CONTINUUM ANALYTICS

Hassle-Free Data Science Apps with Bokeh

ANACONDA

CONTINUUM ANALYTICS

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 form 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.

CONTINUUM ANALYTICS

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

CONTINUUM®
ANALYTICS

Overview of Anaconda

CONTINUUM ANALYTICS

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

CONTINUUM 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

CONTINUUM ANALYTICS

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

CONTINUUM ANALYTICS

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

CONTINUUM ANALYTICS

Write Once, Deploy Anywhere

CONTINUUM®
ANALYTICS

Bokeh Overview & Tour

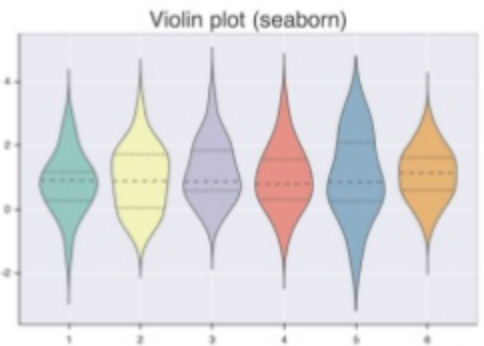
Bokeh

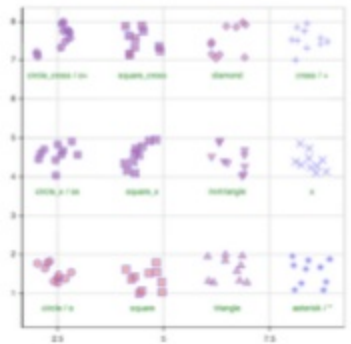
- Interactive visualization
- Novel graphics
- Streaming, dynamic, large data
- For the browser, with or without a server
- No need to write Javascript

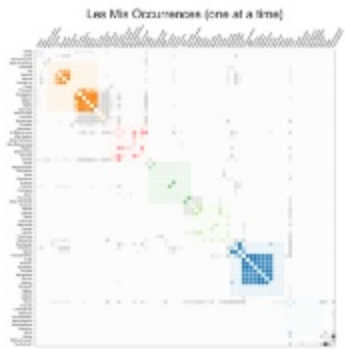
http://bokeh.pydata.org

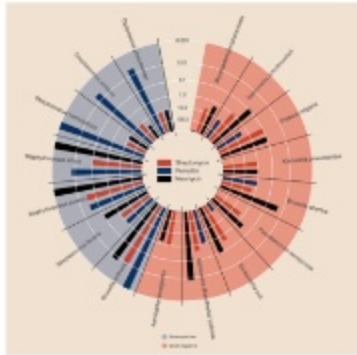
CONTINUUM ANALYTICS

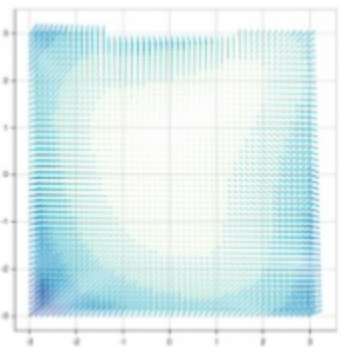
Versatile Plots

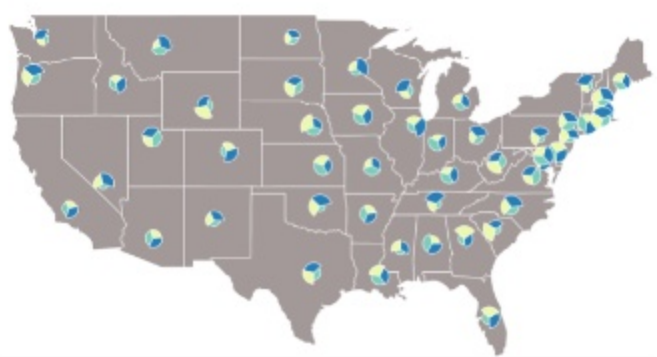




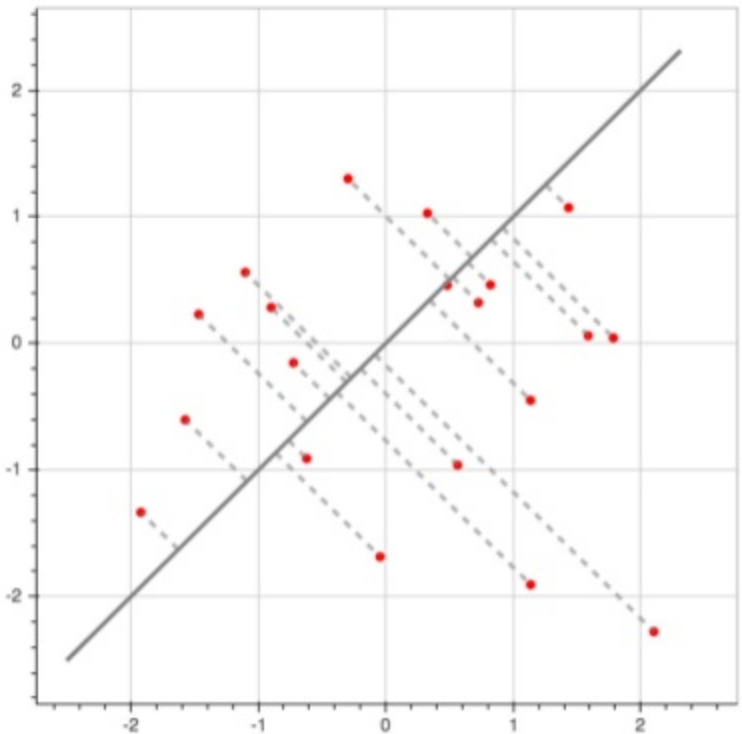






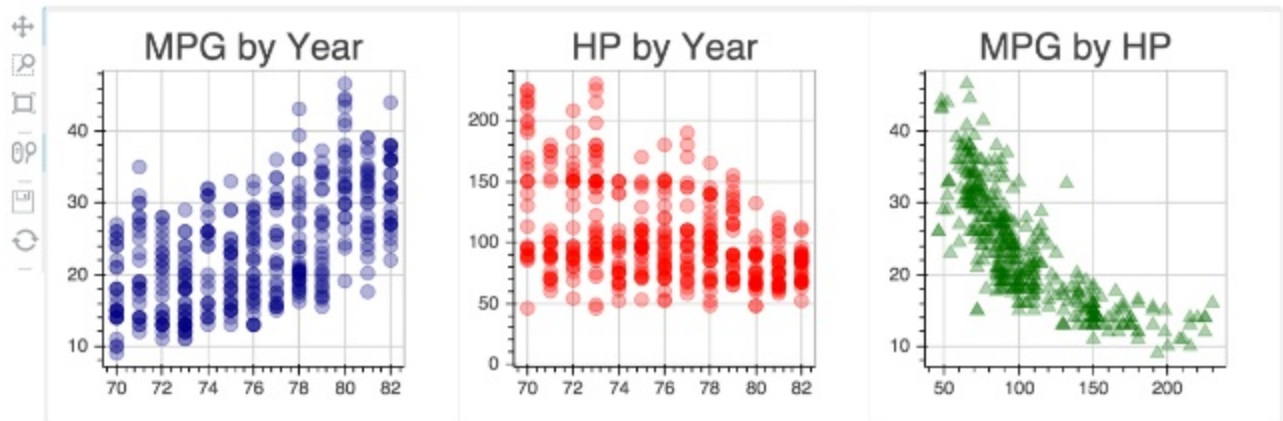


Novel Graphics



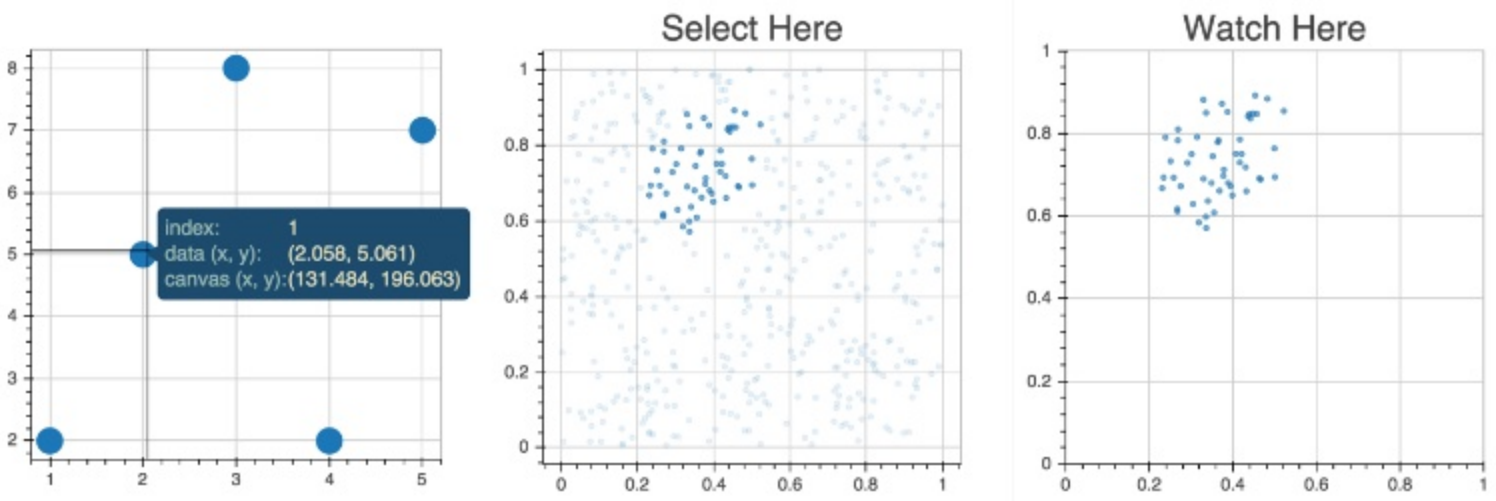

```
import numpy as np
from bokeh.plotting import *
N = 20
x = np.random.normal(size=N)
y = np.random.normal(size=N)

output_file("petertest.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

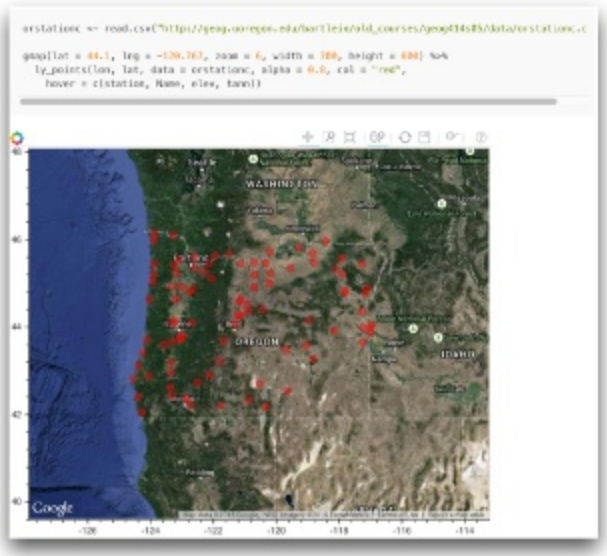
Flexible Tools (Notebook 3)

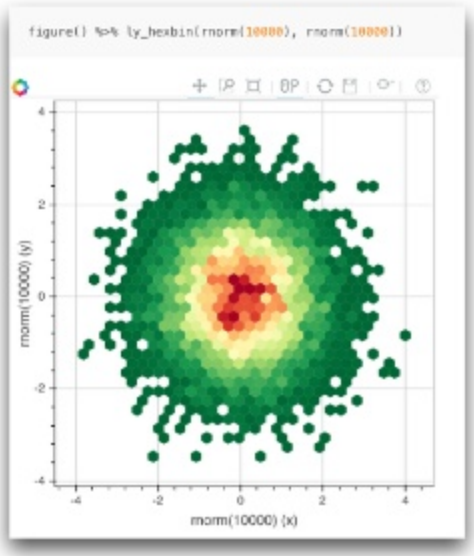


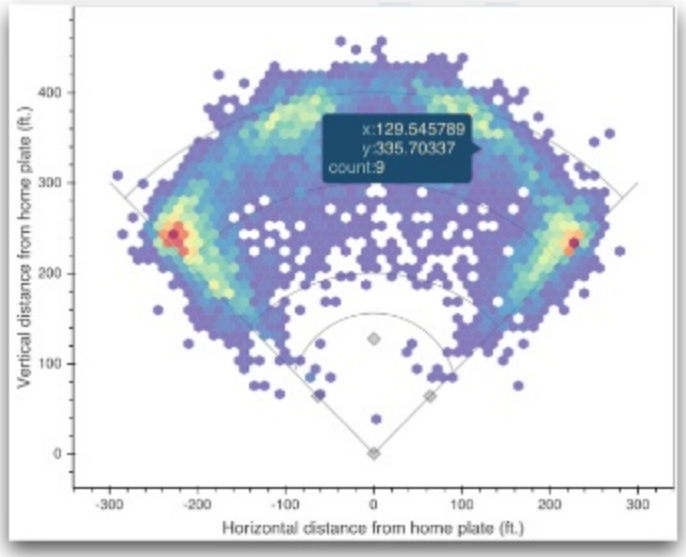
- Many useful tools with built-in functionality
- Easy to extend with Javascript, if so inclined

rBokeh

Plays well with R ecosystem: HTMLwidget, RMarkdown…



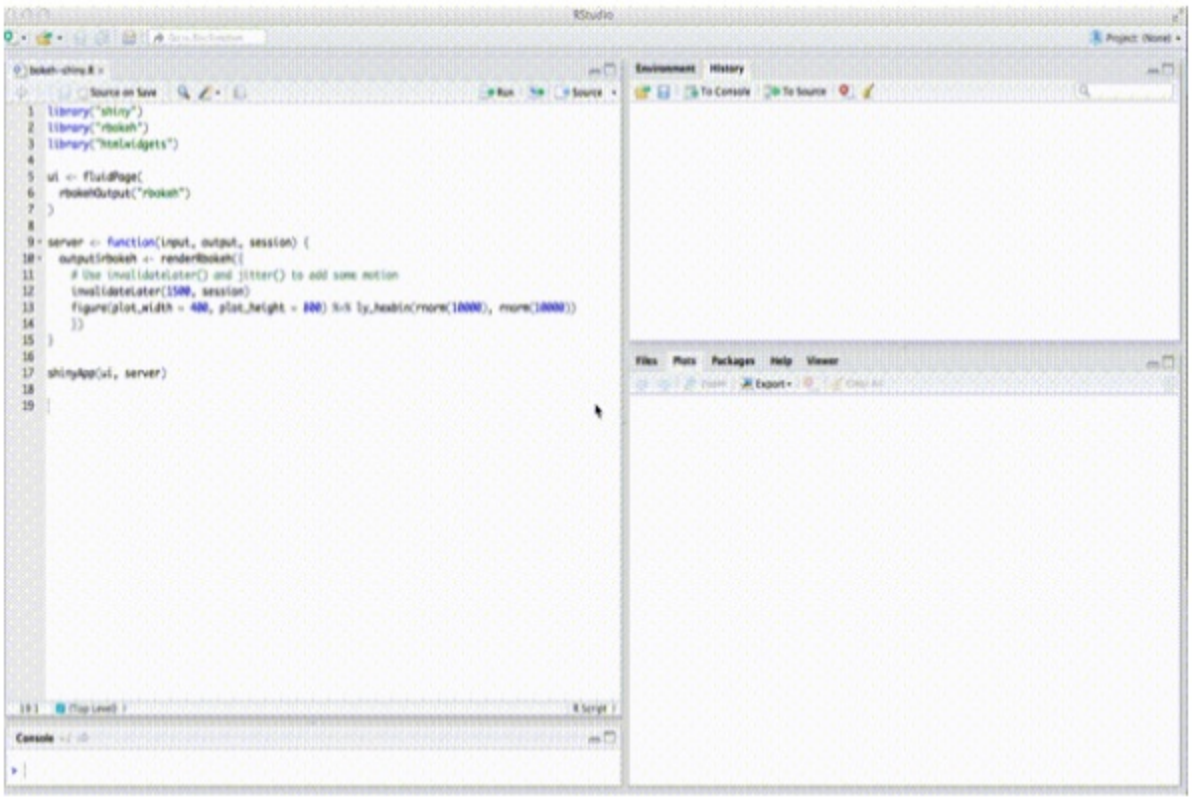




http://hafen.github.io/rbokeh

CONTINUUM ANALYTICS

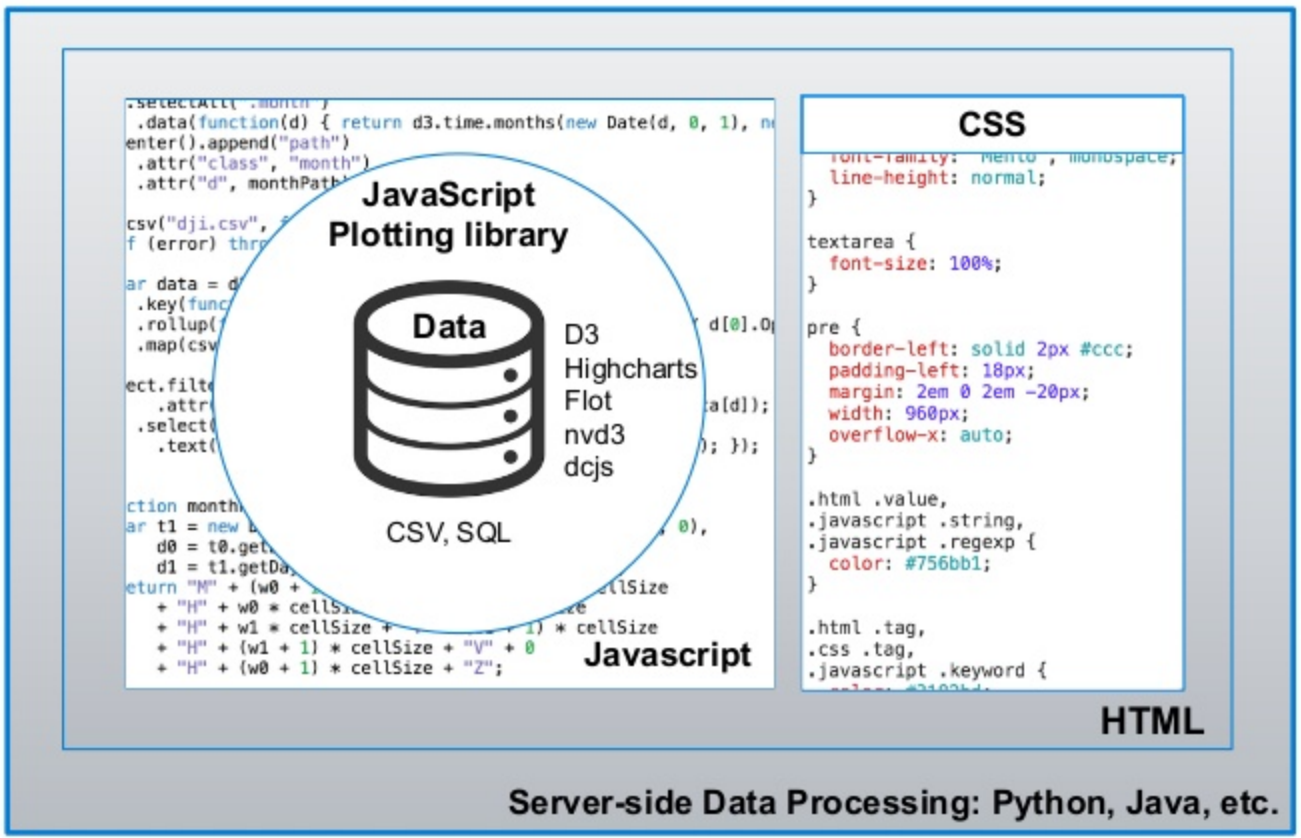
rBokeh with RStudio & Shiny



CONTINUUM®
ANALYTICS

Architecture

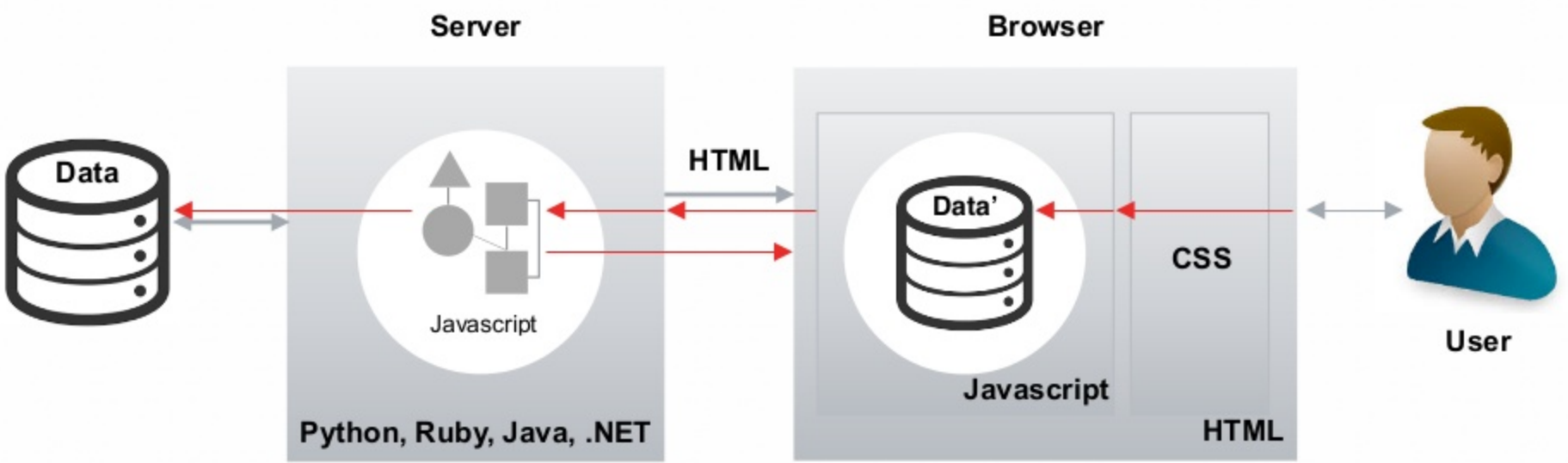
Traditional Web Visualization



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