Zeppelin

Parking Dataset

```scala
1 // Imports
2 import org.apache.spark.sql.functions._
3 import org.joda.time.format.DateTimeFormat
4 import org.apache.spark.rdd.RDD
```

```
import org.apache.spark.sql.functions._
import org.joda.time.format.DateTimeFormat
import org.apache.spark.rdd.RDD
```

FINISHED

```scala
 1 // Load data - adjust the path to the location of your data
 2 val inputPath =  "/Users/joannariascos/data/aarhus_parking.csv"
 3 val park = sqlContext.read
 4         .format("com.databricks.spark.csv")
 5         .option("header", "true") // Use first line of all files as header
 6         .option("delimiter", ",")
 7         .option("inferSchema", "true") // Automatically infer data types
 8         .load(inputPath)
 9
10         park.registerTempTable("park")
```

```
inputPath: String = /Users/joannariascos/data/aarhus_parking.csv
park: org.apache.spark.sql.DataFrame = [vehiclecount: int, _id: int ... 2 more fields]
warning: there was one deprecation warning; re-run with -deprecation for details
```

FINISHED

```sql
1 %sql
2 select vehiclecount, count(1) value
3 from park
4 group by vehiclecount
5 order by vehiclecount
```

FINISHED

vehiclecount	value
0	3,627
1	300
2	292
3	391
4	340
5	281
6	222
7	186
8	266

Results are limited by 1000.

1 %sql select * from park

FINISHED ▷ ⌖ 📖 ⚙

settings ▲

All fields:
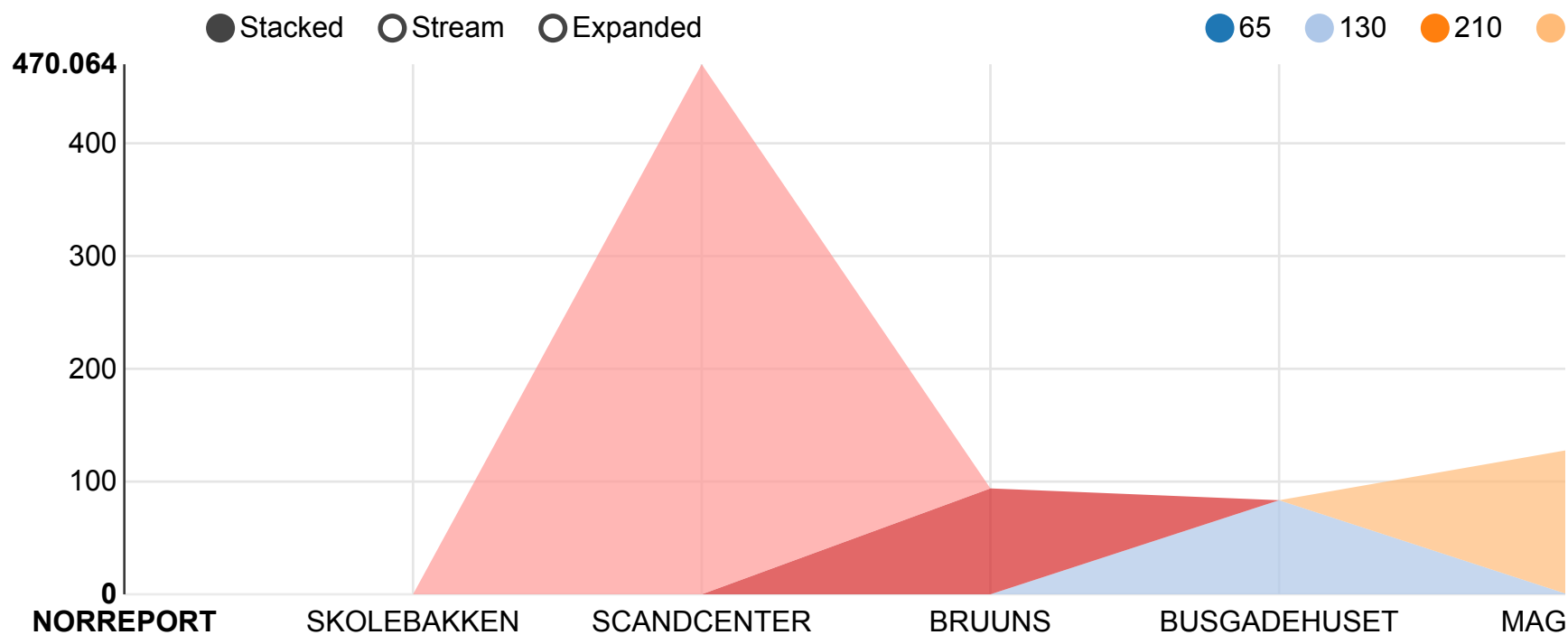
vehiclecount   _id   totalspaces   garagecode

Keys

garagecode ✖

Groups

totalspaces ✖

Values

vehiclecount AVG ✖

○ Stacked   ○ Stream   ○ Expanded

● 65   ● 130   ● 210   ●

470.064

400

300

200

100

0

NORREPORT   SKOLEBAKKEN   SCANDCENTER   BRUUNS   BUSGADEHUSET   MAG

Results are limited by 1000.

```
1 val inputPath =  "/Users/joannariascos/Desktop/algorithm/pollution.csv"
2 val pol = sqlContext.read
3         .format("com.databricks.spark.csv")
4         .option("header", "true") // Use first line of all files as header
5         .option("delimiter", ",")
6         .option("inferSchema", "true") // Automatically infer data types
7         .load(inputPath)
8
9         pol.registerTempTable("pol")
```

inputPath: String = /Users/joannariascos/Desktop/algorithm/pollution.csv
pol: org.apache.spark.sql.DataFrame = [ozone: int]
warning: there was one deprecation warning; re-run with -deprecation for details