

HW 2

Denys Osmak

2024-01-25

Problem 1

Part A Create a histogram to display the overall data distribution of course evaluation scores.

```
#plot2()
```

Part B Use side-by-side boxplots to show the distribution of course evaluation scores by whether or not the professor is a native English speaker

Part C Use a faceted histogram with two rows to compare the distribution of course evaluation scores for male and female instructors.

Part D Create a scatterplot to visualize the extent to which there may be an association between the professor's physical attractiveness (x) and their course evaluations (y).

Problem 2

Part A a line graph showing average hourly bike rentals (total) across all hours of the day (hr).

Part B a faceted line graph showing average bike rentals by hour of the day, faceted according to whether it is a working day (workingday).

Part C a faceted bar plot showing average ridership (y) during the 9 AM hour by weather situation code (weathersit, x), faceted according to whether it is a working day or not. (Remember that you can focus on a specific subset of rows of a data set using filter.)

Problem 3

Part A

One faceted line graph that plots average boardings by hour of the day, day of week, and month. You should facet by day of week. Each facet should include three lines of average boardings (y) by hour of the day (x), one line for each month and distinguished by color. Give the figure an informative caption in which you explain what is shown in the figure and also address the following questions, citing evidence from the figure. Does the hour of peak boardings change from day to day, or is it broadly similar across days? Why do you think average boardings on Mondays in September look lower, compared to other days and months? Similarly, why do you think average boardings on Weds/Thurs/Fri in November look lower? (Hint: wrangle first, then plot.)

Part B

One faceted scatter plot showing boardings (y) vs. temperature (x), faceted by hour of the day, and with points colored in according to whether it is a weekday or weekend. Give the figure an informative caption in which you explain what is shown in the figure and also answer the following question, citing evidence from the figure. When we hold hour of day and weekend status constant, does temperature seem to have a noticeable effect on the number of UT students riding the bus?

Problem 4

Part A

Make a table of the top 10 most popular songs since 1958, as measured by the total number of weeks that a song spent on the Billboard Top 100. Note that these data end in week 22 of 2021, so the most popular songs of 2021 onwards will not have up-to-the-minute data; please send our apologies to The Weeknd.

Your table should have 10 rows and 3 columns: performer, song, and count, where count represents the number of weeks that song appeared in the Billboard Top 100. Make sure the entries are sorted in descending order of the count variable, so that the more popular songs appear at the top of the table. Give your table a short caption describing what is shown in the table.

(Note: you'll want to use both performer and song in any group_by operations, to account for the fact that multiple unique songs can share the same title.)

Part B

Is the “musical diversity” of the Billboard Top 100 changing over time? Let's find out. We'll measure the musical diversity of given year as the number of unique songs that appeared in the Billboard Top 100 that year. Make a line graph that plots this measure of musical diversity over the years. The x axis should show the year, while the y axis should show the number of unique songs appearing at any position on the Billboard Top 100 chart in any week that year. For this part, please filter the data set so that it excludes the years 1958 and 2021, since we do not have complete data on either of those years. Give the figure an informative caption in which you explain what is shown in the figure and comment on any interesting trends you see.

There are number of ways to accomplish the data wrangling here. We offer you two hints on two possibilities:

- 1) You could use two distinct sets of data-wrangling steps. The first set of steps would get you a table that counts the number of times that a given song appears on the Top 100 in a given year. The second set of steps operate on the result of the first set of steps; it would count the number of unique songs that appeared on the Top 100 in each year, irrespective of how many times it had appeared.
- 2) You could use a single set of data-wrangling steps that combines the length and unique commands.

Part C

Let's define a “ten-week hit” as a single song that appeared on the Billboard Top 100 for at least ten weeks. There are 19 artists in U.S. musical history since 1958 who have had at least 30 songs that were “ten-week hits.” Make a bar plot for these 19 artists, showing how many ten-week hits each one had in their musical career. Give the plot an informative caption in which you explain what is shown

Part D