So now we're going to try plotting a world map with a new data set that has the number of international students from each country.

So first of all, we're going to need to use the ggmap package, which you may need to install.

And we're going to load in the data set, which is called intlall.csv.

So read.csv(intlall.csv).

And I'm going to do stringsAsFactors = FALSE.

OK?

Let's look at the first few rows of intlall.

So you see that each row corresponds to a country.

There's a citizenship column that's the country name, number of undergraduates, number of graduates, special undergraduates and graduates, exchange or visiting, and a total column.

Now there's these NAs in here, but they're not really NAs.

They're just 0's.

So what we're going to do is say, all these NAs should be 0's.

So in intlall, all entries that are NA, should be 0.

And let's look at the first few rows again.

OK.

Much better.

Right, so next step is to load the world map.

So let's call it world_map = map_data("world").

We did something similar in the lecture with the state data.

So let's look at the structure of the world_map.

So the first two columns are the longitude and latitude; third column is something called group -- that's actually a group for each country, using a different number for each country; order, we'll get to that later; region is just the country name, and subregion is sometimes used for some countries to describe islands and other things like that.

So we want to shove the world_map data frame and the intlall data frame into one data frame, so we can use it for ggplot.

So let's say world_map is a merge of world_map and intlall.

Now, in world_map, the country name is just called region, as you can see right here.

And in the intlall, the country name is actually called Citizenship.

OK.

So let's look at the structure of world_map just to make sure it makes sense.

Looks good.

OK.

So to plot a map, we use the geom_polygon geometry.

So start off ggplot(world_map, aes(x = long, y = lat, group = group)).

We want to use geom_polygon as the geometry.

Countries will be filled in in white, and their borders will be in black.

And we'll use a Mercator projection.

There's a few other options in there, as well.

OK.

So that looks kind of like a world map.

There's a few things going on here.

So first of all, all the countries look like big black blobs.

What on earth is going on, you might say.

Well, sometimes the merge can reorder the data.

And it turns out that what the world_map data frame really is is actually a list of latitude and longitude points that define the border of each country.

So if we accidentally reorder our data frame they no longer make any sense.

And as it goes from point to point, the points might be on the other side of the country as it defines the polygon.

So, we have to reorder the data in the correct order.

So this command is a little bit complicated looking, but when you break it down, it's not so bad.

So, we take the world_map, and we're going to reorder it.

So world_map, we're going to reorder the rows.

We're going to order the rows based on, first of all, the group, which is pretty much equivalent to the country, and then the order variable, which is just the correct order for the border points.

And we're going to take all the columns, of course.

Done.

So if we go and try plotting it again -- so ggplot-- I guess I should go up, up.

There we go, much easier.

Right, so now we have the map, and it looks far more reasonable.

OK, next problem.

Some of the countries are missing.

Now of course, the USA is missing because MIT is in the USA, so that wouldn't be an international student coming from the USA.

And some parts of Africa are missing, presumably because there are no students at MIT right now who are from those countries.

But you'll also notice that Russia is missing, and a lot of countries near it, as well as China.

Which is definitely not true because I have many friends at MIT who are from Russia and China.

So, what do we do about that?

The reason China is missing is that it has a different name in the MIT data frame than in the world_map data frame.

So when we merged them, it was dropped from the data set because it didn't match up.

So to see what it's called in the MIT data frame, let's just do a table.

There's a few ways to do this, but this is pretty easy.

OK, so we get a list of all the names.

If we scroll all the way up, we'll see it says "China (People's Republic Of)".

Now, in the world_map data frame, it's simply called "China".

So, what we can do is change the MIT data frame.

So let's say the citizenship column, the one row where it equals "China (People's Republic Of)" should just be "China".

OK, let's check.

Do the table again.

Scroll all the way up.

There it is, China.

So we've fixed that.

So now the MIT data frame is consistent with the world_map data frame.

So now we have to go through the merge again.

So let's say world_map is a merge of a fresh copy of the map_data, the intlall data frame with China fixed.

It's called region in the world_map data, and it's called Citizenship in the MIT data.

Alright, now we need to do the reordering again.

So press up a few times until we find it.

There it was.

So there's the reordering command.

OK.

And we should be good to go, now.

So let's try plotting it.

So ggplot, the world_map data frame.

The aesthetic, x is the longitude, y is the latitude.

We need to group countries together, so it doesn't all crisscross over the map.

We're going to use geom_polygon again.

This time though, let's actually fill them with a color that's proportional to the total number of students.

We'll still outline them in black, though.

And we'll use the Mercator projection.

Much better.

So Russia is missing for similar reasons, but we won't deal with that now because it's a little bit annoying.

But you get the idea.

This is pretty interesting actually.

So we can see that Canada, and China, and India supply a large number of international students to MIT.

But it is a little bit confusing doing it on a per country basis, because Europe, presumably, has quite a few students at MIT.

But because Europe is made up of many small countries, it doesn't look very impressive.

Maybe if all the European countries were grouped together, it would look about the same color as Canada.

But it's hard to tell.

There are also other projections we can look at.

So this is a Mercator projection.

What I want to show you is an orthographic projection that allows you to sort of view the map in 3D, like a globe.

So let's try that out.

ggplot, world_map, aesthetics are the same.

Actually, let me do this the right way.

I'll just press up.

OK.

Let's change it to orthographic projection.

And I want to find, now, an orientation.

And this is almost like thinking about where in the world you want to focus on.

So this is a latitude and longitude, 20 degrees and 30 degrees.

If we run this, we should get a map centered above North Africa.

That's quite a nice visualization because if you want to look just at Africa and Europe, this is the way to go.

We can still see China, and Canada, and South America in there, as well.

Let's do something a little bit more personal.

I want to change the coordinates, now, to -37 and 175.

Now it's centered on my hometown of Auckland, New Zealand.