Table 1. Study Data Sources

Data Source	*Description of Source and Potential Relationship to Study*	*Potential for Use in Study?*	*Rationale*
Longitudinal Employer-Household Dynamics (LEHD) employment data	US Census database of employers and employees, including earnings and demographics. LEHD data could be used to track changes in employment and businesses near newly installed bike facilities, relative to a control group. LEHD Origin-Destination Employment Statistics dataset with workplace area characteristics can be used to analyze employment by census block.	Yes	Due to data privacy, obtaining access to LEHD at a level with more granularity than the census block level would involve long lead times. Data at the census block level does not allow the level of detail necessary to identify businesses facing a particular roadway improvement. However, census blocks could be assigned to a corridor based on a radius from the roadway. Data at the annual level is more aggregated than would be preferred and does not allow for precise measurement of changes in employment relative to the completion date of a project. Employment levels tend to be a lagging and indirect indicator of changes in retail activity but may be used to validate control areas where sufficient data are available. These data are available in 2011 through 2020, so sufficient data is not available for more recent installations. This data may be useful in identifying existing trends in employment overall and by industry in the City of Cambridge relative to the surrounding area and in identifying existing trends in employment for census blocks near specific bike corridors compared to a control area.
Pedestrian and Bike Counts, Parking information	City-collected counts of pedestrian and bike traffic at multiple locations, including before-and-after studies at bike facilities. Parking availability information or studies, as available.	Limited use	These data do not provide direct evidence on changes in business sales but allow other information to be viewed in context.
Business establishments opened/closed and by type of establishment	City-collected data on business openings and closings	Limited use	Business openings and closings are affected by many factors and can be difficult to associate with changes in bike facilities or other projects. However, the data may be useful in providing context for other findings.

Data Source	*Description of Source and Potential Relationship to Study*	*Potential for Use in Study?*	*Rationale*
Sidewalk and in-person customer intercept surveys	City-collected survey data from shopper intercepts at bike project locations and controls. These data provide information on mode of travel and purpose of trip.	Limited use	These data do not provide direct evidence on changes in business sales but allow other information to be viewed in context.
Business surveys	City-collected data from surveys of business owners in affected locations versus control groups. This will provide information on changes in retail activity relative to a pre-covid baseline.	Yes	Because the survey is voluntary and the projects are public in nature, there are several sources of potential bias including nonresponse bias. In addition, many macro- and micro-economic factors other than bike projects can affect sales volumes. Nonetheless, the survey data, as a supplement to other data sources, will a provide direct source of information on changes in the business environment in addition to information regarding perceived impacts.
Commercial real estate lease rate and/or occupancy/vacancy data	Private sector providers such as CoStar estimate rental rates and vacancy rates for commercial properties. Data on properties near bike projects can be compared to a control group.	Yes	Because the value of urban real estate hinges on the commercial desirability of its location, changes in rents can provide a useful indirect measure of changes in retail activity and overall accessibility. Occupancy/ vacancy data also provide a secondary measure. These datasets are credible as they are used widely in the real estate industry, and they permit geographic analysis. Although property values are also available through the city assessor, these involve longer time lags, and the assessment methodology may not account for smaller changes in neighborhood access. Use of data is dependent on availability for identified corridors, and some data review and cleaning will be necessary to permit analysis.

Data Source	*Description of Source and Potential Relationship to Study*	*Potential for Use in Study?*	*Rationale*
Point of interest (POI) transaction data	Information from electronic transactions can be purchased from a POI data company such as SafeGraph.	Yes	POI transaction data provides anonymized credit and debit card transaction information which can be used to track local sales trends directly. Use of this data would require a purchase of the dataset. Key limitations are the exclusion of non-credit card transactions, and the lack of transparency regarding potential gaps in the data. The earliest available data is from 2019, so it is likely that this data could only be used to assess trends in beginning in the latter half of 2019 or later. Based on initial discussion with SafeGraph, the data contains primarily full-service restaurants, snack and nonalcoholic beverage bars, limited-service restaurants, hair/nail/skin care services, and fitness centers. The dataset includes both chain and independent businesses at the store level by month. Potential variables of interest could be total sales amount, spend per customer, number of transactions, or number of customers. Some data review and cleaning will be necessary to permit analysis.
Bike lane information	Geographic and time information regarding bike lane installation	Yes	Assessing bike lane impacts requires location and timing information for impacted corridors.

Table 2. Other Data Sources Considered

Data Source	*Description of Source and Potential Relationship to Study*	*Potential for Use in Study?*	*Rationale*
Payment processor transaction data	Information from credit card and other electronic transactions could be purchased from a financial institution and used to track changes in retail sales directly.	No	Unable to find a private sector partner willing to provide such data. There would also be key limitations, such as not providing information on cash or out-of-network transactions.
City of Cambridge local-option meals tax	These data can be used to shed light on changes in restaurant sales, but not on other retail sectors.	No	Lack of information on other retail sectors makes the use of this dataset as a primary source of information insufficient. In addition, data are only available at aggregate level from the State.

Data Source	*Description of Source and Potential Relationship to Study*	*Potential for Use in Study?*	*Rationale*
Massachusetts Department of Revenue meals and sales tax data	Sales and meals tax receipts are based on retailers' own tax documentation, can be analyzed geographically, and provide direct evidence of changes in retail activity. Most retail activity is covered (except for groceries and clothing).	No	The Massachusetts Department of Revenue is unable to release the data at a granular level due to taxpayer privacy considerations. It may be possible to review information from businesses that "opt in" to provide tax data. However, options for this "opt in" process appear to be too broad.
Quarterly Census of Employment and Wages employment and wage data	Data from reports filed by employers subject to unemployment compensation laws, produced publicly at the city/town and county level by NAICS industries.	No	The level of detail for publicly available data is too aggregated to be useful. Massachusetts is a non-signatory state, indicating that projects are approved based on individual state laws. Obtaining data at the establishment level would involve approvals and long lead times.
National Establishment Time Series employment and sales data	This product was established as a potential source via literature review of past studies.	No	Unable to find a current source for this data product.