

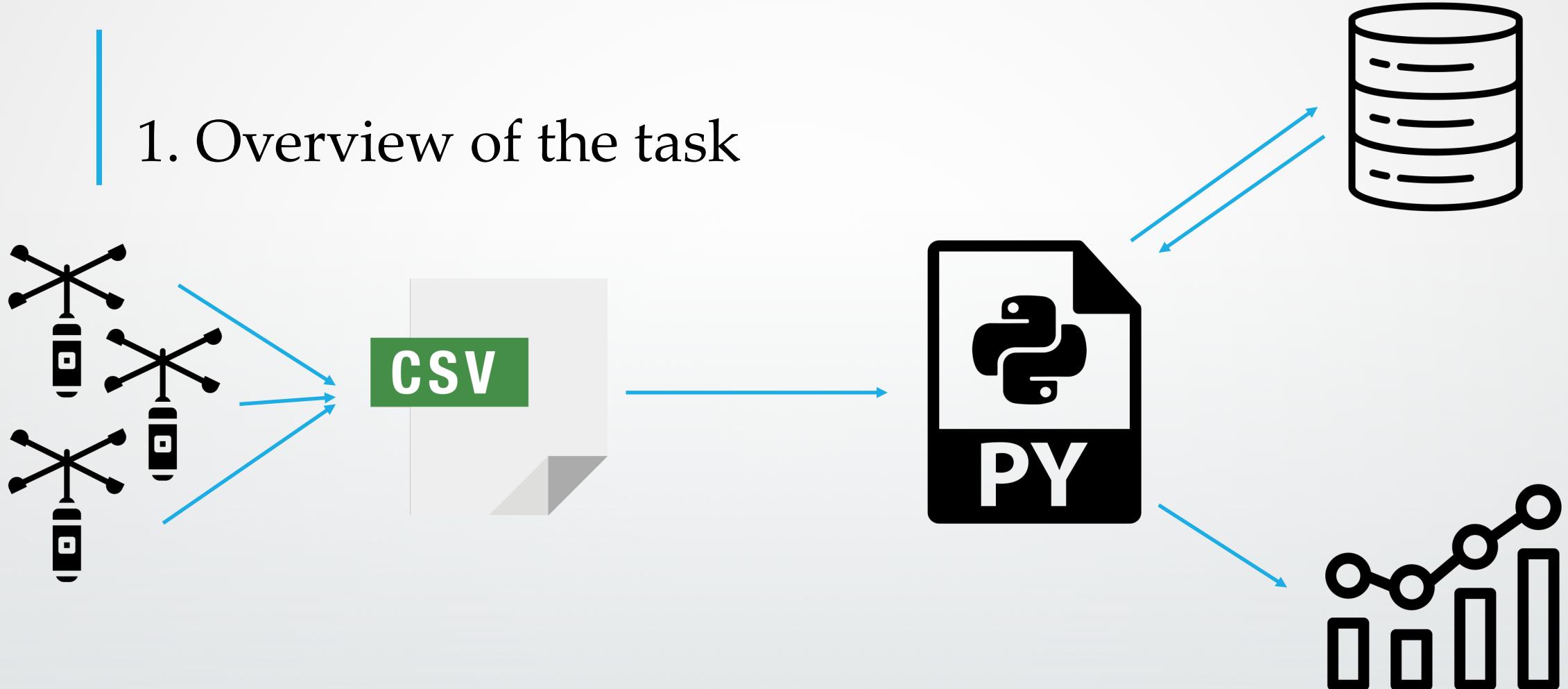
Programming and Databases

BENEDIKT MIETHSAM – PROJECT PRESENTATION

Agenda

1. Overview of the task
2. Data import and preparation
3. Relational model and database connection
4. Meteorological analyses

1. Overview of the task



2. Data import and preparation

- Data source consists of 39.389 data entries (lines)
- Parsing the stations and observations with the csv.py library

```
def parseStations():
    readCSV = csv.reader(f1, delimiter=',')
    # Skip the first entry (descriptors) of the csv
    iterCsv = iter(readCSV)
    next(iterCsv)

    for row in iterCsv:
        lat = row[7]
        lon = row[8]
        stationID = row[9]
        stationName = row[10]
        values = (lat, lon, stationID, stationName)
        if values not in stations:
            stations.append(values)
    f1.close()
    return stations
```

2. Data import and preparation

- Missing values for wind direction ('---') & wrong values for wind speed (-99)

```
41,0,-99,29.2,---,-99,2015-08-03T16:00:00,38.65435,-9.067241,928,Barreiro
```

→ Replacing the values by NULL to be able to filter them

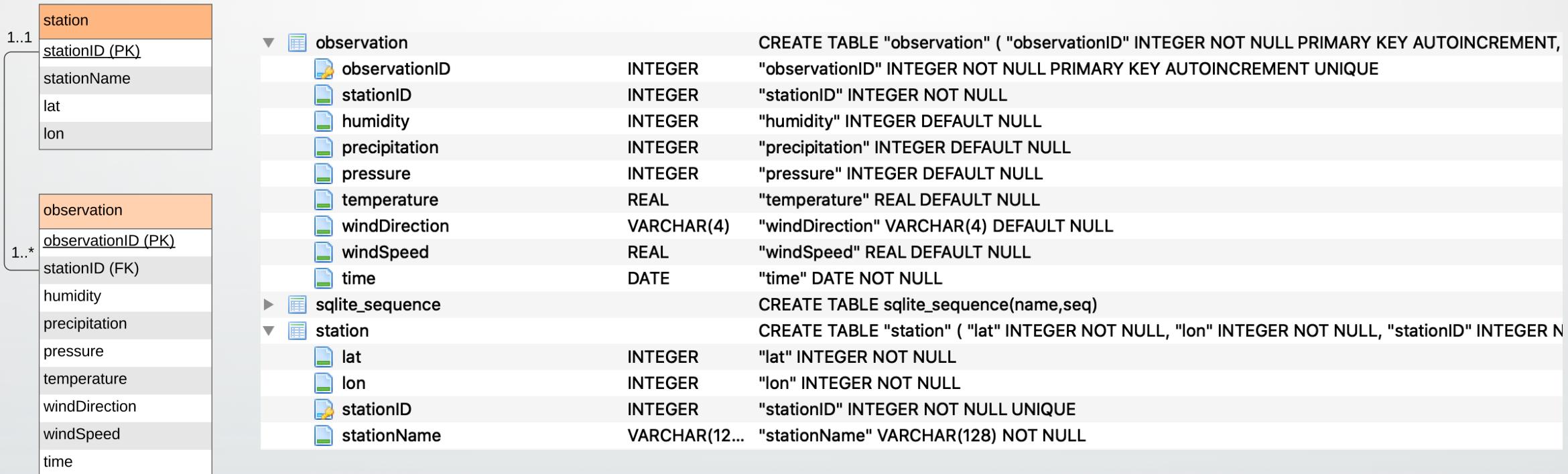
- Repeating rows in the dataset

→ if values not in stations:

```
stations.append(values) (Python)
```

→ „INSERT OR IGNORE INTO station ...“ (SQL)

3. Relational model and database connection



3. Relational model and database connection

```
import sqlite3

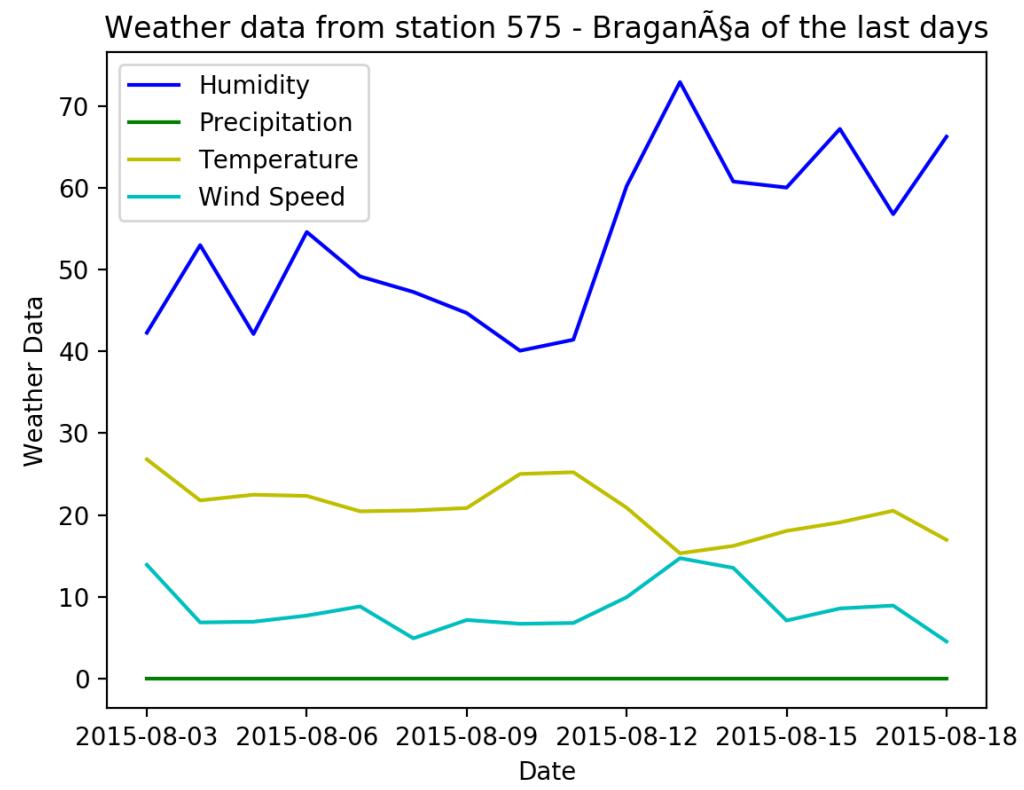
def openConnection():
    conn = sqlite3.connect('./resources/ema-201508.db')
    return conn

def createTables(conn):
    c = conn.cursor()
    # Create table
    c.executescript("DROP TABLE IF EXISTS station;
                    DROP TABLE IF EXISTS observation;
                    CREATE TABLE IF NOT EXISTS "station" (
                        "lat" INTEGER NOT NULL,
                        "lon" INTEGER NOT NULL,
                        "stationID"  INTEGER NOT NULL,
                        "stationName" VARCHAR(128) NOT NULL,
                        PRIMARY KEY("stationID")
                        UNIQUE(stationID)
                    );")
```

```
def insertStationValues(conn, station_values):
    for values in station_values:
        query = (
            "INSERT OR IGNORE INTO station('lat', 'lon',
            'stationID', 'stationName') VALUES (?,?,?,?,?)")
        with conn:
            cur = conn.cursor()
            cur.execute(query, values)
```

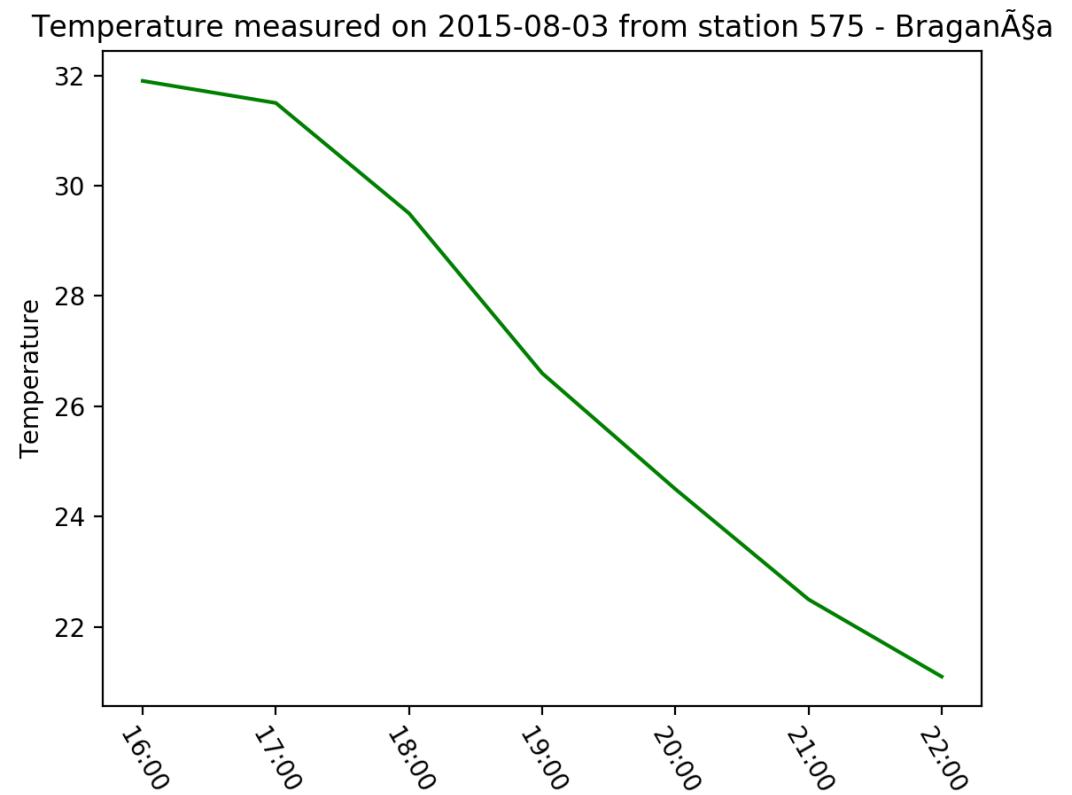
4. Meteorological analyses (using matplotlib)

- Station Overview



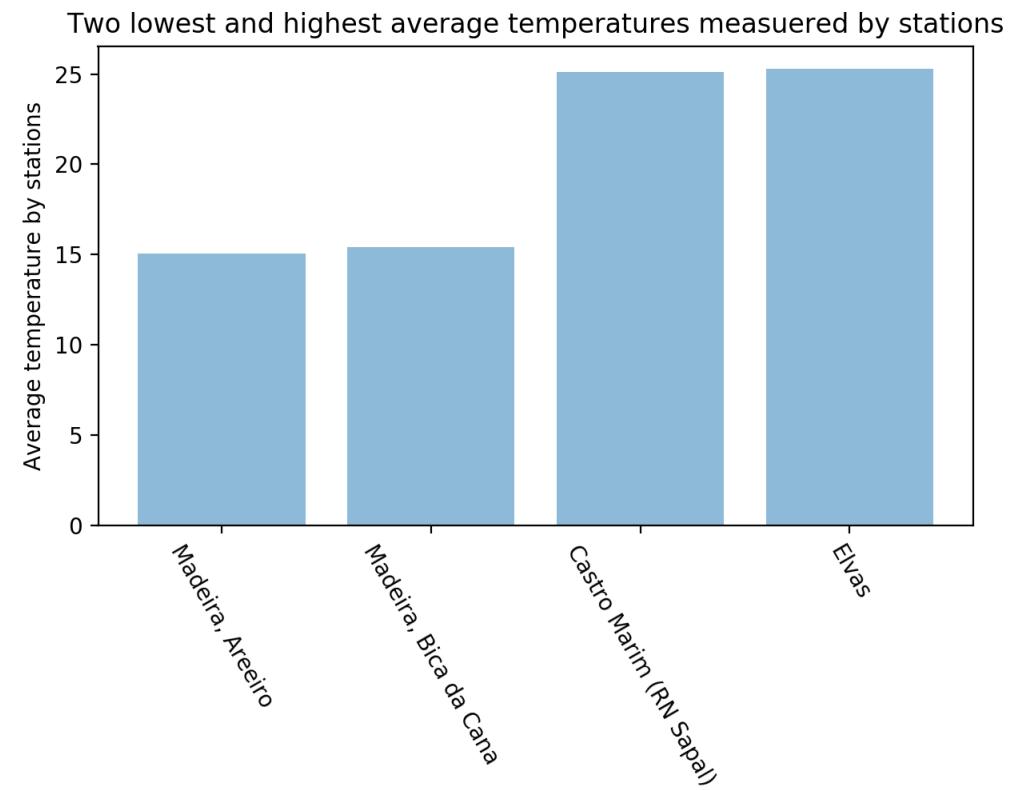
4. Meteorological analyses (using matplotlib)

- Detailed temperature curve by day



4. Meteorological analyses (using matplotlib)

- Lowest and highest temperature



Any Questions?

