Write Up Draft: Green Space Evolution in US Capitals

By 2050, the World Bank[1] projects that 70% of the global population will live in cities. As more of the world's population moves to cities, green space will be crucial to help maintain the social, mental, economic, and climate impacts in cities. Previous research has shown that urban trees have direct health benefits through the reduction of air pollution and improving mental health, they can reduce violence and aggression, improve academic performance, and have direct economic benefits through energy conservation and food security[2]. As such, we investigate the evolution of green space in major US cities from 1984 to present day.

One of the most common methods for identifying the amount of vegetation and green space in a given area is with a metric called NDVI, or the Normalized Difference Vegetation Index. This metric is calculated mathematically from the visible and near-infrared light reflected by vegetation as given by the following equation:

$$NDVI = (NIR - RED) / (NIR + RED)$$

Where NDVI, as stated previously, is the Normalized Difference Vegetation Index. NIR is the near-infrared light and RED is the visible red light. It measures the ratio of (the reflective difference in the red and the near-infrared light) to (the sum of red and near-infrared reflectance). This metric is useful because green vegetation reflects near-infrared light, while absorbs red light (Hafen, 2021). Thus, if there were lots of green vegetation in an area, we would expect the perfect outcome of NDVI to be:

$$NDVI = (NIR - RED) / (NIR + RED) = (1 - 0)/(1 + 0) = 1$$

The results of this calculation range from -1 to +1. The scale is interpreted as follows:

- Anything below 0 signifies no vegetation
- Anything close to 0 signifies there are no green leaves
- Anything close to 1 (0.8-0.9) signifies the highest possible density of green leaves.

We selected this metric as our method to calculate the change of green space in a city largely because of the industry standard in using this, "nearly all satellite Vegetation Indices employ this difference formula to quantify the density of plant growth on the Earth" according to NASA (Levy and Przyborski, 2000).

With data from 1984 to present day, the Landsat satellite record created by NASA and USGS provides the longest continuous satellite image record of the Earth's surface in existence. The Google Earth Engine API provides access to the Landsat datasets in an accessible way for computation. You can [apply for access] to the Google Earth Engine API for academic or research

[1] https://www.worldbank.org/en/topic/urbandevelopment/overview#1
[2] https://nph.onlinelibrary.wiley.com/doi/full/10.1002/ppp3.39

reasons. Here, we use a variety of Landsat datasets including Landsat 5 (1984-2012), Landsat 7 (1999-2022) and Landsat 8 (2013-2022), giving us access to results from 1984 to present day. The Google Earth Engine API offers alternative datasets such as MODIS, which provides similar aerial data but with improved temporal resolution, but worse spatial resolution, at 250 meters. The Landsat dataset offers improved spatial resolution, at 30 meters, compared to MODIS, which is important for a project like this where we analyze the green space detail of cities.

Literature Review of Other attempts exploring this topic

As mentioned previously, the standard for green space calculations is NDVI, measuring the visible and near-infrared light to determine the green space in a given area, traditionally using satellite images to assess the green space. Previous research attempts used the Landsat data but investigated different questions, such as the impact of continuous deforestation in New England (Olofsson, Holden, Bullock, & Woodcock, 2016). Others used different satellite images but similarly applied NDVI as the metric of choice to quantify urban greenness, taking a similar approach to our case study by investigating affluent and poor suburb neighborhoods in Johannesburg (Abutaleb, Freddy Mudede, Nkongolo, & Newete, 2021).

Similar work can be found classifying satellite images in Kaggle competitions and on Medium for Amazon rainforest deforestation that use Google Earth Engine satellite imagery and NDVI to calculate the deforestation.

Exploratory analysis with a few interesting graphs[3]

Using the Landsat dataset from the Google Earth Engine API, we were able to obtain the NDVI data for select US capitals over the course of 1984 to 2021. Due to the exhibition of seasonality, we averaged data of NDVI by years, then plotted all of them out, shown below in Figure 1. An interesting thing to note is that most cities experienced some sort of increase in their NDVI values. Out of the 35 state capital cities that's in our analysis, only 8 experienced a decrease, and 6 of these decreases are smaller than 0.02. To put it into perspective, 7 cities experienced an increase smaller than 0.02, while 19 cities experienced an increase larger than 0.02.

Some of them had large fluctuations over the years that did not seem to make sense (such as Atlanta, which is the third graph in the top row), while others had very consistent values that did not change very much over the years (such as Phoenix, which is the second graph in the 6th row). Additional possibilities for the fluctuations are available in the assumptions section.

[3] Code for EDA, but also in the GitHub folder: https://github.com/HcaZreJ/ds-for-sg-tutorial/blob/main/notebooks/Assignment%202%20EDA.ipynb

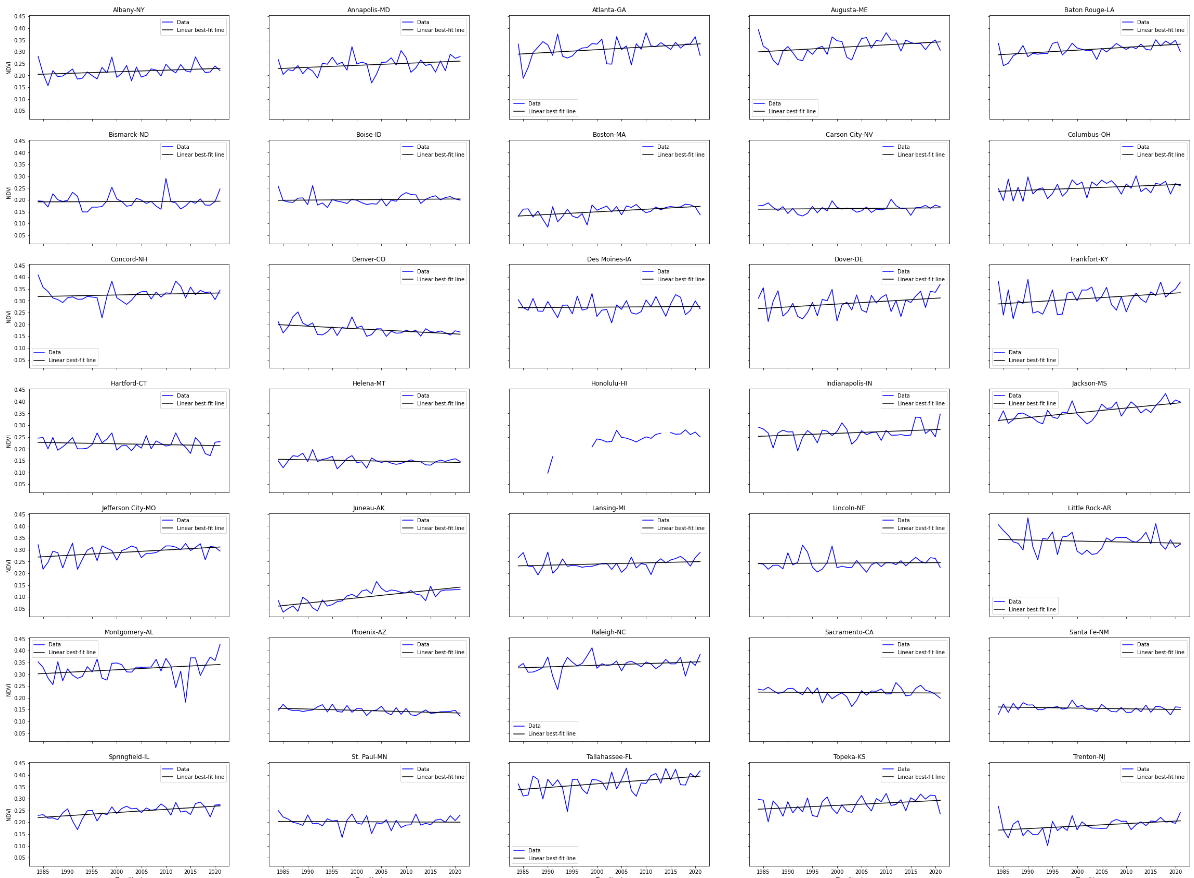

Figure 1 - Yearly averaged NDVI plotted against time in years for 35 state capital cities in the United States. Blue lines represent the actual data values, while black lines represent the best-fit line. Graphs are all on the same scale, meaning that the x-axis ticks (1984 - 2021) and y-axis ticks (0.05 - 0.45) are all the same.

The three cities that experienced the largest change in NDVI values over these 26 years are exhibited below - all of them are increasing. Although the value might seem small, they are actually a lot in perspective to the scale of NDVI. This positive increase indicates that either the amount of vegetation in the city has increased, or that the existing vegetation is healthier, or both. If the reason was solely due to an increase in the amount of vegetation, then the city of Juneau evolved to have 8% more healthy vegetation coverage over the course of 26 years!

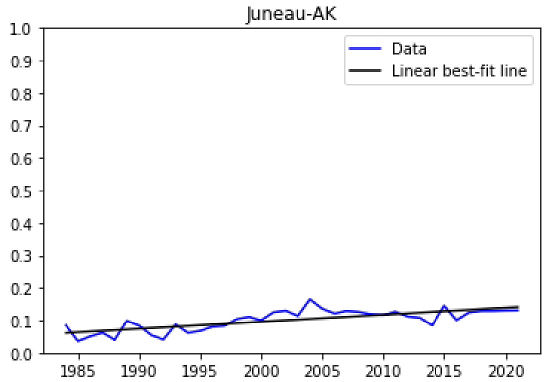

Figure 2 - Yearly averaged NDVI plotted against time in years for Juneau, Alaska, an increase by 0.080

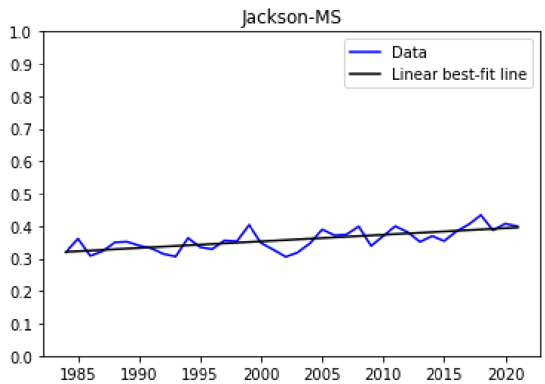

Figure 3 - Yearly averaged NDVI plotted against time in years for Jackson, Mississippi, an increase by 0.075

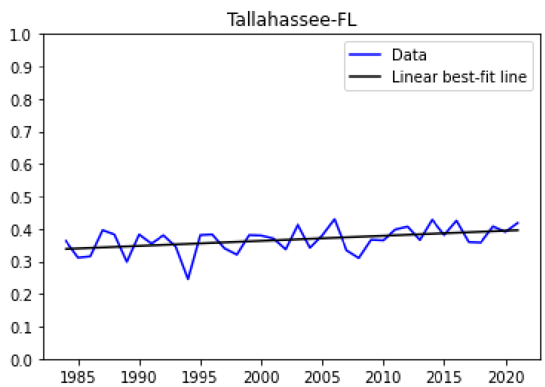

Figure 4 - Yearly averaged NDVI plotted against time in years for Tallahassee, Florida, an increase by 0.057

Additional analysis could be conducted by investigating connections between the amount of green space in a city and wellness indicators, such as mental health, violence & aggression, academic performance, crime rate, etc.

Assumptions for data collection and data analysis

Google Earth Engine API provides access to the Landsat datasets which offer access to satellite images across the entire world. Here, we use a variety of Landsat datasets including Landsat 5 (1984-2012), Landsat 7 (1999-2022) and Landsat 8 (2013-2022), giving us access to results from 1984 to present day. The Google Earth Engine API offers alternative datasets such as MODIS, which provides similar aerial data but with improved temporal resolution, but worse spatial resolution, at 250 meters. The Landsat dataset offers improved spatial resolution, at 30 meters, compared to MODIS, which is important for a project like this where we analyze the green space detail of cities.

To obtain the city boundaries, we used the Boundaries.io API which allows us to define the border of cities based on the US census postal codes for a given city. This could lead to some cities that appear geographically as one city (e.g. Kansas City, as seen below, is split in half by the Kansas and Missouri state boundaries, and though functionally operating as a single city, officially is separated by borders).

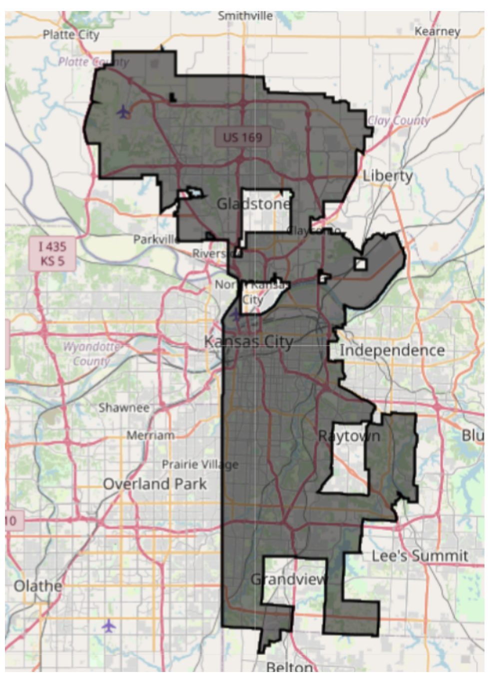

Figure 5: Example of a city that is officially separated by US state boundaries but that functionally operates as a single city, a potential limitation to our chosen method.

As mentioned previously, some of the cities had large fluctuations. These might be due to the order in which we calculated the NDVI, noted as follows:

1. We averaged the NDVI over the city, then
2. Averaged the data over the course of one year.

The justification for why these steps were done is that:

1. A simple filtering leads to a large loss in data.
2. There needs to be a metric for the entire city, comparing pixel by pixel is not realistic.
3. Seasonality problems prevent longer-term trends from being visible.

References

Abutaleb, K., Freddy Mudede, M., Nkongolo, N., & Newete, S. W. (2021). Estimating urban greenness index using Remote Sensing Data: A case study of an affluent vs poor suburbs in the City of Johannesburg. *The Egyptian Journal of Remote Sensing and Space Science*, *24*(3), 343–351. https://doi.org/10.1016/j.ejrs.2020.07.002

Earth Observing System. (2019, August 30). NDVI FAQS: Top 23 frequently asked questions about ndvi. EARTH OBSERVING SYSTEM. Retrieved from https://eos.com/blog/ndvi-faq-all-you-need-to-know-about-ndvi/#:~:text=Consider%20using%20EVI%20(Enhanced%20Vegetation,crops%20or%20other%20dense%20crops

Hafen, K. (2021, February 9). Remote Sensing with QGIS: Calculate ndvi. OpenSourceOptions. Retrieved from https://opensourceoptions.com/blog/remote-sensing-with-qgis-calculate-ndvi/

Levy, R. and Przyborski P. (2000, August 30). Measuring Vegetation (NDVI & EVI). Retrieved from https://earthobservatory.nasa.gov/features/MeasuringVegetation/measuring_vegetation_2.php

Olofsson, P., Holden, C. E., Bullock, E. L., & Woodcock, C. E. (2016). Time series analysis of satellite data reveals continuous deforestation of New England since the 1980s. *Environmental Research Letters*, *11*(6), 064002. https://doi.org/10.1088/1748-9326/11/6/064002