Transcript of Bad Graphs: Assessing Graphs for Readability, Fairness and Impact

Robin Marwick

robinmarwick.ca

22 October 2024

00:00:03:29 - 00:00:34:05
Speaker 1
Thank you so much, Michelle, for the lovely greeting. And hi everybody. I'm. I'm really excited to be here today. Let me share my screen. Let me know when you can see. Oh. Oh, good. Okay, let's get started. So. Yeah. Why are we here? When I tell people I give presentations on graphs, I get two reactions.

00:00:34:07 - 00:01:05:10
Speaker 1
Either people get really excited, or they look completely blank and ask why. And I'm hoping the fact that you're here, thank you again, means that you're at least leaning towards the first group. We are living in an era of metrics. Graphs are everywhere. These days, science, social science, business, finance, education, government, journalism. Some of them are good, and some of them are, not so good.

00:01:05:12 - 00:01:28:15
Speaker 1
And we will be looking at those later. Our job as communicators includes both the words and the pictures, which means that part of what we're hired to do is make those graphs as good as they can be. I am an editor, and I normally give this presentation to other editors, but effective communication is a team effort, and I hope you'll be able to use at least some of what you'll learn today.

00:01:28:16 - 00:01:51:29
Speaker 1
No matter what your role is, what I hope to do today is equipped you with some tools so that you'll feel more confident that you can spot problems with a graph and make suggestions on

how to fix it. Or even take a basically okay graph and suggest a few tweaks to make it better. Or indeed create good graphs from scratch.

00:01:52:02 - 00:02:23:03
Speaker 1
Quick math disclaimer there's probably a fairly wide range of math knowledge, in this audience and comfort, as in any audience. You actually don't need that much math to work with simple graphs anyway. Just what I call everyday grade school math. It's just as important to have a solid. “Does this make sense?” meter. I have tried to make the math minimal in this presentation, but if I say something that makes no sense to you at any point, let me know.

00:02:23:03 - 00:02:51:06
Speaker 1
Michelle mentioned that I'll be answering questions at the end, but yeah, if you need a clarification on something, please, you know, shout out in the moment and I will try to do that. So let's take a step back. What is a graph? What does it form? Basically a graph is a visual that shows a relationship between two or more sets of data.

00:02:51:08 - 00:03:26:11
Speaker 1
That makes graph sound really boring, but a good graph can catch the audience's attention and show patterns and trends in a way that has immediate visual impact. I could tell you I have eaten 27% of this pie and it was delicious. Or I could show you with a graph. There are basically two types of data. There is qualitative or categorical data, meaning ways to describe or classify things, and quantitative or numerical data, meaning stuff you can measure or count.

00:03:26:13 - 00:03:47:26
Speaker 1
On the left are some examples of categories. Think of them as buckets. They can be pretty much anything, including numbers or number ranges. And on the right I have some examples of numerical data, things that you can count or measure or do math on. All of these would be numbers with or without units of measure.

00:03:47:28 - 00:04:23:27
Speaker 1
A good habit to get into is when you look at any graph, figure out what question it's trying to answer and what data it's using to try and answer it. And specifically, what are the categories and what are the numbers that are being looked at? Something else to keep in mind is that all graphs are the outcome of a series of subjective decisions that were made along the way to to whittle down the vast richness of reality into a visual display about a tiny portion of the world that's hopefully informative and accurate.

00:04:23:29 - 00:04:50:16
Speaker 1

Once you have data from somewhere, there are a bunch of steps you need to make it, usable or useful to to a reader. So you hunt around the data for interesting data and relationships, and you find a way to turn it all into a visual that tells a story. This is quite a lot to ask of a graph, so it's no wonder that sometimes they go a bit wrong.

00:04:50:18 - 00:05:14:20
Speaker 1
This cartoon is about entire scientific papers, but it's equally true of graphs. So while you're working with a graph, it can be good to take a step back from time to time and think, okay, what decisions led to this point? How do we how do we know this information? Is there any way that this graph could be biased or skewed or have missing information or stuff?

00:05:14:20 - 00:05:40:23
Speaker 1
We don't know. There are a few general principles for looking at graphs with your communicator hat on. I'm going to review those and then we'll try those skills on some examples. Just like we do for text, it's important to keep the medium and the audience in mind when we're assessing graphs. What's right for one situation may not be right for another.

00:05:40:26 - 00:06:10:08
Speaker 1
Scientists, for example, are used to interpreting graphs, and they like a lot of detail and complexity. Whereas a lay audience might just need the big picture, a print document with a highly motivated reader can have more detail than, say, a presentation where you only have a few seconds to make an impression. And people who are reading just for interest have different needs than people who need to make decisions based on this information.

00:06:10:11 - 00:06:31:14
Speaker 1
So when you start assessing a graph, take in the overall effect. What's your first impression? Where is your eye drawn first and is the most important information in the graph at that point? Is your eye having to bounce around the graph to make sense of it all? Do you have to keep one finger on the legend to see what's going on?

00:06:31:17 - 00:06:55:03
Speaker 1
An instructor I had called this reading with your fingers, which I thought was great. How easy is it to understand? And if you're working with a document, if there are multiple graphs in the document, do they look like they belong together? Your next step is to look at what the graph means. What are the graph and the text it supports trying to achieve?

00:06:55:03 - 00:07:23:11
Speaker 1

And do they succeed? What relationship is the graph describing? What's the source of the data? Is the data appropriate and accurate? Is the type of graph appropriate? Is the presentation appropriate? Do you have the axes, the scale, the data labels and so forth? Are those all working? Is there context that's maybe missing that people will need to make sense of, of this graph?

00:07:23:13 - 00:07:47:06
Speaker 1
And finally, it's time to make sure that all the information in the graph of versus accurate. You need to cross-check all the numbers against the data table. Hopefully you have that and the rest of the document. You know, if the document says 38%, then you want to make sure that the graph says the same thing. Do the descriptions of the graph in the document match what's actually in the graph?

00:07:47:08 - 00:08:22:18
Speaker 1
Are the axes in the data points correctly labeled? Are there outliers? Which means a point that or a few points that are that are way off the to sort of general trend of the graph. Are they really outliers? In which case, what caused them or are they mistakes in data entry? Maybe, I should mention as well that graphs live somewhere in the intersection of statistics, graphic design and communications with a side of technology.

00:08:22:20 - 00:08:52:03
Speaker 1
There are a ton of different ways to create graphs and data visualizations. As an editor, what this often means that unlike with text, we may not be able to fix a problem graph directly. We may not have access to the program that was used to create the graph or the technical skill. So yeah, over over on the side here, you know sure I have Excel, I have Google Docs and Sheets.

00:08:52:05 - 00:09:29:06
Speaker 1
But these are, you know, getting into, you know, sort of more specialized tools and all the way into actual programing languages. And I, as an editor, I kind of touch those. So, if I'm working closely with an author or a writer and I have access to their files, sometimes I can make changes directly. But, yeah, if an author or a designer is using something from the right hand side of this graph, or if they just send images, I can't necessarily go in and just quickly fix something.

00:09:29:06 - 00:09:58:11
Speaker 1
So all I can do is make suggestions and query. All of this is to say that making graphs is often very collaborative work and making changes. If you didn't create the graph, may take some care and tact. For editors in particular, the more you can show that you understand what the author,

artist, or designer is trying to achieve and that you want to help with that, the better things will probably go.

00:09:58:13 - 00:10:24:09
Speaker 1
Another complication that can arise is that some workflows keep words and visuals very separate until quite late in the editorial process. So if possible, you need to have several points from start to finish where everyone looks at the words and pictures at A as a whole. Because if it all comes together right at the end, you may have reviewers saying, oh wait, no, this is totally the wrong graph.

00:10:24:09 - 00:10:49:26
Speaker 1
We should use this. And like, no, I precious timeline. Or you may not have a chance to fix things that really should be fixed. Okay, so what are we looking for in a graph, exactly? What makes a good graph? Actually, this talk is called bad graph. So let's lower our standards a bit. What are we looking for in a baseline?

00:10:49:26 - 00:11:20:10
Speaker 1
Okay. Graph. There are basically four elements. First it's accurate. Obviously, mistakes make everyone look bad and damage the documents credibility. So make sure your graph doesn't have mistakes. Second, it's readable. You don't have to squint or stand on your hand or contort your neck to make sense of it. That sounds basic, but it's astonishing how often it doesn't happen.

00:11:20:12 - 00:11:48:25
Speaker 1
And third, it's fair. What I mean by that is, sure, a graph can have a point of view. All communication has some kind of goal, but it's important to present the available data in a way that doesn't distort the facts. Graphs are amazing persuasive tools, so it's important that they're used responsibly. And finally, the graph needs to support the story you're telling or the point you're trying to make.

00:11:48:27 - 00:12:10:24
Speaker 1
What do I mean by a story? It could be a very factual, scientific paper where the author is reporting their findings and how they interpret those findings. It could be literally a news story or a magazine article. It could be a report or a grant proposal where the author wants to raise money or change a policy or justify the existence of their organization.

00:12:10:27 - 00:12:42:02
Speaker 1

Any time you're trying to engage or persuade an audience, you're telling a story. So I chose this three legged stool metaphor for a reason. Problems with accuracy, readability, or fairness undermine the credibility of a graph and diminish its support for the story. But even if the graph is fair and accurate and looks good, if it doesn't add anything to your story, why give it precious space in your document?

00:12:42:05 - 00:13:05:01
Speaker 1
And this is probably the most subjective aspect of a graph. An otherwise good graph that supports one story might be useless for a different story. Maybe it's irrelevant. Maybe it's too technical or not technical enough for the audience. There are a lot of possible reasons. And of course, this is why it's so important to know your audience and read the graphs together with the text.

00:13:05:03 - 00:13:29:28
Speaker 1
Okay, so what happens when a graph falls short of one or more of these expectations? Let's have a look. So we're going to have some polls. In this next section, I want to show you some terrible examples and talk a bit about not only what's wrong with them, but why they maybe happen. We're going to have some polls to vote on.

00:13:29:28 - 00:13:55:27
Speaker 1
The problem with each graph I'm about to show you, please vote. If you're watching this as a recording, I'm sorry you won't be able to vote, but do take a few seconds to study the graph and decide what the problem is. Are we ready? Okay, so this is definitely a good advertisement for a university, right? So your options are bad data.

00:13:55:29 - 00:14:29:17
Speaker 1
Failure to math. Graphic design is my passion. Shenanigans. And shenanigans, by the way, is a great word, which I love. It's it's a devious trick, used especially for an underhand purpose. Tricky or questionable practices or conduct, or high spirited or mischievous activity. According to Merriam-Webster. Wrong graph. And of course, there's always a this graph is fine option that, we could use.

00:14:29:19 - 00:14:42:21
Speaker 1
Okay. How are we doing?

00:14:42:23 - 00:14:43:17
Speaker 1
Okay, people.

00:14:43:18 - 00:14:45:13
Speaker 2
Ten more seconds.

00:14:45:15 - 00:15:13:19
Speaker 1
Oh. We've ended the poll. Oh, sorry. By by an overwhelming majority. We have failure to math and. Hang on. Indeed. That is the case. Yeah. Somebody thought, oh, 20 million is one third of 60 million. Great. I'll just put that in a pie chart. Apply our brand colors nice and clear. Done. But the total shown here is 80 million.

00:15:13:21 - 00:15:47:18
Speaker 1
In other words, the text says one third, but the graph says one quarter, and both of them can't be right. So this is sort of embarrassing. And, It's not accurate. Sure. It's it's perfectly readable. Although I would take issue with the, you know, the legend repeating the labels that are, you know, on the graph itself, you don't really need both of those.

00:15:47:20 - 00:16:16:20
Speaker 1
Is it fair? Who knows? Without having the university's budget? We can't say, And it doesn't support the story. It isn't accurate. And because of that, it loses credibility. So it definitely doesn't support the story the university is trying to tell, which is we are going to give you a good education and give you financial aid while you're doing it.

00:16:16:22 - 00:16:47:05
Speaker 1
Okay. Next one. This one is a bit sneaky. So take a moment and, decide what you think the problem is.

00:16:47:08 - 00:17:26:11
Speaker 1
Nobody's voting. Okay. Just. Just pick something. Nobody knows what you picked.

00:17:26:14 - 00:18:00:27
Speaker 1
Right. Okay. So, we have three people voted. Shenanigans. Two voted bad data and one voted wrong graph. I would personally, you could certainly make a strong argument for both bad data and shenanigans. I would personally, I put this under bad data because, that is sort of the root cause of the problem.

00:18:00:27 - 00:18:19:01
Speaker 1

And it if it was malicious, you could certainly call it shenanigans. But in this case, my cousin edits for an investing website, and one of the writers handed a graph like this to her one day, and she looked at it for a while and then said, what happened in 2020? And he said they didn't pay a dividend in 2020.

00:18:19:03 - 00:18:53:06
Speaker 1
She said that is pertinent information and you have to include that. And he said, oh, really? Okay. So really bad data for whatever reason. It's a common problem. Either the data gets mangled when it's put into a graph or there's something wrong with it to begin with. Certainly if you were if you were a company and you were trying to, just gloss over the fact that you didn't pay a dividend in 2020, I would definitely call that shenanigans as well.

00:18:53:08 - 00:19:20:11
Speaker 1
So yeah, the value shown may be accurate, but the graph is missing data. So it's pretty readable, but it's not entirely fair, which reduces its credibility and that undermines the story. Okay. Next one. Everyone loves a pie chart. That means they're used a lot. And that means this particular error shows up all the time.

00:19:20:14 - 00:19:55:10
Speaker 1
Let's have a look.

00:19:55:12 - 00:20:29:29
Speaker 1
Amazing. Okay, so, half the audience says wrong graph. And another 40% say failure to math. With a bad data, runner up. So, absolutely. This is the wrong graph. Probably because of a failure to math. So, yeah, this is the wrong type of graph for this data. The number one rule of pie charts is that they have to add up to 100%.

00:20:30:02 - 00:20:50:00
Speaker 1
There's one thing. There's one pie. If you remember our lemon meringue pie from earlier. So you can use it for, say, poll results where there was only one answer lab for question or spending from a budget or in this case, the number of people who have or have not tried marijuana at one specific point in time.

00:20:50:02 - 00:21:22:28
Speaker 1
You can't use it to smush together values from a bunch of different categories. Which in this case are people who tried marijuana today, people who tried it as of last year, and people who had tried it as of 1997. So, yeah. Wrong graph. Yeah. The trouble is that Excel will let you do this. You know, it will just say, sure, here's some numbers.

00:21:22:29 - 00:21:49:06
Speaker 1
I can do numbers. I'll put them all on a pie chart for you. It there's no clippy popping up to say these don't add up to 100%. Are you sure you want to do this? So, yeah, it's not really accurate. It's certainly not very readable. And doesn't support the story because. Yeah, if you're trying, the goal is to make comparisons between, you know, points in time.

00:21:49:06 - 00:23:12:02
Speaker 1
And it's, it's quite hard to make comparisons between angles. You would be much better off with, say, a bar graph or a little, you know, a set of pictograms. Okay, this next one's fun. I love this one. Let's have another poll.

00:23:12:05 - 00:23:43:05
Speaker 1
Okay. So, the answers for this one, were sort of all over. We've got a strong, minority in favor of bad data, followed by shenanigans. Failure to math. Graphic design is my passion and wrong graph. And again, you could certainly make an argument for most of these. I chose this as an example of, you know.

00:23:43:07 - 00:24:12:27
Speaker 1
They've smushed together millions of dollars, numbers of people and numbers of days on the same graph. The data itself may well be fine for all we know. There's no particular reason to think it is not. Definitely. There's a failure to math involved because someone clearly didn't realize that you can't just add all those things together.

00:24:13:00 - 00:24:41:14
Speaker 1
Probably not shenanigans as such. Unless, you know, I don't know exactly what time point they were trying to make. You can definitely say it's the wrong graph. There would be a better way, better ways to to show this if you're making a nice infographic. But really, yeah, I call this one graphic design is my passion because it covers issues all over the design spectrum.

00:24:41:14 - 00:25:11:01
Speaker 1
Sometimes the creator has no design skills. Whatever. Sometimes they have graphic skill, but no understanding of how graphs work, as in this case. Sometimes they're so keen to decorate the graph or make it fit a particular metaphor that it's almost unreadable. This is actually interesting information. But it's so badly presented that, honestly, all you can do is is laugh.

00:25:11:03 - 00:25:44:12
Speaker 1

Yeah. So, so very many better ways to do this. And it's a real shame because, you know, it is kind of interesting information. Yeah. We assume it's accurate. It's pretty unreadable. It's not really fair because you're comparing apples to oranges to persimmons. And. Whatever story they were trying to tell here, I don't think it supports it.

00:25:44:15 - 00:26:42:14
Speaker 1
So, yeah. Okay. Here is another one.

00:26:42:16 - 00:27:11:21
Speaker 1
Okay. So we've got a few votes for wrong graph. Graphic design is my passion. Bad data and shenanigans. So the immediate impression given by this visual is that opinions are pretty evenly split between the two options. The question is, is truncating the y axis misleading? You know, half of the graph is saying no and half of it is saying yes.

00:27:11:24 - 00:27:38:11
Speaker 1
But in fact, if you look at the y axis, this is kind of a joke. It has in fact been truncated. At 98%. So there's a whole huge yes block. This blue that isn't shown at all. So that's why. Yeah. Our last category is shenanigans. I hope you can't hear my dog going nuts downstairs.

00:27:38:12 - 00:28:13:18
Speaker 1
Sorry about that. Basically, yeah. This covers any manipulation of data or visuals that's purposeful and actively misleading. Often this involves messing with the y axis, as was done here, to distort proportions or make a change look very small or very large. So yeah, this graph is technically accurate. It is readable, and it probably does support the story the author wants to tell, but it's not fair.

00:28:13:20 - 00:28:37:15
Speaker 1
And this is a key balance that a graph can and should support your story. But it still has to be fair. If you can't support the story without manipulating the data or the presentation. Then maybe the story is wrong and it needs to go in a different direction. Shenanigans come in many forms, and I encourage you to be on the lookout for them.

00:28:37:18 - 00:29:05:16
Speaker 1
I do want to show you another example. Because of a lot of unfair and shenanigans, E graphs distort their scale. So this means that a graph can be technically correct, but still misleading. And this graph in particular appears on a lot of climate denier blogs. It originated with the National Review in 2015, and it's usually brought out to say, look, the climate has been completely stable since 1880.

00:29:05:16 - 00:29:41:27
Speaker 1
All this hysteria about global warming is completely overblown. So it's showing annual average global temperatures since 1880, and the Y axis has a range of about 0 to 110°F, which is roughly -18 to 43°C. Here's NASA's version of the same information, and it's showing temperature fluctuations around the baseline. So changes from year to year rather than the actual average temperature.

00:29:41:29 - 00:30:07:11
Speaker 1
The time range is the same, but the y axis only has a range of 1.5 Celsius degrees. And that means the scale of the National Review's y axis is literally 40 times bigger than NASA's. And here's why that's important. This is a different version of the first graph in Celsius and starting a bit earlier. It's got a range of 30°C.

00:30:07:11 - 00:30:38:20
Speaker 1
So the y axis is half the size of the National Review's. But that still puts nearly two thirds of the Y axis outside the temperature range of the past 500 million years around when the earliest known vertebrates were appearing. Remember, the red line is average global temperature. And as you can see, a good 90% of this graph the y axis is outside the range of the Holocene, which is roughly the past 12,000 years.

00:30:38:22 - 00:31:05:01
Speaker 1
Around the time humans first started domesticating cattle and cultivating barley and wheat. And this is the kind of thing I'm talking about when I say that some graphs don't pursue their agenda in an honest way, and it actually makes me furious when I see something like this. That graph was widely mocked in 2015, but unfortunately this kind of treatment doesn't go away.

00:31:05:03 - 00:31:35:27
Speaker 1
This is a bar graph of basically the same data tweeted out in 2019, shortly before I gave this talk originally. So yeah, same idea. Huge y axis. So masking small but significant fluctuations in temperature. So that's a sample of graphs that are either inaccurate, unreadable, unfair or all three. And this is my, you know, here's our personal taxonomy of why graphs go wrong.

00:31:35:27 - 00:31:58:10
Speaker 1
And as we've seen, there's a lot of overlap between these categories. The Bad graph creator's evil sidekick is often excel because it won't ever tell you, hey, this data looks weird or you should really consider a different type of graph. Plus, its default options are often pretty ugly. You can do a heck of a lot with it, but it's not magic.

00:31:58:11 - 00:32:24:29
Speaker 1
You have to know what you're doing. You have to have a
You have to understand what you're trying to show. And yeah, this leads me to what I think is the matter problem in many cases. The person who made the graph didn't think it through. So it's up to someone else, often the editor, to gently help them see the problem and fix it before the graph gets out in the wild and people start making fun of it.

00:32:25:02 - 00:32:50:00
Speaker 1
Or if the author didn't think it through and shenanigans are involved, it's it's our responsibility to push back on that. Okay. So how we how do we do that? I'm going to go through this quickly, and sorry, I'm moving a bit slower than I hoped, but we will try and make it up. So there are basically two options for fixing a bad graph.

00:32:50:02 - 00:33:16:02
Speaker 1
One, you can present the data differently, or two, you can present different data. Simple. We're done. You can leave early. Know which option you pick? Depends a great deal. Of course. Again, on the nature of the problem, the purpose of your document and where you are in the document development process. You can look at graph problems on a continuum.

00:33:16:04 - 00:33:44:09
Speaker 1
Some problems are very straightforward. This is wrong or misleading. And we have to fix it. No matter what. Others are more subjective, and you know, whether you can fix those will depend a lot on where you are in the editorial and publication process. You know, if it's just if it's basically fine, you know, all the information is correct, but you just don't like the look of it.

00:33:44:11 - 00:34:05:27
Speaker 1
But you're on a deadline. Sometimes you have to let things go. For example, if you're proofreading a document that's already been laid out and you happen to think a graph is really cluttered and hard to read, there's not necessarily much you can do at that stage other than point it out and suggest they do better next time.

00:34:05:29 - 00:34:33:06
Speaker 1
The earlier you were in the process, the more you can suggest bigger, bolder changes. We will be circulating a handout after, I think it's going out to everyone who registered. And I do have a table of questions to ask at each stage, which you are welcome to make your own. So let's look at fixing each category of problem.

00:34:33:08 - 00:34:58:29
Speaker 1
There are a few different ways that mistakes can sneak into a graph. Maybe someone made a mistake when turning the data into a graph. Sometimes the graph is faithful to the data, but there's a mistake in the data table. Sometimes, especially if the graph is one of a series of similar graphs, the designer has duplicated the file and missed updating one or more components like the source or the axes.

00:34:59:01 - 00:35:21:22
Speaker 1
Sometimes the graph itself is fine, but the axes or the data points are mislabeled or some essential elements are missing. So checking to the best of your ability that the graph is accurate is a key task. Even if you can't fix anything else. Sometimes you can see there's a mistake in a graph, but you can't tell exactly what it is without checking against the data table.

00:35:21:23 - 00:35:43:28
Speaker 1
So back to this one. Maybe the university's total budget is $60 million, in which case the graph is wrong, which is what the headlines suggest. But maybe the graph is right and the headline is wrong. Maybe, you know, maybe the university's total budget is $80 million, and that means a quarter of their operating budget goes towards financial aid.

00:35:43:29 - 00:36:15:26
Speaker 1
So if you don't have the data table or the information isn't in the text, all you can do is query and explain what the issue is. And I earlier mentioned the, you know, the repetition of 20,000,060 million. You know, the legend should add information if you're going to use one. And for a very simple graph like this, it's totally fine to just label the segment, this graph relatively easy fix.

00:36:15:28 - 00:36:36:08
Speaker 1
The writer just needed to insert a new line for 2020 in the data table. Always remember that zero is also a number. So yeah, make sure all the numbers and labels are correct and always label your axes.

00:36:36:10 - 00:37:10:17
Speaker 1
So next category is readability problems. You can have an otherwise great graph that spoiled by readability problems. And these problems can include ordinary layout issues like bad font choices, bad alignment not enough contrast. Some graphs have a lot of visual clutter, like heavy grid lines or purely decorative elements. Edward Tufty, who's a very well known and respected, data visualization guy, refers to these as chart junk.

00:37:10:20 - 00:37:34:24

Speaker 1
Or maybe they just have way too much information to read easily. This graph is actually very interesting. It's from a scientific paper whose author was trying to estimate when the beaver went extinct in Britain by looking at animals mentioned in medieval texts. But as you can see, there's there's a lot of rotated text on this graph.

00:37:34:27 - 00:37:57:25
Speaker 1
Which is a major readability issue on, and in fact, it's rotated a full 90 degrees in either direction, so the labels on the bars or you have to tilt your head 90 degrees to the right. And for everything else, you have to tilt it 90 degrees to the left. So to make sense of it, that that's sort of an extra load.

00:37:57:27 - 00:38:23:25
Speaker 1
It's quite a lot of visual clutter. You know, the category labels on the bars, it's the same ones every time. That is a good opportunity for just a legend. Where you could have these date ranges once, each with little colored boxes. So the data itself is good, and the choice of graph is appropriate. It's just very cluttered and hard to read.

00:38:23:27 - 00:38:58:20
Speaker 1
And if you were, you know, if you were adapting this information for more of a lay audience, you would probably want to change the, look, the Latin binomial names of species to, you know, English names. But for a scientific audience who, who knows what you're talking about, this is probably fine. So. Yeah. And if you can reduce or eliminate the rotation on the remaining text, maybe have these data labels at a 45 degree angle instead of 90, 90 degrees.

00:38:58:22 - 00:39:29:21
Speaker 1
You could put this x axis, label potentially at the top, or indeed you don't necessarily need it because it's in the it's in the caption. You have a lot of options. Also a lot going on here. This graph is quite hard to make sense of. You have to look back and forth between the title and the graph and the legend to figure out what's going on.

00:39:29:24 - 00:40:03:17
Speaker 1
I'm not sure that these trendlines actually convey useful information. If if they are useful, then why are the colors different for the bars and the and the, lines in the same category? What is sundry revenues if, if it's so tiny that we can't even see it? Is it worth including, do we necessarily need all these grid lines?

00:40:03:19 - 00:40:33:19
Speaker 1

Another issue is that this graph only uses color to distinguish between categories and all these colors have very similar saturation. And that can exclude people with limited color vision. So that's something else. Or, you know, people who want to print the report in black and white. And so we need to make sure that there are multiple ways of getting the information.

00:40:33:22 - 00:40:55:11
Speaker 1
Sorry, I don't know why I have so many different clicks there. And with readability issues in particular, you may need to work with the graphic designer or the author. Maybe you are the graphic designer or the author. But these are some general tips. It needs to be clear what the most important elements of the graph are, and what order to read them in.

00:40:55:13 - 00:41:16:09
Speaker 1
If you're constantly going back and forth between different parts of the graph, to make sense of it, maybe you need to make the title more informative. Maybe you need to ask the designer to move the data labels so that they're right next to the data, instead of off in a box by themselves. Sometimes a legend is useful, sometimes it's not.

00:41:16:11 - 00:41:47:23
Speaker 1
Some, you know, sometimes you can reduce visual clutter, by deleting or de-emphasizing elements that are just decorative or for wayfinding, things like heavy grid lines, patterns, gradients, 3D, using too much color or too many colors, nonstandard shapes, things like that. I mentioned earlier that this issue of supporting the story is very context dependent.

00:41:47:26 - 00:42:17:13
Speaker 1
Whether a graph is useful depends on the audience and the story itself is the text talking about one thing while the graph is talking about something else? Alternatively, if information is good and potentially useful, is it obscured by a bad presentation? Here, for example, the the story, according to the legend, focuses on dull donut flavors, but the graph doesn't actually show individual flavors, just presumably overall donut sales.

00:42:17:16 - 00:42:46:18
Speaker 1
So if we want to focus on flavors, here's one possibility. We can break our graph bar graph down by category and highlight the two items of interest. This is a very simple example of using data to support a story. Although there are some potential shenanigans. In this case, the same data is being used to support two different stories on the right.

00:42:46:20 - 00:43:11:23
Speaker 1

We've compressed the x x axis and stretched the y axis, so the line looks like it's rising much more steeply. Either of these presentations is actually fine, with an important caveat. If you have multiple graphs with similar data, for example, if you're comparing donut sales to cake, you want to make sure the scales are all comparable.

00:43:11:26 - 00:43:46:17
Speaker 1
Otherwise, if you show these side by side, it looks like cake is a huge growth area compared with donuts, whereas in fact they're both doing about the same. Another point is this type of heading where you lead the audience in the direction you want to go is great. In some contexts, like a sales presentation, but it would be frowned upon in a scientific paper, where you're expected to let the data speak for itself to some extent with factual and opinion free captions.

00:43:46:20 - 00:44:28:11
Speaker 1
So for a context like that, you might want to make your caption something like donut sales 2015 to 2018. It's not very exciting, but it is descriptive. So yeah, again, this is an example of that very, just the facts descriptive title. We could certainly, you know, make this more exciting by highlighting some interesting trends. As I mentioned, use the common names instead of or as well as scientific names.

00:44:28:13 - 00:44:55:24
Speaker 1
And certainly make the title more gripping. So back to bad data treatment or presentation, which, as we've seen, can undermine your credibility and thus the story. This, as you'll recall, is an inappropriate and sadly, common use of a pie chart. Here's an example of the same problem that I found on a website for healthcare professionals.

00:44:55:27 - 00:45:26:21
Speaker 1
It shows rates of diabetes. Diagnoses in the US in different racial and ethnic groups. So yeah, the numbers don't add up to 100%. Different categories have been smushed into a single pie chart. So to display this information correctly, you'd need five separate pie charts, which, incidentally, is one reason why data viz people kind of hate pie charts because they take a lot of space to give not very much information.

00:45:26:24 - 00:45:58:19
Speaker 1
There's also plenty of space. You know, you have to bounce back between the legend and the, the graph to see what segment refers to who. There's certainly plenty of space. If it was appropriate to label these segments with the category name. So yeah, the trouble with pie charts in particular is Excel will happily take whatever you numbers you give it and turn them into a pie chart, regardless of whether it actually makes any sense to fix it.

00:45:58:21 - 00:46:20:08
Speaker 1
Basically, if the numbers in a pie chart don't add up to 100%, it needs to be a different kind of graph. So in this case, I turned it into a horizontal bar graph. Makes much more sense this way. It's much easier to read and make comparisons between categories. Category labels are right next to the data.

00:46:20:10 - 00:46:45:15
Speaker 1
If you were giving a presentation or a report and you wanted to talk about a particular racial or ethnic group, you could use a contrasting color to highlight just that group. Here's an issue that can come up when you have multiple related graphs in a document. And that's that. These are all from the same document, but they're inconsistent.

00:46:45:17 - 00:47:07:02
Speaker 1
So you're making the reader do extra work if they've just gotten used to one type of graph meaning one thing. Don't make them. You learn a whole new system just because someone got bored with a particular graph design. You. Unless there's a good reason you want to pick one type of graph and stick with it to make it easier to compare.

00:47:07:04 - 00:47:29:05
Speaker 1
So to recap, if you've determined that a graph doesn't support your story, you first need to decide why that is. And that may guide you to the best course of action to fix it. And there are several resources on the last page of the handout that are really great guides to using graphs to tell your story, and I encourage you to check those out.

00:47:29:08 - 00:47:56:10
Speaker 1
The last leg of our stool is, of course, unfair graphs. And as we've seen, an unfair graph may be technically accurate and support your story, but it's not fair. It's a graph with an agenda, and it's not pursuing that agenda in an honest way. Maybe the graph is using cherry picked data. It only uses the information that's most favorable to one side of the story and ignores the rest.

00:47:56:12 - 00:48:28:09
Speaker 1
So yeah, we've seen this. This is another example. Sometimes a data treatment is chosen that obscures the story or pushes a particular version of a story. If you don't account for inflation when you're graphing how wages are changing over time, they're going to look great. But that doesn't show the real picture. So here we've got a graph showing median weekly earnings, change between 2000 and 7 and 2019.

00:48:28:11 - 00:49:03:17

Speaker 1
But this is without inflation and this is real wage growth over, a smaller portion, but still, it does not match at all. You know, you can see this is happily going up. And this is the real the real picture. Real wages accounting for inflation have gone down. So always think about whether the data treatment that's chosen is appropriate, and if it's what the audience is used to seeing for that type of data.

00:49:03:19 - 00:49:25:25
Speaker 1
Always make sure that the graph and the document provide enough context for the reader to assess the information. Is a number low or high without context? How do you know? This is a graphic design issue. The choice of a pie chart is fine. The numbers add up to 100 because these are mutually exclusive categories.

00:49:25:28 - 00:49:47:06
Speaker 1
But the extreme tilt makes the orange slice of the pie look huge, when in fact it's smaller than the blue slice. If you just glance at it and don't take in the numbers, you could get the sense that this party was comfortably in the lead. You know, pretty adds visual appeal, especially for TV or social media.

00:49:47:06 - 00:50:17:24
Speaker 1
But the trouble is, as soon as you tilt the pie, it's distorted and it's no longer giving an accurate picture of your data. Really? Again, Excel is to blame. Simple solution. All pie charts must be circles and always question any 3D effect you see in any graph. So fairness might be the trickiest element to check, and it's often the most delicate element to fix.

00:50:17:26 - 00:50:44:15
Speaker 1
So approach unfair graphs the same way you would if you encountered bias or cherry picking in the text. Highlight issues to the author and if necessary, the supervising editor. Be tactful. Explain what makes the graph misleading or unfair, and how it will detract from the credibility of the work. Suggest a fix if you can, such as an alternate presentation or a different data source, or adding missing context.

00:50:44:17 - 00:51:09:26
Speaker 1
But basically if something seems off, always query. Remember that fairness is a major credibility issue. If you or whoever is putting the graph out into the world doesn't check these things, there are knowledgeable and informed people in the audience who will notice and and they will judge and probably mock you on the internet. I'm just going to show you one last graph here.

00:51:09:29 - 00:51:53:28

Speaker 1
Because, well, actually, I'll show you a few last graphs very quickly. Never do this. Do many categories limit pie charts to maybe 5 or 6 categories at most? And honestly, did this need to be a graph? I love this one. The, the tiny, tiny Indian ladies, cowering behind the gigantic Latvian ladies. The tweet, if you can't read it, says, as an Indian woman, I can confirm that too much of my time is spent hiding behind a rock, praying the terrifying gang of international giant ladies and their Latvian general.

00:51:53:28 - 00:52:19:08
Speaker 1
Don't find me. So the issue with using shapes this way, even when the y axis isn't cropped like this is, it's a visual perception problem. People find it very difficult to accurately compare the sizes of to irregular shapes and shapes. They don't scale up linearly, so something that's twice the height in this one is twice the height roughly of this one.

00:52:19:11 - 00:52:39:10
Speaker 1
But visually it it's four times the area. So it looks even huger. Yeah. This is very. Yeah. For once, this is not Excel's fault. Someone had to work at this.

00:52:39:13 - 00:53:05:11
Speaker 1
This is just funny. Yeah. Graphic design. Definitely someone's passion. You may notice that graphic design is my passion. Is kind of overrepresented in these examples. And I have to admit, that's usually because they're the funniest ones. So, yeah, this is showing that 47% of people's favorite pie is apple, but that's a little quite small sliver, maybe 30%.

00:53:05:14 - 00:53:40:26
Speaker 1
And this is why people selected their three favorites. And, this is the result. This graph adds up to 300%. Yeah, it's a problem. This, it turns out, is shenanigans, which eventually led to suspicions about their data, which proved to be well founded. In case you don't work with scientific data a lot. Let me briefly explain.

00:53:40:26 - 00:54:10:22
Speaker 1
Error bars are used to show the uncertainty or variability in a data set. They're also data. You can't just throw them onto a graph using capital TS. This is from a real paper that was actually retracted late in 2022 after scientists pointed out this issue and people started looking harder at it. Excel can totally put in proper error bars if you know how and if you care to do so.

00:54:10:24 - 00:54:35:22
Speaker 1

As it turns out, the data was fabricated. So yeah, to recap, and I'm sorry we've gone so long. I hope we still have time for a few questions. Graphs or communication? Trust your instincts. Keep the audience and the medium and the story in mind. If it doesn't make sense to you, it probably won't make sense to the audience.

00:54:35:25 - 00:54:58:14
Speaker 1
That said, you won't always be able to fix everything, so accuracy is the priority. And finally, if something seems off, it might be so be alert for shenanigans. Thank you very much. I apologize for, going a bit longer than I meant to, but are there any questions? And should I stop sharing? Probably.

00:54:58:17 - 00:55:27:16
Speaker 2
Please go ahead and stop sharing. Robert. Thank you so much. That was a wild ride. We did start a couple of minutes late, so we'll take a few minutes for questions. And we have one question waiting for you from Harry, who asks with regards to colors. Are there certain color themes to avoid? I'm thinking about inclusivity around color blindness as well as color bias such as red.

00:55:27:16 - 00:55:33:26
Speaker 2
It's bad and green is good. Or any tips around use of color in general?

00:55:33:28 - 00:56:16:24
Speaker 1
Great question. So, color is quite culturally specific. So, you want to take into account, you know, what, what colors mean in your, in your local context. Because they can mean quite different things to, to different groups of people. Definitely. Inclusivity around colorblindness is an issue. There are, some great, accessibility checkers on online that you can, you can upload images to and, you know, check whether they're, you know, readable or not.

00:56:16:27 - 00:56:58:11
Speaker 1
Basically the general rule is don't make color. The only way to distinguish between different categories or different pieces of data. It's fine to have color. Color is great. But it can't be the, the only thing that, that distinguishes between between things. So you want to include, you know, a pattern or labels or something else that will that will help readers to distinguish between, between different things.

00:56:58:13 - 00:57:23:01
Speaker 1
I hope that's helpful. You know, a lot of, a lot of, organizations have their own visual style guides that specify, their, their brand colors. And often you'll, you know, be wanting to use those.

Definitely. If you if you're working with a designer. Absolutely. Consult them about this. They probably know a lot more than I do about this issue.

00:57:23:03 - 00:57:35:02
Speaker 2
Okay. I think we have time for one more question. Lisa asks if you could share a top 5 to 10 data visualization books you'd recommend. Is there something like that in the handout?

00:57:35:04 - 00:57:47:15
Speaker 1
There is. Yes. Okay. That is there's a list in the handout. And definitely there are some great ones out there. Depending depending on your needs. New ones coming out all the time.

00:57:47:21 - 00:58:06:12
Speaker 2
So and I'm going to pronounce this name wrong. I apologize in advance. Gina wants to know if she can, talk about this webinar on her LinkedIn and, tell people what she learned.

00:58:06:14 - 00:58:27:14
Speaker 1
Certainly happy for you to do that for my, for my perspective. Feel free to tag me. I'm. I'm pretty findable. I'm, there are very few Robin Marwick out there. There's a couple of others in Scotland, but I'm I'm, you know, I'm beating them. The SEO game. So, you should be able to find me fairly easily.

00:58:27:16 - 00:58:32:13
Speaker 1
So. Yeah. And feel free to connect if, if that's of interest. Okay.

00:58:32:14 - 00:58:59:23
Speaker 2
Excellent. I think that takes us to the end of our hour. On behalf of everybody else. So I just wanted to say thank you again to Robin. We certainly learned a lot. And remind everybody that there will be a, recording of this available. In case you want to go back and listen again. And in the chat, Frieda has just posted our evaluation link.

00:58:59:26 - 00:59:12:18
Speaker 2
If you could take a moment to go in there and tell us what you thought of this. We would greatly appreciate it so that we can keep targeting these webinars to the right subjects.

00:59:12:21 - 00:59:13:12
Speaker 2
So thank you.

00:59:13:12 - 00:59:21:15
Speaker 1
Everybody, so much. Thank you, Michelle, for and for Rita for having me. And, and I hope I hope everyone got something out of it.