3) Experiment

Vitor Kamada

January 2018

Reference

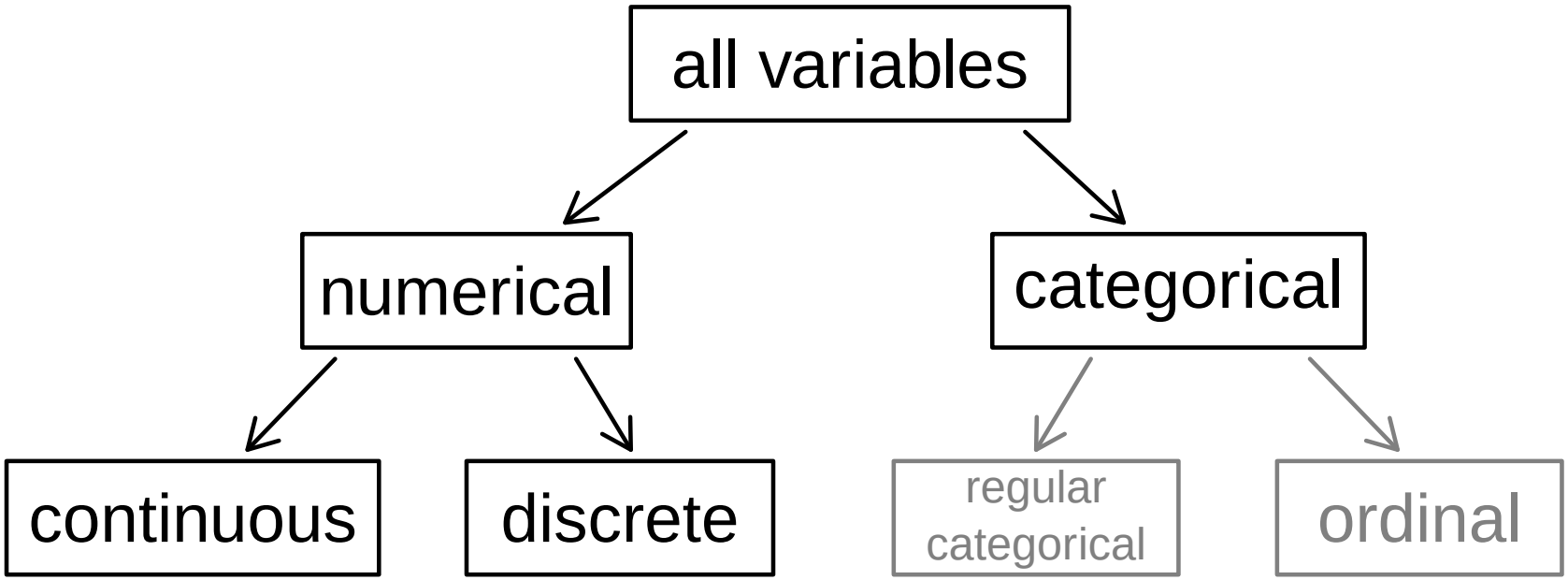
Tables, Graphics, and Figures from

Introductory Statistics with

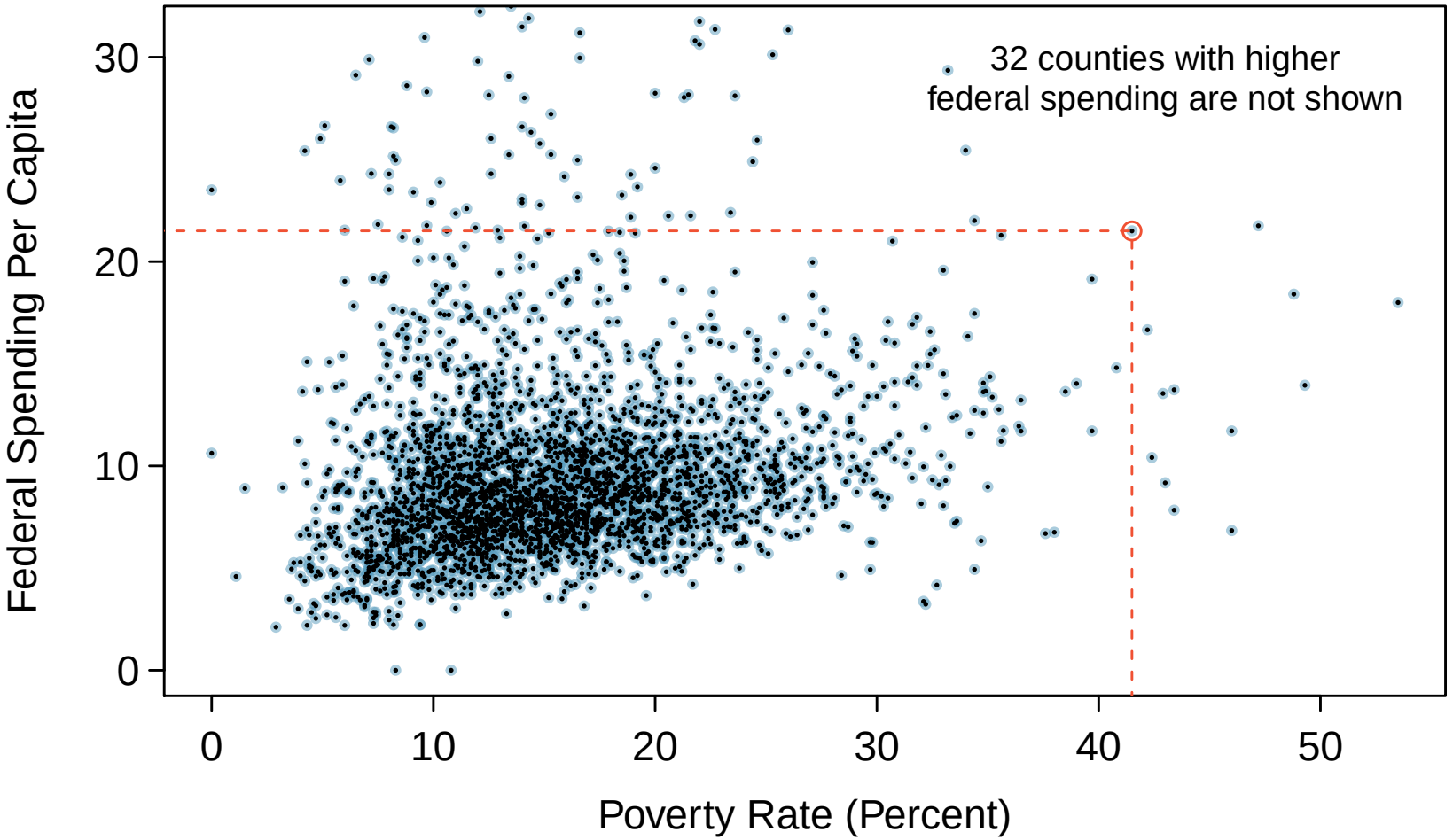
Randomization and Simulation

Diez et al. (2014): Chapter 1 - Introduction to Data

Types of Variables



Relationships between Variables



Anecdotal Evidence

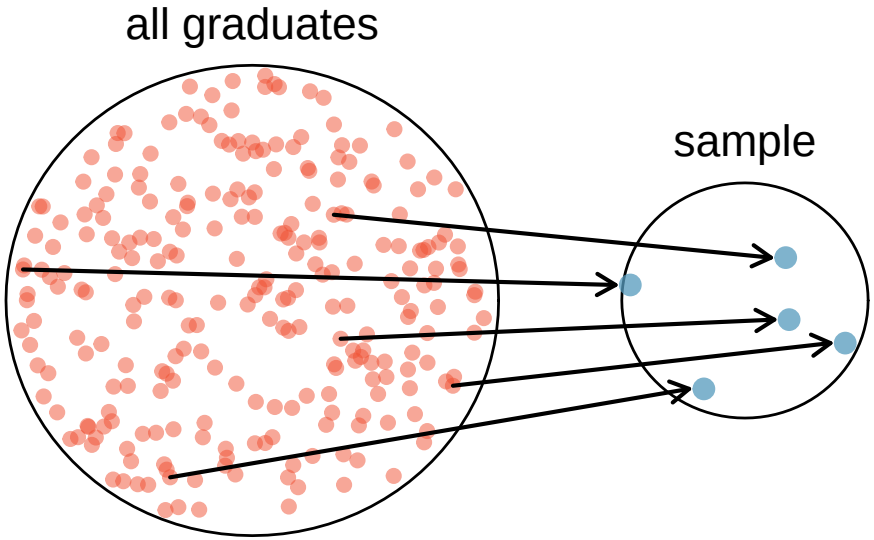
In Feb 2010, some media cited one large snow storm as evidence against global warming.

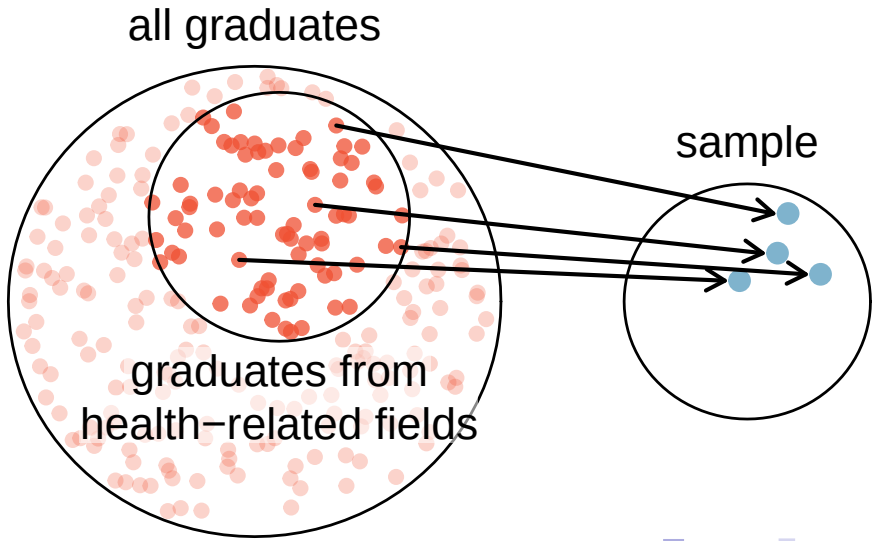
As comedian Jon Stewart pointed out: "*It's one storm, in one region, of one country.*"

Population vs Sample

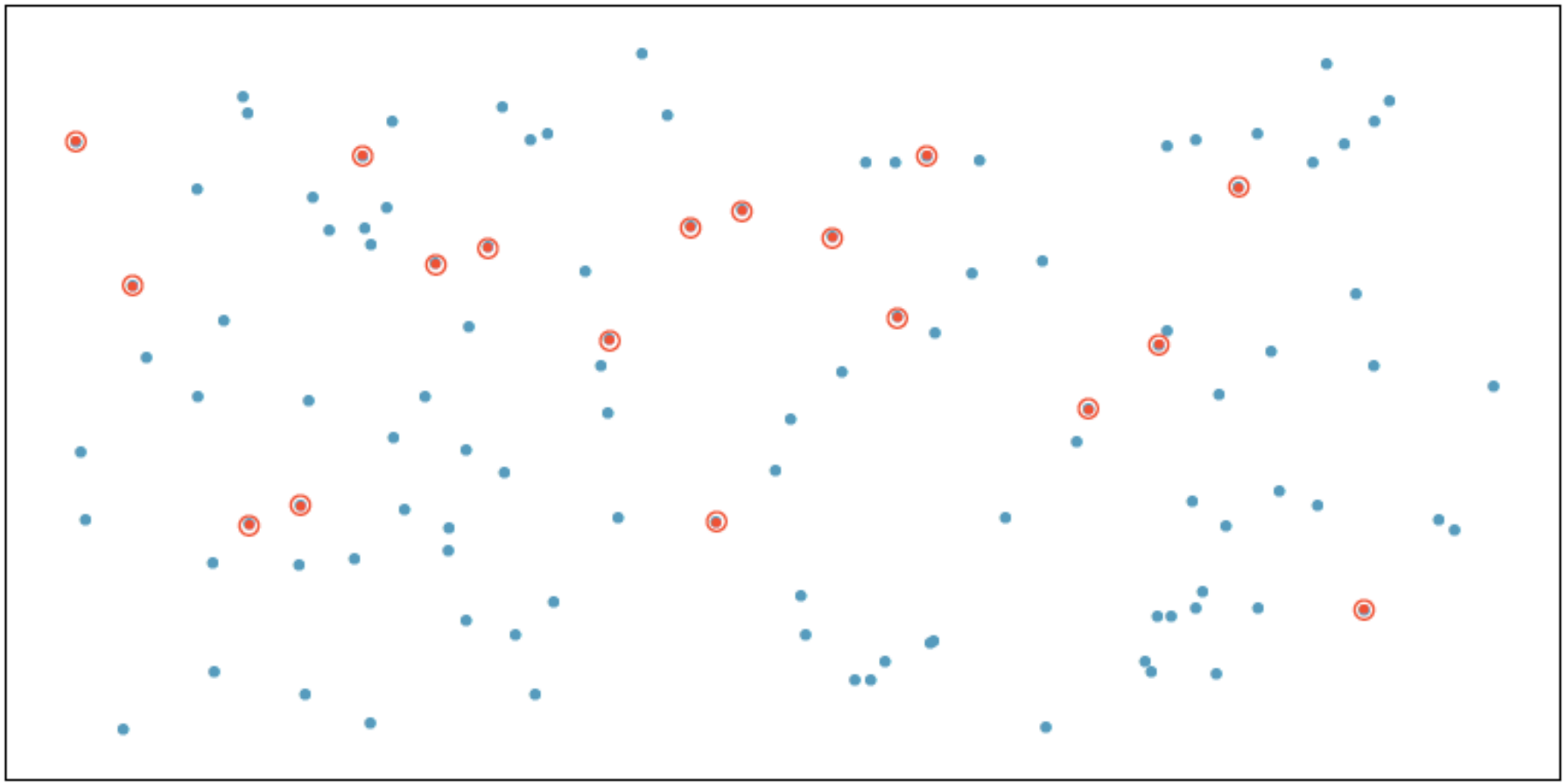
	Parameter/Random Variable	Statistic/Data
Mean	$\mu = \sum_{i=1}^{k} x_i P(X = x_i)$	$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$
Variance	$\sigma^2 = \sum_{j=1}^{k} (x_j - \mu)^2 P(X = x_j)$	$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$
SD	σ	s
Covariance	$\sigma_{xy} = E[(X - \mu_x)(Y - \mu_y)]$	$s_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1}$
Correlation	$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$	$r = \frac{Cov(x,y)}{s_x s_y}$
Proportion	p	$\hat{p}$

Sampling from a Population

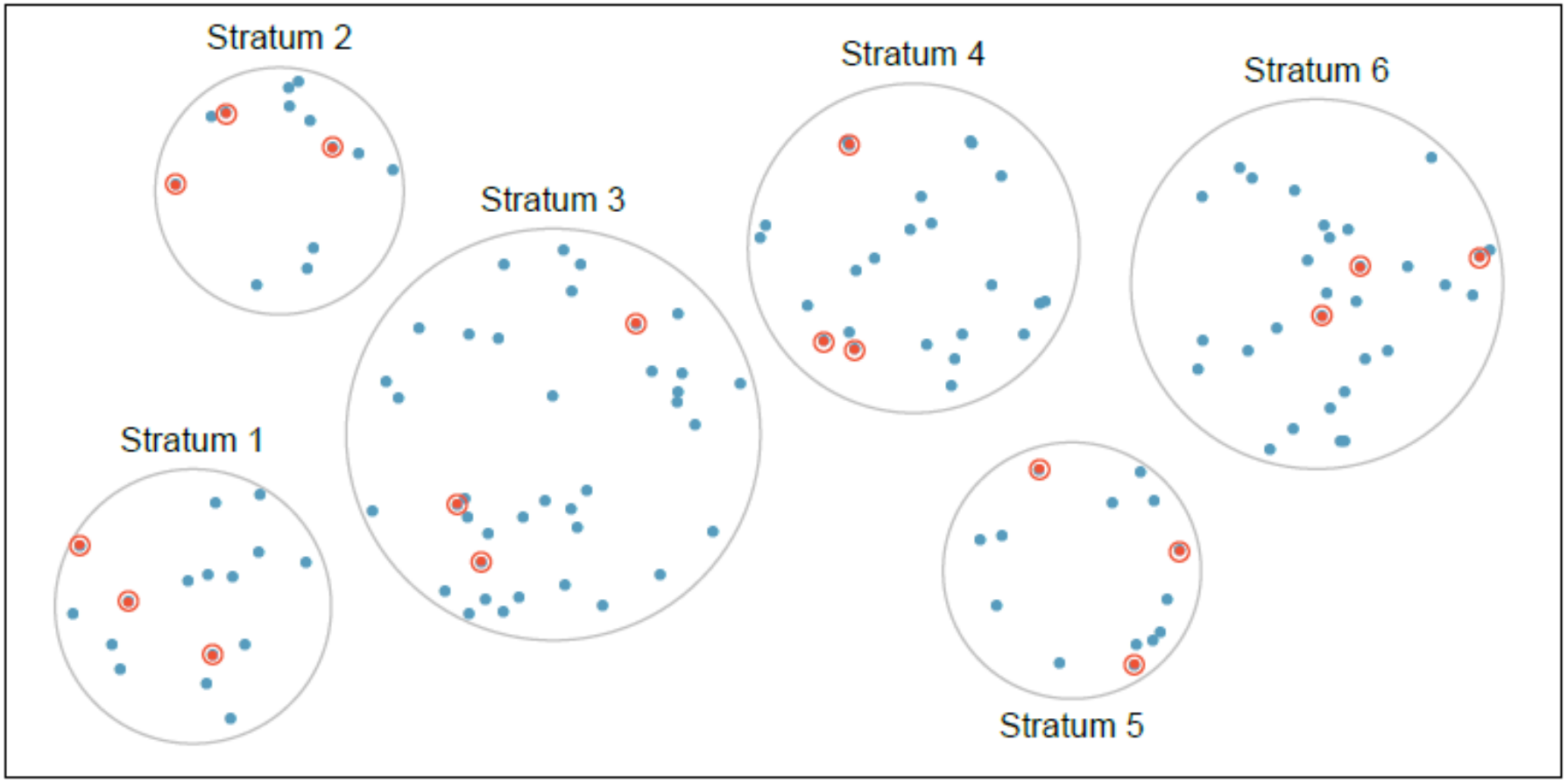




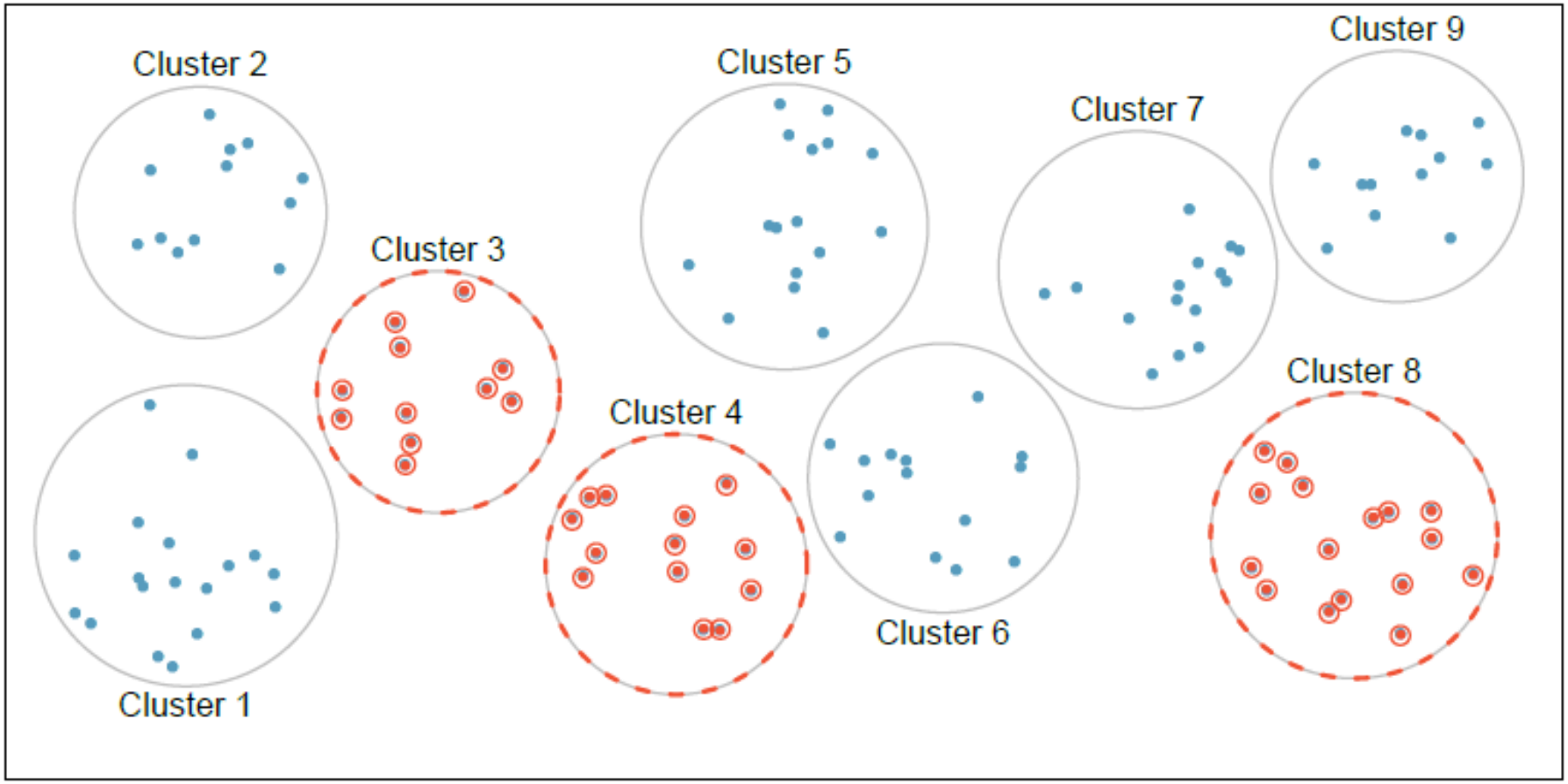
Simple Random Sampling



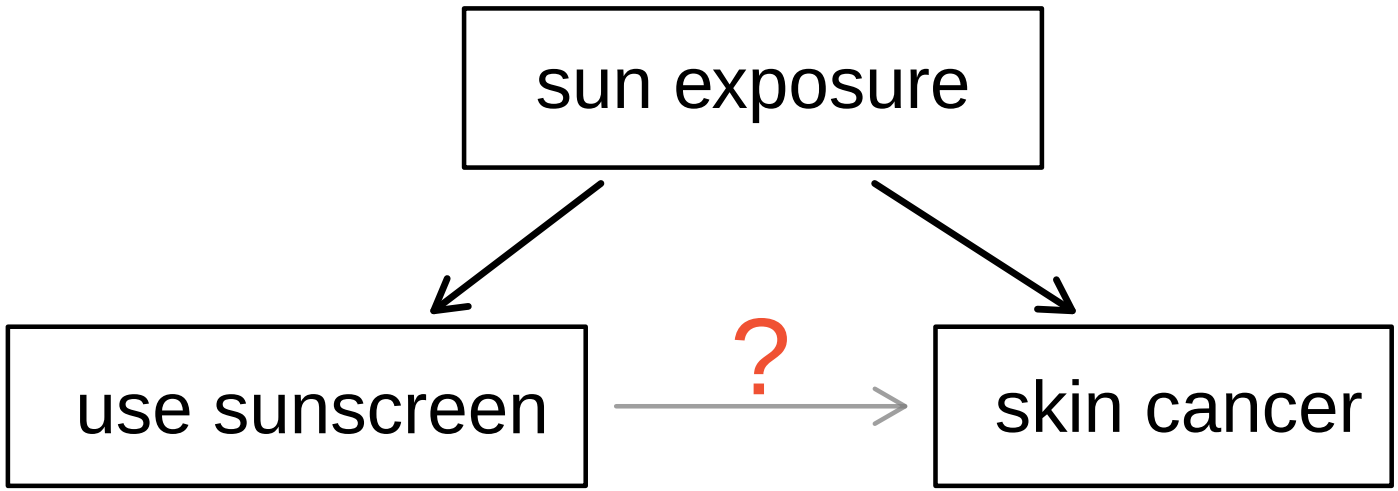
Stratified Sampling



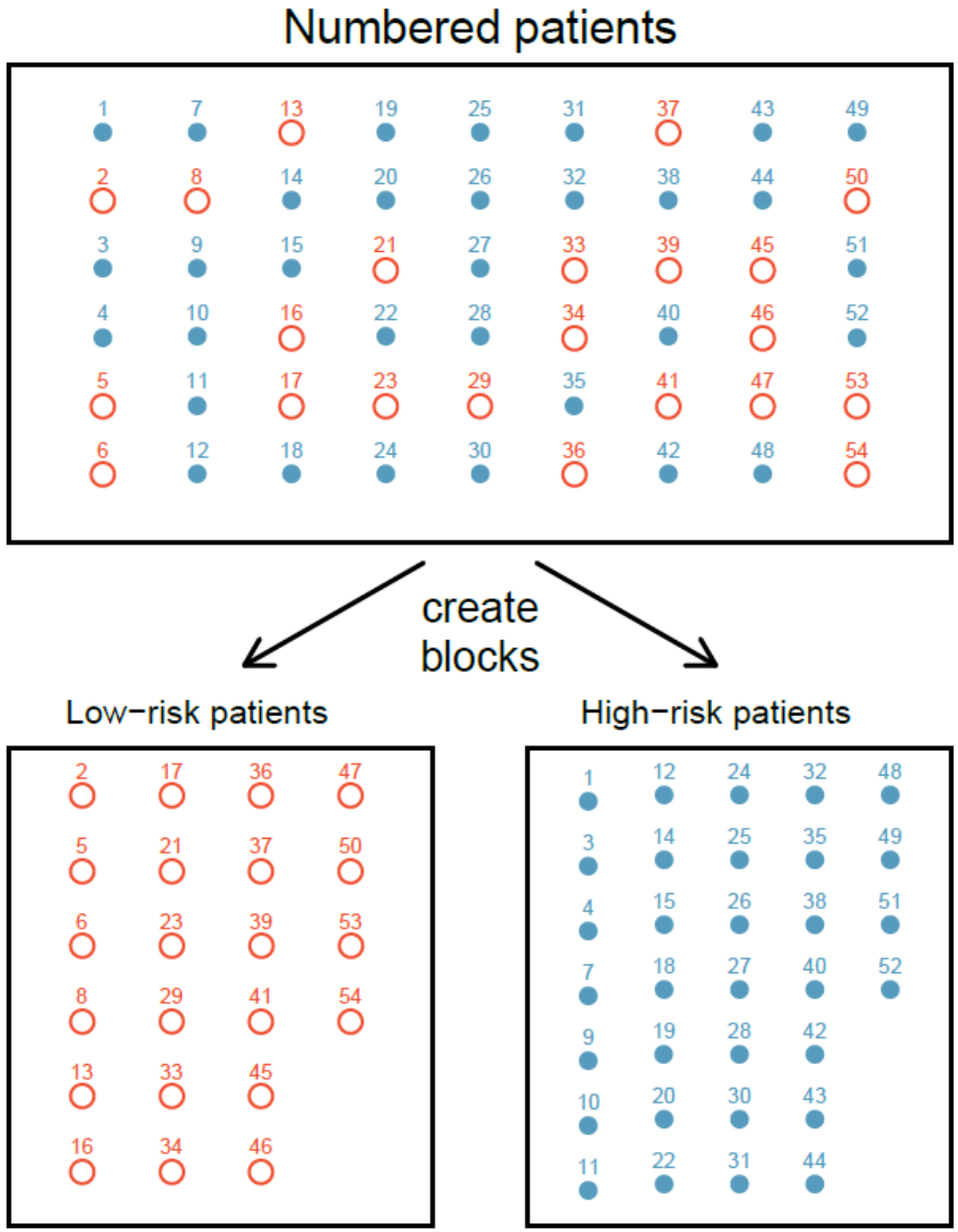
Cluster Sampling



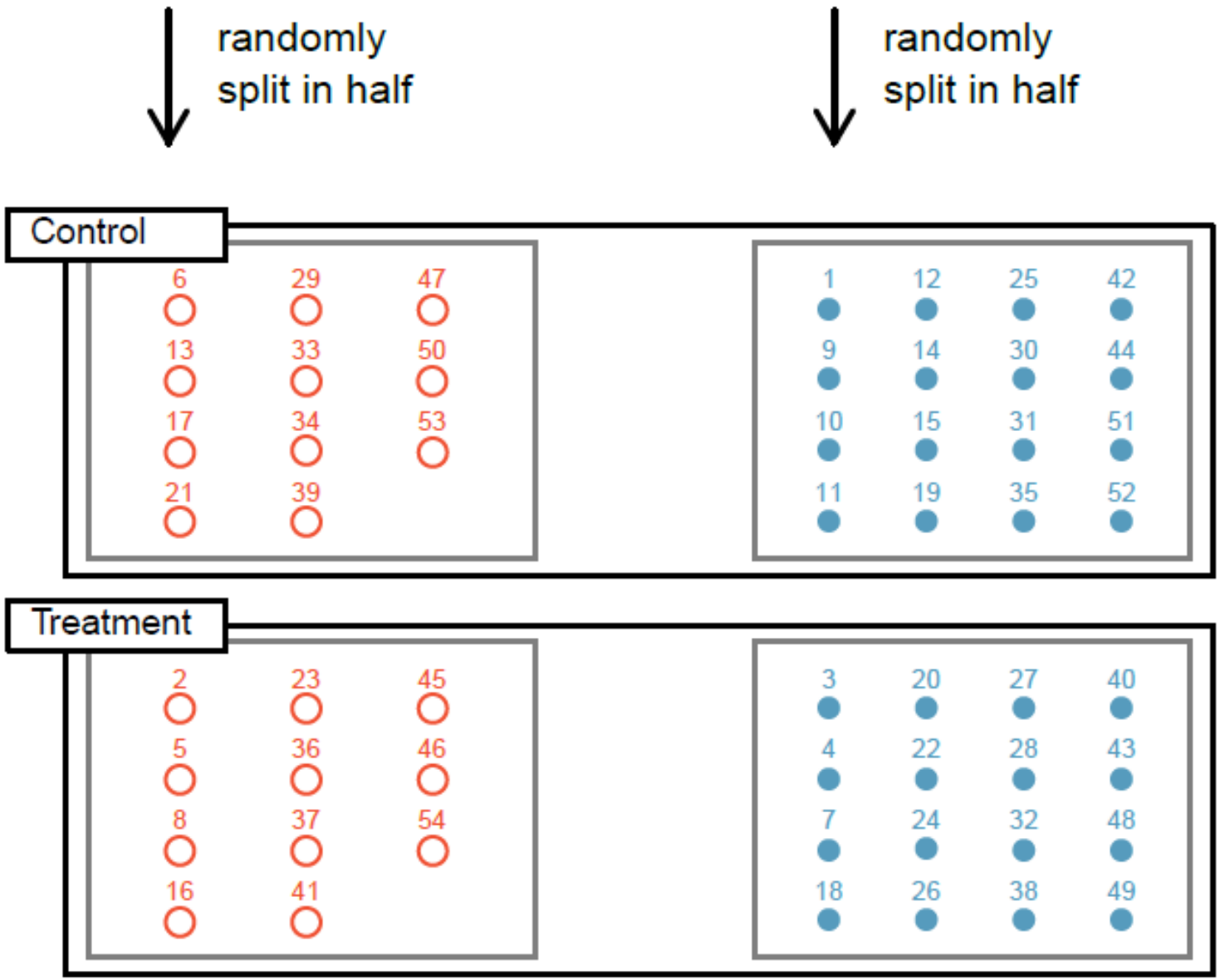
Confounding Factor or Lurking Variable



Experiment: Blocking



Randomized Experiments



Using Stents to Prevent Strokes

Chimowitz MI, Lynn MJ, Derdeyn CP, et al. 2011. **Stenting versus Aggressive Medical Therapy for Intracranial Arterial Stenosis.** *New England Journal of Medicine* 365:993-1003.

	0-30 days		0-365 days	
	stroke	no event	stroke	no event
treatment	33	191	45	179
control	13	214	28	199
Total	46	405	73	378

$$P(Stroke|treatment) = \frac{45}{224} = 20\%$$

$$P(Stroke|control = \frac{28}{227} = 12\%$$

US Bureau of Transportation Statistics (2016)

Arrivals to Four Destinations				
		Airline		
Count Column %		Delta	American	Total
Arrival	Delayed	659 (20%)	1,685 (22%)	2,344
	On Time	2,596 (80%)	5,966 (78%)	8,562
	Total	3,255	7,651	10,906

Simpson's Paradox

On Time %	Delta	American
Boston	80.1%	81.7%
Orlando	80.5%	84.5%
Philadelphia	70.5%	74.3%
San Diego	84.2%	85.4%

Airlines by Destination

Row %	Destination				
Arrival	**Boston**	**Orlando**	**Philadelphia**	**San Diego**	**Total**
Delta	1,409	1,168	312	366	3,255
	43%	36%	10%	11%	
American	1,826	970	4,423	432	7,651
	24%	13%	58%	6%	
Total	3,235	2,138	4,735	798	10,906