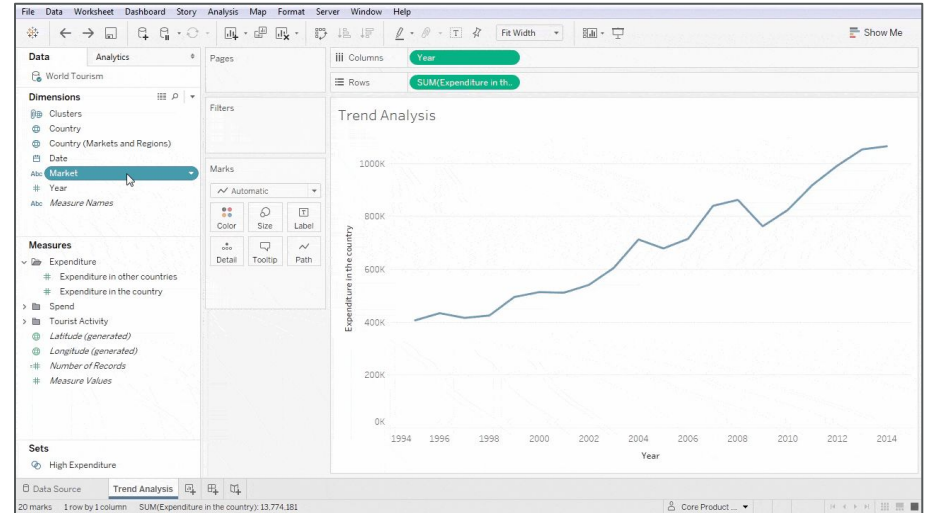


Data Visualization

Suraj Geddam, Charlotte Hettrich, Sujay Jagadeesh,
David Keefe, and Sanjay Mohandas

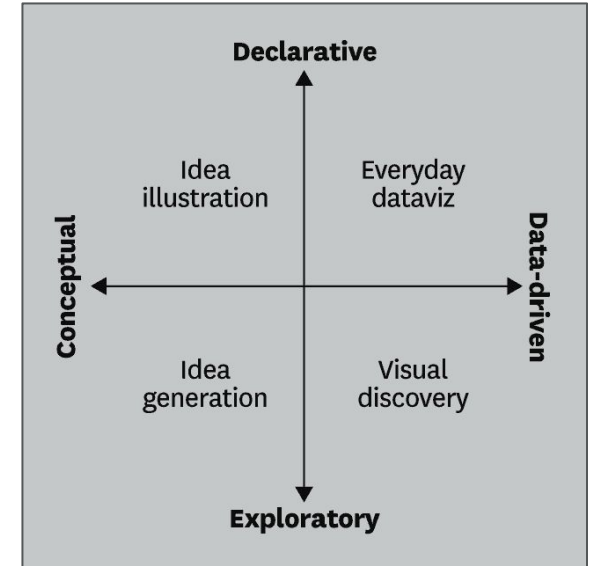
Overview of Data Visualization

- ❖ Data visualization is the representation of data through the use of common graphics, such as charts, plots, infographics, and animations.
- ❖ Data visualization can be used for a variety of purposes, not just for data teams.
- ❖ In the world of Big Data, data visualization tools and technologies are essential to analyze massive amounts of information and make data-driven decisions.



4 Key Purposes of Data Visualization

- 1. Idea Generation**
 - a. Frequently leveraged during brainstorming sessions at the start of a project
- 2. Idea Illustration**
 - a. Assists in conveying an idea
 - b. Used in learning settings like tutorials and courses
- 3. Visual Discovery**
 - a. Helps data analysts, data scientists, and other data professionals identify patterns and trends within a dataset
- 4. Everyday Dataviz**
 - a. Critical step in the data science process helping teams and individuals convey data more effectively to colleagues



Common Terms in Data Visualization

❖ General Types of Visualizations

- **Chart:** information presented in a tabular, graphical form with data displayed along two axes
- **Table:** a set of figures displayed in rows and columns
- **Graph:** a diagram of points, line segments, curves, or areas that represents certain variables in comparison to each other
- **Geospatial:** a visualization that shows data in map form using different shapes and colors to show the relationship between pieces of data and specific locations
- **Infographic:** a combination of visuals and words that represent data
- **Dashboards:** a collection of visualizations and data displayed in one place to help analyzing the presenting data

❖ More Specific Examples

- **Gantt Chart:** typically used in product management, a bar chart depiction of timelines and tasks
- **Histogram:** a type of bar that split a continuous measure into different bins to help analyze the distribution

Visualization Tools

Pros

- ❖ High industry relevance
- ❖ Drag and drop user interface to create powerful visualizations and dashboards without coding
- ❖ Relatively easy to create simple visualizations
- ❖ Easy to share and embed results on web pages

Cons

- ❖ Not open source
- ❖ License required if results cannot be published
- ❖ Becomes more complicated and difficult when you need more advanced visualizations



Visualization Libraries

Pros



















- ❖ High industry relevance
- ❖ Great expressive power and flexibility
- ❖ Thousands of examples and tutorials
- ❖ Open source
- ❖ Primarily suitable for implementing interactive web-based visualizations

Cons

- ❖ Steep learning curve
- ❖ Requires strong understanding of a variety of coding languages depending on the library
 - JavaScript
 - CSS
 - Python



Product Families + Business Context

- ❖ **Business Intelligence (BI) Tools**  + a b l e a u  Power BI  QlikSense
- ❖ **Charting Libraries and Frameworks**   matplotlib  seaborn 
- ❖ **Dashboard Platforms**  Klipfolio  geckoboard
- ❖ **Geospatial Visualization Tools**  ArcGIS  mapbox  Google Maps
- ❖ **Data Storytelling Platforms**  Datawrapper  Observable
- ❖ **Data Exploration and Analytics Platforms**  Looker  DOMO®
- ❖ **Reporting Tools**  Microsoft®
SQL Server®
Reporting Services  JasperReports

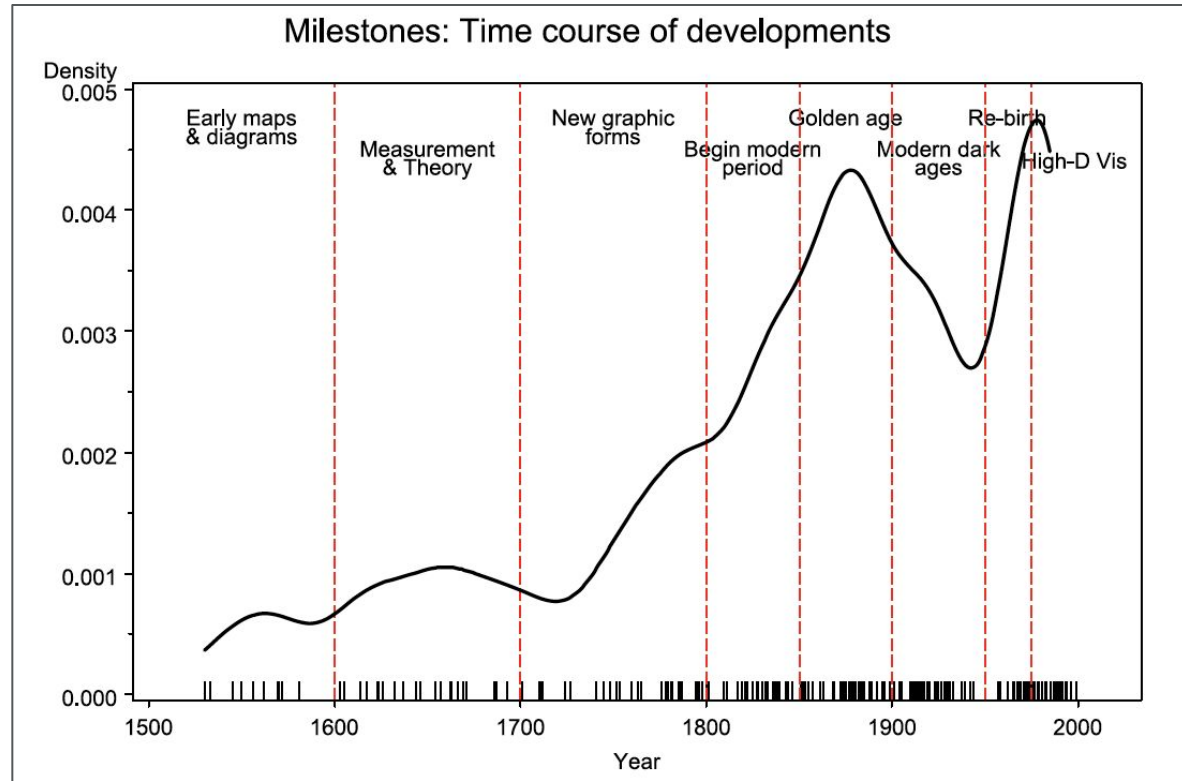
Acknowledgement: *Unleashed - business reporting tools*, <https://www.unleashedsoftware.com/blog/why-business-reporting-is-important-for-business-success>.
Product School - data visualization tools for product managers,
<https://productschool.com/blog/product-fundamentals/data-visualization-platforms-product-manager>.
Google - data visualization tools logos, <https://www.google.com/search?q=data+visualization+tools+logos>.
Google - data visualization tools logos, <https://www.google.com/search?q=data+visualization+tools+logos>.

A Brief History

Time distribution of events considered milestones in the history of data visualization, shown by a rug plot and density estimate

Popular Applications Release Years

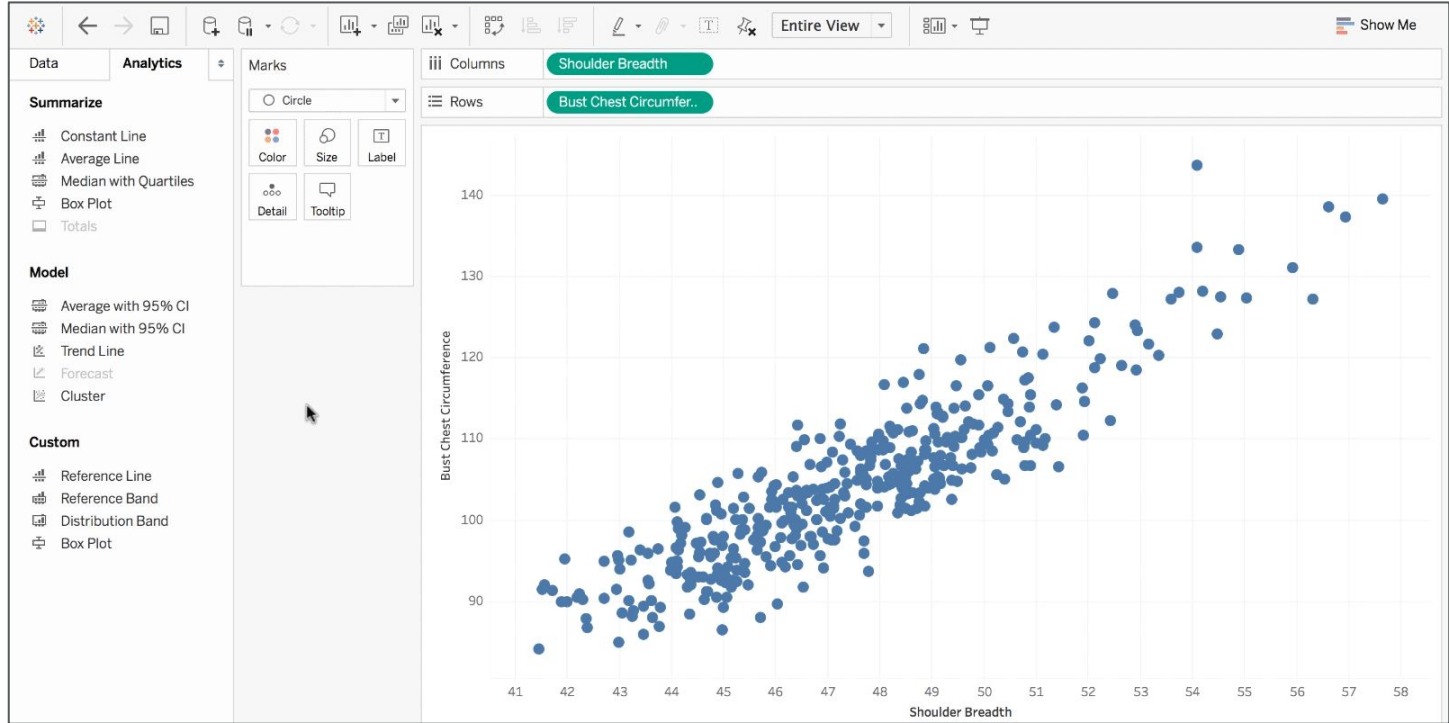
- ❖ Tableau - 2003
- ❖ PowerBI - 2015
- ❖ D3.js - 2011
- ❖ Plotly.js - 2013
- ❖ Vega-Lite - 2015



Visualization Tools Features/Functions

- ❖ **Dataset Hosting**
- ❖ **Automatic Visualization Creation**
- ❖ **Query Optimization**
- ❖ **Scalability**
- ❖ **Security and Governance**
- ❖ **AI Capabilities**

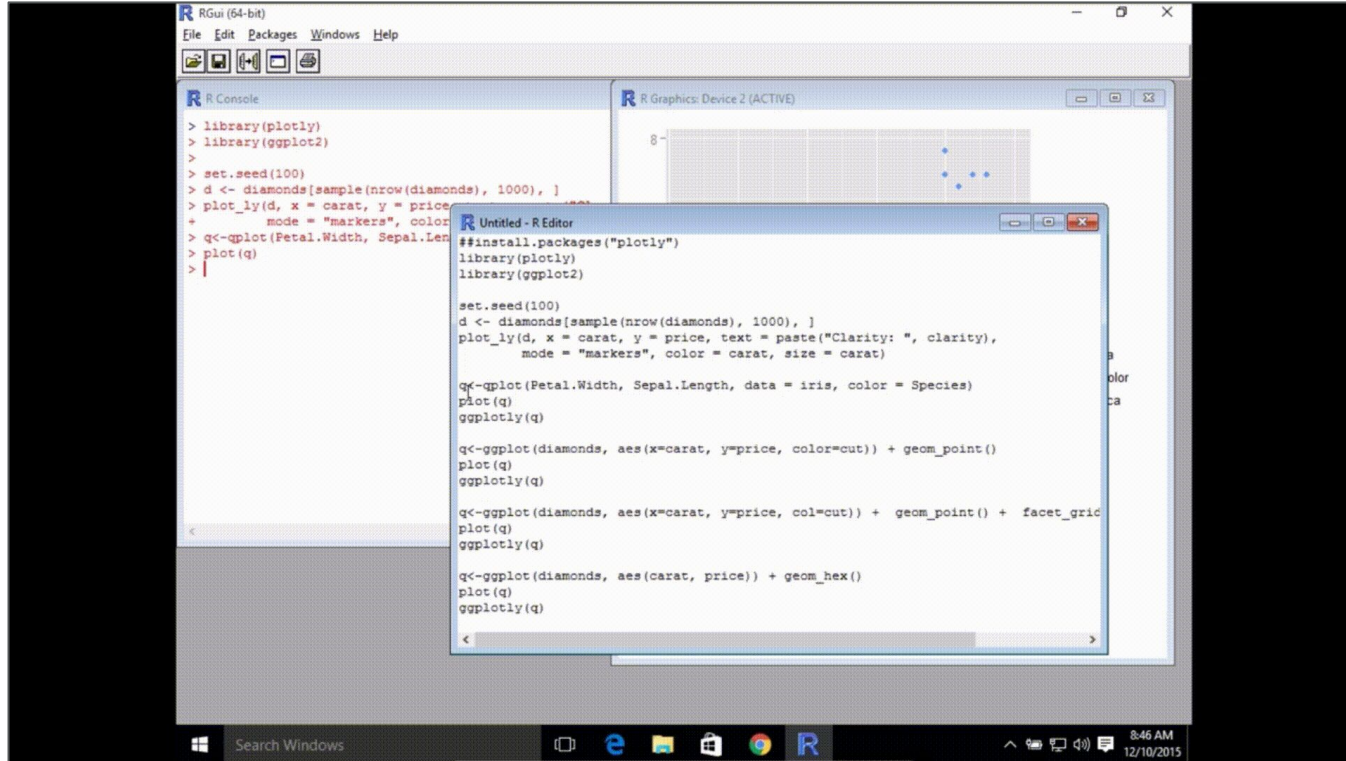
Visualization Tool Features Example



Visualization Libraries Features/Functions

- ❖ **Data Input**
- ❖ **Manual Visualization Creation**
- ❖ **Flexibility**
- ❖ **Figure Exporting**
- ❖ **Support for 1 Language**
- ❖ **Integration with Other Libraries**

Visualization Library Features Example



The screenshot displays the RGui (64-bit) interface. The R Console window shows the following code being executed:

```
> library(plotly)
> library(ggplot2)
>
> set.seed(100)
> d <- diamonds[sample(nrow(diamonds), 1000), ]
> plot_ly(d, x = carat, y = price, color = cut, mode = "markers", color = carat, size = carat)
> q<-qplot(Petal.Width, Sepal.Length, data = iris, color = Species)
> plot(q)
> |
```

The R Graphics window shows a scatter plot of carat vs price with points colored by cut and sized by carat. The R Editor window shows the following code:

```
##install.packages("plotly")
library(plotly)
library(ggplot2)

set.seed(100)
d <- diamonds[sample(nrow(diamonds), 1000), ]
plot_ly(d, x = carat, y = price, text = paste("Clarity: ", clarity),
        mode = "markers", color = carat, size = carat)

q<-qplot(Petal.Width, Sepal.Length, data = iris, color = Species)
plot(q)
ggplotly(q)

q<-ggplot(diamonds, aes(x=carat, y=price, color=cut)) + geom_point()
plot(q)
ggplotly(q)

q<-ggplot(diamonds, aes(x=carat, y=price, col=cut)) + geom_point() + facet_grid
plot(q)
ggplotly(q)

q<-ggplot(diamonds, aes(carat, price)) + geom_hex()
plot(q)
ggplotly(q)
```

Tools Comparison Chart

	 Power BI	 Tableau	 Qlik Sense	 ThoughtSpot	 Looker
Full-featured Free Version	Yes	Separate tool	Separate tool	No	No
Development Environment	Desktop	Desktop	Web Browser	Implementation	Cloud
R and Python Supported	Yes	Yes	Yes	R Only	Yes
Dynamic Cross-filtering	Yes	Yes	Yes	No	No
AI-enabled Analytics	Yes	Yes	Yes	Yes	No
Search Analytics with NLP	Yes	Yes	Yes	Yes	No
Data Prep Tools	Yes	Separate tool	Separate tool	Yes	No
Data Modeling Tools	Yes	Separate tool	Yes	Yes	Yes
Preferred Data Model	Star-schema	Flat	Snowflake	Star-schema	Flat
Database Independent	Yes	Yes	Yes	Yes	No
Built in Row Level Security	Yes	Yes	Yes	No	No
Mixed Model Types	Yes	No	No	Yes	No
Third-party Data Model Access	Yes	No	No	No	No
Commenting & Collaboration	Yes	Yes	Yes	Yes	No
Embedded Analytics	Yes	Yes	Yes	Yes	Yes
Open-source Custom Visualizations	Yes	No	Yes	No	Yes
Native Mobile App	Yes	Yes	Yes	Yes	No

Libraries Comparison Chart

Table 3. Comparison of data visualization libraries

Plot type	Matplotlib	seaborn	Plotly	Bokeh	ggplot
Line chart	+	+	+	+	+
Histograms	+	+	+	-	+
Bar	+	+	+	+	+
Scatterplots	+	+	+	+	+
Boxplot	+	+	+	-	-
Contours	+	+	+	-	-
Filled polygons	+	-	+	+	-
Spectrogram	+	-	+	-	-
Violin plot	+	+	+	-	-
Pairplot	-	+	-	-	-
Heatmap	-	+	+	+	-
Matrix clustermap (dendrogram)	-	+	+	-	-
Regression plot	-	+	-	-	+
Joint plot	-	+	+	-	-
Polar plot	+	-	+	-	-
3D	+	-	+	-	-
Interactive graphs and animations	+	-	+	+	-
Others	+	+	+	+	-

NOTE: If there is minus in some column, it does not necessarily mean that it is not possible to do it, but that there is no direct function for a wanted plot (for example, pairplot is possible to create with Matplotlib with several lines of code and a scatterplot)

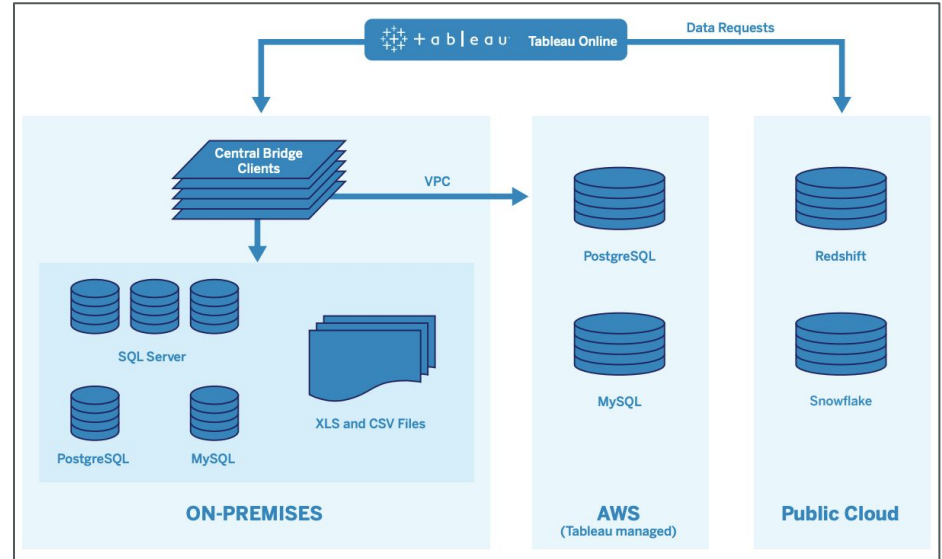
Tableau

- ❖ Founded in 2003, acquired by Salesforce in 2019
- ❖ User-friendly interface
- ❖ **Key Features:**
 1. Interactive dashboards
 2. Real-time data analysis
 3. Large dataset handling
 4. Integration with various data sources

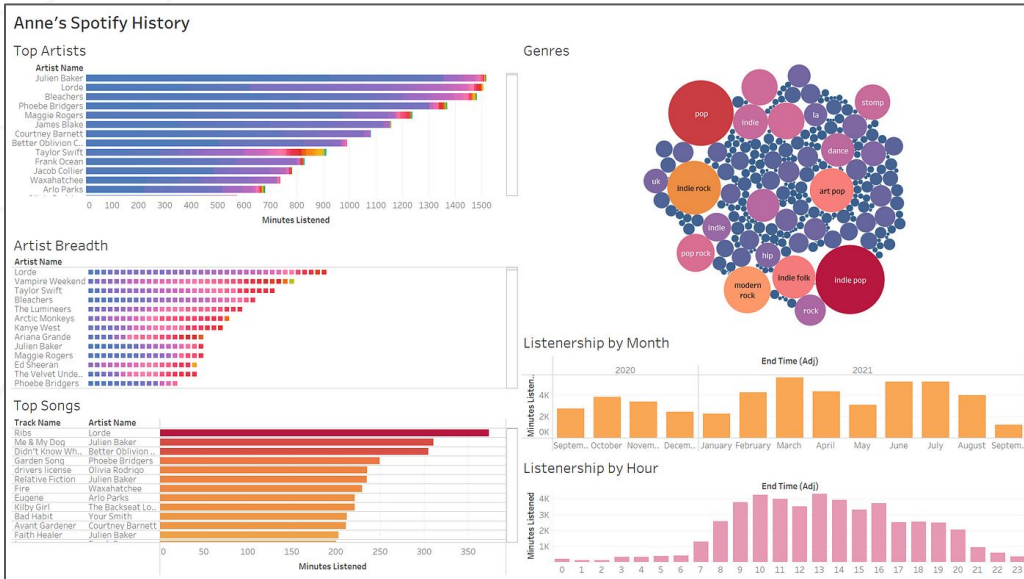


Architecture of Tableau

- ❖ **Tableau Desktop:** Where visualizations and dashboards are created.
- ❖ **Tableau Server:** An on-premise solution for sharing and collaborating on dashboards.
- ❖ **Tableau Online:** A cloud-based platform offering similar functionalities to Tableau Server.
- ❖ **Tableau Prep:** Dedicated tool for data preparation tasks.



Data Connection in Tableau



Types of Data Connections

- ◆ **Live Connection:** Real-time querying, best for up-to-date data but can slow down with large datasets.
- ◆ **Data Extract:** Static snapshot for improved performance and offline access.

Supported Data Sources

- ◆ **SQL Databases:** Connect to databases like MySQL, PostgreSQL, and SQL Server.
- ◆ **Cloud Databases:** Access cloud storage like Amazon Redshift, Google BigQuery, and Azure SQL Database.
- ◆ **Excel Files:** Import data from Excel spreadsheets for quick analysis.
- ◆ **Web Data Connectors:** Fetch data from web-based APIs or HTML tables for online data integration.

Creating Visualizations and Interactive Dashboards

Visualization Creation

- ❖ Drag-and-drop interface for easy visualization creation (bar charts, line graphs, maps).
- ❖ Customizable options for colors, labels, and tooltips.

Interactive Dashboards

- ❖ Combine multiple visualizations into a single dashboard.
- ❖ Interactive elements like filters and actions for dynamic analysis.
- ❖ Shareable dashboards for collaboration and presentation.

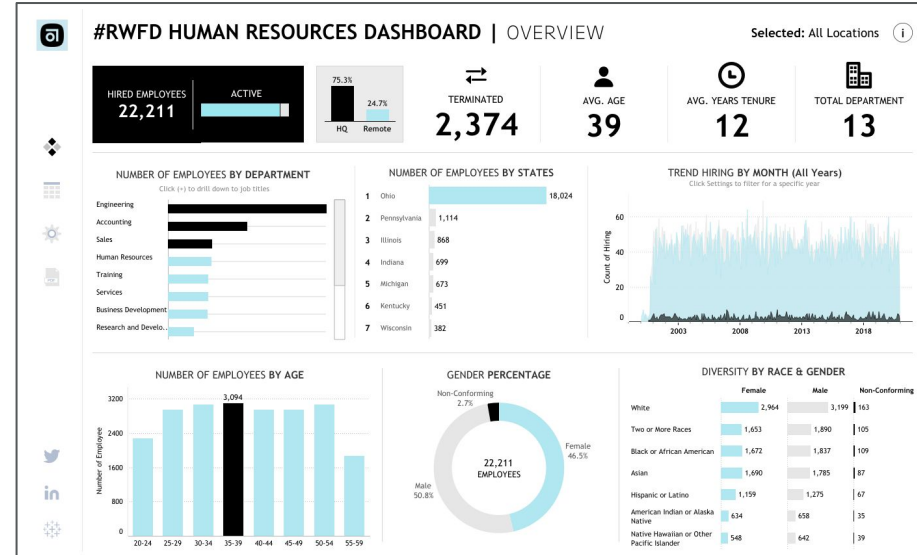
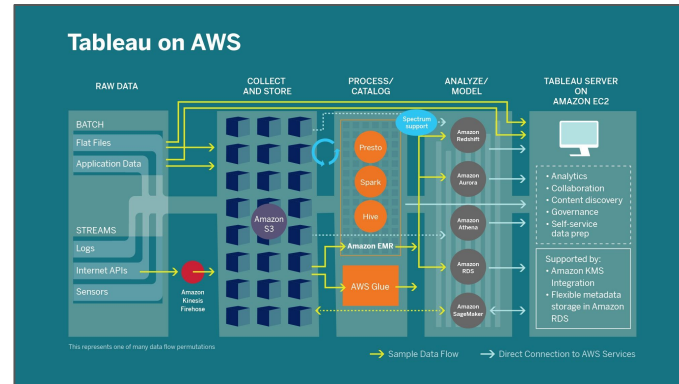
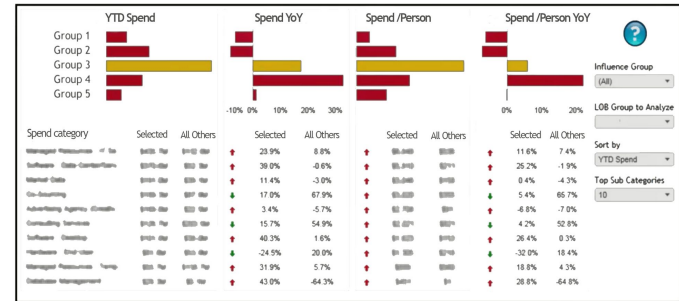


Tableau Used in the Business World

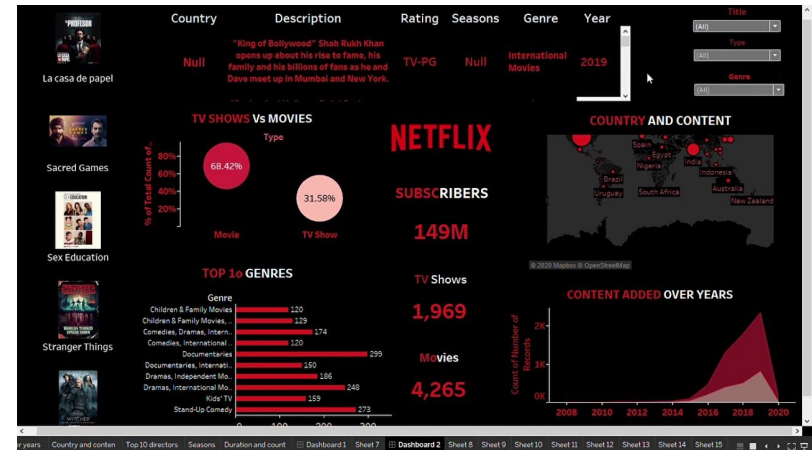
- ❖ **Netflix:** Optimizes content and gains customer insights for personalized recommendations.
- ❖ **Amazon:** Analyzes sales data and customer behavior to improve retail operations.
- ❖ **Verizon:** Utilizes data visualization for network performance monitoring and customer service improvement.
- ❖ **Bank of America:** Employs Tableau for financial analysis and risk management.
- ❖ **Coca-Cola:** Uses Tableau for supply chain optimization and market trend analysis.

How does my LOB compare to others by subcategory?



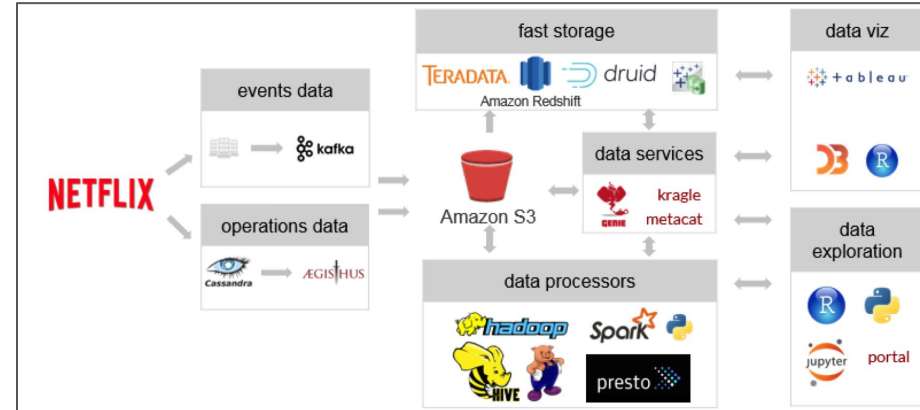
Case Study: Netflix using Tableau - Content Optimization

- ❖ **Content Performance Analysis:** Uses Tableau to monitor key performance indicators (KPIs) for shows and movies.
- ❖ **Personalized Recommendations:** Analyzes viewing patterns to tailor recommendations for each user.
- ❖ **A/B Testing:** Employs Tableau to evaluate the effectiveness of different content recommendation algorithms.



Case Study: Netflix using Tableau - Data Integration/Impact

- ❖ **Big Data Integration:** Combines Tableau with big data platforms like Hadoop for comprehensive analysis.
- ❖ **Real-Time Analytics:** Utilizes Tableau for real-time monitoring of viewer engagement and content trends.
- ❖ **Business Impact:** Data-driven insights lead to informed decision-making, improved content strategy, and enhanced user satisfaction.



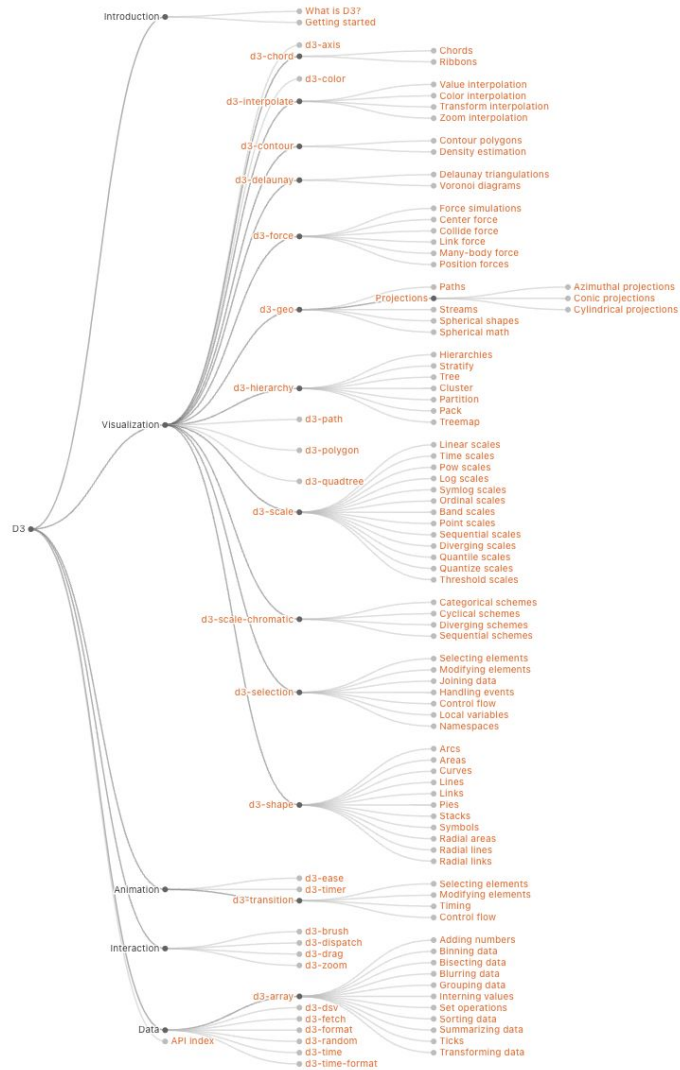
D3.js

- ❖ Data-Driven Documents
- ❖ JavaScript library for manipulating documents based on data
- ❖ Enables binding of data to DOM elements
- ❖ Offers powerful visualization components and a data-driven approach to DOM manipulation



Architecture of D3.js

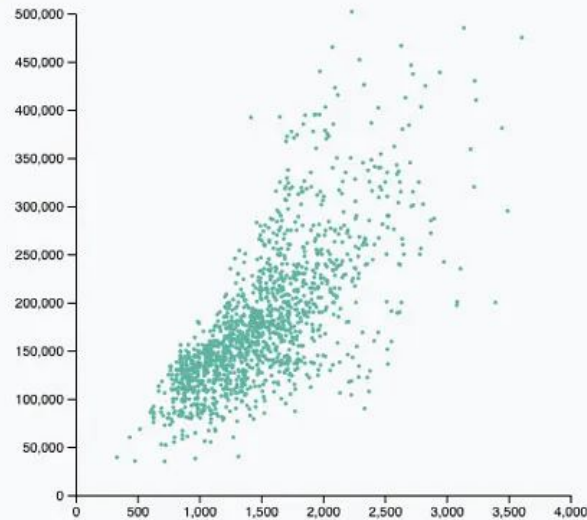
- ❖ **Selections:** Manipulate elements in the document
- ❖ **Data Binding:** Associate data with DOM elements
- ❖ **Scales:** Convert data values to visual representation
- ❖ **Axes:** Generate SVG axes for graphs
- ❖ **Transitions:** Animate changes in the visualization



Working with Data in D3.js

- ❖ **Loading Data:** CSV, JSON, XML, etc.
- ❖ **Binding Data to Elements:** Enter, Update, Exit pattern
- ❖ **Dynamic Updates:** Responding to user interactions and data changes

- ❖ Setting up the SVG container
- ❖ Scaling and mapping data to visual elements
- ❖ Adding axes and labels
- ❖ Customizing styles and adding interactivity

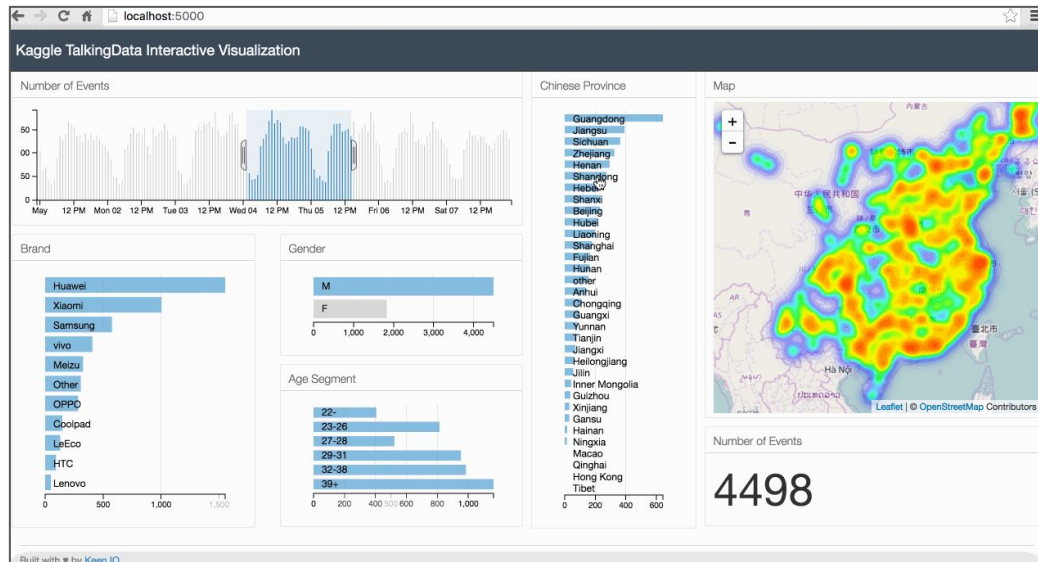


```
// Add Y axis
var y = d3.scaleLinear()
  .domain([0, 500000])
  .range([height, 0]);
svg.append("g")
  .call(d3.axisLeft(y));

// Add dots
svg.append("g")
  .selectAll("dot")
  .data(data)
  .enter()
  .append("circle")
  .attr("cx", function (d) { return x(d.GrLivArea); })
  .attr("cy", function (d) { return y(d.SalePrice); })
  .attr("r", 1.5)
  .style("fill", "#69b3a2")
```


Advanced Features of D3.js

- ❖ Creating complex visualizations like force-directed graphs and choropleth maps
- ❖ Leveraging D3.js modules for specialized tasks
- ❖ Integrating with other web technologies for interactive applications



Companies Using D3.js

The New York Times: Interactive journalism and data storytelling

Uber: Visualizing urban mobility and geospatial data

Airbnb: Data-driven decision making for property listings and market trends

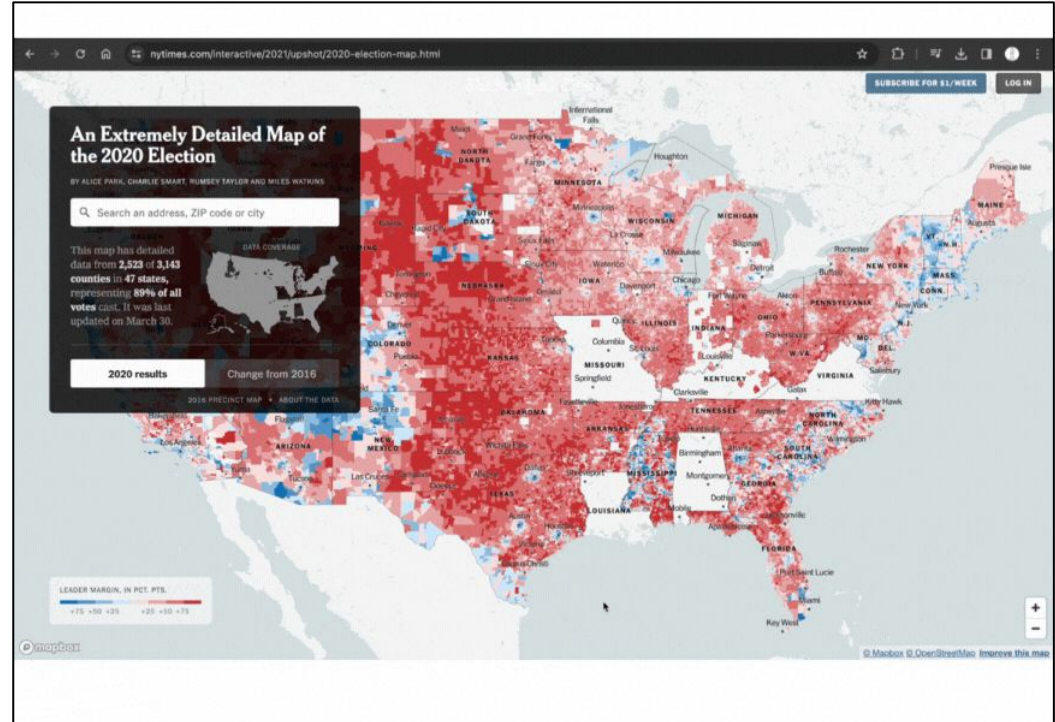
Netflix: User engagement analytics and content performance visualization

Case Study: New York Times

Objective: To provide a dynamic and interactive visualization of the 2020 presidential election results

Implementation: Used D3.js to create a map that displayed real-time election results with detailed state-by-state breakdowns

Features: Zoomable map, tooltip information on hover, color-coded states based on party affiliation

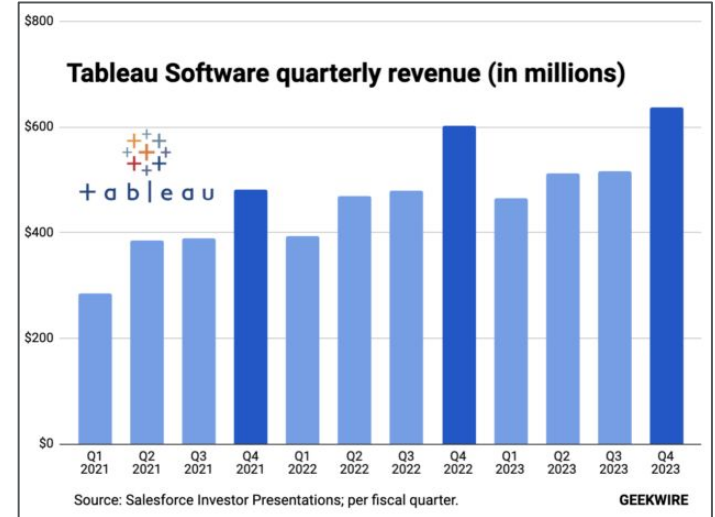


Impact and User Engagement

- ❖ **Enhanced Reader Engagement:** Interactive visualizations led to increased time spent on the site
- ❖ **Improved Understanding:** Readers could explore data at their own pace, leading to a deeper comprehension of election dynamics
- ❖ **Positive Feedback:** The visualization received widespread acclaim for its innovative approach to data presentation

Visualization Tools Marketing Data

- ❖ **Tableau 2022 Revenue: \$2.1 billion**
- ❖ **Power Platform 2022 Revenue: >\$2 billion**
- ❖ **Domo 2022 Revenue: \$258 million**
- ❖ **ThoughtSpot 2022 Revenue: \$150 million**
- ❖ **Looker 2019 Revenue: ~\$140 million**



Visualization Libraries Marketing Data

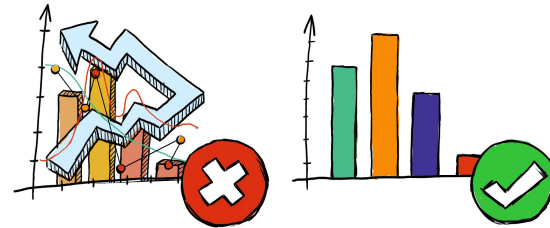
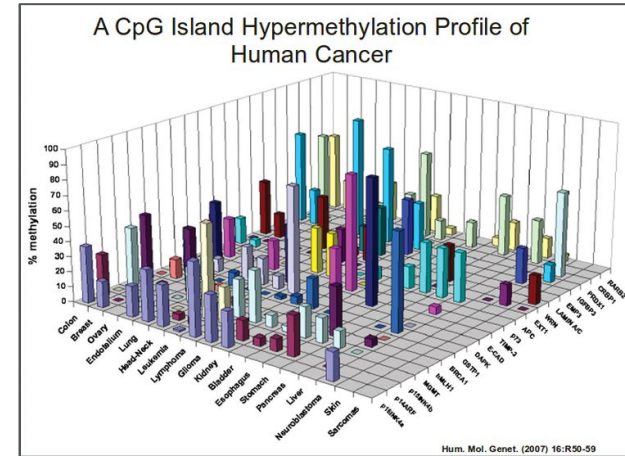
- ❖ **Libraries are primarily open-source and free to use**
- ❖ **Some companies offer paid platforms with more features and customer support**
- ❖ **Download statistics not publicly available**

Market Share By Product

Product	Customers	Market Share
Power BI	80,621	16.00%
Tableau	76,493	15.18%
D3.js	43,965	8.73%
Grafana	19,826	3.93%
Heap Analytics	19,087	3.79%
Others	263,874	52.37%

Current Relevant Problems

- ❖ Data Overload and Clutter
- ❖ Balancing Aesthetics and Functionality
- ❖ Accessibility and Inclusivity
- ❖ Misinterpretation and Misrepresentation
- ❖ Integration with Emerging Technologies



Competitors & Emerging Technology

- ❖ AI-Driven Visualization Tools
- ❖ Augmented Reality (AR) and Virtual Reality (VR)
- ❖ Real-Time Data Visualization Platforms
- ❖ Alternative Visualization Libraries



UNREAL
ENGINE

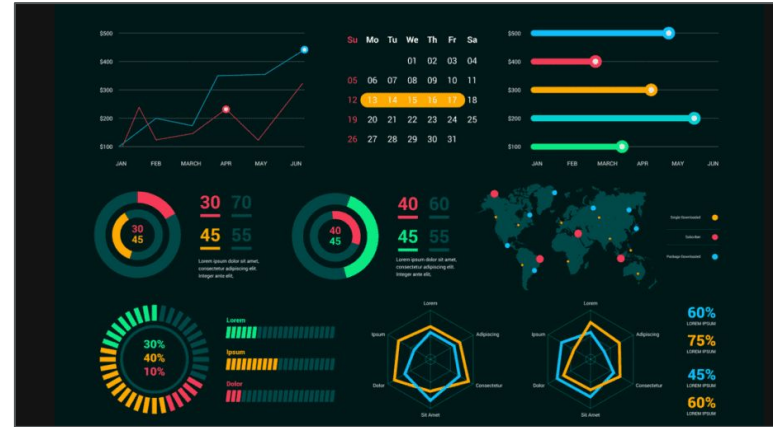


einstein analytics



Future Prognosis

- ❖ User Experience and Storytelling
- ❖ Predictive and Prescriptive Visualizations
- ❖ Integration with IoT and Big Data Analytics
- ❖ Ethical Considerations and Data Privacy



Ongoing Research & Literature Insights

- ❖ Adoption of Natural Language Processing (NLP)
- ❖ Evolution of Data Democratization
- ❖ Emphasis on Ethical Design and Data Privacy
- ❖ Ongoing Research Highlights



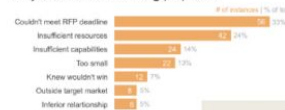
Client Agency	Project Name	Start Date	End Date	Status	Phase 1 - Objective	Phase 2 - Objective	Phase 3 - Objective	Phase 4 - Objective	Phase 5 - Objective	Phase 6 - Objective	Phase 7 - Objective	Phase 8 - Objective	Phase 9 - Objective	Phase 10 - Objective	Phase 11 - Objective	Phase 12 - Objective	Phase 13 - Objective	Phase 14 - Objective	Phase 15 - Objective	Phase 16 - Objective	Phase 17 - Objective	Phase 18 - Objective	Phase 19 - Objective	Phase 20 - Objective
Client Agency	Project Name	Start Date	End Date	Status	Phase 1 - Objective	Phase 2 - Objective	Phase 3 - Objective	Phase 4 - Objective	Phase 5 - Objective	Phase 6 - Objective	Phase 7 - Objective	Phase 8 - Objective	Phase 9 - Objective	Phase 10 - Objective	Phase 11 - Objective	Phase 12 - Objective	Phase 13 - Objective	Phase 14 - Objective	Phase 15 - Objective	Phase 16 - Objective	Phase 17 - Objective	Phase 18 - Objective	Phase 19 - Objective	Phase 20 - Objective

Timing & resource issues are top reasons we aren't submitting/winning proposals

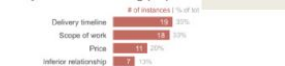
Global project summary



Why we aren't submitting proposals



Why we aren't winning proposals



Conclusion

- ❖ Overview of Data Visualization and Its Importance
- ❖ Key Purposes and Common Terms in Data Visualization
- ❖ Features and Functions of Visualization Tools and Libraries
- ❖ Product Families and Business Context
- ❖ Current Relevant Problems and Competitors in Data Visualization
- ❖ Future Prognosis and Insights from Literature
- ❖ Ongoing Research Highlights

Sanjay's References

LinkedIn. (2024). LinkedIn.com. <https://www.linkedin.com/pulse/data-visualization-market-analysis-unraveling-qc6te/>

Meeks, E. (2018, December 7). *3rd Wave Data Visualization - Nightingale - Medium*. Medium; Nightingale.

<https://medium.com/nightingale/3rd-wave-data-visualization-824c5dc84967>

Now you see it : simple visualization techniques for quantitative analysis - Catalog - UW-Madison Libraries. (2024). Wisc.edu.

<https://search.library.wisc.edu/catalog/9910206923802121>

Shen, L., Shen, E., Luo, Y., Yang, X., Hu, X., Zhang, X., Tai, Z., & Wang, J. (2023). Towards Natural Language Interfaces for Data Visualization: A Survey.

IEEE Transactions on Visualization and Computer Graphics, 29(6), 3121–3144. <https://doi.org/10.1109/tvcg.2022.3148007>

The Truthful Art. (2022). Thefunctionalart.com. <http://www.thefunctionalart.com/p/the-truthful-art-book.html>

What are the latest trends and innovations in data visualization tools and platforms? (2024). LinkedIn.com.

<https://www.linkedin.com/advice/3/what-latest-trends-innovations-data-visualization-1c>

Sujay's References

What is D3? | D3 by Observable. (2022). D3js.org. <https://d3js.org/what-is-d3>

Learn D3. (2023, December 15). CORP-MIDS1 (MDS). <https://www.mastersindatascience.org/learning/programming-languages/javascript/d3/>

Interactive Data Visualization of Geospatial Data using D3.js, DC.js, Leaflet.js and Python // Adil Moujahid // Bridging Tech and Art. (2016). Adilmoujahid.com. <https://adilmoujahid.com/posts/2016/08/interactive-data-visualization-geospatial-d3-dc-leaflet-python/>

An Extremely Detailed Map of the 2020 Election (Published 2021). (2024). The New York Times. <https://www.nytimes.com/interactive/2021/upshot/2020-election-map.html>

d3/d3: Bring data to life with SVG, Canvas and HTML. :bar_chart::chart_with_upwards_trend::tada: (2023, June 3). GitHub. <https://github.com/d3/d3>

Suraj's References

Tableau Software. (n.d.). *Tableau: Business Intelligence and Analytics Software*. <https://www.tableau.com/>

Rouse, M. (2019). *What is Tableau?*. TechTarget.
<https://searchbusinessanalytics.techtarget.com/definition/Tableau-Software>

Tableau Software. (n.d.). *Tableau Desktop: Connect to Data*.
https://help.tableau.com/current/pro/desktop/en-us/connect_basic.htm

Tableau Software. (n.d.). *Get Started with Tableau Desktop*.
<https://help.tableau.com/current/guides/get-started-tutorial/en-us/get-started-tutorial-home.htm>

Feldman, B. (2018). *9 Companies Using Tableau to Drive Results*. Grow.
<https://www.grow.com/blog/companies-using-tableau>

Tableau Software. (n.d.). *Netflix: Revolutionizing Data-Driven Decision-Making with Tableau*.
<https://www.tableau.com/solutions/customer/netflix-revolutionizing-data-driven-decision-making-tableau>

Tableau Software. (n.d.). *Industries*. <https://www.tableau.com/solutions/industries>

Charlotte's References

S. Berinato, "Visualizations That Really Work," in Harvard Business Review, pp..92-100, June 2016.

M. Friendly, "A Brief History of Data Visualization," Chapter 1 in "Handbook of Data Visualization," Springer Handbooks Comp. Statistics, Springer, Berlin, 2008.

Google - data visualization tools logos, <https://www.google.com/search?q=data+visualization+tools+logos>.

Google - data visualization libraries logos, <https://www.google.com/search?q=data+visualization+libraries+logos>.

IBM - "What is Data Visualization?," <https://www.ibm.com/topics/data-visualization>.

Investopedia - product family overview, <https://www.investopedia.com/terms/p/product-family.asp>.

*Product School - data visualization tools for product managers,
<https://productschool.com/blog/product-fundamentals/data-visualization-platforms-product-manager>.*

Tableau - "What Is Data Visualization?," <https://www.tableau.com/learn/articles/data-visualization>.

The University of British Columbia - guide to choosing a data visualization library or tool, <https://www.cs.ubc.ca/~tmm/courses/547-20/tools/>.

Unleashed - business reporting tools, <https://www.unleashedsoftware.com/blog/why-business-reporting-is-important-for-business-success>.

Venngage - overview of data visualization, <https://venngage.com/blog/data-visualization/>.

David's References

Analyzing PyPI package downloads. (2024). Python. <https://packaging.python.org/en/latest/guides/analyzing-pypi-package-downloads/>

Dash Enterprise. (2024). Plotly. <https://plotly.com/dash/>

Data Visualization. (2024). 6sense. <https://6sense.com/tech/data-visualization>

Melvin L. (2015, December 10). *Plotly using R* [Video]. YouTube. https://www.youtube.com/watch?v=n4tx4-RRhdU&list=PPSV&ab_channel=MelvinL.

Noves, V. (2022, November 29). *Mastering Data Visualization in the AEC Industry.* e-verse.
<https://e-verse.com/learn/an-introductory-guide-for-aec-data-visualization/>

Soper, T. (2023, March 1). *Salesforce stock spikes 14% after beating Q4 estimates.* GeekWire.
<https://www.geekwire.com/2023/salesforce-stock-spikes-14-after-beating-q4-estimates-tableau-revenue-grows-3-to-636m/>

Stančin, I., & Jović, A. (2019, May 1). *An overview and comparison of free Python libraries for data mining and big data analysis.* IEEE Xplore.
<https://doi.org/10.23919/MIPRO.2019.8757088>

The Tableau Platform. (2024). Tableau. <https://www.tableau.com/products/our-platform>.

What is Data Visualization? (2024). Tableau. <https://www.tableau.com/learn/articles/data-visualization>.