



Digital Scientific Illustration & Design

Jacopo Niccolò Cerasoni

MPI-SHH
SUMMER SCHOOL
2021

Doorway
to Human History

Lecture Structure



SUMMER
SCHOOL
2021



Doorway to
Human
History



1. Framework of Scientific Imaging

1. Files, Tools, Stroke, Text & Formats
2. Basic 3s of Figure Design
3. Color Theory

2. Data Visualization

1. Basic and complex visualization techniques
2. Uncommon visualization techniques

3. Cartography

1. Coordinate systems
2. Softwares and applications
3. Datasets and tools

4. Photography and Digital Processing

1. Basics of photography
2. Digital processing

5. Beautification

1. Practical examples



Framework of Scientific Imaging



SUMMER
SCHOOL
2021



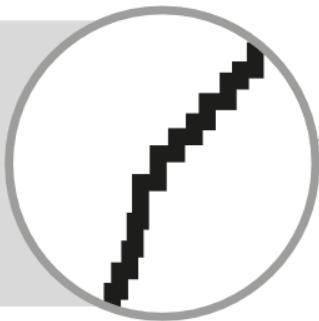
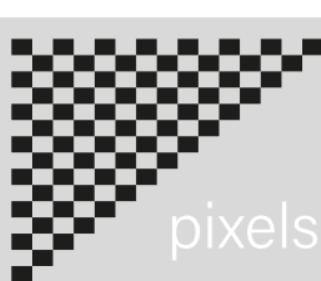
Doorway to
Human
History



Raster



Pixels arranged to form an image



Resolution constraints

Complex colour blends

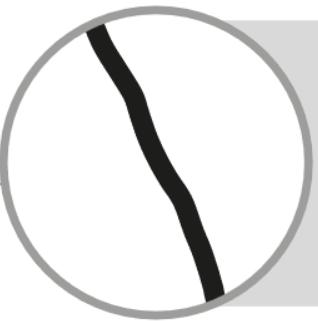
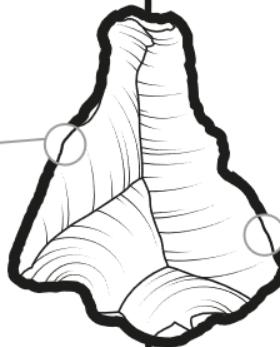
Large file sizes

Detailed editing



Vector

Paths dictated by mathematical formulas



paths

Small file sizes

Infinitely scalable

Difficult colour blends

Precise path editing



Tools

The image displays a variety of tools and features from the Adobe Illustrator interface:

- PEN TOOL** and **CURVATURE TOOL**: Tools for creating vector paths.
- SHAPE TOOLS**: A panel showing icons for square, oval, hexagon, and star shapes.
- CLIPPING MASK**: A tool for creating masked areas.
- OPACITY**: A tool for adjusting the transparency of selected objects.
- PATHFINDER**: A panel showing four modes: ADD, SUBTRACT, INTERSECT, and EXCLUDE.
- WIDTH TOOL**: A tool for adjusting stroke widths.
- ARRANGE**: A tool for managing object stacking order.
- EYEDROPPER**: A tool for picking colors from existing objects.
- TRANSFORM**: A panel showing icons for scale, rotate, reflect, and shear transformations.
- CLICK + ALT/OPTION = DRAG (COPY AND PASTE)**: A note about keyboard shortcuts.
- CREATE/MOVE/CLICK + SHIFT = MULTI-SELECT/REGULAR/SQUARE ACTIONS**: Another note about keyboard shortcuts.
- WIDTH PROFILE/TOOL**, **STROKE DASHED LINES**, and **FILLS**: Panels showing options for stroke styles, dashed lines, and fill patterns.



Stroke & Text

Every journal has set ranges for **stroke** and **character** sizes

Always double-check

Print figures to check legibility

5pt	_____
4pt	_____
3pt	_____
2pt	_____
1pt	_____
0.75pt	_____
0.5pt	_____
0.25pt	_____



Lorem ipsum 14pt

Lorem ipsum 12pt

Lorem ipsum 10pt

Lorem ipsum 8pt

Lorem ipsum 6pt

Lorem ipsum 4pt



Formats

RASTER

JPG/JPEG

- standard and default format of many softwares

Tagged Image File (TIF/TIFF)

- Large raster file which doesn't loose quality
- Uses "lossless" compression
- Not to use on web, ideal for printing

Raw Image Format (RAW; CR2, CRW, NEF, PEF...)

- Least processed image type
- No processing or loss of small visual details
- Largest files type

VECTOR

Portable Network Graphics (PNG)

- Ideal for interactive docs and web pages, but NOT printing
- Can be edited with no loss of quality ("lossless"), but low resolution only

Portable Document Format (PDF)

- Ideal for printing documents and exchanging electronic documents (manuscripts, forms, images, web pages...)
- Can be saved into multi-layer format, ideal for saving images and figures while keeping original hierarchy layer



The 3 basics of figure design

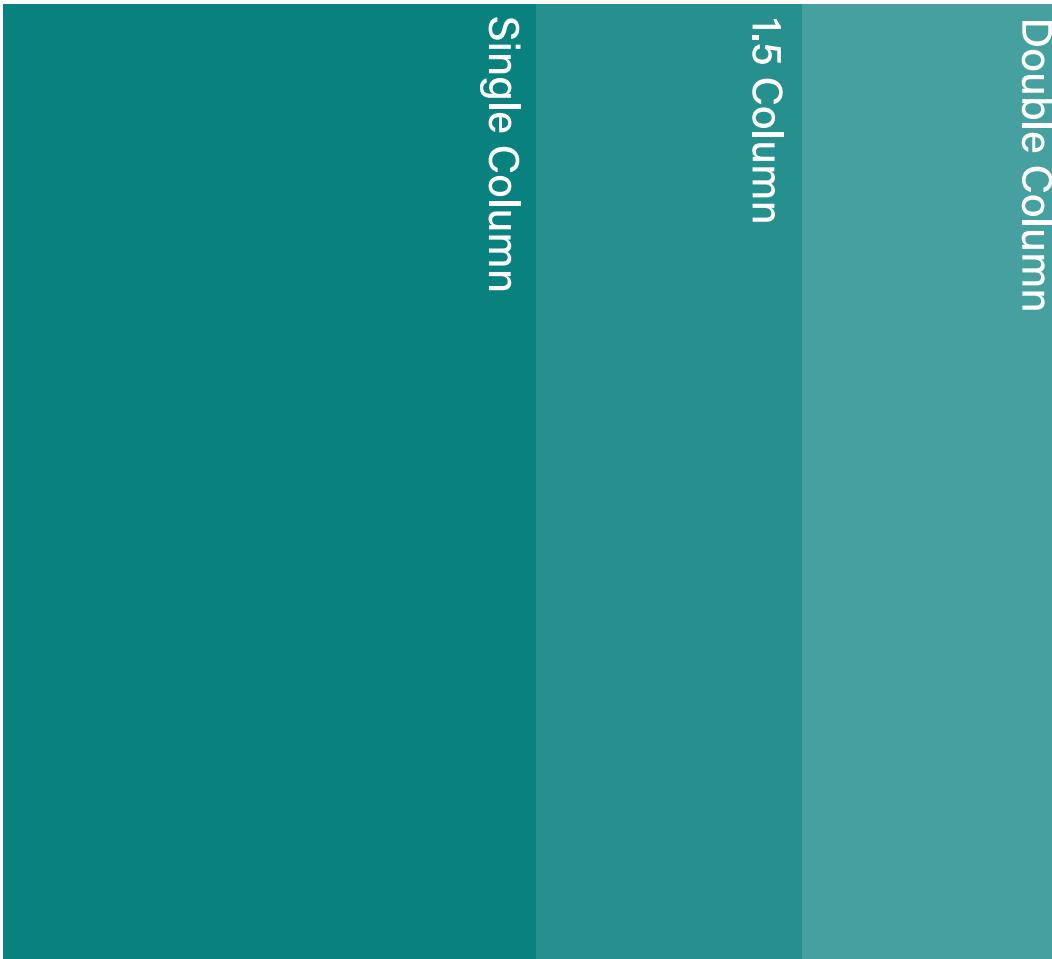
Simplicity

Direction

Consistency

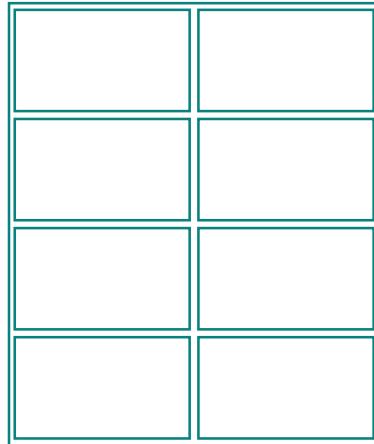


Direction

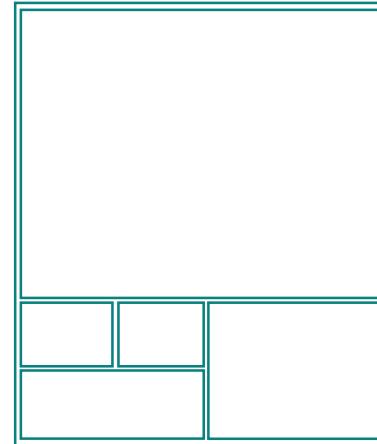


Direction

Recurrent

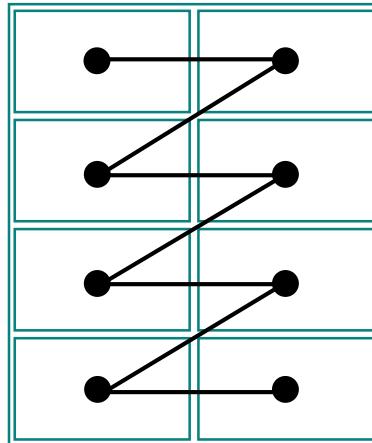


Eccentric

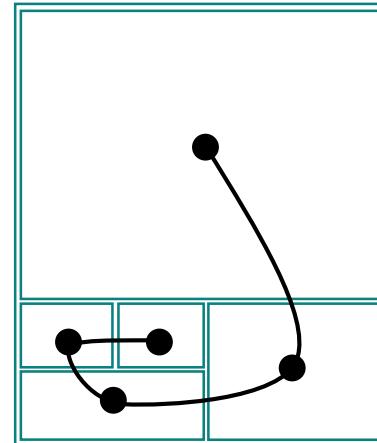


Direction

Recurrent



Eccentric



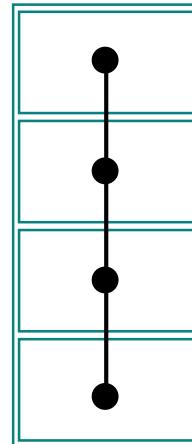
Quantitative data

Chronological graphs

Multi variable analyses

Repetitions

Comparable plot graphs



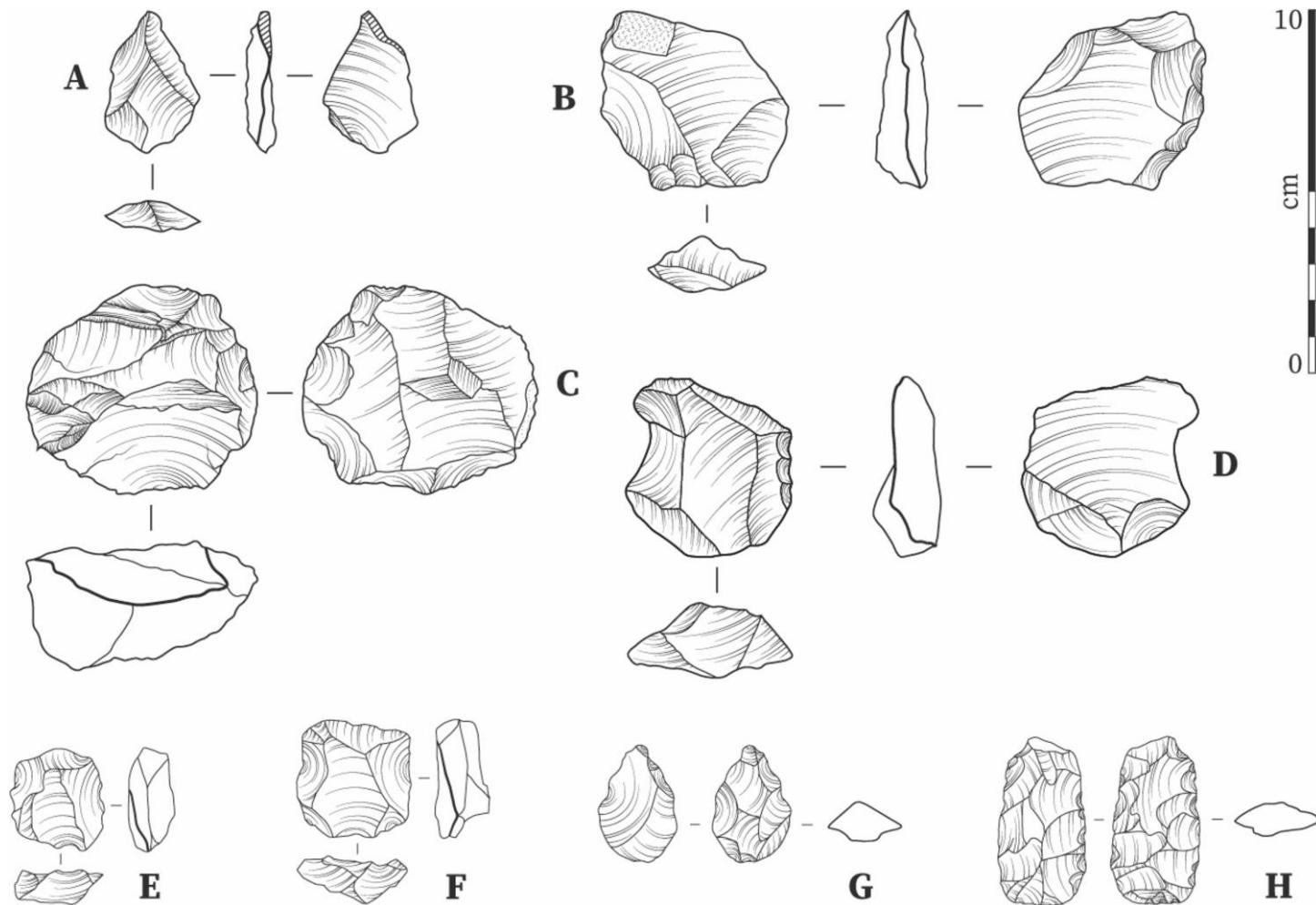
Qualitative data

Maps

Contextual
(site/stratigraphy/landscape...)



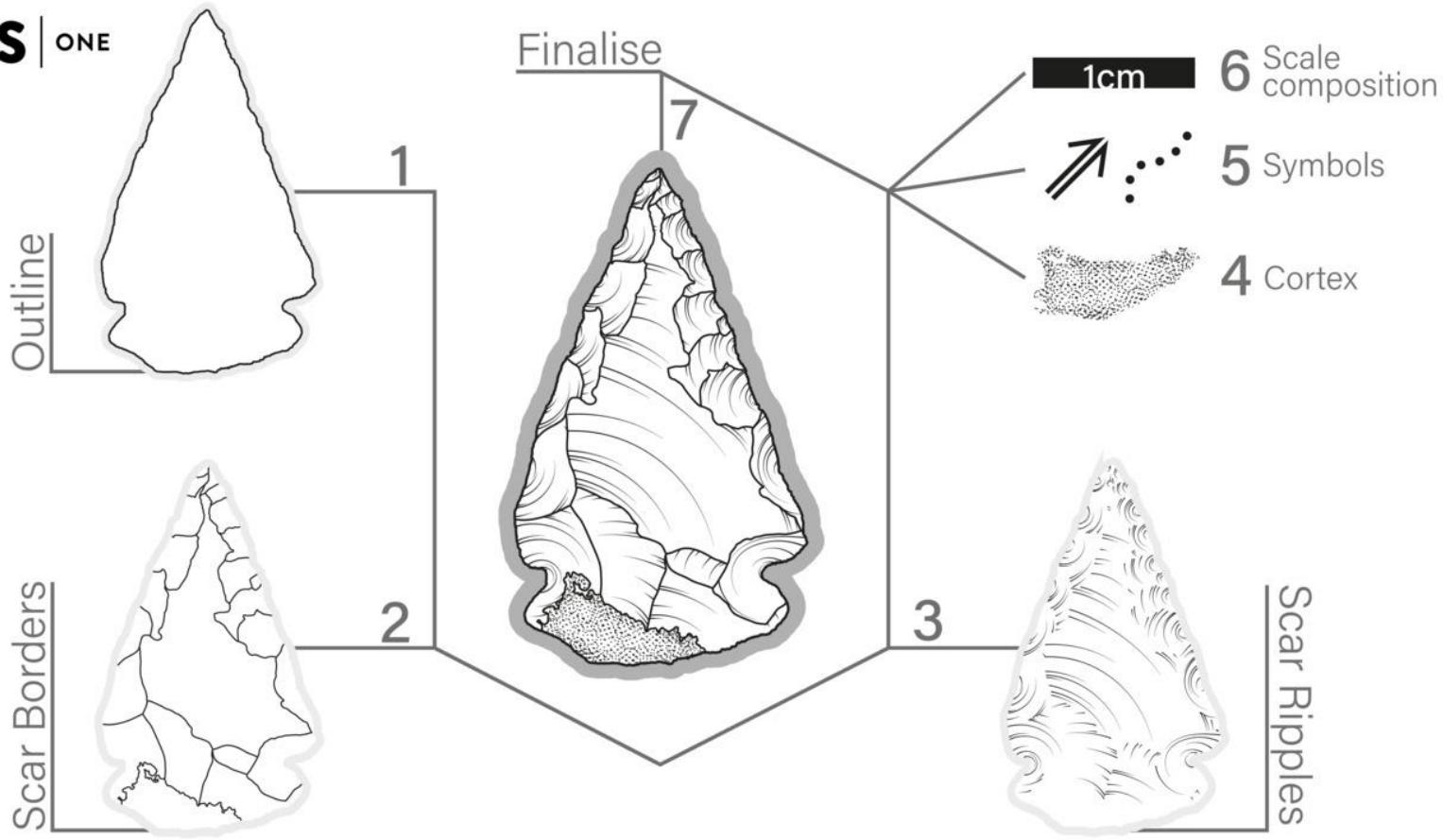
Direction



'STIVA' METHOD

stone Tools Illustrations with Vector Art

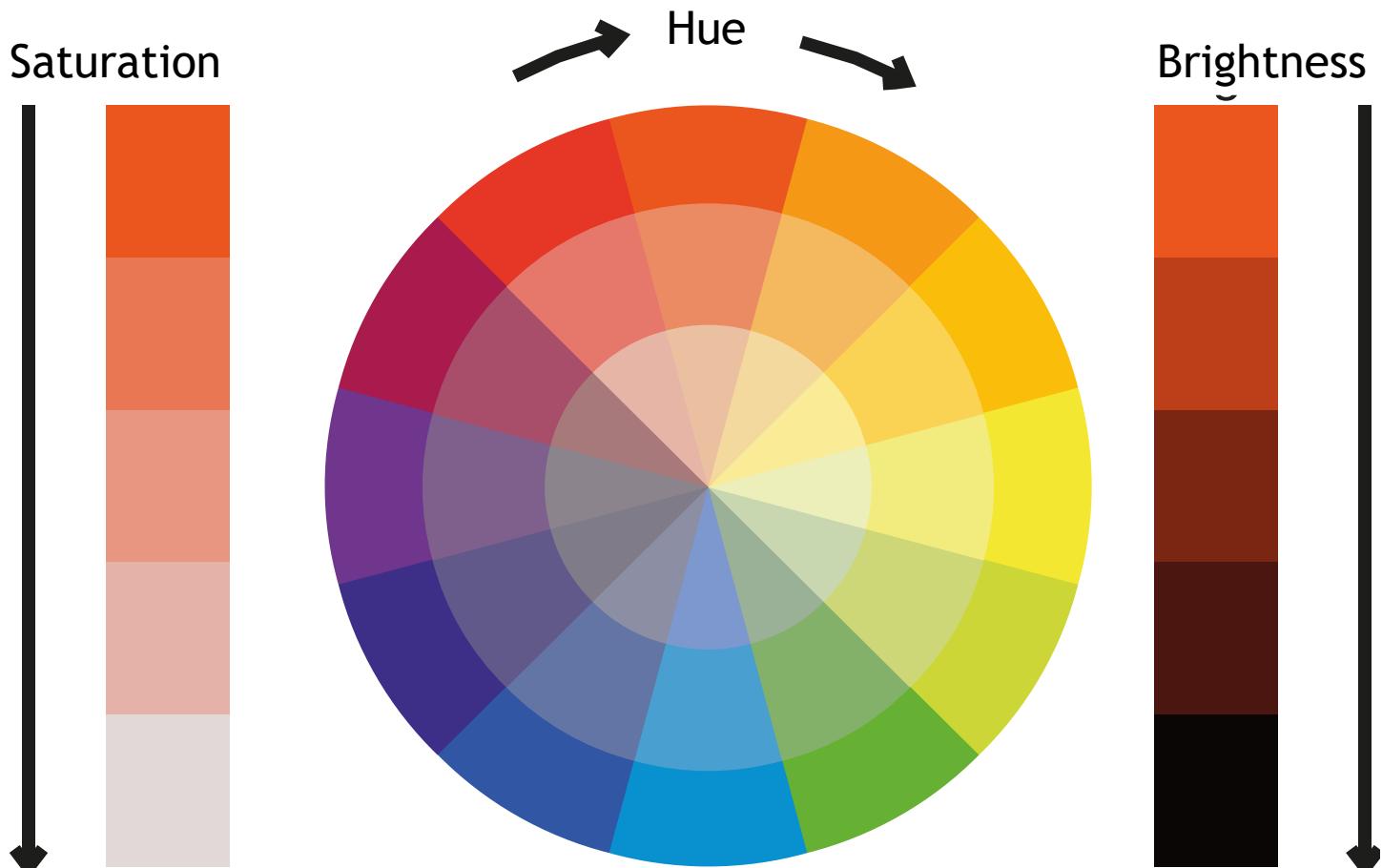
PLOS ONE



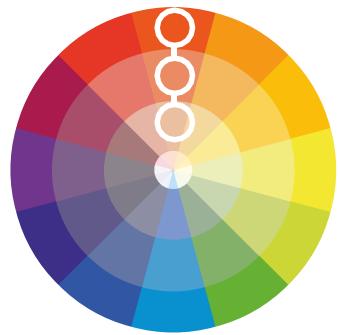
Color Theory



Color Theory



Color Theory



monochromatic



analogous



complementary



split complementary



triadic



tetradic



quadratic



Color Theory

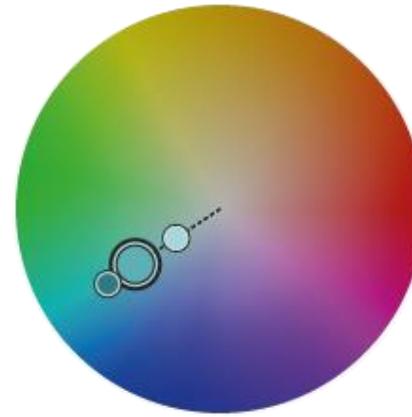
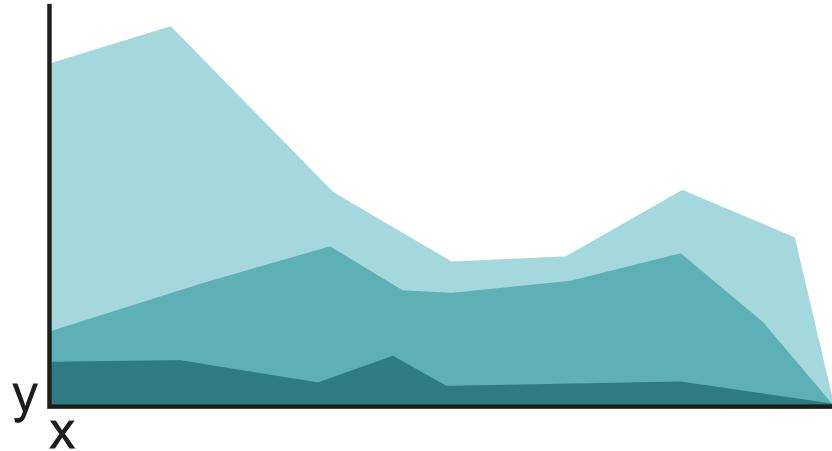
Fewer colors are always better - MAX 5 colours!

Keep consistent palettes between figures of same manuscript

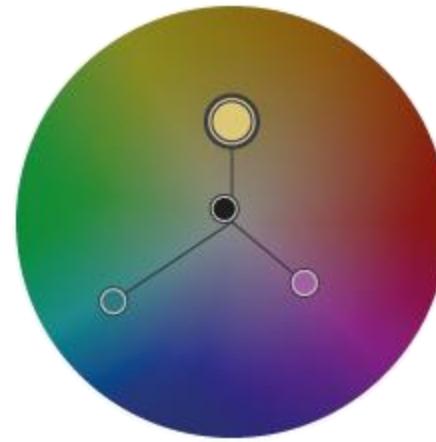
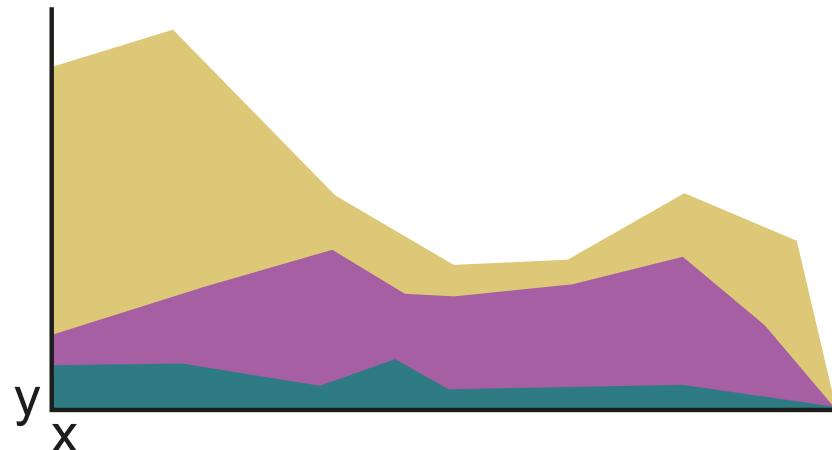
What relationship between the data do you want to show?



Color Theory - What do you want to show?



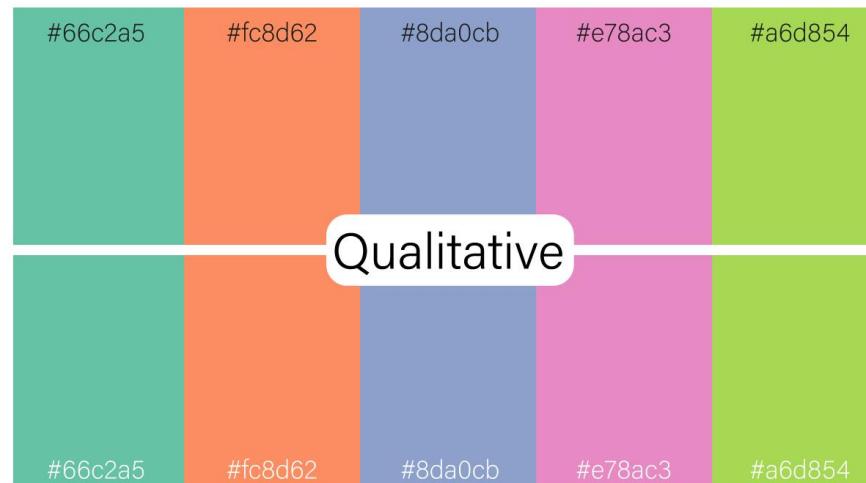
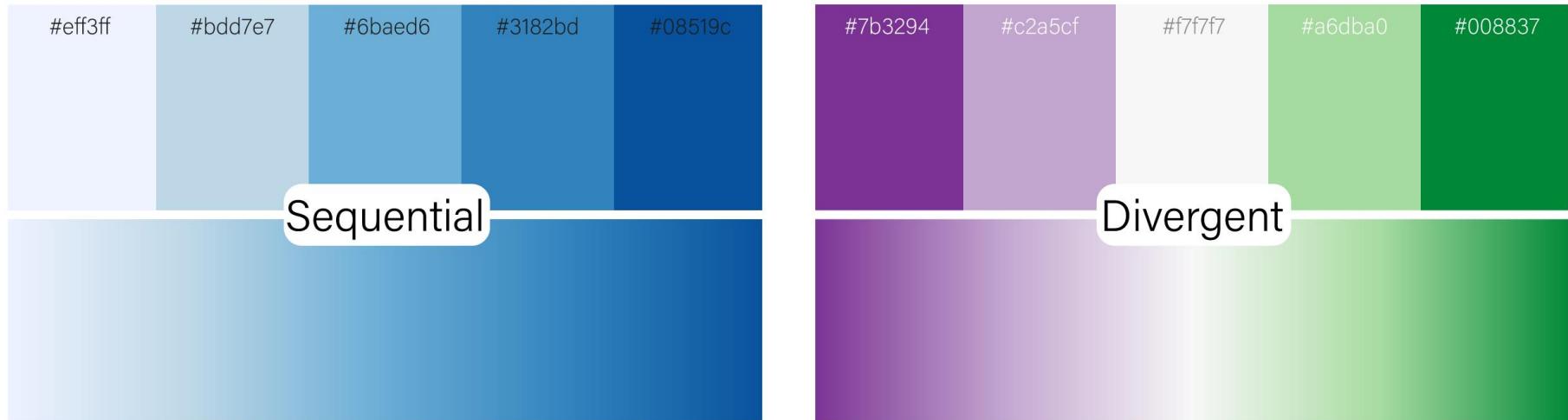
Similarities



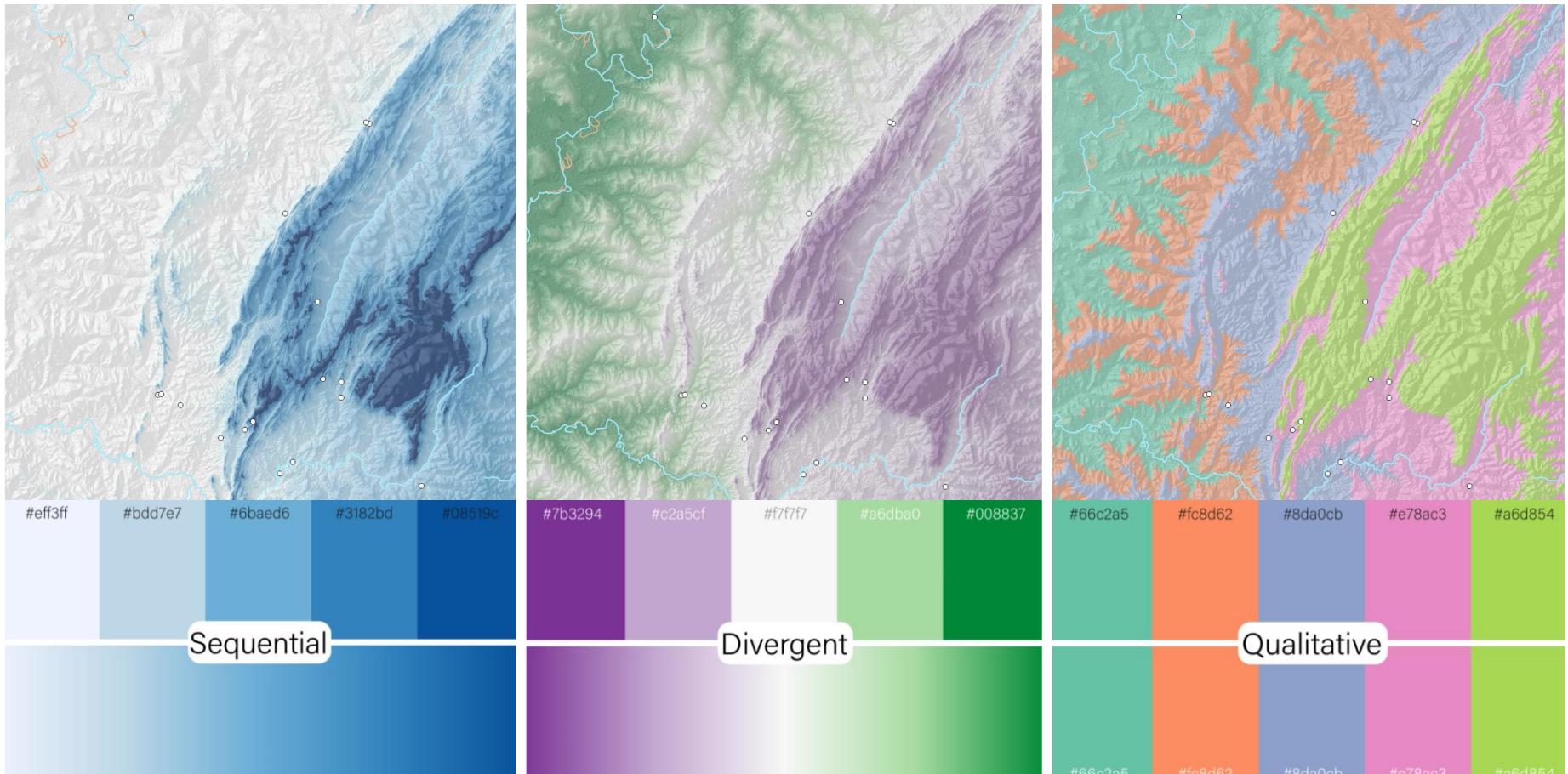
Differences



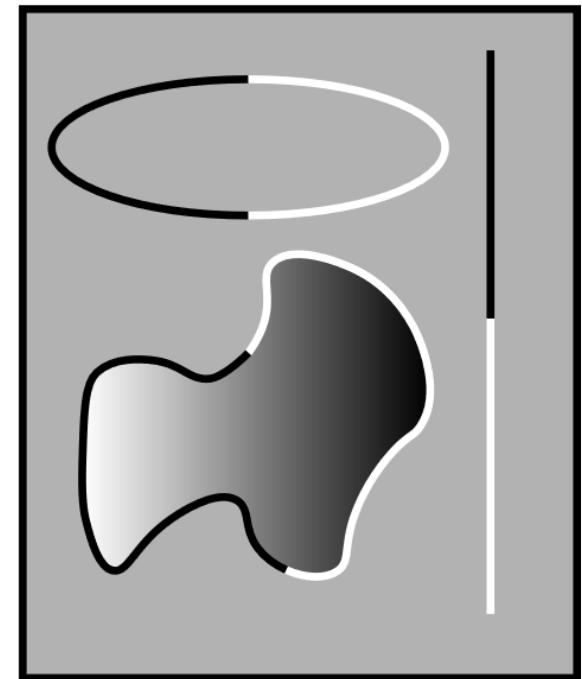
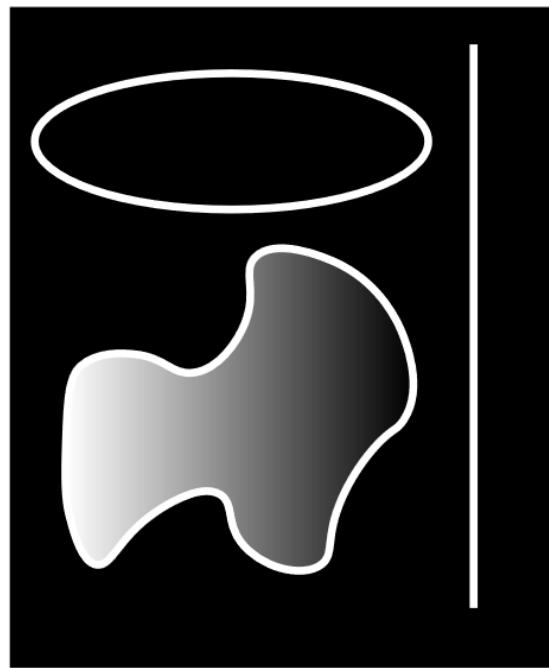
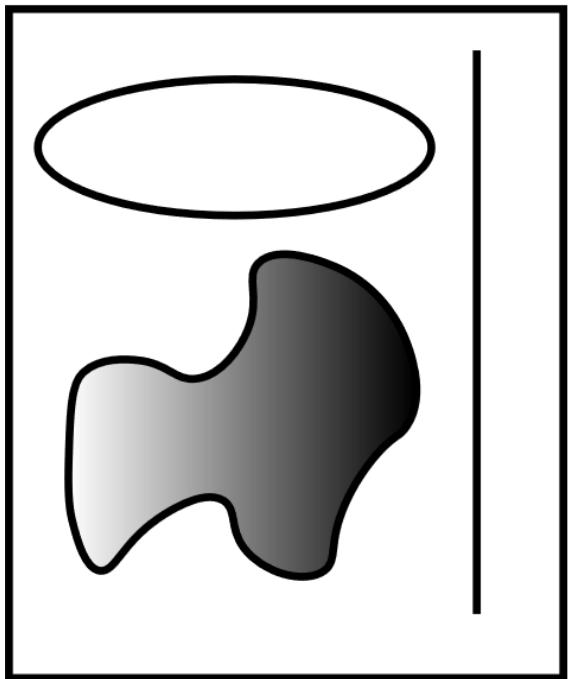
Color Theory - Gradients



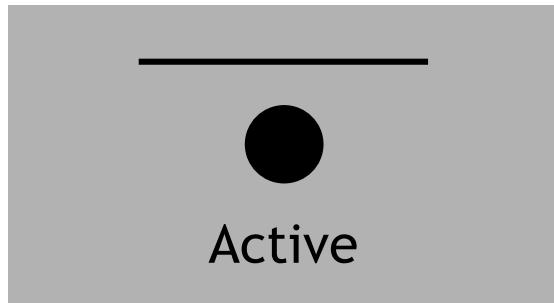
Color Theory - Gradients



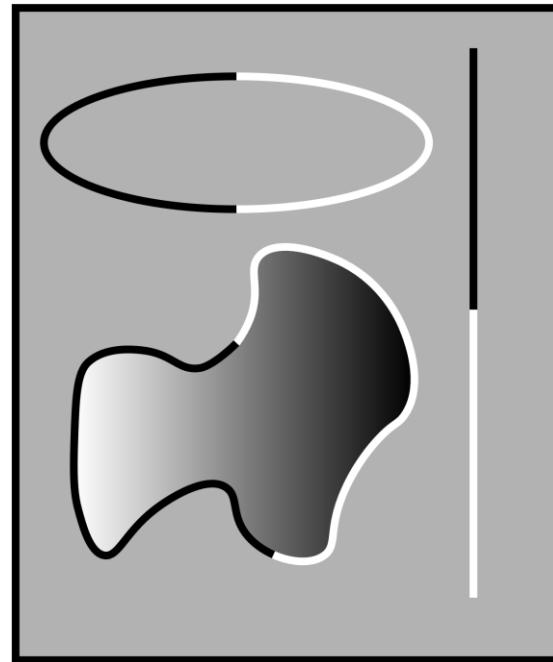
Color Theory - Contrast



Color Theory - Contrast



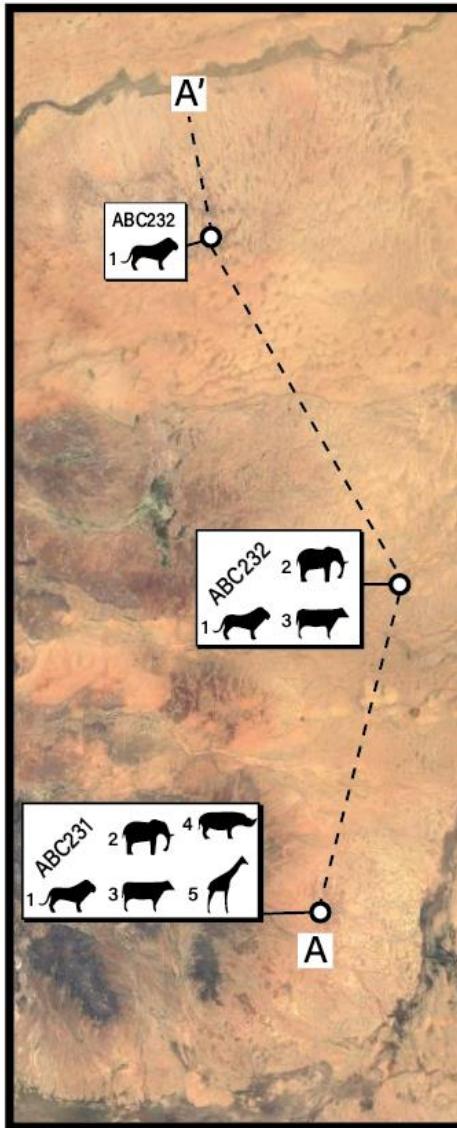
Active



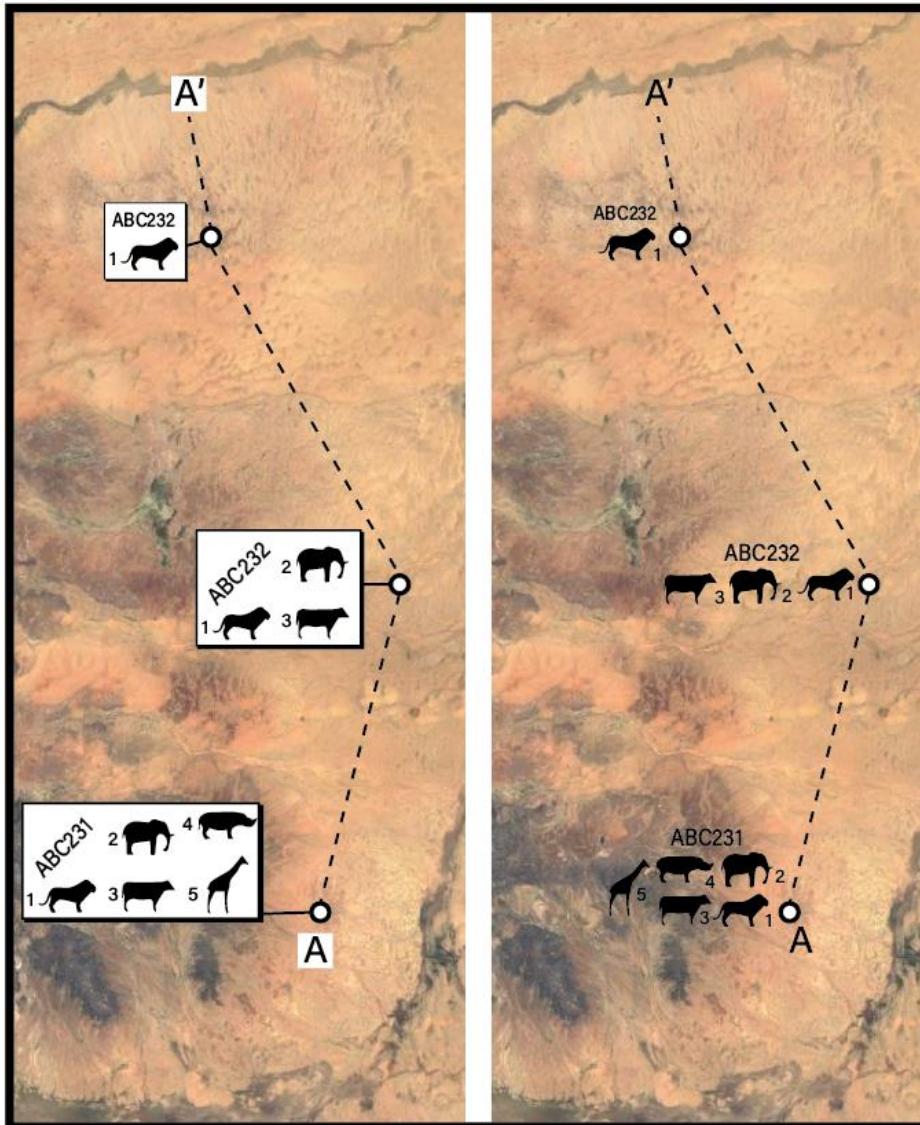
Passive



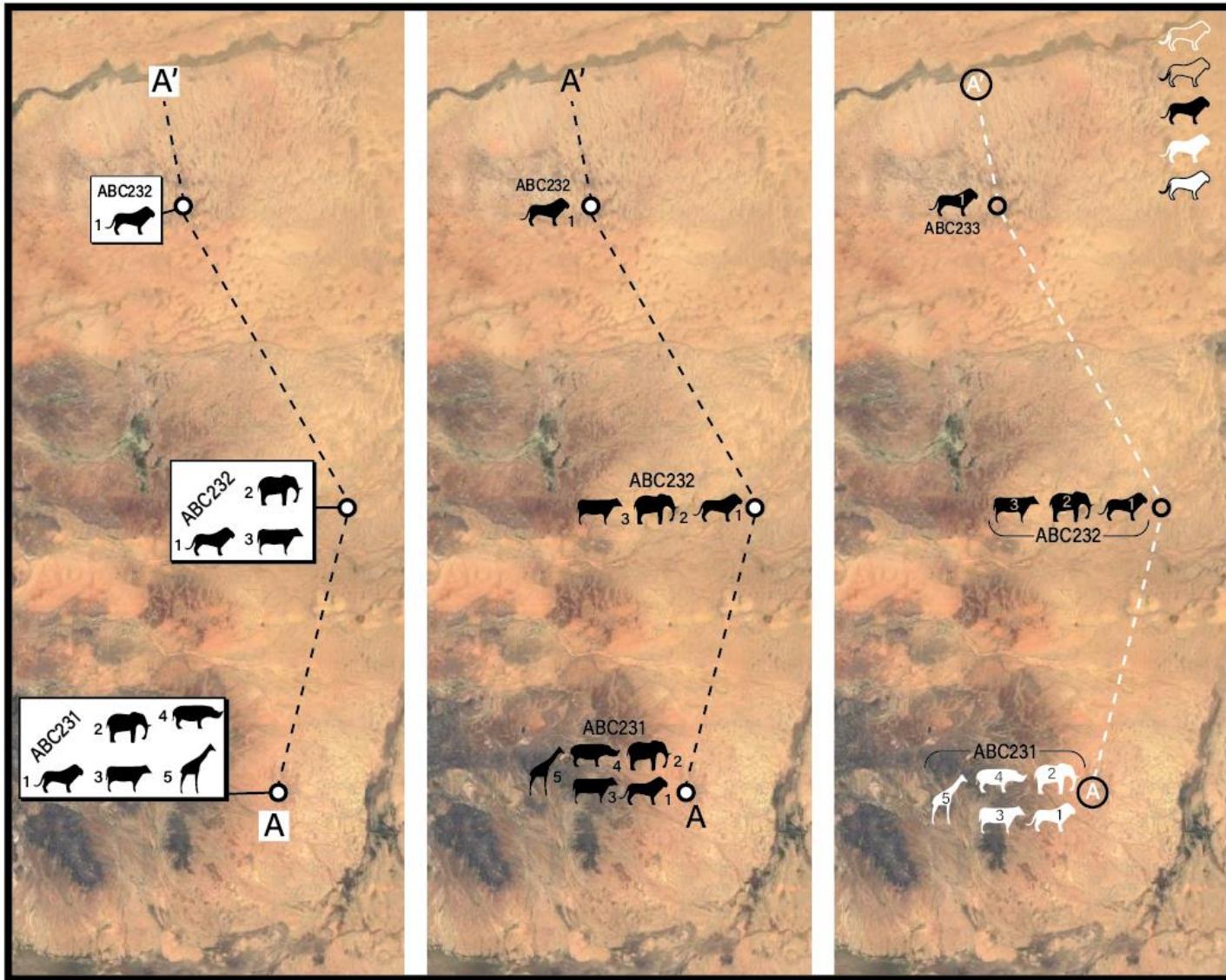
Color Theory - Contrast



Color Theory - Contrast



Color Theory - Contrast



Data visualization



SUMMER
SCHOOL
2021



Doorway to
Human
History



Visual vocabulary

Designing with data

There are so many ways to visualise data – how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

FT graphic: Alan Smith; Chris Campbell; Ian Bell; Li Fausto; Graham Ferrier; Billy Elshorbagy; Shannon Paul; McCullum; Martin Ribeiro
Inspired by the Graphic Continuum by Jon Scherbach and Trevor Ribeiro



ft.com/vocabulary

Deviation

Emphasise variations (+/-) from a fixed reference point. Typically the reference point is zero but it can also be a target or a mean value. Can also be used to show sentiment (positive/negative).

Example FT uses:
Inflation and unemployment, income and life expectancy

Diverging bar
A simple standard bar chart that can handle both negative and positive magnitude values.

Diverging stacked bar
Perfect for presenting survey results which involve sentiment (eg. disagree/agree).

Spine
Splits a single value into two contrasting components (eg. male/female).

Surplus/deficit filled line
The shaded area of these charts allows a balance to be shown – either between a baseline or between two series.

Bubble
Like a scatterplot but adds additional detail by sizing the circles according to a third variable.

XY heatmap
A good way of showing the pattern between 2 categories of data, less effective at showing fine differences in amounts.

Slope
Perfect for showing how ranks have changed over time or very different categories.

Lollipop
Lollipops are great for showing the data value than standard bars because they can also show rank and value effectively.

Bump
Effective for showing changing rankings and data values. For large datasets, consider grouping lines using color.

Population pyramid
A standard way for showing the age and sex distribution of a population distribution; effectively back-to-back histograms.

Cumulative curve
A good way of showing how unequal a distribution is; x is always cumulative (increasing) while y is always a measure.

Frequency polygon
For displaying multiple distributions of data. Like a regular line chart, but the lines add to a maximum of 3 or 4 datasets.

Besswane
Used to emphasise individual points in a distribution. Points can be sized to an absolute value. Best with medium-sized datasets.

Vertical timeline
Presents time on the Y axis. Good for displaying sequential time series that work especially well when scrolling on mobile.

Seismogram
Another alternative to the circle timeline for showing series where there are big variations in the data.

Streamgraph
A type of area chart: used when seeing the relationships over time is more important than individual values.

Correlation

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume that the correlation they see to be causal (i.e. one causes the other).

Example FT uses:
Wealth, deprivation, league tables, constituency election results

Scatterplot
The standard way to show the relationship between two continuous variables, each of which has its own axis.

Column + line time
A good way of showing the relationship between an amount (column) and a rate (line).

Connected scatterplot
Usually used to show the relationship between 2 variables has changed over time.

Bubble
Like a scatterplot but adds additional detail by sizing the circles according to a third variable.

XY heatmap
A good way of showing the pattern between 2 categories of data, less effective at showing fine differences in amounts.

Slope
Perfect for showing how ranks have changed over time or very different categories.

Lollipop
Lollipops are great for showing the data value than standard bars because they can also show rank and value effectively.

Bump
Effective for showing changing rankings and data values. For large datasets, consider grouping lines using color.

Population pyramid
A standard way for showing the age and sex distribution of a population distribution; effectively back-to-back histograms.

Cumulative curve
A good way of showing how unequal a distribution is; x is always cumulative (increasing) while y is always a measure.

Frequency polygon
For displaying multiple distributions of data. Like a regular line chart, but the lines add to a maximum of 3 or 4 datasets.

Besswane
Used to emphasise individual points in a distribution. Points can be sized to an absolute value. Best with medium-sized datasets.

Vertical timeline
Presents time on the Y axis. Good for displaying sequential time series that work especially well when scrolling on mobile.

Seismogram
Another alternative to the circle timeline for showing series where there are big variations in the data.

Streamgraph
A type of area chart: used when seeing the relationships over time is more important than individual values.

Ranking

Show where an item's position in an ordered list is more important than its absolute or relative value. Don't afraid to highlight the points of interest.

Example FT uses:
Inflation and unemployment, income and life expectancy

Ordered bar
Standard bar charts display the ranks of values much more easily when sorted into order.

Dot plot
See above.

Ordered proportional symbol
Use when there are big differences in the size of values and/or seeing fine differences between data is not so important.

Dot strip plot
Good for showing the relationship over time between an amount (dot) and a rate (strip).

Barcode plot
Like dot strip plots, good for displaying all data in a table, especially when highlighting individual values.

Slope
Good for showing changing data as long as the data can be separated into 2 or 3 points without missing a key part of story.

Borplot
Summarise multiple distributions at once by showing the median (Centre) and range (the ends) of the data.

Violin plot
Similar to a box plot but more effective with complex distributions (skewness, outliers, similar to box plot but can be summarised with simple average).

Lollipop
Lollipops are great for showing the data value than standard bars because they can also show rank and value effectively.

Bump
Effective for showing changing rankings and data values. For large datasets, consider grouping lines using color.

Population pyramid
A standard way for showing the age and sex distribution of a population distribution; effectively back-to-back histograms.

Cumulative curve
A good way of showing how unequal a distribution is; x is always cumulative (increasing) while y is always a measure.

Frequency polygon
For displaying multiple distributions of data. Like a regular line chart, but the lines add to a maximum of 3 or 4 datasets.

Besswane
Used to emphasise individual points in a distribution. Points can be sized to an absolute value. Best with medium-sized datasets.

Vertical timeline
Presents time on the Y axis. Good for displaying sequential time series that work especially well when scrolling on mobile.

Seismogram
Another alternative to the circle timeline for showing series where there are big variations in the data.

Streamgraph
A type of area chart: used when seeing the relationships over time is more important than individual values.

Ranking

Show values in a dataset and how often they occur. The shape or 'skew' of a distribution can be a memorable way of thinking about the levels of uniformity or equality in the data.

Example FT uses:
Income distribution, population

Histogram
The standard way to show a statistical distribution - keep the gaps between columns small to highlight the 'shape' of the data.

Dot plot
A simple way of showing the change or range (minimum and maximum) between multiple categories.

Dot strip plot
Good for showing the relationship over time between an amount (dot) and a rate (strip).

Barcode plot
Like dot strip plots, good for displaying all data in a table, especially when highlighting individual values.

Slope
Good for showing changing data as long as the data can be separated into 2 or 3 points without missing a key part of story.

Borplot
Summarise multiple distributions at once by showing the median (Centre) and range (the ends) of the data.

Violin plot
Similar to a box plot but more effective with complex distributions (skewness, outliers, similar to box plot but can be summarised with simple average).

Lollipop
Lollipops are great for showing the data value than standard bars because they can also show rank and value effectively.

Bump
Effective for showing changing rankings and data values. For large datasets, consider grouping lines using color.

Population pyramid
A standard way for showing the age and sex distribution of a population distribution; effectively back-to-back histograms.

Cumulative curve
A good way of showing how unequal a distribution is; x is always cumulative (increasing) while y is always a measure.

Frequency polygon
For displaying multiple distributions of data. Like a regular line chart, but the lines add to a maximum of 3 or 4 datasets.

Besswane
Used to emphasise individual points in a distribution. Points can be sized to an absolute value. Best with medium-sized datasets.

Vertical timeline
Presents time on the Y axis. Good for displaying sequential time series that work especially well when scrolling on mobile.

Seismogram
Another alternative to the circle timeline for showing series where there are big variations in the data.

Streamgraph
A type of area chart: used when seeing the relationships over time is more important than individual values.

Change over Time

Give emphasis to changing trends. These can be short (intra-day) movements or extended series (over months or years). Choosing the correct time period is critical to make sure it is suitable for the reader.

Example FT uses:
Share price movements, economic time series, price movements, economic time series, sectoral changes in a market

Line
The standard way to show a changing time series. These can be short (intra-day) movements or extended series (over months or years).

Column
Columns work well for showing change over time, but don't beat with one time series of data at a time.

Bar
See above. Good when the data are not time series and labels have long category names.

Paired column
As per standard bar charts, but good for multiple series. Can become tricky to read with more than 2 series.

Pie
A common way of showing parts of a whole. But be aware that it's difficult to accurately compare the size of the segments.

Donut
Similar to a pie chart – but the centre can be a good way of showing data at the same time – as long as the data are not too complicated.

Treemap
Use for hierarchical part-to-whole relationships. Can be difficult to read when there are many small segments.

Proportional symbol
Use when there are big differences in the size of values and/or seeing change in components can be very difficult.

Isotype (pictogram)
Excellent solution in some instances – use numbers and symbols (numbers do not slice up to represent a decimal).

Lollipop
Lollipop charts draw more attention to the data value than standard bar charts – does not need to start at zero (but preferable).

Radar
A space-efficient way of showing values of multiple variables – but make sure the axes are organised in a way that makes sense to reader.

Parallel coordinate
An alternative to radar charts – the arrangement of the variables is important. Usually better with highlighting values.

Circle timeline
Good for showing discrete values of varying sizes across different categories (eg. earnings/revenue by continent).

Vertical timeline
Presents time on the Y axis. Good for displaying sequential time series that work especially well when scrolling on mobile.

Grouped symbol
An alternative to bar charts when data is grouped – can highlight data or highlight individual elements is useful.

Seismogram
Another alternative to the circle timeline for showing series where there are big variations in the data.

Streamgraph
A type of area chart: used when seeing the relationships over time is more important than individual values.

Magnitude

Show size comparisons. These can be relative (just being able to see large/lagged or absolute) need to see the difference in size. Choosing the correct time period is critical to make sure it is suitable for the reader.

Example FT uses:
Commodity production, market capitalisation, volumes in general

Stacked column/bar
A simple way of showing part-to-whole relationships but can be difficult to read with more than 5 components.

Marimekko
A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.

Flow map
For showing unidirectional movement across a map.

Contour map
For showing areas of equal value on a map. Can use deviation from the mean to include more information about the data (low/high).

Equalised cartogram
Converting each unit on a map to a square which can illustrate equal values.

Scaled cartogram (value)
Scaling and stretching map so that each area is sized according to a particular value.

Arc
A hemicycle, often used for visualising data that needs composition by number of seats.

Gridplot
Good for showing % information, they work best when used on whole numbers and are great for small multiple layout forms.

Venn
Generally only used for schematic representation.

Waterfall
Can be used for showing part-to-whole relationships where some of the components are negative.

Part-to-whole

Show how a single entry can be broken down into its component elements. If the reader's interest is solely in the size of the total, then this is not the best choice.

Example FT uses:
Political budgets, company structure, national election results

Basic choropleth (area/height)
The standard approach for putting data on a map – should always be applied after totals and use the entire base geography.

Propotional symbol (area/magnitude)
Use for totalling rates – be wary that small differences in data will be hard to see.

Waterfall
Designed to show the sequencing of data flows in a process, typically budgets. Can include +/- components.

Chord
A complex but powerful way of showing movement across a map.

Network
Used for choosing the sequencing of data flows in a process of varying types.

Spatial

Aside from location maps only used when precise locations or geographical patterns in data are more important to the reader than anything else.

Example FT uses:
Population density, natural disaster risk, locations, natural disaster risk, impact, catchment areas, regional budgets.

Satellite
Shows changes in flow from one location to at least one other; good for tracing the eventual outcome of a complex process.

Flow

Show the reader volumes or intensity of movement between two or more locations or conditions. These might be logical sequences or geographical locations.

Example FT uses:
Movement of funds, trade, migrants, living/demographic relationships graphs.

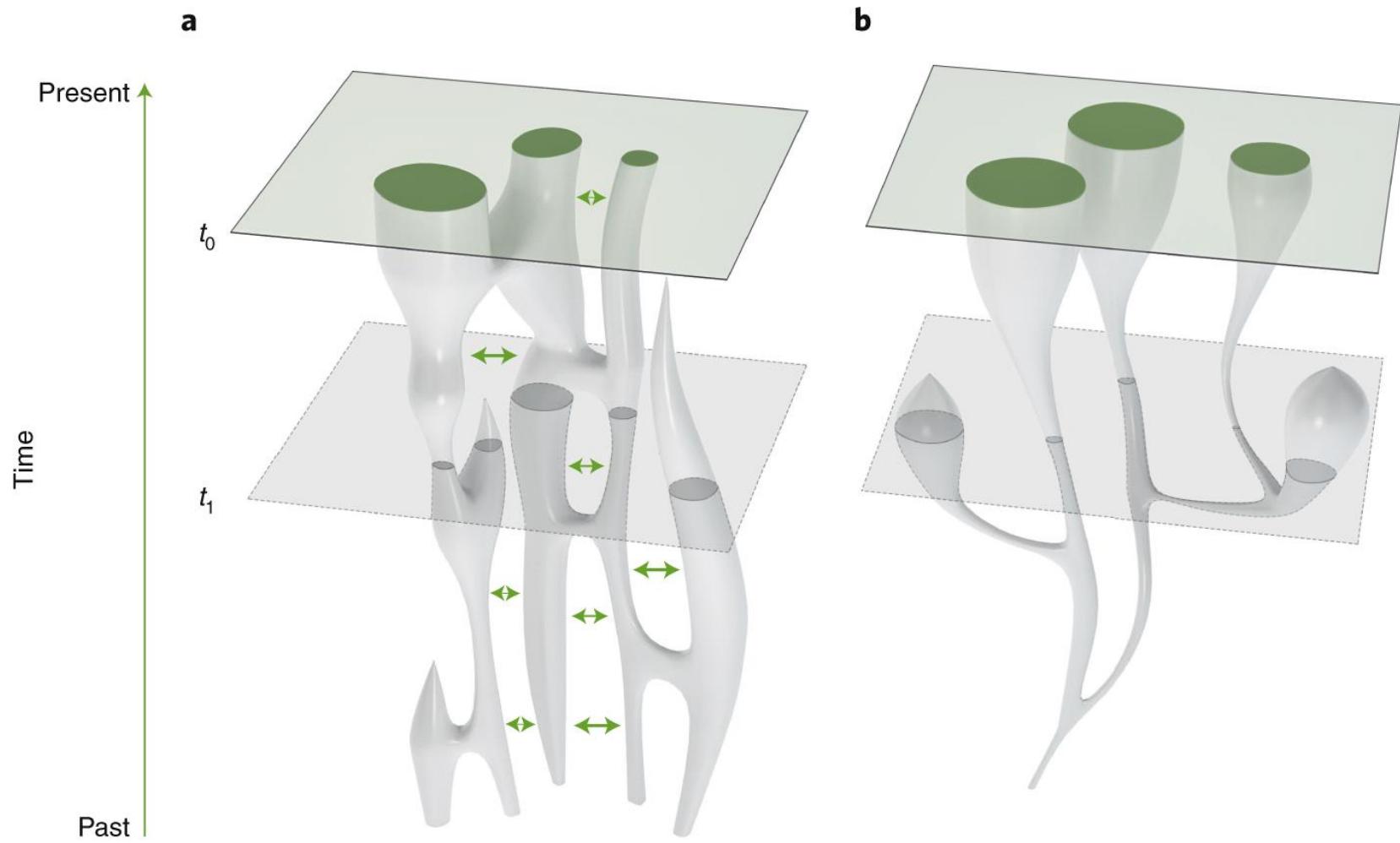
Waterfall
Designed to show the sequencing of data flows in a process, typically budgets. Can include +/- components.

Sankey
Shows changes in flow from one location to at least one other; good for tracing the eventual outcome of a complex process.

Waterfall
Designed to show the sequencing of data flows in a process, typically budgets. Can include +/- components.

Network
Used for choosing the sequencing of data flows in a process of varying types.

FT
© Financial Times 2018-2019
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License



Cartography



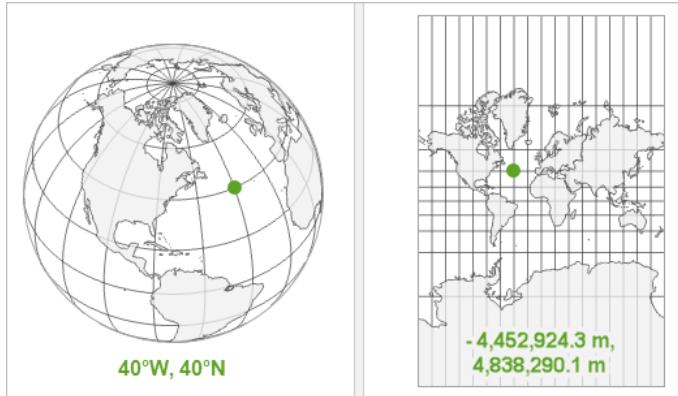
SUMMER
SCHOOL
2021



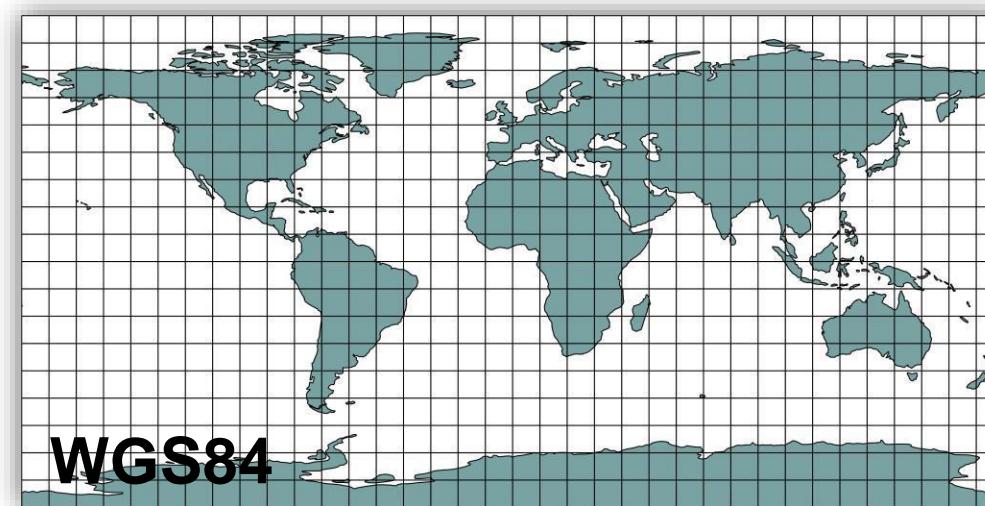
Doorway to
Human
History



Coordinate systems



A **projection** is the means by which you display the coordinate system and your data on a **flat surface**. (1)



Tools

The Basic 7 Geoprocessing Tools

Buffer

Clip

Merge

Dissolve

Intersect

Union

Erase

raster (GeoTiffs) vs. vector (Shapefiles)



Datasets



Natural Earth

Large scale data, 1:10m



Cultural Physical Raster

The most detailed. Suitable for making zoomed-in maps of countries and regions. Show the world on a large wall poster.

1:10,000,000
1" = 158 miles
1 cm = 100 km

Medium scale data, 1:50m



Cultural Physical Raster

Suitable for making zoomed-out maps of countries and regions. Show the world on a tabloid size page.

1:50,000,000
1" = 790 miles
1 cm = 500 km

Small scale data, 1:110m



Cultural Physical

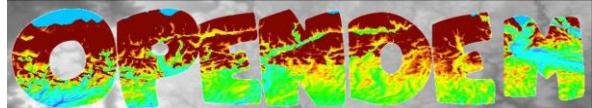
Suitable for schematic maps of the world on a postcard or as a small locator globe.

1:110,000,000
1" = 1,736 miles
1 cm = 1,100 km



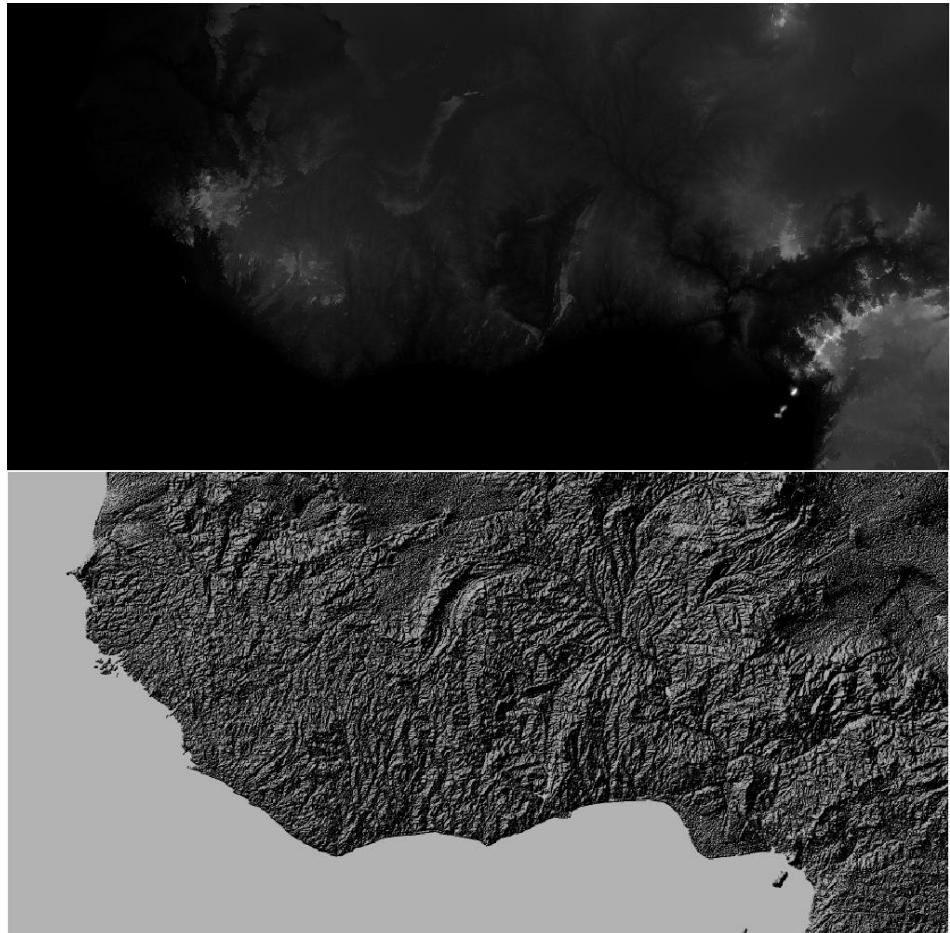
Datasets

Digital Elevation Models (DEMs)

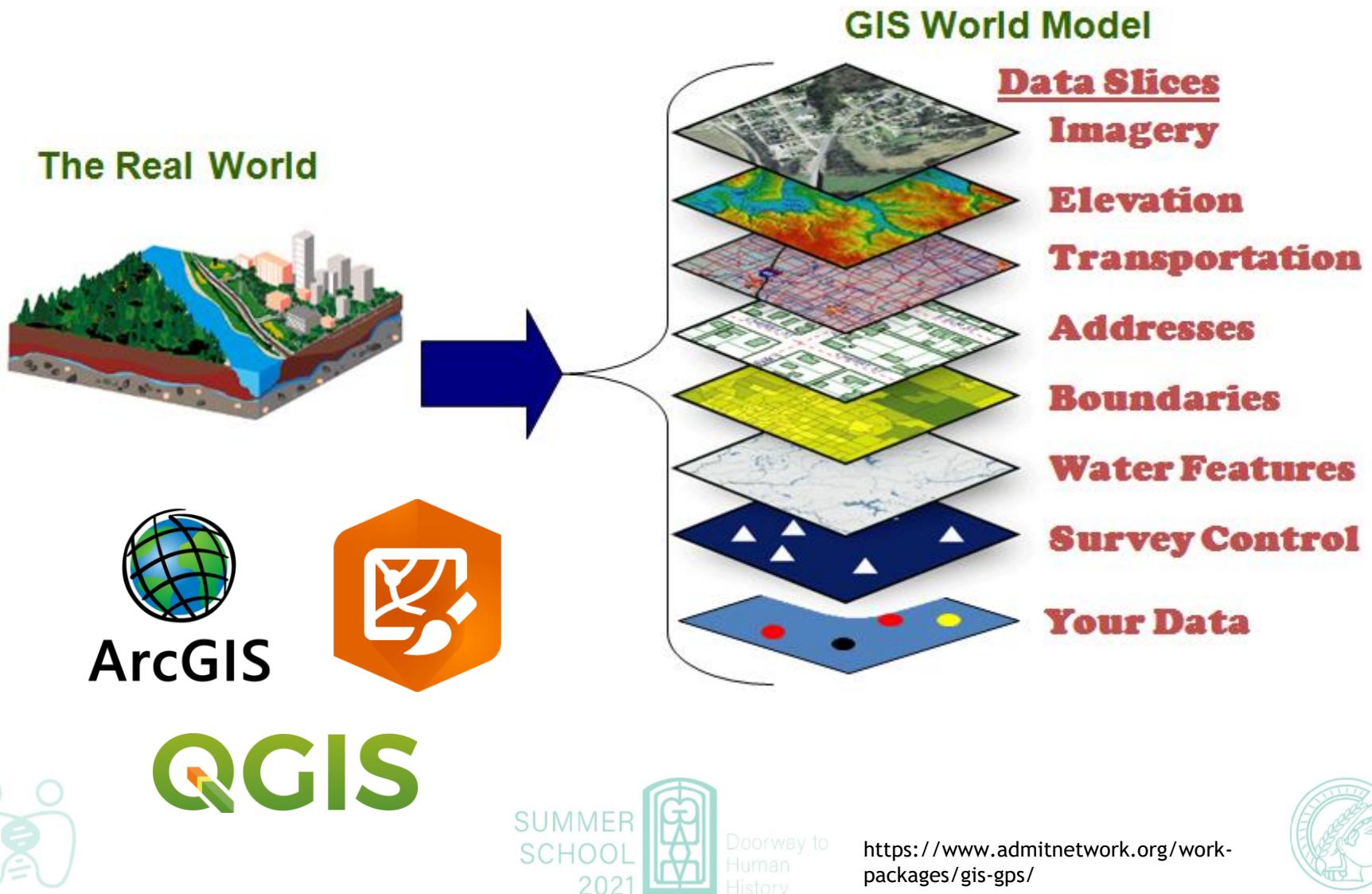


https://www.opendem.info/link_dem.html

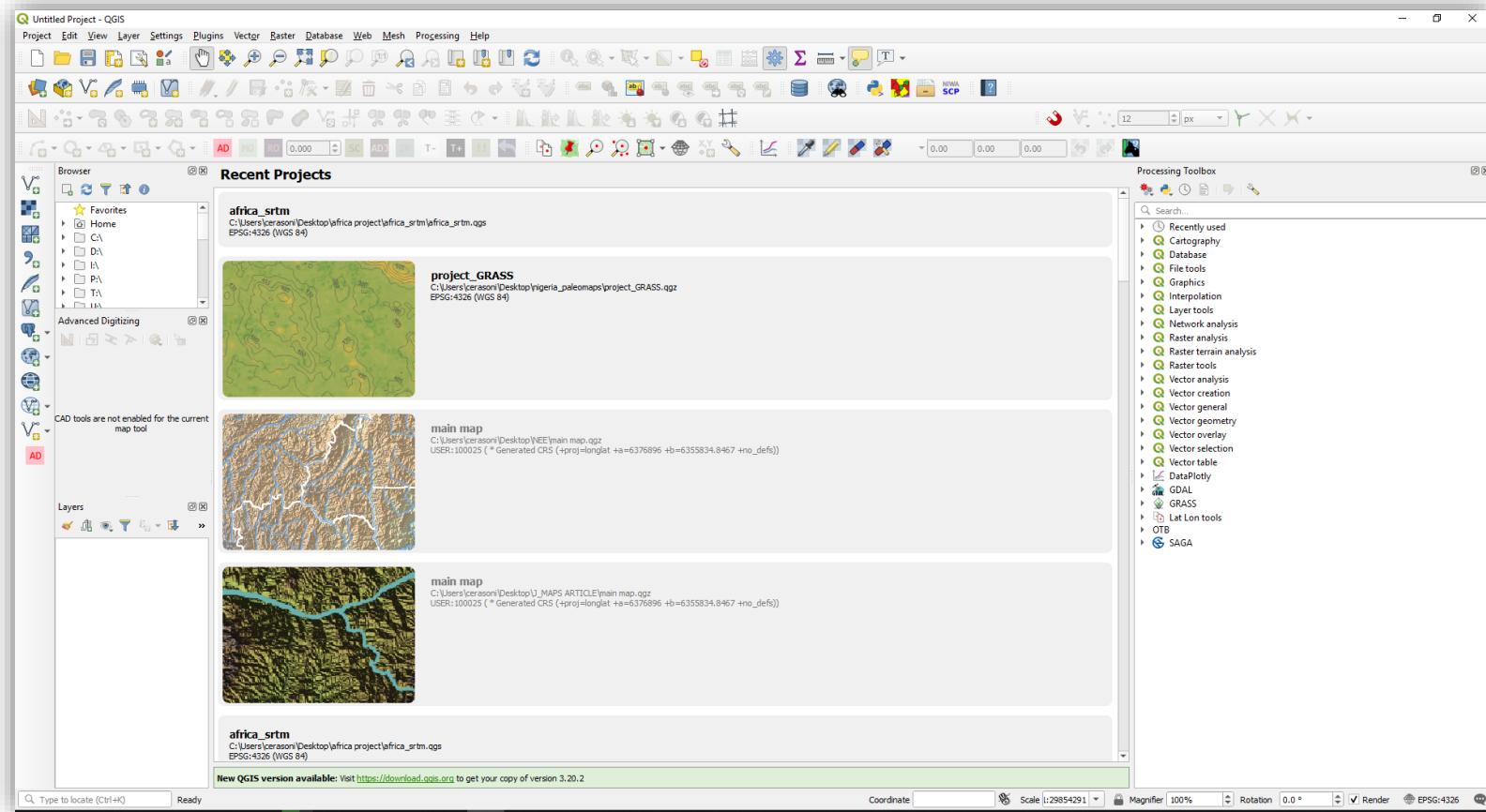
Jarvis, A., H.I. Reuter, A. Nelson, E. Guevara, 2008, Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database (<http://srtm.csi.cgiar.org>).



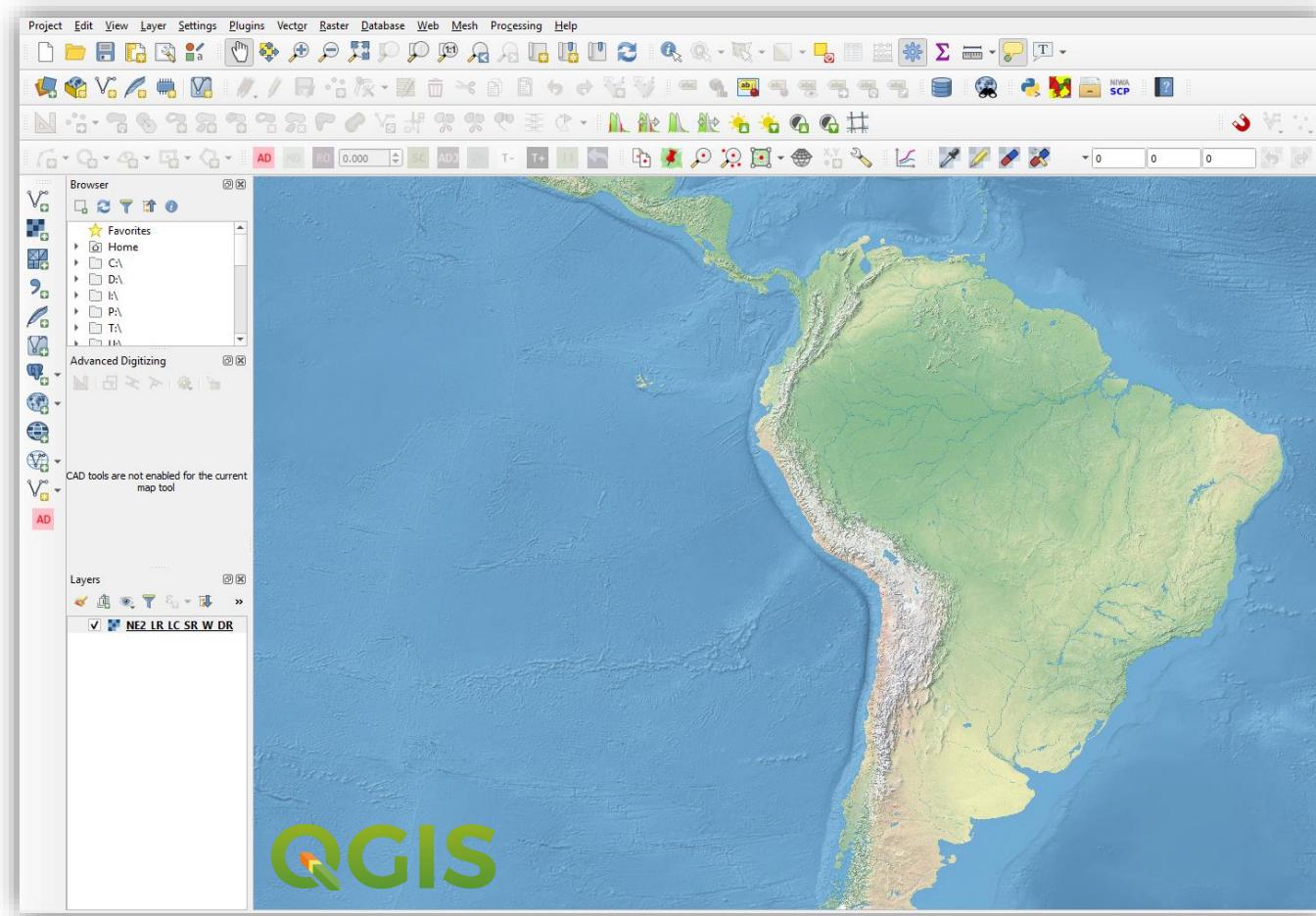
Softwares & Applications



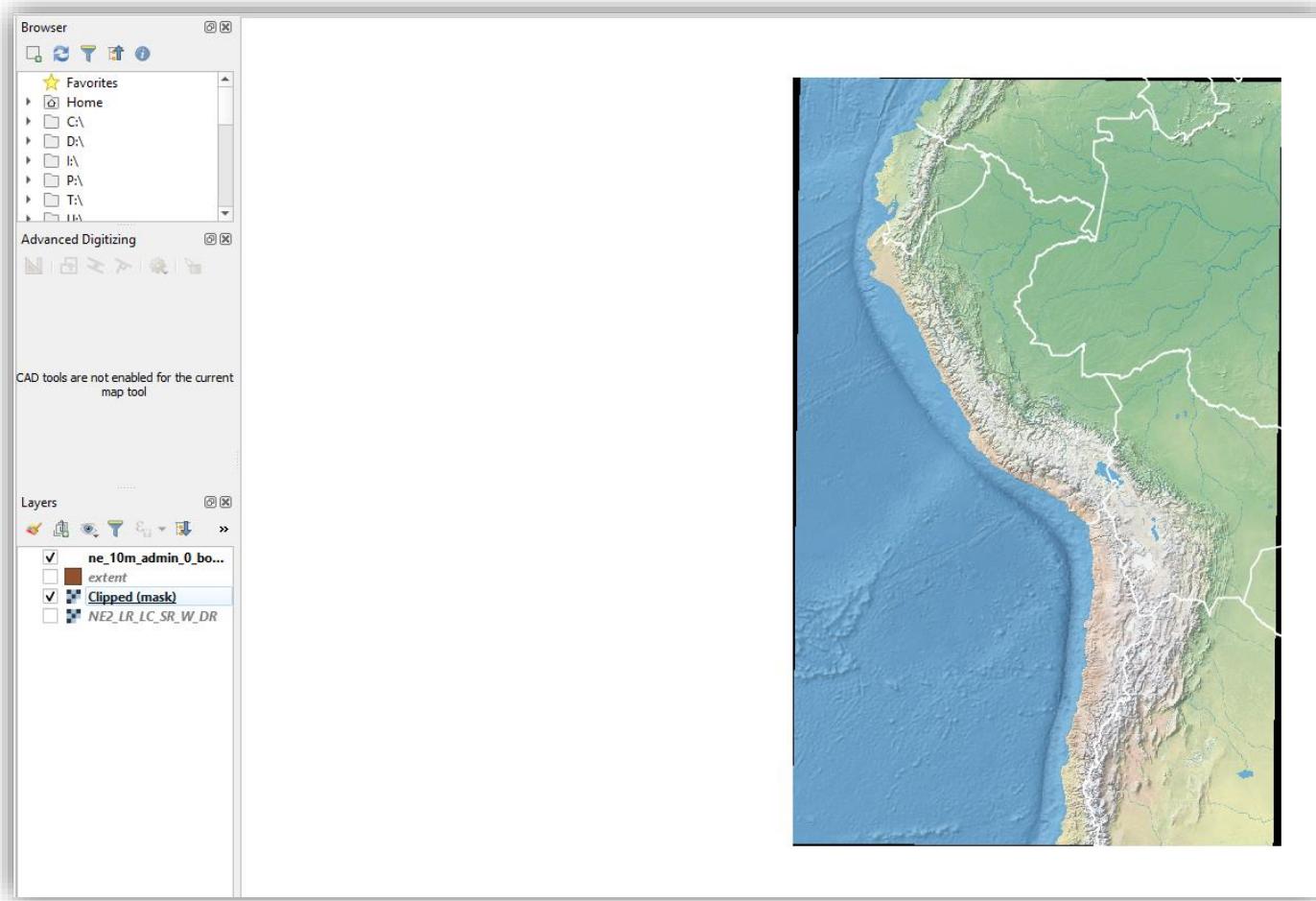
Softwares & Applications



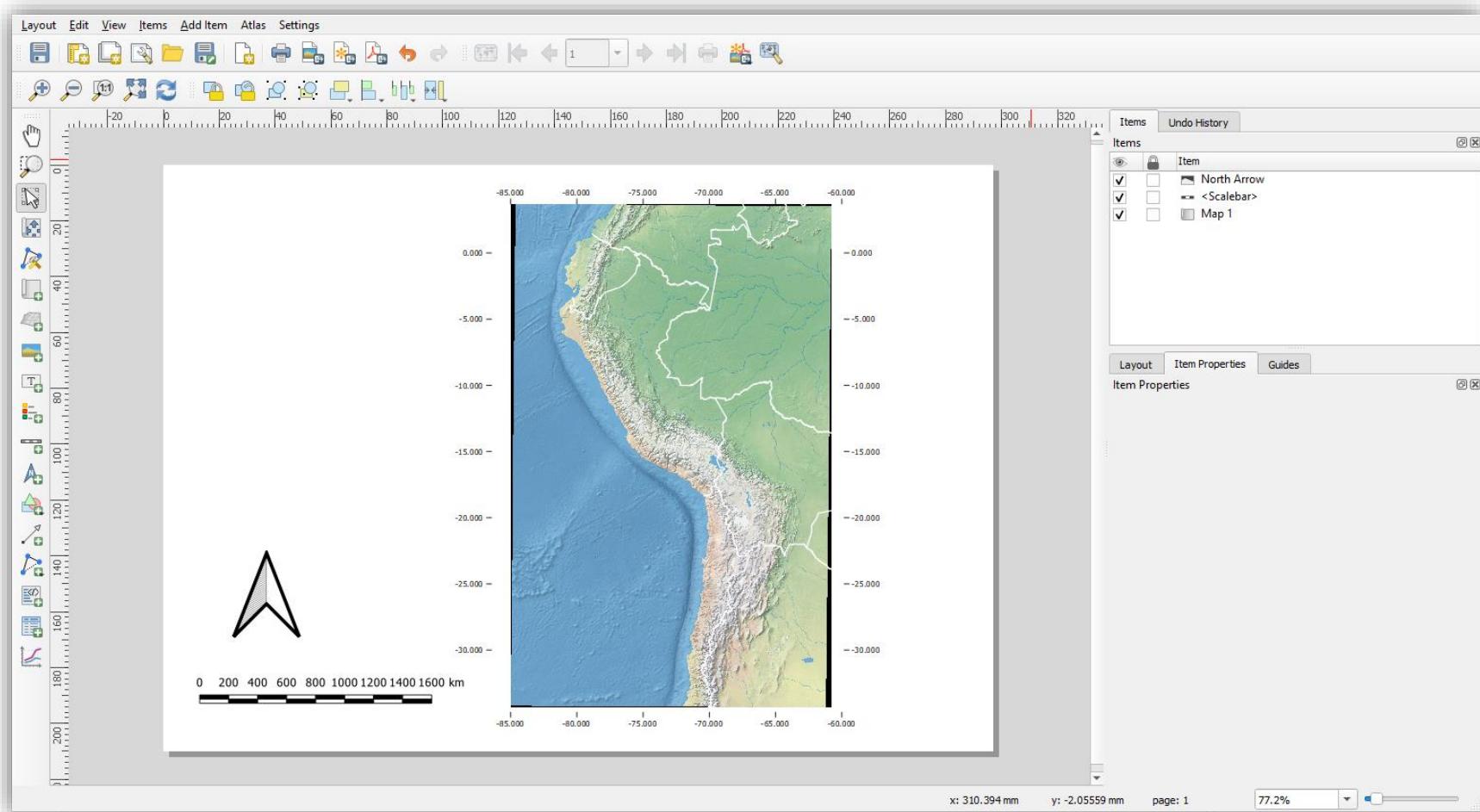
Softwares & Applications



Softwares & Applications



Softwares & Applications



Credit: Roberta Davidson



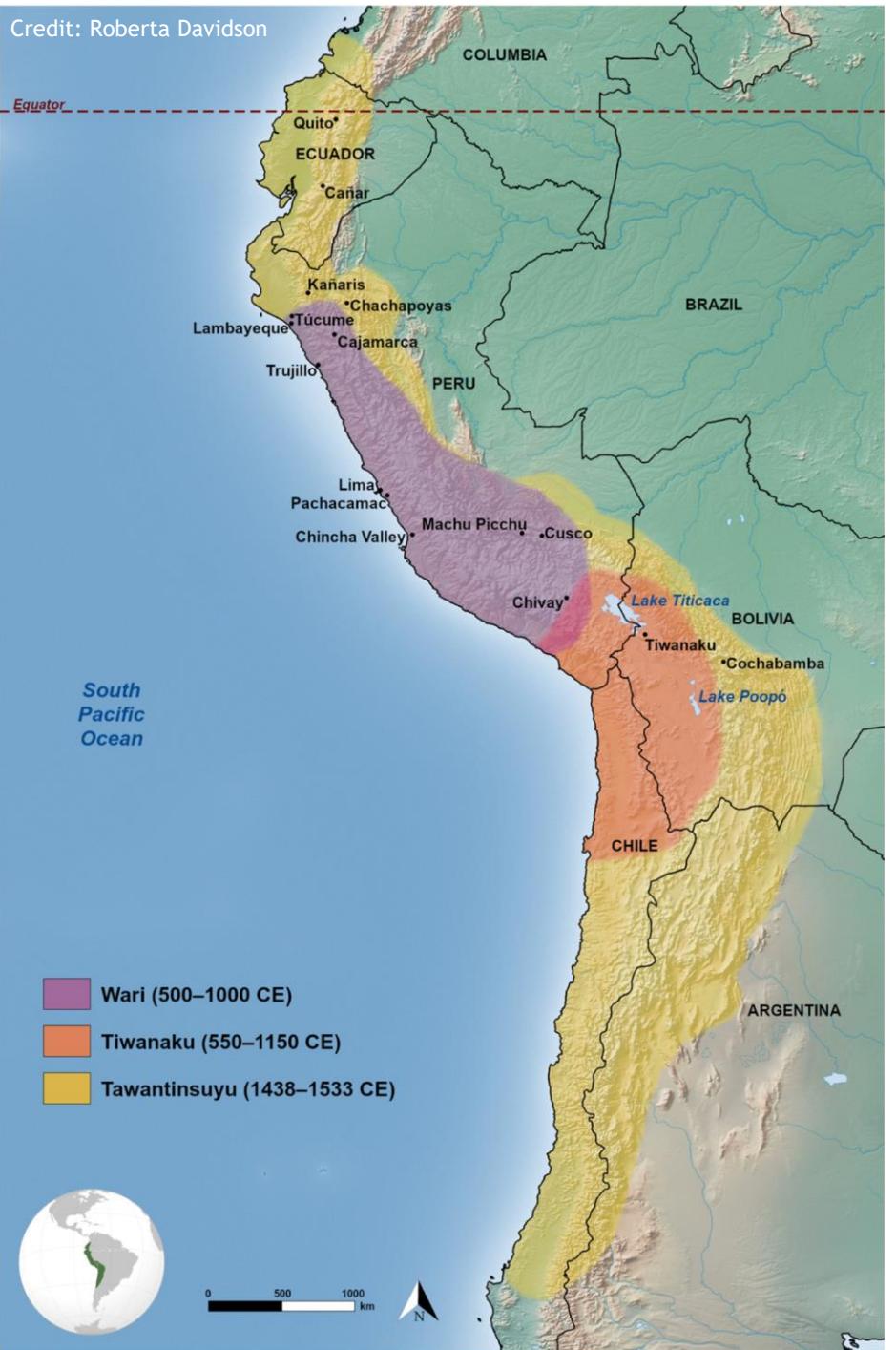
0 500 1000 km



South
Pacific
Ocean



Credit: Roberta Davidson



-10.000

0.000

10.000

10.000

WEST AFRICA

© J.N. Cerasoni



250 0 250 500 750 km

1:12608709

-10.000

0.000

10.000

Photography and Digital Processing



SUMMER
SCHOOL
2021



Doorway to
Human
History



What is the role of photography in scientific literature?

subjectivity vs. objectivity

Fiction vs. non-fiction

PRETTY DOESN'T MEAN NON-SCIENTIFIC



Basics of photography

1. Camera handling

1. Use a tripod whenever possible
2. Grip camera properly
3. Stabilise whole body

2. Lenses

1. Prime lenses (fixed focal length, faster and sharper)
2. Zoom lenses (series of lenses within same lens case, more flexible, bigger and heavier)

3. Depth of field

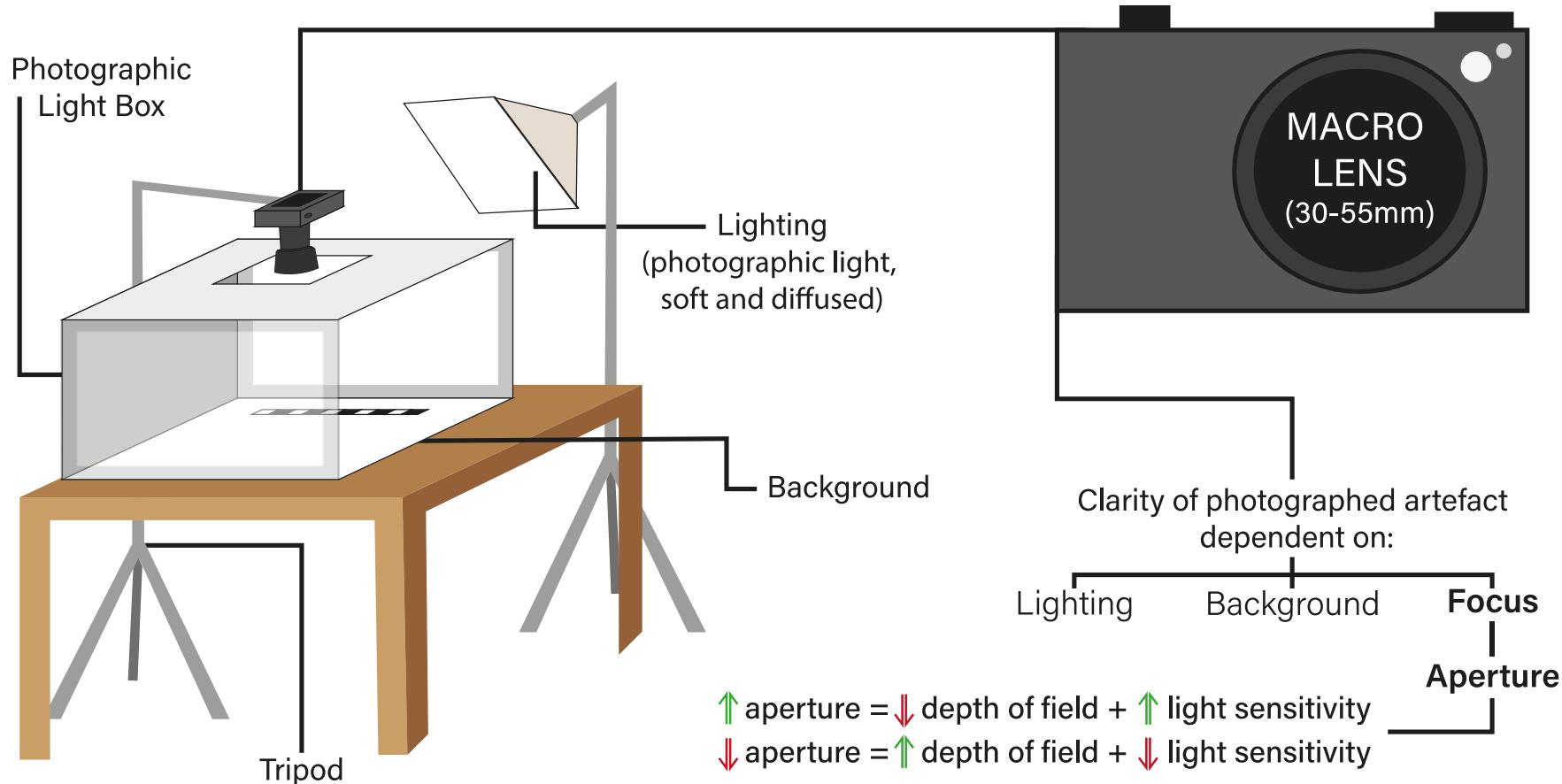
1. Focal length
2. Maximum aperture

4. Exposure

1. ISO
2. Shutter Speed or Time Value
3. Aperture



Basic set up



Depth of Field

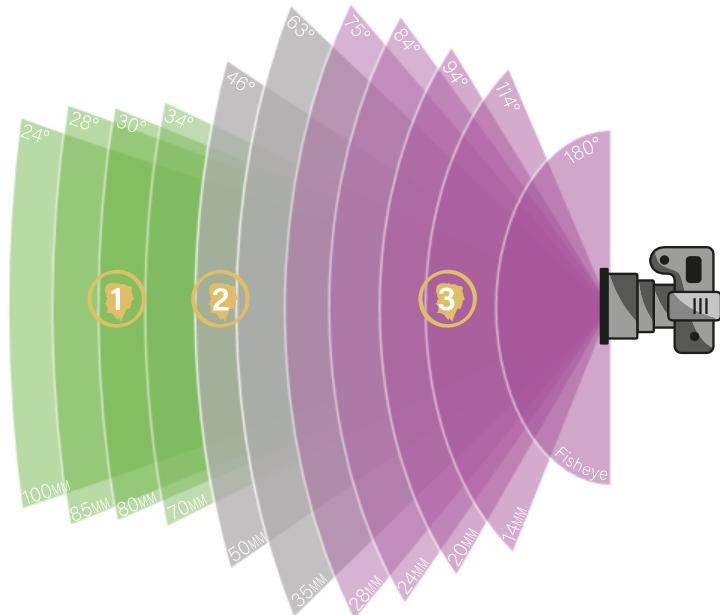
Measure of distance between
point of convergence
of the lens and the *sensor*

Measured in mm

How much of your scene you will capture

Small number -> wider angle of view

Larger number -> narrower angle of view



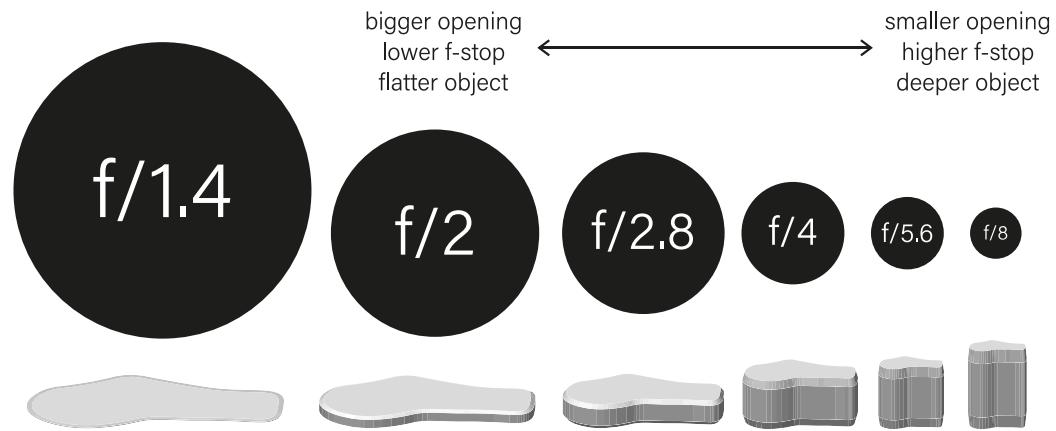
Aperture

Expressed in **f-stops** (larger the number, smaller the opening)

Affects depth of field:

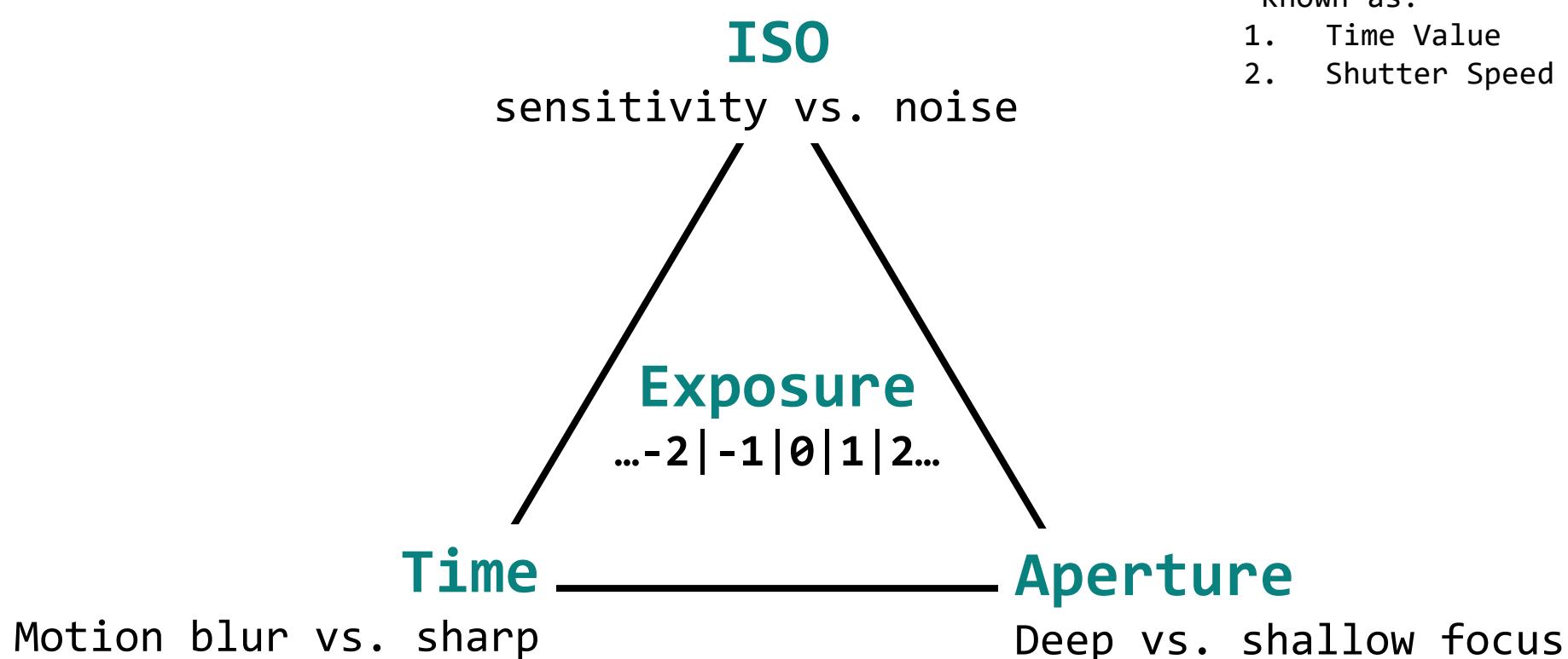
Larger opening -> shallower depth of field

Smaller opening -> larger depth of field



Exposure

Aperture + *Time + ISO



Digital Processing



Example: Iho Eleru Calvaria

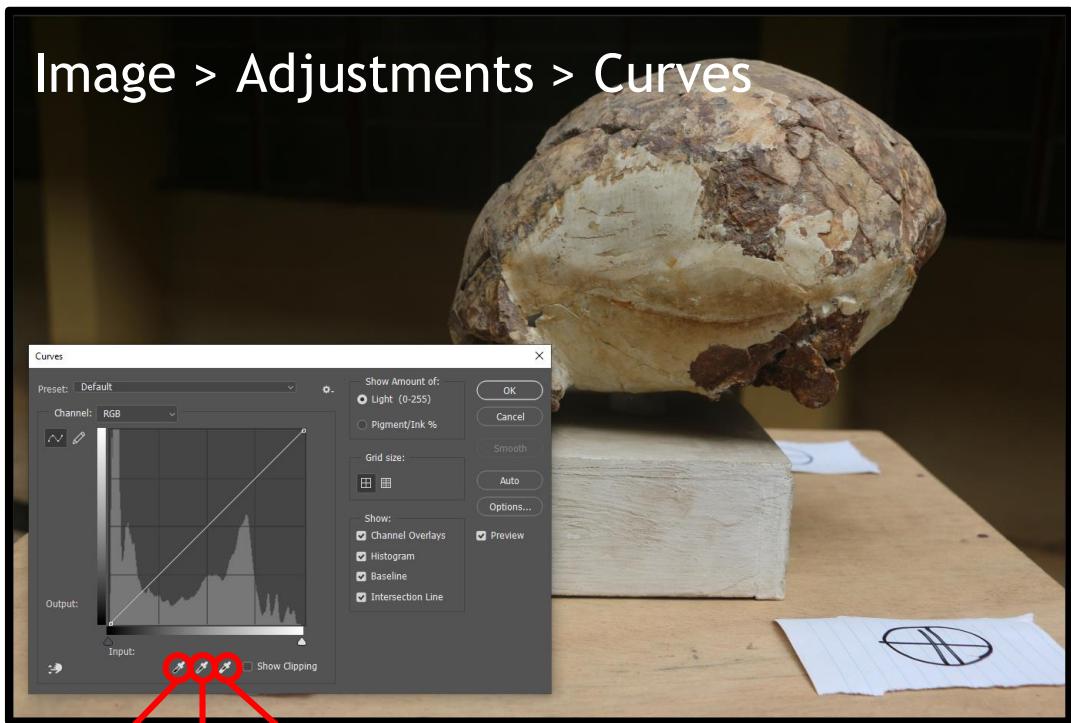


White Balancing



White Balancing

Image > Adjustments > Curves



black
grey
white

Before



After



Background removal

Subject isolation

1. Magic Wand Tool
2. Quick Selection Tool
3. **Select > Subject**

Edges enhancement and clean-up

1. Select object (use Object Selection Tool)
2. Feather edges
3. **Select > Modify > Feather**



Beautification



SUMMER
SCHOOL
2021



Doorway to
Human
History



**What makes a figure efficiently communicate data,
while still being attractive?**

Simplicity

easy to follow

straightforward colours

strong contrast

Balance

coherent stroke and text size

thoughtful composition



Always keep in mind the purpose of your graphics:

Scientific Publication?

Stroke thickness (0.5-2pt)

Text font and size (sans-serif, 7-12pt)

Column width (1, 1.5, 2)

Colors (11-88% black, CYMK)

Graphical Abstract?

Telling a story

Simple yet comprehensive

Press Release?

Simple, eye-catching

Photographs

Public Outreach?

Artistic

Colorful

Not necessarily precise or scientific



Thank you for your attention!

Any questions?

