

PA 446: Coding for Civic Data Applications
Fall 2025

Midterm Exam

Total Points: 116 + 12(Bonus)

You are required to submit the midterm exam on October 19th, 2025, at 11:59 PM on BlackBoard. Please read through the entire exam before attempting the questions.

Submissions to be done on BlackBoard

1. A PDF document containing
 - a. Link to the GitHub repository containing all associated files used to create the webpage. (Part I)
 - b. Code and results/answers for Part II
 - c. Code and results/answers for Part III
2. A .rmd file created to make the webpage (submitted separately)

PART I: GitHub using R Markdown [30 points]

1. Create a webpage (GitHub Page using R Markdown) that:
 - a. Contains an .rmd answering the following question:
 - i. Describe the civic problem and its implications for the community. Be as specific as possible. (*Ex. food wastage*) [300 words] [20 pts]
 - b. Has a theme of your choice to make the webpage attractive. [10 pts]

Part II: Web scraping [20 points]

Go to the website <https://www.scrapethissite.com/pages/simple/> and scrape the data to create a table with four variables: Country, Capital, Population, and Area. The table will have a total of 250 observations.

PART III: API Access [62 points]

Public administrators often use Census data to understand how civic resources - such as income, broadband, and education - vary across regions. The American Community Survey (ACS) provides detailed demographic and socioeconomic data accessible through the tidycensus package, which connects directly to the Census API.

You will use tidycensus and tidyverse tools to explore income inequality and the digital divide across Illinois counties.

Step 1: Identify Relevant Variables [6 points]

Go to the ACS 5-Year Data Dictionary and find the variables that represent the following. I am providing you with one of the variables. Find the remaining two and fill in the blanks.

Hint: To confirm which variables to use, use

<https://api.census.gov/data/2023/acs/acs5/groups/B19013.html>. Replace "B19013" with relevant names for other variables.

- Median household income (in the past 12 months): _____
- Households with broadband Internet: _____
- Total households with any type of internet access: *B28002_001E*

These variables will help you measure both income and digital access across Illinois counties.

Step 2: Retrieve Data [8 points]

Use `get_acs()` to retrieve the following variables for *all counties in Illinois*:

- Median household income (in the past 12 months)
- Households with broadband Internet
- Total households with any type of internet access

Step 3: Clean and Transform Data [14 + 4 points]

- Use `pivot_wider()` to create one row per county (with income and broadband side-by-side). [6 points]
 - **BONUS:** Split "County, Illinois" into County and State [4 points]
- Calculate a new variable: `broadband_rate = broadband/total_households * 100` [4 points]
- Arrange counties from highest to lowest broadband access. [4 points]

HINT: Your final clean data will look something like this:

	county	state	income	total_households	broadband	broadband_rate
1	Kendall County	Illinois	110474	44526	42382	95.18484
2	McHenry County	Illinois	102836	116329	110646	95.11472
3	DuPage County	Illinois	110502	349497	329798	94.36361
4	Lake County	Illinois	108917	256660	240027	93.51944
5	Will County	Illinois	107799	241310	224935	93.21412
6	DeKalb County	Illinois	69022	39314	36455	92.72778
7	Kane County	Illinois	100678	183196	169377	92.45671
8	Monroe County	Illinois	101635	13830	12654	91.49675
9	Grundy County	Illinois	93060	20518	18549	90.40355
10	Madison County	Illinois	74800	109385	98374	89.93372
11	Effingham County	Illinois	75380	14030	12595	89.77192
12	Clinton County	Illinois	82314	14463	12972	89.69094

Step 4: Analyze patterns [8 points]

- a) Compute the mean and median broadband rate across all Illinois counties. [4 points]
- b) Identify the top 5 counties with the highest broadband access and the bottom 5 counties with the lowest. [4 points]

Step 5: Visualize the results [14 + 8 points]

1. **Scatterplot:** Income (x-axis) vs Broadband Rate (y-axis). [8 points]
 - a. Add a regression line using `geom_smooth(method = "lm")`.
 - b. Label axes and add an informative title.
2. **Bar chart:** Show the top 10 and bottom 10 counties by broadband access. HINT: Use the `reorder` function to order bars. [6 points]
3. **BONUS:** Create a single ordered bar chart of ALL counties by broadband [8 points]

Step 6: Reflection [16 points]

Write 3-4 sentences per question:

1. What patterns do you observe between income and broadband access? [4 points]
2. What might explain the variation in broadband access across counties? [4 points]
3. How could public administrators use this data to inform digital inclusion policies? [4 points]
4. What are some limitations of using ACS data for local decision-making? [4 points]