

Hassle-Free Data Science Apps with Bokeh



ANACONDA[™]

Presenters



Peter Wang is the CTO and Co-founder of Continuum Analytics and the creator of Bokeh.

He has been developing commercial scientific computing and visualization software for over 15 years.

As a creator of the PyData conference, he devotes time and energy to growing the Python data community, and advocating and teaching Python at conferences worldwide.



Bryan Van de Ven is the lead developer on the Bokeh project.

He holds an undergraduate degree in Computer Science & Mathematics from UT Austin, and a Masters degree in Physics from UCLA.

Previously Bryan developed data exploration and visualization software for sonar feature detection, financial risk modeling, and fluid mixing simulation.

Overview

- What is Bokeh?
- Overview and tour of major features
- Demo 1: Scikit-learn clustering
- Demo 2: Gapminder
- Demo 3: Streaming data
- Really big data: Preview of data shading
- Q&A

Overview of Anaconda



ANACONDA[™]

ANACONDA[®] is....

the modern open source analytics platform
powered by Python

the fastest growing open data science language

- Easy to Build, Maintain & Deploy Analytics
- Talks with Everything, Runs Anywhere
- High Performance, Scalable Analytics

Anaconda

Accelerating Adoption of Python for Enterprises

ANACONDA[®]

ENTERPRISE DATA INTEGRATION

with optimized connectors & out-of-core processing

NumPy &
Pandas

Numba

PERFORMANCE

with compiled Python for lightning fast execution

COLLABORATIVE NOTEBOOKS

with publication, authentication, & search

Jupyter/
IPython

Bokeh

VISUAL APPS

for interactivity, streaming, & Big

PYTHON & PACKAGE MANAGEMENT

for Hadoop & Apache stack

Spark

Conda

SECURE & ROBUST REPOSITORY

of data science libraries, scripts, & notebooks

Anaconda for Data Science

Empowering Everyone on the Team

Data Scientist

- Advanced analytics with Python & R
- Simplified library management
- Easily share data science notebooks & packages

Developer

- Support for common APIs & data formats
- Common language with data scientists
- Python extensibility with C, C++, etc.

Ops

- Validated source of up-to-date packages including indemnification
- Agile Enterprise Package Management
- Supported across platforms



Data Engineer

- Powerful & efficient libraries for data transformations
- Robust processing for noisy dirty data
- Support for common APIs & data formats

Business Analyst

- Collaborative interactive analytics with notebooks
- Rich browser based visualizations
- Powerful MS Excel integration

Computational Scientist

- Rich set of advanced analytics
- Trusted & production ready libraries for numerics
- Simplified scale up & scale out on clusters & GPUs

Modern Analytics Stack

APP	Notebooks		Embeddable Dashboards		Data Services		Visual Apps		
VIZ	Plots	Interactive Viz		Big Data	Maps & GIS	3D	Streaming	Graphs	
STORYBOARD	Notebooks		Interactive Exploration		Visual Programming		Data IDEs		
ANALYTICS	Data Prep		Stats	ML & Ensembles		Deep Learning		Simulation & Optimization	
	Geospatial		Text & NLP		Video/Image/Audio Mining		Graph & Network		
DATA	Hadoop & Hive		Spark	NoSQL	DW & SQL		Files & Web Services		
HW	Servers		Clusters	GPUs & High End Workstations					

Write Once, Deploy Anywhere



Bokeh Overview & Tour



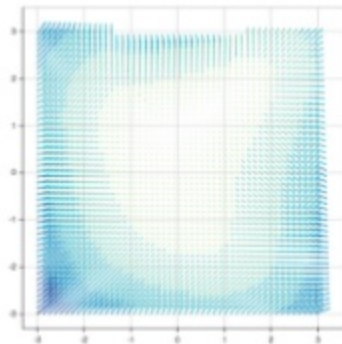
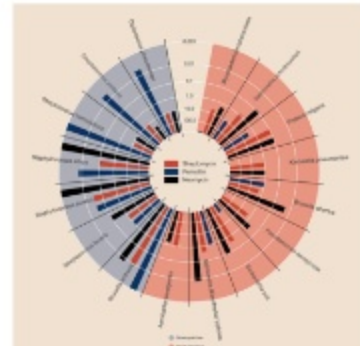
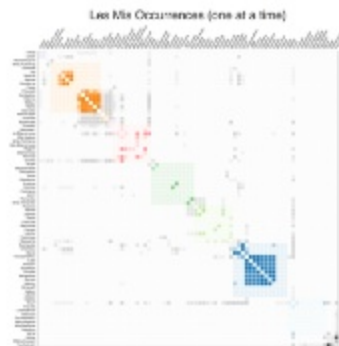
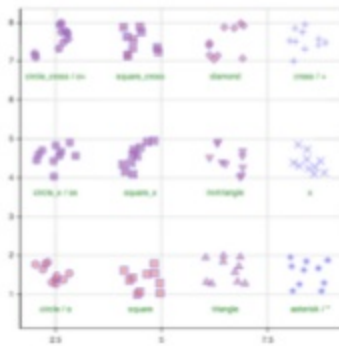
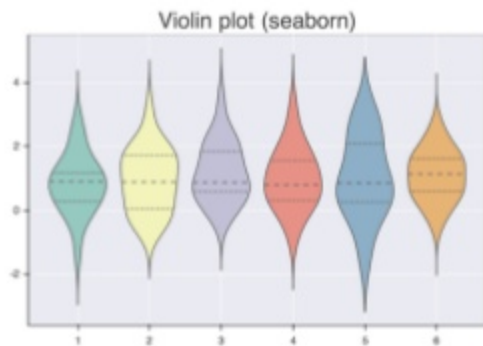
ANACONDA[™]

Bokeh



- Interactive visualization
- Novel graphics
- Streaming, dynamic, large data
- For the browser, with or without a server
- No need to write Javascript

<http://bokeh.pydata.org>

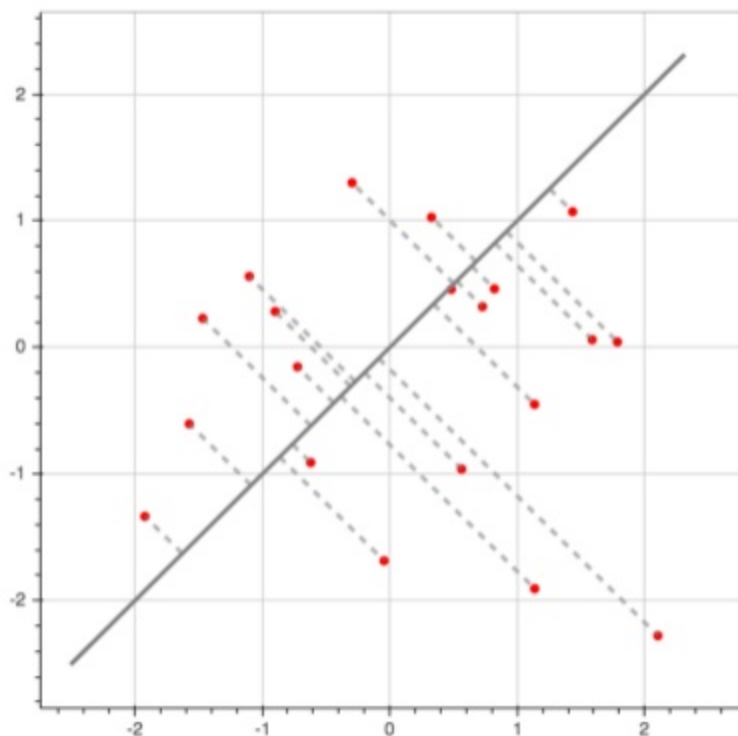


Periodic Table

1 H Hydrogen (1.008)																	79 Au Gold (196.967)					
2 He Helium (4.003)	3 Li Lithium (6.941)	4 Be Beryllium (9.012)																	80 Hg Mercury (200.598)			
5 B Boron (10.811)	6 C Carbon (12.011)	7 N Nitrogen (14.007)	8 O Oxygen (15.999)	9 F Fluorine (18.998)	10 Ne Neon (20.180)																	81 Tl Thallium (204.387)
11 Na Sodium (22.990)	12 Mg Magnesium (24.305)																	82 Pb Lead (207.2)				
13 Al Aluminum (26.982)	14 Si Silicon (28.086)	15 P Phosphorus (30.974)	16 S Sulfur (32.065)	17 Cl Chlorine (35.453)	18 Ar Argon (39.948)																	83 Bi Bismuth (208.980)
19 K Potassium (39.098)	20 Ca Calcium (40.078)	21 Sc Scandium (44.956)	22 Ti Titanium (47.883)	23 V Vanadium (50.942)	24 Cr Chromium (51.996)	25 Mn Manganese (54.938)	26 Fe Iron (55.845)	27 Co Cobalt (58.933)	28 Ni Nickel (58.693)	29 Cu Copper (63.546)	30 Zn Zinc (65.38)	31 Ga Gallium (69.723)	32 Ge Germanium (72.64)	33 As Arsenic (74.922)	34 Se Selenium (78.96)	35 Br Bromine (79.904)	36 Kr Krypton (83.798)					
37 Rb Rubidium (85.468)	38 Sr Strontium (87.62)	39 Y Yttrium (88.906)	40 Zr Zirconium (91.224)	41 Nb Niobium (92.906)	42 Mo Molybdenum (95.94)	43 Tc Technetium (98.906)	44 Ru Ruthenium (101.07)	45 Rh Rhodium (102.91)	46 Pd Palladium (106.36)	47 Ag Silver (107.87)	48 Cd Cadmium (112.41)	49 In Indium (114.82)	50 Sn Tin (118.71)	51 Sb Antimony (121.76)	52 Te Tellurium (127.6)	53 I Iodine (126.91)	54 Xe Xenon (131.29)					
55 Cs Cesium (132.91)	56 Ba Barium (137.33)	57 La Lanthanum (138.91)	58 Ce Cerium (140.12)	59 Pr Praseodymium (140.91)	60 Nd Neodymium (144.24)	61 Pm Promethium (144.91)	62 Sm Samarium (150.36)	63 Eu Europium (151.96)	64 Gd Gadolinium (157.25)	65 Tb Terbium (158.93)	66 Dy Dysprosium (162.50)	67 Ho Holmium (164.93)	68 Er Erbium (167.26)	69 Tm Thulium (168.93)	70 Yb Ytterbium (173.05)	71 Lu Lutetium (174.967)	72 Hf Hafnium (178.49)					
73 Ta Tantalum (180.948)	74 W Tungsten (183.84)	75 Re Rhenium (186.207)	76 Os Osmium (190.23)	77 Ir Iridium (192.22)	78 Pt Platinum (195.084)	79 Au Gold (196.967)	80 Hg Mercury (200.598)	81 Tl Thallium (204.387)	82 Pb Lead (207.2)	83 Bi Bismuth (208.980)	84 Po Polonium (209)	85 At Astatine (210)	86 Rn Radon (222)									
87 Fr Francium (223)	88 Ra Radium (226)	89 Ac Actinium (227)	90 Th Thorium (232)	91 Pa Protactinium (231)	92 U Uranium (238)	93 Np Neptunium (237)	94 Pu Plutonium (244)	95 Am Americium (243)	96 Cm Curium (247)	97 Bk Berkelium (247)	98 Cf Californium (251)	99 Es Einsteinium (252)	100 Fm Fermium (257)	101 Md Mendelevium (258)	102 No Nobelium (259)	103 Lr Lawrencium (262)						



Novel Graphics



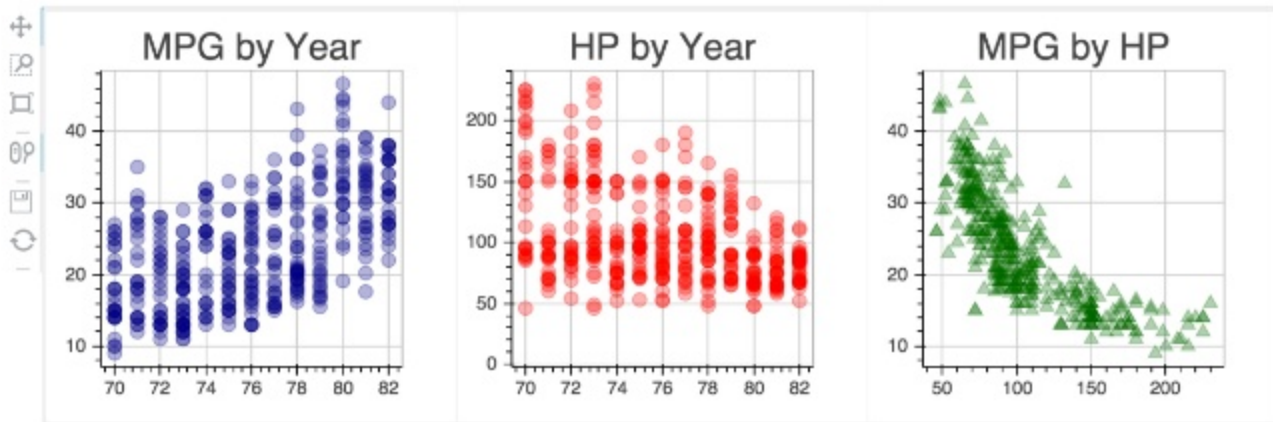
```
import numpy as np
from bokeh.plotting import *
N = 20
x = np.random.normal(size=N)
y = np.random.normal(size=N)

output_file("peter test.html", title="peter test")
figure()
hold()
scatter(x,y, marker="circle", color="red", size=6)

minval = min(x.min(), y.min()) * 1.1
maxval = max(x.max(), y.max()) * 1.1
line([minval, maxval], [minval, maxval], color="gray",
     line_width=3)

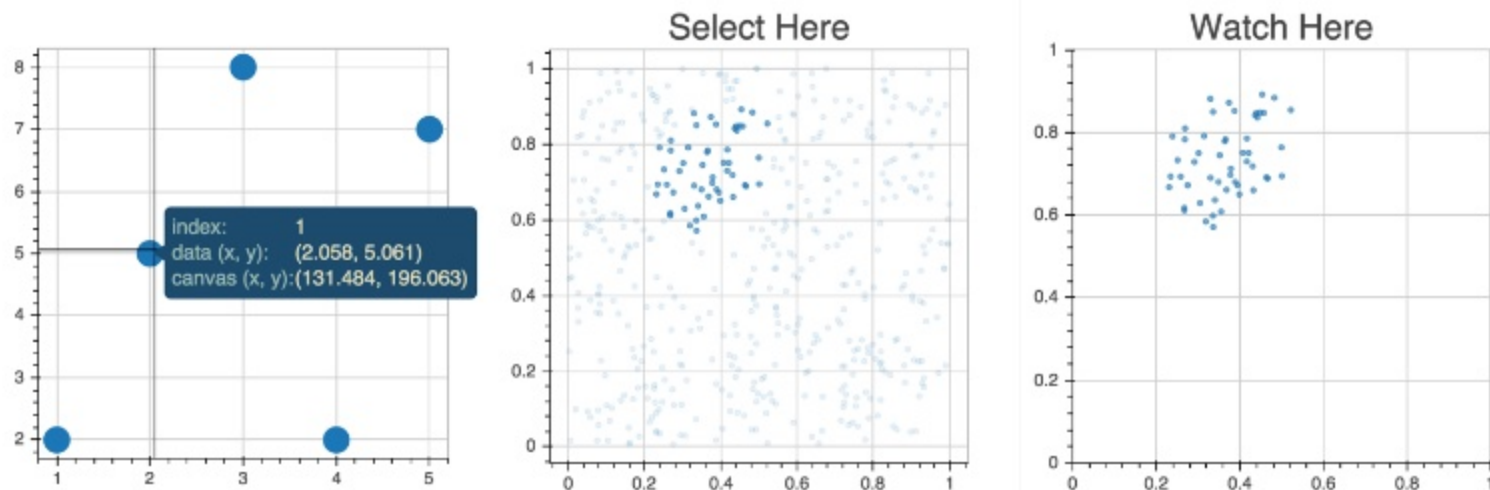
mids = (x+y)/2
segment(mids, mids, x, y, color="gray", alpha=0.6,
       line_width=3, line_dash="dashed")
show()
```


Linked Plots (Notebook 2)



- Easy to show multiple plots and link them
- Easy to link data selections between plots
- Can easily customize the kind of linkage straight from Python, without needing to fiddle around with JS

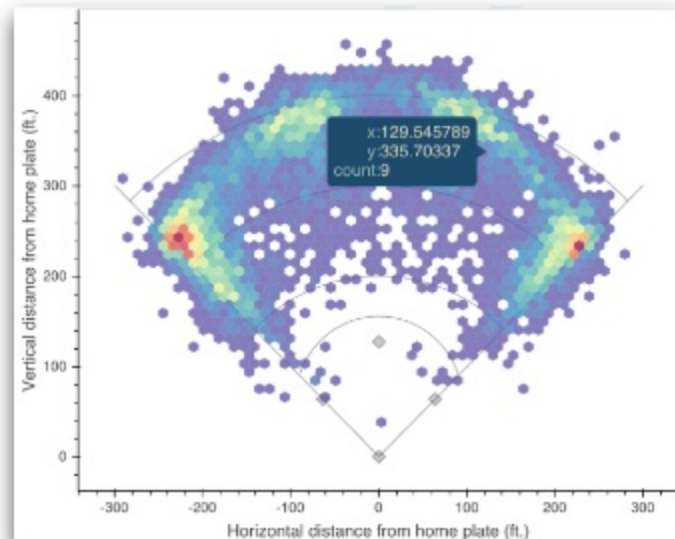
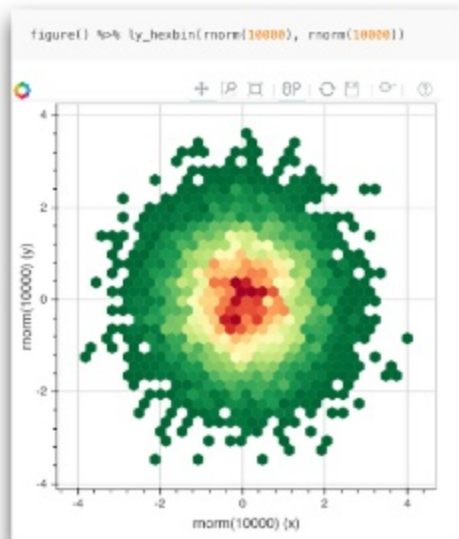
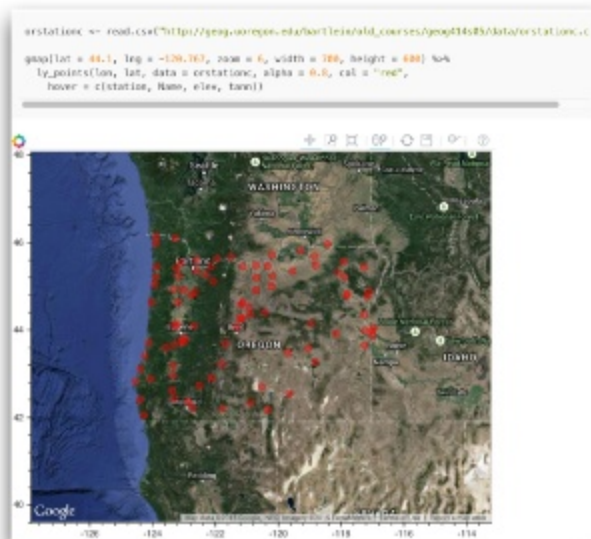
Flexible Tools (Notebook 3)



- Many useful tools with built-in functionality
- Easy to extend with Javascript, if so inclined

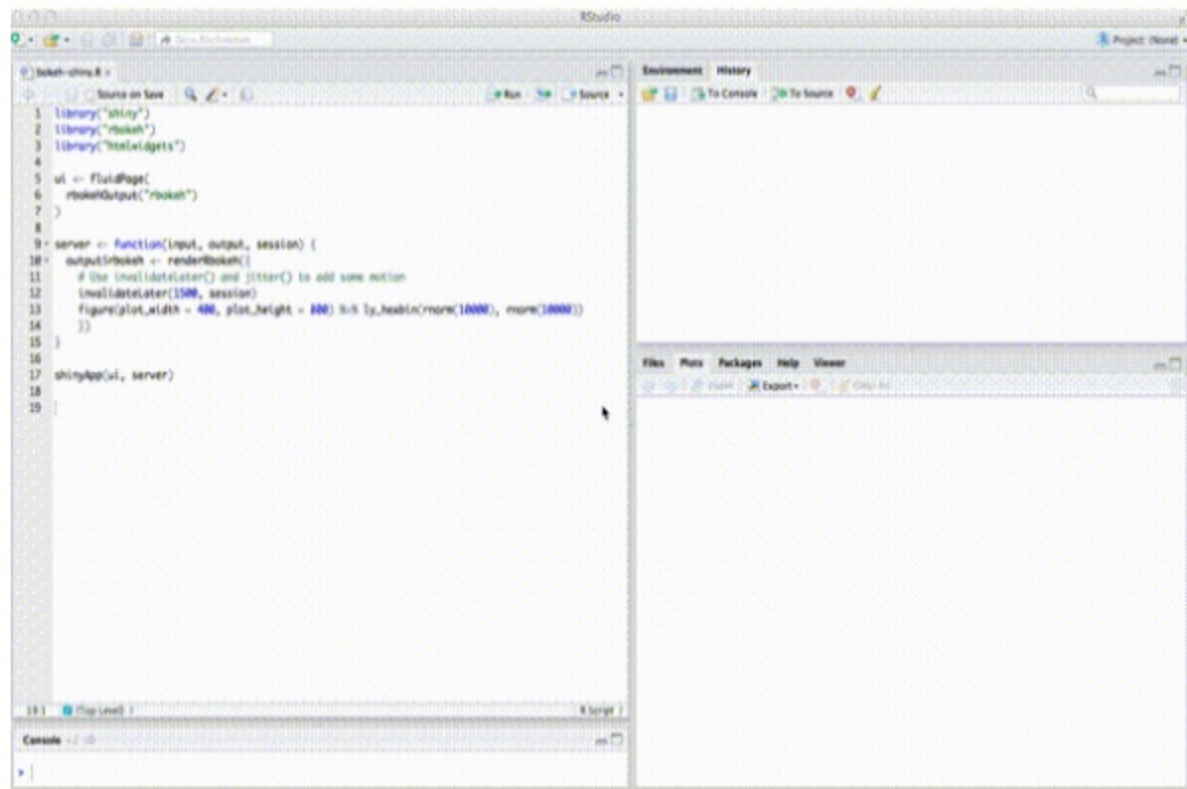
rBokeh

Plays well with R ecosystem: HTMLwidget, RMarkdown...



<http://hafen.github.io/rbokeh>

rBokeh with RStudio & Shiny

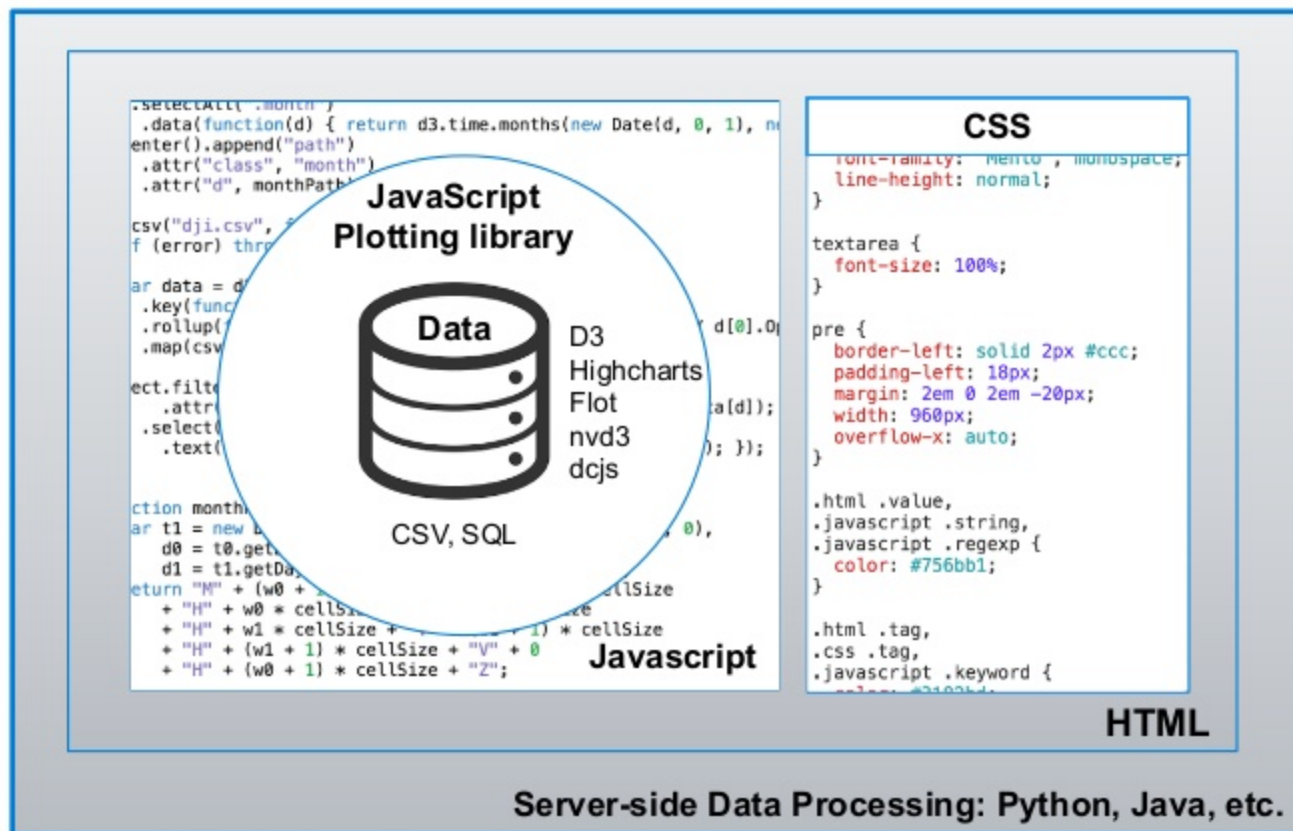


Architecture



ANACONDA[™]

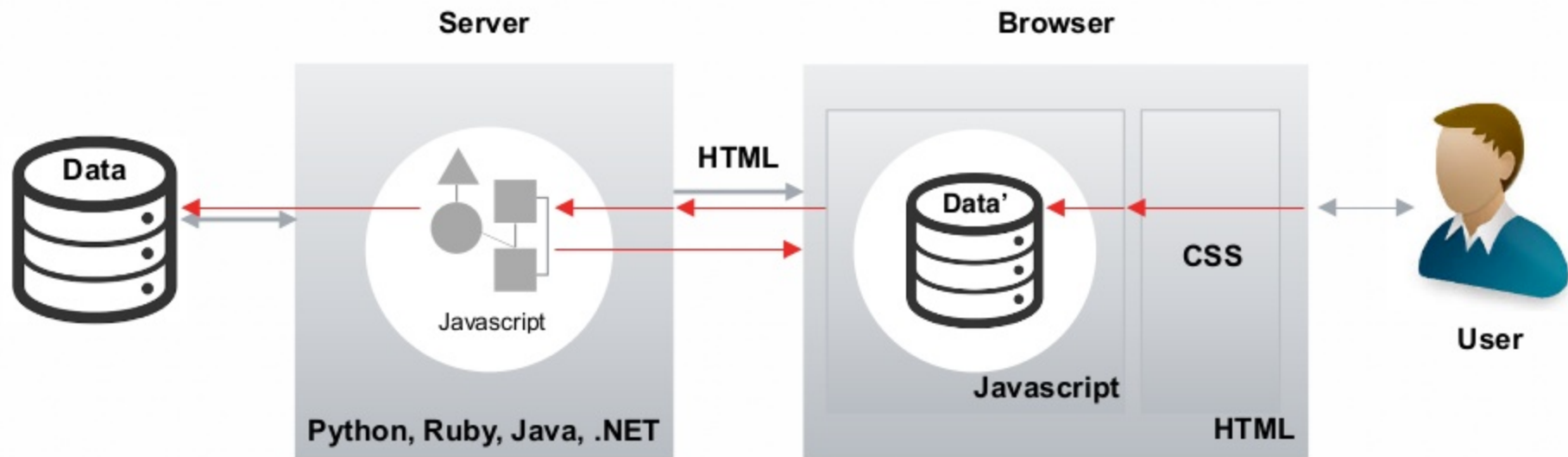
Traditional Web Visualization



Tech:

- Python/R/Java
- HTML & browser compat
- CSS/LESS/Sass
- JS plotting library API
- Javascript
 - jQuery, underscore
 - svg, canvas2D
 - webGL, three.js
 - React
 - Angular
 - node.js, browserify, gulp, grunt, npm, ...

Traditional Web Viz - Interaction



Simple dashboard: Server language generating HTML, JS, CSS styling, subset of data

Handling user interaction: Custom Javascript, calling Server endpoint, which generates updated JSON or JS that gets pushed back to client via websocket