# **Parcels - Extract, Transform, Load**

This workflow is performed using the **parcels\_finalproject\_etl script** found in the scripts folder.

# **Data Acquisition**

In RStudio, connect to the MD Open Data Portal using sf R package and download the necessary parcels & sales dataset, transit, county datasets will be merged during data transformation steps.

# **Data Cleaning and Transformation**

n RStudio, parcel data was extracted using a Socrata (SoQL) query to select relevant fields, filter valid records, and generate two flags indicating new development (yr\_built\_flag, for properties built in or after 2020) and residential land use (lu\_code\_flag). The raw data was then cleaned and transformed using tidyverse packages (dplyr, stringr) and spatially processed with the sf package. This included address standardization, ZIP code validation, and projection to Maryland State Plane (EPSG:26985). The final dataset, with standardized fields and valid geometries, was written into a PostgreSQL/PostGIS database for further spatial analysis., please refer to the [Data Model Justifications](#_61bxlpqseze), [ETL Logic Model](https://docs.google.com/document/d/1EEOONWZlpRRa67cm7Dln5GdU_pkXy5rGWUQzWgvzAAk/edit?tab=t.evygnoxyk6e2), and [Scripts Appendix](https://docs.google.com/document/d/1EEOONWZlpRRa67cm7Dln5GdU_pkXy5rGWUQzWgvzAAk/edit?tab=t.i5eg5b7x65l3).

## **Data Loading**

From RStudio, a database connection is established using the DBI and RPostgres R packages. The cleaned dataset is then loaded into the development database. The first time the data loading process is performed, the development database acts as a shell; no previous data is in the database. Subsequent database restorations will first truncate/delete the previous parcel dataset in the database, then load the new version.

# **Counties - Extract, Transform, Load**

This workflow is performed using the **county boundaries ETL script** found in the scripts folder.

# **Data Acquisition**

In RStudio, connect to the MD Open Data Portal using arcgislayers R package and download the necessary parcels & sales datasets, per county; county datasets will be merged during data transformation steps.

# **Data Cleaning and Transformation**

Continuing in RStudio, using tidyverse R packages (e.i. dplyr & stringr), create a function that cleans the county names to match those in the domain dCountyName. For more information about the data cleaning steps, please refer to the [Data Model Justifications](https://docs.google.com/document/d/1KXxr9GekCMRrQwykF5NYDAVm5CdMHyZdQOAt1D5LLBE/edit?tab=t.o50w2vggkct8), [ETL Logic Model](https://docs.google.com/document/d/1EEOONWZlpRRa67cm7Dln5GdU_pkXy5rGWUQzWgvzAAk/edit?tab=t.evygnoxyk6e2), and [Scripts Appendix](https://docs.google.com/document/d/1EEOONWZlpRRa67cm7Dln5GdU_pkXy5rGWUQzWgvzAAk/edit?tab=t.i5eg5b7x65l3).

## **Data Loading**

From RStudio, a database connection is established using the DBI and RPostgres R packages. The cleaned dataset is then loaded into the development database. The first time the data loading process is performed, the development database acts as a shell; no previous data is in the database. Subsequent database restorations will first truncate/delete the previous parcel dataset in the database, then load the new version.

### **Transit Stations – Extract, Transform, Load**

This workflow is performed using the transit data ETL script found in the scripts folder.

**Data Acquisition** In RStudio, transit facility datasets (MARC, Amtrak, WMATA, MTA Bus, and RideOn) were acquired from Maryland’s Open Data Portal and ArcGIS Feature Services using the arcgislayers R package. Each dataset was downloaded individually as a point layer containing spatial and attribute information about transit stops and stations.

**Data Cleaning and Transformation** Using tidyverse packages (dplyr, stringr) and the sf package, each dataset was standardized to match the project’s spatial data model. Cleaning steps included renaming columns, extracting ZIP codes, removing null geometries, converting coordinates to the Maryland State Plane projection (EPSG:26985), and formatting field types and character lengths according to schema constraints. please refer to the [Data Model Justifications](#_ax66zihqgz5b), [ETL Logic Model](https://docs.google.com/document/d/1EEOONWZlpRRa67cm7Dln5GdU_pkXy5rGWUQzWgvzAAk/edit?tab=t.evygnoxyk6e2), and [Scripts Appendix](https://docs.google.com/document/d/1EEOONWZlpRRa67cm7Dln5GdU_pkXy5rGWUQzWgvzAAk/edit?tab=t.i5eg5b7x65l3).

**Data Loading** With DBI and RPostgres, each cleaned dataset was written into its respective table within the trans schema of the development database. For each dataset load, the target table is first truncated to remove any existing records, ensuring only the latest data is retained. Spatial indexes were applied to improve performance in spatial queries and analysis.

# **Mirroring Database Environments**

This process is performed in RStudio using the **production mirror ETL** in the scripts folder. Once the database has been quality checked in the development environment, this process copies the development database into the production database. This process ensures data integrity and security. In order for this process to work without error, it is necessary to follow the same database/schema/table/user role naming conventions when setting up the production environment database.

### **Data Model Justifications**

### **Overview**

The proposed data structure for the **Transit Access Project** is a relational database that utilizes role-based access control to ensure data security and the use of view tables for spatial analysis and visualization. A [data model template](https://docs.google.com/spreadsheets/d/1JcjCYup_uH2cekN8rZemdVAyW8TC2EOO/edit?gid=487211832#gid=487211832) was used, courtesy of Mike Haxel. The [MD\_Transit\_Data\_Model](https://docs.google.com/spreadsheets/d/1Ewg8PTCsSB5PEE6FlmcX6Muojs2XkK3W/edit?gid=668135550#gid=668135550) documents the transformation of data from our sources into the new database. The database design was implemented using pgAdmin, and the records were transformed and populated using R Statistical Software.

### **Data Normalization Strategy**

The following general decisions were implemented to standardize the data:

* Remove unnecessary fields from source tables.
* Transform and clean spatial and attribute data using R.
* Create and apply custom domains for county names, land use codes, and flags.
* Store all parcel points in a single table for efficient spatial filtering.
* Create flags to evaluate new development status (yr\_built\_flag) and land use type (lu\_code\_flag).
* Use standardized schema names (e.g., acplan, bndy, trans) for organizational clarity.
* Create view tables for spatial proximity analysis between parcels and transit facilities.

### **Multi-Database Environment**

A two-database strategy was implemented to streamline development and production workflows. The **transit\_access\_project** development database enables testing and revision of the data model and queries, while the production database stores finalized data. Role-based access ensures that developers and analysts operate with appropriate privileges.

### **Spatial Reference System**

All spatial data are standardized to the **Maryland State Plane (EPSG:26985)** using st\_transform(). This ensures consistency when conducting spatial joins, buffers, and proximity calculations across datasets like parcels, transit stations, and county boundaries.

### **Database Schemas**

Twelve schemas were defined to separate datasets thematically and restrict access by role. Key schemas for this project include:

* ACPLAN = Planning and property data
* TRANS = Public transportation data (MARC, Amtrak, MTA Bus, WMATA, RideOn)
* BNDY = County boundary polygons
* Other schemas (e.g., econ, soci, envi) are reserved for future use.

### **Core Attributes**

The ETL workflow was used to transform the Maryland parcel dataset into core attributes needed to evaluate development status and residential land use. The following format is used to document how each field was created or transformed:

**Format:**

* new\_field\_name = new field alias
  + Legacy..Field.Name.
  + Data transformation step / Field notes
* **account\_id = Parcel Account ID**
  + Account.ID..MDP.Field..ACCTID.
  + Converted to character; invalid values such as NULL, 0, or blank were excluded.
* **premise\_addr = Premise Address (concatenated)**
  + PREMISE.ADDRESS..Number..MDP.Field..PREMSNUM..SDAT.Field..20
  + PREMISE.ADDRESS..Direction..MDP.Field..PREMSDIR..SDAT.Field..22.
  + PREMISE.ADDRESS..Name..MDP.Field..PREMSNAM..SDAT.Field..23.
  + PREMISE.ADDRESS..Type..MDP.Field..PREMSTYP..SDAT.Field..24.  
     Concatenated components; standardized street suffixes (e.g., “STREET” → “ST”) using regex rules and stringr.
* **premise\_city = Premise City**
  + PREMISE.ADDRESS..City..MDP.Field..PREMCITY..SDAT.Field..25.
  + Converted to uppercase and cleaned; missing values replaced with "NA".
* **premise\_zip = Premise ZIP Code**
  + PREMISE.ADDRESS..Zip.Code..MDP.Field..PREMZIP..SDAT.Field..26.
  + Converted to character; only valid 5-digit ZIP codes retained.
* **premise\_state = Premise State**
  + Derived
  + Manually assigned as "MD" for all parcels.
* **mailing\_addr = Mailing Address**
  + MDP.Street.Address..MDP.Field..ADDRESS.
  + Converted to uppercase and standardized using suffix replacements.
* **mailing\_city = Mailing City**
  + MDP.Street.Address.City..MDP.Field..CITY.
  + Converted to uppercase; missing values replaced with "NA".
* **mailing\_zip = Mailing ZIP Code**
  + MDP.Street.Address.Zip.Code..MDP.Field..ZIPCODE.
  + Converted to character; invalid or missing values set to "NA".
* **jurscode = Jurisdiction Code**
  + Jurisdiction.Code..MDP.Field..JURSCODE.
  + Mapped to county codes using a lookup table in J\_COUNTY\_CODE\_JURSCODE.
* **county\_name = County Name**
  + [County.Name](http://county.name)..MDP.Field..CNTYNAME.
  + Cleaned using stringr; matched to domain dCountyName.
* **yr\_built = Year Built**
  + C.A.M.A..SYSTEM.DATA..Year.Built..YYYY...MDP.Field..YEARBLT..SDAT.Field..235.
  + Converted to integer; years < 1000 set to 0.
* **lu\_code = Land Use Code**
  + Land.Use.Code..MDP.Field..LU.DESCLU..SDAT.Field..50.
  + Used to classify parcels by land use type (e.g., Residential, Commercial).
* **real\_prop\_link = SDAT Property Search Link**
  + Real.Property.Search.Link
  + Retained as-is for linking to public property record.
* **yr\_built\_flag = Year Built Flag**
  + Derived
  + 1 if yr\_built >= 2020, else 0.
* **lu\_code\_flag = Land Use Flag**
  + Derived
  + 1 if lu\_code matches residential categories, else 0.
* **geom = Parcel Point Geometry**
  + MDP.Longitude..MDP.Field..DIGXCORD.converted.to.WGS84.
  + MDP.Latitude..MDP.Field..DIGYCORD.converted.to.WGS84.
  + Converted to sf point geometry, projected to EPSG:26985 (Maryland State Plane).

### **Transit Data Attributes**

Each transit dataset (MARC, AMTRAK, WMATA, MTA Bus, and RideOn) was loaded into its own table within the trans schema. Spatial fields were standardized to SRID 26985 to ensure alignment with other spatial datasets. Below are field-level justifications for each dataset:

#### **MARC Train Stations**

**Table:** trans."MARC"  
 Format:  
 new\_field\_name = new field alias  
 Legacy Field Name  
 Transformation/Notes

* objectid = Station ID  
   OBJECTID  
   Retained as unique identifier
* station\_name = Station Name  
   Name  
   Renamed for clarity
* station\_address = Street Address  
   ADDRESS  
   Renamed
* station\_city = City  
   CITY  
   Standardized to uppercase
* station\_state = State  
   STATE  
   Assumed to be "MD"
* station\_zip = ZIP Code  
   ZIP  
   Converted to character; truncated to 5 digits if needed
* transit\_mo = Transit Mode  
   Transit\_Mo  
   Identifies rail mode
* line\_name = Line Name  
   Line\_Name  
   Used for route identification
* facility\_t = Facility Type  
   Facility\_T  
   Useful for classifying stops
* geom = Point Geometry  
   Derived from geometry field  
   Reprojected to EPSG:26985

#### **AMTRAK Stations**

**Table:** trans."AMTRAK"

* objectid = Station ID  
   OBJECTID  
   Kept as primary key
* station\_name = Station Name  
   STNNAME  
   Renamed
* station\_type = Station Type  
   STNTYPE  
   Indicates classification
* station\_code = Station Code  
   STNCODE  
   Useful for lookups
* station\_addr = Street Address  
   ADDRESS1  
   Renamed
* station\_city = City  
   CITY  
   Standardized
* station\_state = State  
   STATE  
   Assumed "MD"
* station\_zip = ZIP Code  
   ZIP  
   Truncated to 5 characters
* geom = Geometry Point  
   Reprojected to EPSG:26985

#### **WMATA Metro Stations**

**Table:** trans."WMATA"

* gis\_id = Station ID  
   GIS\_ID  
   Unique key
* station\_name = Station Name  
   NAME  
   Renamed
* station\_addr = Address  
   ADDRESS  
   Retained
* metroline = Metro Line  
   MetroLine  
   Identifies rail service
* geom = Geometry  
   Reprojected to EPSG:26985

#### **MTA Bus Stops**

**Table:** trans."MTABUS"

* objectid = Stop ID  
   OBJECTID  
   Used as unique row identifier
* stop\_name = Stop Name  
   stop\_name  
   Renamed
* routes\_served = Served Routes  
   Routes\_Served  
   Converted NA to “NA” string
* mode = Transit Mode  
   Mode  
   Helps distinguish type  
  county = County Name  
   County  
   Standardized
* stop\_id = Stop ID Code  
   stop\_id  
   Useful for joins or comparison
* geom = Geometry  
   EPSG:26985

#### **RideOn Bus Stops**

**Table:** trans."RIDEON"

* objectid = Stop ID  
   Derived from arcgislayers output
* stop\_code = Stop Code  
   stop\_code  
   Converted to character
* stop\_name = Stop Name  
   stop\_name  
   Standardized
* town = Town Name  
   Town  
   NA values converted to “NA” string
* geom = Geometry  
   EPSG:26985

**Data Sources**

**All datasets were acquired from the Maryland iMAP open data catalog**

* Parcels & Sales assessments prepared by Maryland Department of Planning (MDP)
  + **Point type feature class**
  + [Montgomery County Parcels & Sales](https://opendata.maryland.gov/Business-and-Economy/Montgomery-County-Real-Property-Assessments-Hidden/kb22-is2w/about_data)
  + These datasets have been synthesized from the Maryland State Department of Assessments and Taxation (SDAT) Parcel and Sales datasets. It contains over 200 fields.
  + It is the most recently updated version of the data.
* [Generalized county boundaries prepared by Maryland Department of Transportation (MDOT)](https://data.imap.maryland.gov/datasets/maryland-physical-boundaries-county-boundaries-generalized/explore?location=39.330957%2C-76.420360%2C11.13)
  + **Polygon type feature class**
  + The generalized county boundaries were chosen to ensure efficient data querying and clearer data visualization.
  + These county polygons are more detailed than the MDOT county political boundaries,
  + And less detailed than the MDOT county physical boundaries, which was synthesized from the political boundaries + the National Hydrology Data (NHD)
* [Maryland Political Boundaries - ZIP Codes - 5 digit](https://data.imap.maryland.gov/datasets/046ba81433b94c0d90465adec040de1a_4/explore?showTable=true)
  + **Polygon type feature class**
  + The 5-digit Maryland zipcodes prepared by MDoIT.
  + This dataset will be used to compare between mailing zip and the list of all MD zips.
* [MARC Train Stops](https://opendata.maryland.gov/Transportation/MD-iMAP-Maryland-Transit-MARC-Train-Stops/qmkd-vkf9/about_data) provided by Maryland Transit Administration, Washington Metropolitan Area Transit Authority, United States Bureau of Transportation Statistics and DC Office of the Chief Technology Officer. (MTA,WMATA, U.S. BTS, FRA)
  + **Point type feature class**
  + Used for identifying commuter rail access points throughout Maryland.
* [Amtrak Rail Stops](https://opendata.maryland.gov/Transportation/MD-iMAP-Maryland-Transit-Amtrak-Rail-Stops/yiyd-dnxa/about_data) provided by MTA, U.S. BTS, FRA
  + **Point type feature class**
  + Represents intercity rail stations for national connectivity.
* [WMATA Metro Stops](https://opendata.maryland.gov/Transportation/MD-iMAP-Maryland-Transit-WMATA-Metro-Stops/5ttq-ykbc/about_data) provided by MTA, U.S. BTS, FRA
  + **Point type feature class**
  + Used to map heavy rail metro access across the DC metropolitan area
* [MTA Bus Stops](https://opendata.maryland.gov/Transportation/MD-iMAP-Maryland-Transit-MTA-Bus-Stops/j2zf-ej96/about_data) provided by MTA, U.S. BTS, FRA
  + **Point type feature class**
  + Represents statewide local and commuter bus stop locations.
* [Montgomery County Ride On Routes](https://data.imap.maryland.gov/datasets/maryland::maryland-local-transit-montgomery-county-ride-on-routes/about) provided by MTA, U.S. BTS, FRA
  + **Point type feature class**
  + This layer shows all Montgomery County Ride On bus routes including all variations. The information was retrieved from January 2016 MCRO GTFS files. The GTFS file was converted to ArcGIS format by CMRT.
  + Includes stops along local bus routes specific to Montgomery County.
* [Maryland Census Boundaries - Census Tracts 2020](https://data.imap.maryland.gov/datasets/maryland::maryland-census-boundaries-census-tracts-2020/explore?location=39.011405%2C-77.142732%2C10.51)
  + Polygon type feature class
  + MD iMAP Data Catalog (DoIT)

Study area: Montgomery County  
  
Montgomery County was chosen due to its steady suburban growth, proximity to Washington, D.C., and active efforts in transit-oriented development. Its mix of new and old neighborhoods makes it ideal for studying transit access in emerging residential areas.

This project uses 500 meters for bus stops and 1609 meters/ 1Mile for rail stations (MARC, WMATA, Amtrak) to evaluate transit availability. These distances align with standard walkability guidelines used in urban and suburban planning.

Q1. Where are the newly developed parcels (post-2020) located in Montgomery County? And What proportion of these new parcels are residential?

CREATE OR REPLACE VIEW ACPLAN."V\_NEW\_DEV" AS

SELECT \* FROM acplan."PARCELS"

WHERE yr\_built >= 2020 AND lu\_code\_flag = 1;

SELECT COUNT(\*) FILTER (WHERE lu\_code\_flag = 1)::FLOAT / COUNT(\*) \* 100 AS residential\_pct

FROM acplan."PARCELS"

WHERE yr\_built >= 2020;

Parcels built after 2020 in Montgomery County were queried, and the residential proportion was determined using the lu\_code\_flag. Results showed that approximately 95% of newly developed parcels are residential, indicating that recent development in the county is primarily focused on housing.

Q2. How many newly developed residential parcels are within 500 meters of a bus stop?

CREATE OR REPLACE VIEW ACPLAN."V\_RESI\_BUS\_ACCESS" AS

SELECT

p.account\_id,

p.geom::geometry(Point, 26985) AS geom,

CASE

WHEN b.account\_id IS NOT NULL THEN 1

ELSE 0

END AS bus\_access\_flag

FROM ACPLAN."V\_NEW\_DEV" p

LEFT JOIN (

SELECT DISTINCT p.account\_id

FROM ACPLAN."V\_NEW\_DEV" p

JOIN trans."MTABUS" mta ON ST\_DWithin(p.geom, mta.geom, 500)

UNION

SELECT DISTINCT p.account\_id

FROM acplan."V\_NEW\_DEV" p

JOIN trans."RIDEON" r ON ST\_DWithin(p.geom, r.geom, 500)

) b ON p.account\_id = b.account\_id;



The SQL view acplan.V\_RESI\_BUS\_ACCESS is created by spatially joining newly developed residential parcels (acplan.v\_new\_dev) with bus stop datasets (trans.MTABUS and trans.RIDEON). For each parcel, the ST\_DWithin() function checks if there’s at least one bus stop within 500 meters. If so, the bus\_access\_flag is set to 1; otherwise, it's set to 0.

* 4,672 parcels (≈ 73%)have access to a nearby bus stop.
* 1,704 parcels (≈ 27%) do not have nearby bus access.

Q3. How many newly developed residential parcels are within 1000 meters of a train station?

CREATE OR REPLACE VIEW ACPLAN."V\_RESI\_TRAIN\_ACCESS" AS

SELECT

p.account\_id,

p.geom::geometry(Point, 26985) AS geom,

CASE

WHEN EXISTS (

SELECT 1 FROM (

SELECT geom FROM trans."MARC"

UNION ALL

SELECT geom FROM trans."AMTRAK"

UNION ALL

SELECT geom FROM trans."WMATA"

) AS trains

WHERE ST\_DWithin(p.geom, trains.geom, 1609)

)

THEN 1

ELSE 0

END AS train\_access\_flag

FROM acplan."V\_NEW\_DEV" AS p;

The view acplan.V\_RESI\_TRAIN\_ACCESS was created to identify whether each newly developed residential parcel in Montgomery County (from acplan.v\_new\_dev) is within **1609 meters** of any train station. This includes MARC, Amtrak, and WMATA stations, which were combined using a UNION ALL to create a single collection of rail transit points.

Using ST\_DWithin(), the query checks the proximity between each parcel and the nearest train station. A flag (train\_access\_flag) is added to each parcel:

A count of these flags shows:

* 1,274 parcels (≈20%) have nearby train access,
* 5,102 parcels (≈80%) do not have nearby train access.

Q4. Which newly developed residential parcels in Montgomery County since 2020 have no access to any transit facilities within a reasonable walking distance (500 meters for bus, 1 mile for train)

CREATE OR REPLACE VIEW ACPLAN."V\_RESI\_NO\_TRANSIT\_ACCESS" AS

SELECT

p.account\_id,

p.geom

FROM ACPLAN."V\_NEW\_DEV" p

WHERE NOT EXISTS (

-- Nearby bus stops (within 500 meters)

SELECT 1

FROM trans."MTABUS" b

WHERE ST\_DWithin(p.geom, b.geom, 500)

UNION

SELECT 1

FROM trans."RIDEON" r

WHERE ST\_DWithin(p.geom, r.geom, 500)

)

AND NOT EXISTS (

-- Nearby train stations (within 1 mile = 1609 meters)

SELECT 1

FROM (

SELECT geom FROM trans."MARC"

UNION ALL

SELECT geom FROM trans."AMTRAK"

UNION ALL

SELECT geom FROM trans."WMATA"

) AS trains

WHERE ST\_DWithin(p.geom, trains.geom, 1609)

);



This view identifies newly developed residential parcels in Montgomery County that lack nearby access to public transit. It selects parcels from the V\_NEW\_DEV view and filters those that do not fall within 500 meters of any MTA or RideOn bus stop and are also more than one mile (1,609 meters) away from any MARC, AMTRAK, or WMATA train station. The resulting output highlights parcels that are entirely disconnected from both local and regional transit services, helping to pinpoint areas with potential gaps in public transportation coverage.  
  
Parcels with Bus Access: 4,672  
Total New Developments: 6,376

Q5. Which newly developed residential parcels in Montgomery County since 2020 are located more than 5 miles away from the nearest train station (MARC, AMTRAK, or WMATA)?

-- residential areas that do not have transit stations within 5 miles

CREATE OR REPLACE VIEW ACPLAN."V\_NOTRAINSTATION\_5MI" AS

SELECT p.account\_id, p.geom

FROM ACPLAN."V\_NEW\_DEV" p

WHERE NOT EXISTS(

SELECT 1 FROM(

SELECT geom FROM TRANS."MARC"

UNION ALL

SELECT geom FROM TRANS."AMTRAK"

UNION ALL

SELECT geom FROM TRANS."WMATA"

) AS all\_train\_transit

WHERE ST\_Dwithin(p.geom, all\_train\_transit.geom, 8046.72) -- 5 miles in meters

This SQL view identifies newly developed residential parcels (built since 2020) in Montgomery County that are located more than 5 miles (8046.72 meters) from the nearest train station, including MARC, AMTRAK, and WMATA. It uses ST\_DWithin to check for proximity and NOT EXISTS to filter out parcels that do not fall within that 5-mile radius of any transit point. The result highlights areas that may face challenges in regional rail connectivity and can support planning decisions around infrastructure gaps.

Parcels Beyond 5 Miles from Train Access: 457

Q6. Which census tracts in Montgomery County have the highest and lowest percentages of new residential parcels with access to bus transit within 500 meters?

-- Percentage of tracts with parcels access to bus facilities

SET ROLE ACPLAN;

CREATE OR REPLACE VIEW ACPLAN."V\_TRACT\_BUS\_ACCESS\_PCT" AS

SELECT

t.tractfips,

COUNT(p.account\_id) AS total\_parcels,

SUM(p.bus\_access\_flag) AS parcels\_with\_bus\_access,

ROUND(SUM(p.bus\_access\_flag)::numeric / COUNT(p.account\_id) \* 100, 2) AS percent\_bus\_access,

t.geom

FROM bndy."TRACT\_BOUNDARY" t

JOIN ACPLAN."V\_RESI\_BUS\_ACCESS" p

ON ST\_Within(p.geom, t.geom)

GROUP BY t.tractfips, t.geom;

This SQL view summarizes the accessibility of newly developed residential parcels to bus transit at the census tract level. By spatially joining new development points with tract boundaries, it calculates the total number of new parcels within each tract and how many of those are within 500 meters of a bus stop. The result includes the percentage of parcels with bus access per tract, enabling a clear understanding of how well different parts of the county are served by local and regional bus services.

Q6. Percentage of Tracts with Access to Train Transit Facilities

-- Percentage of tracts with parcels access to train facilities

CREATE OR REPLACE VIEW ACPLAN."V\_TRACT\_TRAIN\_ACCESS\_PCT" AS

SELECT

t.tractfips,

COUNT(p.account\_id) AS total\_parcels,

SUM(p.train\_access\_flag) AS parcels\_with\_bus\_access,

ROUND(SUM(p.train\_access\_flag)::numeric / COUNT(p.account\_id) \* 100, 2) AS percent\_train\_access,

t.geom

FROM bndy."TRACT\_BOUNDARY" t

JOIN ACPLAN."V\_RESI\_TRAIN\_ACCESS" p

ON ST\_Within(p.geom, t.geom)

GROUP BY t.tractfips, t.geom;

This query creates a spatial view that calculates the percentage of newly developed residential parcels (since 2020) within each census tract that have access to train transit stations (MARC, AMTRAK, WMATA) within 1 mile. It joins parcel points with tract polygons, summarizes the count of parcels with train access per tract, and computes the percentage to support spatial visualization of transit equity across the county.

Parcels with Train Access: 1,274  
Total New Developments: 6,376

Q7. Census Tracts with No Transit Access for New Residential Parcels ?

-- Find the tracts with parcels with no access to transit facilities

CREATE OR REPLACE VIEW ACPLAN."V\_TRACTS\_WITH\_NO\_TRANSIT" AS

SELECT

t.tractfips,

t.countyfips,

COUNT(p.account\_id) AS no\_transit\_parcels,

ROUND(COUNT(p.account\_id)::numeric \* 100 / 6376, 2) AS pct\_no\_transit,

t.geom

FROM BNDY."TRACT\_BOUNDARY" t

JOIN ACPLAN."V\_RESI\_NO\_TRANSIT\_ACCESS" p

ON ST\_Within(p.geom, t.geom)

GROUP BY t.tractfips, t.countyfips, t.geom;

This view identifies census tracts where newly developed residential parcels (since 2020) lack access to both bus and train transit services. It joins the no-access parcel data with tract boundaries and counts how many such parcels fall within each tract. The percentage is calculated against the total number of new developments (6,376), allowing for clear spatial representation of underserved areas in terms of public transit.  
  
Parcels Without Any Transit Access: **1,653**

Recommendations:

* The analysis revealed that only 20% of newly developed residential parcels have train access within 1 mile (1,609 meters). According to transit planning literature, a 1,000-meter threshold is considered more appropriate for walkable access.
* This indicates that even with a generous buffer, many parcels still fall outside accessible train ranges, emphasizing the need for expanded or new train infrastructure in areas of recent development.
* Local planners should consider enhancing multimodal connectivity, especially in northern and western Montgomery County where transit access is limited.
* The disparity in access should guide equity-focused investments in public transportation to better serve newly developed communities.

Sources:  
https://www.researchgate.net/figure/Mechanisms-by-which-transit-use-and-automobile-use-are-associated-with-active\_fig3\_277842737

https://montgomeryhistory.org/mary-kay-harper-center-for-suburban-studies/css-suburbanization-overview/#:~:text=Proximity%20to%20Washington%2C%20D,county%20%26%20local%20government%20leaders

#### **1. Topical Area: Transportation Planning**

Transit Access in New Residential Areas: Identifying Gaps in Public Transport Coverage

As cities grow and new housing is built, public transit often doesn't expand at the same pace. This can leave new neighborhoods without convenient or reliable access to transportation. My project focuses on identifying where these service gaps exist, using spatial analysis to support better planning and decision-making. The work aligns with goals from Maryland’s Statewide Transit Plan to improve accessibility and equity in transit.

### **2. Data Sources**

### [Maryland Department of Transportation (MDOT)](https://www.mdot.maryland.gov/)

* Provides transit stop and route data. Official source; reliable for transportation networks. May require format conversion (e.g., GTFS to shapefile).
* Authoritative source for statewide GIS data. Quality is good, but update frequency varies by dataset.

### [Maryland iMAP GIS Data Portal](https://imap.maryland.gov/)

* It offers statewide GIS layers, including zoning, land use, and administrative boundaries. Data is authoritative; update frequency may vary by dataset.
* Official and reliable. May require conversion from GTFS format. Some datasets may not reflect recent service changes.

### [Local Government Open Data Portals(Open Baltimore)](https://data.baltimorecity.gov/datasets/baltimore::vacant-building-notices/explore)

* Provide parcel and building data, including attributes like construction year. Useful for identifying new residential development. Data quality and update frequency may vary.
* Accurate for identifying new developments. Data quality depends on local update schedules. Attribute formats may vary.

### [U.S. Census Bureau (ACS 5-Year Estimates)](https://data.census.gov/)

* Source of demographic data such as population, housing, and income. Highly reliable and regularly updated; limited detail in small geographic units.
* Census data is reliable for population insights but it may not offer detailed insights at smaller geographic levels like neighborhoods.

### [OpenStreetMap via Geofabrik](https://download.geofabrik.de/north-america/us/maryland.html)

* Offers open-source data for roads, land use, and points of interest of regional data. Good for supplemental analysis; community-maintained, so accuracy may vary.
* It is helpful for adding roads and land use data when official sources are missing. Since it's community-sourced, coverage can be inconsistent in some areas.

### **3. Documentation and Normalization**

Data model development

* Conceptual Model: Identify residential areas and transit stops. Analyze access based on distance between them.
* Logical Model: Tables for parcels (parcel\_id, year\_built, geom) and transit stops (stop\_id, stop\_name, geom). Use functions like ST\_DWithin to get within data.
* Physical Model: Implement in PostGIS. Import data as point/polygon geometries

Metadata will include title, source, date, coordinate system, and field descriptions. FGDC or ISO 19115 format will be followed using ArcGIS Pro tools.

SOP considerations for data updates, maintenance, and access.

* Check for data updates quarterly. Store in PostGIS with standard naming. Limit edit access; log all changes.

### **4. Database Development**

The database will include the following key tables:

* residential\_areas – new or recently built housing units
* transit\_stops – public transit locations (bus, rail, etc.)s
* access\_results – calculated access data (e.g., distance, accessibility )

Types of spatial and non-spatial attributes.

* Spatial attributes:  
   geometry fields (Point or Polygon) in residential\_areas and transit\_stops
* Non-spatial attributes:  
   Parcel ID, year built, stop name, stop ID, distance in meters, boolean for accessibility
* Expected relationships and integrity constraints (primary and foreign keys).
  + Primary Keys: parcel\_id (residential\_areas), stop\_id (transit\_stops)
  + Foreign Keys: parcel\_id and stop\_id in access\_results
  + Relationships based on spatial joins using ST\_DWithin and ST\_Distance
  + Enforce referential integrity through foreign key constraints in access\_results

### **5. Analysis & Visualization**

* Propose initial spatial analysis ideas or questions your project aims to answer.
  + Which new residential areas are within 500 meters of a public transit stop?
  + Which areas are outside this range and considered underserved?
  + What percentage of new housing has access to transit?
  + Where should new transit stops be prioritized based on service gaps?
* Describe how you plan to visualize results (e.g., ArcGIS Pro).
  + Buffer maps showing service areas around transit stop
  + Symbolized residential points (served vs underserved)
  + Thematic maps highlighting coverage gaps
  + Labeling or charts to show statistics

### **6. Organizational Context**

* Role: Lead data analyst and database administrator.
* Organization: Government transportation planning agency.
* Purpose and Impact: The analysis will support transit planning by identifying service gaps in new residential areas, helping guide decisions on where to expand public transportation infrastructure.

### 7. Workplan

|  |  |  |
| --- | --- | --- |
| Current Schedule | Focus | Planned Work |
| Week 8 | Spatial Functions in POSTGIS | Finalize datasets, load into PostGIS, apply SRID, and test spatial functions. |
| Week 9 | Spatial Query Development | Write and test queries to check transit access using ST\_DWithin, ST\_Intersects. |
| Week 10 | Refine Queries | Clean up and finalize spatial queries. Validate results with sample cases. |
| Week 11 | Final Spatial Queries | Complete all queries and prepare access results (e.g., served vs underserved areas). |
| Week 12 | Database Structure and Access | Review table structure, define keys, and apply basic user access if needed. |
| Week 13 | Map and Output Preparation | Finalize visual outputs in ArcGIS Pro. Create maps for analysis results. |
| Week 14 | Project Documentation | Complete metadata, SOPs, and organize workflow and deliverables. |
| Week 15 | Final Presentation | Present final project, spatial queries, maps, and submit all deliverables. |

**Notes**  
Improved value or new and improved structured value filter above $10000

Specific counties  
Arundale,