Part III	**Gathering Data**
Chapter 11	**Understanding Randomness**
What is it about random selection that makes it seem fair?	1. Nobody can guess the outcome in advance. 2. Outcomes are equally likely.
Random event/phenomenon	We know what outcomes could happen, but not which particular values will happen. Outcomes that we cannot predict but that nonetheless have a regular distribution in very many repetitions.
If our goal as statisticians is to uncover the truth about the world around us, then randomness is both . . .	our greatest enemy and our most important tool.
Simulation	A sequence of random outcomes that model a situation, often difficult to collect data on and with a mathematical answer hard to calculate. Models random events by using random numbers to specify event outcomes with relative frequencies that correspond to the true real-world relative frequencies we are trying to model. An artificial representation of a random process used to study its long-term properties.
Component	The most basic situation in a simulation in which something happens at random. [random happening]
Outcome	An individual result of a component [result of random happening]
Trial	The sequence of several components representing events that we are pretending will take place.
Response variable	The result of each trial with respect to what we were interested in.
Chapter 12	**Sample Surveys**
Population	The entire group of individuals or instances about whom we hope to learn, but examining all of them is usually impractical, if not impossible.
Sample	A (representative) subset of a population, examined in hope of learning about the population.
Sample survey	A study that asks questions of a sample drawn from some population in the hope of learning something about the entire population.(Polls)
Statistic They are written in _____	Any summary calculated from the (sampled) data. Latin ($\bar{x}$, s, r, b, $\hat{p}$)
Parameters They are written in _____	Key numbers in mathematical models used to represent reality. Greek (μ, σ, ρ, β, p)
Population parameter	A numerically valued attribute of a model for a population, often unknowable and estimated from sampled data.
Sample statistic	Correspond to, and thus estimate, a population parameter.
Representative Sample	Statistics computed from it accurately reflect the corresponding population parameters.
Bias	Any systematic failure of a sampling method to represent its

	population. It is almost impossible to recover from.
5 common bias errors:	1. Voluntary response – individuals can choose on their own whether to participate in the sample. Always yields invalid samples. 2. Convenience – when the sample is comprised of individuals readily available. Always yields a non-representative sample. 3. Undercoverage – when individuals from a subgroup of the population are selected less often than they should be. 4. Nonresponse – when a large fraction of those sampled will not or cannot respond. 5. Response – when respondents' answers might be affected by suvey design, such as question wording or interviewer behavior.
____ is often the best use of time and resources when sampling or surveying.	Reducing biases
Randomization	The best defense against bias. (stirring to make sure that on average the sample looks like the rest of the population)
Simple random sample (SRS)	A sample in which each set of n elements in the population has an equal chance of selection. The standard method of utilizing radomization to make the sample representative of the population of interest.
Sampling variability Sampling error	The natural tendency of randomly drawn samples to differ from each other.
The precision of the statistics of a sample depend on _____ not _____	the sample size (soup spoon) its fraction of the larger population.
Census	A sample that consists of the entire population.
Sampling frame	A list of individuals, which clearly defines but may not be representative of the entire population, from which the sample is drawn.
Stratified samples	These samples can reduce sampling variability by identifying homogeneous subgroups and then randomly sampling within each.
Cluster samples	These samples randomly select among heterogeneous subgroups that each resemble the population at large, making our sampling tasks more manageable.
Systematic samples	These samples can work, when there is no relationship between the order of the sampling frame and the variables of interest, and are often the least expensive method of sampling. But we still want to start them randomly.
Multistage sample	A sampling scheme that combines several sampling methods.
Identify the W's: Why What Who When, Where, and How /* previously Who < What */	 Population and associated sampling frame. Parameter of interest and variables measured. Sample actually drawn. Given by the sampling plan.

Chapter 13	**Experiments and Observational Studies**
Observational study	A study based on data in which no manipulation of factors has been employed (researchers don't assign choices). Usually focuses on estimating differences between groups but is not possible to demonstrate a causal relationship. Often used when an experiment is impractical.
Retrospective	Subjects are selected and then their previous conditions or behaviors are determined.
Prospective	Subjects are followed to observe future outcomes. No treatments are deliberately applied.
To prove a cause-and-effect relationship we need to perform _____	a valid experiment.
An experiment _____ to create treatments, _____ to these treatment levels, and then _____ across treatment levels.	manipulates factor levels randomly assigns subjects compares the responses of the subject groups (boxplots are often a good choice for displaying results of groups)
Factor	A variable whose levels are controlled by the experimenter.
Level	The specific values that the experimenter chooses for a factor.
Treatment	The process, intervention, or other controlled circumstance applied to randomly assigned experimental units. Treatments are the different levels of a single factor or are made up of combinations of levels of two or more factors.
_____ are individuals on whom an experiment is performed. Usually called _____ or _____ when human.	Experimental units Subjects Participants
Response	A variable whose values are compared across different treatments.
The 4 principals of experimental design:	1. **Control** sources of variation other than the factors we are testing by making conditions as similar as possible for all treatment groups. 2. **Randomize** subjects to treatments to even out effects that we cannot control. 3. **Replicate** over as many subjects as possible. Would like to get results from a representative sample of the population of interest. 4. **Block** and then randomize within to reduce the effects of identifiable attributes of the subjects that cannot be controlled.
Control group	The experimental units assigned to a baseline treatment level, typically either the default treatment, which is well understood, or a null, placebo treatment. Their responses provide a basis for comparison.
Statistically significant	When an observed difference is too large for us to believe that it is likely to have occurred naturally (only by chance).
Placebo	A (fake) treatment known to have no effect, administered so that all groups experience the same conditions.
Placebo effect	The tendency of many human subjects (often 20% or more of experimental subjects) to show a response even when administered a

	placebo.
Blinding	Individuals associated with an experiment are not aware of how subjects have been allocated to treatment groups.
2 main classes of individuals who can affect the outcome of an experiment: Single-blind Double-blind	1. those who could influence the results (subjects, treatment administrators, or technicians) 2. those who evaluate the results (judges, treating physicians, etc.) When every individual in ***either*** of these classes is blinded. When everyone in ***both*** classes is blinded.
Block	Same idea for experiments as stratifying is for sampling. Group together subjects that are similar and randomize within those groups as a way to remove unwanted variation (of the differences between the groups so that we can see the differences caused by the treatments more clearly) (Doing parallel experiments on different groups.)
Matching	In a retrospective or prospective study, subjects who are similar in ways not under study may be paired and then compared with each other on the variables of interest as a way to reduce unwanted variation in much the same way as blocking.
Designs: Randomized block design Completely random design	 The randomization occurs only within blocks. All experimental units have an equal chance of receiving any particular treatment.
The best experiments are usually:	Randomized, comparative, double-blind, placebo-controlled.
Lurking (Confounding) variables are outside influences that make it _____ we are modeling with _____	 harder to understand the relationship regression and observational studies (a designed experiment).
Lurking variable	Creates an association between two other variables that tempts us to think that one may cause the other. [regression analysis or observational study]
Confounding	Some other variable associated with a factor has an effect on the response variable. [experiments] Arises when the response we see in an experiment is at least partially attributable to uncontrolled variables.