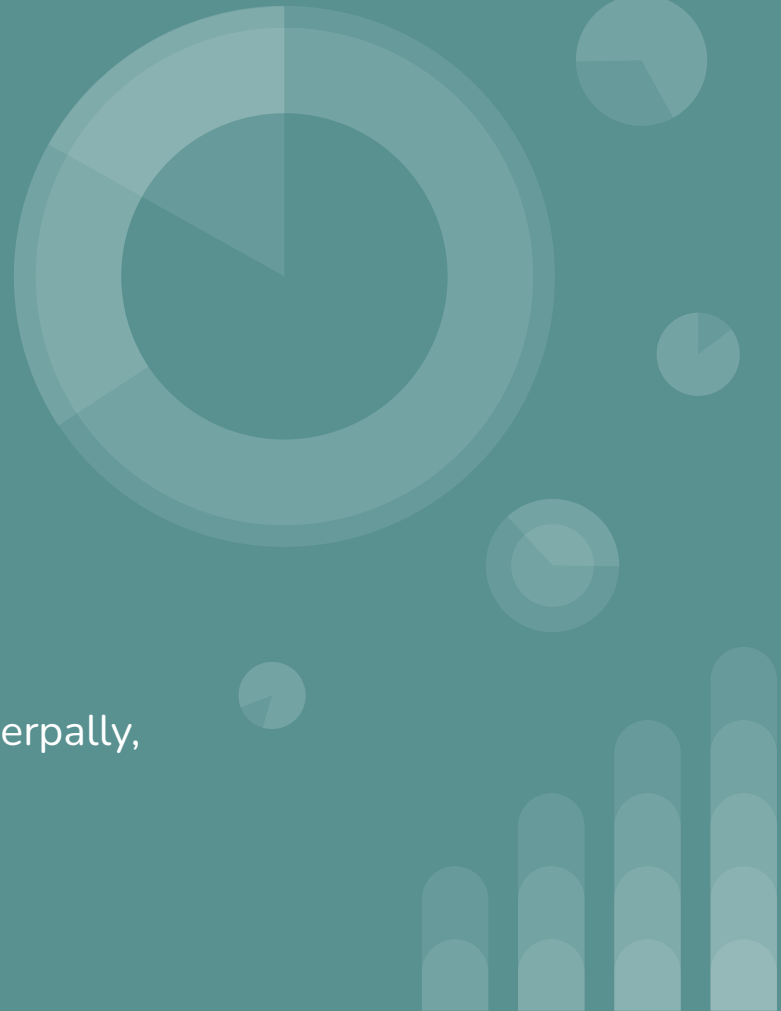# Data Visualization

Yousef Ibrahim, Chrystabel Sunata, Sai Arun Daverpally, Jerry Song, Melissa Leng

Introduction

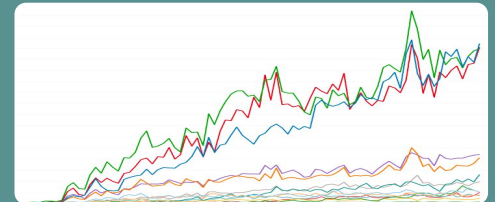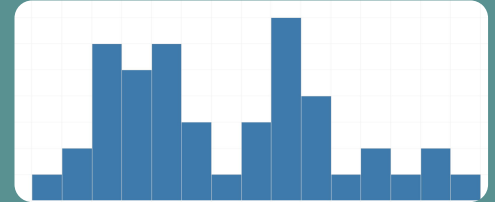# What is Data Visualization?

The graphical representation of data using visual formatting such as charts, graphs, and maps

Purpose:

- Converts complex datasets into clear, visual formats
- Help identify patterns, trends, and outliers

Benefits

- Enhances decision-making by making data more accessible
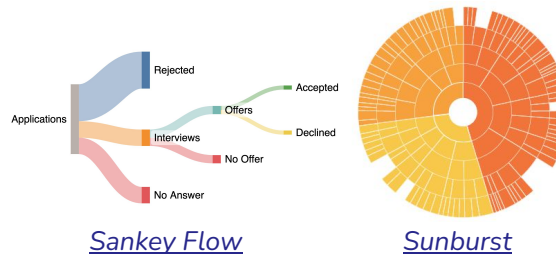- Communicates insights effectively through visual storytelling

Introduction to Data Visualization

Key Terms

- **Data Connectivity**
- **User Interface**
- **Visualizations**
- **Flexibility**
- **Collaboration**
- **Analytics**

Theme River
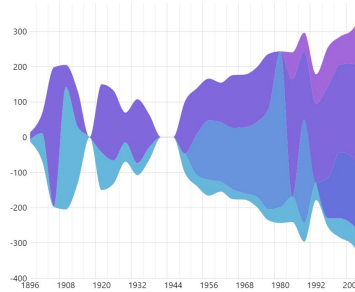
Sankey Flow

Sunburst

Role in Business Applications

- **Enhancing Data Interpretation –** *Simplify complex data*
- **Improving Decision-Making –** *Identify trends and outliers*
- **Boosting Collaboration –** *Share insights across teams*
- **Performance Monitoring –** *Track key business metrics*

Product Overview

Key Features

- Data Connectivity
  - Support multiple data sources
  - Real-time data integration
- User-Friendly Interface
  - Intuitive drag-and-drop interface
  - Pre-built templates for reports and dashboards
- Interactive Visualizations
  - Apply filters
  - Drill-down functionality
  - Cross-chart interactivity

- Customization and Flexibility
  - Various types of charts
  - Custom design, layout, fonts, etc.
  - Custom scripting
- Collaboration and Sharing
  - Export in multiple formats
  - Collaboration features
- Advanced Analytics
  - Statistical functions
  - Predictive analytics
  - AI-powered insights

Comparative Chart

	Tableau	Power BI	D3.js
Ease of Use	High with user-friendly interface, drag-and-drop functionality	Moderate with Excel-like interface, drag-and-drop functionality	Low, requires JavaScript knowledge
Data Connectivity	Widest range of data sources (databases, spreadsheets, etc.)	Best for Microsoft ecosystem, supports APIs and cloud services	Requires manual integration of APIs and databases
Visualization Capabilities	Highly customizable and interactive dashboards	Good, but less flexible than Tableau	Full control over custom visualizations
Scalability	Best for large-scale enterprises	Scalable for mid-large businesses	Unlimited
Interactivity with Data	Very strong, features cross-chart filtering, drill-downs	Strong, but limited customized filter actions	Very strong, but requires coding
Pricing	Expensive ($70/month)	Affordable ($10/month)	Free (open-source)

Technical Analysis
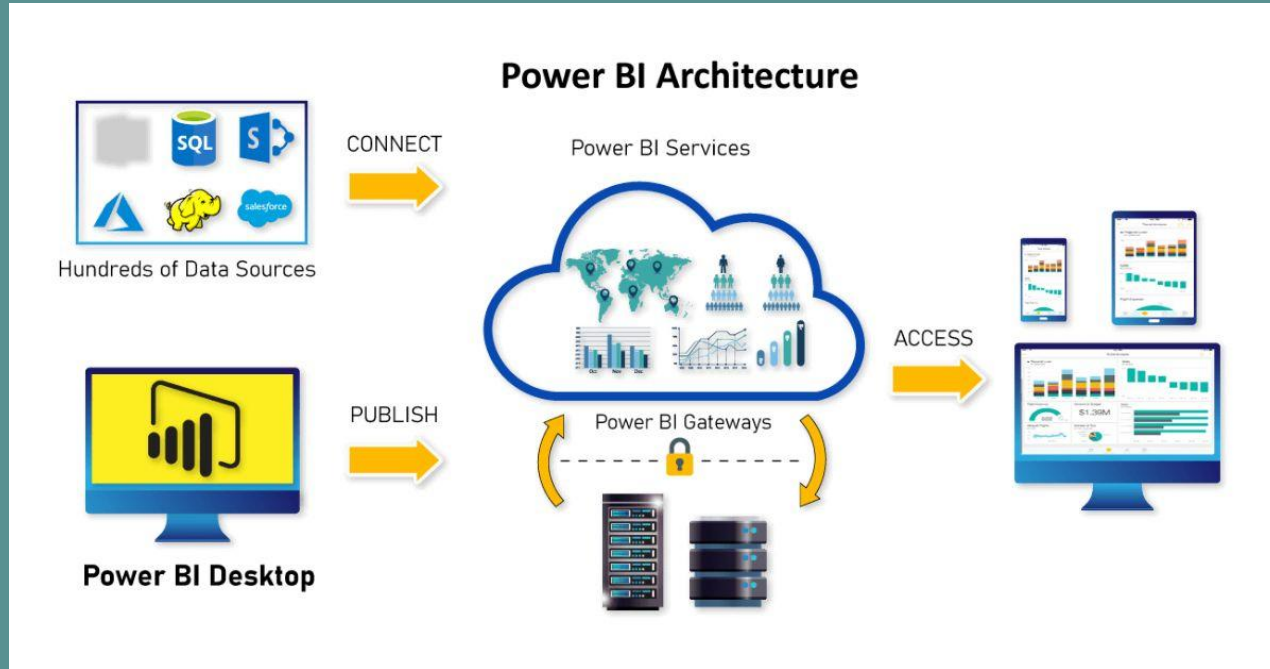
Technical Analysis: Power BI

- PowerBI is a data visualization platform offered by Microsoft. It's part of a collection of low-code/no-code tools called **Power Platform**, which includes Power Apps, Power Automate, and Copilot Studio.
- The architecture of PowerBI consists of several components:
  - **PowerBI Desktop** - A desktop application used by individuals to create visualizations; the starting point.
    - Report - Detailed visualizations created from data obtained from connected datasources, like Excel/CSV, and No-SQL, SQL, and cloud databases.
    - Dashboard - High-level, single-page visualization built using created reports, and it involves the visualization of more generic, aggregated information.
  - **PowerBI Service** - A cloud-based version of Desktop, where individuals can access and share their published reports and dashboards. It's connected to OneDrive, which makes it easier for organizations to share files.
    - **PowerBI Gateway** - A middleware used for connecting on-premises data sources with PowerBI Service.
  - **PowerBI Mobile** - A mobile-based version of Desktop used for quick-access of reports, not development.
  - **PowerBI Premium** - A subscription that allows for more advanced features to cloud, including:
    - Asynchronous refresh, Automatic Aggregations, Backup and Restore
    - Autoscale - Adds more v-cores when capacity is overloaded.
    - Direct Lake and DirectQuery - Connect directly to your datatable and load data in real-time, without importing the data, which might return an instance of the database.
    - **PowerBI Report Server** - An on-premises solution that allows organizations to monitor and maintain their published reports directly, and extract useful information, like total incidents, total sales, and profit margin.
  - **PowerBI Embedded** - An Azure service that allows individuals to embed published reports and interactive charts directly into their application via an API.
- PowerBI allows for **Dax** (Data Analysis Expressions), which is a formula expression language used to perform advanced data transformations.
- PowerBI encrypts all data by controlling access using row-level security, and all persisted data is protected using Microsoft-managed keys.

Technical Analysis: Power BI

- **Typical Use Cases**
  - Analyze supply chain KPIs (Key Performance Indicators), like on-time delivery, purchase order tracking, monthly sales growth.
  - Facilitate financial reports using interactive charts, which can be shared online to upper management within an organization.
  - Understand healthcare data, like patient records, treatment outcomes, and operational costs, using charts that can present this data in one page, while adhering to HIPAA regulations.
- **Key Technical Differentiators**
  - Dashboards can be created with existing reports, and are updated automatically after report updates, which allows for efficient organization of visualizations.
  - PowerBI allows users to combine data from both NoSQL and SQL databases in a single workflow by allowing users to apply data transformations on them.
  - Power Query allows individuals to perform data cleaning and transformation directly in PowerBI, removing the need to perform it elsewhere.
  - Offers an on-premises solution, PowerBI Report Server, which gives organizations more power in maintaining their published reports in-house.
  - PowerBI's low-code/no-code structure expands the user base to non-technical individuals, including business analysts.

Technical Analysis: Power BI

Power BI Architecture

CONNECT

Hundreds of Data Sources

Power BI Desktop

PUBLISH

Power BI Services

Power BI Gateways

ACCESS

Technical Analysis:

- Built by Observable in 2011, D3.js, or commonly known as D3 (Data-Driven Documents), is an open-source JS library used for visualizing data using SVG (Scalable Vector Graphics) and Canvas.
- D3 works as an API collection of modules, which can be used independently or combined together to build larger projects.
- D3 is a low-level toolbox of modules, which span over multiple categories:
    - Visualization - Contains all modules that allow granular access to each part of building a visualization.
        - d3-shape - Contains shape generators for complex/generic shapes. Includes: Arcs, Areas, Curves, Pies, Stacks, Radial Areas/Lines/Links (in polar coordinates).
        - d3-scale - Allows for creating scales for visual encoding. Includes: Linear, Time, Threshold, Sequential, Band, Point.
        - d3-hierarchy - Allows for hierarchical visualizations of micro-observations of individual elements and macro-observations of large groups. Includes: Hierarchies, Tree, Cluster, Partition, Treemap.
    - Animation - Contains modules needed to create animations.
        - d3-transition - Allows for smooth transitions to animating changes in the DOM (Document Object Model) over a given duration. Includes: Timing, Control Flow, Selecting/Modifying Elements
        - d3-ease & d3-timer - Allow for creating smoother animations, and contains a queue for managing consistent, synchronized, concurrent/staged animations, respectively.
    - Interaction - Allows for user interactions for the created SVGs, enabling users to interact with the data and modify it for their needs. Contains d3-zoom, d3-drag.
    - Data - Contains modules that allow for retrieving and storing data in multiple different formats, including performing minor data cleaning/transformations.
        - d3-array - Allows for array manipulation and operations with array as the underlying data structure. Includes: blurring, bisecting, binning, grouping, sorting, summarizing, and transforming data.
        - d3-dsv - Provides parser and formatter for delimiter-separated values, like CSV or TSV.
        - d3-random - Allows for generation of random values from various distributions. Includes: Uniform, Normal, Bates, Exponential, Geometric, Bernoulli, Binomial distributions.

Technical Analysis:

- **Typical Use Cases**
    - Perform exploratory story-telling using D3.js with dynamic visualizations with transitions into different states. This is useful for many fields, especially healthcare and journalism.
    - Create geographic maps or charts with interactive scales to allow users to track changes over a given time period.
- **Key Technical Differentiators**
    - Offers an equivalent high-level sister library called Plot, which can be used to create quick visualizations without requiring extensive background of D3's architecture and modules.
    - Allow high flexibility for building dynamic charts using a bottom-up approach, which ensures that final visualization meets all development requirements.
    - Follows modern web standards, which allows developers to integrate D3 into any project using modern web technologies, like React and Django.
    - Being open-source, D3.js is free to use, is actively being developed with new features, and fosters a community of developers who share the goal of enhancing the library to be more efficient.

Sample Applications

Coca-Cola and Power BI

1) Real Time Sales and Market Performance
2) Supply Chain and Logistics Optimization
3) Consumer Insights and Marketing

Coca-Cola: Market Performance and Customer Insights

Measuring Market Performance

- Uses Power BI dashboards to track market share trends and inventory level sales
- DAX formulas native to Power BI calculate pricing elasticity, sales velocity etc.
- Enables real-time adjustments in pricing and promotions
- Broken down into regions, product, category, distributor, etc.

Visualizing Customer Insights

- Integrates CRM and social media sentiment data for real-time analytics.
- Uses AI in Power BI to predict customer behavior
- Show customer sentiment via Twitter API, Google Trends, etc.
- Personalized marketing campaigns increase sales effectiveness.

Coca-Cola: Supply Chain Optimization and Integration

- Data sourced from vendors, independent bottlers, smart vending machines, distributors and retailers
- Uses Azure SQL and SAP HANA to query large datasets
- Uses ETL processing to move and cleanse data from source

- Uses IoT sensors in bottling plants for real-time tracking.

- Monitors machine uptime, production efficiency, and maintenance schedules.

- Integrates GPS tracking for fleet optimization.

- AI-driven forecasting models reduce waste and improve efficiency.

Netflix: Tableau & D3.js

Tableau is mainly used for Business Intelligence but Netflix typically uses D3.js to aid in interactive visualizations

1) Performance of Content (TV shows and Movies)
2) Customer Interaction and Marketing
3) Streaming Quality and Availability

Netflix: Content Performance Analytics

Tableau

• Processes data from AWS Redshift, Presto, and Apache Spark.

• Uses Tableau dashboards to analyze watch time, drop-off rates, and binge behavior etc.

• Tracks regional, demographic etc. differences in content performance

• Helps executives and content strategists make renewal and content investment decisions

D3.js

•Key visualizations:

• - Network Graphs for viewer similarity and content clustering

• - Time-Series Graphs for drop-off rate analysis per episode

• - Stacked Bar Charts for genre performance predictions

Netflix: Customer Retention & Churn Prediction

Tableau

• Integrates Amazon SageMaker ML models with Tableau dashboards

• Predicts churn probability based on subscriber engagement and identify user groups at risk

• Tableau helps visualize retention strategies and optimize interventions and plan promotional content/engagement

• Data driven engagement tactics reduces churn to 2-3%, saving Netflix $1B annually

D3.js

• Netflix continuously tests UI changes and recommendation strategies.

• D3.js visualizations allow:

• - Heatmaps to analyze user engagement (CTR, watch time, abandonment)

Netflix: Streaming Quality & Day to Day Operations

Tableau

• Uses Tableau to visualize streaming latency location by location

• Engineers monitor buffering rates and network performance.

• Helps prevent and resolve streaming issues before they impact users

• Tableau is used in a preventative measure to ensure Netflix has the highest availability with the highest quality

• Optimize performance to reduce network costs for company

D3.js enables:

• - Live Streaming Latency Maps (Choropleth visualizations)

• - Bitrate Adaptation Line Charts

• - Bubble Charts for Streaming Error Analysis

• Engineers use WebSockets + Kafka to push real-time data into D3.js visualizations.

• Reduced streaming issue response times by 40%.

Future Trends

Improving Accessibility and Interactivity

Accessibility

• Generating easy-to-understand visualizations

• Making the process of creating visualizations accessible to non-experts and easier to learn

• Creating data that is accessible through different platforms (like on mobile devices)
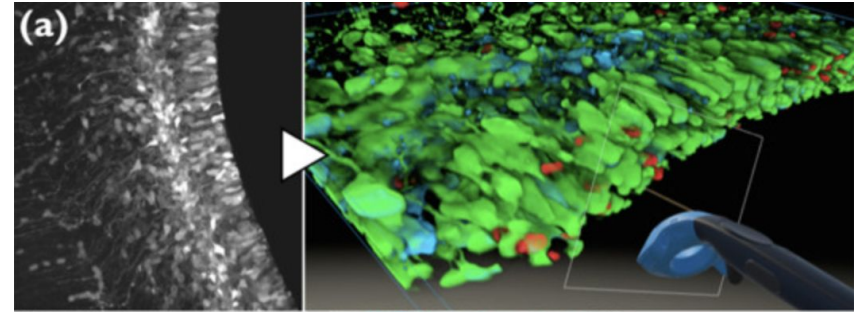
Interactivity

• Creating new ways to interact with data

- • AR/VR to visualize data in a 3D space

• Adapting to different user inputs

- • reacting to a user's emotions

Visualizations in VR

Analyzing data in 3D

• Currently: data in 3D can be viewed in 2D by flattening it or encoding the 3rd dimension in a different way like color

• With VR: 3D geometry of data can be seen

example: On the left, a maximum intensity projection of an image stack of neurons imaged by spinning disk confocal microscopy. On the right, a capture of the same data set in VR.

Virtual Reality: Beyond Visualization

Visualizations in VR

- VR allows for an intuitive measurement of space/relative positioning in the data
- VR also allows for users to easily navigate through the data
    - easy scaling, positioning, rotating

Applications:

- Navigation: martian landscapes, caves
- Used in paleontology, physics, organic chem, MRI, shape perception, brain tumors

Generative AI

• Generative AI has shown
potential in creating
visualizations, even producing
3D ones
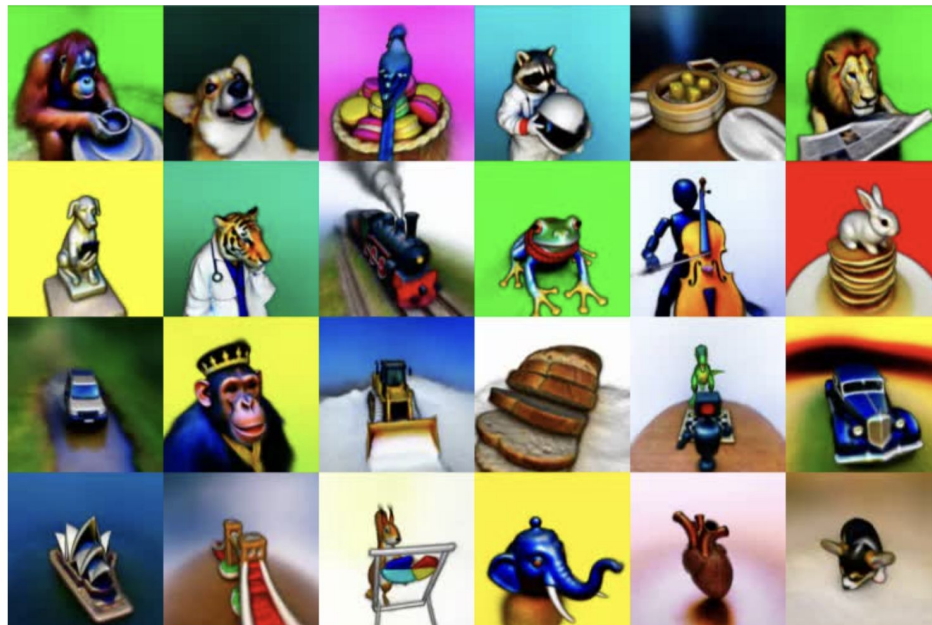
Given a caption, DreamFusion generates relightable 3D objects with high-fidelity appearance, depth, and normals. Objects are represented as a Neural Radiance Field and leverage a pretrained text-to-image diffusion prior such as Imagen.

DreamFusion

Generative AI

Data Processing

- perform data processing and augmentation quickly
- transform data: data inference or embedding
- generate visualization images
- stylize visualizations, create a storyline, and annotate data/visualizations
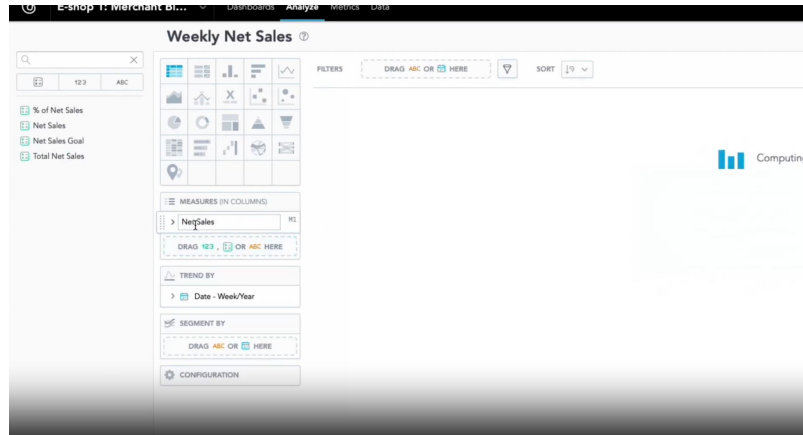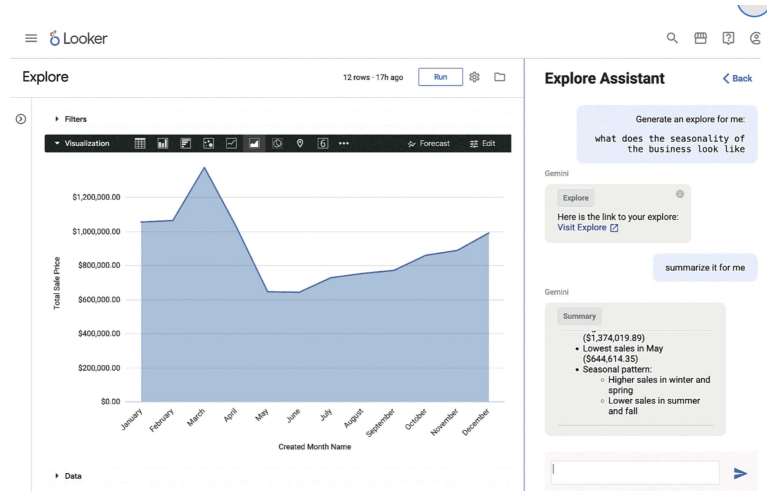- answer questions about content

Generative AI

Challenges

- Generating visualization images is much harder than creative images
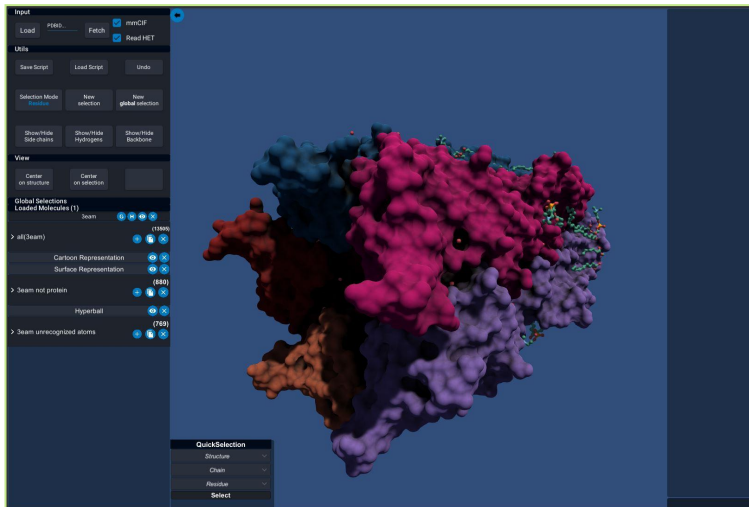- AI needs to be trained on more complex data than general AIs

Examples

GoodData

Looker

Examples

UnityMol

iViz

Thank You!

Questions?