First, let's make sure we have ggplot2 loaded.

So library(ggplot2).

Now let's load our data frame, which is in households.csv.

So read.csv(households.csv).

If we look at the structure of households, we see that there is a year column and then six other columns for each of the different household types.

So this is actually a problem for us.

Given this structure of a data frame, what would we put in the aesthetic for our ggplot command?

It's not obvious, to me at least, and in fact, I don't think it's really possible.

The reason is that ggplot needs it in the form of: year, group, and fraction.

The solution is to use the melt function from the reshape package.

Melt will take a 2-dimensional data frame like ours, and convert it into exactly the right form we need for ggplot2.

So first, let's load reshape2 -- library(reshape2).

Now, let's look at the first two columns of our households data frame -- households[,1:2].

So there's a Year and a MarriedWChild for each year.

Now, let's look at the first few rows of our melted households data frame.

So head(melt(households, id="Year")).

And there you have it.

So, basically, what's happened is that each value of MarriedWChild has turned into its own row in the new data frame.

To make it more clear, perhaps, let's look at the first three columns of households.

Now we have MarriedWOChild.

Now let's look at, instead of just the first rows of our melted data frame, let's look at the first 10 rows.

So rows 1 to 10, all columns.

There we go.

So there you can see the eight values of MarriedWChild, and the first two values of MarriedWOChild.

So there's that 30.3 up there, gone down to 30.3 here, 29.9 gone to down here.

So every value in our data frame now corresponds to a new row in our melted data frame, which is exactly what we need for ggplot.

So let's try plotting this melted data frame -- ggplot, melt, households, using the Year column as an id column, the key column.

Our aesthetic is going to be now to use Year on the x-axis.

Our y-axis will be the value column of our melted data frame.

And the color of the line will depend on the group, which is called variable in the melted data frame.

So that's our aesthetic.

Our geometry is going to be lines.

And I want to make the lines a little bit thicker.

So let's say line size is 2.

And I also want to have points for each year in the data frame, so I'm going to have lines and points.

So geom_point.

And I'm going to make the size of these a little bit bigger than normal too, size = 5.

And we should put a y-axis label.

So ylab("Percentage of Households").

And there you go.

Now, this is actually quite interesting when we compare it back to the chart we had in the slides.

Now you can see just how quickly MarriedWChild is decreasing as a relative share.

You can also more clearly see that MarriedWOChild is pretty much flat, and that the differences being made up by the other four types of households is steadily increasing over the years.

So there you have it, the same data, plotted in two different ways.

Now, I'm not saying one of these is better than the other one.

For example, if I want to compare inside a given year, say 1970, it's not the most easy thing, at a glance, to see just how much of a total hundred percent is taken up by each.

But if I want to see across years, it's far superior.

And I can clearly see that the last data point is pretty much right next to the second to last data point, which is something that was hard to tell with the other visualization.

So I hope this has made you think a little bit more about the different ways you can plot the same data.

And hopefully improved your ggplot2 skills a little bit more.

Thanks for watching.