

Assignment 3 Report

Dataset: <https://www.kaggle.com/rajsengo/texas-covid19-vaccine-allocation>

Dataset obtained from Kaggle, an online and free repository of datasets.

Dataset format: Tabular

The original dataset is a static, flat table which contains data on COVID-19 vaccinations in Texas by county. The amount of vaccinations is also divided by demographics such as race, age, and also phase (1a, 1b, 1c). The dataset is multivariate, as there are multiple attributes for each county, which serves as the key.

Data Type: Item-based

The events are laid out in a tabular dataset so each record within the dataset is considered an item, and each item has its own attributes. Each item is a county in Texas.

Attribute Types & Semantics

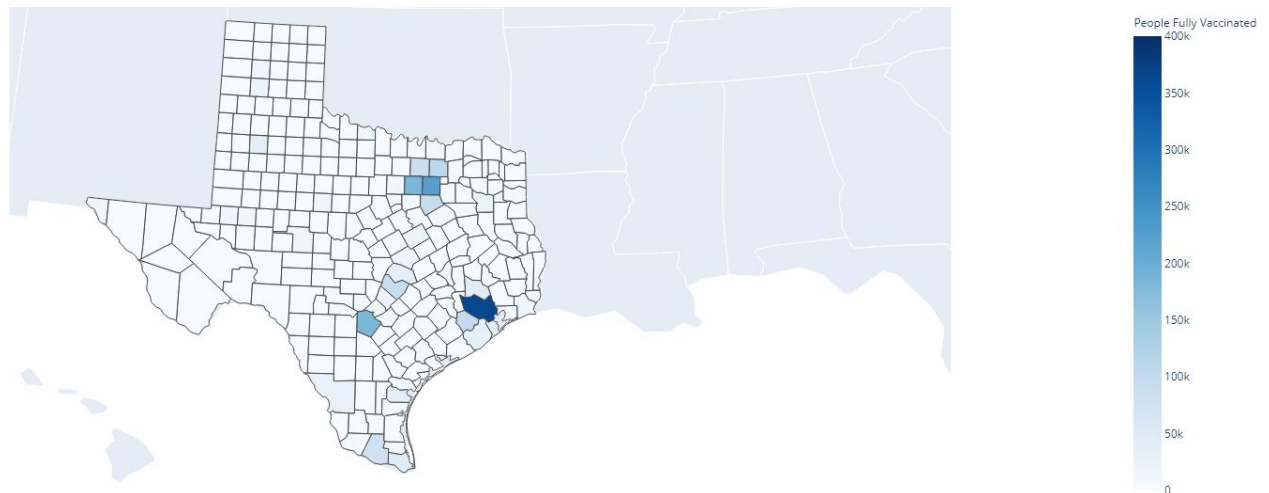
There are twelve attributes: county name (serves as key), public health region, total doses allocated, vaccine doses administered, people vaccinated with at least one dose, people fully vaccinated, population 16+, population 65+, phase 1a healthcare workers population, phase 1a long-term care residents population, age 16-64 with medical condition population, and education and childcare personnel population.

Preprocessing

To preprocess the data, I removed attributes that were not necessary or would not make sense for visualization. The attributes I kept were county name (since it served as the key and would allow me to create a choropleth map), and people fully vaccinated. These two attributes were necessary for creating a choropleth map that visualizes vaccinations by county in Texas. I also needed FIPS codes for each county, so I had to manually import those.

Visualization

I decided to represent this data using a choropleth map in Python.



Analysis

Based on the choropleth map, we can observe that counties with greater populations have more people vaccinated. The darkest colored counties correspond to Dallas, Austin, San Antonio and Houston. Houston, as the most populated city in Texas, has the most vaccinations. Surprisingly, despite having a population with ~600,000 more residents than Austin, San Antonio seems to not have very many more vaccinations than Austin. San Antonio and Dallas have comparable population sizes and vaccinations.

Dataset: <https://data.world/vizwiz/employment-unemployment-by-state-and-month>

Dataset obtained from data.world, an online and free repository of datasets.

Dataset format: Tabular

The original dataset is a static, flat table which contains data on the United States unemployment rate per state on a monthly basis for each year from 1976-2020.

Data Type: Item-based

The events are laid out in a tabular dataset so each record within the dataset is considered an item, and each item has its own attributes. Each item is an unemployment rate for a specific state in a specific month and year.

Attribute Types & Semantics

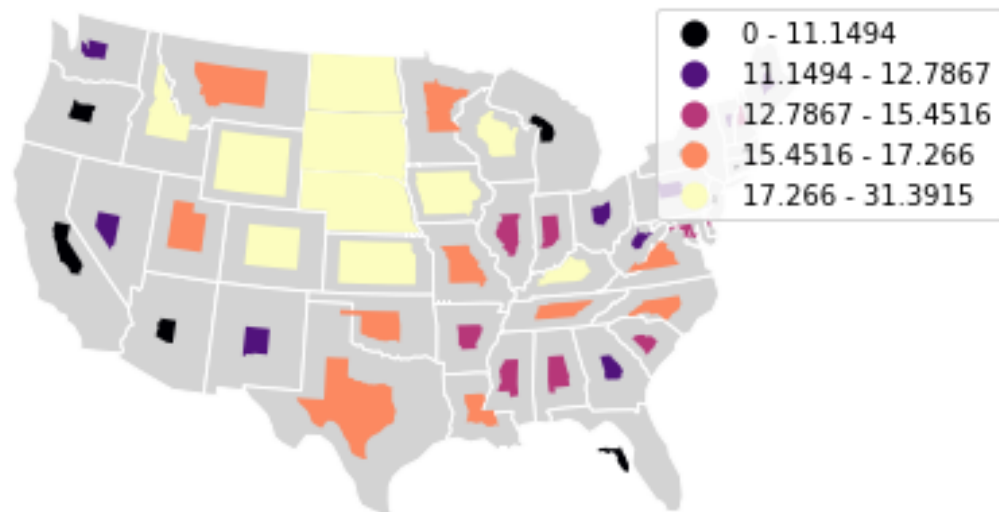
There are seven attributes in this dataset: FIPS code, state, year, month, employable population, employed, and unemployed. The employment numbers are all whole numbers.

Preprocessing

To preprocess the data, I removed attributes that were not necessary or would not make sense for visualization. The attributes I kept were FIPS, state, employable population and unemployed. Since I am wanting to visualize unemployment rates per state, I needed to divide the employable population by the unemployed to create the main attribute I wanted to visualize that I named Unemployment Rate.

Visualization

I decided to represent this data using a non-contiguous cartogram in Python.



Analysis

Observing the generated cartogram, we can see that North Dakota, South Dakota, and Nebraska have the highest rates of unemployment in 1976. The most populous states such as California, New York, Washington, and Oregon (as well as the other northeastern states) appear to have relatively low unemployment rates. It appears that the midwestern states including the Dakotas and Nebraska have the highest unemployment rates.

Dataset:

<https://waupacacounty.maps.arcgis.com/apps/SimpleViewer/index.html?appid=b6dd1313a3944c84ad965dbcbac197d3>

Dataset obtained from ArcGIS's LiDAR data for Waupaca County in Wisconsin.

Dataset format: Tabular

The original dataset is a point cloud created by the LiDAR method.

Data Type: Item-based

The events are laid out in a tabular dataset so each record within the dataset is considered an item, and each item has its own attributes. Each item is a parking event.

Attribute Types & Semantics

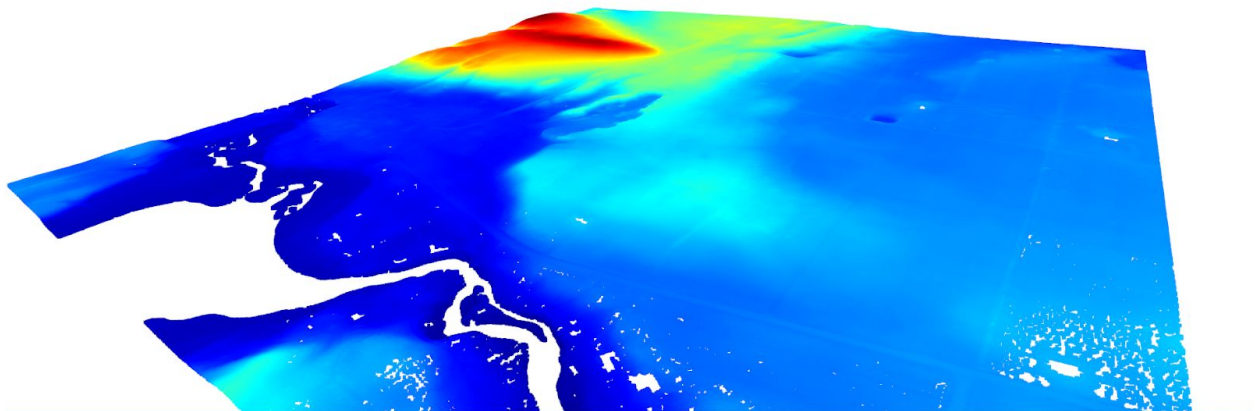
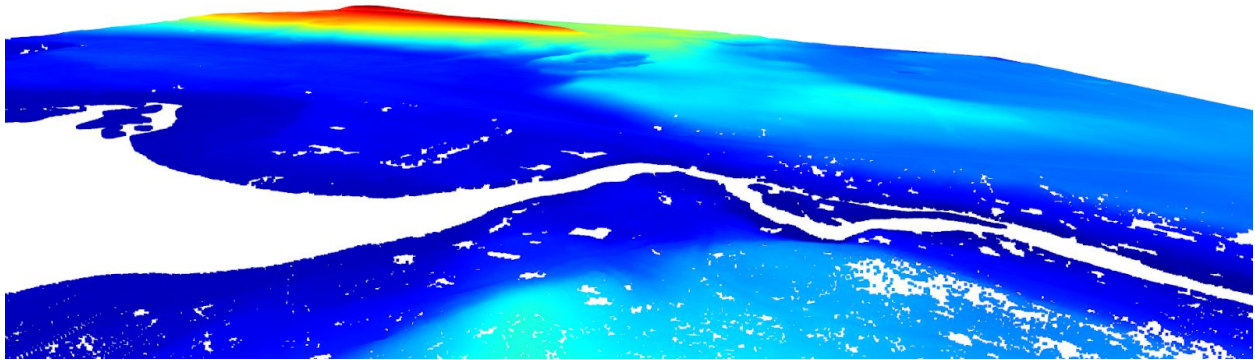
As this is a geometry dataset that contains information on the contours of a specific area in Wisconsin, there are no attributes.

Preprocessing

There was no preprocessing necessary for this data visualization.

Visualization

I decided to represent this data using a 3d point cloud in open3d.



Analysis

Observing the generated 3d point cloud, we can observe the variation in elevation in this particular area not only based on the shape the points take, but also their color. It appears that lower elevation is indicated by darker blue points and higher elevation is indicated by darker red points. The white spaces, or gaps between the points, appear to represent bodies of water as we can clearly see a lake with a river extending outward from it.