Project Proposal and EDA

Gunpreet Singh, Gurleen Pelia and Mariyam Ahmed

Urban Sprawl using Land Cover Data of Canada

Purpose of Project

Land Cover Data : Canada's land cover data is important for several reasons. By offering comprehensive information about ecosystems and forest cover, it plays a crucial role in environmental management and conservation, enabling the preservation of biodiversity and sustainable forest management. Studies on climate change require these data in order to evaluate the effects on ecosystems, vegetation patterns, and carbon sequestration. Land cover data helps manage the risk of disasters, identify locations that are ideal for development while maintaining green spaces, and promote sustainable growth in urban planning and development. All things considered, land cover data is critical to comprehending Canada's landscapes, aiding in conservation efforts, and advancing sustainable development in a number of industries.

Land cover data and urban sprawl are intricately related in thus project that involve studying urban expansion and its impacts on the environment. Understanding this relationship is crucial for analyzing patterns, predicting future changes, and planning sustainable urban development.

Urban Sprawl : Urban sprawl refers to the unplanned and uncontrolled growth of urban areas into neighboring rural areas, which leads to dispersed, low-density development. Longer commutes, a greater reliance on cars, and higher traffic congestion are the results of this phenomenon. It destroys open areas, farms, and natural habitats, upsetting biodiversity and local ecosystems. Infrastructure is also strained by urban sprawl, necessitating large-scale, expensive expenditures in public utilities, roads, and other services. It also adds to environmental problems including pollution and higher emissions of greenhouse gases. Overall, urban sprawl poses serious obstacles to environmental preservation and sustainable growth.

How Land Cover Data Helps Study Urban Sprawl?

1. Identifying Urban Areas:

- Land cover data allows you to identify and delineate urban areas from other land cover types.
- By classifying satellite images, you can determine the extent of urbanization.

2. Temporal Analysis:

- By comparing land cover data from different time periods, you can analyze changes in land use and identify trends in urban expansion.
- This helps in understanding the rate and pattern of urban sprawl.

3. Change Detection:

- Land cover change detection techniques enable you to quantify the conversion of non-urban areas to urban areas.

- This information is crucial for assessing the impact of urban sprawl on natural resources and agricultural land.

4. Modeling and Prediction:

- Using historical land cover data, you can develop models (like CA-Markov) to predict future urban growth.
- This helps in planning and implementing sustainable urban development strategies.

5. Impact Assessment:

- Analyzing land cover changes allows you to assess the environmental impacts of urban sprawl, such as deforestation, loss of biodiversity, and soil degradation.

About the dataset

The CCRS Land Cover dataset is crucial for our project due to its high accuracy, detailed resolution, and comprehensive national coverage. This dataset's 30-meter spatial resolution provides precise and reliable land cover information, essential for a wide range of environmental applications such as climate impact studies, emergency response planning, and wildlife habitat analysis. The dataset spans multiple years (2010, 2015, and 2020), enabling temporal analysis of land cover changes over time. Additionally, its integration into the North American Land Change Monitoring System (NALCMS) underscores its credibility and relevance for both national and regional studies. The solid and accurate data foundation provided by the CCRS Land Cover dataset guarantees that our project is founded on a dependable resource, improving the caliber and significance of our research and assisting with well-informed environmental management and policy decisions.

Description of Data Sources

Primary Data Source:

- Source Name: Canada Centre for Remote Sensing (CCRS) Land Cover Map
- Description: This dataset provides detailed land cover information for Canada at a 30 m spatial resolution. It is produced using observations from the Operational Land Imager (OLI) Landsat sensor and covers the years 2010, 2015, and 2020. The data is essential for a wide range of environmental applications, including climate impact studies, emergency response, and wildlife habitat analysis.
- Citations: Canada Centre for Remote Sensing. "Land Cover Map of Canada (2010, 2015, 2020)." Produced under the North American Land Change Monitoring System (NALCMS).

Secondary Data Source:

- Source Name: National Census Data (2023)
- Description: This dataset includes demographic information for residents of Canada, such as age distribution, household income, and population density. It is used to analyze the relationship between land cover and population demographics.
- Citations: Government of Canada. "National Census Data, 2023." https://www.census.gc.ca

Collection of Land Cover products for Canada as produced by Natural Resources Canada using Landsat satellite imagery. This collection of cartographic products offers classified Land Cover of Canada at a 30 metre scale, updated on a 5 year basis.

- Landcover of Canada 2010
- Landcover of Canada 2015
- Landcover of Canada 2020

Data Pre-Processing

1.Loading and Preparing the Raster Data:

- High-Resolution Data for Canada: Initially, you worked with high-resolution raster data that covered the entire country of Canada. This dataset likely contained detailed land use or land cover information across the country. *Cropping to Fit Ontario: Since the analysis was focused on Ontario, you cropped the Canada-wide raster dataset to fit the geographic boundaries of Ontario. Cropping helps in reducing the size of the data and focusing the analysis on the region of interest.

2.Downsampling Using the Aggregate Function:

- Aggregate Function: To manage the data size and make it more computationally feasible, you downsampled the raster data using the aggregate function. Downsampling reduces the resolution of the raster by combining several adjacent pixels into a single pixel, thereby reducing the data size while preserving the overall spatial patterns.
- Process: The aggregate function groups adjacent cells and calculates a summary statistic (such as mean, sum, or majority) for the new, larger cells. This process retains the essential spatial information but with less detail.

3.Setting the Coordinate Reference System (CRS):

- CRS for Bounding Box of Canada: The CRS defines how the two-dimensional, projected map in your GIS relates to real places on the earth. You set the CRS for the bounding box of Canada using the st_transform function. This ensures that the spatial data aligns correctly with geographic coordinates.
- Transform Function: The st_transform function from the sf package is used to transform the spatial coordinates of the dataset to a specified CRS. This is crucial for spatial operations like cropping and overlaying different spatial datasets.

4.Collecting Data for Toronto Boundaries Using OpenStreetMap (OSM):

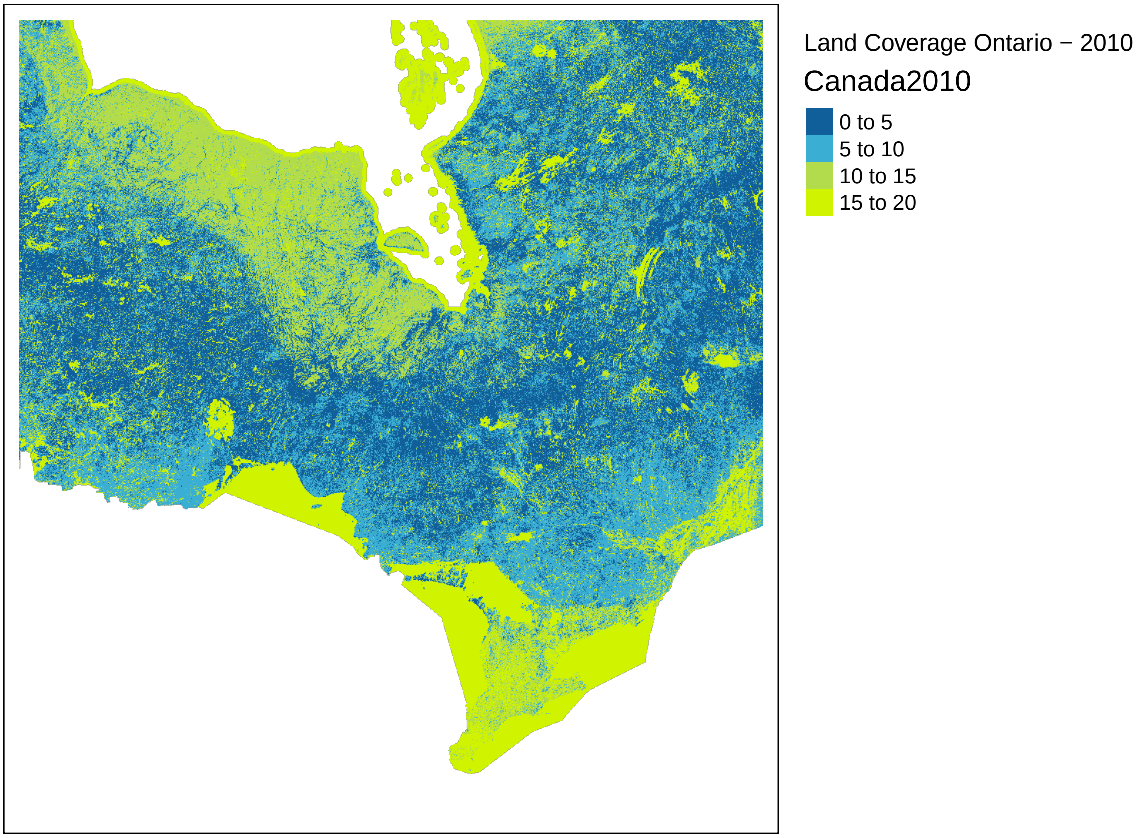
- OpenStreetMap (OSM): OSM is a collaborative project that creates a free, editable map of the world. You used OSM to gather the boundary data for Toronto.
- Key-Value Pair for Toronto: In OSM, geographical features are described using key-value pairs. You likely used a query with a specific key (e.g., boundary) and value (e.g., Toronto) to extract the boundary data for Toronto.
- Data Retrieval: The boundary data for Toronto was retrieved and converted into a spatial object that could be used for further geospatial operations.

5.Cropping and Masking the Ontario Land Use Data:

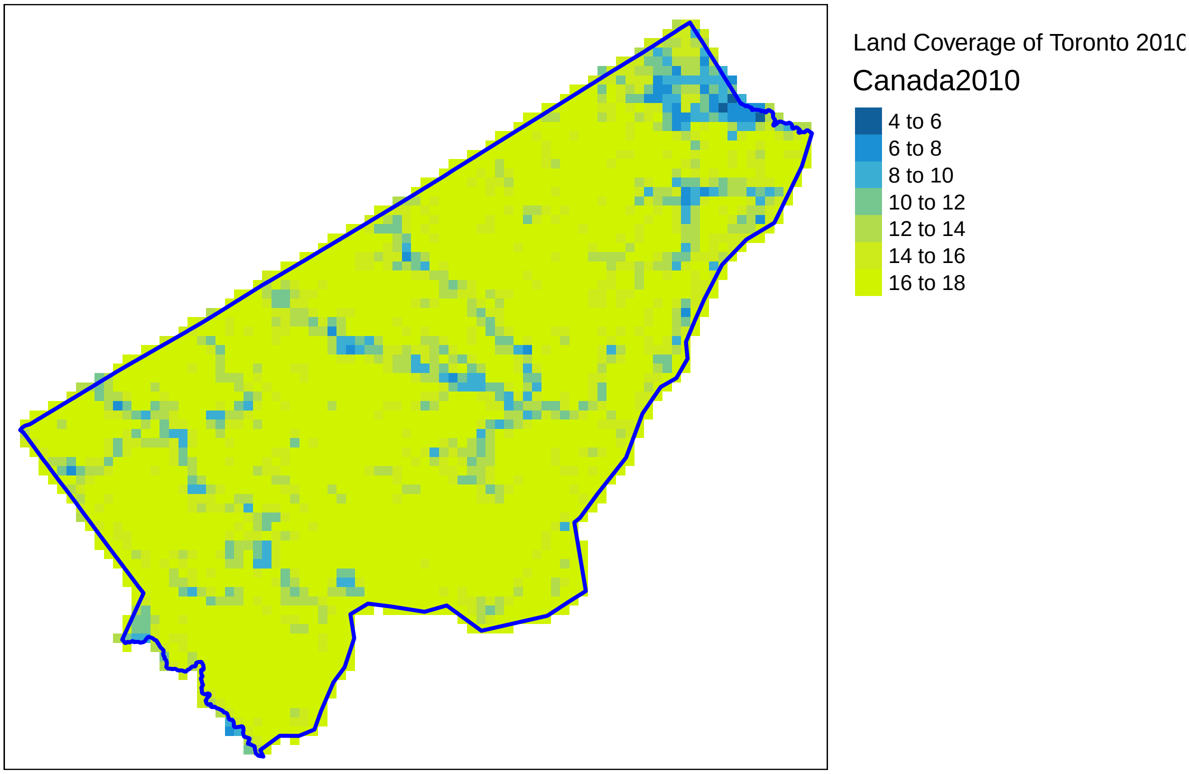
- Cropping: The Toronto boundary data obtained from OSM was used to crop the downsampled Ontario raster data. Cropping confines the raster data to the area within the Toronto boundaries, removing data outside this area.
- Masking: Masking further refines the raster data by setting values outside the specified mask (in this case, the Toronto boundary) to NA or some other value indicating absence of data. This ensures that the analysis is strictly limited to the area of interest.
- Process: The combination of cropping and masking results in a dataset that contains detailed land use information specifically for the Toronto area, extracted from the larger Ontario dataset.

Visualizations

```
## stars object downsampled to 967 by 1035 cells. See tm_shape manual (argument raster.downsample)
```



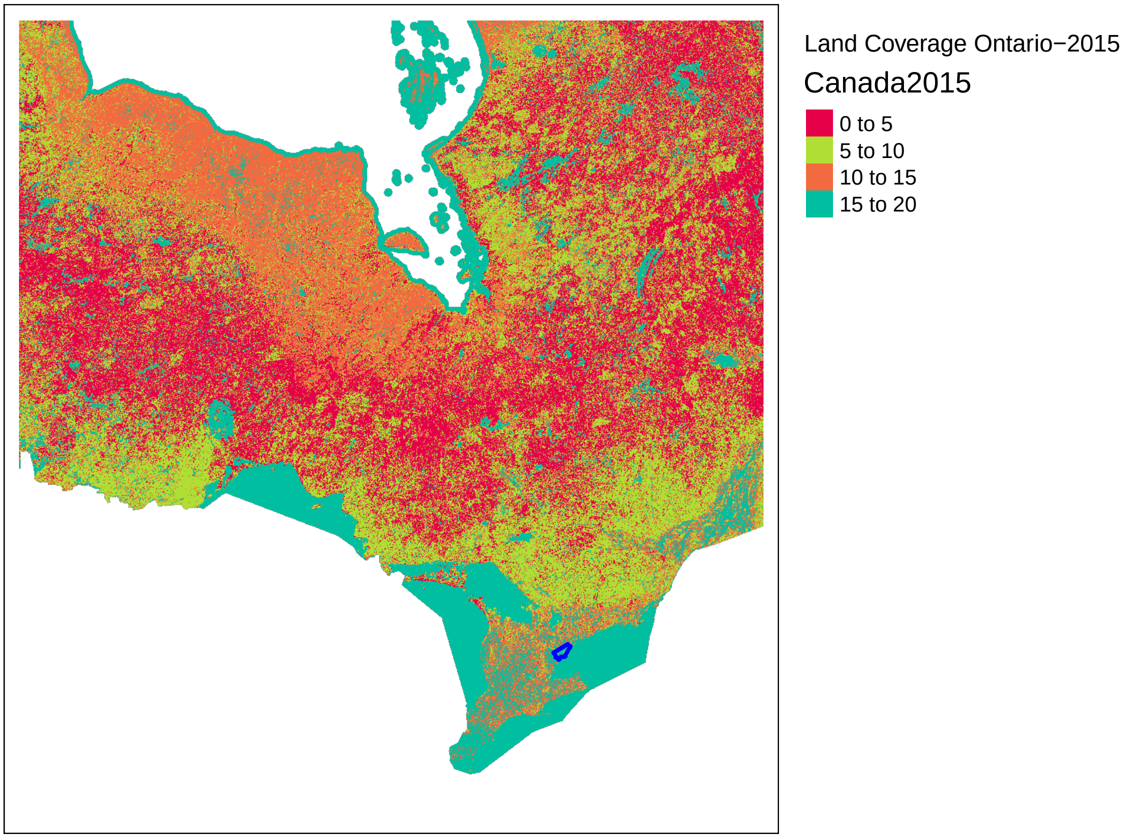
This map shows land coverage in Ontario, Canada, for the year 2010, using a blue-to-yellow color palette to represent different coverage ranges. Dark blue indicates areas with low coverage (0-5), while yellow represents areas with high coverage (15-20). The detailed legend and color scheme highlight spatial variations, with the southern part of Ontario, around the Great Lakes, shown in more detail due to its higher population density and diverse land use patterns. The map serves as a useful tool for environmental analysis, urban planning, and research by providing a clear visual representation of land cover distribution across the region.



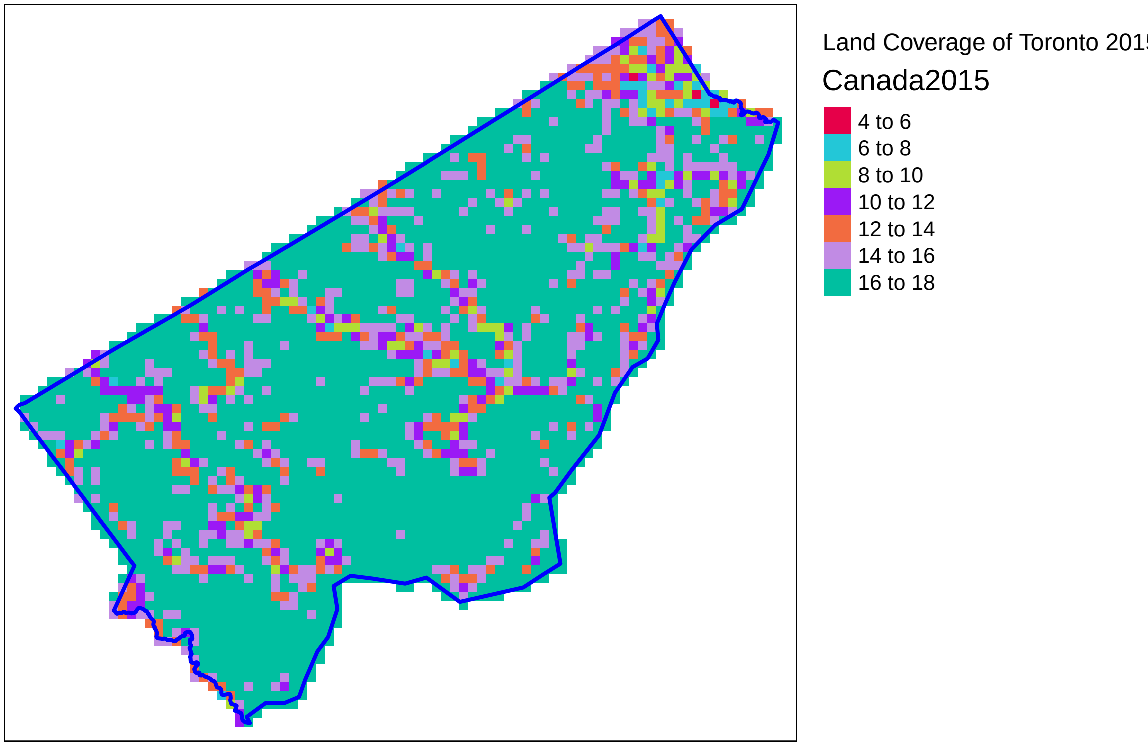
This map displays land coverage in Toronto, Canada, for the year 2010, using a blue-to-yellow color gradient. Dark blue represents areas with lower land coverage values (4 to 6), while light yellow indicates higher coverage values (16 to 18). The map highlights spatial variations across Toronto, with detailed insights into specific regions. Blue areas likely correspond to urban spaces or water bodies, whereas yellow areas indicate dense vegetation or green spaces. This visualization aids urban planners, environmental analysts, and researchers in understanding land use patterns for informed decision-making.

2015

```
## stars object downsampled to 967 by 1035 cells. See tm_shape manual (argument raster.downsample)
```



This map illustrates land coverage in Ontario, Canada, for the year 2015, using a color palette ranging from red to green. The legend indicates that red represents areas with the lowest land coverage values (0 to 5), green represents moderate coverage values (5 to 10 and 10 to 15), and blue-green represents the highest coverage values (15 to 20). The map provides detailed spatial variations across Ontario, highlighting regions with different land cover densities. Red areas likely correspond to urban or less vegetated regions, while blue-green areas indicate dense vegetation or green spaces. This visualization aids urban planners, environmental analysts, and researchers in understanding land use patterns and making informed decisions.

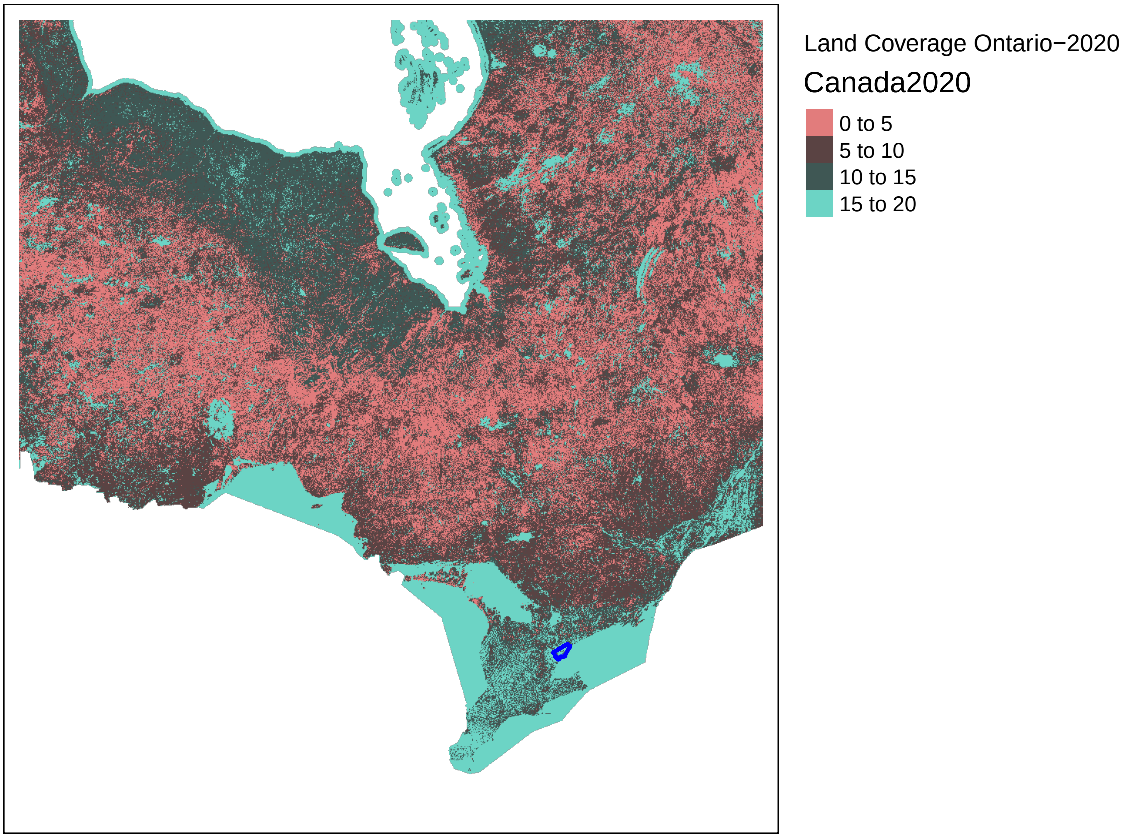


This map depicts land coverage in Toronto, Canada, for the year 2015, using a varied color palette to represent different ranges of land coverage values. The legend indicates:

Red: 4 to 6 Light blue: 6 to 8 Purple: 8 to 10 Orange: 10 to 12 Lavender: 12 to 14 Green: 14 to 16 Teal: 16 to 18 The color scheme helps to identify spatial variations in land coverage across Toronto. Lower values, shown in red and light blue, likely correspond to urban areas or less vegetated spaces. Higher values, depicted in green and teal, indicate areas with dense vegetation or green spaces. This map is useful for urban planners, environmental analysts, and researchers, providing a visual tool for understanding land use patterns and making informed decisions.

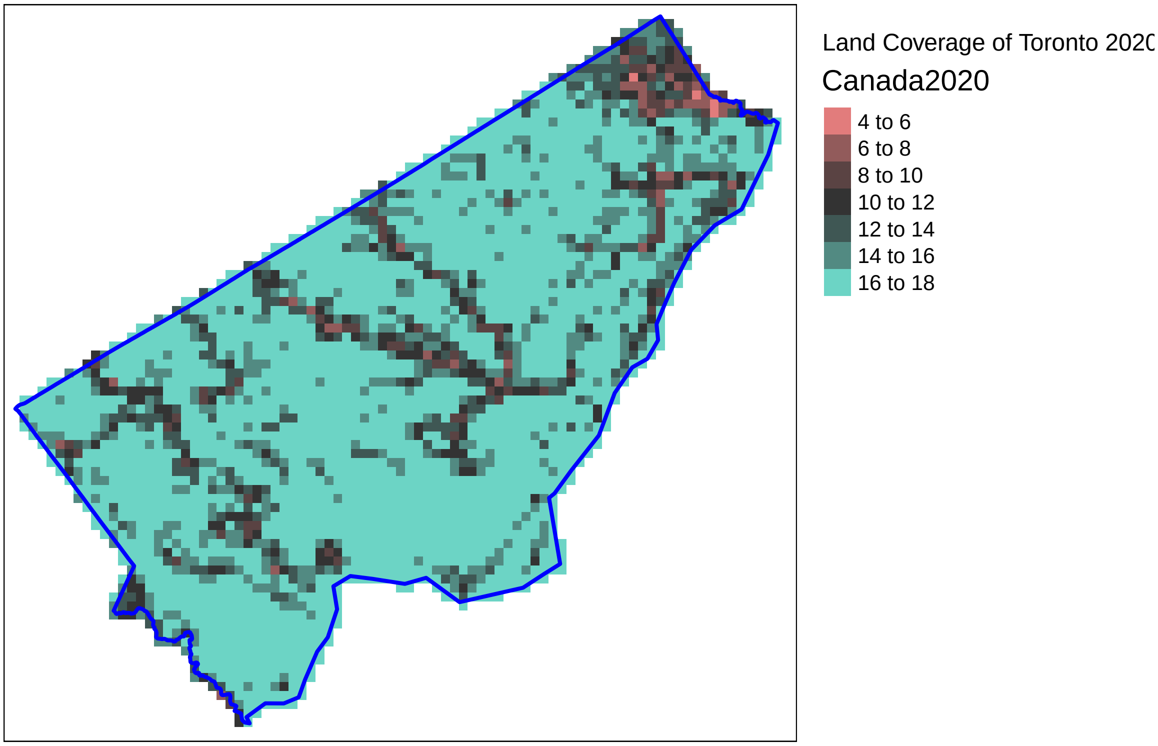
2020

```
## stars object downsampled to 967 by 1035 cells. See tm_shape manual (argument raster.downsample)
```



This map presents land coverage in Ontario, Canada, for the year 2020, using a color scheme that ranges from red to teal. The legend specifies the land coverage values associated with each color:

Red (0 to 5) Dark gray (5 to 10) Dark teal (10 to 15) Light teal (15 to 20) The map highlights spatial variations across Ontario, with red areas representing regions with the lowest land coverage values, likely corresponding to urban areas or regions with less vegetation. The teal areas denote regions with the highest land coverage values, indicative of dense vegetation or green spaces. This visualization is valuable for urban planners, environmental analysts, and researchers, as it provides a clear visual representation of land use patterns and aids in informed decision-making related to environmental management and urban development.



This map displays land coverage in Toronto, Canada, for the year 2020, using a color scheme from red to teal. Red areas represent the lowest land coverage values (4 to 6), various shades of gray indicate moderate values (6 to 12), and teal areas show the highest values (12 to 18). The map highlights spatial variations in land coverage across Toronto, with lower values likely corresponding to urban areas and higher values indicating dense vegetation or green spaces. This visualization aids urban planners, environmental analysts, and researchers in understanding land use patterns for informed decision-making.

Analysis Methods/Modelling

We are planning to use A-Markov modeling by taking reference from 6 different research paper which are mentioned below. This model is a powerful method for simulating and predicting land use changes over time by combining the strengths of Cellular Automata (CA) and Markov Chains.

Cellular Automata (CA) : Cellular Automata are dynamic models that simulate the evolution of spatial patterns over time. They work on a grid of cells, where each cell can be in one of a finite number of states (e.g., different land cover types). The state of each cell at the next time step is determined by a set of rules that consider the current state of the cell and the states of its neighboring cells.

Mathematical Formulation:

- **State Space:** $S = \{s_1, s_2, ..., s_n\}$
  Each cell c can be in one of the states s_i at time t.
- **Neighborhood:** The neighborhood $N(c)$ of cell c includes the cells that influence the state of c.
- **Transition Rules:** The transition rules T determine the state of a cell based on the states of its neighbors:
$$s_c(t+1) = T(s_c(t), N(s_c(t)))$$

Markov Chains: Markov Chains are stochastic models that describe the probability of transitioning from one state to another over time. They are used to model the likelihood of land cover change based on historical data. The transition probabilities are typically derived from observed changes in land cover over a specified period.

Mathematical Formulation:

- **State Vector: $\mathbf{p}(t)$**

  A vector representing the probability distribution of the states at time t.

- **Transition Matrix: $\mathbf{P}$**

  A matrix where P_{ij} represents the probability of transitioning from state i to state j:

$$\mathbf{P} = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{pmatrix}$$

- **Next State Probability:** The state vector at time $t+1$ is obtained by multiplying the state vector at time t with the transition matrix:

$$\mathbf{p}(t+1) = \mathbf{p}(t)\mathbf{P}$$

CA-Markov Model: The CA-Markov model integrates Cellular Automata and Markov Chains to predict future land cover changes. The Markov Chain component provides the transition probabilities, while the Cellular Automata component spatially allocates these changes based on local neighborhood rules.

Mathematical Formulation:

1. **Initial State:** Start with an initial land cover map S_0.
2. **Transition Matrix:** Compute the transition matrix $\mathbf{P}$ from historical land cover maps.
3. **Neighborhood Influence:** Apply the CA rules to incorporate spatial dependencies.

The state of a cell at time $t+1$ depends on:

- The transition probabilities from the Markov Chain.
- The neighborhood configuration from the CA rules.

Steps to Implement CA-Markov Modeling

1. Data Preparation:

- Collect temporal and spatial land cover data
- Preprocess the data

2. Land Cover Classification:

- Classify the land cover data into different categories

3. Transition Probability Matrix:

- Calculate the transition probabilities using historical land cover data.

4. Define CA Rules:

- Establishe rules that determine how cells change state based on their neighbors.

5. Model Calibration and Validation:

- Validate the model using a known time period to ensure accuracy.
- Adjust parameters to improve the model's predictive capability.

6. Prediction:

- Use the calibrated model to predict future land cover changes

References

1. Shikhar Deep, Akansha Saklani. (2014). "Urban sprawl modeling using cellular automata." The Egyptian Journal of Remote Sensing and Space Sciences, 17, 179–187. DOI: 10.1016/j.ejrs.2014.07.001.

2. Mussie G. Tewolde, Pedro Cabral. (2011). "Urban Sprawl Analysis and Modeling in Asmara, Eritrea." Remote Sens., 3, 2148-2165. DOI: 10.3390/rs3102148.

3. Parvinnezhad, Davoud, et al. (2021). "A modified spatial entropy for urban sprawl assessment." Geocarto International, 36(16), 1804-1819.

4. Zeng, Chen, et al. (2015). "Characterization and spatial modeling of urban sprawl in the Wuhan Metropolitan Area, China." International Journal of Applied Earth Observation and Geoinformation, 34, 10-24.

5. Chen, Danling, et al. (2021). "How urban sprawl influences eco-environmental quality: Empirical research in China by using the Spatial Durbin model." Ecological Indicators, 131, 108113.

6. Baqa, Muhammad Fahad, et al. (2021). "Monitoring and modeling the patterns and trends of urban growth using urban sprawl matrix and CA-Markov model: A case study of Karachi, Pakistan." Land, 10(7), 700.