

FIT3179 Week 10 Homework

IMPORTANT: I have added text annotations - they are only viewable after a filter is applied, as they would clutter the visualisation otherwise.

Student Details	
Name	Andre Pham
Email	apha0019@student.monash.edu.au
Student Number	31448232
Studio	Studio 03
Tutor	Dr Jesmin Nahar

URL

TODO - UPDATE URL

[VISUALISATION URL](#)

Screen Capture

AIR QUALITY IN THE U.S.

By Andre Pham 31448232 for FIT3179, Homework Week 10.

U.S. Emissions of Volatile Organic Compounds

The total estimated emissions of volatile organic compounds (more info [here](#)) produced by the U.S. between 1970 and 2019. Emissions have been going down, improving air quality.

*Annotations applied during filter.

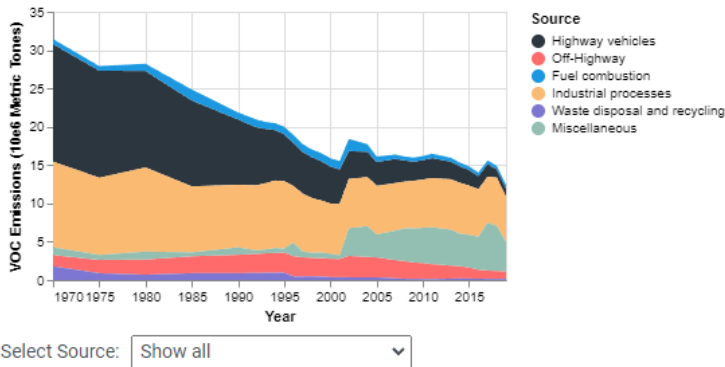


Figure 1: Task 1 (No filter applied) (Note - annotations are applied after a filter is).

AIR QUALITY IN THE U.S.

By Andre Pham 31448232 for FIT3179, Homework Week 10.

U.S. Emissions of Volatile Organic Compounds

The total estimated emissions of volatile organic compounds (more info [here](#)) produced by the U.S. between 1970 and 2019. Emissions have been going down, improving air quality.

*Annotations applied during filter.

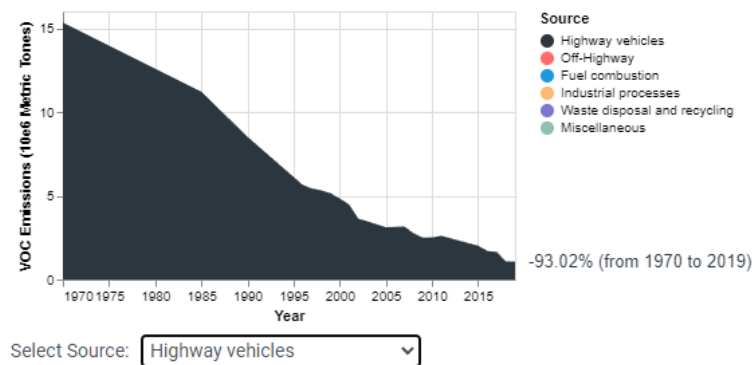


Figure 2: Task 1 (Filtered to "Highway vehicles") (Annotation visible bottom right).

Domain

This visualisation presents the state of air quality in the United States. It provides context on the number of tones of pollutants that are being produced annually in the U.S. (and how that has changed over the past six decades) as well as the current state of the United State's air quality, geographically.

Dataset

The stacked area chart's dataset is available [here](#). It contains the estimated annual U.S. emissions of volatile compounds from 1970 to 2021. It is hosted by the [Bureau of Transportation Statistics](#) and was originally sourced from the U.S. Environmental Protection Agency's Air Emissions Inventories, available [here](#).

The dataset is of the table dataset type, where the items are annual U.S. VOC emissions. The source, year and measurement attributes are categorical, ordinal and quantitative respectively.

The choropleth map's 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](#).

Data Transformation

The stacked area chart's dataset required the rows to be pivoted, the filtering of 2020 and 2021 data (due to incomplete measurements), and the conversion from U.S. short tones to metric tones.

The choropleth map's dataset required the filtering of items to 2021 only, the removal of columns not required, and the grouping of CBSA measurements to then identify the mean.

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

Idiom Justification

A stacked area chart was used to visualise the change of VOC emissions annually produced by the U.S. The use of a stacked area chart means that the total VOC emissions can be visualised, as well as broken down by their sources. For example, if a line chart was to be used to show the sources, the net VOC emissions wouldn't be identifiable - only by stacking the sources via a stacked area chart is this information conveyed. The option to filter by source also negates the largest drawback of the stacked area chart (readability of each category) by allowing each source to be read in isolation.

The justification for the choropleth idiom was covered in homework week 9 and hence will not be re-stated.