

DS 521 Final Exam Practice

Part I: Adapted from a JMP Case Study: Film on the Rocks

The data set for Part I of the Final Exam Practice consists of question responses to a survey administered to 330 attendees at a film festival “Film on the Rocks”. You can **download it from Western Online/Assignments/Final Exam Practice**. You can **use 4 pages of notes** (either handwritten or typed) during the exam.

Background

Film on the Rocks is a summer movie series held at the world-renowned Red Rocks Amphitheatre, which is situated on a hillside of the Rocky Mountains in Morrison, Colorado fifteen miles west of Denver. The film series features classic films, and pre-show entertainment including bands and comedians. Among the features that have made Red Rocks an internationally-famous concert stage are sweeping views of Denver, outstanding acoustics, and cool, dry Colorado summers.

The series is jointly promoted by the Denver Film Society (DFS) and the City and County of Denver’s Division of Theatres and Arenas (DT&A). It is marketed through various outlets including newspaper, radio, and the Red Rocks and Denver Film Society websites. Film on the Rocks patrons also benefit from corporate sponsorship. In return for on-site posters and banners at Red Rocks, and recognition in preshow marketing materials, corporations donate funds that keep ticket prices low.

Although the Red Rocks Amphitheatre provides a cinematic experience unlike any other venue, there are tradeoffs. Red Rocks is a farther commute for most people than the local movie theater or movie rental store. Given the uphill walk to the amphitheater from the parking lot, getting there can be challenging. And, as an outdoor venue, the viewing experience is dependent on the weather.

Patron satisfaction with Red Rocks as the venue for the film series is critical to its success. But, the series promoters would also like to increase attendance at the film series, and are unsure how to do this. Promoters recognize that they need a better understanding of the customer base, and of the current level of satisfaction. Knowing the demographics of those who attend the film series will help attract and expand corporate sponsorship. In addition, knowing which media outlets are most effective will provide information about how best to target future marketing campaigns.

To this end, the promoters conducted surveys during a recent Film on the Rocks season. Questionnaires were handed out at the entrance. Volunteers walked through the crowd to remind people about the free soft drink given to those who returned the survey.

The Task

Use the survey results to address the following questions:

- What is the demographic profile of Film on the Rocks patrons?
- What is the level of customer satisfaction
 - with Signage
 - with Parking
 - with Cleanliness
 - Overall
- What factors are linked to satisfaction?
- In what media outlet(s) should the film series be advertised?

After creating appropriate visualizations, create a Tableau Story answering the questions above. Throughout your work, **implement best practices of data visualization** (e.g. adequate labeling, ordering, colorblind friendly palette, rounding of %, etc.)

Data

In addition to the typical variables of

- Resp ID
- Age Group: The patron's age group numerically coded: 1 = "1-12"; 2 = "13-30"; 3 = "31-60"; 4 = "60+"
- Gender: Male, Female.
- Income Group: The patron's annual household income group numerically coded: 1 = "Less than \$50,000"; 2 = "\$50,000-\$100,000", 3 = "100,000+"
- Marital Status: Married, Single

the data have been reshaped to facilitate the information visualization, thus creating the variables

- Question ID
- Question Grouping
- Question type
- Wording
- Value
- Label

These variables contain the information for the following questions:

- **Hear About:** The patron's response to this question: "How did you hear about Film on the Rocks?" Respondents could **check any of the following that applied:**
 - Television
 - Newspaper
 - Radio
 - Website
 - Word of mouth

- The survey also contained **four Likert-scaled** questions about **satisfaction**; each of the following questions is coded (**Note the REVERSED Likert scale and work accordingly!**):

1 = Excellent; 2 = Good; 3 = Average/Fair; 4 = Poor; 5 = Very Poor.

- Signage “How was the signage directing you to Red Rocks?”
- Parking “How was the venue’s parking?”
- Clean “How was the cleanliness of the venue?”
- Overall “How was your overall customer service experience?”

Problem #1: Before you begin any work make sure that when importing the data into Tableau, the **variables’ actual data type is correct** (make any **changes as necessary**).

Visualize each of the demographic variables

- Age Group
- Gender
- Income Group
- Marital Status

using a **sorted** bar chart. Make sure to create the first bar chart with almost all (if not all) desirable attributes, so that you can save time by duplicating the sheet and replacing the existing demographic variable with a new one. Your visualizations must conform to the following specifications:

- ✓ Round all percent to 0 digits after the decimal point.
 - ✓ Have the distinct counts displayed in addition to the bar labels in percent.
 - ✓ Display a prominent Axis ruler for rows.
 - ✓ Use a hidden reference line set at a suitable constant value in order to have the bar charts on the same scale.
 - ✓ Remove unnecessary headers and axis titles.
- Create a **Sample Size** sheet.
 - Make a Demographics dashboard using the bar charts and the Sample Size sheet created above.

Problem #2: Make a gap chart **given a breakdown by demographic information** for the check-all-that-apply question **Hear About**. Note that you have to first create a **parameter** called **Select Breakdown**, then a **calculation** called **Breakdown** with which you will tie the parameter to controlling the viz. Do not forget to show the parameter. Only **one** viz sheet is needed for problem 2.

Make sure that your viz conforms to the following specifications:

- ✓ Round all percent to 0 digits after the decimal point.
- ✓ Display labels only upon highlight.
- ✓ Use a reference line based on the overall percentage for the check-all-that-apply responses.
- ✓ Remove unnecessary headers and axis titles.

In addition, using **Breakdown**, make a sheet **Demographics Legend**, to be used to control the dashboard for the two check-all-that-apply questions.

In the end, make a **dashboard using the sheet for the check-all-that-apply question viz and the Demographics Legend** sheet in place of the color legends.

Problem #3: Visualize the **Likert Scale Questions** using a **divergent bar chart with neutrals separated** (make sure both the divergent bars and neutrals are on the **same scale**, and the negatives are labeled **without** the negative sign, rounded to 0 digits after the decimal point, remove unnecessary headers and titles). Recall that first you have to create the calculated fields for %Positive, % Negative, %Neutral.

Problem #4: Explore what demographic factors are linked to **satisfaction** by creating a **gap chart given a breakdown by demographic information** for a **new calculated** variable **%Top N Boxes** based on the **Likert scale** ratings.

Problem #5: Using the visualizations created above, **create a Tableau Story answering the questions** posed at the beginning of this case study, namely:

- What is the demographic profile of Film on the Rocks patrons?
- What is the level of customer satisfaction
 - with Signage
 - with Parking
 - with Cleanliness
 - Overall
- What factors are linked to satisfaction?
- In what media outlet(s) should the film series be advertised?

Upload your Tableau Packaged Workbook (.twbx) to the Final Exam Practice submission folder.

**Part II: Georgia 2000 Presidential Elections: Are Uncounted Votes Randomly Distributed Across Voters?
A Case Study based on data from the 2000 Presidential Election in Georgia.**

Adapted from Mary C. Meyer (2002), “Uncounted Votes: Does Voting Equipment Matter?”, *Chance*, Volume 15, No 4.

The data set **Georgia 2000 Presidential Election.xlsx** (posted on **WO** under **Assignments/Fianl Exam Practice**) contains the results from the 2000 presidential election in all 159 counties in the state of Georgia. The following table describes the available information.

Column #	Column Name	Description
1	County	Name of the county
2	equip	Type of voting method/equipment used in the corresponding county. There are five such methods. 1. Optical scan, precinct count (OS/PC) 2. Optical scan, central count (OS/CC) 3. Lever machine (LEVER) 4. Punch card (PUNCH) 5. Paper ballot (PAPER)
3	econ	Overall economic status of the corresponding county: poor, middle, rich.
4	Percent Afr. Am	Proportion of the population in the county that is African American.
5	rural	1 if a county is in a rural area, 0 if not.
6	Atlanta-metro	1 if a county is in the Atlanta Metro area, 0 if not.
7	gore	Number of votes counted for Gore.
8	bush	Number of votes counted for Bush
9	other	Number of votes counted for a candidate other than Gore or Bush.
10	votes	Number of votes counted.
11	ballots	Number of ballots cast in the county.
12	Atlanta/Urban not-Atlanta/Rural	0 if a county is in the Atlanta Metro area, 1 if a county is urban but not in the Atlanta Metro area, 2 if a county is in a rural area.

The main research questions we ask are what variables are adequate in predicting the percentage of uncounted votes and whether the type of voting equipment used alters the chance a vote is uncounted. Based on your analysis make a recommendation which voting method is best in terms of reducing the percentage of invalid ballots.

Note that there are only two counties using paper ballots. In this actual data there is an occurrence of Simpson’s paradox (if we look at the relationship between two variables, we see certain association, but when we

disaggregate the data according to the levels of a third variable, the nature of the association changes). **Hint:** The Simpson paradox occurrence here directly affects our conclusions regarding what voting method would be best to use in order to reduce the percentage of invalid ballots. You can discover the occurrence of Simpson's paradox by making side-by-side jitter plots, which are the best way to compare the distribution of a quantitative variable across levels of categorical variables. Since in this data set the confounding variable will have more than two categories, it wouldn't be useful to make the type of plot we studied at the beginning of the course. (however, the same method to compute weighted averages for the percent invalid ballots that will correct for the paradox).

For the purpose of this exercise, making EDA visualizations addressing the main research question and documenting that Simpson's paradox does occur is enough. You can also put a short story with your recommendations and supporting arguments for the adoption of specific types of voting equipment.

Upload your Tableau Packaged Workbook (.twbx) to the Final Exam Practice submission folder.