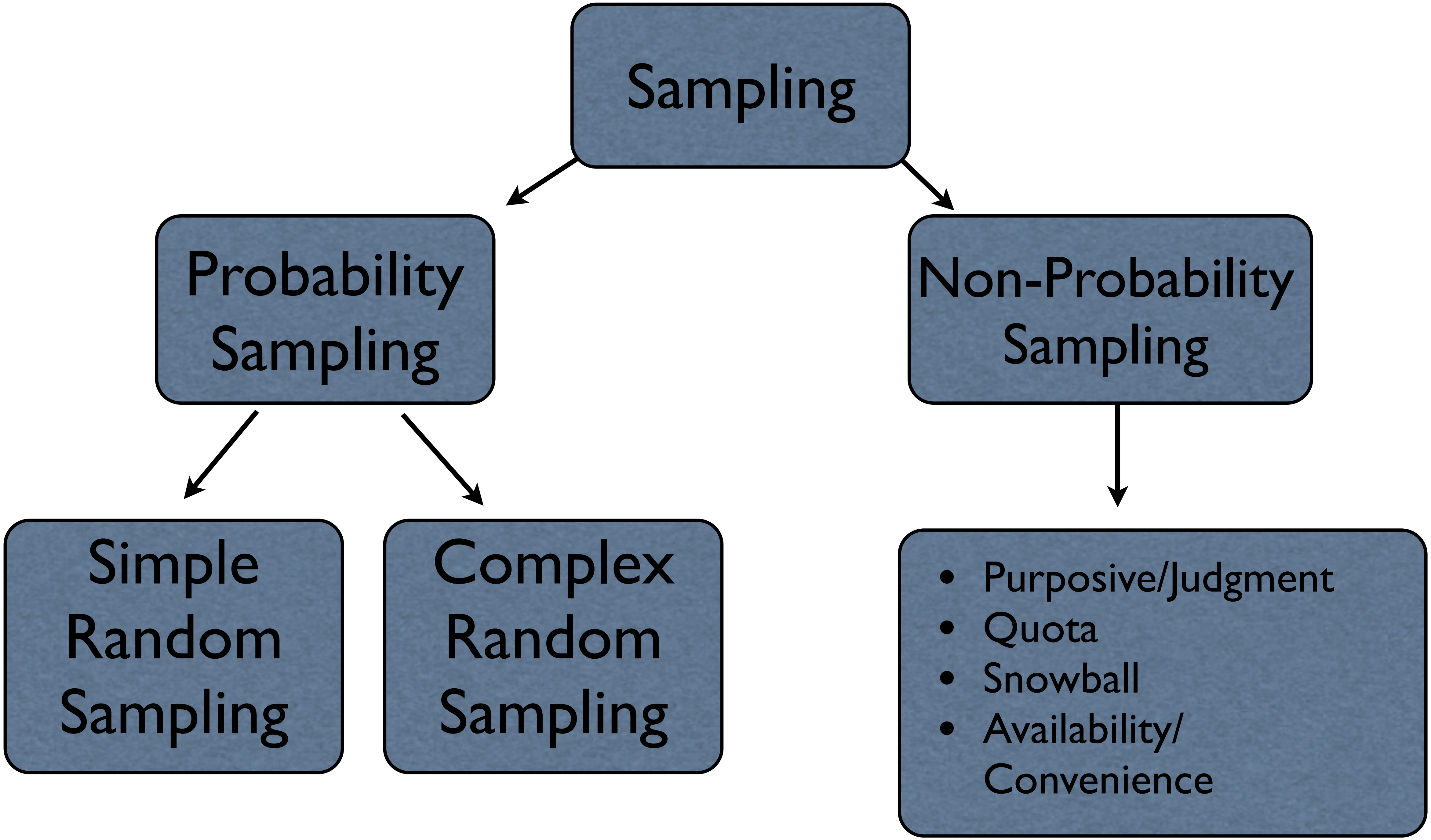
Lesson 4

Tuesday 2/6/24

Summary of Different Sampling Schema



Varieties of Research Involving Statistics

- Descriptive Analysis: "How many" or "how often" questions; documenting patterns.
- Explanatory Studies: measuring the relationship between independent and dependent variables; testing hypotheses.
- Program and policy evaluation: analyzing the effectiveness of various policies and interventions on scientifically interesting criminal justice outcomes.

Examples

- Descriptive Analysis: Which states have the highest rates of police killings?
- Explanatory Studies: Some criminological theory hypothesizes that poor school performance is a cause of crime. A basic question we could answer with data is whether school performance is correlated with crime?
- Program and policy evaluation: do batterer intervention programs reduce future domestic violence?

Police Killings by State

- Within each state, we would need a way to count the number of times someone was killed by the police in a well-defined period of time.

- There is no nationally mandated reporting system for police killings.

- Most research in this area relies on data collected by journalists (see, the Washington Post database for example).

School Performance and Crime

- If poor school performance is a cause of crime, then age should be correlated with crime.
- If school performance is a cause of crime, then the correlation should be measurable under a wide range of circumstances.
- If school performance is a cause of crime, then no other variable should be able to explain the correlation.

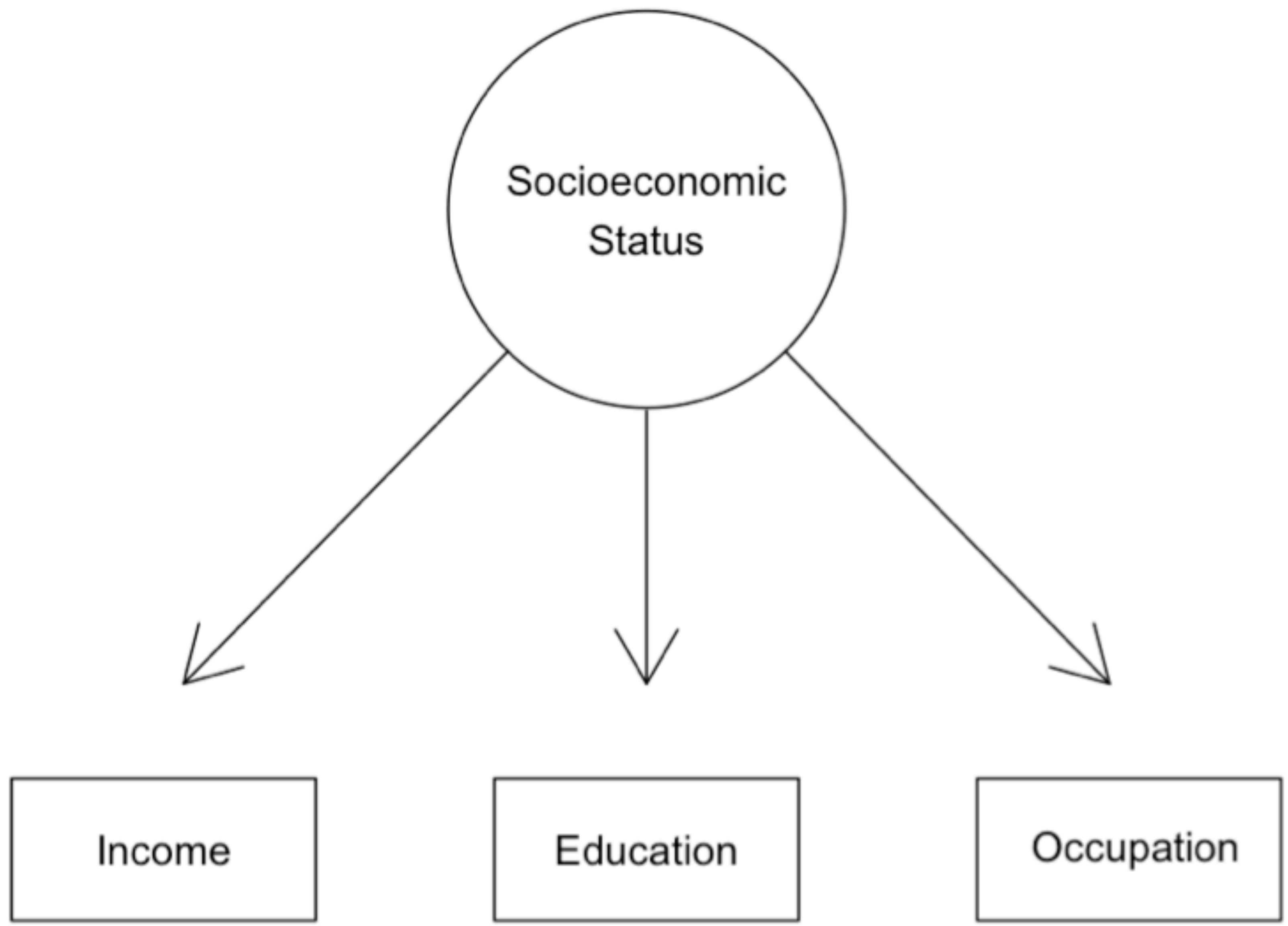
Batterer Intervention and Domestic Violence

- If a batterer intervention program is effective, it should be correlated with reduced rates of domestic violence.
- We could randomly assign abusers coming into court during some time period to 2 different groups: treatment and control. The treatment group gets the program and the control group gets whatever the business as usual condition is.
- Then, we can follow people in both groups (for the same length of time) to see whether the treatment group has better outcomes than the control group.
- This is an example of a randomized controlled experiment or trial (pp. 7-8 in your textbook).

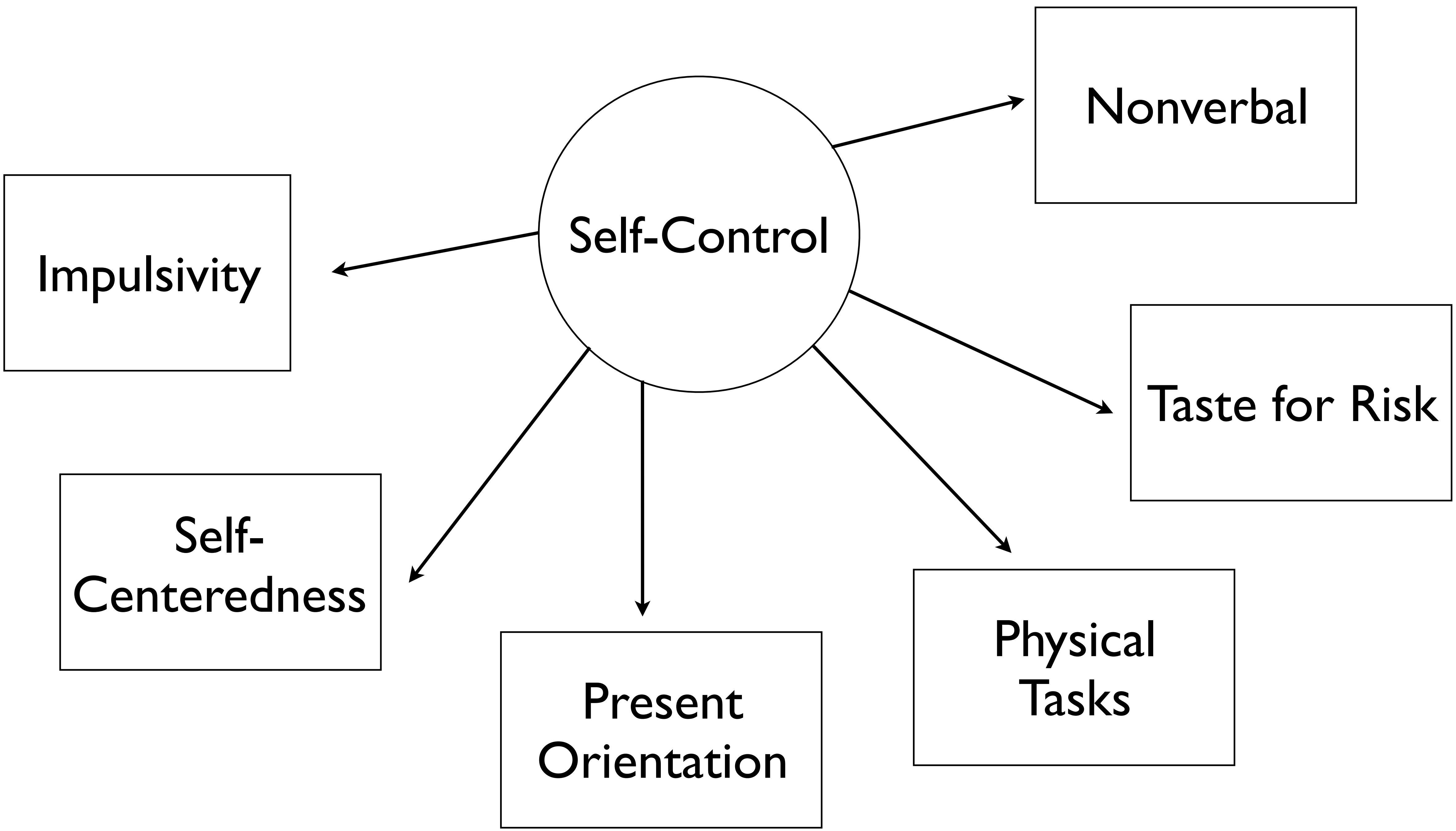
Measurement Validity

- Validity is a synonym for accuracy. In research we say a measure is valid if the measure accurately reflects the concept that is being measured.
- Example: we might use the word "recidivism" to mean "new offending among people who have offended in the past." This is a conceptual definition.
- For purposes of conducting an actual study, we might define recidivism among a cohort of prison releasees as "any arrest for a new crime within 3 years of release from prison." This would be an operational definition.

Another Example (Measurement)



Another Example (Measurement)

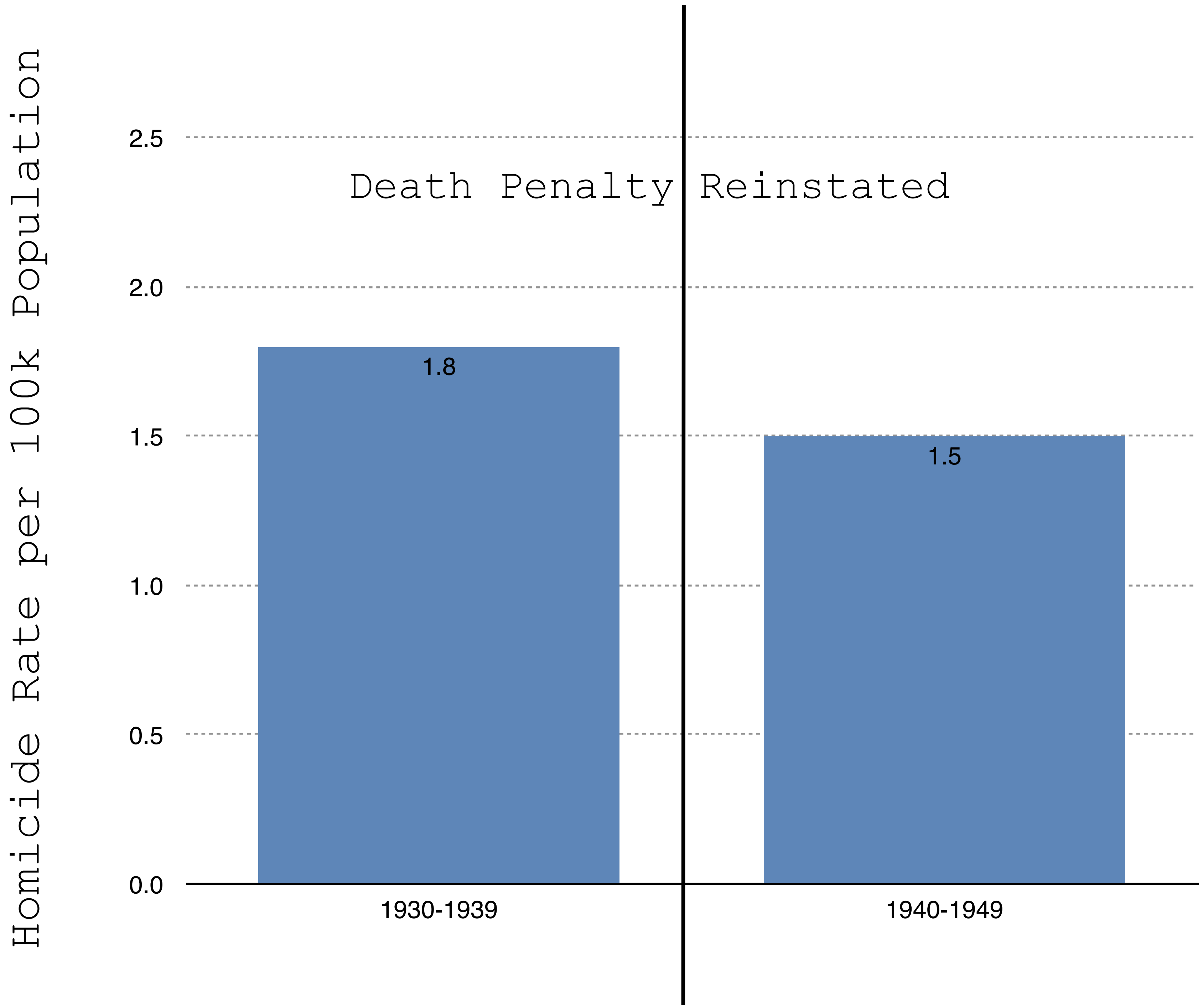


Source: Michael R. Gottfredson and Travis Hirschi (1990). A General Theory of Crime. Stanford, CA: Stanford University Press (see page 90).

Causal Validity

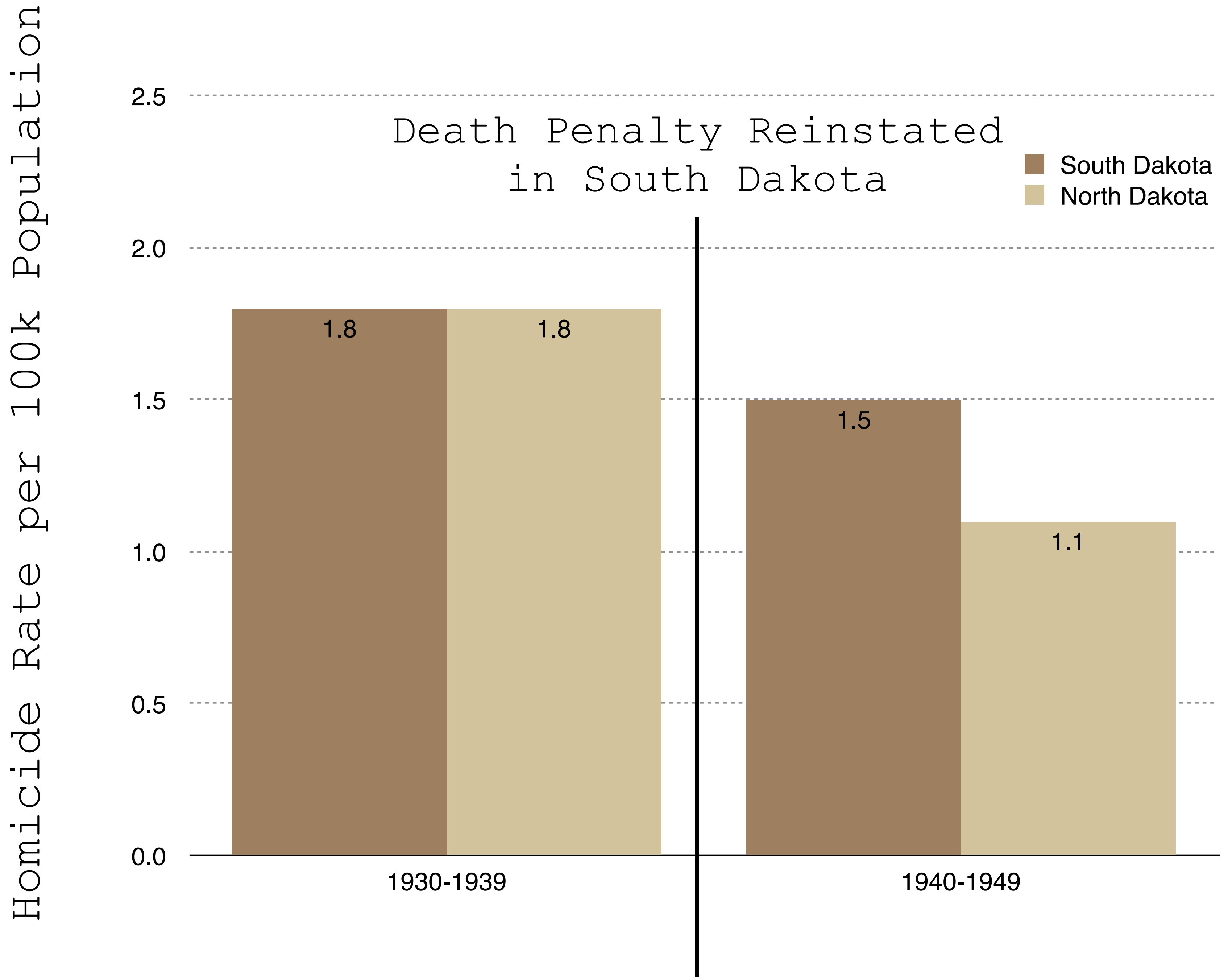
- Causal validity is a synonym for another term -- internal validity -- that is often used in research methods (note: your book has a typo on page 17 and refers to <u>interval</u> validity; this should say "internal" validity instead).
- The concern here (internal validity) is around whether a study provides convincing evidence of that a causal effect has been measured.
- Three conditions must be met to show that x is a cause of y: (1) x must precede y; (2) x and y must be correlated; (3) the correlation between x and y is not spurious.
- Establishing convincing evidence of cause-and-effect is <u>very difficult</u> as our next 3 examples will show.
- Another type of validity which we will consider later is "external validity"; has to do with the generalizability of our results to a larger population.

Homicide Rates in South Dakota



<u>Source</u>: Karl Schuessler (1952). The deterrent influence of the death penalty. <u>Annals of the American Academy of Political and Social Science</u>, 284:54-61 (at page 58).

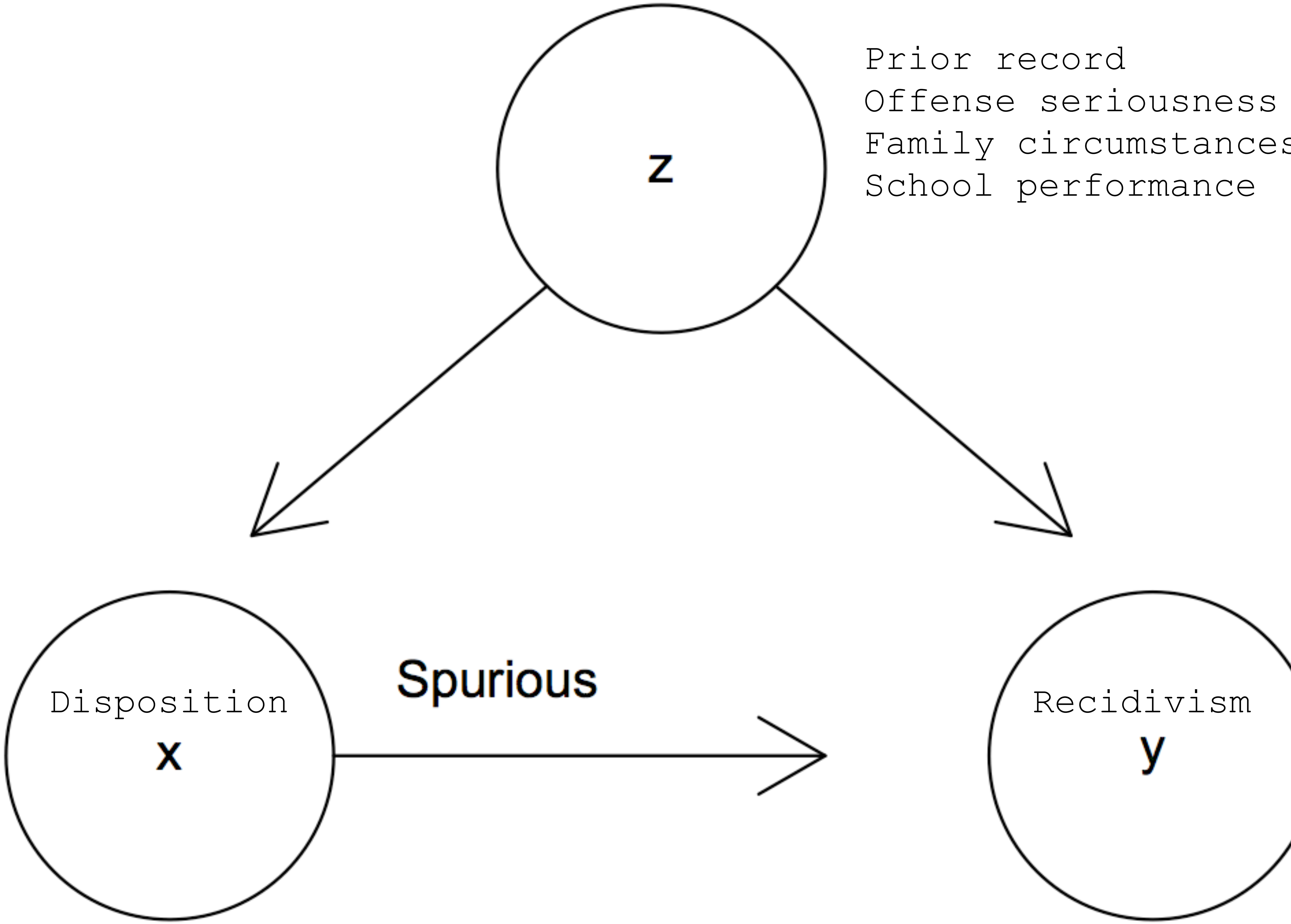
Homicide Rates in North & South Dakota

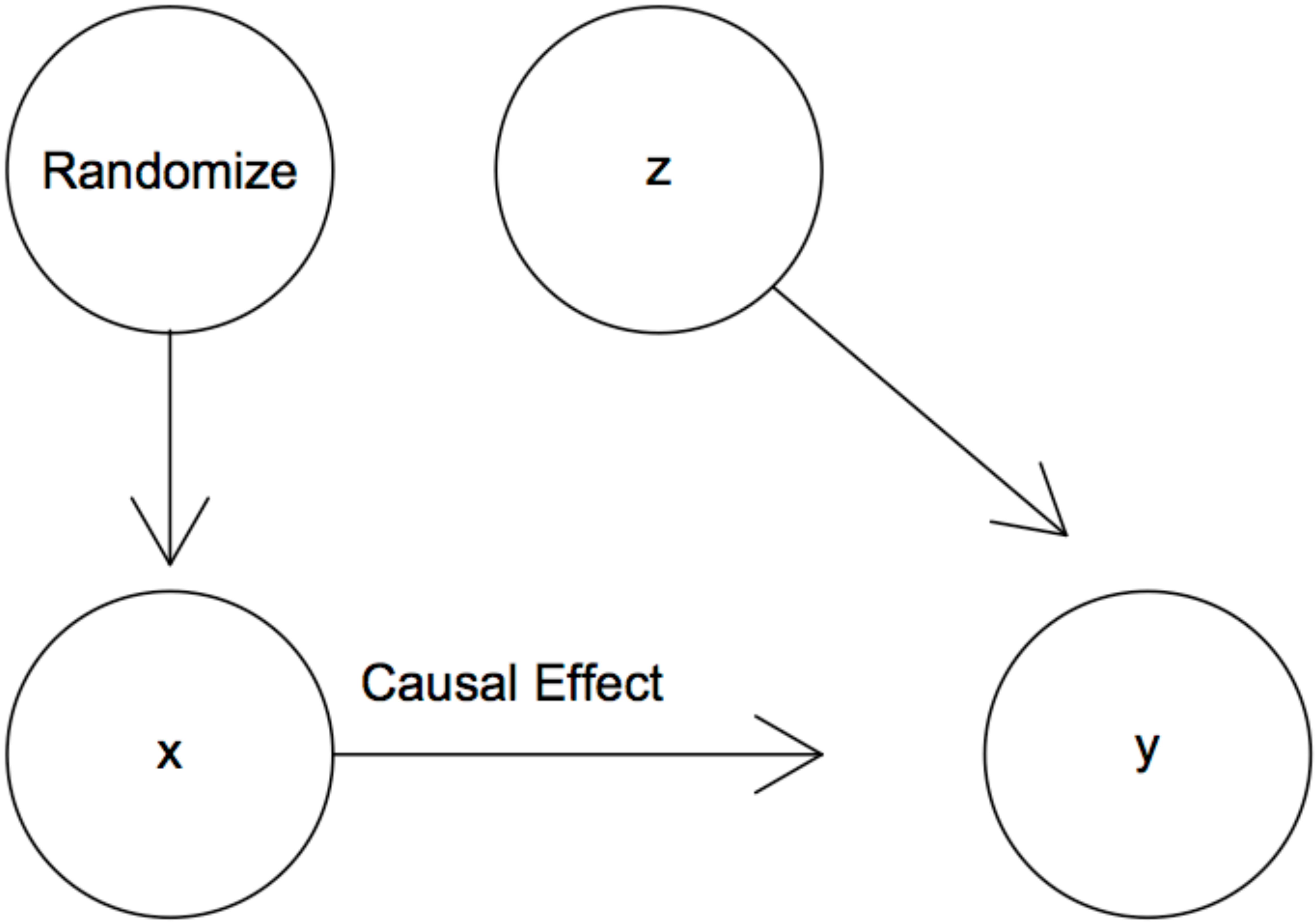


Source: Karl Schuessler (1952). The deterrent influence of the death penalty. Annals of the American Academy of Political and Social Science, 284:54-61 (at page 58).

Manski and Nagin (1998)

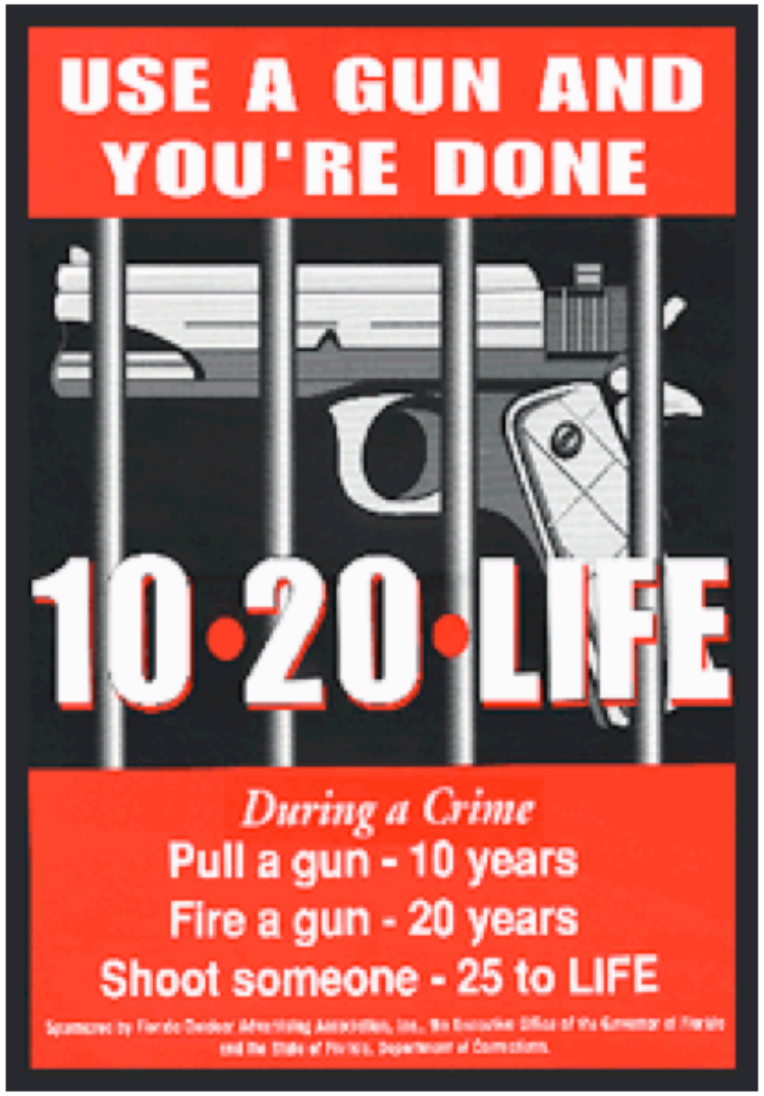
- This research study (intentionally) exposes the ambiguity that arises when we use observational data to infer causal effects.
- The study included 13,197 kids in Utah who were processed in the juvenile court for a criminal offense.
- About 11% of the kids received a "residential treatment" disposition while the other 89% of the kids remained in the community for their disposition.
- The recidivism rate (recidivism was operationally defined as a return to the juvenile court within 2 years) for residential treatment kids was 77% while the recidivism rate for the kids who remained in the community was 59%.
- What can we conclude about the causal effect of the disposition on recidivism?


z
Prior record
Offense seriousness
Family circumstances
School performance
Disposition
x
Spurious
Recidivism
y


Randomize
z
Causal Effect
x
y

Another Example: Florida's 10-20-Life Law

- Mandates a minimum 10 year prison term for certain felonies, or attempted felonies in which the offender possesses a firearm or destructive device
- Mandates a minimum 20 year prison term when the firearm is discharged
- Mandates a minimum 25 years to LIFE if someone is injured or killed
- Mandates a minimum 3 year prison term for possession of a firearm by a felon
- Mandates that the minimum prison term is to be served consecutively to any other term of imprisonment imposed



Florida’s Summary of the Law’s Effects (as recounted by Piquero (2005:786-787)

"The results under 10-20-Life are impressive. In only five years, from 1998-2003, 10-20-Life has helped drive down violent gun crime rates 28 percent statewide. During the 10-20-Life era, compared to 1998 statistics, 8,134 fewer people were robbed by armed criminals. More importantly, compared to 1998 statistics, 346 fewer people in Florida were killed by armed criminals under 10-20-Life. These crime decreases occurred even as Florida’s population increased over 2 million (14 percent) between 1998 and 2003. Punishing criminals who use guns is making our state safer."

Source: Alex R. Piquero (2005). Reliable information and rational policy decisions: does gun research fit the bill? Criminology & Public Policy, 4:779-798.

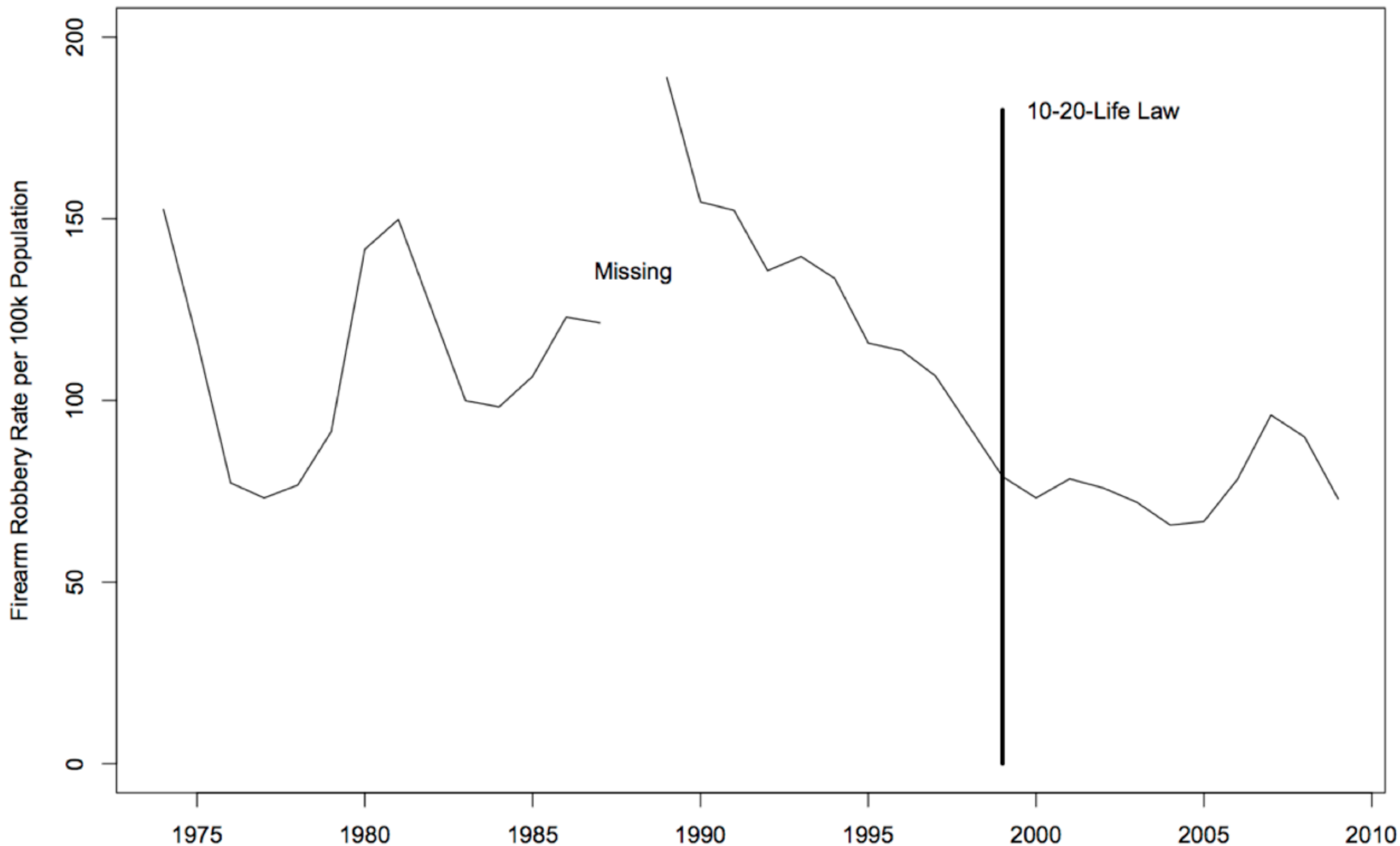
Florida Robbery Data

Statewide Robbery by Firearm, 1971 - 2009.

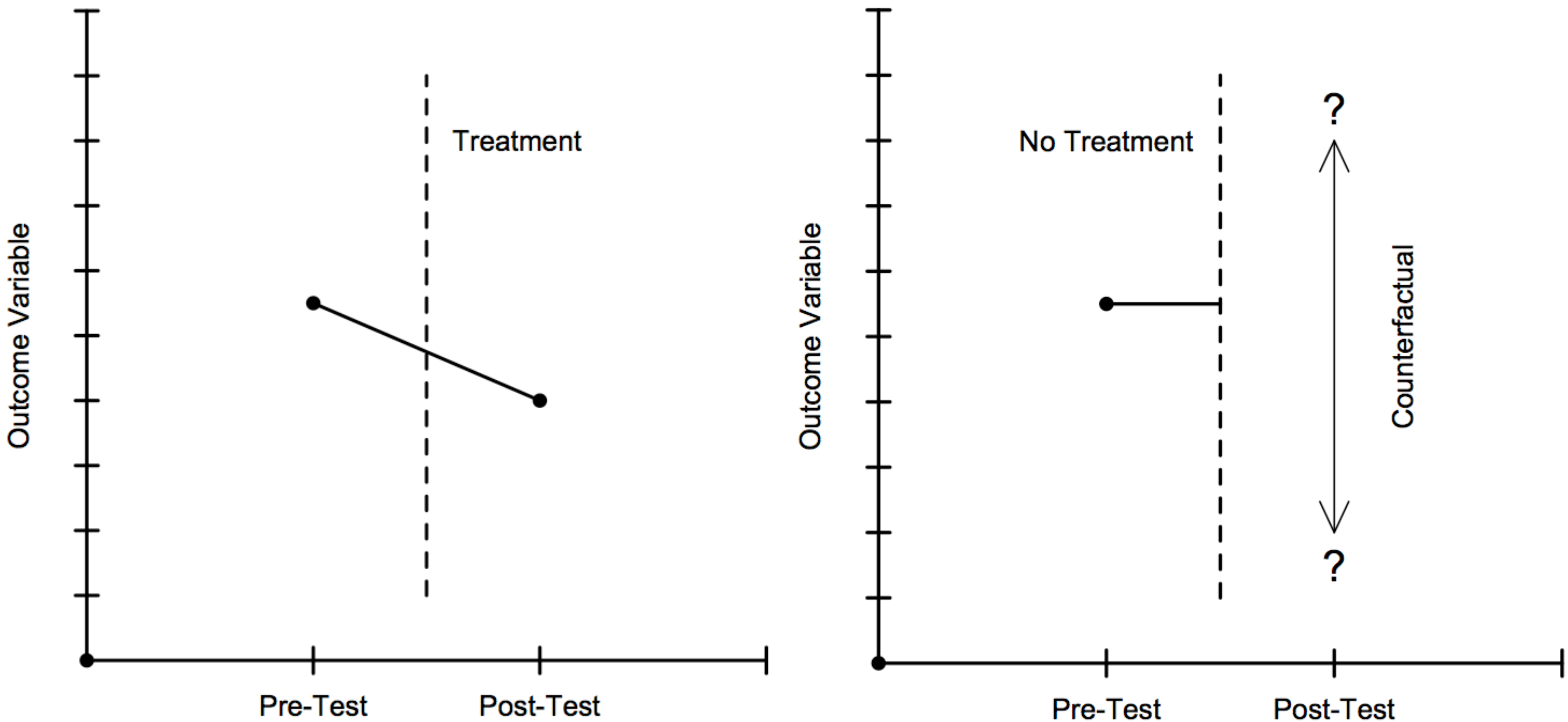
Year	Population	Robbery Total	Robbery Rate per 100,000	Total By Firearm	Total Handgun	Total Other Firearm	Total Percent by Firearms	Percent by Handguns	Percent by Other Firearms
1971	7,041,074	13,420	190.6	--	--	--	--	--	--
1972	7,441,545	13,746	184.7	--	--	--	--	--	--
1973	7,845,092	17,076	217.7	--	--	--	--	--	--
1974	8,248,851	22,261	269.9	12,585	--	--	56.5	--	--
1975	8,485,230	20,036	236.1	9,889	--	--	49.4	--	--
1976	8,551,814	15,684	183.4	6,610	--	--	42.1	--	--
1977	8,717,334	15,881	182.2	6,376	--	--	40.1	--	--
1978	8,967,206	17,700	197.4	6,876	--	--	38.8	--	--
1979	9,245,231	22,097	239	8,453	--	--	38.3	--	--
1980	9,579,497	34,015	355.1	13,565	--	--	39.9	--	--
1981	10,097,754	35,470	351.3	15,127	--	--	42.6	--	--
1982	10,375,332	31,001	298.8	12,946	--	--	41.8	--	--
1983	10,591,701	28,127	265.6	10,585	--	--	37.6	--	--
1984	10,930,389	30,320	277.4	10,734	--	--	35.4	--	--
1985	11,278,547	35,508	314.8	12,027	--	--	33.9	--	--
1986	11,657,843	42,817	367.3	14,332	--	--	33.5	--	--
1987	12,043,608	42,830	355.6	14,618	--	--	34.1	--	--
1988	12,417,606	--	--	--	--	--	--	--	--
1989	12,797,318	51,188	400	24,168	20,415	3,753	47.2	39.9	7.3
1990	13,150,027	54,015	410.8	20,334	17,018	3,316	37.6	31.5	6.1
1991	13,195,952	53,076	402.2	20,100	17,124	2,976	37.9	32.3	5.6
1992	13,424,416	48,957	364.7	18,221	15,566	2,655	37.2	31.8	5.4
1993	13,608,627	47,742	350.8	18,993	16,057	2,936	39.8	33.6	6.1
1994	13,878,905	45,263	326.1	18,541	15,963	2,578	41	35.3	5.7
1995	14,149,317	42,142	297.8	16,384	13,979	2,405	38.9	33.2	5.7
1996	14,411,563	41,643	289	16,385	--	--	39.3	--	--
1997	14,712,922	40,703	276.6	15,707	--	--	38.6	--	--
1998	15,000,475	36,130	240.9	13,937	--	--	38.6	--	--
1999	15,322,040	31,996	208.8	12,111	--	--	37.9	--	--
2000	15,982,378	31,392	196.4	11,692	--	--	37.2	--	--
2001	16,331,739	32,808	200.9	12,804	--	--	39.03	--	--
2002	16,674,608	32,413	194.4	12,656	--	--	39.05	--	--
2003	17,071,508	31,512	184.6	12,288	--	--	38.99	--	--
2004	17,516,732	29,984	171.2	11,504	--	--	38.4	--	--
2005	17,918,227	30,092	167.9	11,945	--	--	39.7	--	--
2006	18,349,312	34,123	186.0	14,362	--	--	42.1	--	--
2007	18,680,367	38,112	204.0	17,926	--	--	47.0	--	--
2008	18,807,219	36,232	192.6	16,917	--	--	46.7	--	--
2009	18,750,483	30,881	164.7	13,668	--	--	44.3	--	--

SOURCE: Florida Statistical Analysis Center: FDLE, Crime in Florida, Florida uniform crime report, 1971-2009. Tallahassee, FL.

Florida Gun Robberies (1971-2009)



Fundamental Problem of Causal Inference



Reliability

- Synonyms for reliability include words like "reproducibility" or "repeatability."
- Test-retest: give a test at time 1; repeat at time 2; how often do I get the same answer?
- Different raters scoring the same phenomenon; how often do the raters produce the same answer?
- Survey questions that ask about the same phenomenon but in somewhat different ways; how correlated are the answers to the different survey questions?
- Note: a measure must be reliable to be valid; the reverse is not true, however. A measure can be reliable and invalid.

Practice Analysis Exercise 1

```
# read the dataset

year=seq(from=1990,to=2022,by=1)
hr=c(9.4,9.8,9.3,9.5,9,8.2,7.4,6.8,6.3,
  5.7,5.5,5.6,5.6,5.7,5.5,5.6,5.8,
  5.7,5.4,5,4.8,4.7,4.7,4.5,4.4,4.9,
  5.3,5.3,5.0,5.0,6.5,6.8,6.3)
df=data.frame(year,hr)

# part 1: create a chart

plot(x=df$year,y=df$hr,
  type="l",
  ylim=c(0,10),
  xlab="Year (1990-2022)",
  ylab="FBI Murders per 100k Population",
  main="FBI Murder Rate by Year (1990-2022)")
points(x=df$year,y=df$hr,pch=19)

# part 2: get some information about the data

subset(df,hr==min(hr))
subset(df,hr==max(hr))
mean(df$hr)
subset(df,hr<mean(hr))
subset(df,hr>mean(hr))
```

Practice Analysis Exercise 2

```
# read the dataset

age=c(rep(16,19),rep(17,161),rep(18,492),rep(19,480),rep(20,624),
  rep(21,599),rep(22,580),rep(23,468),rep(24,537),rep(25,443),rep(26,432),
  rep(27,338),rep(28,415),rep(29,292),rep(30,324),rep(31,254),rep(32,234),
  rep(33,179),rep(34,187),rep(35,167),rep(36,177),rep(37,132),rep(38,152),
  rep(39,117),rep(40,119),rep(41,93),rep(42,113),rep(43,102),rep(44,85),
  rep(45,75),rep(46,90),rep(47,72),rep(48,86),rep(49,62),rep(50,78),
  rep(51,61),rep(52,57),rep(53,50),rep(54,44),rep(55,49),rep(56,55),
  rep(57,34),rep(58,34),rep(59,25),rep(60,21),rep(61,18),rep(62,19),
  rep(63,11),rep(64,16),rep(65,7),rep(66,5),rep(67,13),rep(68,5),rep(69,3),
  rep(70,1),rep(71,3),rep(72,5),rep(73,3),rep(74,4),rep(75,2),rep(77,2),rep(78,2))
n=length(age)
n

# part 1: create a chart

barplot(table(age),
  xlab="Age (in years) at Time of Release",
  ylab="Number of People",
 main="Age at Release from Prison (1978 NCDOC)")

# part 2: get some information about the population

mean(age)
median(age)
```

Practice Analysis Exercise 2 (Continued)

```
# part 3: study a simple random sample

s=sample(1:n,size=100,replace=T)
sample.age=age[s]
mean(sample.age)
median(sample.age)

# part 4: create a chart showing the ages in the sample

barplot(table(sample.age),
  xlab="Age (in years) at Time of Release",
  ylab="Number of People",
 main="Age at Release for Random Sample")
```