Essential Empirical Methods

AF

June 9, 2019

Outline

- Part 1: Concept Measurement and Construction
- Part 2: Describing Data
- Part 3: Hypotheses and Making Comparisons

Part 1: Concept Measurement and Construction

Concept Construction: Objectives

- This class aims to improve the analyst's ability to define, construct, and measure concepts

Concept Construction: Objectives

- This class aims to improve the analyst's ability to define, construct, and measure concepts
- Analysts will not only be more certain about the concepts they create and what constitutes them, but analysts will also be more certain that they are collecting data against these concepts validly and reliably

Concepts as Phenomena

- All research begins with a focus on phenomena, such as wars, political regimes, coups, or military tactics

Concepts as Phenomena

- All research begins with a focus on phenomena, such as wars, political regimes, coups, or military tactics
- These phenomena are **concepts** which represent something in the real world

Concepts as Phenomena

- All research begins with a focus on phenomena, such as wars, political regimes, coups, or military tactics
- These phenomena are **concepts** which represent something in the real world
- Our first challenge is figuring out what *exactly* constitutes our concept

Concepts as Phenomena

- All research begins with a focus on phenomena, such as wars, political regimes, coups, or military tactics
- These phenomena are **concepts** which represent something in the real world
- Our first challenge is figuring out what *exactly* constitutes our concept
- After all, we need to know what we are looking for and what cases our concept applies to

Concepts: A Constitutive Approach

- Say for instance we are interested in studying **democracies**

Concepts: A Constitutive Approach

- Say for instance we are interested in studying **democracies**
- What constitutes a democracy?

Concepts: A Constitutive Approach

- Say for instance we are interested in studying **democracies**
- What constitutes a democracy?
- To be a democracy is to allow citizens to vote, parties to compete, and checks-and-balances of accountability to exist

Concepts: A Constitutive Approach

- Say for instance we are interested in studying **democracies**
- What constitutes a democracy?
- To be a democracy is to allow citizens to vote, parties to compete, and checks-and-balances of accountability to exist
- To **be** a *democracy* **is** to have these features

Concepts Building: A Basic Approach

- Taking a constitutive approach to concepts, this first section provides a 3 step process to building concepts

Concepts Building: A Basic Approach

- Taking a constitutive approach to concepts, this first section provides a 3 step process to building concepts
- Later, we will review some advanced considerations

Concept Construction: Step 1

- **Step 1**: Clarify the features or empirical meaning of the concept

Concept Construction: Step 1

- **Step 1**: Clarify the features or empirical meaning of the concept

- Example: We are interested in determining whether or not women are more **liberal** than men

Concept Construction: Step 1

- **Step 1**: Clarify the features or empirical meaning of the concept
- Example: We are interested in determining whether or not women are more **liberal** than men
- To evaluate this question, we must first know what **liberalism** means before we can collect and examine data

Concept Construction: Step 1

- **Step 1**: Clarify the features or empirical meaning of the concept
- Example: We are interested in determining whether or not women are more **liberal** than men
- To evaluate this question, we must first know what **liberalism** means before we can collect and examine data
- To clarify the features of a concept is to create a list of properties associated with the concept as well as a list of properties associated with the concepts reverse or opposite

Concept Construction: Step 1

- **Step 1**: Clarify the features or empirical meaning of the concept
- Example: We are interested in determining whether or not women are more **liberal** than men
- To evaluate this question, we must first know what **liberalism** means before we can collect and examine data
- To clarify the features of a concept is to create a list of properties associated with the concept as well as a list of properties associated with the concepts reverse or opposite
- We do this because we must know what our concept **is** and what our concept **is not**

Concept Construction: Step 1

- We can use a brainstorming heuristic to help us come up with the features, or concrete measurable properties, associated with our concept

Concept Construction: Step 1

- We can use a brainstorming heuristic to help us come up with the features, or concrete measurable properties, associated with our concept

- First, imagine an ideal type of your concept. An ideal type is a perfect encapsulation of your concept which may not exist in the real world, but if it did, you would instantly identify and label it as your concept.

Concept Construction: Step 1

- We can use a brainstorming heuristic to help us come up with the features, or concrete measurable properties, associated with our concept
- First, imagine an ideal type of your concept. An ideal type is a perfect encapsulation of your concept which may not exist in the real world, but if it did, you would instantly identify and label it as your concept.
- Next, imagine its (ideal) reverse or opposite

Concept Construction: Step 1

- We can use a brainstorming heuristic to help us come up with the features, or concrete measurable properties, associated with our concept
- First, imagine an ideal type of your concept. An ideal type is a perfect encapsulation of your concept which may not exist in the real world, but if it did, you would instantly identify and label it as your concept.
- Next, imagine its (ideal) reverse or opposite
- Then, start listing concrete and measurable traits associated with the ideal type

Concept Construction: Step 1

- Let's apply this brainstorming method to the concept we want to figure out: liberalism

Concept Construction: Step 1

- Let's apply this brainstorming method to the concept we want to figure out: liberalism

- We can imagine a very liberal person as taking on our ideal type. To ground our thinking even further, we might imagine Supreme Court Justices such as Justice Ginsburg or Justice Sotomayor.

Concept Construction: Step 1

- Let's apply this brainstorming method to the concept we want to figure out: liberalism
- We can imagine a very liberal person as taking on our ideal type. To ground our thinking even further, we might imagine Supreme Court Justices such as Justice Ginsburg or Justice Sotomayor.
- The opposite of a very liberal person would be a very conservative person. Again, we might imagine Justice Thomas or Justice Alito.

Concept Construction: Step 1

- Let's apply this brainstorming method to the concept we want to figure out: liberalism
- We can imagine a very liberal person as taking on our ideal type. To ground our thinking even further, we might imagine Supreme Court Justices such as Justice Ginsburg or Justice Sotomayor.
- The opposite of a very liberal person would be a very conservative person. Again, we might imagine Justice Thomas or Justice Alito.
- Now, what ideological traits might these Justices exhibit?

Concept Construction: Step 1

Starting with our concept of interest, list its first measurable trait

Liberal	**Conservative**
Supports gov-funded health care	

Concept Construction: Step 1

Next, list the first measurable trait's opposite

Liberal	**Conservative**
Supports gov-funded health care	Opposes gov-funded health care

Concept Construction: Step 1

Repeat

Liberal	**Conservative**
Supports gov-funded health care	Opposes gov-funded health care
Opposes tax cuts	Support tax cuts

Concept Construction: Step 1

Repeat

Liberal	**Conservative**
Supports gov-funded health care	Opposes gov-funded health care
Opposes tax cuts	Support tax cuts
Opposes restrictions on abortion	Supports restrictions on abortion

Concept Construction: Step 1

Repeat

Liberal	**Conservative**
Supports gov-funded health care	Opposes gov-funded health care
Opposes tax cuts	Support tax cuts
Opposes restrictions on abortion	Supports restrictions on abortion
Supports restrictions on guns	Opposes restrictions on guns

Concept Construction: Clarifying Concrete Features

Why are these entries not concrete properties associated with Liberalism/Conservatism?

Liberal	**Conservative**
Opposes free markets	Supports free markets

Concept Construction: Step 2

- **Step 2**: Create dimensions by grouping like features

Concept Construction: Step 2

- **Step 2**: Create dimensions by grouping like features
- Dimensions help us better understand the generalities that constitute our concept while also simplifying its communication

Concept Construction: Step 2

- **Step 2**: Create dimensions by grouping like features
- Dimensions help us better understand the generalities that constitute our concept while also simplifying its communication
- Instead of claiming that a lengthy table constitutes our concept, we can say that a few dimensions which capture the essence of our table, constitutes our concept

Concept Construction: Step 2

- **Step 2**: Create dimensions by grouping like features
- Dimensions help us better understand the generalities that constitute our concept while also simplifying its communication
- Instead of claiming that a lengthy table constitutes our concept, we can say that a few dimensions which capture the essence of our table, constitutes our concept
- Dimensions are created by noticing patterns in our concept's concrete properties (Step 1)

Concept Construction: Step 2

- To create dimensions, we look for patterns in our concept's features

Liberal	**Conservative**
Supports gov-funded health care	Opposes gov-funded health care
Opposes tax cuts	Support tax cuts
Opposes restrictions on abortion	Supports restrictions on abortion
Supports restrictions on guns	Opposes restrictions on guns

Concept Construction: Step 2

- To create dimensions, we look for patterns in our concept's features
- Notice that *opposition to tax cuts* and *support for government health care* are *economic* traits

Liberal	**Conservative**
Supports gov-funded health care	Opposes gov-funded health care
Opposes tax cuts	Support tax cuts
Opposes restrictions on abortion	Supports restrictions on abortion
Supports restrictions on guns	Opposes restrictions on guns

Concept Construction: Step 2

- To create dimensions, we look for patterns in our concept's features
- Notice that *opposition to tax cuts* and *support for government health care* are *economic* traits
- We can now say that these features tell us something more general about our concept: **liberalism** consists of **economic liberalism**

Liberal	**Conservative**
Supports gov-funded health care	Opposes gov-funded health care
Opposes tax cuts	Support tax cuts
Opposes restrictions on abortion	Supports restrictions on abortion
Supports restrictions on guns	Opposes restrictions on guns

Concept Construction: Step 2

- Also notice that *opposition to abortion restrictions* and *support for gun restrictions* are *social* traits

Liberal	**Conservative**
Supports gov-funded health care	Opposes gov-funded health care
Opposes tax cuts	Support tax cuts
Opposes restrictions on abortion	Supports restrictions on abortion
Supports restrictions on guns	Opposes restrictions on guns

Concept Construction: Step 2

- Also notice that *opposition to abortion restrictions* and *support for gun restrictions* are *social* traits

- We can also say that these features contribute to a more general understanding of our concept: **liberalism** consists of **social liberalism**

Liberal	**Conservative**
Supports gov-funded health care	Opposes gov-funded health care
Opposes tax cuts	Support tax cuts
Opposes restrictions on abortion	Supports restrictions on abortion
Supports restrictions on guns	Opposes restrictions on guns

Concept Construction: Step 2

- Instead of using a lengthy table of features to describe what our concept consists of, we can now say that two dimensions constitute **liberalism**: *economic liberalism* and *social liberalism*

Concept Construction: Step 2

- Instead of using a lengthy table of features to describe what our concept consists of, we can now say that two dimensions constitute **liberalism**: *economic liberalism* and *social liberalism*

- We can now communicate this simplification in a definition:

Concept Construction: Step 2

- Instead of using a lengthy table of features to describe what our concept consists of, we can now say that two dimensions constitute **liberalism**: *economic liberalism* and *social liberalism*
- We can now communicate this simplification in a definition:
- The concept of *liberalism* is defined as the extent to which *individuals* exhibit the characteristic of *economic* and *social liberalism*.

Concept Construction: Step 2

- Instead of using a lengthy table of features to describe what our concept consists of, we can now say that two dimensions constitute **liberalism**: *economic liberalism* and *social liberalism*
- We can now communicate this simplification in a definition:
- The concept of *liberalism* is defined as the extent to which *individuals* exhibit the characteristic of *economic* and *social liberalism*.
- The reader now knows what the concept is, what units/cases it applies to, and generally what traits it consists of

Concept Construction: Definition Template

When writing concept definitions, an excellent template that will not lead you astray is the following:

- The concept of [*the variation within a measurable concept*] is defined as the extent to which [*the unit of analysis to which the concept applies*] exhibits the characteristic of [*how the characteristic is measured*]

Concept Construction: Step 3

- **Step 3**: Attach concept features (indicators) to its dimensions

Concept Construction: Step 3

- **Step 3**: Attach concept features (indicators) to its dimensions
- The final step involves simply re-arranging the work we have already done in somewhat of a reverse order

Concept Construction: Step 3

- **Step 3**: Attach concept features (indicators) to its dimensions
- The final step involves simply re-arranging the work we have already done in somewhat of a reverse order
- We identified our concept (step 1), then we created a table of its concrete measurable traits, or indicators (step 3), and then we grouped these concrete traits into more general categories to simplify (step 2)

Concept Construction: Step 3

- **Step 3**: Attach concept features (indicators) to its dimensions
- The final step involves simply re-arranging the work we have already done in somewhat of a reverse order
- We identified our concept (step 1), then we created a table of its concrete measurable traits, or indicators (step 3), and then we grouped these concrete traits into more general categories to simplify (step 2)
- The important step here is ensuring the concrete features of our dimensions are aligned correctly such that if we collect information about an individual's opposition to tax cuts, we count this as evidence of *economic* and not *social* liberalism

Concept Construction: Step 3

Table: Liberalism: Dimensions and Indicators

Economic Liberalism	**Social Liberalism**
Supports gov-funded health care	Opposes restrictions on abortion
Opposes tax cuts	Supports restrictions on guns

Review

Advanced Considerations

Concept Construction: Components

Review: There are three building blocks involved in concept construction:

- **Concept** - Nouns or nouns with adjectives. We may be interested in battles, or perhaps *naval* battles, *aerial* battles, or *land* battles

Concept Construction: Components

Review: There are three building blocks involved in concept construction:

- **Concept** - Nouns or nouns with adjectives. We may be interested in battles, or perhaps *naval* battles, *aerial* battles, or *land* battles

- **Dimensions** - Reflect the theoretical and constitutive features of the concept in a generalized way

$$\text{Economic Liberalism} = \begin{cases} \text{Supports gov-funded health care} \\ \text{Opposes tax cuts} \end{cases}$$

Concept Construction: Components

Review: There are three building blocks involved in concept construction:

- **Concept** - Nouns or nouns with adjectives. We may be interested in battles, or perhaps *naval* battles, *aerial* battles, or *land* battles

- **Dimensions** - Reflect the theoretical and constitutive features of the concept in a generalized way

$$\text{Economic Liberalism} = \begin{cases} \text{Supports gov-funded health care} \\ \text{Opposes tax cuts} \end{cases}$$

- **Indicators** - Measurable characteristics of dimensions that are specific enough such that we can gather data against them

Advanced Considerations: Concepts

What is the relationship between the concept's positive and negative ideal type? The relationship is always theoretical, but it is something we should consider

- The relationship can be viewed in two ways: dichotomous or continuous. You will find most concepts can be constructed in either way.

Advanced Considerations: Concepts

What is the relationship between the concept's positive and negative ideal type? The relationship is always theoretical, but it is something we should consider

- The relationship can be viewed in two ways: dichotomous or continuous. You will find most concepts can be constructed in either way.
- Take our liberalism example. We can readily build a dichotomous concept, such that we classify individuals as liberal or conservative. We can also build a continuous concept, such that individuals move along a spectrum of liberal-conservative depending on their mix of left-right policy preferences.

Advanced Considerations: Concepts

What is the relationship between the concept's positive and negative ideal type? The relationship is always theoretical, but it is something we should consider

- The relationship can be viewed in two ways: dichotomous or continuous. You will find most concepts can be constructed in either way.
- Take our liberalism example. We can readily build a dichotomous concept, such that we classify individuals as liberal or conservative. We can also build a continuous concept, such that individuals move along a spectrum of liberal-conservative depending on their mix of left-right policy preferences.
- While continuous concepts are richer in precision, dichotomous concepts are significantly easier to build, manage, and observe

Advanced Considerations: Dimensions

After specifying our dimensions, we must specify our concept's structure such that it is clear which dimensions are necessary and sufficient

- **Necessary and Sufficient** - Sufficiency by necessity. This structure is implicit in the basic concept development guide. If we do not observe economic and social liberalism, we also do not observe our concept liberalism. This structure generates strong *intension* but weak *extension*.

Advanced Considerations: Dimensions

After specifying our dimensions, we must specify our concept's structure such that it is clear which dimensions are necessary and sufficient

- **Necessary and Sufficient** - Sufficiency by necessity. This structure is implicit in the basic concept development guide. If we do not observe economic and social liberalism, we also do not observe our concept liberalism. This structure generates strong *intension* but weak *extension*.

- **Family Resemblance** - Sufficiency without necessity. Concepts have no common (necessary) feature because one dimension can step in for the absence of others. The nature of our work pushes us towards using this structure because it has weak *intension* but strong *extension*.

Advanced Considerations: Dimensions

We specify that our dimensions follow a necessary and sufficient structure by including the word "and"

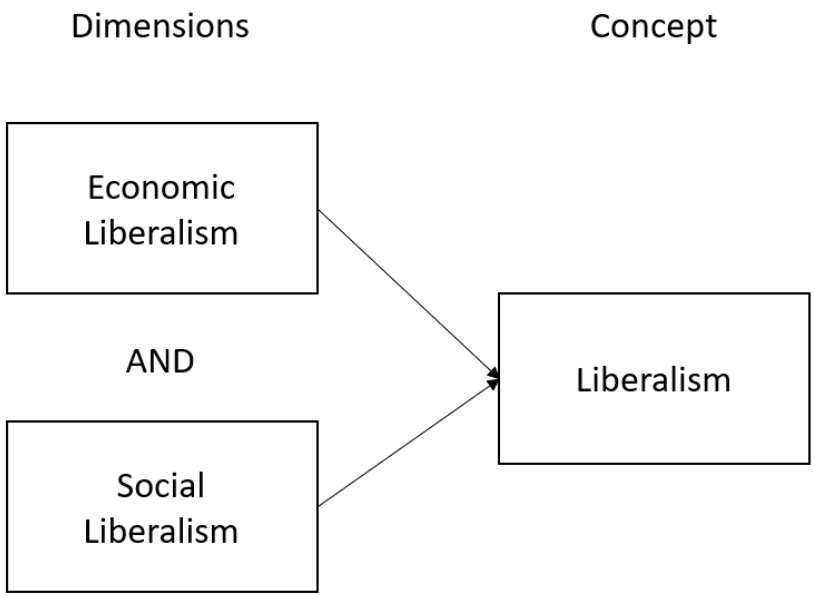


Figure: Dimension Structure: Necessary Conditions

Advanced Considerations: Dimensions

We specify that our dimensions follow a family resemblance structure by including the word "or"

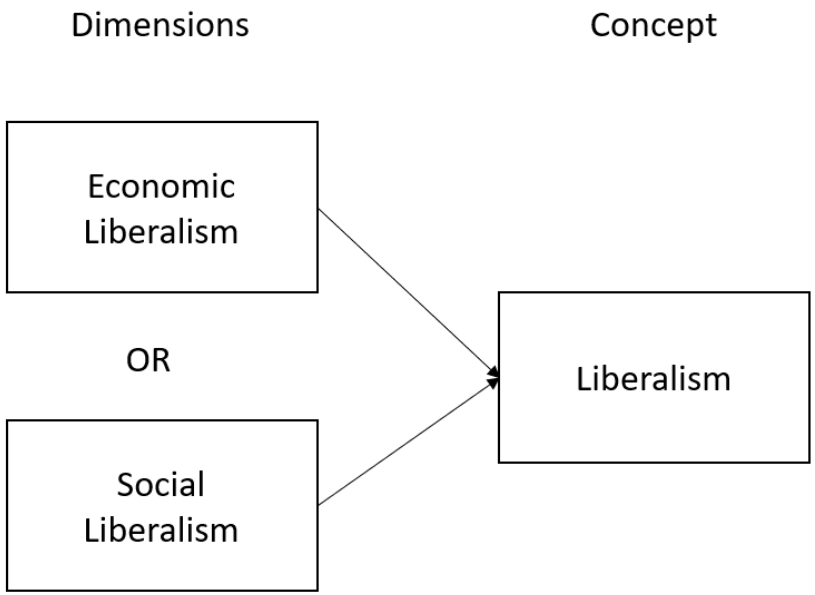


Figure: Dimension Structure: Family Resemblance

Advanced Considerations: Dimensions

We can also assign numerical weights to our dimensions that emphasize the relative importance of one dimension over another

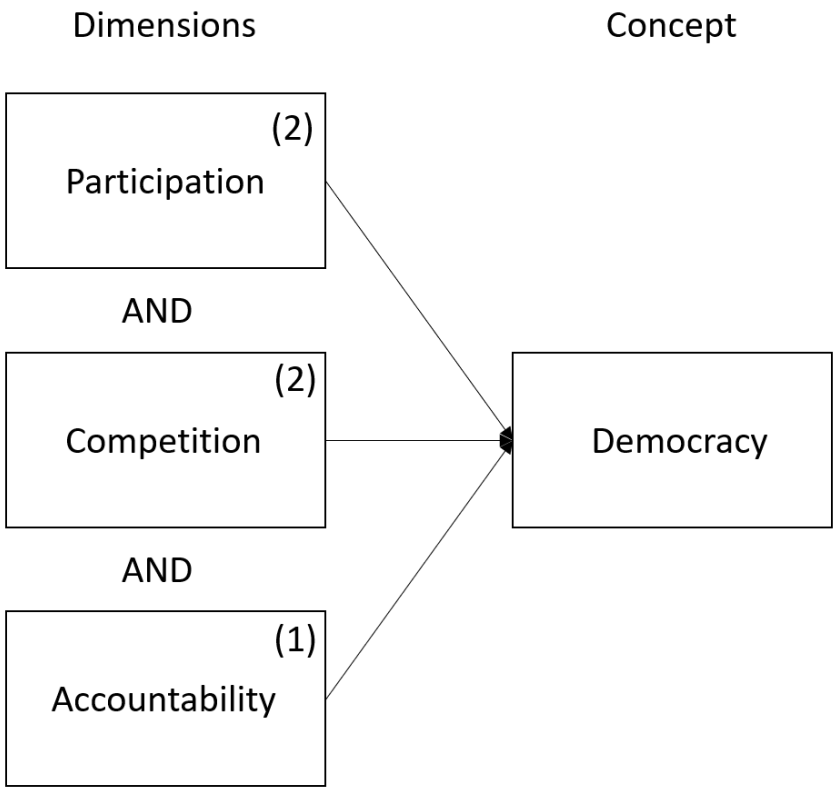


Figure: Dimension Structure: Weights

Advanced Considerations: Indicators

Indicators can also follow the necessary and sufficient or family resemblance structure

Indicators can also be weighted

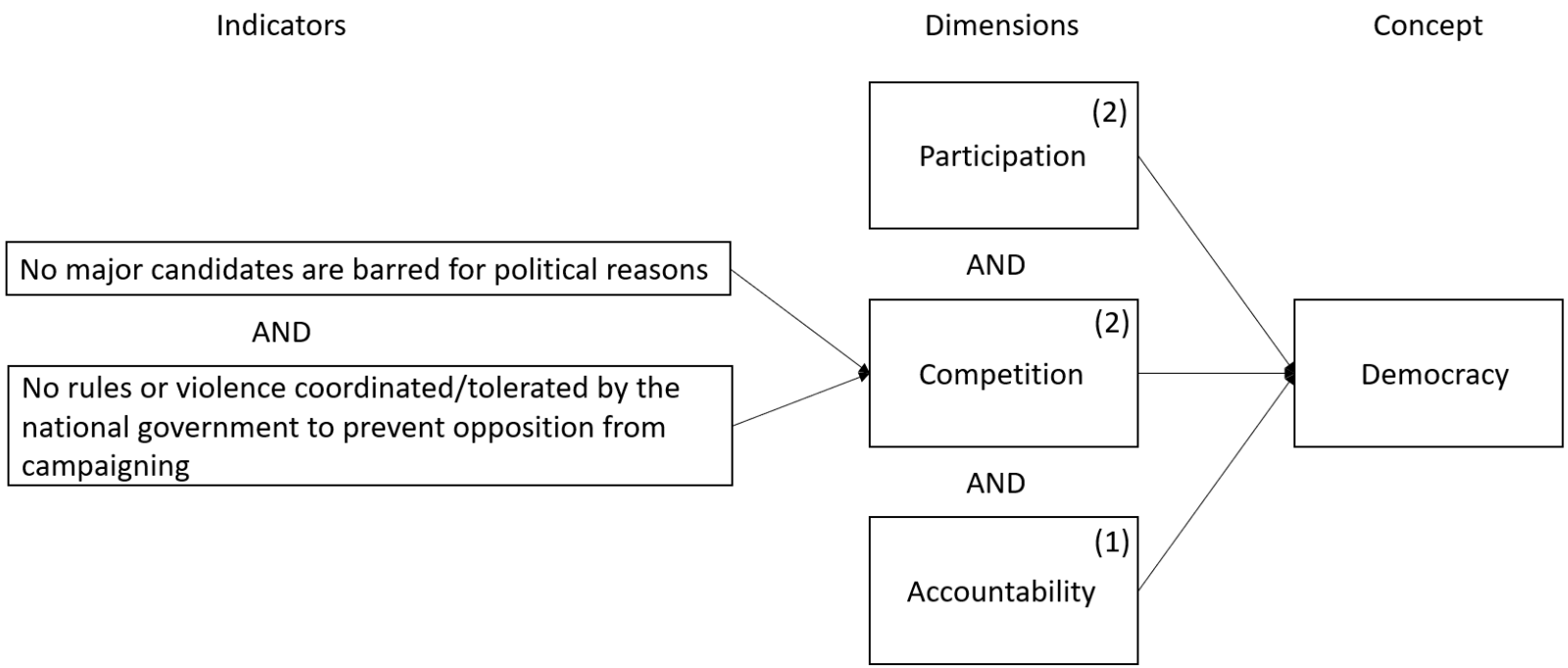


Figure: Indicator Structure and Weights

Advanced Considerations: Aggregation

- To generate a score for our concept given the evidence, we aggregate in specific ways depending on our concept structure

Advanced Considerations: Aggregation

- To generate a score for our concept given the evidence, we aggregate in specific ways depending on our concept structure

- **Family Resemblance** - For each dimension, we add the number of indicators found giving each dimension its own score. We then add each dimension to produce the concept's overall score. Alternatives including taking the max of the indicators.

Advanced Considerations: Aggregation

- To generate a score for our concept given the evidence, we aggregate in specific ways depending on our concept structure

- **Family Resemblance** - For each dimension, we add the number of indicators found giving each dimension its own score. We then add each dimension to produce the concept's overall score. Alternatives including taking the max of the indicators.

- **Necessary-Sufficient** - For each dimension, we add the number of indicators. For our concept's overall score, we take the minimum score across the dimensions. The logic here is that the overall score is as strong as its weakest link.

Advanced Considerations: Aggregation

Welfare states coded on a scale from 0 to 4; the more indicators present, the greater the welfare state

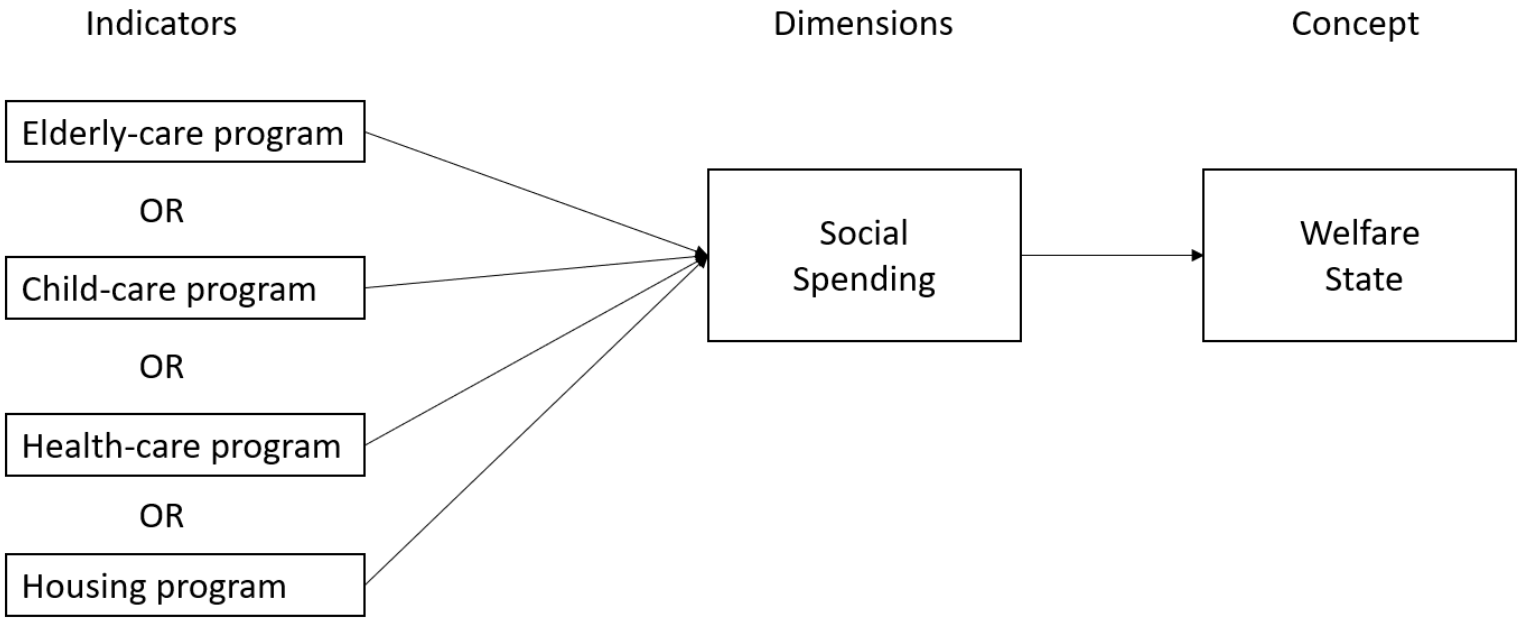


Figure: Family Resemblance Aggregation

Review

Part 2: Describing Data

Describing Data: Objectives

- This class aims to improve the analyst's ability to summarize and interpret large amounts of information, admittedly at the cost of detail

Describing Data: Objectives

- This class aims to improve the analyst's ability to summarize and interpret large amounts of information, admittedly at the cost of detail
- This summation process is a simplified version/beginning of a more complicated discussion on **descriptive inference**

Describing Data: Objectives

- This class aims to improve the analyst's ability to summarize and interpret large amounts of information, admittedly at the cost of detail
- This summation process is a simplified version/beginning of a more complicated discussion on **descriptive inference**
- Ultimately what we are doing in this class is learning how to summarize the facts we know to learn more about the facts we do not know

Describing Data: Objectives

- This class aims to improve the analyst's ability to summarize and interpret large amounts of information, admittedly at the cost of detail
- This summation process is a simplified version/beginning of a more complicated discussion on **descriptive inference**
- Ultimately what we are doing in this class is learning how to summarize the facts we know to learn more about the facts we do not know
- Analysts will be able to summarize information on a basic but powerfully adequate level, thereby allowing them to think more broadly about the trends (or outliers) they can observe and what this might mean for the trends they cannot observe

Describing Data: Summarizing Data

- Now that we have created a variable and collected data against it, our next task is to summarize it

Describing Data: Summarizing Data

- Now that we have created a variable and collected data against it, our next task is to summarize it
- Simplification has been an integral part of every known scholarly work and will probably always be: There is no choice but to simplify

Describing Data: Summarizing Data

- Now that we have created a variable and collected data against it, our next task is to summarize it
- Simplification has been an integral part of every known scholarly work and will probably always be: There is no choice but to simplify
- We will go over ways in which we can interpret simple statistics to help us understand our data

Describing Data: Summarizing Data

- Now that we have created a variable and collected data against it, our next task is to summarize it
- Simplification has been an integral part of every known scholarly work and will probably always be: There is no choice but to simplify
- We will go over ways in which we can interpret simple statistics to help us understand our data
- We use and interpret statistics because its almost always easier than understanding our entire dataset

Describing Data: Variables

- The concept construction process yields a variable

Describing Data: Variables

- The concept construction process yields a variable
- A variable is an empirical measurement of a concept which we use to describe and analyze the world

Describing Data: Variables

- The concept construction process yields a variable
- A variable is an empirical measurement of a concept which we use to describe and analyze the world
- A variable has one name, at least two values, and often has numeric codes for use in computer analysis

Describing Data: Variables

- The concept construction process yields a variable
- A variable is an empirical measurement of a concept which we use to describe and analyze the world
- A variable has one name, at least two values, and often has numeric codes for use in computer analysis
- Before summarizing our data, we must first decide which type of variable we have

Describing Data: Unordered Categorical

- A unordered categorical variable offers the least precision because it simply describes differences between units that otherwise have no ordered difference between them.

Describing Data: Unordered Categorical

- A unordered categorical variable offers the least precision because it simply describes differences between units that otherwise have no ordered difference between them.

- Examples include: `gender` (male/female), `religion` (Catholic, Protestant, Muslim), `region` (South, North, West), and `martial status` (single, married, divorced).

Describing Data: Unordered Categorical

- A unordered categorical variable offers the least precision because it simply describes differences between units that otherwise have no ordered difference between them.

- Examples include: `gender` (male/female), `religion` (Catholic, Protestant, Muslim), `region` (South, North, West), and `martial status` (single, married, divorced).

- Put quantitatively, we might code `gender` in the following way: male $= 0$, female $= 1$. Alternatively, we might code `gender` as: male $= 33$, female $= 12$. The numbers have no substantive meaning other than they distinguish a difference.

Describing Data: Ordered Categorical

- A ordered categorical variable offers increasing precision because it describes relative differences between units in such a way that they can be ranked.

Describing Data: Ordered Categorical

- A ordered categorical variable offers increasing precision because it describes relative differences between units in such a way that they can be ranked.

- For example, if we wanted to estimate voter preferences for gun control, we might ask them how strongly they felt about the issue by offering four categories to choose from: strongly oppose, oppose, support, strongly support.

Describing Data: Ordered Categorical

- A ordered categorical variable offers increasing precision because it describes relative differences between units in such a way that they can be ranked.

- For example, if we wanted to estimate voter preferences for gun control, we might ask them how strongly they felt about the issue by offering four categories to choose from: strongly oppose, oppose, support, strongly support.

- Viewed quantitatively, the numbers matter for an ordered categorical variable. We might code `gun control` in the following way: strongly oppose $= 0$, oppose $= 1$, support $= 2$, and strongly support $= 3$.

Describing Data: Ordered Categorical

- A ordered categorical variable offers increasing precision because it describes relative differences between units in such a way that they can be ranked.

- For example, if we wanted to estimate voter preferences for gun control, we might ask them how strongly they felt about the issue by offering four categories to choose from: strongly oppose, oppose, support, strongly support.

- Viewed quantitatively, the numbers matter for an ordered categorical variable. We might code `gun control` in the following way: strongly oppose $= 0$, oppose $= 1$, support $= 2$, and strongly support $= 3$.

- Here we can see that these numeric codes communicate the ordered ranking, with 0 being the least supportive and 3 being the most supportive.

Describing Data: Continuous

- A continuous variable is the most precise because it describes exact differences between units.

Describing Data: Continuous

- A continuous variable is the most precise because it describes exact differences between units.

- Examples include a country's GDP, a person's age, the number of terrorist attacks, temperature, volume of liquid, and so on. These types of variables of course require no alternative numerical coding because they are already numeric.

Describing Data: Central Tendency and Variable Type

- Each variable type has appropriate measures of central tendency, or the typical value in the distribution of our data

Describing Data: Central Tendency and Variable Type

- Each variable type has appropriate measures of central tendency, or the typical value in the distribution of our data
- We want to use a central tendency statistic because it is almost always easier to interpret the statistic than it is to interpret the dataset in its entirety

Describing Data: Central Tendency and Variable Type

- Each variable type has appropriate measures of central tendency, or the typical value in the distribution of our data
- We want to use a central tendency statistic because it is almost always easier to interpret the statistic than it is to interpret the dataset in its entirety
- We have three options: mean, median, and mode

Describing Data: Central Tendency and Variable Type

- The **mean** tells us the central tendency of the distribution of data and closely resembles our everyday use of the word "average"

Describing Data: Central Tendency and Variable Type

- The **mean** tells us the central tendency of the distribution of data and closely resembles our everyday use of the word "average"
- The **mode** tells us the most common value found in the distribution of data

Describing Data: Central Tendency and Variable Type

- The **mean** tells us the central tendency of the distribution of data and closely resembles our everyday use of the word "average"
- The **mode** tells us the most common value found in the distribution of data
- The **median** tells us the middle value in the distribution of the data

Describing Data: Central Tendency and Variable Type

Table: Variable Type and Appropriate Measure of Central Tendency

Variable Type	**Appropriate Measure**
Unordered Categorical	Mode
Ordered Categorical	Median, Mode
Continuous	Mean, Median, Mode

Describing Data: Central Tendency

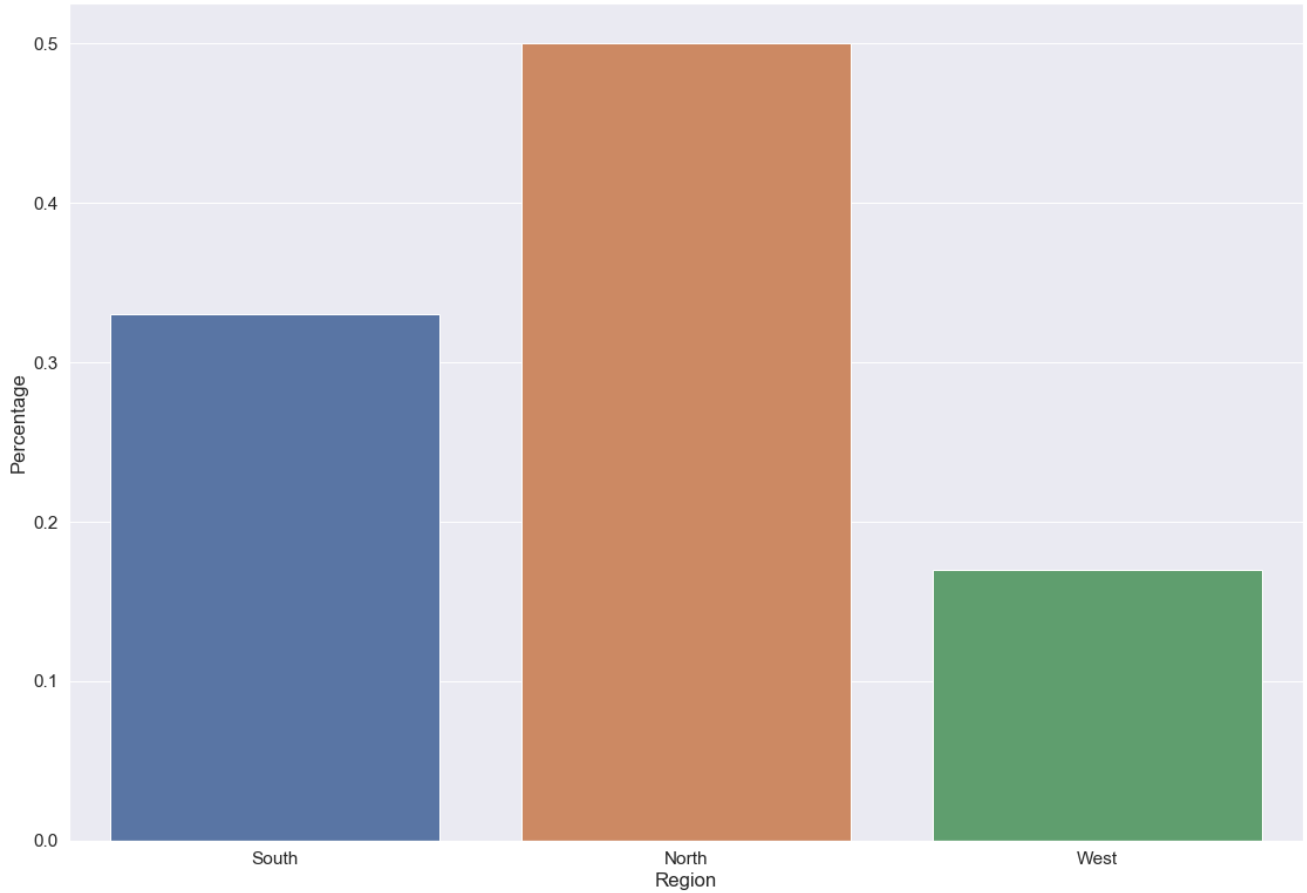
- Bar plots are the preferred way to summarize our data since most of our data tends to be unordered or ordered categorical

Describing Data: Central Tendency

- Bar plots are the preferred way to summarize our data since most of our data tends to be unordered or ordered categorical
- After plotting the data, write two statements: (1) describing the variable's distribution in its entirety, and (2) identifying the central tendency

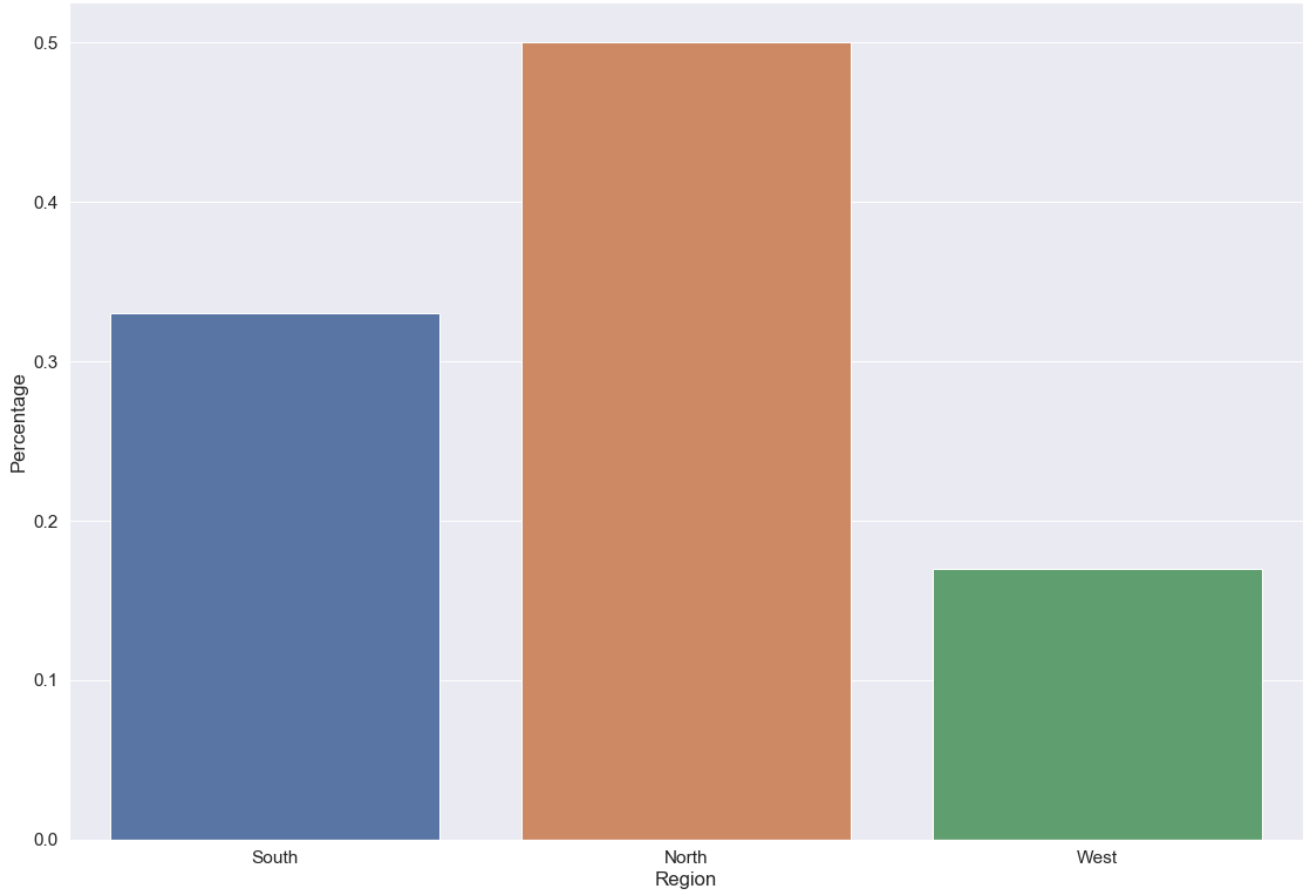
Describing Data: Unordered Categorical Central Tendency

- **Descriptive Statement**: In a survey of 90 people, 50% of the respondents were located in the North, while over 80% of the respondents were located in the North or South



Describing Data: Unordered Categorical Central Tendency

- **Central Tendency**: Among the 90 individuals surveyed, the mode is the North, with 50% of the respondents having this value



Describing Data: Ordered Categorical Central Tendency

- We have to be a little more careful in our presentation and interpretation of ordered categorical data

Describing Data: Ordered Categorical Central Tendency

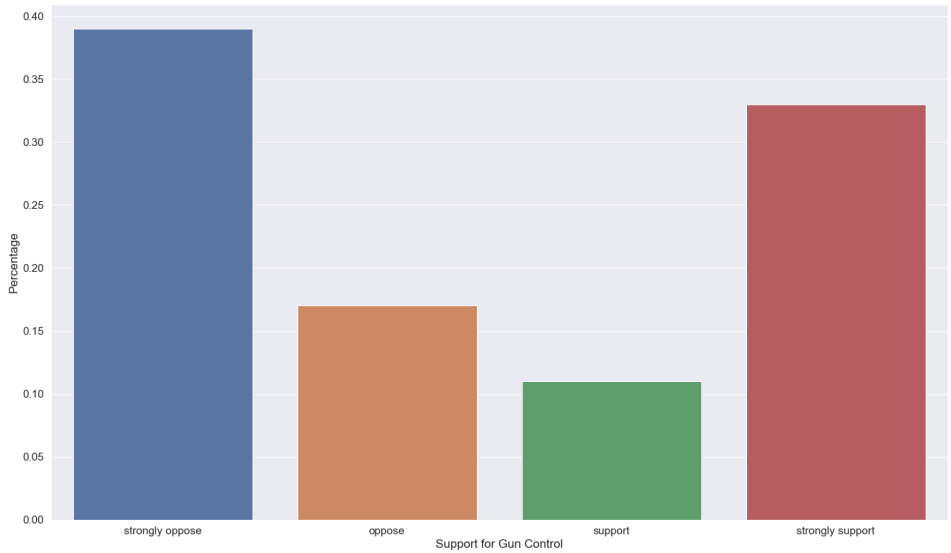
- We have to be a little more careful in our presentation and interpretation of ordered categorical data
- Two ways to derive central tendency: median and mode

Describing Data: Ordered Categorical Central Tendency

- We have to be a little more careful in our presentation and interpretation of ordered categorical data
- Two ways to derive central tendency: median and mode
- The same process applies: after plotting, we should write a statement describing our data

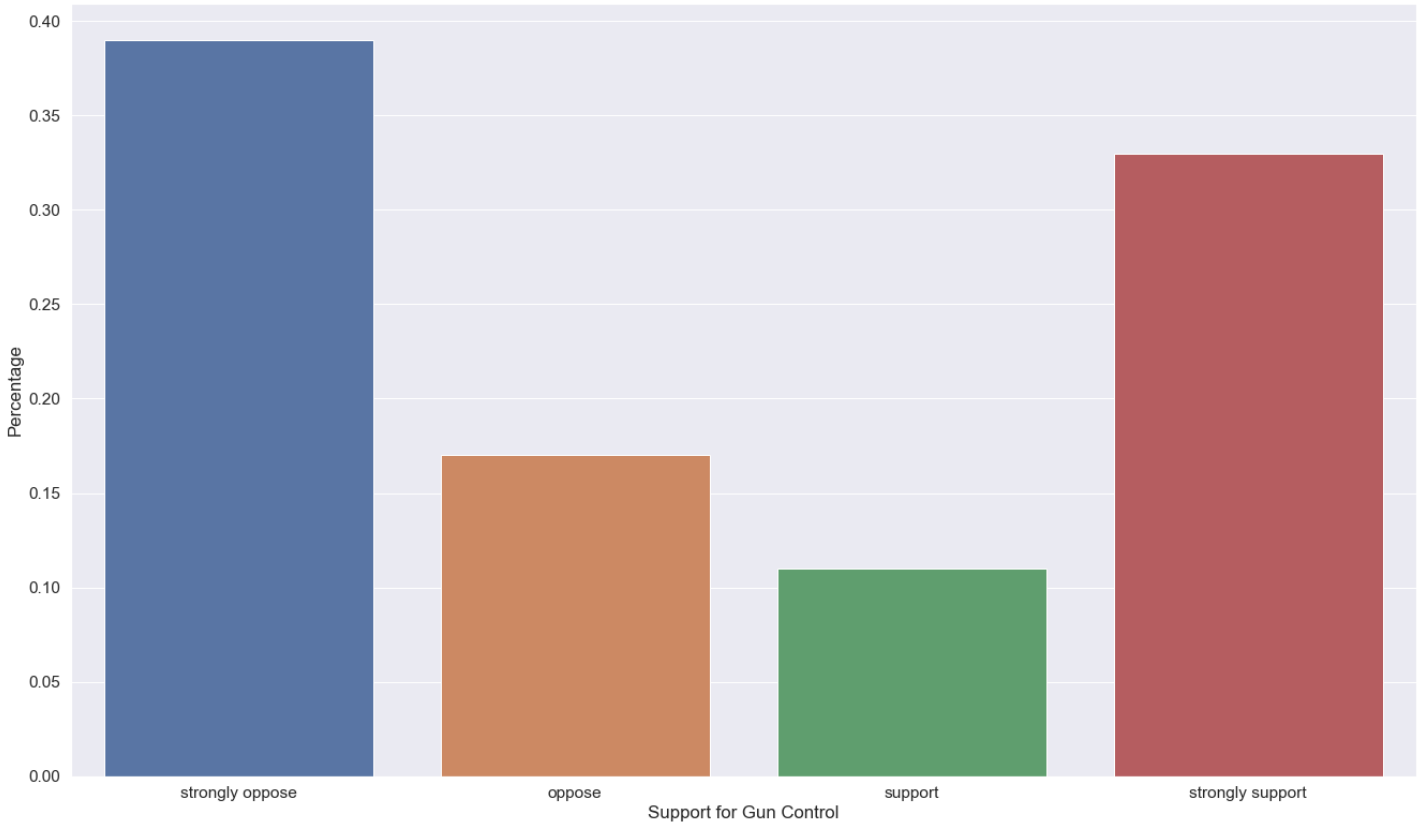
Describing Data: Ordered Categorical Central Tendency

- **Descriptive Statement**: In a survey of 90 people, nearly 35% of respondents strongly oppose gun control. Although opposition to gun control is most common, over 30% of respondents strongly support gun control. Most respondents fall into one of the two extremes while the rest are distributed across the middle values of the variable. There may be a slight bias towards gun control opposition: over half of the sample opposes it



Describing Data: Ordered Categorical Central Tendency

- **Central Tendency**: Among the 90 individuals surveyed, the mode is `strongly oppose` and `strongly support` (bimodal) with 72% of respondents having these values



Describing Data: Ordered Categorical Central Tendency

- We can also find the median; the center of the data's distribution

Describing Data: Ordered Categorical Central Tendency

- We can also find the median; the center of the data's distribution
- Median is part of a family known as percentiles

Describing Data: Ordered Categorical Central Tendency

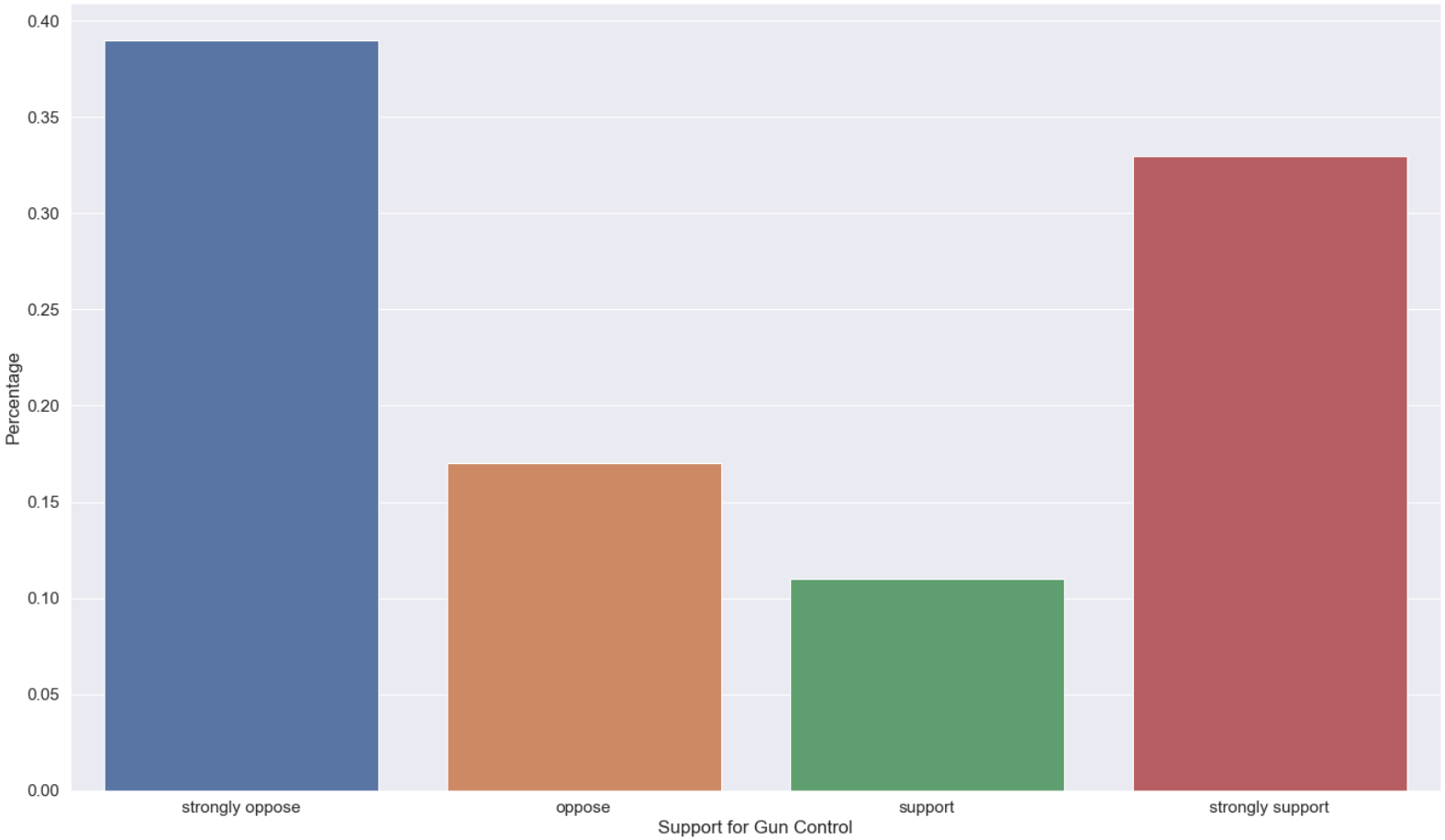
- We can also find the median; the center of the data's distribution
- Median is part of a family known as percentiles
- A percentile reports the percentage of cases in a distribution that lie below it

Describing Data: Ordered Categorical Central Tendency

- We can also find the median; the center of the data's distribution
- Median is part of a family known as percentiles
- A percentile reports the percentage of cases in a distribution that lie below it
- Since the median is the 50th percentile, it divides the distribution in half

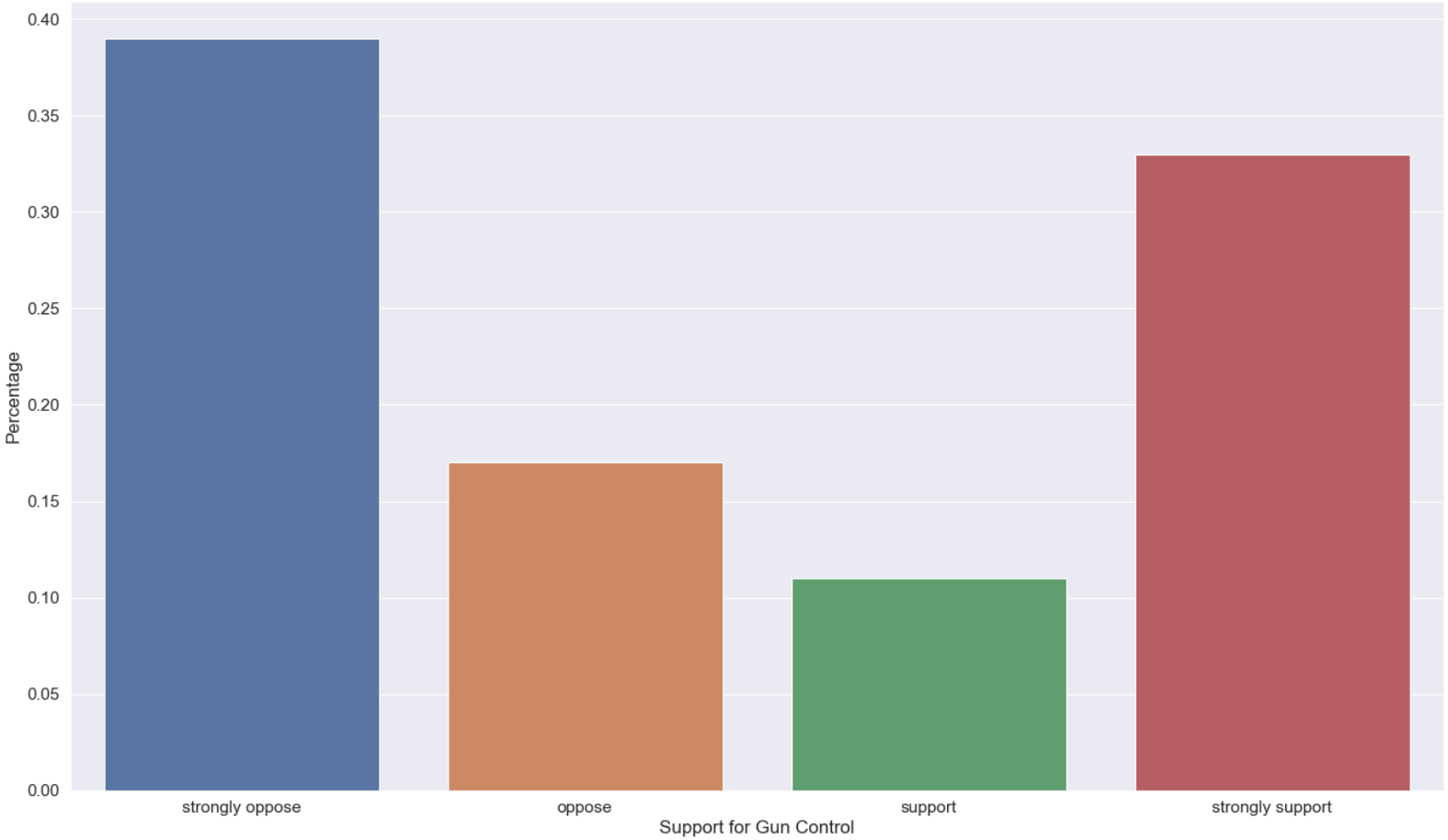
Describing Data: Ordered Categorical Central Tendency

- What is the median of the `gun control` variable? We can visually find the median from our bar plot



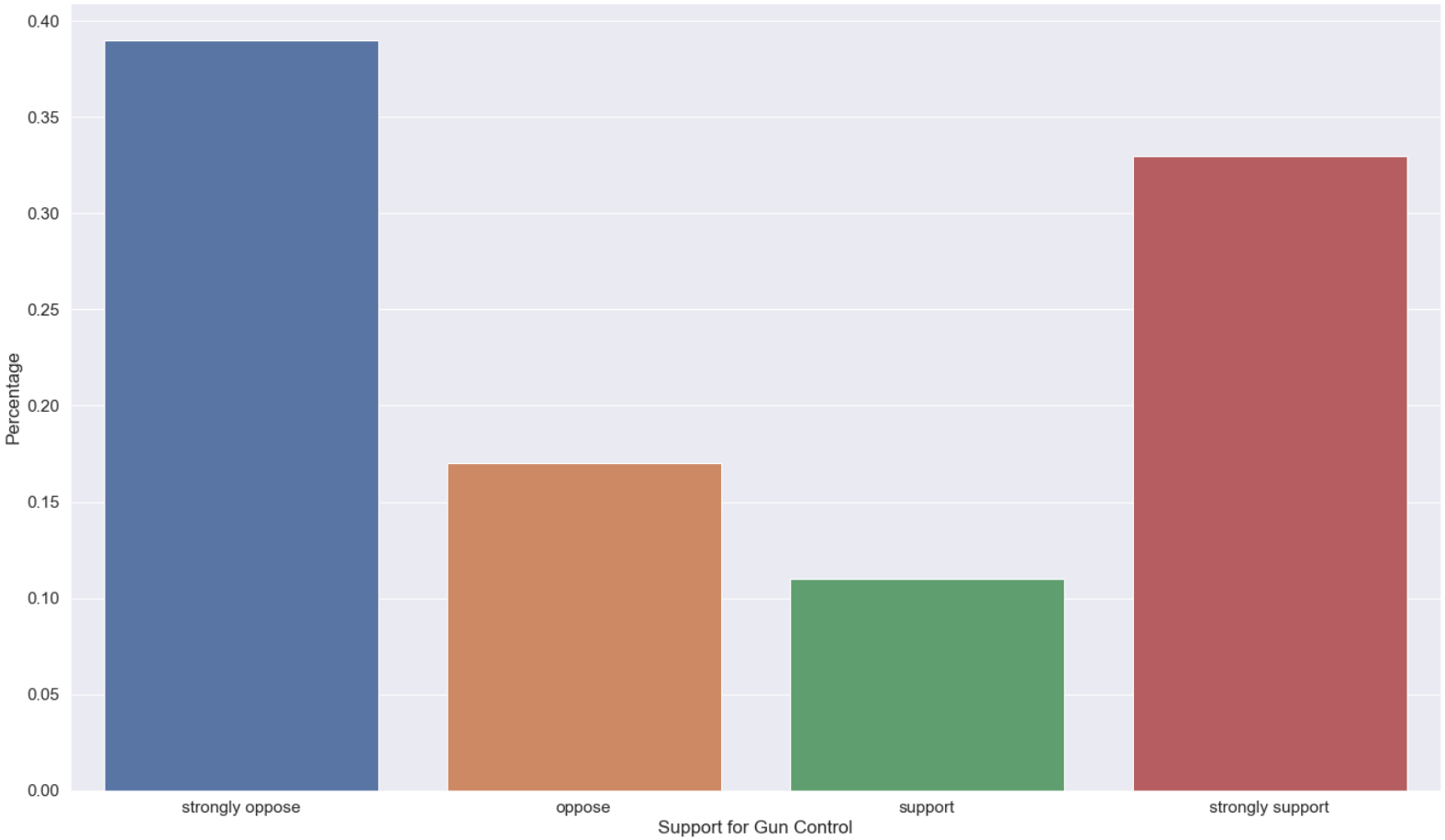
Describing Data: Ordered Categorical Central Tendency

- strongly oppose is not the median because only 39% of the cases have this value



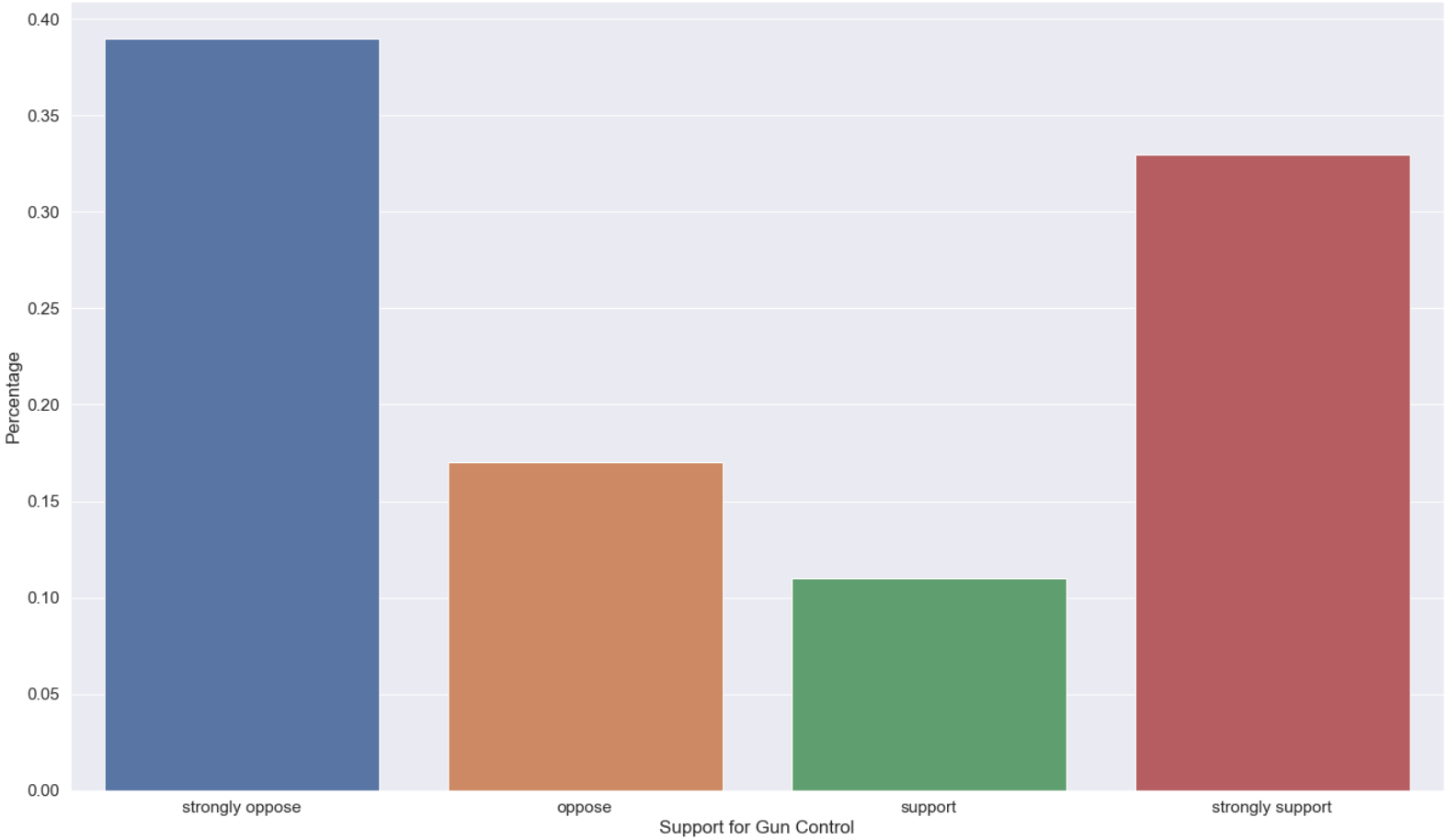
Describing Data: Ordered Categorical Central Tendency

- If we jump up to `support`, we can see we went too far because 67% of respondents are at, or below, this value (by adding each category's percentage)



Describing Data: Ordered Categorical Central Tendency

- The median is `oppose` because 56% of the cases fall in or below this value



Describing Data: Continuous Data Central Tendency

- We have all options (mean, median, mode) available to us in determining our data's central tendency

Describing Data: Continuous Data Central Tendency

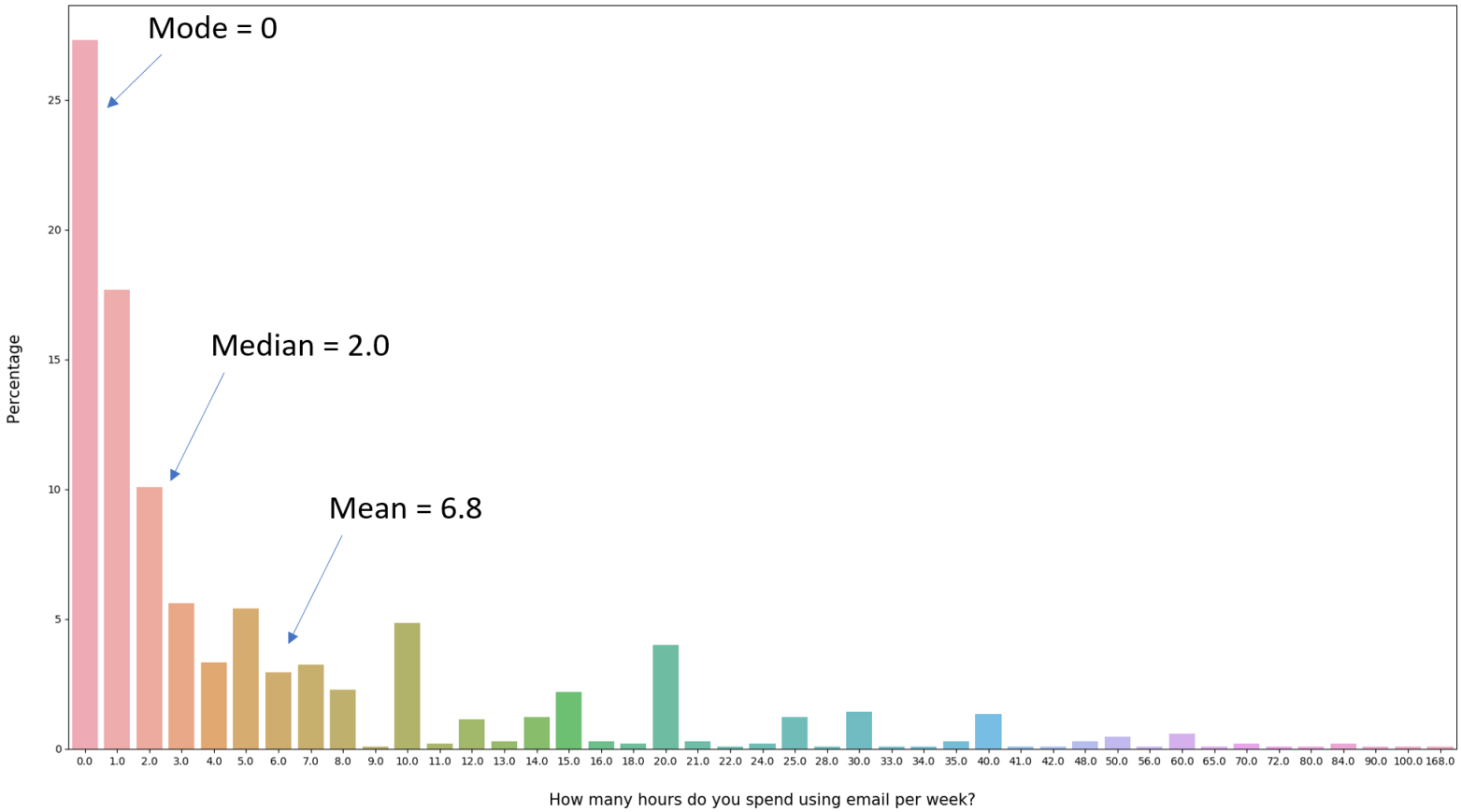
- We have all options (mean, median, mode) available to us in determining our data's central tendency
- Like the other variable types, we will use bar plots to better understand and interpret our data

Describing Data: Continuous Data Central Tendency

- We have all options (mean, median, mode) available to us in determining our data's central tendency
- Like the other variable types, we will use bar plots to better understand and interpret our data
- The same process applies: after plotting, we should write a statement describing our data

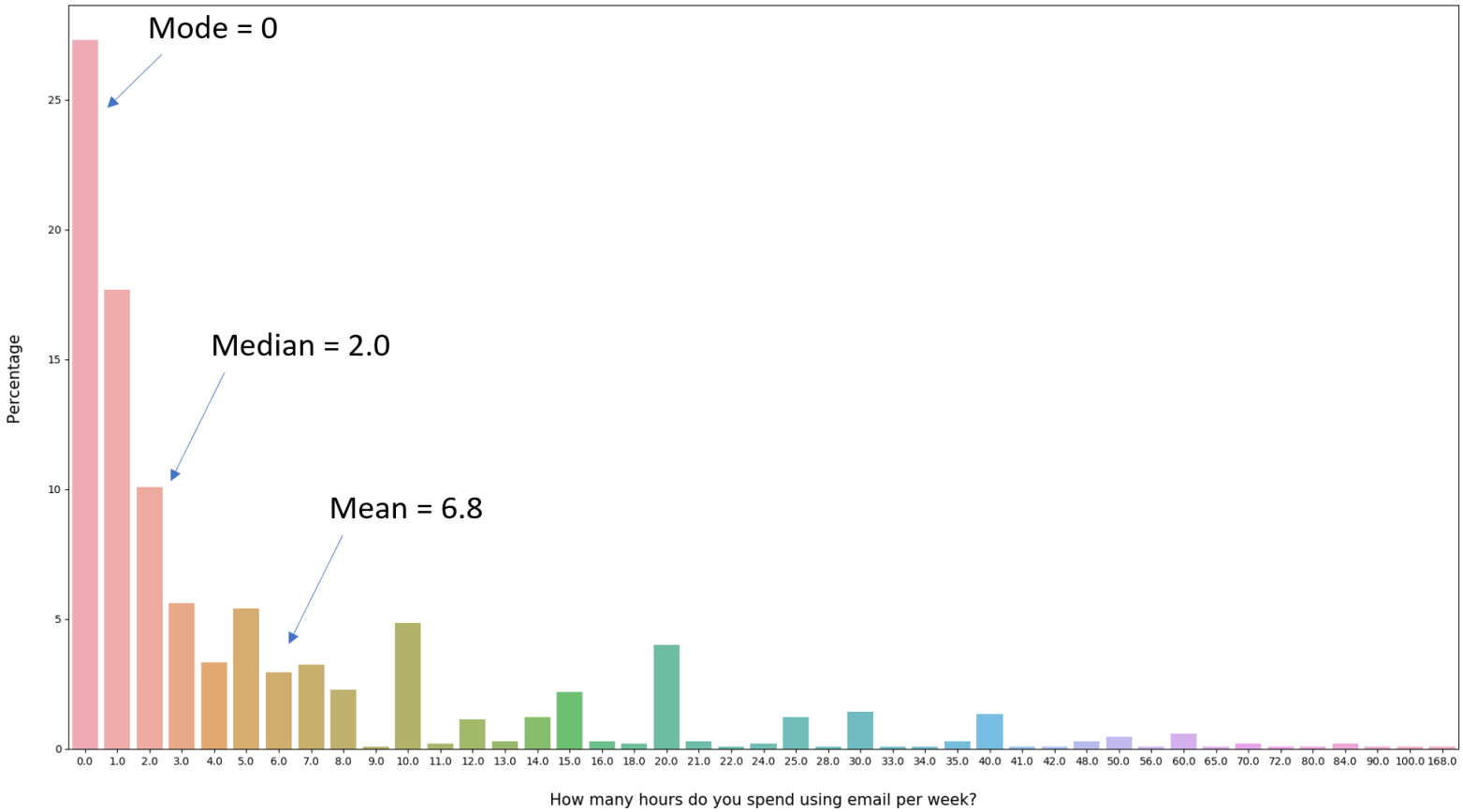
Describing Data: Continuous Data Central Tendency

- Since continuous data has many more values (potentially infinite), our data's distribution is much different from the categorical examples



Describing Data: Continuous Data Central Tendency

- Our data has a positive skew, which can be seen visually as a long right-hand tail, or mathematically, the mean $>$ median



Describing Data: Continuous Data Central Tendency

- Like most things, we have to exercise judgment when choosing our measure of central tendency for a continuous variable

Describing Data: Continuous Data Central Tendency

- Like most things, we have to exercise judgment when choosing our measure of central tendency for a continuous variable
- The distribution and presence of substantive outliers can have strong effects on our measures of central tendency

Describing Data: Continuous Data Central Tendency

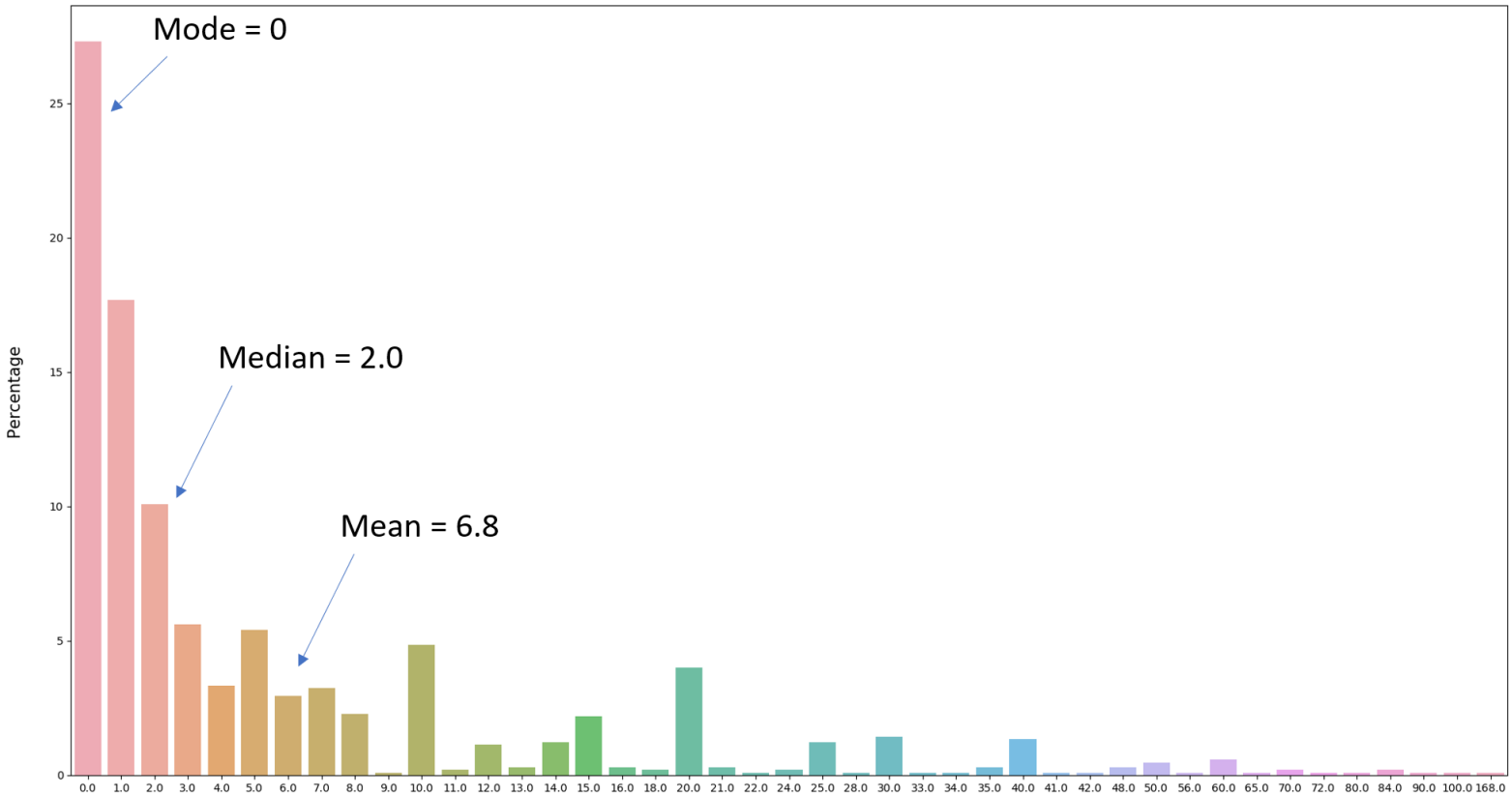
- Like most things, we have to exercise judgment when choosing our measure of central tendency for a continuous variable
- The distribution and presence of substantive outliers can have strong effects on our measures of central tendency
- Ask yourself, is it misleading to use the mean as the central tendency of this distribution?

Describing Data: Continuous Data Central Tendency

- Like most things, we have to exercise judgment when choosing our measure of central tendency for a continuous variable
- The distribution and presence of substantive outliers can have strong effects on our measures of central tendency
- Ask yourself, is it misleading to use the mean as the central tendency of this distribution?
- Since the long tail pulls the mean right-ward, and since there is a substantive difference between the median and mean, we should use the median to describe our data

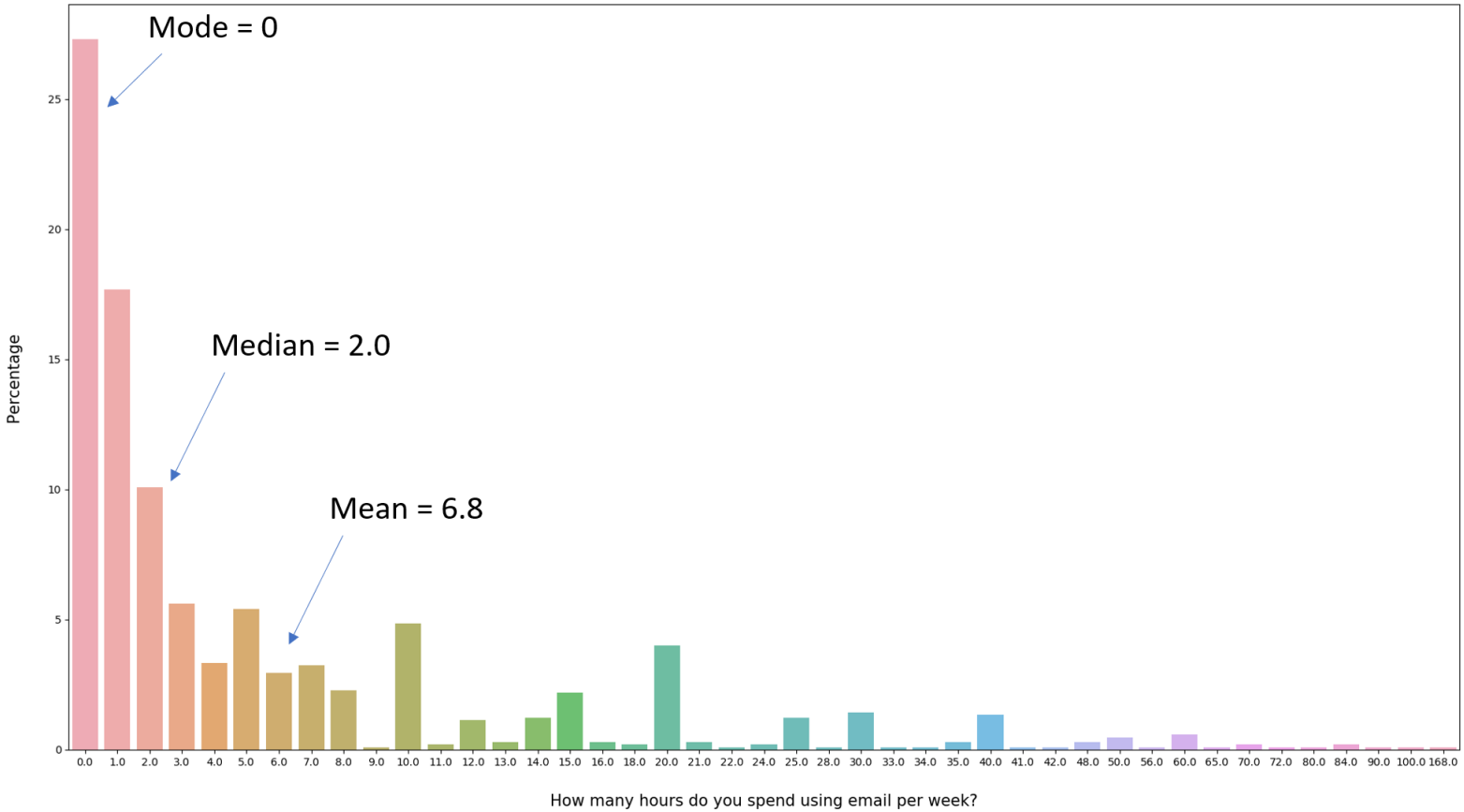
Describing Data: Continuous Data Central Tendency

- **Descriptive Statement**: In a survey of 1000 people, roughly 45% of the respondents spent `0` to `1` hours a week using email. While the survey reveals a positive skew, most respondents spent very little time using email each week.



Describing Data: Continuous Data Central Tendency

- **Central Tendency**: Among the 1000 individuals surveyed, the median is 2.0, with 55% of the respondents having this value



Review

Part 3: Hypotheses and Making Comparisons

Hypotheses and Making Comparisons: Objectives

- This class aims to improve the analyst's ability to formulate and test hypotheses and compare data

Hypotheses and Making Comparisons: Objectives

- This class aims to improve the analyst's ability to formulate and test hypotheses and compare data
- While hypotheses are generalized workhorse models worthy of in-depth attention, this class is mostly focused on *correctly* making comparisons with an incredibly simple but powerful two-way table

Hypotheses and Making Comparisons: Objectives

- This class aims to improve the analyst's ability to formulate and test hypotheses and compare data
- While hypotheses are generalized workhorse models worthy of in-depth attention, this class is mostly focused on *correctly* making comparisons with an incredibly simple but powerful two-way table
- Analysts will be able to present their data in tables correctly, interpret them correctly, and use these tables to further ground their argumentative logic and evidentiary basis for their claims

Hypothesis Testing: Introduction

- Now that we have constructed, described, and summarized our dependent variable, our next goal is to write and test hypotheses associated with it

Hypothesis Testing: Introduction

- Now that we have constructed, described, and summarized our dependent variable, our next goal is to write and test hypotheses associated with it

- "Why" questions occur to us naturally

Hypothesis Testing: Introduction

- Now that we have constructed, described, and summarized our dependent variable, our next goal is to write and test hypotheses associated with it
- "Why" questions occur to us naturally
- We ask ourselves why when we are trying to explain human behavior or outcomes that unfold across the world

Hypothesis Testing: Introduction

- Now that we have constructed, described, and summarized our dependent variable, our next goal is to write and test hypotheses associated with it
- "Why" questions occur to us naturally
- We ask ourselves why when we are trying to explain human behavior or outcomes that unfold across the world
- We focus on "why" questions because they make an explicit observation about a characteristic that varies

Hypothesis Testing: Introduction

- Now that we have constructed, described, and summarized our dependent variable, our next goal is to write and test hypotheses associated with it
- "Why" questions occur to us naturally
- We ask ourselves why when we are trying to explain human behavior or outcomes that unfold across the world
- We focus on "why" questions because they make an explicit observation about a characteristic that varies
- It is that difference, that variation, we want to understand

Hypothesis Testing: Introduction

- In the U.S., why is it that only 50% of the population votes?

Hypothesis Testing: Introduction

- In the U.S., why is it that only 50% of the population votes?
- We want to know why some people vote and some people do not vote

Hypothesis Testing: Introduction

- In the U.S., why is it that only 50% of the population votes?
- We want to know why some people vote and some people do not vote
- "Why" questions implicitly call for a causal explanation for the difference

Hypothesis Testing: Introduction

- In the U.S., why is it that only 50% of the population votes?
- We want to know why some people vote and some people do not vote
- "Why" questions implicitly call for a causal explanation for the difference
- In light of this, we might re-frame the question as: What causes differences between voters in their turnout?

Hypothesis Testing: Introduction

- In the U.S., why is it that only 50% of the population votes?
- We want to know why some people vote and some people do not vote
- "Why" questions implicitly call for a causal explanation for the difference
- In light of this, we might re-frame the question as: What causes differences between voters in their turnout?
- While summarizing data in Part 2 is the beginning of **descriptive inference**, hypothesis testing is the beginning of **causal inference** (i.e., learning about one variable's effects on another)

Hypothesis Testing: Introduction

- Proposing explanations is the essence of social science research

Hypothesis Testing: Introduction

- Proposing explanations is the essence of social science research
- It is a creative practice that allows us to imagine different explanations for observed differences between units

Hypothesis Testing: Introduction

- Proposing explanations is the essence of social science research
- It is a creative practice that allows us to imagine different explanations for observed differences between units
- Explanations, however, is a not a random or free-for-all activity

Hypothesis Testing: Introduction

- Proposing explanations is the essence of social science research
- It is a creative practice that allows us to imagine different explanations for observed differences between units
- Explanations, however, is a not a random or free-for-all activity
- We propose explanations inside hypotheses which are conditional if-then statements that specify the relationship between the cause and effect

Hypotheses

- Hypotheses are generalized workhorse models and simplified theories that propose explanations for our "why" questions. It tells us what we should find when we examine our data

Hypotheses

- Hypotheses are generalized workhorse models and simplified theories that propose explanations for our "why" questions. It tells us what we should find when we examine our data
- Hypotheses state that as a unit's value on the independent variable changes, so too does its value on the dependent variable

Components of Hypotheses

Hypotheses are comprised of the following four components:

- **Unit of Analysis** - the cases or units to which the hypothesis applies to

Components of Hypotheses

Hypotheses are comprised of the following four components:

- **Unit of Analysis** - the cases or units to which the hypothesis applies to

- **Independent Variable** - The proposed cause of the effect

Components of Hypotheses

Hypotheses are comprised of the following four components:

- **Unit of Analysis** - the cases or units to which the hypothesis applies to
- **Independent Variable** - The proposed cause of the effect
- **Dependent Variable** - The proposed effect of the cause

Components of Hypotheses

Hypotheses are comprised of the following four components:

- **Unit of Analysis** - the cases or units to which the hypothesis applies to
- **Independent Variable** - The proposed cause of the effect
- **Dependent Variable** - The proposed effect of the cause
- **Conditional Statement** - A specification of what happens to the dependent variable when the independent variable changes

Components of Hypotheses

Hypotheses are framed in specific ways:

H_1: In a comparison of *individuals*, those who are *more educated* will be *more likely* to vote in elections than those who are *less educated*

Components of Hypotheses

Hypotheses are framed in specific ways:

H_1: In a comparison of *individuals*, those who are *more educated* will be *more likely* to vote in elections than those who are *less educated*

- The hypothesis is fully-specified because:

Components of Hypotheses

Hypotheses are framed in specific ways:

H_1: In a comparison of *individuals*, those who are *more educated* will be *more likely* to vote in elections than those who are *less educated*

- The hypothesis is fully-specified because:
- It clearly identifies the units under analysis – *individuals*

Components of Hypotheses

Hypotheses are framed in specific ways:

H_1: In a comparison of *individuals*, those who are *more educated* will be *more likely* to vote in elections than those who are *less educated*

- The hypothesis is fully-specified because:
- It clearly identifies the units under analysis – *individuals*
- It clearly identifies the independent variable - *education*

Components of Hypotheses

Hypotheses are framed in specific ways:

H_1: In a comparison of *individuals*, those who are *more educated* will be *more likely* to vote in elections than those who are *less educated*

- The hypothesis is fully-specified because:
- It clearly identifies the units under analysis – *individuals*
- It clearly identifies the independent variable - *education*
- It clearly identifies the dependent variable - *likelihood of voting*

Components of Hypotheses

Hypotheses are framed in specific ways:

H_1: In a comparison of *individuals*, those who are *more educated* will be *more likely* to vote in elections than those who are *less educated*

- The hypothesis is fully-specified because:
- It clearly identifies the units under analysis – *individuals*
- It clearly identifies the independent variable - *education*
- It clearly identifies the dependent variable - *likelihood of voting*
- It clearly makes a conditional if-then statement: if we increase an individuals level of education, then their likelihood of voting increases

Hypothesis Template

The following hypothesis template will not lead you astray:

- In a comparison of *[unit of analysis]*, those having *[one value of the independent variable]* will be [*more/less likely*] to have *[one value of the dependent variable]* than will those having *[a different value on the independent variable]*

Making Comparisons

- Hypotheses suggest a comparison because if we separate units according to their values on the independent variable and compare their values on the dependent variable, we should find a difference in the hypothesized way

Table: Appropriate Methods for Making Comparisons

Dependent Variable	Independent Variable	Method
Unordered or Ordered Categorical	Unordered or Ordered Categorical	Cross-tabulation
Continuous	Unordered or Ordered Categorical	Mean Comparison
Continuous	Unordered or Ordered Categorical or Continuous	Linear Regression
Unordered or Ordered Categorical	Unordered or Ordered Categorical or Continuous	Non-Linear Regression

Making Comparisons

- Hypotheses suggest a comparison because if we separate units according to their values on the independent variable and compare their values on the dependent variable, we should find a difference in the hypothesized way
- Depending on the type of variables that we have, we make comparisons in different ways:

Table: Appropriate Methods for Making Comparisons

Dependent Variable	Independent Variable	Method
Unordered or Ordered Categorical	Unordered or Ordered Categorical	Cross-tabulation
Continuous	Unordered or Ordered Categorical	Mean Comparison
Continuous	Unordered or Ordered Categorical or Continuous	Linear Regression
Unordered or Ordered Categorical	Unordered or Ordered Categorical or Continuous	Non-Linear Regression

Making Comparisons: Cross-tabulation

- Since we work extensively with unordered and ordered categorical variables, we are going to practice cross-tabulation or cross-tabs

Making Comparisons: Cross-tabulation

- Since we work extensively with unordered and ordered categorical variables, we are going to practice cross-tabulation or cross-tabs

- Cross-tabs are powerful tools that we can easily and quickly use to test hypotheses

Making Comparisons: Cross-tabulation

- Since we work extensively with unordered and ordered categorical variables, we are going to practice cross-tabulation or cross-tabs
- Cross-tabs are powerful tools that we can easily and quickly use to test hypotheses
- In this class, we are going to practice building and interpreting cross-tabs correctly

Making Comparisons: Framing a Hypothesis

- Before we make a comparison, we first want to specify a hypothesis

Making Comparisons: Framing a Hypothesis

- Before we make a comparison, we first want to specify a hypothesis

- H_1: In a comparison of *individuals*, those who are *married* will be *less likely* to favor the legalization of marijuana than those who are *single*

Making Comparisons: Framing a Hypothesis

- Before we make a comparison, we first want to specify a hypothesis

- H_1: In a comparison of *individuals*, those who are *married* will be *less likely* to favor the legalization of marijuana than those who are *single*

- We will test the hypothesis by using the General Social Survey (2012)

Making Comparisons: Framing a Hypothesis

- Before we make a comparison, we first want to specify a hypothesis
- H_1: In a comparison of *individuals*, those who are *married* will be *less likely* to favor the legalization of marijuana than those who are *single*
- We will test the hypothesis by using the General Social Survey (2012)
- Since 1972, the survey provides academics and politicians the opinions of Americans on a range of issues

Making Comparisons: Framing a Hypothesis

- Before we make a comparison, we first want to specify a hypothesis
- H_1: In a comparison of *individuals*, those who are *married* will be *less likely* to favor the legalization of marijuana than those who are *single*
- We will test the hypothesis by using the General Social Survey (2012)
- Since 1972, the survey provides academics and politicians the opinions of Americans on a range of issues
- Variables: `grass` (favor legalization) and `absingle` (are you single)

Making Comparisons: Data

- Sample Data:

	grass	absingle
0	NaN	NaN
1	legal	yes
2	legal	no
3	legal	NaN
4	not legal	NaN

Creating and Interpreting Cross-tabs

- **Step 1**: List each value of the dependent variable on its own row

Creating and Interpreting Cross-tabs

- **Step 1**: List each value of the dependent variable on its own row
- **Step 2**: List each category of the independent variable on its own column

Creating and Interpreting Cross-tabs

- **Step 1**: List each value of the dependent variable on its own row
- **Step 2**: List each category of the independent variable on its own column
- **Step 3**: Insert the total count for value

Creating and Interpreting Cross-tabs

- **Step 1**: List each value of the dependent variable on its own row
- **Step 2**: List each category of the independent variable on its own column
- **Step 3**: Insert the total count for value
- **Step 4**: Calculate the percentage by **column**: $\frac{part}{whole}$

Creating and Interpreting Cross-tabs

- **Step 1**: List each value of the dependent variable on its own row
- **Step 2**: List each category of the independent variable on its own column
- **Step 3**: Insert the total count for value
- **Step 4**: Calculate the percentage by **column**: $\frac{part}{whole}$
- **Step 5**: Compare columns *only* on the same row

Creating and Interpreting Cross-tabs

- **Step 1**: List each category of the dependent variable on its own row

Legalize Marijuana
Favor
Oppose

Creating and Interpreting Cross-tabs

- **Step 2**: List each category of the independent variable on each column

	Marital Status	
Legalize Marijuana	Married	Single
Favor		
Oppose		

Creating and Interpreting Cross-tabs

- **Step 3**: Insert the total count for each value

	Marital Status	
Legalize Marijuana	Married	Single
Favor	(127)	(177)
Oppose	(213)	(79)

Creating and Interpreting Cross-tabs

- **Step 4**: Calculate the percentage by **column**: $\frac{part}{whole}$

	Marital Status	
Legalize Marijuana	Married	Single
Favor	37% (127)	69% (177)
Oppose	63% (213)	31% (79)

Creating and Interpreting Cross-tabs

- **Step 5**: Compare **column percentages** *only* on the same row

	Marital Status	
Legalize Marijuana	Married	Single
Favor	37% (127)	69% (177)
Oppose	63% (213)	31% (79)

Interpreting Cross-tabulations

- Template: A comparison of individuals in the married category with those in the single category reveals an increase from 37 to 69 in the percentage of individuals who favor marijuana legalization

	Marital Status	
Legalize Marijuana	Married	Single
Favor	37% (127)	69% (177)
Oppose	63% (213)	31% (79)

Interpreting Cross-tabulations

- Improved: Married people are 32% less likely to favor marijuana legalization as compared to single people

	Marital Status	
Legalize Marijuana	Married	Single
Favor	37% (127)	69% (177)
Oppose	63% (213)	31% (79)

Revisiting the Hypothesis

- H_1: In a comparison of *individuals*, those who are *married* will be *less likely* to favor the legalization of marijuana than those who are *single*

Revisiting the Hypothesis

- H_1: In a comparison of *individuals*, those who are *married* will be *less likely* to favor the legalization of marijuana than those who are *single*

- Results: Married people are 32% less likely to favor marijuana legalization as compared to single people

Revisiting the Hypothesis

- H_1: In a comparison of *individuals*, those who are *married* will be *less likely* to favor the legalization of marijuana than those who are *single*
- Results: Married people are 32% less likely to favor marijuana legalization as compared to single people
- We find strong support for our hypothesis

Review

	Gun Permits	
Vote	Favor	Oppose
McCain	30%	57%
	(175)	(119)
Obama	70%	43%
	(414)	(88)

Advanced Considerations

Controlled Comparisons

- Cross-tabulations are so powerful that we can go even further

Controlled Comparisons

- Cross-tabulations are so powerful that we can go even further
- We can estimate the effects of our independent variable while controlling for another variable

Controlled Comparisons

- Cross-tabulations are so powerful that we can go even further
- We can estimate the effects of our independent variable while controlling for another variable
- This is called a **controlled comparison**

Controlled Comparisons: Introduction

- A controlled comparison introduces another factor, or potential cause, into the mix

Controlled Comparisons: Introduction

- A controlled comparison introduces another factor, or potential cause, into the mix
- We introduce additional factors because we are always thinking: how else?

Controlled Comparisons: Introduction

- A controlled comparison introduces another factor, or potential cause, into the mix
- We introduce additional factors because we are always thinking: how else?
- How else might married and single people differ such that it accounts for their opinions on legalizing marijuana?

Controlled Comparisons: Introduction

- A controlled comparison introduces another factor, or potential cause, into the mix
- We introduce additional factors because we are always thinking: how else?
- How else might married and single people differ such that it accounts for their opinions on legalizing marijuana?
- How else might people who favor or support gun permits differ such that it accounts for their presidential vote?

Controlled Comparisons: Introduction

- A controlled comparison introduces another factor, or potential cause, into the mix
- We introduce additional factors because we are always thinking: how else?
- How else might married and single people differ such that it accounts for their opinions on legalizing marijuana?
- How else might people who favor or support gun permits differ such that it accounts for their presidential vote?
- Controlled comparisons reveal whether or not the original bi-variate relationship is: spurious, additive, or interactive

Controlled Comparisons: Introduction

- **Spurious**: The control variable (Z) defines large differences across values of the independent variable (X). (Z) explains the dependent variable (Y)

Controlled Comparisons: Introduction

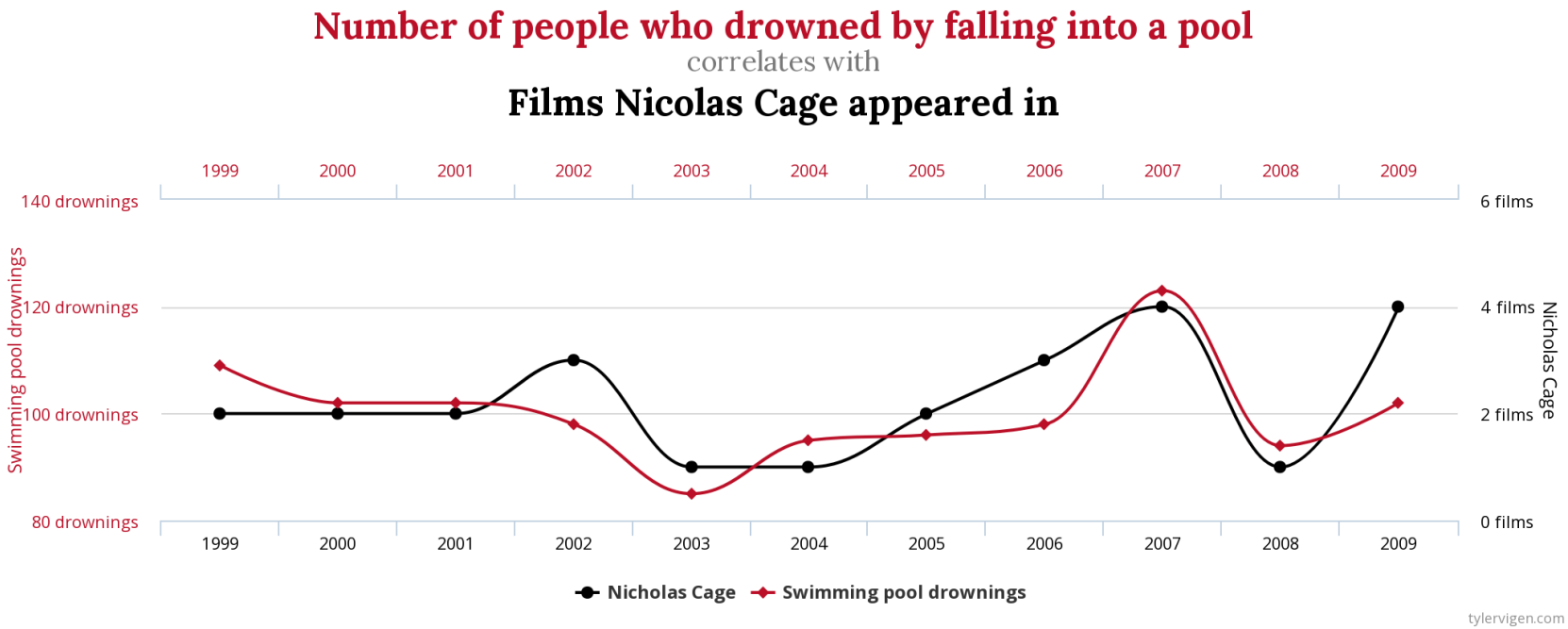
- **Spurious**: The control variable (Z) defines large differences across values of the independent variable (X). (Z) explains the dependent variable (Y)

- **Additive**: The control variable (Z) defines small differences across values of the independent variable. (Z) helps explain the dependent variable (Y)

Controlled Comparisons: Introduction

- **Spurious**: The control variable (Z) defines large differences across values of the independent variable (X). (Z) explains the dependent variable (Y)
- **Additive**: The control variable (Z) defines small differences across values of the independent variable. (Z) helps explain the dependent variable (Y)
- **Interactive**: The differences across values of the independent variable (X) on the dependent variable (Y) depends on the value of the control variable (Z)

Spurious Relationship: A Fun Example



Controlled Comparisons Example: A Familiar Hypothesis

- H_1: In a comparison of *individuals*, those who are *married* will be *less likely* to favor the legalization of marijuana than those who are *single*

Controlled Comparisons Example: A Familiar Hypothesis

- H_1: In a comparison of *individuals*, those who are *married* will be *less likely* to favor the legalization of marijuana than those who are *single*

- We previously found strong support for our hypothesis

Controlled Comparisons Example: A Familiar Hypothesis

- H_1: In a comparison of *individuals*, those who are *married* will be *less likely* to favor the legalization of marijuana than those who are *single*
- We previously found strong support for our hypothesis
- How else might people differ differ such that it accounts for their preferences over marijuana legalization?

Controlled Comparisons Example: A Familiar Hypothesis

- H_1: In a comparison of *individuals*, those who are *married* will be *less likely* to favor the legalization of marijuana than those who are *single*
- We previously found strong support for our hypothesis
- How else might people differ differ such that it accounts for their preferences over marijuana legalization?
- An alternative explanation: political preferences

Controlled Comparisons Example: A Familiar Hypothesis

- H_1: In a comparison of *individuals*, those who are *married* will be *less likely* to favor the legalization of marijuana than those who are *single*
- We previously found strong support for our hypothesis
- How else might people differ differ such that it accounts for their preferences over marijuana legalization?
- An alternative explanation: political preferences
- We can test this hypothesis by including another variable from the GSS survey: `pres08` (Did you vote for Obama or McCain)

Controlled Comparison Table

- **Step 1**: List each category of the DV on its own row

Legalize Marijuana
Favor
Oppose

Controlled Comparison Table

- **Step 2**: List each category of the IV on its own column twice, one set for each control variable value

Legalize Marijuana	Married	Single	Married	Single
Favor				
Oppose				

Controlled Comparison Table

- **Step 3**: List each value of the control variable above each IV pair, thereby creating two tables side-by-side

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor				
Oppose				

Controlled Comparison Table

- **Step 4**: Calculate the percentage by **column**: $\frac{part}{whole}$, conditional on the value of Z

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparison Table

- **Step 5**: Interpret

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparison Table

- **Step 5**: Interpret
- **Controlled effect**: the effect of X on Y given one value of Z

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparison Table

- **Hold Z Constant**: Compare values within each value of Z

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparison Table

- **Hold Z Constant**: Compare values within each value of Z
- Interpretation: 40% of married Obama voters favor marijuana legalization, compared with 72% of single Obama voters

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparison Table

- **Hold Z Constant**: Compare values within each value of Z
- Interpretation: 40% of married Obama voters favor marijuana legalization, compared with 72% of single Obama voters
- Controlled Effect: 40%-72% = -32%

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparison Table

- **Hold Z Constant**: Compare values within each value of Z
- Interpretation: 40% of married Obama voters favor marijuana legalization, compared with 72% of single Obama voters
- Controlled Effect: 40%-72% = -32%
- The controlled effect of martial status among Obama voters: -32%

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparison Table

- **Hold Z Constant**: Compare values within each value of Z

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparison Table

- **Hold Z Constant**: Compare values within each value of Z
- Interpretation: 35% of married McCain voters favor marijuana legalization, compared with 68% of single McCain voters

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparison Table

- **Hold Z Constant**: Compare values within each value of Z
- Interpretation: 35% of married McCain voters favor marijuana legalization, compared with 68% of single McCain voters
- Controlled Effect: 35%-68% = -33%

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparison Table

- **Hold Z Constant**: Compare values within each value of Z
- Interpretation: 35% of married McCain voters favor marijuana legalization, compared with 68% of single McCain voters
- Controlled Effect: 35%-68% = -33%
- The controlled effect of martial status among McCain voters: -33%

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparison Table

- **Hold X Constant**: Compare values within each value of X

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparison Table

- **Hold X Constant**: Compare values within each value of X
- Interpretation: Among married Obama voters, 40% favor marijuana legalization, compared with 35% of married McCain voters

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparison Table

- **Hold X Constant**: Compare values within each value of X
- Interpretation: Among married Obama voters, 40% favor marijuana legalization, compared with 35% of married McCain voters
- Controlled Effect: 40%-35% = 5%

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparison Table

- **Hold X Constant**: Compare values within each value of X
- Interpretation: Among married Obama voters, 40% favor marijuana legalization, compared with 35% of married McCain voters
- Controlled Effect: 40%-35% = 5%
- The controlled effect of presidential vote among married people: 5%

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparison Table

- **Hold X Constant**: Compare values within each value of X

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparison Table

- **Hold X Constant**: Compare values within each value of X
- Interpretation: Among single Obama voters, 72% favor marijuana legalization, compared with 68% of single McCain voters

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparison Table

- **Hold X Constant**: Compare values within each value of X
- Interpretation: Among single Obama voters, 72% favor marijuana legalization, compared with 68% of single McCain voters
- Controlled Effect: 72%-68% = 4%

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparison Table

- **Hold X Constant**: Compare values within each value of X
- Interpretation: Among single Obama voters, 72% favor marijuana legalization, compared with 68% of single McCain voters
- Controlled Effect: 72%-68% = 4%
- The controlled effect of presidential vote among single people: 4%

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparisons: Calculate Partial Effects

- **Controlled effects** are summarized into a single statistic called a **partial effect**. A partial effect is obtained quickly by averaging the two controlled effects or more carefully by calculating a weighted average. The weights are derived from the size of each control group.

Controlled Comparison Table

- We found two controlled effects: 5% and 4%

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparison Table

- We found two controlled effects: 5% and 4%
- Quick partial effect calculation: $\frac{5+4}{2} = 4.5\%$

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparison Table

- We found two controlled effects: 5% and 4%
- Quick partial effect calculation: $\frac{5+4}{2} = 4.5\%$
- What is the partial effect of presidential vote on marijuana legalization, controlling for marital status?

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Controlled Comparison Table

- We found two controlled effects: 5% and 4%
- Quick partial effect calculation: $\frac{5+4}{2} = 4.5\%$
- What is the partial effect of presidential vote on marijuana legalization, controlling for marital status?
- Controlling for marital status, the partial effect of presidential vote on marijuana legalization opinions is roughly 4.5%

	Voted Obama		Voted McCain	
Legalize Marijuana	Married	Single	Married	Single
Favor	40%	72%	35%	68%
Oppose	60%	28%	65%	32%

Review

Source List

These courses are primarily based on the following academic works:

Goertz, Gary. Social science concepts: A user's guide. Princeton University Press, 2006.

Pollock III, Philip H. The essentials of political analysis. Cq Press, 2015.

King, Gary, Robert O. Keohane, and Sidney Verba. Designing social inquiry: Scientific inference in qualitative research. Princeton university press, 1994.