Visualizing ecological concepts in the context of the Serengeti ecosystem

A data visualization project

Author

Shobhit Kenath
Interaction Design - 206330004

Supervisor

Prof. Venkatesh Rajamanickam

IDC School of Design
अभिकल्प विद्यालय

0 Glossary of terms

0.1 Ecological terms

Ecosystem	An ecosystem is a complex network of species living in a given physical environment, under the influence of factors such as climate, habitats, and water relations.
Regulation	This principle implies that the population size of a species in an ecosystem is always kept under check by something else. Regulation prevents the population count from exceeding a certain limit.
Direction of Regulation	The direction (Top-Down or Bottom-Up) of regulation for a given species.
Bottom-Up Regulation	A subprinciple that implies the regulation of the population of a given species because of food availability.
Top-Down Regulation	A subprinciple that implies the regulation of the population of a given species because of predation.
Density Dependence	The dependence of the change in population of a species on its density.
Carrying Capacity	The maximum limit for the population of a species which can be sustained by the available resources in the physical environment.
Equilibrium	A situation where the number of births in a given population equals the number of deaths, thus keeping the population in a state of constancy.

Trophic levels	The various levels in the hierarchy of a given ecosystem, based on directionality of consumption.
Trophic interactions	An interaction which takes place between two trophic levels in an ecosystem.
Migration	A seasonal cyclical movement pattern which can be seen in some animals.
Ungulates	A classification of hoofed animals that used the tips of their hooves to support their body weight. Some examples are wildebeest, zebra, Thomson's gazelle, hippopotamus, buffalo, giraffe, and rhino.
The Keystone Effect	An series of effects created by the presence of one species, which affects the population of other species in the community.
Keystone Species	A species that has a disproportionately large effect on the ecosystem, relative to its abundance.
Biodiversity	The variety of living species in a given ecosystem. The greater the number of interdependent species within a community/ecosystem, the more the biodiversity.
Stability	The constancy in the population size or the number of species in an ecosystem.
Resilience	The ability of an ecosystem to buffer disturbances from the environment.
Habitat	A part of the physical environment that basically includes all the resources available to a species, that can be used in favor of their survival and reproduction.
Disturbance	An event or force that causes changes in the

	environment and in turn the affected ecosystem. Common examples are El Nino events, volcanic eruptions, tornadoes, or glacial advances.
State of an Ecosystem	The configuration of species within a given ecosystem, at a given point in time.

Contents

0 GLOSSARY OF TERMS 1

0.1 Ecological terms 4

1. ABSTRACT 8

2. LITERATURE REVIEW 8

2.1 A Place Like No Other 8

2.1.1 Abiotic components 8

2.1.2 Biotic components 10

2.1.3 Underlying rules of the Serengeti ecosystem 11

2.2 Visualization Analysis and Design 13

2.2.1 Data Abstraction 13

2.2.2 Task Abstraction 16

2.2.3 How idioms can be designed 17

2.2.4 The What-Why-How Analysis 18

2.2.5 The nested model: The 4 levels of validation 18

2.3 Storytelling with data 19

2.3.1 Analysis of information 19

2.3.2 Key lessons from the book 19

3. SCOPING 24

4. DESIGN OF NARRATIVE 24

5. DESIGN EXPLORATIONS 24

5.1 Data 24

5.1.1 Population 24

5.1.2 Maps 27

5.1.3 Climate 28

5.1.4 Species and trophic interactions 29

5.2 Concepts 31

5.2.1 Generic species 31

5.2.2 Migration 31

5.2.3 Keystone effects 33

5.2.4 Regulation 33

6. THE PROTOTYPE 34

6.1 Summary of narrative 34

6.1.1 Context 34

6.1.2 Rules that govern ecosystems 34

6.1.3 An interesting community of animals 34

6.1.4 Conclusion 35

6.2 Section 1 - Contextual Information 35

6.3 Section 2 - The diversity of species 36

6.4 Section 3 - Rules that govern ecosystems 36

6.5 Section 4 - An interesting community of animals 37

7. PLAN FOR EVALUATION 38

7.1 Evaluation with end users 38

7.1.1 Detailed evaluation plan 38

7.2 Findings from the evaluation 41

7.2.1 Limitations 41

8. FUTURE SCOPE 41

9. REFLECTIONS & THOUGHTS 42

10. ANNEXURE 43

10.1 Feedback from Stage 3 43

10.1.1 Justification of design decisions 43

10.1.2 Correlation VS causation 43

10.1.3 Increase in complexity 43

10.1.4 Integration of seemingly disconnected windows 43
10.1.5 Task-based comparative evaluations 43

10.2 Plan for evaluation from an ecology perspective 52

10.3 Detailed findings from evaluation with users 44
10.3.1 Findings from user#1 44
10.3.2 Findings from user#2 48

BIBLIOGRAPHY 53

1. Abstract

The aim of this project is to explain ecological concepts to a general audience, using visualizations. The project spanned over a period of five (from January to May 2022) months. Literature associated with the domain of ecology and data visualization was used to guide design decisions. The scope & complexity of the content, narrative and visualizations were designed and iterated upon in the second half of the project. The final outcome was an interactive digital prototype hosted on the internet. It was evaluated with end users, and the findings list out the current limitations. Finally, opportunities for redesign have been mentioned.

Please use this link (https://shobhit17s.github.io/) to access the prototype.

Note: The prototype is meant to be viewed on a laptop/desktop, using Google Chrome or Microsoft Edge. It is not compatible with Safari.

---

2. Literature review

This project combines ecology with data visualization. Literature from both the domains has been detailed out in this section.

2.1 A Place Like No Other

The original direction of this project sprung from the intent of explaining the phenomenon of **the Great Wildebeest Migration** to a general audience. A video on YouTube by the name "*Serengeti: Nature's Living Laboratory | HHMI BioInteractive Video*" provided a basic understanding of this phenomenon. However, the book (written by Anthony R E Sinclair) "*A Place Like No Other*" was referred to for understanding it in greater detail.

In his book, Sinclair explains how the entire Serengeti ecosystem works, and how the Great Migration plays a pivotal role in its maintenance. There are **abiotic and biotic components, principles, and subprinciples that interact with each other, causing the ecosystem to exist in a given state**. However, the state of the ecosystem can also switch. An undesirable outcome of this switch is something called an irreversible state. Hence, it is important to understand the inner workings of the ecosystem in order to make mindful human interventions.

2.1.1 Abiotic components

Before understanding the rules of the Serengeti ecosystem, it is imperative to conceptually map the components that constitute it. The diagram below shows a schematic map of the ecosystem, and can be used to map the following abiotic components:

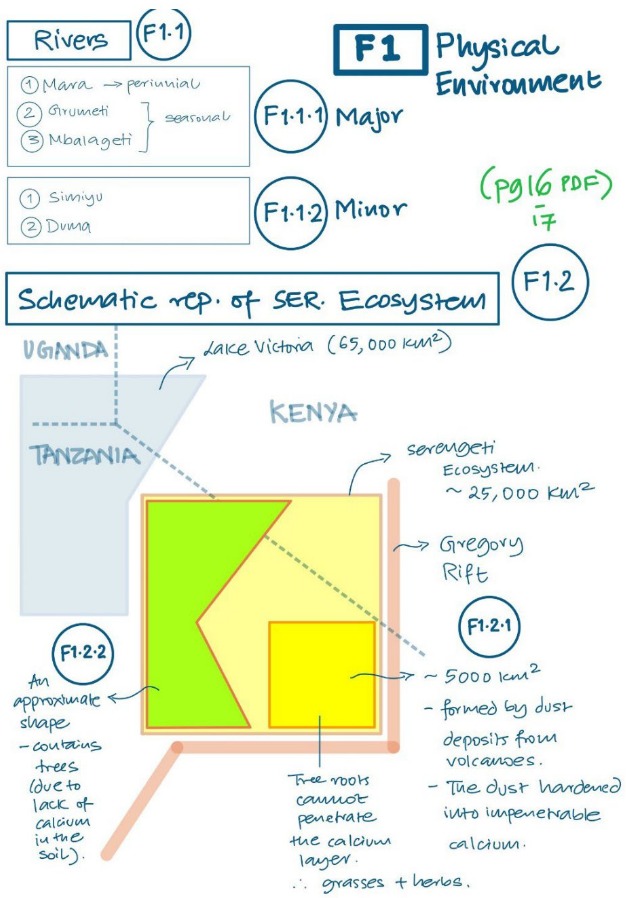


Fig. 1 – Notes on physical environment

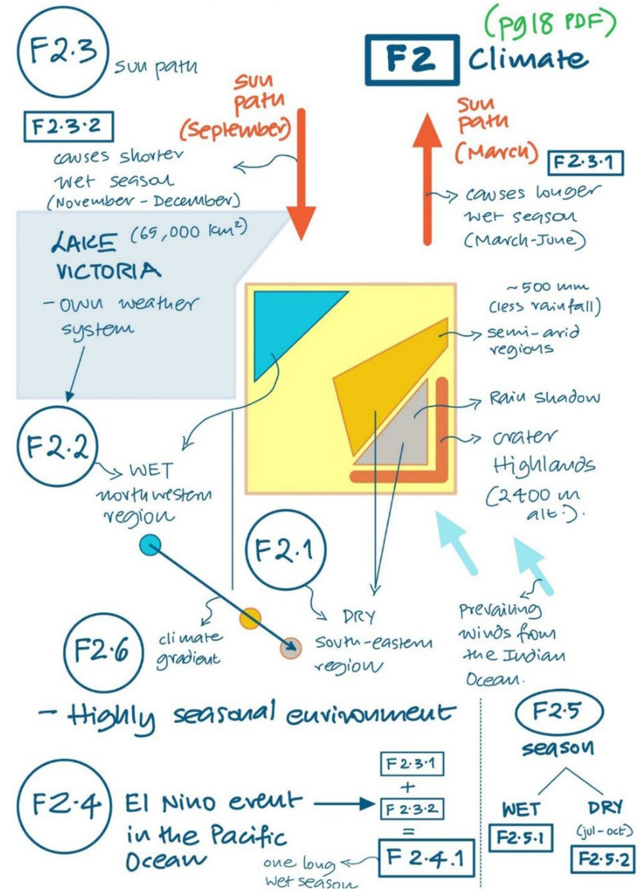


Fig. 2 – Notes on climate

Physical Environment

The soil distribution determines growth of vegetation in the ecosystem. There are three major rivers (Mara, Grumeti, and the Mbalageti) that sustain wildlife. The neighboring lake Victoria is massive enough to be able to create a climatic environment of its own.

Climate

The typical climate of Serengeti varies between two seasons, namely wet and dry. The dry season usually lasts from July to October. The lake causes a climatic gradient to exist over the ecosystem, ranging from a wet north-western region to a dry south-eastern region. On a global scale, an El Nino event in the Pacific Ocean causes changes in distribution of seasons across the year. This in turn impacts the vegetation.

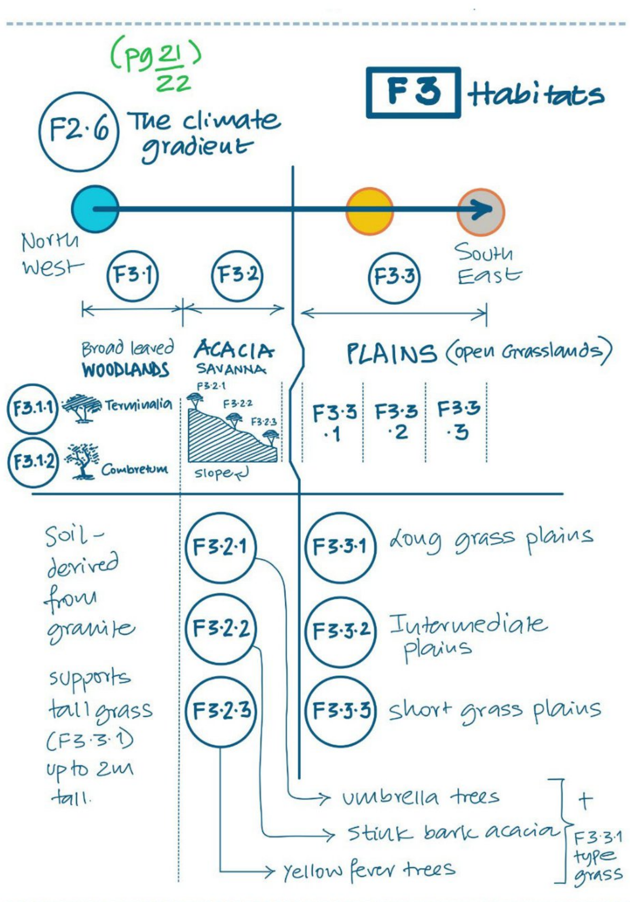


Fig. 3 – Notes on habitats

Habitats

Habitats indicate the preference a species has over the physical environment. In Serengeti, they starkly vary along the climatic gradient. As Sinclair mentions in his book, there are three major types of habitats, excluding the 'special' habitats.

Major habitats

1. Plains
    a. Long Grass Plains
    b. Intermediate Plains
    c. Short Grass Plains
2. Acacia savanna
    a. Umbrella Trees
    b. Stink Bark Acacia
    c. Yellow Fever Trees
3. Broad-leaved woodlands
    a. Terminalia
    b. Combretum

Special habitats

1. Floodplain (western end)
2. Hills forming a plateau (center)
3. Montane forest (northeast)
4. Riverine forest (northwestern end)
5. Kopjes

2.1.2 Biotic components

Serengeti has a very high biodiversity, meaning that there are several species that can be studied. However, not all of them lie within the scope of this project. Since the idea is to understand the effect a keystone species has on an ecosystem, the prime focus shall remain on the wildebeest and associated species.

Species

The following species have been considered for this project. The filtration is based on their relevance to the narrative.

No.	Trophic level	Species
1	Predators	Lion
2		Hyena
3	Meso-predators	Leopard
4		Cheetah
5	Mega-herbivores	African buffalo

6		Giraffe
7		Rhino
8		Elephant
9	Herbivores / Consumers	Topi
10		Kongoni
11		Wildebeest
12		Zebra
13		Oribi
14		Thomson's gazelle
15		Impala
16		Eland
17	Producers	Short grass
18		Medium grass
19		Long grass
20		Shrub leaves
21		Tree leaves

Table 1 – The initial list of species

Trophic levels and interactions

In such a hierarchical structure, the species in the upper levels prey on the species in the lower levels. Hence, the species at the top becomes the apex predator and the one at the bottom becomes the primary producers. These levels are referred to as *trophic levels*. The interaction that takes place between two trophic levels is called a *trophic interaction*.

Once the apex predator is removed from the ecosystem, the effects of the change percolate through all the levels underneath it. This phenomenon is called a *trophic cascade*. The cascading effect can be traced along the trophic interactions between species.

2.1.3 Underlying rules of the Serengeti ecosystem

Sinclair explains the fundamental principle of regulation, and the various sub-principles associated with it. These 'rules' together determine how the components of the ecosystem change over time.

The fundamental principle - REGULATION

This rule states that the population of every species is 'regulated' by some other biotic component of the ecosystem. For example, the oribi population is kept in check by the lions. However, the elephant population is regulated by the availability of tree leaves.

Regulation is dependent on size of the individual, diet, competition, migration, and accessibility to food (dependent on physical environment and climate).

Bottom-Up regulation – Food regulation

It dictates that the population of a species is maintained at an equilibrium state because of the availability of food per individual among that population. Hence, most deaths within a bottom-up regulated population are due to starvation.

A concept that is related to the subprinciple of Bottom-Up regulation is that of **competition**. The food available within a given ecosystem is limited. Hence the many species are often forced into competition for the available food both within the same species and with other species.

However, it also has to be kept in mind that the environment containing that food is not necessarily uniformly accessible. There is an inverse relation between the threat level (usually predators) and **accessibility of**

food in a given area. Quite often there are areas in the wilderness that have a higher threat level and hence less accessibility to food. Usually, herbivores venture into areas with higher threats to life during the dry season, when availability of resources is low. Thus, the physical environment and climate add an additional layer of complexity to understanding regulation.

Top-Down regulation – Predation

The rule implies that the population of a species is maintained at equilibrium because of predators at a higher trophic level. Hence the regulation happens from above. For example, the population of smaller antelopes (topi, kongoni, impala and oribi) in the Serengeti stay in equilibrium because of predation by lions, hyenas, leopards, and cheetahs.

It is important to note that Top-Down regulation can also be seen in other trophic levels of the ecosystem. For example, Elephants regulate the population of Stink Bark and Yellow Fever Acacias from a higher trophic level. Hence, it isn't necessary that a Top-Down regulation interaction occurs only between a carnivore and a herbivore.

Top-Down regulation can have a **regulatory** (stabilizing) as well as an **anti-regulatory** (destabilizing) effect. Left unchecked, the anti-regulatory effect of predation **can lead to the extinction** of a lower trophic level species. At times Top-Down regulation can also regulate the population of smaller predators.

Migration

This movement pattern results in a seasonal round trip. Animals who follow this pattern of movement are known as **migrant species**. Among the species (fauna) listed above, Thomson's gazelle, zebra, wildebeest, eland, and elephants are migrants. **Migration directly affects the fundamental principle of regulation**. For example, the wildebeest have the advantage of counteracting both Bottom-Up and Top-Down regulation due to migration. Hence, migrant species tend to have (depending on their reproductive cycles as well) large populations.

Animals who do not migrate are known as **resident species**. Among the species (fauna) listed above, the predators, the meso-predators, and herbivores such as African buffalo, giraffe, rhino, topi, kongoni, oribi, and impala are residents. Although the total wildebeest population also has a resident subset, it will not be considered for the project.

It is important to note that migration is directly influenced by the climate.
In Serengeti, migration leads to the following consequences for the ecosystem:

1. A difference in fertilization of soils across the rainfall gradient.
2. Stability in the plant community, due to prevention of overgrazing.
3. Stability in the various other communities (insect, bird, and predator).
4. The keystone effect.

Disturbances

External factors such as fluctuations in the environment or disease cause changes in populations. These are termed as disturbances, and directly affect regulation.

The **El Nino event** is a phenomenon that takes place somewhere north of Australia over the Pacific Ocean. It is detected when there are positive or negative variations from an average value over the ocean surface. Both kinds of variation have different implications for the Serengeti ecosystem. In essence, the El Nino event affects two parameters:

1. Rainfall
2. Fires

A positive El Nino causes more rainfall in the dry season, thus leading to a higher survival rate of wildebeest yearlings. This in turn creates more food for the lions in the dry season as well. A negative El Nino can cause more rains in the shorter wet season or the longer wet season. These pathways also cause changes in the ecosystem, sometimes even resulting in fires during the dry season. The fires in turn have a direct effect on the tree population. Below is a diagram that explains the effect of positive and negative El Nino events on the rainfall and fire regimes in the Serengeti ecosystem. The effects of other parameters such as the habitats, human impact and grazing have also been shown.

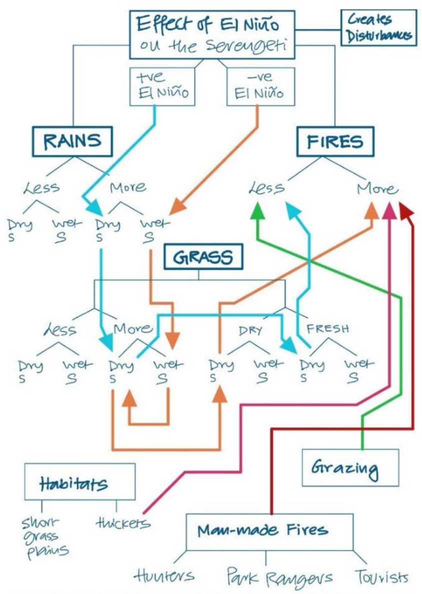


Fig. 4 – Effect of El Nino event on climate

It is important to acknowledge the cascading impact of such phenomena in order to be more mindful of future actions.

2.2 Visualization Analysis and Design

In her book Visualization Analysis and Design, Tamara Munzner very succinctly describes the different types of data and datasets, why vis tools are designed (purpose or objectives), different possibilities in the space of vis idiom designs and also how to analyze a given vis design/tool. She clearly explains the scenarios in which having a human and a computer in the Vis analysis process is beneficial. The task that needs to be performed by a user of the Vis is always a constraint for the designer, and hence needs to be clearly articulated before commencing with the design process. Following is a quick summary of some of the important aspects of Vis design that have been mentioned in the book.

2.2.1 Data Abstraction

This is the process of examining the available dataset and breaking it down to simpler terminology. It provides much needed clarity during the idiom design phase.

Data can either be discrete or continuous in nature. However, there are essentially five types of data. They are as follows:

Items

Items are individual entities that are discrete in nature and typically represent a single entry. For example, each row in a simple table is an item. A node in a network is also an item. Each item can have several attributes.

Attributes

Attributes are specific properties of something. They can be measured, observed, or logged. Attributes can have types such as:

Categorical

Categorical attributes have no order among the entries. They simply fall into a given category. For example, bananas, oranges and apples fall into the category of fruits, but do not have any order among them.

Ordered

As the name suggests, ordered attributes always have some order among the entries. They can be quantitative or ordinal (not necessarily quantitative).

Ordinal

Ordinal ordered attributes are not quantitative. However, they have an intrinsic order among the entries. For example, S, M, L & XL have a clear order in terms of shirt sizes. However, they cannot be arithmetically operated upon.

Quantitative

It is possible to perform arithmetic operations on quantitative ordered attributes. Hence, such attributes usually contain integers or real numbers.

Ordering direction

This concept comes to play while considering the semantics of the ordered data. Hence, it only applies to ordered data.

Sequential

Sequentially ordered attributes have a minimum and a maximum value. The values in between are sequentially arranged.

Diverging

Divergently ordered attributes have a neutral point (or a zero mark). Values to the either side of this mark grow sequentially away from it.

Cyclic

Cyclically ordered attributes wrap around in a loop and are repetitive. For example, days of the week or months of the year. Usually, temporal data falls into this category.

Links

This datatype simply depicts the relationship between items. These are typically found in a network diagram.

Positions

Positions are spatial data, providing a location in a 2D place or 3D space.

Grids

Grids are imaginary points or instances at which samples are drawn or measured. These are more of a ‘sampling strategy’ for continuous data.

Dataset types

The above-mentioned types of data are typically arranged into the following datasets. Sometimes they can be arranged into complex combinations as well.

Tables

These are the most common dataset types. A typical spreadsheet is a table. *Rows* are used to encode *items* and *columns* are used to encode *attributes*. *Cells* that are at the intersection of these data types contain *values*.

Networks

This dataset type is efficient for encoding items and the relationships between them. Each *node* is used to encode an *item* and the *links* are used to encode *relationships*.

Fields

Fields are used to depict *continuous phenomena*. As Tamara articulates it, "this dataset type contains *attribute values* associated with *cells*. Each cell contains measurements or calculations from a *continuous domain*".

Spatial fields

These are very effective in depicting continuous phenomena that *can be regularly sampled in space* in order to attain the values at those intervals. These can be used to *encode attributes into a spatial layout*.

Geometry

This is used to depict *purely geometric data* such as the *boundaries* of a given area on a map. Naturally, *positions* are encoded in such a dataset type. It is important to note that *design decisions are used to create geometric visualizations*. Otherwise, it would simply refer to the generation of computer graphics.

Collections – Clusters, sets & lists

Sets are an *unordered* group of items. Whereas lists are an *ordered* group of items. Clusters are grouped based on the *similarity of attributes*. Paths are an *ordered set of segments* formed through links.

Cardinality of data

This process is aimed at examining the extent of variation of values falling under a given attribute. It helps in getting a sense of the scale of the visualization.

Dataset availability

While thinking of visualizations, it is also important to consider whether the data stays the same (*static*) or it is changing constantly (*dynamic*).

The semantics of data

Data always has an associated meaning, which must be correctly ascertained. For example, a set of numbers could mean anything ranging from positional data to values measured on some continuous scale. It is imperative to ask the question, "what does this data mean?". Broadly, the semantics can be classified into two categories:

Key vs Value semantics

A key is a unique identification meant for looking up an item in a dataset. In a table, typically the keys might be serial numbers or ID numbers. Keys are usually ordinal or categorical. Values are basically the entries at the intersection of the item (representing the key) and the concerned attributes.

Flat tables

Flat tables only have one key attribute but might have one or several value attributes.

Multi-dimensional tables

Multi-dimensional tables have more than one key attribute. A unique combination of keys is required to look up an item in the table.

Fields

Fields perhaps provide more 'useful info' than tables. The same keys + values concept can be applied to fields as well, where the keys are independent variables (IVs) and the values are dependent variables (DVs). In spatial fields, they keys denote spatial positions.

Fields can be either multivariate (based on the number of value attributes) or multidimensional (based on the number of key attributes). Multivariate fields can be either scalar (one value attribute), vector (two value attributes) or tensor (more than three value attributes).

Temporal semantics

Usually time-related data can be deeply multiscale, ranging from nano seconds to millennia. Sometimes, the scales might not even cleanly fit into a strict hierarchy. For example, the number of weeks doesn't fit cleanly into every month.

Time varying data

Here, the timestamp is the key. Hence a variation in values can be seen with respect to the keys. A good example of this is location tracking data.

Time series dataset

A dataset containing time varying data (time-value pairs) is known as a time series dataset. The intervals can be uniform or non-uniform. Such datasets are typically used for tasks related to analysis.

2.2.2 Task Abstraction

This exercise revolves around the question, "why is a Vis tool being used?". Once a Vis designer has a sufficiently complex idea about this, it is possible to choose an appropriate Vis idiom design aimed at achieving task success. Depending on the task, Vis tools can be either specific (user cannot override design choices) or general (user has flexibility in choices). While abstracting tasks, it is important to consider the following components:

1. Actions
2. Targets
3. Who would use the tool?

Actions

The following diagram summarizes the concept of actions.

High-level (Analyze)

This is the most common action performed by a Vis user. Analysis can be categorized as follows:

Consume

A Vis user might want to either present known data, discover patterns or simply enjoy the visualization, using the tool.

Produce

In other cases, the user might want to either annotate, record or derive new data, using the visualization.

Mid-level (Search)

Searching is also a common action performed by a Vis user but lies at a lower level. Search can be categorized as follows:

Lookup

While looking-up something, the 'what' and the 'where' are known.

Locate

While locating something, the 'what' is known but the 'where' is unknown.

Browse

While browsing something, the 'where is known but the 'what is unknown.

Explore

While simply exploring, both the 'what' and the 'where' are unknown.

Low-level (Query)

This is the lowest level of action performed by a Vis user. A query is classified on the basis of search targets:

Identify

A user might want to simply identify a single target. This action returns the characteristics of the target.

Compare

A user might want to compare multiple targets with each other. Such an action requires more 'sophisticated' Vis idioms.

Summarize

A user might want to obtain a summary of all the data. Hence it includes all targets. This action results in an overview of the data.

Targets

A target is simply some aspect of the data that is of interest to the user. Targets can be classified as follows:

All data

Users are typically interested in the trends, outliers, or certain features of the data. These are the most common targets for a user.

Attributes

The user might be interested in knowing about just one particular attribute, specifically the distribution of values, the extremes or even the individual values.
In some cases, the user might be interested in knowing about many attributes, in order to understand their dependencies, correlations, and similarities.

Network Data

The user might be interested in knowing about the topology of a given network structure. At a lower level, they might be interested in knowing about the individual links/pathways.

Spatial data

With spatial data, the user is typically interested in understanding shapes.

Who would use the tool (user / designer)?

It is also important to understand who would be using the Vis tool. In most cases it is designed for a specific user demographic. However, in some cases, the designer themselves might want to be using it.

2.2.3 How idioms can be designed

Basic composition of an idiom

At a broad level, the design of any idiom would contain the following elements:

Marks

These represent the **items** or **links**, that have been extracted from a given dataset and visualized using the idiom.

In most cases, the underlying assumption is that there is *one mark per item* and that there is only a *single view* on display. However, there can be instances with multiple views, multiple marks in a region (known as glyphs) or some items not being represented by marks.

Channels

Channels refer to the **appearance** of the marks, based on their attributes. They provide a way to change this appearance, by **mapping the attribute values**.

The *expressiveness* (matching channel and mark characteristics) and *effectiveness* (accuracy of perception, discriminability, separability, and pop out) of channels must be considered while mapping these values.

Idiom design strategies

Depending on the design objective, various Vis idiom design strategies can be implemented. They can be broadly classified in the following manner:

Encode

This classification has to do with the spatial arrangement and the non-spatial mapping of data. They can be further classified as:

1. Arrangement of data (spatially)
    a. Express
    b. Separate
    c. Order
    d. Align
    e. Use
2. Mapping of data (using non-spatial channels)
    a. Color
    b. Size
    c. Angle
    d. Curvature
    e. Shape
    f. Motion

Manipulate

This classification deals with the manipulation of the data. However, this is usually achieved at the cost of time, legibility and cognitive load. The user can:

1. Change some aspect of the view
2. Select elements
3. Navigate to change the view

Facet

This classification enlists strategies for faceting the data, such as:

1. Juxtapose and coordinate multiple views
2. Partitioning the data between views
3. Superimposition of layers

Reduce

This classification enlists strategies for the reduction of the data. They are:

1. Filtration of data
2. Aggregation of many data elements
3. Embedding of data (contextual information within a single view)

2.2.4 The What-Why-How Analysis

This three-part analysis enables the Vis designer to use a ‘scaffolding’ or framework that asks three broad questions about data abstraction, task abstraction and idiom design (the what, why and how, respectively) for each *instance*, in order to create ‘effective’ visualizations. It is important to note that one instance can be linked to another as well, thus forming *chains*.

2.2.5 The nested model: The 4 levels of validation

Tamara lists out four abstract levels that have a hierarchical (and sequential) relationship, as part of the conceptual model while considering the aspect of validation. These levels are:

Domain situation

Understanding what the users in a given domain (computer science, finance, ecology etc.) might need from the Vis tool. Here the biggest threat is in the misidentification of these *needs*.

Abstraction

Abstraction must be considered only after correctly identifying the needs of the user. Here, the designer looks at the 'what' (data abstraction) and the 'why' (task abstraction) of the three-part analysis process. At this level, the threat lies in showing the 'wrong thing' (irrelevant data for unnecessary tasks) to the user.

Idiom

The idiom level lies underneath the abstraction level and must be only considered after the abstraction phase is complete. This part deals with the 'how' of the design process. Focus lies on the encoding (drawing) of data and any interaction (manipulation) techniques that might need to be implemented. The threat here, lies in the ineffectiveness of the design.

Algorithm

Finally, at the last level, the designer grapples with the construction of the tool. Thus, focus lies in the system time/memory used and the computational complexity of the tool. The threat to validity here, would imply that the code is too slow.

2.3 Storytelling with data

In her book, Cole Nussbaumer Knaflic lays out the concepts that can be leveraged in the design of effective visualizations. At the fundamental level, visualized information falls into either of the following distinctions:

1. Quantitative
2. Categorical

2.3.1 Analysis of information

Analysis is an inherent part of designing visualizations. It is important to note that there are two types of analyses:

Exploratory analysis

During an exploratory analysis session, the user starts off with a hypothesis. Their aim is, as Cole points out, "to dig through the data, trying to find something interesting that somebody else might care about". It is after having identified the 'interesting thing' that one moves into the explanatory space. Hence, it is almost always the case that explanatory analysis follows exploratory analysis.

Explanatory analysis

In this type of analysis, the designer of the visualization has 'something specific' to communicate to 'somebody specific'. In this stage, two of the key lessons discussed below become particularly important. They are '3.2.2.4 focusing the attention of the viewer' & '3.2.2.6 telling a story' using the data.

As can be guessed, the book 'Storytelling with Data' focuses on *explanatory analysis*.

2.3.2 Key lessons from the book

The need for effective communication using data isn't constrained to any single profession. Rather, it is a fundamental requirement while dealing with any type of data. In a nutshell, the book encapsulates six key lessons. They are:

Understanding the context

This step comes at the very beginning. The designer must form a very clear understanding of the *audience* (who they are) and the *content* (what they need to know) that needs to be conveyed.

It is also wise to consider the actions that the designer wants the audience to take, after conveying the content. Another important factor is the *mode of presentation*. Depending on whether the designer is present while the audience views the visualization, the *level of control* and the *level of detail* required vary. *Potential questions* that may need to be addressed must be considered based on this decision.

The *tone* of the communication must also be clear. It can vary from a celebratory one to one that spurs the audience to take correctional action.

Similar to Munzner's three-point analysis, Knaflic suggests a **who-what-how analysis**. The idea is to clearly establish the *audience demographic*, the *content specificity*, and the *choice of presentation* of the information.

Who – understanding the audience

It is important to know the following things about the user(s) of the visualization:

1. Basic background information (decision makers / colleagues / experts).
2. Information that is relevant to them.
3. Present biases.
4. Familiarity with data.

What – understanding the content

It is important to check for the following things, while considering the content:

1. Identification of relevant data (that can strengthen the case).
2. Summary of the idea.
    a. The 3-minute story (time-constrained version).
    b. The big idea (single sentence description).

How – understanding the design

It is important to check for the following things, while considering the design of the display:

1. Desired outcome (for structuring).
2. Storyboarding.

Choosing the appropriate visual display

In this section, Cole discusses various "commonly used displays" that are used to communicate analytics. She also shares several use cases and examples.

Simple text

This proves to be an effective strategy when there are only one or two major numbers that need to be communicated. Using a visual representation here, might undermine the impact of the text.

Tables

Tables are most effective when a set of data is being presented to a mixed audience, where different groups might be interested in different columns or rows (attributes or items). It is important to prevent the design of the table from decreasing the focus on the data. The data must 'take center stage'.

Heatmaps

Heatmaps are a special case of tables. They combine the data present in a table with visual cues that aid the user in quickly interpreting the data.

A change in saturation or brightness is often sufficient to create these visual cues.

Graphs

Graphs work well with the 'visual system' of the user. From a wide range of *graph types*, the ones most commonly used are:

Points

1. Scatterplots – used to encode data simultaneously on x and y axes. Works well with categorical data.

Lines

1. Standard line graph – works well with continuous data, with a temporal attribute.
2. Slope graph – effective when there are two points of comparison for values of a given set of attributes.

Bars

Bar charts can be used to leverage the user's familiarity. These are highly effective, if used correctly. It is important to note that the *baseline of a bar graph must always be at zero*. From an ethical standpoint, the manipulation of scale on a bar chart must be avoided.

1. **Vertical bar charts** – these can be single, two or multiple series. While using multiple series bar charts, attention should be paid towards the loss of focus on a single series as well as the relative order of the categorization.
2. **Stacked vertical bar charts** – This type of bar chart has limited applications and should be used carefully. The main drawback is that the components of two stacked bars become harder to compare due to an inconsistent bassline.
3. **Waterfall charts** – This type of bar chart can be used to either focus on individual components of a stacked vertical bar chart or to show a starting point, followed by increases, decreases and an ending point. The latter case has a temporal aspect as well.
4. **Horizontal bar charts** – These are considerably easier to read than a vertical bar chart. Horizontal bar charts can also be single, two or multiple series. The ordering of categories must aid the interpretation process of the user.
5. **Stacked horizontal bar charts** – These charts can be used to either show parts of an absolute value (bars have varying lengths) or the composition within a uniform whole (all bars have the same length, i.e., 100%).

Area

Area is comparatively harder for human beings to associate with quantitative information. That is the general rule. However, there are instances when an area-based depiction of information can be more effective than one-dimensional (bar charts / line graphs) depictions.

1. **Square area graphs** – these have the ability to encode information using the combination of both vertical and horizontal axes, in an aligned manner. Hence, in most cases, square area graphs are easier to read than other area-based depictions.
2. **Pie charts** – Amusingly enough, Cole has quite the strong opinion about pie charts. She feels "the only thing worse than a pie are two pies" that have to be compared. The primary reason for this bias is the inability of human perception to effectively determine the quantitative differences among pie slices.

   If the differences between the areas of individual slices are large, one can determine which slice is the biggest or bigger than another. However, it is near impossible to gauge the quantitative differences between the areas.

If these differences are small, it becomes extremely difficult to even distinguish a bigger slice from a smaller one.

3. **Donut charts** – This type of area chart is even more difficult to read than a pie chart since it uses an added visual attribute of arc length. Hence, they're *best avoided*.

Elimination of clutter

The essence of this stage lies in the *identification* and *elimination* of elements that don't add any information of significance to the visualization. Clutter adds to the cognitive load on the user. The designer has to be sensitive towards the finite capacity of human beings to mentally process information.

An effective means of identifying clutter is the examination of *data-ink* or the *signal-to-noise* ratio for a visualization. Clutter can also be caused due to the following reasons:

1. Lack of visual order.
2. Improper alignment.
3. Insufficiency of white space.
4. Non-strategic use of contrast.

Gestalt principles of *proximity*, *similarity*, *enclosure*, *closure*, *continuity,* and *connection* can be applied in order to eliminate the clutter within a visualization.

Focusing the attention of the viewer

In order to understand the distribution of focus areas in a given visualization, the designer must administer the "where are your eyes drawn" test upon themselves. This test should also be administered on the intended audience during evaluation.

There are various components to effectively designing for focus. They are:

Preattentive attributes

Preattentive attributes can be used to *direct the focus* of the user and to create a sense of *visual hierarchy*, in turn guiding them through the content in a desired sequence. These tools can be leveraged to specifically cater to the user's *iconic*, *short-term*, or *long-term* memory.

Some such commonly used attributes are color, size, position, outline, text formats (bold, italics, underline etc.).

Applying classic design principles

Some aspects of traditional design, such as affordances, accessibility and building acceptance with the audience, can be leveraged in the design of visualizations.

Visual affordances

Visual affordances can be used to highlight important parts of the data, eliminate distractions, and create a clear hierarchy of information for the user.

Accessibility

While designing for accessibility (to a diverse audience), it is important to not overcomplicate things and use text judiciously. Text is particularly effective when used to emphasize takeaways or conclusions.

Aesthetics

Aesthetics is another important parameter that changes the perception of the visualization. Aesthetic designs are perceived as being easier to use, regardless of their actual effectiveness.

Telling a story

The quality of the story can make a huge difference in the retention of key takeaways in the audience's mind. They are powerful enough to ensure retention even if the visualizations are, as Cole puts it, "crummy". However, the opposite is not true. Hence, the quality of the narrative takes precedence over that of the visualization.

Components of the story

Intelligent manipulation of components can truly add quality to the story. There are three main components that can be employed.

Plot

The context that is essential for the given audience forms the plot of the story. This is composed of the setting, the main character, the imbalance (some form of conflict), the balance (the ideal state), and the solution.

Twists

Twists can be composed of data that acts as an interesting turning point in the story. However, if the data isn't interesting, it should not be included in the twists. This can lead to the audience losing their attention and hence the impact of other meaningful components decreases.

Ending

This part summarizes the main takeaways for the audience. Cole believes that the designer of the visualization should "always want the audience to do something" after having told the story. This is where the designer gets to "call for action", and it is important even if the audience disagrees with the view of the designer. Disagreement with the takeaways is still a response, which is more important that the lack of it. Also, the takeaways should be conveyed in a clear manner.

Narrative flow

There are two ways to design the flow of a narrative.

Chronological

This flow is the most natural one, and often the same as that of the designer of the visualization. It usually starts with the identification of a problem, gathering of data to better understand the situation, analysis of data, constructing a solution and recommending a 'call for action'.

Leading with the ending

Here, the designer starts off with the recommendations. This is effective when the audience inherently wants to know what they should do with the analysis, more than how the recommendations were constructed. It sets the tone for the narrative and retains the attention of the user.

Clarity of the story

There are two prominent ways of ensuring clarity in the story. Before creating anything, the designer must always check for this.

Horizontal logic

This logic dictates that the slides in a presentation or the pages in an interactive prototype, when connected together in a sequential manner, should reveal the overarching story with clarity.

Vertical logic

This logic suggests that the content present within a slide or a page must be coherent or consistent with the title.

Some other ways to check for clarity are *reverse storyboarding* and getting a *fresh perspective*.

3. Scoping

Compared to the video by HHMI BioInteractive, “A Place Like No Other” increases the width (covering more concepts) and depth (higher detail) of the focus area. For example, the book explains concepts such as biodiversity, stability, resilience, states (of an ecosystem), and disturbances in a detailed manner.

For the purposes of this project, the **focus largely remains on migration, the principle of regulation and keystone effects** since they are essential is understanding the phenomenon. Associated concepts such as biodiversity, stability and resilience have been only briefly (reduction in detail) explained, in order to make it more consumable to a general audience.

Thus, both the width and depth of concepts mentioned in the prototype lie somewhere between that of the video and the book.

4. Design of narrative

The video follows a linear narrative, starting with Sinclair’s arrival in the Serengeti National Park. It then discusses the sudden increase in the buffalo and wildebeest populations, owing to the eradication of the Rinderpest disease. Next, it provides an explanation of the principle of regulation (bottom-up and top-down). Post this, the phenomenon of migration is explained as a function of changes in season and vegetation. It is also shown how migration provides an advantage to the wildebeest, explaining their incredibly high numbers. Finally, the keystone effect of the wildebeest on the ecosystem is explained.

The book starts by explaining the physical environment of the Serengeti ecosystem, its climate, and habitats. It then talks about the discovery of the Rinderpest disease, its effects on the ungulate population, and its disappearance. Next, it discusses the principle of regulation, and the direction of regulation (bottom-up & top-down). It then talks about the different movement patterns in animals (migration being one of them) and explains the migration of the wildebeest herds. Next, it explains the concept of a keystone species by stating examples. Post that, the concepts of biodiversity, stability and resilience are explained. It also explains how fluctuations in the environment (such as the El Nino event) affect the ecosystem. Next, it explains the concepts of states (of an ecosystem). Finally future directions, threats, and lessons from the Serengeti are summarized.

The prototype draws from both the sources. Since the users would belong to a general audience (not necessarily aware of the Serengeti N.P.), it was necessary to set the context first. Knowing the concepts would help better understand the chronology of events (starting from Rinderpest to the keystone effect of wildebeest). Hence the concepts were introduced before exploring the narrative. The narrative was then visualized in the final section.

5. Design explorations

Over the course of the project, various designs were tried out. This section documents the explorations until stage 3.

5.1 Data

5.1.1 Population

The initial explorations were **line graphs** that focused the attention of the viewer towards areas relevant to the narrative. For example, the eradication of Rinderpest was the reason for the sudden increase in the wildebeest population. In the visualization below, the timeline before the eradication of Rinderpest has been shaded. Also, the steep slope of the line has been accentuated to draw attention towards the explosion.

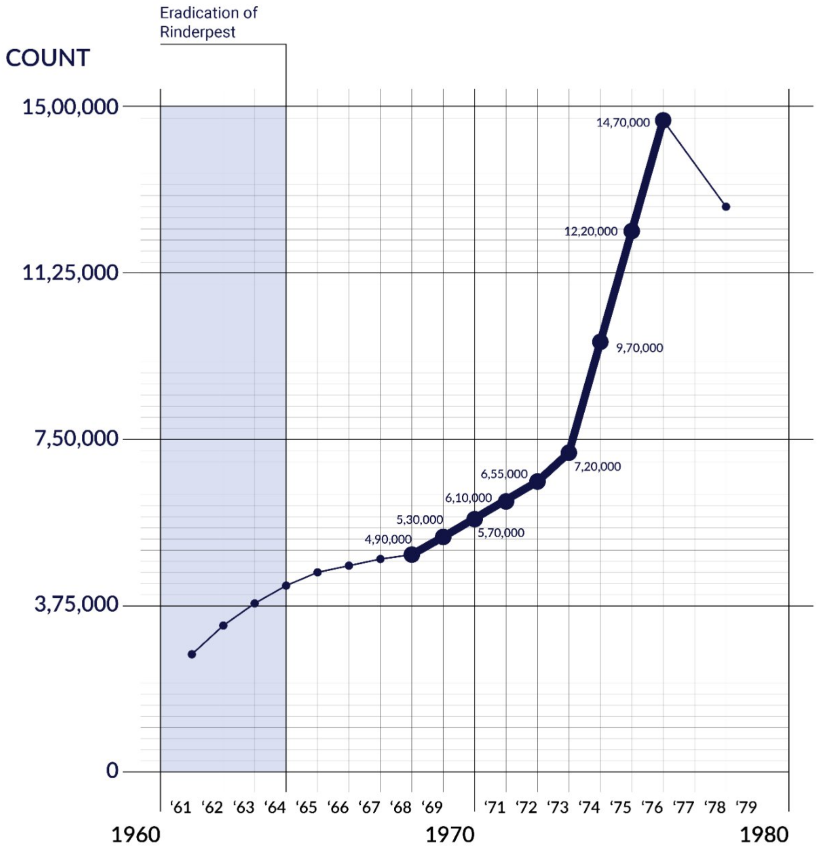


Fig. 5 – Line graph showing wildebeest population

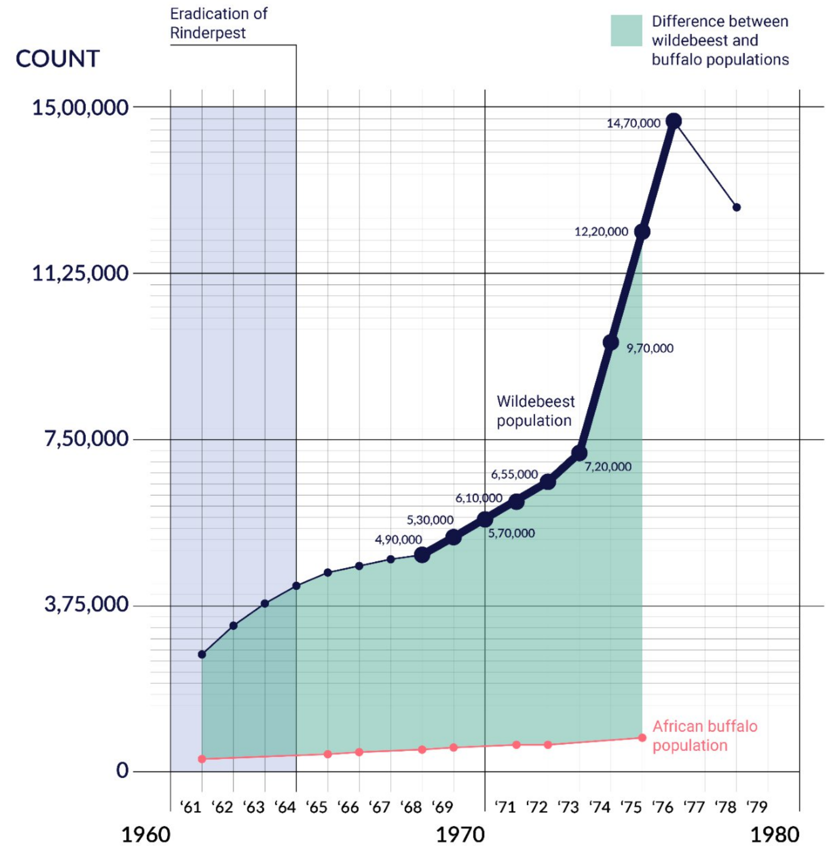


Fig. 6 – Combination of line graphs

For comparing populations between species, the line graphs of different species were **superimposed over a common timeline**. The gap between the two curves was shaded to highlight the difference.

However, these visualizations were **too bland and extremely common**.

In the next iteration, the populations of multiple species were plotted together on a common **circular timeline**. It also used a non-uniform scale along its y-axis in order to save space. This visualization was meant to be part of a **larger composition** of windows, where each window contained the visualization of a concept associated with the explosion in the wildebeest population. Hence the user would be able to mentally connect different concepts relevant to the narrative.

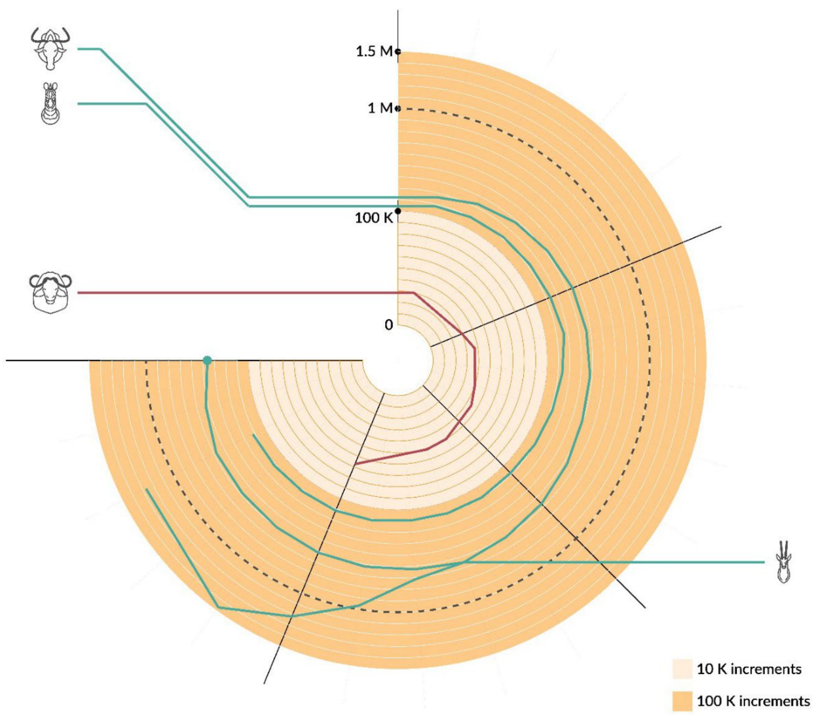


Fig. 7 – Visualizing population on a circular timeline

However, the **cognitive load on the user was too high**. It was quite difficult to make sense of what was going on and the composition became chaotic.

In another iteration, the population was encoded using **line thickness**. The thickness was mapped along a **logarithmic scale** (base 2), to capture a wider range of population.

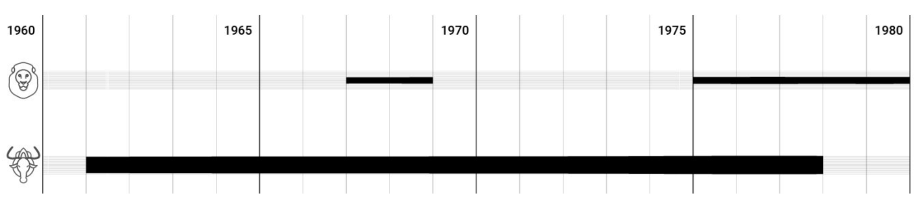


Fig. 8 – Thickness used to encode population

Note that the log values were centrally aligned, i.e., the thickness grows outwards from the center. The difference in population (thickness) between a herbivore (wildebeest) and a carnivore (lion) is quite perceivable. However, the same was hard to differentiate among the ungulates (such as zebra, wildebeest, and Thomson's gazelle). Also, the drastic change in the wildebeest population (especially during the mid-1970s) was near impossible to perceive on this scale. This iteration proved to be inefficient because logarithmic scales are only effective while depicting exponential data. A similar visualization, but bottom-aligned, also failed in the same areas.

Another problem with using logarithmic scales was that generally, people did not know how to read them. Hence, there was the possibility of the user misinterpreting the data (The scale of COVID-19 graphs affects understanding, attitudes, and policy preferences, 2020).

In the next iteration, colour (saturation to be specific) was used to encode population.

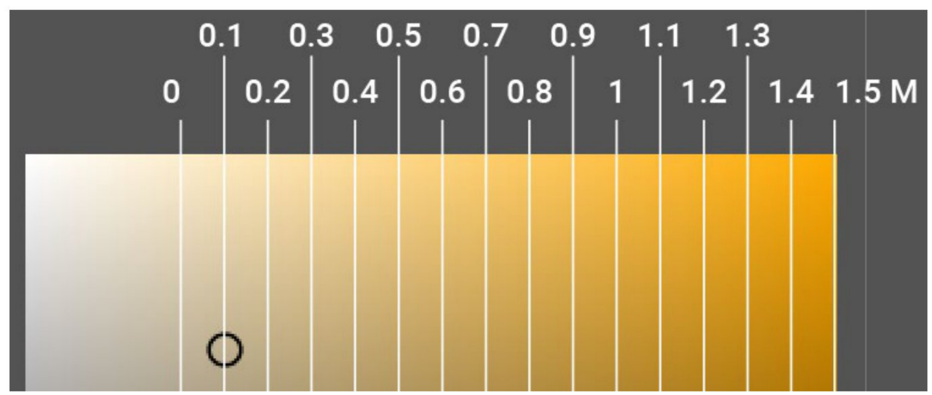


Fig. 9 – Colour used to encode population data

Naturally, this was not a very good option. Firstly, it was extremely difficult to clearly distinguish between values falling into adjacent (0.8M-0.9M and 0.9M-1M) buckets. Secondly, it wasn't efficient in depicting changes in the population.

In the next iteration, **stacked area graphs** were used to visually compare the overall change in population (total population of all species combined) over time. This was discarded as it became difficult to check the population of one species in contrast to another in terms of approximate numerical values. Also, the change in the overall population (all species combined) was not relevant to the narrative.

In this regard, **overlapping area graphs** seemed to be more effective since they allowed for approximate numerical comparisons between species.

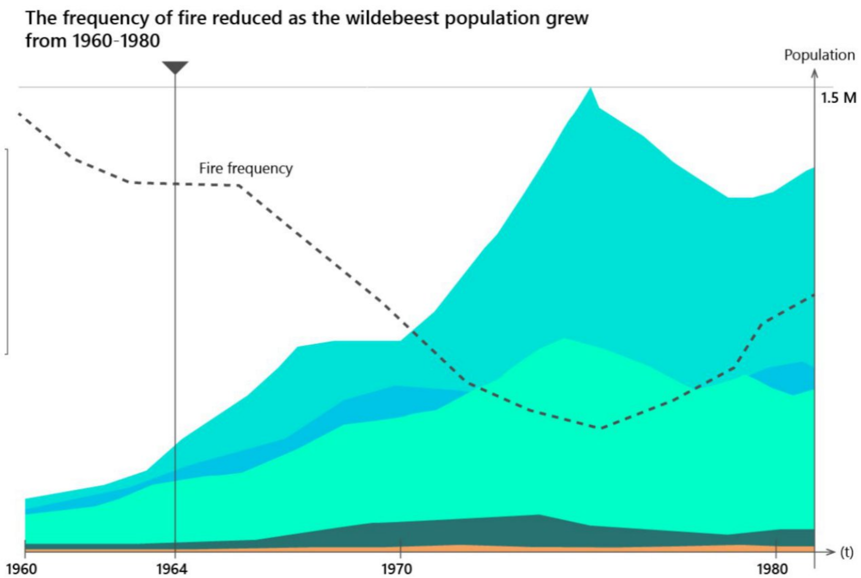


Fig. 10 – Overlapping area graphs

To show the distribution of population between species in the ecosystem, **stacked bar charts** and **tree maps** were explored as well.

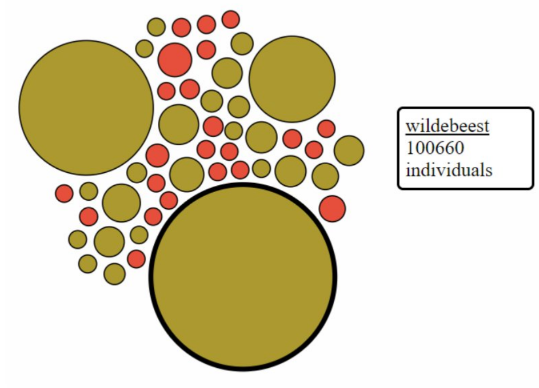
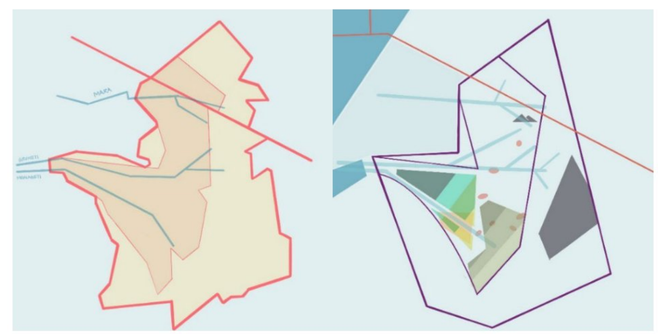


Fig. 11 – Circular tree map

5.1.2 Maps

This set of visualizations was meant to encode information about the geographical boundaries, surroundings, habitats, and vegetation. Schematization served the purpose of reducing the irrelevant details (rugged edges) and visually simplifying the shape of the boundary. Different degrees of abstraction were tried out.



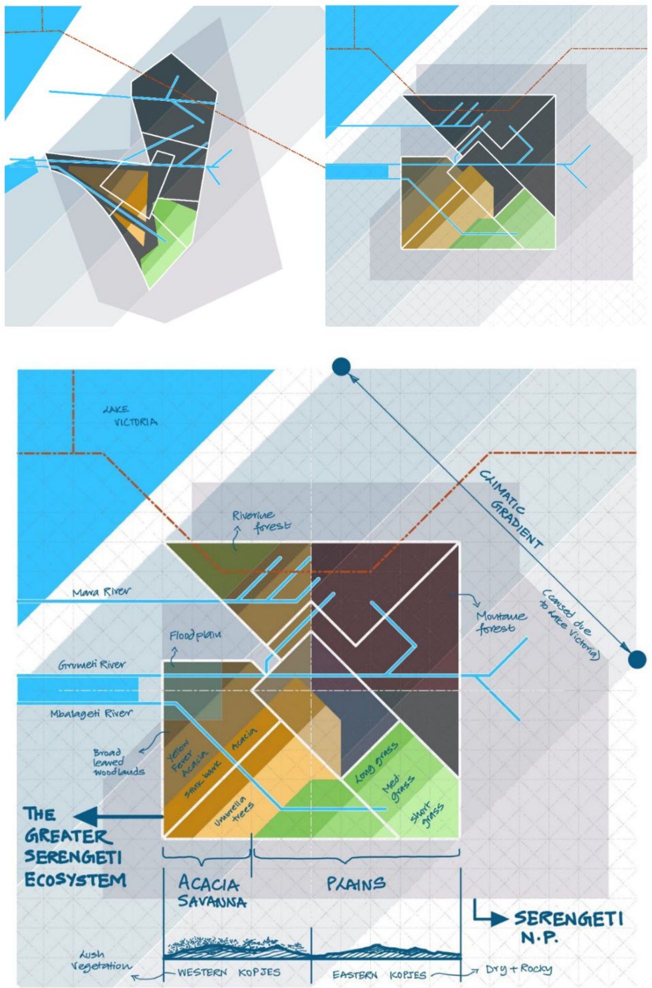


Fig. 12 – Different degrees of schematization

A square-shaped map visualization was explored in accordance with Sinclair's own description of the ecosystem being "roughly a square". The relations between the zones, waterbodies and habitats were maintained. However, this level of abstraction created a lot of confusion and led to the misinterpretation that the national park is actually in the shape of a square. Hence, this version was eventually discarded.

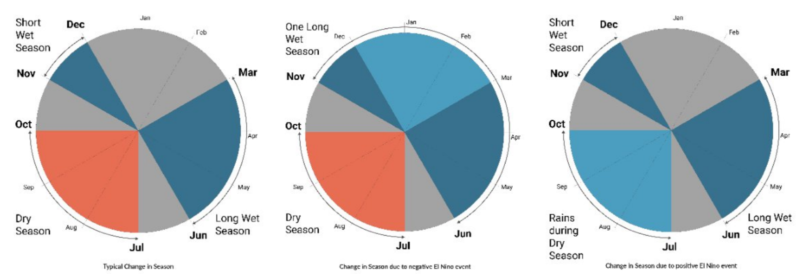
A more reasonable balance was struck where the abstraction did not skew the user's interpretation and also aided visual simplicity.

There are three major rivers (the Mara, the Mbalageti and the Grumeti) in the Serengeti ecosystem. They play an important role in the sustenance of several species. However, they do not hold any relevance to the narrative of the prototype. Hence, they have not been included in the final visualization.

5.1.3 Climate

The seasonal changes are cyclical in nature. There are attributes of climate that had to be visualized, namely season and duration. Seasons were encoded using colour and duration using length.

The initial set of visualizations explored a closed loop idiom, to express the repetitive nature of a cycle. Other possibilities such as horizontal stacked bar graphs were also explored. The advantage of the latter was that it encoded the duration of seasons along just the length. The pie diagrams also encoded the same using length. However, in case of the latter, an additional channel of area was unnecessarily added. Hence pie diagrams were discarded.



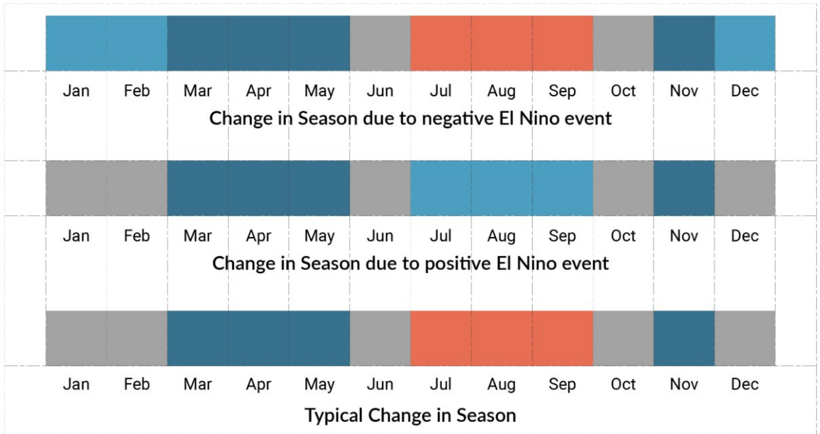


Fig. 13 – Pie diagrams VS stacked bar charts

Since climate is the driver of change (in vegetation) in the ecosystem, the visualization must also express this relationship. Hence an integrated composition of map and climate visualizations was explored.

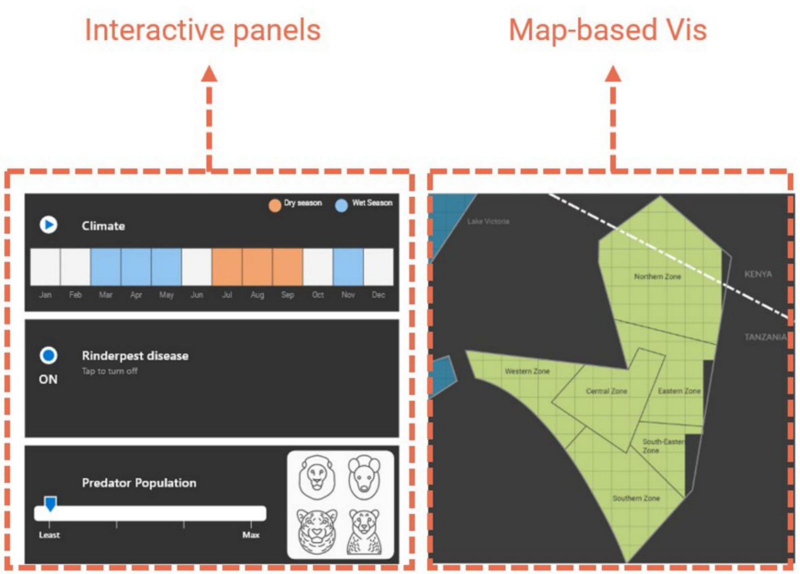


Fig. 14 – Integrated compositions

In the next iteration, the attribute of duration was made less explicit by letting the user perceive it while interacting with the annual timeline.

5.1.4 Species and trophic interactions

This set of visualizations convey necessary information about the 'living' component of the ecosystem, namely the flora, the fauna, their respective trophic levels, and the interactions between them.

The first iteration used simple graphical representations to help maintain a visually familiar semantic. However, labeling was required due to unfamiliarity with the less known species. For the trophic levels, Gestalt principles of proximity and enclosure were applied, and species were clubbed together in horizontal bands.

The trophic interactions are encoded using links between appropriate nodes. Arrows encode the directionality of consumption. For example, if a lion consumes a zebra, the arrow will point towards the zebra.

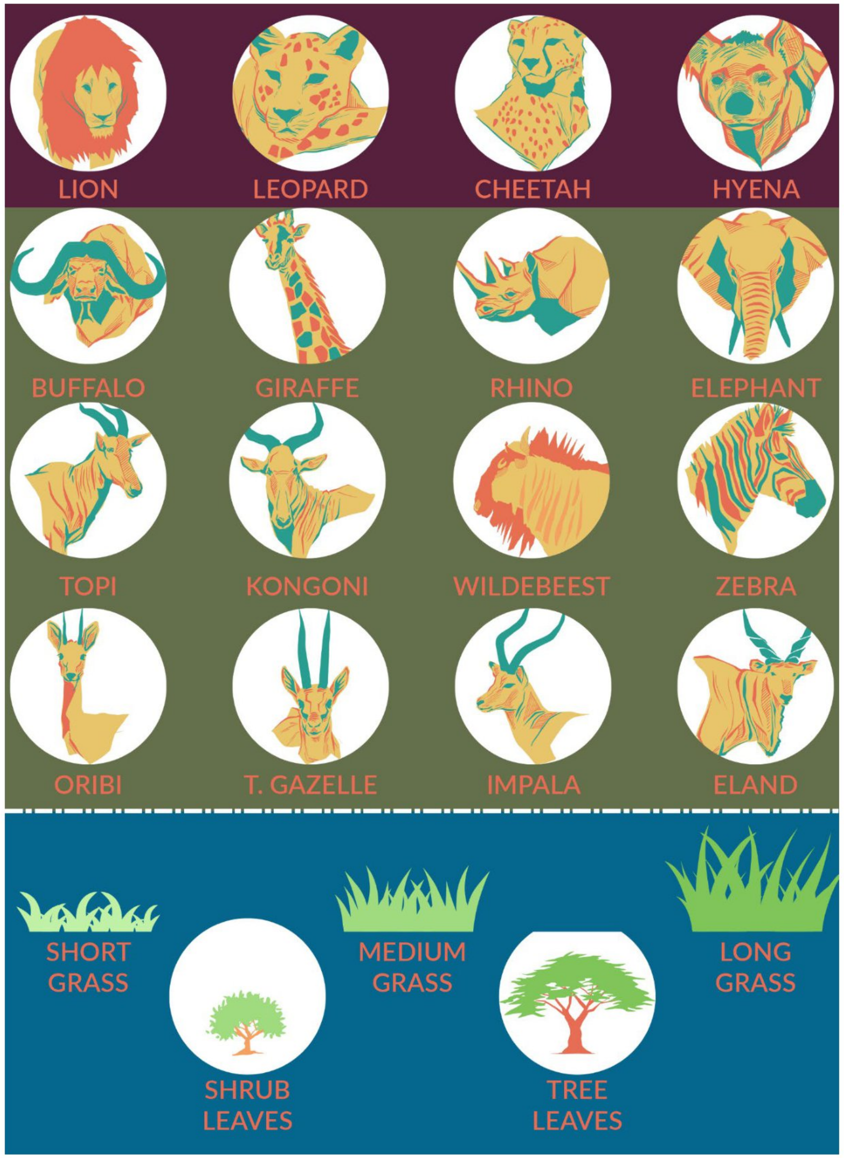


Fig. 15 –Species

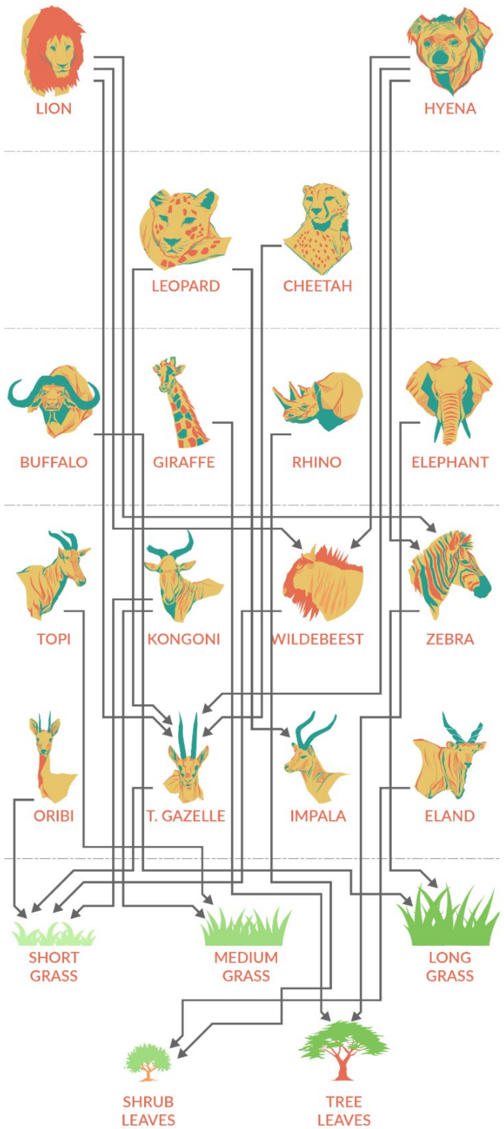


Fig. 16 – trophic interactions

However, not all species were really relevant to the narrative. Hence some of them had to be removed. Also, the visual representations lacked consistency and had the potential to be made mode simple. The trophic interactions were quite difficult to read. For example, it was hard to quickly point out the species that were consumed by any one predator.

In the next iteration, simple and consistent iconography was used to depict the species. The orientation of the trophic interactions signified whether they were carnivorous or herbivorous. However, the visual complexity of the interactions was still making it difficult to read individual interactions.

In the final iteration, the trophic interactions were encoded using thickness of line. It was observed in the earlier iterations that there were common prey or targets. Hence the thickness of the common areas was obtained by multiplying the thickness of individual interactions with the total number of species sharing the common path. This seemed to largely minimize the visual clutter.

5.2 Concepts

A conceptual approach (Berinato, 2016) has been taken to visualize the following visualizations.

5.2.1 Generic species

For explaining some of the concepts such as regulation, it was necessary to use a generic icon that collectively represented species. Their inherent nature and movement pattern had to be integrated into the visual as well. The following icon was used to collectively denote this. Movement pattern has been encoded using the directionality of the arrow-like shape and their inherent nature using colour.

Fig. 17 – Directionality encodes movement pattern (left – residential, right - migratory)

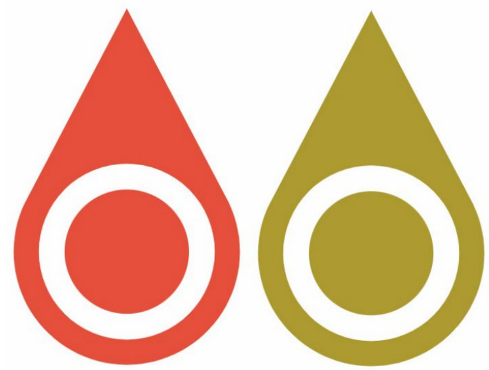


Fig. 18 – Colour encodes inherent nature (red – carnivore, green - herbivore)

5.2.2 Migration

In order to understand migration, it was necessary to visualize the **relationship between the movement of animals and the change in vegetation**. The advantage migratory species have over residential ones can only be understood by observing at least one complete migration cycle. Hence, a progression of visualizations had to be used.

The example below shows a composition of different visualizations such as the map, climate, habitats, species, and their populations. Circles on the map denote fauna. The **colour of the circle indicates individual species** (wildebeest being indigo & buffalo red). **Horizontal bar charts** at the bottom indicate their respective **populations**, starting from 100%. As the cycle progresses, this value increases to ~125% (for the wildebeest) and ~108% (for the buffalo). The **availability of short grass was encoded using opacity**. For example, during the dry season, the opacity of grasslands comes down to 0% whereas in the wet season it remains at a 100%.

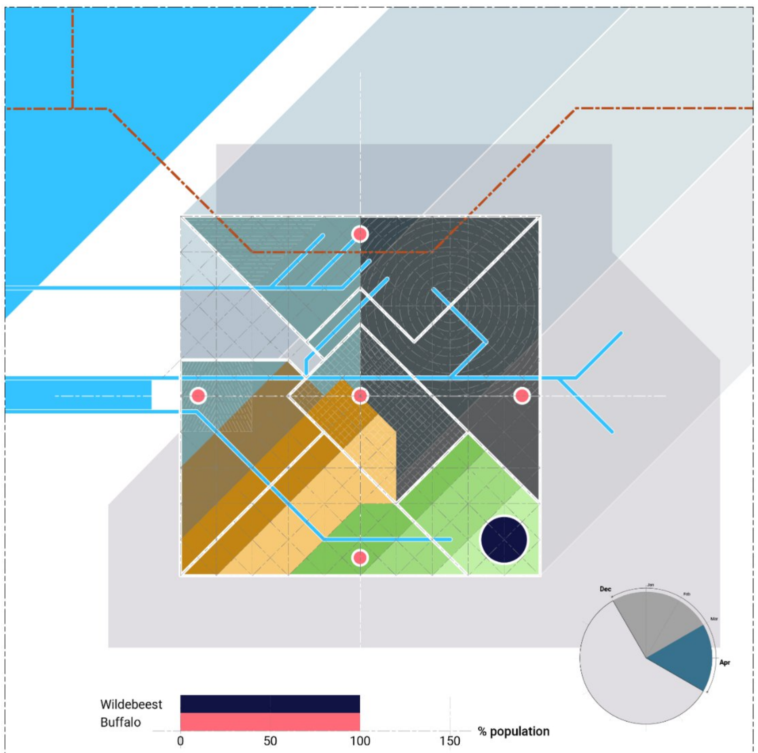


Fig. 19 – Composition using different visualizations

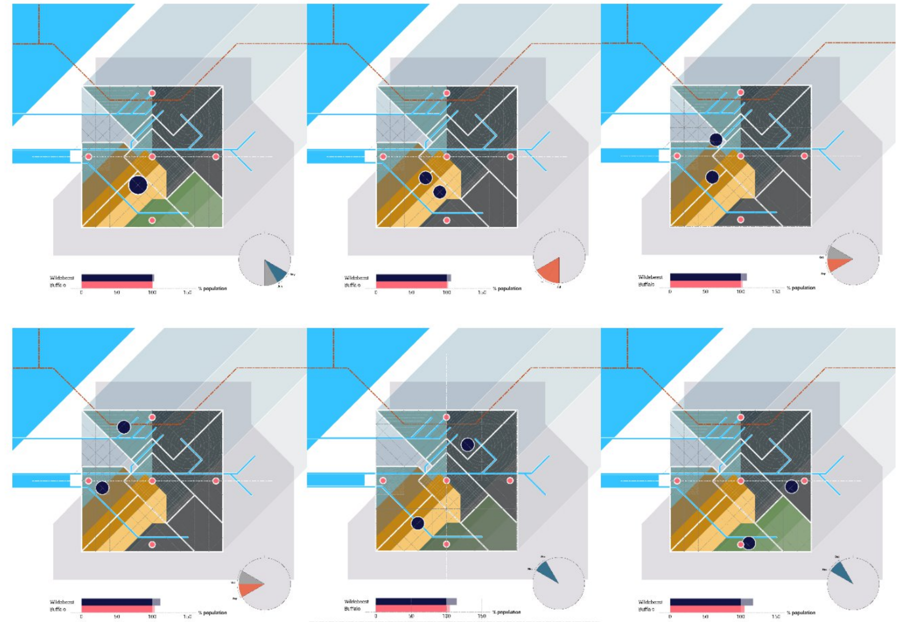


Fig. 20 – Migration cycle

However, the **visual complexity** of this visualization made it quite **difficult to follow the cycle and in turn understand the advantage** of migration. Hence, it needed further simplification.

In the latest iteration, a **conceptual approach** was taken to visualization. Instead of a map, a simple **rhombus** was used to denote the **migratory path**. It was divided into a 3X3 grid layout. At any point in time, **only one of these grid boxes would have grass**. The species were visualized using the simple icons from before. **Predation** was shown using **red arrows and circles that enclosed the prey**.

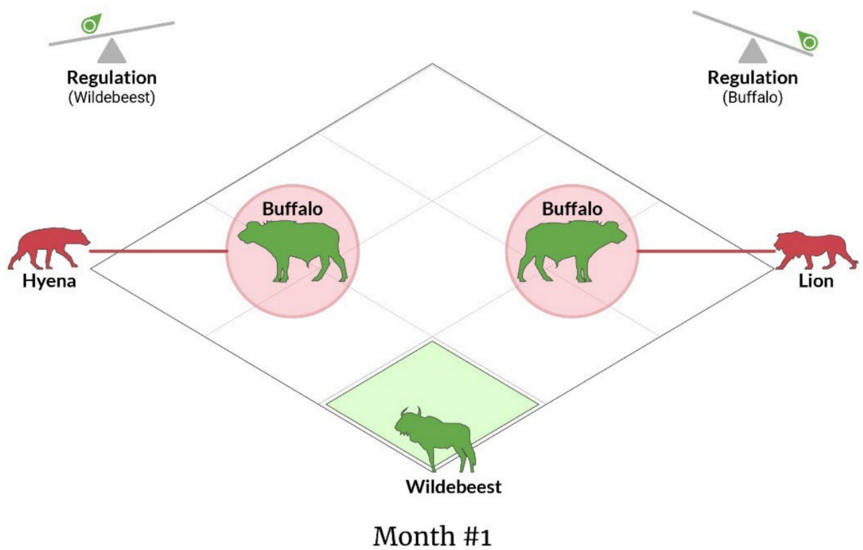


Fig. 21 – Simplified conceptual visualization

5.2.3 Keystone effects

The concept of a keystone species heavily draws from the metaphor of an actual keystone in a masonry arch. Hence, the **visualization mimics the cross-section of an arch**, with the keystone species at the center. However, the keystone effect is a cascading chain of relationships. Each step in the cascade needs to be explained so that the overall effect can be understood. Hence the diagram was designed to **conceptually show the different stages through interactivity**.

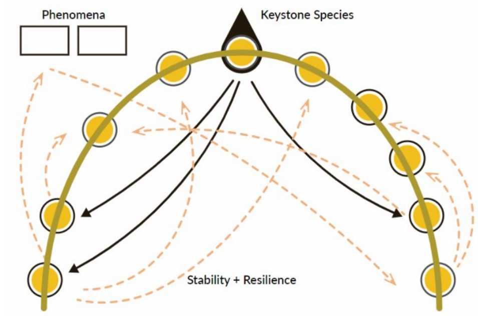


Fig. 22 – Keystone diagram

5.2.4 Regulation

There are three main factors that influence regulation. They are **movement pattern, body size, and the inherent nature** of a species. Based on the **combination of these three factors**, a species (generic) is either top-down or bottom-up regulated. Migratory species, however, are able to counter both types of regulation but lean towards bottom-up regulation.

The **metaphor of a tipping scale** seemed appropriate to depict this phenomenon. Also, the visualization was made more engaging through **interactivity**, such that the user is able to see the outcome based on their own inputs.

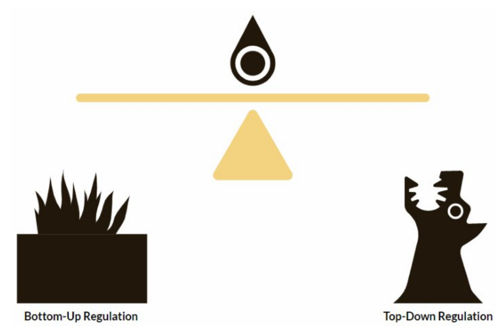


Fig. 23 – Regulation diagram

6. The prototype

Please use this link (https://shobhit17s.github.io/) to access the prototype.

Note: The prototype is meant to be viewed on a laptop/desktop, using Google Chrome or Microsoft Edge. It is not compatible with Safari.

The prototype was developed using HTML, CSS, and JavaScript. It is presented as an interactive web article, currently hosted on GitHub. Most of the visualizations have been made interactive to show different layers of information, navigate through the steps in an ecological process, and make the experience more engaging for the user.

6.1 Summary of narrative

6.1.1 Context

The contextual information includes the **geography, surroundings, climate (including microclimate), and habitats**. The **changes in vegetation in accordance with the climate** (crucial for understanding why the wildebeest migrate in the first place) is also introduced at this point.

The user also gets an idea of the **diversity** (inherent nature and population) of species found in the ecosystem.

6.1.2 Rules that govern ecosystems

This section **includes the key ecological concepts** required for understanding the chronological events, i.e., the main narrative.

Regulation

The main principle at play is that of regulation. Every species in the ecosystem is regulated one of two (either **bottom-up or top-down**) ways.

The **directionality** of regulation mainly depends on factors such as **body size** and **movement** (residential or migratory) pattern, and their **inherent nature**.

Migration

Migration, in particular, proves to be extremely **advantageous for countering the forces of regulation**. For example, species that are small compared to their predators (and hence make for easy prey) can avoid predation if they continuously move around the ecosystem throughout the year. It also allows them to acquire food throughout the year, irrespective of changes in the season. Hence, it raises the equilibrium level of a given species, allowing for a much larger population.

The keystone effect

Every species **directly affects** various other species in the ecosystem. For example, the abundance of one species might impact the population of their predators. However, there can be other **indirect effects** as well. The grazing activity of the wildebeest herds affects the availability of short dry grass, which acts as fuel for forest fires. This gives rise to the concept of a keystone species. Any species that plays a key role in maintaining the state of a given ecosystem is called a keystone species.

Biodiversity, stability, and resilience

It has been proven that a keystone species can increase the biodiversity (the number of species) in the ecosystem. Hence, the effect of a keystone can be measured by the number of species that are affected by its presence. An increase in biodiversity leads to increased stability and resilience, implying that the ecosystem becomes naturally resistant to changing states and develops a higher tolerance to disturbances.

6.1.3 An interesting community of animals

This section provides the set of chronological events that transpired over a 20-year period, from 1960-1980. It includes a brief introduction to the Rinderpest disease, its eradication, and consecutive effects. The population data of three migratory ungulates (wildebeest, Thomson's gazelle, and zebra) and one residential ungulate (African buffalo) illustrates the effect of Rinderpest, migration, and bottom-up regulation at different stages in the time period.

Finally, the keystone effect of the wildebeest population on the community is explained using visualizations.

6.1.4 Conclusion

At the end, this section briefly summarizes the key concepts and their implications for humanity.

6.2 Section 1 - Contextual Information

This section contains basic information about the Serengeti National Park, such as its location & constitution (geographical boundaries, surrounding areas etc.), the microclimatic zones, distribution of habitats, and variation of climate.

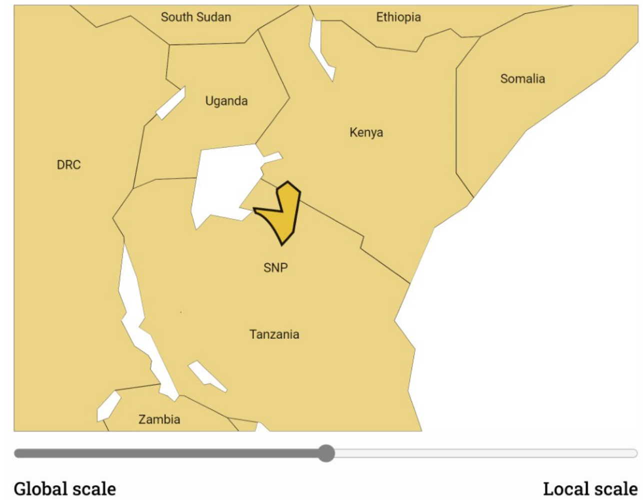


Fig. 24 – Location of Serengeti National Park

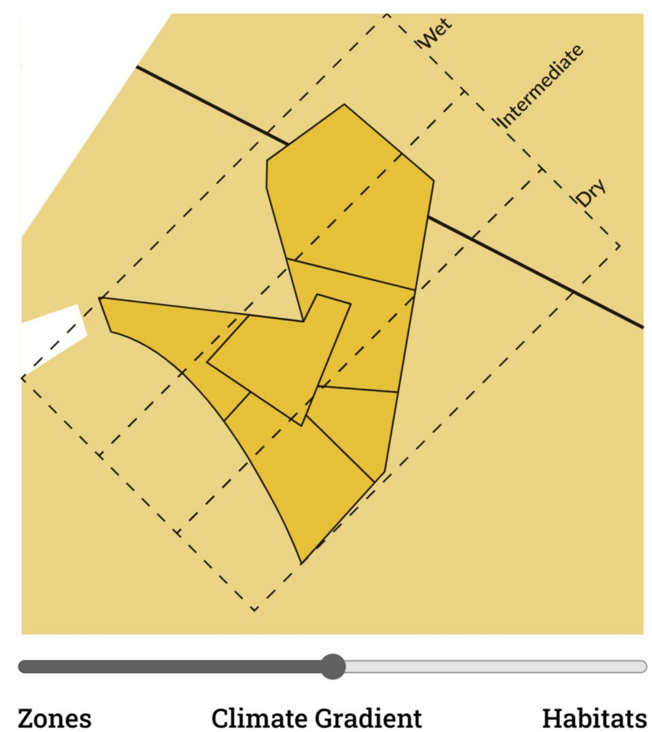


Fig. 25 - Part 1 - Microclimate affects the layout of habitats

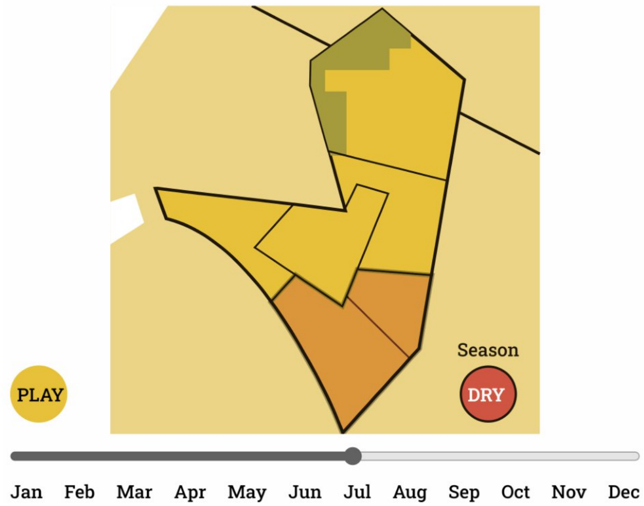


Fig. 26 - Part 2 - Effect of climate on the environment

6.3 Section 2 - The diversity of species

This section introduces the user to the different kinds of animals within the ecosystem. It also visualizes their trophic levels, trophic interactions, population, and inherent nature.

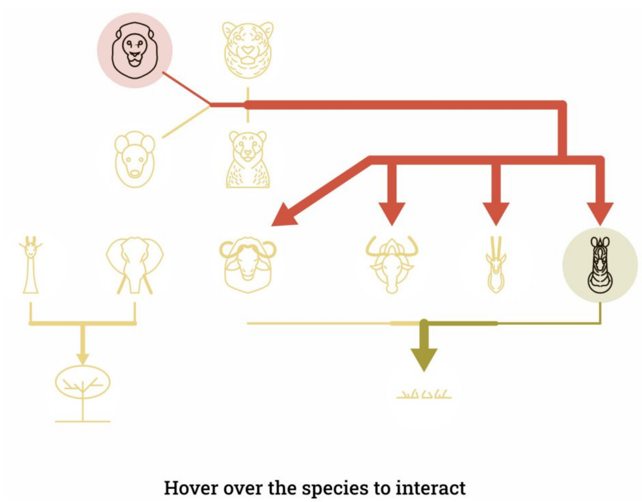


Fig. 27 - Part 3 – Trophic levels and interactions

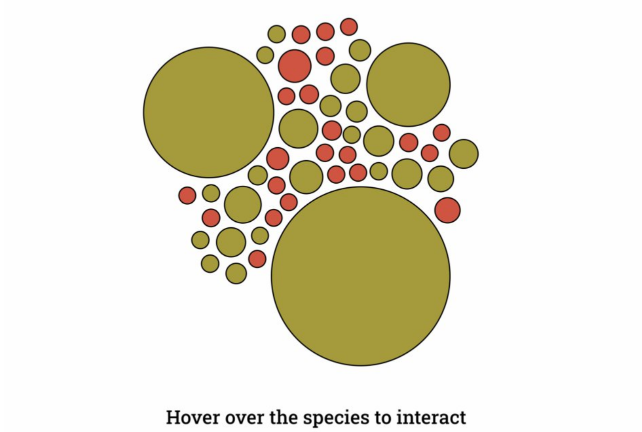


Fig. 28 - Part 4 - Mapping the population

6.4 Section 3 - Rules that govern ecosystems

This section is of higher importance than the previous two, since it explains the ecological concepts such as regulation, migration, and keystone effects. It also touches upon the concepts of biodiversity, stability, and resilience.

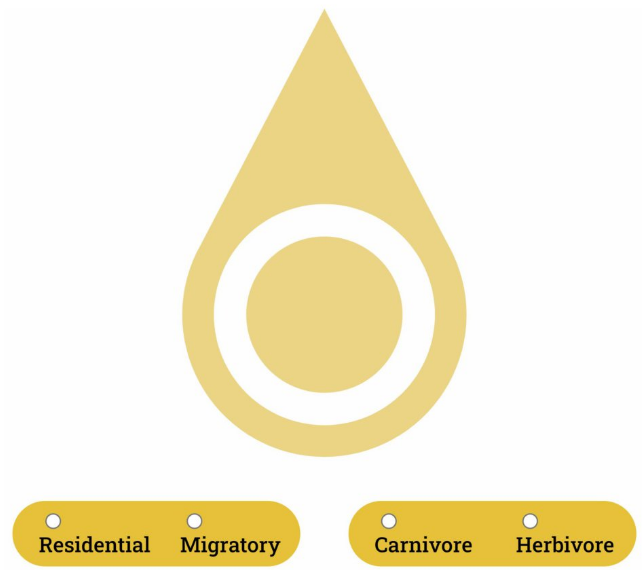


Fig. 29 - Part 5 - Movement patterns in animals

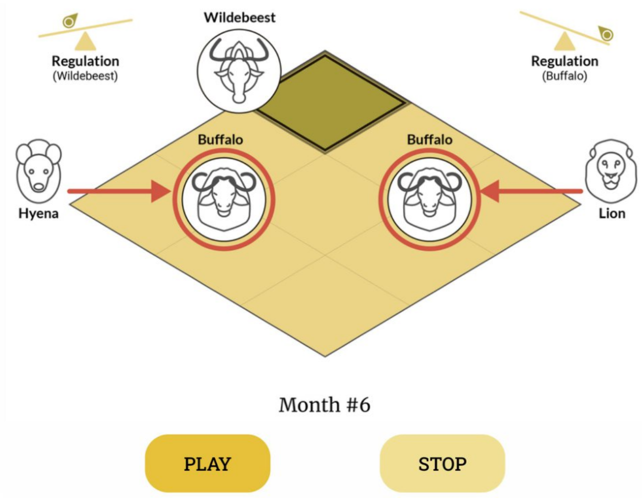


Fig. 30 - Part 6 - The advantage of migration

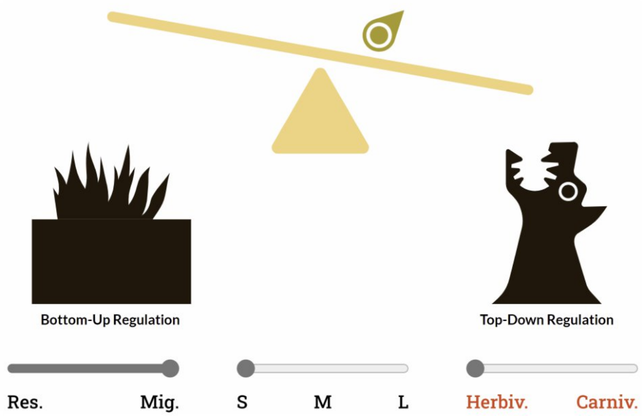


Fig. 31 - Part 7 – The principle of Regulation

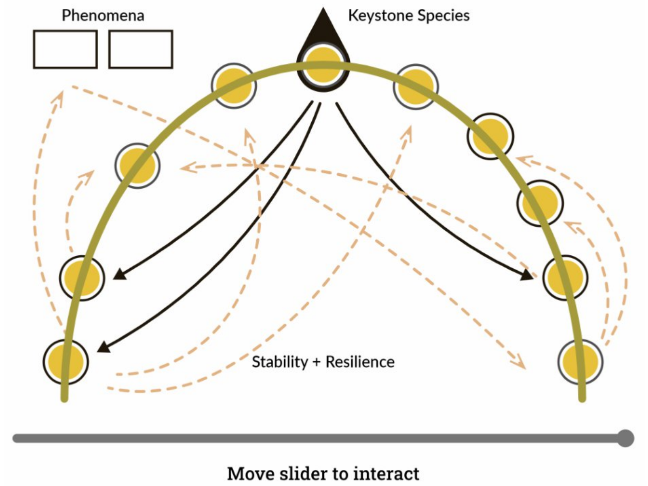


Fig. 32 - Part 8 – Keystone effects

6.5 Section 4 - An interesting community of animals

This section presents the chronological events that originally led to the discovery of the principle of regulation. It visualizes the variation in the population of four ungulate species (wildebeest, zebra, gazelle, and buffalo) across a 20-year time period. At each point in the timeline, the respective events and associated concepts are presented. Short and crisp descriptive text (Knaflic, 2015) is used to communicate the takeaway.

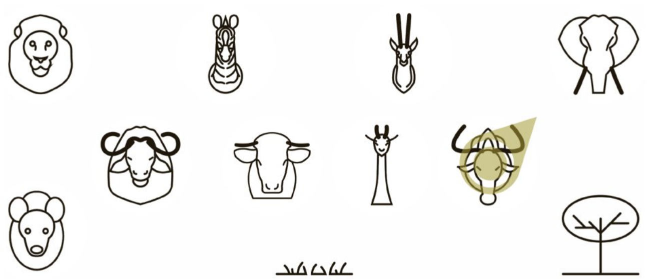


Fig. 33 - Shift of focus to a smaller community

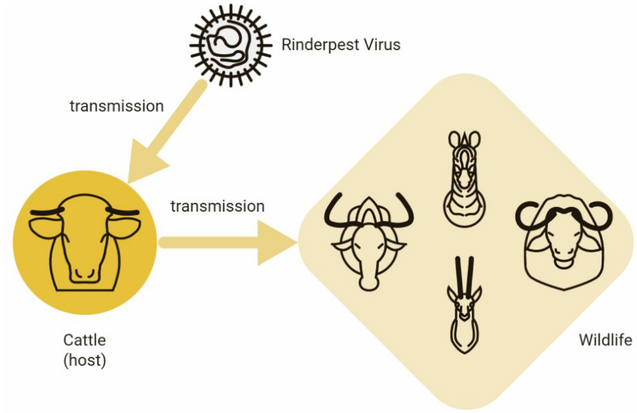


Fig. 34 - Part 10 - The Rinderpest Outbreak

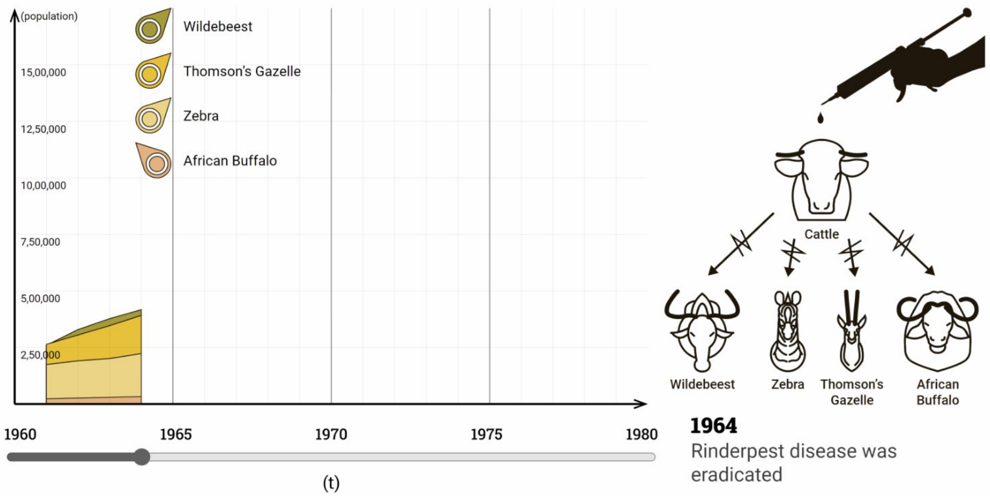


Fig. 35 - Part 11 - Population graph (ungulates)

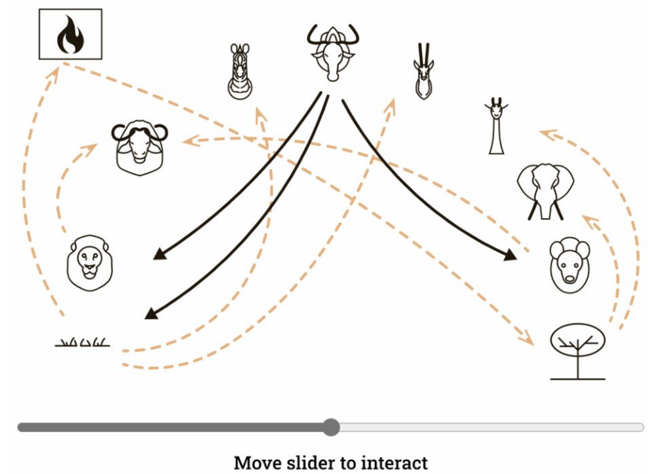


Fig. 36 - Part 12 - Consequences of so many wildebeest

At the end, it explains why the wildebeest became the keystone species in this community of animals.

7. Plan for evaluation

7.1 Evaluation with end users

In this round, the prototype was presented to intended users, i.e., a general audience with an interest in ecology. Since the project has to do with explanatory analysis, the goal of this round was to **test the comprehension** of the user. The retention of concepts, in this case, doesn't take priority.

Two users were recruited for this round of evaluation.

7.1.1 Detailed evaluation plan

Tasks performed by the designer

1. Engage in an initial conversation with the user and discuss their interest in ecology. The purpose of this exercise is to get the user's motivation up to a level that simulates an actual use case.
2. Give a brief introduction of the project, without explaining any of the concepts in too much depth.
3. Present the prototype to the user. Start the evaluation at this point. Do not give them extra time to examine the prototype at this stage.
4. Ask the user to **think out loud** as they are consuming the information. Make **notes about the comprehension level for each part**. Do not stop the flow of the user at any point unless they stop by themselves. Provide any necessary assistance to the user.
5. Be attentive when the user is engaging with the visualizations. Note whether there are any pain points that hinder comprehension.
6. At the end, ask the user to **summarize their learnings**. Check for **concepts that could not be remembered** against their level of importance.
7. Ask the user **what concepts they would apply, while trying to examine a given ecosystem**. Check for the **overlap** between the concepts mentioned by the user and the ones mentioned in the prototype.

Sec.	Parts	Indicator of success	Success criteria
1	**Microclimate affects the layout of habitats**	**Summarize**	The microclimate affects the layout of habitats.
2	**Effect of climate on the environment**	**Summarize**	The availability of vegetation changes in accordance with seasons.
	HORIZONTAL LOGIC OF SECTION#1	**Summarize**	Factors such as microclimate, habitats, macroclimate and change in vegetation must be considered to understand distribution of animals and their movement patterns.
3	**Trophic levels and interactions**	**Browse**	Interact with the visualization to examine the trophic interactions.
		Identify	Trophic levels demonstrate a direction of consumption, with predators being on top and producers being at the bottom.
4	**Mapping the population**	**Compare**	The populations of different species (preferably out loud during the exercise).
		Summarize	The wildebeest, zebra and the Thomson's gazelle have quite a large population compared to others.

			Herbivores seem to have larger populations compared to carnivores.
5	**Movement patterns in animals**	**Browse**	Use the radio buttons to explore all the combinations (total 4 in number).
		Compare	The different combinations of radio inputs to familiarize self with the
		Summarize	The directionality and colour-coding of the species icon.
6	**The advantage of migration**	**Browse**	At least watch one full cycle of the animation (month 1-12).
		Compare	The movement patterns of wildebeest and buffalo.
		Summarize	The advantages of migration correctly.
7	**The principle of Regulation**	**Browse**	Use the sliders to change the settings.
		Compare	Try out all combinations and compare outcomes.
		Summarize	Migratory species have an advantage of countering top-down regulation.
8	**Keystone effects**	**Browse**	Use the sliders to view each phase of the cascade.
		Identify	The distinctive features of each phase.
		Summarize	The cascade.
9	**<IN BETWEEN>**	**SHIFT OF FOCUS**	Realize that a smaller community within the ecosystem will be analyzed.
10	**The Rinderpest Outbreak**	**Identify**	The host of the Rinderpest infection as cattle.
		Summarize	The transmission path and its effect on the population of ungulates (mentioned in the body of text).
11	**Population graph (ungulates)**	**Browse**	Use the slider to examine the change in population over the years.
		Identify	The events and the concepts associated with them, by examining the juxtaposed views.
		Summarize	The events and their underlying reasons (at a broad level) over the 20-year time period.
12	**Consequences of so many wildebeest**	**Browse**	Use the slider to view the different stages of the cascade.
		Compare	The similarity of the current visualization to “keystone

			effects". Memory be relevant in this step of evaluation.
		Identify	The changes that took place in the community in each stage.
		Summarize	The overall progression of events (along with the reason) broadly.

Table 2 – Plan for evaluation with users

7.2 Findings from the evaluation

After conducting the evaluation, areas that worked and did not work were clearly identified. However, the following text only discusses the drawbacks. For viewing the complete and detailed evaluation, please navigate to the annexure.

7.2.1 Limitations

1. The **takeaway** (i.e., microclimate affects the layout of habitats) of the visualization with the climatic zones was **not clear** to one of the users.
2. The visualization with the change in vegetation mapped to the changes in season is hard to read because of **overlaps in colors and a lack of labelling/legend**.
3. The **visualization for a generic species** was not understood by one of the users. She especially struggled with remembering the **directional encoding** (left vs right) of the movement pattern.
4. The animation **visualizing the basic advantage of migration** (that it is able to counter both bottom-up and top-down regulation) was either **completely misinterpreted or not understood** at all by the users.
5. The visualization about the **keystone effect** (where the concept is explained) was **difficult for both the users to understand**. The generic species in the diagram weren't consistent with earlier visualizations and hence created confusion. Similarly, the diagram that visualizes the keystone effect of the wildebeest on the rest of the ecosystem **did not explain the exact nature of the relationships between the nodes**. For example, it was understood that wildebeest affected the lion population, but the user had to guess how. The original intent was to explain the concept of keystone effects using a certain diagram, and then leveraging the same diagram to explain the keystone effects of the wildebeest. This discontinuity, however, between the two diagrams did not serve the intended purpose.
6. The **order of the some of the content was inappropriate** in certain areas. For example, the concept of regulation must come before that of migration. This is because the advantage of migration is that it counters both types of regulation. However, movement pattern is one of the parameters affecting regulation. Hence an adequate explanation of both concepts is needed before viewing either of them.

8. Discussion

The prototype explored both conceptual and data-driven approaches to visualizing ecological concepts in general, and in the context of the Serengeti ecosystem. Findings from the qualitative evaluation revealed areas that act as pain points for the end user, with a further scope for improvement. To summarize:

1. There are **some parts that require labeling and better visual design** (especially concerning the colour).
2. **Better channels** need to be explored for **visualizing the movement (migratory vs residential) pattern** in species.
3. The visualization that aims to conceptually explain the advantage of migration caused a lot of confusion. **Better ways to explicitly convey the avoidance of**

predation and acquisition of temporary food by migratory species need to be designed in the future.

4. The **keystone diagrams** need to **clearly express the nature of relationships** between the species. Also, proper labeling is required.

9. Personal learnings

My own interest in learning about phenomena related to ecology & nature drove this project. Hence, I started reading the book “A Place Like No Other”. However, mid-way I realized the importance of consulting a domain expert from the start of the project, since learning on my own proved to be time consuming. It reduced the time I could spend on designing the visualizations themselves. The scope & direction of the project also changed through the timeline due to lack of guidance from a domain expert, resulting in further delays and confusion. Also, certain types of data were not readily available, as I’d hoped. Such contingencies need to be anticipated beforehand. These are lessons I shall remember in order to increase my productivity in the future.

10. Annexure

10.1 Feedback from Stage 3

In summary, the feedback received from stage 3 (18th April 2022) was regarding the general justification of decisions, integration of the panels into a single interactive visualization, choosing the appropriate complexity of content, and method of evaluation for the final prototype. The feedback can be categorized as follows.

10.1.1 Justification of design decisions

A common theme of concern was regarding the circularity of the timeline used in the prototype. It was suggested that timelines are inherently read in a linear fashion, and that a strong justification needs to be provided for the circularity. Although the circular form helped fit the entire composition into a compact window, the size and layout of the windows themselves were also in question.

Another point of concern was regarding the scale (logarithmic VS linear) of the graphs. Depending on the range of data required to be shown, the choice of scale becomes crucial (Khan Academy, 2011). It was suggested that more thought be given to the appropriateness of the scale being used, with the necessary justification.

10.1.2 Correlation VS causation

In the previous prototype, the term 'causal' was used to explain various steps in the cascading effects. However, a mere correlation does not imply causation (Australian Bureau of Statistics).

Although there is a strong correlation between many components (for example the growth in grass availability and the increase in wildebeest population), the term 'causality' shouldn't be used to describe them unless specified in the concerned literature.

10.1.3 Increase in complexity

A few questions were raised regarding the role of other animals (excluded from the narrative) in the ecosystem and the contribution of various other components (water resources and weather) that were absent in the schematic diagram, towards the effects highlighted in the cascade. It is important to note that climate (mentioned in the schematic diagram) was one of the primary drivers of change. However, this relation was not identified by the jury.

10.1.4 Integration of seemingly disconnected windows

It was felt that the panels and windows were not appropriately linked with each other, resulting in an increase in the overall time taken to process the information. The aim should be to minimize this time interval.

The windows might seem more cohesive and connected if the design is able to call attention to a change so that the viewer is aware about the relation of the change to its corresponding value. The idea is to call attention through some kind of treatment like highlight or animation, so that people do not miss it. Hence, it makes for an interesting design challenge.

It was suggested that efforts be directed towards integrating the various separate windows into a single seamless visualization.

10.1.5 Task-based comparative evaluations

From the perspective of evaluation, formulation of essential tasks (to be performed by the user of the tool) and a comparative study between other similar tools were suggested.

10.2 Detailed findings (end users)

10.2.1 Findings from user#1

Sec.	Parts	Indicator of success	Success criteria	Observations	Result
1	Microclimate affects the layout of habitats	Summarize	The microclimate affects the layout of habitats.	The user was able to clearly identify the wet and dry regions in the climatic gradient. However, they found it hard to read the habitat labels. In the end, he was **able to summarize that short & long grass grow in the dry areas whereas the "bigger" things grow in the wet areas**.	Success.
2	Effect of climate on the environment	Summarize	The availability of vegetation in the grasslands changes in accordance with seasons.	The user couldn't understand why there were no seasons in between the wet and dry seasons. He asked if some transitional season like autumn existed in between these. He **got confused with the colour-coding in the South & South-Eastern region** of the map. He also **misidentified the vegetation** as trees rather than grass. He **could only understand the visualization after a lot of assistance**.	Failure.
3	Trophic levels and interactions	Browse	Interact with the visualization to examine the trophic interactions.	The user **interacted** with the visualization **without any nudges**.	Success.
		Identify	Trophic levels demonstrate a direction of consumption, with predators being on top and producers being at the bottom.	The user **could clearly understand the direction of consumption and the hierarchy of species**. However, the user said that he had seen lions kill giraffes before. This visualization shows the preferred diets of species, which needs clarification.	Success.
4	Mapping the population	Compare	The populations of different species (preferably out loud during the exercise).	The user **intuitively hovered** over the visualization and started to compare species.	Success.

		Summarize	The wildebeest, zebra and the Thomson's gazelle have quite a large population compared to others. Herbivores seem to have larger populations compared to carnivores.	The user was able to instantly realize the enormity of the wildebeest population compared to other species. However, **he did not summarize the required content** and hence did not entirely meet the success criteria. He understood the colour coding of inherent type but **did not make any observations about herbivores and carnivores**.	Partial success.
5	**Movement patterns in animals**	**Browse**	Use the radio buttons to explore all the combinations (total 4 in number).	The user **interacted** with the visualization **without any nudges**.	Success.
		Compare	The different combinations of radio inputs to familiarize self with the	The user **compared different combinations and reflected** on what they meant. However, he did not understand why directionality was used to encode movement pattern and not icons.	Success.
		Summarize	The directionality and colour-coding of the species icon.	User was **able to summarize the directionality and type coding**.	Success.
6	**The advantage of migration**	**Browse**	At least watch one full cycle of the animation (month 1-12).	The user **interacted** with the visualization **without any nudges**.	Success.
		Compare	The movement patterns of wildebeest and buffalo.	The user **compared the movement patterns of wildebeest and buffalo without any nudges**.	Success.
		Summarize	The advantages of migration correctly.	The user **completely misinterpreted the visualization** and inferred that the buffalo actually have an advantage because the wildebeest "come to their rescue" and share the predation.	Failure.
7	**The principle of Regulation**	**Browse**	Use the sliders to change the settings.	The user **interacted** with the visualization **without any nudges**.	Success.
		Compare	Try out all combinations and compare outcomes.	The user **compared the different combinations intuitively, without any nudges**.	Success.

		Summarize	Migratory species have an advantage of countering top-down regulation.	The user was able to understand that carnivores are only residential in the Serengeti. The user was **able to correctly identify the advantage small & medium migratory herbivores have** in terms of regulation. It is interesting to note that he perceived the species as being afraid of going into the predator's mouth (top-down regulation).	Success.
8	**Keystone effects**	**Browse**	Use the sliders to view each phase of the cascade.	The user **interacted** with the visualization **without any nudges**.	Success.
		Identify	The distinctive features of each phase.	The user was able to **clearly identify the distinctive changes in each step**. However, he had trouble in interpreting certain icons (like the ones for phenomena). He also felt that there are two types of phenomena, whereas the icons represented phenomena in general.	Success.
		Summarize	The cascade.	The user got tired of reading the textual description and rushed through it. The user **did not summarize anything** from this diagram.	Failure.
9	**<IN BETWEEN>**	**SHIFT OF FOCUS**	Realize that a smaller community within the ecosystem will be analyzed.	The user **did not really understand that the focus had shifted**. He did engage with the content to see which species were migratory and residential. However, he incorrectly thought there was some hierarchy among the species due to their alignment.	Failure.
10	**The Rinderpest Outbreak**	**Identify**	The host of the Rinderpest infection as cattle.	The user **correctly identified the cause of the spread** of Rinderpest.	Success.
		Summarize	The transmission path and its effect on the population of ungulates (mentioned in the body of text).	The user was able to **correctly summarize the transmission path**. However, he wanted to know when the disease had started spreading. Upon telling him, he exclaimed that the ungulates suffered for quite a long time due to the spread.	Success.

11	**Population graph (ungulates)**	**Browse**	Use the slider to examine the change in population over the years.	The user **interacted** with the visualization **without any nudges**.	Success.
		Identify	The events and the concepts associated with them, by examining the juxtaposed views.	The user **paused and took time to examine** the visualization whenever a concept appeared.	Success.
		Summarize	The events and their underlying reasons (at a broad level) over the 20-year time period.	The user incorrectly assumed that the wildebeest population was the highest due to its legend being placed at the top. He was also wondering why nothing happened in 1960 & 1961. This was because the data was only officially recorded starting from 1961. The user was able to **correctly summarize that the wildebeest population boomed due to eradication of Rinderpest and that they stabilized due to bottom-up regulation**.	Success.
12	**Consequences of so many wildebeest**	**Browse**	Use the slider to view the different stages of the cascade.	The user **interacted** with the visualization **without any nudges**.	Success.
		Compare	The similarity of the current visualization to "keystone effects". Memory would be relevant in this step of evaluation.	The user **did not make any explicit comparisons** in this regard.	Failure.
		Identify	The changes that took place in the community in each stage.	The **user was not really able to correctly understand all the changes** because he could not figure out what was decreasing or increasing. He just knew that things were related.	Failure.
		Summarize	The overall progression of events (along with the reason) broadly.	The user **completely misunderstood the diagram** and drew incorrect conclusions.	Failure.

10.2.2 Findings from user#2

Sec.	Parts	Indicator of success	Success criteria	Observations	Result
1	Microclimate affects the layout of habitats	Summarize	The microclimate affects the layout of habitats.	The user had some confusion regarding what the zones were. She felt like they were direction indicators within the national park. The angular text labelling the climatic gradient was hard to read for her. The user **did not summarize anything** at all because she did not seem to draw any inference out of the habitat diagram.	Failure.
2	Effect of climate on the environment	Summarize	The availability of vegetation in the grasslands changes in accordance with seasons.	The user **was able to correctly infer and summarize that the grasslands change across the year**. However, she incorrectly identified the entire green area as the grasslands (and not just the South & South-Eastern zones). This user also inquired about the third (white) season that lied in between the wet & dry seasons.	Success.
3	Trophic levels and interactions	**Browse**	Interact with the visualization to examine the trophic interactions.	The user **interacted** with the visualization **without any nudges**.	Success.
		Identify	Trophic levels demonstrate a direction of consumption, with predators being on top and producers being at the bottom.	The user **was correctly able to identify the trophic levels and hierarchy** (although she called it "food chain"). She also drew the inference that diets vary between herbivores.	Success.
4	Mapping the population	Compare	The populations of different species (preferably out loud during the exercise).	The user started to **compare the population of different species** within the visualization, **without any nudges**.	Success.
		Summarize	The wildebeest, zebra and the Thomson's gazelle have quite a large population compared to others. Herbivores seem to have larger populations compared to carnivores.	The user **did not voluntarily draw any inferences** about the largeness of the wildebeest population in comparison to the others OR anything about herbivores. Maybe the user should have been probed at this	Failure.

				point.	
5	**Movement patterns in animals**	**Browse**	Use the radio buttons to explore all the combinations (total 4 in number).	The user **interacted** with the visualization **without any nudges**.	Success.
		Compare	The different combinations of radio inputs to familiarize self with the	The user **intuitively tried out different combinations** of movement patterns, and inherent type.	Success.
		Summarize	The directionality and colour-coding of the species icon.	The user **did not understand that the icon indicated species** (generic). After nudging her, she struggled to understand the encoding for the movement pattern.	Failure.
6	**The advantage of migration**	**Browse**	At least watch one full cycle of the animation (month 1-12).	The user **interacted** with the visualization **without any nudges**.	Success.
		Compare	The movement patterns of wildebeest and buffalo.	The user **intuitively started to compare the wildebeest with all the other species**, instead of just the buffalo. However, the user did remain quiet for a while, trying to understand what was happening.	Partial success. It isn't clear which species must be compared.
		Summarize	The advantages of migration correctly.	The **user couldn't summarize** the advantages of migration. Needed a lot of assistance for comprehension.	Failure.
7	**The principle of Regulation**	**Browse**	Use the sliders to change the settings.	The user **interacted** with the visualization **without any nudges**.	Success.
		Compare	Try out all combinations and compare outcomes.	User **intuitively started to compare** different combinations of parameters.	Success.

		Summarize	Migratory species have an advantage of countering top-down regulation.	The user **was able to articulate the summary correctly**. She remarked that she was able to make sense of the concept of migration after having interacted with the visualization for regulation.	Success.
8	**Keystone effects**	**Browse**	Use the sliders to view each phase of the cascade.	The user **interacted** with the visualization **without any nudges**.	Success.
		Identify	The distinctive features of each phase.	The **user was unable to identify several distinct features** associated to the steps. She required a lot of assistance to comprehend this diagram. She too incorrectly interpreted that there were two phenomena.	Failure.
		Summarize	The cascade.	The user **couldn't summarize the cascade** correctly at all. She felt that it lacked labeling. Hence the central text at the bottom of the visualization was ineffective.	Failure.
9	**<IN BETWEEN>**	**SHIFT OF FOCUS**	Realize that a smaller community within the ecosystem will be analyzed.	The user **did not realize that the focus had shifted** from the ecosystem to a community. However, she did interact with the visualization and correctly interpreted the inherent nature and movement patterns of the animals.	Failure.
10	**The Rinderpest Outbreak**	**Identify**	The host of the Rinderpest infection as cattle.	The user **could correctly identify** the host of the virus and the path of transmission.	Success.
		Summarize	The transmission path and its effect on the population of ungulates (mentioned in the body of text).	She **wasn't able to summarize entirely** as she did not identify what happened to the ungulate population.	Partial success.
11	**Population graph (ungulates)**	**Browse**	Use the slider to examine the change in population over the years.	The user **interacted** with the visualization **without any nudges**.	Success.

		Identify	The events and the concepts associated with them, by examining the juxtaposed views.	The colours in the visualization misled the user into thinking that buffalo was a carnivore. However**, she correctly identified & interpreted the events and associated concepts**.	Success.
		Summarize	The events and their underlying reasons (at a broad level) over the 20-year time period.	The user was able to **correctly summarize the progression of events and their reasons**.	Success.
12	**Consequences of so many wildebeest**	**Browse**	Use the slider to view the different stages of the cascade.	The user **interacted** with the visualization **without any nudges**.	Success.
		Compare	The similarity of the current visualization to "keystone effects". Memory would be relevant in this step of evaluation.	The user noticed the similarity and **drew a comparison with the earlier diagram**.	Success.
		Identify	The changes that took place in the community in each stage.	The user intuitively started to compare the stages in the cascade. She **did correctly identify the stages, but the interpretation of the effects were wrong**.	Partial success.
		Summarize	The overall progression of events (along with the reason) broadly.	She **could not correctly summarize** the overall progression of events.	Failure.

10.3 Plan for evaluation from an ecology perspective

In this round, the prototype was meant to be presented to domain experts. The aim was to **assess the correctness of the content** and whether **any essential concepts have been neglected**. Also, the scope needs to be checked for appropriateness, keeping in mind a general audience.

Tasks that needed to be performed by the designer

1. Give a brief explanation about the project and the **intended purpose** of the prototype.
2. Present the prototype to the expert.
3. Clearly mention the **scope & concepts** that have been covered in the prototype.
4. Ask the expert to go through the prototype and **assess the correctness** of the content. Do not interfere unless the expert requires any assistance.
5. Note any feedback & comments.

Questions that needed to be answered

1. Are there any technical discrepancies throughout the prototype? If yes, where exactly?
2. Does the prototype cover an appropriate width of ecological concepts that can be consumed by a general audience?
3. Which concepts should a layman apply while trying to examine a given ecosystem?
4. Is there an inherent weightage among these concepts, in terms of points to be remembered by the users?

Unfortunately, this round of evaluation could not be executed in its entirety, since not many domain experts were available. Out of the four experts approached for feedback, only one responded. Although positive, the feedback was quite broad and rather vague. Hence, it hasn't been included in the main portion of the report. None had the time for a one-on-one discussion.

References

Ahluwalia, Harshit. 2021. How to choose the Right Chart for Data Visualzation. *Analytics Vidhya.* [Online] Septembet 1, 2021. [Cited: April 29, 2022.] https://www.analyticsvidhya.com/blog/2021/09/how-to-choose-the-right-chart-for-data-visualization/.

Aniruddha. 2020. 8 Data Visualization Tips to Improve Data Stories. *Analytics Vidhya.* [Online] July 27, 2020. [Cited: April 29, 2022.] https://www.analyticsvidhya.com/blog/2020/07/8-data-visualization-tips-to-improve-data-stories/.

Australian Bureau of Statistics. Statistical Language - Correlation and Causation. [Online] [Cited: April 23, 2022.] https://www.abs.gov.au/websitedbs/D3310114.nsf/home/statistical+language+-+correlation+and+causation#:~:text=A%20correlation%20between%20variables%2C%20however,relationship%20between%20the%20two%20events..

Berinato, Scott. 2016. *Good Charts: The HBR Guide to Making Smarter, More Persuasive Data Visualizations.* s.l. : Harvard Business Review Press, 2016. p. 228.

Khan Academy. 2011. YouTube - Logarithmic scale | Logarithms | Algebra II | Khan Academy. [Online] August 26, 2011. [Cited: April 23, 2022.] https://www.youtube.com/watch?v=sBhEi4L91Sg.

Kirk, Andy. 2016. *Data Visualisation: A Handbook for Data Driven Design.* Thousand Oaks, CA : SAGE Publications Ltd., 2016.

Knaflic, Cole N. 2015. *Storytelling with Data.* Hoboken, New Jersey : John Wiley & Sons, 2015.

Lima, Manuel. 2017. *The Book of Circles.* Hudson, NY : Princeton Architectural Press, 2017.

Munzner, Tamara. 2015. *Visualization Analysis & Design.* Boca Raton, FL : Taylor & Francis Group, 2015.

Rosenberg, Daniel and Grafton, Anthony. 2012. *Cartographies of Time.* New York, NY : Princeton Architectural Press, 2012.

The scale of COVID-19 graphs affects understanding, attitudes, and policy preferences. **Romano, Alessandro, et al. 2020.** 1-13, London : John Wiley & Sons Ltd., 2020.

Tufte, Edward R. 2001. *The Visual Display of Quantitative Information.* Cheshire, Conneticut : Graphics Press LLC, 2001.