Transcript: Live Intro to Exploratory Data Analysis

Exploratory data analysis, you might remember this being called descriptive statistics in earlier classes that you've taken, either way is fine.

It consists of organizing and summarizing our data; trying to get our raw data to be something that's meaningful and we can glean some information from it

And then we want to look at what are the important features or patterns in the data so I'm comparing males and females do I see a difference in males and females at this stage

And then we're going to interpret the findings in context these are pretty simple findings at this point meaning we're not going to make any broad conclusions about the population at this stage we're just going tosay this is what our data says males tended to in our data have a higher whatever than females by x points.

And the important features of unit 1 we're going to look at one variable a time - examining distributions and then we're going to look at two variables at a time - examining relationships

We will almost always show you both the numerical measures that would be associated with that analysis - means medians frequencies percentages that kinda stuff

And the visual display that goes with it pie chart bar chart histogram box plot scatterplot all that good stuff

We want to have a picture -which to me is the most exciting part when I have a hypothesis, and I don't really have to many hypotheses actually in real life, but if I did have hypotheses in real life, I would be excited when I look at the picture not when I read the means - the means are just there to give me some numbers to go with my picture but the picture is showing me

In the simple world which we're in this semester - as soon as you get to regression and you have four variables well there's no visual for that anymore 3, I can do a visual still I can get a three-dimensional scatterplot but for 4, I'm done I can't look at it anymore

Then I have to trust the ideas that I learned in one, two, and maybe three variables to help me believe the answers I get for five variables or 10 variables or whatever.

I like simple and I think it's always good to go back and think about simple as you start to analyze data.