

Lab 7: Tableau and Categorical Data Analysis

KEY

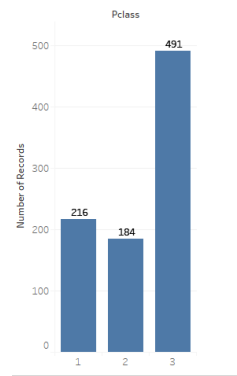
Javier E. Flores

April 10, 2019

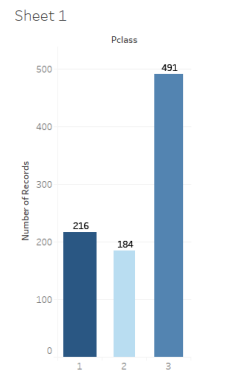
Total Possible Points: 38

Data Visualization with Tableau

Q1) [2 pts] In Tableau find a way to convert the variable "Pclass", which describes the class of each passenger, into a categorical variable. Create a bar chart displaying the number of passengers in each class. Be sure to add a label of counts above each bar indicating the exact amount of passengers represented in a given bar.

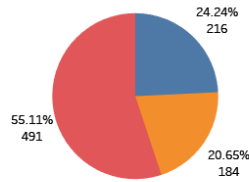


Q2) [3 pts] Drag and drop the variable "Survived" over the "Size" and "Color" options. Describe the change that this induced in your graph. From this graph, what can you say about survival across each class?

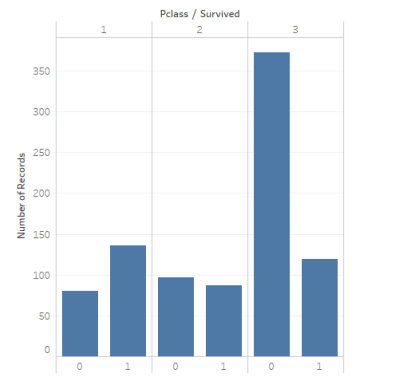


The width and color of the bars now vary with the count of individuals who survived. The lighter the color and thinner the bar, the fewer individuals that survived. This considered, second class passengers had the fewest amount of individuals who survived. However, this is misleading in that the third class passengers had the lowest *proportion* of survivors.

Q3) [3 pts] Create a pie chart for "Pclass". Label each section of the chart with the total counts and the corresponding percentage for each category.

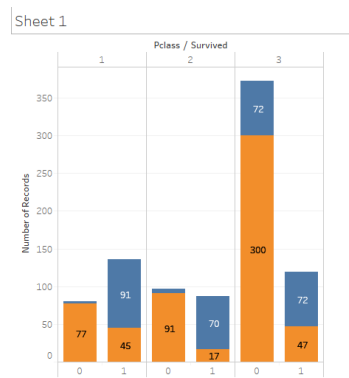


Q4) [3 pts] Create a clustered bar graph with "Pclass" as the outer variable and "Survived" as the inner variable. What do you observe about survival across each class?



This graph more clearly depicts survival across classes. The survival proportion is lowest in third class, and highest in first class. This agrees with expectation.

Q5) [4 pts] Create a clustered bar graph with "Pclass" as the outer variable and "Survived" as the inner variable. Within each displayed bar, display stacks corresponding to the amount of male and females. Add labels of the number of records corresponding to each stack.



Q6) [2 pts] Proportionally, which class had the most women? How does the proportion of women relate to the survival across each class? (I am not asking for you to perform a test, just what you might conclude).

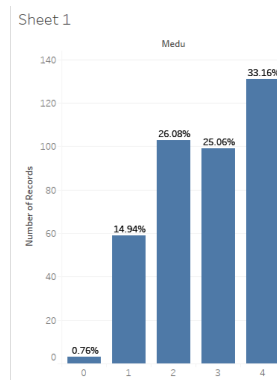
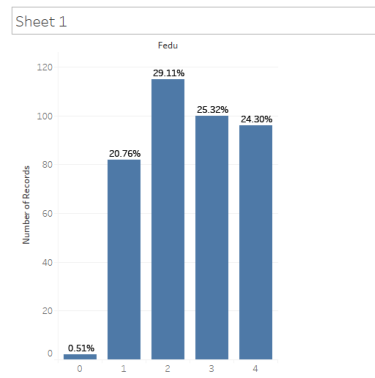
Proportionally, first class had the most women. Accordingly, first class had the greatest proportion surviving. Third class had the lowest proportion of women and also the lowest proportion surviving.

Q7) [9 pts] Using either the titanic or student alcohol data, explore three of the other data visualization options provided under the "Show Me" tab in Tableau. You may choose any variable(s) to use. Describe and interpret each of the three figures generated.

Answers will vary. Check that three different figures are generated, each with a description and interpretation

Q8) [6 pts] According to [Pordata](#), a base of certified statistics on Portugal, the 2005 national proportions of educational attainment among those older than 15 were 13.6% for no schooling, 31.2% for primary education, 32.4% for 5th to 9th grade, 13.4% for secondary education, and 9.4% for higher education. Using Tableau,

- Generate figures appropriate for visualizing the educational attainment levels of father's and mother's separately (from the student alcohol consumption data).
- Perform a statistical test to determine whether the distribution of educational attainment among fathers within our sample is consistent with national proportions. Be sure to state your null and alternative hypotheses, provide a test statistic, find a p-value, and interpret your results.
- Repeat the same analyses for the educational attainment of mothers.



The appropriate test is a χ^2 goodness of fit test. The null and alternative hypotheses (whether dealing with the mother's or father's educational attainment) are:

$$H_0 : p_{ns} = 0.136, p_{pe} = 0.312, p_{5to9} = 0.324, p_{se} = 0.134, p_{he} = 0.094$$

H_A : At least one proportion different from what is specified under the null

The test statistic corresponding to mother's educational attainment is $\chi^2 = 363.665$ and $\chi^2 = 200.11$ for father's educational attainment. Both are highly statistically significant with p-values near 0. The educational attainment proportions of mother's and father's in our sample is different from the national proportions. There proportion of individuals with higher education among parents of children attending these schools is higher than the national proportion.

Q9) [6 pts] Perform a statistical test to determine whether there is an association between a student's address type and their desire to pursue higher education. Be sure to state your null and alternative hypotheses, provide a test statistic, find a p-value, and interpret your results. Use Tableau to generate a figure which best communicates your findings. Describe and interpret the chosen figure.

Address	Higher	
	no	yes
R	6	82
U	14	293

Using the above table, we perform a χ^2 test for association. The null hypothesis is that rural status and the desire for higher education are not associated, and the alternative hypothesis is that they are. The value of the test statistic is $\chi^2 = 0.332$ and the p-value is 0.565. We fail to reject the null hypothesis and conclude that there is not enough evidence to disprove independence between rural status and the desire for higher education. Various figures could be used to communicate these findings, so several answers are possible. One example is provided below.

