Hello, I'm Mori Jamshidian. I'm a faculty member at California state university Fullerton. I will present to you today why use software and what software to use in our introductory stats courses

so perhaps many of us agree that the two primary goals for our students in our introductory courses are to teach statistical literacy and develop our students as statistical investigation skills

well statistical literacy involves learning how to critically assess information, for example, the many statistics that are thrown at us from the media being the television, social media, and so on and then, of course, it in turn this requires that we teach a statistical concept so these days there is much emphasis on teaching statistics conceptually, and I refer you to gays which is an invaluable document that is published by the American associate school association while achieving both of these goals requires use of software here I will focus on a statistical investigation skill generally we should abandon the traditional ways of teaching stats, for example, these include plugging in numbers into formulas or constructing graphs by hand teaching how to use probability tables such as z table t table f chi-squared, etc. or using calculators to do statistics I'm not against using calculators to do arithmetic, but calculators are not made to do a statistical analysis with

So let me go back to the statistical investigation skills that I want to focus on. What is this entails

so there are several steps here identify the question that we want to answer, then we collect evidence or data, then we explore our data

then we analyze and model our data and, of course, make conclusions and discoveries and communicate our discoveries now, especially the parts on exploring data and analyzing and modeling data really requires the use of the statistical software

so with that i thought it would be useful for me to give you an example of what we do in our introductory stats courses at cal state fullerton we teach our courses based on investigations in other words for each concept we have an investigation that students go through and i will present one example to you one of the investigations that we go through is called used car investigation and the question is predicting a car's price according to some other variables so here the whole thing starts with collecting data of course before we start collecting data we discuss on how we are going to collect data whether the results that at the end we get can be generalized to a larger population etc and students collect data in groups and they will choose a car or a type of car we'll just leave that up to them to collect data on they have to collect prices for the cars that they're for that they are choosing and other related variables and again the related variables is up to them it could be like mileage it could be uh the year of the car etc now each member of the group collects a certain number of data points and they input it in the same google sheet and once that they are finished with the data collection then we upload the

data into software and of course the software i'm going to show you in a minute is called Rguroo that's the one that we use at cal state florida then we explore the data and we go ahead and model the car prices as a function of the variables that we have collected and we make conclusions so let me show you an example of a data set that i was collected recently in my stats course these data were collected in november of 2021

and the students collected data on honda's they recorded the model of the car the year of the car the mileage on each of these cars the color of the car the price and the source where they got their their information so for example truecar conv carvana etc now let's go through this investigation so let me go ahead and open up the software the software is called Rguroo that we use and you can get information about it at Rguroo.com so let me just go ahead and sign in the nice thing about this software is that you don't need to download anything or install anything all you need is a browser and internet in order to use it so i'm going to log in of course you need to register for an account and when you log in you see data plots analytics there are different portions but let me just go ahead and talk about data here well we need to import data first so i'm going to import my data set it happens that i have the data sets on my desktop so i'm going to just go ahead and get the used card data and i'm going to say open an upload and when i upload this data it automatically gives me the numerical versus categorical variables and the numerical variables in the data set remember was year mile mileage and price and these are the categorical variables in the data set now if i double click on the data set i can see the data set itself and you can see that that is the same data set as that we looked at before okay well so now we begin our exploration we want to look at price for example as a function of mileage or year so the point of a start is to use a to look at a scatter plot so i'm going to go ahead and go to plots create plot and i'm going to just say scatter plot one thing that i want to point out before i go there is that in order to do that students don't really need to have a lot of instructions because they see plots they click on it and then you know there is create plot they click on and then they get scatter plot right now a menu shows up all of the menus in this particular software have basics and details and details is for customizing graphs and so on but for the most part students only need to use the basics menu so here we begin by selecting our data set which in our case is the used car data all right and our response variable is price and we could choose a predictor variable and let's say let's choose year as the predictor variable i'm going to go ahead and close this window here so i have more room and you can see that you have the newer the car the higher the price that is not surprising at all but then we just go ahead and dig in a little bit more the next step is for example to look at outliers

fortunately this software can actually identify outliers for us so i'm going to go ahead and click outliers here and if i just say preview here you can see that cases number 1 7 and 24 are identified as outliers the software is using a method called mahalanobis distance this tends to to identify outliers this is not something that the students need to know also you can ask the software to identify outliers based on other criteria or determine more outliers or less number of outliers and those are all in the details but for the most part that should be sufficient what is interesting here is that you can see that we have a car here that is old but has a high price but on the other hand a newer car with a low price and then we need to investigate why well so with that we can id variable so maybe it has to do with the model of the car then the students

can investigate and when they look well these are all civic lx this one is civic lx this one is the civic lx this one is civic lx sedan which is a little bit different so that doesn't seem to explain the situation the next step is to look at some other variables so for example let's look at the mileage of the car the cars and if i do that we see that the mileage for this car is 207 610 and that explains why the price is so low on the other hand for this older car you can see that it has a very low mileage 38000 miles and that explains why the price of this car is high so now we we discovered something here but can we discover more let's let's for example investigate whether there's any difference in terms of pricing by by source so here now you can see that there's carmax carbona honda dealers and truecar here and overall you know all the data is kind of pretty mixed up here in this part of the the um the graph and it's not quite clear what's going on but what about if we just summarize the data by using these squares by each factor and if we do that all of a sudden it seems like there is a pattern well what is the pattern it seems like the car prices for carmax are more expensive than for example for truecar now we went through an investigation and just by looking at data we discovered a number of things right in particular we found out that carmax prices are more expensive of course this is only based on these data and whether it is generalizable we have to look at that more i just wanted to show you that if you were to do this on a calculator here well you know calculators are very difficult to use in my opinion so let's say you want to do a scatter plot here is a set of instructions that i got on the internet go to second stat plot make make sure that only plot one is on right right here then um you know then go to y1 and clear any functions so here comes the most challenging part inputting data in general inputting data is going to be difficult here and at the end what you are going to get really is a graph that is not going to be a very descriptive graph so

Why not calculators hopefully, by now, it has become clear to you as to why I think we should not be using calculators. It is difficult to input real data, not intuitive to use poor data visualization, and there's, of course, no interactions, inability to export graphs and reports, and of course, they are expensive, so hopefully, the example that I just showed you indicates how useful a software can be in teaching our students in our introductory stats courses, but here I list a few of the specific advantages of software easy to input real data and work with real data intuitive graphical user interface

Powerful data visualizations, and it obviously allows interactions. As I showed you, expensive and easy to read output reports, you can export graphs and reports into, for example, word document or PowerPoint document or any other application that you you need to present your report with and most importantly

the ability to use a comprehensive set of tools to carry out an investigation is a big plus in using software data visualization is only one part of a statistic but might want to continue with, for example doing inference and so on why we chose our group in our department well it adheres to all gays technology guidelines it is cloud-based students don't need to download or install anything the software is specifically designed for teaching statistics there is a reproducibility aspect to it where you can save everything on the cloud and come back to it, and of course, it's reasonably priced

now, how does our group compare to other statistics software I will talk about this in my next video I will also explain a little bit more about how we chose Rguroo and why we chose Rguroo in the next video as well thank you for watching, and I hope that this was useful