

19ECS747: DATA SCIENCE

Unit V:

Applications of Data Science,
Technologies for visualization,
Bokeh (Python)

Data Visualization Techniques and Tools

- The ever-growing volume of data and its importance for business make data **visualization an essential part** of business strategy for many companies.

What determines data visualization choices

- **Visualization** is the first step to make sense of data.
- To translate and present data and data correlations in a **simple way**, data analysts use a wide range of techniques — charts, diagrams, maps, etc.
- Choosing the **right technique** and its setup is often the only way to make data understandable.
- **Poorly selected** tactics won't let to unlock the full potential of data or even make it irrelevant.

5 Factors that Influence data Visualization choices

1. **Audience:** It's important to adjust data representation to the specific target audience.
2. **Content:** The type of data you are dealing with will determine the tactics.
 - For example, if it's time-series metrics, you will use line charts to show the dynamics. To show the relationship between two elements, scatter plots are often used. In turn, bar charts work well for comparative analysis.

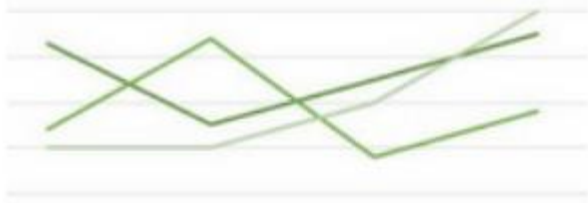
3. **Context:** You can use different data visualization approaches and read data depending on the context.
4. **Dynamics:** There are various types of data, and each type has a different rate of change.
 - For example, financial results can be measured monthly or yearly, while time series and tracking data are changing constantly.
5. **Purpose:** The goal of data visualization affects the way it is implemented.
 - In order to make a complex analysis, visualizations are compiled into dynamic and controllable dashboards that work as visual data analysis techniques and tools.
 - However, dashboards are not necessary to show a single or occasional data insight.

Data Visualization Techniques

- Depending on the above factors, you can choose different data visualization techniques and configure their features.
- Here are the common types of visualization techniques:
 - Charts
 - Plots
 - Maps
 - Diagrams and matrices

1. Charts

- The easiest way to show the development of one or several data sets is a chart.
- Charts vary from bar and line charts that show the relationship between elements over time, to pie charts that demonstrate the components or proportions between the elements of one whole.



Line



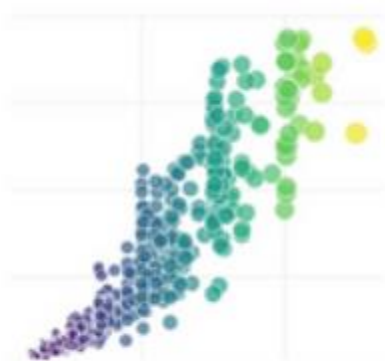
Pie



Bar

2. Plots

- Plots allow to distribute two or more data sets over a 2D or even 3D space to show the relationship between these sets and the parameters on the plot.
- Plots also vary.
- Scatter and bubble plots are some of the most widely-used visualizations.



Bubble



Scatter

3. Maps

- Maps are popular ways to visualize data used in different industries.
- They allow to locate elements on relevant objects and areas — geographical maps, building plans, website layouts, etc.
- Among the most popular map visualizations are heat maps, dot distribution maps.



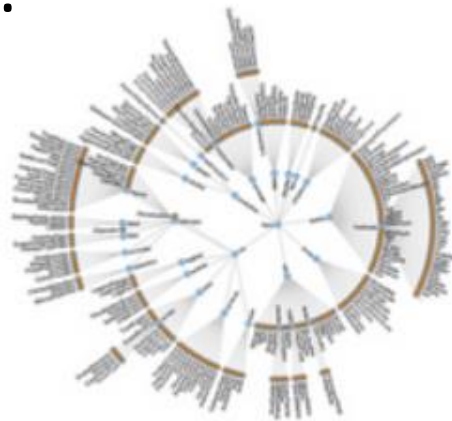
Heat



Dot distribution

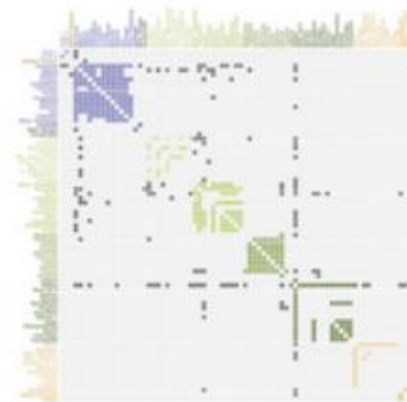
4. Diagrams and matrices

- Diagrams are usually used to demonstrate complex data relationships and links and include various types of data on one visualization.
- They can be hierarchical, multidimensional, tree-like.
- Matrix is one of the advanced data visualization techniques that help determine the correlation between multiple constantly updating (steaming) data sets.



Tree

Instructor S.Md.Mujeeb



Matrix

Data Visualization Tools

- Here are **three categories** of data visualization tools for different types of users and purposes.

1. Data visualization tools for everyone:

- **Tableau** companies rely on this platform to derive meaning from data and use insights for effective storytelling.

Also in this category:

- Among other popular data visualization tools in this category are easy-to-learn **Visme**, **Fusioncharts** with varied integration capabilities, free and open source **Datawrapper** and **ZingChart** for JavaScript and HTML5 charts.

2. Data visualization tools for coders:

- **Plotly** It's more complex than Tableau, however, comes with analytics perks.

Also in this category:

- **Sisense** is another data visualization tool with full-stack analytics capabilities.
- **IBM Watson Analytics** is known for its NLP capabilities.

3. Tools for complex data visualization and analytics:

- **Power BI** is exceptional for its highly intuitive drag-and-drop interface, short learning curve and large integration capabilities
- **Kibana** is the part of the Elastic Stack that turns data into visual insights.
- **Grafana** — a professional data visualization and analytic tool that supports up to 30 data sources, including AWS, Elasticsearch and Prometheus.

Data Visualization Tools Comparison

- There are **dozens**, if not hundreds, of applications, tools, and scripts available to create visualizations of large data sets.
- Many are very basic and have a **lot of overlapping** features.
- But there are standouts that either have more capability for the **types of visualizations** they can create or are significantly **easier to use** than the other options out there.

1. Tableau (and Tableau Public)

- Tableau has a variety of options available, including a desktop app, server and hosted online versions, and a free public option.
- There are hundreds of data import options available, from CSV files to Google Ads and Analytics data to Salesforce data.
- Output options include multiple chart formats as well as mapping capability.
- The **public version of Tableau** is free to use for anyone looking for a powerful way to create data visualizations that can be used in a variety of settings.

BLOCKBUSTER

Analyzing the 10 Highest Grossing Actors of All Time

i Include Bit Parts

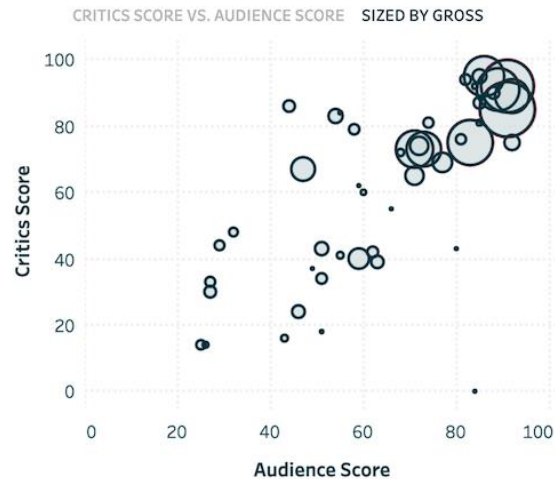
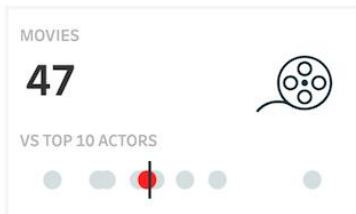
i Domestic (Unadjusted)

CLICK AN ACTOR TO ANALYZE



Scarlett Johansson

CAREER TIMELINE *BIT PART*



SOURCES:
Box Office Mojo
Rotten Tomatoes

Created by Ryan Sleeper



- **Pros**
 - Hundreds of data import options
 - Mapping capability
 - Free public version available
 - Lots of video tutorials to walk you through how to use Tableau
- **Cons**
 - Non-free versions are expensive (\$70/month/user for the Tableau Creator software)
 - Public version doesn't allow you to keep data analyses private
- **Bottom Line**
 - Tableau is a great option for those who need to create maps in addition to other types of charts.
 - Tableau Public is also a great option for anyone who wants to create public-facing visualizations.

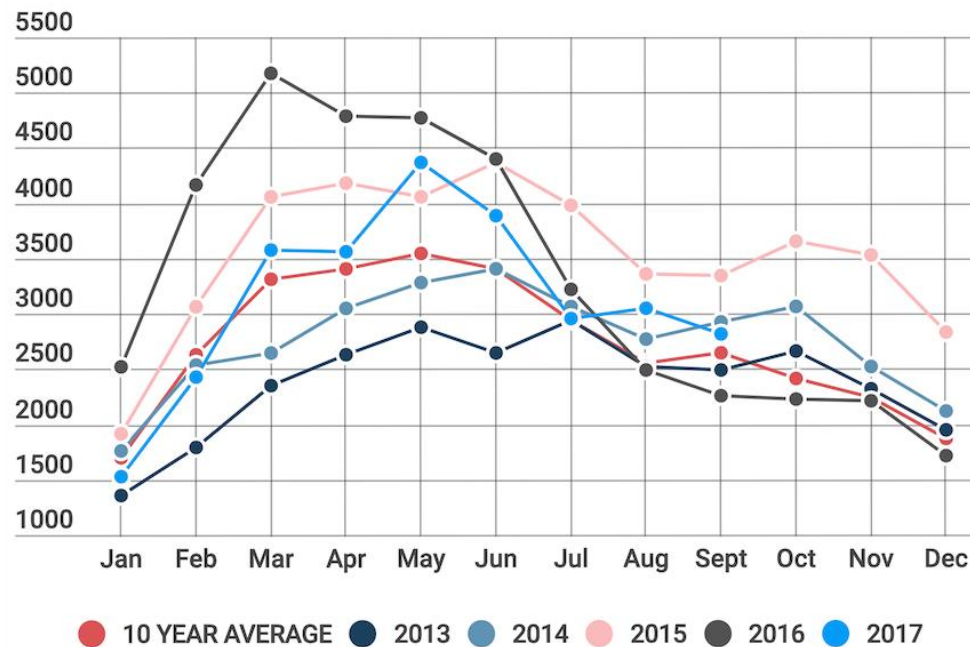
2. Infogram

- Infogram is a fully-featured drag-and-drop visualization tool that allows even non-designers to create effective visualizations of data for marketing reports, infographics, social media posts, maps, dashboards, and more.
- Finished visualizations can be exported into a number of formats: .PNG, .JPG, .GIF, .PDF and .HTML.
- Interactive visualizations are also possible, perfect for embedding into websites or apps.

GREATER VANCOUVER

REAL ESTATE STATS: SEPTEMBER 2017

OF SALES BY YEAR



- **Pros**

- Tiered pricing, including a free plan with basic features
- Includes 35+ chart types and 550+ map types
- Drag and drop editor
- API for importing additional data sources

- **Cons**

- Significantly fewer built-in data sources than some other apps

- **Bottom Line**

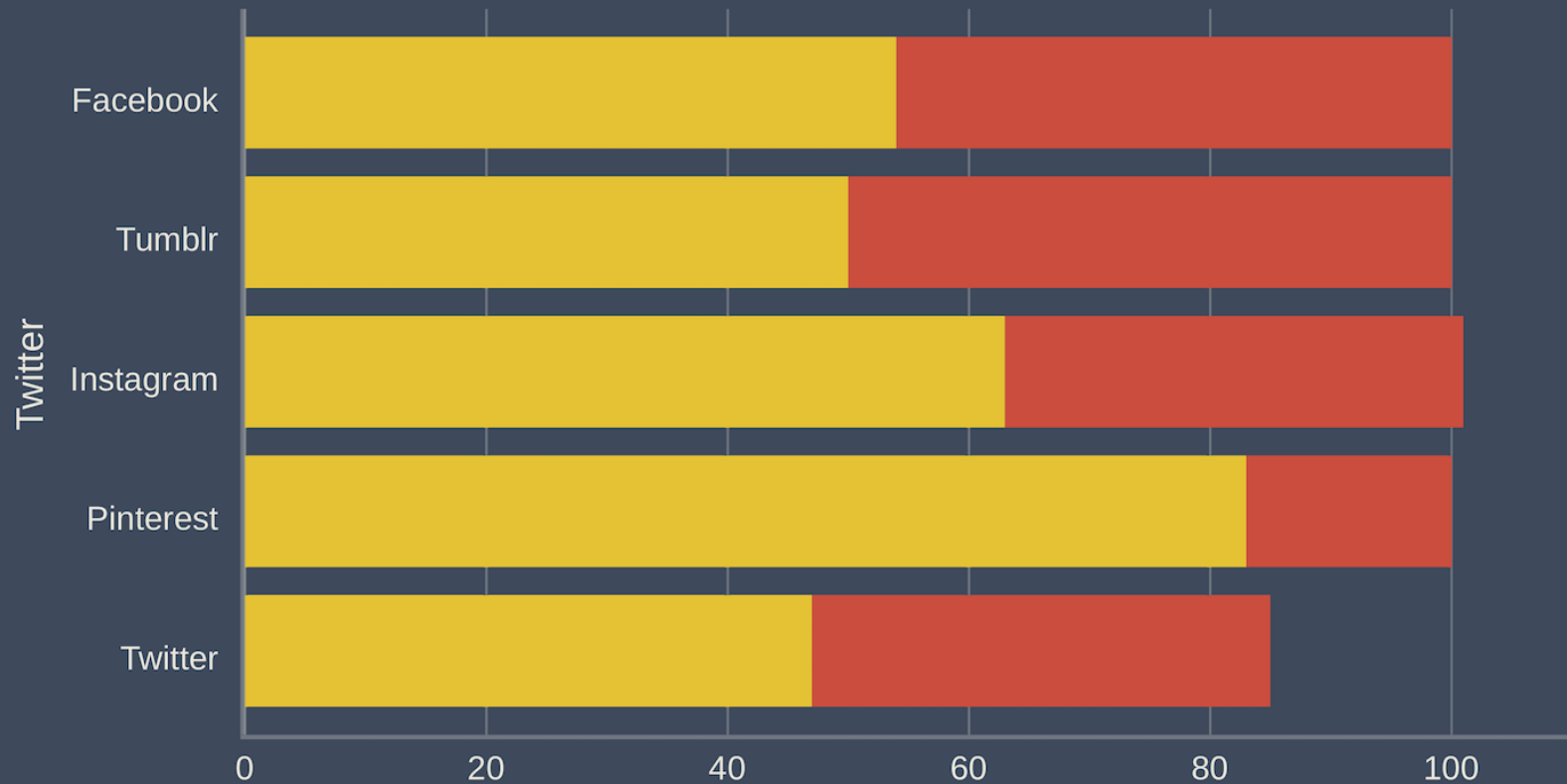
- Infogram is a great option for non-designers as well as designers.
- The drag-and-drop editor makes it easy to create professional-looking designs without a lot of visual design skill.

3. ChartBlocks

- ChartBlocks **claims** that data can be imported from “**anywhere**” using their API, including from live feeds.
- The app allows for extensive customization of the final visualization created, and the chart building wizard helps users pick exactly the right data for their charts before importing the data.

Gender Breakdown of Social Networking Users

Female
Male



- **Pros**

- Free and reasonably priced paid plans are available
- Easy to use wizard for importing the necessary data

- **Cons**

- Unclear how robust their API is
- Doesn't appear to have any mapping capability

- **Bottom Line**

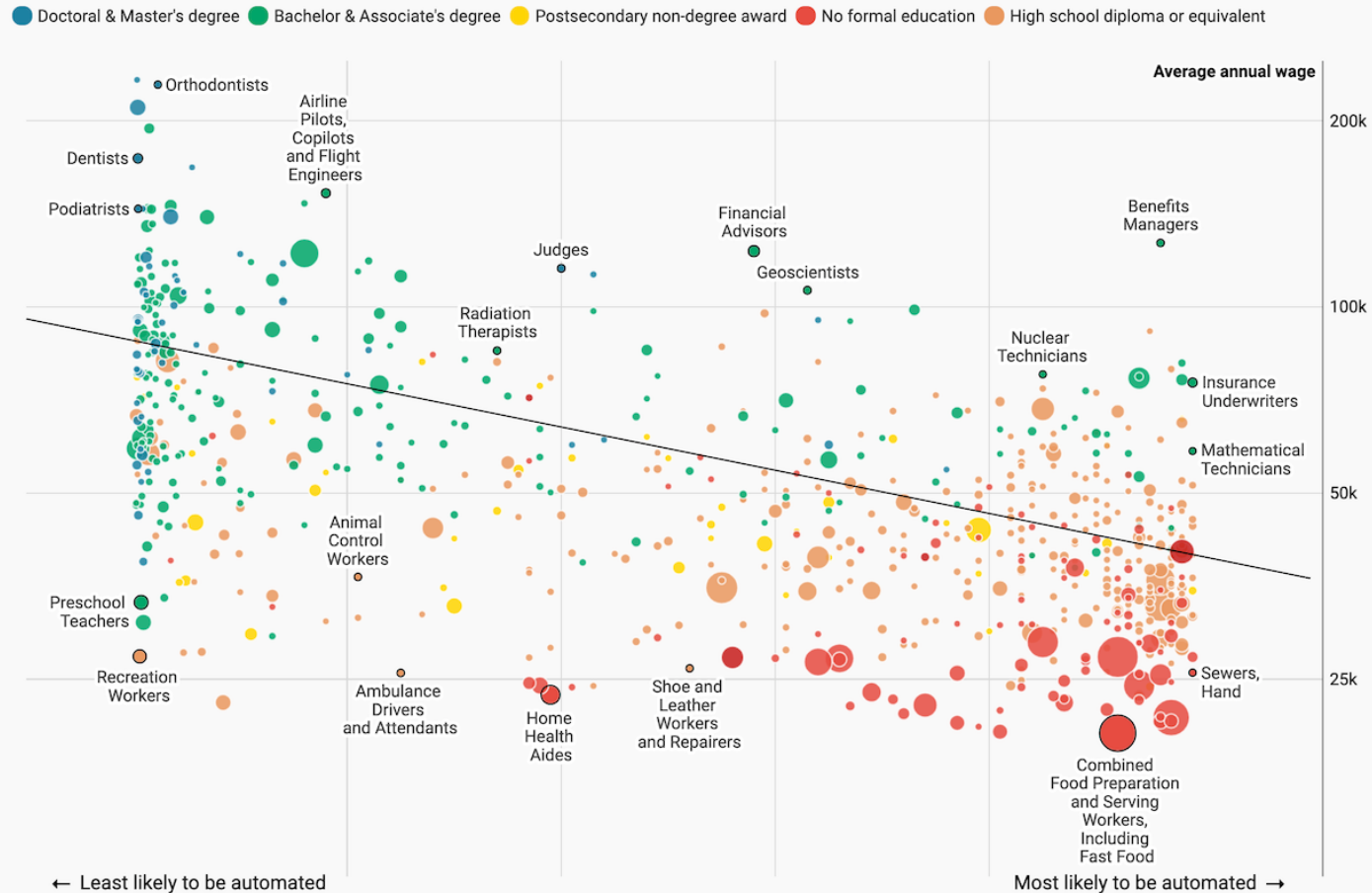
- ChartBlocks has an excellent free plan, which is a big plus.
- The ease of use for creating basic charts and graphs is also outstanding.

4. Datawrapper

- Datawrapper was created specifically for adding charts and maps to news stories.
- The charts and maps created are interactive and made for embedding on news **websites**.
- Once data is imported, charts can be created with a single click.
- Their visualization types include column, line, and bar charts, election donuts, area charts, scatter plots, choropleth and symbol maps, and locator maps, among others.
- The free plan is perfect for embedding graphics on smaller sites with limited traffic, but paid plans are on the expensive side, starting at \$39/month.

Our charts feel at home on the web. They're interactive, responsive and embeddable in your website.

Higher Risk of Job Automation in Lower Paying Jobs



Inspired by a [Bloomberg chart](#).

Source: FREY & OSBORNE, BUREAU OF LABOR STATISTICS • [Get the data](#)

- **Pros**

- Specifically designed for newsroom data visualization
- Free plan is a good fit for smaller sites
- Tool includes a built-in color blindness checker

- **Cons**

- Limited data sources
- Paid plans are on the expensive side

- **Bottom Line**

- Datawrapper is an excellent choice for data visualizations for news sites.
- Despite the price tag, the features Datawrapper includes for news-specific visualization make it worth it.

5. Google Charts

- Google Charts is a **powerful, free data** visualization tool that is specifically for creating interactive charts for embedding online.
- It works with dynamic data and the outputs are based purely on **HTML5 and SVG**, so they work in browsers without the use of additional plugins.
- Data sources include Google Spreadsheets, Google Fusion Tables, Salesforce, and other SQL databases.



Overview

Hello, Charts!

Quickstart

Load the Charts Library

Prepare the Data

Customize the Chart

Draw the Chart

Draw Multiple Charts

Chart Types

Chart Gallery

Annotation Charts

Area Charts

Bar Charts

Bubble Charts

Calendar Charts

Candlestick Charts

Column Charts

Combo Charts

Diff Charts

Donut Charts

Gantt Charts

Gauge Charts

GeoCharts

Histograms

Intervals

Line Charts

Maps

Org Charts

Pie Charts

Sankey Diagrams

Scatter Charts

Stepped Area Charts

Table Charts

Timelines

Tree Map Charts

Trendlines

Waterfall Charts

Word Trees

Miscellaneous Examples

Visualization: Combo Chart

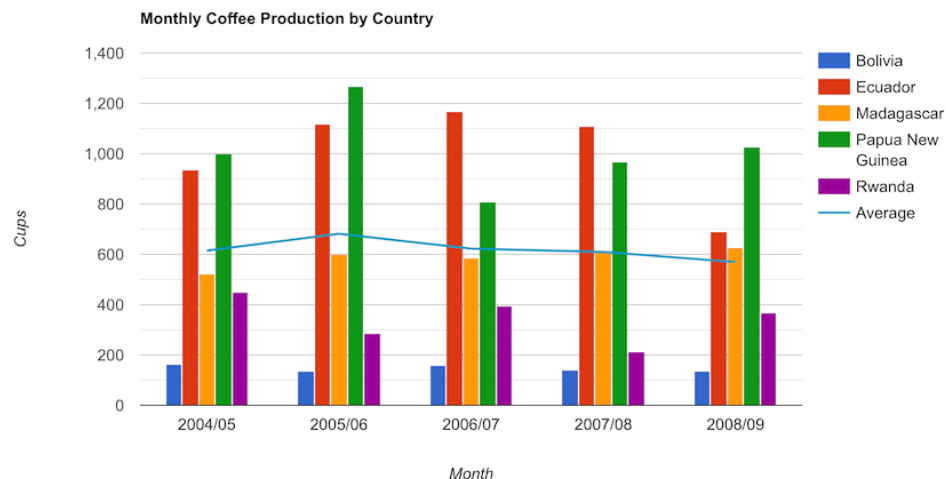


Overview

A chart that lets you render each series as a different marker type from the following list: line, area, bars, candlesticks, and stepped area.

To assign a default marker type for series, specify the `seriesType` property. Use the `series` property to specify properties of each series individually.

Example

[CODE IT YOURSELF ON JSFIDDLE](#)

```
<html>
<head>
  <script type="text/javascript" src="https://www.gstatic.com/charts/loader.js"></script>
  <script type="text/javascript">
    google.charts.load('current', {'packages':['combo']});
    google.charts.setOnLoadCallback(drawChart);

    function drawChart() {
      var data = new google.visualization.DataTable();
      data.addColumn('string', 'Month');
      data.addColumn('number', 'Bolivia');
      data.addColumn('number', 'Ecuador');
      data.addColumn('number', 'Madagascar');
      data.addColumn('number', 'Papua New Guinea');
      data.addColumn('number', 'Rwanda');
      data.addColumn('number', 'Average');
      data.addRows([
        ['2004/05', 150, 950, 500, 1000, 450, 610],
        ['2005/06', 150, 1100, 600, 1250, 300, 680],
        ['2006/07', 150, 1150, 600, 800, 400, 640],
        ['2007/08', 150, 1100, 600, 950, 200, 610],
        ['2008/09', 150, 700, 600, 1000, 350, 570]
      ]);
      var chart = new google.visualization.ComboChart(document.getElementById('chart'));
      chart.draw(data, {
        series: {
          0: {type: 'bar', color: '#4f81bd'},
          1: {type: 'bar', color: '#e34a33'},
          2: {type: 'bar', color: '#ffb74d'},
          3: {type: 'bar', color: '#4caf50'},
          4: {type: 'bar', color: '#9c27b0'},
          5: {type: 'line', color: '#42a5f5', lineWidth: 2}
        }
      });
    }
  </script>
</head>
<body>
  <div id="chart"></div>
</body>
</html>
```

instructor S.Md.Mujeeb

- **Pros**
 - Free
 - Wide variety of chart formats available
 - Cross-browser compatible since it uses HTML5/SVG
 - Works with dynamic data
- **Cons**
 - Beyond the tutorials and forum available, there's limited support
- **Bottom Line**
 - Google Charts is a great option if a designer is somewhat comfortable with coding and wants a powerful, free solution.
 - Being able to use any SQL database as a data source makes it a good option for large data sets, too.

6. FusionCharts

- FusionCharts is a JavaScript-based option for creating web and mobile dashboards. It includes over 150 chart types and 1,000 map types.
- It can integrate with popular JS frameworks (including React, jQuery, React, Ember, and Angular) as well as with server-side programming languages (including PHP, Java, Django, and Ruby on Rails).
- FusionCharts gives ready-to-use code for all of the chart and map variations, making it easier to embed in websites even for those designers with limited programming knowledge.
- Because FusionCharts is aimed at **creating dashboards** rather than just **straightforward data visualizations** it's one of the most expensive options.
- But it's also one of the most powerful.

Energy Dashboard

March 2019



Dashboard



Cost



Appliances



Usage-by-rooms



Emissions

COST PREDICTED



CHANGE IN COST



USAGE ESTIMATE



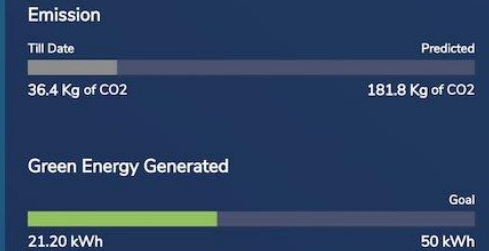
ACTIVE APPLIANCES



ENERGY INTENSITY



CARBON FOOTPRINT



Energy Tip: On warm days, setting a programmable thermostat to a higher setting when you are not at home can help reduce your energy costs by approximately 10 percent.

- **Pros**

- Huge number of chart and map format options
- More features than most of the other visualization tools
- Integrates with a number of different frameworks and programming languages

- **Cons**

- Expensive (starts at almost \$500 for one developer license)
- Overkill for simple visualizations outside of a dashboard environment

- **Bottom Line**

- For creating dashboards, nothing else really compares to FusionCharts.
- If that's(dashboards) the project at hand, this is undoubtedly the most powerful choice.

7. Grafana

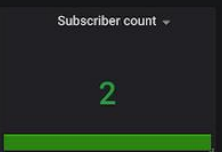
- Grafana is open-source visualization software that lets users create dynamic dashboards and other visualizations.
- It supports mixed data sources, annotations, and customizable alert functions, and it can be extended via hundreds of available plugins.
- That makes it one of the most powerful visualization tools available.



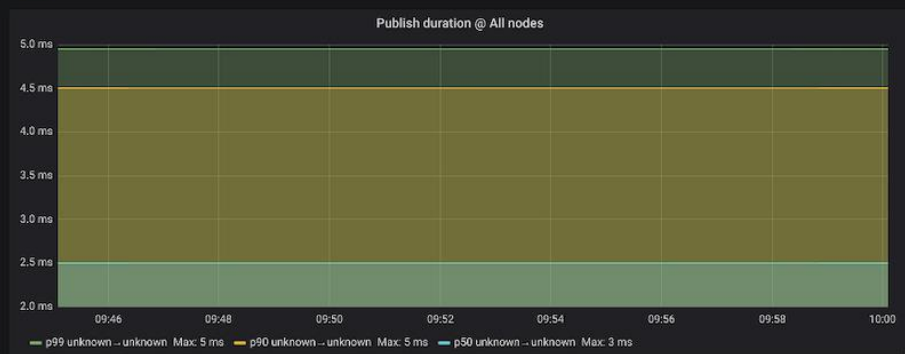
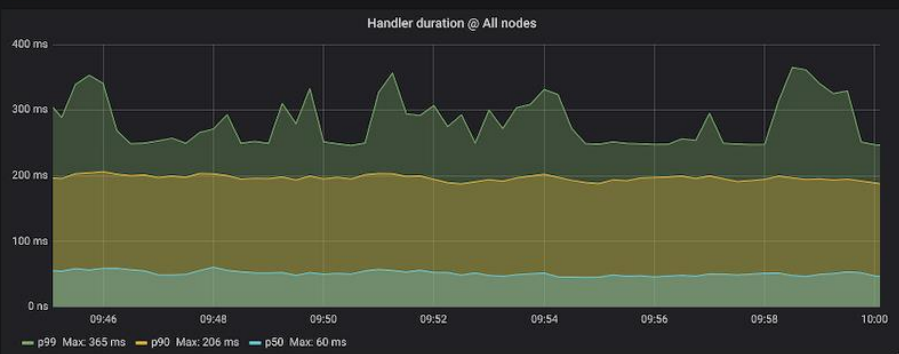
1 Watermill -

interval 1m ▾ node All ▾ handler All ▾ publisher All ▾

Current stats



Handler & publisher operations



- **Pros**

- Open source, with free and paid options available
- Large selection of data sources available
- Variety of chart types available
- Makes creating dynamic dashboards simple
- Can work with mixed data feeds

- **Cons**

- Overkill for creating simple visualizations
- Doesn't offer as many visual customization options as some other tools
- Not the best option for creating visualization images
- Not able to embed dashboards in websites, though possible for individual panels

- **Bottom Line**

- Grafana is one of the best options for creating dashboards for internal use, especially for mixed or large data sources.

Conclusion

- There is such a huge variety of visualization tools available to designers that it can be hard to decide which one to use.
- Data visualization designers should keep in mind things like ease of use and whether a tool has the features they need.
- Selecting the most powerful tool available isn't always the best idea
- Most data visualization tools include free trials (if the entire tool isn't free), so it's worth taking the time to try out a few before deciding on a single solution.

- <https://www.digiteum.com/data-visualization-techniques-tools/>
- <https://www.toptal.com/designers/data-visualization/data-visualization-tools>

Bokeh (Python)

- Bokeh is a Python interactive data visualization.
- Unlike Matplotlib and Seaborn, Bokeh renders its plots using HTML and JavaScript.
- It targets modern web browsers for presentation providing elegant, concise construction of novel graphics with high-performance interactivity.

Features of Bokeh

- Some of the important features of Bokeh are given below:
 - **Flexibility:** Bokeh can be used for common plotting requirements and for custom and complex use-cases.
 - **Productivity:** Its interaction with other popular Pydata tools (such as Pandas and Jupyter notebook) is very easy.
 - **Interactivity:** It creates interactive plots that changes with the user interaction.
 - **Powerful:** Generation of visualizations for specialized use-cases can be done by adding JavaScript.
 - **Shareable:** Visual data are shareable. They can also be rendered in Jupyter notebooks.
 - **Open source:** Bokeh is an open source project.

Interface Level

- Bokeh supports different interface levels can be used by users:
 - **a low-level:** `bokeh.models` interface provides the most flexibility to application developers.
 - **an intermediate-level:** `bokeh.plotting` interface is composing of all visual glyphs.
 - **a high-level:** `bokeh.charts` interface is used to build complex plots easily.

- To begin with,
import following functions from bokeh.plotting modules–
- **from bokeh.plotting import figure, output_file, show**
- The **figure()** function creates a new figure for plotting.
- The **output_file()** function is used to specify a HTML file to store output.
- The **show()** function displays the Bokeh figure in browser on in notebook.

- <https://www.youtube.com/watch?v=o4TB6LTPDaY&feature=youtu.be>
- <https://www.tutorialspoint.com/bokeh/index.htm>
- <https://mybinder.org/v2/gh/bokeh/bokeh-notebooks/master?filepath=tutorial%2F00%20-%20Introduction%20and%20Setup.ipynb>