

# Databases Autumn 2020

## Data Analysis Project P0: Project Idea

*Visualizing Traffic density data and comparing them  
to air pollution- and meteorological data in Basel,  
London and Los Angeles*

October 4, 2020

Pascal Kunz, Etienne Mettaz

### 1 Introduction

During 2017-2019 the ETH Zurich collected traffic data by using over 23541 stationary detectors on roads in urban areas across 40 cities worldwide. Therefore it is the largest multi-city traffic dataset that is currently publicly available with a size of about 8GB.

The Dataset provides information about the vehicle flow as well as the vehicular density on roads which are not only identifiable by their names but also their WGS84 coordinates.

We want to focus our Project on three Major cities, namely Basel, London and Los Angeles. For the data Analysis we intend on comparing meteorological data as well as air pollution data to the UTD19 Dataset.

The air pollution data is sourced from 3 trustworthy data sources publicly available which are produced from stationary measuring units in the concerning cities.

## 2 Datasets

### 2.1 Traffic Information

The largest publicly available traffic dataset provided from the ETH Zurich.

- Source: ETH Zurich, <https://www.research-collection.ethz.ch/handle/20.500.11850/437802>
- Size: 7.81 GB
- Format: CSV
- Temporal: 2017 - 2019
- Spatial: WGS84 coordinates, Basel, London, Los Angeles

### 2.2 Air quality, Station Feldbergstrasse, St. Johannisplatz, Basel-Binnigen, Chrischona

Using the stationary air quality measuring entities on public roads, distributed across Basel.

- Source: Canton Basel-Stadt, <https://data.bs.ch/explore/dataset>
- Size: approximately 200mb combined
- Format: JSON or CSV or GeoJSON
- Temporal: 2000-2020, half an hour Intervals
- Spatial: Name and address of stationary air pollution measuring unit.

### 2.3 Air quality London

Using the stationary air quality measuring entities on public roads, distributed across London. Note that there are a lot of roads that are being monitored and for each road / area a CSV document can be downloaded accordingly.

- Source: London Air: <https://www.londonair.org.uk/london/asp/datadownload.asp>
- Size: approximately 200mb combined, depending on how many roads will be analysed.
- Format: CSV
- Temporal: 2000-2020, half an hour Intervals
- Spatial: Name of of Location of recorded data.

## 2.4 Air quality Los Angeles

Using the stationary air quality measuring entities on public roads, distributed across the entire United States. The datasets are publicly available and can be download for each specific year.

- Source: EPA.gov: [https://aqs.epa.gov/aqsweb/airdata/download\\_files.html#Raw](https://aqs.epa.gov/aqsweb/airdata/download_files.html#Raw)
- Size: approximately 200mb combined, depending on how many roads will be analysed.
- Format: CSV
- Temporal: 2000-2020, half an hour Intervals
- Spatial: Name of of Location of recorded data.

## 2.5 Weather specific data Basel, London, Los Angeles

Using a public meteorological database which holds the information about Temperature, Air Pressure, Humidity, Amount of rain etc.

- MeteoBlue: <https://www.meteoblue.com/de/wetter/archive>
- Size: approximately 10mb each
- Format: CSV
- Temporal: 2000-2020, hourly Intervals
- Spatial: Basel, London, Los Angeles

## 3 Analysis Goals

- Comparing the traffic density over certain time intervals and comparing them to environmental data, for example by using the stationary air pollution measurement across Basel, London and Los Angeles. The goal of the analysis is that we are able to identify clear peaks of air pollution which correlate with the vehicular density.
- How is the air pollution affected by meteorological circumstances such as rain, temperature, wind. We want to be able to determine whether a clear dip of air pollution is visible for example when it has been heavily raining for several days.
- compare areas with similar vehicular density. Do all these similar areas have a similar air quality?
- Rush hour analysis across different cultures. Do all people get to work at the same times or do Swiss People for example get to work later.
- Is there a correlation between heavy rain and the use of vehicles? E.g. Are there more cars on the road when it is raining.
- Yet undecided: Map the vehicular density into a map by using the geographical data provided in the UTD19 Dataset. As of now we have been able to find some tools which require payment unfortunately. We will check back with the assistants in order get further information about the feasibility of this idea with regards to eventual access permissions to certain tools.

## 4 Tools

- mySQL as DataBase-Tool
- Python, for the data visualization
- Maybe: Batch Geo, <https://de.batchgeo.com/>, for the map visualization. However it is not for free.
- Git
- Gant-Project Plan