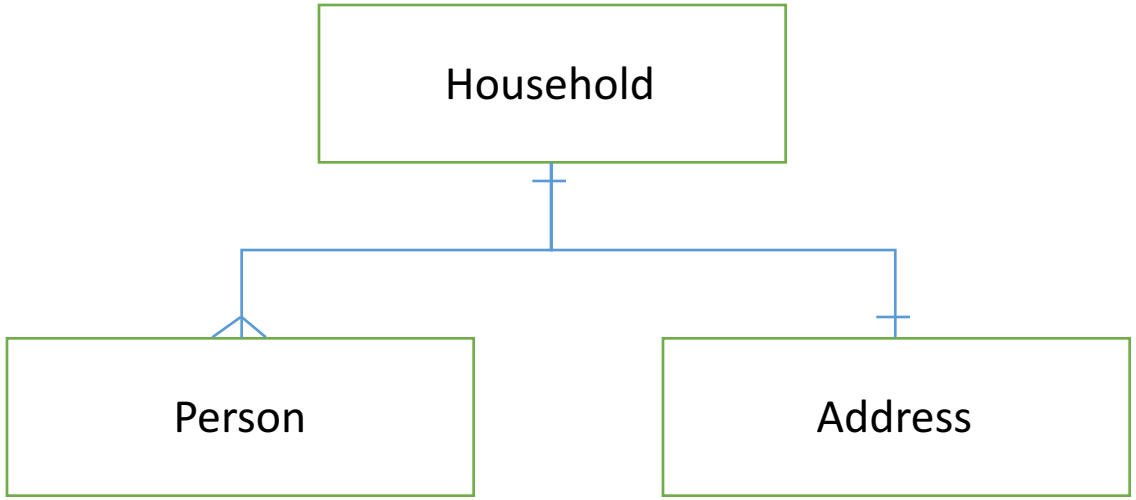
Database
&
Data Visualization

Agenda

- Databases
- Midterm
- Matplotlib
- Tableau

A Model: Entity Relations



- A “Household” has one or more “Person”
- A “Household” has exactly 1 ”Address”
- An “Address” belongs to exactly 1 “Household”

Creating a database

```
import sqlite3
query = """
    CREATE TABLE test
        (a VARCHAR(20), b VARCHAR(20),
         c REAL, d INTEGER );"""
con = sqlite3.connect(':memory:')
con.execute(query)
con.commit()
```

Inserting data to a database

```
data = [('Atlanta', 'Georgia', 1.25, 6),
('Tallahassee', 'Florida', 2.6, 3),
('Sacramento', 'California', 1.7, 5)]

stmt = "INSERT INTO test VALUES(?, ?, ?, ?)"

con.executemany(stmt, data) con.commit()
```

Finding data in a database

```
cursor = con.execute('select * from test')

rows = cursor.fetchall()
```

In [580]: rows
Out[580]:
[(u'Atlanta', u'Georgia', 1.25, 6),
(u'Tallahassee', u'Florida', 2.6, 3), (u'Sacramento', u'California', 1.7, 5)]

Creating a DataFrame

```
rows = cursor.fetchall()
```

In [580]: rows
Out[580]:
[(u'Atlanta', u'Georgia', 1.25, 6),
(u'Tallahassee', u'Florida', 2.6, 3), (u'Sacramento', u'California', 1.7, 5)]

Finding data in a database

```
DataFrame(rows, columns=zip(*cursor.description)[0])
```

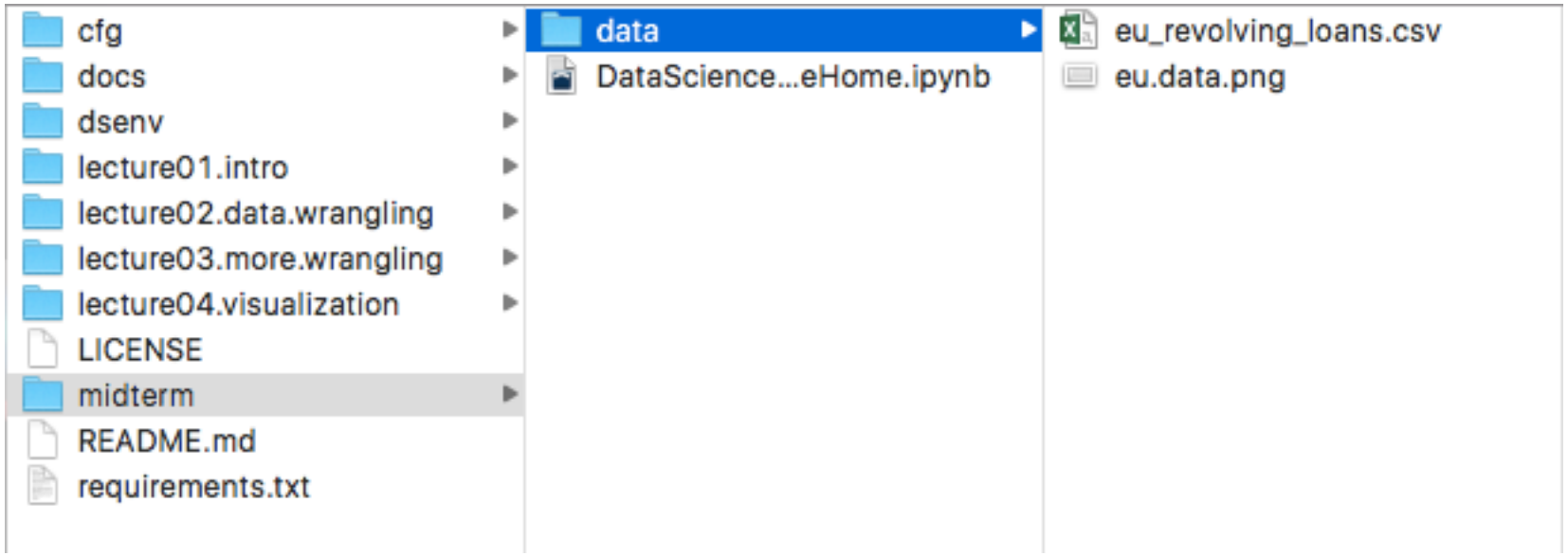
```
In [582]: DataFrame(rows, columns=zip(*cursor.description)[0])
Out[582]:
             a           b     c  d
0      Atlanta     Georgia  1.25  6
1  Tallahassee     Florida  2.60  3
2   Sacramento  California  1.70  5
```

Agenda

- Databases
- **Midterm**
- Matplotlib
- Tableau

Midterm – 2 parts

- In-class (20 questions, 1 hour)
- Take-home (1 dataset)



Agenda

- Databases
- Midterm
- **Matplotlib**
- Tableau

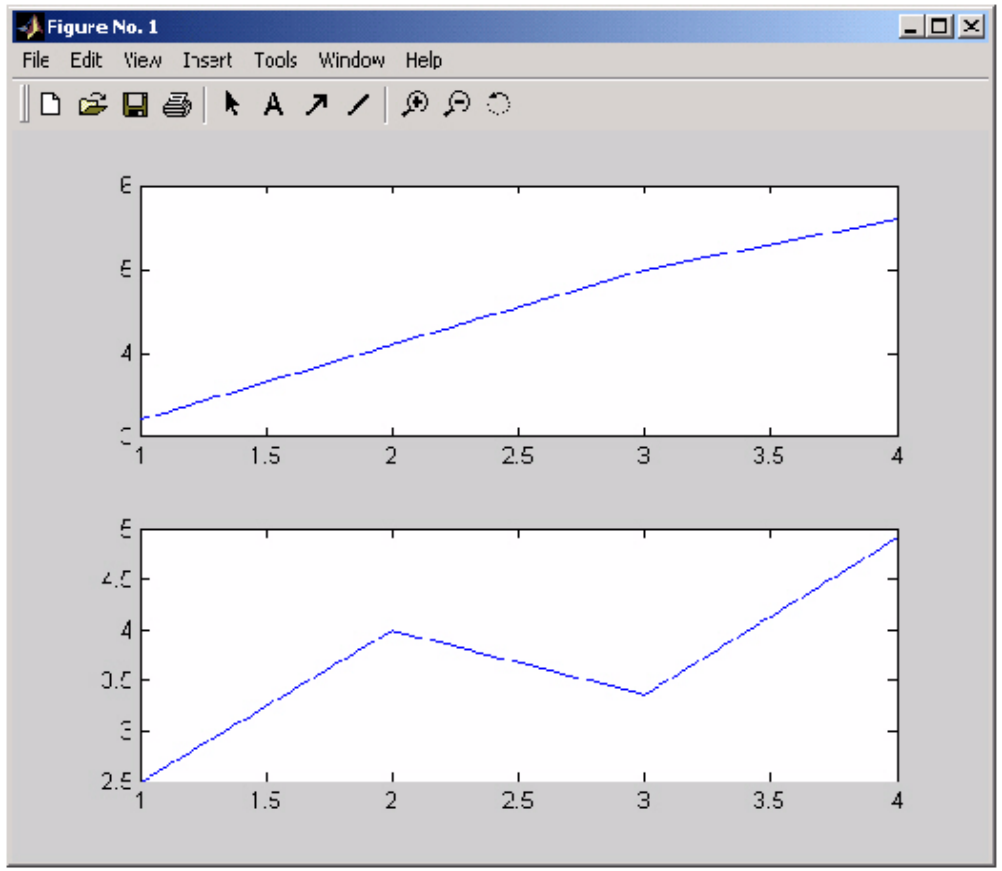
Matplotlib

- **Matplotlib** is a python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms.
- **Matplotlib** is the whole package; **pylab** is a module in matplotlib that gets installed alongside matplotlib; and **matplotlib.pyplot** is a module in matplotlib
- **Pyplot** provides the state-machine interface to the underlying plotting library in matplotlib.

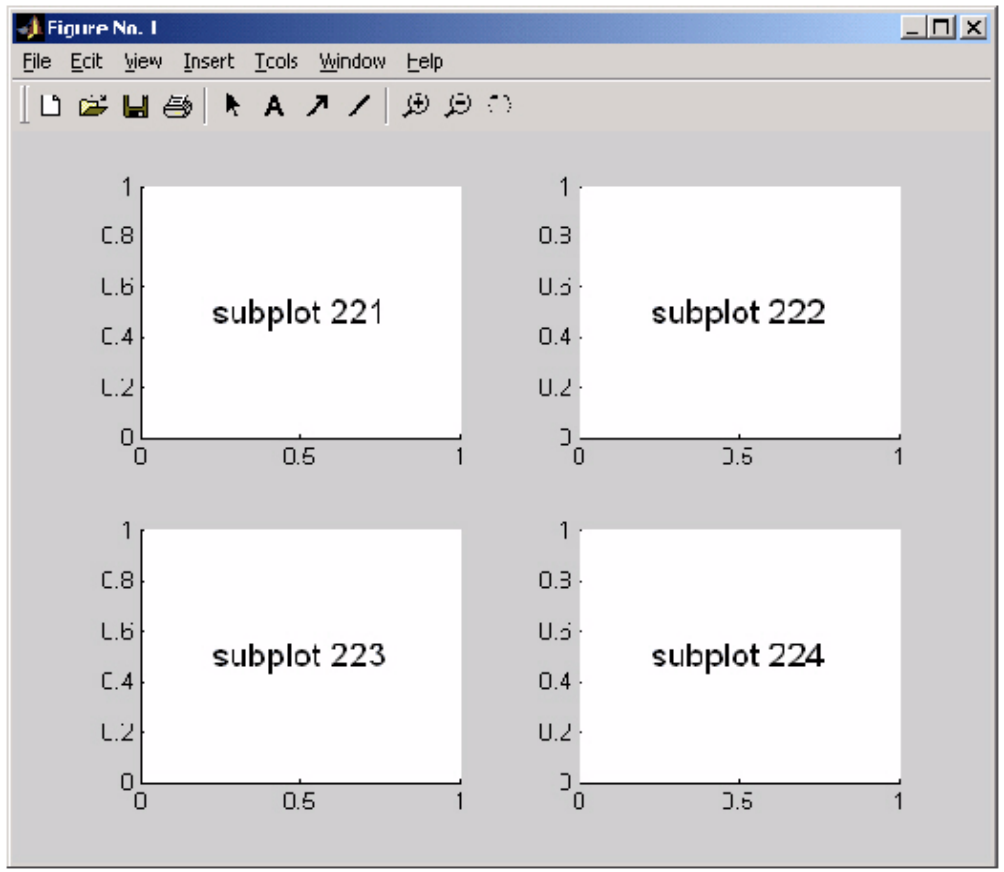
Subplot

- Subplot divides the current figure into rectangular panes that are numbered row wise.
- Each pane contains an axes object. Subsequent plots are output to the current pane.

Subplot grid 2X1



Subplot grid 2x2



Matplotlib Gallery

http://matplotlib.org/examples/index.html

Agenda

- Databases
- Midterm
- Matplotlib
- **Tableau**

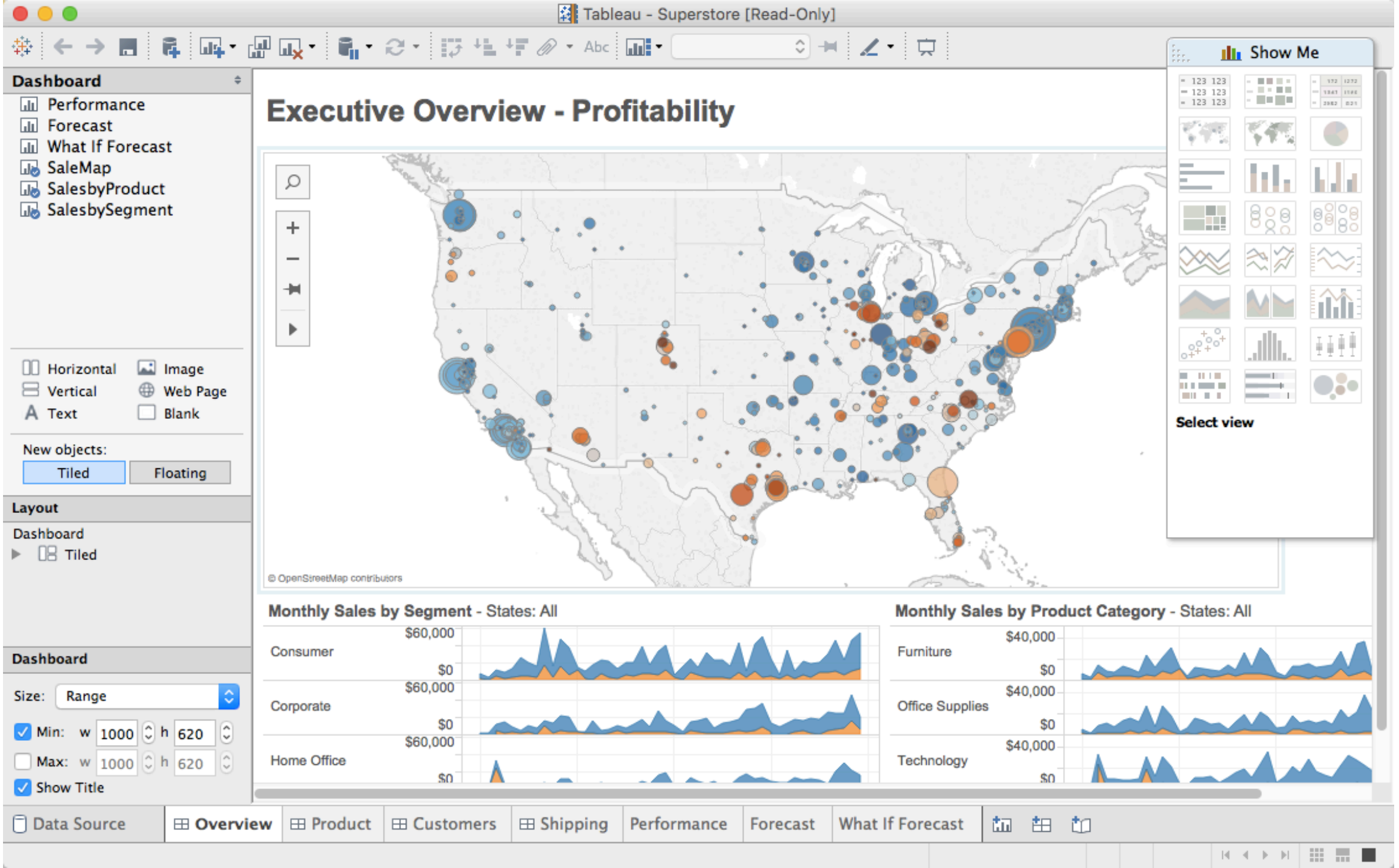


Tableau - Superstore [Read-Only]
Dashboard
Performance
Forecast
What If Forecast
SaleMap
SalesbyProduct
SalesbySegment
Horizontal
Vertical
Text
Image
Web Page
Blank
New objects:
Tiled
Floating
Layout
Dashboard
Tiled
Dashboard
Size:
Range
Min: w 1000 h 620
Max: w 1000 h 620
Show Title
Executive Overview - Profitability
© OpenStreetMap contributors
Show Me
Select view
Monthly Sales by Segment - States: All
Consumer
Corporate
Home Office
$60,000
$0
Monthly Sales by Product Category - States: All
Furniture
Office Supplies
Technology
$40,000
$0
Data Source
Overview
Product
Customers
Shipping
Performance
Forecast
What If Forecast

1 Year Tableau License

http://tableau.com/students