Data meets Law

Understanding Political Numbers

April 17, 2019

But first, some review and tips

Rachael Meager
@economeager

I would say a good 60% of statistical programming time for data processing scripts is just emotionally metabolizing the existential grief about how dumb everything is

♡ 650 9:44 PM - Apr 15, 2019

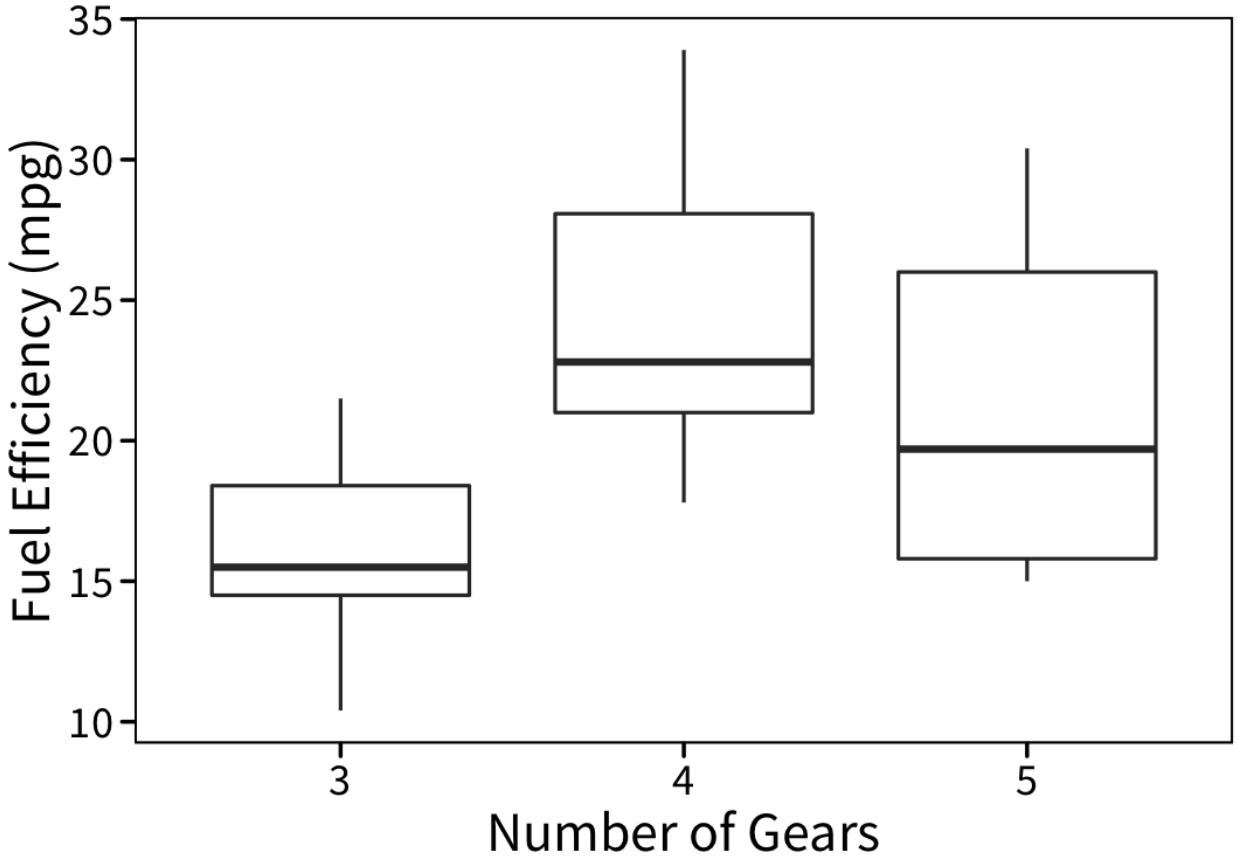
102 people are talking about this

Review: unit of analysis

A dataset has ONE unit of analysis for *all variables*

Variables describe the units

Dummy variables from multiple categories



Dummy variables from multiple categories

Slow but careful: make individual dummies

```
new_cars <- mtcars %>%
  mutate(three_gears = case_when(gear == 3 ~ 1,
                                 TRUE ~ 0),
         five_gears = case_when(gear == 5 ~ 1,
                                TRUE ~ 0))

# four gears are excluded category
dummy_mod <- lm(mpg ~ three_gears + five_gears,
                data = new_cars)

# yhat = a + b1(three_gears) + b2(five_gears)
tidy(dummy_mod)
```

```
## # A tibble: 3 x 5
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)    24.5       1.36     18.1  2.59e-17
## 2 three_gears    -8.43      1.82     -4.62 7.26e- 5
## 3 five_gears     -3.15      2.51     -1.26 2.18e- 1
```

Dummy variables from multiple categories

Fast and scary: use factor variables

```
fct_model <- lm(mpg ~ as.factor(gear), data = mtcars)

tidy(fct_model)
```

```
## # A tibble: 3 x 5
##   term              estimate std.error statistic  p.value
##   <chr>                <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)         16.1       1.22     13.2  7.87e-14
## 2 as.factor(gear)4     8.43      1.82      4.62 7.26e- 5
## 3 as.factor(gear)5     5.27      2.43      2.17 3.84e- 2
```

Dummy variables from multiple categories

Fast and scary: use factor variables

```
fct_model <- lm(mpg ~ as.factor(gear), data = mtcars)

tidy(fct_model)
```

```
## # A tibble: 3 x 5
##   term             estimate std.error statistic  p.value
##   <chr>               <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)         16.1       1.22     13.2  7.87e-14
## 2 as.factor(gear)4     8.43      1.82      4.62 7.26e- 5
## 3 as.factor(gear)5     5.27      2.43      2.17 3.84e- 2
```

R decides which category to exclude alphanumerically. To control which category is dropped, create (`as.factor()`) and reorder the factor (`fct_relevel()`) beforehand

Dummy variables from multiple categories

Augmenting works but makes me nervous

```
pre_augment_data <-
  tibble(gear = c(3, 4, 5)) %>%
  print()
```

```
## # A tibble: 3 x 1
##    gear
##   <dbl>
## 1     3
## 2     4
## 3     5
```

Dummy variables from multiple categories

Augmenting works but makes me nervous

```
pre_augment_data <-
  tibble(gear = c(3, 4, 5)) %>%
  print()
```

```
## # A tibble: 3 x 1
##    gear
##   <dbl>
## 1     3
## 2     4
## 3     5
```

```
# augment knows that 'gear' should be
#   interpreted as a series of dummies
augment(fct_model, newdata = pre_augment_data)
```

```
## # A tibble: 3 x 3
##    gear .fitted .se.fit
##   <dbl>   <dbl>   <dbl>
## 1     3    16.1    1.22
## 2     4    24.5    1.36
## 3     5    21.4    2.11
```

Evidence (Law) and Evidence (Science)

Outline of legal inquiry

What's the question?

Outline of legal inquiry

What's the question?

What facts bear on that question?

Outline of legal inquiry

What's the question?

What facts bear on that question?

What evidence is there in favor of/against those facts?

Outline of legal inquiry

What's the question?

What facts bear on that question?

What evidence is there in favor of/against those facts?

Other issues of procedure:

- demonstration of harm/standing to sue
- is the question "justiciable"

Outline of legal inquiry

What's the question?

What facts bear on that question?

What evidence is there in favor of/against those facts?

Other issues of procedure:

- demonstration of harm/standing to sue
- is the question "justiciable"

How different from science?

Is Political Science Relevant? Ask an Expert Witness

Kenneth R. Mayer

Abstract

Working as an expert witness is one of the most direct ways for a political scientist to affect actual policy. However, the worlds of academia and litigation are vastly different environments, with different norms, goals, and rules. I explore some of these differences, as well as the implications for straddling the two worlds as an expert. Ultimately, the parties in election law, redistricting, and voting rights litigation have found political scientists to be valuable partners. It is a very satisfying experience to find that our methods and skills have value outside the friendly confines of the ivory tower.

What's the question?

When the legal question isn't the same as the *interesting* question

When the legally justifiable view isn't the same as the *best* view

Equal population counts for Congressional districts

> During one of my first experiences as an expert a decade ago, I gave a fluent and informed disquisition on how the flaws in census data mean that there should be more latitude permitted in drawing districts with equal population. The known biases in the data and the passage of time between the collection of data in April and the drawing of the districts two years later meant that it was simply misleading to insist on equal population: there were errors in the original data, people had moved in the years since the census data were recorded, and they would continue to do so between then and the subsequent election.

Equal population counts for Congressional districts

> During one of my first experiences as an expert a decade ago, I gave a fluent and informed disquisition on how the flaws in census data mean that there should be more latitude permitted in drawing districts with equal population. The known biases in the data and the passage of time between the collection of data in April and the drawing of the districts two years later meant that it was simply misleading to insist on equal population: there were errors in the original data, people had moved in the years since the census data were recorded, and they would continue to do so between then and the subsequent election.
>
> It was, therefore, unnecessary to minimize population differences among districts, and we should be willing to live with larger population deviations as a tradeoff for other key redistricting criteria (compactness and respect for existing political boundaries, for example). All of which would have been a terrific lecture, about the nature of representation, the problems of relying on artificial bright-lines, and the tension among the many goals of redistricting. Yet when I had finished, the attorneys looked at me as if I had been talking about the merits of the designated hitter rule in Major League Baseball. Even though what I said was true, my argument and conclusion were unrelated to the question at hand, and the issue did not come up again.

Equal population counts for Congressional districts

> […] In an academic setting, by contrast, there is an ongoing debate over the accuracy of the census data and the implications of different types of errors and counting rules. The bias and errors in the census process have been well known for years, although for apportionment purposes, the Bureau itself is prevented from correcting the data using statistical techniques such as sampling. The Census Bureau estimates that the 1990 census had an undercount of between 4.0 and 5.3 million. The error in the 2000 census was estimated at between an undercount of 3.3 million and an overcount of 1.8 million.

Equal population counts for Congressional districts

> [...] In an academic setting, by contrast, there is an ongoing debate over the accuracy of the census data and the implications of different types of errors and counting rules. The bias and errors in the census process have been well known for years, although for apportionment purposes, the Bureau itself is prevented from correcting the data using statistical techniques such as sampling. The Census Bureau estimates that the 1990 census had an undercount of between 4.0 and 5.3 million. The error in the 2000 census was estimated at between an undercount of 3.3 million and an overcount of 1.8 million.

Courts deal with narrow questions, not always *best* questions

Methods evolve to address the narrow questions

"Majority-Minority" Congressional Districts

> A second example is the generally accepted judicial rule for establishing the existence of racial-bloc voting, or the degree to which members of minority communities vote as a bloc for candidates of the same racial or ethnic background. Racial bloc voting is one necessary condition for a finding that minorities must have opportunities to elect candidates of their choice, and that some legislative districts must therefore be drawn to provide such opportunities (usually by drawing districts so that they contain a threshold number of minority voters, historical between 55 and 60%). But most techniques for identifying racial bloc voting are cumbersome and error-prone, because while we have aggregate outcomes, we lack any information about the voting behavior of individual voters. This is the classic “ecological inference” problem: how do we infer the behavior of individuals when all we have is aggregate data?

"Majority-Minority" Congressional Districts

> A second example is the generally accepted judicial rule for establishing the existence of racial-bloc voting, or the degree to which members of minority communities vote as a bloc for candidates of the same racial or ethnic background. Racial bloc voting is one necessary condition for a finding that minorities must have opportunities to elect candidates of their choice, and that some legislative districts must therefore be drawn to provide such opportunities (usually by drawing districts so that they contain a threshold number of minority voters, historical between 55 and 60%). But most techniques for identifying racial bloc voting are cumbersome and error-prone, because while we have aggregate outcomes, we lack any information about the voting behavior of individual voters. This is the classic “ecological inference” problem: how do we infer the behavior of individuals when all we have is aggregate data?
>
> In a landmark voting rights case, *Thornburg v. Gingles* 478 U.S. 30 (1986), the Supreme Court specifically endorsed a technique—usually called "double regression" or "bivariate ecological regression"—to show the existence of racially polarized voting. But the Court’s preferred method has well-known shortcomings: it relies on some very specific assumptions which often do not hold, and in those instances can produce absurd results.

"Majority-Minority" Congressional Districts

> The debate over the proper method of approaching the ecological inference problem is robust and healthy, and has already led to better methods. But that debate is only a hindrance to a judge, who needs the answer now and will have little patience with arguments over which method is superior. The Supreme Court says double regression is the right method, and double regression it therefore is (although a good expert witness will confirm those results with other methods, or attempt to show that the assumptions are wrong and that other methods will produce better results)

Legal standards & precedent more *more slowly* than research

Standards of evidence

The data world

- What is probably true?
- Diverse methods for reaching conclusions
- "If we knew what we were doing, it wouldn't be called 'research'"

Standards of evidence

The data world

- What is probably true?
- Diverse methods for reaching conclusions
- "If we knew what we were doing, it wouldn't be called 'research'"

The legal world

- What is there *hard evidence* for?
- Probabilities aren't *positive* evidence
- Innocent until *proven guilty*, not "did they probably do it?"

Honesty and (vs?) Advocacy

> While a Ph.D. is taught to subject his or her favored hypothesis to every conceivable test and data source, seeking out all possible evidence against his or her theory, an attorney is taught to amass all of the evidence for his or her hypothesis and distract attention from anything that might be seen as contradictory information. An attorney who treats a client like a hypothesis would be disbarred; a Ph.D. who advocates a hypothesis like a client would be ignored (Epstein and King 2002)

Honesty and (vs?) Advocacy

Averages (systematic data) vs. Anecdotes ("key cases", witnesses)

Honesty and (vs?) Advocacy

Averages (systematic data) vs. Anecdotes ("key cases", witnesses)

Selection biases in the data-generating process

Honesty and (vs?) Advocacy

Averages (systematic data) vs. Anecdotes ("key cases", witnesses)

Selection biases in the data-generating process

Problem of Induction, verification vs. falsification

Honesty and (vs?) Advocacy

Averages (systematic data) vs. Anecdotes ("key cases", witnesses)

Selection biases in the data-generating process

Problem of Induction, verification vs. falsification

Causality, *post hoc ergo propter hoc*

SUPREME COURT OF CALIFORNIA

THE PEOPLE *v.* MALCOLM RICARDO COLLINS

APPEAL FROM A JUDGMENT OF THE SUPERIOR COURT OF LOS ANGELES COUNTY

Crim. No. 11176. Decided March 11, 1968

JUSTICE SULLIVAN delivered the opinion of the Court, in which CHIEF JUSTICE TRAYNOR and JUSTICES PETERS, TOBRINER, MOSK, and BURKE concurred.

We deal here with the novel question whether evidence of mathematical probability has been properly introduced and used by the prosecution in a criminal case. While we discern no inherent incompatibility between the disciplines of law and mathematics and intend no general disapproval or disparagement of the latter as an auxiliary in the fact-finding processes of the former, we cannot uphold the technique employed in the instant case. As we explain in detail, *infra*, the testimony as to mathematical probability infected the case with fatal error and distorted the jury's traditional role of determining guilt or innocence according to long-settled rules. Mathematics, a veritable sor-

The facts:

- Elderly woman robbed in an LA alleyway
- Suspect: woman with blond ponytail
- Suspect flees scene in yellow car
- Driven by black man with a mustache

The facts:

- Elderly woman robbed in an LA alleyway
- Suspect: woman with blond ponytail
- Suspect flees scene in yellow car
- Driven by black man with a mustache

Couple arrested

- Black man, white woman
- Woman: white w/ blond ponytail.
- Man: black w/ beard
- Yellow car

The facts:

- Elderly woman robbed in an LA alleyway
- Suspect: woman with blond ponytail
- Suspect flees scene in yellow car
- Driven by black man with a mustache

Couple arrested

- Black man, white woman
- Woman: white w/ blond ponytail.
- Man: black w/ beard
- Yellow car

Mathematician's testimony: the particular combination of features is 1 in 12 million

The facts:

- Elderly woman robbed in an LA alleyway
- Suspect: woman with blond ponytail
- Suspect flees scene in yellow car
- Driven by black man with a mustache

Couple arrested

- Black man, white woman
- Woman: white w/ blond ponytail.
- Man: black w/ beard
- Yellow car

Mathematician's testimony: the particular combination of features is 1 in 12 million

> [...] the testimony and the manner in which the prosecution used it distracted the jury from its proper and requisite function of weighing the evidence on the issue of guilt, encouraged the jurors to rely upon an engaging but logically irrelevant expert demonstration

The facts:

- Elderly woman robbed in an LA alleyway
- Suspect: woman with blond ponytail
- Suspect flees scene in yellow car
- Driven by black man with a mustache

Couple arrested

- Black man, white woman
- Woman: white w/ blond ponytail.
- Man: black w/ beard
- Yellow car

Mathematician's testimony: the particular combination of features is 1 in 12 million

> [...] the testimony and the manner in which the prosecution used it distracted the jury from its proper and requisite function of weighing the evidence on the issue of guilt, encouraged the jurors to rely upon an engaging but logically irrelevant expert demonstration

Probability is not the same as positive evidence

Problems with probability

A suspect says something fishy, x

Problems with probability

A suspect says something fishy, x

$$p(x \mid \text{Guilty}) > p(x \mid \text{Not Guilty})$$

Problems with probability

A suspect says something fishy, x

$$p(x \mid \text{Guilty}) > p(x \mid \text{Not Guilty})$$

What are we actually trying to calculate?

$$p(\text{Guilty} \mid x) > p(\text{Not Guilty} \mid x)$$

Problems with probability

A suspect says something fishy, x

$$p(x \mid \text{Guilty}) > p(x \mid \text{Not Guilty})$$

What are we actually trying to calculate?

$$p(\text{Guilty} \mid x) > p(\text{Not Guilty} \mid x)$$

What's the probability that they are guilty, *given that* they share features with the suspects?

$$p(\text{Guilty} \mid \text{Description}) = \frac{p(\text{Description} \mid \text{Guilty})\, p(\text{Guilty})}{p(\text{Description})}$$

Problems with probability

A suspect says something fishy, x

$$p(x \mid \text{Guilty}) > p(x \mid \text{Not Guilty})$$

What are we actually trying to calculate?

$$p(\text{Guilty} \mid x) > p(\text{Not Guilty} \mid x)$$

What's the probability that they are guilty, *given that* they share features with the suspects?

$$p(\text{Guilty} \mid \text{Description}) = \frac{p(\text{Description} \mid \text{Guilty})\, p(\text{Guilty})}{p(\text{Description})}$$

Is "probability of guilt" even the right question?

Problems with probability

A suspect says something fishy, x

$$p(x \mid \text{Guilty}) > p(x \mid \text{Not Guilty})$$

What are we actually trying to calculate?

$$p(\text{Guilty} \mid x) > p(\text{Not Guilty} \mid x)$$

What's the probability that they are guilty, *given that* they share features with the suspects?

$$p(\text{Guilty} \mid \text{Description}) = \frac{p(\text{Description} \mid \text{Guilty})\, p(\text{Guilty})}{p(\text{Description})}$$

Is "probability of guilt" even the right question?

Julia Galef: Visual Demonstration of Bayesian Updating

G'bye