Zeppelin

Graph2

default

FINISHED

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

// Load data - adjust the path to the location of your data
val inputPath =  "/Users/neha/Documents/Capstone/pollution*"
val data1 = sqlContext.read
        .format("com.databricks.spark.csv")
        .option("header", "true") // Use first line of all files as header
        .option("delimiter", ",")
        .option("inferSchema", "true") // Automatically infer data types
        .load(inputPath)
data1.toDF().registerTempTable("data2")
```

```
import org.apache.spark.sql.functions._
import org.joda.time.format.DateTimeFormat
inputPath: String = /Users/neha/Documents/Capstone/pollution*
data1: org.apache.spark.sql.DataFrame = [ozone: int, particullate_matter: int ... 6 more field
s]
warning: there was one deprecation warning; re-run with -deprecation for details
```

Took 23 sec. Last updated by anonymous at April 16 2017, 5:08:07 PM.

FINISHED

```
%pyspark
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns

import StringIO
def show(p):
  img = StringIO.StringIO()
  p.savefig(img, format='svg')
  img.seek(0)
  print "%html " + img.buf
```

Took 0 sec. Last updated by anonymous at April 16 2017, 5:32:29 PM.

FINISHED

```
%pyspark
DF = sqlContext.sql("select avg(ozone) as avg_ozone, avg(particullate_matter) as particulate,
, avg(nitrogen_dioxide) as nitrogen from data2 group by longitude,latitude")
data = DF.toPandas()
```

Graph2

Took 2 sec. Last updated by anonymous at April 1 2017, 5:32:38 PM. Untitled Untitled Untitled Untitled Untitled Untitled Untitled Untitled Untitled

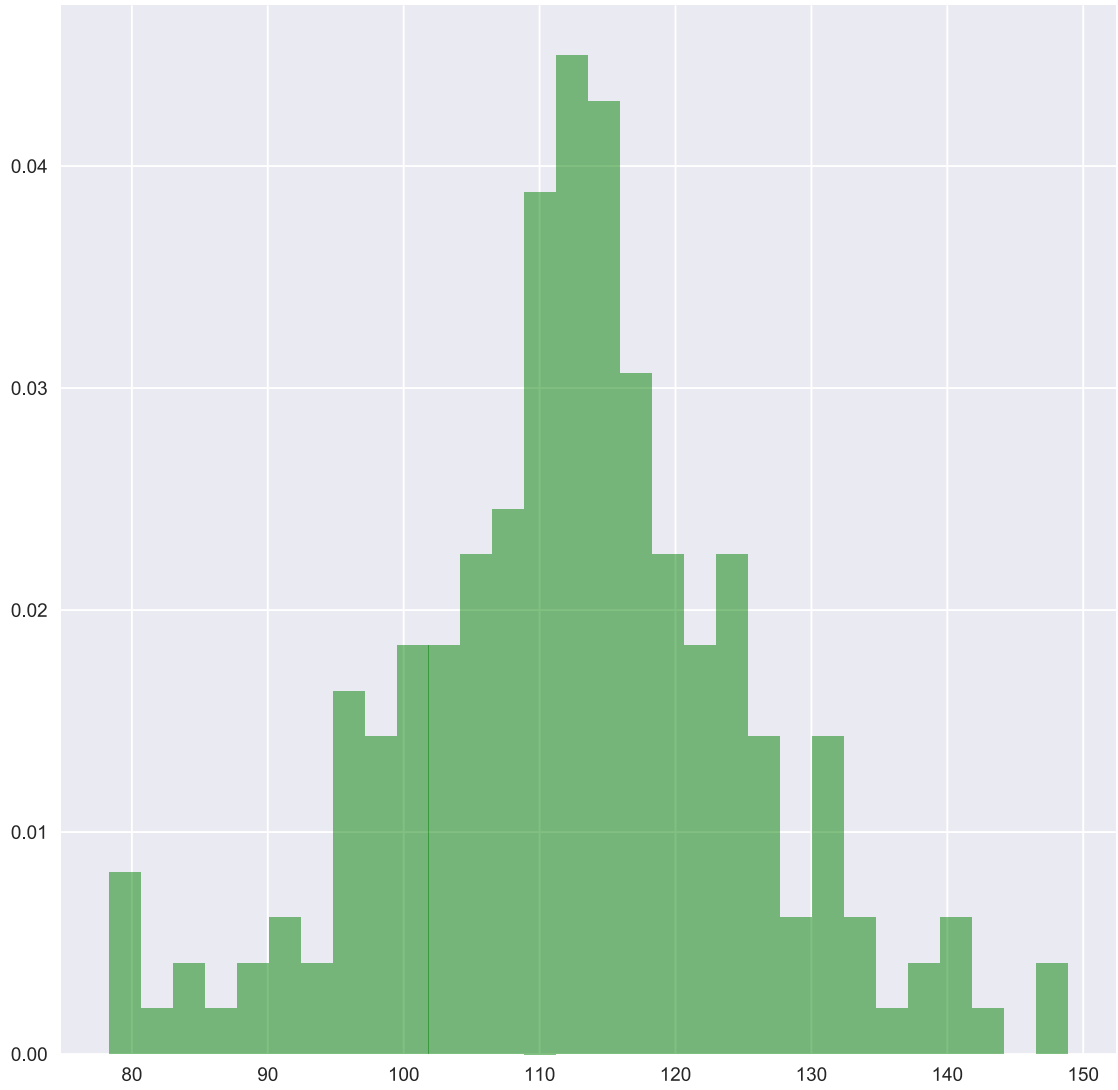
Zeppelin

Graph2

```
%pyspark
import matplotlib.pyplot as plt1
plt.clf()
num_bins = 30
#counts, bins, patches = plt.hist('avg_ozone', num_bins)
n, bins, patches = plt.hist('avg_ozone', num_bins, normed=1, facecolor='green', alpha=0.5, da
show(plt)
```

FINISHED

default



Took 1 sec. Last updated by anonymous at April 16 2017, 5:33:06 PM.

Graph2 Untitled Untitled Untitled Untitled Untitled Untitled Untitled Untitled Untitled

```
%pyspark

import matplotlib.pyplot as plt1
```

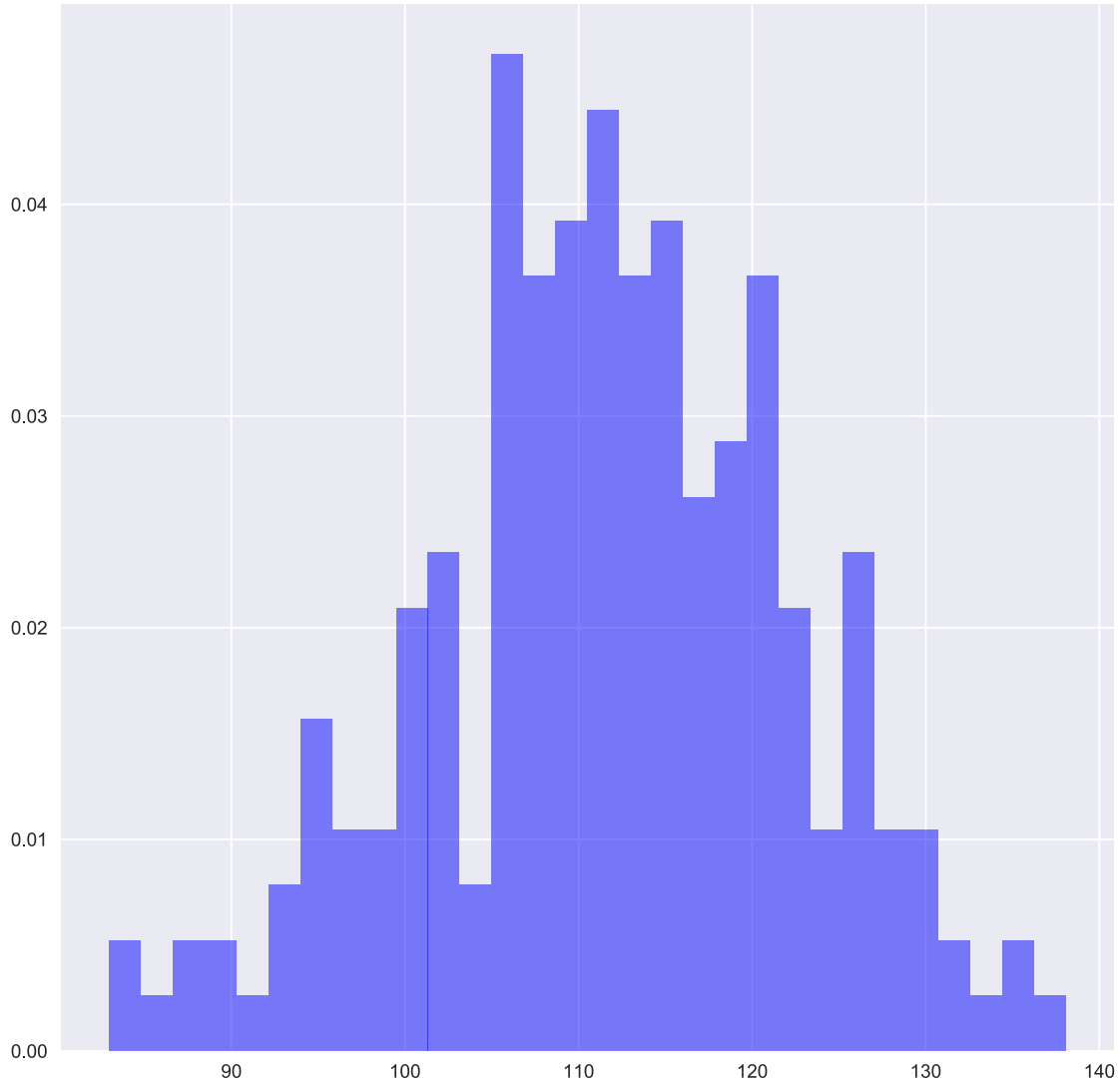
FINISHED

```
plt.clf()
num_bins = 30
n, bins, patches = plt.hist('particulate', num_bins, normed=1, facecolor='blue', alpha=0.5, d
show(plt)
```

Zeppelin

Graph2

default



Took 0 sec. Last updated by anonymous at April 16 2017, 7:53:05 PM.

FINISHED

```
%pyspark

import matplotlib.pyplot as plt1
plt1.clf()Untitled Untitled Untitled Untitled Untitled Untitled Untitled Untitled Untitled
num_bins = 30
#data.hist('monoxide',num_bins)
```

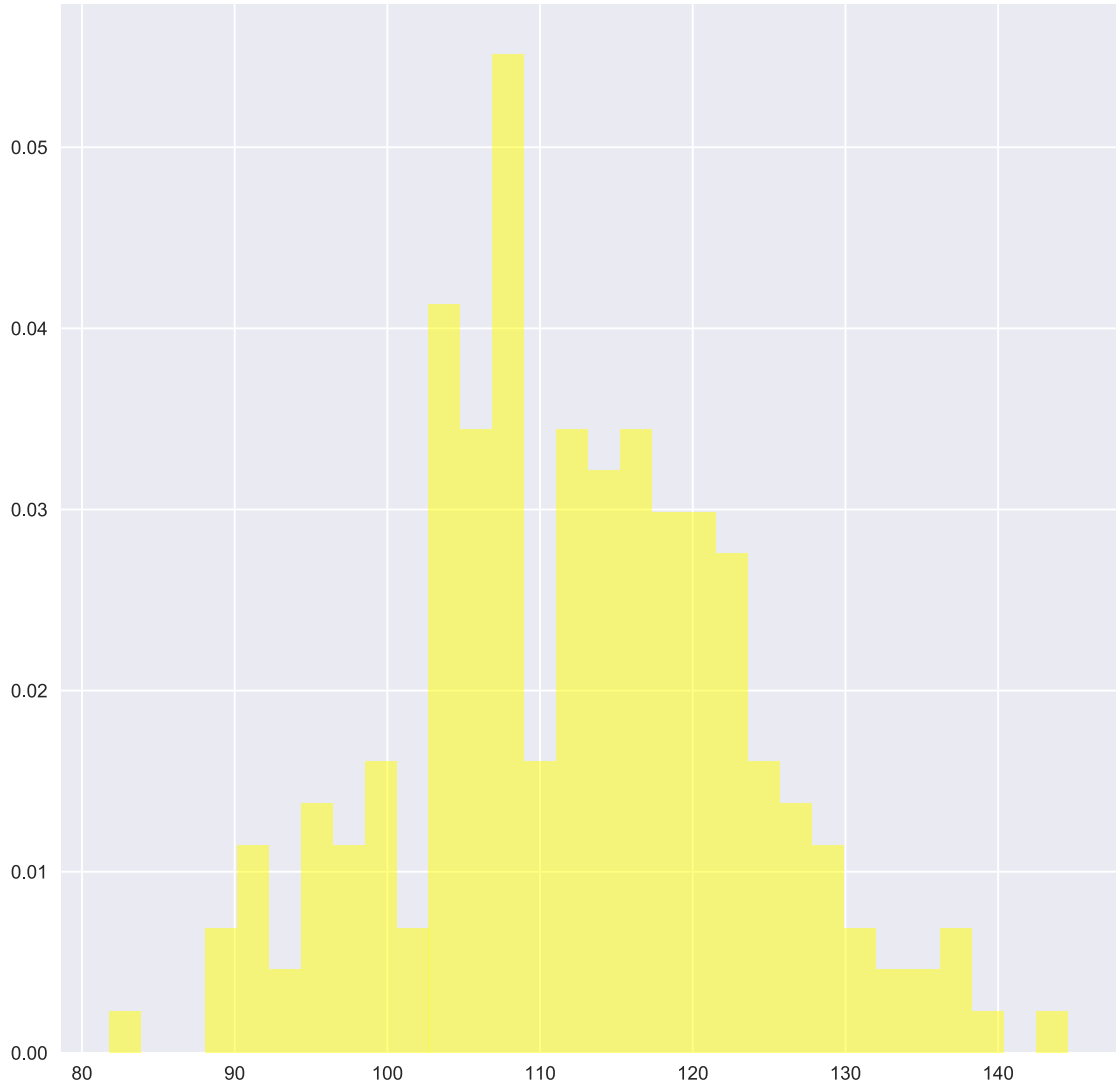
Graph2

```
n, bins, patches = plt.hist('monoxide', num_bins, normed=1, facecolor='yellow', alpha=0.5, da
```

Zeppelin

Graph2

default



Took 0 sec. Last updated by anonymous at April 16 2017, 7:50:28 PM.

FINISHED

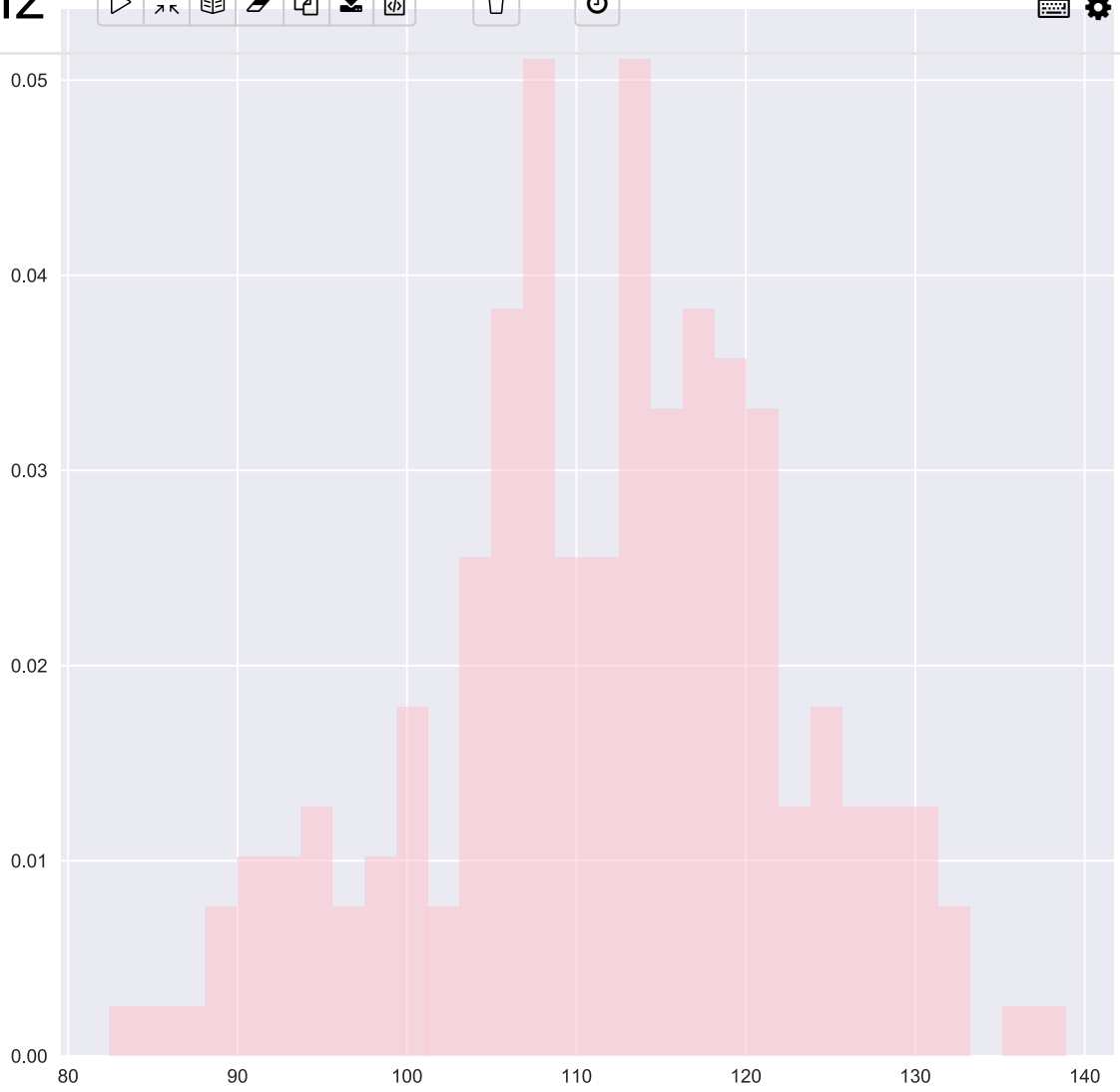
```
%pyspark

import matplotlib.pyplot as plt1
plt.clf()
num_bins = 30
#data.hist('sulfur',num_bins)
n, bins, patches = plt.hist('sulfur', num_bins, normed=1, facecolor='pink', alpha=0.5, data =
show(plt)
```

Graph2

Untitled Untitled Untitled Untitled Untitled Untitled Untitled Untitled Untitled

Zeppelin

Graph2



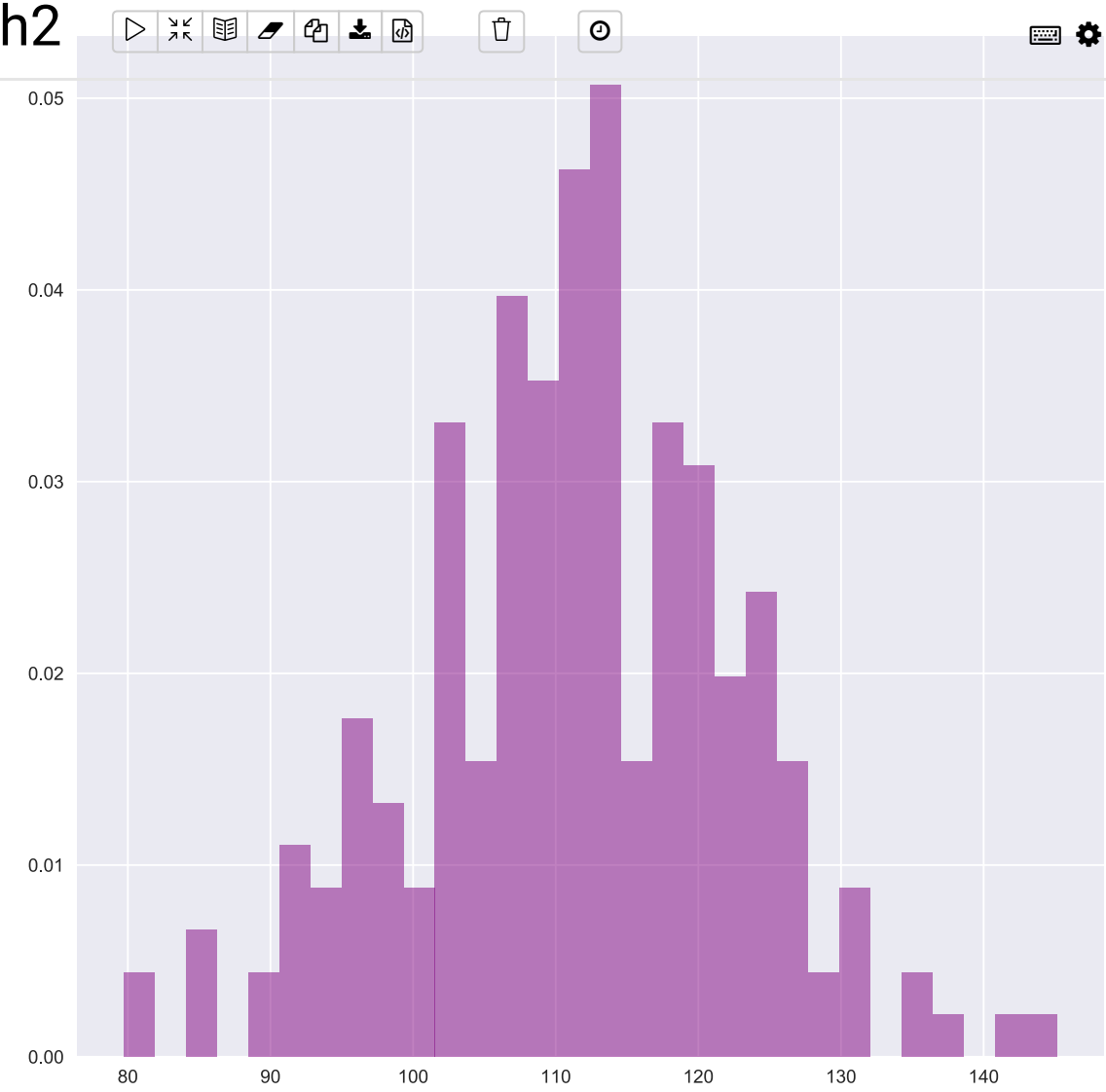
Took 0 sec. Last updated by anonymous at April 16 2017, 7:52:47 PM.

FINISHED

```
%pyspark

import matplotlib.pyplot as plt1
plt.clf()
num_bins = 30
#data.hist('nitrogen',num_bins)
n, bins, patches = plt.hist('nitrogen', num_bins, normed=1, facecolor='purple', alpha=0.5, da
show(plt)
```

Zeppelin

Graph2



Took 1 sec. Last updated by anonymous at April 16 2017, 7:52:18 PM.

READY