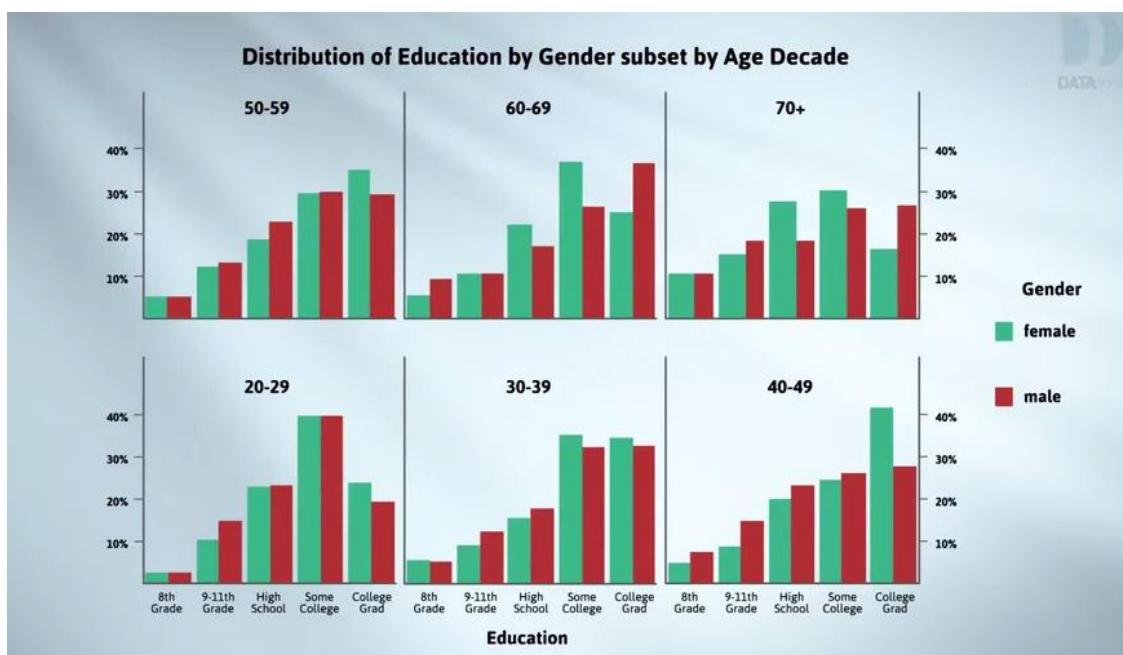


WEEK 3

3.4 CHANGES ACROSS SUBGROUPS by Chris Wild

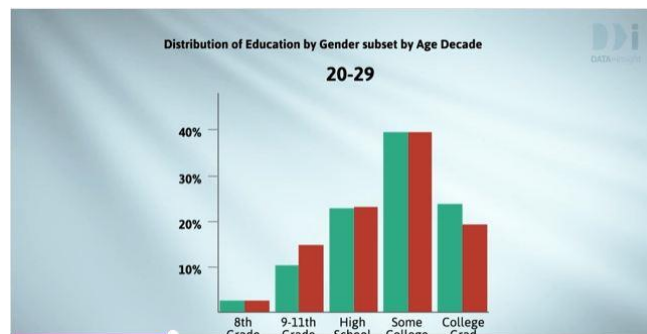
In the last video, we saw how American educational outcomes have changed with age. With things that generally happen early in life like formal education, looking at older age groups is like looking back in time to the way things were for previous generations.

So what we saw about educational outcomes changing with age is, to a large extent, a reflection of changes over time. But what about gender differences in educational outcomes? Have the gender differences in education also changed over time?

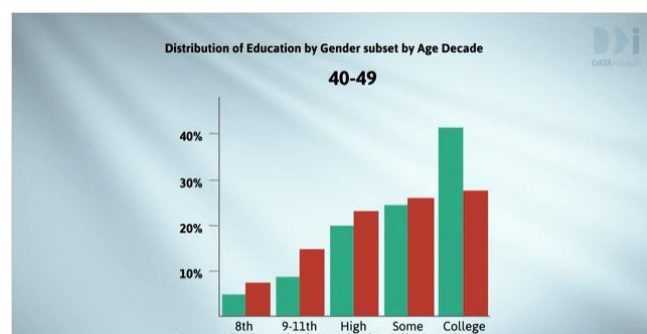
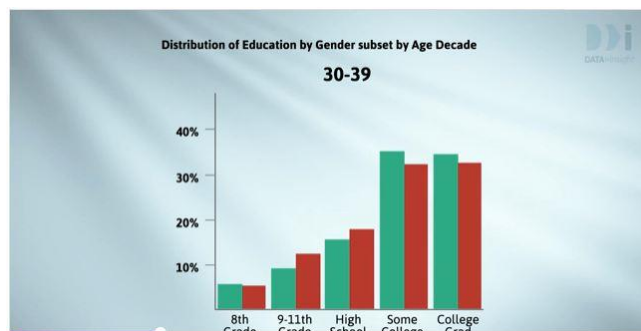
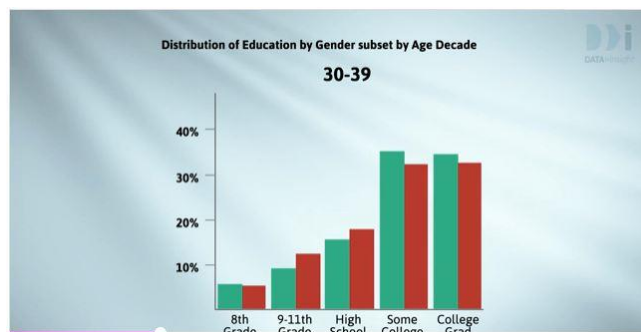


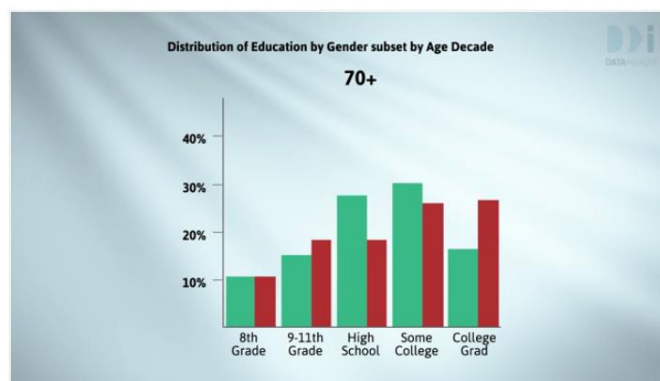
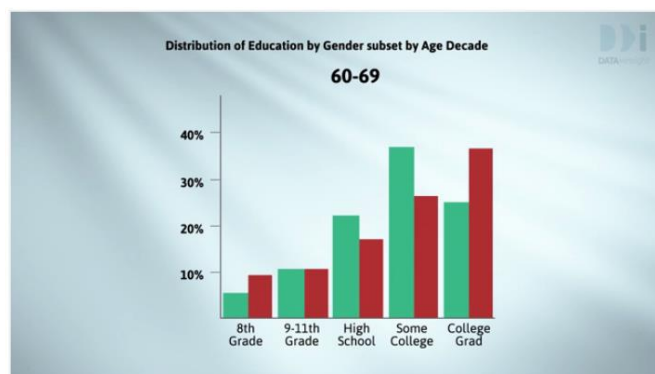
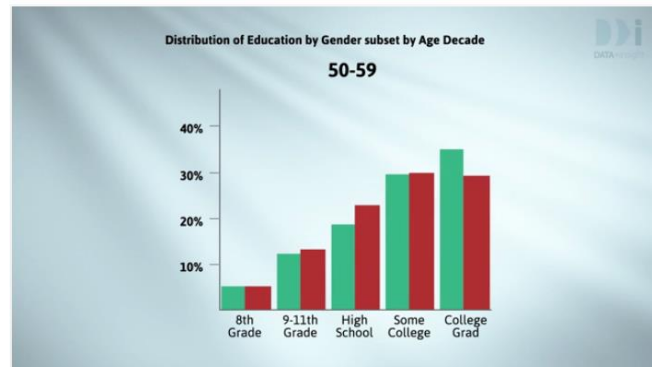
Here are separate graphs of the education by gender relationship for each decade of age. We won't dive straight into this.

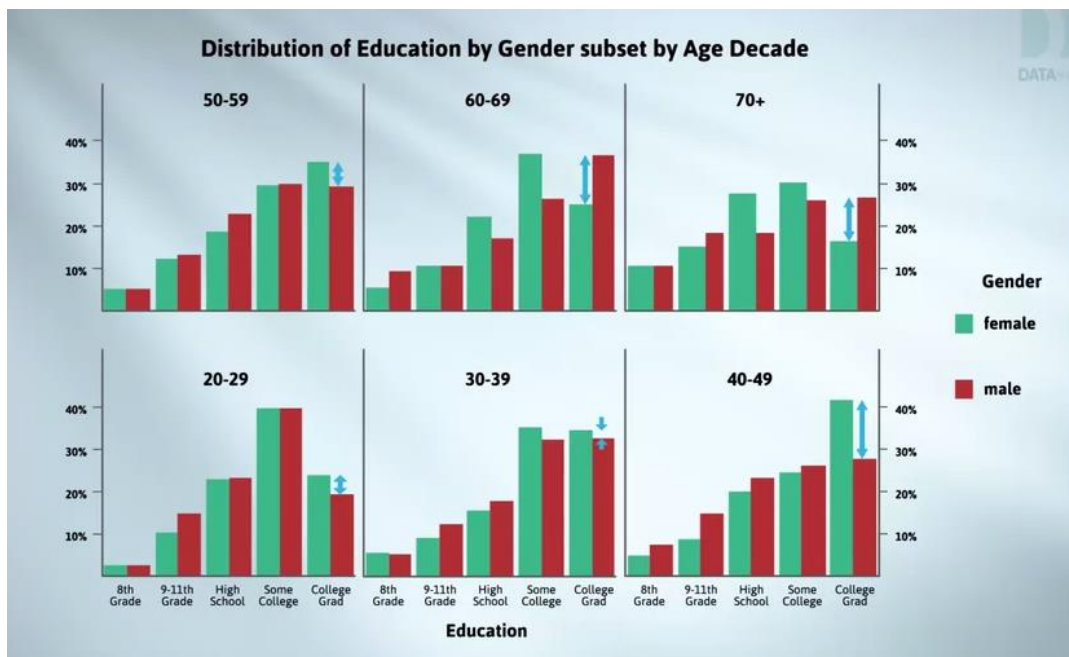
Changes often jump out at you much more readily when you step through the set of graphs than they do in a tiled set of plots. iNZight does this for us when we use the slider or play button.



Here's what it looks like with this data.... Again....



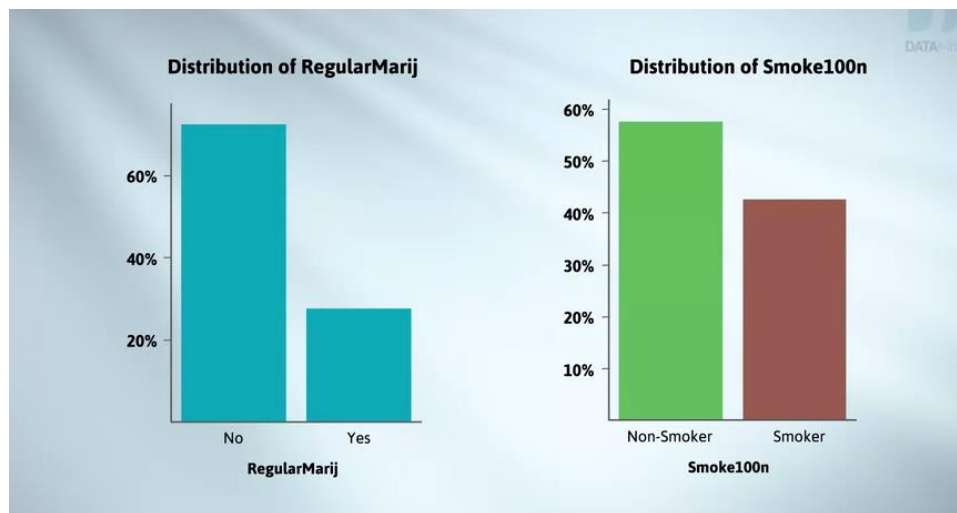




Whatever way you see it, the most interesting changes have to do with gender differences in the CollegeGrad category. There are proportionately more female than male college graduates for each of the age groups, dramatically so for the 40-49s. But at 60 that switches and, from then on, substantially more men graduated. At the same time, particularly for the 60- 69 group, an unusually large percentage of women had some college but didn't graduate. I'd guess that's largely due to women in those generations not completing after they got married in college.

To cement what we've been learning, let's look at a group of variables relating to things that we might loosely think of as risky behaviour-- smoking, drinking alcohol, using marijuana or hard drugs, and early sex. We'll find that these things tend to go together and that people who've tried one are more likely to try another. We'll only look at people aged between 20 and 59, because several of these variables have only been recorded for these age groups.

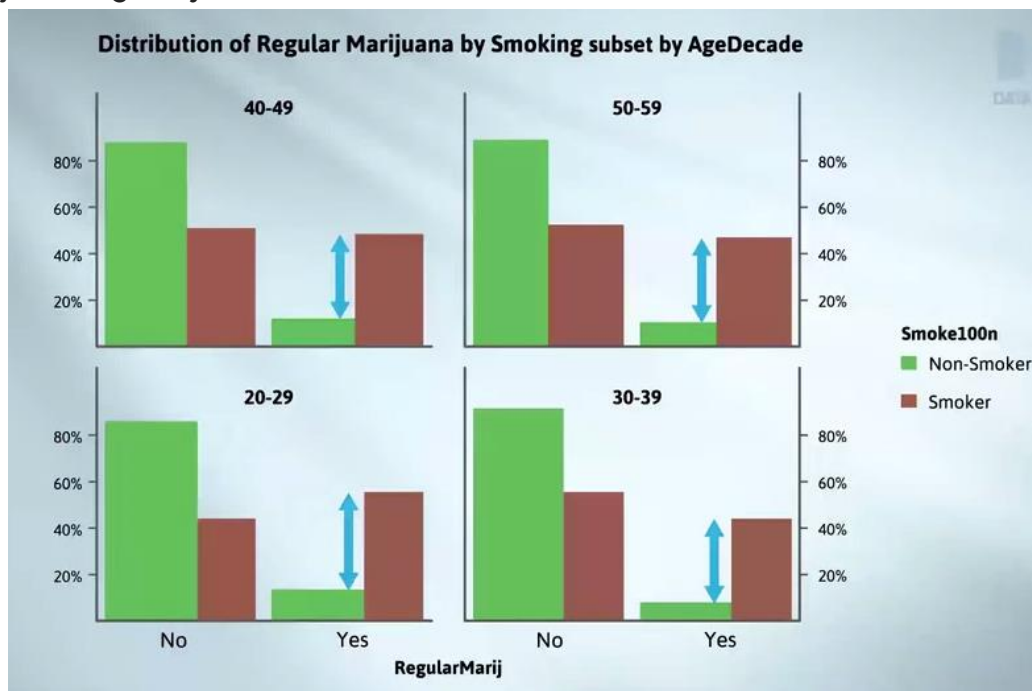
Let's start with regular marijuana use as our outcome variable and cigarette smoking as the predictor variable. For regular marijuana, participants were asked whether they had at some point used marijuana at least once a month for a year. Approximately 25% of respondents said "Yes" to the regular marijuana question, while about 42% said they'd smoked more than 100 cigarettes in their lifetime. We'll refer to those who said "Yes" to the smoked 100 question as smokers and the "Nos" as non-smokers.



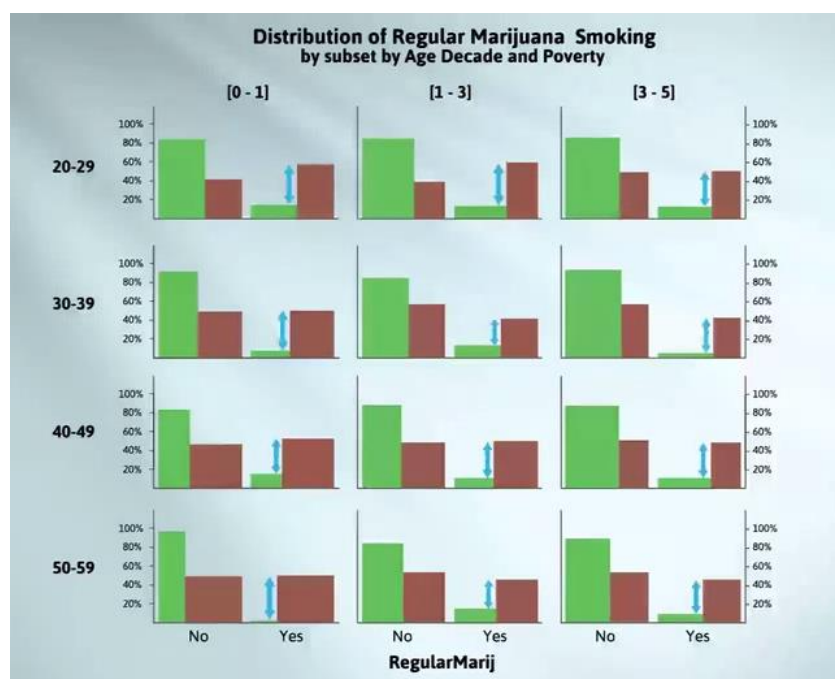
Here, we've coloured the smoking plot by smoke100n. From now on, smokers will be brown and non-smokers green. We're now looking at the relationship between regular marijuana (as the outcome variable) and smoke100 (the predictor variable) as separate graphs **and** as side-by-side bar charts.

It's very clear that the cigarette smokers are much more likely to say "Yes" to the marijuana question than the non-smokers (more than 40% for smokers versus less than 10% for non-smokers). The side-by-side bars (the right hand plot) make it much easier to gauge visually how much more likely the smokers (brown) are to say "Yes" to the marijuana question than the non smokers (green).

The brown "Yes" bar looks more than four times as high as the green "Yes" bar. So the cigarette smokers are more than four times as likely to say "Yes" to having used marijuana regularly than the non smokers.



Our next question is, "Does this pattern change with age?" The pattern of smokers being much more likely to say "Yes" to the regular marijuana question is remarkably consistent across all of the age groups. The brown smokers "Yes" bars are always much taller than the green non-smokers "Yes" bars.



Our next question is, "Does the pattern change over socioeconomic groups. We'll use the variable, poverty, to explore this. Smaller values of the poverty variable correspond to increased poverty.

The age groups are the rows. The poverty groups are the three columns. People get poorer as you move to the left.

It is easy to see that the pattern of smokers being much more likely to say "Yes" to regular marijuana than non-smokers is remarkably consistent across both age (down rows) and economic circumstances (across columns).

Finally, I'll leave you with these questions to remind you of the ideas we've just covered.

QUESTIONS

What basic ideas allow us to explore how a relationship changes with a 3rd (and perhaps a 4th) variable?

Why is stepping through a set of graphs (like playing a movie) useful?