SOC-5811 Week 2: Quantitative research methods

Nick Graetz

University of Minnesota, Department of Sociology

9/10/2025

Quantitative methods

- Empiricism: social theories imply empirical claims that are falsifiable.
- Data doesn't speak for itself; it must be carefully interpreted, summarize, and analyzed.

Quantitative methods

- Statistics is about providing a concise summary of empirical data.
- This involves analyzing **samples** drawn from **populations**.

Samples and populations

Say I have a bag filled with 50 marbles of different colors. I reach in the bag and pull out 10 marbles. I see that 6 marbles are blue and 4 are red.

What is the population and sample here?

What can I say about the sample? What can I say about the population?

Generalization

Summarizing in-sample

- **Generalizing from sample to population.**

Generalization

Summarizing in-sample

- **Generalizing from sample to population**.
  - Is what I see in the sample also true in the population?

Generalization

Summarizing in-sample

- **Generalizing from sample to population**.
  - Is what I see in the sample also true in the population?
  - *I ask 30 people whether they graduated from college; what can I say about the population from which I sampled these people?*

Generalization

Summarizing in-sample

- **Generalizing from sample to population**.
  - Is what I see in the sample also true in the population?
  - *I ask 30 people whether they graduated from college; what can I say about the population from which I sampled these people?*
- **Measurement**: generalizing from **observed measurements** to the **underlying constructs of interest**.

Generalization

Summarizing in-sample

- **Generalizing from sample to population**.
  - Is what I see in the sample also true in the population?
  - *I ask 30 people whether they graduated from college; what can I say about the population from which I sampled these people?*
- **Measurement**: generalizing from **observed measurements** to the **underlying constructs of interest**.
  - Is what I'm measuring what I actually care about?

Generalization

Summarizing in-sample

- **Generalizing from sample to population**.
  - Is what I see in the sample also true in the population?
  - *I ask 30 people whether they graduated from college; what can I say about the population from which I sampled these people?*
- **Measurement**: generalizing from **observed measurements** to the **underlying constructs of interest**.
  - Is what I'm measuring what I actually care about?
  - *I ask 30 people whether they graduated from college; what am I measuring?*

Generalization

Summarizing in-sample

- **Generalizing from sample to population**.
  - Is what I see in the sample also true in the population?
  - *I ask 30 people whether they graduated from college; what can I say about the population from which I sampled these people?*
- **Measurement**: generalizing from **observed measurements** to the **underlying constructs of interest**.
  - Is what I'm measuring what I actually care about?
  - *I ask 30 people whether they graduated from college; what am I measuring?*
  - Measurement theory: very central to psychology, education, etc.

Generalization

- **Forecasting**: generalizing from the present to the future.

Generalization

- **Forecasting**: generalizing from the present to the future.
  - *I have data on whether on not it has rained every day for the past ten years. Will it rain tomorrow?*

Generalization

- **Forecasting**: generalizing from the present to the future.
  - *I have data on whether on not it has rained every day for the past ten years. Will it rain tomorrow?*
- **Causal inference**: generalizing from treatment group to control group.

Generalization

- **Forecasting**: generalizing from the present to the future.
  - *I have data on whether on not it has rained every day for the past ten years. Will it rain tomorrow?*
- **Causal inference**: generalizing from treatment group to control group.
  - *I measure the rate of a specific disease among a group that was vaccinated for that disease and one that was not. Was the vaccine effective at reducing the disease?*

Generalization: Summary

1. Generalizing from sample to population.
2. Measurement.
3. Forecasting.
4. Causal inference.

Quantitative data

```
## # A tibble: 6 x 7
##   msa                                population percapita
##   <chr>                                   <dbl>      <dbl>
## 1 Washington-Arlington-Alexand~            5949.       47.4
## 2 San Jose-Santa Clara-Sunnyva~            1919.       40.4
## 3 Atlanta, Georgia MSA                     5545.       25.3
## 4 Austin-San Marcos, Texas MSA             2271.       24.5
## 5 Little Rock-North Little Roc~             724.       20.3
## 6 Bellingham, Washington MSA                167.       20.0
```

Quantitative data

```
## # A tibble: 5 x 2
##   population region
##        <dbl> <fct>
## 1      5949. 1_Northeast
## 2      1919. 4_West
## 3      5545. 3_South
## 4      2271. 3_South
## 5       724. 3_South
```

Levels of measurement

- **Continuous variables:** infinite number of values in a given interval (e.g., time)

Levels of measurement

- **Continuous variables:** infinite number of values in a given interval (e.g., time)
- **Discrete variables:** limited to a specific set of outcomes (e.g., die roll)

Levels of measurement

- **Continuous variables:** infinite number of values in a given interval (e.g., time)
- **Discrete variables:** limited to a specific set of outcomes (e.g., die roll)
  - Ordinal: ordered, discrete categories (e.g., ranks in a race)

Levels of measurement

- **Continuous variables:** infinite number of values in a given interval (e.g., time)
- **Discrete variables:** limited to a specific set of outcomes (e.g., die roll)
  - Ordinal: ordered, discrete categories (e.g., ranks in a race)
  - Nominal: unordered, discrete categories (e.g., field of study)

Levels of measurement

- **Continuous variables:** infinite number of values in a given interval (e.g., time)
- **Discrete variables:** limited to a specific set of outcomes (e.g., die roll)
  - Ordinal: ordered, discrete categories (e.g., ranks in a race)
  - Nominal: unordered, discrete categories (e.g., field of study)
  - Binary/dichotomous: special case of nominal variable that only has two values (e.g., coin flip)

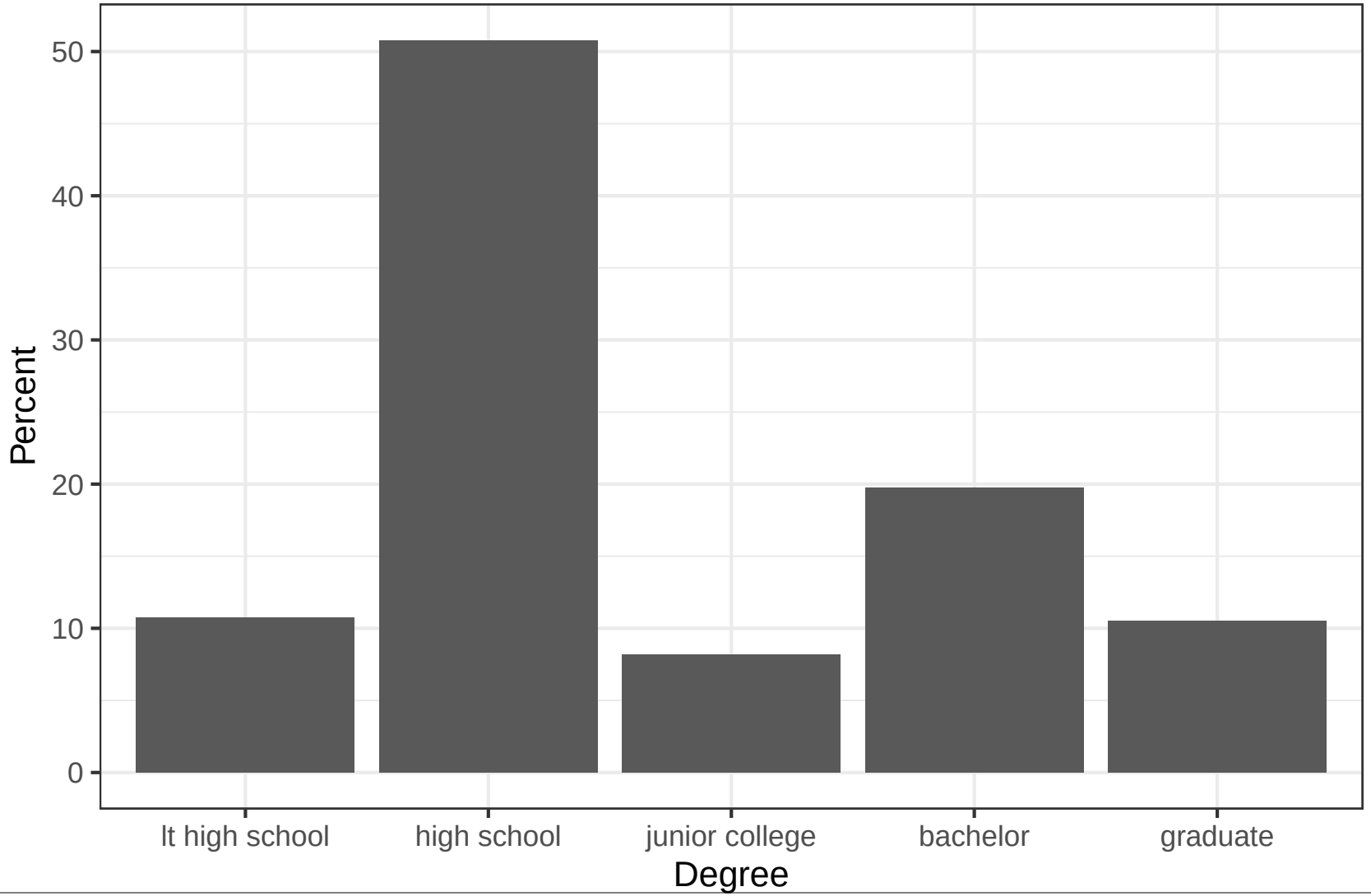
Levels of measurement

- **Continuous variables:** infinite number of values in a given interval (e.g., time)
- **Discrete variables:** limited to a specific set of outcomes (e.g., die roll)
  - Ordinal: ordered, discrete categories (e.g., ranks in a race)
  - Nominal: unordered, discrete categories (e.g., field of study)
  - Binary/dichotomous: special case of nominal variable that only has two values (e.g., coin flip)
- Measurement decisions are not always clear cut (e.g., age)

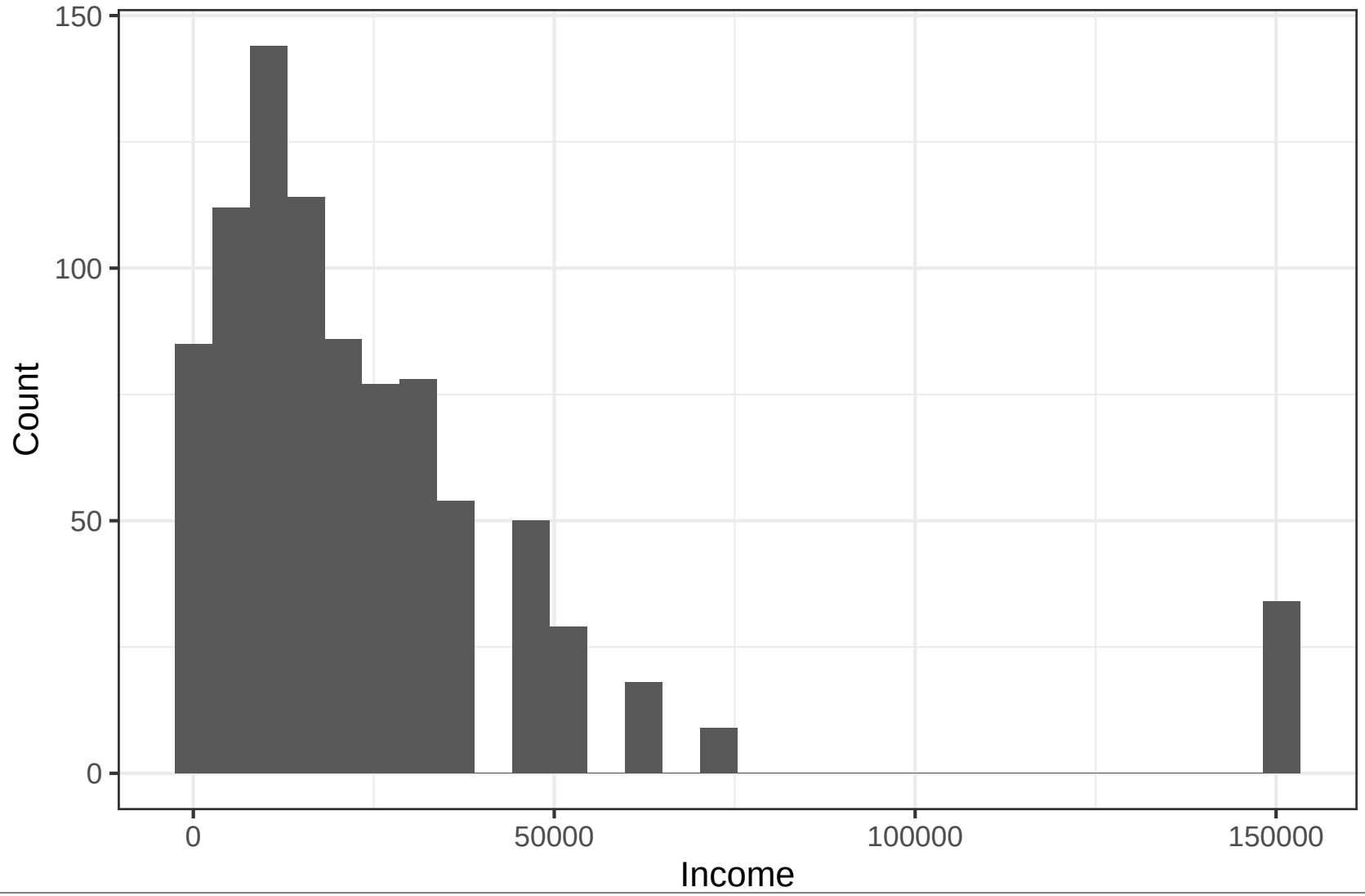
SUMMARIZING DATA

```
## # A tibble: 6 x 8
##   degree          agekdbrn  realrinc race        age pr
##   <fct>        <dbl+lbl> <dbl+lbl> <fct> <dbl+lb> <d
## 1 bachelor            35     45400  white       42
## 2 bachelor            32     54480  white       63
## 3 lt high scho~       17       908  white       62
## 4 high school         30     45400  white       55
## 5 graduate            30     54480  white       59
## 6 high school         20      8512. other       34
```

Summarizing data



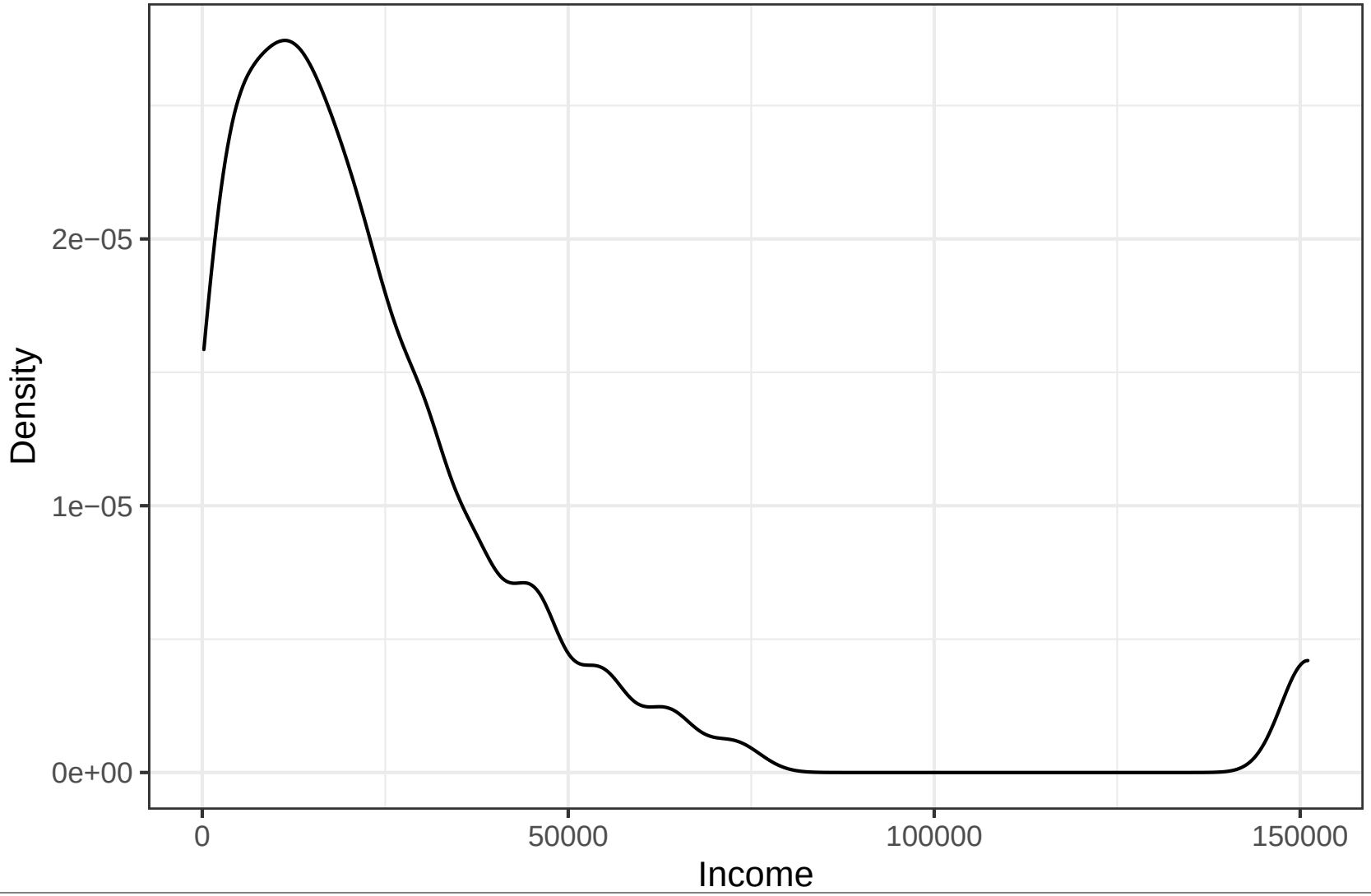
Summarizing data



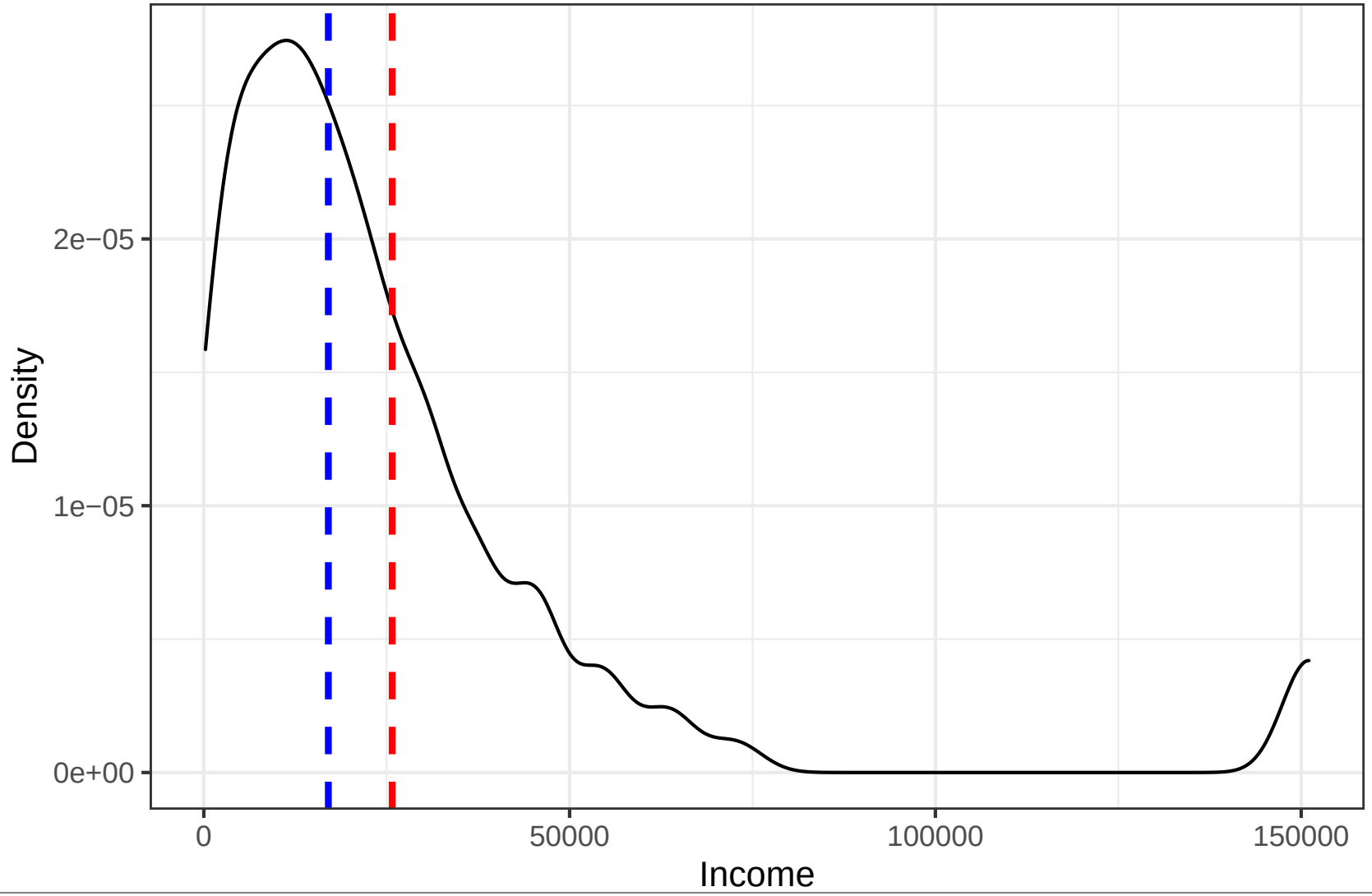
Summarizing data

- Histograms of continuous data: why do we have to bin?
- How do we choose bins?

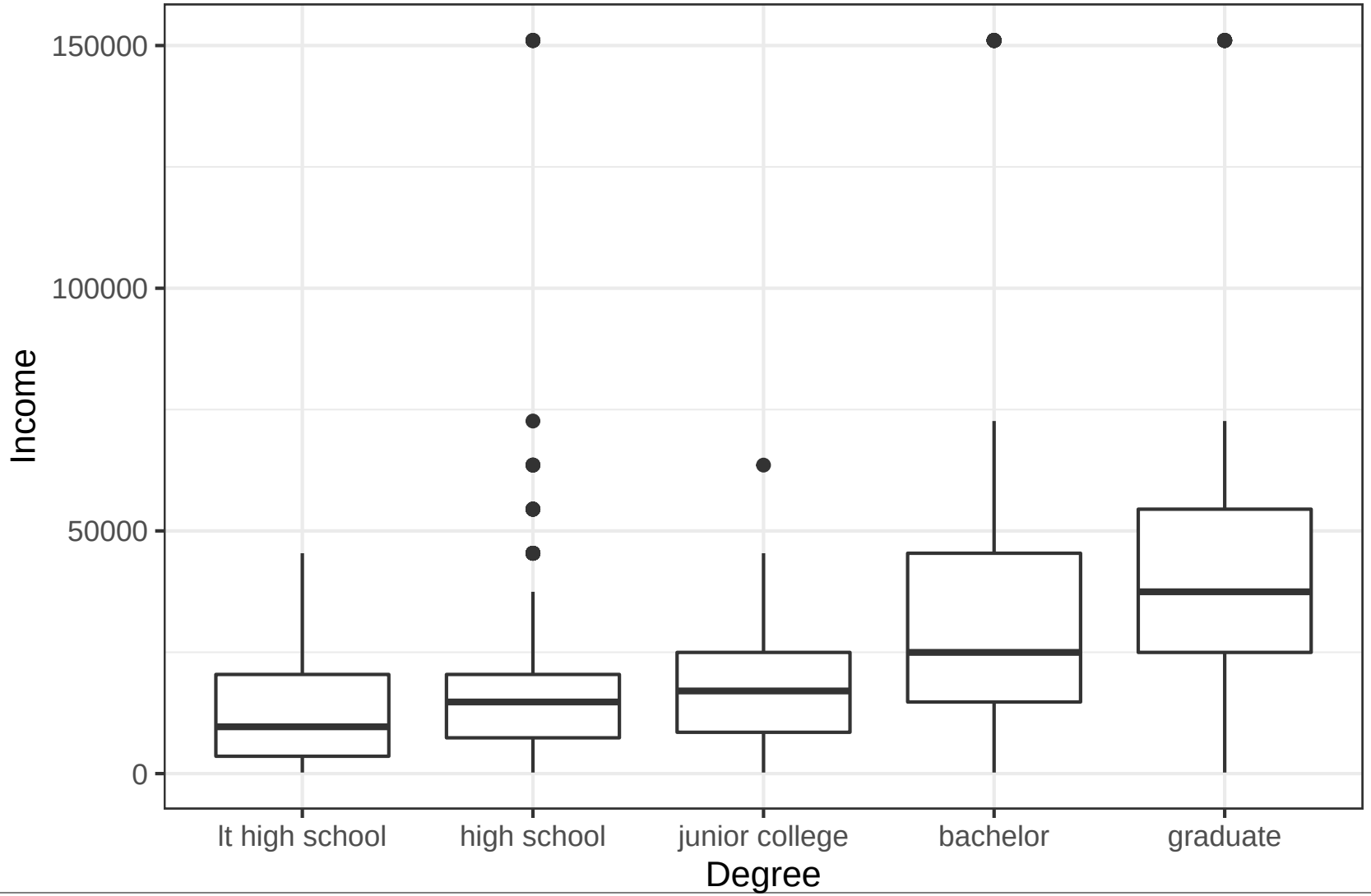
Summarizing data



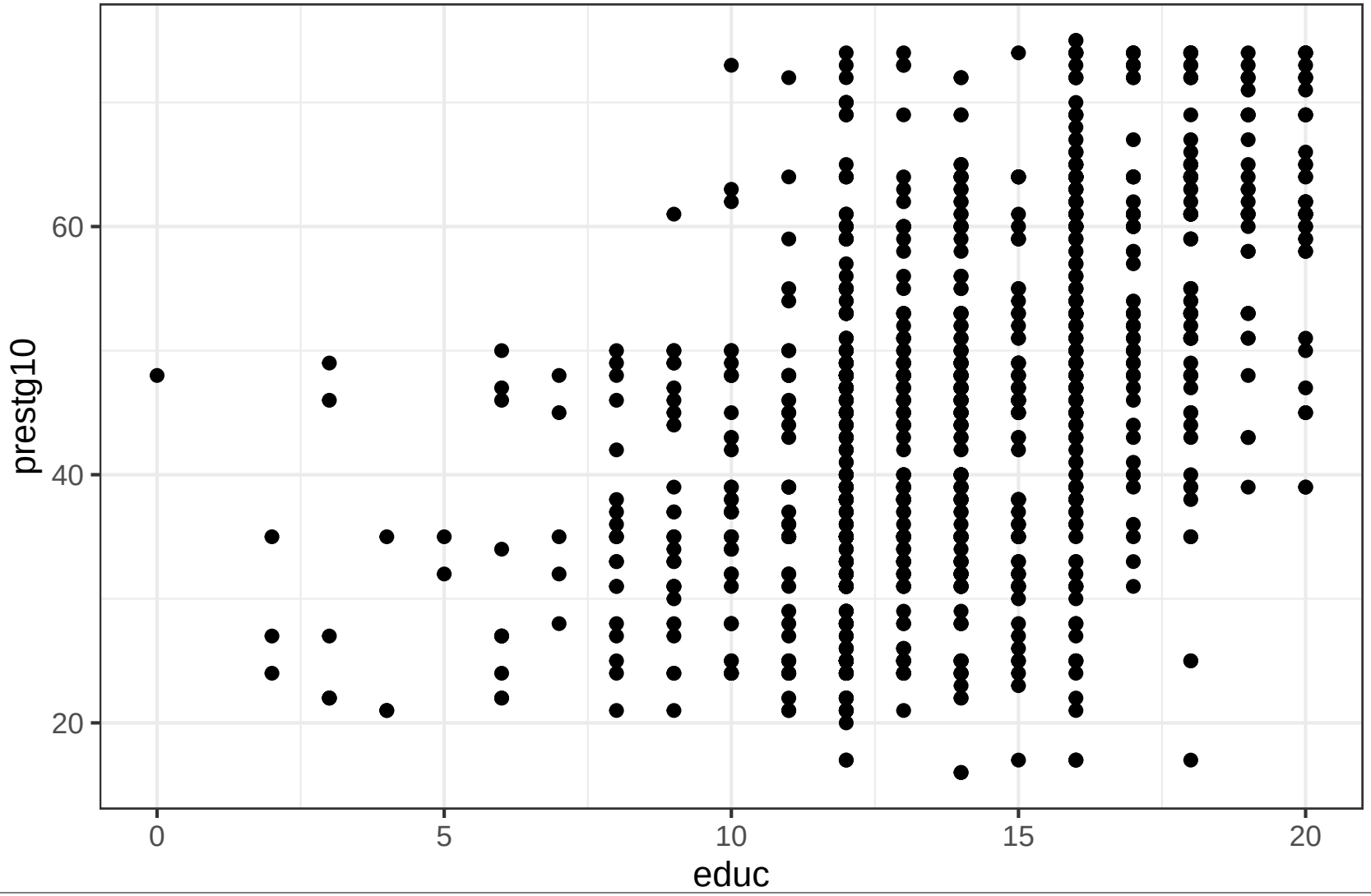
Summarizing data



Univariate summary statistics



Bivariate relationships



Bivariate relationships

```
##              educ prestg10 realrinc agekdbrn
## educ         1.00     0.49     0.35     0.42
## prestg10     0.49     1.00     0.30     0.37
## realrinc     0.35     0.30     1.00     0.22
## agekdbrn     0.42     0.37     0.22     1.00
```

BIVARIATE RELATIONSHIPS

```
##                      race
## degree               white black other
##   lt high school      0.06  0.02  0.03
##   high school         0.36  0.09  0.05
##   junior college      0.06  0.01  0.00
##   bachelor            0.15  0.03  0.02
##   graduate            0.08  0.01  0.01
```

Stats and social science theory

How do we create models?

- Deterministic processes

Stats and social science theory

How do we create models?

- Deterministic processes
- Stochastic processes