There are a lot of visualizations in the world and we don't have time for them all.

So let's focus on one particularly abused plot type, the pie chart.

We have a specimen right here.

This is a pie chart of phone application crashes, showing what percentage of all crashes took place in each mobile operating system.

This data set contains information for all versions of Apple's iOS, which is used in the iPhone, as well as the various versions of Google's Android.

There are many things wrong with this plot, but let's break down exactly what.

Putting aside, for a moment, that there are far too many labels, check out the ordering of the labels corresponding to iOS.

Two sensible ways of ordering iOS data might be by decreasing percentage or by version number.

Instead, we start at the top with iOS 3.13, with 0%, and then jump to iOS 4.2.10, with 12.64%, before going back down to iOS 3.2, with 0.00% again.

Which brings us to the number of labels.

Many of the segments are so narrow that they can't be seen, although technically, all data is retained, because every segment is labeled.

If we look at iOS, we see that there are only three major versions, 3, 4, and 5, suggesting we can compress down the iOS segments to just three segments, while retaining most of the information.

At the least, the versions that differ in the third number should be combined, and all data points of 0% should be removed.

The more fundamental concern of this visualization is that it might really be showing the percentage of the phone market using each operating system, and says nothing about whether one operating system crashes more than the other, which is the focus of this visualization.

Our next pie chart has its own share of problems.

This is a plot of how many shark attacks have been attributed to each type of shark.

Firstly, the pie chart is, for some reason, plotted on a hemisphere, a graphical effect that adds nothing, but has the effect of vertically compressing the pie chart.

Next, there is the issue of label orientation.

While the caption, "Shark species (total/deaths)", and the label, "White shark", are horizontal, the rest are vertical and hard to read.

They are in order, however, which does help.

Although the "Others" segment is unfortunately large, which is unclear if it is due to there being a lot of attacks by many species, or if the species is not known for many attacks.

Finally, at a glance, it is hard to distinguish the magnitude of differences between the orange, green, blue, and brown segments in the top part of the pie chart, and we must resort to the labels to distinguish between them.

There is no meaning in the colors, they are arbitrary.

Finally, we'll look at a pie chart I made, of the origins of the international students at MIT.

I made this chart with the default settings in Google Sheets.

First of all, not all of the segments are labeled, so that data is lost, for the Middle East, Africa, Oceania, and the unknown regions.

Second, again, we have colors that are arbitrary and almost close enough to be confusing.

The difference between Asia and Africa's colors is subtle.

And of course, the 3D-effect on the pie chart adds nothing, but does play a subtle trick on the eye.

Due to the 3D-effect, the blue and red segments are actually larger looking, which at a glance, may lead the viewer to overestimate their size.

What we are going to do now is, switch over to R and plot this data more appropriately, using ggplot.

And then we'll return to the slides, to discuss some more possibilities for this data.