Introduction To Quantitative Political Science

Aleksandr Fisher

Empirical Political Science

- Politics is full of claims
- The credibility of claims depends on the strength of evidence and argument
- This class aims to give you tools to:
  - make credible claims, and
  - evaluate claims made by others

Figure 1: Immigration and Unemployment.

Claims in the Media

Figure 2: Do American's Support Impeachment?

Inference and Methodology

- **Inference**: a belief based on evidence and rules for processing that evidence
- **Methodology**: "tools for gathering and analyzing data to try to make valid inferences

Questions

- Does increased immigraton increase unemployment?
- Does democracy cause economic growth?
- does climate change increase the probability of civil war?

Two Categories of Inference

- *Descriptive* Inference
  - What are the facts?
- *Causal* Inference
  - Why does soemthing occur?

Descriptive Inference

- Seeks to describe the existance of something
- Examples:
  - Is the United States polarizing?
  - Is global terrorism increasing?
  - Is Russia an autocracy?

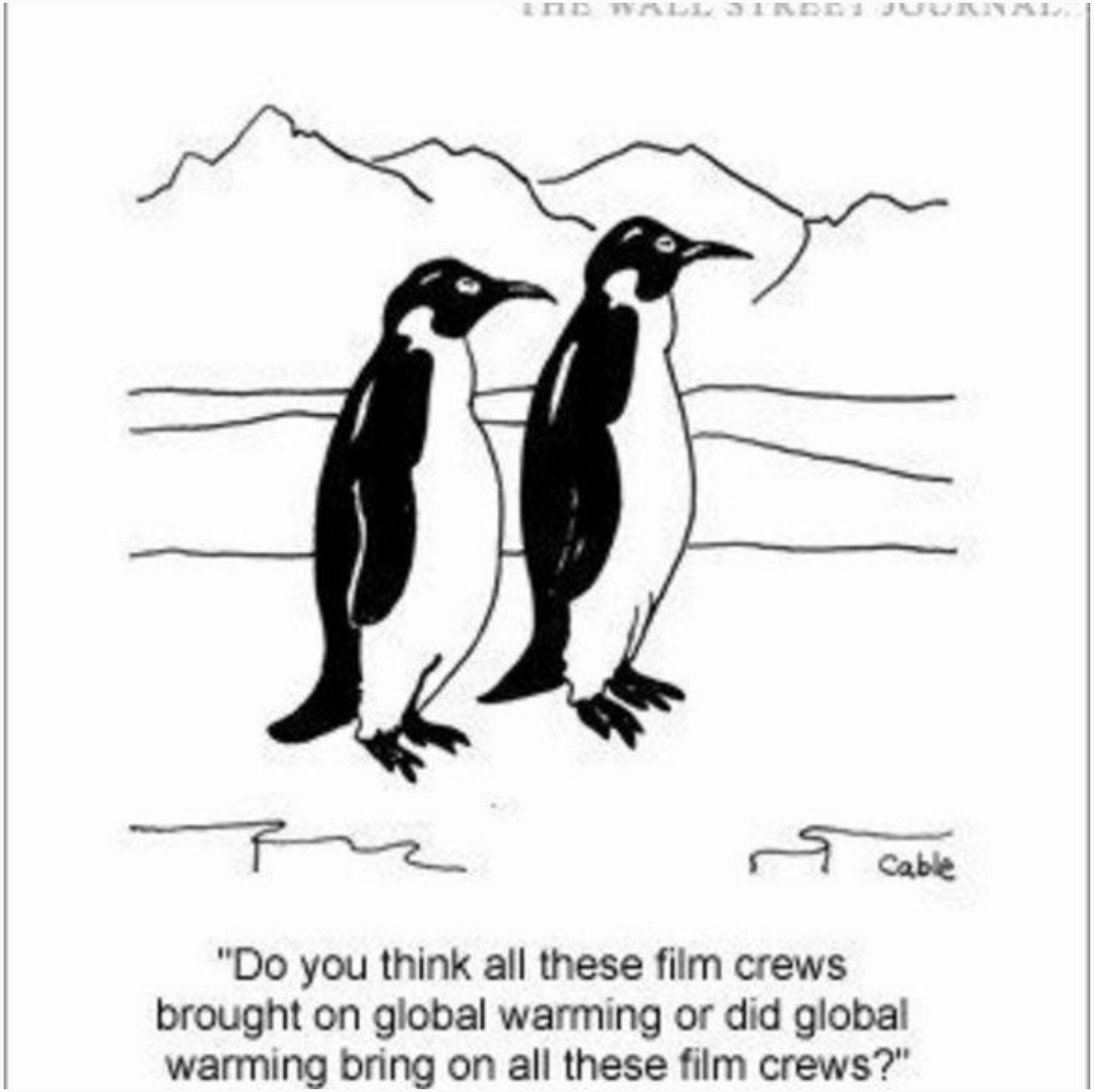
Description = "What?"

- A common feature of descriptive research and descriptive research questions is a focus on what questions
  - What is this?
  - What other things are (un)like this?
  - What features does this have?
  - What people, institutions, and ideas does this involve?
  - Where is this? When is this? What happened before and after?
  - Examples?

Beyond Description

- Sequencing
- Characterisation of processes
- Policy, content, or discourse analysis
- Conceptualisation
- Causal hypothesis generation

What makes something a cause?



Causal Inference

- Seeks to understand the effect of some variable(s) on some other variables(s)
- Questions about why:
  - Why is the United States polarizing?
  - Why is global terrorism increasing/decreasing?
  - Why is Russia not a democracy?

Causal Inference (continued...)

- Can start with either:
  - A **dependent** variable (outcome)
  - An **independent** variable (cause)

Causal Inference (continued...)

- What causes Y?
  - Associated with search for causes
  - *What causes political polarization?*
- What happens if X?
  - Associated with 'experiments'
  - *What happens when people recieve most of their news from social media networks?*

Which of these is a causal research question?

What makes a good research question

- Start from political problem or puzzle
- Builds on an existing research literature
- Non-obvious

Which is a better Research Question?

The Dataset and you

- A rectangular, case-by-variable dataset
    - dataset observations (DSOs')
- Clear unit of analysis
- Quanaitive and qualiative measures
- Calculation of summary statistics

Happiness Dataset

```
# Read Happiness Data
happ2019 = read.csv("C:/Users/afisher/Documents/R Code/Resources/Data/Happiness/2019.csv")
# First 6 observations
head(happ2019)
```

```
##   Overall.rank Country.or.region Score GDP.per.capita Social.support
## 1            1           Finland 7.769          1.340          1.587
## 2            2           Denmark 7.600          1.383          1.573
## 3            3            Norway 7.554          1.488          1.582
## 4            4           Iceland 7.494          1.380          1.624
## 5            5       Netherlands 7.488          1.396          1.522
## 6            6       Switzerland 7.480          1.452          1.526
##   Healthy.life.expectancy Freedom.to.make.life.choices Generosity
## 1                   0.986                        0.596      0.153
## 2                   0.996                        0.592      0.252
## 3                   1.028                        0.603      0.271
## 4                   1.026                        0.591      0.354
## 5                   0.999                        0.557      0.322
## 6                   1.052                        0.572      0.263
##   Perceptions.of.corruption
## 1                     0.393
## 2                     0.410
## 3                     0.341
## 4                     0.118
## 5                     0.298
## 6                     0.343
```

Quanitative vs. Qualitative research

- This divide is illusory because all research is qualitative and some involves quantitative data description

An Example: Opinion

- *Opinion* is a summay evaluation of a particular object
- Only one necessary feature: evaluation/favorability
- How do we measure this?

Operationalization

- Measure features
  - Level of measurement
  - How to score each case on each feature
  - Be concrete
- Aggregate feature measurements
  - Sum? Average? AND logical?
  - Range of possible values
  - Justify against criticisms/alternatives

Operationalization

- To study concepts, we need to be able to observe those concepts and encode them as variables
- The definition of *variable*: A dimension that describes an observation or, the operationalization of a concept

Operationalization

- Definition
  - Feature
    - Indicator(s)

Examples

- What are concepts that we use often in politics that are difficult to measure?

Activity!

- Concept: Democracy
- Attribute: Free and fair elections
- Measure:
  - Categorical
  - Ordinal
  - Numeric

Assessing Measurement Quality

- Conceptual clarity
- Construct validity
  - Convergent validity
  - Divergent validity
- Accuracy and precision

Assessing Measures

- *Conceptual clarity* is about knowing what we want to measure
- Sloppy concepts make for bad measures
  - Ambiguity
  - Vagueness

Assessing Measures

- *Construct validity* is the degree to which a variable measures a concept
- Construct validity is **high** if a variable is a measure of the concept we care about
- Construct validity is **low** if a variable is actually a measure of something else

Example Polity

Institutionalized Democracy: *Democracy is conceived as three essential, interdependent elements*. One is the **presence of institutions and procedures through which citizens can express effective preferences about alternative policies and leaders**. Second is the existence of **institutionalized constraints on the exercise of power by the executive**. Third is the **guarantee of civil liberties to all citizens in their daily lives and in acts of political participation**. Other aspects of plural democracy, such as the rule of law, systems of checks and balances, freedom of the press, and so on are means to, or specific manifestations of, these general principles. We do not include coded data on civil liberties.

Assessing Construct Validity

- Multiple Measures
- Look for:
  - Convergence (Convergent validity)
  - Discrimination (Discriminant validity)
- Convergent validity tests whether constructs that should be related, are related.
- Discriminant validity tests whether believed unrelated constructs are, in fact, unrelated.

Using Multiple Indicator

- Choose the “best” one
- Must have all indicators to be coded “1”
- Scale the indicators (sum or mean)

Accuracy and Precision

- **Accuracy** is how close a measured value is to the actual (true) value.
- **Precision** is how close the measured values are to each other.

Accuracy vs. Precision

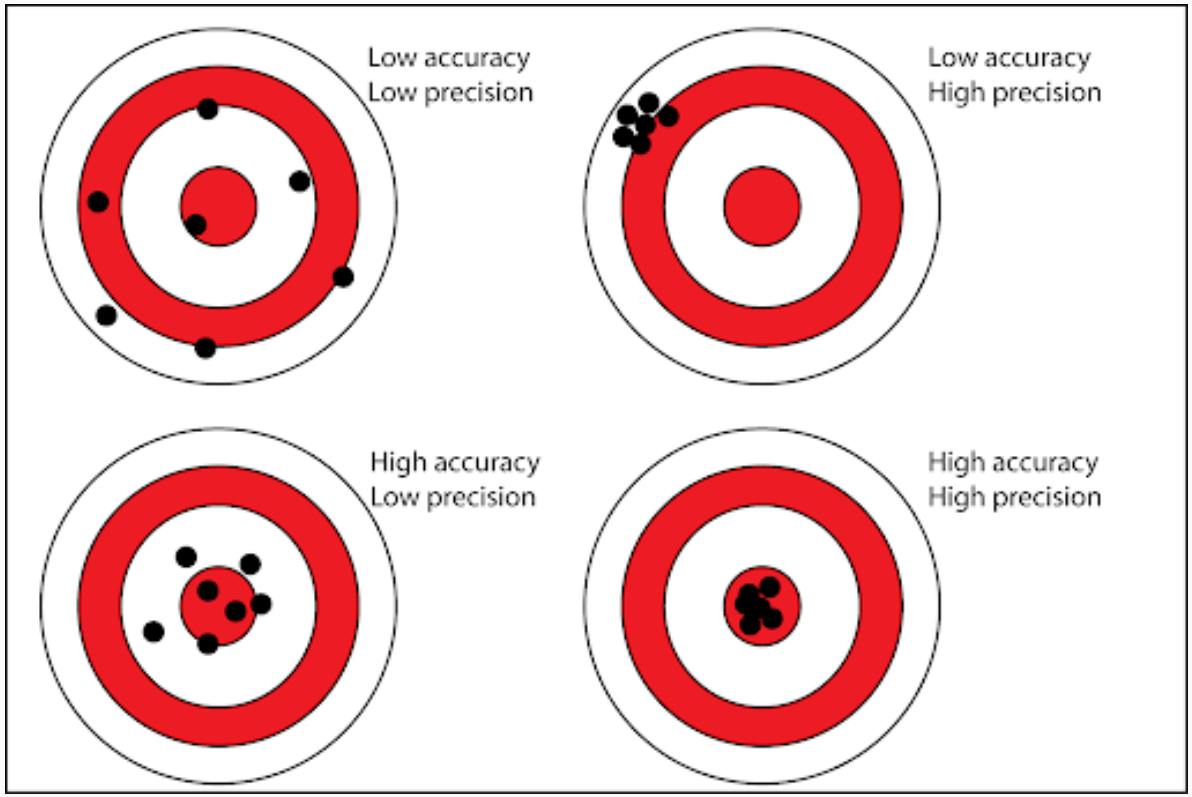


Figure 3: Accuracy vs Precision

Reliability

- **Reliability**: To what extent would our measure yield the same results if we went out and collected more data?
- The more consistent the results, the higher the reliability
- Example:
  - "Will you vote for Trump in 2020?"
  - "On a scale from 0 (negative) to 100 (positive), what is your opinion of Trump?"
- Second question is likely to be less reliable.

Examples of Reliability Concerns

- *Converse (1964)* found that most people's opinions on issues as measured by survey questions appeared to vary randomly over time. His conclusion: people have "non-attitudes," are ignorant of even basic political issues.
- *Achen (1975)* argued this was actually a reliability problem – the apparent attitude instability was due to unreliable measures of political attitudes.
- Debate is still unresolved today.

Data Types

- **Numeric**:
  - Discrete (can be counted)
  - Continuous (can't be counted, i.e. decimals)
- **Categorical**:
  - generally not recorded as numbers
  - Party identification
- **Ordinal**:
  - categories with a specified order
  - think survey responses (strongly agree, somewhat agree,. . .)

Why do we care?

- Once we have measured variables for observations, we can conduct analysis!
- And once we have analysis, we can draw inferences and make evidence-based claims.

Now for some R...

Why R

- R is the most comprehensive statistical analysis package, as new technology and ideas often appear first in R.
- R is an open-source that's why you can run R anywhere any time, and even sell it under conditions of the license.
- It is cross-platform which runs on many operating systems. It's best for GNU/Linux and Microsoft Windows.
- In R, everyone is welcomed to provide bug fixes, code enhancements, and new packages.
- Jobs are nice

Section 1.3: Introduction to R

Numbers as data

- R can be used as a calculator

```
2+2
```

```
## [1] 4
```

- Everything you will use in R is saved in objects.
  - This can be everything from a number or a word to complex datasets
- These are equivalent:

```
x <- 2
x = 2
```

- Now `x` will return the number 2 whenever we write `x`

Numbers as data

- When you are working with scripts, try to save as much you can in objects, so you only need to change information once

```
y=x+7
y
```

```
## [1] 9
```

- Wrapping object in parenthesis tells R that we do not only want to save some information in the object y, but that we also want to see what is saved in y.

```
(y=x+7)
```

```
## [1] 9
```

More than one number

- Not limited to save only one number in an object.
- The code below will return a row of numbers from 1 to 10.

```
1:10
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10
```

Getting the Basics

- Don't forget to check and set your working directory
- R can't find files that aren't there

Arithmetic Operations

```
5 + 3
```

```
## [1] 8
```

```
5 - 3
```

```
## [1] 2
```

```
5 / 3
```

```
## [1] 1.666667
```

```
5 ^ 3
```

```
## [1] 125
```

```
5 * (10 - 3)
```

```
## [1] 35
```

```
sqrt(4)
```

```
## [1] 2
```

Objects

```
result <- 5 + 3
result
```

```
## [1] 8
```

```
print(result)
```

```
## [1] 8
```

```
result <- 5 - 3
result
```

```
## [1] 2
```

```
alex <- "instructor"
alex
```

```
## [1] "instructor"
```

```
alex <- "instructor and author"
alex
```

```
## [1] "instructor and author"
```

Objects

```
Result <- "5"
Result
```

```
## [1] "5"
```

```
result
```

```
## [1] 2
```

```
class(result)
```

```
## [1] "numeric"
```

```
Result
```

```
## [1] "5"
```

```
class(Result)
```

```
## [1] "character"
```

```
class(sqrt)
```

```
## [1] "function"
```

Vectors

```
world.pop <- c(2525779, 3026003, 3691173, 4449049, 5320817, 6127700, 6916183)
world.pop
```

```
## [1] 2525779 3026003 3691173 4449049 5320817 6127700 6916183
```

```
pop.first <- c(2525779, 3026003, 3691173)
pop.second <- c(4449049, 5320817, 6127700, 6916183)
pop.all <- c(pop.first, pop.second)
pop.all
```

```
## [1] 2525779 3026003 3691173 4449049 5320817 6127700 6916183
```

```
world.pop[2]
```

```
## [1] 3026003
```

```
world.pop[c(2, 4)]
```

```
## [1] 3026003 4449049
```

```
world.pop[c(4, 2)]
```

```
## [1] 4449049 3026003
```

```
world.pop[-3]
```

```
## [1] 2525779 3026003 4449049 5320817 6127700 6916183
```

Vectors

```
pop.million <- world.pop / 1000
pop.million
```

```
## [1] 2525.779 3026.003 3691.173 4449.049 5320.817 6127.700 6916.183
```

```
pop.rate <- world.pop / world.pop[1]
pop.rate
```

```
## [1] 1.000000 1.198047 1.461400 1.761456 2.106604 2.426063 2.738238
```

```
pop.increase <- world.pop[-1] - world.pop[-7]
percent.increase <- (pop.increase / world.pop[-7]) * 100
percent.increase
```

```
## [1] 19.80474 21.98180 20.53212 19.59448 15.16464 12.86752
```

```
percent.increase[c(1, 2)] <- c(20, 22)
percent.increase
```

```
## [1] 20.00000 22.00000 20.53212 19.59448 15.16464 12.86752
```

Functions

```
length(world.pop)
```

```
## [1] 7
```

```
min(world.pop)
```

```
## [1] 2525779
```

```
max(world.pop)
```

```
## [1] 6916183
```

```
range(world.pop)
```

```
## [1] 2525779 6916183
```

```
mean(world.pop)
```

```
## [1] 4579529
```

```
sum(world.pop) / length(world.pop)
```

```
## [1] 4579529
```

Functions

```
year <- seq(from = 1950, to = 2010, by = 10)
year
```

```
## [1] 1950 1960 1970 1980 1990 2000 2010
```

```
seq(to = 2010, by = 10, from = 1950)
```

```
## [1] 1950 1960 1970 1980 1990 2000 2010
```

```
seq(from = 2010, to = 1950, by = -10)
```

```
## [1] 2010 2000 1990 1980 1970 1960 1950
```

```
2008:2012
```

```
## [1] 2008 2009 2010 2011 2012
```

```
2012:2008
```

```
## [1] 2012 2011 2010 2009 2008
```

Functions

```
names(world.pop)
```

```
## NULL
```

```
names(world.pop) <- year
names(world.pop)
```

```
## [1] "1950" "1960" "1970" "1980" "1990" "2000" "2010"
```

```
world.pop
```

```
##    1950    1960    1970    1980    1990    2000    2010
## 2525779 3026003 3691173 4449049 5320817 6127700 6916183
```

Functions

```
## myfunction <- function(input1, input2, ..., inputN) {
##
##     DEFINE `output' USING INPUTS
##
##     return(output)
## }

my.summary <- function(x){ # function takes one input
  s.out <- sum(x)
  l.out <- length(x)
  m.out <- s.out / l.out
  out <- c(s.out, l.out, m.out) # define the output
  names(out) <- c("sum", "length", "mean") # add labels
  return(out) # end function by calling output
}
```

Functions

```
z <- 1:10
my.summary(z)
```

```
##    sum length   mean
##   55.0   10.0    5.5
```

```
my.summary(world.pop)
```

```
##       sum   length     mean
## 32056704        7  4579529
```

Data Files

```
# setwd("qss/INTRO")
# getwd()
```

Reading Files

- R can read files of many types and formats
- Usually, data are either in `.csv` or `.Rdata` format
- R can also read Excel spreadsheets with the `readxl` package

```
# Read in csv
# name_object <- read.csv("data_path/file.csv")
UNpop <- read.csv('C:/Users/afisher/Documents/R Code/qss/INTRO/UNpop.csv')

# Get type of object
class(UNpop)
```

```
## [1] "data.frame"
```

```
# Names of columns
names(UNpop)
```

```
## [1] "year"      "world.pop"
```

```
# Number of rows
nrow(UNpop)
```

```
## [1] 7
```

```
# Number of columns
ncol(UNpop)
```

```
## [1] 2
```

Grabbing Data

```r
# extract the column called "world.pop"
UNpop[, "world.pop"]
```

```
## [1] 2525779 3026003 3691173 4449049 5320817 6127700 6916
```

```r
# alternative way to extract column
UNpop$world.pop
```

```
## [1] 2525779 3026003 3691173 4449049 5320817 6127700 6916
```

```r
# extract the first three rows (and all columns)
UNpop[c(1, 2, 3),]
```

```
##   year world.pop
## 1 1950   2525779
## 2 1960   3026003
## 3 1970   3691173
```

Saving Objects

```
## save.image("qss/INTRO/Chapter1.RData")

## save(UNpop, file = "Chapter1.RData")
## save(world.pop, year, file = "qss/INTRO/Chapter1.RData")

## write.csv(UNpop, file = "UNpop.csv")

## load("Chapter1.RData")
```

Explore the data

- After loading the data and converting it into a tibble, one should inspect the data to get some understanding about the structure and content. Common funtions for these tasks are:
- `<name-of-data-tibble>`: Display the first 10 rows and all columns that fit on one screen. It also prints an abbreviated description of the column type.
- `head(<name-of-df>)`, `tail(<name-of-df>)`: Return the first or last part. Use these commands if it is not a tibble but a data frame
- `dim()`: Retrieve the dimension
- `names()`: Get the names

Explore the data

- `str()`: Display compactly the internal structure
- `glimpse()`: is the dplyr-version of str() showing values of each variable the whole sceen width, but does not display the number of levels and names of factor variables. But this feature of str() cannot be displayed completly with either many or long levels names.
- `View()`: With RStudio you can see and inspect the data set comfortably. The View() function invokes a spreadsheet-style data viewer.

Install Packages

- When you download R from the Comprehensive R Archive Network (CRAN), you get that "base" R system
- The base R system comes with basic functionality; implements the R language
- One reason R is so useful is the large collection of packages that extend the basic functionality of R
- R packages are developed and published by the larger R community

Install Packages

- Packages can be installed with the `install.packages()` function in R
- To install a single package, pass the name of the lecture to the install.packages() function as the first argument
- You can install multiple R packages at once with a single call to install.packages()
- `install.packages(c("dplyr", "ggplot2", "devtools"))`

R Scripts

- Why should we write scripts?
  - Save time: automate boring tasks
  - Reproducibility
  - Allow complex tasks to be performed in small steps
  - Faster to run

Loading R Packages

- Installing a package does not make it immediately available to you in R; you must load the package
- The library() function is used to load packages into R
- The following code is used to load the ggplot2 package into R

```
library(ggplot2)
```

- NOTE: Do not put the package name in quotes!

Adding Labels to our plots

- To add a title to your plot, add the code:
  - `+ggtitle("Your Title Here")` to your line of basic ggplot code.
- Note: You can also use:
  - `+labs(title = "Title")‘`
- To alter the labels on the axis, add the code:
  - `+labs(y= "y axis name", x = "x axis name")`
- Can also use:
  - `+xlab("x axis name"` and `+ylab("y axis name")`

Looking at the Data (dim and glimpse)

```
## Cyanide and Happiness 2017 Politics Poll
```

Looking at the Data (head)

```r
# head
head(cah) # first 6 observations
```

```
##   Income Gender Age AgeRange PoliticalAffiliation ApproveTrump       education
## 1 192000 Female  35    35-44    Strong Republican       DK/REF  College degree
## 2  54000 Female  58    55-64           Independent   Disapprove    Some college
## 3  20000   Male  50    45-54  Not Strong Democrat      Approve           Other
## 4  21000 Female  40    35-44           Independent   Disapprove  College degree
## 5 164000 Female  42    35-44      Strong Democrat       DK/REF Graduate degree
## 6   9000 Female  35    35-44      Strong Democrat   Disapprove     High school
##     race AgreeWhiteNationalists RepublicansAgreeWhiteNationalists
## 1  White                 DK/REF                                NA
## 2  White                 DK/REF                                NA
## 3  White                  Agree                                10
## 4  White                 DK/REF                                NA
## 5  Black                 DK/REF                                NA
## 6 Latino                  Agree                                50
##   Would.you.say.that.you.love.America. DemocratsLoveAmerica GovHelpPoor
## 1                                  Yes                   40         Yes
## 2                                  Yes                   80         Yes
## 3                                  Yes                   60         Yes
## 4                                  Yes                   30         Yes
## 5                                  Yes                   NA         Yes
## 6                                  Yes                  100         Yes
##   WhitePeopleRacist CivilWarNextDecade hunting kalesalad VoteTheRockPres
## 1                No           Unlikely      No        No              No
## 2                No             Likely      No        No             Yes
## 3               Yes             Likely     Yes        No             Yes
## 4               Yes             Likely      No       Yes              No
## 5           DK/REF           Unlikely      No       Yes             Yes
```