Mine too, but last week was I was in Oxford for a seminar and had to prepare the courses.  
So this was quite stressful and it seems this semester for the new students is even more confusion with getting getting registered or something like this, but at least good that most of you found it already in the registration and also to class today.  
So we can start with our this is your last year basically right.  
And then and the visualization and communication, the visual communication and data story telling cores.  
Umm, let me share.  
I know I shared already this screen.  
Umm, I provided another GitHub organization, but I found the problem that.  
I failed more or less for some technical reason to invite you, and now it's blocked for 24 hours.  
So I will try to invite you today.  
So you will get umm, you will get the get the email by GitHub hopefully if not everything has changed.  
So worse that it doesn't work anymore, but the idea is that you will get an invitation today.  
I will not will not be that little.  
There will be massive massive repositories.  
Probably there will be one sandbox where positive and then you can also have the project the project Zurich and similar to the Data Science Lab course and also I provide some repository there, this one with some more information.  
When you go to this page, you can already Click to the syllabus site.  
I think this is online.  
Yeah, I hope that not too many mistakes on it, but it has already a rough schedule which will be more precise or maybe make more precise in the in the in the, in the internal, in the internal Rep Course organization repository.  
Yeah, but some some general structure is already there.  
Maybe we have to adjust it a bit depending on what's happening while you come to this again a bit later, or a hopefully toward the end of today's sessions.  
So umm that is how you should find it.  
I think also this link to to to this page.  
I think I also provided this encompass net so you can also get there from there or now it's a link is also also in the team umm.  
So some students are here.  
Hybrid or not, just one.  
Again, there was Anastasiia.  
Before.  
And I guess maybe some others arrive, so I will try to provide this hybrid setup.  
So, but just to make this clear course is designed for in person presence.  
I have some sessions, I think four or something like this planned as more like these individual meetings.  
A bit similar because also this course is a lot.  
The main assessment is project work so.  
Uh.  
Uh, so it may make some sense to have a more direct interaction with, with with you individually on on the project work.  
So.  
So we'll do that and I think hybrid participation, it can work.  
I hope it doesn't end up me sitting here alone and entertaining screen, but yeah, but there are also some sessions where it doesn't not not really work at all.  
I think, for example, today we will have some some sessions where I ask you, yeah, discuss with your neighbor briefly, and then we'll discuss a bit together.  
In principle, it may be organizable online, but it's just much harder.  
Not so easy on the spot.  
So.  
So but yeah, if if there is a need, I'll try to just just just offer the hybrid setup.  
So so if it is the only way that it can work.  
So yeah, I think it can be possible.  
It's just not the best setup of.  
That's that's how it is.  
OK, let me show my my slides for today to get into the topic.  
It's a very end. Umm.  
So like this?  
OK.  
And.  
Yeah.  
So, umm, today should be about what is visual communication and what is data storytelling.  
Umm, let's go right into the contact. Umm.  
This is, yeah, some somewhat.  
A lot of so communication.  
Usually we think of as verbal communication.  
That's what you mostly have in mind, but in in reality, a lot of the communication is visual.  
So we'll provide something visually.  
Drawings, maps, tables, graphs, whatever.  
A photographs or even a 3D objects.  
You provide them, umm?  
Not only to build something for us, but to communicate with others.  
So and so as then.  
So it's also part of the communication and we can think about all the things in communication.  
So we'll sending information, others will perceive it.  
Maybe they perceive it differently, or maybe they perceive it.  
Maybe we'll provide it in a very complicated way, then it will be more complicated to decode.  
So all these questions generally about communications come that come come into place when we when we provide a visual, so to say for our data science work.  
So.  
Yeah.  
And there are these.  
There are these, so I will not go super deep into these things.  
It's more like a a general general introduction.  
So just there is this saying.  
What? What?  
What people remember is is not so much of what we hear, but much more of what we'll what we'll see.  
Not really sure where these numbers come from, but it's yeah, I think it is just very striking that a lot of the things we remember better is in the visuals and the words are.

Yeah.  
When did this type of visual communication start?  
So arguably, probably also animals have this type of visual communication, leaving some marks as type for others, but also many animals do what?  
How do you all factory communication with yeah, with leaving leaving all factory marks that others can smell them, but we'll do a lot with visual sense because this is the strongest sense of us humans, I would say.  
By the way, what do you what do you think visual and visual visual perception is the strongest sense?  
What the second strongest of the senses we have to your opinions.  
Not really a full knowledge question.  
Really.  
But what is the second?  
Second, after we after vision, another smell of touch.  
You don't think here.  
I think hearing is where get more information.  
Of course, all the smelling and touching also important.  
But what is?  
What is?  
What is special about?  
What is the words about touch is just not so far reaching, right?  
I cannot touch the wall on the other side, but maybe I can hear something and just what do you think?  
So there there is something you can perceive these things as, OK waves of information or something.  
But what would you say it's a structural difference between vision and hearing?  
So both things work somehow far reaching, so we can hear someone something from far away.  
We can see something far away.  
What is?  
What is the structural difference?  
When you look at how we can store this information and and record it.  
Just in the data, when you have when you have a visual pictures stored in data or when you have a sound stored in data, what is the structural difference?  
Yeah, something.  
And yet, anyone as an idea is a bit, huh?  
Yeah, the the.  
I don't converted into some other something.  
He's also the same as audio.  
It's more the what is the format I think about dimensionality.  
So how when you see a picture, how many dimensions does it have to when you hear some noise?  
What is the dimension?  
Just one in some sense, right?  
So we and it is also very directed.  
So when we when we have the vision, we can look here and there and there at all makes some sense.  
So when we hear what, when we hear backward, it's first is not really possible.  
Technically it's possible, but doesn't make any sense, right?  
So it's just, it's just very much.  
Yeah, very, very linear and directed.  
While with vision we can, we can.  
Watch around how we want and this is also some implications.  
So when I will talk to you in this linear way, it goes step-by-step.  
Everyone has to follow in that order, but when I saw you a visual, everybody can look where?  
He or she wants and so, but there's all still.  
Of course, from some structure in the visual perception.  
So maybe most people look at the same thing first or something like this, and these are the things to.  
Uh to we want to look a bit deeper into and and learn a bit about that, but I think it's also nice topic to yeah to practice ohh anytime you do some some better data work or data visualization work.  
So when I when I create a figure, when I create a figure, what do my?  
What does my audience look at first?  
Or these type of things.  
OK.  
Anything from this slide we need to we need to discuss, OK.  
It's just there.  
OK, visual communication starts very early with with even this cave paintings some let's say 40,000 years ago.  
OK. Umm.  
OK, this is somehow other forms of of of of visual communication ideograms logograms like like like like like alphabets developing into into alphabets.  
Books and fonts.  
OK, so many fonts have have been developed for to to to show our alphabets.  
So the alphabet is not really.  
The visual communication is alphabet is for textual communication and coding, let's say speech.  
But when you also write it down, it becomes also some visual communication and fonts.  
So this type of writing in a or other types of writing in a become then also crucial will not go deep into that.  
Some people are really into typography.  
Yeah.  
I think also it's interesting, interesting, interesting field, but let's not go too deep into that.  
Other things which have which have been developed is statistical visualization, graphics design, or nowadays even video.  
These are even more advanced and more modern form of visual communication.  
And of course, this is mainly our focus.  
Statistical statistical visualizations.  
But it's of course somehow.  
Yeah, in this larger surrounding of general visual communication.  
Umm.  
Good.  
So, so much for this general intro.  
So now we'll go to to to some some some things in visual communication.  
So first question, how do you read articles or scientific papers?  
This is paper from maybe the 70s.  
So how?  
How do you usually read then when you want to get the information and the paper?  
What do you read first?  
Probably the title right and then.  
Huh.  
Once again, I if they could give something.  
Yeah, right.  
Right.  
Yeah, absolutely.  
Yeah.  
Scientific paper abstract those often executive summary and and then the rest step by step.  
This is the linear way.  
You don't.  
Do I just read the complete, you know?  
Ah, OK then you read the conclusion.  
Just go back and then you go back to the beginning and see or you stopped because you think, OK, this conclusion but interesting enough for me or yeah. OK.  
Yeah.  
Also, I do it often, but of course I skim the figures right?  
So and then I'll try to sense I can maybe show this is just my, but I think it's fairly similar.  
I think it's also not not bad to do it like that, so it gets some sense of what this thing should be about.  
And then you skim and look at all the figures and try to perceive the message of the whole article by the figures.  
Maybe if you then get already some idea what these figures are or telling, then you but have some questions and you'll try to read the figure caption.  
So therefore, often the captions are not so informative, but some depends a lot on this dialogue of the of the article.  
But sometimes caption are really long and very dense information, and often I like it a lot because then I'll get released detail, especially if I know already a bit about the topics then I have already my ideas.  
What it should be about, so I want to have some background information, so I think caption texts are much more red than some text in the middle of the paper and so, so and then I jumped to the introduction and then I start the normal reading and maybe to the conclusion as Emma was saying.  
So in that sense, we don't really, we don't don't even read the text very often.  
Superlinear sometimes it makes a lot of sense to do it, but even that is not is not always the case, so we use the the visual communication.  
Features so to say that we can jump back and forth also for, for reading or perceiving text.  
So and also.  
And yeah, so visuals are super essential.  
So, and I think many authors, including me, we spend much more time when we craft such an article, much a lot of the time is on crafting the figures much more than on the text.  
I would say, or at least the same.  
Of course also writing the text takes a lot of time and also that needs to have needs to be done with care, but it all goes.  
It all goes with this idea that I want to have a compelling figures which bring a large part of the message around, and that and then I'll write the text somehow a bit even around this.  
So that is quite a.  
So that's why I always have these problems with formal requirements for your assessments with the word count, because of course I don't want to have a lot of text which is very generic.  
I want to have the information very condensed and the graphics very cool and this is much more value I think also for you than very generic text.  
Umm so but so that's why the word count really is not is not what I'm looking at.  
So although we somehow need to have it formally somehow, OK.  
OK, look at this slide.  
If you read everything on this slide and did it work for you and that way or?  
You will read this first and you will read this and then you will read this, then this and then this was it like that for you.  
Nine, what did you read first?  
Like Karen?  
1.  
This one. Uh, OK.  
Yeah.  
And you just skip this because it's something anyway, the main headline, so it's not fully precise, but for you it was that way.  
I I use.  
Uh-huh.  
Yeah.  
OK.  
So we see it's not fully precise, but there is.  
It's not that it goes from top to bottom right, so it it goes some some some other way, but Karen has has a sense for your first reading.  
The subtitle good idea sometimes because it's more interesting.  
The main titles of very generic, but anyway, you read all of this, but there is the main point of this is there is some order which is there is some order but it's not a fully precise order.  
So we cannot control it fully as as as when we do graphics or do something like like design, we cannot fully control who who looks at who looks where first, but we have some control.  
Or, umm, if we can emphasize it somehow.  
Uh, so now I'll what I want to do now is go through some examples of of visual communication.  
Umm, let me remember.  
I just have to get back into the times, so we'll have.  
You shouldn't make a break at 11, right? About. OK.  
Yeah.  
OK so I have I have these 5 examples and I'll have some yeah.  
Talk to your neighbor like exercises and then and then discuss a bit with it.  
So I have something called memes.  
Something very short on 3D videos, data dashboards with interactive graphics and even multi page data stories and.  
My idea for the next two weeks, and I hope I give you some framework to to do this, is that you also provide a few of these examples, and then we'll discuss them, discuss them together in the next two classes until we'll and and and on the side we'll go a bit already into what is the projects of you can be.  
So if you have something so something spots up in your mind, I think so.  
Just remember, in the next weeks we'll may we want to look at some more of these examples, but now let's look a bit more into detail.  
I'm not sure if this is the best choice.  
This is a meme.  
Can you read it?  
Can you see it?  
What does it tell to you guys need before?  
Yeah, you see also ready to go.  
You can see on the slide show view is edging linearly.  
You'll have the cover of the actor.  
I mean, yes.  
And then the the tier bluish is the edge of his friends at the time at the deer.  
For example, in he was 24 in 1999, his girlfriend was 18, right?  
And you can see that another exceeded Erfan, but you staging right?  
Yes.  
And that technology 49, he's getting once again the connection you.  
Yeah, I think this is not really updated, but I think the pattern has continued.  
OK.  
And there was a meme I found I found on Twitter, and I think this good sign.  
OK.  
It's just Sonia.  
Even the bit of a like like a red cut down the road.  
OK, some like maximum point because yeah, in the late 30 seconds of turnover.  
OK, maybe you don't get go to deepen the depth, but the point is that it's really some way of how statistical graphics also are very mean, memorable.  
So to say in the and and and and nowadays I'll come come back, come back, come back to this in a minute.  
OK, some more, some more fancy type of visual communication in 3D.  
We are used to do everything in 2D and maybe if you once produced a 3D data visualization, probably you have found that it's quite hard to comprehend.  
So I typically advise not to use 3D visualizations on paper, but these are 3D visualizations really in 3D.  
I found them in.  
Yeah, just just a few days ago in an exhibition.  
In Leslie Park, anyone knew knows what?  
Bletchley Park is in in the UK.  
It's it's a.  
It's a museum nowadays, but it was a spot in the in the rural area in the UK where the code Breakers are working to break the Enigma code of Nazi Germany.  
Maybe you saw this movie imitation game where there's a nice story about Alan Turing founder, one of the founder of Computing, and they were breaking these codes and this was a big part in defeating Nazi Germany.  
And yeah, as it is a big, big mix of mathematics, computer science, linguistics and everything, there's also a Museum of computing there.  
And there's also there was also these exhibition of about the art of data and what do you think is visualized in this things you you can, can you see it?  
What is it?  
It's all I've cities, actually, I don't know exactly what the, what, the orange and the white is.  
I haven't checked, but probably maybe it's inner town.  
Any any ideas of what cities can this be?  
Why Dubai?  
No, it's not.  
It's not there.  
What?  
What would you think is to buy this one?  
Probably right.  
Is the largest towers, So what I was quite impressed that this these are much much larger than than this New York.  
No, no.  
So I think this is Mexico City.  
You've been there.  
Is it large buildings?  
Not that.  
Not that much.  
Yeah, I think here on the site is probably London.  
I think it's even a bit lower, although there are still quite large buildings from my sense living in premium.  
OK, it's OK.  
This makes the city already Support buildings, but these are much smaller.  
He's Mumbai and this is Hong Kong.  
OK.  
Thank you.  
They have to think.  
They also had London, but I only have a video.  
It's much even smaller.  
Umm, OK, now I'll have have a video for you, which we will watch.  
And then another video on the similar topic.  
Anyone of you know hunts Rosling?  
No, and this video? No.  
OK, then even better, even if you know it, I would anyway have shown it, but it's it's a very prominent video.  
So I think if you take another data visualization course, not necessarily this video, but this story is telling us quite a common 1.  
So it's one of the yeah.  
From where is it?  
And I don't know exactly, but yeah, it's quite a quite a quite an important data visualization video.  
Unfortunately guys already dead, but let's let's listen to him.  
Visualization is right at the heart of my own work too.  
I teach global health, and I know having the data is not enough.  
I have to show it in ways people both enjoy and understand.  
Now I'm going to try something I've never done before animating the data in real space with a bit of technical assistance from the group.  
So here we go.  
First, the next is for health.  
Life expectancy from 25 years to 75 years and down here an access were wealth income per person 404,000 and $40,000.  
So down here is four and six, and up here is rich and healthy.  
Now I'm going to show you the world 200 years ago in 1810.  
Here come all the countries Europe, brown Asia, Red, Middle East, Green Africa, South of Sahara Blue and the Americas Yellow and the size of the country bubble show the size of the population and in 1810 it was pretty crowded down.  
There.  
Wasn't it?  
All countries were sick and poor life expectancy were below 40 in all countries and only UK and the Netherlands were slightly better off, but not much.  
And now, why start the world?  
Industrial Revolution makes countries in Europe and elsewhere move away from the rest.  
But the colonized countries in Asia and Africa, they are stuck down there and eventually the Western countries get healthier and healthier.  
And now we're slow down to show the impact of the First World War and the Spanish flu epidemic.  
What a catastrophe.  
And now I speed up through the 1920s and the 1930s.  
And in spite of the Great Depression, Western countries forge on towards greater wealth and health.  
Japan and some others try to follow, but most countries stay down here.  
Uh, after the tragedies of the Second World War, we stop a bit to look at the world. In 1948.  
1948 was a great year.  
The war was over.  
Sweden topped the medal table at the Winter Olympics and I was born, but the difference is between the countries of the world was wider than ever.  
United States was in the front.  
Japan was catching up.  
Brazil was way behind.  
Iran was getting a little richer from oil, but still had short lives, and the Asian giants China, India, Pakistan, Bangladesh and Indonesia.  
They were still more than six down here, but look what is about to happen.  
Here we go again.  
In my lifetime, former colonies gained independence and then finally they started to get healthier and healthier and healthier.  
And in the 1970s, then countries in Asia, Latin America started to catch up with the Western countries.  
They became the emerging economies.  
Some in Africa follows.  
Some Africans were stuck in civil war and others hit by HIV, and now we can see the world today in the most up to date statistics.  
Most people today live in the middle, but there are huge difference at the same time between the best of countries and the worst of countries.  
And there are also huge inequalities within countries.  
These bubbles show country averages, but I can split them.  
Take China.  
I can split it into provinces.  
There goes Shanghai.  
It has the same wealth and health as Italy today, and there is the pouring line.  
Problems apply.  
Show it is like Pakistan and if I split it further, the rural parts are like all in Africa.  
And yet, despite the enormous disparities today, we have seen 200 years of remarkable progress.  
That's huge historical gap between the West and the rest is now closing.  
We have become an entirely new converging world and I see a clear trend into the future with aid to trade, green technology and peace, it's fully possible that everyone can make it to the healthy, wealthy corner.  
Well, what you have seen in the last few minutes is a story of 200 countries shown over 200 years and beyond.  
It involved plotting or 120,000 numbers.  
Pretty neat, huh?  
OK.  
Any comments on that?  
I don't question Mark 18th century China, Bangladesh.  
Five, I have looked it up how it is done here, but probably it's of course quite unsure how how the numbers really were, but I think there are tools to do it so you can just try to estimate them but.  
So probably was various ways traits that probably they were already there and this times.  
But of course there's also inflation, whatever, but they have some measure to unify it until the 18th.  
But I think there's a lot of work involved in deciding what, what, what, what to take.  
Probably.  
Maybe the life span data is a bit easier to get than the income data, but also that that is probably quite hard.  
OK, you you come or months we haven't come.  
Yeah.  
Yeah, right.  
I think it's GDP per capita and I'm not so deep in economics, but I think many, many economists say when you have the GDP per capita, it's basically transferable into average income.  
Huh.  
Per year. Per year?  
Yeah.  
OK, per year?  
Yeah, probably per year.  
Yeah, GDP is typically per year statistic.  
Yeah, typically, yeah.  
Umm, you know, $40,000 average income.  
Data, and let's say information in a way that it was easy to digest and very simple graph and what do you think was the purpose of Hans Rosling or on the purpose?  
Or what is his main message so to say?  
Yeah, that's one.  
And his face, the message is mainly OK.  
It's it's there are some good trends better than you often think.  
I think that was his message.  
Somehow, in the longer perspective, but OK, let's look at the other video.  
I think it's another Swedish guy, umm.  
How do I get back here?  
My mic also seeing this right a minute.  
In.  
OK, OK.  
This is another guy from from Sweden and he made and this video not this, but this might also seen visualizations like this one with the capita income along this axis.  
Life expectancy along this axis and all the different nations represented by bubbles, by color and size, mirror the continents and the number of inhabitants, and you might have seen how all nations during the last two centuries have enjoyed an ever increasing lifespan and more earnings to buy things.  
What you've seen is true.  
These are facts that can't be denied, and they do say something important about the world today.  
However, there are being used for propaganda purposes to justify the existing state of affairs and hail Western capitalism as the ideal society.  
Both politicians and scientists do this, and they do it by emphasizing some facts to the extreme, like completely ignoring other facts.  
Take a look at this in bubbles for instance.  
This is really what the world looks like.  
Look at the US at all.  
It's inhabitants neatly packed in the small little bubble.  
Why is it so small?  
Because it represents the mean values of income and lifespan, not reality.  
In reality, all national income is unequally distributed.  
So among the bottom half of the American population, the average income is $16,000 per year, much lower than the national average for this half of the US, income has remained roughly the same for four decades.  
So why does the national average income increase?  
To understand, we have to look at the entire distribution up to the top 1% and to do that we have to zoom out a bit.  
Among the American top 1%, the average income is $1.3 million.  
Now listen carefully, since the share of all national income going to the top 1% is almost twice as large as the share going to the bottom half of the population.  
Whatever happens up there as a huge impact on the national average.  
And so during the last 35 years, the top 1% has more than tripled its income on the poorest half of the population has remained just as poor.  
That's why national averages can be so misleading.  
Inequalities like this affect how long people live.  
An American woman belonging to the top 1% has an average life expectancy of 89 years, whereas an American man belonging to the bottom 1% has an average life expectancy of 73 years, about the same as in Bangladesh.  
Mind the gap.  
Yes, they have a gap at about 16 years and the US is not that exceptional in this respect.  
If we zoom in on Sweden, where I come from, a relatively egalitarian society, you can take the Stockholm tube from the well educated inhabitants of down the road to the less educated inhabitants on voice record.  
This price will only take a few minutes, but when you arrive, life expectancy will improve 18 year.  
OK, OK.  
So the videos a bit longer so you can can can watch it in full if you want some other time.  
What are your comments on this video?  
If you there are some umm.  
Yeah, that is a problem with the average right there is everywhere that whenever people were talking about the average, they can use it to.  
That's their side of the story.  
On what they have on to what they want to say to you, that's right.  
But yeah, so it's it's so every dot is very much aggregated into into one value on two dimension.  
So I think he's a very good, good point.  
But he said in in reality it's different.  
This is of course a bit the question what is what really is reality?  
Of course, the average is also some reality, but it decomposes into a lot of individuals, and main point is from the average we don't know what the inequality is, so to say right?  
So is there a way to also put this information in here?  
Yeah, I'm can think about very clever data visualization, but probably it's not possible to put put everything in.  
But of course, yeah.  
What?  
What is he doing?  
How is he showing the inequality?  
In what way?  
Just looking at the at the data visualization ideas.  
Insulin in inside the average, which are the maximum, the minimum right?  
So this is this is something which is nowadays possible, but still a lot of work with something like interactive graphs, right?  
So you can he's he's looking, focusing on one thing and then showing some some ranges or some percentiles and thereby showing giving you a sense of the variation because it can also be maybe it's even for some countries the variation is not so large, but he's saying mostly for every country is large.  
And I think it's true, because all countries have large wealth differences.  
Some people or income differences, some more, some less, but all have a lot of a lot of income differences and also lifespan differences.  
Anyway, you off?  
Usually between men and women, but also also between different occupations and so on.  
Umm OK.  
I have also more another another question.  
These are two screenshots from from the from the video, Umm and.  
This is I think this is more or less the same data.  
Maybe the the the the the the other guy.  
Let's check name.  
Powelson has has umm it.  
Maybe he has slightly updated data set so but so here you can compare.  
So this is 2010, this 2009.  
So roughly the same the same spot.  
But why are the bubble clouds looking so differently?  
So maybe you can can briefly discuss with your neighbors and then come up with all the differences you spotted.  
The more the more precise the better.  
So please, these are the things which we have access scales, color, size got.  
Which is so.  
Yeah, I think it's much better.  
There you go because.  
The right one is easier.  
Yeah, absolutely.  
Yeah.  
Why are you?  
Yeah, income from zero to 100 carriers.  
There's you got what?  
Yeah, cool.  
Tons of them.  
That's all the color.  
Yeah.  
Also, the come to me, but yeah, the on the white Board, the distance you, uh.  
So when when you're like, why access spends 100K, then a small distance between two points means something different than the entire scales.  
One kid, that's all.  
But there's different does not make.  
Yeah, it's the same.  
It's not that all.  
You.  
That's right.  
That's, I mean, they're really the right to privacy you have.  
Some more.  
Yes, conference.  
This which?  
Are different questions you know and well what is?  
The message that. Yeah. Yeah.  
Yeah, exactly.  
It's not gonna.  
Actually, maybe realize how the which?  
Yeah, among the.  
And I think that what is the problem?  
What?  
Yeah, I think it's.  
Close.  
They probably and Blessing the presidents can be like somehow at least.  
I mean that that's very fast.  
Would that impress?  
That's right.  
Now that's what, like some stuff, it's.  
How?  
Yeah.  
Awesome.  
Let's so they have the right.  
Yes, yes, yes.  
Times, which is almost redundant, right?  
And there is no human at all.  
Right.  
And and dollar, I guess that you could also be EUR for what you know on the right, I mean, so they are always doing this weakness mistakes.  
And OK, let's let's collect together.  
So what difference is did you spot in the back?  
Yeah, so.  
With this, we're talking about how, like condensed this his honest it's really hard to tell or put into fractions while income each continent or each country has.  
Here you mean?  
Yeah, one I can clearly see in the press wasn't blue does have like $500 to $4000 and you can easily see that in compared to this one.  
OK.  
So yeah, and and what?  
So what is in the difference?  
Just just technically, so to say in the in looking at the data, yeah, the sorry. OK.  
OK. Yeah.  
So the on the left, the income access goes from zero to 40K and on the right it goes from zero to 100 K, meaning on the left everything is put more pots compared to the right one where everything seems to be more close to each other because of the the entire scale is latter right?  
OK, yeah.  
But it makes sense from 2010, with the income has progressively been higher and we can see what, like a dot which is in the 90K income and if it here are there.  
Yeah, yeah.  
Everyone can see that one, umm, it's freezing dead, but I think it would have been better for the lifespan to become 40 to 80 because we don't have anything that there shouldn't.  
True.  
And we would see the size better I think.  
OK, I will summarize it in the table soon.  
Yeah, and.  
What about the other two other two things?  
I think 1 essential part missing on scales where you got already.  
OK, there are different scales.  
I think you stayed at his data is going a bit further.  
So that's one reason, maybe because when he has different limits, but OK, any any comments on counter and size, what is it the background, the first one, the windows are quite bright.  
OK, right.  
Yeah, 7/1 the ground is quite darker and you can see it more clearly.  
And also the other of the access, I think this was a TV recording and this is more than more modern.  
So this is much this is just not so good quality.  
1010 orders and something.  
OK.  
But I meant colors as a visual.  
So what do you see?  
Any different?  
What does it measure and do you see any difference?  
Yeah, the.  
No, no, no, I mean just the colour of the different, the different colors.  
What do they mean?  
Just very basic, so to say.  
What?  
What is the color representing continent, continent, continent and this is the same.  
I guess it was actually.  
She's and.  
OK, they've got some presenting. Yeah.  
Yeah, yeah, yeah.  
I think the color is not exactly, but it's hard to.  
It's hard to compare.  
So for example this, this, this, this Falk, maybe you remember because it was talking about this a lot.  
You would say where is it there just to give it a while?  
Which one?  
This one?  
Yeah.  
So OK, so probably there is justice.  
The America has probably green there and yellow there, and the rest I think maybe they match, although I don't really see the orange ones here on the orange ones.  
OK.  
But probably blue and red is Africa, Asia and yellow and green.  
Maybe they are.  
There changed.  
OK.  
But so it's not, it's not so hard.  
Not so easy to to spot what is there.  
OK.  
And size what is size representing?  
Uh, everything the most right size of the dot of the dots.  
That's what I mean.  
Populations population, right, OK and.  
And what do you get so with size?  
Besides, they are.  
That's always the question.  
OK, what do I match to the size?  
The radius of the dot or the area of the dot, so that will be the radius and that will be the area.  
And actually I'm not perfectly sure, but do you get the difference between area and radius?  
So it can be so that the population of a country, let's say a billion for China roughly or India.  
And I think at that time not for India.  
But anyway, so matching it to the radios or matching it to the area, what would be the?  
What would be?  
What would be the difference?  
So let's assume.  
Let's assume the radius for for China.  
But, and this is China and this is the radius.  
Funds that it's previous if it's four times that it's the area.  
What?  
Wait a second.  
11 million thousand million and let's say what?  
What this?  
Germany it's it's 18,000,000.  
So it's just.  
Probably like like this.  
From the radius roughly.  
Radius.  
What would happen if you go if you change this now making it proportional to the area?  
So now it's proportional to the radius and Germany is much smaller, much much smaller than China.  
And what do you think?  
Is this proportionally?  
What should it be?  
It's so this should be somehow 8% right from compared to China and this would be 100%.  
Is it matching?  
Maybe it's not fully up to scale, but my point is what is?  
What is it better?  
What is the better measure and what is the better visual representations or radios or the area for the population?  
And what would change so if this is this is a correct proportional representation mapping the percentage to the radius, what would change which is German bubble be a bit larger compared to China or bit smaller?  
If you go to areas.  
Small other times more 100 times smaller.  
This one.  
If you go to area or why then we don't have now all the matches.  
The matches between the area 1000% and 8%.  
Umm.  
And if you go to the area, it should look like.  
This should be larger I think.  
But maybe I should make it more precise with a good visual and we'll reconsider it again. Umm.  
So let's let's go.  
Let's go to squares.  
So if we have, if we have China here, then we'll may have.  
What is this 1000 \* 1000 or it doesn't work? Umm.  
So if this is the 1000.  
Uh.  
And then Germany would be 8%.  
So.  
But so sorry, hundred hundred percent, 100% and 8%, right?  
So now it's more or less up to scale, but this is this is matching a square with the side hundred with the square with the side 8.  
But this is not proportional to the area, right?  
So this looks much larger than this, so the areas are not proportional.  
If you go to proportional areas, it should be probably more like more like more like this or something like this, right?  
And what do you think is what do you think is better radius or area for this for this size of the dots?  
If they should match something like a total population.  
Useful area I mean area is more intuitive that would say, but a bit harder to Arthur to what what is the yeah, that's right.  
The construct.  
Kind of, yeah.  
Yeah, right.  
I think we'll reconsider that again.  
But I remember it's an important topic and I think for the proportionality aspect is the area is much better.  
So that is typically the choice you'll do.  
Because yeah, when we'll think of, OK, so when when this should be.  
So this should be 100% and this should be 8%.  
So 8 patches of it should fit a or you know, so the patches of this should should should fit into this.  
So if you go to the if you just go on the one dimensional measure, then this the large numbers get way too large.  
So say because we will square everything, right?  
When you go through area, we'll square everything.  
So when we'll have 100 side lengths, then the whole thing is 10,000 in area.  
So in the middle here we have 10,000 and in the middle of here we have 64.  
So the question is 100 to 8 or 10,000 to 64.  
And of course, 100 to 8.  
This is much larger than this, right?  
So so here are the difference looks much larger.  
And then so now here the difference looks much larger than here, and probably the best way to visualize when you use use size of two dimensional object is to go with the area.  
And what does it mean to?  
It has to do something with this square root.  
Typically, when you're when you're match it, when you when you want to, when you want to make the the the area of a dot proportionate to the actual number you have to.  
You have to take the square root of all the, all the, all the values for this size.  
I think in ggplot, if you remember it, you probably use it soon.  
There you even have something like size, area, visual and this is typically what you should use if you use it like that.  
OK.  
But anyway, it's a good idea to look up the mass and make it clear for yourself.  
If with, with, with circles you may be just a bit confused because maybe you have forgotten the formula right, it's like this.  
I'll grab hands pie.  
So, but there's always a square with square.  
It's just easier if you see the square root immediately.  
So, but it's essentially the same, so you can just take the square root of your values and then then it's proportional.  
But OK, I wanted to check if you already in the in the break.  
Few minutes.  
I think your your reconsider that at some point again.  
But I wanted to check here.  
The fluffy.  
So when we look at the data set behind that, what is the data set behind it?  
What variables do we have?  
Number of operations once again, number of operations population, yeah.  
What else?  
Or life span? Yeah.  
And GDP per capita?  
Let's see people capita.  
Or what you called income.  
Also, we have a country and continent.  
I think that's called it.  
OK, that you come.  
And what not mean.  
Yeah.  
In countries both right in and OK and country names to say yeah.  
The also the years, years, right?  
OK.  
And now we have everything right.  
OK, then let's first look at.  
Ross Ling.  
So maybe you remember the grammar of graphics of of which is the basis of the GGPLOT package.  
So and in the ggplot package, you always you'll start with some data and then you will match this match your variables to athletics.  
So when you look at Rosling, what is he putting where?  
So population he puts to what?  
Esthetic.  
Maybe you you remember all the aesthetics?  
Maybe you will come to them.  
You'll find them out.  
OK, what does population matched to?  
But it's called size.  
Yeah, size that can be for any object that can be a size.  
Typically for dots or squares you have size.  
What is the matching lifespan to what?  
Why umm and ecom?  
And continent color.  
Country no.  
No, and and here's.  
It's.  
The same as last time, right?  
Why?  
I think it's not a visual, it's not a visual in the standard ggplot, but it's we can think about what?  
Where is it matched to the year label?  
OK, label.  
But there's another thing which is not not the static plot.  
But actually we don't in this.  
In this plot there is.  
There's just one year, right?  
So it's not really a very, it's not really a some variation in it, but there is variation in in what dimension.  
And the number of bits in the white.  
No.  
So the number of years it's, it's just time at the same if it's that's except for the year, the year is just is in, in, in the video it's just progressing with time.  
So that it's that point.  
So it's it's not.  
It's not something that you can put directly into ggplot, but if you make a movie, you can also always think about putting some dimension on time.  
And of course, it's a very natural one, because it's also time.  
So to have the years over time, but in principle, we could also say let's go through the GDP with the video time.  
But it's a weird thing.  
Not often done, but so this is that this is this is more than so.  
The label, of course is.  
It's some correct, but it's not.  
It's not the whole we see only a snapshot of the data, so for this.  
For this static visual, it's just a filter for the data, but not the variable this variation, but this comes with time OK before the break.  
Maybe we'll can also check it for Paulson.  
So population.  
It's the same right size.  
So what is the different?  
The the access this two things and the rest is the same, right?  
So it's here X and here Y and it's the same color, just a slightly different axis.  
And then here we have again.  
Time.  
OK.  
There is something like a A, A.  
This would be the esthetic as ethic, and this would be the.  
There could also be something about the scale.  
Umm.  
And it's scale.  
There can be several things that can be data transformations.  
There can be limits.  
You already discussed the limit the limit.  
Yes, thetic.  
We just have to not, OK?  
We have this so we have discussed that we are probably sure that the scale here is with the area and not just the radios, but I haven't really double checked it, but I'll guess it is like that.  
Because if you compare, let's say, China, India with the US, I think the the differences in the size should be the US should be four times smaller or three times smaller, more roughly.  
And I think this matches the area probably well, right?  
The area of the report is more or less three times the area of the yellow but not the radius and gets.  
Yeah, radius is just two times at most.  
So therefore, I'm pretty sure it is the area and this is also the the the correct choice.  
OK, we have already that you had this point that this is more or less 25 to 75 and here it's more here he has this from zero to no it's here it's sorry the lifespan.  
Yeah, that's good.  
Yeah, it's from zero to what, to 8080 roughly.  
And income is from.  
From zero to 100.  
OK.  
And here it's from.  
Umm what 400?  
Once again, I need the but there are some points to the left of the 400 though. Yeah.  
Yeah, right.  
Yeah, that's right.  
That's roughly so, yeah, you're not by probably here.  
It's a bit, yeah, something that's beyond that.  
Right.  
And also below that.  
But there's one very important difference between has not discussed yet in the scales.  
Any idea if you reconsider it?  
So if it only so is it only the fact that OK, here the artificially says.  
OK, I'll extend it because there could be really 0 lightspan, so I want to have it the full range.  
OK.  
So is it just the difference that we need to we need to cut it here and cut it here to get to the to the.  
So if you do that, if you if you would think it to this, then still the variation here would look different than the variation here, right and white.  
This is the income also.  
There's some different scale in the income which we haven't discussed.  
It has to do with data transformations.  
Yeah, like one of the linear and the other one is like, like, yeah.  
Yeah, yeah.  
At least for the first, let's take this.  
It's logarithmic, that's right, but the most important first recognize this is not linear, right?  
So there is 400 and then we go here, it's 4000 to 10 times.  
Let me go same same same distance and now it's even again 10 times more.  
So it's it logarithmic scale in the sense that you 10 \* 10 times.  
Yeah, while here it is plus 200 + 20 K plus 20K plus 20K.  
But here's times 10 \* 10, and that is the typical logarithmic scale, but it's a logarithmic scale, so we'll still have the labels in our original our original value, which is often quite useful for data visualization.  
So yeah, but what is here?  
So probably the scale is logarithmic, sort of cold, but here with nice labels that are still interpretable.  
This is something quite important often for data visualization.  
When you make it technically in your, in your data exploration, you often just take the logarithm.  
Makes the plot, then look at it again, but then you have values which are not easy to interpret, only with your background knowledge, but for making it out to the public.  
It makes a lot of sense like this, but of course there's also some importance to say, OK looks these people have 10 times more than this.  
Yeah.  
So and this dude is 10 times less.  
So these are 100 times less, so it's much, it's much the yeah, the the absolute differences are deported.  
So this is here lock scale and this is linear scale.  
So how could we then?  
In the Handwaving style transform, transform this data.  
Let's take a look to this data.  
Let's say if you have the cloud like this, what you need to do like this?  
What we have to do to bring it to this graph?  
You have to somehow flip it, flip the axis.  
So so mirror it like here.  
So all this should go up there and yeah, and then you also have to, when you have to feel, then you have flipped the X and the Y when you when you mirror it along this axis, but then you also have to switch to to to switch this somehow right to bring to bring to bring bringing this in a bit more up but also stretches of say right so in that way it is the same it is the same data OK.  
The only thing which we haven't here that there is slight different in this color scale is this Americans in Europe.  
But I think you know not so essential.  
OK.  
And this is this is how you can make with the same data quite quite different graphics.  
And of course, there are even more more options to change the graphics and you can play with this.  
And this is the first entry point, then now conceptually thinking about how to present it and why to present in this certain ways, let's discuss a bit more more of these type of examples than after a short break. OK.  
You.  
But that's great.  
Oh yeah.  
55 after 11 if it's OK for you.  
Yeah.  
The thing is like for most of the coffee line.  
Ohh yeah, get the something.  
OK, we have.  
OK.  
Yep, Max. Sorry.  
So yeah, so with my registration, when I tried to registrar.  
Umm.  
What is the problem?  
OK, you are in at least.  
Yeah, I need.  
On two data science tools to register for this.  
So you miss a requirement or yeah, with wine they could sense tools.  
There should be a great right.  
You're doing this.  
What do you think?  
Like I can read it, I can read it.  
Unfulfilled.  
Since tools.  
Don't know, don't means past.  
Uh.  
And have you already contacted create this trust on this algebra?  
Yeah, if you want by email.  
Probably right.  
Yeah, you can see series, so I can confirm it looks like an error on their side because it's also.  
Yeah, it's also confusing messages because push Jan is German word for past, so you should have the requirements so it should not be missing.  
Yeah.  
And you also have.  
You also have the great and your records.  
Yeah, I see.  
Yeah, yeah, we have this communication unfortunately wasn't possible.  
Yeah.  
So OK, go to the office to see.  
If not, I've just been.  
Yeah, you can send an email and you can can see.  
See me then I can maybe when they have questions back, I can confirm it that I don't know because I don't know anything which I can do to solve it because I think everything is done.  
Even when you say you have, this looks like weird error.  
So I think this should be done.  
So what?  
You know the times you can just come.  
I can add you to the team, then probably right.  
Yeah, yeah, yeah.  
Then you are maybe in.  
Umm, OK, I'll do that.  
And then the other should be solved and probably soon.  
OK.  
And by the way, you also you got already information for the contract.  
No, no.  
OK, I think for Stefano it worked already, but this was the old call so I think mine are new.  
I think they are in the processing.  
Same for Leona and so I hope that they will contact you within this week about the contracting and hopefully they could start 16th of September.  
But anyway, the thing was the only Student seems to be a bit delayed, but I think bit of September I think we will be able to to to really start.  
OK, that's all it works out.  
Student.  
So coffee like this, too slow.  
I see if it would work with 50 minutes, but I think even that it's not feasible.  
The one that I hope it will contact me will start this week, but in some.  
Yeah.  
We certainly want to step on all, but I guess it is because this thing was approved a bit earlier the the cost center from central funds for these big course and the other positions are from from where I just made the.  
Called the.  
The request to have this have this have this once assign this bit later so I hope I guess not starts this week.  
Maybe I should put an ask, but I guess maybe I hope this week it will come off that you're very short notice because of course you can imagine probably that often.  
Like maybe in the in this time of the year.  
So it would be good.  
What?  
What?  
So there will be.  
Delays for?  
Yeah, it was.  
It's it's just, there's a lot of.  
I think they have a lot of Peter process.  
Typically they try to make it until exactly the state 16th of September we're starting, but I hope it works and I think it should.  
It should come then information to you this weekend when you are very fast and have your materials ready.  
I think then it should.  
It should work so, but let's see this when.  
But there's no contract.  
There's no conflict yet, of course.  
Yeah, but I think it it is still, it's still normal that it's it's quite tight in the schedule.  
So that when they see use, believing me to have a two year, two weeks before, but they have the next posting there approved and they have your names, so they should contact you asking you for the materials, but maybe they have even more so.  
So you could get everything so still still in the OK timeline.  
Yeah.  
And also somehow I'm still missing all the online students.  
There should be a few, but they haven't appeared on campusnet yet.  
So and also for the new students on site, they are just I think 9 which is basically I think there should be a bit more but maybe it's just that so so there will be something to do in September but?  
At the moment it's more like, OK, that's fine.  
That's for the students.  
Find out.  
So guys, where are you?  
It cannot only work with.  
Silly owners or I can why not?  
Well, it's already 15 minutes.  
So just doesn't work, even with 15 Scottish.  
Good, no.  
So.  
So hmm.  
No one of the.  
So I've already you are late, right?  
Even the 15 minutes.  
Yeah, 15 minutes exactly. Yeah.  
So.  
I think what we'll do is I'll just continue anyway.  
The next story is probably the adult go super deep into, so it's not so graphic.  
Skip it and then the others when the others are back, we come to the to the next topic. Umm.  
Otherwise, they need to watch the video.  
It's still running.  
Yeah.  
OK. So.  
I wanted to build on the what the the the Hans Rosling and the Powelson.  
I think you've got a bit this that this is also in some sense of different political message, so to say, right?  
One was OK.  
See progress in world development.  
Other ones you also problems and world development and but it's the same data, so to say in some sense.  
So you all have some German course, right?  
You know what this means?  
And the German book.  
Did something?  
Yeah, Rickster is like, right, huh?  
Right Rickster is right.  
Data.  
Yeah, and link is left.  
And what do you think it means?  
Of Falk, political yeah.  
Political politically, right?  
Yeah.  
Be like.  
Yeah.  
Let me think.  
Think you the idea is you can look through this on the same data was from different sides and you'll see different messages depending on how you present it.  
Umm, how about this?  
What does this mean?  
But whom do you know?  
What does the end must be, as in home? Why?  
See why we only see what we see, what we want, what we want to see, right? Yeah.  
Where we only see what we want to see.  
That's that's the point.  
Yeah.  
OK, by the way, short fun fact I once heard an interview with the author and she said some some Gen Z or people got the title wrong because Dayton this means starting it means data.  
But Dayton is a German word of to date, someone so so you could also read this as stating left wing people and dating right wing people.  
But it's not, so it's about the data.  
Yeah.  
OK, but this was one graphic from this book, and it's called the elephant graph.  
Do I have it here?  
Yeah, the elephant graph.  
And what we see here.  
You see here percentile of global income distribution.  
So this is a top as a bottom 5% of worldwide income distribution people.  
So probably here in Germany we are all more on this side.  
Even the poor people in Germany and the rich people in Africa probably are also more only up to here, except for the Super rich there at the same scale.  
Maybe so this is the top 5% of world income owners and here that it's it's something about state change from from 1988 to 2008 and here you have the the increase in percentage I think but it was so percentage change in real income.  
So how does the income of the top five, top 5% more than 5%, middle 5%, so to say of the population, how does did this change in percentage uh from 1988 to 2008, so in certain 1020 years, so to say.  
So I'm gonna say, what do you think is what for you, is the message of this graph.  
So there is an increase in income almost for everyone, but it depends on what part are you.  
Are you among the poor in the world, among the middle poor, among the middle?  
So say middle income, middle class, upper middle class or or or really really upper class.  
What is?  
What is?  
What is the message of this graph?  
Who has who has the highest and who has the lowest increase of income?  
Or and like it's the countries it's supply people.  
But of course, it's really the countries.  
Yeah.  
One second, I think grease for us, the richer site.  
Yeah.  
Has doesn't have.  
Actually it hasn't.  
Case right?  
This one this guys.  
So something like upper, upper, middle class and the world has has not really increased.  
Super rich have increased and also also super poor, no increase, but the all the lower lower middle class were not fully poor but still poor.  
They have the highest increase, so and who do you think and not necessarily left and right, but who do you think?  
Who, who, what, what type of people could use it for political political message?  
So to say or what message could you propagate with it?  
They should be the honest.  
In what message in Forex and I think for for for middle class people can can be informed in some different way because they there's not really differences between income, right?  
The section of that which one you talking about, these people, the middle. Yeah.  
Yeah, no.  
This one.  
Yeah.  
This one that that's great.  
Maybe because because baby, baby.  
Somebody go purpose of some political policy might not really affect through the for more we we do right it site and lift it right.  
Yeah.  
So and what what could be?  
Yeah, you had an idea for political message.  
He grabbed the tension from the majority of the populations.  
And say that the income is increasing and ignore the.  
I'm like, OK, let's phrase it a bit more concrete.  
So if you want to have a political message, we are suffering the most.  
Who would be?  
Who would be this given to the given to the given this graph and whatever political message I want to, I want to say, OK, I'm for I'm here for you and you are the one who suffered the most.  
Though I have to make you, you you elect me, I will do something for you.  
That would be a message for for people in this range, right?  
Because they from this perspective, they are.  
They have no increase in the income the last year.  
All these people have increased all these poor people.  
They have so much increase of income, but we are suffering so that could be a negative political message trying to motivate these people to stand up sort of saying the upper middle class because they were they were losing.  
So in that sense, it could be a message for them.  
Of course, they also be a message for them.  
You are there also suffering that anyway super poor.  
So it really these are really suffering people and of course then the question is if you have, if you want to address these people and say you are suffering, you can phrase it.  
OK, let's take it from the Super rich because they have gained a lot.  
Let's take it from the of these tours.  
They don't need to be more Support now because they had so much increase, right?  
That could be the story.  
Not be salaries.  
Operation once again I'm I want to umm decrease the difference between the incomes and our.  
Yeah.  
And just try to like grab the attention from before population and say I'm going to make your income higher just so there will be no differences between.  
Yeah.  
And the rich things like that, yeah.  
Right.  
But here it's it's very it's very it's very both accesses are very not Support trivial right?  
Because this access is percentage of the income, so it's classes which changed all the time and we don't know the absolute value.  
We don't know the absolute, just know these other bottom 5%, these are the top five already hard to rest and this is not just income it's income increase.  
So percentage increase, so they would say, OK, the last 20 years we didn't gain the last 20 years.  
These people gained OK, that could be the political message.  
So for people who come late, just can confirm even with 15 minutes break, it's not possible to get a coffee.  
So get it before.  
I'm actually thought somebody next.  
OK.  
We just started because I would not go so deep into that.  
So it's a bit sight topic similar to Rosling and Paulson.  
Paulson, this is about politicized visuals.  
So we can see very similar data with different political messages.  
And this is another example with a complicated graph which is called the elephant graph, and it's mainly saying let me see if I have it.  
Yeah.  
So so I think this graphic is that often used in Vestal European.  
So here and US middle class to send messages to them.  
OK, you are suffering the most in the world because you are the rest of little classes and see all the African middle classes in the Chinese middle class.  
They have gained so much income, they have 60% more at you have 0% more compared.  
So that is why how this data can be politically used in some way?  
Umm yeah.  
And it was. It was.  
Yeah, of course, Sean somewhere.  
And now let's look at.  
Another version of the graph, so this the other one.  
The black one was called the elephant because some people see militants.  
There were upward trunk once again.  
Yeah, 2 bit, but then the other one is the famous hockey stick.  
Also used a lot in climate graphic and this temperature increase, but now it is this hockey stick which is just going up here super much.  
So when the red it's not really so this is red and this is black.  
So I wanted to the red.  
It's the same idea.  
We have this not so easy to grasp the percentiles here, so lowest 5% of income worldwide, upper 5, middle middle 50%, middle 5% of income.  
Umm, but the the the difference is the is the is the Y axis and this is the.  
The red One is the absolute gain in dollars, so how much more dollars you have?  
So and here you see a different picture.  
What does this communicate more than the black one?  
What is the? What is?  
Yeah.  
When you belong to the 90% or more of the highest income, then you can see the the crazy increase in yeah you have.  
You have a lot more in absolute dollars.  
While all these people, they may have gained in percentage, so they have, let's say 60% more, but in absolute terms this is just yeah, just a few blocks, right?  
So, and that's that's that's saying that.  
Yeah, OK, this this black black graph has has some interesting message in some sense so that people there were increases in the in the in the in the lower and in the middle class or Chinese middle class African middle class they have to be in these increases.  
But compared as the total wells of the world, so to say or the total income of wealth is still there, it's still tiny absolute numbers and even these people here, they have the same total increase of even have more increase, it's just more increases small increase percentage wise.  
But in nominal terms, it's even more than the lower, the lower, the lower class. So.  
Umm, you get the point so.  
It's not, not not really.  
Someone lying with graphics, it's not the issue of lying with statistics here.  
Some people say it often.  
It's anyway overstated.  
The lying thing is just that you'll look at data in a certain way, which is not necessarily wrong, and can we have interesting information, but for a complete picture.  
It's always good to have different types of looking at data.  
Umm, one question can be inferred when we know this graph and being for this graph from it or you only know this graph and the data for it can be infer this graphic from it.  
Just technically, you know, we have to think quite backward what is probably behind it.  
Cannot be just a guess.  
What do you think when you only have this information?  
So only have this data points to say can we?  
Can we infer what this is?  
By the way, these are typical typical questions you could learn to ask.  
Ask to to data or to think.  
Think about what?  
What it is?  
OK.  
In the black line, we'll know the the change, so it's this axis, so plus 20% + 40%.  
So we know, OK, the lower 5% people, they have gained 15% and here we know the lower 5% they have gained essentially 0 or just one 2-3 dollars.  
So and we cannot infer this from that because we don't know, we just know these are percent percentiles.  
We don't even know the income distribution here.  
We just know this is from the from the bottom 5%, but we don't know how much they have and also here we don't know how much they have and what do you guess is of course probably against some very school distribution.  
So probably this has very little and then it's more that it increases the absolute value.  
What is the?  
What is the average income in this group average income in this group?  
Every income in this group, probably this has also a logarithmic scale, so probably it's multiplying over here to some 100,000 or millions or billions.  
Millions.  
I'm going over there and it's not just that there's always $100.00, so it's so we're looking at percent percentiles, often quite.  
Interesting way to do, but it's also quite processed data.  
Wrong room, but this is the IC.  
No, I don't know.  
CC. Let's see.  
Umm OK so this.  
Are the black, is the the black, is the relative income growth, meaning percentage growth.  
So when this guy is hit at uh, when when these guys had had $1000, they have in 2000, in 1988, they have now 40% more than means 1400 in 2008 percentage change.  
You look at the right and then you look at the right.  
Yeah, because it's OK.  
Another topic which just can also come up here, often in data visualization task is this is called do will do will Y axis.  
So we have one Y axis for the absolute change we have one Y axis for the percentage change I.  
Typically it's not very advisable to do this.  
There may be cases where it is interesting to do it.  
Compare certain trends, but normally it's quite dangerous because there's no particular meaning.  
So that's at this point, but the the, the, the magnitude between this because it's arbitrary, how we scale, how we scale both of these axes, right?  
So therefore it must, of course we can compare this factor here.  
We see this large elephant and absolute and relative increase of income, but here we see an absolute numbers.  
It's very little, so that became compare with these two graphs.  
But to read it properly and quantitatively, we'll have to look for red here.  
And for black here.  
So and basically the.  
Yeah, the black line can be used to say, OK, we, the Western middle classes, we have suffered the most or we have not not gained much the last years.  
But when you look at the red line, you can still see, OK, the poor people, they are still quite cool.  
They have some gains, but they are still much poorer than us.  
So which is very natural because it's percentile.  
OK, so this is this can be two graphics of of.  
Left wing comically left wing data and politically wiping data.  
It's the same data, but it's just different way of presenting and there's some truth in votes and.  
For a complete pictures good to look at more more versions of it.  
Let me see how we are in time.  
So we should also talk a bit about the the course progresses.  
OK. Umm.  
OK.  
Then maybe it's still some time.  
OK so.  
And OK, let's let's discuss the second point a bit without discussing this neighbors.  
Maybe you have some comments on that.  
So let's assume so here we have data.  
Here we have data percentiles or five percentiles and relative absolute income or relative absolute income change or relative income change.  
Now let's assume we have raw data.  
Raw data would be we have the data set of all incomes of all, just the incomes, really the incomes of all people once, and all people of the world, or the it's a five or five 5 billion year, let's say 6-7 billion year.  
Umm, so I assume we have all the data, all the incomes.  
So how could we make them the red and the black graph?  
Just what?  
What steps we need to process them?  
Of course, we don't have the data, so that's a bit of a problem.  
But if we have all the.  
Income names.  
What do we have to do to to compute to compute these graphs so this is now a question to the data processing data transformation.  
So what do we need to?  
Yeah, several steps if you have one in mind, just come up with one.  
We need this exercise, of course.  
Right.  
How do we get them?  
How do we get them?  
What's the functions?  
What we need to do, so I mean for the income growth, we just need the you can divide the income in year n + 1 by the income in year N then you get to income growth, right?  
OK, so we have years.  
Let's say we have years 1988 and here we have 2008.  
Yeah.  
So to get the income change, yeah, just need to divide the income in one by the other.  
I think it's better to, yeah, but.  
I think both broke right.  
Yeah, this is not.  
This is not an ideal data format right for this.  
So yeah, so we want to have a new data data form.  
Thank you.  
So what we will need is we go from, we transform it.  
In what way that we'll have that?  
We'll have a 1988 here in 2008 here and here we have then one and yeah.  
And then we can hear compute the percentage change.  
Yeah.  
So you can use pivot wider from the type and then you can right?  
But is it right?  
Is it that what we need that we have for each individual?  
How much it has gained, of course.  
Some of these people are dead.  
Some of these people are new in the labor force, so probably will not fully work.  
And is it what we need?  
It's the right step, but.  
It's just just fine or not, but you need one step before, right?  
Because we want, we have to.  
We also want to have the income change of the percentiles and not the individual change.  
The present access.  
Yeah, right.  
So, so.  
Oh, this is not good here, but in principle was a good idea.  
This is one thing which we need some change, but first we need to compare first compute compute percentiles.  
For each years, yeah.  
For each year.  
OK.  
So we need to condense this data set from 1 to 1,000,001 to a million.  
We have to convert it into into percentiles into something like a per cent tile.  
So let's say the 5% five percent, 10% depends on how detailed you want to have it. Uh.  
So probably it's zero to 5%, five to 10 percent, 95.  
Typically you want to have this even even smaller chunks.  
The top 99, the top 1% this as the top 5% of the world works the top 1% or whatever percent.  
OK, so we have the percentile and then we have here something like the income of this or average or median income of this group.  
So we have this, you know.  
And then we have this for 1988 and then we have it for 2008 and then we'll can compute the change.  
Yeah, but dividing, I would say dividing 2008 by 1998.  
Yeah, but then you get you get, you get the percentage change in the divide this by this change 2008 / 1998 and then we have another this the relative change enough and what is the absolute change and that would be 2008 -, 1998.  
This is something else.  
OK, so we have to do this and then this is.  
This is then from here to there and there.  
Right, this is our.  
This is our Step 2, right?  
And that's real quick.  
Do these two graphs, so it's already it's already highly processed and the problem is the bit it's percentiles is reduced.  
It's a bit hard to conceptualize, for me at least, because these 5% in 1988 need not be the same 5%.  
So some people have become richer, some have become died or whatever.  
So it's not so easy to process, so there can be several changes happening happening in the meanwhile.  
Also, the income distribution may have changed a lot, so maybe the income has become much more unequal or much more equal at.  
In the meanwhile, it would have an absolute numbers and have some drastic changes which we don't see in this graph.  
So there can be can be.  
So looking at this graphs and interpreting them in full context requires a lot of background knowledge.  
Of course, when we have this numbers, you could also first look at the distribution of the income or something like this, which is all a hidden here.  
Still, it's interesting to see, and often I think with income, often you don't have the incomes of all the individuals you have this thing in Texas over to start not for all individuals present.  
Please.  
Yeah, yeah.  
Do you think how often does it happen that somebody who is in one in 1988 in the lower 25 percentile and how many make it out of that percentile into one of the upper percentile?  
Now, because it's also worldwide.  
Yeah, of these numbers, it's really, it's really, I don't know, but I think there's typically a normal a normal trajectory that people get a bit more richer overall life span and become not necessarily, not necessarily. Yeah.  
Yeah, I don't know what it actually is.  
Not much.  
OK.  
Yeah.  
Also, I don't know.  
I saw also for example, I would be interested to see OK how do countries differ?  
Right.  
Maybe some have poorer countries that you can distribution is just here and these just don't exist and maybe I don't know.  
Germany from me.  
Just it's just here or here.  
Yeah.  
So also we don't know exactly right. Umm.  
OK, let's let's move to other topics point here was OK, sometimes graphics it's a skill as a data scientist, you should practice basically all the time to deconflict such graphics and see OK, what data, how the data could be have been processed.  
Often it's it's quite hard work involved, and often it fulfills a certain message, and this is also interesting to decode.  
It need not be that that the data, probably the data is not really wrong, and the researchers did the best they can to estimate these numbers.  
Probably it's quite hard, but still presenting it this way or that way can also have also have a specific specific specific purpose.  
So yeah, but already just this conceptual work is quite, quite quite valuable tool to be able to do as data scientist, but it's multi criteria of.  
OK.  
OK, I'll come to this soon.  
Let's just look at this.  
Dashboard.  
Maybe I showed it already once.  
Umm yeah, this is Germany.  
OK, what are what?  
What are the visuals we're seeing that to say population permit or edge permits?  
What is it called again?  
Yeah, like this edge permit or OK bye.  
Yeah, you can see the edge distribution of men and women, I guess, right?  
And between 1930 and 2015, so and and the the last of the horizontal path are the.  
The more people are in that in that in that course.  
So it's right, yeah.  
So it's not, it's very much.  
It's very raw data so to say because it has here just the count of Germany for 400,000 people in Germany in what we have, 2018 are male.  
So, 400,000 million babies you have and 390,000 many other female babies, so to say in 2018?  
OK.  
And so we see here we see here two also two axes.  
But this is quite quite nice.  
To put it these two numbers, maybe the other small for you to see.  
But So what?  
What access do we have here, Rosenthal?  
No, no, no, this one, just zero, 200, alright.  
And this year, it's just the year, of course.  
So it's just one minus.  
It's all this 2018 minus.  
This is the edge just, but of course it's some relevant information, so we'll see here.  
So in 2018, these were the people around 40 and this is the first year around 1975.  
Umm, OK.  
And so we have it, we have it separated to males and females.  
It's good because we have some we can we can compare.  
But of course the absolute numbers are not so easy to compare because we have to add this and this, but you'll see it's almost almost a bit in parallel.  
You can see that women live longer than that.  
Right.  
And how can we see it?  
OK.  
I'll just leave it on the right side.  
Mirrors.  
Yeah.  
Yeah.  
You see, there are more in the, especially in the older ages there are more females because it's bigger there.  
And how could you see it if you just look at the manual?  
OK, looks a bit bigger, but it's not so clear.  
You can see on the image is more complex than.  
Yeah.  
And what do you think?  
How is it produced?  
What does it mean?  
The darker red and the darker blue.  
Here there is a similar minus the right.  
So basically what is happening there are mirroring the mails here and whenever the females are more they make it radish and the same here whenever the males are more they make it targetable the the number it's a it's already it's.  
I think it's a nice.  
It's a nice trick.  
It's not distracting much anything in the in the general perception of the numbers, but it gives you a clue how to compare.  
OK.  
When we have more males, when we have more females, OK, let's let's look at the look at.  
Look at the movies.  
So now we also have time and remember to time which is the most natural thing.  
Umm, OK, let's start from the 50s.  
So this is 1950.  
So the 90 year olds, here they are born in 1870 or 80% eighty years old, are born in 1870.  
So and these are this is they'll have parents are born, so they're going to go.  
So let's let's see how this then evolved.  
Umm, OK, now we see.  
OK.  
Every year you you children are born and now we are in this year.  
In 1960, we see some, we see some increase and OK, Google, right.  
Success is pretty yeah, this is in Western Europe and the US, I think it's called the baby boomers.  
I don't know if they exist in all parts of the world.  
We have come to that in a minute.  
So now I'm already 20 now coming 20 now and maybe you were born already.  
I don't know exactly and.  
And now we'll see you.  
You can check your cohort.  
If you compare it to Germans, if you are in the lower though, I'm somewhere here. Uh.  
Here comes my daughter.  
Something.  
OK, good.  
There's actually a projecting or what?  
Because we aren't as the color changed.  
If you seen this, no, I mean no, it's it's making predictions.  
Yeah, yeah, it makes predictions.  
It's right.  
And also there's a color change.  
So here the color slightly change through.  
This is because this is the future and it's a projection, and now we clearly see, OK, the projection something has changed.  
I don't know why they have this flaw here with the projection.  
Why real babies more positive for?  
I think that really have 51% chance of getting some right.  
That's.  
Yeah, let's discuss it in a minute, OK?  
Umm, I don't know if you have access to the graph.  
Maybe I can also put it up again, because now I want to have.  
I want to now how can I come back no.  
Umm.  
OK.  
God.  
The full screen.  
I don't get it, OK.  
OK, short short input.  
I don't know if you this was an interactive graphic, many interactive graphics with real time data coming in is what in business is called a data dashboard.  
It's quite commonly used quite common.  
Could also be quite common job for data scientists on both sides on the side of constructing the dashboards, but also on the side of reading dashboards and making and making making reports of it or informed decisions.  
Explaining it to others, to managers or something like this.  
So I just want to mention it here because it is something which she can relate to.  
Also, you can relate to this in your projects, but in this course the main part today is now just look more basically on certain visuals, just to point out that says very practical relevance nowadays for many businesses a data dashboards is quite a quite a common thing and the one question is when question is OK what?  
What graphics do you need and do you have some hierarchy of simple short messages?  
And then go deeper into it.  
So all these questions, OK, how we want users to perceive data and how we can structure data as there's some hierarchy to first start with some basic facts and then go deeper.  
So this is all quite relevant in all businesses today.  
Umm OK but.  
And we'll come to the to the 8 pyramids in a minute.  
Just another source of of visual communication is what I call multiply multi page data stories and I think you saw the page outward and data once.  
Have you have you mentioned it already?  
No.  
If not, you can visit it.  
You'll find very good stories there, and it's typically that they have some data and they have interactive visualizations. Here.  
It's democracy in these indices and you can select various visualizations well.  
He was in.  
That was also time changes.  
Maybe also the bubble graphs of rustling.  
I think these guys are very much in the transition of in the tradition of Hans Rosling, the positive guy from the two videos.  
So they are also.  
They just want to map the world and want to make us not overlook positive or general trends, which are often often overlooked when you just look at the negative current views.  
So classical success stories are typically literacy rate around the world or childbirth.  
So you see, it has basically in in in my lifetime it has.  
It has declined tremendously.  
Child deaths at below at age below 5 has declined tremendously the last 100 years.  
Basically almost finishes.  
OK, so this is, this is another thing about already goes into into data storytelling.  
But let's start.  
Let's go into data storytelling now for the last part, let me check saves in 50 minutes for for.  
Organization.  
So this is how this this is now data story.  
So data story goes a bit beyond just the visual of the use more visuals, and of course it also has something like a main story point or main point.  
So here this is how I solve this graphic once and on Twitter I think on the other side I had to.  
I had the link and these were the tweets.  
And this is how this was umm, this was invented in the story.  
So these are the three snapshots from the tree.  
So this was the core and then it was basically related to this current news, I think from two years ago, Leonardo DiCaprio has broken up after four years and then there was this graphic and then there was some confirmation.  
OK, this woman just turned 25, so it was just looking new data points, still confirming the theory of these 25 years upper limit.  
OK.  
But the point here was, OK, this is this is how one visual is the core of the story relating to current news.  
So to say.  
OK, I'll.  
Because another meaning what's the story?  
What's the story of the mean context?  
Context we will be.  
Yeah, that you you only have a fully functional data story.  
If you provide some context, which is when you have to pause.  
Yeah, right.  
That's not everything, just steps, but the story comes in when you provide compliant. Yeah.  
And the created pictures and then.  
So when we just rang all our data and typically we don't make much meaning out of it, you can sort it, you can present it visually that way.  
What it does still has no meaning, only here it has a meaning.  
So I like the meme a bit, but I think there's also a bit of a mismatch.  
Uh, in this meeting between the story of this meme and the data of this meme, because basically then, this meme itself, not such a great data story.  
You get what I mean.  
The mismatches happening here from here to there.  
You know, what do you think?  
What do you think?  
What I mean, yeah, yeah.  
Right.  
Or almost no yellow.  
So the kind of just don't really match, right?  
You're yellow.  
Also, the green this is just special green.  
It's not.  
It's not these these these parts cannot be composed to this thing, so in that sense I think it's not not the greatest need for this.  
The general point is it's interesting.  
Of course, the main point is when you just have data, you have basically nothing.  
Umm.  
And you cannot get any meaning out of it.  
And you can very and by just doing technical manipulations you don't get much more of the meeting, but you can get much more meaning here, but it's not the right, it's not decomposable to that.  
OK.  
But anyway umm I have last group work so.  
I think you should you.  
You should.  
You should try uh to get some meaning of this of this phenomena.  
So important here is we have different years right?  
195019762 thousand 21 the current almost current and the future projection.  
Umm so.  
I'll propose.  
I yes it is.  
That's you once again.  
That's the bottom line.  
The bottom Texas?  
Ohh OK yeah, give some hints.  
Thinking both births and and age of parents and they speak about two world wars, which is maybe initiated and also part of origin and also many people died in that wars.  
Umm, so let's say.  
A.  
Let's see you take group it, question a quickly, discuss and then make the story.  
After this question you can take B this question what is the story of these two gaps?  
1960s nineteen 76 your story is what is what is what is.  
Why is this area so, so large?  
The Dark red area.  
OK.  
Then for for you.  
What is the story of this gap for?  
For you too.  
What you do and then maybe for you these two questions? Umm.  
This is most often there's no gap.  
Just just on what this is going off on this projections of to say what the story can be, what is happening and why is it happening or just what is first step, what is happening and 2nd step maybe an idea why it's happening.  
OK, so you have some 5 minutes to discuss or let's say 10 minutes and then you just go through through the groups.  
Got the question.  
I don't know.  
It's a bit, I think that was a Spanish really nice American, right?  
Yeah.  
Additional to the war also the Spanish flu.  
Yeah, yeah, yeah.  
Yeah, right.  
It's, but it's a bit more related directly.  
Are you and the question is and unfortunately you cannot see the the the birth years so well here.  
So I think that can be important.  
Because.  
Maybe I can help.  
This is not correcting or let just let me know if you want to know.  
So I think you can see the age here, but of course it comes with a certain year.  
So, but it's not so visible.  
You can see it on the main the main on the main line on the, on the, on the, on the, on the website, and more women if if the move interesting, we've got to get.  
Umm, yeah, yeah.  
I think that was let me see.  
But yeah, want to go through this.  
Thank you.  
And young man, that's right.  
Yeah.  
So that's pretty much doesn't like, no, we're watching.  
That's.  
OK.  
Yes, they.  
19461914 years it was it wasn't, you know the gun, but and then.  
Umm. All right.  
Yeah, I don't know, to be honest.  
How?  
What was the like?  
Did you get drafted?  
Then we're like 60.  
I don't know what's the maximum age to do.  
I don't know.  
Makes this movie.  
Yeah, because with my father can just.  
Yeah.  
Isn't that he wants to go, but she is pretending.  
Yeah, he actually gets arrested.  
Here. Yeah.  
The what if?  
But we are we have to go.  
There's, I heard.  
I heard with that, so for group, for group PNC I just, I just mark you around which which birds years the gaps are.  
So for Group B, because you the gaps are around 1960 and around 1945, these two things, right, and this would be.  
And for the 2021 that's it's it's around 1975.  
And then you have.  
Looks like has occurred.  
The direction that affects so when when one generation cool lot of women cannot find a man.  
But of course, in the next generation we have fewer kids right in the what is the front runner than that?  
I think we have, that's yeah.  
My God.  
First one person.  
You know.  
Thank you.  
Take them.  
We.  
Nothing.  
OK.  
I don't.  
Choose.  
So, so your story should be mainly what do we see and and then why do we see it?  
If you have some some ideas, maybe someone.  
Thank you very much to make.  
Once again, what's talking?  
OK, right.  
I look like, yeah, looks like.  
You can make OK, yeah.  
We'll see some eight to close some years or over there.  
Yeah, you could read them.  
You have to suspect it so 1950 might have it's or in this MINIKEEVA.  
They've been found.  
Yeah, because there's.  
The actual story is very.  
Well, roughly.  
OK, then let's let's start.  
Why saying group 80?  
So what?  
What?  
What do you see and what and why do we see it?  
So then so so you could listen, please.  
No, you're stunning.  
This booklet, OK, we see the doctor red.  
Your.  
Population is women and since it's 1950, after the World War Two.  
Also, if we analyze the the range between zero and five, you see that it is shrinking.  
So with the point that most likely is, if there's a lot of widow women we can think about it, then less possibility of them finding a match and so less kids.  
From the region, right?  
In one generation, a lot of women cannot find a man because there's a surplus of women.  
Since so many men apparently felt in war, then in the next generation, we're gonna have fewer kids, right?  
So this is basically what you're seeing.  
OK, there's already projecting a bit further, but probably it's also true, but the main point is why do we have so much more so much more females here, but not anymore here?  
But why we have so much fewer females here?  
That was the first question because typically we have more boys being born and if you look at the at the currency, you have more old females.  
But typically in this age group we have either equal or even slightly more mad, but here we have just much more females also here.  
Yes, quick question is why that?  
I think you're making right points, but this is more basic question.  
Why do we have so many so few females, so much more?  
Females in this lesson, uh, all of those 8/8, right?  
But wait.  
Less men?  
Yeah, nothing worse.  
Give me go to the bar and at that time it's much less.  
Yeah.  
Yeah.  
So that's mainly that's mainly we have this the mail missing here typically.  
Typically, we'll have this time of mail they're just missing because they were all buying on the front of the off the of what war.  
So these are age groups 19/20/1925 woman.  
So when does it work?  
Start 1939 and 1945 so.  
So yeah, the worst.  
Just five years over, so the missing the missing males here are just the yeah 20 year olds and 25 year olds.  
Typical conscripts, right?  
And, you know, German war history.  
Better even.  
Yeah.  
In the end time of the war, even very young boys were sent to the frontline.  
So these are all the people dying and probably all the soldiers, but this probably consequent it's it's so we are missing, we are missing them here and then we'll see that this gap of course they get transferred upwards but.  
Umm they are then?  
Yeah.  
Then it comes to a bit closer to the normal excess females, which you'll see also in countries without without reason board histories.  
OK, but do you want some interesting question about generations?  
But maybe this is already the next to the next questions would do you want to repeat it again, or should we just go to question two?  
You mean what I said about generational effects or yeah.  
For reasons.  
Lot of women cannot find a man or vice versa.  
I guess then you can see a drop in childbirth.  
Of course, in the next generation, yeah, I guess be then here.  
Right.  
OK.  
But actually it's all of the the the the access mail, it's all over the place, so there's not too many big unless the excess emails or less males.  
It's all over the place here, right?  
It's not such such a drastic gap.  
It's just all or all, all all the ages.  
They are more females here, starting with 20.  
OK.  
But maybe it's related to to these gaps, yeah.  
Let's let's let's ask let's ask you what, what do you think for the explanation for for these guys, it's ready to this gap.  
Of course, this is this gap is, I think this gap right.  
26 years later.  
So this this guy is the same as this gap?  
Umm, what is your destination of?  
So basically these two gaps, they go to these two gaps.  
You have something, right?  
So these two guys come like come from the because the men, most of the men, has to join the army.  
And then it's like effect to the potential parents during the maybe lead to about fewer but say fewer birth rates maybe dropped.  
Yeah.  
And and you can see after the the second gap is is is slightly to be.  
A bird rate has like increased here.  
Yeah, yeah, yeah.  
That identified such a simulation shouldn't get for yeah from 7:00 and we select 5055 for the book Baby Boom generation.  
Yeah.  
Yeah, it's called the baby boom generation.  
But you're really it's really, it's really clear that, but this gaps are so because we saw here this is the best of the males basically, but they are all over the age groups.  
But here we have very, very tight gap, very, very narrow.  
That's this five 4-5 years and same here, just essentially just just just the three years or something.  
And this is exactly how you said it's because of the war times and the males.  
They were just gone at the front lines, so there were even if they had their family and the wife or whatever, then they could just not not produce produce offsprings because they were just just at the front and maybe they're also, yeah, maybe, I don't know, there were some not the most called contraception pill was not available that time.  
Of still, you could also do contraception in some ways.  
So or do not have children in more times, but also it's just the availability.  
Just just that they were not there.  
So that's why we have these these extreme gaps.  
Actually, my mother is 1944 so and I think was produced during some front holiday I guess.  
And then there was more prisoners ship and my mother was telling when my father and my grandfather can barely know, came back from the front 1955 so or 1950 or something.  
My daughter was saying no, no, my my grandfather is just a photograph on the on the board, and my father is just a photograph.  
Not a real person.  
OK, that's fine.  
Yes, you see then my my, my parents generation is one of these.  
One in this or this gap?  
This gap this is the same gap from the second from the First World War, right?  
And the first blood board was the same phenomenon.  
This doing this for four years.  
They were just not so much.  
Most males were just on the front and the females on the home front to say OK yeah the the the anti baby we just mentioned it right.  
Do you know when that was like?  
Which year the pill was invented?  
17 and coincides with this rock.  
Yeah, this this drop.  
Yeah, this is the sexual revolution right there.  
But the interesting point is what I want to hurt, but I don't have any data yet, but it just wants hurt from the inventor of the pill.  
Now these death.  
But his before he died, he had to get to talk here.  
And Bremen at once.  
And it was at that talk.  
As he said, yeah, it's the story is a bit overstated with the pill because in Japan it was only introduced 2005 and they have the same drop of births in more or less the same year without without large scale access to the build.  
But I think Japan is having problems here.  
Well, OK, just story goes further.  
Why the pillow was introduced so late in Japan, but OK, I don't remember exactly.  
Also, funny story.  
Let's go to the next gap.  
Umm, there was you? Yeah.  
That's boring.  
There were the 45 year olds that were in that gap, so it means they were born in the 1975, which is that gap which we just discussed, don't it's yeah.  
But in the. Yeah.  
Yeah. Here.  
Same one, yeah, because we saw that online that in the 70s there any localized also portions may Brewer maybe East Germany anyway, but also West Germany.  
So yeah, that is the same as in the graph.  
So it means there were no deaths, but just the low birth rate.  
Yeah, that's for sure.  
Low birth rates, but then it's like, well, yeah, it's slightly increasing here again with the now 30 late 30 or mid 30 year olds, they are a bit more, let's say that my generation actually I'm in this gap, this gap.  
So they're slightly more so.  
One thing is of course, yeah.  
Abortion, but also contraception.  
The fill is also explaining a lot of this decline, but there's another, even simpler story.  
Education, even simpler, relating to the to the story of Europe.  
So typically, what is the typical age when you get kids?  
Ohh 20 to 40.  
Let's say in Germany.  
Now it's average year is for first child is even 30, but let's say 25 to 30.  
So when you think about, OK, this guys.  
So when these guys, when my parents have kids, they're just left.  
So that's why these people 30 years later or 25 years, because they are also a bit less because they're just generational.  
This is maybe a story of the generational transition.  
We had to get here, so we should expect to get another gap 25 to 35 years later, because they're just less families in that typical age of child was around and I think we even see we even see the gap again here in a bit.  
But it's just washing up, right?  
So yeah, so this is, yeah, this more generation is even another gap here.  
Let's pronounce and even another gap here.  
Let's pronounce here.  
Again, we have a small increase.  
I think it's still still nothing sent propagating from this, but watching out because there's no exact exact here when you get a child.  
So of course there's a lot of rain, 18 to 45 or something.  
OK.  
Final group most open.  
What?  
What can you tell us about D&E?  
There's no virus or.  
2017.  
Yeah, like to see it's kind of you don't see it first.  
I'm looking for for FX, so everything was the school.  
Yeah, actually I'm not really sure, but they should be some corona effect.  
Some what?  
Maybe it's not large enough, or maybe this is it's not, but it's kind of really be.  
Yeah, I have this small off here and there is the entire pandemic.  
I'm gonna maybe, but they took no be is it should not be that year or during the pandemic that he will want to burst.  
Or no, I think it doesn't affect the children.  
So yeah, but this is what?  
What year?  
50 These are the ones.  
Yeah, but it's not the 1st 2020, you know, it's like it gets maybe they were less first in that.  
So we should be the should be the people born during the pandemic 2020.  
And this should be me and I don't know if I'm still alive.  
OK.  
Yeah, but what we what the projection is that there is a steady decline and some say the pyramid is standing on top.  
It's not really.  
Prioritize more portal.  
You said, right? Yeah.  
Skills.  
Uh, But what diction?  
Yeah.  
Yeah.  
And generally it's it's a phenomenon which is, yeah, worldwide phenomenon happening happened in almost all countries during the development.  
There is this demographic transition.  
First, first child child death decline, then worse than population explodes and then it celebrates somehow, because for trilogy goes down.  
Some of this is very, very solid.  
In Germany 19-8 around or you saw it in.  
That's in America in 50 Sixties, 70s, I think still now most of the nation American countries have stabilized in population.  
Same China in India is on the way I think now still growing or not, but you see already the part of it is not like this anymore.  
Like many African countries still, but in India is like this now.  
Like like this and and and Algeria still like this.  
So seems to be, but actually I once talked to an anthropologist and it seems to be that for this generic pattern there's still no real super hard scientific explanation.  
But it's such a generic fact happening all over the place, there should be some general rule behind it, but there are also some competing, competing stories about it.  
OK.  
Umm good.  
Very well done.  
So I think this is this is a bit of a way of how data storytelling works.  
So we have some real data and then we relate it to some historical facts and explain the story around what what is happening and why and what what we what we really see.  
OK, let me see what the time is.  
So yeah, we have 50 minutes left.  
Let's come to what the whole term should be about so.  
OK, data visualization.  
In.  
Got.  
We already used it a lot.  
I think you have already some proficiencies.  
I hope you are interested to develop it further and towards not only this, this is where we have used it a lot.  
It's very basic skill to analyze new data, but it's also essential here in the communication and in the reporting.  
And yeah, and storytelling, so and this is the focus of this course.  
So now it's much more not just getting a graphic out and interpreting it yourself, but getting a graphic out and then thinking about how can I change it?  
How can I improve it for the purpose of communicating some messages so and as you then may guess, it goes back to the data transformation, it goes back to the data origin, so you have to think about OK what is this data about?  
And maybe you should I transform it somehow to make a certain point, but when I transform it, how do we explain it?  
For example, when I do logarithmic transformation, is it really necessary or not?  
And if if I think it's necessary, how do I show it best?  
Should I show the locked numbers?  
Probably not.  
I should show the actual incomes, but then on the these kind of questions are now another of this is technical and it's it's it's your chance then to increase your skills in that.  
And we'll we'll look at some graphics next week.  
How basic improvements can be made?  
Maybe I'll look at your data, data science lab reports.  
Maybe I found some examples we have that are improvements possible.  
Let me see where my own examples, but yeah, the point is the point is another thing is technical, but it all goes down to what you really want.  
So therefore the the first step is to develop a sense of what you want to visualize, and therefore also in the next two to to weeks we want to look at more examples and also I hope that you bring you bring some.  
I'll send a message in the next days.  
Around what?  
What I think you you can bring and then we'll have similar discussions as this on some of your examples.  
Maybe you'll find some interesting meme, or you'll find in our world and data and short visual story which you say.  
Yeah, I think this is interesting and this can also lead already to where you want to make a data story about as a project in the future.  
But maybe I was a bit too fast.  
Umm, OK.  
I'll think I'll.  
I'll show you also a bit of the G plotting again, but as most of you were working with Python, I think it's very transferable or there's even a ggplot interface for Python.  
Or you can use it with the mud plot lip and I can show you a good example of the most common really that visualization I saw with my lip, but of course it was Mark.  
Probably you can do great graphics also, it's just it's just a bit of our technology and the way of thinking in the grammar of graphics style mapping, mapping variables to aesthetics and transforming them before or scaling them.  
This is all the same way of speaking, so to say.  
Uh, and also some graphics are not like really made with ggplot, so there's not that you're bound due to this, but for basic thinking it's makes a lot of sense.  
Also, you can think of course, in principle you can exchange all of this, right?  
You can also have continuous color map.  
You can put lifespan on a color you could play with that and see how it looks and if it makes sense.  
If it gives, conveys some other message, or if it's just confusing.  
So that's what we'll do in this in this term.  
Umm, because also a lot about this is some conventions which you can break or use.  
OK.  
Umm ah, another interesting point.  
This, which is new to courses you have taken with me. I'm.  
Ohh you have this book?  
Yeah.  
Yeah. Good.  
I provide the PDF of this book in the course organization repository where you don't have access to now, but I provided PDF of this book.  
There I found it on the Internet and I didn't have a clear indication that it was illegal to get it, so I just provided it there and so I'm not absolutely sure you can also, I don't know if they have it in the library.  
So and the idea is that you read that book during the course chapter by chapter.  
Our indicate in the syllabus which chapter 3, when when it's not super strict.  
I also provide some guidelines in the syllabus, some where I think close reading is more important than in other chapters.  
For example, there are two chapters where I say OK mainly optional or you can skim through it when you're interested.  
Of course, I think I read everything of the book at least once, at least Blues and many clothes.  
So, and I think all stuff is generally generally interesting, it's also it also may be valuable because it's written a bit from a more business perspective.  
So as you know, I'm not so much a business person, but I think from that perspective the, the, the book is also valuable, but it's also valuable in general for data storytelling.  
And if you want to make your project a bit more businesslike, it's fine.  
If you want to make your project on some societal phenomenon societal aspect comparing many countries similar to rustling on other dimensions or whatever it is, all also fine.  
Yeah.  
So this is there's a lot of a lot of flexibility there.  
You can start thinking about it.  
What you want to do?  
Umm and yeah so so the the the, the the book will work for for both but maybe if you want to also have some some input on data visualization of data storytelling for business.  
I think also good resource there's these other these other.  
These are two of the prominent books and also have this also good stuff in it.  
Also quite practical to learn things.  
Maybe I'll have some examples on this, but I'll select it then this tool to go through this.  
OK, there is this website.  
There's so many visuals around I think you'll have some time in the next weeks to look at it, so this is this is starting from the technicalities.  
So bar charts, density plots, box plots, whatever.  
So you cannot really fully bubble bubble plots like the rustling there are so many different types of visuals for different purposes.  
It's good to look a bit around and get inspired and then try some.  
So I think whenever you have some data and look for some ideas, it could make sense to just try something, but my experience is even if you are quite advanced in the technicality in the technical way of doing it, it's still often I have an idea and then I try it and then it doesn't look so good as like how I start or it doesn't work.  
So what I thought I could put on another aesthetic.  
I put this on size, but then it just it just doesn't doesn't work as I thought.  
So there's a lot of experimenting with it.  
And also communicating it.  
So I think this is a lot of the trial and error and trying new things.  
So and I would be happy if you if you go into this and try yourself and try to discuss it in class.  
So I would be very happy to provide it.  
Yeah, they also have course snippets.  
I think both Python And right.  
Yeah.  
So I think then with that way you can quickly at least try a bit, but it's also just inspirational.  
So what could do, right?  
Not just the technical what?  
How it could look?  
How much time do we have?  
Still some minutes.  
OK, this is a bit something to.  
This is soap is a showcase of uh.  
The final step of data processing here also very much from the business world.  
This is from this other book.  
Storytelling with data and this is typical challenge that they said.  
If you have this this visualization you have in green, the direct sales with all the numbers we have in red, the indirect sales and not really sure what it is actually.  
But, umm, different types of sales of company probabilities.  
And here we have time running for a year and 9:00 to close. Deal.  
OK, umm and the and the and the Y axis along it takes to close the deal and the goal is 90 days and we'll see.  
OK, sometimes we have more than 90 days and most times we have less than 90 days.  
But there's a difference in direct sales and indirect sales.  
So in that way we could decompose decompose the message, but if it is in a business presentation, there may be a better way to to show this to show this information which is just waiting summarized and the two, let's say primitives to do it in a sense, or to activities to do is eliminate clutter.  
So remove what is not so important.  
Umm.  
And focus attention.  
So focus something like like like this.  
So the question is a bit OK. What?  
What clutter could we eliminate here?  
Get some suggestions.  
Yeah, yeah.  
The days and what is it called?  
Days don't have decimal science and so we can remove the .0 on the Y axis from each one.  
Is that that here?  
Yeah.  
Plus, .0 can go away.  
Yeah, also.  
But then when the entire year 2019, so we can remove all of the nineteens on the on the look, I'm just as in sitting all. Yeah.  
Caption I would remove the text the numbers on top of the bus as well because we have this one right.  
We have this also.  
There are those lines in the.  
The Gridlines can also go away I think, yeah, but OK.  
When we do and you don't have this number, then we don't know yet.  
Then maybe, then maybe keep them.  
Yeah, change the many once again, yeah.  
Opacities in what way?  
Uh, on the good times?  
Ohh makes 3 \* A bit a bit lighter so that they don't.  
They're not.  
Restart blue, but maybe light blue or Gray or something.  
Yeah, that's a good point.  
Because then they don't distract, but they still give the opportunity to check more clearly where we are, because this is sometimes important.  
Yeah.  
And so we always want to close the deal in a maximum of 90 days or what?  
Yeah, that's right.  
That we can highlight those that maybe exceeding 90 days, right?  
Sure.  
Yeah.  
And actually the anymore suggestions or ideas?  
Just there's no right or wrong in here.  
Is just.  
It's just ideas and then you can try it.  
I think the the, the, the, the, the the the better solution offered here also has a different type of visualization, not the bar chart design lines, right?  
Yeah.  
So because I think I've also found it often not so easy to compare this to this.  
So actually we want to see the trend in green and then red or compare, but also so then this was the proposal.  
Umm, so it has put both on the line, has highlighted the indirect because they are.  
The problem is and highlights every data point which is above the dot dashed goal line something and then we have a reference line for the direct sales.  
We don't have the problem, so we can have an idea.  
OK.  
Maybe we can, we can compare it or we can learn from the direct sales.  
While it's faster there and the indirect sales we have it and then we can think, OK, is there something particular in June happening actually I don't know any I know any idea where the business case comes from, but here we can also do it in a bit more, a bit more technical.  
So the same so this times of steps come come later.  
But this is also something to keep in mind, so I would never do this.  
Do this type of things when I don't have a clear story point, but of course it's in the end it's still a lot of work to go to this to go to this stage, OK.  
5 minutes.  
So what's happening next is I hope I'll get around you some.  
Some message about Git organization, the invitation and then also what I want you to prepare for next week.  
Mainly it's searching a few examples and then maybe for next next next week there is something playing with some get minder data set with these bubbles of countries and try a few things or this would be probably the for the next week the exercise and it should all go a bit towards thinking about what what your project is about just.  
OK, just let me show you the syllabus once.  
If the schedule once syllabus, you should of course read.  
So you see, if you have, I have some you three classes and then should already be something on individual meetings about project topics.  
Again you you can form teams.  
You can also work alone.  
I leave this, I leave this open just a bit.  
It is when you all work alone, we have little bit less time to talk about each project than in class, but it's feasible so I can so it's just.  
I would then make the schedule, but if you if you are happy to form teams, I'm totally supported.  
If you really want to go alone, I'm also support that so this is also here.  
It's absolutely fine.  
Yeah, yeah.  
Before it was a bit more forced to have some some teams, but here it's it's more free.  
Umm, OK then we have some project pitches.  
Maybe they will also come some other point.  
It depends a bit how it develops, but the idea is that this this meetings where individual meetings, they stay at our schedule meetings with you during first time with the groups, then we have many individual protect, not so much.  
When you have more groups a bit more, umm, but that this should stay, but every there's class there is some either presentation where you should also be there to watch your your your your fellows and comment on that and discuss with them or I have some classroom activities like drafting some visuals.  
Ideally in the data you already have by then and some storyboarding activity, so these kind of these kind of activities and then in the end we have some presentations and just that you have an idea you can already now consult.  
Of course, I will quickly go through it.  
So in the syllabus you'll find here some milestones from zero to six.  
OK, some addition problem.  
I'll solve it.  
And so there is 6 milestones and you'll find them on this other website on this other website.  
There's a short description of it, so they are all related to some classroom activities and it's it should be a bit like OK, when I construct a project here, which should be very much communication oriented, we fairly fast should try to fix the data set.  
So this is really important to to try to fix data set and topic white the quite fast and we'll have a project pitch and then we'll have here these things like, OK, isolate what should be the main insight.  
Then there are many options of it.  
Sometimes, depending on what what data you have.  
So, but because the core of the project should not be necessarily the data analysis, of course there's a bit of this, of course, involved in every project and probably also interested in some questions.  
So there is a bit of exploration, but what you will find out need not be super deep.  
It can be some, some, some, some more or less known fact or or even well known fact.  
But you will just visualize it then or make a story about this to communicate this to some audience.  
You're also discuss about, so there are then some milestones like fixing main inside, making your first visuals, then make the data story with the kind of storyboard which can be Rave into your slide deck and so on.  
So you can consult it a bit, but we'll go into this also during during the course next time.  
So, but it's good to consult this once in a while.  
Often, sometimes you can be faster, but often you will also lag behind in some milestones is not a big problem, but it gives you some.  
It gives you some guidance.  
OK, we have to speed up or we have to.  
Where to?  
Where to be and what and what part in the in the term?  
OK, then I hope I'll meet you next week and also the other students, if you will see some motivate them to come.  
What?  
OK, whatever.  
OK.  
See you next time.  
Are you going to getting?  
What should?  
Yeah, some of us also, we have no problem.  
You messed up.  
Ah, yeah.  
Yeah.  
Yeah, yeah, yeah.  
I'm writing hundreds only for time time that this gauge, the diamond job description.  
It was approved as being.  
That's right, Falk, by another off, I used to say, yeah.  
Different.  
Yeah.  
My cooking?  
Yes, if you're going to, it's an infrastructure.  
Is this?  
You're not actually, no.  
I need money to see.  
That's the ones finding online.  
Didn't always. Kind of strange.  
So OK, but it's so they're working.  
That's good.  
Yeah, right.  
But it is for focuses their fund, which is chairman.  
I knew I had to to write a the the other, the other form that you mentioned at the beginning of this application for changes function.  
Yeah, because yeah, this part as my elective.  
Yeah, I remember that.  
But typically I signed this kind of phone.  
It's another page for me to sign.  
No, no, there's only one page and they say that's the signature, which is.  
That should be someone different viewing, right?  
Yeah.  
OK.  
What about nothing if they can see.  
Yeah.  
OK.  
Maybe then it's fine.  
So for me it's fine.  
As I said, the only thing is a bit this probably GitHub, you know.  
Yeah.  
Yeah, I think that the only thing now is like waiting the thing Professor Ketterman to side.  
Yeah, you can add me to three team like, yeah, cannot manually.  
Yeah, they can.  
They can get, bring it to the team now.  
Umm.