Share


# Looking at Associations with Multivariate Categorical Data

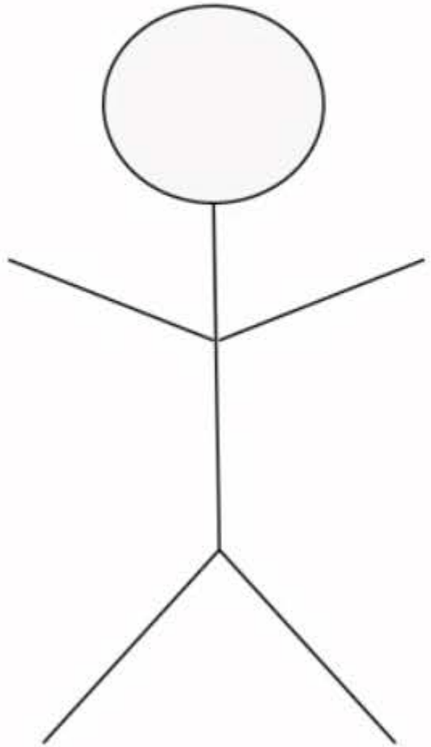
Julie Deeke

Statistics with Python Course Developer



Waiting for cache...

# Gathering Multivariate Categorical Data



What is your **gender**?

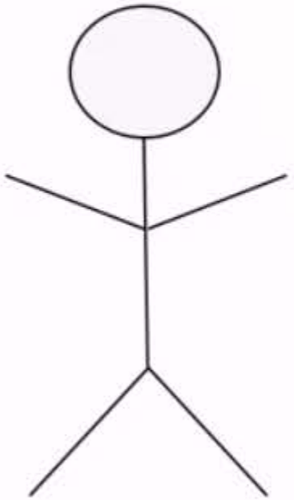
What is your **marital status**?

What is your highest **education level**?

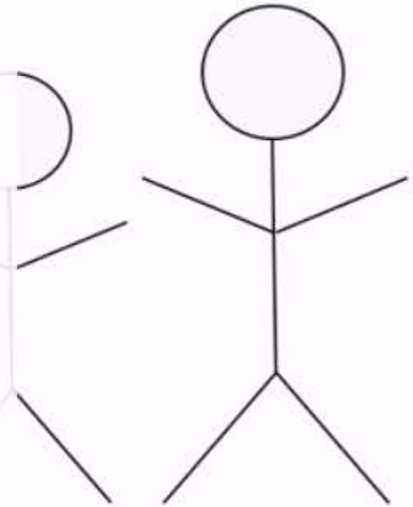
What is your **age grouping**?

We can think about looking at the responses to

# Recording Multivariate Categorical Data



| Gender | Education | Partner_Status | Age |
|---|---|---|---|
| Male | HS grad or GED | Never | 18to29 |
| Female | Some college or AA | Current | 30to44 |
| Male | HS grad or GED | Never | 18to29 |
| Female | HS grad or GED | Never | 30to44 |
| Male | College grad | Current | 60plus |
| Female | College grad | Current | 30to44 |
| Male | HS grad or GED | Current | 45to59 |
| Male | HS grad or GED | Past | 60plus |
| Male | College grad | Current | 45to59 |
| Male | College grad | Current | 30to44 |
| Male | College grad | Current | 60plus |
| Male | HS grad or GED | Never | 18to29 |
| Female | Some college or AA | Current | 30to44 |
| Male | HS grad or GED | Past | 60plus |
| Male | Some college or AA | Current | 30to44 |



we could choose to record our variables in the form of tables.

# Univariate Categorical Data Table

| | | |
|---|---|---|
| **Highest Education Level Attained** | Less than HS degree | 1331 |
| | HS degree or GED | 1166 |
| | Some college or AA | 1656 |
| | College degree or above | 1396 |

**Research Question:**
What factors influence the highest education level attained?

So, we'll split up

# Two-way Table of Education and Gender

| | | Female | Male | **Total** |
|---|---|---|---|---|
| **Highest Education Level Attained** | Less than HS degree | 644 (11.6%) | 687 (12.4%) | 1331 |
| | HS degree or GED | 552 ( 9.9%) | 614 (11.1%) | 1166 |
| | Some college or AA | 894 (16.1%) | 762 (13.7%) | 1656 |
| | College degree or above | 724 (13.0%) | 672 (12.1%) | 1396 |
| | **Total** | 2814 | 2735 | 5549 (100%) |

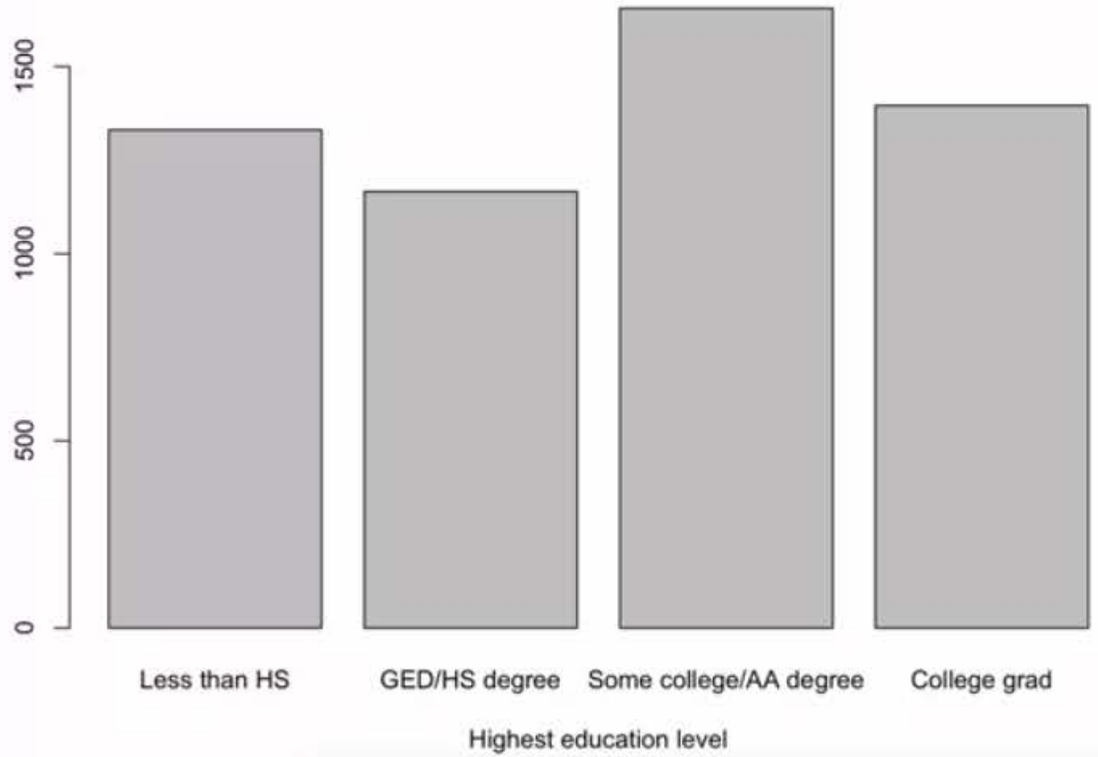
Let's return back to that original table that I showed you a few slides back.

# Marginal Distribution of Education Level

| | | Female | Male | **Total** |
|---|---|---|---|---|
| **Highest Education Level Attained** | Less than HS degree | 644 | 687 | 1331 (24.0%) |
| | HS degree or GED | 552 | 614 | 1166 (21.0%) |
| | Some college or AA | 894 | 762 | 1656 (29.8%) |
| | College degree or above | 724 | 672 | 1396 (25.2%) |
| | **Total** | 2814 | 2735 | 5549 (100%) |

our sample was distributed across our four categories.

# Univariate Bar Chart



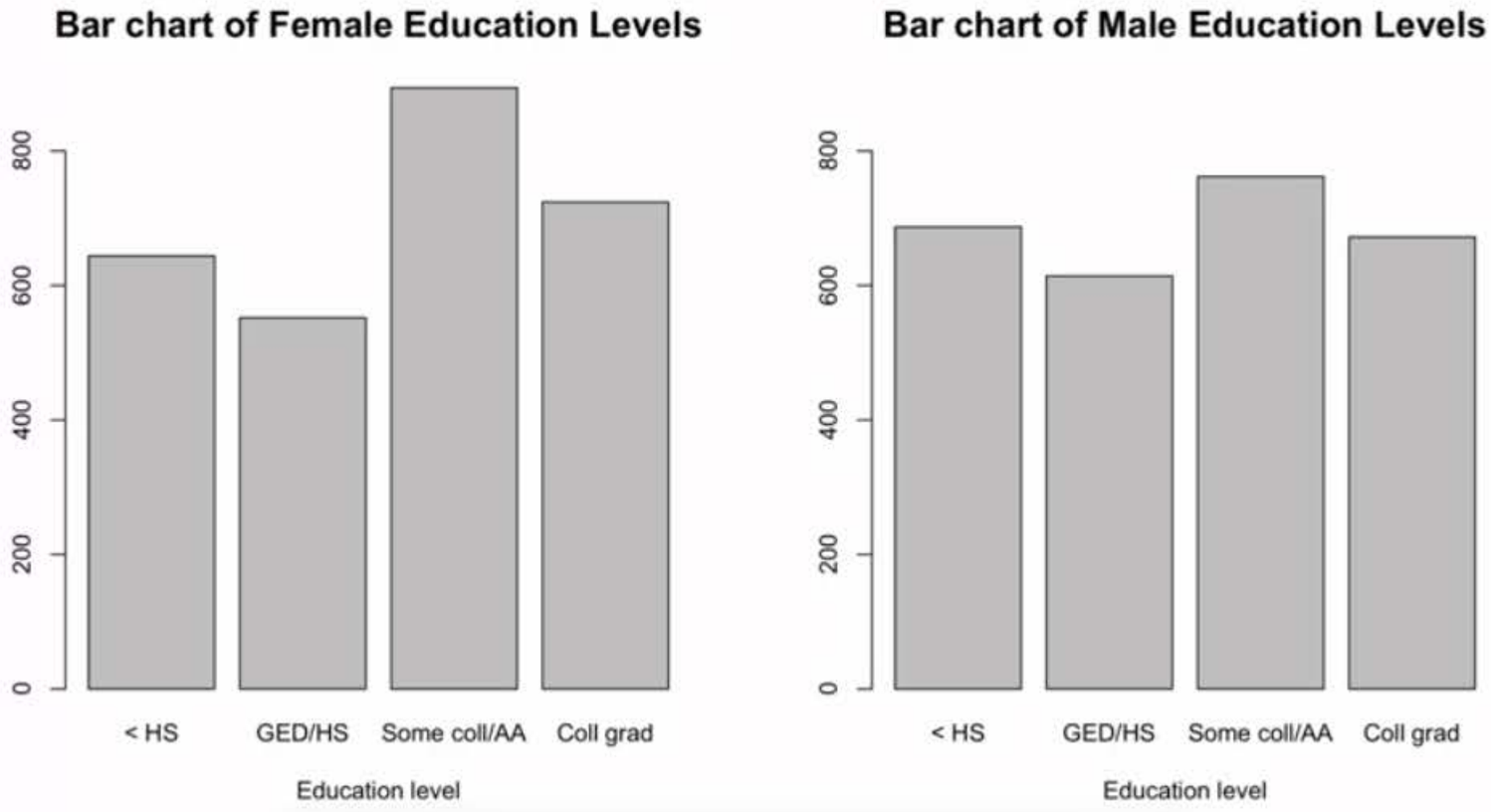

This helps us see this very quickly and at a glance in a way

# Conditional Distributions of Education Level

| | | Female | Male | **Total** |
|---|---|---|---|---|
| **Highest Education Level Attained** | Less than HS degree | 644 | 687 | 1331 |
| | HS degree or GED | 552 | 614 | 1166 |
| | Some college or AA | 894 | 762 | 1656 |
| | College degree or above | 724 | 672 | 1396 |
| | **Total** | 2814 | 2735 | 5549 |

the conditional distributions of our education level based on our gender.

# Two Univariate Bar Charts



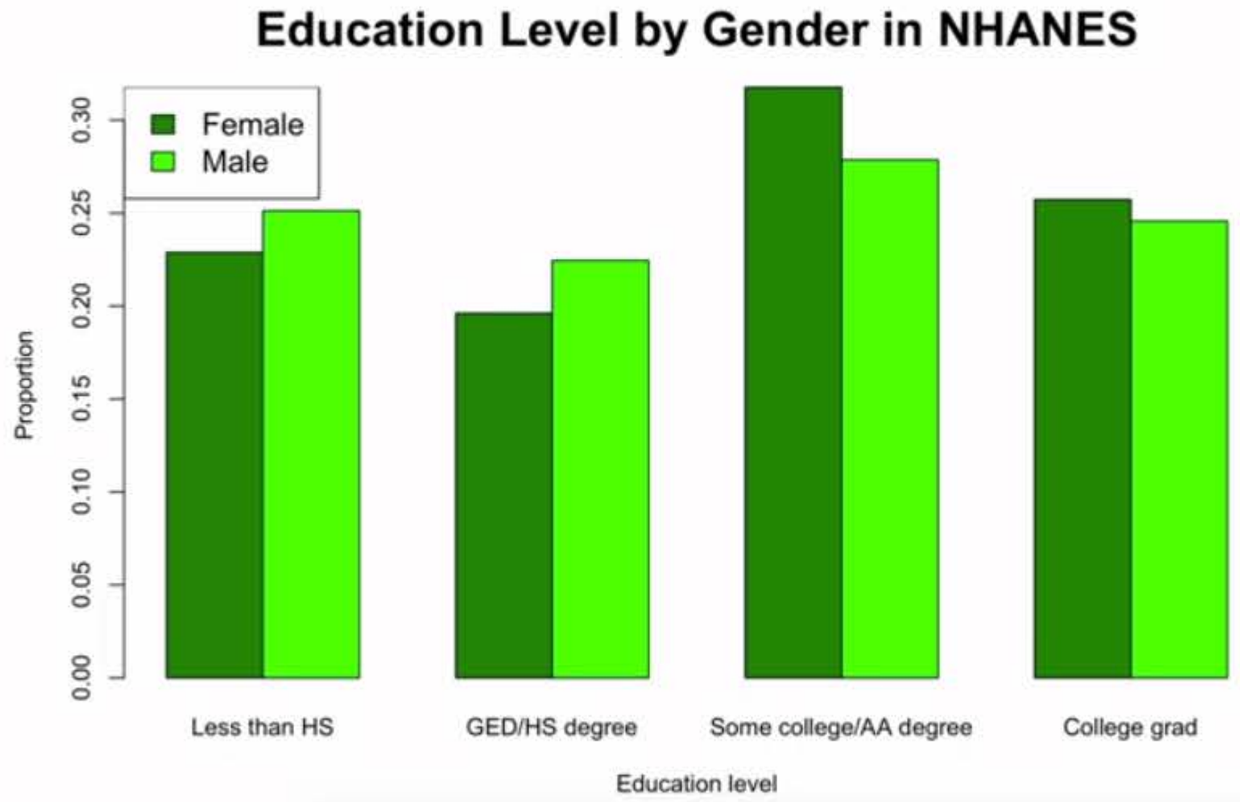

that some college or an associate's degree compared to males,

# Conditional Distributions of Education Level

| | | Female | Male | **Total** |
|---|---|---|---|---|
| **Highest Education Level Attained** | Less than HS degree | 644 (22.9%) | 687 (25.1%) | 1331 |
| | HS degree or GED | 552 (19.6%) | 614 (22.4%) | 1166 |
| | Some college or AA | 894 (31.8%) | 762 (27.9%) | 1656 |
| | College degree or above | 724 (25.7%) | 672 (24.6%) | 1396 |
| | **Total** | 2814 (100%) | 2735 (100%) | 5549 |

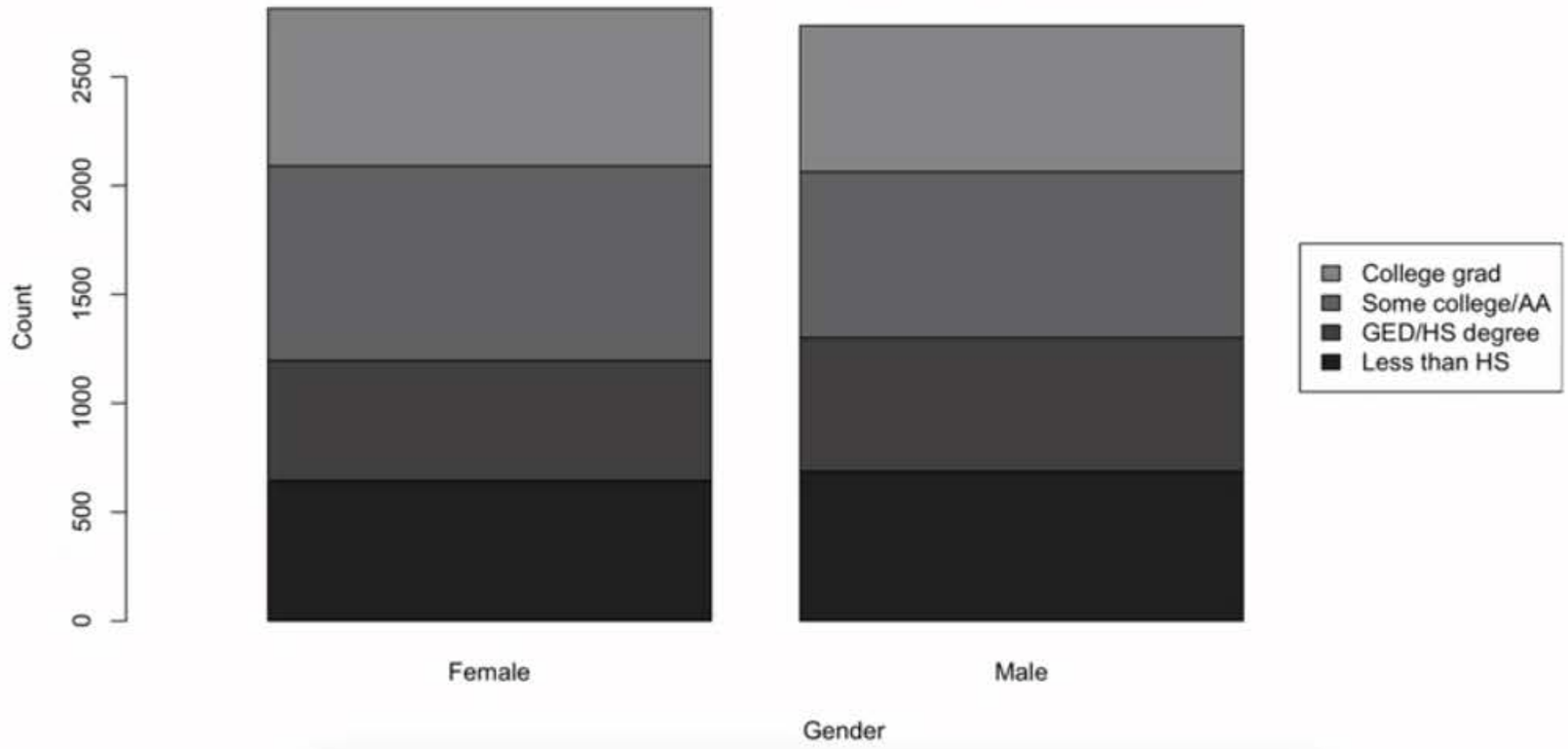
We can also display this in a side-by-side bar chart.

# Side-by-side Bar Chart of Education Level



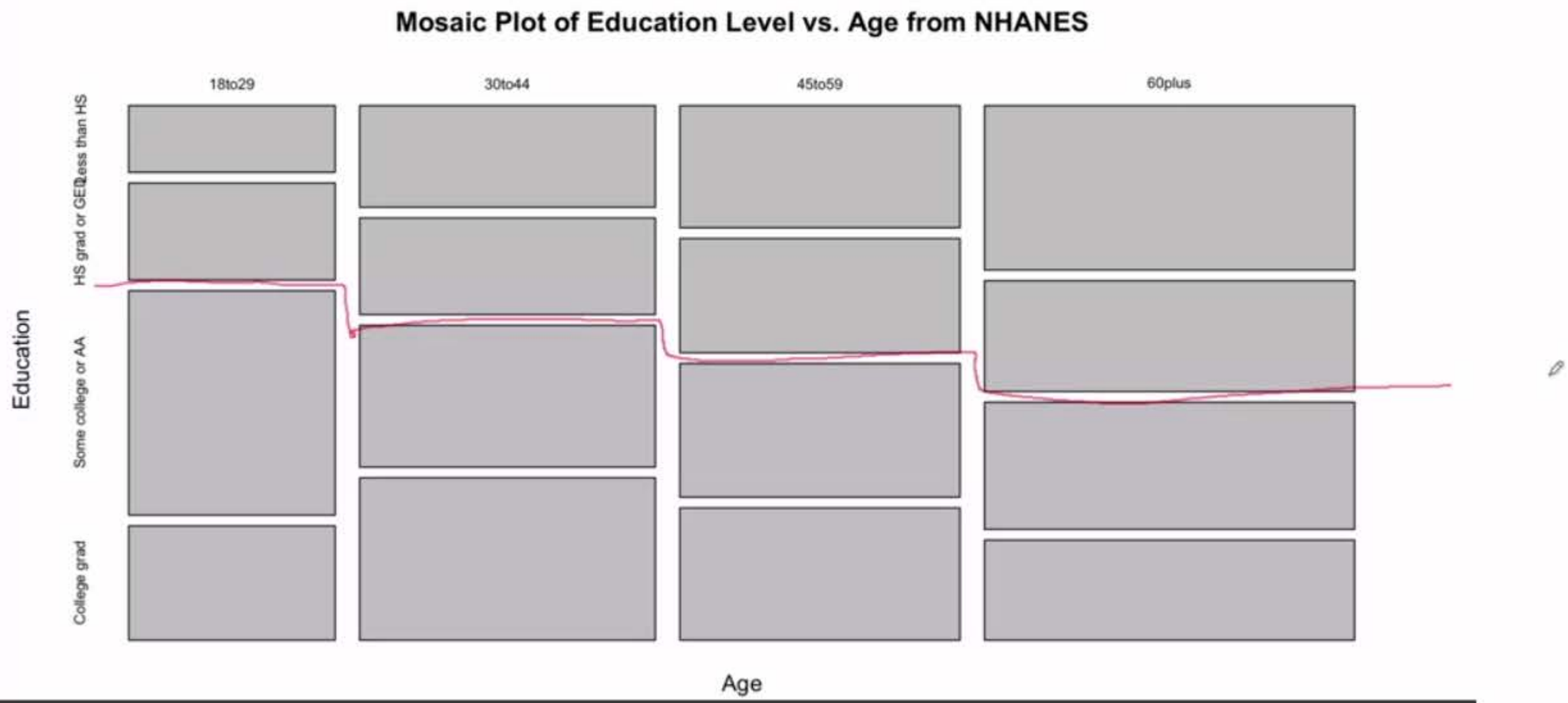

and then this is the largest difference between the female and the male education levels.

# Stacked Bar Chart of Education Level



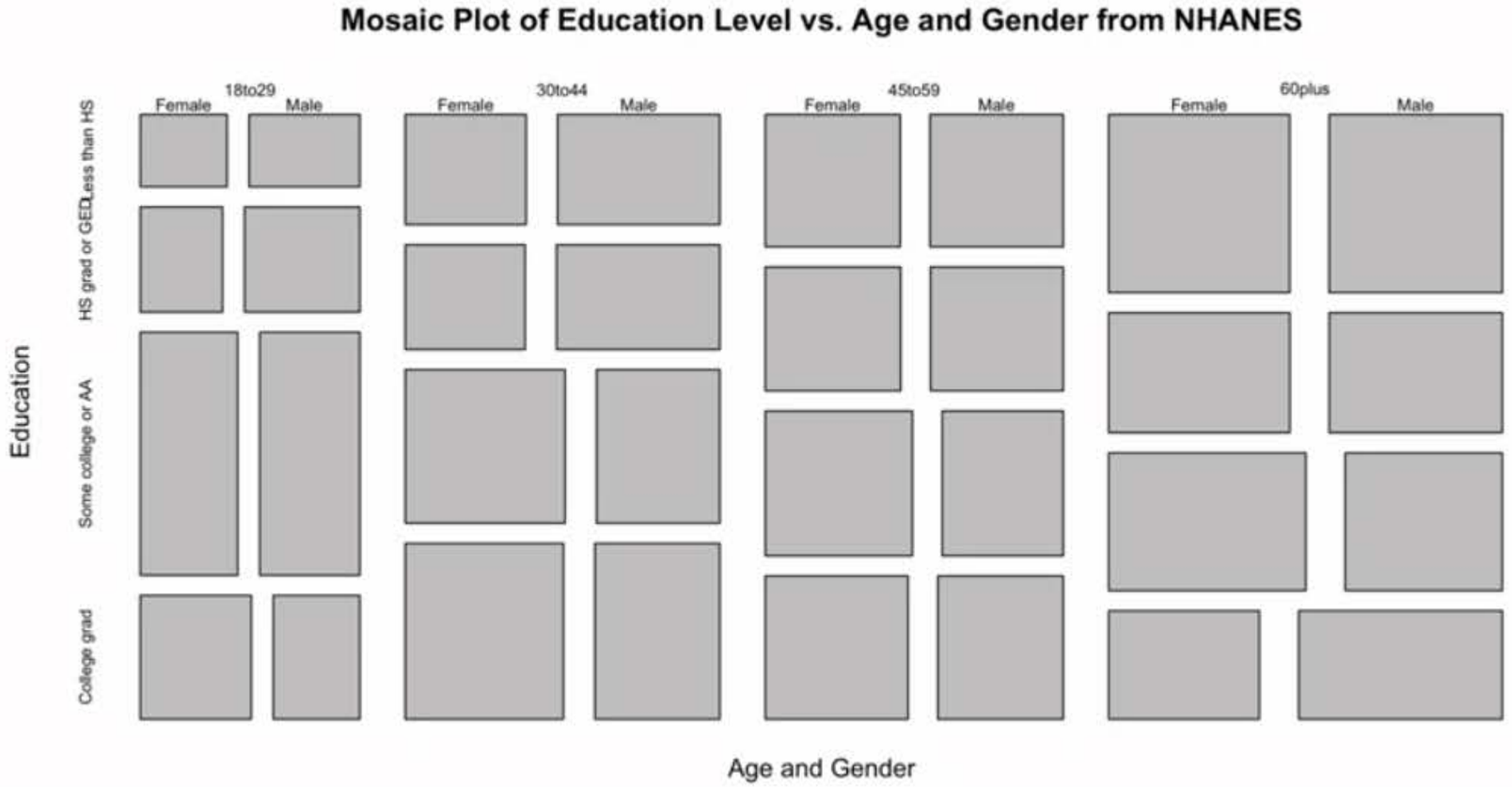

we can also see how those two groups compare to each other.

# Mosaic Plot of Education Level and Age




and is a little bit jaggered as we go through the mosaic plot.

# Mosaic Plot of Education Level, Age, and Gender




Again, we can see the same jaggedness as we look at the different age groups.

# Summary

- Multivariate Categorical Data
- Two-way or Contingency Tables
- Marginal and Conditional Distributions
- Graphs: Bar Charts, Side-by-side Bar Charts, Stacked Bar Charts, and Mosaic Plots

Overall today, we've talked about what multivariate categorical data are,