

# FIT3179 Week 9 Homework

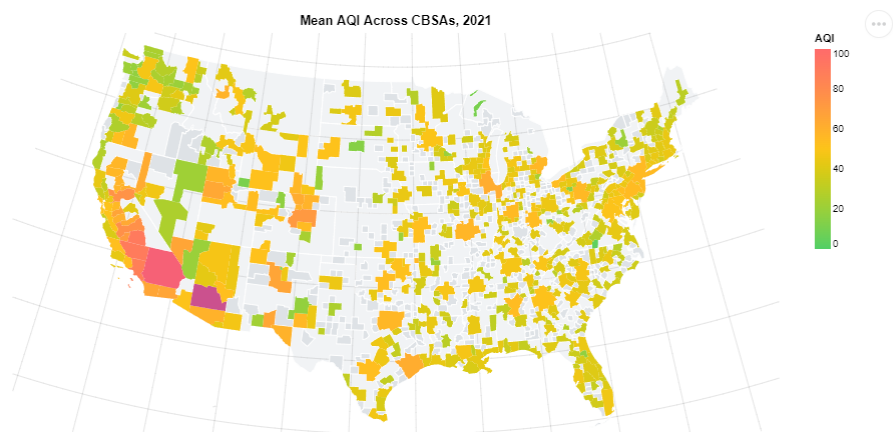
| Student Details |  |
|-----------------|--|
| Name            | Andre Pham   |
| Email           | <a href="mailto:apha0019@student.monash.edu.au">apha0019@student.monash.edu.au</a> |
| Student Number  | 31448232   |
| Studio          | Studio 03  |
| Tutor           | Dr Jesmin Nahar  |

## URL

[VISUALISATION URL](#)

## Screen Capture

### AIR QUALITY IN THE UNITED STATES



The mean AQI (Air Quality Index) across CBSAs (core-based statistical areas). AQI range from 0-500, greater numbers indicating poorer quality air. Graticule spacing is 5° between meridians and parallels.

This visualisation was created by Andre Pham 31448232 for FIT3179. Licence CC by 4.0.  
[Dataset available here.](#)

## Domain

This visualisation presents what areas of the United States in 2021 had the poorest air quality and which had the best. It indicates the current state of the United State's air quality to help inform which areas are most safe and which need improvement. This is achieved by indicating the mean AQI of measurements taken within CBSAs (Core-based statistical areas) within the United States in 2021, by using the colour channel to indicate the magnitude.

## Dataset

The visualised dataset is available on Kaggle [here](#). It contains approximately 5.72 million AQI measurements across the United States from 1980 to 2022. It was uploaded to Kaggle by user [Caleb Reigada](#), where the [original locational data](#) was sourced from [simplemaps](#) and the [original aqi data](#) from the [United States Environmental Protection Agency](#).

The dataset is of the table dataset type, where the items are AQI measurements. Each item is made up of 14 attributes (including the bracketed attribute types): CBSA Code (quantitative), Date (ordinal), AQI (quantitative), Category (ordinal), Defining Parameter (categorical), Number of Sites Reporting (quantitative), city\_ascii (categorical), state\_id (categorical), state\_name (categorical), lat (quantitative), lng (quantitative), population (quantitative), density (quantitative), and timezone (categorical).

## Data Transformation

To transform the original dataset, I conducted the following steps:

1. Filtered the measurements to only include those from 2021
2. Grouped measurements by their CBSA
3. Identified the mean AQI measurement value for each CBSA
4. Output a record for each unique CBSA, with only the relevant columns

Python was used in conjunction with the pandas module to process and export the data.

## Idiom Justification

The unclassed choropleth idiom was selected for the map visualisation.

Choropleths are required to be used with data that is attached to enumeration units, standardised, and supports application on a continuous statistical surface. These three conditions are met by the data being visualised. The AQI measurements are recorded and grouped within specific CBSAs and the AQI measurements are tied to their respective CBSA and hence are attached to enumeration units. The data has been standardised by calculating the mean of measurements within each CBSA, so that the data isn't skewed by measurement count. And finally, the data is applied on a continuous statistical surface, hence AQI can be measured at any location, continuously.

The choropleth idiom provides a basis for encoding a quantitative attribute (the AQI) to regions on a geographic map. This allows quick and intuitive identification for which geographical areas have the best and worst air quality. It also provides a basis of grouping measurements into areas to provide better readability - if a dot map was to be used, each measurement would have to be plotted onto the map, hence cluttering the map, and reducing the amount of useful information by not being able to provide the AQI of areas, and rather forcing the user to navigate through the sea of measurements and their unique AQI values. Hence the choropleth idiom provides storytelling through being mapped to areas.

The unclassed choropleth was selected due to the nature of AQI being a continuous value - grouping AQI values is abstract and doesn't create for fair comparison on a visualised map. For example, if one location had an AQI of 49, and an adjacent location had an AQI of 50, it would be misleading to classify these as different colours despite having virtually the same air quality. Hence making the choropleth be unclassed provides clarity in what is being shown.