Week 1, Class 1: Practice Exercises

Introduction to Quantitative Political Analysis

2024-12-31

1 Non-AI Exercises

1.1 1. Vocabulary & Concepts

1.1.1 1.1 Code Detective

Explain what each line of this code does:

```
electoral_votes <- 270
winner <- "Biden"
margin <- 81283501 - 74223975
percentage <- margin / 155507476 * 100
```

Line 1: _____ Line 2: _____ Line 3: _____ Line 4: _____

1.2 2. Historical Example: John Snow

1.2.1 2.1 The conventional wisdom in 1854 was that cholera spread through:

a) Contaminated water
b) Bad air (miasma)
c) Person-to-person contact
d) Poor nutrition

Answer: _____

1.2.2 2.2 Data-Driven Decision Making

John Snow challenged the conventional wisdom about cholera transmission. What made his approach "quantitative" rather than just observational? Why was mapping the data crucial to his discovery?

Answer:

1.3 3. Critical Thinking: AI and Analysis

1.3.1 3.1 Critical Thinking with AI

Why is it important to verify AI-generated analysis rather than accepting it automatically? Give an example of how an AI might produce technically correct code that leads to a misleading conclusion.

Answer:

1.4 4. Applications of Quantitative Political Analysis

2 AI Exercises

For each AI exercise: - Write your prompt in the designated space - Record Claude's response - Run the code and document the results

2.1 5. Introduction to Political Data

Dataset: nat_pol_attitudes.csv

Description: Simulates a nationally representative survey measuring political attitudes, ideology, and demographics.

Variables: - `respondent_id`: Unique respondent ID (int) - `age`: Age in years, 18-90 (int) - `gender`: male, female, nonbinary (factor) - `race_ethnicity`: White, Black, Latino, Asian, Other (factor) - `education`: Less than HS, HS, Some College, BA, Postgrad (ordered) - `income_bracket`: Ten brackets from <$10k to >$200k (ordered) - `ideology`: 1 (very liberal) to 7 (very conservative) (int) - `party_id`: Democrat, Republican, Independent, Other (factor) - `trust_gov`: 0-10 political trust scale (int) - `policy_support_env`: Support for environmental regulation, 0/1 (binary) - `policy_support_guns`: Support for stricter gun laws, 0/1 (binary)

2.1.1 5.1 Data Exploration

```
# Load the dataset
nat_pol_attitudes <- read_csv("nat_pol_attitudes.csv")
```

```
Rows: 1200 Columns: 11
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (3): gender, race_ethnicity, party_id
dbl (8): respondent_id, age, education, income_bracket, ideology, trust_gov,...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Explore the structure of the data
glimpse(nat_pol_attitudes)
```

```
Rows: 1,200
Columns: 11
$ respondent_id       <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,~
$ age                 <dbl> 36, 34, 32, 36, 44, 41, 81, 40, 60, 73, 69, 23, 87~
$ gender              <chr> "male", "female", "female", "female", "female", "f~
$ race_ethnicity      <chr> "White", "White", "Latino", "Other", "White", "Whi~
$ education           <dbl> 4, 5, 2, 2, 4, 1, 5, 3, 1, 3, 4, 2, 4, 1, 4, 4, 5,~
$ income_bracket      <dbl> 2, 10, 5, 10, 2, 1, 7, 4, 4, 7, 8, 7, 9, 10, 4, 6,~
$ ideology            <dbl> 5, 5, 3, 2, 6, 6, 5, 4, 4, 1, 3, 2, 4, 4, 6, 1, 3,~
$ party_id            <chr> "Republican", "Independent", "Independent", "Repub~
$ trust_gov           <dbl> 2, 2, 5, 0, 5, 6, 5, 4, 4, 4, 2, 4, 9, 3, 3, 4, 0,~
$ policy_support_env  <dbl> 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1,~
$ policy_support_guns <dbl> 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0,~
```

Using Claude, explore this political attitudes dataset. Ask it to help you understand if ideology varies by income bracket. (Bonus: Try to do this without using Claude.)

Prompt: I am working with a dataset in R called nat_pol_attitudes.csv. I am usin tidyverse. I want to know if ideology varies by income bracket. Write code to explore this and explain each step.

Response:

Interpretation:

2.1.2 5.2 Basic Summary Statistics

Work with Claude to calculate the average age and political trust score by party affiliation. Record both your prompt and the code Claude provides. (Bonus: Try to do this without using Claude.)

Prompt:

Response:

Interpretation:

2.1.3 5.3 Understanding Relationships

Ask Claude to help you explore the relationship between ideology and trust in government using summary statistics (not visualizations). What patterns do you discover? (Bonus: Try to do this without using Claude.)

Prompt:

Response:

Interpretation:

2.2 6. Understanding Election Data

Dataset: precinct_elections.csv

Description: Precinct-level election returns with demographics.

Variables: - `precinct_id`: Unique precinct identifier (int) - `county_name`: County name (string) - `state`: Two-letter state abbreviation (factor) - `total_votes`: Total votes cast (int) - `dem_votes`: Democratic candidate votes (int) - `rep_votes`: Republican candidate votes (int) - `third_votes`: Third-party votes (int) - `registered_voters`: Number of registered voters (int) - `median_income`: Precinct median household income (num) - `pct_college`: Percentage with college degree (num) - `pct_white`: Percentage White population (num) - `pct_urban`: Percentage urban population (num)

2.2.1 6.1 Loading and Initial Analysis

```
# Load the dataset
precinct_elections <- read_csv("precinct_elections.csv")
```

```
Rows: 3000 Columns: 12
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (3): state, county, precinct_id
dbl (9): year, reg_voters, turnout, dem_votes, rep_votes, median_income, pct...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Examine the data
glimpse(precinct_elections)
```

```
Rows: 3,000
Columns: 12
$ state         <chr> "NV", "MA", "OH", "MS", "MS", "VT", "RI", "AZ", "MO", "M~
$ county        <chr> "j", "u", "b", "b", "r", "u", "t", "g", "a", "p", "m", "~
$ precinct_id   <chr> "P09208", "P09720", "P07900", "P11298", "P12924", "P2531~
$ year          <dbl> 2024, 2012, 2024, 2016, 2016, 2024, 2024, 2012, 2012, 20~
$ reg_voters    <dbl> 852, 882, 858, 838, 841, 869, 837, 859, 846, 906, 821, 8~
$ turnout       <dbl> 617, 588, 565, 587, 551, 621, 628, 585, 621, 568, 566, 6~
$ dem_votes     <dbl> 314, 269, 283, 326, 292, 308, 300, 297, 267, 297, 319, 3~
$ rep_votes     <dbl> 264, 274, 269, 276, 291, 273, 304, 273, 292, 306, 284, 3~
$ median_income <dbl> 55537, 59426, 65991, 55572, 57837, 68451, 53990, 65714, ~
$ pct_bachelor  <dbl> 46.4, 44.5, 27.0, 36.2, 20.9, 32.0, 11.7, 39.1, 15.0, 65~
$ race_black    <dbl> 48.2, 55.3, 69.7, 5.7, 38.3, 57.6, 66.7, 61.2, 49.7, 15.~
$ race_hispanic <dbl> 73.0, 58.8, 8.7, 23.9, 71.5, 9.7, 44.3, 57.7, 58.4, 25.6~
```

Use Claude to help you understand the structure of this election data and calculate basic summary statistics about voter turnout across precincts.

2.2.2 6.2 Calculating Turnout

Ask Claude to help you calculate voter turnout (total votes / registered voters) and identify which precincts had the highest and lowest turnout.

Prompt:

Response:

Interpretation:

2.3 7. Congressional Data Analysis

Dataset: congress_press.csv

Description: Corpus of press releases issued by U.S. legislators.

Variables: - `release_id`: Unique press-release ID (int) - `member_id`: Legislator ID (int) - `chamber`: House, Senate (factor) - `party`: Democrat, Republican, Independent (factor) - `ideology_score`: DW-NOMINATE first dimension (num) - `state`: Two-letter abbreviation (factor) - `date`: Release date (date) - `topic`: Ten topics like Health, Economy, Foreign Policy, etc. (factor) - `sentiment_score`: -1 to 1 sentiment scale (num) - `contains_attack`: Indicator of partisan attack language (binary)

2.3.1 7.1 Understanding Press Release Patterns

```
# Load the dataset
congress_press <- read_csv("congress_press.csv")
```

```
Rows: 2200 Columns: 10
-- Column specification --------------------------------------------------------
Delimiter: ","
chr  (4): chamber, party, state, topic
dbl  (5): release_id, member_id, ideology_score, sentiment_score, contains_a...
date (1): date

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Look at the data structure
glimpse(congress_press)
```

```
Rows: 2,200
Columns: 10
$ release_id      <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,~
$ member_id       <dbl> 91, 112, 123, 467, 113, 38, 192, 1, 215, 213, 252, 523~
```

```
$ chamber         <chr> "House", "House", "House", "House", "House", "House", ~
$ party           <chr> "Democrat", "Republican", "Democrat", "Republican", "D~
$ ideology_score  <dbl> 1.00833230, 0.03292990, -0.29090319, -0.18102718, 0.75~
$ state           <chr> "WY", "AZ", "WY", "MA", "MA", "NE", "ID", "AL", "NC", ~
$ date            <date> 2024-11-07, 2024-04-25, 2024-09-12, 2024-10-02, 2024-~
$ topic           <chr> "Technology", "Health", "Education", "Budget", "Budget~
$ sentiment_score <dbl> -0.20, 0.12, 0.18, -0.16, 0.35, -0.44, -0.19, 0.60, 0.~
$ contains_attack <dbl> 1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, ~
```

Work with Claude to explore patterns in Congressional press releases. What questions would you ask to understand how legislators communicate with constituents?

2.3.2 7.2 Party Differences

Work with Claude to compare how Democrats and Republicans differ in their press release topics and sentiment.

Prompt:

Resposne:

Interpretation:

2.4 8. Economic Indicators and Politics

Dataset: county_econ.csv

Description: Balanced panel of U.S. counties, 2010-2020, with economic & demographic metrics.

Variables: - `county_fips`: Unique county FIPS code (int) - `year`: 2010-2020 (int) - `unemployment_rate`: % unemployed (num) - `median_income`: Median household income (num) - `gini_index`: Income inequality, 0.2-0.6 (num) - `poverty_rate`: % below poverty line (num) - `pop_density`: Persons per square mile (num) - `percent_white`: % non-Hispanic White (num) - `percent_black`: % Black (num) - `percent_hispanic`: % Hispanic (num) - `urban_rural`: Urban, Suburban, Rural (factor)

2.4.1 8.1 Economic Trends Over Time

```
# Load the dataset
county_econ <- read_csv("county_econ.csv")
```

```
Rows: 3500 Columns: 11
-- Column specification --------------------------------------------------------
Delimiter: ","
chr  (1): urban_rural
dbl (10): county_fips, year, unemployment_rate, median_income, gini_index, p...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Examine the panel structure
glimpse(county_econ)
```

```
Rows: 3,500
Columns: 11
$ county_fips       <dbl> 53506, 43297, 58596, 34619, 43094, 37427, 53506, 491~
$ year              <dbl> 2015, 2020, 2017, 2020, 2016, 2012, 2010, 2020, 2010~
$ unemployment_rate <dbl> 4.2, 8.7, 7.3, 4.9, 2.4, 12.2, 3.3, 2.4, 12.4, 11.4,~
$ median_income     <dbl> 63262, 61006, 48323, 55327, 66234, 57395, 55070, 462~
$ gini_index        <dbl> 0.35, 0.49, 0.36, 0.54, 0.38, 0.51, 0.40, 0.54, 0.45~
$ poverty_rate      <dbl> 15.1, 24.1, 26.0, 5.7, 30.0, 25.6, 25.3, 22.5, 27.4,~
$ pop_density       <dbl> 201.3, 229.9, 123.7, 372.2, 273.8, 522.0, 266.8, 94.~
$ urban_rural       <chr> "Urban", "Rural", "Rural", "Urban", "Urban", "Suburb~
$ percent_white     <dbl> 84.4, 79.1, 21.3, 34.1, 73.2, 71.7, 68.8, 41.1, 41.2~
$ percent_black     <dbl> 3.1, 56.3, 32.7, 3.6, 19.5, 33.3, 42.8, 5.9, 51.5, 1~
$ percent_hispanic  <dbl> 23.7, 17.8, 49.2, 56.1, 47.3, 10.4, 32.7, 52.9, 14.9~
```

This is panel data (same counties observed over multiple years). Ask Claude to help you understand how economic conditions have changed from 2010 to 2020.

2.4.2 8.2 Urban vs Rural Differences

Ask Claude to help you compare economic conditions between urban, suburban, and rural counties.

Prompt:

Resposne:

Interpretation:

2.4.3 8.3 Creating a Summary Report

Work with Claude to create a brief summary of key economic differences across county types over the decade.

Prompt:

Resposne:

Interpretation: