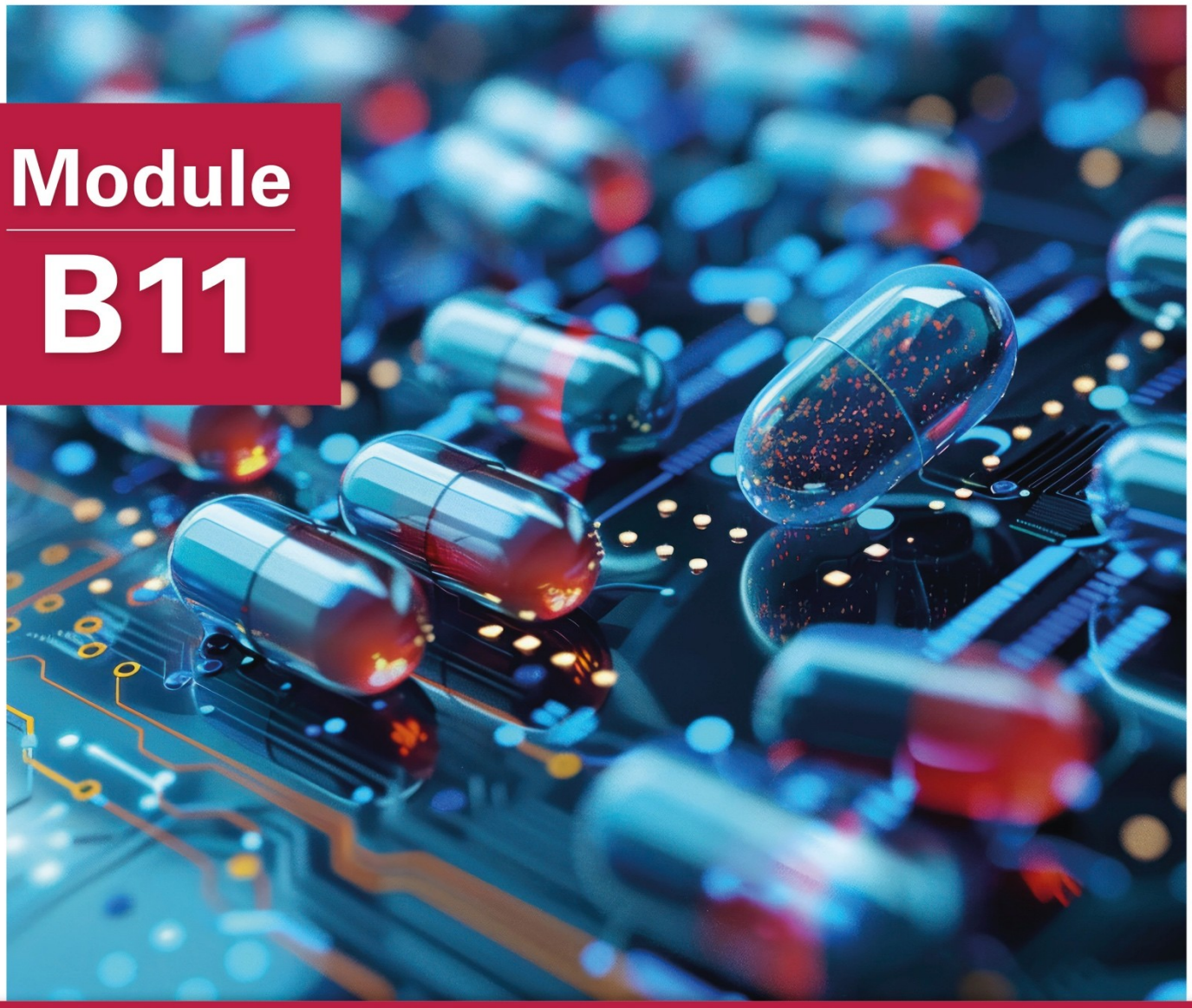
AfricaCDC
Centres for Disease Control and Prevention

Safeguarding Africa's Health



# Module B11

# Bioinformatics Foundational Course

## Data Visualization and Report Generation

NGS Academy for the Africa CDC

ASLM
AFRICAN SOCIETY FOR LABORATORY MEDICINE

H3ABioNet
Pan African Bioinformatics Network for H3Africa

ASLM ACADEMY

# Module B11

# Data Visualization and Report Generation

back to the table of modules

**Module last updated:**
December 2024

| | |
|---|---|
| **Suggested or approximate number of sessions** | 2 |
| **Suggested or approximate total learning time** | 4 hours |
| **Target audience** | Bioinformaticians |
| **Delivery format** | Lectures, videos |
| **Level of the module** | Introductory and intermediate |

## Contributors

Yiqun Li and Perceval Maturure.

## Module description

This module introduces participants to the fundamental principles and practical applications of data visualization, combining theoretical foundations with hands-on experience. Starting with historical context and cognitive principles, the course explores various visualization techniques, from basic static charts to interactive dashboards and geographic representations. Students learn essential data preparation methods, master industry-standard tools including programming libraries like Python's Matplotlib and business intelligence platforms like Tableau, while understanding crucial design principles around color theory, typography, and layout.

The module then advances to the strategic aspects of data visualization, emphasizing storytelling techniques and ethical considerations in visual communication. Participants develop skills in creating compelling narratives through data, designing effective presentations and dashboards, while learning to address critical concerns such as representing uncertainty, avoiding bias, and ensuring accessibility. Through practical exercises and real-world applications, students learn to transform complex data into clear, actionable insights while maintaining high standards of ethical representation and technical excellence. In this module, participants are introduced to the following topics and/or concepts:

1. **Data Management Foundation**
   - Support for multiple input formats (CSV, JSON, SQL databases)
   - Data cleaning, preprocessing, and transformation capabilities
   - Handling missing data and data quality issues
   - Data aggregation and summarization features
   - Efficient handling of large datasets with caching mechanisms

2. **Visualization Components**
   - Core chart library (bar, line, pie, scatter plots, etc.)
   - Advanced visualizations (heatmaps, box plots, geographic maps)
   - Interactive visualization capabilities
   - Customization options for visual elements
   - Accessibility features and color theory implementation
   - Real-time preview functionality

3. **Report Generation and Management**
   - Structured report creation combining narratives and visualizations
   - Multiple output format support (PDF, HTML, PowerPoint)
   - Template system for common report types
   - Automated and scheduled report generation
   - Version control and change tracking
   - Parameterized reports for dynamic updates

4. **User Interface and Experience**
   - Intuitive dashboard creation interface
   - Drag-and-drop functionality
   - Real-time visualization preview
   - Mobile-responsive design
   - User guidance and tooltips
   - Interface customization options

5. **Collaboration and Sharing**
   - Report and dashboard sharing capabilities
   - Collaborative editing features
   - Commenting and annotation tools
   - Version history and change tracking
   - User permissions and access controls
   - Embedding support for external applications

6. **Analytics and Intelligence**
   - Basic statistical analysis tools
   - Trend detection and forecasting
   - Pattern recognition capabilities
   - Automated insights generation
   - Custom metric creation
   - Data-driven alerts and notifications

7. **System Architecture**
   - API-first design for integration
   - Plugin architecture for extensibility
   - Performance optimization for large datasets
   - Scalability for concurrent users
   - Caching and query optimization
   - Security implementation (authentication, authorization)

8. **Best Practices and Guidelines**
   - Visualization selection guidelines
   - Data storytelling principles
   - Performance optimization techniques
   - Security and privacy considerations
   - Accessibility standards
   - Documentation and training materials

## Module learning outcomes

On completion of this module, the participants will have a basic knowledge of, or will be able to:

1. **Data Understanding & Preparation**
   - Evaluate data quality, structure, and relationships to determine appropriate visualization approaches
   - Apply data cleaning, transformation, and preprocessing techniques to prepare datasets for visualization
   - Implement data aggregation and summarization methods to extract meaningful insights
   - Design effective data structures optimized for visualization performance

2. **Visualization Techniques**
   - Create appropriate visualizations based on data types, relationships, and intended message
   - Develop interactive visualizations that allow users to explore and understand complex datasets
   - Apply color theory, typography, and layout principles to enhance visual communication
   - Evaluate and select appropriate visualization libraries and tools for different scenarios

3. **Report Design & Generation**
   - Design structured reports that effectively combine visualizations with narrative context
   - Implement automated report generation systems with parameterized inputs
   - Create reusable report templates that maintain consistency across an organization
   - Integrate multiple data sources and visualization types into cohesive reports

4. **Technical Implementation**
   - Write efficient code to process and visualize large datasets
   - Implement caching and optimization strategies for improved performance
   - Develop APIs for integrating visualization systems with external applications
   - Build scalable solutions that handle concurrent users and real-time updates

5. **Communication & Best Practices**
    - Analyze audience needs and context to create targeted visualizations and reports
    - Apply accessibility standards to ensure visualizations are inclusive
    - Present complex data insights to both technical and non-technical stakeholders
    - Implement data privacy and security best practices in visualization systems

6. **Professional Skills**
    - Collaborate effectively in teams to create visualization solutions
    - Document visualization systems and processes for maintainability
    - Debug and troubleshoot common visualization challenge
    - Evaluate and critique visualizations based on effectiveness and accuracy

## Module assessments

Module practical: Not applicable
Module quiz: Assessment questions available on the ASLM platform

## Module resources

- Babraham Institute | Webpage - Core Bioinformatics Skills: Training Resources

## References


- ClaudeAI. (2024). ClaudeAI response on Data Visualisation and Reporting multiple choice questions. Retrieved Sept 16, 2024, from https://claude.ai/
- ClaudeAI. (2024). ClaudeAI response on Data Visualisation and Reporting Module content. Retrieved Aug 06, 2024, from https://claude.ai/


## Acknowledgements


We would like to thank the following individuals, in alphabetical order of last name, for their valuable time and effort spent in designing (i.e., drafting, reviewing, and refining) this module: **Yiqun Li and Perceval Maturure**.

Furthermore, we would like to thank the following institutions, societies, journals and individuals from whom we sourced open-access resources, used in this module:
Babraham Institute.