Slide 1 – Title Slide

Hello and welcome to Week 5, Part 2 of EGM101: Presenting Data. In this lesson, we'll discuss how we can display and present our data to other people.

Slide 2 – Why do we present data?

Why do we want to present our data, though? Because we're not doing our research in a vacuum, or off on our own in some basement lab somewhere – we have to be able to communicate our research to other people. You will definitely have to communicate your research to other researchers, or to your lecturers as you progress through your course. But, you will most likely also need to be able to communicate your research to people who aren't specialists – it's important to keep your "audience" in mind when you're planning how you're going to communicate.

As part of planning how you're going to communicate, you also need to think about why your audience should care – what is it about your research that's important to the audience?

When you are communicating your research, you need to be able to explain it clearly to the audience – they need to know basic things like what your research is, what you found out in your research, or your "results" and why those results matter.

Presenting data is a big part of being able to effectively communicate with an audience, and that's the topic of this lesson – how we present data, and ways that we can do this in a way that helps get your message across.

Slide 3 – Ways of presenting data

In this module, we're going to discuss three main methods that we have of presenting our data, beginning with written reports. Written reports are arguably the most common way of communicating data that you will encounter, because it encompasses forms like press articles. Open up your news website, or newspaper, or even news TV program, and you will very likely see a number of articles presenting data in some form or other.

They might be reporting on diverse topics such as economic conditions, government policies, or scientific research, but they will in all likelihood include some sort of presentation of data.

As researchers, you will also encounter reports in the form of peer-reviewed articles – either those that you read in the course of your studies or career, or those that you write.

Another place that you will find data presented is in so-called "gray" literature, which is a term that covers reports or other documents produced by government agencies, institutions, or commercial companies that aren't peer-reviewed.

All of these written forms of presenting data also make use of elements such as tables or charts, figures, or graphs – in fact, this press article includes a bar chart showing how the energy price cap has risen since 2018-19; most written reports that present data will include some kind of visualization like this.

For some particularly impressive examples of data visualization, I recommend browsing the "dataisbeautiful" subreddit at https://reddit.com/r/dataisbeautiful – there, users post any number of visualizations that they have made themselves.

Slide 4 – Tables: general rules

A table is a form of data presentation where the data is presented in some sort of grid – here, we have a table showing the area of the world's oceans, with one column for the area in thousands of square kilometers, and another column for the area as a percentage of the entire ocean surface area.

When you present your data in a table, it's important to remember to include either a title that explains what the table is showing (if you're presenting it in a way where it's on its own), or with a caption and a label if it's part of a report or article.

You will also need to remember to include headings that explain what each row or column represents. And you should always remember to include the units for each column – without this information, it's very difficult for your audience to understand what the data in the table actually represent.

And, because we're often showing numerical data in a table, keep significant figures in mind! You most likely won't need to include 10 decimal places for each of the values in the table – you should only include numbers that are actually needed.

Slide 5 – Charts/figures/graphs: general rules

Before we get into different kinds of charts/figures/graphs, we'll start by introducing some general rules to keep in mind when you make them, using this wonderfully generic graph.

Starting from the top, the first rule to remember is that the graph needs some kind of explanation – either a title, if the graph is presented on its own, or with a caption, if it's included in a report or article.

It's also very important to remember to label the axes of the graph, including the units displayed on each axis. Just like in a table, without this information, it's very difficult for the audience to make sense of what you're trying to show.

Finally, just like on a map, if we have multiple lines, or multiple symbols showing different variables, then we need to have a key or legend that explains the colors or symbols.

All in all, these rules could also just be summed up as "make sure that you label things" so that your audience knows what they're looking at.

Slide 6 – A word on color

Before moving on, I also want to say something about color. Color is an incredibly effective tool for communicating in charts and figures. But, like all tools, it can be misused or abused. Not only that, but it's important to remember that not everyone sees color in the same way – around 1 in 12 men and 1 in 200 women have some form of color vision deficiency, which means that they have difficulty distinguishing between different colors – most often, between red and green.

So, when you are putting together your beautiful figures for your reports or your dissertation, try to avoid using color maps that distort, like in the example shown here, from this 2020 paper by Crameri and others. The commonly-used “jet” color map, which ranges from dark blue to dark red, is a great example of a color map that distorts color. Have a look at the original black and white images shown here, and compare those to the same images presented using the “jet” color map. You can still make out that this is an image of a woman (Marie Skłodowska-Curie) or an apple, but you might have trouble picking out that this is an image of the Earth from space. It is also more difficult to pick out facial features – the sharp red/green boundary on the woman’s cheek is not at all apparent in the original image, or in the non-distorting color map shown here on the right.

Make sure that you avoid using inaccessible color combinations – if your audience has trouble distinguishing between red and green, they will have a harder time picking out the details that you might be hoping to show in your figure, or they may see something completely different to what you had intended.

Above all, though, think about the data that you are presenting, and what you are trying to show – be sure to use color effectively to help you communicate!

Slide 7 – Bar charts

The first type of chart that we’ll introduce here is the bar chart, where data values are plotted as a series of bars, where the height or length of the bar is proportional to the value it represents. This example here shows the Ocean Area example that we saw in a table earlier.

Looking at this, we can quickly see that the Pacific Ocean is about twice as large as the next largest ocean, the Atlantic and Indian Oceans are about the same size, and together these three Oceans make up most of the global ocean area.

There are a number of other variations on the basic bar chart shown here, including grouped bar charts, where we group bars belonging to different categories; component bar charts, where each bar has sections with different colors/shading depending on the proportion of each bar represented by different variables, or percent component bar charts, where each bar is the same size, with sections sized according to the percent of the total that each variable represents.

Most importantly, though, bar charts tell a story about magnitudes. If we look at this example here, the way that I have cropped the y-axis makes it look like the Atlantic Ocean is 3 to 4 times larger than the Indian Ocean, which is very clearly not true. By only showing a portion of the entire bar, though, I’ve managed to distort the picture in a very misleading way. We want our communication to be effective, not misleading, so make sure that the axis of a bar chart always shows the entire bar.

Slide 8 – Scatter plots

Scatter plots are used to display paired values of two (or more) variables. In the top example here, I’ve plotted values of creek stage at different points in time as black circles, and values of discharge as blue squares. From both of these examples, you can see how when values of discharge increase, the stage

goes up. This makes sense, since more water flowing in a river typically means that the water level goes up. This is an example of a relationship between two variables, something that scatter plots are often used to help illustrate. Of course, there may not be a relationship between the two variables, but that's something that we'll talk about more next week.

This bottom plot is an example of a logarithmic scale, which is something that you will often see when the values of a variable cover a very large range. Typically, you will see this as a "log 10" scale, where the values plotted correspond to the base-10 logarithm of the data. It's important to note that the ticks aren't evenly spaced on a logarithmic graph – you can see on the horizontal axis here how the gap between 10 and 50 is much larger than the gap between 50 and 100.

Slide 9 – Line plots

Very similar to scatter plots, line plots are used to communicate information about changes – either between two variables, or over time.

The example here shows how global annual temperatures have changed over the period 1850 to 2022, relative to the 1961-1990 baseline. The gray shading on either side of the line shows the estimated error of the estimate. From this plot, we can quickly see that the global annual temperature now is about 1 degree warmer than it was in the 1960s, and almost 1.4 degrees warmer than it was in the 1860s.

In this plot, we have values for all of the years covered by our dataset, so there aren't any gaps. If we had more sporadic observations, it wouldn't make sense to connect the observations with a line. As a general rule of thumb, if you're not sure if there's a relationship between your variables, or there are gaps in between your observations, you should stick to a scatter plot instead of a line plot.

Slide 10 – Bo(& whisker) plots

If we want to show the concentration, or spread, of data, we can use a bo(or boand whisker) plot. The boxes in the boand whisker plot show the middle 50% of the data, something called the interquartile range. The line in the middle is the median, or the exact middle value of the data. If you're not familiar with these terms, don't worry too much - we'll cover both of these terms, and more, in the coming lessons.

From this plot, we can see that the discharge in this creek was very high in 2013 and 2014 based on where the median values are, but we can also see from the size of the boxes that there was a wide range of values in both years, whereas in years like 2016, there was very little variation in either variable.

The other part of the boand whisker plot is the whiskers, which are typically either lines showing where the maximum and minimum values are, or lines indicating 1.5 times the interquartile range, with values outside of that range plotted individually as outliers, or values that don't quite fit in with the rest of the data.

Slide 11 – Pie charts

When you want to show what proportion of a whole different categories represent, you can use a pie chart. The most important thing to remember about using pie charts is that if your data don't add up to a whole of something, you need to pick a different kind of visualization.

The example here shows the percentage of the world's ocean surface each ocean represents. Just like we saw with the bar chart, we can see that the Pacific Ocean represents the largest slice, at nearly half of the area, while the Atlantic and Indian Oceans cover a roughly equal area, and the Southern, Arctic, and "other" ocean surfaces filling out the remaining slices.

A commonly-encountered issue with pie charts is that humans aren't always great at spotting small differences in angles. Without the values printed on the chart here, I don't know that I could tell the difference between the orange and the green "slices" of the pie – usually, it's a good idea to include the numbers to make interpreting the chart easier.

Alternatively, you can use a bar chart, as we saw earlier, or something called a tree map, which works the same as a pie chart, but with rectangular pieces.

A final note about pie charts – if you use them, be sure to only use two-dimensional pie charts like the example here. As discussed in some of the suggested reading at the end of this lesson, three-dimensional pie charts can distort the areas shown, which makes them a great way to mislead or lie with your data, but not so great for effective and honest communication.

Slide 12 – Summary

In this lesson, we've discussed how a big part of research is communicating the results to different audiences.

We've also discussed how there are a number of ways to present/communicate data effectively, including reports, tables, and charts/figures. These are all important tools that we can use, and we'll get plenty of practice with them over the course of this module.

When making charts/figures, remember to choose the right tool for the job, as different charts display information in different ways. Always remember to label your charts to make them easier to interpret, and finally, try to think about color, and how you can use it effectively and honestly to convey your message.

Slide 13 – Additional resources

You can read more about the topics we've discussed here in the textbooks – Illowsky and Dean, Chapters 2.1 and 2.4; Caswell, Chapter 5; and Weiss, Chapters 2.2 – 2.5.

In addition to the textbooks, I also recommend reading these chapters from two other books on the reading list, both discussing how graphs and data visualization be used to mislead or deceive. These books aren't math textbooks, but they talk about how math and statistics more specifically can be used

deceptively. They're both quick reads and available from the university library, or you can find the scanned copies of these chapters on blackboard.

I've also linked to two articles here – the first, from Scientific American, includes a bit more information about the importance of effective communication for science (and research more generally). The second is a paper that discusses the use (and misuse) of color in scientific communication, which we briefly touched on earlier in the lesson.

That's all for this lesson – I hope you found it interesting, and you have any questions, please don't hesitate to e-mail me or post in the discussion forum on blackboard. Bye!