Module 5: Data Analysis & Visualisation for Decision-Making

Inclusion Data Quest

WE ARE OPEN

Module Objective

To understand advanced analysis and visualization techniques to transform raw data into actionable insights.

Additionally, to empower stakeholders to make informed, data-driven decisions that enhance program design and optimize outcomes.

Agenda

1. Introduction to Data Analysis Tools
2. Data Visualization Techniques
3. Integrating visualization for Socio-Economic Data
4. Visualizing intersectionality in data
5. Power of effective communication
6. Effective storytelling

1. Introduction to Data Analysis Tools

What is Data Analysis

Data analysis is the systematic process of inspecting, cleaning, transforming, and interpreting raw data to uncover meaningful patterns, trends, and insights. It involves applying statistical, computational, and visualization techniques to answer questions, solve problems, or inform decisions.

Why is Data Analysis Important?

1. **Evidence-Based Decision-Making**:
   - Transforms raw numbers into actionable insights for policymakers.
   - Example: Using disease prevalence data to prioritize funding for HIV clinics.
2. **Improved Efficiency**:
   - Identifies inefficiencies in resource allocation (e.g., underused health facilities).
   - Example: Analyzing patient wait times to optimize staff schedules.
3. **Accountability & Transparency**:
   - Tracks progress toward development goals (e.g., SDGs) and holds governments accountable.
   - Example: Public dashboards showing budget vs. actual health spending.
4. **Equity & Inclusion**:
   - Reveals disparities (e.g., access to care for disabled populations) to design targeted interventions.
   - Example: Mapping prenatal care access gaps in rural wards.
5. **Predictive Capabilities**:
   - Forecasts outbreaks (e.g., cholera) using historical data, enabling proactive responses.

What is your objective?

Why do you want to communicate this data? What do you hope to achieve afterwards? (engagement, awareness, mindset shift, action)

Table 2.6a: Distribution of Population Age 3 Years and Above by School Attendance Status, Sex and Special Age Groups

		Total	3-5	4-5
KENYA	**Total**	**43,739,906**	**3,739,232**	**2,499,258**
	Male	**21,574,077**	**1,881,994**	**1,263,032**
	Female	**22,164,531**	**1,857,146**	**1,236,160**
	Intersex	**1,298**	**92**	**66**
At School/Learning Institution	**Total**	**17,780,277**	**2,123,269**	**1,800,642**
	Male	9,040,318	1,050,475	898,744
	Female	8,739,524	1,072,746	901,857
	Intersex	435	48	41
Left School/Learning Institution After Completion	**Total**	**11,568,544**	**10,080**	**8,292**
	Male	5,910,664	5,021	4,183
	Female	5,657,565	5,059	4,109
	Intersex	315	-	-
Left School/Learning Institution Before Completion	**Total**	**6,854,538**	**3,502**	**2,806**
	Male	3,172,382	1,750	1,402
	Female	3,681,910	1,752	1,404
	Intersex	246	-	-
Never Been to School/Learning Institution	**Total**	**7,120,542**	**1,570,861**	**671,146**
	Male	3,223,856	808,502	350,251
	Female	3,896,417	762,316	320,871
	Intersex	269	43	24
Don't Know	**Total**	**406,887**	**30,988**	**15,991**
	Male	222,301	15,975	8,256
	Female	184,558	15,012	7,734
	Intersex	28	1	1
Not Stated	**Total**	**9,118**	**532**	**381**
	Male	4,556	271	196
	Female	4,557	261	185
	Intersex	5	-	-

Data Analysis Tools

Excel

Learn essential functions. Perform basic calculations and charting.

SPSS

Explore statistical analysis. Conduct hypothesis testing.

R

Dive into advanced analytics. Use powerful statistical libraries.

Python

Automate data tasks. Build custom analysis workflows.

R

R is a free, open-source programming language and software environment specifically designed for statistical computing, data analysis, and visualization. Developed in the 1990s, it has become a cornerstone tool in academia, research, and industries like public health, finance, and bioinformatics.

Key Features

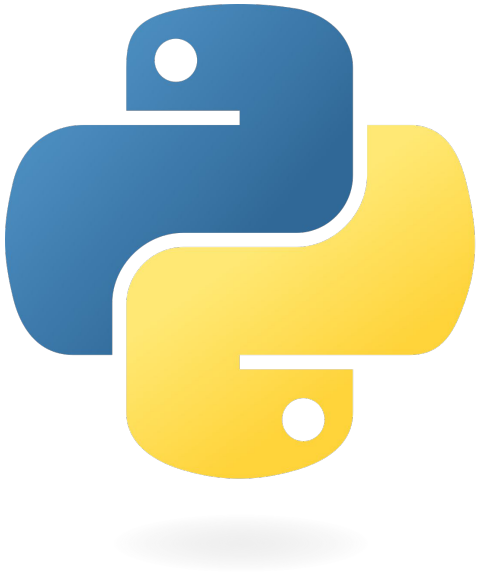
- **Statistical Power**: Built for advanced statistical modeling (regression, hypothesis testing, machine learning).
- **Data Visualization**: Packages like ggplot2 create publication-quality graphs (e.g., disease trend maps, demographic dashboards).

Download R from here: https://cran.r-project.org/

Getting Started with R

1. Import any data → Clean with `dplyr` (remove duplicates, filter outliers).
2. Disaggregate by gender and disability using `group_by()`.
3. Visualize trends with `ggplot2` (line graphs) and spatial disparities with `QGIS` integration.
4. Share findings via an interactive `shiny` dashboard for county health teams.

Python



Python is a versatile, open-source programming language renowned for its simplicity, readability, and broad applicability in data science, web development, automation, and more. Unlike R, Python is a general-purpose language, making it ideal for end-to-end workflows—from data ingestion to machine learning and deployment.

Key Features

- **Readability**: Clean syntax (e.g., minimal brackets) accelerates learning and collaboration.
- **Scalability**: Handles large datasets (e.g., KHIS records) efficiently with libraries like `Pandas` and `Dask`.
- **Machine Learning**: Libraries like `scikit-learn` and `TensorFlow` enable predictive modeling (e.g., disease outbreak forecasting).
- **Integration**: Seamlessly connects to APIs, databases (e.g., PostgreSQL), and tools like QGIS or DHIS2.

Getting Started with Python

1. **Install**: Download Python from python.org and use IDEs like Jupyter Notebook or VS Code.
2. **Learn Basics**:
   - Import/clean CSV files with `Pandas` (e.g., KHIS data).
   - Calculate health metrics (e.g., vaccination coverage) using `NumPy`.
   - Visualize trends with `Matplotlib` or `Seaborn` (e.g., malaria incidence by ward).
3. **Key Libraries for Health Data**:
   - `Pandas`: Data manipulation (filtering, merging datasets).
   - `Statsmodels`: Statistical analysis (regression, hypothesis testing).
   - `Scikit-learn`: Predictive modeling (e.g., forecasting hospital admissions).
   - `Requests`: Pull data from DHIS2/KHIS APIs.

Additional analysis/visualization tools

ArcGIS

2. Data Visualization Techniques

What is Data Visualization?

The graphical representation of information and data. It involves creating visual elements like charts, graphs, maps, and other visual formats to present complex data sets or information in a more accessible and understandable manner.

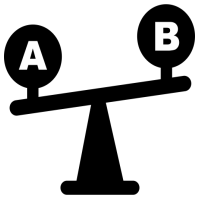
The primary goal of data visualization is to communicate information effectively, enabling users to comprehend patterns, trends, and insights hidden within the data.

It allows for the exploration and interpretation of data, facilitating better decision-making and understanding of complex relationships within the information.

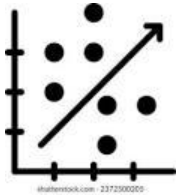
Why is visualization a must have?

1. **Simplifying Complex Information**: Data visualization helps in simplifying complex datasets, making information more accessible and understandable to a wider audience.
2. **Facilitating Decision-Making**: Visualizations aid in decision-making processes by presenting data trends, patterns, and correlations in a clear and concise manner.
3. **Increasing Engagement and Understanding**: Visuals have the potential to engage users more effectively than raw data.
4. **Highlighting Key Trends and Patterns**: Visual representations can uncover hidden trends and patterns within datasets that might not be immediately apparent in raw data. They help in identifying correlations, outliers, and important relationships within the data.
5. **Supporting Storytelling**: Data visualization aids storytelling by presenting data in a narrative format.

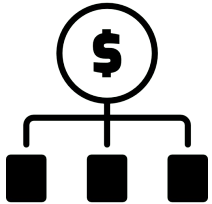
Types of Visualization



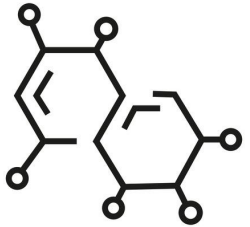
Comparison Visualization



Correlation Visualization

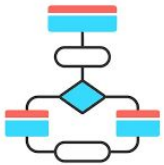


Distribution Visualization

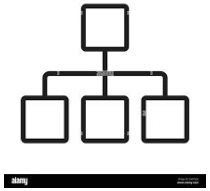


Composition Visualization

Geographic Visualization



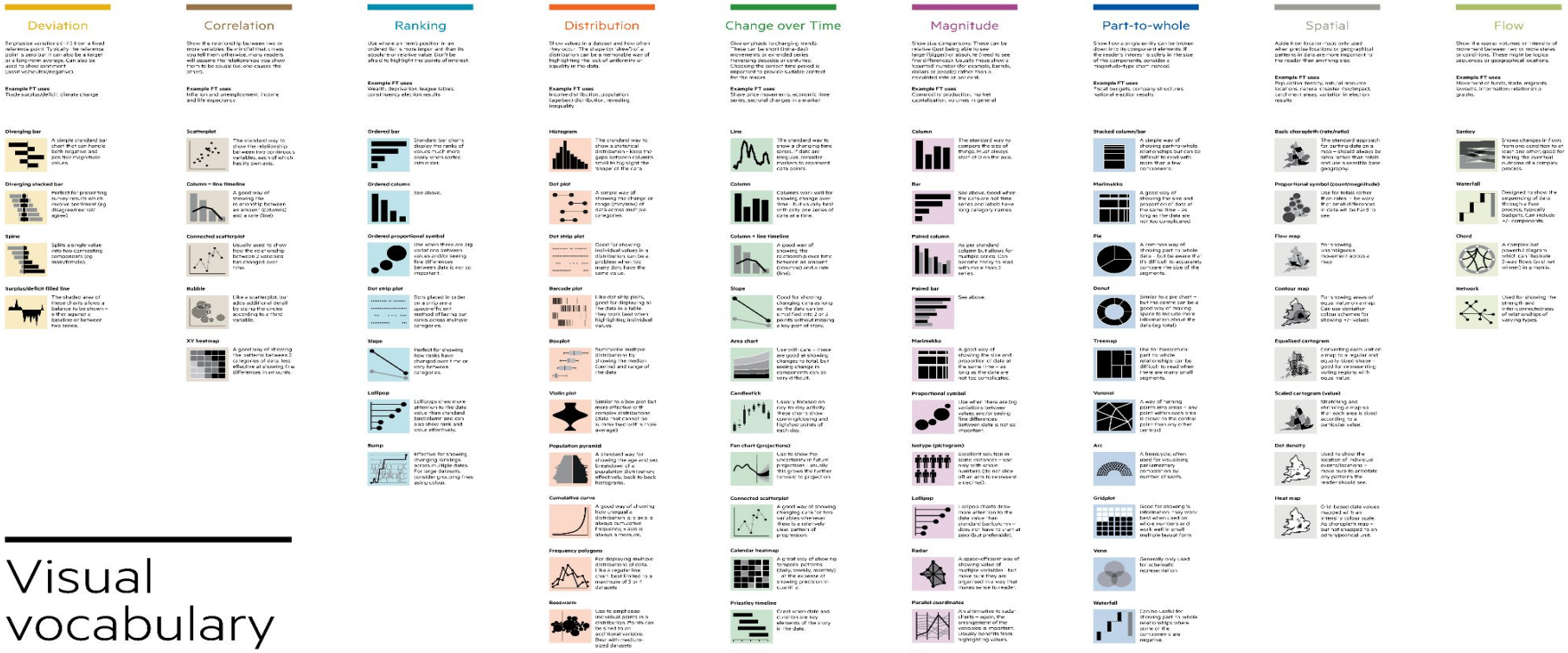
Flow Visualization



Hierarchical Visualization

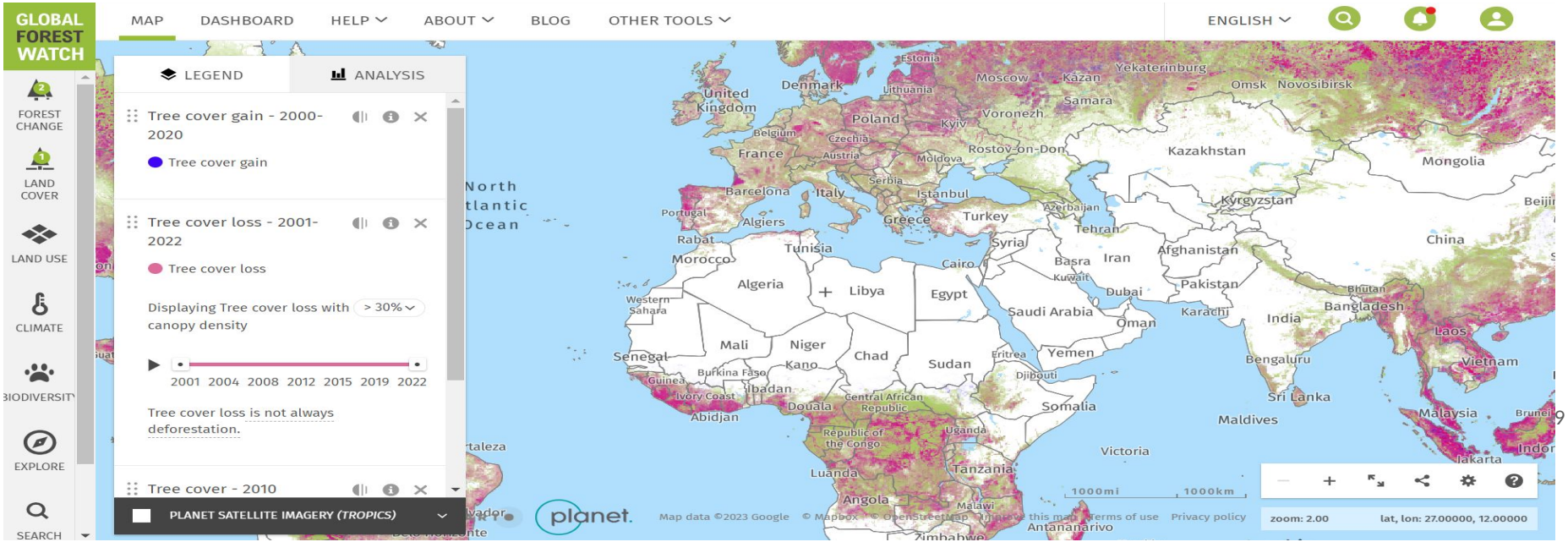
Creative, Innovation Visualization

Visualize with Intention



https://github.com/Financial-Times/chart-doctor/blob/main/visual-vocabulary

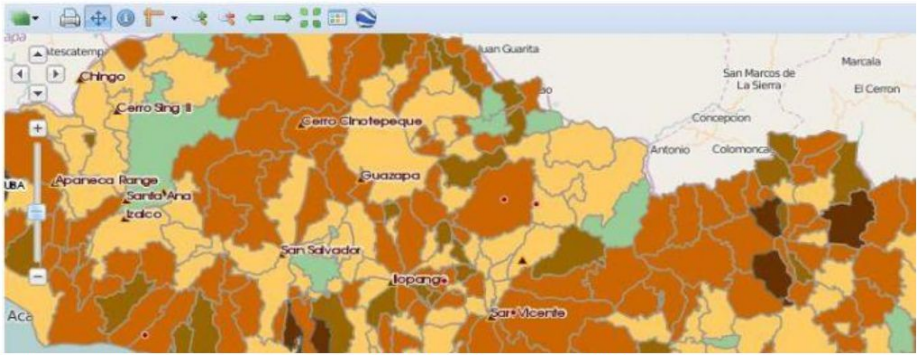
Spatial example: Global Forest Watch



https://www.globalforestwatch.org/map/

Spatial example - Geo Node: UN World Food Programme

WFP GeoNode



Data provided by:	World Food Programme (WFP)
Data accessibility:	export data, export map, visualization of data (e.g. web GIS or real time monitoring)
Link to the data:	GIS Layers GIS Maps
File type:	kml1, kml, wms
Data type:	baseline data, elevation, hazard specific data
Hazard:	Drought, Flood, Severe Storm, Extreme Temperature

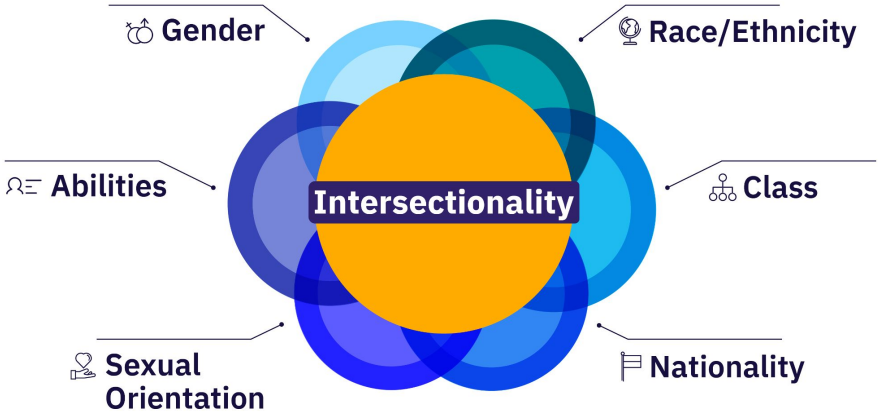
https://www.un-spider.org/links-and-resources/data-sources/wfp-geonode

Let’s Pause

In what ways do you currently visualize your data and what has been the impact?

3. Visualizing intersectionality in data

How do we visualize intersectionality?



How to visualize?

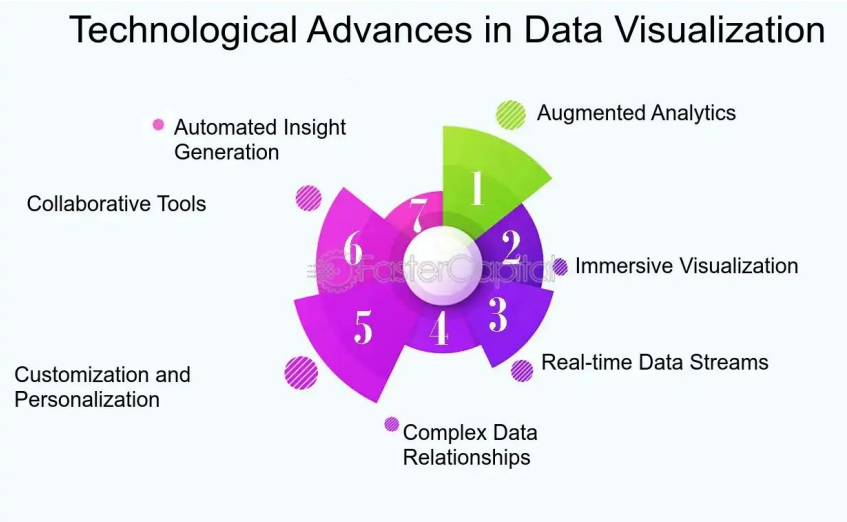
- **Heatmaps**: Show overlapping disparities (e.g., maternal mortality rates across gender × income × disability).
- **Sankey Diagrams**: Map flows between intersecting identities (e.g., education access for women with disabilities vs. non-disabled men).
- **Stacked/Grouped Bar Charts**: Compare outcomes across layered categories (e.g., HIV prevalence by gender + age + location).
- **Intersectional Scatterplots**: Plot two variables (e.g., healthcare access vs. poverty) with points colored by gender/identity.
- **Interactive Dashboards**: Allow users to filter by multiple dimensions (e.g., gender, disability, ethnicity) to explore compounded inequities.

Some Resources

1. Gender-ICT. (2024, March). *Seminar on Relief Maps: New visual methods for intersectional data collection and analysis on discriminations*. Gender-ICT. https://gender-ict.net/2024/03/seminar-on-relief-maps-new-visual-methods-for-intersectional-data-collection-and-analysis-on-discriminations/
2. Pollicy. (2024, May 15). *Harnessing the power of feminist data visualization for gender equality*. Medium. https://pollicy.medium.com/harnessing-the-power-of-feminist-data-visualization-for-gender-equality-dd5d696c8dd7
3. Relief Maps. (n.d.). *Relief Maps: New visual methods for intersectional data collection and analysis on discriminations*. Relief Maps. https://reliefmaps.cat/en/
4. data.org. (n.d.). *Recommended approaches to data analysis and visualization on gender and climate*. data.org Playbooks. https://data.org/playbooks/gender-data-and-climate/c/recommended-approaches-to-data-analysis-and-visualization-on-gender-and-climate/

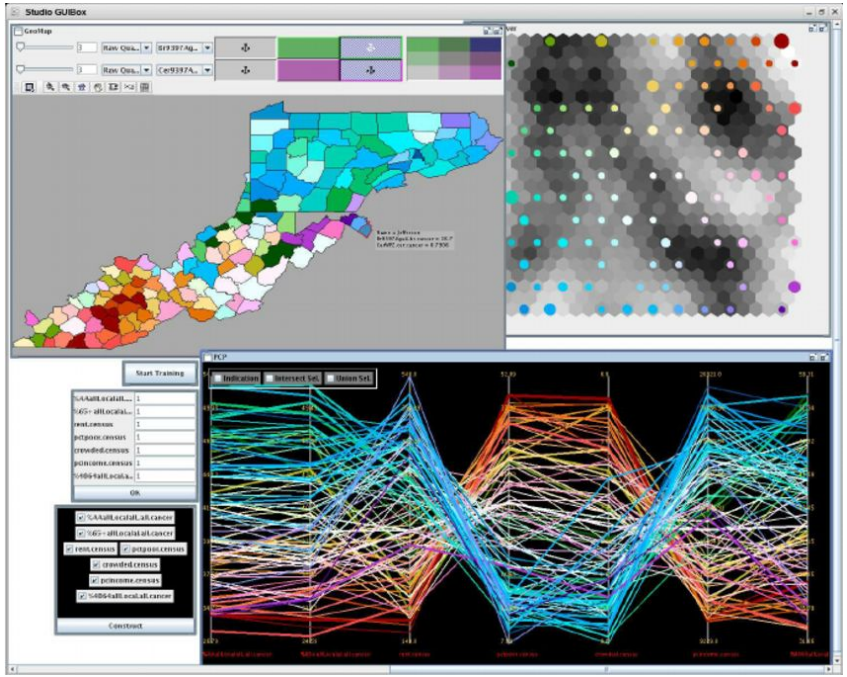
4. Integrating visualization for Socio-Economic Data

Introduction



Visualizing socio-economic data transforms complex datasets into intuitive, actionable insights, enabling stakeholders to identify patterns, advocate for equity, and drive evidence-based policies.

Why Visualize Socio-Economic Data?



- **Uncover Hidden Patterns**: Reveal correlations (e.g., poverty vs. school dropout rates) that raw numbers alone cannot convey.
- **Advocate for Equity**: Highlight disparities (e.g., gender gaps in employment) to prioritize marginalized groups.
- **Engage Stakeholders**: Simplify data for policymakers, citizens, and NGOs through accessible formats (e.g., interactive dashboards).
- **Monitor Progress**: Track SDG indicators (e.g., poverty reduction) spatially and temporally.

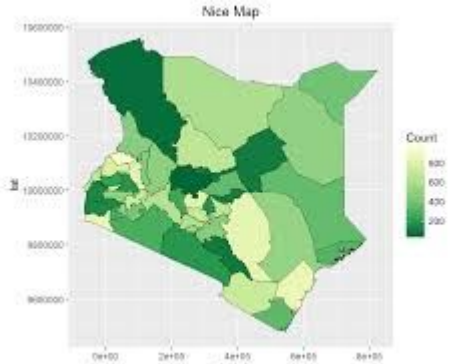
Visualization techniques for socioeconomic data

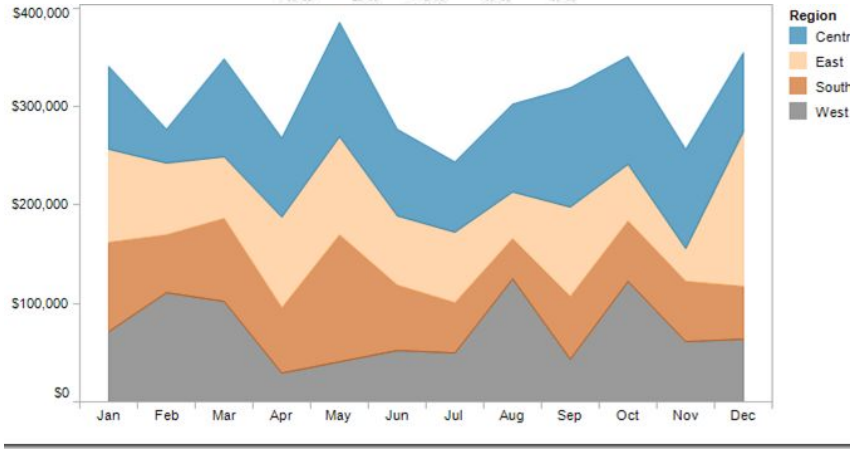
Spatial Mapping

- **Choropleth Maps**: Color-code regions by socio-economic metrics (e.g., income levels, literacy rates). *Example*: Map Kilifi's sub-counties by poverty index to target welfare programs.
- **Heatmaps**: Overlay density of variables (e.g., unemployment clusters near urban centers).
- **Point Maps**: Plot infrastructure (schools, clinics) against socio-economic indicators (e.g., maternal education levels).

Temporal Trends

- **Line/Area Charts**: Track changes over time (e.g., GDP growth vs. healthcare spending).
- **Animated Maps**: Show shifts in poverty or migration patterns across decades.





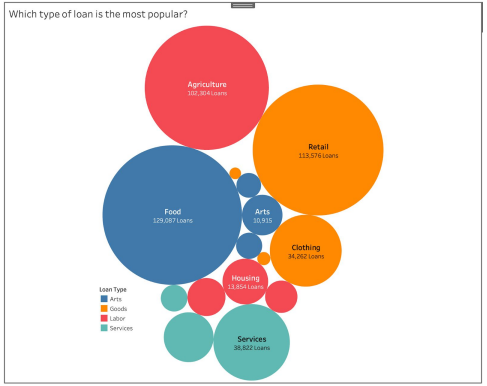
Visualization techniques for socioeconomic data

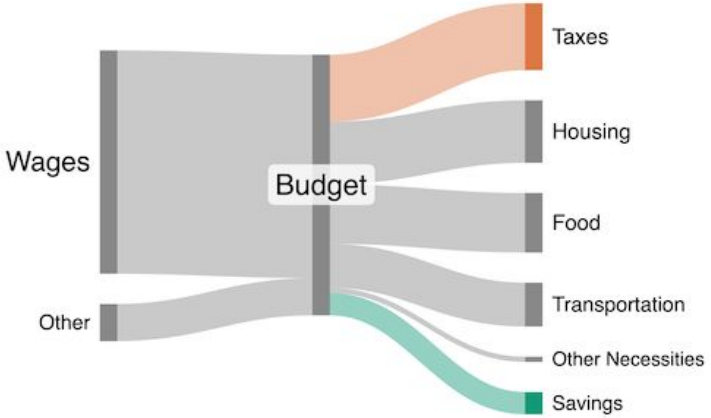
Multivariate Analysis

- **Bubble Charts**: Compare three variables (e.g., population size, income, and access to clean water).
- **Parallel Coordinates**: Visualize interactions between multiple socio-economic factors (e.g., education, employment, and gender).

Intersectional Visuals

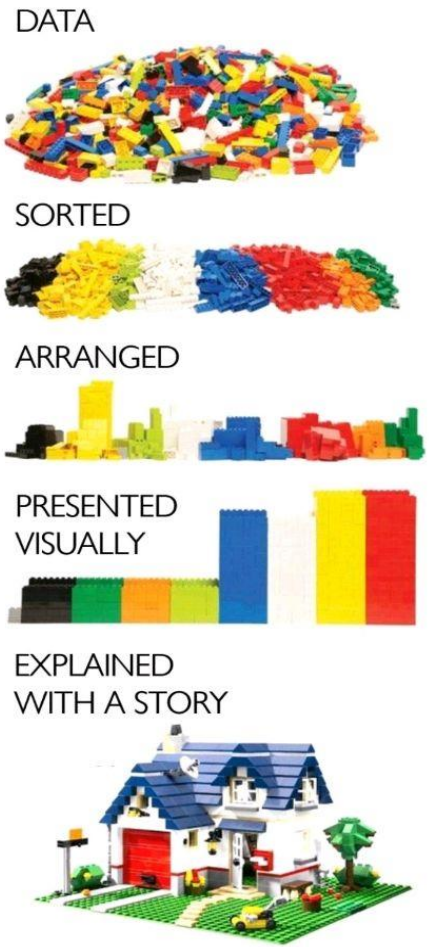
- **Sankey Diagrams**: Trace flows (e.g., education-to-employment pathways for women vs. men).
- **Stacked Bar Charts**: Disaggregate data by gender, disability, and income (e.g., access to microloans).





5. Power of effective communication

Which of these data sets do you currently have in your systems?



Importance of data communication

To provide evidence: presenting data as evidence provides a basis to support decision making

Making a case with data: data provides deeper insight into an issue allowing for more meaningful engagement

To inspire action: communicating data effectively can result in effective change

How to effectively communicate

What is your objective?

Who is your audience?

What is your key message?

What channels will you use?

Know your audience

- Who is the primary or secondary audience? What is the audience demographic—i.e., their age, gender, cultural background, educational level, and beliefs?
- What are their literacy levels and how much do they already know about the topic or message?
- What Does Your Audience WANT and NEED To Hear.
- What Do you REALLY WANT to say and WHY does it matter?
- What do you want from them?

Let’s Pause

How will you frame your message to make it most effective?

- What keywords matter to your audience?
- What aspects of the data are most crucial to highlight the communication objective?

Example

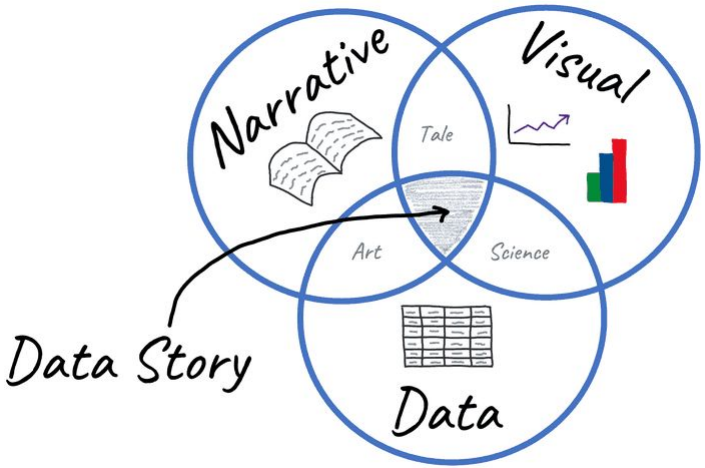
Audience	Minister
Type of audience	Decision Maker
Concerns & interests	Ministry performance
Desired outcome	Clear demonstration of performance over the period
Key Message	Improved revenue collection

6. Effective storytelling

Effective Storytelling

Developing a central theme and story structure

1. Identify trends in the data
2. Use rankings: Rankings tell a story using data about the relationship between items on a list.
3. Draw comparison: tell a side-by-side story
4. Surprise! Counterintuitive data (e.g. sales increased greatly over COVID period)
5. Relationships between data points (rise in teenage pregnancies attributed to school closures due to COVID)



- Start with what's important. Presenting data should be like a newspaper article—you know what happens by the time you finish reading the headline.


SATURDAY NATION

April 11, 2020 KSh60/00 (TSh1,700/00 : US$2,700/00 : Rf=900/00) No. 20005 /dailynation www.nation.co.ke

Xenophobia > Stranded student says she can't leave her house or put her lights on

Kenyans in China: Rescue us from hell

We gave them our masks, and now Chinese radicals have turned against our compatriots

Laws: Why your five-seater car can now only carry one passenger P.4

Plus: UN seeks Sh28 billion in emergency aid to Kenya P.6

Page 6

- **Be able to answer How and Why**.
  - How was the information gathered?
  - How or by whom were the statistics and calculations run?
  - Why is this important?

Let's Pause

Where is your audience? Where are they most likely to be engaged?

Example

Audience	County Governor
Type of audience	Decision Maker
Concerns & interests	County performance
Desired outcome	Clear demonstration of performance over the period
Key Message	Improved revenue collection
Channel	**????**

Key Takeaways from Module 5

- **Effective Data Communication is Audience-Centric**
  Tailor data narratives to stakeholders' needs (e.g., ministers vs. county governors). Focus on clarity, relevance, and actionable insights to drive engagement and policy action.
- **Leverage Diverse Tools for Robust Analysis**
  Use tools like **Excel** (basic analysis), **R** (advanced statistics/visualization), **Python** (automation/ML), and **QGIS** (spatial mapping) to clean, analyze, and visualize complex datasets
- **Visualization Simplifies Complexity**
  Transform raw data into intuitive visuals (e.g., choropleth maps for poverty indices, Sankey diagrams for intersectional disparities) to uncover trends, advocate for equity, and monitor SDG progress
- **Storytelling Amplifies Impact**
  Structure data-driven narratives using rankings, comparisons, and counterintuitive insights. Start with key messages (like headlines) to highlight urgency and answer *how* and *why* questions for stakeholders.