![Utrecht University logo]

Visualisation in Earth Science

Resource modules for MSc students in Earth Sciences

Departments of Earth Science and Physical Geography

Utrecht University

2021

Written by: Anouk van Boxtel, Ayla von Essen, Jasper Hupkes, Marjolein Naudé, Stan Schouten, Steye Verhoeve, Lea de Vries, Jorien van der Wal

Coordinated by: Paul Mason, Marcel van der Perk

Table of contents

1 Introduction

This resource module is intended to help provide background knowledge and theory on visualisation within academia, and Earth Science in particular. Data visualisation is a key aspect of research and science communication. In general, practice with data visualisation is often learn-as-you-go, and as such specific techniques and good form are learned case by case. There are many valuable resources available on this topic such as textbooks, peer-reviewed studies, and academic feedback, a selection of which have been compiled here. The topics covered in this module range from vocabulary and theory, to determining which figures to use, to specific figure formats. Chapter 2 will give you some background on specific vocabulary and visualisation techniques in Earth Science. While chapter 3 covers theory on the main figure types found in Earth Sciences, the other remaining chapters will give you more practical information on the process of creating figures themselves. Use this guide to explore the topics you're interested in focusing on but note that much of the theory can be widely applied. Understanding how to best visualise using one format can provide insights on how to visualise others.

Each chapter will start by listing the learning outcomes for that chapter. Do they correspond to those you set for yourself? Read through the sections and review at the end of each chapter whether you have achieved the outcomes predicted at the beginning. If not, check out the references of subjects you would like to learn more about or go back to the parts you have not yet mastered and see what you can do to improve.

2 Understanding scientific visualisation

Scientific communication is an integrated and valuable facet of the scientific process, though one not often trained during academic studies. To conduct good science communication, communicators need to (a) know their audiences and consider the impact the

> **Learning outcomes:**
> At the end of this chapter you will:
> - Understand the relevance of figures in scientific communication
> - Be acquainted with the vocabulary used in this module

information will have, (b) communicate both the successes and limitations of their research, and (c) continue to train their communication skills[1].

2.1 Vocabulary

As we work on training our communicative skills and to help provide some clarity, the following are some brief definitions for some of the concepts covered in this module.

Though it often seems straightforward, this is a useful exercise in good preparation and good science.

What is scientific communication?

Scientific communication is an umbrella term which includes all forms of communicating science; the internal dialogue between scientists of different fields, and the external dialogue between the scientific community and the public.

What is scientific visualisation?

Scientific visualisation refers to the process of transforming raw data into graphic representations. The main goal is to communicate information to your audience more succinctly. As scientific research in geosciences often deals with large data sets, visualisation is an invaluable tool for scientists to gain insight into the trends, outliers, and behaviour of their data. Moreover, as a universal language, figures are eye-catching and can be an efficient way to communicate research to scientific audiences as they simplify your message without the use of much text.

There are four major benefits to data visualisation[2, 3]. Data visualisation:
1. Provides access to complex information
2. Promotes visual thinking (which may help to prioritise and structure your thoughts)
3. Facilitates communication
4. Improves memorability of information (by stimulating different parts of the brain)

What are scientific figures?

Figures are any type of graphical illustration, *e.g.*, photographs, maps, or data plots. Scientific figures and tables are both space and time effective ways to represent numerical and statistical data. In science, figures can be communicative or exploratory. Communicative figures are used to help the audience understand your research, data, and analyses. Exploratory figures on the other hand are used by scientists to study and interpret their own data. In Earth sciences, exploratory figures can be based on field data and data from laboratory analyses or experiments, but also on models.

Due to the broad nature of Earth sciences, numerous figure types are used in this discipline. The main figure types used for data visualisation in Earth sciences are maps and graphs. However, various other types of figures such as field and microscope images, schematic representations of cross sections, processes, and infographics are also often used. The characteristics and purpose of various types of figures are discussed in Chapter 3.

Figures vs tables

Tables are technically not figures as they contain text and numerical data in column form. They do not visualise patterns or relationships. *Figures* on the other hand can be graphical

representations of patterns and relationships. In scientific text, *Table captions* are located right above the table. *Figures captions* are located below a figure.

2.2   Visualisation techniques in the Earth Sciences

From on-site data collection to satellite-based remote sensing, geoscientists use data visualisation to investigate, monitor, and represent Earth systems. The techniques employed depend on the type of data collected.

Spatial data

The visualisation of geospatial data is achieved via *cartography* a.k.a. *mapmaking*. In this process, variables correlated to geographical locations are layered over maps, in what is called georeferencing. When the visuals accurately depict the data, maps are an efficient and intuitive way to communicate spatial processes[4]. Nowadays, most geospatial data has been acquired with remote sensing devices, satellite, or aerial imagery. Recently the addition of drone footage and higher resolution (satellite) cameras have provided Earth scientists with daily and detailed images of the planet[5, 6]. This has transformed our field considerably over the last decades, increasing the amount of geospatial visualisation programs and allowing for more detailed geospatial analysis.

Several figure types can be used to represent spatial data:
- Maps
- Graphs
- Images/pictures
- Animation/videos

Temporal data

The visualisation of temporal (time series) data is a powerful tool in the geosciences, as we often study changes trough time. Temporal data is mostly represented by graphs but can also be visualised using process visualisation techniques.

Combination of spatial and temporal data

Observing changes in (spatio-)temporal factors lends insights into the forecasting of future events as well as the reconstruction of the past. Despite the complexity of spatiotemporal data, recent decades have seen shift towards real-time visual analysis and data integration[7]. This has been accomplished through the development of platforms such as ArcGIS and R statistical package[8].

Statistical data

After the collection of spatial and temporal data, statistical analysis can be applied. Statistics is the science of analysing, interpreting, validating, and characterizing systems that are

incompletely known. Visualising statistical data is an important part of data analysis. To do so, we must ensure that our databases have a clear data-structure. This includes adequate documentation, effective naming conventions, and an understanding of the project's needs[9].

3   Figure types

> **Learning outcomes:**
> At the end of this chapter you:
> - Can identify the main figure types in Earth Sciences
> - Know what specific type of figure you need to visualise your data and what you need to incorporate into those figures
> - Know how to correctly and accurately visualise your data

Due to the broad nature of the Earth science program and the variety of research disciplines and data types, you will encounter many types of figures during your studies. This chapter aims to describe the main figure types found in Earth sciences, what to use them for and when, and provides tips on designing specific figures. It does not contain tutorials on <u>how</u> to create the specific figures using specific software. There are many options for software to create graphs in and the exact method varies between different applications. If you'd like to learn more about which software to choose, take a look at Chapter 4.4. Here, you will find an overview of the different types of programs used to create graphs. To learn <u>how</u> to use these programs to create graphs we recommend looking our other resource modules such as *Excel basics* or find how-to guides online.

3.1   Graphs

Graphs show the relationship between two or more variables, generally along an x- and y-axis. Graph types that are generally used in scientific articles are bar graphs, line graphs, and scatter plots, but many more exist all with their own characteristics. The type of graph most suitable for your data depends on the message you want to convey (*e.g.*, the relationship you want to show[10]) and what type of data and variables you have[11]. For more information about different types of variables, see the information box below.

This chapter gives you some general information on designing graphs and discusses some of the most common graph types in Earth sciences. If you want to learn more on data visualisation in graphs, we recommend you to also take a look at <u>this website</u>, an open-

source e-book on the fundamentals of data visualisation. Although it is specifically written for creating graphs in R, it contains a lot of relevant information on general graph design.

A variable is any factor that can take on different values, *e.g.*, age, species, or concentration. Different types of variables can be distinguished, based on the type of data they contain:

Quantitative variables contain numerical data, i.e. amounts, and can be split into **discrete** and **continuous** variables.
- **Discrete variables** consist of only whole numbers. This type of data is obtained by counting, *e.g.*, species counts.
- **Continuous variables** can consist of any real number. This type of data is obtained by numeric measurements, *e.g.*, temperature, age etc.

Categorical variables contain qualitative data and represents groupings. This group can be split into **nominal**, **ordinal**, and **binary** variables.
- **Nominal variables** consist of groups without a rank or order in them
- **Ordinal variables** consist of groups ranked in a specific order
- **Binary variables** represent yes/no outcomes

Note that categorical variables can appear numerical when groups are given specific numbers, but the numbers represent groups.

3.1.1 Common graph elements

Although each graph type has its own layout and elements, there are some common elements all graphs generally need to contain:
- Axis labels including units
- A legend or labelled curves when there are multiple datasets or variables plotted[12]
- A legible font size[12]
- A figure caption at the bottom of the graph

Additionally, graphs may contain other elements, such as tick marks, labels, and grid lines. The exact layout of your graph depends on the purpose of your figure and the message you want to convey. Do not automatically use the default settings of your plotting library or software as they are not fine-tuned for a specific plot-type. Instead, think carefully about wat your graph needs to include and what will work best to convey your message[13]. For example, in scientific papers, graphs are often used to show relationships instead of details about individual data points[12]. If you only want to show a trend in your data, gridlines are

superfluous and may make your graph look more cluttered. Therefore, you can often exclude elements such as gridlines from your figure. However, if your audience needs to be able to read exact values from your figures, you will need to include them.

3.1.2 General guidelines for graphs

<u>Visualise your data as accurately as possible</u>

Visualising your data as precisely and accurately as possible prevents you from misleading your audience, and, possibly even more importantly, yourself since you will also base your interpretation on your (exploratory) graphs. As different graph types and graphical elements affect the way we perceive information, they will also affect your and the audience's interpretation of the data. Therefore, when you are designing a graph, it is important to always **keep your data in mind.**

Some things to specifically pay attention to are:
- Always consider where your data comes from. What does it tell you and what would using different visualisations on it imply?
- Check your graphs and plots for errors and outliers. They may often still need to be included in your final graph, but it is important you are aware of them.
- Keep your data visible as much as possible. Often, it better to plot all data points instead of their averages for example, as this is more transparent.
- When using colour, use an appropriate colour scheme. See Chapter 4.3 (colour use) for more information.
- When you want to compare variables, make sure your axis ranges are similar.
- X-Y graphs can (un)intentionally misrepresent the effect size and therefore mislead your audience. This occurs when zero is used as the origin of the dependent variable's axis even though your data has a natural zero point. To prevent this, you can add a zero-break, or you can start your y-axis at zero. The latter solution depends on whether this makes sense with your variable[15].

<u>Keep it clean</u>

Keep your graphs as simple as possible without losing relevant aspects of your message. This means that you should avoid any 'chart-junk' in your graphs, *i.e.* elements that are unnecessary to comprehend the graph. These will make it more difficult for your audience to understand the message you're trying to convey[13]. For example, this means that you do not add colours or patters to your graph just to make it look nice, but consider carefully whether gridlines and small tick marks are necessary or not.

Additionally, do not use the default settings of your software as these are never adapted to <u>your</u> specific type of graph, data, and message. When you are making graphs using a programming language, specific data visualisation packages will often lead to nicer looking

visuals than using the default data plotting settings. For more information on software for data visualisation, see Chapter 4.4.

3.1.3   Scatter plots

Scatter plots are the most basic type of graph. They display observations for two variables as a collection of <u>points</u> on the axes of variation. Scatter plots can be used to study the <u>relationship</u> or <u>correlation</u> between <u>two quantitative variables</u>. Measurements of x-axis variables do not have a consecutive order. If your data does, you are better of using a <u>line graph</u> (see Chapter 3.1.4).

Trendlines

When you use a scatter plot to study the correlation between two variables, you can add a trend line or line of best fit to assess the strength of the correlation[14]. Trend lines can be linear or according to another formulae (power, quadratic, polynomial etc.), depending on the type of relationship you expect and observe. However, always be aware that a correlation does not automatically indicate a causal link between your variables: an unknown third factor might play a role as well.

Always display a formula and an R squared coefficient to your trendline. Furthermore, gridlines are most often left out of a scatterplot as it's the scale of one variable that is of interest.

Overlying datapoints

Especially when you are dealing with large datasets, your data points may overlap. To visualise your data as accurately as possible, keep all your datapoints visible. You can do this by making the outline of your datapoints bold, keeping them small, or adding uncertainty bars.

Independent variables

Scatter plots differ from line graphs as, for the former, it is often unsure if one or which of the variables is the dependent or independent variable. If you know which of your variables is the independent variable, plot this one on the x-axis.

Multiple datasets

If you want to plot the same variables for more than one dataset in one plot, *e.g.*, data from different study sites, this can be easily done by giving the datasets different colours or symbols, or by using multiple windows/facets. Make sure to always label the datasets so your audience can easily see which data is from what dataset. When there is enough space in your graph, it is best to label your datasets in the plot instead of adding a separate legend.

For an example of a scatter plot with multiple windows and trend lines, see Fig. 5 from Marret *et al.* (2001).

Multiple variables
There are three main ways to a third quantitative variable (*i.e.* a third axis) to a scatter plot:

1. The third variable can be represented by a colour gradient. However, be aware that it is easier for the human eye to interpret separate axes than colour spectra or sizes.
2. You can also use different sizes of the data points to add a third variable. Then, your scatter plot will be called a bubble chart. However, in this way, your audience will need to interpret surface area instead of a location on an axis. Interpreting surface areas is less intuitive however, so this may make understanding your graph more complicated than need be.
3. Use multiple windows with separate plots. This is often best because these are easier to interpret.

3.1.4   Line graphs
Line graphs show quantitative data along two axes of variation. Consecutive data points are connected by a straight line, indicating an order in measurements, such as in a time series visualizing trends over time. They are mostly used to plot time series data or spatial transects. In contrast to scatter plots, line graphs always plot a dependent variable on the y-axis with a consecutive order on the x-axis, usually time or age.

Depending on the resolution, you either plot datapoints separately (low resolution, little data points) or plot them together (high resolution, many datapoints). When there is a gap in your data, *e.g.*, due to a missing measurement or incomplete recovery of a core, include a line break to indicate the gap.

When your data is measured as an average of a continuous time or space interval, *e.g.*, calculated sedimentation rates, connecting data points does not accurately represent your data. In this case, you could use horizontal lines instead. In this case your line graph becomes known as a stairs graph.

Proportional data
When you want to visualise a time series of proportional data from a variable with different categories (adding up to 100%), you can make a stacked line graph.

Multiple datasets
You can plot multiple datasets in one graph as individual lines[14]. However, adding too many lines may make your graph look cluttered. If this is the case, it is better to use multiple windows.

Multiple variables

In Earth sciences, you often want to visualise multiple variables of one time series, *e.g.*, temperature, precipitation, and evaporation in one graph. In general, the best way to do this is by using multiple vertically aligned windows (stacked subplots). This makes it possible to spot similar patterns in your data and is a common way to plot data in Earth sciences, see this paper by Capron *et al.* (2019) for an example.

Do not plot multiple variables (y-axes) in one window when your lines are intersecting or interfering with each other. This can confuse your audience. If you want to assess synchroneity of two variables, you could plot two y-axes in one graph. This is specifically useful when you want to study *e.g.*, cyclicity of lake levels and rainfall with astronomical forcing. For an example, see Fig. 2 and Fig. 3 from this paper by Martínez-García *et al.* (2014).

Although this way of plotting multiple variables is common practice in Earth sciences, it might lead to incorrect interpretations and an unintentional misleading of your audience. Therefore, be careful with how you plot your data. There are two important concepts to keep in mind:

1. When you plot multiple facets with the same variable from **different datasets**, *e.g.*, two temperature time series from different study sites, make sure the y-axis range is the same for both graphs. Imagine the time-series temperature on the y-axis ranges from 10-15 degrees, whereas it ranges from 10-30 degrees on the other. This can lead to a seeming trend in temperature when there isn't one, or conversely hide an existing trend. For this reason, avoid default settings for axis range – it is better to set the axis range manually to make sure it is the same across all your facets.
2. Be careful with your interpretations if you are plotting multiple **variables** of one time series in facets (windows). When you are plotting different variables, you will automatically have different y-axes. This could create a misleading figure: you may observe a similar trend in two variables, but if the relative change in one of the variables is much larger than in the other, this trend might not be meaningful. There is no easy fix for this, just always keep asking yourself: 'Is this a meaningful axis range and trend?'.

Stratigraphic plots

A stratigraphic plot is a special type of line graph which is often used in Earth sciences to plot data from sedimentary records. In a stratigraphic plot, the axes are often reversed compared to a regular line graph. The independent variable (age or depth) is plotted in the vertical direction, and the dependent variable(s) is (are) plotted in the horizontal direction. This way, the y-axis with depth or age resembles the vertical location of the layers. Stratigraphic plots are often combined with stratigraphic columns, schematic illustrations that indicate the

lithology of the section. For an example of a stratigraphic plot, see Fig. 2 in <u>this paper by Knies *et al.* (2018)</u>.

In theory, when plotting count-data, *e.g.*, species counts, you should use bar graphs as your dependent variable is **discrete**. However, species counts are also plotted as line graphs in a stratigraphic plot.

3.1.5   Bar graphs

Bar graphs can be used to visualise the relationship between a quantitative variable (the dependent variable) and different categories (the independent variable). Bar graphs show categorical data. The height of the bar is proportional to the dependent (quantitative) variable, such as counts, proportions, and averages of measurements.

<u>Multiple datasets</u>
You can show multiple datasets binned in the same categories in one bar graph by using different shading or colours and clearly labelling them. When the different categories are showing proportions (and add up to 100 %), you can use a <u>stacked bar plot</u>.

<u>When not to use bar graphs</u>
Although bar graphs are a good option to visualise discrete variables (count data) on the y-axis, they conceal the data distribution and sample size when the y-axis represents continuous data, *e.g.*, averages[16]. Therefore, it is often better to plot continuous data using scatter plots, box plots, violin plots, or histograms. These formats visualise the data distribution better. If you want to learn more about this, and for examples of the respective graph types, see <u>this paper by Weissgerber *et al.* (2019).</u>

3.1.6   What graph do I need?

The table below gives an overview of the purpose and variables used in the graphs covered in this module.

Graph type	Purpose	Variables
Scatter plot	• Visualising and studying relationships between variables • Assessing correlations between variables • Study data distribution	• X-axis: dependent or independent, continuous • Y-axis: dependent, quantitative (most suitable for continuous, less for discrete)
Line graph	• Identifying trends over time or space • Comparing multiple dependent variables of one common independent variable	• X-axis: independent, continuous with a <u>consecutive order</u> • Y-axis: dependent, continuous
Bar graph	• Comparing categories • Displaying proportions	• X-axis: independent, categorical (often nominal) • Y-axis: dependent, quantitative (generally discrete)

3.1.7 Statistical plots

When you are analysing large data sets, you will often have to perform statistics as well. It is especially helpful in the initial data analysis stage to plot your data to study the data distribution. Although statistical analysis is beyond the scope of this module, we will discuss some basic statistical visualisations here, focused on data distribution and errors bars.

<u>Visualising data distribution</u>

Several ways exist to visualise the distribution of your data, the display frequency, and comparisons between groups. The most common ones are discussed below[14]. For more information on visualising data distribution, we recommend you read <u>this paper by Weissgerber *et al.* (2019).</u>

Box plots display data distribution of quantitative data through quartiles. They can be used to compare medians and distribution of categorical data. They indicate medians, symmetricity of the data, spread in the data, and skewness. Since they take up less space than other types of data distribution graphs, they are specifically useful if you want to visualise a multitude of groups. However, box plots conceal significant details about how your data is distributed (gaussian, bimodal, or multimodal).

Violin plots are a more advanced version of a box plot, that also indicates probability density. Therefore, they also show the distribution shape of the data. However, they are more complicated to interpret.

Histograms display distribution of quantitative data binned in bars indicating the frequency at each interval. They give insight into what type of distribution your data has, where values are concentrated, and what the extremes are. They give more information about distribution than box plots, but also take up more space and make it difficult to compare different categories.

<u>Error bars</u>

Error bars represent variability and are used to indicate the uncertainty or error in your graphic data. Generally, error bars are based on the statistical standard deviation of a mean, represented by 1 sigma or 2 sigma error margins (68% and 95% confidence interval respectively). However, error bars can also indicate other uncertainties. Therefore, when you use error bars, think about what they indicate and state this in the figure caption.

3.2   Tables

Although tables are technically not figures, they may also be used to present large amounts of data in an effective way[12]. Typically, you use them when the exact values rather than the trends of your data are important. Make sure your table looks neat and is easy to understand.

As with graphs, think about the message you want to convey and what you want to include. Do not put all your raw data in a table, this clouds your message and makes your table difficult to understand[11].

General guidelines for tables:
- Provide units in the top row/header
- Use a legible font type and size
- Use sufficient white space in your table to prevent it looking cluttered[12]
- The figure caption is placed <u>above</u> the table (in contrast to figures, where it is placed below)

3.3   Maps

Geoscientific communication often includes spatial presentations of study areas. Such maps often display a wealth of information that immediately informs an audience of the geoscientific setting involved in the presented research. Maps may show the outcropping lithologies in an area, or they may highlight locations that samples were collected from,

where analyses were done, measurements were taken, etc. For the reproducibility of scientific findings, it is important that such maps adhere to certain requirements. This way, the audience of your presentation or reader of your text can verify your observations, either through satellite imagery or in the field. In general, maps require:

- **A metric scale**: in m or km, depending on the size of your map.
- **Spatial reference**: Always include a coordinate system and north arrow, and mention the geographical/projected coordinate system used to display your maps (*e.g.*, WGS 84 or others).
- **Legend:** must include all features on the map, *e.g.* coloured/shaded areas, symbols etc., including the units of values you may be presenting.
- **Appropriate symbology** and colour use.

Apart from these requirements, maps often include:
- **Abbreviations and annotations** that indicate names of mountain ranges, rivers, towns, etc. All abbreviations are explained in the figure caption, rather than in the legend.
- **Insets** of larger scale maps highlighting the regional/global tectonic setting, or other valuable information that cannot be included in the main map.
- **Base maps**, *i.e.* satellite imagery, terrain, hillshade, or slope maps that add a three dimensional or spatial perspective to the mapping presented on top. Choose the type of base map you use consciously and with its purpose in mind. Note that for the base map to increase value of the presented map, top layers such as shaded areas and other map projections need to have partial transparency.

Remember to always make maps with a purpose in mind. Consider what it is that you want to communicate and decide what the bare minimum of information you need to include to convey your message is. Avoid including all available information of a certain area in your map, but consciously and critically select the map layers necessary to inform your audience as efficiently as possible. For an example of a clear and targeted map, see Fig. 2 from this paper by Walk *et al.* (2020).

More detailed information on visualising maps can be found online, for example here and here.

3.3.1   Tips for each map component
<u>Scale</u>
The maximum value of a scale is generally a round number, and they often include one or two smaller values to indicate shorter distances. For example, if your scale bar has a maximum length of 1 km, you may include extra divisions at 0.5 km and 0.25 km. Keep in mind, however, that values with more than two decimal numbers are uncommon and can be perceived as distracting, so try to avoid this.

Spatial reference

Be consistent in the type of coordinate notation you use to present latitude and longitude, *i.e.* choose to either use decimal degrees (DD); degrees, minutes, seconds (DMS); or other notations (see *e.g*, Archison, 2012). In terms of style, you may choose to include latitude and longitude values on all sides of a map or only on one side (*e.g.*, top and right). Similarly, you may choose to include a grid inside or tick marks on either side of the map. You may also decide to include or exclude zero values of minutes and seconds in the DMS notation, *i.e.* 32°0"0' N vs 32° N. Values on the left- and right-hand side of a map are generally vertical. If you are unsure of the style you need to use for your assignment, check what the conventions are in peer-reviewed literature in your research field, or ask the lecturer/editorial board for their preferences.

Legend

Organise your legend by grouping certain similar features together and ordering them in a logical way. Legends are part of your figure and may be included as insets in your map (if they do not cover important information), or they can be set outside of your reference system if that is more appropriate. Try to keep your figure clean and uncluttered, but also minimalise unused space in your figure when you include a separate legend below your map, for example. Adapt the layout of your legend accordingly.

Symbology

Each sub-field of geoscience uses field-specific symbology to indicate common features mapped for the purposes of that field. Adhere to the symbology in your field of research and look up whether any open-source tools are available to simplify their implementation. For example, the United States Geological Survey (USGS) has made a geological mapping symbology package available for Adobe Illustrator.

Colour use

Be considerate of your audience's perception of colour use in your map and make sure your colours can only be interpreted in the way you want them to be. Keep maps simple and include symbols where possible to avoid confusion. For general information on colour use in data visualisation, see Chapter 4.3. Additionally, be aware of the colour use conventions in your field. For example, geological maps often use specific colours for certain rock types (*e.g.*, igneous rocks are often pink) or ages. These are conventions, not rules, but keep an eye out for patterns used in your research field and apply them to your own maps for convenient and clear communication with your peers.

Line features and polygonal mapping

Geoscientific maps may include multiple aspects, such as line features (*e.g.*, drainage networks or faults) and polygonal features (*e.g.*, geomorphological or geological features),

but it is often favourable to focus on a single aspect for clarity. For example, if you are studying the geomorphology of an area and you also want to highlight its geology as background information, it will not be beneficial to display both aspects in a single map as the shaded areas of geomorphological and geological units will overlap. In such cases you may choose to present two separate maps, a geomorphological and a geological map of the same area. Keep in mind that if you want your audience to compare both maps and understand the link between geomorphology and geology, your maps should both have the same reference frame and scale, and they should include some details that can be used for location recognition (*e.g.*, place names or distinct rivers/lakes).

3.3.2 Software

Most scientific maps are made with software that is specifically intended for map design, such as ArcMap (ArcGIS) or QGIS. These programs are preferred as they allow for georeferenced plots of multiple layers of raster and vector files. However, the final design and editing of maps made in GIS software is often done in a vector editing program such as Adobe Illustrator, or InkScape (see Chapter 4.4).

As a beginning MSc student, you may or may not have experience with using GIS software. Depending on the type of course or project you need to make a map for, it may be necessary to learn how to work with GIS. In many other cases, it may be sufficient to edit an existing map and include your own observations. In that case, make sure you have a high-quality image to begin with and keep in mind the information you need to include to make it legible and useful for your purpose. If necessary, you may ask your supervisor or a colleague to include the most important aspects for your map in GIS and to send you their map so you can edit it in a different type of software.

3.4   Images/photos

In your work you can include images of *e.g.*, field sites, structures, or microscope pictures to help your reader to visualise parts of your research that are difficult to describe concisely in words[12]. When including an image, always make sure the resolution is sufficient for the medium you are using. Preferably use your own photographs. Pay attention to copyright regulations when you use photographs from the internet in your work.

Checklist images:
- Images in scientific writing must always contain scale bars[12]
- Label important features[12]
- Add a legend when you are using colours, symbols etc. to clarify features[12]

3.5   Animations/videos

As spatiotemporal geologic process often span thousands to millions of years and are not directly observable, an intuitive understanding of these abstract process can often be achieved through animation. For the deep sea or deep Earth, photos, videos, and simulations are required for us to access them visually. Informative videos and animations of the Earth facilitates the transfer of knowledge to more diverse audiences than traditional reporting. The creation of a global archive of Earth system animations can help to stimulate public and scientific interest in the Earth sciences, and bring greater appreciation of the field[17].

You may find that an animation or video is the ideal visual element to help explain or understand a particular topic. There are some libraries for geological animation online (*e.g.*, from USGS) but they are often quite topic specific and may not be adaptable to your needs. Though it is a significant undertaking, it is possible to create animations and videos that suit our own needs.

Some rules to keep in mind when making your own geoscience videos[17]:

- **Rule #1**: Animations are complex. Creating a scientifically correct animation is a field in of itself. It is often outside the scope of a master's student to create original animations. In this regard, we'd advise to focus on video and use animation sparingly.
- **Rule #2**: High quality sound pays off. Good sound is critical to making videos and animation that have a high production quality. Recording good sound should be a high priority.
- **Rule #3**: Keep them short and sweet. To maintain the attention of your audience, videos of three to five minutes fare best. If you're aiming for longer videos, keep your intended audience in mind and consider which content is necessary.
- **Rule #4**: Include closed captions. For the hearing-impaired this is essential, but closed captioning can also serve to help your audience follow the video even in noisy environments or in situations where they don't want to disturb those around them.
- **Rule #5**: Provide resources and give credit. The websites, textbooks, or contacted experts you used to research the video's content may prove useful to your audience. List these references at the end of the video and be sure to thank those who helped in the video production.

3.6   Other figures

3.6.1   Process visualisation

There are times when instead of visualizing data we may want to illustrate a process, such as the methodology for an experiment or the rational towards a hypothesis. **Process visualisation**, a.k.a. **process mapping,** or **schematics**, serve to identify the steps or decisions taken to achieve a particular outcome. They are in diagrammatic form, using simple shapes which helps to visualise the flow of information, documents, or materials. They can also highlight the interdependence between steps in the process. For an example of such a figure, see Fig. 2 in this paper by Ioja *et al.* (2013).

Process visualisation can also be used to visualise Earth science processes that are otherwise difficult to explain in text. You will find them mostly in textbooks, but they also occur in scientific papers where they are often used to visualise complex processes such as the interconnectedness of different phenomena or visualise the implications of your results. See Fig. 2 of this paper by Palmer *et al.* (2019) for a visual representation of a large scale earth scientific process, or Fig. 2 in this paper by Reinfelder (2011) for an example of a visualisation of a small-scale process.

3.6.2   Infographics

An 'infographic' or information graphic is a way to create a narrative using icons, illustrations, and minimal texts. Infographics condense large amounts of complex data or data processes into a visual story. Within the sciences, infographics are often used as a way to communicate information to non-academic audiences. Amongst scientists however, infographics can serve to communicate research at conferences or within departments outside of the formalities of a journal article[18].

3.6.3   Geoscientific figures

Many geoscientific fields of research incorporate visualisations of field-specific results or conceptual models. Think of stereoplots for structural data, rose diagrams for sedimentological data, (balanced) cross sections, AKF diagrams for mineralogy etc. In each research field and with each type of geoscientific figure there are certain conventions that researchers adhere to, and there is often (continuously evolving) software that is used to construct such figures. We advise you to thoroughly research the type of figure you want to make, and to ask supervisors, peers, and other researchers or staff members to find out more about this.

4   Creating figures

> **Learning outcomes:**
> - After this chapter, you will be able to:
> - Follow the steps required to design a good scientific figure
> - Use colour appropriately in your figures
> - Make an informed decision on what software to use to create your figure

In this chapter, you will find guidelines for creating figures, an overview of useful software and what figure types to use it for, and specific information on the use of colour. This chapter does not contain tutorials for creating figures using certain software or programs. Many of such tutorials can be found online. If you are looking for a tutorial on making graphs in Excel, take a look at the *Excel* module. For a tutorial on creating graphs in Python using the matplotlib and seaborn libraries, see the *Programming* module.

4.1   A short remark on data transformation

A major part of data visualisation occurs prior to the visualisation itself: In order to visualise your **raw data** in the form of a graph or map, most of the time you will need to **transform** it to get it in the correct form. Although data transformation is beyond the scope of this module as it can use its own complete module, it is important to think about this whenever you want to visualise data.

Data transformation can consist of various steps, *e.g.*, importing data, renaming variables, reordering observations, filtering and selecting data, and grouping. For all of these steps, you will need to make decisions. As these all affect the way your audience sees and interprets your data (and, in the case of exploratory graphs, the way you interpret your own data), you will have to think carefully about this to prevent misleading your audience and yourself. Therefore, data visualisation is strongly connected to scientific integrity. Additionally, in the steps of data transformation, it is easy to lose track of the changes you've made. Therefore, it is important to log everything you change to your raw data, and always keep a copy of the original data. Keep this in mind when you are making figures yourself.

If you want to learn more about data transformation, we recommend you look at chapter 5, Data Transformation from the e-book R for Data Science by Hadley Wickham and Garret Grolemund.

4.1.1 Forethought of making figures

<u>When to start making figures</u>

Although making figures is a time-consuming process, it is well worth spending time on it. Making figures is a very important aspect of both communicating your research and data to your audience and your own data analysis. Using figures in your work helps you to communicate complicated research by simplifying and visualising your data[11]. Moreover, when making figures you're required to focus on certain elements of your research which may cause re-evaluation of your data interpretation or collection methods. These selective thoughts help you take your research and analyses to a higher level. Therefore, we recommend you to already start making your figures early in your data analysis stage. This will save you time when you start a write-up, as you have already thought about your data extensively beforehand.

<u>When to use figures</u>

Including professional figures in your work makes it more attractive to a potential reader. However, do not add figures just to increase the visual interest of your product. Keep in mind that the main function of using figures is to help your reader to understand your writing or presentation. Using too many and/or unclear figures will only decrease your audience's understanding. In general, you use a figure in your writing if it shows something that is relevant, and if it is difficult to describe in words. When you need to explain a complicated lab procedure in your methods for example, or a complicated process, you might use a conceptual figure to simplify it to your reader. You can also add a map of the location of your study site or a picture from a field site showing specific features relevant to your study, but do not insert a picture just because it 'looks nice'.

More specifically, creating graphs or maps to visualise your research data is a good way of communicating your research results to your audience. However, if your data is not too complicated, it might be better to present it in text (*e.g.*, for simple statistics like p-values). As a rule of thumb, if you can present your results in one or two sentences, it is better to use text[10, 11]. If your data however is more complicated or too extensive, you should use a figure or a table as this saves space and prevents your text from cluttering up with too much information[10]. Furthermore, graphs and maps can reveal trends, correlations and patterns that are otherwise impossible to detect[10].

<u>Determine your message</u>

The main purpose of figures in scientific writing is to communicate your research to your audience[10]. Therefore, each figure needs to contain a core message that is relevant to your research question[19]. Think about the message you want to convey with your figure and keep the core message in mind when you are creating it. Without a core message, your audience will likely have trouble understanding your figure. One tip to make sure your graph has a core message is to check whether it provides the answer for a specific question.

Since each of the figures you use needs to convey your message about your research, it is preferred to make your own figures instead of using figures from other sources. If you do use figures from other sources, be sure to adapt them to fit your own message.

<u>Select what you want to show</u>
As you will likely be dealing with complex data, you cannot show all the data you have without considering the purpose of the graph or figure. Therefore, **keep it simple,** and make a careful selection of what you want to show. Students (and scientists in general) often want to be as complete and unambiguous as possible. This often means they want to include every detail in their figures. Moreover, scientists want to show how much research went into their paper because otherwise it might look 'under researched' or 'non-scientific'. Nonetheless, keeping your figures simple is very important for your audience to understand your message. If your figure does not have a clear message your audience may come up with a message themselves, which can be completely different from the story you wanted to tell. However, always be aware of (un)intentionally misleading your audience by leaving out important information.

If you have multiple messages you want to convey, it is often better to make multiple graphs. Overview figures can be helpful to show similarities or differences in trends and patterns, but do not blindly add unrelated aspects together.

<u>Choosing a figure type</u>
Once you have determined your message, you need to choose what type of figure you are going to make. The figure type depends on the message you want to convey and your data. For graphs specifically, the type variables you want to plot determine the most suitable graph type as well. To help you choose a suitable figure, several types of graphs and their characteristics were discussed, as well as mapping conventions, and the use of photos. If you are looking for a more specific graph type, you can use online databases such as Matplotlib library. You can also look at the types of figures used in peer-reviewed papers, especially when they are in your research field and use similar data to yours.

4.2   General guidelines for creating figures

Once you have determined the type of figure that best fits your data and message, you can start to design it. One good way to start is to first draft your figure using pen and paper. After this, you can use the software of your choice (see Chapter 4.4) to make the actual graph. Adapting your figures is time-consuming, so designing your figure before creating it will save you time in the end. Additionally, in this way, your figure design will not be limited by your skills with certain software.

Make sure your figures are interpretable on their own

Figures often are the first part of a text or presentation that draws the attention of an audience. In fact, many readers of scientific texts jump to look at the figures in a published article before reading other parts of the text. Therefore, the message you convey with your figure needs to be as clear as a stand-alone component of your presentation (whether written or oral). All information must be incorporated in the colour use, symbology, line- or polygon features, in annotations and within the figure caption. The message must be clear from the figure and its caption without reading your accompanying text.

Show your data

Try to show the data points behind your graph or map as much as possible, especially when you are first presenting your data. This means, for example, that you do not only show average values of measurements, but also show the datapoints on which the averages are based. This creates more transparency in what the data is behind your interpretations.

Check your figures for errors

Check your figures for errors as these might completely change the interpretation of your data[10]. This is important when you are communicating to an audience, but even more so when you are making graphs to analyse your data.

4.2.1   Layout

Be consistent

Use the same style, font, and colour scheme for all your figures. This will make your end-product look more uniform and professional.

Use high resolution figures

Only use high-resolution figures to avoid blurring or pixelating. If possible, export your figures in a vector (SVG) format as this allows unlimited scaling. When you are creating a figure consisting of different components from other programs, make sure the separate figures you are importing are also in a vector format.

Use clear figure labels

Always add labels to your figures if you are using colours, symbols, or other things to visualise multiple variables. However, keep your labels and legends clean and only insert the essential information. For graphs, labels within the graph area are preferred over external legends as this makes it easier for your audience to interpret your graph. For maps, a separate legend is often more suitable.

4.3   Colour use in data visualisation

Using colour in your figures can be a very powerful tool in making your figures clearer for your audience. However, misuse use of colour can lead to an inaccurate representation and manipulation of your data, thereby misleading your audience and yourself[20]. In this subchapter, you will find some guidelines for using colour in your data visualisation figures (graphs and maps). Although colour use can also play an important role in other figures types, *e.g.*, process visualisation and infographics, the information below is specifically applicable to figures that visualise data.

4.3.1   General tips for colour use

Only add colour with a reason

When you want to use colour in your figures, always ask yourself first if there is a reason for it[13]. Some good reasons to use colour in your figures is to add a third axis to a 2D scatter plot, *e.g.*, to indicate different categories in a scatterplot, or to indicate different features in a map. However, if adding colour will just add visual interest to your figures, keep your figures in black and grey tones as adding colours without a specific reason will only be confusing for your audience.

Keep it clear

Do not use too many colours since this will distract from your message, and it is difficult for colour-blind people to distinguish between them. It is important that your colours are easy to distinguish from each other and the background.

Do not use default colour settings

If you do use colour, think about what colours to use. Default colours will often look unattractive and can make your figures look cluttered[19]. Moreover, our human perception of colour is complicated, and colours often have an intuitive meaning to people depending on the context[21]. For example, on a map, people will intuitively interpret blue as water, whereas in a temperature plot, blue intuitively indicates low temperatures and red indicates high temperatures. Take these connotations into account when you are creating a figure, and do not just use default colour settings of the software you are using as this may complicate your message. If you want to highlight certain features, it is best to use 'pure colours' to highlight important aspects in your figures[21]. For an example of intuitive colour use in a map, see Fig. 1 in this paper by Prebble *et al.* (2016).

Use colours that are distinguishable for visually impaired people

About 8 to 10% of the population is colour blind[22]. Avoid green-red palettes as these are not distinguishable for visionally impaired people. More and more thoughtful colour palettes are available in colour packages for programming languages and other software. Pay extra attention to this when you are making your own colour palette. As an alternative, it's

possible to use different patterns instead of different colours. Be aware, however, that many patterns may unnecessarily clutter your graphs, and it may be sufficient to critically choose one of the many other available colour palettes.

4.3.2 Colourmaps

When you are using multiple colours in your figures, you can use a colourmap. Colourmaps are a premade selection of colour gradients that you can apply easily in your figures. Generally, three categories of colourmaps exist[13]:

- **Sequential colourmaps** are appropriate for visualising sequential data. They can consist of different hues of one colour, or an intuitively increasing sequence of colours.
- **Diverging colourmaps** indicate deviation from a standard, or median value. They generally consist of two contrasting colours decreasing in hue to the middle.
- **Qualitative colourmaps** are appropriate for visualising different categories (discrete numerical data or categorical data) instead of continuous numerical values. Therefore, they consist of different colours without a gradient in between.

Depending on what software you use, you can add colour libraries to apply colourmaps to your figures. Always use a colourmap that is appropriate for your data. This is especially important when you are visualising **gradients** in maps or graphs. Be aware that although many colourmaps exist, not all of them are good to use.

Colourmaps to avoid: rainbow

For categorical data, the exact type of colourmap you use does not matter as long as it is suitable for visionally impaired people and acknowledges any possible intuitive colours for your data. However, for sequential data there are many more things to consider. Often, scientist still use rainbow colourmaps to visualise sequential data. Although these might be visually attractive, this type of colourmap has several problems.

1. The arrangement of colours does not follow an **intuitive order**: to the human eye, yellow is the brightest colour. Therefore, we automatically interpret values with this colour as highest or most important. However, in the rainbow spectrum, yellow indicates intermediate values[20].
2. Not all shifts between colours in this type of colour appear equal as some look more abrupt than others. This results in an inaccurate representation of underlaying (continuous) data with a consistent rate of change.
3. Abrupt changes in the rainbow colour scale are intuitively interpreted as a meaningful boundary, while in the data there is none[21].
4. The uneven colour gradients in the rainbow colourmap are unreadable to people with colour-vision deficiency[20].

As graphs and maps are not only used to communicate information, but also to analyse data, this can lead to wrong interpretations. You might also miss important details in your data. Therefore, choose a suitable colourmap for your data and figure.

<u>Use appropriate, scientifically derived colourmaps</u>
Good colour bars for sequential or diverging data need to contain colours in an intuitive order and even colour gradients, *i.e.* a homogeneous gradient without any abrupt changes. Scientifically derived colour maps with a balance of intuitive sequences of hue, luminance, and saturation now exist. These accurately represent data variation, reduce complexity, and are appropriate for colour-blind people[20]. Do not change the scale of a scientifically derived colour bar as this would distort the data, similar to when you would change the spacing between some of the axis ticks of an x-y graph[20]. Remember to always add a colour bar in figures that use a colour scale.

We recommend you to read <u>this paper by Crameri *et al.* (2020)</u> on the misuse of colour in science communication. It contains several examples of how to (not) use colour in scientific figures, and a flow-chart to decide what type of colourmap or scheme you need for your figure. It also contains several links to downloads of scientifically derived colourmaps to use in your software of choice.

4.4   Tools for creating figures

 In this subchapter, we present some useful programs for making scientific figures in Earth sciences. We make a distinction between software for 1) creating data graphs, 2) creating maps, and 3) creating schematic figures and figure editing. The necessary program depends on what type of figure you want to make.

Often, you will use more than one type of software to produce your final figure. You can visualise your data into graphs by using Excel or a programming language for example, and then finalise your figure using a vector graphics editor. This will result in more attractive figures and will save you from fine-tuning figures using a programming language, which can quite time-consuming.

4.4.1   Creating graphs
<u>Excel</u>
You can use Excel to plot data in simple graphs. Often, students use Excel when they have to make a graph as it is a familiar interface and it is therefore easy to use. When you are already handling/structuring/saving your data in Excel, you can easily and quickly plot your data to explore it, and spot if you have made any big mistakes.

Using Excel to make scientific graph has some disadvantages however: the graph options in Excel are limited to the default functions, and there is no option to add packages for specific, more advanced graph types. If you want to make more advanced graphs, it might be better to use a different program. Additionally, layout options are limited. The standard layout is well-known and might look unprofessional. Therefore, we recommend that you not use the default style. Instead, develop your own style with an appropriate colour scheme.

A license for MS Office 365 is available for free through UU with a Solis-id.

Resources:
- The *Excel* module offers video tutorials for making graphs in Excel, specifically scatter plots and histograms in Excel.
- For a step-by-step guide for creating graphs in Excel, see this website.

4.4.2   Creating figures (graphs) using programming languages

Using a programming language to create your figures might initially take more time than creating your figures in Excel, especially if you do not have any programming experience. However, making your graphs using a programming language has several advantages over the use of MS Excel. It is worth the time to learn it, especially if you want to make more complicated figures, *e.g.*, during your MSc thesis or when you are writing a manuscript for publication. Note that some supervisors are fine with the use of Excel figures, but others are not, so discuss this with your supervisor. They might also have tips for what programs and packages work best for your data and research.

The main advantages of using a programming language for creating figures are:
- Using a programming language will give you a lot more freedom in the layout and format of your figure.
- Using code to make your graphs allows you to save a script with command lines to produce a graph. This makes your workflow more transparent. Moreover, you can later use the same script when you want to make a similar figure with different dataset, saving you time.
- A lot of packages and libraries for specific scientific fields and graph types are available on the internet, including help pages.
- Some languages allow you to create 3D or animated figures and maps.

Several programming languages exist, and people might prefer different languages. However, each language has its own point of focus and advantages. Useful packages of the three most used languages at the Earth Sciences department at Utrecht University will be discussed below.

Python

Python does not have its own built-in function for data visualisation. There are however several plotting libraries available that allow you to create high-quality figures. This gives you many possibilities for creating graphs. An extensive library with example plots and corresponding code is available [here](). When you use the python software (see *Programming* module), the website also introduces the python syntax to produce such figures yourself.

Libraries:
- The **Matplotlib** library has functions for 2D and 3D plotting, and for interactive figures. It provides a lot of freedom for data visualisation. Various other data visualisation libraries are built on Matplotlib.
- **Pandas visualisation** is developed for data manipulation and data cleaning, but also has some visualisation features. It is built on Matplotlib but has a more user-friendly interface and you will need to write less code to make your graphs.
- **Seaborn** is suitable for advanced data visualisation. It has a lot of built-in functions and is particularly useful for visualising time series. It also has some built-in styling formats.
- **Ggplot** is based on the ggplot2 package, which is built on the grammar of graphics. This makes sure your graph contains some basic elements and simplifies your code.
- **Plotly** is an online platform for data visualisation, but has also an extension for Python and R. It allows you to create interactive plots and 3D graphs.

Python and the libraries mentioned above are all open source. For more information on making figures in Python, you can take a look at the *Programming* module. Many free tutorials, including complete beginner courses, are available online.

R

R is a programming language focused on statistical computing. It is connected to RStudio, which is a software with a more user-friendly interface that has several built-in functions for data visualisation. R and RStudio are specifically designed for statistics and data analysis. However, various additional data visualisation packages are available.

Packages:
- **ggplot2** is one of the most popular visualisation packages. It is based on grammar of graphics (R4DS). This package makes creating custom plots fairly easy, because it takes care of small details so that you do not have to write commands for those.
- **Lattice** is specifically useful for visualising multivariate data. It allows you to create plots with different variables in panels.
- **Analogue** is developed to work with paleoecological data. It is specifically useful to create species composition diagrams, as a part of stratigraphic plots.
- **RColorBrewer** helps you to easily manipulate colours in your graphs. It has several sequential, diverging, and qualitative colour palettes.

- **Plotly** is an online platform for data visualisation, but has also an extension for Python and R. It allows you to create interactive plots and 3D graphs.

R, R studio and all packages are open-source. Many free tutorials and books are available online.

Tutorials/books:
- The website 'R for data science' offers a good, free, e-book with step-by-step tutorials to teach you the most important data science tools in R, including a chapter on basic visualisation techniques in R using ggplot2.
- The website 'R for Paleolimnology' offers similar step-by-step tutorials for visualizing paleolimnological data, including creating stratigraphic charts.

MATLAB

Matlab is mainly used for data analysis and modelling. The MATLAB visualisation app allows you to visualise data, and to view and explore data using interactive visualisations of MATLAB plots.

In contrast to Python and R, you need a license to download Matlab. As a UU student, you can use it for free by using SolisWorkspace, or on UU computers. You can also download it for free by using your @students.uu.nl email address. You can find more information on this here.

4.4.3   Creating maps
GIS software is the best option to create your maps in, especially if you are also including data.

ArcGIS pro

With ArcGIS pro, you can visualise, analyse, and explore spatial data. It also has functions for making 3D scenes. ArcGIS online is a cloud-based mapping and analysis software, which you can use to share and collaborate in making maps and analysing data.

ArcGIS pro and ArcGIS online are not open source, but you free access with your Solis-id through MyWorkplace or download it through the GIS Community on Blackboard. See this website for more information.

QGIS

QGIS is an open source alternative to ArcGIS. Tutorials are available through the QGIS website and through GIS OpenCourseWare.

4.4.4 Creating figures and finalising graphs and maps

Vector graphics software is useful for editing and finalising graphs you produced in Excel or using a programming language; when you want to combine various figure types, *e.g.*, data plots, photographs and schematics in one figure; and when you are making more extensive figures such as stratigraphic charts. Fine-tuning your graphs and maps will also be a lot easier and less time-consuming when you use a vector graphics editor.

Using vector graphics software has two main advantages. Firstly, it will keep the separate layers you add in your graph apart instead of flattening them into a collection of pixels[23]. In this way it is easier to keep editing/adapting your figure. Secondly, it will make sure your figures will have a good resolution and will not be pixelated in your end-product.

You can use vector graphics editors to:
- **Create** schematic figures or infographics
- **Adjust** graphs and maps (data visualisations) you made in other programs (Excel, Python, R, Matlab) easily
- **Add elements** such as scale bars etc. to field or microscope photos
- **Adapt** figures from scientific papers or your own work for presentations, *e.g.*, removing superfluous elements
- **Produce** scientific posters

Scientific figures can generally be separated into **vector** and **raster/bitmap** figures.

Vector images consist of mathematically defined lines and curves. Therefore, they these figures can be scaled unlimited without becoming blurred and losing their original colour and proportions. They are however not suitable for realistic drawing as they come across very clean and artificial.

Raster images on the other hand consist of pixels, I.e. coloured rectangular dots. At high resolutions, they look very realistic, and you can adapt colour and shape very detailed. However, their quality decreases when you edit/manipulate a raster image after creation. They cannot be scaled unlimitedly: When you enlarge them too much, they will become blurred and when you shrink them too much, the lines and colours will blur into each other.

For the creation of scientific figures in Earth Sciences, vectors are generally preferred. However, for photos, *e.g.* from a field site, raster images are more suitable.

Adobe suite (Illustrator)

Adobe Illustrator is a well-known and often-used program for creating and editing vector graphics. If you are already familiar with it, it is a great tool to use. However, adobe suite software is not open source. It is installed on all public UU student computers, but the UU does not offer a free download for your own computer. However, you can get a discount when you download it through SURFspot. If you want to use graphic design software on your own computer and are not yet familiar with Adobe Illustrator, it might be better to use open-source software which offer similar possibilities, *e.g.*, Inkscape.

Inkscape

[Inkscape](#) is an open-source program for creating and editing vector graphics. This makes it specifically useful in creating schematic figures, *e.g.*, infographics and process visualisations. It is also very suitable for editing and finalising graphs you made using other software and combining them to make complex and professional scientific figures.

Inkscape has similar functions to Adobe Illustrator but is free of charge and you do not have to sign up. If you are not yet familiar with Inkscape, it might take a bit of time to learn it due to the large number of functions the software contains. However, it has a user-friendly interface, so learning is easy. Moreover, it has many more functions than MS PowerPoint, resulting in better and more professional figures. Therefore, it is worth spending some time on learning Inkscape.

Many tutorials for Inkscape for beginners are available online. Inkscape also has various written and video tutorials available on [their website](#).

Microsoft PowerPoint

PowerPoint is not primarily designed for designing and editing figures and therefore offers fewer possibilities than other vector graphics software. However, PowerPoint includes some shape tools that you can use to visualise processes or create infographics. You can also use it to edit or combine graphs you made in other programs, and to add lines and vectors to them. Similarly to more advanced vector graphic editors, PowerPoint is vector based, *i.e.* it keeps the layers/elements of your graph separated. This means you can move them independently and save your figures as in a vector (SVG) format as well. This makes it a much better alternative to Microsoft Paint. Thus, PowerPoint is a good alternative to Inkscape or Illustrator if you do not have the need to create very complex figures and do not have sufficient time to learn Inkscape or Illustrator.

MS Office (including PowerPoint) is not open source, but you can download and use the office 365 version for free with your Solis-id and password. Most students are already familiar with MS PowerPoint and creating/editing figures works quite intuitively.

GIMP

GIMP is developed for editing raster images instead of vectors. It provides tools for high quality image manipulation. For Earth scientists, it is specifically useful to edit field or microscope photos, but not for editing or creating other scientific figures.

GIMP is open source. It also offers many free tutorials on [its website](#).

5　General tips for implementing figures

> **Learning outcomes:**
>
> At the end of this chapter you will:
>
> - Understand where and how to most effectively implement figures in your work

5.1　The dos and don'ts of data visualisation

The following techniques can be generally applied to a variety of visualisation techniques. When creating a figure, consult this list to make sure you've covered the basics.

1.　Keep it simple, avoid clutter

For figures to convey your messages most efficiently, they can't try and explain too many ideas at once. Make sure you have a specific message or purpose in mind for your figure – this makes it easier to cut down on any redundant aspects. Consider how an observer without your background knowledge would approach your figure for the first time. Be aware that this is more difficult than it seems. Scientists are often afraid that by making a simple figure, their work might seem basic or uncomplete. However, conveying your message correctly and accurately to your audience should always be your first priority.

2.　Use high resolution figures/photos

As word processing software keeps our text looking sharp, we need high quality figures to go along side it. There are three common formats that our images are saved in. JPEG is best for saving photographs but should be avoided for graphics (such as lines, drawings, typography). For scientific figures, PNG is an ideal format as the details of the original image or conserved after compression. PNG also supports transparent background layers. When creating images using graphics software, the best format is SVG which scales endlessly while conserving sharpness.

3.　Practice consistency in form

Whether they are multiple figures for an essay or presentation, the look and feel should be consistent throughout. Make sure the same font and sizing can be seen across different figures. If one colour convention is used, make sure to remain consistent.

4.　Maintain diversity in message

Though the figures should look the same, there should be diversity in what they are trying to explain. Avoid creating figures that express the same conclusion. This adds redundancy, taking up space in your work without helping to explain anything new. Give each of your figures a purpose, and make sure you employ them to tell your whole story.

5.2   Adapt your figures to the medium you are using

You also need to adapt your figures to the medium you are using (see table below). On a PowerPoint in an oral presentation, your audience will look at your figure in a different way than in a scientific paper. For oral presentations, you want to keep your figures simple and you want your message to stand out. Your audience will only look at your figure for a short time, and from a distance. This means that your graph needs to contain less details and needs a bigger font size and line thickness for an oral presentation than for a scientific paper[13]. For an example of this, see Fig. 3 in this paper from Rougier *et al.* (2014). It also contains other tips for making graphs.

If you are making a figure for a specific journal, always start with checking the boundary conditions for size and work from there. This prevents your figure from becoming unreadable when it is shrunken by the journal editor to fit the size restrictions of the journal. Also check in what format the presented text and figures need to be, *e.g.,* in A4 or A3 format.

Description	Complexity	Font	Referencing	Placement
Essays The most formal implementation of figures is found in thesis and journal writing. There are very specific rules from copyright and referencing to formatting. For a complete overview on how to incorporate figures in your essays, including citation and caption guidelines for the geosciences, see the Academic Writing module.	Allows for the most complexity.	Font size should be the same as your main body text size. Generally, around 12pt.	There is completeness and concreteness in captions. They are located below figures and above tables.	Figures should not be free-standing. They appear after being mentioned in-text.
Presentations Figures play an important role in a scientific presentation. They aid in simplifying complex processes and are invaluable in keeping your audience engaged. For a complete overview on how to incorporate figures in your presentations, including best practice guidelines, see the Presentation module.	It is more important to emphasise clarity in a presentation figure than complexity. Keep the audience in mind.	Font size needs to be large enough to be seen across the room. Generally, 24pt.	Figures should include a small caption with the source author, date of publication, and if it's been edited or adapted. A full APA-formatted bibliography goes on the last slide.	Figures may be accompanied by minimal text or stand on their own on a slide.
Posters As opposed to in presentations where figures and the presenter work in tandem to communicate a message, in posters the figures often have to stand completely on their own. This requires them to be self-explanatory but not overly cluttered or complex.	There can be quite some complexity in a poster figure, but the intended audience must be kept in mind.	Font size needs to be large enough to be seen across the room. Generally, 24pt.	To decrease clutter, figures in posters may be simply numbered. The numbers correspond to a citation in a reference section in the bottom right column.	Text and figures are often equally distributed.

6 References

[1] Wingens *et al.*, 2018: Science Communication in a Complex World

[2] Ostergen *et al.*, 2013: Paper: How Scientists Develop Competence in Visual Communication.

[3] Nayak and Isawa, 2019 Paper: Preparing scientists for a visual future

[4] Visualizing geospatial data on a map

[5] Visualization analysis of geotemporal data

[6] Open Topography

[7] Visualizing temporal data in GIS

[8] What is temporal data?

[9] Basis statistical concepts and methods for Earth Scientists

[10] Writing Center UNC

[11] Communicating results with scientific graphs - Queensland University

[12] Springer manuscript guidelines

[13] Rules for better figures

[14] Data visualization catalogue

[15] Misleading graphs

[16] Weissgerber *et al.*, 2019

[17] Geoscience videos: How to make them and how to use them in the classroom

[18] Designing and creating science infographics

[19] How to improve the quality of your scientific figures

[20] The misuse of colour in science communication

[21] Why scientists need to be better at data visualization

[22] Colourblindness

[23] Learning Inkscape