TESLC Dataset: Comprehensive Analysis of Factors Affecting City Carbon Emissions

Miles Hua
Palo Alto Senior High School
Palo Alto, CA 94301
mileshua06@gmail.com

Abstract

Climate change poses a serious threat to future human conditions. Understanding the causes of air pollution is crucial for developing effective strategies to mitigate its harmful effects. To help understand the causes of emissions, we align nationwide datasets on **T**ransportation, carbon **E**mission, **S**olar energy, **L**and cover for major **C**ities into a combined dataset, TESLC. We select 15 major American cities with diverse demographic, geographic, and economic profiles. TESLC dataset introduces the factors that contribute to carbon emissions, from which we calculate correlations between carbon emissions of 15 cities and their corresponding transportation, land cover and solar capacity data. The results demonstrate that transportation is the major influential factor of carbon emission, therefore, strategies and policies of reducing the transportation impact should be considered and regulated.

1 Introduction

Since the Industrial Revolution, air pollution, primarily from carbon emissions, has accelerated climate change to unprecedented levels, creating an issue with worldwide implications. In the short term, emissions can lead to asthma and numerous other negative health effects. In the long term, emissions contributing to global warming pose a unique and significant threat. Current projections show that by the year of 2100 temperatures will rise by as much as 12 degrees Fahrenheit((USGCRP, 2014)), causing increases in natural disasters, heat strokes, rising sea levels, and many other dangerous risks. Understanding and reducing carbon emissions is crucial for ensuring long-term human success. Transitioning away from a carbon-based society may seem challenging, so this paper aims to shine a light on which activities have the greatest impact on emissions, thereby revealing the key areas to focus on to achieve a carbon-neutral United States.

We have curated a dataset TESLC, combining **T**ransportation, daily carbon **E**mission, **S**olar energy production, and **L**and cover for major **C**ities. This dataset includes transportation data about frequency of short to long trips, usage of bridges and public transportation, daily carbon emission data for thousands of cities globally, land cover data about the coverage of built-up lands, trees, and water, and solar capacity per city. Analysis of the data and their correlations demonstrates that transportation is the most influential factor in carbon emissions for major American cities. The dataset is publicly available and serves as a starting point for analyzing city carbon emissions and their causal factors.

2 TESLC dataset

We introduce TESLC, a comprehensive dataset merging the source data of carbon emission, transportation, land cover and solar capacity, as summarized in Table 1.

Preprint. Under review.

Table 1: Summary of source datasets

Dataset	Size	Year	Frequency	Source
Carbon Emission by City	1500	2019-2021	Daily	CDP
Transportation by County	3222	2015-2019	Yearly	Transportation Bureau
Land Cover	299.2 B[1]	2000-2020	16 Days	GLAD
Solar by City	20	2019-2022	Yearly	Frontier Group

Carbon Carbon data is pulled from CDP's carbon monitor cities, with daily emissions data for 1500 major cities (211 in the US) between 2019 and 2021 (Huo et al., 2022). Figure 1 demonstrates the data for Boston, representative of the rest of the data. The violin plots illustrate consistent trends among months, with winter showing the highest emissions. We select 15 major cities across America with diverse demographic and economic profiles, as shown in Figure 2, with different emission ranges.

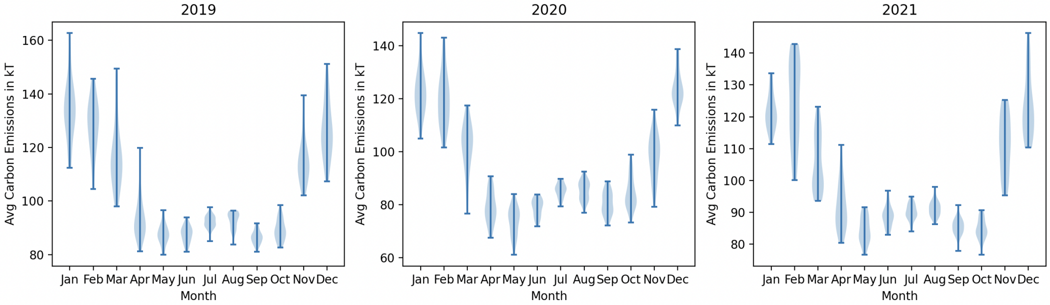

Figure 1: Violin plots for emissions in Boston for the years 2019-2021.

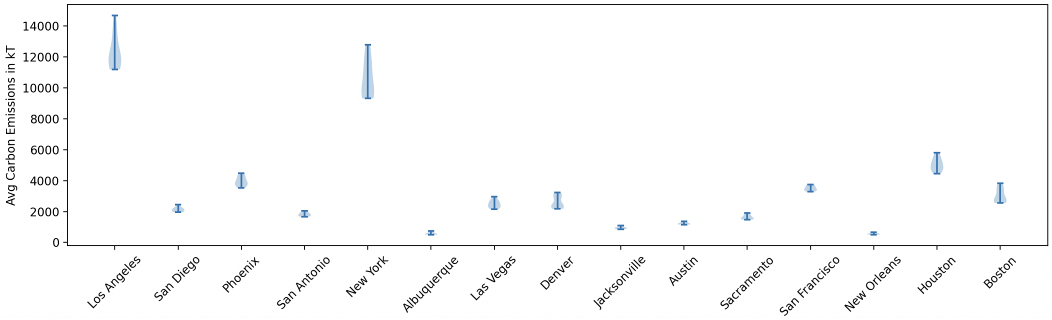

Figure 2: Violin plot of the 15 selected cities in 2019.

Transportation The Bureau of Transportation (of Transportation Statistics, 2019) presents transportation data for every county in the United States, such as the number of trips taken and the number of public transportation users. We use county data to estimate the data for major cities. For cities with multiple counties, we merge them together.

Land cover Data is pulled from the Global Land Analysis & Discovery Landsat tools (Potapov et al., 2022). The data is given in numerous tif files representing 30m by 30m regions. We obtain the shape and locations of cities(Reid, 2014) and map it to the tif, to get the area of buildings, forests, and water in each city.

[1] Calculated by the number of pixels in the tif files, which correspond to the number of 30m grids globally

Solar capacity The Frontier Group (Pforzheimer, 2020) analyzes data from the Department of Energy, providing the installed solar capacity for top cities in America.

TESLC dataset We selected 15 major U.S. cities based on available data in all four categories, aiming to represent diverse demographic, geographic, and economic statuses. Figure 2 shows the varying range of carbon emissions among these cities. We aligned the transportation, land cover, and solar capacity data into a comprehensive dataset for 15 major cities, which is available at https://github.com/mileshua/TESLC. We keep working on adding more cities and additional data sources, such as EV usage, population, and income, to further expand the dataset.

3 Experiment

To identify the most influential factors for city carbon emissions, correlations between carbon emissions and the other three types of data in TESLC are calculated. This serves as a baseline benchmark to demonstrate the dataset's impact. For the remainder of the experiment, we will use the carbon emission data from 2019, as this is the only year unaffected by Covid. The code for the experiments can be found at https://github.com/mileshua/TESLC/tree/main/experiment

Transportation The Bureau of Transportation logs many pieces of information about transportation. As county data is the only available data, we approximate major cities by their respective counties, which is largely representative. For example, Boston, MA, contains 84% of the population of Suffolk county. We calculate the r^2 correlation between emissions and each type of transportation data. The highest r^2 ia 0.80, for the number of trips less than 1 mile taken.

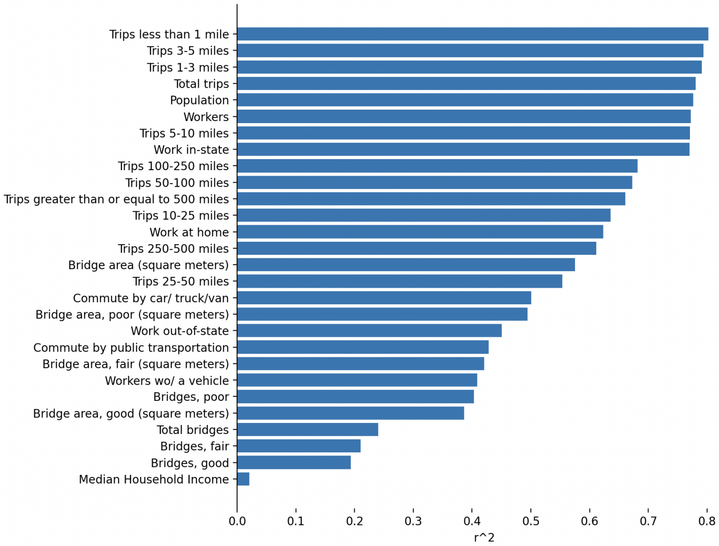

Figure 3: The coefficients of correlation r^2 for each category given in the transportation dataset and carbon emissions.

Land cover GLAD provides comprehensive land cover data globally. Polygons of the cities are obtained from Reid (2014) in the form of geojsons, which are then intersected with the land cover tif, where given bounds are parsed to determine the area in m^2 of buildings, water, and tree cover. Each data point is normalized by the area of the city, and carbon emissions are converted to per capita emissions. The r^2 values are 0.12, 0.07, and 0.08 for building, tree, and water, respectively, indicating no significant correlation.

Solar capacity The solar data obtained from Pforzheimer (2020) provides the total generated energy from solar production. Normalizing for population and applying to carbon data, we obtain a

r^2 of 0.32, signifying a moderate correlation, indicating that more solar power generation leads to less carbon emissions.

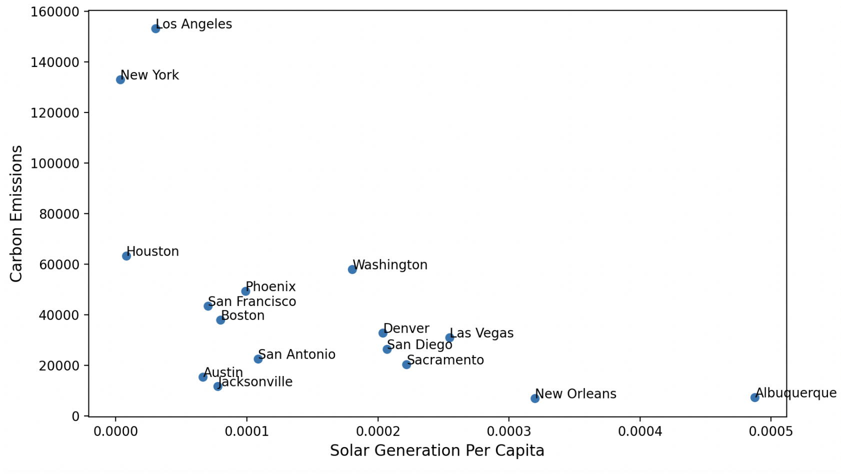

Figure 4: Carbon emissions under varying solar energy generation across different cities.

4 Conclusion and future work

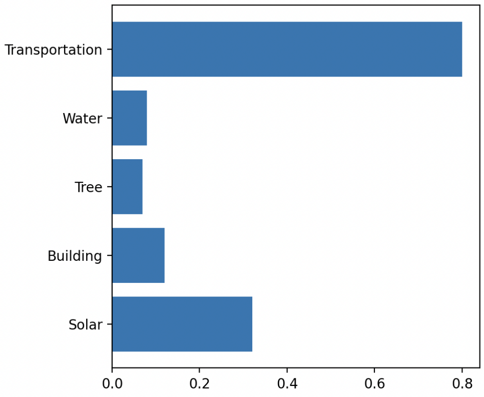

Figure 5: Comparison of correlations between carbon emissions and the causal factors

The coefficients of correlation calculated for the three datasets indicate that transportation is the best predictor of emissions, as shown in Figure 5. The small correlation values for land cover suggest that within cities, the amounts of building, water, and tree cover do not significantly affect carbon emissions. This paper finds that short trips, specifically those under one mile by car, have the strongest impact on emissions. Therefore, we recommend focusing on urban planning and improving public transportation to effectively combat emissions in cities.

Understanding the drivers of our emissions is a crucial step in reducing them. By identifying the causes of emissions, we can develop targeted strategies to combat and reduce them. This paper provides a dataset that can aid researchers in understanding the underlying causes of emissions.

The correlation analysis demonstrates an example of an effective use case of this dataset. Other deep learning methods can be leveraged to provide better insights of the correlations, or even predictions of carbon emissions. Models like KAN (Liu et al., 2024) could even offer interpretable neural network parameters to better understand the causal effects.

References

USGCRP. Climate change impacts in the united states: The third national climate assessment, 2014. URL https://nca2014.globalchange.gov/.

Da Huo, Xiaoting Huang, Xinyu Dou, Philippe Ciais, Yun Li, Zhu Deng, Yilong Wang, Duo Cui, Fouzi Benkhelifa, Taochun Sun, Biqing Zhu, Geoffrey Roest, Kevin R. Gurney, Piyu Ke, Rui Guo, Chenxi Lu, Xiaojuan Lin, Arminel Lovell, Kyra Appleby, Philip L. DeCola, Steven J. Davis, and Zhu Liu. Carbon monitor cities near-real-time daily estimates of co2 emissions from 1500 cities worldwide. *Scientific Data*, 9(1):533, Sep 2022. ISSN 2052-4463. doi: 10.1038/s41597-022-01657-z. URL https://doi.org/10.1038/s41597-022-01657-z.

Bureau of Transportation Statistics. County transportation profiles, 2019. URL https://www.bts.gov/ctp.

Peter Potapov, Matthew C. Hansen, Amy Pickens, Andres Hernandez-Serna, Alexandra Tyukavina, Svetlana Turubanova, Viviana Zalles, Xinyuan Li, Ahmad Khan, Fred Stolle, Nancy Harris, Xiao-Peng Song, Antoine Baggett, Indrani Kommareddy, and Anil Kommareddy. The global 2000-2020 land cover and land use change dataset derived from the landsat archive: First results. *Frontiers in Remote Sensing*, 3, 2022. ISSN 2673-6187. doi: 10.3389/frsen.2022.856903. URL https://www.frontiersin.org/journals/remote-sensing/articles/10.3389/frsen.2022.856903.

Mathew Reid. Geojson world cities, 2014. URL https://github.com/drei01/geojson-world-cities.

Adrian Pforzheimer. Shining cities 2020, 2020. URL https://frontiergroup.org/resources/shining-cities-2020/.

Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y. Hou, and Max Tegmark. Kan: Kolmogorov-arnold networks, 2024. URL https://arxiv.org/abs/2404.19756.