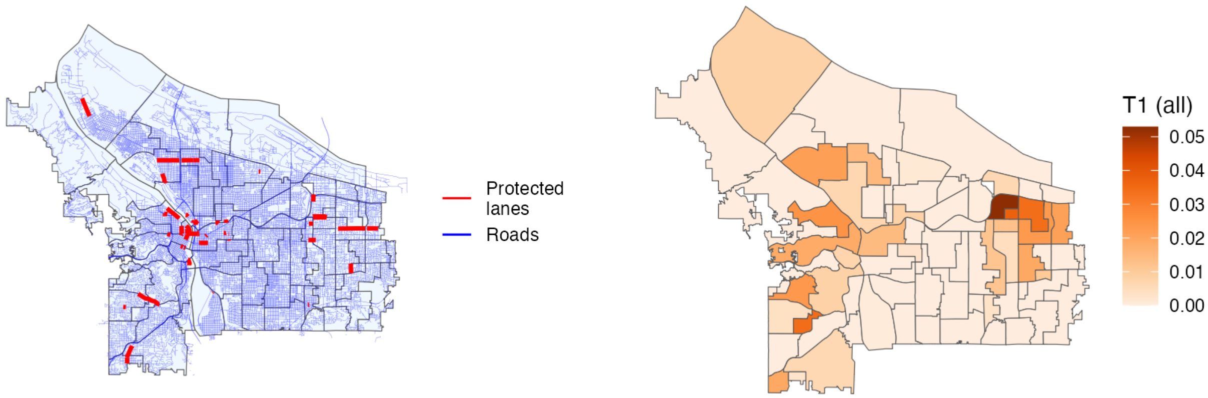
Treatment 2 (T2) is defined as the proportion of each precinct area within a 15-minute walk of bike infrastructure. The complexity of this process meant I calculated T2 for protected *Type 1* and *Type 2* lanes only, option to use only T1 for painted lanes throughout. I first imputed points on each line at 500 meter intervals, or at the line's centroid if it were sufficiently short. I then used the `r5r` package[18] to compute 15-minute *walksheds*: polygons that represent a 15-minute walk radius from each point, calculated by crawling the sidewalk network. Walksheds are useful because they avoid features like water or highway impasses, more accurately capturing *experienced* proximity to an urban feature. Admittedly, 15 minutes may seem an arbitrary threshold, but this number is commonly cited in urbanist circles as a key metric for one's core neighborhood.[19] For separate analyses, I joined these polygons either by year, or overall: if I were ignoring *timing*, for instance, I first merge all walksheds regardless of implementation year, then calculate overlap. This way, like the set of T1 variables, the range of T2 is zero (no walkshed overlap) to one (full overlap). For i in 78 precincts, and for category $= c \in \{protected\text{: } all\text{; } protected\text{: } Type\ 1 \text{ by } year\text{; } protected\text{: } Type\ 2 \text{ by } year\}$, *Treatment 2* is defined as:

$$\frac{(\text{Bike lane walkshed overlap area})_{c,i}}{(\text{Total area})_i} = (\text{T2})_{c,i}$$

These treatments are imperfect, but their underlying data is high-quality, and they capture different aspects of one's experience with bike infrastructure. *Treatment 1* captures the potential compounding effects of bike infrastructure, where under *Treatment 2*, two nearby projects are coded the same as one, since their overlap is a joined walkshed polygon. *Treatment 2*, however, has the advantage of capturing proximity to bike infrastructure, as walksheds can cross precinct boundaries, such as in the case of a project being near the edge of a precinct. Below, Figure 3.9 illustrates the translation of lanes to precinct-level treatment via T1; Figure 3.10 illustrates the same for T2.

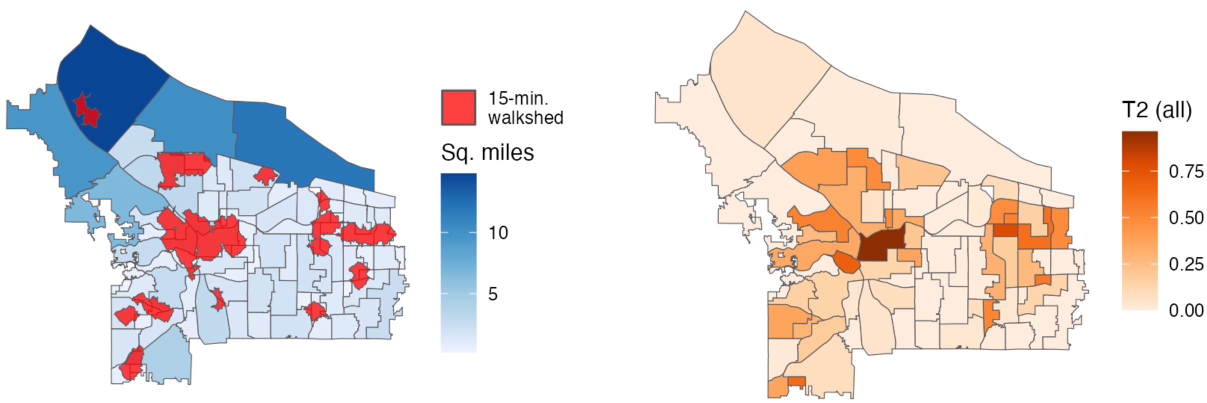
[18] `https://ipeagit.github.io/r5r/`

[19] Take, for instance, the popular concept of the "15 minute cite": `https://www.15minutecity.com`



(a) Protected lanes overlaid on street grid (b) T1: lane coverage of street grid by precinct

Figure 3.9: *Treatment 1*, from all protected lanes to street grid proportions



(a) Protected lanes overlaid on precinct area (b) T2: 15-min. walkshed overlap by precinct

Figure 3.10: *Treatment 2*, from all protected lanes to walkshed overlap proportions

Again, note the subtle differences between T1 and T2. In the former, the center downtown area – with visually high bike lane coverage – is not the most "treated" of all precincts, since downtown has a similarly dense street grid. Instead, the more residential northwest precincts are at the T1 upper end. In T2, the downtown area is the most "treated," since it is almost completely covered by protected lane walksheds. T1 is also a much smaller variable numerically, since Portland's street network is universally larger than its bike network. In summary, T1 captures concentration, while T2 captures coverage.

Confounders: Car-commuting, age, gentrification, etc.

For a full picture of electoral bikelash, I considered both the factors cited in *H1* and others that may correlate with changes in incumbent performance. The former includes a measure of gentrification and variables capturing motorist conservatism including motor use and age. The latter whittles down to population size and density.

The first confounder crucial to my analysis is gentrification. The City of Portland came under scrutiny in the late 2000s as urban scholars noticed its population growth increasingly amounted to displacement, especially for nonwhite, lower-income residents. In response, the city has conducted numerous studies documenting neighborhoods at-risk or actively gentrifying. One such study in 2018[20], uses pre-treatment data from 2016 and earlier and considers whether a Census tract "has a vulnerable population, has experienced demographic change, and [...] housing market conditions," categorizing some tracts as susceptible or in early gentrification stages, and others as actively gentrifying. I recreated this tract index as shapefiles, grouped "early" and "susceptible" areas as pre-gentrification, and kept these as well as actively gentrifying areas [see Appendix # for visualization]. Census tracts do not align perfectly with precincts, so I created variables for *percent overlap with pre-gentrification* (pGA) and *gentrification areas* (GA). For i in 78 precincts, these are:

$$\frac{(\text{Pre–gent. overlap area})_i}{(\text{Total area})_i} = (\text{pGA})_i \qquad \frac{(\text{Gent. overlap area})_i}{(\text{Total area})_i} = (\text{GA})_i$$

Next, I wanted to capture the demographic profile of conservative bikelash: older motorists. The American Community Survey (ACS) releases five-year estimates for a number of demographic variables at the block group level, including *percent commuting by automobile* and *median age*. These estimates are averages over a pre-treatment period from 2011-2016.

[20] https://www.portland.gov/sites/default/files/2020-01/gentrification_displacement_typology_analysis_2018_10222018.pdf

Using the tidycensus[21] package, I collected several ACS variables to calculate an estimated employed population (population commuting to work plus population working from home), estimated workforce (total population subtracting population under age 16 and non-labor force participating population above age 16), and estimated employment rate (the ratio of the two) by Census block group (BG). I then calculated the percent of a block group's population commuting to work via the following formula:

$$\frac{\text{Pop. auto–commuting}}{\text{Pop. workforce}} \times (\text{Employment rate}) = \%\ \text{auto–commuting [BG]}$$

With total proportion commuting by automobile calculated and median age collected for each Census *block group*, I then had to convert these to the precinct level. Importantly, precincts are constructed from Census blocks, as are Census block groups, meaning precinct and block group boundaries often overlap (Amos et al., 2017). My solution to this is what I term *population-weighted block group (pwBG) projection.*

First, I collected each Census block's 2010 populations using the censable[22] package. I use 2010 populations as block-level counts are only available in Census years, 2020 data is post-treatment, and 2020 data accuracy is impacted by differential privacy[23]. Next, I multiplied statistics of interest – here, BG *percent commuting by automobile* and *median age* – by each contained block's 2010 population. This is akin to weighting each one.

Finally, matching each 2010 Census block to the precinct encompassing it, I sum all population-weighted statistics, and divide by the sum of all block populations within the precinct. Formally, to find numerical variable V for precinct i with N_i blocks $\{b_1, ..., b_{N_i}\}$, each with corresponding block group V statistics for $\{v_{b_1}, ..., v_{b_{N_i}}\}$ and population $\{p_{b_1}, ..., p_{b_{N_i}}\}$,

[21] https://walker-data.com/tidycensus/

[22] https://christophertkenny.com/censable/

[23] https://www.census.gov/content/dam/Census/library/factsheets/2021/differential-privacy-and-the-2020-census.pdf

I used the following function:

$$f_{pwBG}(V, i) = \sum_{j=1}^{N_i} \left(\frac{v_{b_j} \cdot p_{b_j}}{p_{b_j}} \right)$$

To find pwBG-projected precinct proportion auto-commuting (Auto) and median age (Age), for i in 78 precincts, we have:

$$f_{pwBG}(\text{Prop. auto–commuting}, i) = (\text{Auto})_i \qquad f_{pwBG}(\text{Median age}, i) = (\text{Age})_i$$

Our pwBG-projected median age and proportion commuting by automobile variables are estimates using estimates: our block group estimates are projected onto blocks, and precinct estimates are generated by weighting these by block population. This introduces caveats to the analysis, but it is reasonable to assume these variables accurately capture precinct characteristics. The accuracy of *projecting* block group features onto blocks relies on the assumption that these features are homogenous within a block, which is reasonable since block groups are in part selected based on common neighborhood features. The "average of medians" variable for age, for instance, was both the best option and an accurate one: it captures age across Portland well when compared to Census tracts and other data points[24].

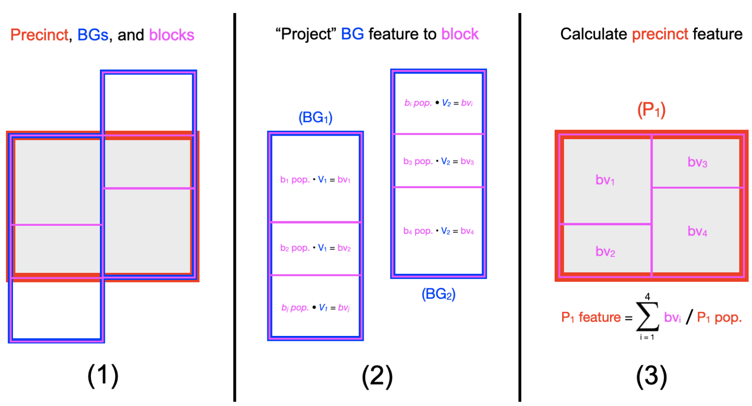


Figure 3.11: The intuition and math behind pwBG-projected variable V.

[24] I spent time verifying precinct maps matched Census tracts colored by age. I also investigated outliers, finding that older precincts contained nursing homes, younger ones were university housing, and more.

Calculating these algorithmically is complicated but made easier using functions from several aforementioned packages. Visualization makes the process intuitive, however; Figure 3.11 above illustrates this formula for one precinct.

Finally, I computed a *population density* (PD) variable for 2020, as denser neighborhoods may be more urbanist and thus more receptive to pro-biking policies, and these same neighborhoods may have higher renting rates and thus more residential turnover not captured in population change. This meant dividing 2010 precinct population from Census blocks by precinct area. For i in 78 precincts:

$$\frac{(\text{Population})_i}{(\text{Total area})_i} = (\text{PD})_i$$

Note that I initially gathered variables for race and income for this analysis; however, the gentrification variables largely captured race in Portland, given racial demographics were a major input for the classification. Income, too, correlated strongly with the gentrification metrics and somewhat with age. I was interested in employing a thorough, but elegantly simple suite of variables, and decided only those enumerated so far were necessary.

Table 3.1 summarizes the variables I collected, their sources, an overview of the processes by which I cleaned and formatted these, and their histograms. The outcome variable histogram displays a somewhat normal distribution and shows Wheeler performed worse in most precincts in 2020. The majority of precincts have near-zero exposure to T1 or T2 as these are right-skewed[25]. The Auto, Age, and PD variables appear normally distributed as is ideal for linear regression. The gentrification metrics of pGA and GA are bimodal and skewed, respectively, but for sake of interpretation, I did not transform these; pGA functions much like an indicator variable in this way.

[25] I use various iterations of T1 and T2, aggregating or separating protected lanes by quality or year. I also use T1 for aggregated painted lanes. *Importantly*, the histograms in Table 3.1 show protected lane treatments across all years and both qualities – the subsets have similar distributions.

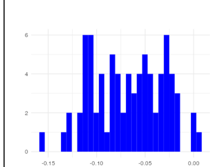
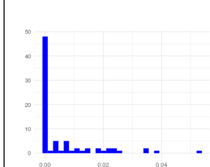
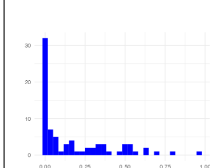
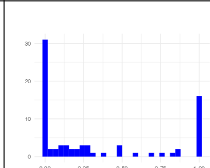
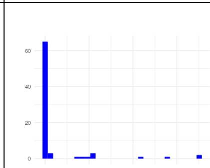
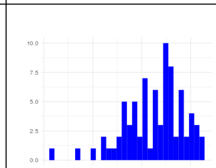
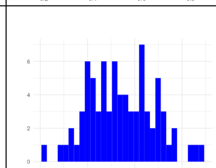
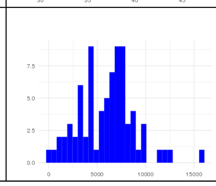
	Variable[26]	Sources	Process	Histogram
Outcome	**ΔWheeler**: Wheeler vote share change, (‘16-‘20)	Oregon counties; GIS analysts	Filtering; calculation	
Treatment	**T1**: % new bike lanes to street length	Portland; Census GIS via `tinytiger`	Verification; calculation	
Treatment	**T2**: % bike lane 15-min. walkshed overlap	Portland; walksheds via `r5r`	Verification; imputation; walkshed build; calculation	
Confounder	**pGA**: % overlap with pre-gentrification areas (‘16)	Portland	Study map build; calculation	
Confounder	**GA**: % overlap with gentrification areas (‘16)	Portland	Study map build; calculation	
Confounder	**Auto**: *pwBG* % auto commuting (‘16)	5-year ACS via `tidycensus`	Calculation; *pwBG* computation	
Confounder	**Age**: *pwBG* median age (‘16)	5-year ACS via `tidycensus`	*pwBG* computation	
Confounder	**PD**: Population density (‘10) (people/mi^2)	Census via `censable`	Calculation	

Table 3.1: Variables for analysis of *electoral bikelash* in Portland, OR

[26]In all cases, “%” refers to *proportion*; these variables have range [0, 1]

Regression and results

I test several linear regression models that address the nuances in bike infrastructure type, quality, and timing, all of which are measured in Tables 3.2 and 3.3. The baseline model, *Model 1*, includes 2016 pre-gentrifying (pGA) and gentrifying area (GA), 2011-2016 pwBG-projected median age (Age) and percent auto commuting (Auto) (and an interaction variable between these two variables), and 2010 population density (PD)[27].

Models 2, *3*, and *4* add variables for iterations of Treatment 1 (T1), bike lane projects as a proportion of the street grid, to the baseline model. *Model 2* incorporates total painted projects and total protected projects by precinct across the election time period, as well as interactions between each of these with pGA, GA, and the interaction variable for Age and Auto. *Model 3* acknowledges that different types of protected lanes are more salient than their painted counterparts, and distinguishes between types 1 and 2 lanes (flex posts or flex posts and some curbs versus full curb or grade-separation). Finally, given more recent lanes may have a larger effect on one's perception of an incumbent's bike contribution, I used only projects completed in 2019 or 2020 for *Model 4*.

Table 3.2[28] displays these, excluding the numerous interaction variables included in *Models 2*, *3*, and *4*. Arbitrarily few of these have statistically significant[29] associated coefficients. None of the models show statistically significant regression coefficients for bike infrastructure of any kind, though, combined with interaction variables, these are negative.

Models 2, *3*, and *4* provide little evidence that electoral bikelash occurred in Portland in 2020. In *Model 2*, we find regression coefficient 0.116 for the aggregate protected T1, or proportion of protected bike lane overlap with a precinct's street grid across all years and quality types. This means that Wheeler is expected to have performed 11.6% worse from 2016 to 2020 in a precinct with *no* protected lanes than in one one with protected lanes

[27]Note that in the regression, I divided the units by 10^3, such that a one-point increase in the variable used in each model corresponds to a $1,000$-person increase in population density, so coefficients are viewable.

[28]All regression tables constructed using the `stargazer` (Hlavac, 2022) package

[29]I use the significance threshold of $\alpha = 0.05$ unless stated otherwise.

Table 3.2: *Model 1* and T1 *Models 2*, *3*, and *4*

	Dependent variable: Wheeler vote share change from 2016 to 2020			
	Model 1	Model 2	Model 3	Model 4
Prop. gentrifying area	-0.058^{***}	-0.062^{**}	-0.060^{**}	-0.060^{**}
	(0.020)	(0.024)	(0.024)	(0.024)
Prop. pre-gentrifying area	-0.059^{***}	-0.032^{**}	-0.033^{**}	-0.029^{**}
	(0.010)	(0.012)	(0.013)	(0.013)
Prop. auto commuting (pwBG)	-0.321	-0.734^{**}	-0.891^{**}	-0.897^{**}
	(0.325)	(0.326)	(0.343)	(0.388)
Median age (pwBG)	-0.008	-0.014^{**}	-0.016^{***}	-0.016^{**}
	(0.006)	(0.006)	(0.006)	(0.007)
Auto-age interaction var.	0.010	0.020^{**}	0.023^{**}	0.024^{**}
	(0.009)	(0.008)	(0.009)	(0.010)
Pop. density (10^{-3}/mi.^{2})	0.005^{***}	0.005^{***}	0.005^{***}	0.006^{***}
	(0.002)	(0.002)	(0.002)	(0.002)
Painted projects (as prop. of street grid)		-2.061 (1.329)	-1.815 (1.387)	-2.279 (1.422)
Protected projects (as prop. street grid)		**0.116** (1.898)		
Type 1 protected projects			**1.077**	
			(2.102)	
Type 2 protected projects			**$-$21.759**	
			(13.977)	
Type 1 protected projects (2019)				**1.436** (3.551)
Type 2 protected projects (2019)				**0.332** (6.082)
Type 1 protected projects (2020)				**$-$4.037** (15.192)
Type 2 protected projects (2020)				**492.211** (341.441)
Constant	0.206	0.425^{*}	0.511^{**}	0.512^{**}
	(0.218)	(0.215)	(0.223)	(0.253)
Observations	78	78	78	78
R^2	0.414	0.548	0.573	0.600
Adjusted R^2	0.365	0.447	0.443	0.440
Residual Std. Error	0.029 (df = 71)	0.027 (df = 63)	0.027 (df = 59)	0.027 (df = 55)
Note:				*p<0.1; **p<0.05; ***p<0.01

completely covering its street grid, *ceteris paribus*. This goes against *H1*, which suggests the latter would punish Wheeler more, but the coefficient is not significant. When separating this effect by quality as in *Model 3* or quality and year as in *Model 4*, I find statistically insignificant coefficients for bike lane treatments, including the painted lane variable.

Instead, incorporating more covariates actually strengthens the argument that confounders are the true determinants of incumbent performance; note the negative relationship between a precinct's gentrification risk and Ted Wheeler's change in vote share. Additionally, higher median age and auto commuting areas seem to have punished Wheeler more in 2020, even when examining the interaction variable (which mathematically dampens, but does not cancel out this effect; its magnitude is similar to that of the median age coefficient, so multiplying age by proportion auto commuting results in an even smaller effect). Population density, too, is statistically significantly associated with higher Wheeler vote change between the precincts.

An ANOVA test on *Models 2*, *3*, and *4* supports that *Model 1*, with only the original confounders, may be a better fit [see Appendix # for details]. Only the F-statistic for measuring *Model 2* against *Model 1* is significant at the 95% confidence level, though since I perform many comparisons, a more conservative p-value than 0.05 may be needed. Thus, I do not reject the null that any model including some form of T1 is better than *Model 1*.

Treatment 2 (T2), or proportion overlap with protected lane 15-minute walksheds, provides a better metric of voters' proximity to bike infrastructure at the expense of measuring its concentration. For the T2 set of models, I take the T1 set and replace only its protected infrastructure variables with their T2 walkshed counterparts. I use the original T1 variable for painted lanes given aforementioned computational issues: painted lanes are so common that their walksheds would cover most precincts completely. *Model 5* replaces the protected variable and interactions from *Model 2* with the T2 walkshed variable for all lanes; *Model 5* is a similar walkshed counterpart to *Model 3*, and *Model 7* is the same to *Model 4*. Table 3.3 displays these, again excluding most interaction variables.

Table 3.3: *Model 1* and T2 *Models 5*, *6*, and *7*

	Dependent variable: Wheeler vote share change from 2016 to 2020			
	Model 1	Model 2	Model 3	Model 4
Prop. gentrifying area	-0.058^{***}	-0.066^{***}	-0.075^{***}	-0.060^{***}
	(0.020)	(0.023)	(0.023)	(0.022)
Prop. pre-gentrifying area	-0.059^{***}	-0.048^{***}	-0.046^{***}	-0.044^{***}
	(0.010)	(0.013)	(0.013)	(0.012)
Prop. auto commuting (pwBG)	-0.321	-0.531	-0.971^{**}	-0.764^{**}
	(0.325)	(0.373)	(0.382)	(0.378)
Median age (pwBG)	-0.008	-0.011^{*}	-0.018^{***}	-0.015^{**}
	(0.006)	(0.006)	(0.006)	(0.007)
Auto-age interaction var.	0.010	0.015	0.026^{***}	0.023^{**}
	(0.009)	(0.009)	(0.010)	(0.010)
Pop. density (10^{-3}/mi.^{2})	0.005^{***}	0.005^{***}	0.005^{***}	0.007^{***}
	(0.002)	(0.002)	(0.002)	(0.002)
Painted projects (as prop. of street grid)		0.337 (0.218)	0.409^{*} (0.223)	0.436^{*} (0.224)
Protected projects (as prop. walkshed overlap)		**−0.012** (0.058)		
Type 1 protected projects			**0.058**	
			(0.082)	
Type 2 protected projects			**−0.218***	
			(0.115)	
Type 1 protected projects (2019)				**0.431**** (0.186)
Type 2 protected projects (2019)				**−0.288** (2.661)
Type 1 protected projects (2020)				**0.165** (0.241)
Type 2 protected projects (2020)				**−18.144***** (6.799)
Constant	0.206	0.325	0.579^{**}	0.426^{*}
	(0.218)	(0.254)	(0.254)	(0.253)
Observations	78	78	78	78
R^2	0.414	0.467	0.551	0.643
Adjusted R^2	0.365	0.378	0.442	0.509
Residual Std. Error	0.029 (df = 71)	0.029 (df = 66)	0.027 (df = 62)	0.026 (df = 56)
Note:				*p<0.1; **p<0.05; ***p<0.01

Models 5, *6*, and *7* provide contradictory evidence that electoral bikelash occurred in Portland in 2020. In *Model 5*, we find regression coefficient -0.012 for the aggregate protected T2, or proportion overlap with 15-minute walksheds for protected bike lanes across all years and quality types. This means that Wheeler is expected to have performed 1.6% better from 2016 to 2020 in a precinct with *no* protected lanes within a 15-minute walk than in one one with full 15-minute walking access to protected lanes, *ceteris paribus*. This agrees with *H1*, but the coefficient is not significant. When we separate this effect by quality as in *Model 6* or quality and year as in *Model 7*, we find some statistically significant coefficients for protected bike lane walksheds. For instance, in *Model 7*, Wheeler is expected to have performed 43.1% worse from 2016 to 2020 in a precinct with no type 1 (flex post plus or minus some curb) protected lanes constructed in 2019 within a 15-minute walk than in one one with full 15-minute walking access to type 1 protected lanes constructed in 2019. A result like this goes against H1; *Model 6* and *Model 7* suggest that bikelash occurs against type 2 (full curb or grade-separation) bike lanes but the opposite occurs for type 1 lanes.

Interestingly, an ANOVA test finds a statistically significant F-statistic at the 95% confidence level for *Model 7* measured against the baseline *Model 1*, suggesting it may be a better fit. This model finds a positive effect for the former and a negative for the latter, which suggests that some infrastructure may "age well" with voters. However, ANOVA tests on *Models 5* and *6* fail to reject the null hypothesis that *Model 1* is a better fit for the data.

In the end, that *Model 7* produced significant results tells us little in the end, as lower p-value threshold than 0.05 may be necessary. A very conservative approach is similar to a Bonferroni correction: dividing the original p-value threshold by the number of models. Neither the significant p-value for type 1 protected projects in 2019 nor that for type 2 projects in 2020 meet this low threshold of $p^* = 0.05/6 \approx 0.008$ [see Appendix # for results]. Instead, as in the T1 set, a main takeaway from *Models 5*, *6*, and *7* is that the original set of confounders – gentrification, age, driving – were negatively associated with Wheeler vote change as expected even when controlling for bike lane geography.

Discussion

Overall, I conclude that if any *electoral bikelash* occurred in Portland's 2020 mayoral primary, the geographical placement of infrastructure did not play a role. Treatments as constructed failed to produce convincing evidence that voting behavior in this election had a geographical relationship to bike infrastructure. Null results, however, are results, and these offer an interesting story. For one, the variables often associated with communities exhibiting bikelash – those living in gentrifying or risk-of-gentrifying areas, as well as older, driving areas – correlated negatively with change in Wheeler vote share in 2020, with high statistical significance across all models. This leads to an alternative hypothesis:

Hypothesis 1_A ($H1_A$): Bike lanes are unpopular among (a) older, car-commuting and (b) gentrifying neighborhoods; but they punish incumbents regardless of proximity to new lanes.

$H1_A$ suggests bikelashing communities' disapproval is toward a *citywide*, not hyper-local, bike agenda. It is reasonable to assume in the four years between elections, people gather some information about what is happening elsewhere in a city – through local news (perhaps bikelash reporting), traveling through other neighborhoods, and more. The take-away for pro-cycling politicians like Wheeler may be that incumbents need not worry about the *placement* of bike infrastructure, but rather bike lane *proliferation* and voters' perception of their wider agenda.

This idea requires further testing, and survey or experimental research may fill in the gaps. Polling about important issues in a local contest, which often focuses on standard issues like crime, could gauge whether these communities are truly aware of bike politics regardless of proximity.

Research decisions I made may impact the results. I avoided post-treatment variables in these models, but many other changes within and across precincts likely took place between elections, therefore these models far from capture the entire picture of change that may impact incumbent performance. Perhaps a different experimental design, such as looking at difference-in-differences changes between closer elections (a primary and a general election) may alleviate this. The variables I used present other problems; *pwBG* projection leaves perfect accuracy out of the question. The timing of these variables, too, isn't ideal (block population, for instance, is taken from 2010); even a new apartment building in a Census block can mean hundreds of new residents, skewing any statistic.

Portland itself presents further challenges. Ideally, its precincts would be smaller and greater in number (like Seattle or NYC). Yet its ample bike lane data and the timing of Wheeler's elections make this worth accepting. The question of generalizability remains; Portland is more wealthy, younger, less Black, and less Hispanic than the median U.S. city with over 300,000 residents [see Appendix #]. Future research into electoral bikelash may want to investigate cities with various demographic, cultural, and economic characteristics. Still, I believe this is a strong analysis that rose to the challenge of numerous data and conceptual hurdles; it is, to my knowledge, the first attempt to quantify the oft-discussed phenomenon of *bikelash* in an electoral context.

Ted Wheeler advanced to and eventually won the general election in 2020. Bike lanes have not exited the conversation, but they are still wrapped up in bikelash issues: gentrification and cars. For instance, in a gentrifying northeast Portland neighborhood, the city recently removed a bike lane after residents called out insufficient public engagement[30]. Soon-retiring Mayor Wheeler, however, is still moving forward with his cycling agenda, as indicated in a supportive speech presented to bike advocates in September 2023[31].

In the next chapter, we travel north to Seattle, WA, where similar bike infrastructure expansion in the 2010s allows us to test *Hypothesis 2* (H2) that lanes boost local turnout by increasing the visibility of local government in action.

[30]https://bikeportland.org/2023/12/18/crews-have-scrubbed-off-ne-33rd-ave-bike-lanes-382601

[31]https://bikeportland.org/2023/09/14/mayor-wheeler-comes-out-at-councils-biggest-advocate-for-cycling-379437

Chapter 4

Bike lane heuristics in Seattle

This chapter aims to quantify the relationship between the geographic distribution of new, protected bike lanes and local turnout via a government *visibility* mechanism; by this, bike lanes are *heuristics* reminding residents of local influence. Specifically, I test:

Hypothesis 2 (H2): Bike lanes amplify local government visibility, increasing the salience of municipal elections and thus increasing local turnout.

I examine change in voter turnout in two Seattle, WA, election cycles: parallel mayoral races in 2013 and 2017 [Figure 4.1], and City Council races in 2015 and 2019 [Figure 4.2]. I first cleaned precinct and election data, algorithmically merging changing precinct shapes across cycles. I then projected electoral factors including race, age, income, and election district fixed effects onto precincts. To test the impact of bike infrastructure, I categorized bike lanes by quality and year using several sources. My treatment uses precinct percent overlap with 15-minute "walksheds" for points imputed on bike lanes, or shapes representing areas that are a realistic 15-minute walk away from this infrastructure [Figure 4.3].

My set of analyses consists of linear regression models including confounders, treatment effects distinguishing by infrastructure implementation and quality, and interactions between these. To test whether whether voters assign bike lanes visibility increases to *local* government, I compare these to corresponding analyses of Seattle turnout change in two federal election cycles: 2014 and 2018 midterms, and 2016 and 2020 presidential elections.

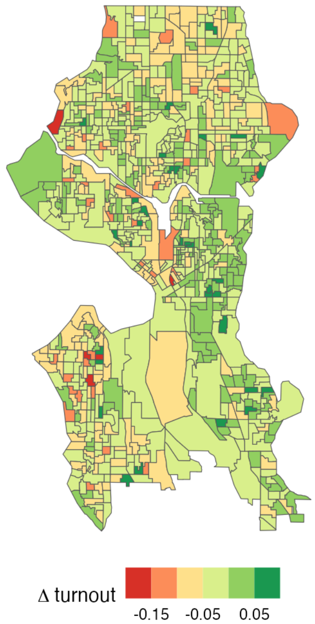


Figure 4.1: Seattle mayoral election turnout change, 2013-2017

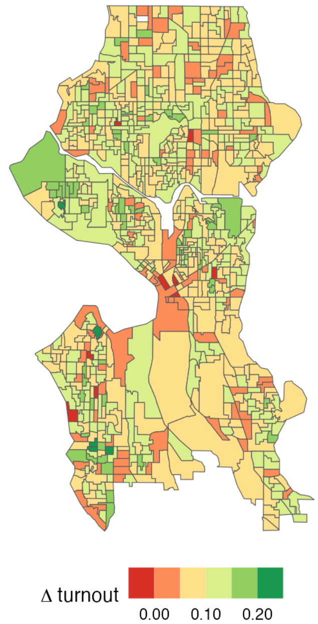


Figure 4.2: Seattle City Council election turnout change, 2015-2029

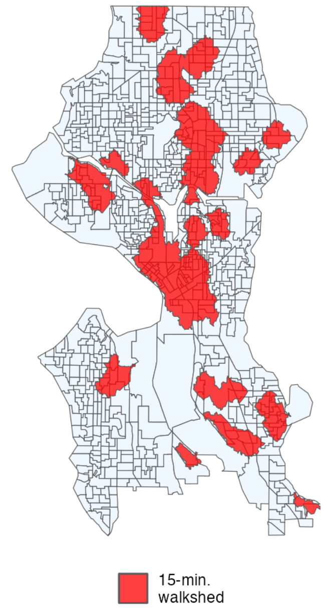


Figure 4.3: Protected bike lane aggregated 15-minute walksheds, all years

My results may agree with *H2*, indicating that precincts receiving bike lanes saw significantly higher turnout decline in federal elections than in local contests. However, these results are far from definitive; a different research design may be needed to isolate the effects of bike infrastructure placement and other neighborhood change.

I end by discussing the implications of these findings, including that the placement of protected bike lanes appears to correlate with turnout decline in any case. If my alternative hypothesis is the case, future research may be important in reassuring residents concerned with neighborhood change that bike lanes do not *cause* this dilution of democracy, and, if anything, they dampen its effect at the local level.

Why Seattle, Washington?

Seattle, Washington, saw a bike infrastructure boom in the 2010s that arose not only from bike-boosting politicians like former Mayor Mike McGinn[1], but also from grassroots activism and popular concern with rising traffic deaths in the city. *Seattle Bike Blog*[2] founder Tom Fucoloro (2023, 127) describes Seattle's transportation policy shift in the 2010s as being primed by "a decades-long struggle to push back against a cars-first mentality." First, sharrows (shared roads) became standard, then came painted lanes, and finally, beginning in 2013, protected bikeways emerged. The political culture shift was swift: when activists installed an illegal "guerilla" lane on Cherry Street in 2013 using cheap pylons and adhesives, City Traffic Engineer Dongho Chang literally thanked them in a statement[3].

The story of Seattle's bike lane boom illustrates a feedback loop between local activism and increased local government *visibility.* A city with responsive government like Seattle is ideal for analyzing whether citizens respond back – namely, by turning out in local elections.

From a research perspective, Seattle shares many of the features that made Portland ideal for analysis. The city built an impressive network of bike lanes in the 2010s, and it releases accurate shapefiles of the infrastructure, categorized by quality. The city does not release clean lanes implementation dates, but these exist elsewhere on the Internet, on expired planning pages, blogs, local news, and more.

Seattle also has many election precincts – up to 1,000 by the end of the decade. As a result, when I recycle the 15-minute walkshed treatment method used previously, treatment variables function much more like indicator variables, since it takes a smaller walkshed to encompass an entire Seattle precinct, on average. A large sample size and simpler treatment distribution make analysis in Seattle more precise and intuitive.

[1] `https://www.thebicyclestory.com/2014/11/mayor-mike-mcginn/`

[2] *Seattle Bike Blog* (`https://www.seattlebikeblog.com`) is arguably the most prominent source for bike infrastructure news in the city, and I used it extensively in verifying data

[3] `https://www.seattlebikeblog.com/2013/04/04/guerrilla-road-safety-group-politely-installs-illegal-bike-lane-protectors-on-cherry-street/`

Outcome: Voter turnout change

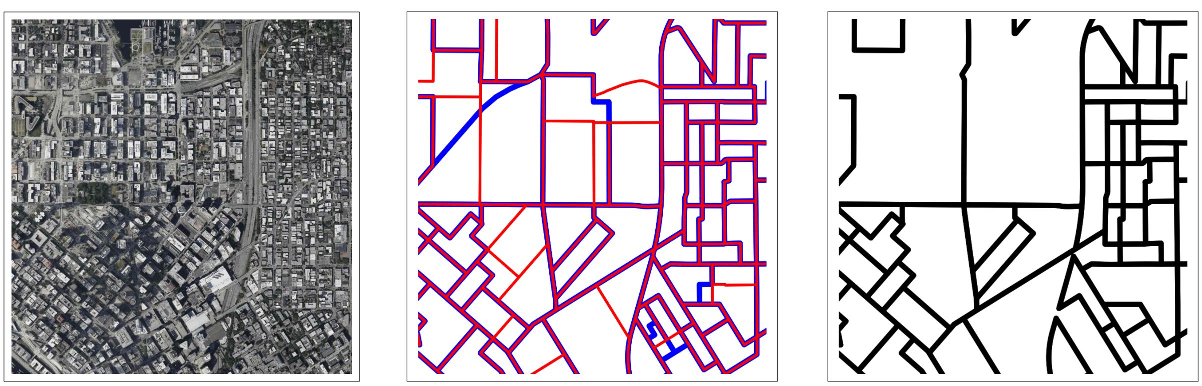
Seattle's local elections occur in November in odd years, and turnout fluctuated at around 50% of all registered voters in these from 2013 to 2019. The federal elections couched between these saw turnout around 80% of registered voters, with the exception of 2014 when midterm turnout tanked nationally[4]. I compare general election turnout across two odd-year election cycles, 2013 to 2017 and 2015 to 2019; the former corresponds to mayoral races, while the latter only had City Council candidates at the top of the ticket. This controls for turnout fluctuations based on the stakes of the local race as well as diverging demographic mobilization effects that may occur in different cycles.

Election data came from King County, Washington[5]. I calculate voter turnout as a fraction of registered voters. Using turnout as a percent of registered voters means the outcome variable is not only impacted by eligible voters' decision to turn out, but also by changes in who registers, how voter lists are managed, and more. However, the net number of registered voters increased smoothly across odd-year elections in the city, from about 410,000 in 2013 to 475,000 in 2019, in line with population growth. The registered voter roll for the average precinct, as I define them next, grew by about 13% between these years; the standard deviation of voter roll growth was about 18%. This suggests a somewhat stable change in voter registration rates across the city.

Inspecting some of the precincts that saw the highest increase in voter registration, I found these are usually large redevelopment areas whose populations increased. Controlling for variables pre-treatment that may correlate with a neighborhood's future redevelopment, as I describe later, may help alleviate these outliers. Overall, this outcome variable calculation is reasonable and common, but keeping in mind the role of registered voter counts as an input is important in selecting confounders and analysis.

[4] `https://www.electproject.org/national-1789-present`
[5] `https://kingcounty.gov/en/legacy/depts/elections/results`

Election precincts for Seattle came from the Washington Secretary of State[6]. These presented a second challenge, since many splits or small border changes occurred between every cycle in the decade. As a result, the number of precincts across the city grew from 942 to 1,000 precincts between 2012 and 2020. Since precincts are constructed from Census blocks, I identified each block with the precinct that encompassed it for each election. Then, I combined precincts that "swapped" Census blocks at any point; these came from the `censable`[7] package. I did this separately for odd-year local elections and even-year federal elections. For instance, if a given Census block was assigned to a hypothetical *Precinct 1* for all odd-year elections but switched to *Precinct 2* in the latest due to population shifts, I combined these over the entire set into a larger *Precinct 1*. The resulting two longitudinally-consistent, citywide datasets contained 833 precincts or combined precincts for odd-year elections and 815 for even-year elections. Figure 4.4 illustrates this process and what these combined precincts look like when focusing on the downtown area from 2013 to 2019.



(a) Downtown Seattle satellite image, from Apple Maps

(b) 2019 (red) precincts overlaid on 2013 (blue) precincts

(c) Data set with precincts merged across all elections

Figure 4.4: Precincts in the window above split or morphed between 2013 and 2019, and I merged these shapes - along with their features - based on a common outline across all years. Not all changes are visible in this example; in a few cases, precincts in 2013 and 2019 matched, while differences within these appeared in 2015 and 2017 sets.

[6] https://www.sos.wa.gov/elections/data-research/election-data-and-maps/reports-data-and-statistics/precinct-shapefiles

[7] https://christophertkenny.com/censable/

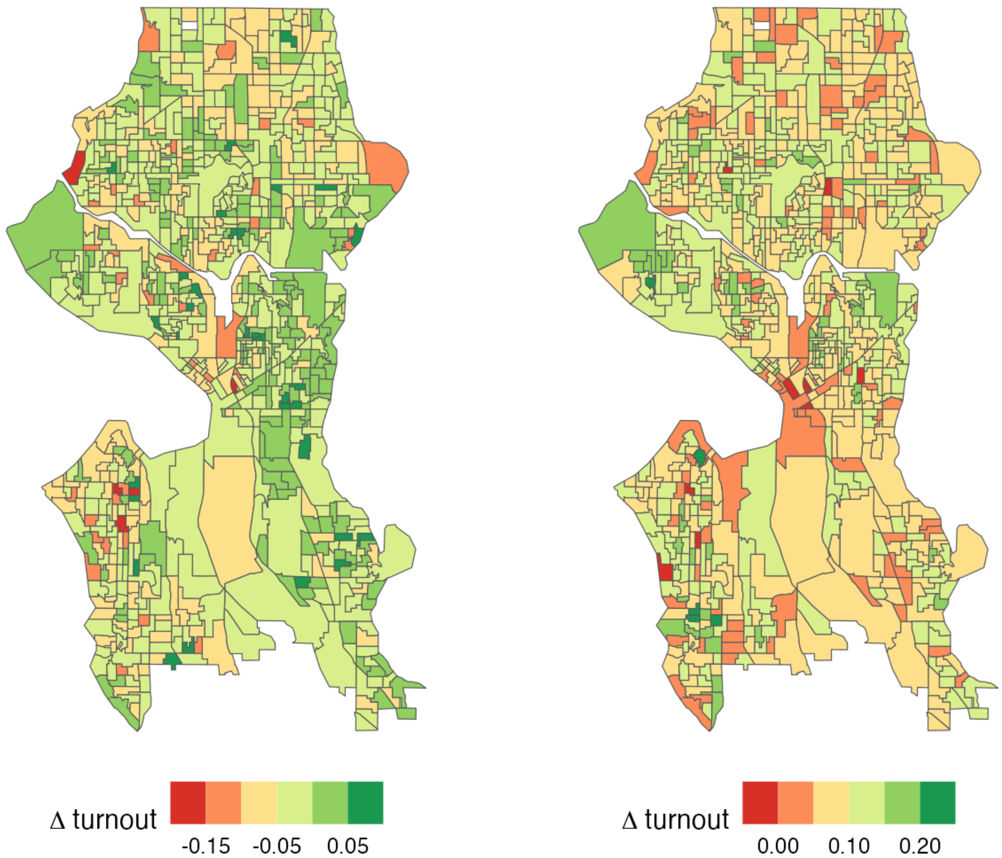
To validate these original algorithms, I inspected the odd-year set and merged precincts manually. This meant creating a *match-overlap* algorithm: I matched election precincts for a given cycle by ID and removed those with 99% overlap[8], which left many splits or precincts with small border changes – the *leftover* set. I then tediously hard-coded a column for joins across election years by inspecting side-by-side maps of leftover sets and the original precinct maps. Finally, I merged precincts until the match-overlap algorithm produced no leftovers.

The algorithmic and validation (semi-algorithmic) processes for odd-year cycles were similar, though the validation method resulted in slightly fewer combined precincts. I determined this is because my algorithm tested for block overlap with precincts by year, and in some cases, a freeway or trivial (empty) block switched precincts. There was no way to fix this in the algorithmic approach: for instance, I needed to maintain separate identification for trivial blocks in 2010 since redevelopment over the decade resulted in some of these gaining residents between cycles[9]. Though validation ended up performing better, this was marginal; I opted for the algorithmic approach for consistency and replicability.

It is worth noting that precincts which split across cycles – or those that I algorithmically merged for the final set of precincts – may have different features than others. In the odd-year set, for instance, there are 86 *merged* precincts, and 747 *original* precincts whose boundaries did not change (many of these are the same in the even-year set) [see Appendix # for visualization]. A two-sample t-test finds that in 2010, the merged set were younger, less affluent, and less white than the original set at the 95% statistical significance level; this suggests that precincts with splits changed more over the course of the decade as compared to wealthier, suburban-style neighborhoods. Whether redevelopment growth or gentrification drives these is unclear; these two concepts are intertwined. In any case, this result is not a significant issue, since I control for all three confounders in my analyses.

[8]This was a conservative threshold; some precincts included exhibited miniscule visible change.

[9]This included major roadways converted to developable land: `https://wsdot.wa.gov/construction-planning/major-projects/alaskan-way-viaduct-replacement-program`



(a) $C_{O,1}$: 2013-2017

(b) $C_{O,2}$: 2015-2019

Figure 4.5: Proportion turnout change for both odd-year election cycles; turnout generally declined from 2013-2017 and generally rose from 2015-2019

My precinct merge algorithm was not only a novel approach, it also maintained full precinct coverage of the city. In creating it, I avoided the the problematic but common choice of simply filtering out mismatched precincts. For the odd-year set, for i in 833 precincts or merged precincts [Figure 5.6], and cycle $= C_O \in \{2017, 2019\}$ corresponding to the mayoral and City Council elections, our outcome variable is defined[10]:

$$\frac{(\text{Counted})_{C_O,i}}{(\text{Registered})_{C_O,i}} - \frac{(\text{Counted})_{C_O-4,i}}{(\text{Registered})_{C_O-4,i}} = (\Delta\text{turnout})_{[C_O-4 \text{ to } C_O],i}$$

[10]I used the same formula, of course, for the even year elections, with i in 815 precincts or merged precincts and cycle $= C_E \in \{2018, 2020\}$

Treatment: Protected bike lanes

As in the previous chapter using bike infrastructure in Portland, I am primarily concerned with the geographical placement of *high-quality* lanes or cycleways around the city between each election cycle in analyzing the effect of bike infrastructure on voter turnout in Seattle. I make a few changes from the Portland analysis, however. First, I do not include painted lanes; this variable turned up insignificant in Portland, and as I explain, the Seattle data is much less organized at the outset.

Raw bike lane data comes from the City of Seattle[11], which releases shapefiles covering all bike routes throughout the city, categorized by infrastructure quality. However, implementation dates for these are often missing or inaccurate. I filtered these for protected bike lanes and buffered bike lanes, as the city describes each, since these covered the high-quality protected lane types I used in the previous chapter. From the previous chapter, I recycle my definitions of *quality* and *timing* of bike lanes. Protected lanes are either of *type 1* (flex post or flex post / curb combination) or *type 2* (full curb or grade-separation). I also again define the "introduction" of these lanes to communities as their implementation year.

Using Google Street View, I classified all high-quality protected lanes as either *type 1* or *type 2*. I also used historical Google Street View data to find years before and after the infrastructure was installed [Figures 5.7 and 4.7], and filtered out those installed before late 2013 or after 2020 as evidenced by Street View images. Then, I used historical City of Seattle construction pages or the Wayback Machine[12] for each of the lanes, which contained their year of completion. For years where this was difficult, I was able to find posts on the popular *Seattle Bike Blog*[13] or several other sources (news sites, for instance) documenting

[11] `https://data-seattlecitygis.opendata.arcgis.com/search?tags=bike`

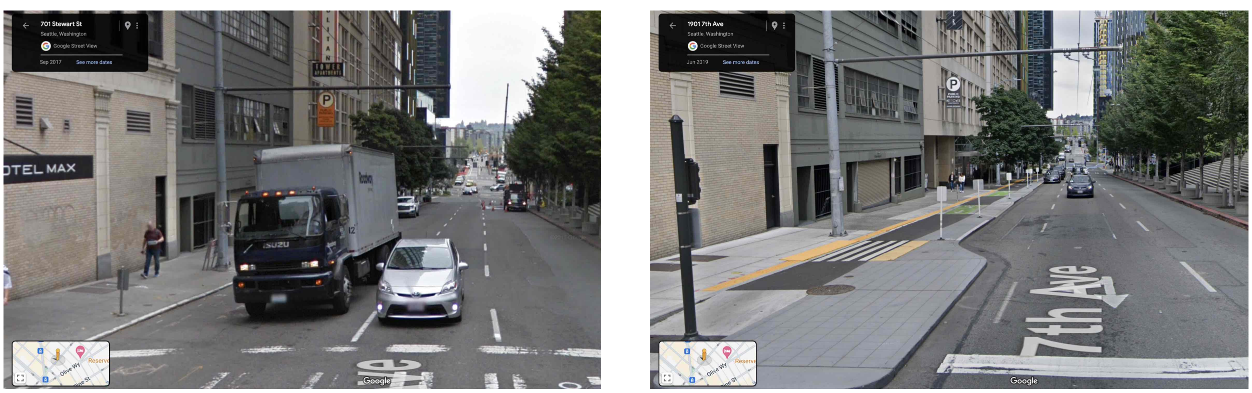
[12] Example: official city data cited 2010 as the installation date for Roosevelt Way NE lanes, but this planning page helped me correct it to 2015: `https://www.seattle.gov/Documents/Departments/SeattleBicycleAdvisoryBoard/presentations/SBABBikeMasterPlan6modraftmap.pdf`

[13] Example: official city data cited 2011 as the installation year for Mercer St. lanes, but this blog post helped me correct this to 2015: `https://www.seattlebikeblog.com/2015/05/05/mercer-street-king-of-the-stroad/`

the opening dates for remaining lanes. Sifting through planning documents or articles for each lane also allowed me to verify that most were constructed in a few months' time and all were complete before a November general election within its coded year due to seasonal construction timelines. I compiled these sources in an original data set [see Appendix # for more information]. At the end of this meticulous process, I had a shapefile data set consisting of 108 *type 1* lanes and 27 *type 2* projects constructed between 2013 and 2020 with accurate implementation years.

(a) Before: street parking, faded painted lanes (b) After: flex post protection, parking removal

Figure 4.6: Flex post (*type 1*) protected lanes installed on S Myrtle St. in 2019[14]



(a) Before: three traffic lanes, two-side parking (b) After: two traffic lanes, one-side parking

Figure 4.7: Grade-separated (*type 2*) protected lanes installed on 7$^{\text{th}}$ Ave. in 2017[15]

[14]https://www.seattle.gov/transportation/projects-and-programs/programs/maintenance-and-paving/current-paving-projects/swift-myrtle-othello

[15]https://www.seattle.gov/transportation/projects-and-programs/programs/bike-program/protected-bike-lanes/7th-ave-mobility-improvements

In this analysis, I use the walkshed treatment method described in the previous chapter (where I called it *T2*) to address who the *recipients* of bike lanes are. For each lane, I imputed points at 500 meter intervals (or centroids if the lanes were less than this threshold), and I used the `r5r` package to find polygons representing the area within a 15-minute walk from each point. This treatment is a more accurate capture of *proximity* to bike infrastructure projects, even though it doesn't capture *concentration* as accurately as the street grid proportion option (*T1*) did in the previous chapter. I only use the walkshed approach here primarily because utilizing street grids is less feasible for smaller precincts: an extra block of a protected bike lane could mean an additional 10% of the precinct's street grid is covered, even though the marginal impact of such a bike lane extension may be small in the minds of nearby residents.

Finally, I created variables for whether a lane's implementation falls into each of the four election cycles (two odd- and two even-year). Depending on the election cycle I use for each regression, I group either all that fall between the two election years together, or group them by *type*, and compute the proportion of overlap between the walkshed and precinct area [see Appendix # for grouping information].

For the odd-year set, for i in 833 precincts or merged precincts, cycle $= C_O \in \{2017, 2019\}$ to filter lanes implemented between mayoral or City Council elections[16], and $c \in \{all; \, type\ 1; \, type\ 2\}$, our treatment T is:

$$\frac{(\text{Bike lane walkshed overlap area})_{C_O,c,i}}{(\text{Total area})_{C_O,c,i}} = (\text{T})_{C_O,c,i}$$

Figure 4.7 shows lanes constructed between the 2013 and 2017 mayoral elections overlaid on precinct area. Next to this, Figure 4.8 shows the treatment variables, or overlap with protected lanes off both types, by precinct for this cycle.

[16] Again, I used the same formula for the even year elections, with i in 815 precincts or merged precincts and cycle $= C_E \in \{2018, 2020\}$

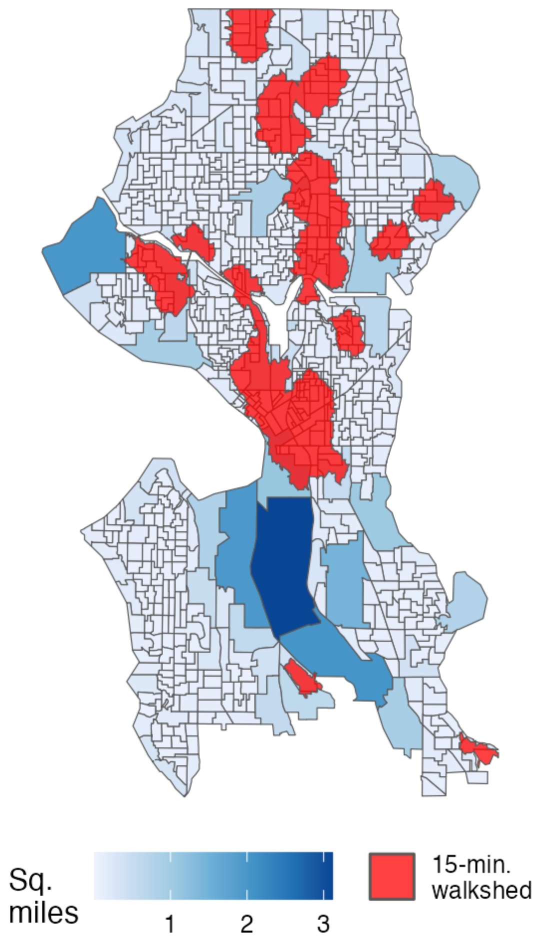


Figure 4.8: Protected bike lane aggregated 15-minute walksheds, 2013-2017, overlaid on precinct area

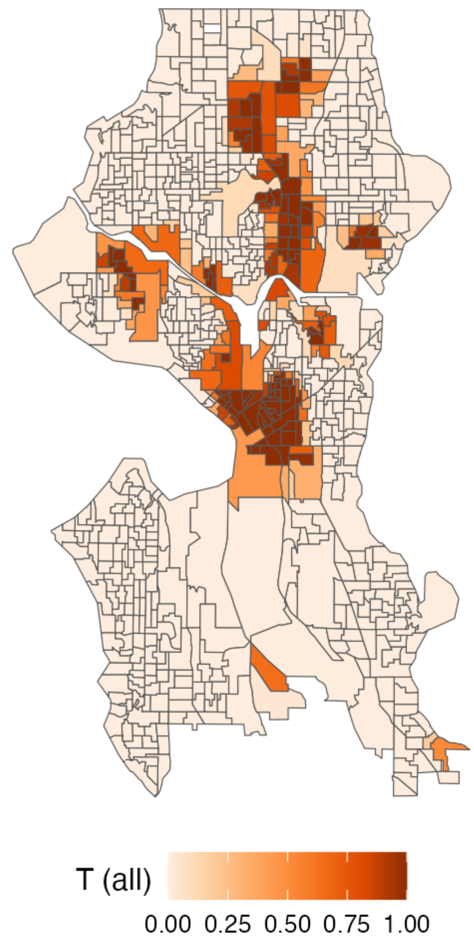


Figure 4.9: Protected bike lane aggregated 15-minute walksheds, 2013-2017, converted to proportion overlap with precinct

Confounders: Race, age, income, etc.

To isolate the effect of *H2* on bike lanes on turnout (via an increased local government *visibility* mechanism), I control for precursors of precinct change that may affect the turnout calculation: race, income, and age. These double as crucial controls for cyclical election effects; unique groups are affected differently by candidates and issues across cycles. I also included previous turnout and district fixed effects as confounders.

The first confounder is race. Using the `censable` package, I collected population counts by race for each Census block in 2010 and combined these at the precinct level. I found multicollinearity between percent white, percent Black, and percent Hispanic was too high as indicated by these variables' respective variance inflation factors (VIF) in early models, so I included only percent white by precinct as a confounder. The average precinct in my merged data set for odd-year elections is about 70% white, although some are as low as 6% white. The most diverse precincts in the city are concentrated in the south and southeast neighborhoods; Seattle's 2nd City Council district encompassing these was only about 30% white in 2015[17]. Neighborhoods that were more white in 2010 were less at risk of rapid gentrification, given these were many of the lower-density, suburban-style neighborhoods.

I included median income as a second confounder, as lower income neighborhoods are by definition at higher risk of gentrification. I used the *population-weighted block group (pwBG) projection* [Figure 3.11] process outlined in the previous chapter to collect these: I gathered 5-year American Community Survey (ACS) median age estimates for Seattle block groups in 2013, "projected" these medians onto Census blocks (with a homogeneity assumption), weighted these by 2010 block population, and took their average at the precinct level. I divide the median income statistic by 10^3 for interpretability in the regression table.

I also gathered median age because younger areas – universities, downtown, and more

[17]Council districts were first drawn in 2015 when the city switched from at-large voting; I do not use post-treatment data from this year, however. For demographics, see: `https://projects.seattletimes.com/2015/city-council-districts/`

– generally see more turnover in a city like Seattle, and these areas turn out at lower baseline rates regardless of race. This variable was available at the block level in the 2010 Census, although calculating this at the precinct level still meant taking an average of medians.[18]

For electoral confounders, I first included the original cycle's turnout. Past turnout is a strong predictor of future turnout, and mathematically, precincts with lower initial turnout have higher turnout ceilings for the next election (and vice versa). H2 implies that we should see larger changes in precincts with lower initial turnout, since these are precincts where local government is not as relatively salient to its voters before the introduction of infrastructure, and using baseline turnout as a confounder helps isolate this impact.

Finally, for the non-mayoral election years, where the top of the ticket likely varied across the city in terms of candidate quality and campaign efforts, I control for district fixed effects using variables for a precinct's overlap with each of the seven City Council districts. Most are zeros or ones, but a few precinct splits occurred across district borders, so the variable is technically not an indicator. I did the same for the federal midterm election merged precinct set to account for the two congressional districts that bisect Seattle. I constructed these fixed effects variables using the original precinct data.

Table 4.1 summarizes the variables I collected, their sources, an overview of the processes by which I cleaned and formatted these, and their histograms. All correspond to the first local election cycle, or the mayoral races; the Dist histogram uses the indicator variable for council district 1. The outcome variable histogram is normally distributed centered near -5% turnout change from 2013 to 2017. Nearly all precincts have near-zero or near-complete exposure to the treatment, though a decent fraction have full walkshed overlap; this holds for subsets by type not shown below. The Auto and Income variables appear normally distributed as is ideal for linear regression, while the White variable is left skewed.

[18]This is both the best option and reasonable on its own; Census blocks have similar populations, and blocks grouped into the same precinct often have similar demographic characteristics due to urban segregation.

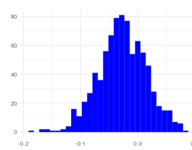
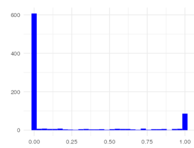
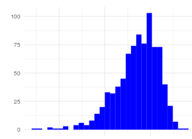
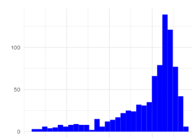
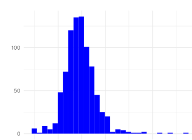
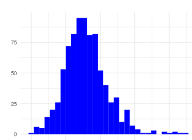
	Variable[19]	Sources	Process	Histogram
Outcome	**Δturnout**: Turnout change by cycle	King County; Census GIS	Merging precincts across cycles, calculation	
Treatment	**T**: % bike lane 15-min. walkshed overlap	Seattle; local news; walksheds via `r5r`	Manual imputation; walkshed build; calculation	
Confounder	**bTO**: Baseline (election year one) turnout proportion	King County	Calculation	
Confounder	**White**: % population white only ('10)	Census via `censable`	Calculation	
Confounder	**Age**: Median age, weighted[20] ('10)	Census via `tidycensus`	Weighted calculation	
Confounder	**Income**: *pwBG* median income ('13)	5-year ACS via `tidycensus`	*pwBG* computation	
Confounder	**Dist**: District fixed effects by cycle	King County	Calculation	

Table 4.1: Variables for analysis of *bike lane visibility turnout* in Seattle, WA

[19] In all cases, "%" refers to *proportion*; these variables have range [0, 1]; in my regression, however, I use percentages with range [0, 100]

[20] Note that this variable was not computed via *pwBG* projection; these were available at the block level for 2010 and weighted by population up to the precinct level

Regression and results

For each election of the two odd-year local election cycles, I test regressions that address nuance in bike infrastructure quality on the outcome variable (Δturnout). The baseline models, *Model 1a* for 2013-2017 and *Model 2a* for 2015-2019, include baseline election year turnout (bTO) 2010 percent white (White), 2010 population-weighted median age (Age), 2011-2016 pwBG-pwBG-projected median income (Income). The latter city council cycle also includes district level fixed effects (Dist). Importantly, across all regressions, I multiply Δturnout, bTO, and White by 100, such that these are in percentage point units with theoretical range $\{-100, 100\}$.

Models 1b / 2b and *Models 1c / 2c* add variables for our treatment to *Models 1a / 2a*, respectively. *Models 1b / 2b* include aggregated 15-minute walksheds for all infrastructure constructed between election years. *Models 1c / 2c* distinguish these walksheds by type (flex posts or flex posts and some curbs versus full curb or grade-separation).

The results for the *Model 1* set for 2013-2017 and the *Model 2* set for 2015-2019 are displayed in Tables 4.2 and 4.3, respectively.

The confounders are revealing in both sets of elections. All six models find statistically significant[21], positive associations between 2010 median income with turnout change, and a statistically significant, negative relationship between baseline turnout and change in turnout. The former is surprising, suggesting, *ceteris paribus*, areas that were more affluent in 2010 saw turnout rise in local elections over the course of the decade. The latter is unsurprising, since precincts with higher turnout to begin with have a lower ceiling for turnout improvement. Other confounding variable relationships are interesting; for instance, some City Council districts, like District 2, saw significant increases in turnout across the second set of elections, along with precincts that were more white in 2010.

[21] I use the significance threshold of $\alpha = 0.05$ throughout.

Table 4.2: *Models 1a*, *1b*, and *1c* for mayoral cycle 2013-2017

	Dependent variable: Difference in % turnout, 2013-2017		
	Model 1a	Model 1b	Model 1c
2013 % turnout	-0.146^{***}	-0.152^{***}	-0.156^{***}
	(0.024)	(0.025)	(0.025)
% white, 2010	-0.010	-0.008	-0.007
	(0.010)	(0.011)	(0.010)
Median income (pwBG) $\times 10^{-3}$	0.051^{***}	0.050^{***}	0.050^{***}
	(0.006)	(0.006)	(0.006)
Median age (pop.-weighted)	-0.018	-0.019	-0.016
	(0.026)	(0.026)	(0.026)
Protected lanes (walkshed overlap divided by area)		**−0.393** (0.440)	
Type 1 protected lanes			**−0.377**
			(0.458)
Type 2 protected lanes			**−1.012**
			(0.977)
Constant	2.641^{***}	2.968^{***}	3.063^{***}
	(0.951)	(1.019)	(1.015)
Observations	833	833	833
R^2	0.108	0.109	0.110
Adjusted R^2	0.104	0.103	0.104
Residual Std. Error	3.991 (df = 828)	3.991 (df = 827)	3.990 (df = 826)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4.3: *Models 2a*, *2b*, and *2c* for City Council cycle 2015-2019

	Dependent variable: Difference in % turnout, 2015-2019		
	Model 1a	Model 1b	Model 1c
2015 % turnout	-0.151^{***}	-0.166^{***}	-0.169^{***}
	(0.023)	(0.024)	(0.023)
% white, 2010	0.094^{***}	0.099^{***}	0.092^{***}
	(0.013)	(0.014)	(0.014)
Median income (pwBG) $\times 10^{-3}$	0.037^{***}	0.035^{***}	0.033^{***}
	(0.006)	(0.006)	(0.006)
Median age (pop.-weighted)	-0.058^{**}	-0.056^{**}	-0.036
	(0.026)	(0.026)	(0.026)
Council district 1	0.669	0.241	-0.325
	(0.485)	(0.504)	(0.525)
Council district 2	3.050^{***}	3.070^{***}	2.065^{**}
	(0.778)	(0.775)	(0.821)
Council district 3	1.599^{***}	1.653^{***}	1.076^{*}
	(0.547)	(0.544)	(0.563)
Council district 4	-0.609	-0.648	-1.262^{**}
	(0.526)	(0.524)	(0.548)
Council district 5	0.159	-0.171	-0.840
	(0.520)	(0.529)	(0.559)
Council district 6	0.479	0.170	-0.296
	(0.497)	(0.505)	(0.519)
Protected lanes (walkshed overlap divided by area)		$\mathbf{-1.289}^{***}$ (0.436)	
Type 1 protected lanes			$\mathbf{-0.587}$
			(0.472)
Type 2 protected lanes			$\mathbf{-3.404}^{***}$
			(0.945)
Constant	8.623^{***}	9.537^{***}	10.165^{***}
	(1.010)	(1.051)	(1.058)
Observations	833	833	833
R^2	0.144	0.153	0.165
Adjusted R^2	0.133	0.142	0.153
Residual Std. Error	3.739 (df = 822)	3.721 (df = 821)	3.695 (df = 820)
Note:			*p<0.1; **p<0.05; ***p<0.01

More important for my analysis, however, are the protected cycling infrastructure variables. These coefficients – both those for all infrastructure and those split bys type – are insignificant for the 2013 to 2017 mayoral cycle. Interestingly, however, for the lower-stakes City Council cycle from 2015 to 2019, this is not the case. In *Model 2b*, we find a statistically significant coefficient for the aggregated treatment variable (all protected infrastructure walkshed overlap) of -1.29. This suggests that, all else held equal, a precinct with complete overlap with new protected infrastructure walksheds averaged a 1.29 percentage point decrease in turnout in the next election, compared to a precinct with no walkshed overlap. *Model 2c* shows that the coefficient for type 2 protected lane project walksheds is statistically significant, while that for type 1 infrastructure is not, and the magnitude of its negative association with turnout change is even greater. All coefficients for our treatment variables, significant or not, are negative. This contradicts H2, which says that proximity to new bike infrastructure would increase local turnout.

Next, I run the same analyses on federal elections to examine whether association between bike infrastructure and turnout is truly *local*. Table 4.4 summarizes the baseline models for the federal midterm cycle from 2014 to 2018, while Table 4.5 shows these for the higher-turnout presidential cycle from 2016 to 2020 (*Models 3a* and *4a*, respectively) in addition to corresponding models containing covariates and walkshed treatments, either aggregated (*Models 3b* and *4b*) or separated by type 1 versus type 2 protected infrastructure type (*Models 3c* and *4c*) between respective election years. The *Model 3* set also includes a fixed effects indicator for Washington's 7^{th} congressional district in the 2010s since the city is split by two districts.

Again, the relationship between our confounders and the outcome variable, turnout change between cycles, is unsurprising. Older precincts in 2010 saw relatively lower turnout change between the 2014 and 2018 midterms, the latter of which some described as a "blue wave" midterm of young voters. Previous election cycle turnout is strongly predictive of turnout change, likely in part because there's more room to grow for low-turnout precincts.

Table 4.4: *Models 3a*, *3b*, and *3c* for midterm cycle 2014-2018

	Dependent variable: Difference in % turnout, 2014-2018		
	Model 3a	Model 3b	Model 3c
2014 % turnout	-0.391^{***}	-0.399^{***}	-0.404^{***}
	(0.021)	(0.021)	(0.022)
% white, 2010	0.153^{***}	0.156^{***}	0.157^{***}
	(0.013)	(0.013)	(0.013)
Median income (pwBG) $\times 10^{-3}$	0.007	0.006	0.005
	(0.005)	(0.005)	(0.005)
Median age (pop.-weighted)	-0.210^{***}	-0.210^{***}	-0.202^{***}
	(0.021)	(0.021)	(0.022)
Congressional district 7	-2.224^{***}	-2.243^{***}	-2.242^{***}
	(0.529)	(0.527)	(0.528)
Protected lanes (walkshed overlap divided by area)		$\mathbf{-0.814}^{**}$ (0.374)	
Type 1 protected lanes			$\mathbf{-0.087}$
			(0.407)
Type 2 protected lanes			$\mathbf{-1.608}^{**}$
			(0.766)
Constant	43.734^{***}	44.237^{***}	44.115^{***}
	(0.884)	(0.912)	(0.908)
Observations	815	815	815
R^2	0.487	0.490	0.491
Adjusted R^2	0.484	0.487	0.486
Residual Std. Error	3.361 (df = 809)	3.353 (df = 808)	3.354 (df = 807)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4.5: *Models 4a, 4b, and 4c for presidential cycle 2016-2018*

	Dependent variable: Difference in % turnout, 2016-2020		
	Model 4a	Model 4b	Model 4c
2016 % turnout	-0.191^{***}	-0.201^{***}	-0.200^{***}
	(0.021)	(0.021)	(0.021)
% white, 2010	-0.024^{***}	-0.023^{***}	-0.023^{***}
	(0.007)	(0.007)	(0.007)
Median income (pwBG) $\times 10^{-3}$	0.014^{***}	0.012^{***}	0.012^{***}
	(0.003)	(0.003)	(0.003)
Median age (pop.-weighted)	-0.012	-0.014	-0.015
	(0.014)	(0.014)	(0.014)
Protected lanes (walkshed overlap divided by area)		$\mathbf{-1.133}^{***}$ (0.276)	
Type 1 protected lanes			$\mathbf{-1.242}^{***}$
			(0.313)
Type 2 protected lanes			**0.455**
			(0.656)
Constant	21.313^{***}	22.413^{***}	22.418^{***}
	(1.421)	(1.433)	(1.433)
Observations	815	815	815
R^2	0.263	0.278	0.279
Adjusted R^2	0.259	0.274	0.273
Residual Std. Error	2.444 (df = 810)	2.420 (df = 809)	2.421 (df = 808)
Note:			*p<0.1; **p<0.05; ***p<0.01

Interestingly, in federal elections, I find that new high quality bike infrastructure walksheds may be even *more* predictive of turnout change between election cycles. Between 2014 and 2018, a precinct having complete overlap with new protected infrastructure walksheds was associated with a statistically significant average decline in turnout by 0.8 percentage points compared to a precinct with no overlap. Between the presidential elections of 2016 and 2020, this significant association was a decline in turnout of about 1.1 percentage points. Separating walksheds by protected infrastructure type indicates this association was only significant for type 2 infrastructure walkshed overlap between midterms and type 1 infrastructure between presidential races.

Across all except for the *Model 1* set, an ANOVA test found that at least one of the models with the treatment variables was a statistically significant improvement on the baseline model [see Appendix # for results]. This suggests that proximity to high-quality protected lanes are informative predictors of both local *and* federal election turnout, though in a direction I did not anticipate in developing H2.

<u>FOR CHRIS:</u> Any ideas on how to fill this whitespace (if necessary)??

Discussion

Overall, the results present a perplexing picture of the relationship between proximity to new bike infrastructure and change in turnout, but the disparity between local and federal turnout effects may align with a rephrasing of Hypothesis 2, which originally said proximity to new, protected bike lanes would increase turnout between local election cycles, in contrast to a smaller or absent effect in federal cycles. In the higher-turnout, federal election cycles, proximity to new bike infrastructure appeared to have a stronger, more negative association with turnout change between election cycles when compared to lower-turnout local cycles. An alternative hypothesis may be:

Hypothesis 2_A ($H2_A$): Bike lanes amplify local government visibility, increasing the salience of municipal elections and thus dampening changing neighborhoods' local turnout decline.

One possible interpretation of this evidence supports the idea that proximity to bike lanes indeed has a positive, *local* effect on turnout. Generally, I found that treated precincts saw larger relative declines in turnout for federal elections when compared to untreated precincts, whereas in local elections, this was less of the case. The treatment of bike lanes, then, could be interpreted as the *reason* treated precincts saw less of a "turnout hit" in local elections than in federal elections. In other words, features of a precinct I could not measure – post-treatment features touching on demographic change – may have caused its relatively higher overall turnout decline; bike lanes and their contribution to the *visibility* of local government may have been what dampened this turnout decline in local elections.

To investigate this, I gathered 2020 Census block population data, merged this by precinct, and calculated 2013 precinct population change over the decade. I found a statistically significant positive association between precinct population growth and overlap with infrastructure walksheds in the decade [see Appendix # for a visualization].

In short, treated precincts grew faster; the confounders I included may not have served as ideal precursors to post-treatment change. This or other changes associated with my treatment variable may be causing turnout to fall more, and bike lanes may be working against this in local elections as residents notice neighborhood changes and turn out.

There is a possibility, on the other hand, that bike lanes reduce turnout in federal elections more than local ones by some other mechanism – perhaps those who generally vote in federal elections and not local ones are associate lanes with higher levels of government, for instance. But I believe an explanation like this is unlikely. Public engagement on these projects is led by local governments on social media, through mailers, and in-person; local candidates, unlike those at the federal level, campaign on bike lane issues.

$H2_A$ requires further testing, and an experimental design may be ideal to isolate neighborhood changes occurring at the same time bike lanes are installed. Identifying demographically-similar precinct pairs that saw little or no growth, differing only in one of each pair receiving a bike lane, may be one option.

Finally, this analysis also provides subtle nuance to Hypothesis 1. Seattle was the fastest growing U.S. city with a population over 300,000 in the 2010s [see Appendix #], adding over 150,000 new residents in this period; its growth rate eclipsed those of Austin, Fort Worth, and Denver at the time. *Change* is likely on the minds of its residents, particularly those more likely to engage in *bikelash.* That bike lanes, population growth, and turnout decline are all related, regardless of the causal direction if any exists, helps explain why residents may resist bike infrastructure improvements in the first place.

Seattle's growth has since slowed while its bike expansion continues in full force. But more research on the relationship between bike lanes and turnout is important, particularly when residents may be justified in pointing to a negative association between the two at several election levels.

In the next chapter, we cross the country to Manhattan, New York City, where a vast network of bike share stations allows us to test *Hypothesis 3* that bike infrastructure improves democracy by making the act of voting itself easier.

Chapter 5

Biking to the polls in Manhattan

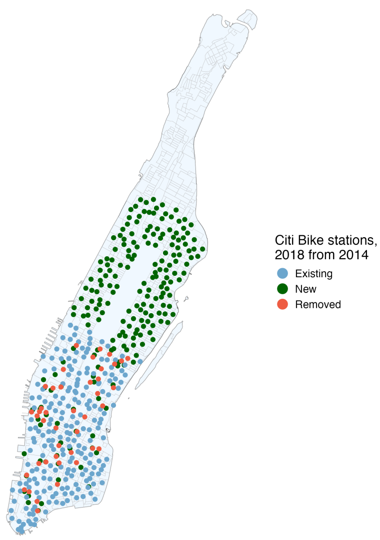


Figure 5.1: *A1* looooooooooooooks atttttttttttttt someeeeeeeeeeething

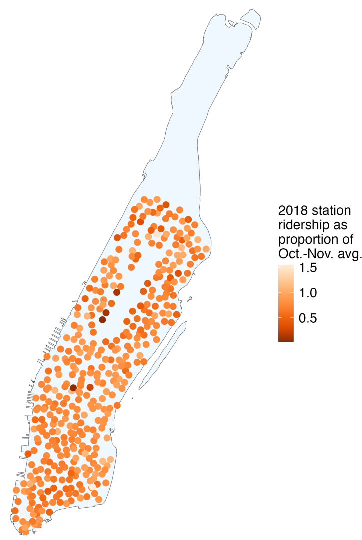


Figure 5.2: *A1* looooooooooooooks atttttttttttttt someeeeeeeeeeething

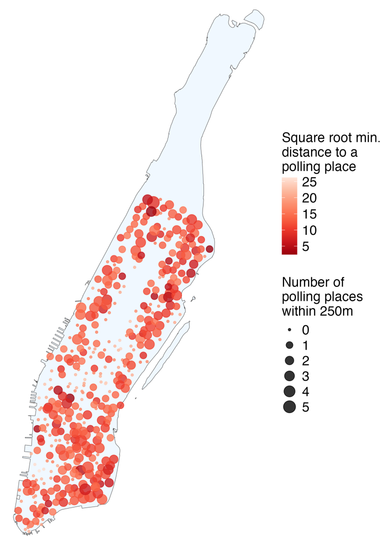


Figure 5.3: *A3* looks at the relationship between Citi Bike station proximity to polling places in 2018

Why Manhattan, New York City?

Citi Bike station data and analyses

Citi Bike data for New York City is extensive, stretching back as far as 2013, and files are managed by system operator Lyft Bikes and Scooters, LLC[1]. Publicly available annual or monthly files contain *every single Citi Bike ride*: from the moment a rider "unlocks" their bike to the moment they dock it. I refer these as the ride "start" and "stop."

Citi Bike data is a trove of information on movement in New York City. A single row, which represents a Citi Bike ride, includes not only exact start and stop date and timestamps, starting and stopping stations, and coordinates for each station, but also sparse demographic information on the rider. For many rows, date of birth and gender are available. This data is available under Lyft's commitment[2] to "supporting bicycling as an alternative transportation option," but that this data is released in raw, non-differentially private form is surprising. Presumably, it would be possible to identify an individual by the available information and their commuting patterns. Lyft may want to consider adding noise in the future to protect privacy. That said, the data has incredible research potential.

In examining H3, I am concerned with three aspects of the Citi Bike data; correspondingly, I performed three analyses utilizing it in different forms. For all analyses, however, I use a subset of Citi Bike data centered on Election Day, the first Tuesday after the first Monday of the month, in the years from 2014 (Nov. 4$^{\text{th}}$) to 2018 (Nov. 6$^{\text{th}}$). I filtered the data for rides up to two weeks before Election Week to rides up to two weeks after it. I also filtered out weekends, since leisure ridership is higher on these days and Election Day is mid-week. In summary, the data I used included five weeks of weekday rides, or 25 days, for each year from 2014 to 2018. Though my five weeks do not span all of October and November, I refer to this as "Oct.-Nov." ridership data throughout. Since I examine Citi Bike Election Day ridership, which I hypothesize may be relating to voting, I filtered all rides *ending* between 5:30 AM and 9:00 PM (30 minutes before polls open to poll close).

[1] `https://citibikenyc.com/system-data`
[2] `https://ride.citibikenyc.com/data-sharing-policy`

Additionally, in my third analysis, I am concerned with rides *to* the polls on Election Day; I group all data by date, year, and trip *end* station (where the bike was docked to end a trip), and count the total number of rides[3]. This gives me my final, overarching subset of Citi Bike ridership data: the total count of rides between 5:30 AM to 9:00 PM by station, date, and year, for all weekdays in five weeks centered on Election Week, from 2014 to 2018.[4]

I next prepared the Citi Bike data subset for each of three analyses exploring unique ways in which an expansive bike share system may relate to voting: new station access increasing turnout (A1), system use above weekday-weather modelled expectation on Election Day (A2), and higher station use on Election Day when located closer to a polling place (A3). I begin by preparing the Citi Bike data subset for each.

First, in *Analysis 1* (A1), I am interested in exploring whether there is an association between turnout change and new Citi Bike stations in a precinct; H3 suggests that this association would be positive, or Citi Bike stations would increase turnout as getting to the polls via bicycle is easier. Since I use the 2014 and 2018 elections in my analysis, I defined my treatment (T_{A1}) as *whether a precinct had a new station in 2018 and no station in 2014.* This gives a simply metric of whether a precinct's voters had increased access to Citi Bikes. Note that a polling place is not always located *in* a precinct, but my grouping of the subsetted Citi Bike data here includes all existing stations, so T_{A1} still measures geographical access to a bike – whether one uses it to leave *for* or arrive home *from* the polls. To prepare Citi Bike data for computing T_{A1}, I simply created two data sets of points: coordinates for all 2018 stations and the same for all 2014 stations. The regression using T_{A1} as a treatment includes turnout change from 2014 to 2018 as an outcome variable with demographic and electoral fixed effects confounding variables.

Second, in *Analysis 2* (A2), I am interested in exploring whether there is an association

[3]In Analyses 1 and 2, I explain, this grouping by trip end is less important; I use station *presence* (where ridership count is unimportant) or *direction-independent volume* on Election Day (where I am not concerned with whether one is arriving or departing from a station, only whether total usage is higher), respectively.

[4]Since my analysis applies to Manhattan, New York City, this data is also filtered to stop stations within New York County.

between expected system use (controlling for weather and day of week) and actual system use on Election Day; while unrelated directly to H3, this helps answer the *extent* to which bikes are used for transportation in traveling to the polls, and thus helps explain the results of A1. For this analysis, I first calculated *average* stops by station for each year (in the Oct.-Nov.) period. I group by year because system ridership increased overall annually as the network expanded; taking averages by year accounts for this annual fixed effect. Next, for each station and day, I compute the proportion of total stops to Oct.-Nov. annual average stops; this gives a relative metric of that station's use each day. This metric, *the proportion of daily station stops to annual Oct.-Nov. average* (P_S), is my outcome variable in a regression I run across all days, stations, and their proportions, using weather conditions and weekday as predictors. From this model, I calculate what the *expected* proportion of Election Day rides in 2014 and 2018 should be, and compare the actual proportion to this.

Third, in *Analysis 3* (A3), I am interested in exploring whether there is a relationship between station *proportion of Election Day station stops to Oct.-Nov. average* (P_S) and proximity to polling places in 2018; H3 suggests that there will be, since it rests on the idea that transportation, including walking, is a barrier to voting, and station use may be lower on Election Day if it is less convenient to access a polling place by docking at it. I use two treatments to calculate distance: square root of distance from a station to the nearest polling place ($T1_{A3}$), and the count of polling places within a 250m radius ($T2_{A3}$). I chose the 250m cutoff, as this seems reasonable; polling places within this are very convenient to access, about a 2-minute walk from a station, meaning these stations may be Election Day "hotspots." The only Citi Bike data preparation this section required was creating the 250m buffers around each station for $T2_{A3}$.

These three analyses – which respectively address new station access and turnout, Election Day system use compared to modeled volume, and Election Day station use and proximity to polling places – are summarized in Table 5.1 below. The following sections describe additional data required for each and present their results.

FOR CHRIS: FULL PAGE WITH CITI BIKE MAP?

Description	Sources	Data preparation[5]
Analysis 1 (*A1*): Linear regression, comparing merged precinct turnout change between 2014 and 2018 elections against the presence of new bike share stations (treatment) and demographic confounders	State of NY; VEST; Census via `censable`; 5-year ACS via RDH and `tidycensus`;	Precinct merges across 2014 and 2018; Confounding variable collection (*pwBG*, etc.); Treatment calculation (*new* station presence with no *existing* stations)
Analysis 2 (*A2*): Investigating Election Day turnout as proportion of Oct.-Nov. mean by station and year; comparing value to predicted proportion via weather/weekday annual model	NWS Central Park historical data via Kaggle	Calculating average station stops, Oct.-Nov. by year; Calculating proportion daily stops to average; Regressing proportion on weather and weekday; Calculating predicted Election Day proportion
Analysis 3 (*A3*): Linear regression, comparing 2018 Election Day turnout as proportion of Oct.-Nov. mean by station in against two treatment options: square root distance to nearest polling place (*T1*) and count of polling places within a 250m radius (*T2*)	NYC Open Data; city addresses via `tidygeocoder`; Google Earth	Imputing missing polling place coordinates via OSM data or manually; Identifying nearest polling place for each station and computing distance; Computing 250m buffers for each station and calculating number of polling places within (*T2*)

Table 5.1: Analyses 1, 2, and 3, applying Citi Bike station data to various aspects of 2014 and 2018 elections in Manhattan

[5]Not including initial preparation of Citi Bike station data

A1: Turnout effect of new Citi Bike stations

In *Analysis 1* (A1), I begin by testing the impact of new Citi Bike stations within a precinct on its voting turnout change from 2014 to 2018. H3 expects this relationship to be positive; that is, a new station makes it easier to get to the polls, thereby making voting less costly and increasing turnout.

Data required for this analysis, aside from Citi Bike station data for 2014 and 2018, includes *election precinct shapefiles and returns for these years*. For 2014, I gathered precincts from cartographer Nathaniel Kelso[6] and results from OpenElections[7]. For 2018, both precincts and results came from the University of Florida Voting and Science Election Team (VEST)[8] for 2018. I filtered both for New York County, or Manhattan.

The initial Manhattan election precinct shapes were difficult to clean and visually nonsensical. Some precincts were disjoint shapes; several snaked through neighborhoods as if they were gerrymandered. And, of course, many precinct shapes did not match across the years, as election officials modified their borders to serve changing populations. As in the previous chapter (where I describe the precinct merge process more extensively [Figure 4.4]), I identified the Census blocks that constituted each of the 2014 and 2018 precincts[9] and merged precinct shapes that shared Census blocks across the two data sets. The process worked slightly less smoothly than in the previous chapter, and a handful of precincts – including those including parkland or whose shapes erroneously cut through block borders – did need to be removed due to lack of block alignment or incorrect raw data[10]. However, these were few in number, and the resulting longitudinally-consistent data set covered over 95% of all populated Manhattan Census blocks. The resulting, final data set contained 775 original or combined precincts.

[6] `https://github.com/nvkelso/election-geodata/tree/master`

[7] `https://github.com/openelections`

[8] `https://election.lab.ufl.edu/data-archive/`

[9] Again, using blocks from `censable`: `https://christophertkenny.com/censable/`

[10] There appeared to be no solution to this, as blocks covering these showed no demographic information. Issues like this are common in New York election data, but the source of the problem is unknown here.

To build the treatment for my A1 regression (T_{A1}), I found whether each precinct contained at least one 2018 Citi Bike station and whether each contained at least one 2014 Citi Bike station, using the data described earlier. Then, if the former was true and the latter was not, I coded the precinct as treated, since my treatment is *whether a precinct had a new station in 2018 and no station in 2014.* Figure 5.5 visualizes that most Citi Bike expansion during this time period occured along Central Park on the Upper East and West Sides. Expressed technically with indicator variables, where $I = 1$ if true and 0 otherwise, and $n_{y,i}$ is the number of stations within precinct i in year y, for i in 775 precincts, the indicator treatment variable $\mathrm{T}_{\mathrm{A1},i}$ is:

$$I\big(n_{2018,i} > 0\big) \cdot \Big[I\big(n_{2018,i} > 0\big) - I\big(n_{2014,i} > 0\big)\Big] = \mathrm{T}_{\mathrm{A1},i}$$

As in the previous chapter, I computed turnout by dividing the number of votes at the highest common level of office, or the gubernatorial election, by the number of active registered voters. Change is computed as the difference between this statistic for 2018 minus that for 2014. Figure 5.4 shows relatively uniform turnout increase across Manhattan. For i in 775 precincts, the outcome variable is:

$$\frac{\big(\mathrm{Counted}\big)_{2018,i}}{\big(\mathrm{Registered}\big)_{2018,i}} - \frac{\big(\mathrm{Counted}\big)_{2014,i}}{\big(\mathrm{Registered}\big)_{2014,i}} = (\Delta\mathrm{turnout})_{[2014 \text{ to } 2018],i}$$

Finally, I included many of the same confounding variables as the previous chapter: demographic characteristics race, income, and age, and electoral controls for baseline year turnout and Congressional district fixed effects. The former are meant to control for fluctuations in voter mobilization from external factors, including salient political issues, outreach, and more. The latter control for either technical aspects of the variable construction (baseline lower-turnout precincts have more “room for improvement” mathematically) or more campaign and election cycle effects than may differ, for instance, by competitiveness of a congressional seat. I gathered median income and median age via *population-weighted block* *group-projection* (pwBG) using 5-year ACS data from 2009-2014 from `tidycensus`[11] (an in-depth description can be found in the Portland chapter [3.11]). I included 2010 percent white, Black, and Hispanic in this model given the complex diversity of Manhattan and results indicating this does cause an overfit baseline model; these data points came from the underlying Census blocks used to combine precincts. Finally, Congressional district shapefiles came from the Redistricting Data Hub[12]; their units are proportion overlap with each precinct, which in all but a few cases was 0 or 1.

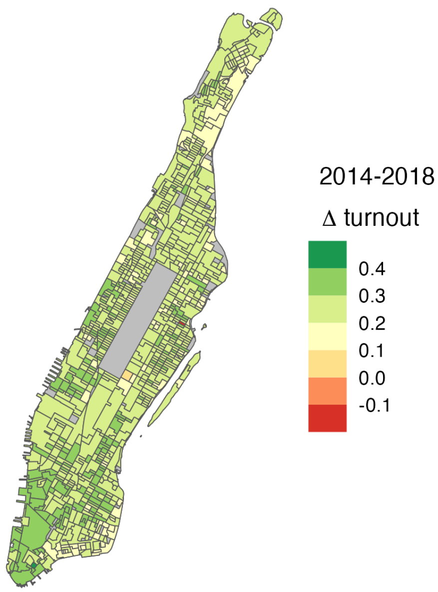


Figure 5.4: *A1* outcome variable

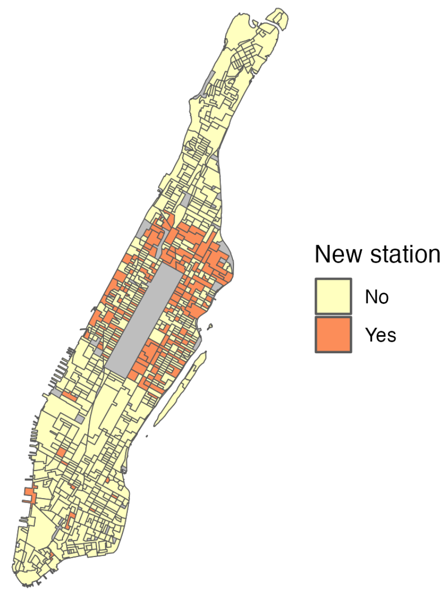


Figure 5.5: *A1* treatment variable

Below, Table 5.2 displays the results of baseline regression *Model 1a*, with only confounders, and *Model 1b* including the Citi Bike station treatment variable. *Model 1b* suggests there is no statistically significant relationship between the presence of a new Citi Bike station and voter turnout change between these election years. Instead, some confounders help explain fluctuations in turnout: precincts that were older or more Hispanic saw statistically significant, relatively higher turnout declines, all else held equal, and those that were more white in 2010 saw a significant turnout increase.

[11] `https://walker-data.com/tidycensus/`
[12] `https://redistrictingdatahub.org`

Table 5.2: *Models 1a* and *1b*

	Difference in % turnout, 2014-2018	
	Model 1a	Model 1b
2014 % turnout[13]	−0.241***	−0.244***
	(0.029)	(0.029)
% white, 2010	0.100***	0.102***
	(0.020)	(0.020)
% Black, 2010	0.002	0.004
	(0.020)	(0.020)
% Hispanic, 2010	−0.071***	−0.071***
	(0.017)	(0.017)
Median income (pwBG) $\times 10^{-4}$[14]	−0.003	−0.003
	(0.005)	(0.005)
Median age (pwBG)	−0.198***	−0.196***
	(0.027)	(0.027)
CD-07[15]	−65.056	−63.127
	(136.959)	(136.840)
CD-10	−61.895	−59.922
	(136.962)	(136.843)
CD-12	−63.751	−61.816
	(136.965)	(136.846)
CD-13	−60.611	−58.589
	(136.967)	(136.849)
New Citi Bike station		**−0.597**
		(0.386)
Constant	99.194	97.233
	(136.969)	(136.850)
Observations	775	775
R^2	0.399	0.401
Adjusted R^2	0.391	0.392
Residual Std. Error	0.040 (df = 764)	0.040 (df = 763)
Note:		*p<0.1; **p<0.05; ***p<0.01

Given there is no apparent relationship between the placement of new Citi Bike stations and turnout, I now turn to Analyses 2 and 3, which uniquely examine Citi Bike volume on Election Day to test the potential role of bikes in going to the polls in the first place.

[14]Baseline turnout and race variables are multiplied by 100, so their units are percentages with theoretical range [0, 100], not proportions.

[15]Median income is divided by 10^{-5} for coefficient interpretability.

A2: Modelled system use versus Election Day

In *Analysis 2* (A2), I test for differences between expected Citi Bike system volume on Election Day and expected volume on an ordinary day in Manhattan. This analysis is related to H3 in that it is meant to explain the mechanism behind A1; it investigates the extent to which bikes play a role in transportation on Election Day.

Election Day is impacted by several variables, so a model predicting "expected" ridership must include important factors of *weather* and *weekday*. Poor weather makes biking unpleasant and may suppress ridership; the day of week may explain work week-based cyclical commuting patterns (for instance, people may have more energy to bike to work on a Monday). Before controlling for these, comparing average Oct.-Nov. turnout compared to Election Day turnout for 2014 and 2018 yields contradictory results; in 2014, stations saw above-average ridership on Election Day, and the opposite was true for 2018.

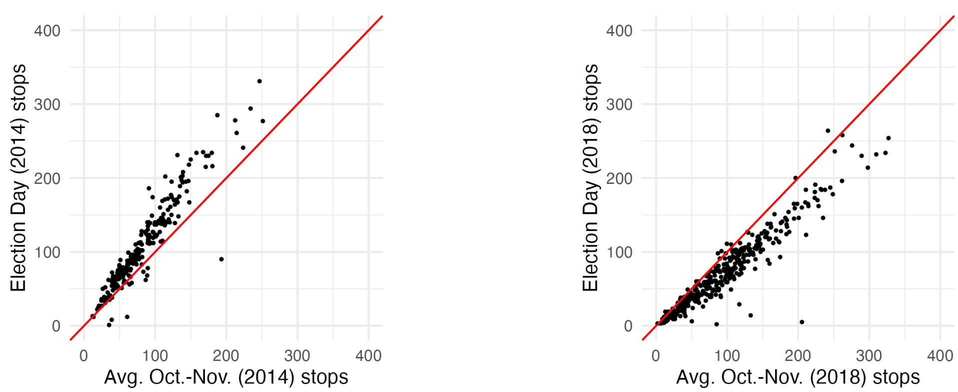


Figure 5.6: Average Oct.-Nov. turnout compared to Election Day turnout, by station and election year, with red $y = x$ line.

Weather data comes from the National Weather Service (NWS) via Kaggle[16]; this includes daily precipitation, snowfall, and snow depth in inches, as well as high and low temperatures, measured from Central Park since 1869. Weekday was already included in my cleaned Citi Bike data set.

[16] `https://www.kaggle.com/datasets/danbraswell/new-york-city-weather-18692022`

As described, for each station, day, and year, I calculated the *proportion of Election Day station stops to that year's Oct.-Nov. average* as my outcome variable, which gives a measure of ridership. Regressing on this, then, for a training set including all days *except for annual Election Days* each year from 2014 to 2018, produces a model explaining why station ridership is higher or lower than its average based on weather and day of week.

Table 5.4 shows my *Weather-weekday model*. Rain and ridership were negatively correlated, while minimum and maximum temperatures are associated with higher volume. Station ridership is was higher earlier in the week. The positive association for snowfall is confusing[17], but the regression otherwise holds logically.

	Prop. daily stops to annual Oct.-Nov. avg.
Precipitation (in.)	-0.453^{***}
	(0.004)
Snowfall (in.)	0.050^{***}
	(0.002)
Snow depth (in.)	-0.002
	(0.003)
Min. temperature (°F)	0.005^{***}
	(0.0003)
Max. temperature (°F)	0.013^{***}
	(0.0003)
Monday	0.040^{***}
	(0.004)
Tuesday	0.074^{***}
	(0.004)
Wednesday	0.064^{***}
	(0.004)
Thursday	-0.033^{***}
	(0.004)
Constant	0.079^{***}
	(0.008)
Observations	39,362
R^2	0.445
Adjusted R^2	0.445
Residual Std. Error	0.259 (df = 39352)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5.3: *Weather/weekday model*

[17] The snowfall variable isn't cancelled out by snow depth, since the former is of higher magnitude and the latter is usually first measured after a snow day. Perhaps leisure ridership is higher on these days, or this effect is "cancelled out" by much lower temperatures and their variable effects.

Next, I use this model to predict ridership proportion to annual average for Election Day on 2014 and 2018. The former saw clear skies and a high temperature near 70°F, and the model predicts the proportion of 2014 Election Day ridership compared to 2014 Oct.-Nov. average to be 1.29, or 29% above-average ridership. For Election Day 2018, with weather in the fifties and almost an inch of rain, this proportion is predicted to be about 0.85, or 85% of average 2018 Oct.-Nov. ridership.

Finally, I look at the distribution of the proportion of Election Day rides to average Oct.-Nov. rides for 2014 and 2018 at all stations. These are shown below in Figure **??**, where the red vertical line is the modelled proportion. A 95% confidence interval for the center of the 2014 distribution is $\left[1.273, 1.332\right]$, which contains the modelled proportion of 1.29. On the other hand, a 95% confidence interval for the center of the 2014 distribution is $\left[0.718, 0.778\right]$, which does not contain the modelled proportion of 0.85, telling us that ridership on Election Day in 2018 was statistically significantly *lower* than expected. This is clear in the histogram, which shows a 2018 distribution shifted left of the red line.

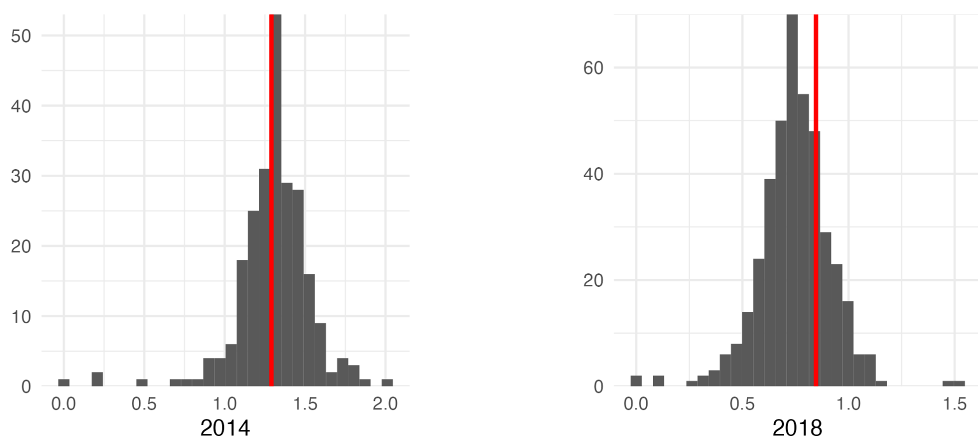


Figure 5.7: Election Day 2014 and 2018 ridership as a proportion of average annual Oct.-Nov. ridership by station, with red line *weather-weekday modelled* proportion

As this result suggests that Election Day Citi Bike ridership is lower than average, if anything, Analysis 3 explores if the placement of Citi Bike stations in Manhattan places a role in voters' choice to find other means of transportation.

A3: Election Day station use & polling place proximity

In *Analysis 3* (A3), I test for a relationship between Citi Bike station ridership as a proportion of Oct.-Nov. average and proximity to polling places in 2018. This analysis helps explain the results of the previous two contradicting H3, or why I find that Citi Bike stations don't seem to boost turnout, nor do people use them disproportionately more on Election Day.

To do this, I gathered polling place data from NYC Open Data[18]; these were only available for 2018. The data includes approximately variables with the polling station name and address, but for about a sixth of the locations, the latitude and longitude variables were missing. To find missing coordinates, I used `tidygeocoder`[19] and polling place addresses, which recovered around half of the missing data. Using polling place names, I manually collected the remaining coordinates using Google Maps; these addresses matched those in the data. Polling places were most frequently set up in schools, community centers, and large apartment buildings or complexes.

Next, I construct two treatment options that capture a Citi Bike station's *proximity* to a polling place. For the first treatment option ($T1_{A3}$), I find the minimum distance (in meters) to a polling place from each Citi Bike station and take its square root[20] For the second treatment option ($T2_{A3}$), I construct a 250m buffer – which represents a few minutes in walking time – around each Citi Bike station and count the number of polling places within this buffer. Note that the latter of these two, the buffer option, is similar to the walkshed method used in previous chapters as a measure of proximity. Here, I use buffers, as Manhattan sits on a consistent street grid, so the shapes are similar. I use a several-minute walk instead of fifteen, as I am looking at the *convenience* of getting to a polling place by

[18] `https://data.cityofnewyork.us/City-Government/Voting-Poll-Sites/mifw-tguq/about_data`

[19] `https://jessecambon.github.io/tidygeocoder/`

[20] I do this as the non-transformed statistic is right-skewed; using the right-skewed statistic, however, does not change the high-level results nor their statistical significance.

docking a Citi Bike nearby. $T1_{A3}$ can be thought of as capture the *maximum* convenience of a station, while $T2_{A3}$ comes closer to capturing its *aggregate* convenience on Election Day in getting to a polling place. I reuse the 2018 Citi Bike station data set from A1 to compute these. Figure 5.8 shows the distribution of polling places overlaid on Citi Bike stations, and Figure 5.9 shows Citi Bike stations colored and sized based on $T1_{A3}$ and $T2_{A3}$, respectively.

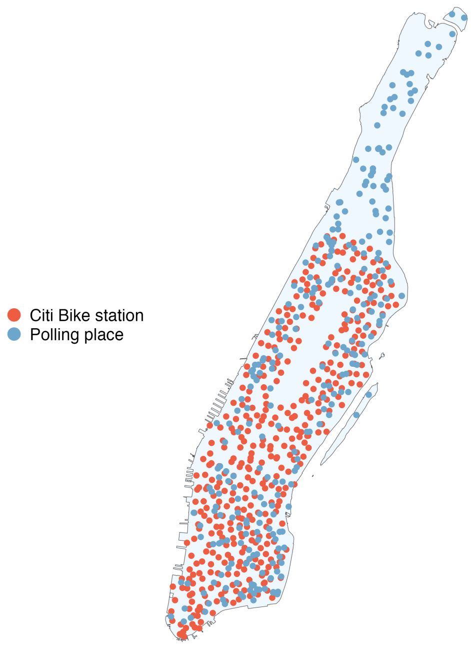


Figure 5.8: Polling places overlaid on Citi Bike stations, 2018, used to construct treament options $T1_{A3}$ and $T2_{A3}$

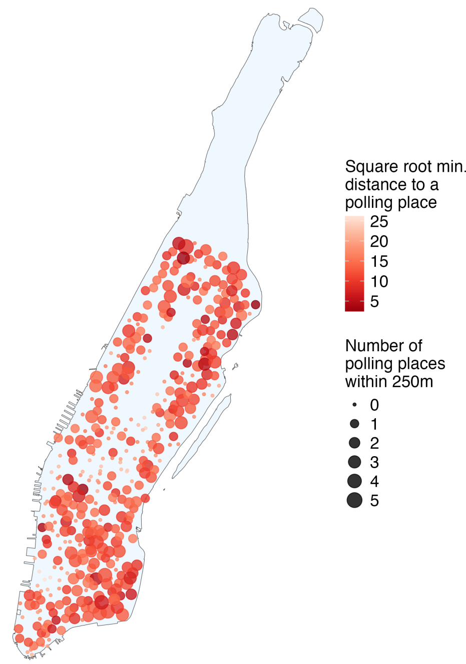


Figure 5.9: *Min. distance to* ($T1_{A3}$) and *nearby concentration of* ($T2_{A3}$) 2018 polling places from stations

Finally, I run a simple regression on Election Day station ridership as a proportion of 2018 Oct.-Nov. average ridership using each treatment: *Model 3a* uses the minimum distance option ($T1_{A3}$) while uses the nearby polling places option ($T2_{A3}$). Table 5.4 shows the results of these models, and Figure **??** visualizes each relationship below.

Table 5.4: *Models 3a* and *3b*

	Prop. E-Day stops to annual Oct.-Nov. avg.	
	Model 3a	Model 3b
[Min. distance to a polling station (m)]$^{0.5}$	$-\mathbf{0.007}^{***}$	
	(0.002)	
Number of polling places within 250m radius		$\mathbf{0.032}^{***}$
		(0.008)
Constant	0.844^{***}	0.713^{***}
	(0.028)	(0.012)
Observations	407	407
R^2	0.031	0.035
Adjusted R^2	0.028	0.033
Residual Std. Error (df = 405)	0.168	0.168
F Statistic (df = 1; 405)	12.827^{***}	14.816^{***}
Note:		*p<0.1; **p<0.05; ***p<0.01

In *Model 3a*, I find a statistically significant negative relationship between square root minimum distance to the nearest polling station from a station and its stop volume as a proportion of the Oct.-Nov. average. The coefficient of −0.007 is very small but suggests that for a 100 meter increase in a station's minimum distance to a polling place – about a one-minute walk – the proportion of Election Day ridership to Oct.-Nov. average is expected to fall by about 0.07; if Election Day ridership matches the average, this change amounts to a 7% decrease. In *Model 3b*, I find a similar result: the coefficient for the number of nearby polling places to a station is 0.032, suggesting that an additional polling place amounts to an increase of Election Day ridership of about 3% if it initially matched the Oct.-Nov. average [see Appendix # for a visualization of these relationships].

These results offer some insight into why A1 and A2 produced little evidence regarding the impact of or use of Citi Bikes on Election Day. Proximity to polling places matters in a station's Election Day ridership, as *Models 3a* and *3b* find that closer distance to polling places and more polling places nearby are associated with higher ridership. This suggests the Citi Bike system may not be ideally set up to facilitate transportation to the polls, meaning H3 is difficult to measure even using the United States' most prominent bike share system.

Analyses and results

Discussion

Chapter 6

Conclusion

Bibliography

John H. Aldrich. Rational choice and turnout. *American journal of political science*, 37(1): 246–278, 1993. ISSN 0092-5853.

Brian Amos, Michael P. McDonald, and Russell Watkins. When Boundaries Collide: Constructing a National Database of Demographic and Voting Statistics. *Public Opinion Quarterly*, 81(S1):385–400, 04 2017. ISSN 0033-362X. doi: 10.1093/poq/nfx001. URL `https://doi.org/10.1093/poq/nfx001`.

Muskaan Arshad. Separated bike lanes divide cambridge city council candidates ahead of nov. 7 election, November 2023. URL `https://www.thecrimson.com/article/2023/11/3/bike-lanes-elections-2023/`.

John J. Betancur. The politics of gentrification: The case of west town in chicago. *Urban affairs review (Thousand Oaks, Calif.)*, 37(6):780–814, 2002. ISSN 1078-0874.

Lindsay M. Braun, Daniel A. Rodriguez, and Penny Gordon-Larsen. Social (in)equity in access to cycling infrastructure: Cross-sectional associations between bike lanes and area-level sociodemographic characteristics in 22 large u.s. cities. *Journal of transport geography*, 80:102544, 2019. ISSN 0966-6923.

Tim Burke, Aakriti Kaul, and Jake Mainwaring. The effect of moving precincts on voting. May 2019. URL `https://www.ischool.berkeley.edu/sites/default/files/sproject_attachments/paper_-_voting_capstone.pdf`.

Craig M. Burnett and Vladimir Kogan. The politics of potholes: Service quality and retrospective voting in local elections. *The Journal of politics*, 79(1):302–314, 2017. ISSN 0022-3816.

Justin de Benedictis-Kessner. How attribution inhibits accountability: Evidence from train delays. *The Journal of politics*, 80(4):1417–1422, 2018. ISSN 0022-3816.

Adam M. Dynes. Which citizens do elected officials target with distributive spending? a survey experiment on u.s. municipal officials. *American politics research*, 48(5):579–595, 2020. ISSN 1532-673X.

Katherine Levine Einstein. *Neighborhood defenders :participatory politics and America's housing crisis*. Cambridge University Press, Cambridge, 1st ed. edition, 2020. ISBN 1-108-84819-2.

J.G. Gimpel and J.E. Schuknecht. Political participation and the accessibility of the ballot box. *Political geography*, 22(5):471–488, 2003. ISSN 0962-6298.

Zoltan Hajnal and Paul G. Lewis. Municipal institutions and voter turnout in local elections. *Urban affairs review (Thousand Oaks, Calif.)*, 38(5):645–668, 2003. ISSN 1078-0874.

Zoltan Hajnal and Jessica Trounstine. Where turnout matters: The consequences of uneven turnout in city politics. *The Journal of politics*, 67(2):515–535, 2005. ISSN 0022-3816.

Zoltan L. Hajnal, Vladmir Kogan, and G. Agustin Markarian. Who votes: City election timing and voter composition. *The American political science review*, 116(1):374–383, 2022. ISSN 0003-0554.

Moshe Haspel and H. Gibbs Knotts. Location, location, location: Precinct placement and the costs of voting. *The Journal of politics*, 67(2):560–573, 2005. ISSN 0022-3816.

Henry Grabar. *Paved paradise : how parking explains the world*. Penguin Press, New York, 2023. ISBN 9781984881144.

Marek Hlavac. *stargazer: Well-Formatted Regression and Summary Statistics Tables*. Social Policy Institute, Bratislava, Slovakia, 2022. URL https://CRAN.R-project.org/package=stargazer. R package version 5.2.3.

Thomas M. Holbrook and Aaron C. Weinschenk. Campaigns, mobilization, and turnout in mayoral elections. *Political research quarterly*, 67(1):42–55, 2014. ISSN 1065-9129.

Daniel J. Hopkins, Susanne Schwarz, and Anjali Chainani. Officially mobilizing: Repeated reminders and feedback from local officials increase turnout. *The Journal of politics*, 85 (2):771–777, 2023. ISSN 0022-3816.

Derek Hyra. The back-to-the-city movement: Neighbourhood redevelopment and processes of political and cultural displacement. *Urban studies (Edinburgh, Scotland)*, 52(10):1753–1773, 2015. ISSN 0042-0980.

J. Eric Oliver. *Local Elections and the Politics of Small-Scale Democracy.* Princeton University Press, Princeton, NJ, course book edition, 2012. ISBN 1-280-49406-9.

Boqian Jiang. Homeownership and voter turnout in u.s. local elections. *Journal of housing economics*, 41:168–183, 2018. ISSN 1051-1377.

Alexandra Kanik. How many of your neighbors voted in the 2020 election? see voter turnout by harris county zip codes, October 2022. URL https://www.houstonchronicle.com/politics/election/2022/article/2020-election-turnout-ZIP-codes-17515500.php.

Amy Lubitow. Mediating the 'white lanes of gentrification' in humboldt park: Community-led economic development and the struggle over public space. In *Bicycle Justice and Urban Transformation*, pages 249–259. Routledge, United Kingdom, 1 edition, 2016. ISBN 9781138950245.

Melissa Marschall and John Lappie. Turnout in local elections: Is timing really everything? *Election law journal*, 17(3):221–233, 2018. ISSN 1533-1296.

Rodrigo Mora, Ricardo Truffello, and Gabriel Oyarzún. Equity and accessibility of cycling infrastructure: An analysis of santiago de chile. *Journal of transport geography*, 91:102964, 2021. ISSN 0966-6923.

Richard Nunno. Making way for bicycle infrastructure, 2021. URL `https://www.eesi.org/articles/view/making-way-for-bicycle-infrastructure`.

J. Eric Oliver. *Local Elections and the Politics of Small-Scale Democracy.* Princeton University Press, Princeton, NJ, course book edition, 2012. ISBN 1-280-49406-9.

William H. Riker and Peter C. Ordeshook. A theory of the calculus of voting. *The American Political Science Review*, 62(1):25–42, 1968. ISSN 00030554, 15375943. URL `http://www.jstor.org/stable/1953324`.

Matt Sledge. Whitmire's lopsided win in the houston mayoral runoff came amid low turnout, results show, December 2023. URL `https://houstonlanding.org/whitmires-lopsided-win-in-the-houston-mayoral-runoff-came-amid-low-turnout-results-show/`.

Charles Stewart. 2016 Survey of the Performance of American Elections, 2017. URL `https://doi.org/10.7910/DVN/Y38VIQ`.

Caroline Tait, Roger Beecham, Robin Lovelace, and Stuart Barber. Is cycling infrastructure in london safe and equitable? evidence from the cycling infrastructure database. *Journal of transport health*, 26:101369, 2022. ISSN 2214-1405.

Tom Fucoloro. Biking uphill in the rain. University of Washington Press, 2023. ISBN 0295751584.

Jessica Trounstine. Turnout and incumbency in local elections. *Urban affairs review (Thousand Oaks, Calif.)*, 49(2):167–189, 2013. ISSN 1078-0874.

Kirsty Wild, Alistair Woodward, Adrian Field, and Alex Macmillan. Beyond 'bikelash': engaging with community opposition to cycle lanes. *Mobilities*, 13(4):505–519, 2018. ISSN 1745-0101.