30538 Problem Set3: Parking Tickets

Peter Ganong, Maggie Shi and Alex Lan

2026-01-25

1. **PS3:** Due Jan 31 at 5:00PM Central.

"This submission is my work alone and complies with the 30538 integrity policy." Add your initials to indicate your agreement: **___**

Github Classroom Assignment Setup and Submission Instructions

1. **Accepting and Setting up the PS3 Assignment Repository**

   - Each student must individually accept the repository for the problem set from Github Classroom ("ps3") – https://classroom.github.com/a/fViHl3QT
     - You will be prompted to select your cnetid from the list in order to link your Github account to your cnetid.
     - If you can't find your cnetid in the link above, click "continue to next step" and accept the assignment, then add your name, cnetid, and Github account to this Google Sheet and we will manually link it: https://rb.gy/9u7fb6
   - If you authenticated and linked your Github account to your device, you should be able to clone your PS3 assignment repository locally.
   - Contents of PS3 assignment repository:
     - `ps3_template.qmd`: this is the Quarto file with the template for the problem set. You will write your answers to the problem set here.

2. **Submission Process**:

   - Knit your completed solution `ps3.qmd` as a pdf `ps3.pdf`.
     - Your submission does not need runnable code. Instead, you will tell us either what code you ran or what output you got.
   - To submit, push `ps3.qmd` and `ps3.pdf` to your PS3 assignment repository. Confirm on Github.com that your work was successfully pushed.

Grading

- You will be graded on what was last pushed to your PS1 assignment repository before the assignment deadline
- Problem sets will be graded for completion as: {missing (0%); - (incomplete, 50%); (complete, 90%); + (excellent, 100%)}

    – The percent values assigned to each problem denote how long we estimate the problem will take as a share of total time spent on the problem set, not the points they are associated with.

- In order for your submission to be considered complete, you need to push both your `ps3.qmd` and `ps3.pdf` to your repository. Submissions that do not include both files will automatically receive 50% credit.

Background Recap

Read **this** article and **this** shorter article. If you are curious to learn more, **this** page has all of the articles that ProPublica has done on this topic. This problem set is a continuation of PS2 using the same data. Please start by loading the data in the same way as PS2.

Data cleaning continued

1. Some of the other articles on the propublica website discuss an increase in the dollar amount of the ticket for not having a city sticker. What was the old violation code and what is the new violation code?

2. How much was the cost of an initial offense under each code? (You can ignore the ticket for a missing city sticker on vehicles over 16,000 pounds.)

Revenue increase from "missing city sticker" tickets

1. Using pandas, create a new value for violation codes which combines the two codes that you found in the previous question. Again using pandas, collapse the data to capture the number of "missing city sticker" tickets by month. Then, using Altair, plot the number of tickets over time.

2. Suppose that your reader wants to be able to use the plot to deduce when the price increase occurred. Add frequent or custom date labels on the x-axis of your plot such that the date of the price increase is readily apparent. We haven't covered Altair's date labeling features in class so you'll first need to find the relevant help page in the documentation. Which help page did you use?

3. The City Clerk said the price increase would raise revenue by $16 million per year. For now, ignore the fact that many tickets are not paid and assume that the number of tickets issued is the same before and after the policy change. Using only the data available in the calendar year prior to the increase, how much of a revenue increase should they have projected? Remember that you are working with a one percent sample of the data.

   Assume that the number of tickets of this type issued afterward would be constant and you can assume that there are no late fees or collection fees, so a ticket is either paid at its face value or is never paid.

4. What happened to repayment rates (percentage of tickets issued that had payments made) on this type of ticket in the calendar year after the price increase went into effect? Suppose for a moment that the number of tickets issued was unchanged after the price increase. Using the new repayment rates in the year after the price increase occurred, what would the change in revenue have been?

5. Make a plot with the repayment rates on "missing city sticker" tickets and a vertical line at when the new policy was introduced. Interpret.

6. Suppose that the City Clerk were committed to getting more revenue from tickets. What three violation types would you as an analyst have recommended they increase the price of? Consider both the number of tickets issued for each violation type and the repayment rate for each violation type. You may assume there is no behavioral response to price changes (i.e., people continue to commit violations at the same rate and repay at the same rate). Make a plot to support your argument and explain in writing why it supports your argument.

Headlines and sub-messages

1. The City Clerk has now begun to wonder... maybe raising ticket prices will lead to a decline in repayment rates after all. Make a data frame where each row is a violation description, the fraction of time that the ticket is paid, and the average level 1 fine. Sort this dataframe based on how many total tickets of each type have been issued. Print the rows for the 5 most common violation descriptions.

2. Make a scatter plot which shows the relationship between fine amount and the fraction of tickets that are paid. Focus only on violations that appear at least 100 times. There will be one outlier with a high fine and you can exclude that ticket type from the plot. Then make two other plots which show the same relationship in different ways. For all three plots, write out what are the headlines and what are sub-messages.

3. The City Clerk doesn't understand regressions and only has time to look at one plot. Which plot are you going to bring to them and why?

Exploration vs. Production

1. Go back to your scatter plot on level 1 fine and fraction paid from the previous section. We want to add labels to each dot. Implement this in two ways:

   a. Label every dot with adjacent text or
   b. Put the text in a legend.

2. Either way, you will find the same problem – there are too many labels and the plot is illegible.

   a. Write a proposal for improvement in words.
   b. Implement your proposal for the improvement in code and make a new plot.